Stationery Sales Analysis Using OLAP

Myat Myitzu Kyaw
University of Computer Studies,Yangon
nemesis.myitzu@gmail.com

Abstract

Today, business organizations need to know what's going on in the market, and make the right foreseeing decision. To overcome this requirement, this system implements useful tools to analyze the Stationery Shop sales records on multidimensional database and data cube by using on-line analytical processing (OLAP) that can view or browse information from different angles. Data mining tools can analyze enormous sets of data and then extract the meaning of the data and the resulting information can provide for the decision making process. A data warehouse, a repository of long-term storage of data from multiple data sources, organized sources as to facilitate management decision making. It makes better and faster decisions and expected to present right information in the right place at the right time with the right cost to make the right decision. It can present implementation of data warehouse for stationery shop. OLAP is a methodology that support analysis and decision support system with aggregate values extract from daily transaction database. Data cube will be used for faster reports since they store the pre-computation of count of the transactions.

Keywords: Data Warehousing, OLAP, Business Intelligence, Analysis

1. Introduction

Business transactions are more complex and more competitive today. Many corporations are actively looking for the new technologies that will assist them in becoming more profitable and competitive. Data can be stored in many different types of databases. A data warehouse that is designed for the query and analysis rather than transaction processing.

Data warehouse technology includes data cleaning, data integration and on-line analytical processing (OLAP) that is, analysis technique with functionality such as summarization, consolidation and aggregation, as well as the ability to view information from different angles.

A data warehouse is a repository of information collected from multiple sources. A data warehouse provides OLAP tools for the interactive analysis of multidimensional data of varied granularities, which facilities effective data mining. Data warehouse are constructed via a process of Extract, Transform and Load (ETL). The most popular data model for data warehouse is multidimensional data model. Data warehouse and OLAP tools are based on multidimensional data model. OLAP functionality is characterized by dynamic and multidimensional analysis of an enterprise data for end user activities. There are three logical database schemas for OLAP; star schema, snowflake schema and fact constellation schema. Typical OLAP operations in multidimensional data model are roll-up, drill-down, slice and dice and pivot.

An OLAP server helps the user to generate fast reports regardless of database size and complexity and users are allowed to define new adhoc calculations in any desired way. OLAP is implemented in multi-user and client server modes regardless of database size and complexity and offers quick response to queries. This helps users to go through historical and projected data in different data model scenarios and these are achieved only through use of OLAP server.

This system is to develop Stationery Sales Analysis Using OLAP. By using OLAP, the analyst or manager of the Stationery Shop can quickly analyze the information that has been summarized into multidimensional views and hierarchies. Data warehouse is an integrated database, used in decision making. Different operational data sources are integrated into a central repository. In this system, the data warehouse for previous sales records of stationery shop will be built. Different data sources will be prepared in the form of Microsoft Access and will all be homogeneous. By using this system, the analyst or manager can know the sales condition of each stationery item according to the time dimension

and category dimension and also know monthly or yearly sales reports of Stationery Shop.

3. Motivation

In the traditional script- based analytical reports are usually generated by extracting data from two-dimensional relational database tables. These two dimensional reports are formatted and delivered to users but these reports have disadvantages such as:

- These reports cannot be easily modified; they are predefined, built-in reporting applications
- Depending on the business reporting requirements only IT professionals can modify the report or build new report.
- Business data is in different place from the reporting user.

On the other hand, OLAP reporting integrates the complex issues of an enterprise data structures, procedures and algorithms into its dimensional data structures. These are presented to end users in easy to understand dimensional information views in a very instinctive way. Users can easily navigate the predefined reports and read the business data to build new adhoc reports with small support from IT professionals. By this way users become aware of existence business dimensions and get near to business data to answer the queries.

4. Background Theory

4.1. Three-Tier Data Warehouse Architecture

Data warehouse adopt 3-tier architecture.

The bottom tier is a warehouse data server that is almost always a relational database system. Data from operational database and external sources are extracted using application program interfaces known as gateways. A gateway is supported by the underlying DBMS and allows clients programs to generate SQL code to be executed at a server. The middle tier is an OLAP server that is typically implemented using either a relational OLAP (ROLAP) model or a multidimensional OLAP (MOLAP) model. The top tier is a client, which contains query and reporting tools, analysis tools, and/or data mining tools.

