Bindia / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue 3, May-Jun 2012, pp.2988-2990

Metadata Repository Architecture based on Lifecycle of DW 2.0 Environment

Bindia*
*(M.Tech, Lovely Professional University,
Computer Science Department Phagwara, Punjab, India

ABSTRACT

A data warehouse has become an important part of any organization. DW 2.0 recognizes the life cycle of data within the data warehouse, need for having textual data in the DW and recognizes that metadata is an essential component of the data warehouse environment. In first generation data warehousing, metadata was ignored. But in today's world, metadata is very critical component of data warehouse 2.0 environments. For management of metadata in life cycle of Data warehouse 2.0, we propose a new architecture regarding metadata repository.

Keywords – **Data warehouse, metadata repository, DW 2.0 environment, lifecycle.**

I. INTRODUCTION

Data Warehouses integrate data from multiple heterogeneous representations for decision support applications. Data Warehousing is a collection of decision support technologies, aimed at enabling the knowledge worker to make better and fast decisions [11]. Data Warehouse 2.0 was basically designed to solve problem of unstructured data that was present in first generation data warehouses.

The Data Warehouse 2.0 includes four life cycles sectors of data: Interactive Sector, Integrated Sector, near line Sector, and Archival Sector as shown in figure. The data enter in interactive system rapidly. When the data is settled, the data is sent to integrated sector. Till its probability of access declines, it remains in integrated sector. Most probably, the access declines after three to four years. Data is then move onto one of two sectors i.e. archival sector or near line sector. Near line sector is an extension of integrated sector. Basically it is an optional sector. There is no need to place the data necessary. The data is placed in near line sector only when there is large amount of data and the probability of access data is significantly changes. The data in archival sector has very less probability of access. Data in this sector is mostly seven to years old. This sector is fed by near line or integrated

sector. Also the probability of accessing data in this sector is very low.

Metadata plays a very important role in Data warehouse 2.0 environments and it is required in each sector. So all sectors will have metadata. Metadata in archival sector is different from others. Data enter in Data warehouse 2.0 environments either through ETL from other application or from direct applications that are housed in interactive sector.

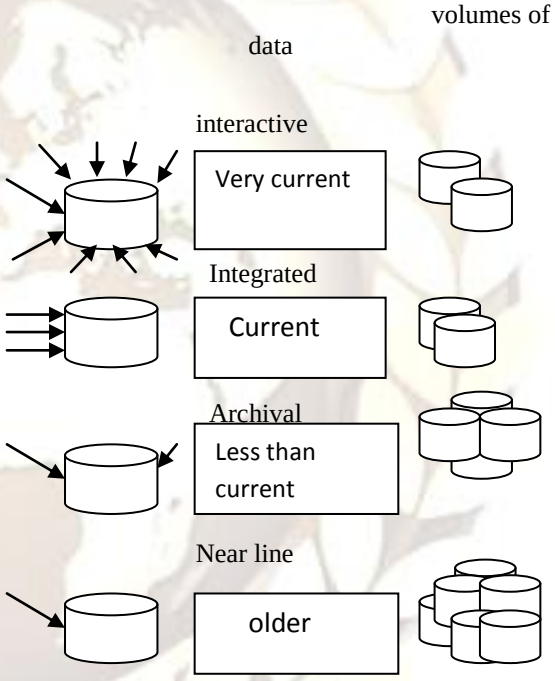

Fig.1: Life cycle of data within DW 2.0

II. EXISTING METADATA REPOSITORY ARCHITECTURE

All of the other sectors excluding archival sector have their metadata stored in a repository. A metadata repository contains all the data about data warehouse. The data is collected from various sources and compiled in repository that provides centralized location and maintain the organization's data. This data from repository can be accessed from any location. A metadata repository can be passive or active. A passive repository acts as a reporting hub that takes metadata from multiple tools as read only information for use in metadata reporting functions

such as data lineage and impact analysis. An active metadata repository manages the metadata for applications or pushes metadata out to applications and databases.

The active repository is the place where enterprise metadata is kept when the metadata is actively used during the development process. A passive repository is one in which the metadata does not interact in any direct manner with the development.

