Business Intelligence and Data Warehousing (ANL408)

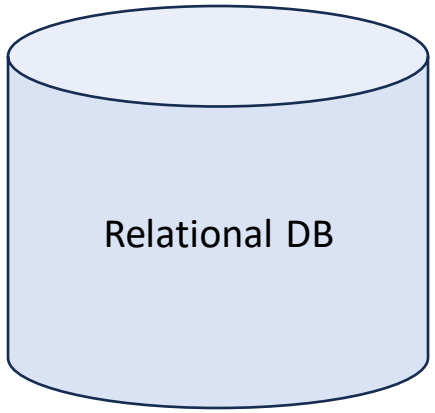
- By Sabarish Nair

Recap from last week....

- Traditional Data Warehouse
- Modern Data Warehouse
- Operational Data Store (ODS)
- In-Memory Databases
- Metadata
- Types of Metadata
- Benefits of Metadata
- Best Practices: Metadata
- Data Warehousing Approaches
- Inmon's Data warehousing Approach
- Kimball's Data warehousing Approach
- Kimball's front and back room analogy

Relational Database

Tables with Rows and Columns



Columns

Rows

Column 1	Column 2	Column 3
Row 1...		
Row 2...		

Primary Key – Foreign Key Relationship

Primary Key

Product _ID	Product_Name	Date	Customer_Id
1	Bat	23/02/2024	1
2	Ball	22/02/2024	2
3	Computer	21/02/2024	4

Foreign Key

Customer_id	Name	City
1	John	Cork
2	Jill	Dublin
4	Bill	Gallway

What is dimensional modeling?

Method of organizing data

A design technique used in data warehousing to structure and organize data for easy and efficient querying and analysis.

Primary Components of Dimensional Modeling

Fact Tables (Quantitative Data)

Dimension Tables (Descriptive Attribute)

Star Schema (Fact at center, dimensions around, Star shaped)

Snowflake Schema (Dimension Tables normalized)

Benefits of Dimensional Modeling

SIMPLIFIED QUERYING

PERFORMANCE OPTIMIZATION

INTUITIVE DESIGN

SCALABILITY

Dimensional Modeling (Basic Example)

