**APPLICATION MONITORING AND LOG ANALYSIS USING SPLUNK**

**ABSTRACT**

We present an in-depth study of over 500K log analysis queries from Splunk, a platform for data analytics. Using these queries, we quantitatively describe log analysis behavior to inform the design of analysis tools. The work’s purview was to go through over 500K logs analysis queries using Splunk, a platform for data analysis. We measure, create dashboards on basis of various error scenarios in the system and simultaneously measure the hardware utilization of the tomcat servers. We find that log analysis primarily involves filtering, reformatting, and summarizing data and that non-technical users increasingly need data from logs to drive their decision making.

**Keywords**: log analysis, query logs, user modeling, Splunk, user surveys

**TABLE OF CONTENTS**

**S.No. Contents Page No**

1. **Introduction 1**
   1. Why Splunk? 1
   2. Aim 1
   3. Objectives 1
   4. Sub Objectives 2
   5. Process Chart Legend 2
2. **System Analysis 3**
   1. Existing System 3
   2. Proposed System 3
   3. Modules 3

2.4.1. Log Configuration 3

2.4.2. Log Input To Splunk Enterprise 3

2.4.3. Log Searching In Slunk Enterprise 3

2.4.4. Create Dashboards 3

2.4.5. Trigger Alerts 3

2.4.6. Tomcat and JMX Integration with Splunk ITSI 3

2.4.7. Hardware Utilization monitoring 3

1. **Design 4**
   1. Modelling 4

3.1.1. Flowchart Representation 4

1. **Application Monitoring and Log Analysis Using Splunk 5**
   1. Overview 5
   2. Technical Specifications 5

4.2.1 Hardware Requirements 5

4.2.2 Software Requirements 5

4.2.3 Features 5

4.3 Programming the model 5

1. **Implementation 6**
   1. Scenarios 6
2. **Output screens 7**
3. **Limitations and Future Enhancements 8**
4. **Conclusion**
5. **References 8**

**TABLE OF FIGURES**

**S. No. Figure Page No**

1. **Introduction 1**

Fig 1: The Process Chart Representation of the work flow 2

1. **Design 4**

Fig 12.1 Flowchart Diagram 4

1. **Output Screens 7**
2. **INTRODUCTION**

Today’s datacenter has evolved and IT environments are now more complex than ever. Legacy systems can be expensive and fail to adapt to changing IT environments. Because they usually operate in data silos, these systems can struggle to collect and correlate information from multiple technologies, making it difficult for IT to monitor infrastructure and rapidly troubleshoot problems.

Splunk software changes this all-too-familiar scenario by enabling one to examine their data in depth and in real time from across all of their IT environment—whether on-premises, in the cloud or hybrid. Splunk software collects and correlates the machine or fabric data needed, so one can quickly troubleshoot issues and outages, monitor end-to-end service levels and detect anomalies. Reduce MTTR, lower monitoring costs, improve the system uptime and support strategic initiatives like datacenter optimization and tool consolidation.

Splunk majors in the following capabilities:

* Turn silos of machine data generated in your datacenter into integrated and actionable information. Integrate Splunk apps with existing server monitoring and provisioning tools for one-click granular data analysis.
* Monitor, audit, secure and analyze your messaging and identity infrastructures and workloads in one place in real time.
* Use a customizable Active Directory interface, providing role-based access to support professionals in issue analysis and resolution, while also maintaining compliance.
* Gain visibility across the virtual stack, make faster event correlations, and search transactions spanning virtual and physical components, all from one place.
* Use out-of-the-box, customizable reports to improve your storage monitoring efficiency and proactively plan storage capacity allocations.
* Eliminate silos with complete operational visibility across heterogeneous environments made of multiple flavors of Windows and Linux operating systems, regardless of deployment model.
* Correlate system metrics and events data with data from other technology tiers to prevent performance slowdowns and capacity bottlenecks, as well as secure your environment.
* Troubleshoot problems and outages quickly and detect early warning signs. Collect all logs, configurations, messages, traps and metrics, so you can search, alert and report across your servers in seconds.
  1. AIM

The primal goal of this project is to analyze, configure, monitor and build dashboards based on application log files, and also measure the hardware utilization of app servers.

