Contents

What's New in Tableau Prep Builder ................................................................. 1

Added in version 2019.2.3 ................................................................. 1

Examine and Filter Your Data ................................................................. 1

Select specific values to filter ................................................................. 1

Duplicate fields in one click ................................................................. 2

Clean and Shape Data ........................................................................... 3

Fix invalid data role values with fuzzy matching and cleaning recommendations .... 3

Added in version 2019.2.2 ................................................................. 6

Connect to Data ...................................................................................... 6

Connect to more data with the Other Databases (ODBC) connector .......... 6

Specify an Initial SQL command when connecting to your data ................. 6

Examine and Filter Your Data ................................................................. 7

Keep only the fields you need while cleaning your data ......................... 7

Clean and Shape Data ........................................................................... 7

Rearrange steps in your flow ................................................................ 7

Added in version 2019.2.1 ................................................................. 9

Connect to Data ...................................................................................... 9

Connect to more data with new database connectors .............................. 9

Keep Flow Data Fresh ........................................................................... 9

View file safe list locations when publishing a flow .................................. 9

Additional enhancements ..................................................................... 10

Select from 10 different languages for your user interface ................... 10

Quickly find your flow on Tableau server with flow thumbnail views ........ 11

Added in version 2019.1.4 ................................................................. 12
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connect to Data</td>
<td>12</td>
</tr>
<tr>
<td>Connect to more data with new database connectors</td>
<td>12</td>
</tr>
<tr>
<td>Clean and Shape Data</td>
<td>12</td>
</tr>
<tr>
<td>Group similar values in a field by data role</td>
<td>12</td>
</tr>
<tr>
<td>Adjust your results when grouping field values</td>
<td>14</td>
</tr>
<tr>
<td>Added in version 2019.1.3</td>
<td>15</td>
</tr>
<tr>
<td>Connect to Data</td>
<td>15</td>
</tr>
<tr>
<td>Quickly join your data based on table relationships</td>
<td>15</td>
</tr>
<tr>
<td>Clean and Shape Data</td>
<td>16</td>
</tr>
<tr>
<td>See all cleaning recommendations in the toolbar</td>
<td>16</td>
</tr>
<tr>
<td>Other enhancements</td>
<td>17</td>
</tr>
<tr>
<td>Find specific changes in your flow</td>
<td>17</td>
</tr>
<tr>
<td>Manage recently viewed flows from the Start page</td>
<td>18</td>
</tr>
<tr>
<td>Added in version 2019.1.2</td>
<td>19</td>
</tr>
<tr>
<td>Keep Flow Data Fresh</td>
<td>19</td>
</tr>
<tr>
<td>Schedule flows to run automatically in Tableau Server</td>
<td>19</td>
</tr>
<tr>
<td>Help Design</td>
<td>19</td>
</tr>
<tr>
<td>Added in version 2019.1.1</td>
<td>20</td>
</tr>
<tr>
<td>Clean and Shape Data</td>
<td>20</td>
</tr>
<tr>
<td>Pivot rows to columns</td>
<td>20</td>
</tr>
<tr>
<td>Use wildcard search to pivot fields</td>
<td>20</td>
</tr>
<tr>
<td>Use descriptions to add context to changes</td>
<td>21</td>
</tr>
<tr>
<td>Use cleaning recommendations to quickly split data in fixed-width text files</td>
<td>22</td>
</tr>
<tr>
<td>Added in version 2018.3.3</td>
<td>24</td>
</tr>
<tr>
<td>Connect to Data</td>
<td>25</td>
</tr>
<tr>
<td>Automatically refresh input unions when new files are added</td>
<td>25</td>
</tr>
</tbody>
</table>
Clean and Shape Data ................................................................. 25
Reuse cleaning operations ......................................................... 25
Remove extra whitespace .......................................................... 26
Examine and Filter Your Data .................................................... 26
Quickly find where fields are used in a flow .................................. 26
Other Enhancements ................................................................. 27
Open multiple windows of Tableau Prep Builder at once ................. 27
Turn off usage reporting after install ......................................... 27
Added in version 2018.3.2 ........................................................... 28
Clean and Shape Data ................................................................. 28
Improve data quality fast using cleaning recommendations .............. 28
Other Enhancements ................................................................. 29
Added flexibility when running flows from the command line .......... 29
Performance improvements ....................................................... 30
Added in previous versions ....................................................... 30
Install and Deploy Tableau Prep .................................................. 31
Change your default repository location ...................................... 31
Connect to Data ........................................................................ 31
Clean data stored in PDF files ...................................................... 31
Wildcard union for database tables ............................................. 32
Clean and Shape Data ................................................................. 33
Replace field values with Null ...................................................... 33
Apply cleaning operations in any step ........................................... 34
Connect to Data ........................................................................ 36
Connect to data stored in MongoDB Business Intelligence (BI) ....... 36
Examine and Filter Your Data ..................................................... 36

Tableau Prep Help
Specify a data role for your field values .................................................. 36
Join or Union Data .................................................................................. 38
Fix mismatched fields directly in the join clause .................................... 38
Install and Deploy Tableau Prep ............................................................... 40
Set your display language ....................................................................... 40
Connect to Data ...................................................................................... 41
Connect to data stored in Microsoft Access ............................................ 41
Clean and Shape Data ............................................................................ 41
Change the color scheme for your flow steps .......................................... 41
Add descriptions to your steps ............................................................... 42
Use fuzzy match to find and fix spelling errors ....................................... 43
Save and share your work ..................................................................... 44
Run flows from the command line ......................................................... 44
Install and Deploy Tableau Prep ............................................................... 46
Deactivate Tableau Prep from the command line .................................... 46
Use virtual desktop support to optimize Tableau Prep Installations ......... 46
Connect to Data ...................................................................................... 46
Connect to cloud data sources and Hadoop Hive ................................... 46
Explore Your Data .................................................................................. 47
New filter options to keep only the data you want .................................. 47
Clean and Shape Data ............................................................................ 47
Use the ISO-8601 date standard in calculated fields .............................. 48
Apply cleaning operations in the data grid ............................................. 48
Use multi-select to group values in the Profile pane .............................. 49
Other enhancements ............................................................................. 49
Join or Union Data ................................................................................ 50
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use union recommendations to clean mismatched fields</td>
<td>50</td>
</tr>
<tr>
<td>Identify mismatched fields for all join types</td>
<td>52</td>
</tr>
<tr>
<td>Install and Deploy Tableau Prep</td>
<td>53</td>
</tr>
<tr>
<td>Activate and register Tableau Prep from the command line (Windows)</td>
<td>53</td>
</tr>
<tr>
<td>Connect to Data</td>
<td>53</td>
</tr>
<tr>
<td>Connect to data stored in statistical files or on Presto</td>
<td>54</td>
</tr>
<tr>
<td>Union sub-tables found by Data Interpreter in the Input step</td>
<td>54</td>
</tr>
<tr>
<td>Better feedback when loading tables</td>
<td>55</td>
</tr>
<tr>
<td>Explore Your Data</td>
<td>56</td>
</tr>
<tr>
<td>Reorder fields in the Profile pane and the Data grid</td>
<td>56</td>
</tr>
<tr>
<td>Clean and Shape Data</td>
<td>56</td>
</tr>
<tr>
<td>Use drag-select to remove multiple steps in your flow</td>
<td>56</td>
</tr>
<tr>
<td>Pivot multiple groups of fields in a single action</td>
<td>57</td>
</tr>
<tr>
<td>Improved field naming when merging fields</td>
<td>57</td>
</tr>
<tr>
<td>Other enhancements</td>
<td>58</td>
</tr>
<tr>
<td>Get Started with Tableau Prep Builder</td>
<td>59</td>
</tr>
<tr>
<td>In this article</td>
<td>60</td>
</tr>
<tr>
<td>Here's the story...</td>
<td>60</td>
</tr>
<tr>
<td>1. Connect to data</td>
<td>61</td>
</tr>
<tr>
<td>Check your work: Watch &quot;Connect to data&quot; in action.</td>
<td>64</td>
</tr>
<tr>
<td>2. Explore your data</td>
<td>65</td>
</tr>
<tr>
<td>3. Clean your data</td>
<td>67</td>
</tr>
<tr>
<td>Clean Orders_Central</td>
<td>67</td>
</tr>
<tr>
<td>Clean data with calculated fields</td>
<td>68</td>
</tr>
<tr>
<td>Interact directly with fields to clean your data</td>
<td>72</td>
</tr>
<tr>
<td>Review your changes</td>
<td>73</td>
</tr>
</tbody>
</table>
Connecting to SAP HANA (version 2019.2.1 and later) .............................................. 112
Connecting using ODBC (version 2019.2.2 and later) .................................................. 114
Before you connect ............................................................................................................. 114
Connect using Other Databases (ODBC) ......................................................................... 116
Use Initial SQL to query your connections (version 2019.2.2 and later) .................. 118
Working with Tableau data extracts .................................................................................. 118
Refresh data in the Input step ............................................................................................ 119
File input step types ........................................................................................................... 119
File, database or Tableau extract input step types ............................................................. 119
Use Data Interpreter to clean your files ............................................................................ 121
Union files and database tables in the Input step ............................................................... 122
Union files .......................................................................................................................... 123
Union database tables (version 2018.3.1 and later) .............................................................. 126
Merge fields after a union ..................................................................................................... 128
Join data in the Input step ..................................................................................................... 128
Configure your data set ........................................................................................................ 130
Connect to a custom SQL query ........................................................................................ 131
Select fields to include in the flow .................................................................................... 132
Configure field properties .................................................................................................. 132
Configure text settings in text files ................................................................................... 132
Change data types .............................................................................................................. 133
Change field names ............................................................................................................ 133
Set your data sample size ................................................................................................... 134
Apply filters to fields in the Input step .............................................................................. 135
Examine and Filter Your Data ............................................................................................ 136
See size details about your data .......................................................................................... 136
Review the data types assigned to your data .................................................. 138
Assign data roles to your data ........................................................................ 139
See the distribution of values or unique values ................................................ 142
Sort values and fields ...................................................................................... 144
Highlight fields and values in a flow ............................................................... 145
  Trace fields in a flow ....................................................................................... 145
See related values ............................................................................................. 146
Highlight identical values ................................................................................ 146
Keep or remove fields ....................................................................................... 147
Duplicate fields (version 2019.2.3 and later) .................................................. 148
Search for fields and values ............................................................................ 149
Filter values ...................................................................................................... 151
  Filters available for each data type ................................................................. 152
Use a Calculation filter ................................................................................... 152
Use a Selected Values filter (version 2019.2.3 and later) ................................. 153
Use a Range of Values filter ........................................................................... 154
Use Range of Dates filter .............................................................................. 154
Use a Relative Date filter ............................................................................... 155
Use a Wildcard Match filter .......................................................................... 155
Use a Null Values filter .................................................................................. 156
Clean and Shape Data ..................................................................................... 157
Build your flow ............................................................................................... 157
  Change the color scheme .............................................................................. 158
Add a description .............................................................................................. 159
Reorganize the layout of your flow (version 2019.2.2 and later) ...................... 160
  Watch "Reorganize flow steps" in action ......................................................... 161
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remove steps from the flow</td>
<td>162</td>
</tr>
<tr>
<td>Pivot your data</td>
<td>162</td>
</tr>
<tr>
<td>Pivot columns to rows:</td>
<td>163</td>
</tr>
<tr>
<td>Watch &quot;pivot on multiple field&quot; in action.</td>
<td>165</td>
</tr>
<tr>
<td>Use wildcard search to pivot</td>
<td>165</td>
</tr>
<tr>
<td>Pivot rows to columns</td>
<td>167</td>
</tr>
<tr>
<td>Apply cleaning operations</td>
<td>170</td>
</tr>
<tr>
<td>Apply cleaning operations in an input step</td>
<td>173</td>
</tr>
<tr>
<td>Apply cleaning operations in other steps</td>
<td>174</td>
</tr>
<tr>
<td>View your changes</td>
<td>176</td>
</tr>
<tr>
<td>Add a description to a change entry (version 2019.1.1 and later)</td>
<td>178</td>
</tr>
<tr>
<td>Apply cleaning operations using recommendations</td>
<td>181</td>
</tr>
<tr>
<td>Merge fields</td>
<td>183</td>
</tr>
<tr>
<td>Cleaning (fixing) variations of the same value</td>
<td>185</td>
</tr>
<tr>
<td>Edit a single value</td>
<td>185</td>
</tr>
<tr>
<td>Edit multiple values</td>
<td>186</td>
</tr>
<tr>
<td>Edit multiple values using quick cleaning operations</td>
<td>186</td>
</tr>
<tr>
<td>Group and edit multiple values inline</td>
<td>187</td>
</tr>
<tr>
<td>Replace one or more values with Null</td>
<td>188</td>
</tr>
<tr>
<td>Edit multiple values manually using Group and Replace</td>
<td>189</td>
</tr>
<tr>
<td>Edit multiple values using Group and Replace with fuzzy match</td>
<td>191</td>
</tr>
<tr>
<td>Group similar values using fuzzy match</td>
<td>192</td>
</tr>
<tr>
<td>Adjust your results when grouping field values</td>
<td>193</td>
</tr>
<tr>
<td>Group similar values by data role</td>
<td>195</td>
</tr>
<tr>
<td>Add and identify values that aren't in the data set</td>
<td>197</td>
</tr>
<tr>
<td>Aggregate and group values</td>
<td>199</td>
</tr>
</tbody>
</table>
Join or Union Data ................................................................. 201
Join your data ........................................................................ 201
Inspect the results of the join .................................................. 203
Common join issues ............................................................... 204
Fix mismatched fields and more .............................................. 205
Union your data ..................................................................... 206
Inspect the results of the union ............................................... 207
Fix fields that don’t match ...................................................... 209
Additional merge field options .............................................. 212
Save and Share Your Work .................................................... 213
Save a flow ........................................................................... 213
View your data sample in Tableau .......................................... 213
Create and publish data extracts and data sources .................. 214
Create an extract file ............................................................. 215
Publish as a data source ....................................................... 217
Refresh output files from the command line ......................... 218
Before running the flow ........................................................ 219
Tips for creating your credentials file .................................... 220
Run the flow ......................................................................... 221
Command options ................................................................ 223
Syntax examples .................................................................... 224
The flow connects to and publishes to local files .................... 224
The flow connects to databases and publishes to a server ......... 225
The flow publishes to a server and the credentials file is stored on a network share ...................................................... 225
Version Compatibility with Tableau Prep ............................... 226
Step 1 (New Install): Install Tableau Server with Tableau Prep Conductor

Who can do this?

Before you install

Install Tableau Server and enable Tableau Prep Conductor

Configure public gateway settings

Enable Tableau Prep Conductor

Verify Tableau Prep Conductor is enabled and running

Dedicate a node for Tableau Prep Conductor

Multi-node installations

Next step

Step 1 (Existing Install): Enable Tableau Prep Conductor

Who can do this?

Before you upgrade

Prepare for upgrade:

Configure public gateway settings

Tableau Server Installations using User-Based licenses

Tableau Server single-node installations

Tableau Server multi-node installations

Tableau Server Installations using Core-Based licenses

Tableau Server single-node installations

Tableau Server multi-node installations

Next step

Step 2 - Configure Flow Settings for your Tableau Server

Who can do this?
Publishing, Scheduling, and Credential Settings ........................................... 264
Implication of disabling Tableau Prep Conductor ........................................... 265
Configure notifications for flow failures ......................................................... 266
To enable the server-wide email notification .................................................... 266
Set notification values ................................................................................. 267
To configure email notification for a site: ....................................................... 268
Next step ............................................................................................... 268

Step 3- Create Schedules for Flow Tasks ......................................................... 268
Who can do this? .............................................................................. 268
Create a new schedule: ........................................................................ 269
Next step ........................................................................................... 270

Step 4 - Safe list Input and Output locations .................................................. 270
Who can do this? .......................................................................... 270
How to safe list input and output locations .................................................... 271
Next step ........................................................................................... 272

Step 5 - Optional Server Configurations ......................................................... 272
Who can do this? ........................................................................... 272
Set the timeout period for flows ................................................................. 272
Set the threshold for suspended flow tasks ................................................... 273

Publish a Flow to Tableau Server ................................................................. 273
Who can do this? ........................................................................... 273
Before you publish ....................................................................... 274
Publish a flow .................................................................................. 278

Schedule a Flow Task .............................................................................. 281
Who can do this? ........................................................................... 281
Schedule a flow task ..................................................................... 282
Manage a Flow .................................................................................................................. 283
Who can do this? .................................................................................................................. 283
Managing your flows .......................................................................................................... 290
Monitor Flow Health and Performance ............................................................................. 291
Detect issues as they occur and resolve them ................................................................. 291
Who can do this? .................................................................................................................. 291
Get email notifications when a flow fails: ........................................................................ 292
View and resolve errors ....................................................................................................... 292
Flow Overview page ........................................................................................................... 293
Connections page ............................................................................................................... 293
Scheduled Tasks page ........................................................................................................ 294
Run History page ................................................................................................................ 294
Alerts .................................................................................................................................. 295
Tableau Prep Conductor process status ............................................................................. 296
Monitor flow activity and performance with Administrative Views ................................ 296
Who can do this? .................................................................................................................. 297
Action by all users ............................................................................................................... 297
Action by Specific User ......................................................................................................... 297
Action by Recent Users ........................................................................................................ 298
Backgrounder Task Delays ................................................................................................. 299
Background Tasks for Non Extracts ................................................................................... 299
Performance of Flow Runs .................................................................................................. 300
Stats for Space Usage .......................................................................................................... 301
Developer Resources - REST APIs ...................................................................................... 302
Day in the Life Scenarios ..................................................................................................... 303
Hospital Bed Use with Tableau Prep .................................................................................. 303
LOD Driver Infractions ................................................................. 355
Troubleshoot Tableau Prep Builder .................................................. 359
Common errors when using the command line to run flows .................. 359
Error: "These features were found that prevent this version of the application from using this file" ................................................................. 363
Error: "Failed to parse response from Tableau Server" when publishing Tableau Prep output ............................................................................. 364
Error: "You are using Server version: null..." when signing in to an SSL-enabled Tableau Server using Tableau Prep ................................................ 365
Maintain Licenses for Tableau Desktop and Tableau Prep .......................... 365
View data about your license ................................................................. 366
Track Tableau Desktop license usage and expiration data ....................... 367
Additional resources ........................................................................... 368
What's New in Tableau Prep Builder

Browse summaries of new features for currently supported versions.

**Note:** Starting in version 2019.1.2, Tableau Prep is now called Tableau Prep Builder.

Added in version 2019.2.3

Examine and Filter Your Data

Select specific values to filter

With the the new **Selected Values** filter, you can now pick and choose the values that you want to keep or exclude for a field, even values that aren’t in your sample. In the Profile pane, Results pane or data grid click the **More options** button then select **Filter > Selected Values**.

![Selected Values Filter](image)

This new visual indicator lets you select the values that you want to keep or exclude and visually shows you what you've selected. If you are working with a sample of your data and want to include or exclude values that aren’t in the sample you are working with, simply add the values you need and Tableau Prep Builder applies the keep or exclude option to these new values too.
For more information about filtering field values, see Filter values on page 151.

Duplicate fields in one click

In prior releases if you wanted to experiment with your cleaning operations on a field but didn’t want to change the original data, you would have to write a calculation to duplicate the field. Now you can easily duplicate your fields in one click. Select the field you want to duplicate, then from the More options menu, select Duplicate Field.

A new field is created and placed in the far left of the Profile pane, data grid, or Results pane with the same name and a modifier. For example, "Ship Date -1".
For more information about duplicating fields and other cleaning options, see Duplicate fields (version 2019.2.3 and later) on page 148 Apply cleaning operations in other steps on page 174

Clean and Shape Data

Fix invalid data role values with fuzzy matching and cleaning recommendations

You can use data roles to quickly identify field values that aren’t valid. After you assign the data role you can use one of two fuzzy matching algorithms; Spelling and Spelling + Pronunciation to quickly group and match invalid values for the data role to similar standardized values for that data role.
In addition to suggesting data roles for your fields, Tableau Prep Builder can now also suggest how to automatically fix invalid values for fields assigned to a data role using the Spelling and Spelling + Pronunciation algorithms. If a suggested fix is detected, a light bulb icon appears on the profile card. Click the icon and apply the suggested change to automatically group and replace invalid values with similar standardized values for that data role.
This feature is only available for the following data roles:

- Airport
- City
- Country/Region
- County
- State/Province

For more information about applying data roles to your fields, see Assign data roles to your data on page 139. For more information about using group and replace to fix invalid values for data roles, see Group similar values by data role on page 195.
Added in version 2019.2.2

Connect to Data

Connect to more data with the Other Databases (ODBC) connector

If you need to connect to data sources that aren’t listed in the Connections pane, you can connect to any data source using the Other Databases (ODBC) connector that supports the SQL standard and implements the ODBC API. Connecting to data using the Other Databases (ODBC) connector works similarly to how you might use it in Tableau Desktop, however there are a few differences:

- You can only connect using the DSN (data source name) option.
- To publish and run your flow in Tableau Server, the server must be configured using a matching DSN.
- There is a single connection experience for both Windows and MacOS. Prompting for connection attributes for ODBC drivers (Windows) isn’t supported.
- Only 64-bit drivers are supported by Tableau Prep Builder.

For more information about connecting to data using the Other Databases ODBC connector, see Connect to Data on page 109.

Specify an Initial SQL command when connecting to your data

You can now specify an Initial SQL command, when connecting to databases that support it, that will run when a connection is made to the database. For example when connecting to Amazon Redshift, you can enter a SQL statement to apply a filter when connecting to the database just like adding filters in the Input step. The SQL command will apply before data is sampled and loaded into Tableau Prep Builder.

If you change the Initial SQL command and refresh the Input step, the Initial SQL command is run again and the Input step is updated. When you run the flow, the initial SQL command is run before processing all of the data and if you schedule flows on Tableau Server, the initial SQL is run every time as part of the data loading experience.

For more information, see Connect to Data on page 109.
Note: Parameters are not yet supported for Initial SQL in Tableau Prep Builder.

Examine and Filter Your Data

Keep only the fields you need while cleaning your data

In prior releases you could remove unwanted fields in the Input step by clearing the check boxes after connecting to your data. In a cleaning or action step you could select one or more fields and simply right-click to remove them. Now, you can select one or more fields in the Profile pane or data grid and right-click to keep only the selected fields and all other fields are removed.

For more information, see Keep or remove fields on page 147.

Clean and Shape Data

Rearrange steps in your flow

Have you ever built a large complicated flow, but the default layout that Tableau Prep Builder uses makes your flow hard to follow? Well now you can select and move your steps in the canvas to organize your flow in a way that makes sense to you. For example, you can fix crossed flow lines, move your flow steps to clean up extra white space, or rearrange your flow steps to show a clear sequence of events.

The following flow is confusing and hard to follow:
Using this new feature, you can select one or more steps in the flow and drag them up, down, left or right and drop them right where you want them.

Note: Some restrictions do apply and flow steps can’t be moved to a position that disrupts the left-to-right process flow. When dragging flow steps to an allowed location, an orange box displays. If the location isn't allowed, no orange box displays and the steps return to their original location when you try to drop them.

For more information, see Build your flow on page 157.
Added in version 2019.2.1

Connect to Data

Connect to more data with new database connectors

We’ve added more connectors to help you connect to and clean your data. Now you can connect to data stored in the following databases:

- Amazon Athena
- Other Databases (JDBC)

For more information about connecting to specific connectors, see Supported Connectors in the Tableau Desktop and Web Authoring help. For more information about connecting to data to build a flow see Connect to Data on page 109.

Keep Flow Data Fresh

View file safe list locations when publishing a flow

You can publish a flow to Tableau Server to automatically run your flows on a schedule and refresh the flow input or output using Tableau Prep Conductor. If your flow includes input or output steps with connections to a network share, your system administrator must include the file locations in your organization's safe list ahead of time, otherwise when you try to run the flow in Tableau Server, it will fail.

In Tableau Prep Builder version 2019.2.1 and later, if you publish a flow that includes input or output steps with connections to a network share, you are notified at publishing time if the files need to be included in your organization's safe list. To see safe listed locations, simply click the link in the warning message and move your files to one of the locations shown.
For more information about publishing flows, see Publish a Flow to Tableau Server on page 273. For more information about how to add file locations to your organization’s safe list, see Step 4 - Safe list Input and Output locations on page 270.

Additional enhancements

Select from 10 different languages for your user interface

Tableau has added two additional languages that you can choose from to configure your user interface (menus, messages, etc.):

- en_GB (British English)
- zh_TW (Traditional Chinese)
For more information about configuring your display language, see Set your display language in the Tableau Desktop and Tableau Prep Builder deployment guide.

Quickly find your flow on Tableau server with flow thumbnail views

When you publish your flows to Tableau Server, you can now switch to the thumbnail view and view a graphic representation of your flow layout to help you quickly spot your flow at a glance.
Added in version 2019.1.4

Connect to Data

Connect to more data with new database connectors

We've added even more connectors to help you connect to and clean your data. Now you can connect to data stored in the following databases:

- SAP HANA
- Azure SQL Data Warehouse
- MariaDB

**Note:** When connecting to data using SAP HANA, prompting for variables and parameters when opening a flow and using Initial SQL to query the connection isn't supported.

For information about connecting to specific connectors, see connection details in the Supported Connectors section of the Tableau Desktop and Web Authoring help. For more information about connecting to data to build a flow see Connect to Data on page 109.

Clean and Shape Data

Group similar values in a field by data role

In Tableau Prep Builder version 2018.2.3 we introduced data roles to help you standardize field values and quickly find values that weren't valid. Now if you assign a data role to your fields, you can use that data role to match and group values with the standard value defined by your data role. Simply assign a data role to your field, then select Data Role Matches from the Group and Replace menu. This new option then matches invalid values to the most similar valid value based on spelling and pronunciation.
For example if you have a field that lists countries such as Saint Martin and Saint Maarten, and you assign the **Country/Region** data role to the field, you can group fields using the new **Data Role Matches** option and map these values to the standardized data role value "Saint-Martin". If the standardized spelling of the value doesn't exist in your data set, Tableau Prep Builder automatically adds it to the group for you and groups the matching values using this standard value.

This feature applies to the following data roles and is currently only supported for U.S. English.

- Airport
- City
Country/Region
County
State/Province

For more information about assigning data roles to a field, see Assign data roles to your data on page 139. For more information about using data roles to group field values, see Edit multiple values using Group and Replace with fuzzy match on page 191.

Adjust your results when grouping field values

When you use the Group and Replace cleaning option to group similar values by Spelling or Pronunciation, you can now use a simple slider on the field to adjust how strict the grouping parameters are, giving you more control over your results.

Select from 5 threshold settings to control the number of values included in a group and the number of groups that get created.

For more information about grouping fields using fuzzy matching, see Cleaning (fixing) variations of the same value on page 185.
Added in version 2019.1.3

Connect to Data

 Quickly join your data based on table relationships

When you connect to databases that include tables where table relationships are defined, such as Vertica or Microsoft SQL Server, Tableau Prep Builder can now detect and show this information.

For example, you can now see which fields in a table are unique identifiers and which fields are related fields (foreign key) as well any related tables for these fields.

A new column called **Linked Keys** shows in the Input pane with an icon next to the fields that are unique identifiers, related fields or a combination of both.

![Linked Keys in Input Pane](image)

In the **Linked Keys** field, click on a related field or combination field to view related tables, then in one click, add the related table to your flow or create a join.

![Related Tables in Fields](image)

When creating a join, Tableau Prep Builder uses the defined field relationship to join your tables and shows a preview of the fields that define the join clause. From there you can click on the new Join step to clean and edit your join conditions as needed.
For tables that contain related fields, you can also see a list of all related tables and create a join from the menu in the Flow pane. From the **Add Join** menu option, select the table that you want to join on from the list and Tableau Prep Builder creates the join based on the fields that make up the relationship between the two tables.

This feature is available for any supported database connector where table relationships are defined. For more information about using **Linked Keys** to add or join tables, see [Join data in the Input step](#) on page 128.

**Clean and Shape Data**

**See all cleaning recommendations in the toolbar**

If Tableau Prep Builder detects and provides cleaning recommendations for your fields, a light bulb icon appears on the Profile card for the field. Now you can see all cleaning recommendations for any field from the toolbar in a selected step in one convenient list.
Simply click on an item in the list and Tableau Prep Builder takes you right to the field where you can apply the suggested change. For more information about cleaning recommendations, see *Apply cleaning operations using recommendations* on page 181.

**Note:** This option isn't available for Join, Input and Output step types.

### Other enhancements

#### Find specific changes in your flow

You can now quickly find specific changes that you made in a flow by clicking on the change icon on a step in the Flow pane or on a profile card in the Profile or Results pane. Simply click on the icon that represents the change you are looking for then select the annotation from the list of changes. The change is highlighted in the **Changes** pane, **Profile** or **Results** pane and data grid if applicable.
For more information about viewing your changes, see View your changes on page 176.

Manage recently viewed flows from the Start page

When you save a flow and close Tableau Prep Builder, a thumbnail for the flow is saved to your Start page so you can quickly find your most used flows. But as a data prep rock star, your Start page can get crowded fast! To help you quickly find the flow you need and reduce the clutter when you’re finished working with a flow, we’ve added a few new features to the Start page.

- Improved graphics on the flow thumbnail more accurately represents the flow layout to help you spot your flow at a glance.
- If the graphic doesn’t give you enough information, hover over a flow thumbnail to instantly view the input and output details for the flow as well as when it was last opened.
- Clear the clutter by removing recently viewed flows from the start page that you no longer need. This option removes the flow from the Recent Flows view only and won't delete the flow.
Added in version 2019.1.2

Keep Flow Data Fresh

Schedule flows to run automatically in Tableau Server

Add the new Data Management Add-on to Tableau Server version 2019.1 and later and you can use Tableau Prep Conductor to schedule and manage your flow runs. Publish a flow to Tableau Server, then schedule tasks to run the flow and automatically update your flow inputs and outputs. Use the error and alert options and Administrative views to track and manage your flow processing health.

For more information, see Keep Flow Data Fresh on page 233.

Help Design

Updates have been made to help you navigate our content, including a new right-hand menu to quickly scan and select different topics on a page.
Added in version 2019.1.1

Clean and Shape Data

Pivot rows to columns

If you work with data in a cross-tab or spreadsheet it might not be in the format you need to easily analyze it in Tableau. Sometimes your data may have many columns of data that need to be converted into rows, or you may have fewer columns but the data values represent multiple things.

To make shaping your data easier than ever, you can now choose the pivot option that is right for you. In a pivot step, you can now select one of the following options:

- **Columns to Rows**: Drag one or more fields to the **Pivoted Fields** pane to pivot columns to rows.

- **Rows to Columns**: Drag a field to the **Pivoted Fields** pane to create new columns using the field values, then drag in the field to use to populate those columns and choose how you want to aggregate them.

For more information, see **Pivot your data** on page 162
Use wildcard search to pivot fields

If you have a data set with a large number of fields or if fields are frequently added or removed over time, you can use a wildcard search when pivoting columns to rows to instantly pivot your data based on a wildcard pattern match. If new fields are added or removed that match the pattern, Tableau Prep Builder detects the schema change when you run the flow and the pivot results are automatically updated.

To use wildcard search to pivot your data, add a pivot step then click the **Click here to create wildcard pivot** link to open the wildcard search field. Enter a wildcard search pattern to find fields with names that match the values you entered. Click the **Search Options** ... button to select whether you want to include fields that contain, start with or end with the pattern you entered.

![Tableau Prep Builder screenshot](image)

Click enter to apply the pattern and Tableau Prep Builder automatically pivots the fields.
For more information, see **Pivot your data** on page 162

**Use descriptions to add context to changes**

Add descriptions to any change in the **Changes** pane to add context to your cleaning operations. Simply open the **Changes** pane, right-click or Ctrl-click (Mac) on a change in the list and select **Add Description**. Then type a description in the text field that appears.
The description appears below the generated text for the change with a comment icon.
For more information see *Clean and Shape Data* on page 157.

Use cleaning recommendations to quickly split data in fixed-width text files

In prior releases if you work with data stored in fixed-width text files, files that contain strings of data values in a single column constrained by a fixed character limit, you might have had to perform some pre-cleaning on the file or use complex calculations to separate out the values into separate columns.

Now when you connect to this type of data source, the cleaning recommendations feature detects these types of columns and suggests split options for you. Simply connect to your data source then add a cleaning step. In the Profile pane, click on the recommendation icon on the profile card.

Click **Apply** to split the fields, then check the results and clean or rename each new field as needed.
For more information, see Apply cleaning operations using recommendations on page 181.

Added in version 2018.3.3

Connect to Data

Automatically refresh input unions when new files are added

If you use a wildcard pattern to find and union Microsoft Excel or text (.csv) files in the input step, new files that are added to the same folder that match the pattern are automatically included in the union the next time you open the flow or run it from the command line.

**Note:** Packaged flow files (.tflx) won’t automatically pick up new files because the files are already packaged with the flow. To include new files, open the flow file (.tfl) to pick up the new files then repackage the flow to include the new file data.

For more information about unioning files in the Input step, see Union files and database tables in the Input step on page 122.

Clean and Shape Data

Reuse cleaning operations

When cleaning your data you often perform the same cleaning operations or actions over and over throughout your flow. To help make cleaning and shaping your data more efficient, we’ve
added the ability to copy and paste these operations or actions throughout your flow so you can perform a cleaning operation or action once, then reuse it where you need it.

- Copy one or more steps to use them in another area of your flow. This option is not available for Input steps that include a wildcard union.

- Copy an operation from the Changes pane in one step and paste it in the Changes pane for another step to apply that same operation in that step.

- Drag and drop an operation from the Changes pane to a field in the Profile pane to apply that operation to a different field. You can also copy an operation in the Changes pane and drag and drop it to other fields in the step to apply that operation to multiple fields. This option is not available for operations that impact multiple fields, such as calculated fields.

For more information, see Clean and Shape Data on page 157.

Remove extra whitespace

We've added two new quick cleaning operations to help you remove extra whitespace in your field values:

- **Remove Extra Spaces**: Use this option with text fields values to remove leading and trailing whitespace and replace extra whitespace in-between characters with a single space.

- **Remove All Spaces**: Use this option to remove all whitespace from text values, including leading and trailing whitespace and any whitespace in between characters.

For more information, see Cleaning (fixing) variations of the same value on page 185.

Examine and Filter Your Data

Quickly find where fields are used in a flow

When building your flow, if you aren't seeing all the field values that you expect, you can now easily trace everywhere a field was used - even back to where a field originated. Just select a field in the Profile pane or Results pane in a step and the steps in the flow pane will highlight. This shows you the path that field has taken in your flow, making troubleshooting your flow easier than ever.
For more information, see Trace fields in a flow on page 145.

Other Enhancements

Open multiple windows of Tableau Prep Builder at once

In Windows or on the Mac click File > New to open a new window for a flow instead of replacing the current window you are working in. When opening a flow file (.tfl or .tflx) this also opens a new window so you can easily work on or compare multiple flows at the same time.

Turn off usage reporting after install

To help us improve Tableau Prep Builder, you can share usage pattern data with us. We never collect any specific user data, only anonymous usage data. For example data source connection types (such as .csv, sql or Amazon Redshift), large data sets versus small data sets, number of steps in a flow or types of operations used (but no data specifics such as step names or formulas used), and so on.

Your data is confidential and we will never share this information with anyone. But if you want to turn this option off after installation, you can. Go to Help > Settings and Performance and click on Share Usage Data To Improve Tableau Prep Builder to clear the check mark. For more information, see Usage reporting in the Tableau Desktop and Tableau Prep Builder Deployment guide.
Added in version 2018.3.2

Clean and Shape Data

Improve data quality fast using cleaning recommendations

To help you find and fix problems with your data fast, Tableau Prep can now examine your data fields and suggest changes you can make to fix the problem. Tableau Prep examines your data fields and when a recommended change is identified, a light bulb icon appears on the Profile card for the field.

Click the light bulb icon to see the recommended changes, then hover over the recommendations card and click Apply to automatically apply them to the field. The change is recorded in the Changes pane and can be removed at any time.

In the example below Tableau Prep identified the field values as URL's and recommends applying a data role to the field to help identify values that aren't valid.
This feature is available in all step types except Input, Output and Join step types. In this release, recommendation types include:

- Data roles
- Filter
- Replace invalid values with Null values
- Remove fields
- Trim spaces

For more information, see *Apply cleaning operations using recommendations* on page 181.

**Other Enhancements**

**Added flexibility when running flows from the command line**

When running flows from the command line, if you connect to databases or publish output files to a server, you need a credentials .json file that includes all required credentials. In past releases if you only connected to databases or published the flow output to a server, you would still need to include the empty array in the credentials .json file or the process would fail.

Now you no longer need to include the empty array in the credentials .json file. For example, if you only connect to databases you no longer need to include the outputConnection array in the .json file.
For more information about running flows from the command line, see Refresh output files from the command line on page 218.

Performance improvements

To help make working with your flows faster and more efficient, we've made the following enhancements:

- Shutting down the product returns the memory space Tableau Prep was using right away.
- If you receive an unexpected error you can now retry your action or process to keep you in the flow.
- Improved performance when loading data or viewing data in a flow to reduce processing time.
- If you connect to the following connectors, you can now use a search field to quickly find your schema:
  - Amazon EMR Hadoop Hive
  - Cloudera Hadoop
  - Hortonworks Hadoop Hive
  - MapR Hadoop Hive
  - Presto
  - Spark SQL

Added in previous versions

Added in version 2018.3.1

Install and Deploy Tableau Prep

- Change your default repository location on the facing page

Connect to Data
Install and Deploy Tableau Prep

Change your default repository location

By default, when you install Tableau Prep, the My Tableau Prep Repository is automatically installed in your Documents folder on your computer. But if you have restrictions about where you can store your flows and data sources, you can now move your My Tableau Repository to a different directory after Tableau Prep is installed.

To move your directory, open Tableau Prep, then from the top menu, select File > Repository Location. Select the folder where you want to move the repository.

For more information, see Change your My Tableau Prep Repository location in the Tableau Desktop and Tableau Prep Deployment Guide.

Connect to Data

Clean data stored in PDF files

Analyzing data stored in PDF files can be difficult to work with. Now you can connect directly to your PDF files and clean your data using Tableau Prep.
For more information about how to connect to your data in PDF files, see PDF file in the Tableau Desktop help.

**Note:** Data connectors are not backward compatible. Flows that include these connectors may open in a prior version of Tableau Prep, but will have errors or can't run unless the data connections are removed.

### Wildcard union for database tables

If you work with database tables stored in a single data source you can now union your tables in the input step using a wildcard search. You can then include all of the data from the selected tables in the Input step using the same process that you use to union multiple files in the Input step.

To use this type of union, the tables must be in the same data source and you must connect to a data source that supports wildcard union.

The following data sources support this type of union:

- Amazon Redshift
- Microsoft SQL Server
- MySQL
- Oracle
- PostgreSQL
For more information, see Union files and database tables in the Input step on page 122.

Clean and Shape Data

Replace field values with Null

If you have data rows that you want to include in your analysis but you want to exclude certain field values you can change them to a Null value. Simply select the field values that you want to replace, then right-click (Ctrl+click on Mac) on the selected value and select Replace with Null from the menu.
For more information about replacing field values, see Cleaning (fixing) variations of the same value on page 185.

Apply cleaning operations in any step

Fixing problems in your flow just got more efficient. Now if you spot problems in your data as you join, union, pivot or aggregate, you can fix these problems directly in that step without having to add an additional cleaning step.

Since the order of operations matters, the cleaning operation is placed either before or after the join, union, pivot or aggregate and this information is tracked in the Changes pane for that step.

Annotations for the change are also added to the step in the Flow pane. For Input steps, annotations show to the left of the step and are added to the Input field list so you can easily
see which field had changes applied. For all other step types, the annotations show above the step in the Flow pane.

Some cleaning operations must still be performed in a cleaning step and only some cleaning operations are available for Input and Output steps. The table below lists the cleaning operations available for each step type:

<table>
<thead>
<tr>
<th></th>
<th>Input</th>
<th>Clean</th>
<th>Aggregate</th>
<th>Pivot</th>
<th>Join</th>
<th>Union</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Group and Replace</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Clean</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Split Values</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Rename Field</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Edit Value</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Change Data Type</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Create Calculated Field</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Remove Field</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

For more information, see Apply cleaning operations on page 170.

