Installing and Configuring Splunk on Red Hat Enterprise Linux or CentOS Servers

Richard Davis
May 2014

Introduction to Splunk:

Ask any network or server admin about the importance of logs as they relate to performing his or her job and chances are good they'll tell you it would be difficult, if not impossible, to operate without them. The problem is, networks aren't getting any smaller. For every router, switch, firewall, server or other device with an IP address that we add to our network, we've got that many more logs to sift through when something goes wrong. A few years ago, accessing and parsing that log data meant we had to SSH or RDP to the device in question and utilize our skills using Windows Event Viewer, or more commonly grep on a Linux/UNIX machine.

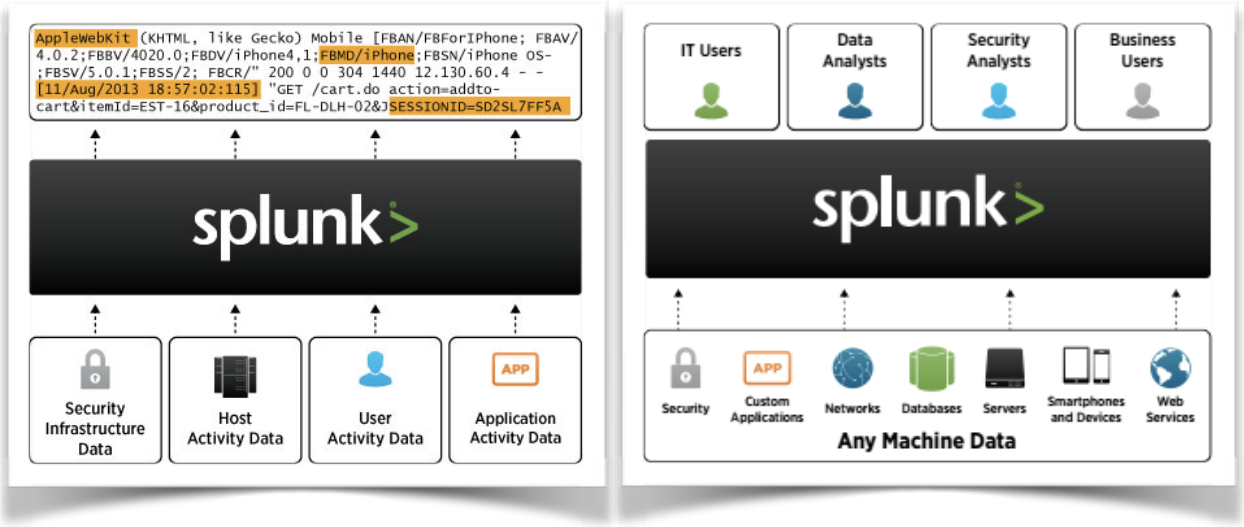
Splunk is log aggregation software that runs on Windows, OS X, Linux and UNIX. For every device on our network that generates log data, be it via SNMP traps or informs, Syslog, NetFlow, or any kind of file written to a file system, a Splunk Universal Forwarder (a small daemon running on the system) will take that data and send it to a Splunk indexer. The Splunk indexer will serve as the actual collector of the data. The indexer also provides the web GUI with which you interact to search through the various data that has been indexed. The Universal Forwarders can collect any data you tell them via modification of a file called "inputs.conf." This file points to the files on the file system that should be sent to the indexer. The "outputs.conf" file specifies the location of the indexing server. In the case of Syslog data, a Universal Forwarder isn't even necessary because Splunk will natively function as a Syslog server. That is, you can send Syslog data directly to the indexer, and it will happily receive it.

The limit of the amount of data you can send to the Splunk indexer depends on the license you have. When you are logged in to the Splunk GUI, the data is shown in real time. You can use the search blank just as you would with Google to search for literally anything the logs may contain. Because Splunk has its own search language, you can build very granular searches to isolate very specific events. Splunk will automatically extract some fields from the log data and you can easily train the software to extract other fields via RegEx. For example, you could take some proprietary log file generated for a custom piece of software and train Splunk that the 3rd comma separated value is the IP address of the client establishing a connection to that software. You could then have Splunk grab that information and perform GeoIP lookups to determine the geographical location of that IP address. By installing a free "Google Maps" app into Splunk, that data could then be plotted, in real time, on a global map.

Splunk is very extensible and scalable. By searching through Splunk Base, you can find dozens of free apps, which can extend the functionality by adding things like Google Maps or even Cisco firewall apps that show firewall data in real time. Because most of Splunk is written in Python, you can easily use Python to extend the functionality even further by writing your own apps.

Log-aggregation software is just as beneficial for Information Security and eDiscovery as it is for a network or server admin troubleshooting a particular problem. Splunk can be used within security investigations to correlate data and obtain forensic evidence that

can be used to track down specific events. Additionally, let's assume a "bad guy" compromises a server. When this happens, the attacker often tries to cover his or her tracks by removing log data. With Splunk, even if the original log data is deleted, chances are it's already been sent to the indexer and is safely stored in Splunk's database. Alerting capability is also included, such that you could tell Splunk to search in real time for a particular event and send an email immediately upon detection of such an event.