Facts

Measurement by profit

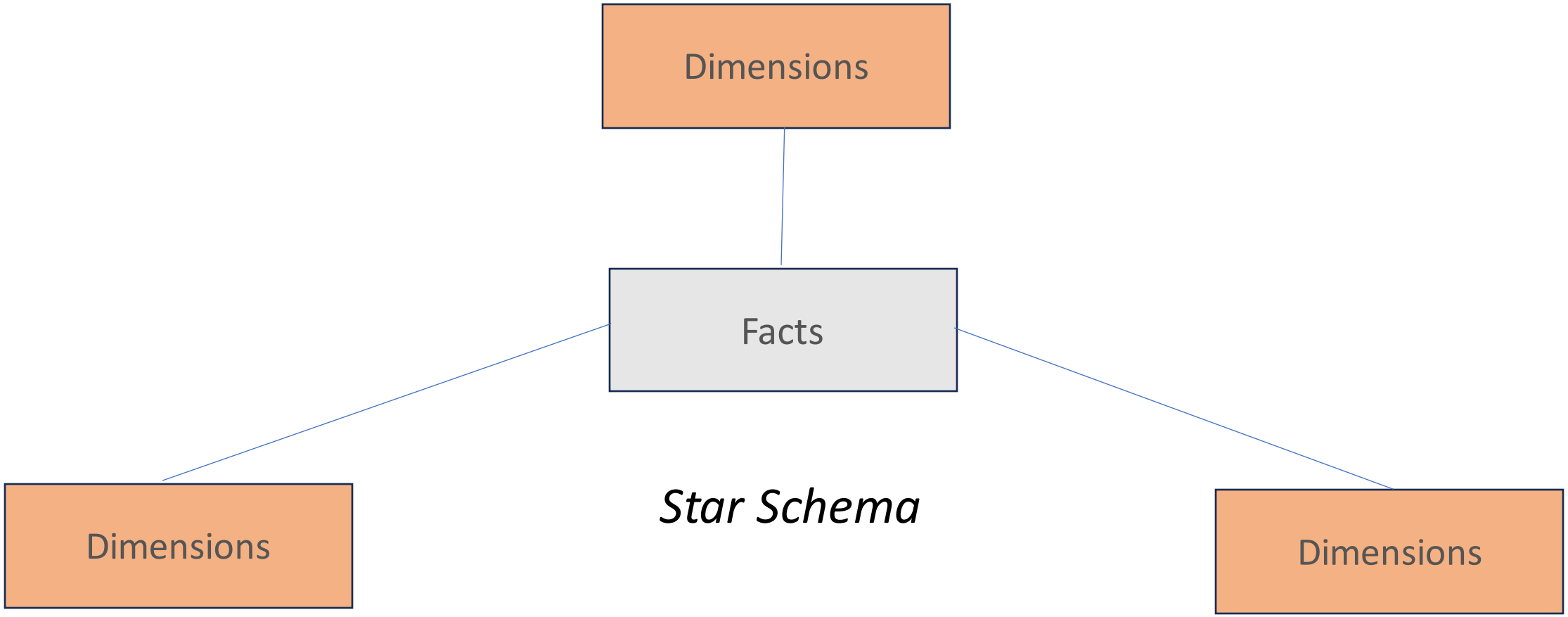
Dimensions

Context like period or category

Profit By Department

Profit By Month

Dimensional Modeling (Star Schema)



Star Schema

Fact Table

