SLIIT

COMPUTING | BUSINESS | ENGINEERING

Smart Resume CV selector for IT Industry

System Requirements Specification

Project ID: 18-005

Name – Madhushika K.A

Student ID- IT15097798

B.Sc. Special (Hons) in Information Technology specialized in Software Engineering

Submitted on - 08-May-2018

DECLARATION

I hereby declare that the submitted project Software Requirements Specification document for Smart Resume CV selector is an original work done by Madhushika K.A. This document is proprietary and an exclusive property of the SLIIT project group 18-005. List of references I referred for preparation of this document is given references at the end of the document.

Member : IT15097798 – Madhushika K.A

Signature: …….

Approval Block

	Name	Partner	Date	Signature
From				
Verified By				
Approved By				

Revision History

Version	Date	Summary of Changes	Author

Table of Contents

Figures

Tables

1 INTRODUCTION

1.1 Purpose

This project intended to illustrate the requirements of the Project **Smart Resume CV Selector.** In this document, it describes all the necessary information related to the research project. The main objective of this document is to illustrate a description of software requirements for **Smart Resume CV Selector** business intelligence tool. In this SRS document, it further analyzes user interfaces, the flow of the project, user requirements, purpose and features of the system, what the system will do, the constraint under which it must operate and other factors necessary to provide a complete and comprehensive description of requirements for the project

This document is mainly targeted to the project supervisor, co-supervisor, and the research team members. The document is written in a form that any person can read and understand the content of this document. Also, it will be a benefit for former researchers who are interested to implement this kind of an application.

1.2 Scope

Smart Resume CV Selector is a Business intelligent tool which analyzes and classify operational data with algorithms and presents a web application that facilitates task automating to select most suitable and most qualified candidates depending on attributes given by the user. An **ETL** (Extract Transform and Load) tool will analyze the data and load them into a data warehouse then prediction model will design the optimal and feasible solution in **IT industry**. Depending on the attributes **Smart Resume** will dynamically visualize the candidate list.

The document is organized into several sections in order to give a clear understanding of project requirements. And it gives the detailed description of both functional and non-functional requirements for the above-mentioned system. Functional behaviors are associated with the specific functions, tasks or behaviors of the system. Nonfunctional requirements are the constraints on various

attributes of the function which need to be considered. This SRS covers the requirements for the **Data warehousing using ETL tool**. If the requirements changes in the future it is possible to change the specification accordingly. Also covers the hardware and software requirements need to implement the **ETL tool** and covers the limitations.

1.2.1 Objectives

Creating an ETL (Extract Transform and Load) Tool

- **Extract Data**
- **Transform Data**
- **Load Data**

The main objective is to create an ETL (Extract Transform and Load) tool to map the CSV (comma separated values) files data into Data warehouse created using MySQL [2]. ETL is a tool that maps data into data warehouse. From this tool, data can be map into data warehouse without inserting manually into the database. Data can be cleaned automatically by removing redundant data make unorganized data into organized data formats and validate the data in CSV files. Finally, map them into relevant attributes in database tables. So this provides most accuracy database mapping.

Other Objectives

- To make the Business intelligent tool easy to understand and easy to use for non-techy people
- Develop a system with high accuracy, efficiency, flexibility, and support for other non-functional requirements

1.2.2 Benefits

- Smart Resume is a comprehensive product with ETL tool and visualization features so there is no need to buy separate ETL tool
- Small or large all IT industry firms can be use this product
- Reduce time to check CV and analyze them
- Reduce time use to insert data one by one.

1.3 Definitions, Acronyms, and Abbreviations

SRS	Software Requirement Specification
ETL	Extract Transform Load Tool
DW	Data Warehouse
MySQL	MySQL database
CSV	Comma separated value
CV	Curriculum vitae
GUI	Graphical user interface
IT	Information Technology
BI	Business Intelligence
HR	Human Resource

Table 1- Definitions, Acronyms, and Abbreviations

1.4 Overview

This SRS document intent to cover all the functional and non-functional requirements of the **Smart Resume Data Warehousing.** Each of them has been discussed clearly in details. All are described under three chapters. It will also give an overall description that presents a background of general factors that effect to the requirements, a summary of main functions that system will perform, and general characteristics of the users of the product. It provides the purpose of the SRS, users, objectives, goals, and benefits of the system.

The second chapter contains software and hardware interfaces that system contains, requirements of the system, summary of major functionalities, an overview of the system.

The third chapter contains references, indexes, and appendices. This document can be used as a guide by the development team in the development phase.

