

IST 722: DATA WAREHOUSE

SUPPLEMENT 2

NOTE: MATERIAL HERE IS FROM A MIX OF SOURCES AS LISTED IN REFERENCES SLIDE(S)

DATA WAREHOUSE / BI

Begin : Introductions

Part 1 : Assessing Business Process Needs

Part 2 : Designing the Data Warehouse

Part 3 : Building the Data Warehouse with ETL

Part 4 : Analysis Services and Business Intelligence

End : Project – DW Integration of 2 companies

02 AGENDA

Poll questions

CIF Reference Architecture

DW Architecture = Technical + System

Breakout: Questions

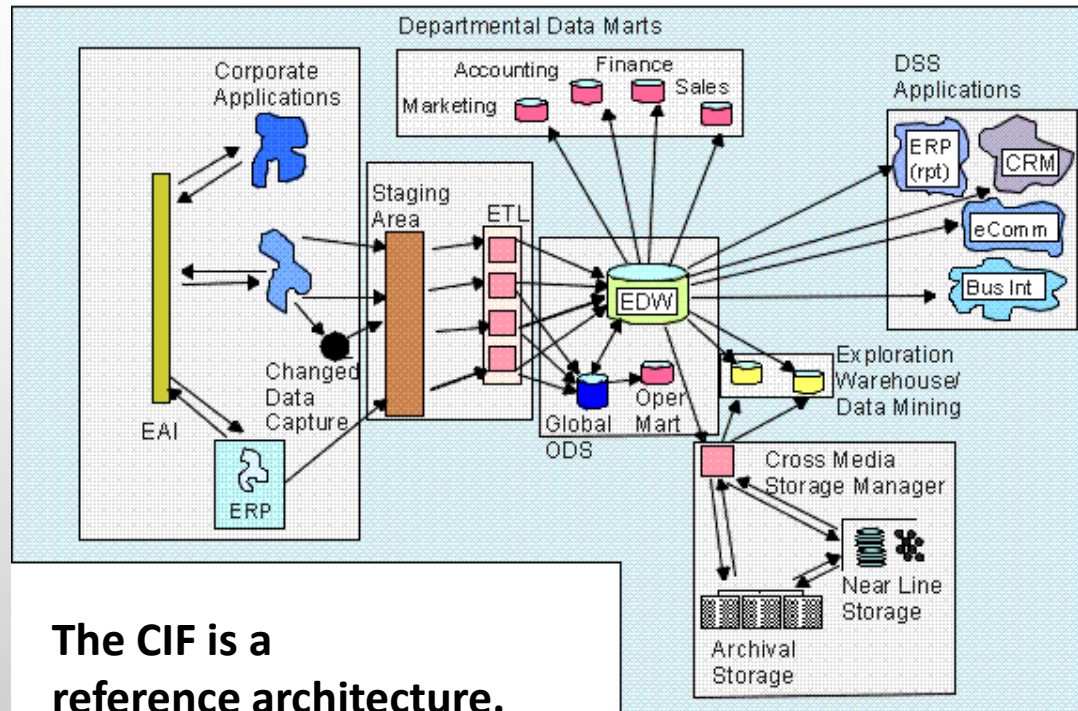
Report Out: Discussion

Assignment 2

Key Points

02 CORPORATE INFORMATION FACTORY

Corporate Information Factory



**The CIF is a
reference architecture.**

by Bill Inmon and Claudia Imhoff
Copyright ©2001, all rights reserved.

