International Journal of Information Technology and Management

Vol. VIII, Issue No. XII, May-2015, ISSN 2249-4510

CONCEPTUAL FRAMEWORK OF DATA WAREHOUSING

AN INTERNATIONALLY INDEXED PEER REVIEWED & REFEREED JOURNAL

www.ignited.in

International Journal of Information Technology and Management
Vol. VIII, Issue No. XII, May-2015, ISSN 2249-4510

Conceptual Framework of Data Warehousing

Ms. Pooja Sharma

Assistant Professor, Govt. College, Ateli Mahender Garh

Abstract – One recent advance in database management is data warehousing. Data Warehouses are an important asset for organizations to maintain efficiency, profitability and competitive advantages. The basic premise of data warehouse is "to create a clearinghouse in which to gather and organize critical business data. It is a store of integrated data obtained from various internal and external sources, which represents events or facts as of a given point in time. The data collected have degrees of value and business relevance. As data is collected, it is passed through a 'conveyor belt', call the Data Life Cycle Management.

----------X----------

INTRODUCTION

One recent advance in database management is data warehousing. Data Warehouses are an important asset for organizations to maintain efficiency, profitability and competitive advantages. The basic premise of data warehouse is "to create a clearinghouse in which to gather and organize critical business data. It is a store of integrated data obtained from various internal and external sources, which represents events or facts as of a given point in time. The data collected have degrees of value and business relevance. As data is collected, it is passed through a 'conveyor belt', call the Data Life Cycle Management.

DATA LIFE CYCLE MANAGEMENT

Data becomes active as soon as it is of interest to an organization. Data life cycle begins with a business need for acquiring data. Active data are referenced on a regular basis during day-to-day business operations. Over time, this data loses its importance and is accessed less often, gradually losing its business value, and ending with its archival or disposal.

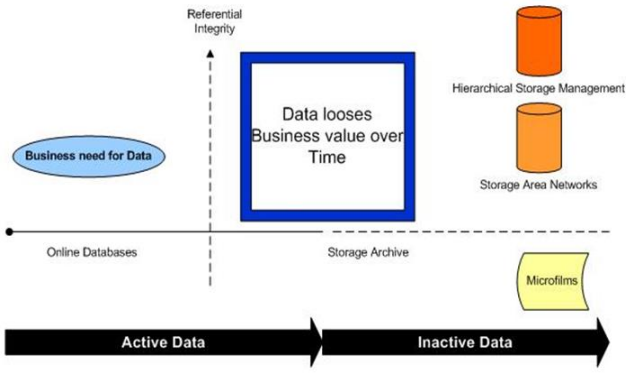

Figure 1: Data Life Cycle in Enterprises

ACTIVE DATA

Active data is of business use to an organization. The ease of access for business users to active data is an absolute necessity in order to run an efficient business. The simple, but critical principle that all data moves through life-cycle stages is a key to improve data management. By understanding how data is used and how long it must be retained, companies can develop a strategy to map usage patterns to the optimal storage media, thereby minimizing the total cost of storing data over its life cycle.

INACTIVE DATA

Data are put out to pasture once they are no longer active i.e. there are no longer needed for critical business tasks or analysis. Prior to the mid-nineties, most enterprises achieved data in Microfilms and tape back-ups. There are now technologies for data archival such as **Storage Area Networks (SAN),Network Attached Storage (NAS**) and **Hierarchical Storage Management**. These storage systems can maintain referential integrity and business context.

DATA WAREHOUSING INFRASTRUCTURE

The goal of Data Warehousing is to generate front-end analytics that will support business executives and operational managers.

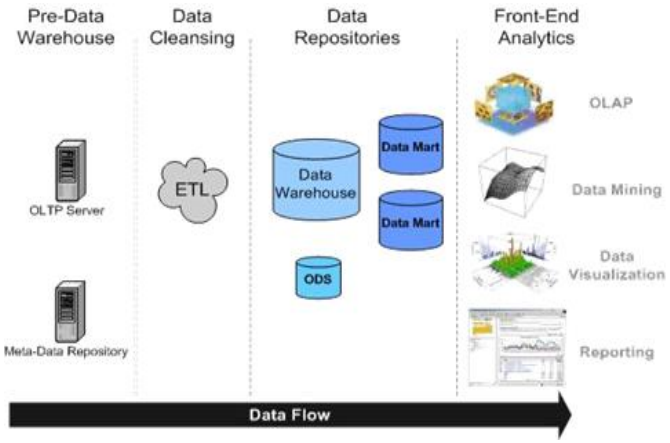

Figure 2: Overview of Data Warehousing Infrastructure

Pre-Data Warehouse

The Pre-Data Warehouse zone provides the data for data warehousing. Data Warehouse designers determine which data contains business value for insertion.

OLTP databases are where operational data are stored. OLTP databases can reside in transactional software applications such as Enterprise Resource Management (ERP), Supply Chain, Point of Sale, and Customer Serving Software. OLTPs are design for transaction speed and accuracy.

Metadata ensures the sanctity and accuracy of data entering into the data lifecycle process. Meta-data ensures that data has the right format and relevancy. Organizations can take preventive action in reducing cost for the ETL stage by having a sound Metadata policy. The commonly used terminology to describe meta data is "data about data".

DATA CLEANSING

Before data enters the data warehouse, the extraction, transformation and cleaning (ETL) process ensures that the data passes the data quality threshold. ETLs are also responsible for running scheduled tasks that extract data from OLTPs.

DATA REPOSITORIES

The Data Warehouse repository is the database that stores active data of business value for an organization. The Data Warehouse modeling design is optimized for data analysis. There are variants of Data Warehouses - Data Marts and ODS. Data Marts are not physically any different from Data Warehouses. Data Marts can be thought of as smaller Data Warehouses built on a departmental rather than on a company-wide level.

Data Warehouses collects data and is the repository for historical data. Hence it is not always efficient for providing up-to-date analysis. This is where ODS, **Operational Data Stores**, come in. ODS are used to hold recent data before migration to the Data Warehouse. ODS are used to hold data that have a deeper history that OLTPs. Keep large amounts of data in OLTPs can tie down computer resources and slow down processing - imagine waiting at the ATM for 10 minutes between the prompts for inputs.

FRONT-END ANALYSIS

The last and most critical potion of the Data Warehouse overview are the front-end applications that business users will use to interact with data stored in the repositories.

OLAP, Online Analytical Processing, is used to analyze historical data and slice the business information required. OLAPs are often used by marketing managers. Slices of data that are useful to marketing managers can be - How many customers between the ages 24-45, that live in New Yorkstate, buy over $2000 worth of groceries a month?

Reporting tools are used to provide reports on the data. Data are displayed to show relevancy to the business and keep track of key performance indicators (KPI).

REFERENCES

1. Arif Ghafoor, “Web-based multimedia databases: prospects and challenges”, 2nd ACM international workshop on Multimedia databases 2004.

2. W. I. Grosky, Managing Multimedia Information in Database Systems, Comm. ACM 40,12 (Dec. 1997), pp. 73-80.

3. Hirzalla, N.B., Karmouch , A . - A multimedia query specification language, in Nwosu K., Thuraisingham B.,

4. Bruce Berra P., Multimedia Database Systems, Design and Implementation Strategies, Kluwer Academic Publishers, 1996.

5. Khoshafian, S . - Multimedia and Imaging Databases, Morgan Kaufmann, 1995

6. Faloutsos, C. (2005), Searching Multimedia Databases by Content. Springer

7. Budford John F., Multimedia Systems, Addison- Wesley, ACM Press.1944.