



The British College
KATHMANDU



Coursework Submission Coversheet
(individual coursework only)

Faculty of Arts, Environment and Technology

LBU Student Id:

FOR CHECKING BY THE STUDENT:

Please ensure all information is complete and correct and attach this form securely to the front of your work before posting it in a coursework collection box.

Award name: BSc (Hons) Computing

Module code:

Module name: Production Project

Module run:

Coursework title:

Due Date:

Module leader: (In LBU)

Module Supervisor: (In TBC)

TURNITIN Checked: YES NO (*please circle*)

Submission date & time: Date: Time:

Total Word Count:

Total Number of Pages (including this front sheet):

In submitting this form with your assignment, you make the following declaration:

I declare, that the coursework submitted is my own work and has not (either in whole or part) been submitted towards the award of any other qualification either at LBU or elsewhere. I have fully attributed/referenced all sources of information used during the completion of my assignment, and I am aware that failure to do so constitutes an assessment offence.

Signed:

Date:

You are strongly advised to retain a second copy of your work in case of any query about the assignment.

FOR COMPLETION BY THE FACULTY:

Teacher's Signature: _____

Date: _____



Production Project

360 Degree Customer Management through
Datawarehouse

SAMPLE

Dedication

I fully dedicate this production project research to my beloved family and specially to my beloved mother, who always shows me the morals of discipline, honesty and hard-work. I momentarily appreciate the datum that she introduced the highly importance of education in this era and thoroughly monitored the progress of my education. All grandeur and decency to God, for his adequate provision and for engaging the precise individual, by the precise duration through the project.

SAMPLE

ACKNOWLEDGMENT

I would like to thank precious and reliable directions and supervision of my supervisor Dr. Pranita Upadhyaya, the meaningful lectures' astute reproaches and persistent inspiration, Mr. Arun Lal Joshi (Associate Dean, IT & Computer Science), Dr. Mahesh Maharjan and whole instruction and non-instruction staff of The British College.

I would similarly like to recognize the entire staff of Big Mart, especially the senior administration including Information & Communications Technology leader Miss Prabina Poudel, who initiate time during their rush schedule to cooperate with me throughout the collection of Customers sales Data and approved access to the precious data foundations which empowered me derive up with the exploration verities. Likewise, I acknowledge the folks who contributed throughout the system's recital appraisal, who keenly cooperate me.

Similarly, I would like to express gratitude towards my beloved ones, classmates and friends, who always have braced me through the entire progression.

SAMPLE

DEFINITION OF TERMS/ ABBREVIATION AND ACRONYMS

BM- Big Mart

ETL – Extract, Transform and Load

EMM - Entity-Mapping Methodology

UML – Unified Modelling Language

OLTP – Online Transaction Processing

OLAP - Online Analytical Processing

SAMPLE

Abstract

Every business leader wants to gain the behavior of their customer so that they can promote their business in upcoming days. The operative methods and efficient of a data warehouse establish of a source rigorous workflow, creating the vital portion of the System design. To settle the concentrated workflow and to accomplish the data warehouse operative procedures, Extraction Transformation Loading (ETL) methods are used. A methodology deployed for the execution of Big Mart's data warehouse was designated throughout this project. Big Mart Pvt. Ltd. is the Grocery Retail Chain within the Grocery Supermarket Industry of Nepal. The purpose of this project was to deliver the resolution of data from file, i.e. Excel sheet for analyzing and reporting. The valuation of studies and associated inspections aimed by categorizing the key foundations and features of active ETL development on the execution of a data warehouse. The research method employed expressive and case study plans. The collection of data was directed to create the technological state of dealings at Big Mart, specifying the existing deviating sources of data, architecture of System and management of data. Questionnaires managed to staff and corresponding managers were used to collect data. The analysis of data is done using Excel and Oracle App with charts and tables. The process of ETL met the potentials of dealing in generating reports in a flexible and friendly manner.

SAMPLE

Contents

Dedication	3
ACKNOWLEDGMENT	4
DEFINITION OF TERMS/ ABBREVIATION AND ACRONYMS	5
Abstract	6
Chapter 1.....	10
Introduction	10
Running sources of data at Big Mart.....	11
1.2. The Declaration of the Problem	11
1.3. Projected Solution	11
1.4. Objectives	12
1.5. Questions for Research.....	12
1.6. Rationalization	12
Chapter 2.....	13
2.0 Review of Literature	13
2.1 Introduction.....	13
2.2 The Concept of Datawarehouse.....	14
2.3. Modeling of Datawarehouse.....	14
2.4. Datawarehouse flow or process.....	15
2.5. Design of Data Warehouse.....	15
2.6. Reviews and Reports from Referring Firms.....	19
Chapter 3.....	21
3.0 Review of Technology.....	21
SOFTWARE.....	21
Chapter 4.....	26
4.0 Proposed and Applied Methodologies	26
4.1. Proposed and Applied Methodologies	26
Field Visit.....	27
4.2. Product Design.....	28
Fig: Use case diagram for Big Mart Data Warehouse	32
Fig: Entity Relationship Diagram	33
Requirements Analysis	34
System Users and Their Roles.....	34

Chapter 5.....	35
5.0 Product Implementation and Testing.....	35
5.1 Product Implementation.....	35
Reports of Tableau	57
5.3 Product Testing.....	59
Chapter 6.....	60
6.0 Product Evaluation.....	60
6.1 Introduction.....	60
6.2 Exploration Strategy	60
6.3 Targeted Population	60
6.4 Data Collection Approaches.....	61
6.5 Analysis of Data	63
6.6 Consistency and Validity	63
6.7 Potential of the project.....	63
6.8. Limitations of the product	64
Chapter 7.....	66
7.0 Project Management and Evaluation.....	66
7.1 Introduction.....	66
7.2 Value of Project Management.....	66
7.3 Applied Gannt.....	66
7.4 Project Evaluation.....	67
7.5 Field Methodologies.....	68
7.6 Usability.....	68
7.7 Report on several portion after final product and project implemented.....	71
7.8 Limitations of the Research	75
Chapter 8.....	76
8.0 Summary and Conclusion.....	76
8.1 Summary	76
8.2 Conclusion	76
8.3 Recommendations.....	77
8.4 Future Research and Practices	78
"	78
References.....	79

APPENDICES	81
Appendix I: Initial Proposal	81
360 Degree Customer Management through Data Warehouse	81
Appendix II: Data Collection Authorization Letter	86
Appendix III: Field Visit Letter from Big Mart	87
Appendix IV: Meeting Records with Supervisor	88

SAMPLE

Chapter 1

Introduction

1.1 Background

We are living in the Customer age. Customer Age means controlling the customer relationship of any organization which has changed. If we compare the business from past, we can see lots of changes. The story of Big Mart Pvt. Ltd. starts in 2000 with only one store at City Centre. Since 2000 Big Mart Pvt. Ltd. now has expanded with over 32 retail small-medium-large format stores across Kathmandu. The vision of Big Mart is to put Nepal into the world of super-marketing map. At present Big Mart uses Ginesys software application with oracle database to do all of its tasks. The main server is in Head office, while the store servers are located at each respective stores. Data between head office and stores are exchanged using intranet. The development of 360 Degree Customer Management will help Big Mart creating data warehouse and dashboard.

Reliable and quality data is essential as the key foundation for any organization to endure competitive in today's era of business. Big Mart Pvt. Ltd in Nepal is implementing emergent and erudite technology to confirm strategy and planned decision making is accomplished. Making a Data Warehouse produce imposing decision support. Data warehouse can be denoted to a user expressive high-level span as an assortment of methodology and tool. The goal is to presenting the fact's base outcomes on time.

For effective execution of a data warehouse, the tool for ETL is used to accomplish the following three main jobs:

1. Extraction of data from deviating data foundations.
2. Transition towards staging of the data environment aimed at cleaning and transformation, and
3. Loading of the data towards the data warehouse.

The complications faced to rectifying, transforming and loading information towards the data warehouse is accomplished through ETL tools. Big Mart flat file holds huge records of the month, a bill date, customer name, customer mobile number, item code, category1, a bill qty sum, net amount sum. The sales data of January and February was nearly collected around 1 million records. This production project aimed at detecting a prescribed and methodical representation framework for producing Extract, Transform and Load data warehouse functionalities. The mapping of data from various branches of Big Mart within a satisfactory format to discern the aim of loading data towards the targeted data warehouse. The abstract version was projected that was used to model the different

phases of the ETL. The abstract version revised the strategy of the ETL process, adaptation and mapping among the traits of the foundations of data collected with their consistent to the focused data warehouse. This research used Big Mart flat file as the case study which leads to the development of ETL process design in its own data warehouse execution for the persistence of cultivating decision-making abilities which finally result in enhancing Big Mart's data warehouse services and satisfaction of the organization.

Running sources of data at Big Mart

Since its beginning, Big Mart has been using Ginesys software application with an oracle database to do all of its tasks. Data of Big Mart required to make knowledgeable verdicts are inside fragmented systems not appropriately integrated and not abundantly operated. Producing and enhancing information by these sources of data has been existing and time consuming. For the current deliberate strategy, Big Mart has a strong ingenuity to boost its business processes, incorporating and implementation by approaching with the data warehouse.

1.2. The Declaration of the Problem

To expand productivity in service and improvement in the making of decision, Big Mart has decisive job to integrate entirely flows of work and methods. Almost all data using Ginesys software is placed in the Excel sheet. All the Point of Sales (POS) of Big Mart information is not transparently accessible across the range to support the different business units of Big Mart which access reliable and precise data towards modest edge. The current system and Ginesys system under execution lack the facility of regaining data and storing centrally by fast, effective and exterior databases without analyzing the transactional of databases. To counter the problematic and accomplish the precise goal of approaching up through a data warehouse to merge all sources of data, an ETL process requires to be used to perform the Extract Transform and Load. Anyhow, the implication of the research process with maximum research in this discipline has been directed due to the implication and compactness. Acceptance of standard design is missing -to signify the states of ETL. This is important to present the Extract Transform and Load process in standard proper method.

1.3. Projected Solution

This production project offers to provide an inclusive Extract Transform and Load process which will guide drawing of information from varied data stores towards the planned data

warehouse in a well-applied manner, which will pivot on the mapping and demonstrating of the entire ETL process. The projected solution utilizes an ETL process model for the implementation of the data warehouse, which will be appropriate for assimilating entire fragmented data in Big Mart for producing reports and analytical goals. Enhancement of previous design was used to make the projected model.

1.4. Objectives

- To do research on existing flat file of Customer Sales.
- To research more on various types of product promotions, schemes and other factors that decrease the number of unprofitable customers.
- To design a system that analyses customer behavior and produce results for customer promotion.
- To explore the approach engaged by a user analyst in coming along with the ETL model.

1.5. Questions for Research

1. What are the possible technical and functional necessities perilous based on current literature and research?
2. What space identified in current research will be maintained in approaching an Extract Transform and Load abstract design?
3. Will the execution of the data warehouse involves data mining for strategy of decision?

1.6. Rationalization

In today's era of business, data warehousing is very vital which consists of correct, secure and convenient information. The execution of an inclusive ETL method will support cleaning and retrieving the data from varied foundations inside the data warehouse in an applied manner, which will pivot on the mapping and demonstrating of the entire ETL process. In literature, this type of model could not be acknowledged.

Chapter 2

2.0 Review of Literature

2.1 Introduction

Producing high level management along with useful information, past and present is the crucial role of the Datawarehouse system. For making a knowledgeable decision without making any interruption of the day to day operations done through OLTP systems, Big Mart has accomplished competition and burden, calling for the interdependence of Datawarehouse as the solution to participate and preserve its valuable customers, facing the chance of perspective details about the market ([Han and Kamber, 2008](#)). The usage of Datawarehouse design has assisted to explore and keep the huge quantity of records from different data ([Alenazi et al. 2013](#)). Data Warehouse is motivated to be a dynamic mechanism to integration of information which is essential as a strategic method for administration to make verdicts through the usage of storing zone of precise data cross worth chain ("[Architecture for Real Time Analytical Data Integration and Data](#)", "2014").

The collection of mechanism aimed at allowing the verdict producer to produce improved and quicker decisions. Data warehouse varies from the functional database as they are oriented on subjects, integrated, invariant of time, non-volatile, abridged, huge, unnormalized which perform OLAP. There are three layers consisting data warehouse architecture. They are ETL, data sources and Datawarehouse primary (Simitis). The activities of data warehouse consist data sourcing, staging of data (ETL) and growth of decision support system ([Senapat and Kumar 2015](#)). The Extract, Transform and Load process is necessary for data warehouse hands as this process is responsible for loading of whole data into the warehouse. Without the process of ETL, data warehouse cannot occur, which contributes towards the supervisory of whole data. ETL tools consist of heterogeneously filtering the data and loading into the data warehouse which makes the tool specific in manner. At an initial stage the phase of extraction makes all the data inside a single arrangement which is further taken for the process of transformation. The filtration of data eliminates the duplicate records, detects variation and error source within the data. From any source the data needed in ETL processes can be obtained, it can be obtained from POS transactions, website embedded databases, customer relationship management tools or any flat files. Though Extract, the Transform and Load process is very essential. It contains research as it is tough and lacks a basic model for presenting the Extract, the Transform and Load process that shows the information from various sources of data which can be beneficial for the loading data into the focused data-warehouse ([Mawilmadia 2012](#)).

For effective execution of a data warehouse, the tool for ETL is used to accomplish the following three main jobs:

1. Extraction of data from deviating data foundations.
2. A proliferation towards staging of the data environment aimed at cleaning and transformation, and
3. Loading of the data towards the data warehouse.

This Production Project research will stab to invent a proper illustration for catching the process of ETL which chart the external data from various sources of data for keeping in perfect set-up, which will lead to load within the targeted data warehouse.

2.2 The Concept of Datawarehouse

The main goal of data warehouse technology is to give proper structure to the whole data in a very appropriate method to give access to the data and practice in a very effective and competent way ([Mawilmada 2012](#)). As defined by [Kerri, Quintana, Alberta, Cotline, Chaves, Juanito and Youngman's \(2002\)](#), the data warehouse system is responsible for the reliability of information. The combination of tools like querying tools, tools for reporting and tools for analysis offers changes to get the rationality of information. The key point of a data warehouse is to establish the congregation of an inclusive series of data and keep it within the data repository ([Kerkri et al., 2002](#)). At present, data warehouse plays a very vital role in the community of business at a small or large scale.

2.3. Modeling of Datawarehouse

The modeling of data warehouse is usually separated in three main methods: analysis of requirements, logical design and the Extract, Transform and Load process. All the raw data directed to the data warehouse has various set-ups and all the process requires for each of these arrangements need precise methods in order to put up inside the definite mapping. Management of character, safe transactions, parsing of text and alphanumerical signals processing require to reach towards the data warehouse along with its whole requirements. The definition of data warehouse is an important portion in designing the data warehouse. Within various transaction systems, necessary data is needed to design the data warehouse. All the important information based on joint dimension is collected from various sources, then filtered and transferred to meet the final requirements of the users. The data warehouse process has basically four stages.

- 1) Stage of Database: Extract, Transform and Load Stage
- 2) Stage of Warehouse

- 3) Stage of tools
- 4) Stage of Interface

2.4. Datawarehouse flow or process

The foremost process inside the data warehouse must be stated before constructing the architecture of the data warehouse. The distinctive drift within data warehouse for handling the data warehouse location can be mentioned as below ([Ananhor and Murrey, 2012](#)):

- i) Backup of data and Archiving process

Failure of software and hardware can occur any time. These types of sudden accident can hamper the organization in very huge amount. If the system is collapsed all the data will be destroyed. Backing up of the data warehouse is must require portion as it will ensure that the data can be achieved after failure of the system too. Archiving of the data is process in which all the old data are removed from the system to a such way which can be restored anytime when required. To recover data in Oracle database redo archived is used.

- ii) Process of Query and Management

“The process of system which handle all the queries and boost them up by giving direction to all queries towards the most effective data source is known as process of query and management” ([Anehory and Muray, 2012](#)). Process of query and management always confirms the execution of a query is flowing in right manner which won't let the system fail. While loading the data into data warehouse query management is not operated. Beside the loading process, it runs whole time to ensure the quality of the system.

2.5. Design of Data warehouse

Conceptual Design:

The very first stage of business modelling is conceptual design. This design explains the whole entire business. All the main entities are included in this design. All the relationship of the system including entities are specified in this portion. Unfortunately, it lacks the representation of an attribute of the entities.

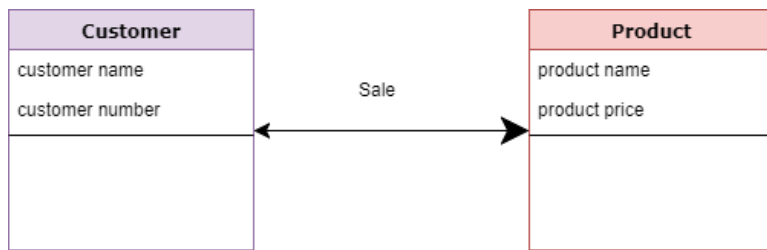


Fig 1: Conceptual design of data model

Logical Design:

Logical design sums of some extra information towards the conceptual model. The main goal of the logical design is to give informatic design to the whole system. All the possible entity and information are shown in this. Logical design is also known as blue print of the database.

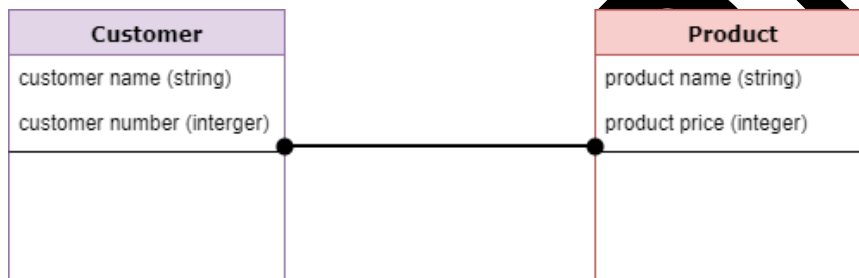


Fig 2: Logical Design of data model.

Dimension Design

Dimension design or dimension modeling is the way of rationally modeling an OLAP architecture or data warehouse system. This design is also known for the easiest architecture for highly performance access ([Kimball, 1998](#)).

2.6. Architecture of Data Warehouse

There are basically three standard types of data warehouse. They can be illustrated as below down:

1) Data warehouse using Basic

This architecture is so least in use as its main aim is to lessen the quantity of data which is being stored. From various sources data, is collected and processed within the data warehouse ([Lane, 2003](#)).

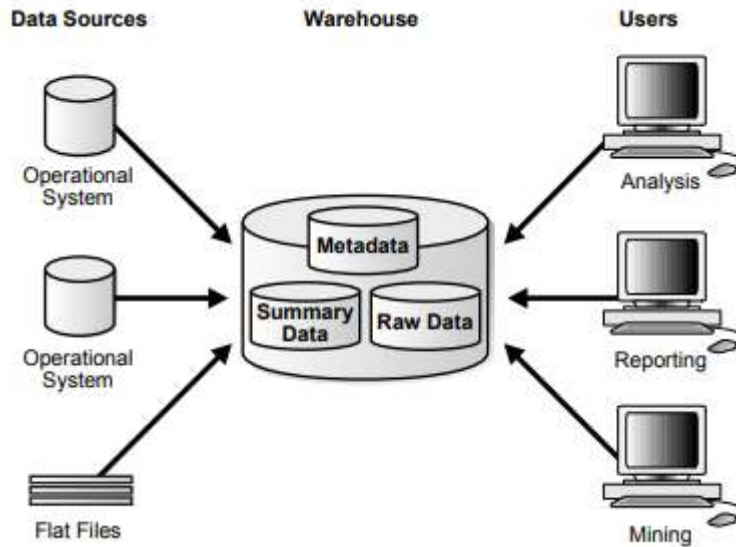


Fig 3: Data Warehouse Basic Architecture

Above figure demonstrate that data collected for data warehouse are collected from working systems. For real time business and transaction this architecture only carries OLTP tasks. The basic downside of this architecture is that this fails in differentiating transaction and analytical reports ([Golfarelli and Rizzi, 2010](#)).

