IUPD (Intelligent University Performance Dashboard)

NGO CHEN BANG

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

IUPD – INTELLIGENT UNIVERSITY PERFORMANCE DASHBOARD

NGO CHEN BANG

This report is submitted in partial fulfillment of the requirements for the Bachelor of Computer Science (Software Development)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2017

DECLARATION

I hereby declare that this project report entitled

IUPD – INTELLIGENT UNIVERSITY PERFORMANCE DASHBOARD

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT : _____ Date: _____

(NGO CHEN BANG)

SUPERVISOR : _____ Date: _____

(PN. NOOR AZILAH MUDA)

DEDICATION

To my beloved parents. teachers and friends.

ACKNOWLEDGEMENTS

I would like to show my greatest appreciation to my beloved parents who have been giving me support and motivation throughout my project.

Apart from the efforts of me, the success of this thesis depends largely on the encouragement and guidelines from my supervisor. I would also like to thank Pn. Noor Azilah for giving me assistant to complete this project successfully.

Last but not least, for those who are not included, which help me formally or informally, I surely very grateful for all the help and assistance I could get. Thank you.

ABSTRACT

Intelligent University Performance Dashboard is a set of modern intelligent analytic tools in the form of RESTful web service, namely an intelligent analytic dashboard, an analytic engine and a flexible analytic NoSQL database aimed to assist the higher education institutions in Malaysia to improve their university performance rating which is evaluated through MyRA assessment. MyRA is an acronym for the Malaysia Research Assessment which is a comprehensive system developed to assess the research capacity and performance of all higher education institutions in Malaysia. The problem statement for this project is that some higher education institutions in Malaysia are having a hard time trying to improve their rating in MyRA assessment so they might be interested in a system that can assist them in any decision-making for improving the performance rating of university. Therefore, the objectives of this project are to design an intuitive browser based dashboard capable of visualizing both the raw MyRA data and the result from analysis for exploration purpose, to investigate the best way to build a scalable analytic engine capable of processing large data sets efficiently to identify the most significant cause for a problem using causal effect analysis, to design a reliable database that can store a huge amount of raw MyRA data and at the same time can run basic analytic operations on them and to design a robust RESTful architecture to facilitate the communication among the tools above through their REST service API. Academically, this project can be considered as one of the reliable means to help improving university performance. It can visualize the MyRA data efficiently which allows more accurate critical decision making from the higher education institutions as they can explore through and obtain insight from their MyRa data. The causal effect analysis functionality provided by this project also allows HEIs to obtain valuable insights into the reasons why their universities are doing badly in certain area and provide suggestions on how to improve these areas. Besides, this causal effect analysis can also be employed on other domains such as business, agriculture, marketing, healthcare and so on.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

UTeM	Universiti Teknikal Malaysia Melaka
IUPD	Intelligent University Performance Dashboard
REST	Representational State Transfer
WWW	World Wide Web
HTML	HyperText Markup Language
API	Application Programming Interface
DOM	Document Object Model
WebGL	Web Graphic Library
CSS	Cascading Style Sheet
XHTML	Extensible HyperText Markup Languange
SPA	Single Page Application
MPA	Multi Page Application
URL	Unified Resource Language
JSON	Javascript Object Notation
HEI	Higher Education Institution

CHAPTER I

INTRODUCTION

1.1 Introduction

IUPD is a set of modern intelligent analytic tools comprising of a browser based analytic dashboard interface, data processing analytic engine and NoSQL database aimed to assist the HEIs in Malaysia to improve their university performance rating which is evaluated through MyRA assessment.

MyRA is an acronym for the Malaysia Research Assessment which is a comprehensive system developed to assess the research capacity and performance of all HEIs including the research university, comprehensive university and focused university in Malaysia. Basically, HEI is evaluated by MyRA through a set of criteria and the data for those criteria must be collected and provided by the HEI itself according to the requirements stated in the MyRA guidelines. However, the HEI will undergo document-audit and site-audit by a panel of trained auditors every year to ensure that the data provided by the HEI is valid and the HEI is being assessed by

MyRA thoroughly and accurately. Some examples of the main criteria used in MyRA are quantity and quality of researchers, quantity and quality of research, quantity of postgraduate, quality of postgraduates, innovation and so on. In short, there is a total of nine main criteria in MyRA where the final score of the assessment is based on the sum of the score from each of these main criteria and the calculation of the score for each of main criteria is also stated in the MyRA guideline. Under each main criterion, there are children criteria in which some of the children criteria may, in turn, consists of their own children criteria and these children criteria have its own mark that contributes to some part of the score calculation.

