Business Intelligence and Data Warehousing (ANL408)

- By Sabarish Nair

Recap from last week....

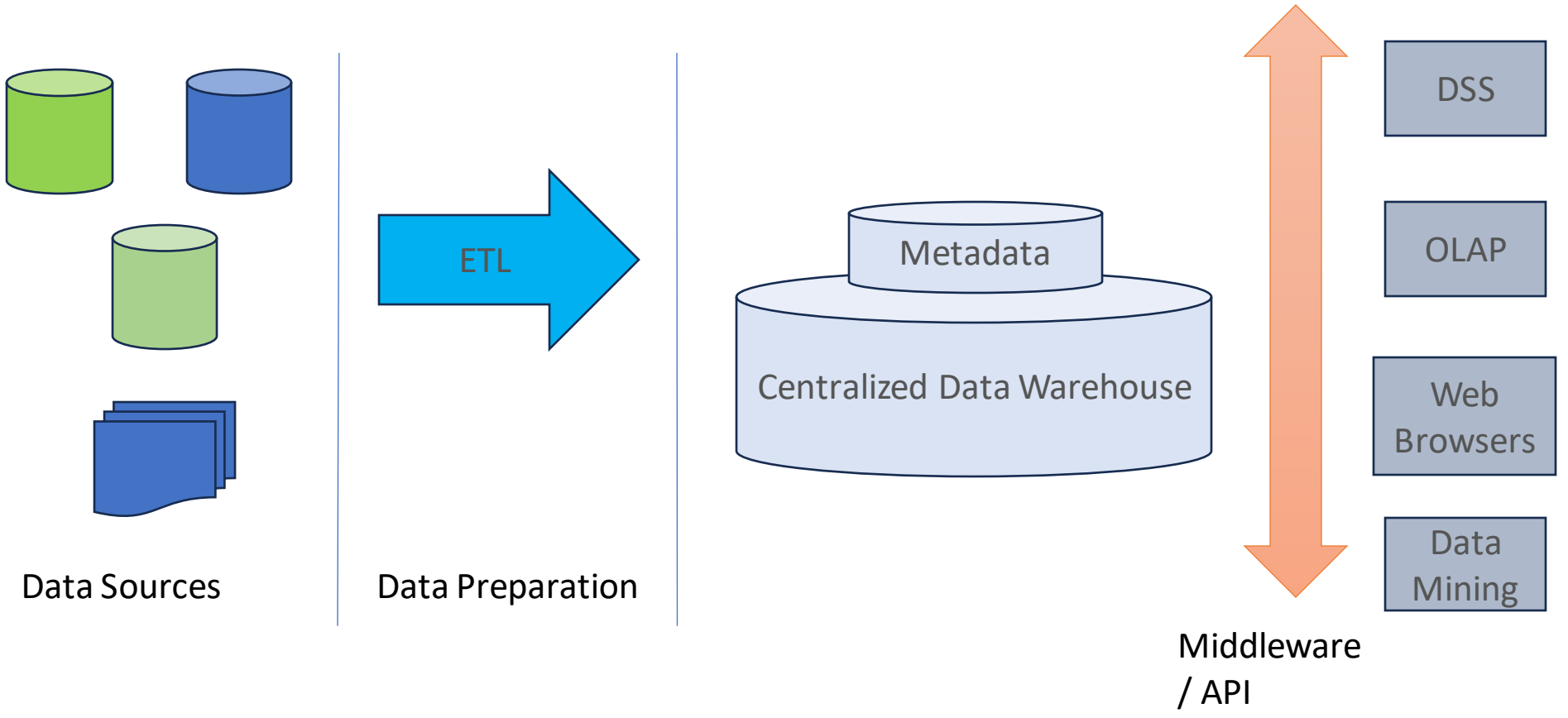
- Data Explosion
- Types of Data
- Database and its type
- Why do we need data warehouse?
- What is a data warehouse?
- Components of data warehouse
- Software for data warehouse
- Business Intelligence
- BI Example
- BI Process
- BI Components
- BI Tools
- Practical example of BI

Data Warehouse Architecture

Three major data warehouse architecture types:

- Enterprise or Centralized Architecture
- Federated Architecture
- Multi-Tiered Architecture

