



splunk >

tutorialspoint

SIMPLY EASY LEARNING

www.tutorialspoint.com



<https://www.facebook.com/tutorialspointindia>



<https://twitter.com/tutorialspoint>

About the Tutorial

Splunk is a software used to search and analyze machine data. This machine data can come from web applications, sensors, devices or any data created by user. It serves the needs of IT infrastructure by analyzing the logs generated in various processes but it can also analyze any structured or semi-structured data with proper data modelling. It has built-in features to recognize the data types, field separators and optimize the search processes. It also provides data visualization on the search results.

Audience

This tutorial targets IT professionals, students, and IT infrastructure management professionals who want a solid grasp of essential Splunk concepts. After completing this tutorial, you will achieve intermediate expertise in Splunk, and easily build on your knowledge to solve more challenging problems.

Prerequisites

The reader should be familiar with querying language like SQL. General knowledge in typical operations in using computer applications like storing and retrieving data and reading the logs generated by computer programs will be an highly useful.

Copyright & Disclaimer

© Copyright 2019 by Tutorials Point (I) Pvt. Ltd.

All the content and graphics published in this e-book are the property of Tutorials Point (I) Pvt. Ltd. The user of this e-book is prohibited to reuse, retain, copy, distribute or republish any contents or a part of contents of this e-book in any manner without written consent of the publisher.

We strive to update the contents of our website and tutorials as timely and as precisely as possible, however, the contents may contain inaccuracies or errors. Tutorials Point (I) Pvt. Ltd. provides no guarantee regarding the accuracy, timeliness or completeness of our website or its contents including this tutorial. If you discover any errors on our website or in this tutorial, please notify us at contact@tutorialspoint.com

Table of Contents

About the Tutorial	ii
Audience.....	ii
Prerequisites.....	ii
Copyright & Disclaimer	ii
Table of Contents	iii
1. Splunk – Overview	1
Product Categories	1
Splunk Features	1
2. Splunk – Environment	3
Linux Version	3
Windows Version.....	6
3. Splunk – Interface	9
Administrator Link	9
Settings Link.....	10
Search and Reporting Link	11
4. Splunk – Data Ingestion	13
Selecting Source Type.....	14
Input Settings	15
Review Settings	17
5. Splunk – Source Types.....	19
Supported Source Types.....	19
Source Type Sub-Category.....	20
Pre-Trained Source Types.....	21
6. Splunk – Basic Search.....	22
Combining Search Terms.....	23
Using Wild Card	24

Refining Search Results	25
7. Splunk – Field Searching.....	27
Choosing the Fields.....	28
Field Summary	29
Using Fields in Search	30
8. Splunk – Time Range Search	31
Selecting a Time Subset.....	32
Earliest and Latest	33
9. Splunk – Sharing Exporting	35
Sharing the Search Result.....	35
Finding the Saved Results.....	36
Exporting the Search Result	37
10. Splunk – Search Language	39
Components of SPL.....	39
11. Splunk – Search Optimization	44
Analysing Search Optimisations	44
Turning Off Optimization.....	46
12. Splunk – Transforming Commands.....	49
Examples of Transforming Commands.....	49
13. Splunk – Reports.....	53
Report Creation	53
Report Configuration	54
Modifying Report Search Option.....	56
14. Splunk – Dashboards.....	58
Creating Dashboard	58
Adding Panel to Dashboard	60
15. Splunk – Pivot and Datasets.....	64

Creating a Dataset	64
Selecting a Dataset	64
Choosing Dataset Fields.....	65
Creating Pivot	67
Choose the Pivot Fields	68
16. Splunk – Lookups	70
Steps to Create and Use Lookup File	70
17. Splunk – Schedules and Alerts.....	77
Creating a Schedule	77
Schedule Actions	79
Alerts	79
18. Splunk – Knowledge Management.....	84
Knowledge Object	84
Uses of Knowledge Objects	84
19. Splunk – Subsearching	86
Example	86
20. Splunk – Search Macros	89
Macro Creation.....	89
Macro Scenario.....	90
Defining the Macro.....	90
Using the Macro	92
21. Splunk – Event Types	94
Creating Event Type.....	94
Using New Event Types	96
Viewing the Event Type	98
Using the Event Type	100
22. Splunk – Basic Chart.....	101

Creating Charts	102
Changing the Chart Type	103
Formatting a Chart	104
23. Splunk – Overlay Chart.....	105
Chart Scenario	105
Creating Chart Overlay	107
24. Splunk – Sparklines	110
Selecting the Fields.....	110
Creating the Sparkline	111
Changing the Time Period	112
25. Splunk – Managing Indexes.....	113
Checking Indexes	113
Creating a New Index	115
Indexing the Events	116
26. Splunk – Calculated Fields.....	118
Example	118
Using the eval Function	119
Adding New Fields	120
Displaying the calculated Fields.....	120
27. Splunk – Tags	122
Creating Tags	123
Search Using Tags	124
28. Splunk – Apps	126
Listing Splunk Apps.....	126
App Permissions	127
App Marketplace	128
29. Splunk – Removing Data	130

Assigning Delete Privilege.....	130
Identifying the data to be removed	131
Deleting the Selected Data	132
30. Splunk – Custom Chart.....	135
Axis Customization	136
Legend Customization	136
31. Splunk – Monitor Files	138
Add files to Monitor	138
32. Splunk – Sort Command.....	142
Sorting by Field Types.....	142
Sorting up to a Limit	143
Using Reverse	145
33. Splunk – Top Command	146
Top Values for a Field	146
Top Values for a Field by a Field	147
Show Options	148
34. Splunk – Stats Command	149
Finding Average	149
Finding Range	150
Finding Mean and Variance	151

1. Splunk – Overview

Splunk is a software which processes and brings out insight from machine data and other forms of big data. This machine data is generated by CPU running a webserver, IOT devices, logs from mobile apps, etc. It is not necessary to provide this data to the end users and does not have any business meaning. However, they are extremely important to understand, monitor and optimize the performance of the machines.

Splunk can read this unstructured, semi-structured or rarely structured data. After reading the data, it allows to search, tag, create reports and dashboards on these data. With the advent of big data, Splunk is now able to ingest big data from various sources, which may or may not be machine data and run analytics on big data.

So, from a simple tool for log analysis, Splunk has come a long way to become a general analytical tool for unstructured machine data and various forms of big data.

Product Categories

Splunk is available in three different product categories as follows:

- **Splunk Enterprise:** It is used by companies which have large IT infrastructure and IT driven business. It helps in gathering and analysing the data from websites, applications, devices and sensors, etc.
- **Splunk Cloud:** It is the cloud hosted platform with same features as the enterprise version. It can be availed from Splunk itself or through the AWS cloud platform.
- **Splunk Light:** It allows search, report and alert on all the log data in real time from one place. It has limited functionalities and features as compared to the other two versions.

Splunk Features

In this section, we shall discuss the important features of enterprise edition:

Data Ingestion

Splunk can ingest a variety of data formats: JSON, XML and unstructured machine data such as web and application logs. The unstructured data can be modeled into a data structure by the user as and when needed.

Data Indexing

The ingested data is indexed by Splunk for faster searching and querying on different conditions.

Data Searching

Searching in Splunk involves using the indexed data for the purpose of creating metrics, predicting future trends and identifying patterns in the data.

Using Alerts

Splunk alerts can be used to trigger emails or RSS feeds when some specific criteria are found in the data being analyzed.

Dashboards

Splunk Dashboards can show the search results in the form of charts, reports and pivots, etc.

Data Model

The indexed data can be modelled into one or more data sets that is based on specialized domain knowledge. This leads to easier navigation by the end users who analyze the business cases without learning the technicalities of the search processing language used by Splunk.

2. Splunk – Environment

In this tutorial, we will aim to install the enterprise version. This version is available for a free evaluation for 60 days with all features enabled. You can download the setup using the below link which is available for both windows and Linux platforms.

[https://www.splunk.com/en_us/download/splunk-enterprise.html.](https://www.splunk.com/en_us/download/splunk-enterprise.html)

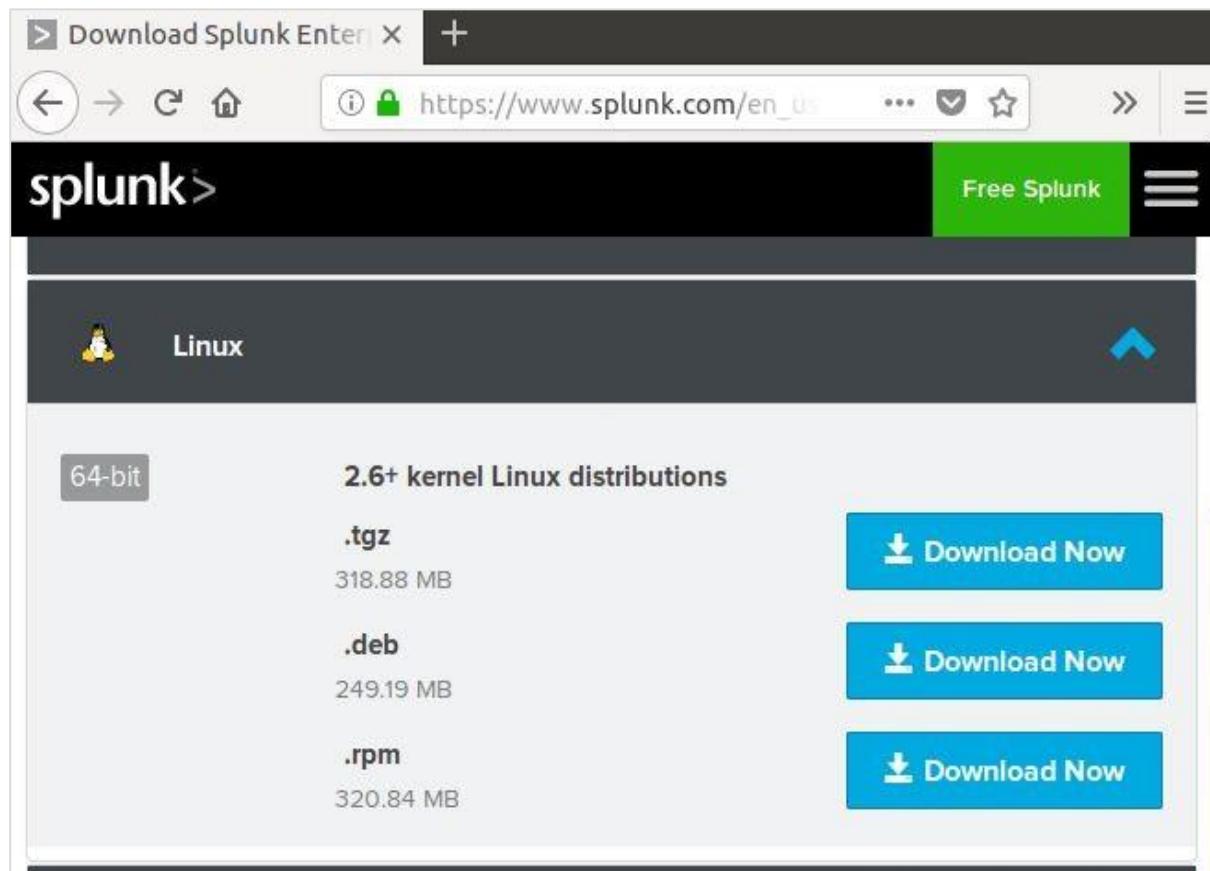
Linux Version

The Linux version is downloaded from the download link given above. We choose the .deb package type as the installation will be done in a Ubuntu platform.

We shall learn this with a step by step approach:

Step 1

Download the .deb package as shown in the screenshot below:



Step 2

Go to the download directory and install Splunk using the above downloaded package.

```
ubuntutrain@ubuntu:~/Downloads
ubuntutrain@ubuntu:~/Downloads$ ls
splunk-7.2.0-8c86330ac18-linux-2.6-amd64.deb
ubuntutrain@ubuntu:~/Downloads$ sudo dpkg -i splunk-7.2.0-8c86330ac18-
-linu...6-amd64.deb
[sudo] password for ubuntutrain:
Selecting previously unselected package splunk.
(Reading database ... 176940 files and directories currently installed.)
Preparing to unpack splunk-7.2.0-8c86330ac18-linux-2.6-amd64.deb ...
Unpacking splunk (7.2.0) ...
Setting up splunk (7.2.0) ...
complete
ubuntutrain@ubuntu:~/Downloads$ █
```

Step 3

Next, you can start Splunk by using the following command with accept license argument. It will ask for administrator user name and password which you should provide and remember.

```
ubuntutrain@ubuntu:/opt/splunk/bin$ sudo ./splunk start --accept-lic
ense
This appears to be your first time running this version of Splunk.

Splunk software must create an administrator account during startup.
Otherwise, you cannot log in.
Create credentials for the administrator account.
Characters do not appear on the screen when you type in credentials
.

Please enter an administrator username: admin
Password must contain at least:
    * 8 total printable ASCII character(s).
Please enter a new password:
```

Step 4

The Splunk server starts and mentions the URL where the Splunk interface can be accessed.

```

Starting splunk server daemon (splunkd)...
Generating a 2048 bit RSA private key
...Terminal .+++
.....+writing new private key to 'privKeySecure.pem'
-----
Signature ok
subject=/CN=ubuntu/O=SplunkUser
Getting CA Private Key
writing RSA key
Done

Waiting for web server at http://127.0.0.1:8000 to be available....
..... Done

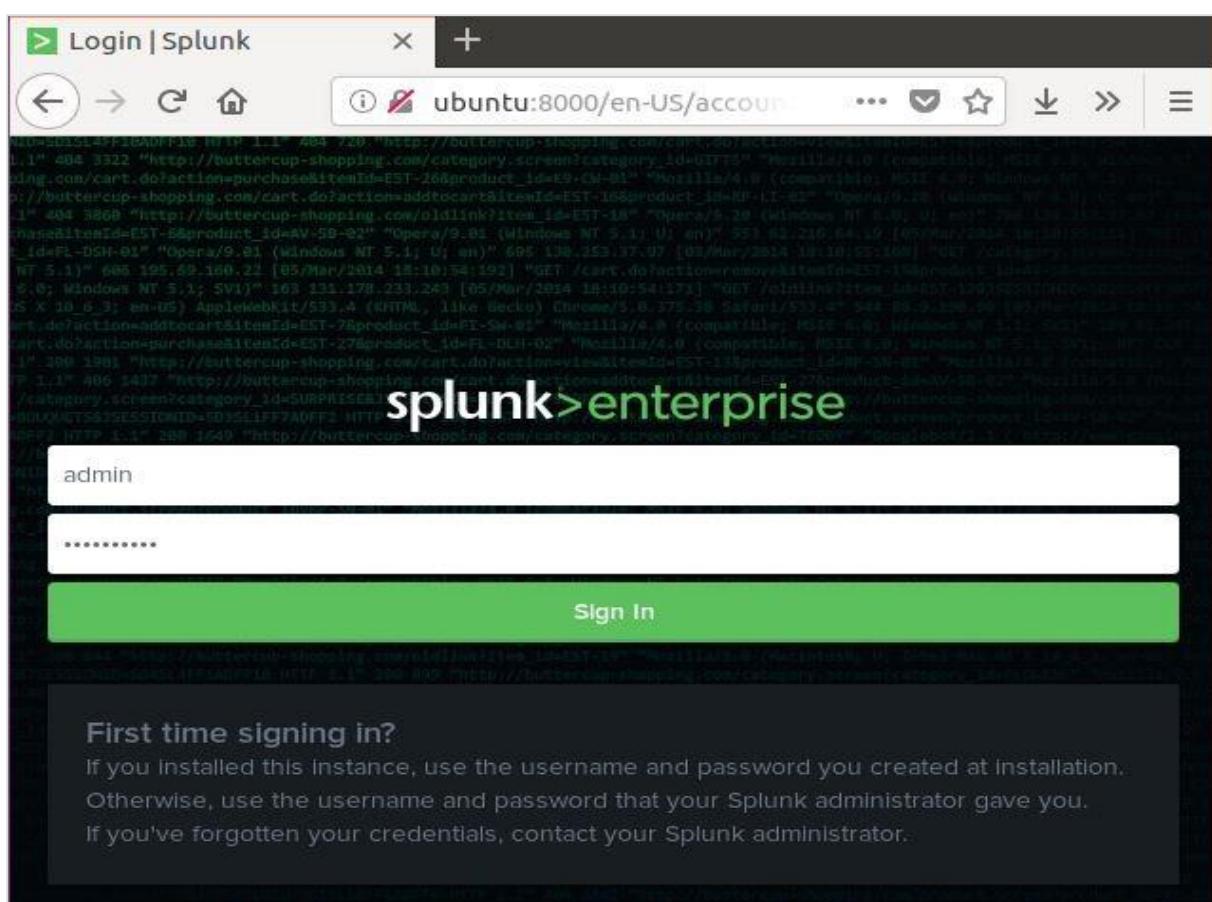
If you get stuck, we're here to help.
Look for answers here: http://docs.splunk.com

The Splunk web interface is at http://ubuntu:8000

```

Step 5

Now, you can access the Splunk URL and enter the admin user ID and password created in step 3.



Windows Version

The windows version is available as a msi installer as shown in the below image:

The screenshot shows the Splunk website's "Choose Your Installation Package" section. At the top, there is a navigation bar with the Splunk logo, a "Free Splunk" button, and a menu icon. Below the navigation bar, there is a heading "Choose Your Installation Package". The main content area displays three sections: 1) Windows (64-bit), 2) Windows (32-bit), 3) Linux, and 4) Mac OS. Each section includes a small icon, the operating system name, a "Download Now" button, and a "More Options" dropdown arrow.

Platform	Architecture	Operating Systems	File Type	File Size	Action
Windows	64-bit	Windows 8.1, and 10 Windows Server 2012, 2012 R2, and 2016	.msi	224.09 MB	Download Now
	32-bit	Windows 8.1, and 10	.msi	196.18 MB	Download Now
Linux					
Mac OS					

Double clicking on the msi installer installs the Windows version in a straight forward process. The two important steps where we must make the right choice for successful installation are as follows.

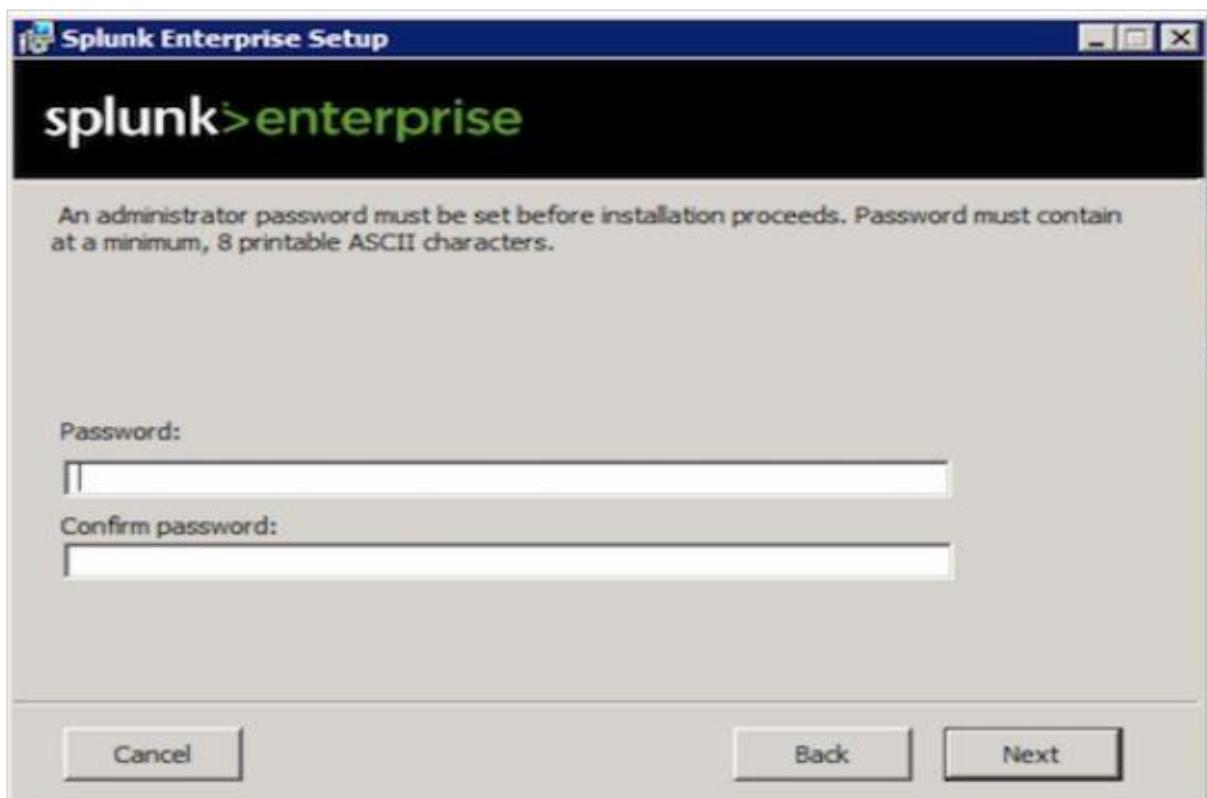
Step 1

As we are installing it on a local system, choose the local system option as given below:



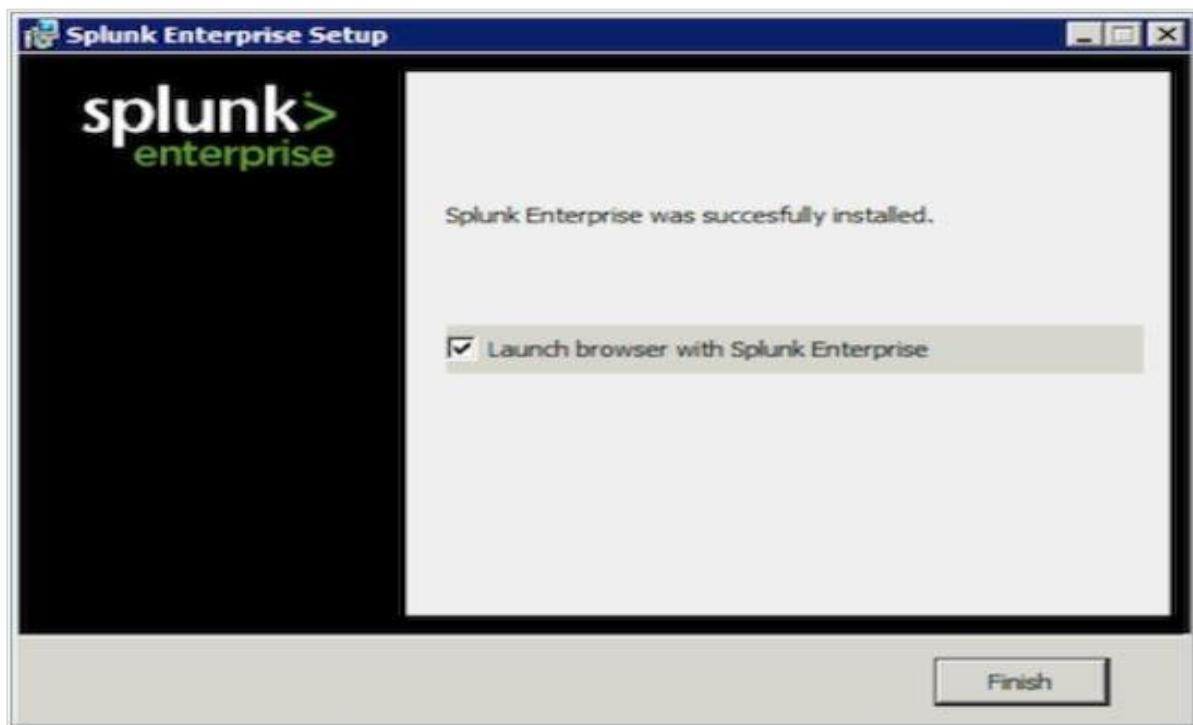
Step 2

Enter the password for the administrator and remember it, as it will be used in the future configurations.



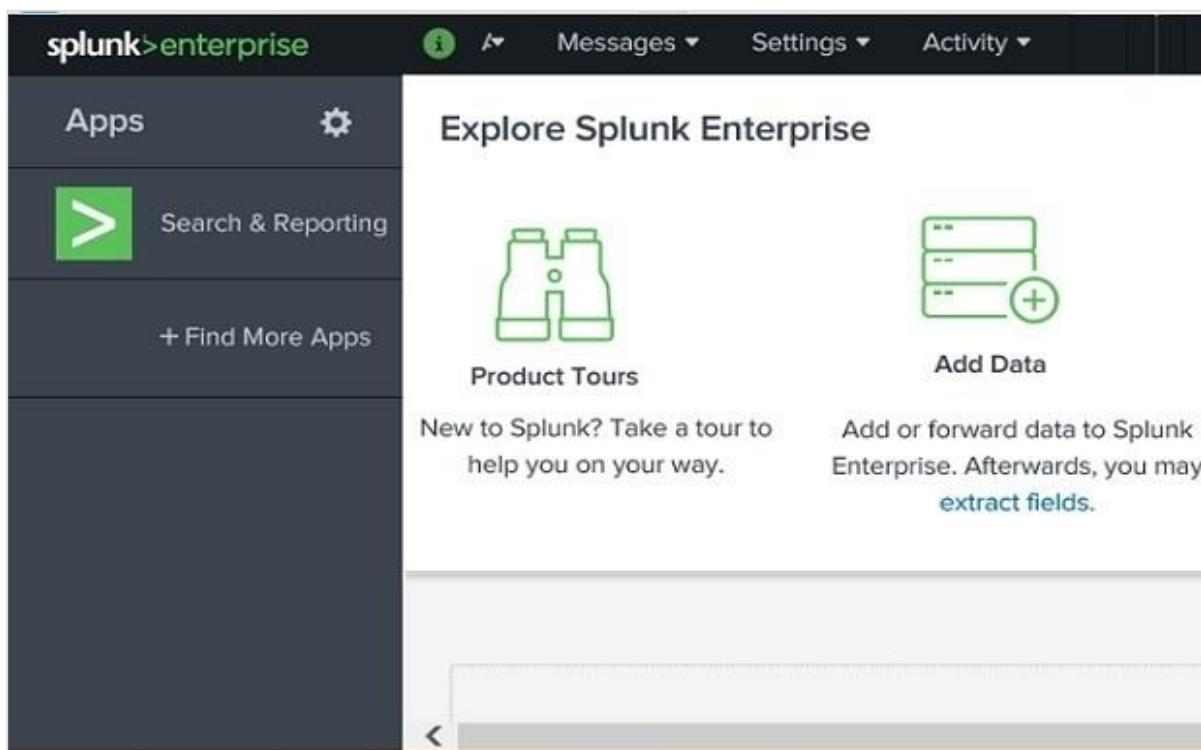
Step 3

In the final step, we see that Splunk is successfully installed and it can be launched from the web browser.



Step 4

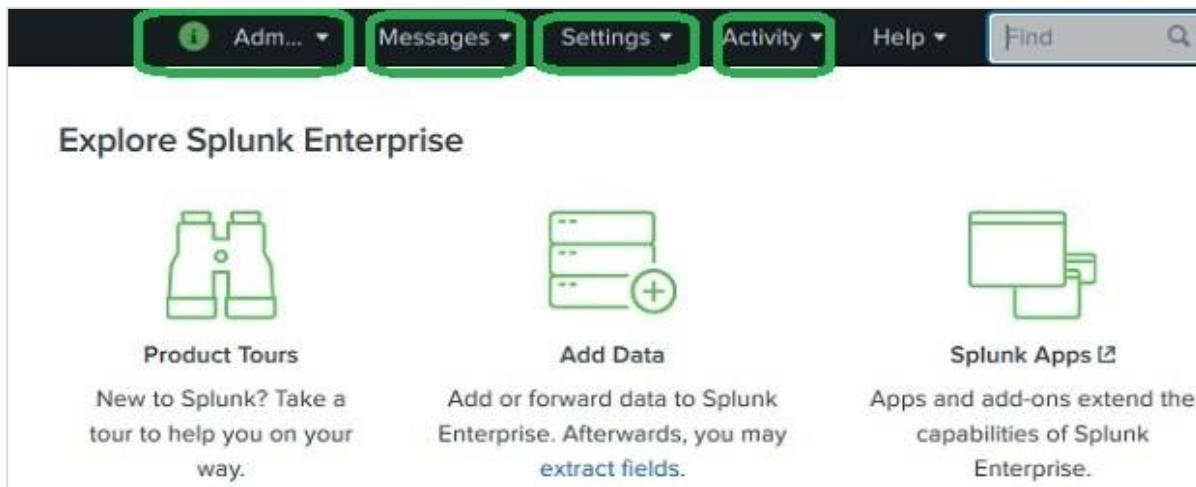
Next, open the browser and enter the given url, <http://localhost:8000>, and login to the Splunk using the admin user ID and password.



3. Splunk – Interface

The Splunk web interface consists of all the tools you need to search, report and analyse the data that is ingested. The same web interface provides features for administering the users and their roles. It also provides links for data ingestion and the in-built apps available in Splunk.

The below picture shows the initial screen after your login to Splunk with the admin credentials.



Administrator Link

The Administrator drop down gives the option to set and edit the details of the administrator. We can reset the admin email ID and password using the below screen:

A screenshot of the Splunk Admin interface, specifically the 'Personal' settings screen. The screen has a 'Personal' heading at the top. It contains five input fields: 'Full name' (value: 'Administrator'), 'Email address' (value: 'changeme@example.com'), 'Old password' (placeholder: 'Old password'), 'Set password' (placeholder: 'New password'), and 'Confirm password' (placeholder: 'Confirm new password'). Below these fields is a note: 'Password must contain at least 8 characters'. At the bottom right is a green 'Save' button.

Further from the administrator link, we can also navigate to the preferences option where we can set the time zone and home application on which the landing page will open after your login. Currently, it opened on the Home page as shown below:

The screenshot shows the 'Preferences' page with two tabs: 'Global' and 'SPL Editor'. The 'Global' tab is selected. A descriptive text block says: 'Use these properties to set your timezone, default application, and default search time range picker. You can also specify if background jobs should restart when Splunk software restarts.' Below this, there are three settings: 'Time zone' (dropdown menu set to 'Default System Timezone'), 'Default application' (dropdown menu set to 'Home'), and 'Restart background jobs' (a toggle switch that is off). At the bottom right are 'Cancel' and 'Apply' buttons.

Settings Link

This is a link which shows all the core features available in Splunk. For example, you can add the lookup files and lookup definitions by choosing the lookup link.

We will discuss the important settings of these links in the subsequent chapters.

The screenshot shows the Splunk web interface. At the top, there is a navigation bar with links for Adminis..., Messages, Settings, Activity, Help, Find, and a search icon. On the left, there is a sidebar with a tree view. The tree view has three main branches: 'Explorers' (selected), 'Monitoring', and 'System'. Under 'Explorers', there are 'Add Data' and 'Monitoring Console'. Under 'Monitoring', there is 'Logs' (selected). Under 'System', there are 'Server settings', 'Server controls', 'Health report manager', 'Instrumentation', 'Licensing', and 'Workload management'. To the right of the sidebar, there is a large grid of links categorized into four groups: KNOWLEDGE, DATA, DISTRIBUTED ENVIRONMENT, and USERS AND AUTHENTICATION. The 'KNOWLEDGE' group includes Searches, reports, and alerts, Data models, Event types, Tags, Fields, Lookups, User interface, Alert actions, Advanced search, and All configurations. The 'DATA' group includes Data inputs, Forwarding and receiving, Indexes, Report acceleration summaries, and Source types. The 'DISTRIBUTED ENVIRONMENT' group includes Indexer clustering, Forwarder management, and Distributed search. The 'USERS AND AUTHENTICATION' group includes Access controls.

KNOWLEDGE	DATA
Searches, reports, and alerts	Data inputs
Data models	Forwarding and receiving
Event types	Indexes
Tags	Report acceleration summaries
Fields	Source types
Lookups	DISTRIBUTED ENVIRONMENT
User interface	Indexer clustering
Alert actions	Forwarder management
Advanced search	Distributed search
All configurations	USERS AND AUTHENTICATION
Server settings	Access controls
Server controls	
Health report manager	
Instrumentation	
Licensing	
Workload management	

Search and Reporting Link

The search and reporting link takes us to the features where we can find the data sets that are available for searching the reports and alerts created for these searches. It is clearly shown in the below screenshot:

splunk>enterprise    Messages  Settings 

Search Datasets Reports Alerts Dashboards

Datasets

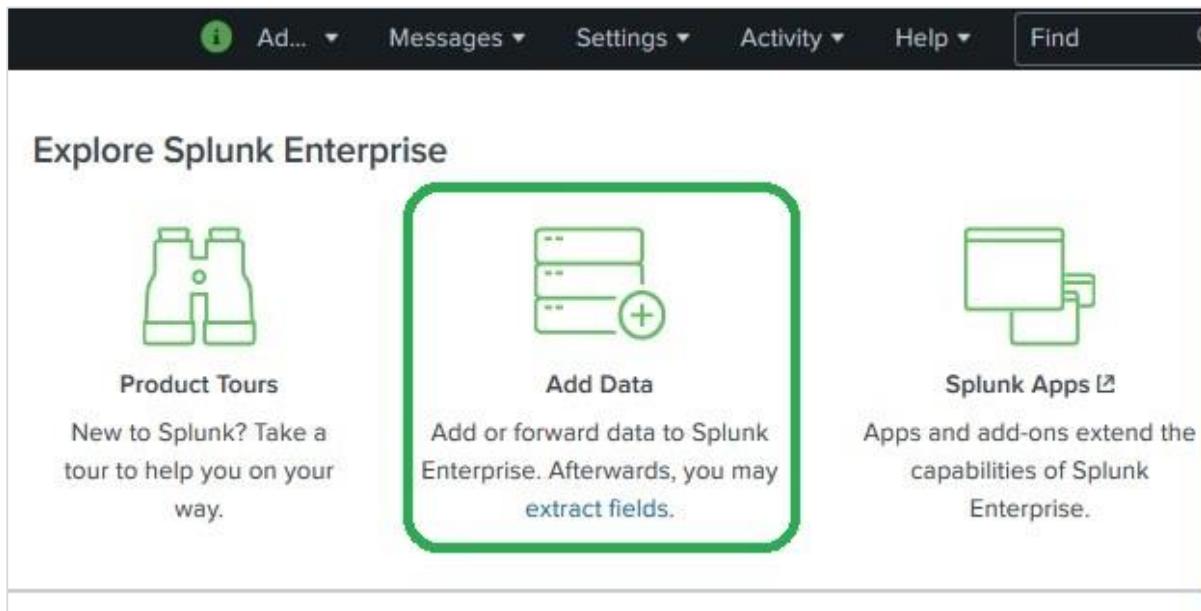
Use the Datasets listing page to view and manage your existing datasets. Click a dataset name to view its contents. Click Pivot to design a visualization-rich report based on the dataset. Click Explore in Search to search the dataset in Search and save it as a new report, alert, or dashboard panel.

