Selecting the Right Architectures for Successful Data Warehouses

Deciding a suitable architecture is very important activity in the Data warehouse life cycle. Architecture is critical in setting up the abilities and the limitations of a data warehouse. Finding the way through the confounding array of architectural choices and the various approaches can be a daunting task. The paper reviews and summarizes the basic types of architectures, their pros and cons, and discusses the various approaches to build a data warehouse system, concentrating on the two main options, top down and bottom up.
About the Author

Srinivasan Kanchi
Srinivasan Kanchi has a Masters degree in Business Management with specialization in IT Management. He has an experience of twenty years in IT industry and has worked with various clients, in different roles and through variety of consulting assignments. Currently, he is the lead consultant with IT Architecture Consulting (ITAC) practice and is based in the UK office.
Table of Contents

1. Introduction          3
2. Classic Enterprise Data Warehouse          3
3. Federated Data Warehouse - Data Mart Systems    4
4. Non-Architected Data Marts (LegaMarts)      5
5. Incremental Architected Data Marts      6
6. Approaches to Achieve the Goal of the Enterprise Data Warehouse 6
7. Choosing the Right Path          10
8. Summary          12
9. Acknowledgements          12
10. References          12
Introduction
Data warehousing has evolved from a simple theoretical vision to a very complex world full of the variations and refinements that mark the real world. Whereas the classic Enterprise Data Warehouse (EDW) vision consists of a world of simple black and white, the real world of data warehouse and data mart system implementation is made up primarily of potentially confusing shades of gray. Where once there had only been the EDW option, now there is the incremental Architected Data Mart (ADM), Enterprise Data Mart Architecture (EDMA), Data Stage/Data Mart (DSDM), Distributed Data Warehouse/Distributed Data Mart (DDW/DDM), and lately the Federated Data Warehouse/Data Mart (FDW/FDM). In the subsequent sections, the basic types of various architectures that exist today are outlined along with the advantages and disadvantages of each type and their suitability (or adaptability) in an enterprise.

Classic Enterprise Data Warehouse
The classic Enterprise Data Warehouse (EDW) is a common and unique repository for enterprise information. It is a read only environment made up of detailed and aggregated data that is fully cleansed and integrated; and includes extensive detailed history of transaction level data. Architecturally, the classic enterprise data warehouse delivered that priceless and elusive goal, a single version of the truth, primarily through subset, or dependent, data marts.

The advantages of the classic EDW architecture include:
- Single version of the truth
- One set of extraction processes and business rules
- Common semantics
- Centralized, controlled environment
- Easily created and populated subset data marts
- Single Meta data repository

The disadvantages of the classic EDW architecture include:
- Expensive to implement
- Very resource intensive
- Inherent enterprise scale requires enterprise-scale systems and resources
- High risk scenario

![Fig 1 Classic Enterprise Data Warehouse System](image-url)
Federated Data Warehouse - Data Mart Systems

In many organizations, multiple teams undertake data warehouse projects, resulting in multiple data warehouse systems across the enterprise. Although in the strictest sense, there is only one Enterprise Data Warehouse (EDW) with all other entities being subset or incremental data marts, not many organizations are too strict with semantics. Thus we have many enterprises around the world with two, six, or a dozen or more “data warehouse” systems.

This proliferation of data warehouses has led to the next generation of the EDW architecture, that of a federated system of data warehouses or data marts. The Federated Data Warehouse - Federated Data Mart (FDW-FDM) system is marked by the characteristics of sharing common data points between multiple data warehouse or data mart systems, thus eliminating redundancy, and ensuring a consistent and unique version of the “truth” throughout the organization.

The advantages of the Federated Data Warehouse-Data Mart system are:

- Common semantics and business rules
- Single set of extraction processes
- Decentralized resources and control
- Parallel development

The disadvantages of the Federated Data Warehouse-Data Mart system are:

- Challenging to coordinate efforts
- Difficult to overcome political and data ownership issues
- Requires consensus on architecture, business rules and semantics among disparate teams
- Complex and challenging technical environment
- Often have multiple Metadata repositories
Non-Architected Data Marts (LegaMarts)

A side effect of the degree of difficulty associated with achieving the goal of an EDW or FDW system has been the rise of the non-architected, non-integrated data mart. These quick and dirty data sets are not architected in any way, and consequently are not capable of being integrated with any other data warehouse system resource.

They quickly become legacy systems, stand-alone stovepipes of information that simply add to the problem, not to the solution. Multiple extraction processes, multiple business rules, fragmented data, lack of architecture and multiple versions of the truth mark these legacy data marts, or LegaMarts.

