|  |  |
| --- | --- |
| **Server & Cloud Security**  Diploma in CSF/IT  Oct 2022 | Week 12 |
| Practical |
| **Practical: Splunk Search, Correlation and Attack Detection** | |

## Learning Objectives

You will learn how to use the Search app to add data to your Splunk deployment, search the data, and analyze the data in order to identify some basic attacks.

**Part 1: Installation & Preparation**

**1.1 Windows installation instructions**

For this practical you will install Splunk Enterprise using the default installation settings, which run the software as the Local System user, admin.

Firstly, download the Splunk installer from PoliteMall, or from the following Google Drive / One Drive links. **Transfer the installer to your MS1 server after downloading.**

Google Drive:

<https://drive.google.com/file/d/12vWuNV-jheKtoxUvQnkTnwP_eEHq5bUp/view?usp=sharing>

One Drive:

[splunk-9.0.3-dd0128b1f8cd-x64-release.msi](https://connectnpedu.sharepoint.com/:u:/s/CSF.SCS/EfyLvGGnsH1HiDPPRspVT5sBBkaoLEtjyHqQFpX1avhGJw?e=UEgnbY)

**You are recommended to install the Spunk Enterprise on your MS1 server.**

1. Double-click the installer.
2. In the Welcome panel, read the License Agreement and click **Check this box to accept the license agreement**.
3. Click **Next**.
4. A terminal window appears and you are prompted to specify a User ID and password to use with the Splunk Trial.

The password must be at least 8 characters in length. The cursor will not advance as you type.

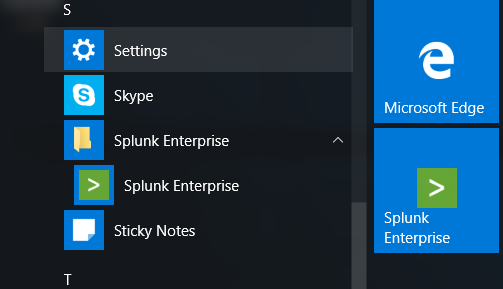
**Make note of the User ID and password. You will use these credentials to login Splunk Enterprise.**

1. Confirm that the **Launch browser with Splunk** check box is selected.
2. The installation finishes, Splunk Enterprise starts, and Splunk Web launches in a browser window.

**1.2 Login to Splunk Web**

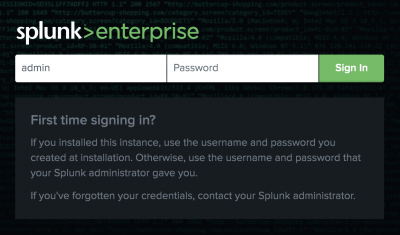
Open your browser and use the URL to access Splunk Web is <http://localhost:8000>.

Or your can click the shortcut icon to launch it.



When you launch Splunk Enterprise for the first time, this login screen appears.

Login using the username and the password that you specified when you installed Splunk Enterprise.

**[](https://docs.splunk.com/File:7.3.0_first_time_logon-compressor.png)**

## 1.3 Download the tutorial data files

|  |
| --- |
| Understand the Data Files |
| This lab practical uses a fictitious game store, called Buttercup Games, that sells games and related items in an online store.  The data files contain web access log files, secure formatted log files, and sales log files. |

[Download](http://docs.splunk.com/images/Tutorial/tutorialdata.zip) the tutorialdata.zip file from the below link. Do not un-compress the file.

<http://docs.splunk.com/images/Tutorial/tutorialdata.zip>

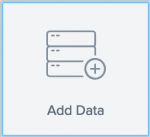
## 1.4 Understanding the Data in Splunk

The Splunk platform accepts any type of data. In particular, it works with all IT streaming and historical data. The source of the data can be event logs, web logs, live application logs, network feeds, system metrics, change monitoring, message queues, archive files, and so on.

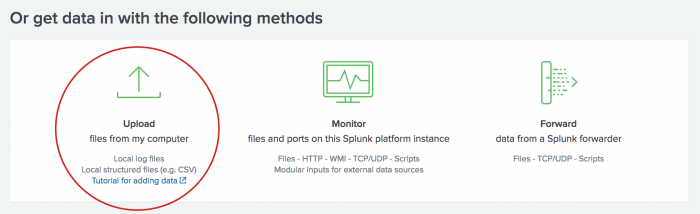
|  |
| --- |
| Knowledge & Tips |
| The process of transforming the data is called indexing. During indexing, the incoming data is processed to enable fast searching and analysis. The processed results are stored in the index as events.  Events are stored in the index as a group of files that fall into two categories:   * Raw data, which is the data that you add to the Splunk deployment. The raw data is stored in a compressed format. * Index files, which include some metadata files that point to the raw data. |

## 1.5 Upload the tutorial data

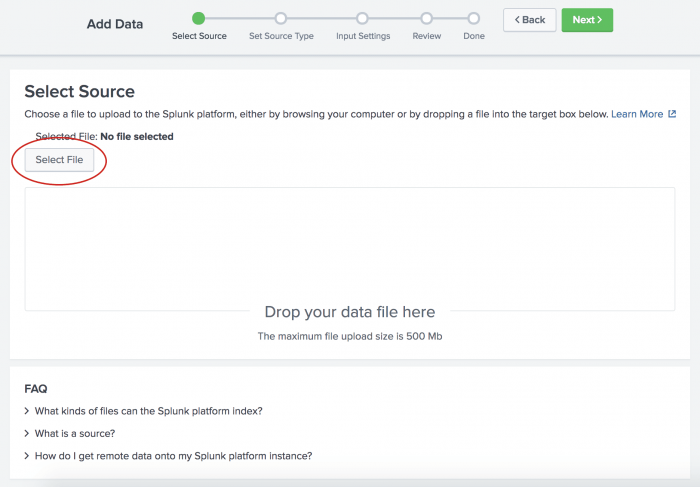
1. Click **Settings > Add Data**.

