VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI - 590018

A TECHNICAL SEMINAR REPORT

On

"SPLUNK"

Submitted in the partial fulfilment of the Requirements for the award of the Degree

B.E. in Computer Science and Engineering

Submitted by

AHISH D **4BD16CS003**

Under the Guidance of

Prof. Waseem Khan M.Tech.,
Assistant Professor,
Department of CS&E,
B.I.E.T., Davangere

Seminar Coordinator

Prof. Raghu B R M. Tech.

Head of the Department

Dr. Nirmala C R Ph.D.

Department of Computer Science and Engineering

Bapuji Institute of Engineering and Technology

Davangere- 577004

June- 2020

Bapuji Institute of Engineering and Technology
Davangere- 577004

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that the technical seminar entitled **"SPLUNK"** is a bonafide work carried out by **AHISH D** bearing USN **4BD16CS003** has successfully conducted the technical seminar in the partial fulfilment of the award of the degree of the Bachelor of Engineering in Computer Science and Engineering of the Visvesvaraya Technological University, Belagavi for the academic year 2019-2020.

SEMINAR GUIDE

Prof. Waseem Khan M. Tech

SEMINAR COORDINATOR

Prof. Raghu B R M. Tech

HEAD OF THE DEPARTMENT

Dr. Nirmala C R Ph. D

ACKNOWLEDGMENT

Salutations to my beloved and highly esteemed institute, "**BAPUJI INSTITUTE OF ENGINEERING AND TECHNOLOGY**" for having well-qualified staff and labs furnished with the necessary equipment.

I express my sincere thanks to my guide **Prof. Waseem Khan** for giving me constant encouragement and support for the seminar without whose stable guidance this seminar would not have been completed.

I express my special thanks to the Seminar Coordinator **Prof. Raghu B R** for giving me valuable suggestions and support throughout the course of the seminar.

I express my whole hearted gratitude to **Dr. Nirmala C R** who is our respectable H.O.D of Computer Science and Engineering Department. I wish to acknowledge her who made my task easy by providing with her valuable help and encouragement.

I also express my whole hearted gratitude to our respected Principal, **Dr. Aravinda H B** for his moral support and encouragement.

I would like to extend my gratitude to all staff of **Department of Computer Science and Engineering** for the help and support rendered to me. I have benefited a lot from the feedback, suggestions given by them.

I would like to extend my gratitude to all my family members and friends especially for their advice and moral support.

AHISH D (4BD16CS003)

ABSTRACT

Handling a huge amount of data is one of the biggest challenges, as there is a rapid development in the IT sector and its machines. Splunk is a tool that is used to analyze machine data. It is a software mainly used for searching, monitoring, and examining machine-generated Big Data in real time through a web-style interface. It can monitor and read different type log files and stores data as events in indexers. Splunk allows you to visualize data in forms of dashboards. Splunk turns data into business outcomes and this powerful platform and unique approach to data have empowered companies to improve service levels, reduce operation costs, mitigate risks, enhance DevOps collaboration etc.

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction to Splunk

Splunk is an advanced, scalable, and effective technology that indexes and searches log files stored in a system. It analyzes the machine-generated data to provide operational intelligence. The main advantage of using Splunk is that it does not need any database to store its data, as it extensively makes use of its indexes to store the data. This tool allows you to visualize data in various forms of dashboards.

Splunk is a software mainly used for searching, monitoring, and examining machine-generated Big Data through a web-style interface. Splunk performs capturing, indexing, and correlating the real-time data in a searchable container from which it can produce graphs, reports, alerts, dashboards, and visualizations. It aims to build machine-generated data available over an organization and is able to recognize data patterns, produce metrics, diagnose problems, and grant intelligence for business operation purposes. Splunk is a technology used for application management, security, and compliance, as well as business and web analytics.

With the help of Splunk software, searching for a particular data in a bunch of complex data is easy. As you might know, in the log files, figuring out which configuration is currently running is challenging. To make this easier, there is a tool in Splunk software which helps the user detect the configuration file problems and see the current configurations that are being utilized.

1.1.1 Why Splunk?

Splunk is a digitized platform that assists in accessing machine-generated data, which will be useful and worthwhile for everyone. Handling a huge amount of data is one of the biggest challenges, as there is a rapid development in the IT sector and its machines. In this situation, Splunk plays a vital role to deal with the situation.

1.1.2 Brief History of Splunk

Rob Das and Eric Swan co-founded this technology in the year 2003 as a solution to all the queries raised while examining the information caves faced by most of the companies. The name ‘Splunk’ is derived from the word ‘spelunking,’ which means exploring the information caves. It was developed as a search engine for the log files that are stored in the infrastructure of a system.