* 1. OBJECTIVES

In this project, we looked after:

* Indexing of various logs from all backend components.
* Configure alerts and thresholds.
* Build dashboards on various scenarios.
* Monitor CPU and Memory utilization of tomcat servers.
* Generate Alerts for Production Issues.
  1. SUB OBJECTIVES

In this project, we looked after the following sub objectives:

* Understand the different types of logs such as server generated, fabric data logs, etc.
* Study Log Structure and Log Alert types.
* Load silos of machine data into Splunk for indexing and analyzing purposes.
* Integrate Splunk Enterprise with Splunk Add-Ons for JMX(Java Monitoring Extension) and Tomcat.
* Integrate Splunk Enterprise with Splunk IT Service Intelligence app for monitoring purposes.
  1. PROCESS CHART

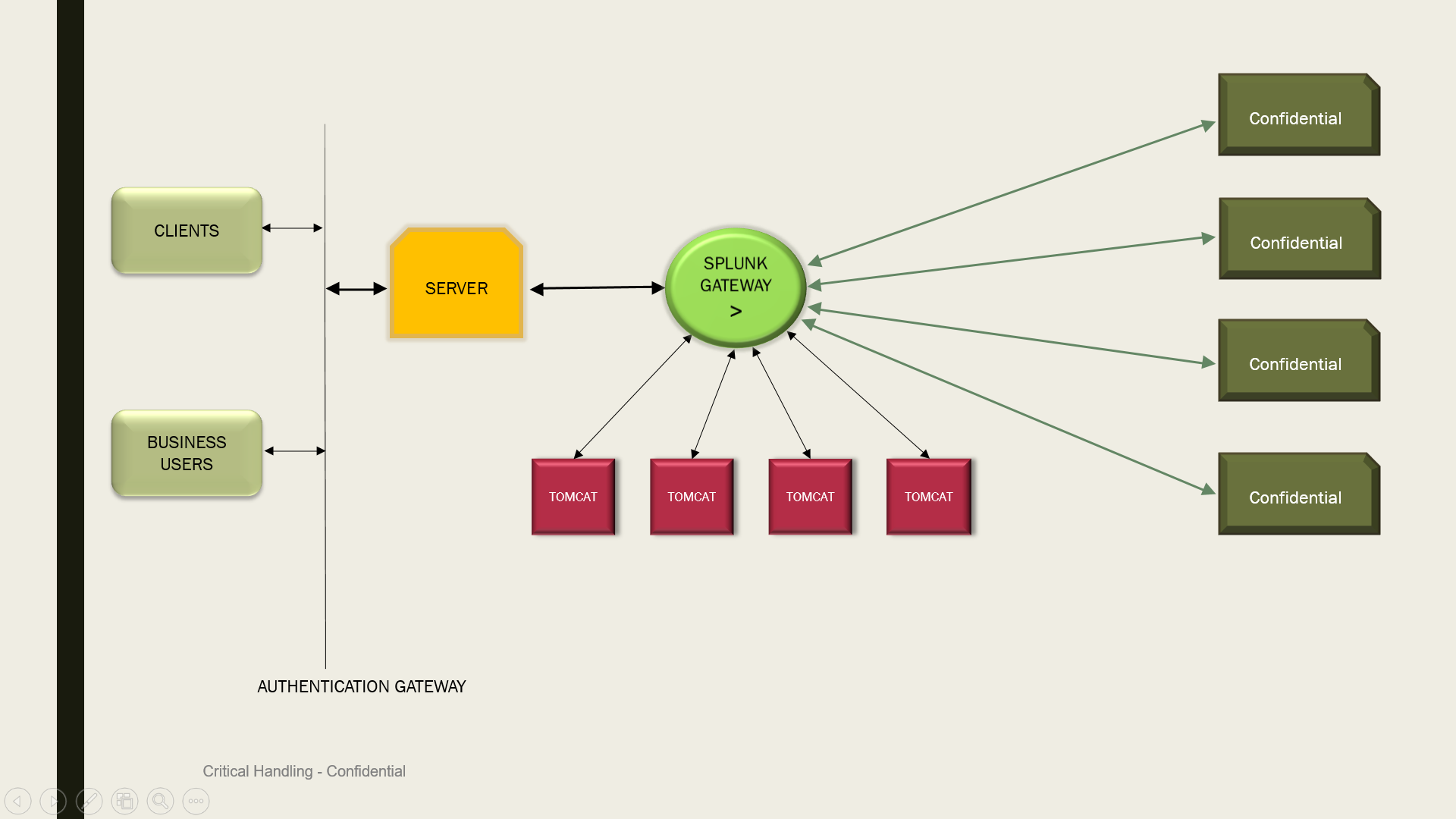


Fig 1. The Process Chart Representation of the work flow.

