

What the Splunk?

By Bill Varhol, Security+, CEH, LPT

By now you're probably wondering, "What in the world is Splunk?" Well guess what? I have your answer. In its simplest statement, Splunk is by far one of the coolest log analysis/indexers available! Anything you have that generates logs of some shape, size, or form (and I really mean anything) can be sent to Splunk for indexing and future analysis.

Sure there are other things out there that do something similar, but nothing that indexes every piece of data the way Splunk does. By every piece, I mean you have the ability to search for IP addresses, time, date, web requests, error messages and more! This makes troubleshooting all sorts of different issues a breeze. Don't believe me? Watch the demonstration video of "Search IT" available from the Splunk Website. This is a great example of how powerful Splunk can be.

del.icio.us

Discuss in Forums {mos_smf_discuss:/root}

Introduction

First let's get some quick, technical details out of the way. Splunk can be installed on a variety of operating systems. The reviewed version (3.2.6) supports Linux, Solaris, Mac OS, FreeBSD, AIX, and Windows. Splunk is a high-performance application that has some pretty hefty hardware requirements. The currently recommended configuration for non-Windows platforms is 2 x 3.4GHz processors and 4GB RAM. For Windows systems, multi-core Xeon or equivalent at 3GHz and 4GB RAM is the current recommended configuration. Keep in mind these recommendations are for running Splunk in an active, production environment. If you're looking to use Splunk on a much smaller scale, such as personal use, then you should consider the minimum requirements. The minimum requirements for running Splunk on a non-Windows system is a single 1.4GHz processor and 1GB RAM. For a Windows system, a Pentium 4 or equivalent at 2GHz and 2GB RAM should be sufficient. It is important to note that on the Windows platform, Splunk is only supported on 32-bit Windows systems at this time. Splunk provides 64-bit support for FreeBSD 6.2 and some Linux platforms. For specific details, see the Splunk website.

If you find that you are having trouble with your Splunk environment, you have several different options for seeking support. The first (and often stated in this review) is the Support section of the Splunk website. The website contains all of the needed documentation for installing, administering and using Splunk. There are also deployment and developer manuals available. In the event that you are unable to find the needed answer in the Splunk documentation, there are community support forums and chats (via IRC) available, as well as the ability to submit a case to a support engineer online. You can also get phone help if you have enterprise support. Splunk also offers educational programs for training and professional services for assisting you with design, installation, configuration, customization and integration of Splunk into your network infrastructure.

So what can Splunk do for you?

Splunk has a variety of special features that set it apart from the rest of the IT world. According to the Splunk website, it can help you with operations, security, business intelligence and even compliance. Splunk has 7 core features that make up its fundamentals.

The first is its indexing ability. Splunk can index logs, configurations, traps and alerts, messages, scripts, and code and performance data from all your applications, servers and network devices. Because the data could be coming from virtually any device on your network that you choose, it has very flexible and easy to use input methods. You can send your data to Splunk via common methods such as SNMP or syslog, or over custom TCP/UDP ports that you define. However you get the data to Splunk, it will log and index it for you.

The second core feature (which fits hand-in-hand with indexing) is its search capability. Searching with Splunk is incredibly fast. You can search through 100% of the data that has been fed to Splunk immediately after it has received it. You can even include various search operators like Boolean, nested, quoted and wildcards. Adding or subtracting from your search query is as easy as clicking on some of the returned data.

The next feature of Splunk is its ability to provide alerts. By creating a custom schedule to run searches, you can tell Splunk to notify you if any of the returned data matches the criteria you've provided. Splunk can send you the notifications by RSS, email or SNMP.

The fourth core feature of Splunk is its reporting ability. Again by combining its search feature, Splunk can quickly create incredible summaries of your data with interactive charts, graphs and tables. You really have to try this feature hands-on to really see the powerful reporting capabilities of Splunk.

Next on the list is the Splunk sharing feature. What exactly does this mean? The easiest way to think of it is like a "solutions" database. Each user can add their knowledge to Splunk. Maybe you know that a certain string of code is due to a specific error. You can easily add this into Splunk, so that another user will be able identify what it is they're looking at and also how to solve the problem the next time it comes up.