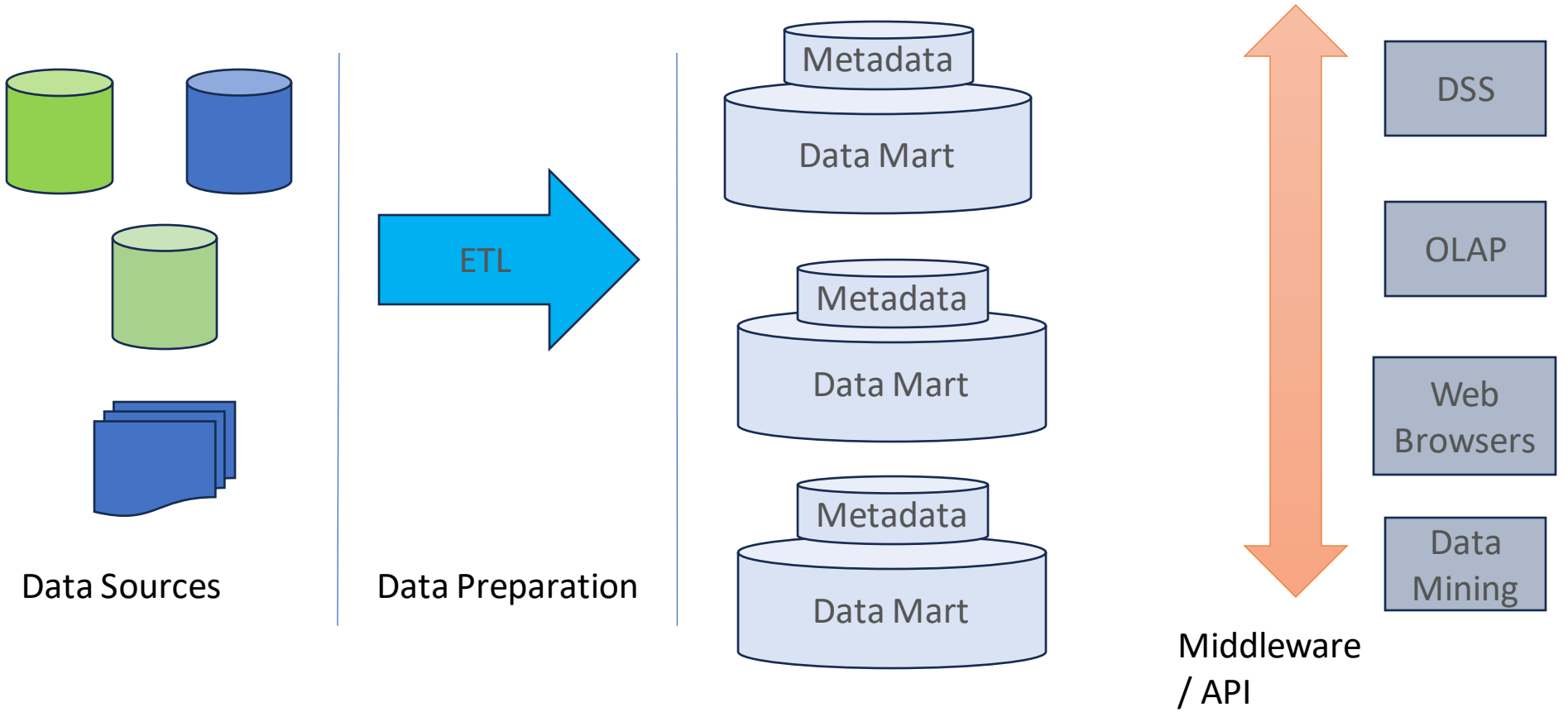
Enterprise/ Centralized Data warehouse



Benefits of Centralized Architecture

- Unified Data Management System
- 360-degree view of company's operation
- Easier data governance
- Centralized and single point of control

