CHAPTER 2

Working with Single and Multiple Data Sources

"Most of us need to listen to the music to understand how beautiful it is. But often that's how we present statistics: we just show the notes, we don't play the music."

— Hans Rosling, co-founder and chairman of the Gapminder Foundation, who developed the Trendalyzer software system

Chapter 1 familiarized us with the basic concepts of visualization, the need and significance of visualization, the features of Tableau, the Tableau product line and the various file formats in Tableau. This chapter will help us to understand how to work with single and multiple data sources in Tableau. We will explore the following:

- Desktop architecture
- Tableau environment
- Connect to a file
- Connect to a server
- Metadata grid
- Joins
- Custom SQL
- Data blending
- Data extracts

Let us start with the desktop architecture.

2.1 Desktop architecture

Tableau architecture is based on an n-tier client server architecture (Shown in Fig. 2-1.) Tableau serves as a desktop installed software, web client and mobile client. Tableau Desktop is an authoring and publishing tool. It is used to create shared views on Tableau Server. Tableau offers a scalable solution to create and deliver desktop, web and mobile analytics. Tableau Desktop allows one to explore data and share insights.

S. Acharya and S. Chellappan, *Pro Tableau*, DOI 10.1007/978-1-4842-2352-9_2

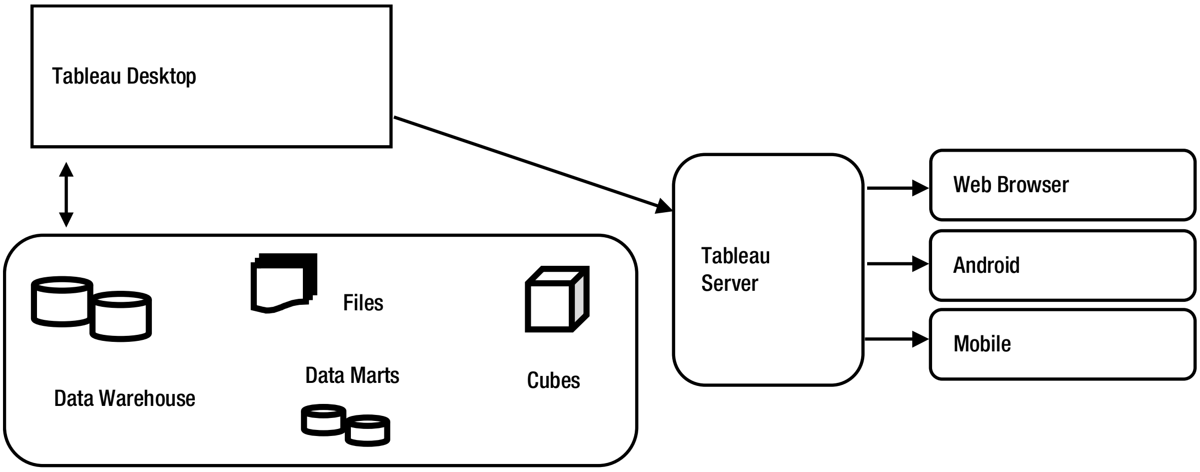


Figure 2-1. *Tableau Desktop Architecture*

Let us discuss the various layers of Tableau desktop architecture in brief.

2.1.1 Data layer

The bedrock of Tableau is its data layer. Tableau allows you to work with a heterogeneous data environment. You can work with databases, servers, data warehouses, cubes and flat files such as Excel, Access, etc. In Tableau, it is not necessary to bring all your data into memory unless it is required. Tableau allows you to leverage your existing environment by applying the database features to answer your questions.

2.1.2 Data connectors

Tableau provides various data connectors to work with databases such as Microsoft SQL Server, Oracle, Teradata, Vertica, Cloudera Hadoop, and many more. In addition to this there are generic ODBC connectors to connect to any system without having a native connector. In Tableau, there are two modes to interact with data: (i) live connection (ii) in-memory. Tableau users can switch between these two with ease.

2.1.3 Live connection

Tableau's data connectors allows you to leverage your existing data infrastructure. This is done by sending dynamic SQL or MDX statements to the source database directly instead of importing all the data. It means, if you have invested in a fast, analytics-optimized database like Vertica, you can get the benefits of those by connecting live to your data. This leaves the detail data in the source system and send the aggregate results of queries to Tableau. Tableau can also utilize unlimited amount of data. Tableau is the front-end analytics client to many of the largest databases in the world. Each connector is optimized to take the unique characteristics of each data source.

2.1.4 In-memory

Tableau has a fast, in-memory data engine for analytics. Tableau allows you to connect your data with one click, extract and bring it in memory. Tableau's data engine fully utilizes the entire system to achieve fast query response on hundreds of millions of rows of data on commodity hardware. Because the data engine can access disk storage as well as RAM and cache memory, it is not limited by the amount of memory on a system. There is no requirement that an entire data set be loaded into memory to achieve its performance goals.

In-memory is ideal when:

- your database is too slow for interactive analytics.
- you need to take load off a transactional database.
- you need to be offline and can't connect to your data live.

But live connections can be preferable when:

- you have a fast database, like Vertica, Teradata, or another analytics-optimized database.
- you need up-to-the minute data.

Refer to link below to learn about in memory and live data: Which is better?

http://www.tableau.com/learn/whitepapers/memory-or-live-data

2.2 Tableau environment

Let us try to understand the Tableau environment.

2.2.1 To open

Double click the Tableau icon on the desktop (Shown in Fig. 2-2).

Figure 2-2. *Tableau shortcut icon*

2.2.2 To close

Click on close button on the right side of the application (Shown in Fig. 2-3).

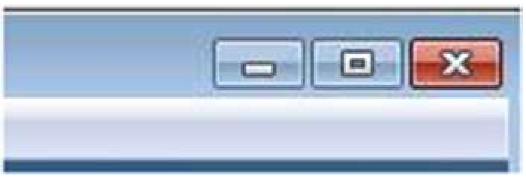


Figure 2-3. *Close button*

Next, we will learn about the start page.

2.2.3 Start page

The start page is a central location to help connect to data sources, access recent work books and explore tutorials provided by the Tableau community (Shown in Fig. 2-4).

Figure 2-4. *The Tableau start page*

There are three panes in the start page.

1. Connect**:** Using Connect, you can connect to various data sources such as connect to a file and connect to a server. Also you can open the saved data sources. “Sample - Superstore” is the default saved data source that comes with the Tableau Desktop Edition. Refer to Fig. 2-5.

Figure 2-5. *Connect page*

2. Open: When you open Tableau for the first time, this pane will be empty. As you start creating workbooks, you can see the most recently opened workbooks in this pane. You can also open sample workbooks to explore the functionality of Tableau. You can also pin workbooks to the start page by clicking the pin icon that appears in the top-left corner of the workbook thumbnail. Pinned workbooks always appear on the start page, even if they weren't opened recently. To remove a recently opened or pinned workbook, hover over the workbook thumbnail, and then click on the “x” that appears. The workbook thumbnail is removed immediately but will show again with your most recently used workbooks the next time you open Tableau Desktop. Refer to Fig. 2-6.

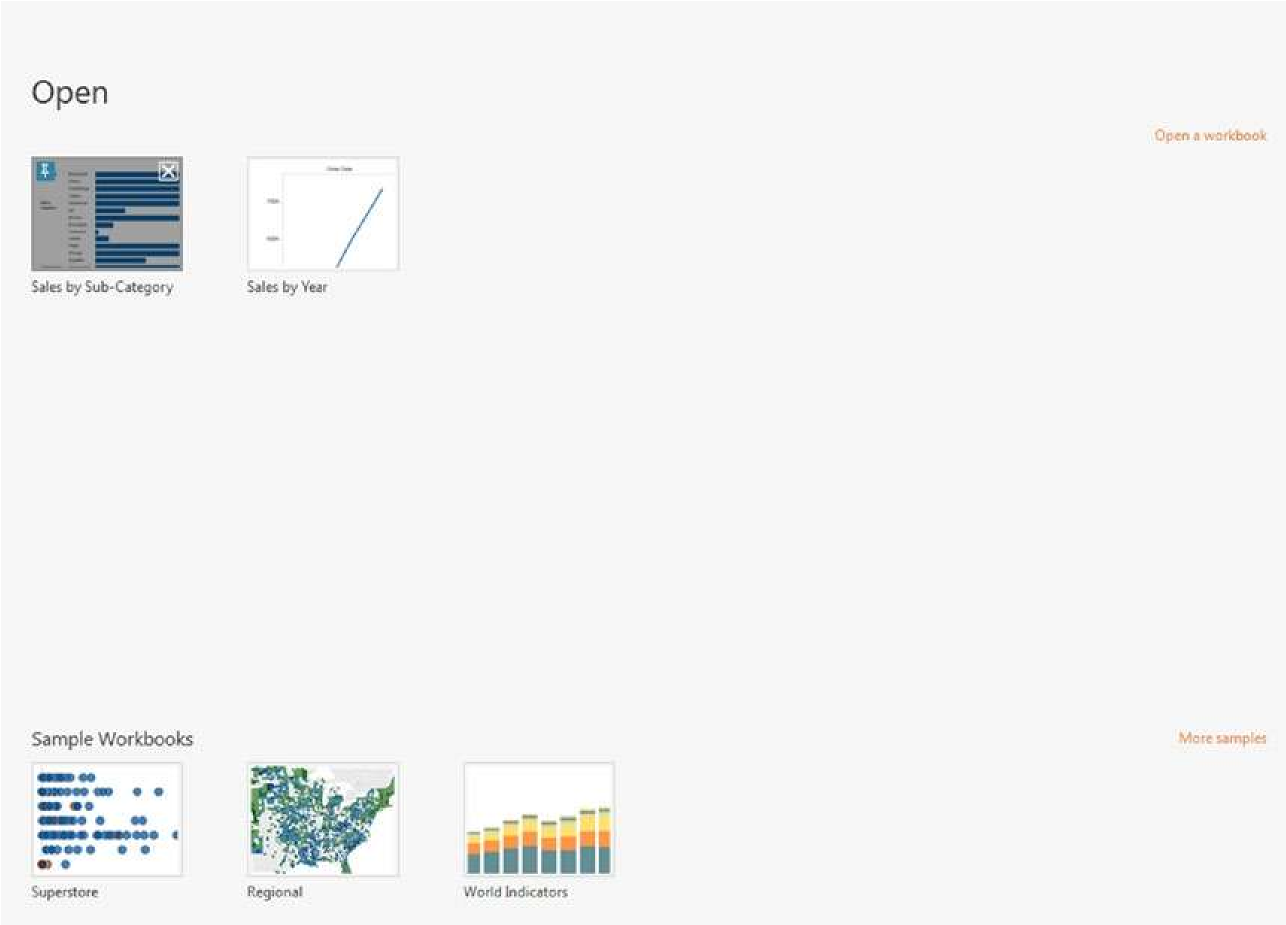


Figure 2-6. *Open page showing pin option*

3. Discover: You can view details about training provided by Tableau, blogs, conferences and references, etc. Refer to Fig. 2-7.

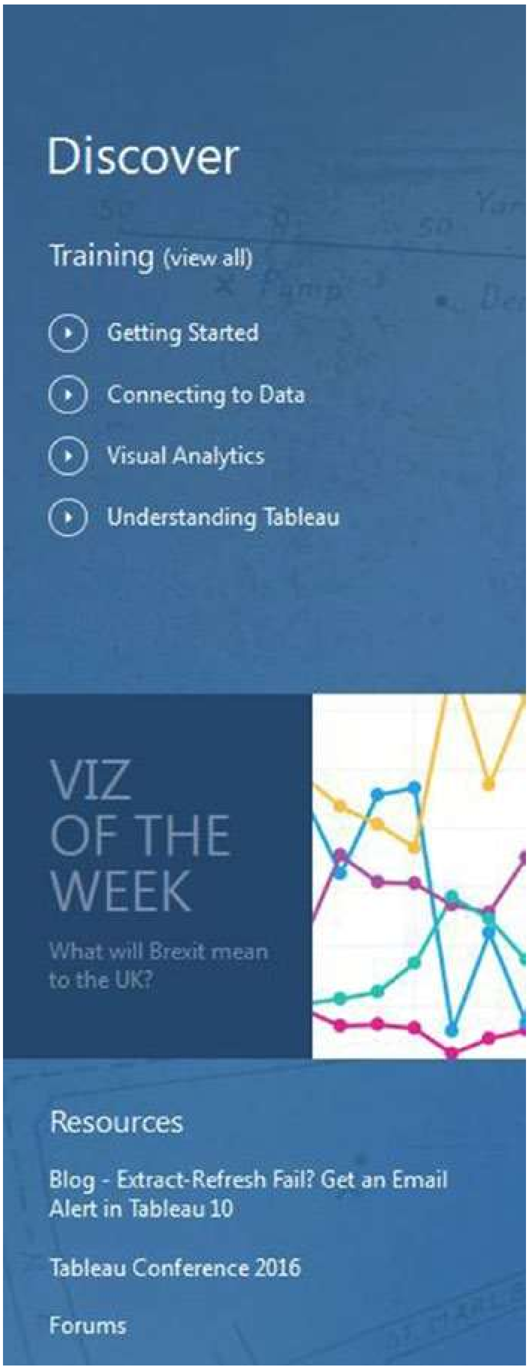


Figure 2-7. *"Discover" Page*

2.2.4 Data Source Page

You need to establish an initial connection to your data to get the data source page. You can follow the steps below to connect to an Excel file (Sample – Superstore).

Follow these steps:

2.2.4.1 Step 1

On the start page, under "Connect", select To a File ➤ Excel (Shown in Fig. 2-8).

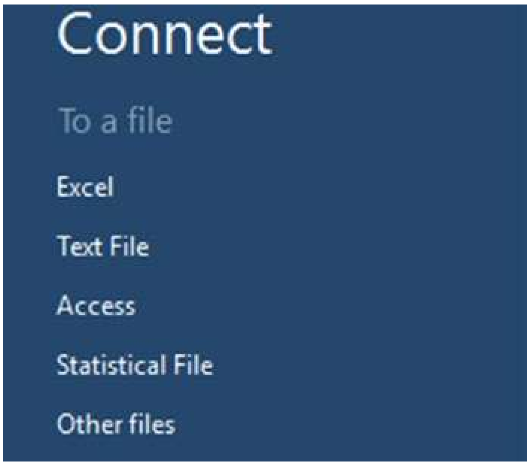


Figure 2-8. *"Connect" Section*

2.2.4.2 Step 2

"Open" dialog box will be opened. Navigate to the folder where Sample-Superstore excel file is present. In Tableau Desktop, the default path is "C:\Users\Username\Documents\My Tableau Repository\Datasources\9.3\en_US-US" (Shown in Fig. 2-9).

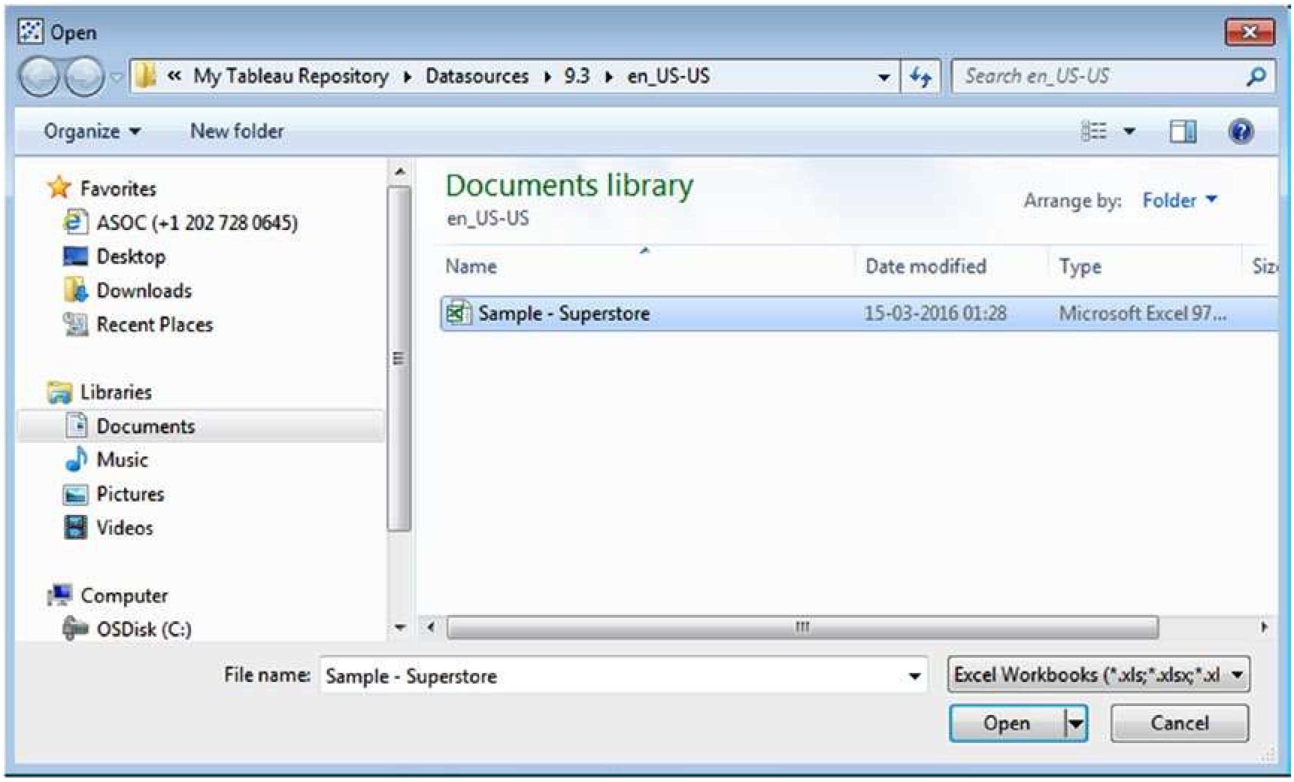


Figure 2-9. *"Open" dialog box to open "Sample - Superstore"*

2.2.4.3 Step 3

When you click on “Open” button, you can see “Processing Request” window as shown in Fig. 2-10.

Figure 2-10. *Processing request window*

2.2.4.4 Step 4

Now, you will be able to view the “Data Source Page”. (Shown in Fig. 2-11).

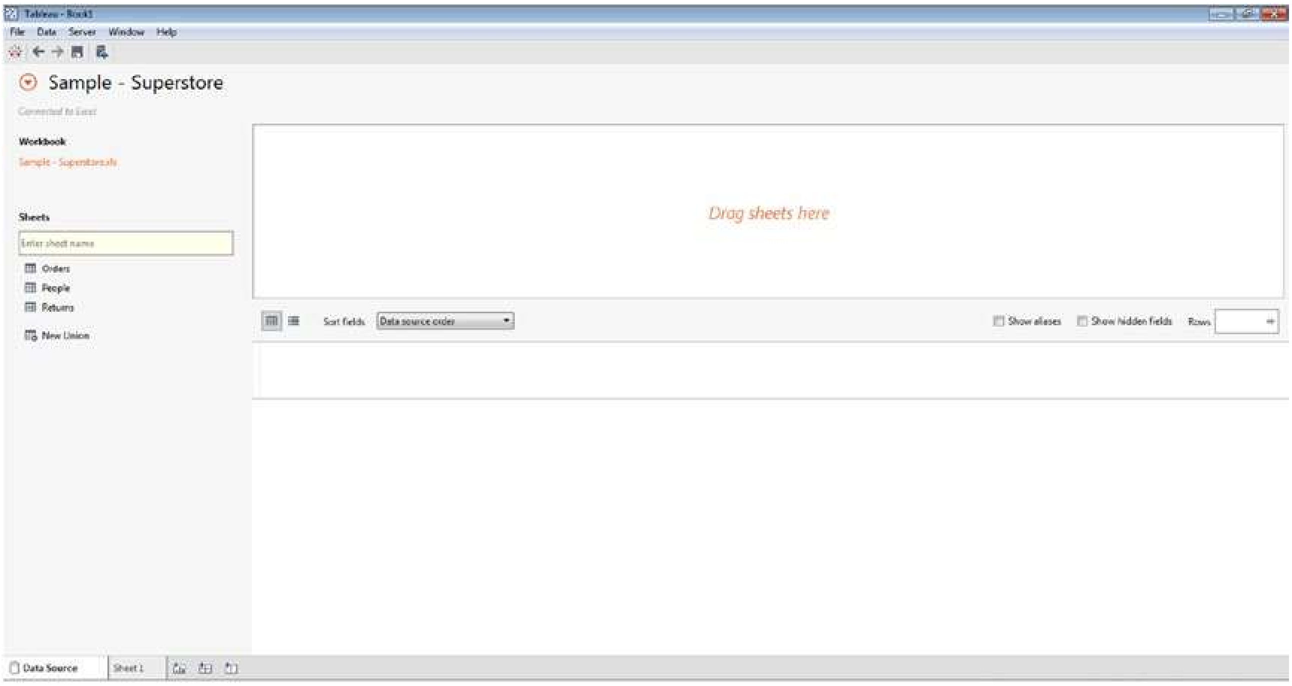


Figure 2-11. *Data source page*

2.2.4.5 Step 5

Drag “Orders” sheet from the left pane to the canvas area as shown in Fig. 2-12. You will be able to preview the data as well.

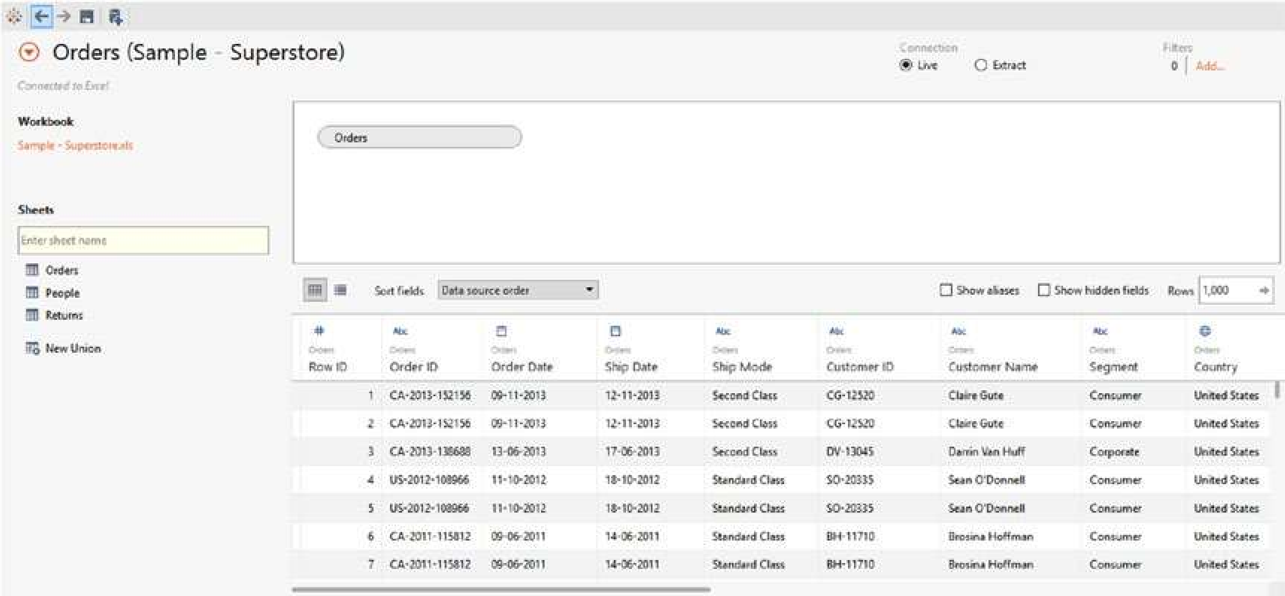


Figure 2-12. *"Orders" sheet placed on the canvas area*

There are four main areas in the Data Source Page. They are (i) left pane, (ii) canvas, (iii) grid, and (iv) metadata grid (Shown in Fig. 2-13).

1. Left pane**:** displays details about the data to which you are connected. For example, for file-based data, the file name and worksheets will be displayed.
2. Canvas: allows you to drag and drop one or more tables to the canvas area to set up your data source.
3. Grid: allows you to review first 1,000 rows of data that is present in your data source. It also allows you to modify your data source like renaming field names, sorting, creating a field, etc.
4. Metadata grid**:** allows one to click on the metadata grid to display fields in your data source.

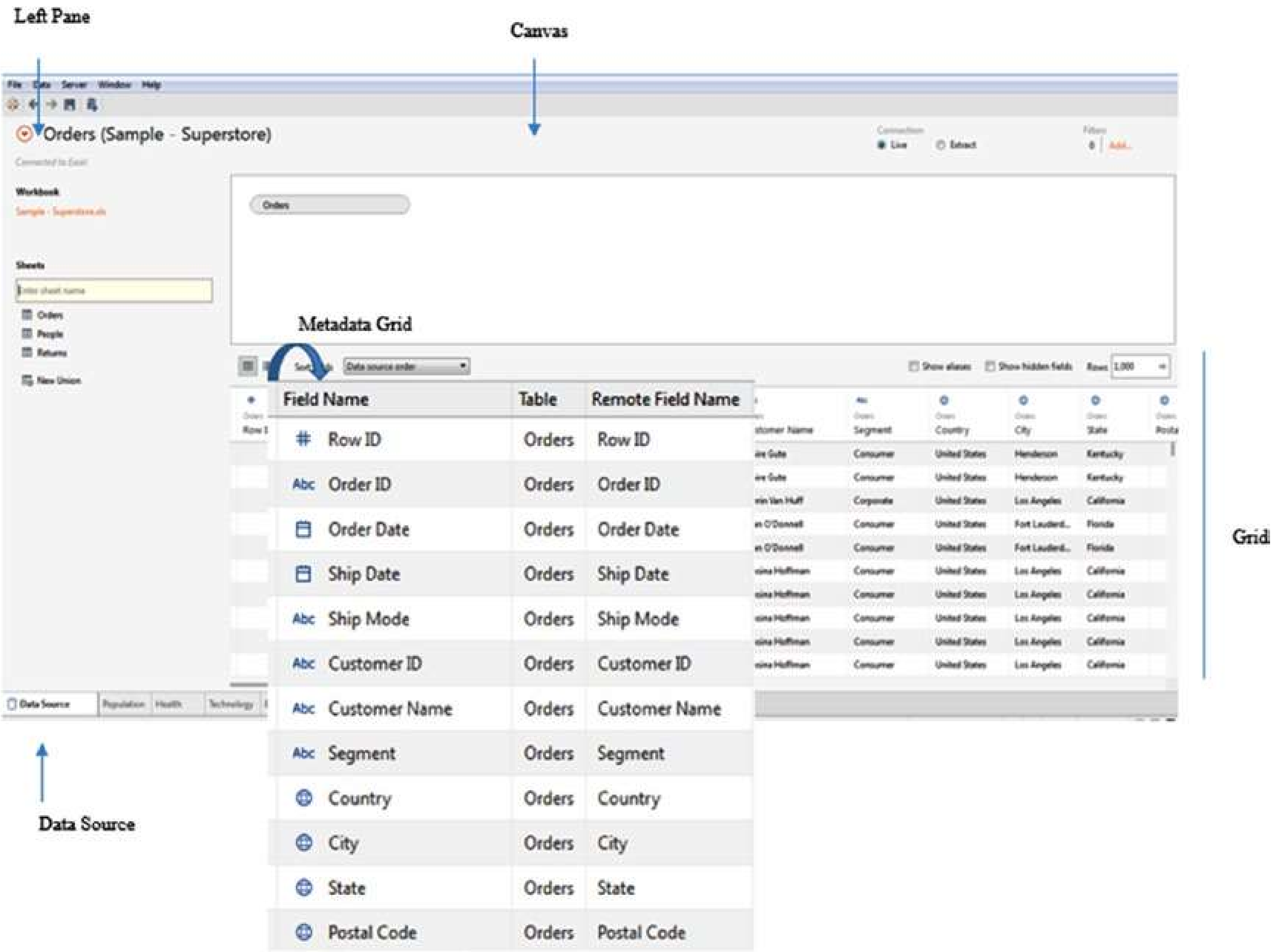


Figure 2-13. *Four main areas on data source page*

Tableau provides two types of connections:

- Live**:** This allows extracting data in real time.
- Extract**:** This is about extracting data at regular frequencies.

