Elasticsearch Logstash Kibana

A Project Report

Submitted in partial fulfillment of the
Requirement for the award of the degree of

Master of Computer Applications
(Session 2016- 2019)

To

By

Anjali Hora
51610087

Under the Supervision
Of

Mr. Vishal Kamble
(Manager)

DEPARTMENT OF COMPUTER APPLICATIONS
NATIONAL INSTITUTE OF TECHNOLOGY, KURUKSHETRA
May/June/July 2019

intellect®
Design for Digital

28 May 2019

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Ms.**Anjali Hora** (Emp ID – **X41**) is working with Intellect Design Arena as Stipendiary Trainee since 07 Jan 2019 as '**Product Engineer'** in Mumbai.
During this period she worked as part of 'Liquidity Management' Product of iGTB unit

Her conduct and competency was good during this period.

For Intellect Design Arena,

James Jerry Issac
AVP – Human Resources

28 May 2019

DECLARATION

I hereby declare that the work which is being presented in this project report entitled **"ElasticSearch Logstash Kibana",** in partial fulfillment of the requirement for the award of the degree of **MASTER OF COMPUTER APPLICATIONS** submitted to Department of Computer Applications, National Institute of Technology, Kurukshetra is an authentic work done by me during a period from January 2, 2019 to January 29, 2019 under the Guidance of **Mr. Vishal Kamble** in **Intellect Design Arena Limited, Mumbai, Maharashtra.**

The work presented in this project report has not been submitted by me for the award of any other degree of this or any other Institute/University.

Anjali Hora
51610087

This is to certify that the above statement made by the candidate is correct to best of my knowledge and belief.

Date :
Place: Mumbai

Mr. Vishal Kamble
Manager

ACKNOWLEDGEMENT

I would like to express my deepest appreciation to all those who provided me the possibility to complete this report. A special gratitude I give to our final year project manager, Mr. Vishal Kamble, whose contribution in stimulating suggestions and encouragement, helped me to coordinate my project especially in writing this report.

Furthermore I would also like to acknowledge with much appreciation the crucial role of Senior Vice President, Mr. Prasad M Natarajan, who gave the permission to use all required equipment and the necessary materials to complete the Project. A special thanks goes to my team mate, Mr. Akhilesh Lodhi, who helped me to integrate the parts and gave suggestions about the Project. Last but not least, many thanks go to the company, Intellect Design Arena, for giving me the opportunity to work for them. I have to appreciate the guidance given by other supervisor, Mr. Harcharanjeet B and Mrs. Deeksha Prakash, as well as the panels especially in our project presentation that has improved our presentation skills. Thanks to their comment and advices.

ABSTRACT

With Digitization, the compulsion for making a better product or application in IT sector is increasing day by day. Every application in its initial stage requires a lot of features to be added. Even when application goes live, many changes are to be done as customer feedback demands it. In production or deployment stage when application is about to go live and customers are using it, any error in application can make the situation worst as whole code needs to be checked from starting. Making use of checkpoints in code can help but cannot resolve the problem of looking at the code again and again. APIs like log4j can also help to an extent by providing the structured log files to analyse but manually analysing log files will be time consuming. So There is an urgent need to analyse the log files in bulk. ElasticSearch Logstash Kibana is an open source tool for analysing the data, and visualizing it. In the Project, Log files in huge will be analysed using ELK and a dashboard will be created that will help to find out the insights from the logs.

LIST OF FIGURES

Figure No.	Figure Caption	Page No.
Figure 1.1	Operating Cycle of Business	6
Figure 1.2	Inter Company Loans	8
Figure 1.3	Code generation through T-Line Framework	12
Figure 4.1	Complete Flowchart	23
Figure 4.2	RAG Summary Page	24
Figure 4.3	OS Metrics Monitoring	25
Figure 5.1	Features in kibana	30
Figure 5.2	Discover page for LMS Logs in kibana	33
Figure 5.3	Bar Graph that counts number of documents in kibana	34
Figure 5.4	Dashboard for LMS Logs in kibana	35
Figure 5.5	User Interface sample	37
Figure 5.6	Login Page	38
Figure 5.7	User Interface Sample with Embedded Dashboard created in Kibana	39

LIST OF TABLES

Table No.	Table Caption	Page No.
Table 1.1	Definitions, Acronyms, and Abbreviations	9-10
Table 1.2	Terminologies	13-14

TABLE OF CONTENTS

Company Certificate		
Declaration		i.
Acknowledgement		ii.
Abstract		iii.
List of Figures		iv.
List of Tables		v.
Table of Contents		vi.

Chapter No.	**Description**	**Page No.**
1	Introduction	1-4
.	1.1 About Company	5-15
	1.2 About Product	
2	Literature	
	2.1 Log Analysis Tools	16
	2.1.1 Commercially available log analysis tools	16-17
	2.2.2 Open Source log analysis tools	18-19
3	Project Objectives	20-21
4	ElasticSearch Logstash Kibana	22-25
5	Implementation Details	
	5.1 Storing the data in ElasticSearch	26
	5.1.1 Structure of Configuration file	26-29
	5.2 Discovering the data in Kibana	30-33
	5.3 Visualizing the data in kibana	33-34
	5.4 Creating the Dashboard in Kibana	34
	5.4.1 Building a Dashboard	34-35
	5.4.2 Arranging Dashboard Element	35-36
	5.4.3 Inspecting the visualization from the Dashboard	36
	5.4.4 Sharing a Dashboard	36
	5.5 Embedding the dashboard in UI sample	36-39
6	Conclusion	40
7	References	41

CHAPTER 1
INTRODUCTION

1.1 About Company

Intellect Design Arena Ltd. is a Polaris Group company, a global leader in Financial Technology for Banking, Insurance and other Financial Services. A uniquely focused Products business, Intellect addresses the needs of financial institutions in varying stages of technology adoption.Intellect embodies rich Intellectual Property and robust platforms & products across Global Consumer Banking (iGCB), Central Banking, Risk & Treasury Management (iRTM), Global Transaction Banking (iGTB), Insurance (Intellect SEEC), and Wealth Management (iWealth). With over a decade of continuous and significant research and development investment, the Intellect suite is the largest in the industry.

Intellect powers complexity reduction™ in banking and insurance and the key is design thinking. Seamlessly joining the dots between Business, Technology and Operations, it unleashes unprecedented value. 8012 FinTech Design Center is the nerve centre for digital transformation, where better customer experience and operational excellence are delivered through Digital IN and Digital OUT technologies.

World's first full spectrum banking products suite

iGCB

The most advanced Consumer Banking Platform built on current technologies. Seamless omnichannel banking with life cycle assurance optimizes first time cost of ownership and technology running costs. Peerless productivity, at 70% lower post implementation costs! Intellect Quantum CBS, the specialist Core Banking Solution for the unique requirements of central banks is trusted by the central banks of India, Seychelles, Ethiopia and now in

Europe. The solution deploys a formidable array of advanced technology frameworks. Running active balance sheets for nations on real-time enterprise GLs, the solution enables a single source of truth, and has a proven track record for the fastest and most cost efficient implementation.

iRTM

Shifting gears from managing risk to leveraging risk for business advantage, banks turn to iRTM. The largest Treasury operations in the world run on Intellect. The Intellect Basel III LRM Solution with Risk Visualisation across 64 dimensions of risk, addresses the most arduous LRM challenges, enabling 360 swivel views and transaction- level drill downs.

iGTB

The world's first complete Global Transaction Banking Platform. It enables the customer to seize the tremendous global transaction banking opportunity, conservatively estimated at $509bn by 2021. The formidable third generation iGTB with built-in omnichannel Corporate Banking Exchange powers the customer's way to Principal Banker position. Delivering the financial technology the customer has always needed to leverage expertise and on-field innovation without constraints.

