simplilearn | PURDUE UNIVERSITY

Tableau 10

Data Connections

What's In It For Me

- List the methods to prepare data for Tableau
- Describe how to connect data to Tableau
- Explain how the Tableau data engine optimizes performance
- Understand how to create and use data extracts

Data Connections

Topic 1: Preparing Data for Tableau

TABLEAU DESKTOP 10

simplilearn

Preparing Data for Tableau

TABLEAU TOOLS

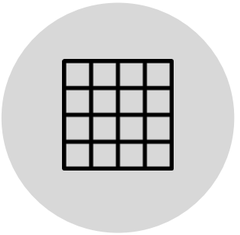
It is important to format data before it can be analyzed using Tableau; this helps to save time and prevent errors.

Tableau offers the following tools to help prep data for analysis:

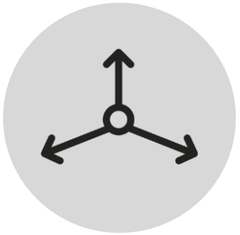
Data Joins (From one or more databases)

Data Blending

Splits

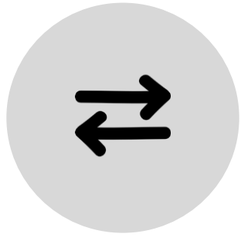


Metadata Grid View



Pivot

Union

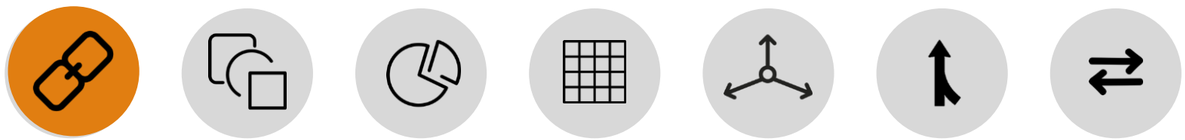


Data Interpreter

Preparing Data for Tableau

DATA JOINS

- A dataset is typically made up of a collection of tables related by specific fields or columns.
- The Joining method is used to combine the related data in those common fields.
- A Join results in a virtual table that is typically extended horizontally by adding columns.



Preparing Data for Tableau

DATA JOINS: EXAMPLE

Shown here is the analysis of data on product sales with two files:

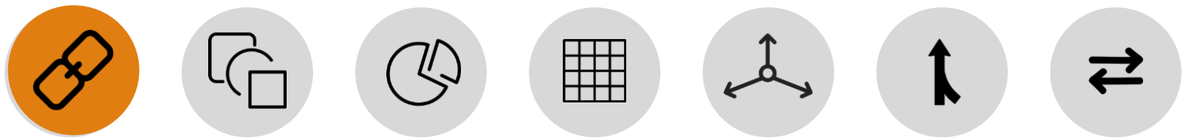
Product Sales: **LEFT** Table

Product Sales		
Product ID	Sales	Number of Records
6	985	2
7	243	15
8	652	17
9	1123	25
10	851	33
11	665	15
12	458	12
13	784	9
14	965	8
15	1033	29

Product Profits: **RIGHT** TABLE

Product Profits		
Product ID	Profits	Number of Records
1	246	15
2	61	12
3	163	9
4	281	8
5	213	29
6	166	22
7	115	6
8	196	41
9	241	13
10	258	21

The product ID field serves as the primary key to join the data from the two sets.



Preparing Data for Tableau

TYPES OF JOINS

- **Inner**
- **Left**
- **Right**
- **Full Outer**

The resulting table includes data that is present in BOTH data sets.

Venn Diagram

Product Sales		
Product ID	Sales	Number of Records
6	985	2
7	243	15
8	652	17
9	1123	25
10	851	33
11	665	15
12	458	12
13	784	9
14	965	8
15	1033	29

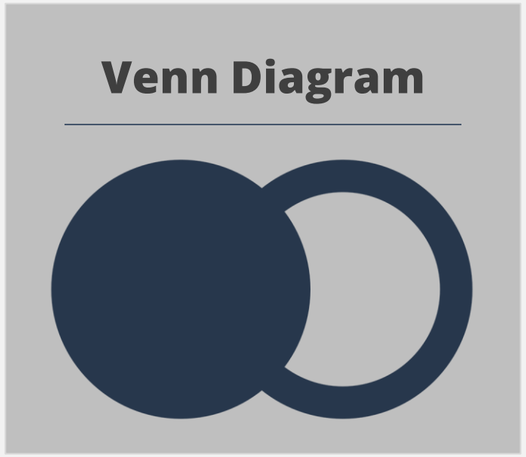
Product Profits		
Product ID	Profits	Number of Records
1	246	15
2	61	12
3	163	9
4	281	8
5	213	29
6	166	22
7	115	6
8	196	41
9	241	13
10	258	21

Preparing Data for Tableau

TYPES OF JOINS

- Inner
- Left
- Right
- Full Outer

The resulting table contains all values from the LEFT table and any matches from the RIGHT table. When a value in the LEFT table doesn't have a corresponding match in the RIGHT table, you see a null value in the data grid.



Product Sales		
Product ID	Sales	Number of Records
6	985	2
7	243	15
8	652	17
9	1123	25
10	851	33
11	665	15
12	458	12
13	784	9
14	965	8
15	1033	29

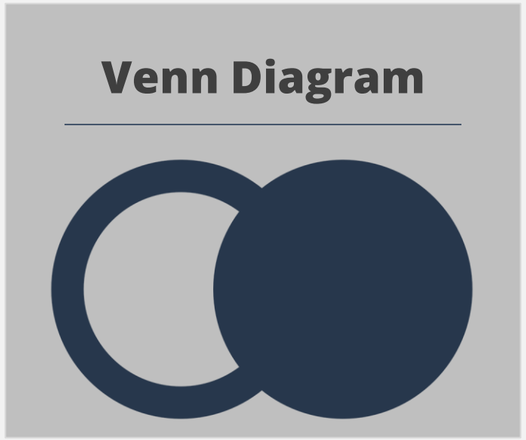
Product Profits		
Product ID	Profits	Number of Records
1	246	15
2	61	12
3	163	9
4	281	8
5	213	29
6	166	22
7	115	6
8	196	41
9	241	13
10	258	21

Preparing Data for Tableau

TYPES OF JOINS

- Inner
- Left
- **Right**
- Full Outer

The resulting table contains all values from the RIGHT table and any matches from the LEFT table. When a value in the RIGHT table doesn't have a corresponding match in the LEFT table, you see a null value in the data grid.