2.2.5 Workspace

Workspace contains data pane, cards and shelves, and one or more sheets. These sheets can either be worksheets, stories or dashboards. Cards and shelves can be used to build views (Shown in Fig. 2-14).

The workspace for creating a story is quite different from the workspace for creating a dashboard.

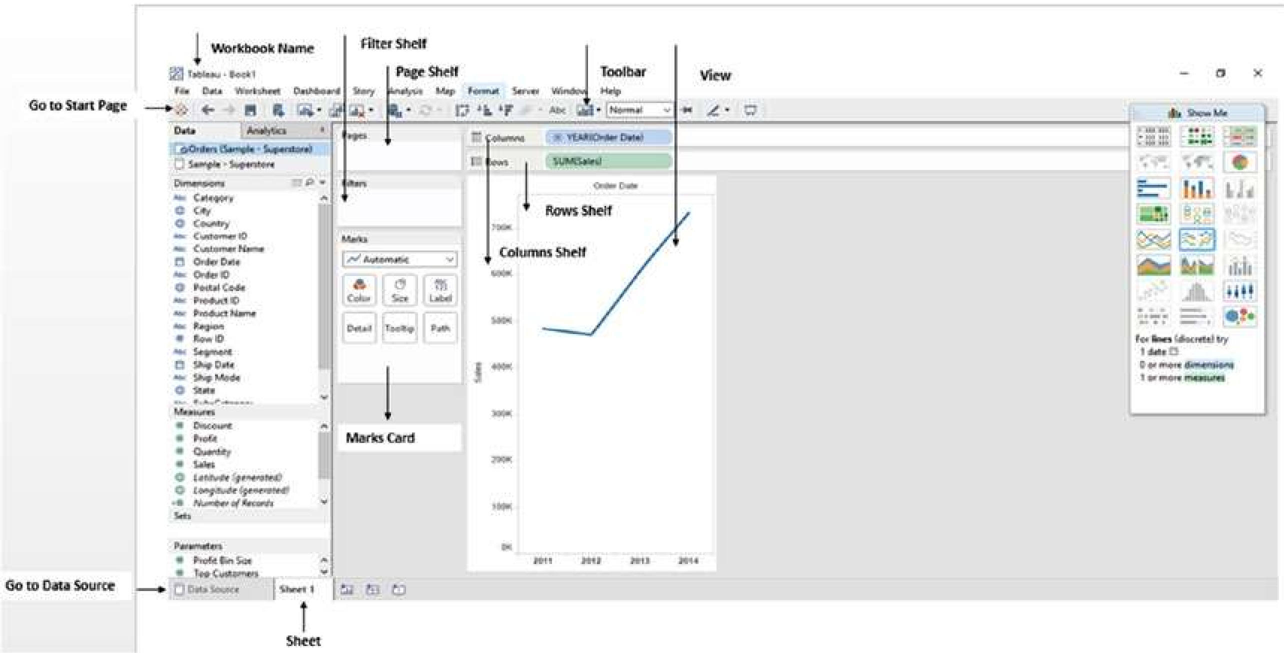


Figure 2-14. Tableau Workspace

2.2.6 Workbooks and Sheets

A Tableau workbook is quite similar to a Microsoft Excel workbook. A workbook can contain one or more sheets. These sheet can be a worksheet(s), a story or a dashboard(s). A workbook is a container for all your work. Workbook will help you to organize, perform analysis, and save and share your results.

- Worksheet represents single view with data pane, cards, shelves, and legends.
- Dashboard represents a collection of views from multiple worksheets.
- Story represents a sequence of worksheets or dashboards to convey certain information.

2.2.7 Visual Cues and Icons in Tableau

Let us discuss few visual cues and icons in Tableau.

2.2.7.1 Data Sources in Data Pane

Refer to Table 2-1 for data source icons.

Table 2-1. *Data Source Icons*

Visual Cue	Description
	Primary Data Source.
	Secondary Data Source.

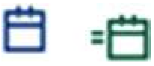
2.2.7.2 Fields in Data Pane

Let us discuss the icons displayed in the data pane.

- Discrete field is indicated by the “blue” color.
- Continuous field is indicated by the “green” color.
- User-defined function is indicated by an equal sign.

Refer to Table 2-2 for icons in data pane.

Table 2-2. *Icons in Data Pane*

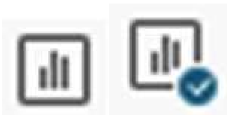
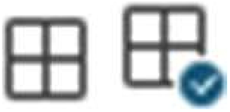
Visual Cue	Description
	Text values.
	Numeric values.
	ONLY date values.
	Both date and time values.
	Geographical data.
	User-defined set.
	Numeric bin.
	Group.
	Relational hierarchy.
	The field is blended with a field from another data source.
	The field is not blended with a field from another data source.

2.2.7.3 Sheets in the Dashboard and Worksheet Pane

Let us discuss the sheet(s) which are used in a story.

A blue check mark indicates that sheet is being used in one or more stories. Refer to Table 2-3.

Table 2-3. *Sheets in the Dashboard and Worksheet Pane*

Visual Cue	Description
	Worksheet.
	Dashboard.

Refer to the link below to learn more about the visual cues and icons in Tableau.

https://onlinehelp.tableau.com/current/pro/desktop/en-us/tips_visualcues.html

2.3 Connect to a File

Let us explore how to connect to the below-mentioned files.

- Text
- Microsoft Access
- R data file

2.3.1 Connect to a Text File

Follow the steps to connect to a text file.

2.3.1.1 Steps to connect to a text file

2.3.1.1.1 Step 1

On the start page, under Connect, select To a File ➤ Text File (Shown in Fig. 2-15)

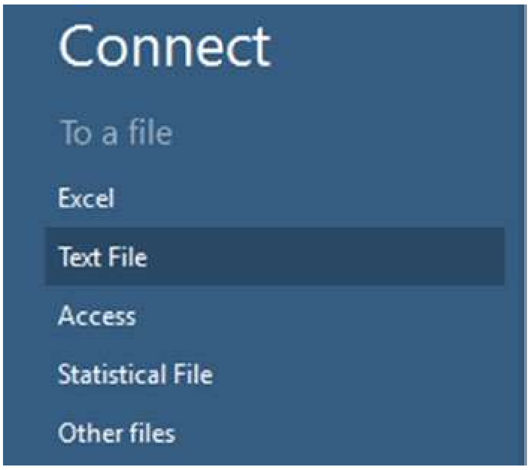


Figure 2-15. *Connect to a text file*

2.3.1.1.2 Step 2

"Open" dialog box will show up. Navigate to the folder where "Sample-Superstore" text file is present. Select the file and click on the "Open" button (Shown in Fig. 2-16).

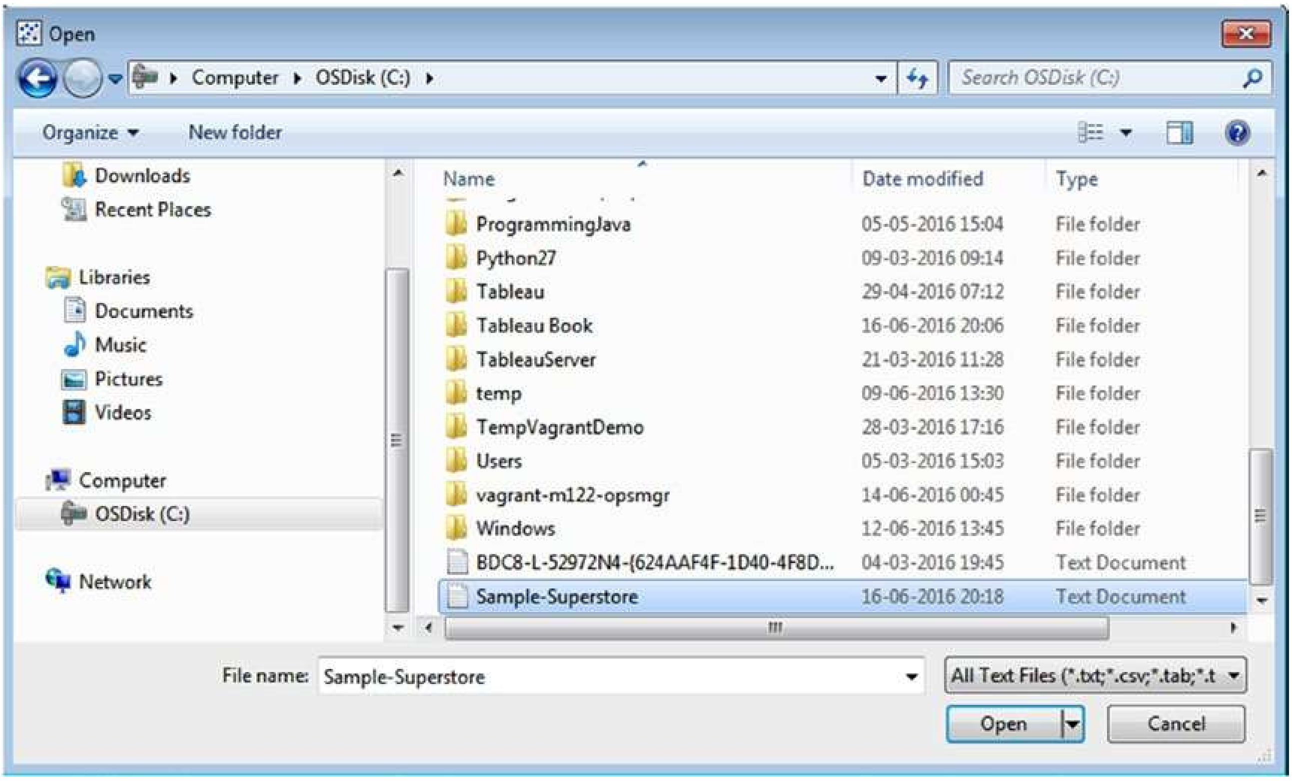


Figure 2-16. *Open dialog box to open the "Sample-Superstore" text file*

2.3.1.1.3 Step 3

Data source page will open as shown in Fig. 2-17.

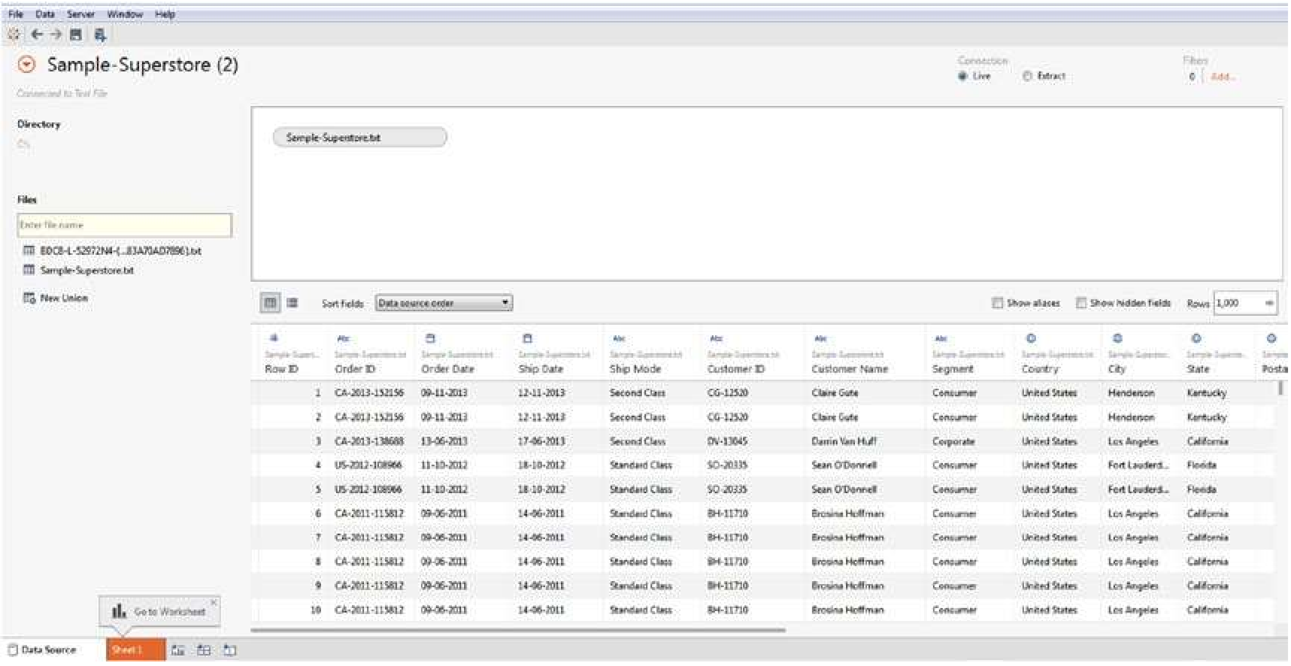


Figure 2-17. *Data source page showing the connection to "Sample-Superstore.xls" data source*

2.3.2 Connect to MS Access

Follow the steps to connect to MS Access.

2.3.2.1 Steps to connect to MS Access

2.3.2.1.1 Step 1

On the start page, under Connect, select To a File ➤ Access (Shown in Fig. 2-18).

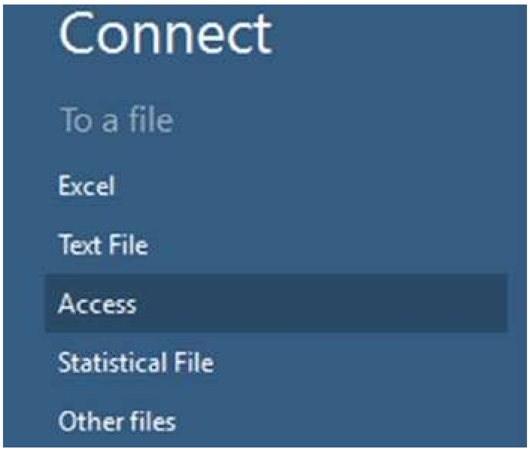


Figure 2-18. *Connect to "Access"*

2.3.2.1.2 Step 2

Access connection wizard will open (Shown in Fig. 2-19). Click on the "Browse" button. "Open dialog box" will show up.

Figure 2-19. *"Access Connection" window*

2.3.2.1.3 Step 3

Navigate to the folder where the required access file is present. Select the file and click on "Open" button (Shown in Fig. 2-20) to open an Access file (Shown in Fig. 2-21).

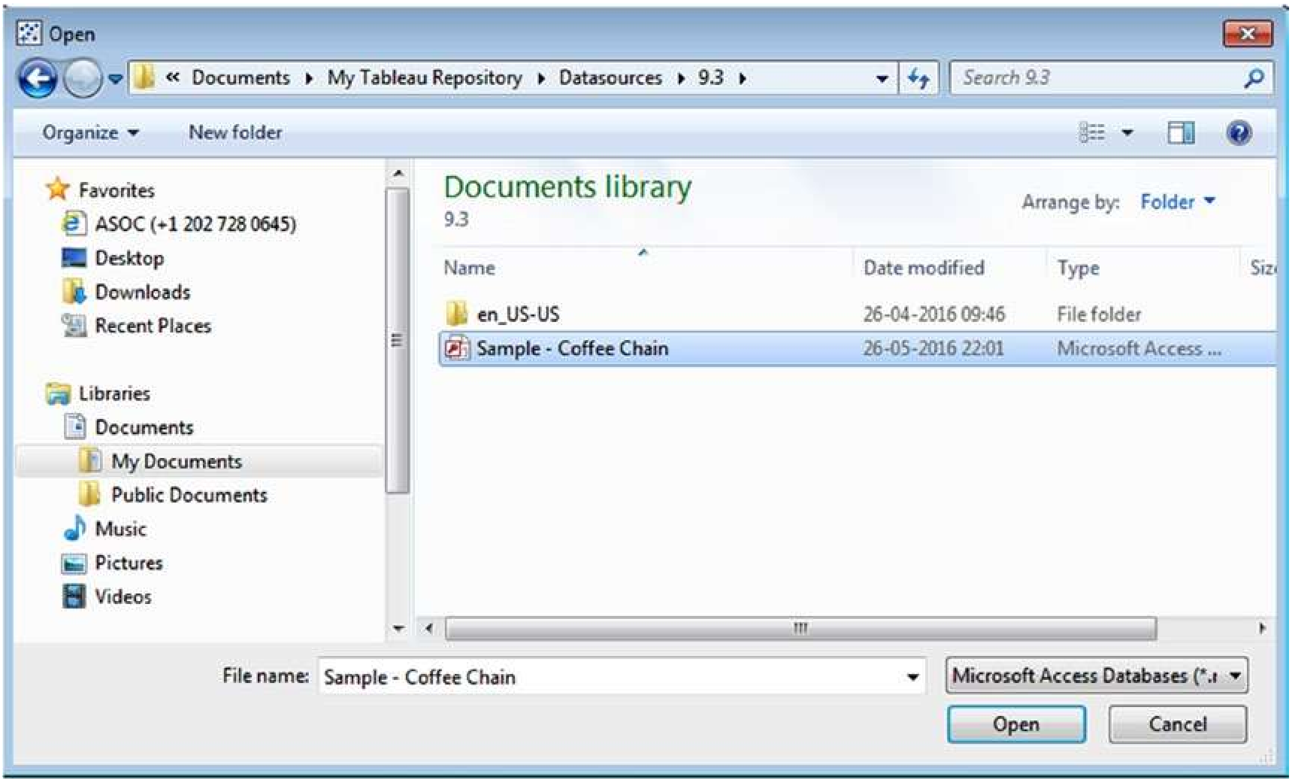


Figure 2-20. *Open dialog box to open Sample - Coffee Chain (Microsoft Access database)*

Figure 2-21. *Access Connection window showing the selected "Sample - Coffee Chain.mdb" file*

2.3.2.1.4 Step 4

Click on “OK” to connect to Access file (Shown in Fig. 2-22).

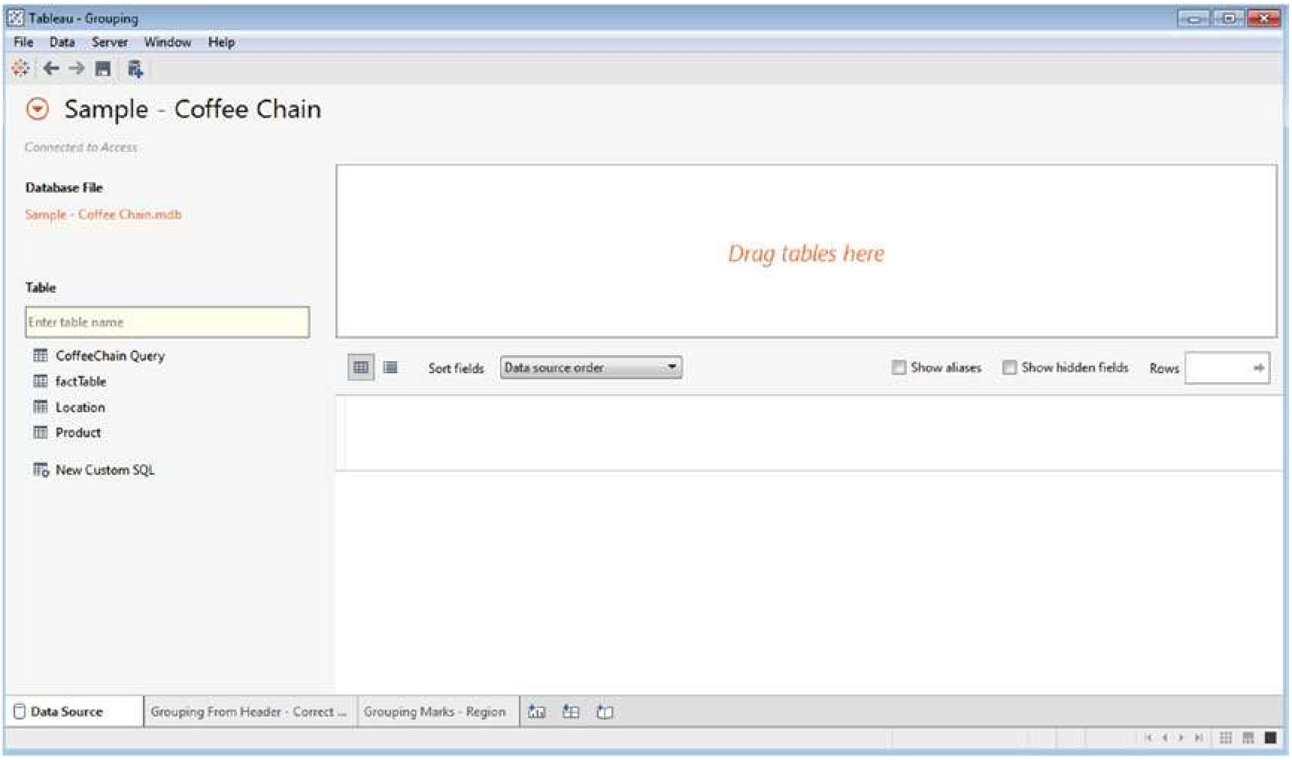


Figure 2-22. *Data source page showing connection to Sample-Coffee Chain Access data source*

2.3.3 Connecting to RData files

Perform the following steps to bring in the data stored in .RData file (created in R Statistical Programming Language) into Tableau.

2.3.3.1 Steps to connect to RData file

We will look at a .CSV file that we will read inside the R interface and save it into a dataset. Then this dataset will be saved to an .RData file. This file will then be read by tableau.

2.3.3.1.1 Step 1

DataSetIris.csv is the comma separated value file, available in the D: drive. Let us look at the data in the .csv file (Shown in Fig. 2-23).

	A	B	C	D	E	F
1		Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
2	1	5.1	3.5	1.4	0.2	setosa
3	2	4.9	3	1.4	0.2	setosa
4	3	4.7	3.2	1.3	0.2	setosa
5	4	4.6	3.1	1.5	0.2	setosa
6	5	5	3.6	1.4	0.2	setosa
7	6	5.4	3.9	1.7	0.4	setosa
8	7	4.6	3.4	1.4	0.3	setosa
9	8	5	3.4	1.5	0.2	setosa
10	9	4.4	2.9	1.4	0.2	setosa
11	10	4.9	3.1	1.5	0.1	setosa
12	11	5.4	3.7	1.5	0.2	setosa
13	12	4.8	3.4	1.6	0.2	setosa
14	13	4.8	3	1.4	0.1	setosa
15	14	4.3	3	1.1	0.1	setosa
16	15	5.8	4	1.2	0.2	setosa
17	16	5.7	4.4	1.5	0.4	setosa
18	17	5.4	3.9	1.3	0.4	setosa
19	18	5.1	3.5	1.4	0.3	setosa

DataSetIris

Figure 2-23. *DataSetIris.csv Data Set*

Let us explore how to create an .RData file.
Steps to read the DataSetIris.csv file into R

2.3.3.1.2 Step 2

Start the R interface. At the R command prompt issue the following command (Shown in Fig. 2-24 and Fig. 2-25).

Figure 2-24. *Command to read data from .csv file into a Dataset, "RDataSet"*

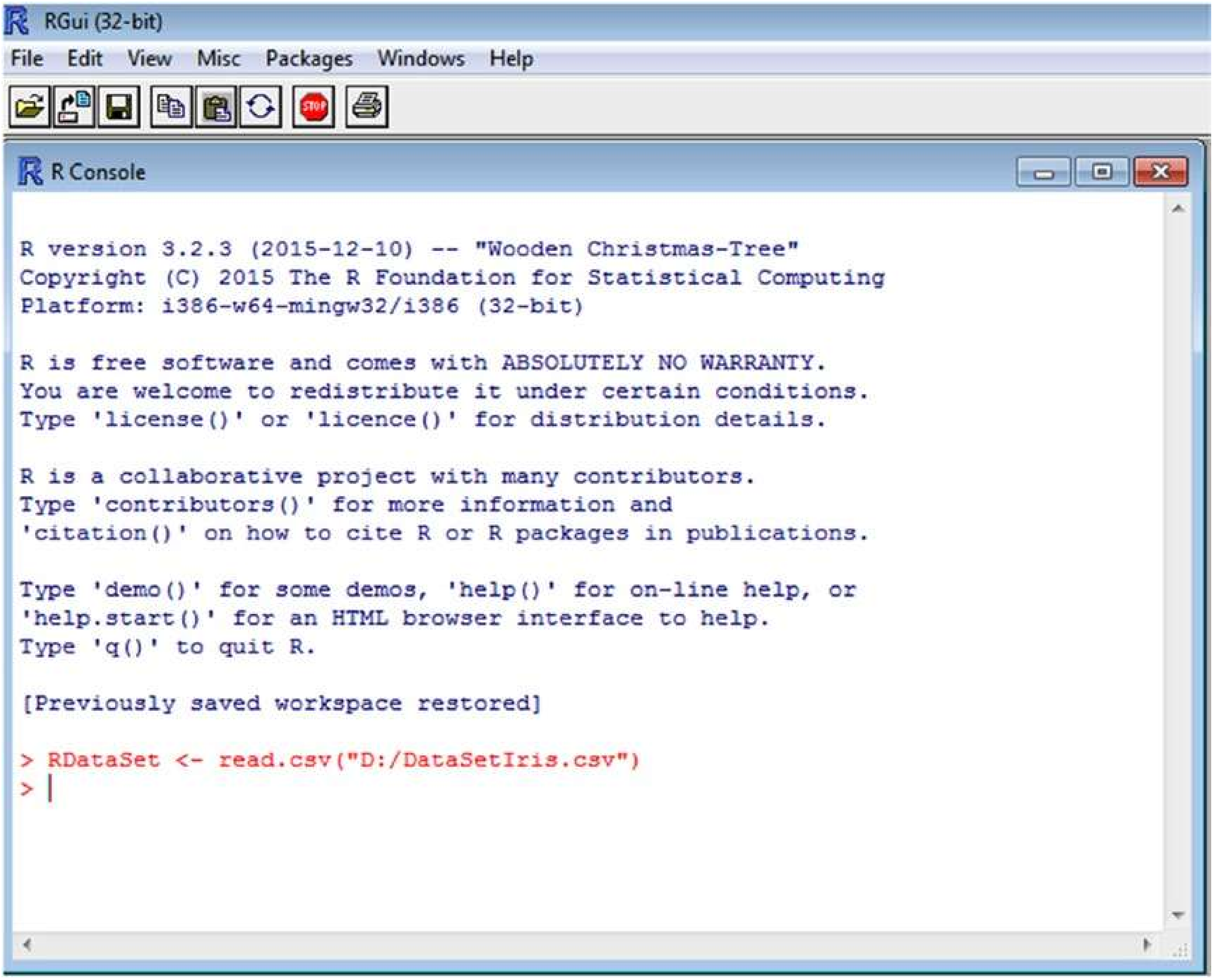


Figure 2-25. *R Interface*

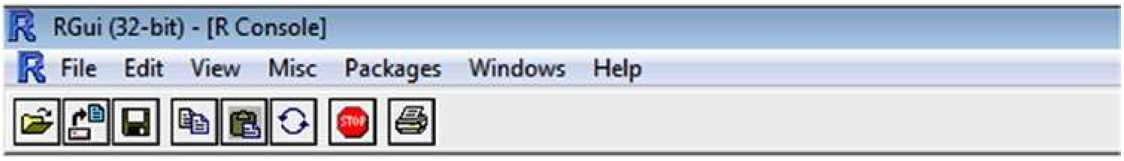
The above command creates a dataset by the name, "RDataSet". And into this data set is read the data values from "DataSetIris.csv" file which is available in the D: drive.

