

Aim: Connecting and extracting with various data resources in tableau.

Background Information:

Data Sources

Tableau can extract data from all the popular data sources. These include:

1) File System

The simplest data source you can use with Tableau is a file. These could be files like an Excel spreadsheet, a CSV file or a text file.

2) Cloud System

You can also source data from popular cloud sources. Some of the options are:

- Google Analytics
- Google BigQuery
- Windows Azure
- Amazon Redshift

3) Relational systems

You can connect to many types of relational databases such as SQL Server, Oracle, and DB2.

4) Live Data Sources

Connect live is a feature of Tableau that allows you to connect real-time data. Tableau does this by constantly reading the data, so your visualizations are constantly up to date.

5) Using In-Memory Data

The alternative to connecting to a live data source is to load one into memory. This is a better option for static data that won't change anytime soon, as it will only be loaded once. The in-memory database will then be analyzed by Tableau.

6) Connecting Multiple Data Sources

One of the great features of Tableau is the ability to combine data sources. You can work with data from a file system and data from a relational database all at the same time. All you need to do is define multiple data connections.

Data Extraction Techniques

- Once you've decided on your data sources, the next step is to extract the data you need from those sources.
- Whether you are connecting to a live database or storing your data in memory, you may well want to cut it down to only what you need for your application. This will mean you'll have less data to extract from a live source or a smaller amount of data to store in memory.
- It also converts the data to a form that works well with the Tableau engine, meaning things will speed up even more.
- With Tableau, this is done with data extracts.
- A data extract is simply a subset of a total data source. When extracting data, you can choose exactly what you want and how much of underlying data to extract using extract data dialog box.
- To create a new Tableau data extract, go to Data -> Extract Data. You'll be presented with many options to limit the number of rows and aggregate for dimensions. Here is where you can use filters to cut down your data to just the things you need.

Filtering Extracted Data

- You might not need every single field and row in the data you've extracted. By cutting it down to just the things you need, you can improve performance and make life easier for yourself.
- There are three main types of filters to use in Tableau
 - 1) Dimension filter
 - 2) Measure filter
 - 3) Date filter
- Each works on a different type of data field. To apply a filter, simply drag a field into the filter pane, it looks like this.
- Then you'll be prompted with some options for your filter. Choose the ones you need and click apply.
- Once you've created a data extract, you can add more data to it from the data pane. Do this by going to Data -> Extract -> Append to File. You can do this with new data types, just make sure they are the same type and have the same number of fields as the original data.
- It's possible to work on large data sets using Tableau. Things do, however, get a little more difficult if your dataset doesn't fit in memory. This is where data extracts and filters really come in handy. If your data is still too big to fit in RAM after extracting and filtering it down, it will still work but will run a lot more slowly.

Excel File Data Source

The screenshot shows the Tableau Public interface with a data source named 'Employee Sample Data' connected to an Excel file. The data is displayed in a table with 14 fields and 1000 rows. The fields are: Eeid, Full Name, Job Title, Department, Business Unit, Gender, Ethnicity, Age, and Hire Date. The table shows data for employees like Penelope Jordan, Austin Vo, Joshua Gupta, Ruby Barnes, and Luke Martin.

Eeid	Full Name	Job Title	Department	Business Unit	Gender	Ethnicity	Age	Hire Date
E02832	Penelope Jordan	Computer Systems Manager	IT	Manufacturing	Female	Caucasian	26	9/27/2019
E01639	Austin Vo	Sr. Analyst	Finance	Manufacturing	Male	Asian	55	11/20/1995
E00644	Joshua Gupta	Account Representative	Sales	Corporate	Male	Asian	57	1/24/2017
E01550	Ruby Barnes	Manager	IT	Corporate	Female	Caucasian	27	7/1/2020
E01550	Luke Martin	Analyst	Finance	Manufacturing	Male	Black	25	5/16/2020

Text File Data Source

The screenshot shows the Tableau Public interface with a data source named 'Employee Sample Data' connected to a text file. The data is displayed in a table with 14 fields and 1000 rows. The fields are: Eeid, Full Name, Job Title, Department, Business Unit, Gender, and Ethnicity. The table shows data for employees like Emily Davis, Theodore Dinh, Luna Sanders, Penelope Jordan, Austin Vo, and Joshua Gupta.

Eeid	Full Name	Job Title	Department	Business Unit	Gender	Ethnicity
E02387	Emily Davis	Sr. Manager	IT	Research & Development	Female	Black
E04105	Theodore Dinh	Technical Architect	IT	Manufacturing	Male	Asian
E02572	Luna Sanders	Director	Finance	Specialty Products	Female	Caucasian
E02832	Penelope Jordan	Computer Systems Manager	IT	Manufacturing	Female	Caucasian
E01639	Austin Vo	Sr. Analyst	Finance	Manufacturing	Male	Asian
E00644	Joshua Gupta	Account Representative	Sales	Corporate	Male	Asian

JSON File Data Source

Tableau Public - Book1

FileDataWindowHelp

iris

iris

Filters
0Add

Connections

iris
JSON file

Files

iris.json

New Union

New Table Extension

Table Details

iris.json6 fields 150 rows100 rows

Document Index (gener...	Petal Length	Petal Width	Sepal Length	Sepal Width	Species
1	1.40000	0.20000	5.10000	3.50000	setosa
2	1.40000	0.20000	4.90000	3.00000	setosa
3	1.30000	0.20000	4.70000	3.20000	setosa
4	1.50000	0.20000	4.60000	3.10000	setosa
5	1.40000	0.20000	5.00000	3.60000	setosa
6	1.70000	0.40000	5.40000	3.90000	setosa
7	1.40000	0.30000	4.60000	3.40000	setosa
8	1.50000	0.20000	5.00000	3.40000	setosa
9	1.40000	0.20000	4.40000	2.90000	setosa
10	1.50000	0.10000	4.90000	3.10000	setosa
11	1.50000	0.20000	5.40000	3.70000	setosa
12	1.60000	0.20000	4.80000	3.40000	setosa

Data Source

Sheet1