Exploring Splunk Queries

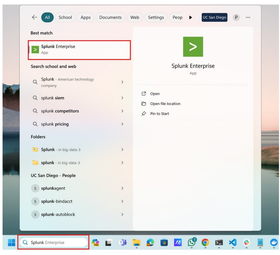
By the end of this activity, you will be able to:

- Import CSV files into Splunk.
- Query, filter, and plot data.
- Perform statistical calculations.

Step 1. Open Splunk.

Windows

Go to the search menu, and search for *Splunk*. Once you locate the application, click on it, an you will be redirected to http://localhost:8000/ where is Splunk will be running.



macOS

First, open your terminal and run the following command to move into your *splunk* directory:

```
1  cd /Applications/splunk
```

Then, run the following command to start *Splunk*:

```
1  ./bin/splunk start
```

Once you start Splunk, open your browser and go to *http://localhost:8000/*.

Step 2. Login. Login to *Splunk* by entering your credentials:

Step 3. Import census data. Let's import the census data CSV file to Splunk. First, click on *Settings* in the top right, then click on *Add Data*:

Next, click on *Upload*:

Click on *Select File*:

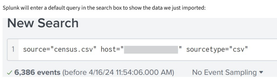
Navigate to your *big-data-3/Splunk* directory and select *census.csv*. Then click *Next>*:

On the left, make sure the *Source type* is *csv*.

If the *Source type* is not *csv*, click on *Source type*, go down to *Structured*, and select *csv*:

The table on the right is a preview of the CSV data. It shows the values for different fields:

Next, click on *Next*, click on *Review*, and then click *Submit*. Splunk will say the file is successfully uploaded:



Step 4. View data. Click on *Start Searching*:

Splunk will enter a default query in the search box to show the data we just imported:

This query shows all the data from the *census.csv* file and whose data type is CSV. In general, we can use *source=* to query from different file names, and *sourcetype=* to query from different formats.

The table shows the results matching this query:

Step 5. Filtering for specific values. We can filter the results by looking for a field with a specific value. For example, we can find the entries where the state is California:

The field *STNAME* contains the name of the state, and the above query only shows the results where the state is California. We can use an OR to search for multiple values on the same field:

We can search multiple fields with specific values by adding them to query. For example, let's search for state name California and 2010 population greater than one million people:

We can filter our results to just show a single column. For example, let's just show the county names of the previous query:

The | (pipe) is the syntax for sending the results from one query to the next, and the *table* command creates a table using only the specified column name(s).

We can sort the results based on any of the fields, such as population, and order them in either ascending or descending order. The image below shows an example of a search for all items with a population greater than 100000, sorts the results in descending order, and creates a table containing the population and state name. (To sort in ascending order you would replace "desc" with "asc").

Instead of using "desc" you can use a dash before the sorting field, e.g. "... | sort -CENSUS2010POP | table ..." for the above query.

We can view plots of search results by clicking on the *Visualization* tab. For example, if we use our last query and add the 2010 population value to the table:

We can click on the *Visualization* tab to see a chart of the results:

Step 6. Perform statistical calculations. The *Splunk stats* command is used to perform statistical calculations on the data. Let's count the results where the state is California:

We can sort based on the count by adding "| sort count" to the above query. This would sort in ascending order. If we want to sort in descending order we would use " | sort -count".

Next, let's compute the total 2010 population for California:

Finally, let's compute the average 2010 population for California:

Note: If you close Splunk and want to work with the data later, **do not load the data again.** Loading the data again will create a duplicate version, causing all your queries to return doubled results.

To access the data again, open Splunk, click *Apps* in the upper left corner, and select *Search and Reporting*.

Once you open the query engine again, make sure to set the **time picker to "All time"** to ensure that you can access the data, regardless of how many days have passed since you uploaded it.

Mark as completed

Like Dislike Report an issue