**Submit your document as firstname\_INFT 1202.LAB3**

**Introduction to the Lab**

You can’t succeed in business without a smart data preparation strategy. If you don’t clean, validate, and consolidate your raw data the right way, you can’t ask deeper questions of it to get meaningful answers.

Splunk is an innovative technology which searches and indexes log files and helps organizations derive insights from the data. The main benefit of Splunk is that it uses indexes to store data, and so does not require a separate database to store its information.

Splunk is used for monitoring and searching through big data. It indexes and correlates information in a container that makes it searchable, and makes it possible to generate alerts, reports, and visualizations. It can recognize data patterns, create metrics, and help diagnose problems, for business challenges like IT management, security, and compliance.

In this lab, we will be configuring a Splunk server which will analyze and visualize the logs uploaded to the server.

**Lab Requirements:**

• Ubuntu Virtual machine o 2 GHz dual core processor or better (I would prefer 4 processors) o 4 GB system memory o 25 GB of free hard drive space o Internet access

o Either a DVD drive or a USB port for the installer media

**Steps of the lab:**

To get our VM up and running, we will need to execute the following steps:

**Step 1: Download and install Splunk Enterprise on ubuntu machine.**

**Step 2: Start Splunk and access its web interface.**

**Step 3: Getting Data in the Splunk**

**Step 4: Business Case – Dashboard Creation**

**Step 5: Discuss four differences between Security Onion and Splunk SIEM solution. Attach the reference links to your solution.**

**Execution of the lab (Saving your screenshots here):**

# Step 1: Download and install Splunk Enterprise on ubuntu machine [5Marks]

1. Create your account in https://www.splunk.com/ and select software Splunk Enterprise for download.
2. Select your installation package as per Linux OS and use **wget string** to download the the software.
3. Once the .tgz is downloaded, make sure it’ there by **using ls command**. **[Attach your screenshot]**
4. Splunk likes to live in **opt directory** [ Just in case, you do not know, /opt/splunk is considered as home directory for splunk]. So lets **move this .tgz in /opt directory by using mv command**.Type **sudo** first. After it’s done, **check .tgz presence in /opt directory by using ls command.**
5. Next, let’s unzip this .tgz file. We will used **sudo tar -xvzf filename.tgz. [4 Marks]** 
   1. What does option x means?
   2. What does oprtion v means? **iii.** What does option z means?

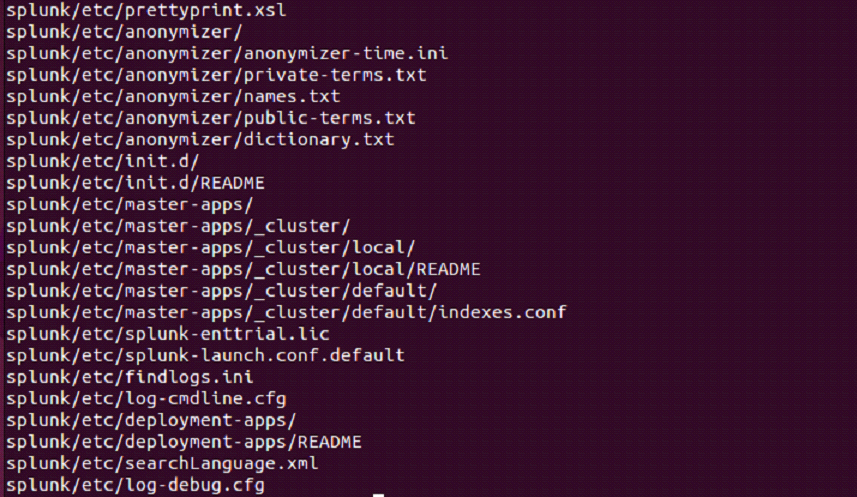
**iv.** What does option f means?

