Splunk Enterprise Deployment Practical Lab

Scenario

You have been asked by a customer to create a Proof of Concept (PoC) to demonstrate Splunk.
The customer would like to see the following demonstrated:

- **Three dashboards displaying information of interest to them.**
- **The ability to parse different data formats appropriately.**
- **No bad or extraneous data in the required indexes.**
- **High availability. Should one indexer go down, they will want to still be able to search all data.**
- **Monitoring of the entire Splunk environment.**
- **A scalable deployment approach.**
- **Application of the latest Splunk best practices.**

The customer has provided some more detailed requirements for this Proof of Concept, which you will find in the instructions below.

If you are uncertain about something, never hesitate to ask the customer! Asking questions is a good thing! For this class, the instructor is your customer.

However, when it comes to performing the work, **you are on your own.**

- None of your coworkers are available to assist you in any way. You must work alone.
- The customer does not know Splunk very well. They are counting on you to know Splunk best practices. If asked, however, they will gladly clarify their requirements for you.
- You have access to the materials from the Splunk courses that you have taken.
- You also have access to the Internet, including Splunk Docs, Splunk Answers, and the Splunk Community Forum, and any other online resource **other than another living person.**

Machines

The customer has provided eight Linux 64-bit machines, each running an AWS version of Ubuntu Linux. Note that Splunk has not yet even been installed on these machines.

The following Splunk best practices have already been done for you by the Linux system administrator.

You do **NOT** need to:

- Modify any Linux ulimit settings.
- Turn off Transparent Huge Pages (THP)
- Implement time synchronization.

These best practices are recommended prior to installing Splunk on a Linux environment. However, they have already been done for you.

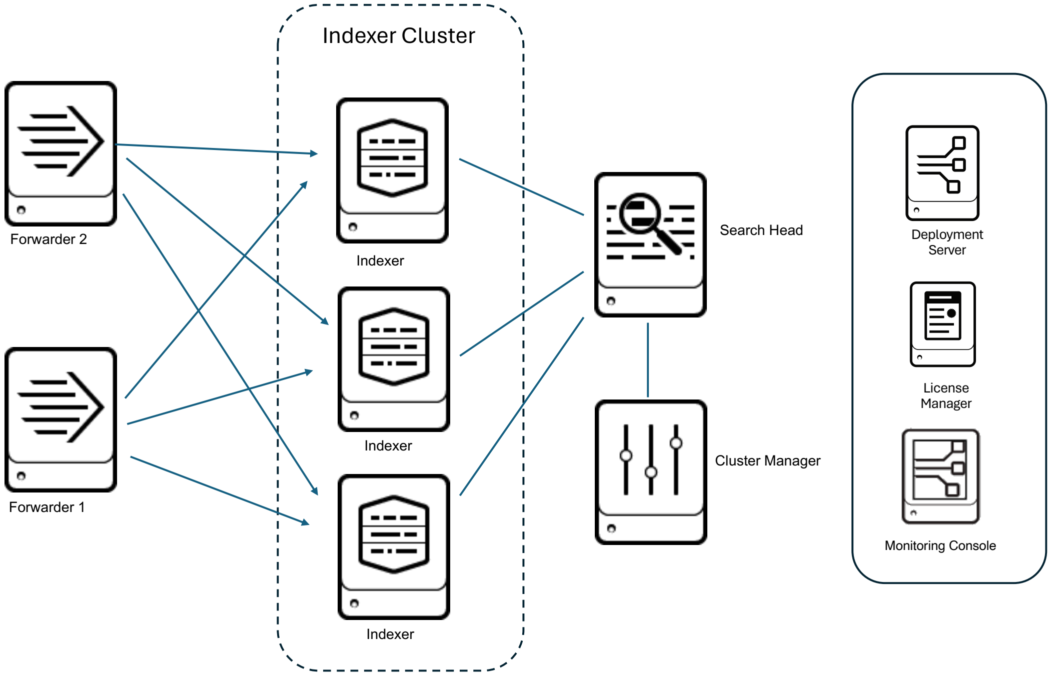
IP Addresses

You must use the **Internal IP** addresses in your configuration files, wherever IP addresses are needed. The internal IP addresses exist within a private network, which allows your servers to communicate with each other. In the diagram below, all communication between the servers occurs on the private network, using the internal IP addresses.

The **external IP** addresses allow you to access your servers via the command line or browser interface. The external IPs allow you to login or browse to your servers via the public internet.

If you have questions about which IP address to use in any situation, ask your instructor.