The sixth core feature of Splunk was mentioned earlier, and that's its ability to scale to your environment. Since the enterprise version of Splunk can send data from one Splunk server to another, you can easily configure multiple Splunk servers in your environment. However you create your network, Splunk can be built to accommodate.

Lastly is the security included with Splunk. The Splunk website summarizes it best, "Splunk gives you secure data handling, fine grain access controls, auditability, assurance of data integrity and integration with existing authentication systems." All of these core features make Splunk what it is today and are the reason Splunk is more than "just another log analyzer."

Windows Install

The installation of Splunk on Windows is very straightforward. You can acquire the .msi installer directly from the Splunk website. After that, it's pretty much "next, next, finish." Let's quickly walk through the standard installation.

After double-clicking the installer and clicking "next" on the initial screen, the first thing you are met with is the License Agreement. Take your time and read through this, then click "Agree" and move forward. You are then asked for your username and organization. The default value of "administrator" for the username is probably fine. Whether you want to fill in your organization is up to you, then click "next." Select your install location (default is C:\program files\splunk) and move on. You are then asked whether you want to use the local system user or another user to run Splunk. If you choose to use a user other than the local system user, you will need to make sure that the account meets the following requirements: ability to run as a service, ability to read whatever files you will be configuring Splunk to monitor and the ability to write to the Splunk directory. Once you've determined the user click "next" and then "install" to begin the Splunk installation.

The installation typically runs very quickly and is done within a minute or two. The last screen of the installer includes some final checkboxes for Splunk configuration. The first is whether you want to run Splunk as soon as the installer is finished. Typically you will leave this checked unless you want to make some changes to the Splunk configuration files first. The next option is whether you want to open a browser to your Splunk. The last option asks if you'd like Splunk to index local events. If so, you can select which events you would like Splunk to pull data from (Application, Security, System). Once you've made your final decisions, go ahead and click "Finish." That's it! That's all there is to it! You now have a working installation of Splunk!

Linux Install

Installing Splunk on Linux is just as simple as installing on Windows. You have three options for downloading/installing Splunk on Linux: RPM, DEB, or tarball. Each can be downloaded from the Splunk website. Be sure to choose the one that best suits your system.

To install via RPM: `rpm -i splunk_package.rpm`

The default directory is `/opt/splunk`

If you'd like to change the install directory, add `--prefix=/opt/new/directory`

To install via DEB: `dpkg -i splunk_package.deb`

The default directory is `/opt/splunk` (this cannot be changed at this time)

To install via tarball: expand into the appropriate directory

The default directory is /opt/splunk

There are a couple of notes to keep in mind when installing with the tarball. The first is that the default user, splunk, is not created automatically. If you would like Splunk to run as a different user, you must create that user manually. The second is to be sure there is enough space on the install partition for the expanded Splunk files.

To start Splunk (and accept the license agreement; must be performed at first start), you can run the following command:

```
$SPLUNK_HOME/bin/splunk start --accept-license
```

After running your appropriate command(s) to install Splunk, it will identify that it has been installed in the proper directory and will advise you on how to start Splunk as well as how to access it via your web browser. Don't forget, the first time you start Splunk you'll have to add the --accept-license to your command to accept the agreement. Once you load Splunk, you'll see it copy some files, validate its databases, verify some pre-requisites and create the security certificates. After that, it's away you go!

Administration and Usage Overview

Before we get into my personal experience with using Splunk, let's go through some standard usage and administration. I can not put enough emphasis on how great the documentation on the Splunk website is for helping you understand how to run Splunk. There are two guides that are very beneficial; the Admin Manual and the User Manual.

At first login you're met with the Getting Started dashboard. From this page, you can link to video demonstrations on the core features of Splunk, quickly access support through methods mentioned earlier or access some of the available add-ons for extra functionality. Towards the upper-right of the page, almost to the corner, there is a drop-down box to switch your dashboard view. From here you can switch to the admin or main view, or even customize your own. If you select the admin view, you will be shown several different charts (or panels) highlighting some important statistics of your Splunk installation. The chart at the top shows our "messages per minute for the past 3 hours." This chart gives us a quick summary of not just how many messages we're receiving, but also from where those messages are coming. Each source of data has its own color in the chart. The second chart is the "KB indexed per hour in the last 24 hours." This title should be pretty self-explanatory, but this tells us how much actual data is coming into our Splunk indexer. The next chart shows us our "Splunk errors for the past 24 hours." This is literally referring to the errors that are generated by Splunk itself. You hopefully should have a very minimal amount of these at this point, and they should not be anything critical. The last chart gives us our "Daily indexing volume by server."