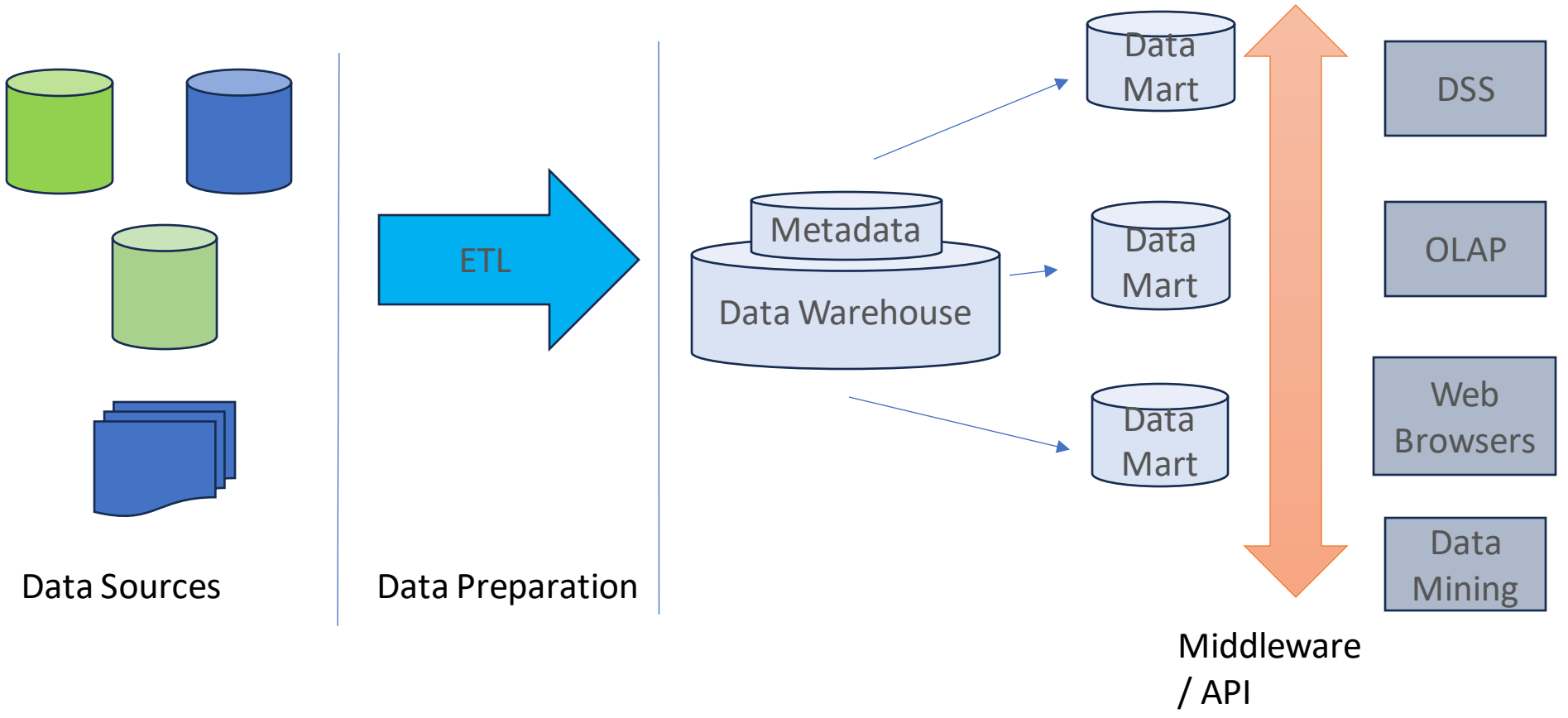
Federated Data warehouse/ Single Department Data Mart



Benefits of Federated System

- Flexibility and agility
- Scalability
- Enhanced collaboration and data sharing
- Improved fault tolerance and resiliency
- Governance and regulatory compliance