(source: http://www.splunk.com/view/splunk/SP-CAAAG57)

Getting Started:

This manual will serve as your guide to installing and configuring Splunk on a Red Hat Enterprise Linux® (RHEL) or CentOS server. There are two major components of Splunk on which we will focus: the indexer, and the Universal Forwarder. The indexer is the head-end device that runs the Splunk GUI and collects log data. When you think of Splunk, this is the main component you will be using. The Universal Forwarders send data from the various sources (Linux/UNIX, OS X, Windows, etc.) to the indexer. The first part of this guide is going to focus on the initial download, installation, and configuration of the indexer. After we are up and running, we will transition to the Universal Forwarder. Please note, this manual assumes you have a basic to intermediate knowledge of Linux, specifically RHEL or CentOS distributions. You need not have any prior knowledge or experience with Splunk. We will cover everything you need to get a basic deployment up and running within this manual.

Let's get started! Please visit www.splunk.com and click the *Login* link to access your Splunk.com account. If you do not have an account, please choose "*Sign Up Now*" from the login page to create a new account.

After you have logged in with your Splunk.com account, click the "*Free Download*" button to be redirected to the Splunk download page. As of this writing, the current major version of Splunk Enterprise is 6. Choose the 64-bit (or 32-bit if applicable) Linux RPM download. The filename should be similar to: splunk-6.*x*.*x*-*xxxxxx*-linux-2.6-x86_64.rpm or splunk-6.*x*.*x*-*xxxxxx*.i386.rpm, where *x* will change depending on the current minor version and build number.

On the next page, click the "***Got wget? Get this URL.***" link on the right side of the page. All base installations of RHEL and CentOS include the *wget* utility. Copy the URL to your clipboard, and then sign in to the Linux server on which the Splunk indexer will be installed. As a best practice, it is recommended you not install Splunk as the root user. Paste the contents of the clipboard to your terminal window and execute the command. This should start the download of the Splunk indexer RPM. Alternatively, you may manually download the file from another computer and transfer it to the Linux server via SCP, SFTP, or some other means, or download the file directly on the Linux server via XWindows.

Indexer Installation:

Once you have successfully downloaded the Splunk indexer RPM installation package, enter the following command:

rpm –ivh *filename.rpm* (Where *filename.rpm* is the name of the file you just downloaded.)

Example:
rpm –ivh splunk-6.1.1-207789-linux-2.6-x86_64.rpm
This will initiate the installation process. Once the process is complete, you will need to start Splunk for the first time. To do so, enter the following command:

/opt/splunk/bin/splunk start

Please read and accept the license agreement, and wait for the Splunk initialization to complete.

Indexer Initial Configuration (CLI):

Once the installation and initialization process is complete, Splunk should be successfully running on your system. It is recommended you create an init script so that Splunk can then be controlled with the *service* command. To do so, enter the following command:

/opt/splunk/bin/splunk enable boot-start
This command will create an *init* script in /etc/init.d, and will allow you to control the Splunk daemon using as follows:

service splunk stop Stop collecting data and safely stop the Splunk daemon

service splunk start Start the Splunk daemon

service splunk restart Stop and start the Splunk daemon
This is commonly used to enable certain configuration Changes to take effect.

You may also disable or enable the Splunk daemon from starting on boot using the *chkconfig* command:

chkconfig splunk on Enable the Splunk daemon to start on boot.
chkconfig splunk off Disable the Splunk daemon from starting on boot.

Note: the Splunk boot-start command mentioned above not only creates the init.d service script, it also tells the daemon to start on boot (as in *chkconfig splunk on*).

Next, we need to configure the Linux firewall, *iptables* by default on RHEL and CentOS, to allow inbound access for ports required by Splunk.

While we have not yet configured Splunk to listen for Universal Forwarder connections (explained later) on any ports or accept Syslog traffic, we will provision the firewall accordingly so that our configuration will work when we reach those steps.

Enter the following commands:

iptables -A INPUT -p udp -m udp --dport 514 -j ACCEPT
iptables -A INPUT -p tcp -m tcp --dport 8000 -j ACCEPT
iptables -A INPUT -p tcp -m tcp --dport 9997 -j ACCEPT

The first line allows UDP connections on port 514 from any source. This is important if you intend to use Splunk to collect Syslog data.

The second line allows TCP connections on port 8000 from any source. By default, Splunk runs on TCP/8000, and in order to access the Splunk GUI we will need to allow this port.

The third line allows TCP connections on port 9997 from any source. This port is commonly used by Universal Forwarders to send data to the Splunk indexer (the machine we are currently configuring).