2) Data warehouse using Staging Architecture

Basically, it contains four stages of data flow. The basic stages of staging architecture are mentioned below:

- i) Data Sources
- ii) Area of staging
- iii) Layer of data warehouse
- iv) Layer of Analysis

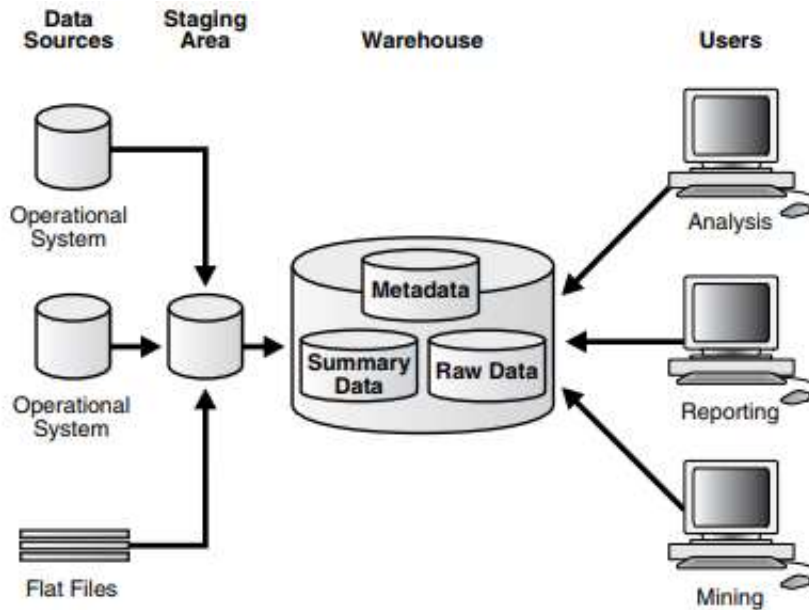


Fig 4: Staging Area architecture of Data Warehouse

In this architecture, the initial stage is the staging of data sources. For data source stage, operational data are the basic sources. For staging area, the operational data are extracted, filtered and almost processed for load inside the staging area. The staging area always helps out in making simplification while preparing for the summary data and also helps out in handling data warehouse ([Khachan, n.d.](#)).

3) Data warehouse using staging area and Data Marts

By comparing all three architecture of data warehouse this is the most appropriate and used architecture. This architecture match with one of the above architectures but it has some addition layers i.e. layers of data mart.

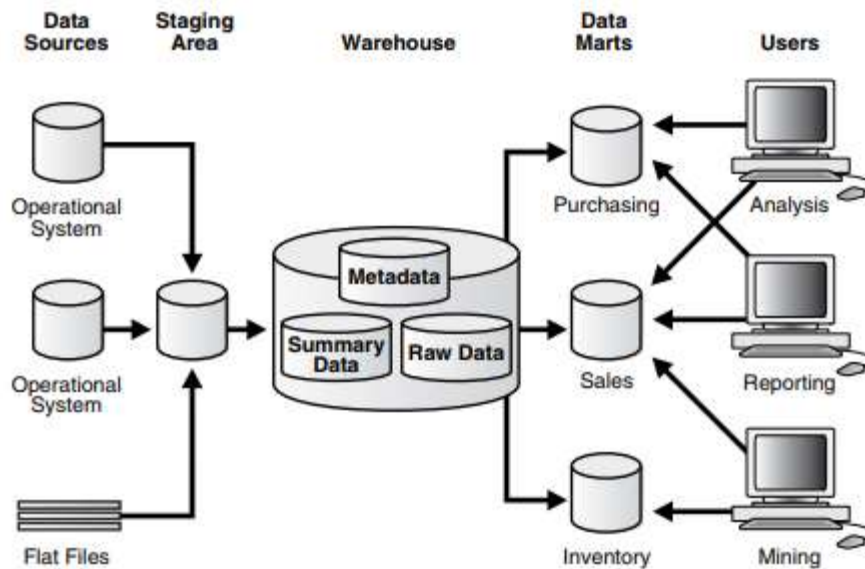


Figure 5: Staging Area and data marts architecture of data warehouse

2.6. Reviews and Reports from Referring Firms

As stated in del ([Hoyo-Barbolla and Less \(2003, p.45\)](#), it is applicable in health care sector as well. There is also mentioned "Within modest weather, if health care establishments are to retain the patients perspective and handling information regarding about patient it is necessary and establishment realizes that this is so vital to access practical and sensible data". Additionally, "mixing data from various sources and renovating all these data inside the reliable information, competitive knowledge is obtained (deli Hoyo-Barbolla and Lees 2003). There are three main database merchants which virtually provides Extract, Transform and Load solution at free of cost. They are listed as Oracle data warehouse with Oracle database builder [Oracle 04], Microsoft database with overall Microsoft Data Transformation Services [Microsoft 04] and IBM database with IBM Data Warehouse System [IBM 04]. In spite of these three main database merchants, over the world, the foremost merchant in database filed some were Informatica power center [Info 0 4] and Ascentiale's data stage collections [Asce 04 , Asce 04 A](Further collaged with IBM for recommendations of all process carried out in Extract, Transform and Load].The feasible case study for the propose of upcoming technology trails and forecasts which enrolls the all over combination of ETL including adapters of XML ,Enterprises Applications and Integration (EAI) few tools like MQI Series II ,customization of several quality due data tools, the further move of parallel processing towards workflows of ETL.

The above-mentioned review was braced from third case study [Gart 04, 05], where the publishers renowned the deterioration in certificate revenue for complete ETL tools, especially because it lacks information technology spending and the proper advent of

extract, transform and load process from very moderated database and business intelligence merchants. The study of Gartner mentioned about the specific roles of all those three basic databases merchants (Oracle all version, Microsoft all version and IBM all version) and further piercing that they would yield the part of Extract, Transform and Load marketplace throughout the database management system in built solution. Through the end of 2007, more than 51 % of newly established data warehouse distributions had used tools for ETL which has been distributed by main Database management system merchants (Oracle, Microsoft and IBM) (0.8 approximately) [Fie 05].

SAMPLE

Chapter 3

3.0 Review of Technology

In an overall valuation viable tool for ETL were responsible for the execution of the flow of data inside the data warehouse. Many commercial tools are of generally two types. They are of engine generating based and code generating based. The earlier implicit describes that entirely data needs to go through an engine for the process of transformation. Furthermore, the appliances took over the extract and load process which goes for final ETL process.

In today's era, we can find out lots of tools and technology which is taken to do 'Techniques' for the successful result of any project or research. For appropriate selection of tools and technique for 360 Degree Customer Management, an immense research was carried out. For the outcome of better product, various technology was implemented. All the tools and technology implemented in the product are described below down:

SOFTWARE:

- i) Oracle Database version 11g EE (Express Edition):

In present world, Oracle database is termed to be the world's most famous and reliable database management system. Basically, for Online Transaction Process (OLTP) and development of data warehouse, it is used ([DB Engine, 2017](#)). As Oracle Database 11g Express Edition is freely available and is web based, I selected this for development of data warehouse using ETL process. Oracle Database 11g is based on fully released of Oracle Release 2 code. All the CRUD operations were done using Oracle database. All the activities of importing and exporting of the database was carried out within this software. As my topic was 360 Degree Customer Management, all the roles and privileges were easily carried out using this database. All queries of SQL statement along with report was done using Oracle database. From extraction of data by OLTP system to keeping all those data inside staging table, filtering and transforming the data into filtered table and finally saving all those reliable data inside the data warehouse all work was done inside this platform only.

ORACLE Oracle Database XE 11.2

Home	Storage	Sessions	Parameters	Application Express
------	---------	----------	------------	---------------------

At initial stage, MYSQL database along with php myadmin tool was preferred for the development of the data warehouse. PhpMyAdmin cannot create application, so Oracle Apex was selected for creation of application. We don't need to install oracle Apex as it comes pre-installed with Oracle Database 11g Express Edition. For the report generation and development of data warehouse both Oracle database and Oracle Apex was used.

ii) Oracle Application Express

Apex is the platform which allows to create and maintain the entire data through database. This tool was selected to create application where all the records of customer and sales was stored within the dashboard. For the report of business intelligence, this is most appropriate tools over any other software.

ORACLE Oracle Database XE 11.2 Username: SYS@123456 Logged

Home	Storage	Sessions	Parameters	Application Express
------	---------	----------	------------	---------------------

Home Parameters

Q: Go Actions:

Parameter	Value	Type	Description	Dynamic
audit_file_dest	C:\ORACLE\APP\ORACLE\ADMIN\ADUMP	String	Directory in which auditing files are to reside.	No
compatible	11.2.0.3.0	String	Database will be completely compatible with this software version.	No
control_files	C:\ORACLE\APP\ORACLE\ORADATA\CONTROL.DBF	String	control file names list.	No
db_name	XE	String	database name specified in CREATE DATABASE.	No
db_recovery_file_dest	C:\oracledb\app\oradata\fast_recovery_area	String	default database recovery file location	Yes
db_recovery_file_dest_size	1073741824	Big Integer	database recovery file size limit	Yes
diagnostic_dest	C:\ORACLE\APP\ORACLE	String	diagnostic base directory	Yes
dispatchers	(PROTOCOL=TCP) (SERVICE=XE)	String	specifications of dispatchers	Yes
job_queue_processes	4	Integer	maximum number of job queue slave processes	Yes
memory_target	1073741824	Big Integer	target size of Oracle SGA and PGA memory	Yes
open_cursors	300	Integer	max # cursors per session	Yes
remote_login_passwordfile	EXCLUSIVE	String	password file usage parameter	No
sessions	172	Integer	user and system sessions	No
shared_servers	4	Integer	number of shared servers to start up	Yes
spfile	C:\ORACLE\APP\ORACLE\PRODUCT\11.2.0\SERVER\ORSPFILE.XE.ORA	String	server parameter file	Yes
undo_management	AUTO	String	undo space management mode	No
undo_tablespace	UNDOTBS1	String	undo tablespace	Yes

1 - 17

iii) Command Prompt (cmd):

Command prompt also known as cmd is basically a line interpreter within the system. All the command was executed within cmd for various purpose. Through command prompt the oracle database was accessed using correct query. Oracle database has pre-installed SQL * plus which is accessed by command and all the SQL, PL/SQL commands are executed over there.

Various command for database administration, creating/updating and viewing of objects it was used.

```
Command Prompt - sqlplus

Microsoft Windows [Version 10.0.17134.471]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\User>sqlplus

SQL*Plus: Release 11.2.0.1.0 Production on Sun Dec 23 01:52:11 2018

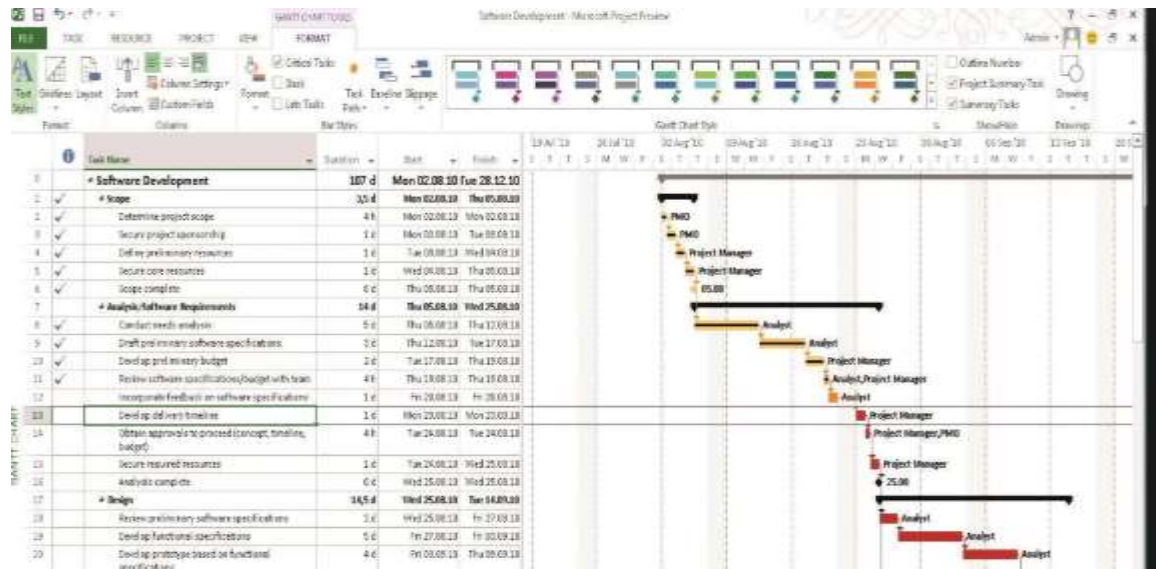
Copyright (c) 1982, 2010, Oracle. All rights reserved.
```

- iv) QSEE Super Lite: QSEE- Super Lite is most useful and easy software for data modeling. For the representation of business activities in graphically blue print it is used. All the useful diagrams for the product was done using QSEE Super Lite. For creating of ER Diagram, star schema and Unified Modeling Language use case diagram was done using this software.



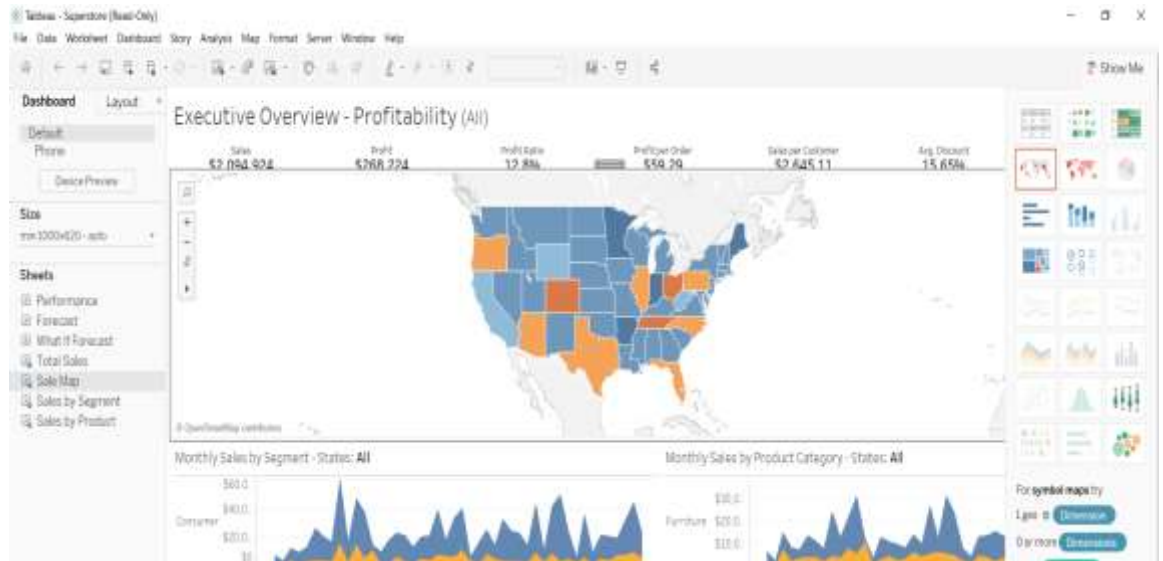
- v) Microsoft project:

Microsoft project is the product of Microsoft which is project management system. It helps for proper planning and managing for the project. For creation of Gantt Chart, it was used. Gantt chart helped to follow the timing by showing timeframe given for each portion of the project.



vi) Tableau:

Tableau is one of the best software for visualization of data and Business Intelligence (BI) which is developed by the company Tableau itself. Gartner has listed tableau as best business intelligence tool in the last five years. All the bar charts and useful report for BSN Mart was done through this software. The oracle database was connected to tableau for visualization of data from data warehouse for generating useful promotional and development report for BSN Mart.



HARDWARE:

Asus K54U model PC was used for the development of this project. As development of data warehouse requires RAM more than 2 GB, I upgraded my RAM to 6 GB so that system failure won't take place. NVIDIA GeForce Graphics with intel i5 processor (7th Generation) was pre-available in the system. Windows 10 was selected as operating system.

Chapter 4

4.0 Proposed and Applied Methodologies

4.1. Proposed and Applied Methodologies

Following a precise methodology for the development of any product is so important in today's technological world. Basically, methodology helps to plan, design and control the process of development. It makes easier to describe everyone's responsibilities and gives structure to the entire project. While development of the product, it is necessary to have proper meeting with supervisor, module leader for better outcome of product. Hence development of data warehouse was the product of this module, the movement of data inside the data warehouse never stops. There are continuous deviations in the project of data warehouse. After knowing these all agile methodology was selected over waterfall methodology.



Fig: Agile Methodology cycle

Agile methodology and water fall methodology both follow similar Software Development Life Cycle (SDLC) process. Waterfall methodology follow sequential order for the development of software whereas agile methodology follows iterative way. In agile methodology all the phases of SDLC are executed in repetition. This is the main point to select agile methodology over water fall We can make changes in any portion when it requires. While developing the data warehouse we need to go back to the previous steps to execute data and load into the system. At the end of iteration process, testing is done.

This confirms satisfaction of end user and also supports to debug the system when require. This makes agile methodology perfect selection over waterfall methodology. There are different types of customer of Big Mart whose shopping nature changes seasonally. The obligation variations addressed by the regular customer are discoursed in the meeting and then after executed in iteration manner. This is also the plus point to select agile methodology.

If we compare both methodology we can find out, waterfall methodology lacks enrollment of much customer. The enrollment of the customer is typically only in initial phases to deliver necessary document and in the final phases to display the outcome. The feedback is taken at the end of final product. When the enrollment of customer increases in Big Mart, the understanding and visualizing nature of data also increases. Similarly, there may be requirement for the fresh sources of data to be inserted. While inserting the rows for new data, the system gets updated time and again. Additionally, the variations in the data process will be essential if there are many variations in reports of the customer. Customers are generally indistinct about the needs of business unless the data is stored in the data warehouse. There are very high probabilities of variation in business ethics and logics. Moreover, in the long-term business, the object of data warehouse is typically faltering as data warehouse remains long duration in the organization.

Field Visit

The purpose of this project was to deliver the resolution of data from file, i.e. Excel sheet for analyzing and reporting. The evaluation studies and associated inspections aimed by categorizing the key foundations and features of active ETL development on the execution of a data warehouse. The research method employed expressive and case study nature. The collection of data was directed to create the technological state of dealing at Big Mart, displaying the existing deviating sources of data, architecture of System and management of data. Questionnaires managed to staff and corresponding manager to collect data. The analysis of data is done using Excel and Oracle Apex with charts and tables. The process of ETL met the potentials of dealing in generating reports in a flexible and friendly manner.

Since its beginning, Big Mart has been using Ginesys software application with an oracle database to do all of its tasks. Data of Big Mart required to make knowledgeable verdicts are inside fragmented systems not appropriately integrated and not abundantly operated. Producing and enhancing information by these sources of data has been exacting and time consuming. For the current deliberate strategy, Big Mart has a strategic ingenuity to boost its business processes, incorporation and implementation by approaching with the data warehouse.

The sales data of January and February was nearly collected around 1 million records. This production project aimed at detecting a prescribed and methodical representation.

n framework for producing Extract, Transform and Load data warehouse functionalities. The mapping of data from various branches of Big Mart within a satisfactory format to discern the aim of loading data towards the targeted data warehouse. The abstract version was projected that was used to model the different phases of the ETL. The abstract version revised the strategy of the ETL process, adaptation and mapping among the traits of the foundations of data collected with their consistent to the focused data warehouse. This research used Big Mart flat file as the case study which leads to the development of ETL process design in its own data warehouse execution for the persistence of cultivating decision-making abilities which finally result in enhancing Big Mart's data warehouse services and satisfaction of the organization.

4.2. Product Design

- i) Identify the appropriate data sources:

The initial step for designing the data warehouse is to identify the appropriate data and analysis of the data. For specific sets of data warehouse, concept of big mart fact table was discovered where all the customer sales record was shown from various data source.