MyRA was first developed in 2006 but it is considered as stable only recently. This is because all HEIs in Malaysia is compulsory to participate in this annual assessment exercise starting in 2014 and the format of the assessment has been stabilized. This means there is only around two to three years of recent years' data available for the analytic purpose for most of the HEIs. Since MyRA is becoming one of the main assessment tools for the HEIs in Malaysia to evaluate their performance rating and research capability, a lot of effort has been invested by both public and private HEIs to improve their current star rating or even trying to attain the highest 6-Star rating in MyRA assessment.

Generally, there are a lot of ways that can help an HEI to achieve better MyRA rating but one of the most effective, quick and convenient way is definitely through the use of Information Technology which is also the main purpose of this IUPD project. This project is going to leverage the combination of various fields, techniques and latest technologies available in the market such as Intelligent Information Systems, Business Intelligence, Software Technology, Database Technology, Computer System Technology, Distributed Computing and so on to develop a set of intelligent analytic tools versatile enough to be used and employed by every Malaysia's HEI as one of their means to improve the MyRA rating. However, for the purpose of project development and testing, the MyRA data from UTeM will be used.

For an HEI to improve its overall rating in MyRA assessment, one of the most efficient strategies is that the HEI must focus on improving its score rating continuously for every main assessment criterion used by MyRA. So, it is critical to understand the HEI performance in each of the main criterion for every year. For

example, the trend for each of the main criteria in the recent years, the comparison of the main criteria score between two selected years, the main criteria in which the HEI is excel at in the recent years (strength), the main criteria in which the HEI is bad at in the recent years (weakness) and so on. It is hard to get all this information by just looking at the raw MyRA data which is usually stored in an excel-formatted for CSV file. IUPD can help the HEIs to overcome this problem by providing an excellent analytic dashboard capable of data exploration and visualization. Besides, some of the HEIs may not be satisfied with the data exploration functionality only. They may be more interested in what information or insights that they can further extract from their current set of raw MyRA data. IUPD also offers data analysis functionality to perform operation like prediction, rules mining and so on. Since the number of years of MyRA data available for analysis is quite limited for some of the HEIs, some analytic operations may not be possible in these HEIs.

1.2 Problem Statement

The problem statement for this project is some HEIs in Malaysia are having a hard time trying to improve their rating in MyRA assessment. They might be interested in an efficient, cost-effective and convenient way in the form of information technology system in which they can rely on to improve the performance rating of a university. This system can visualize the MyRA data from year to year to feature easier data exploration so any decision making can be done accurately based on the visualized data. By using the same dataset, many analysis techniques can also be done such as finding the reasons that are hindering the improvements for one of the main criteria, looking for potential improvement area for one of the main criteria and so on.

1.3 Objective

This project embarks on the following objectives:

- To design an intuitive browser based analytic dashboard capable of visualizing both the raw data for exploration and the data resulted from an analysis efficiently so the HEIs can gain more valuable insight and understand better of their MyRA data.
- To investigate the best way to build a scalable analytic engine that is fast and efficient in processing large data set in the form of web service and provide the capability of RESTful communication.
- To design a reliable and flexible database that can store a huge amount of raw MyRA data and at the same time run any query and analytic operation on them efficiently.
- To assess the possibility of allowing the more advanced users such as developers to interact directly with the analytic engine service through RESTful API with a specific authorization mean.

1.4 Scope

This project focuses on developing three main analytic tools

- Browser based analytic dashboard that is responsible for visualizing data through the use of the chart, graph, table and so on to let the HEIs explore and understand their data easily and act as the client interface in which can be used to control the operation of other tools.
- Analytic engine service that is responsible for running analysis algorithm on the provided data to extract valuable insights and providing the analytic functionality through RESTful API.

- NoSQL database that is responsible for storing mostly unstructured MyRA data collected by the HEI based on the requirements state in MyRA guideline in the way that allows any query and analytic operation to be carried out smoothly.

1.5 Project Significance

The project significance can be divided into two sub-topics which are academic significance and industrial significance.

Academically, this project can be considered as one of the reliable means to improve the university performance. It can visualize the MyRA data efficiently which allows critical decision making from the HEIs as they can explore through and understand their MyRA data better. The analytics functionality provided by this project will also allow HEIs to obtain valuable insights into the reasons why their universities are doing badly in some of the main criteria while excel at the rest of the main criteria.

Industrial significance can be explained for future business decision making. This is all due to the importance of the intelligent analytics in business where they will encourage more effective communication with current and potential customers of an organization, identify opportunities in their current business area, minimize investment risks, help to plan by uncovering and identifying potential problems in business processes, establish markets trends, set up more robust market positioning and a lot more of benefits.