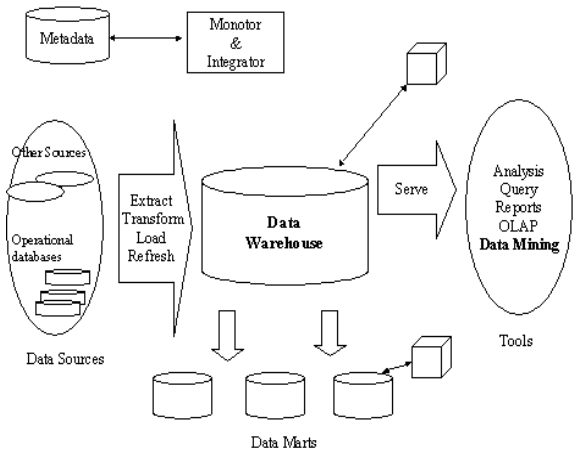

Figure 1. Three-tier Data Warehouse Architecture

4.2. ETL Process

Extract, transform, and load (ETL) in database usage and especially in data warehousing involves:

- Extracting data from outside sources .
- Transforming it to fit operational needs (which can include quality levels) .
- Loading it into the end target (database or data warehouse)

The advantages of efficient and consistent databases make ETL very important as the way data actually gets loaded. Programmers can set up ETL processes using almost any programming language.

4.3. OLAP Architecture

MOLAP (multidimensional OLAP)

The underlying data for a cube is stored along with aggregation data in a high-performance multidimensional structure. MOLAP storage provides excellent performance and data compression.

ROLAP (relational OLAP)

The underlying data for a cube is stored along with the aggregation data in a relational database. ROLAP storage enables you to take advantage of your investment in relational technology and enterprise data management tools.

HOLAP (hybrid OLAP)

The underlying data for a cube is stored in a relational database and the aggregation data is stored in a high-performance multidimensional structure. HOLAP storage offers the benefits of MOLAP for aggregations without necessitating duplication of the underlying detail data.

Virtual cubes and partitions are other forms of hybrid OLAP that enable you to tailor cube storage alternatives to meet your needs.

5. Proposed System

5.1. Proposed System Overview

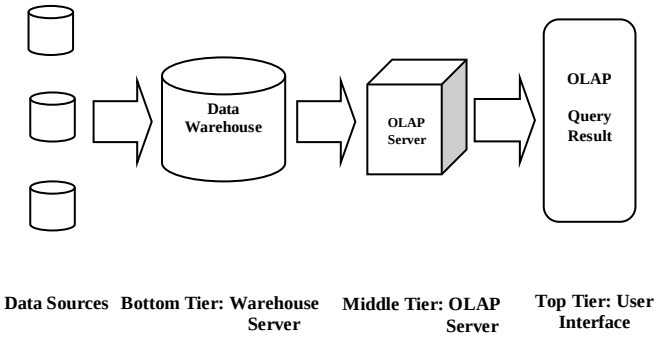

Figure 2. System Overview

5.2 System flow of the Proposed System

In this system, the analyst or manager of the Stationery Shop can utilize the following steps.

- When the system is run, the Admin Login Page will appear.
- Enter the Admin name and passwords.
- If the name and passwords are valid, choose the desired query option. If not , the system will return to the Admin Login Page.
- There are two main query options according to the time dimension and category dimension.
- If the analyst or manager chooses query option according to the time dimension, it will extract the Stationery sales report at user interested time period.
- The analyst or manager can also query another options.

