Business Intelligence and Data Warehousing (ANL408)

- By Sabarish Nair

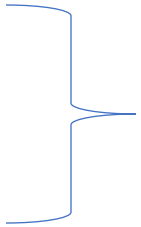
Recap from last week....

- Data Warehouse Architecture
- Enterprise Architecture
- Federated Architecture
- Multi-tiered Architecture
- Layers of data warehouse
- Staging Layer
- Temporary Staging Layer
- Persistent Staging Layer
- Data Mart
- Types of Data Mart
- Advantages and Drawbacks of Data Mart
- Data Mart vs Data Warehouse

Traditional Data Warehouse

Centralized warehouse repositories **within an organization**. They are designed to support:

- Reporting
- Analytics
- Business Intelligence



DECISION – MAKING!!!

Traditional Warehouse - Characteristics

- Structured Data Storage
- ETL Processes
- Dimensional Modelling
- Query and Reporting
- Data Marts
- Historical Data Storage
- Centralized Architecture
- Scalability and Performance

Modern Data Warehouse

Advanced data management platforms that are designed to handle large volumes of structured and unstructured data.

REAL-TIME INSIGHTS!!!

Modern Warehouse - Characteristics

- Cloud Based Architecture
- Massively Parallel Processing (MPP)
- Columnar Storage
- In-Memory Processing
- Integration with Big Data Technologies
- Advanced Analytics and Machine Learning
- Self-Service Analytics

Traditional vs Modern Data Warehouse

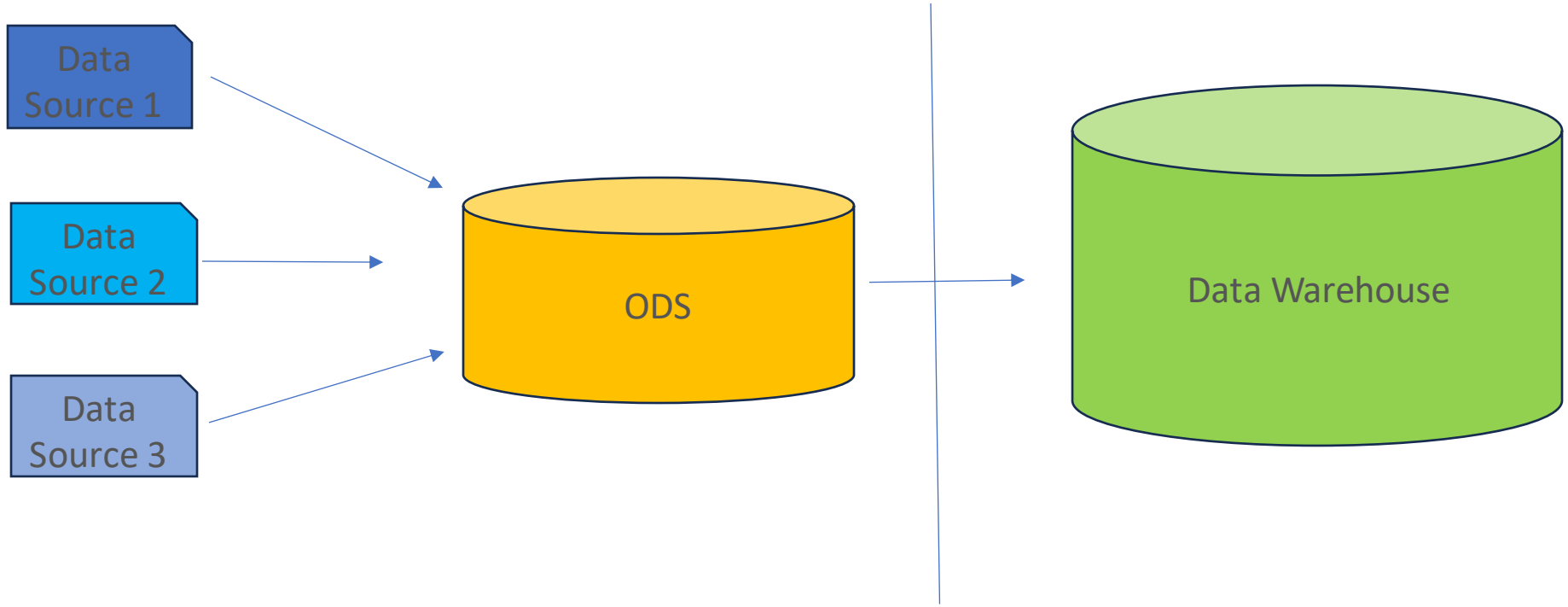
Parameter	Traditional Data Warehouse	Modern Data Warehouse
Architecture	On-premises or Private Cloud	Cloud-based (private, public or hybrid cloud)
Scalability	Limited scalability, requires upfront capacity planning	Elastic scalability, can scale resources up or down as needed
Processing Model	Single-node architecture with shared resources	Distributed, massively parallel processing (MPP) architecture
Data Processing	Row-based storage and processing	Columnar storage and processing for optimized query performance
Integration	Limited integration with external data sources	Integrates with big data technologies and external data sources
Real-time Analytics	Limited support for real-time analytics	Support for real-time analytics and streaming data processing
Cost	Upfront hardware and software costs, ongoing maintenance	Pay-as-you-go pricing model, based on resource usage
Maintenance and Updates	Manual maintenance and updates	Automated maintenance and updates through cloud providers

