ST. XAVIER'S COLLEGE

(Affiliated to Tribhuvan University)

Maitighar, Kathmandu

FINAL YEAR PROJECT REPORT
ON
"SCALABLE BUSINESS INTELLIGENCE FOR MEDIUM SCALE ENTERPRISES"
[CSC 404]

For the partial fulfillment of Bachelor's Degree in Computer Science and Information Technology awarded by Tribhuvan University

Under the Supervision of
Mr. Vishnu Kumar Rana
Head of the Department
Department of Computer Science

Submitted by
Dixit Bhatta (T.U. Exam Roll No.1455)
Dirban Singh Lama (T.U. Exam Roll No.1454)

Submitted To

ST.XAVIER'S COLLEGE

Department of Computer Science
Maitighar, Kathmandu, Nepal
August, 2015

SCALABLE BUSINESS INTELLIGENCE FOR MEDIUM SCALE ENTEPRISES

[Course Code: CSC 404]

Submitted by

Dixit Bhatta (T.U Exam Roll No. 1455)

Dirban Singh Lama (T.U Exam Roll No. 1454)

A Final Year Project Report submitted in partial fulfillment of the requirements for the Bachelors degree of Science in Computer Science and Information Technology awarded by Tribhuvan University

Submitted To

ST. XAVIER'S COLLEGE

DEPARTMENT OF COMPUTER SCIENCE

Maitighar, Kathmandu, Nepal

August, 2015

ST. XAVIER'S COLLEGE
MAITIGHAR, KATHMANDU, NEPAL
Post Box :7437
Contact: 4221365,4244636
Email: ktm@sxc.edu.np

सेन्ट जेभियर्स कलेज
माईतीघर, काठमाडौं, नेपाल
पो.ब.नं. : ७४३७
फोन : ४२२१३६५,४२४४६३६
ईमेल : ktm@sxc.edu.np

CERTIFICATE OF APPROVAL

The undersigned certify that they have read and recommended to the Department of Computer Science for acceptance, the project report entitled "**Scalable Business Intelligence for Medium Scale Enterprises**", submitted by **Dixit Bhatta (TU Roll No. 1455)** and **Dirban Singh Lama (TU Roll No. 1454)** in partial fulfillment for the degree of **Bachelors of Science in Computer Science and Information Technology** awarded by Tribhuvan University.

.....

Mr. Vishnu Kumar Rana
Supervisor
Department of Computer Science
St. Xavier's College
Maitighar, Nepal

.....

External Examiner
Tribhuvan University

.....

Mr. Vishnu Kumar Rana
Head of the Department
Department of Computer Science
St. Xavier's College
Maitighar, Nepal

ABSTRACT

Business Intelligence is a popular concept in global market today. Many large enterprises frequently use a number of products that help in business analysis and decision making. However, such systems prove really costly for medium scale enterprises. One of the main reasons for this is that the enterprises have to pay for the tools whose many features are either unnecessary or never used. Those enterprises have the potential to grow and need good decision support for that, but affordability comes as an important issue. Due to this, the current practices are mostly limited to analysis of manually created charts in meetings and boardroom discussions.

This project explores the tools and techniques that can be used to build Business Intelligence system for such enterprises which is both inexpensive and scalable when the enterprise grows. With simple data analysis methods and integration of various low cost components, a stable and powerful system can be created. Hence, the project discusses different statistical techniques, and free and open-source platforms which can be used to develop a secure, interactive and sustainable Business Intelligence system that enables data visualization and predictive analysis.

Keywords: Business Intelligence, Scalability, Affordability, Data Visualization, Predictive Analysis

ACKNOWLEDGEMENT

We would like to express our deepest appreciation to all those who provided us the possibility to complete this project. We give special gratitude give to our Project Supervisor and Head of the Department, **Mr. Vishnu Kumar Rana**, whose contribution in initiation and encouragement helped to coordinate the project especially in understanding the real life scenario of the project.

Furthermore we would also like to acknowledge with much appreciation the crucial role of, **Er. Rajan Karmacharya**, whose role as the mentor right from the project proposal phase was influential in making things easier and methodical for us to follow. Many thanks go to **Mr. Jeetendra Manandhar**, Lecturer at Department of Computer Science, who worked in background to help and encourage us.

We have to appreciate the guidance given by all other involved parties, especially our subject lecturers **Mr. Sansar Dewan, Mr. Nitin Malla, Er. Anil Shah, Mr. Saugat Sigdel** and **Mr. Balkrishna Subedi** for our project which has improved our research and presentation skills. Thanks to their comments and advices.

Dixit Bhatta (1455/068)

Dirban Singh Lama (1454/068)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

BI	Business Intelligence
ERP	Enterprise Resource Planning
CART	Classification and Regression Trees
CBS	Case-Based Reasoning
NoSQL	Not only SQL
JVM	Java Virtual Machine
JRI	Java/R Interface
SCM	Source Code Management
WAR	Web application ARchive
CSF	Critical Success Factors
ETL	Export, Transform and Load
VCS	Version Control System

CHAPTER 1: INTRODUCTION

1.1 Background

Business Intelligence is a popular term among the business elites today and many companies are continuously being attracted toward it. It is basically a set of tools that aid businesses in taking decisions by drawing information from raw data [1]. They enable summarization and data visualization to provide quick insights from even unstructured data. The importance of business intelligence is further consolidated by the fact that there are a number of ways it can used to improve a company's revenues. Aspects such as predictive analytics, social sentiments and location can be interpreted by business intelligence systems to create specific strategies that boost profit [2].

Likewise, in present global context where companies are growing rapidly and entrepreneurs making the most from their startups, medium scale enterprises are the ones that have real potential to turn into big players. Hence, they can be benefitted a lot from business intelligence tools. But again, problems arise due to the fact that these companies are not stable like huge enterprises. So, a scalable solution is needed for them that can accommodate both growth and affordability when they shrink due to market fluctuations [3]. The system should be dynamic enough to serve the changing business environment. Thus, for medium scale enterprises, business intelligence systems should address scalability without the system being too expensive to sustain.

1.2 Problem Definition

When we talk about business intelligence tools, the first and foremost thing that comes in people's mind is that they are for only big enterprises. Clearly, the price imposed on such products justifies the fact that are they are indeed developed and marketed by keeping large enterprises in mind. But what about the medium scale enterprises those want to be competitive and use the technology to mature themselves. Even if they decide to buy some products, they might not require a number of features that add to the cost. Investing on just a single hot feature is never a good idea for medium scale businesses [4].

So, our motivation is to find a solution that provides features such as data summarization, data visualization and predictive analysis to medium scale businesses without adding significant installation and operational costs. As we are also moving into the age of Big Data, organizations will continuously generate data. The bigger the data the better will be predictions [5]. So, the solution also addresses the issue of scalability and serves as a platform that will support growing data no matter the organization grows or declines.

The solution seeks to build upon the existing resources at the enterprises. As there might not be enough data at the medium scale enterprises, the techniques will be centered on finding as precise results as possible from limited data. In the same way, the tools used for development will be based on cost-effective tools, but they are also suited for enterprise level solutions. All in all, it is about providing the medium scale enterprises reliable decision support that is scalable and sustainable.

1.3 Scope and Objectives

As suggested by the project name itself, the specific targets of the project are medium scale enterprises. By medium scale enterprises we mean businesses that bridge the gap between large and small enterprises. They own most of their resources and have fair presence in the market. At the moment, the application is not fully generic and hence cannot be readily used in all medium scale organizations. Nonetheless, regardless of the data practices in the organization, the kind of systems they use, and the domain they represent, the solution can be used to generate insights from data after minor migration and modification steps. Though developed specifically using data provided by a single institution, the application can be rapidly reprogrammed and configured to be usable at other organizations with similar needs.

The key objectives further clarify the intent of the project which can be listed as:

1. Identifying key factors that assist medium scale enterprises in business decisions
2. Migrating the data to a scalable database system
3. Finding suitable statistical methods for precise predefined and predictive analysis
4. Enabling data summarization and quick insights from data

5. Creating an interactive, platform-independent system that helps in visualization of the analyzed results, summaries and insights.

So, the project aims to develop a solution through a series of related steps.

1.4 Project Features

The project is based on research on suitable data analysis techniques that help in precise decision support to medium scale businesses. The research includes data analytics along with scalability issues. The end-product is a fully functional web application that provides business intelligence features such as data summarization, data visualization, pre-defined analysis and predictive analysis. The web application has all relevant components of a good business intelligence system such as dashboard, charts, reports, and guided application tour. As per the research motivation, the application is platform independent and scalable.

The application is platform independent because it can run equally well on the most popular Operating System families i.e. Windows and Linux. This is further supported by the fact that the application runs on Apache Tomcat, which provides good performance across different platforms [6]. In the same way, MongoDB enables this application to operate on any scale of data. It has low-latency, high throughput and continuous availability as its main strengths [7]. Hence, the application has all it takes to be a good business intelligence product.

1.5 Requirements Analysis

Finding the requirements is one of the most important and critical aspects for a successful software project. It requires good amount of research and discussions with relevant people to extract a set of requirements. The methodology used for requirements analysis in this project is a **Qualitative research** approach. Instead of focusing on large number of data sources and drawing results from them, the project is based on carefully analyzing the elements of interest in an exploratory manner. So, the research is focused on collecting data less in quantity but more in depth and relevance. The requirements are

finally realized and implemented as a business intelligence tool. They are also separated into functional, non-functional, and hardware/software requirements to make the development process efficient and organized so as to meet the specific objectives of the project. Approaches such as below were followed in the process:

- Conversation with potential clients
- Team discussions
- Discussion sessions with the project supervisor
- Consulting academic sources such as books and journals
- Consulting various white papers and research papers

This can be summarized as the REQST model which can be visualized using Figure 1:

Figure 1 The REQST model of Requirements Analysis

1.6 Feasibility Study

Talking about feasibility of the system, it is feasible in all regards as it fulfills all **technical, economic, legal**, and **operational** constraints. All the technologies required to develop and use this system are easily viable (free and open-source), and cost being one of the major implementation factors, the system is designed to fit well inside the budget. Considering legal issues, the system does not violate any copyright laws and does not use

any illegal components. Proper attribution will be made if required while using various libraries and interfaces.

The expected system is lightweight and easy to use, hence it can easily adapt in a working scenario. It does not have huge maintenance costs and its intended benefits easily outweigh its operational expenses. Finally, the **schedule** for the project has been created after careful planning and analysis. Thus, as far as there are no extraordinary circumstances, it will be completed within promised time.

1.7 System Requirements

Business Intelligence systems comprise of a set of tools that help any organization for making effective decisions. Obviously, each organization has its own needs and the set of tools required by it are different from others. However, when it comes to the basic requirements of such systems, there are a set of common tools that are useful to all. Being specific to medium scale enterprises and their needs we have selected proper tools for development. When implemented using those tools, the system will have following hardware and software requirements. These are minimal requirements and must be fulfilled for smooth running of the software.

- Any current Linux distribution or Windows OS (released after 2012)
- Grails 2.4.4
- Apache Tomcat 7.x
- MongoDB 3.0
- 1 GB System Memory (2 GB for Windows platforms)
- 500 MB of free disk space

CHAPTER 2: LITERATURE REVIEW

2.1 Background

Business Intelligence tools can be seen in most big enterprises today. However, taking help of technology to make decisions is nothing new, even in smaller businesses. Even without business intelligence packages, tools like Excel are used to visualize data and perform calculations [4]. Likewise, presentations are made to disseminate the information during meetings and discussions. While it is important and very likely that the meetings and discussions will continue, organizations are increasingly being attracted to business intelligence tools. In a competitive market scenario, all organizations need a collection of tools that provide them quick and comprehensive insights from data [5]. Therefore, it cannot be better than to have a set of tools exclusively made for organizations of particular size and common needs. This literature review will explore how this applies to medium scale enterprises, and propose a set of tools and techniques that can used to develop such a system specifically suited to them.

2.2 Study of Existing Practices

Considering the current practices in most enterprises, it would not take a second thought before imagining a classy boardroom full of nicely dressed decision makers who discuss the ways to get ahead in the market. For medium scale enterprises too, the scenario is similar if not the same. In some organizations, there are regular meetings where employees manually prepare charts and business agenda to put in their understanding in front of the directors [8]. In other cases, it is limited to analyzing the raw data in closed boardroom discussions. It is hard to find a data scientist or a proper statistical analysis tool in such organizations, and most are still in the Excel mindset [4]. Hence, even if they are attracted to business intelligence products, using them has not been that popular.

Likewise, those who use them exploit only a small percentage of their features. Most medium scale enterprises use it without considering the business parameters they want to analyze. This is simply waste of their finances [9]. Some tools are so comprehensive that

they are not just expensive to buy but also to train the staff. Other problems are related with scalability and adaptability of the products. So, a tool is needed that is specifically made for them, which is more useful and wastes less of their money. Similarly, the system should be simple to learn, and usable with current hardware and software resources [10]. Thus, not just in operational practices, change is also needed in technological practices to make business intelligence effective.

2.3 Results Sought

The main goal of the end-product, like all projects, is to improve over existing practices. This is primarily based on the project objectives that were set previously. Nonetheless, it is important to sort out the minute aspects within those objectives to visualize the result wanted at the end of the development cycle [11]. For this, BI requirements checklist from M87Systems was very useful as it contains all popular business intelligence features along with their descriptions [12]. The checklist has been modified here for analysis of the results relevant to end-product and presented in Table 1.

Aspect	**Description**	**Relevance to Medium Scale Enterprises**
Simple Ad Hoc real time inquiry	The ability to retrieve information containing simple sums, counts and averages	Relevant but simple queries can be covered items in dashboard [13]
Scheduled data extracts	The ability to have queries executed at predetermined times or related to business events.	Not needed. New details can be updated at each login.
Computed Columns	The ability to have columns of information that are calculated and not stored.	Important for creating graphs and charts.
External Sources of Information	The ability to integrate information from outside the Data Warehouse.	Not needed. Internal data is more important and basically enough.

Use Summary Information	Information that is aggregated to pre-determined levels.	Important for creating quick insights [14] and observing trends.
Retrieve large amounts of information	A few thousand or more rows of information	This feature is important as the end-product has to be scalable.
Forecasting	Prediction of future values	Some predictive analysis is always necessary [5]
Budgeting	Help in breaking down organization's budget	Not relevant. ERP systems are more suited to this.
Time Series Analysis	The ability to perform a time based analysis	Relevant but basic insights can be given using bar charts.
Regression Analysis	The ability to use data to predict outcomes for other data series.	Important and simple technique to make predictions.
Charts	Various bar, pie, doughnut charts and graphs for data visualization	Data visualization is a must in every BI tool [13].
Task automation	The ability to automate frequently used tasks. Usually in the form of macros and small applications that serve as assists	Not important. The application basically performs same tasks every time.
Remote access	The ability to use BI tools from remote locations.	Somewhat important. Mobile support should be considered (responsive application) [5].
Stand Alone access	The ability to use the tools when not connected to the network or internet	Important in the situations when there are problems with the internet

Table 1: Analysis of different aspects of BI relevant to Medium Scale Enterprises

2.4 Comparison with Market Products

There are many business intelligence tools in the market that provide an array of features. Most of them are made by known tech giants and are of high quality. Some of them provide services even provided over the cloud [5]. While the target of this project is not to directly compete with all of these products, it is worth comparing it with some of the market leaders to identify the hot features that must be included. It is also useful to indentify the strengths of the proposed system and build around them. The products compared here are IBM Cognos, Pentaho and Sisense.