Product Sales		
Product ID	Sales	Number of Records
6	985	2
7	243	15
8	652	17
9	1123	25
10	851	33
11	665	15
12	458	12
13	784	9
14	965	8
15	1033	29

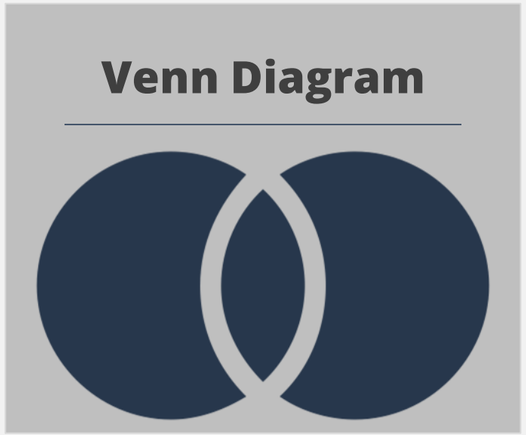
Product Profits		
Product ID	Profits	Number of Records
1	246	15
2	61	12
3	163	9
4	281	8
5	213	29
6	166	22
7	115	6
8	196	41
9	241	13
10	258	21

Preparing Data for Tableau

TYPES OF JOINS

- Inner
- Left
- Right
- **Full Outer**

The resulting table contains all values from BOTH tables. When a value from EITHER table doesn't have a match with the other table, you see a null value in the data grid.



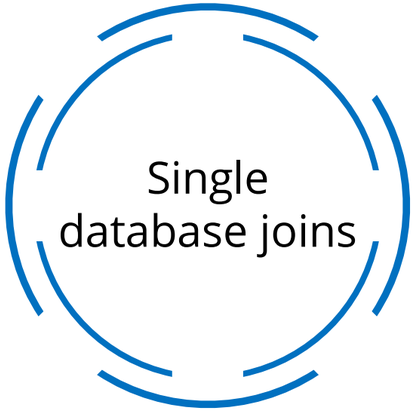
Product Sales		
Product ID	Sales	Number of Records
6	985	2
7	243	15
8	652	17
9	1123	25
10	851	33
11	665	15
12	458	12
13	784	9
14	965	8
15	1033	29

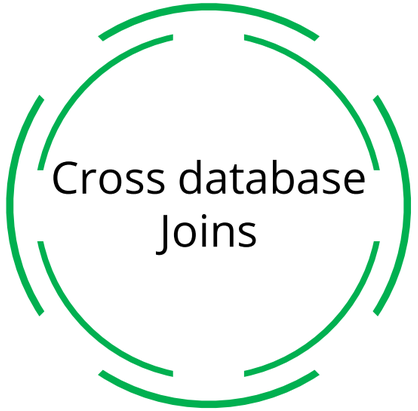
Product Profits		
Product ID	Profits	Number of Records
1	246	15
2	61	12
3	163	9
4	281	8
5	213	29
6	166	22
7	115	6
8	196	41
9	241	13
10	258	21

Preparing Data for Tableau

JOINS FROM DATABASE

Tableau facilitates creating joins in two ways:

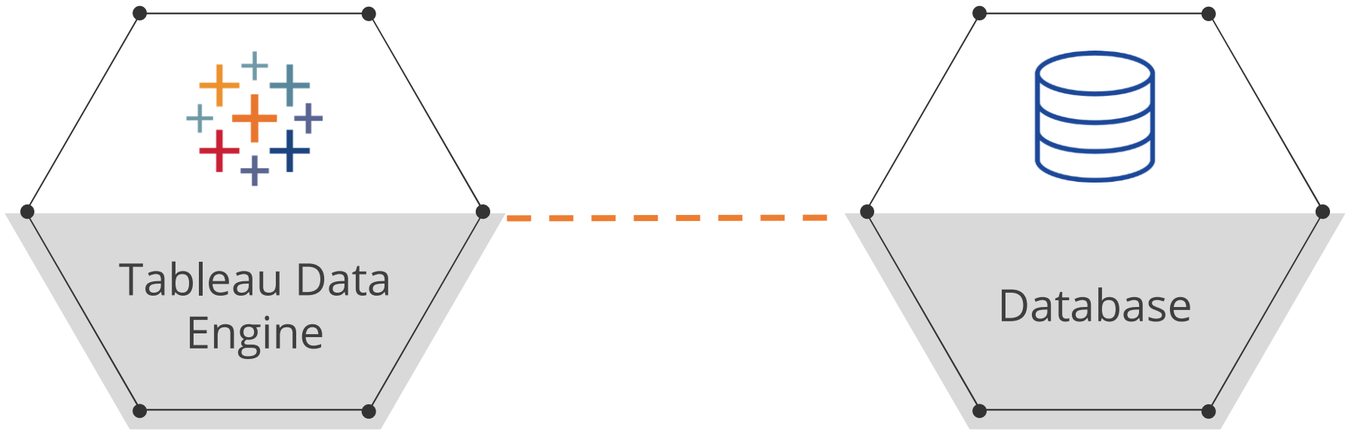




Preparing Data for Tableau

SINGLE DATABASE JOINS

Joining tables from the same database requires only a single connection in the data source.



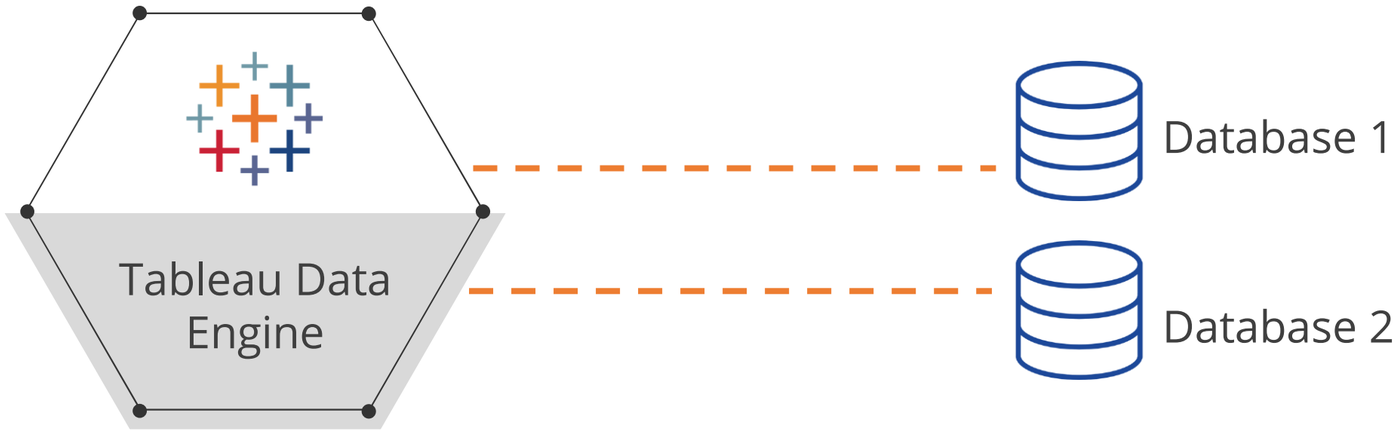
Demo—Perform a Single Database Join

TABLEAU DESKTOP 10

Preparing Data for Tableau

CROSS DATABASE JOINS

- Cross-database Joins require setting up a multi-connection data source by creating a new connection to each database.
- Multi-connection data sources are helpful when different internal systems are used.