[](https://docs.splunk.com/File:7.2.0_adddata-compressor.png)

1. At the bottom of the window, click **Upload**.

[](https://docs.splunk.com/File:7.2.0_adddata_upload-compressor.png)

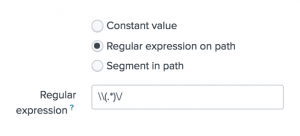
1. Under **Select Source**, click **Select File**.

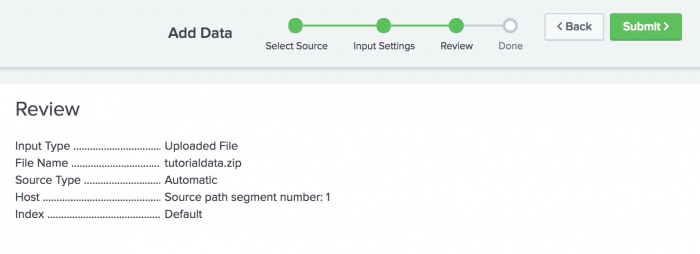
[](https://docs.splunk.com/File:7.2.0_adddata_selectsource-compressor.png)

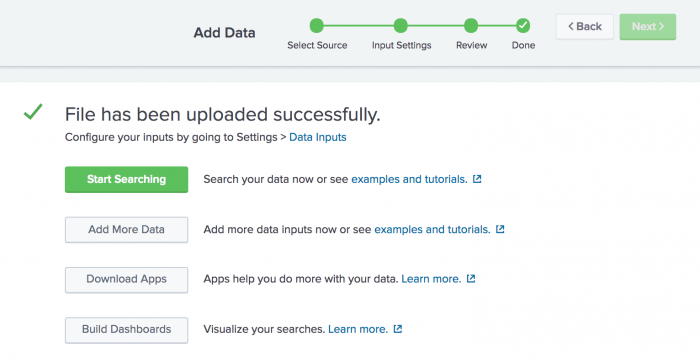
1. In your download directory, select the tutorialdata.zip file and click **Open**.
2. Click **Next** to continue to **Input Settings**.
3. Under **Input Settings**, you can override the default settings for Host, Source type, and Index.  
   Because this tutorial uses a ZIP file, you are going to modify the **Host** setting to assign the host values by using a portion of the path name for the files included in the ZIP file.

a. Select **Regular expression on path**.

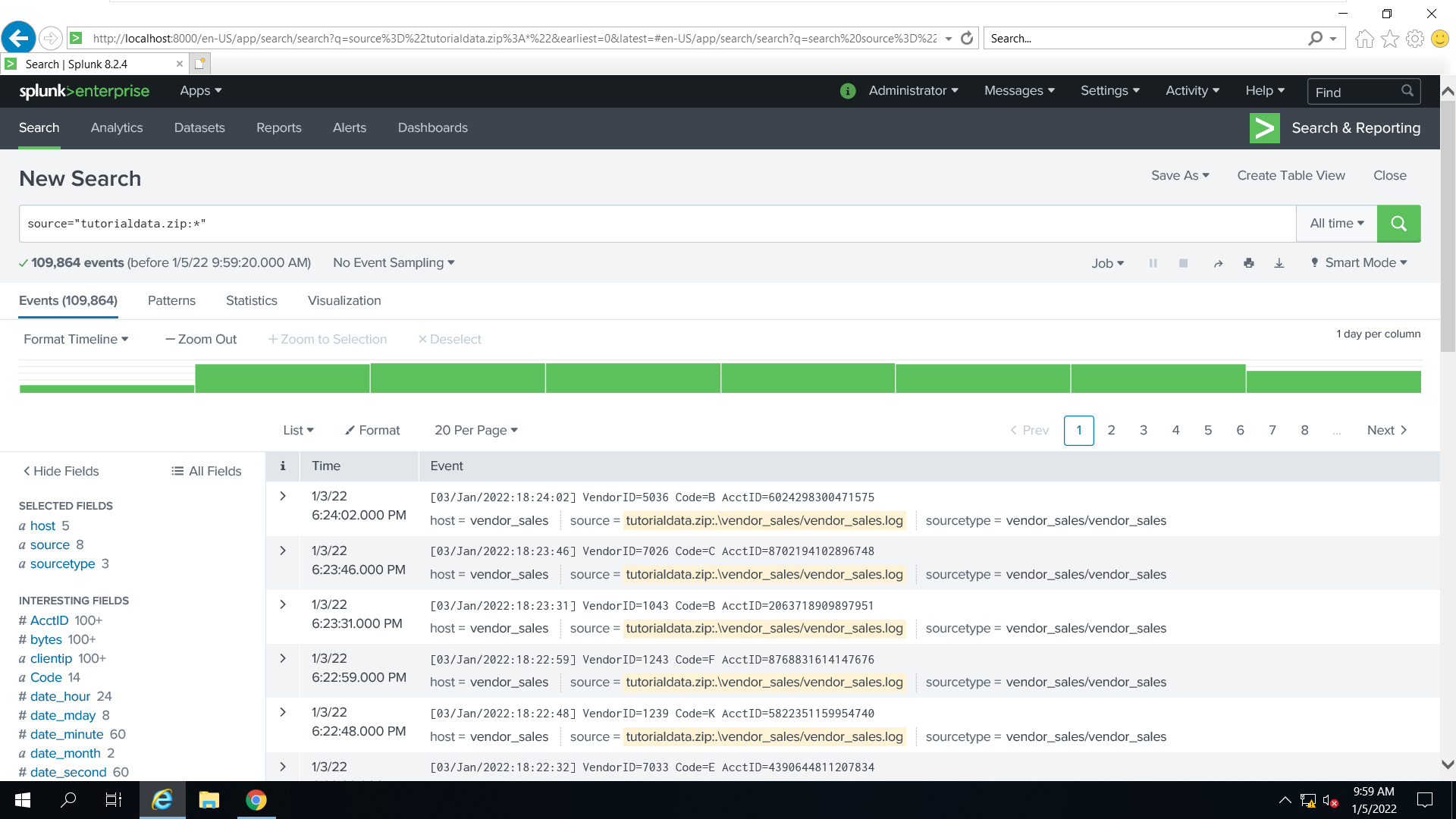
b. Type  for the regex to extract the host values from the path.