2.3.3.1.3 Step 3

To display the values contained in RDataSet, type the following command (Shown in Fig. 2-26 and Fig. 2-27).

```
> RDataSet
```

Figure 2-26. *Command to display data in "RDataSet"*




```
> RDataSet
    X Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
1   1          5.1         3.5          1.4         0.2     setosa
2   2          4.9         3.0          1.4         0.2     setosa
3   3          4.7         3.2          1.3         0.2     setosa
4   4          4.6         3.1          1.5         0.2     setosa
5   5          5.0         3.6          1.4         0.2     setosa
6   6          5.4         3.9          1.7         0.4     setosa
7   7          4.6         3.4          1.4         0.3     setosa
8   8          5.0         3.4          1.5         0.2     setosa
9   9          4.4         2.9          1.4         0.2     setosa
10 10          4.9         3.1          1.5         0.1     setosa
11 11          5.4         3.7          1.5         0.2     setosa
12 12          4.8         3.4          1.6         0.2     setosa
13 13          4.8         3.0          1.4         0.1     setosa
14 14          4.3         3.0          1.1         0.1     setosa
15 15          5.8         4.0          1.2         0.2     setosa
16 16          5.7         4.4          1.5         0.4     setosa
17 17          5.4         3.9          1.3         0.4     setosa
18 18          5.1         3.5          1.4         0.3     setosa
19 19          5.7         3.8          1.7         0.3     setosa
20 20          5.1         3.8          1.5         0.3     setosa
21 21          5.4         3.4          1.7         0.2     setosa
22 22          5.1         3.7          1.5         0.4     setosa
23 23          4.6         3.6          1.0         0.2     setosa
24 24          5.1         3.3          1.7         0.5     setosa
25 25          4.8         3.4          1.9         0.2     setosa
26 26          5.0         3.0          1.6         0.2     setosa
27 27          5.0         3.4          1.6         0.4     setosa
28 28          5.2         3.5          1.5         0.2     setosa
29 29          5.2         3.4          1.4         0.2     setosa
30 30          4.7         3.2          1.6         0.2     setosa
31 31          4.8         3.1          1.6         0.2     setosa
32 32          5.4         3.4          1.5         0.4     setosa
33 33          5.2         4.1          1.5         0.1     setosa
34 34          5.5         4.2          1.4         0.2     setosa
35 35          4.9         3.1          1.5         0.2     setosa
36 36          5.0         3.2          1.2         0.2     setosa
```

Figure 2-27. *Data in "RDataset"*

There are 150 such rows in the data set.

2.3.3.1.4 Step 4

You can make an .RData file by issuing the following command at the R prompt.
Save (RDataSet, file="D:/TableauDataSet.RData")

■ **Note** The Data Set, "RDataSet" is saved to the file, "TableauDataSet.RData".

2.3.3.1.5 Step 5

Start Tableau Desktop. Click on "Statistical File" under "To a File" (Shown in Fig. 2-28. Choose the statistical file that you wish to open within R (Shown in Fig. 2-29).

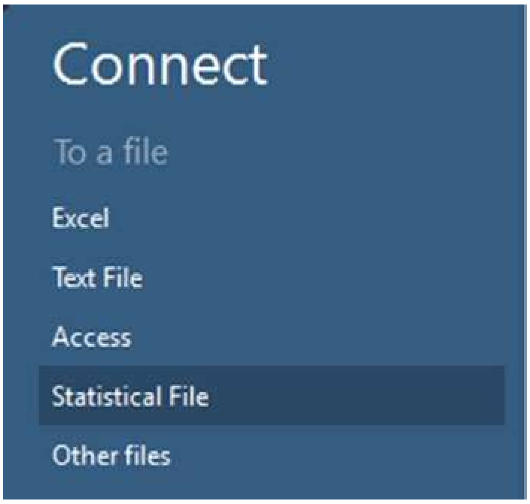


Figure 2-28. *Connecting to a statistical file*

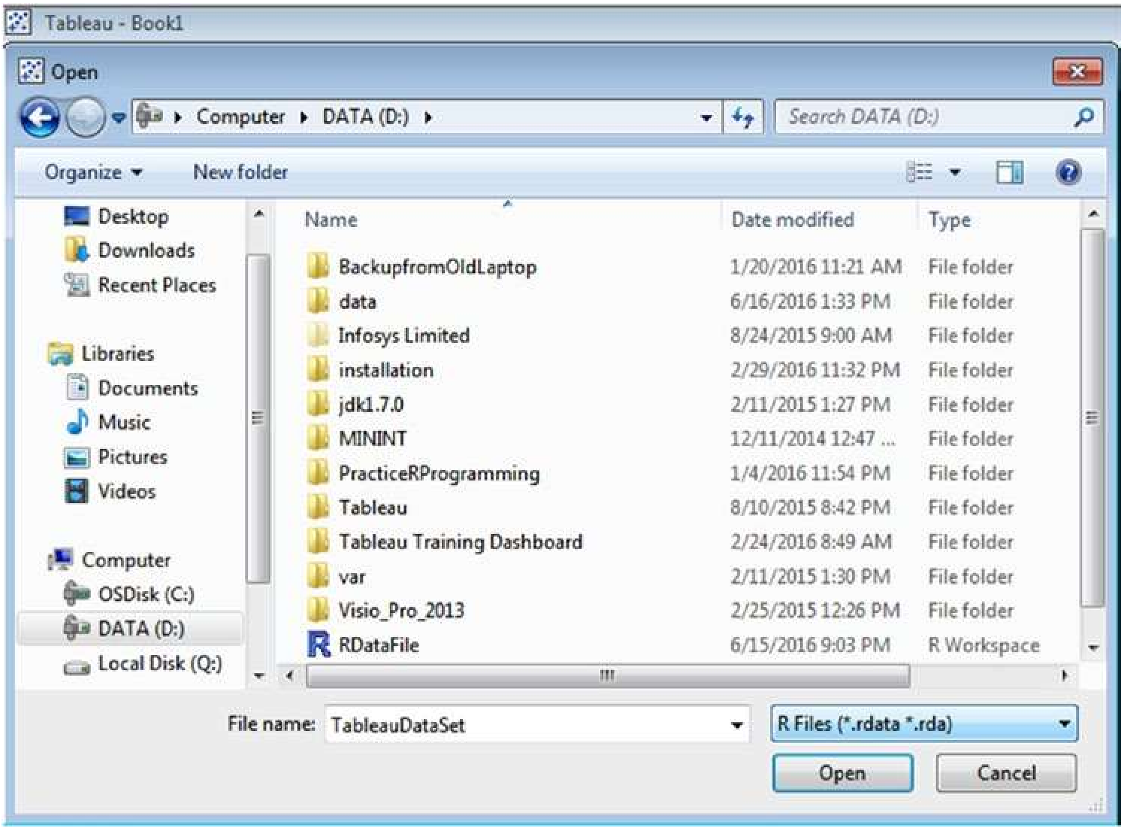


Figure 2-29. Open dialog box to open RDataFile

2.3.3.1.6 Step 5

Now, you can view the RDataset in the data source page (Shown in Fig. 2-30).

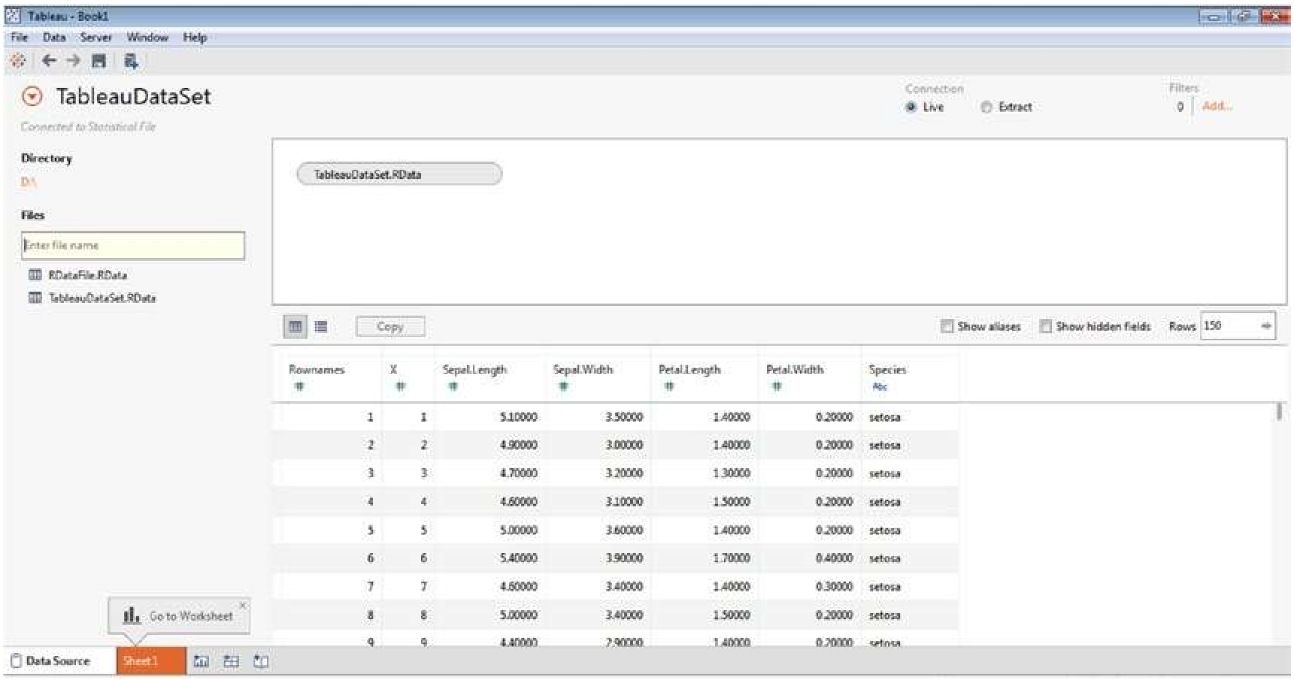


Figure 2-30. Data source page showing the connection to the statistical file

2.4 Connect to a Server

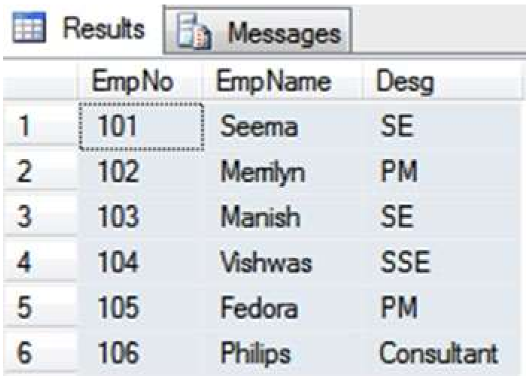
Let us explore how to connect to the below data sources:

- Microsoft SQL Server
- MySQL
- NoSQL Databases
    - MongoDB
    - Cassandra

2.4.1 Connecting to MS SQL Server 2014 Management Studio

We have a table with the name “Employee” in the “Test” Database in MS SQL Server 2014.

We have six records in the table “Employee” as displayed below (Shown in Fig. 2-31).



	EmpNo	EmpName	Desg
1	101	Seema	SE
2	102	Merrilyn	PM
3	103	Manish	SE
4	104	Vishwas	SSE
5	105	Fedora	PM
6	106	Philips	Consultant

Figure 2-31. *“Employee” Table*

The objective is to read these six records inside tableau.

Follow the steps below.

2.4.1.1 Steps to connect to MS SQL Server

2.4.1.1.1 Step 1

Open a Tableau workbook on the “Connect” page.

Click on Microsoft SQL Server under “To the server” (Shown in Fig. 2-32) to bring up the dialog box displaying the server connection to Microsoft SQL Server (Shown in Fig. 2-33).

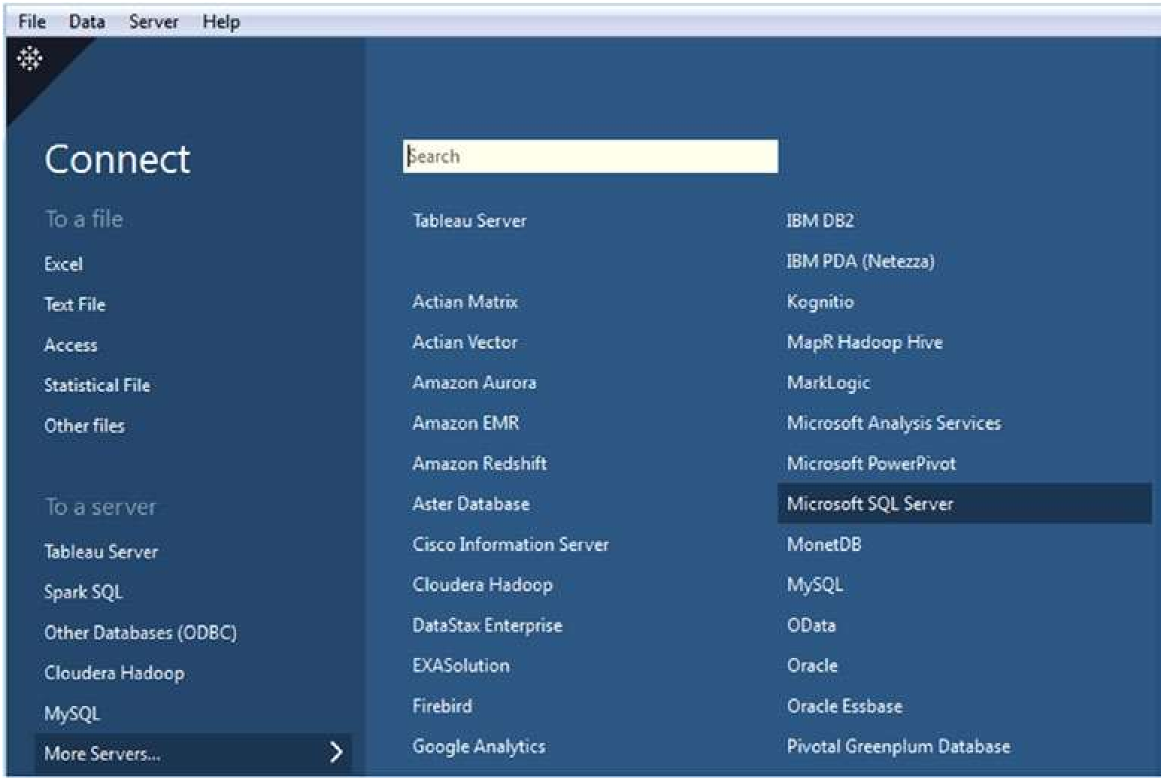


Figure 2-32. *Connect to Microsoft SQL Server*

Figure 2-33. *Microsoft SQL server connection window*

2.4.1.1.2 Step 2

Fill in the details about the database server such as "server name". Select either "Windows Authentication" or provide a specific username and password (Shown in Fig. 2-34).

Figure 2-34. *Microsoft SQL Server "Server Connection" details*

2.4.1.1.3 Step 3

If the connection is successful, it shows the screen below to allow one to select the desired database (Shown in Fig. 2-35).

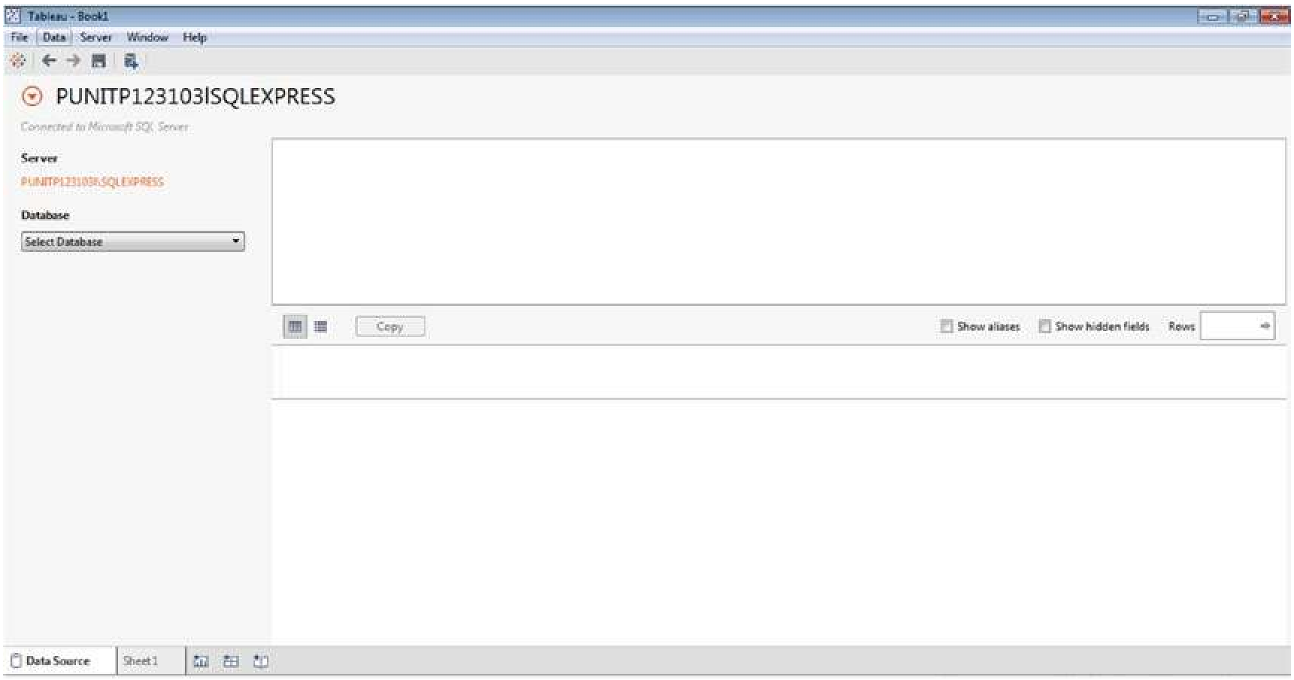


Figure 2-35. *Data Source Page showing connection to "Microsoft SQL Server"*

2.4.1.1.4 Step 4

The table "Employee" that we wish to work with is in the "Test" database (Shown in Fig. 2-36).

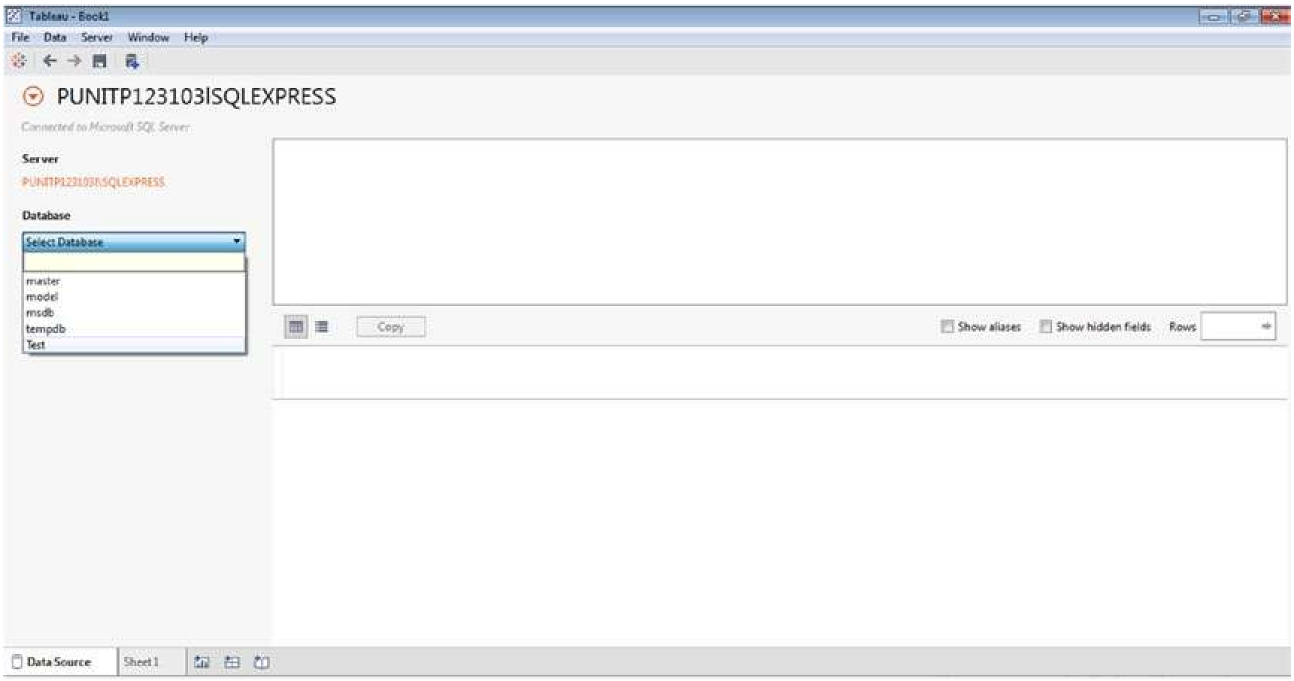


Figure 2-36. *Selection of "Test" database*

2.4.1.1.5 Step 5

Currently, there is only one table, "Employee" in the "Test" database. Drag the table to the canvas area. Select how you wish to have the records updated, either "Update Now" or "Automatically Update" (Shown in Fig. 2-37).

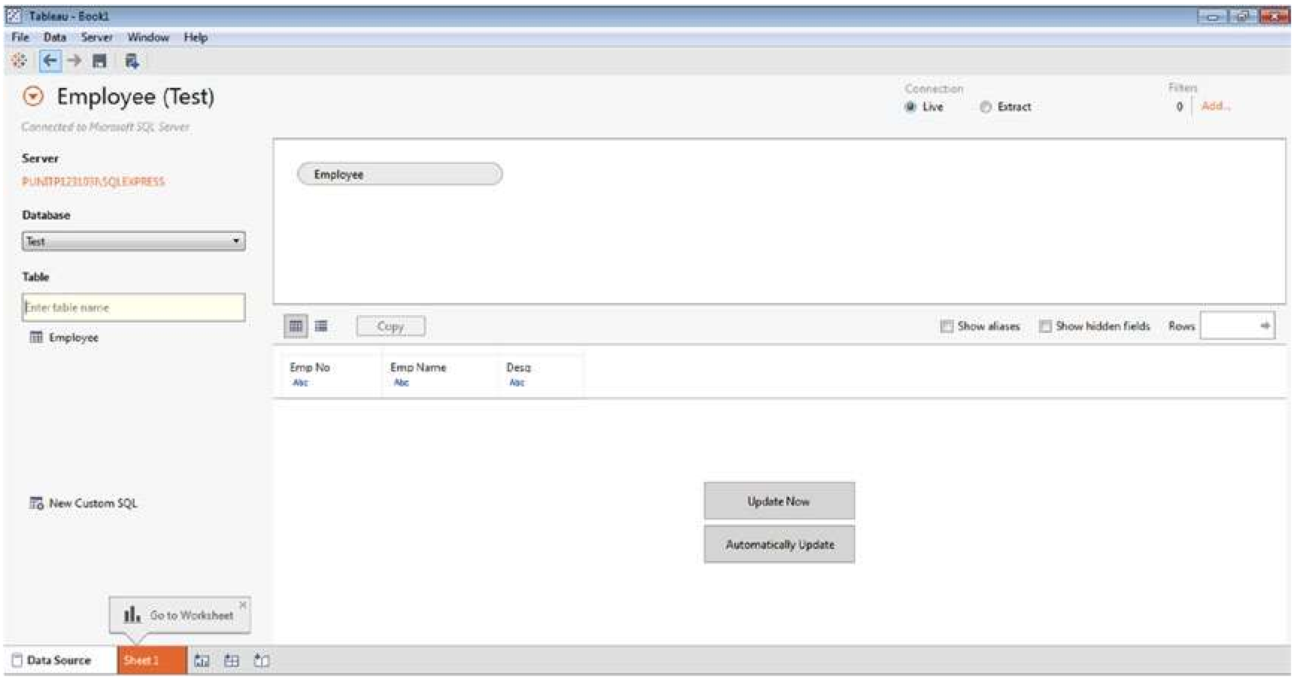


Figure 2-37. *"Employee (Test)" table placed on canvas area*

2.4.2 Connecting to MySQL

Follow the steps to connect to a MySQL Database.

2.4.2.1 Steps to connect to MySQL

2.4.2.1.1 Step 1

Download MySQL installer for windows from below mentioned link and install it.

https://dev.mysql.com/downloads/installer/

2.4.2.1.2 Step 2

Go to All Programs ➤ MySQL Server 5.7 ➤ MySQL Command Line Client (Shown in Fig. 2-38). Click on MySQL Command Line Client to start MySQL Server (Shown in Fig. 2-39).

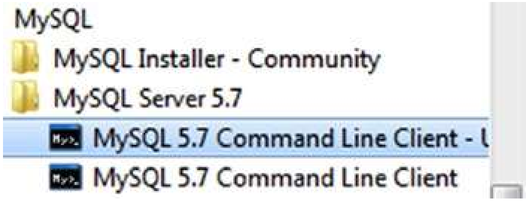


Figure 2-38. *MySQL Command Line Client*

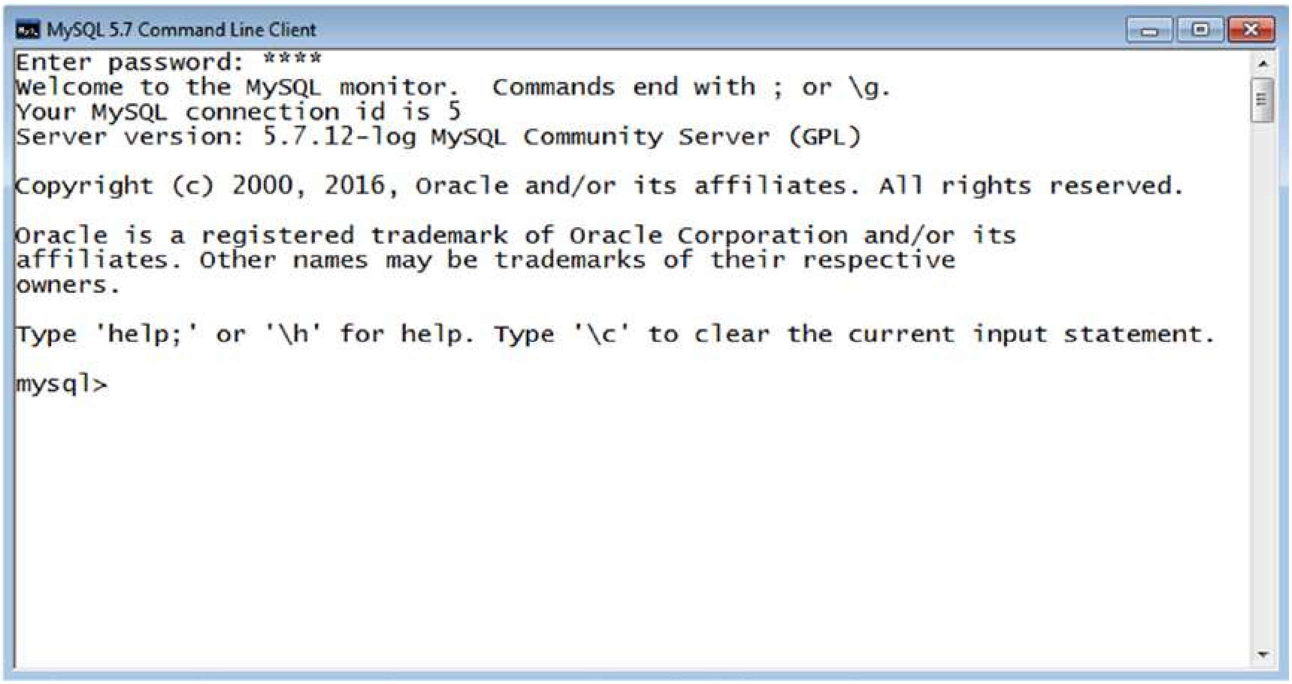


Figure 2-39. *MySQL command prompt*

2.4.2.1.3 Step 3

Download MySQL Driver for Tableau from link below and install it.

`http://www.tableau.com/support/drivers`

2.4.2.1.4 Step 4

Open Tableau Desktop, Under Connect ➤ Select MySQL (Shown in Fig. 2-40).