Learn more about Datasets.  Don't have the Splunk Datasets Add-on? Download it here. 

28 Datasets		All	Yours	This App's	Filter by title, description, file type	
i	Title	Dataset Type	↓	Actions	↓	Owner
>	Splunk's In...	data model	↓	Manage  Explore 	↓	nobody
>	Splunk's In...	data model	↓	Manage  Explore 	↓	nobody
>	Splunk's In...	data model	↓	Manage  Explore 	↓	nobody
>	Splunk's In...	data model	↓	Manage  Explore 	↓	nobody

4. Splunk – Data Ingestion

Data ingestion in Splunk happens through the **Add Data** feature which is part of the search and reporting app. After logging in, the Splunk interface home screen shows the **Add Data** icon as shown below.



On clicking this button, we are presented with the screen to select the source and format of the data we plan to push to Splunk for analysis.

Gathering The Data

We can get the data for analysis from the Official Website of Splunk. Save this file and unzip it in your local drive. On opening the folder, you can find three files which have different formats. They are the log data generated by some web apps. We can also gather another set of data provided by Splunk which is available at from the Official Splunk webpage.

We will use data from both these sets for understanding the working of various features of Splunk.

Uploading data

Next, we choose the file, **secure.log** from the folder, **mailsv** which we have kept in our local system as mentioned in the previous paragraph. After selecting the file, we move to next step using the green coloured next button in the top right corner.

The screenshot shows the 'Add Data' wizard with five steps: 'Select Source', 'Set Source Type', 'Input Settings', 'Review', and 'Done'. The 'Select Source' step is active, indicated by a green dot. The interface includes a file selection area with a 'Select File' button, a large empty box for dropping files, and a note about the maximum upload size (500 Mb). A 'Done' button is at the bottom.

Add Data

Select Source Set Source Type Input Settings Review Done

Select Source

Choose a file to upload to the Splunk platform, either by browsing your computer or by dropping a file

Selected File: **secure.log**

Select File

Drop your data file here

The maximum file upload size is 500 Mb

Done

Selecting Source Type

Splunk has an in-built feature to detect the type of the data being ingested. It also gives the user an option to choose a different data type than the chosen by Splunk. On clicking the source type drop down, we can see various data types that Splunk can ingest and enable for searching.

In the current example given below, we choose the default source type.

ID	Time	Value
1	10/8/18 12:15:05.000 AM	1
2	10/8/18 12:15:05.000 AM	1
3	10/8/18 12:15:05.000 AM	1
4	10/8/18 12:15:05.000 AM	1
5	10/8/18 12:15:05.000 AM	1

Input Settings

In this step of data ingestion, we configure the host name from which the data is being ingested. Following are the options to choose from, for the host name:

Constant value

It is the complete host name where the source data resides.

regex on path

When you want to extract the host name with a regular expression, enter the regex for in the regular expression field.

segment in path

When you want to extract the host name from a segment in your data source's path, enter the segment number in the Segment number field. For example, if the path to the source is **/var/log/** and you want the third segment (the host server name) to be the host value, enter "3".

Next, we choose the index type to be created on the input data for searching. We choose the default index strategy. The summary index only creates summary of the data through aggregation and creates index on it while the history index is for storing the search history. It is clearly depicted in the image below:

The screenshot shows the Splunk Data Input configuration interface. At the top, there is a navigation bar with icons for 'Messages', 'Settings', 'Activity', 'Help', and a 'Find' button. Below the navigation bar, a progress bar indicates the current step: 'Set Source Type' (green dot), 'Input Settings' (green dot), 'Review' (grey dot), and 'Done' (grey dot). To the right of the progress bar are 'Back' and 'Review' buttons, with 'Review' being green.

The main configuration area has several sections:

- Host field value:** A dropdown menu with three options: 'Constant value' (selected and highlighted with a green box), 'Regular expression on path', and 'Segment in path'.
- Host field value input:** A text input field containing the value 'mailsecure_log'.
- Index selection:** A section titled 'Index' with a 'Default' dropdown menu. The 'Default' option is checked. Other options listed are 'history', 'main', and 'summary'. This entire section is also highlighted with a green box.
- Text on the left:** A sidebar with descriptive text about setting up a data source, mentioning that each event receives a host value and providing links for more information.
- Text at the bottom:** A question 'What are multiple indexes?'.

Review Settings

After clicking on the next button, we see a summary of the settings we have chosen. We review it and choose Next to finish the uploading of data.

The screenshot shows the Splunk Enterprise interface with the title 'splunk>enterprise'. In the top right corner, there are icons for help, messages, and settings. Below the title, the 'Add Data' process is shown with five steps: 'Select Source', 'Set Source Type', 'Input Settings', 'Review', and 'Done'. The 'Review' step is currently selected, indicated by a green circle. The main content area displays the configuration details:

Setting	Value
Input Type	Uploaded File
File Name	secure.log
Source Type	securelogsource
Host	mailsecure_log
Index	Default

On finishing the load, the below screen appears which shows the successful data ingestion and further possible actions we can take on the data.

Add Data

Select Source Set Source Type Input Settings Review Done

File has been uploaded successfully.

Configure your inputs by going to [Settings > Data Inputs](#)

Start Searching Search your data now or see examples and tutorials. [Learn more](#)

Extract Fields Create search-time field extractions. [Learn more about fields](#).

Add More Data Add more data inputs now or see examples and tutorials. [Learn more](#)

Download Apps Apps help you do more with your data. [Learn more](#).

Build Dashboards Visualize your searches. [Learn more](#).

5. Splunk – Source Types

All the incoming data to Splunk are first judged by its inbuilt data processing unit and classified to certain data types and categories. For example, if it is a log from apache web server, Splunk is able to recognize that and create appropriate fields out of the data read.

This feature in Splunk is called source type detection and it uses its built-in source types that are known as "pretrained" source types to achieve this.

This makes things easier for analysis as the user does not have to manually classify the data and assign any data types to the fields of the incoming data.

Supported Source Types

The supported source types in Splunk can be seen by uploading a file through the **Add Data** feature and then selecting the dropdown for Source Type. In the below image, we have uploaded a CSV file and then checked for all the available options.

_index	_score	_type	_id	_index	_score	_type	_time
1	▲	11/23/1					12:05:1
2	▲	11/23/1					12:05:1
3	▲	11/23/1					12:05:1
4	▲	11/23/1					12:05:1
5	▲	11/23/1					12:05:1
6	▲	11/23/1					12:05:1
7	▲	11/23/1					12:05:1
8	▲	11/23/1					12:05:1

Source Type Sub-Category

Even in those categories, we can further click to see all the sub categories that are supported. So when you choose the database category, you can find the different types of databases and their supported files which Splunk can recognize.

Add Data

Select Source Set Source Type Input Settings

Set Source Type

This page lets you see how the Splunk platform sees your data before indexing. If the even "Next" to proceed. If not, use the options below to define proper event breaks and timestamp your data, create a new one by clicking "Save As".

Source: productidvals.csv

Source type: csv ▾

filter

- > Default Settings
Splunk's default source type settings
- > Application
- > Custom
- > Database
- > Email
- > Metrics
- > Miscellaneous
- > Network & Security
- > Operating System
- > Structured
- > Uncategorized
- > Web

	4	5	6	7	8
mysqld_error	11/23/12:05:	11/23/12:05:	11/23/12:05:	11/23/12:05:	11/23/12:05:

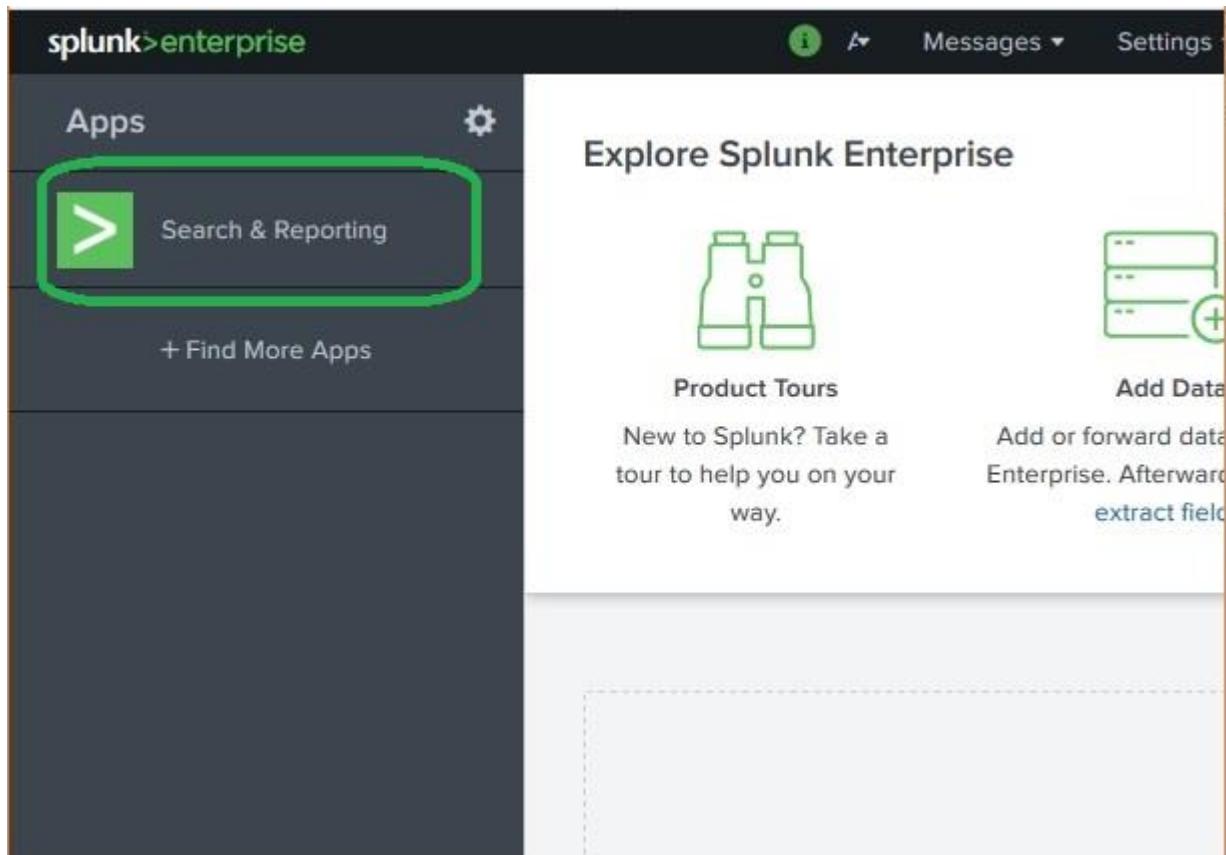
Pre-Trained Source Types

The below table lists some of the important pre-trained source types Splunk recognizes:

Source Type Name	Nature
access_combined	NCSA combined format http web server logs (can be generated by apache or other web servers)
access_combined_wcookie	NCSA combined format http web server logs (can be generated by apache or other web servers), with cookie field added at end
apache_error	Standard Apache web server error log
linux_messages_syslog	Standard linux syslog (/var/log/messages on most platforms)
log4j	Log4j standard output produced by any J2EE server using log4j
mysqld_error	Standard mysql error log

6. Splunk – Basic Search

Splunk has a robust search functionality which enables you to search the entire data set that is ingested. This feature is accessed through the app named as **Search & Reporting** which can be seen in the left side bar after logging in to the web interface.



On clicking on the **search & Reporting** app, we are presented with a search box, where we can start our search on the log data that we uploaded in the previous chapter.

We type the host name in the format as shown below and click on the search icon present in the right most corner. This gives us the result highlighting the search term.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk>enterprise' logo, user profile icon, 'Messages' and 'Settings' dropdowns, and tabs for 'Search', 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. Below the navigation is a search bar containing the query 'host="mailsecure_log"'. A green oval highlights this search term. Underneath the search bar, it says '9,829 events (before 10/20/18 9:17:05.000 AM) No Event Sampling'. There are buttons for 'Job', 'Smart Mo...', and other search controls. Below these are tabs for 'Events (9,829)', 'Patterns', 'Statistics', and 'Visualization', with 'Events' being selected. A timeline bar shows the search range. At the bottom, there are filters for 'Show Fields', 'List', 'Format', and '20 Per Page', along with a page number indicator '1 2 ...'. The main table lists three log entries from October 15, 2018, at 00:15:06, all from 'mailsecure_log' host, showing failed password attempts for SSH connections.

i	Time	Event
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[5276]: Failed password for im... 3351 ssh2 host = mailsecure_log source = secure.log sourcetype = securelogso...
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[1039]: Failed password for ro... host = mailsecure_log source = secure.log sourcetype = securelogso...
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[5258]: Failed password for im... 626 ssh2 host = mailsecure_log source = secure.log sourcetype = securelogso...

Combining Search Terms

We can combine the terms used for searching by writing them one after another but putting the user search strings under double quotes.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk>enterprise' logo, a gear icon, 'Messages' dropdown, and 'Settings' dropdown. Below the bar, there are tabs for 'Search' (which is selected), 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. The main area is titled 'New Search'.

In the search bar, the query is: `source="secure.log" host="mailsecure_log" sourcetype="securelogsource"`. Below the search bar, it says '9,829 events (before 10/20/18 9:30:24.000 AM) No Event Sampling'.

Below the search bar are various controls: 'Job' dropdown, zoom controls ('Zoom Out', '+ Zoom to Selection'), and a 'Smart Mode' button.

The search results are displayed in a table with columns: 'Time' and 'Event'. There are three rows of results:

i	Time	Event
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[5276]: Failed password for im 3351 ssh2 host = mailsecure_log source = secure.log sourcetype = securelogso
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[1039]: Failed password for ro host = mailsecure_log source = secure.log sourcetype = securelogso
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[5258]: Failed password for im 626 ssh2 host = mailsecure_log source = secure.log sourcetype = securelogso

Below the table, there are buttons for 'Show Fields', 'List', 'Format', '20 Per Page', and page navigation (1, 2, ...).

Using Wild Card

We can use wild cards in our search option combined with the **AND/OR** operators. In the below search, we get the result where the log file has the terms containing fail, failed, failure, etc., along with the term password in the same line.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk>enterprise' logo, a gear icon, 'Messages' dropdown, and 'Settings' dropdown. Below the bar, there are tabs for 'Search' (which is selected), 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. The main area is titled 'New Search' and contains a search bar with the query 'fail* AND password'. Below the search bar, it says '✓ 66,272 events (before 10/20/18 9:36:32.000 AM) No Event Sampling'. There are various controls like 'Job', 'Smart Mode', and zoom tools ('Format Timeline', 'Zoom Out', 'Zoom to Selection', 'Deselect'). The results section is titled 'Events (66,272)' and includes tabs for 'Events', 'Patterns', 'Statistics', and 'Visualization'. The results table has columns for 'Time' and 'Event'. The first few results are:

i	Time	Event
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[5276]: Failed password for im 3351 ssh2 host = solunkhost source = secure.log sourcetype = mailsecurelogdata
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[5276]: Failed password for im 3351 ssh2 host = mailsecure_log source = secure.log sourcetype = securelogsource
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[1039]: Failed password for ro host = mailsecure_log source = secure.log sourcetype = securelogsource

Refining Search Results

We can further refine the search result by selecting a string and adding it to the search. In the below example, we click over the string **3351** and select the option **Add to Search**.

After **3351** is added to the search term, we get the below result which shows only those lines from the log containing 3351 in them. Also mark how the time line of the search result has changed as we have refined the search.

splunk>enterprise

Search Datasets Reports Alerts Dashboards

New Search

fail* AND password 3351

✓ 21 events (before 10/20/18 9:53:14.000 AM) No Event Sampling ▾

Job ▾ II ⌂ ⌂ ⌂ Smart Mo

Events (21) Patterns Statistics Visualization

Format Timeline ▾ — Zoom Out + Zoom to Selection X Deselect

i	Time	Event
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[5276]: Failed password for user 3351 ssh2 host = solunkhost source = secure.log sourcetype = raw
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[5276]: Failed password for user 3351 ssh2 host = mailsecure_log source = secure.log sourcetype = raw
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[3351]: Failed password for user 4856 ssh2 host = solunkhost source = secure.log sourcetype = raw
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[3351]: Failed password for user 4856 ssh2 host = mailsecure_log source = secure.log sourcetype = raw

7. Splunk – Field Searching

When Splunk reads the uploaded machine data, it interprets the data and divides it into many fields which represent a single logical fact about the entire data record.

For example, a single record of information may contain server name, timestamp of the event, type of the event being logged whether login attempt or a http response, etc. Even in case of unstructured data, Splunk tries to divide the fields into key value pairs or separate them based on the data types they have, numeric and string, etc.

Continuing with the data uploaded in the previous chapter, we can see the fields from the **secure.log** file by clicking on the show fields link which will open up the following screen. We can notice the fields Splunk has generated from this log file.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a search bar with the query "fail* AND password". Below the search bar, it says "66,272 events (before 10/21/18 6:48:13.000 AM)" and "No Event Sampling". The interface includes tabs for "Events (66,272)", "Patterns", "Statistics", and "Visualization". Below these are buttons for "Format Timeline", "Zoom Out", "Zoom to Selection", and "Deselect". The main area displays a timeline at the bottom and a table of event results above. The table has columns for "Time" and "Event". The first few rows of the table are:

i	Time	Event
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 ms id user appserver from 194. host = solunkhost source =
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 ms id user appserver from 194. host = mailsecure_log sour sourcetype = securelogsource
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 ms from 194.8.74.23 port 3768 host = mailsecure_log sour

On the left side, there's a sidebar with sections for "SELECTED FIELDS" containing "a host 4", "a source 3", and "a sourcetype 4"; and "INTERESTING FIELDS" containing "# date_hour 24", "# date_mday 30", "# date_minute 60", "a date_month 2", and "# date_second 60". Two green arrows point from the "SELECTED FIELDS" and "INTERESTING FIELDS" sections towards the "Time" column in the event table.

Choosing the Fields

We can choose what fields to be displayed by selecting or unselecting the fields from the list of all fields. Clicking on **all fields** opens a window showing the list of all the fields. Some of these fields have check marks against them showing they are already selected. We can use the check boxes to choose our fields for display.

Besides the name of the field, it displays the number of distinct values the fields have, its data type and what percentage of events this field is present in.

Select Fields					
		Select All Within Filter	Deselect All	Coverage: 1% or more	X
i	✓ ▾	Field	# of Values	Event Coverage	Type
>	<input checked="" type="checkbox"/>	host	4	100%	String
>	<input checked="" type="checkbox"/>	source	3	100%	String
>	<input checked="" type="checkbox"/>	sourcetype	4	100%	String
>	<input type="checkbox"/>	date_hour	24	100%	Number
>	<input type="checkbox"/>	date_mday	30	100%	Number
>	<input type="checkbox"/>	date_minute	60	100%	Number
>	<input type="checkbox"/>	date_month	2	100%	String
>	<input type="checkbox"/>	date_second	60	100%	Number
>	<input type="checkbox"/>	date_weekday	7	100%	String
>	<input type="checkbox"/>	date_year	1	100%	Number
>	<input type="checkbox"/>	date_zone	1	100%	String
>	<input type="checkbox"/>	index	1	100%	String
>	<input type="checkbox"/>	linecount	1	100%	Number
>	<input type="checkbox"/>	pid	>100	75.23%	Number

Field Summary

Very detailed stats for every selected field become available by clicking on the name of the field. It shows all the distinct values for the field, their count and their percentages.

The screenshot shows the Splunk search interface with the following details:

- Top Bar:** Includes icons for info, gear, Messages, Settings, Activity, Help, and Find, followed by a green search bar labeled "Search & R".
- Header:** Shows "Alerts" and "Dashboards" buttons, a "Save As" dropdown, and a time range selector set to "All time".
- Search Bar:** Displays the query "(8:13:00 AM) No Event Sampling" and includes Job, Smart Mode, and visualization mode buttons.
- Time Range:** Set to "1 day".
- Panel Headers:** Statistics and Visualization.
- Panel Content:**
 - sourcetype:** A summary card showing "4 Values, 100% of events" with "Selected" buttons for Yes and No, and an "X" button to close the card.
 - Reports:** Options for "Top values", "Top values by time", and "Rare values".
 - Events with this field:** A table showing the top values and their counts and percentages:

Values	Count	%
linux_secure	49,858	75.232%
mailsecurelogdata	8,154	12.304%
securelogsource	8,154	12.304%
access_combined_wcookie	106	0.16%
 - Event Preview:** A detailed view of an event with timestamp (10/15/18 12:15:06.000 AM), source (Thu Oct 15 2018 00:15:06 mailsv1 sshd[1039]: Failed password from 194.8.74.23 port 3768 ssh2), host (host=colunkhost), source (source=secure log), and sourcetype (sourcetype=mailsecure).

Using Fields in Search

The field names can also be inserted into the search box along with the specific values for the search. In the below example, we aim to find all the records for the date, 15th Oct for the host named **mailsecure_log**. We get the result for this specific date.

The screenshot shows the Splunk Enterprise interface with the following details:

- Search Bar:** The search bar contains the query: `fail* AND password host="mailsecure_log" date_mday`.
- Event Count:** 8,154 events (before 10/21/18 7:22:58).
- Filter:** A dropdown menu is open under the search bar, listing values for `date_mday`: "1", "10", "11", "12", "13", "14", "15", and "16". A green arrow points to the value "15".
- Event View:** The main area displays a table of event results. The columns are labeled: `i`, Time, and Event.
- Event Data:**

	Time	Event
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv id user appserver from 194.8.74 host = mailsecure_log source = sourcetype = securelogsource
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv from 194.8.74.23 port 3768 ssh2 host = mailsecure_log source = sourcetype = securelogsource
>	10/15/18	Thu Oct 15 2018 00:15:06 mailsv
- Left Sidebar:** Shows selected fields (`host 1`, `source 1`, `sourcetype 1`) and interesting fields (`# date_hour 1`, `# date_mday 8`, `# date_minute 1`, `a date_month 1`).

8. Splunk – Time Range Search

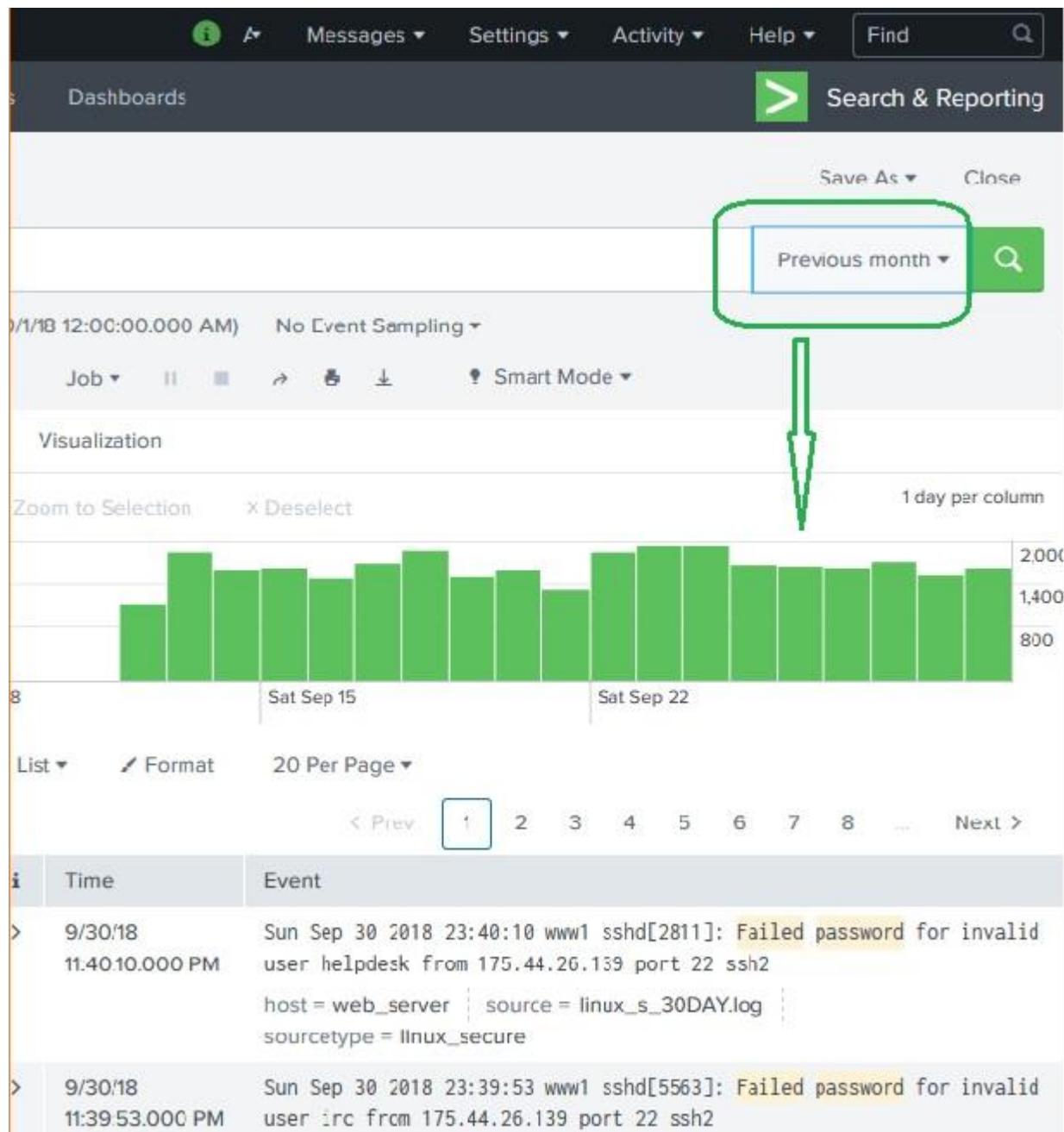
The Splunk web interface displays timeline which indicates the distribution of events over a range of time. There are preset time intervals from which you can select a specific time range, or you can customize the time range as per your need.

The below screen shows various preset timeline options. Choosing any of these options will fetch the data for only that specific time period which you can also analyse further, using the custom timeline options available.

The screenshot shows the Splunk web interface with a dark header bar containing 'Messages', 'Settings', 'Activity', 'Help', 'Find', and a magnifying glass icon. Below the header, there are tabs for 'Dashboards' and 'Search & Reporting'. A green box highlights the 'All time' dropdown in the top right corner of the main search area. A green arrow points to the 'Previous month' option in the expanded timeline dropdown menu. The menu includes sections for 'REAL-TIME', 'RELATIVE', and 'OTHER', with various time intervals listed. A sidebar on the left lists navigation options like 'Relative', 'Real-time', 'Date Range', 'Date & Time Range', and 'Advanced'. At the bottom, a log entry is displayed: '10/15/18 12:15:05.000 AM Thu Oct 15 2018 00:15:06 mailsv1 sshd[5276]: Failed password for invalid user appserver from 194.8.74.23 port 3351 ssh2'.

REAL-TIME	RELATIVE	OTHER
30 second window	Today	Last 15 minutes
1 minute window	Week to date	Last 60 minutes
5 minute window	Business week to date	Last 4 hours
30 minute window	Month to date	Last 24 hours
1 hour window	Year to date	Last 7 days
All time (real-time)	Yesterday	Last 30 days
	Previous week	
	Previous business week	
	Previous month	
	Previous year	

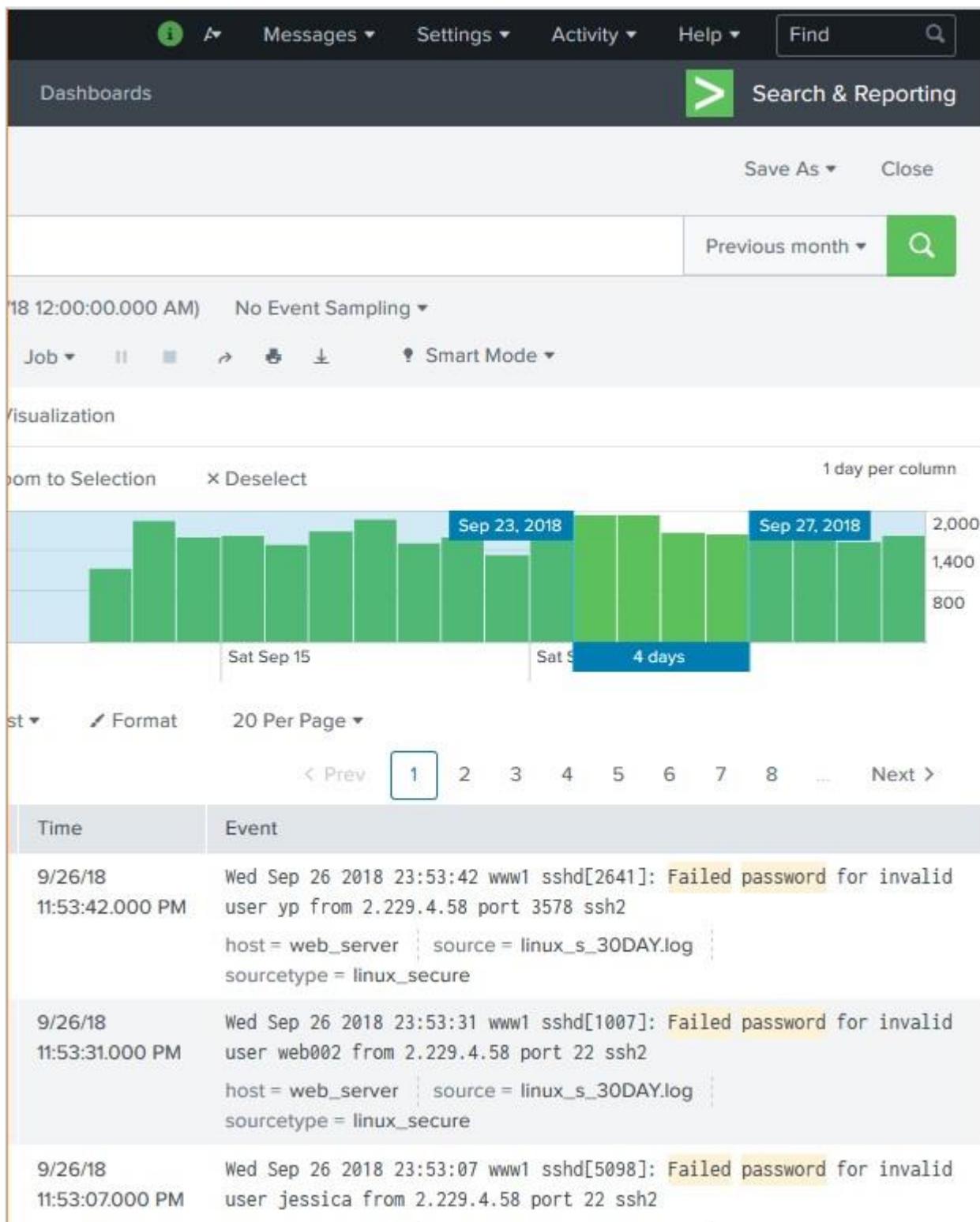
For example, choosing the previous month option gives us the result only for the previous month as you can see the in spread of the timeline graph below.



Selecting a Time Subset

By clicking and dragging across the bars in the timeline, we can select a subset of the result that already exists. This does not cause the re-execution of the query. It only filters out the records from the existing result set.

Below image shows the selection of a subset from the result set:



Earliest and Latest

The two commands, earliest and latest can be used in the search bar to indicate the time range in between which you filter out the results. It is similar to selecting the time subset, but it is through commands rather than the option of clicking at a specific time line bar. So, it provides a finer control over that data range you can pick for your analysis.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk>enterprise' and icons for 'Messages' and 'Settings'. Below it is a secondary navigation bar with 'Search', 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. The main area is titled 'New Search'.

In the search bar, the query is: `host=mailsecure_log earliest=-15d latest=-7d`. A green oval highlights this query. Below the search bar, it says '8,858 events (before 10/14/18 10:34:18.000 AM) No Event Sampling'.

Below the search bar are buttons for 'Job', 'Smart Mo', and other search controls.

The main content area has tabs for 'Events (8,858)', 'Patterns', 'Statistics', and 'Visualization'. The 'Events' tab is selected. It includes buttons for 'Format Timeline', 'Zoom Out', 'Zoom to Selection', and 'Deselect'.

A timeline visualization shows event counts over time. Red arrows point upwards from the dates 'Sun Oct 7 2018', 'Tue Oct 9', and 'Thu Oct 11' towards the event count line. The y-axis ranges from 600 to 1,800 events.

Below the timeline are buttons for 'List', 'Format', and '20 Per Page'. A page number '1' is highlighted in a box.