The first version of Splunk was launched in 2004 which was largely appreciated by its end-users. Slowly and gradually, it became viral among most of the companies, and they started to buy

its enterprise licenses. The main goal of the founders is to market this developing technology in bulk so that it can be deployed in almost all kinds of use cases possible.

1.2 Splunk Products

Splunk is available in three different versions.

- Splunk Enterprise
- Splunk Light
- Splunk Cloud

1.2.1 Splunk Enterprise

Splunk Enterprise edition is used by large IT business. It helps you to gather and analyze the data from applications, websites, applications, etc.

1.2.2 Splunk Cloud

Splunk Cloud is a hosted platform. It has the same features as the enterprise version. It can be availed from Splunk or using AWS cloud platform.

1.2.3 Splunk Light

Splunk Light is a free version. It allows search, report and alter your log data. It has limited functionalities and feature compared to other versions.

1.3 Log File

A log file is a file that keeps a registry of events, processes, messages and communication between various communicating software applications and the operating system. Log files are present in executable software, operating systems and programs whereby all the messages and process details are recorded. Every executable file produces a log file where all activities are noted.

The phenomenon of keeping a log is called logging, whereas a record file itself is called a log file. The most commonly used logging standard is syslog, which is short for "system log." The log messages in a log file can be recorded and analyzed later, even after the program has been closed.

1.3.1 Why Explore Logs?

Logs are the go-to-achieves for gaining company-wide Operational Intelligence.

1. **Server Logs**

- Source of IP traffic

- Security threats
- Network vulnerability
- Network traffic and spike

2. System Logs

- System performance
- CPU usage & load
- User access logs
- App performance monitoring

1.3.2 Use of Splunk in Exploring Logs

Splunk is the ultimate log collection and analysis tool.

- Real-time log forwarding
- Real-time syslog analysis
- Real-time server monitoring
- Real-time alerts/notifications
- Historical data/log store and analysis

CHAPTER 2

SPLUNK ARCHITECTURE

2.1 Components of Splunk Architecture

- **Universal Forwarder (UF):** It is a lightweight element that assists in pushing the data to the heavy Splunk forwarder. The principal task of this element is to just forward the log data from the server. You can easily install Universal Forward at the client-side or on the application side.
- **Load Balancer (LB):** In computing terms, Load balancing enhances the distribution of workloads over multiple computing resources. A load balancer is an element that distributes the network or the application traffic over a cluster of servers.
- **Heavy Forwarder (HF):** It is recognized to be the heavy element. This Splunk component enables you to filter the data. For instance, it will help in accumulating only the error logs.
- **Indexer:** The chief task of an indexer is to store and index the filtered data. It helps in improving Splunk's performance. By default, Splunk automatically implements the indexing like hosts, sources, date, and time.
- **Search Head (SH):** It is simply a Splunk instance that helps in distributing the searches to the other indexers, and it normally doesn't have any instance of its own. It is essentially used to achieve intelligence and perform reporting.
- **Deployment Server (DS):** It helps in deploying the configuration like updating the UF (Universal Forwarder) configuration file. You can use a DS to share data between the components.
- **License Master (LM):** A license slave is a Splunk Enterprise state which is controlled by a License Master. If you have a single Splunk Enterprise instance, it assists as its License Manager (once you have installed an Enterprise license on it). The license is based on quantity and usage. For example, for 50 GB per day usage, Splunk examines the licensing details daily.