1. **SYSTEM ANALYSIS**
   1. EXISTING SYSTEMS

There are no existing system for the proposed project.

* 1. PROPOSED SYSTEM:

Splunk Enterprise Software (“Splunk”) is probably the single most powerful tool for searching and exploring data that you will ever encounter. Splunk is often used by system administrators, network administrators, and security gurus, but its use is not restricted to these audiences. There is a great deal of business value hidden away in corporate data that Splunk can liberate.

The fastest way to understand the power and versatility of Splunk is to consider two scenarios: one in the datacenter and one in the marketing department.

* 1. MODULES:
     + 1. Log Configuration

Different types of Logs (confidential) need to be extracted, cleaned, and transformed for further processing

* + - 1. Log Input To Splunk Enterprise

The cleansed logs are added to Splunk via Splunk Web GUI along with other specifications like how to monitor the data and which type of indexing is preferred.

* + - 1. Log Searching In Slunk Enterprise

The logs are parsed by Splunk and are ready for queries for different types of alerts, errors, successful and debug scenarios.

* + - 1. Create Dashboards

The logs that are filtered by search can be presented in form of different kinds of visualizations such as dashboards, pivot charts, quick reports and statistics.

* + - 1. Trigger Alerts

Various kinds of alerts can be created via the Splunk Tool for local or remote monitoring of apps and processes.

* + - 1. Tomcat and JMX Integration with Splunk ITSI

Splunk has a plethora of apps and add-ons from which tomcat and JMX (java monitoring extension) additions are done for tomcat and CPU monitoring.

* + - 1. Hardware Utilization monitoring

The local and remote hardware usage percentage can be measured by the Splunk ITSI Extension.