[](https://docs.splunk.com/File:7.1.0_adddata_inputsettings2-compressor.png)

1. Click **Review**. The following screen appears where you can review your input settings.  
   [](https://docs.splunk.com/File:7.1.0_adddata_review-compressor.png)
2. Click **Submit** to add the data.

[](https://docs.splunk.com/File:7.1.0_adddata_done-compressor.png)

1. To see the data in the Search app, click **Start Searching**. You might see a screen asking if you want to take a tour. You can take the tour or click **Skip**. The Search app opens and a search is automatically run on the tutorial data source.

  
  
Success! The results confirm that the data in the tutorialdata.zip file was indexed and that events were created.

1. Click the **Splunk** logo to return to Splunk Home.

## 1.6 Time ranges and the tutorial data

When you run a search using the tutorial data, if no events are returned, it is probably because you downloaded the tutorialdata.zip file more than one day ago. When you download the ZIP file, timestamps are generated at that moment in time and are added to the data.

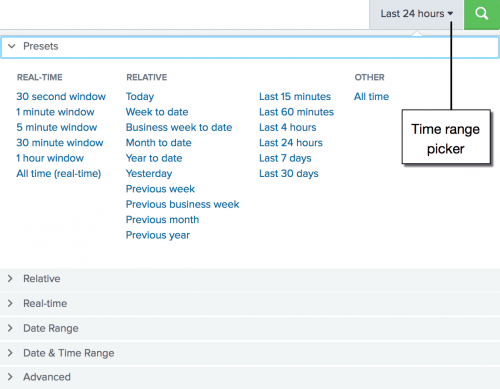
## Preset time ranges

The time range picker has many preset time ranges that you can select from. Click the time range picker to see a list of the time range options.

The **Presets** option contains **Real-time**, **Relative**, and **Other** time ranges.

* + **Real-time searches** display a live, streaming view of events. You can specify a window over which to retrieve events.
  + **Historical searches** display events from the past. You can restrict your search by specifying a relative time range or a specific date and time range.

Because the data for the Buttercup Games online store is a snapshot of historical data, you will not use the '''Real-time''' preset time ranges in this tutorial.

[](https://docs.splunk.com/File:7.1.0_timerange_presets-compressor.png)