Rectangles: Components
Blobs: Applications
Boxes: Processes
Cylinders: Data Stores

02 IMPORTANCE OF STAGING DATA

Resource contention

- reduce impact on OLTP and other source systems

Consolidation

- bring data from multiple sources for integration

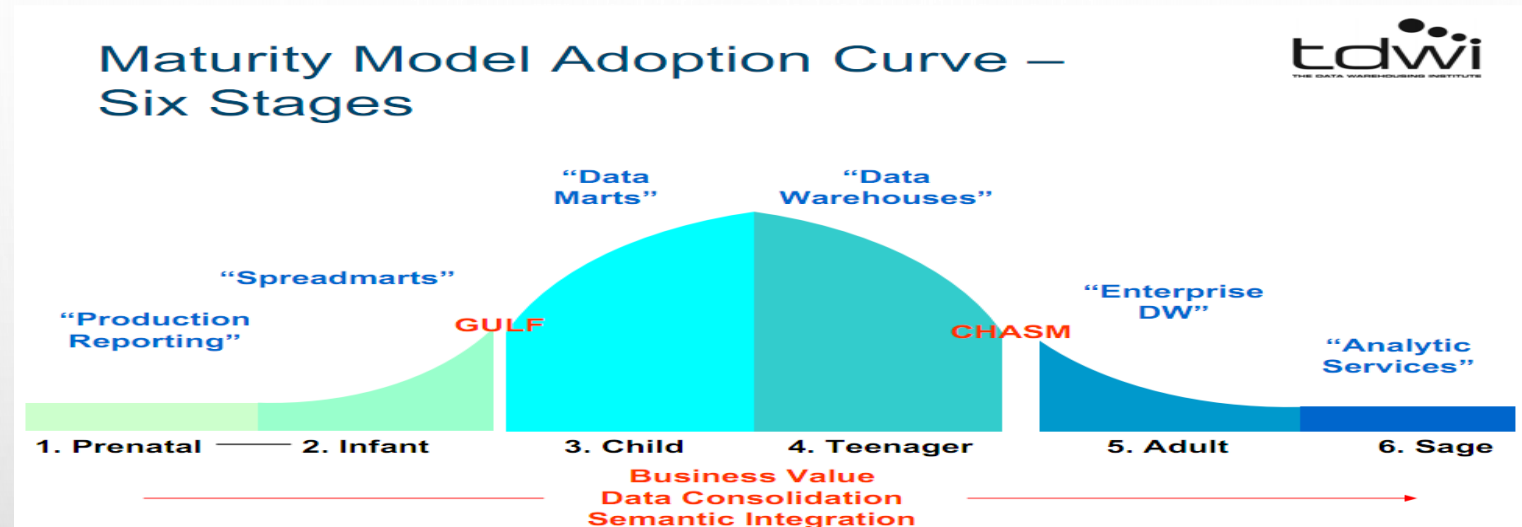
Change detection

- track changes to existing data or processed data

Snapshotting

- build time variance in the data

02 MATURITY MODEL ADOPTION CURVE



Transforming Data With Intelligence: The **Gulf** and **Chasm** represent the series of obstacles and challenges that early- and later-stage programs encounter during their BI journey, respectively. Companies in the Gulf struggle with sponsorship, funding, data quality, project scope, and spreadmarts. Companies in the Chasm struggle with the politics, logistics, and dynamics of delivering an enterprise BI environment.

<https://tdwi.org/blogs/tdwi-blog/2009/10/chasm-1.aspx>

02 BREAKOUT

- [1] What is DW Technical Architecture? Give examples.
- [2] What is DW System Architecture? Give examples.
- [3] What are the 4 types of data stores found in technical architectures?
- [4] Review the 5 technical architectures discussed throughout the coursework.
- [5] Discuss the comparative success of the 5 technical architectures. Be brief.

02 TECHNICAL AND SYSTEM ARCHITECTURES

Technical Architecture is concerned with data and data flows

- The *data flow architecture* is about how the data stores are arranged within a data warehouse and how the data flows from the source systems to the users through these data stores.
- A *data store* is one or more databases or files containing data warehouse data, arranged in a particular format and involved in data warehouse processes.
- *Data architecture* is about how the data is arranged in each data store and how a data store is designed to reflect the business processes.
- The activity to produce data architecture is known as *data modeling*.

System Architecture is concerned with supporting technical architecture

- The *system architecture* is about the physical configuration of the servers, network, software, storage, and clients.