1. **DESIGN**
   1. Flow Chart

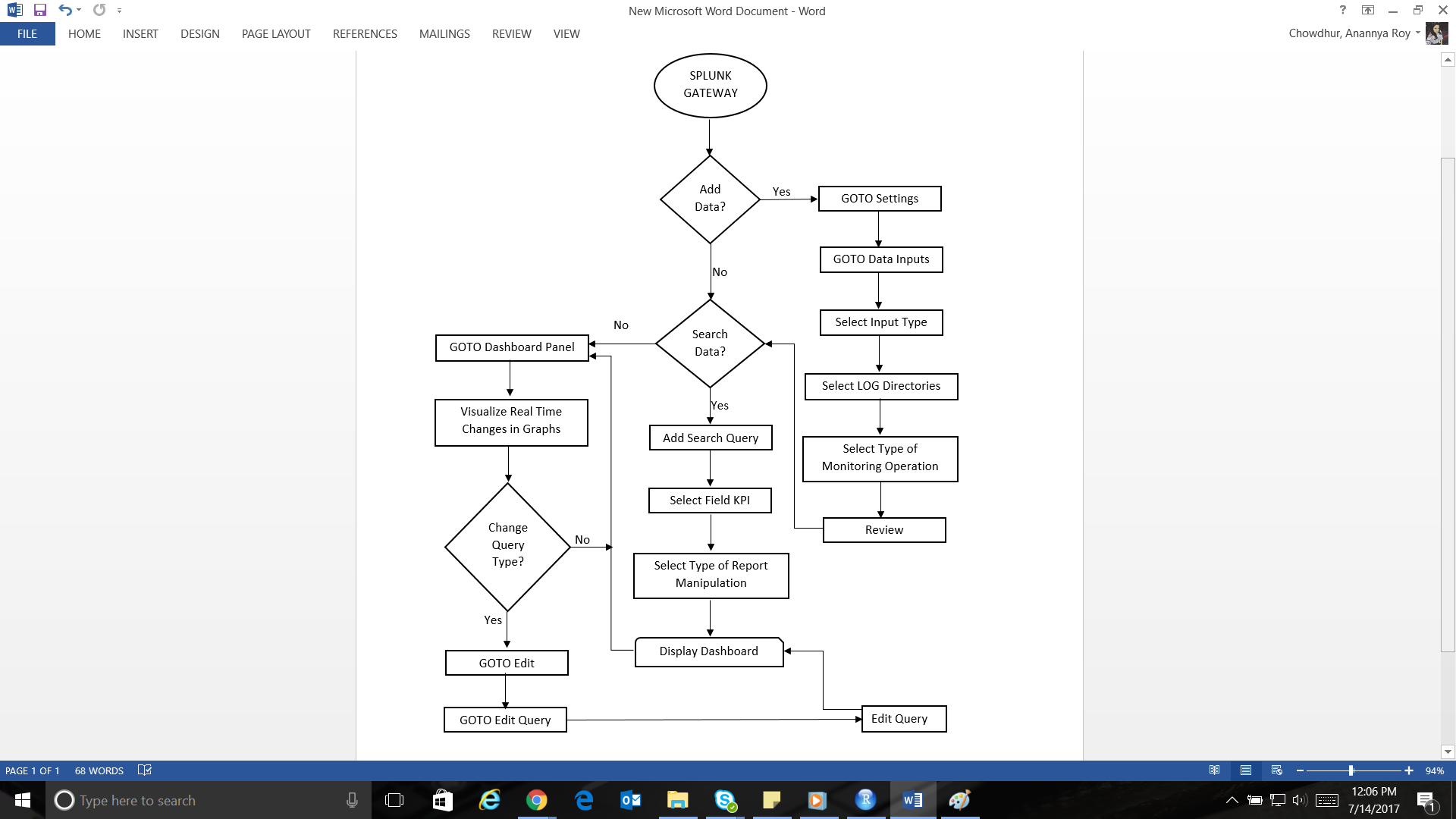


Fig: 2. Flowchart Diagram

1. **APPLICATION MONITORING AND LOG ANALYSIS USING SPLUNK**
   1. OVERVIEW:

The general purpose of a log file is to maintain data of the state of the machine and give reports weather to use it or not. Consider Splunk a Google for log files. It indexes all the log files, fed or automatically collected, and helps in creating analytical dashboards depending on the scenarios. The Splunk Intelligent service tool manages the KPI indicators well and gives memory usage and utilizations results better. Using the enterprise and ITSI, we work out our way to formalize the server logs to preempt the service providing.

* 1. TECHNICAL SPECIFICATIONS:
     1. HARDWARE REQUIREMENTS
* Stations**:** 2 PCs
* Minimum RAM requirement: 32 GB
* Minimum storage space requirement: 100 GB
  + 1. SOFTWARE REQUIREMENT
* Operating System: Windows 8.1/10/OS X/Linux
* Front-End: Splunk Enterprise Software
* Integrated Development Environment: Splunk Web & Splunk ITSI
  + 1. FEATURES:
* **Google for Log Files**: It stores all your logs or machine data by indexing them and provides very fast search capabilities roughly in the same way Google does for the internet.
* **Apps, add-on and Data sources**:  If the data you need cannot be found in any log you can write a script and direct Splunk to digest its output. If that still is not enough you should check Splunk’s App directory for an add-on that collects the necessary data.
* **No schema architecture**:  It performs field extraction at search time. Many log formats are recognized automatically, everything else can be specified in configuration files or written in the search expression.
* **No Backend**:  There is no backend to manage, no database to set up, nothing. Splunk stores data directly in the file system.
* **Smart Operations:** Indexing, Investigating, Knowledge Extraction, Automated Monitoring, Alert Generation, and Reporting.

1. **IMPLEMENTATION**
   1. SCENARIOS:
      1. Log Processing

Configuration of Server/FRS logs onto Splunk (using Splunk web).

Search for the sourcetype and launch the same onto the UI.

In the search bar, enter the error code or the query name you wish to look.

