

Database design for Data Analysis

⋮ Tags

Database

I. Dim table

- Dimension tables add more detail to fact table
- Dimension / attributes a bout a process of :
 - + Hold reference data
- The value are more likely to be duplicated (solve it by using SDC)

Table Name: Customer_Dim

| Keys | ColumnName |
|------|------------|
| PK | CustomerID |
| | Name |
| | AccountNum |
| | LoyaltyID |
| | Country |
| | Email |

Legend: PK = Primary Key

II. Fact table

- Measurements, metrics or fact about an organization
- Links to **dimension tables** for more details

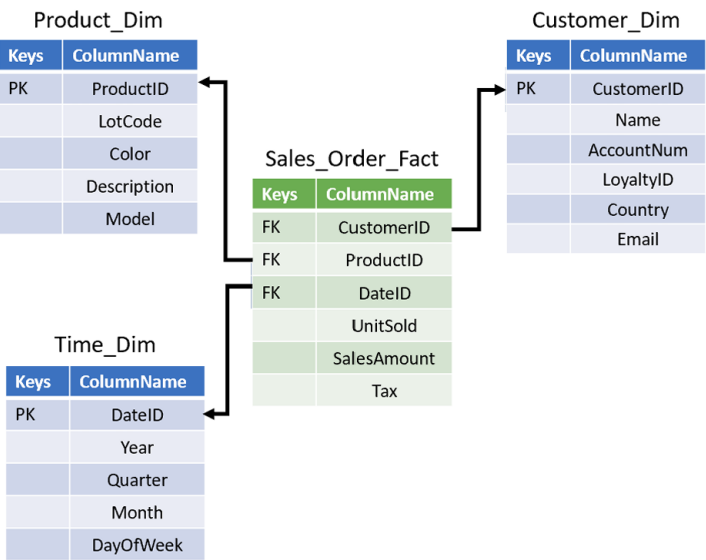
Table Name: Sales_Order_Fact

| Keys | ColumnName |
|------|-------------|
| FK | CustomerID |
| FK | DateID |
| FK | ProductID |
| | UnitSold |
| | SalesAmount |
| | Tax |

Legend: FK = Foreign Key

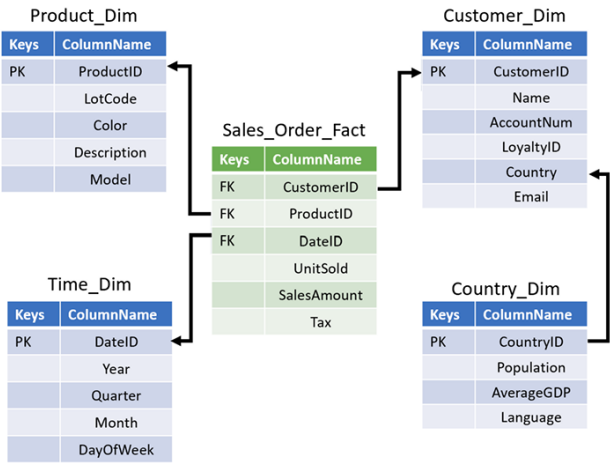
III. Star schema

- A central fact table, with one or more dimensional tables
- Easy for business users



