Cloud Computing for Beginners

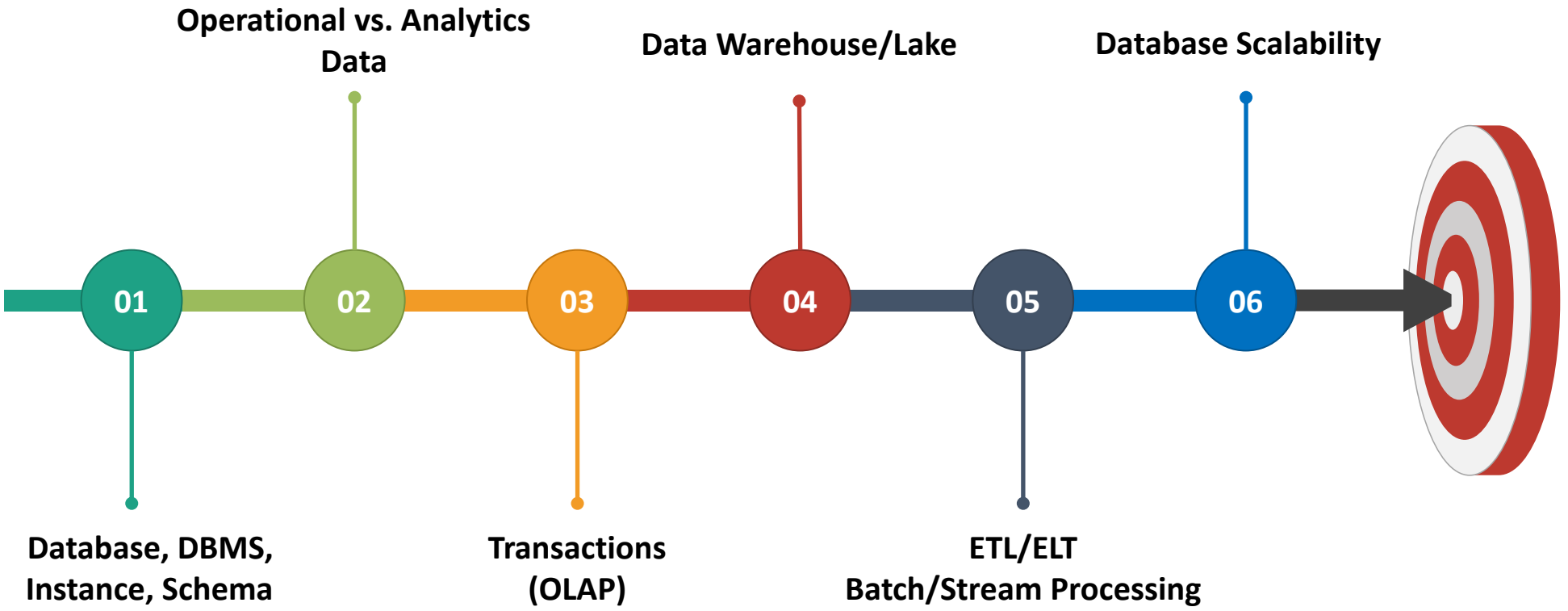
Database Technologies

By Idan Gabrieli

Basic Database Terminology

DBMS, Schema, OLTP/OLAP, ETL/ELT….

Basic Database Terminology



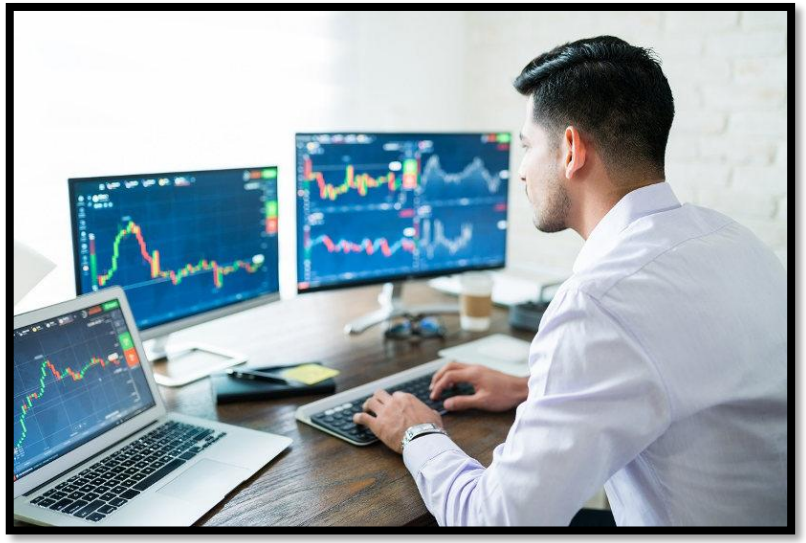
DBMS, Database Instance and Schema

What is a Database?

A database is an organized collection of information that is stored and accessed by applications.

DBMS, Database Instance and Schema

What is a Database?