**Part 2: Understanding the Basic Search and Field Search**

# **2.1 Basic searches and search results**

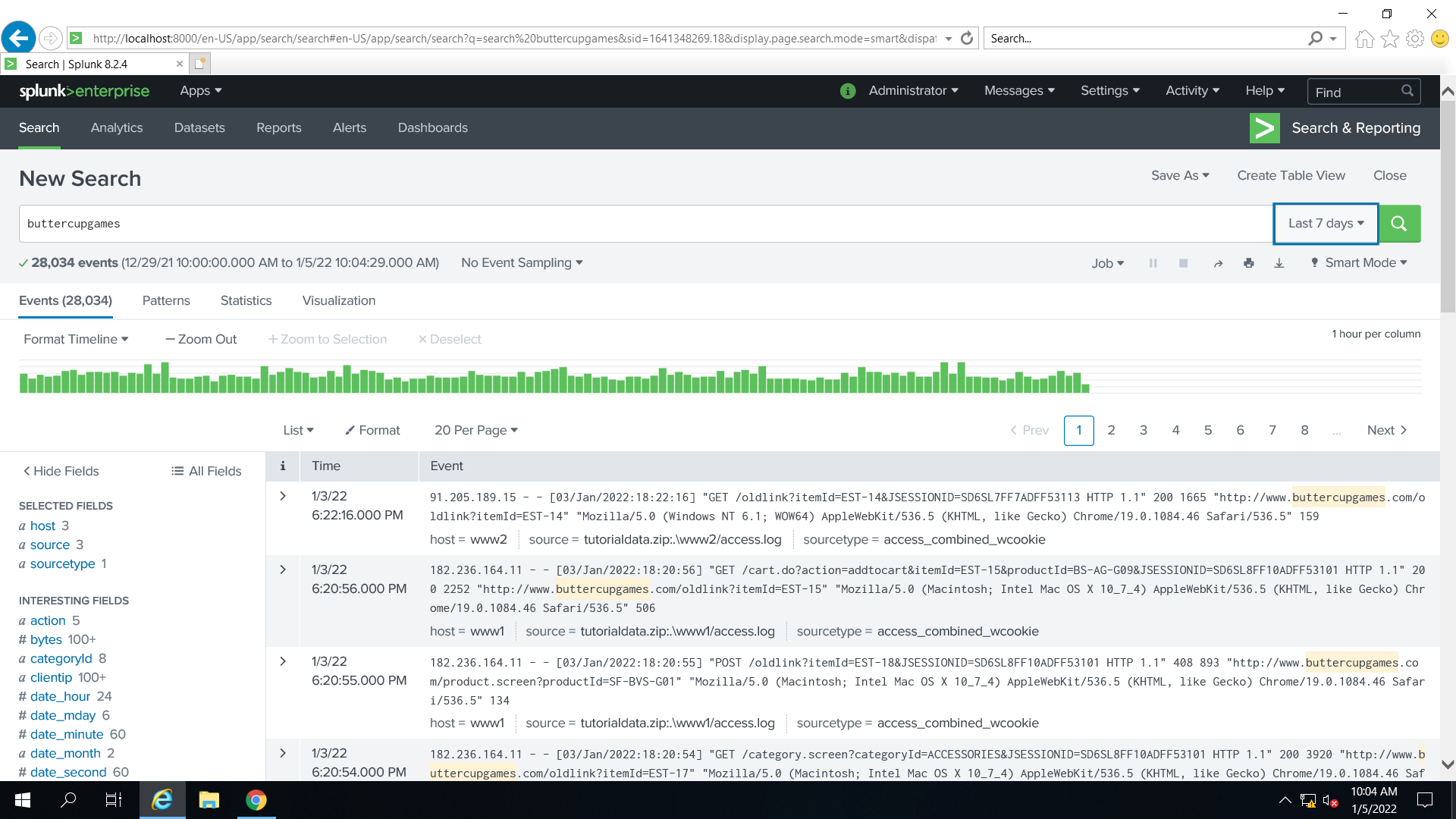
In this section, you create searches that retrieve events from the index.

The Search Assistant is a feature in the Search app that appears as you type your search criteria. The Search Assistant is like autocomplete, but so much more.

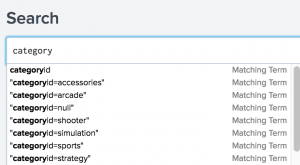
1. Click **Search & Reporting** in the App bar to start a new search.
2. Type **buttercupgames** in the Search bar.

When you type a few letters into the Search bar, the Search Assistant shows you terms in your data that match the letters that you type in.

**If no result is displayed, you can adjust the time range until you see the results.**



1. Click **Search** in the App bar to start a new search.
2. Type **category** in the Search bar. The terms that you see are in the tutorial data.

**[](https://docs.splunk.com/File:7.1.0_tutorial_searchassistant1-compressor.png)**

1. Select **"categoryid=sports"** from the Search Assistant list.
2. Press **Enter**, or click the **Search icon** on the right side of the Search bar, to run the search.

# **2.2 Use fields to search**

To take advantage of the advanced search features in the Splunk software, you must understand what fields are and how to use them.

|  |
| --- |
| Knowledge & Tips |
| What are fields?  Fields exist in machine data in many forms. Often, a field is a value with a fixed, delimited position on a line, or a name and value pair, where there is a single value to each field name. |

A field can be multivalued, that is, a field in a single event can have multiple values in a field.

* Some examples of fields are clientip for IP addresses accessing your Web server, \_time for the timestamp of an event, and host for domain name of a server.
* One of the more common examples of multivalue fields is email address fields. While the From field will contain only a single email address, the To and Cc fields have one or more email addresses associated with them.

## Search with fields

When you search for fields, you use the syntax field\_name=field\_value.