1. The above command is going to go through and build all of the directories that Splunk requires and all the default files that comes in those directories. **Attach the screenshot** for successful unzip operation. **[1Mark]**
2. Now , lets do **ls** and check out what it did. A new directory splunk has been created. Lets check what’s there in the splunk directory. Use **cd splunk** and then **ls.**  You will see **bin folder** from where we will be starting splunk in the next section

# Step 2: Start Splunk and access its web interface. [12 Marks]

1. Start Splunk and first time you start splunk, you need to accept the license agreement. And if you don’t accept like this, you will be forced to read the whole thing before you accept it.

What is the command to start splunk? **[Attach your screenshot] [ 1+2 Marks]**

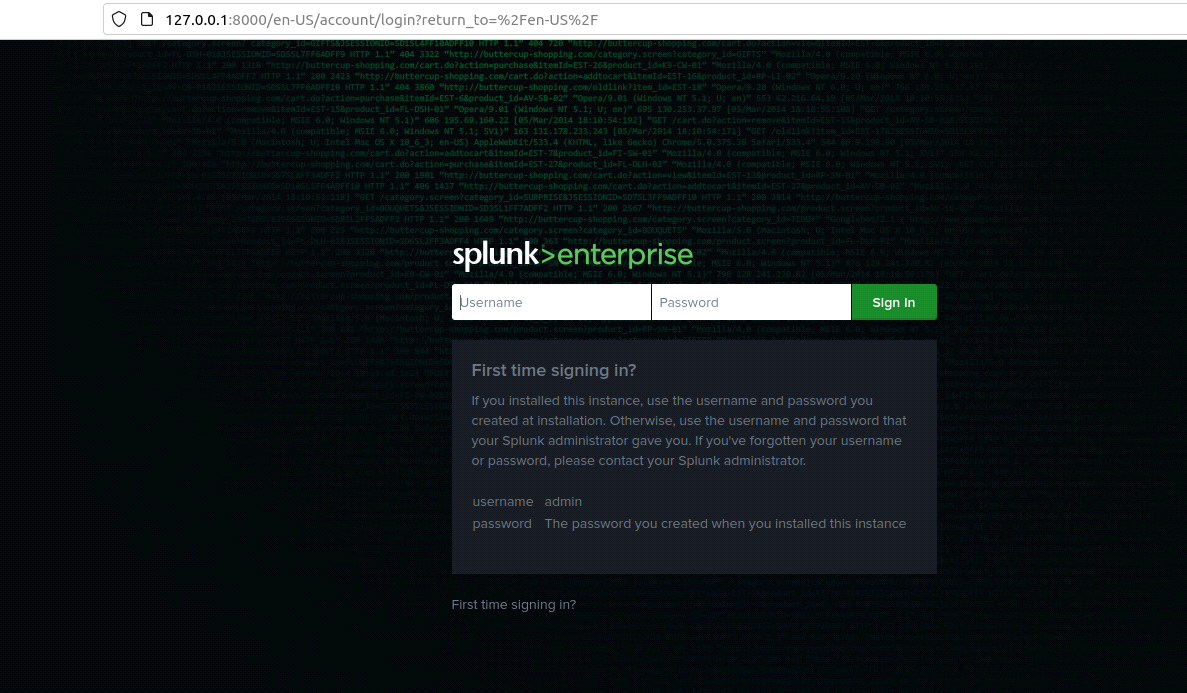


1. Splunk wants you to create a user name for the administrator and a password and then Splunk will go through its startup procedure. Please use **firstname\_admin** as your administrator username.
2. After initial start up process, it will provide you web interface to assess splunk- loop back address 127.0..1:8000 or [https://yourmachinename:8000](https://yourmachinename:8000/) or https://ipaddress:8000.