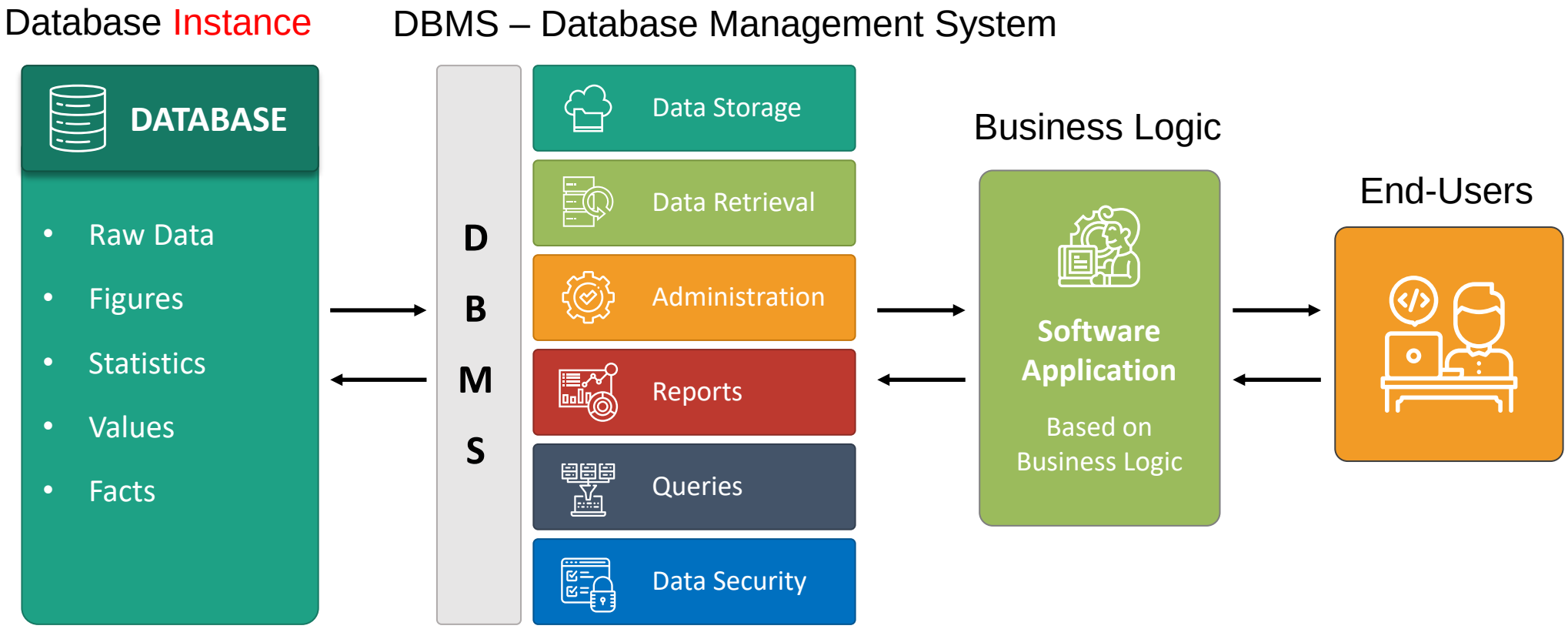


End-users profiles

Historical prices of stocks

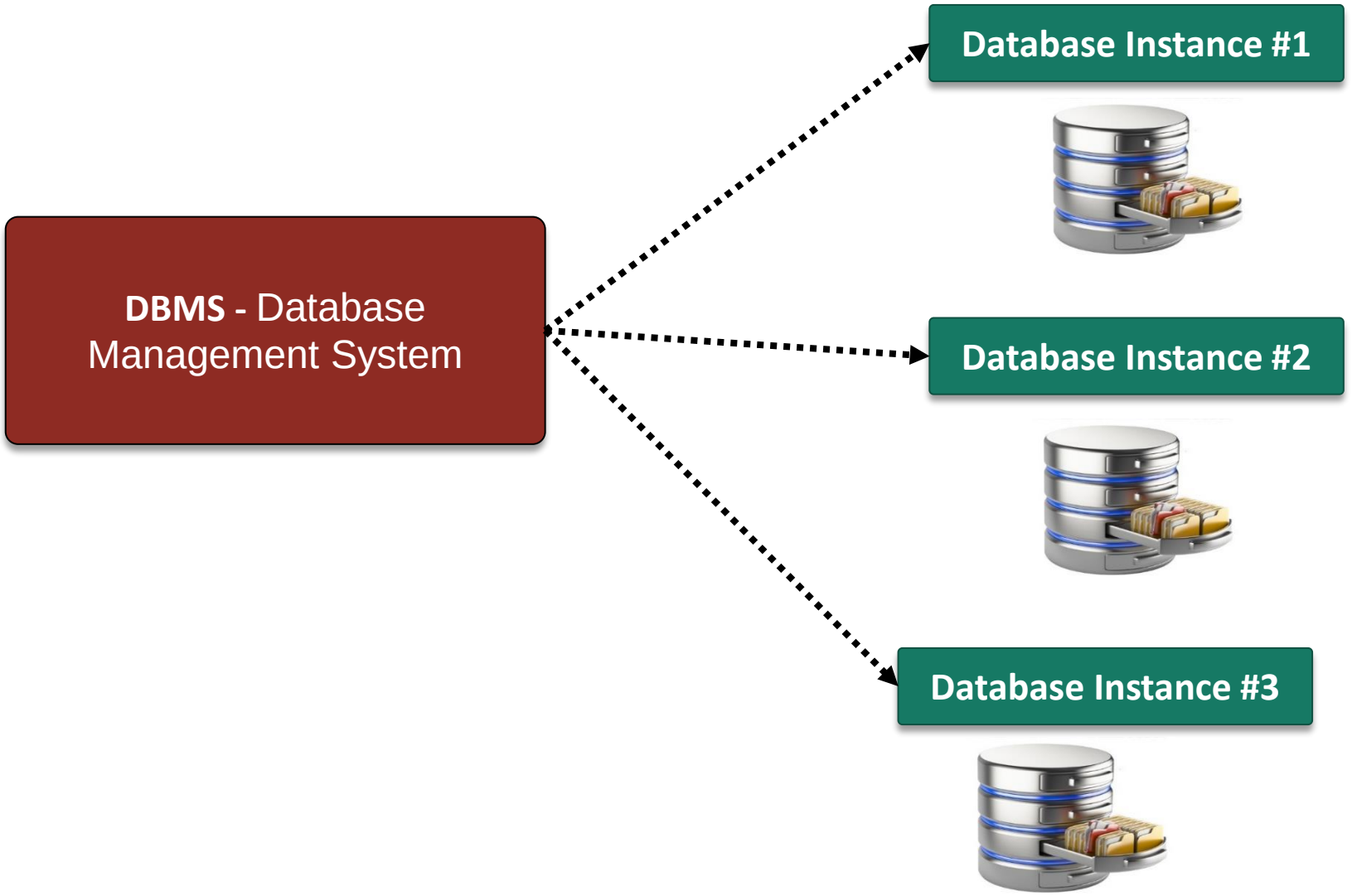
DBMS, Database Instance and Schema

What is a Database?