The bottom section shows a table of event details:

	Time	Event
>	10/14/18 12:15:06.000 AM	Wed Oct 14 2018 00:15:06 p from 193.33.170.23 port host = mailsecure_log sc sourcetype = securelogsource
>	10/14/18 12:15:06.000 AM	Wed Oct 14 2018 00:15:06 id user jabber from 193.3 host = mailsecure_log sc sourcetype = securelogsource
>	10/14/18 12:15:06.000 AM	Wed Oct 14 2018 00:15:06 id user rightscale from 1

On the left side, there are sections for 'SELECTED FIELDS' (host 1, source 1, sourcetype 1) and 'INTERESTING FIELDS' (#date_hour 1, #date_mday 7, #date_minute 1, #date_month 1, #date_second 2).

In the above image, we give a time range between last 7 days to last 15 days. So, the data in between these two days is displayed.

Nearby Events

We can also find nearby events of a specific time by mentioning how close we want the events to be filtered out. We have the option of choosing the scale of the interval, like – seconds, minutes, days and week etc.

9. Splunk – Sharing Exporting

When you run a search query, the result is stored as a job in the Splunk server. While this job was created by one specific user, it can be shared across with other users so that they can start using this result set without the necessity of building the query for it again. The results can also be exported and saved as files which can be shared with users who do not use Splunk.

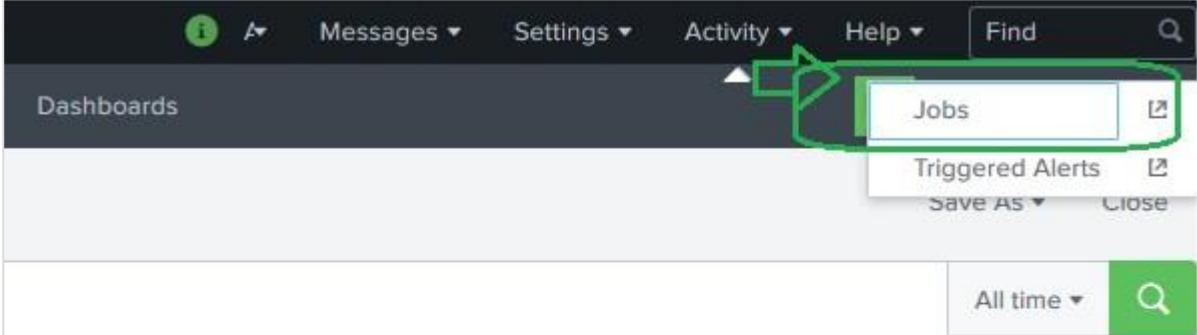
Sharing the Search Result

Once a query has run successfully, we can see a small upward arrow in the middle right of the web page. Clicking on this icon gives a URL where the query and the result can be accessed. There is a need to grant permission to the users who will be using this link. Permission is granted through the Splunk administration interface.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk>enterprise' and icons for user status, messages, and settings. Below the bar, there are tabs for 'Search', 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. The 'Search' tab is selected. The main area is titled 'New Search' and contains a search bar with the query: 'host=mailsecure_log earliest=-15d latest=-7d'. Below the query, it says '✓ 8,858 events (before 10/14/18 10:34:18.000 AM) No Event S...' followed by a 'Share' button. A green arrow points from this 'Share' button down to a 'Share Job' modal window. The modal window has a title 'Share Job' and a message: 'The job's lifetime has been extended to 7 days and read permissions have been set to Everyone. Manage the job via Job Settings.' It contains a 'Link To Job' field with the URL 'http://localhost:8000/en-US/app/search/search?sid=154' and a note: 'Copy or bookmark the link by right-clicking the icon, or drag the icon into your bookmarks bar.' On the left side of the search results, there's a sidebar with 'Event' and 'For' filters, and a chart showing event counts from 600 to 1,800. At the bottom, there's a table showing selected fields: 'host 1', 'source 1', and 'sourcetype 1'. The table also shows some log entries with fields like 'host', 'source', 'sourcetype', and timestamp.

Finding the Saved Results

The jobs that are saved to be used by all users with appropriate permissions can be located by looking for the jobs link under the activity menu in the top right bar of the Splunk interface. In the below image, we click on the highlighted link named jobs to find the saved jobs.



The screenshot shows the Splunk web interface. At the top, there is a navigation bar with links for 'Messages', 'Settings', 'Activity', 'Help', and a search bar labeled 'Find'. A green arrow points from the text 'highlighted link named jobs' to the 'Jobs' link in the 'Activity' dropdown menu, which is also highlighted with a green box. Below the navigation bar, there is a 'Dashboards' section. Further down, there is a visualization area showing event sampling and a timeline from Oct 10 to Oct 14. At the bottom, there is a table listing events with columns for 'Time' and 'Event'. The first event listed is:

```
10/15/18 12:15:06 AM Thu Oct 15 2018 00:15:06 mailsv1 sshd[5276]: Failed password for invalid user appserver from 194.8.74.23 port 3351 ssh2
host = mailsecure_log | source = secure.log
sourcetype = securelogsource
```

The second event listed is:

```
10/15/18 12:15:06 AM Thu Oct 15 2018 00:15:06 mailsv1 sshd[1039]: Failed password for root from 194.8.74.23 port 3768 ssh2
```

After the above link is clicked, we get the list of all the saved jobs as shown below. Here, we have to note that there is an expiry date post where the saved job will automatically get removed from Splunk. You can adjust this date by selecting the job and clicking on Edit selected and then choosing Extend Expiration.

The screenshot shows the Splunk Enterprise interface with the title "splunk>enterprise". The top navigation bar includes "Messages" and "Settings". The main section is titled "Jobs" with the sub-instruction "Manage your jobs. Learn More". Below this, there are filters for "2 Jobs", "App: Search & Reporting (search)", "Filter by owner", and "Status: All". A "filter" button is also present. An "Edit Selected" button is highlighted with a red box. A table lists two search jobs:

	<input type="checkbox"/>	Owner	Application	Events	Size	Created at	Expires
>	<input type="checkbox"/>	admin	search	9,829	364 KB	Oct 23, 2018 10:30:59 AM	Oct 30, 2018 10:53:58 AM
>	<input type="checkbox"/>	admin	search	8,858	300 KB	Oct 21, 2018 10:34:18 AM	Oct 28, 2018 10:45:39 PM

Below the table, search queries are displayed:

- host=mailsecure_log [before 10/23/18 10:30:59.000 AM]
- host=mailsecure_log earliest=-15d latest=-7d [before 10/14/18 10:34:18.000 AM]

Exporting the Search Result

We can also export the results of a search into a file. The three different formats available for export are: CSV, XML and JSON. Clicking on the Export button after choosing the formats downloads the file from the local browser into the local system. This is explained in the below image:

splunk>enterprise 

Search Datasets Reports Alerts Dashboards   Messages  Settings

New Search

host=mailsecure_log

✓ 9,829 events (before 10/23/18 10:30:59.000 AM) No Event Sampling **Export**

Job       Smart Mo

Events (9,829) Patterns Statistics Visualization

Format Timeline  - Zoom Out + Zoom to Selection   

3,000
2,000
1,000

Export Results

Format **CSV**  2

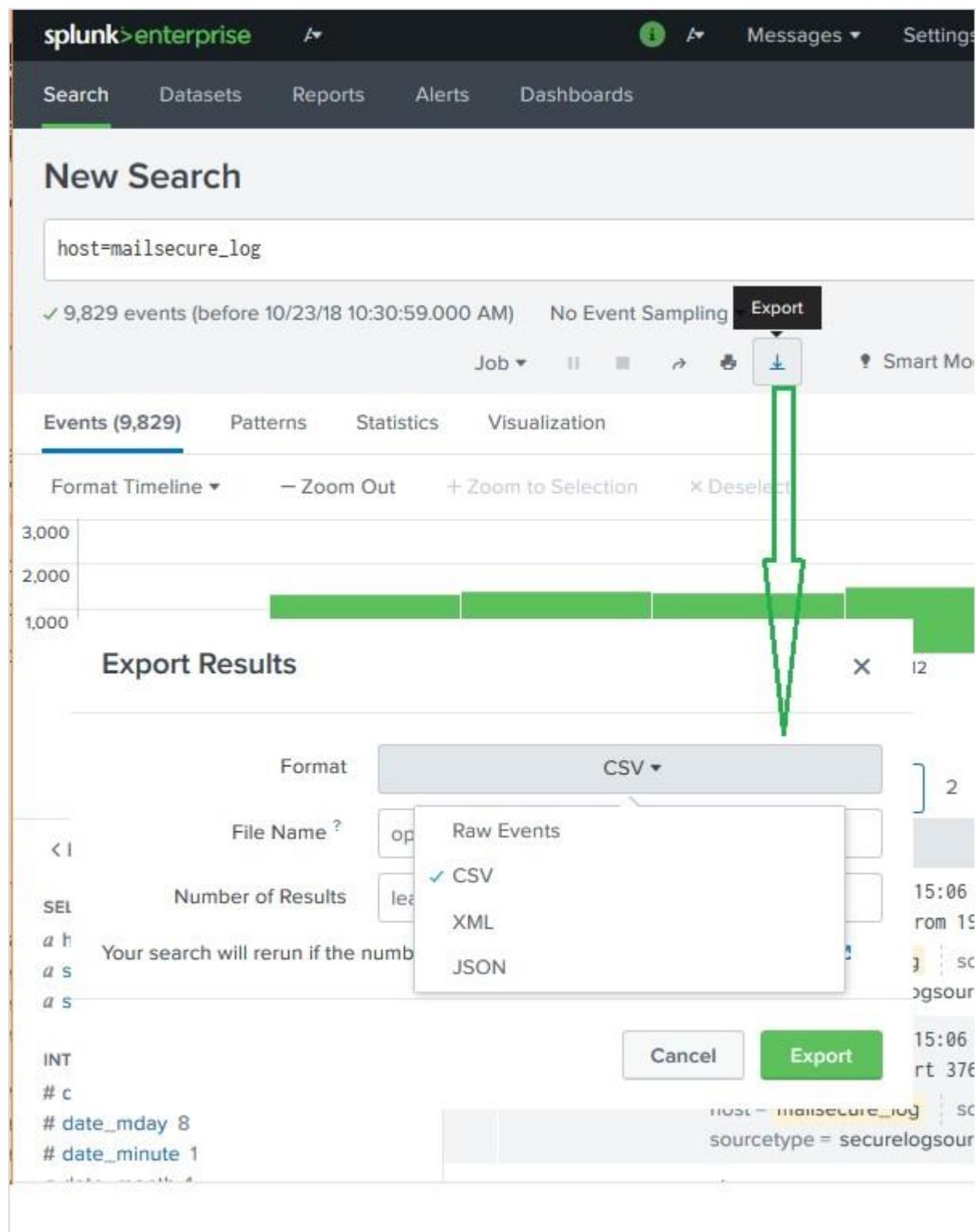
File Name ?  op  15:06
SEL lead rom 15
a h Your search will rerun if the numbr sc
as  15:06
as  rt 376

Number of Results  15:06

Raw Events
 CSV
XML
JSON

Cancel **Export**

host=mailsecure_log sc
sourcetype=securelogsource



10. Splunk – Search Language

The Splunk Search Processing Language (SPL) is a language containing many commands, functions, arguments, etc., which are written to get the desired results from the datasets. For example, when you get a result set for a search term, you may further want to filter some more specific terms from the result set. For this, you need some additional commands to be added to the existing command. This is achieved by learning the usage of SPL.

Components of SPL

SPL has the following components:

- **Search Terms** – These are the keywords or phrases you are looking for.
- **Commands** – The action you want to take on the result set like format the result or count them.
- **Functions** – What are the computations you are going to apply on the results. Like Sum, Average etc.
- **Clauses** – How to group or rename the fields in the result set.

Let us discuss all the components with the help of images in the below section:

Search Terms

These are the terms you mention in the search bar to get specific records from the dataset which meet the search criteria. In the below example, we are searching for records which contain two highlighted terms.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with links for Search, Datasets, Reports, Alerts, and Dashboards. Below that is a search bar containing the query `host="mailsecure_log" nsharpe`, with the results count of 376 events displayed. A timeline visualization shows event counts for each day from October 8 to October 12, 2018. Below the timeline is a table listing the top 20 events. The first three events are listed as follows:

i	Time	Event
>	10/13/18 12:15:06.000 AM	Tue Oct 13 2018 00:15:06 mailsv1 sudo: nsharpe ; TTY=pts/0 ; host = mailsecure_log source = secure.log sourcetype = securel
>	10/13/18 12:15:06.000 AM	Tue Oct 13 2018 00:15:06 mailsv1 sshd[10249]: failed password host = mailsecure_log source = secure.log sourcetype = securel
>	10/13/18 12:15:06.000 AM	Tue Oct 13 2018 00:15:06 mailsv1 sshd[65318]: pam_unix(sshd:se:0)

Commands

You can use many in-built commands that SPL provides to simplify the process of analysing the data in the result set. In the below example we use the head command to filter out only the top 3 results from a search operation.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk>enterprise' and various dropdown menus like 'Messages' and 'Settings'. Below the bar, a menu bar includes 'Search', 'Datasets', 'Reports', 'Alerts', and 'Dashboards', with 'Search' being the active tab.

New Search

In the search bar, the query 'host="mailsecure_log" | head 3' is entered. Below the search bar, it says '3 events (before 10/24/18 9:23:59.000 AM) No Event Sampling'.

Below the search bar are several controls: 'Job', 'Smart Mode', and buttons for 'Events (3)', 'Patterns', 'Statistics', and 'Visualization'. There are also buttons for 'Format Timeline', 'Zoom Out', 'Zoom to Selection', and 'Deselect'.

A green timeline bar highlights the events at 12:15:06.000 AM on Monday, October 15, 2018. The timeline shows four points labeled 1, 2, 3, and 4 from left to right, corresponding to the event times: 12:15:06.000 AM, 12:15:06.000 AM, 12:15:06.000 AM, and 12:15:06.000 AM.

Below the timeline, there are buttons for 'Show Fields', 'List', 'Format', and '20 Per Page'. The main area displays a table of events:

i	Time	Event
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[5276]: Failed password f 3351 ssh2 host = mailsecure_log source = secure.log sourcetype = securel
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[1039]: Failed password f host = mailsecure_log source = secure.log sourcetype = securel
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[5258]: Failed password f 626 ssh2 host = mailsecure_log source = secure.log sourcetype = securel

Functions

Along with commands, Splunk also provides many in-built functions which can take input from a field being analysed and give the output after applying the calculations on that field. In the below example, we use the **Stats avg()** function which calculates the average value of the numeric field being taken as input.

host="web_application" | stats avg(bytes)

✓ 131,645 events (before 10/25/18 6:25:14.000 AM) No Event Sampling ▾

Job ▾ || ■ ↻ ⌂ ⌂ Smart Mo

Events Patterns Statistics (1) Visualization

20 Per Page ▾ Format Preview ▾

avg(bytes) ▾

2178.2865965285428

Clauses

When we want to get results grouped by some specific field or we want to rename a field in the output, we use the **group by** clause and the **as** clause respectively. In the below example, we get the average size of bytes of each file present in the **web_application** log. As you can see, the result shows the name of each file as well as the average bytes for each file.

splunk>enterprise ▾

Search Datasets Reports Alerts Dashboards

New Search

host="web_application" | stats avg(bytes) by file

All time ▾ 

✓ 131,645 events (before 10/25/18 6:32:00.000 AM)

Job ▾ II ■

Events Patterns Statistics (30) Visualization

20 Per Page ▾ ✓ Format Preview ▾ < Prev 1 2 Next >

file	avg(bytes)
ADMIN	3406
Admin	3406
account	2119
adm	3406
admin	3406
administration	3406

11. Splunk – Search Optimization

Splunk already includes the optimization features, analyses and processes your searches for maximum efficiency. This efficiency is mainly achieved through the following two optimization goals:

- **Early Filtering:** These optimizations filter the results very early so that the amount of data getting processed is reduced as early as possible during the search process. This early filter avoids unnecessary lookup and evaluation calculations for events that are not part of final search results.
- **Parallel Processing:** The built-in optimizations can reorder search processing, so that as many commands as possible are run in parallel on the indexers before sending the search results to the search head for final processing.

Analysing Search Optimisations

Splunk has given us tools to analyse how the search optimization works. These tools help us figure out how the filter conditions are used and what is the sequence of these optimisation steps. It also gives us the cost of the various steps involved in the search operations.

Example

Consider a search operation to find the events which contain the words: fail, failed or password. When we put this search query in the search box, the built-in optimizers act automatically to decide the path of the search. We can verify how long the search took to return a specific number of search results and if needed can go on to check each and every step of the optimization along with the cost associated with it.

We follow the path of **Search -> Job -> Inspect Job** to get these details as shown below:

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk>enterprise' and links for Search, Datasets, Reports, Alerts, and Dashboards. Below the search bar, it says 'New Search'. The search query is 'fail* AND password' with a time range of 'All time'. It shows 66,272 events found before 11/21/18 4:51:00.000 PM, with 'No Event Sampling'. A histogram at the bottom displays event counts by day from September 15 to 29, 2018. A context menu is open over the histogram, with the 'Inspect Job' option highlighted. The table below lists two log entries related to failed password attempts.

i	Time	Event
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[5276]: Failed password for user 3351 ssh2 host = mailsecure_log
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsv1 sshd[5276]: Failed password for user 3351 ssh2 host = solunkhost

The next screen gives details of the optimization that has occurred for the above query. Here, we need to note the number of events and the time taken to return the result.

Search job inspector

This search has completed and has returned **1,000** results by scanning **66,272** events in **3.747** seconds

(SID: 1542799259.474) [search.log](#)

Execution costs

Duration (seconds)	Component	Invocations
0.00	command.fields	28
2.08	command.search	28
0.09	command.search.expand_search	2
0.00	command.search.calcfields	25
0.00	command.search.expand_search.calcfield	2
0.00	command.search.expand_search.fieldaliaser	2
0.00	command.search.expand_search.kv	2
0.00	command.search.expand_search.lookup	2
0.00	command.search.expand_search.sourcetype	2

Turning Off Optimization

We can also turn off the in-built optimization and notice the difference in the time taken for the search result. The result may or may not be better than the in-built search. In case it is better, we may always choose this option of turning off the optimization for only this specific search.

In the below diagram, we use the No Optimization command presented as **noop** in the search query.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk>enterprise' on the left, a dropdown arrow, and 'Messages' with a count of 2 on the right. Below the navigation bar are links for 'Search', 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. The main area is titled 'New Search'.

In the search bar, the query is: `fail* AND password |noop search_optimization=false`. The results indicate 66,272 events found before 11/21/18 5:28:57.000 PM, with 'No Event Sampling' selected.

A context menu is open over the search results, with 'Edit Job Settings...' highlighted. Other options in the menu include 'Send Job to Background', 'Inspect Job', and 'Delete Job'. The background shows a timeline visualization with green bars representing event counts over time, specifically for September 15, 22, and 29, 2018.

Below the timeline, there are buttons for 'Show Fields', 'List', 'Format', and '20 Per Page'. The 'List' view is currently selected, displaying two event entries:

i	Time	Event
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsrv1 sshd[5276]: Failed password for user 3351 ssh2 host = mailsecure_log
>	10/15/18 12:15:06.000 AM	Thu Oct 15 2018 00:15:06 mailsrv1 sshd[5276]: Failed password for user 3351 ssh2 host = solunkhost

The next screen gives us the result of using no optimization. For this given query, the results come faster without using in-built optimizations.

Search job inspector

This search has completed and has returned **1,000** results by scanning **66,272** events in **1.344** seconds
(SID: 1542801536.478) [search.log](#)

Execution costs

Duration (seconds)	Component	Invocations
0.00	command.addinfo	30
0.00	command.fields	29
0.74	command.search	29
0.03	command.search.expand_search	2
0.00	command.search.calcfields	26
0.00	command.search.expand_search.calcfield	2
0.00	command.search.expand_search.fieldaliaser	2
0.00	command.search.expand_search.kv	2
0.00	command.search.expand_search.lookup	2

12. Splunk – Transforming Commands

These are the commands in Splunk which are used to transform the result of a search into such data structures which will be useful in representing the statistics and data visualizations.

Examples of Transforming Commands

Following are some of the examples of transforming commands:

- **Highlight** – To highlight the specific terms in a result.
- **Chart** – To create a chart out of the search result.
- **Stats** – To create statistical summaries from the search result.

Highlight

This command is used to **highlight specific terms in the search result set**. It is used by supplying the search terms as arguments to the highlight function. Multiple search terms are supplied by separating them with comma.

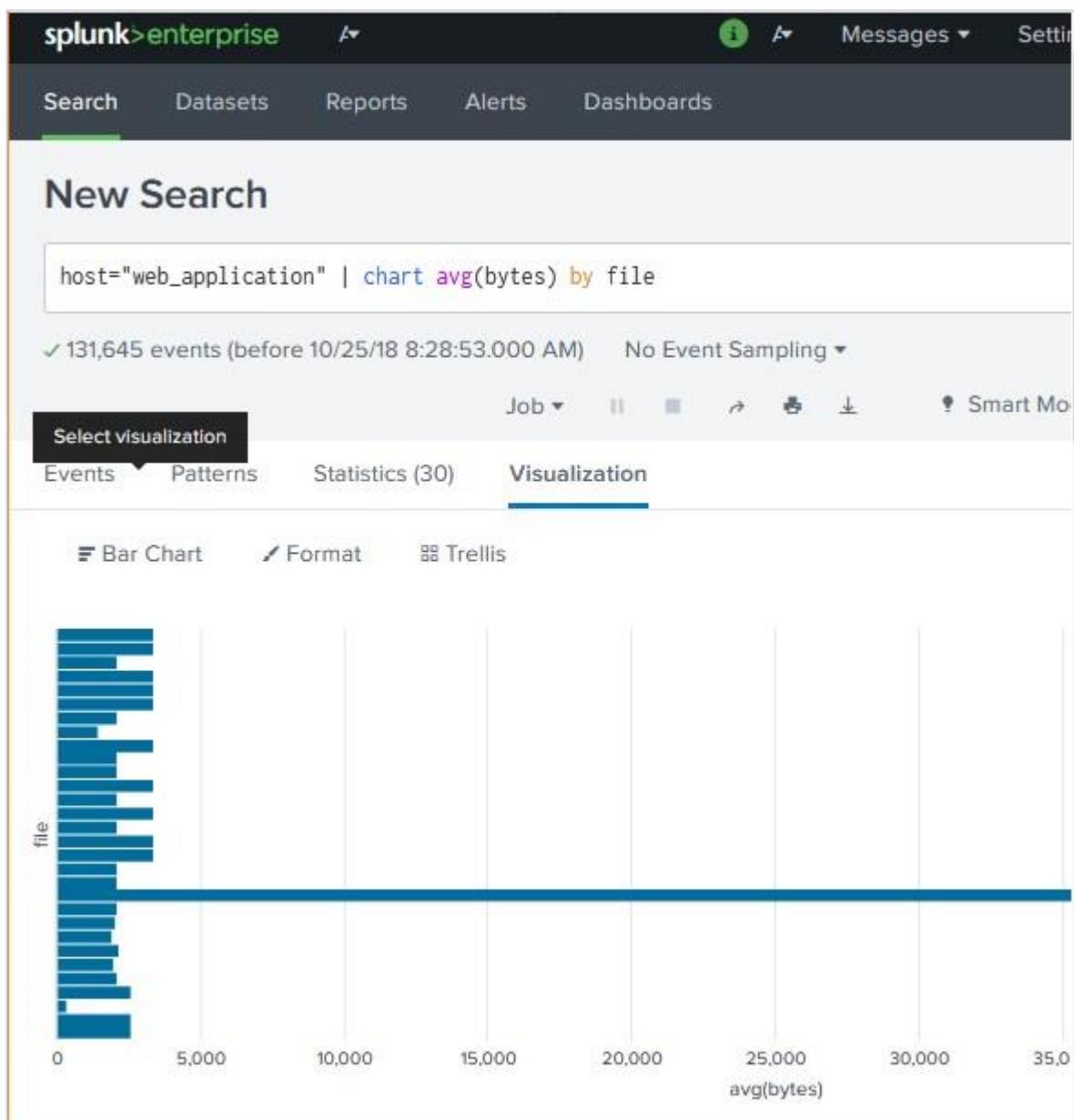
In the below example, we search for the terms, **safari** and **butter** in the result set.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with links for Search, Datasets, Reports, Alerts, and Dashboards. Below that is a search bar containing the query: host="web_application" | highlight Safari, butter. Underneath the search bar, it says "✓ 131,645 events (before 10/25/18 7:45:15.000 AM) No Event Sampling". A toolbar above the results includes "Select visualization", "Events (131,645)", "Patterns", "Statistics", and "Visualization". The "Events (131,645)" tab is selected. Below the toolbar, there are buttons for "Format Timeline", "Zoom Out", "Zoom to Selection", and "Deselect". The main area displays a horizontal bar chart representing the number of events over time. The x-axis shows dates from Sat Sep 15, 2018, to Sat Sep 29, 2018. The y-axis shows event counts from 0 to 1000. Below the chart is a table with two columns: "Time" and "Event". The first event row is expanded, showing detailed log data. The second event row is partially visible.

i	Time	Event
>	10/12/18 11:59:45.000 PM	192.188.106.240 - - [12/Oct/2018:23:59:45] "GET /category.screen?categoryId=ARCADE&productId=MB-AG-G07" 200 2958 "http://www.buttercupgames.com/category.screen?categoryId=ARCADE&productId=MB-AG-G07" "Mozilla/5.0 (iPad; CPU OS 5_1_1 like Mac OS X; rv:53.0) AppleWebKit/536.25 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.25" bytes = 2958 file = category.screen host = web_application source = access_combined_wcookie sourcetype = access_combined_wcookie
>	10/12/18 11:59:43.000 PM	212.235.92.150 - - [12/Oct/2018:23:59:43] "POST /success.do?action=purchase&G07&JSESSIONID=5D4SL6FF7ADFF4963 HTTP/1.1" 503 2198 "http://www.buttercupgames.com/categoryId=ARCADE&productId=MB-AG-G07" "Mozilla/5.0 (iPad; CPU OS 5_1_1 like Mac OS X; rv:53.0) AppleWebKit/536.25 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.25" bytes = 2198 file = success.do host = web_application productId = MB-AG-G07 sourcetype = access_combined_wcookie

Chart

The **chart** command is a transforming command that returns your results in a table format. The results can then be used to display the data as a chart, such as column, line, area, etc. In the below example, we create a horizontal bar chart by plotting the average size of bytes for each file type.



Stats

The Stats command transforms the search result data set into various statistical representations depending on the types of arguments we supply for this command.

In the below example, we use the stats command with count function which is then grouped by another field. Here, we are counting the number of file names created on each week day. The result of the search string come out in a tabular form with rows created for each day.

splunk>enterprise ▾

Search Datasets Reports Alerts Dashboards

New Search

host="web_application" | stats count(file) by date_wday All time 

✓ 131,645 events (before 10/25/18 9:00:03.000 AM)

Job ▾ || ⏪ ⏩

Select visualization

Events Patterns Statistics (7) Visualization

50 Per Page ▾ ✓ Format Preview ▾

date_wday	count(file)
friday	22775
monday	17754
saturday	16899
sunday	17217
thursday	21541
tuesday	17515
wednesday	17943

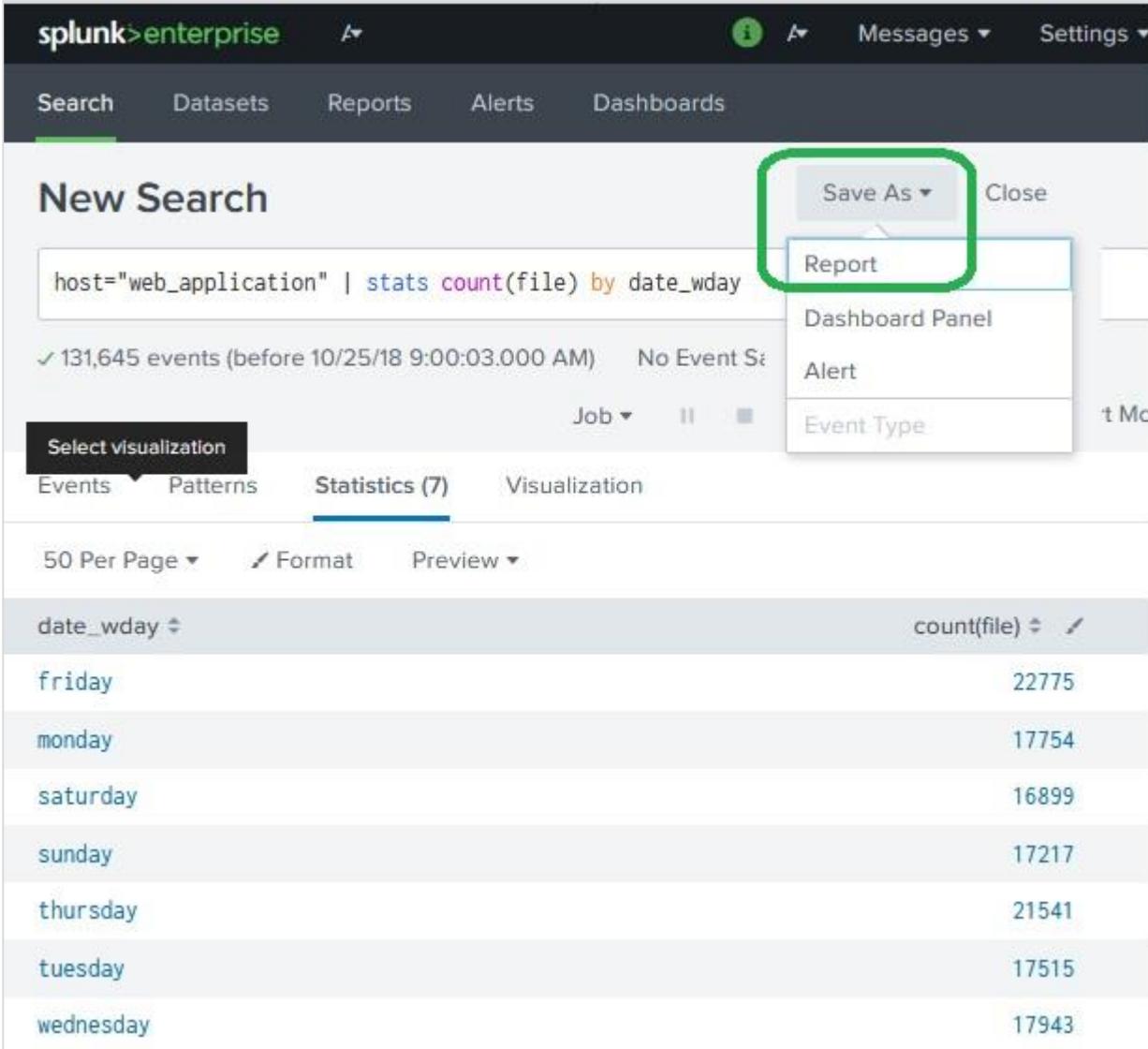
13. Splunk – Reports

Splunk reports are results saved from a search action which can show statistics and visualizations of events. Reports can be run anytime, and they fetch fresh results each time they are run. The reports can be shared with other users and can be added to dashboards. More sophisticated reports can allow a drill down function to see underlying events which create the final statistics.

In this chapter, we will see how to create and edit a sample report.

Report Creation

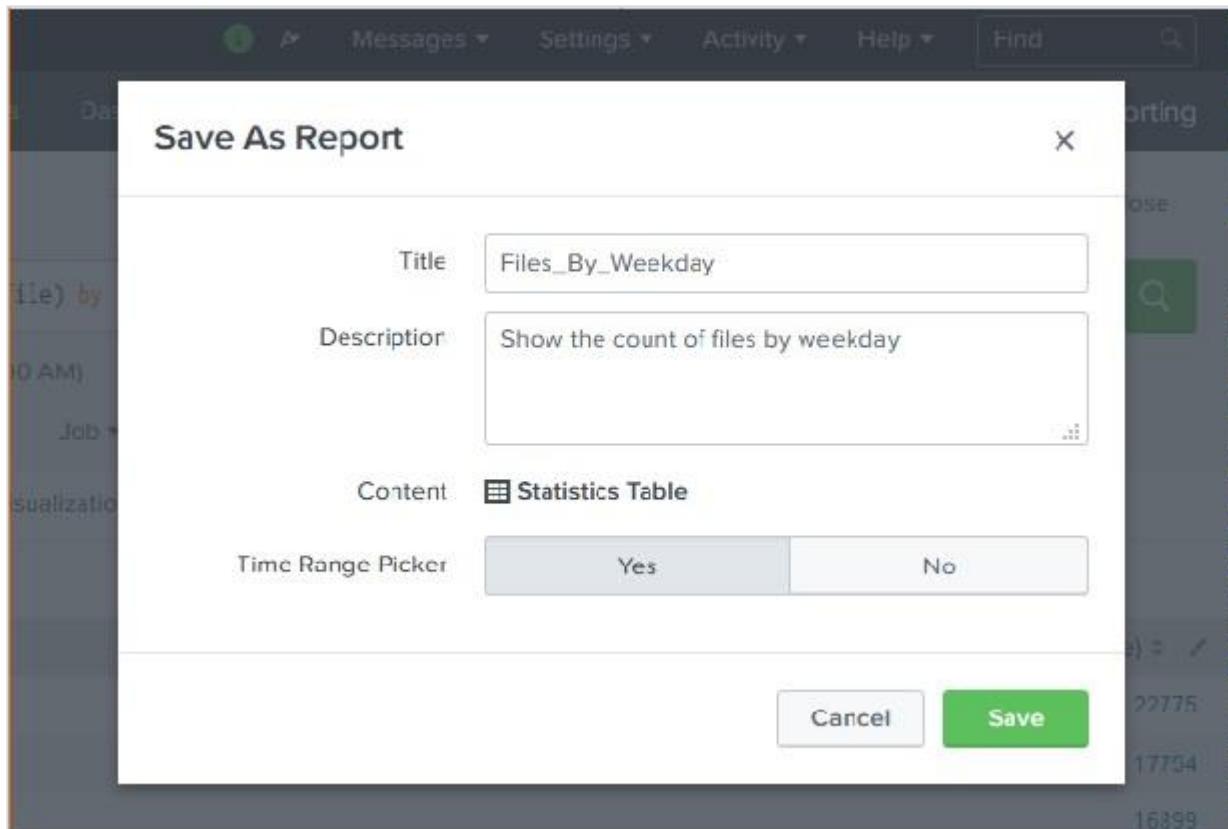
Report creation is a straight forward process where we use the **Save As** option to save the result of a search operation choosing the Reports option. The below diagram shows the **Save As** option.



