

TABLE OF CONTENT

Title	Page No
ABSTRACT.....	I
LIST OF FIGURES.....	V
LIST OF TABLES.....	Vi
ABBREVIATIONS.....	Vii
CHAPTER 1 INTRODUCTION.....	1
1.1 Project Summary.....	2
1.2 Project Purpose.....	2
1.3 Project Scope.....	2
1.4 Objectives.....	3
1.5 Technology and Literature Overview.....	3
1.6 Synopsis.....	6
CHAPTER 2 LITERATURE SURVEY.....	7
2.1 Introduction of Survey.....	8
2.2 Why Survey ?.....	8
CHAPTER 3 PROJECT MANAGEMENT.....	9
3.1 Project Planning Objectives.....	10
3.1.1 Software Scope.....	10
3.1.2 Resource.....	11
3.1.2.1 Human Resource.....	11
3.1.2.2 Reusable Software Resources.....	11
3.1.2.3 Environment Resource.....	11
3.1.3 Project Development Approach.....	11
3.1.3.1 Agile software development.....	12
3.1.3.2 Agile methods.....	12
3.1.3.3 Extreme programming.....	12
3.1.3.4 Scrum.....	12
3.1.3.5 Feature Driven Development.....	13
3.2 Project Scheduling.....	14
3.2.1 Basic Principle.....	15
3.2.2 Compartmentalization	15

3.2.3	Work breakdown structure.....	16
3.2.4	Project Organization.....	16
3.2.5	TimeLine Chart.....	17
3.2.5.1	Time Allocation.....	17
3.2.5.2	Task Sets.....	17
3.3	Risk Management.....	18
3.3.1	Risk Identification.....	19
3.3.1.1	Risk Identification artifacts.....	20
3.3.2	Risk Projection.....	20
CHAPTER 4 SYSTEM REQUIREMENTS.....		21
4.1	User Characteristics.....	22
4.2	Functional Requirement.....	22
4.3	Non Functional Requirement.....	23
4.4	Hardware and Software Requirement.....	24
4.4.1	Hardware Requirement.....	24
4.4.2	Software Requirement	24
CHAPTER 5 SYSTEM ANALYSIS.....		25
5.1	Study of Current System.....	26
5.2	Problems in Current System.....	26
5.3	Requirement of new System.....	26
5.3.1	Requirement Engineering process.....	26
5.3.2	Requirement Analysis process.....	26
5.4	Process Model.....	27
5.5	Feasibility Study.....	29
5.5.1	Technical Feasibility.....	29
5.5.2	Operational Feasibility.....	29
5.5.3	Economical Feasibility.....	29
5.5.4	Schedule Feasibility.....	29
CHAPTER 6 DETAIL DESCRIPTION.....		30
6.1	Admin Module.....	31
6.2	Customer Wizard Module.....	36

CHAPTER 7 Testing.....	39
7.1 Black-Box Testing.....	40
7.2 White-Box Testing	41
7.3 Test Cases	42
CHAPTER 8 SYSTEM DESIGN.....	47
8.1 ER Diagram.....	48
8.2 Use – Case Diagram.....	52
8.3 Activity Diagram.....	53
8.4 Data Flow Diagram.....	55
8.5 Sequence Diagram.....	58
8.6 Screenshots.....	60
CHAPTER 9 LIMITATION AND FUTURE ENHANCEMENT...	72
9.1 Limitation.....	73
9.2 Future Enhancement.....	73
CHAPTER 10 CONCLUSION.....	74
10.1 Conclusion.....	75
BIBLIOGRAPHY.....	76

WHAT COMPANY DO

➤ **Web Development**

- Web Designing Services
- E-Commerce Development.
- Web Application Development
- Content Management System

➤ **Mobile Application Development**

- Android App Development
- IOS App Development
- Windows App Development

➤ **Cloud Solutions**

➤ **Internet of Things (IoT)**

➤ **Bots & Cognitive Services**

➤ **Data Science**

➤ **Machine Learning**

➤ **Artificial Intelligence**

➤ **Deep Learning**

➤ **Smart Device Development**

➤ **Enterprise Mobility Solutions**

LIST OF FIGURES

Figure No.	Title	Page No.
Figure 3.1	Project Resource Pyramid	8
Figure 3.2	Prototyping Model	12
Figure 3.3	Work Breakdown Structure	14
Figure 3.4	Project Organization	14
Figure 5.3	Process Life Cycles	23
Figure 8.1	Use-Case diagram of Admin	41
Figure 8.2	Use-Case diagram of Customer	42
Figure 8.3	DFD Level-0	42
Figure 8.4	DFD Level-1	43
Figure 8.5	DFD Level-2	44
Figure 8.6	Sequence Diagram	45
Figure 8.7	Activity Diagram	46
Figure 8.8	Admin ER Diagram	47
Figure 8.9	Customer ER Diagram	48
Figure 8.10	Cleaning ER Diagram	49
Figure 8.11	Template ER Diagram	50
Figure 8.12	Screenshots	51

LIST OF TABLES

Table No.	Title	Page No.
Table 1.1	Synopsis	6
Table 3.1	Time Line Chart	16
Table 3.2	Risk Type	17
Table 3.3	Risk Category	18
Table 7.3	Test Cases	36

ABBREVIATION

Abbreviations used throughout this whole document are:

SRS	Software Requirements Specification
FDD	Feature Driven Development
DSDM	Dynamic System Development Method
XP	Extreme Programming
HOD	Head of the Department
HR	Human Resource
WBT	White Box Testing
WWW	World Wide Web
HTML	Hypertext Markup Language
CSS	Cascading Style Sheet
DBMS	Database management system
MySQL	My Structured Query Language
BI	Business Intelligence
ETL	Extraction Transformation Loading

CHAPTER 1

INTRODUCTION

- PROJECT SUMMARY
- PROJECT PURPOSE
- PROJECT SCOPE
- OBJECTIVES
- TECHNOLOGY AND
LITERATURE OVERVIEW
- SYNOPSIS

1.1 PROJECT SUMMARY

The following section provides an overview of the derived Software Requirements Specification (SRS) for the subject **Automated ETL & BI**. To begin with, the purpose of the document is presented and its intended audience outlined. Subsequently, the scope of the project specified by the document is given with a particular focus on what the resultant web-application will do and the relevant benefits associated with it. To conclude, a complete document overview is provided to facilitate increased reader comprehension and navigation.

1.2 PROJECT PURPOSE

The ability to analyze and act on data is increasingly important to businesses. The pace of change requires companies to be able to react quickly to changing demands from customers and environmental conditions. Effective business intelligence (BI) tools assist managers with decision making. Companies need cleaned and analyzed data and for that they hire data scientists who manually clean their data. For common datasets, instead of hiring a data scientist for doing the same job repeatedly, this product “Automated ETL & BI” actually does the same with minimal effort.

1.3 PROJECT SCOPE

Automated ETL & BI is a web application which focuses mainly on cleaning, analyzing and visualizing the customer datasets. The product provides following functionalities to end users:

- They can create their sub-users.
 - Their data will be cleaned automatically.
 - They can create their own datasets.
 - They can create their own data source.
 - They can view generated data source in form of graphs.
-

1.4 OBJECTIVES

In current scenario, for maintaining and cleaning various end user files, there is a need to process them manually, organize and create a dataset according to the needs of an end user, which is performed by a data scientist. Instead of performing the above mentioned manual process, Automated ETL provides a platform for completing the same automatically.

There are mainly two user roles for this system:

- Admin
- End User/Customer

An end user has to upload the files along with few details about the file being uploaded. There is also an option for end user to select the template according to which industry his dataset is based on. End user can manage and view his activities as well as his sub user's activities on activity dashboard.

Admin will manage various activities related to customer wizard such as: customer package management, template management, customer privileges, etc.

1.5 TECHNOLOGY AND LITERATURE OVERVIEW

1.5.1 Programming Languages

❖ Python

Python is an interpreted high-level programming language for general-purpose programming. Python has a design philosophy that emphasizes code readability and provides constructs that enable clear programming on both small and large scales. It supports multiple programming paradigms and has a large standard library.

❖ R

R is a programming language and free software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis.

Database**❖ MySQL**

MySQL is an open-source relational database management system. MySQL is written in C and C++. The MySQL development project has made its source code available under the terms of the GNU General Public License.

❖ MongoDB

MongoDB is a free and open-source cross platform document-oriented database program. MongoDB uses JSON-like documents with schemas. MongoDB is developed by MongoDB Inc. and it is classified as a NoSQL database program.

1.5.2 Software**❖ PyCharm**

PyCharm is an Integrated Development Environment (IDE) used in computer programming, specifically for the Python language. It is developed by the Czech company JetBrains.

❖ phpMyAdmin

phpMyAdmin is a free and open-source administration tool for MySQL and MariaDB. As a portable web application written primarily in PHP, it has become one of the most popular MySQL administration tools, especially for web hosting services. There are many features provided by phpMyAdmin including exporting and importing data.

❖ Jupyter

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

❖ Microsoft Office Word 2007

We use Microsoft Office Word 2007 to do our documentation of this final year project. We have used the feature of text box to draw the hierarchical chart to describe several subsystem, modules and sub-modules in the system. It also used to check spelling and grammar and to justify all the words. This tool is very useful for correct formation of document.

❖ Atlassian Bitbucket

Bitbucket is a web-based version control repository hosting service owned by Atlassian, for source code and development projects that use either Mercurial or Git revision control systems. Bitbucket offers both commercial plans and free accounts. Bitbucket integrates with other Atlassian software like Jira, HipChat, Confluence and Bamboo. It is similar to GitHub which primarily uses Git. Bitbucket has 3 deployment models: Cloud, Bitbucket Server and Data Center.

❖ WinSCP

WinSCP is a free and open-source SFTP, FTP, WebDAV, Amazon S3 and SCP client for Microsoft Windows. Its main function is secure file transfer between a local and a remote computer. Beyond this, WinSCP offers basic file manager and file synchronization functionality. For secure transfers, it uses Secure Shell (SSH) and supports the SCP protocol in addition to SFTP.

1.6 SYNOPSIS

Project Title	Automated ETL & BI
Project Description	Automated ETL & BI is a web application which aims to automate cleaning of data and represents information visually through graphs.
Objective	File upload, dataset and data-source generation
Purpose	Automated cleaning and Graph plotting
Anticipated Outcome	Web Application
Time Frame	4 months
Software Specification	PyCharm, phpMyAdmin 4.7.5, Mongodb 3.6, MySQL 5.7, Python 3.6.4, R 3.4.4, Django 2.0.1
Division of Responsibility	The entire project was divided among the 4 project developers.

Table 1.1 Synopsis

CHAPTER 2

LITERATURE SURVEY

- INTRODUCTION OF SURVEY
- WHY SURVEY?

2.1 INTRODUCTION OF SURVEY

A literature survey or a literature review in a project report is that section which shows the various analyses and research made in the field of the project and the results already published, taking into account the various parameters of the project and the extent of the project.

The survey of a software which already exists or similar product gives us idea of what our product should be like. Features that enhance future requirements are considered for further development.

2.2 WHY SURVEY?

A literature survey helps us to find the limitations of similar kind of projects and thus enables us to enhance our project by working on those limitations and providing the users a better experience.

Manual data cleaning is a very laborious task for data analysts. Often, there are chances of data remaining un-cleaned while using manual cleaning approach. The need for automatic data cleaning increases with the size of data files. The automated cleaning process prevents the need for any extra person to do the cleaning task and also the speed being faster, saves a lot of time.

It is much easier to understand information when it is represented graphically. The relation between two or more than two data sets can also be represented visually with softwares available in the market like Tableau and SAS. But, the availability of tools performing both the tasks of data cleaning and data visualization is very low. Our project will provide better experience to users working with bulky data by providing automated cleaning and by providing the feature to represent data source in various graphical formats.

CHAPTER 3

PROJECT MANAGEMENT

- PROJECT PLANNING
OBJECTIVES
- PROJECT SCHEDULING
- RISK MANAGEMENT

3.1 PROJECT PLANNING OBJECTIVES

The objective of software project planning is to provide a framework that enables the manager to make reasonable estimates, cost, and scheduling. These estimates are made within a limited time frame at the beginning of a software project and should be updated regularly as the project progresses.

During the project development period we have presented report to the internal guide on alternate Saturday for review and inspection.

There are many parameters in the software project planning as follows:

3.1.1 Software Scope

The project being a live project and simulation concept implementation being complex, the project planning is done in a way to get the maximum of resources for timely completion of the project.

Function and performance allocated to software during system engineering should be assessed to establish a project scope that is unambiguous and understandable at the management and technical levels.

3.1.2 Resource

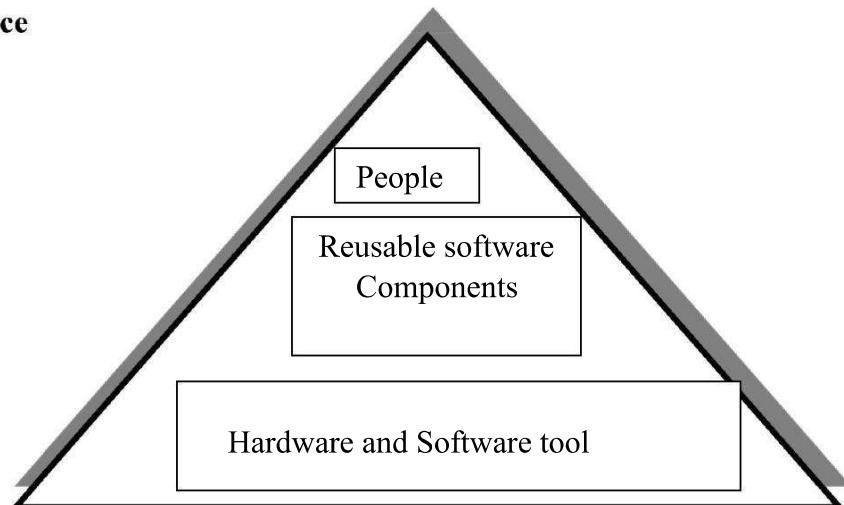


Figure 3.1: Project Resource Pyramid

The second software planning task is the estimation of the resources required to accomplish the software development effort. The above figure illustrates development resources as a pyramid. The development environment – hardware and software tools-sits at the foundation of the resources pyramid and provide the infrastructure to support the development effort. At a higher level, we encounter reusable software components-software building blocks that can dramatically reduce development costs and accelerate delivery. At the top of the pyramid the primary resource is people.

3.1.2.1 Human Resources

The human resources required are

1. Database, network administrator.
2. Project Guides.
3. Developers

3.1.2.2 Reusable Software Resource

Reusability is possible as and when require in this application. We can update it when new company register or company modified its data. Reusable software reduces design, coding and testing cost by amortizing effort over several designs .All other software components are building as per the need.

3.1.2.3 Environment Resource

The environment that supports the software project, often called the software engineering includes software and hardware.

3.1.3 Project Development Approach

As we came to know that traditional software development approaches are more mechanistic which concentrate more on processes, tools, contracts and plans. In contrast to traditional methods, agile methods keep emphasis on interaction, working software, embracing change at any moment of the project, customer relationships. The method can be agile if it is Incremental and Straightforward.

“Agile view is more people centric rather than plan-centric.” Agile methods are not defined by a small set of principles, practices and techniques. It creates a strategic capability which has capability of responding to change, capability to balance the structure and flexibility, capability of innovation and creations through development team and uncertainty.

Other different Agile Software development models such as

- XP (Extreme programming)
- Scrum
- FDD (Feature driven development)
- DSDM (Dynamic systems development method)

3.1.3.1 Agile Methods

Agile methods are designed to produce the first delivery in weeks, to achieve an early win and rapid feedback. These methods invent simple answers so that change can be less. These also improve design issues and quality as they are based on iteratively incremental method.

What makes a method an Agile? When the process are:

- Incremental: Small releases with rapid iterations
- Cooperative: Customer and developer relationships
- Straight: The method which is easy to learn and modify with documentation
- Adaptive: Able to embrace changes instantly

Different Agile Software Development methods:

- Extreme programming
- Scrum
- Feature driven development

3.1.3.2 Scrum

The term ‘SCRUM’ originally derives from a strategy in the game of rugby where it denotes “getting an out of play ball back into the game” with teamwork. Scrum concentrates on how the team members should function in order to produce the system flexibly in a constantly changing environment.

Scrum is extremely simple model used by different software companies from long time, which works with existing engineering practices and is scalable.

Scrum process includes three phases

- Pre-game
- Development
- Post-game

1. Pre-game phase includes two sub-phases

Planning and Architecture design.

Planning phase includes the development of the required system. A Backlog list is created, which contains all the requirements that are known at that moment.

Architecture phase: In this phase an abstract view of the model is designed by viewing Backlog list.

2. The Development phase:

This phase takes care of the different variable like time frame, quality, requirements, resources, technologies and tools. The system is developed in Sprints. Sprints are the iterative cycles where functionality is developed or enhanced to produce new increments.

Each Sprint includes the traditional phases of software engineering.

