|  |
| --- |
|  |



# Data Manipulation with Splunk

We will be showing you how to manipulate data in Splunk, as well as adding data and creating different Splunk features.

**In this section, we will go over the following:**

* Adding in data to Splunk
* Create different searches
* Create customizable dashboards
* Create lookups



Before we begin, please make sure you have all of the required fields for this activity. They should be the *airports.csv* and *countries.csv* files. If you do not have them, please let us know right away!

## Adding Data

1. The first thing we want to make sure of is that we’re logged into Splunk. If you have any trouble logging in, please let one of us know.
2. Once logged in, we can get started. Let’s load out dataset for today. We’ll be working with airport data from around the world.

In your Splunk app, go to Settings 🡪 Add Data 🡪 Upload 🡪 airports.csv 🡪 Next

****

 Now we need to set the sourcetype. Splunk is trained to automatically recognize many sourcetypes, but we may need to label our customer sourcetypes (or we might want to override the default). Our default is correct here , and the sourcetype should be labeled as **csv**. If not, select from the drop down.

 Sourcetype 🡪 Structured 🡪 csv

 Notice that we have a warning that Splunk is not recognizing the timestamp of our data. Splunk will automatically look for a timestamp at the front of the event that meets epoch time—in cases where our data is not formatted like this, we can tell Splunk how to recognize the timestamp by writing the string in strptime, or telling Splunk what field the timestamp exists in (note that Splunk has already shown us what fields it is able to pull from our data).

 Because we’re working with sample data that doesn’t need to have a **timestamp** evaluated, we can choose **Current Time** for Today.

 Click **Next 🡪**

**‘Name’ = ‘airports.csv’ 🡪**

**App Context = ‘Search & Reporting’ 🡪 Save**

We’re now on the input settings page. Here we can verify that the host value is correct and determine where to send our data. A best practice is to group our data into separate indexes by whatever category we want to apply—this could be vendor specific, related to compliance mandates, by operational or security use case, etc.

**Create New Index 🡪 airport\_information**

We’re fine with the defaults here, but note that we can change the paths, bucket size, and default app at this location.

**Review 🡪 Submit 🡪 Start Searching**

## Creating Different Searches

1. You should see your data displayed in the Splunk Search Bar. Now let’s try running a search. Delete the text in the search bar and enter

**\*small\_airport\***

Time Range should be 24 hours when searching!

‘\*’ is the Splunk wildcard feature, and if we look at our data we know that it contains the phrase ‘small\_airport’ many times over. But when we run the query, we get no results.

This is because we need to make a new index, Splunk will only search it if specifically told to do so. We need to tell Splunk that we want to search this index by default. To do so, please go to

**Settings 🡪 User and Authentication 🡪 Access Controls 🡪 Roles 🡪 Admin 🡪 Indexes 🡪 Searched by Default 🡪 Add airport\_information 🡪 Save**

****

Now when we run the same query (\*small\_airport\*), we get the results we’re looking for.

## Creating Customizable Dashboards

Let’s start building dashboards that show us what this data has to tell us. On the left-hand side of the Search & Reporting App we see a list of fields. Host, Source, and Sourcetype are all pulled for every source. The rest of these fields are fields that Splunk has recognized—in this case by delimiters. In cases where there are not necessarily delimiter, Splunk will still recognize most fields that are repeated over 20% of the time and automatically extract them. If there is a field that we know is contained in our data but isn’t pulled out, we can use Splunk’s RegEx wizard at the bottom left to extract a new field.

1. Let’s look at the average elevation by type of airport. We don’t want to look at closed airports since we’re not sure what category they would have fallen into if they were open, so we’ll make sure to exclude those.

**index=airport\_information type!= closed |chart avg(elevation\_ft) by type**

Now I have an easily digestible statistics table that shows me the average elevation of the different types of airports we have data on.

**Save As 🡪 Dashboard Panel**

**Dashboard Title = Airport Statistics**

**Panel Title = Elevation by Type 🡪 Save**

1. Next let’s learn a little more about the distribution of airports across the continents.

**sourcetype=”airports.csv” |stats count by continent**

 Make sure Time Range is All Time when searching!

Save this as a dashboard panel to our existing dashboard. Now that we look at it though, this isn’t very flashy. If we want to change this to something a little more eye-catching, like a visualization, we can easily do that without going back to the search bar.

**Edit 🡪 Select Visualization (top right in bar) 🡪 Pie Chart 🡪 Save**

Great! Unless I want to get really picky and think the colors are a little too boring for my taste. A little granular maybe, but Splunk lets us be as flexible as we want so why not take advantage of it?