IBM is one of the global leaders in technology and their "Cognos" is obviously one of the most popular business intelligence products. It has a number of excellent features that make it a comprehensive business intelligence tool. It has sleek reporting, interactive dashboards, predictive capability and ability to integrate with different tools and platforms. It can be used over web and mobile too, with a dedicated mobile version [13]. However, the presence of a whole lot of features and some very complex ones makes it difficult to learn.

In short span of establishment, Pentaho has established itself as a renowned brand that specifically deals in open source business intelligence products. It has a good collection of features that include dashboards, charts, predictive analysis and even time series analysis [15]. It is probably the best tool in overall among the ones discussed here. Nevertheless, the high price has kept it beyond the reach of small and most medium scale enterprises.

Another popular and relatively less expensive tool is Sisense. It has all the basic features that must be in a good business intelligence tools. However, the main focus is on dashboards and data visualization. Likewise, the major drawback is that it does not have predictive analysis [16]. While this has reduced its edge over other products, it is still doing well in the market and surprising bigger competitors. This proves that a simpler tool can do great when it is ingenious. So, the proposed system will be an attempt to emulate this idea too.

Table 2 aims to summarize the comparison of different systems based on the features they have with the proposed system [13] [15] [16] [17]:

Feature	IBM Cognos	Pentaho	Sisense	Proposed System
Dashboard	Yes	Yes	Yes	Yes
Data Visualization	Yes	Yes	Yes	Yes
Predefined Analysis	Yes	Yes	Yes	Yes
Predictive Analysis	Yes	Yes	No	Yes
Platform Independence	Yes	Yes	Yes	Yes
Learning	Steep learning curve	Can be difficult for some users	Some complex features	Easy
Target Enterprises	Medium and Large	Medium and Large	All sizes	Medium
Price	Very High	Very High	High	Affordable

Table 2: Comparison of Similar Business Intelligence Systems

As we can see in above table, the proposed system does well to match the most important features required in a good business intelligence tool. In terms of performance and reliability, the proposed system may still have a long way to go, but the point is that it will focus strictly on medium scale enterprises and offer an affordable option.

2.5 Techniques for Predictive Analysis

Predictive analysis is one of the salient features of the proposed system, and easily one of the most complex ones. All other features like insights on the dashboard, plotting the

graphs and charts, and finding the aggregated items can be implemented in a straightforward way. However, predictive analysis requires a good method to have an equally good result [3]. Selecting appropriate methods for predictive analysis is hence a vital aspect of the research. So, a set of data mining and machine learning techniques will be explored, analyzed and compared in this section.

It is often said that the accuracy of prediction does not depend upon the complexity of the algorithm or method applied but the size of data available. Thus, we can actually make better predictions using a simple algorithm and large data instead of a complex algorithm and less data [18]. Though every institution starts with small amount of data, using a complex method only adds computation overhead to the application. Hence, sticking to simple methods and focusing on both the quantity of data is more desirable.

Considering the predictive techniques, there are many data analysis techniques such as regression analysis, decision trees, time-series analysis, neural networks, case-based reasoning, CART, rule induction and Genetic Algorithms [19]. Among these techniques, there is no rule of thumb to conclude that one is better than another. It basically depends on the type of application or field in which it is applied. Therefore, a careful analysis of these techniques is necessary to determine the ones that can be useful to provide reasonable predictions quickly from a large amount of data. Table 3 scrutinizes these different techniques and their usability measures with respect to the proposed system [12] [14] [18] [19].

Technique	**Application Area**	**Comments**
Regression Analysis	Simple predictions	Useful and less computational overhead
Decision Trees	Simple classifications	Useful and less computational overhead
Time-Series Analysis	Forecasting based on time intervals	Useful but basic insights can be given using bar charts and histograms

Neural Networks	Pattern Recognition, Forecasting	Useful but adds high computational overhead
Case-Based Reasoning	Predictions and Classifications	Cases are hard to gather and model for large data
CART	Financial predictions	Useful but adds high computational overhead
Rule Induction	Electronics, Physics and Engineering	Not useful
Genetic Algorithms	Biological research	Not useful

Table 3: Different prediction techniques and their relevance to proposed system

It is clear from the above table that the two techniques that are both useful and lightweight are Regression Analysis and Decision Trees. Though they both can be critiqued for being simple, it is the simple and quick that needs to be used in the proposed system. Furthermore, with proper implementation, they are more than capable of providing good predictions.

2.6 Tools suitable for the Solution

The real culmination of the project is with a working solution. The solution is not only the proof of concept but also an integral part of the project objective. And clearly, right tools must be selected to make the solution scalable and low-cost. The basic criteria for this start with free and open source tools. After that, the scalability issue becomes a prime importance and other minor issues follow. Like every software project, many tools are required to develop the solution and they must address above issues very seriously [11].

Keeping the project objectives in mind, the main tools used for the project are Grails Framework, MongoDB (the NoSQL database) and R (statistical package), and all of these work behind a number of front-end tools and libraries to realize our business intelligence tool. The descriptions of the tools and the reasons behind choosing them are as follows:

Front-End Tools

Tool	Description	Comments
HTML5 (Hyper Text Markup Language)	HTML5 is the latest version of HTML. It has cool new features and it fully optimized for modern browsers.	It is clean and more powerful than previous versions. It is easier to build eloquent user interfaces using HTML5. Validation of forms and designing is easier in HTML5.
CSS (Cascading Style Sheet)	CSS is used for presenting a web document in a better way look wise [20].	CSS is used in the system to give styles to the designs of the web application. Styling remains separate from the layout which makes it easy to design and maintain.
Bootstrap	Bootstrap is a free framework for designing great looking user interfaces.	It contains a collection UI of features like icons, font, buttons, etc [21]. It also allows us to build responsive designs for multi-device support. So, it is important to provide mobile support.
JavaScript	An interpreted computer programming language used to write client-side scripts to make dynamic websites.	JavaScript has been used for scripting or to implement some core functionalities of the system [22]. External JavaScript libraries like Chart.js are useful for charts and graphs.

jQuery	A client side web development library for easy interface design [23].	jQuery has been used to provide some dynamic behavior to the application mainly related with the user interface components such as menus and navigation.

Table 4: Front-end tools used in the project

The backend tools were carefully selected in order to make a low-cost and quality solution. Issues such as integration of tools with each other during and after development also hold great importance. All the selected technologies should be able to work well with each other for a stable system. Incompatibility and crashes are the last things we need here. So, these issues were also analyzed for smooth operation. The description of each tool including reasons for using them can be seen below:

Grails

Grails is a full stack JVM based framework providing support for all aspects of developing web applications. It bundles the dynamic power of the Groovy language and existing successful frameworks like Spring, Hibernate, Sitemesh, Tomcat, H2 and several others combined into a powerful framework [24]. Grails provides the developers with all the tools needed to achieve the tasks in an efficient and effective manner without having to deal with configurations or low level details.

As it is based on the JVM framework, it also delivers scalability and sustainable performance. The applications made in it are robust and stable. Specifically in this project, Grails allows easy integration with other backend tools with the same benefits. It is simpler and allows quick development compared to other Java frameworks. The simplicity and short development time of Grails applications has inspired many web developers to create the well-known web applications like Netflix Asgard and Vodafone Music [25]. Hence, Grails is a good match for developing the proposed system.

R

R is a free and open-source language and environment for statistical computation and graphics. It is an interpreted language which allows branching and looping as well as object-oriented programming using functions. R is highly extensible through functions and extensions [26]. There are many other statistical packages such as SPSS, SAS, MATLAB, STATISTICA, Minitab, etc. which can give competitive edge to R. However, R has major advantage over others in terms of affordability. R is freely available where as others aren't.

In terms of the features, R provides a lot of features as its counterparts [27]. R can be easily integrated into Java with the use of Java extensions of R. So R can be easily used with the applications where Java is being used. Most commercial software being standalone packages cannot be directly integrated into Java. R is up-to-date and updates are released twice a year. R has tons of documentation and various community supports and discussion forums for problem solving.

The R variant used in this project is the Renjin (pronounce "R-engine") script engine which like Java runs on JVM. Therefore, it can be very easily integrated with Grails. Since both the Java code and R script will run on JVM, it will perform faster calculations and provide faster results. Even though Renjin is said not to be enterprise ready, it has been updated to support many native R libraries [28]. It is free like R, and supports bigger data and better performance. Hence, it is the most logical choice for predictive calculations in this project over others like RServe, RCaller and JRI (Java/R Interface).

MongoDB

MongoDB is a NoSQL database that allows us to store data in form of "documents", a kind of data structure. It is free and open-source, and the most popular NoSQL database system. They key to use MongoDB in this project is that a RDBMS database can be seamlessly migrated to MongoDB [29] and there are SQL equivalent queries that make it easier for adapting into NoSQL paradigm. Additionally, it has excellent caching facilities to support faster access to frequently accessed data [30]. So, it is well suited for web applications and moreover for real-time data analytics for business intelligence.

2.7 Chapter Conclusion

In conclusion, the proposed system is a combination of research with a usable solution. In the process, a lot of data analysis techniques were scrutinized along with a number of business intelligence features. All the steps for this were carefully planned and after assessment of different alternatives, all suitable tools have been selected. Keeping everything in mind, the rough solution has already been visualized. The project is research-oriented that has objectives directed towards an interactive business intelligence tool. The investigation is complete and development is to follow.

CHAPTER 3: SYSTEM DEVELOPMENT

3.1 Project Management Strategy and Tools

The project management part begins right from the project initiation where we have to identify the tools and technologies required to develop our system. We had a number of options to consider and research given that new and better technologies are coming our way almost every day now. The first part about this was specifying the project teams and roles and then selecting the developing methodology.

3.1.1 Project Team

Resource Person	Role
Vishnu Kumar Rana	Project Supervisor
Dixit Bhatta	Designer/Developer
Dirban Singh Lama	Designer/Developer

Table 5: Project Team

3.1.2 Roles and Responsibilities

The specific responsibilities of the project personnel are based on their roles in the project. The supervisor has responsibilities related with enhancing project quality through effective mobilization and encouragement of the team members as well as controlling overall project. The team members have the responsibility to do the research, realize the solution and finally implement it. The summary of their responsibilities are shown in Table 6:

Role	Responsibilities
Supervisor	Feedback and Suggestions Sharing Information Project Progress Tracking Motivating team members

Team Members	Research and Analysis Design and Development Testing and Optimization Implementation Documentation

Table 6: Roles and Responsibilities of the Project Team

3.1.3 Selected Methodology

The methodology used in this project is Extreme Programming with "Test Driven Development". The method was used due to following reasons:

- The development team was small (Extreme programming usually has two-person programming team) [31]
- It helps to maintain software quality and addresses changing requirements
- The quick incremental cycles were well suited for the project because the functionalities could be implemented quickly one by one
- It leads to high quality code implying a robust and stable system
- The overall advantage is that it improves productivity

3.1.4 Comparing Alternative models

Certainly, there are many other methodologies which could have been deployed in this project, the most probable ones being the other two models discussed above: Prototyping and Agile methodologies.

The Prototyping model was not used because it relies heavily on the cost of the developer. It should be conducted at the least possible cost and hence the budget constraint may come into play at the middle of the development phase [32]. It is also considered as a slow process where client is frequently involved. Keeping on revising the prototype can be very difficult for the developers especially when a lot of new requirements emerge between prototype revisions.

The next model under discussion was Scrum. It has become a kind of cliché for the software developers who boast about their software development practices. Their

bragging right is well justified though because it has tried to accommodate all the good software development practices under a single method [33]. However, it requires more communication with less documentation, and hence it requires more time to take out facts from the client and difficult to review the projects later.

3.1.5 Project Management Tools

The specific tools used for effective project management are described below:

Project Scheduling and Task Management (Using Asana)

Asana enables teamwork done through fast and flexible web and mobile applications. The friction of communicating the right amount of information, to the right people, at the right time, in the right place, gets in the way of teamwork [34]. This friction lowers our collective productivity. Asana improves the productivity of teams and increase the potential output of every team's effort.

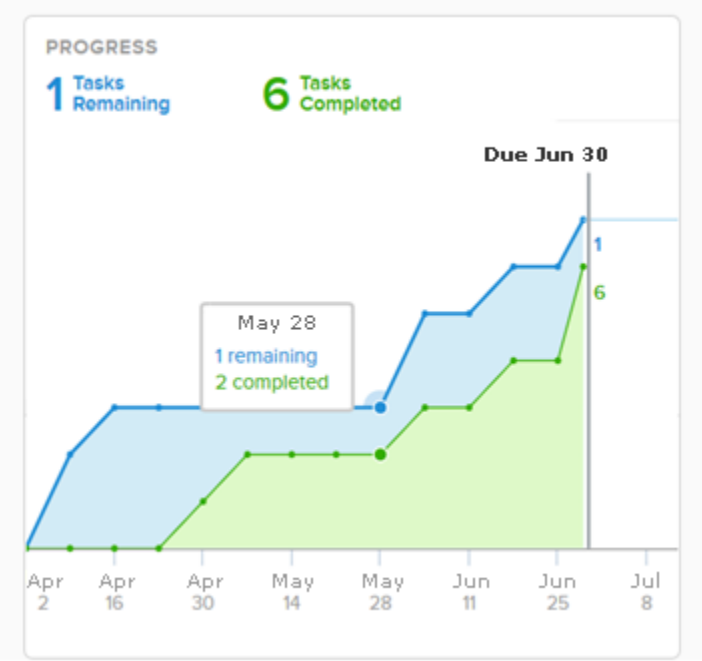


Figure 2 Work Progress of D2 Business Intelligence (Asana Dashboard)

Figure 2 displays the work progress of the project. We can observe the delay caused due to earthquake in April/May period.

Online Code Repository: Github

To have better management of coding part and effective collaboration, we used GitHub as the online code repository and version management system. GitHub is a web-based Git repository hosting service, which offers all of the distributed revision control and source code management (SCM) functionality of Git as well as adding its own features [35]. It also provides access control and several collaboration features such as wikis, task management, and bug tracking for projects.

3.2 System Analysis

The main stakeholders in this system are the end-users and administrator. End-users can be anyone in the client organization, the only thing that matters is to provide correct insights to the stakeholders. This system creates a hassle free and easy way for the end-users to visualize the patterns and predict some parameters using relevant data. For that we needed to identify the requirements of the system based on the objectives and features defined previously in this document.