- Requirement
- Analysis
- Design
- Evolution and delivery

3. The Post-game phase close to release.

Roles in Scrum:

- SCRUM master
 - Product owner
 - Scrum Team
 - Customer
 - User
 - Management
-

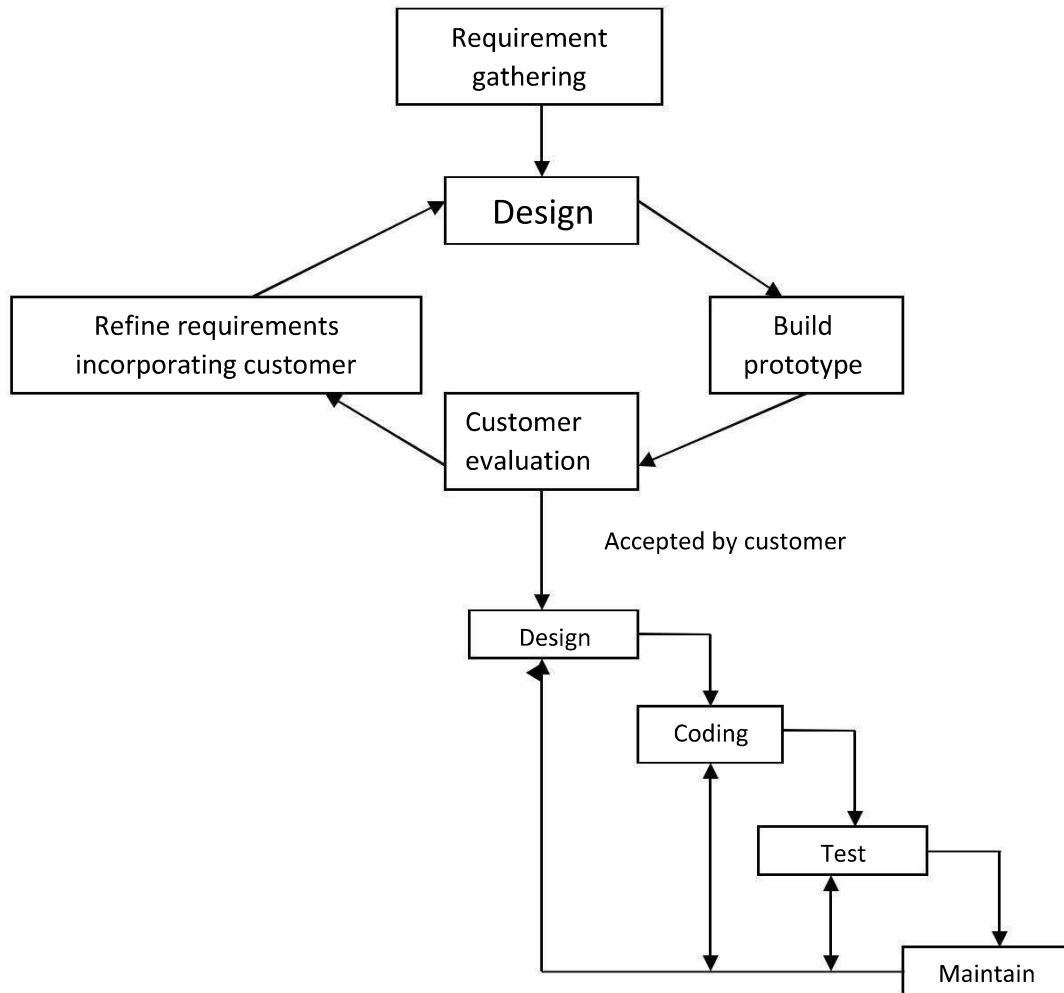


Figure 3.2: Prototyping Model

3.2 PROJECT SCHEDULING

Project scheduling is one of the main key aspects of any project. Any project must be scheduled before developing it.

When project developer works on scheduled project it is more advantageous for him/her to compare to unscheduled project. It gives us timeline for finishing the particular activity. Scheduling gives us idea about project length, its cost, its normal duration of completion and we can also find out the shortest way to complete the project with less overall cost of project.

Project schedule describes dependency between activities. The estimated time required to reach each milestones and allocation of people to activities.

3.2.1 Basic Principle

“Software project scheduling is an activity that distributes estimated effort across the planned project duration by allocating the effort to specific software engineering tasks.”

Proper scheduling requires:

- All tasks appear in network and dependent on some of other.
- Effort and timing are intelligently allocated to each task.
- Interdependencies between tasks are properly indicated.
- Resources are allocated for the work to be done.

3.2.2 Compartmentalization

Our software project is compartmentalized into the following tasks,

- Designing the GUI
- Query used for designing pages
- Coding
- Validations
- Testing

3.2.3 Work Breakdown Structure

It is used to decompose a given task and set recursively in small task.

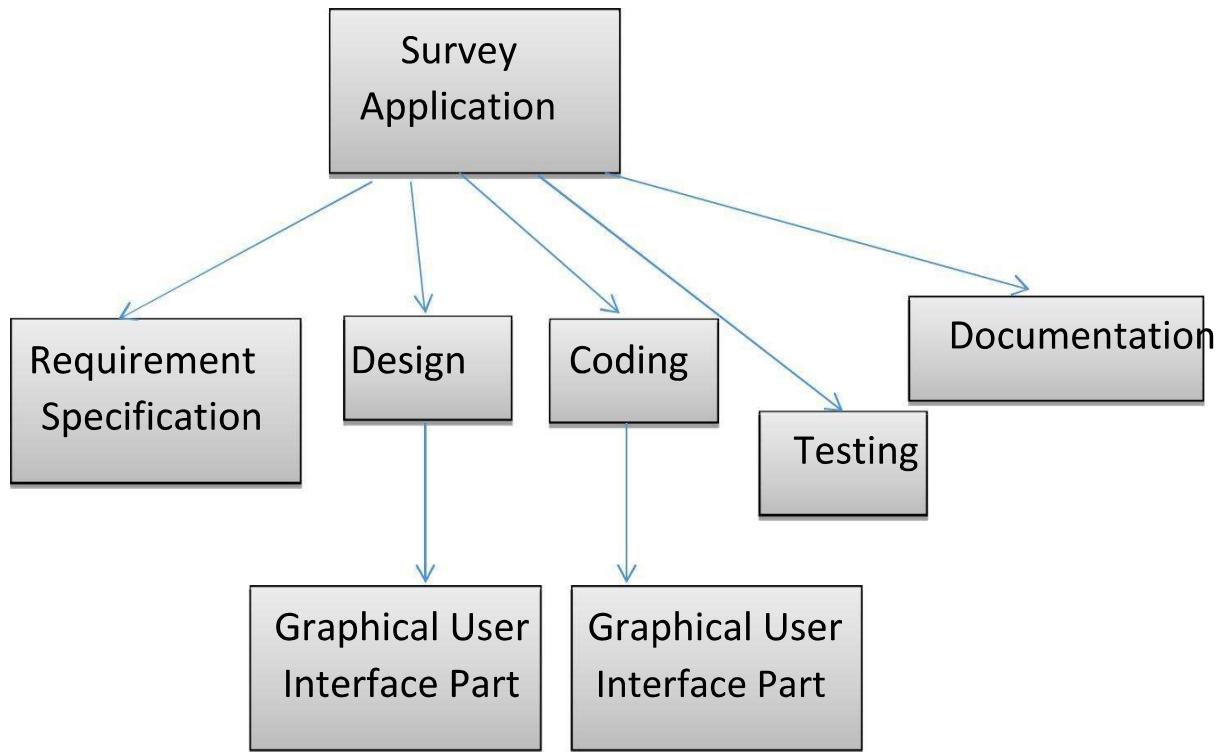


Figure 3.3: Work breakdown structure

3.2.4 Project Organization

This describes the way in which the development team is organized, the people involved and their roles in team. Here Project Organization Chart is shown.

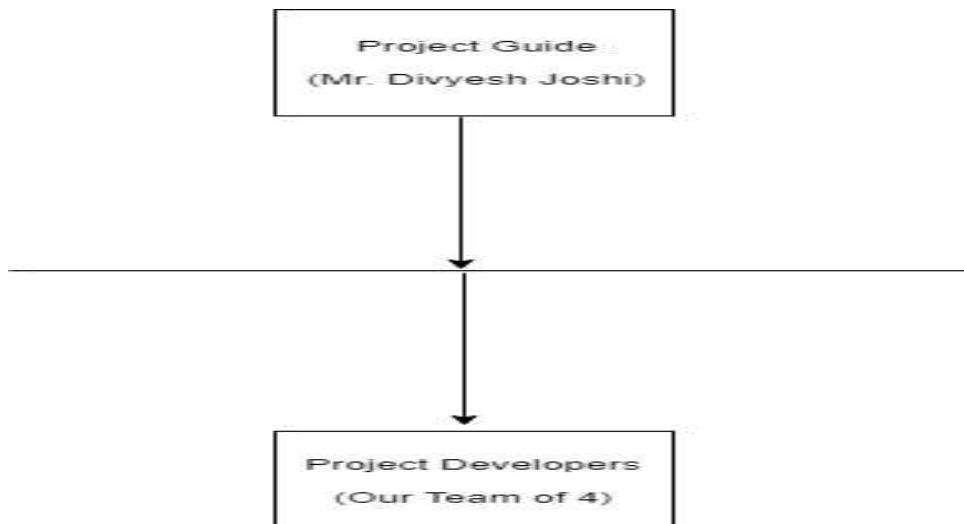


Figure 3.4: Project Organization

3.2.5 Timeline Chart

3.2.5.1 Task Sets

We have selected the “Prototype Process Model” so that there are six different work tasks to work together. A tasks set is a collection of software engineering work tasks, milestones, and deliverables that must be accomplished to complete a particular project. Tasks set are designed to accommodate different types of projects and different degree of rigour. Most software organizations encounter the following projects.

- Concept Development.
- New Application development.
- Application enhancement.
- Application maintenance.
- Re-engineering project.

Refinement of major tasks

The above tasks are again refined as follows:

Analysis of required system

1. Study of required flow
2. Study of methodology

Defining goals and objectives

1. Preparing the goal
2. Defining the flow of the project

Finding out resources required

1. Identifying the resources
2. Making arrangement for getting the resources

Coding and Testing.

1. Designing the tables and creating them.
2. Coding for the Forms.
3. Validation of the forms.
4. Testing

Documentation

1. Divide document in small parts
2. Documentation of each part
3. Integration of all parts
4. Review of project documentation.

Table 3.1: Time Line Chart

ID	Task	Start	Finish	Days
1	Study of overall system	2 nd January 2018	12 th January 2018	10
2	Requirement Analysis	12th January 2018	18 th January 2018	5
3	System Diagrams	12 th January 2018	22 nd January 2018	10
4	System Designing	20 th January 2018	20 th February 2018	30
5	Coding	20 th January 2018	20 th April 2018	3months
6	Testing	20 th April 2018	30 th April 2018	10
7	Documentation	22 nd April 2018	2 nd May 2018	10

3.3 RISK MANAGEMENT

Software is a difficult undertaking. Lots of things can go wrong and frankly, many often do. It's for this reason that being prepared understanding the risks and taking proactive measure to avoid or manage them is a key element of good software project management. Recognizing what can go wrong is the first step called 'Risk Identification'. Next each risk is analyzed to determine the likelihood that it will occur and the damage that it will do if it does occur. Once this information is established, risks are ranked, by probability and impact. Finally a plan is developed to manage those risks with high probability.

General Risks:

The general risks that can affect the development of software are as follows:

- **Lack of resources:** The resources which are needed for the development of this project are not available during project.
- **Time Duration:** We have limited time period so it takes well analyzed time chart to implement correctly and completely.
- **Lack of information:** Lack of information and knowledge can also consume much of time.

3.3.1 Risk Identification

Risk identification is a systematic attempt to specify threats to the project plan. By identifying known and predictable risks, we take first step towards avoiding them when possible and controlling them when necessary.

3.3.1.1 Risk Identification Artifacts

We considered the following types of risk to identify the risk in proper manners. The next table shows the type of risks.

Table 3.2 Risk Type

Risk Type	Description
Project Risks	<p>This type of risk can threaten the project plan. That is, if project risks become real, it is likely that project schedule and personal requirement problems and their impact on a software project.</p> <p>Example: Requirement Change, Specification delay</p>
Technical Risks	<p>This type of risk can threaten the quality and timeliness of software to be produced. If a technical risk becomes a reality then</p>

	<p>implementation may become impossible or difficult.</p> <p>Example: Technology, Hardware unavailability, Hardware failure</p> <p>Heap Space goes out of range</p>
Business Risks	<p>This type of risk can threaten the viability of the software to be built. Example: project size under estimation, lack of business funds</p>

3.3.2 Risk Projection

Risk projection, also called risk estimation, attempts to rate each risk in two ways—the likelihood or probability that risk is real and consequences of the problems associated with the risk should it occur. The following table shows the artifacts used in the risk projections.

Table 3.3 Risk Category

Risk Categories	Description
Catastrophic	Risk can cause the whole system to fail.
Critical	Significant degradation of the system may occur.
Minor	The risk can be easily recovered from the system failure.
Negligible	The risk can be negligible and shall not affect in the performance at all.

CHAPTER 4

SYSTEM REQUIREMENT

- USER CHARACTERISTICS
- FUNCTIONAL
REQUIREMENT
- NON FUNCTIONAL
REQUIREMENT
- HARDWARE AND
SOFTWARE REQUIREMENT

4.1 USER CHARACTERISTICS

The end-users of the Automatic ETL fall into three categories- Admin, Customer and Sub-Users.

➤ ADMIN

Admin will handle all the masters of the Admin panel. Only Admin will have the privilege to enter information in the master pages for customer.

➤ CUSTOMER

Customers will be able to select appropriate templates for their industries, select datasets, upload files for cleaning purpose, generate data source for graphs.

➤ SUB-USERS

Sub-users will have lesser privileges as compared to the Customers. They will be able to do the tasks similar to the tasks done by Customers but according to the privileges assigned to them.

4.2 FUNCTIONAL REQUIREMENTS

- Web-Application has to run simultaneously on various devices. It must be able to operate for long periods, without error.
- The server must be able to operate unattended indefinitely. It should not need physical interaction except for upgrades and failure of hardware elements.
- Backup and recovery should be handled by the DBMS and operating system, or external software running on a timed backup system.

- Representation of graph should be proper and well-understood.
- Data-cleaning script should be faultless.

4.3 NON- FUNCTIONAL REQUIREMENTS

Safety:

The system shall log every state and state change of every surface tablet and display to provision recovery from system failure. The system shall be capable of restoring itself to its previous state in the event of failure (e.g. a system crash or power loss).

Reliability:

Specifies the factors required to establish the required reliability of the software system at time of delivery.

Performance Requirements:

Performance requirements define acceptable response times for system functionality. The load time for user interface screens shall take no longer than five seconds. The log in information shall be verified within five seconds. Queries shall return results within five seconds.

Security:

The system provides the password security access control to avoid unauthorized user to login to the system and also Google recaptcha to avoid login from bots.

Consistency:

The application provides consistency to user interface design to the end-user. The designs of the screen are standardize and consistent that make the end-user feel comfortable to use it.

4.4 HARDWARE AND SOFTWARE REQUIREMENT

Software Requirement:

The software requires the support of the following softwares for the database and other requirements.

- Python
- Django
- Jupyter
- PyCharm
- phpMyAdmin
- R
- MySQL
- MongoDb
- Bitbucket
- CSS
- HTML
- Bootstrap
- JavaScript
- JQuery

Hardware Requirement:

- Laptop
- Server
- Internet Connection

CHAPTER 5

SYSTEM ANALYSIS

- STUDY OF CURRENT SYSTEM
- PROBLEMS IN CURRENT SYSTEM
- REQUIREMENT OF NEW SYSTEM
- PROCESS MODEL
- FEASIBILITY STUDY
- FEATURES OF NEW SYSTEM

5.1 STUDY OF CURRENT SYSTEM

- The current system which is “Automated ETL & BI” is used to clean the files and to represent information visually.
- This system will clean the files with automated cleaning script.
- This system will require users to select industry template from the given templates to map their files.
- The users will be able to generate data source by setting relationship between two or more than two datasets.

5.2 PROBLEMS OF CURRENT SYSTEM

- The problem will arise when the customer has to select any template other than the default templates available in our web-application.
- If the server crashes or has any problem, data will be lost.
- If there is connection problem while communicating, session will be lost and new session must be started.
- Any problem in connectivity can affect the communication between the modules.
- Novice users will need some time to understand the working of the whole system.

5.3 PROCESS MODEL

- The basic idea behind iterative enhancement is to develop a software system incrementally, allowing the developer to take advantage of what was being learned during the development of earlier, incremental, deliverable versions of the system. Learning comes from both the development and use of the system, where possible. Key steps in the process were to start with a simple implementation of a subset of the software requirements and iteratively enhance the evolving sequence of versions until the full system is implemented. At each iteration, design modifications are made and new functional capabilities are added.

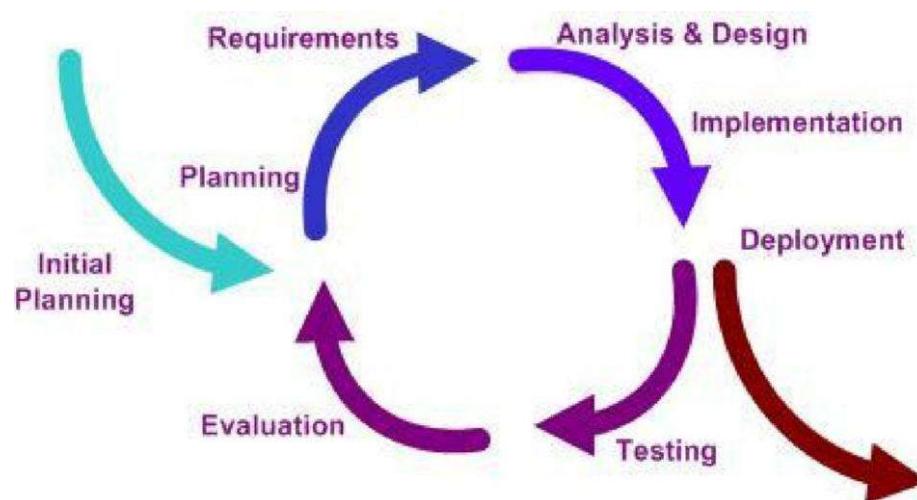


Figure 5.3 Process Life Cycles

- The Procedure itself consists of the Initialization step, the Iteration step, and the Project Control List. The initialization step creates a base version of the system. The goal for this initial implementation is to create a product to which the user can react. It should offer a sampling of the key aspects of the problem and provide a solution that is simple enough to understand and implement easily. To guide the iteration process, a project control list is created that contains a record of all tasks that need to be performed. It includes such items as new features to be implemented and areas of redesign of the existing solution. The control list is constantly being revised as a result of the analysis phase.

