

ETL Basics

- Lesson 1: Basic Concepts

Lesson Objectives

- On completion of this lesson on ETL basics, you will be able to:
 - Understand Data warehousing strategies and architecture
 - Know the meaning and need of ETL

Datawarehouse

- A single, complete and consistent store of data obtained from a variety of different sources made available to end users in a format that they can understand and use in a business context.

Datawarehousing Strategies

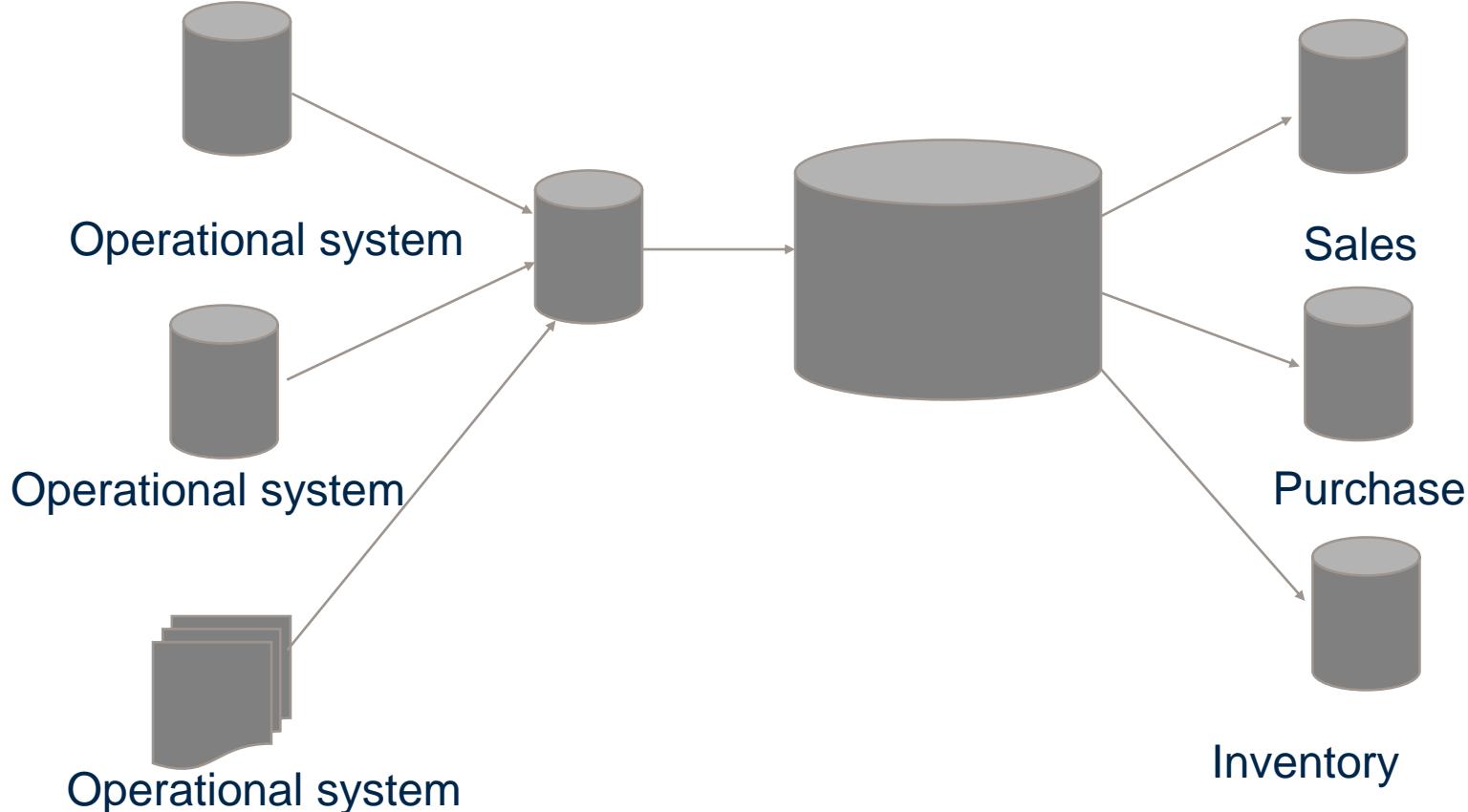
- Enterprise-wide warehouse, top down, the Inmon methodology
- Data mart, bottom up, the Kimball methodology
- When properly executed, both result in an enterprise-wide data warehouse