Intellect SEEC

One of the world's leading providers of insurance software with an extensive portfolio covering distribution, underwriting and claims. Intellect SEEC offers the right mix for Life and Non-Life Insurers to focus on their strategic imperatives while reducing time-to-market of technology initiatives by up to 60%. Intellect SEEC builds its innovative, low-cost solutions on a firm belief that while the underlying business and technology of insurance are complex, their application should not be.

iWealth

iWealth is an integrated front, middle and back-office solution for Private Banks, Wealth Management Firms, Advisory Firms, Broker Dealers Independent Financial Ad-visors. The platform supports the complete spectrum of Wealth Management functions but can be deployed in modular fashion over multiple delivery channels. With great expectations from clients, shrinking margins and multiplication of local tax and regulatory requirements, Wealth Management has become much more difficult in uncertain times. Wealth management firms must adapt the value proposition and delivery model to serve the new generation of clients. They must leverage the digital opportunity to reduce the cost to serve and restore trust in the ability to deliver superior investment advice. Intellect Wealth Suite supports Prospect Client Management, Asset allocation, Financial planning,

Portfolio management, Performance analytics, Ad-visor workstation, Lending against collaterals, Accounting and Reporting. Intellect Wealth, part of the Intellect Suite of products, is an integrated front, middle, and back office solution for managing the complete wealth management life cycle. Intellect was launched in 2007. Retail and private banks, financial advisers, and wealth management companies use Intellect Wealth. Half of its clients are in Asia, while Europe and the Middle East account for 25 percent each.

Wealth Management System solution has a number of functionality. It has built-in financial planning module which supports goal-based, cash flow-based and retirement planning, income and expenses and capture of lump sum overflow. It also supports customer profiling, portfolio suggestion, and asset allocation based on a user-definable questionnaire, investment manager database with access to subordinates and customers, and a customizable financial planning report tool.

Design for digital differentiation

As banks and insurance companies embrace digital for competitive advantage, they're all focused on the same goal. To provide their customers the same experience at every touch point. It takes Intellect to make this an extraordinary digital experience. Inside and out.

Digital encompasses everything of all types of banking. Its holistic adoption covers Digital OUTSIDE, the experience driver; and Digital INSIDE, the operational excellence enabler.

Uniquely total customer centricity Intellect design philosophy ensures a dramatic shift from disjointed digital activities to strategically aligned digital outcomes.

Digital IN technology manifests in the Apps Vault' An advanced ecosystem with the most comprehensive repository of financial industry apps, engineered for coexistence of legacy, custom build and third party apps, wired for complete front-, mid- and back-end delivery.

Digital OUT technology powers the Ops Hub, Six powerful Digital Ops Hubs that make rapid STP all-pervasive, orchestrate work flows and the optimize the way services are delivered, while ensuring the system remains flexible for change. (Payments Services Hub, Credit Services Hub, Risk Hub, Customer On boarding Hub, Funds Control Hub and Collateral & Limits Hub, as well as universal technology product processors that protect customer legacy investments).

Successful digital design harmonizes seven essential COPARIS architectures. The four architectures for business include the customer in the centre, design of operations, design of decision making, and design of risk associated with the business. The three enabling architectures encompass technical architecture (apps based, omnichannel), integration architecture (pre-built components) and security architecture (entitlement engine built with the apps).

The full spectrum advanced intellect suite of products is built on iDigital, which delivers a uniquely comprehensive digital platform to financial institutions.

Unleashing the power of collaborative design at the world's first design center for financial institutions

At 8012 FinTech Design Center progressive financial institutions realize their transformation ambitions. Identify white spaces in a stimulating environment engineered for collaborative design thinking. Work with leading product and solutions specialists who have led the most complex and ambitious change initiatives at banks around the world.

Banking is a brutally complex business and Intellect is invested in complexity reduction. We observe patterns across our lines of business, from generations of technology trying to relate

with each other. At the Design Centers in Chennai, Mumbai and soon to open in the US, the emphasis is on making design thinking replicable and instilling a design mindset that enables connecting the dots between business, technology and operations. The 8012 FinTech Design Center is a physical manifestation of the enterprise commitment to continuous innovation and value maximization for customers.

With a formidable array of full spectrum banking and insurance products, 'unitized services' can now be assembled to create powerful customized products, or be applied to existing applications. 8012 FinTech Design Center is the nerve center for digital transformation, where superior customer experience and high order operational excellence are delivered through digital technologies.

1.2 About Product

Liquidity Management is a sub-set of the overall activity of cash working capital management performed by a corporate. Businesses procure raw material from their suppliers, add other inputs and convert them into finished goods. This activity is typically called the working capital cycle, or the operating cycle of the business. Liquidity refers to the property of an asset to be quickly converted into cash. Here cash refers to account balances in operating accounts such as current accounts from which cheques can be issued. Assets such as liquid funds or short term fixed deposits may also be considered liquid in so far as money from them can be quickly withdrawn and used for making payments. Liquidity Management Product modules are as follows:

Figure 1.1 Operating Cycle of Business

There are 4 basic modules on which Liquidity works. These Modules are as follows:

- Sweeps
- Notional Pooling
- Investment Sweeps
- Inter Company Loans

SWEEPS:

Sweeps, also sometimes called as "cash concentration" is an arrangement which automatically transfers amounts from an account that exceed (or fall short of) a certain level to another account at a pre-scheduled time, such as the close of each business day. Commonly, the excess cash from multiple accounts of a company may be aggregated into a single account and then invested into higher yielding deposits or money market funds. Excess cash may also be swept into other accounts which have a debit balance, thus causing substantial interest savings. Bank customers stand to gain many unique benefits from using sweeps. Some of the benefits are: Optimizes cash availability: The treasury manager gets full visibility and access to all the funds lying in multiple accounts of the company. He is then able to make better utilization of his money. Reduction in interest costs: Sweep structures

allow corporates to sweep funds from cash surplus accounts to overdrawn accounts, thus bringing substantial interest savings to the company. Convenience: Once set up, sweep structures operate in an automatic way. This frees the corporate and the treasury team of constant monitoring of accounts and performing manual transfers.

NOTIONAL POOLING:

Notional Pooling is a method of calculating interest on the combined credit and debit balances of the Participating Accounts without physically transferring funds. Interest is calculated on the aggregated net account balance of all accounts implemented in the cash pool. In a notional pool structure, the debit and credit balances of the accounts within a pool are offset on a ”virtual” basis (i.e. there is no physical movement of funds). As the debit balances are fully or partially offset by the credit balances, the corporate assumes lower debit interest or may receive higher credit interest. Typically bank interest is charged on a single position after notional aggregation of the balances on the pool accounts. This is often combined with a reallocation of interest to the underlying participating accounts. In an interest compensation or interest enhancement arrangement, credit balances help to cover short positions, but interest is still paid to or from the participating account. Corporates and banks however find the regulatory environment for notional pooling to be more complex than for cash concentration (sweep) services.

INVESTMENT SWEEPS:

We have looked at Sweeps (cash concentration) previously, where funds are transferred from one account to another in an automated way. Investment Sweeps is an extension of the Sweeps functionality, whereby funds are transferred from the bank to a higher yielding investment avenue such as a Time Deposit account or a Money Market Fund (MMF). The reason for transferring funds to such an entity is to get higher returns than what would accrue to the account holder if they were lying overnight in the bank accounts. As the reader may be aware, current / checking accounts in banks offer no or negligible returns. However,

investments in time deposits or MMFs yield better returns and offer a high level of safety as well.

INTERCOMPANY LOANS:

Globally, large corporations operate as "conglomerates", creating multiple companies belonging to the same group, each of which is a separate legal entity. The creation of such group companies may be on different logical lines, such as catering to different product lines or geographical areas or joint venture partners or for any other reason. Each of these companies however, would have one thing in common: a common parent, which may act as a holding company and providing the overall strategic leadership and vision to the group of companies. Some of these companies at times may be deficit in cash while others could be in surplus. Cash deficit companies would ordinarily be expected to raise lines of credit from their banks, while cash surplus companies would ordinarily be expected to deploy their surplus funds in higher yield instruments such as term deposits and MMFs. However, if the companies belong to the same conglomerate group, they can come together and create a win – win model of inter-company loans, subject to regulatory restrictions. A hypothetical example of a corporate group is shown below:

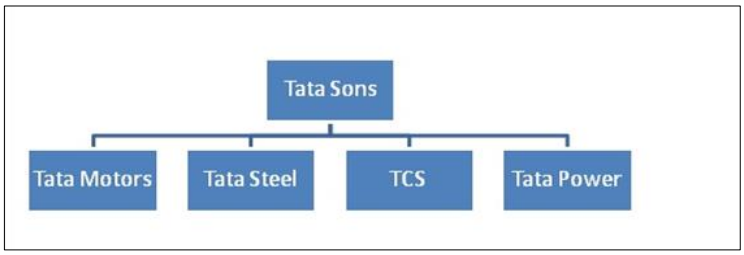

Figure 1.2 Inter Company Loans

The win-win situation can be created because a wide “spread’ exists between the rates at which banks lend to their borrowers and the rates which they offer to deposit holders. For example, if bank deposits earn 8% and loans cost 12%, a “spread” of 4% exists between the lending and the borrowing rates of the banks. If the cash surplus company were to lend to cash deficit company at any rate between this range (say, 10%), it would be a win-win situation for both the entities.