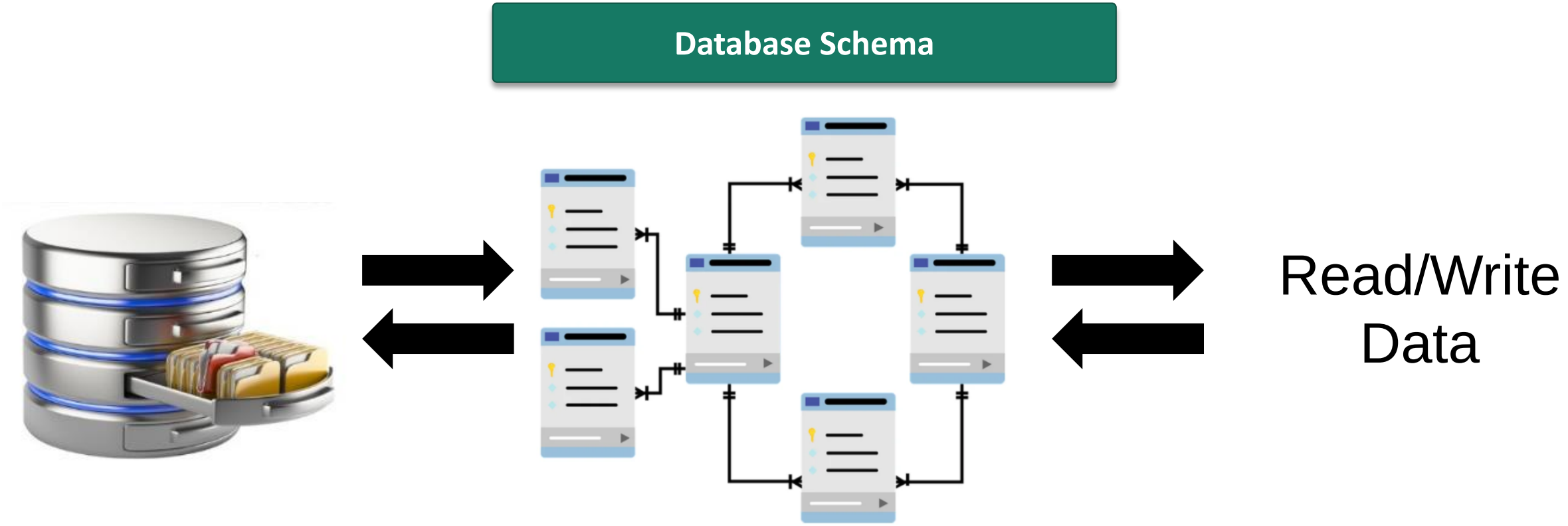
DBMS, Database Instance and Schema

What is a Database?



DBMS, Database Instance and Schema

What is a Database?



Operational Data and OLTP

Operational
Data

Analytics
Data

Operational Data and OLTP

- **Operational data** is the data that is produced by the organization's day-to-day operations
  - Understand what's going on with the organization?
    - Orders
    - Inventory
    - Customers/Suppliers
    - Products/Services
    - Trouble Tickets
    - Sales
  - **Up-to-date, real-time** information
  - Operational activities are recorded in databases using **transactions**

Operational Data and OLTP

- **Transactions**
  - A database **transaction** is a unit of work performed within the database
  - Stored in two-dimensional structure (columns and rows)
  - Transactions are typically handled by systems called **OLTP**
- **OLTP** - Online Transactional Processing
  - E.g. ERP, CRM, Payments….
  - OLTP application → OLTP database
  - Databases that can support **transactions**
  - ACID Model (Atomic, Consistent, Isolation, and Durable)
    - Transactions
    - Moving from state X to state Y

Operational Data and OLTP

ATM

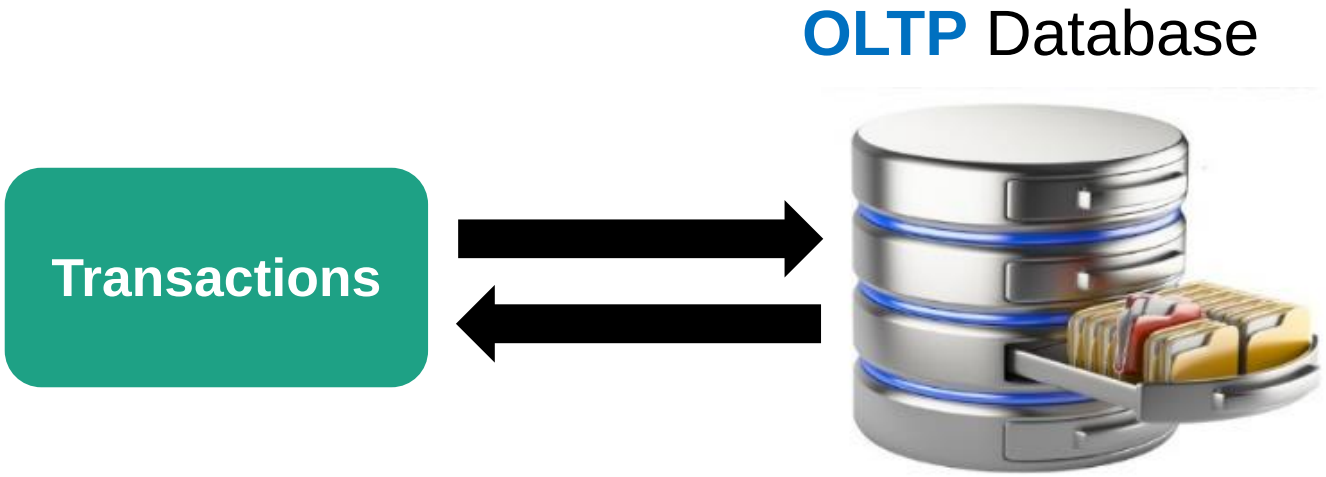
Withdraw Money

Transaction Confirmed/Not Confirmed

OLTP Database

Operational Data and OLTP

- **Transactions in OLTP Systems**
  - Operational activities are recorded in databases using **transactions**
  - **High-volume** on the database
  - **Low-latency** - fast access to the stored data
  - Data can be **changed frequently** with updated transactions

