

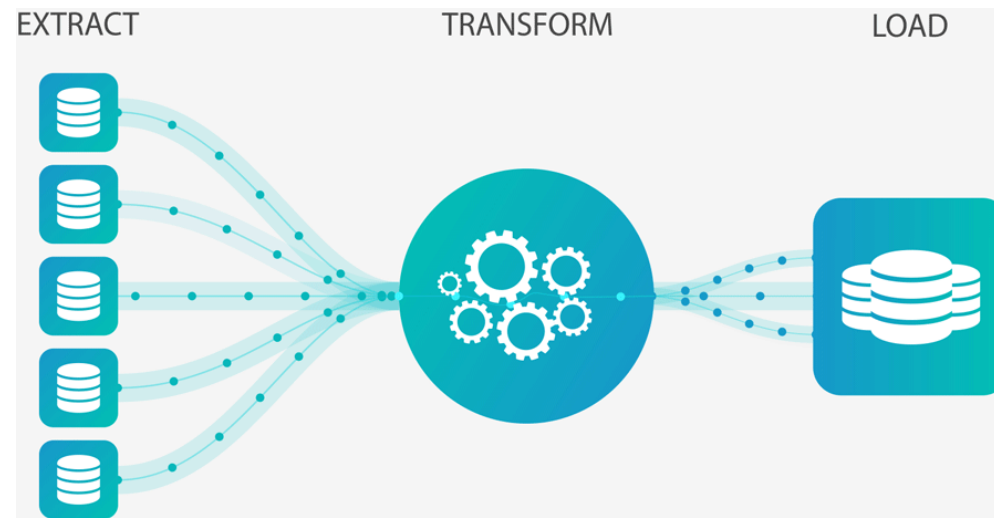


Start-Tech Academy

ETL

What

1. **ETL** stands for Extract, Transform and Load
2. **Extract** is reading data from data sources
3. **Transform** is processing the data
4. **Load** is writing the data to the destination



Data Warehouse

What

1. A **data warehouse** collates data from a wide range of sources within an organization.
2. **NOT** operational database.
3. **NOT** updated frequently
4. For the purpose of **analytics**



Data Warehouse

Differences with OLTP

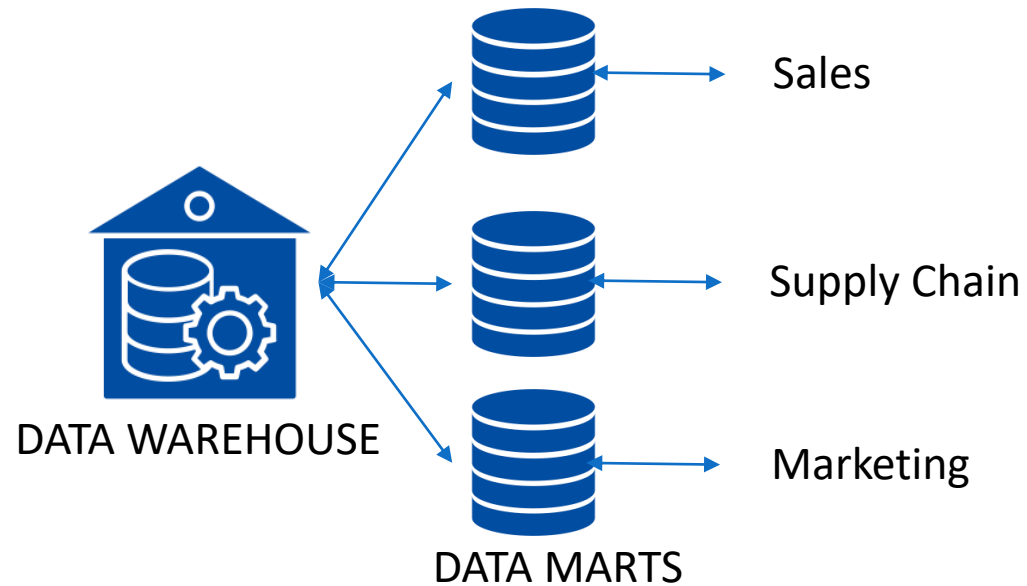
OLTP stands for OnLine Transaction Processing. Used for adding/ updating one/ few rows of data at a time

