Journal of Information Systems & Information Technology (JISIT)
ISSN: 2478-0677; Vol. 6 No.2, 2021; [Pages 62- 71] -

Data Warehouse Implementation: Cost Effective Approach for Small Businesses

Ahamed Rameez Mohamed Nizzad [1], Mohideen Bawa Mohamed Irshad[2]

[1] Department of Computing, School of Computing, British College of Applied Studies
[2] Department of Management and Information Technology, South Eastern University of Sri Lanka

[1] nizzad@bcas.lk, mohamednizzad@gmail.com
[2] mbmirshad@seu.ac.lk, mbmirshad@gmail.com

Abstract

Data is the oil of this era where businesses compete among in order to own the most informative data for gaining competitive advantages and business success. It is where the necessity of using a sophisticated data warehouse arises. However, though there are high end Data warehouses and modern data analytics tools are in the market. It is empirical that most of the small businesses locally and internationally have not implemented a proper data warehouse considering the overhead burden of such an implementation and the technical difficulties involved. Therefore, in this study, it is proposed to provide a low cost, cost effective less complicated data warehouse implementation for any small business entities so as to enable them to leverage the benefit of their business data through the analytics to support the decision making and make the business data driven. The tools employed in this research are office automation tools with other freeware. The results suggest that it is worthwhile for small businesses to implement this architecture for their data analytics purpose. Further research in this area will enable automating the entire process.

Keywords: Data warehouse, Data Analytics, ETL Process, Small business, OLTP, OLAP

[Received: 4th June, 2021; Revised and Accepted: 20th December, 2021]

1 Introduction

Data is the oil of this era where businesses compete among in order to own the most informative data for gaining competitive advantages and business success. Therefore, managing data in the most efficient way is essential for the core existence of any business. It is where the emergence of data warehouse comes in. Data warehouse is a system that gathering and accumulating data at specific period of time from the OLTP database and then data modified into a dimensional or standard form of data repository [1]. Commonly data warehouse holds historical data and used of data query to business intelligence or analytical process. Data warehouse can be understood as a centralized storage which facilitate to collect information from several sources, and it manages data for efficient storage and retrieval to meet decision support and business intelligent requirements. Therefore, it is essential to have a data warehouse regardless of business domain or the size of the business. It is very much important to prepare data warehouse by using the proper design methodology and process. This is because data warehousing provides users with large amounts of clean, organized, and summarized data [2]. The basic idea behind the design and implementation of data warehouse involves in transforming data from Online Transaction Processing (OLTP) to Online Analytical Processing (OLAP). The process is known as ETL (Extraction, Transform and Loading). The process is outlined in Fig.1

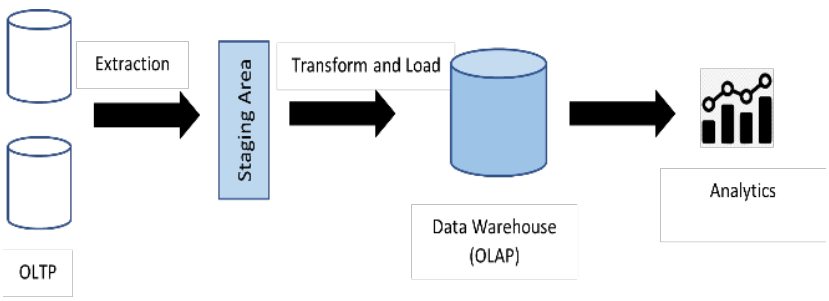

Fig.1. The Process of Extraction, Transform & Loading

[Received: 4th June, 2021; Revised and Accepted: 20th December, 2021]

In this study, an open-source dataset has been used from Kaggle [3] which is about the transaction data of a typical ecommerce website which has been illustrated in Fig. 2