IV. Snowflake schema

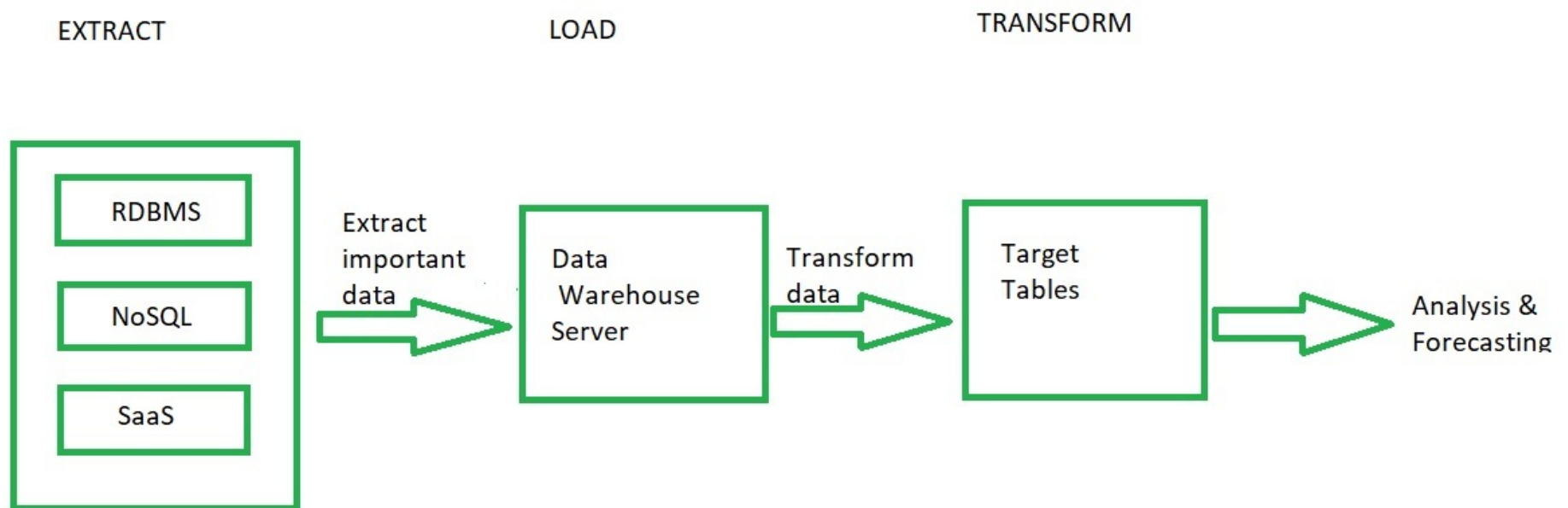
- Dimension table connected through another dimension table



V. ELT vs ETL

1. ELT (Extract - Load- Transform)

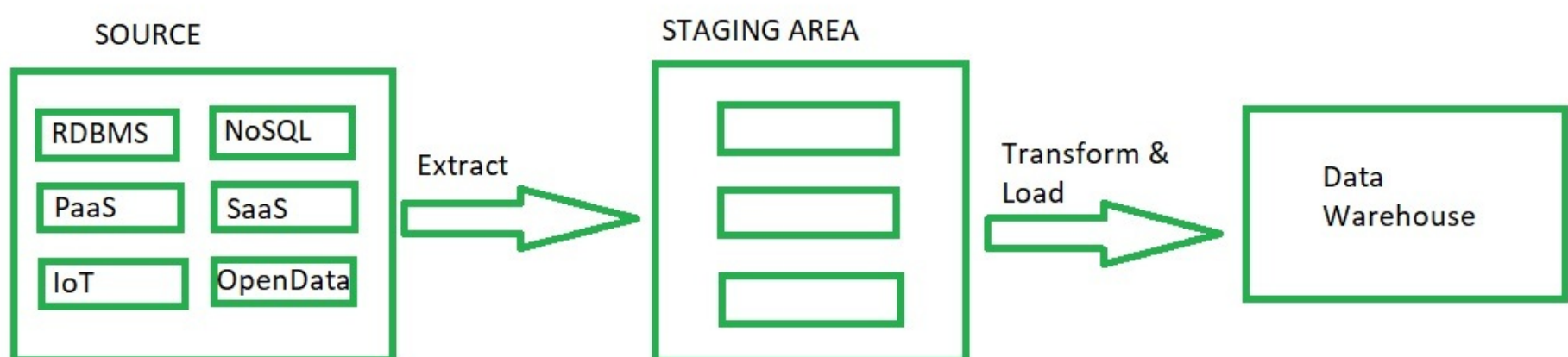
- **Extract** : the technique of identifying data from one or more sources. The sources may be database, files, ERP, CRM or others
- **Load** : the process of storing the extracted raw data in data warehouse or data lakes
- **Transform** : Data transformation is the process in which the raw data source is transformed to the



- **Drawback** : once transformed data is stored in the data warehouse, it can not be modified again, While in ELT, a copy of raw is always available and only the required data is transformed when needed

2. ETL (Extract - Transform - Load)

- **Extract** : the technique of identifying data from one or more sources. The sources may be database, files, ERP, CRM or others
- **Transform** : The extracted data is immediately transformed as required user
- **Load** : then it is loaded into the data warehouse from where user can access it



Note : the transformation are performed in the staging area. Once this transformed successfully, this will be stored in data warehouse