Table 1.1 Definitions, Acronyms, and Abbreviations

Term	Description
iGTB	Global Transaction Banking
iRTM	Central Banking, Risk & Treasury Management
iGCB	Global Consumer Banking
GLE	Global Liquidity Engine
iWealth	Wealth Management
IDE	Integrated Development Environment
FT	FinTech
LMS	Liquidity Management System
SDLC	Systems Development Life Cycle

UI	User Interface
EOD	End of Day
MVC	Model, View, and Controller.
JSP	Java Server Pages
ARX	Armour Applicaion
WSDL	Web-Services Description Language
XML	Extensible Mark-up Language
XSD	XML Schema Definition

Tools and frame work of LMS

Tools used

1.T-line frame work

2.EJB

3.JAVA

4.JSP

5.JMS FRAME WORK

What is T-line frame work?

Facilitates Rapid Application Development (RAD) of web-based J2EE applications conforming to MVC architecture.

Provide high customization.

Generates almost 70% of code required for implementing transaction use-cases, resulting in significant productivity boost

Code generated is defeact free and easy to maintain.

Code re-generaration is made easy

its Relaible ,offers high performance scalability and flexibility.

T-line Application frameworks

It is based on standard MVC Frame work follows means it separate areas from

1.Module

2.View

3.Controller

Features

- We have used struts framework to provide MVC frameworks
  Struts process all the incoming requests through 'front controller'.
- A view pages consists of several JSP pages.
- A from defined tool is used in screen design and provide relability,statbility,robustness.
- Provide automation for static entiites ,build maker and checker framework.
- Provides Support for User Defined Fields (UDF) without any code change
- Provides integration with any Security Provider (default Armor) to automate authentication, L2 checks and entitlements

T-line Application framework

- Framework for implementing any transaction use-case.

- Static Maintainence framework : This framework again uses UDF / Picklist / List / Menu / Securty / Logging / Exception and Message handling frameworks) to provide complete automation for static maintenance of entities

T-line Framework Events

- Login Event
- Refresh Menu Event
- List Events
- Picklist Events
- Transaction use-case Events

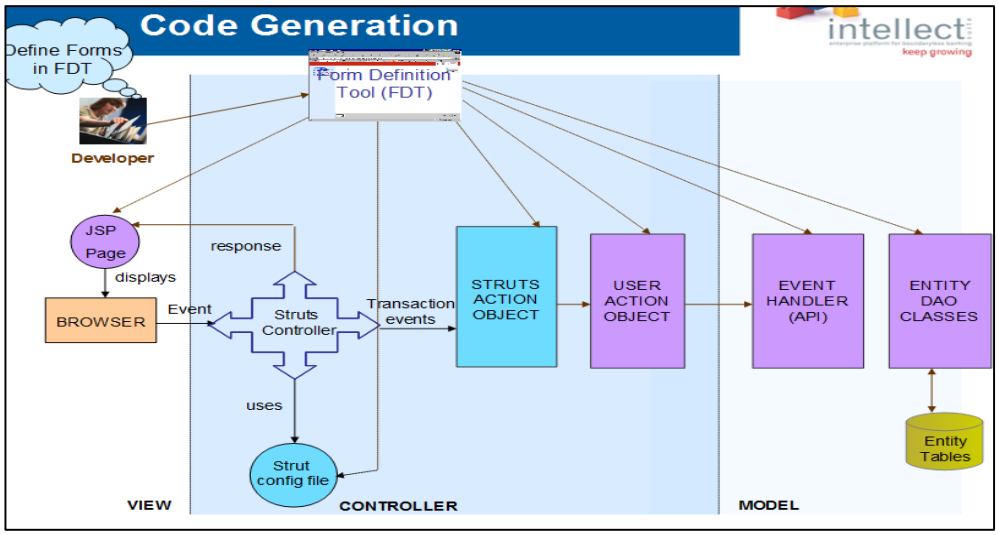

Figure 1.3 Code generation through T-Line Framework

Table 1.2 Terminologies

S.No	TERMS	DESCRIPTION
1	IGTB	Global Banking Platform
2	GLE	Global Liquidity Engine is an Intellect LMS Product works as a solution for cash management and opt by various banking Institute such as HSBC
3	SWIFT MESSAGES	Swift message is used to transact between cross boarder
4	ARMOUR	Armour is with security with login and passwords
5	NON-ARMOUR	Non-Armour is application without login and password
6	BVT	Back Value Transaction is done when there is some changes in amount and calculation needs to be dine for current amt adding the changed amount form back date,like total amount should be present adding 1000 rupees from 10 day back so system use BVT.
7	Vestro Account	Vestro accounts are accounts of the bank but outside country boundry
8	Nostro Account	Nostro Account is account which is account from our bank but inside country boundries

9	Control Account	The account on which the sweep rule will be executed
10	Contra Account	The account which has the other leg of the transaction
11	Sweep reversal	Reversal will be performed if the next working day of the ControlAccount, Contra Account and the Currency is the same.
12	EOD	END OF DAY this is the execution mode for sweep instruction
13	BOD	BEGINNING OF DAY this is the execution mode for sweep instruction
14	EOI	END OF INPUT

LMS product is build on T-line framework which is their own customizable framework.It uses java as its core language. The product is used by banks and helps in managing the unlimited transactions taking place on timely basis. Transactions are occurring by accounts owned by customers and companies. Company's requirement was to visualize the LMS logs to calculate the health of the application. The technology used by us to visualize the logs is Elasticsearch Logstash Kibana.

Elastisearch is used for storing the logs. Logstash is used for filtering the logs and Kibana is used for visualizing the logs. One can also use other preprocessing tools like streamset (GUI based) for filtering the logs present in Elasticsearch. Kibana is used to visualize the logs present in elasticsearch.

CHAPTER 2
LITERATURE

2.1 Log Analysis Tools

A log is an automatically produced and time-stamped documentation of events related to a particular event. Log analysis helps us in understanding what has happened and derive useful metrics in monitoring, performance, digital marketing etc. Log analytics helps us in performing real time analysis of large scale data and obtain insights for a wide variety of applications such as digital marketing, application monitoring, fraud detection, ad tech, IoT etc. Log analysis tools are available both commercially or as open source tools that can be used for free.

2.1.1 Commercially available log analysis tools

Splunk

Splunk is a software platform which is used to search, analyse and visualise machine generated data. The data can be gathered from websites, applications, sensors etc which make up your IT infrastructure and business. Splunk allows real time processing of data which is it's biggest selling point. Splunk can be used to create Alerts or Event notifications depending on the state of the machine. We can create visualisations using Splunk for better representation of the data.

Retrace

A node is a single server that is part of your cluster, stores your data, and participates in the cluster's indexing and search capabilities. Just like a cluster, a node is identified by a name which by default is a random Universally Unique IDentifier (UUID) that is assigned to the node at startup.You can define any node name you want if you do not want the default. This name is important for administration purposes where you want to identify which servers in your network correspond to which nodes in your Elasticsearch cluster. A node can be

configured to join a specific cluster by the cluster name. By default, each node is set up to join a cluster named elasticsearch which means that if you start up a number of nodes on your network and—assuming they can discover each other—they will all automatically form and join a single cluster named elasticsearch.

Logentries

An index is a collection of documents that have somewhat similar characteristics. For example, you can have an index for customer data, another index for a product catalog, and yet another index for order data. An index is identified by a name (that must be all lowercase) and this name is used to refer to the index when performing indexing, search, update, and delete operations against the documents in it.

Logmatic

A document is a basic unit of information that can be indexed. For example, you can have a document for a single customer, another document for a single product, and yet another for a single order. This document is expressed in JSON (JavaScript Object Notation) which is a ubiquitous internet data interchange format. Within an index/type, you can store as many documents as you want. Note that although a document physically resides in an index, a document actually must be indexed/assigned to a type inside an index.