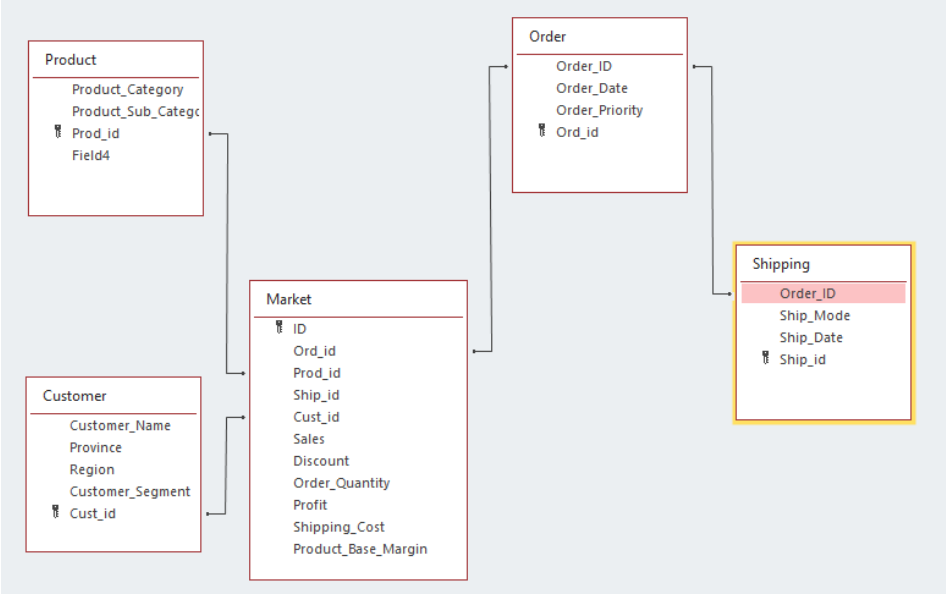

Fig 2. ER Diagram of the OLTP Database

The above ER Diagram has 5 tables namely Product, Customer, Order, Shipping and Marketing.

2 Literature Review

In this section previous works are reviewed in view of understanding the direction for further research and conduct this research as a continuation.

2.1 Data Warehouse design framework in view of Business Analyst

It is important to understand the outcome of the data warehouse in business analysts' (BA) point of view. Basically, having a sophisticated data warehouse will provide competitive advantage and would help make better decisions than businesses that do not employ any such data warehouse. Secondary, it will

[Received: 4th June, 2021; Revised and Accepted: 20th December, 2021]

increase the productivity of the organization due to the availability of data and description about the organization at present. In addition, data warehouse will reduce the overall cost with its ability to provide information related to the trends, patterns, exceptions, and other key information. Therefore, in this perspective, there are four different design views are to be considered when doing a data warehouse design [2]. Those views are expressed in the Table 1.

Table 1 - Four Different Views on Data Warehouse Design

Top-down view	It allows the selection of the relevant information necessary for the data warehouse. This information matches current and future business needs
Data Source view	This exposes the information being captured, stored, and managed by operational system. This information may be documented at various levels of detail and accuracy, from individual data source tables to integrate at various levels of detail and accuracy, form individual data source tables to integrated data source tables. Data sources are often modeled by traditional data modeling techniques, such as the E-R model or DASE tools.
Data warehouse view	This includes fact tables and dimension tables. It represents the information that is stored inside the data warehouse, including precalculated totals and counts, as well as information regarding the source, date and time of origin added to provide historical context.
Business Query View	This view is the data perspective in the data warehouse form the end-user's viewpoint

2.2 Data warehouse usage Information Processing

At the core, data warehouse is primarily used to generate reports and answering queries that we predefined and diverse in nature and requirements. At later stage, it can be used to analyze, summarize and detail data where the output can be presented in the form of reports, charts and so on. Finally, they can be used for strategic direction of the organization. Taking this into consideration, the tools for data warehouse can be categorized and chosen. As such the categories include Access and retrieval tools, database reporting tools, data analysis tools and data mining tools. Hence, there are three basic data warehouse applications such as Information Processing, Analytical Processing, and Data Mining [2].

