Comparative Study and Analysis of Federated and Non-federated Data Warehouse Architecture

Bikramjit Pal[1], Sunanda Bharti[2], Mallika De[3]
[1]Research Scholar, Department of Engineering and Technological Studies, University of Kalyani, Kalyani, Nadia-741235, West Bengal
[2]Student, Department of Computer Application, JIS College of Engineering, Kalyani, Nadia-741235, West Bengal
[3]Ex-Professor, Department of Engineering and Technological Studies, University of Kalyani, Kalyani, Nadia-741235, West Bengal

Abstract: Data warehouse is a repository of an organization's electronically stored data. Data warehouses are designed to facilitate reporting and analysis. A data warehouse houses standardized, consistent, clean and integrated form of data sourced from various operational systems in use in the organization, structured in a way to specially address the reporting and analytic requirements. In this paper we performed a comparative analysis of federated vs other architectures to find their usefulness in business. We investigated thoroughly the advantages and disadvantages to clearly distinguish the various organizationally implemented data warehouse architectures.

Keywords: *Data warehouse, reporting, analytic, federated, operational, consistent, standardized, integrated.*

*****

1 Introduction: Data warehousing has come a long way since its inception. In the current scenario, data warehouses are growing into rich systems capable of delivering key performance metrics to top management, strong analytical capabilities to middle management and corrective data back to operational systems based on information derived from the analytical system. The DW market is currently driven by business-driven solutions focused on domain specific challenges.

The current business environment of global village, consolidation in numerous industries and cutthroat competition originate newer and stiffer challenges for the data warehouse architects. Although, there are various methodologies available in the market to handle these challenges, most of the data warehouses fail to deliver the expected results. There are solutions available in the market which are infallible in theory and appear to address most of the challenges but are ridden by the implementation nightmares.

2. Federated Data Warehouse: A federated data warehouse is the integration of heterogeneous business intelligence systems set to provide analytical capabilities across the different function of an organization. It's a realistic method to achieve the "single version of the truth" across the organization considering the political and implementation challenges. It aims to integrate the key business metrics, measures and dimensions. But it doesn't aspire to create a single platform to carry out all the functional analysis.

The foundation of federated DW is the common business model of the organization. Common business model is a continuously evolving semantic understanding of the business in consent with all business units. This common business model is the initiation point for an iterative process of building the different business intelligent subsystem based on a common staging area.

2.1 The Need of Federated Data Warehouse

- Mergers & acquisitions – The way business has adopted the inorganic growth path has put many architects in unenviable positions. Merger & acquisitions have become routine events in today's business scenario. Each acquisition brings in a big information bank along with it but the biggest challenge is to integrate this information with the existing BI system. It doesn't make sense to abandon a fully functioning data warehouse infrastructure, so the DW teams are forced to adopt a federated BI architecture. It's certainly not an easy task to completely imbibe the new system with the existing BI architecture. The best way to handle it is to federate and integrate the two systems.
- Cross-functional requirements – Cross function analysis has become an everyday need for most of the enterprises; the need of cross-functional analysis must be emphasized while finalizing the architecture of the data warehouse. A federated DW stores the information of most of the functions, if not all, that an organization deals with. This information could be stored in one or more than one BI systems. FDW support the cross functional analysis using the common dimension across the different systems. Common dimension is an outcome of the common business model which defines the collective nature of business functionality. A federated DW is an active

 ISSN: 2321-8169
94 – 99

cooperation of multiple business intelligence systems where each system can talk to other BI system and full fill cross-functional requirements.

- A speedy cost effective solution – The build time involved in federated data warehouse is enormously less when compared with enterprise wide data warehouse. One of the main reasons is that the federated DW tries to integrate the existing system by providing a common framework. It does not aspire to build a uniform foundation which is a tedious and lengthy process. The incremental nature of the federated DW reduces the long waiting period associated with most of the big DW, eventually reducing the cost as well.
- Ease of implementation – When building an enterprise wide data warehouses, architects have experienced a lot of politics and vested interests trying to mould the crucial decisions. In the real world scenario these factors cannot be sidelined and at times, influence the implementation in a larger manner. Federated DW approaches the problem in a more pragmatic manner and the idea to use the old BI systems by integrating them with the newer system prevents the major point of conflict.