Sumo logic

An index can potentially store a large amount of data that can exceed the hardware limits of a single node. For example, a single index of a billion documents taking up 1TB of disk space may not fit on the disk of a single node or may be too slow to serve search requests from a single node alone. To solve this problem, Elasticsearch provides the ability to subdivide your index into multiple pieces called shards. When you create an index, you can simply define the number of shards that you want. Each shard is in itself a fully-functional and independent ”index” that can be hosted on any node in the cluster.

2.1.2 Open Source log analysis tools

Graylog

Graylog is defined in terms of log management platform for collecting, indexing, and analyzing both structured and unstructured data from almost any source. Nowadays most of the applications following microservice architecture .where many of microservices are hosted on different machines. So, it will take very huge time for the user to get the logs from each microservice on different machines every time. To avoid this we have an application called Graylog which works by configuring at one place and get the logs of all microservices at one centralized location.

Logstash

Logstash is an open source data collection engine with real-time pipelining capabilities. Logstash can dynamically unify data from disparate sources and normalize the data into destinations of your choice. Cleanse and democratize all your data for diverse advanced downstream analytics and visualization use cases. The following are the features of logstash: The ingestion workhorse for Elasticsearch and more Horizontally scalable data processing pipeline with strong Elasticsearch and Kibana synergy Pluggable pipeline architecture Mix, match, and orchestrate different inputs, filters, and outputs to play in pipeline harmony Community-extensible and developer-friendly plugin ecosystem Over 200 plugins available, plus the flexibility of creating and contributing your own.

Fluentd

Fluentd works on Unified Logging Layer means it tries to structure logs as JSON as much as possible. The idea is to provide an interface which can be used by almost any producer or any consumer of the logs. This helps in all phases of log processing like Collection, Filter, and

Output/Display. it also provides 300+ plugins out of which only a few are provided by official Fluentd repo and a majority of them are maintained by individuals. Fluentd also makes use of Regex patterns for the logs whose format is not known or is not already available with Fluentd. Those patterns can be verified on Fluentular. Fluentd is written in CRuby and is maintained by Treasure Data Inc.

Logz.io

Logz.io provides a cloud-based log analysis service which is based on the open source log analysis platform - the ELK Stack (Elasticsearch, Logstash, Kibana). The features provided by Logz.io include: alerting, user control, parsing services, support, integrations, and audit trail. The platform uses machine learning algorithms to identify and reveal critical errors hidden within the log data.

Elasticsearch Logstash Kibana

The ELK Stack is a collection of three open-source products, Elasticsearch, logstash,kibana. Elasticsearch is a search engine based on the Lucene library.It is ”a solution built from the ground up to be distributed” and used a common interface, JSON over HTTP, suitable for any programming language. Data is stored in indexes. Logstash is a tool for filtering the events or log messages by parsing it. Logstash is an open source, server-side data processing pipeline that ingests data from a multitude of sources simultaneously, transforms it, and then sends it to elasticsearch. Kibana is an open source analytics and visualization platform designed to work with Elasticsearch. You use Kibana to search, view, and interact with data stored in Elasticsearch indices. You can easily perform advanced data analysis and visualize your data in a variety of charts, tables, and maps.

CHAPTER 3
PROJECT OBJECTIVES

Data is important in any form be it a customer data, Employee data, etc. Even the list of household items or a memo has data in some form. With this data, they are taking many big decisions that ultimately increases their ROI (Return on Investment). Data Visualizations, on the other hand, is a powerful way to explore data with presentable results. Its Primary use is the preprocessing portion of the data mining process. It supports in data cleaning process by finding incorrect and missing values. It also uses for variable derivation and selection means to determine which variable to include and discarded in the analysis. It also play role in combining categories as part of the data reduction process. Logs are type of data that contains information about events that occur in an operating system or other software. There are different types of logs, event logs, transaction logs, message logs etc. Logs play a very important role for product based company as it calculates the health of the application post production.

While any incident raised post-production, the L1 support has to go via multiple applications/screens to verify the error/exception that would have occurred. It leads to following set of steps to occur:

1. Raise Fire ID Tickets.
2. Work with Application server admin to see where the issue is.
3. Work with Database admins to find where the issue is.
4. If no access to logs, ask admin to send the logs (File Size, PI

Data, Network, Citrix/VPN constraints).

5. Consolidate all the analysis at one place.

This requires huge effort and also a cumbersome process of trying to bring entire data for a given time period in one screen/excel/notepad. When a new business/customer/major

release is being on boarded, there are no statistics available to predict how the behavior of the system will be. This leads to visualization of logs so that the whole statistics can be shown and easily understandable. Following project objectives after concluding the problem statement are:

1. To manage bulk of logs coming from company's applications on timely basis.
2. To visualize the logs such that it is easy to understand the type of error in the applications.
3. To manage all types of logs like server logs, system logs, Application logs etc.
4. To inform in prior the risk of failure in application by machine learning algorithm.
5. To inform about the loop-holes in applications with the help of statistics.

CHAPTER 4
ELASTICSEARCH LOGSTASH KIBANA

Elasticsearch Logstash kibana

Our proposed solution is to use the utility called ELK STACK which is the combination of the three open source tools called Elasticsearch, Logstash, Kibana. ELK STACK is open source and version is updating as new feature is adding frequently. These three tools can be used as an individual entity and can be integrated with other different tools. Using ELK stack in a customized way to fulfill the need of the company is the motivation of the project. Also, Integrating it with other technologies is another important task to fulfill.

Elastic Search:

Elastic Search is an open source and distributed database that is used to store, search and analyze big volume of data quickly and in real time. Our Problem statement requires the need to collect log and transaction data to analyze and to mine this data to look for trends, statistics, summarizations, or anomalies. Elastic Search (part of Elastic Search/Log stash/Kibana)is used to store large volumes of data.

Log stash:

Log stash (part of Elastic Search,/Log stash/Kibana) is open source tool and used to parse data so that data enters elastic search in structured form. It has real time pipeline capability with strong elasticsearch and kibana synergy. It has over 200 plugins available that can handle data of all shapes and sizes. It also ingest the data shipped from mobile devices to intelligent homes, connected vehicles, healthcare sensors and many other industry specific applications.

Kibana:

Kibana(part of Elastic Search/Logstash/Kibana) is open source tool and used to create dashboard which helps in finding the solution where the problem is coming in product or application. Also, Visualization helps in depicting the current scenario of product to almost anyone as it is easy to understand. Customizing the dashboard according to the need is another important aspect of kibana as it interacts with data stored in Elasticsearch. We are able to visualize the data in a variety of charts, tables, and maps.