- The iteration involves the redesign and implementation of a task from project control list, and the analysis of the current version of the system. The goal for the design and implementation of any iteration is to be simple, straightforward, and modular, supporting redesign at that stage or as a task added to the project control list. The code can, in some cases, represent the major source of documentation of the system. The analysis of iteration is based upon user feedback, and the program analysis facilities available. It involves analysis of the structure, modularity, usability, reliability, efficiency, and achievement of goals. The project control list is modified in light of the analysis results.

Guidelines that drive the implementation and analysis include:

- Any difficulty in design, coding and testing a modification should signal the need for redesign or re-coding.
- Modifications should fit easily into isolated and easy-to-find modules. If they do not, some redesign is needed.
- Modifications to tables should be especially easy to make. If any table modification is not quickly and easily done, redesign is indicated.
- Patches should normally be allowed to exist for only one or two iterations. Patches may be necessary to avoid redesigning during an implementation phase.
- The existing implementation should be analyzed frequently to determine how well it measures up to project goals.
- Program analysis facilities should be used whenever available to aid in the analysis of partial implementations.
- User reaction should be solicited and analyzed for indications of deficiencies in the current implementation.

5.4 FEASIBILITY STUDY

Objective of Feasibility Study

An important outcome of the preliminary investigation is the determination that the system requested is feasible. The feasibility study is carried out to examine the likelihood that the system will be useful to the organization. There are four aspects in the feasibility study namely.

- Operational Feasibility
- Technical Feasibility
- Economic Feasibility
- Schedule Feasibility

5.4.1 Technical Feasibility

The main purpose of checking Technical Feasibility is to examine whether the current technology is sufficient for the development of the system. The outcomes of the technical feasibility are as follows:

5.4.2 Operational Feasibility

The main purpose of checking Operational Feasibility is to find out whether the system will be functional after its development and installation or not. The outcomes of the operational feasibility are as follows:

5.4.3 Economical Feasibility

The main purpose of checking Economical Feasibility is to examine whether the financial investment in the system will meet the organization's requirements or not.

5.4.4 Schedule Feasibility

This type of the feasibility includes a measure of how reasonable the project is with respect to time aspect. When developing software it is difficult to measure such things as software complexity, quality and to estimate the amount of effort it will take to complete the project.

CHAPTER 6

DETAIL DESCRIPTION

- ADMIN PANEL MODULE
- CUSTOMER WIZARD

6.1 ADMIN PANEL MODULE

Admin Panel Module contains 10 masters which will contain information about the company of a particular customer. Customers will be able to buy a package for their company according to their requirements and they will be given privileges according to their selection. Admin will enter the information in all the masters according to the requirements provided by the Customers.

Functional Requirements

➤ R1: Login

State: Login page will be displayed. Only Admin can login.

Input: Username and password are entered.

Output: If username and password are correct, it will be redirected to Admin Panel. If it is incorrect, an error message will be displayed on the login page.

➤ R2: Admin Panel

Description: Admin Panel contains all the masters such as Department Management, File Grouping management, Data Source management, Package management, Country management, State management, City management etc. which are handled by Admin.

➤ R3: Master Pages

Description: These pages are used to list, edit, delete or add new data to the database by admin. Search functionality is also provided for each page to increase simplicity for finding an existing record. Master pages include File grouping master, Data Source master, Template master, Employee strength master, Industry master etc.

← R3.1: Department management Master

State: List of all departments will be displayed along with their name, description and status.

Input: By clicking on add department button, new department information can be added to database. By clicking on edit icon in action column, admin can edit the stored information about the selected department and update the changes by clicking on submit button. By clicking on delete icon, admin can delete selected department information from database.

Output: This new department will be added to database and will be displayed in the list on clicking submit button.

← R3.2: File Grouping management Master

State: List of all file groups will be displayed along with their name, description and status.

Input: By clicking on add file group button, new file group information can be added to database. By clicking on edit icon in action column, admin can edit the stored information about the selected file group and update the changes by clicking on submit button. By clicking on delete icon, admin can delete selected file group information from database.

Output: This new file group will be added to database and will be displayed in the list on clicking submit button.

← R3.3: Feature Master management

State: List of all features will be displayed along with their name, description and status.

Input: By clicking on add feature button, new feature information can be added to database. By clicking on edit icon in action column, admin can edit the stored information about the selected feature and update the changes by clicking on submit button. By clicking on delete icon, admin can delete selected feature information from database.

Output: This new feature will be added to database and will be displayed in the list on clicking submit button.

← **R3.4: Role management Master**

State: List of all roles will be displayed along with their name, description and status.

Input: By clicking on add role button, new role information can be added to database. By clicking on edit icon in action column, admin can edit the stored information about the selected role and update the changes by clicking on submit button. By clicking on delete icon, admin can delete selected role information from database.

Output: This new role will be added to database and will be displayed in the list on clicking submit button.

← **R3.4: Employee Strength Master**

State: List of all strengths will be displayed along with their status.

Input: By clicking on add employee strength button, new employee strength can be added to database. By clicking on edit icon in action column, admin can edit the stored information about the selected strength and update the changes by clicking on submit button. By clicking on delete icon, admin can delete selected strength from database.

Output: This new employee strength will be added to database and will be displayed in the list on clicking submit button.

← **R3.5: Package management Master**

State: List of all packages will be displayed along with their name, price, duration and status.

Input: By clicking on add package button, new package information can be added to database. By clicking on edit icon in action column, admin can edit the stored information about the selected package and update the changes by clicking on submit button. By clicking on delete icon, admin can delete selected package information from database.

Output: This new package will be added to database and will be displayed in the list on clicking submit button.

← **R3.6: Industry type management Master**

State: List of all industry will be displayed along with their name, description and status.

Input: By clicking on add industry type button, new industry information can be added to database. By clicking on edit icon in action column, admin can edit the stored information about the selected industry and update the changes by clicking on submit button. By clicking on delete icon, admin can delete selected industry information from database.

Output: This new industry will be added to database and will be displayed in the list on clicking submit button.

← **R3.7: Data Source management Master**

State: List of all file groups will be displayed along with the file names, file format and status.

Input: By clicking on add data source button, new data source information can be added to database. By clicking on edit icon in action column, admin can edit the stored information about the selected data source and update the changes by clicking on submit button. By clicking on delete icon, admin can delete selected data source information from database.

Output: This new data source will be added to database and will be displayed in the list on clicking submit button.

← **R3.8: Country management Master**

State: List of all countries will be displayed along with their status.

← **R3.9: State management Master**

State: List of all states will be displayed along with their status.

← **R3.10: City management Master**

State: List of all cities will be displayed along with their status.

← **R3.11: Template management Master**

State: List of all template names will be displayed along with the industry name and status.

Input: By clicking on add template button, new template information can be added to database. By clicking on edit icon in action column, admin can edit the stored information about the selected template and update the changes by clicking on submit button. By clicking on delete icon, admin can delete selected template information from database.

Output: This new template will be added to database and will be displayed in the list on clicking submit button.

6.2 CUSTOMER WIZARD MODULE

Customer Wizard allows customers and sub-users to select template, select dataset type and upload file for cleaning purpose. Customers will be able to set relationships between two or more than two datasets and generate data source. The data source generated will be used to plot graphs and various operations like sum, average etc. can be performed on these graphs.

← **R4.1: Login**

Status: Login page will be displayed. Only Customers can login here.

Input: Users will be prompted to enter their username and password. Their credentials will be validated and if and only if their role is customer then they will be able to successfully login.

Output: After successful login, the user will be redirected to either dashboard or wizard.

← **R4.2: Select Dataset or Template**

Customer will be able to select either Dataset or Template in the first page of wizard. Dataset option will allow user to select one dataset type from multiple dataset types. Template option will allow user to select one template from multiple templates of different industries.

← **R4.3: Select Dataset type**

Customer will be able to select one dataset type from multiple dataset types such as Excel, CSV, Text etc. available in the package master.

← **R4.4: Give Dataset Name**

Customer will be required to give appropriate Dataset name where all the uploaded files will get stored.

← **R4.5: Upload files**

Customer will be able to upload multiple files from their system on which cleaning task will be performed. The extension of the files uploaded should match the dataset type selected otherwise an error message will be displayed. The files on successful upload will get stored in a folder with name same as the dataset name.

← **R4.6: Data Mapping**

Customer will now map the fields of the file with the appropriate type and proper data format will be required to be selected. Customer has to select whether the field is required or not and whether the data should be unique or non-unique. Customer can also give optional description about the field.

← **R4.7: Select Template**

Customer has now to select proper template from multiple templates on basis of required industry. E.g. If the industry of the customer is pharmaceutical, then selection of pharmaceutical template must be done.

← **R4.8: Template Column Mapping**

Customer has to map file columns with template columns in this step. The mapping will be with the columns of the template selected in the previous step and the file selected by the customer.

← **R4.9: Dashboard**

Dashboard will display the status of the cleaning process of all the files uploaded. The status can be completed, pending, in process or completed with errors. The error-file will display what errors were found during the cleaning process.

← R5.0: Data-source

Data-source is a set of relations between two or more than two datasets. User will be able to create relations between datasets by selecting fields in each dataset and through various join-queries, data-source can be generated.

← R5.1: KPI Dashboard

KPI dashboard will generate graphs on the basis of the data- source. Various operations like min, max, sum, count, average can be performed on the data-source.

CHAPTER 7

TESTING

- BLACK-BOX TESTING
- WHITE-BOX TESTING
- TEST CASES

7.1 BLACK-BOX TESTING

- Black box testing treats the system as a '**black-box**', so it doesn't explicitly use Knowledge of the internal structure or code. Or in other words the Test engineer need not know the internal working of the "Black box" or application.
- Main focus in black box testing is on functionality of the system as a whole. The term '**behavioral testing**' is also used for black box testing and white box testing is also sometimes called '**structural testing**'. Behavioral test design is slightly different from black-box test design because the use of internal knowledge isn't strictly forbidden, but it's still discouraged.
- Each testing method has its own advantages and disadvantages. There are some bugs that cannot be found using only black box or only white box. Majority of the application are tested by black box testing method. We need to cover majority of test cases so that most of the bugs will get discovered by black box testing.
- Black box testing occurs throughout the software development and testing life cycle i.e. in Unit, Integration, System, Acceptance and regression testing stages.

Advantages of Black Box Testing

- Tester can be non-technical.
- Used to verify contradictions in actual system and the specifications.
- Test cases can be designed as soon as the functional specifications are complete

Disadvantages of Black Box Testing

- The test inputs needs to be from large sample space.
-

- It is difficult to identify all possible inputs in limited testing time. So writing test cases is slow and difficult chances of having unidentified paths during this testing.

7.2 WHITE-BOX TESTING

White box testing (WBT) is also called **Structural or Glass box testing**. White box testing involves looking at the structure of the code. When you know the internal structure of a product, tests can be conducted to ensure that the internal operations performed according to the specification. And all internal components have been adequately exercised.

Why we do White Box Testing?

To ensure:

- That all independent paths within a module have been exercised at least once.
- All logical decisions verified on their true and false values.
- All loops executed at their boundaries and within their operational bounds internal data structures validity.

Limitations of White-Box Testing:

- Not possible for testing each and every path of the loops in program. This means exhaustive testing is impossible for large systems.
- This does not mean that WBT is not effective. By selecting important logical paths and data structure for testing is practically possible and effective.

7.3 TEST CASES

Login:

Table 7.3.1 Login test case

Sr. No.	Test Case	Expected Output	Actual Output	Test Case Status
1.	Enter correct Username, Password and User Type	Redirect to dashboard/Wizard	Redirect to Dashboard/Wizard	Pass
2.	Enter incorrect Username, Password and User Type	Display Error Message	Display Error Message	Pass

Add new data in Master Pages:

Table 7.3.2 Add new data in Master pages test case

Sr. No.	Test Case	Expected Output	Actual Output	Test Case Status
1.	Check validation of add master details Page	Show error message for validation.	Show error message for validation.	Pass
2.	Enter correct detail for a add master details page	Add new data to database and successful message	Add new data to database and successful message	Pass

Edit data in Master Pages:**Table 7.3.3 Edit data in Master pages test case**

Sr. No.	Test Case	Expected Output	Actual Output	Test Case Status
1.	Check Validation of edit master details page	Show error message for validation	Show error message for validation	Pass
2.	Enter correct detail for edit master details page	Update data to database and display successful message	Update data to database and display successful message	Pass

Delete data in Master Pages:**Table 7.3.4 Delete data in Master pages test case**

Sr. NO.	Test Case	Expected Output	Actual Output	Test Case Status
1.	Click on delete icon	Show successful message and delete data from database.	Show successful message and delete data from database.	Pass

Upload files in Customer wizard module:**Table 7.3.5 Upload files in Customer wizard module test case**

Sr. NO.	Test Case	Expected Output	Actual Output	Test Case Status
1.	Upload proper file matching selected dataset type and file size not exceeding the limit	File Uploaded successfully	File Uploaded successfully	Pass
2.	Upload file not matching dataset type or file size is exceeding	Display Error Message	Display Error Message	Pass

Data Column Mapping:**Table 7.3.6 Data Column Mapping test case**

Sr. No.	Test Case	Expected Output	Actual Output	Test Case Status
1.	Map user-defined data type to system data type	Redirect to template selection	Redirect to template selection	Pass

Template Selection:**Table 7.3.7 Template Selection test case**

Sr. NO.	Test Case	Expected Output	Actual Output	Test Case Status
1.	Select One Template	Redirect to template column mapping	Redirect to template column mapping	Pass

Template Column Mapping:**Table 7.3.8 Template column mapping test case**

Sr. No.	Test Case	Expected Output	Actual Output	Test Case Status
1.	Map file columns with template columns	Redirect to Dashboard	Redirect to Dashboard	Pass

File Cleaning:**Table 7.3.9 File cleaning test case**

Sr. No.	Test Case	Expected Output	Actual Output	Test Case Status
1	File cleaning completed	File status must be Completed	File status must be Completed	Pass
2	File cleaning completed with errors	File status must be Completed with Errors	File status must be Completed with Errors	Pass

Data Source:**Table 7.3.10 Data Source test case**

Sr. No.	Test Case	Expected Output	Actual Output	Test Case Status
1	Proper Query fired	Data Source Generation	Data Source Generation	Pass
2	Improper Query fired	Invalid Query	Invalid Query	Pass