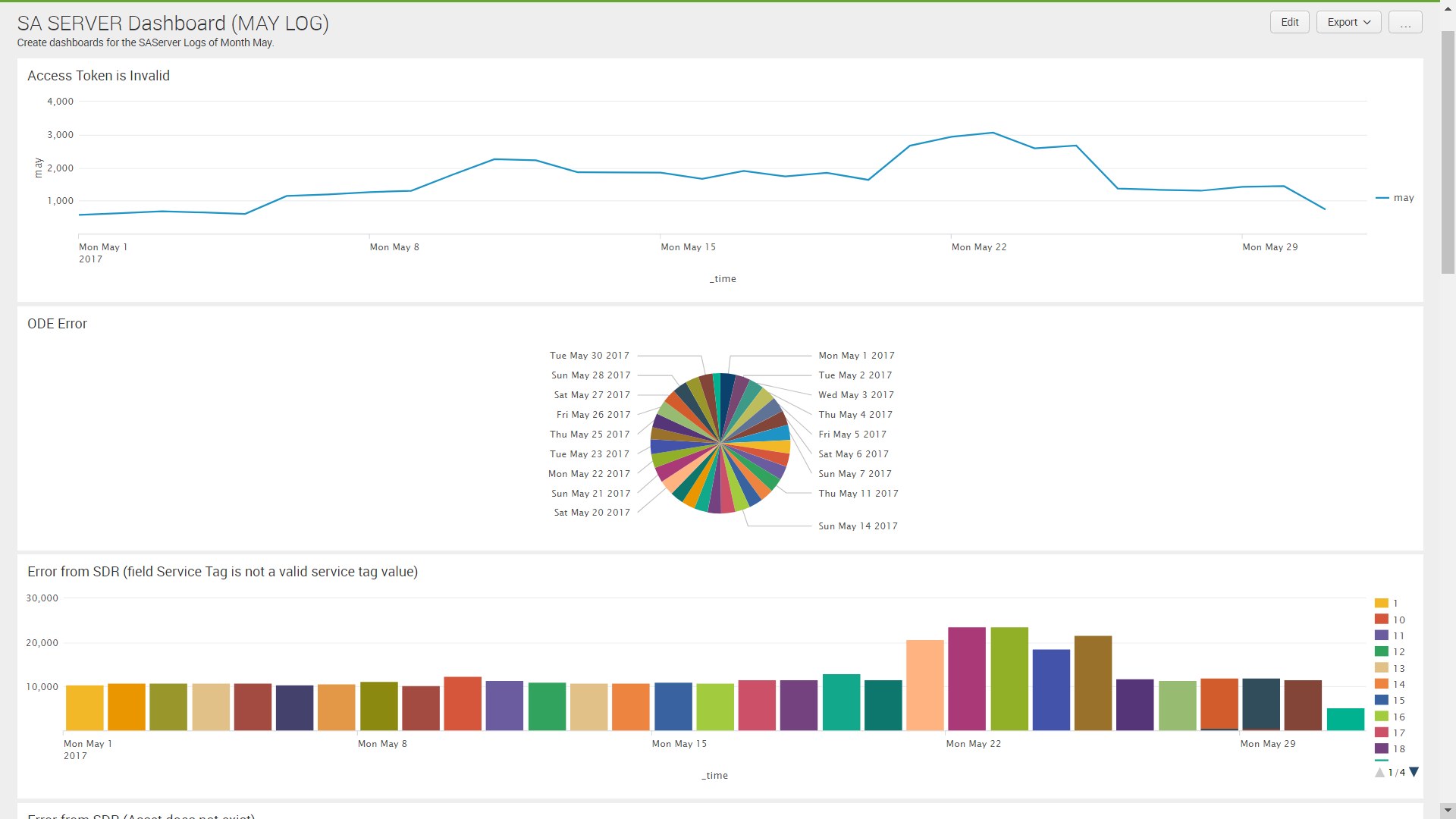
It generates all the possible logs consisting of the errors.

Later multiple reports can be created using the searches enabled.

