IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Compare Data Warehouse to Data Marts

Surya Lakshmisri

Department of Information Technology

Data Warehouse

It is a relay database system that stockpiles historical data as w el as metadata from a process system for efficient reporting and examination of the same. It encompasses a wide variety of data across numerous subject areas. The data is uploaded from diverse systems prior to manipulation. It is used by managers to assist in decision making. The data warehouse environment is made up of extraction, transportation, transformation and loading solution, an online analytical processing engine. Additionally, client analysis tools as well as other system applications that control the course of data gathering, ultimately conveying the results to business users (Zacek & Hunka, 2015). The characteristics of an efficient data warehouse facility include:

- Subject orientation. They are designed to give results pertaining to a particular topic matter. In order to know the sales performance of an entity, the warehouse system defines the data related to sales such as the region that offers the largest sales numbers. It will aid decision making on segmentation of the markets for better customer experience.
- Integration. It is the combination of different sources into a consistent format. The units of measure should conform to the nature of the result needed for strategic decisions.
- Non-volatile. It means that, upon entry, the data input must be the same as the output. It will assist analysis of the transactions.
- Time-variant. The information should represent active trends that are deciphered through historical data. It should be a reliable medium for focusing on changes experienced in operations over time.

The data warehouse is typically housed on a mainframe server and in recent times, in the cloud. It eases data selection for handling user queries.

It upholds a copy of the information from transactions in order to:

- Collect data from several sources into a single database so that a particular query engine is utilized when presenting data.
- Alleviate the dilemma of database isolation levels in transaction processing systems whenever the attempts to run extensive analysis queries
- Preserve data history for future reference
- Amalgamate data from several systems for better and centralized scrutiny across all fields
- Perks up the quality of data through the presentation of consistent codes and depictions as well as repairing corrupted data services.
- Add the significance of operational entity activities such as customer relationship management.

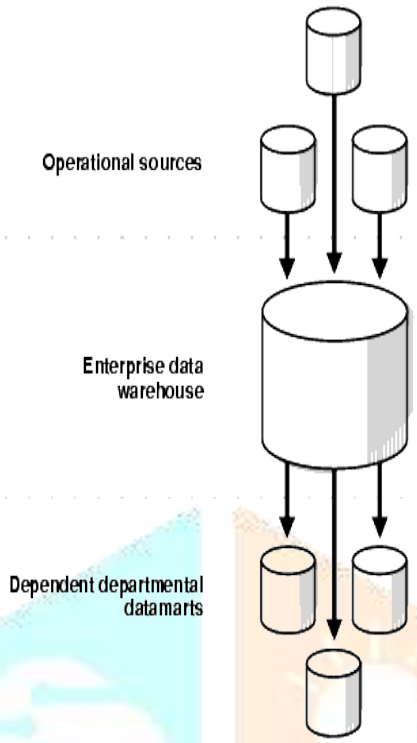
There are two approaches to storing data in a data warehouse-dimensional approach and the normalized approach. The dimensional approach was proposed by Ralph Kimball. It partitions transaction data, which are generally numerals (facts) and dimensions that are the reference points. For example, a sales transaction could be broken up to the number of units sold as well as price charged for the same. On the other hand the normalized approach proposed by Bill Inmon, and states that the data warehouse should be modeled by use of the third normal form (Chenoweth, Schuff & Louis, 2003).The primary problem addressed by a data warehouse is that the ultimate users have a difficult time producing impromptu queries and reports. It is necessary because obtaining the data usually requires a programmer to develop the report. Moreover, the data stores were designed for transaction processing and not ad –hoc systems (ARIYACHANDRA & WATSON, 2008).