1.4.1 Goals and Tasks

- Identify the Relevant and most suitable attributes to store in the data warehouse(Extract)
- Validate the data in CSV rows
- Remove redundant data, cleaned unorganized data (Data Cleansing)
- Transform data into a normalized format.
- Map data into a data warehouse
- Create interfaces for the user

1.4.2 Users

- HR workers in IT firms
- DB Administrator

1.4.3 System Overview

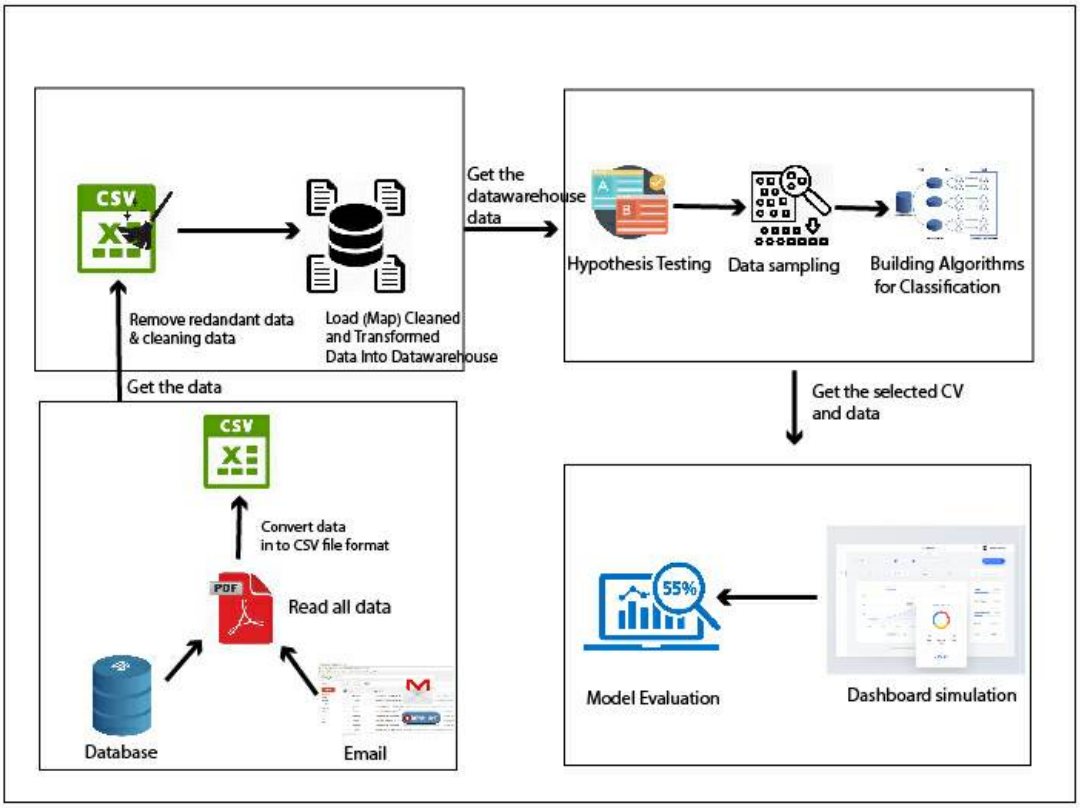

Figure 1-System Overview

2. OVERALL DESCRIPTION

Smart Resume- Is a Business intelligence toot that selects most suitable candidates for the given user inputs. The system involves data warehousing, data mining, and predictive modeling. **Smart Resume** turn data into the decision.

Smart Resume ETL Tool involve to store data from CSV files and store them in a data warehouse for further analysis and decision-making processes. To get data set for the predictive model's data should be stored in a data warehouse. To store those set of data will create an ETL tool to get the Dataset of CSV files.
In order to save data in data warehouse, data should be in a normalized format and fit into the data warehouse. For this purpose, data must be Extracted, Transform and d cleansing the data by removing redundant data and validating data sets. Lastly, dataset map into a data warehouse for further processing.

2.1 Product Perspective

- SQL Server Integration Service (SSIS)

  SSIS[4] is a component of a SQL server Database which performs integration on windows environment. It is quite expensive ETL tool in the market. Main disadvantage of this tool is it only operate on windows platforms only.