Measurable, numeric data that businesses want to analyze.

Quantitative measures or metrics that can be aggregated, summed, or averaged.

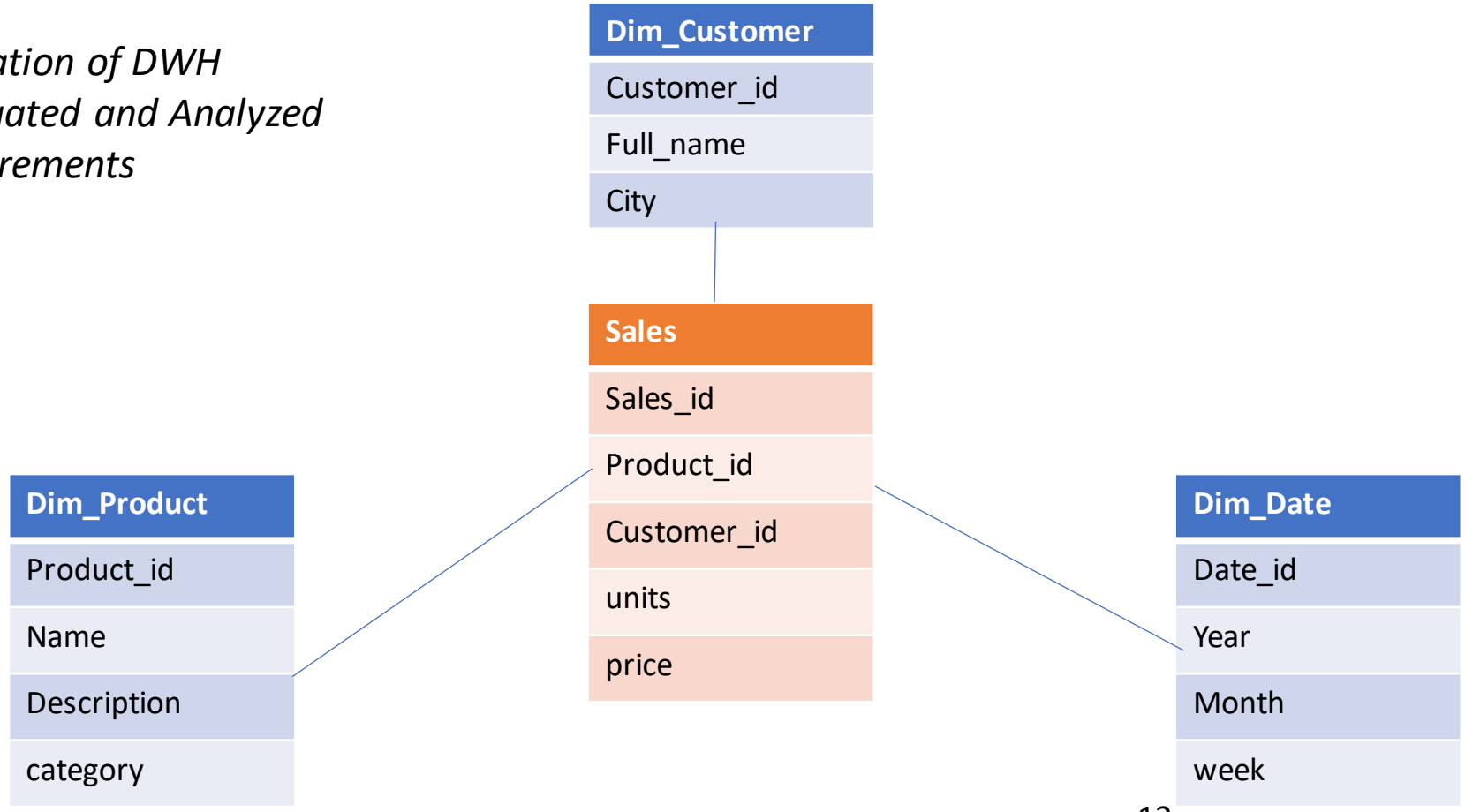
Example: Sales Revenue, Quantity Sold, Profit Margin, Customer Count, Website Visits