After you have entered the iptables rules, we will save the configuration and restart the iptables daemon. To do so, enter the following commands:

service iptables save
service iptables restart

To view the iptables configuration at any time, use the command:

iptables –L

If you wish to restrict the source IP addresses from which connections can be initiated to the ports listed above, you may use the –s option as in the following example:

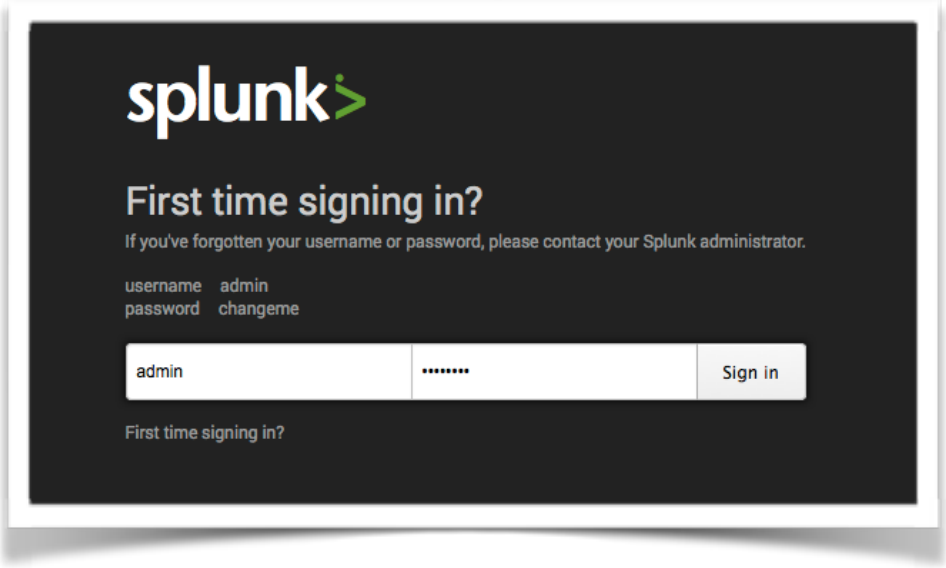
iptables -A INPUT –s 192.168.1.0/24 -p udp -m udp --dport 514 -j ACCEPT

The above command allows Syslog connections sourced from computers on the 192.168.1.0/24 subnet to reach UDP/514 on this server.
At this point, we have reached the end of the basic CLI indexer configuration. The remaining configuration will be performed via the Splunk web interface.

Indexer Initial Configuration (Web Interface):

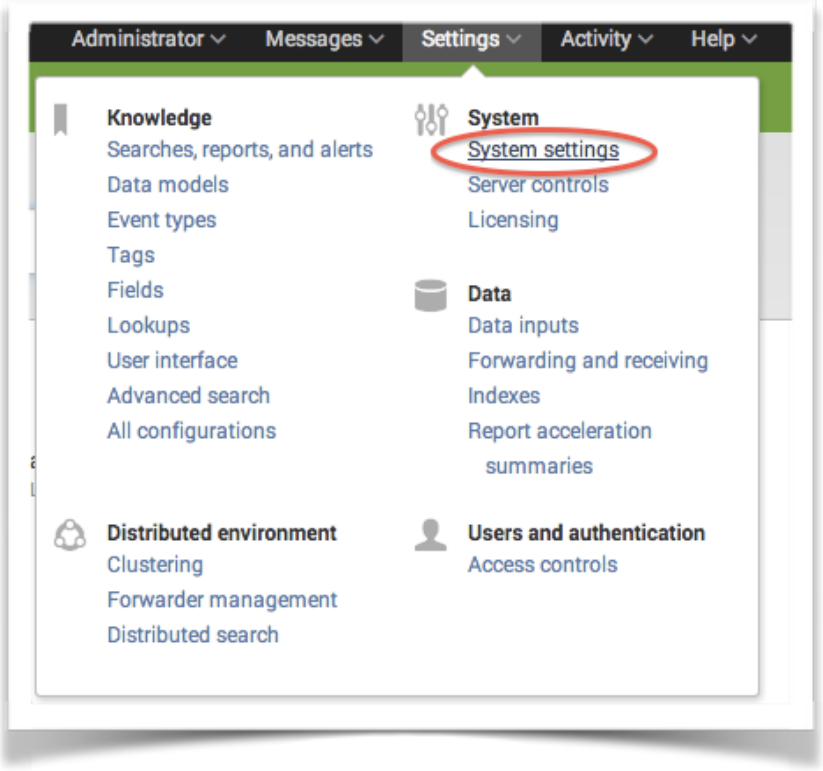
Using your favorite browser, please navigate to the IP address or DNS name of the server on which we have been configuring Splunk, specifying http using port 8000.

At the Splunk login screen, enter the default credentials as shown in the screenshot below and click *Sign In*. You will be prompted to change the password during this initial login.



(source: http://docs.splunk.com/Documentation/Splunk/6.1.1/SearchTutorial/StartSplunk)

The first thing we will want to do is change the Splunk indexer to use HTTPS. To do so, click the **Settings** menu in the upper right-hand corner, and then choose **System Settings** under the **System** section. Then, choose **General Settings** from the next page.



