**CUSTOMER INFORMATION SYSTEM**

**A PROJECT REPORT**

*Submitted in partial fulfilment for the award of the degree*

*of*

**Master of Technology**

***In***

**Information Technology**

*by*

**Suryakant Kashinath Dubalgunde**

**17MIN0395**

*Under the guidance of*

**Prof. Animesh Giri**

**Asst.Professor**



**School of Information Technology and Engineering**

February, 2021



**School of Information Technology and Engineering**

**DECLARATION BY THE CANDIDATE**

I hereby declare that the thesis entitled “**CUSTOMER INFORMATION SYSTEM”** submitted by me to Vellore Institute of Technology University Vellore, in partial fulfillment of the requirement for the award of the degree of **Master of Technology** in **Information Technology** is a record of bonafide project work carried out by me under the supervision of  **Professor Animesh Giri.** I further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

**Place**:   Bangalore                  Suryakant .K. Dubalgunde

**Date**:    06/02/2021                    **Signature of the Candidate**

# 

**School of Information Technology and Engineering**

**BONAFIDE CERTIFICATE**

This is to certify that the project work entitled “**CUSTOMER INFORMATION SYSTEM”** by **SURYAKANT KASHINATH DUBALGUNDE(17MIN0395),** to Vellore Institute of Technology, Vellore, in partialfulfillment of the requirement for the award of the degree of **Master of Technology** in **Information Technology**, is a project bonafide work carried out by him under my supervision. The project fulfills the requirement as per the regulations of this Institute and in my opinion meets the necessary standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this Institute or any other Institute or University.

**Prof. Animesh Giri**

**Internal Supervisor**

**Asst. Professor**

**VIT University**



**Internal Examiner** **External Examiner**

**ACKNOWLEDGEMENT**

It is great pleasure for me to acknowledge the assistance and support our professor, teachers and mentor for the successful completion of this project.We take this opportunity to express our sincere gratitude to **Vellore Institute of** **Technology** for having provided us with a great opportunity to pursue ourMaster’s Degree in this institution.We are highly indebted to our internal guide **Prof. Animesh Giri,** Asst.Professor, Vellore Institute of Technology for his constant support and guidance.He has been a great source of inspiration throughout the course of this project.Finally, we gratefully acknowledge the support of our families during the completion of the project.

**Place**: Bangalore Suryakant Kashinath Dubalgunde

**Date**: 6th February,2021 **Signature of the Candidate**

**ABSTRACT**

CIS(CUSTOMER INFORMATION SYSTEM) is a System suite that transforms Customer Raw Data into Practical Intelligence and Knowledge. CIS combines externally derived data and internally company sourced data to create data analysis and reporting and helps the company for rational decision making. This system helps to lists the strategies, effective decision-making processes, technologies, etc. and supports the organization to make business operational decisions. The main objective of developing this system is to structure and analyze the historical data within a company smart insights providing scope for rational decision making.CIS plays a pivotal role in the strategic planning of any organizations decision-making process which includes performance progress, quantitative analysis, reporting, data sharing and understanding customer insights. The CIS system involves the use of computing technologies to identify, analyze and forecast or predict customer views. It can help in efficient a quick decision-making process by analyzing the vitality of the various parameters based on the legacy customer data and help the enterprise come up with effective solutions based on the performance indicators.CIS also removes the requirement for manual data entry and interpretations and provides with quick reporting features and data visualizations

### TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| Chapter No. | Title | Page No. |
|  | Abstract | 5 |
|  | List of Figures | 7 |
|  | List of Tables | 8 |
|  | List of Abbreviations | 9 |
| **1.** | **INTRODUCTION** | 10 |
|  | 1.1. Background | 10 |
|  | 1.2 Project Overview | 10 |
|  | 1.3. Scope | 10 |
|  | 1.4. Objectives | 11 |
|  | 1.5. Assumptions | 11 |
| **2.** | **LITERATURE SURVEY** | 12 |
|  | 2.1. System Planning | 12 |
|  | 2.2. Feasibility Study | 13 |
|  | 2.3. Techniques for Implementation | 13 |
|  | 2.4. System Requirement Specification | 14 |
| **3.** | **PROPOSED SYSTEM** | 17 |
|  | 3.1 Benefits of proposed system | 17 |
| **4.** | **DETAILED DESGIN OF PROJECT** | 18 |
|  | 4.1. System Design and Architecture | 18 |
|  | 4.2. Structural & Behavioral Diagrams | 19 |
| **5.** | **IMPLEMENTATION OF SYSTEM** | 23 |
|  | 5.1. Module Description | 23 |
|  | 5.2. Tools & Technologies Used | 26 |
|  | 5.3. Type of Input | 29 |
|  | 5.4. Description of dataset and source for it | 30 |
|  | 5.5 Snapshots | 34 |
|  | 5.6 Sample Code | 41 |
| **6.** | **TESTING AND RESULTS** | 43 |
| **7.** | **CONCLUSION & FUTURE WORK** | 44 |
|  | 7.1 Conclusion | 44 |
|  | 7.2 Future Work | 44 |
| **8.** | **LIST OF REFERENCES** | 45 |

**LIST OF FIGURES**

|  |  |
| --- | --- |
| **No** | **Title** |
| 1 | Fig:2.1.1 Software Development Plan |
| 2 | Fig 2.3. An Iterative and Incremental Approach for CUSTOMER INFORMATION SYSTEM |
| 3 | Fig 4.1: Architecture diagram of CUSTOMER INFORMATION SYSTEM |
| 4 | Fig 4.1.1: Working of MSBI System |
| 5 | Fig. 4.2.1 Dataflow Diagram |
| 6 | Fig 4.2.2 Sequence diagram |
| 7 | Fig.4.2.3 System Diagram |
| 8 | Fig. 4.2.4 Use Table case diagram |
| 9 | Fig 5.4 Database Tables |
| 10 | Fig 5.4.1 Country Table Design |
| 11 | Fig 5.4.2 Product Table Design |
| 12 | Fig 5.4.3 Time Table Design |
| 13 | Fig 5.4.4 CUSTOMER\_INFORMATION\_SYSTEM Table Design |
| 14 | Fig 5.4.5 Conversion\_Truncation\_Error\_Records\_Table Table Design |
| 15 | Fig 5.4.6 CUSTOMER\_INFORMATION\_SYSTEM\_File\_BadRecords Table Design |
| 16 | Fig 5.4.7 DataLoad\_Log Table Design |
| 17 | Fig 5.4.8 SSIS\_Error\_Log Table Design |
| 18 | Fig 5.5.1 Solution Explorer |
| 19 | Fig 5.5.2.1. Main Package |
| 20 | Fig 5.5.2.2. Country Package |
| 21 | Fig 5.5.2.3 Product Package |
| 22 | Fig 5.5.2.4 Customer Package (Control Flow) |
| 23 | Fig 5.5.2.5 Customer Package(Data Flow Design) |
| 24 | Fig 5.5.3.1 SSAS CUBE DESIGN |
| 25 | Fig 5.5.3.2 SSAS CUBE |
| 26 | Fig 5.5.4.1. SSRS in Textual Format |
| 27 | Fig 5.5.4.2. SSRS in Graphical Format |
| 28 | Fig 5.5.5.1 No File Found |
| 29 | Fig 5.5.5.2 Data Load Fails |
| 30 | Fig 5.5.5.3 Data Load Success |
| 31 | Fig 5.6.1 Checks File Existence @Location |
| 32 | Fig 5.6.2. Send Mail if File Not Found |
| 33 | Fig 5.6.3 Send Mail if Data Load Fails |
| 34 | Fig 5.6.4 Send Mail if Data Load happens |

**LIST OF TABLES**

|  |  |
| --- | --- |
| **No** | **Title** |
| 1 | Table 2.4.3. Functional Requirement Specification table |
| 2 | Table 5.3 Type of Input |
| 3 | Table 6. Test Result |

|  |  |  |
| --- | --- | --- |
| **ACRONYM** | **EXPANSION** |  |
| CIS | Customer Information System |  |
| SSIS | Sql Server Integration Service |  |
| SSAS | Sql Server Analysis Services |  |
| SSRS | Sql Server Reporting Services |  |

# LIST OF ABBREVIATIONS

**1.INTRODUCTION**

**1.1 Background**

In a real world,whether it is a small scale business or high scale enterprises,they are facing failure or a huge loss in terms of goodwill or profit due to lack of effective decision and due to failure in identifying, analyzing and forecasting or predicting customer views..An organisation is just performing its business operations and not structuring and analyzing the historical data of the organisation which can help it to lists the strategies,take effective decision and supports the organization to make business operational decisions.

**1.2 Project Overview**

CIS is a System suite that transforms Customer Raw Data into Practical Intelligence and Knowledge.This system helps to lists the strategies, effective decision-making processes, technologies, etc. and supports the organization to make business operational decisions. The main objective of developing this system is to structure and analyze the historical data within a company smart insights providing scope for rational decision making.CIS plays a pivotal role in the strategic planning of any organizations decision-making process which includes performance progress, quantitative analysis, reporting, data sharing and understanding customer insights. The CIS system involves the use of computing technologies to identify, analyze and forecast or predict customer views. It can help in efficient a quick decision-making process by analyzing the vitality of the various parameters based on the legacy customer data and help the enterprise come up with effective solutions based on the performance indicators.CIS also removes the requirement for manual data entry and interpretations and provides with quick reporting features and data visualizations.

We are designing a system comprising of MSBI(Microsoft Business Intelligence)Suite. We are going to use three tools from MSBI Suite i.e. SSIS,SSAS,SSRS.The first tool i.e. SSIS is used to perform an ETL activity(Extracting the data from the various source systems,Transforming the data and Loading the data in datawarehouse).The second tool i.e. SSAS is used to perform an Analysis activity(Creating the Cubes of the data available in the datawarehouse for better Analysis.The third tool i.e. SSRS is used to perform an Reporting activity(Creating the Reports of the analysed data).

