

TECHNICAL NOTE

Available Online at www.jgrcs.info

DATA WAREHOUSE AND ITS METHODS

Sandeep Singh ^{1*} and Sona Malhotra ²

- ^{*1}. M.Tech Scholar, CSE Department, U.I.E.T. Kurukshetra University, Kurukshetra, India
san_nan_mam@hotmail.com
2. Assistant Professor, CSE Department, U.I.E.T. Kurukshetra University, Kurukshetra, India

Abstract: Data warehousing is a collection of *decision support* technologies, aimed at enabling the knowledge worker to make better and faster decisions. A data warehouse is a subject-oriented, integrated, time varying, non-volatile collection of data that is used primarily in organizational decision making. Data warehouse supports on-line analytical processing, the functional and performance requirements of which are quite different from those of the on-line transaction processing applications traditionally supported by the operational databases. In this paper author suggest that for a strategic solution, the hub and spoke / centralised architecture is the more likely choice.

Keywords: On-line transaction processing, On-line analytical processing, Data marts, global metadata.

INTRODUCTION

Data warehouses is a useful tool, gives benefit from the ability to store and analyze data, and this can allow in making sound business decisions. It is also important to make sure that the correct information is published, and it should be easy to access by the people who are responsible for making decisions. There are two elements that make up the data warehouse environment, and these are presentation and staging. The staging could also be known as the acquisition area. It is composed of ETL operations, and once the data has been prepared, it will be sent to the presentation area. To build an effective data warehouse, it is important to understand the data warehouse design principles. If data warehouse is not built correctly, it run into a number of different problems. The proper methods for building a powerful data warehouse are based on information technology tactics, it is important that for an individual or concern organization to understand the importance of having a data warehouse.[1,2,3]

BENEFITS OF DATA WAREHOUSE METHODS

- With data warehousing, you can provide a common data model for different interest areas regardless of data's source. In this way, it becomes easier to report and analyze information.
- Find out inconsistencies and resolved before loading of information in data warehousing, this makes the reporting and analyzing process simpler.
- The best part of data warehousing is that the information is under the control of users, so that in case the system gets purged over time, information can be easily and safely stored for longer time period.

- Because of being different from operational systems, a data warehouse helps in retrieving data without slowing down the operational system.
- Data warehousing enhances the operational business applications values and customer relationship management systems.
- Data warehousing give a way to proper functioning of support system applications like trend reports, exception reports and the actual performance analyzing reports.
- Precisely, a data warehouse system proves to be helpful in providing collective information to all its users. It is mainly created to support different analysis, queries that need extensive searching on a larger scale.[4,5,6]

TYPES OF DATA WAREHOUSE ARCHITECTURES

There are predominantly five Architectures – independent data marts, bus architecture, hub and spoke, centralized, and federated.

Independent Data Marts-

- Generally developed by individual organizational departments, which operate in isolation. Organizations' with a number of data marts will find data definitions across the data marts inconsistent and lacking in conformity.

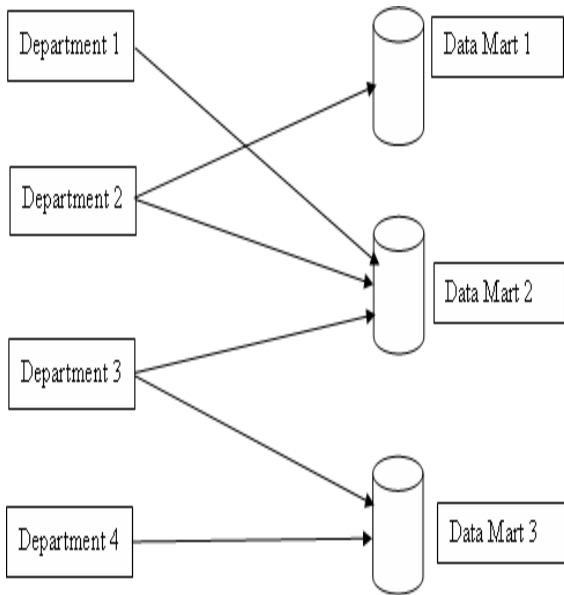


Figure 1: Independent Data Mart Architecture

- Accident, not architecture
- Sourced directly from operational systems
- Redundant data
- Redundant processing
- Not scalable
- “Doesn’t require that a business come to terms with their data and business procedures”

• **Data Mart Bus architecture** - this architecture is rooted in specific business processes but the use of conformed dimensions and facts enables the incremental integration of additional data marts to form an organisation wide view of the organisation. Data is modelled dimensionally in a star schema.[7,14]

- Start at the ground level rather than the enterprise level – “Bottoms up” approach
- Pick business processes and model them
- Dimensional modeling (star schema) rather than ERD
- Data marts uses “standardized, conformed dimensions”
- Warehouse is “conceptual” created by the “bus” of conformed dimensions

• **Hub and Spoke architectures** – the aim of this architecture is to iteratively develop, subject by subject, an enterprise wide view of data where atomic level data is maintained in the warehouse in 3rd normal form i.e. the hub. The vast majority of users will access the data from dependant dimensionally modeled data marts (spokes).[8,13]

- Addresses need for dependent data marts
- Marts receive data from central source-the warehouse.
- Medium and large contexts
- Scalable, often enterprise wide
- Sometimes called the Corporate Information Factory.