Furthermore, the authorities can use the analytics engine service in this project to assist in any of their decision making especially if the data collaborates with the unstructured data. This project is also useful in terms of natural disaster issues because it can give alert to the certain rescues to the most needed place, or the media company can cover as many news as possible at the same time.

1.6 Expected Output

The expected result of this project is that an analytic tool leveraging a combination of modern technologies like Scala, Akka, NoSQL database, RESTful architecture, Javascript view library React and so on will be developed to help higher education university in Malaysia to improve their rating in MyRA assessment. The success of this project will also prove the fact that the computer technology nowadays is already mature and versatile enough to simplify and support any complex human works such as storing complex unstructured data, visualizing data, analyzing data for insights and assisting in decision making.

1.7 Conclusion

To conclude this chapter, the main purpose of this project is to develop a set of analytic tools comprising of the latest technology available in the market to serve as one of the most convenient, cheapest and efficient ways for the HEIs in Malaysia to attain better rating in the MyRA assessment. In the next chapter, the literature review and project development methodology used will be discussed.

CHAPTER II

LITERATURE REVIEW AND PROJECT METHODOLOGY

2.1 Introduction

This chapter will discuss in depth the literature review of the IUPD project which includes several facts and findings necessary to uncover the important concepts, theories and knowledges related to the system and system domain involved throughout the project development. Besides, the project development methodology employed in this project will be introduced which include the general description about the methodology, the number of phases involved in the methodology, description for each phase and the overview of activities carried out in each phase. Finally, project requirements including software and hardware requirements will be listed down in detail together with the planned project development schedule and milestone.

2.2 Facts and Findings

The domains related to this project are web application, single page application, RESTful architecture, NoSQL database and intelligent analytics technology. The explanation is included in the sections, from 2.2.1 to 2.2.6.

2.2.1 Web Application

The analytic dashboard of this project will be developed in the form of web application. This section discusses in detail the emergence of web application, the definition of modern web application, how does web application work, the common usages of web application, the business advantages of web application.

2.2.1.1 The Emergence of Web Application

During the days where Internet was first introduced, the World Wide Web comprised of only sites which also known as information repositories which store static documents and the invention of web browsers was one of the means to retrieve and display those documents. All these processes of accessing static documents represents the one-way flow of information from the server to the browser. Since the information held on the server was usually not sensitive and already open to the public view, most sites did not practice any authentication nor authorization on their users and each of their users was treated similarly by showing a similar view of information (Stuttard & Pinto, 2011).

The other type of client-server computing model in the early day was each of the users must install the application own pre-compiled client program on their personal computer in order to have access to the application's user interface and all

the processing works for the application are typically distributed between code located on the server and code installed separately on each client machine. Any upgrades to the server-side code of the application would require huge support cost and effort since all the client-side code installed on each user workstation must be upgraded too. In addition, the components from both the server and the client of the application were usually tightly coupled to a certain type of computer technology and architecture and migrating them to others was often tremendously expensive.

In the time when each individual web page was served to the browser client as static document, interactive experience could still be provided when the user input was returned to the server using HTML form elements embedded in the HTML markup of the static documents. However, for this to happen, a round trip back to the server would be required even for the tiniest change to refresh the page.

The evolution of web application started in 1995 when a client-side scripting language known as Javascript was introduced by Netscape which allows dynamic elements to be added to the client browser interface. These embedded dynamic elements can perform various powerful functions such as input validation, allow simple animations like showing or hiding part of the web page and so on where most of these functions must trigger round trip of sending data back to the server, generating new web page and delivering the web page to the client's browser previously. In 1996, the introduction of Flash, a vector animation player that could directly embed complex animations on the web page by Macromedia opens the possibility of providing interactions on the browser client by using scripting language entirely without having to communicate with the server. However, Flash must be added to the browser as a plug in to make all these to happen which at the same time can be considered as a drawback because plug in is known to slowing down the browser.

In 2005, the introduction of Ajax allows scripting language to be able to contact the server directly for storing or retrieving data which is impossible previously. This has eliminated the need for the client browser to download and refresh the entire web page just to make a small change to happen. The finalization of HTML5 specifications in 2011 provides multimedia capabilities and enriches the semantic content of the documents on the client's browser through its clean APIs and DOM without the need of plug in. Furthermore, the WebGL API which is also one of the

fundamental parts of the HTML5 specification introduces the possibility of advanced 3D graphics on the browser.

In short, these evolutions have significant impacts in creating a modern web application widely used nowadays which is truly platform and browser independent.

