Master Data Management and Data Warehousing

Zahra Mansoori

1. Preference

IT landscape growth

IT landscapes have grown into complex arrays of different systems, applications, and technologies over the last several decades and creates significant data problems

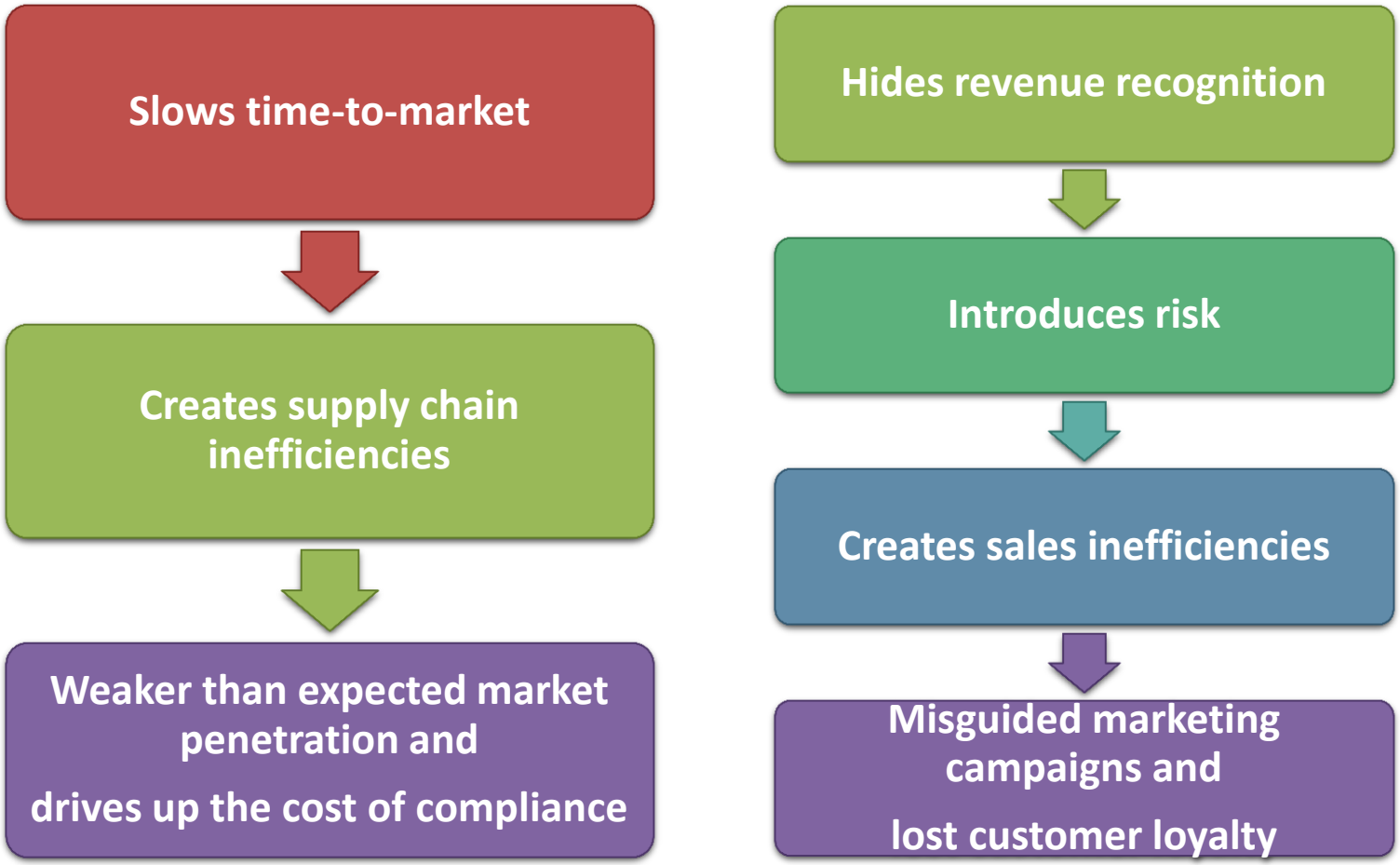
Impeding initiatives of:

- Customer relationship management (CRM)
- Enterprise resource planning (ERP)
- Supply chain management (SCM)

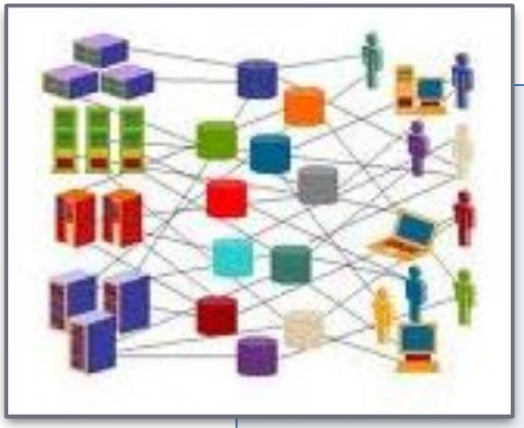
Corrupting analytics

Costing corporations billions of dollars a year

Fragmented inconsistent Product data defects



A critical question arises

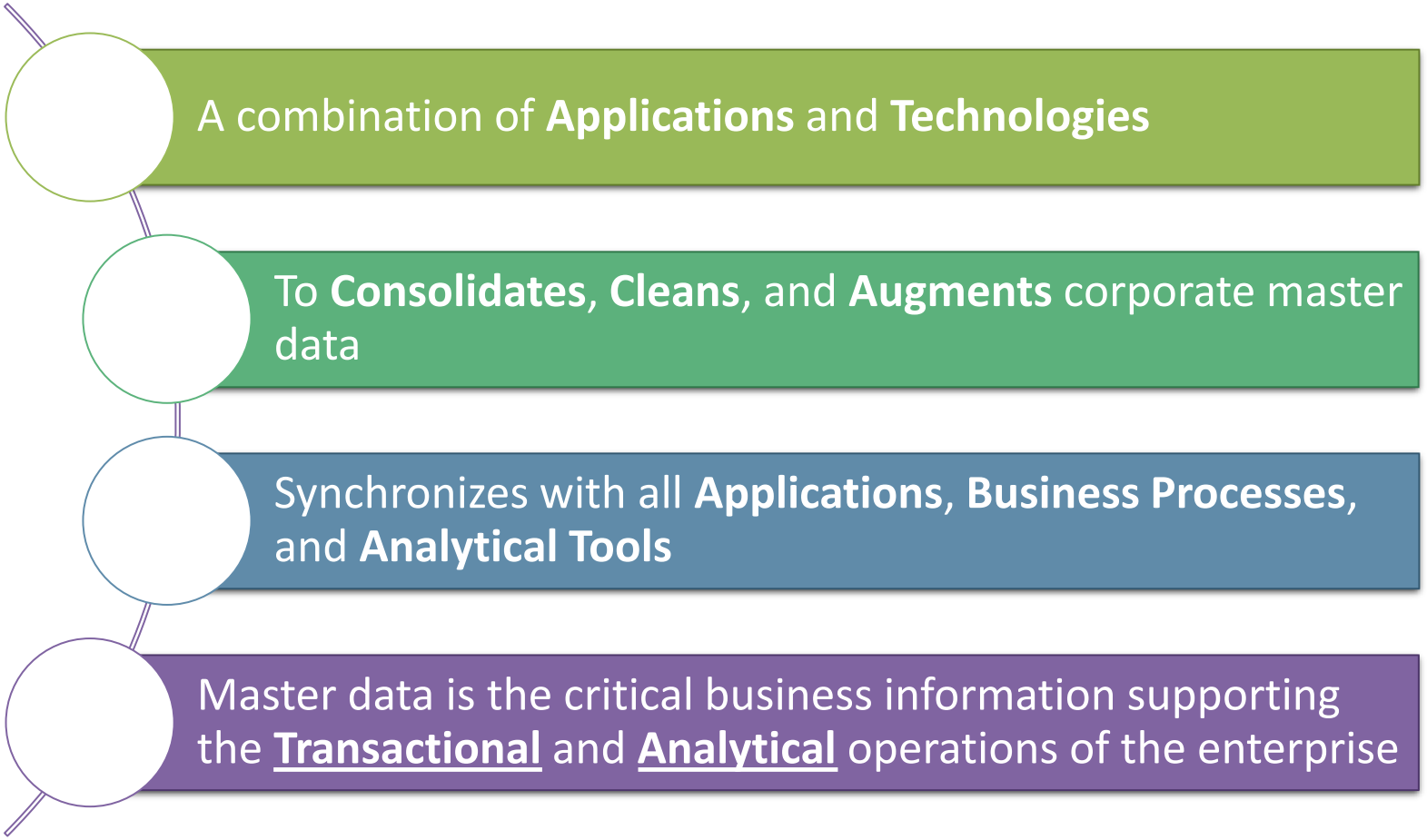


How do you get from a thousand points of data entry to a single view of the business?