[Received: 4th June, 2021; Revised and Accepted: 20th December, 2021]

- Information Processing supports querying, basic statistical querying, basic statistical analysis, and reporting using cross tabs, tables, charts or graphs. A current trend in data warehouse information processing is to construct low-cost web-based accessing tools that are then integrated with web browsers.
- Analytical Processing supports basic OLAP operations, including slice-and-dice, drill-down, roll-up, and pivoting. It generally operates on historic data in both summarized and detailed forms. The major strength of online analytical processing over information processing is the multidimensional data analysis of data warehouse data.
- Data Mining supports knowledge discovery by finding hidden pattern and association constructing analytical models, performing classification and prediction, and presenting the mining results using visualizations tools. Therefore, these are the three various data warehouse applications which will help to design and use of data warehouse.

3. Proposed Model

3.1 Setting up a Star Schema for the identified Database

At this point the OLTP database must be converted to OLAP Data warehouse design. Therefore, careful decision has to be made in order to create the most suitable Fact table and Dimension tables. Fact table is the key table in Data warehouse which contains measures that are numeric or quantitative. On the other hand, Dimension tables are subset of fact table. The dimension table measures are textual performance measurements of business [1]. In addition to these, there are instances where we need to introduce a separate key than natural keys in order to preserve the history of the records. In such instances, a unique sequential number known as surrogate key (SK) is used as the base key in the dimension tables. Using this SK, rows are uniquely identified in dimension table [4]. Below is the ER Diagram in Fig. 3 is the converted OLAP Data warehouse design.

[Received: 4th June, 2021; Revised and Accepted: 20th December, 2021]

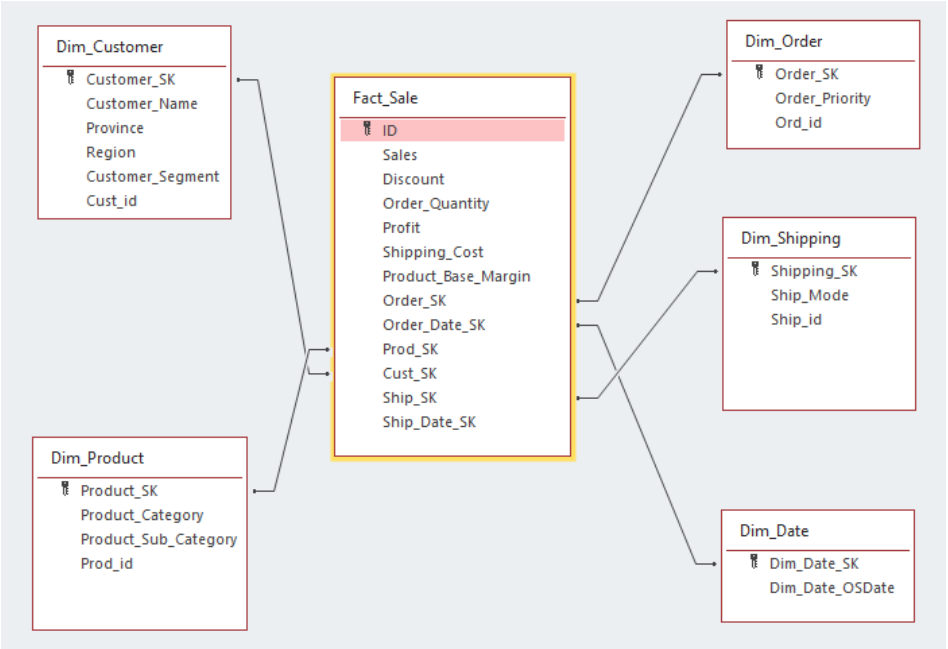

Fig. 3: ER Diagram for the Data Warehouse

Based on the above diagram, it has been identified that the sales table as the fact table and other tables are dimension tables where the entire Data warehouse has 1 Fact table and 5 Dimension tables as shown below in Table. 2

Table 2 - List of Fact Table and Dimension Tables

Fact Table	**Sale**
Dimension Table	Date
Dimension Table	Customer
Dimension Table	Product
Dimension Table	Order
Dimension Table	Shopping

Each dimension table accompany surrogate key as the primary key of each respective dimension table in order to preserve the history in case of any slowly changing dimensions (SCD) [5]. Then, all the primary keys of the

[Received: 4th June, 2021; Revised and Accepted: 20th December, 2021]

dimension tables are included in the fact table as the foreign key in order to create the relationship between fact table and dimension table and also to create link between the fact table and all the dimension tables. In this process, it had been notified that there are two different dates are involved in this database. They are order date and shipping date where it is possible to receive the order on a date and shipment on another date. Therefore, a dimension table Date has been employed to address this issue where the attribute Dim_Date_SK (Dim_Date) will serve as both Order date and Shipping date in the fact table as required.