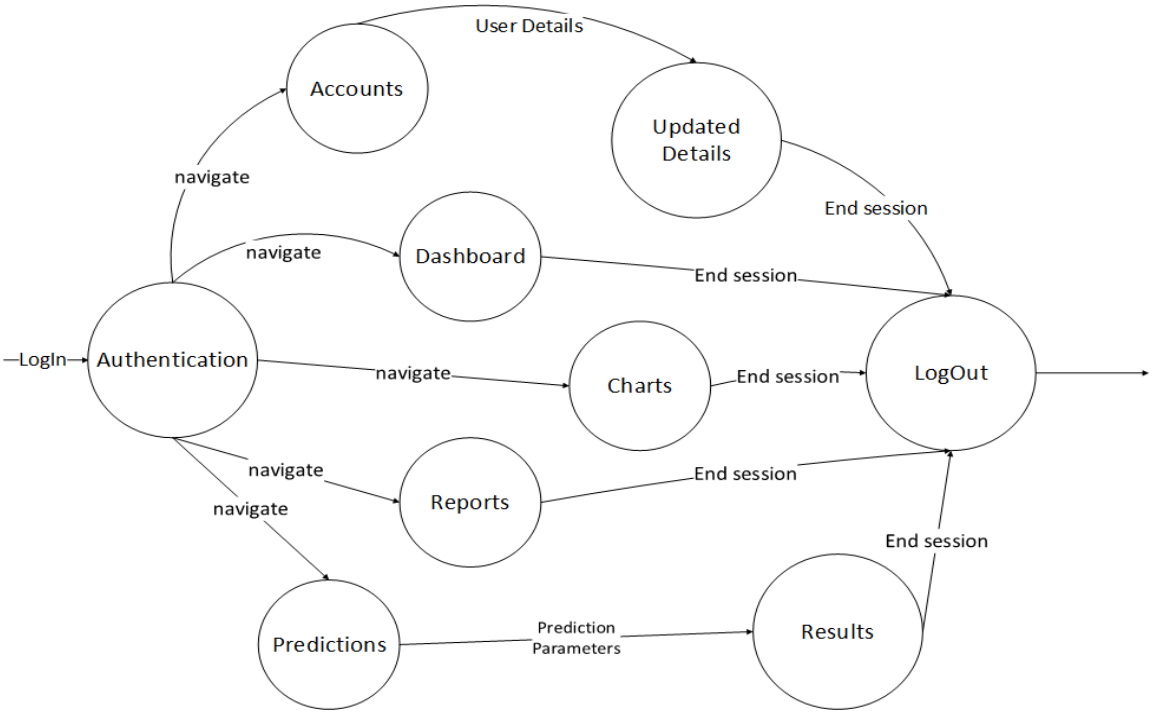


Figure 3 High Level State Diagram of the system

So, the analysis was done by designing a high-level state diagram depicting the overall system and the core requirements that must be implemented by it. The diagram (Figure 3) focuses mainly on how the application should work and the navigation between the states (pages and results provided by the application).

Apart from the state diagram, we also need a set of functional and non-functional requirements as a part of comprehensive analysis. With well defined requirements, it is easier to work on the design and development phases. So the functional and non-functional requirements of the system are listed below:

3.2.1 Functional Requirements

- The system must have a dashboard showing the most important insights
- The system should have data visualization features using charts/graphs
- The system should make reasonable predictions
- The system must be able to handle increasing data
- Users must be authenticated before using the system
- Users must have control over their individual accounts

3.2.2 Non-functional Requirements

- The system must be simple to use
- The system must provide quick insights from data within seconds
- The system must work with existing resources without adding major costs
- The system must be stable and should not crash at runtime

3.3 System Design

For proper development, all the analyzed requirements should be transformed into designs. The designs provide us graphical representation of the system's architecture, flow of data, features and involvement of stakeholders in different aspects of the system. Various diagrams such as system architecture, data flow diagrams, use case diagram, sequence diagram and system flowchart were created in the design phase to aid development process. The diagrams are presented and explained below:

3.3.1 System Architecture

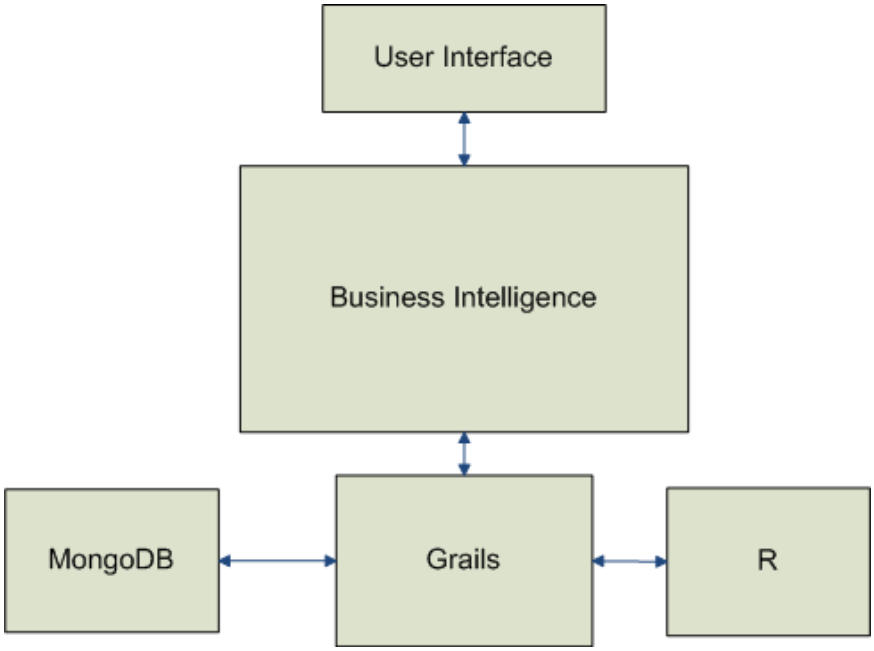


Figure 4 System Architecture

The system is a collection of components built over the Grails Framework. The user interface developed using the front-end tools are also embedded in Grails as "views". Tools like "R" and "MongoDB" are necessarily separate entities of the architecture but are interfaced together by Grails to give us a look of a single Business Intelligence system. So, it is a 3 tiered architecture with Grails serving as the application server.

To be precise, the variant of R used in this system is "Renjin" which is based on JVM. So, all we had to do for integration was to include the JAR file in the Grails environment. The architecture is thus more like R and Grails both reside inside the JVM and work as a single combined entity. The communication between them is very smooth, and clearly faster than with MongoDB. Still, MongoDB being a separate entity does not introduce much lag in the system. We can easily establish multiple connections to the database and get the required items in real-time. The Mongo Java Driver enables us to perform CRUD (Create, Read, Update and Delete) operations in a structured manner. So, the interaction between the components in the architecture is organized well.

3.3.2 Data Flow Diagrams

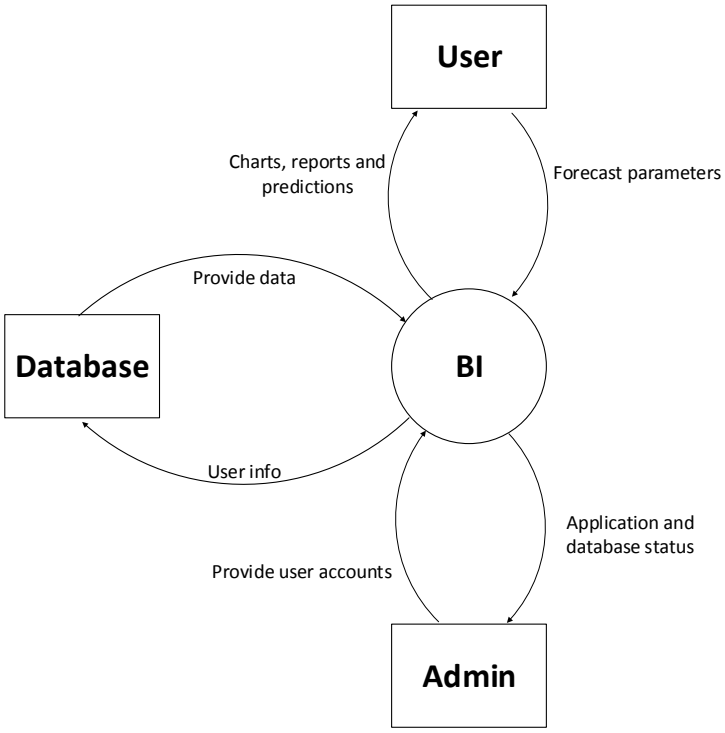


Figure 5 Context Diagram

Figure 5 is the context diagram of the system. The high level view of the data flow in the system is shown by the diagram in terms of simple exchanges. As we can see in it, BI system interacts with the user, database, and the administrator, the main stakeholders of the system. Regardless of the way the interaction between the system and external entities takes place, it is data that will be flowing throughout the process. Each individual entity has different data on offer and expects different output data from the system.

The database stores all the required data and provides them whenever the system requires. Admin is responsible for managing the user accounts and also monitors the status of the application and database. User can view the charts, reports and other visual representations of the company's database using the system.

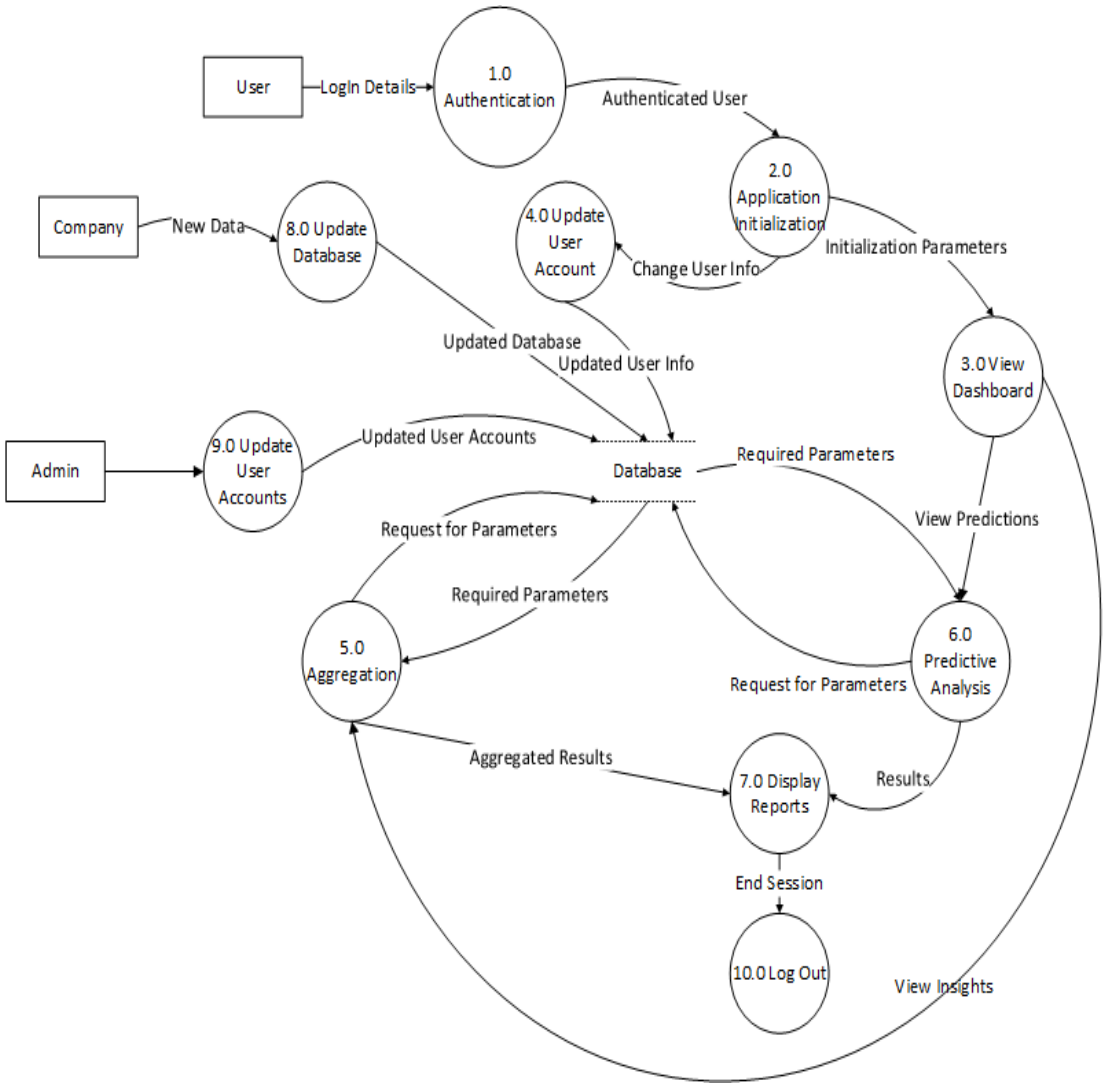


Figure 6 DFD Level 0

Figure 6 represents a Level 0 DFD which represents the overall flow of data in the system. There are mainly 3 actors in the system: User, Company and Admin. The company is responsible for providing the updated database to the system administrator (Admin). Admin can add or remove the users account from the database and also aid in synchronizing the database.

An authenticated user can view the insights of the company's database in a dashboard. After the application initializes, the user can choose to update the user details in the database. The application aggregates the database and provides the aggregated result for report generation. The user can also view the predictive analysis made by the system.

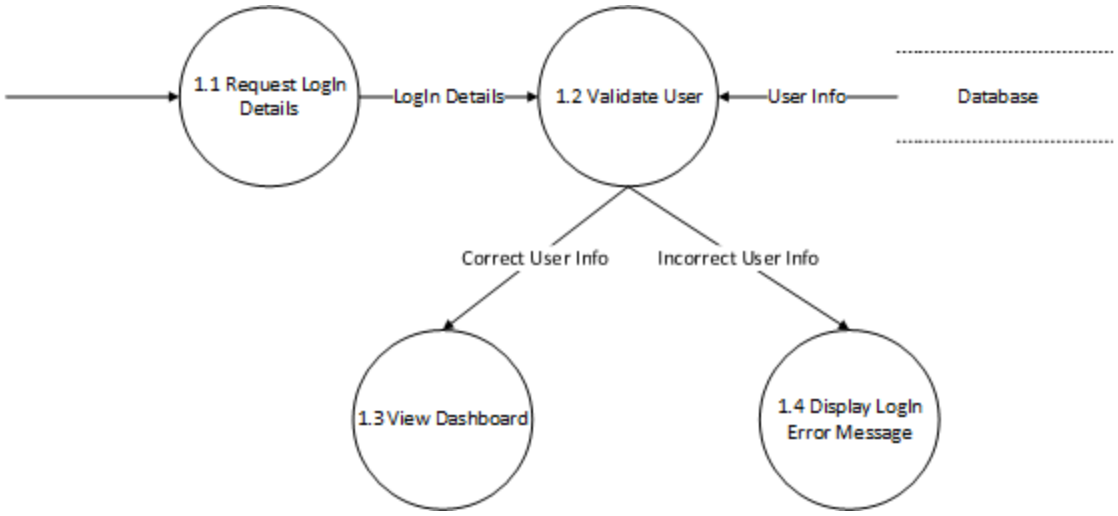


Figure 7 DFD Level 1 (Authentication)

Figure 7 shows the Level 1 DFD diagram for the authentication process. The authentication process involves requesting the login details from the user. The login detail is corroborated with the user information stored in the database. The validated user can view the dashboard otherwise the error message is displayed.

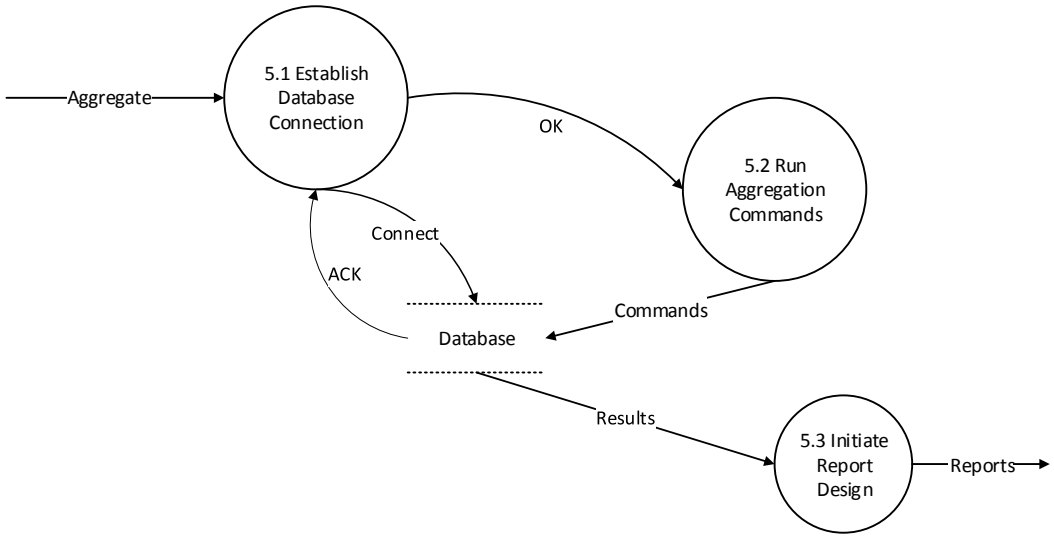


Figure 8 DFD Level 1 (Aggregation)

As we can see in Figure 8, the connection with the database should be established first for aggregation. After the database acknowledges the connection, aggregation commands are run. Then the results are used to initiate the designing of the reports.