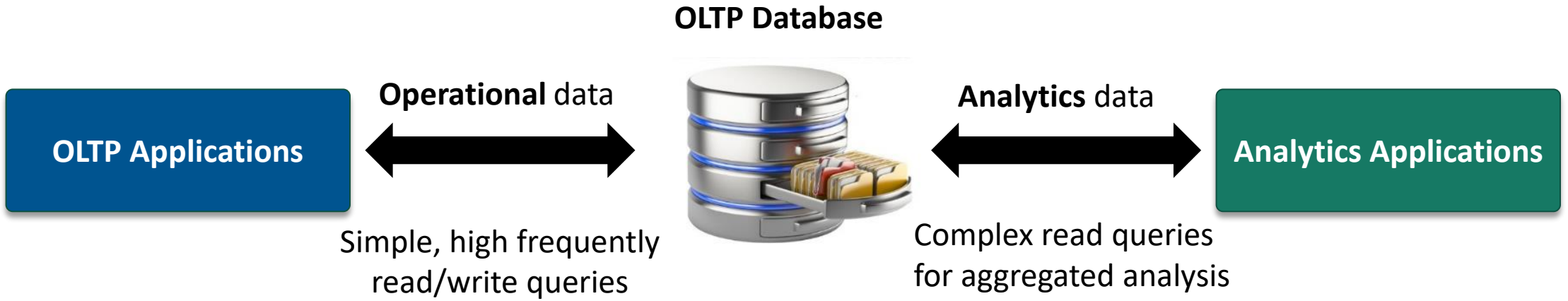


Analytics Data and OLAP

- **Analytics Data**
  - Make ongoing business decisions by **querying historical data**
  - Input for Business Intelligence (BI)
  - High-level **aggregated** analysis
    - Tune the business
    - Focus on emerging new opportunities
    - Identify trends in the market
    - Detect anomalies