We are going to answer this question…

2. Introduction to MDM

Master Data Management (MDM)



Master Data Management (MDM) Characteristics

Master data management has two architectural components:

- The technology to profile, consolidate and synchronize the master data across the enterprise
- The applications to manage, cleanse, and enrich the structured and unstructured master data

Integrate with modern service oriented architectures (SOA)

- And bring the clean corporate master data to the applications and processes that run the business

Integrate with data warehouses and the business intelligence (BI) systems

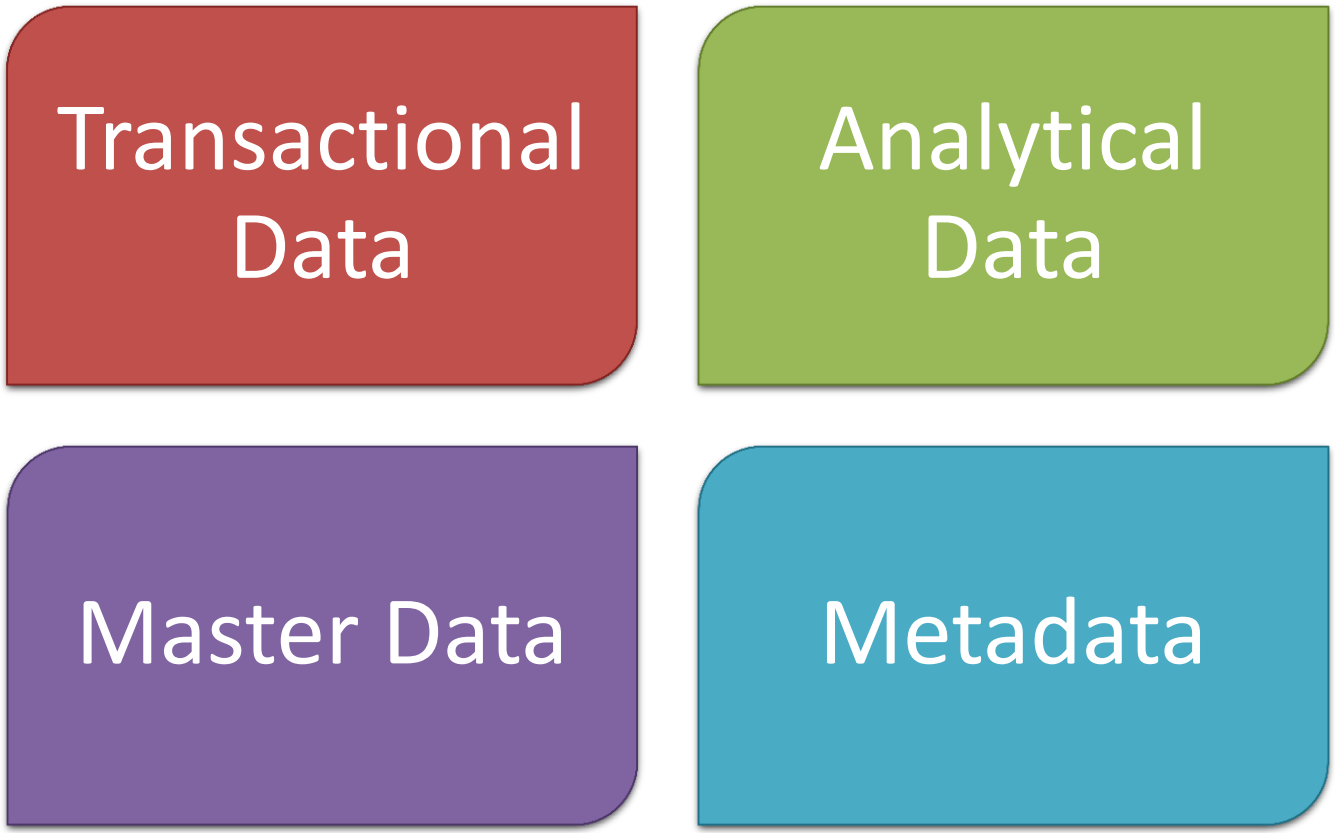
- Bring the right information in the right form to the right person at the right time

Support data governance

- Enables orchestrated data stewardship across the enterprise

3. Enterprise Data

Enterprise Data



Transactional Data : OLTP

- Significant amounts of data caused by a company's operations:
  - Sales, service, order management, manufacturing, purchasing, billing, accounts receivable and accounts payable
- The objects of the transaction are the customer and the product
- Data stores in **OnLine Transaction Processing (OLTP)** tables
- Support high volume low latency access and update
- Master data solution is called: **Operational MDM**

Analytical Data: OLAP

- Support a company's decision making
- Identify
  - Churn, profitability, and marketing segmentation
  - Suppliers categorization based on performance, for better supply chain decisions
  - Product behavior over long periods to identify failure patterns
- Data is stored in large data warehouses and possibly smaller data marts with table structures
- Data stores in **OnLine Analytical Processing (OLAP)** tables
- Master data solution is called: **Analytical MDM**
  - Lack the ability to influence operational systems

Master Data:
A Single Version Of The Truth

- Master Data represents
  - Business objects that are shared across more than one transactional application
  - Key dimensions around which analytics are done
- Must support high volume transaction rates

Master Data Hub (Also called Dimensions)