Demo—Perform a Cross Database Join

TABLEAU DESKTOP 10

Data Blending

- Blending is a method of combining related data from multiple sources in a single view in order to analyze it.
- There is always one primary data source, while the rest become secondary data sources.

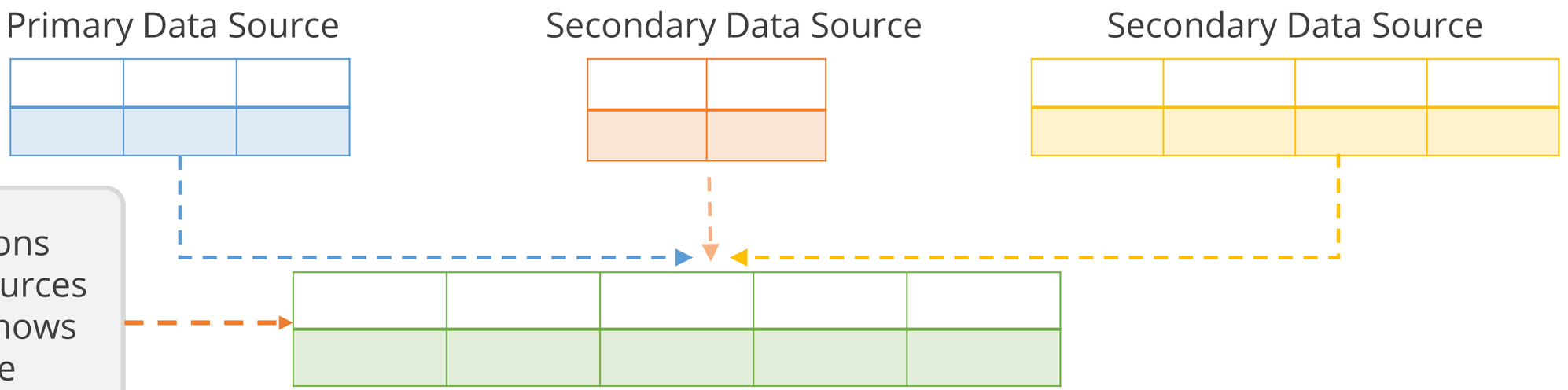


Tableau identifies common dimensions within the data sources so that the user knows which fields can be used.

Blending does not create joins at the row level; instead, it automatically creates an outer join to the secondary source(s).

Data Blending

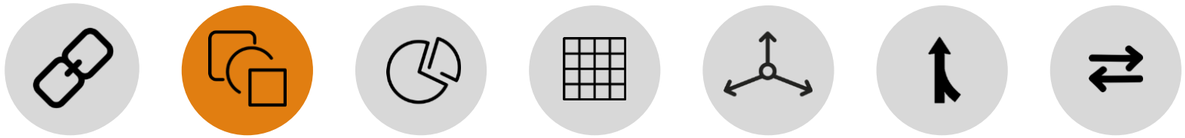
BLEND VS JOIN

When to use a Blend:

- You have to combine data from different databases that do not support cross-table joins.
- Data within the different sources are at different levels of detail.
- Using a Join causes duplicate rows.
- You are working with large amounts of data

When to use a Join:

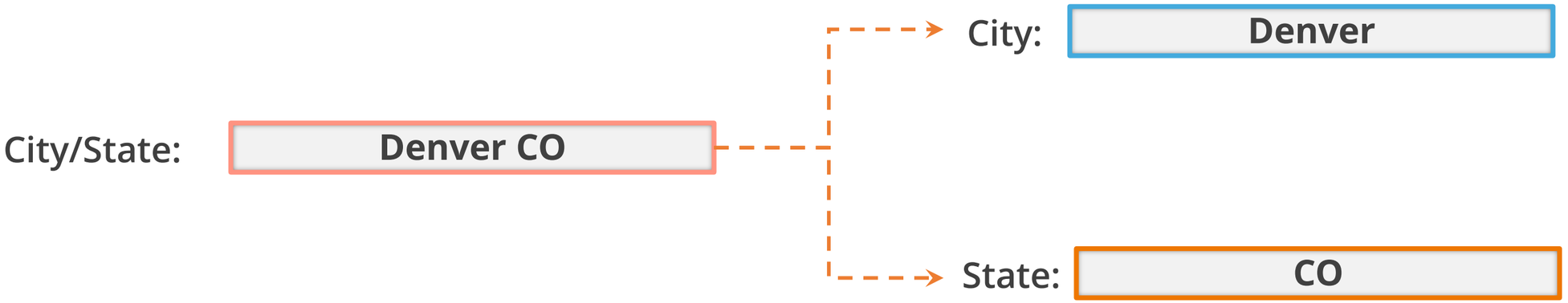
- Your data format is consistent across all sources.
- You are working with relatively small amounts of data.
- Data within each source is at the same level of detail.



Splits

Splitting data from one field into multiple columns is used often in data preparation.

Example:

