Survey paper on data warehouse architecture

Shubha Singh
shubhasingh815@gmail.com
Babu Banarsi Das Institute of Technology, Ghaziabad, Uttar Pradesh

ABSTRACT

A data warehouse is a storage which is used to store the data. Data warehouse and Data mining are technologies that are used to deliver valuable information for making a decision. This survey paper defines the architecture of data warehouse and different types of data warehouse, which supports the many colleges and universities in making a decision. Types of the data warehouse are used in education to extract transform and load the data. All these techniques improve the benefits of the data warehouse in the education system.

Keywords— Data Warehouse, ETL (Extract, transform and load) System, Online Analytical Processing (OLAP), Decision Support System

1. INTRODUCTION

A data warehouse is a relational or multidimensional database designed for query and analysis [4]. They are not optimized for transaction processing which is a domain of OLTP system. A data warehouse is storage where all types of data are stores and processes when anyone requests for that data [1].

A data warehouse usually consolidates the historical and transactional data derived from the multiple resources [4]. A data warehouse is used to the data for the long-term process.

2. DATA WAREHOUSE ARCHITECTURE

Data warehouse includes the following layers:

- Data Source Layer
- Data Staging Layer
- Data Storage Layer
- Data Presentation Layer

![Architecture of data warehouse](image)

Fig. 1: Architecture of data warehouse
2.1 Data Source Layer
Data source layer is a layer where the data is collected from the various external or internal sources resides in the relational database. Following types of data is stored in the Data source Layer [2].

- Operational Data
- Social media Data
- Third-party Data

2.2 Data Staging Layer
Data staging layer resides between data warehouse and data sources. In this layer, data is extracted from the different External and internal sources. Multiple technologies and tools are used to at this layer to extract and load the data. Once the extracted data is loaded it is sent to higher quality check [2].

The staging layer contains the following components:-

- Landing Database and Staging Area
- Data Integration Tool (ETL)

2.3 Landing Database and Staging Area
The landing databases are used to store the data retrieved from the data source and staging is used to apply quality checks on the data before moving it to the data. Staging is an important step in the data warehouse.

2.4 Data integration Tool
Extract, Transform and Load is a data integration tool which is used for integration of the data. The process of extracting the data from the source system and load it into the data warehouse is called ETL [2].

2.5 Working of Extract, Transform and Load (ETL) Process of Data Warehouse
The process of ETL involves the following steps in extracting the data from external resources and load it into the data warehouse.

- Extraction of data: During the extraction of data we extract the data from the different external sources such as a database, application, and another repository system. There are a large amount of data stored at the external sources and extraction process is used to extract the meaning full data from the external sources [3].
- Transportation of data: After extracting the data from the external sources we have to transfer it to the data warehouse. We transport the data to the data warehouse. The transportation of data is physical. After transforming the data we use the load process.
- Loading of data: After transforming we load the data in the data warehouse. The process of loading and storing the data in the data warehouse is the last process of ETL.

2.6 Data Storage Layer
Data storage layer is a layer where the data is stored. The Data Storage may be Data Mart, Data Warehouse and Operational Data Store [2]. Data stored in data storage are used for future processing.

2.7 Data Presentation Layer
Data Presentation Layer is a layer where users interact with the Data warehouse. At this Layer, the user provides a query to interact with the data warehouse when the user interacts with data warehouse it provides the data required to the users.

3. LITERATURE REVIEW
A decision support system is required for an education system which should gather the data (like student data, teachers data, financial data from different colleges and universities) and on the basis of that data perform decision (like no of student increases per year or not?, no of teacher increases per year, subject prefered by students ) [1].

In this paper, we discuss how the education-related decision support system uses the data warehouse techniques and OLAP to make a decision. Data of all the colleges/universities are stored in a data warehouse. At the time of designing and management of metadata. They use warehouse studio and for storing data they use Sybase IQ [5][6]. Request of the data is done by filling the specific form. Once the data request is done the data warehouse tools read the query and sends the data according to the uses request. OLAP and Data Mart perform an important role in making a decision. So both have been described below

OLAP: OLAP stands for Online Analytical Processing. It is a computing methodology which enables the users to extract the selective data and analyze from the different point of view [8]. The operation performed by OLAP are given below:

- Roll-up
- Slice
- Dice
- Drill-down

Data Mart: Data mart is a subset of the data warehouse. It focuses on a single function of data. Similar types of data are stored in a single data mart which is easy to use. If we want to access the similar types of data then we have to access the single data mart. It is very helpful in making a decision [9].

For making decision data warehouse provides the information which is required for making a decision and data mart is used to store the similar type of data at a single place so that we can easily access that data. ETL is used to extract the data from different
external sources and load in the data warehouse. OLAP also has an important role in making the decision for the university or college, it performs OLAP operation when we request for a specific data.

Guan et al. (2000) proposed Data warehousing is a key element in the “paradigm shift” that is needed in higher education decision making [7]. There is a sign that prompted this view is being changed, while survey the college/universities it was found that data warehouse is initially low in use but they are showing interest in the data warehouse to store their data so that they can store their data at a safe place for a long time for the future use.

There is a very little research on the data warehousing for making decision tree in colleges/universities. A decision tree is made for synchronization of data of any college or university. Decision-making planning requires a lot of analysis data for college/university. This survey paper provides facilities to utilize the data warehouse.

4. CONCLUSION
This paper presents a complete overview of the data warehouse. It describes the architecture of the data warehouse and the techniques that are helpful for making the decision for college/university. In this paper, we have described the technique which will be a benefit for college and university to implement the data warehouse. ETL tools are used in the data warehouse to improve the efficiency of the data warehouse. The future work on a data warehouse is to study on the OLAP operation and to enhance the use of the OLAP tools in the data warehouse. The data warehouse has a lot of scope in the education systems and other organization for storing their data. For being familiar with physical storage; we have to use Oracle, SQL, MySQL, web mining etc; in a data warehouse so that students can easily interact with the data warehouse.

5. REFERENCES
[1] https://pdfs.semanticscholar.org/bc2f/2a5cc55fb1c4555e31b42aaabf540c4e8451.pdf
[7] https://digitalcommons.andrews.edu/cgi/viewcontent.cgi?article=1432&context=dissertations
[8] https://searchdatamanagement.techtarget.com/definition/OLAP
[12] https://pdfs.semanticscholar.org/i700/2215fa1991347ac97bc2f74e8311a021c1d0.pdf