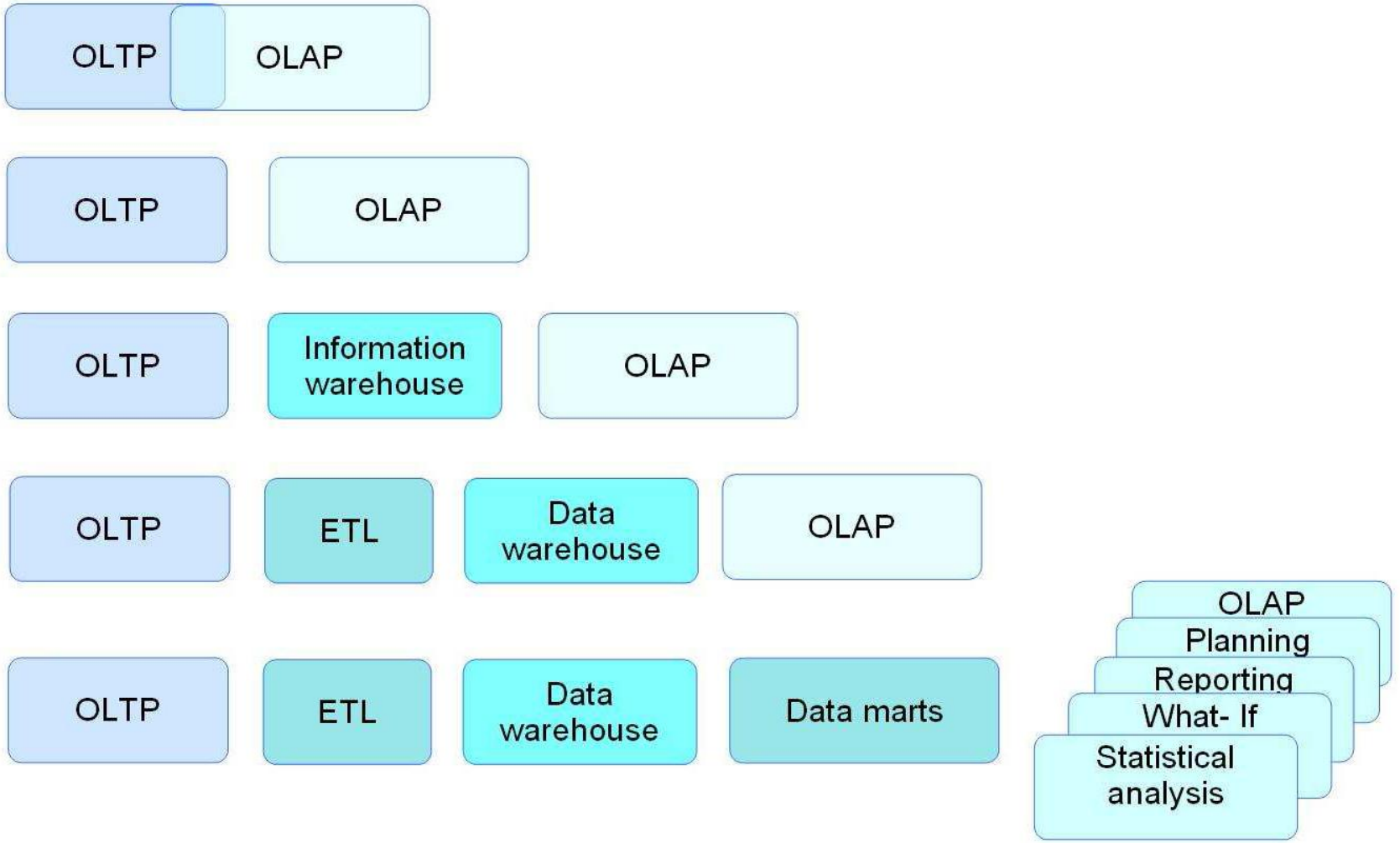
Enterprise MDM

- Maximum business value comes from managing both **Transactional** and **Analytical** master data
- Operational data cleansing:
  - improves the operational efficiencies of the applications and the business process
- Analytical analysis:
  - true representations of how the business is actually running
- The insights realized through analytical processes, are made available to the operational side of the business

4. Data Warehouse

Data Warehousing

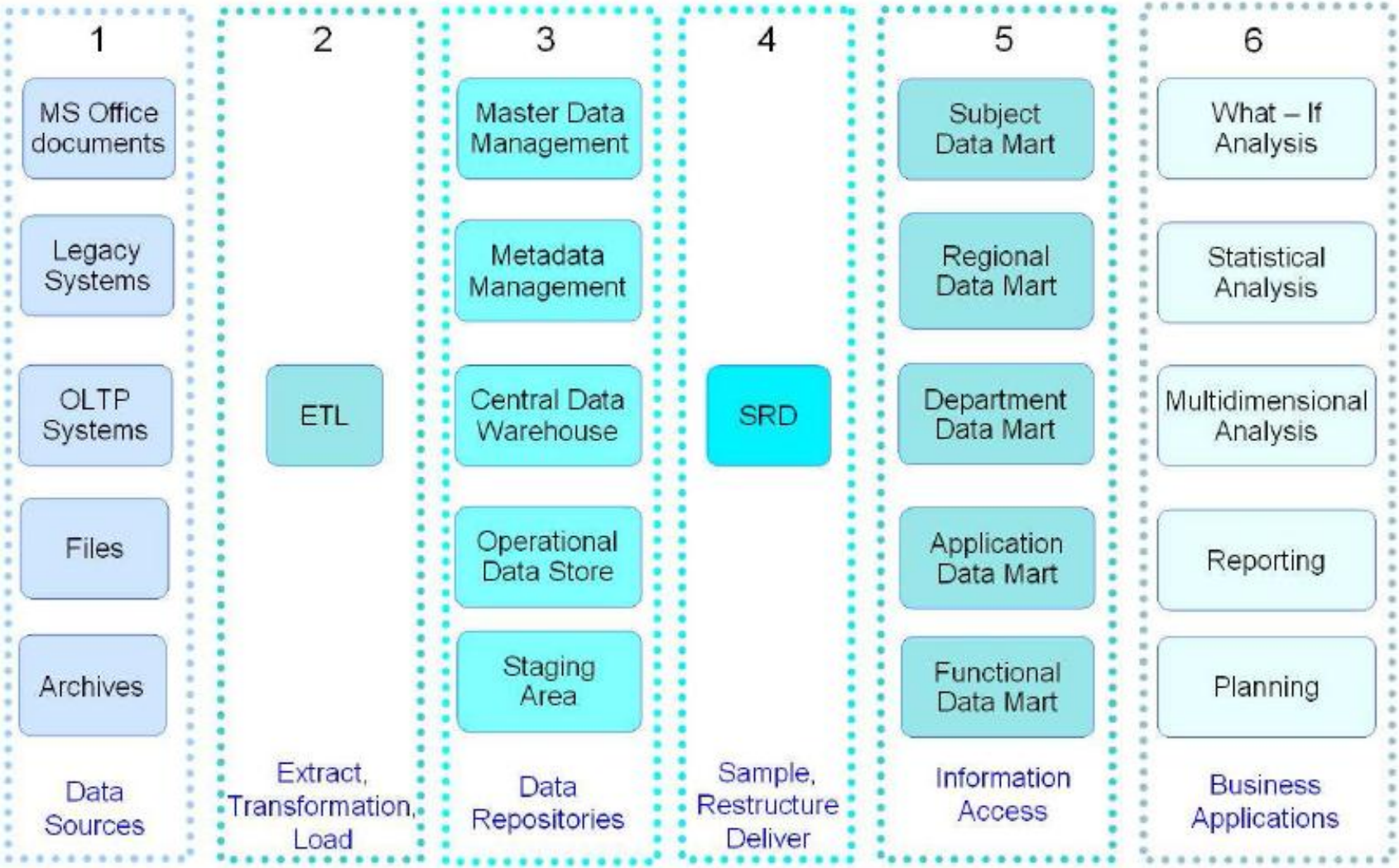
- Somewhat similar to a water purification system.
- Water with different chemical composition is collected from various sources.
- specific cleaning and disinfection methods are applied for each case of water source
- Water delivered to the consumers meets strict quality standards