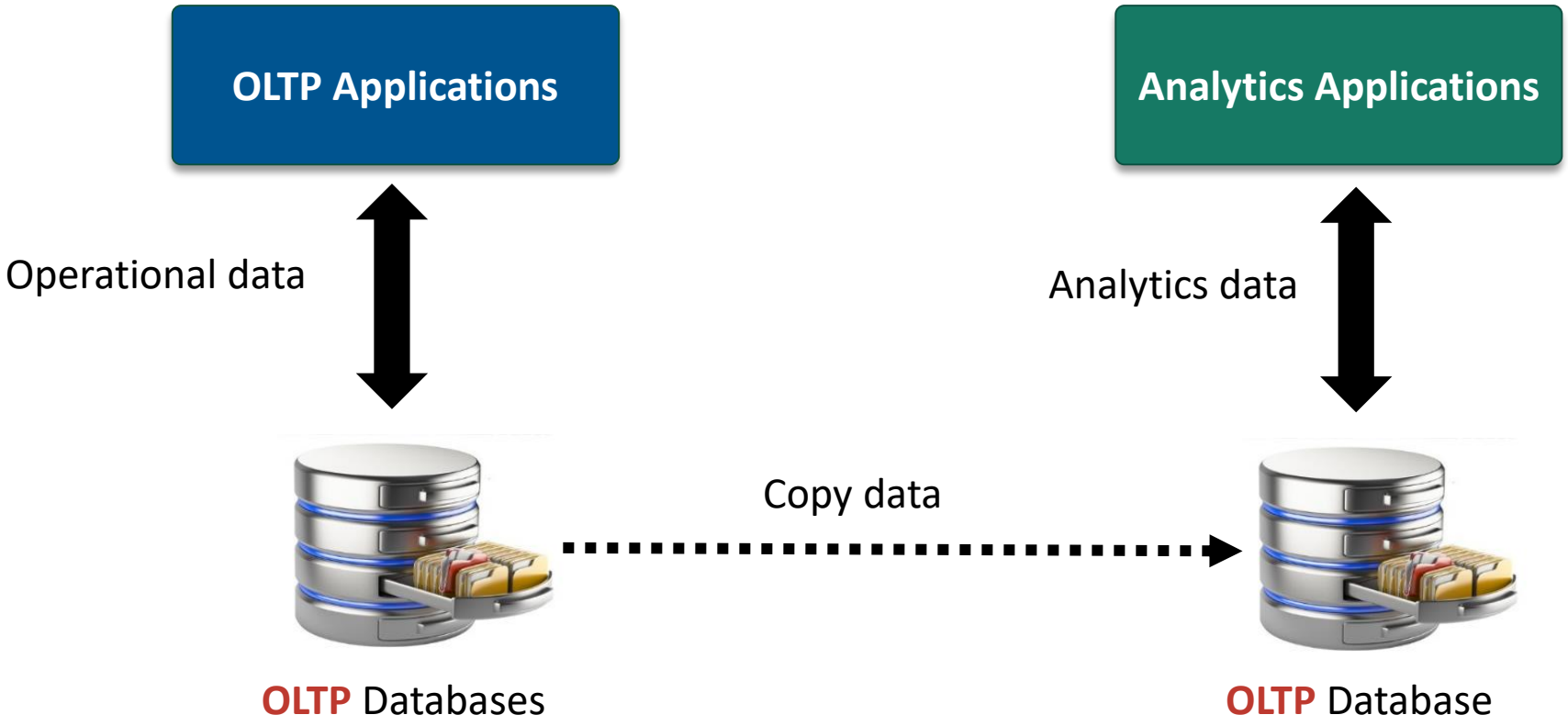
Analytics Data and OLAP

Option #1 – The Same OLTP Database



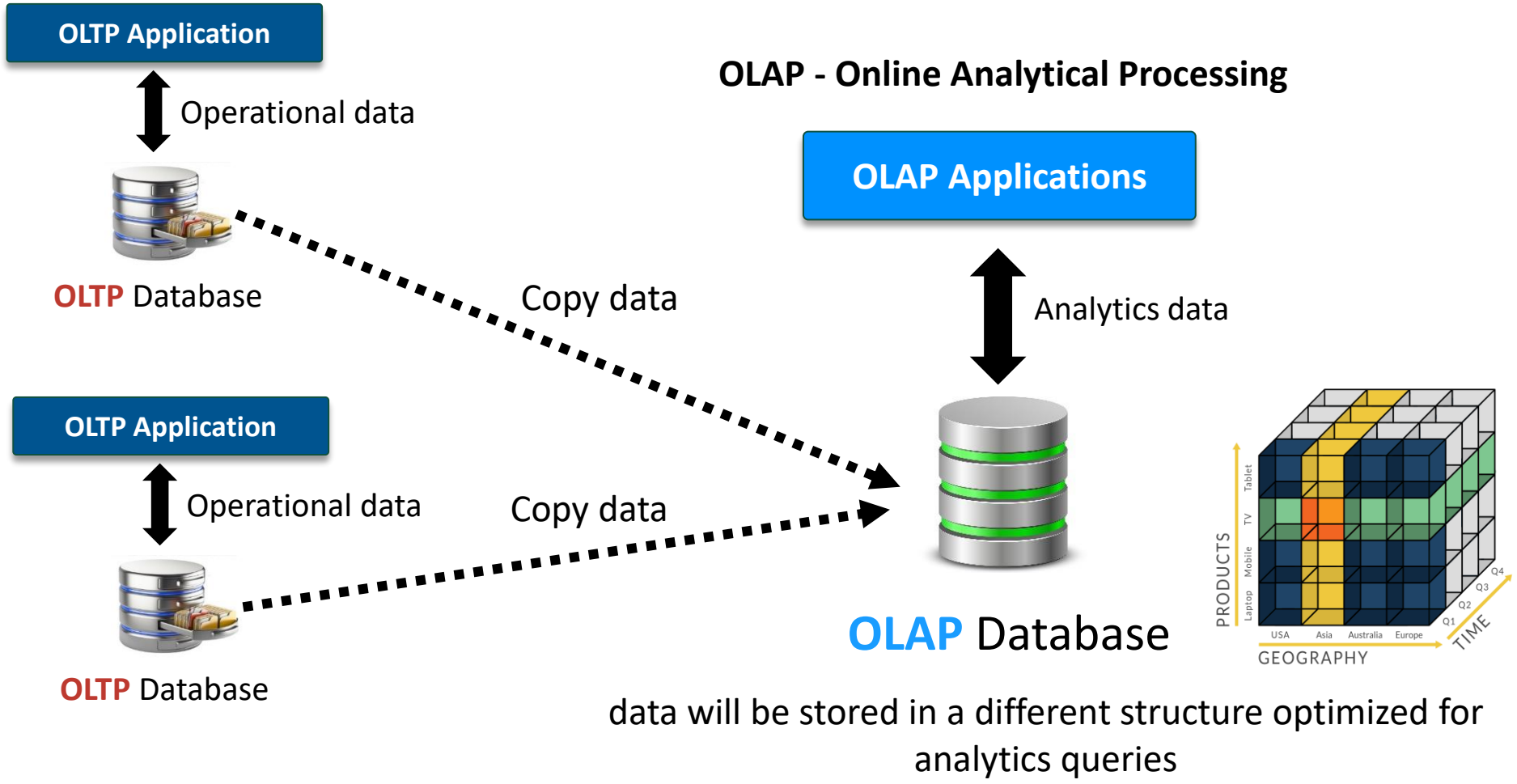
Analytics Data and OLAP

Option #2 – Dedicated OLTP Database



Analytics Data and OLAP

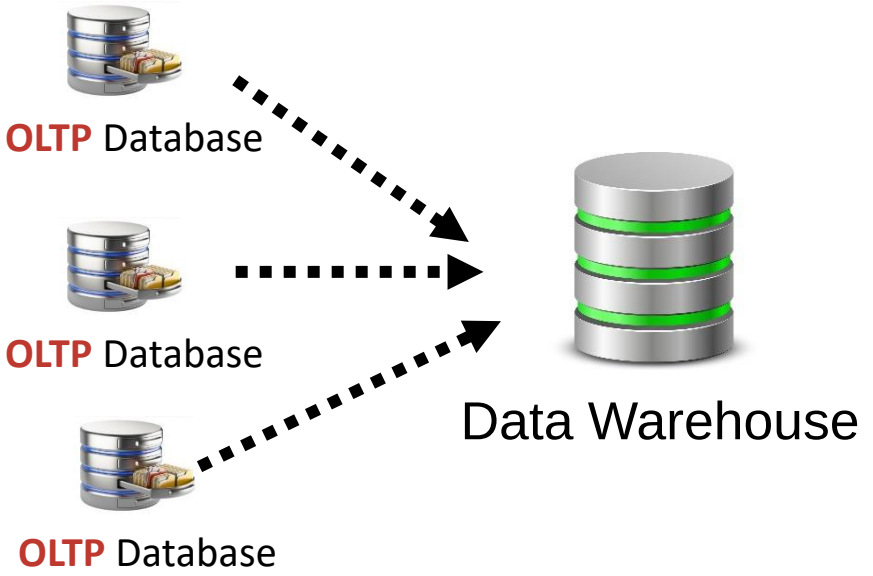
Option #3 – OLAP Database

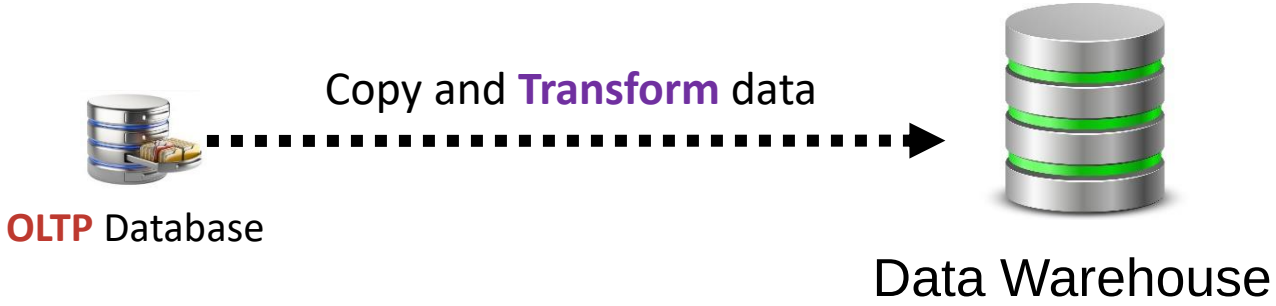


data will be stored in a different structure optimized for analytics queries

Data Warehouse and Data Lake

- **Data Warehouse**
  - Consolidate in one single place data from multiple data sources
  - Ingest large amounts of data
  - Optimized for OLAP use cases (complex ad-hoc read-queries)
  - Store historical data for a long retention time
  - Terabytes and even petabytes
  - Raw data is going to be transformed to a new structure





Data Warehouse and Data Lake

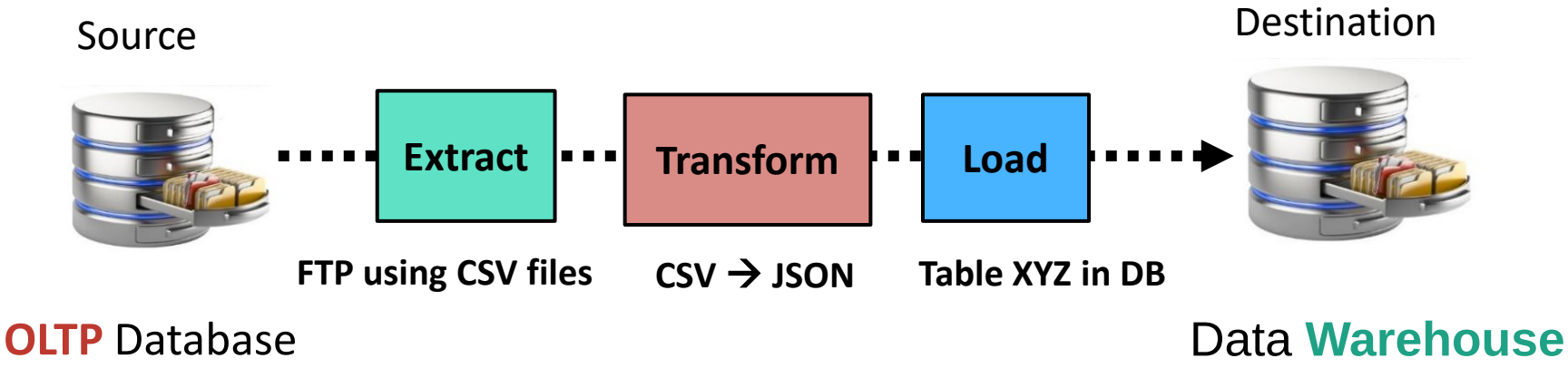
- **Data Lake**
  - Centralized repository for storing all structured and unstructured data