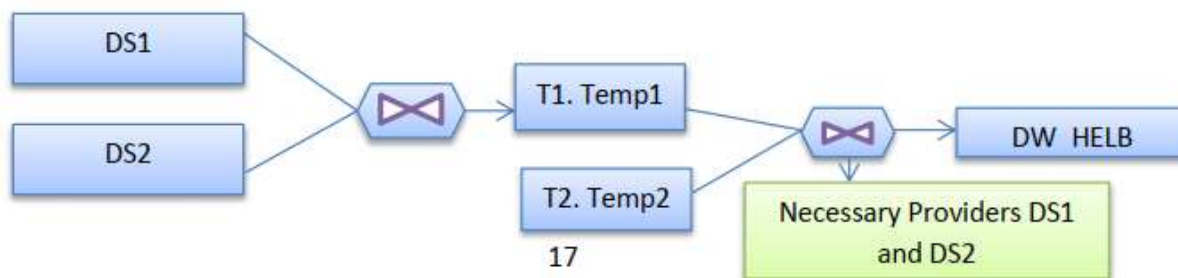



Diagram Key	
DS:	Data source (Ds1-Datasource 1, Ds2 – Data source 2)
T1:	Temporary tables (T1- Temporary table1, T2, temporary table 2)
	Union or combination
DW:	target Data warehouse

ii)

In EMM, the modeling of the ETL process is explained which gives mapping of data from sources towards the targeted schema of the data warehouse. The fig. 6 conceptually defines all simple model for an ETL process. The shown layer represents the data stores which involves in the whole process. Usually, data suppliers are relational database and file. The required data from all sources are extracted by extraction process, which deliver either whole polaroid or discrepancies of the data sources. Then, these data are circulated to the area of data staging. Inside data staging, the data are transformed and filtered before they are being loaded to the data warehouse. The data warehouse is shown in the correct portion of the data store layer and contains the targeted data stores. The loading of the change data warehouse is achieved from the loading actions shown on the above portion of the figure.

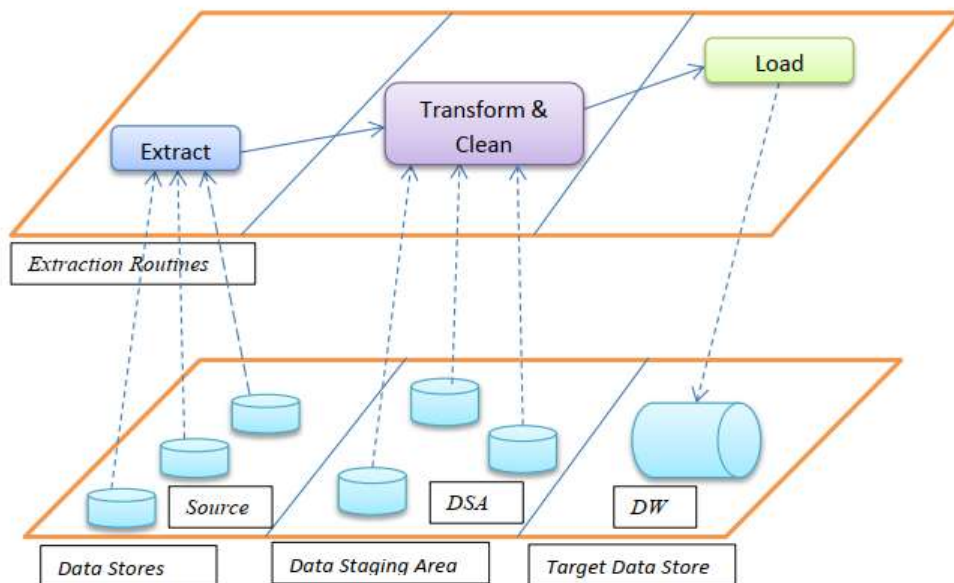


Figure: Entity Mapping Design

For a proposed methodology, the design is kept in step wise method. Further that lead to reach basic target i.e., the relationship of the attribute. These phases established the methodology for the strategy of the conceptual portion of the inclusive ETL process. These all could be described as follows:

- classification of the suitable data stores.
- the applications and access customer for the complicated data stores.
- the characteristics designing between the suppliers and the customers and
- interpreting the whole diagram with runtime constrictions.

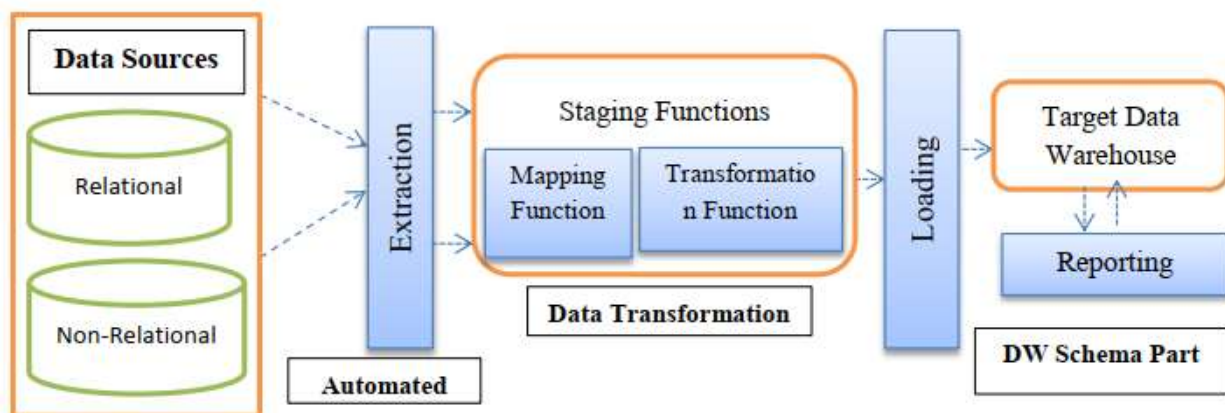


Fig: Relational and non-relational Entity Mapping Framework

Inside the data source portion: the contributed data sources were strained. The sources of data could be either structured or non-structured. Inside structured source; the joined databases and their joined table & characters were used unswervingly as the base source, and in case of un-structured source; an alteration step was used. Another step was to change the un-structured source into structured source. From the view of design, there was single alteration paradigm that could renovate un-structured foundation into structured database. After viewing from implementation way, each and every type of un-structured foundation has its own alteration segment which was known as wrapper. Wrappers are generally dedicated program that inevitably extract all data from various sources of data.

Wrapper distinctive duty can be explained as:

- i) Collecting data from an isolated source
- ii) Penetrating for, identifying and mining specified data
- iii) Storing the data in well-structured arrangement to enhance further guidance.

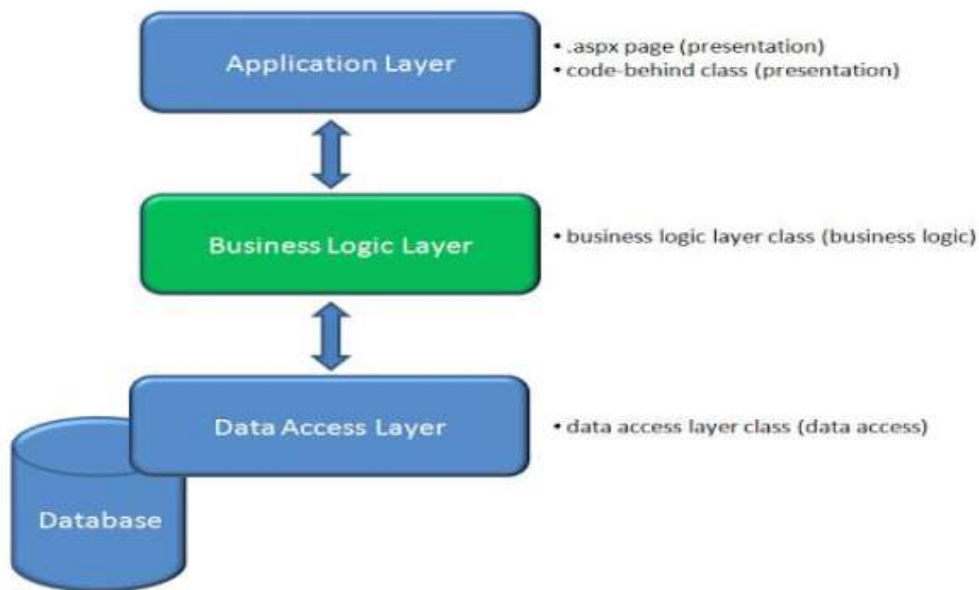


Fig: The n-tier system architecture

System Requirements:

a. Functional requirements

ID	Requirement
FR-1	Ability to extract data from heterogeneous sources.
FR-2	Data cleansing ability for detecting and correcting (or removing) corrupt or inaccurate records from mined table.
FR-3	Map entities using entity modelling.
FR-4	Load data into the target data warehouse schema for data warehousing activities
FR-5	Decision support capabilities by the data warehouse
FR-6	Reporting functionality of the system

b. Non-functional requirements

ID	Requirement
NFR-1	The system should guard against accidental deletion and erroneous update of stored data.
NFR-2	The system should provide for user authentication.
NFR-3	The system should check and verify that entered data is in the appropriate format
NFR-4	The system should have adequate understandability, testability, maintainability, and reusability.

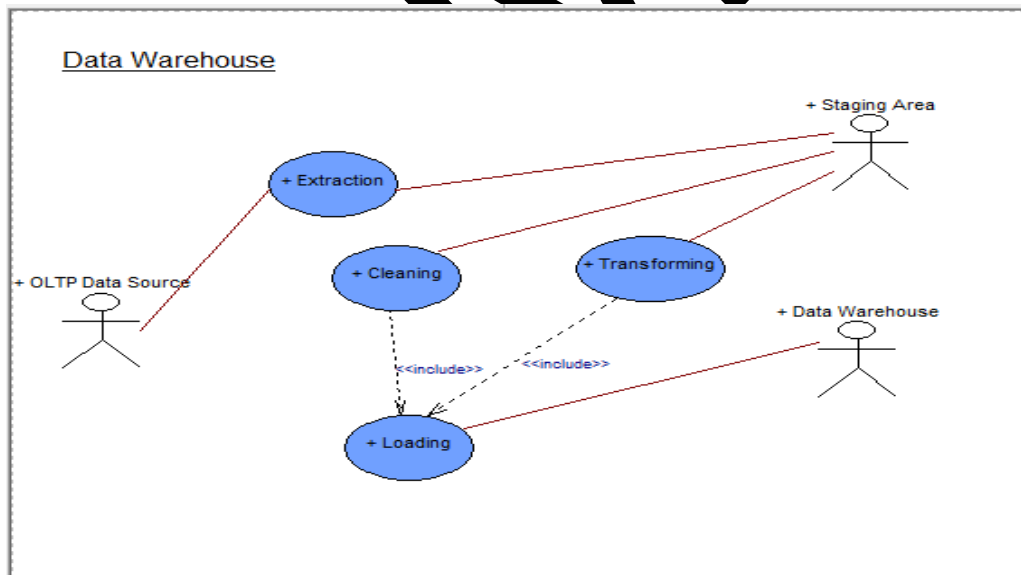


Fig: Use case diagram for Big Mart Data Warehouse

Entity Relationship Diagram

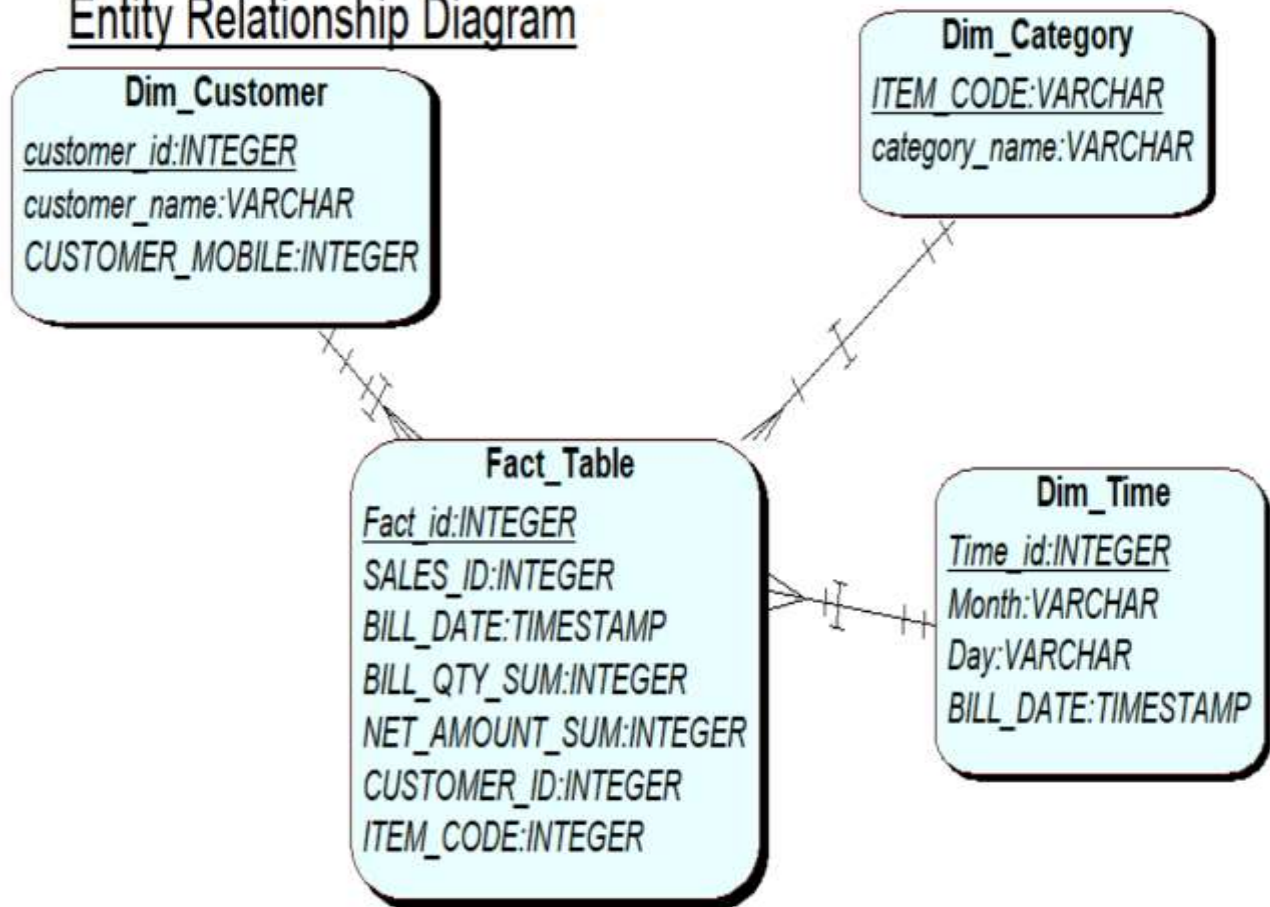


Fig: Entity Relationship Diagram

Requirements Analysis

This section describes the different users of the system and their roles, and subsequently presents system requirements both functional and non-functional), based on user needs and roles.

System Users and Their Roles

a) **System Administrator:** This User has administrative rights in the System. The roles of the administrator include:

- Creating User accounts; includes setting system privileges to users.
- Managing user accounts (editing and deleting user accounts).

b) **Database Administrator:** This User is involved in the mining of data using the ETL tool to the target data warehouse. The specific functions of the system manager include:

- Data extraction from the various databases and files
- Transforming extracted data from different sources in preparation for loading.; including data cleansing and
- Loading data to the target data warehouse.

c) **Management:** This User is in management and is interested in decision-making activity by facilitation for the data warehouse reporting services. The specific functions of the manager include:

- Reporting activities
- Decision making activities based on the data residing in data warehouse
- Strategic functions

Chapter 5

5.0 Product Implementation and Testing

5.1 Product Implementation

Inside the product implementation, entirely collected data and intended information are curved into a definite product. Development of “360 Degree Customer Management through data warehouse” was the product required for this project. After generation of data warehouse, the useful reports were generated using Oracle Apex and business Intelligence tool, Tableau. Following Kimball’s data warehouse architecture along with staging area concept and fact and dimension process, the data ware house created successfully. To develop the data warehouse, the customer sales data of Big Mart was taken. The data was massive in volume which was further filtered for the ETL process and making data warehouse for the Big Mart. The complications allied to rectifying, transforming and loading information towards the data warehouse is accomplished through ETL tools. Big Mart file holds huge records of the month, a bill date, customer name, customer mobile number, item code, category1, a bill qty sum, net amount sum. The sales data of January and February was nearly collected around 1 million records. This presentation primarily aimed at detecting a prescribed and methodical representation framework for producing Extract, Transform and Load data warehouse functionalities. The mapping of data from various branches of Big Mart within a satisfactory format to discern the form of loading data towards the targeted data warehouse.

For effective execution of a data warehouse, the tool for ETL is used to accomplish the following three main steps:

1. Extraction of data from deviating data foundations.
2. Data proliferation towards purging of the data environment aimed at cleaning and transformation, and
3. Loading of the data towards the data warehouse.

All the filtered data from recognized OLTP system were mined towards the staging stage following ETL process. All the required data were filtered and transformed from the staging area. All the filtered data were loaded towards the clean tables and the dirty data or bad data were loaded towards the error table. This process was done using ETL flow. Star schema was mapped for the creation of fact tables and dimension table to develop data warehouse. Using ETL process all the clean tables were denormalized. After denormalization process the clean data were loaded in the clean tables. For the creation of application for 360 Degree Customer Management through data warehouse, Oracle Application Express was used. The application contains report and dashboard as Big Mart new data warehouse and application. The further

required reports were generated using Tableau. All the evidence for implementing of the product are given below down along with screenshots:

a) Process of Source Data

The initial stage was to insert the source data of Big Mart into the oracle database. The scripts were sub divided into three scripts file as the volume was massive as a result the Oracle database 11 g Express edition doesn't support that much volume

Oracle Application Express interface showing the results of a SQL script execution. The script is named "dim and fact tables" and its status is "Complete". The view is set to "Summary" and shows 15 rows. The table below lists the executed statements, their elapsed time, the feedback message, and the number of rows affected.

Number	Elapsed	Statement	Feedback	Rows
1	2.01	drop table fact_table cascade constraints	Table dropped.	0
2	0.06	drop table Dim_Customer cascade constraints	Table dropped.	0
3	0.02	drop table Dim_Category cascade constraints	Table dropped.	0
4	0.03	drop table Dim_Time cascade constraints	Table dropped.	0
5	0.30	CREATE TABLE Fact_Table(Fact_id INTEGER NOT NULL,	Table created.	0
6	0.06	drop sequence fact_loadd	Sequence dropped.	0
7	0.02	create sequence fact_loadd start with 1 increment by 1	Sequence created.	0
8	0.01	CREATE TABLE Dim_Customer(customer_id INTEGER NOT NU	Table created.	0
9	0.02	CREATE TABLE Dim_Category(ITEM_CODE VARCHAR(30) NOT	Table created.	0
10	0.04	CREATE TABLE Dim_Time(Time_id INTEGER NOT NULL,	Table created.	0
11	0.01	drop sequence seq_time	Sequence dropped.	0
12	0.01	create sequence seq_time start with 1 increment by 1	Sequence created.	0
13	0.27	ALTER TABLE Fact_Table ADD CONSTRAINT fk1_Fact_Table_to_Dim_	Table altered.	0
14	0.01	ALTER TABLE Fact_Table ADD CONSTRAINT fk2_Fact_Table_to_Dim_	Table altered.	0
15	0.01	ALTER TABLE Fact_Table ADD CONSTRAINT fk3_Fact_Table_to_Dim_	Table altered.	0

row(s) 1 - 15 of 15

Download

Statements Processed 15
Successful 15
With Errors 0

Workspace: PP_FINAL User: PP_FINAL

Fig: Creation of table with 0 error

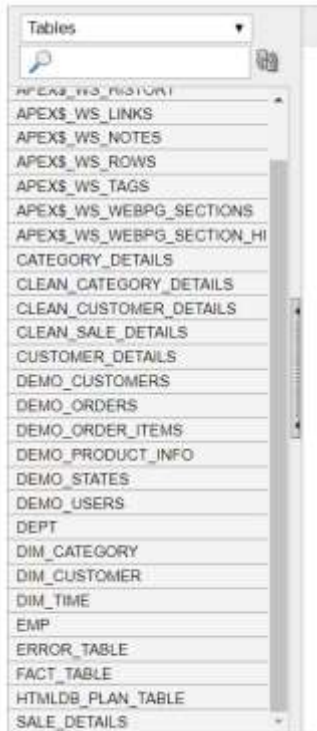


















Fig: Tables created for source data

b) Process for ETL



Fig: Query to create Customer_detail table for Extraction Process

c)

CUSTOMER_DETAILS			
EDIT	CUSTOMER_ID	CUSTOMER_NAME	CUSTOMER_MOBILE
	1	Ashristya Bhandaryi	1234567890
	2	Dbinod Kumar Thakura	1234567890
	3	Rpuja Shakyi	1234567890
	4	Apragya Shresthai	1234567890
	5	Aasha Upadhyayi	1234567890
	6	Syubraj Dhakali	1234567890
	7	Aseema Thokeri	1234567890
	8	Ateja Shree Sharmai	1234567890
	9	Ariku Tamangi	1234567890
	10	Amala Rajbhandarii	1234567890
	11	AT.N Bhattraai	1234567890
	12	Apushpa Bhattaraii	1234567890
	13	Aajay Prakash Chaudharii	1234567890
	14	Taabita Upadhayayei	1234567890
	15	Bmriti Upadhyau	1234567890
row(s) 1 - 15 of 417 			