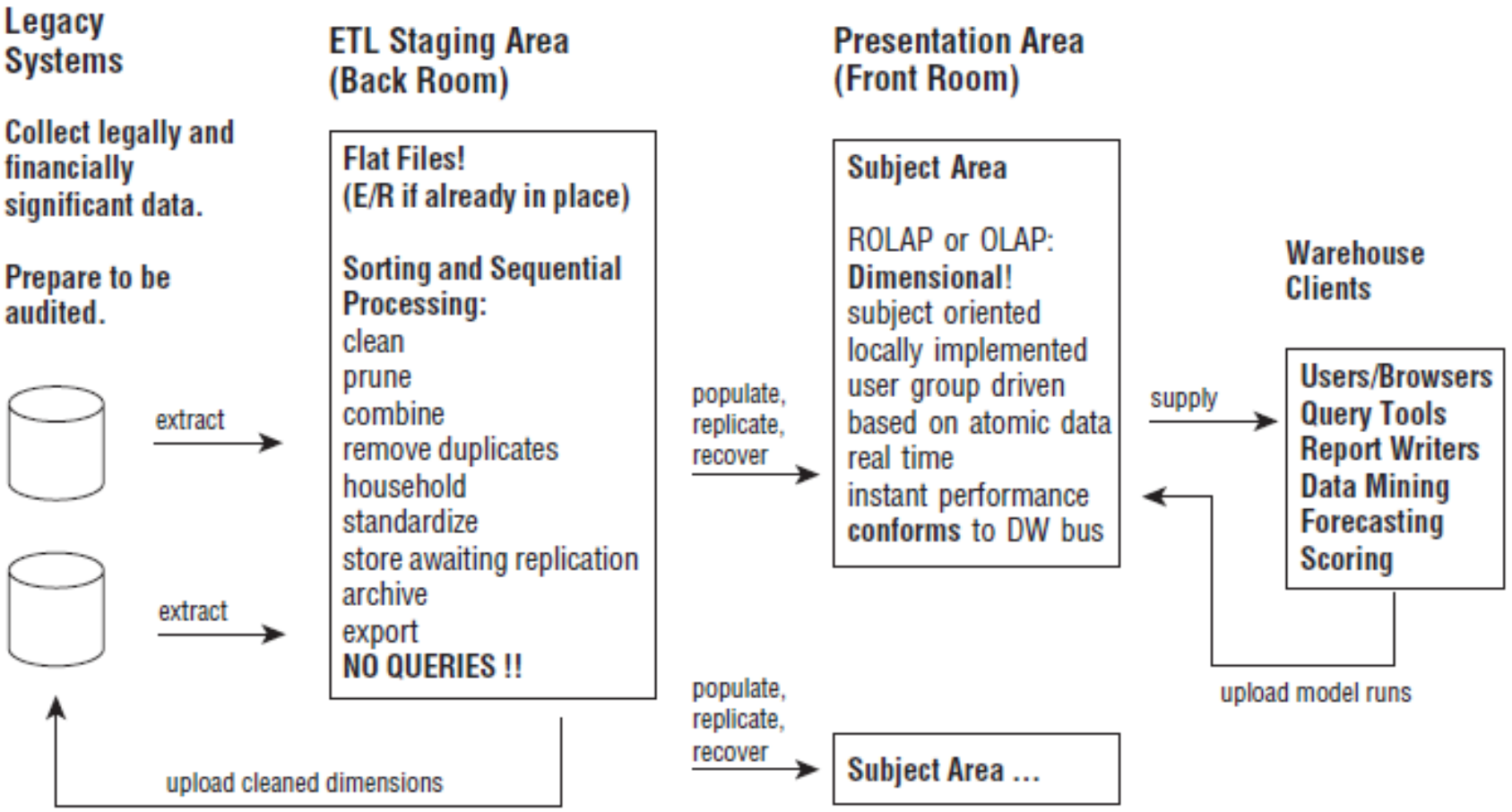
Evolution of OLTP and OLAP understandings

ETL (Extraction, Transformation, and Loading)

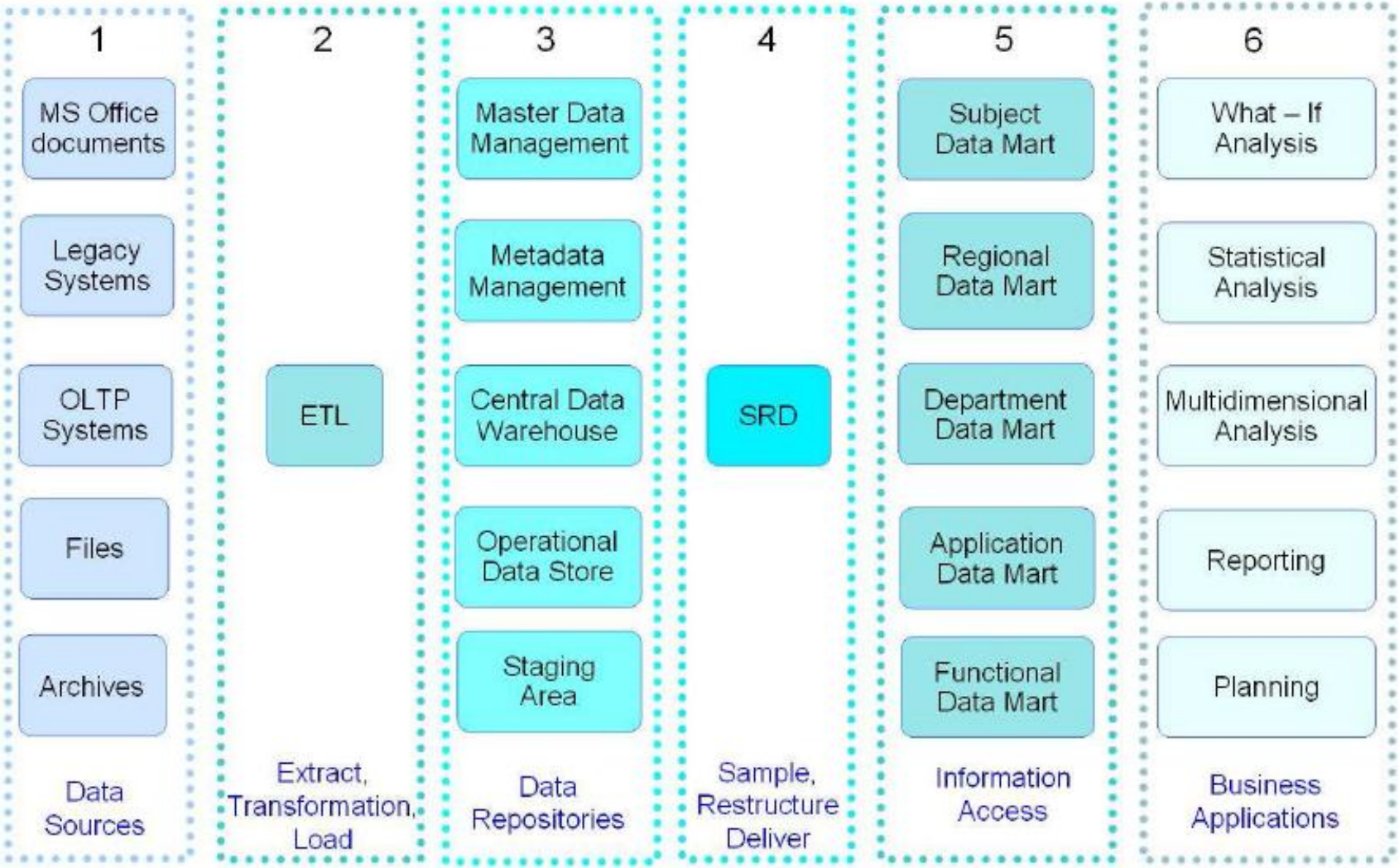
- main objective of ETL is to extract data from multiple sources to bring them to a consistent form and load to the DW



Six layers of DW architecture


Legacy Systems
Collect legally and financially significant data.
Prepare to be audited.
extract
extract
ETL Staging Area (Back Room)
Flat Files!
(E/R if already in place)
Sorting and Sequential Processing:
clean
prune
combine
remove duplicates
household
standardize
store awaiting replication
archive
export
NO QUERIES !!
upload cleaned dimensions
populate, replicate, recover
populate, replicate, recover
Presentation Area (Front Room)
Subject Area
ROLAP or OLAP:
Dimensional!
subject oriented
locally implemented
user group driven
based on atomic data
real time
instant performance
conforms to DW bus
Subject Area ...
supply
Warehouse Clients
Users/Browsers
Query Tools
Report Writers
Data Mining
Forecasting
Scoring
upload model runs