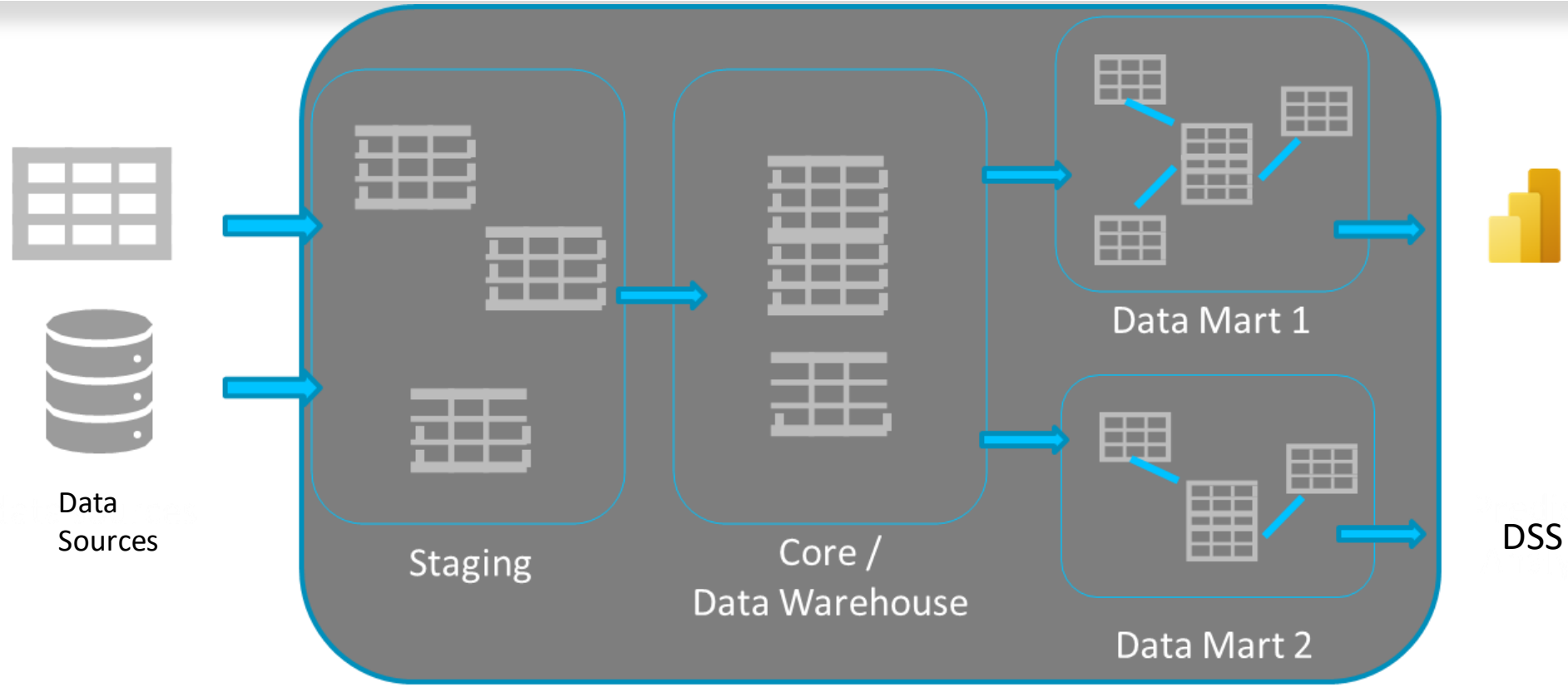
Multi-tiered Architecture



Benefits of Multi-tiered architecture

- Scalability
- Better Performance
- Modularity
- Security
- Improved resource management
- Easier Maintenance
- Improved Reliability

Layers of a data warehouse



Staging Layer

"A staging area, or landing zone, is an intermediate storage area used for data processing during the extract, transform, and load process."

"An intermediate area where raw data from various sources is initially loaded before being transformed and integrated into the data warehouse."