CHAPTER 8

SYSTEM DESIGN

- USE-CASE DIAGRAM
- DATA FLOW
DIAGRAM
- SEQUENCE DIAGRAM
- ACTIVITY DIAGRAM
- ER DIAGRAM
- SCREENSHOTS

8.1 USE-CASE DIAGRAM

ADMIN

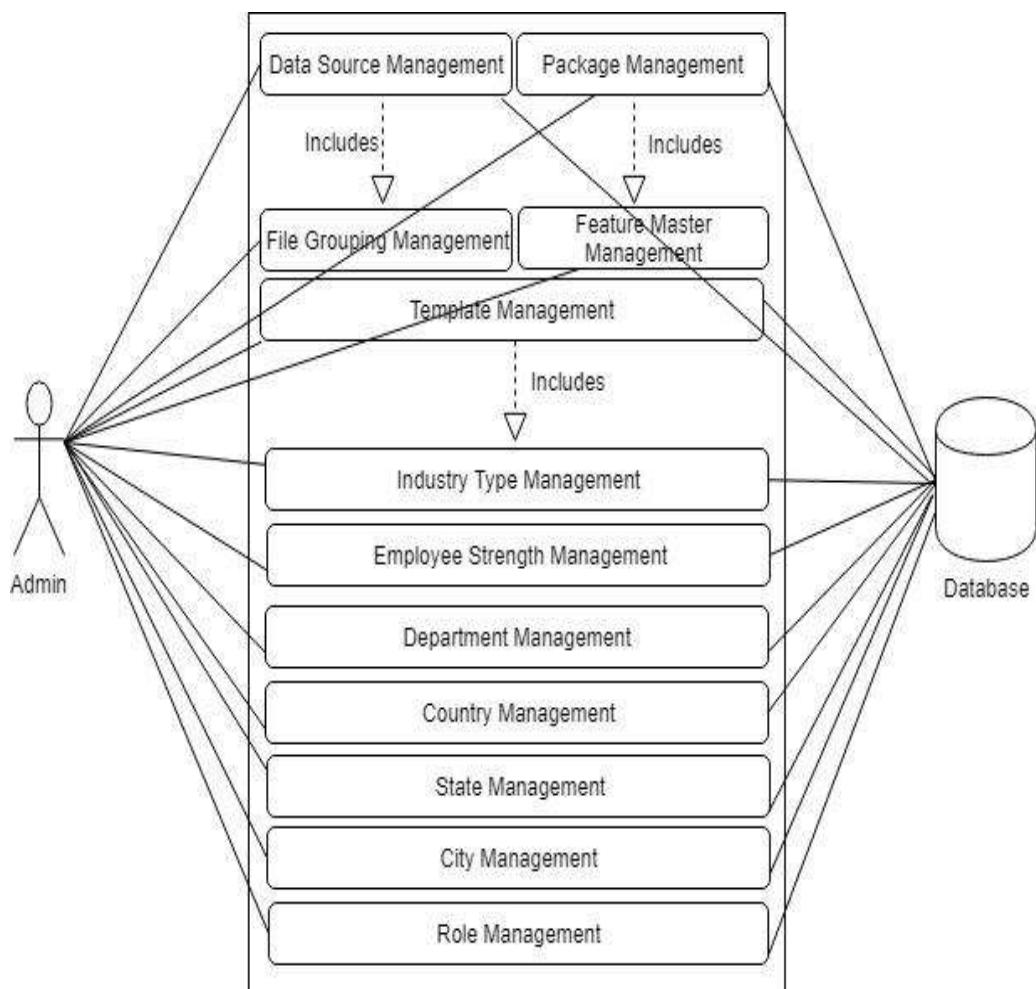


Figure 8.1: Use-Case diagram of Admin

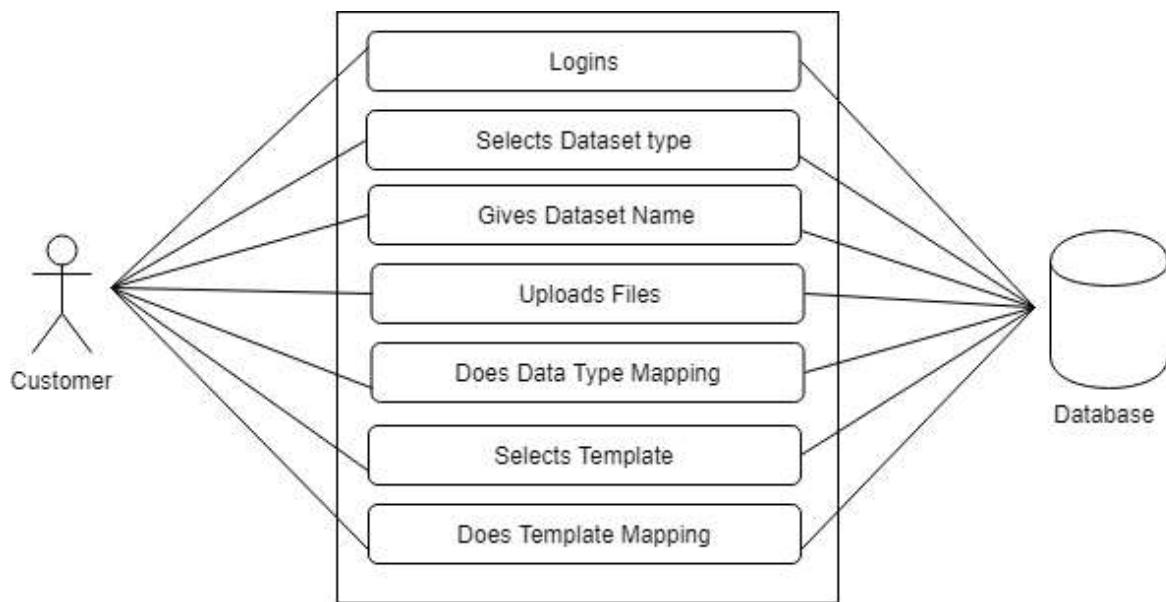
CUSTOMER

Figure 8.2: Use-case diagram of Customer

8.2 DATA FLOW DIAGRAM

DFD Level-0

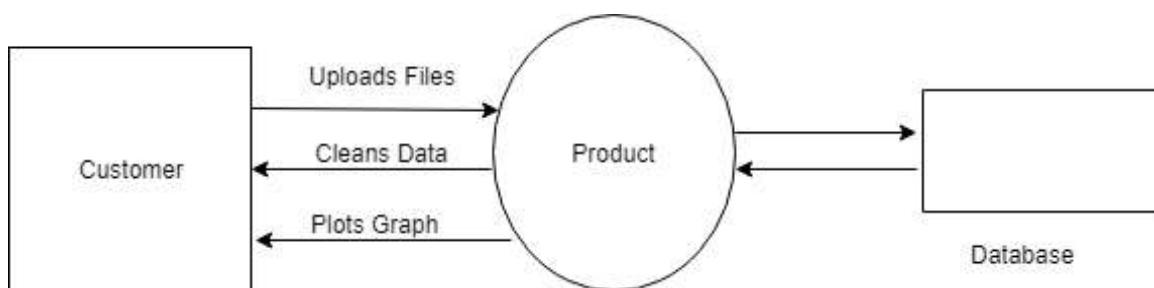


Figure 8.3: DFD Level-0

DFD Level-1

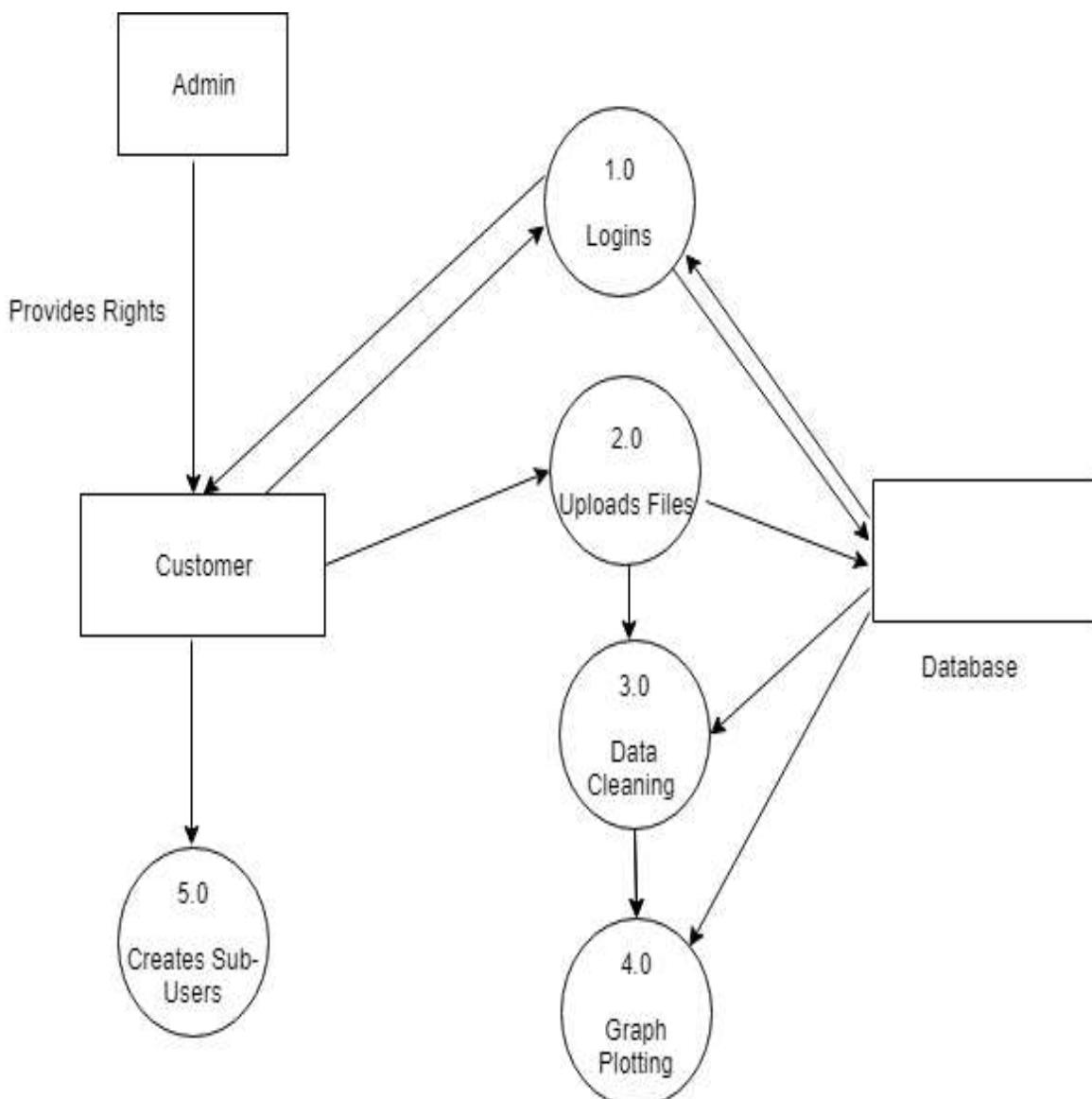


Figure 8.4: DFD Level-1

DFD Level-2

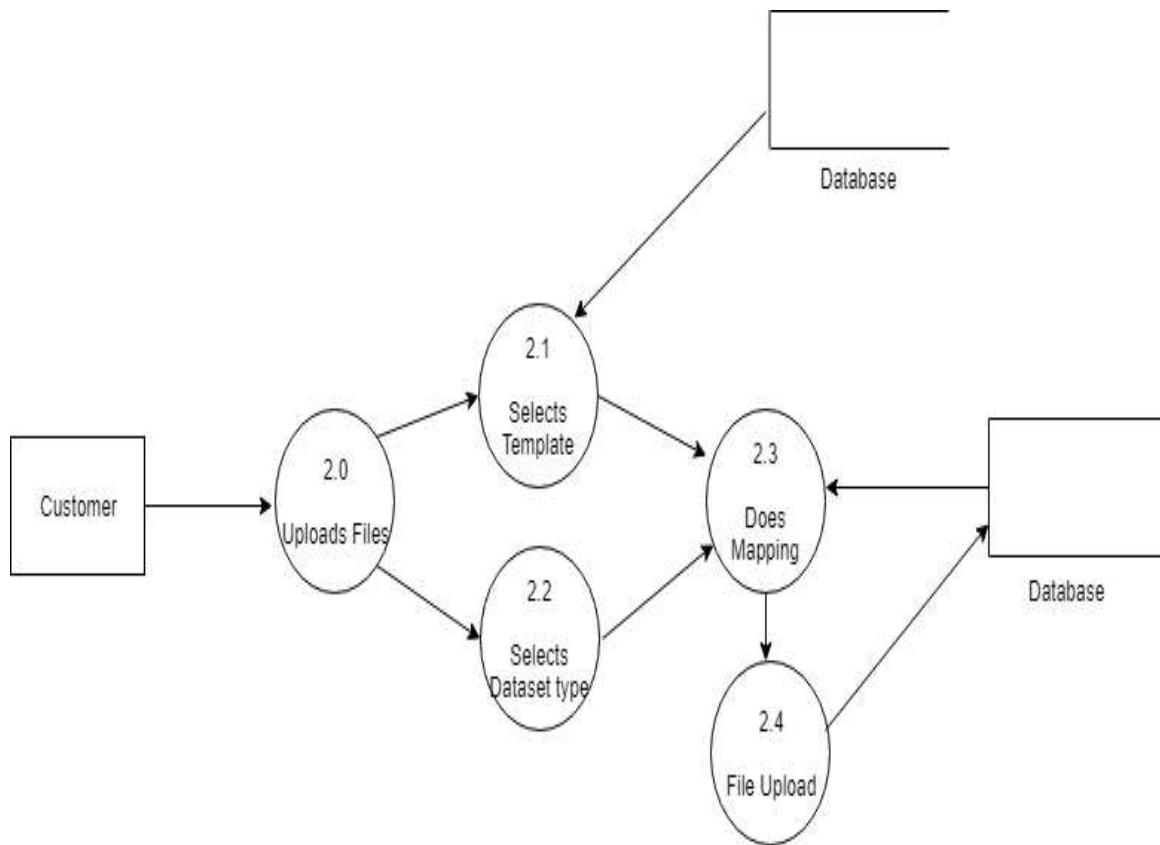


Figure 8.5: DFD Level-2

8.3 SEQUENCE DIAGRAM

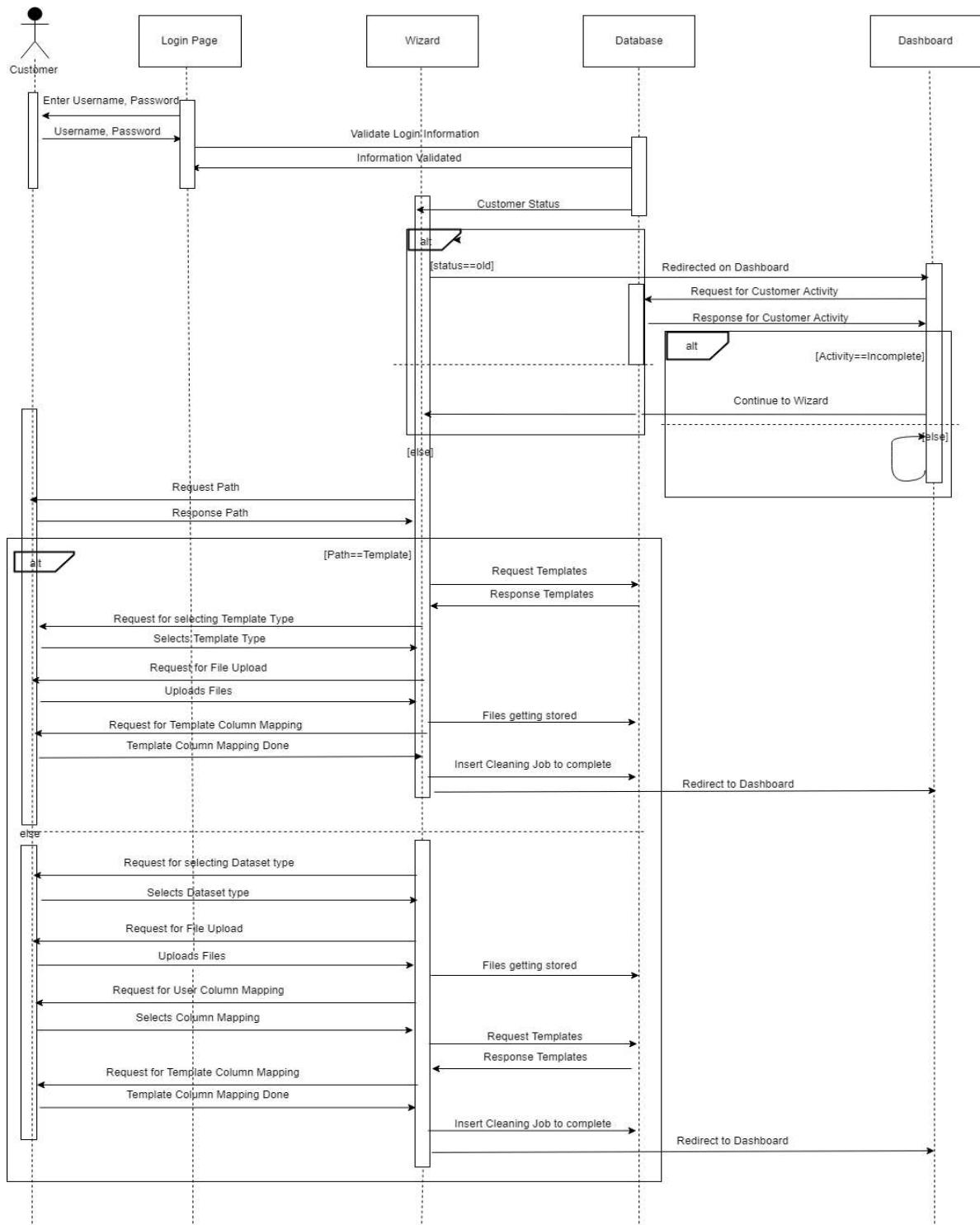


Figure 8.6: Sequence Diagram

8.4 ACTIVITY DIAGRAM

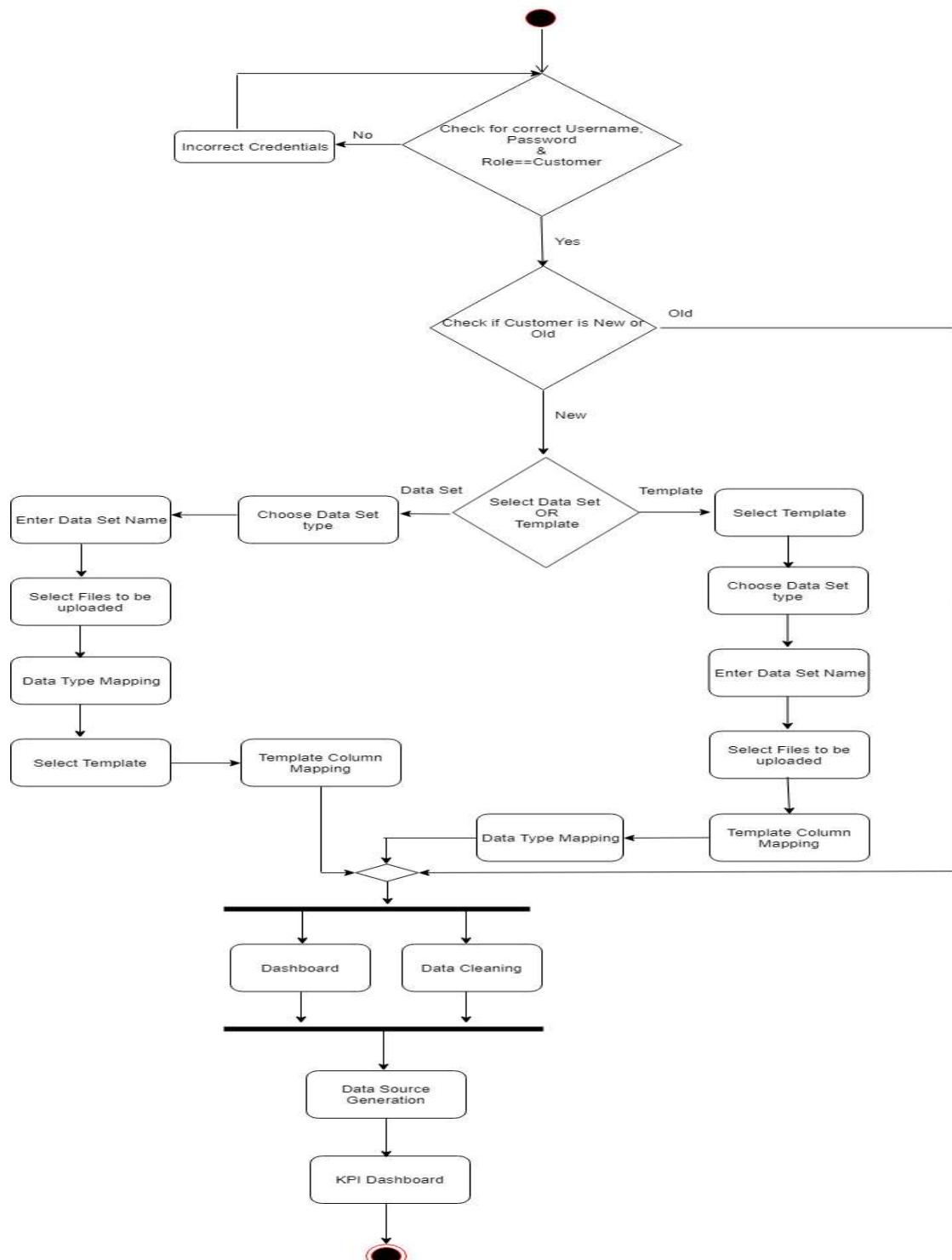


Figure 8.7: Activity Diagram

8.5 ER DIAGRAM

ADMIN ER DIAGRAM

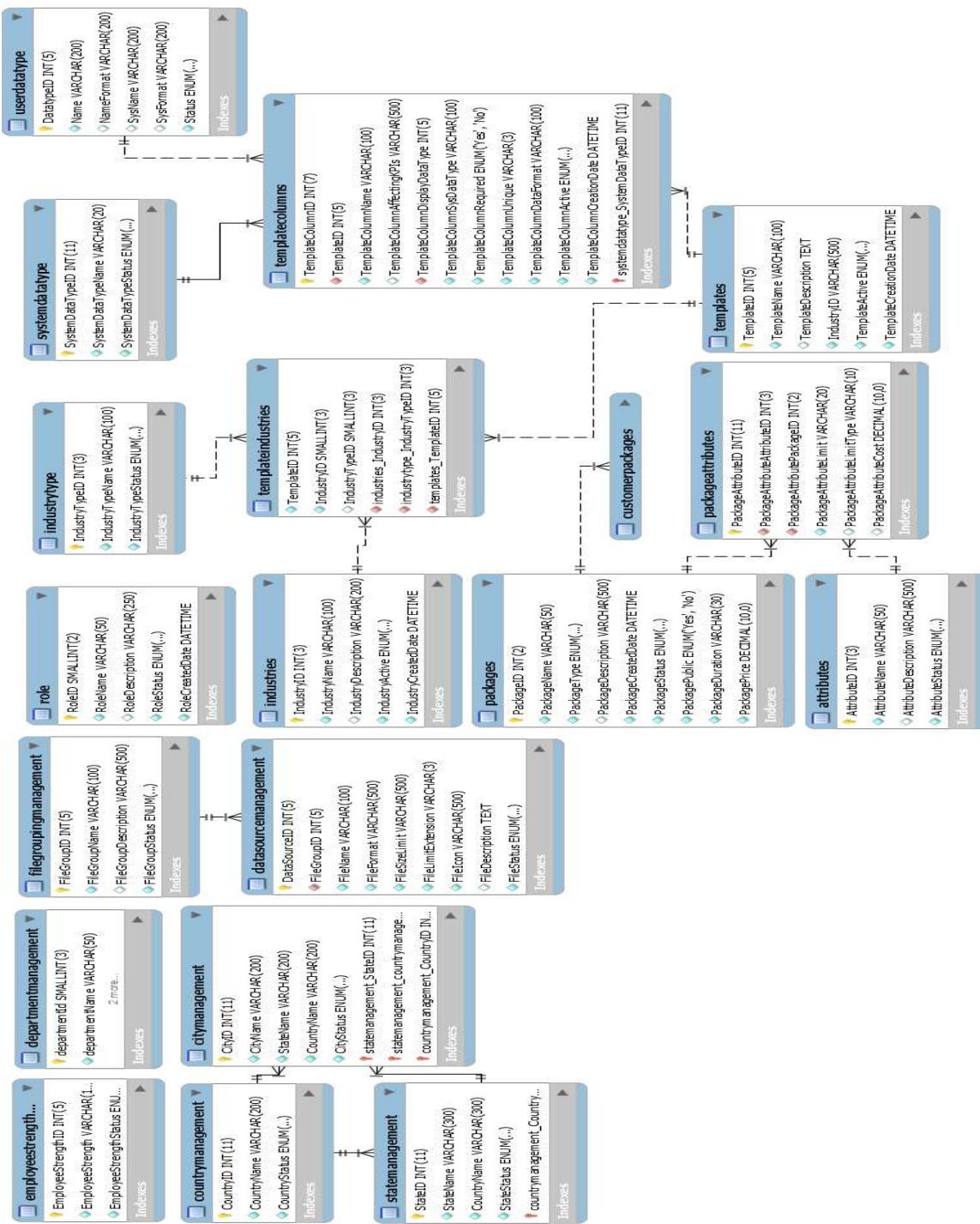