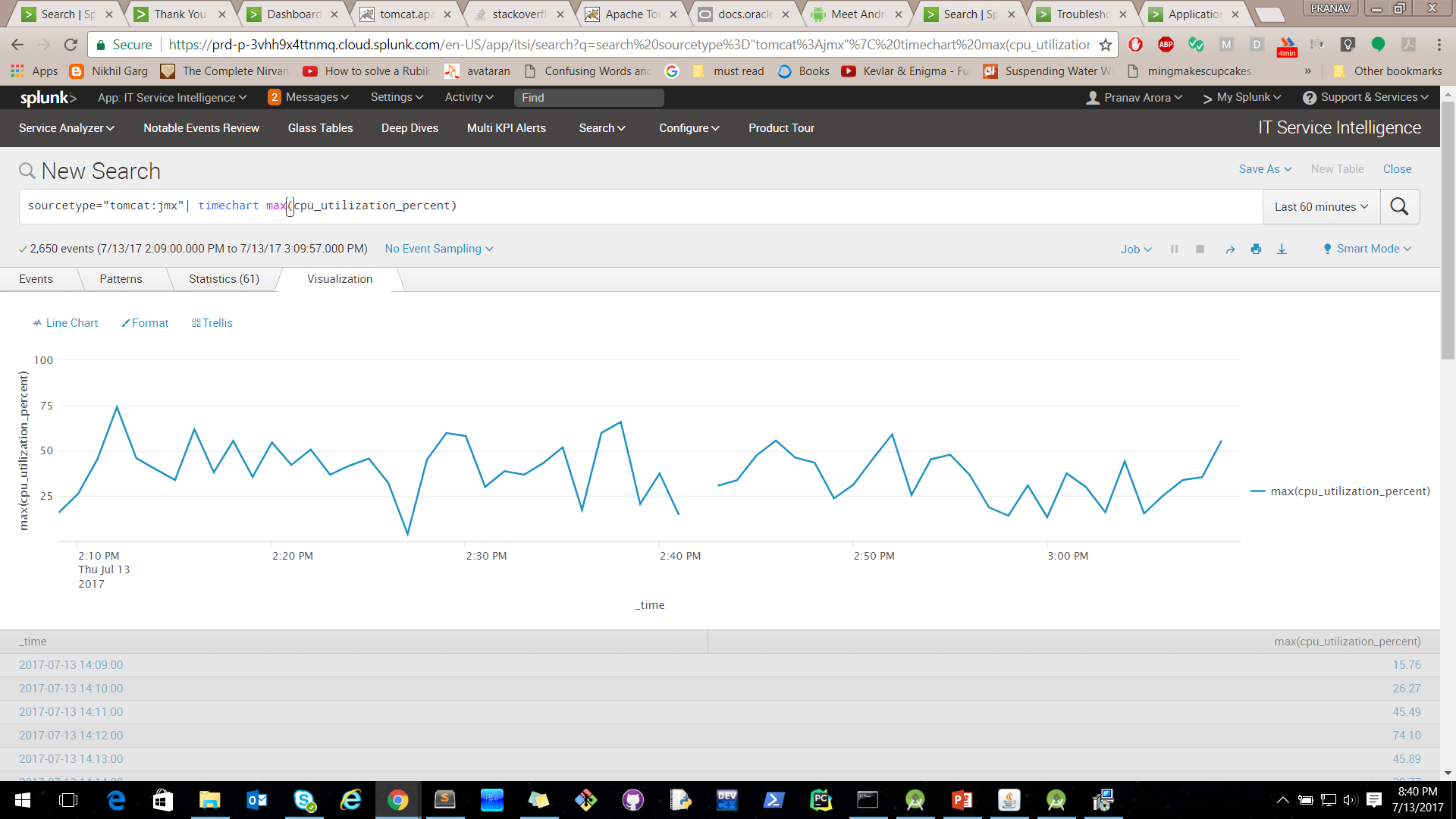
* + 1. Tomcat Utilizations

1. Acknowledge the add-ons onto Splunk enterprise of, Splunk tomcat (server, local or remote) and the JMX monitoring add on for initializing the managing of the server utilizations.
2. Set up the tomcat add-on by adding the localhost details and the username and password for the JMX tool.
3. Fire up the IT service Intelligence Splunk and manually search for “sourcetype=tomcat:jmx”.
4. It will provide you with a screen showing all the memory, CPU usage and many other fields relating to the localhost running on your machine.
5. **OUTPUT SCREENS**

* Server Log



* TOMCAT Utilization



1. **LIMITATIONS AND FUTURE ENHANCEMENTS**
2. Most of the Apps and Add-Ons will work only on the licensed version of the software.
3. The local host is not acknowledged for certain add-ons on free trial.
4. The enterprise tool is not user friendly to measure the CPU utilizations for the server.
5. Possibly the above limitations would be removed in the next free trials.
6. **CONCLUSION**

The present work has been developed with objectives to analyze, configure, monitor and build dashboards based on application log files, and also measure the hardware utilization of app servers while generating alerts for production issues.

1. **REFERENCES**
2. Platform for digital transformation. (2005). Retrieved from https://www.splunk.com/