A few commonly changed settings on the General Settings page include:

Splunk server name:
By default, this will be the local hostname of the Linux server on which the indexer has been installed. You may change the name here.

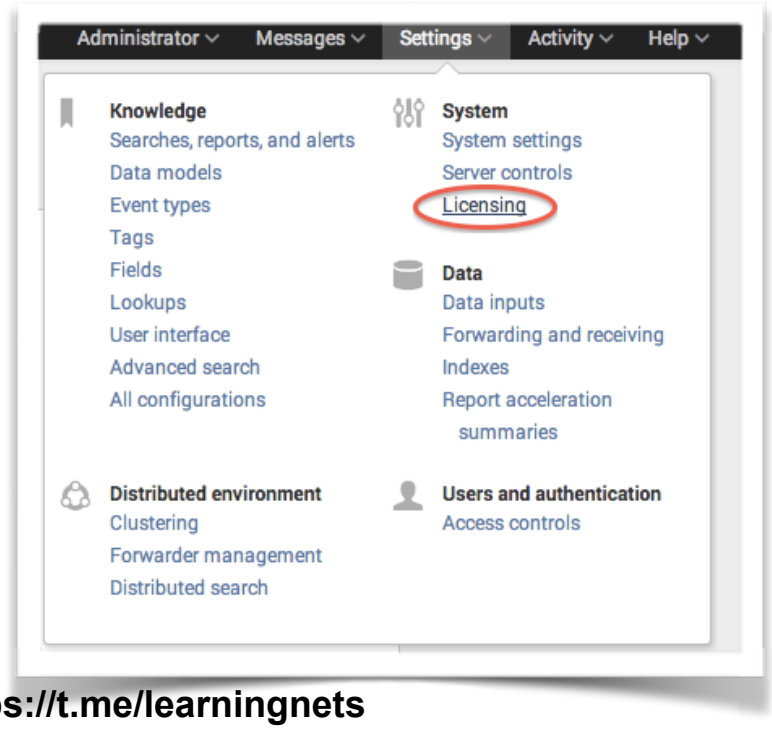
Splunk Web > Enable SSL (HTTPS) in Splunk Web?:
By default, this option will be set to No. Change the radio button to **Yes**.

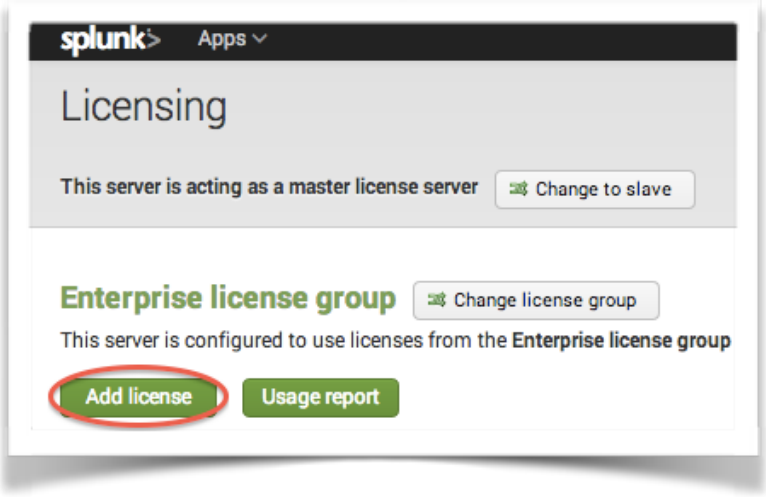
Web port:
The default web port is **8000**, and can be changed here as well. Keep in mind that should you change the default port, you will need to modify the iptables firewall entry we created earlier to allow inbound connections for that port.

Session timeout:
By default, inactive sessions will timeout in **1 hour**.

Once you have made your changes, click the **Save** button at the bottom of the page.

Next, we need to License the Splunk indexer. Splunk offers a free 500MB per day license that can be great for home use or testing, but you will want to purchase an enterprise license from Splunk to use the product in an enterprise environment. The free version does not have any authentication options (i.e. anyone can access the Splunk web interface, is capped at 500MB per day maximum, and has no included technical support.)