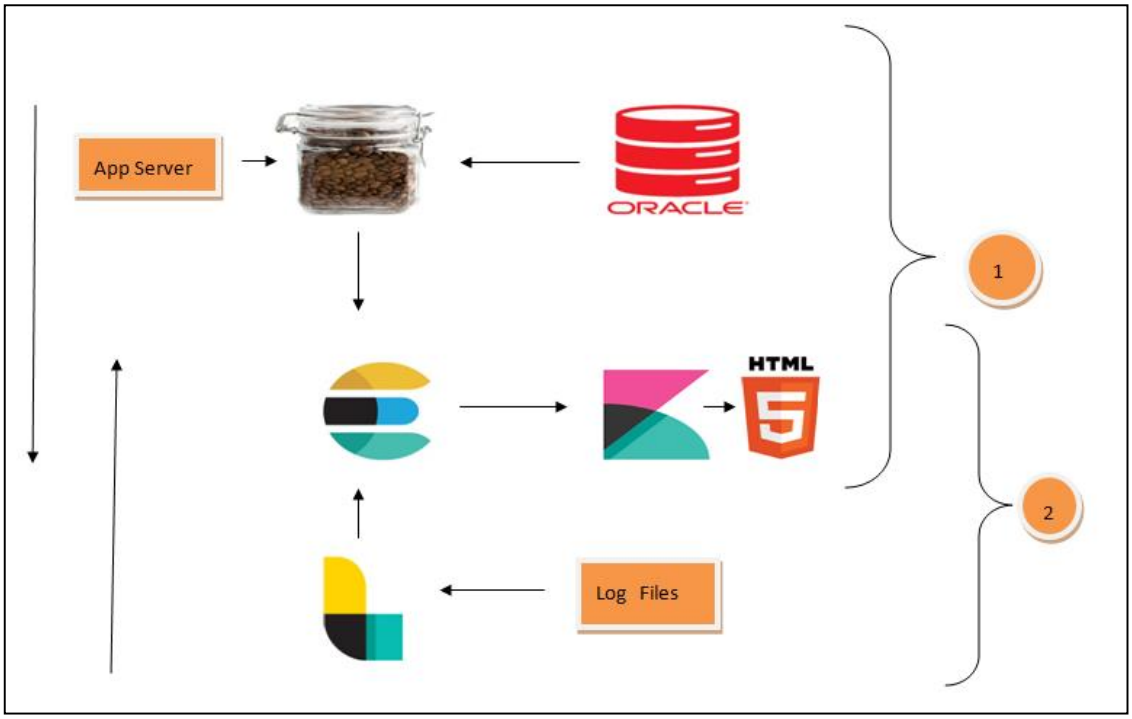

Figure 4.1 Complete Flowchart

Initially, Company want to monitor the server so that the health of the application can be predicted. For that, a clustering algorithm was written in java that takes app server metrics and put it into java fat jar and store the data in elasticsearch. Application's database logs were also monitored. Conditions are applied through algorithm that takes logs specific to the type of servers. The logs will be visualized in kibana. Our module was to store the log files in elasticsearch, pre-process it using logstash, and visualize it using kibana. Other task was to create User Interface that embeds dashboard created in kibana. Section 1 in figure 5.1 is completed by the other team. Section 2, on the other hand, is completed by our team.
Following are the visualizations created through kibana:

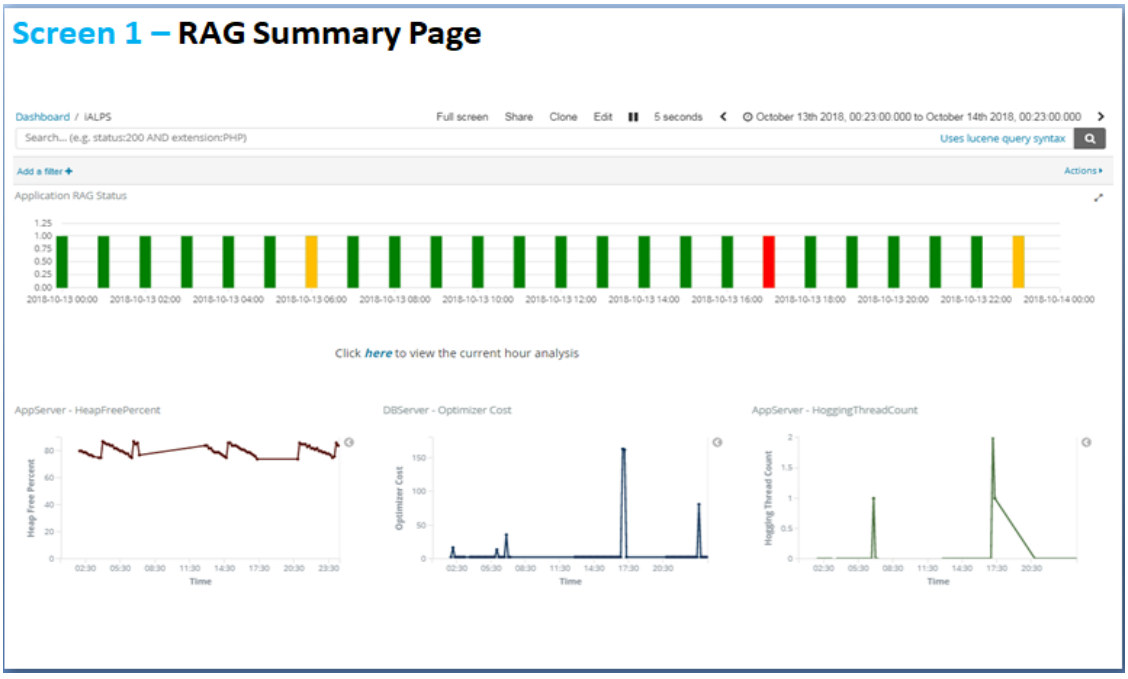

Figure 4.2 RAG Summary Page

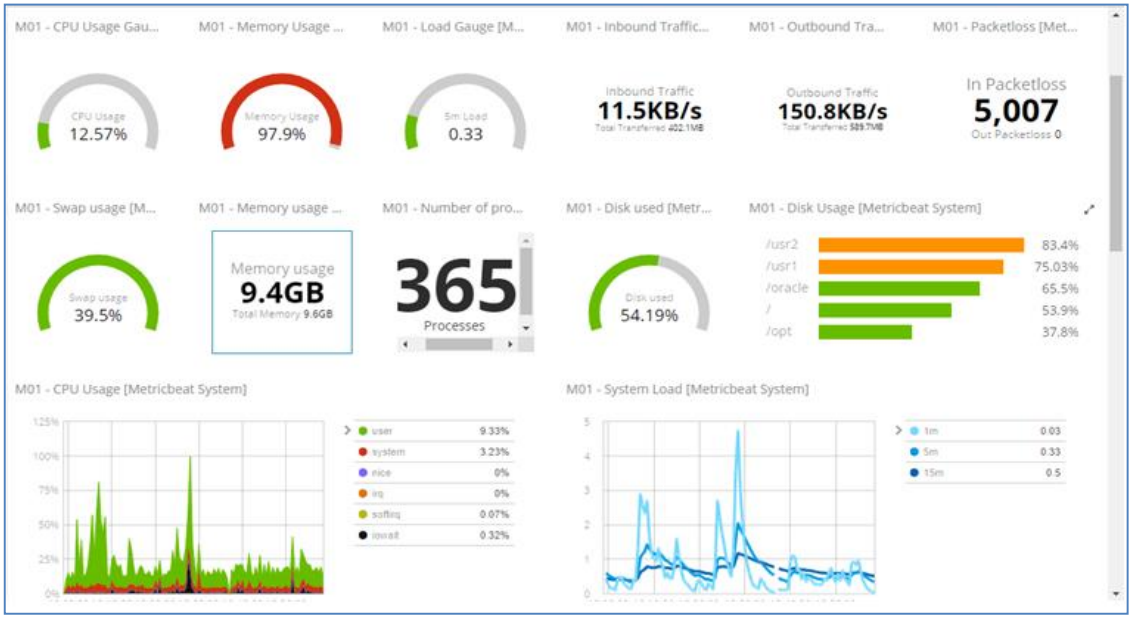

Figure 4.3 OS Metrics Monitoring

CHAPTER 5
IMPLEMENTATION DETAILS

Elasticsearch is a highly scalable open-source full-text search and analytics engine. It allows you to store, search, and analyze big volumes of data quickly and in near real time. It is generally used as the underlying engine/technology that powers applications that have complex search features and requirements.

The following five steps shows the Project's implementation:

5.1 Storing the data in elasticsearch

Bydefault, the port number for elasticsearch is 9200 and the port number for kibana is 5601.

To run elasticsearch in command line,

./bin/elasticsearch

To run kibana in command line,

./bin/kibana

Changes are to be done in the yml files of elasticsearch, logstash, kibana for making it run in other servers. Data can be stored in elasticsearch using logstash config file.

5.1.1 Structure of configuration file

```
Input
{
...
}

filter
{
```

...

}

output

{

...

}

input

{

file{

path=>"/var/log/messages"

type=>"syslog"

}

file{

path=>"/var/log/apache/access.log"

type=>"apache"

}

}

path is the path for log files. Files with any extension can be stored in elasticsearch but It will be converted into json format when viewing it in kibana.

type is the type of server from where log is coming.

filter

{

grok { match => ["message", "%{HTTPDATE:[@metadata][timestamp]}"] }

```
 date { match => [ "[@metadata][timestamp]", "dd/MMM/yyyy:HH:mm:ss Z" ]  }
}
```

filter tag helps in filtering the events and logs. Filters help in joining the data as there is unavailability of joins in elastic search.

```
 output {
 elasticsearch {
 action => "%{[@metadata][action]}"
 document_id => "%{[@metadata][_id]}"
 hosts => ["example.com"]
 index => "index_name"
 protocol => "http"
 }
}
```

action depicts what kind of action we want to perform in elasticsearch.
document id depicts the id of document.
hosts is the hostname.
index is the indexname.

After creating the log file, run the command in command line

bin/logstash -f logstash-simple.conf

After running the command, Log files will be stored in elasticsearch. Elasticsearch uses indexing for fast searching. Bydefault, logstash runs in port 5044.
Filtering on data present in elasticsearch can be done using the filters in kibana. It helps in getting only the data which is important. Once the data stored in elasticsearch and filtered in

kibana, we can view it in kibana by visualizing it. Visualizing the filtered data in kibana helps in understanding the data easily. Thus helps in finding the insights from the data.