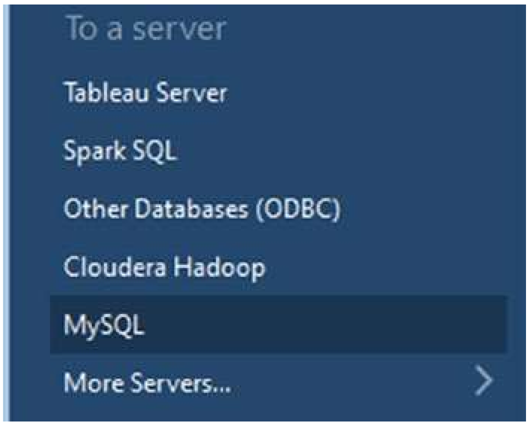


Figure 2-40. *Connect to MySQL server*

2.4.2.1.5 Step 5

MySQL connection wizard will open. Provide inputs for "Server", "Username" and "Password" and click "OK" to connect to MySQL Server (Shown in Fig. 2-41).

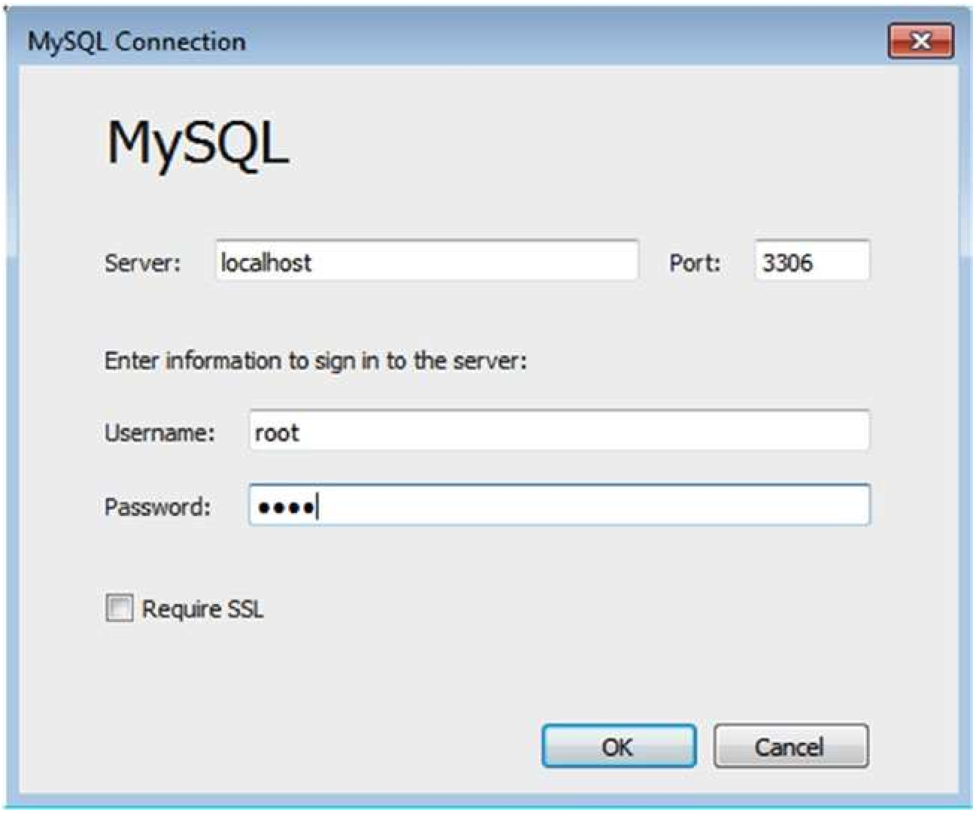


Figure 2-41. *MySQL server connection details*

2.4.2.1.6 Step 6

If the connection is successful, it will show the screen below to allow one to select the desired database (Shown in Fig. 2-42).

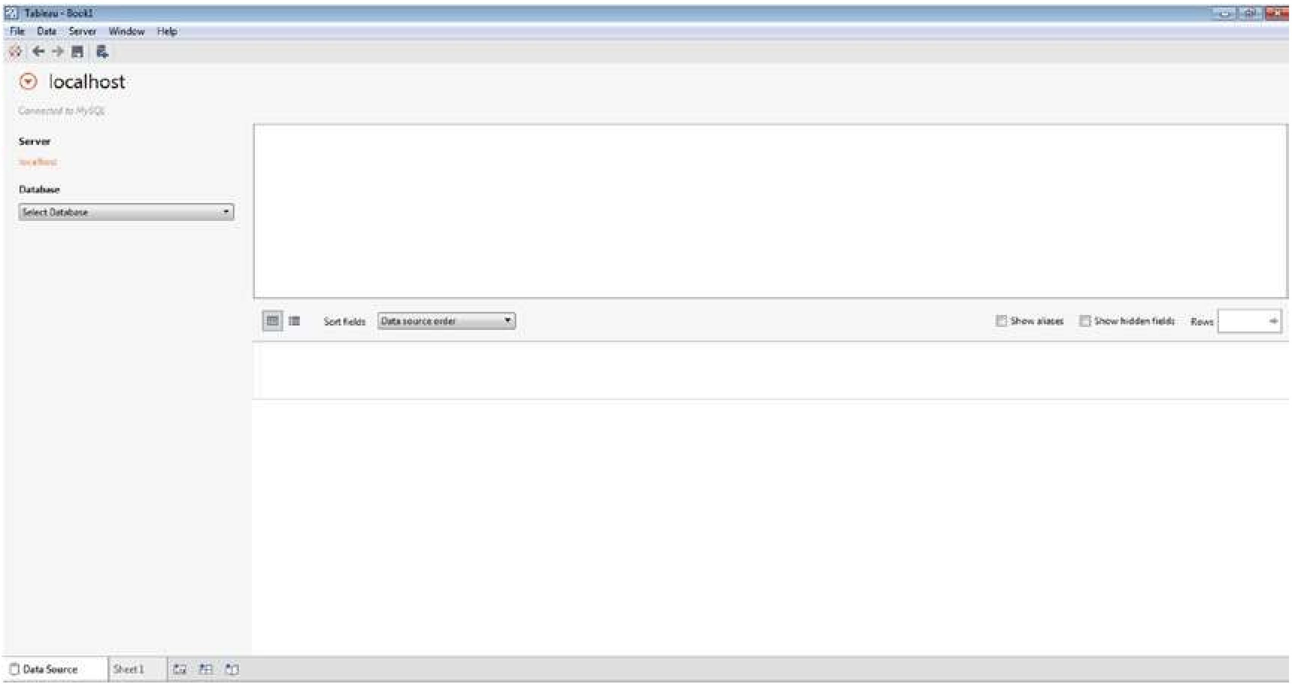


Figure 2-42. *Data source page showing a successful MySQL server connection*

2.4.3 Connecting to NoSQL Databases

Let us discuss how to connect to NoSQL databases such as Cassandra and MongoDB.

2.4.3.1 Connecting to Cassandra

Follow the below steps to connect to a Cassandra NoSQL Database.

2.4.3.1.1 Steps to connect to Cassandra NoSQL database

2.4.3.1.1.1 Step 1

Download DataStax Community Edition for Windows from the below-mentioned link and install it.

`https://downloads.datastax.com/community/`

Java 1.8 is required to work with Apache Cassandra.

2.4.3.1.1.2 Step 2

Select All Programs ➤ DataStax Community Edition ➤ Cassandra CQL Shell (Shown in Fig. 2-43) to start Cassandra CQL Shell (Shown in Fig. 2-44).

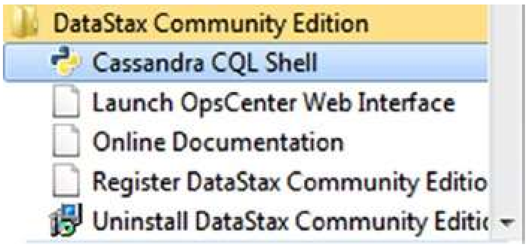


Figure 2-43. *Cassandra CQL Shell option*

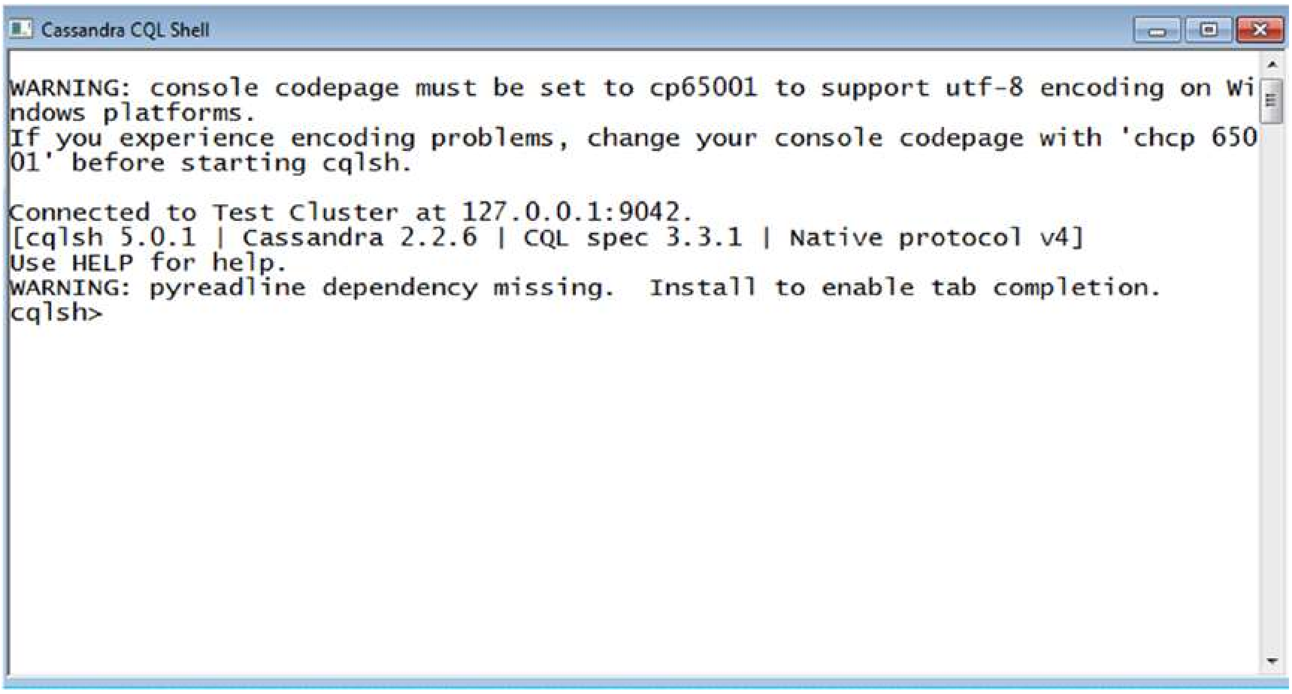


Figure 2-44. *Cassandra CQL shell*

2.4.3.1.1.3 Step 3

Download Cassandra ODBC and JDBC drivers with SQL connector from the below link and install it.

http://www.simba.com/drivers/cassandra-odbc-jdbc/

You can download 30 days trial version, SimbaApacheCassandraDriver.lic file which will be sent to your registered email.

2.4.3.1.1.4 Step 4

To check Cassandra ODBC and JDBC driver installation, Select All Programs ➤ Simba Cassandra ODBC Driver 2.2 ➤ 64 bit ODBC Driver (Shown in Fig. 2-45).

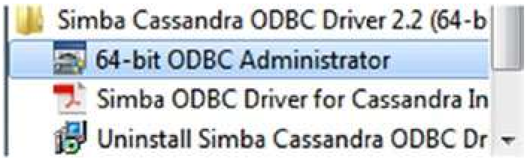


Figure 2-45. *Cassandra ODBC Driver option*

2.4.3.1.1.5 Step 5

Select "System DSN" to see "Simba Cassandra ODBC DSN" (Shown in Fig. 2-46).

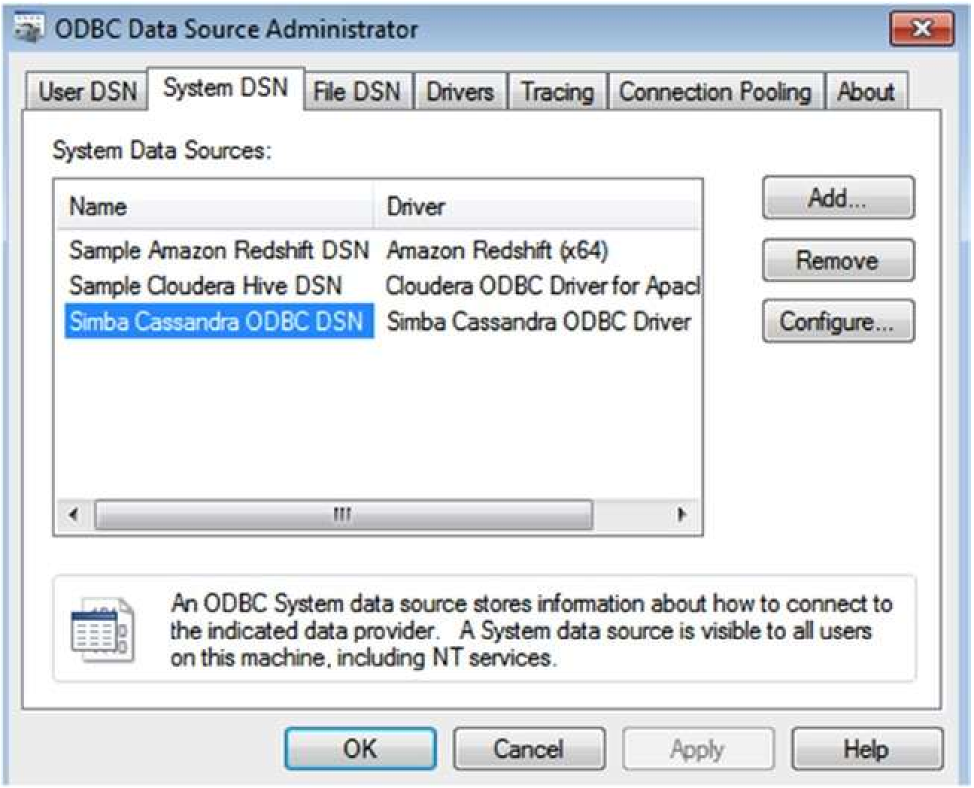


Figure 2-46. *Simba Cassandra OBDC DSN (highlighted)*

Copy SimbaApacheCassandraDriver.lic file to the lib folder of SimbaCassandraODBC Driver. You can find SimbaCassandraODBC Driver folder inside the Program Files.

2.4.3.1.1.6 Step 6

Click on the "Configure" button to configure the Simba Cassandra ODBC DSN (Shown in Fig. 2-47). Specify the host as 127.0.0.1 and click "Test" to check the connectivity (Shown in Fig. 2-48). Once the connection comes through click on the "OK" button.

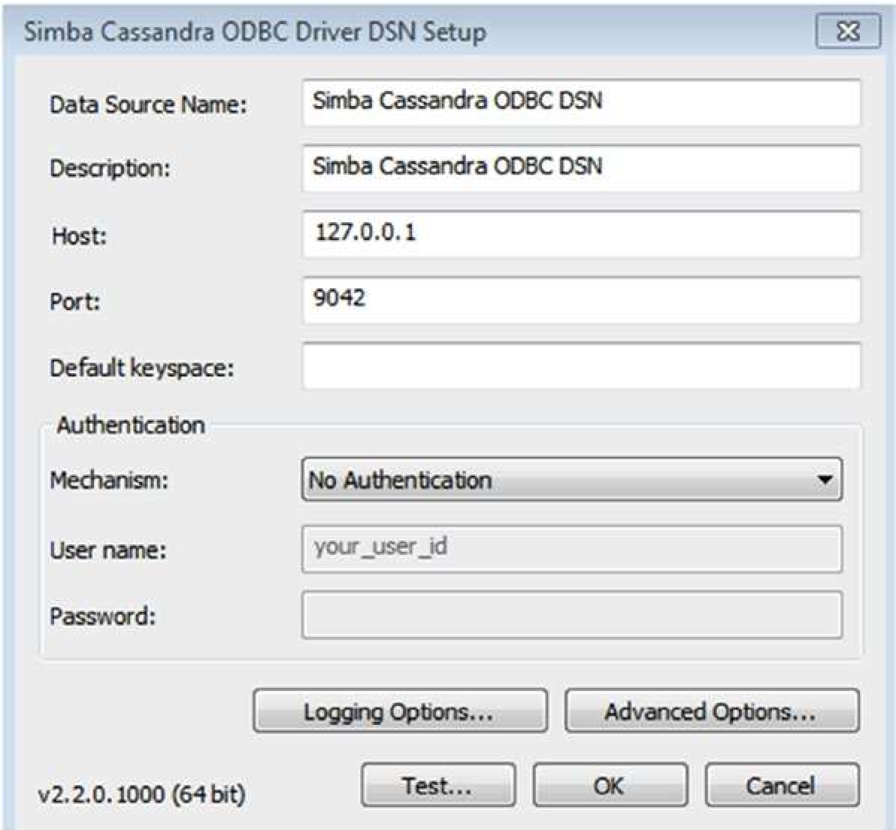


Figure 2-47. Simba Cassandra ODBC driver DSN setup

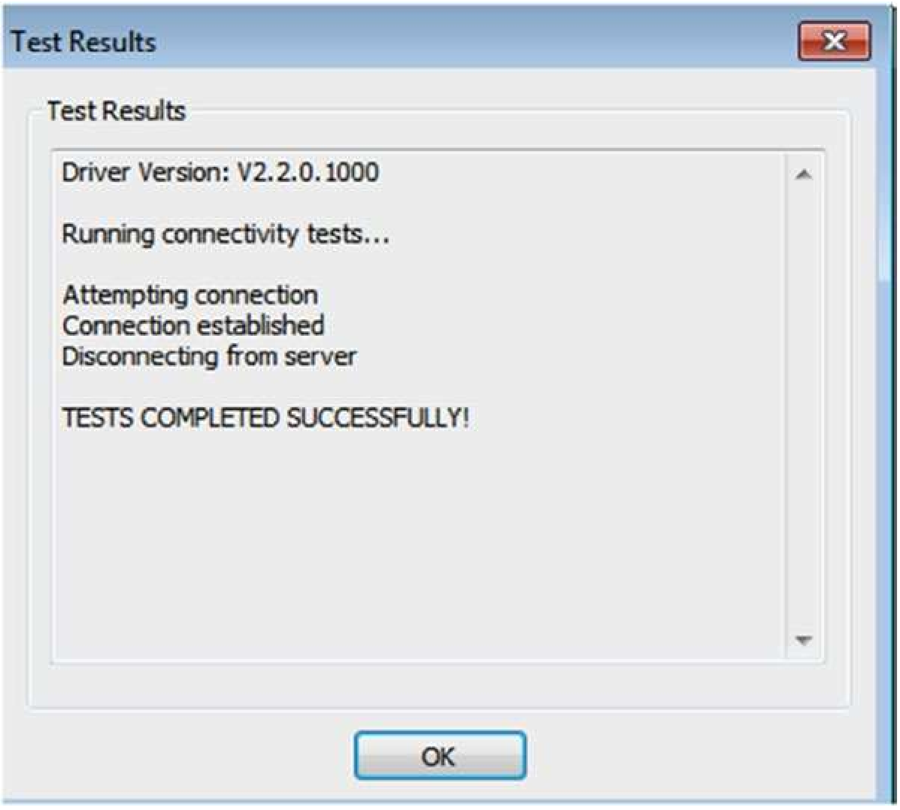


Figure 2-48. *Connection "Test Results" window*

2.4.3.1.1.7 Step 7

Open Tableau Desktop, From Connect, Select More Servers ➤ Other Databases (ODBC) (Shown in Fig. 2-49).

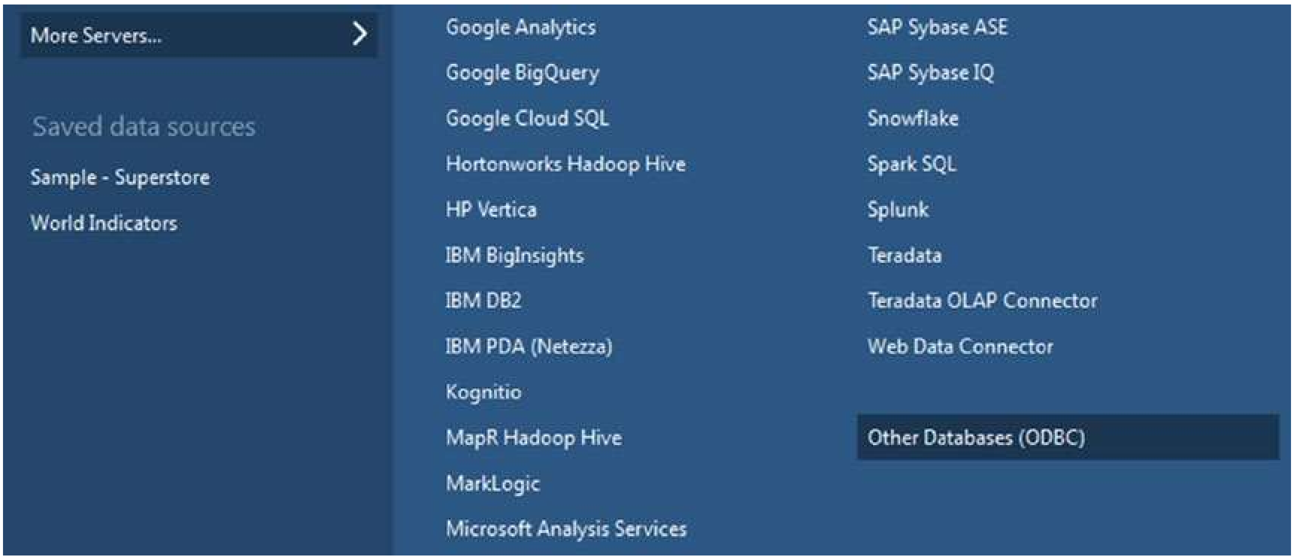


Figure 2-49. *Connect to "Other Databases (ODBC)"*

2.4.3.1.1.8 Step 8

Other Databases(ODBC) Connection Wizard shows up. Select Simba Cassandra ODBC DSN (Shown in Fig. 2-50) and click on “Connect” button. Next, click on “OK” button to connect to the Cassandra database.

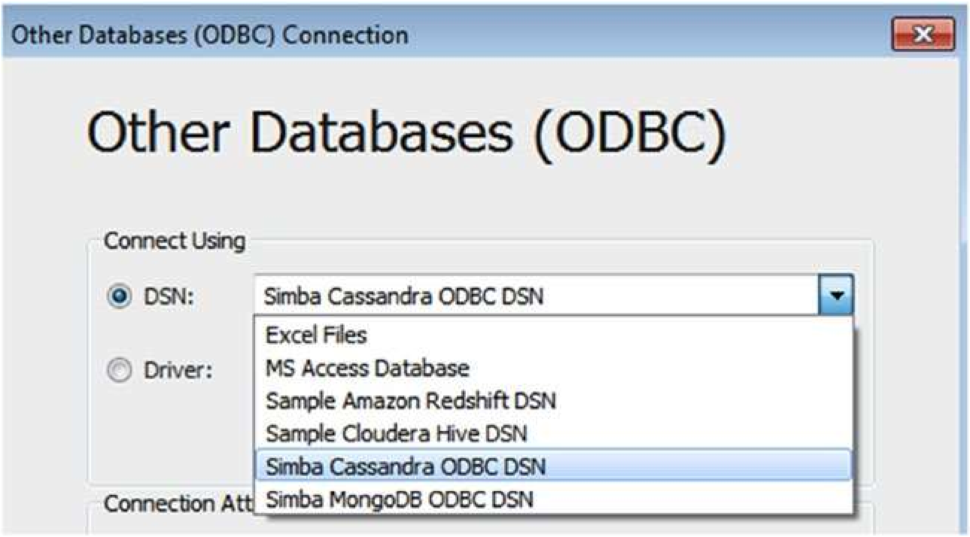


Figure 2-50. *Other Databases (ODBC) connection window*

2.4.3.1.1.9 Step 9

If the connection is successful, it shows the screen below to allow one to select the desired database (Shown in Fig. 2-51).

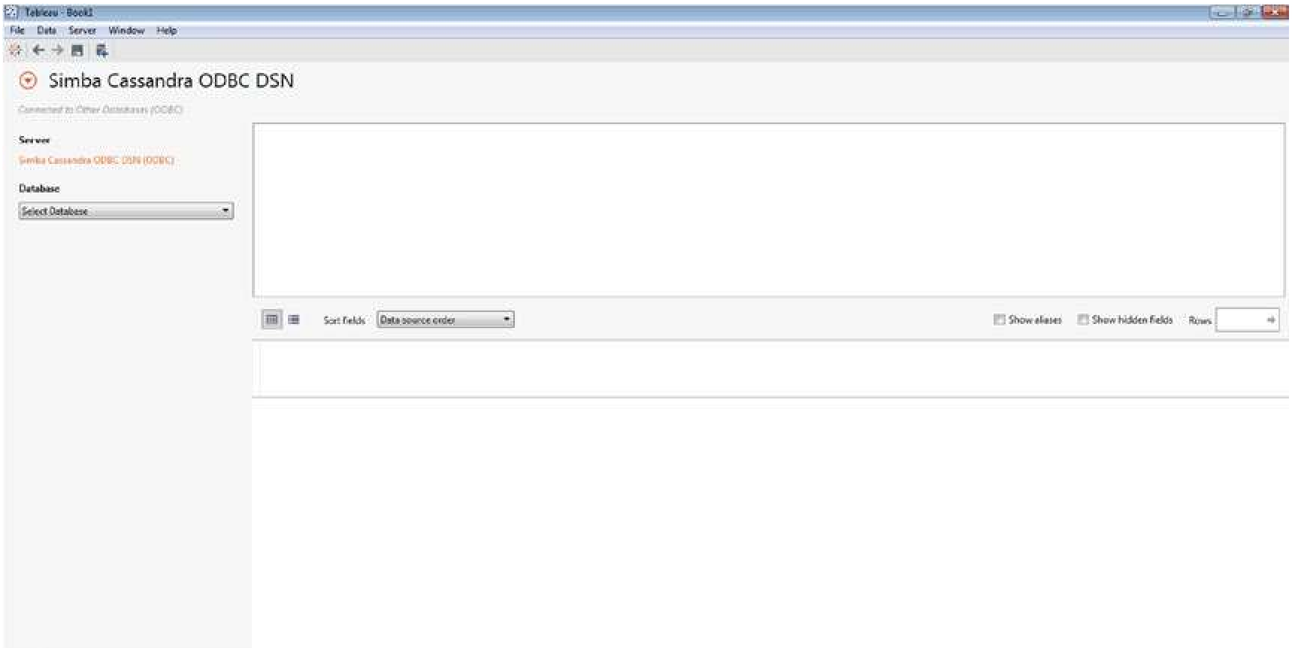


Figure 2-51. *Data source page showing the Cassandra connection*

2.4.3.2 Connecting to MongoDB

Follow the below steps to connect to a MongoDB NoSQL Database.

2.4.3.2.1 Steps to connect to MongoDB NoSQL Database

2.4.3.2.1.1 Step 1

Download MongoDB for Windows from the below-mentioned link and install it.

https://www.mongodb.com/download-center#community

2.4.3.2.1.2 Step 2

To start the MongoDB Server, open the command prompt and navigate to the installation folder of MongoDB as shown in Fig. 2-52.

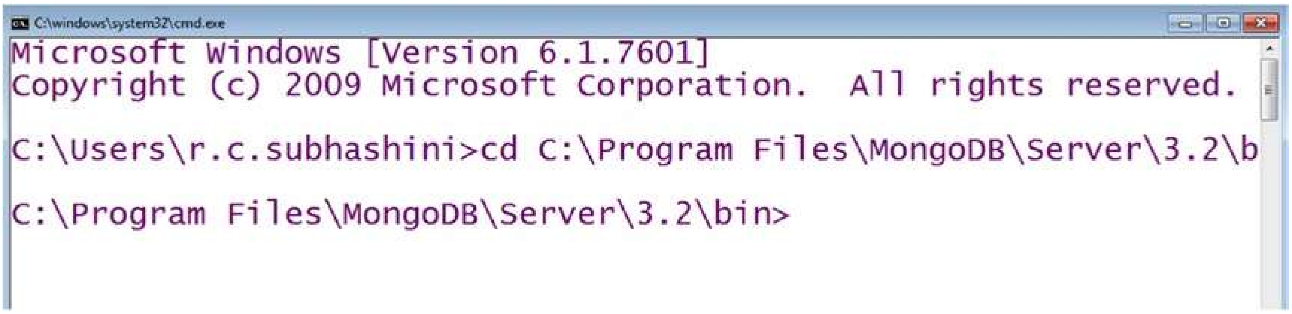


Figure 2-52. *MongoDB installation directory path*

2.4.3.2.1.3 Step 3

Type "mongod.exe" to start the server (Shown in Fig. 2-53).