The screenshot shows the Splunk Enterprise interface. At the top, there's a navigation bar with 'splunk>enterprise' on the left and 'Messages' and 'Settings' on the right. Below the navigation bar, there are tabs for 'Search', 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. The 'Search' tab is currently selected. The main area is titled 'New Search' and contains a search bar with the query 'host="web_application" | stats count(file) by date_wday'. Below the search bar, it says '✓ 131,645 events (before 10/25/18 9:00:03.000 AM)' and 'No Event S...'. There are buttons for 'Job', 'Events', 'Patterns', 'Statistics (7)', 'Visualization', '50 Per Page', 'Format', and 'Preview'. A context menu is open over the search bar, with options 'Save As', 'Report' (which is highlighted with a green box), 'Dashboard Panel', 'Alert', and 'Event Type'. The 'Report' option is the second item in the list. The 'Statistics (7)' tab is currently active, displaying a table of event counts by day of the week:

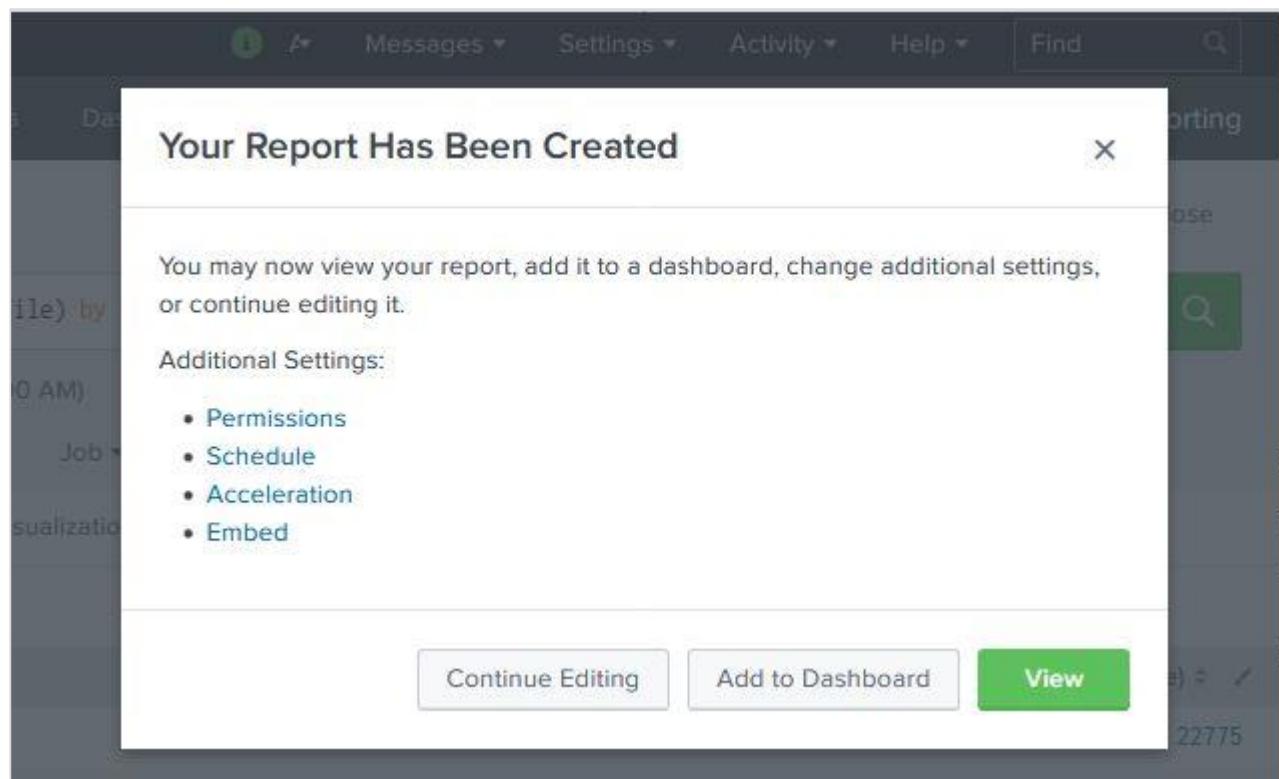
date_wday	count(file)
friday	22775
monday	17754
saturday	16899
sunday	17217
thursday	21541
tuesday	17515
wednesday	17943

By clicking on the Reports option from the dropdown, we get the next window which asks for additional inputs like the name of the report, the description and choosing the time picker. If we choose the time picker, it allows the time range to be adjusted when we run the report. Below diagrams show how we fill the required details and then click save.



Report Configuration

After clicking save to create the report in the above step, we get the next screen asking for configuring the report as shown below. Here, we can configure the permissions, scheduling the report, etc. We also get an option to go to the next step and add the report to a dashboard.



If we click on **View** in the above step, we can see the report. We also get configuration options after the report is created.

The screenshot shows the Splunk Enterprise interface with the title 'Files_By_Weekday'. The search results table has 7 results per page. A context menu is open over the first row, showing options like 'Edit Description', 'Edit Permissions', 'Edit Schedule', 'Edit Acceleration', 'Clone', 'Embed', and 'Delete'. The table data is as follows:

		count(file) ▾
friday		22775
monday		17754
saturday		16899
sunday		17217
thursday		21541
tuesday		17515
wednesday		17943

Modifying Report Search Option

While we can edit the permissions, schedule, etc., sometimes we need to modify the original search string. This can be done by choosing the **Open in Search** option as given in the above image. That will open the original search option again which we can be edited to a new search. Refer to the below image:

splunk>enterprise App: Sear... Administr... Help Find

Search Datasets Reports Alerts Dashboards > Search & Rep

Files_By_Weekday

host="web_application" | stats count(file) by date_wday

All time

✓ 131,645 events (before 10/25/18 10:08:00.000 AM) No Event Sampling ▾

Events Patterns Statistics (7) Visualization

50 Per Page ▾ Format Preview ▾

date_wday	count(file)
friday	22775
monday	17754
saturday	16899
sunday	17217
thursday	21541
tuesday	17515
wednesday	17943

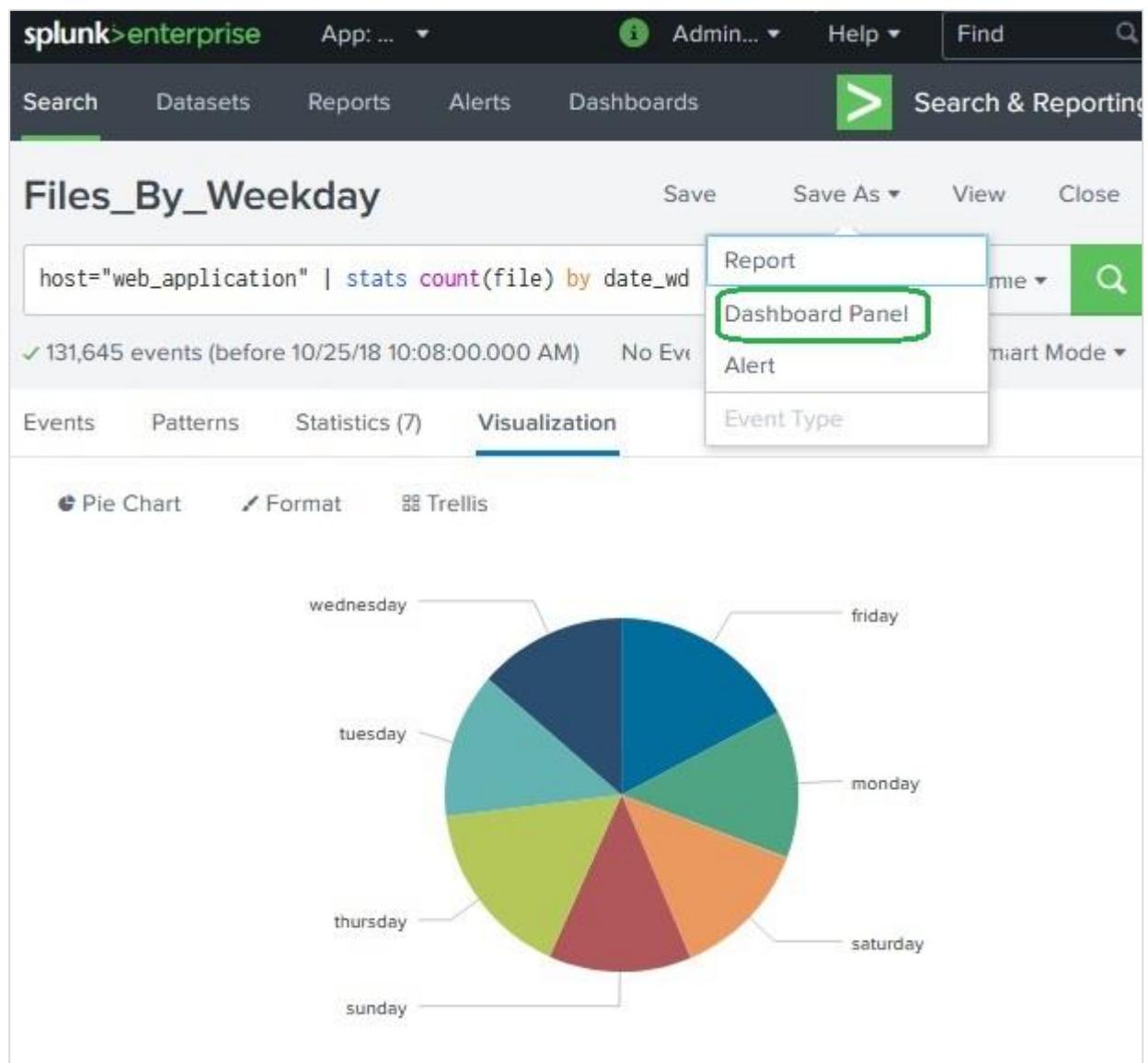
14. Splunk – Dashboards

A dashboard is used to represent tables or charts which are related to some business meaning. It is done through panels. The panels in a dashboard hold the chart or summarized data in a visually appealing manner. We can add multiple panels, and hence multiple reports and charts to the same dashboard.

Creating Dashboard

We will continue with the search query from the previous chapter which shows the count of files by week days.

We choose the Visualization tab to see the result as a pie chart. To put the chart on a dashboard, we can choose the option **Save As -> Dashboard Panel** as shown below.

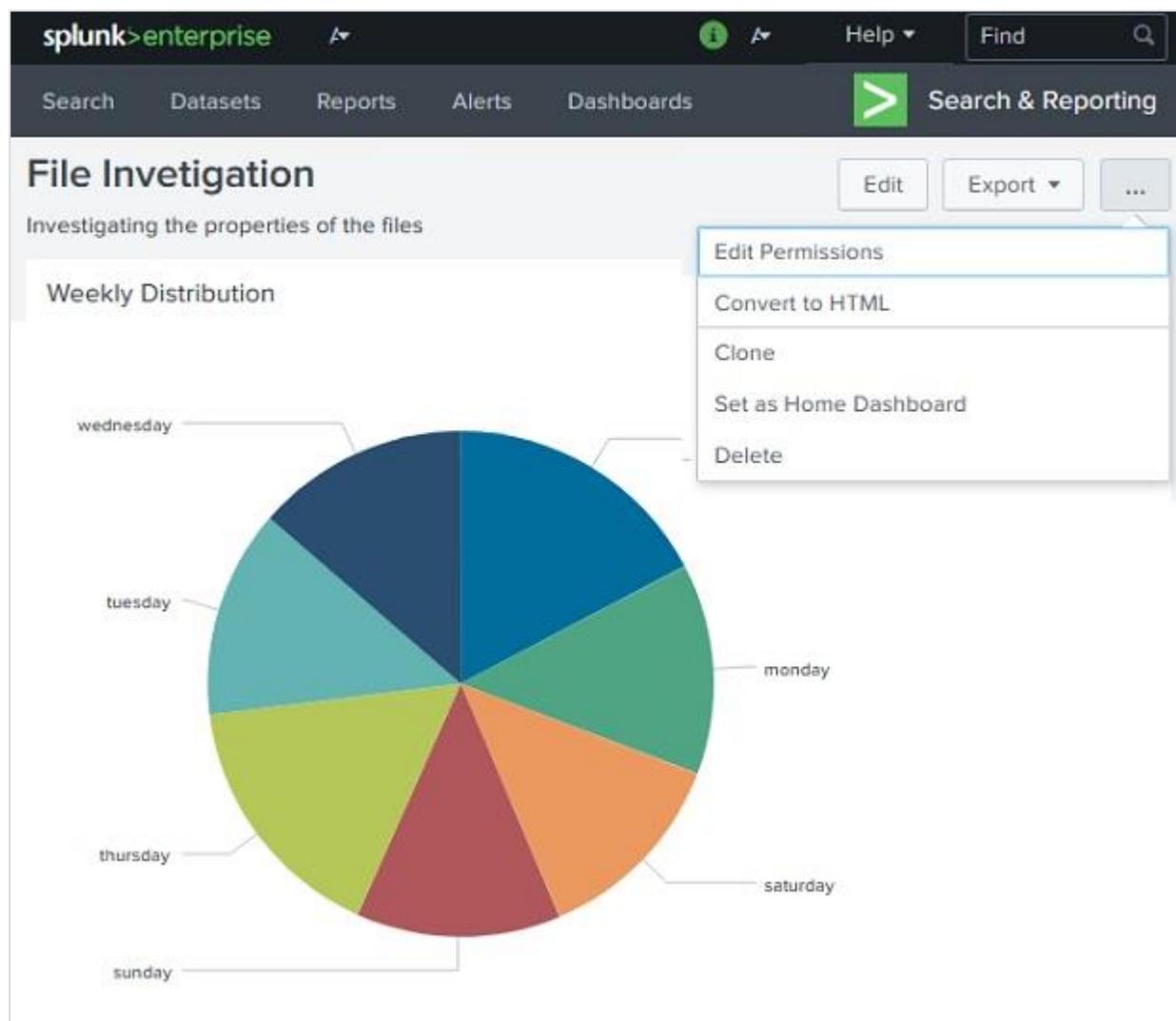


The next screen will ask for fillings the details of the dashboard and the panel in it. We fill the screen with details as shown below.

Save As Dashboard Panel

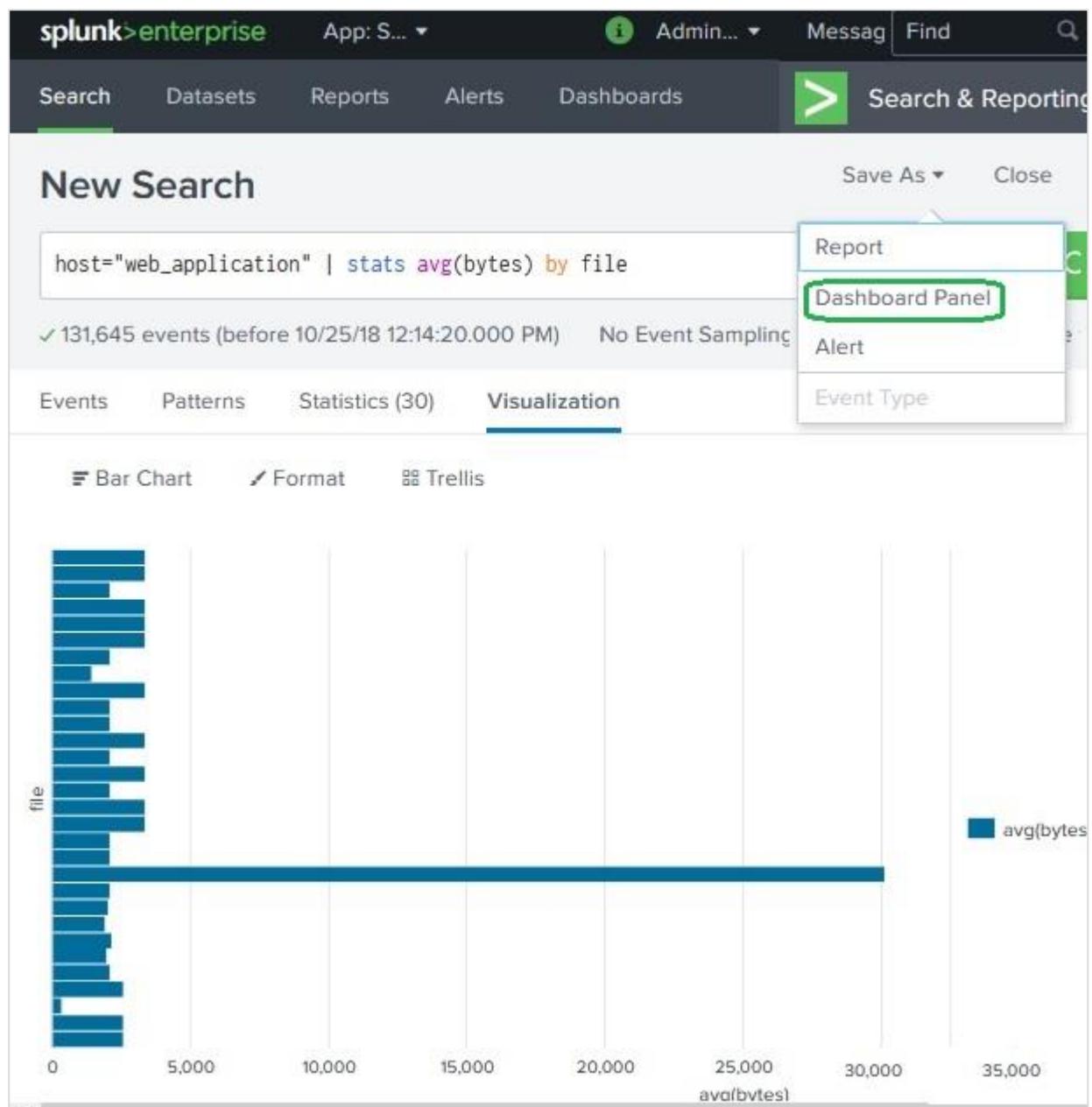
Dashboard	<input checked="" type="radio"/> New	<input type="radio"/> Existing
Dashboard Title	File Investigation	
Dashboard ID ?	file_invetigation	
	Can only contain letters, numbers and underscores.	
Dashboard Description	Investigating the properties of the files	
Dashboard Permissions	<input checked="" type="radio"/> Private	<input type="radio"/> Shared in App
Panel Title	Weekly Distribution	
Panel Powered By	<input type="radio"/> <input checked="" type="radio"/> Inline Search	<input type="radio"/> Report
Drilldown ?	No action	
Panel Content	<input type="radio"/> Statistics	<input checked="" type="radio"/> Pie Chart
	<input type="button" value="Cancel"/>	<input type="button" value="Save"/>

On clicking on Save button, the next screen gives an option to view dashboard. On choosing to view dashboard, we get the following output where we can see the dashboard and options to edit, export or delete.



Adding Panel to Dashboard

We can add a second chart to the dashboard by adding a new panel containing the chart. Below is the bar chart and its query which we are going to add to the above dashboard.



Next, we fill up the details for the second chart and click **Save** as shown in the below image:

Save As Dashboard Panel

Dashboard

File Investigation ▾

Panel Title File Average Size

Panel Powered By ?

Drilldown ? No action

Panel Content

Finally, we get the dashboard which contains both the charts in two different panels. As you can see in the image below, we can edit the dashboard to add more panels and you can add more input elements: Text, Radio and Dropdown buttons to create more sophisticated dashboards.

splunk>enterprise

Search Datasets Reports Alerts Dashboards

Edit Dashboard UI Source + Add Panel + Add Input Dark Theme

File Investigation

Investigating the properties of the files

Weekly Distribution

File Count on Different days

T Text
Radio
Dropdown
Checkbox
Multiselect
Link List
Time
Submit

A pie chart titled "File Count on Different days" showing the distribution of files across the seven days of the week. The slices represent the count of files for each day, with labels pointing to the slices: wednesday, tuesday, monday, saturday, thursday, sunday, and friday.

Day	Approximate File Count
wednesday	~15,000
tuesday	~10,000
monday	~10,000
saturday	~5,000
thursday	~5,000
sunday	~5,000
friday	~10,000

File Average Size

Average file size distribution

A histogram titled "Average file size distribution" showing the distribution of file sizes in bytes. The x-axis is labeled "avg(bytes)" and ranges from 0 to 35,000. The y-axis is labeled "file". The distribution is highly skewed, with most files being very small (under 5,000 bytes) and a few outliers reaching up to 35,000 bytes.

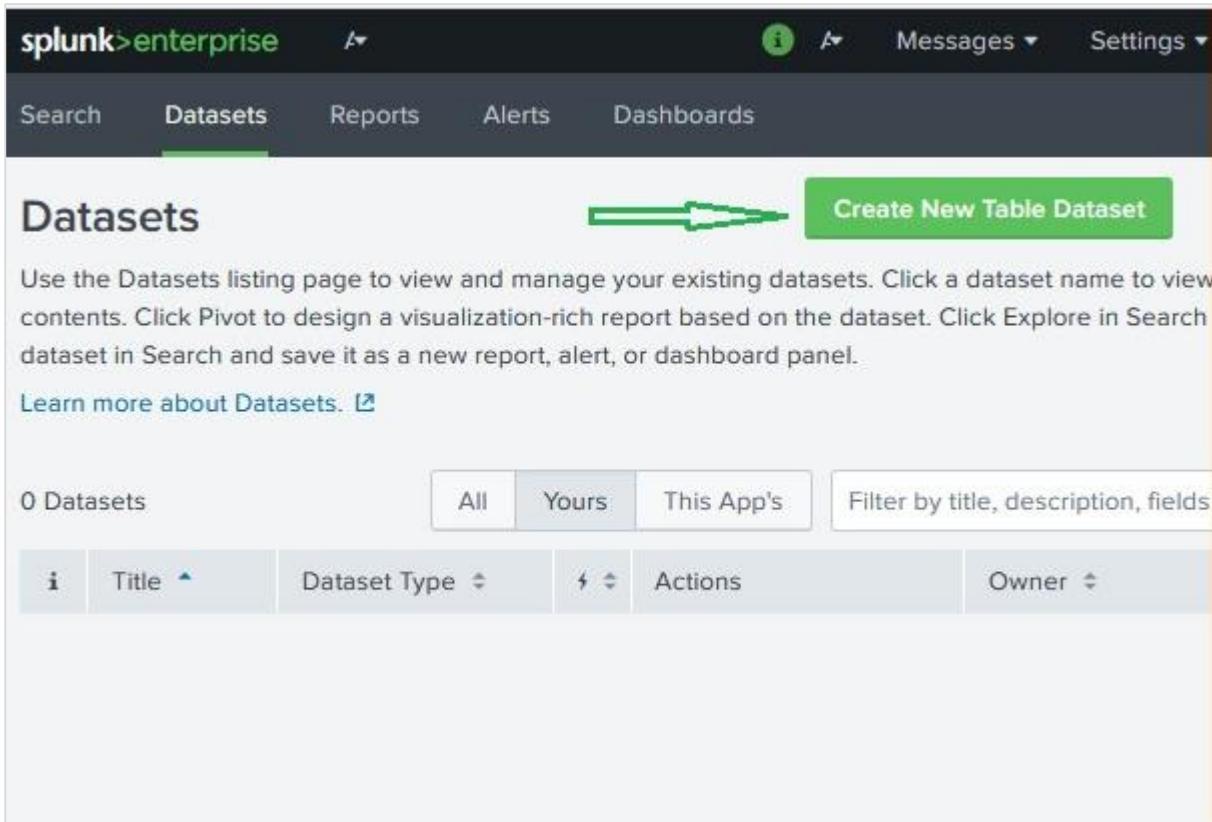
avg(bytes)	Count
0 - 5,000	~100
5,000 - 10,000	~100
10,000 - 15,000	~100
15,000 - 20,000	~100
20,000 - 25,000	~100
25,000 - 30,000	~100
30,000 - 35,000	~100
> 35,000	~10

15. Splunk – Pivot and Datasets

Splunk can ingest different types of data sources and build tables which are similar to relational tables. These are called **table dataset** or just **tables**. They provide easy ways to analyse and filter the data and lookups, etc. These table data sets are also used in creating pivot analysis which we learn in this chapter.

Creating a Dataset

We use a Splunk Add-on named **Splunk Datasets Add-on** to create and manage the datasets. It can be downloaded from the Splunk website, <https://splunkbase.splunk.com/app/3245/#/details>. It has to be installed by following the instructions given in the details tab in this link. On successful installation, we see a button named **Create New Table Dataset**.



The screenshot shows the Splunk Enterprise web interface. At the top, there's a navigation bar with 'splunk>enterprise' on the left, followed by dropdown menus for 'Messages' and 'Settings'. Below the navigation bar, there are tabs for 'Search', 'Datasets' (which is currently selected), 'Reports', 'Alerts', and 'Dashboards'. The main content area is titled 'Datasets'. On the right side of this title, there's a green button labeled 'Create New Table Dataset' with a green arrow pointing towards it. Below the title, there's a brief description: 'Use the Datasets listing page to view and manage your existing datasets. Click a dataset name to view contents. Click Pivot to design a visualization-rich report based on the dataset. Click Explore in Search dataset in Search and save it as a new report, alert, or dashboard panel.' There's also a link 'Learn more about Datasets.' with a small icon. At the bottom of the page, there's a search bar with '0 Datasets' and filters for 'All', 'Yours', 'This App's', and 'Filter by title, description, fields'. There are also columns for 'Title', 'Dataset Type', 'Actions', and 'Owner'.

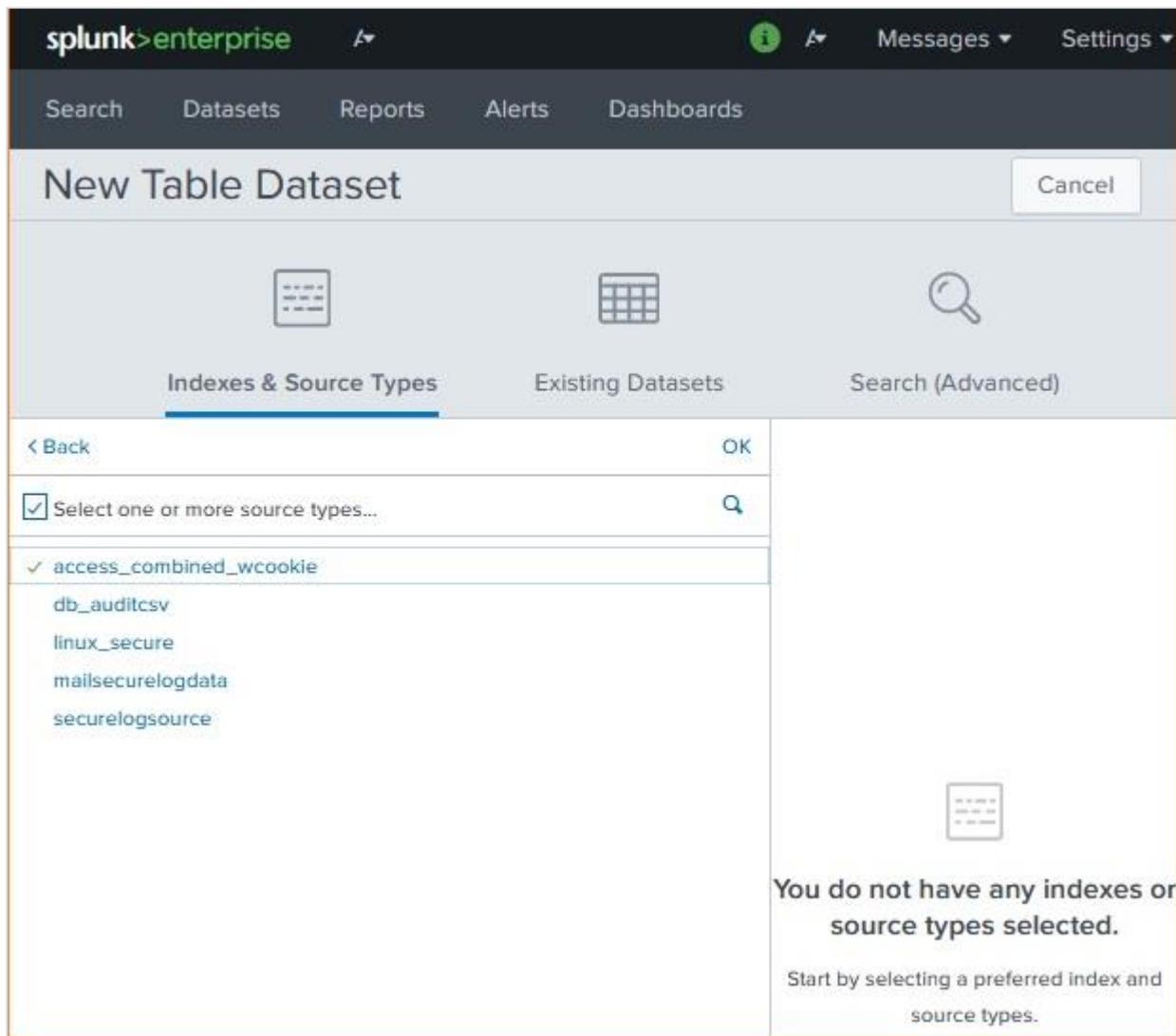
Selecting a Dataset

Next, we click on the **Create New Table Dataset** button and it gives us the option to choose from the below three options.

- **Indexes and Source Types** – Choose from an existing index or source type which are already added to Splunk through Add Data app.

- **Existing Datasets** – You might have already created some dataset previously which you want to modify by creating a new dataset from it.
- **Search** – Write a search query and the result can be used to create a new dataset.

In our example, we choose an index to be our source of data set as shown in the image below:



Choosing Dataset Fields

On clicking OK in the above screen, we are presented with an option to choose the various fields we want to finally get into the Table Dataset. The `_time` field is selected by default and this field cannot be dropped. We choose the fields: **bytes**, **categoryID**, **clientIP** and **files**.

The screenshot shows the Splunk Enterprise web interface. At the top, there's a navigation bar with links for Search, Datasets, Reports, Alerts, and Dashboards. On the far right of the top bar are icons for a user profile, Messages, and Settings. Below the navigation bar, the title "New Table Dataset" is displayed, along with a "Cancel" button. There are three main tabs: "Indexes & Source Types" (which is selected), "Existing Datasets", and "Search (Advanced)". Under "Indexes & Source Types", there are two input fields: "index = main" and "sourcetypes = access_combined_wcookie". Below these fields is a link "+ Add an index and one or more source types...". Under "Existing Datasets", there is a table with a single column labeled "# bytes" containing several numerical values. At the bottom of the configuration screen is a green "Done" button.

bytes
2958
2198
669
2223
1911
1130
3833
3906
006

On clicking done in the above screen, we get the final dataset table with all the selected fields, as seen below. Here the dataset has become similar to a relational table. We save the dataset with **save as** option available in the top right corner.

New Table Dataset

*	_time	# bytes	a categor...	IP clientip
1	2018-10-1 2T23:59:4 5.000+05 :30	2958	TEE	192.188.106.240
2	2018-10-1 2T23:59:4 3.000+05 :30	2198	ARCADE	212.235.92.150
3	2018-10-1 2T23:59:4 1.000+05: 30	669	null	212.235.92.150
4	2018-10-1 2T23:59:3 9.000+05 :30	2223	ARCADE	212.235.92.150
5	2018-10-1 2T23:59:	1911	null	192.188.106.240

Creating Pivot

We use the above dataset to create a pivot report. The pivot report reflects aggregation of values of one column with respect to the values in another column. In other words, one column's values are made into rows and another column's values are made into rows.

Choose Dataset Action

To achieve this, we first select the dataset using the dataset tab and then choose the option **Visualize with Pivot** from the Actions column for that data set.

The screenshot shows the Splunk Enterprise web interface. At the top, there's a navigation bar with links for Search, Datasets (which is highlighted with a green box), Reports, Alerts, and Dashboards. Below the navigation is a main content area titled "Datasets". A sub-instruction says: "Use the Datasets listing page to view and manage your existing datasets. Click a dataset name to view its contents. Click Pivot to design a visualization-rich report based on the dataset. Click Explore in Search to search for the dataset in Search and save it as a new report, alert, or dashboard panel." There's a link "Learn more about Datasets." with a help icon. Below this, a summary says "1 Datasets". A table lists one dataset: "Access Co..." (Type: table). To the right of the table are buttons for "Manage" and "Explore", and the word "admin" indicating ownership. A dropdown menu is open next to the "Actions" button, containing options like "Visualize with Pivot" (which is circled with a red box) and "Investigate in Search".

Choose the Pivot Fields

Next, we choose the appropriate fields for creating the pivot table. We choose category ID in the **split columns** option as this is the field whose values should appear as different columns in the report. Then we choose File in the **Split Rows** option as this is the field whose values should be presented in rows. The result shows count of each categoryid values for each value in the file field.