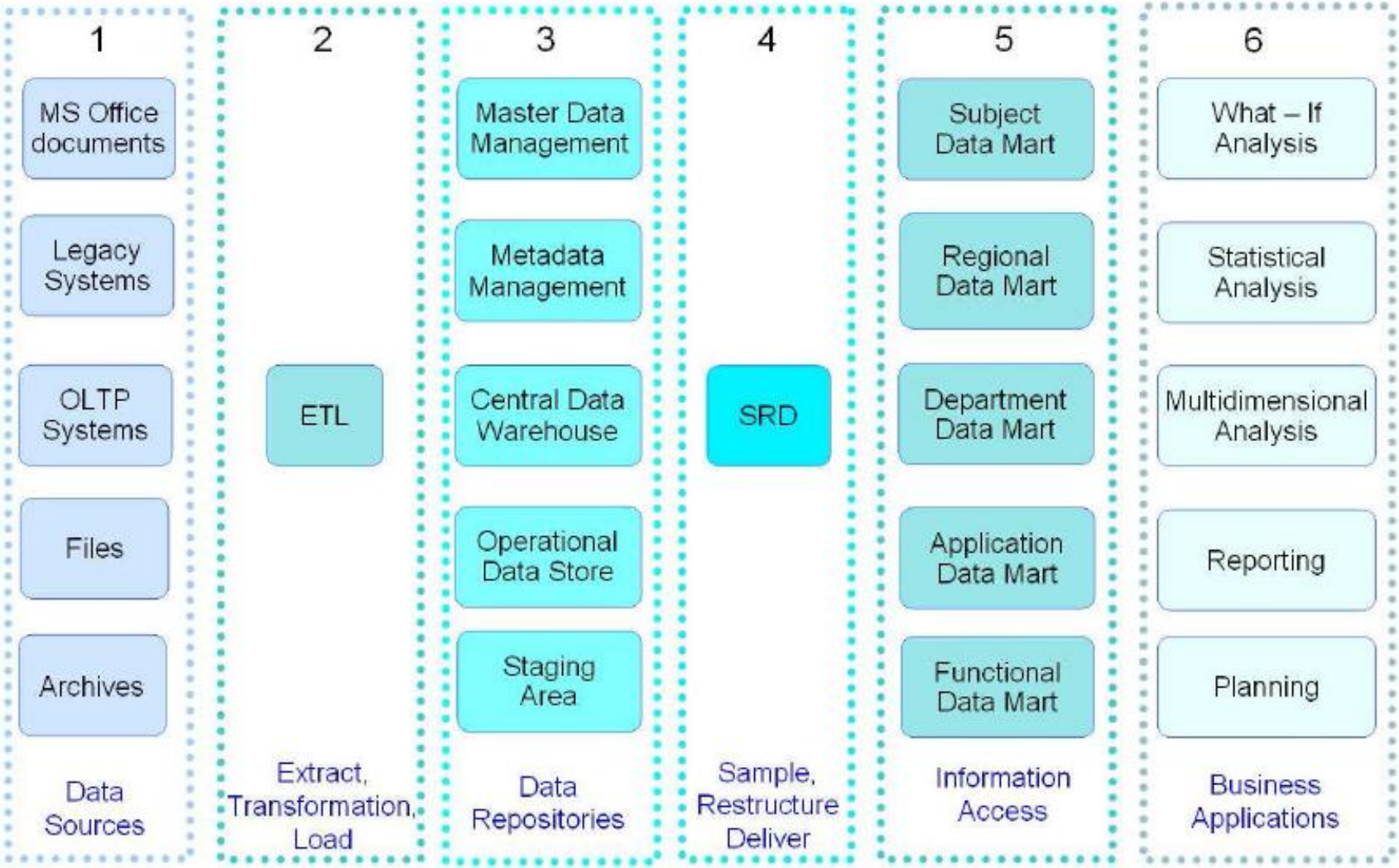
ETL



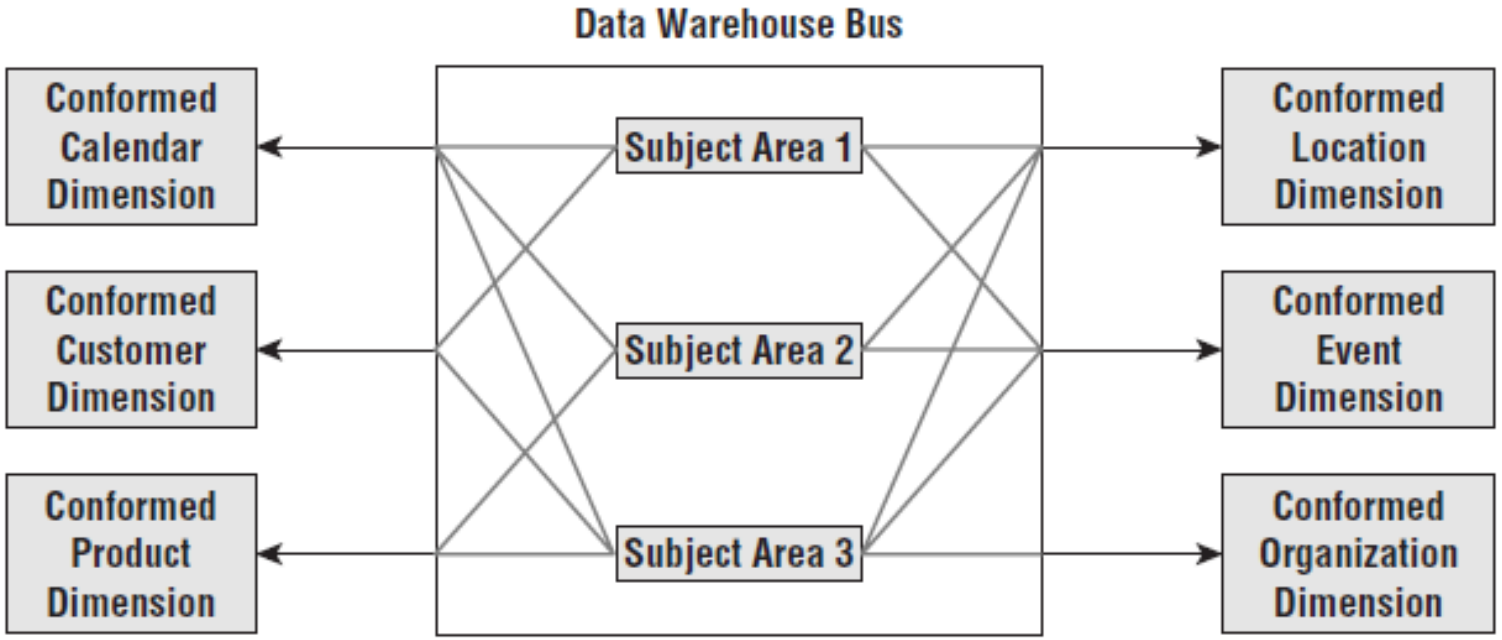
Six layers of DW architecture - SRD

SRD vs. ETL

- **SRD** can be simply called **ETL**, as well as the system of extraction, transformation, and loading on the second layer
- To emphasize the differences from **ETL**, **SRD** are sometimes called **ETL-2**
- tasks of **SRD** differ significantly from the tasks of **ETL**, namely, sampling, restructuring, and data delivery (**SRD** - Sample, Restructure, Deliver)
- **ETL** extracts data from a variety of external systems, but **SRD** selects data from a single **DW**
- **ETL** receives inconsistent data that are to be converted to a common format, but **SRD** has to deal with purified data
- **ETL** loads data into a central **DW**, but **SRD** shall deliver the data in different data marts in accordance with the rights of access, delivery schedule and requirements for the information set



Six layers of DW architecture – Data Marts



The data warehouse bus architecture

Showing a series of data marts connecting to the conformed dimensions of the enterprise

5. Information Architecture

Contents

4.1. Operational Application

- Heterogeneous Applications
- Transactional data exists in the applications local data store
- Data needs to be synchronized in order to support business processes that cross these application boundaries

The n^2 Integration Problem

- complexity grows geometrically with the number of applications
- Some companies have been known to call their data center connection diagram a “hair ball”
- IT projects can grind to a halt
- costs quickly become prohibitive

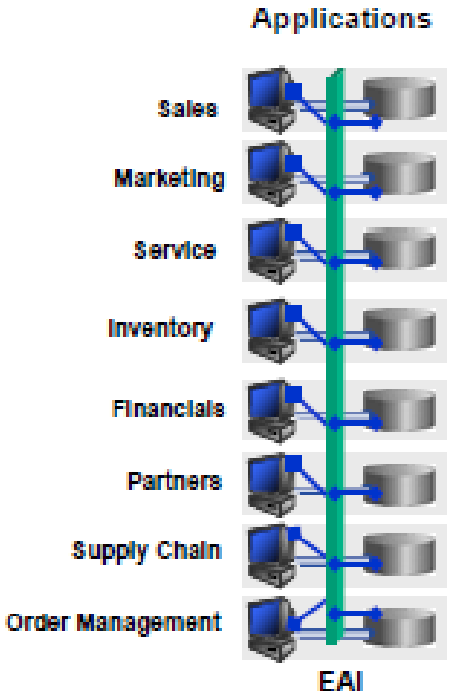
This problem literally drove the creation of Enterprise Application Integration (EAI) technology

Enterprise Application Integration (EAI)

