

Research on the Advantages of Data Warehouse in Avoiding Delay and Reducing Data Complexity

ICO Carlos Attai

Computación Centro Universitario UAEM Valle de Chalco

ABSTRACT

We have a major challenge when migrating data to a data warehouse is to preserve the same investment you have thought of from the beginning, to store the same and to maximize capacity with a minimum of interruption when making this migration for the business. In this paper, the data warehouse was carried out within the facilities to avoid latencies and reduce risks in the complexity of the data, since there are also variable costs in the network in different regions that would affect having it on network lines, improved performance as choosing or processing very robust data could be done with Analytics database and apart could make use of our own hardware. The main advantage lies in the structures of how it stores information as it can be stored by models of star boards, snowflake, relational cubes, among others. We show some of the other advantages that data warehouse.

KEYWORDS

Data warehouse; Petabytes; Business intelligence.

1. Introduction

The data warehouses first appeared in 1980 but the data warehouse arose in 1985, was built by William H. Inmon (Bill Inmon) built the first warehouse for a bank in Colorado, the purpose was to help operating systems with data flow and decision-making support (Decision Support System - DSS), to help integrate and organize data that was scattered across multiple databases, The first warehouses required a huge amount of redundancy as most companies had multiple DSS environments that served their users. DSS used much of the data, data collection, cleaning and integration, and often replicated for each system or environment.

In the period from 1990 to 2000 the concept of data warehouse has been improved which brings more storage capacity, new data query tools and internet connection or an intranet. The data warehouse is a type of unified data management or repository, is normally hosted on the server of a company or in the cloud which is made or designed to be able to solve and solve business intelligence activities in order to integrate and debug information from one or more different sources, so that they can be processed to allow analysis from different perspectives and with high response speeds, it is nourished by different sources including internal operational

systems through ETL(Extract, Transform and Load).The main advantage lies in the structures of how it stores the information since it can store it by models of star tables, snowflake, relational cubes, among others.

We show some of the other advantages of the data warehouse.

- They facilitate access to a wide variety of data
- Provides key information for decision-making in the enterprise such as trend or exception reports and these data are useful in the medium and long term
- Allows for working together and increasing the operational value of applications because if data sources are clear they are easy to install, contributes with greater flexibility and speed in the access to information
- Useful for storing analysis and historical queries

Disadvantages of the data warehouse is the high price that it can cost throughout its instance or useful life because it needs constant maintenance.

The history of data warehouses begins with the evolution of information systems and systems for decision-making support.

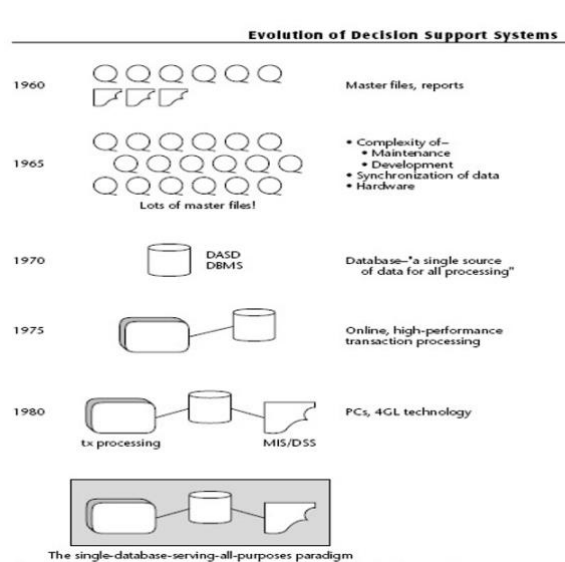


Figure 1. The early evolutionary stages of the architected environment (carlosproal, 2020).

The architecture of a data warehouse is based on five levels:

Data Source Layer - This is all the data sources that come in to enter the Data warehouse. **Layer ETL(extract-transform-load)**- It is all the process that takes with the data warehouse as the extraction of data from the sources already obtained, transform the data to load them to the data warehouse.

Storage layer- After getting them, transforming and extracting them is where they are stored. **Data marts layer** - These are the smallest warehouses as they allow greater accessibility to data warehouse information.

Data exploitation layer - It is the front-end client layer that presents the results of the analysis using data visualization tools, allowing data queries such as data marts and data warehouse.

The life cycle of a data warehouse has several stages which begins with its planning this is based on seeing if they have the solid or need of a data warehouse within the company will see the scope and improvements that can provide such a process and see economic resources, technicians and humans.

The second step for the cycle is to define the requirements that are carried in the business with the use of the data warehouse, this can be validated with the help of some users and see the satisfaction or deficiencies that are taken in the course of the process.

Then we have the technology part, the data part and the BI application part.

The technology has to do with the tools that will be used software, hardware, etc., the data is the creation of a diagram where the data warehouse will be stored, having the diagram the data will be created in the database, tables, views, etc. The design and structure of the ETL processes will then be created, and they will be responsible for loading the data from the data warehouse. The final process is that of deployment and production.