New Pivot

✓ 131,645 events (before 10/28/18 10:28:19.000 AM)

Filters		Split Columns			
All time			categoryId		
Split Rows		Column Values			
file			Count of Acc...		
file	ACCESSORIES	ARCADE	NULL	SHOOTER	SIMULATION
ADMIN	0	0	3	0	0
Admin	0	0	3	0	0
account	0	0	2	0	0
adm	0	0	1	0	0
admin	0	0	2	0	0
administration	0	0	1	0	0
anna_nicole.html	0	0	235	0	0
api	0	0	1	0	0
bdoor	0	0	1	0	0
cart.do	566	747	28341	377	372
category.screen	2793	3750	3062	1834	1775
door	0	0	1	0	0
error.do	0	0	1796	0	0

20 per page ▾

Next, we can save the pivot table as a Report or a panel in an existing dashboard for future reference.

16. Splunk – Lookups

In the result of a search query, we sometimes get values which may not clearly convey the meaning of the field. For example, we may get a field which lists the value of product id as a numeric result. These numbers will not give us any idea of what kind of product it is. But if we list the product name along with the product id, that gives us a good report where we understand the meaning of the search result.

Such linking of values of one field to a field with same name in another dataset using equal values from both the data sets is called a lookup process. The advantage is, we retrieve the related values from two different data sets.

Steps to Create and Use Lookup File

In order to successfully create a lookup field in a dataset, we need to follow the below steps:

Create Lookup File

We consider the dataset with host as web_application, and look at the productid field. This field is just a number, but we want product names to be reflected in our query result set. We create a lookup file with the following details. Here, we have kept the name of the first field as **productid** which is same as the field we are going to use from the dataset.

```
productId,productdescription
WC-SH-G04,Tablets
DB-SG-G01,PCs
DC-SG-G02,MobilePhones
SC-MG-G10,Wearables
WSC-MG-G10,Usb Light
GT-SC-G01,Battery
SF-BVS-G01,Hard Drive
```

Add the Lookup File

Next, we add the lookup file to Splunk environment by using the Settings screens as shown below:

The screenshot shows the Splunk administrative interface. At the top, there is a navigation bar with links for 'Administrat...', 'Messages', 'Settings' (which is highlighted with a green box), 'Activity', 'Help', and 'Find'. Below the navigation bar, there are two main columns of settings. The left column, under 'Add Data', includes 'Knowledge' (Searches, reports, and alerts; Data models; Event types; Tags; Fields; Lookups - which is highlighted with a green box), 'Monitoring Console' (User interface; Alert actions; Advanced search; All configurations), and 'System' (Server settings; Server controls; Health report manager; Instrumentation; Licensing; Workload management). The right column, under 'DATA', includes 'Data inputs', 'Forwarding and receiving', 'Indexes', 'Report acceleration summaries', 'Source types', 'DISTRIBUTED ENVIRONMENT' (Indexer clustering; Forwarder management; Distributed search), and 'USERS AND AUTHENTICATION' (Access controls).

After selecting the Lookups, we are presented with a screen to create and configure lookup. We select lookup table files as shown below.

The screenshot shows the 'Lookups' configuration page. At the top, there is a header with the 'splunk>enterprise' logo, 'Apps', 'Administrat...', 'Find', and a search icon. Below the header, the title 'Lookups' is displayed, followed by the sub-instruction 'Create and configure lookups.' The page is divided into three sections: 'Lookup table files' (List existing lookup tables or upload a new file. + Add new), 'Lookup definitions' (Edit existing lookup definitions or define a new file-based or external lookup. + Add new), and 'Automatic lookups' (Edit existing automatic lookups or configure a new lookup to run automatically. + Add new). The '+ Add new' button in the 'Lookup table files' section is highlighted with a green box.

We browse to select the file **productidvals.csv** as our lookup file to be uploaded and select search as our destination app. We also keep the same destination file name.

The screenshot shows the Splunk Enterprise web interface with the following details:

- Header:** splunk>enterprise, Apps ▾, Admin... ▾, Messages ▾, Settings ▾, Activity ▾.
- Title:** Add new
- Breadcrumb:** Lookups » Lookup table files » Add new
- Fields:**
 - Destination app:** search
 - Upload a lookup file:** Browse... prodcutidvals.csv
 - Destination filename *:** prodcutidvals.csv
- Help Text:** Select either a plaintext CSV file, a gzipped CSV file, or a KMZ/KML file. The maximum file size that can be uploaded through the browser is 500MB.
- Buttons:** Cancel (gray), Save (green)

On clicking the save button, the file gets saved to the Splunk repository as a lookup file.

Create Lookup Definitions

For a search query to be able to lookup values from the Lookup file we just uploaded above, we need to create a lookup definition. We do this by again going to **Settings -> Lookups -> Lookup Definition -> Add New.**

splunk>enterprise Apps ▾ Admin... ▾ Help ▾ Find

Add new

Lookups » Lookup definitions » Add new

Destination app: search

Name *: Productid_descriptions

Type: File-based

Lookup file *: prodcutidvals.csv

Create and manage lookup table files.

Configure time-based lookup

Advanced options

Cancel **Save**

Next, we check the availability of the lookup definition we added by going to **Settings -> Lookups -> Lookup Definition**.

Successfully saved "Productid_descriptions" in search.

Showing 1-6 of 6 items

Name	Type	Supported fields	Lookup file
Productid_descriptions	file	productId,productdescription	prodctidvals.csv
dnslookup	external	clienthost,clientip	
geo_attr_countries	file	country,region_wb,region_un,subre	geo_attr_countries.csv
geo_attr_us_states	file	state_name,state_fips,state_code	geo_attr_us_states.csv
geo_countries	geo	None	geo_countries.kmz

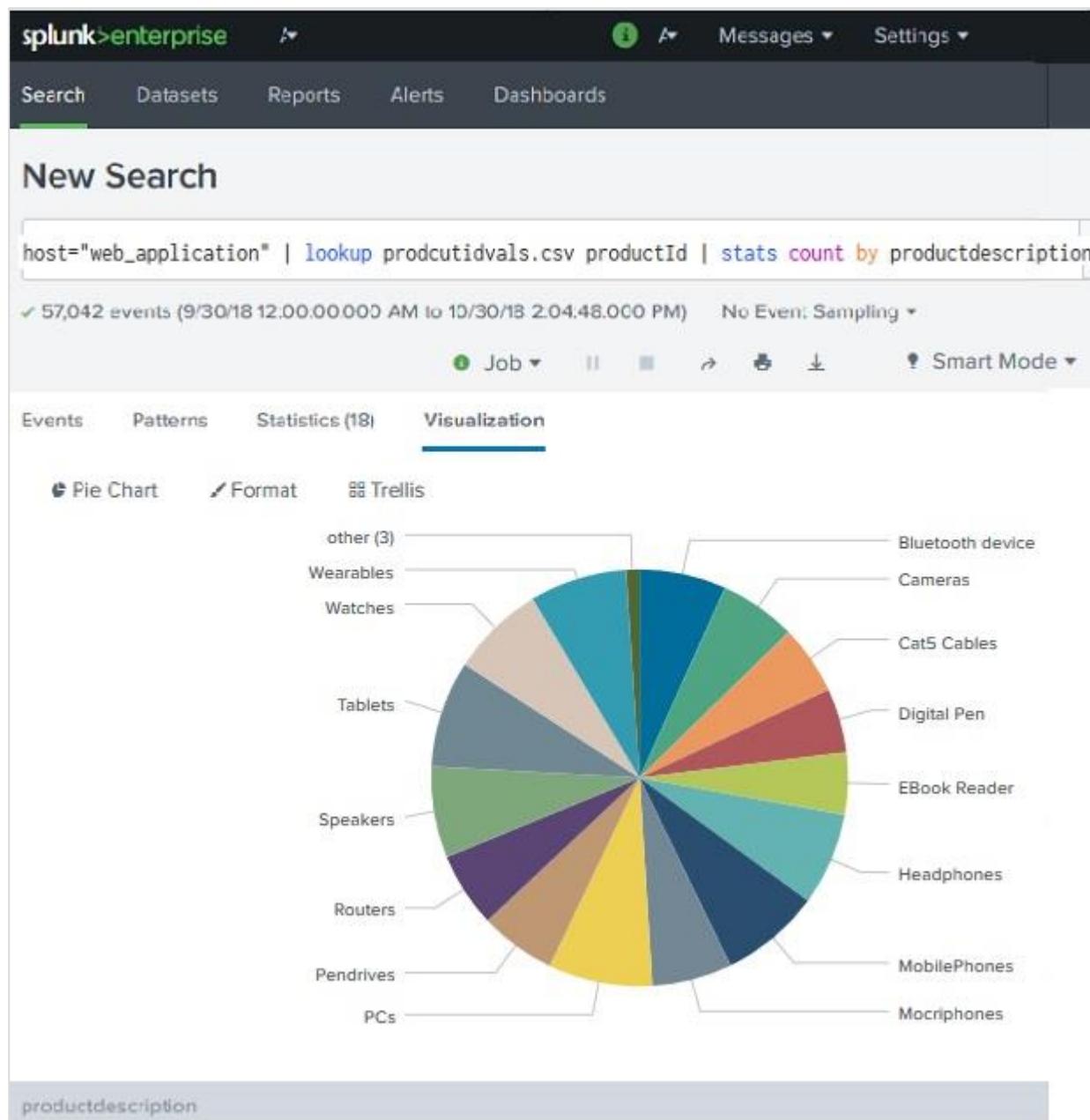
Selecting Lookup Field

Next, we need to select the lookup field for our search query. This is done by going to **New search -> All Fields**. Then check the box for **productid** which will automatically add the **productdescription** field from the lookup file also.

Select Fields				X
	Select All Within Filter	Dese	Filter	Q
i	✓ ▾ Field	# of Values	Event Coverage	Type
>	<input checked="" type="checkbox"/> bytes	>100	100%	Number
>	<input checked="" type="checkbox"/> date_wday	7	100%	String
>	<input checked="" type="checkbox"/> file	30	100%	String
>	<input checked="" type="checkbox"/> host	1	100%	String
>	<input checked="" type="checkbox"/> productId	18	74.49%	String
>	<input checked="" type="checkbox"/> productdescription	18	74.49%	String
>	<input checked="" type="checkbox"/> source	1	100%	String
>	<input checked="" type="checkbox"/> sourcetype	1	100%	String
>	<input type="checkbox"/> JSESSIONID	>100	99.6%	String
>	<input type="checkbox"/> action	5	49.54%	String
>	<input type="checkbox"/> categoryId	8	43.57%	String
>	<input type="checkbox"/> clientip	>100	100%	String
>	<input type="checkbox"/> date_hour	24	100%	Number
>	<input type="checkbox"/> date_mday	13	100%	Number
>	<input type="checkbox"/> date_minute	60	100%	Number
>	<input type="checkbox"/> date_month	2	100%	String

Using the Lookup Field

Now we use the Lookup field in the search query as shown below. The visualization shows the result with productdescription field instead of productId.



17. Splunk – Schedules and Alerts

Scheduling is the process of setting up a trigger to run the report automatically without the user's intervention. Below are the uses of scheduling a report:

- By running the same report at different intervals: monthly, weekly or daily, we can get results for that specific period.
- Improved performance of the dashboard as the reports finish running in the background before the dashboard is opened by the users.
- Sending of reports automatically via email after it finishes running.

Creating a Schedule

A schedule is created by editing the report's schedule feature. We go to the **Edit Schedule** option on the Edit button as shown in the image below.

The screenshot shows the Splunk Enterprise search interface. At the top, there is a navigation bar with links for 'splunk>enterprise', 'Messages', 'Settings', 'Activity', 'Help', 'Find', and a search icon. Below the navigation bar, there is a secondary navigation bar with links for 'Search', 'Datasets', 'Reports', and 'Alerts'. To the right of this is a 'Search & Reporting' section with a green arrow icon. The main content area displays a search result titled 'Files_By_Weekday' with the subtitle 'Show the count of files by weekday'. A green button labeled 'All time' is visible. To the right of the search results, there is a context menu with options: 'Edit', 'More Info', and 'Add to Dashboard'. The 'Edit' option is highlighted with a blue box. A sub-menu is open under 'Edit' with the following options: 'Open in Search', 'Edit Description', 'Edit Permissions', 'Edit Schedule' (which is highlighted with a green oval), 'Edit Acceleration', 'Clone', 'Embed', and 'Delete'. Below the search results, there is a table showing the count of files for each weekday. The table has columns for the day name ('date_wday') and the count ('count(file)'). The data is as follows:

date_wday	count(file)
friday	22775
monday	17754
saturday	16899
sunday	17217
thursday	21541
tuesday	17515
wednesday	17943

On clicking the edit schedule button, we get the next screen which lays out all the options for creating the schedule.

In the below example, we take all the default options and the report is scheduled to run every week on Monday at 6 AM.

Edit Schedule

Report Files_By_Weekday

Schedule Report

Learn More ↗

Schedule

On at

Time Range

Schedule Priority ?

Schedule Window ?

Trigger Actions

Important Features of Scheduling

The following are the important features of scheduling:

Time Range – It indicates the time range from which the report must fetch the data. It can be last 15 minutes, last 4 hours or last week etc.

Schedule Priority – If more than one report is scheduled at the same time then this will determine the priority of a specific report.

Schedule Window – When there are multiple report schedules with same priority then we can choose a time window which will help the report to run at anytime during this window. If it is 5 minutes, then the report will run within 5 minutes of its scheduled time.

This helps in enhancing the performance of the scheduled reports by spreading their run time.

Schedule Actions

The schedule actions are meant to take some steps after the report is run. For example, you may want to send an email stating the run status of the report or run another script. Such actions can be carried out by setting the option by clicking on **Add Actions** button as shown below:

Report Files_By_Weekday

Schedule Report

- Log Event
Send log event to Splunk receiver endpoint
- Output results to lookup
Output the results of the search to a CSV lookup file
- Output results to telemetry endpoint
Custom action to output results to telemetry endpoint
- Run a script
Invoke a custom script
- Send email

+ Add Actions ▾

Cancel Save

Alerts

Splunk alerts are actions which get triggered when a specific criterion is met which is defined by the user. The goal of alerts can be logging an action, sending an email or output a result to a lookup file, etc.

Creating an Alert

You create an alert by running a search query and saving its result as an alert. In the below screenshot, we take the search for daywise file count and save the result as an alert by choosing the **Save As** option.

The screenshot shows the Splunk Enterprise interface. The search bar contains the query: `host="web_application" | stats count(file) by date_wday`. Below the search bar, it says `✓ 131,645 events (before 10/25/18 4:07:08.000 PM)` and `No Event S`. The results table is titled "Statistics (7)" and lists the day of the week and the count of files. The "Alert" option in the "Save As" dropdown menu is highlighted with a green box.

date_wday	count(file)
friday	22775
monday	17754
saturday	16899
sunday	17217
thursday	21541
tuesday	17515
wednesday	17943

In the next screenshot, we configure the alert properties. The below image shows the configuration screen:

Save As Alert

Settings

Title	Alert File Size	
Description	Email Alert when the file size report is run	
Permissions	Private	Shared in App
Alert type	Scheduled	Real-time
Run every week ▾		
On	Monday ▾	at 6:00 ▾

Trigger Conditions

Trigger alert when	Number of Results ▾	
	is greater than ▾	0
Trigger	Once	For each result
Throttle ?	<input type="checkbox"/>	

Trigger Actions

+ Add Actions ▾

The purpose and choices of each of these options is explained below:

- **Title:** It is the name of the alert.
- **Description:** It is the detailed description of what the alert does.
- **Permission:** Its value decided who can access, run or edit the alert. If declared private, then only the creator of the alert has all the permissions. To be accessed

by others the option should be changed to **Shared in App**. In this case everyone has read access but only power user has the edit access for the alert.

- **Alert Type:** A scheduled alert runs at a pre-defined interval whose run time is defined by the day and time chosen from the drop downs. But the other option on real-time alert causes the search to run continuously in the background. Whenever the condition is met, the alert action is executed.
- **Trigger condition:** The trigger condition checks for the criteria mentioned in the trigger and sets off the alert only when the alert criteria is met. You can define number of results or number of sources or number of hosts in the search result to trigger the alert. If it is set for once, it will execute only once when the result condition is met but if it is set to **For each Result**, then it will run for every row in the result set where the trigger condition is met.
- **Trigger Actions:** The trigger actions can give a desired output or send an email when the trigger condition is met. The below image shows some of the important trigger actions available in Splunk.

Save As Alert

Settings

Title: Alert File Size

Description: Email Alert when the file size report is run

Permissions: Private | Shared in App

Alert type: Scheduled | Real-time

Schedule: Run every week ▾
On: Monday ▾ at: 6:00 ▾

Triggered Actions:

- Add to Triggered Alerts
- Add this alert to Triggered Alerts list
- Log Event**
Send log event to Splunk receiver endpoint
- Output results to lookup**
Output the results of the search to a CSV lookup
- Output results to telemetry endpoint
Custom action to output results to telemetry endpoint
- Run a script

+ Add Actions ▾

Number of Results ▾
0
For each result

Cancel Save

18. Splunk – Knowledge Management

Splunk knowledge management is about maintenance of knowledge objects for a Splunk Enterprise implementation.

Below are the **main features of knowledge management**:

- Ensure that knowledge objects are being shared and used by the right groups of people in the organization.
- Normalize event data by implementing knowledge object naming conventions and retiring duplicate or obsolete objects.
- Oversee strategies for improved search and pivot performance (report acceleration, data model acceleration, summary indexing, batch mode search).
- Build data models for Pivot users.

Knowledge Object

It is a Splunk object to get specific information about your data. When you create a knowledge object, you can keep it private or you can share it with other users. The examples of knowledge object are: saved searches, tags, field extractions, lookups, etc.

Uses of Knowledge Objects

On using the Splunk software, the knowledge objects are created and saved. But they may contain duplicate information, or they may not be used effectively by all the intended audience. To address such issues, we need to manage these objects. This is done by classifying them properly and then using proper permission management to handle them. Below are the uses and classification of various knowledge objects:

Fields and field extractions

Fields and field extractions is the first layer of Splunk software knowledge. The fields automatically extracted from the Splunk software from the IT data help bring meaning to the raw data. The manually extracted fields expand and improve upon this layer of meaning.

Event types and transactions

Use event types and transactions to group together interesting sets of similar events. Event types group together sets of events discovered through searches. Transactions are collections of conceptually-related events that span time.

Lookups and workflow actions

Lookups and workflow actions are categories of knowledge objects that extend the usefulness of your data in various ways. Field lookups enable you to add fields to your data from external data sources such as static tables (CSV files) or Python-based

commands. Workflow actions enable interactions between fields in your data and other applications or web resources, such as a WHOIS lookup on a field containing an IP address.

Tags and aliases

Tags and aliases are used to manage and normalize sets of field information. You can use tags and aliases to group sets of related field values together, and to give extracted field tags that reflect different aspects of their identity. For example, you can group events from set of hosts in a particular location (such as a building or city) together by giving the same tag to each host.

If you have two different sources using different field names to refer to same data, then you can normalize your data by using aliases (by aliasing clientip to ipaddress, for example).

Data models

Data models are representations of one or more datasets, and they drive the Pivot tool, enabling Pivot users to quickly generate useful tables, complex visualizations, and robust reports without needing to interact with the Splunk software search language. Data models are designed by knowledge managers who fully understand the format and semantics of their indexed data. A typical data model makes use of other knowledge object types.

We will discuss some of the examples of these knowledge objects in the subsequent chapters.

19. Splunk – Subsearching

Subsearch is a special case of the regular search when the result of a secondary or inner query is the input to the primary or outer query. It is similar to the concept of subquery in case of SQL language. In Splunk, the primary query should return one result which can be input to the outer or the secondary query.

When a search contains a subsearch, the subsearch is run first. Subsearches must be enclosed in square brackets in the primary search.

Example

We consider the case of finding a file from web log which has maximum byte size. But that may vary every day. Then we want to find only those events where the file size is equal to the maximum size, and is a Sunday.

Create the Subsearch

We first create the subsearch to find the maximum file size. We use the function **Stat max** with the field named bytes as the argument. This identifies the maximum size of the file for the time frame for which the search query is run.

The below image shows the search and the result of this subsearch:

The screenshot shows the Splunk Enterprise search interface. The search bar contains the following command:

```
host="web_application"
| stats max(bytes) as bytes
```

The results pane shows the following information:

- 131,645 events (before 10/31/18 6:13:16.000 PM)
- No Event Sampling
- Statistics (1) tab selected
- bytes: 47251

Adding the Subsearch

Next, we add the subsearch query to the primary or the outer query by putting the subsearch inside square brackets. Also the search clause is added to the subsearch query.

The screenshot shows the Splunk Enterprise web interface. At the top, there's a navigation bar with 'splunk enterprise' logo, user profile, 'Messages', 'Settings', and other links. Below it is a secondary navigation bar with 'Search', 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. The main area is titled 'New Search'.

In the search bar, the following SPL query is entered:

```
host="web_application" date_wday="sunday"
[search host="web_application"
| stats max(bytes) as bytes ]
```

Below the search bar, it says '38 events (before 10/31/18 3:41:41.000 PM)' and 'No Event Sampling'. There are buttons for 'Job', 'Smart Mode', and other search controls.

The timeline visualization shows three green bars representing event counts for different days in September 2018. The x-axis labels are 'Mon Sep 17 2018', 'Fri Sep 21', 'Tue Sep 25', and 'Sat Sep 29'.

Below the timeline, there are filters: 'Format Timeline', 'Zoom Out', 'Zoom to Selection', and 'Deselect'. The event list is set to '20 Per Page'.

The event list table has columns: 'Time' and 'Event'. The first few events are listed:

Time	Event
10/7/18 3:39:41.000 PM	27.1.11.11 - - [07/Oct/2018:15:39:41] "GET /passwords.pdf HTTP 1.1" 200 type:pdf+passwords&start=90" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/41.0.2228.118 Safari/536.5" 908 bytes = 47251 date_hour = 15 date_mday = 7 date_wday = sunday file host = web_application source = access_30DAY.log sourcetype = access_c
10/7/18 3:39:10.000 PM	27.1.11.11 - - [07/Oct/2018:15:39:10] "GET /passwords.pdf HTTP 1.1" 200 type:pdf+passwords&start=90" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/41.0.2228.118 Safari/536.5" 511 bytes = 47251 date_hour = 15 date_mday = 7 date_wday = sunday file host = web_application source = access_30DAY.log sourcetype = access_c
10/7/18 2:13:02.000 PM	203.45.206.135 - - [07/Oct/2018:14:13:02] "GET /passwords.pdf HTTP 1.1" 200 filetype:pdf+passwords&start=90" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; .NET CLR 2.0.50727; .NET CLR 3.0.4506.2152; .NET CLR 3.5.30729)" 47251 bytes = 47251 date_hour = 14 date_mday = 7 date_wday = sunday file host = web_application source = access_30DAY.log sourcetype = access_c

As we see, the result contains only the events where the file size is equal to the max file size found by considering all the events, and the event day is a Sunday.

20. Splunk – Search Macros

Search macros are reusable blocks of Search Processing Language (SPL) that you can insert into other searches. They are used when you want to use the same search logic on different parts or values in the data set dynamically. They can take arguments dynamically and the search result will be updated as per the new values.

Macro Creation

To create the search macro, we go to the **Settings -> Advanced Search -> Search Macros -> Add New**. This brings up the below screen where we start creating the macro.

The screenshot shows the Splunk Enterprise web interface with the following details:

- Header:** splunk>enterprise, i, Messages ▾, S
- Title:** Add new
- Breadcrumb:** Advanced search > Search macros > Add new
- Destination app:** search
- Name:** (Placeholder: Enter the name of the macro. If the search macro takes an argument, indicate this by appending the number of arguments to the name. For example: mymacro(2))
- Definition:** (Placeholder: Enter the string the search macro expands to when it is referenced in another search. If arguments are included, enclose them in dollar signs. For example: \$arg1\$)
 Use eval-based definition?
- Arguments:** (Placeholder: Enter a comma-delimited string of argument names. Argument names may only contain alphanumeric, '_' and '-' characters.)
- Validation Expression:** (Placeholder: Enter an eval or boolean expression that runs over macro arguments.)
- Validation Error Message:** (Placeholder: Enter a message to display when the validation expression returns)
- Buttons:** Cancel, Save

Macro Scenario

We want to show various stats about the file size from the **web_applications** log. The stats are about max, min and avg value of the filesize using the bytes field in the log. The result should display these stats for each file listed in the log.

So here the type of the stats is dynamic in nature. The name of the stats function will be passed as an argument to the macro.

Defining the Macro

Next, we define the macro by setting various properties as shown in the below screen. The name of the macro contains (1), indicating that there is one argument to be passed into the macro when it is used in the search string. **fun** is the argument which will be passed on to the macro during execution in the search query.

splunk>enterprise ▾ i ↶ Messages ▾ 8

Add new

Advanced search » Search macros » Add new

Destination app search ▾

Name * Enter the name of the macro. If the search macro takes an argument, indicate this by appending the number of arguments to the name.
For example: mymacro(2)

`filesize(1)`

Definition * Enter the string the search macro expands to when it is referenced in another search. If arguments are included, enclose them in dollar signs.
For example: \$arg1\$

`stats fun(bytes) by file`

Use eval-based definition?

Arguments Enter a comma-delimited string of argument names. Argument names may only contain alphanumeric, '_' and '-' characters.

`fun`

Validation Expression Enter an eval or boolean expression that runs over macro arguments.

Validation Error Message Enter a message to display when the validation expression returns

Cancel Save

Using the Macro

To use the macro, we make it a part of the search string. On passing different values for the argument we see different results as expected.

Consider finding the average size in bytes of the files. We pass avg as the argument and get the result as shown below. The macro has been kept under ` sign as part of the search query.

The screenshot shows the Splunk Enterprise search interface. The top navigation bar includes 'splunk>enterprise' on the left, a dropdown menu in the center, and 'Messages' on the right. Below the navigation bar, there are tabs for 'Search' (which is selected), 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. The main area is titled 'New Search' and contains the search command:

```
host="web_application" | `filesize(avg)`
```

Below the search command, a message indicates there are 48,444 events from 10/2/18 to 11/1/18. A 'No Event Sampling' dropdown is present. The interface includes standard search controls like 'Job', 'Events', 'Patterns', 'Statistics (30)', 'Visualization', and a 'Format' dropdown set to '50 Per Page'. The 'Statistics' section displays a table of file names and their average byte sizes:

file	avg(bytes)
ADMIN	3406
Admin	3406
account	2119
adm	3406
admin	3406
administration	3406
anna_nicole.html	1990.788888888888
api	1456
bdoor	3406
cart.do	2083.075435525548

Similarly, if we want the maximum file size for each of the files present in the log, then we use **max** as the argument. The result is as shown below.

21. Splunk – Event Types

In Splunk search, we can design our own events from a dataset based on certain criteria. For example, we search for only the events which have a http status code of 200. This event now can be saved as an event type with a user defined name as **status200** and use this event name as part of future searches.

In short, an event type represents a search that returns a specific type of event or a useful collection of events. Every event that can be returned by the search gets an association with that event type.

Creating Event Type

There are two ways to create an event type after we have decided the search criteria. One is to **run a search** and then save it as an Event Type. Another is to **add a new Event Type from the settings tab**. We will see both the ways of creating it in this section.

Using a Search

Consider the search for the events which have the criteria of successful http status value of 200 and the event type run on a Wednesday. After running the search query, we can choose **Save As** option to save the query as an Event Type.

host="web_application" status=200 date_wday="Wednesday"

12:00:00.000 AM to 11/4/18 1:53:47.000 PM) No Event Samplir

Job ▾ II ■ ➡ 🔍 ⌂ Smart Mo

terns Statistics Visualization

- Zoom Out + Zoom to Selection × Deselect

Oct 8 Mon Oct 15 Mon Oct 22

ist ▾ Format 20 Per Page ▾ < Prev 1 2 3 4 5

Event

```
88.191.83.82 - - [10/Oct/2018:23:57:34] "GET /product.screen?productId=MB-AG-TI
HTTP 1.1" 200 3835 "http://www.buttercupgames.com/category.screen?categoryId=TI
el Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 :
bytes = 3835 | date_hour = 23 | date_mday = 10 | date_wday = wednesday | file =
host = web_application | productId = MB-AG-T01 | source = access_30DAY.log | sta
```

```
88.191.83.82 - - [10/Oct/2018:23:57:15] "GET /cart.do?action=view&productId=MB-
F4953 HTTP 1.1" 200 1569 "http://www.buttercupgames.com/cart.do?action=view&pr
(Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/
bytes = 1569 | date_hour = 23 | date_mday = 10 | date_wday = wednesday | file =
host = web_application | productId = MB-AG-G07 | source = access_30DAY.log | sta
```

The next screen prompts to give a name for the Event Type, choose a Tag which is optional and then choose a colour with which the events will be highlighted. The priority option

decides which event type will be displayed first in case two or more event types match the same event.

Save As Event Type

Name	successful_wed
Tags	Optional
Color	purple ▾
Priority	1 (Highest) ▾

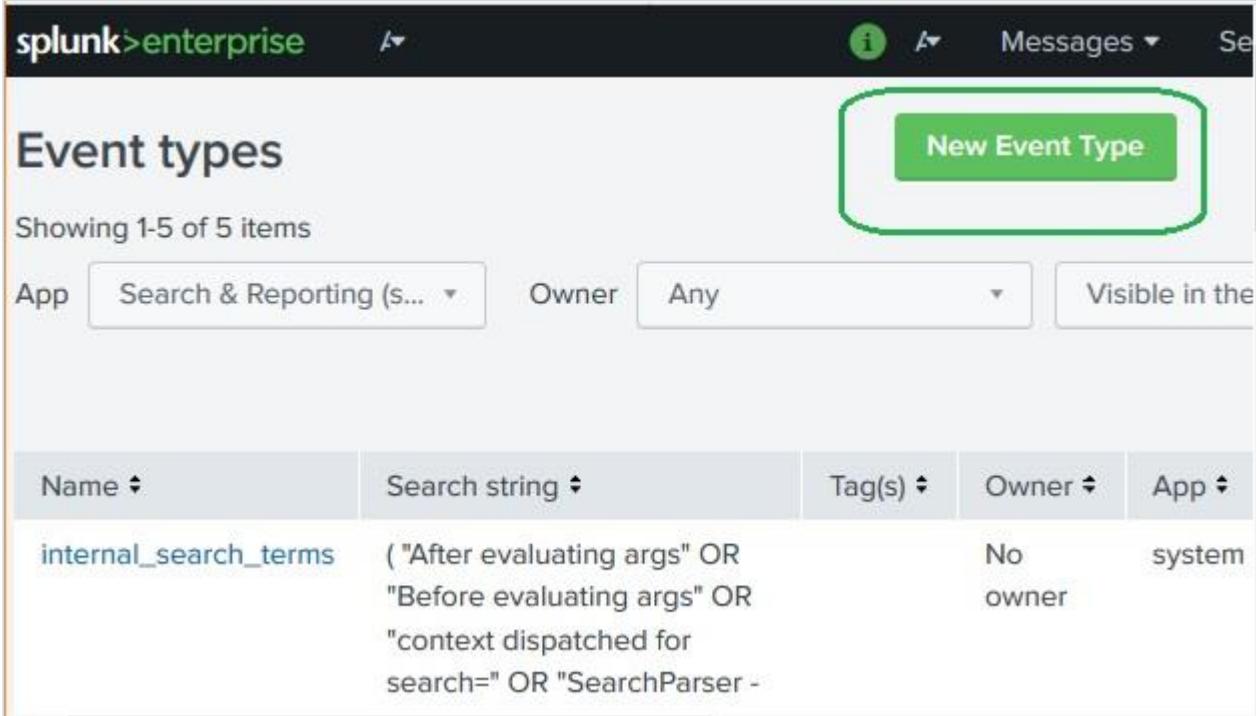
Determines which style wins, when an event has more than one event type.

Cancel **Save**

Finally, we can see the Event Type has been created by going to the **Settings -> Event Types** option.

Using New Event Types

The other option to create a new Event Type is to use the **Settings -> Event Types** option as shown below where we can add a new Event Type:



The screenshot shows the Splunk Enterprise interface for managing event types. At the top, there's a navigation bar with the 'splunk>enterprise' logo, a gear icon, 'Messages' dropdown, and a search bar. Below the header, the title 'Event types' is displayed, followed by a message 'Showing 1-5 of 5 items'. There are several filter options: 'App' set to 'Search & Reporting (s...)', 'Owner' set to 'Any', and 'Visible in the...' dropdown. A prominent green button labeled 'New Event Type' is highlighted with a red rounded rectangle. The main content area is a table with columns: 'Name', 'Search string', 'Tag(s)', 'Owner', and 'App'. One row is visible, showing an event type named 'internal_search_terms' with a complex search string involving 'After evaluating args', 'Before evaluating args', and 'context dispatched for search=' OR 'SearchParser -'. The owner is listed as 'No owner' and the app is 'system'.

On clicking the button **New Event Type**, we get the following screen to add the same query as in the previous section.

splunk>enterprise Apps Admin... Messages Settings

Add new

Event types » Add new

Destination App: search

Name *: successful_wed

Search string *: host="web_application" status=200 date_wday="Wednesday"

Tag(s):
Enter a comma-separated list of tags.

Color: purple

Priority: 1 (Highest)
Highest priority shows up first in a result.