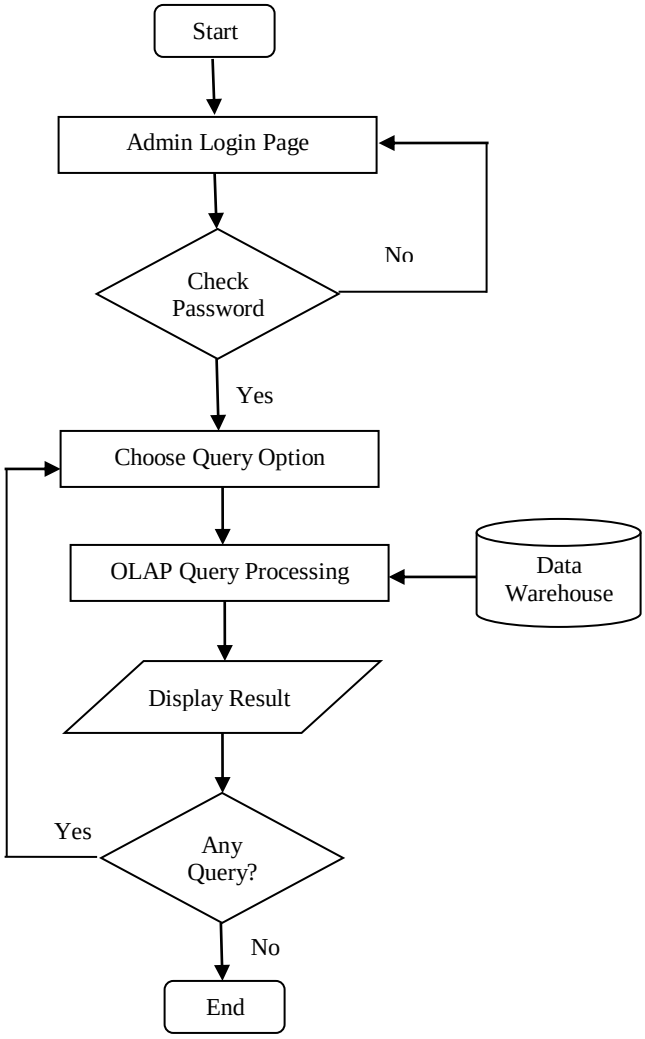

Figure 3. System Flow of the System

Typical OLAP operations used in this system are-

Roll-up: The roll-up operation (also called the drill-up operation) performs aggregation on a data cube, either by climbing up or a concept hierarchy for a dimension or by dimension reduction. For example, consider a sales data cube containing only the two dimensions location and time. Roll-up may be performed by removing, say, the time dimension, resulting in an aggregation of the total sales by location, rather than by location and by time.

Drill-down: Drill-down is the reverse of roll-up (eg. decreasing the level of aggregation/increasing detail).

Slice and dice: The slice operation performs a selection on one dimension of the given cube, resulting in a sub cube. The dice operation defines a sub cube by performing a selection on two or more dimensions.

Pivot (rotate): Pivot (also called rotate) is visualization operation that rotates the data axes in view in order to provide an alternative presentation of the data.

5.3 Star Schema for Proposed System

There may be data rich but information poor situations. To overcome this problem, the system will use Data mining tools. This system is to develop Stationery Shop analysis using OLAP. There are three specialized logical schemas for OLAP: star schema, snowflake schema and fact constellation schema. This system will implement with star schema. Stationery Shop data are complex. So the data store should be multidimensional.