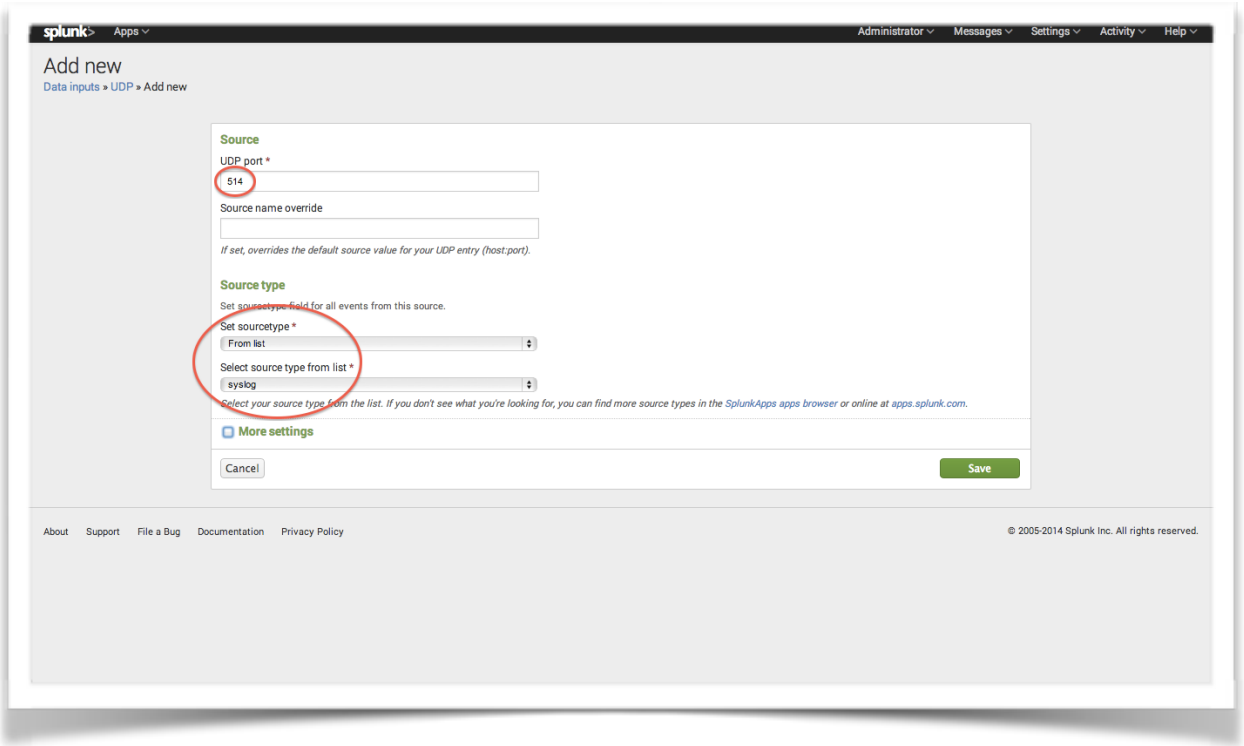


To install a Splunk license, click the **Settings** menu in the upper right-hand corner, and then choose **Licensing** under the **System** section.

From the licensing page, select the **Add license** button to apply the license you received from Splunk. If you have not purchased a Splunk license, you may continue with the using the trial or free license and later license the product.

Next, we will configure Splunk to receive data from two different source types: Syslog and Universal Forwarders.

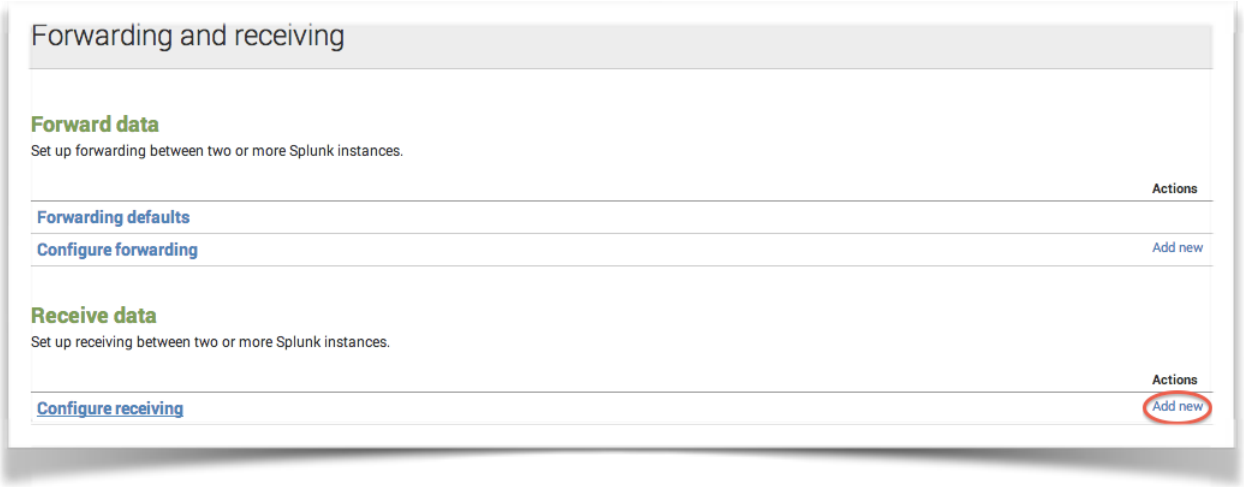
Syslog:
Syslog uses, by default, UDP port 514. To configure Splunk to accept Syslog data, click the **Settings** menu in the upper right-hand corner, and then choose **Data inputs** under the **Data** section.

