Federated Data Warehouse Architecture

Authors: Rajan Jindal, Abhishek Acharya

Deciding the most suitable architecture is the most crucial activity in the Data warehouse life cycle. Architecture is the key factor in setting up the abilities and the limitations of a data warehouse. This paper intends to discuss the Federated approach for DW architecture, it also tries to evaluate the need and cases when implementing a federated data warehouse becomes the most practical approach for architecting a data warehouse.
Table of Contents

INTRODUCTION ................................................................................................... 3
FEDERATED DATA WAREHOUSE ........................................................................ 3
THE NEED OF FEDERATED DATA WAREHOUSE ............................................. 3
ARCHITECTURE OF FEDERATED DATA WAREHOUSE .................................... 4
APPROACH FOR FEDERATION ........................................................................... 6
KEY SUCCESS FACTORS ................................................................................... 7
CONCLUSION ...................................................................................................... 7
ABOUT THE AUTHORS ....................................................................................... 7
ABOUT WIPRO TECHNOLOGIES ........................................................................ 8
WIPRO IN BUSINESS INTELLIGENCE AND DATA WAREHOUSING .............. 8
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.

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 concert 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.

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 cooperation of multiple business intelligence systems where each system can talk to other BI system and fulfill 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.

Architecture of Federated Data Warehouse

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

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

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.

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.
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
Approach 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
Key Success Factors

In order to develop a successful federated data warehouse, certain criteria have to be fulfilled:

- The various component data warehouses should share conformed dimensions. The conformed dimensions represent the dimensions having identical business meaning, identical structure and identical data. However, to keep them physically separated or not, can be decided based on implementation complexities.
- There should well defined, documented and integrated business rules which would be used across the component data warehouses in the whole architecture.
- The component data warehouses should have confirmed facts or measures. Conformed facts have identical business meaning and exactly same data values for same set of dimensions. Identical data value for a confirmed fact will also ensure “single version of truth” which is essentially one of the prime design objectives of a federated data warehouse.
- A single ETL tool should be used in the whole federated data warehouse. Usage of a single ETL tool would also ensure common metadata.

Conclusion

A Federated Data Warehouse offers a very practical solution for implementing and delivering value added solution to a business. The iterative nature of a Federated Date Warehouse reduces the extended delays normally associated with data warehousing solutions. In conclusion, a Federated Data Warehouse, with its lower costs and higher reliability, provides an excellent value proposition to organizations.

About the Authors

Rajan Jindal is a Senior Consultant with Wipro Technologies BI & DW Practice and has six years of experience with multiple end-to-end data warehouse implementations for various clients across the globe. He is also a Wipro certified Data Warehouse Data Architect. His areas of interest include Data Modeling, Architecture, and Data Mining.

Abhishek Acharya is a Senior Consultant with Wipro Technologies BI & DW Practice and has worked extensively in data warehousing and implemented many big DW projects. He is a Wipro certified DW Data Architect.
About Wipro Technologies

Wipro is the first PCMM Level 5 and SEI CMMI Level 5 certified IT Services Company globally. Wipro provides comprehensive IT solutions and services (including systems integration, IS outsourcing, package implementation, software application development and maintenance) and Research & Development services (hardware and software design, development and implementation) to corporations globally.

Wipro’s unique value proposition is further delivered through our pioneering Offshore Outsourcing Model and stringent Quality Processes of SEI and Six Sigma.

Wipro in Business Intelligence and Data Warehousing

Wipro provides end-to-end Data Warehousing and Business Intelligence services to Global Corporate Enterprises. Wipro has implemented Business Intelligence and Data Warehousing solutions for over 90 customers worldwide including 30 Fortune 1000 customers across various verticals like Finance, Insurance, Utilities, Telecom, Retail, Logistics, Manufacturing and Healthcare.

Wipro has evolved its “Insta Intelligence” project management and delivery methodology built around leading edge technologies in the areas of Data Acquisition, Data Modeling, Data Management, OLAP, Data Mining, and Meta-data Management to deliver innovative, surefire solutions to its customers.

For further information visit us at: http://www.wipro.com/datawarehouse

For more whitepapers logon to: http://www.wipro.com/insights

©Copyright 2004. Wipro Technologies. All rights reserved. No part of this document may be reproduced, stored in a retrieval system, transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without express written permission from Wipro Technologies. Specifications subject to change without notice. All other trademarks mentioned herein are the property of their respective owners. Specifications subject to change without notice.