**Edit 🡪 On ‘Edit Dashboard’ 🡪 Toggle from ‘UI’ to ‘Source’**

Here we can see and edit the simpleXML generating the visualization. There are a lot of great things we can do from here. Let’s start by standardizing the color of all our pie charts.

**Add**

**<option name="charting.fieldColors">{ "NA": 0xFF0000, "OC": 0xFF9900, "SA":0x009900, "AF":0xC4C4C0, "AN":0x000000, "AS":0xfff00, "EU":0000FF}</option>**

****

This should have changed the colors of your Pie Chart. Click **Save**. This is the kind of dashboard I was looking for! Only I’d like to be able to get a little more detail from here without having to go pick through all the data.

1. Let’s add a drilldown.

**Edit 🡪 Edit Drilldown 🡪 Link To Search 🡪 custom**

****

Now I can add a search string that will allow me to get more detail on each individual continent I want to dive into by adding the following to search string box:

**index=airport\_information continent=$row.continent$ | stats count by iso\_country**

**Time Range 🡪 Global 🡪 Check ‘Open in new tab’ 🡪 Save**

****

Try clicking on one of the pie chart sections now. You should be able to navigate to the new query we entered to get a breakdown by country! One thing though… that’s a lot of countries, and I have no idea what a lot of these abbreviations stand for.

## Creating a Lookup Table

1. Let’s create a lookup table so that I’m not having to open an encyclopedia every time I pull this information up.



**Settings 🡪 Lookup 🡪 Lookup table files 🡪 add new 🡪**

****

**destination app ‘search’ --> choose file ‘countries.csv’ --> Destination Filename ‘countrycodes.csv’ --> Save**

**Permission --> All Apps --> Roles --> Everyone**

****

****

**Lookups --> Lookup Definitions --> Add New -->**

****

**countrycodes --> Type File-based --> Lookup file countrycodes.csv --> Save**

* **Permission --> All Apps --> Roles --> Everyone**

****

****

Let’s just configure this to apply automatically so that I’m not going through the process of referencing every time. I already know any search I want to run I’m going to want to append the name to the abbreviation.

**Settings --> Lookups --> Automatic Lookups --> Add New -->**

****

**Destination App = Search**

**Name = ‘countrycodes’**

**Lookup Table = ‘countrycodes’**

**Apply to = ‘sourcetype’ ‘airports.csv.’**

countrycodes

countrycodes

****

* **Permission --> All Apps --> Roles --> Everyone**

****

Open up countries.csv and airports.csv on your desktop. Note that both documents have an ‘iso\_country’ fields. So for ‘Lookup input fields’ we want to note that **iso\_country = iso\_country** so that Splunk knows these two fields are equivalent between our data source and lookup file.

Now we want the output to be ‘country\_name’, and we’re fine with it reading as ‘country\_name’ since it’s a similar format to our existing fields. So for ‘Lookup output fields’, we can go ahead and enter

**country\_name = country\_name**

**Save**

1. Run the following query to verify your automatic lookup is working:

 **index=airport\_information |stats count by iso\_country, country\_name**

Perfect! Now we have one more step!

Let’s go to our **airport statistics dashboard, edit the drilldown**, and change the ‘iso\_country’ field in our drilldown to ‘country\_name’. **Save** and test your drilldown.

## Creating Customizable Dashboards 2.0

1. Let’s build another dashboard to illustrate more Splunk visualization capabilities. Our first dashboard looked at the elevation of airports by type, and our next looked at the distribution of airports across continents and countries. Let’s look at which countries have the highest airports and the greatest number of airports.

**index=airport\_information|stats count max(elevation\_ft) by country\_name**

The statistics table here isn’t much better than looking at the raw data. Line charts don’t work too well here either, but if we selected a scatter chart on the visualization tab, we get a great picture of the elevation and number of airports across all the countries we have data on. If we want to limit it to only those countries that have heavier traffic, we can use the where command to set a condition on this search.

**index=airport\_information|stats count max(elevation\_ft) by country\_name|where count>500**

**Save --> Dashboard Panel --> Existing --> Airport Statistics -->**

**Panel Title** = **Elevation & Count by Country**

**Save**

## Creating Customizable Dashboards 3.0

1. Finally, let’s close out by building one more dashboard to look at the distribution of types of airports across the world. Here we’ll use Splunk’s ability to recognize latitude and longitude to build a cluster map.

**index=airport\_information|geostats count by type latfield=latitude\_deg longfield=longitude\_deg**

**Select visualization --> cluster map.** Note that you can see more granular location by zooming in.

**Save As --> Dashboard Panel --> Existing --> Airport Statistics**

**Panel Title** **= Worldwide Type Distribution**

**Save**

Note that if we don’t like the way the panels appear on the screen, we can click edit and drag them into any position we want.