  - Keep the collected data **as-is without changing the data**
    - Text, Images, log files, social media, IoT, Video files
  - Used for different analytics use cases
    - E.g. data science – ML\AI

- **Typical Organizations**
  - **Data Warehouse + Data Lake**
  - Used for different use cases

ETL and ELT

Moving Data between Systems

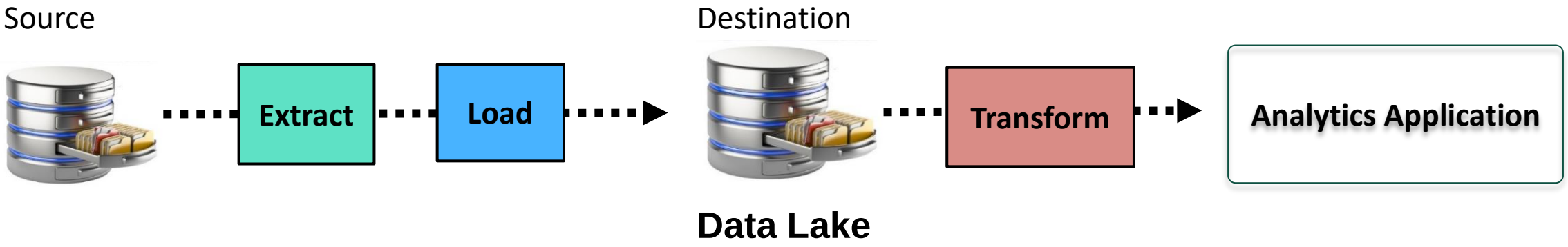
- **ETL - Extract, Transform, and Load**
  - The process of **extracting** data from one system called a data source, **transform** the data into a new structure, and then **load** the data into a destination system
  - Data is transformed **before** being loaded



ETL and ELT

Moving Data between Systems

- **ELT - Extract, Load and Transform**
  - **Extract** the raw data from the data source and **load** it into the database **without** transformation
  - Data will be stored in its original raw format
  - Data is transformed by applications **after** being loaded to the database