Data Mart

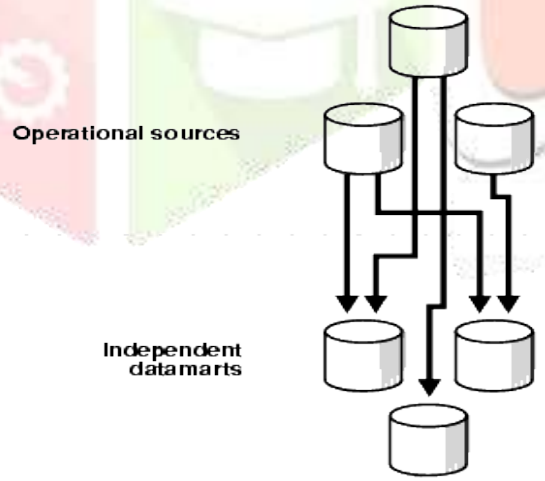
It is a straightforward structure of a data warehouse that focuses on a particular and distinct functional area. It is designed to serve a fastidious community of knowledge workers. These areas are termed as departments in the organizational aspect. They are often designed and organized to draw data from merely a few sources. These sources could be operational systems, centralized data warehouse or peripheral data. The need for data marts came about due to complexities, costs and long development period for undeviating implementation of data warehouses. Additionally, it is a decision support system that slots in part of the entity's data focusing on specific functions of the enterprise. They are software applications relevant whenever a firm wants to measure the impact of a new product offering, as well as forecasting performances. They have evolved to become a popular alternative to data warehouses.

There are three types of data marts, namely: dependent, independent and hybrid marts. These categories are based on the data source that feeds the data mart. Dependent data marts get data from centralized data warehouses that have already been created. Independent data marts -on the other hand, are stand-alone systems designed by directly sourcing data from external sources of data. Thirdly, hybrid data marts can draw records from both operational systems and data warehouses.

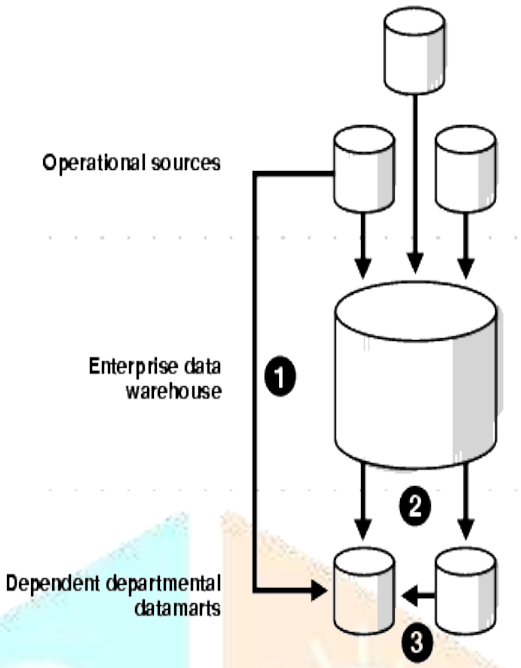
Dependent marts illustration



Independent data marts illustration



Hybrid data marts illustration



The core differences between these types of data marts are the extraction-transformation-transportation (ETT) process, engaging in moving data from day to day activities, filtering it and loading it to the data marts.

Since the data marts are intended to view data in an inimitable, the process always starts with the analysis of client needs. Data virtualization software is applicable when creating implicit data marts, combining them with other data necessary to conform to the requirements of specific companies. It affords the firm's database management team a level of control over the organization's records throughout its lifecycle. (Chenoweth, Schuff, & Louis, 2003).

Steps in implementing a data mart

- Designing. It swathes all tasks done from the moment of initiating the need for a data mart up to the information gathered and development of the logical as well as physical blueprint of the data mart. The activities involved in this stage are as follows: gathering the technical requirements, identifying data sources, selecting appropriate detachment of data and finally the framework of the data mart.
- Constructing. It creates the physical database associated with the data mart to give fast and efficient access to the components. It involves creating diagram representations such as tables and indices. Lastly, the developer should determine the best way to set up the access structures.
- Populating. It encompasses all undertakings related to getting data, editing it and input to the data mart. It involves plotting data sources to target data structures, extracting data, cleansing data, loading data onto the data mart and finally storing the metadata.
- Accessing. It means setting up a transitional layer to translate database structures into business terms for better interaction. Additionally, the maintenance and management of interfaces is done here.
- Managing. The data mart is checked over its lifetime. The relevance of the applications is established for providing secure access to data, optimizing the system for better performance and finally ensuring availability of data even in the event of system failures.