2.2 System Interfaces

- Web- Desktop connectivity Interface
- JVM
- Data Interaction Interface

2.2.1 User interface

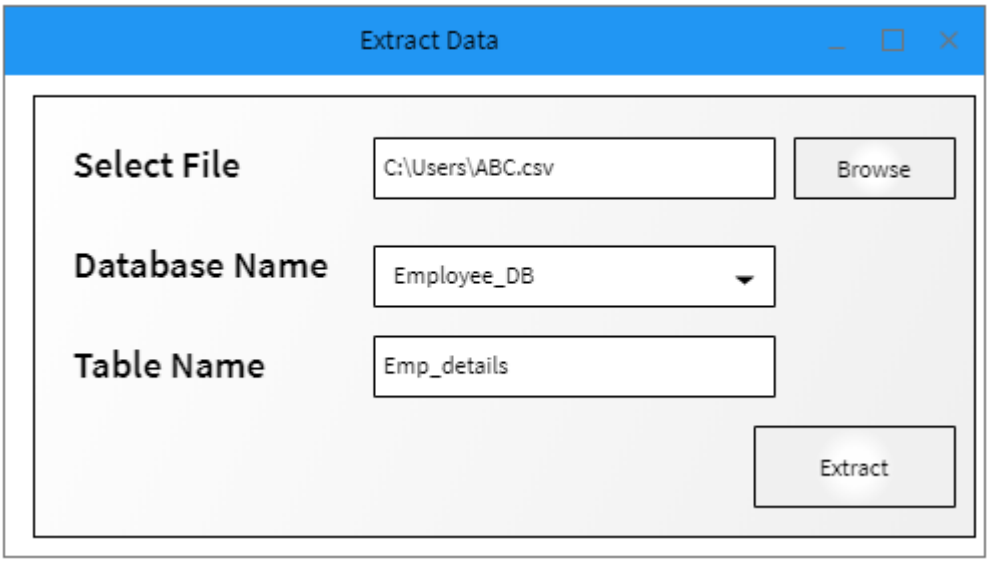

Figure 2- Data Extraction Interface

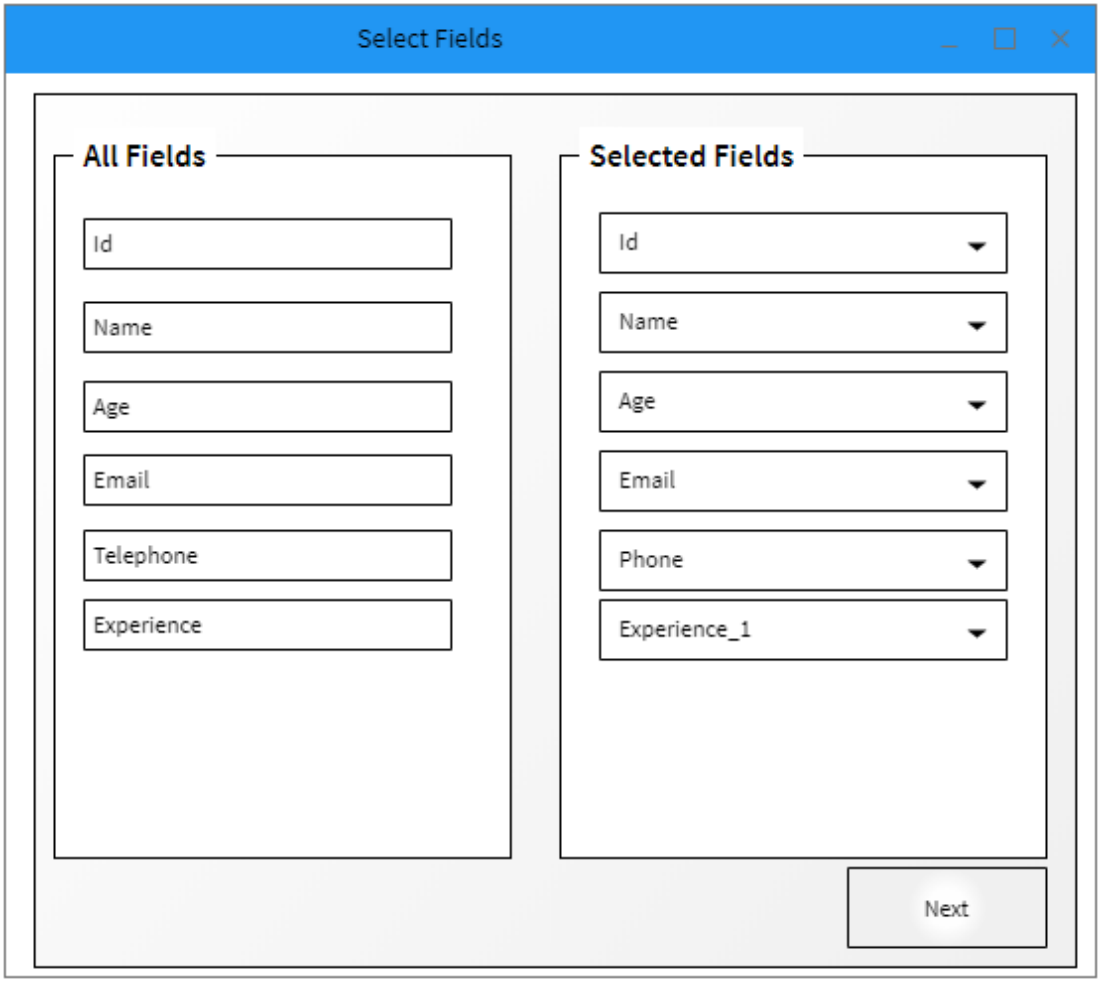

Figure 3- Field Selection Interface