**1.3. Scope**

The traditional manual methods of managing,viewing,processing the customer informations are tedious as the someone have to be manually process the customer data. This is tedious, time consuming and prone to inaccuracies.Use of the face CIS system in lieu of the traditional methods will provide a fast and effective method of managing,viewing,processing customer data accurately. CIS combines externally derived data and internally company sourced data to create data analysis and reporting and helps the company for rational decision making. This system helps to lists the strategies, effective decision-making processes, technologies, etc. and supports the organization to make business operational decisions.

**1.4. Objectives**

The main objective of developing this system is to structure and analyze the historical data within a company smart insights providing scope for rational decision making.CIS plays a pivotal role in the strategic planning of any organizations decision-making process which includes performance progress, quantitative analysis, reporting, data sharing and understanding customer insights.

**1.5. Assumptions**

While creating this system,I have made below assumptions

* I am developing the system for an Ecommerce company which operates in multiple countries and deals with wide range of products.
* This system is developed to process the customer sales data of that company.
* For SSIS Package,we are using customer sales data(CSV file) as a source and Sql Server DW as a target for performing an ETL activity.
* For SSAS Cube,we are using Sql Server DW as a source for performing an analysis activity.
* For SSRS Report,we are using SSAS Cube as a source and for performing reporting activity.

**2. LITERATURE SURVEY**

Main aim of the survey is to generalize the findings. Many problems are faced by the researchers in the survey process.Then systematic literature review results were validated by conducting semi-structured, face- to-face interviews with software engineering researchers.

**2.1. System Planning**

Systems Planning is the first phase of SDLC.During the planning phase, the objective of the project is determined and the requirements of the system are considered. Meeting with managers or stake holders are held to determine the exact requirements of the project. An estimate of resources, such as personnel and costs, is prepared, to either bring changes in the current system or develop a new system.A schedule with tollgates is planned. All of the information is analyzed to see if there is an alternative solution to creating a new product.A feasibility study is conducted of the proposed project in the planning stage.If there is no other viable alternative, the information is assembled into a project plan and presented to management for approval. A rough budget for the project is made.Communication plans, meetings, contracts and potential risks are discussed in this phase. Finally, a Requirement Specification document is created which serves the purpose of guideline for the next phase of the model.

System planning is done by people who have faith in the future and a vision of the future adequate to form the basis for planning. System planning has two major outputs which embody its contributions. These are proposals and design concepts. The proposal is addressed to the decision-maker. One of its main ingredients consists of a statement of the objectives of the system. The objectives may be set by the system planner, or they may be articulated by him as an expression of his understanding of the concensus of other responsible persons. Design concepts evolving from system planning are based on the visualized system, and are addressed primarily to the system engineer. There are many problems and pitfalls associated with system planning which should be kept in mind in system planning. Consideration of these helps to make the plans more realistic, more likely to be accepted, and less likely to bog down in the implementation stages. The development of an adequate science of planning should carry high priority among systems people, as such a science will help to provide effective analytical and communicative tools for system planning.

**2.1.1. Gantt Chart**

A Gantt chart is a bar chart that provides a visual view of tasks scheduled over time. A Gantt chart is used for planning projects of all sizes, and it is a useful way of showing what work is scheduled to be done on a specific day. It can also help you view the start and end dates of a project in one simple chart.

A Gantt chart is a project management tool assisting in the planning and scheduling of projects of all sizes, although they are particularly useful for simplifying complex projects.Gantt charts convey this information visually. They outline all of the tasks involved in a project, and their order, shown against a timescale. This gives you an instant overview of a project, its associated tasks, and when these need to be finished.

# 

# Fig:2.1.1 Software Development Plan

**2.2. Feasibility Study**

Feasibility study is performed to check whether the proposed project is implementable and to know the benefits from the existing system. The most essential tasks performed by a Feasibility Study are the identification and description of systems, the evaluation of the systems and the selection of the best of the systems.

**2.2.1. Technical Feasibility**

Technical Feasibility study is performed to check whether the proposed system is technically feasible or not. Technical feasibility centers around the existing computer system (hardware, software, etc.) and to what extent it can support the proposed addition.

**2.2.2. Economic Feasibility**

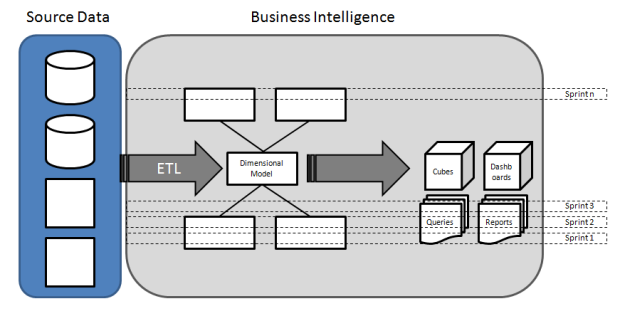
Economic Feasibility Study is the most frequently used method for evaluating the effectiveness of a candidate system. More commonly known as cost/benefit analysis, the procedure is to determine the benefits and savings that are expected from a candidate system and compare them with cost. This analysis phase determines how much cost is needed to produce the proposed system. This system is economically feasible since it does not require any initial setup cost, as the organization has required machines and supporting programs for the application to execute itself. It does not need additional staffing requirements.

**2.2.3. Operational Feasibility**

Operational feasibility is a measure of how well a proposed system solves the problems, and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development. Operational Feasibility study is performed to check whether the system is operationally feasible or not.

**2.3. Techniques** **for** **Implementation**

We are going to use an iterative and incremental approach to build the CIS SYSTEM.The incremental development approach came up with is aligned with the business priorities while allowing business users to further develop the knowledge of their business and better define their expectations.In this approach,we build layers of the entire data warehouse through iterations.



# Fig 2.3. An Iterative and Incremental Approach for CUSTOMER INFORMATION SYSTEM

As can be seen in the diagram, the intent is not to divide the project into phases where data is extracted from the source systems first, then move to a centralized repository and so forth. Instead of the traditional approach, we use a layer approach where each indicator is developed from source system to the presentation layer.

**2.4. SYSTEM REQUIREMENT SPECIFICATION**

Requirement Specification is a collection of the set of all requirements that are to be imposed on the design and verification of the product. The specification also contains other related information necessary for the design, verification, and maintenance of the product.A software requirements specification is a description of a software system to be developed. It is modeled after business requirements specification, also known as a stakeholder requirements specification.Software requirements specification establishes the basis for an agreement between customers and contractors or suppliers on how the software product should function. Software requirements specification is a rigorous assessment of requirements before the more specific system design stages, and its goal is to reduce later redesign.

The software requirements specification document lists sufficient and necessary requirements for the project development. To derive the requirements, the developer needs to have clear and thorough understanding of the products under development.

**2.4.1. Hardware Specification**

Hardware Requirements:

The system should run on Microsoft windows based system.

Processor : Intel Core i3 or higher

Speed : 1.70 GHz or higher

RAM : 4 GB or higher

Hard Disk : 100 GB or higher

**2.4.2. Software Specification**

Software Requirements:

Operating System : Windows 8.1 or higher

Front-End Tool : Visual Studio 2017

Back-End Tool : MS SQL Server Developer 2017

Front-End Technologies : SSIS,SSAS,SSRS,SQL

Back-End Technologies : SQL

**2.4.3. Functional Requirements**

Below section contains a detailed description of Functional requirements.

These are the statements of services that the system should provide, how the system should react for particular inputs and behave in particular situations.

* Should extract the Same Pre-defined Structured Raw Customer data from the CSV file.
* If CSV file is not found in the Pre-defined location,then should trigger the mail automatically.
* Should transform the Raw Customer Data into Customer Information.
* If CSV file has some corrupted data,then the corrupted data should not flow in final target table.
* If interface fails to execute then it should log the exception in Logging table and should notify to client.
* Should successfully load the Customer Information into the Customer Datawarehouse.
* Should analyze the Customer Information from the Customer Datawarehouse with the help of cubes.
* Should able to generate the reports for data visualization.

|  |  |  |
| --- | --- | --- |
| No | Technology | Functional Requirement Description |
| 1 | SSIS | The Package should check whether source csv files are available for processing or not. |
| 2 | SSIS | If files not available,then trigger the mail to client. |
| 3 | SSIS | If files available,then append the Processing Datetime to filename(eg:Customer\_20201128\_2106.csv). |
| 4 | SSIS | After renaming the file,try to load the data in Database.If some data is corrupted then that corrupted data should not flow in final target table.Corrupted data should flow in other tables. |
| 5 | SSIS | If data load fails,then should trigger the mail to client with error details,log the error and exception details and move that file to error file location. |
| 6 | SSIS | If data load is successful,then trigger the mail to client with load details and Archive the processed file. |
| 7 | SSIS | If package fails to execute then it should log the exception in Logging table and should notify to client. |
| 8 | SSAS | Once the data is loaded into database,Should be able to create Cube which is a combination of Dimension and Measure for analysis purpose. |
| 9 | SSRS | Once the data is analyzed,should be able to use that analyzed data for creating Reports like Tabular report or graph-based reports. |

Table 2.4.3. Functional Requirement Specification table

**2.4.4.** **Non-Functional requirements**

Below section contains a detailed description of Non-Functional requirements.

Non-Functional requirements define the needs in terms of performance, logical database requirements, design constraints, standards compliance, reliability, availability, security.

* PERFORMANCE:-System should provide the response within an acceptable time.
* RELIABILITY:- The outcome from the system should be reliable.
* AVAILABILITY:- The system shall be available for all the time.
* SECURITY:-Various location should be accessible by authorized users.
* MAINTAINABILITY:-System should be easily maintainable.
* PORTABILITY:-The System should be easily portable from one system to another.