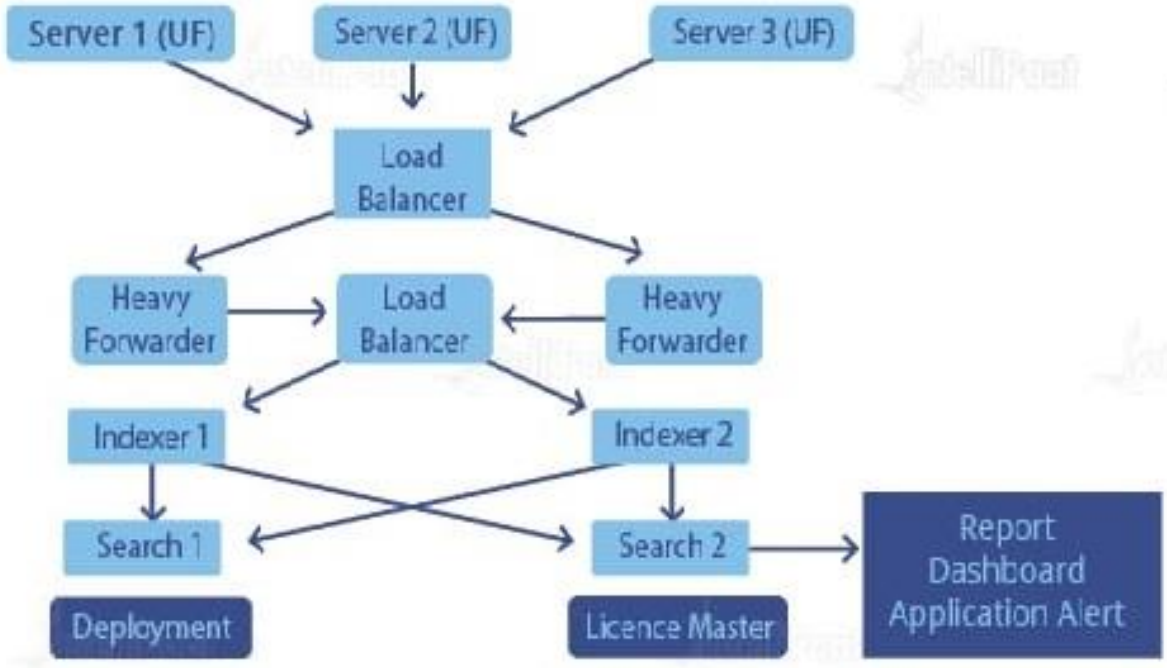

Fig 2.1: Splunk Architecture

Figure 2.1 shows the components of a Splunk Architecture.

2.2 Working of Splunk

Fig 2.2: Working of Splunk

Figure 2.2 shows how a Splunk tool works.

- Forwarder: It assists in collecting the data from the primitive machines, then it forwards the data to the indexer in real-time.
- Indexer: It helps in processing the incoming data in real-time. It also collects and arranges the data on the disk.
- Search Head: With the help of Search Head, end-users can interact with Splunk. It enables users to perform the search, analyze, and visualize functions.

In detail working of Splunk:

- The forwarder can track the data, make a copy of the data and can perform load balancing on that particular data before it sends it to the indexer.
- Cloning can help in producing duplicated copies of any case at the data source whereas load balancing is performed so that even if one case collapses, that data can be carried to another case which is hosting the indexer.
- When the data is obtained from the forwarder, it is then dropped in an Indexer component. In the Indexer, the obtained data is then split into various logical datastores and at every datastore, you can set authorities which will then guide the user's views and accesses.
- When the data is inside the Indexer, you can explore that data and assign those explorations to different search companions and all the results that we will be getting after assigning will be merged and carried forward to the Search Head.
- You can also perform scheduling the search companions and creating the alerts, which will be then activated when some situations will match the saved searches.
- You can also use the knowledge objects only to intensify the existing unstructured data (data which do not have any format).
- The search heads and knowledge objects can be retrieved from a Splunk CLI or a Splunk Web Interface. This interaction happens over a REST API connection.

CHAPTER 3

FEATURES OF SPLUNK

3.1 Features of Splunk

- **Powerful analytics**

Grants faster and easier analysis and visualizations for business users.

- **Intuitive user experience**

Has improved the users' productivity by enabling instant access to relevant apps and content. It is a great productivity feature for end-users.

- **Simplified management**

Produces simplified and scalable management for enterprise Splunk deployment.

- **Rich developer environment**

Helps in rapidly build Splunk apps with the help of approved web languages and frameworks.

3.2 Benefits of Splunk

- Offers enhanced GUI and real-time visibility in a dashboard.
- It reduces troubleshooting and resolving time by offering instant results.
- It is a best-suited tool for root cause analysis.
- Splunk allows you to generate graphs, alerts, and dashboards.
- You can easily search and investigate specific results using Splunk.
- It allows you to troubleshoot any condition of failure for improved performance.
- Helps you to monitor any business metrics and make an informed decision.
- Splunk allows you to incorporate Artificial Intelligence into your data strategy.
- Allows you to gather useful Operational Intelligence from your machine data.
- Summarizing and collecting valuable information from different logs
- Splunk allows you to accept any data type like .csv, json, log formats, etc.
- Offers most powerful search analysis, and visualization capabilities to empower users of all types.
- Allows you to create a central repository for searching Splunk data from various sources.