**Attach the screenshot** for successful start up of Splunk **[3 Marks]**

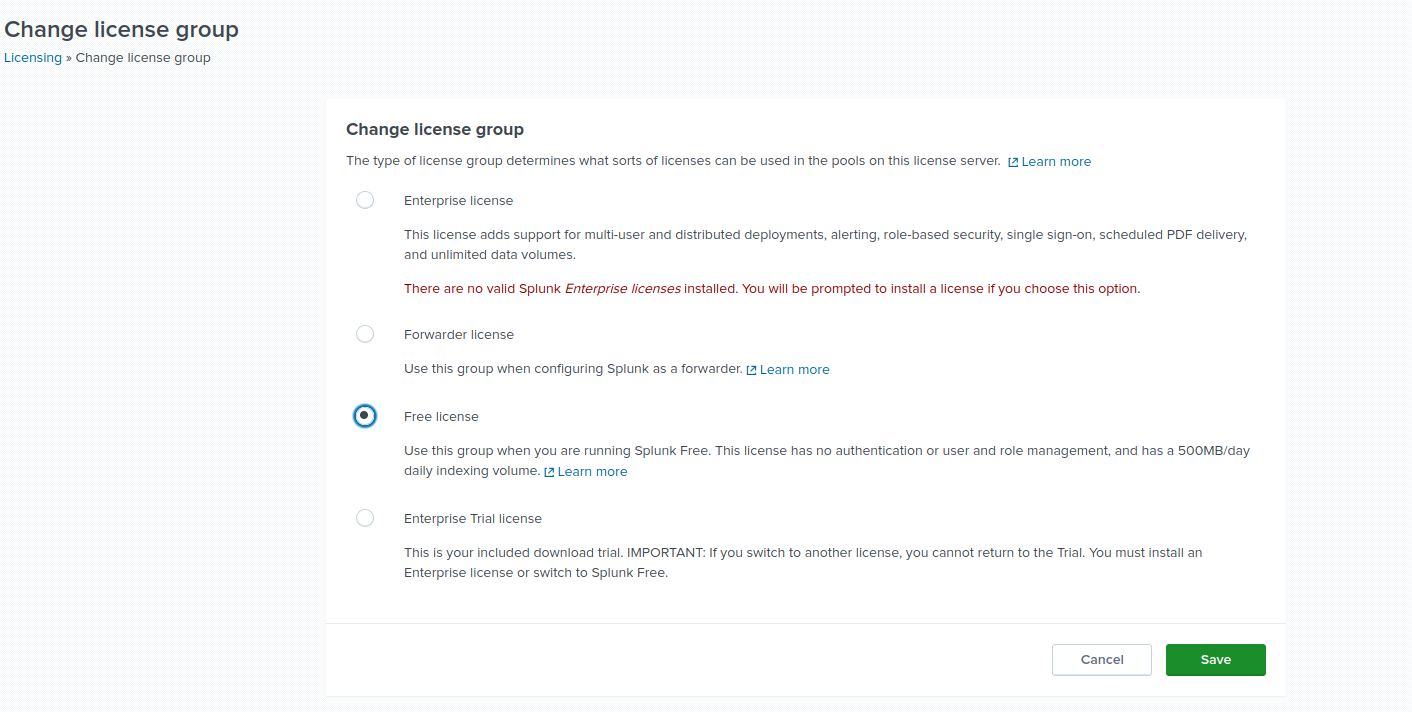
1. Go to web browser and type ipaddress:8000. It will take you to Splunk interface and enter the credentials you created when installing Splunk. It will show the Splunk interface if your

VM is on . **[Attach your screenshot] [3 Marks]**



1. Hurray , you got Splunk Server ready to go. Check the successful start of web interface of Splunk.
2. Check the license of Splunk, before starting uploading data in it. Go to settings -> licensing > change license group to Free license group. This license will provide you daily volume of 500