2.2 Federated Data Warehouse Architecture

We envisage two types of federation possible in a federated data warehouse:

- Regional federation (or more appropriately regional versus corporate/global federation),
- Functional federation

2.2.1 Regional federation

This can be explained with the help of an example. A big organization has regional data warehouses for regional analytical requirements and a global data warehouse for the corporate requirements. The difference between the two systems (regional & global) is based on the nature of the data that is contained in each system. The global data warehouse stores data which is mostly summarized for the purpose of corporate analytics and reference data.

Reference data would contain confirmed dimensions and corporate level data like currency conversions etc. Regional warehouses store data based on the regional analytical requirements which generally has more detailed information. Reference data provide the integration platform for regional and global warehouse. The data flow from regional data warehouses to the global data warehouse is defined as upward federation and the data flow from global data warehouse to regional data warehouses as downward federation. For data consistency and integrity, reference data should be made common across the various data warehouses. Uniform & consistent definition of reference data across the participating data warehouses ensures the 'single version of truth' across the federated architecture.

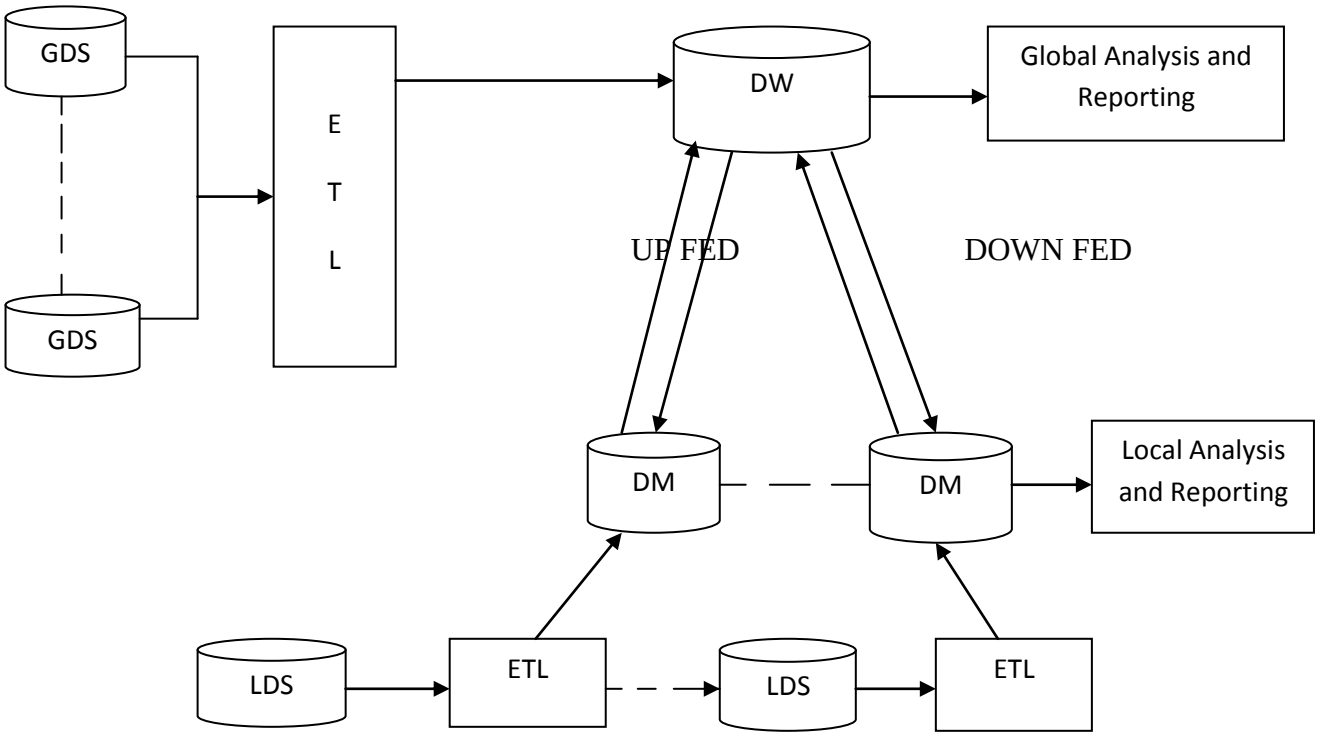

Fig – 1. Federated Data Warehouse Architecture

Data movement during upward federation:

• Upward federation would include the movement of fact data from regional data warehouses to the global data warehouse. If required, this data can be aggregated during movement.