**3. PROPOSED SYSTEM**

CIS is a System suite that transforms Customer Raw Data into Practical Intelligence and Knowledge. CIS combines externally derived data and internally company sourced data to create data analysis and reporting and helps the company for rational decision making. This system helps to lists the strategies, effective decision-making processes, technologies, etc. and supports the organization to make business operational decisions. The main objective of developing this system is to structure and analyze the historical data within a company smart insights providing scope for rational decision making.CIS plays a pivotal role in the strategic planning of any organizations decision-making process which includes performance progress, quantitative analysis, reporting, data sharing and understanding customer insights. The CIS system involves the use of computing technologies to identify, analyze and forecast or predict customer views. It can help in efficient a quick decision-making process by analyzing the vitality of the various parameters based on the legacy customer data and help the enterprise come up with effective solutions based on the performance indicators.CIS also removes the requirement for manual data entry and interpretations and provides with quick reporting features and data visualizations.

Some features of this application are: -

• CIS is a System suite that transforms Customer Raw Data into Practical Intelligence and Knowledge .

• This system helps to lists the strategies, effective decision-making processes, technologies etc. and supports the organization to make business operational decisions.

• This System helps to analyse and maintain the Historical Data of the Company sales in system.

• It helps for data visualization of the analysed data.

• It provides scope for rational decision making through analysed data.

• Mainly,it provides performance progress, quantitative analysis, reporting, data sharing and understanding customer insights.

* 1. **Benefits of proposed system**
* Easy to Use.
* Ensure data accuracy.
* Removes manual data entry.
* No Time Consumption.
* Flexible.
* Used for making rational decisions.
* Maintain historic data.

**4. DETAILED DESGIN OF PROJECT**

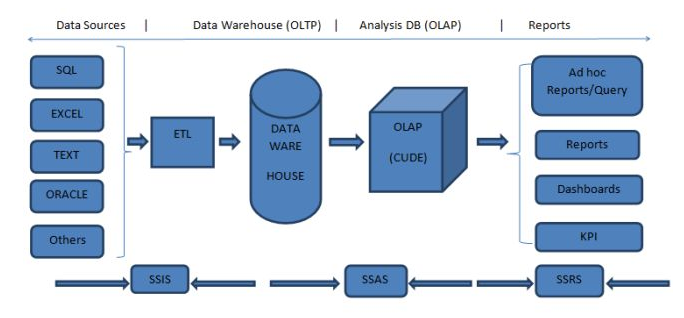
Design is a creative process. A good design is the key to effective system. The term “design” is defined as “the process of applying various techniques and principles for the purpose of defining a process or a system in sufficient detail to permit its physical realization”.

The design phase is the “architectural” phase of system design. The flow of data processing is developed into charts, and the project team determines the most logical design and structure for data flow and storage. The physical system is designed with the help of the logical design prepared by system analysts. The analysts and designers work together and use certain tools and software to create the overall system design, including the probable output.

The system and software design is prepared from the requirement specifications which were studied in the first phase. System Design helps in specifying hardware and system requirements and also helps in defining overall system architecture.

**4.1. System Design and Architecture**

In this design, several related components in terms of functionality have been grouped to form sub- systems which then combine to make up the whole system. Breaking the system down to components and sub-systems informs the logical design of the CIS system.

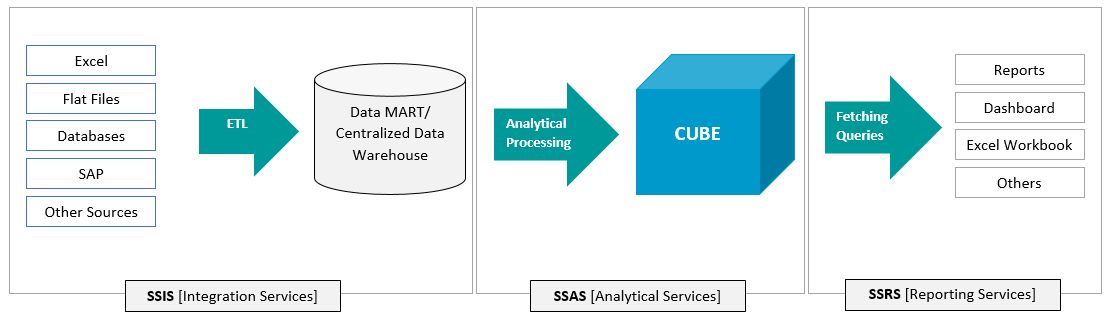
** Fig 4.1: Architecture diagram of CUSTOMER INFORMATION SYSTEM**

Our system follows the three tier architecture. ETL,Analysis,Reporting tier

1)ETL-Should successfully extract,transform and load the data from the source to target.

2)Analysis-Should successfully able to create Dimensions and Cubes for better analysis purpose.

3)Reporting or Display-Should successfully able to create the reports based on the analysed data.



**Fig 4.1.1: Working of MSBI System**

**4.2. Structural & Behavioral Diagrams**

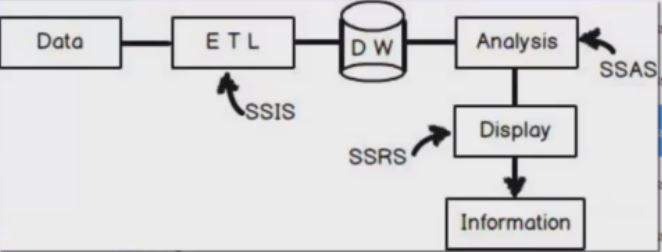
Structural diagram is an important part of the UML.It represents the static aspect of the system and the static parts of the diagrams are represented by classes, interfaces, objects, components and nodes.Theses diagrams show the things and different objects in a system being modeled.Structural diagram is an important part of the UML. It represents the static aspect of the system and the static parts of the diagrams are represented by classes, interfaces, objects, components and nodes. Theses diagrams show the things and different objects in a system being modeled.UML Structural diagrams depict the elements of a system that are independent of time and that convey the concepts of a system and how they relate to each other. The elements in these diagrams resemble the nouns in a natural language, and the relationships that connect them are structural or semantic relationships.example of structural diagrams are Class Diagram. Component Diagram. Deployment Diagram. Object Diagram. Package Diagram. Profile Diagram. Composite Structure Diagram.

Behavioral diagram is used to describe how the objects interact with each other to create a functioning system.These diagrams shows how would happen in a system.It is used to visualize, specify, construct and document the dynamic aspects of a system.It shows how the system behaves and interacts with itself and other entities (users, other systems). They show how data moves through the system, how objects communicate with each other, how the passage of time affects the system, or what events cause the system to change internal states.Behavioral Diagrams are Use Case Diagram. Activity Diagram. State Machine Diagram. Sequence Diagram. Communication Diagram. Interaction Overview Diagram.

The difference between structural and Behavioral diagram is the elements in a structure diagram represent the meaningful concepts of a system, and may include abstract, real world and implementation concepts. Behavior diagrams show the dynamic behavior of the objects in a system, which can be described as a series of changes to the system over time.

**4.2.1. Dataflow Diagram**

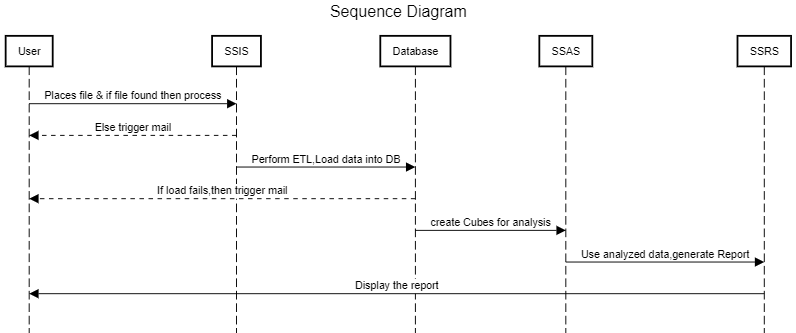
A data flow diagram is the graphical representation of the flow of data through an information system. DFD is very useful in understanding a system and can be efficiently used during analysis.A DFD shows the flow of data through a system. It views a system as a function that transforms the inputs into desired outputs. Any complex systems will not perform this transformation in a single step and a data will typically undergo a series of transformations before it becomes the output.With a data flow diagram, users are able to visualize how the system will operate that the system will accomplish and how the system will be implemented, old system data flow diagrams can be drawn up and compared with a new systems data flow diagram to draw comparisons to implement a more efficient system.



**Fig. 4.2.1 Dataflow Diagram**

**4.2.2. Sequence Diagram**

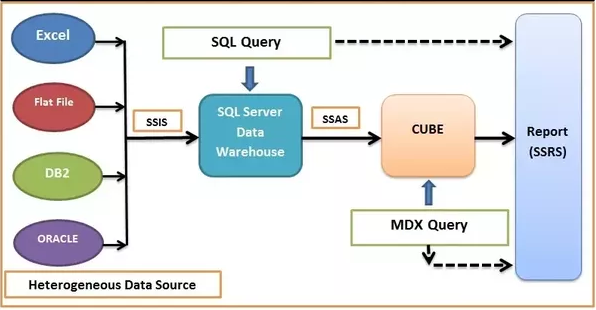
A sequence diagram shows object interactions arranged in time sequence. It depicts the objects involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the [Logical View](https://en.wikipedia.org/wiki/4%2B1_architectural_view_model) of the system under development. Sequence diagrams are sometimes called event diagrams or event scenarios.

