

MANAGED BY
RUBY THALIB AI & DATA CONSULTANT

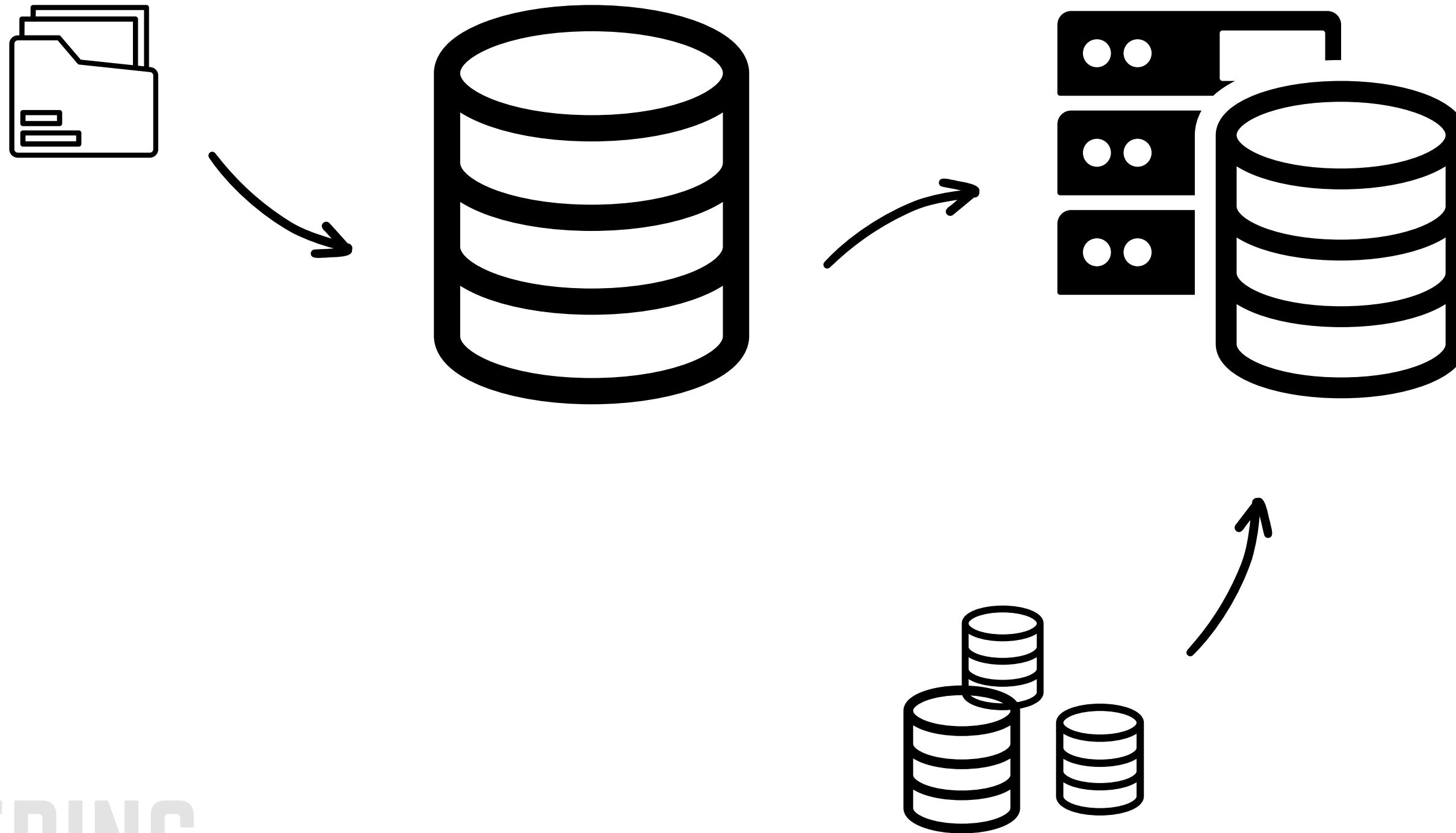
DATA ENGINEERING

MOST SCOPE

INTERACTIVE CLASS, 2024

- **Data Ingestion**
- **Database Introduction**

- **Data Ingestion**



MANAGED BY
RUBY THALIB AI & DATA CONSULTANT

**DATA
ENGINEERING**

MOST SCOPE

INTERACTIVE CLASS, 2024

- # Data Ingestion

Aspects

- Data sources



- Process estimation

- Target ingestion (destination)