ETL and ELT

App X - Detect List of Objects in an Image

Raw images

ETL

Extract

Transform

Load

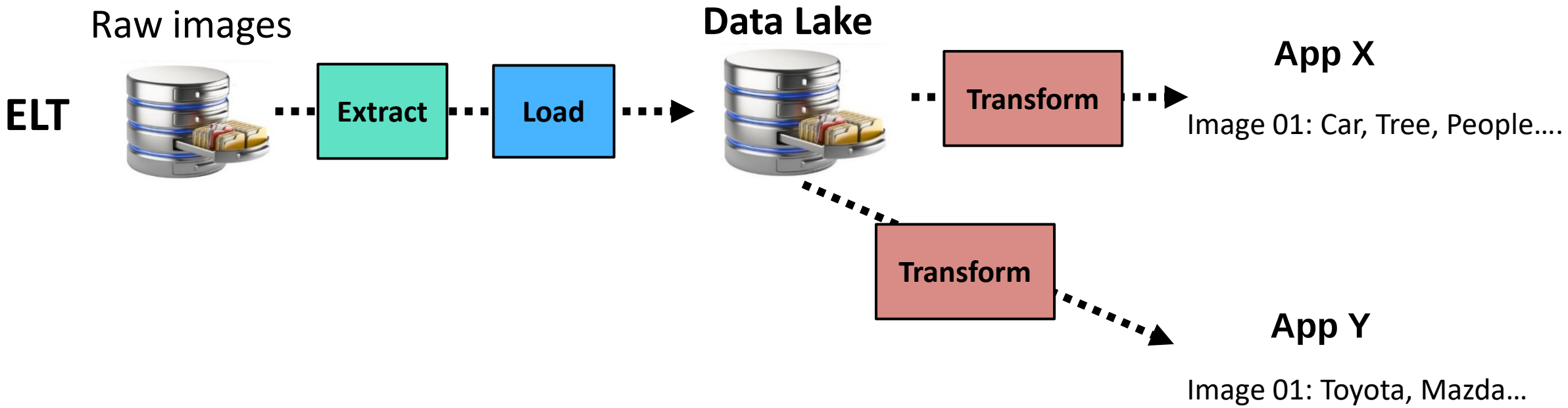
Image → list of objects

Image 01: “Car”, “Tree”, “People”….

Image 02: “Bicycle”, “Tree”…..

ETL and ELT

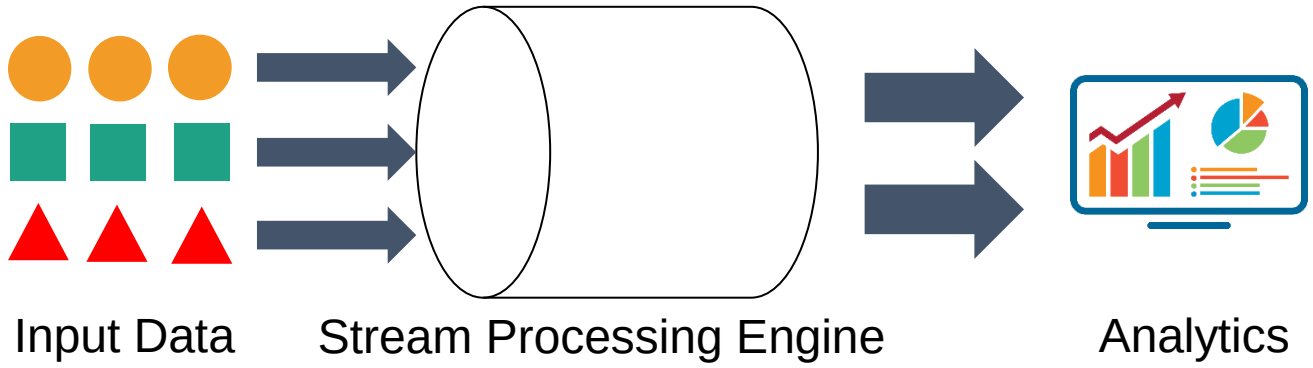
App X and App Y



Batch and Stream Processing

The frequency of moving data between systems

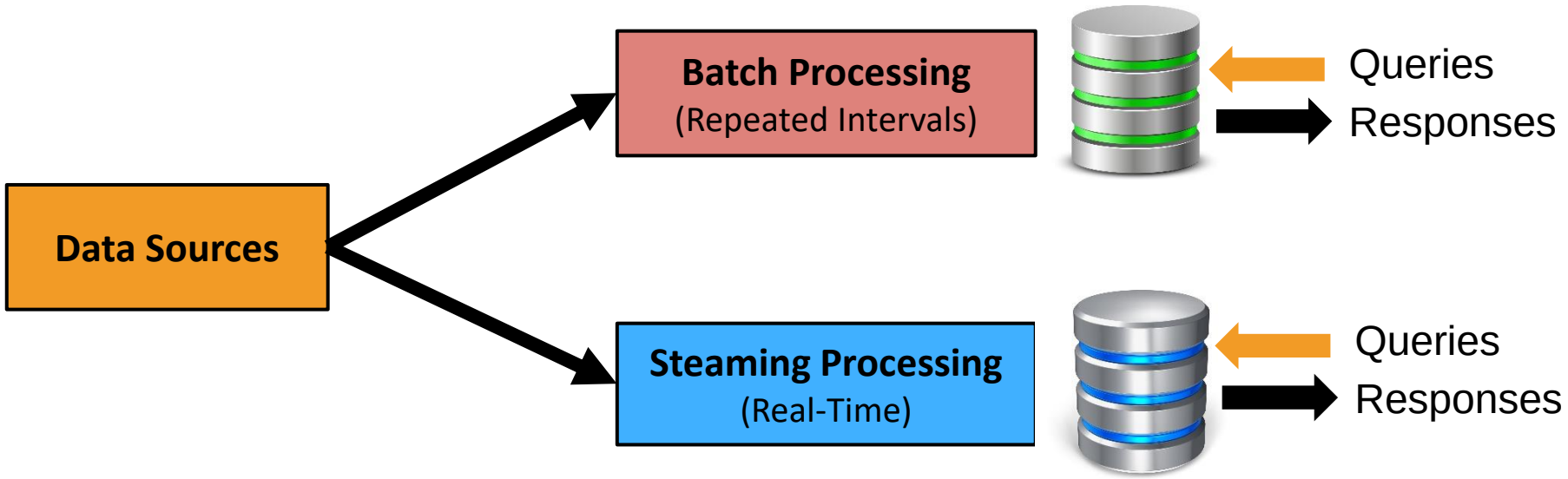
- **Batch processing**
  - Moving large volumes of data as a **chunk** to a target system
  - Typically, during off-peak times, in scheduled repetitive intervals
    - E.g. every 24 hours, at 02:00 AM
  - Useful when the target systems **don't require real-time data**
    - E.g. processing financial data in batches
- **Stream processing**
  - Taking action on a data immediately at the time it is created
  - Real time use cases with **streaming data sources**
  - A stream processing is designed to handle a **constant** stream of data