Cancel Save

Viewing the Event Type

To view the event we just created above, we can write the below search query in the search box and we can see the resulting events along with the colour we have chosen for the event type.

splunk>enterprise ▾ i ▾ Messages ▾

Search Datasets Reports Alerts Dashboards

New Search

Save As ▾ New Table Close

eventtype="successful_wed"

Previous month 🔍

✓ 7,442 events (10/1/18 12:00:00.000 AM to 11/1/18 12:00:00.000 AM) No Event Sampling

Job ▾ Smart Mode

Events (7,442) Patterns Statistics Visualization

Format Timeline ▾ Zoom Out + Zoom to Selection × Deselect

4,000
3,000
2,000
1,000

Mon Oct 1 2018 | Mon Oct 8 | Mon Oct 15

> Show Fields List ▾ Format 20 Per Page ▾ < Prev

i	Time	Event
»	10/10/18 11:57:34.000 PM	88.191.83.82 - - [10/Oct/2018:23:57:34] "GET /product.sc HTTP/1.1" 200 3835 "http://www.buttercupgames.com/catego el Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko bytes = 3835 date_hour = 23 date_mday = 10 date_w host = web_application productId = MB-AG-T01 source =
»	10/10/18 11:57:15.000 PM	88.191.83.82 - - [10/Oct/2018:23:57:15] "GET /cart.do?ac F4953 HTTP/1.1" 200 1569 "http://www.buttercupgames.com/ (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KH bytes = 1569 date_hour = 23 date_mday = 10 date_w host = web_application productId = MB-AG-G07 source =

Using the Event Type

We can use the Event type along with other queries. Here we specify some partial criteria from the Event Type and the result is a mix of events which shows the coloured and non-coloured events in the result.

New Search

Save As ▾ New Table Close

host="web_application" file=cart.do date_wday="wednesday" Last 30 days

✓ 2,020 events (10/1/18 12:00:00.000 AM to 11/1/18 12:00:00.000 AM) No Event Sampling

Job ▾ || ⌂ ⌄ ⌅ ⌆ Smart Mode

Events (2,020) Patterns Statistics Visualization

Format Timeline ▾ — Zoom Out + Zoom to Selection × Deselect

1,200
800
400

Mon Oct 1 2018 | Mon Oct 8 | Mon Oct 15

> Show Fields List ▾ ✓ Format 20 Per Page ▾ < Prev

i	Time	Event
>	10/10/18 11:22:53.000 PM	87.194.216.51 -- [10/Oct/2018:23:22:53] "GET /cart.do?acti F5ADFF4953 HTTP/1.1" 200 2877 "http://www.buttercupgames.co zilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.2.28; 0729; .NET4.0C)" 333 bytes = 2877 date_hour = 23 date_mday = 10 date_wday = 6 host = web_application productId = MB-AG-G07 source = access_30DAY.log status = 200
>	10/10/18 11:22:48.000 PM	194.8.74.23 -- [10/Oct/2018:23:22:48] "GET /cart.do?acti 20 "http://www.buttercupgames.com/category.screen?category ebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/ bytes = 720 date_hour = 23 date_mday = 10 date_wday = 6 host = web_application source = access_30DAY.log status = 200
>	10/10/18 11:22:33.000 PM	194.8.74.23 -- [10/Oct/2018:23:22:33] "GET /cart.do?acti 4963 HTTP/1.1" 200 628 "http://www.buttercupgames.com/cart indows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) bytes = 628 date_hour = 23 date_mday = 10 date_wday = 6 host = web_application productId = CLIPCG06 source = access_30DAY.log status = 200

22. Splunk – Basic Chart

Splunk has great visualization features which shows a variety of charts. These charts are created from the results of a search query where appropriate functions are used to give numerical outputs.

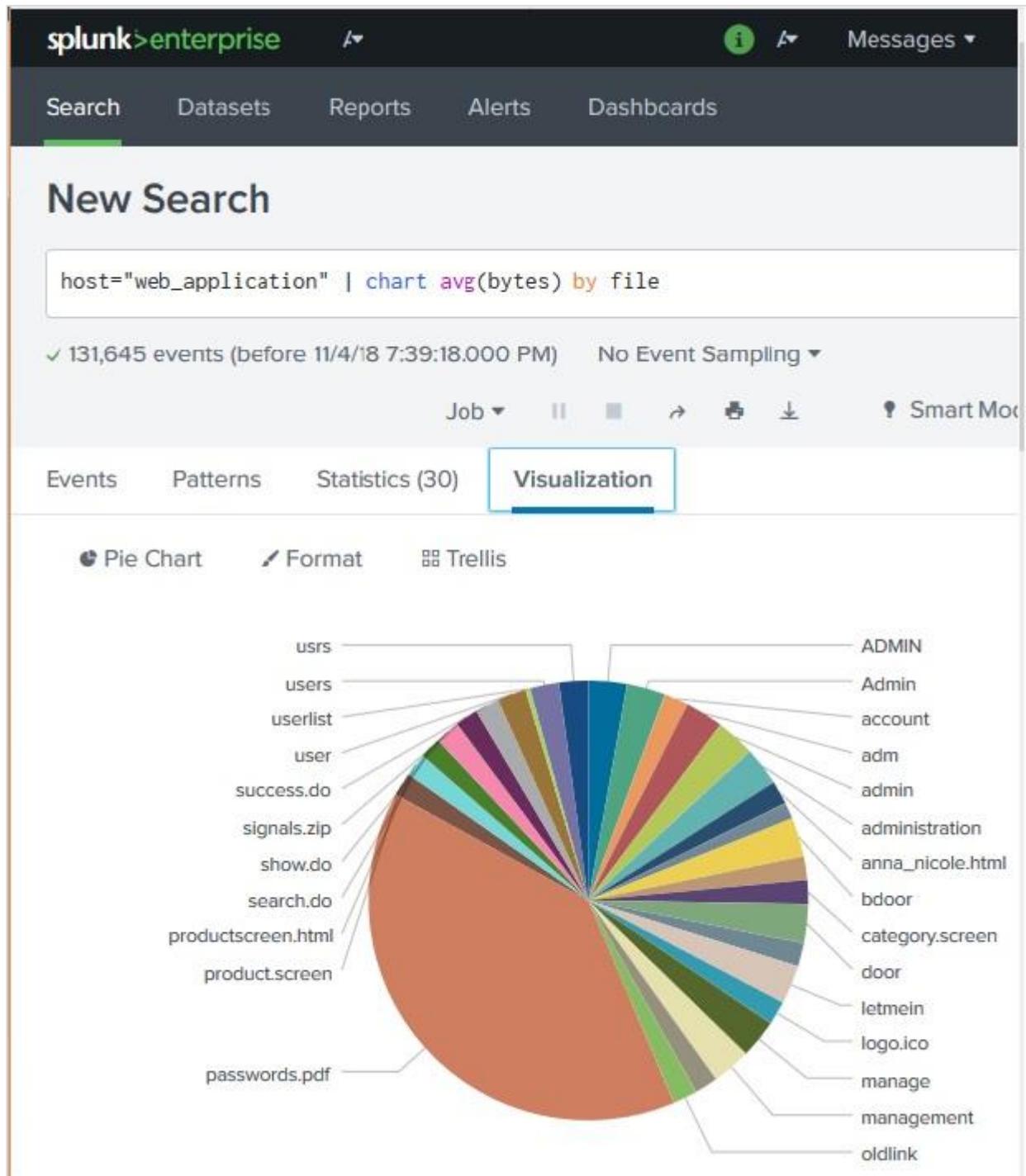
For example, if we look for the average file size in bytes from the data set named web_applications, we can see the result in the statistics tab as shown below:

The screenshot shows the Splunk Enterprise search interface. The search bar contains the command: `host="web_application" | chart avg(bytes) by file`. Below the search bar, it says `✓ 131,645 events (before 11/4/18 3:48:00.000 PM)` and `No Event Sampling`. The Statistics tab is selected, showing a table of file names and their average byte sizes. The table data is as follows:

file	avg(bytes)
ADMIN	3406
Admin	3406
account	2119
adm	3406
admin	3406
administration	3406
anna nicole.html	2102.876595744681

Creating Charts

In order to create a basic chart, we first ensure that the data is visible in the statistics tab as shown above. Then we click on the Visualization tab to get the corresponding chart. The above data produces a pie chart by default as shown below.



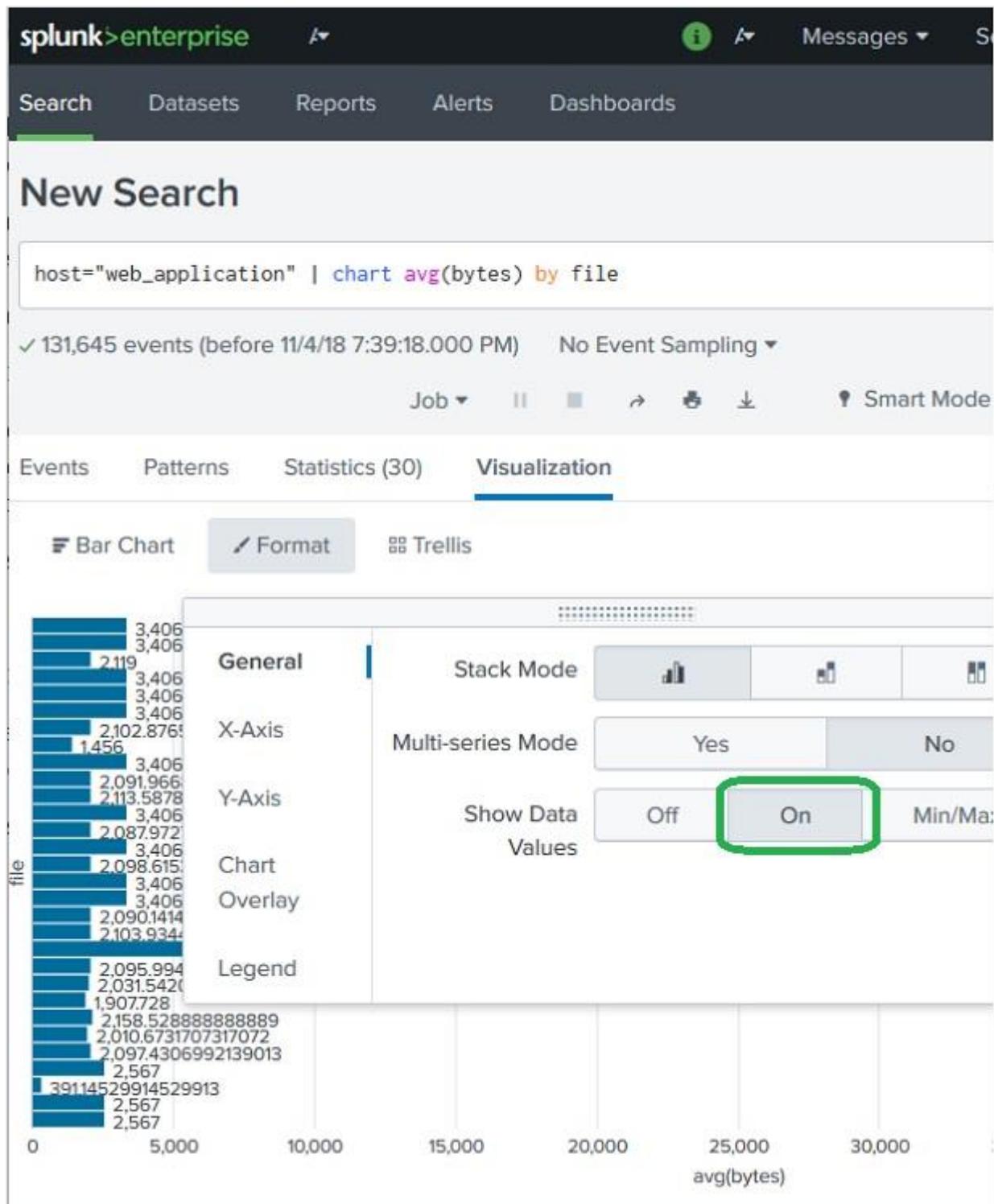
Changing the Chart Type

We can change the chart type by selecting a different chart option from the chart name. Clicking on one of these options will produce the chart for that type of graph.

The screenshot shows the Splunk Enterprise interface. At the top, there's a navigation bar with 'splunk>enterprise' and various links like 'Search', 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. Below the search bar, the title 'New Search' is displayed. The search query is: `host="web_application" | chart avg(bytes) by file`. It shows 131,645 events from before 11/4/18 7:39:18.000 PM with 'No Event Sampling'. The 'Visualization' tab is selected, and under it, the 'Pie Chart' option is also selected. A large pie chart on the right side represents the data from the search query. On the left, there's a 'Splunk Visualizations' section with icons for various chart types like line, area, bar, histogram, and scatter plots, along with a 'Find more visualizations' link. Below this, a 'Pie Chart' section is shown with the text 'Compare categories in a dataset.' and a search fragment: `| stats count by comparison_category`.

Formatting a Chart

The charts can also be formatted by using the Format option. This option allows to set the values for the axes, set the legends or show the data values in the chart. In the below example, we have chosen the horizontal chart and selected the option to show the data values as a Format option.

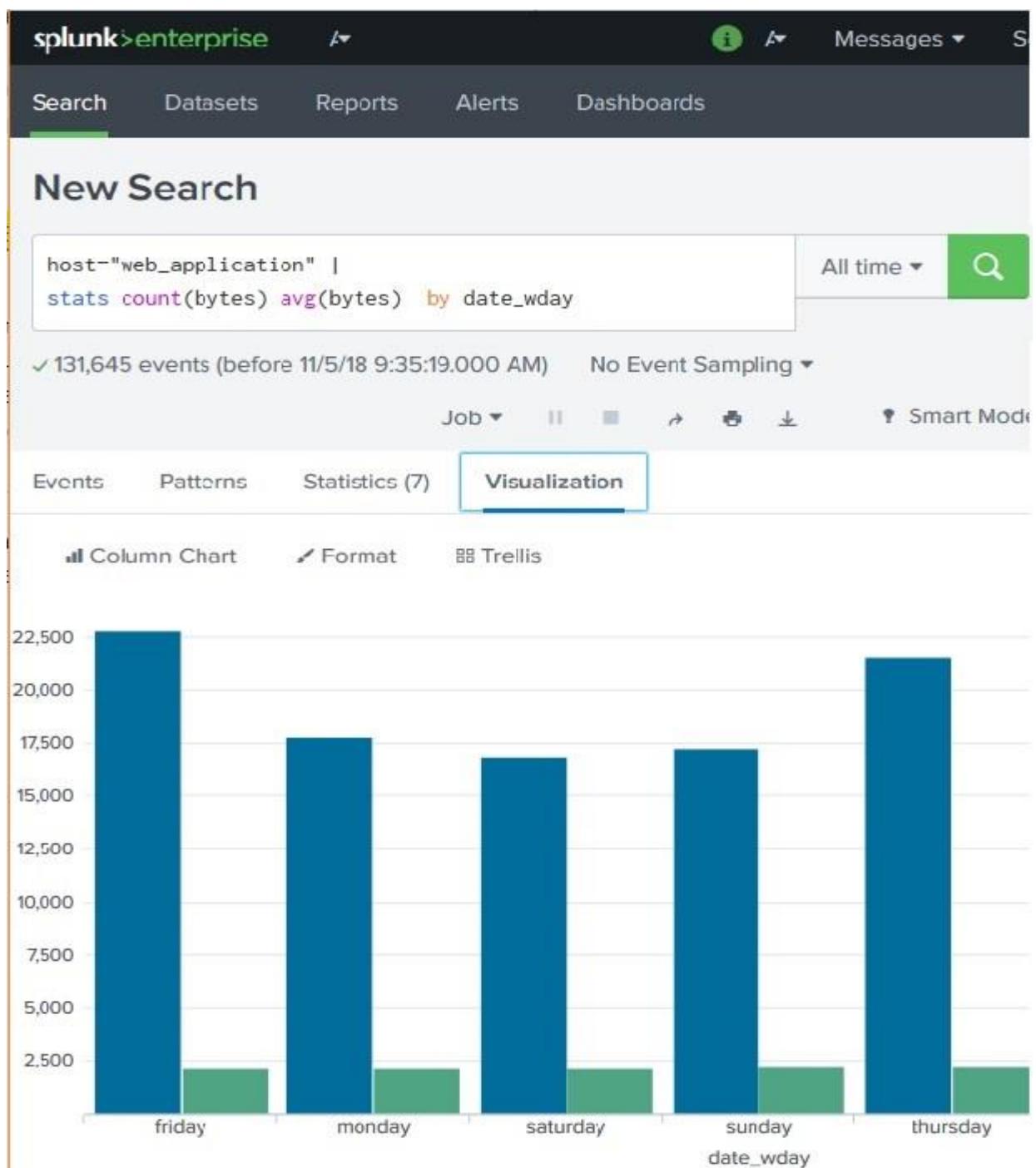


23. Splunk – Overlay Chart

Many times, we need to put one chart over another to compare or see the trend of the two charts. Splunk supports this feature through the chart overlay feature available in its visualization tab. To create such a chart, we need to first make a chart with two variables and then add a third variable which can create the overlay chart.

Chart Scenario

Continuing the examples from previous chapter, we find out the byte size of the files on different week days and then also add the average byte size for those days. The below image shows the chart showing the byte size versus average byte size of files on different days of the week.



Next, we are going to add the statistical function called standard deviation to the above search query. This will bring the additional variable needed to create the chart overlay. The below image shows the statistics of the query result which will be used in the visualization.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk>enterprise' on the left, a user icon, 'Messages', and a gear icon on the right. Below the bar, there are tabs for 'Search', 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. The 'Search' tab is highlighted.

The main area is titled 'New Search'. It contains a search bar with the query:

```
host="web_application" |  
stats count(bytes) avg(bytes) stdev(bytes) by date_wday
```

Below the search bar, it says '131,645 events (before 11/5/18 8:36:56.000 AM)' and 'No Event Sampling'. There are buttons for 'Save As', 'New Table', and 'Close'.

Below the search bar, there are several controls: 'Job', a timeline, and 'Smart Mode'.

The table has four columns: 'date_wday', 'count(bytes)', 'avg(bytes)', and 'stdev(bytes)'. The data rows are:

date_wday	count(bytes)	avg(bytes)	stdev(bytes)
friday	22775	2159.2494840834247	2016.6553106950907
monday	17754	2160.1039202433253	2076.110516511169
saturday	16899	2169.882359902953	2107.12103664981
sunday	17217	2207.1629784515303	2386.1347734331075
thursday	21542	2188.988580447498	2357.4705135356016
tuesday	17515	2186.973222951756	2240.1489907775485
wednesday	17943	2179.3207378922143	2200.784409479441

Creating Chart Overlay

To create the chart overlay, we follow **Visualization -> Format -> Chart Overlay**

This brings up a pop-up window where we need to choose the field which will be the overlay chart. In this case, we choose `stdev(bytes)` as the field as shown in the image below. We can also fill in other values: title, scale and their intervals, minimum values, maximum values, etc. For our example, we choose the default values after selecting the field for the overlay option.

The screenshot shows the Splunk Enterprise interface. At the top, there's a navigation bar with links for Search, Datasets, Reports, Alerts, and Dashboards. Below this is a section titled "New Search" containing a search command:

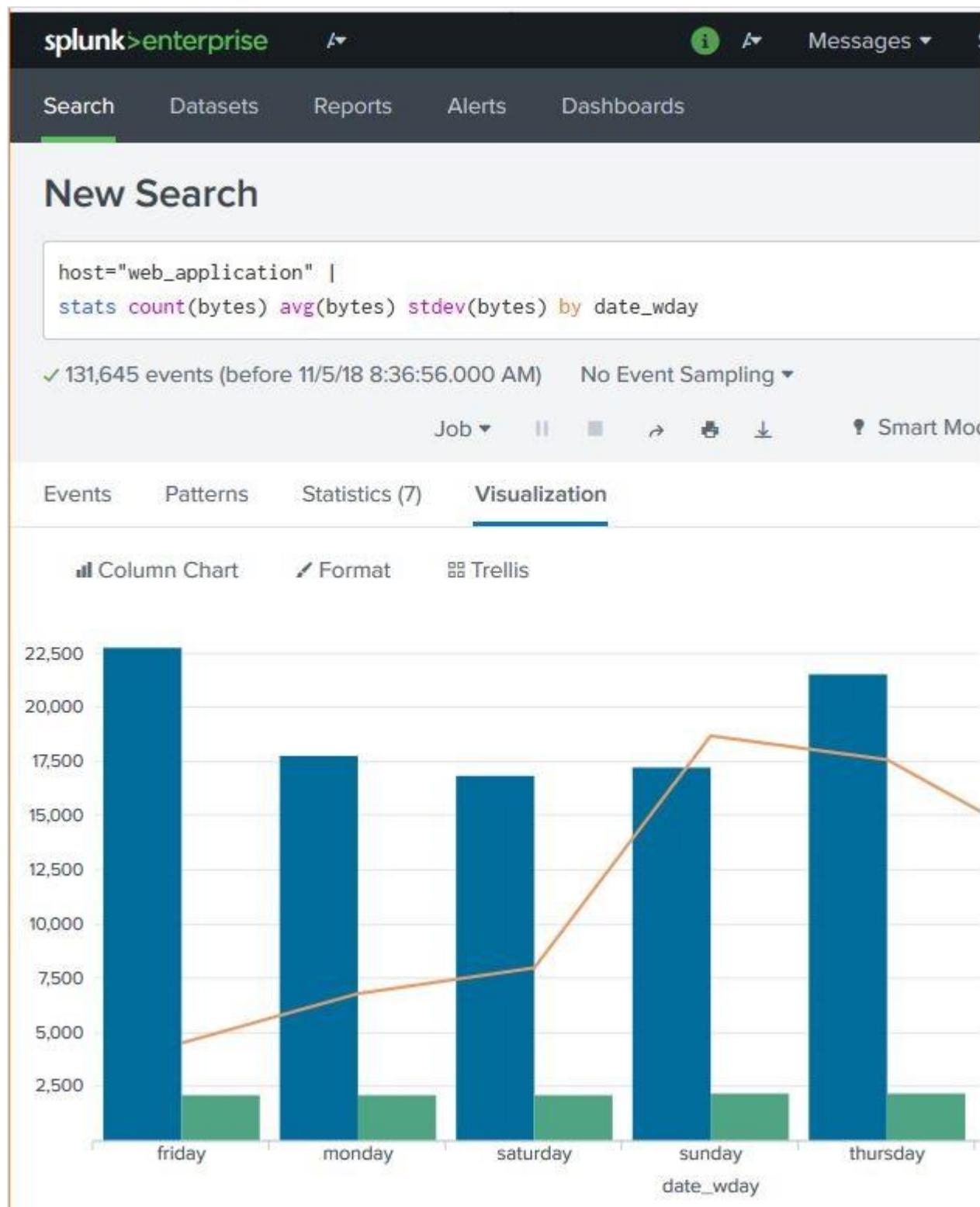
```
host="web_application" | stats count(bytes) avg(bytes) stdev(bytes) by date_wday
```

Below the search command, it says "✓ 131,645 events (before 11/5/18 8:36:56.000 AM)" and "No Event Sampling". There are various job-related buttons like Job, II, and download.

The main content area has tabs for Events, Patterns, Statistics (7), and Visualization. The Visualization tab is selected, showing options for Column Chart, Format (which is highlighted with a green box), and Trellis.

A large pop-up window titled "Chart Overlay" is open. It has a sidebar with categories: General, X-Axis, Y-Axis, Legend, and Chart Overlay (which is also highlighted with a green box). The main content area contains settings for Overlay, View as Axis, Title, Scale, Interval, Min Value, Max Value, and Number Abbreviations. The "Overlay" section has a field where "stdev(bytes)" is entered, which is also highlighted with a green box.

After selecting the above options, we can close the chart overlay pop-up window and see the final chart as shown below:



24. Splunk – Sparklines

A sparkline is a small representation of some statistical information without showing the axes. It generally appears as a line with bumps just to indicate how certain quantity has changed over a period of time. Splunk has in-built function to create sparklines from the events it searches. It is a part of the chart creation function.

Selecting the Fields

We need to select the field and the search formula which will be used in creating the sparkline. The below image shows the average byte size values of the some of the files in the web_application host.

The screenshot shows the Splunk Enterprise interface. At the top, there's a navigation bar with 'splunk>enterprise' on the left, followed by 'Search', 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. On the right, there are icons for help, messages, and a dropdown menu. Below the navigation bar, the title 'New Search' is displayed. The search bar contains the query: 'host="web_application" | chart avg(bytes) by file'. To the right of the search bar are buttons for 'All time' and a magnifying glass icon. Below the search bar, a message indicates '52,881 events (10/1/18 12:00:00.000 AM to 11/1/18 12:00:00.000 AM)' and 'No Event Sampler'. Underneath the search bar, there are buttons for 'Job', 'Events', 'Patterns', 'Statistics (30)', 'Visualization', '50 Per Page', 'Format', and 'Preview'. The 'Statistics (30)' button is highlighted with a blue border. The main area displays a table of file names and their average byte sizes:

file	avg(bytes)
product.screen	2095.9946449218037
productscreen.html	2031.5420168067226
search.do	1907.728
show.do	2158.5288888888889
signals.zip	2010.6731707317074
success.do	2097.4306992139013
user	2567
userlist	391.14529914529913
users	2567

Creating the Sparkline

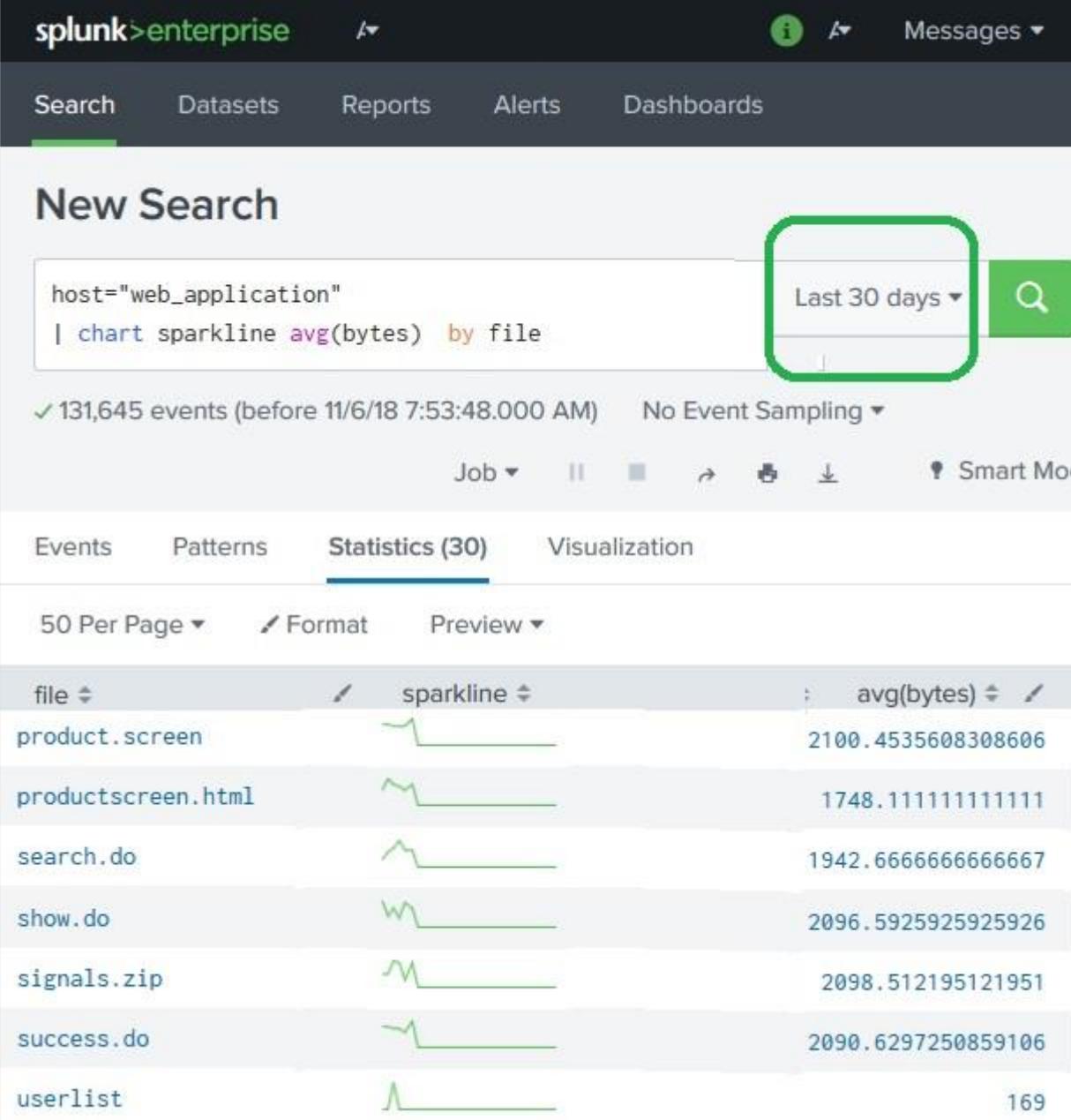
To create the Sparklines from above statistics, we add the Sparkline function to the search query as shown in the image below. The table view of the above statistics now starts displaying the sparklines for average byte size of those files. Here, we have taken **All Time** as the time period for calculating the variation in average byte size of files. If we change this time period, then the nature of the graphs will change.

The screenshot shows the Splunk Enterprise search interface. The search bar contains the query: `host="web_application" | chart sparkline avg(bytes) by file`. The results section displays 131,645 events from before November 6, 2018, at 7:53:48.000 AM, with no event sampling applied. The "Statistics (30)" tab is selected, showing a table with the following data:

file	sparkline	avg(bytes)
product.screen		2095.9946449218037
productscreen.html		2031.5420168067226
search.do		1907.728
show.do		2158.5288888888889
signals.zip		2010.6731707317074
success.do		2097.4306992139013
user		2567
userlist		391.14529914529913
users		2567
usrs		2567

Changing the Time Period

If we change the time period for the above graph from All Time to Last 30 days, we will see the sparklines to be little different as shown below. Here we need to note, how few file names have vanished from the list as those files were not available in that time period.



The screenshot shows the Splunk Enterprise interface with a search bar containing the query: `host="web_application" | chart sparkline avg(bytes) by file`. A green box highlights the time range selector "Last 30 days". Below the search bar, it says "131,645 events (before 11/6/18 7:53:48.000 AM)" and "No Event Sampling". The "Statistics (30)" tab is selected in the results view. The table lists file names with their average bytes values:

file	sparkline	avg(bytes)
product.screen		2100.4535608308606
productscreen.html		1748.111111111111
search.do		1942.6666666666667
show.do		2096.5925925925926
signals.zip		2098.512195121951
success.do		2090.6297250859106
userlist		169

25. Splunk – Managing Indexes

Indexing is a mechanism to speed up the search process by giving numeric addresses to the piece of data being searched. Splunk indexing is similar to the concept of indexing in databases. The installation of Splunk creates three default indexes as follows.

- **main:** This is Splunk's default index where all the processed data is stored.
- **Internal:** This index is where Splunk's internal logs and processing metrics are stored.
- **audit:** This index contains events related to the file system change monitor, auditing, and all user history.

The Splunk Indexers create and maintain the indexes. When you add data to Splunk, the indexer processes it and stores it in a designated index (either, by default, in the main index or in the one that you identify).

Checking Indexes

We can have a look at the existing indexes by going to **Settings -> Indexes** after logging in to Splunk. The below image shows the option.

The screenshot shows the Splunk web interface. At the top, there is a navigation bar with links for Admin..., Messages, Settings (which is highlighted with a green box), Activity, Help, and Find. Below the navigation bar is a sidebar on the left containing icons for Add Data (with a plus sign) and Monitoring Console (with three vertical bars). The main content area is divided into several sections under the Settings menu:

- KNOWLEDGE**: Searches, reports, and alerts; Data models; Event types; Tags; Fields; Lookups; User interface; Alert actions; Advanced search; All configurations.
- DATA**: Data inputs; Forwarding and receiving; **Indexes** (highlighted with a green box); Report acceleration summaries; Source types.
- DISTRIBUTED ENVIRONMENT**: Indexer clustering; Forwarder management; Distributed search.
- SYSTEM**: Server settings; Server controls; Health report manager; Instrumentation; Licensing; Workload management.
- USERS AND AUTHENTICATION**: Access controls.

On further clicking on the indexes, we can see the list of indexes Splunk maintains for the data that is already captured in Splunk. The below image shows such a list.

splunk>enterprise Apps ▾ Admin... ▾ Messages ▾ S

Indexes

A repository for data in Splunk Enterprise. Indexes reside in flat files on the Splunk Enterprise

Name	Actions	Type	App	Current Size	Max Size
_audit	Edit Delete Disable		system	14 MB	488.28 GB
_internal	Edit Delete Disable		system	227 MB	488.28 GB
_introspection	Edit Delete Disable		system	370 MB	488.28 GB
_telemetry	Edit Delete Disable		system	1 MB	488.28 GB
_thefishbuck et	Edit Delete Disable		system	1 MB	488.28 GB
history	Edit Delete Disable		system	1 MB	488.28 GB
main	Edit Delete Disable		system	36 MB	488.28 GB

Creating a New Index

We can create a new index with desired size by the data that is stored in Splunk. The additional data that comes in can use this newly created index but better search functionality. The steps to create an index is **Settings -> Indexes -> New Index**. The

below screen appears where we mention the name of the index and memory allocation etc.