Added in version 2018.2.3
Connect to Data

Connect to data stored in MongoDB Business Intelligence (BI) below

Examine and Filter Your Data

Specify a data role for your field values below

Join or Union Data

Fix mismatched fields directly in the join clause on page 38

Connect to Data

Connect to data stored in MongoDB Business Intelligence (BI) below

If you store your data in MongoDB Business Intelligence, you can now connect to your data and clean it with Tableau Prep.

For more information about how to connect to your data using MongoDB Business Intelligence, see MongoDB BI Connector in the Tableau Desktop help.

Note: Data connectors are not backward compatible. Flows that include these connectors may open in a prior version of Tableau Prep, but will have errors or can't run unless the data connections are removed.

Examine and Filter Your Data

Specify a data role for your field values

You can now assign a data role to your field values and use Tableau Prep to help you find inaccuracies or outliers in your data set.

Data roles tell Tableau Prep what the field values mean or represent, for example email addresses or a geographic role such as city or zip code. When a data role is assigned to a field, Tableau Prep examines the field values and flags the values that don't match so that you can take a closer look.
To view only the values that are valid or not valid, use the new filter option on the drop-down menu:
In this release we support the following data roles:

- Email
- URL
- Geographic roles (Based on current geographic data and is the same data used by Tableau Desktop)
  - Airport
  - Area code (U.S.)
  - CBSA/MSA
  - City
  - Congressional District (U.S.)
  - Country/Region
  - County
  - NUTS Europe
  - State/Province
  - Zip code/Postal code

For more information see Assign data roles to your data on page 139.

Join or Union Data

Fix mismatched fields directly in the join clause

When you join two tables of data you will often have field values that are the same but are mismatched due to data entry errors. Tableau Prep helps identify mismatched fields in your join clauses by turning the mismatched field values red. But wouldn't it be great if you could just fix those field values right in your join clause? Well now you can.

Now when you join two data sets you can edit the field values that you join on, right in the join clause to fix them. Simply double-click in a field value to edit it.

Select field to change  Edit in-line
The cleaning action is automatically pushed back to the previous cleaning step in the flow for the appropriate join data set.

No cleaning step before the join step? No problem. Tableau Prep automatically adds one for you to capture the cleaning operation from the join.
For information about how to clean field values directly in a join clause, see Join or Union Data on page 201.

Added in version 2018.2.2

Install and Deploy Tableau Prep

- Set your display language below

Connect to Data

- Connect to data stored in Microsoft Access on the facing page

Clean and Shape Data

- Change the color scheme for your flow steps on the facing page
- Add descriptions to your steps on page 42
- Use fuzzy match to find and fix spelling errors on page 43

Save and Share your work

- Run flows from the command line on page 44

Install and Deploy Tableau Prep

Set your display language

When you start up Tableau Prep, it now detects the locale set on your computer and displays the user interface, dates, and number formats in the appropriate language. If you want to change the display language for the user interface you can select from the supported languages from the top menu under Help > Choose Language.
For more information see Set your display language in the Tableau Desktop and Tableau Prep deployment guide.

Connect to Data

Connect to data stored in Microsoft Access

If you use Microsoft Access for data entry or to store your data tables you can now connect to your access files (from version 2007 or higher) through a file browser and clean your data with Tableau Prep.

This connector requires a 64-bit driver.

For more information about how to connect Tableau Prep to your data, see Access in the Tableau Desktop help.

Note: Data connectors are not backward compatible. Flows that include these connectors may open in a prior version of Tableau Prep, but will have errors or can’t run unless the data connections are removed.

Clean and Shape Data

Change the color scheme for your flow steps

By default, Tableau Prep assigns each step in your flow a color to help you easily track the changes you make to your data as you build your flow. But you have choices when it comes to this color scheme.
You can now pick from a color palette to change the color scheme for one or more steps. Just select the steps in the Flow pane that you want to change, right-click the selected steps and select **Edit Step Color** from the context menu.

Don't like how it looks? Click **Undo** from the top menu or select **Reset Step Color** from the color palette menu.

For more information, see **Build your flow** on page 157.

**Add descriptions to your steps**

If you share your flows with others, communicating the changes that you made and why can be cumbersome. To make it easier, you can now add a short description to any individual step in your flow and it displays right in the flow pane.
For more information about adding descriptions to flow steps, see *Build your flow* on page 157.

**Use fuzzy match to find and fix spelling errors**

To help you quickly identify and clean up multiple variations of the same value due to spelling errors, we’ve added another option to our fuzzy match cleaning feature. Use the new *Spelling* option to find and group text values that differ because of additional or missing letters. This option works in any supported language.

The *Spelling* option uses the Levenshtein distance algorithm to compute an edit distance between two text values and then groups them together when the edit distance is less than a default threshold value.

For more information, see *Cleaning (fixing) variations of the same value* on page 185.
Save and share your work

Run flows from the command line

To keep data fresh in Tableau Prep you run your flow. However, opening Tableau Prep every time you want to run flows can take time. To help streamline this process, you can now run flows from the command line without having to open it in Tableau Prep.

To run a flow from the command line, you'll need:

- The path to the flow (.tfl) file that you want to run.
- A .json file that contains the database credentials for any databases that the flow connects to for its input steps and the credentials for the server where the output is published.

If your flow connects to or publishes to local files or files that are stored on a network share for inputs or outputs, then this file isn't needed.

**Note:** Connecting to or publishing files that are stored on a network share that are password protected isn't supported.

- Administrator permissions on the machine where you are running the flow.

This option is available on both Windows (Task Scheduler is supported) and Mac machines. To use this process, you need an activated version of Tableau Prep and the process must be run on the same machine where Prep is installed.
For information about how to run flows from the command line, see **Refresh output files from the command line** on page 218.

**Added in version 2018.2.1**

<table>
<thead>
<tr>
<th>Install and Deploy Tableau Prep</th>
</tr>
</thead>
<tbody>
<tr>
<td>• <strong>Deactivate Tableau Prep from the command line</strong> on the next page</td>
</tr>
<tr>
<td>• <strong>Use virtual desktop support to optimize Tableau Prep Installations</strong> on the next page</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Connect to Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>• <strong>Connect to cloud data sources and Hadoop Hive</strong> on the next page</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Explore Your Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>• <strong>New filter options to keep only the data you want</strong> on page 47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clean and Shape Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>• <strong>Use the ISO-8601 date standard in calculated fields</strong> on page 48</td>
</tr>
<tr>
<td>• <strong>Apply cleaning operations in the data grid</strong> on page 48</td>
</tr>
<tr>
<td>• <strong>Use multi-select to group values in the Profile pane</strong> on page 49</td>
</tr>
<tr>
<td>• <strong>Other enhancements</strong> on page 49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Join or Union Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>• <strong>Use union recommendations to clean mismatched fields</strong> on page 50</td>
</tr>
<tr>
<td>• <strong>Identify mismatched fields for all join types</strong> on page 52</td>
</tr>
</tbody>
</table>
Install and Deploy Tableau Prep

Deactivate Tableau Prep from the command line
Like Tableau Desktop, if you no longer need Tableau Prep on your computer you can now deactivate it from the command line using the -return option.

For more information see Deactivate the product key. For more information about other installer properties that are available for Tableau Prep from the command line, see the Installer options and relevant sections in Deploy Tableau Desktop.

Use virtual desktop support to optimize Tableau Prep Installations
Just like Tableau Desktop you can now configure virtual desktop support to optimize your installations of Tableau Prep for non-persistent virtual desktops or for computers that are regularly reimaged. With virtual desktop support, Tableau Prep licenses are automatically deactivated after a predetermined period of time using a Tableau-hosted “Authorization to Run” (ATR) service, eliminating the need to manually deactivate the product key.

For more information about how to configure this option, see Configure Virtual Desktop Support in the Tableau Desktop and Tableau Prep deployment guide.

Connect to Data

Connect to cloud data sources and Hadoop Hive
We’ve added support for the following connectors so that you can connect to cloud data and data stored in Hadoop.

- Snowflake.
- Amazon EMR Hadoop Hive
- Cloudera Hadoop (Hive and Impala)
- Hortonworks Hadoop Hive
- MapR Hadoop Hive
Tableau Prep Help

- Apache Drill
- SparkSQL

For more information about how to connect Tableau Prep to your data, see the topic for your connector under Supported Connectors in the Tableau Desktop help.

These data connectors are not backward compatible. Flows that include these connectors may open in a prior version of Tableau Prep, but will have errors or can't run unless the data connections are removed.

Explore Your Data

New filter options to keep only the data you want

No more writing complex calculations to keep or exclude Null values or to set up a wildcard match on text values. Instead select one of the new filter options on a field to see the impact of your changes right away without having to first write a calculation and then revert your change if it doesn't give you the results you want.

For more information, see Filter values on page 151.

Clean and Shape Data
Use the ISO-8601 date standard in calculated fields

Creating calculated fields to support European calendars just got easier. Tableau Prep now supports the ISO-8601 international date standard for the following date parts:

- "iso-year"
- "iso-quarter"
- "iso-week"
- "iso-weekday"

Use these date parts in functions DATEPART, DATETRUNC, DATENAME, DATEDIFF, and DATEADD.

For example: Week Number = STR(DATEPART('iso-year', [Week Date])) + "-" + STR(DATEPART('iso-week', [Week Date]))

For more information about how to work with date functions in calculated fields, see Date Functions in the Tableau Desktop help.

Apply cleaning operations in the data grid

In prior versions, the data grid showed you a preview of your data, but had few cleaning options available. Now you can act on your data anywhere. If you want to work with the detailed values in the data grid, collapse the Profile pane and perform the same cleaning operations that are available in the Profile pane in the data grid.
For more information, see Clean and Shape Data on page 157.

Use multi-select to group values in the Profile pane

To quickly group a set of values for a field, you can now multi-select the values in the Profile card, then right-click to open the menu and select Group. The values are grouped under the field value that you select when you right-click to open the menu. A paperclip icon shows next to the grouped value.

Right-click the grouped values to open the menu again to ungroup or edit the values.

For more information about grouping values, see Cleaning (fixing) variations of the same value on page 185

Other enhancements

We've also made the following enhancements to improve usability:

- New icons show on the menu when adding steps to your flow to provide visual cues and help you learn the visual language of Tableau Prep.
New animations in the Profile pane help catch your eye to better see the impact of your changes.

Join or Union Data

Use union recommendations to clean mismatched fields

Fixing mismatched fields after unioning two or more files just got easier. A new **Mismatched Fields** section in the **Union summary** pane shows a list of fields that don't match and the data source where they originated. Merge your mismatched fields directly in this section using one of the following options:

- Select a field in the list. If Tableau Prep identifies a field with similar characteristics, based on similar data types and field names, it highlights the field in yellow, suggesting a match. Click the plus button that appears on hover on the suggested matched field to merge the two fields.
Select two or more fields in the list, right-click on one of them and select **Merge Fields**.

For more information about unioning data and resolving mismatched fields, see **Union your data** on page 206.
Identify mismatched fields for all join types

Easily identify fields that don't match no matter how you join tables. In prior versions the Join Clauses tables showed field values that were excluded only when one field equaled another. But now you can see this data using any operator when matching join clauses, such as "End Date >= Modified Date".

For more information, see Join your data on page 201.

Added in version 2018.1.2

Install and Deploy Tableau Prep

- Activate and register Tableau Prep from the command line (Windows) on the facing page

Connect to Data

- Connect to data stored in statistical files or on Presto on page 54
Install and Deploy Tableau Prep

Activate and register Tableau Prep from the command line (Windows)

Like Tableau Desktop, you can now activate and register Tableau Prep from the command line by including a command line with the following properties:

- ACTIVATE_KEY="<key>". The installer runs -activate to apply the product key.
- REGISTER="1". During the installation process, the installer will run the -register process and add the registration information.

For more information about these installer properties and how to activate and register Tableau Prep from the command line, see the Installer options and relevant sections in Deploy Tableau Desktop.

Connect to Data
Connect to data stored in statistical files or on Presto

We've added two new connectors to help you connect to data from more locations.

- Statistical files. Connect to to SAS (*.sas7bdat), SPSS (*.sav), and R (*.rdata) data files.
- Presto. For more information about how to configure your connection to Presto, see [Presto](#) in the Tableau Desktop help.

These data connectors are not backward compatible. Flows that include these connectors will open in a prior version of Tableau Prep, but will have errors or can't run unless the data connections are removed.

Union sub-tables found by Data Interpreter in the Input step

Using Data Interpreter to clean your Microsoft Excel data and now you want to union the resulting sub-tables? You can now use wildcard union to union all of the found sub-tables in the Input step.

Simply drag one of your sub-tables to the Flow pane, and then use the wildcard search criteria to select the remaining sub-tables to union the data and include all the sub-table data in the Input step.

For more information see [Union files and database tables in the Input step](#) on page 122.
Better feedback when loading tables

When you connect to a database for the first time, it can sometimes seem like nothing happened. Now a new indicator tells you that the data is still loading.
Explore Your Data

Reorder fields in the Profile pane and the Data grid

You can now drag and drop to reorder your fields in both the Profile pane and the Data grid and the two panes will stay in sync. We also maintain the field order even when you rename a field. Fields are no longer reordered automatically.

Clean and Shape Data

Use drag-select to remove multiple steps in your flow

Removing whole sections of your flow just got easier. You can now click in the Flow pane and use your mouse to drag and select the section of the flow that you want to remove. Then right-click to remove all of the selected steps at once. For more information see Build your flow on page 157.
Pivot multiple groups of fields in a single action

The pivot option in Tableau Prep has been expanded to better support more complex spreadsheets or text files. You can now perform either a single pivot or select groups of fields to pivot.

For example to pivot a spreadsheet to see sales, tax, and totals by month and year you can pivot each group of fields to get the results you want. For more information, see Pivot your data on page 162.

Improved field naming when merging fields

When you merge fields, the field names are no longer concatenated. Instead the field name of the target field persists. For more information, see Merge fields on page 183.
Other enhancements

We’ve also made the following enhancements to improve usability and performance:

- A new edit icon shows on hover on entries in the **Changes** pane to help you quickly see which items are editable.

  ![Rename Field]

  Rename Field
  [Discount]
  From [Discount] to [Discount]

- Renaming a step will no longer automatically run the flow so your authoring experience is seamless.

- The **Connections** pane automatically opens when you connect to a Microsoft Excel file with multiple sheets to help you quickly get to your data.
Get Started with Tableau Prep Builder

Note: Tableau Prep version 2019.1.2 had changed its name to Tableau Prep Builder. If you are using an earlier version of the product you will still see "Tableau Prep" in the product.

This tutorial introduces you to the common operations that are available in Tableau Prep Builder. Using the sample data sets that come with Tableau Prep Builder, you will walk through creating a flow for Sample Superstore. This tutorial uses the most current version of Tableau Prep Builder. If you are using a previous version, your results may differ.

Watch for tips along the way to gain insights into how Tableau Prep Builder helps you clean and shape your data for analysis.

To install Tableau Prep Builder before continuing with this tutorial, see Install Tableau Desktop or Tableau Prep Builder from the User Interface in the Tableau Desktop and Tableau Prep Builder Deployment guide. Otherwise you can download the free trial.

Note: To complete the tasks in this tutorial, you need to install Tableau Prep Builder, and you need to download the following sample Superstore data files to a single location:

- Orders_South_2015
- Orders_South_2016
- Orders_South_2017
- Orders_South_2018
- Orders_Central
- Orders_East
- Orders_West
- returns_reasons_new

You can also find the files in the following location on your computer after installing Tableau Prep Builder:
In this article

Here’s the story... below
1. Connect to data on the facing page
2. Explore your data on page 65
3. Clean your data on page 67
4. Combine your data on page 80
5. Run your flow and generate output on page 93
Wrap up and resources on page 95

Here's the story...

You work at the headquarters for a large retail chain. Your boss wants to analyze product sales and profits over the last four years for the company. You suggest that he use Tableau Desktop to do that. Your boss thinks that’s a great idea and wants you to get right on that.

As you start gathering all the data you’ll need, you notice that the data has been collected and tracked differently for each region. You also notice a lot of creative data entry in the different files, and that one region even has a separate file for each year!

Before you can start analyzing the data in Tableau, you’ll have to do some serious data cleaning first, and it’s going to be a long night.

As you rummage for restaurant menus to order some dinner, you remember that Tableau just introduced a new product called Tableau Prep Builder that might help you with your Herculean data cleaning task.

You sign up for a free trial and decide to give it a try.
1. Connect to data

The first thing you see when you open Tableau Prep Builder is a Start page with a Connections pane, just like Tableau Desktop.

To get started, the first step is to connect to your data and create an Input step. From there you will start building a workflow or "flow", as it’s called in Tableau Prep Builder, and add more steps to take action on your data as you go.

Tip: The Input step is the ingestion point for your data and the starting point for your flow. You can have multiple Input steps and some might include multiple data files. For more information about connecting to data, see Connect to Data on page 109.

Your sales data files for the different regions are stored in different formats, and your orders from the South are actually multiple files. You check out the Connections pane and see that you have a lot of choices to connect to data. Great!

Since your other regions have one file for all four years worth of data, you decide to tackle the files from the South first.

1. On the Connections pane, click the Add connection button.

2. The files are .csv files, so select Text file in the list of connections.

3. Navigate to the directory for your files, select the first file orders_south_2015.csv and click Open to add it to your flow. (For file location, see Wrap up and resources on page 95.)

After you connect to your first file, the Tableau Prep Builder workspace opens and you see it is divided into two main sections. The Flow pane at the top and the Input pane at the bottom.
In Tableau Prep Builder, the **Flow** pane is a canvas, much like Tableau Desktop, where you can interact with your data visually and build your flow. The **Input** pane contains configuration options about how the data is ingested. It also shows you the fields, data types, and sample values for your data set.

We’ll look at how you can interact with this data in the next section.

**Tip:** For single tables, Tableau Prep Builder automatically creates an Input step for you in the **Flow** pane when you add data to your flow. Otherwise you can use drag-and-drop to add tables to the **Flow** pane.

4. You have three other files for your orders in the South. You could add each file individually, but you want to combine all the files together into one Input step, so you click the **Multiple Files** tab in the **Input** pane.

5. You see an option for **Wildcard union**. Select it.

You notice that the directory where you selected your file is already populated and the other files you need are listed in the **Included files** section in the Input pane.

**Tip:** Using a wildcard union is a great way to connect to multiple files from a single data source with a similar name and structure. To use this option, the files must be in the same parent or child directory. If you don't see the files you need right away, change your search criteria. For more information, see **Union files and database tables in the Input step** on page 122.
6. Click **Apply** to add these files to add the data from these files to the `orders_south_2015` input step.

The files for the other regions are all single table files, so you can select all of the files at once and add them to your flow.

7. Open File Explorer or Finder and navigate to the directory for the files. Ctrl+click (Command+click on Mac) to select the following files and drag-and-drop them onto the **Flow** pane to add them to your flow. (For file location, see Wrap up and resources on page 95.)

   - Orders_Central.csv
   - Orders_East.xlsx
   - Orders_West.csv

**Note:** These are different file types. If you don't see all of these files, make sure your file explorer or finder is set to view all file types.
Check your work: Watch "Connect to data" in action.

Click the image to replay it
2. Explore your data

Now that you have the data files loaded into Tableau Prep Builder, you're pretty sure that you want to combine the files together. But before you do that, it might be a good idea to take a look at them first and see if you can spot any issues.

When you select an Input step in the Flow pane, you can see the settings used to bring in the data, the fields that are included, and some sample values.

This is a good place to decide how much data you want to include in your flow and remove or filter fields that you don't want. You can also change any data types that were assigned incorrectly.

Tip: If you are working with large data sets, Tableau Prep Builder will automatically bring in a sample of the data to maximize performance. If you don't see the data you expect, you might need to adjust the sample. You can do this on the Data Sample tab. For more information about configuring your data options and sample size, see Configure your data set on page 130.

In the Flow pane, as you select each step and look over each data set, you notice a few things that you want to fix later and one thing that you can fix now in the Input step.

- In the Flow pane, click the Orders_Central Input step to select it. In the Input pane, you notice the following issues:
  - The order dates and ship dates are separated out into fields for month, day, and year.
  - Some of the fields have different data types than the same fields in other files.
  - There is no field for Region.

You'll need to do some cleaning on these fields before you can combine this file with the others files. But you can't fix that here in the Input step, so you make a note to do this later.

- Select the Orders_East Input step.

  The fields in this file look like they align pretty well with the other files. But the Sales values all seem to have the currency code included. You'll need to fix that later, too.

- Select the Orders_West Input step. There are some issues in this file too.
The **State** field uses abbreviations for the state name. Other files spell this out, so you'll need to fix that later.

There are a lot of fields that start with **Right_.** These fields appear to be duplicates of the other fields. You don't want to include these duplicate fields in your flow. This is something you can fix here in the **Input** step:

**To fix this now,** clear the check box for all fields that start with **Right_.** This tells Tableau Prep Builder to ignore these fields and not to include them in the flow.

**Tip:** When you perform cleaning operations in a step, like removing fields, Tableau Prep Builder tracks your changes in the **Changes** pane and adds an annotation in the Flow pane to help you keep track of the actions you take on your data. For **Input** steps, an annotation is also added to each field.

In Tableau Prep Builder version 2018.3.1 and later, cleaning operations are available in most steps in your flow. In prior versions, most cleaning operations must be performed in a cleaning step.

Now that you've identified a few troublemakers in your data sets, the next step is to examine your data a bit more closely and clean up any issues that you find so that you can combine and shape your data and generate an output file that you can use for analysis.
3. Clean your data

In Tableau Prep Builder, examining and cleaning your data is an iterative process. After you decide on the data set that you want to work with, the next step is to examine and take action on that data by applying various cleaning, shaping, and combining operations to it. You apply these operations by adding steps to your flow. For more information about cleaning options in Tableau Prep Builder, see Clean and Shape Data on page 157.

Steps come in many flavors, depending on what you are trying to do. For example, add a cleaning step (Add Step) any time you want to apply cleaning operations to your fields like filter, merge, split, rename, and so on. Add an aggregation step (Add Aggregate) to group and aggregate fields and change the level of detail of your data. For more information about the different step types and their uses, see Build your flow on page 157.

Tip: As you add steps to your flow, a flow line is automatically added to connect the steps to one another. You can move these flow lines around and remove or add them as needed.

When you run your flow, these connection points are required so Tableau Prep Builder knows which steps are connected and in which order the steps apply in the flow. If a flow line is missing, the flow will be broken and you'll get an error.

Clean Orders_Central

To address the issues you noticed earlier and to see if there are any other issues, you start by adding a cleaning step to the Orders_Central Input step.

1. In the Flow pane, select Orders_Central, click the plus icon and select Add Step.

   When you add a cleaning step to your flow, the workspace changes and you see the details of your data.
A. Flow pane, B. Toolbar, C. Profile pane, D. Data grid

The workspace is now split into three parts: the Flow pane, the Profile pane with a toolbar, and the Data grid. The Profile pane shows you the structure of your data, summarizing the field values into bins so that you can quickly see related values and spot outliers and null values.

The Data grid shows you the row level detail for your fields.

**Tip:** Each field in the Profile pane is shown on a profile card. Use the More options menu (drop-down arrow in prior versions) on each card to see and select the different cleaning options that are available for that field type. You can also sort the field values, change the data type, assign a data role to the field or drag and drop the profile cards and the columns in the Data grid to rearrange them.

**Clean data with calculated fields**

This data set is missing a field for Region. Since the other data sets have this field you'll need to add it so that you can combine your data later. You'll need to use a calculated field to do this.

2. In the toolbar, click **Create Calculated Field**.

3. Name the calculated field **Region**. Then enter "Central" (including the quotes) and click **Save**.
You love the flexibility of being able to use calculated fields to shape your data. You are pleased to see that Tableau Prep Builder uses the same calculation editor language as Tableau Desktop.

**Tip:** When you make changes to your fields and values, Tableau Prep Builder keeps track of them in the Changes pane. An icon representing the change is also added to the cleaning step in the flow and to the field in the Profile pane. We'll look at the Changes pane after making more changes.

Next you want to address the separate order date and ship date fields. You want to combine them into two single fields, one for **Order Date** and one for **Ship Date** so they align with the same fields in the other data sets.

You can use a calculated field again to do this in one easy step.

4. In the toolbar, click **Create Calculated Field** to combine the **Order Year**, **Order Month**, and **Order Day** fields into one field with the format "MM/DD/YYYY".

5. Name the calculated field **Order Date**. Then enter the following calculation into the Calculation editor and click **Save**:

   \[
   \text{MAKEDATE}([\text{Order Year}],[\text{Order Month}],[\text{Order Day}])
   \]
Now that you have a new field for your order date, you want to remove the existing fields, as you no longer need them.

You have a lot of fields in the Profile pane. You notice a Search box in the top right corner on the toolbar. You wonder if you can use that to quickly find the fields that you want to remove. You decide to give it a try.

6. In the Profile pane, in the search box, type Order.

Tableau Prep Builder quickly scrolls all the fields with Order in the name into view. Cool!

7. Ctrl+click or Command+click (Mac) to select the fields for Order Year, Order Month, and Order Day. Then right-click on the selected fields and select Remove Field from the menu to remove them.
8. Now repeat steps 4 though 7 above to create a single field for **Ship Date**. Try it on your own or use the steps below to help you.

- In the toolbar, click **Create Calculated Field** to combine the **Ship Year**, **Ship Month**, and **Ship Day** fields into one field with the format "MM/DD/YYYY".

- Name the calculated field **Ship Date** and enter the calculation `MAKEDATE([Ship Year], [Ship Month], [Ship Day])`. Then click **Save**.

- Remove the **Ship Year**, **Ship Month**, and **Ship Day** fields. Search for the fields, select them, and select **Remove Field** from the menu.

**Tip:** Tableau Prep Builder summarizes the data in the Profile pane into bins to help you quickly see the shape of your data, find outliers, spot relationships between fields, and so on.

In this scenario, the order and ship dates can now be summarized by year. Each bin represents a year from January of the beginning year to January of the following year and is labeled accordingly. Because there are sales dates and ship dates that fall in the latter part of 2018 and 2019, we get a bin for that data that is labeled with the ending year 2019 and 2020 accordingly.

To change this view to the actual dates, click the **More options** menu (drop-down
arrow in prior versions) in the Profile card and select **Detail**.

Interact directly with fields to clean your data

Your data is starting to look good. But, as you finish removing the extra fields for the order and ship dates, you notice that the **Discounts** field has a couple of issues.

- It’s assigned to a **String** data type instead of a **Number (decimal)** data type.
- There’s a field value **None** instead of a numeric value for no discount.

This will cause a problem when you combine the files, so you better fix that too.

**9.** Clear your search and enter **disc** in the search box to find the field.

**10.** Select the **Discounts** field, double-click the field value **None**, and change it to the numeric value **0**.

**11.** Change the data type for the **Discount** field from **String** to **Number (decimal)**. Click **Abc** and select **Number (decimal)** from the drop-down menu.
12. Finally name your step to help keep track of what you did in this step. In the Flow pane, double-click the step name Clean 1 and type in Fix dates/field names.

Review your changes

You made a lot of changes to this data set and you start to worry that you won’t remember everything you did. As you look over your work, you see a column on the left of the Profile pane called Changes.

You click the arrow to open it and are delighted to see a list of every change you just made. As you scroll through the changes in the list, you notice that you can delete or edit your changes or even move them around to change the order that you did them in.

You love that you can easily find the changes you made in any step as you build your flow and experiment with the order of those changes to get the most out of your data.
Check your work: Watch "Clean Orders_Central" in action.

Click the image to replay it

Now that you've cleaned one file, you take a look at the other files to see what other issues you need to fix.

You decide to look at the Excel file for **Orders_East** next.

**Clean Orders_East**

As you look over the fields for the **Orders_East** file, most of the fields look like they align with the other files, except for **Sales**. To take a closer look and see if there are any other issues to address, you add a cleaning step to the **Orders_East** Input step.
1. In the **Flow** pane, select **Orders_East**, click the plus icon and select **Add Step**.

Looking at the **Sales** field you quickly see that the **USD** currency code has been included with the sales numbers, and Tableau Prep Builder interpreted these field values as a string.

You'll need to remove the currency code from this field and change the data type if you want to get accurate sales data.

Fixing the data type is easy, you already know how to do that. But there are over 2000 unique rows of sales data and fixing every individual row to remove the currency code seems cumbersome.

But this is Tableau Prep Builder, and you decide to check out the drop-down menu to see if there is an option to fix this.

When you click the drop-down arrow for the **Sales** field, you see a menu option called **Clean** and an option under that to remove letters. You decide to give that a try and see what it does.

2. Select the **Sales** field. Click the **More options** menu (drop-down arrow in prior versions) and select **Clean > Remove Letters**.

Wow! That cleaning option instantly removed the currency code from every field. Now you just need to change the data type from **String** to **Number (decimal)** and this file is looking good.

3. Click the data type and select **Number (decimal)** from the drop-down list.
4. The rest of the file looks pretty good. Name your cleaning step to keep track of your work. For example, **Change data type**.

Next you look at your last file for **Orders_West** to see if there are any issues there that you need to fix.

**Clean Orders_West**

As you look over the fields for the **Orders_West** file, most of the fields look like they align with the other files, but you remember seeing that the **States** field used abbreviations for the values instead of spelling out the state name. To combine this file with the other files, you'll need to fix this. So you add a cleaning step to the **Orders_West** Input step.

1. In the **Flow pane**, select **Orders_West**, click the plus icon and select **Add Step**.

Scroll or use Search to find the **State** field. You see that all the state name values use the short abbreviation. There are only 11 unique values for this field. You could manually change each one, but maybe Tableau Prep Builder has another way to do this?

You click the **More options** menu (drop-down arrow in prior releases) for the field and see an option called **Group and Replace**. When you select it you see several options:

- Manual Selection
- Pronunciation
Tip: You can double-click a field name or field value to edit a single value. To edit multiple values you can select all the values and use the right-click menu option Edit Values. But when you want to map one or more values to specific values, use the Group and Replace option in the drop-down menu.

For more information about editing and grouping values, see Cleaning (fixing) variations of the same value on page 185.

2. Select the State field. Click the drop-down arrow and select Group and Replace > Manual Selection.

   ![Group and Replace editor](image)

   A two column card opens. This is the Group and Replace editor. The column on the left shows the current field values and the column on the right shows the fields that are available to map to the fields on the left.

   You want to map your state abbreviations to the spelled out version of the state name, but you don't have those values in the Orders_West data set. You wonder if you can just edit the name directly and maybe add it there, so you give that a try.

3. In the Group and Replace editor in the left pane, double-click AZ to highlight the value and type Arizona. Then press Enter to add your change.
Tableau Prep Builder created a mapped value for your new value **Arizona** and automatically mapped the old value, **AZ** to it. Having a mapped relationship set up for these values will save you time if you get more data from this region entered like this.

**Tip:** You can add field values that aren’t in your data sample to set up mapping relationships to organize your data. If you refresh your data source and new data is added, you can add the new data to the mapping instead of manually fixing each value.

When you manually add a value that isn’t in your data sample, the value is marked with a red dot to help you easily identify it.

4. Repeat these steps to map each state to the spelled out version of its name.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>State Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZ</td>
<td>Arizona</td>
</tr>
<tr>
<td>CA</td>
<td>California</td>
</tr>
<tr>
<td>CO</td>
<td>Colorado</td>
</tr>
<tr>
<td>ID</td>
<td>Idaho</td>
</tr>
<tr>
<td>MT</td>
<td>Montana</td>
</tr>
<tr>
<td>NM</td>
<td>New Mexico</td>
</tr>
<tr>
<td>NV</td>
<td>Nevada</td>
</tr>
</tbody>
</table>
Then click **Done** to close the **Group and Replace** editor.

![Diagram of Tableau Prep interface showing state mapping and changes pane.]

After all the states are mapped, you look at the **Changes** pane and see there is only one entry there instead of 11.

It looks like Tableau Prep Builder groups similar actions for a field together. You like that because it will make it easier to find changes you made to your data set later.

Fixing the **State** field values was the only change you needed to make here.

5. Name your cleaning step to keep track of your work. For example **Rename states**.

You've done a lot of clean up in your files, and you can't believe how quick and easy it was. You might make it home for dinner after all! To make sure that you don't lose all of your work so far, save your flow.

Click **File > Save** or **File > Save As**. Save your file as a flow file (.tfl) and give it a name. For example, **My Superstore**.
Tip: When you save your flow files, you can either save them as a flow file (.tfl) or you can save them as a packaged file (.tflx) and package your local data files with them to share the flow and files with someone else. For more information about saving and sharing your flows, see Save and Share Your Work on page 213.

4. Combine your data

Now that all the files are cleaned up, you are finally ready to combine them all.

Because all the files have similar fields, you want to union the files together to add the rows from each file into a single table.

You remember that there was a step option called Add Union, but you wonder if you can simply drag and drop the steps to union them. You decide to try it and see.

Union your data

1. In the Flow pane, drag the cleaning step Rename states to the Changed data type step and drop it on the Union option.

   ![Diagram of Union step]

   You see that Tableau Prep Builder added a new Union step to your flow. Great! Now you want to add the other files to this union too.

   2. Drag the Fix dates/field names step to the new Union step. Drop it on Add to add it to the existing union.
3. Drag the **orders_south_2015** step to the new **Union** step. Drop it on **Add** to add it to the existing union.

Now all of your files are combined into a single table. In the **Flow** pane, select the **Union** step to see your results.

You notice that Tableau Prep Builder automatically matched up the fields that had the same names and types.

You also see that the **colors** assigned to the steps in the flow are used in the union profiles to indicate where the field came from and also appear in the **colored band** across the top of each field to show you if that field exists in that table.
You notice that a new field called **Table Names** was added that lists the tables where all the rows in the union come from.

A list of mismatched fields also shows in the summary pane and you can see right away that the fields **Product** and **Discounts** only appear in the **Orders_Central** file.

4. To take a closer look at these fields, in the **Union Results** pane, select the **Show only mismatched fields** check box.

![Diagram of Tableau Prep interface showing mismatched fields]

Looking at the field data, you quickly see that the data is the same, but the field name is different. You could simply rename the field, but you wonder if you could just drag and drop these fields to merge them. You decide to try that and see.

5. Select the **Product** field and drag and drop it onto the **Product Name** field to merge the fields. After the fields are merged, they no longer appear in the pane.
6. Repeat this step to merge the **Discounts** field with the **Discount** field.

   The only field that doesn't have a match now is the **File Paths** field. This field shows the file paths for the wildcard union that you did for your sales orders from the South. You decide to leave this field there as it has good information.

   **Tip:** You have several options when fixing mismatched fields after a union.

   Depending on the Tableau Prep Builder version you are using, you can select a field in the **Mismatched Fields** list (version 2018.2.1 and later) and if Tableau Prep Builder detects a possible match, it will highlight it in yellow. To merge the fields hover over the highlighted field and click the plus button that appears.

   For more ways to merge fields in a union, see Fix fields that don't match on page 209.

7. Clear the **Show only mismatched fields** check box to show all the fields included in the union.

8. Name your Union step to represent what this union includes work. For example, **All orders**.

Check your work: Watch "Union your data" in action.

*Click the image to replay it*
As you are admiring the results of your cleaning prowess, your boss calls. He forgot to mention that he also wants you to include any product returns in your analysis. He hopes that won’t be too much trouble. With Tableau Prep Builder in your toolkit, it’s not a problem.

Clean the product returns data

You look over the Excel file that your boss sent you for product returns and it looks a little messy. You add the new file `return_reasons_new` to your flow to take a closer look.

1. In the Connections pane, click Add connection. Select Microsoft Excel and navigate to the sample Superstore data files (see Wrap up and resources on page 95 for the file location).

2. Select `return_reasons_new.xlsx`, and then click Open to add the file to the flow pane.

   There are only 4 fields that you want to include from this file in your flow: Order ID, Product ID, Return Reason and Notes.

3. In the Input pane for `returns_new` clear the check box at the top of the field grid to clear all the check boxes. Then select the check box for the Order ID, Product ID, Return Reason and Notes fields.

4. Rename the Input step to better reflect the data that is included. In the Flow pane,
double-click the Input step name `Returns_new` and type in `Returns (all)`.

Looking at the sample field values, you notice that the Notes field seems to have a lot of different data combined together.

You have some cleaning to do in this file before you can do any further work with the data, so you add a cleaning step to check it out.

5. In the Flow pane, select the Input step `Returns (all)`, click the plus icon, and then select Add Step.

In the Profile pane, click and drag the outer right edge of the field to the right to re-size the Notes field so you can see the entries better.

6. In the Notes field, use the visual scroll bar to the right of the field values to scan the values.

You notice a few things that are problematic:

- Some of the entries have an extra space in the entry. This can result in the field being read as a null value.

- It looks like the name of the approver is included in the return notes entry. To better work with this data you'll want that information in a separate field.

To tackle the extra spaces, you remember that there was a cleaning option to remove trailing spaces, so you decide to try that to see if it can fix that problem.

7. Select the Notes field. Click the More options menu (drop-down arrow in prior releases) and select Clean > Trim Spaces.

Yes! It did exactly what you wanted it to do. The extra spaces are gone.
Next you want to create a separate field for the approver name. You see a Split Values option in the menu, so you decide to try that.

8. Select the Notes field. Click the More options menu (drop-down arrow in prior releases) and select Split Values > Automatic Split.

This option did exactly what you were hoping it would do. It automatically split the return notes and the approver name into separate fields.

Just like Tableau Desktop, Tableau Prep Builder automatically assigned a name to those fields. So you'll need to rename the new fields to something meaningful.

9. Select the field Notes-Split 1. Double-click in the field name and type Return Notes.

10. Repeat this step for the second field and rename it to Approver.

11. Finally remove the original Notes field, as you no longer need it. Select the Notes field, click the More options menu (drop-down arrow in prior releases), and select Remove Field from the menu.

Looking at the new Approver field, you notice that the field values lists the same names but they are entered differently. You want to group them to eliminate multiple variations of the same value.

Maybe the Group and Replace option can help with that?

You remember there was an option for Common Characters. Since these values share the same letters, you decide to try that.

12. Select the Approver field. Click the More options menu (drop-down arrow in prior releases) and select Group and Replace > Common Characters.
This option grouped all of the variations of each name together for you. That's exactly what you wanted to do.

After checking the other names to make sure they are grouped properly, you click Done to close the Group and Replace editor.

This file is looking pretty good.

13. Name your cleaning step to keep track of your work. For example Cleaned notes.

Now that the product return data is all cleaned up, you want to add this data to the orders data in your unioned files. But many of these fields don't exist in the unioned files. To add these fields (columns of data) to your unioned data set, you need to use a join.

Join your data

When you join data, the files must have at least one field in common. Your files share the Order ID and Product ID fields, so you can join on those fields to see all the rows that have those
fields in common. You remember an option to create a join when you created your union using drag and drop, so you give that a try.

1. In the Flow pane, drag the Cleaned notes step on to the All orders Union step and drop it on Join.

When you join files, Tableau Prep Builder shows you the results of your join in the Join Profile.

Working with joins can be tricky. You often want to have a clear view of the factors that are included in the join, such as the fields used to join the files, the number of rows included in the results and any fields that aren’t included or are null values.

As you review the results of the join in Tableau Prep Builder, you are delighted to see so much information and interactivity at your fingertips.
Tip: The far left pane of the join profile is where you can explore and interact with your join. You can also fix values directly in the Join Clauses panes and perform cleaning operations in the Join Results pane.

Choose the fields that you want to join on in the Applied Join Clauses section or add suggested join clauses from the Join Clause Recommendations section.

Click in the Join Type diagram to try different join configurations and see the number of rows included or excluded in your join for each table in the Summary of Join Results section.

For more information about working with joins, see Join or Union Data on page 201.

You see that you have over 13,000 rows excluded from your All Orders files. When you created your join, Tableau Prep Builder automatically joined on the Product ID field, but you wanted to also join on the Order ID field.

As you scan the left pane of the join profile, you see that Order ID is in the list of recommended join clauses, so you quickly add it from there.

2. In the left pane of the Join profile, in the Join Clause Recommendations section, select Order ID = Order ID and click the plus button to add the join clause.
Because the Join Type is set to an inner join (the default setting for Tableau Prep Builder), the join is only including values that exist in both files. But you want all of the data from your Orders files as well as the return data for those files. So you'll need to change the join type.

3. In the Join Type section, click the left side of the diagram to change the join type to a Left join.
Now you have all of the data from the sales order files and any return data that apply to those orders. You review the **Join Clauses** pane and see the distinct values that don’t exist in the other file.

For example there are many order rows (shown in red) that have no corresponding return data. You love being able to explore this level of detail about your join.

You’re anxious to start analyzing this data in Tableau Desktop, but you notice a few results from the join that you want to clean up before you do that. Good thing you know what to do!

