OLTP vs OLAP

- These are two types of processing behind different database servers
  - OLTP (online transactional processing): optimal for processing transactional data
  - OLAP (online analytical processing): optimal for serving downstream analytics applications
- As they are serving different purposes, data in these database servers should have different data structures
- Raw data can be, for instances,
  - stored in OLTP: often highly normalized and 3NF(Third Normal Form) as the data model
  - extracted from API: often One Big Table (OBT)
- The process of creating the new data structure/blueprint of refined layers of data warehouse is called data modelling

Dimensional modeling

- Dimensional modeling is a common choice for analytics purpose
- Under dimension modeling, you can further choose: star schema and snowflake schema
- Dimension model is composed of fact and dimension tables
- Example: a company wants to analyse sales growth(metrics/KPI) by customer country of origin, product types, month of sales (dimensions)

- Advantage:
  - intuitive data structure,
  - efficient by choosing the relevant dimension table for a certain analysis purpose (ex. A team will only analyse sales by export destination)
- Star schema = one fact table + several dimension tables
- Snowflake schema = one fact table + several dimension tables + several 'sub-dimension' tables

Normalization

Raw data as One Big Table

Order no.	Date	Customer code	Customer name	Product code	Product name	Price	Quantity
1101	3-May	12	Company A	1	melon	800	1
1101	3-May	12	Company A	2	strawberry	150	3
1102	4-May	13			melon	900	
1103	4-May	14			melon	900	

Sales table (Fact)

Order no.	Date	Customer code	Product code	Quantity
1101	3-May	12		

Customer table (Dimension)

Customer code	Customer name
12	Company A
13	Company B

Product table (Dimension)

Product code	Product name	Price
1	melon	800
2	strawberry	150