Topology Example



Credentials

Linux server login: **archStudent**

This account has sudo privileges.

When you install each instance of Splunk, create a Splunk user called **admin.** Give it the same password as archStudent.

Your personal password was provided to you in the credentials email that was sent to you prior to class. If you are uncertain as to your password, ask the instructor.

Do NOT change the password. Remember also to assign this password to the Splunk **admin** user that you create. Failure to use the assigned password will likely result in an automatic FAIL.

Phase 1 – Installation

Install and configure the Splunk deployment on the provided Linux 64-bit machines.

For all installation tasks, download a recent release of the installation files for Linux 64-bit Splunk Enterprise (or Splunk Universal Forwarder) from http://www.splunk.com/download.

We suggest that you use the install files with the `.tgz` extension. All servers have **wget** installed, so you can use the `wget` command that is provided on the Splunk download success page.

For every task, keep in mind and use, where applicable, Splunk best practices.
Note that not all the recommended best practices are spelled out for you in the instructions below.

A. Set your user account.

Best practice is to install and run Splunk with a dedicated account that exists solely for this purpose. The Linux system administrator has created one for us on all eight servers: **splunk.** You will need to become this user on each server with the following command:

```
sudo su – splunk
```

To confirm what your current user is, you can run the command: `whoami`

The user **splunk** also has sudo rights, and the same password as archStudent.

B. Install the Deployment Server / License Manager / Monitoring Console

1. On the Deployment Server / License Manager / Monitoring Console machine, create the Splunk home directory:

   ```
   sudo mkdir /opt/splunk
   sudo chown splunk /opt/splunk
   ```

2. Install Splunk in: `/opt/splunk`

   ```
   tar -xzvC /opt -f {splunk_file_downloaded_using_wget}
   ```

3. Assign a unique Splunk server name to this instance that includes your name and the purpose of the instance (i.e., `SMITH-LICENSE`).

4. Configure Splunk using best practices and designate this instance as a license manager. Download and install the big license from https://splk.it/edu-lab-licenses (password: **open.sesam3**).

   Following best practices, what are some additional steps that you should include when installing a Splunk instance?

   You will want to follow a similar set of steps in the installs that follow.

C. Install the Cluster Manager and indexers

5. Install Splunk on the Cluster Manager and each of the indexer machines in: `/opt/splunk`

6. Assign unique Splunk server names to each instance that includes your name and purpose of the instance (i.e., `SMITH-INDEX1`, `SMITH-INDEX2`).

7. Configure the indexer cluster with a replication factor of 2 and search factor of 2.

8. The customer wants their data in three different indexes. These indexes will require High Availability. Configure the indexer cluster for three replicated indexes: **`os`**, **`gamelog`**, and **`network`**.

   Before you create these three indexes, what are some questions that you might have for the customer? This might be a good time to ask them!

D. Install the forwarders

9. Install universal forwarders on the forwarder machines in: `/opt/splunkforwarder`. You will of course want to manage these in a scalable manner.

10. Configure the forwarders to send data to the indexer cluster. You will want to ensure that data is distributed evenly among the available indexers.

E. Install the Search Head

11. Install Splunk on the Search Head machine in: `/opt/splunk`
12. Assign a unique Splunk server name to this instance that includes your last name and purpose of the instance (i.e., `SMITH-SEARCH`).
13. Configure the Search Head to search the indexer cluster.

F. Forward internal logs to the Indexers

14. Configure Search Head Forwarding for the Search Head, DS/LM/MC, and Cluster Manager instances.

> **If you are using version 9.2.x or higher**, configuring the DS/LM/MC instance to forward to the indexers will cause your forwarders to not appear in Forwarder Management. To resolve this, you need to add the DS/LM/MC instance as a Search Head in the cluster. (As you did in step 13 above.)
>
> Be advised that it may be several minutes before your forwarders appear again in Forwarder Management.

G. Configure the Monitoring Console

15. Configure the Monitoring Console (MC) on the DS/LM/MC instance to monitor the ***entire*** environment. This should include the forwarders.

Phase 2 – Configure Data Inputs and Fields

A. Configure and deploy the data inputs

Since your approach must be scalable, you will want to use the Deployment Server to deploy apps for all data inputs.

The customer will be displeased if they see any bad data in any of the three indexes that they have requested. It is therefore an important best practice to test each input before routing the data into these indexes. Good data has proper event boundaries, valid timestamps, and correct host, source, and sourcetype field values.