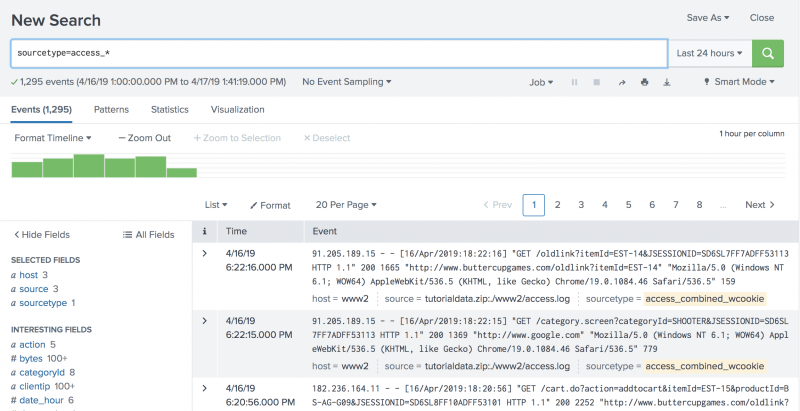
* Field names are case sensitive, but field values are not.
* You can use wildcards in field values.
* Quotation marks are required when the field values include spaces.

Let's try a search.

1. Click **Search** in the App bar to start a new search. Notice that the time range is set back to the default **Last 24 hours**. If your search returns no result, you can change the time to a wider range such as “All Times”.
2. To search the **sourcetype** field for any values that begin with **access\_**, run the following search.

sourcetype=access\_\*

This search indicates that you want to retrieve only events from your web access logs and nothing else.

[](https://docs.splunk.com/File:7.3.0_searchfields-compressor.png)

1. Scroll through the list of events in your search results.

**Part 3: Events Correlation using Subsearch**

|  |
| --- |
| Knowledge & Tips |
| * Event correlation is a technique for making sense of a large number of events and pinpointing the few events that are really important in that mass of information. This is accomplished by looking for and analyzing relationships between events. * Event correlation tools are a fundamental instrument in your toolbox to detect threats from all sources across your organization. * A subsearch is a search that is used to narrow down the set of events that you search on. The result of the subsearch is then used as an argument to the primary, or outer, search. Subsearches are enclosed in square brackets within a main search and are evaluated first. |

Let's find the single most frequent shopper on the Buttercup Games online store, and what that shopper has purchased.

## Example 1: Search without a subsearch

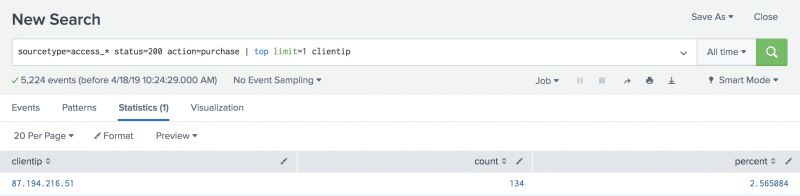
You want to find the single most frequent shopper on the Buttercup Games online store and what that shopper has purchased. Use the top command to return the most frequent shopper.

1. Start a new search.
2. Change the time range to **All time**.
3. To find the shopper who accessed the online shop the most, use this search.

sourcetype=access\_\* status=200 action=purchase | top limit=1 clientip

The limit=1 argument specifies to return 1 value. The clientip argument specifies the field to return.

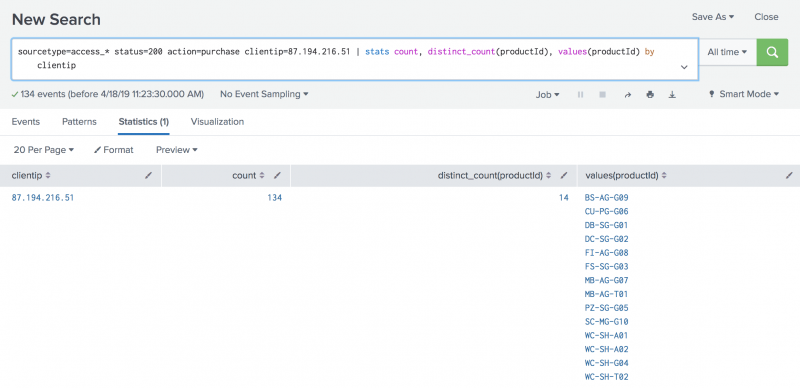
(Note: Status 200 OK is the Standard response for successful HTTP requests.)

[](https://docs.splunk.com/File:7.3.0_nosubsearch1-compressor.png)  
  
This search returns one clientip value, 87.194.216.51, which you will use to identify the VIP shopper. The search also returns a count and a percent. These are the default fields that are returned with the top command.

1. You now need to run another search to determine how many different products the VIP shopper has purchased. Use the stats command to count the purchases by this VIP customer.

sourcetype=access\_\* status=200 action=purchase clientip=87.194.216.51 | stats count, distinct\_count(productId), values(productId) by clientip

This search uses several [statistical functions](http://docs.splunk.com/Documentation/Splunk/8.0.2/SearchReference/CommonStatsFunctions) with the stats command. An alias for the distinct\_count() function is dc().

[](https://docs.splunk.com/File:7.3.0_nosubsearch2-compressor.png)

This search uses the count() function to return the total count of the purchases for the VIP shopper. The dc() function is the distinct\_count function. Use this function to count the number of different, or unique, products that the shopper bought. The values function is used to display the distinct product IDs as a multivalue field.

|  |
| --- |
| Knowledge & Tips |
| Example 1 shows how to find the most frequent shopper without a subsearch.  The drawback to this approach is that you have to run two searches each time you want to build this table. The top purchaser is not likely to be the same person in every time range.  A better approach is demonstrated below in Example 2 which shows how to find the most frequent shopper with a subsearch. |

## Example 2: Search with a subsearch

Let's start with our first requirement, to identify the single most frequent shopper on the Buttercup Games online store.