(Vincent Rainardi)

02 SUCCESS OF THE FIVE ARCHITECTURES

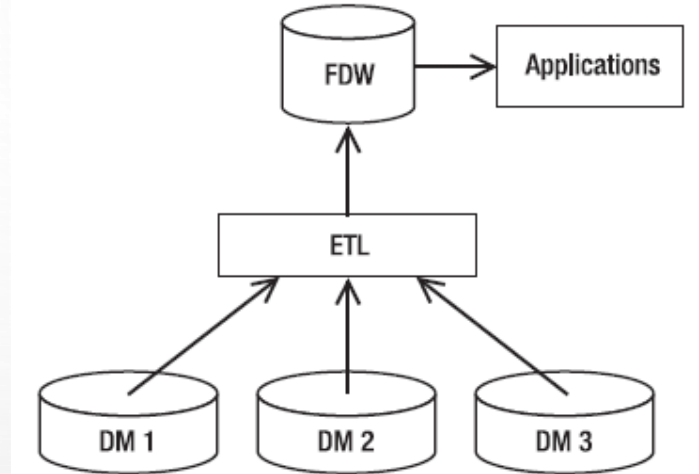
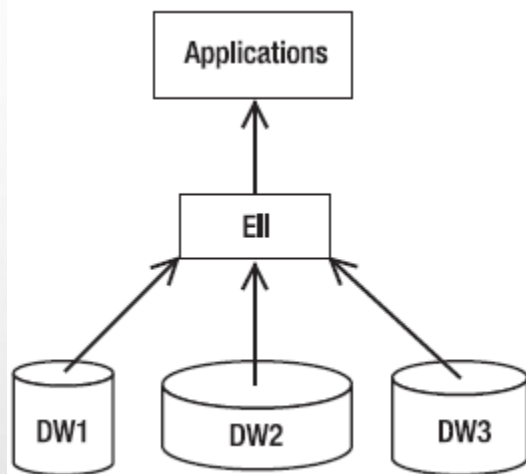
	Independent Data Marts	Bus Architecture	Hub and Spoke	Centralized (No Dependent Data Marts)	Federated
Information Quality	4.42	5.16	5.35	5.23	4.73
System Quality	4.59	5.60	5.56	5.41	4.69
Individual Impacts	5.08	5.80	5.62	5.64	5.15
Organizational Impacts	4.66	5.34	5.24	5.30	4.77

Article: Which Data Warehouse Architecture Is Most Successful? – Thilini Ariyachandra and Hugh J. Watson

Conclusion

The bus, hub-and-spoke, and centralized architectures earned similar scores on the success metrics. This finding helps explain why these competing architectures have survived over time – they are equally successful for their intended purposes. In terms of information and system quality and individual and organizational impacts, no single architecture is dominant.

02 FEDERATED DATA WAREHOUSES



Federated data warehouse = several data warehouses plus a data retrieval layer

Granularity of the FDW data is the same as the highest of the granularities of sources

ETL that extracts data from the sources needs to be aware of the timing of the data

FDW ETL needs to match the frequency of the sources

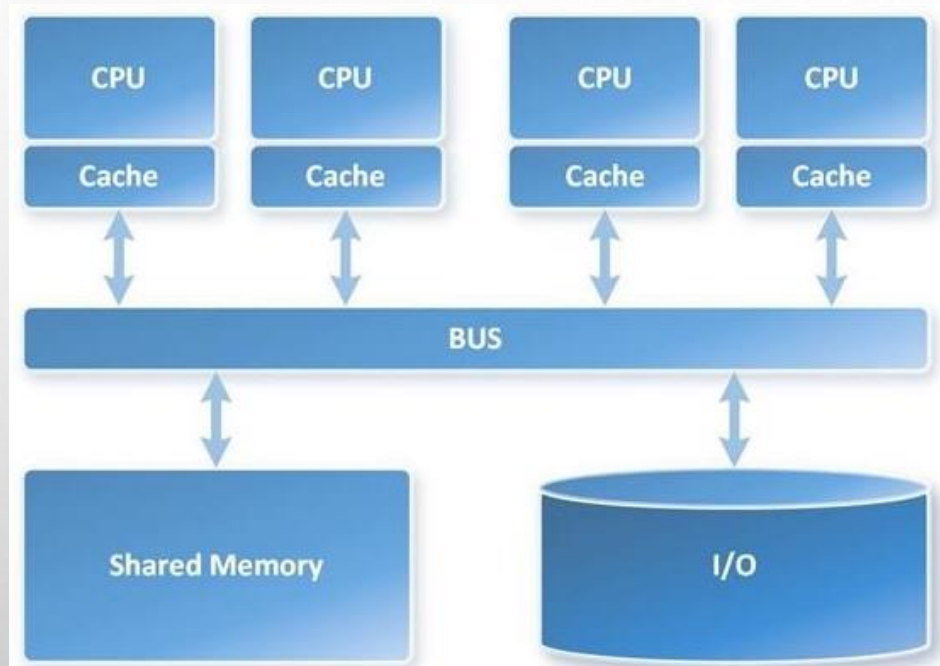
Can use the ETL approach here when integrating the data marts