All of the charts have a couple characteristics in common with each other. If you don't like looking at the data in "chart" form, you can easily select the "table" option from the drop-down box to the upper-right of each chart. This will show your data in spreadsheet-like layout. Next, and this gets us into the search feature, each piece of data within the chart can be clicked on to view specifics about that event. Each panel also has a couple of options at the top of its box. In the upper-left corner you'll see a minimize button that will shrink the panel down to "title bar" size and in the upper-right corner you'll see a close button that will remove the panel from your view. It's also important to note that next to the dashboard view drop-down menu you'll see two links: edit and delete. The edit link allows you to select which panels are viewable on the dashboard. The delete link will remove the dashboard view entirely. Now let's take a quick look at the last dashboard type, the Main dashboard. On the Main dashboard we have 3 panels available. The first provides a list of "All indexed data," further separating by source, source type and host. The next panel again displays our "Errors in the last hour," and the bottom panel shows us our "Saved searches."

Sample Report from Splunk Site

[Click for Full Size Image](#)

As I'm sure you're starting to notice, this is a very robust application. I feel that the Splunk staff has done an excellent job with putting together their online tutorial and the video demonstrations of each of the core features of Splunk. For this reason, I'm not going to go too deep into an explanation of each of those things. I'll leave it up to you to explore those videos and go further with them at your own interest. Instead, I'll go over some of the administration pieces of Splunk, and then I'll finish up with my personal experiences in using the product.

No matter which dashboard view you are looking at, the search bar at the top does not move. It is available from any dashboard and several helpful links in the upper-right of the page. From left to right the links are: Admin, Preferences, and Help. You'll also see a little farther to the left a refresh link that will simply refresh the data in your charts or tables on that page. The Help link simply takes you right over to the Splunk website's Resources page, where you can easily access the various forms of support. Clicking on Preferences will give you a few options about viewing the website such as default time range, max search results per page and the default theme. The Admin link is where you'll spend more of your time configuring and setting Splunk up to work the way you want it to.

After clicking on the Admin link you're presented with a new page that should show your server hostname at the top and six different tabs: server, data inputs, distributed, users, saved searches and license & usage. There are a couple of tabs here that are used only with the Enterprise version of Splunk. The first is the Distributed tab. It does exactly what you'd expect – allows your Splunk server to communicate directly with other Splunk servers. Here you can see if your server is receiving data from other Splunk servers, which servers it may be forwarding that data to and also amongst where searches may be distributed (a pretty cool feature!). The second tab is the Users tab. This would allow you to manage multiple logins and access controls. There is another tab listing your Splunk license and server usage.

The tabs that you'll be mainly interested in are the remaining three: Server, Data Inputs, and Saved Searches. The Server tab allows us to control some general Splunk configurations. Here you can change things such as the port Splunk runs on, the datastore location and whether to use SSL for the web application. You can also restart the Splunk server from the Control sub-tab located here. The final sub-tab, Authentication, is another feature that is only available with the Enterprise version of Splunk. Moving on to the Data Inputs tab, this is where you control what gets logged (what Splunk eats). You have a variety of different options available to you for logging purposes. Splunk can read the local events, read from a log file on the local machine, accept syslog messages, listen on custom ports and more! The first sub-tab, Files & Directories, has several logs already being watched for the Splunk server itself. You have different options including using a 'tail' method to watch active log files, uploading a log file or copy and watching. The next option is FIFO. The Splunk website explains this very simply: 'A FIFO (AKA named pipe) is a queue of data maintained in memory. File systems can write log messages directly to a FIFO. Splunk then accesses the FIFO as though it were a file.' The last option, and probably most commonly used, is the Network Ports option. This is where you can configure Splunk to listen on any TCP/UDP port you would like it to receive data on. It can be something common such as syslog or any other custom port you need in order to work with another application. And finally the Saved Searches tab is pretty self-explanatory. You can create, save and schedule your custom searches here.