**Tip:** Wonder if your data is clean enough? You can preview your data in Tableau Desktop from any step in your flow to check it out.

Just right-click on the step in the **Flow** pane and select **Preview in Tableau Desktop** from the menu.

You can experiment with your data and any changes that you make in Tableau Desktop
won’t write back to your data source in Tableau Prep Builder. For more information see View your data sample in Tableau on page 213.

4. Before you start cleaning your join results, name your Join step Orders+Returns and save your flow.

Clean your join results

**Note:** To clean up the fields in your join, you can perform cleaning operations directly in the Join step. For the purposes of this tutorial we will add a cleaning step so you can clearly see your cleaning operations. If you want to try performing these steps directly in the join step skip steps 1 and 3 below.

When you joined the two steps, the common fields Order ID and Product ID were added for both tables.

You want to keep the Product ID field from all of your orders and the Order ID field from the returns file and remove the duplicate fields that came from those files. You also don’t need the File Paths and Table Names fields in your output file, so you want to remove those fields as well.

**Tip:** When you join tables using fields that exist in both files, Tableau Prep Builder will bring in both fields and rename the duplicate field from the second file by adding a "-1" or a "-2" to the field name. For example Order ID and Order ID-1.

1. In the Flow pane, select Orders+Returns, click the plus icon, and then select Add Step.

2. In the Profile pane, select and remove the following fields:
   - Table Names
   - Order ID
   - File Paths
   - Product ID-1

3. Rename the field Order ID-1 to Order ID.
You have quite a few null values where the product was returned but there was no return note or approver indicated. To make this data easier to analyze, you want to add a field with a value of Yes and No to indicate whether the product was returned.

You don’t have this field, so you can add it by creating a calculated field.

4. In the toolbar, click **Create Calculated Field**.
5. Name the field **Returned?** and then enter the following calculation and click **Save**.

   If ISNULL([Return Reason])=FALSE THEN "Yes" ELSE "No" END

For your analysis you would also like to know the number of days it takes to ship an order, but you don’t have that field either.

You have all the information that you need to create it though, so you add another calculated field to create it.

6. In the toolbar, click **Create Calculated Field**.
7. Name the field **Days to Ship** and then enter the following calculation and click **Save**.

   DATEDIFF('day',[Order Date],[Ship Date])

8. Name your step **Clean Orders +Returns**.
9. Save your flow.

5. Run your flow and generate output

Your data is looking good and you’re ready to generate your output file to start analyzing it in Tableau Desktop. All you need to do is run your flow and generate your extract file. To do this you need to add an **Output** step.

1. In the **Flow pane**, select **Clean Orders +Returns**, click the plus icon and select **Add Output**.

When you add an Output step, the **Output** pane opens and shows you a snapshot of your data. Here you can select the type of output that you want to generate, and specify the name and where you want to save the file.

The default location is in the **My Tableau Prep Builder** repository in your data sources folder.
2. In the left pane select **Save to file**.

3. Click the **Browse** button, then in the **Save Extract As** dialog, enter a name for the file, for example **Orders_Returns_Superstore**, and click **Accept**.

4. In the **Output type** field, select the output type. Depending on the version of Tableau Desktop you use you can choose from the following options:
   - Tableau Data Extract (.hyper) for Tableau Desktop version 10.5 and later.
   - Tableau Data Extract (.tde) for Tableau Desktop version 10.0 through 10.4.
   - Comma Separated Values (.csv) if you want to share the extract with a third party.

**Tip:** You have choices when generating output from your flow. You can generate an extract file, or you can publish your data as a data source to Tableau Server or Tableau Online. For more information about generating output files, see **Create and publish data extracts and data sources** on page 214.

5. Click the **Run Flow** button to generate your output.

6. When the flow is finished running, a status dialog shows whether the flow ran successfully and the time it took to run. Click **Done** to close the dialog.
Wrap up and resources

You are a data prep rock star! You took dirty data and transformed it with ease! In no time, you cleaned and prepped your data from multiple data sets and turned it into a sleek, clean data set that you can now work with in Tableau Desktop to do your analysis.

Want more practice? Try replicating the rest of the sample flow for Superstore using the data files found here:

- Orders_South_2015
- Orders_South_2016
- Orders_South_2017
- Orders_South_2018
- Orders_Central
- Orders_East
- Orders_West
- returns_reasons_new
- Quota

You can also find the files in the following location on your computer after installing Tableau Prep Builder:

- (Windows) C:\Program Files\Tableau\Tableau Prep Builder <version>\help\Samples\en_US\Superstore Files
- (Mac) /Applications/Tableau Prep Builder <version>.app/Contents/help/Samples/en_US/Superstore Files
Want more training? Check out the new training videos for Tableau Prep Builder or take an in-person training course.

Want more information about the topics we covered? Check out the other topics in the Tableau Prep Builder online help.

About Tableau Prep Builder

Tableau Prep Builder is a tool in the Tableau product suite designed to make preparing your data easy and intuitive. Use Tableau Prep Builder to combine, shape, and clean your data for analysis in Tableau.

Note: Tableau Prep version 2019.1.2 had changed its name to Tableau Prep Builder. If you are using an earlier version of the product you will still see "Tableau Prep" in the product.

Using Tableau Prep Builder

Start by connecting to your data from a variety of files, servers, or Tableau extracts. Connect to and combine data from multiple data sources. Drag and drop or double-click to bring your tables into the flow pane, and then use familiar operations such as filter, split, rename, pivot, join, and union to clean and shape your data.

Each step in the process is represented visually in a flow chart that you create and control. Tableau Prep Builder tracks each operation so that you can check your work and make changes at any point in the flow.

When you are finished with your flow, run it to apply the operations to the entire data set.

Tableau Prep Builder works seamlessly with other Tableau products. At any point in your flow, you can create an extract of your data, publish your data source to Tableau Server or Tableau Online, publish your flow to Tableau Server to refresh your data using a schedule or open Tableau Desktop directly from within Tableau Prep Builder to preview your data.

For information about installing Tableau Prep Builder, see Install Tableau Desktop or Tableau Prep Builder in the Tableau Desktop and Tableau Prep Deployment Guide.
See Tableau Prep Builder in action

*Click the image to replay it.*

Ready to try it out? From the **Start** page, click on one of the sample flows to explore and experiment with the steps, try the **Get Started with Tableau Prep Builder** on page 59 hands-on tutorial to learn how to create a flow or try stepping through one of our **Day in the Life Scenarios** on page 303 using Tableau Prep Builder.
Note: You can find the sample data files used in the flows in these locations:

- (Windows) C:\Program Files\Tableau\Tableau Prep Builder <version>\help\Samples\en_US
- (Mac) /Applications/Tableau Prep Builder <version>.app/Contents/help/Samples/en_US

To learn more about Tableau Prep Builder and the different features and functions it offers, review the topics in this guide.

A tour of the Tableau Prep Builder workspace

The Tableau Prep Builder workspace consists of the Connections pane (A) and three coordinated areas that help you interact with and explore your data:

- Flow pane (B): A visual representation of your operation steps as you prepare your data.
- Profile pane (C): A summary of each field in your data sample. See the shape of your
data and quickly find outliers and nulls.

- **Data grid (D):** The row level detail for your data.

After you connect to your data and begin building your flow, you add steps in the **Flow** pane. These steps function as a lens into the structure of your data, as well as a summary of operations that is applied to your data. Each step represents a different category of operations that you define, all as part of your flow.

**Connections pane**

On the left side of the workspace is the **Connections** pane, which shows the databases and files you are connected to. Add connections to one or more databases and then drag the tables you want to work with into the **Flow** pane. For more information see **Connect to Data** on page 109.
You can minimize the **Connections** pane if you need more room in your workspace.

**Flow pane**

At the top of the workspace is the **Flow** pane. This is your canvas. As you connect to, clean, shape, and combine your data, steps appear in the **Flow** pane and align from left to right along the top. These steps tell you what kind of operation is being applied, in what order, and how your data is affected by it. For example, the Join step shows you which join type you’ve applied, the join clauses, recommended join clauses, and the fields of the tables that are included in the join.

You start your flow by dragging tables into the **Flow** pane. Here you can add additional data sets, pivot your data, union or join data, create aggregations, and generate output files in the form of .tde and .tds files or Hyper extract (.hyper) files that you can use in Tableau. For more information about generating output files, see **Save and Share Your Work** on page 213.
Note: If you make changes to the data while in Tableau Desktop, for example renaming fields, changing data types, and so on, these changes aren’t written back to Tableau Prep Builder.

Profile pane

In the center of the workspace is the Profile pane. The Profile pane shows you the structure of your data at any point in the flow. The structure of your data can be represented in different ways depending on the operation you want to perform on your data or the step that you select in the Flow pane.

At the top of the Profile pane is a toolbar that shows you the cleaning operations that you can perform for each step in your flow. An options menu also appears on each card in the Profile pane where you can select the different operations that you can perform on the data.

For example:

- Search, sort, and split fields
- Filter, include, or exclude values
- Find and fix null values
- Rename fields
- Clean up data entry errors using group and replace or quick cleaning operations
- Use automatic data parse to change data types
- Rearrange the order of your field columns by dragging and dropping them where you want them
Select one or more field values in a Profile card and right-click or Ctrl-click (Mac) to see additional options to keep or exclude values, group selected values or replace values with Null.

Tableau Prep Builder keeps track of any changes you make, in the order you make them, so you can always go back and review or edit those changes if needed. Use drag and drop to re-order those operations to experiment and apply changes in a different order.

Click the arrow on the upper right of the pane to expand and collapse the Changes pane for more room to work with the data in the Profile pane.

For more information about applying cleaning operations to your data see Clean and Shape Data on page 157.

Data grid

At the bottom of the workspace is the Data grid, which shows you the row level detail in your data. The values displayed in the Data grid reflect the operations defined in the Profile pane. You can perform the same cleaning operations here as you can in the Profile pane if you prefer to work at a more detailed level.
Click the **Collapse Profiles** icon on the toolbar to collapse (and expand) the **Profile** pane to see your options.

How Tableau Prep Builder stores your data

When you connect Tableau Prep Builder to your data and create a flow, it stores the frequently used data in a .hyper file. For large data sets, this might be a sample of the data. Any stored data is saved under a secure temporary file directory in a file named Prep BuilderXXXXX, where XXXXX represents a universally unique identifier (UUID). After you save the flow, the file is deleted. For more information about how Tableau Prep Builder samples your data, see **Set your data sample size** on page 134.

Tableau Prep Builder also saves data in the Tableau flow (.tfl) file to support the following operations (which can capture entered data values):

- Custom SQL used in Input steps
- Filtering (on data entry)
- Group and replace (on data entry)
- Calculations
Input Step
Start your flow by dragging data to the Flow pane to create an input step. The icon shows you the type of data source:
- Data Source
- Data Source with Wildcard Union
- Excel
- Excel with Wildcard Union
- CSV
- CSV with Wildcard Union
- Tableau Extract
**Tableau Prep Help**

### Clean Step, Changes Pane, and Toolbar

Easily keep track of the changes you make to your data. Find these icons on the Clean steps in the Flow pane, in the Changes pane, and on the toolbar:

- **Calculate Field**
- **Change Data Type**
- **Edit Value**
- **Exclude Values**
- **Filter Values**
- **Group Values**
- **Keep Only**
- **Hide Profile Pane**
- **Show Profile Pane**
- **Merge Fields**
- **Remove Field**
- **Rename Field**
- **Search**
- **Split Fields**

### Union Step

Combine up to ten sources of data with similar fields in a single Union step:

- **Union Data**

### Join Step

Connect two steps to join your data on one or more common fields. Select one of these options to choose the join type:

- **Full Anti Join**
- **Inner Join**
- **Left Inner Join**
- **Left Outer Join**
- **Full Outer Join**
- **Right Inner Join**
- **Right Outer Join**
Output Step

Add an Output step and run your flow to apply the changes to the complete data set and generate output files.

- CSV
- Published Data Source
- Tableau Data Extract
- Run Flow

Pivot Step

Add a Pivot step to change columns to rows.

- Pivot Data

Aggregate Step

Add an Aggregate step to group and aggregate your data, which changes the level of detail of your data.

- Aggregate Data
Notification
If there’s a problem with your flow or something you need to know, check notifications. Errors include a Go to Error link to help you quickly find the problem.

- No Notifications
- Notification Alert
- Error in the Step

Profile Pane
See the exact row count of your data and know when your data is sampled.

- Sampled
  - Only shows when data is sampled

Profile Card
Identify the data type and see the options available to apply to your data when you select a field in the Profile pane.

- Calculated Field
- Change Data Type
- Edit Value
- Exclude Values
- Filter Values
- Group Values
- Keep Only
- Merge Fields
- Remove Field

- Rename Field
- Search
- Split Fields
- Boolean Data Type
- Date Data Type
- Date Time Data Type
- Numeric Data Type
- Text Data Type
Connect to Data

To use Tableau Prep Builder to clean and prepare your data, start a new flow by connecting to your data, just like in Tableau Desktop. You can also open an existing flow and pick up where you left off.

You can see and access your most recent flows right on the Start page, so it's easy to find your work in progress. After you connect to your data, use the different options in the Input step to identify the data that you want to work with in your flow.

Start or open a flow

Tableau Prep Builder supports connections to popular types of data as well as Tableau data extracts (.tde or .hyper). Connectors that are supported in Tableau Desktop might not yet be supported in your version of Tableau Prep Builder. But new connectors are added with each new version, so check the Connections pane to see if your connector is available.

**Note:** If you open a flow in a version where the connector isn't supported, the flow may open but might have errors or won't run unless the data connections are removed.

You can also use custom SQL queries to connect to data just like you can in Tableau Desktop today. For more information, see Connect to a Custom SQL Query in the Tableau Desktop and Web Authoring Help.

To check whether you can connect to your data, open Tableau Prep Builder and click the Add connection button to see if a connector for your data is listed in the left pane under Connect.

**Note:** Some connectors might require you to download and install a driver before you can connect to your data. See the Driver Download page on the Tableau website to get driver download links and installation instructions.
Start a new flow

1. Open Tableau Prep Builder and click the **Add connection** button.
2. From the list of connectors, select the file type or server that hosts your data. If
prompted, enter the information needed to sign in and access your data.

3. From the **Connections** pane, do one of the following:

   - If you connected to a file, double-click or drag a table to the **Flow** pane to start your flow.

     For single tables, Tableau Prep Builder automatically creates an Input step for you in the **Flow** pane when you add data to your flow.

   - If you connected to a database, select a database or schema, and then double-click or drag a table to the **Flow** pane to start your flow.
Open an existing flow

To open an existing flow, on the Start page do one of the following:

- Under Recent Flows, select a flow.
- Click Open a Flow to navigate to your flow file and open it.

Click File > New to open a new workspace.

Connecting to SAP HANA (version 2019.2.1 and later)

Tableau Prep Builder supports connecting to data using SAP HANA just like Tableau Desktop but with a few differences. Connect to the database using the same procedure you would use in Tableau Desktop. For more information see SAP HANA. After you connect and search for your table, drag the table to the canvas to begin building your flow.

Prompting for variables and parameters when opening a flow isn't supported in Tableau Prep Builder. Instead, in the Input pane, click the Variables and Parameters tab and select the variables and operands you want to use, then select from a list of preset values or enter custom values to query your database and return the values you need.

**Note:** Starting in Tableau Prep Builder version 2019.2.2 you can use Initial SQL to query your connection and if you have multiple values for a variable, you can select the value you need from a drop-down list.
You can also add additional variables. Click the plus button in the Variables section and select a variable and operand, then enter a custom value.

**Note:** This connector isn't supported in Tableau Prep Conductor version 2019.1. If you create a flow that uses this connector, you can publish the flow but you won't be able to run the flow using the scheduling features in Tableau Server. Instead you can run the flow using the command line interface to keep your data fresh. For more information about running flows from the command line see Refresh output files from the command line on page 218. For more information about version compatibility, see Version Compatibility with Tableau Prep on page 226.
Connecting using ODBC (version 2019.2.2 and later)

If you need to connect to data sources that aren't listed in the Connections pane, you can connect to any data source using the Other Databases (ODBC) connector that supports the SQL standard and implements the ODBC API. Connecting to data using the Other Databases (ODBC) connector works similarly to how you might use it in Tableau Desktop, however there are a few differences:

- You can only connect using the DSN (data source name) option.
- To publish and run your flow in Tableau Server, the server must be configured using a matching DSN.

**Note:** Running flows from the command line (Tableau Prep Builder) that include the Other Databases (ODBC) connector isn't currently supported.

- There is a single connection experience for both Windows and MacOS. Prompting for connection attributes for ODBC drivers (Windows) isn't supported.
- Only 64-bit drivers are supported by Tableau Prep Builder.

Before you connect

To connect to your data using the Other Databases (ODBC) connector, you must install the database driver and set up and configure your DSN (data source name). To publish and run flows to Tableau Server, the server must also be configured with a matching DSN.

**Important:** Tableau Prep Builder only supports 64-bit drivers. If you have a 32-bit driver already set up and configured, you may need to uninstall it and then install the 64-bit version if the driver doesn't allow both versions to be installed at the same time.

1. Create a DSN using either the **ODBC Data Source Administrator (64-bit)** (Windows) or the using an ODBC Manager utility (MacOS).
   
   If you don't have the utility installed on your Mac, you can download one from (www.odbcmanager.net for example) or you can manually edit the odbc.ini file.

2. In the **ODBC Data Source Administrator (64-bit)** (Windows) or the ODBC Manager utility (MacOS), add a new data source then select the driver for the data source then click **Finish**.
3. In the **ODBC Driver Setup** dialog, enter the configuration information such as server name, port, user name and password. Click **Test** (if your dialog has that option) to verify that your connection is set up correctly, then save your configuration.

**Note**: Tableau Prep Builder doesn't support prompting for connection attributes so you must set this information when configuring the DNS.

This example shows the configuration dialog for a MySQL Connector.
Connect using Other Databases (ODBC)

1. Open Tableau Prep Builder and click the Add connection button.

2. From the list of connectors, select Other Databases (ODBC).

3. In the Other Databases (ODBC) dialog, select a DSN from the drop-down list and enter the user name and password. Then click Sign In.
4. From the **Connections** pane, select your database from the drop-down list.
Use Initial SQL to query your connections (version 2019.2.2 and later)

You can specify an Initial SQL command when connecting to databases that support it, that will run when a connection is made to the database. For example when connecting to Amazon Redshift, you can enter a SQL statement to apply a filter when connecting to the database just like adding filters in the Input step. The SQL command will apply before data is sampled and loaded into Tableau Prep Builder.

To refresh your data and run the Initial SQL command again do one of the following:

• Change the Initial SQL command and refresh the Input step be re-establishing the connection.

• Run the flow. The initial SQL command is run before processing all of the data.

• Schedule the flow to run on Tableau Server. The initial SQL is run every time that the flow is run, as part of the data loading experience.

1. In the Connections pane, select a connector in the list that supports Initial SQL.
2. Click the Show Initial SQL link to expand the dialog and enter your SQL statements.

Important: Parameters are not yet supported for Initial SQL in Tableau Prep Builder. If you try and enter the parameters like you would in Tableau Desktop, you will receive errors.

Working with Tableau data extracts

When you connect to a Tableau data extract, Tableau Prep Builder unpackages the extract and hyper expands, using a lot of temp space as it applies your flow operations to the resulting
raw data.

This means you may need more RAM and disk space to accommodate a file that size. For example, an extract file with 18 columns and 1.2 million rows that is 360MB (8.5 GB uncompressed) may need up to 32GB RAM, 16-core, and 500GB of disk space available to support the file when it is unzipped.

Refresh data in the Input step

If data changes in your input files or tables after you begin working with your flow, you can refresh the Input step to bring in the new data.

File input step types

To refresh file inputs steps, do one of the following:

- In the flow pane on the top menu, click the Refresh button to refresh all Input steps. To refresh a single Input step, click the drop-down arrow next to the refresh button and select the Input step from the list.

- In the flow pane, right-click the Input step you want to refresh and select Refresh from the menu.

File, database or Tableau extract input step types

To refresh database or tableau extract input steps, do one of the following:
Try editing the connection.

**Note:** To maintain performance, Tableau Prep Builder samples large data sets. If your data is sampled, you may or may not see your new data in the profile pane. You can change the settings for how your data is sampled in the Data Sample tab in the Input step, but it may impact performance. For more information about setting your data sample size, see Set your data sample size on page 134.

1. In the **Connections** pane, right-click or Ctrl-click (MacOS) on the data source and select **Edit**.

   ![Connections pane with edit option](image)

2. Re-establish your connection by signing into the database or re-selecting the file or Tableau extract.

   Remove and re-add the Input step to the flow.

   1. In the flow pane, right-click the Input step you want to refresh and select **Remove** from the menu.

   ![Flow pane with remove option](image)

   This will temporarily put your flow in an error state.

   2. Connect to the updated file again.
3. Drag the table to the flow pane on top of the second step in the flow where you want to add the Input step. Drop it on the Add option to reconnect it to the flow.

Use Data Interpreter to clean your files

When working with Microsoft Excel files, you can use Data Interpreter to detect sub-tables in your data as well as remove extraneous information to help prepare your data for analysis. When you turn on Data Interpreter, it detects these sub-tables and lists them as new tables in the Tables section of the Connections pane.

You can then drag them into the Flow pane. If you are using Tableau Prep Builder version 2018.1.2 or later, you can select the Wildcard union option in the Multiple Files tab to include all found sub-tables in your flow. For more information about using Wildcard union in the Input step see Union files and database tables in the Input step on the next page.

If you turn Data Interpreter off, these tables are removed from the Connections pane. If these tables are already used in the flow, this will result in flow errors from the missing data.

**Note:** Currently, Data Interpreter only detects sub-tables in your Excel spreadsheets and doesn't support specifying the starting row for text files and spreadsheets.

The example below shows the results of using Data Interpreter on an Excel spreadsheet in the Connections pane. Data Interpreter detected two additional sub-tables.

<table>
<thead>
<tr>
<th>Before Data Interpreter</th>
<th>After Data Interpreter</th>
</tr>
</thead>
</table>

NYT Feb 18, 2018

Clean 1

Data Interpreter detected two additional sub-tables.
To use Data Interpreter, complete the following steps:

1. Select **Connect to Data** then select **Microsoft Excel**.
2. Select your file and click **Open**.
3. Select the **Use Data Interpreter** check box.
4. Drag the new table to the **Flow** pane to include it in your flow. To remove the old table, right-click the Input step for the old table and select **Remove**.

### Union files and database tables in the Input step

When working with multiple files or database tables from a single data source, you can search for files or tables using a wildcard search and then union the data to include all of the file or table data in the Input step. To union files, the files must be in the same parent or child directory.

To union data tables, the tables must be in the same database and the database connection must support using a wildcard search to union. The following databases support this type of union:
If you add or remove files or tables after you create the union you can refresh the Input step to update your flow with the new or changed data.

**Note:** Currently, this feature applies only to Excel and .csv (text) files and data tables stored in the specific databases listed above. This option is not available for Tableau data extracts.

Wildcard union for files is available in Tableau Prep Builder version 2018.1.2 and later. Wildcard union for database tables is available in Tableau Prep Builder version 2018.3.1 and later. Editing a flow connection with this type of union in a prior version can result in errors.

If you need to union data from different data sources, you can do that using a Union step. For more information about creating Union steps, see [Join or Union Data](#) on page 201.

### Union files

By default, Tableau Prep Builder unions all .csv files in the same directory as the .csv file you connected to or all the sheets in the Excel file you connected to. If you use Data Interpreter to clean Excel files and are using Tableau Prep Builder version 2018.1.2 or later, you can use the wildcard search to union and add any sub-tables that Data Interpreter found.

If you want to change the default union, use the following criteria to find the files or sheets you want to include in the union:

- **Search in:** Select the directory to use to search for files. Select the **Include subfolders** check box to include files in the sub-directory of the parent folder.

- **Files:** Select whether to include or exclude the files that match the wildcard search criteria.

- **Matching Pattern (xxx*)&**: Enter a wildcard search pattern to find files that have those
characters in the file name. For example, if you enter ord* all files that include the file name are returned. Leave this field blank to include all of the files in the specified directory.

To use wildcard search to union files:

1. Click the Add connection button and under Connect, click Text File for .csv files or Microsoft Excel for Excel files, and then select a file to open.

2. In the Input pane, select the Multiple Files tab, and then select Wildcard union.

The example below shows a wildcard union using a matching pattern. The plus sign on the file icon on the Orders_Central Input step in the Flow pane indicates that this step includes a wildcard union. The files in the union are listed under Included files.
3. Use the search, file and matching pattern options to find the files that you want to union.

4. Click **Apply** to union the files.

When you add a new step to the flow, you can see all the files added to the data set in the **File Paths** field in the Profile pane. This field is added automatically.
Union database tables (version 2018.3.1 and later)

1. Click the **Add connection** button and under **Connect**, connect to a database that supports wildcard union.
2. Drag a table to the flow pane.
3. In the **Input** pane, select the **Multiple Tables** tab, and then select **Wildcard union**.
4. Use search, **Tables** and **Matching Pattern** options to find the tables that you want to union.

Only tables that display in the **Connections** pane in the **Tables** section can be included in the union. Wildcard search doesn't search across schemas or across the database connection to find tables.

5. Click **Apply** to union the table data.

When you add a new step to the flow, you can see all the tables added to the data set in the **Tablea Names** field in the Profile pane. This field is added automatically.
Merge fields after a union

After you create a union in the input step, you might want to merge fields. You can do this in any subsequent step, except for the Input or Output steps. For more information, see Additional merge field options on page 212.

Join data in the Input step

In Tableau Prep Builder version 2019.1.3 and later, when you connect to databases that include tables with relationship data, Tableau Prep Builder can now detect and show which fields in a table are identified as the unique identifier and which fields are identified as a related field as well as show the related table names for these fields.

A new column called Linked Keys shows in the Input pane and shows the following relationships if they exist:

- **Unique identifier.** This field uniquely identifies each row in the table. There can be multiple unique identifiers in a table. The values in the fields must be unique and cannot be blank or null.

- **Related field.** This field relates the table to another table in the database. There can be multiple related fields in a table.
Both Unique Identifier and related field. The field is a unique identifier in this table and also relates the table to another table in the database.

You can leverage these relationships to quickly find and add the related tables to your flow or create joins from the Input step. This feature is available for any supported database connector where table relationships are defined.

1. Connect to a database (such as Microsoft SQL Server) that contains relationship data for fields, such as unique identifiers or related fields (foreign key).

2. In the Input pane, click on a field that is marked as a related field or as both a unique identifier and related field.

   A dialog opens that shows a list of related tables.

3. Hover on the table that you want to add or join and click the plus button to add the table to your flow, or click the join button to create a join with the selected table.

If you create a join, Tableau Prep Builder uses the defined field relationship to join the tables and shows you a preview of the join clauses that it will use to create the join.
4. Alternatively, you can join related tables from the menu in the Flow pane. Hover over a step until the plus icon appears, then select **Add Join** to see a list of related tables. Tableau Prep Builder creates the join based on the fields that make up the relationship between the two tables.

**Note:** If your table doesn't have table relationships defined, this option is not available.

For more information about working with joins, see **Join your data** on page 201.

**Configure your data set**

To determine how much of your data set to include in the flow, you can configure your data set. When you connect to your data or drag tables into the **Flow** pane, an Input step is automatically added to the flow. This is always the first step in your flow. You can right-click the Input step to rename or remove it. If you’re connected to an Excel or text file, you can also refresh the data from the Input step.
In the Input step, you can see the metadata profile for your data set. Here you can search for fields, see sample values, and perform actions to reduce the size of your data set, such as selecting the fields to include, selecting the data sample to work with, or applying filters to selected fields or rows.

You can also configure the field properties by changing the data type or field name for fields. For text files, you can also configure the text settings.

**Note:** Field values that include square brackets are automatically converted to parentheses.

Connect to a custom SQL query

If your database supports using custom SQL, you will see **Custom SQL** displayed near the bottom of the **Connections** pane. Double-click **Custom SQL** to open the **Custom SQL** tab where you can enter queries to preselect data and use source-specific operations. After the query retrieves the data set, you can select the fields to include, apply filters, or change the data type before adding the data to your flow.
For more information about using custom SQL, see Connect to a Custom SQL Query in the Tableau Desktop and Web Authoring Help.

Select fields to include in the flow

The Input pane shows you a list of fields in your data set. You can use the Search field to find fields in the list, and then use the check boxes to select the fields to include or exclude. To include or exclude all fields from the flow, toggle the check box at the top left of the grid.

Configure field properties

When you work with text files, you see a Text Settings tab where you can edit your connection and configure text properties, such as the field separator for text files. You can also edit the file connection in the Connections pane.

When you work with text or Excel files, you can correct data types that have been inferred incorrectly before you even start your flow. Data types can always be changed in subsequent steps in the Profile pane after you start your flow.

Configure text settings in text files

To change the settings used to parse text files, select from the following options:

- **First line contains header** (default): Select this option to use the first row as the field labels.

- **Generate field names automatically**: Select this option if you want Tableau Prep Builder to auto-generate the field headers. The field naming convention follows the same model as Tableau Desktop. For example F1, F2, and so on.
**Field Separator:** Select a character from the list to use to separate the columns. Select Other to enter a custom character.

**Text Qualifier:** Select the character that encloses the values in the file.

**Character Set:** Select the character set that describes the text file encoding.

**Locale:** Select the locale to use to parse the file. This setting indicates which decimal and thousand separator to use.

**Change data types**

To change the data type for a field, do the following:

1. Click the data type for the field.
2. Select the new data type from the menu.

![Data Type Selection](image)

**Change field names**

To change the name of a field, in the Field Name column, select the name, and then type the new name in the field. An annotation is added in the field grid and in the flow pane to the left of the Input step. Your changes are also tracked in the Changes pane.
Set your data sample size

By default, Tableau Prep Builder limits the data included in the flow to a representative sample of your data set to maintain peak performance. The data sample is determined by calculating the optimal number of rows based on the total number of fields in the data set and the data types for those fields. Tableau Prep Builder then retrieves the top number of rows for the calculated amount as quickly as possible.

The resulting data sample may include all the rows you need, or it may not, depending on how the sample was calculated and returned. If you don’t see the data that you expect, you can change the data sample settings to run the query again.

**Note:** If your data is sampled, a Sampled indicator shows in the Profile pane and persists for every step you add. Any changes you make apply to the sample you are working with in the flow. All changes apply to your entire data set when you run the flow.

To change your data sample settings, select an Input step, then on the Data Sample tab select from the following options:

- **Default sample amount** (default): Tableau Prep Builder calculates the total number of rows to return.
- **Use all data**: Retrieve all rows in your data set regardless of size. This can impact performance or cause Tableau Prep Builder to time out.

**Note:** To maintain performance, even if you select this setting, a data sample limit of 1 million rows is applied to Aggregate, and Union step types and a data sample limit of 3 million rows is applied to Join and Pivot step types.

- **Fixed number of rows**: Select the number of rows to return from the data set. The recommended number of rows is 1 million or less. Setting the number of rows to more than 1 million can impact performance.
- **Quick select** (default): The database returns the number of rows requested as quickly as possible. This might be the first N number of rows or the rows that the database had cached in memory from a previous query.
- **Random sample**: The database returns the number of rows requested but looks at
every row in the data set and returns a representative sample from all of the rows. This option may impact performance when the data is first retrieved.

Apply filters to fields in the Input step

To filter a field, do the following:

1. In the toolbar click Filter Values.

2. Enter your filter criteria in the calculation editor.

Additional filtering capabilities are available in other steps in the flow. For more information, see Filter values on page 151.
Examine and Filter Your Data

Get a good understanding of the composition of your data to better understand changes you need to make, and the effect of the operations you include in the flow.

See size details about your data

After you connect to your data, add a table to the flow, and then add a step. You can use the Profile pane to see the current state and structure of your data and spot nulls and outliers.

- **Number of fields and rows**: In the upper-left corner of the Profile pane you can find information that summarizes the number of fields and rows in the data at a particular point in the flow. Tableau Prep Builder rounds to the nearest thousand. In the example below, there are 21 fields and 3000 rows in the data set.

![Orders_East Results](image)

When you hover over the number of fields and rows, you can see the exact number of rows (in this example, 2848):
• **Data set size:** Work with a subset of your data by specifying the number of rows to include in the Data Sample tab in the Input pane.

• **Sampled:** To enable you to interact directly with your data, Tableau Prep Builder works with a subset of your raw data. The number of rows is determined by the data types and number of fields that are being rendered. String fields take more storage space than integers, so if you have 10 fields of strings in your data set, you might get fewer rows than if you had 10 fields of integers. A Sampled indicator displays next to the size details in the Profile pane to indicate that this is a subset of your data set. For more information about data samples, see "Set your data sample size" in Connect to Data on page 109.

• **Number of unique values:** The number next to each field header represents the distinct values that are contained within that field. Tableau Prep Builder rounds to the nearest thousand. In the example below, there are 3,000 distinct values that are represented in the Description field, but if you hover over the number, you can see the
exact number of unique values.

Review the data types assigned to your data

Like Tableau Desktop, Tableau Prep Builder interprets the data in your fields when you drag a connection to the Flow pane and automatically assigns a data type to it. Because different databases can handle data in different ways, Tableau Prep Builder's interpretation might not always be correct.

To change a data type, click the data type icon and select the correct data type from the context menu. You can change string or integer data types to Date or Date & Time, and Tableau Prep Builder will trigger Auto DateParse to change these data types. Like Tableau Desktop, if the change is not successful you will see Null values in the fields instead and you can create a calculation to make the change.
For more information about using DateParse, see Convert a Field to a Date Field in the Tableau Desktop and Web Authoring Help.

Assign data roles to your data

To make it easier to identify field values that aren't valid, you can assign a data role to your field the same way you assign a data type. This tells Tableau Prep Builder what your data values represent so it can automatically validate values and highlight ones that aren't valid for that role.
Data roles tell Tableau Prep Builder what the field values mean or represent. For example if you have field values for geographical data, you can assign a data role of **City** and Tableau Prep Builder compares the values in the field to a set of known domain values or patterns to identify values that don’t match.

**Note:** Each field is analyzed independently so a City value of "Portland" in State "Washington" in Country "USA" might not be a valid city and state combination, but it won’t be identified that way because it is a valid city name.

Tableau Prep Builder supports the following data roles:

- Email
- URL
- Geographic roles (Based on current geographic data and is the same data used by Tableau Desktop)
  - Airport
  - Area code (U.S.)
  - CBSA/MSA
  - City
  - Congressional District (U.S.)
Tableau Prep Help

- Country/Region
- County
- NUTS Europe
- State/Province
- Zip code/Postal code

**Tip:** In Tableau Prep Builder version 2019.1.4 and later, if you assign a geographic role to a field, you can also use that data role to match and group values with the standard value defined by your data role. For more information about grouping values using data roles, see [Group similar values by data role](#) on page 195.

To set a data role for your field, do the following:

1. In the Profile pane, Results pane or data grid, click the data type for the field.
2. Select the data role for the field.

Tableau Prep Builder compares the field's data values to known domain values or patterns (for email or URL) for the data role you select and marks any values that don't match with a red exclamation point.

3. Click the drop-down arrow for the field and from the **Show Values** section select an option to show all values or only values that are valid or not valid for the data role.
4. Use the cleaning options on the **More options** menu for the field to correct any values that aren't valid. For more information about how to clean your field values see [Apply cleaning operations on page 170](#).

### See the distribution of values or unique values

By default, Tableau Prep Builder groups numerical, date, and datetime values in a field into buckets. These buckets are also known as bins. The bins ensure that you can see the distribution of values as a whole and quickly identify outliers and null values. The bin size is calculated based on the minimum and maximum values in the field, and null values are always shown at the top of the distribution.

For example, order and ship dates are summarized or "binned" by year. Each bin represents a year from January of the beginning year to January of the following year and labeled accordingly. Because there are sales dates and ship dates that fall in the latter part of 2018 and 2019, a bin is created for the following year for those values.
If a discrete (or categorical) data field contains many rows or has a distribution that is large enough that it can’t be displayed in the field without scrolling, you can see a summarized distribution to the right of the field. You can click and scroll through the distribution to target specific values.

When your data contains numeric or date fields, you can toggle to display the detailed (discrete) version of the values or a summarized (continuous) version of the values. The summarized view shows you the range of values in a field and the frequency with which certain values appear.

This toggle can help you isolate unique values (like the number of “3” records in a field) or the distribution of values (like the sum of all “3” records in a field).

To toggle your view:
1. In the Profile pane, Results pane or data grid, click the **More options** menu for a numeric or date field.

![Image of More options menu](image)

2. In the context menu, select **Detail** to see the detailed version of the values, or **Summary** to see the distributed version of the values.

---

**Sort values and fields**

Sort options on a profile card let you sort the bins (the count of values represented by the distribution bars) in ascending or descending order, or the individual field values in alphabetical order.

![Image of sorting options](image)

If you want to rearrange the order of your fields, in the **Profile** pane, **Results** pane or **Data** grid simply select a profile card or field in the data grid and drag it until you see the black target.
line appear. Then drop it into place. The Profile pane and data grid are synced so the field will appear in the same order in both places.

Highlight fields and values in a flow

Tableau Prep Builder makes it easy to find fields and values in your flow data. Trace where a field originated and where it is used throughout the flow in the flow pane, or click individual values in a profile card or in the data grid to highlight related or identical values.

Trace fields in a flow

In Tableau Prep Builder version 2018.3.3 and later, you can highlight everywhere a field is used in a flow, even where it originated to help you track down missing values or troubleshoot a flow when you aren’t seeing the results you expect.

Click on a field in the Profile pane in a cleaning step or in the Results pane in any other step type and the flow pane will highlight the path where that field is used.

**Note:** This option is not available for Input or Output step types.
See related values

You can use highlighting to find related values across fields. When you click a value in the Profile card in the Profile pane or Results pane, all the related values in the other fields are highlighted in blue. The blue color shows the relationship distribution between the value you selected and the values in the other fields.

For example, to highlight related values, in the Profile pane, click a value in a field. The related values in other fields turn blue and the proportion of the bar highlighted in blue represents the degree of association.

Highlight identical values

When you select a value in the data grid, all identical values are highlighted too. These highlights help you identify patterns or irregularities in your data.
Keep or remove fields

As you work with your data in your flow you might want to remove unwanted fields. In the Profile pane or the data grid in any cleaning or action step, select one or more fields and right-click or Ctrl-click (MacOS) and select **Remove** to remove the selected fields, or select **Keep Only** (version 2019.2.2 and later) to keep only the selected fields and remove all unselected fields.
Duplicate fields (version 2019.2.3 and later)

If you wanted to experiment with your cleaning operations on a field but don't want to change the original data, you can duplicate your fields.

1. In the profile pane, data grid or results pane, select the field you want to duplicate.

2. From the More options menu, select Duplicate Field.

A new field is created with the same name and a modifier. For example, "Ship Date -1".
Search for fields and values

In the Profile pane or Results pane, you can search for fields or values of particular interest to you and use the search results to filter your data.

To search for fields, enter a full or partial search term in the search box on the toolbar.

To search for a value in a field:
1. Click the Search icon for a field, and enter a value.

2. To use advanced search options, click the **Search options...** button.

3. To use the search results to filter the data, select **Keep Only** or **Exclude**.

   In the **Flow** pane, a filter icon appears above affected steps.
Filter values

An easy way to filter a value is to select it in a profile card, data grid or results card, right-click, and then select **Keep Only** or **Exclude**. You can also select **Edit Value** to edit the value in-line.

You can filter data at any step in the flow. To add a filter, click the **More options** ⋯ menu on the profile card, in the data grid or in the results pane.

**Note:** To apply a filter in the data grid **More options** ⋯ menu, click the **Hide profile pane** button and then click **More options** ⋯ for the field you want to filter.
Filters available for each data type

<table>
<thead>
<tr>
<th>Data type</th>
<th>Available filters</th>
</tr>
</thead>
<tbody>
<tr>
<td>String</td>
<td>Calculation, Wildcard Match, Null Values, Selected Values</td>
</tr>
<tr>
<td>Number</td>
<td>Calculation, Range of Values, Null Values, Selected Values</td>
</tr>
<tr>
<td>Date and Date &amp; Time</td>
<td>Calculation, Range of Dates, Relative Date, Null Values, Selected Values</td>
</tr>
</tbody>
</table>

Use a Calculation filter

When you select Calculation, the Add Filter dialog box opens. Enter the calculation, verify that it’s valid, and click Save.
Use a Selected Values filter (version 2019.2.3 and later)

When you select **Selected Values**, you can pick and choose the values that you want to keep or exclude for a field, even values that aren't in your sample. In the right pane, click the **Keep Only** or **Exclude** tab to select your action, then enter search terms to search for values or click **Add a value** to add values that are in your data set but aren't included in your sample. Click **Done** to apply your filter.