The server logs and server metrics went through lightweight shipper called filebeat and metricbeat respectively. Filebeat is a lightweight shipper for forwarding and centralizing log data. Installed as an agent on your servers, Filebeat monitors the log files or locations that you specify, collects log events, and forwards them to either to Elasticsearch or Logstash for indexing.

Metricbeat is a lightweight shipper that you can install on your servers to periodically collect metrics from the operating system and from services running on the server. Metricbeat takes the metrics and statistics that it collects and ships them to the output that you specify, such as Elasticsearch or Logstash.

Logs can be stored in elasticsearch from different end points. For example: Aerospike metrics, Apache logs, AWS metrics, cloudwatch logs, elasticsearch logs, Golang metrics, kafka logs, MySQL logs, nginx logs, PostgreSQL logs, System logs, Docker metrics etc. Multiple servers can run and store logs in elasticsearch but each end point need to configure with different filebeat. The capability of logstash pipelines will help to handle the excess of logs. If you need to run more than one pipeline in the same process, Logstash provides a way to do this through a configuration file called pipelines.yml.

After sending the logs from logstash into elasticsearch, data will be discovered in the kibana. Data will be stored in elasticsearch and uses index for fast searching.We can again filter the data present in elasticsearch using the query in kibana. Once the data stored in elasticsearch, we can view it in kibana by visualizing it.

5.2 Discovering the data in kibana

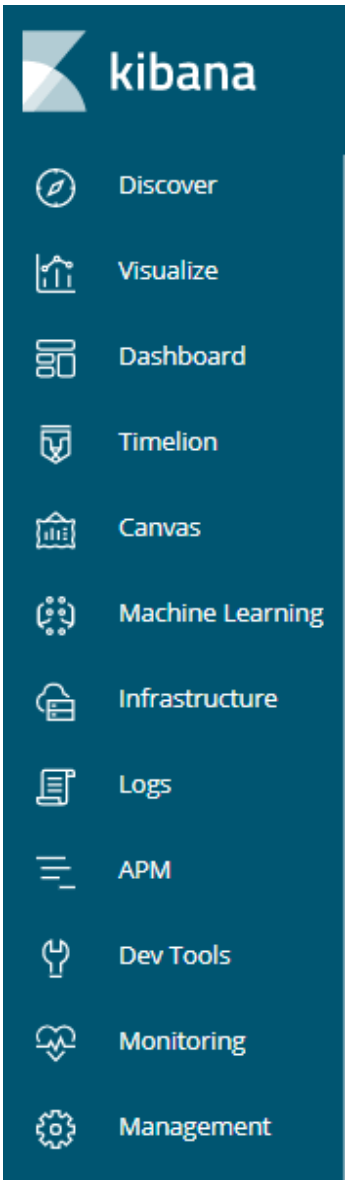

Figure 5.1 Features in kibana

The features in kibana are as follows:

Discover:

Discover enables you to explore your data with Kibana's data discovery functions. You have access to every document in every index that matches the selected index pattern. You can

submit search queries, filter the search results, and view document data. You can also see the number of documents that match the search query and get field value statistics. If a time field is configured for the selected index pattern, the distribution of documents over time is displayed in a histogram at the top of the page.

Visualize:

Visualize enables you to create visualizations of the data in your Elasticsearch indices. Kibana visualizations are based on Elasticsearch queries. By using a series of Elasticsearch aggregations to extract and process your data, you can create charts that show you the trends, spikes, and dips you need to know about. You can create visualizations from a search saved from Discover or start with a new search query.

Dashboard:

Dashboard allows you to build dashboards that display related visualizations. In our project, LMS logs are visualised.

Timelion:

Timelion is a time series data visualizer that enables you to combine totally independent data sources within a single visualization. It's driven by a simple expression language you use to retrieve time series data, perform calculations to tease out the answers to complex questions, and visualize the results.

Canvas:

Canvas is a whole new way of making data look amazing. Canvas combines data with colors, shapes, text, and your own imagination to bring dynamic, multi-page, pixel-perfect, data displays to screens large and small.

Machine Learning:

As datasets increase in size and complexity, the human effort required to inspect dashboards or maintain rules for spotting infrastructure problems, cyber attacks, or business issues becomes impractical. The Elastic machine learning features automatically model the normal

behavior of your time series data — learning trends, periodicity, and more — in real time to identify anomalies, streamline root cause analysis, and reduce false positives.
The machine learning features run in and scale with Elasticsearch, and include an intuitive UI on the Kibana **Machine Learning** page for creating anomaly detection jobs and understanding results.

Infrastructure:
Use the interactive Infrastructure UI to monitor your infrastructure and identify problems in real time. You can explore metrics and logs for common servers, containers, and services.

Logs:
Logs UI are used to explore logs for common servers, containers, and services.

APM:
Elastic Application Performance Monitoring (APM) automatically collects in-depth performance metrics and errors from inside your applications.

Dev Tools:
The **Dev Tools** page contains development tools that you can use to interact with your data in Kibana.

Monitoring:
The Kibana monitoring features serve two separate purposes:

- To visualize monitoring data from across the Elastic Stack. You can view health and performance data for Elasticsearch, Logstash, and Beats in real time, as well as analyze past performance.
- To monitor Kibana itself and route that data to the monitoring cluster.

Management:
The Management application is where you perform your runtime configuration of Kibana, including both the initial setup and ongoing configuration of index patterns, advanced

settings that tweak the behaviors of Kibana itself, and the various "objects" that you can save throughout Kibana such as searches, visualizations, and dashboards.

Data can only be visualized in kibana when it is stored in elastic search. For viewing the data in kibana, index pattern needs to be created. The same index pattern will be created as the one in config file. We can check whether the data is coming by clicking on the “Discover” tab in Kibana.

The discover page for the LMS logs is as follows:

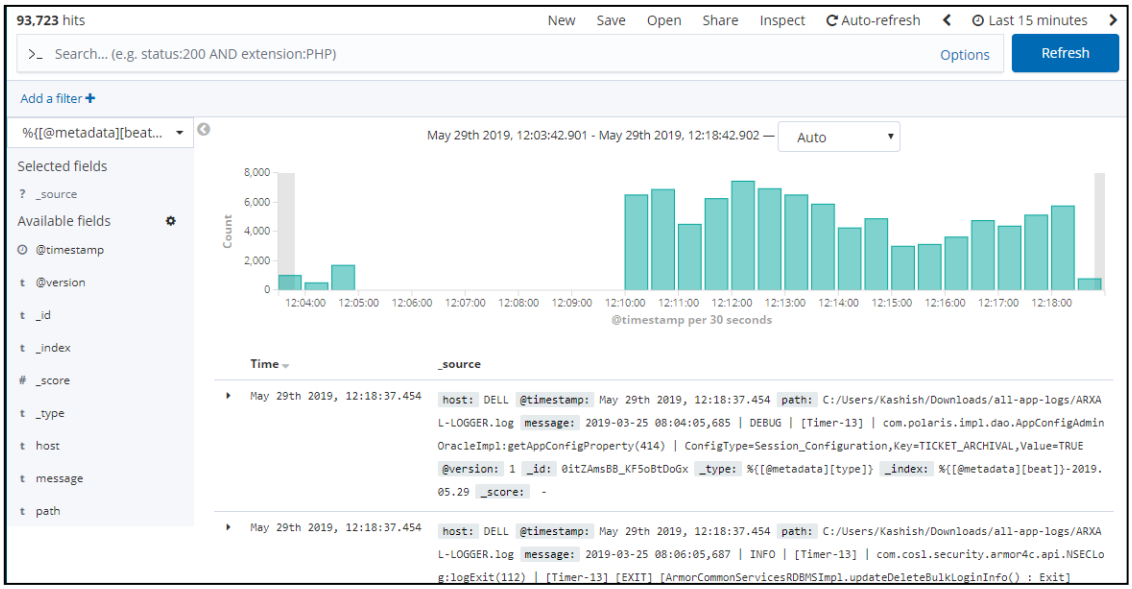

Figure 5.2 Discover page for LMS Logs in kibana

In the above diagram, the data stored in the elasticsearch can be discovered in Discover page. The data can be filtered using filters and by setting the time. The discover page has fields like timestamp, id, hostname, message, path etc.