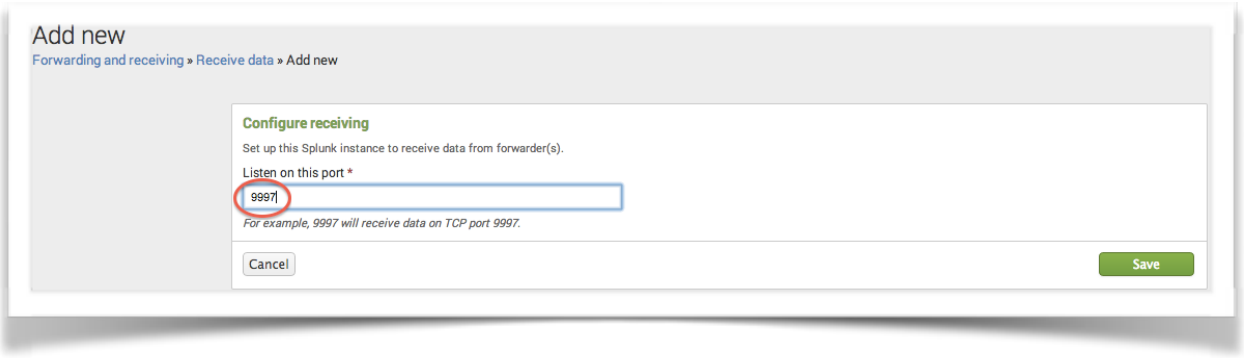


Choose **UDP** from the next page, as that is the layer 4 protocol used by Syslog. Choose **New** to add a new UDP port. Enter **514** in the **UDP port** field, then under the **Source type** section choose Set sourcetype: **From list**, Select source type from list: **syslog**. Once you have made your changes, click the **Save** button at the bottom of the page.

Now we will configure Splunk to also receive data from Universal Forwarders. From Splunk.com: The universal forwarder is a streamlined, dedicated version of Splunk Enterprise that contains only the essential components needed to forward data. It has no searching, indexing or alerting features, does not parse data and does not include a bundled version of Python. Universal forwarders have a default transfer rate of 256Kbps. (source: http://www.splunk.com/web_assets/pdfs/secure/Splunk_Forwarders_Tech_Brief.pdf) The next section of this manual will detail the installation and configuration of a Universal Forwarder on a RHEL or CentOS server. Before that data can be collected by the indexer, we must tell the indexer to listen for and accept that data.

To configure receiving, click the **Settings** menu in the upper right-hand corner, and then choose **Forwarding and receiving** under the **Data** section. On the next page, choose **Configure receiving**. Choose **New**, and then enter **9997** in the **Listen on this port** field. Please note that you can use any available port, but you must change the iptables firewall accordingly.





Once you have made your changes, click the **Save** button at the bottom of the page.

Congratulations! You have completed the basic installation and configuration of the Splunk indexer, and it is now listening for Syslog data on UDP/514 and Universal Forwarder data on TCP/9997.

Next, let's install and configure a Universal Forwarder on another Linux server so we can send data to the indexer. Please note that while this manual only shows the procedure for a Linux server, the Universal Forwarder is available for all platforms and the process is very similar across operating systems.

Universal Forwarder Installation:

The next two sections will be very familiar to you if you followed this manual to install the Splunk indexer. The installation and configuration of the Universal Forwarder has many similarities to that of an indexer. The biggest difference is that the Universal Forwarder is configured via CLI only, and has no web interface. Think of it as a stripped-down "light" version of Splunk designed to use very little memory and resources, and to efficiently and quickly forward data to the indexer.

Please go to https://www.splunk.com/downloads/universalforwarder and choose the 64-bit (or 32-bit if applicable) Linux RPM download. The filename should be similar to: splunkforwarder-6.*x*.*x*-*xxxxxx*-linux-2.6-x86_64.rpm or splunkforwarder-6.*x*.*x*-*xxxxxx*.i386.rpm, where *x* will change depending on the current minor version and build number.
On the next page, click the "***Got wget? Get this URL.***" link on the right side of the page. All base installations of RHEL and CentOS include the *wget* utility. Copy the URL to your clipboard, and then sign in to the Linux server on which the Splunk Universal Forwarder will be installed. As a best practice, it is recommended you not install Splunk as the root user. Paste the contents of the clipboard to your terminal window and execute the command. This should start the download of the Splunk Universal Forwarder RPM. Alternatively, you may manually download the file from another computer and transfer it to the Linux server via SCP, SFTP, or some other means, or download the file directly on the Linux server via XWindows.

Once you have successfully downloaded the Splunk Universal Forwarder RPM installation package, enter the following command:

rpm –ivh *filename.rpm* (Where *filename.rpm* is the name of the file you just downloaded.)