Staging area: Key Points

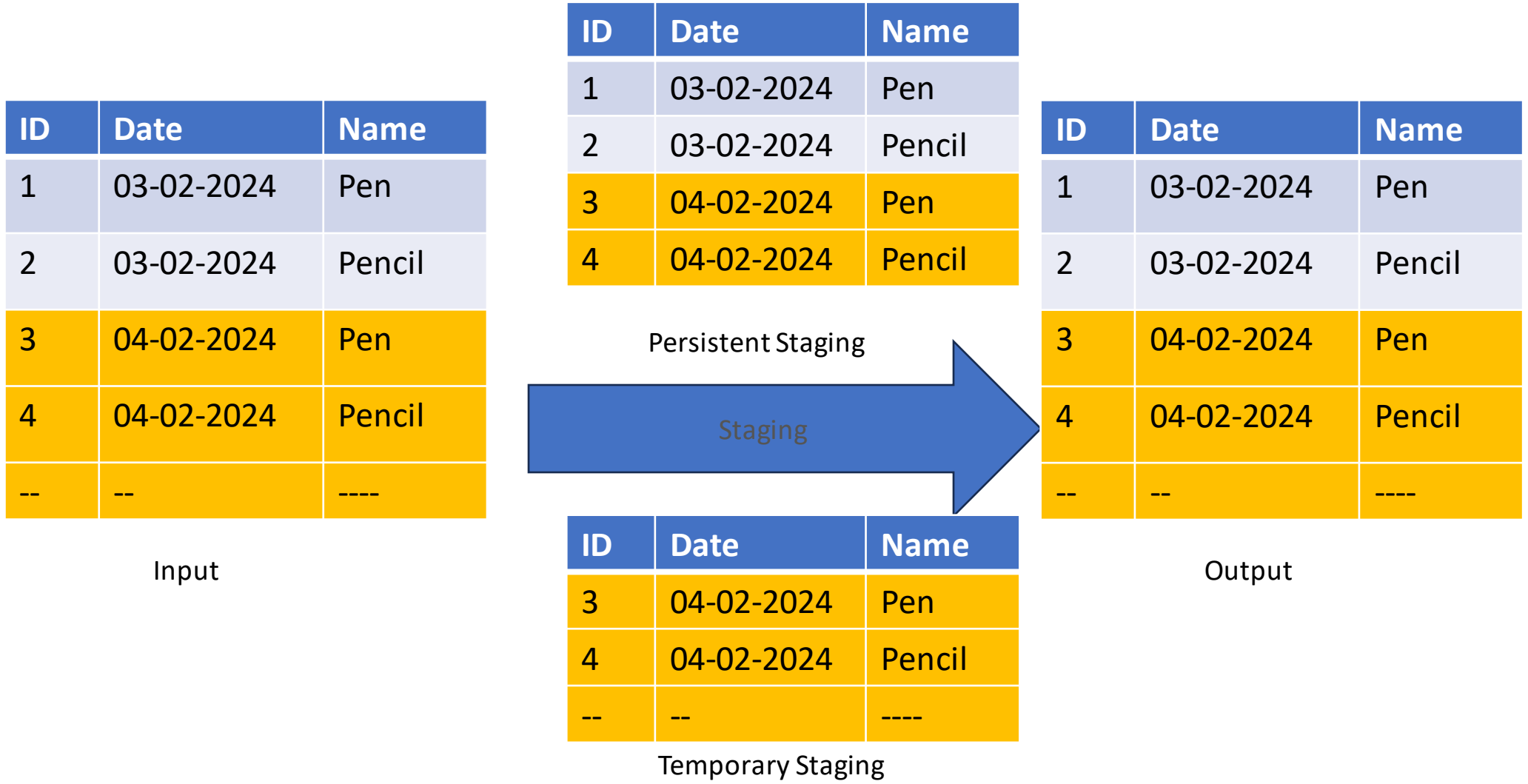
- Data Ingestion
- Data Cleansing
- Data Transformation
- Data Integration
- Temporal Storage
- Performance Optimization

Types: Staging Areas

TEMPORARY

PERSISTENT

Staging Area: Example 1 (Temporary vs Persistent Staging)



Staging Area: Example 2 (Combining Tables)

PID	Date	Name
1	03-02-2024	Pen
2	03-02-2024	Pencil

Input 1: Product Information

SID	PID	Name	Location
1	1	Lidl	Cork
2	2	Dunnes	Dublin

Input 2: Store Information

PID	Date	Name	SID
1	03-02-2024	Pen	1
2	03-02-2024	Pencil	2

Output

Staging Area: Example 3 (Merging Tables)

PID	Date	Name
1	03-02-2024	Pen
2	03-02-2024	Pencil

Input 1: Product Information

PID	Date	Name
3	03-02-2024	Ink
4	03-02-2024	Rubber

Input 2: Product Information

PID	Date	Name
1	03-02-2024	Pen
2	03-02-2024	Pencil
3	03-02-2024	Ink
4	03-02-2024	Rubber