Note: this filter options isn't available for **Aggregation** or **Pivot** step types.
Use a Range of Values filter

When you select **Range of Values**, you can specify a range or set minimum or maximum values.

Use Range of Dates filter

When you select **Range of Dates**, you can specify a range of dates or set a minimum or maximum date.
Use a Relative Date filter

When you select Relative Dates, you can filter the date range based on year, quarter, month, week, or day. You can also configure an anchor relative to a specific date, and include null values.

Use a Wildcard Match filter

When you select Wildcard Match, you can filter the field values to keep or exclude only those values that match your criteria. In the filter editor, select the Keep Only or Exclude tab, enter a value to match and then set the criteria to return the values you are looking for.

The filtered results display in the left pane of the filter editor so that you can review and experiment with your results. Once you have the results you want, click Done to apply your change.
Use a Null Values filter

When you select Null Values you can filter the values in the selected field to show only null values or exclude all null values.
Clean and Shape Data

Clean and shape data to make it easier to combine with other data or to make it simpler for other people to understand.

Build your flow

After you connect to the data that you want to include in your flow, begin cleaning and shaping your data by adding new steps to the flow. Hover over a step until the plus icon appears, and then click the icon to display options.

Use these options to build your flow:

- **Add Branch**: Split your flow into different branches.
- **Add Step** or **Insert Step**: Add a step to a new flow or insert a step into an existing flow to perform cleaning operations.
- **Add Aggregate**: Select the step that includes the data you want to aggregate or group.
- **Add Pivot**: Select the step that includes the data you want to pivot.
- **Add Join**: Select the step that you want to join data with. In Tableau Prep Builder version 2019.1.3 and later, if you connect to databases that include tables with relationship data you can also create a join from the menu in the Flow pane. For more information about joining tables using this method, see [Join data in the Input step](#) on page 128.
As an alternative, you can drag and drop a step to join files. In the following example, we’re dragging the Orders_Central Input step and dropping it on Join:

For more information about creating a join, see Join or Union Data on page 201.

- **Add Union**: Select the step that you want to union data with. As an alternative, you can drag and drop a step to union files. For more information about creating a union, see Join or Union Data on page 201.

- **Add Output**: Select this option to save the output to an extract file (.tde or .hyper), a .csv file, or publish the output as a data source to a server.

### Change the color scheme

Tableau Prep Builder assigns each step in your flow a color by default. This color scheme is applied throughout the flow to help you keep track of your data throughout the flow as you apply cleaning steps, join, union or aggregate the data so you know which files are impacted by your operations.

To select a different color scheme for your steps do the following:

1. Select one or more steps.
2. Right-click on a selected step and select **Edit Step Color**.
3. Click on a color in the color palette to apply it.

To reset the step color back to the default color, do one the following:

- Click **Undo** from the top menu.
- Cntrl+Z or Command-Shift-Z (Mac).
- Select the steps you changed, right-click on a selected step and select **Edit Step Color**, then select **Reset Step Color** from the bottom of the color palette.
Add a description

As you build your flow and perform various cleaning operations, you might want to add a description to help others who might later look at or work with your flow to better understand your steps.

You can add a description to any individual step in your flow directly on the Flow pane. The description can be up to 200 characters long.

When you add a description, a message icon is added underneath the step. Click the icon to show or hide the description text in the Flow pane.

To add a description to a step do the following:

1. In the Flow pane, select a step.
2. Do one of the following:
   - Right-click the step and select Add Description from the menu.
   - Double-click in the name field for the step, then click on Add a description.
3. Type your description in the text box.
4. Click outside the text box or press Enter to apply your changes. By default, the description displays underneath the step. To hide the description click the message icon.

5. To edit or delete the description, right-click on the step or description then from the menu, select Edit Description or Delete Description.

Reorganize the layout of your flow (version 2019.2.2 and later)

As you build a flow, Tableau Prep Builder adds your steps and branches using a default layout. Each flow processes your data from left to right, with Input steps beginning on the far left of the canvas and Output steps ending on the right side of the canvas. However if you build large, complex flows, they can quickly become hard to follow.

To clean up the layout of your flow, you can select and move steps so the flow layout is organized in a way that makes sense to you. For example, you can fix crossed flow lines, move your flow steps to clean up extra white space, or rearrange your flow steps to show a clear sequence of events.

For example, the following flow is confusing and hard to follow:
To clean up this flow you can select and drag steps up, down, left or right and drop them to a new location in the canvas. Flow steps can’t be moved to a position that disrupts the left-to-right process flow. For example, you can’t drag a union step that is positioned before a join step, to a position that is after that join step in the flow.

When dragging flow steps to an allowed location, an orange box displays. If the location isn’t allowed, no orange box displays and the steps return to their original location when you try to drop them.

Watch "Reorganize flow steps" in action

The following example shows rearranging a flow using drag and drop.

To move steps in your flow, do the following:
1. In the Flow pane, select the steps that you want to move. You can click on a specific step, drag to select multiple steps, or Ctrl-click or Cmd+Click (MacOS) to select steps that aren't next to each other.

2. Drag and drop the steps to the new location.

**Note:** If you don’t like the reorganization moves that you make you can click **Undo** in the top menu to reverse them. However, if you perform cleaning actions in between moving steps, you may undo those actions as well. The **Undo** option reverses your actions in the order that you performed them.

### Remove steps from the flow

At any point in the flow, you can remove steps or the flow lines between steps.

- To remove a step or flow line, select the step or line you want to remove, right-click the element, and then select **Remove**.
- To remove multiple steps or flow lines, do one of the following:
  - Use your mouse to drag and select a whole section of the flow. Then right-click on one of the selected steps and select **Remove**
  - Press **Ctrl+A** or **Cmd+A** (MacOS) to select all elements in the flow, or press **Ctrl+click** or **Cmd+Click** (MacOS) to select specific elements, and then press the **Delete** key.

### Pivot your data

Sometimes analyzing data from a spreadsheet or crosstab format can be difficult in Tableau. Tableau prefers data to be "tall" instead of "wide", which means that you often have to pivot your data from columns to rows so that Tableau can evaluate it properly.

However, you may also have scenarios where your data tables are tall and narrow and are too normalized to properly analyze. For example, a sales department that tracks advertising spend in two columns, one called **Advertising** that contains rows for **radio**, **television** and **print** and one column for total spent. In this type of scenario, to analyze this data as separate measures you would need to pivot that row data to columns.
Tableau Prep Help

But what about pivoting larger data sets or data that changes frequently over time? You can use a wildcard pattern match to search for fields that match the pattern and automatically pivot the data.

Use one of the following options when pivoting your data:

- Pivot columns to rows
- Use wildcard search to instantly pivot fields based on a pattern match (Tableau Prep Builder version 2019.1.1 and later).
- Pivot rows to columns (Tableau Prep Builder version 2019.1.1 and later).

No matter how you pivot your fields, you can interact directly with the results and perform any additional cleaning operations to get your data looking just the way you want it. You can also use Tableau Prep Builder’s smart default naming feature to automatically rename your pivoted fields and values.

**Note:** Pivoting columns to rows on multiple groups of fields is not supported in Tableau Prep Builder 2018.1.1. Editing a pivot that includes pivoted columns on multiple field groups in Tableau Prep Builder version 2018.1.1 can result in errors or unexpected results.

**Pivot columns to rows:**

Use this pivot option to go from wide data to tall data. Pivot from columns to rows on one or more groups of fields. Select the fields that you want to work with and pivot the data from columns to rows.

1. Connect to your data source.
2. Drag the table that you want to pivot to the Flow pane.
3. Click the plus icon, and select Add Pivot from the context menu.
4. (Optional) In the Fields pane, enter a value in the Search field to search the field list for fields to pivot.
5. (Optional) Select the **Automatically rename pivoted fields and values** check box to enable Tableau Prep to rename the new pivoted fields using common values in the data. If no common values are found, the default name is used.
6. Select one or more fields from the left pane, and drag them to the Pivot1 Values column.
in the **Pivoted Fields** pane.

7. (Optional) In the **Pivoted Fields** pane, click the plus icon to add more columns to pivot on, then repeat the previous step to select more fields to pivot. Your results appear immediately in both the **Pivot Results** pane and the data grid.

   **Note:** You must select the same number of fields that you selected in Step 5. For example if you selected 3 fields to initially pivot on, then each subsequent column that you pivot on must also contain 3 fields.

8. If you didn’t enable the default naming option or if Tableau Prep couldn’t automatically detect a name, edit the names of the fields. You can also edit the names of the original fields in this pane to best describe the data.

9. (Optional) Rename the new Pivot step to keep track of your changes. For example "Pivot months".

10. To refresh your pivot data when data changes, run your flow. If new fields are added to your data source that need to be added to the pivot, manually add them to the pivot.

**Example:** Pivoting on multiple fields

This example shows a spreadsheet for pharmaceutical sales, taxes and totals by month and year.

![Spreadsheet example](image)

By pivoting the data you can create rows for each month and year and individual columns for sales, taxes and totals so that Tableau can more easily interpret this data for analysis.
Watch "pivot on multiple field" in action.

Use wildcard search to pivot

In Tableau Prep Builder version 2019.1.1 and later if you work with large data sets or if your data frequently changes over time, you can use a wildcard search when pivoting columns to rows to instantly pivot your data based on a wildcard pattern match. If new fields are added or removed that match the pattern, Tableau Prep Builder detects the schema change when the flow is run and the pivot results are automatically updated.

1. Connect to your data source.
2. Drag the table that you want to pivot to the Flow pane.
3. Click the plus icon, and select Add Pivot from the context menu.
4. In the **Pivoted Fields** pane, click on the link **Click here to create wildcard pivot**.

5. Enter a value or partial value that you want to search for. For example, enter `Sales_` to match fields that are labeled as `sales_2017`, `sales_2018` and `sales_2019`. Do not use asterisks to match the pattern unless they are part of the field value that you are searching for. Instead click the **Search Options** button to select how you want to match the value. Then press **Enter** to apply the search and pivot the matching values.
6. (Optional) In the **Pivoted Fields** pane, click the plus icon to add more columns to pivot on, then repeat the previous step to select more fields to pivot.

7. If you didn't enable the default naming option or if Tableau Prep Builder couldn't automatically detect a name, edit the names of the fields.

8. To refresh your pivot data when data changes, run your flow. Any new fields added to your data source that match the wildcard pattern are automatically detected and added to the pivot.

9. If the results aren't what you expect, try one of the following options:
   - Enter a different value pattern in the Search field and press enter. The pivot will automatically refresh and show the new results.
   - Manually drag additional fields to the **Pivot1 Values** column in the **Pivoted Fields** pane. You can also remove fields that were added manually by dragging them off the **Pivot1 Values** column and dropping them in the **Fields** pane.

**Note:** Fields that were added from the wildcard search results can't be removed by dragging them off the **Pivot1 Values** column. Instead try using a more specific pattern to match the search results you are looking for.

### Pivot rows to columns

In Tableau Prep Builder version 2019.1.1 you can pivot rows to columns. Use this option if your data is too normalized and you need to create new columns - going from tall data to wider data.

For example if you have advertising costs for each month that includes all advertising types in one column, if you pivot the data from rows to columns you can then have a separate column for each advertising type instead, making the data easier to analyze.

<table>
<thead>
<tr>
<th>Month</th>
<th>Cost</th>
<th>Advertising Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>12187</td>
<td>Radio</td>
</tr>
<tr>
<td>January</td>
<td>45956</td>
<td>Television</td>
</tr>
<tr>
<td>January</td>
<td>9762</td>
<td>Print</td>
</tr>
<tr>
<td>February</td>
<td>45918</td>
<td>Print</td>
</tr>
<tr>
<td>February</td>
<td>23987</td>
<td>Television</td>
</tr>
<tr>
<td>February</td>
<td>8907</td>
<td>Radio</td>
</tr>
<tr>
<td>March</td>
<td>10987</td>
<td>Print</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Month</th>
<th>Radio</th>
<th>Television</th>
<th>Print</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>12187</td>
<td>45956</td>
<td>9762</td>
</tr>
<tr>
<td>February</td>
<td>8907</td>
<td>23987</td>
<td>45918</td>
</tr>
<tr>
<td>March</td>
<td>Null</td>
<td>Null</td>
<td>10987</td>
</tr>
</tbody>
</table>
You can select one field to pivot on. The field values for that field are then used to create the new columns. Then, select a field to use to populate the new columns. These field values are aggregated and you can select the type of aggregation to apply.

Because aggregation is applied, pivoting columns back to rows won’t reverse this pivot action. To reverse a row to column pivot type, you will need to undo the action. Either click the **Undo** button on the top menu, remove the fields from the **Pivoted Fields** pane or delete the pivot step.

1. Connect to your data source.
2. Drag the table that you want to pivot to the **Flow** pane.
3. Click the plus icon, and select **Add Pivot** from the context menu.
4. In the **Pivoted Fields** pane, select **Rows to Columns** from the drop-down list.
5. (Optional) In the **Fields** pane, enter a value in the **Search** field to search the field list for fields to pivot.
6. Select a field from the left pane, and drag it to the **Field that will pivot rows to columns** section in the **Pivoted Fields** pane.

**Note:** If the field you want to pivot on has a data type of date or datetime, you will need to change the data type to string to pivot it.

The values in this field will be used to create and name the new columns. You can change the column names in the **Pivot Results** pane later.

7. Select a field from the left pane and drag it to the **Field to aggregate for new**
columns section in the Pivoted Fields pane. The values in this field are used to populate the new columns created from the previous step.

A default aggregation type is assigned to the field. Click the aggregation type to change it.

8. In the Pivot Results pane, review the results and apply any cleaning operations to the new columns that you created.
9. If the field being pivoted has a change in its row data, right-click or Command-click (Mac) on the Pivot step in the flow pane and select Refresh.

Apply cleaning operations

You clean data by applying cleaning operations such as filtering, adding, renaming, splitting, grouping, or removing fields. In prior releases, cleaning operations could only be performed in the Clean step type. In Tableau Prep Builder version 2018.2.1 and later, you could also perform cleaning operations in the data grid in a cleaning step.

In Tableau Prep Builder version 2018.3.1 and later, you can perform cleaning operations in most step types in your flow. The one exception is the output step.

The following table shows which cleaning operations are available in each step type:

<table>
<thead>
<tr>
<th></th>
<th>Input</th>
<th>Clean</th>
<th>Aggregate</th>
<th>Pivot</th>
<th>Join</th>
<th>Union</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Group and Replace</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Clean</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Split Values</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Rename Field</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
As you make changes to your data, annotations are added to the corresponding step in the Flow pane and an entry is added in the Changes pane to track your actions. If you make changes in the Input step, the annotation shows to the left of the step in the Flow pane and is also shown in the Input profile in the field list.

The order that you apply your changes matters. Changes made in Aggregate, Pivot, Join and Union step types are performed either before or after those reshaping actions, depending on where the field is when you make the change. Where the change was made is shown in the Changes pane for the step.

For example, if you change a field value in a join clause, the change is performed prior to the join action to give you the corrected results.

The following table shows where the cleaning action is performed in Aggregate, Pivot, Join and Union step types depending on where the field is in the step.
<table>
<thead>
<tr>
<th>Field Location:</th>
<th>Groupe-d fields</th>
<th>Aggreg-at ed fields</th>
<th>Not in pivot</th>
<th>Created from pivot</th>
<th>Included in one table*</th>
<th>Included in both tables*</th>
<th>Mis-matched fields</th>
<th>Combined fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter</td>
<td>Before Aggreg- ation</td>
<td>After Aggreg- ation</td>
<td>Before Pivot</td>
<td>After Pivot</td>
<td>Before Join</td>
<td>After Join</td>
<td>Before Union</td>
<td>After Union</td>
</tr>
<tr>
<td>Group and Replace</td>
<td>NA</td>
<td>NA</td>
<td>Before Pivot</td>
<td>After Pivot</td>
<td>Before Join</td>
<td>After Join</td>
<td>Before Union</td>
<td>After Union</td>
</tr>
<tr>
<td>Clean</td>
<td>NA</td>
<td>NA</td>
<td>Before Pivot</td>
<td>After Pivot</td>
<td>Before Join</td>
<td>After Join</td>
<td>Before Union</td>
<td>After Union</td>
</tr>
<tr>
<td>Split Values</td>
<td>NA</td>
<td>NA</td>
<td>Before Pivot</td>
<td>After Pivot</td>
<td>Before Join</td>
<td>After Join</td>
<td>Before Union</td>
<td>After Union</td>
</tr>
<tr>
<td>Rename Field</td>
<td>Before Aggreg- ation</td>
<td>After Aggreg- ation</td>
<td>Before Pivot</td>
<td>After Pivot</td>
<td>Before Join</td>
<td>After Join</td>
<td>Before Union</td>
<td>After Union</td>
</tr>
<tr>
<td>Duplicate Field</td>
<td>NA</td>
<td>NA</td>
<td>Before Pivot</td>
<td>After Pivot</td>
<td>Before Join</td>
<td>After Join</td>
<td>Before Union</td>
<td>After Union</td>
</tr>
<tr>
<td>Keep</td>
<td>After</td>
<td>After</td>
<td>Before</td>
<td>After</td>
<td>Before</td>
<td>After</td>
<td>Before</td>
<td>After</td>
</tr>
<tr>
<td>Only Field</td>
<td>Aggregation</td>
<td>Aggregation</td>
<td>ore</td>
<td>Pivot</td>
<td>e Join</td>
<td>Join</td>
<td>Union</td>
<td>Union</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-----</td>
<td>-------</td>
<td>--------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Remove Field</td>
<td>Remove(s from Aggregation)</td>
<td>Remove(s from Aggregation)</td>
<td>Before Pivot</td>
<td>After Pivot</td>
<td>Before Join</td>
<td>After Join</td>
<td>Before Union</td>
<td>After Union</td>
</tr>
<tr>
<td>Create Calculated Field</td>
<td>NA</td>
<td>NA</td>
<td>Before Pivot</td>
<td>After Pivot</td>
<td>Before Join</td>
<td>After Join</td>
<td>Before Union</td>
<td>After Union</td>
</tr>
<tr>
<td>Edit Value</td>
<td>NA</td>
<td>NA</td>
<td>Before Pivot</td>
<td>After Pivot</td>
<td>Before Join</td>
<td>After Join</td>
<td>Before Union</td>
<td>After Union</td>
</tr>
<tr>
<td>Change Data Type</td>
<td>Before Aggregation</td>
<td>After Aggregation</td>
<td>Before Pivot</td>
<td>After Pivot</td>
<td>Before Join</td>
<td>Before Join</td>
<td>Before Union</td>
<td>After Union</td>
</tr>
</tbody>
</table>

**Note:** For joins, if the field is a calculated field that was created using a field from one table, the change is applied before the join. If the field is created with fields from both tables, the change is applied after the join.

**Apply cleaning operations in an input step**

Only some cleaning operations are available in an Input step. You can make any of the following changes in the Input field list. Your changes are tracked in the **Changes** pane and annotations are added to the left of the Input step in the **Flow** pane and in the Input field list.

- **Filter:** Click **Filter Values** in the toolbar then enter your filter criteria in the calculation editor.
- **Rename Field:** In the **Field Name** field, double-click (CTRL+click on Mac) on the field...
name and enter a new field name.

- **Change Data Type**: Click on the data type for the field and select a new data type from the menu.

- **Remove Field**: Clear the check box next to the fields that you don’t want to include in your flow.

---

**Apply cleaning operations in other steps**

All of the cleaning operations are available in the other step types. To apply cleaning operations to fields, use the toolbar options or click **More options** on the field profile card to open the menu. In Aggregate, Pivot, Join and Union step types, the **More options** menu is available on the profile cards in the Results pane.

**Profile pane toolbar**

**Drop-down menu**
In a cleaning step you can also apply cleaning operations in the data grid. Just click the **Hide profile pane** button to collapse the Profile pane then click **More options** on a field to open the menu.

**Note:** Not all cleaning operations are available in the data grid. For example if you want to edit a value in-line, you must do this in the Profile pane.

To apply cleaning operations to a field:

1. In the **Profile pane**, **data grid**, or **Results pane** select the field you want to make changes to.

2. From either the toolbar or **More options** for the field, select from the following options:
   - **Filter** or **Filter Values**: Select from one of the filter options, right-click (CTRL + Click on Mac) a field value to keep or exclude values or use the **Selected Values** filter to pick and choose the values to filter included values not in your flow sample.
   - **Group and Replace**: Manually select values or use automatic grouping. You can also multi-select values in the Profile card and right-click (CTRL+click on Mac) to group or ungroup values or edit the group value.
- **Clean**: Select from a list of quick cleaning operations to apply to all values in the field.

- **Split Values**: Select either automatic or custom splits.

Note: Automatic split and custom split work the same as they do in Tableau Desktop. For more information, see Split a Field into Multiple Fields in the Tableau Desktop and Web Authoring Help.

- **Rename Field**

- **Duplicate Field** (version 2019.2.3 and later) Create a copy of your field and values.

- **Keep Only Field** (version 2019.2.2 and later)

- **Create Calculated Field**

- **Remove Field**

3. To edit a value, right-click or Ctrl+click (MacOS) one or more values and select **Edit Value** then enter a new value. You can also select Replace with Null to replace the values with a Null value or double-click in a single field to edit it directly. For more information about editing field values see Cleaning (fixing) variations of the same value on page 185.

4. Review the results of these operations in the Profile pane, Summary panes or data grid.

**View your changes**

The different types of cleaning operations are represented by icons over the steps in your flow. If more than four types of operations are applied to a step, an ellipsis displays over the step. Hover over these icons to view annotations showing applied operations and the order in which they are performed.

Starting in Tableau Prep Builder version 2019.1.3 and later, you can click on an annotation on the change icon on a step in the Flow pane or on a profile card in the Profile or Results pane and the change and field it impacts will be highlighted in the Changes pane and the Profile or Results pane.
You can also select a step and then expand the **Changes** pane to view the details for each change, edit or remove your changes, drag changes up or down to change the order in which they're applied and in Tableau Prep Builder version 2019.1.1, you can add a description to provide context to other users.

**Cleaning annotation**

**Changes pane**
When viewing changes in an Aggregation, Pivot, Join, or Union step, the order that the change is applied shows either before or after the reshaping action. The order of these changes is applied by the system and cannot be changed. You can edit and remove the change.

Add a description to a change entry (version 2019.1.1 and later)

Add a description of up to 200 characters to any change entry in the Changes pane to provide additional context for your changes.

1. Select a step in the flow pane.
2. Open the Changes pane or Changes tab.
3. Right-click or Ctrl-click (Mac) on an entry in the Changes pane and select Add Description.
4. Enter a description for the step.
The description appears below the generated text for the change with a comment icon.

5. To edit or delete the description, right-click or Ctrl-click (Mac) on the change item, and select **Edit Description** or **Delete Description**.
Apply cleaning operations using recommendations

Sometimes it can be hard to identify which cleaning operation you need to use to fix problems in your data. Tableau Prep Builder version 2018.3.2 and later has the ability to analyze your data and recommend cleaning operations that you can apply automatically to help you quickly fix problems in your data fields or help to identify problems so you can fix them.

**Note:** If you don’t want to use this feature, you can turn it off. From the top menu, go to Help > Settings and Performance. Then click on Enable Recommendations to clear the check mark next to the setting.

This feature is available in all step types except Input, Output and Join step types. Currently, recommendation types include:

- Data roles
- Group and Replace (also applies to fields with data roles starting in version 2019.2.3)
- Filter
- Replace values with Null values
- Remove fields
- Split (version 2019.1.1 and later)

**Note:** This option works specifically with data in fixed-width type text files. To use the split recommendation with this file type, after you connect to the data source, in the Input step, in the Text Settings tab, select a Field Separator character that is not used in the data so the data loads as a single field.

- Trim spaces

To use recommendations to clean your data, do the following:
1. Do one of the following:

   - Click the light bulb icon in the top right corner of the profile card.
   - From the toolbar, click the **Recommendations** drop-down arrow to view all recommendations for your data set and select a recommendation from the list.

   This option only appears when recommended changes are identified by Tableau Prep Builder.

2. To apply the recommendation, hover on the Recommendations card and then click **Apply**.
The change is automatically applied and an entry is added to the Changes pane. To remove the change, click Undo in the top menu or hover over the change in the Changes pane and click the X to remove it.

3. If Tableau Prep Builder identifies further recommendations as a result of the change, the light bulb icon remains on the Profile card until no further recommendations are found. Repeat the steps above to apply any additional changes or ignore the suggested change and use the other cleaning tools in Tableau Prep Builder to address the data problems.

Merge fields

If you have fields that contain the same values but are named differently, you can easily merge them into a single field to combine them by dragging one field on top of the other. When you
merge the fields, the target field becomes the primary field and the field name of the target field persists. The field that you merge to the target field is removed.

Example:

Wildcard union results in 3 fields with the same values

Merge 3 fields into 1

When you merge fields, Tableau Prep Builder keeps all of the fields from the target field and replaces any nulls in that field with values from the source fields that you merge with the target field. The source fields are removed.

Example

<table>
<thead>
<tr>
<th>Name</th>
<th>Contact_Phone</th>
<th>Business_Phone</th>
<th>Cell_Phone</th>
<th>Home_Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>123-4567</td>
<td>123-4567</td>
<td>null</td>
<td>null</td>
</tr>
<tr>
<td>Sally</td>
<td>null</td>
<td>null</td>
<td>456-7890</td>
<td>789-0123</td>
</tr>
<tr>
<td>Fred</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>567-8901</td>
</tr>
<tr>
<td>Emma</td>
<td>null</td>
<td>234-5678</td>
<td>345-6789</td>
<td>null</td>
</tr>
</tbody>
</table>

If you merge the Business_Phone, Cell_Phone and Home_Phone fields with the Contact_phone field, the other fields are removed and results in the following:
To merge fields, do one of the following:

- Drag and drop one field onto another. A **Drop to merge fields** indicator displays.
- Select multiple fields and right-click within the selection to open the context menu, and then click **Merge Fields**.
- Select multiple fields, and then click **Merge Fields** on the toolbar.

For information about how to fix mismatched fields as a result of a union, see **Fix fields that don’t match** on page 209.

### Cleaning (fixing) variations of the same value

Multiple variations of the same value can prevent you from accurately summarizing your data. You can quickly and easily correct these variations using the following options in Tableau Prep Builder.

**Note:** Any edits that you make to the values must be compatible with the field data type.

### Edit a single value

1. In the **Profile** card, click the value you want to edit, and enter the new value. A group icon shows next to the value.

   Alternatively, right-click a value and click **Edit Value**. The change is recorded in the **Changes** pane on the left side of the screen.
2. View the results in the **Profile** pane, and data grid.

**Edit multiple values**

You have several options to edit multiple values at once. For example, use quick cleaning operations to remove punctuation for all values in a field, manually group values using multi-select, automatically group values together using fuzzy-match algorithms that find similar values or select multiple values and replace them with Null.

**Note:** When you map multiple values to a single value, the original field shows a group icon next to the value, showing you which values are grouped together.

**Edit multiple values using quick cleaning operations**

This option applies only to text fields.

1. In the **Profile** pane, Results pane or data grid, select the field you want to edit.

2. Click **More options** ..., select **Clean**, and then select one of the following options:
- **Make Uppercase**: Change all values to uppercase text.
- **Make Lowercase**: Change all values to lowercase text.
- **Remove Letters**: Remove all letters and leave only other characters.
- **Remove Numbers**: Remove all numbers and leave letters and other characters.
- **Remove Punctuation**: Remove all punctuation.
- **Trim Spaces**: Remove leading and trailing spaces.

You can stack operations to apply multiple cleaning operations to the fields. For example, first select **Clean > Remove Numbers** and then select **Clean > Remove Punctuation** to remove all numbers and punctuation from the field values.

3. To undo your changes, click the **Undo** arrow at the top of the **Flow** pane, or remove the change from the change list.

**Group and edit multiple values inline**

1. In the **Profile** card, select the field you want to edit.

2. Press Ctrl or Shift+click or Command or Shift+click (MacOS), and select the values that you want to group.

3. Right-click, and select **Group** from the context menu. The value in the selection that you right-click becomes the default name for the new group but you can edit this in-line.
4. To edit the group name, select the grouped field and edit the value or right-click or Ctrl+click (Mac) on the grouped field and select **Edit Value** from the context menu.

5. To ungroup the grouped field values, right-click on the grouped field and select **Ungroup** from the context menu.

**Replace one or more values with Null**

If you have data rows that you want to include in your analysis but you want to exclude certain field values you can change them to a Null value.

1. In the **Profile** card, press Ctrl or Shift+click or Command or Shift+click (on Mac), and select the values that you want to change

2. Right-click or Ctrl+click (Mac), and select **Replace with Null** from the menu. The values
are changed to Null and the group icon ⚅ shows next to the value.

![Image of Tableau Prep interface showing Customer ID field and group options]

**Edit multiple values manually using Group and Replace**

Use **Group and Replace** to map the value of a field from one value to another or manually select multiple values to group them. You can even add new values to set up mapping relationships to organize your data.

For example, let’s say you have three values in a field: My Company, My Company Incorporated, and My Company Inc. All these values represent the same company, My Company. You can use **Group and Replace** to map the values My Company Incorporated and My Company Inc to My Company, so that all three values appear as My Company in the field.

**Map multiple values to a single selected field**

1. In the **Profile** pane or Results pane, select the field you want to edit.

2. Click **More options** ⚙️ and select **Group and Replace > Manual Selection** from the menu.

3. In the left pane of the **Group and Replace** editor, select the field value that you want to use as the grouping value. This value now shows at the top of the right pane.

4. In the lower section of the right pane in the **Group and Replace** editor, select the values
you want to add to the group.

To remove values from the group, in the upper section of the right pane in the Group and Replace editor, clear the check box next to the values.

Create a group by selecting multiple values

1. In the Profile pane or Results pane, select the field you want to edit.

2. Click More options and select Group and Replace > Manual Selection from the menu.

3. In the left pane of the Group and Replace editor, select multiple values that you want to group.

4. In the right pane of the Group and Replace editor, click Group Values.
A new group is created using the last selected value as the group name. To edit the group name, select the grouped field and edit the value or right-click or Ctrl+click (Mac) on the grouped field and select Edit Value from the menu.

Edit multiple values using Group and Replace with fuzzy match

To search for and automatically group similar values, use one of the fuzzy match algorithms. Field values are grouped under the value that appears most frequently. Review the grouped values and add or remove values in the group as needed.

Choose one of the following options to group values:

- **Pronunciation**: Find and group values that sound alike. This option uses the Metaphone 3 algorithm that indexes words by their pronunciation and is most suitable for English words. This type of algorithm is used by many popular spell checkers. This option isn't available for data roles.

- **Common Characters**: Find and group values that have letters or numbers in common. This option uses the ngram fingerprint algorithm that indexes words by their unique characters after removing punctuation, duplicates, and whitespace. This algorithm works for any supported language. This option isn't available for data roles.

  For example, this algorithm would match names that are represented as "John Smith" and "Smith, John" because they both generate the key "hijmnost". Since this algorithm doesn't consider pronunciation, the value "Tom Jhinois" would have the same key "hijmnost" and would also be included in the group.

- **Spelling**: Find and group text values that are spelled alike. This option uses the
Levenshtein distance algorithm to compute an edit distance between two text values using a fixed default threshold. It then groups them together when the edit distance is less than the threshold value. This algorithm works for any supported language.

Starting in Tableau Prep Builder version 2019.2.3 this option is available to use after a data role is applied. In that case, it matches the invalid values to the closest valid value using the edit distance. If the standard value isn't in your data set sample, Tableau Prep Builder adds it automatically and marks the value as not in the original data set.

- **Pronunciation +Spelling**: (version 2019.1.4 and later) If you assign a data role to your fields, you can use that data role to match and group values with the standard value defined by your data role. This option then matches invalid values to the most similar valid value based on spelling and pronunciation. If the standard value isn't in your data set sample, Tableau Prep Builder adds it automatically and marks the value as not in the original data set. This option is most suitable for English words.

For more information see [Group similar values by data role](#) on page 195.

**Note:** In Tableau Prep Builder version 2019.1.4 and 2019.2.1 this option was labeled **Data Role Matches**.

Group similar values using fuzzy match

1. In the **Profile** pane or Results pane, select the field you want to edit.

2. Click **More options** and select **Group and Replace** then select one of these options:
   - **Pronunciation**
   - **Common Characters**
   - **Spelling**
Tableau Prep Builder finds and groups values that match and replaces them with the value that occurs most frequently in the group.

3. Review the groupings and manually add or remove values or edit them as needed. Then click **Done**.

Adjust your results when grouping field values

If you group similar values by **Spelling** or **Pronunciation**, you can change your results by using the slider on the field to adjust how strict the grouping parameters are.

Depending on how you set the slider, you can have more control over the number of values included in a group and the number of groups that get created. By default, Tableau Prep Builder detects the optimal grouping setting and shows the slider in that position.

When you change the threshold, Tableau Prep Builder analyzes a sample of the values to determine the new grouping. The groups generated from the setting are saved and recorded in the **Changes** pane, but the threshold setting isn't saved. The next time the **Group and**
Replace editor is opened, either from editing your existing change or making a new change, the threshold slider is shown in the default position, enabling you to make any adjustments based on your current data set.

1. In the Profile pane or Results pane, select the field you want to edit.

2. Click More options and select Group and Replace then select one of these options:
   - Pronunciation
   - Spelling

Tableau Prep Builder finds and groups values that match and replaces them with the value that occurs most frequently in the group.

3. In the left pane of the Group and Replace editor, drag the slider to one of the 5 threshold levels to change your results.
To set a stricter threshold, move the slider to the left. This results in fewer matches and creates less groups. To set a looser threshold, move the slider to the right. This results in more matches and creates more groups.

4. Click **Done** to save your changes.

**Group similar values by data role**

*Note:* In Tableau Prep Builder version 2019.1.4 and 2019.2.1 this option was labeled **Data Role Matches**.

You can assign data roles to fields to help standardize your field values and quickly find values that aren’t valid. In Tableau Prep Builder version 2019.1.4 and later, if you assign a geographic data role to a field you can use the values in the data role to group and match values in your data field based on spelling and pronunciation to standardize them. In Tableau Prep Builder version 2019.2.3, you can use either **Spelling** or **Spelling + Pronunciation** to group and match invalid values to valid ones.

These options uses the standard value defined by the data role. If the standard value isn’t in your data set sample, Tableau Prep Builder adds it automatically and marks the value as not in the original data set. For more information about assigning data roles to fields, see **Assign data roles to your data** on page 139.

To use data roles to group values, complete the following steps.

1. In the Profile pane, Results pane or data grid, click the data type for the field.
2. Select one of the following data roles for the field:
- Airport
- City
- Country/Region
- County
- State/Province

Tableau Prep Builder compares the field’s data values to known domain values for the data role you select and marks any values that don't match with a red exclamation point.

3. Click **More options** ⋯, select **Group and Replace**, then select one of the following options:
   - **Spelling**: Matches invalid values to the closest valid values that differ by adding, removing, or substituting characters.
   - **Pronunciation + Spelling**: Matches invalid values to the most similar valid value based on spelling and pronunciation.

   Note: In Tableau Prep Builder version 2019.1.4 or 2019.2.1, this option was called **Data Role Matches**.
Tableau Prep Builder compares the values by spelling or spelling and pronunciation and then groups similar values under the standardized value for the data role. If the standardized value isn’t in your data set, the value is added and marked with a red dot.

Add and identify values that aren't in the data set

If you want to map values in your data set to a new value that doesn’t exist, you can add it using Group and Replace. To easily identify any values that are not in the data set, these values are marked with a red dot next to the value name in the Group and Replace editor.

For example in the image below, Wyoming and Nevada aren’t in the data set.
Some reasons why a value might not be in the data set include the following:

- You just added the new value manually.
- The value is no longer in the data.
- The value is in the data, but isn’t in the sampled data set.

To add a new value:

1. In the **Profile** pane or Results pane, select the field you want to edit.

2. Click **More options** and select **Group and Replace > Manual Selection** from the context menu.

3. In the left pane of the **Group and Replace** editor, click the plus to add a new value.

4. Type a new value in the field and press Enter to add it.
5. In the right pane, select the values that you want to map to the new value.

6. (Optional) To add additional new values to your mapped value, click the plus button in the right pane in the **Group and Replace** editor.

**Aggregate and group values**

Sometimes you’ll need to adjust the granularity of some data, either to reduce the amount of data produced from the flow, or to align data with other data you might want to join or union together. For example, you might want to aggregate sales data by customer before joining a sales table with a customer table.

If you need to adjust the granularity of your data, use the **Add Aggregate** option to create a step to aggregate or group data. Whether data is aggregated or grouped depends on the data type (string, number, or date).
1. In the Flow pane, click the plus icon, and select Add Aggregate. A new aggregation step displays in the Flow pane and the Profile pane updates to show the aggregate and group profile.

2. To group or aggregate fields, drag them from the left pane to one of the columns in the right pane.

You can also:

- Drag and drop fields between the two panes.
- Search for fields in the list and select only the fields you want to include in your aggregation.
- Double-click a field to add it to the left or right pane.
- Change the function of the field to automatically add it to the appropriate pane.
- Click Add All or Remove All to bulk apply or remove fields.
- Apply certain cleaning operations to fields. For more information about which cleaning options are available, see Apply cleaning operations on page 170.

Fields are distributed between the Grouped Fields and Aggregated Fields columns based on their data type. Click the group or aggregation type (for example, AVG or SUM) headings to change the group or aggregation type.

In the data grids below the aggregation and group profile, you can see a sample of the members of the group or aggregation.

Any cleaning operations that are made to the fields are tracked in the Changes pane.
Join or Union Data

There are two methods you can use to combine data in Tableau Prep Builder: join and union.

Join your data

The data that you want to analyze is often made up of a collection of tables that are related by specific fields. Joining is a method for combining the related data on those common fields. The result of combining data using a join is a table that's typically extended horizontally by adding fields of data.

Joining is an operation you can do anywhere in the flow. Joining early in a flow can help you understand your data sets and expose areas that need attention right away.

To create a join, do the following:

1. Join two tables using one of the following methods:
   - Add at least two tables to the flow pane, then select and drag the related table to the other table until the **Join** option displays or click the ✚ icon and select Add Join from the menu.
   - Click the ✚ icon and select Add Join from the menu, then manually add the other input to the join and add the join clauses.

**Note:** In Tableau Prep Builder version 2019.1.3 and later, if you connect to a table that table relationships defined and includes related fields, you can select Add Join and select from a list of related tables. Tableau Prep Builder creates the join based on the fields that make up the relationship.
between the two tables.

For more information about connectors with table relationships, see Join data in the Input step on page 128.

A new join step is added to the flow and the profile pane updates to show the join profile.

2. To review and configure the join, do the following:

   a. Review the **Summary of Join Results** to see the number of fields included and excluded as a result of the join type and join conditions.

   b. Under **Join Type**, click in the Venn diagram to specify the type of join you want.

   c. Under **Applied Join Clauses**, click the plus icon or, on the field chosen for the default join condition, specify or edit the join clause. The fields you selected in the join condition are the common fields between the tables in the join.

   d. Alternatively, you can click the recommended join clauses shown under **Join**.
Clause Recommendations to add the clause to the list of applied join clauses.

Inspect the results of the join

The summary in the join profile shows metadata about the join to help you validate that the join includes the data you expect.

- **Applied Join Clauses**: By default, Tableau Prep Builder defines the first join clause based on common field names in the tables being joined. Add or remove join clauses as needed.

- **Join Type**: By default, when you create a join, Tableau Prep Builder uses an inner join between the tables. Depending on the data that you connect to, you might be able to use left, inner, right, or outer joins.

- **Summary of Join Results**: The Summary of Join Results shows you the distribution of values that are included and excluded from the tables in the join.
  
  - Click each **Included** bar to isolate and see the data in the join profile included in the join.
  
  - Click each **Excluded** bar to isolate and see the data in the join profile that are excluded from the join.
  
  - Click any combination of the **Included** and **Excluded** bars to see a cumulative perspective of the data.

- **Join Clause Recommendations**: Click the plus icon next to the recommended join
clause to add it to the **Applied Join Clauses** list.

- **Join Clauses** pane: In the **Join Clauses** pane, you can see the values in each field in the join clause. The values that don't meet the criteria for the join clause are displayed in red text.

![Join Clauses](image)

- **Join Results** pane: If you see values in the **Join Results** pane that you want to change, you can edit the values in this pane.