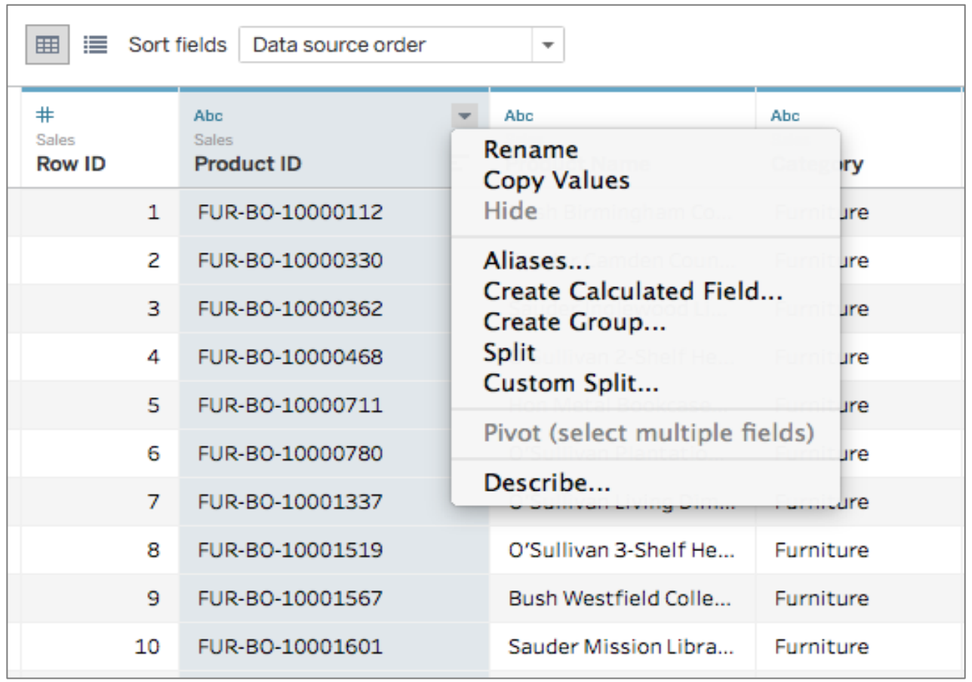


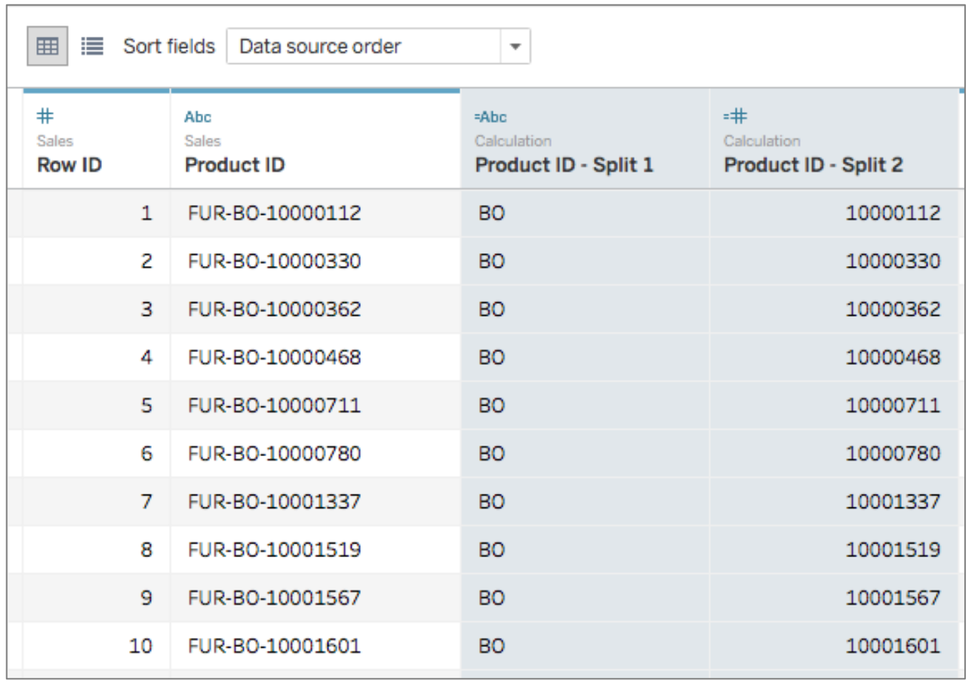
This was typically remedied in Excel with the “Text to Columns” function.

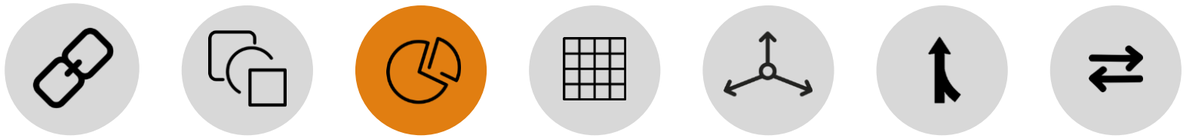
Splits

AUTOMATIC SPLITS

- A string field can be split automatically based on a common separator that Tableau detects (space or underscore).
- This split can be used to automatically separate a field's value into a maximum of ten new fields, depending on the type of data connection.



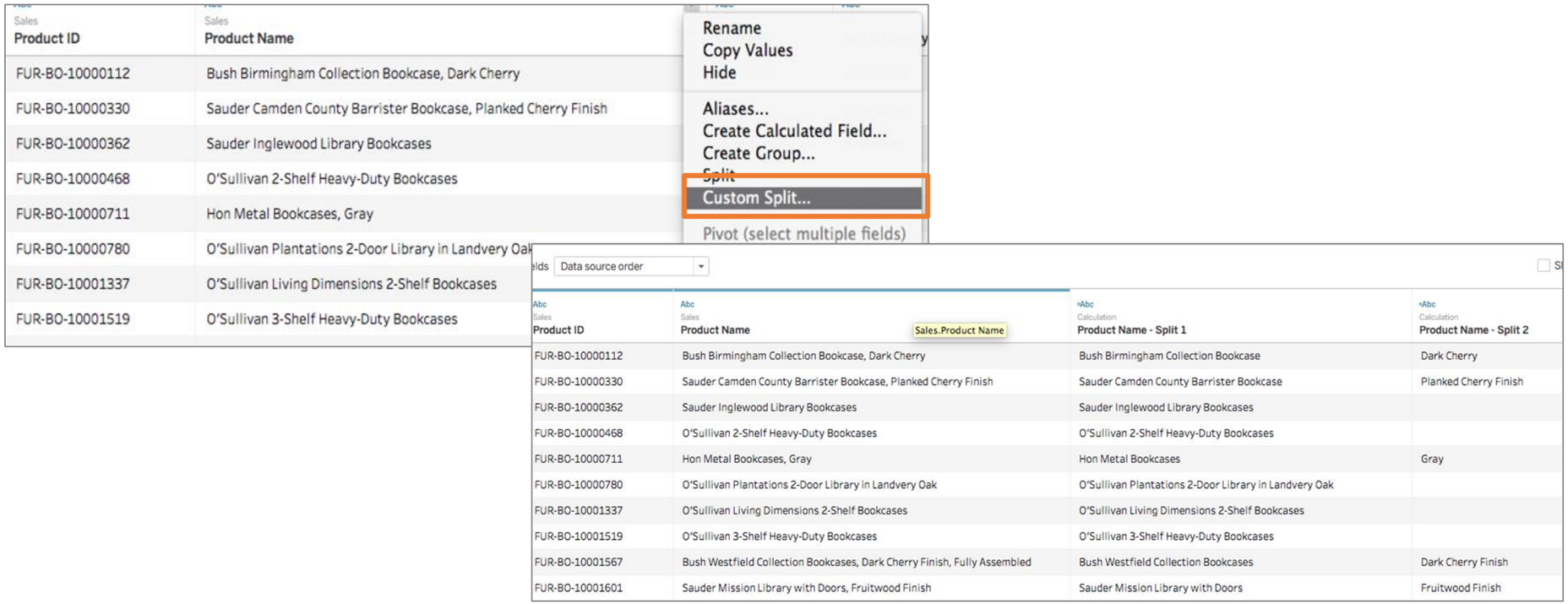




Splits

CUSTOM SPLITS

The custom split can also separate a string field into a maximum of ten new fields based on a separator within the original field.