Inmon methodology - Top Down approach

- Bill Inmon saw a need to transfer data from diverse OLTP systems into a centralized place where the data could be used for analysis
- Inmon's philosophy recommends to start with building a large centralized enterprise-wide data warehouse, followed by several data-marts

Top Down Approach

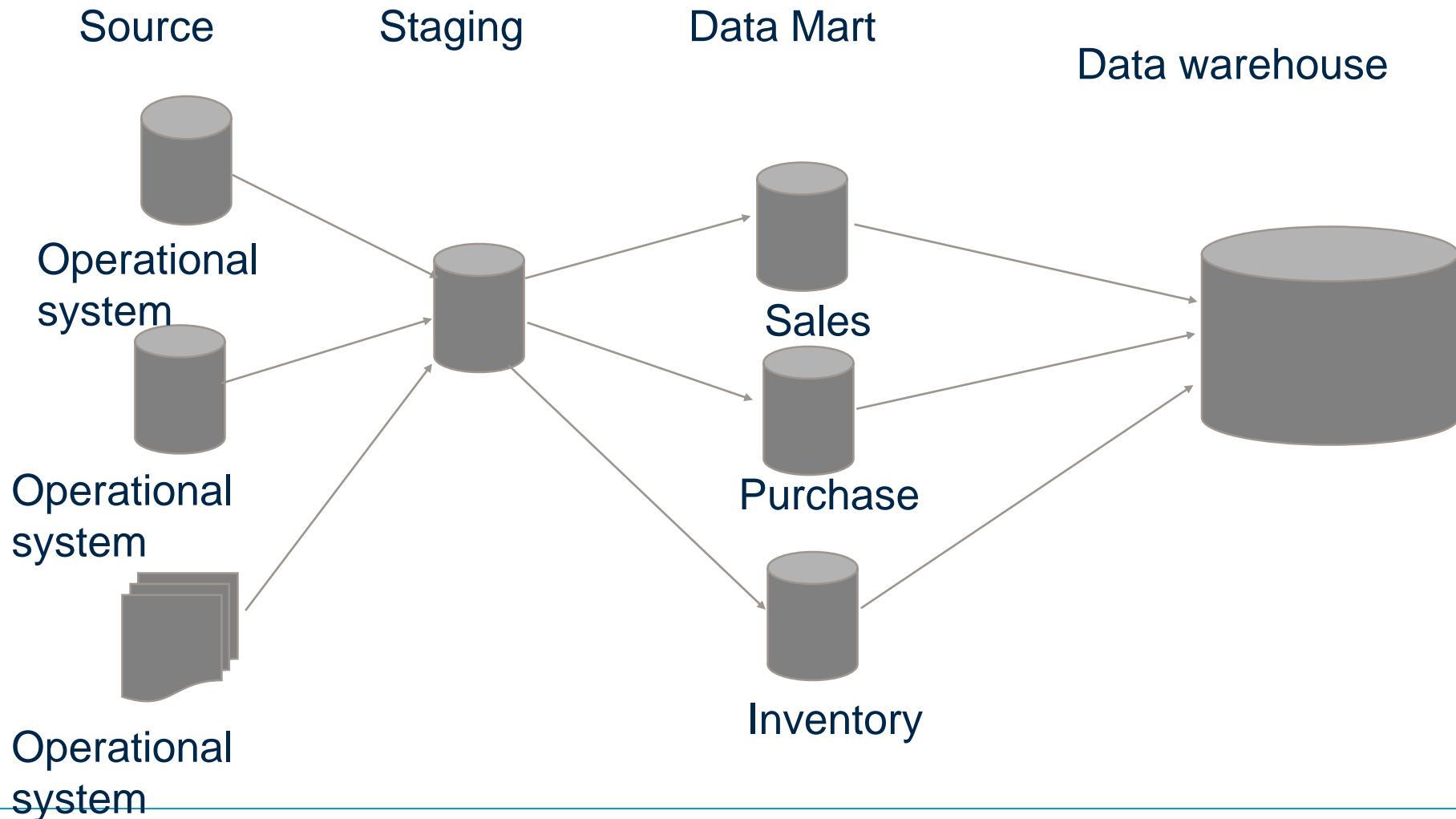
- Data Sources Staging Area Warehouse Data mart