2.2.1.2 The Definition of Modern Web Application

In general, web applications can be defined as software applications that resides on a web server where it will be served to the website visitors through any of their preferred browser and they can use it to submit and retrieve data to or from a database over the Internet.

In a more technically oriented way, when the users initially visit and access a web application using their browsers, the web application will query the content server, also sometimes known as content repository database and dynamically generate a set of web documents that will be served to the users. The generated documents are usually in the form of HTML or XHTML with CSS which is a widely accepted standard format that can be interpreted and presented in all types of browsers easily. With the help of Javascript and Ajax, any further users interactions with the web applications, communications such as data requests from the web server can be done reloading whole web pages because when the requested data is received via Ajax asynchronously, the Javascript scripting language can then uses the DOM to dynamically update selected regions of the web page based on the shape of data fetched, allowing an intuitive and interactive user experience. (O'Reilly, 2005)

In short, by using these simple techniques, web developers can make their web applications to mimic those functionalities usually found in a desktop application, such as word processing, image processing, data processing, and slide-show presentation. These web applications can then be rapidly deployed on any web server and without any installation tasks and special requirements at the user's end.

2.2.1.3 How does a Web Application Work?

The following Figure 2.1 illustrates the components and the relationship between the components found in the simplest web application architecture, namely the three-layered model. The first layer is normally the user interface or a browser, the second layer is the web server where the web application resides and it is usually written in server-side languages such as PHP, Java EE, ASP.NET, Node.js and so on. The third layer is the database used to store important data and content that can be retrieved by the user later.

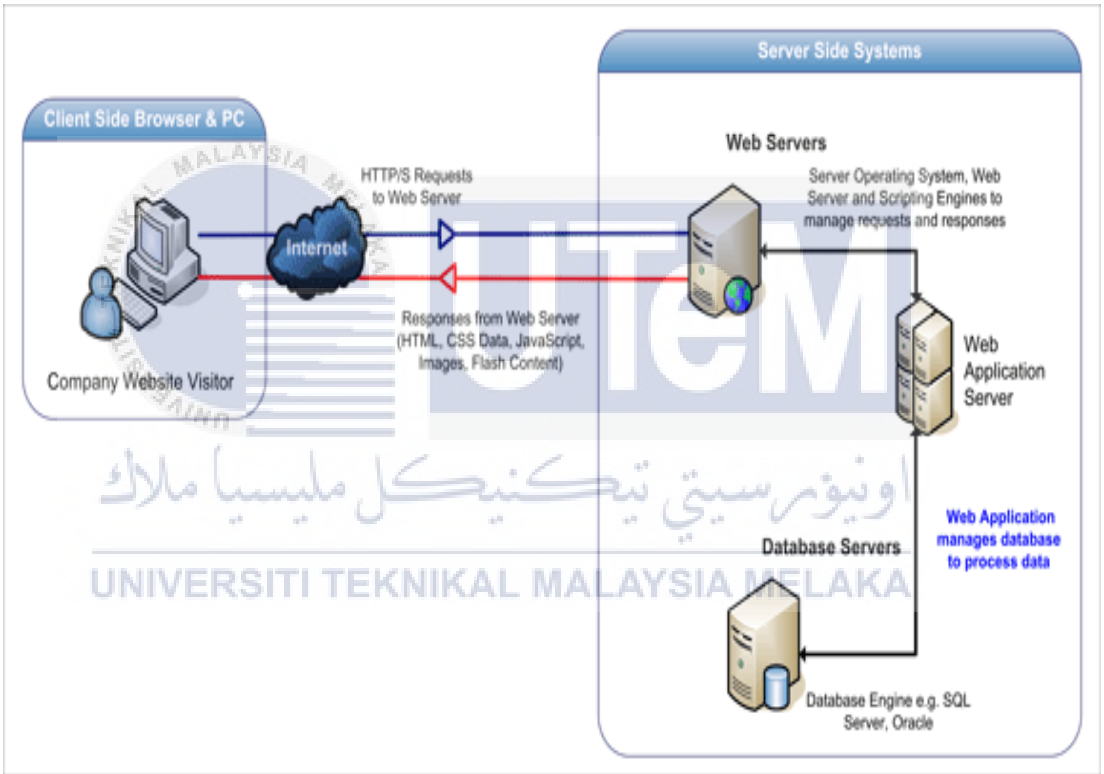

Figure 2.1: Three-Layered Web Application Model (Web Application Structure, 2012)

The following Figure 2.2 illustrates how the web application server handles the request triggered by the users through their browsers over the Internet. The web application server then processes the requests received from the users to determine the type of operations that needs to be done. If any data storing, updating or retrieving operations are required to complete the users' request, the web application will access