Characteristics of Fact Tables

- Numeric Values
- Granularity
- Additivity (Additive, Semi-additive, Non-additive)
- Foreign Key Relationships
- Surrogate Keys
- Time Stamps

Facts

- *Foundation of DWH*
- *Aggregated and Analyzed*
- *Measurements*



Dimensions

Descriptive attributes or context by which facts (numeric, measurable data) are analyzed.

"Who, what, where, when, why, and how" aspects of the data.

Characteristics of Dimensions

Descriptive attributes

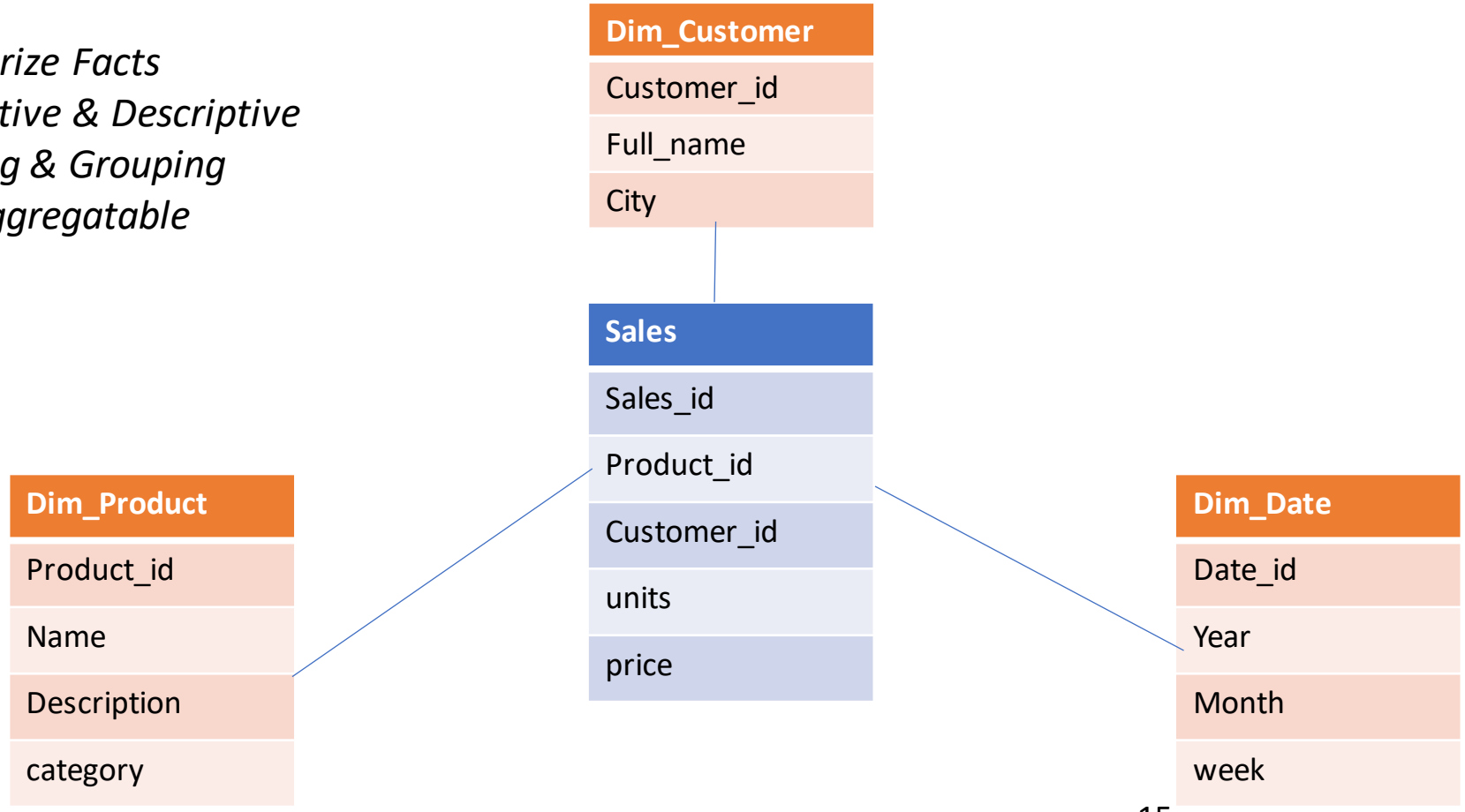
- Time Dimension
- Product Dimension
- Customer Dimension
- Location Dimension