CHAPTER 4

CASE STUDY

4.1 Splunk Use Case: Domino's Pizza

You might be aware that Domino's Pizza is an e-commerce cum fast food giant, but you might be unaware of the big data challenge they were facing. They wanted to understand their customers' needs and cater to them more effectively by using Big Data. This is where Splunk came to the rescue.

Look at the figure 4.1 below which depicts the circumstances that were building up to cause big data problems at Domino's.

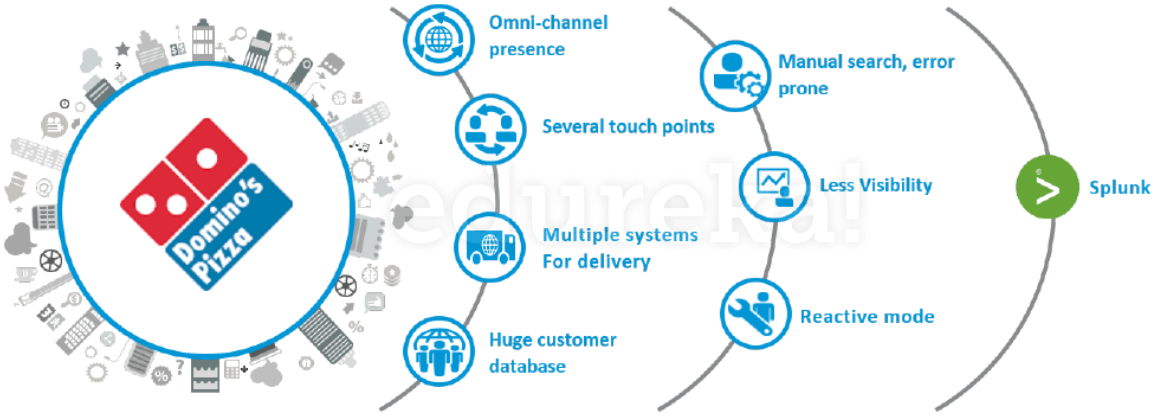

Fig 4.1: Big data problems at Domino's

Lot of unstructured data was generated because:

- They had an omni-channel presence for driving sales
- They had a huge customer base
- They had several touch points for customer service
- They provided multiple systems for delivery: Order food in-store, order via telephone, via their website and through cross-platform mobile applications
- They upgraded their mobile apps with a new tool to support 'voice ordering' and enable tracking of their orders

The excess data generated gave rise to the following problems:

- Manual searches being tedious and error prone
- Less visibility into how customer need/preference varies

- Unpreparedness and thus working in reactive mode to fix any problem

Domino's felt that the solution to these problems would lie in a tool which can easily process data. That was when they implemented Splunk.

Using Splunk for Operational Intelligence in place of a traditional APM tool helped them to lower the cost, search the data faster, monitor performance and get better insights into how customers were interacting with Domino's. If you look at the below figure 4.2, you will find the different applications that were set up by implementing Splunk.

Fig 4.2: Applications set up by implementing Splunk

- Interactive Maps, for showing orders in real time coming from all across US. This brought employee satisfaction and motivation
- Real time feedback, for employees to constantly see what customers are saying and understand their expectations
- Dashboard, used to keep scores and set targets, compare their performance with previous weeks/ months and against other stores
- Payment Process, for analyzing the speeds of different payment modes and identifying error free payment modes
- Promotional Support, for identifying how various promotional offers are impacting in real-time. Before implementing Splunk, the same task used to take an entire day
- Performance Monitoring, to monitor the performance of Domino's in-house developed point of sales systems

Splunk proved to be so beneficial to Domino's that teams outside the IT department started exploring the possibility to use Splunk for gaining insights from their data.

Splunk For Promotional Data Insights

Domino's had no clear visibility into which offer works best – in terms of:

- Offer type (Whether their customers preferred a 10% discount or a flat $2 discount?)
- Cultural differences at a regional level (Do cultural differences play a role in offer choice?)
- Device used for buying products (Do devices used for ordering play a role in offer choices?)
- Time of Purchase (What is the best time for the order to be live?)
- Order revenue (Will offer response change wrt to order revenue size?)

As you can see from the below figure 4.3, promotional data was collected from mobile devices, websites and various outlets of Domino's Pizza(using Splunk Forwarders) and sent to a central location(Splunk Indexers).