CHARACTERISTIC	OLTP	DATA WAREHOUSE
System scope/view	Single business process	Multiple business subjects
Data sources	One	Many
Data model	Static	Dynamic
Dominant query type	Insert/update	Read
Data volume per transaction	Small	Big
Data volume	Small/medium	Large
Data currency	Current timestamp	Seconds to days old
Bulk load/insert/update	No	Yes
Full history available	No	Yes
Response times	< 1 second	< 10 seconds
System availability	24/7	8/5
Typical user	Front office	Staff
Number of users	Large	Small/medium

Data Warehouse

Differences with DataMart

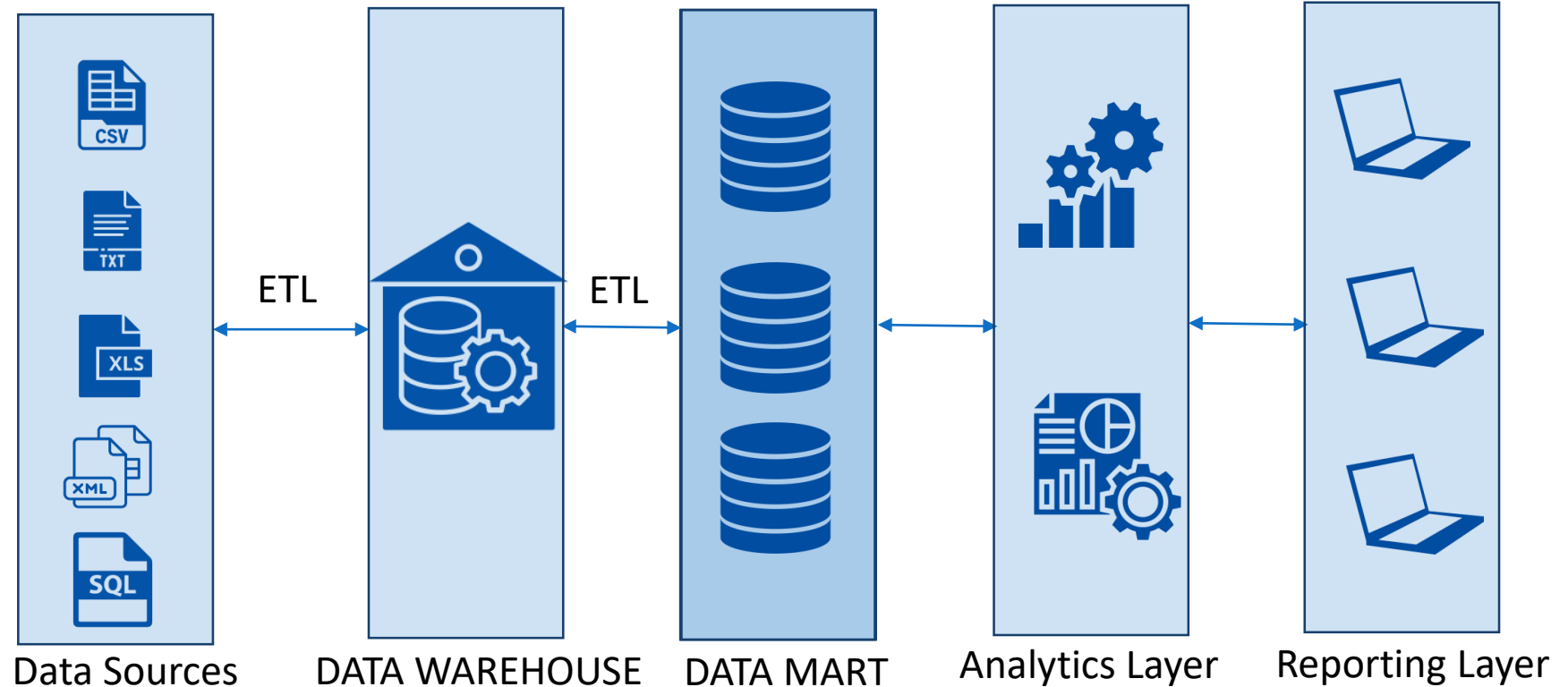
Data mart contain repositories of summarized data collected for analysis on a specific unit within an organization, for example, the sales, finance, operations, marketing department.