### Common join issues

If you don't see the results you expect after joining your data, you may need to do some additional cleaning on your field values. The following issues will result in Tableau Prep Builder reading the values as not a match and exclude them from the join:

- Different capitalization: My Sales and my sales
- Different spelling: Hawaii and Hawai'i
- Mispelling or typo's: My Company Health and My Company Heath
- Name Changes: John Smith and John Smith Jr.
The good news is that if your field values have any of these issues, you can fix the field values in-line directly in the Join Clauses or work with excluded values by clicking in the Excluded bars in the Summary of Join Results and use the cleaning operations in the profile card menu.

For more information about the different cleaning options available in the Join step, see Apply cleaning operations on page 170.

Fix mismatched fields and more

You can fix mismatched fields right in the join clause. Double-click or right-click the value and select Edit Value from the context menu on the field that you want to fix and enter a new value. Your data changes are tracked and added to the Changes pane right in the Join step.

You can also select multiple values to keep, exclude or filter in the Join Clauses panes, or apply other cleaning operations in the Join Results pane. Depending on which fields you change and where they are in the join process, your change is applied either before or after the join to give you the corrected results.
Union your data

Union is a method for combining data by appending rows of one table onto another table. For example, you might want to add new transactions in one table to a list of past transactions in another table. Make sure the tables you union have the same number of fields, the same field names, and the fields are the same data type.

**Tip:** To maximize performance a single union can have a maximum of 10 inputs. If you need to union more than 10 files or tables, try unioning files in the Input step. For more information about this type of union, see Union files and database tables in the Input step on page 122.

Similar to a join, you can use the union operation anywhere in the flow.

To create a union, do the following:

1. After you add at least two tables to the flow pane, select and drag a related table to the other table until you see the **Union** option. You can also click the icon and select **Add Union** from the menu. A new union step is added in the Flow pane, and the Profile pane updates to show the union profile.

For more information about cleaning fields see **Apply cleaning operations** on page 170.
2. Add additional tables to the union by dragging tables toward the unioned tables until you see the Add option.

3. In the union profile, review the metadata about the union. You can remove tables from the union as well as see details about any mismatched fields.

**Inspect the results of the union**

After you create a union, inspect the results of the union to validate that the data in the union is what you expect. There are a number of areas in the union profile that you can check to help you validate the data in the union.
• **Review the union metadata**: The union profile shows some metadata about the union. Here you can see the tables that make up the union, the resulting number of fields and any mismatched fields.

![Union Profile]

- **Review the colors for each field**: Next to each field listed in the Union summary and above each field in the union profile, is a set of colors. The colors correspond to each table in the union.

If all table colors show for that field, then the union performed correctly for that field. A missing table color indicates that you have mismatched fields.

![Union Results]
Mismatched fields are fields that might have similar data but are different in some way. You can see the list of fields that don't match in the Union summary and the tables where they came from. If you want to take a closer look at the data in the fields, select the **Show only mismatched fields** check box to isolate the mismatched fields in the Union profile.

To fix these fields, follow one of the suggestions in the “Fix fields that don’t match” section below.

**Fix fields that don’t match**

When tables in a union don’t match, the union produces extra fields. The extra fields are valid data being excluded from their appropriate context.

To resolve a field mismatch issue, you must merge the mismatched fields together.

There are a number of reasons why fields might not match.

- **Corresponding fields have different names:** If corresponding fields between tables have different names, you can use union recommendations, manually merge fields in the **Mismatched Fields** list, or rename the field in the union profile to merge the mismatched fields together.

  To use union recommendations, do the following:

  1. in the **Mismatched Fields** list, click on a mismatched field. If a suggested match exists, the matching field is highlighted in yellow.

     Suggested matches are based on fields with similar data types and field names.
2. Hover on the highlighted field and click the plus button to merge the fields.

To manually merge fields in the **Mismatched Fields** list, do the following:

1. Select one or more fields in the list.

2. Right click a selected field and if the merge is valid, the **Merge Fields** menu option appears.

   If you see **No options available** when you right-click the field, this is because the fields are not eligible to merge. For example trying to merge two fields from the same input.

3. Click **Merge Fields** to merge the selected fields.

To rename the field in the union profile pane, right-click the field name and click **Rename Field**.
• **Corresponding fields have the same name but are a different type:** By default, when the name of corresponding fields match but the data type of the fields don’t, Tableau Prep Builder will change the data type of one of the fields so they are compatible with each other. If Tableau Prep Builder makes this change, it’s noted at the top of the merged field by the Change Data Type icon.

In some cases, Tableau Prep Builder might not pick the correct data type. If that happens and you want to undo the merge, right-click the Change Data Type icon and select **Separate Inputs with Different Types**.

You can then merge the fields again by first changing the data type of one of the fields and then using the suggestions in **Additional merge field options** on the next page.

• **Corresponding tables have different number of fields:** To union tables, each table
in the union must contain the same number of fields. If a union results in extra fields, merge the field into an existing field.

**Additional merge field options**

In addition to the methods described in the above section for merging fields you can also use one of the following methods to merge fields. You can merge fields in any step, except for the Output step.

For information about how to merge fields in the same file, see [Merge fields](#) on page 183.

To merge fields, do one of the following:

- Drag and drop one field onto another. A **Drop to merge fields** indicator displays.
- Select multiple fields and right-click within the selection to open the context menu, and then click **Merge Fields**.
- Select multiple fields, and then click **Merge Fields** on the context-sensitive toolbar.
Save and Share Your Work

At any point in the flow, you can save your flow, view a preview of the data in your flow in Tableau Desktop, or create an extract of your data that includes all the operations that you’ve applied to your flow. You can also package your data with your flow to share it with others or publish your data extract to Tableau Server or Tableau Online as a data source.

Save a flow

Save your flow to back up your work before performing any additional operations. Your flow is saved in the Tableau Prep Builder flow (.tfl) file format.

You can also package your local files (Excel, Text Files, and Tableau extracts) with your flow to share with others, just like packaging a workbook for sharing in Tableau Desktop. Only local files can be packaged with a flow. Data from database connections, for example, aren’t included.

When you save a packaged flow, the flow is saved as a Packaged Tableau Flow File (.tflx).

- To save your flow, from the top menu, select File > Save.
- To package your data files with your flow, from the top menu, do one of the following:
  - Select File > Export Packaged Flow
  - Select File > Save As. Then in the Save As dialog, select Packaged Tableau Flow Files from the Save as type drop down menu.

View your data sample in Tableau

Sometimes when you’re cleaning your data you might want to check your progress by looking at it in Tableau Desktop. When your flow opens in Tableau Desktop, Tableau Prep Builder creates a permanent Tableau extract (.tde or .hyper depending on your version of Tableau) and a Tableau data source (.tds) file. The files are saved in your Tableau repository in the Datasources file so you can experiment with your data at any time.

Note: While you can experiment with your data, Tableau only shows you a sample of your data and you won’t be able to save the workbook as a packaged workbook (.twbx).
When you are ready to work with your data in Tableau, create an output step for your flow and save the output to a file or as a published data source, then connect to the full data source in Tableau.

When you open the flow in Tableau Desktop, you can see the data sample that you are working with in your flow with the operations applied to it, up to the step that you selected.

To view your data sample in Tableau do the following:

1. Right-click the step where you want to view your data, and select **Preview in Tableau Desktop** from the context menu.

2. Tableau Desktop opens on the **Sheet** tab.

Create and publish data extracts and data sources

To create an extract, run your flow. When you run your flow your changes are applied to your entire data set. Running the flow results in a Tableau Data Source (.tds) and a Tableau Data Extract (.tde or .hyper) file. You can create an extract file from your flow output to use in Tableau Desktop or to share your data with third parties.

**Note:** You can publish data extracts or data sources to Tableau Server version 10.0 and later as well as to Tableau Online.
Tableau Prep Help

You can create an extract file in the following formats:

- **Tableau Data Extract (.tde):** The extract is saved as both a Tableau extract (.tde) and a Tableau data source (.tds) file. Use this file type if you use Tableau Desktop or Tableau Server version 10.0 through 10.4.

- **Hyper Extract (.hyper):** This is the new Tableau extract file type but can only be consumed by Tableau Desktop or Tableau Server version 10.5 and later.

- **Comma Separated Value (.csv):** Save the extract to a .csv file to share your data with third parties. The encoding of exported CSV file will be UTF-8 with BOM.

You can also publish your output as a data source to Tableau Server or Tableau Online to share your data and provide centralized access to the data you have cleaned, shaped, and combined.

**Note:** To publish your output to Tableau Server, the Tableau Server REST API must be enabled. For more information see Rest API Requirements in the Tableau Rest API Help. To publish to a server that uses Secure Socket Layer (SSL) encryption certificates, additional configuration steps are needed on the machine running Tableau Prep Builder. For more information, see the Before you Install in the Tableau Desktop and Tableau Prep Builder Deployment Guide.

**Create an extract file**

1. Click the plus icon on a step and select **Add Output**.

   If you have run the flow before, click the run flow button on the Output step. This overwrites the previous output.

   The **Output** pane opens and shows you a snapshot of your data.
2. In the left pane select **Save to file**.

3. Click the **Browse** button, then in the **Save Extract As** dialog, enter a name for the file and click **Accept**.

4. In the **Output type** field, select the output type. Depending on the version of Tableau Desktop you use you can choose from the following options:
   - Tableau Data Extract (.hyper) for Tableau Desktop version 10.5 and later.
   - Tableau Data Extract (.tde) for Tableau Desktop version 10.0 through 10.4.
   - Comma Separated Values (.csv) if you want to share the extract with a third party.

   **Tip:** You have choices when generating output from your flow. You can generate an extract file, or you can publish your data as a data source to Tableau Server or Tableau Online. For more information about generating output files, see **Create and publish data extracts and data sources** on page 214.

5. Click **Run Flow** to run the flow and generate the extract file.
Publish as a data source

1. Click the plus icon on a step and select **Add Output**.

   If you have run the flow before, click the run flow button on the Output step. This overwrites the previous output.

2. The output pane opens and shows you a snapshot of your data.

3. Select the **Publish as data source** radio button and complete the following fields:
   
   - **Server**: Select the Tableau server where you want to publish the data source and data extract. If you aren’t signed in to a server you will be prompted to sign in.
   
     On the Mac you may be prompted to provide access to your Mac keychain so Tableau Prep Builder can securely use SSL certificates to connect to your Tableau Server or Tableau Online environment.
   
   - **Project**: Select the project where you want to load the data source and extract.
   
   - **Name**: Enter file name
   
   - **Description**: Enter a description for the data source.

4. Click **Run Flow** to run the flow and publish the data source.
Refresh output files from the command line

**Note:** Running flows from the command line that include connections using either the JDBC or ODBC connectors is supported starting in version 2019.2.3.

If you want to refresh output files for your flow you can run the flow from the command line instead of opening it in Tableau Prep Builder and running the flow from there. You can run one flow at a time using this method.

You can also schedule your flow to run automatically in Tableau Server using Tableau Prep Conductor. For more information about Tableau Prep Conductor, see Keep Flow Data Fresh on page 233.

This option is available on both Windows and Mac machines where Tableau Prep Builder is installed. For Windows machines, you can also schedule this process using Windows Task Scheduler. For more information about using Windows Task Scheduler, see Task Scheduler in the Microsoft online help.

When you run flows from the command line, Tableau Prep Builder refreshes all outputs for the flow.

**Note:** The output location for the files is specified in the output step for the flow in Tableau Prep Builder when you run the flow. When you refresh the files from the command line, this process uses that same location and will overwrite any previous output files for the flow with the refreshed version.

For information about how to specify an output location for your flow files, see Create and publish data extracts and data sources on page 214.

To run the flow from the command line, you'll need the following:

- Administrator privileges on the machine where you are running the flow.
- The path where Tableau Prep Builder is installed.
- If connecting to databases and publishing output files to a server, a credentials .json file that includes all required credentials.
- The path where the Tableau Flow (.tfl) file is located.
Before running the flow

If you are running a flow that connects to database files or publishes output files to a server, then you'll need to create a .json file that includes the credentials that are required to connect to these locations.

When you run the process, the hostname, port and username provided in the credentials .json file is used to find the matching connection in the Tableau flow file (tfl) and updated before running the process.

**Note:** Skip this step if the flow connects to and outputs to local files, files stored on a network share or input files that use Windows Authentication (SSPI). For more information about Windows Authentication, see SSPI Model in the Microsoft online help.

If you plan to reuse the file, place it in a folder where it won't be overwritten by the Tableau Prep Builder install process.

The following table lists the credentials that you need to include in the .json file. Port ID and Site ID are optional if your connections don't require this information.

<table>
<thead>
<tr>
<th>Input connections</th>
<th>Output location</th>
</tr>
</thead>
<tbody>
<tr>
<td>• User name</td>
<td>• Server Url</td>
</tr>
<tr>
<td>• Host name (Server name)</td>
<td>• Content Url (Site ID. This appears after /site/ in the URL when you sign into Tableau Server or Tableau Online)</td>
</tr>
<tr>
<td>• Port ID</td>
<td>• User name</td>
</tr>
<tr>
<td>• Password</td>
<td>• Password</td>
</tr>
</tbody>
</table>

The following example shows the syntax to use for the credentials .json file.

In this example the flow connects to two databases (Microsoft SQL Server and Oracle) and outputs files to a server that includes a Site ID.

```json
{
   
   "inputConnections": [

   ]
}
```
"username": "jsmith",
"hostname": "mssql.example.lan",
"port": 1234,
"password": "passw0rd"
},
{
"username": "jsmith",
"hostname": "Oracle.example.lan",
"port": 5678,
"password": "passw0rd"
}
],
"outputConnections": [
{
"serverUrl": "http://MyServer",
"contentUrl": "FinanceTeam",
"username": "jsmith",
"password": "passw0rd$"
}
]

Tips for creating your credentials file

Tableau Prep Builder uses information from the flow file and from the credentials .json file to run the flow when you have remote connections. For example, the database name for your remote connections and the project name for your output files come from the flow, and the server name and the log in credentials come from the .json file.

To avoid errors when running the flow, make sure your credentials file follows these guidelines:

- If using Tableau Prep Builder version 2018.2.2 through 2018.3.1, always include the "inputConnections" and "outputConnections" arrays even if the flow doesn't have remote connections for inputs or outputs. Just leave those arrays blank. If you are using Tableau Prep Builder version 2018.3.2 and later you don't need to include the blank array.
No remote input connection? Include this syntax at the top of the .json file

```json
{
 "inputConnections": [  
 ],
}
```

No remote output connection? Include this syntax at the bottom of the .json file

```json
"outputConnections": [
 ]
}
```

No port ID for your input connection? Don’t include the "port":xxxx, reference in the .json file, not even "port": "".

When referencing the "serverUrl": don’t include a "/" at the end of the address. For example, use this "serverUrl": "http://server" not this "serverUrl": "http://server/".

If you have multiple input or output connections include the credentials for each one in the file.

Run the flow

**Important:** The examples below include the name change for "Tableau Prep" version 2019.1.2 to "Tableau Prep Builder". If you are using an earlier version of the product use "Tableau Prep" instead.

1. Open the command prompt or terminal command prompt (Mac) as an Administrator.
2. Run one of the following commands:

   - The flow connects to local files or files stored on a network share and publishes to local files, files stored on a network share or uses Windows authentication:

     **Note:** If connecting to or outputting to files stored on a network share, use the UNC format for the path: \server\path\file name. It can’t be password protected.

Windows
"\[Tableau Prep Builder install location]\Tableau Prep Builder <version>\scripts\tableau-prep-cli.bat -t "path\to\[your flow file name].tfl"

**Mac**

/Applications/Tableau\ Prep\ Builder\ [Tableau Prep Builder version].app/Contents/scripts/.\tableau-prep-cli -t path/to/[your flow file name].tfl

- The flow connects to databases or publishes to a server:

**Windows**

"\[Tableau Prep Builder install location]\Tableau Prep Builder <version>\scripts\tableau-prep-cli.bat -c "path\to\[your credential file name].json" -t "path\to\[your flow file name].tfl"

**Mac**

/Applications/Tableau\ Prep\ Builder\ [Tableau Prep Builder version].app/Contents/scripts/.\tableau-prep-cli -c path/to/[your credential file name].json -t path/to/[your flow file name].tfl

- The flow file or credentials file is stored on a network share (use the UNC format for the path: \server\path\file name):

**Windows**

"\[Tableau Prep Builder install location]\Tableau Prep Builder <version>\scripts\tableau-prep-cli.bat -c "\server\path\[your credential file name].json" -t "\server\path\[your flow file name].tfl"

**Mac**: Map the network share to /Volumes in Finder so that it is persistent, then use /Volumes/.../[your file] to specify the path:
For examples of sample commands see **Syntax examples on the next page**. For common errors and resolutions see **Common errors when using the command line to run flows** on page 359.

## Command options

If you want to view the help options, include `-h` in the command line.

<table>
<thead>
<tr>
<th>Command options</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>-c</code></td>
<td>The connection path to the credentials file.</td>
<td>Requires the path to where the credentials file is located.</td>
</tr>
<tr>
<td><code>-d</code></td>
<td>Debug the flow process.</td>
<td>Include this option to view more information to help debug a problem with refreshing the flow. Log files are stored in: <code>My Tableau Prep Builder Repository\Command Line Repository\Logs</code></td>
</tr>
</tbody>
</table>
| `-h` | View the help for syntax options | The help option or a syntax error shows the following information: usage: `tableau-prep-cli [-c <arg>] [-d] [-h] [-t <arg>]
-c, --connections <arg> path to a file with all connection information` |
-d,--debug
This option is for debugging

-h,--help
print usage message

-t,--tflFile <arg>
The tableau prep builder flow file

-t The .tfl flow file
Requires the path to where the .tfl flow file is located.

Syntax examples

The command lines below show three different examples for running a flow using the following criteria:

- **Tableau Prep Builder version**: 2019.2.2
  
  **Important**: The examples below include the name change for Tableau Prep version 2019.1.2 to Tableau Prep Builder. If you are using an earlier version of the product use "Tableau Prep" instead.

- **Flow name**: Flow1.tfl

- **Flow location**: C:\Users\jsmith\Documents\My Tableau Prep Builder Repository\Flows

- **Credentials file name**: Flow1.json

- **Credentials file location**: C:\Users\jsmith\Desktop\Flow credentials

- **Credentials file location stored on a network share**: \tsi.lan\files\Flow credentials

The flow connects to and publishes to local files

**Windows**

"\Program Files\Tableau\Tableau Prep Builder 2019.2.2\scripts\tableau-prep-cli.bat  -t "\C:\Users\jsmith\Documents\My Tableau Prep Builder Repository\Flows\Flow1.tfl"

**Mac**
The flow connects to databases and publishes to a server

**Windows**

"\Program Files\Tableau\Tableau Prep Builder 2019.2.2\scripts\tableau-prep-cli.bat -c "\C:\Users\jsmith\Desktop\Flow credentials\Flow1.json" -t "\C:\Users\jsmith\Documents\My Tableau Prep Builder Repository\Flows\Flow1.tfl"

**Mac**

/Applications/Tableau\ Prep\ Builder\ 2019.2.2.app\Contents\scripts/./tableau-prep-cli -c /Users/jsmith/Documents/My\ Tableau\ Prep\ Builder\ Repository/Flows/Flow1.tfl

The flow publishes to a server and the credentials file is stored on a network share

**Windows**

"\Program Files\Tableau\Tableau Prep Builder 2019.2.2\scripts\tableau-prep-cli.bat -c "\\tsi.lan\files\Flow credentials\Flow1.json" -t "\C:\Users\jsmith\Documents\My Tableau Prep Builder Repository\Flows\Flow1.tfl"

**Mac/Applications/Tableau\ Prep\ Builder\ 2019.2.2.app\Contents\scripts/./tableau-prep-cli -c /Volumes/files/Flow credentials/Flow1.json -t /Users/jsmith/Documents/My\ Tableau\ Prep\ Builder\ Repository/Flows/Flow1.tfl
Version Compatibility with Tableau Prep

Tableau Prep Builder releases new versions every month. Like Tableau Desktop, if new features or connectors are introduced in a new version and you are working in an older version of Tableau Prep Builder, compatibility may be an issue if you try to open a flow.

Similarly, if you publish flows to Tableau Server to schedule them to run using Tableau Prep Conductor, if your flows include new features or connectors that aren’t supported in the version of Tableau Server that you publish to, you can run into compatibility errors that might prevent you from running your flows in Tableau Server.

**Note:** To download a specific version of Tableau Prep Builder, open the Product Downloads and Release Notes page and select Tableau Prep Builder from the list on the left side of the page.

Finding our version

Follow the instructions below to find your product version number.

**Tableau Prep Builder**

Tableau Prep Builder release versions are numbered like Tableau Desktop maintenance versions, for example 2019.1.4, except Tableau Prep Builder doesn’t release maintenance versions for this product. Each release represents the latest version of the product including any new features and updates.

To find the release version for your product, open Tableau Prep Builder, then in the top menu, click **Help > About Tableau Prep** or **About Tableau Prep Builder**, depending on your version. The release number displays in the lower left corner of the dialog.
Tableau Prep Help

Tableau Server

Tableau Prep Conductor was introduced as part of the Data Management Add-on to Tableau Server version 2019.1. To publish and schedule flows to run on Tableau Server, you must be using Tableau Server version 2019.1 or later and Tableau Prep Conductor must be enabled.

To find your version of Tableau Server, open Tableau Server in your web browser. In the top menu bar click the information icon in the top right corner and select About Tableau Server. A dialog opens that tells you which version of Tableau Server you are using. For information about how Tableau Prep Conductor is enabled, see Step 2 - Configure Flow Settings for your Tableau Server on page 264.
Compatibility between different versions of Tableau Prep Builder

Generally, a new version of Tableau Prep Builder can open flows created in an older version. However, compatibility issues can occur when you try to open a flow between newer and older versions of Tableau Prep Builder or even when opening flows in the same version of Tableau Prep Builder using different computers.

For example:

- The flow includes input connectors or features that aren't supported in the version where the flow is opened.
- The machine that you use to open the flow doesn't have the required input connectors installed or has a driver version for the connector that isn't compatible. Tableau Prep Builder requires 64-bit drivers to be installed to work with flow input connectors.

If compatibility is an issue, when you try to open the flow, the flow may open but contains errors or the flow won't open and you receive an error message. In the example below, the flow won't open and an error message displays and lists the incompatible features and options for resolving the issue.

Fix compatibility issues with Tableau Prep Builder

To fix compatibility issues, try one of the following:
Tableau Prep Help

- Upgrade to the latest version of Tableau Prep Builder.
  
  Click the update button on the bottom of the Discover pane to download the latest version of the product and follow the instructions to Install Tableau Prep Builder in the Tableau Desktop and Tableau Prep Builder Deployment Guide. If you don't have access to the update button on the Discover pane, instructions about how to download the latest version of the product are included in the Install Tableau Prep Builder topic.
- Make sure your computer is compatible with Tableau Prep Builder. For example, make sure that you have the 64-bit drivers installed for the connectors used by the flow. To install drivers, see the Driver Download page.
- Open a copy of the flow that has the incompatible features removed.

Compatibility between different versions of Tableau Prep Builder and Tableau Prep Conductor

Like Tableau Desktop, publishing from a newer version of Tableau Prep Builder to an older version of Tableau Server can result in compatibility issues. Tableau Prep Conductor detects the features that are included in a flow when it has been published. If it finds features that it doesn't support, the flow can still be published to Tableau Server, but it is blocked from running in Tableau Server and can’t be scheduled or added to a task.

You can still run the flow using the command line. For more information about using this process, see Refresh output files from the command line on page 218.

In Tableau Prep Builder, if you publish a flow and incompatibility is detected, the following message is displayed and lists the features that aren't supported in the version of Tableau Server.

![Incompatible Flow Version]

If you continue to publish the flow, publishing will complete successfully. However, when you open the flow in Tableau Server, you will see the following message:
To schedule and run the flow in Tableau Server, you can do one of the following:

- Upgrade Tableau Server to be compatible with the version of Tableau Prep Builder that you are using. For example, Tableau Prep Builder releases a new build each month. Features released in Tableau Prep Builder version 2019.1.4 may not be supported in Tableau Server version 2019.1 but would be supported in Tableau Server version 2019.2 or later.

- Before publishing the flow, remove the features from the flow that Tableau Prep Conductor doesn't recognize, then publish the flow.

- If it is already published to Tableau Server download the flow and remove the features or create the flow in an older version of Tableau Prep Builder using only the features available in that version.

  **Note:** To download a specific version of Tableau Prep Builder, open the Product Downloads and Release Notes page and select Tableau Prep Builder from the list on the left side of the page.

**Fixing compatibility issues with Tableau Server**

If the flow is already published to Tableau Server, try the steps below to remove the incompatible features using your current version of Tableau Prep Builder. After you remove the features and no longer see the version incompatibility message, republish your flow to schedule it using Tableau Prep Conductor.
Identify and remove incompatible features

If you are working in Tableau Prep Conductor, it doesn’t currently list the incompatible features in your flow. To identify the list of features to remove from the flow, you need to open the flow in Tableau Prep Builder then find and remove them in your flow.

1. Open your flow. If you are in Tableau Prep Conductor, from the More actions menu, click **Download** to download and open the flow in Tableau Prep Builder or simply open the flow in Tableau Prep Builder.
2. If you downloaded the flow, click on the downloaded flow to open it.
3. In Tableau Prep Builder, from the top menu select **Server > Publish Flow**. If you need to sign into the server again, make sure you select the same server that is incompatible with the flow.

   A warning dialog opens that lists the features that are not compatible with the Tableau Server version. Note the features so you can identify and remove them from the flow.

   4. Click **Cancel** to close the **Publish Flow to Tableau Server** dialog.

   5. From the top menu, click **File > Save As** to save a copy of your flow.

Remove incompatible features from the flow

You can use various methods to find and remove features from your flow. This section shows some options to help you resolve incompatibility errors.

Incompatible data sources

If the data source isn’t compatible, for example a new connector was added that isn’t yet supported in Tableau Prep Conductor, you’ll need to connect to a data source that is supported.

To change your data connection do the following:

1. In the **Connections** pane, click the drop-down arrow on the connector and select **Remove** to remove the data source connection.

   This will temporarily put your flow in an error state.

2. Connect to a compatible data source and drag it to the Flow pane. To connect the new Input step with your flow, drag the step to the first step in the flow where you want to connect it and drop it on the connecting step on the **Add** option.
3. Repeat these steps to replace any other connectors. Then save your flow and republish it.

Incompatible features

To remove incompatible features you'll need to find the steps where the features were used and remove them.

1. If the feature is a step type, in the Flow pane click on the step where the feature is used. Right-click or Ctrl-click (Mac) an the step and select **Remove**.

2. If the feature is a cleaning operation, in the Flow pane click on the step where the feature is used. You can hover over the annotations in the **Flow** pane or in the **Profile** or **Results** panes to see a list of changes.

   **Note:** In Tableau Prep Builder version 2019.1.3 and later you can hover on the icon that represents the change you are looking for over a step in the Flow pane or in the profile card then select the annotation from the list of changes. The change is highlighted in the **Changes** pane, **Profile** or **Results** pane and data grid.

3. Open the **Changes** pane if needed, and select the change that matches the feature you need to remove. Click on the change to select it and click **Remove** to delete it from the flow.

4. Repeat these steps to replace any other features. Then save your flow and republish it.
Keep Flow Data Fresh

You’ve built your flow and cleaned your data, but now you want to share your data set with others and you want to keep that data fresh. You can always run your flow in Tableau Prep Builder and publish an extract to Tableau Server, but now there’s a better way.

Meet Tableau Prep Conductor, part of the new Data Management Add-on to Tableau Server version 2019.1 and later. If you add this option to your Tableau Server installation, you can use Tableau Prep Conductor to run your flows automatically and keep your flow data fresh.

With Tableau Prep Conductor you can do the following:

- Configure your Server to use Tableau Prep Conductor
  - Enable or disable Tableau Prep Conductor for individual sites
  - Set up email notifications for flow failures for flows that are run either on-demand or using a schedule
  - Configure flow timeout settings
- Publish a flow from Tableau Prep Builder to Tableau Server
  - Upload data files or connect directly to your files or databases. If connecting to databases, you can either embed the database credentials or require a user prompt.
Note: If you connect to data files through a direct connect or publish your flow output to a file share, the files need to be in a location that Tableau Server can access. For more information see Step 4 - Safe list Input and Output locations on page 270.

- Select from a project hierarchy when publishing your flows
- Enter tags and a description to help others find your flow

- Manage the flow
  - Set permissions
  - Move the flow to a different project
  - Change the flow owner
  - Add or edit tags
  - View the version history and select from the list to restore the flow to a previous version
  - Mark a flow as a favorite and add it to your favorites list
  - Edit an input connection and update credentials
  - View data sources created from a flow and link back to the flow that created it

- Create schedules to run your flows
- Add scheduled tasks to run the flow and select which flow outputs to update
- Run the flow on demand without a schedule

- Monitor the flow
  - Set up email alert notifications
  - View errors
  - Monitor and restart flows that have been suspended
  - View run history
  - Use Admin views
About Tableau Prep Conductor

If you added the Tableau Data Management Add-on to your Tableau Server installation, you can use Tableau Prep Conductor to leverage the scheduling and tracking functionality of Tableau Server to automate running your flows to update the flow output. Otherwise, you can open Tableau Prep Builder and manually run individual flows as your data changes.

Seamlessly publish flows from Tableau Prep Builder to Tableau Server using functionality that is similar to publishing data sources and workbooks with Tableau Desktop today. Package files with the flow or specify a direct connection to data sources to update the flow input as data changes. If your flow connects to databases, specify the authentication type and set credentials to access the data.

You can also publish a flow to share it with others. For example, publish an incomplete flow to Tableau Server and then download it to another computer to continue working on it. Create a flow with only Input steps that are properly configured and share it with co-workers who can then download the flow to their computers and create and publish their own flows. Flows can only be run in Tableau Server if the flow includes output steps and the output steps don’t have any errors.

For more information about publishing a flow, see Publish a Flow to Tableau Server on page 273.

Keeping track of the health of your flows is easy. With Tableau Prep Conductor, you can:

- View and monitor the details about your flow, including recent activity in the Content pages.

- View the results of the flow runs and any errors in the Run History tab.

- Use Administrative Views to monitor server and site activity including a new view that tracks flow performance history.

- View detailed alerts for failed flow runs.

- Set up email notification alerts to send emails to flow owners notifying them when the flow failed to run and why.

  For more information about setting up alerts, see Monitor Flow Health and Performance on page 291.

If a flow fails to run due to connectivity errors, you can fix the error right in Tableau Server by editing the connection. If a flow fails to run due to errors in one or more of the flow steps, such as
a calculation that isn’t valid, simply download the flow to Tableau Prep Builder, fix the problems and then republish the flow to Tableau Server to pick up where you left off.

Enabling Tableau Prep Conductor

Tableau Prep Conductor is available with the Data Management Add-on and can be installed with Tableau Server or added to an existing installation of Tableau Server version 2019.1 or later. For more information about how to install and enable Tableau Prep Conductor, see Enable and Configure Tableau Prep Conductor on Tableau Server on page 244.

Tableau Prep Conductor is designed to work with Tableau Prep Builder version 2019.1.2 and later. You must install or upgrade Tableau Prep Builder to use version 2019.1.2 or later to be able to publish flows to Tableau Server. For information about how to install Tableau Prep Builder, see Install Tableau Desktop or Tableau Prep Builder from the User Interface.

Getting to know the flow workspace

After you publish your flow you can schedule tasks in Tableau Server to automatically run your flow on a regular basis to keep your output data fresh.

Tableau Prep Conductor leverages much of the same functionality for managing flows that you might see when managing workbooks or data sources from Tableau Desktop in Tableau Server. For example, just like extract refreshes, scheduled flow tasks and on-demand flow runs are queued as background tasks. But when it comes to working with flows, there are a few differences.

Flow Overview page

The flow Overview page is the main landing page where you can view data about your flow and schedule, monitor and maintain running the flow.

Open the flow Overview page by clicking on a flow in your list. You can navigate there from Content > Explore (All Flows) or by opening the project that contains your flows.
A. The header lists the name of the flow, the flow owner and the date that the flow was last modified.

Add a flow to your favorites, or from the More actions ••• menu you can also download the flow, set permissions, change the flow owner, restore previous flow versions, and more.

B. View and edit the flow description and set tags to help others find the flows they are looking for.

C. View the output steps for a flow along with the status of the last update, any schedule the output is assigned to, and any errors from the last flow run. You can also click the Run button to run all output steps or individual output steps on-demand.

If the flow has errors, the flow run will fail. Connectivity errors can be resolved directly in Tableau Server by navigating to the Connections tab for the flow and editing the input connections. To resolve any other flow errors, from the More actions ••• menu select Download to download and open the flow in Tableau Prep Builder. Edit the flow then republish it and try running the flow again.

After a flow has run successfully, outputs that are data sources become links that you can click to open the Data Source page to view more information about the data source or edit the flow input connection.

In the Schedule field, view the scheduled tasks that the output step is assigned to. A flow output can be assigned to one or more tasks.
If no schedule has been assigned yet, click **Create new task** to add the output step to a schedule. To immediately run the flow to update a specific output step, click the **Run** button on the left-hand side of the row.

D. View an image of the flow.

**Flow Connections page**

View both the input and output locations for a flow, connection types, authentication settings, input and output steps and any connectivity errors.

For database input types, click the **More actions ...** menu for an input connection to edit the connection and change the server name, port, user name and password.

**Flow Scheduled Tasks page**

View any schedules that the flow is assigned to and the outputs that are included in those schedules. Click the schedule link at the top of the task card to open the **Schedules** page and see a list of flows that are assigned to that schedule. For more information about assigning flows to a schedule, see **Schedule a Flow Task** on page 281.

You can also add new tasks or manage existing ones from this page. To take action on an existing task, select the check box on a task card then click the **Actions** drop-down menu to run, edit, or delete the task. If the flow fails to run after a configured number of consecutive attempts, the flow is automatically suspended. You can see that status on the **Overview** tab as well as this tab. You can resume suspended tasks from this menu.

For information about how to set the threshold for suspended flow tasks, see Step 5 - **Optional Server Configurations** on page 272. For more information about suspended flow tasks, see **View and resolve errors** on page 292.
Schedules page

On the Schedules page, you can view the flows assigned to a schedule and the details about the flow runs. You can run the schedule on-demand and run all flows assigned to it. You can also select one or more flows, then use the Actions menu to change the flow schedule or priority, delete selected flows from the schedule or resume suspended flows.

For information about how to set up a schedule, see Step 3- Create Schedules for Flow Tasks on page 268.

Flow Run History

See, search, and sort through a list of historical runs for a flow. This page also includes details about the flow run such as duration and number of rows that were generated.

If the flow output has an error, hover over the error to view the messages. If applicable, click the Go to Connections link in the error message to navigate to the Connections page to fix connectivity errors. If shown, click the Download the Flow link to download the flow to fix flow errors in Tableau Prep Builder, then republish the flow to continue to manage it using Tableau Prep Conductor.

Note: The run history for a flow will persist unless the flow is deleted.
Flow Revision History

If you need to revert a flow to a previous version, from the More actions ••• menu for the flow, select Revision History. On the Revision History dialog, select the flow version from the list that you want to revert to.

Licensing Tableau Prep Conductor for Tableau Server

Tableau Prep Conductor is licensed through the Data Management Add-on, on a per Deployment basis, which may be User-Based or Core-Based. A Deployment includes a licensed production Tableau Server installation and licensed non-production Tableau Server installations that support the production installation. For more information on Deployment, see the EULA Documentation.

- The Data Management Add-on can only be activated on a licensed Tableau Server Deployment. For more information on how to purchase the Data Management Add-on
and get the product key, contact your account manager.

- When the product key is active and enabled, you can publish flows to the server, schedule or run flows, manage and monitor flows.
- When the Data Management product key is removed or deactivated, or if the Data Management Product Key expires, then the ability to publish, schedule, run flows, or receive flow alerts is disabled. You will not be able to see the flows, but the data will remain on the Tableau Server and will not be deleted.

**Note:** If your Tableau Server product key is still active and valid, you can download the flows using the Tableau Server REST API. For more information, see Flow Methods.

**How Data Management Add-on licensing works**

**User-Based:**

A User-Based license metric allows you to deploy Tableau Server on a single computer or on multiple computers in a cluster. Each user that accesses Tableau Server must be licensed. Administrators add users and license them. The first Creator or Explorer product key that you add to Tableau Server activates Tableau Server, and will be used by a Server Administrator.

Data Management Add-on product key enables Tableau Prep Conductor at the deployment level and Tableau Prep Conductor is licensed for all the users that are already licensed for Tableau Server.

**Core-Based:**

A Core-Based license metric imposes no constraints on the number of user accounts in Tableau Server. Instead, the license specifies the maximum number of computer cores on which you can run Tableau Server.

Typically, the total number of cores in all the computers should not exceed the total number that the Tableau Server license allows. When you add the Data Management product key to your Tableau server, it includes a specific number of Tableau Prep Conductor cores. In this scenario, the total number of cores on all computers should not exceed the total number that the Tableau Server license and the Data Management license together allow.
In this topic we will refer to the cores licensed through Tableau Server license as Tableau Server cores, and the cores licensed through Data Management Add-on as Tableau Prep Conductor cores.

Here are some concepts that apply to how licensing is applied in a Core-Based metric:

- A node can be licensed by only one of type – Tableau Server Cores or Tableau Prep Conductor cores.

- The Tableau Prep Conductor cores are applied to any node that is dedicated to running Tableau Prep Conductor and when the Backgrounder on that node is set to run only flow background jobs. In this case, the total number of cores on this node cannot exceed the number of cores that Data Management Add-on license allows. If this node has any other licensed process besides Tableau Prep Conductor, Backgrounder and Data Engine enabled, then this node will require and use a Tableau Server core license.

- As mentioned above, the Backgrounder node role also affects which license is used by a node. For example, if the Backgrounder node role is set to run jobs of all types (this is the default), then this node will be licensed through the Tableau Server cores. For more information on node roles, see Node Roles in Tableau Server.

See the following table and decision flow to understand how a node is licensed:

<table>
<thead>
<tr>
<th>If a node has...</th>
<th>the core on the node is counted towards...</th>
<th>the node is licensed using...</th>
</tr>
</thead>
<tbody>
<tr>
<td>one of the following processes enabled:</td>
<td>Total count of Tableau Server cores.</td>
<td>Tableau Server cores.</td>
</tr>
<tr>
<td>• Application Server</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Backgrounder (node role is set to run all jobs)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• File Store</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Data Server</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• VizQL Server</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
only the following processes enabled:

- Tableau Prep Conductor
- Backgrounder (node role is set to run only flows)
- Data Engine

<table>
<thead>
<tr>
<th>Total number of Tableau Prep Cores purchased through the Data Management Add-on.</th>
</tr>
</thead>
</table>

**Note:** If there are no Tableau Prep Conductor cores available, but Tableau Server cores are available, then Tableau Server cores are used.

<table>
<thead>
<tr>
<th>Tableau Prep Cores included in the Data Management Add-on.</th>
</tr>
</thead>
</table>

**Note:** If there are no Tableau Prep Conductor cores available, but Tableau Server cores are available, then Tableau Server cores are used.
For more information about licensed processes, see Tableau Server Processes.

To learn more about Tableau Server licensing, see Tableau Server Licensing Overview.

Enable and Configure Tableau Prep Conductor on Tableau Server

Tableau Prep Conductor is supported only on Tableau Server versions 2019.1 or later.
Tableau Prep Help

Tableau Prep Conductor is licensed through the Data Management Add-on, on a per Deployment basis, which is User-Based or Core-Based. A Deployment includes a licensed production Tableau Server installation and licensed non-production Tableau Server installations that support the production installation. For more information on Deployment, see the EULA Documentation.

For more information on how Tableau Prep Conductor licensing works, see Licensing Tableau Prep Conductor for Tableau Server.

Who can do this?

Server administrators can install Tableau Server and enable Tableau Prep Conductor.

Server-level settings can be configured by Tableau Server administrators, and site-level settings can be configured by Tableau Server and Site administrators.

Server Topology

When you install Tableau Server and enable Tableau Prep Conductor, using the Data Management product key, Tableau Prep Conductor is automatically enabled by default by the setup program.