Forwarder 1

1. Deploy the 'Splunk Add-On for Unix and Linux' add-on to monitor all files in `/var/log` and send them to the **`os`** index.

   If the splunk account does not have permission to access this directory, how might you resolve this? If you are not a Linux expert, perhaps the Internet can provide some helpful suggestions.
2. Monitor the file `access.log` from all three `www*` directories (`/opt/log/www*`) and send them to the **`network`** index. The host value for these events should be derived from the third directory segment in the pathname (for example, `www1`).

Forwarder 2

3. Monitor `cisco_ironport_web.log` from `/opt/log/cisco_router1` and send it to the **`network`** index.
4. Monitor `cisco_ironport_mail.log from /opt/log/cisco_router1` and send it to the **`network`** index.
5. Monitor `dreamcrusher.xml` from `/opt/log/crashlog` and send it to the **`gamelog`** index.

   What should events from the xml file look like? Each event should begin with the tag <Interceptor> *and end with the tag* </Interceptor>. (*For this class, you can ignore any header lines in the file. Don't worry if they get indexed.)*

Confirmation

6. From the Search Head, this would be a good point to validate that the data in the three required indexes looks correct. Remember, the customer does not want to see any bad data!

B. Create field extractions.

7. Make sure the following fields are being extracted for the `/var/log/auth.log` file:

   - user
   - src_ip

   If you installed the 'Splunk Add-On for Unix and Linux' add-on correctly, you should see these fields correctly populated.

8. Create the following search-time field extractions for `cisco_ironport_web.log`:
   - user (example: `doc@demo.com`)
   - domain (example: `www.adventureindonesia.com`)
   - url (example: `http://www.adventureindonesia.com/images/komodo/komodo.jpg`)
9. Create the following search-time field extractions for `cisco_ironport_mail.log` (field values should all be numeric):
   - mid
   - icid
   - dcid
10. Create the following search-time field extractions for `dreamcrusher.xml`:
    - Infiltrators
    - AttackVessel

Did you check the permissions on these fields to ensure that everyone can see them?

Phase 3 – Reporting

In this phase, you will create reports and dashboards. The three dashboards are to be called **OS**, **Network**, and **Game Activity.** The dashboard panels are to be powered by the reports.
As you create your searches, keep in mind search best practices!

A. Create the following report and display it on the **OS** dashboard.

1. For the /var/log/`auth.log` input, display a count of failed logins in the last 60 minutes by `user` and `src_ip`.
   Name the dashboard panel: **Failed Logins by User – Last 60 minutes.**

B. Create the following reports and display them on the **Network** dashboard.

2. For `access.log`, display a count of web server errors in the last 24 hours by status code and host.
   Name the panel: **Web Server Errors – Last 24 hours.**
3. For `cisco_ironport_web.log`, list all events in the last 24 hours that contained either `.exe` or `.bat`.
   Name the panel: **Suspect Events Summary – Last 24 hours**
4. For `cisco_ironport_web.log`, count the number of suspect events from the above search for each user and present it in a table.
   Name the panel: **Suspect Events Summary by User – Last 24 hours**
5. For `cisco_ironport_mail.log`, group all events for a particular piece of email together (you can correlate on the fields mid, `dcid`, and `icid`), and then search for the term REJECT.
   Name the panel: **Rejected Email Transactions – Last 24 hours.**

C. Create the following report and add to the **Game Activity** dashboard.

6. For `dreamcrusher.xml`, calculate (sum) the total number of `Infiltrators` for each `AttackVessel`. Name the panel: **AttackVessel usage - All Time**

Did you check the permissions on these reports and dashboards to ensure that other users can access them?

Wrap-up

Scenario Revisited

Did you meet the requirements set forth by the customer for this Proof of Concept?

- **Three dashboards displaying information of interest to them.**

  Are there three dashboards displaying the data that they requested?

- **The ability to parse different data formats appropriately.**

  When they search the three indexes, do they see good data? Correct host, source, and sourcetype values? Proper event boundaries and valid timestamps?

- **High availability. Should one indexer go down, they will want to still be able to search all data.**
  Does the Cluster Manager show that buckets in the three indexes are being replicated?

- **Monitoring of the entire Splunk environment.**

  When you bring up the Monitoring Console, does it show the health of all eight Splunk instances?

- **A scalable deployment approach.**
  Can you explain to the customer how the procedures that you followed in creating this Proof of Concept will scale to handle 100 clustered indexers and 500 forwarders?

- **Application of the latest Splunk best practices.**

  Could you point out to the customer some of the best practices that you implemented?