Data movement during downward federation

• Reference data will flow from global to the regional level data warehouses. This flow would be strictly downward to ensure consistency and integrity of reference data.

• Transactional data that is available in corporate transactional systems (such as corporate ERP) will be sourced at the global level, cleansed and transformed and then moved to the respective regions.

• Summary data – The global summarization will happen in the global data warehouse and will be moved to the regional data warehouses. This can be useful in the analyzing how a particular Region is performing against the rest of the company.

2.2.2 Functional federation: A functional federated data warehouse will be the candidate when an organization has built data warehouses which are either subject specific, packaged solutions or built for a specific enterprise application. The federated data warehouse architecture is the "big umbrella" that provides the foundation and environment to facilitate and enable business analysis and decision support in this heterogeneous environment.

A functional federated data warehouse has room for all the components of a contemporary BI application of a large and complex business entity. Typically it should contain the following components:

• Packaged data warehouses (DWs) and data marts (DMs)
• Custom built data warehouses and data marts
• Real time data store & real time data reporting
• Custom built analytical applications
• Online analytical processing (OLAP) tools
• Extraction, transformation and load (ETL) tools
• Cross functional reporting systems

2.3 Approaches for Federation

The following steps should be taken when architecting a federated data warehouse:

• Define the federated data warehouse goals and business requirements
• Define team structure & assign roles and responsibilities
• Document existing data warehouses & BI systems. The information to be documented is - Business areas addressed by the existing systems, target users, reporting and analytical capabilities provided, the data sourcing strategy and cleansing/transformations methodology while moving the data into these systems
• An analysis needs to be done to identify whether it is a case for federated architecture.
• Define the common business model which should provide the basis for common dimensions.
• Define an integration strategy for federation. Identify inter-dependency between existing and new to-be developed data warehouses.
• Perform the integration of existing data warehouses into the new architecture in small phases, giving first priority to the areas which are critical for business.

Federated Data Warehouses are best in very large organization where development is separated by geography, organizational boundaries, or where multiple data warehouses exists due to mergers & acquisitions.

To make FDWs successful, there needs to be a rapid convergence to standardized technologies. This include:

- Same type of databases and support pack levels (costs and compatibility)
- Same technical platforms Hardware, Backups and Archiving (costs)
- Shared Portal and user interface strategy (reduced training and support)
- Shared security design and centralized administration (risk management)

If the data is federated you gain faster response time to business needs, can execute multiple projects in parallel, and work 24/7 across the globe. But without any standardization, it can also be very costly.

3. Non-Federated Data Warehouse:

3.1. Centralized Data Warehouse: A Centralized Data Warehouse is a single physical repository that contains integrated data extracted from multiple operational systems and merged with data from external information systems. Currently, most organizations build central data warehouses, even if their organizations are geographically distributed.

The Need of Centralized Data Warehouse

Centralized Data Warehouses are great for small and mid-size data warehouses (less than 15-40Tb). There are great benefits in terms of the ease to mange upgrades, support packs, enforcing development standards, transport control, master data management and the overall total cost of ownership.