[Download](#)

Fig: Data populated inside customer_details as extraction process

NOTE: All the first name, last name and contact number of customers were altered for the privacy issues of Customer.

d)

The screenshot shows the SQL Developer interface with the 'TABLES' pane on the left and the 'SQL' pane on the right. The 'CATEGORY_DETAILS' table is selected in the left pane. The SQL pane contains the following code:

```
CREATE TABLE "CATEGORY_DETAILS"
(
  "ITEM_CODE" VARCHAR2(30),
  "CATEGORY_NAME" VARCHAR2(1000),
  CONSTRAINT "CATEGORY_DETAILS_PK" PRIMARY KEY ("ITEM_CODE") ENABLE
);

CREATE OR REPLACE TRIGGER "bi_CATEGORY_DETAILS"
before insert on "CATEGORY_DETAILS"
for each row
begin
  if :new."ITEM_CODE" is null then
    select "CATEGORY_DETAILS_SEQ".nextval into :new."ITEM_CODE" from dual;
  end if;
end;

/

ALTER TRIGGER "bi_CATEGORY_DETAILS" ENABLE;
```

Fig: Query to create Category_detail for Extraction Process

e)

The screenshot shows the SQL Developer interface with the 'TABLES' pane on the left and the 'SQL' pane on the right. The 'CATEGORY_DETAILS' table is selected in the left pane. The SQL pane contains the following code:

EDIT	ITEM_CODE	CATEGORY_NAME
<input checked="" type="checkbox"/>	BM19348	BUBBLES WS 1001 WHISTLE
<input checked="" type="checkbox"/>	BM67616	NESCAFE LATTE CUP COFFEE 25G
<input checked="" type="checkbox"/>	BM49066	NESCAFE CLASSIC COFFEE PKT 50G
<input checked="" type="checkbox"/>	BM21093	SHRESTHA CHILI POWDER (KHURSANI) 500G
<input checked="" type="checkbox"/>	BM58398	BIG CHOICE RAHAR NON POLISH PREMIUM DAAL/KG
<input checked="" type="checkbox"/>	BM8973	SWEET POTATO
<input checked="" type="checkbox"/>	BM27823	QUEEN SWEET RAINBOW 55G
<input checked="" type="checkbox"/>	BM37775	UNIQUE BROKEN WALNUT 100G
<input checked="" type="checkbox"/>	BM19357	PARTY POPPER MEDIUM
<input checked="" type="checkbox"/>	BM46663	GILLETTE 2
<input checked="" type="checkbox"/>	BM47441	BAGRRY'S WHITE OATS REFIL PACK 500G
<input checked="" type="checkbox"/>	BM56602	BRITANNIA GOOD DAY RICH CASHEW COOKIES 200G
<input checked="" type="checkbox"/>	BM6590	COLGATE MAXFRESH RED GEL 150G
<input checked="" type="checkbox"/>	BM89430	SUJALGUM BANANA CHEWINGUM 13.5G
<input checked="" type="checkbox"/>	BM69421	BRITANNIA GOOD DAY BUTTER COOKIES 68GM*17GM

row(s) 1 - 15 of more than 500

Fig: Data populated inside category_details as extraction process

f)

```

SALE_DETAILS

Table Data Indexes Model Constraints Grants Statistics UI Defaults Triggers Dependencies SQL

CREATE TABLE "SALE_DETAILS"
(
  "SALES_ID" NUMBER,
  "BILL_DATE" DATE,
  "CUSTOMER_NAME" VARCHAR2(255),
  "ITEM_CODE" VARCHAR2(30),
  "CATEGORY_NAME" VARCHAR2(4000),
  "BILL_QTY_SUM" NUMBER,
  "NET_AMOUNT_SUM" NUMBER,
  "CUSTOMER_ID" NUMBER,
  CONSTRAINT "SALE_DETAILS_PK" PRIMARY KEY ("SALES_ID") ENABLE
) ;ALTER TABLE "SALE_DETAILS" ADD CONSTRAINT "FK_ITEM" FOREIGN KEY ("ITEM_CODE")
REFERENCES "CATEGORY_DETAILS" ("ITEM_CODE") ENABLE;

CREATE OR REPLACE TRIGGER "bi_SALE_DETAILS"
before insert on "SALE_DETAILS"
for each row
begin
  if :new."SALES_ID" is null then
    select "SALE_DETAILS_SEQ".nextval into :new."SALES_ID" from dual;
  end if;
end;

/
ALTER TRIGGER "bi_SALE_DETAILS" ENABLE;

```

Fig: Query to create sale_details table for Extraction Process

g)

sales report

Sales Id	Bill Date	Customer Name	Category Name	Bill Qty Sum	Net Amount Sum
121	02-JAN-18	Kohit Kumar Pokharely	ORBIT CHEWINGUM MIXED FRUIT 4.4GM	1	10
122	19-JAN-18	Kohit Kumar Pokharely	GULLON SUGAR FREE DIGESTIVE 400GM	1	292
123	31-JAN-18	Karu Baraly	PVC CAP SMALL	2	26
124	10-JAN-18	Ksha Pandey	NESTLE NESCAFE COFFEE & MILK HAZELNUT 180ML	1	50
125	22-JAN-18	Kabi Raj Thapy	BIG CHOICE REGULAR MAAS NON POLISH DAAL/KG	1	94
126	28-JAN-18	Kabi Raj Thapy	UNIQUE CASHEWNUTS 100G	1	202
127	21-JAN-18	Krusha Joshy	DDC KANCHAN CHEESE 500G	1	520
128	14-JAN-18	Kayusha Khedkay	PALPASA BROWN SESAME BALL 90G	2	84
129	12-JAN-18	Kditi Hirawaly	PASEO TOILET 2 ROLL 3 PLY	1	174
130	22-JAN-18	Kditi Hirawaly	REAL CRANBERRY NECTAR 1L	1	220
131	22-JAN-18	Kditi Hirawaly	HIBIS ULTRA THIN NORMAL 20 PADS	1	326
132	30-JAN-18	Kradeep Thakury	HALDIRAMS BHUJIA SEV 350G	1	175
133	13-JAN-18	Kantosh Sharmay	CLEANSING WIPES 80PCS	1	124
134	13-JAN-18	Kantosh Sharmay	IMPERIAL LEATHER CLASSIC SOAP 115G	2	100
135	20-JAN-18	Kantosh Sharmay	MAMY POKO PANTS BABY DIAPER S42	1	798

Previous row(s) 121 - 135 of more than 500 Next

Fig: Data populated inside sale_details as extraction process

NOTE: All the first name and last name of customer were altered for the privacy issues of Customer.

h)

CLEAN_CUSTOMER_DETAILS									
Table	Data	Indexes	Model	Constraints	Grants	Statistics	UI Defaults	Triggers	Dependencies
SQL									
<pre>CREATE TABLE "CLEAN_CUSTOMER_DETAILS" ("CUSTOMER_ID" NUMBER NOT NULL ENABLE, "CUSTOMER_NAME" VARCHAR2(255) NOT NULL ENABLE, "CUSTOMER_MOBILE" NUMBER, CONSTRAINT "PK_CLEAN_CUST" PRIMARY KEY ("CUSTOMER_ID") ENABLE);</pre>									

Fig : Query to create clean_customer_details

i)

CLEAN_CUSTOMER_DETAILS									
Table	Data	Indexes	Model	Constraints	Grants	Statistics	UI Defaults	Triggers	Dependencies
SQL									
<div>Query</div> <div>Count Rows</div> <div>Insert Row</div>									
EDIT	CUSTOMER_ID	CUSTOMER_NAME	CUSTOMER_MOBILE						
	1	Ashristy Bhandaryi	1234567890						
	2	Ainod Kumar Thakuri	1234567890						
	3	Auja Shakyi	1234567890						
	4	Aragya Shresthai	1234567890						
	5	Asha Upadhyai	1234567890						
	6	Aubraj Dhakali	1234567890						
	7	Aeema Thokeri	1234567890						
	8	Aeja Shree Sharmai	1234567890						
	9	Aiku Tamangi	1234567890						
	10	Aala Rajbhandar	1234567890						
	11	AT.N Bhattra	1234567890						
	12	Aushpa Bhattarai	1234567890						
	13	Aay Prakash Chaudhari	1234567890						
	14	Abita Upadhayayei	1234567890						
	15	Amriti Upadhyai	1234567890						

Fig: Data populated inside clean_customer_details

NOTE: All the first name, last name and contact number of customer were altered for the privacy issues of Customer

j)
















CLEAN_CATEGORY_DETAILS									
Table	Data	Indexes	Model	Constraints	Grants	Statistics	UI Defaults	Triggers	Dependencies
									SQL
<pre>CREATE TABLE "CLEAN_CATEGORY_DETAILS" ("ITEM_CODE" VARCHAR2(30) NOT NULL ENABLE, "CATEGORY_NAME" VARCHAR2(1000) NOT NULL ENABLE, CONSTRAINT "PK_CLEAN_CAT" PRIMARY KEY ("ITEM_CODE") ENABLE);</pre>									

Fig: Query to create clean_category Details

k)

CLEAN_CATEGORY_DETAILS

Query | Count Rows | Insert Row

EDIT	ITEM_CODE	CATEGORY_NAME
	BM66275	AMRIT SUNFLOWER OIL 1L
	BM29175	APPLE FUJI
	BM5271	PRASUMA BUFF SAUSAGE 500G
	BM49706	NESTLE LACTOGEN 4 400GM
	BM58400	BIG CHOICE MAAS NON POLISH PREMIUM DAAL/KG
	BM8920	TOMATO SMALL
	BM11469	NANGLO BROWN BREAD 500G
	BM37774	UNIQUE PISATACHIOS 100G
	BM68484	ORBIT CHEWINGUM MIXED FRUIT 4.4GM
	BM20923	GULLON SUGAR FREE DIGESTIVE 400GM
	BM62022	PVC CAP SMALL
	BM66751	NESTLE NESCAFE COFFEE & MILK HAZELNUT 180ML
	BM65347	BIG CHOICE REGULAR MAAS NON POLISH DAAL/KG
	BM37772	UNIQUE CASHEWNUTS 100G
	BM16752	DDC KANCHAN CHEESE 500G
row(s) 1 - 15 of more than 500 >		

Download

Fig: Data populated in Clean_category_details table

l)

CLEAN_SALE_DETAILS									
Table	Data	Indexes	Model	Constraints	Grants	Statistics	UI Defaults	Triggers	Dependencies
SQL									
<pre>CREATE TABLE "CLEAN_SALE_DETAILS" ("SALES_ID" NUMBER NOT NULL ENABLE, "BILL_DATE" DATE, "BILL_QTY_SUM" NUMBER, "NET_AMOUNT_SUM" NUMBER, "CUSTOMER_ID" NUMBER, "ITEM_CODE" VARCHAR2(30), CONSTRAINT "PK_CLEAN_SALE" PRIMARY KEY ("SALES_ID") ENABLE); ALTER TABLE "CLEAN_SALE_DETAILS" ADD CONSTRAINT "FK_CLEAN_SALECUS" FOREIGN KEY ("CUSTOMER_ID") REFERENCES "CLEAN_CUSTOMER_DETAILS" ("CUSTOMER_ID") ENABLE; ALTER TABLE "CLEAN_SALE_DETAILS" ADD CONSTRAINT "FK_CLEAN_SALEITEM" FOREIGN KEY ("ITEM_CODE") REFERENCES "CLEAN_CATEGORY_DETAILS" ("ITEM_CODE") ENABLE;</pre>									

Fig: Query to create clean_sale_details table

m)

CLEAN_SALE_DETAILS						
EDIT	SALES_ID	BILL_DATE	BILL_QTY_SUM	NET_AMOUNT_SUM	CUSTOMER_ID	ITEM_CODE
	243	01/22/2018	1	195	151	BM1038
	244	01/02/2018	1	5	154	BM68247
	245	01/10/2018	1	385	155	BM47442
	246	01/28/2018	1	255	155	BM11435
	247	01/13/2018	2	1900	156	BM66101
	248	01/28/2018	1	240	157	BM11141
	249	01/10/2018	5	80	158	BM7211
	250	01/06/2018	1	136	158	BM50989
	251	01/20/2018	1	679	159	BM41950
	252	01/28/2018	1	70	157	BM66332
	253	01/02/2018	4	352	158	BM67273
	254	01/30/2018	1	470	160	BM66110
	255	01/19/2018	2	140	161	BM52015
	256	01/04/2018	1	350	162	BM9023
	257	01/29/2018	1	125	163	BM59746
row(s) 1 - 15 of more than 500						
Download						

Fig: Data Populated in clean_sale_details table

n)

ERROR_TABLE										
Table	Data	Indexes	Model	Constraints	Grants	Statistics	UI Defaults	Triggers	Dependencies	SQL
<pre>CREATE TABLE "ERROR_TABLE" ("ERROR_ID" NUMBER(*,0) NOT NULL ENABLE, "FROM_TABLE" VARCHAR2(100), "ERROR_DESCRIPTION" VARCHAR2(250), CONSTRAINT "PK_ERROR_TABLE" PRIMARY KEY ("ERROR_ID") ENABLE);</pre>										

Fig :Query to create error_table

o)

DIM_CUSTOMER										
Table	Data	Indexes	Model	Constraints	Grants	Statistics	UI Defaults	Triggers	Dependencies	SQL
<pre>CREATE TABLE "DIM_CUSTOMER" ("CUSTOMER_ID" NUMBER(*,0) NOT NULL ENABLE, "CUSTOMER_NAME" VARCHAR2(255) NOT NULL ENABLE, "CUSTOMER_MOBILE" NUMBER(*,0), CONSTRAINT "PK_DIM_CUSTOMER" PRIMARY KEY ("CUSTOMER_ID") ENABLE);</pre>										

Fig: Query to create dim_customer table

p)

DIM_CUSTOMER			
Table	Data	Indexes	Model Constraints Grants Statistics UI Defaults Triggers Dependencies SQL
Query Count Rows Insert Row			
EDIT	CUSTOMER_ID	CUSTOMER_NAME	CUSTOMER_MOBILE
	234	Prajmol Man Baniye	1234567890
	235	Big Movie Customer 3%	1234567890
	236	Prajwal Man Shrestha	1234567890
	237	Suvechhya Joshi	1234567890
	238	Suman Sharma	1234567890
	239	Anjana Dhakal	1234567890
	240	Dinesh Neupane	1234567890
	241	Ashmita Rayamajhi	1234567890
	242	Suraj Ratna Shakya	1234567890
	243	Shrijana Nepal	1234567890
	244	Anju .	1234567890
	245	Jyoti Ghale	1234567890
	246	Sarita Suyash Dhakm	1234567890
	247	Pratima Manandhar	1234567890
	248	Khushal Bhutoria	1234567890

Fig :Data populated in dim_customer table

NOTE: All the first name, last name and contact number of customer were masked for the privacy issues of Customer.

q)

DIM_TIME			
Table	Data	Indexes	Model Constraints Grants Statistics UI Defaults Triggers Dependencies SQL
<pre>CREATE TABLE "DIM_TIME" ("TIME_ID" NUMBER(*,0) NOT NULL ENABLE, "MONTH" VARCHAR2(30), "DAY" VARCHAR2(30), "BILL_DATE" DATE, CONSTRAINT "PK_DIM_TIME" PRIMARY KEY ("TIME_ID") ENABLE);</pre>			

Fig:Query to create dim_time table

r)

<

Fig: Data Populated in dim_Time table

s)

DIM_CATEGORY										
Table	Data	Indexes	Model	Constraints	Grants	Statistics	UI Defaults	Triggers	Dependencies	SQL
<pre>CREATE TABLE "DIM_CATEGORY" ("ITEM_CODE" VARCHAR2(30) NOT NULL ENABLE, "CATEGORY_NAME" VARCHAR2(1000) NOT NULL ENABLE, CONSTRAINT "PK_DIM_CATEGORY" PRIMARY KEY ("ITEM_CODE") ENABLE);</pre>										

Fig: Query to create dim_category table

t)

DIM_CATEGORY		
Table	Data	Indexes Model Constraints Grants Statistics UI Defaults Triggers Dependencies SQL
Query	Count Rows	Insert Row
EDIT	ITEM_CODE	CATEGORY_NAME
	BM64605	NILESH POTATO CHIPS PLAIN 90G
	BM50558	BIG CHOICE SUGAR PLAIN 1KG
	BM50775	NESTLE EVERYDAY DAIRY WHITENER MILK POWDER 400G
	BM64877	INDICA EASY 10M SHAMPOO COLOUR BLACK 25ML
	BM53409	SANTOOR HAND WASH ESSENTIAL OILS 180ML
	BM23872	EVEREST ICE CUBE 1KG
	BM40185	ASHIRBAD CHANA BESAN 500G
	BM63403	KELLOGGS CHOCOS CHOCOLATE 700G
	BM61883	RAMRO SHAHI NAMKEEN COOKIES 200G
	BM4307	DRUK SOY SAUCE 200G
	BM5888	LAXMI CASHEW NUT PLAIN 200G
	BM8571	HULAS JEERA MASINO 5KG
	BM68705	WHISPER ULTRA CLEAN XL 30 PADS
	BM7189	TOKLA TEA GOLD BOX 200G
	BM60646	MANGALDEEP AGRABATTI SILVER 70STICKS

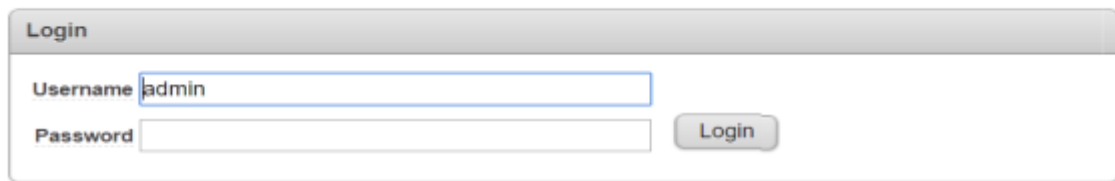
Fig: Data populated in dim_Category table

u)



Fig: Big Mart Final Application

v)



A login form titled "Login" with a title bar. It contains two input fields: "Username" with the text "admin" and "Password" which is empty. A "Login" button is located to the right of the password field.