For multi-node installations, by default, one instance of Tableau Prep Conductor is enabled on any node that has backgrounder installed. In the example below, Tableau Prep Conductor is enabled on node 2 and 3 where the backgrounders are also enabled, but not on node 1, 4, and 5.
Below is a visual representation of that work-flow:
Next step:

New Tableau Server Installations: **Step 1 (New Install): Install Tableau Server with Tableau Prep Conductor** below

Existing Tableau Server Installations: **Step 1 (Existing Install): Enable Tableau Prep Conductor** on page 252

**Step 1 (New Install): Install Tableau Server with Tableau Prep Conductor**

This topic describes how to Tableau Prep conductor on a new installation of Tableau Server.

Tableau Prep Conductor is supported only on Tableau Server versions 2019.1 or later.

Tableau Prep Conductor is licensed through the Data Management Add-on, on a per Deployment basis. A Deployment includes a licensed production Tableau Server installation and licensed non-production Tableau Server installations that support the production installation. For more information on Deployment, see the EULA Documentation.

Who can do this?

Server administrators can install Tableau Server and enable Tableau Prep Conductor.

Before you install

The recommended topology for a production Tableau Server installation is a dedicated node for running flows. If you are currently planning to have a single node Tableau Server installation it is recommend that you add a second node and dedicate it to run flows.

- Review the hardware recommendations for Tableau Server and Tableau Prep conductor.
  - Minimum Hardware Requirements and Recommendations for Tableau Server installation on Windows.
  - Minimum Hardware Requirements and Recommendations for Tableau Server installation on Linux.
Install Tableau Server and enable Tableau Prep Conductor

Use the instructions provided in the following topics to install Tableau Server.

**Windows: Install Tableau Server**

**Linux: Install Tableau Server topic**

When you get to the **Activate** step, use the Tableau Server product keys to activate Tableau Server.

All product keys are available through the **Customer Portal**.

Configure public gateway settings

If your Tableau Server is set up with one of the following:

- Load balancer to distribute requests across gateways.
- Reverse proxy to authenticate external (internet) client requests and offloading SSL-based encryption.

You must configure the following public gateway settings:

```
tsm configuration set -k gateway.public.host -v <name> (This should be the URL that your users are using to access Tableau Server)
```

```
tsm configuration set -k gateway.public.port -v 443
```

For more information on configuring gateway settings, see **Configuring Proxies for Tableau Server**.

Enable Tableau Prep Conductor

**Use the following steps to add the Data Management product key to your Tableau Server:**

**Note:** This process requires a restart of the Tableau Server.

1. If the computer where you are running Tableau Server has been configured to connect to the internet through a forward proxy, follow the procedure in the topic, **Configure Product Key Operations with Forward Proxy**, before continuing.

2. Open TSM in a browser:

   `https://<tsm-computer-name>:8850`
3. Click Licensing on the Configuration tab and click Activate License:

4. Enter or paste your Data Management product key and click Activate:

5. On the Register page, enter your information into the fields and click Register.

6. Follow the prompts and restart Tableau after registration is complete.

Verify Tableau Prep Conductor is enabled and running

When you activate the Data Management product key, a single instance of Tableau Prep Conductor is automatically enabled on any node that has Backgrounder enabled.

Use the following steps to verify that it is enabled and running:

1. Open a browser and enter the Tableau Server URL, and append the dedicated TSM web UI port. Here are some examples of what the URL might look like:

   https://localhost:8850/ (if you’re working directly on the server computer)
   https://MarketingServer:8850/ (if you know the server’s name)
   https://10.0.0.2:8850/ (if you know the server’s IP address)

   In the sign-in page that appears, enter your administrator user name and password.

   **Note:** Tableau Server creates and configures a self-signed certificate during the installation process. This certificate is used to encrypt traffic to the TSM Web UI. Because it’s a self-signed certificate, your browser will not trust it by default. Therefore, your browser will display a warning about the trustworthiness of the certificate before allowing you to connect.

2. In the Tableau Services Manager web interface, click the Status tab to see the status.

   - If Tableau Prep Conductor is enabled and running, you should see Tableau Prep Conductor in the list of processes as Active on at least one node. If Tableau Prep Conductor is not enabled, you will see Tableau Prep Conductor in the list of processes, but with no status information for any of the nodes.

   **Tableau Prep Conductor not enabled:**
Tableau Prep Conductor enabled and running. In the image below, Tableau Prep Conductor is enabled on node1 and node3:
Dedicate a node for Tableau Prep Conductor

On the node you are planning to dedicate to running flows, enable Backgrounder process if it is not already enabled. It is recommended that you do not run other processes like VizQL server on this node.

Because you are dedicating this node to running flows, you must configure Backgrounder to run only flow tasks. By default, the Backgrounder process runs tasks of all types, including flows, extract refreshes, and subscriptions. For more information, see Node Roles in Tableau Server.

Run the following tsm commands on that dedicated node to run only flow tasks:

1. Run the following command to allow Backgrounders on this node to run only flow tasks.
   ```bash
tsm topology set-node-role -n node1 -r flows
   ```
2. Set the node role on the initial node to no flows. The backgrounder on this node will run all jobs except flows:
   ```bash
tsm topology set-node-role -n node1 -r no-flows
   ```
3. Apply the changes and restart Tableau Server:
   ```bash
tsm pending-changes apply
   ```

Multi-node installations

If you have more than 2 nodes in your Tableau Server installation, you can choose to configure other nodes to run all tasks other than flows:

1. Restrict a node to not allow flows. This command removes Tableau Prep Conductor from this node and Backgrounders on this node will not run flow tasks.
   ```bash
tsm topology set-node-role -n node1 -r no-flows
   ```
2. Apply the changes and restart Tableau Server:
   ```bash
tsm pending-changes apply
   ```

Next step

Step 2 - Configure Flow Settings for your Tableau Server on page 264
Step 1 (Existing Install): Enable Tableau Prep Conductor

This topic describes how to enable Tableau Prep conductor on your existing installation of Tableau Server.

Tableau Prep Conductor is supported only on Tableau Server versions 2019.1 or later. If you are using Tableau Server 2018.3 or earlier, you must first upgrade your Tableau Server to 2019.1 before enabling Tableau Prep Conductor on your Tableau Server installation.

Tableau Prep Conductor is licensed through the Data Management Add-on, on a per Deployment basis, which is User-Based or Core-Based. A Deployment includes a licensed production Tableau Server installation and licensed non-production Tableau Server installations that support the production installation. For more information on Deployment, see the EULA Documentation.

This topic describes how to enable Tableau Prep conductor on your existing installation of Tableau Server.

Who can do this?

Tableau Server Administrators can install or upgrade Tableau Server, and enable Tableau Prep Conductor on Tableau Server.

Before you upgrade

Prepare for upgrade:

- Know before you upgrade
- Licensing Tableau Prep Conductor
- Tableau Server Hardware Requirements and Recommendations

Configure public gateway settings

If your Tableau Server is set up with one of the following:

- Load balancer to distribute requests across gateways.
- Reverse proxy to authenticate external (internet) client requests and offloading SSL-based encryption.

You must configure the following public gateway settings:
Tableau Prep Help

tsm configuration set -k gateway.public.host -v <name> (This should be the URL that your users are using to access Tableau Server)

tsm configuration set -k gateway.public.port -v 443

For more information on configuring gateway settings, see Configuring Proxies for Tableau Server.

Tableau Server Installations using User-Based licenses

The recommended topology for a production Tableau Server installation is a dedicated node for running flows. For more information, see Minimum Hardware Requirements and Recommendations for Tableau Server.

Tableau Server single-node installations

If you currently have a single node Tableau Server installation, it is recommended that you add a second node and dedicate it to running flows.

1. Run upgrade on your current Tableau Server installation using the information in the topics below:
   - Windows
   - Linux

   When you get to the Activate step, use the Tableau Server product keys to activate Tableau Server.

   All product keys are available through the Customer Portal.

2. After completing the installation, add the Data Management product key to enable Tableau Prep Conductor on your node. The Data Management product key, like your other server keys, are available through the Customer Portal.

   - In the Tableau Services Manager web interface, click Licensing on the Configuration tab and click Activate License.

   - Enter or paste your new product key and click Activate.

   - On the Register page, enter your information into the fields and click Register.

3. You will be prompted to restart the server. Restart the server and verify that Tableau Prep Conductor is enabled and running.
In the Tableau Services Manager web interface, click the **Status** tab to see the status. If Tableau Prep Conductor is enabled and running, you should see Tableau Prep Conductor in the list of processes as **Active**. If Tableau Prep Conductor is not enabled, you will see Tableau Prep Conductor in the list of processes, but with no status information.

**Tableau Prep Conductor not enabled:**

**Tableau Prep Conductor enabled and running:**

4. Add a second node to your Tableau Server installation. The installer will enable certain required processes like the Cluster Controller. Enable Backgrounder process on it as it is required to run scheduled flow tasks. When you enable the Backgrounder process,
the installer automatically enables a single instance of Data Engine and Tableau Prep Conductor on the node. Do not add any other processes on this node.

5. Run the following commands to dedicate this node to do only flow tasks. For more information on node roles, see Node Roles in Tableau Server.
   - Get the nodeID for your dedicated node to see the list of services on each node:
     
     tsm topology list-nodes -v.
   - Set the node role for the dedicated node using the nodeID that you got from running the command described above:
     
     tsm topology set-node-role -n <nodeID> -r flows.
   - Apply the changes, and restart the server:
     
     tsm pending-changes apply.
   - Review the status to ensure that all the processes are up and running and configured correctly:
     
     tsm status -v.

You have successfully added Tableau Prep Conductor to your Tableau Server installation.

Tableau Server multi-node installations

1. Run upgrade on your current Tableau Server Installation using the information in the topics below:
   - Windows
   - Linux

   When you get to the **Activate** step, use the Tableau Server product keys to activate Tableau Server.

   All product keys are available through the Customer Portal.

2. After completing the installation, add the Data Management product key to enable Tableau Prep Conductor. Tableau Prep Conductor is automatically enabled on the nodes where you already have the Backgrounder process enabled. The Data Management product key, like your other server keys, are available through the Customer Portal.

   - In the Tableau Services Manager web interface, click **Licensing** on the **Configuration** tab and click **Activate License**.
• Enter or paste your new product key and click **Activate**.

• On the **Register** page, enter your information into the fields and click **Register**.

3. You will be prompted to restart the server. Restart the server and verify that Tableau Prep Conductor is enabled and is running.

• In the Tableau Services Manager web interface, click the **Status** tab to see the status of all the processes. If Tableau Prep Conductor is enabled and running, you should see Tableau Prep Conductor in the list of processes as **Active**. If Tableau Prep Conductor is not enabled, you will see Tableau Prep Conductor in the list of processes, but with no status information.

**Tableau Prep Conductor not enabled:**

---

![Tableau Services Manager Screen](image-url)
Tableau Prep Conductor enabled and running:

4. Add a new node to your Tableau Server installation. The installer will enable certain required processes like the Cluster Controller. Enable Backgrounder process on it as it is required to run scheduled flow tasks. When you enable the Backgrounder process, the installer automatically enables a single instance of Data Engine and Tableau Prep Conductor on the node. Do not add any other processes on this node.

Note: The dedicated note counts towards to total count of the Coordination Service ensemble. You may need to deploy a Coordination Service on the new node depending on the total number of nodes you have in your cluster including the new dedicated node. For more information, see Deploy a Coordination Service Ensemble.

5. Run the following command to dedicate this node to only doing flow related operations. For more information on node roles, see Node Roles in Tableau Server.
   - Get the nodeID for your dedicated node to see the list of services on each node:
     - `tsm topology list-nodes -v`
   - Set the node role for the dedicated node using the nodeID that you got from running the command described above:
- `tsm topology set-node-role -n <nodeID> -r flows`.
- Apply the changes and restart the server: `tsm pending-changes apply`.
- Review the status to ensure that all the processes are up and running and configured correctly: `tsm status -v`.

6. At this stage, you may have Tableau Prep Conductor enabled on other nodes. By default, the Backgrounder process on a node performs all tasks of all types including flow tasks. To isolate Tableau Prep Conductor and flow tasks to only certain nodes, you can configure the Backgrounders to do one of the following:

   - To run only flow tasks: `tsm topology set-node-role -n <nodeID> -r flows`.
   - To run all other tasks except flows: `tsm topology set-node-role -n <nodeID> -r no-flows`.

You have successfully added Tableau Prep Conductor to your Tableau Server installation.

**Tableau Server Installations using Core-Based licenses**

The recommended topology for a production Tableau Server installation is a dedicated node for running flows. For more information, see Minimum Hardware Requirements and Recommendations for Tableau Server.

The Data Management Add-on for Core-Based licenses includes product keys that enable Tableau Prep Conductor for your Tableau Server, and Tableau Prep Conductor cores that come in units of four. The Tableau Prep Conductor cores should be applied to the node dedicated to running the flows. These product keys, like your other server keys, are available through the Customer Portal.

To learn more about Tableau Prep Conductor licensing, see Licensing Tableau Prep Conductor for Tableau Server.

**Tableau Server single-node installations**

If you currently have a single node Tableau Server installation, it is recommended that you add a second node and dedicate it to running flows.

1. Run upgrade on your current Tableau Server Installation using the information in the topics below:
2. Activate the product keys. This will enable Tableau Prep Conductor on the nodes where you already have the Backgrounder process enabled.

   - In the Tableau Services Manager web interface, click **Licensing** on the **Configuration** tab and click **Activate License**.
   - Enter or paste your new product key and click **Activate**.
   - On the **Register** page, enter your information into the fields and click **Register**.

3. You will be prompted to restart the server. Restart the server and verify that Tableau Prep Conductor is enabled and is running.

   - In the Tableau Services Manager web interface, click the **Status** tab to see the status. If Tableau Prep Conductor is enabled and running, you should see Tableau Prep Conductor in the list of processes as **Active**. If Tableau Prep Conductor is not enabled, you will see Tableau Prep Conductor in the list of processes, but with no status information.

### Tableau Prep Conductor not enabled:

![Tableau Services Manager screenshot](image-url)
4. Add a second node to your Tableau Server installation. The installer will enable certain required processes like the Cluster Controller. Enable Backgrounder process on it as it is required to run scheduled flow tasks. When you enable the Backgrounder process, the installer automatically enables a single instance of Data Engine and Tableau Prep Conductor on the node. Do not add any other processes on this node.

**Important:** The number of physical cores on this machine must be equal to, or less than the Tableau Prep Conductor cores you purchased. For example, if you purchased four Tableau Prep Conductor cores, your node can only have up to four physical cores. To understand about how Tableau Prep Conductor licensing works, see Licensing Tableau Prep Conductor for Tableau Server.

5. Run the following commands to dedicate this node to only doing flow tasks. For more information on node roles, see Node Roles in Tableau Server.

- Get the nodeID for your dedicated node to see the list of services on each node:
  
  tsm topology list-nodes -v.

- Set the node role for the dedicated node using the nodeID that you got from running the command described above:
  
  tsm topology set-node-role -n <nodeID> -r flows.

- Apply the changes and restart the server: tsm pending-changes apply.

- Review the status to ensure that all the processes are up and running and
configured correctly:

```
tsm status -v.
```

You have successfully added Tableau Prep Conductor to your Tableau Server installation.

**Tableau Server multi-node installations**

1. Run upgrade on your current Tableau Server installation using the information in the topics below:
   - **Windows**
   - **Linux**

2. Activate the product keys. This will enable Tableau Prep Conductor on the nodes where you already have the Backgrounder process enabled.
   - In the Tableau Services Manager web interface, click **Licensing** on the **Configuration** tab and click **Activate License**.
   - Enter or paste your new product key and click **Activate**.
   - On the **Register** page, enter your information into the fields and click **Register**.

3. You will be prompted to restart the server. Restart the server and verify that Tableau Prep Conductor is enabled and is running.
   - In the Tableau Services Manager web interface, click the **Status** tab to see the status. If Tableau Prep Conductor is enabled and running, you should see Tableau Prep Conductor in the list of processes as **Active**. If Tableau Prep Conductor is not enabled, you will see Tableau Prep Conductor in the list of processes, but with no status information.

   **Tableau Prep Conductor not enabled:**
4. Add a new node to your Tableau Server installation. A dedicated node to run flow related operations is recommended for production Tableau Server installations. The installer
will enable certain required processes like the Cluster Controller. Enable Backgrounder process on it as it is required to run scheduled flow tasks. When you enable the Backgrounder process, the installer automatically enables a single instance of Data Engine on the node. Do not add any other processes on this node.

**Note:** The dedicated note counts towards to total count of the Coordination Service ensemble. You may need to deploy a Coordination Service on the new node depending on the total number of nodes you have in your cluster including the new dedicated node. For more information, see Deploy a Coordination Service Ensemble.

**Important:**
The number of physical cores on this machine must be equal to, or less than the Tableau Prep Conductor cores you purchased. For example, if you purchased four Tableau Prep Conductor cores, your node can only have up to four physical cores. To understand about how Tableau Prep Conductor licensing works, see Licensing Tableau Prep Conductor for Tableau Server.

5. Run the following commands to dedicate this node to only doing flow tasks. This will enable Tableau Prep Conductor on your new node. For more information, see Node Roles in Tableau Server.

- Get the nodeID for your dedicated node to see the list of services on each node:
  
  tsm topology list-nodes -v.

- Set the node role for the dedicated node using the nodeID that you got from running the command described above:
  
  tsm topology set-node-role -n nodeID -r flows.

- Apply the changes and restart the server:
  
  tsm pending-changes apply.

- Review the status to ensure that all the processes are up and running and configured correctly:
  
  tsm status -v.

6. At this stage, you may have Tableau Prep Conductor enabled on other nodes that have the Backgrounder process. By default, the Backgrounder process on a node performs all
tasks of all types including flow tasks. To isolate Tableau Prep Conductor and flow operations to only certain nodes, you can configure the backgrounders to do one of the following:

- To run only flow tasks:
  
  tsm topology set-node-role -n <nodeID> -r flows.

- To run all other tasks except flows:
  
  tsm topology set-node-role -n <nodeID> -r no-flows.

Next step

**Step 2 - Configure Flow Settings for your Tableau Server** below.

**Step 2 - Configure Flow Settings for your Tableau Server**

This topic describes the various flow settings that you can configure for your Tableau Server.

Who can do this?

Tableau Server Administrators can configure server and site level settings. Tableau Site Administrators can configure site level settings.

**Publishing, Scheduling, and Credential Settings**

When you activate Tableau Prep Conductor using the Data Management product key, Tableau Prep Conductor is enabled for the entire Tableau Server installation. You can further modify and customize the setting for sites.

Use the following instructions to configure settings related to flows for all your sites or for individual sites:

Use the following instructions to sign in to Tableau Server Admin pages:

- Windows: Tableau Server Admin Pages.
- Linux: Tableau Server Admin Pages.

**Configure whether publishing and scheduling flow should be allowed for a site:**

1. **Enable users to publish and schedule flows:** This setting is enabled by default when you enable Tableau Prep Conductor. If you have multiple sites, you can selectively
turn off Tableau Prep for Server for individual sites. If you disable this setting for a site that once allowed flows, see Implication of disabling Tableau Prep Conductor below to learn more.

On the General page under Settings, scroll to the Flows setting and clear the Allow user to publish and schedule flows check box to turn off Tableau Server for the site.

2. Embed Credentials

- Allow publishers to embed credentials in a data source, flow or workbook: This setting allows publishers to attach passwords to published flows that will automatically authenticate web users.

- Allow publishers to schedule flow runs and data extract refreshes: This option is only available if setting above is enabled. When this setting is enabled, publishers will see scheduling options in the Publish dialog box.

Implication of disabling Tableau Prep Conductor

If you disable Tableau Prep Conductor after using it for a while, you will not be able to see the flows, schedules, tasks, and other things related to flows. The following table gives you more information on what you can and cannot see when you disable Tableau Prep for Server completely or only for specific sites:
<table>
<thead>
<tr>
<th></th>
<th>Prep not enabled at Server level</th>
<th>Prep enabled at Server level, but disabled for a site</th>
<th>Prep enabled for both Server and site</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Show flows</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Show tasks/schedules in Server view</strong></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Show tasks/schedules in Site view</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Show Site setting (only for Server Admins)</strong></td>
<td>Yes (disabled)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Show TSM status</strong></td>
<td>Yes (Tableau Prep Conductor is not shown)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Show TSM settings</strong></td>
<td>Yes (disabled)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Important**: Scheduled tasks will continue to run even when Tableau Prep Conductor is disabled for that site, but will fail.

**Configure notifications for flow failures**

You can configure Tableau Server to send email notifications for flow run failures. The notifications are sent for failures that occur when running the flows through either a scheduled task or a manual run using the **Run now** menu option. You must first enable the server-wide setting, and then configure at the site level.

**To enable the server-wide email notification**

You can either use the Tableau Services Manager (TSM) web interface or TSM CLI as described below:

**Use the TSM web interface**
1. Open TSM in a browser:
   https://<tsm-computer-name>:8850.
2. Click **Notifications** on the **Configuration** tab and click **Email Server**.
3. Enter the email server information.
4. Click the **Events** tab.
5. Under **Content Updates**, select **Send emails for extract refresh failures and flow run failures** if not already turned on by default.
6. Click **Save Pending Changes** after you've entered your configuration information.
7. Click **Apply Changes and Restart**.

---

**Use the TSM CLI**

The notification values can be set individually with the tsm configuration set command:

Windows: `tsm configuration`.

Linux: `tsm configuration`.

**Set notification values**

Use the tsm configuration set command with the following syntax to enable flow failure notifications, run the following command:

```
tsm configuration set -k backgrunder.notifications_enabled -v true
```

**Note:** This will enable email notification for both extract refresh failures and flow failures.

After you are done setting values, you must run the following command:

```
tsm pending-changes apply
```

The pending-changes apply command displays a prompt to let you know this will restart Tableau Server if the server is running. The prompt displays even if the server is stopped, but in
the case there is no restart. You can suppress the prompt using the --ignore-prompt option, but this does not change the restart behavior.

To configure email notification for a site:

On the **General** page under **Settings**, scroll to the **Flow Run and Refresh Failure Notifications** setting and clear the **Send email to data source, flow, and workbook, when flow runs and scheduled refreshes fail** check box to turn off refresh failure and flow run notifications.

Next step

**Step 3- Create Schedules for Flow Tasks** below

**Step 3- Create Schedules for Flow Tasks**

Who can do this?

Tableau Server Administrators can create and modify schedules. Schedules are created at the server level and apply across all the sites on a server.
Create a new schedule:

1. On the Schedules tab, click New Schedule.

2. Enter the following information in the New Schedule dialog box and click Create.
   - **Name**: Enter a descriptive name for the schedule. Typically, this includes the description of the schedule frequency.
   - **Task Type**: Select Flow as the task type.
   - **Default Priority**: You can define a default priority from 1 to 100, where 1 is the highest priority. This value will be assigned to the tasks by default. If two tasks are pending in the queue, the one with the higher priority runs first.
   - **Execution**: Choose whether a schedule will run in parallel or serially. Schedules that run in parallel run on all available backgrounder processes so that they can complete faster.
- **Frequency:** You can define an hourly, daily, weekly, or monthly schedule.

![New Schedule](image)

**Next step**

**Step 4 - Safe list Input and Output locations below**

**Step 4 - Safe list Input and Output locations**

This topic describes the rules that apply to this feature and how to safe list the directories on your network.

Flow input and output connections may need to connect to databases or files in the directories on your network. You must safe list the directories you want to allow access to. Input and Output connections will only be allowed to connect to data in the safe listed locations. By default, no connections are allowed. You can still publish the flows and any data that is embedded in the flow file (tflx) to Tableau Server.

**Who can do this?**

Tableau Server Administrators who also are a member of the tsmadmin group can configure settings using tsm commands.
How to safe list input and output locations

The following rules apply and must be considered when configuring this setting:

- The directory paths should be accessible by Tableau Server. These paths are verified during server startup and at flow run time and are **not** verified at the time of publishing the flow to Tableau Server.

- Network directory paths have to be absolute and cannot contain wildcards or other path traversing symbols. For example, `\\myhost\myShare\*or\\myhost\myShare*` are invalid paths and would result in all the paths as disallowed. The correct way to safelist any folder under `myShare` would be `\\myhost\myShareor\\myhost\myShare\`.

  **Note:** The `\\myhost\myShare` configuration will not allow `\\myhost\myShare1`. In order to safe list both of these folders safe list them as `\\myhost\myShare;\\myhost\myShare1`.

- **Windows:**
  
  - The value can be either `*`, to allow any network directory, or a specified list of network directory paths, delimited by a semicolon (`;`). If the path contains spaces or special characters you will have to either use single or double quotes. Whether you use single or double quotes depends on the shell that you are using.

  - No local directory paths are allowed even when the value is set to `*`.

- **Linux:**
  
  - The value can be either `*` meaning that any path, including local (with the exception of some system paths configured using “native_api.internal_disallowed_paths”), or a list of paths, delimited by a semicolon (`;`).
  
  - You must be using a kernel version of equal to or later than 4.7. Safe listing is not supported on kernel version earlier than 4.7. To check the kernel version, in the Linux terminal, type the command `uname -r`. This will display the full version of the kernel you are running on the Linux machine.

  **Note:** If a path is both on the flows allowed list and internal_disallowed list, internal_disallowed takes precedence.

Use the following commands to create a list of allowed network directory paths:
**For input connections:**

```
tsm configuration set -k maestro.input.allowed_paths -v your_networkdirectory_path_1;your_networkdirectory_path_2
tsm pending-changes apply
```

**For output connections:**

```
tsm configuration set -k maestro.output.allowed_paths -v your_networkdirectory_path_1;your_networkdirectory_path_2
tsm pending-changes apply
```

Important:
These commands overwrite existing information and replace it with the new information you provided. If you want to add a new location to an existing list, you must provide a list of all the locations, existing, and the new one you want to add. Use the following commands to see the current list of input and output locations:

```
tsm configuration get -k maestro.input.allowed_paths
tsm configuration get -k maestro.output.allowed_paths
```

**Next step**

**Step 5 - Optional Server Configurations** below

**Step 5 - Optional Server Configurations**

The options described in this topic are not required to enable flow publishing and scheduling flows on Tableau Server. They can be used to customizing your environment according to your requirements.

**Who can do this?**

Tableau Server administrators can make changes to server configurations.

**Set the timeout period for flows**

You can set time limits for how long a flow can run to make sure that subsequent tasks are not held up due to stalled tasks. The following two `tsm` command options determine how long a
flow task can run before the flow background task is canceled. These two commands together determine the total timeout value for flow tasks.

Use the following tsm commands to change the default timeout period of 4 and half hours:

`backgrounder.extra_timeout_in_seconds`
(Default value: 14400 seconds or 4 hours)

`backgrounder.default_timeout.run_flow`
(Default value: 1800 seconds or 30 minutes)

**Set the threshold for suspended flow tasks**

By default, a flow task is suspended after 5 consecutive flow task failures. To change the threshold number of flow task failures that can occur before they are suspended, use the following tsm configuration set command:

`tsm configuration set -k backgrounder.flow_failure_threshold_for_run_prevention -v <number>`

This sets the threshold for the number of consecutive failed flow tasks necessary before suspending the tasks. This is a server-wide setting.

---

**Publish a Flow to Tableau Server**

Publish your flows to Tableau Server to automatically run them on a schedule and refresh the flow output using Tableau Prep Conductor.

**Who can do this?**

Server Administrator, Site Administrator Creator, and Creator allow full connecting and publishing access.

Explorer (can Publish) and Site Administrator Explorer have limited publishing capabilities, as described in the following topics:

**Tableau Server:**

- Windows: General capabilities allowed with each site role.
- Linux: General capabilities allowed with each site role.
Before you publish

To make sure that you can run your flow in Tableau Server, check the following:

1. Verify that there are no errors in the flow.

   Flows that contain errors will fail when you try to run them in Tableau Server. Errors in the flow are identified by a red exclamation mark and a red dot with an Errors indicator in the upper right corner of the canvas.

2. Verify that your flow doesn't include input connectors or features that aren't compatible with your version of Tableau Server.

   Flows that include connectors or features that aren't yet supported in your version of Tableau Server can be published, but can't be scheduled to run in Tableau Server. To publish and schedule flows to run on Tableau Server, you must be using Tableau Server version 2019.1 or later and Tableau Prep Conductor must be enabled.

   For example, the SAP HANA connector was introduced in Tableau Prep Builder version 2019.1.4 but this connector isn't supported in Tableau Server version 2019.1 for Tableau Prep Conductor. When you publish the flow, you would see a message like the example below.
To be able to run your flow in Tableau Server, you need to take the appropriate actions to make the flow compatible. For more information about working with incompatible flows, see Version Compatibility with Tableau Prep on page 226.

3. Flows that include input or output steps with connections to a network share require safe listing.

Flow input and output steps that point to files stored in a network share (UNC path) aren’t permitted unless the file and path is accessible by the server and are included in your organization’s safe list. If you publish the flow without adding the file location to your safe list, the flow will publish, but you will get an error when you try and schedule or run the flow in Tableau Server.

If the files aren’t stored in a safe listed location, you will see a warning message when you publish the flow.
Click the "list" link in the message to see a list of allowed locations. Move your files to one of the locations in the list, and make sure that you flow points to these new locations. Then try publishing the flow again.
To configure the allowed network paths, use the tsm command options described in **Step 4 - Safe list Input and Output locations** on page 270.

If you don’t want to move your files to a safe listed location, you will need to package the input files with the flow and publish the flow output to Tableau Server as a published data source. For more information about setting these options, see Publish a flow on the next page in this topic.

4. If publishing the flow output to Tableau Server, set each flow output step to publish as a data source. All flow output steps must point to the same server or site where the flow is published. Only one server or site can be selected.

To set the publishing location for all output steps, do the following:

a. In the flow pane, select the output step.

b. In the publishing pane, select Publish as a data source.

c. Select the server or site and the project where you want to publish the flow. Sign in to the server or site if needed.

d. Enter a name and description for each output.

   The output file name should be distinctive enough so that the person running the flow can easily identify which output files to refresh. The file name shows on the Overview and Connections page for the flow in Tableau Server.

e. Save your flow.
For more information about how to configure output steps for publishing, see Create and publish data extracts and data sources on page 214.

Publish a flow

1. Open your flow in Tableau Prep Builder.

2. From the top menu select Server > Publish Flow.

3. In the Publish Flow to Tableau Server dialog complete the following fields:
   - **Project**: Click the drop-down option to select your project from the project hierarchy. This should be the same project that the output files are published to.
   - **Name**: Enter a name for your flow. This name shows on the server on the Flow pages. If you want to overwrite an existing flow, click the drop-down option to select a name from the list.
   - **Description (optional)**: Enter a description for the flow.
Tableau Prep Help

- **Tags (optional):** Click **Add** to type in one or more tags to identify your flow so users can easily find it. Tags can also be added after publishing in the Flow pages in Tableau Server.

- **Authentication:** If your flow connects to one or more databases, select the authentication type that you want to use to connect to the flow input data sources. Click **Edit** to change your authentication type.

The following authentication types are supported:

- **Prompt User:** You must edit the connection and enter the database credentials before running the flow.

- **Embedded Password:** The credentials you used to connect to the data will be saved with the connection and used by everyone who accesses the flow you publish.

- **Server Run As Account:** The server’s **Run As User** account will authenticate all users.

- **Files:** If your flow connects to local files, click **Edit** to select whether to upload or connect directly to the input files.
- Uploaded files are packaged with your flow and aren’t refreshed if the data changes.

- To retrieve the most current data when refreshing the output files, use **Direct Connection** if Tableau Server can connect to the file location and the location is included in your organization’s safe list.

If your input or output steps point to files stored in a network share (UNC path) and the location isn't included in your organization's safe list, you will see a warning message. Click the link in the message to see a list of safe
listed locations, move your files and point your input and output steps to the new file location. For more information, see Step 3 in **Before you publish** on page 274.

For information about how to add locations to your organization’s safe list, see **Step 4 - Safe list Input and Output locations** on page 270.

**Note:** All files must have the same setting, either **Upload** or **Direct Connection**.

4. Click **Publish**. Then click **Done** to close the publishing status dialog. Tableau Server opens automatically in your default browser on the flow **Overview** page.

### Schedule a Flow Task

To schedule flows to run at a specific time or on a recurring basis you can create scheduled tasks. Scheduled tasks rely on pre-configured schedules. Schedules are created by the System Administrator (Tableau Server only). For information about how to create schedules on Tableau Server, see **Step 3- Create Schedules for Flow Tasks** on page 268.

### Who can do this?

- The Server administrator can do this on all sites on the server. The Site administrator can do this on sites they have access to if the site settings to allow users to publish and schedule tasks is enabled.
- Flow owners and project leaders can create flow tasks for flows or projects that they own respectively.

For more information, review the following articles:

- Windows: [Set Users’ Site Roles and Content Permissions and Ownership](#)
- Linux: [Set Users’ Site Roles and Content Permissions and Ownership](#)
Schedule a flow task

1. From the **Scheduled Tasks** tab, click **New Task**.

   If the output step isn’t assigned to a task, you can also create a new task from the **Overview** page. On that page, in the **Schedules** field, click **Create new task**.

2. In the **New Task** dialog, select a schedule from the drop-down list.

3. Select one of the following options:
   - **Automatically include all output steps for this flow**: (default) Select this option to include all current and future output steps for this flow in the scheduled task. As new output steps are added to the flow over time, they are automatically included in the schedule when it runs.
   - **Select the output steps to include in this task**: Select this option to manually
select the output steps to include in this scheduled task

4. Select the output steps to include in the flow task. To include all output steps in the flow task, select the check box next to Output Steps.

   This area can't be edited if the Automatically include all output steps for this flow radio button is selected. Select the other radio button to enable this section.

5. Click Create Task to create the scheduled task.

Manage a Flow

Once you publish a flow to Tableau Server, you can manage your flows and make changes to them as necessary. This topic describes the various actions you can take to manage your flows.

Who can do this?

Tableau Server Administrator

Can do the following tasks across all the sites:

- Run flows
- Delete flows
- Download flows
- Change Owner
- Change Permissions
- Change Project
- Add/Remove Tags
- Change Description
- Change Name
- Update Flow Task
- Delete Flow Task
- Create a Flow Task*
- Version Management*

*Some additional conditions apply to these actions:

- To create a flow task:
  - A flow schedule must be available. Only Server Administrators can create a schedule.
  - The flow must have at least one output step.

- Version management:
  - Revision history must be enabled on the site.
  - User role is allowed to publish to the project.

### Tableau Site Administrator

**Can do the following tasks on flows published to the sites that they are site administrator for:**

- Run flows
- Delete flows
- Download flows
- Change Owner
- Change Permissions
Some additional conditions apply to these actions:

- To create a flow task:
  - A flow schedule must be available. Only Server Administrators can create a schedule.
  - The flow must have at least one output step.

- Version management:
  - Revision history must be enabled on the site.
  - User role is allowed to publish to the project.

**Project Leader**

*Can do the following tasks on flows published to the projects where they have project leader permissions:*

- Run Flows
- Delete
- Download
- Change Owner
- Change Permissions
- Change Project
- Add/Remove Tags
• Change Description
• Change Name
• Update Flow Task
• Delete Flow Task
• Create a Flow Task*
• Version Management*

*Some additional conditions apply to these actions:

• To create a flow task:
  • A flow schedule must be available. Only Server Administrators can create a schedule.
  • The flow must have at least one output step.

• Version management:
  • Revision history must be enabled on the site.
  • User role is allowed to publish to the project.

Project Owner

Can do the following tasks on flows published to the projects that they own:

• Run Flows
• Delete
• Download
• Change Owner
• Change Permissions
• Change Project
• Add/Remove Tags
• Change Description
• Change Name
Tableau Prep Help

- Update Flow Task
- Delete Flow Task
- Create a Flow Task*
- Version Management*

* Some additional conditions apply to these actions:

- To create a flow task:
  - A flow schedule must be available. Only Server Administrators can create a schedule.
  - The flow must have at least one output step.

- Version management:
  - Revision history must be enabled on the site.
  - User role is allowed to publish to the project.

Flow Owner

Can do the following tasks on flows that they own:

- Run flows
- Delete flows
- Download flows
- Change Owner
- Change Permissions
- Change Project
- Add/Remove Tags
- Change Description
- Change Name
- Update Flow Task
- Delete Flow Task
• Create a Flow Task*
• Version Management*

*Some additional conditions apply to these actions:

• To create a flow task:
  • A flow schedule must be available. Only Server Administrators can create a schedule.
  • The flow must have at least one output step.

• Version management:
  • Revision history must be enabled on the site.
  • User role is allowed to publish to the project.

User with Creator License

Can do the following tasks:

• Run flows (with Run flow permissions)
• Delete (with delete permissions)
• Download (with download or Save as, and read permissions)
• Change Permissions (with Change Permissions)
• Change Project (with move permissions, and write permissions on the destination project)
• Add/Remove Tags (with read permissions)
• Change Description (with Save permissions)
• Change Name (with Save permissions)
• Update Flow task (with Execute permissions)
• Delete Flow Task (with Execute permissions)
• Create a Flow Task* (with Execute permissions)
• Version Management* (with view, read, save as, download permissions)

* Some additional conditions apply to these actions:
To create a flow task:
  - A flow schedule must be available. Only Server Administrators can create a schedule.
  - The flow must have at least one output step.

Version management:
  - Revision history must be enabled on the site.
  - User role is allowed to publish to the project.

**User with Explorer License**

**Can do the following tasks:**

- Run flows (with Run flow permissions)
- Delete (with delete permissions)
- Download (with download or Save as, and read permissions)
- Change Permissions (with Change Permissions)
- Change Project (with move permissions and write permissions on the destination project)
- Add/Remove Tags (with read permissions)
- Change Description (with Save permissions)
- Change Name (with Save permissions)
- Update Flow task (with Execute permissions)
- Delete Flow Task (with Execute permissions)
- Create a Flow Task* (with Execute permissions)
- Version Management* (with view, read, save as, download permissions)

*Some additional conditions apply to these actions:

To create a flow task:
  - A flow schedule must be available. Only Server Administrators can create a schedule.
  - The flow must have at least one output step.
  - The flow version must be compatible with the Tableau Server version.
Version management:

- Revision history must be enabled on the site.
- User role is allowed to publish to the project.

User with Viewer License

Viewers cannot manage flows, they can however view the flow and the different versions of the flow.

For more information about the full capabilities you can set on flows, see Set Permissions on Individual Content Resources.

Managing your flows

Following is a list of actions you can take to manage your flows:

- **Run flows:** You can manually run a flow in addition to creating scheduled flows tasks that run at a specific time.
  
  Navigate to the list of flows, select one or more flows you want to run, select Actions and click Run now.

- **Tag:** Tags are keywords you can create for flows to help you find, filter, and categorize content. Authors can add tags to flows when they publish it. But you can also add tags to any workbook, view, or data source that you are allowed to access and you can delete any tags you have added. You can add a tag to a list of flows.
  
  Navigate to the list of flows, select one or more items you want to tag, select Actions and click Tag. To add a tag to a specific flow you can do it from the list of flows as described above. Open the Flow, from the Overview tab, select Actions, and click Tag.

- **Change Owner:** Administrators, flow owners, and project leaders can change owners.

- **Permissions:** You can set permissions for users and specify if they can perform edit actions like save, download, move to a different project and delete. In addition, you can specify who can view and run the flow.

- **Download:** You can download a flow to view or modify it using your Tableau Prep. To download a flow, you need download permissions. You’ll have that by default if you are the owner, but you might need to add it for other users.
Revision History: When you make a change to the flow, and republish it to the same project with the same name, a new version of the flow is created. You can view the revision history by selecting revision history from the actions menu. Flow owners have permissions to restore a previous version of a flow.

Move: You can move flows between projects. To move a flow, users need Move permission. You'll have that by default for flows you own, but you might need to add it for other users.

Rename: You can rename a flow. To rename a flow, users need the Save permission. You'll have that by default for flows you own, but you might need to add it for other users.

Delete: You can delete a flow. To delete a flow, users need the Delete permission. You'll have that by default for flows you own, but you might need to add it for other users.

Monitor Flow Health and Performance

After you publish flows and schedule them to run periodically, you want to know that they are running as expected and resolve any issues as they occur. You will also want to monitor and understand the performance of your flows.

This topic describes the various methods that Tableau Server provides to help you monitor your flows.

Detect issues as they occur and resolve them

You can set up Tableau Server to send email notifications when flows fail, or find and review errors on Tableau Server using the Alerts menu or by reviewing the flow pages for the flows that you are interested in. This type of monitoring allows you to detect problems as they occur.

Who can do this?