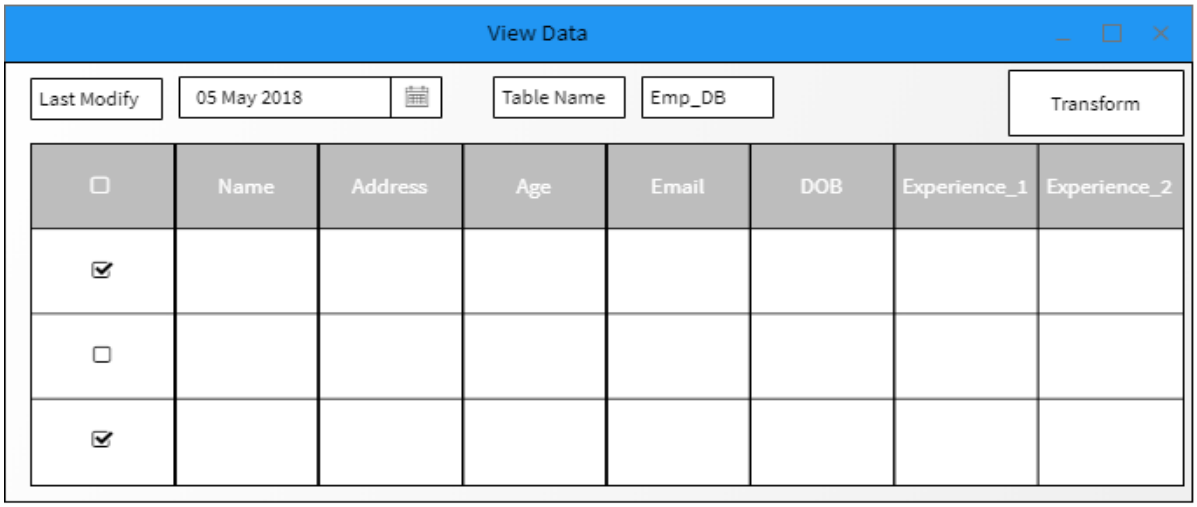

Figure 4- View Data Interface

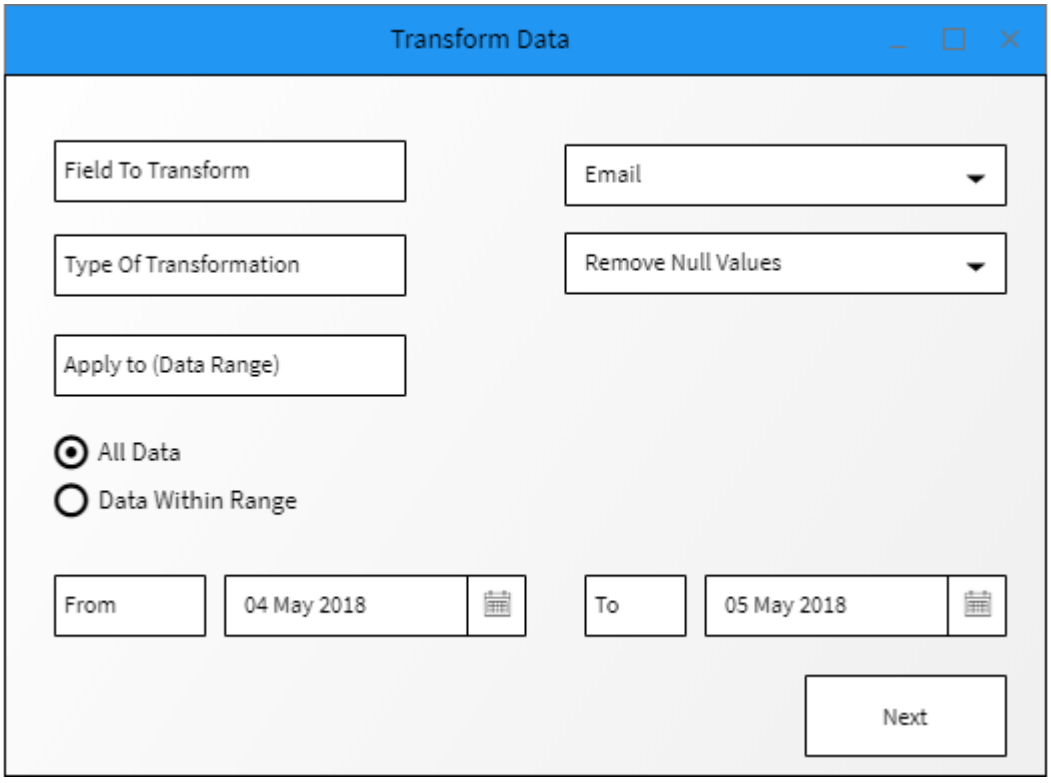

Figure 5- Data Transform Interface

2.2.2 Hardware Interfaces

Smart Resume ETL Tool does not require any special Hardware interfaces apart from a computer.