Fig: Log in Panel for Big Mart App

w)

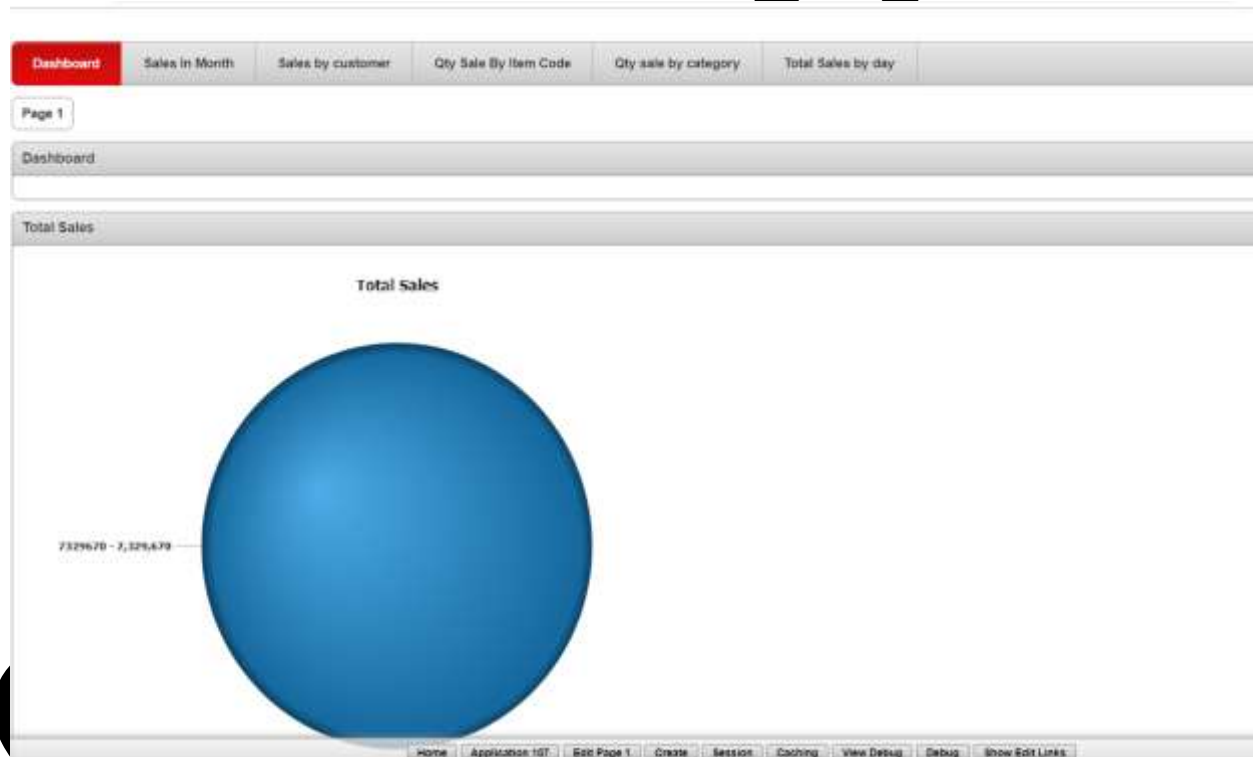


Fig: Main Dashboard for Application

s)

<div><div>Q-</div><div></div><div>Go</div><div>Actions ▾</div></div>		
<u>Customer Id</u>	<u>Customer Name</u>	<u>Customer Mobile</u>
1	AShristy Bhandarye	1234567890
2	ABinod Kumar Thakure	1234567890
3	APuja Shakyae	1234567890
4	APragya Shresthae	1234567890
5	AAsha Upadhyaye	1234567890
6	AYubraj Dhakale	1234567890
7	ASeema Thokere	1234567890
8	ATEja Shree Sharmae	1234567890
9	ARiku Tamange	1234567890
10	AMala Rajbhandarie	1234567890
11	AT.N Bhattraie	1234567890
12	APushpa Bhattaraie	1234567890
13	AJay Prakash Chaudharie	1234567890
14	ABabita Upadhayayee	1234567890
15	ASmriti Upadhyae	1234567890

1 - 15 >

Figure 3. Details in the dashboard part

NOTE: All the first name, last name and contact number of customers were altered for the privacy issues of Customer.

t)

Category Details	
Item Code	Category Name
BM64605	NILESH POTATO CHIPS PLAIN 90G
BM50558	BIG CHOICE SUGAR PLAIN 1KG
BM50775	NESTLE EVERYDAY DAIRY WHITENER MILK POWDER 400G
BM64877	INDICA EASY 10M SHAMPOO COLOUR BLACK 25ML
BM53409	SANTOOR HAND WASH ESSENTIAL OILS 180ML
BM23872	EVEREST ICE CUBE 1KG
BM40185	ASHIRBAD CHANA BESAN 500G
BM63403	KELLOGGS CHOCOS CHOCOLATE 700G
BM61883	RAMRO SHAHI NAMKEEN COOKIES 200G
BM4307	DRUK SOY SAUCE 200G
BM5888	LAXMI CASHEW NUT PLAIN 200G
BM8571	HULAS JEERA MASINO 5KG
BM68705	WHISPER ULTRA CLEAN XL 30 PADS
BM7189	TOKLA TEA GOLD BOX 200G
BM60646	MANGALDEEP AGRABATTI SILVER 70STICKS

row(s) 1 - 15 of more than 500 ▾ Next ➤

Fig: Category Details available in Dashboard

u) Total Sales of month January and February of Big Mart



Fig: Total Sales of January and February

Report was generated after successfully running the system. From above figure we can find out the sales of January are more than sales done in February. So, the mart should more focus on promotional sales rather than doing promotion in any other sector.

v) Sales done by Customer

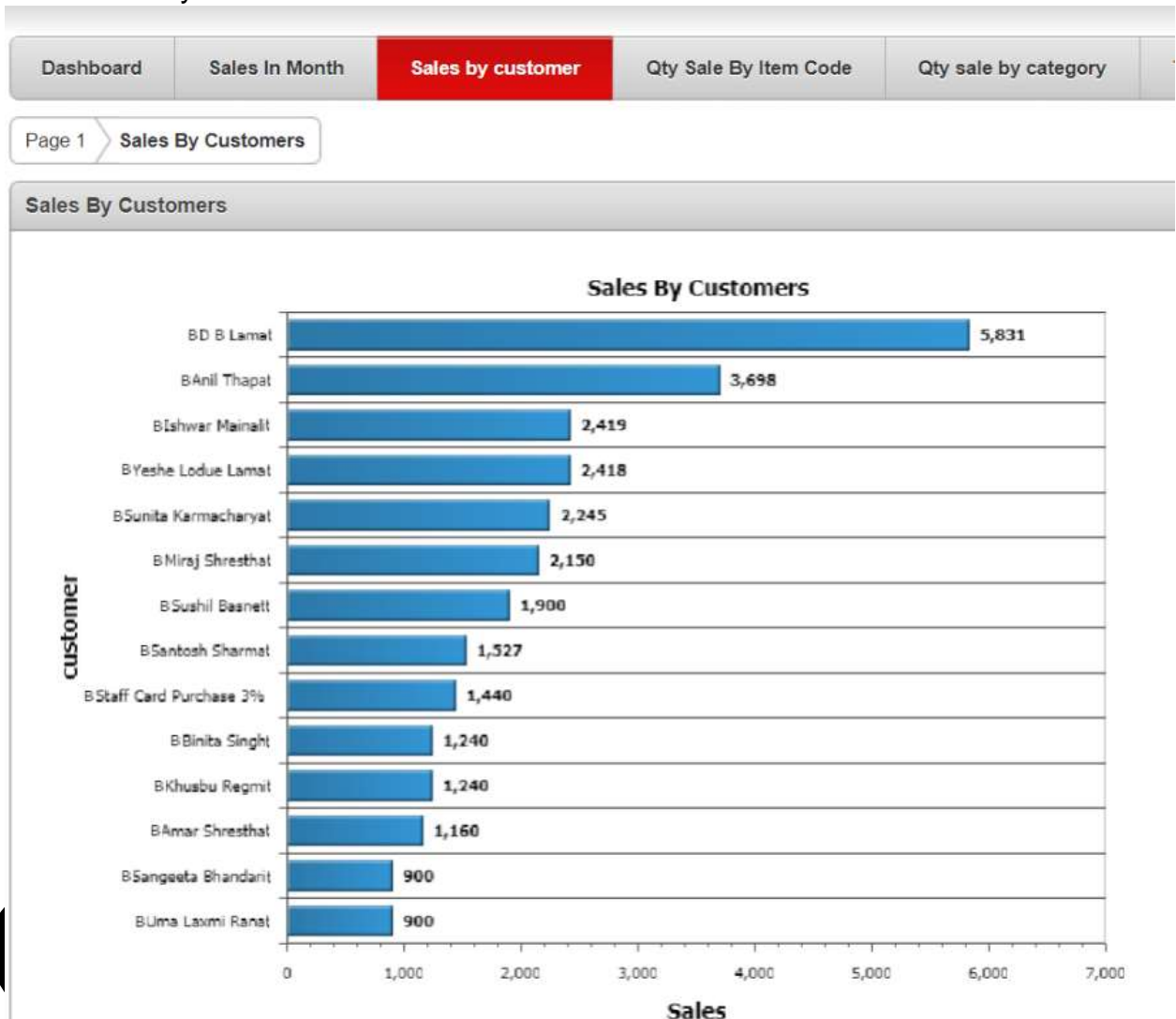


Fig: Sales done by customer

Report was generated after successfully running the system. From above figure we can find the top 14 customer who did more sales in the month of January and February. DB Lama is the best customer and various promotion facility should be provided to him.

NOTE: All the first name, last name and contact number of customers were altered for the privacy issues of Customer

w) Quantity sales done by Item Code

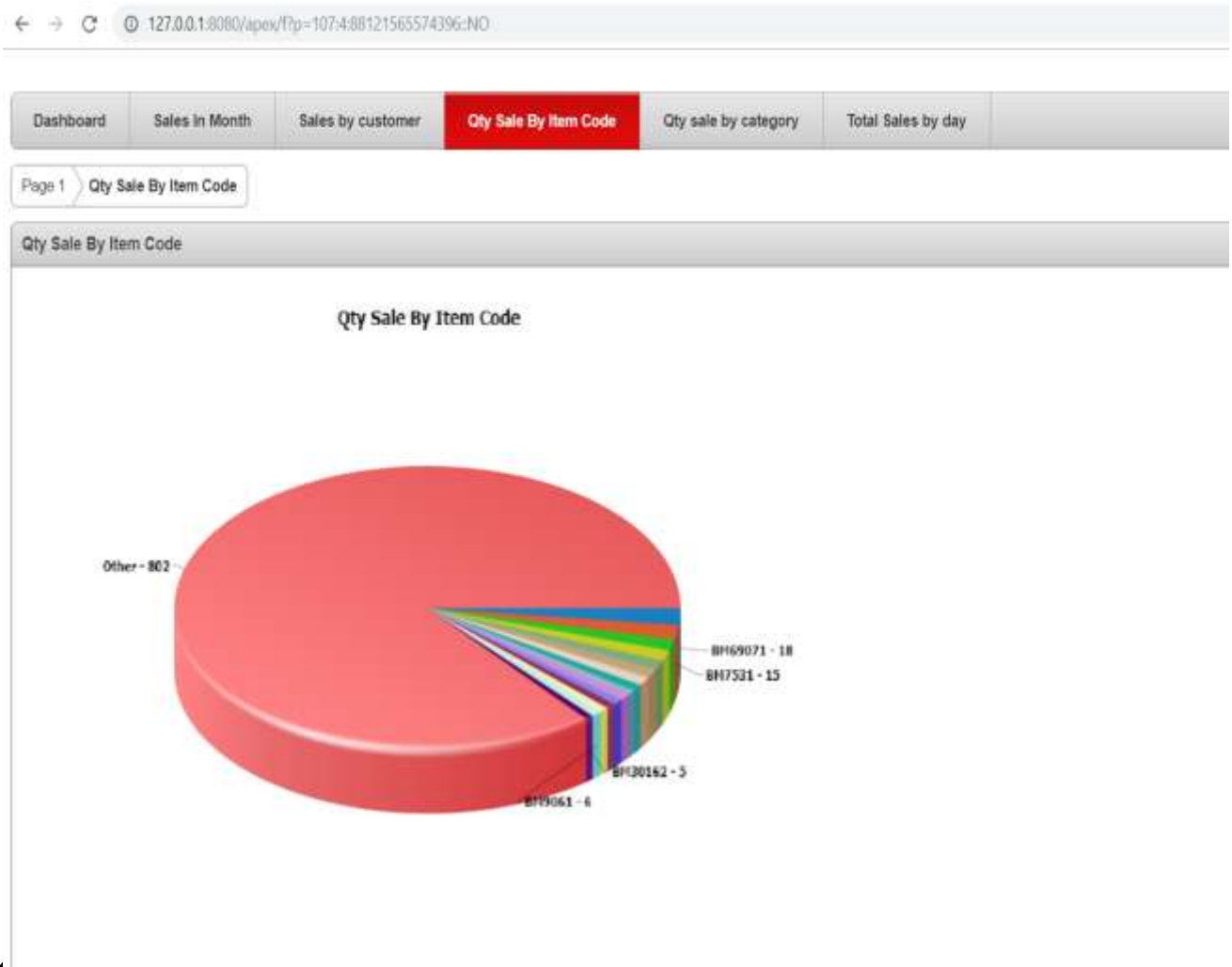


Fig: Quantity sales done by Item Code

Report was generated after successfully running the system. From above figure we can find the top 15 items sold according to their item code.

x) Total Sales by day

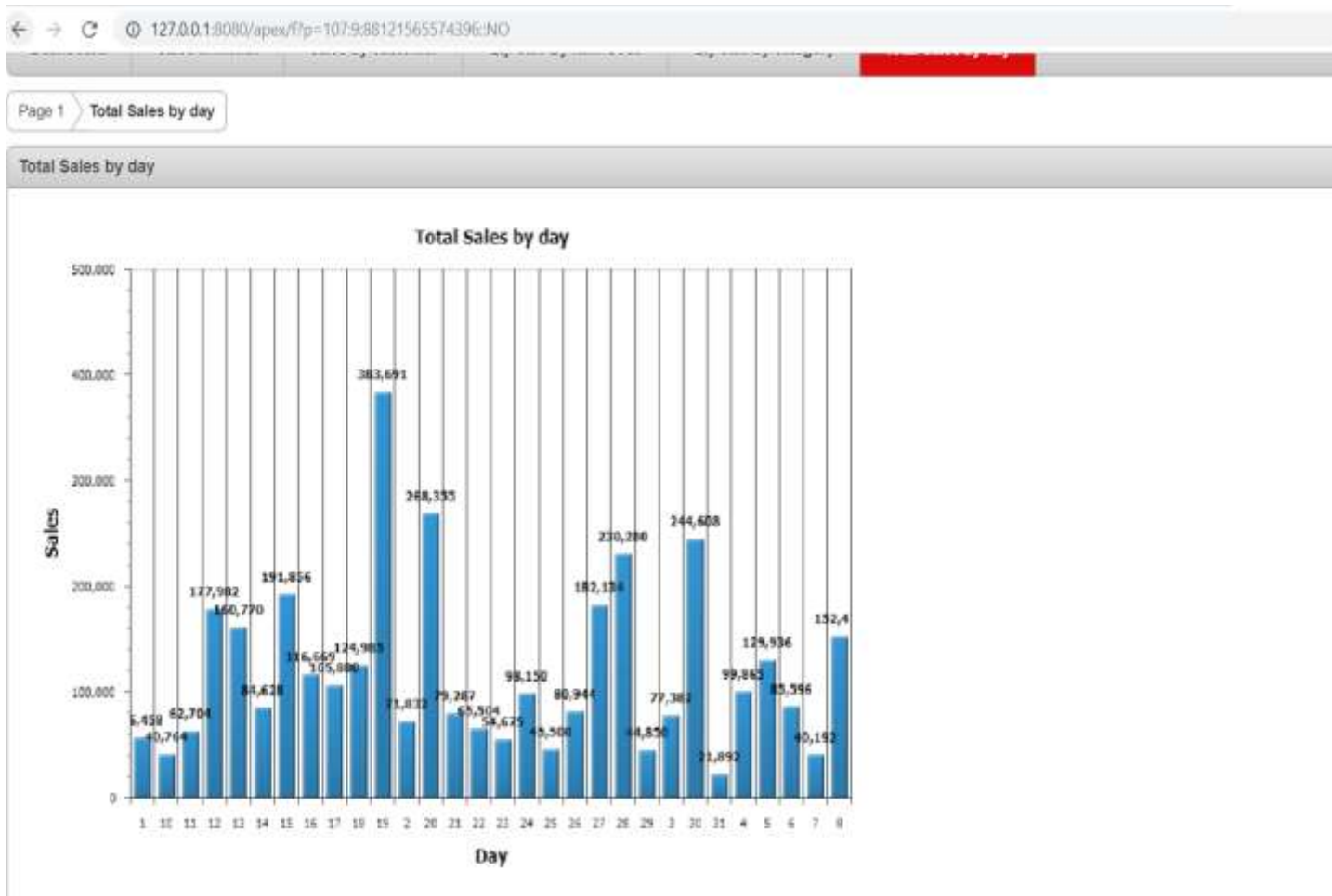


Fig: Total sales by day

Report was generated after successfully running the system. From the above figure we can find the sales of January and February.

y)

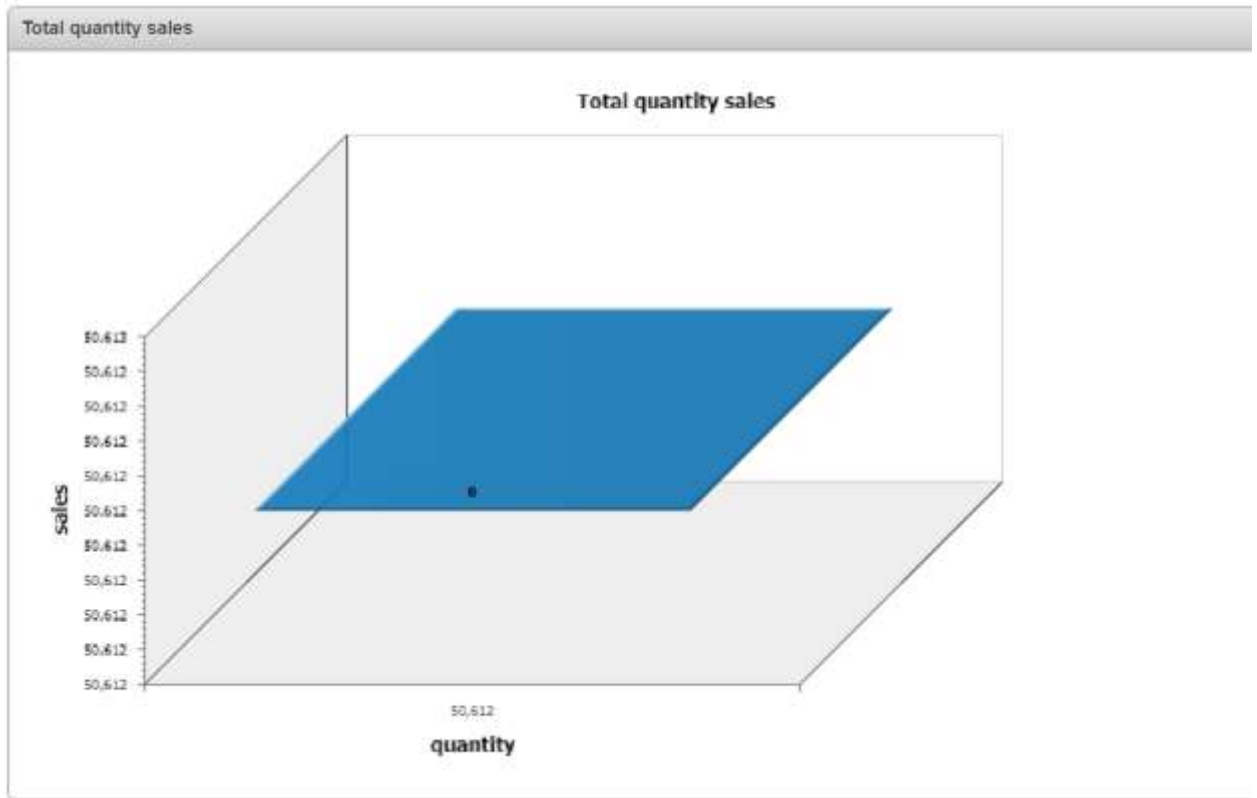


Fig: Total quantity sales report in dashboard

z)

Sales Id	Bill Date	Customer Name	Category Name	Bill Qty Sum	Net Amount Sum
121	02-JAN-18	Kohit Kumar Pokharely	ORBIT CHEWINGUM MIXED FRUIT 4 4GM	1	10
122	19-JAN-18	Kohit Kumar Pokharely	GULLON SUGAR FREE DIGESTIVE 400GM	1	292
123	31-JAN-18	Karu Baraly	PVC CAP SMALL	2	26
124	10-JAN-18	Ksha Pandey	NESTLE NESCAFE COFFEE & MILK HAZELNUT 180ML	1	50
125	22-JAN-18	Kabi Raj Thapy	BIG CHOICE REGULAR MAAS NON POLISH DAAL/KG	1	94
126	28-JAN-18	Kabi Raj Thapy	UNIQUE CASHEWNUTS 100G	1	202
127	21-JAN-18	Krusha Joshy	DDC KANCHAN CHEESE 500G	1	520
128	14-JAN-18	Kayusha Khadkay	PALPASA BROWN SESAME BALL 90G	2	84
129	12-JAN-18	Kditi Hirawaly	PASEO TOILET 2 ROLL 3 PLY	1	174
130	22-JAN-18	Kditi Hirawaly	REAL CRANBERRY NECTAR 1L	1	220
131	22-JAN-18	Kditi Hirawaly	HIBIS ULTRA THIN NORMAL 20 PADS	1	326
132	30-JAN-18	Kradeep Thakury	HALDIRAMS BHUJIA SEV 350G	1	175
133	13-JAN-18	Kantosh Sharmay	CLEANSING WIPES 80PCS	1	124
134	13-JAN-18	Kantosh Sharmay	IMPERIAL LEATHER CLASSIC SOAP 115G	2	100
135	20-JAN-18	Kantosh Sharmay	MAMY POKO PANTS BABY DIAPER S42	1	798