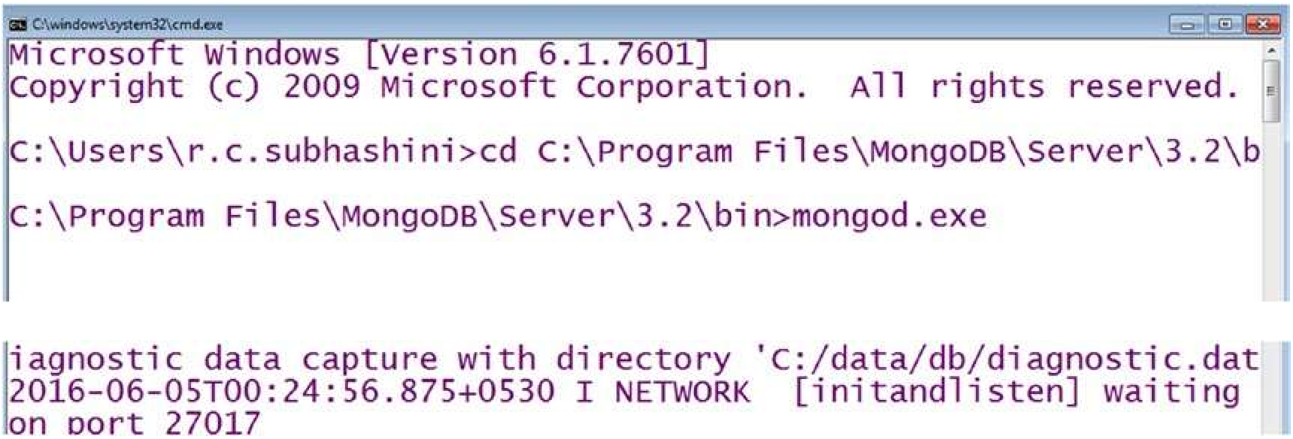


Figure 2-53. *Starting MongoDB Server*

You should get a message stating "waiting on port 27017..."

2.4.3.2.1.4 Step 4

To start the MongoDB Client, open the command prompt and navigate to the installation folder of MongoDB and type mongo.exe as shown in Fig. 2-54.

```
C:\Users\r.c.subhashini>cd C:\Program Files\MongoDB\Server\3.2\b

C:\Program Files\MongoDB\Server\3.2\bin>mongo.exe
2016-06-05T00:28:40.458+0530 I CONTROL  [main] Hotfix KB2731284
is not installed, will zero-out data files
MongoDB shell version: 3.2.3
connecting to: test
>
```

Figure 2-54. *Starting MongoDB Client*

2.4.3.2.1.5 Step 5

Download the MongoDB ODBC and JDBC drivers with SQL connector from the below link and install it.

http://www.simba.com/drivers/mongodb-odbc-jdbc/

You can download 30 days trial version, you will receive the SimbaMongoDBODBDriver.lic file in your registered email.

2.4.3.2.1.6 Step 6

To check MongoDB ODBC and JDBC driver installation, select All Programs ➤ Simba MongoDB ODBC Driver 2.0 ➤ 64 bit ODBC Driver (Shown in Fig. 2-55).

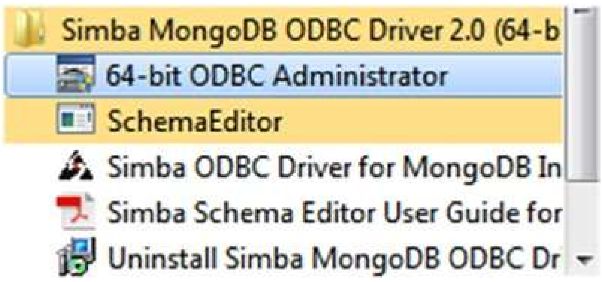


Figure 2-55. *Simba MongoDB ODBC Driver*

2.4.3.2.1.7 Step 7

Select “System DSN” to see “Simba MongoDB ODBC DSN” (Shown in Fig. 2-56).

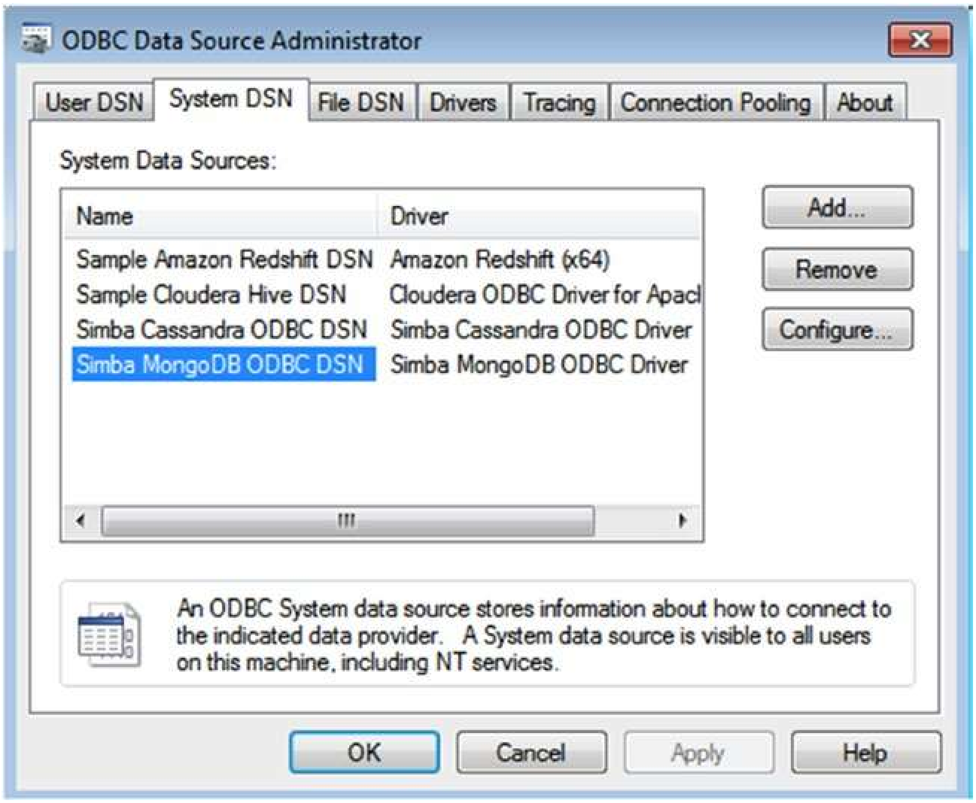


Figure 2-56. *“System DSN” tab showing “Simba MongoDB ODBC DSN”*

Copy SimbaMongoDBODBDriver.lic file to the lib folder of SimbaMongoDBODBC Driver. You can find SimbaMongoDBODBC Driver folder inside the Program Files.

2.4.3.2.1.8 Step 8

Click “Configure...” button to configure the “Simba MongoDB ODBC DSN”. Specify the server as “localhost”, port as “27017” and the database as “test” (default database of MongoDB) (Shown in Fig. 2-57) and click “Test” to check the connectivity (Shown in Fig. 2-58). Once the connection is successful, click on the “OK” button.

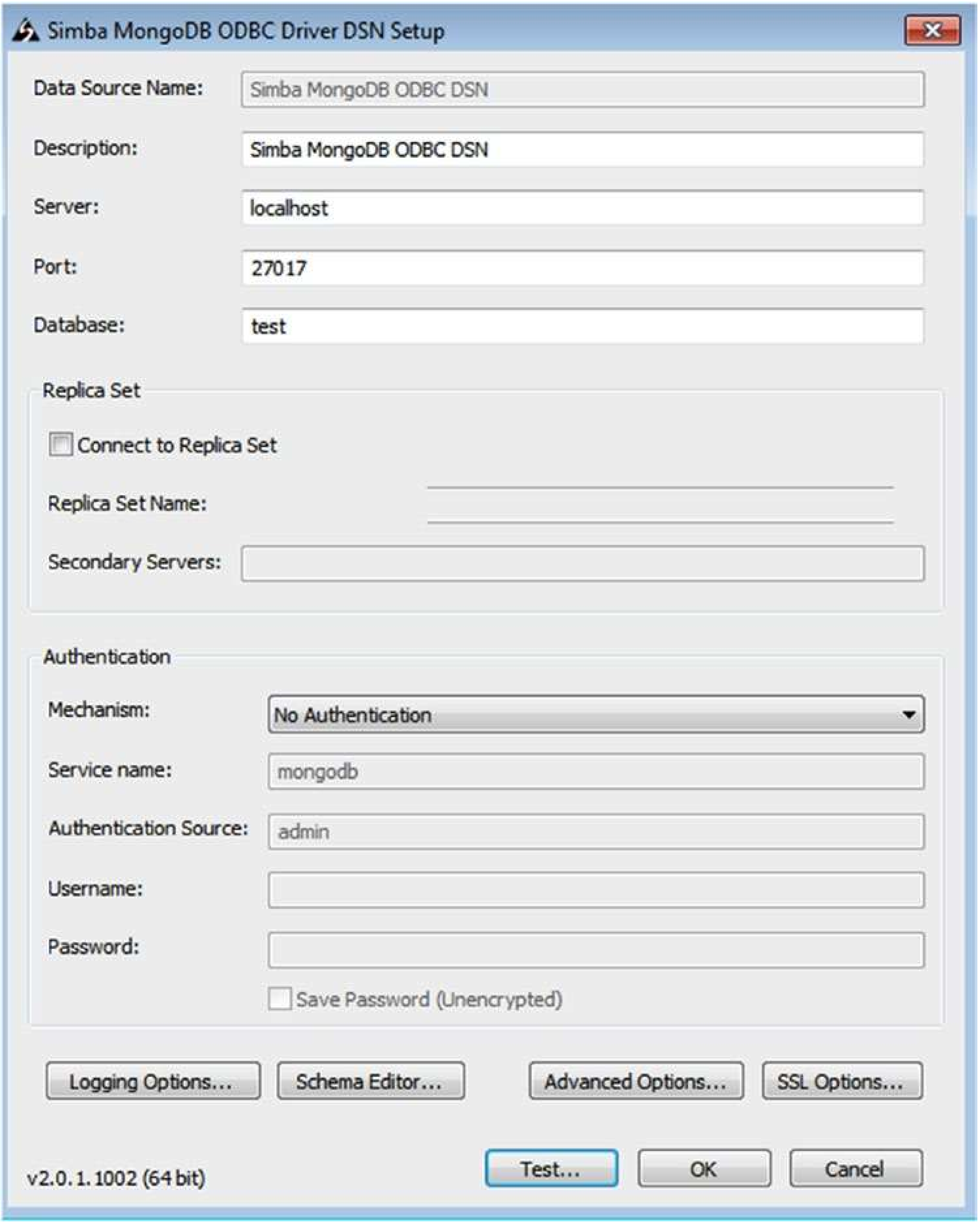


Figure 2-57. *Simba MongoDB ODBC Driver DSN Setup*

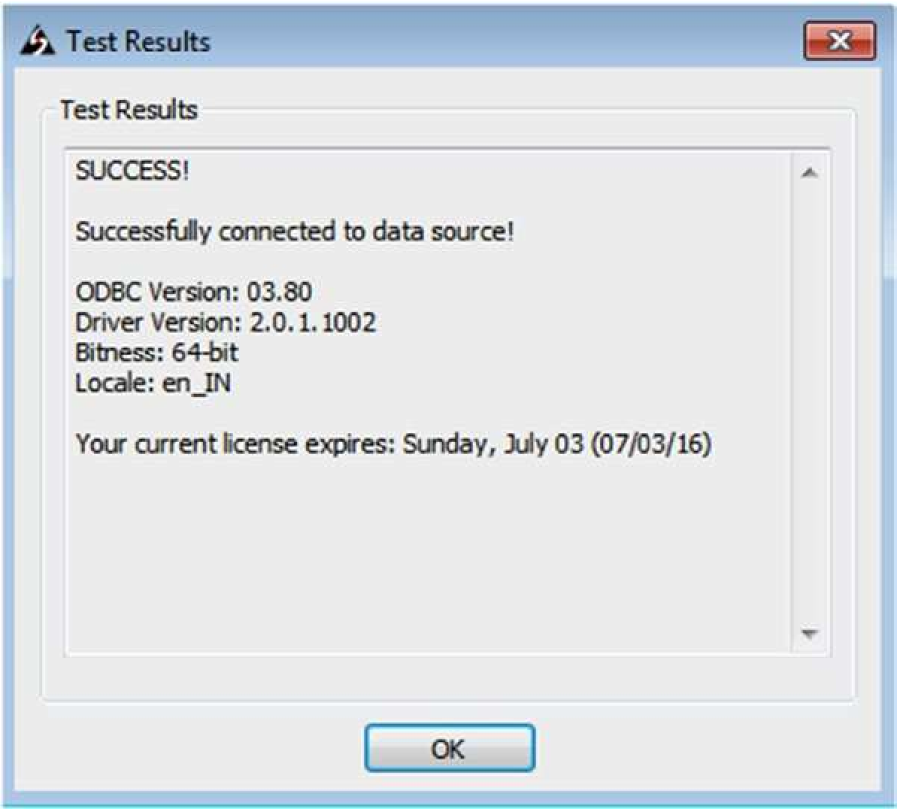


Figure 2-58. Connectivity test results

2.4.3.2.1.9 Step 9

Open Tableau Desktop, From Connect, Select More Servers ➤ Other Databases (ODBC) (Shown in Fig. 2-59).

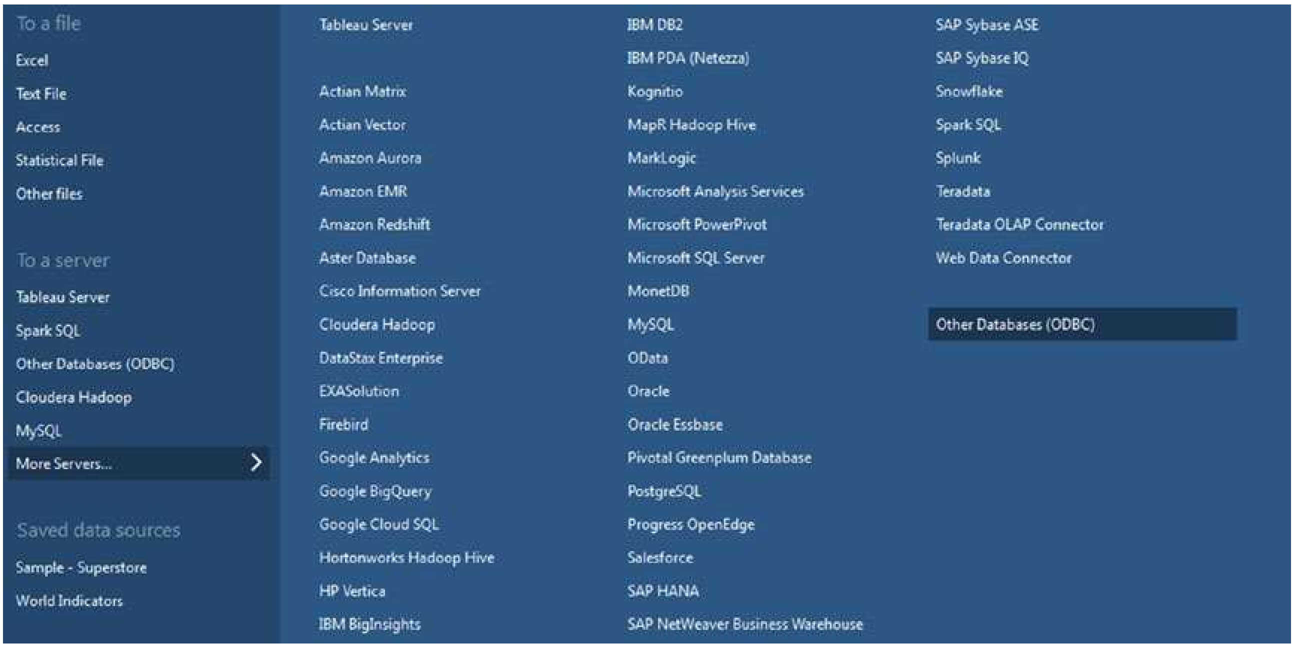


Figure 2-59. Connection to "Other Databases (ODBC)"

2.4.3.2.1.10 Step 10

Other Databases (ODBC) Connection Wizard will show up. Select “Simba MongoDB ODBC DSN” and click on the “Connect” button (Shown in Fig. 2-60). Next, click on the “OK” button to connect to MongoDB Database.

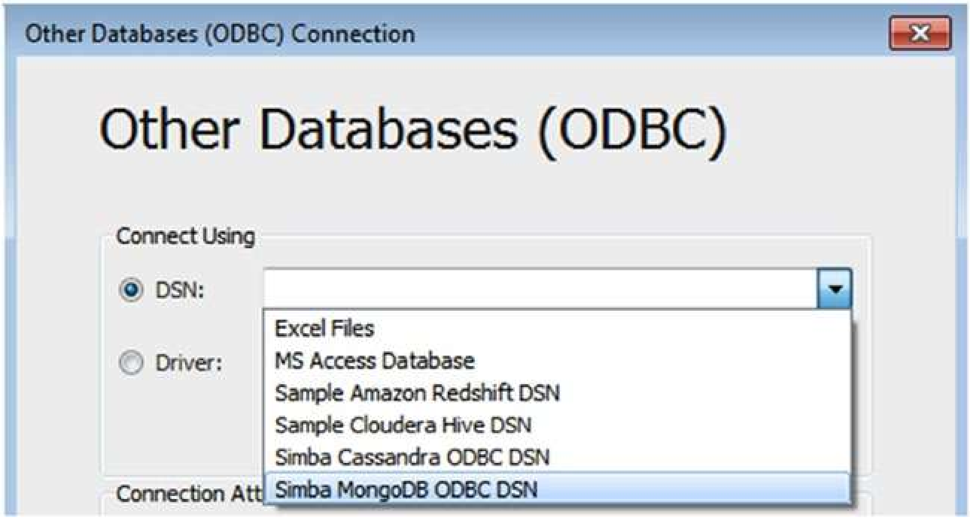


Figure 2-60. *Selection of “Simba MongoDB ODBC DSN”*

2.4.3.2.1.11 Step 11

If the connection is successful, it shows the screen below to allow one to select the desired database (Shown in Fig. 2-61).

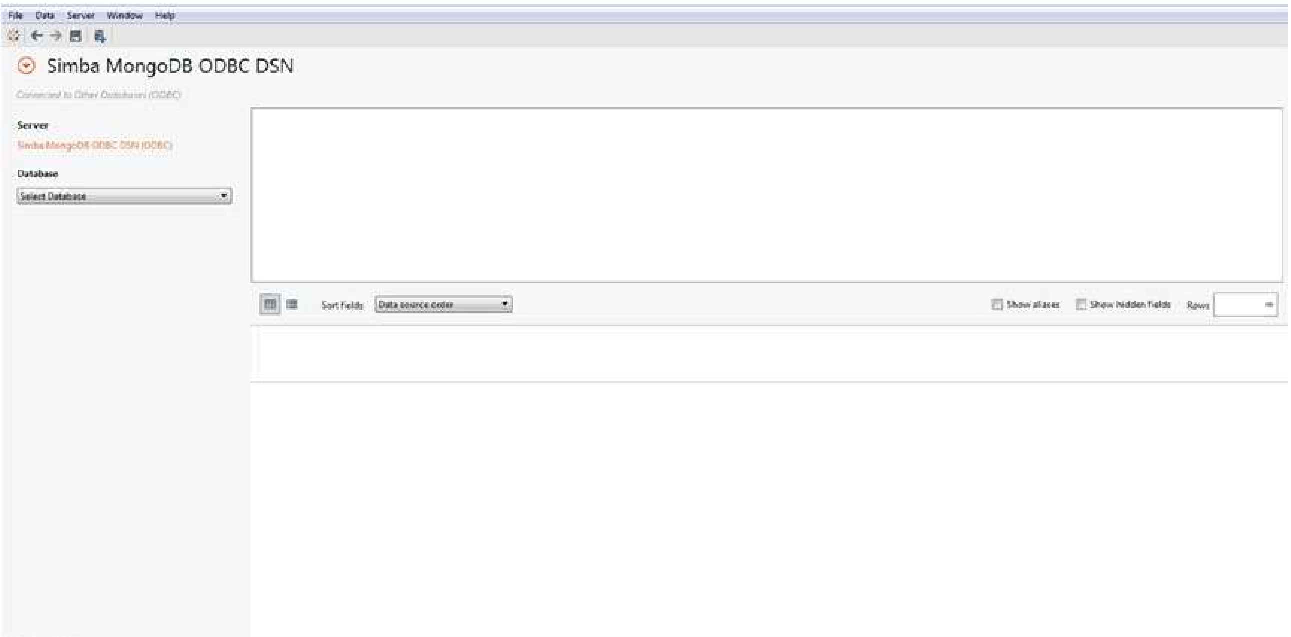


Figure 2-61. *Data source page showing a MongoDB connection*

2.5 Metadata Grid

Click on the metadata grid icon (Shown in Fig. 2-62) to open the metadata grid. Metadata grid displays the fields that are available in the Data Source. Metadata grid helps one to analyse the structure of Tableau data source (Shown in Fig. 2-63), to rename or hide fields at once (Shown in Fig. 2-64), etc.

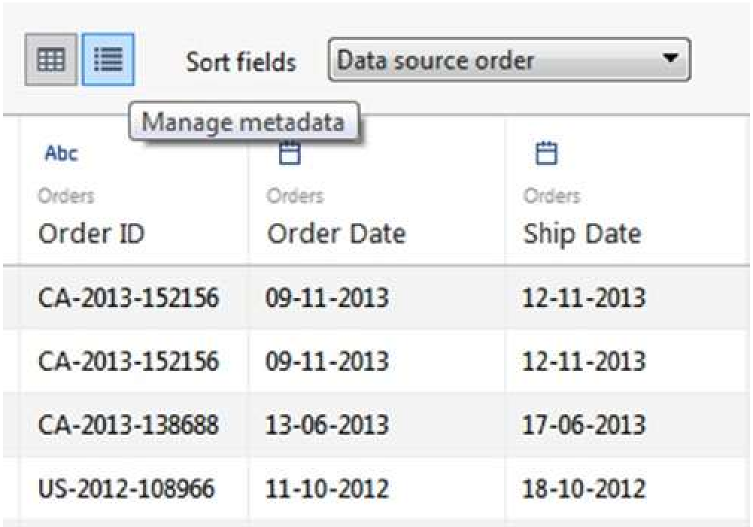


Figure 2-62. *Metadata Grid Icon*

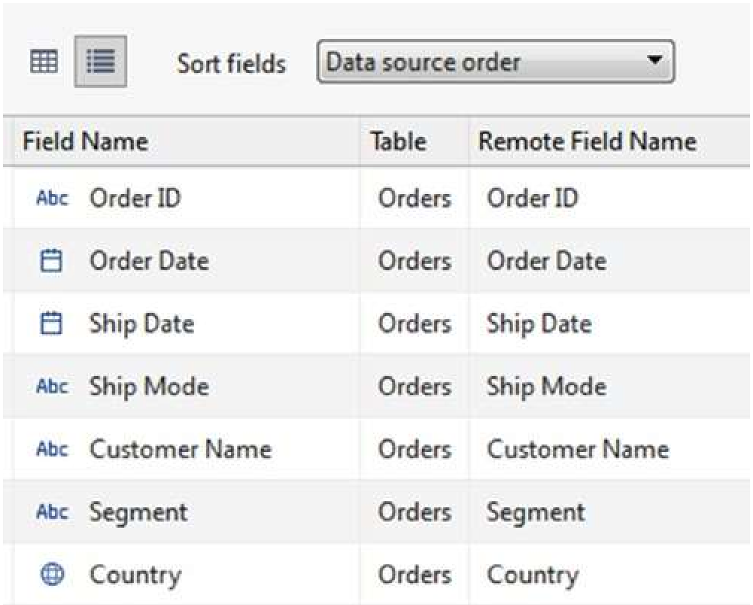


Figure 2-63. *Metadata grid displaying data source fields*

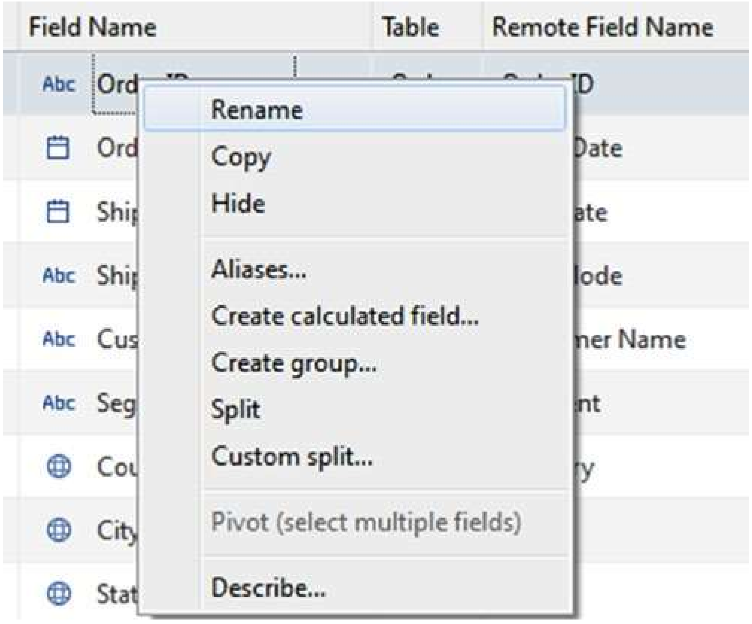


Figure 2-64. *Options to "Rename" or "Hide" field and other options*

2.6 Joins

Relational data source contains collections of tables and tables are related by specific field. For example, let us consider the schema design for a blog website. It contains a table for recording blog entries such as blog id, blog tile, description, URL, likes, and posted by, etc. In addition to this, there could be another table to store details of comments such as blog id, comment id, by user, likes and comments. To analyze and answer questions such as which blog contains the highest likes, you will need to join two tables using a common field such as blog id.

Once you establish a connection, you can use the data source page to connect to multiple tables, and specify joins to perform your analysis.

Tableau Desktop supports different types of joins such as inner, left, right and full outer.

2.6.1 Adding Fields to the Data Pane

You can add or edit a table to add or remove a field, modify join operation from the data pane to specify how your data should look for your analysis.

Follow the steps to add a field to a table.

2.6.1.1 Steps to add a field to a table

2.6.1.1.1 Step 1

Select a data source on the data pane and then right click to get the "Edit data source..." option (Shown in Fig. 2-65).