Example:
rpm –ivh splunkforwarder-6.1.1-207789-linux-2.6-x86_64.rpm

This will initiate the installation process. Once the process is complete, you will need to start Splunk for the first time. To do so, enter the following command:

/opt/splunkforwarder/bin/splunk start

Please read and accept the license agreement, and wait for the Splunk initialization to complete.

Universal Forwarder Configuration:

Once the installation and initialization process is complete, Splunk should be successfully running on your system. It is recommended you create an init script so that Splunk can then be controlled with the *service* command. To do so, enter the following command:

/opt/splunk/bin/splunk enable boot-start
This command will create an *init* script in /etc/init.d, and will allow you to control the Splunk daemon using as follows:

service splunk stop	Stop sending data and safely stop the Splunk daemon
service splunk start	Start the Splunk daemon
service splunk restart	Stop and start the Splunk daemon *This is commonly used to enable certain configuration Changes to take effect.*

You may also disable or enable the Splunk daemon from starting on boot using the *chkconfig* command:

chkconfig splunk on	Enable the Splunk daemon to start on boot.
chkconfig splunk off	Disable the Splunk daemon from starting on boot.

Note: the Splunk boot-start command mentioned above not only creates the init.d service script, it also tells the daemon to start on boot (as in *chkconfig splunk on*).
Before we edit the necessary configuration files for the forwarder, the first thing we will do is change the password. As with the indexer, the default password is ‘changeme’. Obviously, you should never leave any password at its default.

To change the password, enter the following command:
/opt/splunkforwarder/bin/splunk edit user admin -password '*your-password-here***'**
… replacing ‘*your-password-here*’ with the actual password you would like to use.

Note:
This command will likely be logged in your bash history. If you wish to remove the command so the password doesn’t appear in your history, simply edit the file ~/.bash_history and remove the line. Remember, the bash_history file will only be

written when you log out. You would need to log out and back in again, then edit the file and remove the line.

For the purposes of the Splunk Universal Forwarder, all of the configuration files we will need to edit will be located in **/opt/splunkforwarder/etc/system/local**.

We will focus on three files:

- inputs.conf
- outputs.conf
- server.conf

Using your favorite editor (Vi/Vim, Emacs, Nano, Pico, etc.), edit the **inputs.conf** file.

Example:
vi /opt/splunkforwarder/etc/system/local/inputs.conf

This is the most important file you will edit regarding the Splunk Universal Forwarder configuration, as it defines which files will be sent to the indexer for logging. On a RHEL or CentOS box, it is recommended you index, at a minimum, the following files:

- /var/log/messages
- /var/log/secure
- /var/log/yum.log
- /var/log/httpd/* (if applicable)
- /var/log/maillog (if applicable)

Splunk configuration files are organized using stanzas. Bracketed text defines each stanza as shown below in the index.conf sample file. This file contains four stanzas. The first stanza sets some global defaults. Stanzas two through four define the files we are sending to the Splunk indexer.

Sample index.conf file:

```
[default]
host = myserver.local
sourcetype = linux

[monitor:///var/log/messages]
[monitor:///var/log/secure]
[monitor:///var/log/yum.log]
```

The host declaration sets the host name of the machine, which by default matches the Linux hostname. For example, if your server is named “myserver.local”, data received from that server will be classified in the indexer as being received from that name. If you wish to override this, you may do so using the host declaration.

The sourcetype in the sample above is “linux”, because these files are being sent from a Linux server. You may change this to anything you would like. This is what you will use to classify the received log data for the purposes of searches, reports, alerts, etc.
The remaining configuration specifies the necessary input files. Notice the file names begin with monitor:**//** (with two trailing forward slashes), followed by the full path to the file from root, which includes another forward slash for a total of three forward slashes.

There are many other configuration options available, and for detailed configuration you should consult the extensive Splunk documentation library. More information regarding additional resources and help will be listed in the summary section of this manual.

The next file listed, outputs.conf, usually does not need to be manually edited. This file specifies the location to which the collected data should be sent (i.e. the location of the Splunk indexer). To specify the server location, enter the following command:

/opt/splunkforwarder/bin/splunk add forward-server *servername***:9997**
… replacing ‘*servername*’ with the IP address or FQDN of the Splunk indexer.

Once you have entered the following command, the outputs.conf file will be modified as such:

[tcpout]
defaultGroup = *servername*_9997

[tcpout:*servername*_9997]
server = *servername*:9997

[tcpout-server://*servername*:9997]

This file can be manually modified if so desired.
The last file we will discuss is server.conf. This file usually doesn’t need to be edited, but if you wanted to change the server name of the forwarder you would do so here. By default, this will be the local hostname of the Linux server on which the forwarder has been installed. You can edit the serverName field under the [general] stanza. Changing this field will change the forwarder name Splunk sees as it receives data from this server, even though the server’s hostname will remain unchanged.