****

**Fig 4.2.2 Sequence diagram**

**4.2.3. System Diagram**

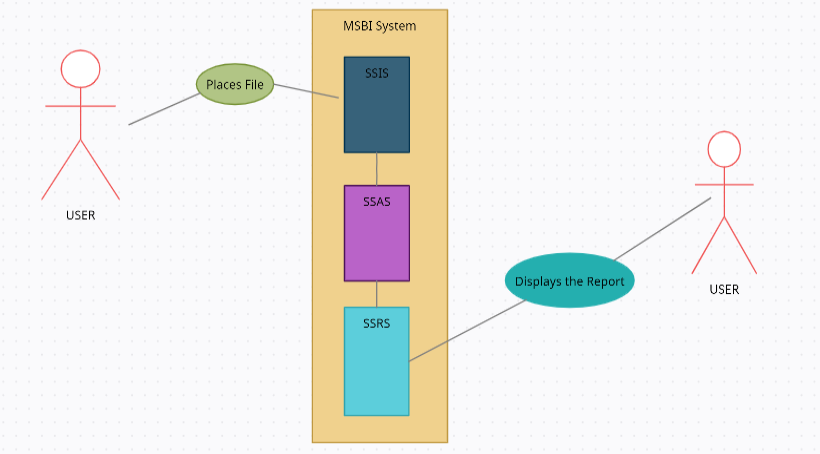
A system context diagram (SCD) in [engineering](https://en.wikipedia.org/wiki/Engineering) is a [diagram](https://en.wikipedia.org/wiki/Diagram) that defines the boundary between the [system](https://en.wikipedia.org/wiki/System), or part of a system, and its environment, showing the entities that interact with it. This diagram is a high level view of a [system](https://en.wikipedia.org/wiki/System). It is similar to a [block diagram](https://en.wikipedia.org/wiki/Block_diagram).

.

**Fig.4.2.3 System Diagram**

**4.2.4. Use case Diagram**

A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different [use cases](https://en.wikipedia.org/wiki/Use_case) in which the user is involved. A use case diagram can identify the different types of users of a system and the different use cases and will often be accompanied by other types of diagrams as well. The use cases are represented by either circles or ellipses.



**Fig. 4.2.4 Use case diagram**

**5. IMPLEMENTATION OF SYSTEM**

**5.1. MODULE DESCRIPTION**

The following modules will be used in the implementation of the designed system.

**5.1.1. ETL**

ETL stands for “extract, transform, and load.” The process of ETL plays a key role in data integration strategies. ETL allows businesses to gather data from multiple sources and consolidate it into a single, centralized location. ETL also makes it possible for different types of data to work together.

In computing, extract, transform, load (ETL) is the general procedure of copying data from one or more sources into a destination system which represents the data differently from the source(s) or in a different context than the source(s). The ETL process became a popular concept in the 1970s and is often used in data warehousing.

Data extraction involves extracting data from homogeneous or heterogeneous sources; data transformation processes data by data cleaning and transforming them into a proper storage format/structure for the purposes of querying and analysis; finally, data loading describes the insertion of data into the final target database such as an operational data store, a data mart, data lake or a data warehouse.

A properly designed ETL system extracts data from the source systems, enforces data quality and consistency standards, conforms data so that separate sources can be used together, and finally delivers data in a presentation-ready format so that application developers can build applications and end users can make decisions.

Since the data extraction takes time, it is common to execute the three phases in pipeline. While the data is being extracted, another transformation process executes while processing the data already received and prepares it for loading while the data loading begins without waiting for the completion of the previous phases.

ETL systems commonly integrate data from multiple applications (systems), typically developed and supported by different vendors or hosted on separate computer hardware. The separate systems containing the original data are frequently managed and operated by different employees. For example, a cost accounting system may combine data from payroll, sales, and purchasing.

**5.1.2. Data Analysis**

Data analysis is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making. The purpose of Data Analysis is to extract useful information from data and taking the decision based upon the data analysis.

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains. In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively.Data mining is a particular data analysis technique that focuses on statistical modeling and knowledge discovery for predictive rather than purely descriptive purposes, while business intelligence covers data analysis that relies heavily on aggregation, focusing mainly on business information. In statistical applications, data analysis can be divided into descriptive statistics, exploratory data analysis (EDA), and confirmatory data analysis (CDA). EDA focuses on discovering new features in the data while CDA focuses on confirming or falsifying existing hypotheses. Predictive analytics focuses on the application of statistical models for predictive forecasting or classification, while text analytics applies statistical, linguistic, and structural techniques to extract and classify information from textual sources, a species of unstructured data. All of the above are varieties of data analysis.Analysis is the process of breaking a complex topic or substance into smaller parts in order to gain a better understanding of it. The technique has been applied in the study of mathematics and logic since before Aristotle (384–322 B.C.), though analysis as a formal concept is a relatively recent development.Analysis, refers to dividing a whole into its separate components for individual examination. Data analysis, is a process for obtaining raw data, and subsequently converting it into information useful for decision-making by users. Data, is collected and analyzed to answer questions, test hypotheses, or disprove theories.

Data integration is a precursor to data analysis, and data analysis is closely linked to data visualization and data dissemination.

**5.1.3. Reporting and Data visualization**

Reporting means collecting and presenting data so that it can be analyzed.When we talk about reporting in business intelligence (BI), we are talking about two things. One is reporting strictly defined. The other is “reporting” taken in a more general meaning.In the first case, reporting is the art of collecting data from various data sources and presenting it to end-users in a way that is understandable and ready to be analyzed. In the second sense, reporting means presenting data and information, so it also includes analysis–in other words, allowing end-users to both see and understand the data, as well as act on it.

Reporting can be classified in many different ways. One is to differentiate reporting by the role of the person(s) preparing the report: managed reporting is reporting prepared by technical personnel such as developers; ad-hoc reporting is instead the realm of the nontechnical end-user. Another way in which reporting can be classified is by identifying the most important features of a report, such as data tables, cross-tab reports, visualization features, etc.A report is a document that presents information in an organized format for a specific audience and purpose. Although summaries of reports may be delivered orally, complete reports are almost always in the form of written documents.

In modern business scenario, reports play a major role in the progress of business. Reports are the backbone to the thinking process of the establishment and they are responsible, to a great extent, in evolving an efficient or inefficient work environment.Reports communicate the planning, policies and other matters regarding an organization to the masses. a report is a concise summary distilled from a larger set of data, intended for a specific audience. For example, reports may detail the findings of an experiment or inquiry. ... Typically most reports are done in a word processor, but could be done in any text editor.The process of organizing data into informational summaries in order to monitor how different areas of a business are performing..The process of exploring data and reports in order to extract meaningful insights, which can be used to better understand and improve business performance.Reports communicate information which has been compiled as a result of research and analysis of data and of issues. Reports can cover a wide range of topics, but usually focus on transmitting information with a clear purpose, to a specific audience. Data visualization is the graphical representation of information and data.

**5.2. TOOLS & TECHNOLOGIES USED**

The following modules will be used in the implementation of the designed system.

**5.2.1. SSIS(SQL SERVER INTEGRATION SERVICE)**-

SSIS is the ETL tool from Microsoft.Integration Services is a platform for building high-performance data integration and workflow solutions, including extraction, transformation, and loading (ETL) operations for data warehousing.We can process the data from various locations and various formats (source locations) and save the data into a centralized repository as a Data Warehouse/Data Mart (destination).It includes graphical tools and wizards for building and debugging packages.SQL Server Integration Services is a platform for building enterprise-level data integration and data transformations solutions. Use Integration Services to solve complex business problems by copying or downloading files, loading data warehouses, cleansing and mining data, and managing SQL Server objects and data.Integration Services can extract and transform data from a wide variety of sources such as XML data files, flat files, and relational data sources, and then load the data into one or more destinations.

Integration Services includes a rich set of built-in tasks and transformations, graphical tools for building packages, and the Integration Services Catalog database, where you store, run, and manage packages.You can use the graphical Integration Services tools to create solutions without writing a single line of code. You can also program the extensive Integration Services object model to create packages programmatically and code custom tasks and other package objects.

**5.2,2. SSAS(SQL SERVER ANALYSIS SERVICE)-**

This is the process of converting two dimensional (rows and columns/OLTP) data into a multi-dimensional data model (OLTP). This will help you to analyze the large volume of data.SQL Server Analysis Services (SSAS) is the technology from the Microsoft Business Intelligence stack, to develop Online Analytical Processing (OLAP) solutions. In simple terms, you can use SSAS to create cubes using data from data marts / data warehouse for deeper and faster data analysis.Cubes are multi-dimensional data sources which have dimensions and facts (also known as measures) as its basic constituents. From a relational perspective dimensions can be thought of as master tables and facts can be thought of as measureable details. These details are generally stored in a pre-aggregated proprietary format and users can analyze huge amounts of data and slice this data by dimensions very easily. Multi-dimensional expression (MDX) is the query language used to query a cube, similar to the way T-SQL is used to query a table in SQL Server.Simple examples of dimensions can be product / geography / time / customer, and similar simple examples of facts can be orders / sales. A typical analysis could be to analyze sales in Asia-pacific geography during the past 5 years. You can think of this data as a pivot table where geography is the column-axis and years is the row axis, and sales can be seen as the values. Geography can also have its own hierarchy like Country->City->State. Time can also have its own hierarchy like Year->Semester->Quarter. Sales could then be analyzed using any of these hierarchies for effective data analysis.  
A typical higher level cube development process using SSAS involves the following steps:  
1) Reading data from a dimensional model  
2) Configuring a schema in BIDS (Business Intelligence Development Studio)  
3) Creating dimensions, measures and cubes from this schema  
4) Fine tuning the cube as per the requirements  
5) Deploying the cube

Some of the advantages:

1. Multi-dimensional analysis
2. Key Performance Indicator (KPI)
3. Scorecard
4. Slice, dice, drill-down functionalities
5. Good performance
6. Security and so on.

**5.2.3. SSRS(SQL SERVER REPORTING SERVICE)-**

Microsoft SQL Server Reporting Services (SSRS) is an enterprise reporting platform supporting traditional and interactive reports delivered over the web or through custom applications. It supports various data sources like two dimensional and multi-dimensional. SQL Server Reporting Services (SSRS) provides a set of on-premises tools and services that create, deploy, and manage mobile and paginated reports.The SSRS solution flexibly delivers the right information to the right users. Users can consume the reports via a web browser, on their mobile device, or via email.SQL Server Reporting Services offers an updated suite of products:

1. "Traditional" paginated reports brought up to date, so you can create modern-looking reports, with updated tools and new features for creating them.
2. New mobile reports with a responsive layout that adapts to different devices and the different ways you hold them.
3. A modern web portal you can view in any modern browser. In the new portal,you can organize and display mobile and paginated Reporting Services reports and KPIs. You can also store Excel workbooks on the portal.