- **Tableau Server Administrators:**
  - Setup email notifications at the server level
  - Set up email notifications for a site
  - View errors
  - Resume suspended tasks
  - View alerts
  - view process status
Tableau Site Administrators:
- Set up email notifications at the site level
- View errors
- Resume suspended tasks
- View alerts

Flow owners, project leaders and any user who is granted permissions to view the flow:
- View errors
- Resume suspended tasks
- View alerts (Flow owners)

Get email notifications when a flow fails:
You can configure Tableau Server to send email notifications to flow owners when a flow fails.
To set up email notifications for your Tableau Server and sites, follow the instructions in Step 2 - Configure Flow Settings for your Tableau Server on page 264.

View and resolve errors
The following errors can happen when running a flow:

- **Connection errors**: Connection errors generally happen when Tableau Server is unable to connect to one or more data inputs or is unable to make a connection in one or more output steps.
  - For Input connection errors, use the **Edit connections** option on the **Connections** tab to make changes to connection details, then run the flow again.
  - For output connection errors, download the flow and check the output location for the flow output steps. If the flow output is going to a network share, make sure the output steps are pointing to a safe listed location. After you make any changes republish the flow and try running it again.

- **Errors in the flow**: If there are errors in one or more steps in the flow, you will see an error message. In such cases, download the flow to Tableau Prep Builder, resolve the errors, republish the flow to the server and run the flow again.

- **Suspended flow tasks**: When a scheduled flow task fails to run after a configured number of attempts, the flow task is suspended. A flow can have multiple scheduled tasks assigned to it, but only the failed tasks are suspended. All other flow tasks will continue to run unless they have errors. To resolve a suspended task, review and
resolve the errors then run the flow on-demand or let the flow run automatically based on the assigned schedule.

**Note:** By default, a flow task is suspended after 5 consecutive flow tasks failures. Server administrators can change this default by using the tsm configuration set option. For more details, see Step 5 - Optional Server Configurations on page 272.

You can view errors on the following pages:

**Flow Overview page**

On this page you can see the status of the most recent flow run and any errors. Hover on the error text to review the error details. If a scheduled task is suspended, a warning icon shows next to the schedule. Hover on the icon to view the status.

After you resolve the error that caused the flow to fail or the task to be suspended, you can run the flow manually or let the flow run based on the assigned schedule. For suspended flow tasks, click the **Go to Scheduled Task** link on the tooltip for the suspended task to navigate to the **Scheduled Tasks** page and click the **Resume Scheduled Tasks** button to resume the suspended tasks.

**Connections page**

The **Connections** page shows the most recent status and any related connectivity errors. To correct input errors, click the **More actions** menu for an input connection to edit the
connection and change the server name, port, user name and password.

To fix output connection errors, download the flow in Tableau Prep Builder, correct the file path, then republish the flow to continue running it.

Scheduled Tasks page

View the scheduled tasks assigned to a flow. If a scheduled task is suspended, you can see the status of that tasks here and you can manually resume the flow tasks from this page. Before resuming a suspended task, resolve any errors in the flow. Errors are not shown on this page, but you can review them on the Overview or Run History pages.

A suspended scheduled task will automatically resume when the flow is republished, if you edit a connection for the flow or manually run the flow tasks. To manually resume a suspended tasks, on the Scheduled Tasks page, click Resume Scheduled Tasks. This resumes all suspended tasks for the flow.

To resume individual tasks, click the More actions menu for a scheduled task and select Resume. You can also click Run Now to run all tasks for the flow immediately.

Run History page

The Run History page shows the details of all the flow runs that have either completed or are in progress for each output. View any error details by hovering over the errors in the Errors column.
Alerts

When a flow fails, the alerts menu is populated with the error details with the option to re-run the flow, or download the flow to troubleshoot.

**Note:** Flow owners, Server or Site Administrators can see this menu.
Tableau Prep Conductor process status

- The Tableau Services Manager (TSM) status page is accessible in TSM and can be viewed by TSM administrators. You must be able to log into TSM to see this page.
- The Tableau Server status page appears in the Tableau Server web UI and is accessible by Tableau Server administrators. If you hover your mouse pointer over the status indicator for a process, a tooltip shows the node name and the port the process is running on. The Tableau Server status page does not show TSM processes.

When Tableau Server is functioning properly, Tableau Prep Conductor will show as Active or Busy:

- **Active**—The process is functioning as intended.
- **Busy**—The process is completing some task.
- **Down**—The process is down. The implications of this differ depending on the process.
- **Status unavailable**—Tableau Server is unable to determine the status of the process.

Monitor flow activity and performance with Administrative Views

Administrative views can be used to monitor the activities related to flows, performance history, and the disk space used. The **Status** page contains an embedded Tableau workbook with
various administrative views that can be used to monitor different types of server or site activity. For **Tableau Server** installations, you must install PostgreSQL drivers before you can see Administrative views. For more information, see [Database Drivers (Linux)](#), [Database Drivers (Windows)](#). Server administrators can use these views to see activity both at the server level (aggregated for all sites) or for a specific site. Only server administrators can filter by site.

**Who can do this?**

- **Tableau Server**: Tableau Server administrators and Tableau Site administrators can both view and work with Administrative Views. Only Server administrators can filter by site.

**Action by all users**

Use this view to gather insight into how flows are being used. This includes actions like publish, download, and flow runs. You can filter the view by actions, by site, and by time range. The Total Users count shows the number of users who have performed an action. This value is not affected by any filtering. The Active user count shows the number of users who have been active during the selected time period and performed one of the selected actions.

![Image](#)

**Action by Specific User**

Use this view to gather insights about how an individual user is working with flows. You can filter the view by user name, the type of action, time range, and by site.
Action by Recent Users

This view shows you which users have been active on Tableau Server over the past 24 hours. This can be useful if you need to do some maintenance activity on the server and want to know how many and which users this will affect, and what they are doing.

The view **Active**, **Recently Active**, and **Idle** users that are currently signed in to Tableau Server.

For this view, an active user is one who took an action in the last 5 minutes, a recently active user is one who last took an action within 30 minutes, and an idle user is one who last took an action more than 30 minutes ago.

Select a user to see only the actions that user performed recently. Hover over an action to see details of the action.
Backgrounder Task Delays

The Background Task Delay view displays the delay for extract refresh tasks, subscription, and flow tasks—that is, the amount of time between when they are scheduled to run and when they actually run. You can use the view to help you identify places you can improve server performance by distributing your task schedules and by optimizing tasks.

Possible reasons for the delays and ways that you might reduce the delays include the following:

- Many tasks are scheduled for the same time. In the example view, tasks that show long delays are clustered at the same time every day, which creates spikes in the wait time. Note that you can set the Timeline filter to a single day to view task delays by hour and identify the hours of the day which have many tasks scheduled at the same time. A solution to this issue can be to distribute the tasks to off-peak hours to reduce load on the server.

- Other server processes are running at the same time and are consuming server resources and slowing down performance. Monitor the CPU and memory usage of server processes to see which processes are consuming the most resources and then adjust the configuration of processes on your server. For more information on monitoring processes, see Collect Data with Windows Performance Monitor.

Background Tasks for Non Extracts

Background Tasks are created to run flows (scheduled and ad hoc). You can use this view to see how many flow tasks succeeded or failed on this site. For details on a task, hover over its icon.
Performance of Flow Runs

Use this view to see the performance history for all the flows on a site. You can filter by Flow Name, the Output Step Name, Flow Owner, Run Type (Scheduled or Ad Hoc), and the time the flow runs were started.

Here are some questions you can answer using this view:

- **What flow tasks are currently scheduled?** – To do this, use the Start Time filter and select the time frame you want to look at. For example, to see flow tasks that are scheduled in the next 3 hours, select Hours -> Next -> and enter 3.

- **What is the duration of flow tasks?** - To answer this, click on a mark in the view and you should see details including the task duration.

**How many flows were run ad hoc, and how many were scheduled runs?** - To answer this, use the Run Type filter and select Ad hoc or Scheduled.

**Note:** This is not functional in this release and will not actually filter the data.

In addition to the questions described above, here are some examples of insights you might be able to gather:
The flows that are running most frequently will have the most marks.

To see how many flows are running at the same time currently, hover over a mark that shows “In Progress” or “Pending and select “Keep Only” to filter all flow runs that are currently running.

To see how many flows are running at the same time during a specific time range, select a range for the Start Time filter. For example, you can choose “Next three hours” to see which flows will be running in the next three hours.

Stats for Space Usage

The Stats for Space Usage view can help you identify which flow outputs are taking up the most disk space on the server. Disk space usage is displayed by user, project, and by the size of flow output and is rounded down to the nearest number.

Use the Min Size filter to control which flow outputs are displayed, based on the amount of space they take up. Use the object type filter for flows.

- **What Users Use the Most Space** – This section shows the users who own flows (when filtered for flows) that are taking up the most space. Click a user name to filter the next two graphs for that user.

- **What Projects Use the Most Space** – This section shows the projects with flows (when filtered for flows) that are using the most space.

- **What Workbooks, Data Source and Flows Use the Most Space** – This section shows the flows (when filtered for flows) that take up the most space.
Developer Resources - REST APIs

Use Tableau Server REST APIs to automate and seamlessly integrate Tableau into your existing workflows. Tableau Server REST API gives you programmatic access to work with your content, users, sites, and now flows. Manage provisioning, permissions, and publishing on Tableau Server via HTTP. The REST API gives you access to the functionality behind the data sources, projects, workbooks, site users, sites, and flows. You can use this access to create custom applications or to script interactions with server resources.

Tableau REST API supports the following flow functionality:

New flow endpoints have been added to support publishing flows, scheduling flows, running flows on demand, managing permissions, downloading flows and such. For a full list of all the new REST API endpoints for flows, see Flow Methods.

In addition, existing endpoints have been updated to support flow functionality such as Creating New Schedules for Flows, Creating new sites, updating existing sites, and managing default permissions.
Day in the Life Scenarios

What does it mean to shape data? How does that impact what visualizations can be built and what analysis can be performed? In the tutorials below, we explore scenarios for analysis and visualization, identify the data limitations holding us back, then see how Tableau Prep can help us shape the data to reach our intended outcome.

Download the data sets and follow along with these day in the life scenarios using Tableau Prep and Tableau Desktop. Learn how to apply the features and functions in Tableau Prep to get your data ready for analysis in Tableau Desktop.

Give us your feedback. We are just starting to build this section of the online help. If there are specific scenarios you’d love to see here, please let us know. Use the feedback bar at the top of the page to tell us more.

To complete the tasks in these tutorials, you need Tableau Prep and Tableau Desktop installed, and you’ll need to download and save the data to your computer.

For information about how to install Tableau Prep and Tableau Desktop, see Install Tableau Prep and Install Tableau Desktop in the Tableau Desktop and Tableau Prep Deployment guide. Otherwise you can download the Tableau Prep and Tableau Desktop free trials.

Hospital Bed Use with Tableau Prep

Reaching capacity in a hospital is problematic but so is an overabundance of resources. It’s important to understand hospital beds from the perspective of the bed as a resource. However, the data is often stored from the perspective of a patient. How can we take data that captures when patients are in beds and determine the bed usage?

Note: To complete the tasks in these tutorials, you need Tableau Prep and optionally Tableau Desktop installed:

To install Tableau Prep and Tableau Desktop see Install Tableau Prep and Install Tableau Desktop in the Tableau Desktop and Tableau Prep Deployment guide. Otherwise you can download the Tableau Prep and Tableau Desktop free trials.

You will also need to download three data files. It is recommended to save them in your
My Tableau Prep Repository > Datasources folder.
- Beds.xlsx
- Hours.xlsx
- Patient Beds.xlsx

The Data

For our four beds, A, B, C, and D, we track which patient was in the bed and their start and end time there. The data looks like this:

<table>
<thead>
<tr>
<th></th>
<th>Hospital Bed</th>
<th>Patient</th>
<th>Start Time</th>
<th>End Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>Person 1</td>
<td>1/1/2018 8:34</td>
<td>1/1/2018 9:34</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>Person 5</td>
<td>1/1/2018 9:55</td>
<td>1/1/2018 10:15</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>Person 9</td>
<td>1/1/2018 10:34</td>
<td>1/1/2018 16:34</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>Person 8</td>
<td>1/1/2018 17:00</td>
<td>1/1/2018 23:00</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>Person 2</td>
<td>1/1/2018 8:45</td>
<td>1/1/2018 13:45</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>Person 6</td>
<td>1/1/2018 15:13</td>
<td>1/1/2018 18:27</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>Person 7</td>
<td>1/1/2018 18:41</td>
<td>1/1/2018 21:56</td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>Person 10</td>
<td>1/1/2018 22:13</td>
<td>1/1/2018 23:43</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>Person 3</td>
<td>1/1/2018 9:05</td>
<td>1/1/2018 9:35</td>
</tr>
<tr>
<td>10</td>
<td>C</td>
<td>Person 4</td>
<td>1/1/2018 9:30</td>
<td></td>
</tr>
</tbody>
</table>

Preliminary Analysis

If we bring this data into Tableau Desktop, we can create a Gantt chart to show when patients are in beds.
This is a useful visual. We can see that there are only small gaps in use for beds A and B, but bed C is very under-used. Bed D's patient has no end time, but we could address that with some calculations. This gives us a visual overview of how the beds are used.

However, what if we wanted to count the hours when a bed was empty? Or compare open bed time before and after a new policy is put in place? There's no easy way to do that with the data as it's currently structured.

**Desired Data Structure**

By creating some very basic data sets and combining them in Tableau Prep, we can modify this data set into a form that will allow us to perform deeper analysis and create even more useful visualizations.

Before we jump into Tableau Prep, let's step back and think about what we need to create to answer the question, "How many hours was each bed empty?"

We need to be able to look at each bed for each hour, and know whether or not there was a patient in the bed. Right now, the data is solely when a patient was in the bed; we haven't given Tableau information about the empty hours.

To create that full matrix of all beds and all hours, we'll create two new data sets. One is simply a list of beds (A, B, C, D) and the other is a list of hours (1, 2, 3, ..., 23, 24). By performing a cross join (joining every row in one data set with every row in the other data set) we'll wind up with every possible combination of beds and hours.

<table>
<thead>
<tr>
<th>The Beds.xlsx data set looks like this:</th>
<th>The Hours.xlsx data set looks like this:</th>
<th>And the cross joined results look like this:</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Beds Data Set" /></td>
<td><img src="image2.png" alt="Hours Data Set" /></td>
<td><img src="image3.png" alt="Cross Joined Results" /></td>
</tr>
</tbody>
</table>

Next, we'll bring in the Patient Beds information, labeling each bed-hour combination as having a specific patient or not. We wind up with a data set that has a row for each bed-hour,
and if a patient was in the bed, their number and start and end times. Null values indicate the bed was unoccupied.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>D</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>A</td>
<td>8</td>
<td>Person 1</td>
<td>1/1/2018 8:34</td>
<td>1/1/2018 9:34</td>
</tr>
<tr>
<td>31</td>
<td>B</td>
<td>8</td>
<td>Person 2</td>
<td>1/1/2018 8:45</td>
<td>1/1/2018 13:45</td>
</tr>
<tr>
<td>32</td>
<td>C</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>D</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>A</td>
<td>9</td>
<td>Person 5</td>
<td>1/1/2018 9:55</td>
<td>1/1/2018 10:15</td>
</tr>
<tr>
<td>35</td>
<td>A</td>
<td>9</td>
<td>Person 1</td>
<td>1/1/2018 8:34</td>
<td>1/1/2018 9:34</td>
</tr>
<tr>
<td>36</td>
<td>B</td>
<td>9</td>
<td>Person 2</td>
<td>1/1/2018 8:45</td>
<td>1/1/2018 13:45</td>
</tr>
<tr>
<td>37</td>
<td>C</td>
<td>9</td>
<td>Person 3</td>
<td>1/1/2018 9:05</td>
<td>1/1/2018 9:35</td>
</tr>
<tr>
<td>38</td>
<td>D</td>
<td>9</td>
<td>Person 4</td>
<td>1/1/2018 9:30</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>A</td>
<td>10</td>
<td>Person 9</td>
<td>1/1/2018 10:34</td>
<td>1/1/2018 16:34</td>
</tr>
<tr>
<td>40</td>
<td>A</td>
<td>10</td>
<td>Person 5</td>
<td>1/1/2018 9:55</td>
<td>1/1/2018 10:15</td>
</tr>
<tr>
<td>41</td>
<td>B</td>
<td>10</td>
<td>Person 2</td>
<td>1/1/2018 8:45</td>
<td>1/1/2018 13:45</td>
</tr>
<tr>
<td>42</td>
<td>C</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>D</td>
<td>10</td>
<td>Person 4</td>
<td>1/1/2018 9:30</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>A</td>
<td>11</td>
<td>Person 9</td>
<td>1/1/2018 10:34</td>
<td>1/1/2018 16:34</td>
</tr>
</tbody>
</table>

With the data in this structure, we can perform analyses like this, which enables us to investigate unoccupied beds as easily as patient beds.
Restructuring the Data

So how do we get there with Tableau Prep? We'll build out the flow in two parts, first building the Bed Hours matrix, then combining it with the Patient Beds data. Make sure you've downloaded all three Excel files (Beds.xlsx, Hours.xlsx, and Patient Beds.xlsx) to follow along.

Bed Hour Matrix

First, we'll connect to the Beds.xlsx file.

1. Open Tableau Prep.
2. From the start screen, click Connect to Data.
3. On the Connections pane, click Microsoft Excel. Navigate to where you saved Beds.xlsx and click Open.
4. The Beds sheet should automatically be brought out to the Flow pane.

Tip: For more information about connecting to data, see Connect to Data on page 109.
Next, we need to create a field we can use to do the cross join with the **Hours** data set. We’ll add a calculation that is simply the value 1.

5. In the **Flow** pane, select **Beds**, click the plus icon, and select **Add Step**.

6. With the **Clean** step we just added, the **Profile** pane will come up. Click **Create Calculated Field** in the toolbar.

7. Name the field **Cross Join** and enter the value 1.

8. The **Data** grid should update to show the current state of the data.

<table>
<thead>
<tr>
<th>Cross Join</th>
<th>Bed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>1</td>
<td>D</td>
</tr>
</tbody>
</table>

Now we'll repeat the process with the Hours data set.

**Click for directions**

9. On the **Connections** pane, click the **Add connection** button to add another data connection.

10. Choose **Microsoft Excel** and then select the **Hours.xlsx** file and click **Open**.

11. In the **Flow** pane, select **Hours**, click the plus icon, and select **Add Step**.

12. From the toolbar in the **Profile** pane, create a calculated field named **Cross Join** and enter the value 1.
Both data sets now have a shared field, **Cross Join**, and can be joined.

13. Join the two cleaning steps by dragging **Clean 2** onto **Clean 1** and dropping it on the **Join** option.

14. In the **Join Profile** below, the join configurations should populate automatically.

   - Because we named both fields **Cross Join**, Tableau Prep automatically identifies them as the shared field and creates the appropriate **Applied Join Clauses**.
   - The default **Join Type** is inner, which is what we want.
   - This join will match all rows from **Beds** with all rows from **Hours**, as seen in the **Data** grid.
We no longer need the Cross Join fields, so we can remove them.

15. In the Flow pane, select Join 1, click the plus icon, and select Add Step.
16. Select the fields Cross Join-1 and Cross Join, and click Remove Fields.
17. Double click on the Clean 3 label and rename that step Bed Hour Matrix.

We now have the Bed Hour Matrix data set that contains all beds and all hours and have finished the first part of building our data set.

**Patient Bed Use**

Part two is bringing in the patient bed usage. To start, we'll connect to the data.

1. On the Connections pane, click the Add connection button to add another data connection.
2. Choose Microsoft Excel and then select the Patient Beds.xlsx file, and click Open.

3. In the Flow pane, select Patient Beds, click the plus icon and select Add Step.

Because the Bed Hour Matrix file is based on hour but Patient Beds is based on actual time, we need to pull the hour out of the Patient Beds start and end times. Additionally, for the end time, we want to ensure that if a patient is still in the bed at the end of the day (midnight, hour 24) we indicate that the bed is occupied even though there's no end time in the data set. We'll add a calculated field in this new step.

4. In the toolbar, click Create Calculated Field.

5. Name the field Start Hour. For the calculation, enter DATEPART('hour',[Start Time]).

   This takes the hour of the start time and pulls it out. Therefore, "1/1/18 9:35 AM" becomes simply "9".

6. Create another calculated field named End Hour. For the calculation, enter IFNULL(DATEPART('hour',[End Time]), 24).

   The DATEPART portion takes the hour of the end time. The IFNULL portion will assign an end time of 24 (midnight) to any missing end time.

Now we're ready to join patient bed usage to the Bed Hour Matrix. This is a bit more complex join than we did previously. An inner join would only return values present in both data sets. Because we want to make sure we keep all the bed-hour slots, regardless of whether or not a patient was in the bed, we'll need to do a left join. This will result in a lot of nulls, but that's appropriate.

We also need to match when a bed-hour slot is taken by a patient (or patients). In addition to matching the bed the patient is in, we need to consider the time. The Bed Hour Matrix data set just has an Hour field, and the Patient Beds data set has Start Hour and End Hour. We'll use some basic logic to determine if a patient should be assigned to a given bed-hour slot: A patient is considered in a bed if their start hour is less than or equal to (<=) the bed-hour slot AND their end hour is greater than or equal to (>=) the bed-hour slot.

Therefore, three join clauses are needed to appropriately match these two data sets together.

9. Join the Clean 3 step with the Bed Hour Matrix step.

10. In the Applied Join Clauses area, the default should be Hour = End Hour. Click the
join clause to change the operator from "=" to "<=".

11. Click the plus button in the upper right corner of the Applied Join Clauses area to add another join clause. Set it to be `Hour >= Start Hour`.

12. Add a third join clause for `Bed = Hospital Bed`.

13. In the Join Type section, click the unshaded area of the graphic next to Bed Hour Matrix to change the join type to a Left join.
**Note:** If you drag the **Bed Hour Matrix** to **Clean 3** instead of the other way around, the desired results can be obtained by using a right join instead of a left join. The order of dragging the steps matters for the orientation of the join. The join clauses will also be in reverse order—be sure to preserve the correct logic of comparing the hours.

Our data is now joined, but we should clean up some artifacts from the join and make sure the fields are tidy. We no longer need **Start Hour** and **End Hour**. **Hospital Bed** and **Bed** are also redundant. Finally, a value of null in the **Patient** field really means the bed is unoccupied.

14. In the **Flow** pane, add a cleaning step so we can tidy up the joined data.

15. Ctrl+click (Command+click on Mac) to multi select the fields **End Hour**, **Start Hour**, and **Hospital Bed**, then click **Remove Fields** in the toolbar.

16. On the **Patient** field profile card, double click the **null** value and type **Unoccupied**.

We now have a data structure with a row for every bed-hour; if there was a patient in bed during that hour, we have the patient information as well. All that remains to do is add an output step and generate the data set itself.
17. In the Flow pane, select Clean 4, click the plus icon, and select Add Output.

18. In the Output pane, change the Output type to .csv then click Browse.

19. Enter Bed Hour Patient Matrix for the name and choose the desired location before clicking Accept to save.

20. Click the Run Flow button at the bottom of the pane to generate your output. Click Done in the status dialog to close the dialog.

Tip: For more information about outputs and running a flow, see Save and Share Your Work on page 213.

The final flow should look like this:

Analysis in Tableau Desktop

To install Tableau Desktop before continuing with this tutorial, you can download the free trial.

Now that we have the data set in the desired structure, we can perform deeper analysis than with the original data.

1. Open Tableau Desktop. In the Connect pane, select Text file, navigate to the Bed Hour Patient Matrix.csv file, and click Open.

2. On the Data source tab, the data should appear in the canvas by default. Click to Sheet 1.

3. In the Data pane, drag Hour from Measures to Dimensions to make it a discrete dimension.

4. Drag Bed to the Rows shelf and Hour to the Columns shelf.

5. Drag Patient to the Color shelf.
Formatting is optional, but may help make the visual more readable.

6. Click on the **Color** shelf and select **Edit Colors**.

7. In the area to the left, select **Unoccupied**. From the drop down on the right, choose the **Seattle Grays** color palette.

8. Select the fourth, lightest gray, and click **OK**.

9. Click the **Color** shelf again, then click the **Border** dropdown. Choose the second gray option at the far right.

10. In the toolbar, from the Size dropdown, change from **Standard** to **Fit Width**.

11. Click the **Format** menu and then **Borders**.

12. For **Row Divider**, click the Pane dropdown and choose a very light gray.

13. Adjust the **Level** slider to the second tick mark.

14. Repeat with the **Column Divider**. Set the **Pane** color to be light gray and the **Level** to the second tick mark.
15. Double click the sheet tab at the bottom and rename it **Bed Use by Hour**.

This view lets us quickly see when a given bed was occupied or open.

But we can go further and count the number of hours each bed was unoccupied.

16. Click the new sheet tab icon at the bottom to open a clean sheet.

17. Drag **Patient** to **Rows**.
18. Drag **Hour** to **Columns**. Right click to open the menu. Choose **Measure > Count**.

19. Drag another copy of the **Patient** field from the **Data** pane to the **Color** shelf.

20. Right click on the axis and select **Edit Axis**. Change the title to **Hours** and close the dialog.

21. Rename the sheet tab **Bed Hours by Patient**.

This view lets us identify how many unoccupied bed hours we had, something we couldn't do with the original data set. What other charts or dashboards can you create? Give it a try now that your data is in the right structure.

**Recap and Resources**

To build this data structure using Tableau Prep, we needed to perform the following actions:

1. Build a data set for each aspect we want to analyze, in this case, **Beds** and **Hours**.

2. Cross join those data sets to create a **Bed Hour Matrix** data set with every possible combination of beds and hours.

3. Join the **Bed Hour Matrix** with the **Patient Bed** data, making sure the join keeps all bed-slot hours and the join clauses appropriately match patient bed data with the bed-hour slots.

We used the following calculations to create fields we could join on. The second and third pull out the hour information from the original datetime fields.

- **Cross Join** = 1
  - This simply assigns the value 1 to every row

- **Start Hour** = `DATEPART('hour', [Start Time])`
Finding the Second Date with Tableau Prep

A common need in analytics is to determine the date a second event happens, such as when a customer made a second purchase—thereby becoming a repeat customer—or when a driver gets a second traffic violation. Finding the date of a first event is easy, it's simply the minimum date. Finding the second date is trickier.

In this two-part tutorial, we'll shape traffic infraction data and answer the following questions:

1. What was the length of time in days between the first and second infraction for each driver?
2. Compare the fine amounts for the first and second infractions. Are they correlated?
3. Which driver paid the most overall? Who paid the least?
4. How many drivers had multiple infraction types?

5. What was the average fine amount for drivers who never attended traffic school?

In the first stage, we'll use Tableau Prep to restructure the data for our analysis. In the second stage, Analysis with the Second Date in Tableau Desktop on page 331, we'll move on to analysis in Tableau Desktop.

The goal of this tutorial is to present various concepts in the context of a real-life scenario and work through options—not prescriptively establishing which is best. At the end, you should have a better sense of how data structure impacts calculations and analysis, as well as greater familiarity with various aspects of Tableau Prep and calculations in Tableau Desktop.

**Note:** To complete the tasks in this tutorial, you need Tableau Prep and optionally Tableau Desktop installed and the data downloaded.

To install Tableau Prep and Tableau Desktop before continuing with this tutorial, see Install Tableau Prep and Install Tableau Desktop in the Tableau Desktop and Tableau Prep Deployment guide. Otherwise you can download the Tableau Prep and Tableau Desktop free trials.

The data set is Traffic Violations.xlsx. It is recommended to save it in your My Tableau Prep Repository > Datasources folder.

### The Data

For this example, we're looking at traffic infraction data. Each infraction is a row. The driver, date, type of infraction, if the driver was required to attend traffic school, and fine amount are recorded.
Desired Data Structure

The data is currently structured such that each *infraction* is a row. A driver with multiple infractions appears on multiple rows, and there's no easy way to tell which was their first or second infraction.

To investigate our repeat offenders, we want a data set that separates out the first and second infraction dates, and the information associated with each of those infractions, and each row is a *driver*.
Restructuring the Data

So how do we get there with Tableau Prep? We'll build out the flow in stages, beginning with pulling out the first infraction date, then the second, then shaping the final data set as desired. Make sure you've downloaded the Excel file (Traffic Violations.xlsx) to follow along.

Initial Aggregation for 1st Infraction Date

First, we'll connect to the Traffic Violations.xlsx file.

1. Open Tableau Prep.
2. From the start screen, click Connect to Data.
3. In the Connections pane, click Microsoft Excel. Navigate to where you saved Traffic Violations.xlsx and click Open.
4. The Infractions sheet should automatically be brought out to the Flow Pane.

Tip: For more information about connecting to data, see Connect to Data on page 109.

Next, we need to identify the first infraction date per driver. We'll use an Aggregate step to do this, creating a mini data set of Driver ID and Minimum Infraction Date.

Note: When using an Aggregate step in Tableau Prep, any field that should define what makes a row is a Grouped Field. (For us, that's Driver ID.) Any field that will be aggregated and presented at the level of the grouped fields is an Aggregated Field. (For us, that's Infraction Date).

5. In the Flow pane, select Infractions, click the plus icon, and select Add Aggregate.
6. Drag Driver ID to the Grouped Fields drop area.
7. Drag Infraction Date to the Aggregated Fields area. The default aggregation is CNT (count). Click CNT and change the aggregation to Minimum.
This identifies the smallest (earliest) date, which is the first infraction date per driver.

**Tip:** For more information about aggregations, see *Aggregate and group values* on page 199.

8. In the **Flow** pane, select **Aggregate 1**, click the plus icon, and select **Add Step** so we can clean up the output of the aggregation.

9. In the **Profile** pane, double-click on the field name **Infraction Date** and change it to **1st Infraction Date**.

*At this stage, the flow and profile should look like this:*
From the Profile pane in this Clean step, we can see that our data now consists of 39 rows and only 2 fields. Any field not used for grouping or aggregation is lost. But we want to be able to keep some of the original information. We could either add those fields to the grouping or aggregation (but doing so would change the level of detail or require the fields to be aggregated), or join this mini data set back to the original (essentially adding a new column to the original data for **1st Infraction Date**). Let's do the join.

10. In the Flow pane, select **Infractions**, click the plus icon, and select **Add Branch**.

This branch has all the original data. We'll join the results of the aggregation to this copy of the full data. By joining on **Driver ID**, we're adding the minimum date from our aggregation into the original data.
11. Select step **Clean 2** and drag it on top of step **Clean 1**, and drop it on **Join**.

12. The default join configuration should be correct: an inner join on **Driver ID = Driver ID**.

![Image of Join configuration]

**Tip:** For more information about joins, see **Join your data** on page 201.

Because some fields may be duplicated during a join, such as the fields in the join clause, it’s often a good idea to clean up extraneous fields after performing a join.

13. In the **Flow** pane, select **Join 1**, click the plus icon, and choose **Add a Step**.

14. In the **Profile** pane, click the card for **Driver ID-1**, then click **Remove Field** in the toolbar.

15. To change the field order, drag the **1st Infraction Date** card between **Driver ID** and **Infraction Date** where you see the black line appear.

*At this stage, the flow should look like this:*
Looking at the data grid below, we can see our new, combined data set. We have the minimum—that is, first—infracation date for each driver added to each row in the data set.

### Second Aggregation for 2nd Infracation Date

We need to also determine the second infraction date. To do this, we want to filter out any row where the infraction date is equal to the minimum—thus removing the first date. We can then take the minimum of the remaining dates using another aggregate step, leaving us with the second infraction date, which we'll rename for clarity.

**Note:** Because we'll want to use the data as it currently is in **Clean 3** later on in the flow, we'll add another **Clean** step to break out the process of getting the second infraction date.

16. In the Flow pane, select **Clean 3**, click the plus icon, and select **Add Step**.
17. On the toolbar in the Profile pane, choose **Filter Values**. Create a filter [Infraction Date] != [1st Infraction Date].
18. Remove the field **1st Infraction Date**.

19. In the **Flow** pane, select **Clean 4**, click the plus icon, and select **Add Aggregate**.

20. Drag **Driver ID** to the **Grouped Fields** drop area. Drag **Infraction Date** to the **Aggregated Fields** area and change the aggregation to **Minimum**.

21. In the **Flow** pane, select **Aggregate 2**, click the plus icon, and select **Add Step**. Rename **Infraction Date** to **2nd Infraction Date**.

*At this stage, the flow should look like this:*

![Flow Diagram](image)

We now have our second infraction date identified for each driver. To get all the other information associated with each infraction (type, fine, traffic school) we again need to join this back to the entire data set.

22. Select **Clean 5** and drag it on top of **Clean 3**, dropping it on **Join**.

23. Again, the default join configuration should be correct: an inner join on **Driver ID** = **Driver ID**.

24. In the **Flow** pane, select **Join 2**, click the plus icon, and select **Add Step**. Delete the fields **Driver ID-1** and **1st Infraction Date** as they are no longer needed.

*At this stage, the flow should look like this:*

![Flow Diagram](image)

Create full data sets for the 1st and 2nd infractions

Before we go any further, let's step back and think about everything we have and how we want to bring it all together. Our desired end state is a data set that looks like this, with a column for
Driver ID, then columns for date, type, traffic school, and fine amount for the 1st and 2nd infractions.

How do we get there from here?

In the step Clean 3, we have our complete data set with a column that repeats the first infraction date for each driver.

We want to eliminate all the rows for a driver that aren't the first infraction, building a data set of only first infractions. That is, we only want to keep the information for a given driver when 1st Infraction Date = Infraction Date. Once we’ve filtered to keep only the row of the first infraction, we can remove the Infraction Date field and tidy up field names.

Similarly, after the second aggregation and join, we have our complete data set with a column for the second infraction date.
We can perform a similar filter of 2nd Infraction Date = Infraction Date to keep only the row of information for each driver's 2nd infraction. Again, we can also remove the now-redundant Infraction Date and tidy up field names.

We'll start with the first infraction data set.

25. In the Flow pane, select Clean 3, click the plus icon, and select Add Branch.

26. With this new Clean step selected, in the Profile pane, click Filter Values in the toolbar. Create a filter [1st Infraction Date] = [Infraction Date].

27. Remove the field Infraction Date.

28. Rename the Infraction Type, Traffic School, and Fine Amount fields to start with "1st".

29. Double-click on the name Clean 7 under the step in the Flow pane and rename it Robust 1st.

Now for the second infraction data set.

30. In the Flow pane, select Clean 6, after the last join.

31. Click Filter Values in the toolbar. Create a filter [2nd Infraction Date] = [Infraction Date].

32. Remove the field Infraction Date.

33. Rename the Infraction Type, Traffic School, and Fine Amount fields to start with "2nd".

34. Double-click on the name Clean 6 under the step in the Flow pane and rename it Robust 2nd.

At this stage, the flow should look like this:
Create the complete data set

Now that we have these two tidy data sets with complete information for the first and second infractions per driver, we can join them back together on **Driver ID** and wind up with our desired data structure.

35. Select **Robust 2nd** and drag it on top of **Robust 1st**, dropping it on **Join**.

36. The default join clause should be correct as **Driver ID = Driver ID**.

37. Because we don't want to drop drivers who didn't have a second infraction, we need to make this a left join. In the **Join Type** area, click the unshaded area of the diagram next to **Robust 1st**, turning it into a **Left** join.

38. In the **Flow** pane, select **Join 3**, click the plus icon, and select **Add Step**. Remove the field duplicate **Driver ID-1**.

The data is in the desired state, so we can create an output and proceed to analysis.

39. In the **Flow** pane, select the newly added **Clean 6**, click the plus icon, and select **Add Output**.

40. In the **Output** pane, change the **Output type** to .csv then click **Browse**. Enter **Driver Infractions** for the name and choose the desired location before clicking **Accept** to save.

41. Click the **Run Flow** button at the bottom of the pane to generate your output. Click **Done** in the status dialog to close the dialog.

**Tip:** For more information about outputs and running a flow, see **Save and Share Your Work** on page 213.

*The final flow should look like this:*
Recap

For the first stage of this tutorial, our goal was to take our original data set and prepare it for analysis involving the first and second infraction dates. The process consists of three phases:

Identify the first and second infraction dates:

1. Create an aggregation that keeps Driver ID and MIN Infraction Date. Join this with the original data set to create an "intermediate data set" that has the first (minimum) infraction date repeated for every row.

2. On a new step, filter out all rows where the 1st Infraction Date is the same as the Infraction Date. From that filtered data set, create an aggregation that keeps Driver ID and MIN Infraction date. Join this with the intermediate data set from the first step. This identifies the second infraction date.

Build out clean data sets for the first and second infractions:

3. Go back and create a branch from the intermediate data set and filter to keep only rows where the 1st Infraction Date is the same as the Infraction Date. This builds a data set for just the first infraction. Tidy it up by removing any unnecessary fields and rename all the desired fields (except Driver ID) to indicate they’re for the first infraction. This is the Robust 1st data set.

4. Tidy the data set for the second infraction date. Clean the join results from step 2 by filtering to keep only rows where the 2nd Infraction Date is the same as the Infraction Date. Remove any unnecessary fields and rename all the desired fields (except Driver ID) to indicate they’re for the second infraction. This is the Robust 2nd data set.

Combine the first and second infraction data into one data set:
5. Join the **Robust 1st** and **Robust 2nd** data sets, making sure to keep all records from **Robust 1st** to prevent losing any drivers without a second infraction.

Next, we want to explore how this data can be analyzed in Tableau Desktop.

**Continue to Analysis with the Second Date in Tableau Desktop below.**

**Note:** Special Thanks to Ann Jackson’s Workout Wednesday topic **Do Customers Spend More on Their First or Second Purchase?** and Andy Kriebel's Tableau Prep Tip **Returning the First and Second Purchase Dates** that provided the initial inspiration for this tutorial. Clicking these links will take you away from the Tableau website. Tableau cannot take responsibility for the accuracy or freshness of pages maintained by external providers. Contact the owners if you have questions regarding their content.

**Analysis with the Second Date in Tableau Desktop**

This is the second stage of the tutorial and assumes the first stage, **Finding the Second Date with Tableau Prep** on page 318, has been completed.

In the first stage, we took our original data set and shaped it to answer the following questions:

1. What was the length of time in days between the first and second infraction for each driver?
2. Compare the fine amounts for the first and second infractions. Are they correlated?
3. Which driver paid the most overall? Who paid the least?
4. How many drivers had multiple infraction types?
5. What was the average fine amount for drivers who never attended traffic school?

As we now explore these questions, it becomes clear that there are some pros and cons to the first data structure we created. We'll go back into Tableau Prep and do some additional reshaping, then see how that impacts the same analysis in Tableau Desktop. Finally, we'll look at a Tableau Desktop-only approach to the analysis using Level of Detail (LOD) expressions with the original data.
The goal of this tutorial is to present various concepts in the context of a real-life scenario and work through options—not prescriptively establishing which is best. At the end, you should have a better sense of how data structure impacts calculations and analysis, as well as greater familiarity with various aspects of Tableau Prep and calculations in Tableau Desktop.

**Note:** To complete the tasks in this tutorial, you need Tableau Prep and optionally Tableau Desktop installed and the data downloaded.

To install Tableau Prep and Tableau Desktop before continuing with this tutorial, see Install Tableau Prep and Install Tableau Desktop in the Tableau Desktop and Tableau Prep Deployment guide. Otherwise you can download the Tableau Prep and Tableau Desktop free trials.

The data set is the output from Driver Infractions.tflx, as built in the first stage.

### Analysis in Tableau Desktop

Now that we have our data configured, we'll bring it into Tableau Desktop. We can easily answer some questions, but others involve a few (or a lot of) calculations. Try your hand at the questions below; you can expand each one for basic information about how to proceed if you get stuck.

**Note:** You can download the workbook Driver Infractions.twbx to look at the solutions in context. Remember that there may be alternative ways to interpret the analysis or pursue answers.

1. What was the length of time in days between the first and second infraction for each driver?

   A. To answer this question in Tableau Desktop, we'll use the DATEDIFF function. This function takes three arguments—the date part, the start date, and the end date. Since we want to know the days between these events, we'll use the date part 'day'. Our start and end dates are in the data set as 1st Infraction Date and 2nd Infraction Date.

   B. The calculation is:
**Time Between Infractions** = DATEDIFF('day', [1st Infraction Date], [2nd Infraction Date])

C. We can plot that against **Driver ID** as a bar chart. Note that seven drivers had no second infraction, so there are seven nulls.

2. Compare the fine amounts for the first and second infractions. Are they correlated?

A. To answer this question in Tableau Desktop, we'll create a scatter plot of **1st Fine Amount** and **2nd Fine Amount**. By bringing **Driver ID** to the **Detail** shelf on the **Marks** card, we can create a mark for each driver.
B. To add a trend line, use the **Analytics** tab in the left-hand pane and bring out a linear trend line. Hovering over the trend line, we can see the R-squared value is practically zero, and the p-value is far above any cutoff for significance. We can determine that there is no correlation between first and second fine amount.