Figure 2. The life cycle of the data warehouse (Calzada, 2020)

2. Methodology

In this research it is of documentary type, Qualitative where the modeling of a structure that carries the data warehouse, its impact on the systems of large companies, consortia and scientific institutions as well as the efficiency of data consumption as it optimizes and guides the data of the organizations for their best exploitation. Also, determine in which cases it is opportune the use of this technology and see the strategies that companies use when applying this, the efficiency and growth as an entity as well as the results obtained. A database design (SQL) is created to be able to carry out the structure that implies the data warehouse, the scope of this modeling in terms of data storage, in this process can be done both physically and virtually.

With this process of data warehouse, it is intended to facilitate the consultation of data by users, since it is easier to generate statistical analyses, reports, qualitative analysis, projections, including for the purpose of better strategic decisions by companies or users. The study of the data warehouse is aimed at large companies as they contain many more consumable data which need to have a good integration, protection and can be optimized to make faster and fluid the use of such data.

Applications begin to accumulate a lot of information, terabytes of big data collected from multiple sources and this growth requires more tasks, have several tables or database can become a big problem by the fact of making queries to such databases or tables without doing the analysis in a fluid way causes this to become slow and unsafe and the use of data warehouse is highly recommended to guarantee the users the management of the system and the database.

Storage data warehouse can be set up on a corporate server or in a cloud warehouse, the storage in the cloud is becoming more common because it is more practical to be able to access the data, it is cheaper, it is easier and faster to install, cloud service providers are constantly updated for data protection. The storage of data on a local server, there are no long times for data query, by servers that are away from the client as the cloud, has its own security control.

3. Results

There are two DW methodologies that are imposed that are those of the Kimball and Inmon.

Inmon described a data warehouse as a topic-specific, integrated, time-variant, non-volatile data collection that supports the decision-making process. (PowerData, 2020) Kimball described the data warehouse as a copy of a data mart (Data Warehouse oriented to a specific area) that describes an event, “the change as a result of a transaction”.

Mentioning in the two main methodologies presented, the two parts are distinguished: that of the relational database which contains all the data to form the data warehouse.

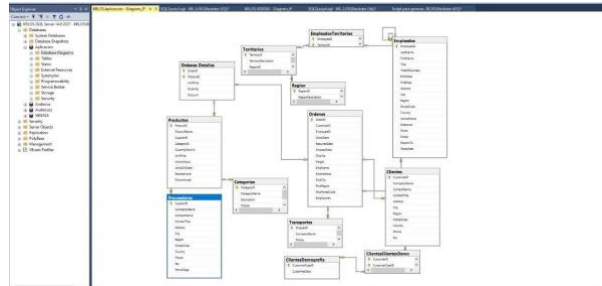


Figure 3. Diagram Entity Database

According to the data warehouse is set up in a system for a company which is seen to decrease tables and is faster access to queries as it can access data in a fluid manner, fast and remote, is a successful management for any company of large mass of information both for production and management reasons for efficient responses to customers using in this case such application.

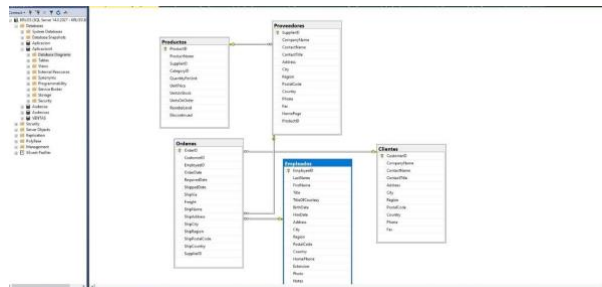


Figure 4. Simplified diagram of the Database.

4. Discussion

We also have a major challenge when migrating data to a data warehouse is to preserve the same investment you have thought of from the beginning, to store the same and to maximize capacity with a minimum of interruption when making this migration for the business. In this section, the data warehouse was carried out within the facilities to avoid latencies and reduce risks in the complexity of the data, since there are also variable costs in the network in different regions that would affect having it on network lines, improved performance as choosing or processing very robust data could be done with Analytics database and apart could make use of our own hardware. The main advantage lies in the structures of how it stores information as it can be stored by models of star boards, snowflake, relational cubes, among others. We show some of the other advantages that data warehouse.

- Facilitates access to a wide variety of data

S.L., Sinergia e Inteligencia de Negocio. (25 de Noviembre 2020). Sinnexus. Obtenido de Manual de BI:
https://www.sinnexus.com/business_intelligence/datawarehouse.aspx
Sierra, Y. (25 de noviembre de 2020). MediaCloud. Obtenido de MediaCloud ADN:
<https://blog.mdcloud.es/dwh-ejemplos-arquitectura-data-warehouse/>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.
This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution
license (<http://creativecommons.org/licenses/by/4.0/>).