MB of quota to handle data. **[Attach your screenshot] [3 Marks]**

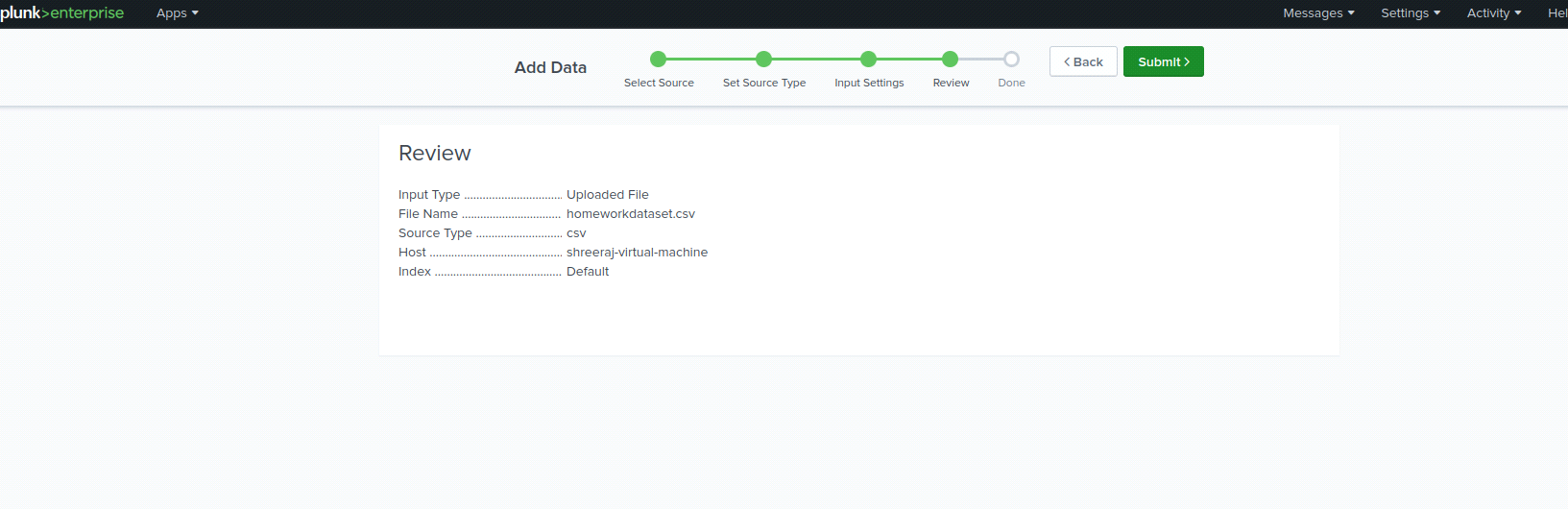


# Step 3: Getting Data in the Splunk [5 Marks]

**a.** Click **Add Data** to access the Add Data page. After you access the Add Data page, choose one of three options for getting data into your Splunk platform deployment with Splunk Web:

o Upload (This lab) o Monitor (Future labs) o Forward (Future labs)

1. Adding data from uploading files from computer. Upload the data set given to you- homeworkdataset.csv
2. Select source – homeworkdataset.csv and click next**.**
3. Set Source type shows how the Splunk platform sees your data before indexing. If the events are correct and have the right timestamps, click next.
4. Optionally set additional input parameters for this data input or keep default. If you keep your host as constant value, you should say your host name, in my case it was sukhwant\_virtual\_machine
5. Review your information and submit **[Attach your screenshot] [5 Marks]**



1. Final Step and click start searching

# Step 4: Business Case – Dashboard Creation [30 Marks]

Let's solve a business problem with a basic search. We're going to use a broad search and some basic commands; the business has asked us **what the backup duration was for each domain over the last 30 days.**

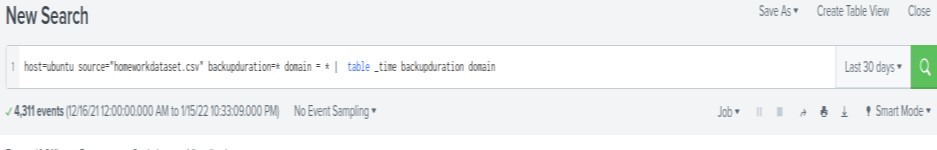
So, let's go above and beyond and create a nice dashboard for the business users that does in fact say backup duration by domain, but **also gives us the average duration over the last 30 days and the domain with the longest backup time over the last 30 days. And then a line chart that shows the trends in backup times over the last 30 days.**