Figure 8.8: Admin ER Diagram

CUSTOMER ER DIAGRAM

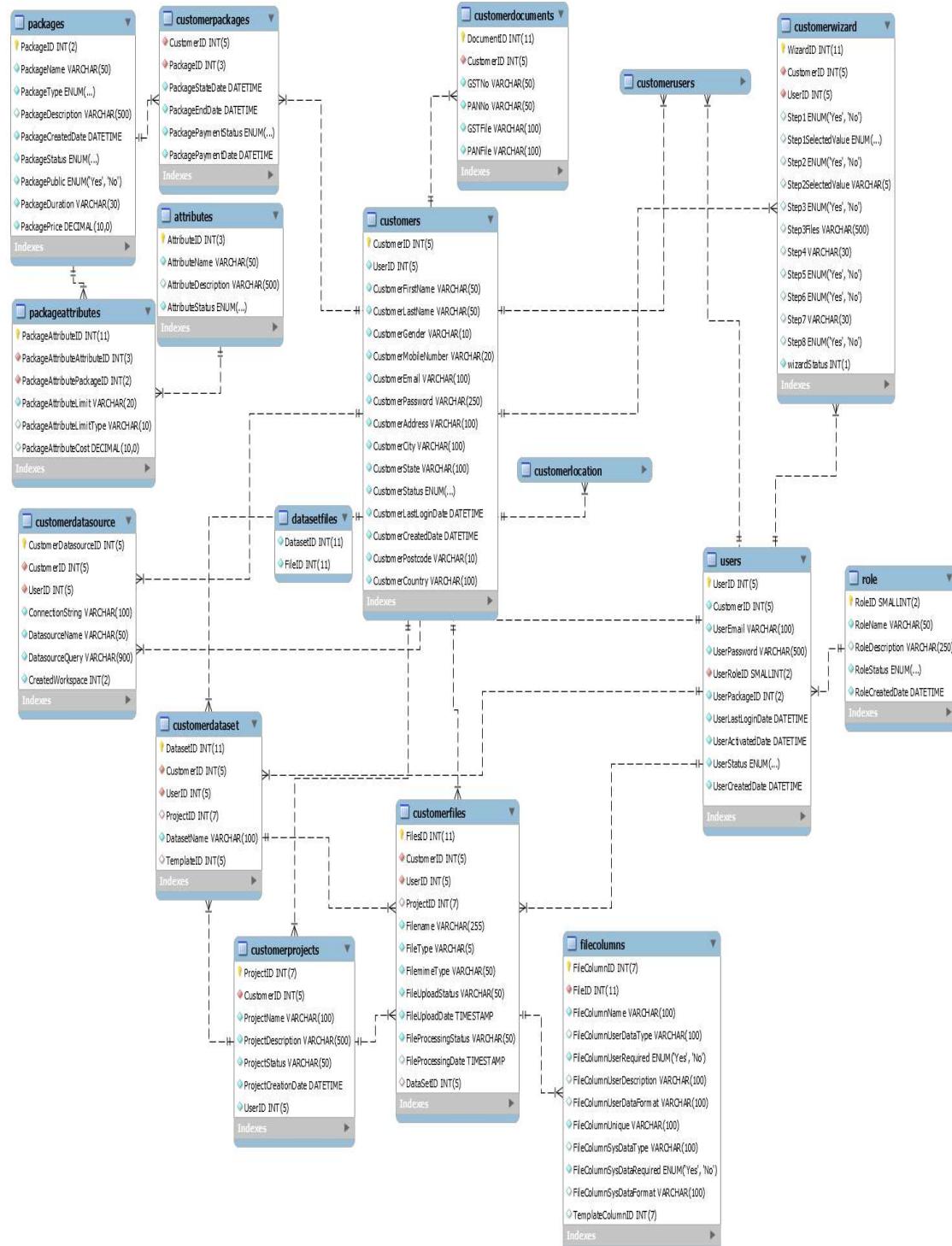


Figure 8.9: Customer ER Diagram

CLEANING ER DIAGRAM

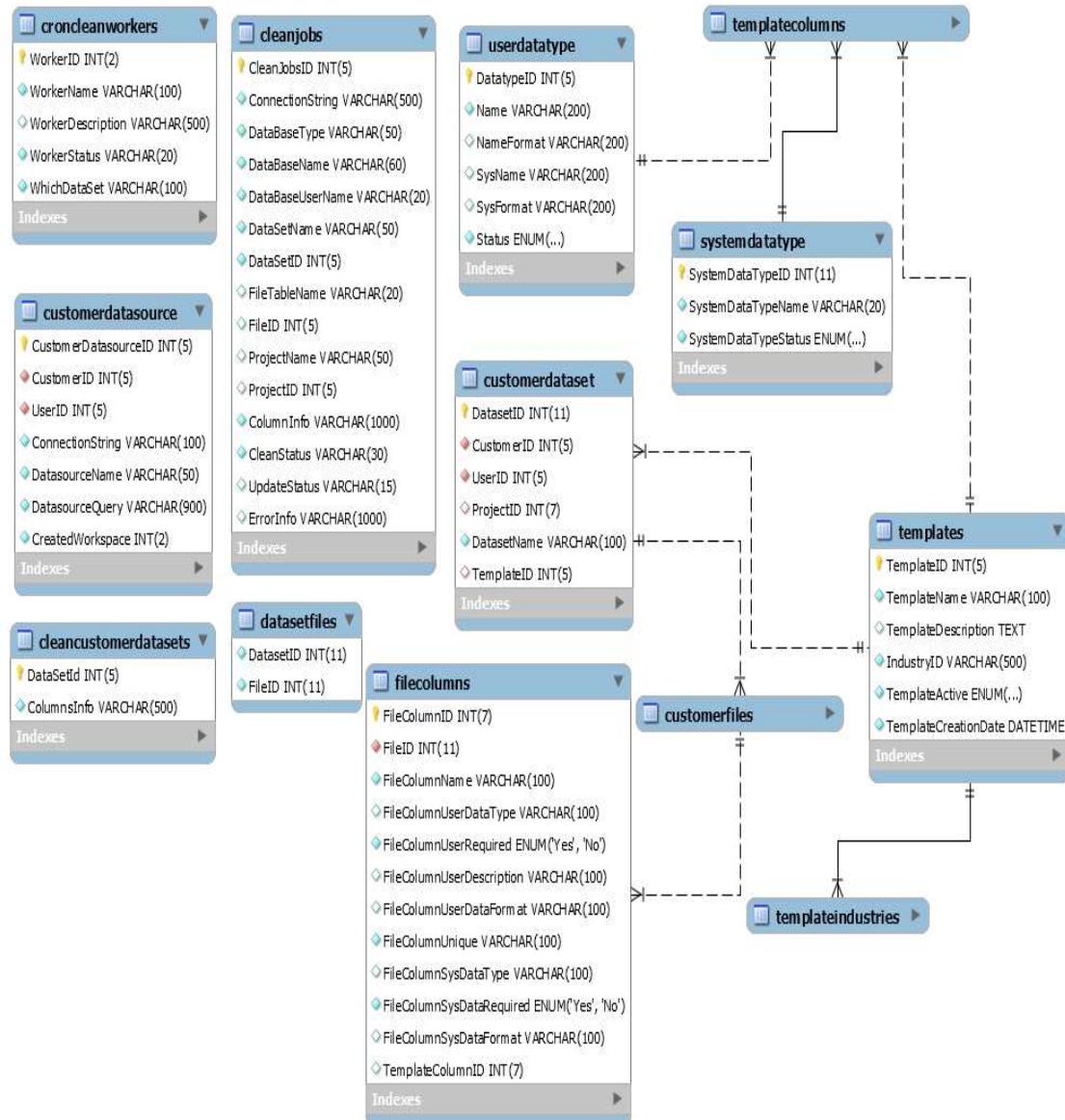


Figure 8.10: Cleaning ER Diagram

TEMPLATE ER DIAGRAM

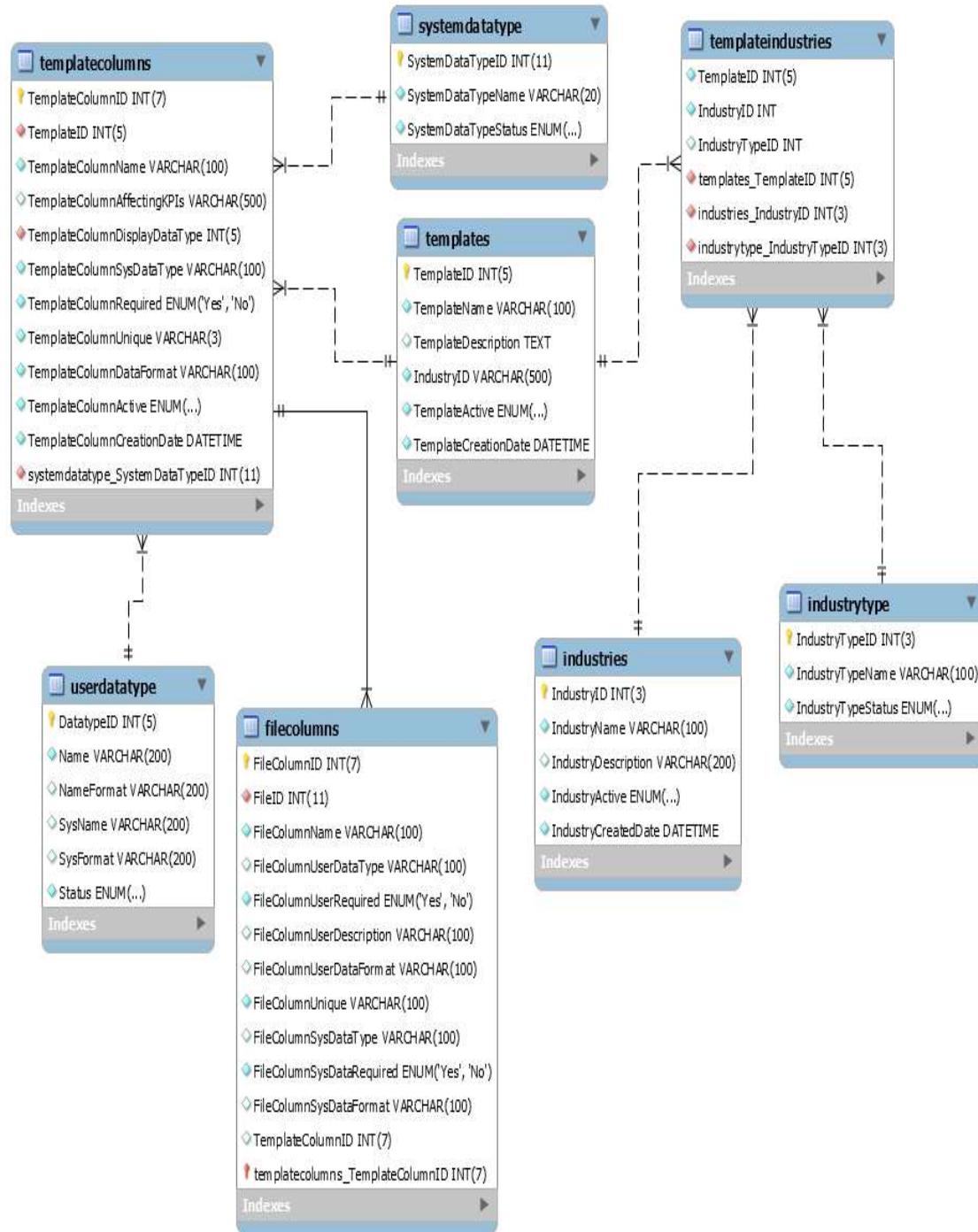


Figure 8.11: Template ER Diagram

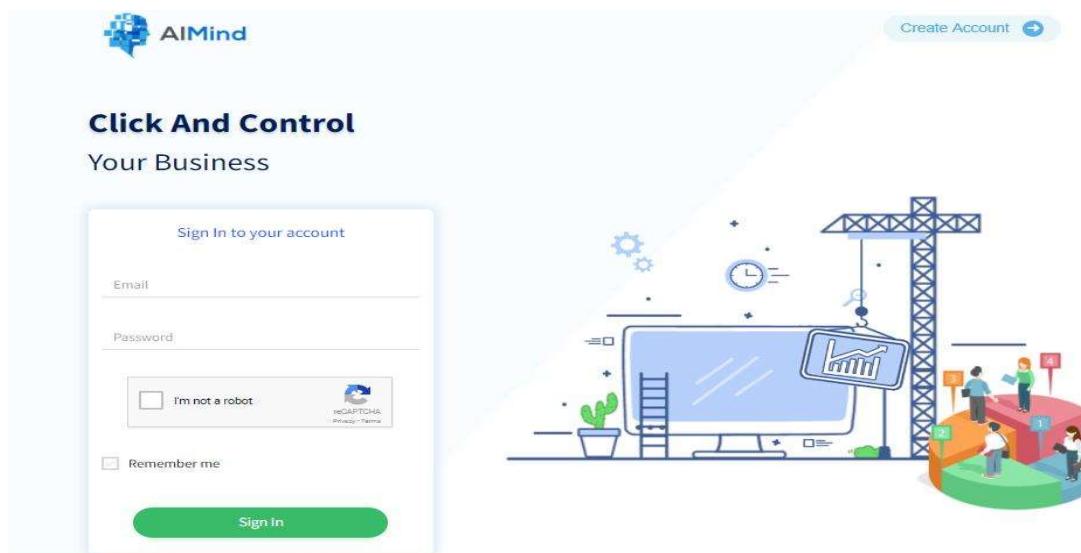
8.6 SCREENSHOTS

Admin Login page



The screenshot shows the AI Mind Admin Login page. At the top is a blue AI head logo and the text "AI Mind". Below it is a "Sign In" form. It has fields for "Email ID" and "Password". Underneath these is a reCAPTCHA box with the text "I'm not a robot" and the reCAPTCHA logo. To the right of the reCAPTCHA box is a link to "Privacy - Terms". Below the reCAPTCHA are two buttons: "Remember me" and "Forgot password?". At the bottom is a large green "Log In" button.