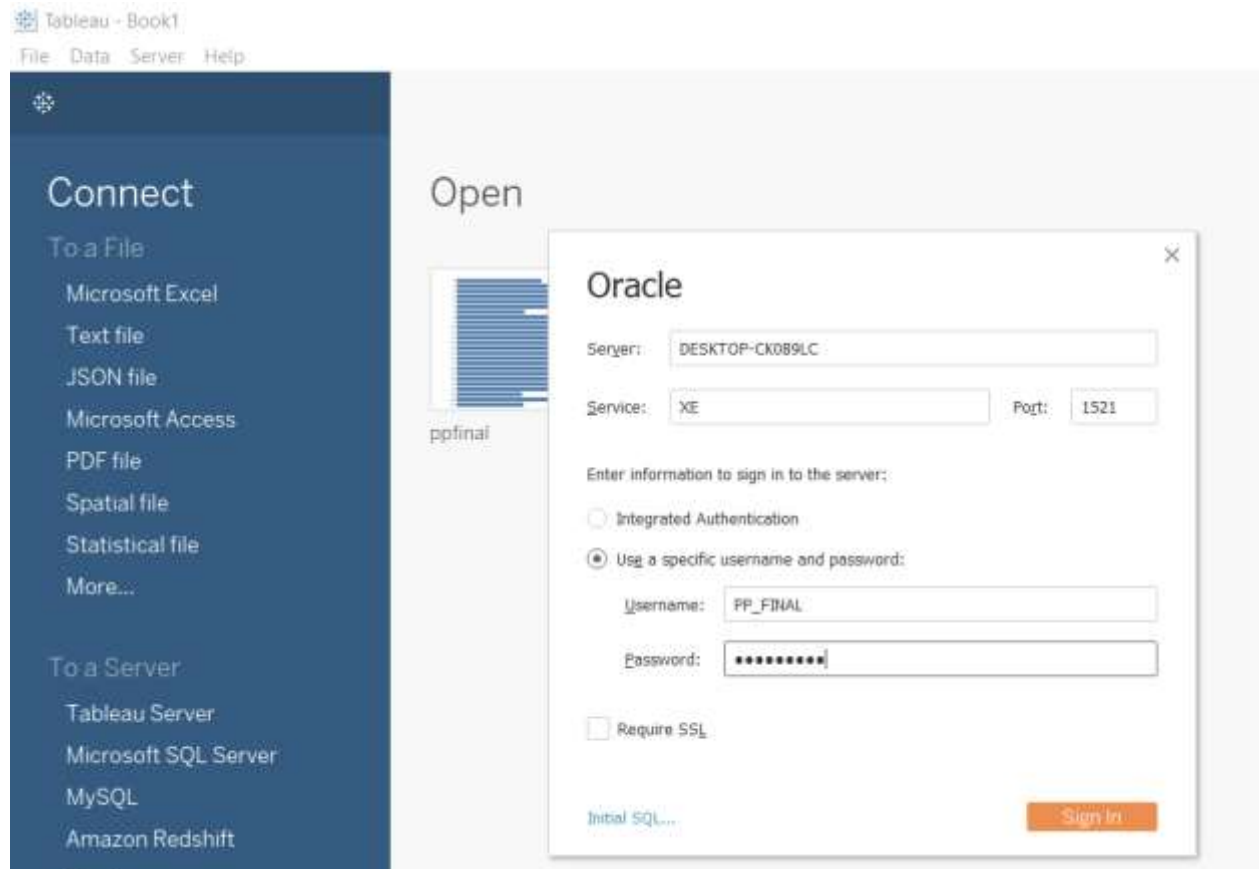
Previous row(s) 121 - 135 of more than 500 Next

Fig: Sales Report in dashboard

NOTE: All the first name, last name and contact number of customer were altered for the privacy issues of Customer .

5.2 Tableau

Tableau software was used for the generation of business intelligence report. The oracle workspace was connected with Tableau for various useful report generation of the Big Mart.



g. Connection with Oracle data warehouse

Reports of Tableau



Fig: Total sales done in January and February was generated from the tableau

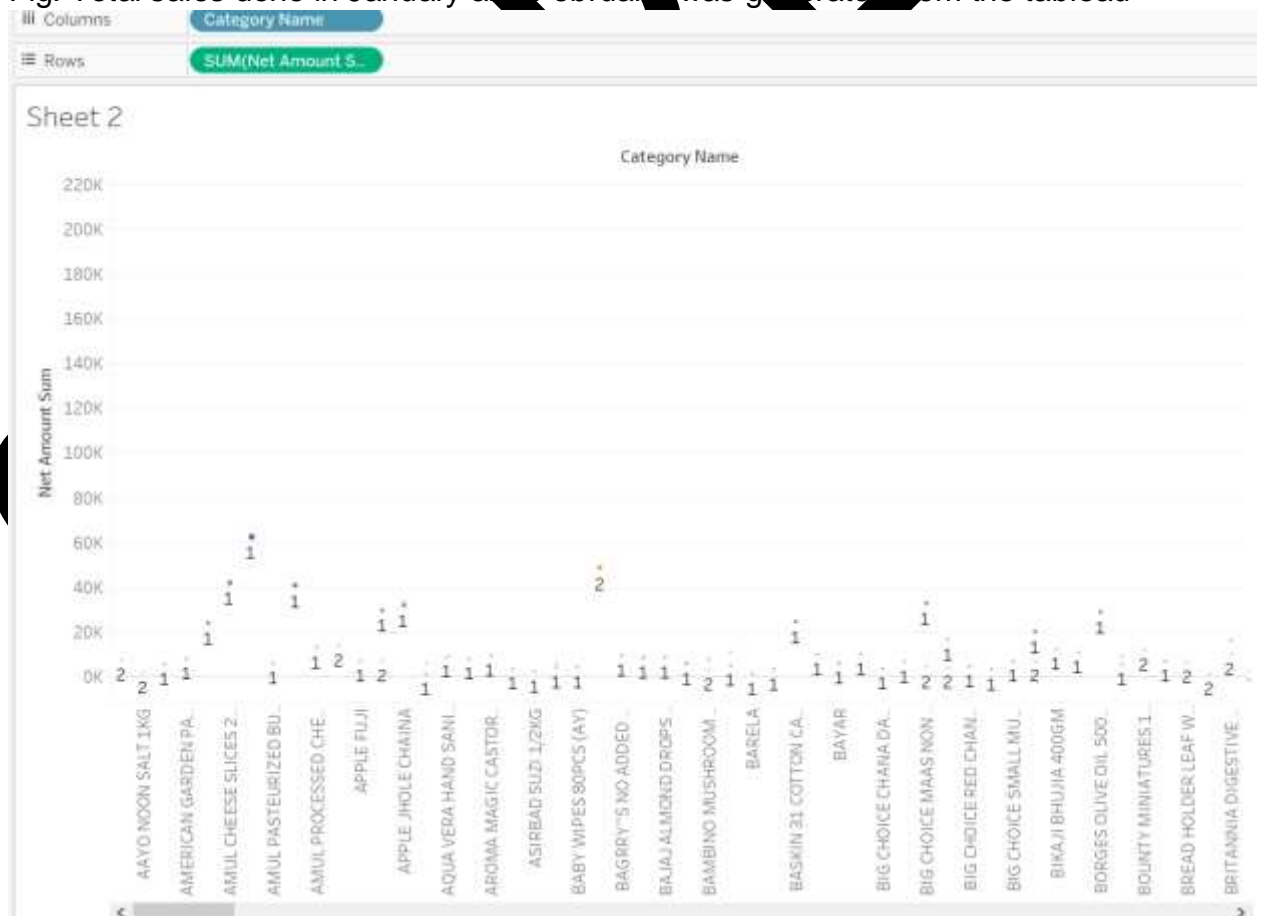


Fig: Tableau report of Category name with net sum sale

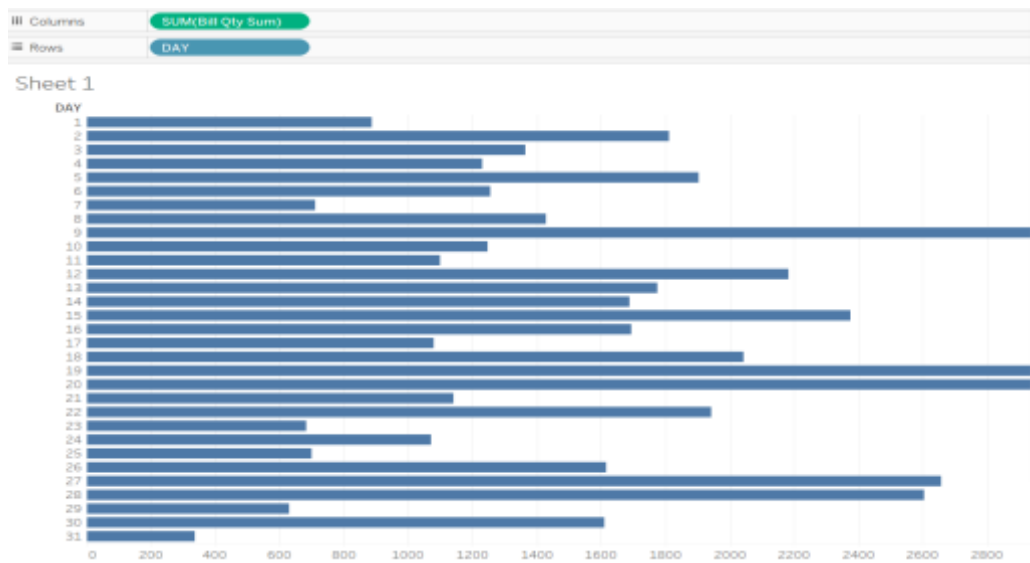


Fig: Tableau Report of total Sum according to per day

SAMPLE

5.3 Product Testing

Testing is very vital part of any product as it defines the outcome from the development of project. For the testing of product, Black Box testing was carried out. After giving input the output was tested. Black box testing is easy method as it doesn't require any programming skills and even the tester can carry out black box testing in very effective way.

Testing of the Data Warehouse:

Activity Number	Description	Results
1.	Identifying Source Data.	Pass.
2.	Loading Source Data to Staging Area.	Pass.
3.	Correct way of identifying each table and its column in Staging Area.	Pass.
4.	Procedure to check data quality of the data in staging area to load clean data into clean tables.	Pass.
5.	Procedure to check data quality of the data in staging area to load bad data into error table.	Pass.
6.	Populating Clean tables.	Pass.
7.	Populating Error tables.	Pass.
8.	Tables designed according to the star schema i.e. Dimensional and Fact tables.	Pass.
9.	Extracting data from clean tables and loading into Dimensional Tables.	Pass.
10.	Extracting data from Dimensional Tables to load into Fact Tables.	Pass.
11.	Populating Dimensional tables.	Pass.
12.	Populating Fact table.	Pass.
13.	Apex Application to display Business Intelligence reports from the Data Warehouse.	Pass.
14.	Tableau software to display Business Intelligence reports from the Data Warehouse.	Pass.

Chapter 6

6.0 Product Evaluation

6.1 Introduction

The goal of this chapter was to collect data, the approaches and necessities required for the construction of data warehouse to promote the business. The exploration of design and development was included throughout the research. The focused member from which data was gained to answer the questions, is mentioned. This production project implemented the method of investigating the issues, creating a proposal to resolve the issue.

6.2 Exploration Strategy

The environments for gathering and analyzing data are organized with the decisive aim of achieving significance to the production project purpose. According to ([Seltize et al,1964](#)) investigation strategy is a conceptual construct which acts as an administrative and plot for the gathering, extending data analysis. The study accepted vivid and case study plans. The expression research strategy proposed to deliver into the research obstacle, achieved through explaining the main variable of interest. The assumed plans were preferred since it is modest to use and suitable.

6.3 Targeted Population

Big Mart ICT staff members were interviewed to describe the current system architecture and management of data in the organization. The organization included the administrator for system, administrator for database, analysts for system and supporting staffs. The procedural specialists consisted to govern the data warehouse system based on arranged data and LAP sub query collected. All those methodologies described the existing and current system architectures. Big Mart management as well as Business faculty departments were interviewed to regulate the process of business. The key aim was to arrange and classify data. All the sales data collected from POS was used to significant business verdicts; the category and magnitude of data to the data warehouse. Three senior heads of departments were interviewed; ICT, Finance and Operations. The panel's management resolute the cut-off months for transactional history as it will be massive in size. Including this all, other targeted members was members of accounts and digitization board, accountable for the digitization of account and information in the organization.

6.4 Data Collection Approaches

The primary data collection bases were soft copy data along with questionnaires. While efficiently using fastened ended questions duration and additional sources were hoarded and enabled an analysis. Flexible questions gave detailed answers and improved view of all applicable evidence of current system architecture and position of data quality. Meeting was conducted on organization target groups. Administration were interviewed for the decision-making method of Big Mart. The assessment of current literature relevant to this plan will be systematically done with significant attention on verdict support, data warehouse development techniques in Big Mart .

Category of Data Information	Foundation of Data	Targeted Cluster	Inhabitants	Data Collected
Past data and inclinations	<ul style="list-style-type: none"> • Inheritance Schemes • External Sources 	<ul style="list-style-type: none"> • System Experts • Archives • Digitization Team 	1 1	<ul style="list-style-type: none"> • Historical and current technologies and tools at the Panel. • Data capacities and architecture. • Nature of data (accuracy and completeness). • Data collected through interviews and questionnaires
Commercial/Process Needs	<ul style="list-style-type: none"> • Big Mart Strategic Plan • ICT Strategy and papers 	<ul style="list-style-type: none"> • Heads and Administration • Departmental Vectors 	3 4	<ul style="list-style-type: none"> • Commercial/Functioning requirements, strategies and procedures • Commercial/Functioning verdicts and procedures • Data composed through soft copy and questionnaires

Current System Architecture	<ul style="list-style-type: none"> • ICT Strategy; Organization Security and Structure Strategy 	<ul style="list-style-type: none"> • Head of Information and Communication Technology • Procedural staff 	1 5	<ul style="list-style-type: none"> • Construction of ERP Structure and other present stages. • Extract Transform and Load Process necessities • Data composed through investigations and current studies
Functional Necessities of Data Warehouse	<ul style="list-style-type: none"> • Data Warehouse Proposed Typically 	<ul style="list-style-type: none"> • Professional Specialists 	3 3 2	<ul style="list-style-type: none"> • Business Requirements and supplies • Data collected through inspections, current studies and representations
Practical Necessities of Data Warehouse	<ul style="list-style-type: none"> • Data Warehouse Planned Model 	<ul style="list-style-type: none"> • System Experts • Organization and Network Overseers • Database Administrator • Additional procedural staff 	3 3 2	<ul style="list-style-type: none"> • Organization Requirements • Logical and Conceptual strategy necessities • Physical strategy necessities • Data collected through inspections and current study

6.5 Analysis of Data

After an explanation of the data gathering procedure and the general features of the defendants, a vivid analysis was directed. To the range that data incorporation to a data warehouse is not presently predominant in Big Mart, the functional conclusions from the vivid analysis supplement the frame of current information on data incorporation and warehousing and its implementation in the Panel for decision support.

Data analysis is the process used to define fact, perceive designs, progress descriptions and test premises. The queries and replies from the plans were implied and arrived into the mainframe using Microsoft Excel to Oracle Application Express (Apex).

6.6 Consistency and Validity

By definition, consistency signifies to in what way a given set of dimensions are obtainable. On the other hand, validity defines the grade to which the investigation reflects to the assumed real exploration complications. Evaluating and associating the validity of outcomes was accentuated. It can similarly be well-defined as the repeatability of conclusions where similar outcomes are acknowledged upon a solitary test path or whereas perceiving actions of some procedure. Similarly, validity states to the reliability or acceptability of the research project. There is a decisive relationship among consistency and validity. For example, if data is effective it can assume that they must be consistent. Though, it is worth observing that consistency is essential but not adequate, it remained a state for validity.

6.7 Potential of the project

The effective execution of a data warehouse product can carry major potential to an organization like Big Mart:

- **Competitive benefit**

The enormous revenues on investment for Big Mart that has effectively implemented a data warehouse product is indication of the enormous competitive advantage that go with this technology. The modest benefit is increased by letting decision-makers access to information that can expose earlier inaccessible, unidentified, and available data on, for sample, customers, trends, and strains.

- **Enlarged production of business decision-makers**

Data warehousing expands the efficiency of business decision-makers by generating a united database of reliable, subject-oriented, past data. It participates data from several mismatched schemes into a procedure that delivers one reliable sight of the Big Mart. Through transforming data into expressive information, a data warehouse product permits corporate managers to achieve more functional, correct, and steady analysis.

- **Extra cost-effective decision making**

Data warehouse helps to decrease the inclusive price of the product by dropping the amount of stations.

- **Improved initiative intelligence.**

It assistances to deliver better initiative intellect.

- Improved customer service.
- It is used to increase customer facility.

The necessity of data warehouse system is demonstrated in figure below:



6.8. Limitations of the product

The limitations connected with developing and handling a data warehouse product are as follows:

Underestimation of properties of data loading process

Sometimes we underrate the period required to extract, transform, and load the data into data warehouse. It might take the important quantity of the over-all expansion period, though few tools are there which are cast-off to decrease the period and exertion consumed on this method.

Unseen complications with basis systems

Sometimes unseen complications related with the basis systems feeding the data warehouse system might be recognized after ages of existence undetected. For instance, when inflowing the facts of an original stuff, confident fields might permit nulls which might upshot in supervise entering unfinished stuff data, even when obtainable and appropriate.

Mandatory data not taken

In some cases, the mandatory data is not taken by the basis systems which might be very significant for the data warehouse persistence. For instance, the day of recording for the stuff might be not cast-off in basis system but then again it might be very significant investigation purpose.

Demand of high maintenance

Data warehouse demand high maintenance. Any restructuring of the commercial procedures and the basis systems might affect the product which results high maintenance charge.

Long-duration schemes

The construction of a huge data warehouse can take up more duration which is why few organizations are unwilling in exploring in to data warehouse product. Approximately the past data of a specific section is taken in the data warehouse resultant data-marts. Data marts care only the necessities of a specific section and incomplete the functionality to that section or part only.

Complication of incorporation

The most significant part for the management of a data-warehouse product is the incorporation competences. Big Mart must apply a momentous quantity of period determining in what way the different data warehouse tools can be combined hooked on the complete resolution that is wanted. This work can be a very problematic job.

SAM

Chapter 7

7.0 Project Management and Evaluation

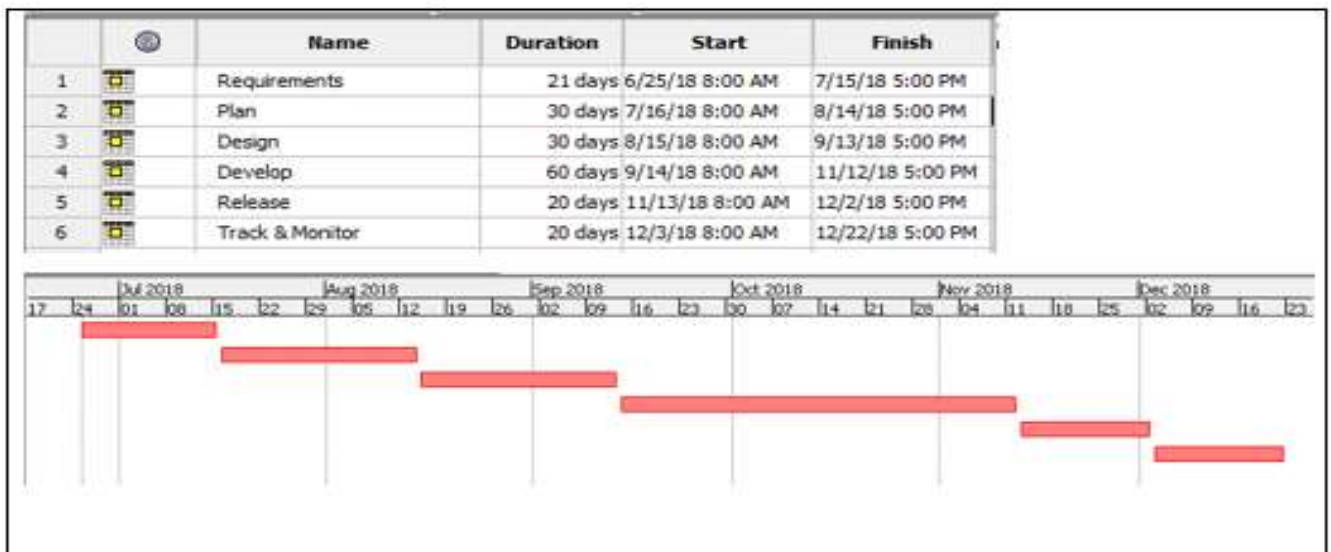
7.1 Introduction

Project management emphasizes on planning and establishing a project & its possession. This contains recognizing and dealing the development to be used, smearing it towards the user-centered project development, expressing the project theme, and professionally administrative the side through all stages until project accomplishment.

