Designing sales data mart system for ease of analysis and developing data mining model to enhance the promotional strategies for Gamma Pizza hut

S.Shamaran

129171M

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Degree of Master of Science in Information Technology.

November 2016

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been duly acknowledged in the text and a list of references is given.

S.Shamaran

Name of Student

S Shamaran

Signature of Student

Date: 26/11/2016

Supervised by

Mr.Saminda Premaratne

Name of Supervisor

UOM Verified Signature

Signature of Supervisor

Date: 21/12/2016

Dedication

THIS THESIS IS DEDICATED TO MY PARENTS AND MY SISTER SHANTHIPAA FAMILY AND BROTHER SHATHULAN FAMILY

Acknowledgement

It is a great pleasure to me thanks the many people who made this thesis possible.

My sincere gratitude and thanks to Mr.Samida Premaratne, the supervisor for this project and Course coordinator MSC in IT, for the tremendous guidance, motivation and corporation given throughout the project.

Heartfelt thanks to my parents, my sister family, and my brother family for their support and encouragement during the development of this project.

Special thanks to previous working colleagues and managers (Buddhika, Chithara, Dinith) of Scienter Technologies Pte Ltd.

Grateful thanks to current working colleagues and Sujeewa Fernando, Head of Dot Net development at Mazarin.

And lastly gratitude thanks to my wife Hamshine and his family for in order to able to continue my project after wedding on this year.

Abstract

Making effective business decisions with the data is the key to succeeding in today's competitive environment. Organizations are now looking to improve their decision-making ability with their current data, but unfortunately operational systems have limited features and various ad-hoc reports for same data. This unsatisfactory & frustration lead the managers and IT industry to find new level of applications. These applications focused on ease of analysis on the single screen to make effective decisions at the time and mining techniques help to generate new business opportunities by providing prediction of trends and behaviors as well as discovery of previously unknown or hidden patterns.

The DSS/BI systems should have more analyzing features and structured data. But current OLTP data and its database design not give much more analyzing power. In order to that OLAP architecture has built from various database vendors to make to use by DSS /BI systems. The developing of a data warehouse database and Data Mart database with suitable schema and approaching with relevant architecture is make a foundation for DSS/BI systems

The Data warehouse database makes on available history data as possible of getting last update record. The fact and dimension structure are used when designing database schema for Data Warehouse. ETL process generate a data to warehouse from various data sources. The Data Marts are used for holding various subject areas like sales, purchase, production, finance, etc. But here only considering about sales and delivery data only. The Data Cube Technology (OLAP technology) is used for end user to viewing data with various dimensional and drill-down drill-up processes within the application.

Finally those data are used to mining frequent patterns, Associations and Correlations between items in menu orders by using apriori algorithm (Microsoft Association algorithm) and forecasting Predictive sales for each item by using ARIMA algorithm (Mircosoft Time Series)

The data warehouse solution can be made from by integrating various database technologies in the middle; those technologies include SQL Server Management Studio (SSMS), SQL Server Integration Services (SSIS), SQL Server Analysis Server (SSAS), SQL Server Report Service (SSRS) and SQL Server Data Tools for Visual Studio used to create Analyzing project and Data mining project. C# language, DMX and MDX queries are used to build the simple mining application.

TABLE OF CONTENTS

	Page
Chapter 1- Introduction	
1.1 Introduction	1
1.2 Background and Motivation	1
1.3 The Problem Domain	2
1.4 The solution address by others	3
1.5 Aim & Objective	3
1.6 Solution in brief	4
1.7 Structure of Dissertation	4
Chapter 2 – Review of Others Work's	
2.1 Introduction	6
2.2 Background	6
2.3 Decision Support System Concept	6
2.4 Database and Data Mining for Coffee shops in Egypt	7
2.5 Data mining on time series:	8
2.6 Summary	9
Chapter 3 - Technology Adapted	
3.1 Introduction	10
3.2 Data warehouse Technology	10
3.3 Data warehouse	10
3.3.1 Basic elements of Data warehouse	11
3.3.2 Data warehouse management	12
3.4 Data base Design	14
3.4.1 Logical design	14
3.4.2 Physical Design	15
3.5 OLAP Technology	16
3.6 OLAP Server/models	17
3.6.1 The ROLAP Model	17
3.6.2 The MOLAP Model	17
3.6.3 The HOLAP Model	17
3.7 Data Warehouse Architecture	18
3.7.1 Conceptual view	18
3.7.2 Physical view	19
3.7.2.1 Centralized Architecture	19
3.7.2.2 Federated Architecture	19
3.7.2.3 Tiered Architecture	20
3.8 Data Warehouse models	20
3.9 DMX	20
3.10 MDX	21
3.11 Summary	21
Chapter 4 – Data Mining Algorithms	22
4.1 Introduction	22
4.2 Frequent Itemset mining	22
4.2.1 Association rules	22
4.2.2 Apriori algorithm	23
4.3 Sales Trend Analysis	24
4.3.1 ARIMA model	24

4.3.1.1 Autoregressive model 25
4.3.1.2 Moving Average model 25
4.4 How to apply 26
4.5 Summary 26

Chapter 5 – Theoretical Foundation of .NET 27
5.1 Introduction 27
5.2 Benefits of .NET framework 27
5.3 Integration with .NET 28
5.4 How to apply 29
5.5 Summary 29

Chapter 6 – Development Methodologies 30
6.1 Introduction 30
6.2 Overview of Approaches 30
6.3 The Bottom-Up approach 30
6.4 The Top-Down approach 31
6.5 Comparison between Top down Vs Bottom Up 32
6.6 Software process model 32
6.7 My Approach 33
6.8 Data warehousing development process 33
9.9 Summary 35