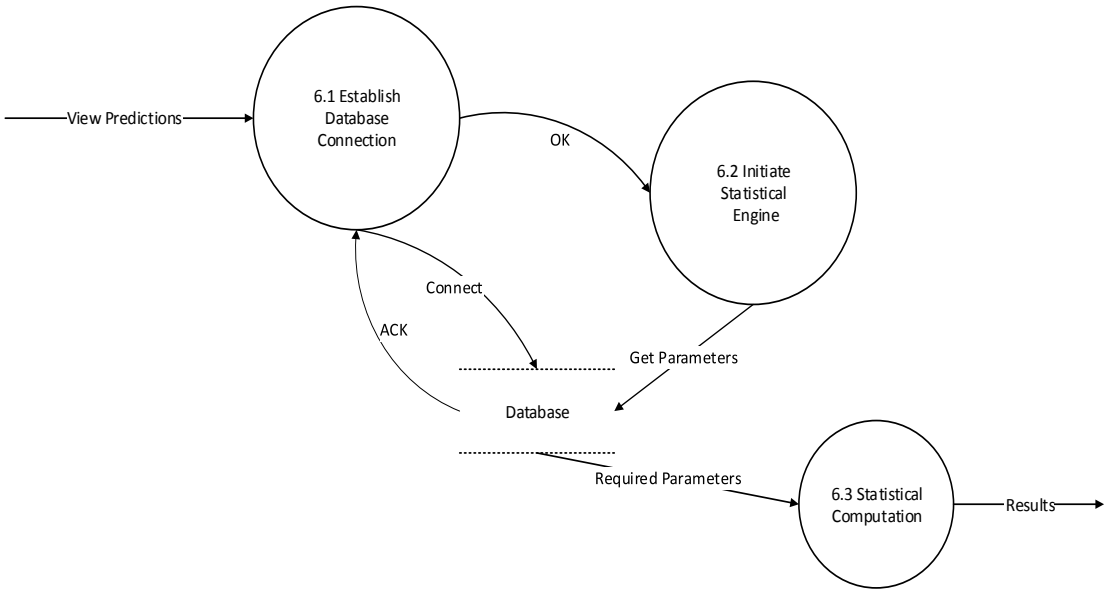


Figure 9 DFD Level 1 (Predictive Analysis)

For predictive analysis too, the database connection has to be established. Then the statistical engine is initiated which obtains the parameters from the database. After obtaining the required parameters, statistical computation is done to obtain the results.

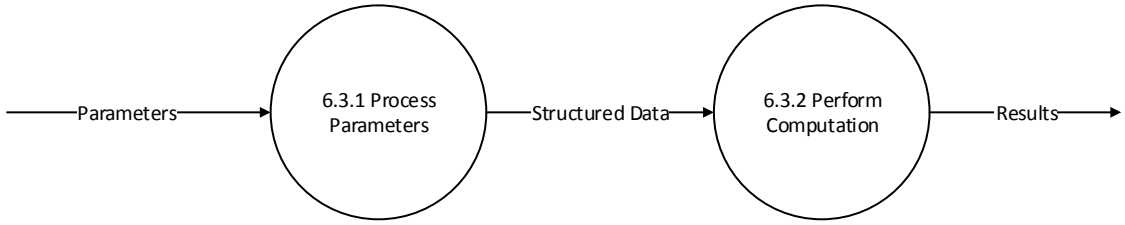


Figure 10 DFD Level 2 (Statistical Computation)

Now, for statistical computation of results, the parameters are processed to obtain the structured data and computation is performed on those data to obtain the results.

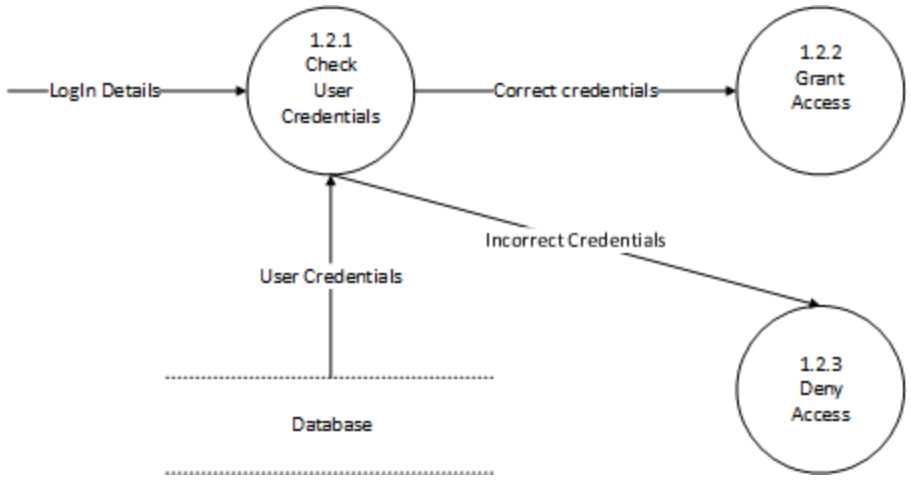


Figure 11 DFD Level 2 (Login Validation)

Figure 11 represents the Level 2 diagram for the login validation process. The user credentials are stored in the database. The user with the correct credentials are granted access to the system otherwise the access is denied.

3.3.3 Use Case Diagram

Figure 12 shows the use case diagram of the system. The diagram depicts how the external factors interact with the system. It also shows the working processes of the system. It is the simplest is a representation of a user's interaction with the system which shows the system of interest, the actors in the system, and the specific roles played by actors. In a broad sense, it also shows the relationships between the actors.

As we can see in the diagram, aspects such as user authentication are done by the system. The user is able to view the reports and predictions made by the system from a synchronized database. The updates to the data are handled by the database system itself. However, the database must be regularly synchronized, which according to the diagram is done by the administrator.

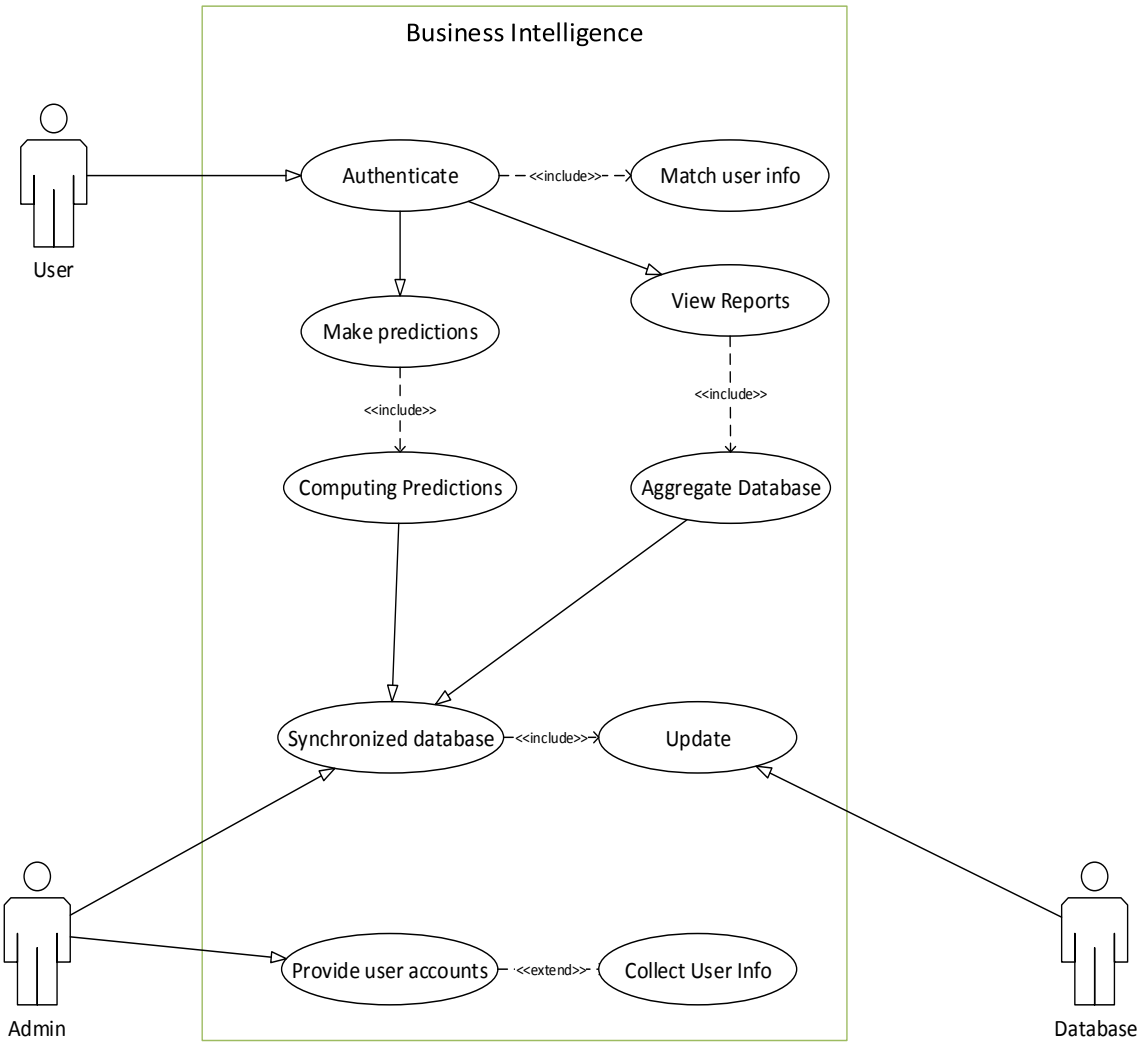


Figure 12 Use Case Diagram

3.3.4 Sequence Diagram

The sequence diagram depicts how the different components of the system activate and interact with each other. The diagram shows the user interaction through the user interface and how other components work behind the scenes. The system regularly interacts with the database for the required data and shows the reports according to the provided data. Figure 13 shows the exact sequence of the activities when a user interacts with the BI system. The user begins by authenticating, and the application automatically interfaces with the database and relevant views to generate required data. All the user has to do is to observe and interpret the results before finally logging out of the application, all of which is an individual session.

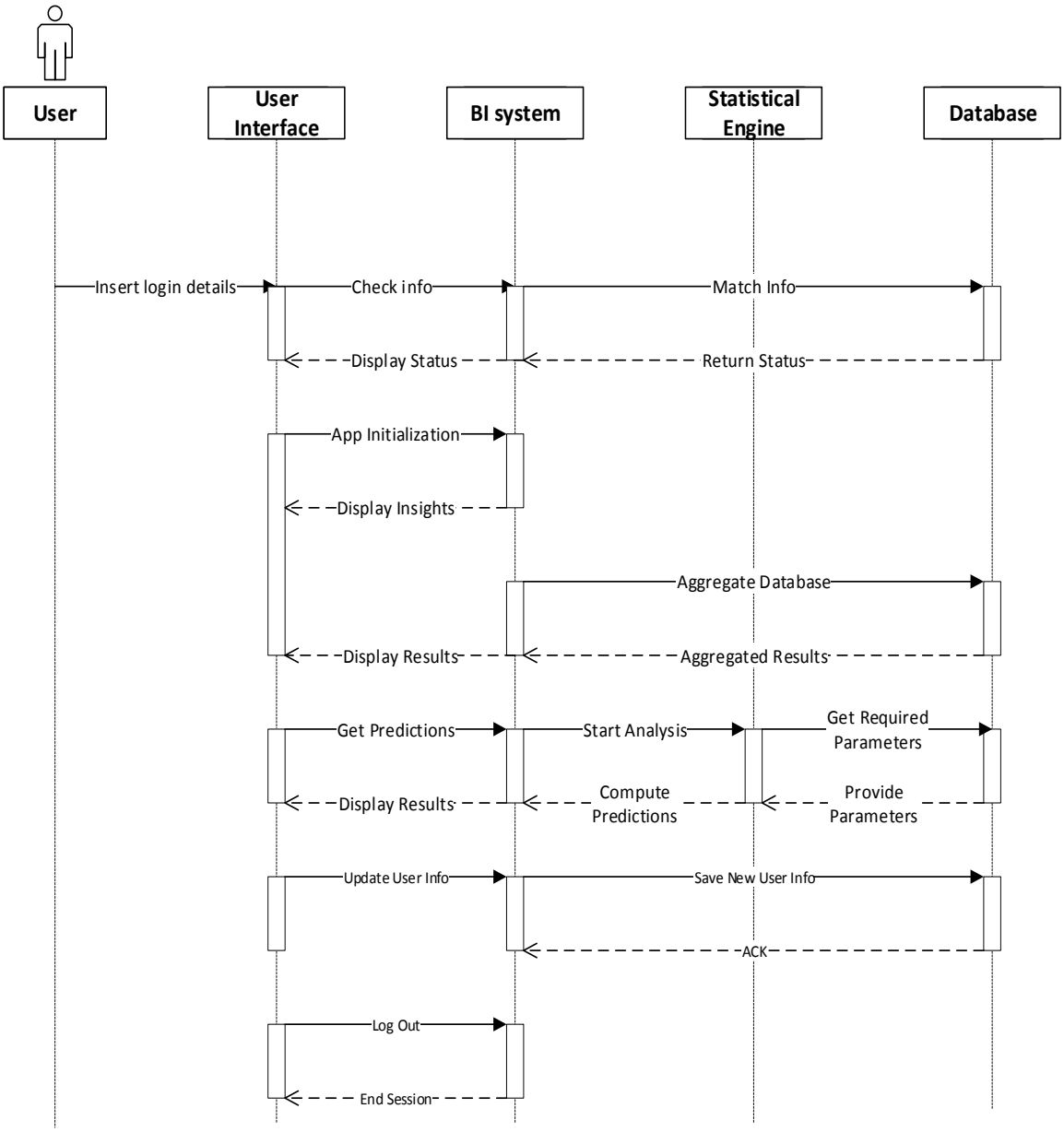


Figure 13 Sequence Diagram

3.3.5 System Flowchart

The system flowchart (Figure 14) depicts the overall flow of the processes in the system. The basic input, output and decision processes are graphically presented below. The major processes like authentication, aggregation, prediction, report preparation and account modification, and the relation between them can be seen in the flowchart below, all of which are shown in terms of ordered blocks from start to the end.