New Index

General Settings

Index Name	INDEX_WEB_APP	
Set index name (e.g., INDEX_NAME). Search using index=INDEX_NAME.		
Index Data Type	<input checked="" type="radio"/> Events	<input type="radio"/> Metrics
The type of data to store (event-based or metrics).		
Home Path	optional	
Hot/warm db path. Leave blank for default (\$SPLUNK_DB/INDEX_NAME/db).		
Cold Path	optional	
Cold db path. Leave blank for default (\$SPLUNK_DB/INDEX_NAME/colddb).		
Thawed Path	optional	
Thawed/resurrected db path. Leave blank for default (\$SPLUNK_DB/INDEX_NAME/thaweddb).		
Data Integrity Check	<input checked="" type="radio"/> Enable	<input type="radio"/> Disable
Enable this if you want Splunk to compute hashes on every slice of your data for the purpose of data integrity.		
Size of Entire Index	100	GB ▾
Maximum target size of entire index.		
Max Size of Warm/Cold Bucket	auto	GB ▾
Maximum target size of buckets. Enter 'auto_high_volume' for high-volume indexes.		
Frozen Path	optional	
<input type="button" value="Cancel"/>		<input type="button" value="Save"/>

Indexing the Events

After creating the index above we can configure the events to be indexed by this specific index. We choose the event type. Use the path **Settings -> Data Inputs -> Files & Directories**. Then we choose the specific file of the events which we want to attach to the newly created event. As you can see in the below image, we have assigned the index named **index_web_app** to this specific file.

splunk>enterprise Apps ▾ i Administrator ▾ Messages ▾ Settings ▾

\$SPLUNK_HOME\var\log\splunk

Data inputs » Files & directories » \$SPLUNK_HOME\var\log\splunk

You can tell Splunk to continuously collect data from a file or directory (keep indexing data as it comes in), or index a static file and then stop.

Host

Tell Splunk how to set the value of the host field in your events from this source.

Set host: constant value

Specify method for getting host field for events coming from this source.

Host field value: localhost

Source type

Set the source type: Automatic

Index

Set the destination index for this source.

Index: idx_web_app

Advanced options

Whitelist:

Specify a regex that files from this source must match to be monitored by Splunk.

Blacklist:

Specify a regex that files from this source must NOT match to be monitored by Splunk

Buttons: Cancel, Save

26. Splunk – Calculated Fields

Many times, we will need to make some calculations on the fields that are already available in the Splunk events. We also want to store the result of these calculations as a new field to be referred later by various searches. This is made possible by using the concept of calculated fields in Splunk search.

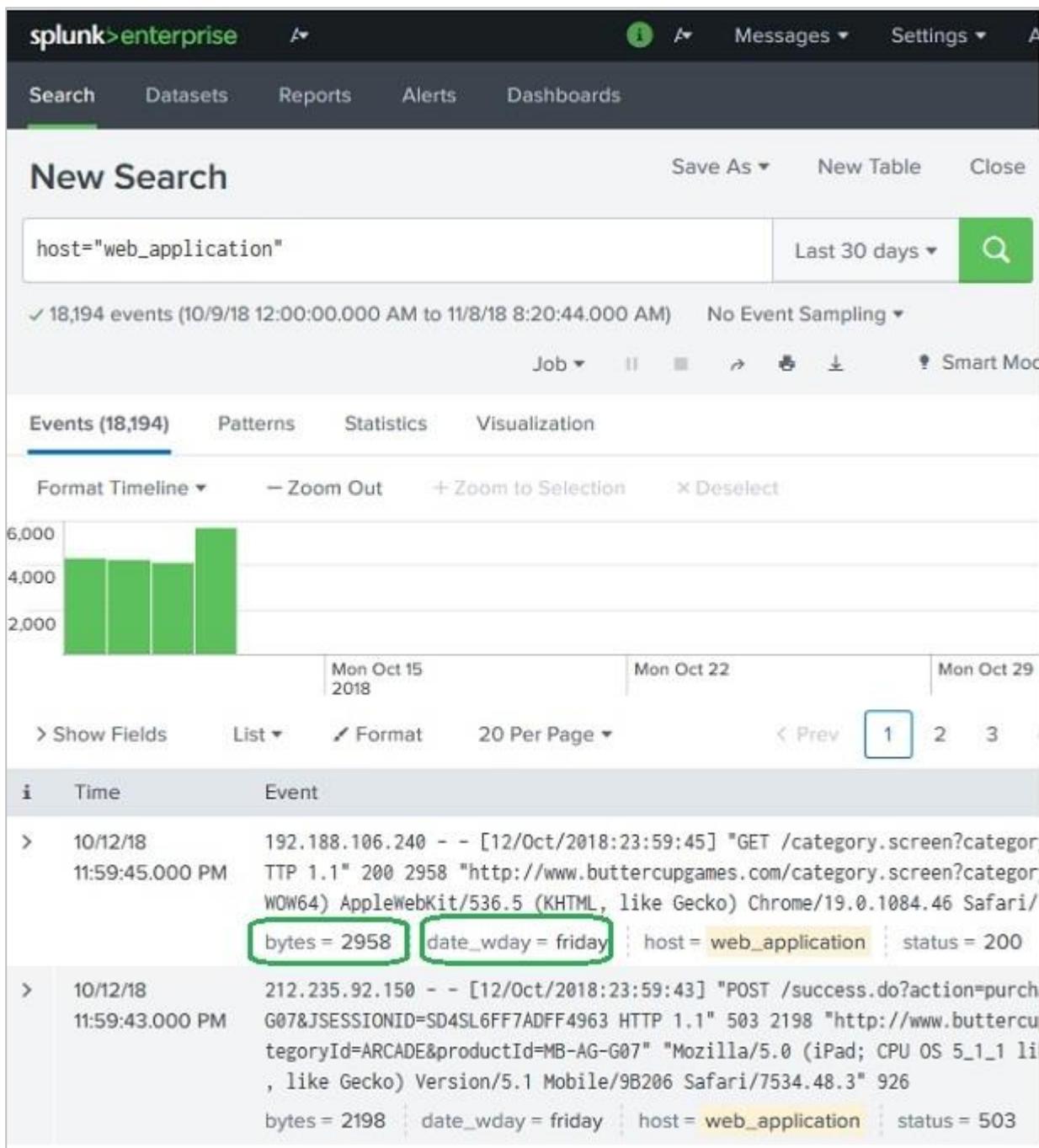
A simplest example is to show the first three characters of a week day instead of the complete day name. We need to apply certain Splunk function to achieve this manipulation of the field and store the new result under a new field name.

Example

The Web_application log file has two fields named bytes and date_wday. The value in the bytes field is the number of bytes. We want to display this value as GB. This will require the field to be divided by 1024 to get the GB value. We need to apply this calculation to the bytes field.

Similarly, the date_wday displays complete name of the week day. But we need to display only the first three characters.

The existing values in these two fields is shown in the image below:



Using the eval Function

To create calculated field, we use the eval function. This function stores the result of the calculation in a new field. We are going to apply the below two calculations:

```
# divide the bytes with 1024 and store it as a field named byte_in_GB
Eval byte_in_GB = (bytes/1024)
```

```
# Extract the first 3 characters of the name of the day.
Eval short_day=substr(date_wday,1,3)
```

Adding New Fields

We add new fields created above to the list of fields we display as part of the search result. To do this, we choose **All fields** options and tick check mark against the name of these new fields as shown in below image:

	Field	# of Values	Event Coverage	Type
> <input checked="" type="checkbox"/>	byte_in	>100	100%	Number
> <input checked="" type="checkbox"/>	host	1	100%	String
> <input checked="" type="checkbox"/>	short_d ay	4	100%	String
> <input checked="" type="checkbox"/>	status	9	100%	Number
> <input type="checkbox"/>	JSESSI ONID	>100	99.81%	String
> <input type="checkbox"/>	action	5	49.66%	String
> <input type="checkbox"/>	bytes	>100	100%	Number
> <input type="checkbox"/>	date_y ear	1	100%	Number

Displaying the calculated Fields

After choosing the fields above, we are able to see the calculated fields in the search result as shown below. The search query displays the calculated fields as shown below:

splunk>enterprise

Search Datasets Reports Alerts Dashboards

New Search

host="web_application" | eval byte_in_GB = (bytes/1024) | eval short_day=substr(date_wday,1,3)

Last 30 days

✓ 18,194 events (10/9/18 12:00:00.000 AM to 11/8/18 8:09:43.000 AM) No Event Sampling

Job ▾ Smart Mode

Events (18,194) Patterns Statistics Visualization

Format Timeline ▾

Mon Oct 15 2018 Mon Oct 22 Mon Oct 29

> Show Fields List ▾ Format 20 Per Page ▾ < Prev 1 2 3

i	Time	Event
>	10/12/18 11:59:45.000 PM	192.188.106.240 - - [12/Oct/2018:23:59:45] "GET /category.screen?category=WOW64" TTP 1.1" 200 2958 "http://www.buttercupgames.com/category.screen?category=WOW64" AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5 byte_in_GB = 2.888671875 host = web_application short_day = fri status = 200
>	10/12/18 11:59:43.000 PM	212.235.92.150 - - [12/Oct/2018:23:59:43] "POST /success.do?action=purchase&productID=ARCADE&JSESSIONID=SD4SL6FF7ADFF4963 HTTP 1.1" 503 2198 "http://www.buttercupgames.com/cart.do?productID=ARCADE&productID=MB-AG-G07" "Mozilla/5.0 (iPad; CPU OS 5_1_1 like Gecko) Version/5.1 Mobile/9B206 Safari/7534.48.3" 926 byte_in_GB = 2.146484375 host = web_application short_day = fri status = 503
>	10/12/18 11:59:41.000 PM	212.235.92.150 - - [12/Oct/2018:23:59:41] "POST /cart.do?action=addtocart&productID=ARCADE&JSESSIONID=SD4SL6FF7ADFF4963 HTTP 1.1" 200 669 "http://www.buttercupgames.com/cart.do?productID=ARCADE&productID=MB-AG-G07" "Mozilla/5.0 (iPad; CPU OS 5_1_1 like Mac OS X) AppleWebKit/534.46 (KHTML, like Gecko) Version/4.0 Mobile/9B206 Safari/7534.48.3" 197 byte_in_GB = 0.6533203125 host = web_application short_day = fri status = 200

27. Splunk – Tags

Tags are used to assign names to specific field and value combinations. These fields can be event type, host, source, or source type, etc. You can also use a tag to group a set of field values together, so that you can search for them with one command. For example, you can tag all the different files generated on Monday to a tag named mon_files.

To find the field-value pair which we are going to tag, we need to expand the events and locate the field to be considered. The below image shows how we can expand an event to see the fields:

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk>enterprise' and icons for help, messages, and settings. Below it is a secondary navigation bar with 'Search', 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. The main area is titled 'New Search' and contains a search bar with the query 'host="web_application" status=503 OR 505'. Below the search bar, it says '✓ 13,921 events (10/10/18 12:00:00.000 AM to 11/9/18 8:46:21.000 AM) No Event Sampling'. There are buttons for 'Job', 'Smart Mo', and other search controls. A histogram is displayed, showing event counts over time from Mon Oct 15 to Mon Oct 25. The x-axis is labeled with dates: Mon Oct 15 2018, Mon Oct 22, and Mon Oct 25. The y-axis ranges from 0 to 6,000 with increments of 2,000. The histogram has three bars, with the third bar reaching approximately 6,000. Below the histogram, there are buttons for 'Format Timeline', 'Zoom Out', 'Zoom to Selection', and 'Deselect'. The 'Events (13,921)' tab is selected, showing a table of event details. The table has columns for 'Time' and 'Event'. The first event is highlighted with a green border. The event details are as follows:

Time	Event
10/12/18 11:59:45.000 PM	192.188.106.240 - - [12/Oct/2018:23:59:45] "GET /category.screen?ca TTP 1.1" 200 2958 "http://www.buttercupgames.com/category.screen?ca WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Sa bytes = 2958 date_wday = friday host = web_application status = : 200
10/12/18 11:59:43.000 PM	212.235.92.150 - - [12/Oct/2018:23:59:43] "POST /success.do?action= G07&JSESSIONID=SD4SL6FF7ADFF4963 HTTP 1.1" 503 2198 "http://www.but tategoryId=ARCADE&productId=MB-AG-G07" "Mozilla/5.0 (iPad; CPU OS 5_1. , like Gecko) Version/5.1 Mobile/9B206 Safari/7534.48.3" 926 bytes = 2198 date_wday = friday host = web_application status = E 200

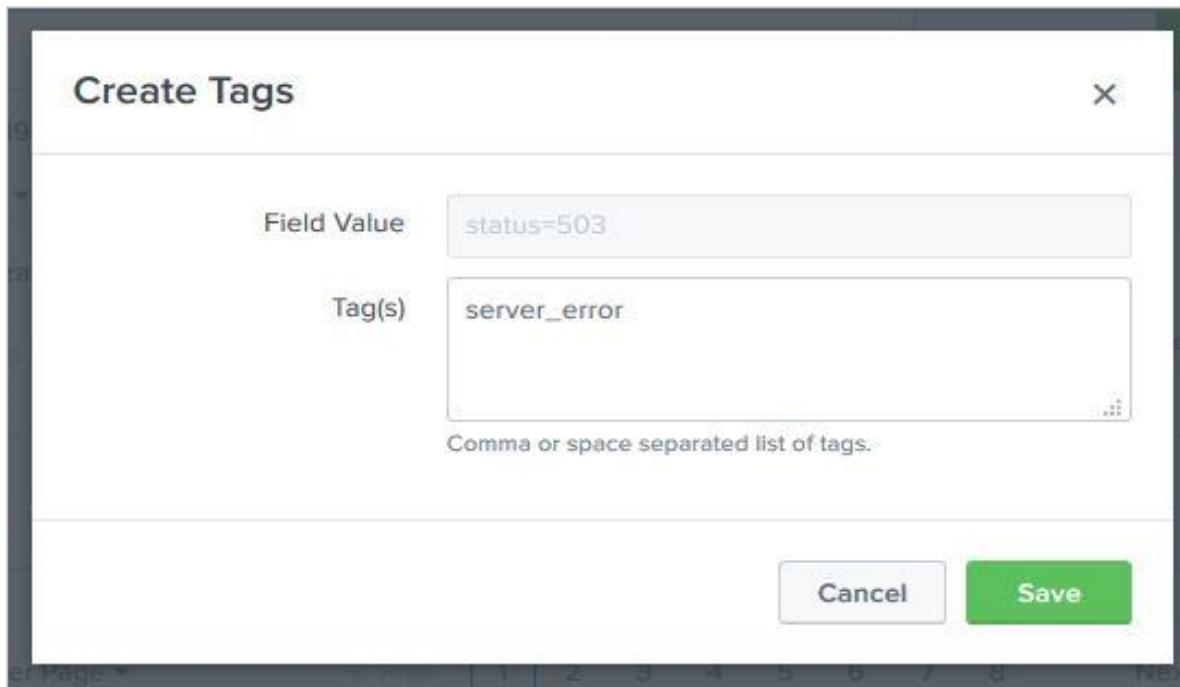
Creating Tags

We can create tags by adding the tag value to field-value pair using **Edit Tags** option as shown below. We choose the field under the Actions column.

The screenshot shows the Splunk Enterprise search interface. At the top, the navigation bar includes 'splunk>enterprise', 'Search' (selected), 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. Below the search bar, the query 'host="web_application" status=503 OR 505' is entered. The results section shows '13,921 events (10/10/18 12:00:00.000 AM to 11/9/18 8:46:21.000 AM)' with 'No Event Sampling'. The 'Events (13,921)' tab is active, displaying a histogram from Mon Oct 15 to Mon Oct 29. Below the histogram, event details for 10/12/18 11:59:45.000 PM are shown, including source IP 192.188.106.240, method GET, URL /category.screen?category, status 200, and user agent AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5. An 'Event Actions' dropdown menu is open, showing fields: bytes (value 2958), date_wday (value friday), host (value web_application, highlighted with a yellow box), and status (value 503). A 'Edit Tags' button is located next to the host field. The bottom right corner of the event table area is also highlighted with a green box.

The next screen prompts us to define the tag. For the Status field, we choose the status value of 503 or 505 and assign a tag named server_error as shown below. We have to do

it one by one by choosing two events, each with the events with status value 503 and 505. The image below shows the method for status value as 503. We have to repeat the same steps for an event with status value as 505.



Search Using Tags

Once the tags are created, we can search for events containing the Tag by simply writing the Tag name in the search bar. In the below image, we see all the events which have status: 503 or 505.

splunk>enterprise

Search Datasets Reports Alerts Dashboards

New Search

Save As ▾ New Table Close

tag::status="server_error"

Last 30 days

✓ 417 events (10/10/18 12:00:00.000 AM to 11/9/18 10:13:01.000 AM) No Event Sampling ▾

Job ▾ Smart Mo

Events (417) Patterns Statistics Visualization

Format Timeline ▾

Show Fields List ▾ Format 20 Per Page ▾ 1 2

i	Time	Event
>	10/12/18 11:59:43.000 PM	212.235.92.150 - - [12/Oct/2018:23:59:43] "POST /success.do?action=G07&JSESSIONID=SD4SL6FF7ADFF4963 HTTP 1.1" 503 2198 "http://www.buttercupgames.com/categoryId=ARCADE&productId=MB-AG-G07" "Mozilla/5.0 (iPad; CPU OS 5_1, like Gecko) Version/5.1 Mobile/9B206 Safari/7534.48.3" 926 host = web_application status = 503 server_error
>	10/12/18 11:48:44.000 PM	27.102.11.11 - - [12/Oct/2018:23:48:44] "GET /product.screen?prod 7 HTTP 1.1" 503 1068 "http://www.buttercupgames.com/category.scre MSIE 7.0; Windows NT 5.1; Trident/4.0; .NET CLR 2.0.50727; MS-RTC L host = web_application status = 503 server_error
>	10/12/18 11:16:54.000 PM	95.130.170.231 - - [12/Oct/2018:23:16:54] "GET /category.screen?cat HTTP 1.1" 505 3831 "http://www.buttercupgames.com/oldlink" "Mozilla Trident/4.0; .NET CLR 2.0.50727; MS-RTC LM 8; InfoPath.2)" 607 host = web_application status = 505 server_error

28. Splunk – Apps

A Splunk app is an extension of Splunk functionality which has its own in-built UI context to serve a specific need. Splunk apps are made up of different Splunk knowledge objects (lookups, tags, eventtypes, savedsearches, etc). Apps themselves can utilize or leverage other apps or add-ons. Splunk can run any number of apps simultaneously.

When you log in to Splunk, you land on an app which is typically, the **Splunk Search app**. So, almost everytime you are inside the Splunk interface, you are using an app.

Listing Splunk Apps

We can list the available apps in Splunk by using the option **Apps -> Manage Apps**. Navigating this option brings out the following screen which lists the existing apps available in Splunk interface.

Name	Folder name	Version	Visible	Sharing	Status
SplunkForwarder	SplunkForwarder		No	App Permissions	Disabled Enable
SplunkLightForwarder	SplunkLightForwarder		No	App Permissions	Disabled Enable
Log Event Alert Action	alert_logevent	7.2.0	No	App Permissions	Enabled Disable
Webhook Alert Action	alert_webhook	7.2.0	No	App Permissions	Enabled Disable
Apps Browser	appsbrowser	7.2.0	No	App Permissions	Enabled
framework	framework		No	App Permissions	Enabled Disable
Getting started	gettingstarted	1.0	Yes	App Permissions	Disabled Enable

Following are important values associated with the Splunk apps:

- **Name:** It is the name of the App and unique for each App.
- **Folder Name:** It is the name to use for the directory in \$SPLUNK_HOME/etc/apps/. The name of the folder cannot contain "dot" (.) character.
- **Version:** It is the app version string. Visible Indicates whether the app should be visible in Splunk Web. Apps that contain a user interface should be visible.

- **Sharing:** It is the level of permissions (read or write) given to different Splunk users for that specific app.
- **Status:** It is the current status of availability of the App. It may be enabled or disabled for use.

App Permissions

A proper setting of permissions for using the app is important. We can restrict the app to be used by a single user or by multiple users including all users. The below screen which appears after clicking on the permissions link in the above is used to modify the access to different roles.

The screenshot shows the Splunk Enterprise interface with the following details:

- Header:** splunk>enterprise, Apps ▾, Administrator ▾, Messages ▾, and a search icon.
- Page Title:** Permissions
- Breadcrumb:** Apps » alert_logevent » Permissions
- Section:** App permissions
- Description:** Users with read access can only save objects for themselves, and require write access to be able to share objects with other users.
- Table:** A grid showing permissions for various roles. The columns are Roles, Read, and Write. The rows include Everyone, admin, can_delete, power, splunk-system-role, and user. The 'Everyone' row has both Read and Write checked. Other rows have both Read and Write unchecked.

Roles	Read	Write
Everyone	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
admin	<input type="checkbox"/>	<input type="checkbox"/>
can_delete	<input type="checkbox"/>	<input type="checkbox"/>
power	<input type="checkbox"/>	<input type="checkbox"/>
splunk-system-role	<input type="checkbox"/>	<input type="checkbox"/>
user	<input type="checkbox"/>	<input type="checkbox"/>

- Section:** Sharing for config file-only objects
- Description:** Set permissions for configurations that have been copied over or added to config files rather than created through the UI.
- Text:** Objects defined in config files only (not in the UI) should appear in
- Radio Buttons:** This app only (system) (selected) and All apps
- Buttons:** Cancel and Save

By default, the check marks for Read and Write option is available for Everyone. But we can change that by going to each role and selecting appropriate permission for that specific role.

App Marketplace

There is a wide variety of needs for which the Splunk search functionalities are used. So, there is a Splunk App market place which has come into existence showing many different apps created by individual and organizations. They are available in both free and paid versions. We can browse those apps by choosing the option **Apps -> Manage Apps -> Browse More Apps**. The below screen comes up.

The screenshot shows the Splunk App Marketplace interface. At the top, there's a navigation bar with the Splunk logo, a dropdown menu, and links for 'Messages', 'Settings', and 'Find'. Below the header, the title 'Browse More Apps' is displayed. On the left side, there are several filter categories with checkboxes:

- CATEGORY:** DevOps, Security, Fraud & Compliance, IT Operations, Utilities, Business Analytics, IoT & Industrial Data.
- CIM VERSION:** 4.x, 3.x.
- SUPPORT TYPE:** Community, Developer, Splunk, Unsupported, Python.
- APP CONTENT:** Inputs, Alert Actions.

In the center, there are two app cards:

- Website Monitoring**: Category: IT Operations | Author: Luke Murphey | Do | Last Updated: 13 minutes ago | View on Splunkbase. Includes a green 'Install' button.
- NLP Text Analytics**: Category: Utilities, Business Analytics | Author: Nathaniel | Last Updated: 5 months ago | Released: 5 months ago | View on Splunkbase. Includes a green 'Install' button.

As you can see, the App name along with a brief description of the functionality of the App appears. This helps you decide which app to use. Also, note how the Apps are categorized in the left bar to help choose the type of App faster.

29. Splunk – Removing Data

Removing data from Splunk is possible by using the **delete** command. We first create the search condition to fetch the events we want to mark for delete. Once the search condition is acceptable, we add the delete clause at the end of the command to remove those events from Splunk. After deletion, not even a user with admin privilege is able to view this data in Splunk.

Removal of data is irreversible. If you still want the removed data back into Splunk then you should have the original source data copy with you which can be used to re-index the data in Splunk. It will be a process similar to creating a new index.

Assigning Delete Privilege

Any user including admin user does not have access to delete the data by default. By default, only the "**can_delete**" role has the ability to delete events. So, we create a new user, assign this role and then login with the credentials of this new user to perform the delete operation. The below image shows how we create a new user with "can_delete" role. We arrive at this screen by following the path **Settings -> Access Controls -> Users -> New User**.

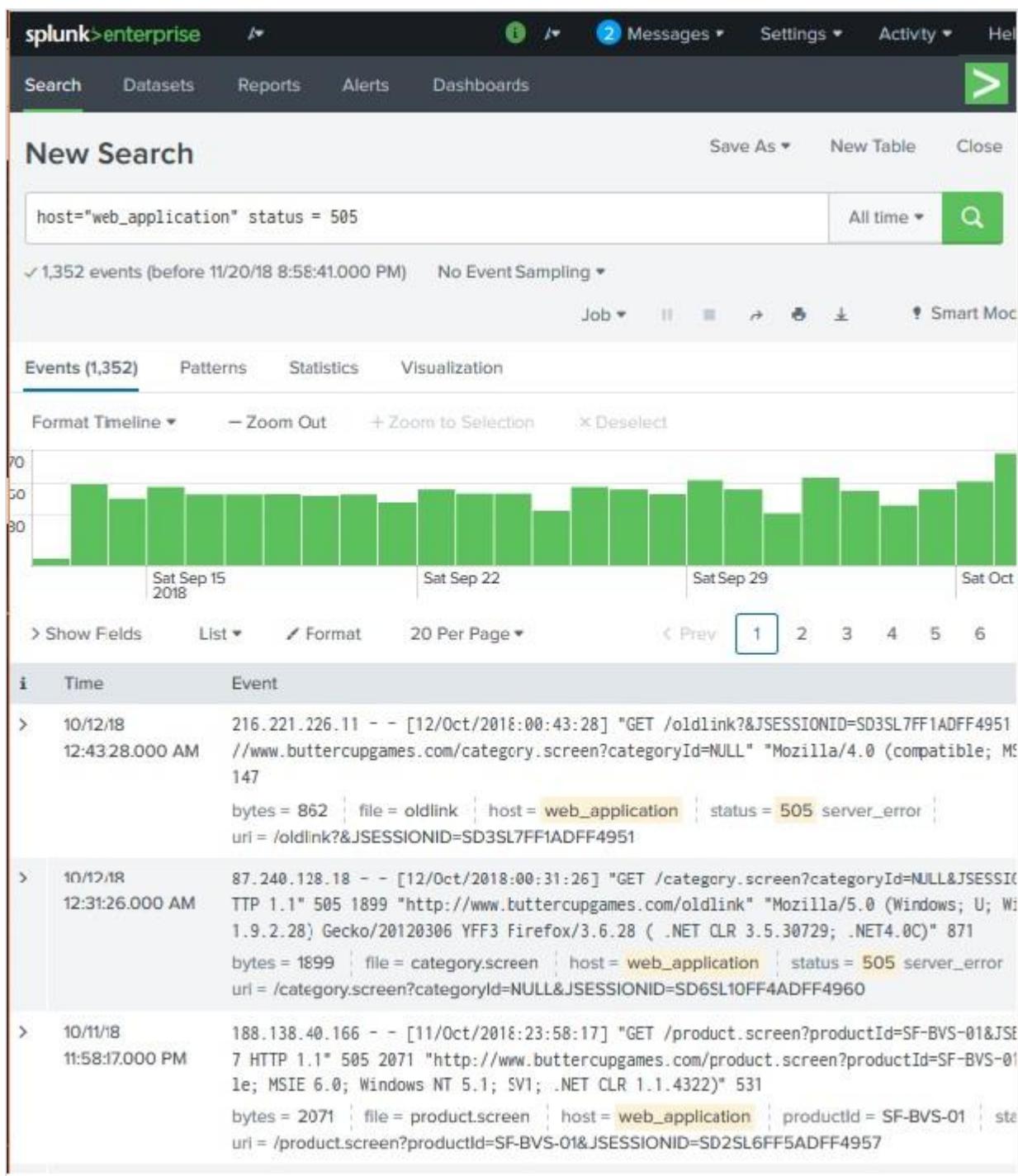
Create User

Name	del_usr	
Full name	optional	
Email address	optional	
Set password	*****	
Confirm password	*****	
Password must contain at least ? <small>✓ 8 characters</small>		
Time zone ?	– Default System Timezone – ▾	
Default app ?	launcher (Home) ▾	
Assign to roles ?	Available item(s) add all > admin can_delete power splunk-system-role <small>USER</small>	Selected item(s) « remove all can_delete user
Create a role for this user	<input type="checkbox"/>	
Require password change on first login	<input checked="" type="checkbox"/>	
Cancel Save		

We then log out of Splunk interface and login back with this newly created user.

Identifying the data to be removed

First, we need to identify the list of events we want to remove. It is done using a normal search query specifying the filter condition. In the below example, we choose to look for the events from the host web_application which has the field http status value as **505**. Our goal is to delete only the set of data containing these values to be removed from the search result. The below image shows this set of data selected.



Deleting the Selected Data

Next, we use the delete command to remove the above selected data from the result set. It involves just adding the word `delete` after '`|`' at the end of the search query as shown below:

The screenshot shows the Splunk Enterprise interface with the following details:

- Search Bar:** host="web_application" status = 505 | delete
- Event Count:** ✓ 1,352 events (before 11/20/18 8:58:41.000 PM) No Event Sampling
- Timeline:** Format Timeline, Zoom Out, Zoom to Selection, Deselect
- Event List:**

i	Time	Event
>	10/12/18 12:43:28.000 AM	216.221.226.11 - - [12/Oct/2018:00:43:28] "GET /oldlink?&JSESSIONID=SD3SL7FF1ADFF495 //www.buttercupgames.com/category.screen?categoryId=NULL" "Mozilla/4.0 (compatible; I 147 bytes = 862 file = oldlink host = web_application status = 505 server_error uri = /oldlink?&JSESSIONID=SD3SL7FF1ADFF4951
>	10/12/18 12:31:26.000 AM	87.240.128.18 - - [12/Oct/2018:00:31:26] "GET /category.screen?categoryId=NULL&JESS TTP 1.1" 505 1899 "http://www.buttercupgames.com/oldlink" "Mozilla/5.0 (Windows; U; 1.9.2.28) Gecko/20120306 YFF3 Firefox/3.6.28 (.NET CLR 3.5.30729; .NET4.0C)" 871 bytes = 1899 file = category.screen host = web_application status = 505 server_error uri = /category.screen?categoryId=NULL&JSESSIONID=SD6SL10FF4ADFF4960
>	10/11/18 11:58:17.000 PM	188.138.40.166 - - [11/Oct/2018:23:58:17] "GET /product.screen?productId=SF-BVS-01&J 7 HTTP 1.1" 505 2071 "http://www.buttercupgames.com/product.screen?productId=SF-BVS- 1e; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322" 531 bytes = 2071 file = product.screen host = web_application productId = SF-BVS-01 s uri = /product.screen?productId=SF-BVS-01&JSESSIONID=SD2SL6FF5ADFF4957

After running the search query above, we can see the next screen where those events have got deleted.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk enterprise' logo, user info (2 messages), and links for Settings, Activity, and Help. Below the bar, a menu bar has 'Search' selected, followed by Datasets, Reports, Alerts, and Dashboards.

The main area is titled 'New Search'. A search bar contains the query: 'host="web_application" status = 505 | delete'. To the right of the search bar are 'Save As', 'New Table', and 'Close' buttons. A time range selector shows 'All time' and a green search button.

Below the search bar, a message indicates '✓ 1,352 events (before 11/21/18 6:12:31.000 AM) No Event Sampling'. There are several filtering and sorting options: 'Job' dropdown, a toolbar with icons for Job, Run, Stop, Refresh, and Save, and a 'Smart Mode' link.

The results section has tabs for 'Events', 'Patterns', 'Statistics (2)', and 'Visualization'. The 'Statistics' tab is active. It shows a table with two rows:

splunk_server	Index	deleted	errors
DESKTOP-JKQCPPLP	_ALL	1352	0
DESKTOP-JKQCPPLP	main	1352	0

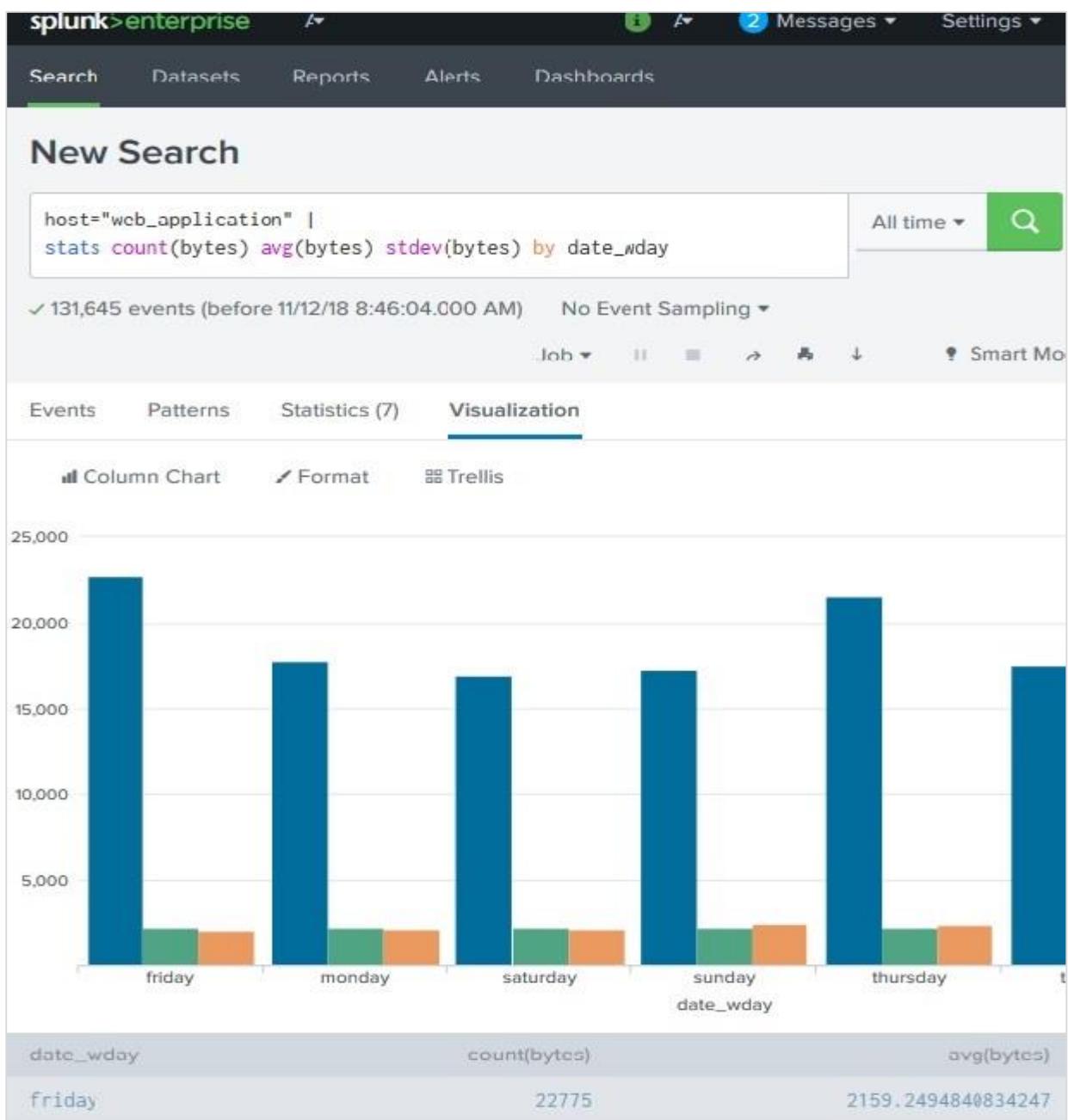
A green box highlights the 'deleted' column header in the statistics table.