Hierarchical structure

- Year > Quarter > Month > Day

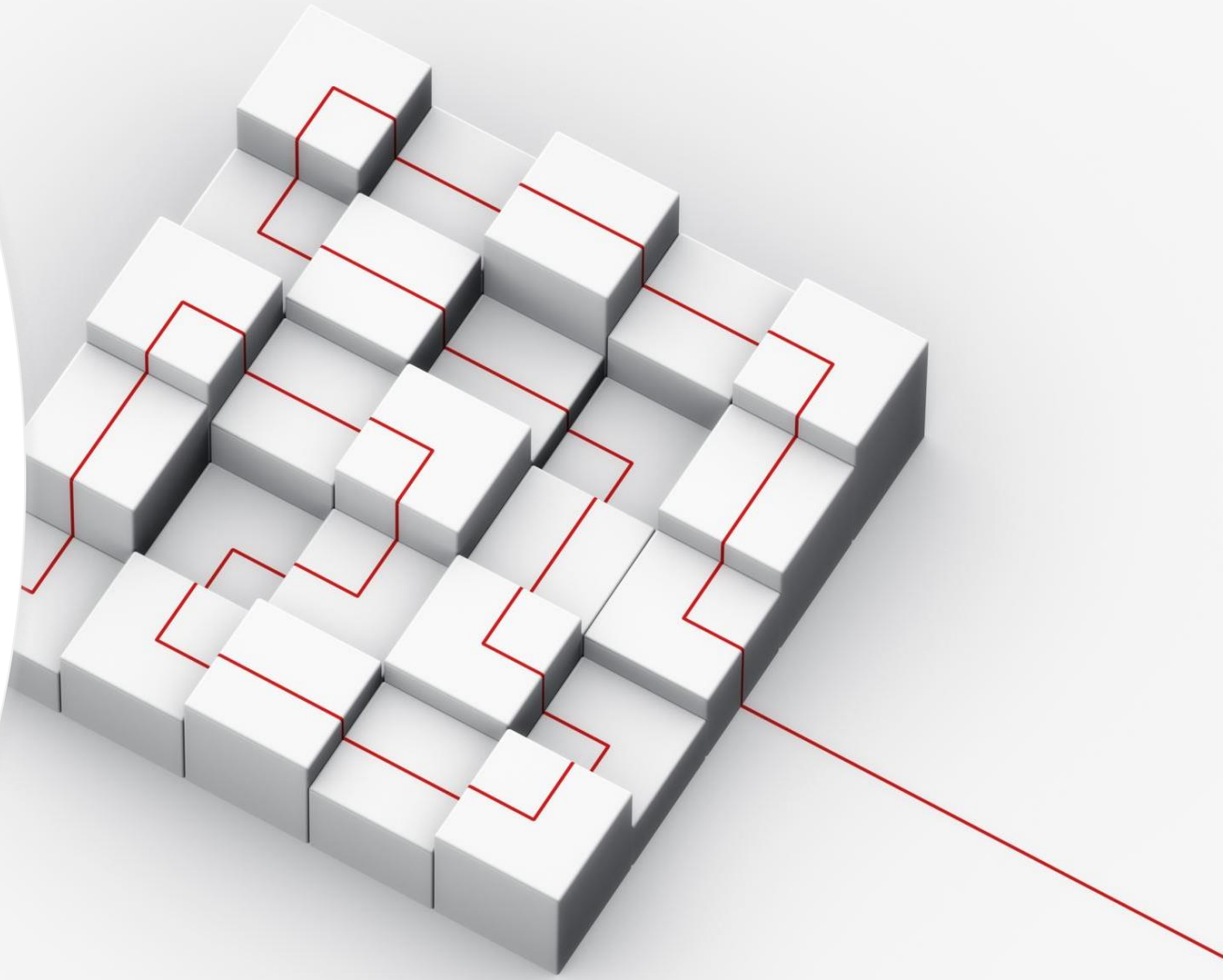
Dimensions

- *Categorize Facts*
- *Supportive & Descriptive*
- *Filtering & Grouping*
- *Non-Aggregatable*
- *Static*



Star Schema

- A central fact table surrounded by multiple dimension tables.
- Denormalized
  - Data Redundancy
  - Query Performance (Read)
  - Optimized to get data out
  - User Experience
- Relationship Type (1:n)



Star Schema

USED COMMONLY IN DATA MART

SIMPLEST FORM (VS. SNOWFLAKE SCHEMA)

TUNED FOR SPECIFIC NEEDS

SPECIFIC USE CASE (READ) [USABILITY AND PERFORMANCE]

Snowflake Schema

MORE Normalized!!

Fact Table

Product_id	Customer_id	Price	date
1	1	20	24/02/2024
2	2	222	25/02/2024

Product_id	Name	Category_id	date
1	Bat	1	Sports
2	Ball	1	Sports

Dimension Table (Level 1)

Category_id	Name
1	Sports
2	Toys

Dimension Table

(Level 2)

Characteristics of Snowflake Schema

- Normalized Structure
- Multiple levels if normalization
- Increased Joins
- Data Integrity
- Storage Efficiency
- Complexity

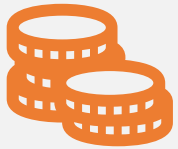
Star Schema vs Snowflake Schema

Parameter	Snowflake Schema	Star Schema
Structure	Normalized	Denormalized
Complexity	Complex	Simpler and Easier
Query Performance	Reduced	Better
Storage Efficiency	Optimized	Higher storage requirements
JOINS	More Joins	Fewer Joins
Use Cases	OLAP	Simpler analytical and reporting needs

Question 1: Identify the Fact and Dimension Tables

You are tasked with designing a data warehouse schema for a retail company that wants to analyze its sales data. The company sells various products through multiple stores across different locations. They want to analyze sales performance based on different dimensions such as time, product, and store.