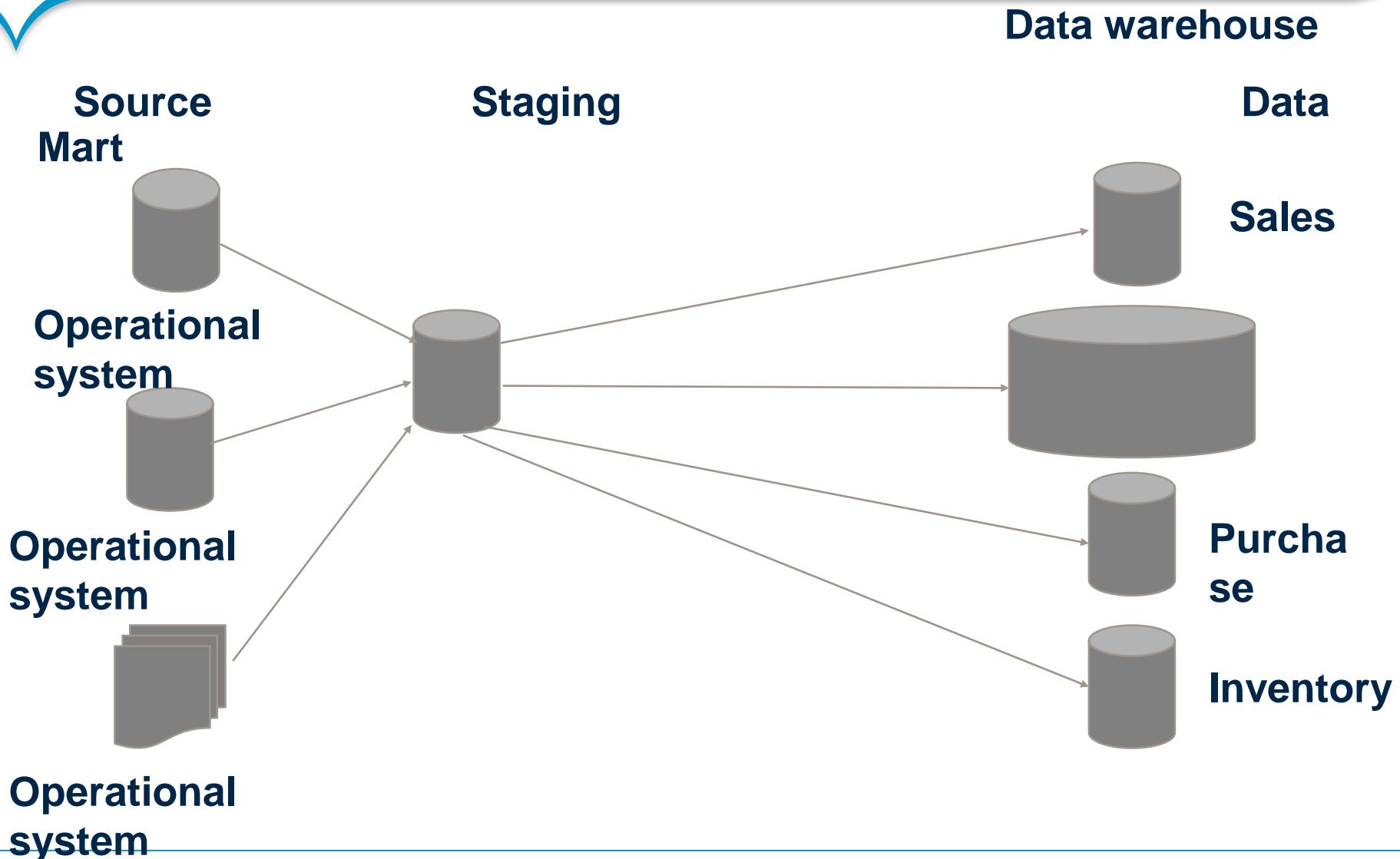
Kimball methodology – Bottom Up approach

- Kimball's philosophy recommends to start with building several data marts that serve the analytical needs of departments, followed by "virtually" integrating these data marts.