3.2 Extraction, Transformation and Loading

In computing, extract, transform, load (ETL) is the general procedure of copying data from one or more sources into a destination system [6]. ETL in Data warehouse involves extraction of data from OLTP, transforming it to support the data warehouse design and then loading the data to the Designed data warehouse. There are different ETL tools are available for commercial user and non-commercial use. In this study, Visual importer ETL [7] is used which supports data transformation and loading from Excel to Access [8], Access to MySQL and so on.

Once the ETP process is completed, the data has been loaded to the MS Access. Once the data has been loaded to the MS Access, Microsoft Power BI [9], a business intelligence software by Microsoft has been used to do visual analysis of the identified the database.

4. Results and Discussion

The resultant impact was successful that all data had been loaded successfully to the Power BI where with the use of Power BI, the following expected questions were answered

- Region based profits
- Future demands in terms of products and categories
- Customer location-based purchasing pattern
- Cost of Goods, Profit Margin and related items

[Received: 4th June, 2021; Revised and Accepted: 20th December, 2021]

- Time based significance in terms of daily and monthly
- Minimum stock required on day basis and month basis

Though the potentiality is not limited to above identified questions that need to be answered through the model. However, a careful design and implementation will give a thorough understanding about the position of the business and key strategic decisions to be ahead of the competitors in the industry. An example visual report generated shown in Fig. 4 from Power BI for the case of answering the question that in which region the particular business has more profits.

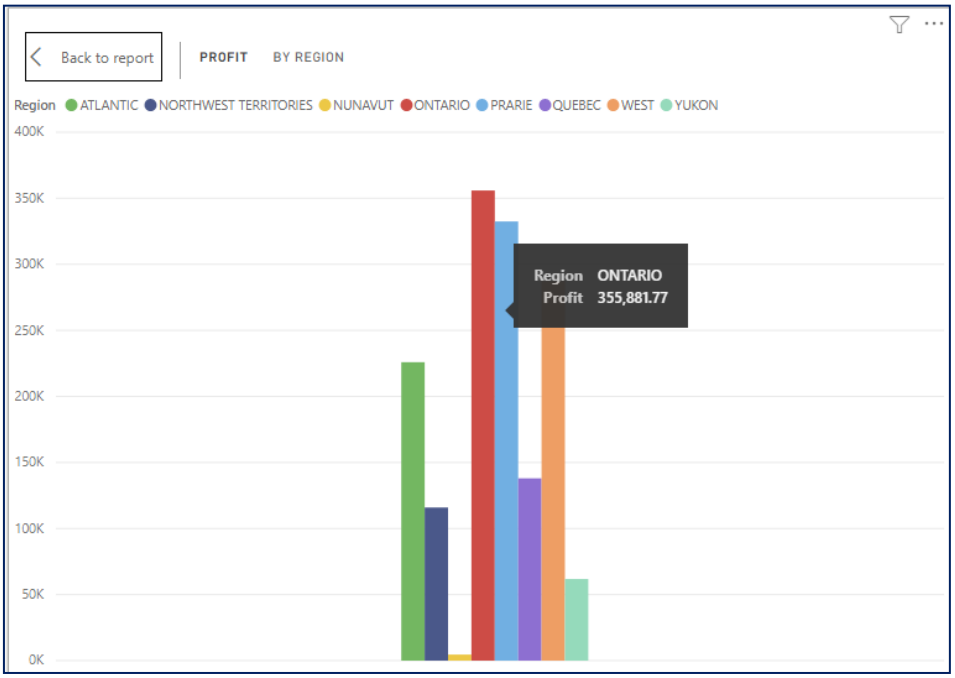

Fig. 4 - Region Based Profits

It is crystal clear that the Ontario region is the one that has more profits whereas the region Nunavut is having a very less profit close to negligible compared to Ontario.