My Splunk Experience

I would not consider myself to be any sort of Splunk expert, but I did enjoy the time spent with this product while playing around with it a bit. Shortly after registering on the Splunk website to proceed with my download I received an email from

a representative thanking me for trying Splunk and offering help/support if needed. I thought this was a nice touch since it made it a little more personalized rather than the generic "Thanks for downloading our product" - support team. The gentleman asked whether I was evaluating Splunk for an IT initiative and I replied letting him know that I was just testing the product out.

When determining where to install Splunk, I had a couple of different options and pretty much used them all. I ran an install in a (sort of) production environment at work on a Windows computer. I also installed on a Windows VM at home and an Ubuntu VM on my laptop. I primarily used the computer at work the most, and it's nothing special; Server 2003 R2, AMD 64 X2 @ 2GHz, 1024MB RAM. However, I also am limited to what I could throw at it, since I can only point to it what's within my control. I ended up logging the local event logs, grabbing VPN logs via a share, and pointing our firewall to it using syslog. Even though I had limited data, I still was able to get a good understanding of the program. One of our network engineers has also been playing around with it a little as well, but I haven't received any feedback from him yet.

Installing Splunk took me a couple of tries due to the confusing documentation about installing as local system as opposed to another user on a Windows machine. Even when clicking on the option for another user, and then selecting back to choose local system, I found that I had errors with the install. Finally I just chose to leave it as the system user, and I had no problems installing past that point.

Once Splunk was installed I tried opening it through my browser and found that it said "checking for updates." After approximately 40 minutes of sitting at this screen I decided to boot up Wireshark to see what was going on. There was no traffic. At this point I started digging around the documentation to determine how to disable the auto-update feature. The answer lied in changing one of the configuration files, pretty simple fix. At this point, I had spent longer than I wanted just getting Splunk installed and called it quits for the night.

My next login involved watching all of the available video demonstrations to see what Splunk could do. I felt that I had a pretty good understanding of its abilities after watching the videos and started to play around with the web interface. I became familiar with the available options and configurations, the charts and tables and how to conduct searches. I only had one question left - how to get data to Splunk. I saw the enterprise only option of connecting multiple Splunk servers and thought that would be the easiest route, but what if you don't purchase the enterprise license. I started digging around the Splunk documentation on the website, and this is also when I thought I'd try out the chat (IRC) room as well. It's worth noting at this point that, in my opinion, some of the Splunk documentation is lacking for Windows support. A large majority of the examples within the documentation are focused on the *nix directory structure and commands. It was a little difficult at times trying to find an answer for a Windows system, but not too big a deal. Once within the IRC channel, I saw about 5-7 operators and several guests. I asked my question about sending data to Splunk via network ports and within 10 minutes or so I received a reply from one of the operators. They gave me a few quick examples and suggestions, and I was then on my way.

Once I had Splunk configured to watch the logs I was feeding it, there wasn't much else to do except sit back a little bit and let it collect some data. So throughout the rest of my evaluation of Splunk I'd login to play around with its searching ability, check out the cool looking graphs and generate a report or two. I even used it a couple times to assist with troubleshooting some VPN issues!

Conclusions

Overall I think that Splunk is a very cool application. I personally don't have enough logs and such that I monitor frequently enough for Splunk to really have a strong advantage, but that's just me. I definitely see how this could significantly help someone out that has many logs they need to keep track of for analysis and troubleshooting. Splunk

really does a great job of indexing all of the data that it receives and works on a variety of platforms (in addition to having incredible scalability) making it an ideal solution for many different environments.

Bill Varhol has been working with computers for over 8 years and started out taking CompTIA A+ and Cisco CCNA courses in high school. Shortly after graduation, he got his start in technology working as a credit card terminal technician for a Fortune 100 card processor. He later found temporary contract work for a non-profit credit and debt counseling company in Michigan, where he was eventually brought on full-time. He is still with the same company working as a Network Administrator where he participates in a variety of security initiatives and projects. With a strong sense of giving back, Bill is also a member of Hackers For Charity and his local Citizen Corps CERT. Bill has earned his CEH, ECSCA, LPT and Security+ certifications. He is currently enrolled in the SANS 560/GPEN course and plans to sit for the CISSP exam in December.