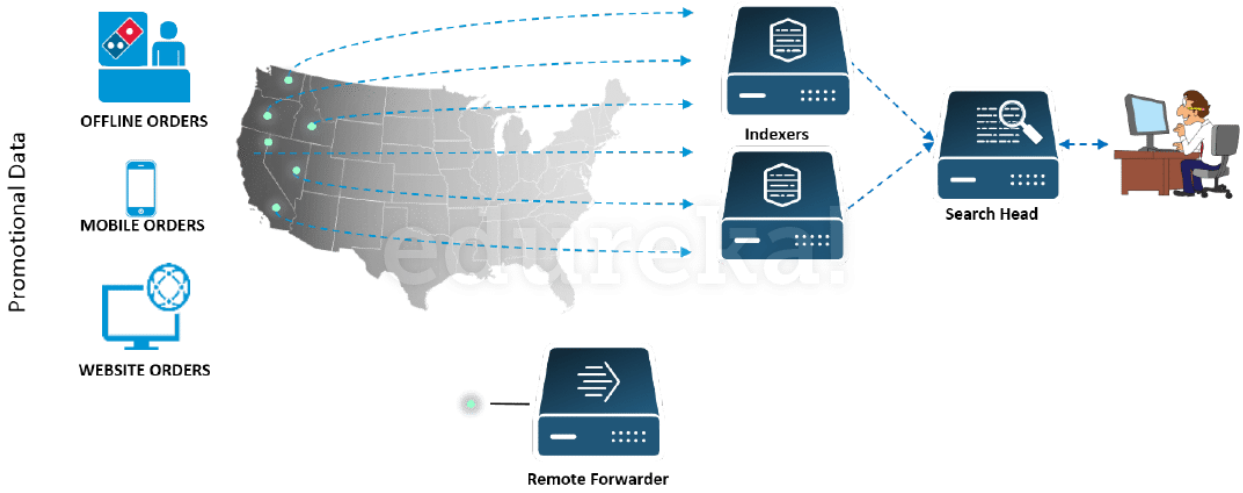

Fig 4.3: Promotional data collection using Splunk forwarders and indexers

Splunk forwarders, would send the promotional data generated in real time. This data contained information about how customers responded when they were given offers, along with other variables like demographics, timestamp, order revenue size and device used.

Customers were divided into two sets for A/B Testing. Each set was given a different offer: 10% discount offer and flat $2 offer. Their response was analyzed to determine which offer was preferred by the customers.

The data also contained the time when customers responded and if they would prefer to buy in-store or do they prefer to order online. If they did it online, then the device they used to make the purchase was also included. Most importantly, it contained Order revenue data – to understand if offer response changes with the order revenue size.

Once the raw data was forwarded, Splunk Indexer was configured to extract the relevant information and store it locally. Relevant information being the customers who responded to offers, time at which they responded and the device used for redeeming the coupons/offers.

Typically, the below information was stored:

- Order revenue based on customer response
- Time of purchase of products
- Device preferred by customers for placing the order
- Coupons / Offers used
- Sales numbers based on Geography

For performing various operations on the Indexed data, Search head was used. It is the component which gives a graphical interface for searching, analyzing and visualizing the data stored in the Indexers. Domino's Pizza gained the below [figure 4.4] insights by using the visualization dashboards provided by the Search head:

Fig 4.4: Visualizations dashboards created by the Splunk Search Head

- In USA and Europe, customers preferred a 10% discount instead of a $2 offer. Whereas in India, customers were more inclined to a flat $2 offer
- 10% discount coupons were used more when the order revenue size was large, whereas flat $2 coupons were used more when order revenue size was small.
- Mobile apps were the preferred device for ordering during the evening and orders coming in from the website was most during the noon. Whereas ordering-in-store was highest during the morning

Domino's Pizza collated these results to customize the offers/coupons with respect to order revenue sizes for customers from a particular geography. They also determined which was the best time to give offers/coupons and targeted the customers based on the device they were using.

CHAPTER 5

IMPLEMENTATION

5.1 Installing Splunk Enterprise on Windows

1. Download the Splunk installer from the Splunk download page.
2. To start the installer, double-click the `splunk.msi` file. The installer runs and displays the **Splunk Enterprise Installer** [figure 5.1] panel.

