II

**ABSTRACT**

Technologies related to data mainly consist of data warehousing, business intelligence using dashboards and data mining. We have implemented all the above mentioned technologies in this project.

An organization manages information in two dominant forms: operational systems of record and data warehouses. Operational systems are designed to support online transaction processing (OLTP) whereas data warehousing systems are designed to support online analytical processing (OLAP).

In the first part of our project, we have implemented the construction of the data mart for OLAP from the operational systems of records. The data mart is built using the star schema and consists of a single fact table and four dimensions table.

Base on the data in the data mart we have implemented two types of operations. The first part is to construct reports and dashboards for the end user to visualize the meaning of the data available in the data mart. The second part is to implement data mining algorithms to find out hidden meanings in the data.

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| No | Name | Page No |
| 1 | Introduction   * 1. Data Warehouse   2. Data Mart   3. Dashboard Reporting   4. Data Mining |  |
| 2 | Problem Definition  2.1 Introduction  2.2 Business Drivers  2.3 Business Objectives  2.4 Business Goals |  |
| 3 | Literature Survey  3.1 Introduction  3.2 Informatica  3.2.1 About Informatica  3.2.2 Informatica Products  3.3 Data Warehousing Concepts  3.3.1 Introduction  3.3.2 Components of Data Warehouse  3.3.3 Dimensional Data Modeling  3.3.4 Bill Inmon vs. Ralph Kimball  3.4 PowerCenter- ETL Tool  3.4.1 PowerCenter Client  3.4.2 Transformations Overview  3.5 Microstrategy  3.5.1 Introduction  3.5.2 Microstrategy Architecture  3.6 Data Mining Concepts  3.6.1 Frequent item set mining using FP-Growth algorithm  3.6.2 Naïve-Bayesian Classification |  |
| 4 | Project Requirement Definition  4.1 Introduction  4.2 Define Business Scope  4.2.1 Identify Source Data Systems  4.2.2 Determine Target Requirements  4.3 Define Functional Requirements |  |
| 5 | System requirements specification  5.1 Introduction  5.2 Design Development Architecture  5.3 Procure Hardware and Software  5.3.1 Hardware Used  5.3.2 Software Used |  |
| 6 | Gantt chart |  |
| 7 | System Design  7.1 Introduction  7.2 Develop Data Model  7.3 Develop Microstrategy attribute hierarchy  7.4 Develop design of mining algorithms  7.4.1 FP-Growth Algorithm  7.4.2 Bayesian Classification |  |
| 8 | Detailed Design  8.1 Introduction  8.2 Develop Source to Target Data flow diagram  8.3 Design Presentation Layer  8.3.1 Design Presentation Layer Prototype  8.3.2 Design algorithms for data mining applications |  |
| 9 | Implementation  9.1 Introduction  9.2 Design and Develop Data Integration Processes  9.2.1 Design High Level Load Process  9.2.2 Design Individual Mappings  9.3 Build Presentation Layer |  |
| 10 | Testing  10.1 Introduction  10.2 Perform Unit Test |  |
|  | Conclusion |  |
|  | Further Enhancement |  |
|  | Bibliography |  |

**List of Figures**

|  |  |  |
| --- | --- | --- |
| No | Name | Page No |
| 2 | 2.1 Sales Module Process Flow |  |
| 3 | 3.1 Informatica Products  3.2 Components of a data warehouse  3.3 PowerCenter Architecture  3.4 MicroStrategy Architecture |  |
| 6 | 6.1 Gantt chart for the month of January  6.2 Gantt chart for the month of February  6.3 Gantt chart for the month of March  6.4 Gantt chart for the month of April  6.5 Gantt chart for the month of May |  |
| 7 | 7.1 Star Schema for Data Mart  7.2 MicroStrategy Attribute hierarchy |  |
| 8 | Mapping to load  8.1 Date from flat file to warehouse  8.2 Customer from Apps to Staging  8.3 Fact table from Apps to Staging  8.4 Product from Apps to Staging  8.5 Customer from Stage to Warehouse  8.6 Product from Staging to Warehouse  8.7 Region from Oracle Apps to Staging  8.8 Region from Staging to Warehouse  8.9 Fact from Staging to Warehouse |  |
| 9 | Snapshots of Mappings  9.1 m\_load\_OrclApps\_Stage\_Customer  9.2 m\_load\_OrclApps\_Stage\_Product  9.3 m\_load\_OrclApps\_Stage\_Fact  9.4 m\_load\_OrclApps\_Stage\_Region  9.5 m\_load\_Stage\_Warehouse\_Region  9.6 m\_load\_Stage\_Warehouse\_Customer  9.7 m\_load\_Stage\_Warehouse\_Product  9.8 m\_load\_Flatfile\_Warehouse\_Date  9.9 m\_load\_Stage\_Warehouse\_Fact  Snapshots of Reports  9.10 Number of Orders Report  9.11 Number of Bookings Report  9.12 Orders to Bookings Close Rate Report  9.13 Product Metric Report  Snapshots of Dash Board  9.14 Number of Orders Placed in different regions for a customer Dash Board  9.15 Customer Information Dash Board |  |

**List of Tables**

|  |  |  |
| --- | --- | --- |
| No | Name | Page No |
| 8 | 8.1 Prototype of Number of Orders Report  8.2 Prototype of Number of Bookings Report  8.3 Prototype of Orders to Bookings Close Rate Report  8.4 Prototype of Product Metric Report |  |
| 9 | 9.1 Load Window for the Data Mart |  |
| 10 | Test Case 1  10.1 Cycle 1  10.2 Cycle 2  10.3 Cycle 3  10.4 Cycle 4 |  |
|  | Test Case 2  10.5 Cycle 1  10.6 Cycle 2  10.7 Cycle 3  10.8 Cycle 4 |  |
|  | Test Case 3  10.9 Cycle 1  10.10 Cycle 2  10.11 Cycle 3 |  |
|  | Test Case 4  10.12 Cycle 1  10.13 Cycle 2  10.14 Cycle 3  10.15 Cycle 4  10.16 Cycle 5  10.17 Cycle 6  10.18 Cycle 7  10.19 Cycle 8 |  |