1. Copy and paste the following search into the Search bar and run the search. Make sure the time range is **All time**.

sourcetype=access\_\* status=200 action=purchase | top limit=1 clientip | table clientip

This search returns the clientip for the most frequent shopper, clientip=87.194.216.51. This search is almost identical to the search in Example 1 Step 1. The difference is the last piped command, | table clientip, which displays the clientip information in a table. Because you specified only the clientip field with the table command, that is the only field returned. The count and percent fields that the top command generated are discarded from the output.

To find what this shopper has purchased, you run a search on the same data. You provide the result of the most frequent shopper search as one of the criteria for the purchases search.

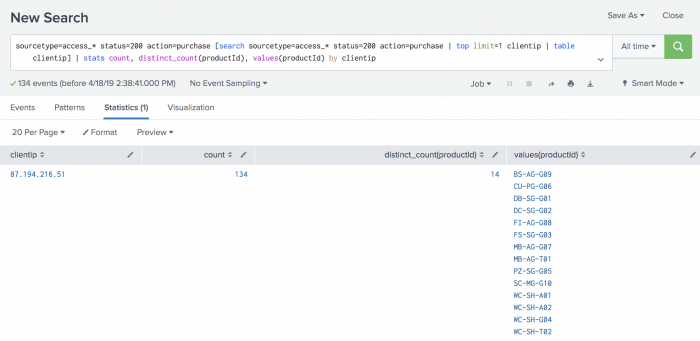
The most frequent shopper search becomes the **subsearch** for the purchases search. The purchases search is referred to as the **outer** or primary search. Because you are searching the same data, the beginning of the outer search is identical to the beginning of the subsearch.

A subsearch is enclosed in square brackets [ ] and processed first when the search criteria are parsed.

1. Copy and paste the following search into the Search bar and run the search.

sourcetype=access\_\* status=200 action=purchase [search sourcetype=access\_\* status=200 action=purchase | top limit=1 clientip | table clientip] | stats count, distinct\_count(productId), values(productId) by clientip

Because the top command returns the **count** and **percent** fields, the table command is used to keep only the clientip value.

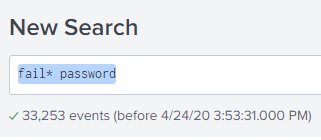
[](https://docs.splunk.com/File:7.3.0_subsearch1-compressor.png)

These results should match the result of the two searches in Example 1, if you run it on the same time range. If you change the time range, you might see different results because the top purchasing customer will be different.

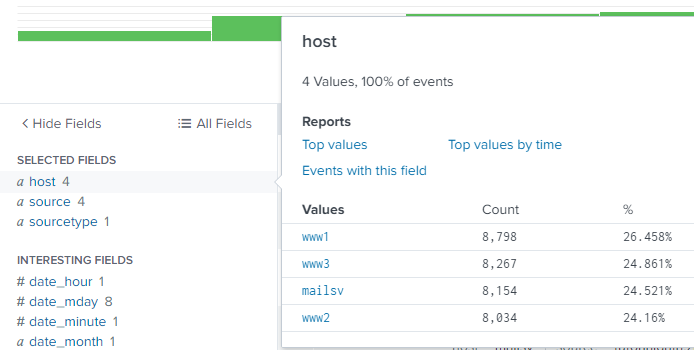
**Part 4: Attack Detection**

|  |
| --- |
| Knowledge & Tips |
| * Splunk enables rapid security investigation and analysis. It can provide the data platform and security analytics capabilities needed to allow organizations to monitor, alert, analyze, investigate, respond, share, and detect known and unknown threats regardless of organizational size or skillset. |

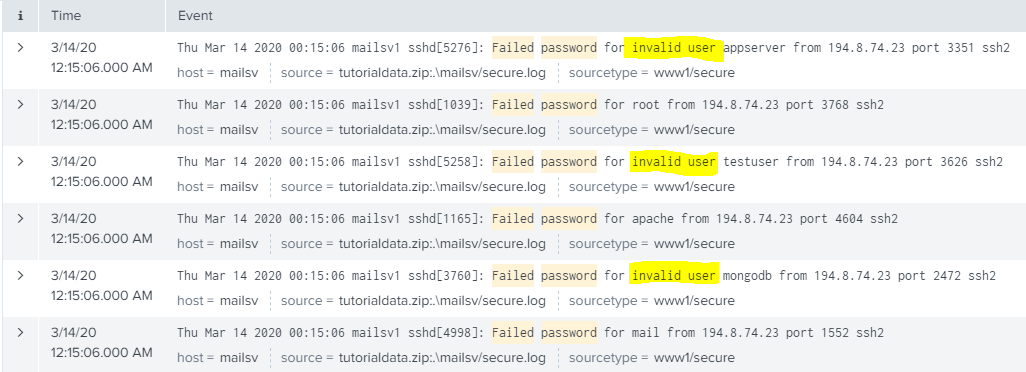
1. Start a new search and type **fail\* password**, change the time period to “All time”.