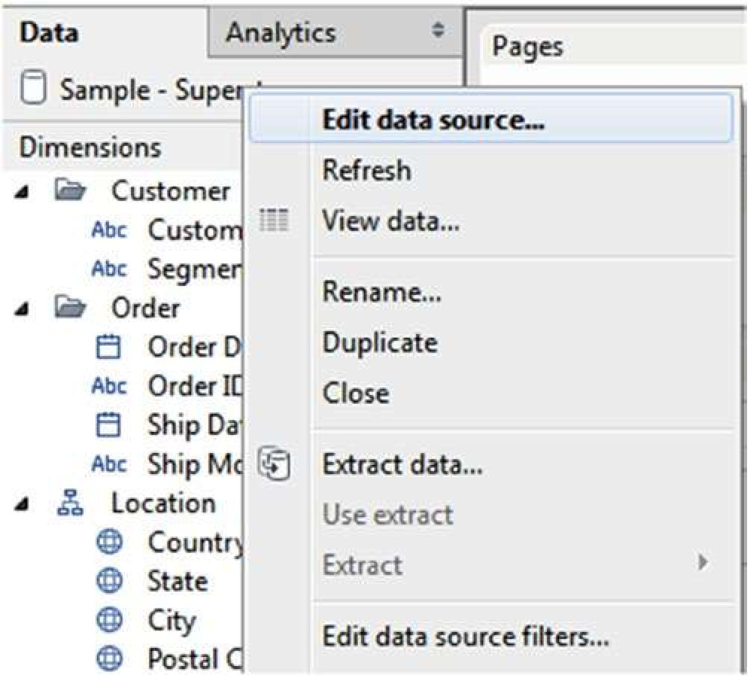
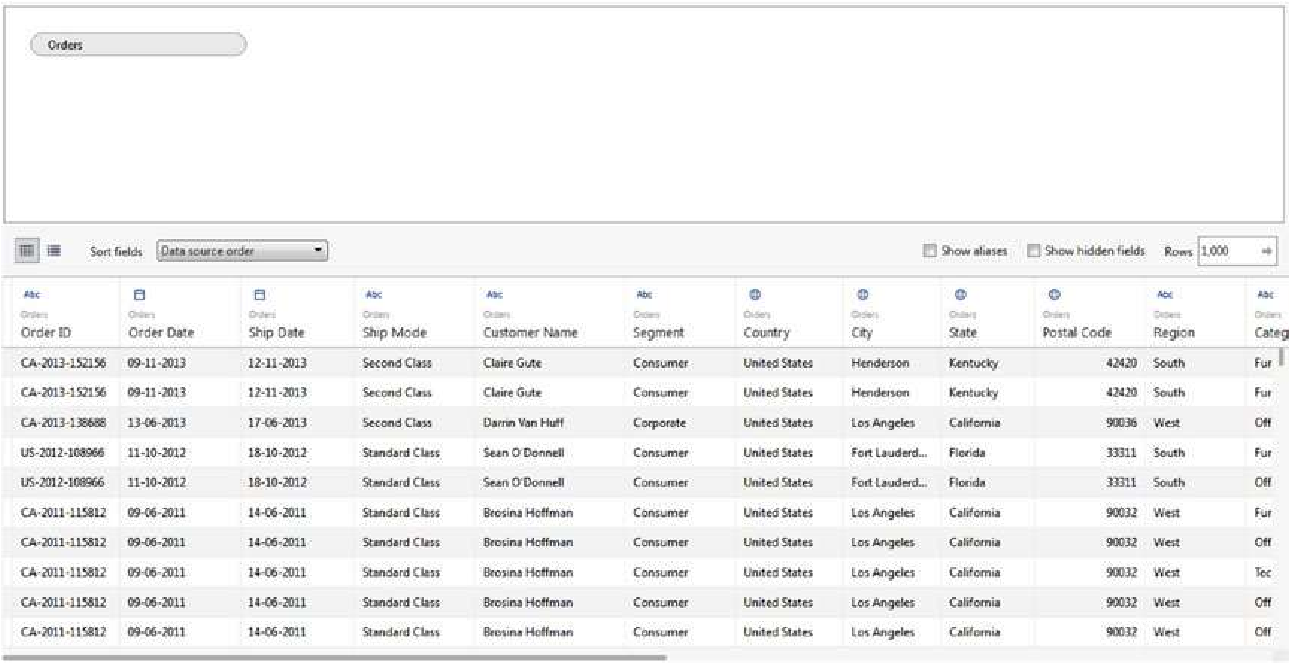


Figure 2-65. "Edit data source..." option

2.6.1.1.2 Step 2

You will be able to view the data source page. Drag the required table to the canvas area to perform the join operation (Shown in Fig. 2-66 and Fig. 2-67).



Order ID	Order Date	Ship Date	Ship Mode	Customer Name	Segment	Country	City	State	Postal Code	Region	Categ
CA-2013-152156	09-11-2013	12-11-2013	Second Class	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	Fur
CA-2013-152156	09-11-2013	12-11-2013	Second Class	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	Fur
CA-2013-138688	13-06-2013	17-06-2013	Second Class	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036	West	Off
US-2012-108966	11-10-2012	18-10-2012	Standard Class	Sean O'Donnell	Consumer	United States	Fort Lauderd...	Florida	33311	South	Fur
US-2012-108966	11-10-2012	18-10-2012	Standard Class	Sean O'Donnell	Consumer	United States	Fort Lauderd...	Florida	33311	South	Off
CA-2011-115812	09-06-2011	14-06-2011	Standard Class	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	Fur
CA-2011-115812	09-06-2011	14-06-2011	Standard Class	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	Off
CA-2011-115812	09-06-2011	14-06-2011	Standard Class	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	Tec
CA-2011-115812	09-06-2011	14-06-2011	Standard Class	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	Off
CA-2011-115812	09-06-2011	14-06-2011	Standard Class	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	Off

Figure 2-66. "Orders" sheet placed on canvas area

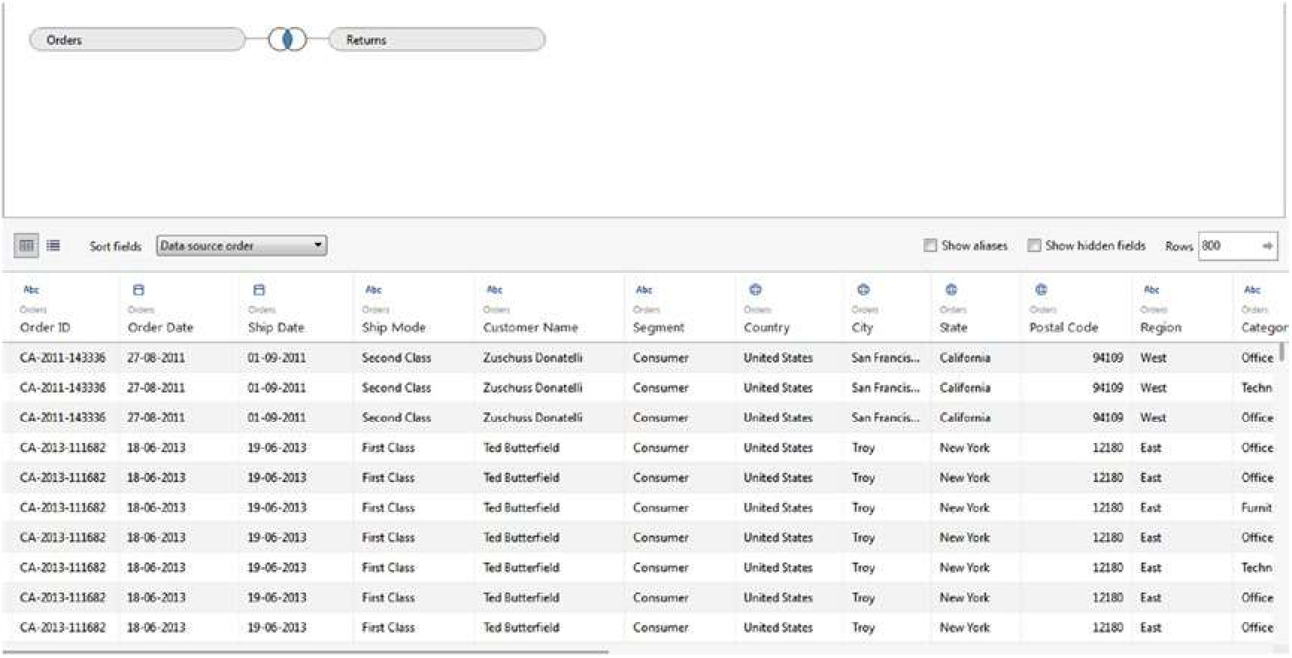


Figure 2-67. *"Returns" sheet placed on canvas area*

2.6.1.1.3 Step 3

Click on join icon to select / edit the type of join operation (Shown in Fig. 2-68).

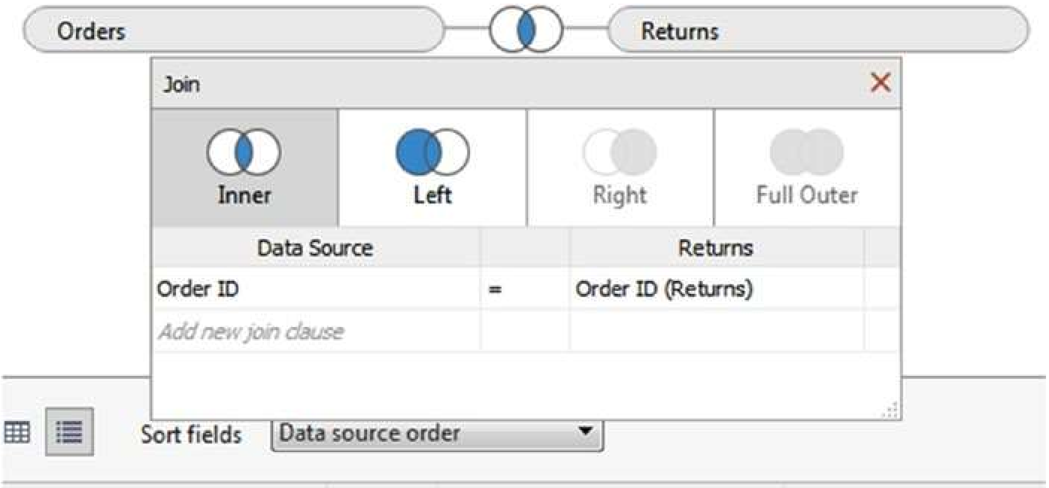


Figure 2-68. *Different types of join operation*

2.6.2 Exploring different types of Join

Let us understand how to perform the Inner Join, Left Join, etc. in Tableau using two tables namely, "Student" and "Grade".

Follow the steps to perform the join operation between two data sources in Tableau.

2.6.2.1 Steps

Follow the below steps.

2.6.2.1.1 Step 1

Create an Excel File and name it as “Sample – Student”. Student sheet contains details about students as shown below (Shown in Fig. 2-69).

	A	B
1	StudNo	Student Name
2	1001	John
3	1002	Jack
4	1003	Smith
5	1004	Joshi

Figure 2-69. *Student sheet showing details about students*

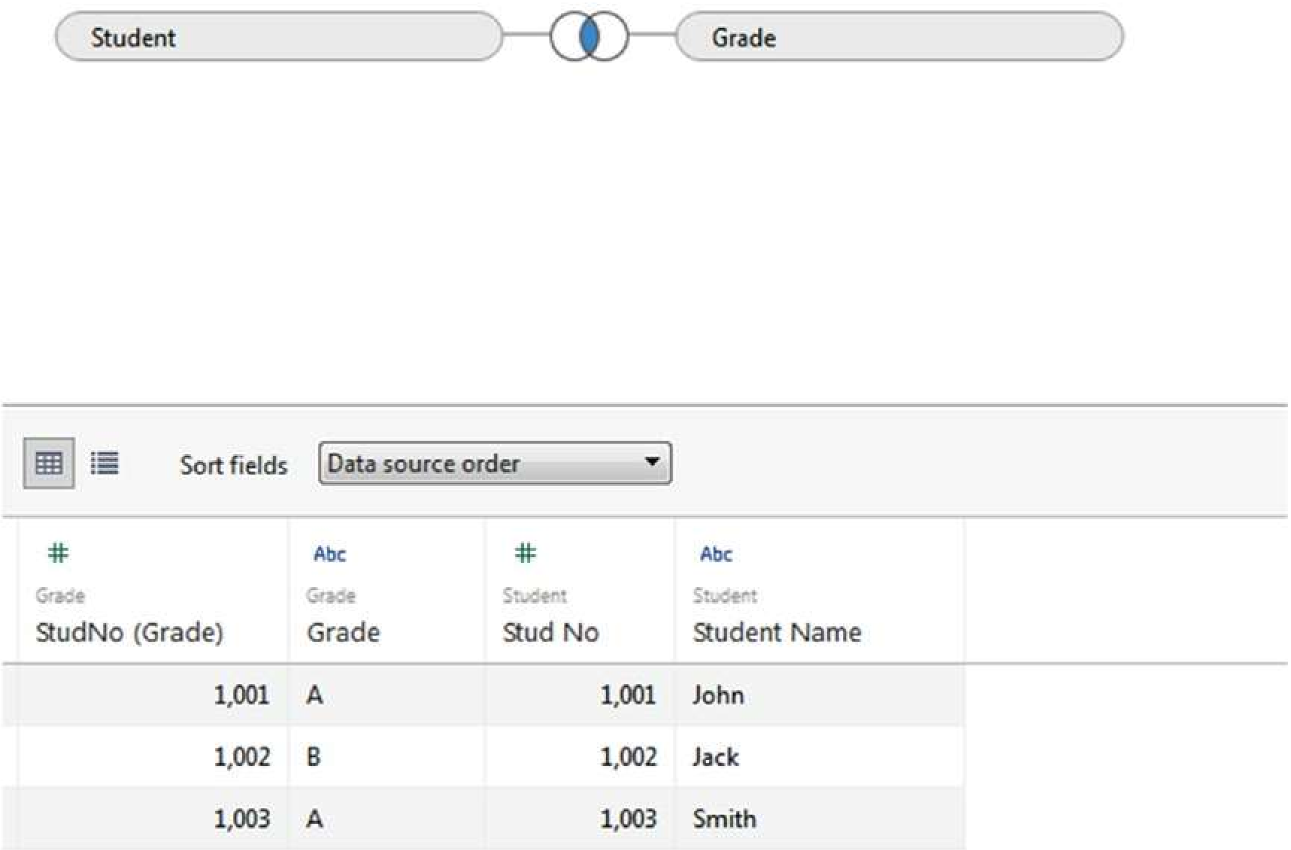
Grade Sheet contains the below details (Shown in Fig. 2-70).

	A	B
1	StudNo	Grade
2	1001	A
3	1002	B
4	1003	A
5	1006	C
6	1007	A
7	1008	B

Figure 2-70. *Grade sheet showing details about the grades scored by students*

2.6.2.1.2 Step 2

Connect to “Sample-Student”. In the Data Source Page, drag and drop the “Student” and “Grade” Sheets (Shown in Fig. 2-71).



Grade StudNo (Grade)	Grade Grade	Student Stud No	Student Student Name
1,001	A	1,001	John
1,002	B	1,002	Jack
1,003	A	1,003	Smith

Figure 2-71. *"Student" and "Grade" tables placed on the canvas*

2.6.2.2 Inner Join

Inner Join fetches all records from one table having a matching entry in the second table based on a common field. By default in Tableau, tables are joined using Inner Join (Shown in Fig. 2-72).

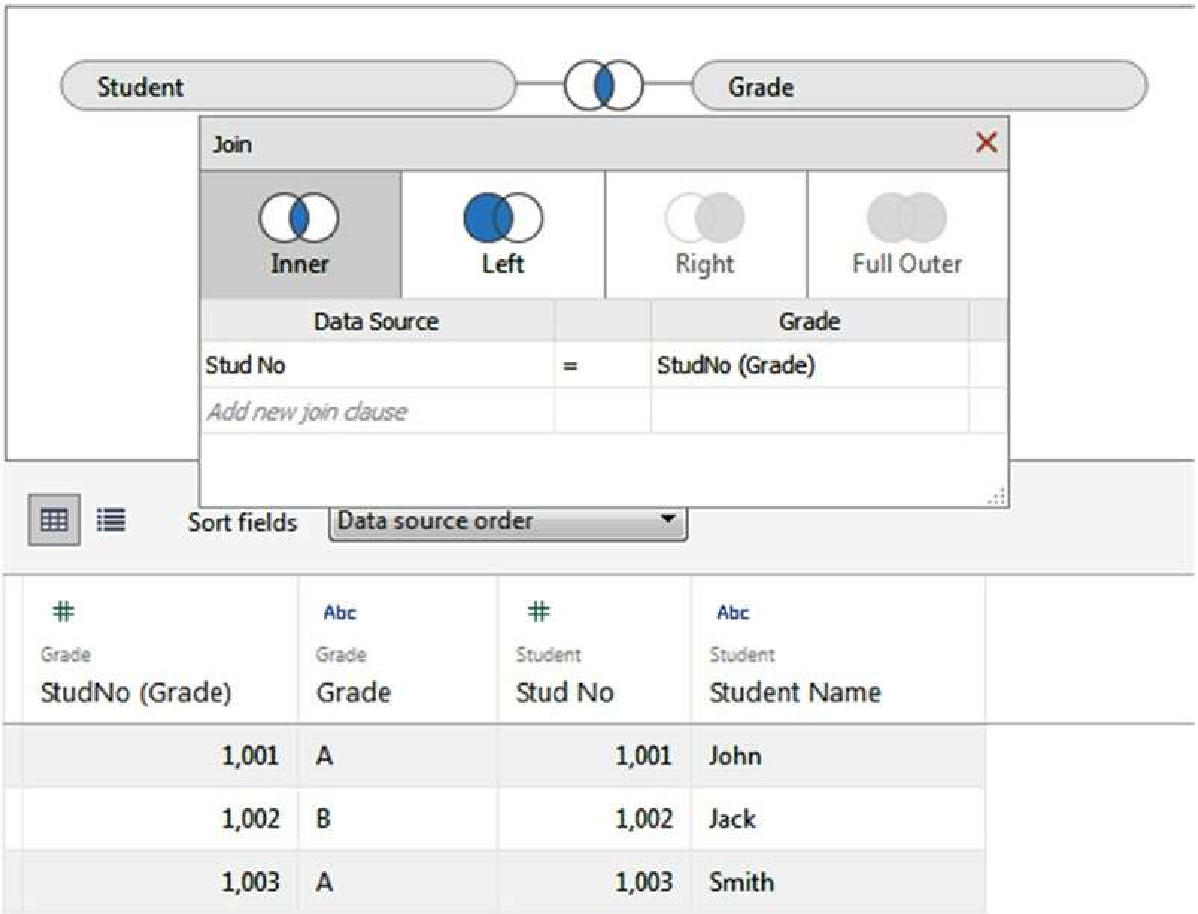


Figure 2-72. *"Inner Join" Operation*

2.6.2.3 Left Join

Left join fetches records from the left table having a matching record(s) in the right table. Null values will be displayed for records where there is no match in the right side table (Shown in Fig. 2-73).

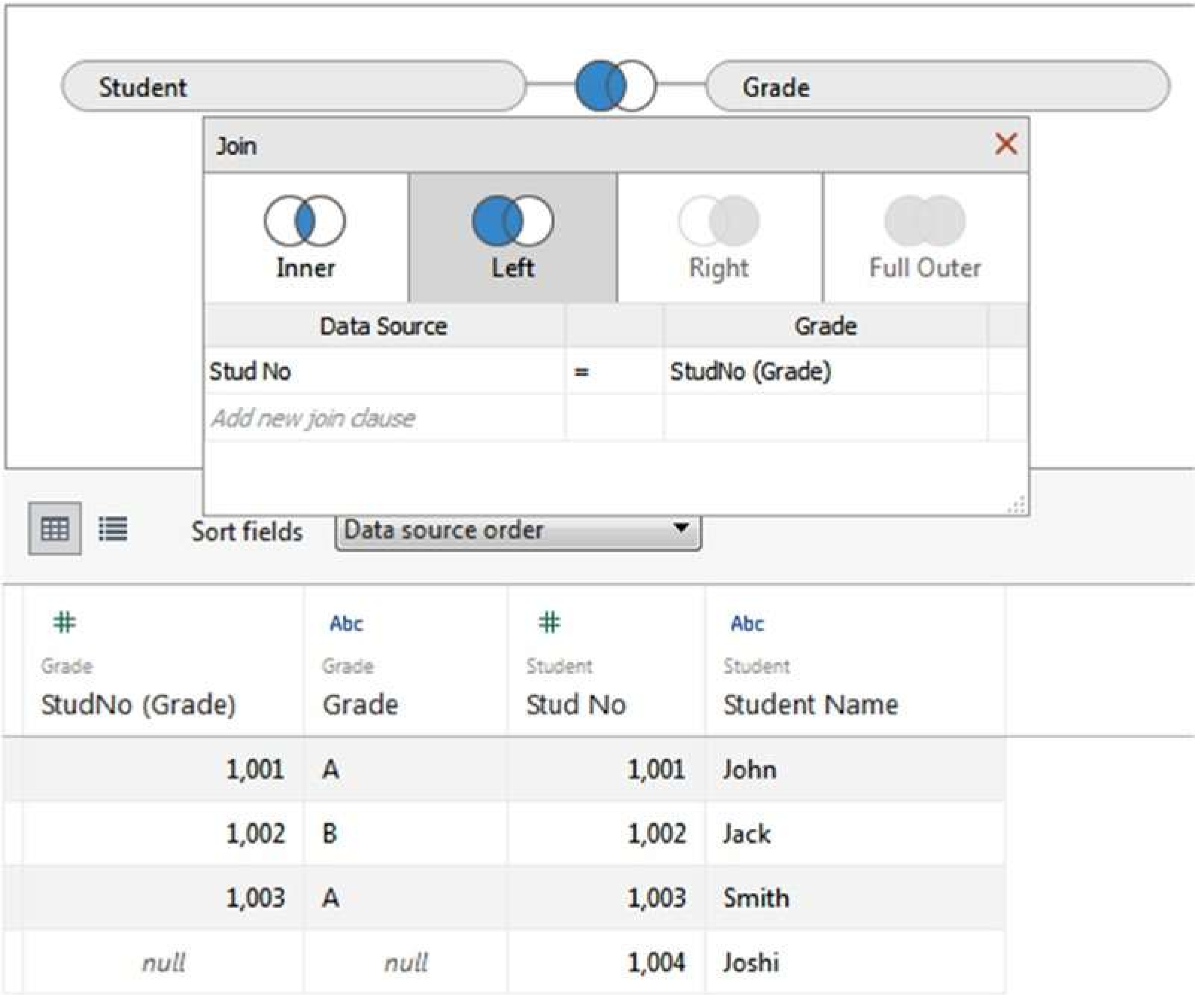


Figure 2-73. "Left Join" Operation

2.6.2.4 Right Join

MySQL data source supports Right Join. Refer to connect to server section to connect to MySQL data source. Create two tables namely "Student" and "Grade" with the same structure as shown in the example on "Inner Join".

Right join fetches records from the right table having a matching record(s) in the left table. Null values will be displayed for records where there is no match in the left side table (Shown in Fig. 2-74).

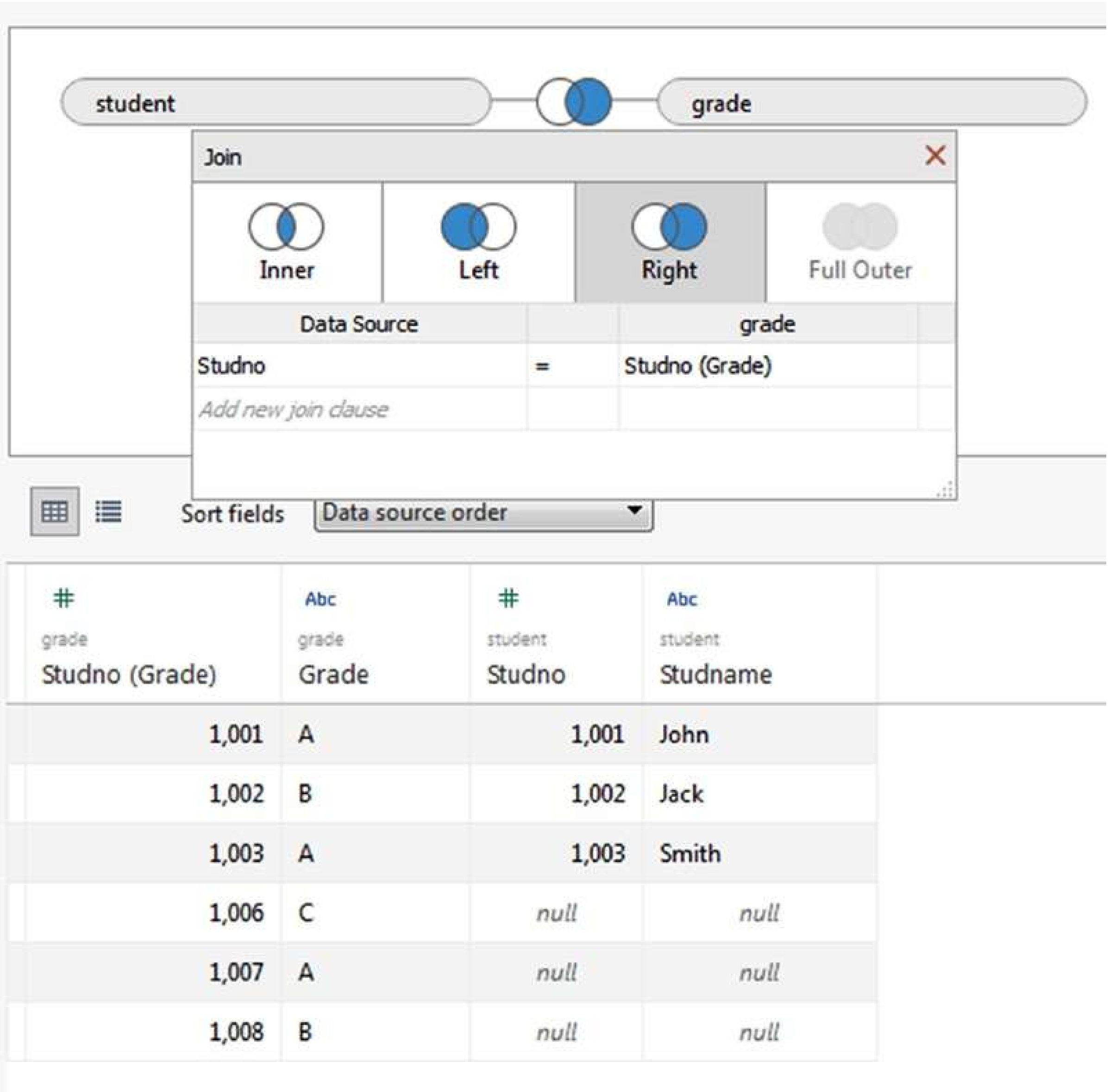


Figure 2-74. *"Right Join" Operation*

2.6.3 Union

In Tableau, "union" operation supports combining data from different files. Consider the "Sample – Student" Excel file used in Inner Join.

Follow the below steps to perform the "union" operation.

2.6.3.1 Steps

2.6.3.1.1 Step 1

Click on "New Union" option as shown below (Shown in Fig. 2-75).

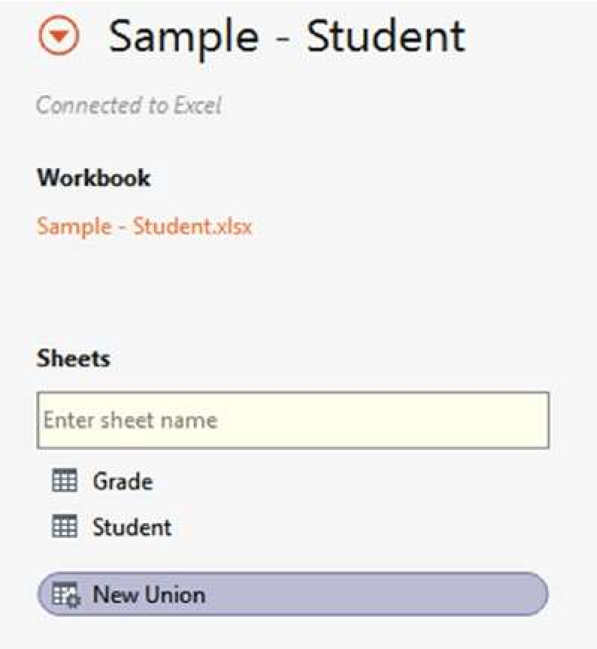


Figure 2-75. *"New Union" option*

2.6.3.1.2 Step 2

The "Union Window" will open. Specify the name as "Student_Grade", drag and drop the "Student" and "Grade" sheets to the window. Click on "OK" button to combine the data from the two sheets (Shown in Fig. 2-76).