You can also further run the search query to verify that these events are not returned in the result set.

30. Splunk – Custom Chart

The charts created in Splunk has many features to customize them as per the user need. These customizations help in displaying the data completely or changing the interval for which the data is calculated. After initially creating the chart, we dive into the customization features.

Let us consider the below search query for getting the statistics of various measurements of byte size of the files by week day. We choose a column chart to display the graph and see the default values in the X-axis and Y-Axis values.



Axis Customization

We can customize the axes displayed in the chart by choosing the **Format -> X-axis** button. Here, we edit the Title of the chart. We also edit the Label Rotation option to choose an inclined label to fit better into the chart. After editing these, results can be seen in the chart as highlighted using the green boxes below.

The screenshot shows the Splunk interface for a new search. The search query is:

```
host="web_application" | stats count(bytes) avg(bytes) stdev(bytes) by date_wday
```

The search results show 131,645 events from before 11/12/18 8:46:04.000 AM, with no event sampling.

The **Visualization** tab is selected. A context menu is open over the chart, with the **X-Axis** option highlighted. The menu options include:

- General
- X-Axis**
- Y-Axis
- Chart Overlay
- Legend

The **X-Axis** menu has the following settings:

- Title: Bytes by Week Day (highlighted with a green box)
- Label Rotation: abc (highlighted with a green box)
- Label Truncation: Yes (highlighted with a green box)

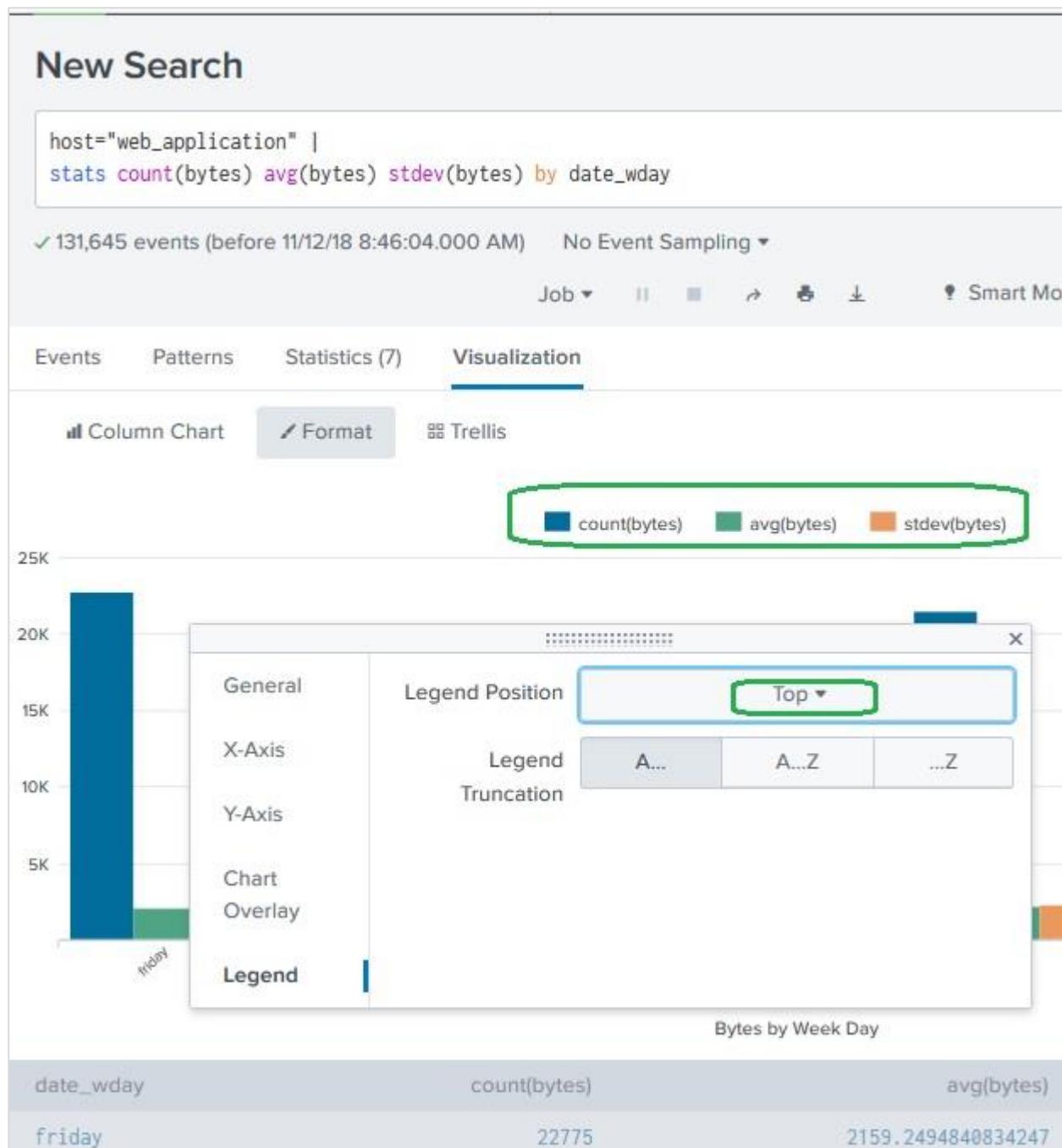
The chart displays data for the week days:

date_wday	count(bytes)	avg(bytes)
friday	22775	2159.2494840834247
saturday		
sunday		
monday		
tuesday		
Wednesday		
Thursday		

A legend at the bottom is labeled "Bytes by Week Day". A green box highlights the "Hidey" label on the x-axis.

Legend Customization

The legends of the chart can also be customized by using the option **Format -> Legend**. We edit the option Legend Position to mark it at Top. We also edit the Legend Truncation option to Truncate the End of the legend if required. The below cart shows the legends displayed at the top with colors and values.



31. Splunk – Monitor Files

Splunk Enterprise monitors and indexes the file or directory as new data appears. You can also specify a mounted or shared directory, including network file systems, as long as Splunk Enterprise can read from the directory. If the specified directory contains subdirectories, the monitor process recursively examines them for new files, as long as the directories can be read.

You can include or exclude files or directories from being read by using whitelists and blacklists.

If you disable or delete a monitor input, Splunk Enterprise does not stop indexing the files: input references. It only stops checking those files again.

You specify the path to a file or directory and the monitor processor consumes any new data written to that file or directory. This is how you can monitor live application logs such as those coming from Web access logs, Java 2 Platform or .NET applications, and so on.

Add files to Monitor

Using Splunk web interface, we can add files or directories to be monitored. We go to **Splunk Home -> Add Data -> Monitor** as shown in the below image:

splunk>enterprise    Messages 

What data do you want to send to the Splunk platform?

Follow guides for onboarding popular data sources

 **Networking**
Get your networking data in to the Splunk platform.
2 data sources

 **Operating System**
Get your operating system data in to the Splunk platform.
1 data source

 **Security**
Get your security data in to the Splunk platform.
3 data sources

3 data sources in total

Or get data in with the following methods


Upload
files from my computer
Local log files
Local structured files (e.g. CSV)
[Tutorial for adding data](#)


Monitor
files and ports on this Splunk platform instance
Files - HTTP - WMI - TCP/UDP - Scripts
Modular inputs for external data sources

On clicking Monitor, it brings up the list of types of files and directory you can use to monitor the files. Next, we choose the file we want to monitor.

splunk>enterprise A i 2 Messages ▾ Settings

Add Data ● Select Source Set Source Type Input Settings Review Done < Back Next >

Local Event Logs
Collect event logs from this machine.

Remote Event Logs
Collect event logs from remote hosts. Note: this uses WMI and requires a domain account.

Files & Directories
Upload a file, index a local file, or monitor an entire directory.

HTTP Event Collector
Configure tokens that clients can use to send data over HTTP or HTTPS.

TCP / UDP
Configure the Splunk platform to listen on a network port.

Configure this instance to monitor files and directories for data. To monitor all objects in a directory, select the directory. The Splunk platform monitors and assigns a single source type to all objects within the directory. This might cause problems if there are different object types or data sources in the directory. To assign multiple source types to objects in the same directory, configure individual data inputs for those objects. [Learn More](#)

File or Directory ? Browse

On Windows: c:\apache\apache.error.log or \\hostname\apache\apache.error.log. On Unix: /var/log or /mnt/www01/var/log.

Continuously Monitor Index Once

Whitelist ?

Blacklist ?

Next, we choose the default values as Splunk is able to parse the file and configure the options for monitoring automatically.

After the final step, we see the below result which captures the events from the file to be monitored.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk>enterprise' on the left, followed by icons for help, messages (2), and settings. Below the bar are tabs for 'Search', 'Datasets', 'Reports', 'Alerts', and 'Dashboards'. The main area is titled 'New Search'.

In the search bar, the query is:

```
source="C:\\\\Users\\\\Documents\\\\TP_AWS\\\\splunk\\\\prodctidvals.csv"
host="DESKTOP-JKQCPLP" sourcetype="csv"
```

Below the search bar, it says '18 events (before 11/13/18 10:13:11.000 AM)' and 'No Event Sampling'. There are various search controls like 'All time', a green search button, and a 'Smart Mode' icon.

The search results are displayed in a table with columns: 'Time', 'Event', and some additional fields. The table has 18 rows. The first few rows show:

i	Time	Event
>	10/30/18 12:27:23.000 PM	SF-BVS-G01,Hard Drive host = DESKTOP-JKQCPLP productdescription = Hard Drive
>	10/30/18 12:27:23.000 PM	GT-SC-G01,Battery host = DESKTOP-JKQCPLP productdescription = Battery
>	10/30/18 12:27:23.000 PM	WSC-MG-G10,Usb Light host = DESKTOP-JKQCPLP productdescription = Usb Light
>	10/30/18 12:27:23.000 PM	CU-PG-G06,EBook Reader host = DESKTOP-JKQCPLP productdescription = EBook Reader

If any of the value in the event changes, then the above result gets updated to show the latest result.

32. Splunk – Sort Command

The **sort** command sorts all the results by specified fields. The missing fields are treated as having the smallest or largest possible value of that field if the order is descending or ascending, respectively. If the first argument to the sort command is a number, then at most that many results are returned, in order. If no number is specified, the default limit of 10000 is used. If the number 0 is specified, all of the results are returned.

Sorting by Field Types

We can assign specific data type for the fields being searched. The existing data type in the Splunk dataset may be different than the data type we enforce in the search query. In the below example, we sort the status field as numeric in ascending order. Also, the field named url is searched as a string and the negative sign indicates descending order of sorting.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk>enterprise' logo, 'Search', 'Datasets', 'Reports', 'Alerts', 'Dashboards', and 'Messages' (with 2 notifications). Below the navigation is a search bar containing the query: `host="web_application" | sort num(status), -str(url)`. A note indicates `✓ 10,000 events (before 11/13/18 8:10:25.000 PM)` and 'No Event Sampling'. The main area has tabs for 'Events (10,000)' (selected), 'Patterns', 'Statistics', and 'Visualization'. The 'Visualization' tab shows a bar chart of event counts over time, with a peak around 12:00 AM on Thursday, Oct 11, 2018. Below the chart is a table of log events, each with a timestamp, event ID, and detailed log message. The table includes columns for 'Time' and 'Event'. The first few events are:

i	Time	Event
>	10/12/18 11:59:45.000 PM	192.188.106.240 - - [12/Oct/2018:23:59:45] "GET /category.screen?categoryId=TEE&JSE TTP 1.1" 200 2958 "http://www.buttercupgames.com/category.screen?categoryId=TEE" "M WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 602 bytes = 2958 file = category.screen host = web_application status = 200 uri = /category.screen?categoryId=TEE&JSESSIONID=SD2SL4FF9ADFF4959
>	10/12/18 11:59:41.000 PM	212.235.92.150 - - [12/Oct/2018:23:59:41] "POST /cart.do?action=addtocart&productId 6FF7ADFF4963 HTTP 1.1" 200 669 "http://www.buttercupgames.com/product.screen?produc (iPad; CPU OS 5_1_1 like Mac OS X) AppleWebKit/534.46 (KHTML, like Gecko) Version/5 .48.3" 197 bytes = 669 file = cart.do host = web_application productId = MB-AG-G07 status : uri = /cart.do?action=addtocart&productId=MB-AG-G07&JSESSIONID=SD4SL6FF7AD...
>	10/12/18 11:59:39.000 PM	212.235.92.150 - - [12/Oct/2018:23:59:39] "GET /product.screen?productId=MB-AG-G07& 3 HTTP 1.1" 200 2223 "http://www.buttercupgames.com/category.screen?categoryId=ARCA OS 5_1_1 like Mac OS X) AppleWebKit/534.46 (KHTML, like Gecko) Version/5.1 Mobile/9 bytes = 2223 file = product.screen host = web_application productId = MB-AG-G07 uri = /product.screen?productId=MB-AG-G07&JSESSIONID=SD4SL6FF7ADFF4963

Sorting up to a Limit

We can also specify the number of results that will be sorted instead of the entire search result. The below search result shows the sorting of only 50 events with **status** as ascending and **url** as descending.

splunk enterprise

Messages 2 Settings Activity

Search Datasets Reports Alerts Dashboards

New Search

Save As

host="web_application" | sort 50 num(status), -str(url)

All time

✓ 50 events (before 11/13/18 8:37:59.000 PM) No Event Sampling

Job Smart Mc

Events (50) Patterns Statistics Visualization

Format Timeline

12
8
4
11:46 PM Fri Oct 12 2018 11:48 PM 11:50 PM 11:52 PM 11:54 PM 11:56 PM

> Show Fields List 20 Per Page

i	Time	Event
>	10/12/18 11:59:45.000 PM	192.188.106.240 - - [12/Oct/2018:23:59:45] "GET /category.screen?categoryId=TEE&JSE TTP 1.1" 200 2958 "http://www.buttercupgames.com/category.screen?categoryId=TEE" "M WOW64" AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 602 bytes = 2958 file = category.screen host = web_application status = 200 uri = /category.screen?categoryId=TEE&JSESSIONID=SD2SL4FF9ADFF4959
>	10/12/18 11:59:41.000 PM	212.235.92.150 - - [12/Oct/2018:23:59:41] "POST /cart.do?action=addtocart&productId 6FF7ADFF4963 HTTP 1.1" 200 669 "http://www.buttercupgames.com/product.screen?produ (iPad; CPU OS 5_1_1 like Mac OS X) AppleWebKit/534.46 (KHTML, like Gecko) Version/5. .48.3" 197 bytes = 669 file = cart.do host = web_application productId = MB-AG-G07 status = uri = /cart.do?action=addtocart&productId=MB-AG-G07&JSESSIONID=SD4SL6FF7AD...
>	10/12/18 11:59:39.000 PM	212.235.92.150 - - [12/Oct/2018:23:59:39] "GET /product.screen?productId=MB-AG-G07& 3 HTTP 1.1" 200 2223 "http://www.buttercupgames.com/category.screen?categoryId=ARCA OS 5_1_1 like Mac OS X) AppleWebKit/534.46 (KHTML, like Gecko) Version/5.1 Mobile/9 bytes = 2223 file = product.screen host = web_application productId = MB-AG-G07 uri = /product.screen?productId=MB-AG-G07&JSESSIONID=SD4SL6FF7ADFF4963
>	10/12/18 11:59:27.000 PM	192.188.106.240 - - [12/Oct/2018:23:59:27] "GET /oldlink?&JSESSIONID=SD2SL4FF9ADFF4 p://www.buttercupgames.com/oldlink" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/5 34.46 (KHTML, like Gecko) Version/5.1 Mobile/9 bytes = 1911 file = oldlink host = web_application status = 200 uri = /oldlink?&JSESSIONID=SD2SL4FF9ADFF4959

Using Reverse

We can toggle the result of an entire search query by using the reverse clause. It is useful to use the existing query without altering and reversing the sort result as and when needed.

The screenshot shows the Splunk Enterprise web interface. At the top, there's a navigation bar with 'splunk>enterprise' logo, user status (2 messages), Settings, and Activity. Below it is a secondary navigation bar with links for Search, Datasets, Reports, Alerts, and Dashboards. The main area is titled 'New Search' with a 'Save As' button. The search bar contains the query: 'host="web_application" | sort 50 num(status), -str(url) | reverse'. The search results indicate 50 events found before 11/13/18 8:44:51.000 PM, with 'No Event Sampling' applied. Below the search bar are buttons for Job, Job History, and Smart Macros. The main content area has tabs for Events (50), Patterns, Statistics, and Visualization. The Events tab is selected, showing a timeline from 11:46 PM on Friday, October 12, 2018, to 11:56 PM. The timeline bars are green, representing event counts. Below the timeline is a table of event details. The table has columns for i (Info icon), Time, and Event. The first few rows of the table are:

i	Time	Event
>	10/12/18 11:46:17.000 PM	203.223.0.20 - - [12/Oct/2018:23:46:17] "POST /product.screen?productId=DB-SG-G01&HTTP/1.1" 200 900 "http://www.buttercupgames.com/product.screen?productId=DB-SG-G01MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0; BOIE9;ENUS" 675 bytes = 900 file = product.screen host = web_application productId = DB-SG-G01 uri = /product.screen?productId=DB-SG-G01&JSESSIONID=SD5SL2FF7ADFF4951
>	10/12/18 11:46:35.000 PM	203.223.0.20 - - [12/Oct/2018:23:46:35] "GET /oldlink?&JSESSIONID=SD5SL2FF7ADFF4951 /www.buttercupgames.com/category.screen?categoryId=TEE" "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0; BOIE9;ENUS)" 884 bytes = 3164 file = oldlink host = web_application status = 200 uri = /oldlink?&JSESSIONID=SD5SL2FF7ADFF4951
>	10/12/18 11:46:39.000 PM	203.223.0.20 - - [12/Oct/2018:23:46:39] "GET /category.screen?categoryId=STRATEGY&HTTP/1.1" 200 2369 "http://www.buttercupgames.com/product.screen?productId=PZ-SG-G01; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0; BOIE9;ENUS" 201 bytes = 2369 file = category.screen host = web_application productId = PZ-SG-G05 uri = /category.screen?categoryId=STRATEGY&JSESSIONID=SD5SL2FF7ADFF4951
>	10/12/18 11:48:35.000 PM	27.102.11.11 - - [12/Oct/2018:23:48:35] "GET /product.screen?productId=FI-AG-G08&HTTP/1.1" 200 966 "http://www.google.com" "Mozilla/4.0 (compatible; MSIE 7.0; Windows CLR 2.0.50727; MS-RTC LM 8; InfoPath.2)" 995 bytes = 966 file = product.screen host = web_application productId = FI-AG-G08 uri = /product.screen?productId=FI-AG-G08&JSESSIONID=SD0SL2FF8ADFF89567

33. Splunk – Top Command

Many times, we are interested in finding the most common values available in a field. The **top** command in Splunk helps us achieve this. It further helps in finding the count and percentage of the frequency the values occur in the events.

Top Values for a Field

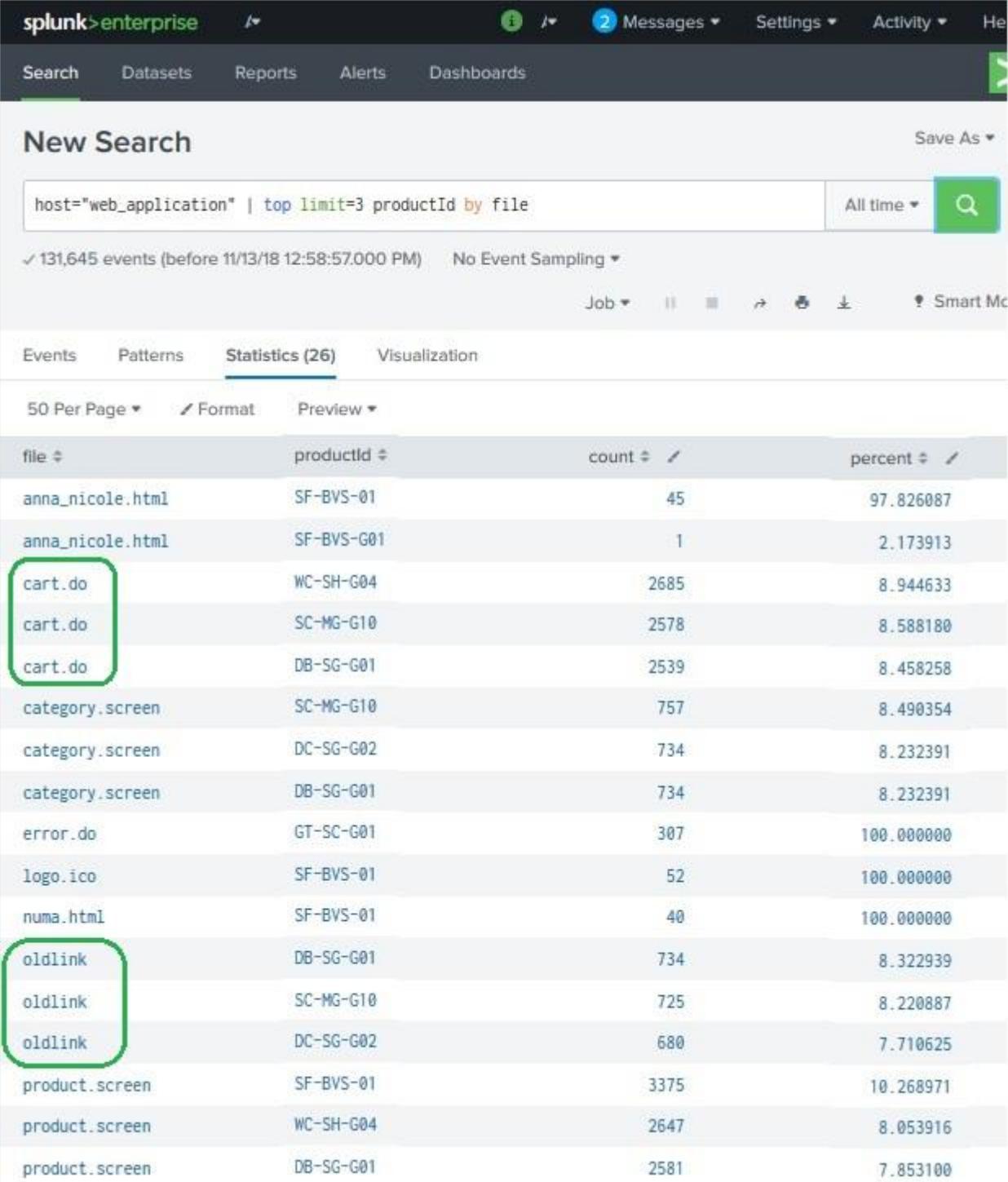
In its simplest form, we just get the count and the percentage of such count as compared to the total number of events. In the below example, we find 8 top most productid values.

The screenshot shows the Splunk Enterprise search interface. The search bar contains the query: `host="web_application" | top limit=8 productId`. The results table displays 8 rows of data, each representing a product ID and its count and percentage:

productId	count	percent
WC-SH-G04	8112	8.253968
DB-SG-G01	7977	8.116606
DC-SG-G02	7703	7.837810
SC-MG-G10	7425	7.554945
WC-SH-A02	7152	7.277167
MB-AG-T01	6977	7.099105
MB-AG-G07	6915	7.036020
FS-SG-G03	6595	6.710419

Top Values for a Field by a Field

Next, we can also include another field as part of this top command's by clause to display the result of field1 for each set of field2. In the below search, we find top 3 productids for each file name. Note how the file names are repeated 3 times showing different productid for that file.



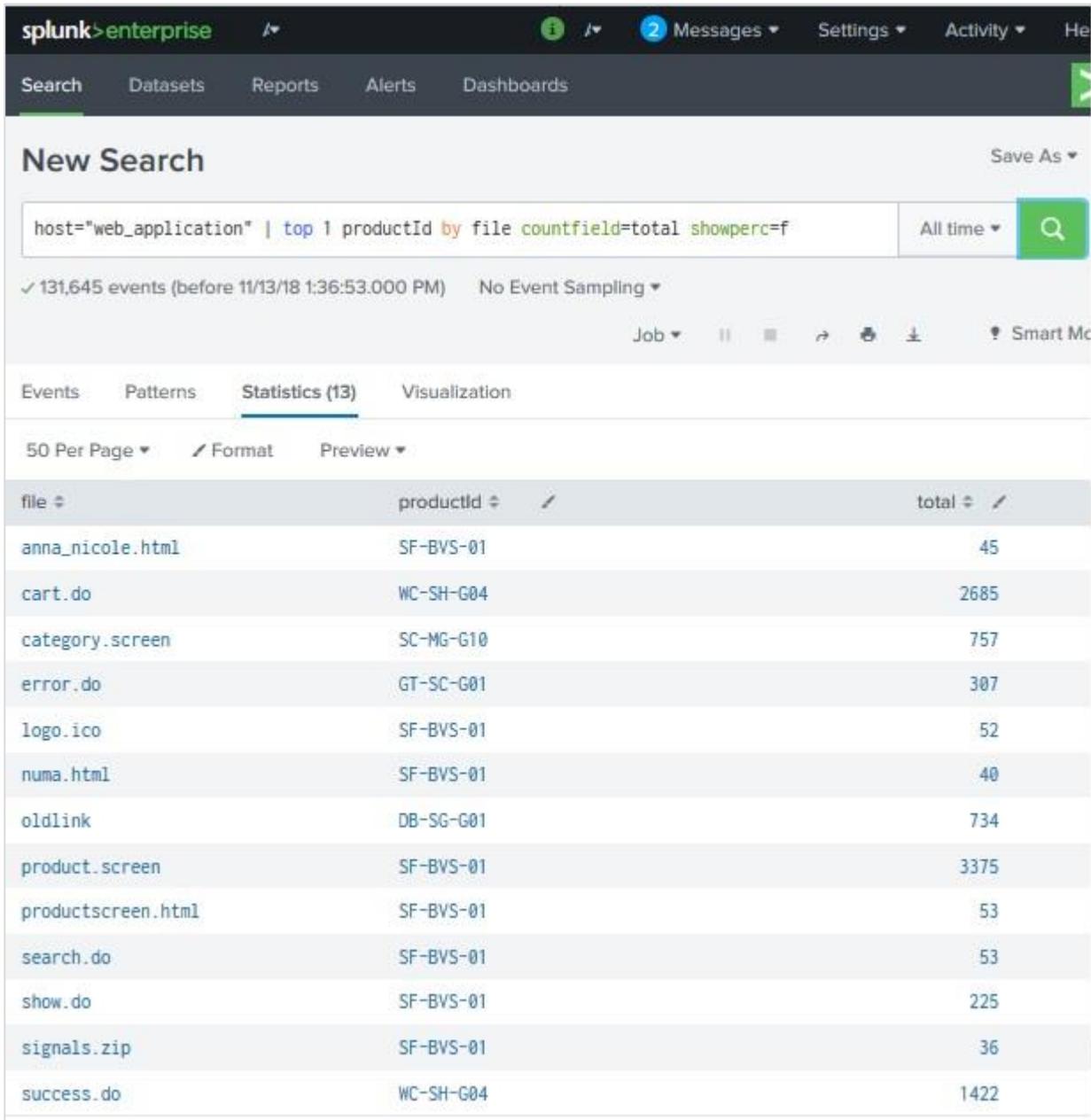
host="web_application" | top limit=3 productId by file

✓ 131,645 events (before 11/13/18 12:58:57.000 PM) No Event Sampling ▾

file	productId	count	percent
anna_nicole.html	SF-BVS-01	45	97.826087
anna_nicole.html	SF-BVS-G01	1	2.173913
cart.do	WC-SH-G04	2685	8.944633
cart.do	SC-MG-G10	2578	8.588180
cart.do	DB-SG-G01	2539	8.458258
category.screen	SC-MG-G10	757	8.490354
category.screen	DC-SG-G02	734	8.232391
category.screen	DB-SG-G01	734	8.232391
error.do	GT-SC-G01	307	100.000000
logo.ico	SF-BVS-01	52	100.000000
numa.html	SF-BVS-01	40	100.000000
oldlink	DB-SG-G01	734	8.322939
oldlink	SC-MG-G10	725	8.220887
oldlink	DC-SG-G02	680	7.710625
product.screen	SF-BVS-01	3375	10.268971
product.screen	WC-SH-G04	2647	8.053916
product.screen	DB-SG-G01	2581	7.853100

Show Options

We can also decide to show specific columns by using additional options available in Splunk with the Top Command. In the below command, we disable to show the percentage option and display only the top product ID by File name.



The screenshot shows the Splunk Enterprise search interface. The search bar contains the command: `host="web_application" | top 1 productId by file countfield=total showperc=f`. The results table displays 13 statistics entries, each with file, productId, and total counts. The table is sorted by total count in descending order. The 'Statistics (13)' tab is selected in the navigation bar.

file	productId	total
anna_nicole.html	SF-BVS-01	45
cart.do	WC-SH-G04	2685
category.screen	SC-MG-G10	757
error.do	GT-SC-G01	307
logo.ico	SF-BVS-01	52
numa.html	SF-BVS-01	40
oldlink	DB-SG-G01	734
product.screen	SF-BVS-01	3375
productscreen.html	SF-BVS-01	53
search.do	SF-BVS-01	53
show.do	SF-BVS-01	225
signals.zip	SF-BVS-01	36
success.do	WC-SH-G04	1422

34. Splunk – Stats Command

The stats command is used to calculate summary statistics on the results of a search or the events retrieved from an index. The stats command works on the search results as a whole and returns only the fields that you specify.

Each time you invoke the stats command, you can use one or more functions. However, you can only use one BY clause. If the stats command is used without a BY clause, only one row is returned, which is the aggregation over the entire incoming result set. If a BY clause is used, one row is returned for each distinct value specified in the BY clause.

Below we see the examples on some frequently used stats command.

Finding Average

We can find the average value of a numeric field by using the **avg()** function. This function takes the field name as input. Without a BY clause, it will give a single record which shows the average value of the field for all the events. But with a by clause, it will give multiple rows depending on how the field is grouped by the additional new field.

In the below example, we find the average byte size of the files grouped by the various http status code linked to the events associated with those files.

The screenshot shows the Splunk Enterprise search interface. The search bar contains the command: `host="web_application" | stats avg(bytes) by status`. The results section displays 131,645 events from before November 19, 2018, at 7:25:05 PM. The results are grouped by status code, with the average bytes for each status listed. The table has two columns: 'status' and 'avg(bytes)'.

status	avg(bytes)
200	2191.187422765344
400	2116.0206929740134
403	1823.4197828709289
404	2070.8220153340635
406	2098.6553537284894
408	2089.915982617093
500	2098.8101761252447
503	2092.354336283186
505	2083.1005917159764

Finding Range

The stats command can be used to display the range of the values of a numeric field by using the **range** function. We continue the previous example but instead of average, we now use the **max()**, **min()** and **range** function together in the stats command so that we can see how the range has been calculated by taking the difference between the values of max and min columns.

The screenshot shows the Splunk Enterprise search interface. At the top, there's a navigation bar with 'splunk>enterprise' logo, user info (2 messages), and links for Settings, Activity, and Help. Below the bar, a menu bar has 'Search' selected, followed by Datasets, Reports, Alerts, and Dashboards. To the right is a green search button with a white arrow.

The main area is titled 'New Search'. A search bar contains the query: 'host="web_application" | stats min(bytes) max(bytes) range(bytes) by status'. Below the search bar, it says '131,645 events (before 11/19/18 8:16:40.000 PM)' and 'No Event Sampling'. There are also 'Save As', 'New Table', and 'Close' buttons.

Below the search bar, there are filter and sort options: 'Job' dropdown, time range 'All time', and a search icon. The table has tabs for 'Events', 'Patterns', 'Statistics (9)', and 'Visualization', with 'Statistics' currently selected.

The table displays 9 rows of statistics for http status codes:

status	min(bytes)	max(bytes)	range(bytes)
200	200	47251	47051
400	202	4000	3798
403	160	3997	3837
404	202	4000	3798
406	200	4000	3800
408	201	4000	3799
500	202	3998	3796
503	202	3998	3796
505	200	3999	3799

Finding Mean and Variance

Statistically focused values like the mean and variance of fields is also calculated in a similar manner as given above by using appropriate functions with the stats command. In the below example, we use the functions **mean()** and **var()** to achieve this. We continue using the same fields as shown in the previous examples. The result shows the mean and variance of the values of the field named bytes in rows organized by the http status values of the events.

splunk>enterprise

Search Datasets Reports Alerts Dashboards >

New Search

host="web_application" | stats mean(bytes) var(bytes) by status

All time 

✓ 131,645 events (before 11/19/18 9:14:30.000 PM) No Event Sampling

Job ▾      

Events Patterns Statistics (9) Visualization

50 Per Page ▾  

status	mean(bytes)	var(bytes)
200	2191.187422765344	5304807.372529001
400	2116.0206929740134	1179873.994275495
403	1823.4197828709289	1462459.2462748317
404	2070.8220153340635	1198257.665290402
406	2098.6553537284894	1217765.0810664936
408	2089.915982617093	1239566.6518748675
500	2098.8101761252447	1183659.0711445606
503	2092.354336283186	1185161.3265968058
505	2083.1005917159764	1200765.841835327