Operational Data Store (ODS)

"A subject-oriented, integrated, volatile, current valued data store containing detailed data."

"Type of database that serves as a central repository for integrating and storing data from various operational systems within an organization."

Visualizing ODS



Characteristics of ODS

- Real-Time or Near Real-Time Data Integration
- Centralized Data Repository
- Transactional Processing
- No need for long history
- Data Quality and Consistency
- Parallel vs Sequential

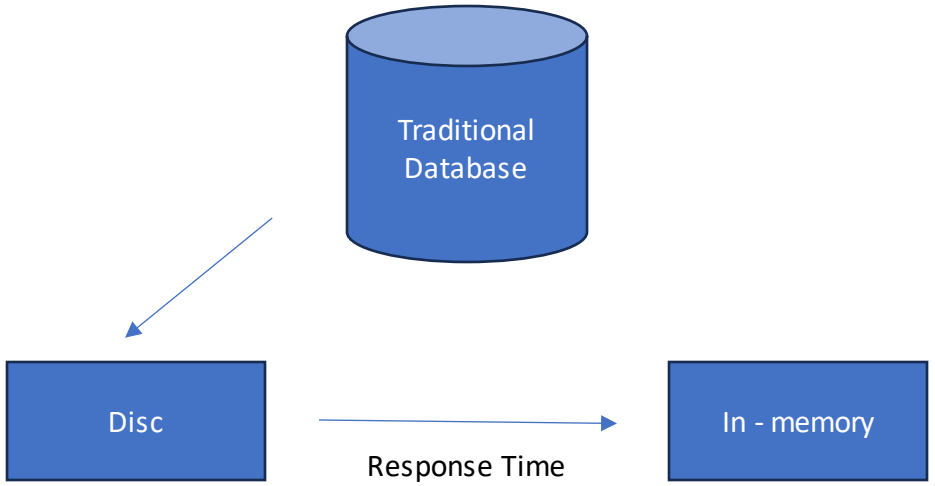
Staging vs ODS

Parameter	Staging Area	ODS
Purpose	Temporary storage for raw data before transformation	Central repository for integrated and current operational data
Data Transformation	Minimal transformation, raw data format	Integrated, cleansed, and transformed data
Data Persistence	Temporary, overwritten or deleted after loading	Persistent, historical data (*data retained as per business needs*)
Performance	Optimized for data ingestion and processing efficiency	Optimized for real-time or near real-time updates and queries
Schema Flexibility	Flexible schema to accommodate changes in source data	Structured and stable schema for consistent data model

