Splunk - Roadmap to Operational Information that Leads to Valuable Insights

Shivani Grover, M. Tech., Computer Science, MIET Mohri, Shahbad Distt, Kurukshetra, Haryana
Parbhat Verma, Assistant Professor, Computer Science, MIET Mohri, Shahbad Distt, Kurukshetra, Haryana

ABSTRACT

There is tremendous growth in the amount of data generated in the world as most modern systems generate abundant and diverse log data. With increasing cost of storage, there are few reasons to summarize or discard data. Traditional data analysis, management and monitoring solutions are simply not engineered for this high volume and highly diverse data. In this paper, we present Splunk, a powerful platform for analyzing machine data, data that machine emits in great volumes but which is seldom used effectively. We share observations, lessons and case studies from real data sets and exhibit Splunk’s power which provides unique visibility across entire IT infrastructure.

Keywords
Operational Intelligence, Splunk, Reports, Dashboards

1. INTRODUCTION

Mining a massive dataset is a non-trivial but a more challenging task is to cross-correlate and mine multiple datasets from various sources. Analyzing data from many disparate sources is challenging enough when attempting to handle and incident. One of the tools used for this is Splunk. Over 1850 enterprises, service providers and government organizations in 68 countries use Splunk to realize new levels of service quality. Splunk is the leading platform for real time operational intelligence. Splunk was founded to pursue disruptive new vision: make machine data accessible, usable and valuable to everyone.

2. SPLUNK OVERVIEW

Splunk enables you to search, report, monitor and analyse all real-time streaming and historical IT data. Now we can troubleshoot application problems and investigate security incidents in minutes instead of hours or days. Splunk software indexes data from any application, operating system, server or network device, enabling us to search and analyse billions of events across IT infrastructure from one location in real time. Splunk allows almost instantaneous visibility into every element of the enterprise datacenter.

3. SEVERAL KEY POINTS ABOUT SPLUNK

- Creating a central repository is vital
  One of the major victories of Splunk is the way that diverse types of data from many different sources are centralized for searching.

- Splunk helps you understand the structure and meaning of data
  Splunk also helps to capture what we learn to make future investigations easier.

- Helps understanding the behavior of the people browsing a website.
  Splunk not only helps to diagnose systems but also to better understand the behavior of people.

- Splunk converts data into answers
  Splunk helps us find the insights that are buried in the data.

4. CRITICAL USES OF SPLUNK ACROSS IT AND BUSINESS

- Application Management
  Application Management provides end-to-end visibility across distributed infrastructures; troubleshoot across application environments; monitor for performance degradation; trace transactions across distributed systems and infrastructure.

- Security and compliance
  It provides rapid incident response, real time correlation and in-depth monitoring across data sources; conduct statistical analysis for advance pattern detection and threat defense.

Fig.1. Splunk use case areas
5. OPERATIONAL INTELLIGENCE AT WORK

Splunk does something that no other product can: efficiently capture and analyze massive amounts of unstructured, time-series textual machine data. Although IT departments generally start out using Splunk to solve technically esoteric problems, they quickly gain insights valuable elsewhere in their business.

Using machine data in Splunk helps solve vexing business problems.

Here are a few examples:

- An operations team implemented a cloud-delivered customer-facing application and used Splunk for diagnostics. They soon realized they could track user statistics and better plan capacity—a metric with profound business implications.
- Web server traffic logs can be used to track shopping carts being filled and abandoned in real time. The marketing department can use this information to determine where consumers are getting stuck and what types of purchases are being abandoned so that any problems can be fixed right away and promotions can target items that are abandoned.
- Organizations using Splunk to monitor applications for troubleshooting have realized that they can easily provide views to their first-line support team to handle customer calls directly, versus escalating those calls to expensive engineering resources.
- A major utility company was able to eliminate costly software maintenance fees by replacing six other monitoring and diagnostic tools withSplunk, while enhancing their NERC and SOX compliance efforts.
- A major public media organization reduced the time it took to capture critical web analytics from months to hours. They were also able to track their digital assets with a granularity and accuracy that they couldn’t have otherwise, resulting in better royalty accounting and content marketing.
- Ultimately, operational intelligence enables organizations to ask the right questions, leading to answers that deliver business insights, using combinations of real-time and historical data, displayed in easily digestible dashboards and graphical tools. There’s a reason for the trend toward calling machine data “big data.” It’s big, it’s messy, and in there, buried somewhere, is the key to the future of your business.