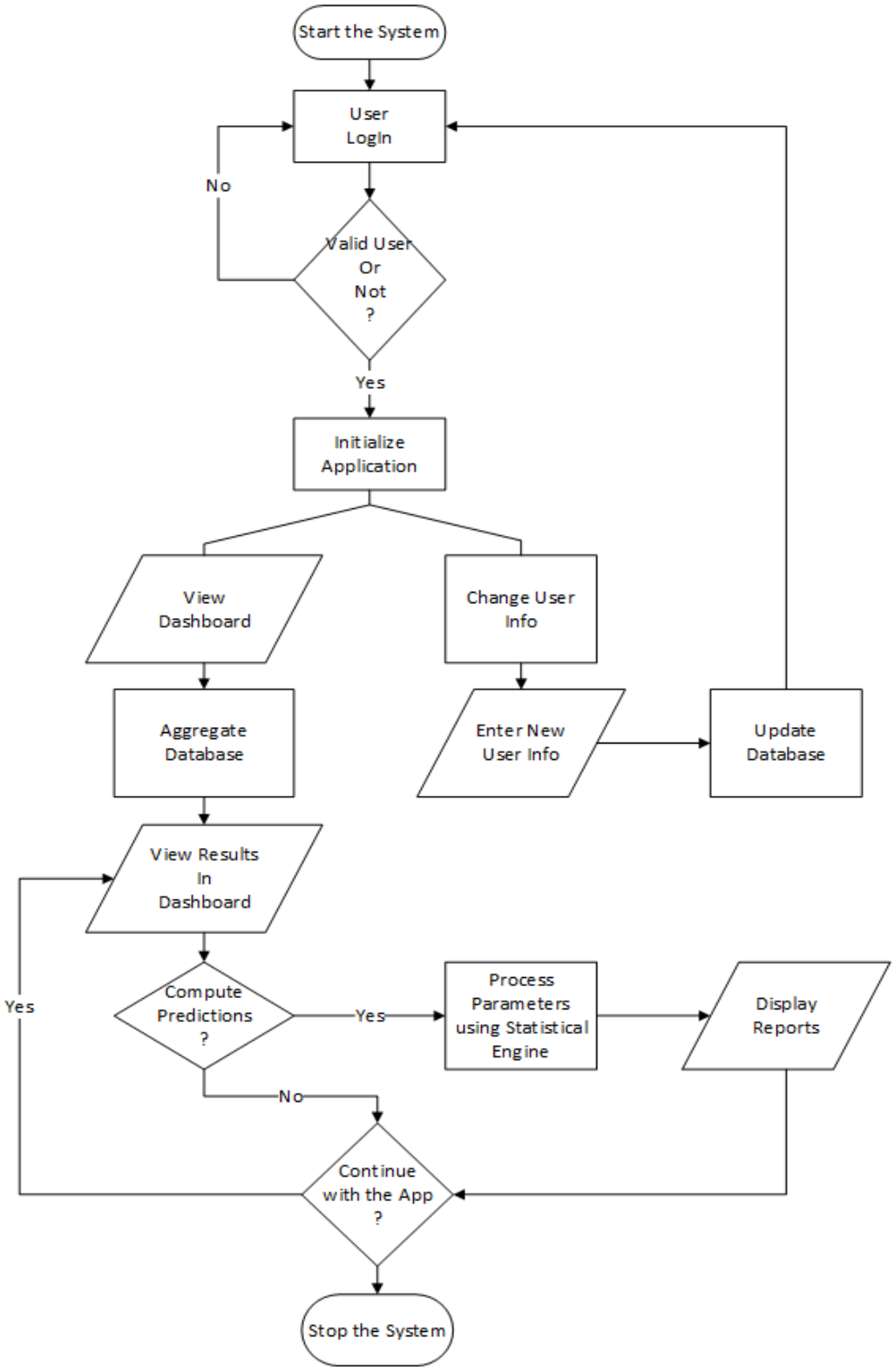


Figure 14 System Flowchart

3.4 Project Schedule

The tentative time schedule of the project is outlined below showing planned schedule and actual time consumed:

S.N.	Task Description	Planned			Actual			Planned Deliverables	Comments
		Start Date	End Date	Hours	Start Date	End Date	Hours		
1	**Project Startup**								
1.1	Topic Selection	3/5	3/5	3	3/5	3/5	3	Project Idea	On schedule
1.2	Initial Planning	3/6	3/7	10	3/6	3/7	10	Basic Plan	On schedule
1.3	Background Research	3/8	3/12	20	3/8	3/12	20	Solution Ideas	On schedule
1.4	Supervisor Meeting	3/12	3/12	5	3/12	3/12	5	Project Advices	On schedule
1.5	Topic Finalization	3/12	3/12	5	3/12	3/12	5	Project Topic	On schedule
1.6	Identification of Project Objectives	3/13	3/15	8	3/13	3/15	8	Project Scope and Objectives	On schedule
1.7	Collection of Resources	3/13	3/18	20	3/13	3/18	20	Research papers, references	On schedule
1.8	Schedule and Gantt Chart Preparation	3/13	5/7	40	3/13	5/7	40	Project Schedule	On schedule
1.9	Proposal Preparation	3/19	3/25	25	3/19	3/25	25	Proposal Draft	On schedule
1.1	Proposal Review/Submission	3/26	3/27	6	3/26	3/27	6	Project Proposal	On schedule

1.11	First Review: Proposal Defence	3/28	3/31	20	3/28	3/31	20	Project Acceptance	On schedule
2	**Research Work**								
2.1	Business Intelligence Features	4/1	4/8	26	4/1	4/8	26	Research draft	On schedule
2.2	Data Analysis Tools	4/3	4/4	16	4/3	4/4	16	Research draft	On schedule
2.3	Predictive Analysis Techniques	4/3	4/10	32	4/3	4/10	32	Research draft	On schedule
2.4	RDBMS to NoSQL Migration	4/3	4/7	17	4/3	4/7	17	Research draft	On schedule
2.5	Integrating development tools	4/8	4/15	18	4/8	5/25	22	Research draft	Delay due to earthquake
2.6	Prepare for development	4/16	4/17	8	5/25	5/28	15	Development Strategy	Cumulative delay
3	**Development Phase**								
3.1	User Interface Design	4/18	4/23	32	5/29	6/2	30	User Interface	Started without team discussion
3.2	Transformation into Basic Grails app	4/23	4/25	16	6/3	6/5	16	Working Prototype	Team reunited/ Cumulative delay
3.3	Test User Interface	4/25	4/26	10	6/5	6/6	10	Test Results and Corrections	Cumulative delay
3.4	Integrate R into application	4/27	4/28	11	6/7	6/8	12	Basic analytical application	Cumulative delay

3.5	Test Integration with R	4/29	4/30	10	6/9	6/10	5	Test Results and Corrections	Cumulative delay
3.6	Second Review	5/1	5/2	10	6/9	6/10	5	Project Progress Status	Review Delayed
3.7	Migrate RDBMS into MongoDB	5/3	5/8	19	6/9	6/10	6	Scalable database	Many tasks in two days
3.8	Integrate application with MongoDB	5/9	5/11	17	6/9	6/11	25	Working analytical application	Second Review performed after this phase
3.9	Touch-ups and optimizations	5/12	5/13	14	6/12	6/13	14	Improved application	Cumulative delay
3.1	Final Testing	5/14	5/16	18	6/14	6/16	18	Test Results and Corrections	Cumulative delay
3.11	Quality Assessment and Critical Analysis	5/17	5/19	18	6/17	6/18	18	Future prospects	Cumulative delay
4	**Documentation Phase**								
4.1	Draft Report Writing	4/5	5/22	36	4/5	6/18	36	Report Draft	Cumulative delay
4.2	Final Report Writing	5/20	5/23	18	-	-	18	Furnished Report	Cumulative delay
4.3	Evaluation and Conclusion	5/23	5/24	8	-	-	8	Findings and conclusion	Cumulative delay

5	**Submission Phase**								
5.1	Submission of Draft Copy	5/25	5/27	6	6/18	6/20	6	Submission for correction	Cumulative delay
5.2	Corrections	5/28	5/28	6	6/21	6/21	6	Correct Document	Cumulative delay
5.3	Printing and Binding	5/29	5/29	6	8/10	8/13	6	Final Project Report	Cumulative delay (Final Exams)
5.4	Final Review: Report Submission	5/30	5/30	6	8/14	8/14	6	Project Completed	Cumulative delay / Finished before presentation day
	Total Time (days)	87						**Planned Hours per day:** 6.20 (3.10 hours/person)	
	Total Hours	540							

Table 7: Time Schedule of the project

Likewise, the Gantt Chart (planned) based on the time schedule of the project is shown in Figure 15:

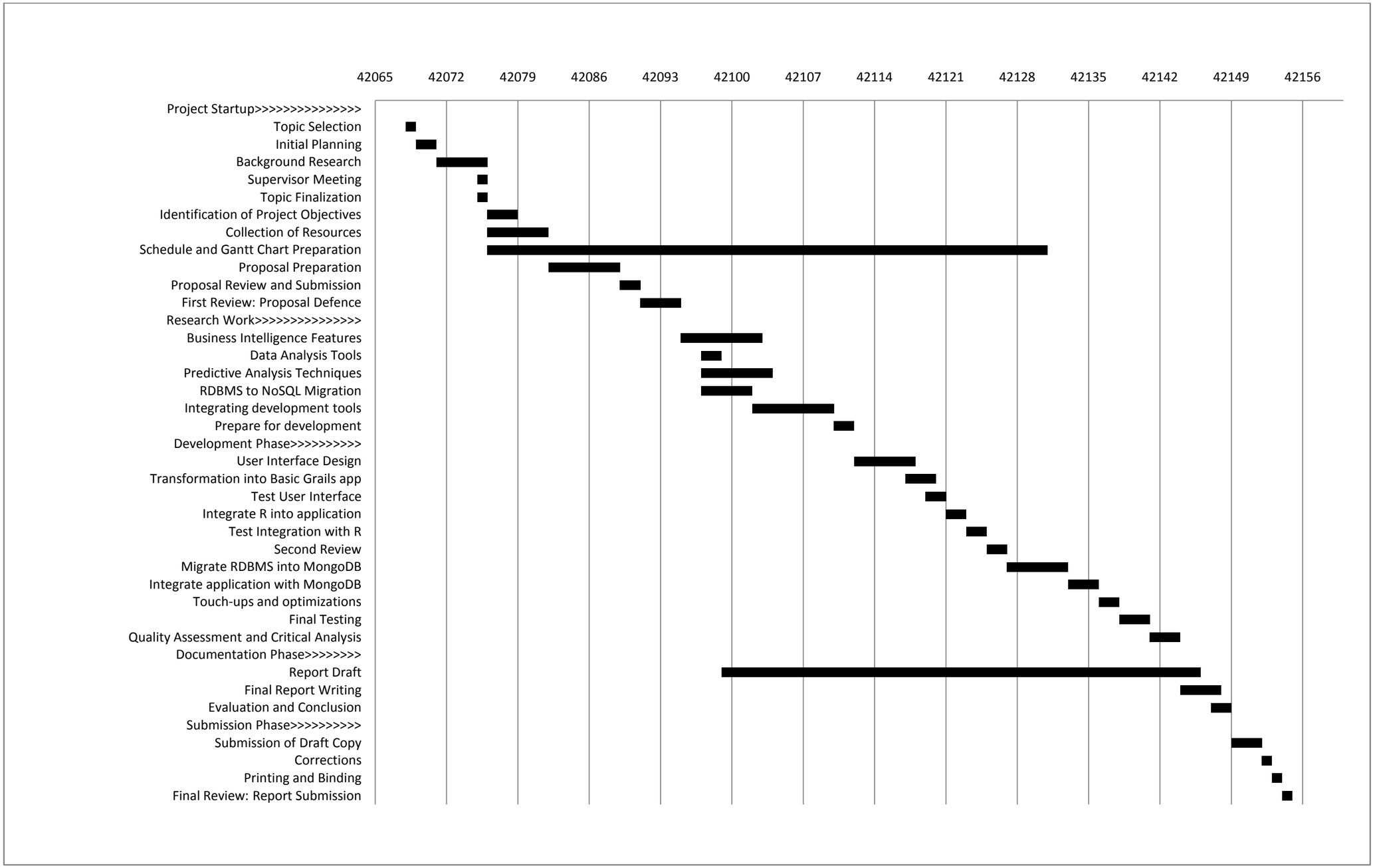


Figure 15 Gantt Chart for the project

3.5 System Testing

Since the selected methodology for the development of the project is based on Test Driven Development, the code is tested every time a new feature is added to the system be it the changes in the user interface or the core functionality. To ensure quality and stability in the end-product comprehensive unit tests and integration tests were performed. The whole system was run and inspected in the end for optimization.

3.5.1 Unit Test:

The purpose of conducting the unit test is to find any errors by conducting the tests on the smallest units or modules of the program. Each class of the code and every page of the application are tested in the process. Following steps are taken in performing unit tests:

- review program unit (e.g. an inspection, or structured walk-through)
- conduct the code execution tests
- identify and resolve any discrepancies (e.g., the types of errors, such as error handling, failure to reinitialize, the number of errors found)
- determine that the test is complete (finish tests and document results)

3.5.2 Integration Test:

Since the project has a number of tools integrated together, integration testing is an important part of testing plan. Addition of every external library and major tool prompted integration testing and it was duly performed before moving into integrating another tool. The major components that were tested for integration were jQuery, Renjin Script Engine and MongoDB. Snippet 1 is an example code snippet that was used to test the integration of Renjin Script Engine and its output showing evaluation of a linear regression.

```
package R;
/**
 * Created by uclid on 6/07/2015.
 */
import javax.script.*;

public class tryRenjin {

    public static void test() throws Exception {
        // create a script engine manager:
        ScriptEngineManager manager = new ScriptEngineManager();
```

```
        // create a Renjin engine:
        ScriptEngine engine = manager.getEngineByName("Renjin");
        // check if the engine has loaded correctly:
        if(engine == null) {
            throw new RuntimeException("Renjin Script Engine not found
on the classpath.");
        }
        // ... evaluation Java code here ...
        System.out.println();
        String result = engine.eval("df <- data.frame(x=1:10,
y=(1:10)+rnorm(n=10))").toString();
        String result2 =engine.eval("print(df)").toString();
        String result3 =engine.eval("print(lm(y ~ x, df))").toString();
        System.out.println(result+ "\n" + result2 + "\n" + result3);
    }
}
```

Snippet 1: Code Snippet of testing Renjin Script Engine

```
Call:

lm(formula = y ~ x, data = df)

Coefficients:

(Intercept) x

2.038       0.735
```

Snippet 2: Output of the test code for Linear Regression

3.5.3 System Testing

Although there was comprehensive unit testing and integration testing, whole system must be tested together to ensure proper functioning. If any bug was not visible after integration, this may show up after the whole system is built. Also, some features need to be optimized to provide better performance before the system is implemented. Hence, multiple walkthroughs of the application and the code must be done, and scanned for discrepancies after the system is ready for deployment.

3.6 Implementation

The implementation of the project is based on parallel implementation plan. Even though the organization to adopt this system may not be using any business intelligence tool, they will surely have some way of analyzing their business. Hence, making the system

work along with the existing one is the best way to evaluate the system and measure its effectiveness. After getting used to the system, the older system can be easily discarded or kept on running as per the needs of the organization.

The basic implementation begins with installation of Apache Tomcat. This can be done on any platform that fulfills the hardware requirements. After that, the Grails Framework must be installed and the app can be run straight from Grails command-line by configuring proper port.

```
grails run-app

grails run-app -https // with HTTPS

grails test run-app

grails -Dserver.port = 9090
```

Snippet 3: Running Grails application

However, it is intended for development only and never recommended for deployment [36]. Actual implementation should be done by making a WAR file, and installing it in Apache Tomcat. This is done using the command

```
grails war
```

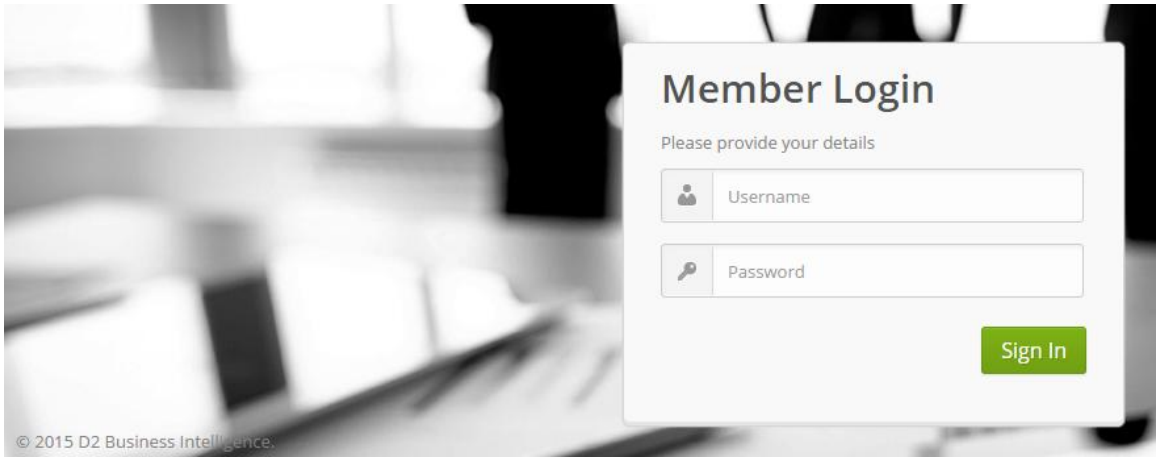
The WAR file will be created using the application and its version. The file can then be deployed in Apache Tomcat by copying it into `TOMCAT_HOME/webapps` folder and restarting it. Any other container like Glassfish, WebSphere and JBOSS can be used for deployment [36].