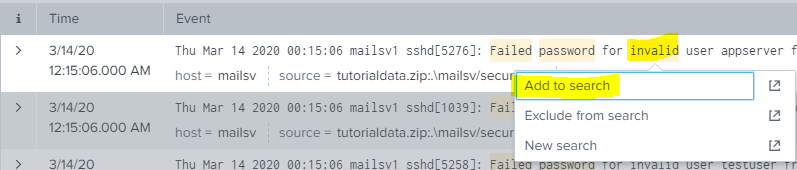
1. Under the “Selected Field”, click the “host” and you can see a summary count of failed password attempts on respective servers.



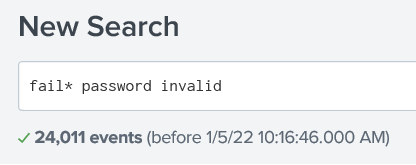
1. From the event details, you may notice that some failed password attempts are for valid users such as “root” or “apache”, while some other failed password attempts are even associated with invalid user account, as highlighted in yellow below.



1. This may be an indication of brute force attack as the attacker may try out many possible user IDs and password combinations.
2. Mouse over to the word “invalid” and double click it. Select “Add to search”, we will fine tune the search to only invalid users.



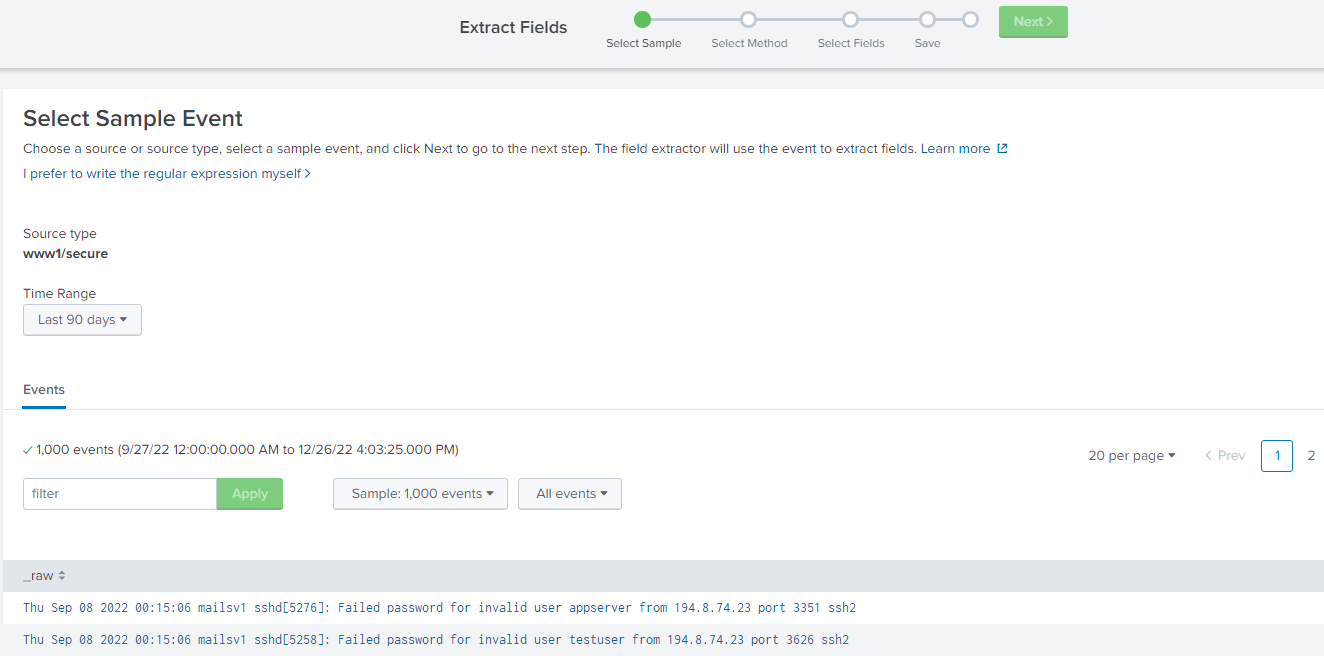
1. After you added “invalid” to search, your search syntax became the below. Alternatiely, you can directly copy this line **fail\* password invalid** inside the search.