- **Enterprise Service Bus** or **Integration Hub**

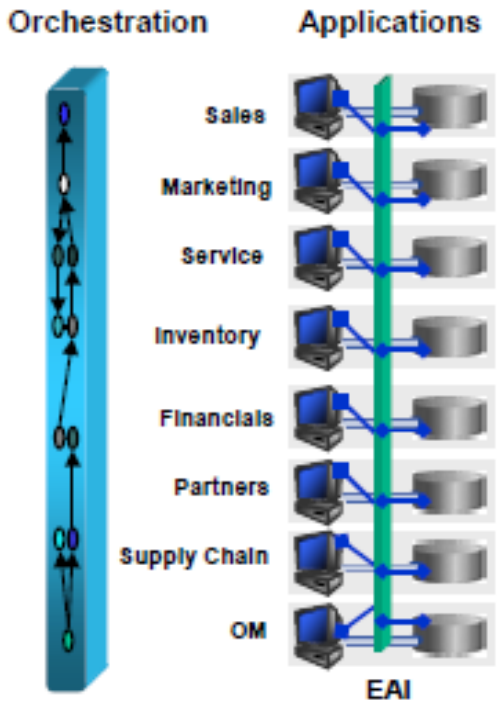
- Uses a metadata driven approach to synchronizing the data across the operational applications at runtime
  - Information about what data needs to move
  - When it needs to move
  - What rules to follow as it moves
  - What error recovery processes to use, etc.

is stored in the metadata repository of the EAI tool



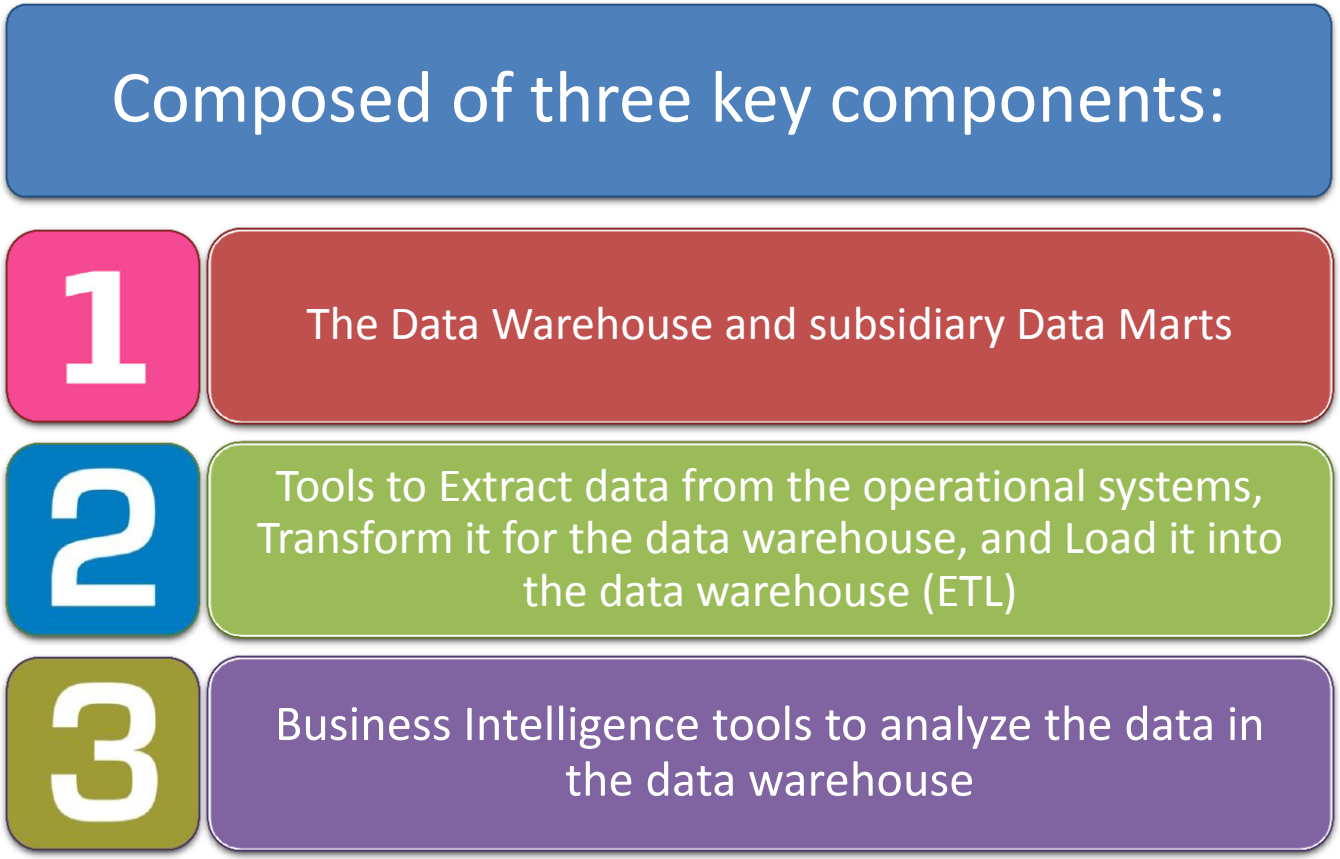
Service Oriented Architecture (SOA)

- The features and functions of the applications are exposed as shared services using standardized interfaces
- End-to-end business processes by a technique called **Business Process Orchestration**
- The Data Quality Problem still remains