**Note: Your screenshots would be different as my dataset ran for longer amount of time and repetitive.**

1. Click on search and reporting -> data summary-> your machine name
2. Select host = your machine, source = homeworkdataset.csv and time as last 30 days, you have

\_\_\_\_\_\_\_\_\_ data. **How much data you see? 4000 data. [2 mark]**

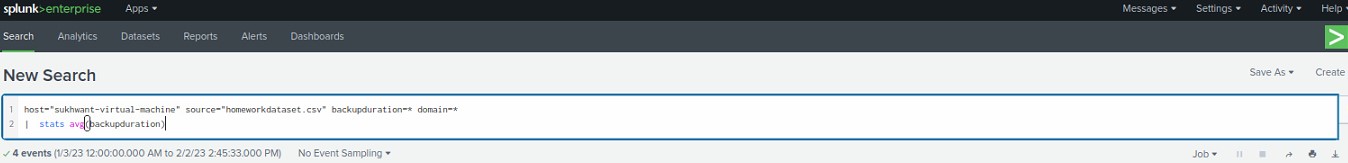
1. We are asked for backupduration and domain, we are going to search these fields with wild cards. Remember, the business wanted backup duration by domain over the last 30 days. **Write down the commands used [2 marks].** My screenshot is for reference. Don’t include in your submission. We have the date and time. The number of hours the back up took and the domain **[Attach your screenshot] [2 marks]**



1. Now, we can create a visualization with this data because we have mostly numerical data, so let's click visualization. And it already builds a column chart for us, the column chart displays the time on the X axis and the duration on the Y axis, and we can hover over any one of these columns.
2. And see the back of duration for that time for fun, let's change this to a line chart and see how that looks. **[Attach your screenshot] [2 marks]**
3. Now we can see a trend in our backup durations if we go to format. We can connect the lines, remember, we can also show data values, minimum maximum values. Your minimum value

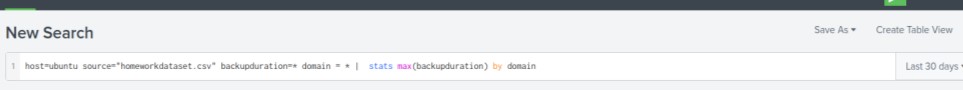
\_\_\_\_\_\_\_\_ and maximum is \_\_\_\_\_\_\_\_\_\_\_\_.  **[Attach your screenshot] [2 + 2 marks]**

1. Domain is not a numerical value, so it's not showing us anything useful. So actually, let's remove that from the table and run the search again.
2. Now, we have a trend in up duration. So, let's save this as a dashboard panel and we'll call the **dashboard backup**. And we'll call **this specific panel back up duration over time**. And let's click save.
3. Don't view the dashboard yet, we're going to add more stuff to it. Remember, the business wanted information about the domain in this chart doesn't show that this chart- just shows a trend of how long backups are taking. So, let's do a simple table with time, backup, duration and domain. And then click on the statistics tab and then run the search again.
4. And now let's save this as a dashboard panel as well. **Save as dashboard panel, choose existing dashboard backups panel title. Let's call it backup duration by domain**. And don't view the dashboard just yet.
5. It would also be useful on our dashboard to display the average time that backups are taking to do that, we're going to use a very popular command called **stats avg(backupduration).** And let's run that search.Youraverage back up duration time of \_\_\_\_\_\_\_ over the last 30 days. **[Attach your screenshot] [1 + 2 Marks]**