5.3 Visualizing the data in kibana

For visualizing the data,”visualize” tab will be used. Visualization can be of any form, Bar graph, pie chart, heat map, data table etc. By clicking on the “visualize” tab, we can create the different visualizations.

Bar graph that counts the number of documents for LMS logs is as follows:

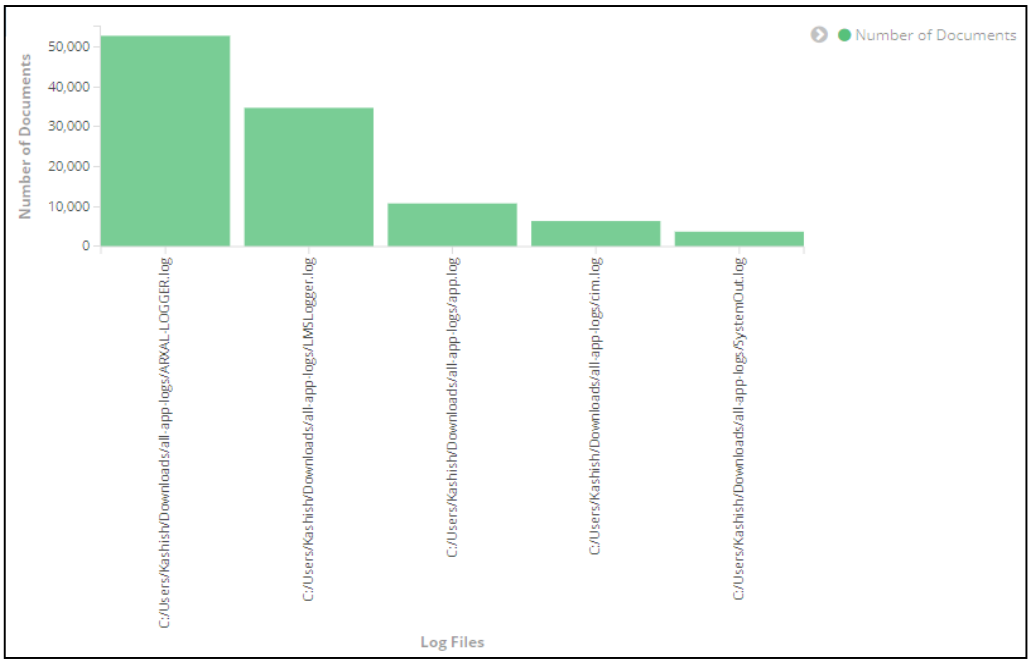

Figure 5.3 Bar Graph that counts number of documents in kibana

5.4 Creating the dashboard in kibana

For creating the dashboard in kibana, ”Dashboard” tab will be used. All the previous visualizations will be embedded in dashboard. A Kibana *dashboard* displays a collection of visualizations and searches. You can arrange, resize, and edit the dashboard content and then save the dashboard so you can share it.

5.4.1 Building a Dashboard

If you haven’t yet indexed data into Elasticsearch or created an index pattern, you’ll be prompted to do so as you follow the steps for creating a dashboard. Or, you can use one of the prebuilt sample data sets, available from the Kibana home page.

1. In the side navigation, click Dashboard.

2. Click Create new dashboard.
3. Click Add.
4. Use Add Panels to add visualizations and saved searches to the dashboard. If you have a large number of visualizations, you can filter the lists.
5. When you're finished adding and arranging the panels, go to the menu bar and click Save.
6. In Save Dashboard, enter a dashboard title and optionally a description.
7. To store the time period specified in the time filter, enable Store time with dashboard.
8. Click Save.

To import, export, and delete dashboards, see Managing Saved Objects. Dashboard for the LMS logs is as follows:

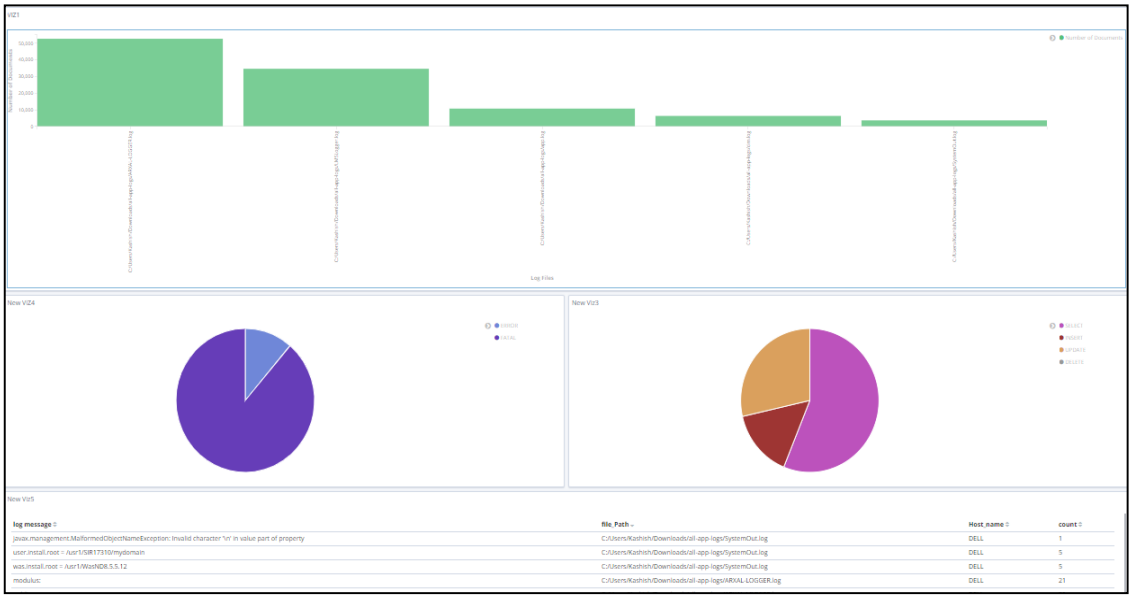

Figure 5.4 Dashboard for LMS Logs in kibana

The Dashboard above shows the bar graph, pie chart, Data table etc. for LMS logs.

5.4.2 Arranging dashboard Elements

The visualizations and searches in a dashboard are stored in panels that you can move, resize, edit, and delete. To start editing, click Edit in the menu bar.

- To move a panel, click and hold the panel header and drag to the new location.
- To resize a panel, click the resize control on the lower right and drag to the new dimensions.

5.4.3 Inspecting the visualization from the Dashboard

Many visualizations allow you to inspect the data and requests behind the visualization.

In the dashboard, expand the visualization's panel menu (or gear menu if in **Edit** mode) and select **Inspect**.

The initial view shows the underlying data for the visualization. To view the requests that were made for the visualization, choose **Requests** from the **View** menu.

The views you'll see depend on the element that you inspect.

5.4.4 Sharing a Dashboard

You can either share a direct link to a Kibana dashboard, or embed the dashboard in a web page. Users must have Kibana access to view an embedded dashboard.

1. Open the dashboard you want to share.
2. In the menu bar, click **Share**.
3. Copy the link you want to share or the iframe you want to embed. You can share the live dashboard or a static snapshot of the current point in time.

5.5 Embedding the dashboard in UI sample

A User Interface will be created in HTML/CSS and interactivity will be provided using javascript. Dashboard created in kibana will be integrated in UI sample.
UI sample for LMS logs is as follows:

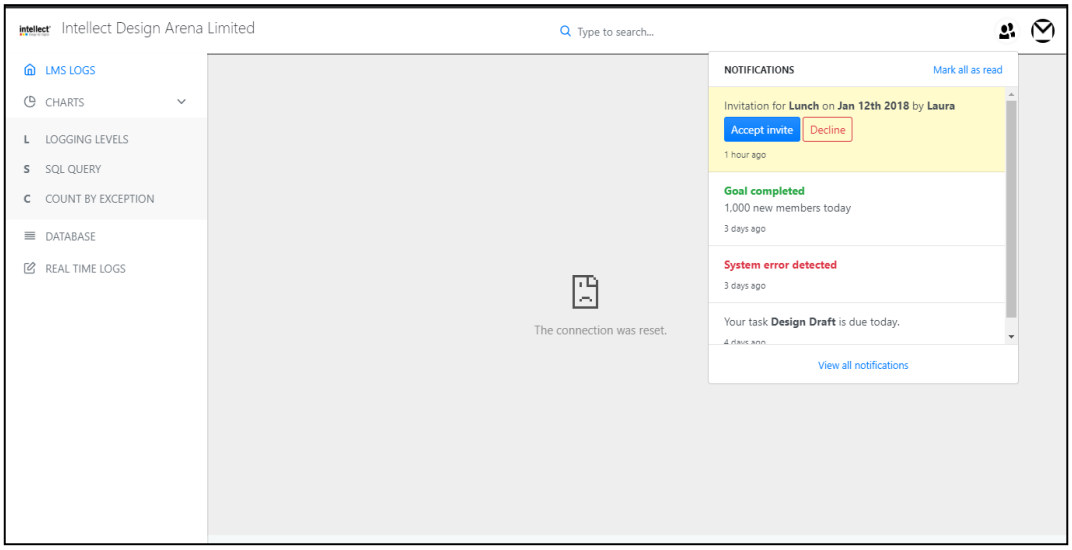

Figure 5.5 User Interface sample

Following are the explanation of each tab in User Interface:

LMS LOGS: This section of the UI will show the complete Dashboard with visualizations of LMS Logs.