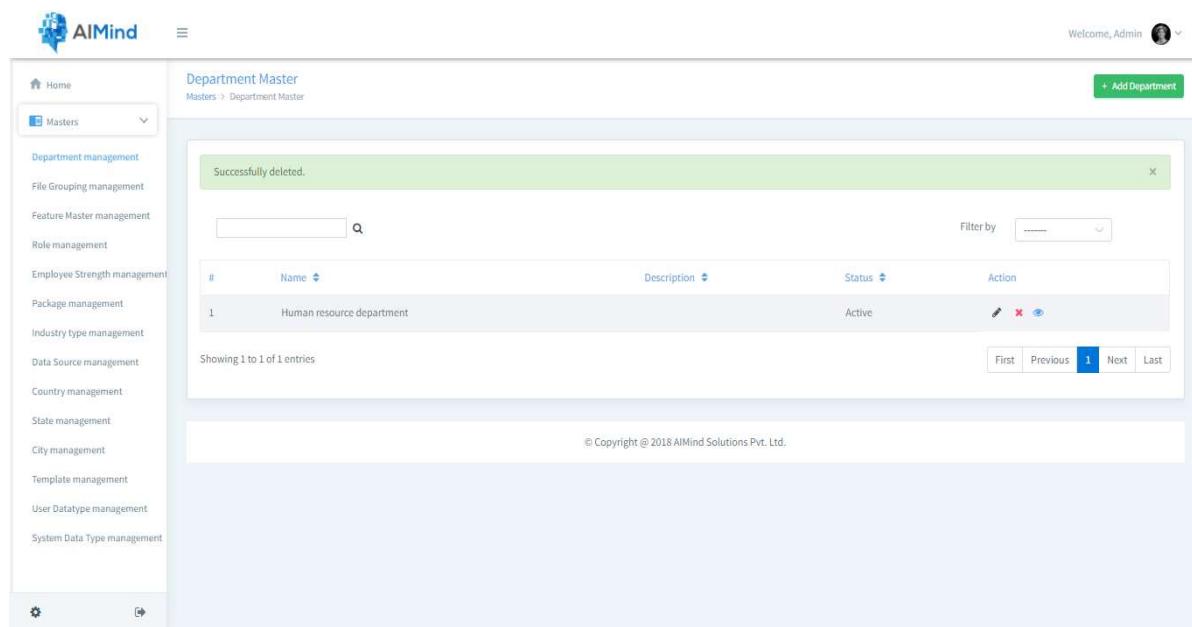
Customer Login Page



The screenshot shows the AI Mind Customer Login page. At the top left is the AI Mind logo. At the top right is a "Create Account" button with a blue arrow icon. Below the logo is a main heading "Click And Control" followed by "Your Business". To the left is a sign-in form with the title "Sign In to your account". It includes fields for "Email" and "Password", a reCAPTCHA box with "I'm not a robot" and "reCAPTCHA Privacy - Terms" links, and "Remember me" and "Sign In" buttons. To the right is a stylized illustration of a construction site. It features a large blue smartphone screen displaying a bar chart, a crane, a ladder, a potted cactus, and several people standing on a pink and green base. There are also small icons like gears and a magnifying glass.

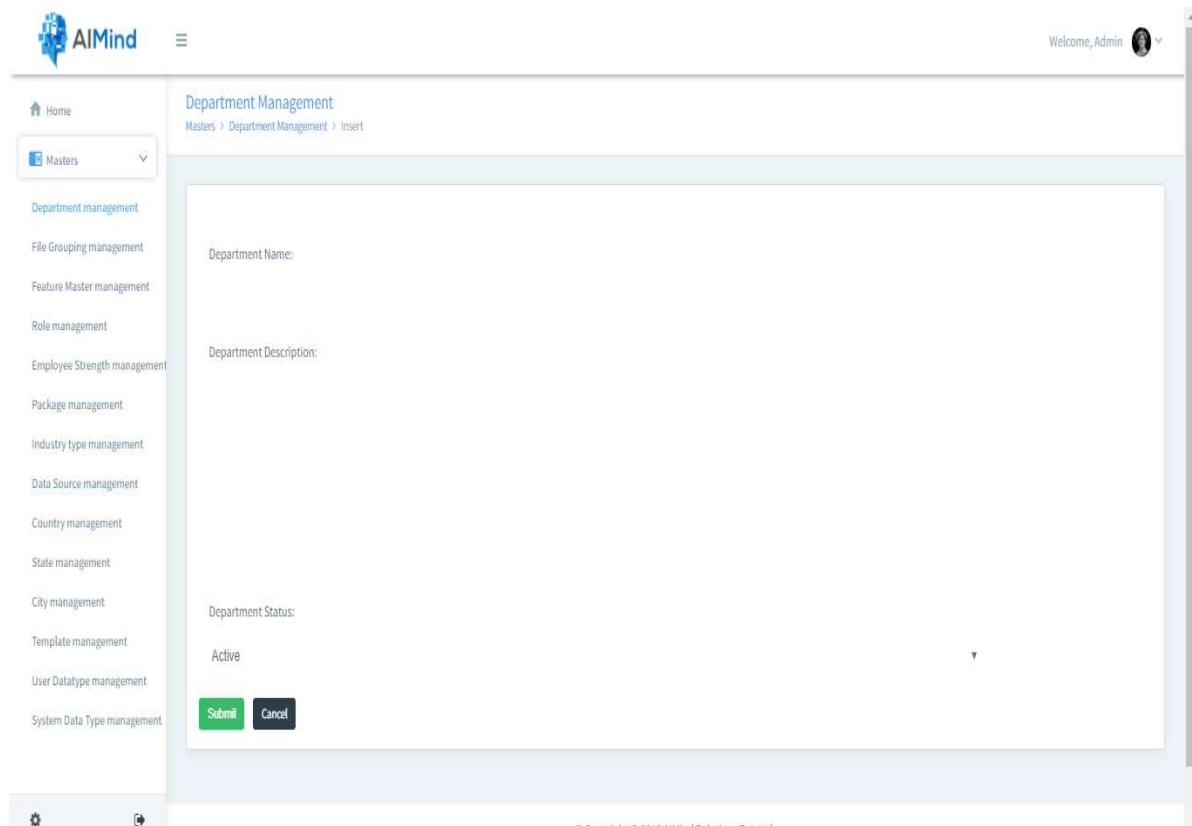
MASTER PAGES:

Department Master Page



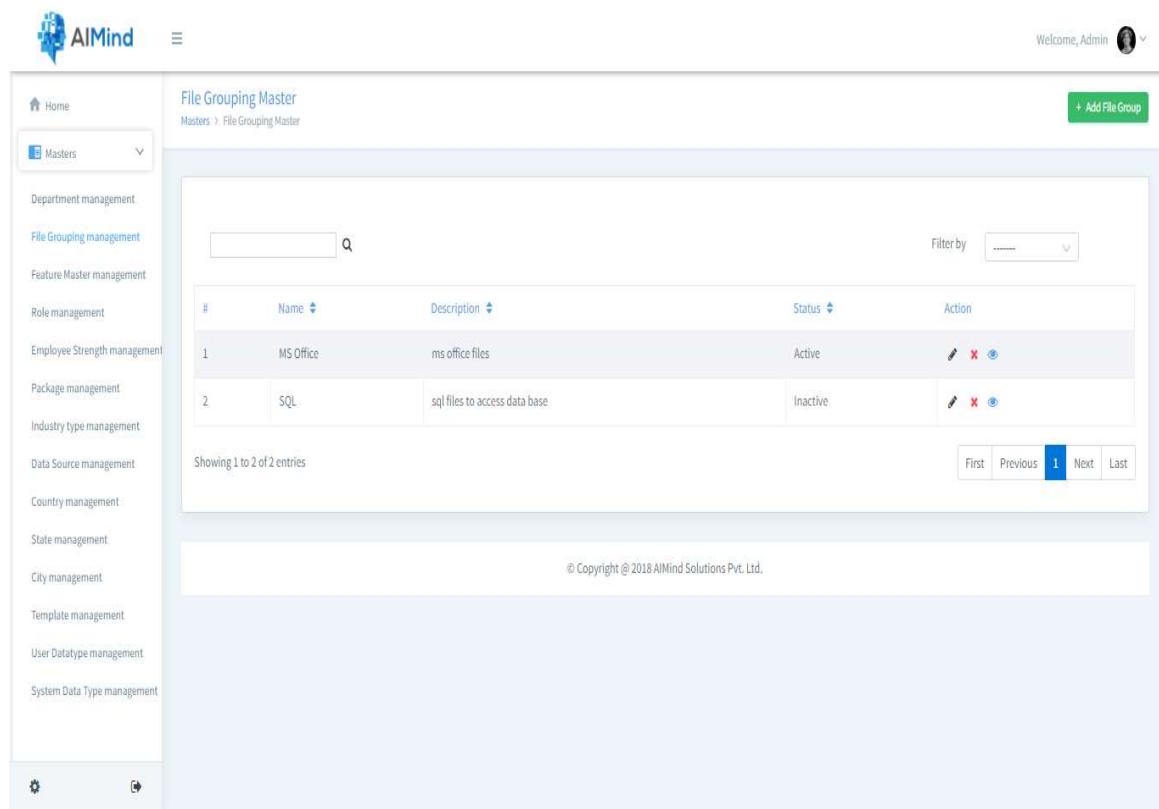
The screenshot shows the 'Department Master' page from the AIMind application. At the top right, it says 'Welcome, Admin' with a user icon. On the left, a sidebar menu under 'Masters' includes 'Department management'. The main content area displays a table with one entry: 'Human resource department' (Status: Active). A green banner at the top indicates a successful delete operation. Below the table, a footer note reads '© Copyright @ 2018 AIMind Solutions Pvt. Ltd.'

Add New Department



The screenshot shows the 'Insert' sub-page of 'Department Management'. The left sidebar has 'Masters' selected. The main form contains fields for 'Department Name' (with placeholder 'Department Name:'), 'Department Description' (with placeholder 'Department Description:'), and 'Department Status' (set to 'Active'). At the bottom are 'Submit' and 'Cancel' buttons.

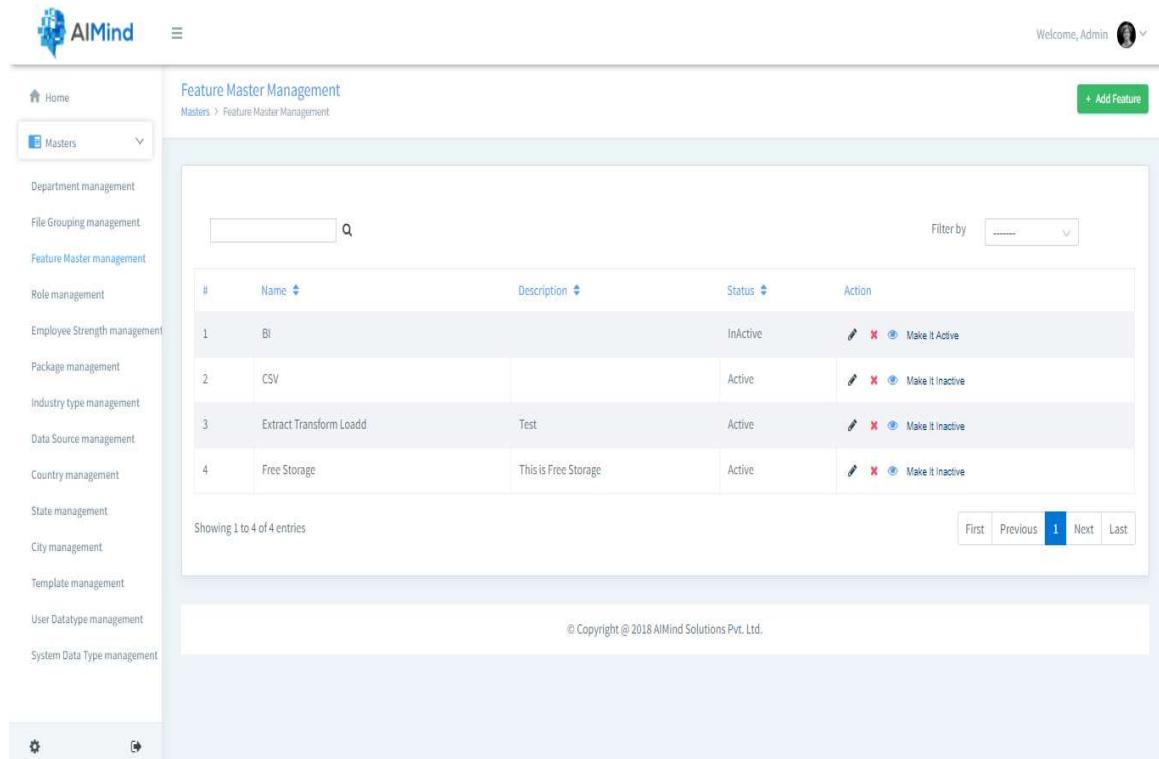
File Grouping Management Master Page



The screenshot shows the 'File Grouping Management' master page. The left sidebar lists various management categories under 'Masters'. The main content area displays a table of file groups with columns for Name, Description, Status, and Action. Two entries are shown: 'MS Office' (Description: ms office files, Status: Active) and 'SQL' (Description: sql files to access data base, Status: Inactive). A search bar and filter dropdown are at the top, and a navigation bar with 'First', 'Previous', '1', 'Next', and 'Last' is at the bottom.

#	Name	Description	Status	Action
1	MS Office	ms office files	Active	
2	SQL	sql files to access data base	Inactive	

Feature Master Management Page



The screenshot shows the 'Feature Master Management' page. The left sidebar lists various management categories under 'Masters'. The main content area displays a table of features with columns for Name, Description, Status, and Action. Four entries are shown: 'BI' (Status: InActive), 'CSV' (Status: Active), 'Extract.Transform Loaddd' (Status: Active), and 'Free Storage' (Status: Active). A search bar and filter dropdown are at the top, and a navigation bar with 'First', 'Previous', '1', 'Next', and 'Last' is at the bottom.

#	Name	Description	Status	Action
1	BI		InActive	Make it Active
2	CSV		Active	Make it Inactive
3	Extract.Transform Loaddd	Test	Active	Make it Inactive
4	Free Storage	This is Free Storage	Active	Make it Inactive

Role Management Master Page

The screenshot shows the 'Role Master Management' page under the 'Masters' section. The left sidebar lists various management categories. The main area displays a table of roles:

#	Name	Description	Status	Action
1	Admin		Inactive	edit delete details
2	Customer	This is customer	Active	edit delete details
3	Senior Project Manager	Test	Active	edit delete details

Below the table, it says 'Showing 1 to 3 of 3 entries'. The footer includes a copyright notice and navigation links: First, Previous, 1, Next, Last.

Employee Strength Management Master Page

The screenshot shows the 'Employee Strength Master' page under the 'Masters' section. The left sidebar lists various management categories. The main area displays a table of strength levels:

#	Strength	Status	Action
1	0-10	Active	edit delete details
2	20-30	Inactive	edit delete details
3	30-40	Inactive	edit delete details

Below the table, it says 'Showing 1 to 3 of 3 entries'. The footer includes a copyright notice and navigation links: First, Previous, 1, Next, Last.

Package Management Master Page

The screenshot shows the 'Package Master' page within the AIMind application. The left sidebar includes links for Home, Masters (selected), Department management, File Grouping management, Feature Master management, Role management, Employee Strength management, Package management (which is selected), Industry type management, Data Source management, Country management, State management, City management, Template management, User Datatype management, and System Data Type management. The main content area displays a table with three entries:

#	Name	Price	Duration	Status	Action
1	Basic	0	Monthly	Active	
2	Basic Package for starter	2000	Semi Annually	Active	
3	Package unlimited	100	Monthly	Active	

Below the table, it says 'Showing 1 of 3 entries'. At the bottom right, there are buttons for First, Previous, Next, and Last. The footer copyright notice is '© Copyright @ 2018 AIMind Solutions Pvt. Ltd.'