Finally, Database migration during implementation can be done seamlessly using the tool MongoVUE. It allows us to establish connection with popular RDBMS and MongoDB server, and migrate the whole data in the same structure as in the existing database. It also helps us to view the data stored in MongoDB in tabular form and exploit various features of MongoDB using a graphical interface.

CHAPTER 4: RESULT ANALYSIS

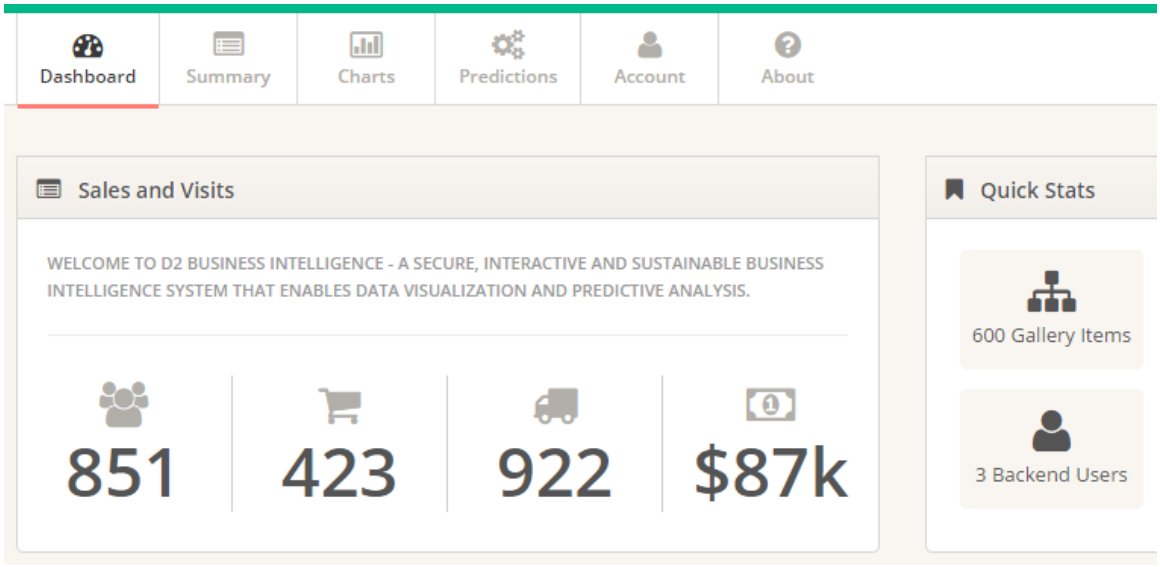
4.1 Results

The first version of the project was completed on time. And the quick upgrade was done to optimize certain features such as chart legends. There is a plan to add more features in near future. The screenshots of the application and their description are below:



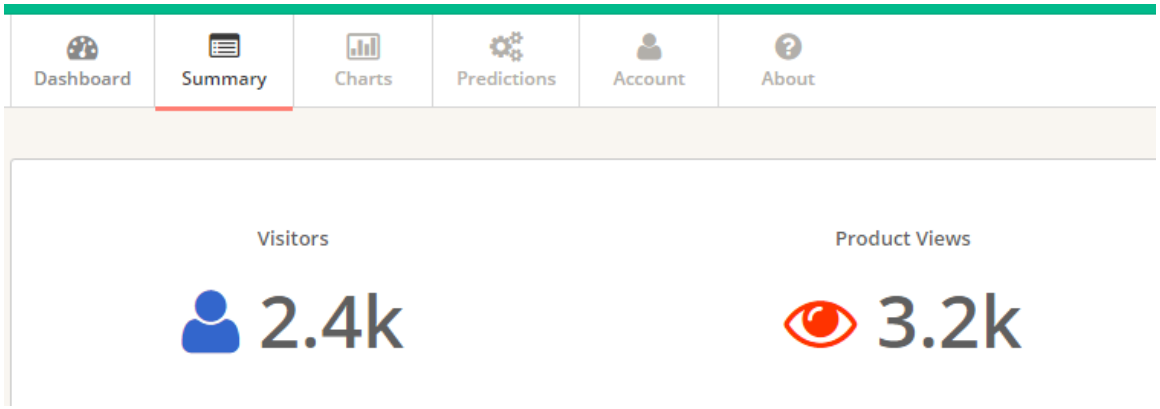
Screenshot 1: Login Page

The users can login through the login page using the username and password provided by the administrator. A separate session is created making the authentication secure.



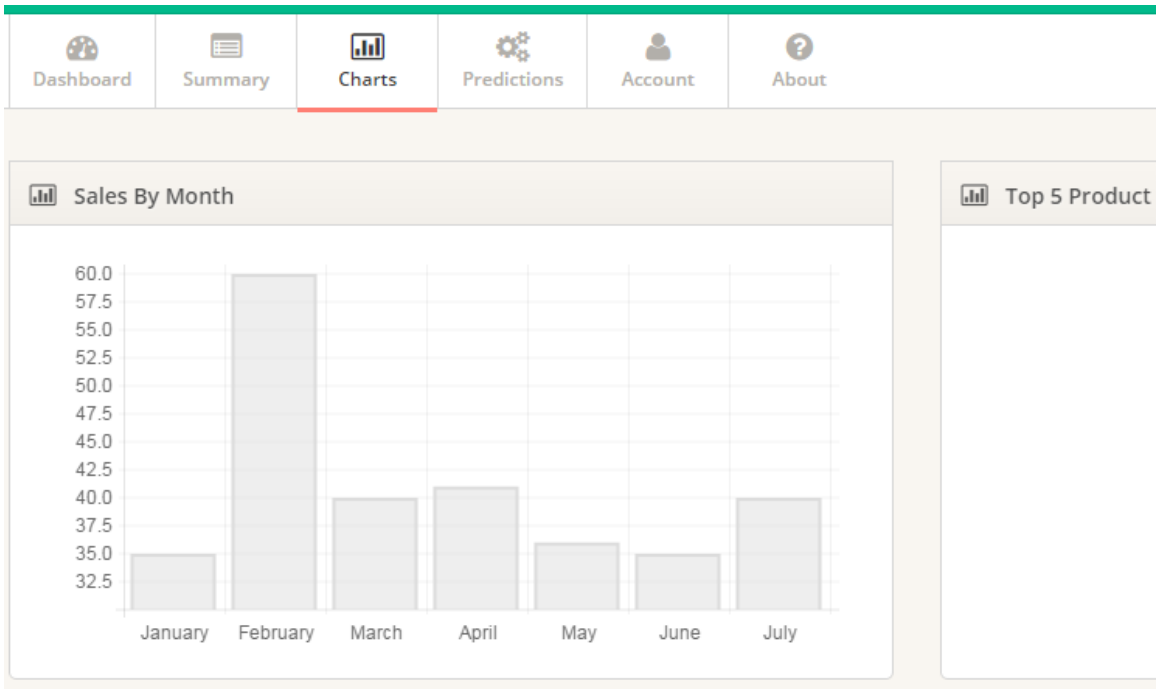
Screenshot 2: Dashboard

As the main page, the dashboard has all the things that should catch your eyes first. From important figures to subtle quick stats, the dashboard displays each and every number that makes sense to business decisions.



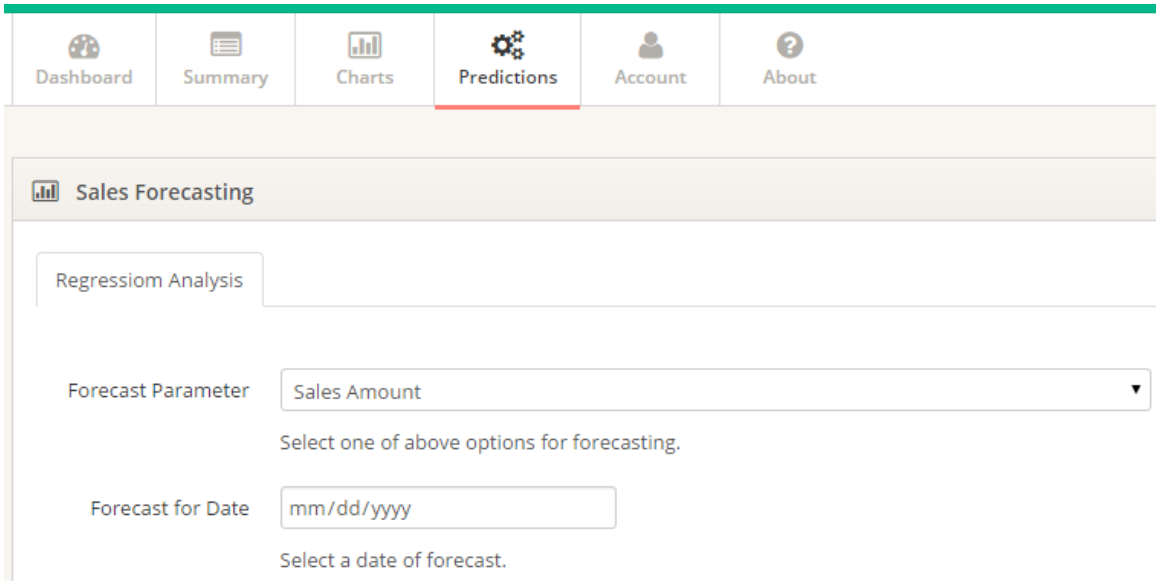
Screenshot 3: Summary Page

The summary page has the monthly summary of trivial details of business. The monthly data can be seen as a trend that can provide a sneak peek into possible outcomes.



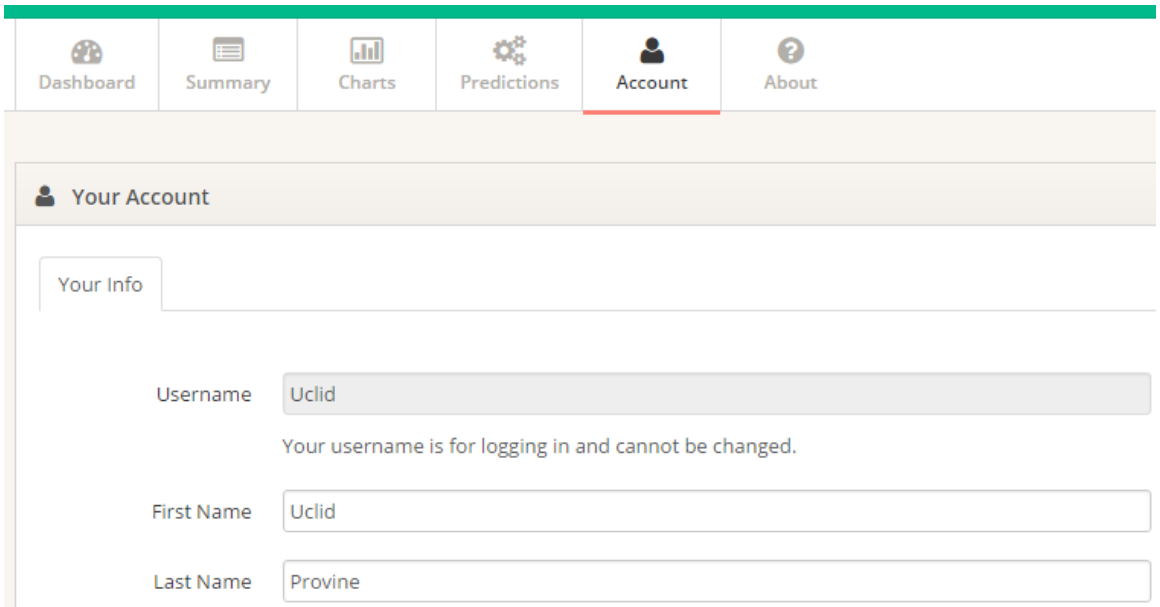
Screenshot 4: Charts Page

The charts provide you an excellent way of visualizing data. Using the charts page, users can see some of the important statistics of their business visually. They can easily take a snapshot of the chart or use the page directly for discussions.



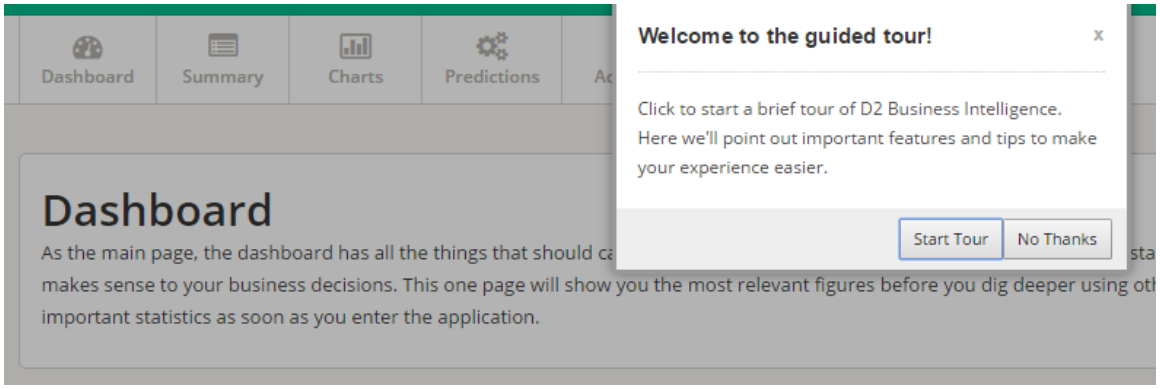
Screenshot 5: Predictions Page

This page predicts the trends of the important aspects of the business based on the patterns in available data and enables taking wiser decisions using them.



Screenshot 6: Account Page

Accounts page allows the users change personal details and password. This is important as the password provided by the administrator should be changed immediately.



Screenshot 7: About Page

This page is designed to give a brief guided tour of the application. It points out important features and tips to make the user experience easier.

4.2 Critical Analysis

The system is smooth and fulfills the objectives stated at the start of the project. It was built using Qualitative research and Test Driven Development methodology. Hence, the facts are well researched and the errors are either handled properly or prevented by all means. The end system has friendly design and a separate guided tour to make the users clear about using the features. Further information for support can be provided during training session. Nonetheless, no application can be guaranteed to be completely bug-free or an out-and-out success. A proper analysis of the research work and also the application is needed to explore the areas of improvement.