7.2 Value of Project Management

Through appropriate project management, we can pledge that the idea and goals of the project are continued and supporting the spectator's responsibilities and purposes. Moreover, we avoid risks successfully and professionally and use our accessible resources. It similarly supports to recognize the tasks, the deliverables projected, and the timetable needs to track to complete the project on exact time

7.3 Applied Gantt chart



7.4 Project Evaluation

Performance Evaluation

The matrix associated the abundant methods for demonstrating Extract Transform and Load processes and evaluated the planned design compared to other design. *P* determines that the design supported the identical limitations, somewhat.

Comparison and evaluation matrix				
Measure	Models			
	UML environment	Conceptual constructs	Mapping expressions	EMM
Design aspects				
Complete graphical model	x	✓	x	✓
New constructs	x	✓	x	✓
Object Oriented concept independent	✓	P	✓	✓
DBMS independent	✓	✓	✓	✓
Mapping operations	✓	✓	✓	✓
User defined transformation	x	x	x	✓
Mapping relationship	✓	✓	✓	✓
Source independent	x	x	x	x
Source converting	x	x	x	✓
Flat model	✓	✓	✓	✓
Implementation aspects				
Generate mapping document	x	x	x	✓
Non-relational handling	x	x	x	x
Generate SQL	x	x	✓	✓
Develop a tool		✓	✓	✓
Evaluation	4	7.5	7	13
✓ = 1; X=0; P: partial=0.5; Grand Total=13.				
Key: ✓ =YES; X= NO				

Fig: Project Evaluation and Assessment

7.5 Field Methodologies

The motivation behind this venture was to convey the goals of information from a record, i.e. Excel sheet for dissecting and revealing. The valuation of studies and related assessments pointed out by sorting the key establishments and highlights of dynamic ETL improvement on the execution of an information stockroom. The exploration strategy utilized expressive and contextual analysis designs. The accumulation of information was coordinated to make the innovative condition of dealings at Big Mart, determining the current going astray wellsprings of information, the design of System and the board of information. Surveys figured out how to operate and the relation controller was utilized to gather data. The analysis of data is finished utilizing Excel and Oracle Apex with outlines and tables. The procedure of ETL met the possibilities of managing in creating reports in an adaptable and benevolent way.

Since its starting, Big Mart has been utilizing Gauss programming application with a prophet database to do the majority of its undertakings. Information at Big Mart required to make educated decisions are inside divided frameworks not properly incorporated and not inexhaustibly worked. Delivering and improving data by these wellsprings of information has been demanding and tedious. For the current purpose of system, Big Mart has vital creativity to support its business procedures joining and merge by drawing closer with the information distribution center.

7.6 Usability

The project research additionally revised the presentation of the Extract Transform and Load tool, as to whether it was user friendly and how accessible it remained to the Big Mart administration. It required to create the user consumption reply in relative to Graphical User Interface (GUI), Navigation and Perceptibility of the system.

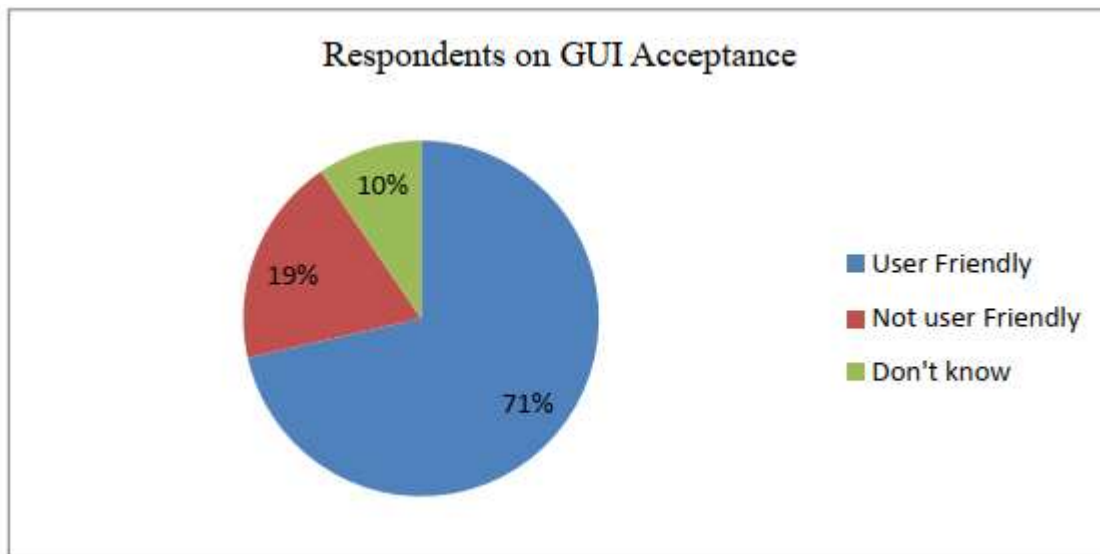
Below are some tables which was taken from administration on the basis usability

The tables below depict some of the feedback from users as per the usability limitations.

i) Acceptance of GUI

Respondents on GUI Acceptance		
Response	Frequency	Percentage
User Friendly	15	71%
Not user Friendly	4	19%
Don't know	2	10%

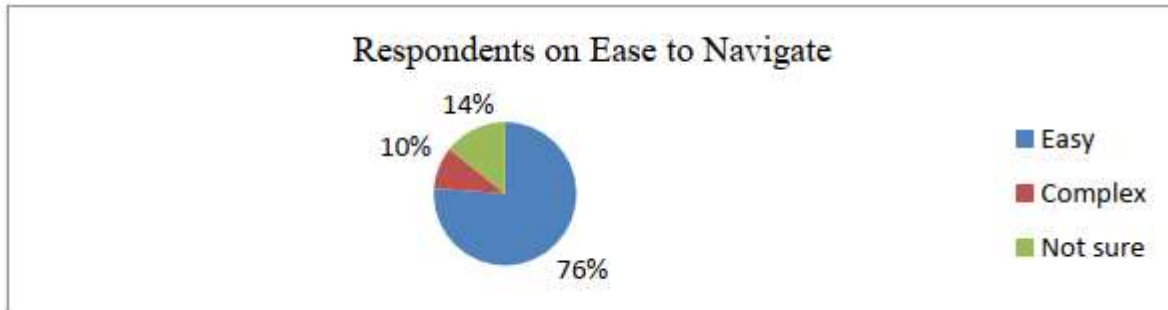
Fig: Response of Acceptance of GUI'



ii) **Comfort to Navigate**

Ease to Navigate		
Response	Frequency	Percentage
Easy	16	76%
Complex	2	10%
Not sure	3	14%

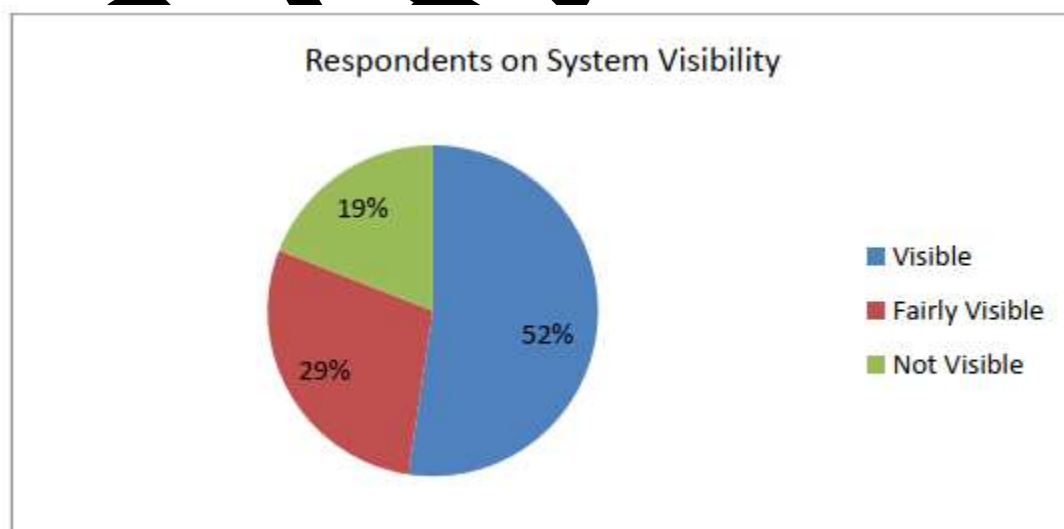
Fig: Response for comfort to navigate



iii) Perceptibility

Ease to Navigate		
Response	Frequency	Percentage
Visible	11	53%
Fairly Visible	6	29%
Not Visible	4	19%

Fig: Response on Perceptibility



System Usability Discussion

From the above usability parameters administered to twenty one (21) system user it can be deduced that from the GUI 71% of the users find the overall system to have a good user interface that is easy to interact, 19% of the system users are of the opinion that the overall system does not have a good interface to interact while 10% of the users have don't know. From this assessment, the research can easily conclude that majority of the users are comfortable using the system without ignoring the remarks raised by the 19% system users' i.e. color scheme of the interface interactive screen and some highlighted that the system is congested. 76% of the users find the system easy to navigate from one point to another while 10% of them find challenges navigating from one point to another, 14 % are not sure as they still require more time to analyze the system. On the visibility of the system especially for the ETL tool 52% of the users have no problem with font size used, 29% are partially comfortable with the screen visibility while 19% have issues with screen visibility. 29% and 19% of the system users is as a result of aging workforce of the employees.

7.7 Reports on several portion after final product and project implemented

a) Total Sales of month January and February of Big Mart

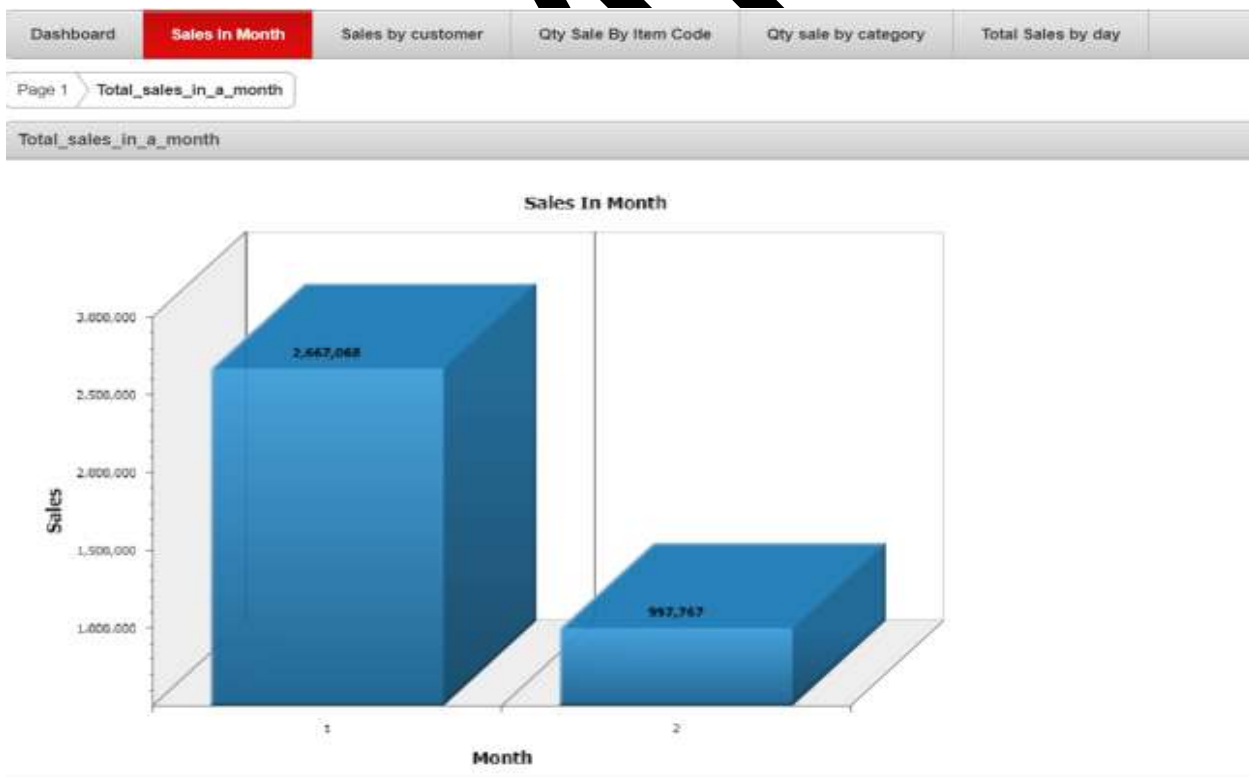


Fig: Total Sales of January and February

Report was generated after successfully running the system. From above figure we can find out the sales of January is more than sales done in February. So, the mart should more focus on promotion of sales rather than doing promotion in any other sector.

b) Sales done by Customer

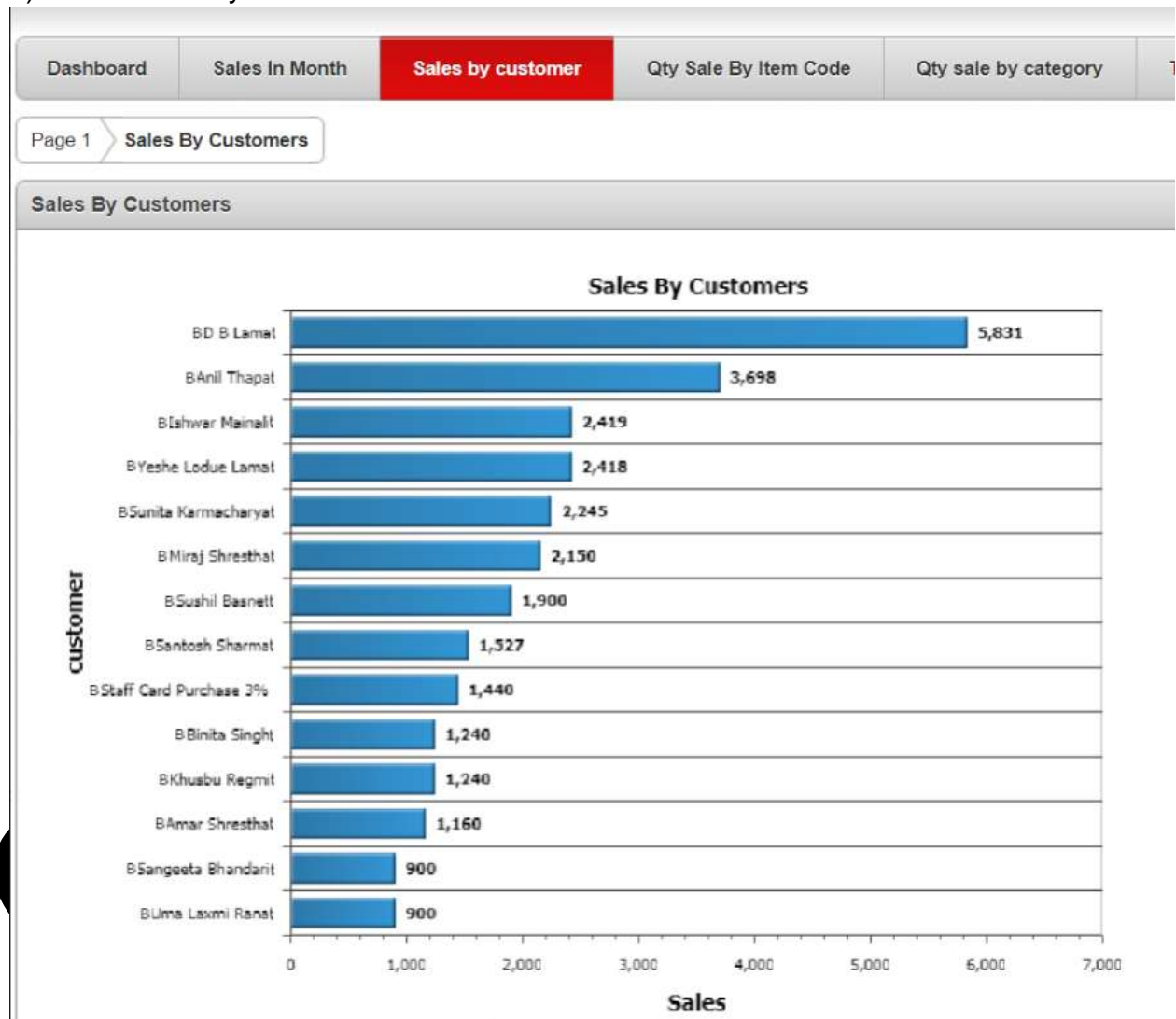


Fig: Sales done by customer

Report was generated after successfully running the system. From above figure we can find the top 14 customer who did more sales in the month of January and February. DB Lama is the best customer and various promotion facility should be provided to him

c) Quantity sales done by Item Code

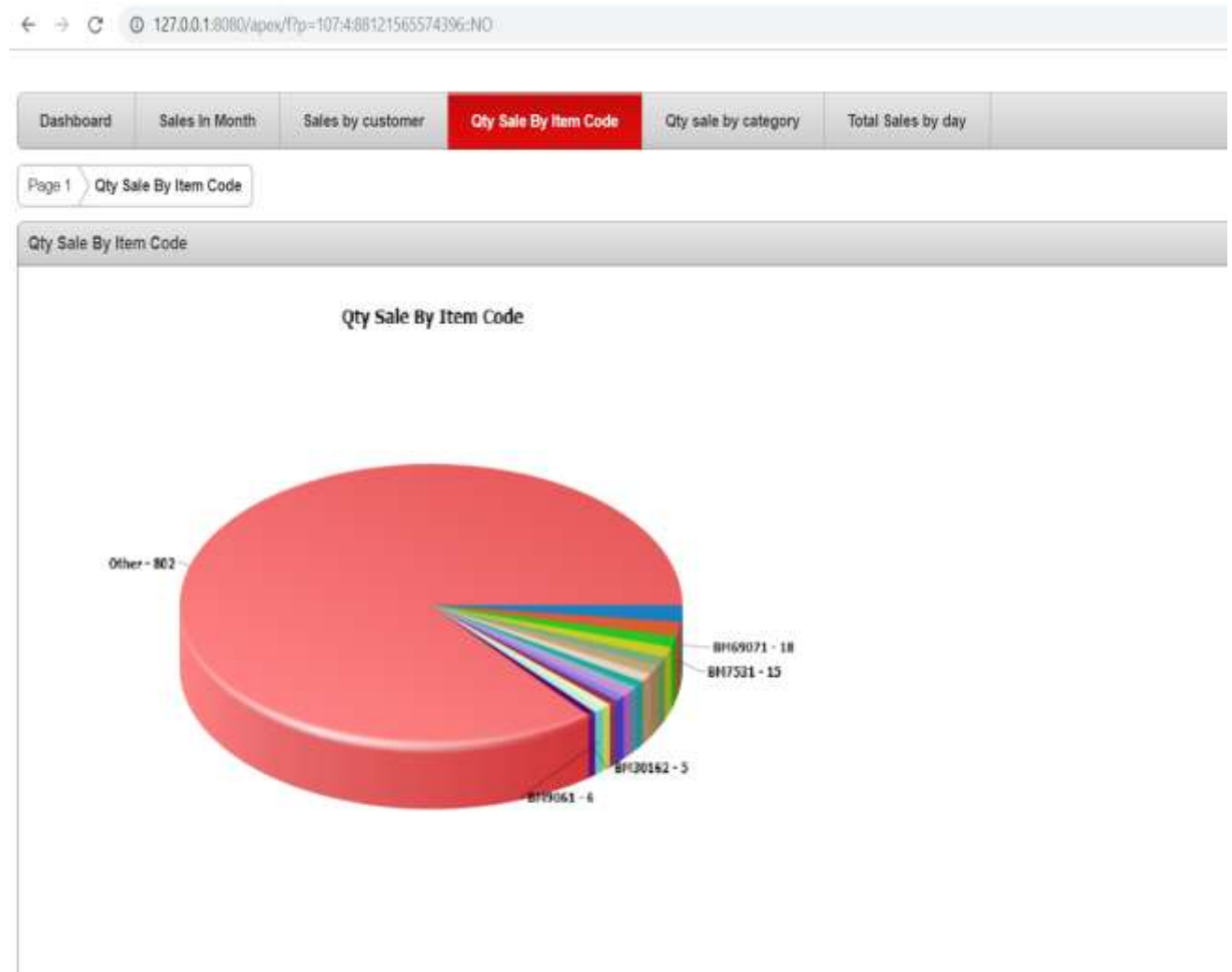


Fig: Quantity sales done by Item Code

Report was generated after successfully running the system. From above figure we can find the top 15 items sold according to their item code.