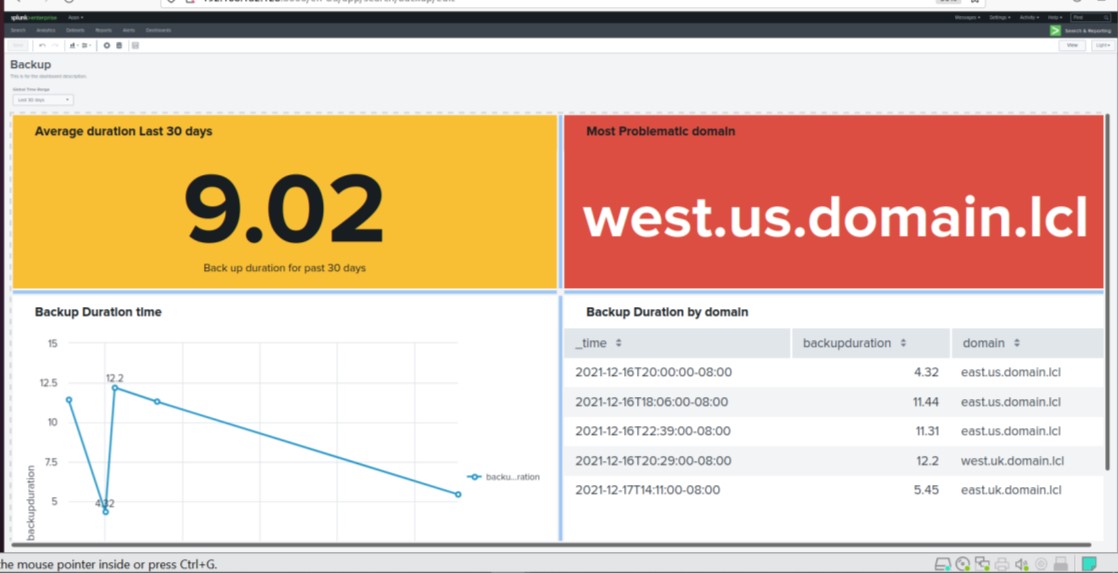


|  |  |
| --- | --- |
| l) | Let's do a visualization on that. And let's change line chart to single value. Do not underestimate the power of a single value visualisations. Let's go to format and we'll change the caption to average back up Duration Last 30 days, let's change the number format to give us two decimal places. Let’s go to color, choose use colors, let's use the color mode, block background. Let's say we want to change the color to yellow if the back up duration is over nine hours. So, under nine hours we're good. Over nine hours, we are in warning status. So, from minimum to nine. We're going to be green. Let's remove blue and let's say from nine to 15, we are in warning status and then from 15 on we are in critical status.  OK, close that box now. |

1. Let's save that panel as a dashboard panel. Existing dashboard backups, panel title, average duration last 30 days. Click save, don't view the dashboard just yet,
2. Let's make a panel that shows the maximum back up duration by domain. So, we're going to **use stats max(backupduration) by domain.** Let's search that and now it's going to give us the domain with backups that are taking the longest. **[Attach your screenshot] [2 Marks]**



1. Let's format this single value and we'll change the caption to **Domain with longest backup duration** and let's close that.
2. We're going to save as dashboard panel existing dashboard in the **panel title is going to be most problematic domain. [Attach your screenshot] [2 Marks]**
3. Now, let's view the dashboard, so we have resolved the business problem and we've also gone the extra mile by creating a dashboard, let's edit the dashboard and make it prettier.
4. Let's move our single values to the top. And let's put our backup duration table next to our line chart. And we'll save that dashboard.
5. Now, to make it prettier, let's edit the dashboard again. Let's **edit this panel so it only shows the top five and then it will line up with the back up duration line chart panel.** We'll do another pipe and then just you head five. Click apply**. [Attach the screenshot] [ 3 Marks]**
6. Same as above - For backupduration. **[Attach the screenshot] [ 3 Marks]**
7. Final Dashboard **[Attach your screenshot] [5 Marks].** My screenshot for reference. Please remove it in your submission.



# Step 5: Discuss four differences between Security Onion and Splunk SIEM solution. Attach the reference links to your solution. [ 08 Marks]

**Things to Explore:**

You are welcome to explore beyond the mandatory requirements if you wish.

**General submission requirements**

* Include an opening comment with your full name, date, and a short description.
* **Do not alter the sequence of steps of this document. Do not delete the words [Attach the screenshot] anywhere from the assignment. Keep the numbering same. Paste the screenshots wherever it’s being asked and answer the questions.**
* Please follow Academic Integrity guidelines.