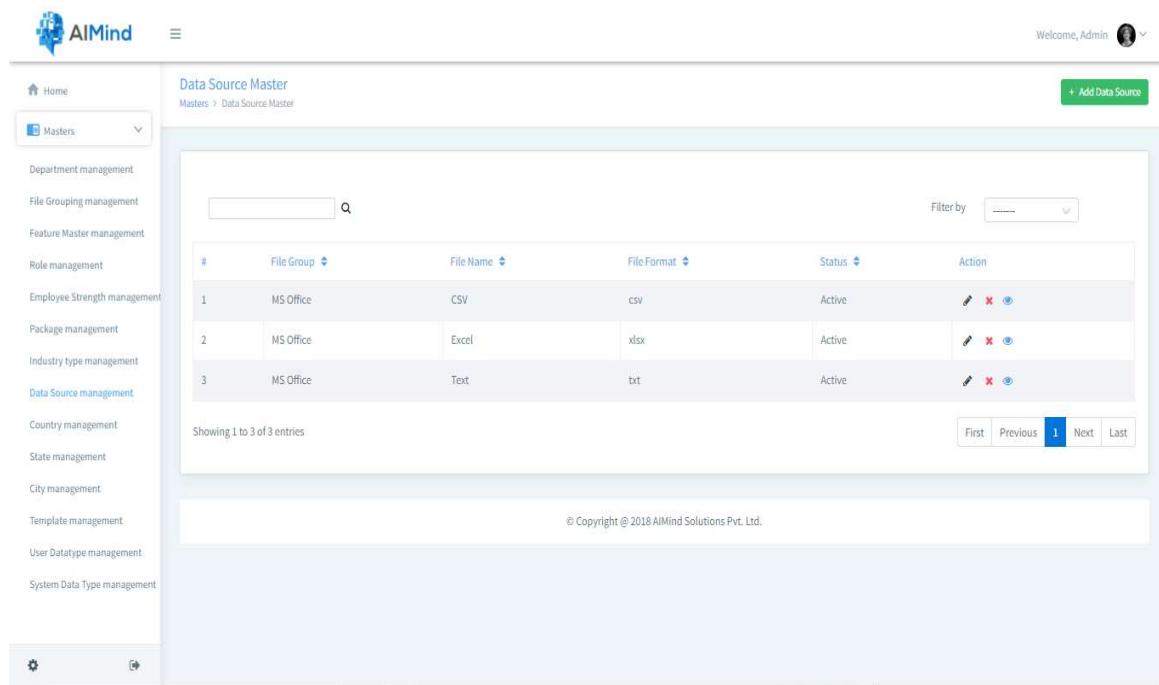
Industry Type Management Master Page

The screenshot shows the 'Industry type Master' page within the AIMind application. The left sidebar includes links for Home, Masters (selected), Department management, File Grouping management, Feature Master management, Role management, Employee Strength management, Package management, Industry type management (which is selected), Data Source management, Country management, State management, City management, Template management, User Datatype management, and System Data Type management. The main content area displays a table with one entry:

#	Name	Description	Status	Action
1	Supermarket	a large self-service shop selling foods and household goods.	Active	

Below the table, it says 'Showing 1 to 1 of 1 entries'. At the bottom right, there are buttons for First, Previous, Next, and Last. The footer copyright notice is '© Copyright @ 2018 AIMind Solutions Pvt. Ltd.'

Data Source Management Master Page

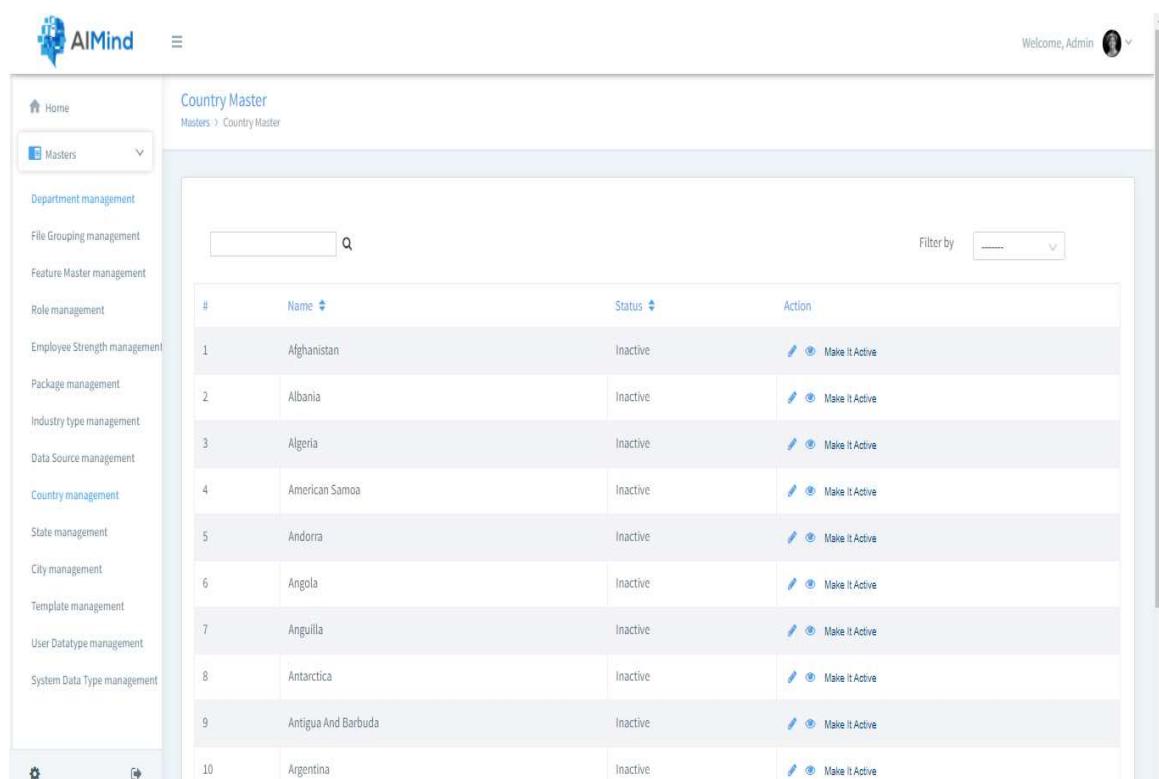


The screenshot shows the Data Source Management Master Page. The left sidebar contains a navigation menu with options like Home, Masters (selected), Department management, File Grouping management, Feature Master management, Role management, Employee Strength management, Package management, Industry type management, Data Source management (selected), Country management, State management, City management, Template management, User Datatype management, and System Data Type management. The main content area is titled "Data Source Master" and shows a table with three entries:

#	File Group	File Name	File Format	Status	Action
1	MS Office	CSV	CSV	Active	
2	MS Office	Excel	xlsx	Active	
3	MS Office	Text	txt	Active	

Below the table, it says "Showing 1 to 3 of 3 entries". At the bottom right, there are buttons for First, Previous, Next, and Last. The footer copyright notice is "© Copyright @ 2018 AIMind Solutions Pvt. Ltd."

Country Management Master Page



The screenshot shows the Country Management Master Page. The left sidebar contains a navigation menu with options like Home, Masters (selected), Department management, File Grouping management, Feature Master management, Role management, Employee Strength management, Package management, Industry type management, Data Source management, Country management (selected), State management, City management, Template management, User Datatype management, and System Data Type management. The main content area is titled "Country Master" and shows a table with ten entries:

#	Name	Status	Action
1	Afghanistan	Inactive	Make It Active
2	Albania	Inactive	Make It Active
3	Algeria	Inactive	Make It Active
4	American Samoa	Inactive	Make It Active
5	Andorra	Inactive	Make It Active
6	Angola	Inactive	Make It Active
7	Anguilla	Inactive	Make It Active
8	Antarctica	Inactive	Make It Active
9	Antigua And Barbuda	Inactive	Make It Active
10	Argentina	Inactive	Make It Active

State Management Master Page

The screenshot shows the 'State Master' page under the 'Masters' section. The left sidebar includes links for Home, Masters (selected), Department management, File Grouping management, Feature Master management, Role management, Employee Strength management, Package management, Industry type management, Data Source management, Country management, State management (selected), City management, Template management, User Datatype management, and System Data Type management. The main content area displays a table titled 'State Master' with columns: #, State Name, Country Name, Status, and Action. The table contains 10 rows of data.

#	State Name	Country Name	Status	Action
1	'Adan	Yemen	Active	
2	'Ajlun	Jordan	Inactive	
3	'Ali Sabih	Djibouti	Inactive	
4	'Ayn Daflah	Algeria	Inactive	
5	'Ayn Tamushanat	Algeria	Inactive	
6	'Isa	Bahrain	Inactive	
7	A Coruna	Spain	Inactive	
8	A'lali-an-Nil	Sudan	Inactive	
9	A'ana	Samoa	Inactive	
10	Aargau	Switzerland	Inactive	

City Management Master Page

The screenshot shows the 'City Master' page under the 'Masters' section. The left sidebar includes links for Home, Masters (selected), Department management, File Grouping management, Feature Master management, Role management, Employee Strength management, Package management, Industry type management, Data Source management, Country management, State management, City management (selected), Template management, User Datatype management, and System Data Type management. The main content area displays a table titled 'City Master' with columns: #, City Name, State Name, Country Name, Status, and Action. The table contains 10 rows of data.

#	City Name	State Name	Country Name	Status	Action
1	Valeggio Sul Mincio	Verona	Italy	Active	
2	'Abadiah	Bashshar	Algeria	Inactive	
3	'Abbin	'Ajlun	Jordan	Inactive	
4	'Abdul Hakim	Punjab	Pakistan	Inactive	
5	'Abullah-as-Salam	al-Kuwayt	Kuwait	Inactive	
6	'Adan	'Adan	Yemen	Inactive	
7	'Afak	al-Qadisiyah	Iraq	Inactive	
8	'Afrif	ar-Riyad	Saudi Arabia	Inactive	
9	'Afrin	Halab	Syria	Inactive	
10	'Afula	Hazafon	Israel	Inactive	

Template Master Page

The screenshot shows the 'Template Master' page within the AIMind application. The left sidebar contains a navigation menu with various management options like Home, Masters, Department management, etc. The main content area displays a table titled 'Template Master' with two entries:

#	Template Name	Industry Name	Status	Action
1	Customers	Supermarket	Active	edit delete details
2	Sales	Supermarket	Active	edit delete details

Below the table, it says 'Showing 1 to 2 of 2 entries'. At the bottom right, there are links for First, Previous, Next, and Last. The footer copyright notice reads '© Copyright @ 2018 AIMind Solutions Pvt. Ltd.'

CUSTOMER WIZARD:

Select Path

The screenshot shows a modal dialog box titled 'Welcome! Jack Smith' with the sub-instruction 'What would you like to do?'. It contains two options:

- Choose a Data Set**: An icon of a document with a green checkmark, with the sub-instruction 'Select File to Upload'.
- Select any Template**: An icon of a document with a purple checkmark, with the sub-instruction 'Lorem Ipsum is simply dummy text'.

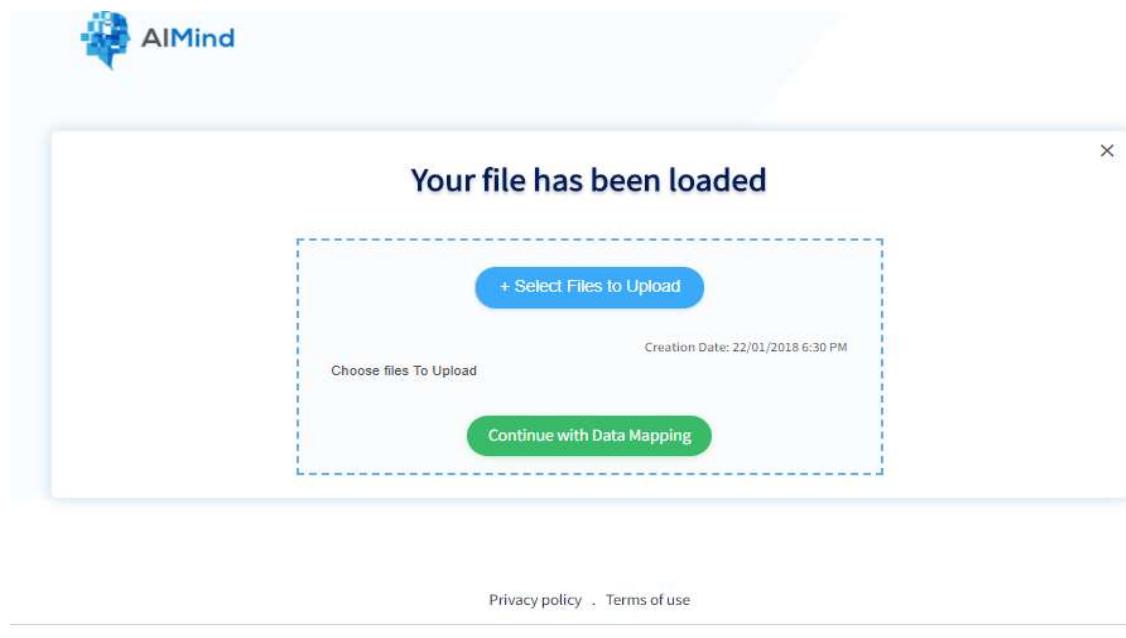
Select File to Upload

The screenshot shows a user interface for selecting a file to upload. At the top, there's a logo for 'AlMind' featuring a blue brain-like icon. Below the logo is the title 'Select File to Upload'. There are three categories of files displayed: 'CSV File' with a CSV icon, 'Excel File' with an Excel icon, and 'Text File' with a Text icon. At the bottom of the interface, there are links for 'Privacy policy' and 'Terms of use', followed by a copyright notice: 'Copyright © 2018 AlMind Solutions Pvt. Ltd.'

Give Dataset Name

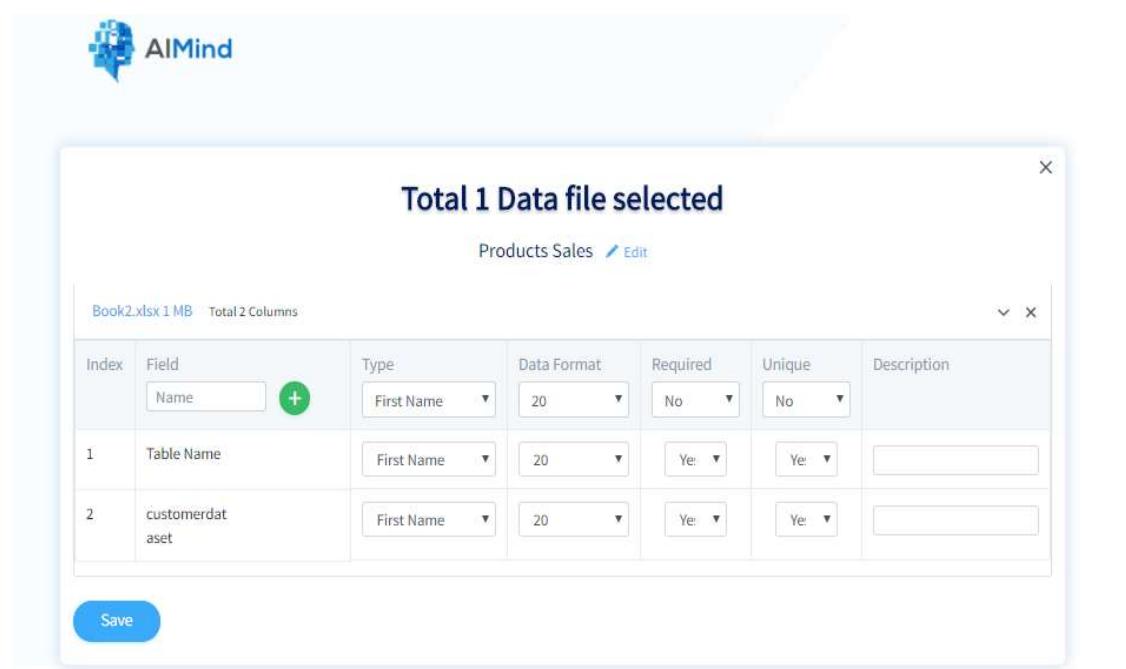
The screenshot shows a user interface for naming a dataset. At the top, there's a logo for 'AlMind' featuring a blue brain-like icon. Below the logo is the title 'Let's give your Data Set Name'. A text input field is present with the placeholder 'Source Name'. At the bottom, there is a green button labeled 'Continue with File Uploading'. At the very bottom of the page, there are links for 'Privacy policy' and 'Terms of use', followed by a copyright notice: 'Copyright © 2018 AlMind Solutions Pvt. Ltd.'

Upload Files



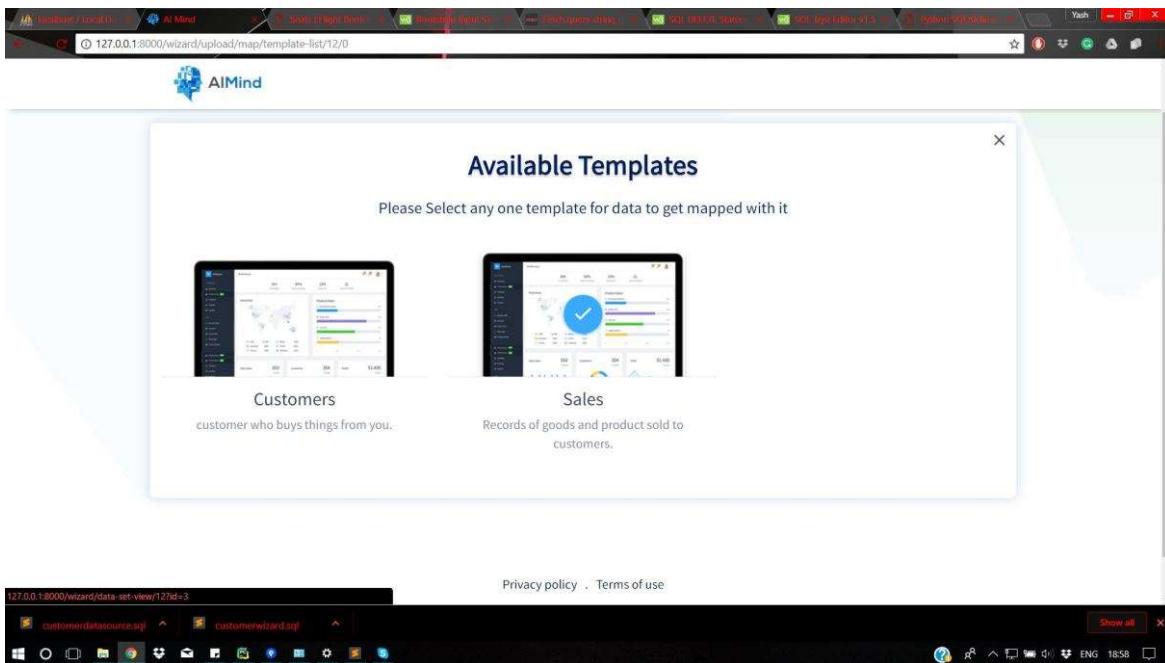
The screenshot shows the AIMind web interface after a file has been uploaded. At the top left is the AIMind logo. In the center, a blue banner displays the message "Your file has been loaded". Below this, a dashed-line box contains a blue button labeled "+ Select Files to Upload" and a green button labeled "Continue with Data Mapping". To the right of the box, the text "Choose files To Upload" is visible. At the bottom right of the box is the creation date "Creation Date: 22/01/2018 6:30 PM". A small "x" icon is located in the top right corner of the main window.