[general]
serverName = myserver.local

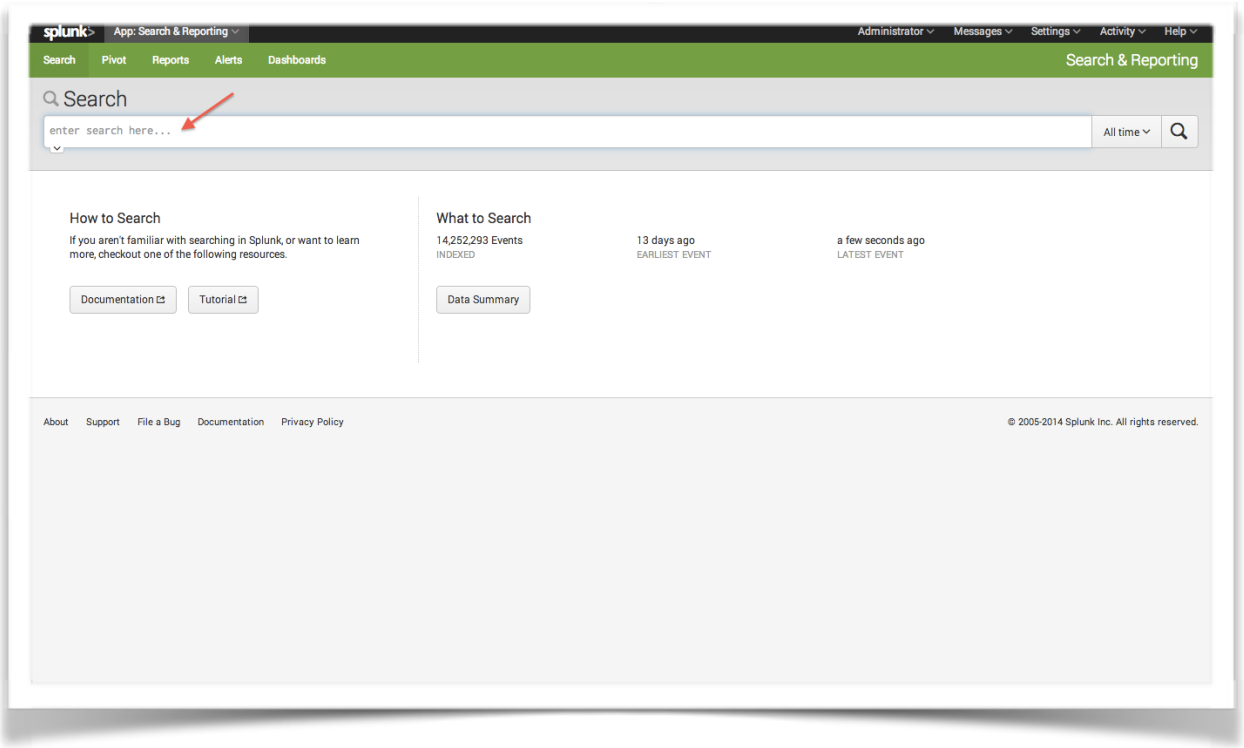
Now that these options have been configured, we must restart the Splunk daemon so the changes can take effect. To do so, enter the following command:

service splunk restart

This concludes the basic configuration of a Splunk Universal Forwarder. You may edit index.conf at any time to add or remove data you send to the Splunk indexer. If you do so, simply restart the Splunk daemon as we did above so the changes can be applied.

Summary:

Congratulations! You now have a basic Splunk installation up and running! Go to the Splunk web interface and try a simple search against the data you are sending to the indexer.



To summarize, here is what we have accomplished:

- Downloaded and installed a Splunk indexer
- Started the daemon and accepted the license agreement
- Enabled boot-start
- Configured iptables to allow inbound connections for Splunk data
- Configured various options including changing the admin password and enabling SSL
- Configured Splunk to receive data via Syslog and forwarders

- Downloaded and installed a Splunk Universal Forwarder
- Changed the admin password
- Edited the inputs.conf file to specify data to be collected and sent to the indexer
- Enabled forwarding to the indexer and examined the purpose of outputs.conf
- Examined the purpose of server.conf

We have only scratched the surface of a very minimal installation and deployment of the indexer and forwarder. It is recommended you use the following resources for additional information as you continue your Splunk education:

Splunk Docs:
http://docs.splunk.com/Documentation/Splunk
The complete documentation collection for Splunk (manuals, tutorials, references, etc.)

Splunk Education Videos:
http://www.splunk.com/view/education-videos/SP-CAAAGB6
Educational videos produced by Splunk on a variety of topics

Exploring Splunk Book (ePub, PDF, Kindle, Hardcopy):
http://www.splunk.com/goto/book
The official Splunk book, available for free download or purchase a hardcopy

Splunk YouTube Channel:
https://www.youtube.com/user/splunkvideos
Splunk's official YouTube channel with marketing videos, product announcements, etc.

Splunk Support:
http://www.splunk.com/support
Splunk's official support portal. Open tech support cases, communicate via IRC, etc.

Splunk Apps:
https://apps.splunk.com/
Extend Splunk by adding apps such as Deployment Monitor, Google Maps, etc.