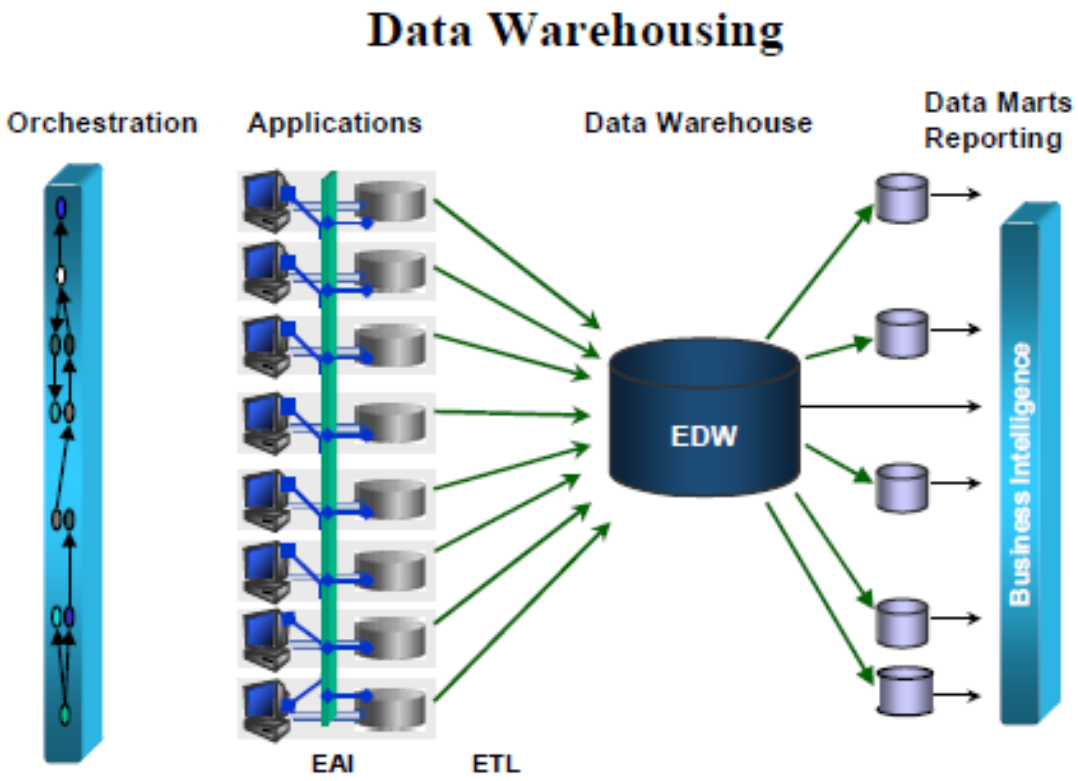
4.2. Analytical Systems

- Data warehousing to as a single view of the truth

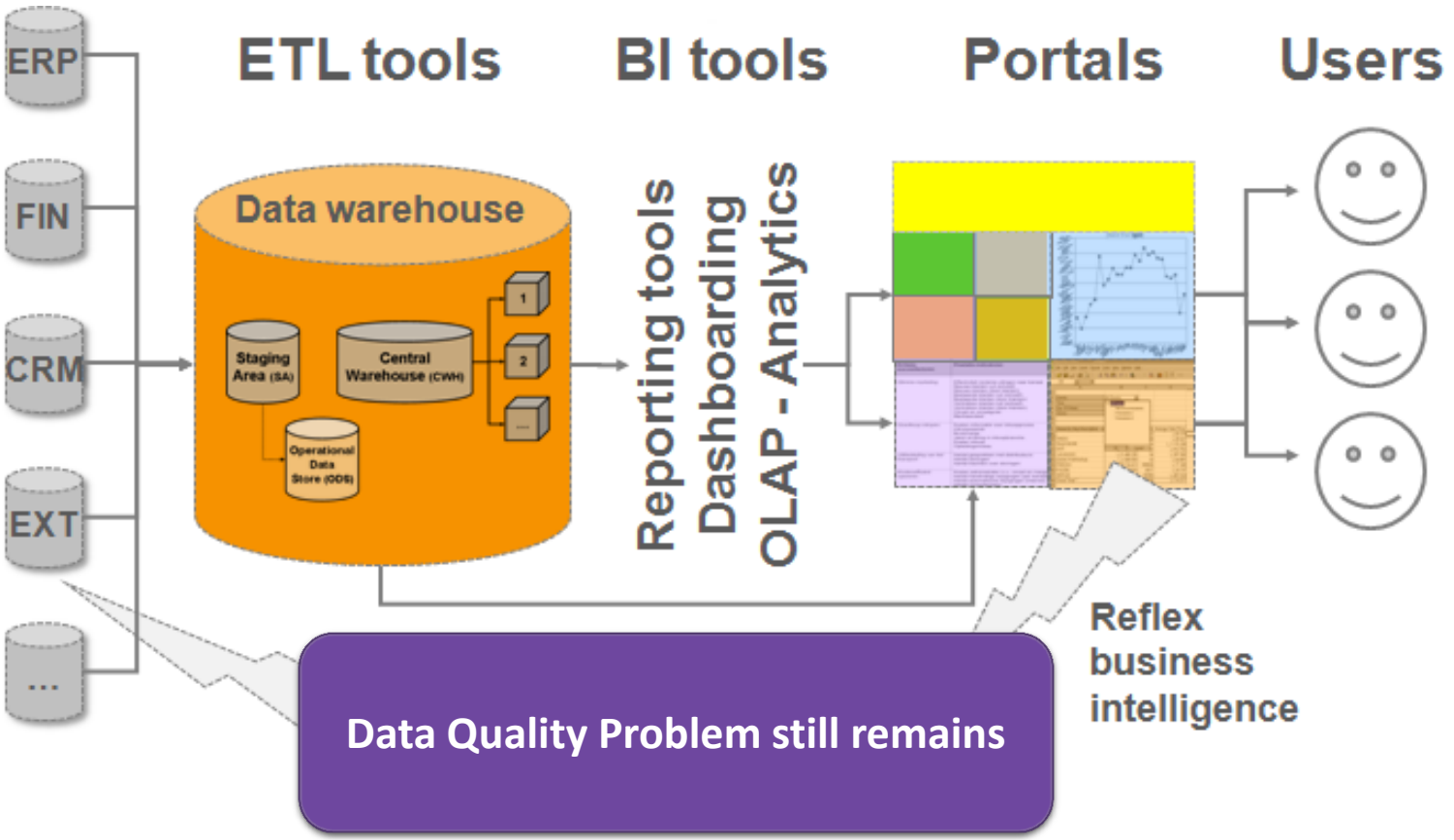


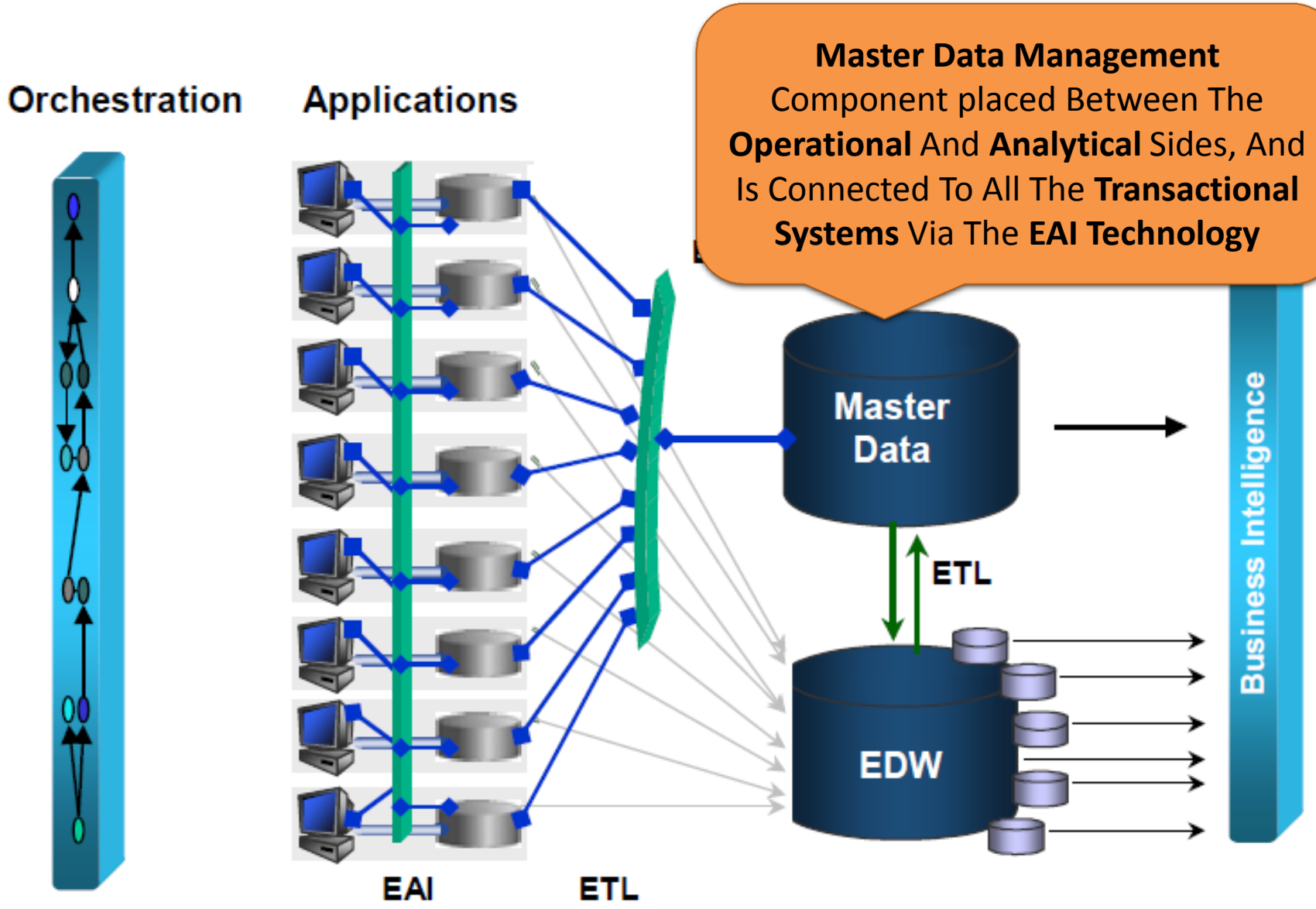
Enterprise Data Warehousing (EDW) and Data Marts

- Carries transaction history from operational applications including key dimensions such as
  - Customer
  - Product
  - Asset
  - Supplier
  - Location