Data Warehouse

Inmon vs Kimball

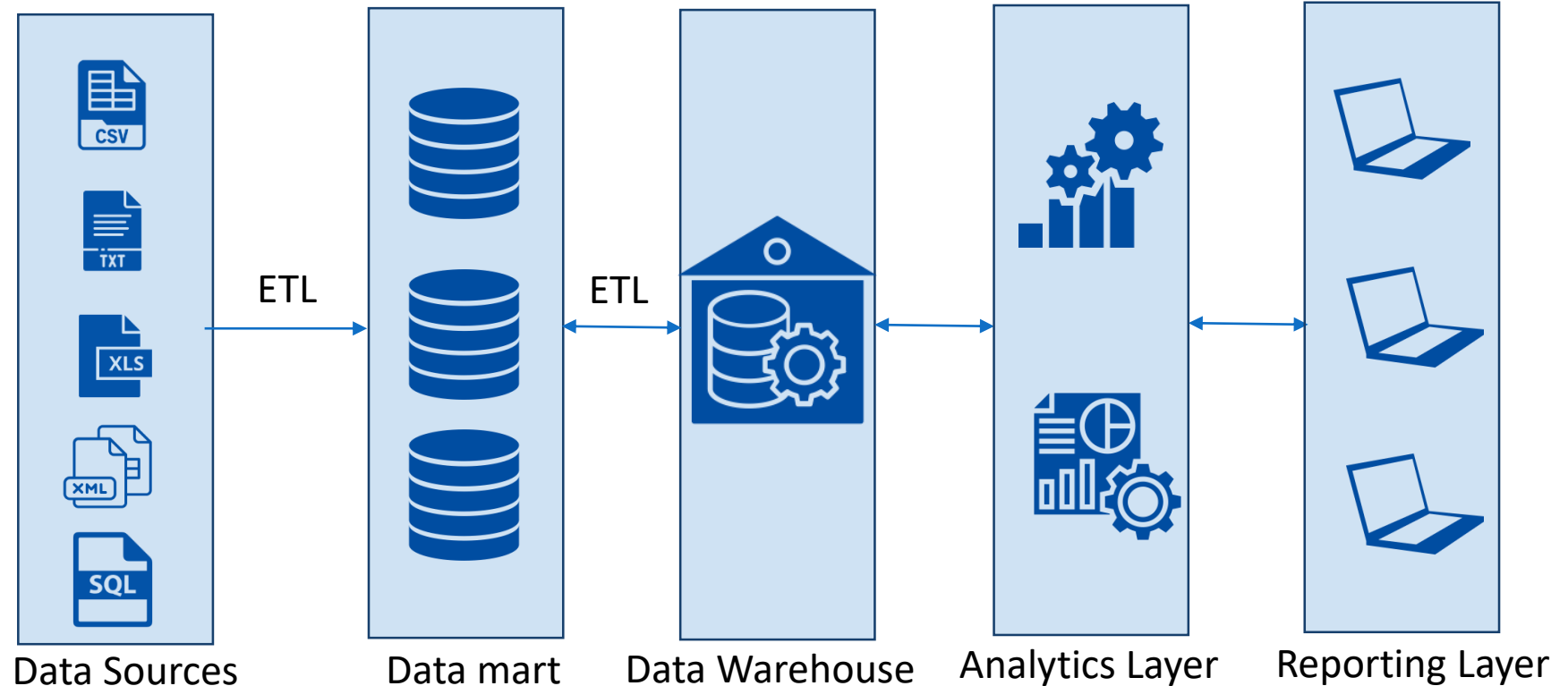
Inmon model: Organizations create DW, Data marts are created from DW, Analytics is applied on Data marts