LOGGING LEVELS: This section will show the logging levels like INFO, DEBUG, WARN, ERROR, FATAL, OFF, TRACE etc present in log files in the form of pie chart.

SQL QUERY: This section will show the occurrence of SQL queries like select, update, delete, insert in log file in the form of pie chart.

COUNT BY EXCEPTION: This section will count the type of exceptions in the form of bar graph where x-axis is the type of exceptions and y-axis is the number of such exceptions.

DATABASE: This section will show the database where the attributes are the path, host name, time stamp, message corresponding to the log files.

REAL TIME LOGS: This section will show the logs on the real time basis.

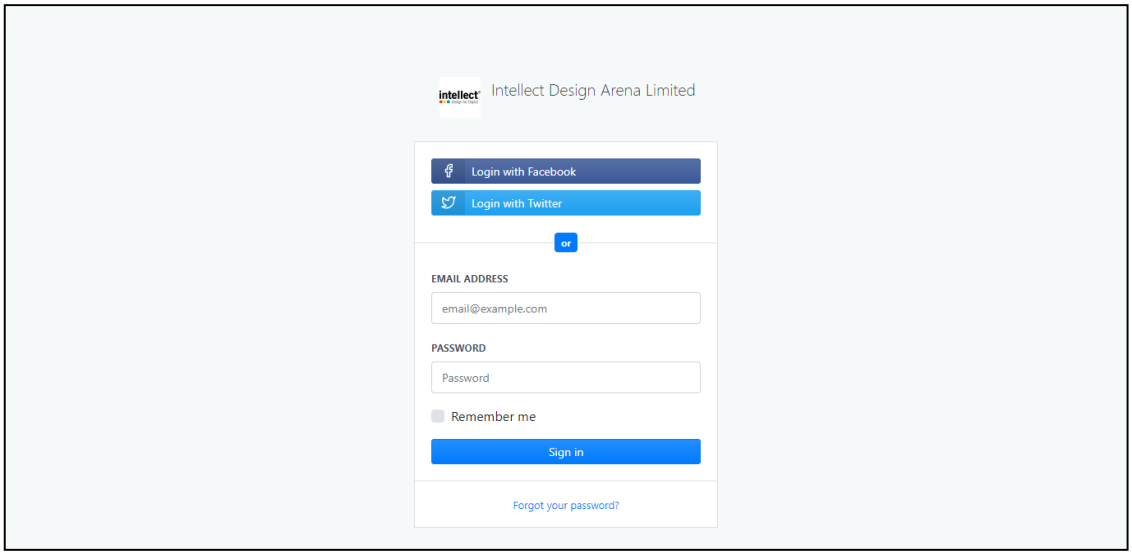

Figure 5.6 Login Page

The whole project will be added in Bit Bucket which is an open source platform for creating private repository for free. Other web based version control and collaboration platform can also be used for sharing and improving the code.

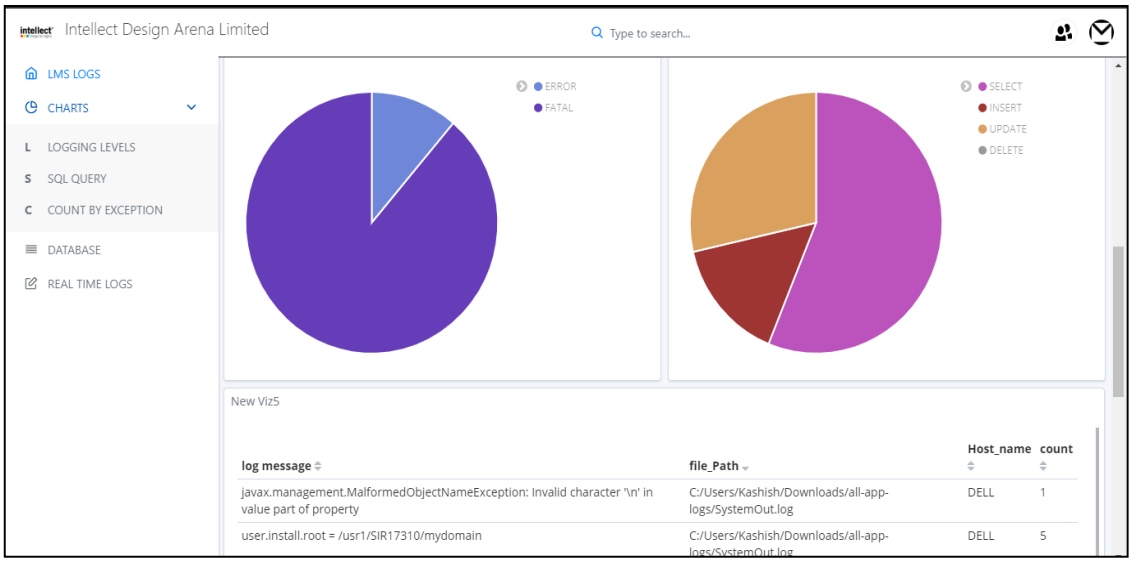

Figure 5.7 User Interface Sample with Embedded Dashboard created in Kibana

CHAPTER 6
CONCLUSION AND FUTURE SCOPE

A log is the automatically produced and time-stamped documentation of events relevant to a particular system. Virtually all software applications and systems produce log files. These files are generated in tonnes. The benefits of log files is to find out the health of a network, who has been accessing the network, how an application is performing etc. By embedding the dashboard created through kibana in UI sample, we are able to dynamically visualise the LMS logs for the company. In future, we will make our UI sample more interactive by adding more functionalities like chatbot, notifications, search queries. Also, we will explore other features of ELK stack.

REFERENCES

[1] Bajer, Marcin. "Building an IoT data hub with Elasticsearch, Logstash and Kibana." *2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*. IEEE, 2017.

[2] Bagnasco, S., et al. "Monitoring of IaaS and scientific applications on the Cloud using the Elasticsearch ecosystem." *Journal of physics: Conference series*. Vol. 608. No. 1. IOP Publishing, 2015.

[3] Reelsen, Alexander. "Using elasticsearch, logstash and kibana to create realtime dashboards." *Dostupné z: https://secure. trifork. com/dl/goto-berlin-2014/GOTO_Night/logstash-kibanaintro. pdf* (2014).

[4] Langi, Pingkan PI, Warsun Najib, and Teguh Bharata Aji. "An evaluation of Twitter river and Logstash performances as elasticsearch inputs for social media analysis of Twitter." *2015 International Conference on Information & Communication Technology and Systems (ICTS)*. IEEE, 2015.

[5] Hamilton, James, et al. "SCADA Statistics monitoring using the elastic stack (Elasticsearch, Logstash, Kibana)." (2018): TUPHA034.

[6] Bagnasco, S., et al. "Towards Monitoring-as-a-service for Scientific Computing Cloud applications using the ElasticSearch ecosystem." *Journal of Physics: Conference Series*. Vol. 664. No. 2. IOP Publishing, 2015.

[7] Gupta, Yuvraj. *Kibana Essentials*. Packt Publishing Ltd, 2015.

[8] Andreassen, Odd, Alicia De Dios Fuente, and Cedric Charrondière. "Monitoring Mixed-Language Applications with Elastic Search, Logstash and Kibana (ELK)." (2015): WEPGF041.

[9] http://delphi.intellectdesign.com/login/index.php