Existing metadata repository architecture creates a uniform and consistent model that organizes metadata to be stored in global Meta data repository and shared local data in local repository. All that shared metadata present in many repositories have to go through the central global repository, but the problem remains i.e. sharing and to have access to local metadata are independent of central repository. This architecture is called decentralized metadata repository architecture given by Macro.D.

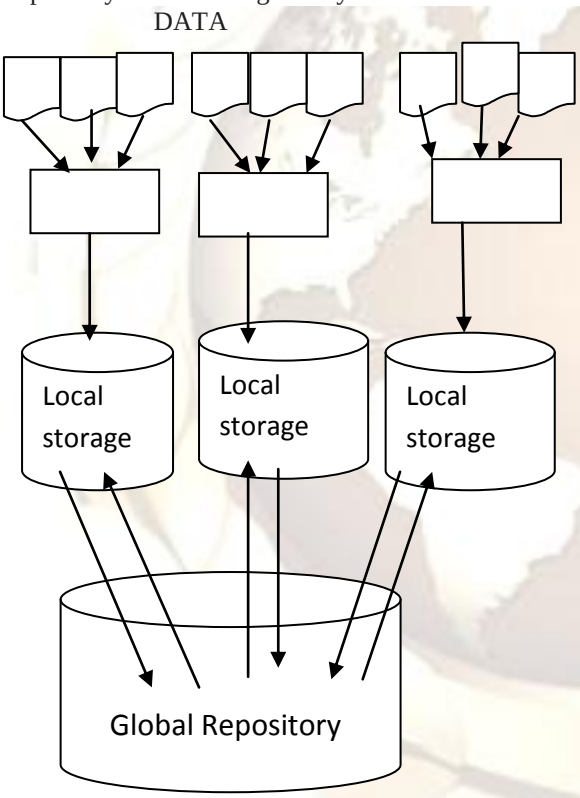

Fig 2: Existing architecture of metadata repository

III. PROPOSED ARCHITECTURE

This centralized metadata repository is not suitable for lifecycle of data in data warehouse 2.0. Here we develop Metadata Repository Architecture based on Lifecycle of DW 2.0 Environment that is decentralized also. Data access is very fast and random in this case. Also the response time can be measured in sub second.

In the interactive sector, Data can be added, deleted and modified. Also historical data is not much, so there is separate data repository. And that metadata repository is managed by metadata manager.