SQL Server Reporting Services (SSRS) is a reporting software that allows you to produce formatted reports with tables in the form of data, graph, images, and charts. These reports are hosted on a server that can be executed any time using parameters defined by the users. It is part of Microsoft SQL Server Services suite.

The following are some features of SSRS:

1. Retrieve data from a different source
2. Web-based access to reports
3. Support for Ad-hoc reporting
4. Report builder helps to customize the reports for end-user.
5. Easy subscriptions options
6. Export functionality with lots of formats.
7. Display reports in various ways like tabular, Chart, Gauge, and so on.

**5.2.4. Visual Studio**

Visual Studio is the most powerful Universal Windows Platform development environment. It brings unparalleled productivity improvements, a streamlined acquisition experience and enhanced debugging tools for Universal Windows Platform developers. Microsoft Visual Studio is an [integrated development environment](https://en.wikipedia.org/wiki/Integrated_development_environment) (IDE) from [Microsoft](https://en.wikipedia.org/wiki/Microsoft). It is used to develop [computer programs](https://en.wikipedia.org/wiki/Computer_program), as well as [websites](https://en.wikipedia.org/wiki/Web_site), [web apps](https://en.wikipedia.org/wiki/Web_app), [web services](https://en.wikipedia.org/wiki/Web_service) and [mobile apps](https://en.wikipedia.org/wiki/Mobile_app). Visual Studio uses Microsoft software development platforms such as [Windows API](https://en.wikipedia.org/wiki/Windows_API), [Windows Forms](https://en.wikipedia.org/wiki/Windows_Forms), [Windows Presentation Foundation](https://en.wikipedia.org/wiki/Windows_Presentation_Foundation), [Windows Store](https://en.wikipedia.org/wiki/Windows_Store) and [Microsoft Silverlight](https://en.wikipedia.org/wiki/Microsoft_Silverlight). It can produce both [native code](https://en.wikipedia.org/wiki/Machine_code) and [managed code](https://en.wikipedia.org/wiki/Managed_code).Visual Studio includes a [code editor](https://en.wikipedia.org/wiki/Code_editor) supporting [IntelliSense](https://en.wikipedia.org/wiki/IntelliSense) (the [code completion](https://en.wikipedia.org/wiki/Code_completion) component) as well as [code refactoring](https://en.wikipedia.org/wiki/Code_refactoring). The integrated debugger works both as a source-level debugger and a machine-level debugger. Other built-in tools include a [code profiler](https://en.wikipedia.org/wiki/Profiling_(computer_programming)), designer for building [GUI](https://en.wikipedia.org/wiki/GUI) applications, [web designer](https://en.wikipedia.org/wiki/Web_designer), [class](https://en.wikipedia.org/wiki/Class_(computing)) designer, and [database schema](https://en.wikipedia.org/wiki/Database_schema) designer. It accepts plug-ins that expand the functionality at almost every level—including adding support for [source control](https://en.wikipedia.org/wiki/Source_control) systems (like [Subversion](https://en.wikipedia.org/wiki/Subversion_(software)) and [Git](https://en.wikipedia.org/wiki/Git)) and adding new toolsets like editors and visual designers for [domain-specific languages](https://en.wikipedia.org/wiki/Domain-specific_language) or toolsets for other aspects of the [software development lifecycle](https://en.wikipedia.org/wiki/Software_development_lifecycle) (like the [Azure DevOps](https://en.wikipedia.org/wiki/Azure_DevOps_Server) client: Team Explorer).

**5.2.5. SQL SERVER MANAGEMENT STUDIO**

SQL Server Management Studio (SSMS) is an integrated environment for managing any SQL infrastructure. Use SSMS to access, configure, manage, administer, and develop all components of SQL Server, Azure SQL Database, and Azure Synapse Analytics. SSMS provides a single comprehensive utility that combines a broad group of graphical tools with a number of rich script editors to provide access to SQL Server for developers and database administrators of all skill levels.SQL Server Management Studio (SSMS) is a software application first launched with [Microsoft](https://en.wikipedia.org/wiki/Microsoft) [SQL Server 2005](https://en.wikipedia.org/wiki/Microsoft_SQL_Server) that is used for configuring, managing, and administering all components within [Microsoft SQL Server](https://en.wikipedia.org/wiki/Microsoft_SQL_Server). It's the successor to the Enterprise Manager in SQL 2000 or before. The tool includes both script editors and graphical tools which work with objects and features of the server.A central feature of SSMS is the Object Explorer, which allows the user to browse, select, and act upon any of the objects within the server. It also shipped a separate Express edition that could be freely downloaded, however recent versions of SSMS are fully capable of connecting to and manage any SQL Server Express instance. Microsoft also incorporated backwards compatibility for older versions of SQL Server thus allowing a newer version of SSMS to connect to older versions of SQL Server instances. It also comes with Microsoft SQL Server Express 2012, or users can download it separately.

**5.2.5. SQL**

SQL is a domain-specific language used in programming and designed for managing data held in a relational database management system (RDBMS), or for stream processing in a relational data stream management system (RDSMS). It is particularly useful in handling structured data, i.e. data incorporating relations among entities and variables.

* SQL stands for Structured Query Language.
* It is designed for managing data in a relational database management system (RDBMS).
* It is pronounced as S-Q-L or sometime See-Qwell.
* SQL is a database language, it is used for database creation, deletion, fetching rows, and modifying rows, etc.

**5.3. Type of Input**

The main input for Customer Information System to run successfully is that Pre-Defined Structured CSV formatted file.

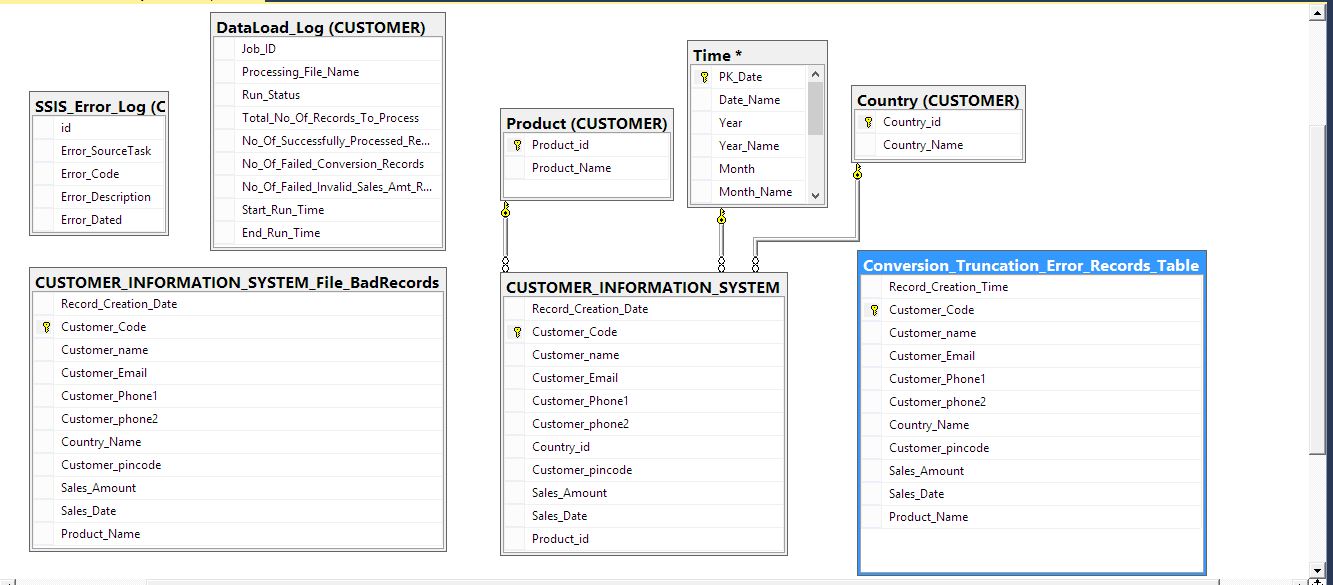
|  |  |
| --- | --- |
| **File Name** | **File Snapshot** |
| **Country.csv** |  |
| **Product.csv** |  |
| **Customer\_Jan2020.csv** |  |

**Table 5.3 Type of Input**

**5.4. Description of dataset and source for it**

•As a dataset for Pre-Defined Structured CSV formatted file,I am using some random data created by me.

•To store the Customer Information in Datawarehouse and that information to use for Analysis purpose,created few Random tables in SQL SERVER instance.

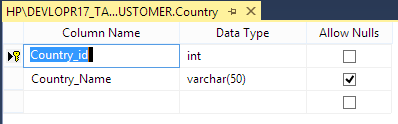
****

**Fig 5.4 Database Tables**

**5.4.1. Country Table**

Country table is used for storing the country name which can be used to analyse the data by country name.

Table design

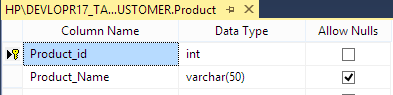


**Fig 5.4.1 Country Table Design**

**5.4.2. Product Table**

Product table is used for storing the Product name which can be used to analyse the data by Product name.

Table design

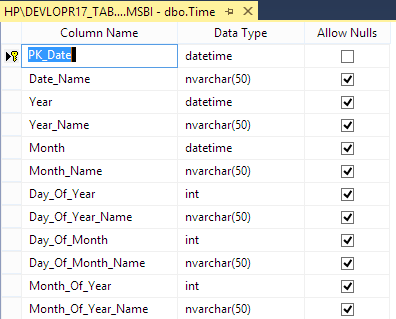


**Fig 5.4.2 Product Table Design**

**5.4.3. Time Table**

Time table is used for storing the Time data which can be used to analyse the data by time.