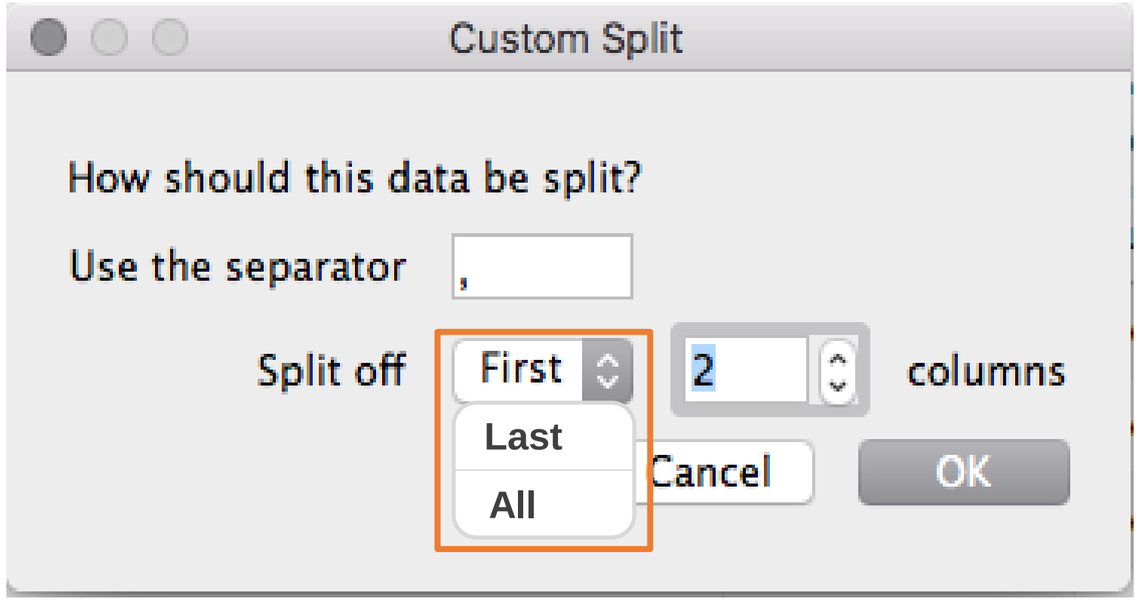


Splits

CUSTOM SPLITS

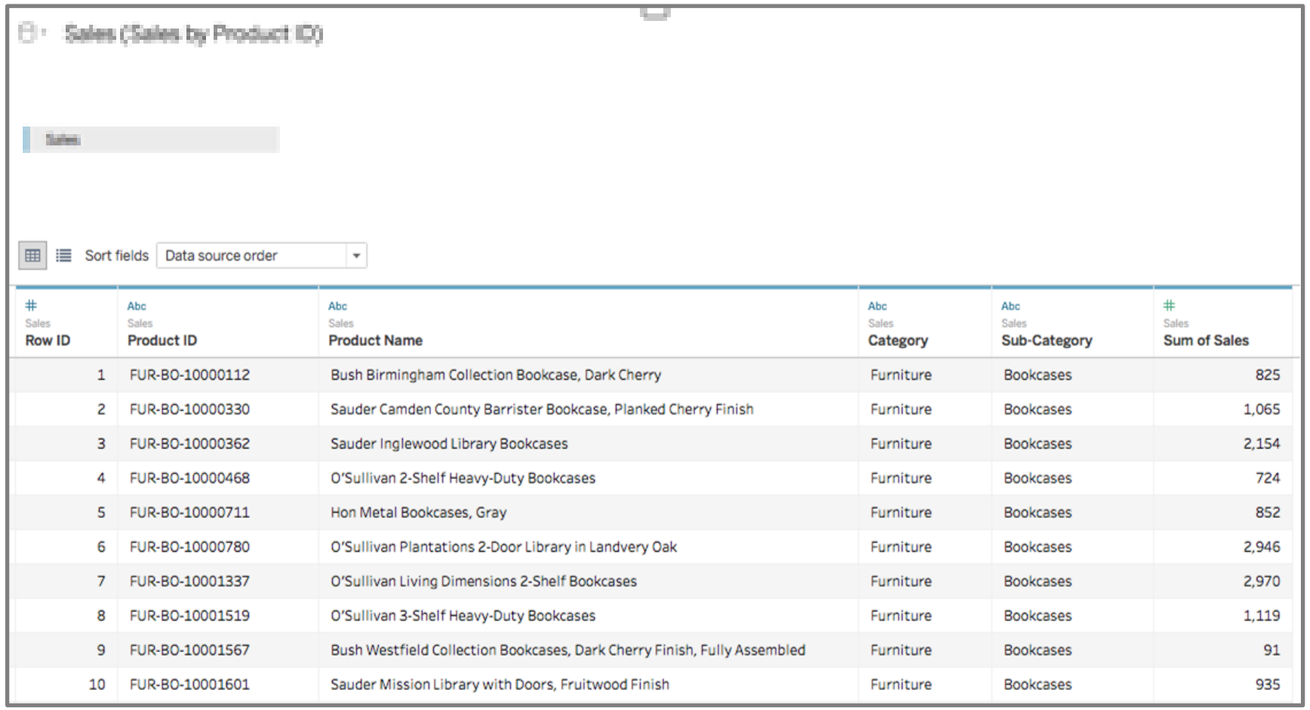
You can choose to split the values at:

- The first n occurrences of the separator
- The last n occurrences of the separator
- All the occurrences of the separator



Metadata Grid

- When preparing data for analysis, a list of fields is sometimes more useful than the data preview.
- The Metadata Grid view in Tableau allows you to quickly perform actions, such as rename, hide, and others, on multiple fields with a single command.



Sales Row ID	Sales Product ID	Sales Product Name	Sales Category	Sales Sub-Category	Sales Sum of Sales
1	FUR-BO-10000112	Bush Birmingham Collection Bookcase, Dark Cherry	Furniture	Bookcases	825
2	FUR-BO-10000330	Sauder Camden County Barrister Bookcase, Planked Cherry Finish	Furniture	Bookcases	1,065
3	FUR-BO-10000362	Sauder Inglewood Library Bookcases	Furniture	Bookcases	2,154
4	FUR-BO-10000468	O'Sullivan 2-Shelf Heavy-Duty Bookcases	Furniture	Bookcases	724
5	FUR-BO-10000711	Hon Metal Bookcases, Gray	Furniture	Bookcases	852
6	FUR-BO-10000780	O'Sullivan Plantations 2-Door Library in Landvery Oak	Furniture	Bookcases	2,946
7	FUR-BO-10001337	O'Sullivan Living Dimensions 2-Shelf Bookcases	Furniture	Bookcases	2,970
8	FUR-BO-10001519	O'Sullivan 3-Shelf Heavy-Duty Bookcases	Furniture	Bookcases	1,119
9	FUR-BO-10001567	Bush Westfield Collection Bookcases, Dark Cherry Finish, Fully Assembled	Furniture	Bookcases	91
10	FUR-BO-10001601	Sauder Mission Library with Doors, Fruitwood Finish	Furniture	Bookcases	935

Pivot

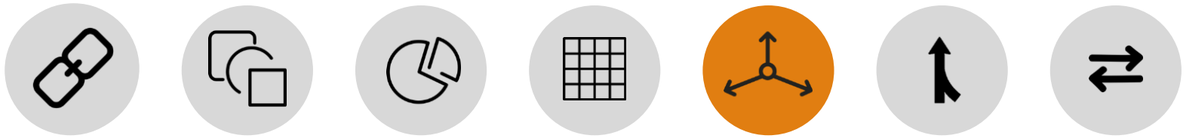
Data is often not organized as a typical data set: field names along the columns and members along the rows.