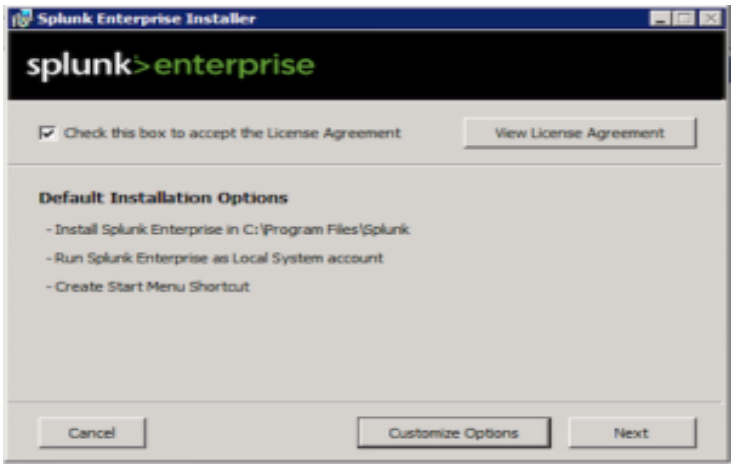

Fig 5.1: Splunk Enterprise Installer

3. To continue the installation, check the "Check this box to accept the License Agreement" checkbox. This activates the "Customize Installation" and "Next" buttons.

5.2 Windows Event logs using Splunk

5.2.1 Splunk Sign-in page

Fig 5.2: Splunk Sign-in Page

Figure 5.2 shows the sign-in page for Splunk Enterprise.

5.2.2 Splunk Home Page

Splunk Home is your interactive portal to the data and apps in your Splunk deployment. The first time you log into your Splunk deployment, you land in Splunk Home. All of your apps appear on this page. Figure 5.3 shows the Splunk Home Page. The main parts of Splunk Home include the navigation bar, the Apps menu, the Explore panel.

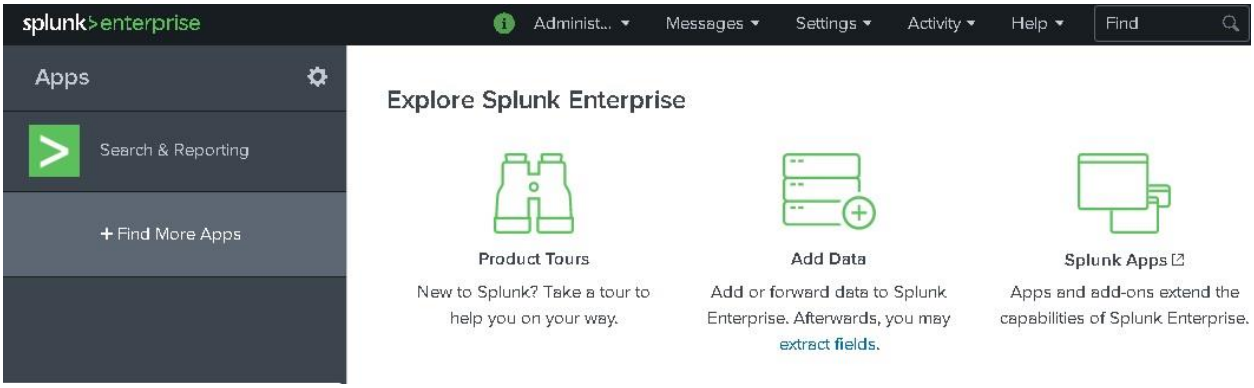

Fig 5.3: Splunk Home Page

5.2.3 Splunk Settings

The Settings menu [figure 5.4] lists the configuration pages for Knowledge objects, Distributed environment settings, System and licensing, Data, and Authentication settings.

Fig 5.4: Splunk Settings

5.2.4 Data Input from local event Log Collection

Windows event Log files are provided as input to the Splunk.