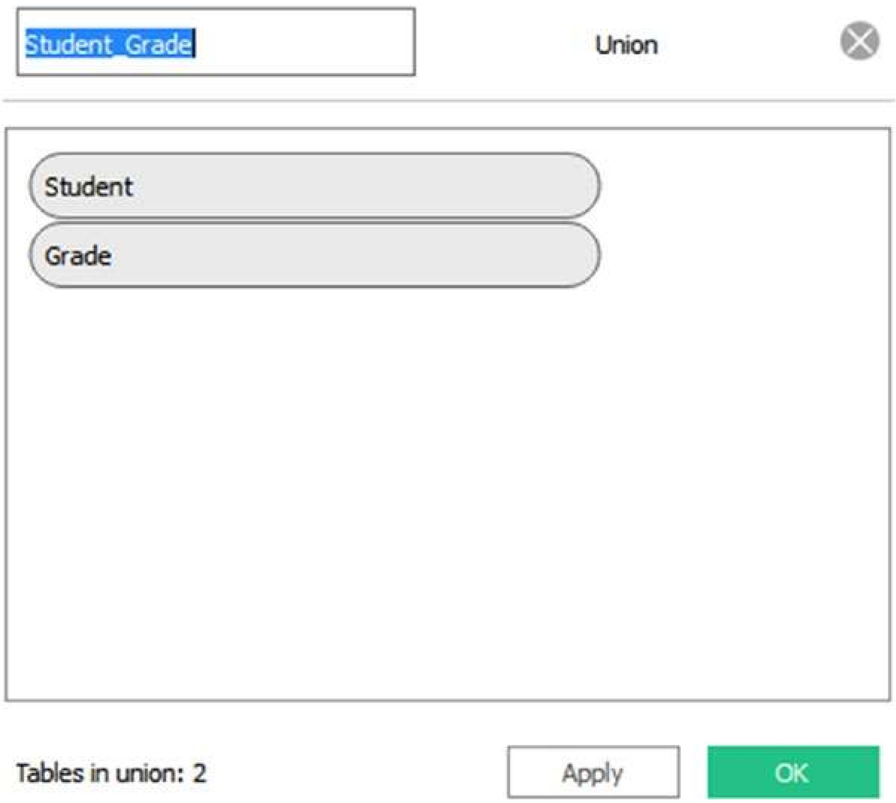
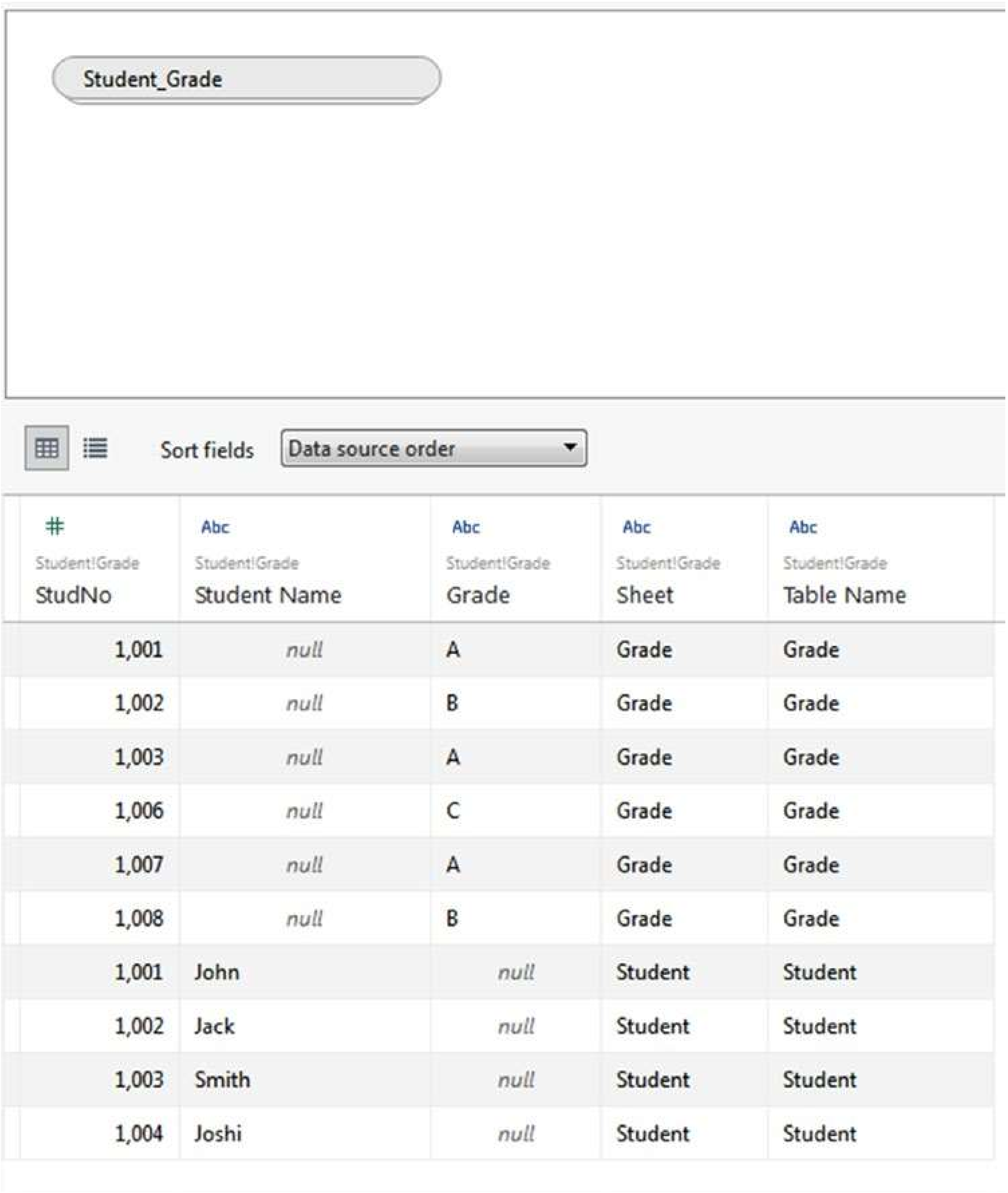


Figure 2-76. *Union window*

2.6.3.1.3 Step 3

You will be able to view the combined data in the grid window. It also displays the corresponding sheet name for each field (Shown in Fig. 2-77).



StudNo	Student Name	Grade	Sheet	Table Name
1,001	null	A	Grade	Grade
1,002	null	B	Grade	Grade
1,003	null	A	Grade	Grade
1,006	null	C	Grade	Grade
1,007	null	A	Grade	Grade
1,008	null	B	Grade	Grade
1,001	John	null	Student	Student
1,002	Jack	null	Student	Student
1,003	Smith	null	Student	Student
1,004	Joshi	null	Student	Student

Figure 2-77. *Union Result*

2.7 Custom SQL

Let us learn to write SQL statements to retrieve data suitable for analysis. Assume, you have connected to a MySQL data source, which has the following structure (Refer to Table 2-4, Table 2-5).

Table 2-4. *Student Table*

StudNo	Student Name
1001	John
1002	Jack
1003	Smith
1004	Joshi

Table 2-5. *Grade Table*

StudNo	Grade
1001	A
1002	B
1003	A
1006	C
1007	A
1008	B

Database Name: student

2.7.1 Demo 1

Follow the below steps to write a Custom SQL.

2.7.1.1 Steps to write custom SQL

2.7.1.1.1 Step 1

Select the "student" database (Shown in Fig. 2-78).

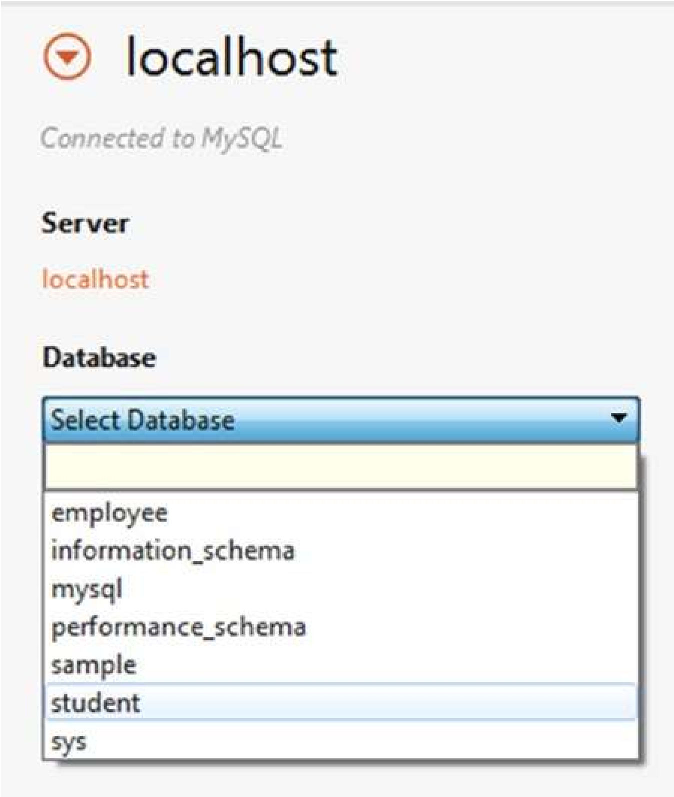


Figure 2-78. *Selection of "student" database*

2.7.1.1.2 Step 2

Double click on "New Custom SQL" (Shown in Fig. 2-79).

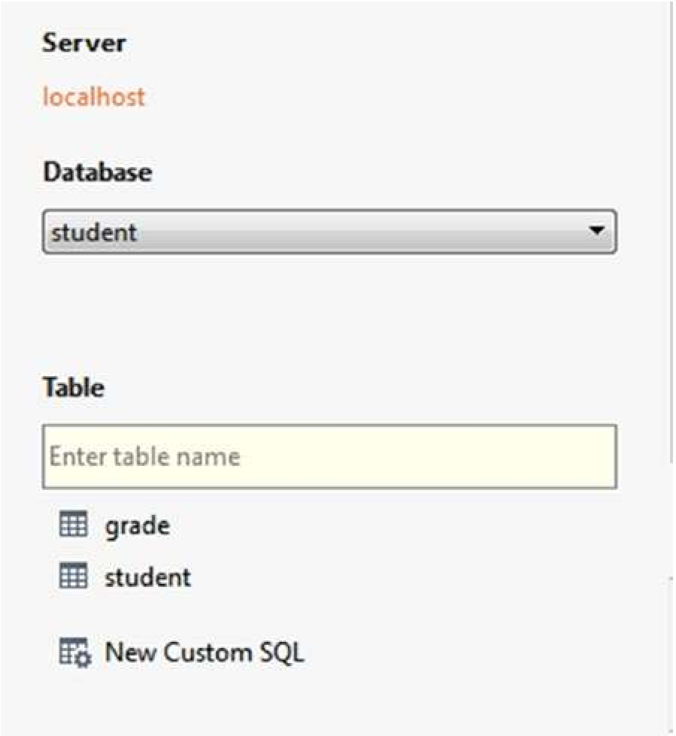


Figure 2-79. *"New Custom SQL" option*

2.7.1.1.3 Step 3

"Edit Custom SQL" dialog box shows up. Type in the required SQL statement (Shown in Fig. 2-80).

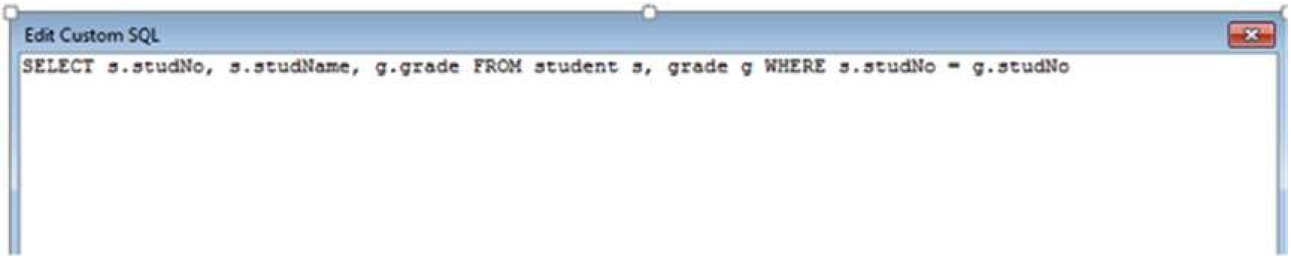


Figure 2-80. *"Edit Custom SQL" Statement*

2.7.1.1.4 Step 4

Click on "Preview Results" to preview the output and click on "OK" to create "Custom SQL" (Shown in Fig. 2-81).

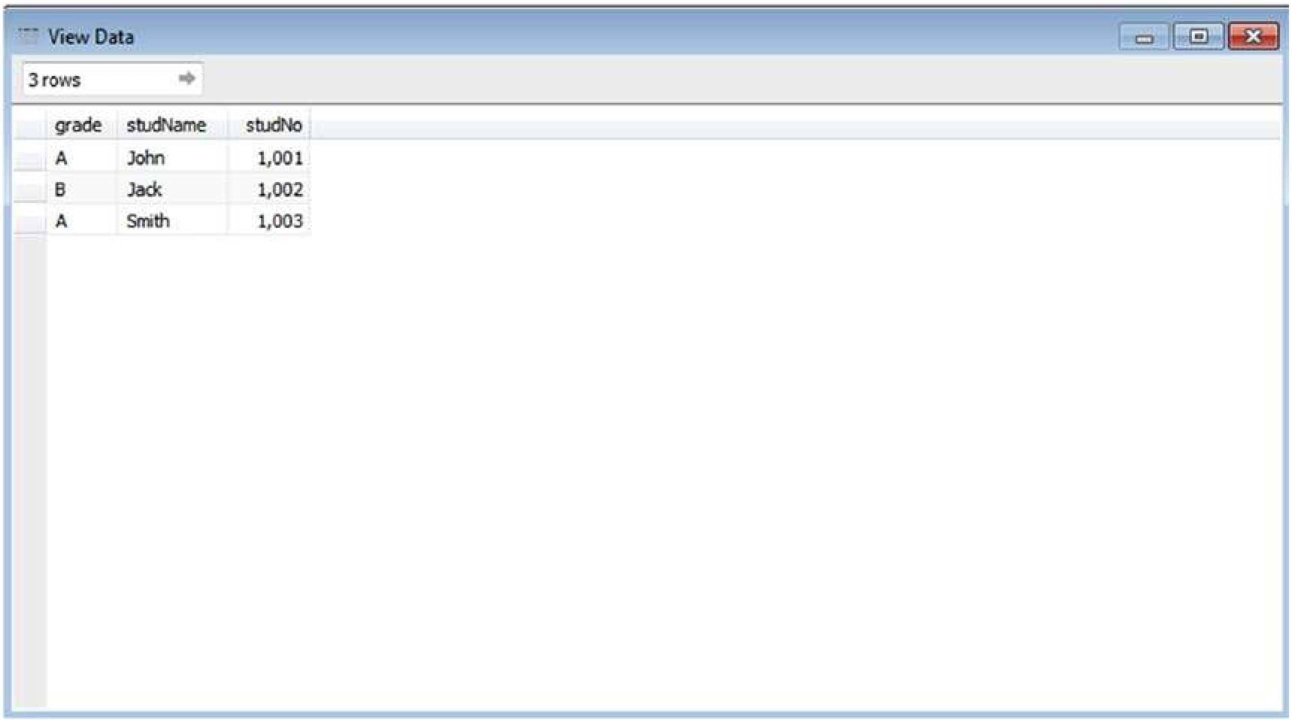


Figure 2-81. *"View Data" window showing preview of the data*

2.7.1.1.5 Step 5

Custom SQL Query is displayed in the canvas area (Shown in Fig. 2-82).

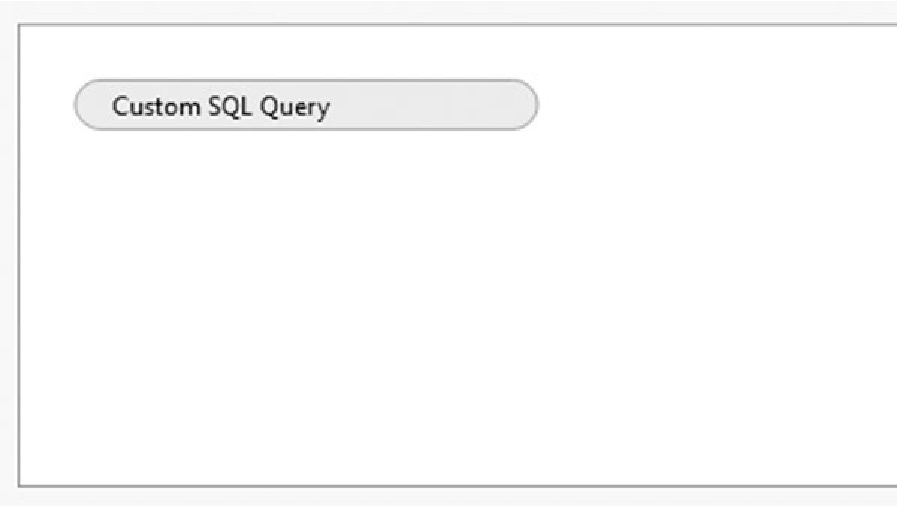


Figure 2-82. *"Custom SQL Query" placed in the canvas area*

2.7.1.1.6 Step 6

Right click on Custom SQL Query, to "Edit Custom SQL Query" (Shown in Fig. 2-83).

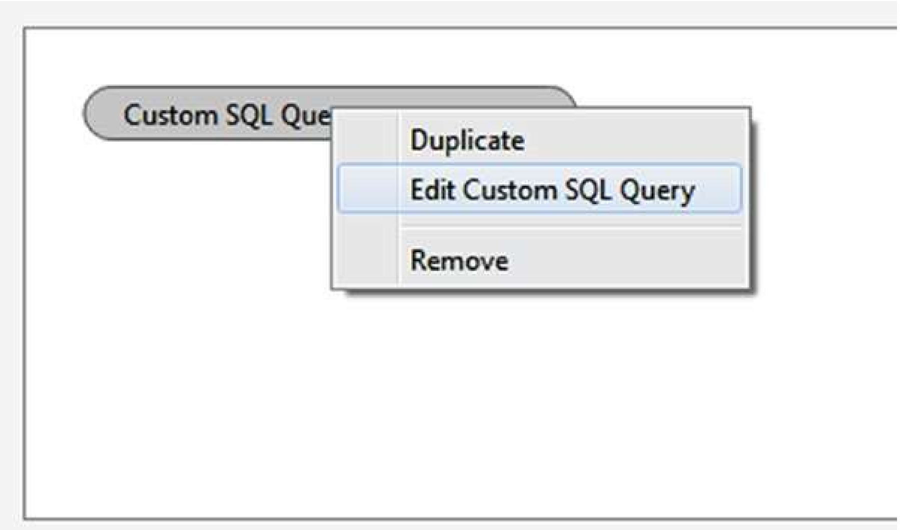


Figure 2-83. *"Edit Custom SQL Query"*

2.8 Data Blending

Data blending is the best choice when you want to use data from more than one data source for your analysis.

To perform data blending, a common field should be present in both the data sources. Here, we will consider "Sample-Superstore Excel data source" and "Sample - CoffeeChain Access data source". They have a common field namely, Market (CoffeeChain) and Region (Superstore).

2.8.1 Demo 1

Follow the below steps to perform Data Blending.

2.8.1.1 Steps

2.8.1.1.1 Step 1

Connect to Sample-Superstore data source.

2.8.1.1.2 Step 2

Drag and drop "Orders" sheet into the Canvas area.

2.8.1.1.3 Step 3

Connect to "Sample-Coffee chain Access" data source.

2.8.1.1.4 Step 4

Drag and drop "CoffeeChain Query" to the Canvas area.

2.8.1.1.5 Step 5

Go to worksheet. To create a relationship, select the data menu, and then select "Edit relationship..." (Shown in Fig. 2-84).

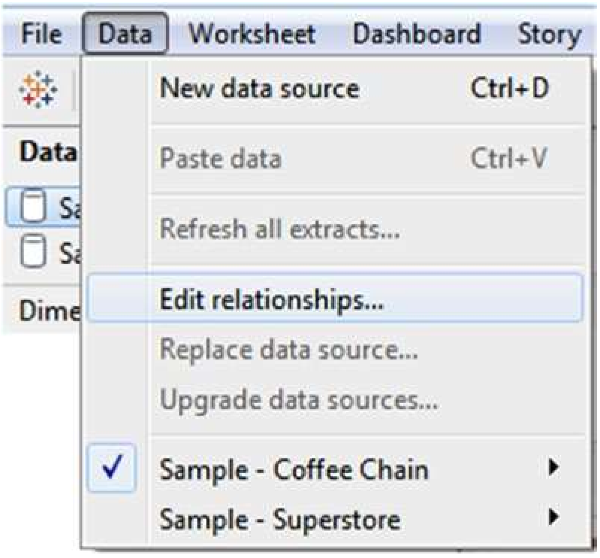


Figure 2-84. *"Edit relationships..." option*

2.8.1.1.6 Step 6

Relationships dialog box will open as shown below (Shown in Fig. 2-85).

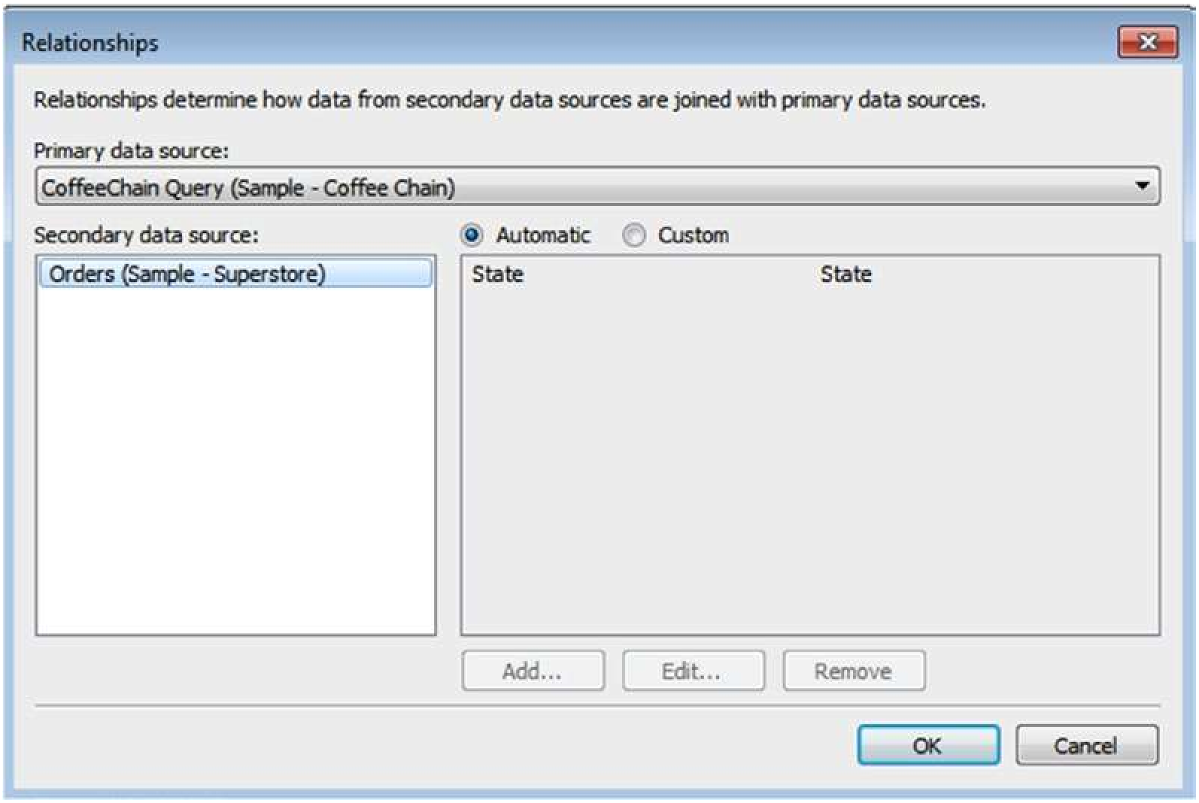


Figure 2-85. *"Relationships" dialog box showing Primary data source and Secondary data source*

2.8.1.1.7 Step 7

Click on the "Custom" radio button to create a custom relationship and then click on the "Add" button to add field mapping (Shown in Fig. 2-86).

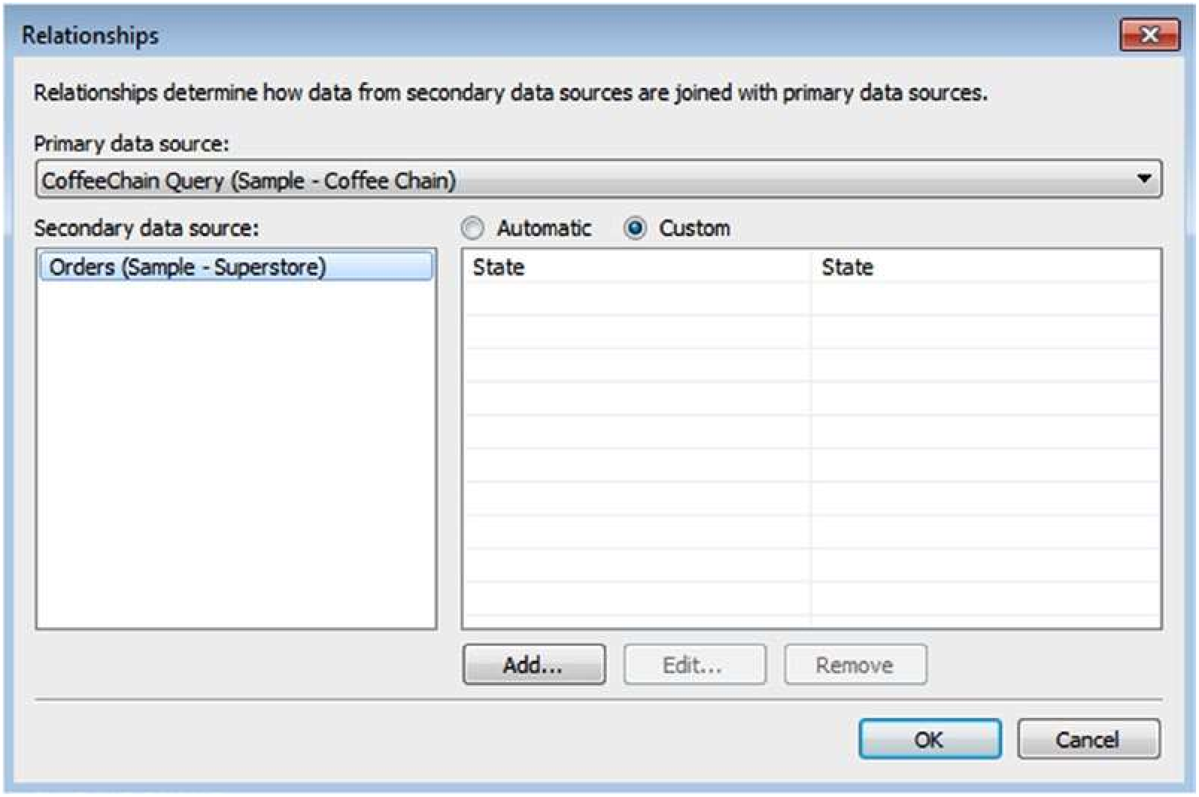


Figure 2-86. *Selected "Custom" option*

2.8.1.1.8 Step 8

From the Add/Edit field mapping dialog box, select "Market" from the primary data source (Coffeechain) and "Region" from the secondary data source (Superstore) and click on "OK" to create the field mapping (Shown in Fig. 2-87).