- Data Ingestion – common data type

tabular source

	ID	Age	Height	Weight	Year
count	271116.000000	261642.000000	210945.000000	208241.000000	271116.000000
mean	68248.954396	25.556898	175.338970	70.702393	1978.378480
std	39022.286345	6.393561	10.518462	14.348020	29.877632
min	1.000000	10.000000	127.000000	25.000000	1896.000000
25%	34643.000000	21.000000	168.000000	60.000000	1960.000000
50%	68205.000000	24.000000	175.000000	70.000000	1988.000000
75%	102097.250000	28.000000	183.000000	79.000000	2002.000000
max	135571.000000	97.000000	226.000000	214.000000	2016.000000

json source

```
{
  "orders": [
    {
      "orderno": "748745375",
      "date": "June 30, 2088 1:54:23 AM",
      "trackingno": "TN0039291",
      "custid": "11045",
      "customer": {
        "custid": "11045",
        "fname": "Sue",
        "lname": "Hatfield",
        "address": "1409 Silver Street",
        "city": "Ashland",
        "state": "NE",
        "zip": "68003"
      }
    }
  ]
}
```

compressed source

MANAGED BY
RUBY THALIB AI & DATA CONSULTANT

DATA
ENGINEERING

MOST SCOPE

INTERACTIVE CLASS, 2024



- Database Introduction

Analytic

Transaction

Purpose	OLAP helps you analyze large volumes of data to support decision-making.	OLTP helps you manage and process real-time transactions.
Data source	OLAP uses historical and aggregated data from multiple sources.	OLTP uses real-time and transactional data from a single source.
Data structure	OLAP uses multidimensional (cubes) or relational databases.	OLTP uses relational databases.
Data model	OLAP uses star schema, snowflake schema, or other analytical models.	OLTP uses normalized or denormalized models.
Volume of data	OLAP has large storage requirements. Think terabytes (TB) and petabytes (PB).	OLTP has comparatively smaller storage requirements. Think gigabytes (GB).
Response time	OLAP has longer response times, typically in seconds or minutes.	OLTP has shorter response times, typically in milliseconds
Example applications	OLAP is good for analyzing trends, predicting customer behavior, and identifying profitability.	OLTP is good for processing payments, customer data management, and order processing.

MANAGED BY
RUBY THALIB AI & DATA CONSULTANT

DATA
ENGINEERING

MOST SCOPE

INTERACTIVE CLASS, 2024

- Database Introduction

Analytic

Transaction

Purpose	OLAP helps you analyze large volumes of data to support decision-making.	OLTP helps you manage and process real-time transactions.
Data source	OLAP uses historical and aggregated data from multiple sources.	OLTP uses real-time and transactional data from a single source.
Data structure	OLAP uses multidimensional (cubes) or relational databases.	OLTP uses relational databases.
Data model	OLAP uses star schema, snowflake schema, or other analytical models.	OLTP uses normalized or denormalized models.
Volume of data	OLAP has large storage requirements. Think terabytes (TB) and petabytes (PB).	OLTP has comparatively smaller storage requirements. Think gigabytes (GB).
Response time	OLAP has longer response times, typically in seconds or minutes.	OLTP has shorter response times, typically in milliseconds
Example applications	OLAP is good for analyzing trends, predicting customer behavior, and identifying profitability.	OLTP is good for processing payments, customer data management, and order processing.

MANAGED BY
RUBY THALIB AI & DATA CONSULTANT

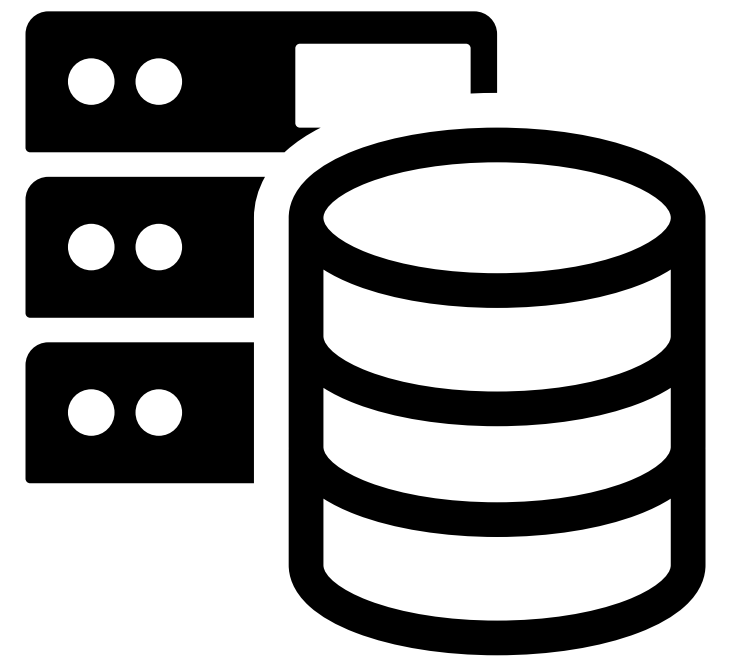
DATA
ENGINEERING

MOST SCOPE

INTERACTIVE CLASS, 2024

- **Database Introduction – Basic Concept**

- Schema
- Tables
- Data types (INTEGER, VARCHAR, TIMESTAMP, etc.)
- Constraints (NOT NULL, UNIQUE, PRIMARY KEY, FOREIGN KEY)



MANAGED BY
RUBY THALIB AI & DATA CONSULTANT

DATA
ENGINEERING

MOST SCOPE

INTERACTIVE CLASS, 2024

MANAGED BY
RUBY THALIB AI & DATA CONSULTANT

DATA ENGINEERING

MOST SCOPE

INTERACTIVE CLASS, 2024

DAY 2 TRIAL :D