3. Which driver paid the most overall? Who paid the least?

When we want to go deeper in our analysis, we may need to create some calculations.

A. To answer this in Tableau Desktop, we need to add the fines for both infractions into a single field. Because some drivers may not have had a second infraction, we need to use the zero null ZN function to turn any nulls for **2nd Fine Amount** into zeros.

B. The calculation is:
Total Amount Paid = [1st Fine Amount] + ZN([2nd Fine Amount])

C. We can plot Total Amount Paid against Driver ID and sort the bar chart.

4. How many drivers had multiple infraction types?

A. To answer this in Tableau Desktop, we need to do a fancier IF calculation, comparing if the first and second infraction types are the same. If they are, we want to assign the value "1". If they are not the same, we'll assign "2". Any other result, such as a null second infraction type, will be assigned "1".

B. The calculation is:

\[
\text{Number of Infraction Types} = \text{IF First Infraction Type = Second Infraction Type, 1, 2}
\]
IF [1st Infraction Type] = [2nd Infraction Type] THEN 1
ELSEIF [1st Infraction Type] != [2nd Infraction Type] THEN 2
ELSE 1 END

C. We can then plot **Number of Infraction Types** against **Driver ID** and sort the bar chart.

5. What was the average fine amount for drivers who never attended traffic school?

A. To answer this in Tableau Desktop, we cannot simply divide the total fine amount by two, since some drivers only had one infraction. We also can't calculate the average fine per
driver and take the average of those values, because averaging averages can lead to inconsistencies. Instead, we need to calculate the total amount paid by drivers who never attended traffic school, then divide by the total number of infractions associated with those fines.

1. First, we need to determine if each driver had a second infraction. We can leverage the fact the information in all the "2nd" fields will be null if there was no second infraction and start building the calculation:

\[
\text{IFNULL}([2\text{nd Infraction Type}], 'no')
\]

This will return an infraction type if it exists, or "no" if there was no second infraction.

2. Next, we need to turn this information into the number of infractions, 1 or 2. If the result of our \text{IFNULL} calculation is "no", then the driver should be marked as having one fine. Any other result should be marked as having two fines. The calculation is:

\[
\text{Number of Infractions} = \begin{cases} 
1 & \text{IF IFNULL}([2\text{nd Infraction Type}], 'no') = 'no' \\
2 & \text{ELSE}
\end{cases}
\]

3. Now we need to consider the total fine amount. Similarly to question 3 above, we'll add the first and second fine amounts, with a \text{ZN} function around the second. However, because we want this to be computed at the level of the entire data set, it's a best practice to specify the aggregations, \text{SUM}, in the calculation itself. The calculation is:

\[
\text{SUM}([1\text{st Fine Amount}]) + \text{SUM}(\text{ZN}([2\text{nd Fine Amount}]))
\]

4. To bring it all together, we'll take this total fine amount and divide it by our new \text{Number of Infractions} calculated field to determine the average fine amount:

\[
\text{Average Fine} = \frac{\left( \text{SUM}([1\text{st Fine Amount}]) + \text{SUM}(\text{ZN}([2\text{nd Fine Amount}])) \right)}{\text{SUM}([\text{Number of Infractions}])}
\]

B. We also need to filter out drivers who ever attended traffic school—but that information is also stored across two fields.
1. Tableau is very efficient at numerical calculations. We'll phrase this with numbers to help performance as much as we can. To combine these two fields, we'll create a calculation for each one that says "Yes = 1" and "No = 0" (null should also = 0, for drivers with no second infraction). By summing the outcome of these calculations, any driver with an overall value of 0 never went to traffic school (and a value of 1 or 2 represents how many times they went). We can then filter to keep only drivers with a value of 0.

2. This time, we'll use a CASE statement instead of IF. They function very similarly but have different syntax. The start of the calculation should look like this:

   ```
   CASE [1st Traffic School] WHEN 'Yes' THEN 1
   WHEN 'No' THEN 0
   ELSE 0
   END
   ```

3. And then we'll do the same thing for 2nd Traffic School. We can add both pieces in the same calculation by wrapping each case statement in parentheses and adding a plus between them. Removing some of the line breaks, it looks like this:

   ```
   Number of Traffic School Attendances =

   (CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No'
   THEN 0 ELSE 0 END)
   +
   (CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No'
   THEN 0 ELSE 0 END)
   ```

4. If we drag Number of Traffic School Attendances to the Dimensions area of the Data pane, the values 0–2 will become discrete.

5. Now if we filter on Number of Traffic School Attendances, we can select just the 0 and know we're getting drivers who have never attended traffic school.

C. To answer the original question, we'll simply bring Average Fine to the Text shelf on the Marks card. Because we built the aggregations into the calculation, the aggregation on the field will be AGG and we cannot change it. This is as expected.
Go Further—Pivoted Data

While the data we've been working with is well structured to address questions specifically around first and second infractions, it isn't the standard structure recommended for use with Tableau Desktop. The more our analysis diverges from basic questions around the infraction dates, the more complicated our calculations become to combine the relevant information into useable form.

Usually, when data is stored with multiple columns for the same type of data (such as two columns for date, two columns for fine amount, etc.) and unique information is stored in the field name (such as whether it's the first or second infraction), this is an indication the data should be pivoted.

Performing a multiple pivot can handle this nicely. Back in Tableau Prep, and working from the end of the Driver Infraction Tableau Prep flow created in the step-by-step tutorial above:

1. From the final clean step, add a Pivot step that pivots by every duplicated field.

   **Tip:** Use the plus icon in the upper right corner of the Pivoted Fields area to add more Pivot Values. Each set of fields (such as 1st and 2nd Fine
Amounts) should be pivoted together.
For more information about pivoting, see Pivot your data on page 162.

The results can be tidied by removing null dates as well as renaming and reordering fields.

2. Add a cleaning step after the pivot. In the Infraction Date column, right-click on the null bar and choose Exclude.

3. Double-click the field name Pivot1 Names and rename it Infraction Number.

4. Drag fields as appropriate to reorder them as below:

   ![Pivot Table Example]

5. From the new, pivoted data, create an output named Pivoted Driver Infractions and bring it into Tableau Desktop. (Don’t forget to run the flow after adding the Output step.)

Now we can look at our five questions again with this pivoted data structure; you can expand each one for basic information about how to proceed if you get stuck.

**Note:** You can download the completed flow file Pivoted Driver Infractions.tflx to check your work, or download the workbook Pivoted Driver Infractions.twbx to look at the solutions in context. Remember that there may be alternative ways to interpret the analysis or pursue answers.

1. What was the length of time in days between the first and
second infraction for each driver?

A. To answer this in Tableau Desktop, as we did with the first data set, we'll use the DATEDIFF function. This function requires a start date and an end date. This information is present in our data, but all in one field. We need to pull it out into two fields.

1. Create two preliminary calculated fields:

   1st Infraction = IF [Infraction Number] = "1st" THEN [Infraction Date] END

   2nd Infraction = IF [Infraction Number] = "2nd" THEN [Infraction Date] END

2. Because we want to make sure both of these values are available to be compared for each driver, we need to fix them to the level of Driver ID.

   **Note:** Don't believe me? Try to do a DATEDIFF calculation with these two fields as they are: Time Between Infractions = DATEDIFF('day', [1st Infraction], [2nd Infraction])

   You'll get null results everywhere, because Tableau is trying to compare across a data structure that looks like this:

<table>
<thead>
<tr>
<th>Driver ID</th>
<th>1st Infraction Date</th>
<th>2nd Infraction Date</th>
<th>Time between infractions</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA-106451404</td>
<td>Null</td>
<td>9/28/2017</td>
<td>Null</td>
</tr>
<tr>
<td></td>
<td>7/5/2017</td>
<td>Null</td>
<td>Null</td>
</tr>
<tr>
<td>AF-108351406</td>
<td>Null</td>
<td>9/2/2018</td>
<td>Null</td>
</tr>
<tr>
<td></td>
<td>5/9/2018</td>
<td>Null</td>
<td>Null</td>
</tr>
<tr>
<td>AJ-107951404</td>
<td>Null</td>
<td>12/31/2017</td>
<td>Null</td>
</tr>
<tr>
<td></td>
<td>10/15/2017</td>
<td>Null</td>
<td>Null</td>
</tr>
</tbody>
</table>

   Here, the row that knows what the first date is doesn’t know what the second date is, and vice versa. To get around this, we'll use a FIXED Level of Detail expression to force these first and second dates to be related by Driver ID.

   Edit each calculation as follows:

   1st Infraction = { FIXED [Driver ID] : MIN ( IF [Infraction Number] = "1st" THEN [Infraction Date] END ) }
2nd Infraction = \{ \text{FIXED [Driver ID] : MIN ( IF [Infraction Number] = "2nd" THEN [Infraction Date] END ) } \}

\textbf{Note:} The original calculation must be aggregated when embedded in an LOD expression. We can use any basic aggregation that will preserve the date value (so aggregations like SUM, AVG, or MIN work, but not CNT or CNTD).

3. Now we can create the \texttt{DATEDIFF} calculation as follows:

\textbf{Time Between Infractions} = \texttt{DATEDIFF('day', [1st Infraction], [2nd Infraction])}

4. If we want to look at weeks or months, simply modify the date part (currently 'day').

The results will be identical to the outcome with the unpivoted data structure.

2. \textbf{Compare the fine amounts for the first and second infractions. Are they correlated?}

A. To answer this in Tableau Desktop, we'll use very similar logic to the previous question. We'll use \textbf{Infraction Number} to identify if a given row is the first or second infraction, then pull out the fine amount accordingly.

1. If all we want to do is make a scatter plot, we can skip the LOD portion and just use the \texttt{IF} calculation:

\textbf{1st Fine Amount} = \texttt{IF [Infraction Number] = "1st" THEN [Fine Amount] END}

\textbf{2nd Fine Amount} = \texttt{IF [Infraction Number] = "2nd" THEN [Fine Amount] END}

2. However, if we wanted to compare and see the difference in amount between the first and second fines for a single driver, we'd run into the same null issue as with the dates. It can't hurt to wrap these calculations in a \texttt{FIXED} LOD, so it might be good just to do so from the start:
Tableau Prep Help

1st Fine Amount = { FIXED [Driver ID] : MIN ( IF [Infraction Number] = "1st" THEN [Fine Amount] END ) }

2nd Fine Amount = { FIXED [Driver ID] : MIN ( IF [Infraction Number] = "2nd" THEN [Fine Amount] END ) }

3. Create a scatterplot and bring out a linear trend line as before.

The results will be identical to the outcome with the unpivoted data structure.

3. Which driver paid the most overall? Who paid the least?

A. To answer this question in Tableau Desktop, the pivoted data structure is ideal. All we need to do is bring out Driver ID and Fine Amount into a bar chart. The default aggregation is already SUM, so the total amount paid by the driver will automatically be plotted.

The results will be identical to the outcome with the unpivoted data structure.

4. How many drivers had multiple infraction types?

A. To answer this question in Tableau Desktop, the pivoted data structure is ideal. All we need to do is bring out Driver ID and a Count Distinct of Infraction Type as a bar chart, and we'll have our answer.

The results will be identical to the outcome with the unpivoted data structure.

5. What was the average fine amount for drivers who never attended traffic school?

A. To answer this in Tableau Desktop, we cannot simply divide the total fine amount by two, since some drivers only had one infraction. We also can't calculate the average fine per driver and take the average of those values, because averaging averages can lead to inconsistencies. Instead, we need to calculate the total amount paid by drivers who never attended traffic school, then divide by the total number of infractions associated with those fines.
1. First, we need to determine if each driver had a second infraction. We can leverage the fact **2nd Infraction Date** will be null if there was no second infraction and start building the calculation:

$$\text{IFNULL}(\text{STR}([2nd ~\text{Infraction}]), ~'no')$$

This will return the date of the second infraction if it exists, or "no" if there was no second infraction.

**Note:** The **STR** portion of this calculation is necessary because **IFNULL** needs consistency of data type in its arguments. Because we want to return the string "no" for null values, we need to convert the date to a string as well.

2. Next, we need to turn this information into the number of infractions, 1 or 2. If the result of our **IFNULL** calculation is "no", then the driver should be marked as having one fine. Any other result should be marked as having two fines. The calculation is:

$$\text{Number of Infractions} = \begin{cases} 
1 & \text{IF \text{IFNULL}(\text{STR}([2nd ~\text{Infraction}]), ~'no') = 'no'} \\
2 & \text{ELSE} \\
\text{END} 
\end{cases}$$

3. Now we need to consider the average fine amount. We already have a single field for **Fine Amount**. All we need to do is divide that by our new **Number of Infractions** field, wrapping both in **SUM**:

$$\text{Average Fine} = \frac{\text{SUM}([\text{Fine Amount}])}{\text{SUM}([\text{Number of Infractions}])}$$

B. We also need to filter out drivers who attended traffic school. It looks like we could use the **Traffic School** field and filter on **Traffic School = no**. However, this would filter on *infractions* not associated with traffic school, not *drivers* who never went to traffic school. If a driver went to traffic school for one infraction but not a second, we don’t want the second infraction to be considered here.

What we want to do is filter out any driver who’s attended traffic school. In terms of the data, we want to filter out any driver who has a "Yes" for **Traffic School** on any row.
Let’s build our calculation in stages, using a simple view to help keep track of what's happening:

1. First, we want to know if a driver has a "Yes" for Traffic School. Drag Driver ID to Rows and Traffic School to Columns. We’ll get a text table with placeholder "Abc" text indicating the relevant values for each driver.

2. Next, we want to build a calculation that will identify if the value of Traffic School is "Yes". The first stage of the calculation is:

   \[ \text{Attended Traffic School} = \text{CONTAINS}([\text{Traffic School}], \text{'Yes'}) \]

   If we bring Attended Traffic School to the Color shelf on the Marks card, we see it accurately labels "False" for every mark in the "No" column, and "True" for every mark in the "Yes" column.

3. However, what we really want is this information at the level of the driver, not the infraction. An LOD expression is a natural fit when trying to compute a result at a different level of detail than the basic structure of the data. We’ll make this a FIXED LOD expression. But, as we know, the aggregate expression portion of an
LOD must be aggregated. Previously, we've used **MIN**. Will that work here? We'll modify the calculation to be:

**Attended Traffic School** = { FIXED [Driver ID] : MIN ( CONTAINS ([Traffic School], 'Yes')) }

With that change applied in the view, we see the opposite of what we want. Any driver that has a "No" is marked as "False" across the board. Instead, we want to carry the "Yes" as a "True" for every record for that driver. What is MIN doing here? It's picking the first response alphabetically, that is, "No".

4. What if we changed it to **MAX**? Would that take the last response alphabetically? We'll modify the calculation to be:

**Attended Traffic School** = { FIXED [Driver ID] : MAX ( CONTAINS ([Traffic School], 'Yes')) }
And here we have it: if a driver has "Yes" anywhere in the data, they are marked as "True" for having attended traffic school, even on the infraction that didn't involve traffic school.

5. If we bring **Attended Traffic School** to the **Filter** shelf and select only "False", we'll be left with only drivers who never attended traffic school.

C. To answer the original question, with our filter in place we'll simply bring **Average Fine** to the **Text** shelf on the **Marks** card. Because we built the aggregations into the calculation, the aggregation on the field will be **AGG** and we cannot change it. This is as expected.

The results will be identical to the outcome with the unpivoted data structure.

**The benefits of pivoted data**

We could stick with the original data structure from the tutorial if we know we'd only need to answer questions that are easy to answer with that structure. However, the pivoted data format is more flexible. Even though it requires some calculations, once they're in place the resulting data set is well suited to answer broader questions.
Go Further Still—Calculations Only

What if you don't have access to Tableau Prep? Are you out of luck entirely if you're stuck with the original data? Not at all!

Tableau Desktop and LOD expressions can answer all of our analytical questions. If we connect to the original Traffic Violations.xlsx, it looks very similar to the pivoted data set—just without the crucial Infraction Number field. We'll need to mimic the outcome of the aggregation steps via LOD expressions.

Note: You can download the workbook LOD Driver Infractions.twbx to look at the solutions in context. Remember that there may be alternative ways to interpret the analysis or pursue answers.

1. What was the length of time in days between the first and second infraction for each driver?

   A. To answer this in Tableau Desktop, we'll again use the DATEDIFF function. This function requires a start date and an end date. This information is present in our data, but all in one field. We need to pull it out into two fields. Because we want to make sure both of these values are available to be compared for each driver, we need to fix them to the level of Driver ID.

      1. To find the first infraction date, we use the calculation:

      \[ 1st \text{ Infraction} = \{ \text{FIXED [Driver ID]} : \text{MIN ( [Infraction Date] )} \} \]

      2. We'll do the second infraction date in stages.

         a. To start, we need to look at just the dates that are larger than the first date:

         \[ \text{IF [Infraction Date]} > [1st \text{ Infraction}] \ \text{THEN} \ [\text{Infraction Date}] \ \text{END} \]

         b. But this will give us every infraction after the first, and we only want the second. So we want the smallest of these dates. Wrap the whole thing in MIN:
MIN( IF [Infraction] : [1st Infraction] THEN [Infraction Date] END )

c. We also want to recalculate the second infraction date for each driver. That’s where LOD expressions come in. We’ll fix this to the level of Driver ID:

\[
2\text{nd Infraction} = \{ \text{FIXED [Driver ID]} : \text{MIN ( IF [Infraction Date] > [1st Infraction] THEN [Infraction Date] END )} \}
\]

3. And we can now create the \texttt{DATEDIFF} calculation:

\[
\text{Time Between Infractions} = \text{DATEDIFF('day', [1st Infraction], [2nd Infraction])}
\]

The results will be identical to the outcomes with the other two data structures.

2. Compare the fine amounts for the first and second infractions. Are they correlated?

A. To answer this in Tableau Desktop, we’ll use similar logic to the pivoted data version of this question. We’ll use the 1\text{st Infraction} and 2\text{nd Infraction} fields we created for question I to identify if a given row is the first or second infraction, then pull out the fine amount accordingly.

1. If all we want to do is make a scatter plot, we can skip the LOD portion and just use an IF calculation:

\[
1\text{st Fine Amount} = \text{IF [1st Infraction] = [Infraction Date] THEN [Fine Amount] END}
\]

\[
2\text{nd Fine Amount} = \text{IF [2nd Infraction] = [Infraction Date] THEN [Fine Amount] END}
\]

2. However, if we want to compare and see the difference in amount between the first and second fines for a single driver, we’d run into issues with nulls, as in the first data structure. It can’t hurt to wrap these calculations in a FIXED LOD, so it might be good just to do so from the start:
1st Fine Amount = \{ \text{FIXED [Driver ID]} : \text{MIN ( IF [1st Infraction] = [Infraction Date] THEN [Fine Amount] END )} \}

2nd Fine Amount = \{ \text{FIXED [Driver ID]} : \text{MIN ( IF [2nd Infraction] = [Infraction Date] THEN [Fine Amount] END )} \}

The results will be identical to the outcomes with the other two data structures.

3. Which driver paid the most overall? Who paid the least?

A. To answer this in Tableau Desktop, we need to first realize something about the LOD-only method. Both methods using Tableau Prep filter out records that are not the first or second infraction for a driver. The LOD method in Tableau Desktop keeps all records. This means that if we were to create a viz of \text{SUM(Amount Paid)} by \text{Driver ID}, the Tableau Desktop-only version will show higher amounts for drivers with more than two infractions. To get a \text{Total Amount Paid} value from the complete data that matches the other methods, instead of using the original \text{Fine Amount} field, we instead need to sum the first and second fines like we did with the first data structure.

B. Using the fields we created for question 2, we’ll add the two fine amounts. \text{ZN} is necessary to prevent a null result for any drivers who only had one infraction. The calculation is:

\text{Total Amount Paid} = [1st \text{ Fine Amount}] + \text{ZN}([2nd \text{ Fine Amount}])

The results will be identical to the outcomes with the other two data structures.

4. How many drivers had multiple infraction types?

A. To answer this question in Tableau Desktop, we can’t simply bring out \text{Driver ID} and a \text{Count Distinct of Infraction Type}. Because this data set has infractions beyond the second, some drivers may have more than two infraction types. To match the results with the other methods, we need to limit the scope to just the first two infractions.
We can to pull out the 1st and 2nd infraction types, wrap them in LOD expressions to make them \texttt{FIXED} to the driver, then use an \texttt{IF} calculation to count the types:

1. **1st Infraction Type** = \{ \texttt{FIXED [Driver ID]} : \texttt{MIN ( IF [1st Infraction] = [Infraction Date] THEN [Infraction Type] END )} \} 

2. **2nd Infraction Type** = \{ \texttt{FIXED [Driver ID]} : \texttt{MIN ( IF [2nd Infraction] = [Infraction Date] THEN [Infraction Type] END )} \} 

3. **Number of Infraction Types** =

   \[
   \text{IF [1st Infraction Type] = [2nd Infraction Type]} \text{ THEN 1}
   \]

   ELSEIF [1st Infraction Type] \text{ != [2nd Infraction Type]} \text{ THEN 2}

   ELSE 1 END

**Note:** It’s also possible to create many of these calculations as a single field by nesting the initial calculations directly in the larger calculation. Here, the combined calculation would look like this:

\[
\text{IF} \\
\{\text{FIXED [Driver ID]} : \text{MIN (IF [1st Infraction] = [Infraction Date] THEN [Infraction Type] END))}
\]

\[
= \\
\{\text{FIXED [Driver ID]} : \text{MIN (IF [2nd Infraction] = [Infraction Date] THEN [Infraction Type] END))}
\]

\[
\text{THEN 1}
\]

ELSEIF

\[
\{\text{FIXED [Driver ID]} : \text{MIN (IF [1st Infraction] = [Infraction Date] THEN [Infraction Type] END))}
\]

\[
!= \\
\{\text{FIXED [Driver ID]} : \text{MIN (IF [2nd Infraction] = [Infraction Date] THEN [Infraction Type] END))}
\]

\[
\text{THEN 2}
\]
Which is a bit harder to make sense of, but works if preferred. (Note that line breaks and some spaces do not impact how a calculation is interpreted by Tableau.)

C. We can then plot **Number of Infraction Types** against **Driver ID** and sort the bar chart.

The results will be identical to the outcomes with the other two data structures.

5. What was the average fine amount for drivers who never attended traffic school?

A. To answer this in Tableau Desktop, we cannot simply divide the total fine amount by two, since some drivers only had one infraction. We also can’t calculate the average fine per driver and take the average of those values, because averaging averages can lead to inconsistencies. Instead, we need to calculate the total amount paid by drivers who never attended traffic school, then divide by the total number of infractions associated with those fines.

1. First, we need to determine if each driver had a second infraction. We can leverage the fact the information in all the "2nd" fields will be null if there was no second infraction and start building the calculation:

   ```plaintext
   IFNULL([2nd Infraction Type], 'no')
   ```

   This will return an infraction type if it exists, or "no" if there was no second infraction.

2. Next, we need to turn this information into the number of infractions, 1 or 2. If the result of our IFNULL calculation is "no", then the driver should be marked as having one fine. Any other result should be marked as having two fines. The calculation is:

   ```plaintext
   Number of Infractions =
   ```
IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1 ELSE 2 END

3. For the Total Amount Paid, we can use the calculation from question 3. To bring it all together, we'll take this total fine amount and divide it by our new **Number of Infractions** calculated field to determine the average fine amount:

   \[
   \text{Average Fine} = \frac{\text{SUM}([\text{Total Amount Paid}])}{\text{SUM}([\text{Number of Infractions}])}
   \]

B. We also need to filter out drivers who attended traffic school. Because this data set contains some drivers with a third or fourth infraction, we can't use the same method as the pivoted data structure. Instead, we'll follow the same method as the unpivoted data, summarized here:

1. First, we need to built two calculations identifying if the first and second infractions involved traffic school or not:

   **1st Traffic School** = \{ FIXED [Driver ID] : MIN (IF [1st Infraction] = [Infraction Date] THEN [Traffic School] END ) \}

   **2nd Traffic School** = \{ FIXED [Driver ID] : MIN (IF [2nd Infraction] = [Infraction Date] THEN [Traffic School] END ) \}

2. Then we'll add those values to get the overall number of traffic school attendances:

   \[
   \text{Number of Traffic School Attendances} = \\
   (\text{CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END}) \\
   + \\
   (\text{CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END})
   \]

3. If we drag **Number of Traffic School Attendances** to the **Dimensions** area of the **Data** pane, the values 0–2 become discrete.

4. Now if we filter on **Number of Traffic School Attendances**, we can select just the 0 and know we're getting drivers who have never attended traffic school.
C. To answer the original question, we'll simply bring **Average Fine** to the **Text** shelf on the **Marks** card. Because we built the aggregations into the calculation, the aggregation on the field will be **AGG** and we cannot change it. This is as expected.

The results will be identical to the outcomes with the other two data structures.

It's important to remember that this solution has a lot of nested calculations and LOD expressions. Depending on the size of the data set and the complexity of the data, performance could be an issue.

**Reflection on Methods**

So which route should you go? That's entirely up to you and the tools at your disposal.

- If you want to steer clear of LODs, there's a data-shaping solution, though calculations might be necessary for some analysis (**Analysis in Tableau Desktop** on page 332).
- If you can shape the data and are comfortable with calculations—including LODs—the middle-of-the-road option provides the best flexibility (**Go Further—Pivoted Data** on page 339).
- If you're comfortable with LODs, there's minimal impact on performance, and/or you don't have access to Tableau Prep, solving this with LODs alone is a viable option (**Go Further Still—Calculations Only** on page 348).

At the very least, it's valuable to understand how aggregation in Tableau Prep and Level of Detail expressions in Tableau Desktop are interrelated and impact data analysis. As with most things in Tableau, there's more than one way to do anything. Exploring all the various options can help bring concepts together and let you pick the best solution for you.

**Calculations used:**

**Driver Infractions**

- **Time Between Infractions** = `DATEDIFF('day', [1st Infraction Date], [2nd Infraction Date])`
- **Total Amount Paid** = `[1st Fine Amount] + ZN([2nd Fine Amount])`
- **Number of Infraction Types** = `IF [1st Infraction Type]=[2nd Infraction Type] THEN 1 ELSE 0`
Infraction Type] THEN 1 ELSEIF [1st Infraction Type]!=[2nd Infraction Type] THEN 2 ELSE 1 END

- **Number of Infractions** = IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1 ELSE 2 END

- **Average Fine** = \( \frac{\text{SUM}([1st\ Fine\ Amount]) + \text{SUM}([2nd\ Fine\ Amount])}{\text{SUM}([Number\ of\ Infractions])} \)

- **Number of Traffic School Attendances** = (CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END) + (CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END)

### Pivoted Driver Infractions

- **1st Infraction** = {FIXED [Driver ID] : MIN(IF [Infraction Number] = "1st" THEN [Infraction Date] END)}

- **2nd Infraction** = {FIXED [Driver ID] : MIN(IF [Infraction Number] = "2nd" THEN [Infraction Date] END)}

- **Time Between Infractions** = DATEDIFF('day', [1st Infraction], [2nd Infraction])

- **1st Fine Amount** = {FIXED [Driver ID] : MIN(IF [Infraction Number] = "1st" THEN [Fine Amount] END )}

- **Number of Infractions** = IF IFNULL(STR([2nd Infraction]), 'no') = 'no' THEN 1 ELSE 2 END

- **Average Fine** = \( \frac{\text{SUM}([Fine\ Amount])}{\text{SUM}([Number\ of\ Infractions])} \)

- **Attended Traffic School** = { FIXED [Driver ID] : MAX( CONTAINS ([Traffic School], 'Yes'))}

### LOD Driver Infractions

- **1st Infraction** = {FIXED [Driver ID] : MIN([Infraction Date])}

- **2nd Infraction** = { FIXED [Driver ID] : MIN(IF [Infraction Date] > [1st Infraction] THEN [Infraction Date] END )}

- **Time Between Infractions** = DATEDIFF('day', [1st Infraction], [2nd Infraction])
Infraction])

- **1st Fine Amount** = {FIXED [Driver ID] : MIN( IF [1st Infraction] = [Infraction Date] THEN [Fine Amount] END ) }

- **2nd Fine Amount** = {FIXED [Driver ID] : MIN( IF [2nd Infraction] = [Infraction Date] THEN [Fine Amount] END ) }

- **Total Amount Paid** = [1st Fine Amount] + ZN([2nd Fine Amount])

- **1st Infraction Type** = {FIXED [Driver ID] : MIN( IF [1st Infraction] = [Infraction Date] THEN [Infraction Type] END ) }

- **2nd Infraction Type** = {FIXED [Driver ID] : MIN( IF [2nd Infraction] = [Infraction Date] THEN [Infraction Type] END ) }

- **Number of Infraction Types** = IF [1st Infraction Type] = [2nd Infraction Type] THEN 1 ELSEIF [1st Infraction Type] != [2nd Infraction Type] THEN 2 ELSE 1 END

- **Number of Infractions** = IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1 ELSE 2 END

- **Average Fine** = SUM([Total Amount Paid]) / SUM([Number of Infractions])

- **1st Traffic School** = {FIXED [Driver ID] : MIN (IF [1st Infraction] = [Infraction Date] THEN [Traffic School] END ) }

- **2nd Traffic School** = {FIXED [Driver ID] : MIN (IF [2nd Infraction] = [Infraction Date] THEN [Traffic School] END ) }

- **Number of Traffic School Attendances** = (CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END) + (CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END)

**Note:** Special Thanks to Ann Jackson’s Workout Wednesday topic Do Customers Spend More on Their First or Second Purchase? and Andy Kriebel’s Tableau Prep Tip
Returning the First and Second Purchase Dates that provided the initial inspiration for this tutorial. Clicking these links will take you away from the Tableau website. Tableau cannot take responsibility for the accuracy or freshness of pages maintained by external providers. Contact the owners if you have questions regarding their content.
Troubleshoot Tableau Prep Builder

This article lists problems you might encounter when using Tableau Prep Builder and suggestions for how to resolve them.

Common errors when using the command line to run flows

You can run flows from the command line to refresh your output files programmatically instead of opening Tableau Prep Builder to run each flow manually. While this process helps build efficiency in your flow process, if your syntax is incorrect, or you are missing credentials for your connections or output locations, you will receive errors when running this process.

The following table describes common errors and how to resolve them. For information about how to run flows from the command line, see Refresh output files from the command line on page 218.

<table>
<thead>
<tr>
<th>Error</th>
<th>Cause</th>
<th>How to fix it</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Missing arguments&quot;</td>
<td>One of the required command line arguments is missing.</td>
<td>Use “tableau-prep-cli -help” to see a list of the arguments for the command line.</td>
</tr>
<tr>
<td>&quot;Unable to read the connections file.&quot;</td>
<td>There are errors in the syntax or format in the credentials.json file for the input connections.</td>
<td>Check the syntax for the input connections in the .json file. For more information and examples, see Refresh output files from the command line on page 218.</td>
</tr>
</tbody>
</table>
There are errors in the flow. Unable to run the flow. Check that the credentials.json file includes all required credentials. Open the flow in Tableau Prep Builder to view error details.

<table>
<thead>
<tr>
<th>Error Message</th>
<th>Cause</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;There are errors in the flow. Unable to run the flow. Check that the credentials.json file includes all required credentials. Open the flow in Tableau Prep Builder to view error details.&quot;</td>
<td>There are missing credentials in the credentials.json file for the input connections or the flow has errors.</td>
<td>Check that the .json file has the credentials for all connections, and open the flow file in Tableau Prep Builder to see if there are any errors in the flow. If the flow has errors, you must fix them and republish the flow to Tableau Server, then try running the process again.</td>
</tr>
<tr>
<td>&quot;Could not find match for &lt;hostname of inputConnections&gt;&quot;</td>
<td>The credentials.json file is missing an entry for the hostname (server name).</td>
<td>Make sure the credentials.json file includes the correct credentials for the hostname (server name). For more information and examples, see Refresh output files from the command line on page 218</td>
</tr>
<tr>
<td>&quot;We don't have credentials of all connections in tfl/flx file. The following connection(s) were not found: &lt;hostname of inputConnections&gt;&quot;</td>
<td>The credentials.json file is missing or has incorrect credentials.</td>
<td>Make sure the credentials.json file includes the correct credentials for the hostname (server name).</td>
</tr>
<tr>
<td>Error Message</td>
<td>Credits File Details</td>
<td>Connection Details</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>&quot;Error signing in server &lt;serverUrl&gt; as a user &lt;userName&gt;. Please check the credentials.&quot;</td>
<td>The credentials.json file has the incorrect credentials for Tableau Server.</td>
<td>Make sure the credentials.json file includes all the correct credentials and elements for the output connection. For more information and examples, see Refresh output files from the command line on page 218.</td>
</tr>
<tr>
<td>&quot;Could not sign in successfully as &lt;userName&gt; to server &lt;serverUrl&gt;(&lt;contentUrl&gt;)&quot;</td>
<td>The credentials.json file has the incorrect credentials for Tableau Server.</td>
<td>Make sure the credentials.json file includes all the correct credentials and elements for the output connection. For more information and examples, see Refresh output files from the command line on page 218.</td>
</tr>
<tr>
<td>Error Message</td>
<td>Explanation</td>
<td>Solution</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>&quot;We don't have credentials for Tableau Server to publish extract for one or more output nodes in tfl/tflx file.&quot;</td>
<td>The credentials.json file was not passed in as a command line argument or it is missing the credentials for the output connection.</td>
<td>Make sure the path to the credentials.json file is included in the command line and verify that the credentials.json file includes all the correct credentials and elements for the output connection. For more information and examples, see <a href="#">Refresh output files from the command line on page 218</a>.</td>
</tr>
<tr>
<td>&quot;Loom rest api server not started&quot;</td>
<td>The installation or environment setup is incorrect.</td>
<td>Make sure that Tableau Prep Builder is installed correctly and that you are running the command as an Administrator. For information about how to install Tableau Prep</td>
</tr>
</tbody>
</table>
Builder, see Install Tableau Desktop or Tableau Prep Builder from the User Interface.

"Error. Flow file does not exist."
The path to the flow file is incorrect.
Make sure that the correct path to the flow file is included in the command line.

"Error. Connections file does not exist."
The path to the credentials.json file is incorrect.
Make sure that the correct path to the credentials.json file is included in the command line.

"Could not find match for <mapr01:5181>,<mapr02:5181>,<mapr03:5181>"
You must specify a specific Port ID when connecting to Apache Drill using ZooKeeper.
Include a credentials.json file in the command line that specifies "port": 31010 for the input credentials.

Error: "These features were found that prevent this version of the application from using this file"

If you open a flow that was created in version 2018.2.1 or later in an earlier version of Tableau Prep Builder, you may see the following error:
Flows that include features that are not supported in earlier releases will result in this incompatibility error. To resolve the error, open the flow in the later version, and save a copy of the flow without the indicated features. In the above example, remove the null filter from the field where it is applied.

Then open the copy that has the feature removed in the earlier version of Tableau Prep Builder.

**Error: "Failed to parse response from Tableau Server" when publishing Tableau Prep output**

To successfully publish output from Tableau Prep Builder to Tableau Server, the REST API must be enabled on Tableau Server. If the REST API is not enabled, you will see the following error:

```
Failed to parse response from Tableau server due
to:javax.xml.bind.UnmarshalException - with linked
exception: [org.xml.sax.SAXParseException; lineNumber: 1;
columnNumber: 10; DOCTYPE is disallowed when the feature
"http://apache.org/xml/features/disallow-doctype-decl" set
to true.]
```

For information about enabling the REST API on Tableau Server, see **REST API Requirements** in the REST API Help.

For information about publishing output from Tableau Prep Builder, see **Create and publish data extracts and data sources** on page 214.
Error: "You are using Server version: null..."
when signing in to an SSL-enabled Tableau Server using Tableau Prep

When you sign in to an SSL-enabled Tableau Server from Tableau Prep Builder, you must have a root certificate installed on the computer where Tableau Prep Builder is installed. If the certificate is not installed, you might see the following error:

You are using Server version: null but the minimum compatible version is: 10.0. Please upgrade to a compatible version

If you see this error, work with your IT department or system administrator to install the required root certificate on the computer where Tableau Prep Builder is installed. For more information, see System requirements in the Tableau Desktop and Tableau Prep Builder Deployment Guide.

Maintain Licenses for Tableau Desktop and Tableau Prep

Important: The Tableau licensing service was moved to a new data center on October 6, 2018. This means that any environments that required special configuration (static IP safe listing for example) to access licensing.tableau.com or licensing.tableausoftware.com will need to be updated before you can activate, refresh, or deactivate a Tableau product key. For more information, see Tableau Community.

Tableau Desktop and Tableau Prep can be licensed under a term license model. Term licenses, also known as subscription licenses, allow you to use and update Tableau Desktop and Tableau Prep for a specified period of time.

Term licenses must be renewed to continue providing uninterrupted service. You can continuously renew the term license as each specified period expires. If you don't renew your term license and the term expires, Tableau will stop working and you will no longer have access to the software. For more information about renewing your license, see How to Renew your Tableau Licenses.
**Note**: Trial licenses for Tableau Desktop or Tableau Prep expire after a set period of time, usually 14 days. After the trial period expires, you'll need to purchase a license to continue using the product.

**View data about your license**

After you install Tableau Desktop or Tableau Prep open the application and then navigate to Help > Manage Product Keys from the top menu to see information about the type of license you have and when it expires.

You can also activate or deactivate a product key or refresh a maintenance product key from this dialog.

**Note**: Tableau offers term licenses that provide a range of capabilities. The type of license that you have is displayed in the Product field. For more information about the different type of user-based licenses that are available, see User-based licenses in the Tableau Server help.

Existing Tableau Desktop users may have a perpetual (permanent) license. Perpetual licenses don't expire. However, to get access to product updates and technical support you must purchase Support and Maintenance services. These services must be renewed to continue receiving the service. Perpetual (permanent) licenses are no longer available for Tableau Desktop.

Use the following buttons to take action on your product key:

- **Refresh** (Tableau Desktop only): Click the Refresh button to refresh a maintenance license that is expiring, then close and restart Tableau Desktop. If the Maintenance
Expires date doesn’t update, check with your license administrator as the key or maintenance agreement may have changed.

To refresh a maintenance key from the command line see Refresh the product key in the Tableau Desktop and Tableau Prep Deployment guide.

- **Deactivate**: Select a product key in the list then click Deactivate to deactivate the product key. Deactivate a product key if you need to move the product key to another computer or when you no longer need the product key on this computer.

  For more information about deactivating a product key, see Move or Deactivate Product Keys in the Tableau Desktop and Tableau Prep Deployment guide.

- **Activate**: After Tableau Desktop or Tableau Prep is installed, click Activate to open the activation dialog and enter your product key. If you get an error and can’t activate Tableau Desktop or Tableau Prep using your product key, contact Tableau Support.

  For more information about activating a product key, see Activate Tableau Desktop or Activate and Register Tableau Prep in the Tableau Desktop and Tableau Prep Deployment guide.

### Track Tableau Desktop license usage and expiration data

If you want to track and view license usage and expiration data for Tableau Desktop in Tableau Server you must configure Tableau Desktop to send license data to Tableau Server on a set interval, and then enable reporting on Tableau Server.

This enables server administrators to access two reports:

- **Desktop License Usage**: This report lets server administrators see usage data for Tableau Desktop licenses in your organization.

- **Desktop License Expiration**: This report gives server administrators information about which Tableau Desktop licenses in your organization have expired or need maintenance renewal.

If Tableau Desktop and Tableau Server are configured for license reporting, when signed in to Tableau Server as an Administrator, you will see these two reports listed on the Server Status page in the Analysis section.
If you don’t see these reports listed, then Tableau Desktop and Tableau Server may not be configured for Tableau Desktop usage reporting.

For information about how to configure Tableau Desktop and Tableau Server for usage reporting, see Manage Tableau Desktop License Usage in the Tableau Desktop and Tableau Prep Deployment guide.

Additional resources

For more information about managing your license refer to the following topics:

- To find your product key and activate Tableau Desktop see Install Tableau Desktop.
- To find your product key and activate Tableau Prep, see Install Tableau Prep.
- To deactivate a product key or move it to another computer, see Move or Deactivate Tableau Desktop.
- To learn more about product keys for non-persistent virtual desktops or for computers that are regularly re-imaged, see Configure Virtual Desktop Support.
- To learn more about product key management for Tableau Server or Tableau Online, see Licensing Overview (Linux | Windows)
To learn more about the license renewal process or to renew a license, see How to Renew your Tableau Licenses.