Output

Summarizing Staging Layer

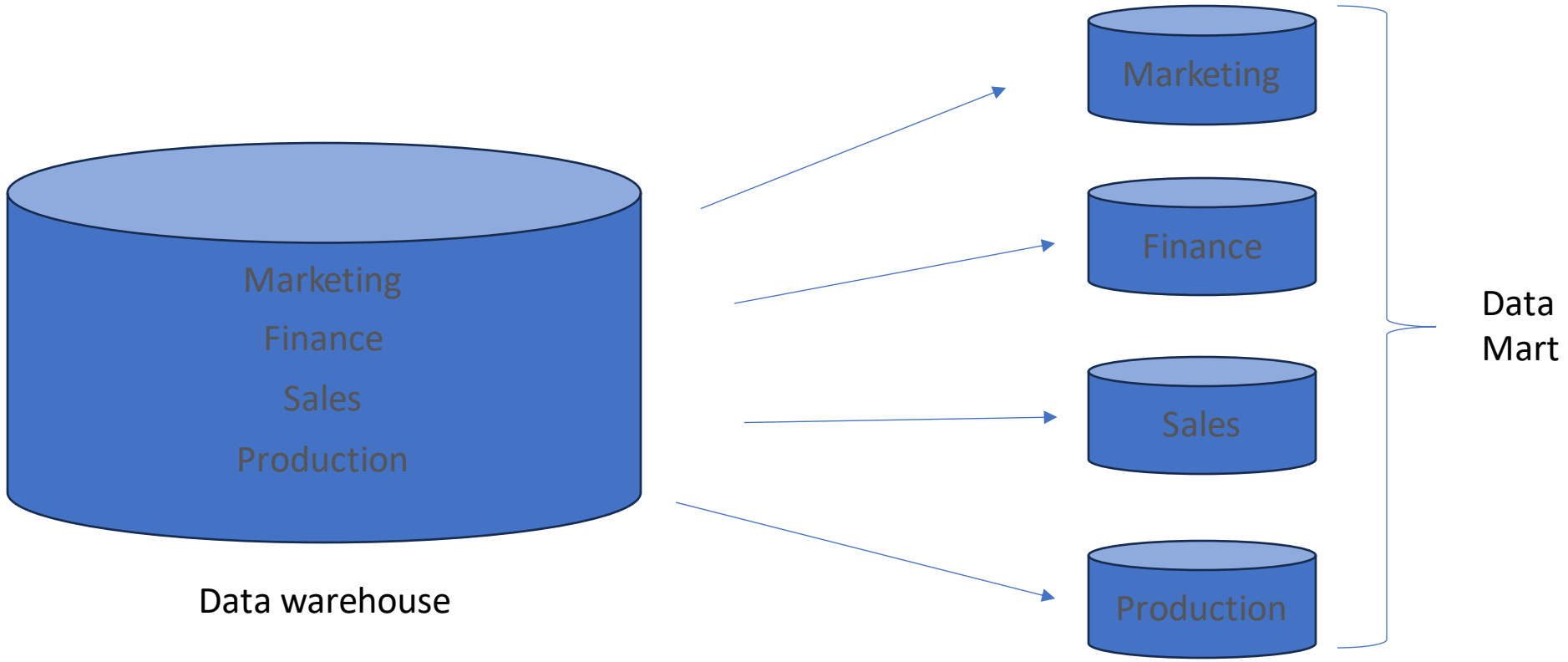
- Landing Zone
- Pulls data from multiple sources
- Temporary or Persistent staging layers
- Separate database

Data Mart

"A data mart is a specialized subset of a data warehouse focused on a specific functional area or department within an organization."

"Small-scale data warehouse that focuses on a specific business problem."

Data Mart: Example



Data Mart: Characteristics

- Restricted Scope
- Usability
- Closely match data requirements
- Life cycle
- Scope of coverage
- Tool dependent

Types of Data Mart

- Independent
- Dependent
- Hybrid

Data Mart: Advantages

- Smaller Scope
- Visualization
- Closely match data requirements
- Linked with other BI tools
- Focused data
- Improved Performance
- Ease of use
- Flexibility and scalability
- Business Agility
- Enhanced Security

Data Mart: Drawbacks

- Limited Scope
- Data Redundancy
- Data Inconsistency
- No common consolidates source
- No common consolidated cleaning
- Issue with cross data marts

Data Mart vs Data Warehouse

Data Warehouse	Data Mart
Centralized System	Decentralised system
Top- Down Model	Botton-Up Model
Long life	Shorter life
Vast Data	Smaller Data
Collects data from various sources	Collects data from warehouse
Optimum efficiency for business analysis	Faster response to requests and information


Thank you