Accommodates existing data warehouses, with shorter development time

Difficult to build a good-quality warehouse from diverse standards typical of DWs

02 SMP SYSTEM ARCHITECTURE

Symmetric Multi-Processing (SMP)



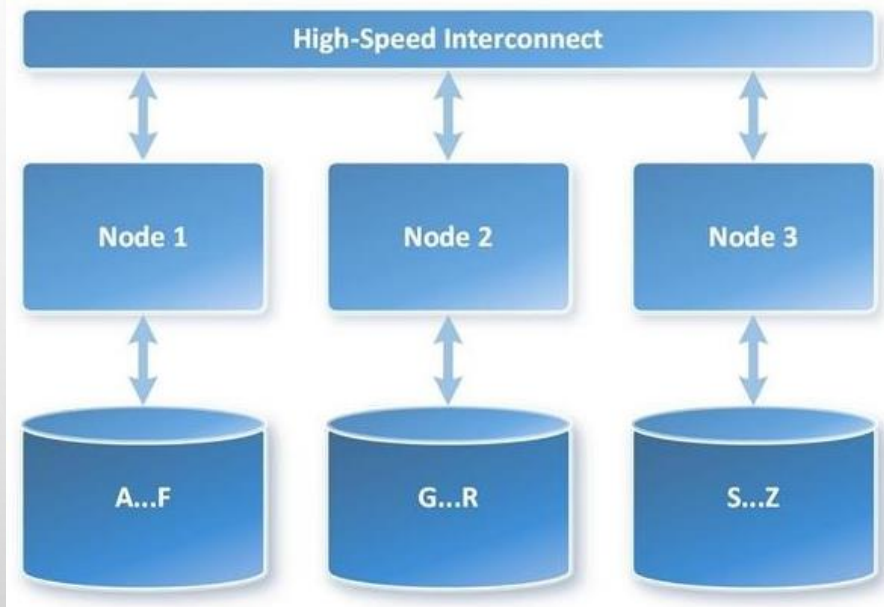
Symmetric Multi-Processing (SMP) is a tightly coupled multiprocessor system

- Single instance of the Operating System (OS), memory, I/O devices
- Connected using a common bus
- The primary parallel architecture employed in servers

<https://cloudblogs.microsoft.com/sqlserver/2014/07/30/transitioning-from-smp-to-mpp-the-why-and-the-how/>

02 MPP SYSTEM ARCHITECTURE

MPP Parallel Data Warehouse (PDW)



<https://cloudblogs.microsoft.com/sqlserver/2014/07/30/transitioning-from-smp-to-mpp-the-why-and-the-how/>

Massively Parallel Processing (MPP) is the coordinated processing of a single task by multiple processors:

- Each processor uses its own OS and memory and communicates via some form of messaging interface
- MPP can be setup with a shared nothing or shared disk architecture
- Shared nothing architecture:
 - no single point of contention across the system
 - nodes do not share memory or disk storage

02 ASSIGNMENT PRODUCTS

Part 3: On Your Own

Research the following product offerings:

1. Azure Synapse
2. Birst
3. Informatica ILM
4. MemSQL
5. Microsoft Dynamics GP
6. Oracle Essbase
7. Pentaho Data Integration
8. PostgreSQL
9. Snaplogic
10. Snowflake.com

02 ASSIGNMENT QUESTIONS

Part 3: On Your Own

For each of the product offerings, answer the following questions:

1. Website?
2. What does it do?
3. Is it a product (hardware/software), service (we do the work for you), or both?
4. Is it installed onsite, or is it in the cloud?
5. Is it systems or technical architecture?
 - a) If systems architecture, is it SMP, MPP, or Hadoop/MapReduce?
 - b) If technical architecture, which part of the corporate information factory does it aim to address? Data mart? IM&T/ETL? ODS/NDS/DDS? External world application? Cross-media storage? And so on.

02 KEY POINTS

DW Architecture = Technical Architecture + System Architecture

Five common technical architectures in order of increasing complexity: Independent data marts, Centralized, Enterprise bus, Hub and Spoke, and Federated (with ETL or EII/EAI).

Four Types of Data Stores: Stage, NDS, ODS, DDS/MDS (multidimensional).

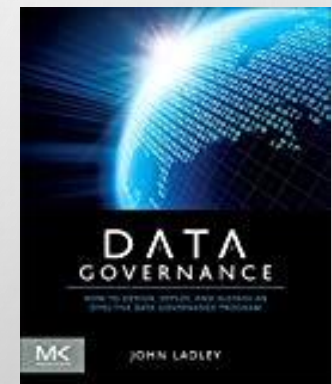
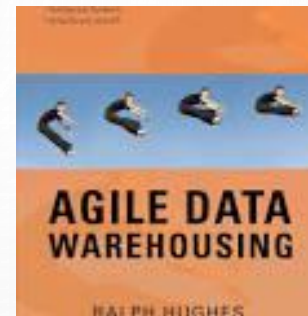
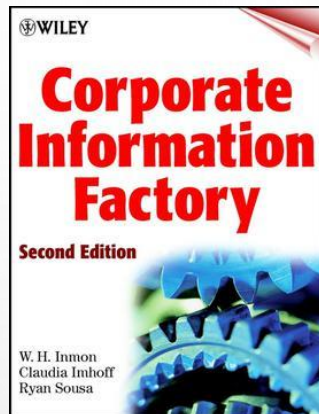
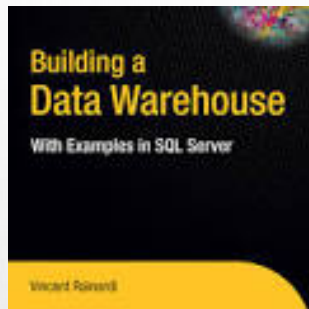
Four Reasons to Stage Data: Resource contention, Consolidation, Change detection, and Snapshotting

Corporate Information Factory (CIF) is a reference architecture (Inmon).

CIF components: Corporate Applications, ETL Processes, EDW (NDS + ODS), Data marts, Decision-Support Systems (DSS), and Cross-Media Storage.

System architecture comprises the software, platform, cloud, server, network and storage – that is the infrastructure on which the Data Warehouse sits.

REFERENCE BOOKS



REFERENCE LINKS

<https://www.1keydata.com/datawarehousing/datawarehouse.html>

<http://tdan.com/data-warehouse-design-inmon-versus-kimball/20300>

<https://www.kimballgroup.com/category/articles-design-tips/>

<https://www.slideshare.net/phanleson/building-the-data-ware-house-chapter-11>

<https://blog.chartio.com/posts/5-tips-for-selecting-the-right-data-warehouse>

https://www.tutorialspoint.com/dwh/dwh_schemas.htm

<http://www.bi-bestpractices.com/view-articles/4737>

<https://www.element61.be/en/resource/master-data-management-mdm-architecture-technology>

<https://tombreur.wordpress.com/2017/04/30/the-past-and-future-of-dimensional-modeling/>

<http://www.jamesserra.com/archive/2017/12/is-the-traditional-data-warehouse-dead/>