2.2.3 Software Interfaces

- MySQL
- Jython Library

2.2.4 Commutation Interface

- Internet
- Database Connection Interface

2.2.5 Memory Constraints

- RAM of 2GB or higher

2.2.6 Operations

- Operations of Database Administrator:
  - Extract Data From Given Source.
  - Cleansing of data with necessary attributes and get it transformed.
  - Load the data set to the data mart.
- Operations of Top Level Management
  - Extract, Transform and Load Dataset to the Data warehouse

2.2.7 Site Adaptation Requirements

- The server must have MySQL server installed on it.
- The user machine should have Java Virtual Machine installed.

2.3 Product functions

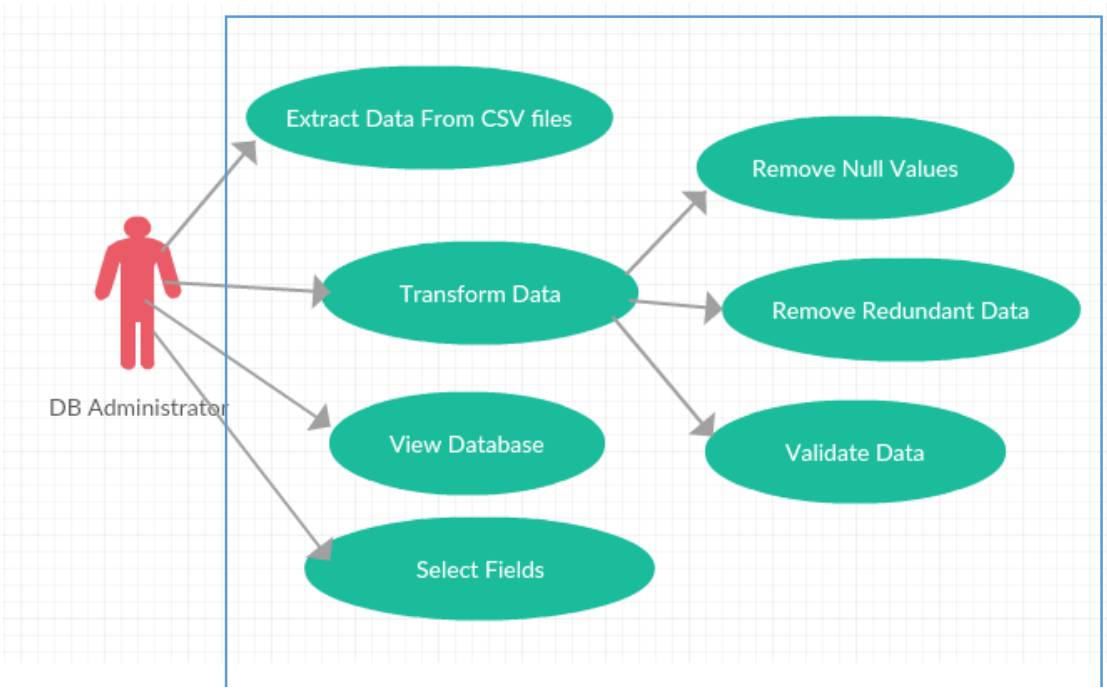

Figure 6- Use case diagram for ETL Tool

Use Case No	01
Use Case Name	Extract Data
Pre-Conditions	Database connection is active
Actors	Database Administrator
Main Success Scenario	1. Select the source csv file 2. Select or check the target database 3. Check the database table name 4. Click “Extract” button
Extensions	1a. Invalid Files 1b. Select Valid csv file 4a. Invalid source files 4b. Select a valid file and click “Extract”

Table 2 - Use case scenario for extracting Data

Use Case No	02
Use Case Name	Select Fields
Pre-Conditions	Csv file is selected Database connection active
Actors	Database Administrator
Main Success Scenario	1. check necessary fields selected 2.Click “Next”
Extensions	1a. Select relevant field from dropdown 1b.Click “Next”

Table 3 - Use case scenario for Selecting Fields

Use Case No	03
Use Case Name	View Database Table
Pre-Conditions	Database connection is active
Actors	Database Administrator
Main Success Scenario	1. User select the Database table 2. Click “Show Data” 3. System display Database table with data
Extensions	

Table 4 - Use case scenario for viewing data

Use Case No	04
Use Case Name	Transform Data(Remove null values)
Pre-Conditions	Database connection active
Actors	Database administrator
Main Success Scenario	1. Select the database table 2.Select Field to Transform 3. select type as “Remove null values” from dropdown list 4.select the data range 5.Click “Next” 6.Specify the default value to be replace 7.Click “OK”
Extensions	7a. No data within the range 7b. specify a different range and Click “OK” 8a. user Click “Cancel” 8b.cancel the process.

