UNIVERSITÉ
Concordia
UNIVERSITY

ENCS 6921 - Industrial Stage and Training

Winter 2023

January Monthly Report

Literature Survey

Submitted By:
Manan Dineshkumar Paruthi - 40192620

Submitted To:
Dr. Rajagopalan Jayakumar
Director, Co-op Program, Associate Professor, Computer Science and Software Engineering

Introduction & Problem Statement

Before starting my Winter 2023 Coop, I had a discussion with my manager about the requirements of this course i.e ENCS 6921 Industrial Stage and Training and what can be done to aim to work above and beyond the scope of my internship.

So, he informed me that application logging is very important and critical in the industry, for Identifying and debugging issues in the production environment. As it contains sensitive and personal customer data, live debugging can not be done. Also, analysis of logs data helps us to gain information which is important to improve the quality of the applications.

Currently, the company is using Splunk and they are willing to migrate it to ELK Stack (Elastic Search, Logstash & Kibana) as it is open source so free of cost while Splunk is costly and requires a license.

But the issue with ELK Stack is that it is more difficult to set up and maintain. So, it would be great if I can make a POC (Proof Of Concepts) on the same which can be demoed to the senior management and based on that we can start the migration to ELK Stack.

The main reasons for migration are :

1) Cost: ELK Stack is open-source, which means it is free to use, whereas Splunk is proprietary software and requires a license, it makes ELK a more cost-effective option for organisations and helps them to reduce the budget expenditure in this economic crisis and layoff environment.
2) Flexibility: ELK Stack is highly customizable and can be tailored to specific needs, whereas Splunk's capabilities are more limited. With ELK, organizations can choose to use different components and add-ons, and can also create custom plugins to extend their functionality
3) Data Ownership: With ELK Stack, organizations have more control over their data and can decide where it is stored and how it is processed, whereas Splunk stores data on its own servers.

What is ELK Stack ?

The ELK Stack, consisting of Elasticsearch, Logstash, and Kibana, is a popular open-source solution for application logging and data analysis. The stack is designed to collect, process, and analyze log data, providing organizations with valuable insights into their systems and applications.

What problems are solved by ELK Stack ?

One of the main problems that the ELK Stack addresses is the ability to search, analyze, and visualize large volumes of log data. Elasticsearch, the core component of the stack, is a powerful search and analytics engine that allows for fast and efficient data querying. Logstash, another component, is a data processing pipeline that can normalize, enrich, and process log data before it is indexed in Elasticsearch. Kibana, the final component, is a visualization tool that allows users to create interactive dashboards and charts to analyze the data.

What are the advantages of ELK Stack ?

The ELK Stack has several advantages that make it a popular choice for application logging and analysis. One of the main advantages is its open-source nature, which allows for greater flexibility and customization. Additionally, the stack has a large and active community, which provides a wealth of resources and support. The stack is also highly scalable and can handle large volumes of data, which makes it well-suited for organizations with large and complex systems.

What are the disadvantages of ELK Stack ?

However, the ELK Stack also has some drawbacks. One of the main cons is that it can be more difficult to set up and maintain compared to other solutions. Additionally, the stack does not have a built-in alerting system, which can make it difficult to detect and respond to issues in a timely manner.

What is the demand for ELK Stack in the industry ?

The demand for the ELK Stack in the industry is high, with many organizations using the stack for application logging and analysis. The stack is well suited for organizations that have large and complex systems and that need more control over their data. Additionally, the stack is popular among organizations that are looking for an open-source solution that can be customized to their specific needs.

Comparison of ELK Stack with Splunk

When compared to other solutions, the ELK Stack is often compared to Splunk, which is a proprietary log management and data analysis tool. While both solutions offer similar capabilities, the ELK Stack is open-source and offers greater flexibility and customization. However, Splunk is known for its ease of use and intuitive interface and offers a wide range of pre-built visualizations and alerts. Both solutions have their own strengths and weaknesses, and the choice between them will depend on the specific needs and resources of the organization.