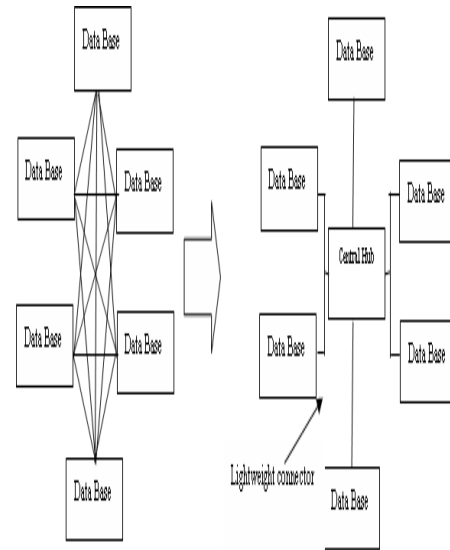


Figure 2: Hub and spoke architectures

• **Centralized Data Warehouse** – this architecture is similar to the hub and spoke architecture but has no dependant data marts.

- No dependent data marts
- Consolidates data marts into data warehouse
- Warehouse contains both atomic (detail) data and summary.[12]

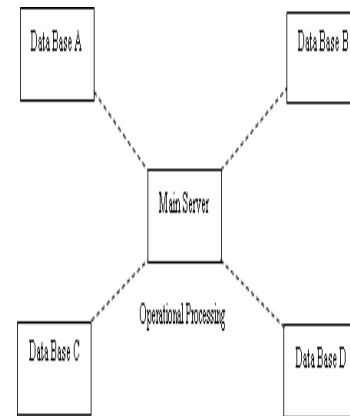


Figure 3: Centralised data warehouse

• **Federated** – the federated architecture draws upon existing decision support structures where the “data is either logically or physically integrated using shared keys, global metadata, distributed queries, and other methods”.[9,10,11]

- Low overhead
- Do not re-architect existing data structures (such as marts, warehouses, or transactional systems)
- Logically or physically integrate data
- Distributed queries and metadata associate the data
- Access data simultaneously across multiple systems
- Useful in mergers and acquisitions.

CONCLUSION

An independent data mart is more likely to be selected if resources are limited. It is observed that in the key factors influencing the choice of a bus architecture is a high need to share data or information between departments. Where as in Hub and Spoke / Centralised architecture tends to be selected where the data warehouse is considered to be an integral part of a strategic solution hence there is a high need for data to be made freely available between business units. While the need for information to be made freely available between business units is important in the choice of either of these architectures, it is the bus architecture that will tend to be chosen. Similarly, if the data warehouse is required quickly the bus architecture has proved popular. However, as a strategic solution, it is the hub and spoke / centralised architecture that is the more likely choice.

REFERNCES

[1]Ralph Kimball, Margy Ross; “The Data Warehouse Toolkit”; Wiley Computer Publishing; 2ndedition, 2002; pp.79-85.

[2]Ralph Kimball, Margy Ross; “The Data Warehouse Toolkit”; Wiley Computer Publishing; 2nd edition,2002; pp.309.

[3]Rajiv Maheshwari;“Need of Data Warehouse”;Article;16 Feb 2010.

[4]Craig Borysowich; “Objectives Of Data Warehouse”; article 6/3/2009;22 Feb 2010.

[5]By Wayne Eckerson, Director of Research, “Advantages and disadvantages of Data Warehouse” TDWIconference 2002.

[6]W. H. Inmon; “Building the Data Warehouse”; Wiley Computer Publishing; Third Edition,2002;pp.300-301.

[7]<http://www.geekinterview.com/kb/Metadata-Warehouse.html>;12 Feb 2010.

[8]<http://www.cssolutionsinc.com/upload/31082006072229.pdf> f18Brian;May“Data WarehousingConsultant” ;Article;http://EzineArticles.com/expert=Brian_May;<http://www.datawarehousingconsultants.com>;19 Feb 2010.

[9]Gray J. et.al. “Data Cube: A Relational Aggregation Operator Generalizing Group-by, Cross-Tab and Sub Totals”

Data Mining and Knowledge Discovery Journal, Vol 1, No 1, 2007.

[10]Database Systems.” CompCon Spring ’91, SanFrancisco, California.Farley, J. D. 2006 May 16. “The NSA’s MathProblem.” New York Times.

[11]Belson, K. 2006 May 16. “BellSouth Denies ItHanded Over Telephone Records to the N.S.A.” TheNew York Times.

[12]Bae, K., Kim, J. and Huh, S. 2003. “FederatedProcess Framework in a Virtual Enterprise Using anObject-Oriented Database and Extensible MarkupLanguage.” Journal of Database Management 14(1):27-47.

[13]Chen, H., Zeng, D., Atabakhsh, H., Wyzga, W. andSchroeder, J. 2003. “COPLINK: Managing LawEnforcement Data and Knowledge.” Communications of the ACM 46: 1.

[14]Fang, D., Hammer, J., McLeod, D. and Si, A. 2001.“Remote-Exchange: An Approach to ControlledSharing Among Autonomous, Hetrogenous