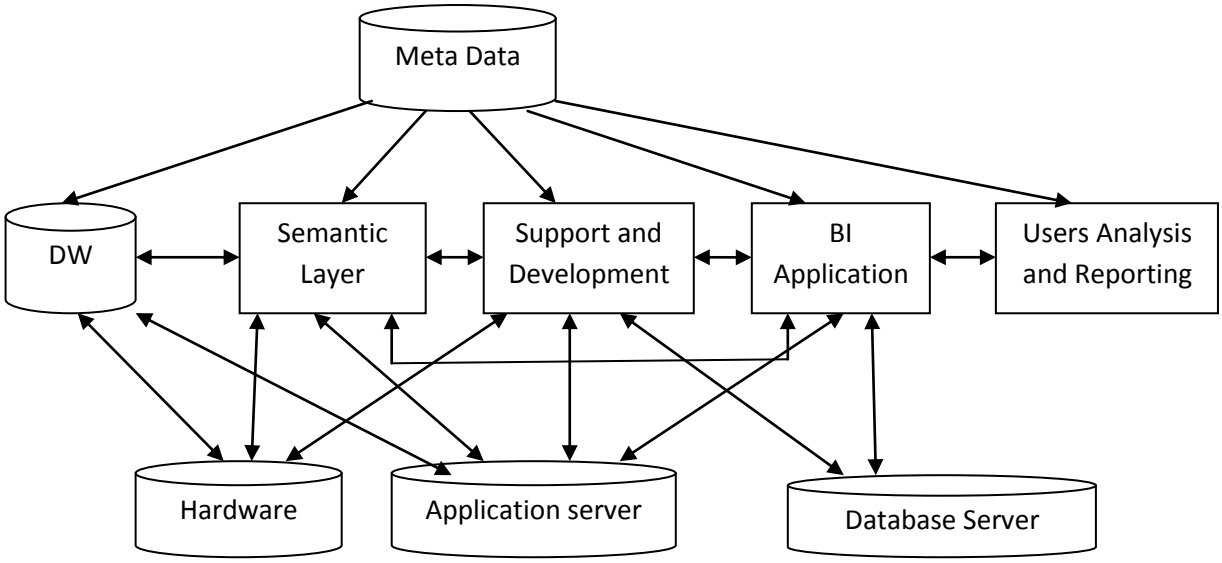

Fig – 2. Centralized Data Warehouse

To make CDWs successful, there needs to be:

- Adequate funding of hardware, application servers, database servers
- Serious consideration should be made to move BI and reporting to BWA
- Focus on using the database capacity on storage and data loads– not queries
- No direct reporting from DSOs (takes too much system resources)
- Broadcasting , caching and performance tuning is a dedicated support effort
- A plan for data partitioning and archiving needs to be in-place as soon as the system exceeds 5-8 TB. If the data is centralized it is faster to develop new solutions for the business and merging from different data sources are easier.

The Advantages of Centralized Data Warehouse

The major advantages of a centralized data warehouse are :

- *Security*: The centralized data warehouse offers a high degree of security and control over data access.
- *Ease of Management*: A centralized data warehouse is easier to manage because operations to analyze and answer complex queries do not involve the additional complexity of a network.
- *Experience*: A centralized data warehouse can draw upon the technical experience available for centralized database systems. Unfortunately, there is very little experience in managing this type of system on a distributed platform.

The Disadvantages of Centralized Data Warehouse

The disadvantages of centralized data warehouse are :

- *Performance*: Many users must compete to access the data, resulting in delays caused by querying requests. In addition, if the users are geographically distributed delays also may occur due to transmitting access requests and responses to and from the central data warehouse location.
- *Expandability*: Expansion is expensive in a centralized data warehouse. For example, if the volume of the data exceeds the capacity of the warehouse, or the amount of processing against the data has increased, then the organization must replace the existing centralized data warehouse with a larger one. This is costly and may not be possible within an organization's fiscal constraints, which may cause the data warehouse to become out dated prematurely.
- *Reliability*: The centralized data warehouse is the single point of failure, when the whole system goes down. It usually takes some time to bring it back.
- *Cost*: Experience over the past decade has shown that it is cheaper to have a number of smaller computers linked together rather than a large central machine.
- *Vendor Dependency*: Vendor dependency is an unfortunate consequence of any centralized approach. Typically, purchasing single centralized systems binds future expansion to the initial choice. Obviously, an inability to use multiple vendors, at least for negotiation can be expensive and migrating to a third party vendor can be cost-prohibitive.

3. 2. De-centralized Data Warehouse: In Decentralized Data Warehouse, a central gateway provides access to

remote data with the help of a logical view. This central gateway processes real time user queries. Users can access and also query the remote data via central gateway.