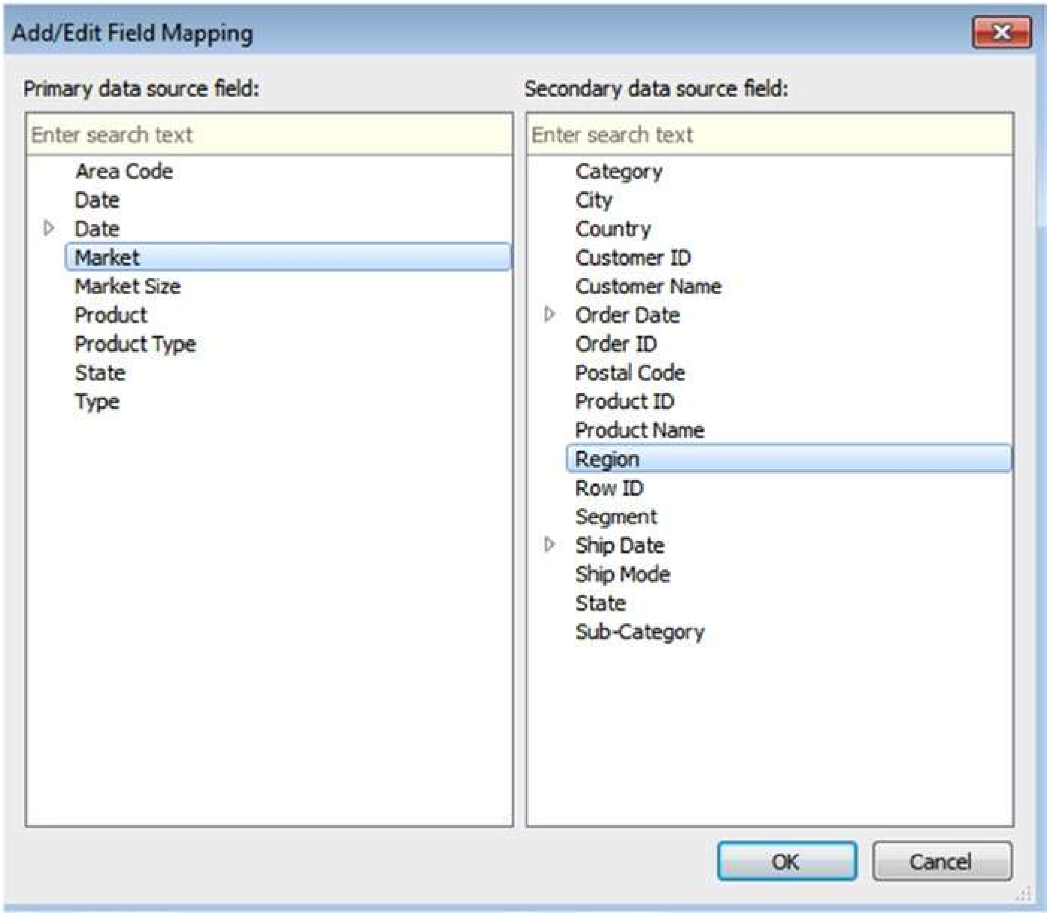


Figure 2-87. *"Market", "Region" mapping*

2.8.1.1.9 Step 9

Observe the mapping field in the "Relationships" dialog box. (Shown in Fig. 2-88).

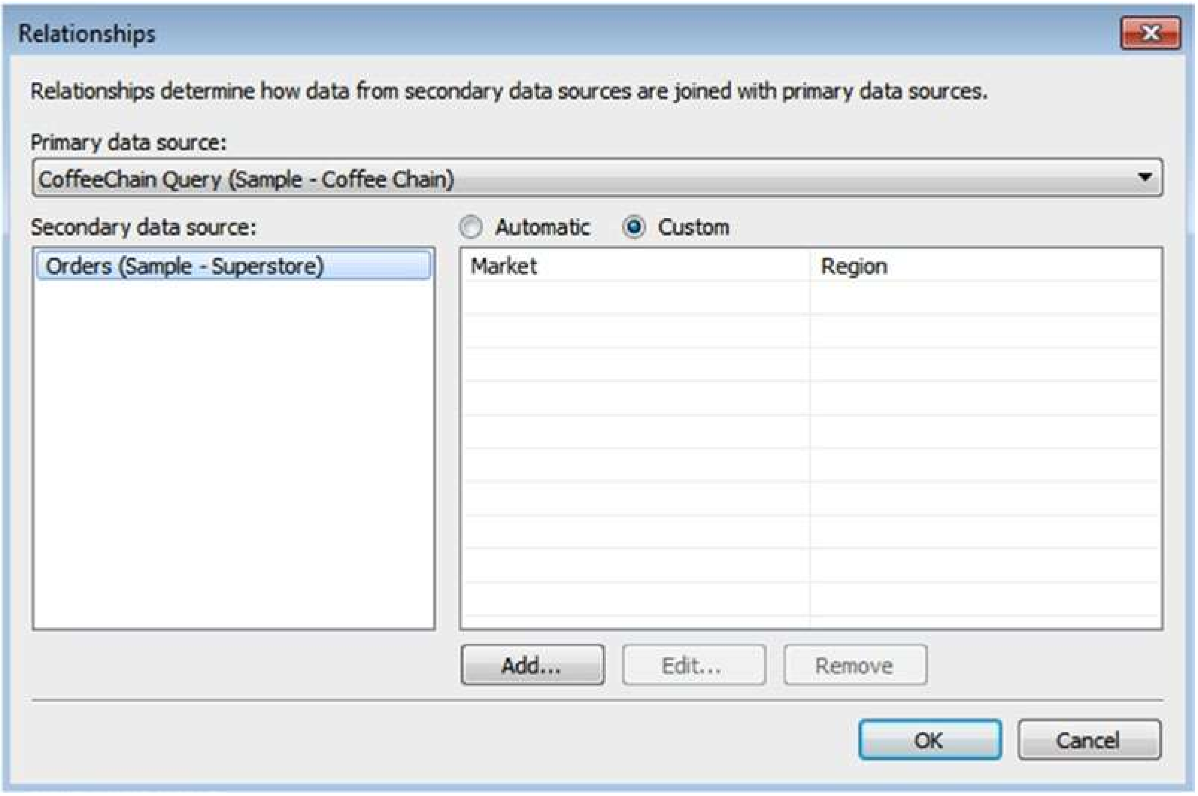


Figure 2-88. *Relationship window showing "Market", "Region" mapping*

2.8.1.1.10 Step 10

From "CoffeeChain Query", select the dimension "Market" and place it on the rows shelf (Shown in Fig. 2-89).

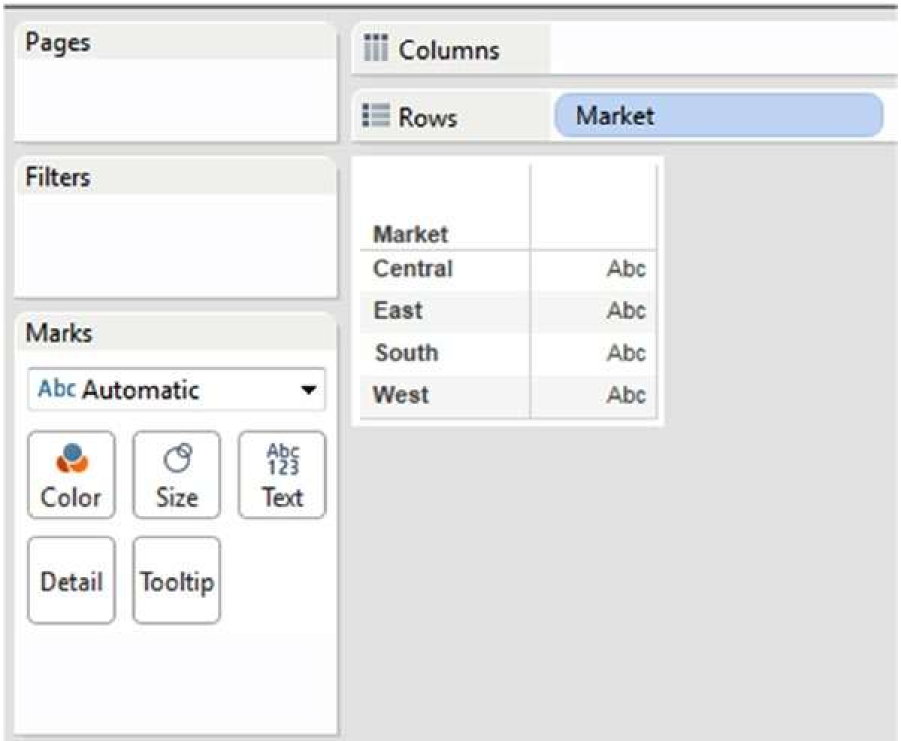


Figure 2-89. *Dimension, "Market" placed on the rows shelf*

2.8.1.1.11 Step 11

Now, click on "Orders" data source. You can see the relationship symbol (paper clip symbol) next to the dimension, "Region" (Fig. 2-90).

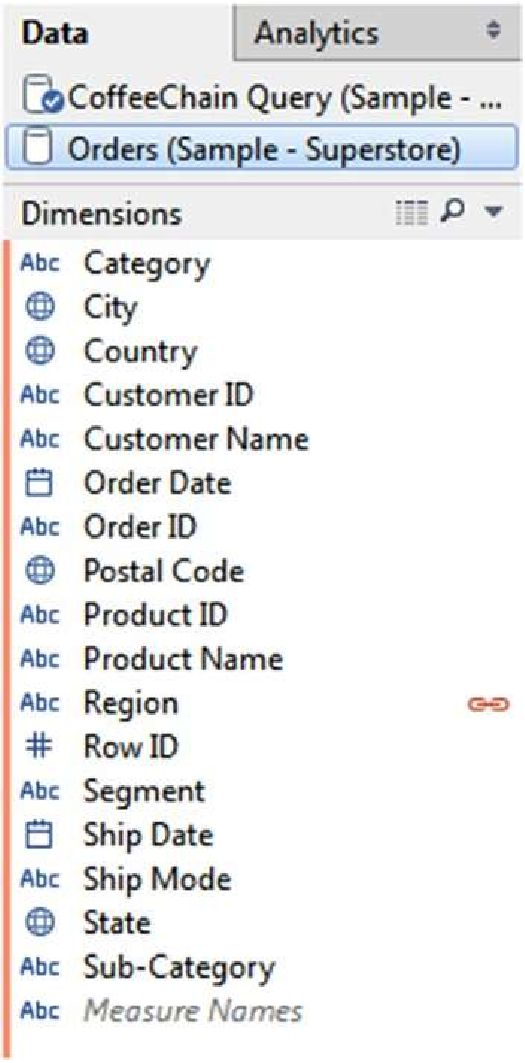


Figure 2-90. *Relationship symbol for the dimension "Region"*

2.8.1.1.12 Step 12

From Orders, select the measure "Sales" and place it on the columns shelf to construct a view. The view below represents the "Sales" by "Market" (Fig. 2-91).

Figure 2-91. *"Sales" by "Market" view*

2.9 Data Extracts

The saved subsets of data source are known as extracts. Data extracts can be used to:

- improve performance: When you want to work with only a subset of data, filter extract helps you to limit the load on the server.
- use Tableau functionality: You can use Tableau functionality such as Count Distinct, which is not available with the original data source.
- to provide an offline access to data: You can extract the data to a local data source when you don't have access to server.

2.9.1 Demo 1

Let us learn how to create an extract.

2.9.1.1 Steps to create an extract

Follow these steps.

2.9.1.1.1 Step 1

On the data source page, select “Extract” and click on “Edit” to open the Extract Data window (Shown in Figure 2-92).

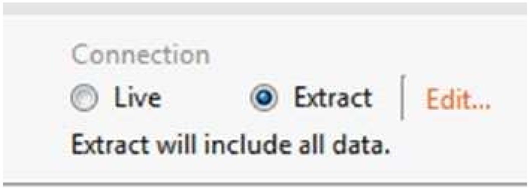


Figure 2-92. *Extract option*

2.9.1.1.2 Step 2

Extract Data dialog box is displayed (Shown in Fig. 2-93).

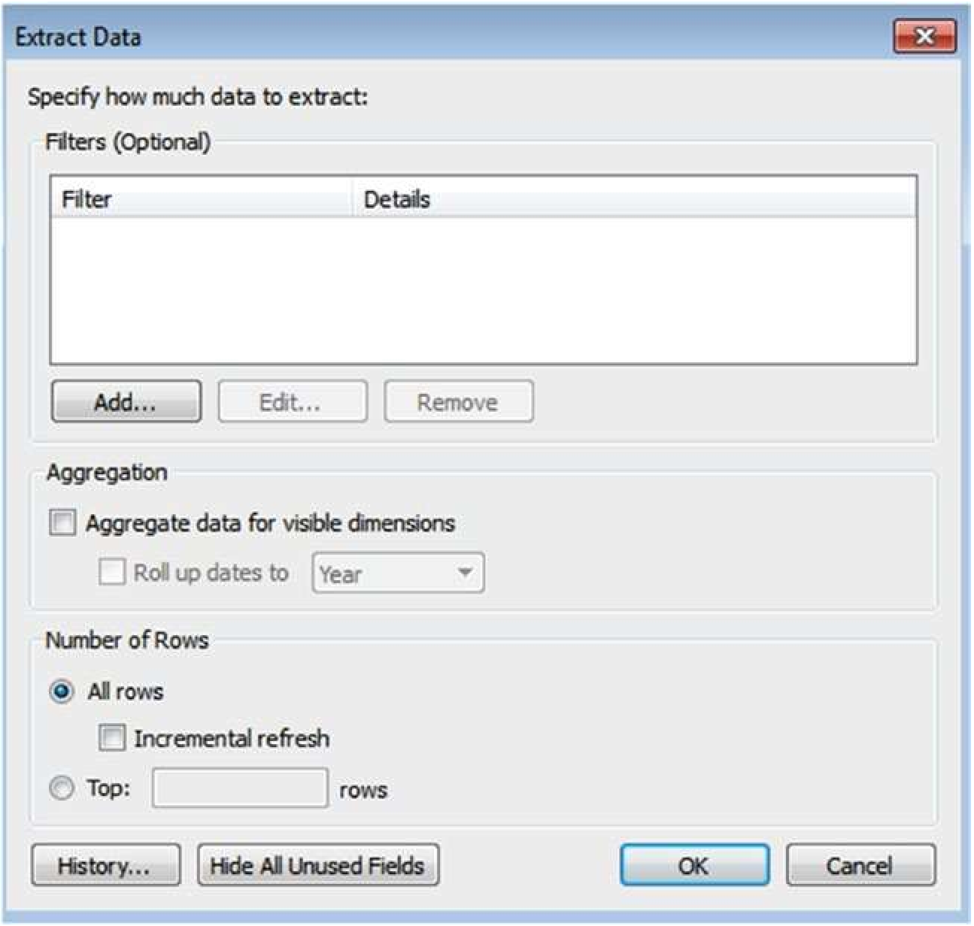


Figure 2-93. *"Extract Data" dialog box*

2.9.1.1.3 Step 3

You can limit data for an extract by adding a filter. Click on "Add" button to add a filter. Select "State" field (Shown in Fig. 2-94) as filter criteria and click OK.

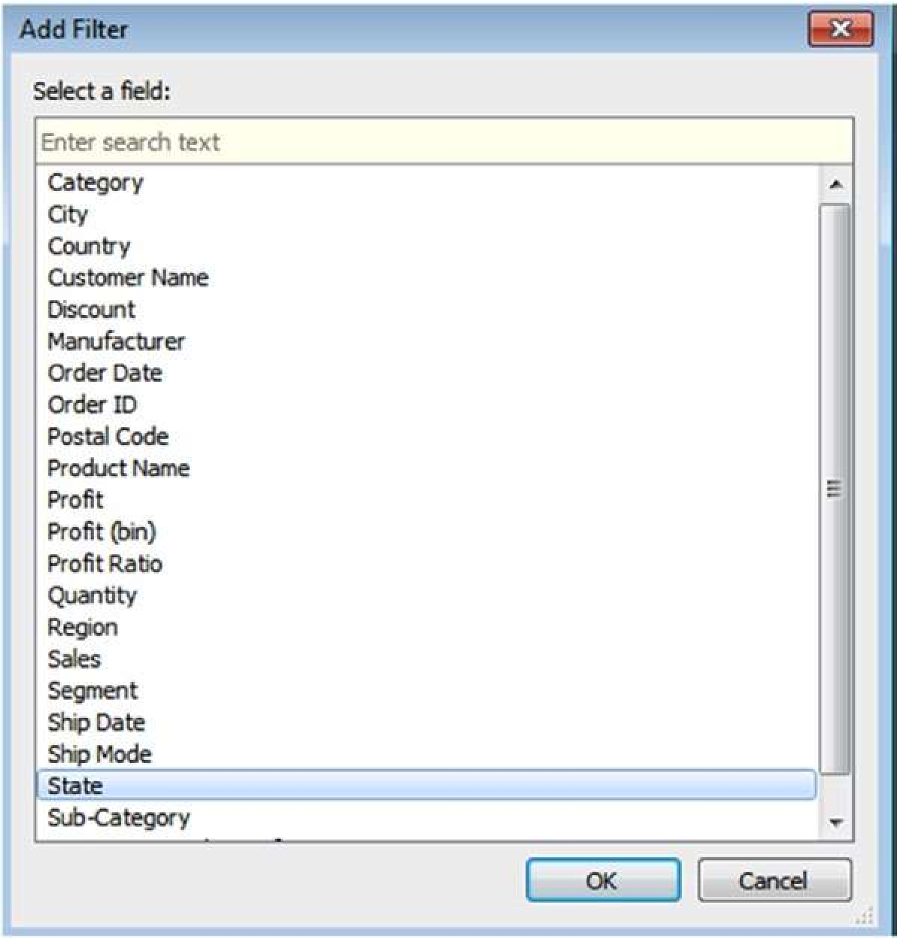


Figure 2-94. *Selected "State" field for filter condition*

2.9.1.1.4 Step 4

Filter [State] dialog box shows up. From the list of states, select only "California" (Shown in Figure 2-95) and click "OK".

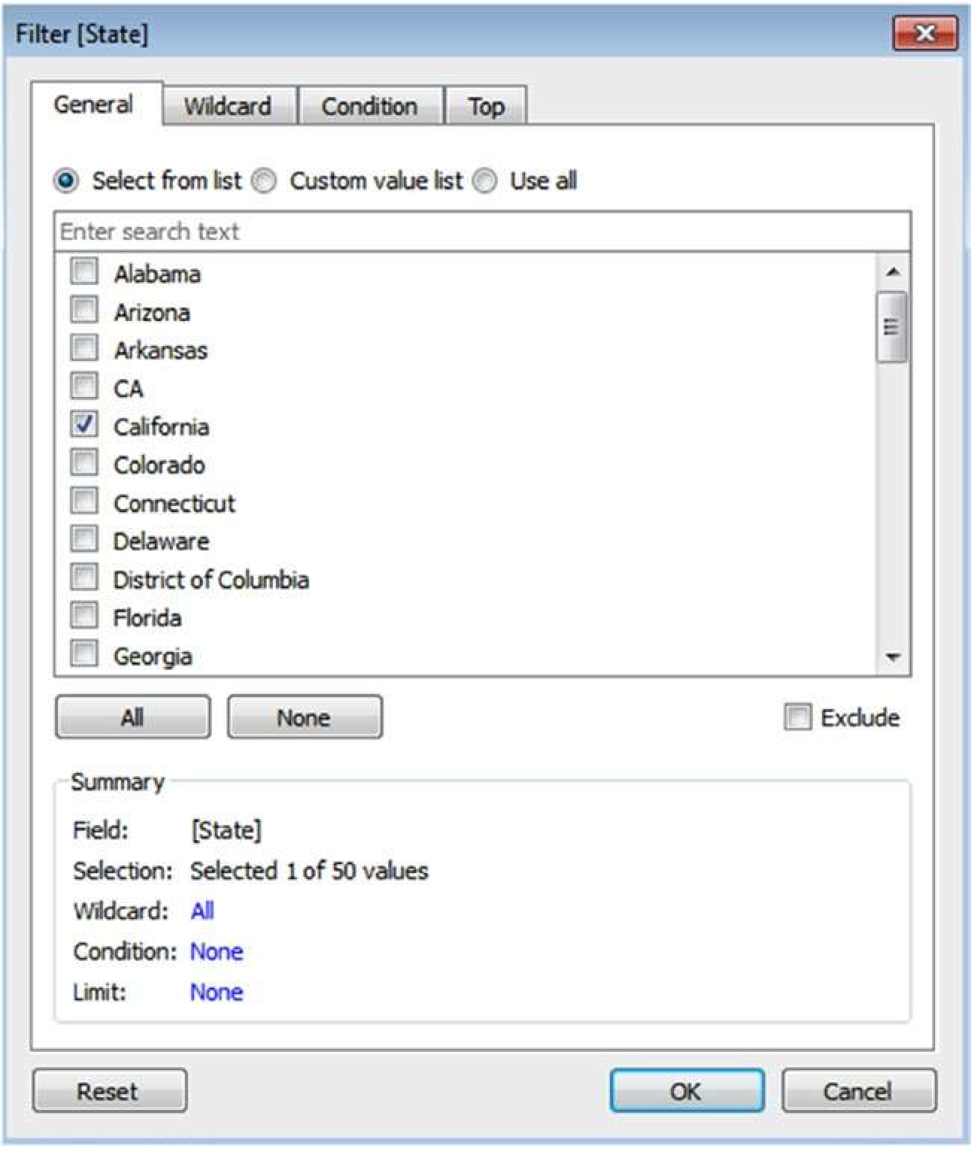


Figure 2-95. *Filter[State] dialog box, California checked*

2.9.1.1.5 Step 5

Check "Aggregate data for visible dimensions" to aggregate measures by their default aggregation. This helps you to minimize the extract file size and to increase performance. Also choose "Roll up dates to" to specify a date level such as Year, Month and select the "Number of Rows" to display a certain number of rows (Shown in Fig. 2-96).

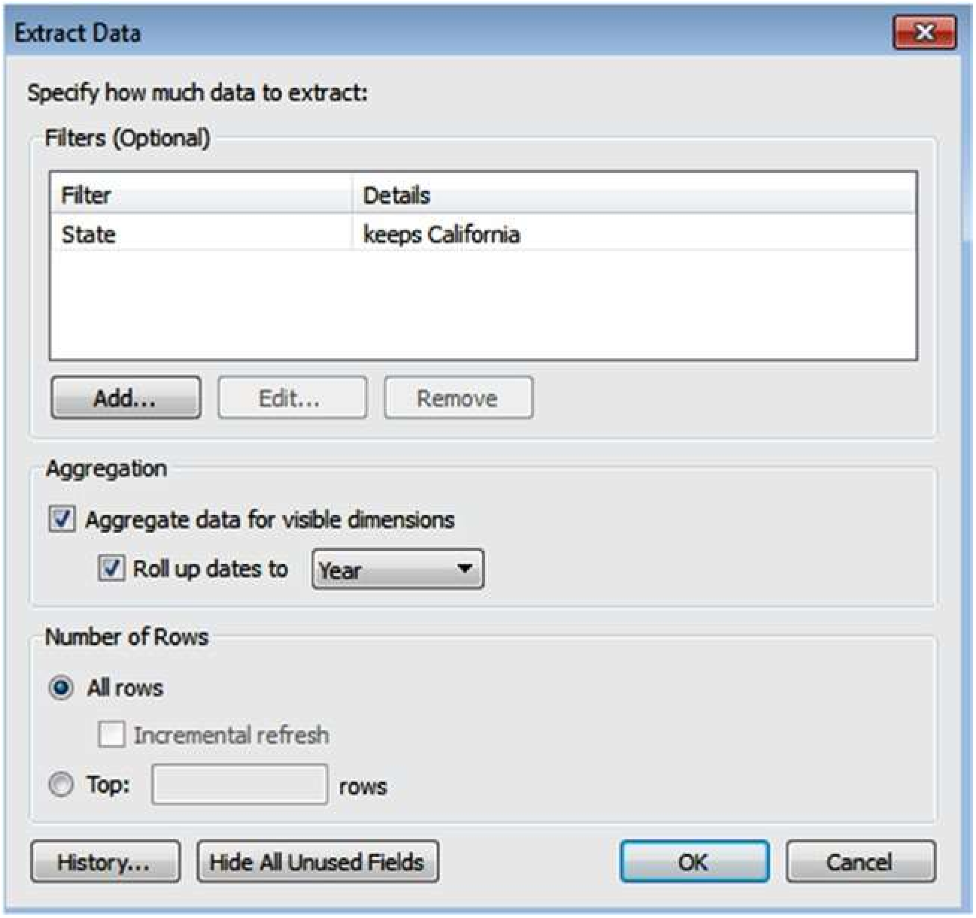


Figure 2-96. *Extract data window showing conditions for Extraction*

2.9.1.1.6 Step 6

Observe details about the state of “California” in the data grid (Shown in Fig. 2-97).

Orders Country	Orders City	Orders State
United States	Mission Viejo	California
United States	Los Angeles	California
United States	Los Angeles	California
United States	Pasadena	California
United States	Los Angeles	California
United States	San Diego	California
United States	Fresno	California
United States	San Francis...	California
United States	Los Angeles	California
United States	Sacramento	California

Figure 2-97. *Data grid showing only state of California details*

2.9.1.1.7 Step 7

When you go to Sheet 1, you will get a “Save Extract As” dialog box to save your extract (Shown in Fig. 2-98).

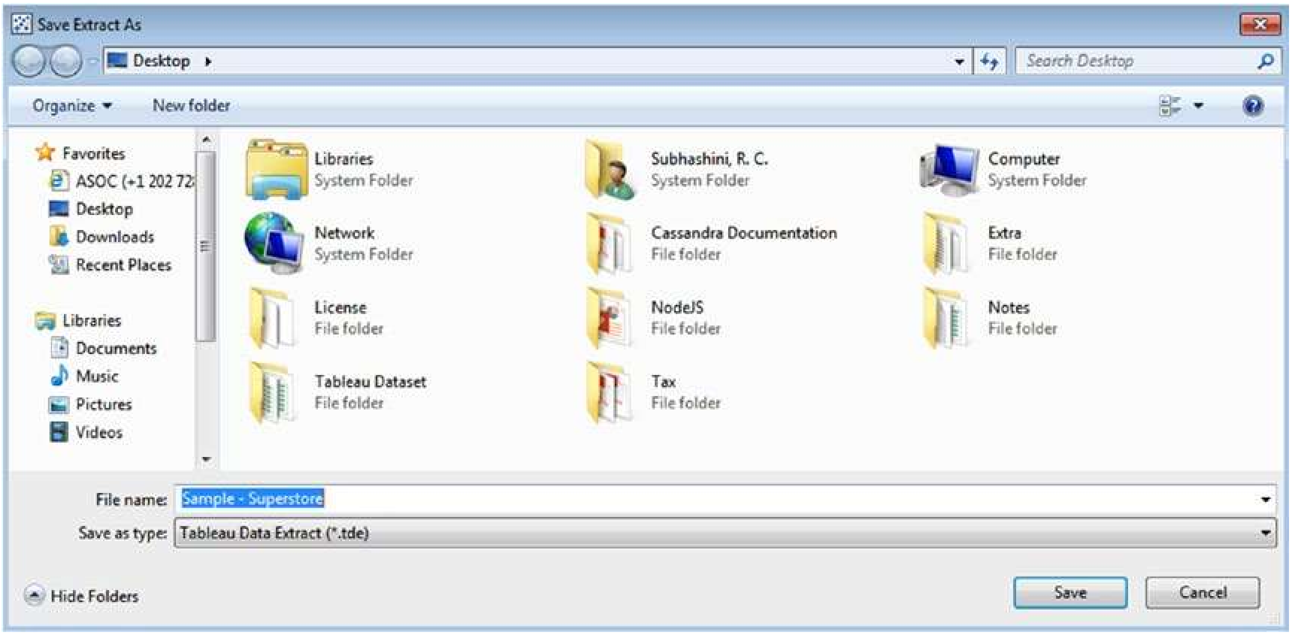


Figure 2-98. *“Save Extract As” dialog box to save extract*

2.10 Points to Remember

- Tableau has a fast, in-memory data engine for analytics.
- Tableau provides various data connectors for databases and a generic ODBC connector to connect to any system not having a native connector.
- All joins are not supported by all databases.
- Tableau extracts helps to improve the performance and provides offline access to your data.

2.11 Next Step

In the next chapter, we will focus on the following:

- Filter
- Sort
- Group
- Hierarchies
- Sets