The advantages of LegaMarts are:
- Speed
- Low cost

The disadvantages of LegaMarts are:
- Multiple versions of the truth
- Multiple extraction processes
- Multiple business rules
- Multiple semantics
- Fragmented Data
- Extremely challenging to integrate
**Incremental Architected Data Marts**

In an effort to overcome the very real challenges associated with an EDW or FDW and at the same time avoid the nuisance of LegaMarts, an alternative architecture has evolved, that of incremental Architected Data Marts (ADM). It is typical that a modern, enterprise class tool is used to facilitate an “extract once, populate many” strategy of populating the incremental ADM.

The advantages of incremental Architected Data Marts are:
- Common semantics and business rules
- Single set of extraction processes
- Accomplishable scope
- Inherently incremental

The disadvantages of incremental Architected Data Marts are:
- Most effective when using an enterprise class tool
- Requires Enterprise Data Mart Architecture (EDMA)
- Requires compliance with EDMA by all teams

**Approaches to Achieve the Goal of the Enterprise Data Warehouse**

Generally, there are two basic approaches to meet the goal of the Enterprise Data Warehouse, top down or bottom up. In a top down scenario, the EDW is architected, designed and constructed in an iterative manner. In a bottom up scenario, a series of incremental Architected Data Marts are created, which forms the basis of the resulting EDW system.
“Top Down” Development
An EDW is composed of multiple subject areas, such as finance, human resources, marketing, sales, manufacturing or core business transactions, and so on. In a top down scenario, the entire EDW is architected, and then a small slice of a subject area is chosen for construction. Subsequent slices are constructed, until the entire EDW is complete. Complete EDW systems take 3-4 years for construction on average and typically cost $5-8 million (US$) for a medium sized organization (based on multiple industry surveys). Large global organizations should expect to spend $10-50 million to build the initial EDW system, which is most likely to be FDW architecture.

Fig 5 EDW Phased Subject Area Development

Fig 6 Top Down Phase One EDW Development
While there has been no technical reason to run away from an EDW approach, the cultural, or “soft” issues have proven very challenging for the average IT team to overcome. Chief among these is the “cross everything” nature of an EDW system. By its very name, “Enterprise” Data Warehouse, it implies and demands that the EDW team crosses every functional, process, ownership, and geographic and so on, boundary in the entire organization.

The successful negotiation of this critical area requires a tremendous level of political acumen, one that is not always present in the EDW team. Add to this mix, the requirement for the team to be extremely flexible, 100% user-oriented, capable of living in constant change, and able to constantly, endlessly re-sell and re-market the EDW system, end up with a set of challenges that very few teams are culturally outfitted to overcome.

Also, EDW systems require an absolute commitment of sustainable support at the highest levels of the organization. These systems take a long time to develop, and may not demonstrate any tangible return on Investment (ROI) for many years. In order to fend off competing initiatives, maintain resources and funding, and maintain sponsorship and commitment of the organization, the project must enjoy sustained support at the CEO level. The sustaining of support, in the face of ever changing executives, ever shifting sands of corporate priorities and constantly growing demands for resources often prove to be impossible for an EDW team.

The advantages to a “top down” approach are:
- Coordinated environment
- Single point of control and development

The disadvantages to a “top down” approach are:
- “Cross everything” nature of enterprise project
- Analysis paralysis
- Scope control
- Time to market
- Risk and exposure

**Fig 7** Top Down Phase Two EDW Development
“Bottom Up” Development

In an effort to realize the tremendous upside potential of data warehousing, but avoid the very difficult challenges inherent with the “top down” model, data warehousing teams have developed the “bottom up” approach to reach the goal of the EDW system. In the “bottom up” approach, an Enterprise Data Mart Architecture (EDMA) is developed to provide a context for development efforts. While it takes the entire system scope at a high level, it is not as detailed as an EDW system architecture, so avoids the “analysis paralysis” so common to those efforts.

Once the EDMA is complete, an initial area of business pain is selected for the first incremental Architectured Data Mart (ADM). The EDMA is expanded in this area to include the detail required for the design and development of the incremental ADM. Subsequent phases fill in the EDMA, until the team and the organization is ready to construct the EDW.

**Fig 8 Iterative “Bottom Up” EDW Development**

The “bottom up” approach allows teams to develop the skills and techniques required for data warehousing in a much lower risk and lower exposure environment than a full scale EDW project. Incremental ADMs are much quicker to develop than EDW systems as well. It is common to develop the first incremental ADM in six to nine months, while it can often take 12 to 18 months to deliver the first phase of an EDW system.

This speed to market is especially important when trying to demonstrate the ROI and intrinsic value of data warehousing to an organization. From the project management standpoint, incremental ADMs greatly assist in keeping the team “on message.” Incremental ADMs are built to solve specific problems for the business; a “focused problem” leads to a “focused team.”