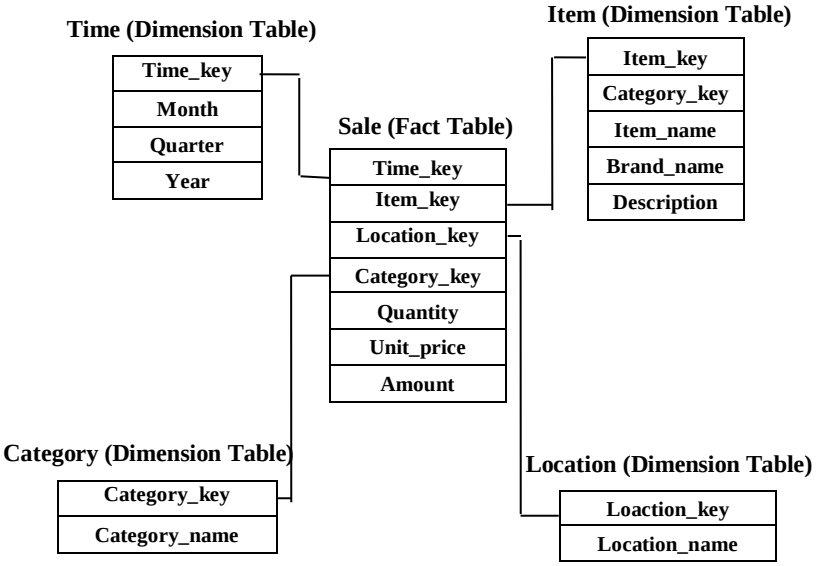

Figure 4. Star Schema for Data Warehouse

6. Implementation

Data warehouse and OLAP tools are based on a multidimensional data model. The goal of this system is to extract and analyze the sale items from the sale records of Stationery Shop data warehouse. Therefore, knowledge workers and decision makers can use this system sale to analyze the sales of interested item or interested category.

This system accepts users required and/or interested analyze or query or report and period as system input. According to these dimensions, required facts are retrieved from the fact table. The resulting information is visualized to the user in tables and graphs.

When the analyst or manager of the Stationery Shop query the stationery sales report according to the time dimension, the OLAP report dialog box will appear and choose the desired options.(for example category: book, brand: unicolor, name: 2-line 80-page book, location: Yangon, year: 2008). According to the user selected options, OLAP SQL query is generated. This SQL query is run on the warehouse and extracts the OLAP report result.

MONTH	NO OF SALES	TOTAL QTY	QTY PER SALE	TOTAL AMOUNT	AMOUNT PER SALE
January	7	28	4	11200	1600
February	7	28	4	11200	1600
March	10	52	5	20800	2080
April	4	15	3	6000	1500
May	6	33	5	13200	2200
June	9	35	3	14000	1555
July	9	41	4	16400	1822
August	8	35	4	14000	1750
September	5	26	5	10400	2080
October	11	44	4	17600	1600
November	5	24	4	9600	1920
December	4	19	4	7600	1900

Figure 5. OLAP Detail Report Result for Book Category of One Brand (Drill Down)

To emphasize only on category: book, location: Yangon and time: 2008, it is a roll-up operation of above table.

MONTH	NO OF SALES	TOTAL QTY	QTY PER SALE	TOTAL AMOUNT	AMOUNT PER SALE
January	10	87	6	33200	2100
February	13	96	6	47500	2600
March	24	123	7	67800	3000
April	12	93	5	45300	1500
May	11	90	5	36500	2200
June	22	110	7	56700	3200
July	15	100	7	52300	3000
August	18	108	7	54300	3100
September	25	130	7	76540	4500
October	21	115	7	57600	3300
November	15	100	7	48900	2400
December	14	99	6	48000	2800

Figure 6. OLAP Report Result for Book Category (Roll Up)

7. Conclusion

This system is to develop data mining process for user required and /or interested query or analyze or report concerning with sale records. This will be a useful tool to analyze the sale condition for stationery item and/or stationery category and which branch is the best seller. This On-Line Analytical Data Mining System will help the analyst or manager to know the sale conditions of each stationery category and branch conditions.

References

[1] A White Paper Prepared by Micro Strategy, "The Case for Relational OLAP"

[2]Abdourahamane FAYE, "From Data warehousing to Data Mining

[3]Berson, Alex & Smith, Ste phen J.,"Data warehousing, data mining & OLAP"

[4]Debecoise, Nelison Thomos,"The Data Warehouse Method"

[5]G.Spencer Brown, "Basic Data Warehousing Distinctions."

[6]Han, Jiawei & Kanber, Micheline, "Data mining concepts and technique"

[7]Leon Gong, Donna Venditti, DB2 Data Warehouse OLAP Services, "Part 1: Starting out with OLAP services released

[8]Logman, Addision Vesley,"Data Warehouse Method"

[9]Nikoskarayannidsi, Aris, Timos Sellia,"Processing Star Queries on Hierarchically-Clustered Fact Tables"

[10]Planet Earth Projects, Inc, "The Purpose of Data Warehousing

[11]Teorey, J.Toby, "Database Modeling and Design"