Differences

The two methodologies of data manipulation and analysis differ significantly in terms of application and characteristics. Both the methodologies have their security modeling aligning the beat practices in agile environment (SM Mohammad & S Lakshmisri, 2018). The following are dissimilar properties of these database management systems:

- Scope. A data warehouse is primarily applied in the corporate sector of business. It is a vital aspect when designing a long-term strategic plan and guides on the implementation of various projects. It is therefore suited to large organizations with complex endeavors and processes. On the other hand, data marts are applied in distinct lines of businesses. It signifies the relevance of its use in internal entities. It examines the

firm's position within an industry. It is therefore perceived to be an integral component of running an enterprise.

- Subjects. A data warehouse can take up several forms of applications. The data mined from this undertaking is derived from a wide variety of sources, provided they will give an accurate reflection of the operational activities. However, data marts deal solely with the singular subject matter involved. The application provides feedback on the component itself, with a narrow view of the problem at hand (Jonker, 2011).
- Data sources. A data warehouse has several sources of input to assist in decision making whereas data marts give feedback when supplied with data from a select few sources. It enables accurate and acute solutions for implementation. Furthermore, the number of sources has no impact on the relevance of the information derived from this venture. Data sources can be integrated with continuous integration and continuous deployment pipeline (SM Mohammad, 2016).
- Size. A data warehouse makes use of a large data field to provide meaningful information. All components related to the day-to-day operations of an entity are considered to make informed decisions. It is not a cheap option for cash-strapped firms. The minimum quantity of data to be used varies from approximately one hundred gigabytes onwards. On the other hand, data marts need slightly fewer labored data to provide meaningful information. It works best with data less than a hundred gigabytes of space.
- Implementation time. Data warehouses take up to several years of examination of the tenets of successful operations. The system is a long-term approach when generating information valuable to top management and other institutions like the government. Data marts the other hand applied few months order give accurate. Even though agile software development methodologies are more and more common, many organizations have discovered that they have various departmental features practicing in data warehouse to data marts (SM Mohammad, 2017).

References

ARIYACHANDRA, T., & WATSON, H. J. (2008). Which Data Warehouse Architecture is Best? Communications Of The ACM, 51(10), 146-147.

Chenoweth, T. Schuff, D, & Louis, R. S. (2003). METHOD FOR DEVELOPING DIMENSIONAL DATA MARTS. Communications Of The ACM, 46(12), 93-98. doi:10.1145/953460.953465

Farooq, F. (2013). The data warehouse virtualization framework in operational business intelligence. Expert Systems, 30(5), 451-472. doi:10.1111/j.1468-0394.2012.00651.x

Jonker, D. (2011). Data Marts Can't Dance to Data's New Groove. Event DV, 24(10), 10-12.

Maté, A. & Trujillo, J. (2014). Tracing conceptual models' evolution in data warehouses by using the model-driven architecture. Computer Standards & Interfaces, 36(5), 831-843.

Sikender Mohsienuddin Mohammad, "DEVOPS AUTOMATION AND AGILE METHODOLOGY ", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.5, Issue 3, pp.946-949, August-2017, Available at :http://www.ijcrt.org/papers/IJCRT1133441.pdf

Sikender Mohsienuddin Mohammad, "CONTINUOUS INTEGRATION AND AUTOMATION", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.4, Issue 3, pp.938-945, July 2016, Available at :http://www.ijcrt.org/papers/IJCRT1133440.pdf

Zacek, J. & Hunka, F. (2015). Applying ETL Fuzzy Filter for Data Warehouse Size Minimization. AIP Conference Proceedings, 1648(1), 1-4. doi:10.1063/1.4912768

Sikender Mohsienuddin Mohammad,Surya Lakshmisri , "SECURITY AUTOMATION IN INFORMATION TECHNOLOGY", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.6, Issue 2, pp.901-905, June 2018, Available at :http://www.ijcrt.org/papers/IJCRT1133434.pdf