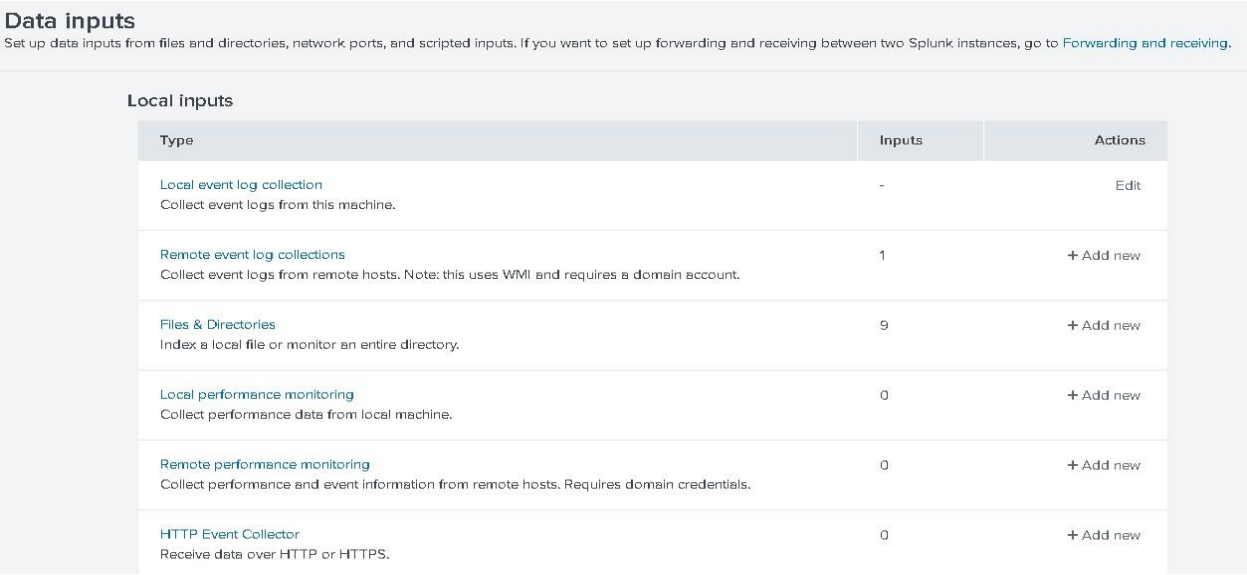

Fig 5.5: Data Input from Windows event files

Figure 5.5 shows the data input window to the Splunk from Windows event files.

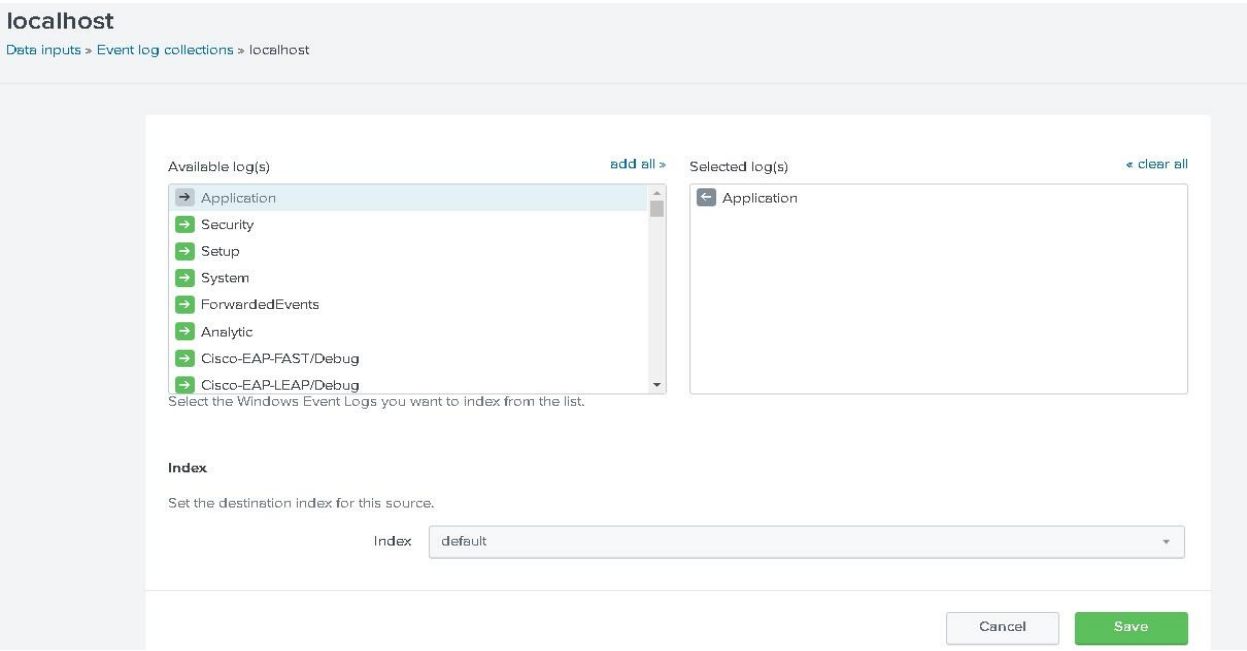

Fig 5.6: Log selection from Windows Event logs

Figure 5.6 shows the log selection window from Windows Event logs.

5.2.5 Searching a keyword and retrieving the result

Enter a query or simply a keyword in Search bar [figure 5.7] which will produce results depending on the time range provided. If it displays “No result found”, then simply change the time range or try searching a different keyword.