d) Total Sales by day

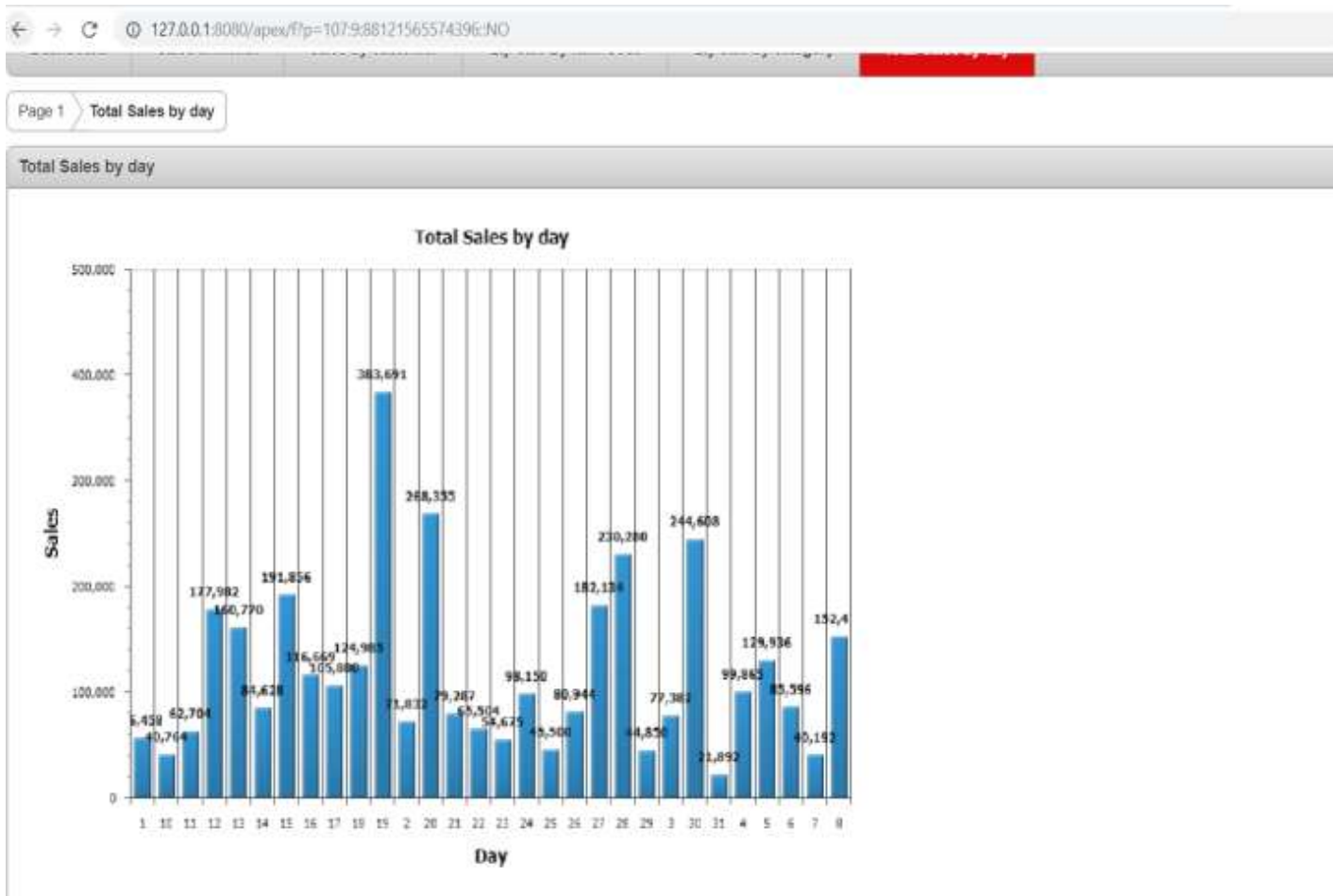


Fig: Total sales by day

Report generated after successfully running the system. From the above figure we can find the total sales of January and February.

7.8 Limitations of the Research

Data warehouse is normally a comparatively new field in Nepal and abundant exploration needs to be done. Some research revisions and opportunities were locally available, applicable to this research and that might further extant applied samples and situations. The necessity to bench mark by associated organization was delayed, missing a frequently approved measures and ethics to achieve the Extract Transform and Load processes.

The collection of Customer sales data was a main problem because of confidentiality and privacy of the organization and Customer. Though I visited various organizations for the collection of data with recommendation letter from the college, it was very difficult task to collect. I visited Bhat-Bhateni Super Market, KK marts and more organizations but they didn't provide me data. It was really tough task but at last I got an opportunity to meet with IT staff, Miss Prabina Poudel at Big Mart. She introduced with the head of Big Mart and I shared all ideas and implementation of my project. The IT head was impressed from my research initial planning and provided me customer sales data of 100000. For the development of data warehouse, filtered data was only used from all the data of January and February, 2018 as the size of data was massive. Additional optimization to advance the product concert was not accomplished and continuation difficulties were practiced. The testing of data warehouse was not 100 % covered and not achieved the full reports for 1 million data because of massive volume.

Chapter 8

8.0 Summary and Conclusion

8.1 Summary

The execution of a data warehouse in Big Mart, through Extract Transform and Load process design was the key research goal. The accomplishment was built on the filtration of data from mixed data foundations, for better timely policymaking and improved productivity in facility distribution. The research project categorically reports the foremost challenge of data remaining in raw nature, and absence of combined past and present data across Big Mart's range to assist the professional administration, accurate and reliable data to uphold modest authority. The intervening goal of this research was to make a data warehouse through ETL by the capability aimed at making of a decision. The review of literature explored prevailing procedures hired by other scholars. The valuation of the studies and associated reviews intended to classify the features and main fundamentals of an actual ETL development design for the execution of a data warehouse. The data played a vital part in classifying the boundaries of prevailing procedures and frameworks, originating in the review of the literature and later controlled to the expansion of design. Big Mart can easily implement the data warehouse in their organization and can manage their customer.

8.2 Conclusion

Usually, the research was designed at investigating current procedures by addressing problems associated with ETL procedure models in Big Mart. The primary objective of this production project was to do research on existing flat file of Customer Sales. Through the assessment of standard studies and current tools, dual main models were recognized and measured. The additional objective intensive was to develop data warehouse using ETL process. The objective was accomplished by associating approaches that occur in Big Mart with the literature review defines as perfect methods. Another precise objective was to design a system that analyses customer behavior and produces results for customer promotion. The production project accomplished this objective by implementing 2 complete jobs; developing dashboard and generating useful reports.

To accomplish, entirely the goals and aims of the project were accomplished effectively. The exploration and the project by way of whole provided a countless platform for exploring about various types of databases, designs of data-warehouse, procedure available in ETL process and usage of commercial intellect equipment in a data-warehouse. The concept of data warehouse including the generation of report using tableau was successfully done.

Achievements of the Research

Generally, the study was aimed at investigating existing methodologies by addressing issues related to ETL process models, in relation to Big Mart scenario. The first objective of the project was to identify and examine the strategies and methods ICT technical experts use to enable non-technical Users extract data from different sources in an accurate but simple way. Through the evaluation of documented studies and existing tools, two major models were identified and assessed. They included; the 'intermodal assertions', mapping and conceptual UML-based metamodel in relation to the backstage and the front-end data warehouse architecture. Three major international database vendors that provided ETL solutions were evaluated: Oracle, Microsoft and IBM. The second objective focused on identifying the shortcomings and gaps of the ETL process strategies/methods. The objective was achieved by comparing methods that exist in industry with what literature describes as ideal techniques. It was established that the models identified were highly subjective, complex and too technical for end users; hence unreliable in extracting data for timely and improved decision-making.

Lastly, a comparison between the new model and the ones existing in literature and in industry was conducted for validating the relevance of the new model.

8.3 Recommendations

- a) The initial cost of implementing a data warehouse based on any existing commercial ETL methods are high and the process requires commitment at all levels. However, the benefits of a data warehouse override the cost and return on investment can be realized if the opportunities are reviewed. The implementation of a data warehouse is a worthwhile investment project that most organization handling voluminous data cannot overlook.
- b) The need for completeness and freshness of data and a near real time ETL process is recommended.
- c) Based on the changing environment in distributed computing, databases and other related technological user needs, further recommendation is made for the algorithmic and theoretical results in data warehousing and underlying components and processes such as the ETL.
- d) To realize the full benefit of a data warehouse implementation, through the Entity Mapping Methodology, Big Mart should continually encourage and support staff in

embracing data warehousing as a way of improving decision-making, using divergent historical and current data.

- e) By analyzing the current technology, the algorithmic and hypothetical consequences inside the data warehouse and fundamental constituent and procedure such as the ETL should be implemented.

8.4 Future Research and Practices

- a) In future the automation of few steps of the methodology for validation is required which will validate automatically in the overall system.
- b) The notification about all categories of customer.
- c) Generation of databased instead of generating tables only.
- d) Data Auditing of whole system.
- e) The opportunity of up sale and cross sale.

References

- Anon, 2015. ARCHITECTURE FOR REAL-TIME ANALYTICAL DATA INTEGRATION AND DATA. , (July).
- Ferreira, N. & Furtado, P., Near Real-Time with Traditional Data Warehouse Architectures : Factors and How-to. Mawilmada, P.K., 2013. IMPACT OF A DATA WAREHOUSE MODEL FOR IMPROVED DECISION- MAKING. , (October).
- Oketunji, T. & Omodara, O., 2011. Design of Data Warehouse and Business Intelligence System. , (June).
- Parmar, V., Yadav, R. & Sharma, M., 2016. REVIEW ARTICLE OF DATA CLEANING MODEL FOR DATAWARE HOUSE. , 4(1). Quality, D. & Line, B., Data quality. Russell, P., TDWI CHECKLIST REPORT Data Integration for Real-Time
- Data Warehousing and Data Virtualization TDWI CHECKLIST REPORT Data Integration for Real-Time Data Warehousing and Data Virtualization.
- Senapati, R. & Kumar, D.A., 2014. A Survey on Data Warehouse Architectures., pp.5235–5239.
- Sharma, N. & Gupta, S.K., 2012. DESIGN AND IMPLEMENTATION OF SUCCESS THE CONTENTS IN THE., 6(1), pp.61–64.
- Simitsis, A. & Vassiliadis, P., A Methodology for the Conceptual Modeling of ETL Processes. Bernstein, P., Rahm, E., 2000. Data warehouse scenarios for model management. In: Proceedings of the 19th International Conference on Conceptual Modeling (ER'00), LNCS, vol. 1920, Salt Lake City, USA, pp. 1–15.
- Berson and Smith, 1997A. Berson, S.J. Smith Data Warehousing, Data Mining, and OLAP. Dobrev, Hakim pour, Dittmar, 2003.
- Operations and classification for data mapping in semantic integration. In: Proceedings of the 22nd International Conference on Conceptual Modeling (ER'03), LNCS, vol. 2813, Chicago, USA, pp. 534–547.
- Wang Aihua, Guo Winqiang, Xu Guoxiong, Jia Jiyou, Wen Dongmao, 2009. "GIS-Based Educational Mohammed, M. A., Hasson, A. R., Shawkat, A. R., & Al-Khafaji, N. J. (2012). E-Government Architecture Uses Data Warehouse Techniques to Increase Information Sharing in Iraqi Universities (pp. 1-5). IEEE Explore.
- Wang, F. (2009). Application Research of Data Warehouse and Its Model Design (pp. 798-801). IEEE computer society.

Yebai, L., & Li, Z. (2009). Interactive Modeling of Data Warehouse on E-business System (pp. 80-83). IEEE computer society.

Yen, D.C., Chou, D.C. and Chang, J. (2002), "A synergic analysis for Web-based enterprise resources planning systems", Computer Standards & Interfaces, 24(4), 337-46

George, S. (2012) **Inmon or Kimball: Which approach is suitable for your data warehouse?**[Online]. Available from: <<https://www.computerweekly.com/tip/Inmon-or-Kimball-Which-approach-is-suitable-for-your-data-warehouse>> [Accessed 30 July 2018].

Golfarelli, M. Rizzi, S. (2009) **Data Warehouse Design: Modern Principles and Methodologies** [Online]. New Delhi: Tata McGraw Hill Education Private Limited. Available from: <<https://cs09lects.files.wordpress.com/2012/12/book-data-warehouse-design-golfarelli-rizzi.pdf>> [Accessed 5 August 2018].

Guru99 (n.d.) **What is Business Intelligence? Definition & Example** [Online]. Available from: <<https://www.guru99.com/business-intelligence-definition-example.html>> [Accessed 15 September 2018].

Inmon, W. (2002) **Building the Data Warehouse** [Online]. Canada: John Wiley & Sons, Inc. Available from: <<http://fit.hcmute.edu.vn/Resources/Docs/SubDomain/fit/ThayTuan/DataWH/Bulding%20the%20Data%20Warehouse%204%20Edition.pdf>> [Accessed 15 September 2018].

APPENDICES

Appendix I: Initial Proposal

BSc (Hons) Computing Course 2018-19	
Level 6 Production Project	
Name: Shiwam Sah	Student I.D. : 193508
Course: BSc (Hons) Computing	Supervisor's Name : Pranish Upadhyaya
FINAL PROJECT INDIVIDUAL AIM & OBJECTIVES	
Title of my Project: 360 Degree Customer Management through Data Warehouse	
The aim of my Final Project: The major aim of my project is to develop a Data warehouse for Big Mart by using Extract, Transform and Load process. Personalizing the Customer Experience records for management work,	
Objectives of my Final Project: The objectives of my project are:	

- To do research on existing flat file of Customer Sales.
- To develop data warehouse using ETL process.
- To design a system that analyses customer behavior and produces results for customer promotion.
- To explore the approach engaged by another analyst in coming up along with ETL model.

Specification of my Final Product: The subsequent final product from this research is a vital approach empowering organizations to offer the best client encounter over all channels, by taking into consideration a brought together perspective of all client touchpoints with all divisions associated with client connections. This bound together view will be made by developing data-warehouse via ETL process through the organization catches about its clients and their cooperation. The final product resulting from this research will look into this objective.

Research

The goal of this chapter is to cover data, the approaches and necessities required for the construction of data warehouse to promote the business. The exploration of design and development is included throughout the research. The focused member from which data is gained to answer the questions is the main point. This production project implements the method of investigating the issues, creating a proposal to resolve the issue.

Methods of Production Evaluation

A description of the method of production evaluation (100 words max).

Must Have:

- Access by authorized only.
- Admin Panel
- Regaining info must be easy and correct.
- Product/Service Activity

Should have:

- Customer details with the description is stored in the database.
- Customer history easily accessible by the authority.
- Accessible customer history by the end user of the system.
- Transaction Behaviour

Could have:

- Notification alert about various Customers
- Upsell or Cross-sell opportunities
- Data Audit

PROJECT PLANNING and METHODOLOGY

Project Planning

A one-page Gantt chart viewing timing and limits for key activities.

Methodology

Following a precise methodology for the development of any product is so important in today's technological world. Basic methodology helps to plot, design and control the process of development. It makes easier to describe everyone's responsibilities and gives structure to the entire project. While development of the product, it is necessary to have a proper meeting with the supervisor, module leader for a better outcome of the product. Hence the development of a data warehouse is the product of this module, the movement of data inside the data warehouse never stops. There are continuous deviations inside the project of the data warehouse. After knowing these all, agile methodology is selected over waterfall methodology.

PHYSICAL RESOURCE

For the successful completion of my project, I require the the following software and hardware:

Item (Hardware or Software)	Source (Faculty, own or specified other organization)
• Personal Computer	Own
• Microsoft Office Package	Own
• Leeds Beckett Blackboard Student Login	Own (Log in credential)
• ProjectLibre Software	Own
• Oracle 11 g	Own
• Oracle Application Express (Apex)	

HUMAN RESOURCE

I am working on my Project with the following people

Dr. Pranita Upadhyaya

Name: *Dr. Pranita Upadhyaya*

Role: Module Leader

The British College Academic Staff


INTERNAL BIBLIOGRAPHY

Famuyide, S. (2018). A List of Requirements Prioritization Techniques You Should Know About. [online] Business Analyst Learnings. Available at: <https://businessanalystlearnings.com/blog/2016/8/18/a-list-of-requirements-prioritizationtechniques-you-should-know-about> [Accessed 6 Apr. 2018].

QASymphony. Agile Methodology: The Complete Guide to Understanding Agile Testing. [online] Airbrake Blog. Available at: <https://www.qasymphony.com/blog/agile-methodology-guide-agile-testing/> [Accessed 5 Apr. 2018].

SAMPLE

Appendix II: Data Collection Authorization Letter

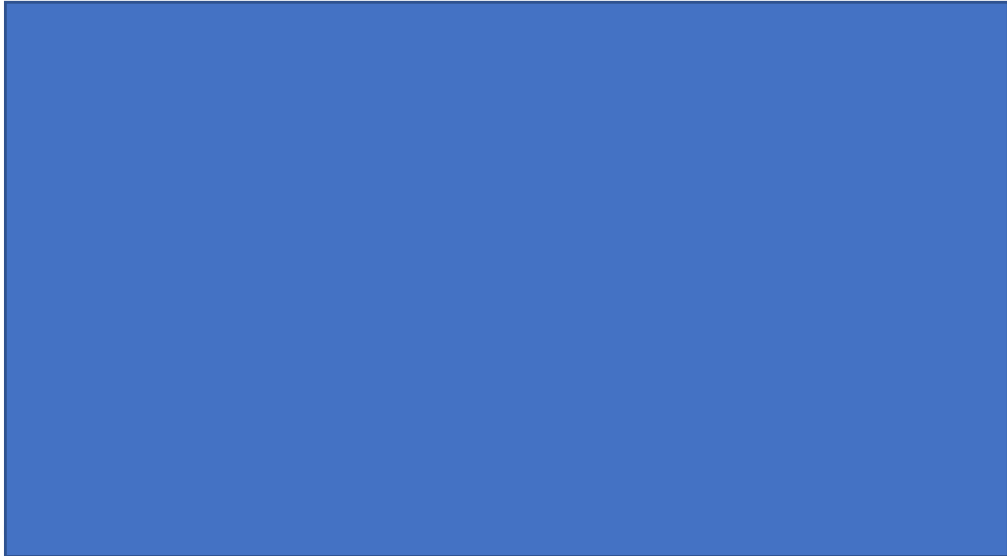


Ref No:-TBC/SSD/452
Date: 09th August, 2018

To whom it may concern



Dear Sir,

Sub: Permission for field visit



Thank You,

Yours faithfully,



Arun Lal Joshi
Associate Dean, IT & Computer Science

The British College, Thapathali, Kathmandu, Nepal
Tel : +977-01-5111101/5111102
Fax : +977-01-5111103
Email : info@thebritishcollege.edu.np
URL : www.thebritishcollege.edu.np

Appendix III: Field Visit Letter from Big Mart



MEGA MART PVT. LTD.
P.O. Box No. 4362
Lazimpat-2, Kathmandu, Nepal
Tele: +977 1 4005240, 4005241, 4005242
Email: info@bigmart.com.np
URL: www.bigmart.com.np



Regards,


23rd Dec 2018

Ashok Karn

Manager – IT Application



Appendix IV: Meeting Records with Supervisor

School of Computing, Creative Technologies and Engineering 2016/17	
Level 6 Computing Production Project	
MEETING RECORD SHEET:	Meeting 1
<i>ABOVE here - student to complete before Meeting with supervisor. BELOW here - complete at the Meeting.</i>	
Next meeting (date/time):	
Agreed Actions to complete before next meeting:	
1	Think about concrete topic research on.
2	Research about what kinds of searches are performed before.
3	Study different project methodologies.
4	Study the project requirement document properly.
5	Start preparing Gantt chart.
6	
Comments of supervisor (if any):	

References:

SAMPLE