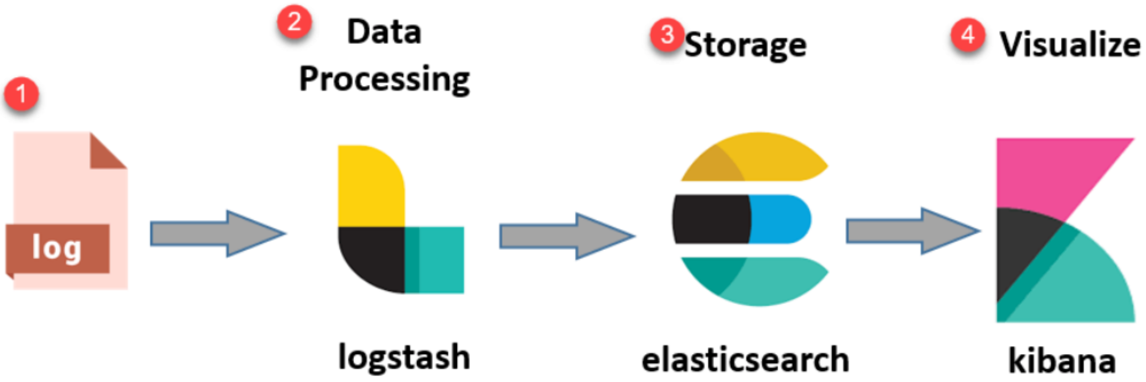
Solution

This project is a web application that utilizes the MEAN stack (MongoDB, ExpressJS, AngularJS, and NodeJS) for the front-end and back-end development, and ELK Stack (Elasticsearch, Logstash, and Kibana) for log management and data analysis.

The application will provide a platform for users to create, read, update, and delete (CRUD) data related to a specific topic. The front end will be built using Angular, which will provide a seamless user experience, and the back end will be built using NodeJS, ExpressJS, and MongoDB, which will provide a robust and scalable framework for data storage and management.

The application will also make use of ELK Stack, which will be used to collect, store, and analyze log data from the application. Logstash will be used to collect log data from various sources, Elasticsearch will be used to store the log data, and Kibana will be used to provide visualizations and analytics.

This application will provide a comprehensive solution for data management and log analysis, making it easier for organizations to gain insights from their data and improve their operations. The main advantages of this application are that it uses technologies that are widely adopted in the industry, it provides a great user experience, it's easy to maintain, it's easy to scale, and it provides great visualization and analytics capabilities.



Project Scope

The MEAN Stack application (MongoDB, ExpressJS, Angular, NodeJS), a full stack application with MongoDB as NoSQL database, Angular as Frontend and ExpressJS & NodeJS as Backend, which will have create account, login & logout features along with logging functionality. And logs parsing and analysis functionality by using ELK stack.

It will simulate an industry-level application environment which includes an application along with logging, parsing, and analysis capabilities using open-source technologies.

Technologies

The MEAN stack refers to a collection of JavaScript-based technologies that are commonly used for building web applications. The acronym stands for MongoDB, ExpressJS, AngularJS, and Node.js.

MongoDB is a NoSQL document-oriented database that stores data in a JSON-like format. It is known for its scalability and performance and is often used in web applications that need to handle large amounts of data.

ExpressJS is a web application framework for Node.js that provides a set of features for building web applications, such as routing and middleware.

Angular is a JavaScript framework for building dynamic, single-page web applications. It allows developers to create reusable UI components and also provides a powerful data binding mechanism.

Node.js is a JavaScript runtime environment that allows developers to run JavaScript code on the server side. It is known for its speed and performance and is often used for building high-performance web applications.

The ELK Stack, on the other hand, is a collection of open-source tools for log management and data analysis. The acronym stands for Elasticsearch, Logstash, and Kibana.

Elasticsearch is a powerful search and analytics engine that stores and indexes large volumes of log data. It allows users to perform fast and efficient data querying.

Logstash is a data processing pipeline that can normalize, enrich, and process log data before it is indexed in Elasticsearch. It allows users to manipulate and process log data before it is stored in the Elasticsearch cluster.

Kibana is a visualization tool that allows users to create interactive dashboards and charts to analyze the data. It provides a web interface for exploring and visualizing log data indexed in Elasticsearch.

Both the MEAN stack and ELK stack are used to build and analyze web applications, but they have different purposes and technologies. The MEAN stack is used for building web applications, while the ELK stack is used for log management and data analysis.

Conclusion

Coop point of view :
A Proof Concept will be made on ELK Stack which will help the company to migrate from Splunk to ELK Stack. It will be above and beyond the scope of my internship.

Professional Growth point of view :
It will help to learn technologies which a Full-Stack developer is expected to know in order to get a full-time job and learn ELK stack which is an open-source technology, due to which many companies are moving to it from costly tools like Splunk, New Relic, etc and many job opening’s descriptions state about it too. It will also give me a chance to showcase my this technical skills with a live demo to my internship company manager in order to give a positive & strong image which will help me to convert my coop to a full-time job opportunity on top of learning very important technical skills which are considered hot technologies in the market.