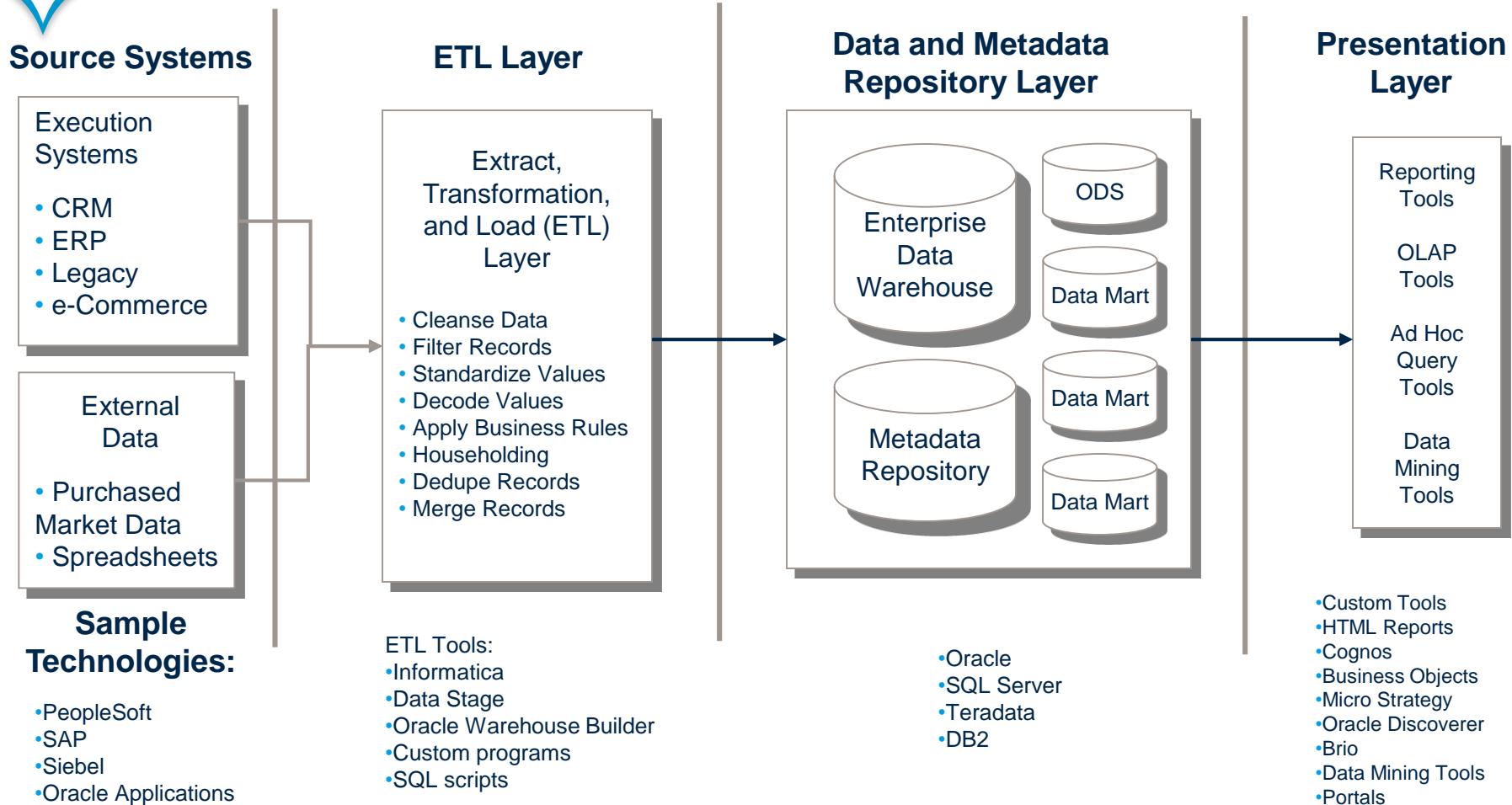
Bottom up Approach



Hybrid Approach



Data Warehouse Architecture Components



What is ETL?

- ETL stands for Extract Transform & Load
- The process of updating the data warehouse
- ETL is the automated and auditable data acquisition process from source system that involves one or more sub processes of data extraction, data transportation, data transformation, data consolidation, data integration, data loading and data cleaning.

Need for ETL

- The process of ETL is required so that data from different heterogeneous sources can be combined and brought into one common source.
- The Advantage of having the process of ETL is that, as data from different sources can be brought together, highly complex and user friendly reports can be generated for decision making

Need for ETL

- Data stored in different formats in different types of databases
- Some data sources might be archives while others may be active operational systems
- Data extraction and cleansing - time-consuming and difficult
Aggregation of data

Summary

- In this module, you learned about the following:
 - Datawarehousing strategies
 - Datawarehousing architecture
 - Need for ETL
 - Meaning of ETL