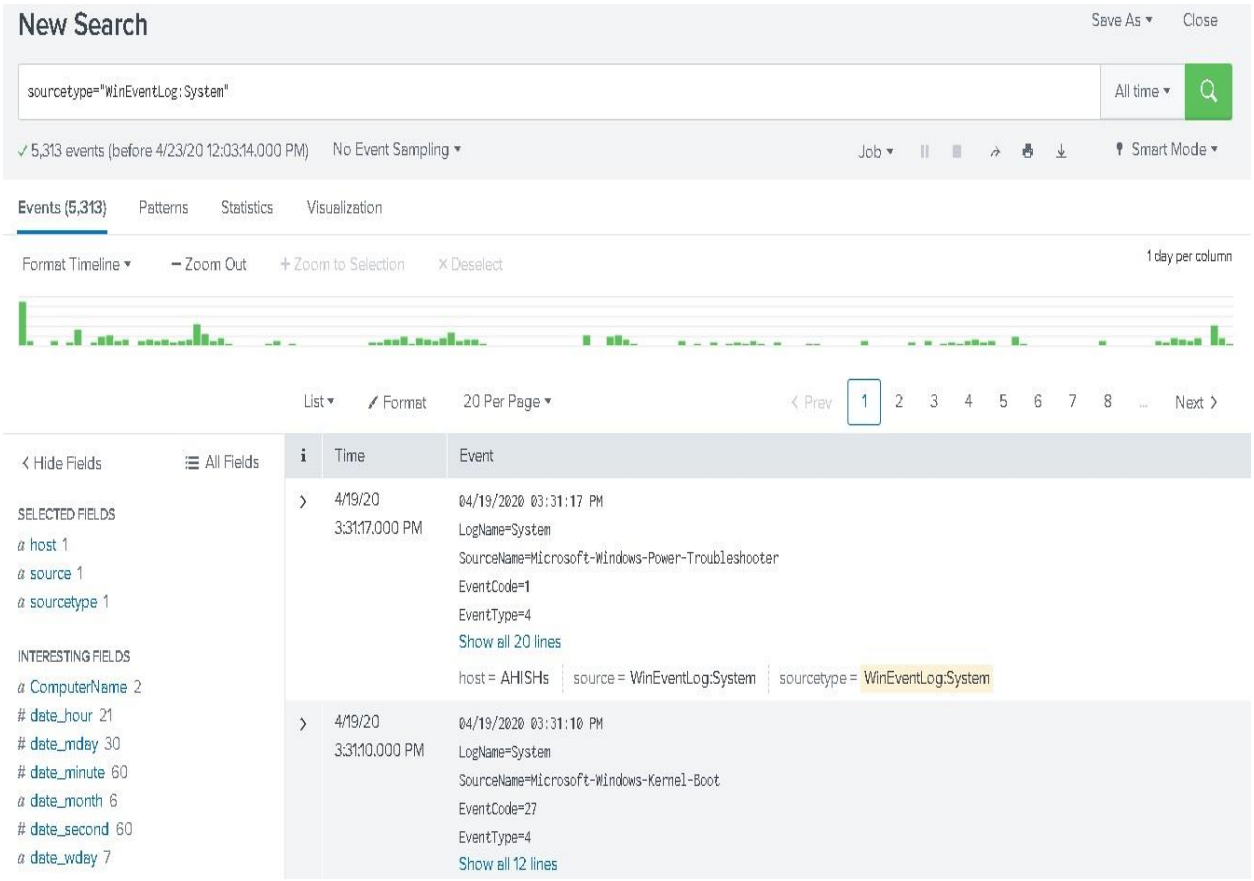

Fig 5.7: Searching and retrieving the required data

CONCLUSION

Today, in the world of machine data, Splunk has become one of the most in-demand tools for Big data professionals. In Big data, there can be various data sources and it can be either in structured or unstructured form, so Splunk like tools help the professionals to drag most important information even from the raw or unstructured data. The company, Splunk is researching and launching new tools and features to make the application powerful. For data-driven organizations, it can be a more profitable and efficient tool.

BIBLIOGRAPHY

[1] "Splunk," Splunk. [Online]. Available: https://www.splunk.com/. [Accessed: 16Apr-2020]

[2] "Splunk Documentation," Splunk. [Online]. Available: https://docs.splunk.com/Documentation/. [Accessed: 14Apr-2020]

[3] "Splunk," Splunk. [Online]. Available: https://intellipaat.com/blog/what-is-splunk/. [Accessed: 12Apr-2020]

[4] "Splunk," Splunk. [Online]. Available: https://guru99.com/splunk-tutorial.html/. [Accessed: 12Apr-2020]

[5] K. Sankari, R. Lavanya, S. Amalagracy, "Real Time Monitoring System using Splunk", IJCSMC, Vol. 4, Issue. 3, pg.434 – 441, March 2015.