Data Mapping

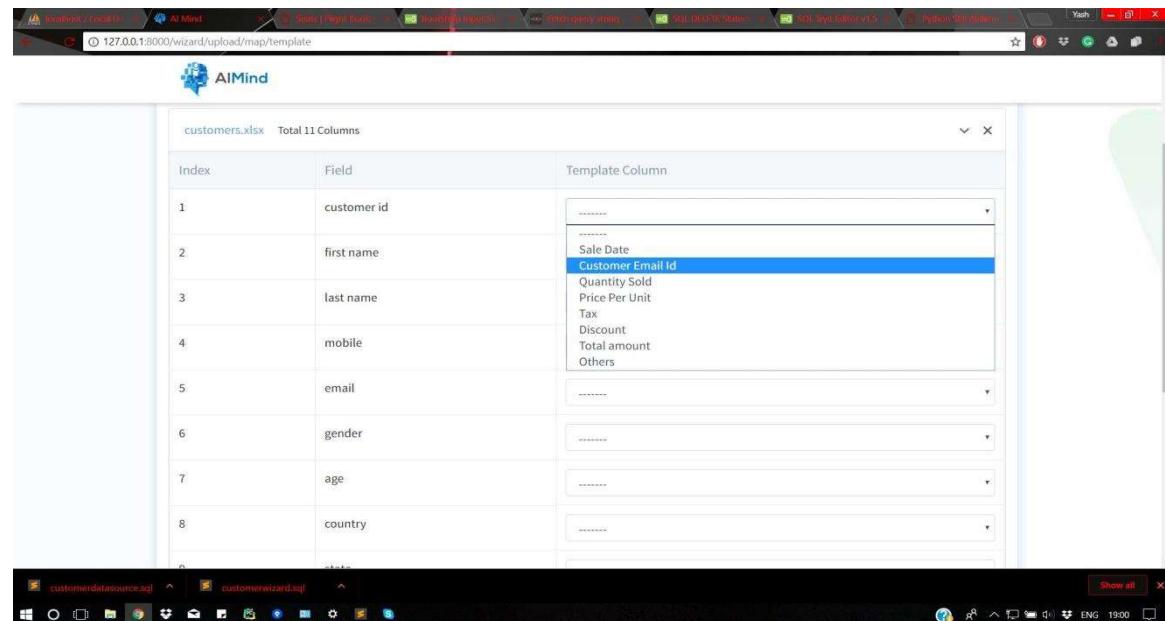


The screenshot shows the AIMind data mapping interface. At the top left is the AIMind logo. In the center, a blue banner displays the message "Total 1 Data file selected". Below this, the file name "Products Sales" is shown with an "Edit" link. A table titled "Book2.xlsx 1 MB Total 2 Columns" lists two columns: "Table Name" and "customerdataset". Each row has dropdown menus for "Type" (First Name), "Data Format" (20), "Required" (No), and "Unique" (No). The "Description" column is empty. A green "+" button is located next to the "Field" header. At the bottom left is a blue "Save" button. A small "x" icon is located in the top right corner of the main window.

Template Selection



Template Column Mapping



Dashboard

The screenshot shows the AIMind dashboard. At the top, there are three cards representing different dashboards:

- Performance**: A brand new dashboard for tracking performance related data.
- Analytics**: States, analytics and beautifully designed charts.
- Performance**: A brand new dashboard for tracking performance related data.

Below these cards is a section titled "Incomplete Wizard". It displays a table of recent datasets:

Data Sets	files	Owner	Status	Last Modified On	Action
Sales	currency_test_STADIN.xlsx	Jack Smith	complete	April 27, 2018, 5:59 a.m.	
Sales	currency_test1.xlsx	Jack Smith	complete_with_error	April 27, 2018, 6:28 a.m.	

At the bottom of the dashboard, there is a footer note: "© Copyright © 2018 AIMind Solutions Pvt. Ltd."

Add Dataset

The screenshot shows the AIMind interface with a browser window open at "127.0.0.1:8000/datasource/". The main area is titled "Create Your Own Datasource." and contains a "SELECTED FIELDS" dropdown menu. To the left, there is a sidebar with navigation links: Home, Data Sources (selected), Reports, and Video Tutorials.

A modal dialog box titled "Add Dataset" is open in the center. It has a sidebar labeled "Add tables" containing the following table names:

- file_2_master
- file_3_master
- invoice_2_master
- Invoice_master

Below the table names are "GO" and "Cancel" buttons.

Generate Relationship between Datasets

Add Dataset

```

SELECT
  file_2_master.Customer_First_Name, file_3_master.Tax, file_3_master.Customer_Last_Name, file_3_master.Discount, file_2_master.Customer_Email_Id, file_3_master.Total_amount
FROM
  (file_2_master INNER JOIN file_3_master ON file_2_master.Customer_Email_Id = file_3_master.Customer_Email_Id)
  
```

Run Edit Save

file_2_master

- Customer_First_Name
- Customer_Last_Name
- Customer_Email_Id
- Customer_Gender
- Customer_City
- Customer_State
- Living_Home_Address
- AlMindConfField

file_3_master

- Customer_Email_Id
- Category_Sort
- Price_Per_Unit
- Tax
- Discount
- Total_amount
- AlMindConfField

SELECTED FIELDS

- file_2_master.Customer_First_Name
- file_2_master.Customer_Last_Name
- file_3_master.Tax
- file_3_master.Discount
- file_3_master.Total_amount
- file_2_master.Customer_Email_Id

JOIN

file_2_master.Customer_Email_Id Inner JOIN file_3_master.Customer_Email_Id

WHERE

ORDER BY

Data Source Generated

Customer_First_Name	Tax	Customer_Last_Name	Discount	Customer_Email_Id
Nishant	2	Bhaviar	20	nishant@gmail.com
Yash	5	Patel	1	yash@gmail.com
Satyak	7	Patel	5	satyak@gmail.com

 The 'Add Dataset' screen below shows the same datasets and relationships as the previous screenshot, with the generated data visible in the preview area."/>

Add Dataset

```

SELECT
  file_2_master.Customer_First_Name, file_3_master.Tax, file_3_master.Customer_Last_Name, file_3_master.Discount, file_2_master.Customer_Email_Id, file_3_master.Total_amount
FROM
  (file_2_master INNER JOIN file_3_master ON file_2_master.Customer_Email_Id = file_3_master.Customer_Email_Id)
  
```

Run Edit Save

file_2_master

- Customer_First_Name
- Customer_Last_Name
- Customer_Email_Id
- Customer_Gender
- Customer_City
- Customer_State
- Living_Home_Address
- AlMindConfField

file_3_master

- Customer_Email_Id
- Category_Sort
- Price_Per_Unit
- Tax
- Discount
- Total_amount
- AlMindConfField

Query Data

Customer_First_Name	Tax	Customer_Last_Name	Discount	Customer_Email_Id
Nishant	2	Bhaviar	20	nishant@gmail.com
Yash	5	Patel	1	yash@gmail.com
Satyak	7	Patel	5	satyak@gmail.com

SELECTED FIELDS

- file_2_master.Customer_First_Name
- file_2_master.Customer_Last_Name
- file_3_master.Tax
- file_3_master.Discount
- file_3_master.Total_amount
- file_2_master.Customer_Email_Id

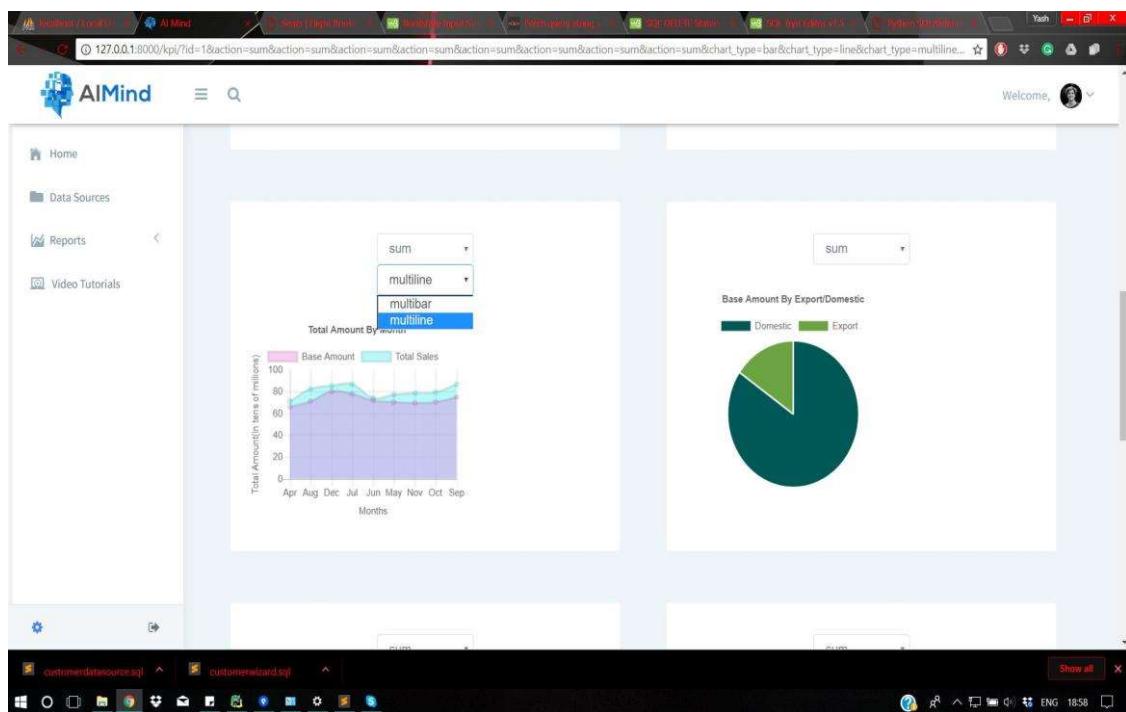
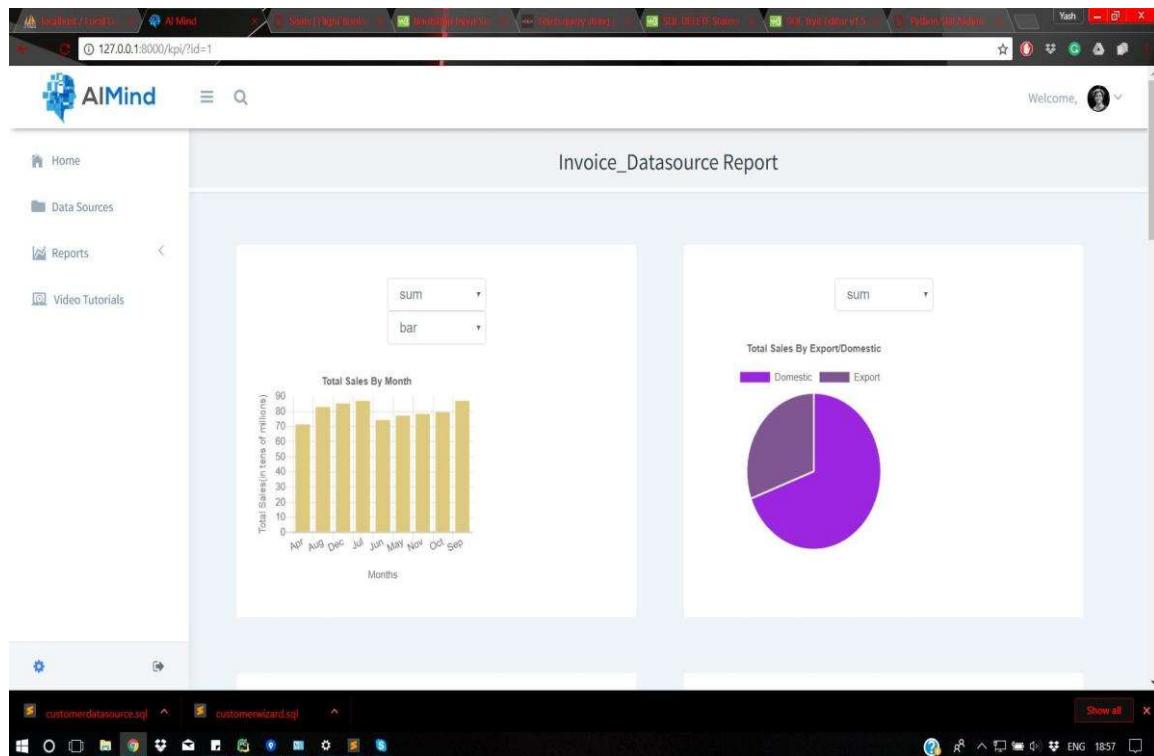
JOIN

file_2_master.Customer_Email_Id Inner JOIN file_3_master.Customer_Email_Id

WHERE

ORDER BY

KPI Dashboard



CHAPTER 9

LIMITATIONS AND FUTURE ENHANCEMENTS

- LIMITATIONS
- FUTURE ENHANCEMENTS

9.1 LIMITATIONS

- This web application contains templates for some industries only.
- Novice users will find it difficult to understand the whole working flow of the system.
- File uploading might take time as it is a web-application.
- Heavy files can take time to get uploaded and cleaned by the web-app.
- Custom graphs are not made at this stage of web-application.

9.2 FUTURE ENHANCEMENTS

- The web-application will be enhanced by Predictive Analysis in the future.
- More industry templates will be added in the future.
- Custom graphs will be made at later stage of the web-application.
- All types of dataset will be handled in the future.

CHAPTER 10

- CONCLUSION
- BIBLIOGRAPHY

10.1 CONCLUSION

Automated ETL & BI is a web application that allows users to visually analyze the data. Users can create and distribute an interactive dashboard which depicts the data in form of graphs and charts. This web application has prevented the need of manual data cleaning by including the feature of automated cleaning script. Automated cleaning script cleans the user data files by replacing the garbage data. The speed of automatic cleaning will vary with the size of the user file but overall the speed will be faster than manual cleaning. The user will be updated with an error report when file cleaning is completed with errors. The user is provided with the facility of generating a data source by setting relationships between datasets when the file cleaning is completed. The data source generated is further used to view the information visually in form of graphs. Thus, this web application helps companies to analyze their company data and make productive and intelligent decisions for their company in the near future.

BIBLIOGRAPHY**Websites**

<https://aws.amazon.com/premiumsupport/knowledge-center/python-boto3-virtualenv>

Django Pagination

<https://simpleisbetterthancomplex.com/tutorial/2016/08/03/how-to-paginate-with-django.html>

<https://medium.com/@sumitlni/paginate-properly-please-93e7ca776432>

Django Session and Cookie

https://docs.djangoproject.com/en/1.7/ref/settings/#std:setting-SESSION_EXPIRE_AT_BROWSER_CLOSE

<https://stackoverflow.com/questions/3024153/how-to-expire-session-due-to-inactivity-in-django>

<https://github.com/yourlabs/django-session-security>

Django Mongodb Logging

<https://pypi.python.org/pypi/log4mongo/1.6.2>

<https://www.ibm.com/developerworks/library/os-django-mongo/>

<https://www.loggly.com/docs/django-logs/>

<https://api.mongodb.com/python/current/tools.html>

<http://engineering.hackerearth.com/2015/02/26/logging-millions-requests-what-it-takes/>

<https://github.com/gnulnx/django-mongolog>

<https://github.com/andreisavu/mongodb-log/blob/master/mongolog/handlers.py>

<https://django-mongodb-engine.readthedocs.io/en/latest/>

Django Cron Job

<http://docs.celeryproject.org/en/latest/django/first-steps-with-django.html>

<https://docs.djangoproject.com/en/dev/dev/howto/custom-management-commands/#howto-custom-management-commands>

<https://github.com/Tivix/django-cron>

<http://django-cron.readthedocs.io/en/latest/installation.html>

https://www.reddit.com/r/Python/comments/m2dg8/explain_like_im_five_why_or_why_not_would_celery/

<https://simpleisbetterthancomplex.com/tutorial/2017/08/20/how-to-use-celery-with-django.html>

Sql Query Builder Jquery

<http://querybuilder.js.org/>

Data Cleaning

<https://www.analyticsvidhya.com/blog/2017/11/flashtext-a-library-faster-than-regular-expressions/>

<https://softwareengineering.stackexchange.com/questions/122440/how-do-regular-expressions-actually-work>

<https://stackoverflow.com/questions/1732348/regex-match-open-tags-except-xhtml-self-contained-tags/1732454#1732454>

https://en.wikipedia.org/wiki/Automata_theory

<http://danielweitzfeld.github.io/passtheroc/blog/2014/10/12/datasci-sqlalchemy/>

<https://dzone.com/articles/django-vs-sqlalchemy-which-python-orm-is-better>

<https://engineering.betterworks.com/2015/09/03/sqlalchemy-and-django/>

<https://stackoverflow.com/questions/6506578/how-to-create-a-new-database-using-sqlalchemy>

<https://stackoverflow.com/questions/29355674/how-to-connect-mysql-database-using-pythonsqlalchemy-remotely>

<https://stackoverflow.com/questions/270879/efficiently-updating-database-using-sqlalchemy-orm>

<https://stackoverflow.com/questions/9667138/how-to-update-sqlalchemy-row-entry>

<https://auth0.com/blog/sqlalchemy-orm-tutorial-for-python-developers/>

https://edgarroman.github.io/zappa-django-guide/walk_database/

<http://spejss.com/index.php/2017/12/21/connect-amazon-rds-sqlalchemy-python/>