A data warehouse is very large repository of a company's historical and current transactional data. In order for the data warehouse to efficiently handle the high volume data while trying to service potentially high number of data consumer, certain mechanisms must be considered in the design and implementation of a data warehouse in order to smoothly operate the whole system.

The Need of Decentralized Data Warehouse

A **Decentralized Data Warehouses** makes sense if there are logical division between business units, geographies and little shared reporting. i.e., in a conglomerate organization with diverse business units. The benefits of DDWs include the flexibility of the FDW with the technology standardization and lower cost of ownership of the CDW. To make DDWs successful, there needs to be:

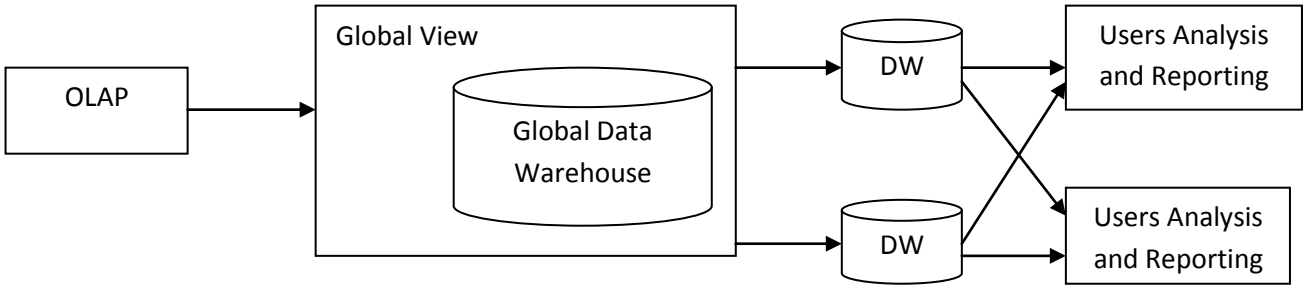

Fig – 3. De-centralized Data Warehouse

- A formal Master Data Management (MDM) strategy with clearly defined standards
- A rule based data cleaning and data integration plan for centralized reporting
- A shared hardware location to keep costs lower
- Tight integration with upgrades, support packs and interface standards

With DDWs there is a risk of creating stove-pipe data marts that cannot be integrated at the corporate level without very high costs.

The Advantages of Decentralized Data Warehouse

The major advantages of a decentralized data warehouse are:

- Controlling and storing function can be accomplished easily.
- Delay in material handling will be eliminated.
- Minimizes the chances of loss by fire.
- No need of internal transportation costs.
- Specific needs of individual departments can be easily fulfilled.
- Saving in material handling cost.

The Disadvantages of Decentralized Data Warehouse

The disadvantages of decentralized data warehouse are:

- Higher cost of supervision.
- More space is required for individual departments.
- Higher amount of investment is required.
- More time for stock taking and taking.
- Higher cost of staff and stationary.
- Improved technique is less possible for controlling of materials.

4. Recommendations FDW, CDW and DDW Architectures

In general, the benefits and risks can be summarized as:

Parameters	FDW (Federated Data warehouse)	CDW (Centralized Data warehouse)	DDW (Decentralized Data warehouse)
Organization	Best for very large organization where development is separated by geography, organizational boundaries, or where multiple data warehouses exists due to mergers & acquisitions.	Best for small and mid-size data warehouses in organizations.	If there are logical division between business units, geographies and little shared reporting i.e. in a conglomerate organization with diverse business units.
Max Size	Virtually unlimited.	40+ Tb	Virtually unlimited.
Success Factor	Use same type of database, ETL, tools and support levels (costs & compatibility).	Adequate funding of hardware, application servers, database servers.	A formal Master Data Management (MDM) strategy with clearly defined standards.
	Use the same O/S, Hardware, Backups and Archiving Systems (costs).	Implement BWA.	A rule based data cleaning and data integration plan for centralized reporting.
	Shared Portal and User interface strategy (reduced training and support).	Use the database capacity data loads not queries.	Use a shared hardware location to keep support costs lower.
	Shared security design and centralized administration (information risk management).	Direct reporting from DSOs should not be allowed.	Tight integration with upgrades, support packs and interface standards.
Issues	Without any standardization, it can be very costly.	Performance can be Poor. An archiving plan is essential when the system exceeds 5-8 Tb.	There is a risk of creating stovepipe data marts that cannot be integrated at the corporate level without very high costs.