Data Warehouse

Inmon vs Kimball

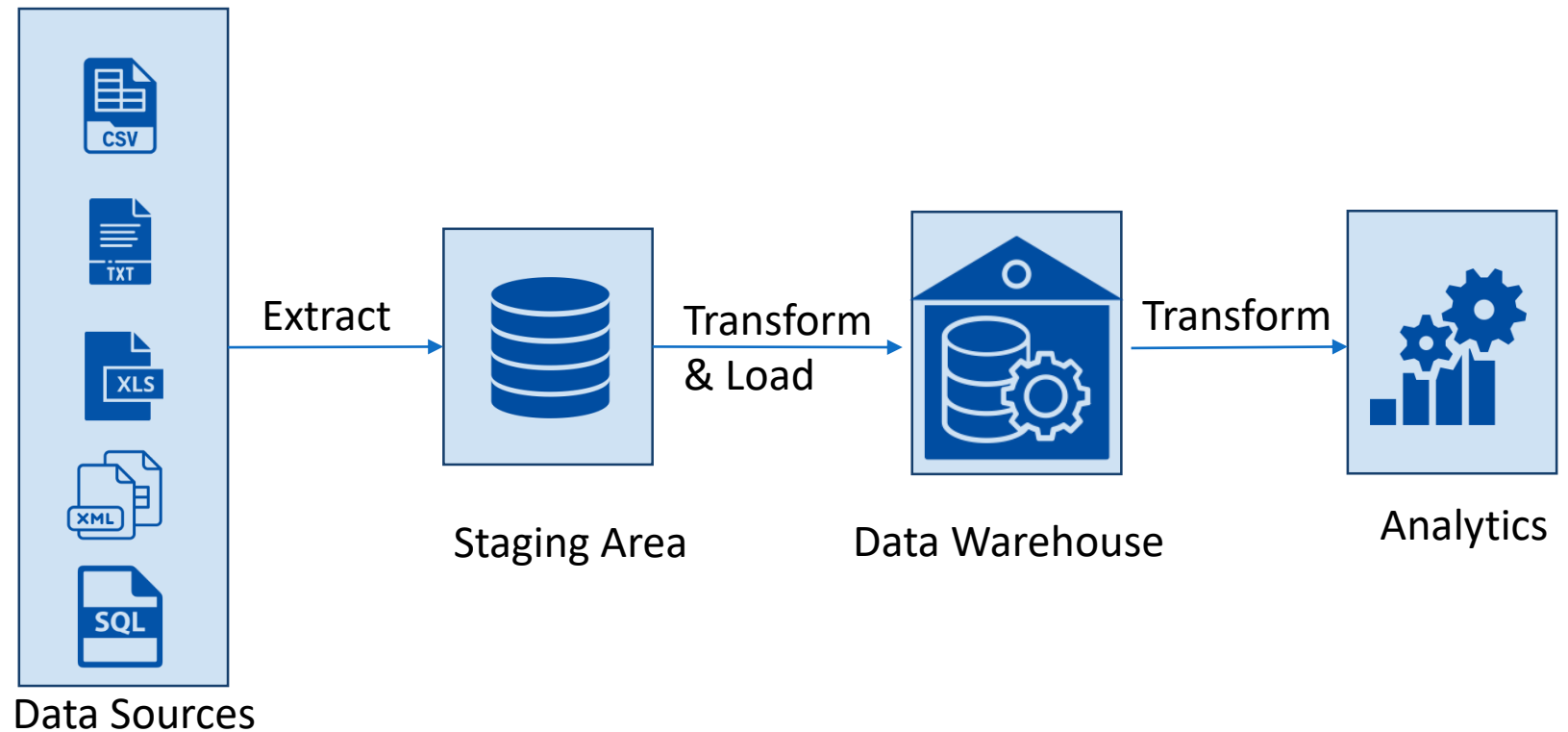
Kimball model: Organizations create Data marts, DW are created from data marts, Analytics is applied on DW



Data Warehouse

ETL vs ELT

ETL – Extract, Transform, Load



Data Warehouse

ETL vs ELT

ELT – Extract, Load, Transform