In this case, it provides the clear facts for the decision makers whether to continue the business to the Nunavut region or they can close down the store on that region and focus on other regions where they can aim more profits. As stated at

[Received: 4th June, 2021; Revised and Accepted: 20th December, 2021]

the introduction, this kind of insights will set a new direction for the business which is data driven. Having above is just an example, the Data analytics with careful design and implementation will provide a definite advantage to any business that is focused on gaining long term benefit and competitive edge over others in the industry. Therefore, the below steps in Fig. 5, summarize the entire process involved in this project.

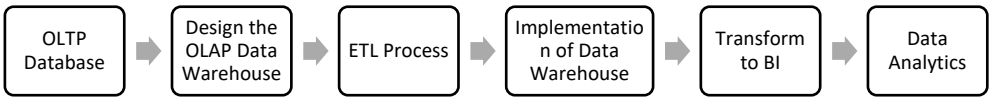

Fig. 5. Processes involved in implementing the proposed Data Warehouse

5. Conclusion and Future Works

The proposed architecture best fits for the small-scale businesses where a low-cost solution for solving the core data analytics issues is met. Therefore, based on this implementation following conclusion has been drawn.

1. Research should focus on reducing the query and response time in order to receive quick response from the proposed system.

2. Star Schema is ideal for the above purpose than snowflake as

- Star schema is denormalized and design flexible than snowflake which is easier for any small businesses to adopt.
- In Star Schema the query is very simple and easy to understand, while Snowflake Schema is more complex due to multiple foreign key which joins between dimension tables.
- Star Schema performance is better than snowflake. Database engine can optimize and boost the query performance based on predictable framework, while Snowflake Schema is more foreign key joins; therefore, longer execution time of query in compare with star schema.
- Star Schema has fewer joins while Snowflake Schema has higher number of joins due to its normalization concept.

Therefore, based on the study and review of literature, though data warehouse is well advanced than those days, it is evident that there are lack of adoption and

[Received: 4th June, 2021; Revised and Accepted: 20th December, 2021]

reluctance to use small business entities. Future studies should focus on traditional or standard theoretical studies on this aspect such as Technology Acceptance Model (TAM), Perceived Ease of Use (PEOU), and Perceived Usefulness (PU) to shed light to better understand the issues and challenges surface among small businesses to propose an ideal model of Data warehouse for them.

In this study, the researchers focused on proposing a model using well known office automation software which is very likely to be used by small businesses at any point and comparatively less expensive. However, analytics tool is less utilized compared to other office automation tools. Therefore, training providers can focus on devising a short-term course on how to utilize business analytics tools such as Power Bi for small businesses.

References

[1] T. Oketunji and O. Omodara, “Design of Data Warehouse and Business Intelligence System,” Master Thesis, no. June, 2011.
[2] D. Mankad and P. Dholakia, "The Study on Data Warehouse Design and Usage," International Journal of Scientific and Research Publications, vol. 3, 2013.
[3] Kaggle, "Kaggle," Kaggle, [Online]. Available: https://www.kaggle.com/datasets. [Accessed 08 10 2020].
[4] K. Orr,” Data quality and systems theory”. Communications of the ACM, 41(2), 1998. pp. 66-71
[5] R. Kimball and M. Ross, The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 1996.
[6] M. Denney, "Validating the extract, transform, load process used to populate a large clinical research database," International Journal of Medical Informatics., vol. 94, pp. 271-274, 2016
[7] D. S. L. LTD, "Visual Importer ETL - Professional," DB SOFTWARE LABORATORY LTD, [Online]. Available: https://www.etl-tools.com/visual-importer-etl-professional/overview.html. [Accessed 10 11 2020].
[8] Microsoft, "Microsoft Office," Microsoft, [Online]. Available: https://www.office.com/. [Accessed 10 11 2020].
[9] Microsoft, "Micrsoft Power Bi," Microsoft, [Online]. Available: https://powerbi.microsoft.com/en-us/. [Accessed 10 11 2020].

[Received: 4th June, 2021; Revised and Accepted: 20th December, 2021]