Table design

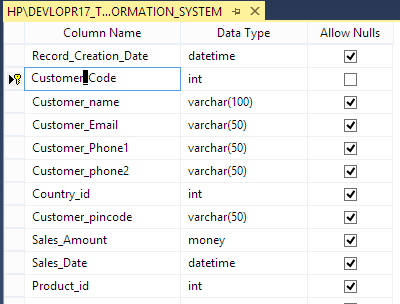


**Fig 5.4.3 Time Table Design**

**5.4.4. CUSTOMER\_INFORMATION\_SYSTEM Table**

CUSTOMER\_INFORMATION\_SYSTEM table is used for storing the customer data which can be used to analyse the data of customer wrt time,product,country etc.

Table design



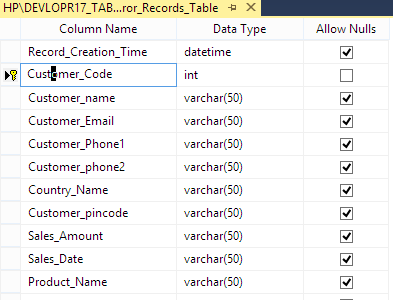
**Fig 5.4.4 CUSTOMER\_INFORMATION\_SYSTEM Table Design**

**5.4.5. Conversion\_Truncation\_Error\_Records\_Table Table**

Conversion\_Truncation\_Error\_Records\_Table table is used to capture the customer data where

Sales amount is not specified as money or in digit.

Table design

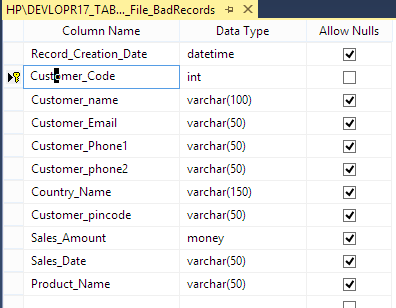


**Fig 5.4.5 Conversion\_Truncation\_Error\_Records\_Table Table Design**

**5.4.6. CUSTOMER\_INFORMATION\_SYSTEM\_File\_BadRecords Table**

CUSTOMER\_INFORMATION\_SYSTEM\_File\_BadRecords table is used to capture the customer data where Sales amount is specified as Null,<0,=0 etc.

Table design

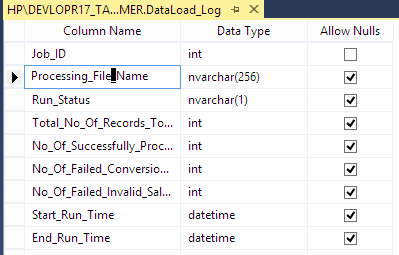


**Fig 5.4.6 CUSTOMER\_INFORMATION\_SYSTEM\_File\_BadRecords Table Design**

**5.4.7. DataLoad\_Log Table**

DataLoad\_Log table is used to capture the run status of the interface..

Table design

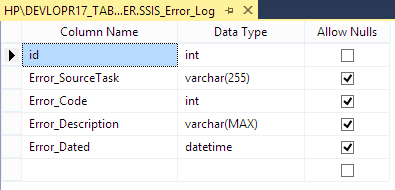


**Fig 5.4.7 DataLoad\_Log Table Design**

**5.4.8. SSIS\_Error\_Log Table**

SSIS\_Error\_Log table is used to capture the failure log of the interface..

Table design

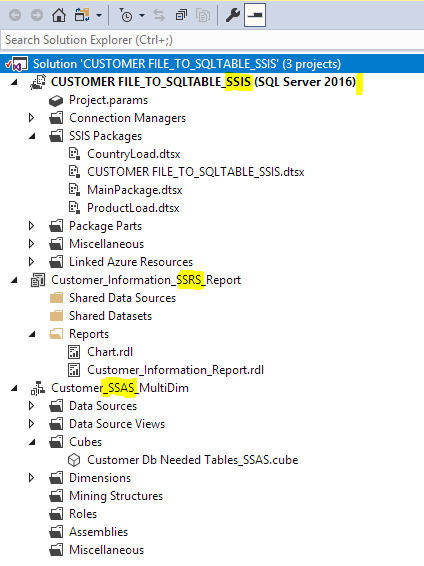


**Fig 5.4.8 SSIS\_Error\_Log Table Design**

**5.5 Snapshots**

**5.5.1 Solution Explorer**

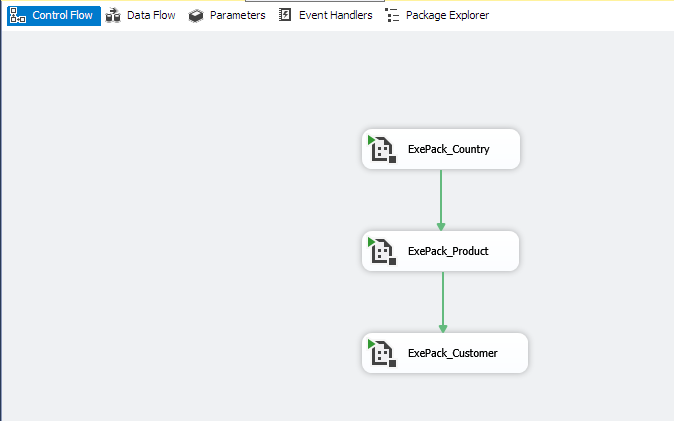
Shows all the Designed system in one window.

****

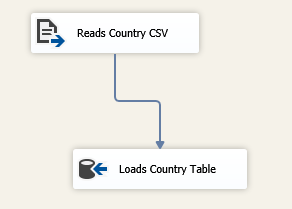
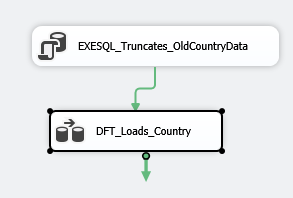
**Fig 5.5.1 Solution Explorer**

**5.5.2. SSIS PACKAGE**

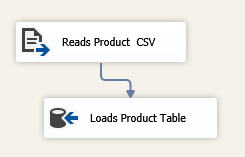
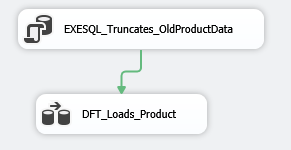
SSIS package is used for performing ETL Activity where it will extract the data from csv files located @location and then it will transform the data and load all the transformed data into the Database.