6. PROPOSED METHODOLOGY

Splunk does something that no other product can: efficiently capture and analyze massive amounts of unstructured, time-series textual machine data. Although

IT departments generally start out using Splunk to solve technically esoteric problems, they quickly gain insights valuable elsewhere in their business. The proposed methodology covers the algorithm used for implementing the concepts of Operational Intelligence using Splunk. Algorithms are discussed as follows:

A. Implementing OI using Splunk

The algorithm depicts that how zeta bytes of data are made useful after indexing into Splunk. The following algorithms show the process of adding data into indexer by either directly uploading into indexer or after installing universal forwarders on two different types of machines. The forwarders send data to the Splunk server.

a. Algorithm for Splunk data Indexer

This specifically states the process to add data into main Splunk indexer.

- There are multiple options to add data into the Splunk. Click Add Data on the home screen.
- Select a data source or source type to Splunk the data.
- If it is OS related data select an appropriate source or source type click next to select the OS parameters to be indexed.
- If we need to read the data specifically from log files then select files & folder at source or source.
- Take a preview or skip preview to validate the file which we have uploaded is the appropriate one to be uploaded.
- Click next and data will be indexed in the Splunk

b. Algorithm for Splunk data Indexer

This specifically states the process to create Dashboards out of data indexed.

- Run a search that generates a report for a dashboard.
- Select Create » Dashboard panel.
- Give your search a name, and click Next.
- Decide if you want this report to go on a new dashboard or on an existing dashboard. If you’re creating a new dashboard, give it a name. Click Next.
- Specify a title for your dashboard and a visualization (table, bar, pie, gauge, etc.), and when you want the report for the panel to run (whenever the dashboard is displayed or on a fixed schedule).
- Click next followed by the View dashboard link or OK.

7. RESULTS

In real time we deal with machine data and this data is generated in terms of log files. There are thousands of such files generated on day to day basis getting achieved without putting any add-ons. Our research provided a way to fetch values out of complex data in really simpler way.
The next figure describes the sample log file before implementing concepts of Operational Intelligence in Splunk.

The next figure describes a screen capture from Splunk. I tried to find a word “error” in the above log file using Splunk & the result provides the complete details whenever error is matched. We can use these details to take intelligent decisions. These details include: the time of error, trace of error, related events.

8. CUSTOM DASHBOARD AND VIEWS

Make more sense of the huge volumes of data at your disposal. Create custom dashboards and views for different types of users, technical and non-technical. Integrate reports, search results and even data from external applications. Edit dashboards using a simple drag-and-drop interface; integrated charting controls mean you can change chart types on the fly. Doing this all through the Splunk UI means that you can empower business users to do the same.

Example: Representing error faced by users for CreditDoesotMatch at the time making payment:

9. CONCLUSION AND FUTURE SCOPE

To make data more useable, add knowledge to it. What do I mean by that?

When I say Splunk how to extract fields from the data, we can start reasoning about those fields and give Splunk the knowledge to classify the data for deeper analysis.

When we save reports and dashboards, the data becomes easier to understand for everyone and after creating alerts, Splunk proactively reveals potential issues so that we don’t have to look for them manually after the fact. This paper has covered three areas:

- Need of operational intelligence in real time scenarios.
- Using Splunk to understand Data shows how to explore, categorize, and become familiar with our data.
- Displaying Data shows the basics of visualizing data.

It’s often said that we can’t prove a negative. We can’t look everywhere and say, what I seek is not there. With Splunk you can do negative searching and in fact you should. The reason it’s hard to see what’s happening with
log files, and many other types of data, is that so much of it is the same, sort of business as usual machine data. With Splunk we can categorize that uninteresting data and tell Splunk to show us only what’s unusual or different. Show me what I haven’t seen before. Some security experts use Splunk in just this way to identify anomalous events that could indicate an intrusion, for example. If they’ve seen it before, they give it a tag and exclude it from their search. After we do this for a while, if anything odd happens, we’ll see it right away.

10. ACKNOWLEDGEMENT

The author Shivani Grover would like to thank co-author Parbhat Verma for his support of, and contribution to, this work.

REFERENCES


AUTHOR PROFILE

Shivani Grover is pursuing masters of technology in computer science and technology from MIET Mohri, Shahbad. She has developed her interest in researching on upcoming and current market trends. She has selected “Operational Intelligence” an emerging scenario and must required area in IT, as a part of dissertation for M-Tech.

Parbhat Verma is an assistant professor working in MIET Mohri, Shahbad. He is also pursuing Ph.D. He has extensive experience in the field of education.