Batch and Stream Processing

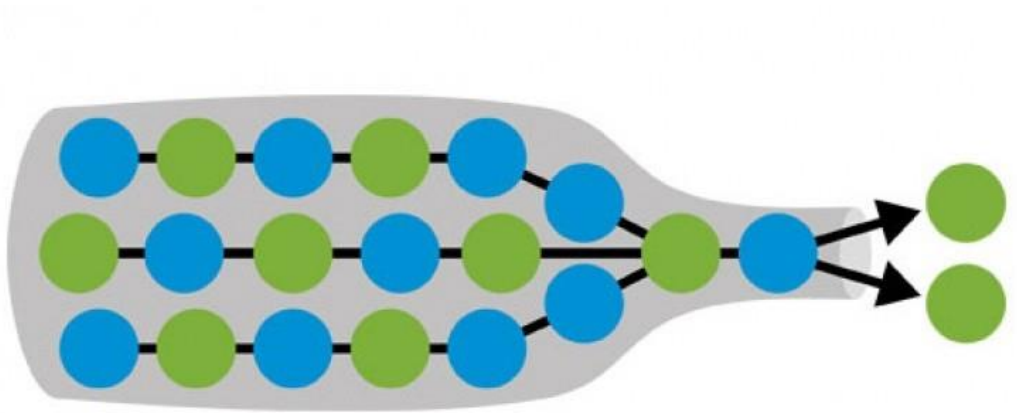
The frequency of moving data between systems

- **Lambda Architecture**
  - A data-processing architecture designed to handle massive quantities of data by taking advantage of both batch and stream-processing methods



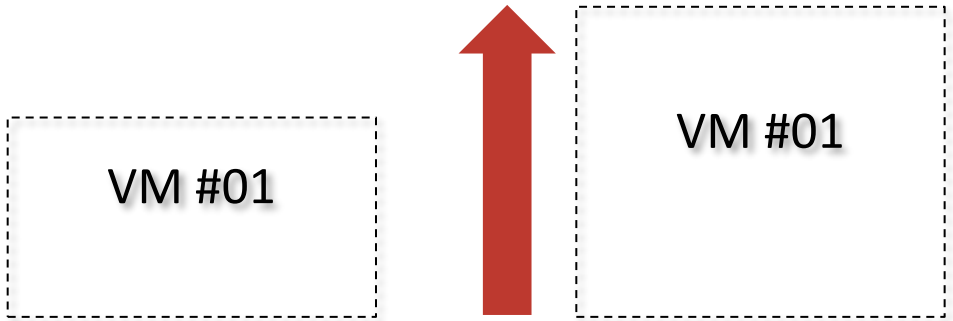
Scaling Up and Out

- **System Scalability**
  - Scaling is the process of managing the underline IT resources being used by the application to meet a set of performance requirements
    - IT Resources - CPU, Memory, Storage….
    - Over-utilized, Under-utilized
    - Ongoing process
- **Scaling Databases**
  - Databases are considered one of the **main bottleneck** of applications
  - Scaling a database system is important!
  - **Two options for scaling**
    - Vertical scaling
    - Horizontal scaling



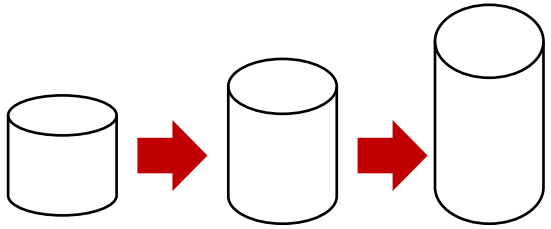
Scaling Up and Out

Vertical Scaling (up/down)



2 x vCPU
1 x GB RAM

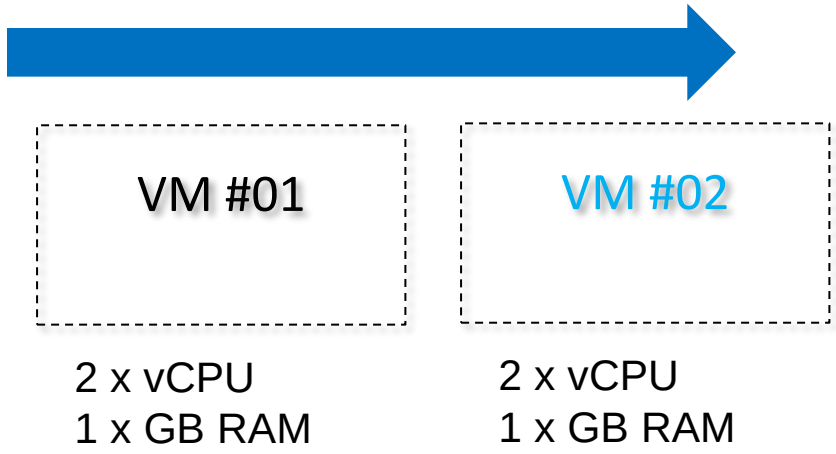
4 x vCPU
2 x GB RAM



2 x vCPU
1 x GB RAM

4 x vCPU
2 x GB RAM

Horizontal Scaling (out/in)



2 x vCPU
1 x GB RAM

2 x vCPU
1 x GB RAM