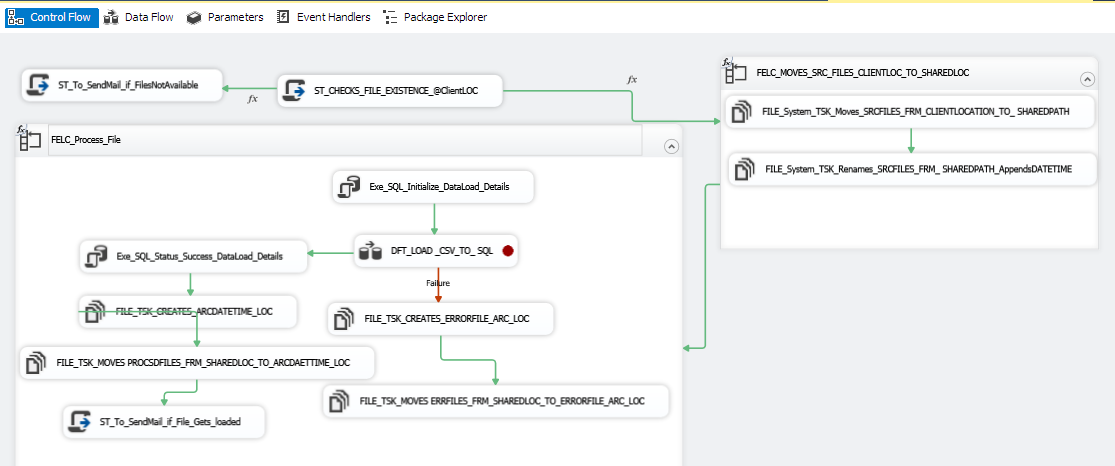
**Fig 5.5.2.1. Main Package**

****

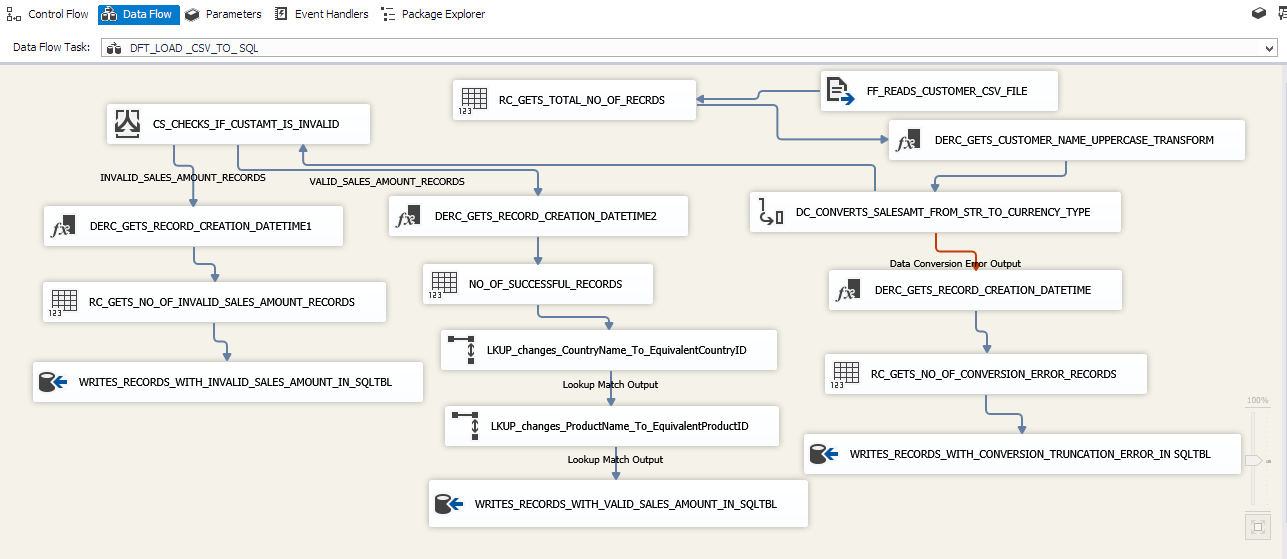
**Fig 5.5.2.2. Country Package**

****

**Fig 5.5.2.3 Product Package**

****

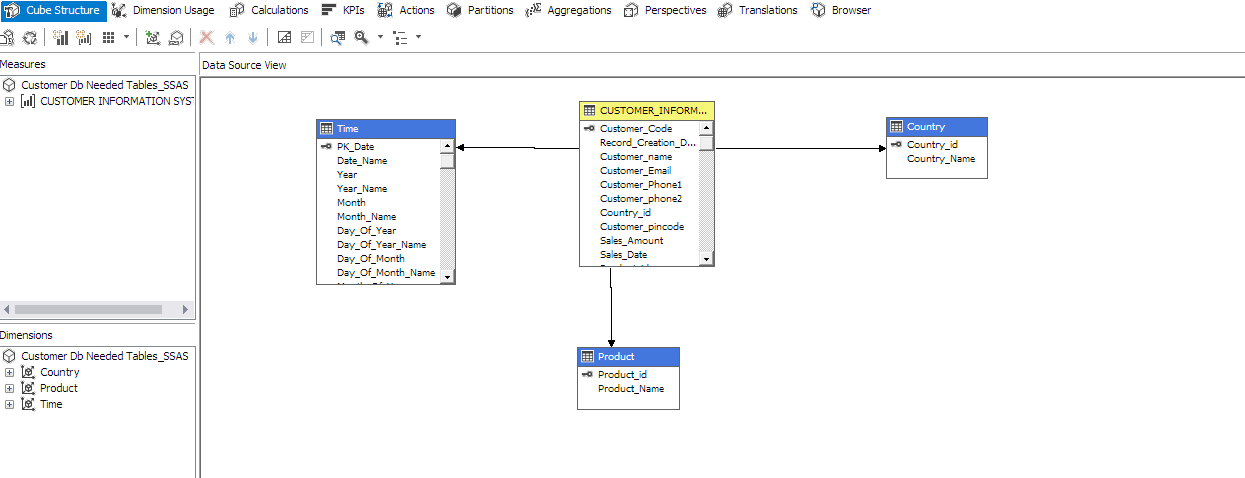
**Fig 5.5.2.4 Customer Package (Control Flow)**

****

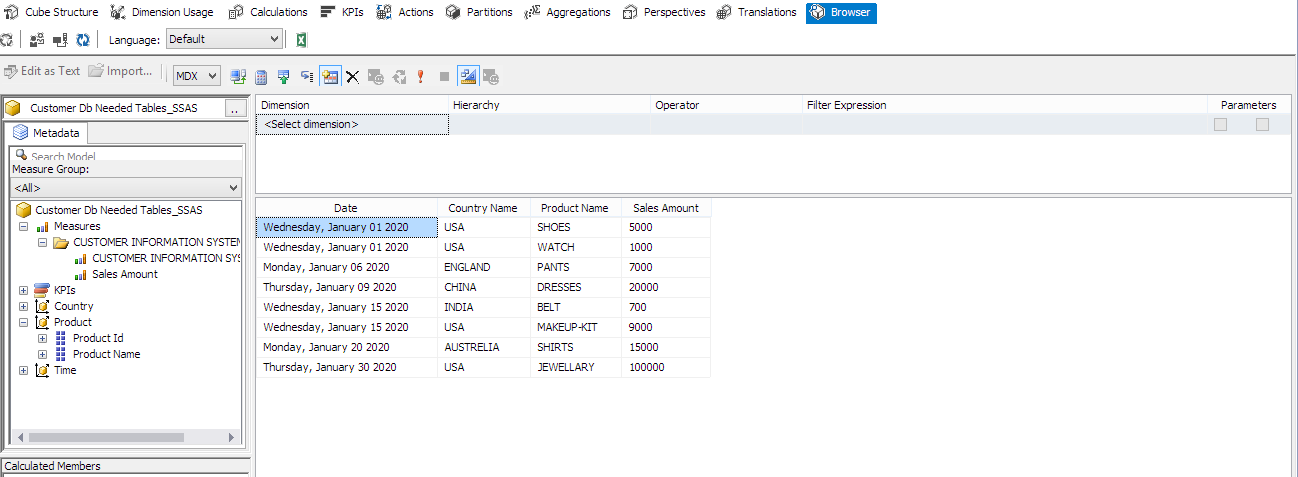
**Fig 5.5.2.5 Customer Package(Data Flow Design)**

**5.5.3. SSAS CUBE**

SSAS extracts the data from Database to create cubes for analysing purpose.



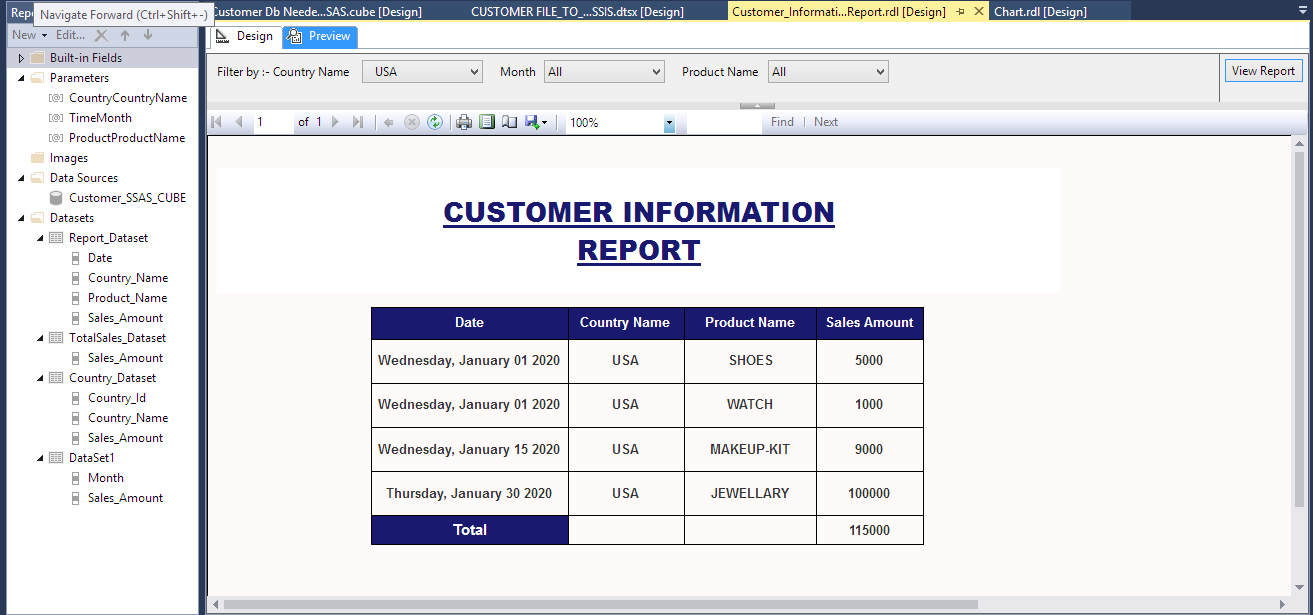
**Fig 5.5.3.1 SSAS CUBE DESIGN**

****

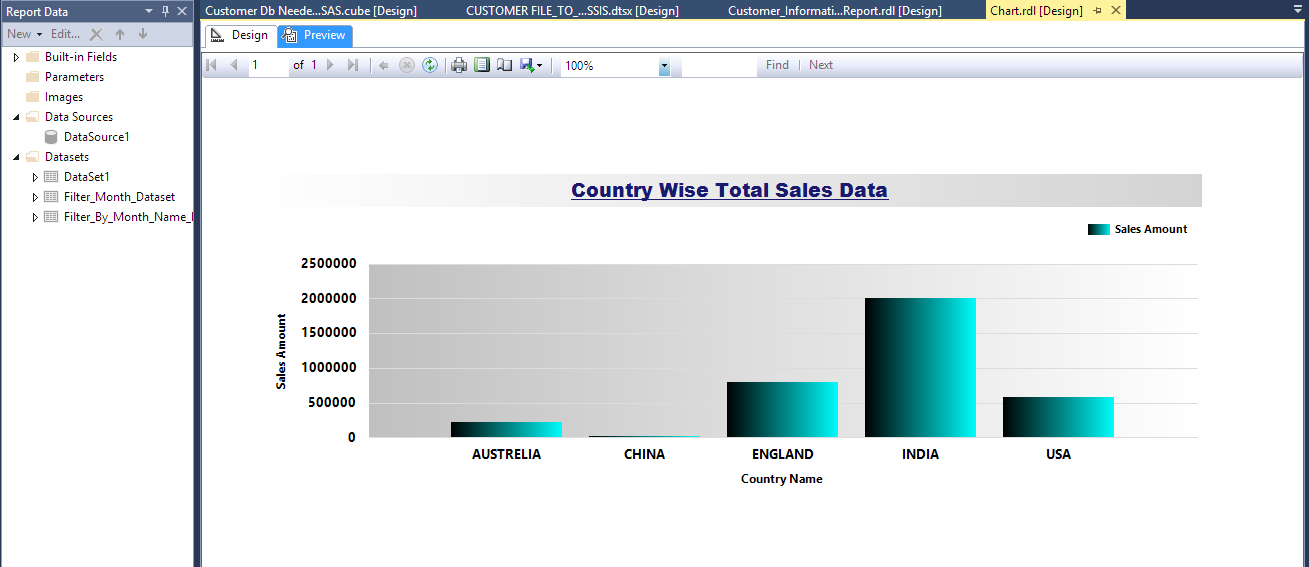
**Fig 5.5.3.2 SSAS CUBE**

**5.5.4. SSRS REPORT**

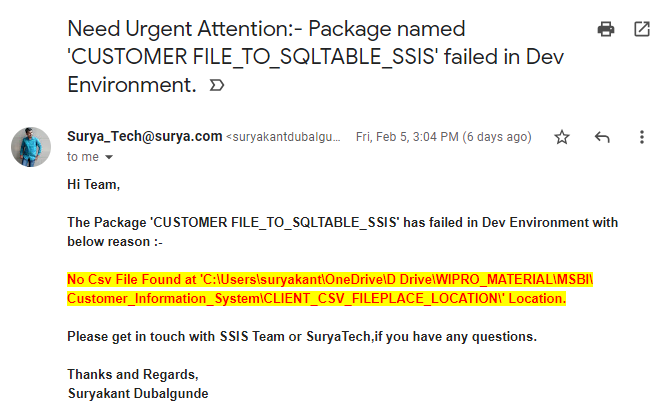
SSRS will be used to create reports.Reports can be displayed in textual format or graphical format.