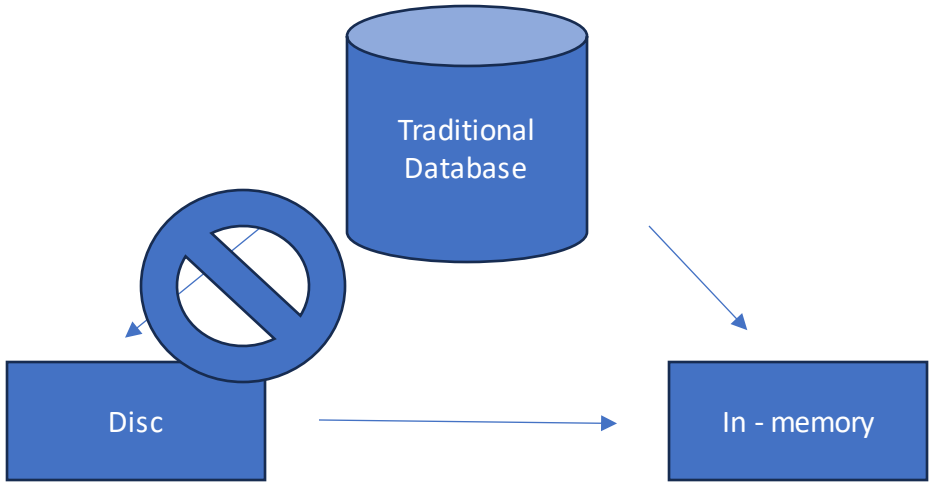
In-Memory Databases

"*Type of database management system (DBMS) that primarily stores and manages data in main memory (RAM) rather than on disk storage.*"

Traditional Databases



In-memory databases

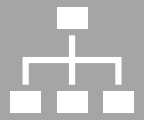


Characteristics : In Memory Databases

- Data Storage in RAM
- High Performance
- Data Compression
- Columnar Storage
- Durability and Persistence (Periodic Snapshots or backups to disc)
- Scalability

Metadata : Overview

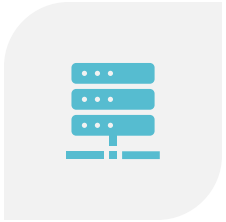
"*Data about Data E.g. Data Dictionary*"



"*Provides information about the structure, content, and context of the data, enabling users and administrators to understand, manage, and utilize the data effectively.*"

Types of Metadata

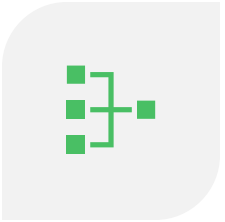
STRUCTURAL METADATA (TABLES, KEYS, RELATIONSHIPS, ETC.)



OPERATIONAL METADATA (ETL, MAINTENANCE, ETC.)

BUSINESS METADATA (BUSINESS CONTEXT AND SEMANTICS)

TECHNICAL METADATA (STORAGE STRUCTURE, INDEXES, ETC.)



DATA LINEAGE METADATA (TRACKS MOVEMENT OF DATA IN DWH)

DATA QUALITY METADATA (PROFILING, VALIDATION RULES, ETC.)

USAGE METADATA (LOGS, QUERY PERFORMANCE, ETC.)

Benefits of Metadata

- Improved Data Understanding
- Enhanced Data Quality
- Increased Data Accessibility
- Streamlined Data Governance
- Empowered Data Analytics

Metadata Best Practices

- Standardize Metadata Definitions
- Document Metadata
- Centralize Metadata management
- Automate metadata capture
- Implement metadata versioning
- Enable search and discovery
- Monitor metadata usage

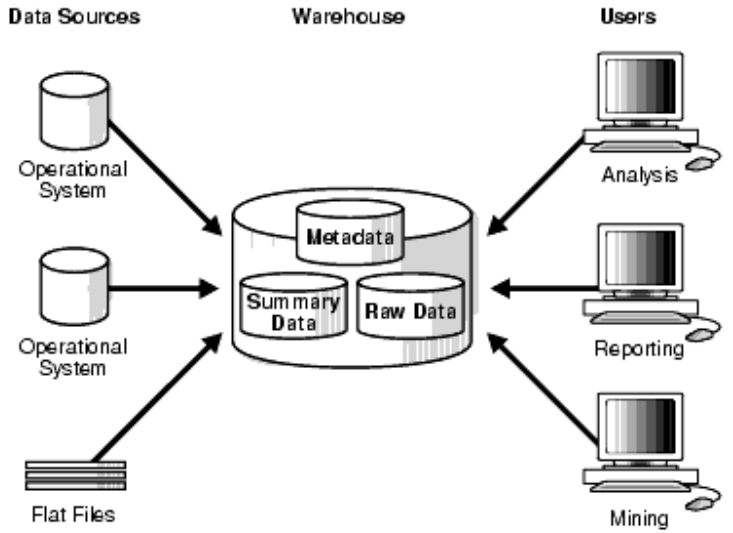
Data Warehousing approach

Methodology or strategy employed in designing, implementing, and managing a data warehouse environment.