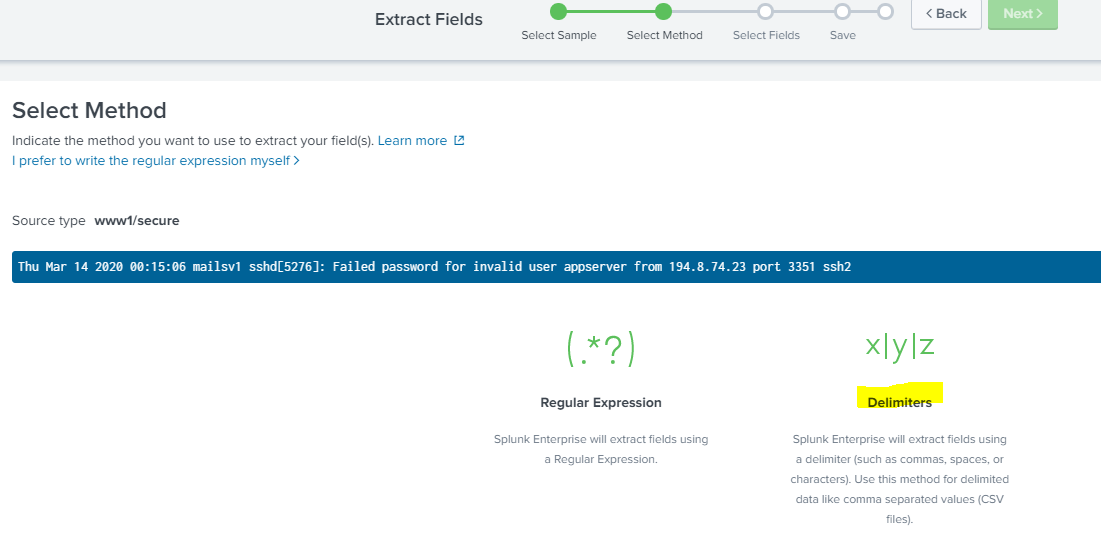
1. On the left hand side of the page, select “Extract New Fields”.



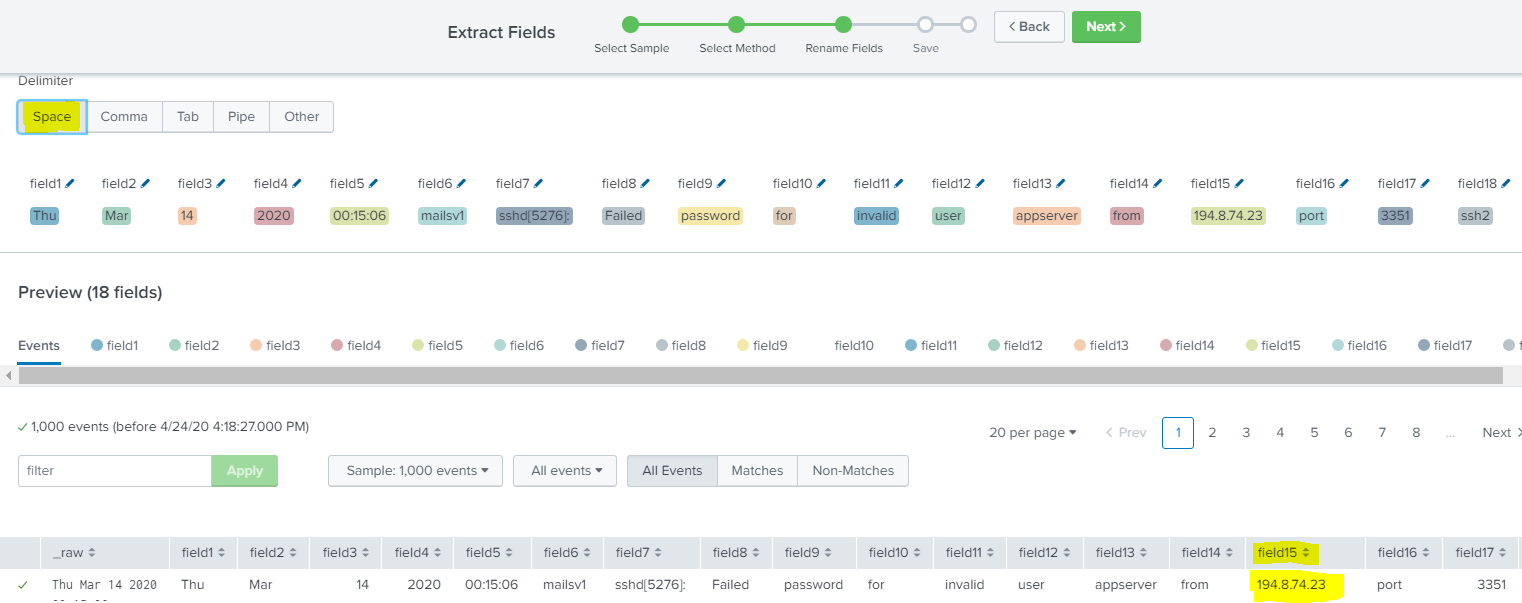
1. Click any line as the sample event. Then click Next.



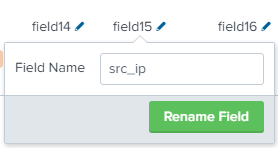
1. Click “delimiters” as the method to extract fields then click Next.



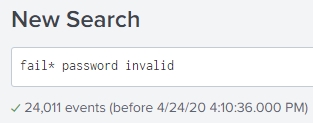
1. Select “Space” to separate the fields and now your sample event is divided into multiple fields. There will be one field containing the IP address, in this example, its field 15.



1. Rename the field with IP address to “src\_ip” or any name easy for you to recognize.



1. Click Next and save your customized extraction with a name.
2. Continue your previous search by copying this line **fail\* password invalid** inside the search.



1. The “src\_ip” created by you is now your interested field. Click on it and you can find out the top IP addresses that accessing your server with the high count of wrong passwords.



1. These IP addresses are potentially from attackers conducting brute force attack. These IP addresses should be reported for further investigation and be blocked by Firewall or border router.

== End of Practical ==