Table 5 - Use case scenario for Transforming Data by Removing Null values

Use Case No	05
Use Case Name	Transform Data – Remove Redundant Data
Pre-Conditions	Database Connection is active
Actors	Database administrator
Main Success Scenario	1.Select the Database table 2.Select type as “Remove redundant Data” 3.Click “OK”
Extensions	3a. No data within the given data range 3b. user click “Cancel” 3c.Cancel the process

Table 6 - Use case scenario for Transforming Data by Removing Redundant Data

Use Case No	06
Use Case Name	Validating Fields
Pre-Conditions	Database Connection is active
Actors	Database Administrator
Main Success Scenario	1.Select the Field that want to validate 2. select type as “Validating” 3.Specify the Default value to be replace with 4.Click ok
Extensions	1a No data within the field 4a. user Click “Cancel” 4b. Cancel the validation process.

Table 7 - Use case scenario for Transforming Data by Validating Data

2.4 User Characteristics

User	Privileges	Activities
Database Administrator	Full Access to the Database system	• Extracting data from CSV files. • Transform Data into normalize format • Load the Data into Data warehouse

Table 8- User Characteristics

User must not be very techy person and not an intelligence professional about business intelligence. But user must have an idea about Database Management.

This system can be used by several users such as

- Database Administrators
- Top Level Management
- HR Managers

2.5 Constraints

- Major constraint will be limitation of available time. ETL Tool expected to complete within 5 month of time period
- All the tools and technologies should be open source.

2.6 Assumption and dependencies

- All the users have basic knowledge of using computer and internet.
- There is an active internet connection
- Servers up and running 24x7 hours
- Sufficient memory and processing powers in all computes.

2.7 Appointment of Requirements

Example: The requirements described in sections 1 and 2 of this document are referred to as primary specifications; those in section 3 are referred to as requirements (or functional) specifications. The two levels of requirements are intended to be consistent. Inconsistencies are to be logged as defects. In the event that a requirement is stated within both primary and functional specifications, the application will be built from functional specification since it is more detailed.

3. SPECIFIC REQUIREMENTS

3.1 External Interface requirements

3.1.1 User Interfaces

- **Data Extraction Interface:**

  This is the first step of extracting data. Data shroud be loaded from CSV files. So data should store in CSV files in order to extract data to the data warehouse. Loaded Database and the table can be seen by user and if they want they can provide Database name in necessary time. [6]

- **Field Selection Interface**

  In this interface user can see what the attributes map into data warehouse are and what are the selected table headings of database table. The can select from dropdown list if there are any changes to apply. Left side list have extracted attributes and right side has mapping attributes in the database table. User should click next to proceed to the next step.

- **View Data Interface**

  In this interface user can see what are the available database tables and its data. On top of the interface user can see last modified data and table name. If they want further transformation they can transform from there.

- **Data Transformation Interface**

  This interface use to transform data if user want to transform. When mapping normally cleaning data is done, but if user want further transformation user can transform by selecting relevant transform type from the dropdown box and also giving relevant attribute and data ranges that they want to transform.

3.1.2 Hardware Interfaces

No special Hardware interfaces required for the system.

3.1.3 Software Interfaces

- **MySQL**

  MySQL is used for database management system for the system. For data transformation and data mapping use MySQL. Also to create data marts.

- **Jython Library**

  Use to connect interfaces between java and python. As all interfaces and validations develop by java and use python for calculations and algorithm developments.

3.1.4 Communication interfaces

- **Internet Connection**

  Data processing is done offline mode but in order to access web application user need an internet connection.

- **Database Connection Interface**

  This interface is use to communicate between database and the system. It acts as the adapter which converts database queries into system and vise-versa.

3.3 Performance Requirements

- Database can handle 50000 records.

3.4 Design Constraints

- The system based on IT industry only.
- Limitation of available time to develop the system.
- Performance, accuracy, reliability, security should achieve.

3.5 Software system attributes

3.5.1 Reliability

Reliability is the probability that an application will accurately perform its task under stated environmental conditions. It means that system support to allow user to work normally in many environmental conditions. Application must be tested unit vise and integrated vise in order to eliminate errors and give user a proper reliability. All technologies and configuration is used to provide better reliability to the system.

3.5.2 Availability

Database servers are running 24x7 hours. So that ETL tool can be performed in every time. [5] Immediate server's aggregation to data warehouse can increase high availability

3.5.3 Security

It indicates how to protect application from unauthorized people. Database has usernames and passwords in order to grant security. Only authorized people can access to the database and work with ETL Tool.