Example:

Row ID	Product Name	2012	2013	2014	2015
1	Bush Somerset Collection Bookcase			42	
2	Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back			220	
3	Self-Adhesive Address Labels for Typewriters by Universal			7	
4	Bretford CR4500 Series Slim Rectangular Table		-383		
5	Eldon Fold 'N Roll Cart System		3		
6	Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood	14			
7	Newell 322	2			
8	Mitel 5320 IP Phone VoIP phone	91			
9	DXL Angle-View Binders with Locking Rings by Samsill	6			
10	Belkin F5C206VTEL 6 Outlet Surge	34			

The Pivot function in Tableau allows you to select the columns you want to manipulate and format them into a typical data set ready for analysis.

Row ID	Product Name	Year of Order Date	
1	Bush Somerset Collection Bookcase	2014	42
2	Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	2014	220
3	Self-Adhesive Address Labels for Typewriters by Universal	2014	7
4	Bretford CR4500 Series Slim Rectangular Table	2013	-383
5	Eldon Fold 'N Roll Cart System	2013	3
6	Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood	2012	14
7	Newell 322	2012	2
8	Mitel 5320 IP Phone VoIP phone	2012	91
9	DXL Angle-View Binders with Locking Rings by Samsill	2012	6
10	Belkin F5C206VTEL 6 Outlet Surge	2012	34



Union

- Data often also resides in multiple, separate files and may need to be combined into a “master file.”
- Tableau’s “Union” feature helps you assemble data from multiple small files into one large file.

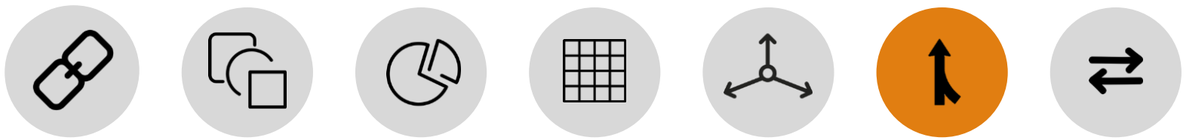
Region	Category	Profit	Sales
West	Furniture	$11,505	$252,613
	Office Supplies	$52,610	$220,853
	Technology	$44,304	$251,992

Region	Category	Profit	Sales
East	Furniture	$3,046	$208,291
	Office Supplies	$41,015	$205,516
	Technology	$47,462	$264,974

Region	Category	Profit	Sales
Central	Furniture	($2,871)	$163,797
	Office Supplies	$8,880	$167,026
	Technology	$33,697	$170,416

Region	Category	Profit	Sales
South	Furniture	$6,771	$117,299
	Office Supplies	$19,986	$125,651
	Technology	$19,992	$148,772

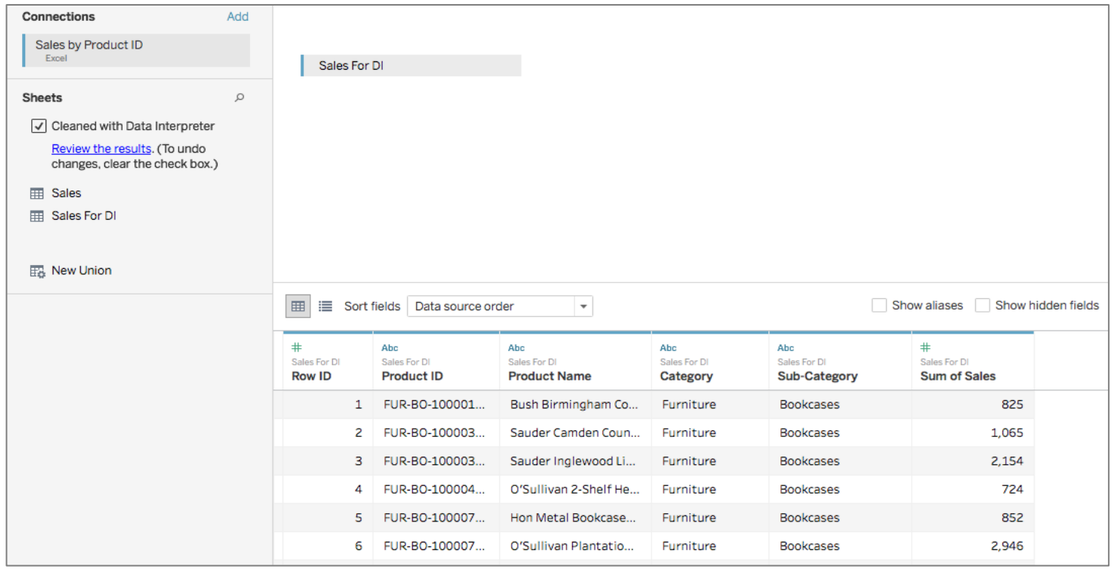
Region	Category	Profit	Sales
Central	Furniture	($2,871)	$163,797
	Office Supplies	$8,880	$167,026
	Technology	$33,697	$170,416
East	Furniture	$3,046	$208,291
	Office Supplies	$41,015	$205,516
	Technology	$47,462	$264,974
South	Furniture	$6,771	$117,299
	Office Supplies	$19,986	$125,651
	Technology	$19,992	$148,772
West	Furniture	$11,505	$252,613
	Office Supplies	$52,610	$220,853
	Technology	$44,304	$251,992

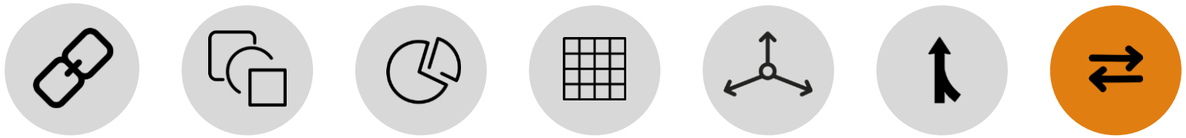


Data Interpreter

- This function automatically “cleans” your data and preps it for analysis.
- Examples of items that need to be cleaned prior to analysis:
    - Merged cells
    - Titles
    - Footnotes
    - Blank rows or columns