While cost is rarely the deciding factor, incremental ADMs are less expensive than EDW systems. A scaleable, industrial strength ADM can be constructed for $1 M to $1.5 M (US$), this compares favorably with EDW system costs of several million dollars (US$). *(Source-Gartner Reports)*

The advantages to a “bottom up” approach are:

- Quick ROI
- Low risk, low exposure learning and development environment
- Lower level, shorter-term determination required
- Fast delivery
- Focused problem, focused team
- Inherently incremental

The disadvantages to a “bottom up” approach are:

- Overwhelming success overwhelms resources
- Multiple team coordination
- Must have an EDMA to integrate incremental data marts
Choosing the Right Path

Choosing the correct and appropriate path for an enterprise, “top down” or “bottom up”, is always a daunting task for management. Before making a quick, imprudent decision based on which methodology or which approach seems easier and better, it is critical to understand the characteristics of the unique organization or site targeted for implementation of EDW. There are no panaceas in data warehousing, nor one solution or architecture that is correct for everyone. A good practice is to start by determining the type of organization that you are in:

- Think Globally, Act Globally;
- Think Globally, Act Locally;
- Think Locally, Act Locally

Early adopters of data warehousing were in the “Think Globally, Act Globally” group. These organizations are willing to make huge investments in technology and take large risks in the hopes of realizing significant competitive advantages or productivity gains. This group of early EDW adopters skewed the perceptions of analysts and some tool companies, who thought that the entire data warehousing market would think and act like these customers.

As the market has developed, however, it has been revealed that the vast majority of the market is in the “Think Globally, Act Locally” group. Other major players like Microsoft’s entry into the data warehousing market, with extremely low price points and packaged solutions promises to greatly expand this segment of the market.

While there will always be a “Think Globally, Act Globally” component crowded with vendors fighting over those high end budgets, the middle tier will become the dominant market segment in data warehousing.

![Fig 9 Iterative “Bottom Up” EDW Development](image)
Once the overall type of organization is determined, it is time to look at some more specific evaluation criteria such as time, cost, risk, impact, resources, ROI parameters and the scale and gravity of the business pain to solve.

Rate an organization based on each of these criteria, and develop a weighted score based on the priorities and business needs. In developing this score, it is critical that the “soft” issues are recognized as more critical for a long term success or failure than any “hard” technical issue associated with a data warehouse or data mart system.

![Fig 10 Development Strategy Evaluation Factors](image)

For instance, if the evaluation reveals a long term, very high level, CEO level pain, a good level of resources and an organization that gives plenty of time to demonstrate ROI, then a “top down” enterprise data warehouse project is a viable approach. On the other hand, if the pain is at a relatively low level, small in scale, and needs to demonstrate measurable ROI very quickly, then an incremental Architected Data Mart approach is more suitable to that site’s characteristics. It is essential that an approach should not be force fed based on preferences. Both approaches are viable and are being used at thousands of organizations around the world, but the chances of success will be very slim if an approach that does not match the site’s characteristics is used.

A round peg cannot be force fit into a square hole, just because the look and feel of a round peg is preferred.
Summary
Data warehousing has evolved rapidly and continues to be a very fast moving, fast changing market segment. While once there were no tools and there was only one way to design and architect a data warehouse system, now there is a variety of tools and technologies to choose from and a variety of viable data warehouse system architectures. Where once there was only a “top down” approach to the goal of the data warehouse, now there is a viable architected “bottom up” approach that greatly enhances the chances of success for those sites that are not well suited for the “top down” approach.

Given an appropriate architecture and a suitable approach to the goal of the enterprise data warehouse, any team can deliver a high impact, high value, high ROI, sustainable data warehouse system that will entirely change the range of potential outcomes for the enterprise.

Acknowledgements
The author would like to thank Mr. A S Lakhmi, Country Manager, TCS UK and Mr. Anantha Sekar, ATC Practice, TCS UK for their valuable support and guidance.

References
1. Félix Saltor, Marta Oliva, Alberto Abelló and José Samos Building Secure Data Warehouse Schemas from Federated Information Systems
3. Info on Data warehouse and Technologies: The Govtech.net Library by Bitpipe, Inc
About TCS’ IT Architecture Consulting Practice
The IT Architecture Consulting Practice assists customers to define, deploy and govern their enterprise and domain architectures. Spread across the globe, this practice has a strong pool of architects involved in delivery of large solutions across various customers. It has developed strong alliances with architecture tools vendors, training institutions and service providers.

About Tata Consultancy Services
Tata Consultancy Services (TCS) is among the leading global information technology consulting, services and business process outsourcing organizations. Pioneer of the flexible global delivery model for IT services that enables organizations to operate more efficiently and produce more value, TCS focuses on delivering technology led business solutions to its international customers across varied industries.

For more information contact
Ameya Vanjari
Tata Consultancy Services
Gateway Park, Rd. No 13
Andheri (E), Mumbai
India  4000093

Phone:  91 22 6750 5929
Fax: 91 22 6750 6855

Email: ameya.vanjari@tcs.com
Website : www.tcs.com/consulting