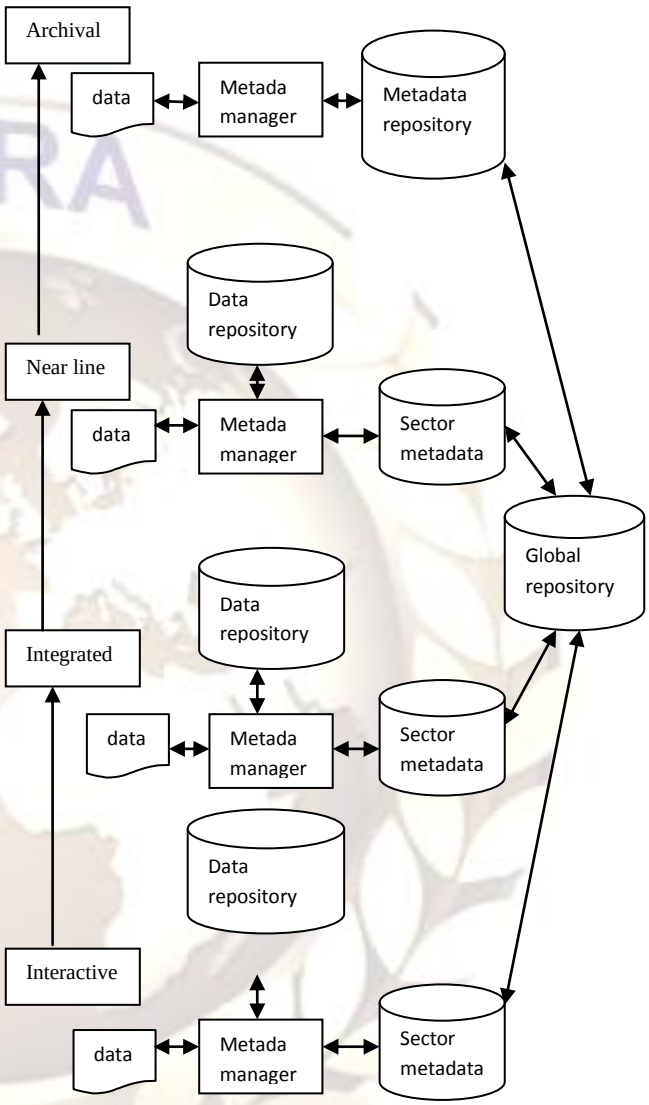

Fig.3: Metadata Repository Architecture based on Lifecycle of DW 2.0 Environment

In integrated sector, many conversions are done here. Subject oriented and summary data, continuous time span architecture, continuous snapshot data and profile data are present here in this sector. It contains historical and granular data because data comes from interactive sector is passed through ETL processing. Accesses of data is fewer than interactive

sector but request more data. Response time is from 10s to much in this sector. This sector is also having sector metadata and separate data repository.

The Near line sector is also having separate data repository and sector metadata. It is basically a form of caching for integrated data. When the integrated sector of data warehouse is very large, then only data with the low probability of access is stored in near line sector. The structure of near line sector is same as integrated sector, but the near line storage is sequential data storage on cartridges whose cost is less than disks. So here is also a independent sector metadata and data repository.

In archival sector, data basically comes from integrated and near line sector. The data present in archival sector is accessed very infrequently. Data can be stored sequentially to lower the storage cost only. Because it stores very large amount of data, It is very difficult to keep the metadata safe in separate repository. So in this sector metadata and data are managed in the same repository.

CONCLUSION

Metadata was ignored in first generation data warehouse that is now very critical component of data warehouse 2.0 environments. Without metadata, the data in DW2.0 would be a big pile of essentially useless data. Due to disadvantages of centralized metadata repository architectures, our proposed metadata repository architecture is developed which reduces the cost also. In addition to integrated data, it contains unstructured data in different forms, integrated data for both business metadata and technical metadata, online high performance data that can be updated, profile data records, summary and subject oriented data etc. the data structure of archival sector is different from other sectors. It is very difficult to keep the metadata safe in separate repository in case of archival sector, so in this sector metadata and data are managed in the same repository.

ACKNOWLEDGEMENT

I am very thankful to the IJERA for giving me opportunity to get my research completed. I am very thankful to my parents, who support me time to time in developing my research related work. They left no stone unturned to help me. Due to them, it becomes possible to complete my research. I also want to say thanks to some of my friends who helped me when I was unable to move further in my research topic. And finally to God, so that I am able to stand at this stage in all respects. Thanks to all of you.

REFERENCES

Book:

[1] W.Inmon,(1992) *Building the data Warehouse*, QED Technical Publishing Group**.**

Journals:

[1] G. Satyanarayana Reddy *"Data Warehousing, data mining, OLAP and OLTP technologies are essential elements to support decision-making process in industries"*, (2010), pp. 2865-2873.

[2] Surajit Chaudhuri, Umeshwar Dayal *"An Overview of Data warehousing and OLAP Technology"*, ACM, (1997), pp. 23-33.

[3] Jixue Liu, Millist Vincent "An Architecture for data warehouse systems", In the Proc. of *IEEE Intl. Conf. on Global Connectivity in Energy, Computer, Communication and Control (TENCON'98)*, (1998), pp.107-110.

[4] Yu QianCheng,(2010) "Metadata Integration Architecture in Enterprise Data Warehouse System".

[5] Hugh J. Watson (2005) " Data Warehouse Architectures: Factor in the Selection Decision and the Success of the Architectures", pp.10-22.

[6] Amer Nizar AbuAli , Haifa YouslAbu-Addose (2010)"Data Warehouse Critical Success Factors" Vol.42), pp.326-335.