Solution to Question 1

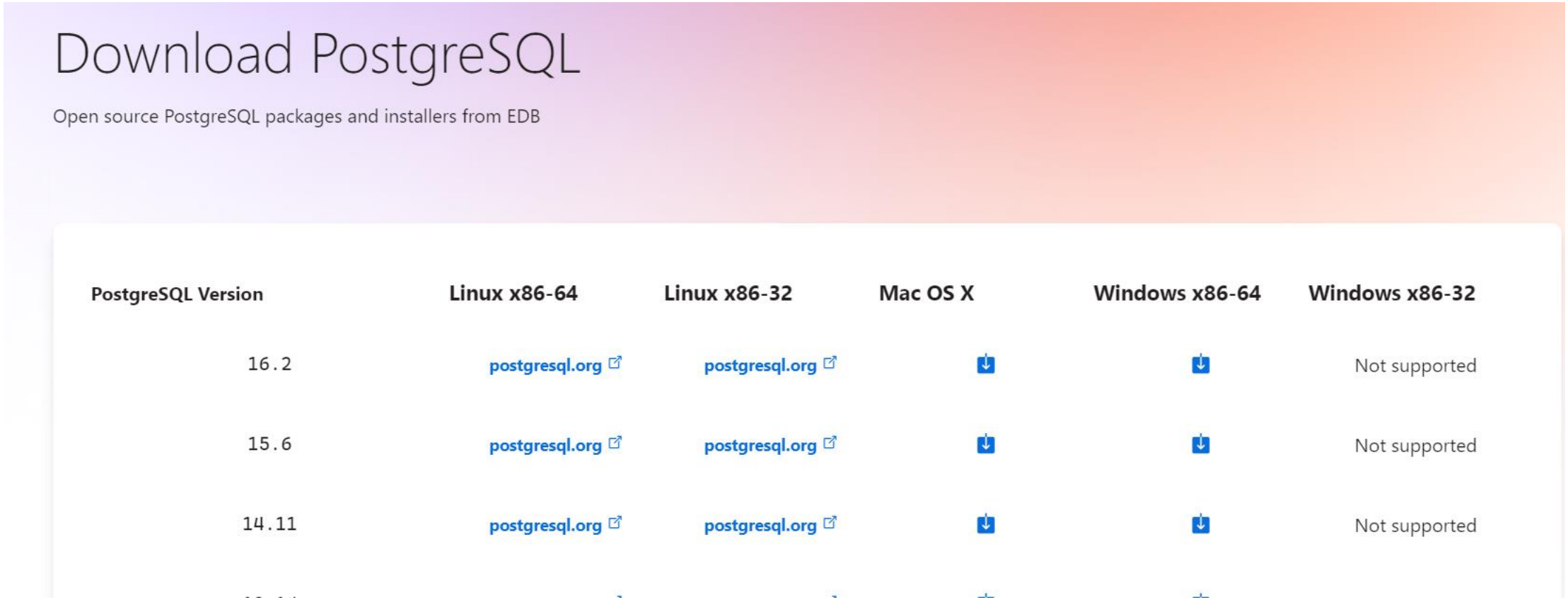


Fact Table: Sales

Dimension Table: Date, Product, and Store

Install PostgreSQL

https://www.enterprisedb.com/downloads/postgres-postgresql-downloads




Thank you