The literature review was the main focal point of the research. So, the concentration of the references was very high in that section. From a quantitative perspective, there are about 2 references per page in the literature review, and almost 1 reference per page in the whole document. Thus, the research reflects more or less acceptable academic standards quantitatively. However, since the research is qualitative, good attention has been paid to the quality of the references. So, it should demonstrate good scholarship and academically soundness. There was a great deal of learning regarding the research types,

analyzing facts, and developing a set of suggestions/solutions from the research. Nonetheless, there are still some places for improvement in terms of better summarization of the facts, and use of multiple citations.

The Critical Analysis of the system must be done by considering the Critical Success Factors (CSF) of the application which comprises a multitude of aspects that range from trivial needs to complex processes. These aspects can be listed out and elaborated as:

- **Capability of the Client**: How much investment can the client make?
- **Information Architecture**: How does the client store and manage information?
- **User Adoption**: How ready is the client to adopt the system?
- **Relevance of BI features**: What BI features mean the most to the client?
- **Product Performance**: How does the product perform in actual working environment?
- **Hidden Costs**: What are the hidden costs of using the application?
- **Competitors**: What are the competitive alternatives to the application?

Talking about the capability of the client, the application was developed using free tools and hence the final cost is lower than popular market products. Affordability is one of the main objectives of the system and the product price solely depends on minor implementation costs. Likewise, the tools used for transforming the information to a structure supported by the system are also free. So, information architecture can be maintained without adding major costs in implementation.

The readiness of the clients is a key issue and it has been demonstrated previously in this document that businesses are open to using business intelligence tools when the barriers to their adoption such as costs and complexity are overcome. A detailed analysis of the features relevant to medium scale businesses was also performed to ensure that the complexity is reduced to the acceptable margins. So, these issues have been addressed properly by the application.

Coming to the performance of the system, there were many problems faced during integration of development tools and libraries. Because of this, it was hard to make the system stable. After a lot of integration testing and tune-ups, the application performs

well as a bunch of different tools. The test data provided by Sagar International from their ecommerce website Baleyo.com worked well with the application. Nevertheless, the recent earthquake meant that there were no transactions in their website for more than a month. This disabled us from testing the predictive analysis feature of our application using real data. We just used the available data and tried to fit in that existing data into the predictive model to check the correctness of this feature. This must be one of the key limitations of this application.

Finally, there are no hidden costs related with the system. The system can be implemented in the existing resources providing great flexibility in deployment. Unless there are serious hardware limitations, there is no need to purchase anything. This makes it competitive on economic grounds. In depth comparison of the system with leading market products has already been done in the document showing that it can be a successful product. Therefore, even with its limitations, the product can do well in the targeted market. It can clearly improve and evolve later on to enhance its credibility.

4.3 Applications

It has been repeated a number of times in this document that the main application area of this system are medium scale enterprises. So, we focus on specific application of this system in this section. The system can be used to draw insights from different kinds of data sources. For the time being, it uses data from an ecommerce website based on Magento but it can be configured to work with different systems such as:

- Inventory Management System
- Budgeting System
- Task Management and Scheduling System
- Simple databases

4.4 Limitations and Future Enhancements

Like every application in its first version, the application has some problems and limitations. The limitations are mainly due to the need to keep the application simple to

use and low in cost. Many limitations were also identified during the development phase. So, the limitations of the system are:

- It cannot automatically migrate and synchronize the databases used for drawing information and results.
- There is no provision for customized reporting or publishing reports.
- Chart labeling in the UI could have been more interactive.
- Login and Sessions are not fully secure from latest hacking techniques.
- Predictive analysis features are very basic.
- Does not support multiple types of user accounts.

Based on the limitations of the system, the future enhancements to the system can be:

- Export, Transform and Load (ETL) feature
- Customized reporting and publishing
- Login and Sessions secured using additional security plugins.
- Powerful predictive features.
- Multiple user types and specific views for them.

4.5 Conclusion

In conclusion, the whole project presented a collection of ideas and a proof of concept for developing scalable, interactive and low-cost business intelligence system specifically for medium scale enterprises. It analyzed the features that are relevant to medium scale enterprises when it comes to having an affordable business intelligence solution. Likewise, different methods for predictive analysis were also scrutinized. In the end, a simple and interactive application was created with multiple deployment options. It has smooth operation with proper error handling without delay in rendering or processing. No major bugs have been found thus far. So, we can call this project a success.

REFERENCES

[1] Achmad Nizar Hidayanto, Robertus Kristianto, and Muhammad Rifki Shihab. (2015) Measuring Business Intelligence (BI) Implementation Readiness: A Framework Development and Its Application to Small Medium Enterprises (SMEs). [Online]. http://www.academia.edu/1525313/Measuring_Business_Intelligence_BI_Implementation_Readiness_A_Framework_Development_and_Its_Application_to_Small_Medium_Enterprises_SMEs_

[2] Information Builders. (2012) Eight Ways to Generate New Revenue With Business Intelligence and Analytics. [Online]. http://www.informationbuilders.com/about_us/whitepapers/download_form/8549

[3] Paul Phillips. (1999, July) White Paper: An insight into Business Intelligence. [Online]. http://www.computerweekly.com/feature/White-Paper-An-insight-into-Business-Intelligence

[4] Information Builders. (2013) Top Six Worst Practices in Business Intelligence. [Online]. http://www.informationbuilders.com/about_us/whitepapers/download_form/4599

[5] Information Builders. (2012) Anatomy of the New Decision. [Online]. http://www.informationbuilders.com/about_us/whitepapers/download_form/8464

[6] Frank C. Ziglar. (2006) Comparing Apache Tomcat Performance Across Platforms. [Online]. http://www.webperformance.com/library/reports/windows_vs_linux_part1/

[7] MongoDB. (2015, February) Performance Best Practices for MongoDB. [Online]. http://s3.amazonaws.com/info-mongodb-com/MongoDB-Performance-Best-Practices.pdf

[8] Roger Schwarz. (2015, March) How to Design an Agenda for an Effective Meeting. [Online]. https://hbr.org/2015/03/how-to-design-an-agenda-for-an-effective-meeting

[9] Patrick Yurgosky. (2012, June) Beyond Dashboards: Business Intelligence Tools for Program Analysis and Reporting. [Online]. http://www.idealware.org/articles/beyond-dashboards-business-intelligence-tools-program-analysis-and-reporting

[10] Jennifer Lonoff Schiff. (2014, August) 9 Common BI Software Mistakes (and How to Avoid Them). [Online]. http://www.cio.com/article/2464167/business-intelligence/9-common-bi-software-mistakes-and-how-to-avoid-them.html

[11] Claudia Imhoff, "Best Practices for BI: What We Have Learned in 20 years," Intelligent Solutions, 2009.

[12] M87Systems. (n.d.) BI Requirements Checklist. [Online]. http://www.m87systems.com/Assets/BIRequirementsChecklist.pdf

[13] IBM Corporation. (2013, October) IBM Cognos Enterprise. [Online]. http://www.olsps.com/analytic/pdf/Cognos/IBM-Cognos-Enterprise.PDF

[14] Tableau Software. (2011) Business Intelligence Requirements for IT. [Online]. https://www.tableau.com/sites/default/files/whitepapers/bi-requirements-it.pdf

[15] Mark Kerzner and Sujee Maniyam, "Business Intelligence Tools For Hadoop and Big Data," in *Hadoop Illuminated*.: Elephant Scale LLC, 2013, ch. 13. [Online]. http://hadoopilluminated.com/hadoop_illuminated/BI_Tools_For_Hadoop.html

[16] Daniel Humphries. (2015, March) Compare Business Intelligence (BI) Software Tools. [Online]. http://www.softwareadvice.com/bi/

[17] Douglas Henschen. (October, 2014) 2015 Analytics & BI Survey. [Online]. http://reports.informationweek.com/abstract/81/12544/Business-Intelligence-and-Information-Management/2015-Analytics---BI-Survey.html

[18] Yanchang Zhao, *R and Data Mining: Examples and Case Studies*.: Elsevier, 2013.

[19] Yogesh Singh, Pradeep Kumar Bhatia, and Omprakash Sangwan, "A Review of Studies On Machine Learning Techniques," *International Journal of Computer Science and Security (IJCSS)*, vol. I, no. 1, pp. 70-84, May/June 2007.

[20] W3Schools. (2015) CSS Introduction. [Online]. http://www.w3schools.com/css/css_intro.asp

[21] W3Schools. (2015) Bootstrap Get Started. [Online]. http://www.w3schools.com/bootstrap/bootstrap_get_started.asp

[22] W3Schools. (2015) JavaScript Introduction. [Online]. http://www.w3schools.com/js/js_intro.asp

[23] W3Schools. (2015) jQuery Introduction. [Online]. http://www.w3schools.com/jquery/jquery_intro.asp

[24] B. Beckwith, *Programming Grails*. Sebastopol, CA: O'Reilly Media, 2013, pp. 35-38.

[25] Grails. (2015) Sites using grails. [Online]. https://grails.org/websites

[26] Kurt Hornik. (n.d.) Frequently Asked Questions on R. [Online]. http://CRAN.R-project.org/doc/FAQ/R-FAQ.html

[27] R.A. Muenchen. (n.d.) The Popularity of Data Analysis Software. [Online]. http://r4stats.com/articles/popularity/

[28] Renjin. (2013) About. [Online]. http://www.renjin.org/

[29] MongoDB. (2015, February) RDBMS to MongoDB Migration Guide. [Online]. http://www.mongodb.com/thank-you/white-paper/migration-rdbms-nosql-mongodb

[30] MongoDB. (2015) SQL to MongoDB Mapping Chart. [Online]. http://docs.mongodb.org/manual/reference/sql-comparison/

[31] D Gadish, *Modern System Analysis and Design*.

[32] M. Rouse. (2005) Prototyping Model. [Online]. http://searchcio.techtarget.com/definition/Prototyping-Model

[33] TechnologyPlex. (2013).NET Agile Method of Software Development.. [Online]. http://tplex.com/services/dot-net-agile-development

[34] Asana. OUR STORY. [Online]. https://asana.com/company

[35] GitHub Inc. (2015) About GitHub. [Online]. https://github.com/about

[36] Grails. (2015) Deployment. [Online]. https://grails.org/wiki/Deployment

Appendix

Project Folder Structure

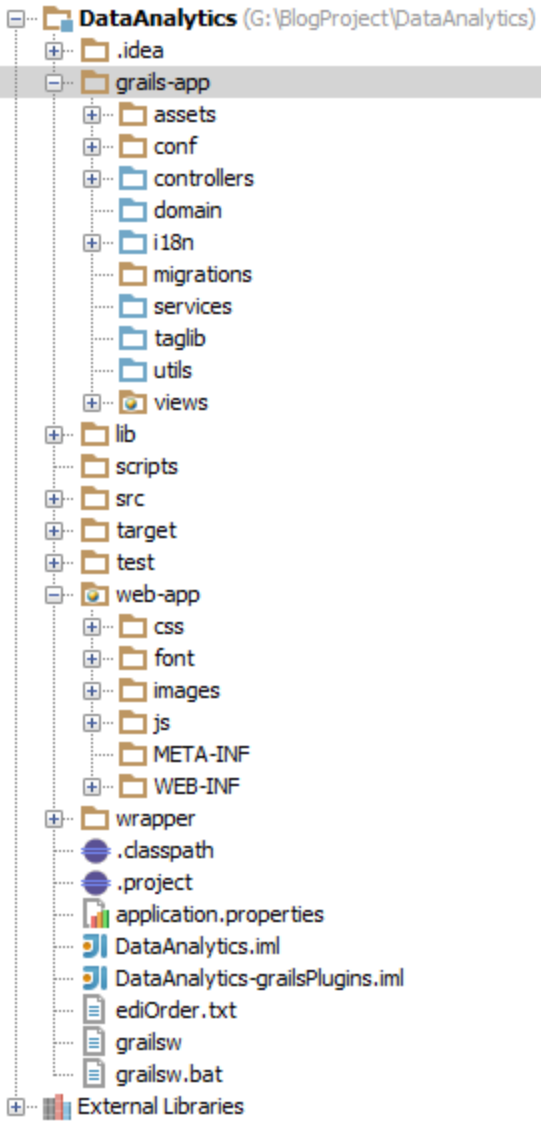


Fig: Project Folder Structure as shown in IntelliJ IDEA Ultimate 14

Sample Code Snippet: Main Controller

```
package dataanalytics

import R.LinearRegression
import R.DecisionTree
import MongoDB.tryMongoDB
import java.text.SimpleDateFormat

class IndexController {

    def index() {
    }

    def dashboard(){
        if(session["user"] != null){
            flash.message = session["user"];
        }
        else {
            redirect(action: 'index');
        }
        int[] results = tryMongoDB.testMongo();
        int totSales = results[3]/1000;
        int[] orderSold = tryMongoDB.SalesOrders();
        [
                webVisitors: results[0],
                orders: results[1],
                itemsInStock: results[2],
                totalSales: totSales,
                galleryItems: results[4],
                visibleItems: results[5],
                activeDiscounts: results[6],
                searches: results[7],
                backendUsers: results[8],
                itemsViewed: results[9],
                compared: results[10],
                wishlist: results[11],
                g: orderSold        ]
    }

    def logout(){
        session.invalidate()
        redirect(action: "index")
    }
}
```

Sample Code Snippet: Using MongoDB

```
package MongoDB;

import java.util.ArrayList;
import java.util.List;
import org.json.*;
import com.mongodb.*;

public class tryMongoDB {
    public static MongoClient mongoClient;
    public static DB db;

    public static void connect() throws Exception{
        mongoClient = new MongoClient("localhost" , 27017);
        db = mongoClient.getDB("local");
    }

    public static boolean authenticate(String user, String pass) throws
Exception{

        connect();
        Boolean result = false;
        DBObject cred = db.getCollection("users").findOne();
        String users = cred.toString();
        JSONObject items = new JSONObject(users);
        String username = items.getString("user");
        String password = items.getString("password");

        System.out.println(user + " "+ username + " " + pass + " " +
password);
        if((user.equals(username)) && (pass.equals(password))){
            result = true;
        }
        mongoClient.close();
        return result;
    }

    public static int[] testMongo() throws Exception {
        connect();
        int[] results = new int[12];

        //Sales and Visits
        results[0] = (int) db.getCollection("log_visitor").getCount();
//webVisitors

        //orders
        BasicDBObject orQuery = new BasicDBObject();

        List<BasicDBObject> obj = new ArrayList<BasicDBObject>();
        obj.add(new BasicDBObject("state", "new"));
        obj.add(new BasicDBObject("state", "pending_payment"));
        orQuery.put("$or", obj);
        DBCursor cursor1 =
```

```
db.getCollection("sales_flat_order").find(orQuery);
        results[1] =
(int)db.getCollection("sales_flat_order").getCount();
        cursor1.close();

        //items in stock
        DBObject stockItems = new BasicDBObject("$group", new
BasicDBObject("_id", "$qty")
                .append("count", new BasicDBObject("$sum", 1)));
        AggregationOutput output =
db.getCollection("cataloginventory_stock_item").aggregate(stockItems);

        JSONObject items;
        int count = 0;
        int qty = 0;
        int counter = 0;
        for (DBObject result : output.results()) {
            items = new JSONObject(result.toString());
            count = items.getInt("count");
            qty = (int)items.getDouble("_id");
            counter = counter + (count*qty);
            //System.out.println(result);
        }
        results[2] = counter;

        //total sales worth
        int total = 0;
        DBObject salesWorth = new BasicDBObject("$group", new
BasicDBObject("_id", "$state")
                .append("total", new BasicDBObject("$sum",
"$base_grand_total")));
        AggregationOutput output1 =
db.getCollection("sales_flat_order").aggregate(salesWorth);
        for (DBObject result : output1.results()) {
            items = new JSONObject(result.toString());
            total = items.getInt("total"); //total comes at last
        }
        results[3] = total;

        //Quick Stats
        results[4] = (int)
db.getCollection("catalog_product_entity_media_gallery").getCount();
//items in gallery
        //visible items
        BasicDBObject whereQuery = new BasicDBObject();
        whereQuery.put("is_visible_on_front", 1);
        DBCursor cursor3 =
db.getCollection("catalog_eav_attribute").find(whereQuery);
        results[5] = cursor3.count();
        cursor3.close();

        //active discounts
        BasicDBObject discQuery = new BasicDBObject();
        discQuery.put("is_active", 1);
        DBCursor cursor4 =
```

```
db.getCollection("catalogrule").find(discQuery);
        results[6] = cursor4.count();  //active discounts
        cursor4.close();

        results[7] = (int)
db.getCollection("catalogsearch_query").getCount();  //searches

        results[8] = (int)
db.getCollection("customer_entity").getCount();  //backend users

        List l1 =
db.getCollection("report_viewed_product_index").distinct("product_id");
//most viewed products
        results[9] = l1.size();

        List l2 =
db.getCollection("report_compared_product_index").distinct("product_id"
); //most compared products
        results[10] = l2.size();
        results[11] = (int)

db.getCollection("wishlist_item").getCount();    //wishlist
        mongoClient.close();
        return results;
    }
}
```