- Inmon's approach (Top-Down Approach)
- Kimball's approach (Bottom-Up Approach)

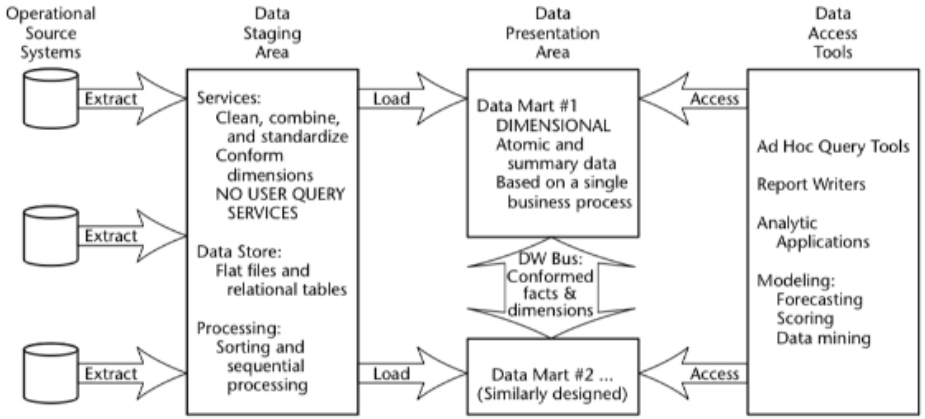
Inmon's Data warehousing Approach

- Enterprise DW
- Normalized Data Model
- Data Integration
- Data Staging
- Data Marts (subset of DW)
- Long Term Strategic Perspective



Kimball's Data warehousing Approach

- Dimensional Data Modelling
- Star and Snowflake Schema
- Data Mart-Centric
- Iterative Development
- Business User Empowerment
- Botton-Up Scalability

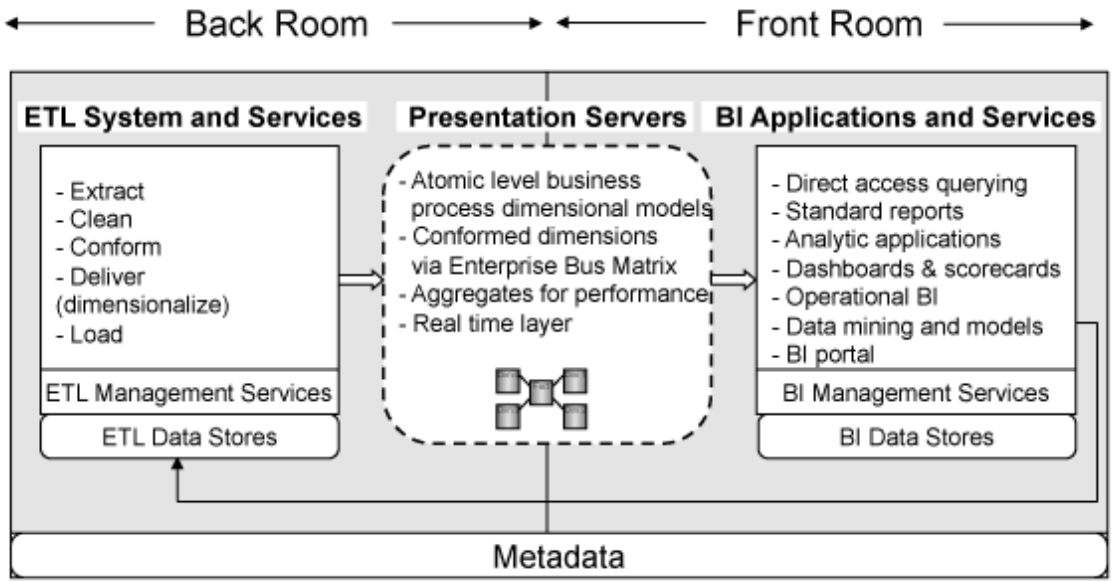


Kimball's vs Inmon's Approach

Parameter	Kimball's Approach	Inmon's Approach
Philosophy	Bottom-Up Approach	Top-Down Approach
Centralization	Emphasizes decentralized data marts	Advocates for a centralized Enterprise Data Warehouse (EDW)
Data Integration	Data Marts are independently integrated	Extensive data integration process
Data Marts	Business specific and tailored	Derived from centralized DW
Development Approach	Iterative and Incremental	Long-term strategic
User Empowerment	Prioritizes business user empowerment	Focuses on building a robust foundation for EDW
Data Modelling	Dimensional	Normalized
Scalability	Botton-Up	Top-Down
Agility	Agile Approach	Strategic Approach

Kimball's front and back room analogy

- Front Room (User-Facing)
- Back Room (Technical)




Thank you