Chapter 7 – Analysis 36
7.1 Introduction 36
7.2 Requirement Analysis 36
7.2.1Existing system architecture 36
7.2.2 Business parameters 37
7.3 Source data analysis 39
7.3.1 Source conceptual schema 40
7.3.2 Hierarchical levels 41
7.3.3 Drill down analysis 41
7.4 ETL process 42
7.5 Data changing process 42
7.6 Tool Selection 43
7.6.1 Microsoft Corporation 43
7.6.2 Microsoft Data Warehousing Framework 43
7.6.3 Framework components 44
7.6.4 Business Intelligent Development Studio 44
7.6.5 SQL Server Integration Services 45
7.6.6 Microsoft Data mining Life Cycle 45
7.7 Summary 46

Chapter -8 Design 47
8.1 Introduction 47
8.2 Use Case Diagram 47
8.3 Matrix Design 48
8.3.1 Dimensions 48
8.3.2 Facts 48
8.3.3 Measures 48
8.4 Design decisions 49
8.5 The Constellation Schema 49
8.5.1 Data warehouse conceptual schema 50

8.6 Proposed architecture 51
8.6.1 The Logical Overview 52
8.7 Summary 52

Chapter – 9 Implementation 53
9.1 Introduction 53
9.2 ETL Process 53
9.2.1 Develop SSIS Package 53
9.3 Build & Populate Data Cubes 54
9.3.1 Develop SSAS project 55
9.3.1.1 Create Sales Cube 55
9.3.1.2 Create Delivery Cube 56
9.4 Data mining project 56
9.4.1 Mining data with Microsoft Association Rule 56
9.4.1.1 Mining results 57
9.4.1.2 Dependency Network 58
9.4.2 Mining data with Microsoft Time Series 60

9.5 User Interface developments for Data Mining 62
9.6 User Interface developments for ETL process 65
9.7 Summary 66

Chapter – 10 Testing 67
10.1 Introduction 67
10.2 Testing Sales value with predicted values 67

Chapter – 11 Conclusion 69
11.1 Introduction 69
11.2 Achievements 69
11.3 Problem Encountered 69
11.4 Limitations 70
11.5 Lesson Learnt 70
11.6 Further works 70

Appendix A – Matrix Design 73

Appendix B – Diagrams 74

Appendic C – Stored proc 75

Appendix D- MDX Queries 77

Appendix E – Implementing steps of Association rule 79

Appendix F – Strongest to Lowest Associate itemset 81

Appendix G - Coding for ETL package execution through Win App 82

Appendix H - Coding of Mining user interface 83

Appendix I – Test results 91

List of Tables

Table 3.1 Definitions of basic elements of the data warehouse 12

Table 6.1 Comparison of essential features of Inmon's and Kimbal's 32

Table 9.1 Association Rule -Mining Structure 57

Table 9.2 Association Rule - Algorithm Parameters 57

Table 9.3 Mining rules from Association algorithm 58

Table 9.4 Mining rules Item set 58

Table 9.5 Time Series - Mining Structure 61

Table 9.6 Time Series - Mining Algorithms 61

Table 10.1 Test Results 68

List of Figures

	Page
Figure 2.1: Decision Support system architecture	7
Figure 3.1: Collection of Data warehouse Tools	12
Figure 3.2: General features of OLAP	16
Figure 3.3: MOALP vs RLOAP	17
Figure 3.4: Conceptually Comparison of Data Warehouse Architecture	18
Figure 3.5: Physical View of Centralized Architecture	19
Figure 3.6: Physical View of Federated Architecture	19
Figure 3.7: Physical View of Tiered Architecture	20
Figure 6.1: The Bottom-up approach	31
Figure 6.2: The Top-Down approach	31
Figure 6.3: Data warehouse Development process	34
Figure 7.1 Existing system architecture	37
Figure 7.2Source conceptual schema	40
Figure 7.3 Hierarchical levels	41
Figure 7.4 Drill down analysis steps	41
Figure 7.5 Data warehouse Database Vendors	43
Figure 7.6Data changing process	44
Figure 7.7 Microsoft Data Mining Life Cycle	45
Figure 8.1 Use Case Diagram	47
Figure 8.2 Dimension, facts, matrix	48

Figure 8.3 Data warehouse conceptual schema 50

Figure 8.4 Proposed Data Warehouse Architecture 51

Figure 8.5 Logical Overview of the proposed solution 52

Figure 9.1 Conceptual Data flow 53

Figure 9.2 SSIS- Data Flow Task for migrating sales data 54

Figure 9.3 Sales Cubes – Data Source View 55

Figure 9.4 Date Hierarchy 55

Figure 9.5 Delivery Cubes – Data Source View 56

Figure 9.6 Dependency Network 59

Figure 9.7 Sum of Item sales by Month 60

Figure 9.8 Time Series Graph 62

Figure 9.9 Mining Results Interface 62

List of Abbreviations

BI – Business Intelligence

DSS – Decision Support System

KDD – Knowledge Discovery Data

ARIMA - Autoregressive –Integrated moving average

MDX – Multi Dimensional Query

DMC – Data Mining Query

OLAP – On Line Analytical Processing

ROLAP – Relational On Line Analytical Processing

MOLAP – Multidimensional On Line Analytical Processing

HOLAP – Hybrid on Line Analytical Processing

ETL – Extract, Transform, Load

SSIS – SQL Server Integration Service

SSAS – SQL Server Analytical Server

SSRS – SQL Server Reporting Service

BIDS – Business Intelligent Development Studio

DLL – Dynamic Link Library