Sample Code Snippet: Testing Linear Regression using Renjin

```
import javax.script.*;

public class LinearRegression {

    public static double test(int val) throws Exception {
        double[] results = new double[2];
        // create a script engine manager:
        ScriptEngineManager manager = new ScriptEngineManager();
        // create a Renjin engine:
        ScriptEngine engine = manager.getEngineByName("Renjin");
        // check if the engine has loaded correctly:
        if(engine == null) {
            throw new RuntimeException("Renjin Script Engine not found
on the classpath.");
        }
        // ... evaluation Java code here ...
        System.out.println();

        String x = "(1,2,5,7,9,11)";
        String y = "(2,4,10,14,18,21)";
        engine.eval("x <- c" + x);
        engine.eval("y <- c" + y);
        engine.eval("df <- data.frame(x, y)");
        // String result2 =engine.eval("print(df)").toString();
        engine.eval("z <-(lm(y ~ x, df))");
        String result3 =engine.eval("summary(z)$coefficients[1,
1]\n").toString();
        String result4 =engine.eval("summary(z)$coefficients[2,
1]\n").toString();

        results[0] = Double.parseDouble(result3);
        results[1] = Double.parseDouble(result4);
        double result = results[0] * val + results[1];
        return result;
    }
}
```

Sample Code Snippet: UI Layout

```
<%--
  Created by IntelliJ IDEA.
  User: uclid
  Date: 3/17/2015
  Time: 7:33 AM
--%>

<%@ page contentType="text/html;charset=UTF-8" %>
<!DOCTYPE html>
<html lang="en">
    <head>
        <meta charset="utf-8">
        <title>Dashboard</title>
        <link rel="shortcut icon" href="${resource(dir: 'images', file:
'favicon.ico')}" type="image/x-icon">
        <meta name="viewport" content="width=device-width, initial-
scale=1.0, maximum-scale=1.0, user-scalable=no">
        <meta name="apple-mobile-web-app-capable" content="yes">
        <link href="${resource(dir: 'css', file: 'bootstrap.min.css')}"
rel="stylesheet">
        <link href="${resource(dir: 'css', file: 'bootstrap-
responsive.min.css')}" rel="stylesheet">
        <link href="${resource(dir: 'css', file: 'font-awesome.css')}"
rel="stylesheet">
        <link href="${resource(dir: 'css', file: 'style.css')}"
rel="stylesheet">
        <link href="${resource(dir: 'css/pages', file:
'dashboard.css')}" rel="stylesheet">
        <!-- Le HTML5 shim, for IE6-8 support of HTML5 elements -->
        <!--[if lt IE 9]>
          <script
src="http://html5shim.googlecode.com/svn/trunk/html5.js"></script>
        <![endif]-->
    </head>
    <body>
        <div class="navbar navbar-fixed-top">
            <div class="navbar-inner">
                <div class="container"> <a class="btn btn-navbar" data-
toggle="collapse" data-target=".nav-collapse"><span
                        class="icon-bar"></span><span class="icon-
bar"></span><span class="icon-bar"></span> </a><a class="brand"
href="dashboard.gsp"><img src ="${resource(dir: 'images', file:
'applogo.png')}" alt="logo"/> D2 - Business Intelligence</a>
                    <div class="nav-collapse">
                        <ul class="nav pull-right">
                            <li class="dropdown"><a href="#"
class="dropdown-toggle" data-toggle="dropdown"><i
                                    class="icon-user"></i>
                                <g:if test="${flash.message}">
                                    ${flash.message }
                                </g:if>
                                <b class="caret"></b></a>
                                <ul class="dropdown-menu">
```

```
                                    <li><g:link
action="account">Account</g:link></li>
                                    <li><g:link
action="logout">Logout</g:link></li>
                                </ul>
                            </li>
                        </ul>
                    </div>
                    <!--/.nav-collapse -->
                </div>
                <!-- /container -->
            </div>
            <!-- /navbar-inner -->
        </div>
        <!-- /navbar -->
        <div class="subnavbar">
            <div class="subnavbar-inner">
                <div class="container">
                    <ul class="mainnav">
                        <li class="active"><g:link
action="dashboard"><i class="icon-
dashboard"></i><span>Dashboard</span></g:link></li>
                        <li><g:link action="summary"><i class="icon-
list-alt"></i><span>Summary</span></g:link></li>
                        <li><g:link action="charts"><i class="icon-bar-
chart"></i><span>Charts</span></g:link></li>
                        <li><g:link action="prediction"><i class="icon-
cogs"></i><span>Predictions</span></g:link></li>
                        <li><g:link action="account"><i class="icon-
user"></i><span>Account</span></g:link></li>
                        <li><g:link action="guidely"><i class="icon-
question-sign"></i><span>About</span></g:link></li>
                    </ul>
                </div>
                <!-- /container -->
            </div>
            <!-- /subnavbar-inner -->
        </div>
        <!-- /subnavbar -->
        <div class="main">
            <div class="main-inner">
                <div class="container">
                    <div class="row">
                        <div class="span6">
                            <div class="widget widget-nopad">
                                <div class="widget-header"> <i
class="icon-list-alt"></i>
                                    <h3>Sales and Visits</h3>
                                </div>
                                <!-- /widget-header -->
                                <div class="widget-content">
                                    <div class="widget big-stats-
container">
                                        <div class="widget-content">
                                            <h6
```

```
class="bigstats">Welcome to D2 Business Intelligence - A secure,
interactive and sustainable Business Intelligence system that enables
data visualization and predictive analysis.</h6>
                                        <div id="big_stats"
class="cf">
                                            <div class="stat"> <i
title="Visitors" class="icon-group"></i> <span
class="value">${webVisitors}</span> </div>
                                            <!-- .stat visitors-->

                                            <div class="stat"> <i
title="Orders" class="icon-shopping-cart"></i> <span
class="value">${orders}</span> </div>
                                            <!-- .stat orders-->

                                            <div class="stat"> <i
title="Items in Stock" class="icon-truck"></i> <span
class="value">${itemsInStock}</span> </div>
                                            <!-- .stat stock-->

                                            <div class="stat"> <i
title="Total Sales Worth" class="icon-money"></i> <span
class="value">Rs.${totalSales}k</span> </div>
                                            <!-- .stat sales-->
                                        </div>
                                    </div>
                                    <!-- /widget-content -->

                                </div>
                            </div>
                        </div>
                        <div class="widget widget-table action-
table">
                            <div class="widget-header"> <i
class="icon-th-list"></i>
                                <h3>Most Sold Products</h3>
                            </div>
                            <!-- /widget-header -->
                            <div class="widget-content">
                                <table class="table table-striped
table-bordered">
                                    <thead>
                                    <tr>
                                        <th> Product Name </th>
                                        <th> Product Value</th>
                                        <th class="td-actions"> In
Stock </th>
                                    </tr>
                                    </thead>
                                    <tbody>
                                    <tr>
                                        <td> Kaspersky Antivirus
</td>
                                        <td> Rs. 1,200 </td>
                                        <td class="td-actions"><a
```

```
href="javascript:;" class="btn btn-success btn-small"><i class="btn-
icon-only icon-ok"> </i></a></td>
                                        </tr>
                                        <tr>
                                            <td> Brother Printer </td>
                                            <td> Rs. 10,000 </td>
                                            <td class="td-actions"><a
href="javascript:;" class="btn btn-success btn-small"><i class="btn-
icon-only icon-ok"> </i></a></td>
                                        </tr>
                                        <tr>
                                            <td> Prestigio Mobile </td>
                                            <td> Rs. 17,000 </td>
                                            <td class="td-actions"><a
href="javascript:;" class="btn btn-small btn-success"><i class="btn-
icon-only icon-ok"> </i></a></td>
                                        </tr>
                                        <tr>
                                            <td> NEC Projector </td>
                                            <td> Rs. 40,000 </td>
                                            <td class="td-actions"><a
href="javascript:;" class="btn btn-danger btn-small"><i class="btn-
icon-only icon-remove"> </i></a></td>
                                        </tr>
                                        <tr>
                                            <td> Prestigio Tablet </td>
                                            <td> Rs. 10,000 </td>
                                            <td class="td-actions"><a
href="javascript:;" class="btn btn-small btn-success"><i class="btn-
icon-only icon-ok"> </i></a></td>
                                        </tr>
                                        <tr>
                                            <td> USB Mouse</td>
                                            <td> Rs. 700 </td>
                                            <td class="td-actions"><a
href="javascript:;" class="btn btn-danger btn-small"><i class="btn-
icon-only icon-remove"> </i></a></td>
                                        </tr>

                                        </tbody>
                                    </table>
                                </div>
                                <!-- /widget-content -->
                            </div>
                        </div>
                        <!-- /span6 -->
                        <div class="span6">
                            <div class="widget">
                                <div class="widget-header"> <i
class="icon-bookmark"></i>
                                    <h3>Quick Stats</h3>
                                </div>
                                <!-- /widget-header -->
                                <div class="widget-content">
                                    <div class="shortcuts">
```

```
                                        <a href="javascript:;"
class="shortcut"><i class="shortcut-icon icon-sitemap"></i>
                                            <span class="shortcut-
label">${galleryItems} Gallery Items</span></a>
                                        <a href="javascript:;"
class="shortcut"><i class="shortcut-icon icon-eye-open"></i>
                                            <span class="shortcut-
label">${visibleItems} Visible Items</span> </a>
                                        <a href="javascript:;"
class="shortcut"><i class="shortcut-icon icon-gift"></i>
                                            <span class="shortcut-
label">${activeDiscounts} Active Discounts</span> </a>
                                        <a href="javascript:;"
class="shortcut"> <i class="shortcut-icon icon-search"></i>
                                            <span class="shortcut-
label">${searches} Searches</span> </a>
                                        <a href="javascript:;"
class="shortcut"><i class="shortcut-icon icon-user"></i>
                                            <span class="shortcut-
label">${backendUsers} Backend Users</span> </a>
                                        <a href="javascript:;"
class="shortcut"><i class="shortcut-icon icon-bar-chart"></i>
                                            <span class="shortcut-
label">${itemsViewed} Items Viewed</span> </a>
                                        <a href="javascript:;"
class="shortcut"><i class="shortcut-icon icon-check"></i>
                                            <span class="shortcut-
label">${compared} Compared</span> </a>
                                        <a href="javascript:;"
class="shortcut"> <i class="shortcut-icon icon-pushpin"></i>
                                            <span class="shortcut-
label">${wishlist} in Wishlist</span> </a>
                                    </div>
                                    <!-- /shortcuts -->
                                </div>
                                <!-- /widget-content -->
                            </div>
                            <!-- /widget -->
                            <div class="widget">
                                <div class="widget-header"> <i
class="icon-signal"></i>
                                    <h3> Orders vs Sold Products</h3>
                                </div>
                                <!-- /widget-header -->
                                <div class="widget-content">
                                    <canvas id="area-chart"
class="chart-holder" height="250" width="538"> </canvas>
                                    <div id="bar-legend" class="chart-
legend"></div>
                                    <!-- /area-chart -->
                                </div>
                                <!-- /widget-content -->
                            </div>
                        </div>
                        <!-- /span6 -->
```

```
                    </div>
                    <!-- /row -->
                </div>
                <!-- /container -->
            </div>
            <!-- /main-inner -->
        </div>
        <!-- /main -->
        <div class="extra">
            <div class="extra-inner">
                <div class="container">
                    <div class="row">
                        <div class="span3">
                            <h4>
                                Powered By:</h4>
                            <ul>
                                <li><img src= "${resource(dir:
'images', file: 'uclid.png')}" alt="uclid"/></li>
                            </ul>
                        </div>
                        <!-- /span3 -->
                        <div class="span3">
                            <h4>
                                Data Provided By:</h4>
                            <ul>
                                <li><img src= "${resource(dir:
'images', file: 'logo.png')}" alt="uclid"/></li>
                            </ul>
                        </div>
                        <!-- /span3 -->
                        <div class="span3">
                            <h4>
                                Specific Data Source:</h4>
                            <ul>
                                <li><img src= "${resource(dir:
'images', file: 'baleyo.png')}" alt="uclid"/></li>
                            </ul>
                        </div>
                    </div>
                    <!-- /row -->
                </div>
                <!-- /container -->
            </div>
            <!-- /extra-inner -->
        </div>
        <!-- /extra -->
        <div class="footer">
            <div class="footer-inner">
                <div class="container">
                    <div class="row">
                        <div class="span12"> &copy; 2015 D2 Business
Intelligence. </div>
                        <!-- /span12 -->
                    </div>
                    <!-- /row -->
```

```
                </div>
                <!-- /container -->
            </div>
            <!-- /footer-inner -->
        </div>
        <!-- /footer -->
        <!-- Placed at the end of the document so the pages load faster
-->
        <script src="${resource(dir: 'js', file: 'jquery-
1.7.2.min.js')}"></script>
        <script src="${resource(dir: 'js', file:
'excanvas.min.js')}"></script>
        <script src="${resource(dir: 'js', file: 'chart.min.js')}"
type="text/javascript"></script>
        <script src="${resource(dir: 'js', file:
'bootstrap.js')}"></script>
        <script src="${resource(dir: 'js', file: 'base.js')}"></script>
        <script>

            var lineChartData = {
                labels: ["January", "February", "March", "April",
"May", "June",
"July","August","September","October","November","December"],
                datasets: [
                    {
                        fillColor: "rgba(220,220,220,0.5)",
                        strokeColor: "rgba(220,220,220,1)",
                        pointColor: "rgba(220,220,220,1)",
                        pointStrokeColor: "#fff",
                        data:
[${g[0]},${g[1]},${g[2]},${g[3]},${g[4]},${g[5]},${g[6]},${g[7]},${g[8]
},${g[9]},${g[10]},${g[11]}],
                        label: "Orders"
                    },
                    {
                        fillColor: "rgba(151,187,205,0.5)",
                        strokeColor: "rgba(151,187,205,1)",
                        pointColor: "rgba(151,187,205,1)",
                        pointStrokeColor: "#fff",
                        data:
[${g[12]},${g[13]},${g[14]},${g[15]},${g[16]},${g[17]},${g[18]},${g[19]
},${g[20]},${g[21]},${g[22]},${g[23]}],
                        label:"Sold"
                    }
                ]
            }

            var myLine = new Chart(document.getElementById("area-
chart").getContext("2d")).Line(lineChartData);
            document.getElementById('bar-legend').innerHTML =
myLine.generateLegend();

        </script>
    </body>
</html>
```