5. Conclusion and Result: At the end of the evolution, we have come to the conclusion that the data warehouse and the data mart have a co-existing relationship by adhering to the user analysis and reporting methodology, designed and conceptualized from user perspective and which has very prominent practical application in the real world. Also, Federated Architecture proves to be very affordable and cost effective for implementation in a large scale business organization.

6. References:

[1] Data Warehouse and OLAP: A research-oriented bibliography by a.Mendelzon, C.Hurtado and D.Lemire.
[2] http://db.stanford.edu/pub/papers/warehouse-reserach.ps
[3] Pal B., "Comparison of Data Warehouse Architecture Based on Data Model", IJITKM, July-December 2010, 2, Number 2,pp. 303-304
[4] Chowdhury R., Pal B., "Proposed Hybrid Data Warehouse Architecture Based on Data Model", International Journal of Computer Science and Communication, Vol. 1, No. 2, 2010, pp. 211-213
[5] Federated Data Warehouse Architecture-ZenTut *www.zentut.com/data-arehouse/federated-data-warehouse-architecture/*
[6] Federated Vs. Centeralized Vs. De-centeralized Data warehouse s*apinsider.wispubs.com/.../Federated-Vs-dot-Centeralized-Vs-dot-De-centeralized-Data Warehouse*
[7] *https://en.wikipedia.org/wiki/Federated_database_system*
[8] *lyle.smu.edu/~mhd/7330f09/reid.pptx*
[9] Federated data warehouse approach to support the national - WIT.pdf
[10] Four Ways to Build a Data Warehouse www.bi-bestpractices.com/view-articles/4770
[11] Federated Vs. Centeralized Vs. De-centeralized Data warehouse
[12] Centralized Data Warehouse | Online Learning www.learn.geekinterview.com/data-warehouse/data.../centralized-data-warehouse.html
[13] Selecting a Centralized vs. De-centralized Data Warehousing Solution *datawarehouse.ittoolbox.com/.../selecting-a-centralized-vs-decentralized-data-warehouse*
[14] Centralized-data-warehouse-system : - ITC Infotech *www.itcinfotech.com/centralized-data-warehouse-system*
[15] Benefits of a Data Warehouse - BI-Insider *bi-insider.com/portfolio/benefits-of-a-data-warehouse/*
[16] 16. Federated-, Decentralized - and Centralized DataWarehouse *www.svenvanleuken.com/federated-decentralized-and-centralized-datawarehouse/*
[17] Is Data Decentralization the New Trend? | Sandhill *sandhill.com/article/is-data-decentralization-the-new-trend/*
[18] Centralization vs decentralization - Toolbox for IT ... - Data Warehouse *datawarehouse.ittoolbox.com/groups/.../centralization-vs-decentralization-583259*
[19] Data Warehouse Architectures for Multinational Organizations: Part 2 *www.ibmbigdatahub.com/.../data-warehouse-architectures-multinational-organization*
[20] Effective Database Management : Centralized vs. decentralized data *effectivedatabase.com/centralized-vs-decentralized-data-entry/*
[21] Decentralized Warehouse | Online Learning *www.learn.geekinterview.com/data-warehouse/dw.../decentralized-warehouse.html*
[22] FederatedDataWarehouse2.ppt - Capita *capita.wustl.edu/capita/capitareports/.../FederatedDataWarehouse2.ppt*
[23] Corporate Data Architecture in a Federated World - IRMAC *www.irmac.ca/0102/InergiIRMAC2002final.ppt*
[24] Federated Data Warehouse | prateek gaur - Academia.edu *www.academia.edu/5240170/Federated_Data_Warehouse*
[25] Data warehouse architecture - SlideShare *https://www.slideshare.net/pcherukumalla/data-warehouse-architecture-7636477*
[26] Data Warehousing Architecture - TutorialsPoint *https://www.tutorialspoint.com/dwh/dwh_architecture.html*
[27] Overview of a Data Warehousing Architecture - Data-Warehouses.net *data-warehouses.net/architecture/*