	B	C	D	E	F
1	SALES BY PRODUCT ID				
2					
3					
4	Product ID	Product Name	Category	Sub-Category	Sum of Sales
5	FUR-BO-10000112	Bush Birmingham Collection Bookcase, Dark Cherry	Furniture	Bookcases	$825
6	FUR-BO-10000330	Sauder Camden County Barrister Bookcase, Planked Cherry Finish	Furniture	Bookcases	$1,065
7	FUR-BO-10000362	Sauder Inglewood Library Bookcases	Furniture	Bookcases	$2,154
8	FUR-BO-10000468	O'Sullivan 2-Shelf Heavy-Duty Bookcases	Furniture	Bookcases	$724
9	FUR-BO-10000711	Hon Metal Bookcases, Gray	Furniture	Bookcases	$852
10	FUR-BO-10000780	O'Sullivan Plantations 2-Door Library in Landvery Oak	Furniture	Bookcases	$2,946
11	FUR-BO-10001337	O'Sullivan Living Dimensions 2-Shelf Bookcases	Furniture	Bookcases	$2,970
12	FUR-BO-10001519	O'Sullivan 3-Shelf Heavy-Duty Bookcases	Furniture	Bookcases	$1,119



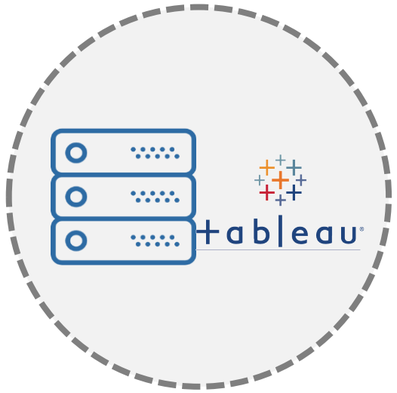


Data Connections

Topic 2: Connecting Data to Tableau

Connecting Data to Tableau

We can connect data to Tableau in one of the following ways:



Data sources published to Tableau Server

Queries

Web data through web data connectors

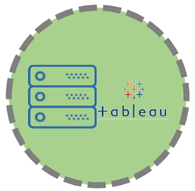
Local files

Connecting Data to Tableau

TABLEAU SERVER

With a Tableau Server deployment, you can:

- post and access data sources published to an instance of Tableau server
- connect to the data “live” or generate an extract from the data source
- schedule automatic data extract refreshes



Data sources published to Tableau Server

Queries

Web data through web data connectors

Local files

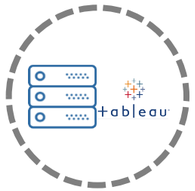
Demo—Connecting to a Data Source Through Tableau Server

TABLEAU DESKTOP 10

Connecting Data to Tableau

SQL QUERIES

You can create a connection to an SQL database such as MySQL or Microsoft SQL Server.



Data sources published to Tableau Server

Queries

Web data through web data connectors

Local files

Connecting Data to Tableau

WEB DATA CONNECTORS

- With Tableau, you can connect to applications such as Google Analytics, Market, Salesforce, and others.
- For other applications, Tableau includes a wizard for building custom web data connectors.

Data sources published to Tableau Server

Queries

Web data through web data connectors

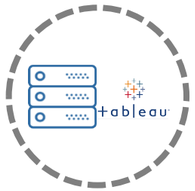
Local files

Connecting Data to Tableau

LOCAL FILES

These can be

- Text files
- Spreadsheets
- Statistical files such as SAS, tab/character delimited
- Tableau extracts/workbooks/data sources



Data sources published to Tableau Server

Queries

Web data through web data connectors

Local files

Demo—Connecting to Flat Files

TABLEAU DESKTOP 10

simplilearn

Data Connections

Topic 3: Methods of Performance Optimization

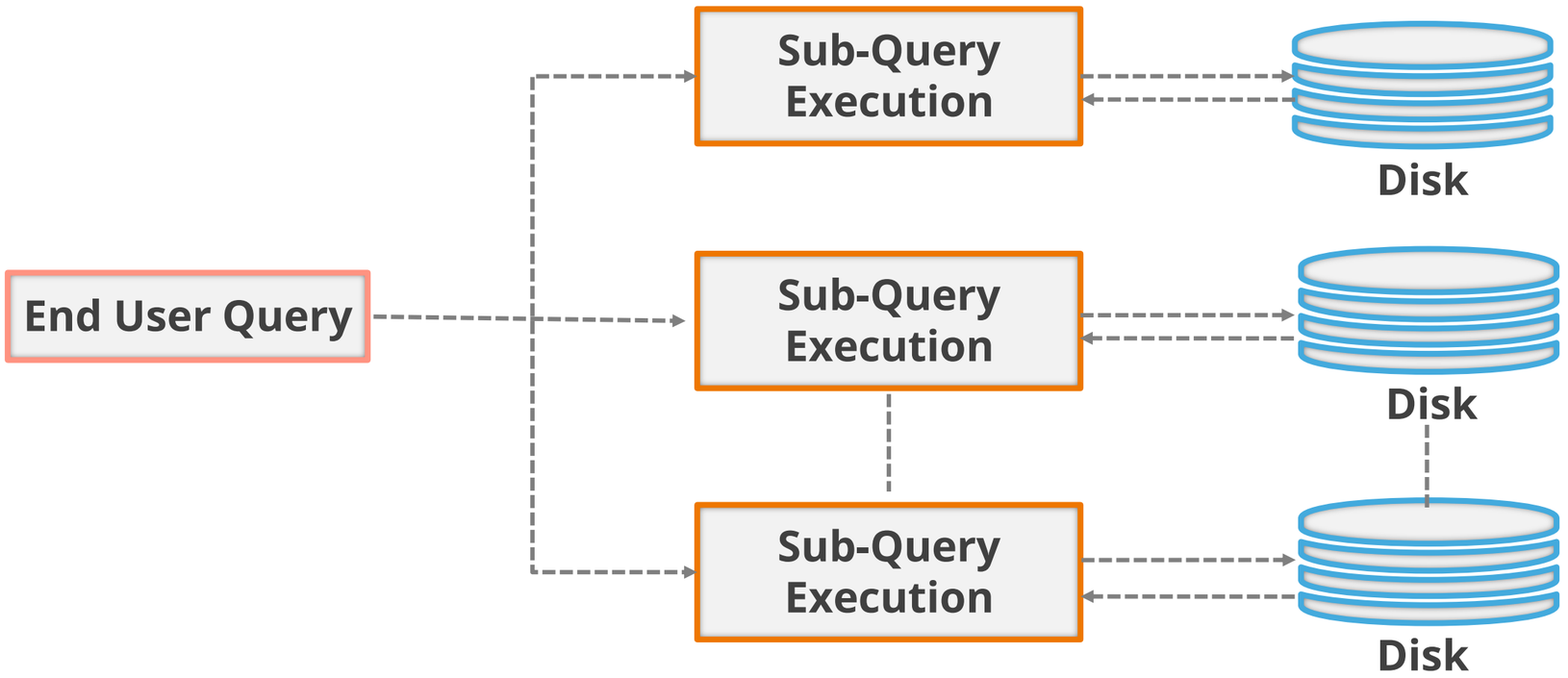
TABLEAU DESKTOP 10

Methods of Performance Optimization

- Running Queries in Parallel
- Data Engine Vectorization
- External Query Caching
- Query Fusion

RUNNING QUERIES IN PARALLEL

Running queries in parallel instead of serially is an effective way to boost performance.