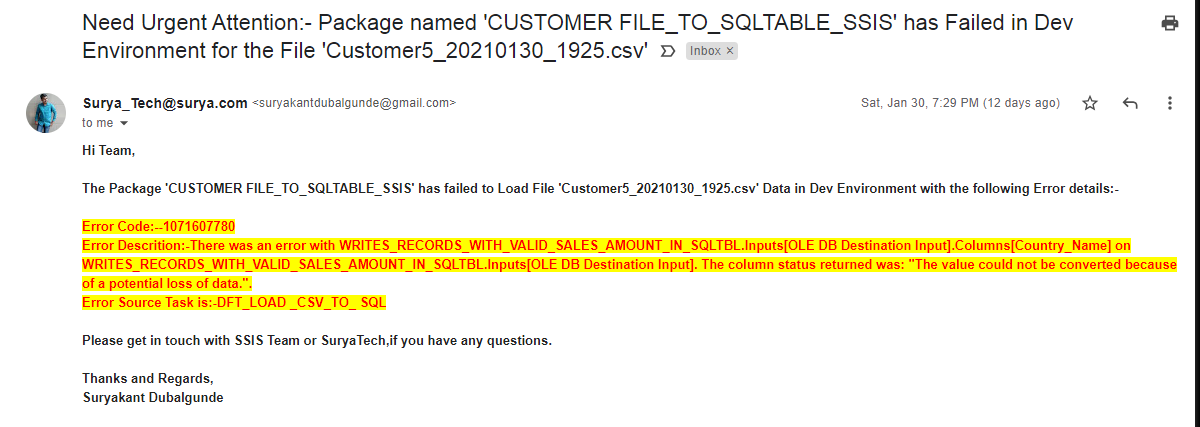
**Fig 5.5.4.1. SSRS in Textual Format**

****

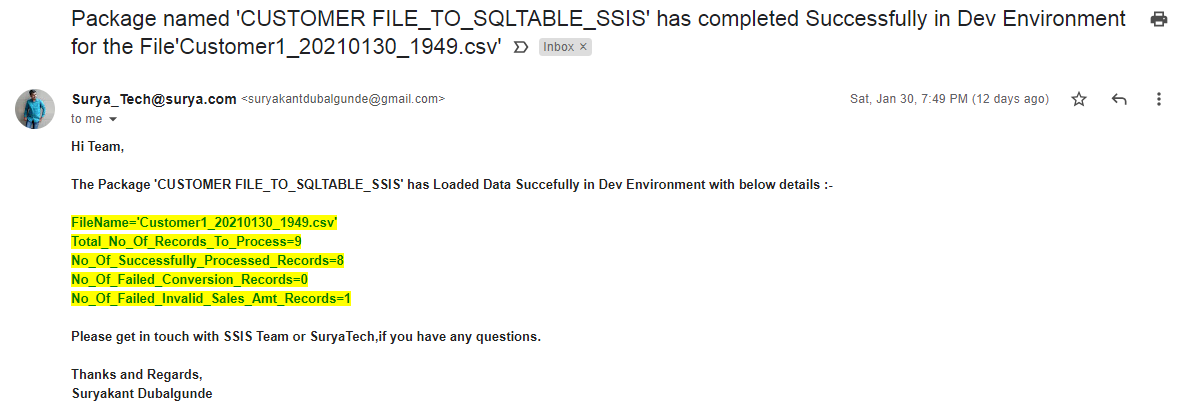
**Fig 5.5.4.2. SSRS in Graphical Format**

**5.5.5. Emails Received**

**Fig 5.5.5.1 No File Found**

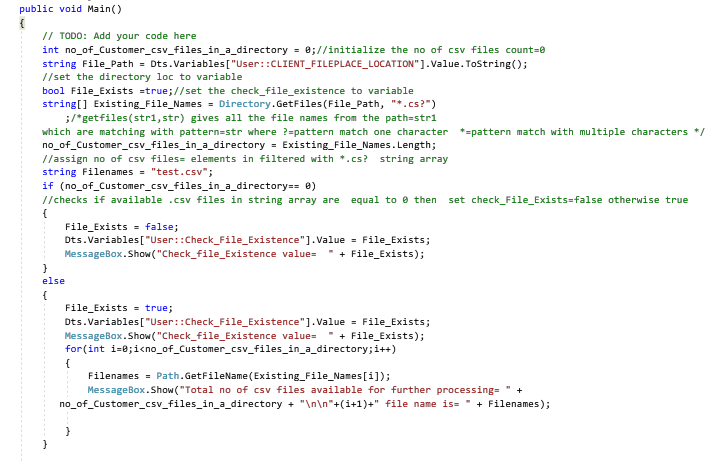
****

**Fig 5.5.5.2 Data Load Fails**

****

**Fig 5.5.5.3 Data Load Success**

**5.6 SAMPLE CODE**

****

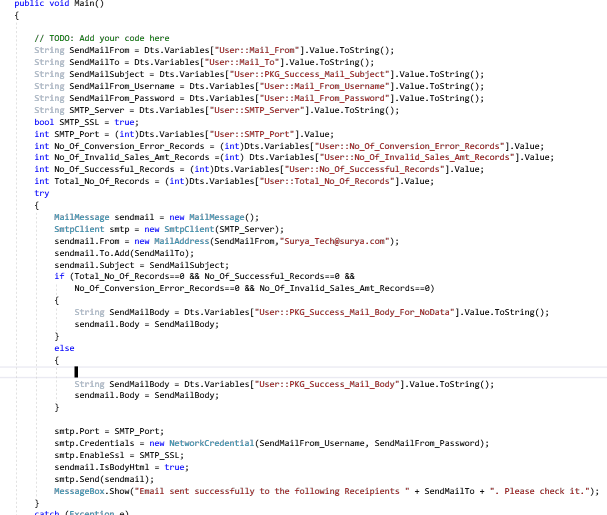
**Fig 5.6.1 Checks File Existence @Location**

****

**Fig 5.6.2. Send Mail if File Not Found**

****

**Fig 5.6.3 Send Mail if Data Load Fails**

****

**Fig 5.6.4 Send Mail if Data Load happens**

# TESTING AND RESULTS

Performed Unit Testing,Integration Testing,System Testing etc. For below scenarios

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Step | Expected Behaviour | Actual Behaviour | Result |
| 1 | The SSIS Package should trigger the mail to client if file not available. | should trigger the mail. | Triggered the mail | PASS |
| 2 | The SSIS Package should append the Processing Datetime to filename(eg:Customer\_20201128\_2106.csv), If files available. | Should append the Processing Datetime to filename | Appended the Processing Datetime to filename | PASS |
| 3 | If file data is corrupted then that corrupted data should not flow in final target table | corrupted data should not flow in final target table | corrupted data didn't flow in final target table | PASS |
| 4 | If file data load fails,then should trigger the mail. | should trigger the mail | Triggered the mail. | PASS |
| 5 | If file data load fails,move that file to error file location | move that file to error file location | Moved that file to error file location | PASS |
| 6 | If file data load is successful,then should trigger the mail. | should trigger the mail. | Triggered the mail. | PASS |
| 7 | If file data load is successful,then should move that file to archive location. | should move that file to archive location | Moved that file to archive location | PASS |
| 8 | If package fails to execute then it should log the exception | should log the exception | Logged the exception | PASS |
| 9 | If package fails to execute then should notify the client by email. | should notify the client by email. | Notified the client by email. | PASS |
| 10 | Once the data is loaded into database,Should be able to create Cube for analysis. | Should be able to create Cube for analysis. | Created the Cube | PASS |
| 11 | Once the data is analysed,should be able to use that analysed data for creating Reports | should be able to use that analysed data for creating Reports | used that analysed data for creating Reports | PASS |

**Table 6. Test Result**

**7. CONCLUSION & FUTURE WORK**

**7.1 CONCLUSION**

This project serves to automate the prevalent traditional tedious and time-wasting methods of providing the Customer information in minimum time. The use of this system will help to reduce the manual work lists the strategies, effective decision-making processes, technologies, etc. and supports the organization to make business operational decisions. The main objective of developing this system is to structure and analyze the historical data within a company smart insights providing scope for rational decision making.CIS plays a pivotal role in the strategic planning of any organizations decision-making process which includes performance progress, quantitative analysis, reporting, data sharing and understanding customer insights. The CIS system involves the use of computing technologies to identify, analyze and forecast or predict customer views. It can help in efficient a quick decision-making process by analyzing the vitality of the various parameters based on the legacy customer data and help the enterprise come up with effective solutions based on the performance indicators.CIS also removes the requirement for manual data entry and interpretations and provides with quick reporting features and data visualizations.he system makes overall business management much easier and more flexible.It offers high security with a different level of authentication.The new system will provide accurate results. The new system will be much better in performance,speed,reliable,flexible as well as efficient compared to the existing one.

**7.2 FUTURE WORK**

* Can be deployed in any other machine's SQL SERVER instance
* Can be Scheduled to run at specific time.
* Can be added more transformations to the data.
* Can be applied more calculations and formulas in analysis .
* Can be applied more Reporting tools.

**8. LIST OF REFERENCES**

1)Wikipedia-https://en.wikipedia.org/wiki/

2)C-Sharp Cornor-https://www.c-sharpcorner.com

3)Udemy-www.udemy.com

4)Youtube-www.youtube.com

5)SqlServerCentral-www.sqlservercentral.com