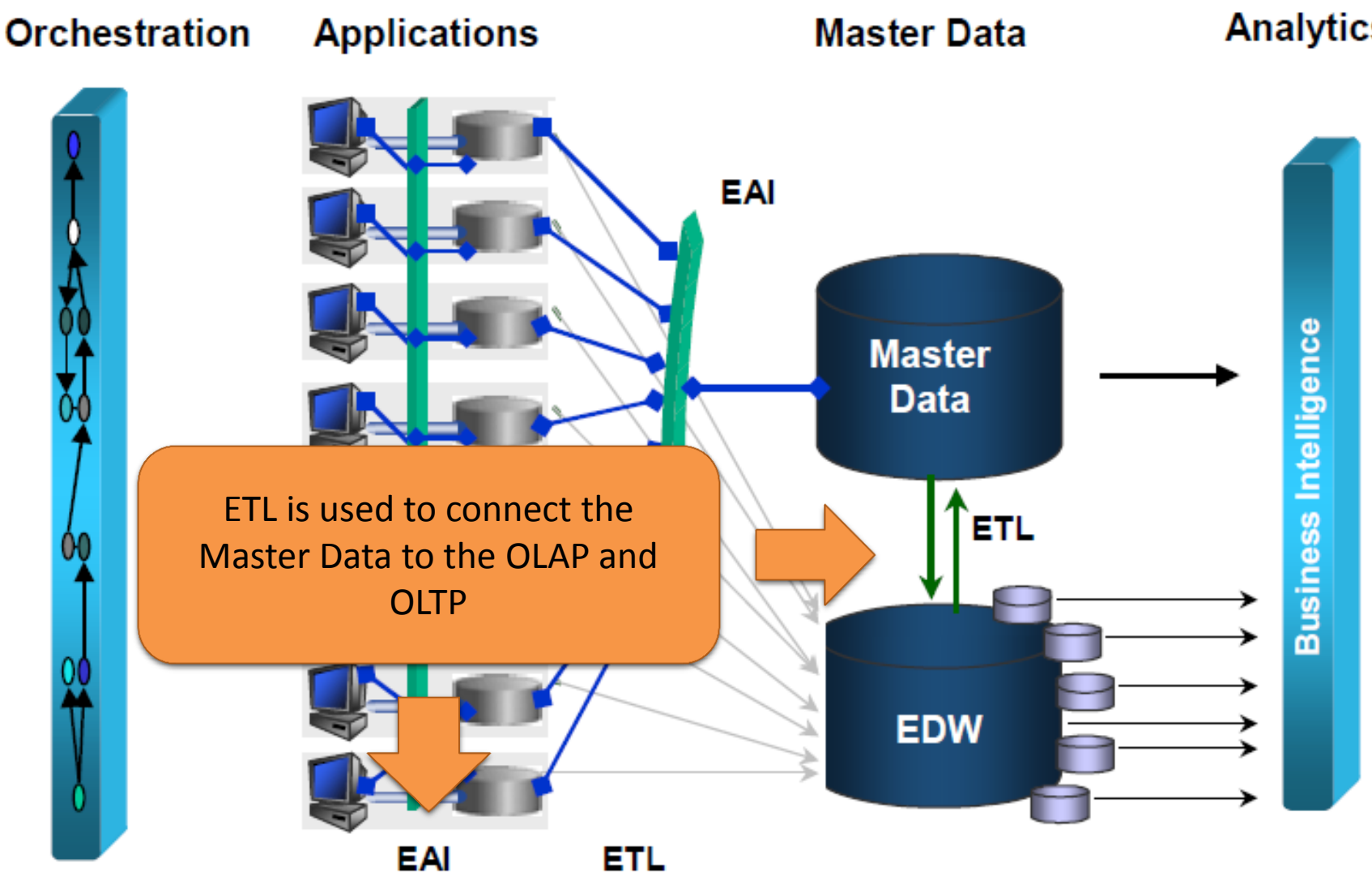


Business Intelligence (BI)

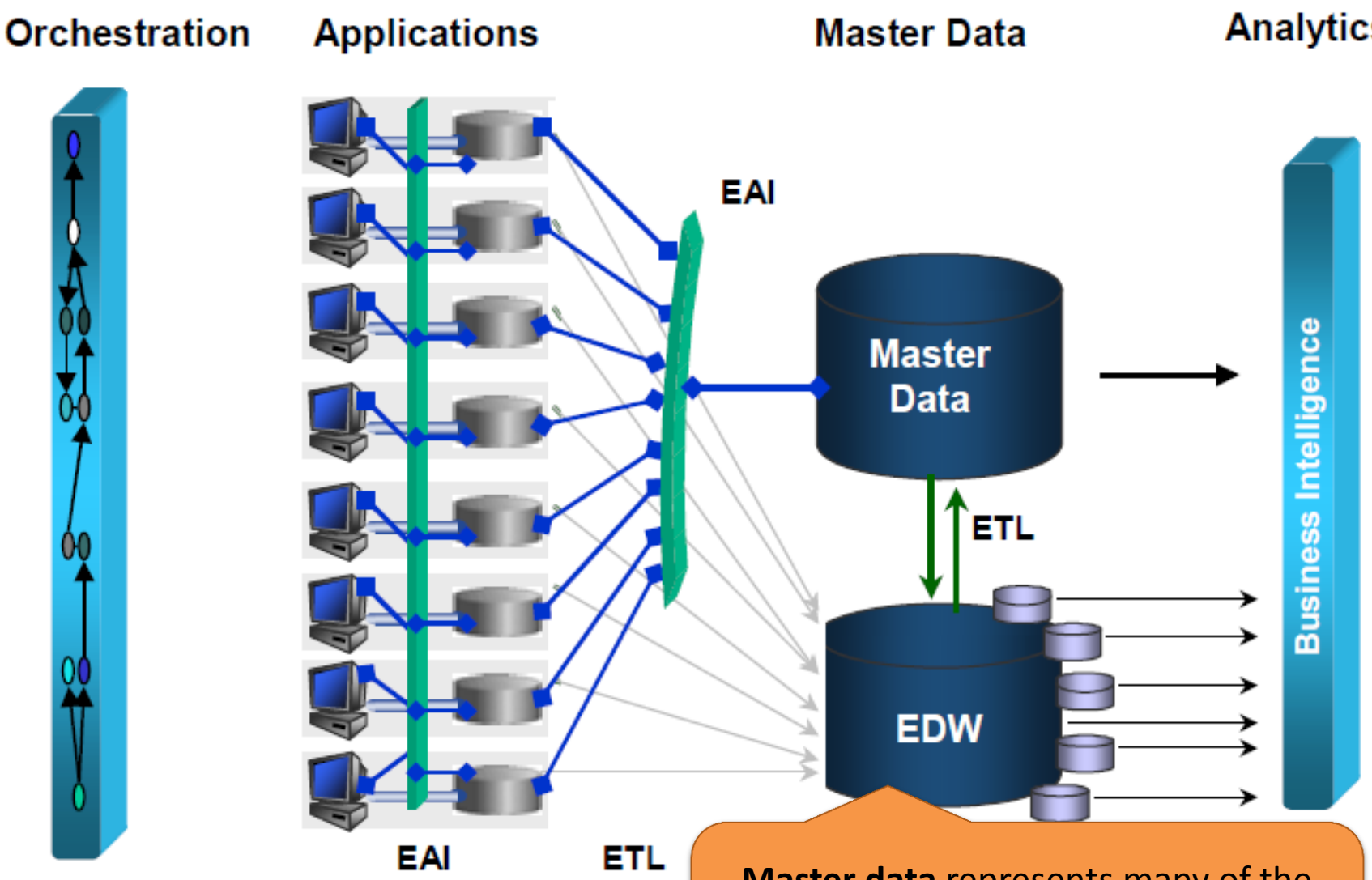




4.3. Ideal Information Architecture



4.3. Ideal Information Architecture



4.3. Ideal Inform

Master data represents many of the **Major Dimensions** supported in the **Data Warehouse**

References

I. David Butler, Bob Stackowiak, "*Master Data Management, An Oracle White Paper",* June 2009

II. Ralph Kimball, Margy Ross, "*The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence*", Wiley Publications, March 2010

III. Sabir Asadullaev ,*Learning course Data warehouse architectures and development strategy*, Open Group