Methods of Performance Optimization

- Running Queries in Parallel
- Data Engine Vectorization
- External Query Caching
- Query Fusion

DATA ENGINE VECTORIZATION

Tableau's data engine takes advantage of vector instructions on current processors.

The data engine uses SIMD instructions to perform low-level operations such as plus, minus, divide, min, max, sum, etc., on multiple data in parallel.

This means that basic computations can be performed more quickly.

Methods of Performance Optimization

- Running Queries in Parallel
- Data Engine Vectorization
- External Query Caching
- Query Fusion

EXTERNAL QUERY CACHING

- Tableau saves query results from the previous time the dashboard was opened.
- A single short query is run to fetch the cache data when the workbook is opened.

Methods of Performance Optimization

- Running Queries in Parallel
- Data Engine Vectorization
- External Query Caching
- Query Fusion

QUERY FUSION

This is a technology for database connections that looks at all of the queries in the dashboard and finds ways to consolidate them into fewer queries.

Data Connections

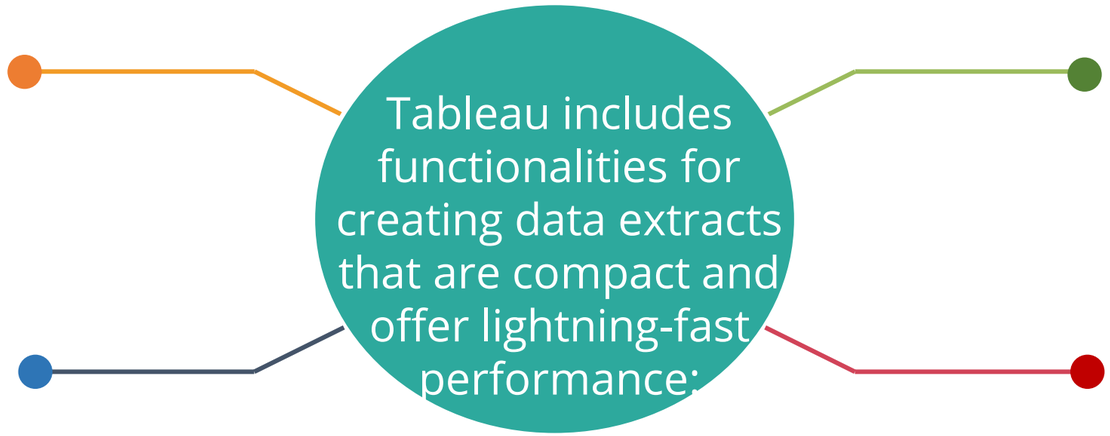
Topic 4: Tableau Data Extract Capabilities

Tableau Data Extracts

Data from your data source can be “extracted” into a file called a Tableau data extract; it transforms data into a Tableau-friendly format, improving query efficiency.

Custom filters can limit the scope of information included in an extract.

Extract refreshes can be executed through a schedule (with Tableau Server).



Custom aggregations can minimize the amount of data loaded into the extract.

Extracts can also be published to Tableau Server for sharing.

Demo—Creating a Data Extract

TABLEAU DESKTOP 10

simplilearn

Quiz

TABLEAU DESKTOP 10

QUIZ
1

When should you use a Blend instead of a Join to combine data?

a. Data within the sources are at different levels of detail

b. You are working with large data sets

c. You need to combine data from sources that don't support cross-table joins

d. All of the above

QUIZ
1

When should you use a Blend instead of a Join to combine data?

a. Data within the sources are at different levels of detail

b. You are working with large data sets

c. You need to combine data from sources that don't support cross-table joins

d. All of the above

The correct answer is **d.**

In each of these situations, you should consider using a Blend instead of a Join to combine data from different databases.

QUIZ
2

The Tableau _____ function separates a string field into multiple string fields.

a. Union

b. Join

c. Split

d. Data Interpreter

QUIZ
2

The Tableau _____ function separates a string field into multiple string fields.

a. Union

b. Join

c. Split

d. Data Interpreter

The correct answer is **c.**

The Split function is used to separate a string field into multiple string fields.

QUIZ
3

Technology for database connections that will look at all of the queries in your dashboard and find ways to consolidate them into fewer queries is called _____.

a. Query fusion

b. Parallel aggregation

c. Data Engine Vectorization

d. Query optimization

QUIZ
3

Technology for database connections that will look at all of the queries in your dashboard and find ways to consolidate them into fewer queries is called _____.

a. Query fusion

b. Parallel aggregation

c. Data Engine Vectorization

d. Query optimization

The correct answer is **a.**

Query fusion is a technology that examines all queries in your dashboard and consolidates them to reduce the number of queries hitting your processors.

Guided Exercise

Tableau DESKTOP 10

Problem Statement

Genelia needs to analyze Sales data for her company, and the data she needs is coming from multiple databases. The data sources include Excel spreadsheet files, an SQL database, and Salesforce.com. Additionally, the data is at different levels of detail and forms an extremely large data set.

- Which tools should she consider to prepare the data for analysis?
- Should she attempt to combine the data with a Blend or a Join?
- What steps should she consider to ensure that queries are executed efficiently?

Solution

- Which tools should she consider to prepare the data for analysis?

  Genelia should consider the Union functionality to combine the Excel-based data.
  She should use the Pivot functionality to align the levels of detail within the different databases.

- Should she attempt to combine the data with a Blend or a Join?

  Given that she is working with a large set of data, and the data sources are at different levels of detail, she should use a Blend to combine the data.

- What can she do to ensure his queries are executed efficiently?

  Genelia should consider creating a Tableau data extract and filter out data that is not necessary for this analysis. This will improve the performance of her queries.

Key Takeaways

- Data is rarely clean, tidy, and complete, and it is typically not in one file or even one type of file.
- A good BI professional takes her time to understand data prior to using it for analysis with a BI tool like Tableau.
- Tableau offers a suite of functions that help the user automate as much of the data preparation as possible.
- Tableau can point to several different types of files—from a simple Excel file to a custom, complex SQL database
- Data extracts in Tableau come in very handy as they can boost performance, get refreshed automatically, and limit the scope of your data.

This concludes “Data Connections.”