3.5.4 Maintainability

- Proper coding standards and naming conventions will be used at development time.
- Each attributes are validated using proper transformation functions.
- Creating user friendly interfaces with some standards.

3.6 Other Requirements

- Data that Map into Database are acquire from valid sources for further processing
- Use open source technologies.

4. SUPPORTING INFORMATION

4.1 Appendices

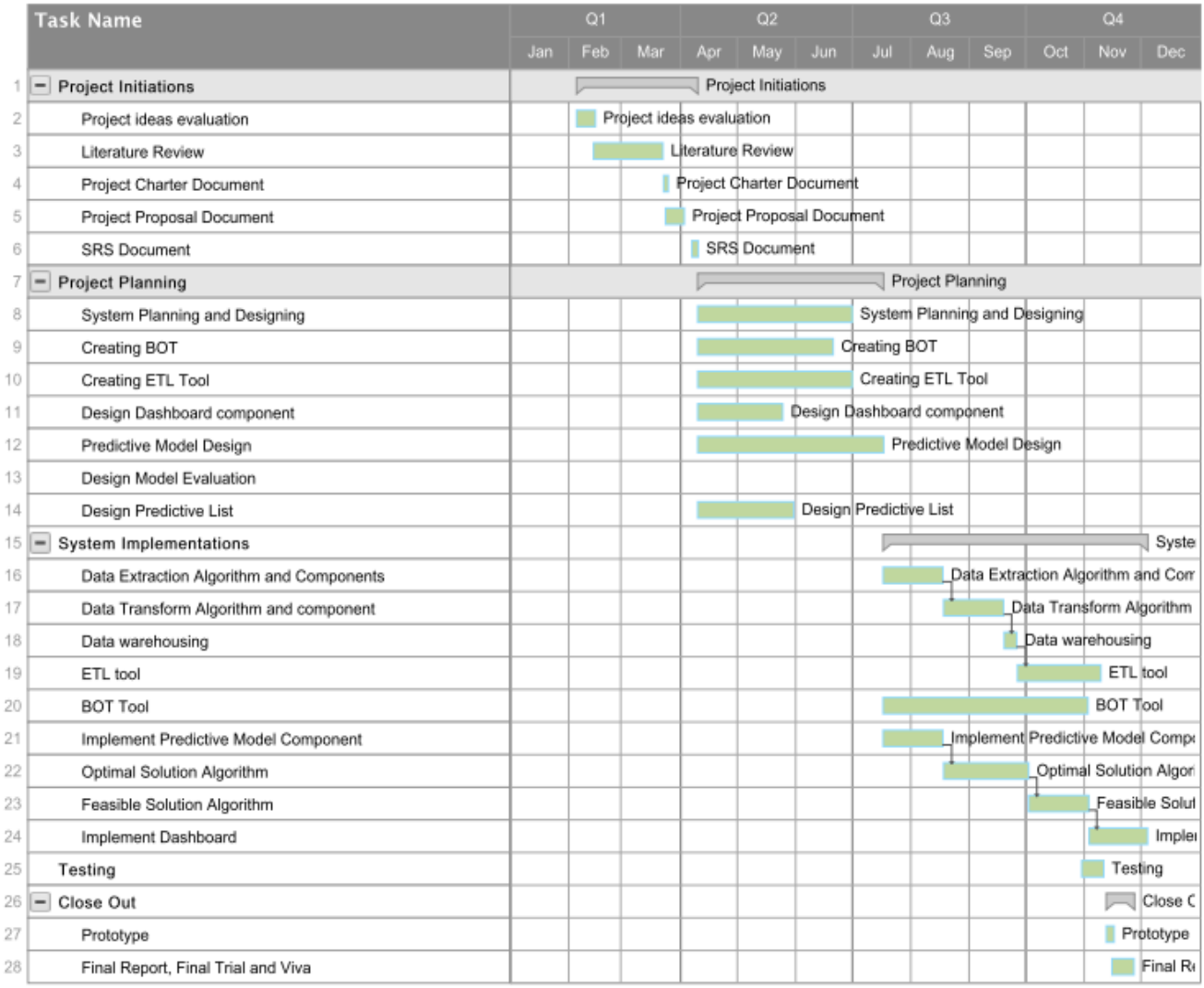

Figure 7- Grant chart

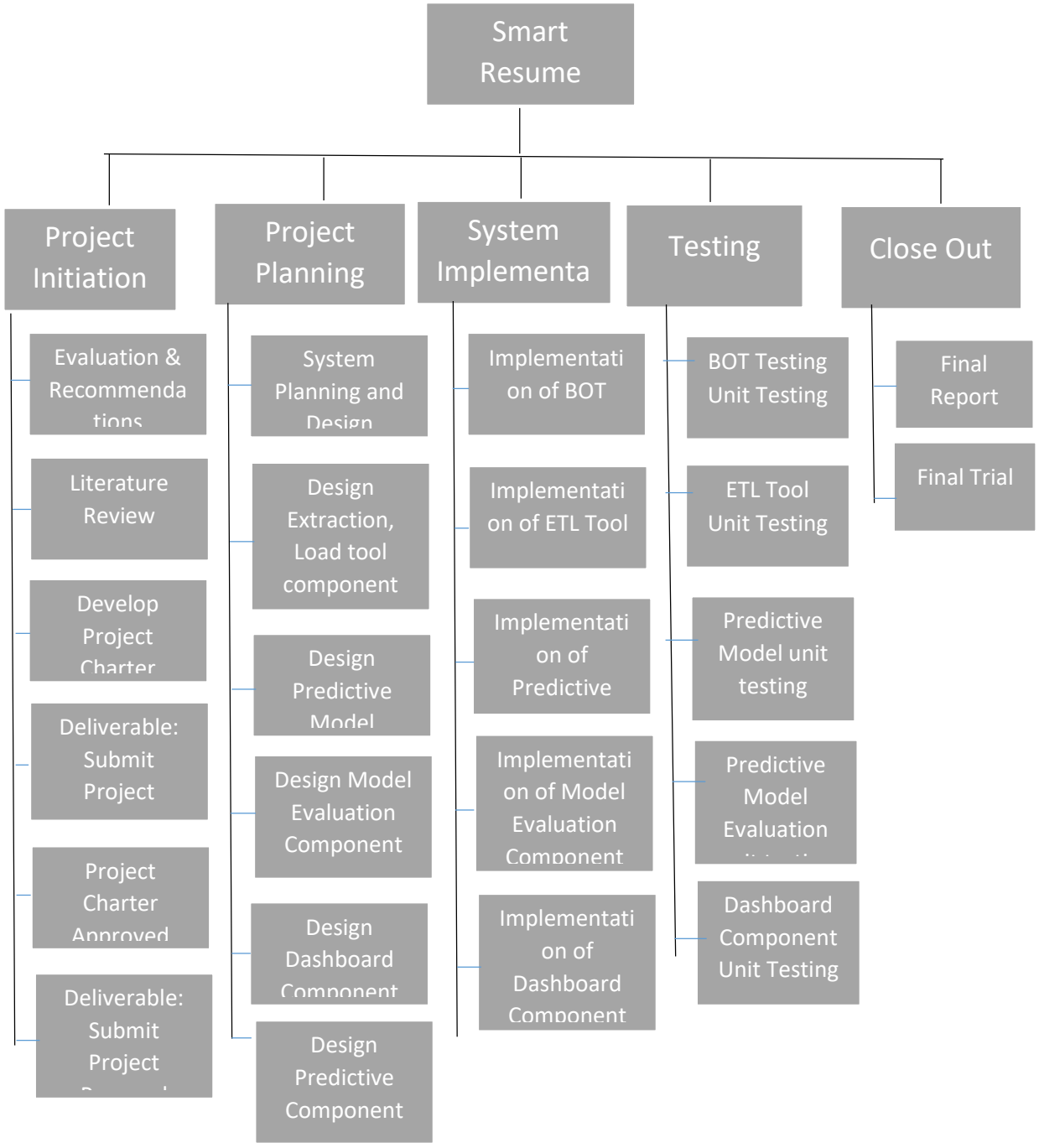

Figure 8 - Work Breakdown Structure

4.2 References

[1] Ephraim Turban, Decision Support and Business Intelligence Systems, 10th edition, 2014, pp.6-15.

[2] [Online]. Available: http://www.bitool.net/software/mysql.html

[3]C. Shrinivasan, "Data migration from a product to a data warehouse using ETL tool," *Proc. Eur. Conf. Softw. Maint. Reengineering, CSMR*, pp. 63–65, 2011.

[4]R. Mukherjee, "A Comparative Review Of Data Warehousing ETL Tools With New Trends And Industry Insight," *2017 IEEE 7th Int. Adv. Comput. Conf.*, pp. 944–949, 2017.

[5]A. Sabtu *et al.*, "The challenges of Extract, Transform and Loading (ETL) system implementation for near real-time environment," *Int. Conf. Res. Innov. Inf. Syst. ICRIIS*, pp. 3–7, 2017.

[6]P. Figueiras, R. Costa, G. Guerreiro, H. Antunes, A. Rosa, and R. Jardim-Goncalves, "User interface support for a big ETL data processing pipeline an application scenario on highway toll charging models," *2017 Int. Conf. Eng. Technol. Innov.*, pp. 1437–1444, 2017.