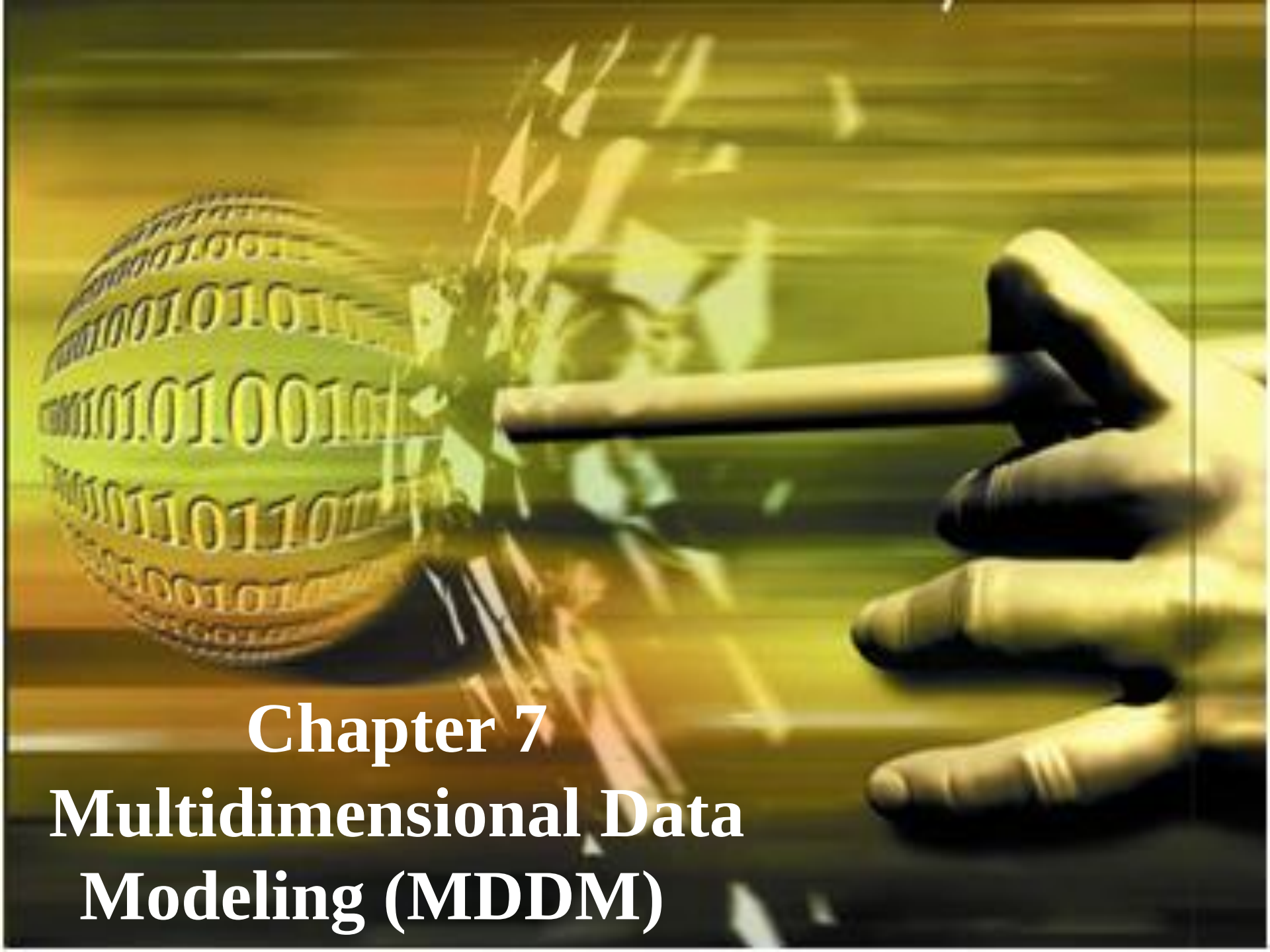


Chapter 7
Multidimensional Data Modeling (MDDM)

Learning Objectives and Learning Outcomes

Learning Objectives	Learning Outcomes
1. To assess the capabilities of OLTP and OLAP systems	(a) Appreciate the differences between OLTP and OLAP systems
2. To understand Dimensional Modeling (DM)	(b) Understanding of DM, basics of data warehousing and related terminology; also an overview of what facts and dimensions are
3. To learn data warehousing in further detail through a case-study	(c) Understand how to convert an OLTP schema into a dimensional schema model through various techniques of data warehousing

Session Plan

Lecture time	:	90 minutes approx.
Q/A	:	15 minutes

Agenda

- Database Concepts – Recap
- Introduction to On-line Analytical Processing (OLAP)
- Multidimensional Data Modeling (MDDM)

  – To Answer Why ? Where ? When ? and How ?

Recap

- Databases and Tables

- Normalization and Keys
  - ACID Properties
  - Primary, Foreign and Surrogate Keys
  - Cardinality

- Transactions
  - On-Line Transaction Processing
  - On-Line Analytical Processing

Recap (contd.)

- Difference between OLTP and OLAP

	OLTP	OLAP
Definition	On Line Transaction Processing	On Line Analytical Processing
Data	Dynamic (day to day transaction / operational data)	Static (historical data)
Data Atomicity	Data is stored at microscopic level	Data is aggregated or summarized and stored at the higher level
Normalization	Normalized Databases to facilitate insertion, deletion and updation	De-normalized Databases to facilitate queries and analysis
History	Old data is purged or archived	Historical data stored to enable trend analysis and future predictions
Queries	Simple queries and updates Queries use small amounts of data (one record or a few records) Example: update account balance enroll for a course	Complex queries Queries use large amounts of data Example: Total annual sales for north region Total monthly sales for north region

Comparison of OLTP and DSS

OLTP Capability Examples

- Search & locate student(s)
- Print student scores
- Filter students above 90% marks
- Update student Grade
- Group by Batch and compute average score
- Find top 10 high performance students

DSS Capability Examples

- Which courses have productivity impact on-the-job?
- Which colleges need to be rewarded for supplying students with consistent high on-the-job performance?
- What is the customer satisfaction improvement due to extended training?
- How project level profitability is influenced by certification?
- How much training is needed on future technologies for non-linear growth in BI?

Still a little fuzzy about what OLAP can do?

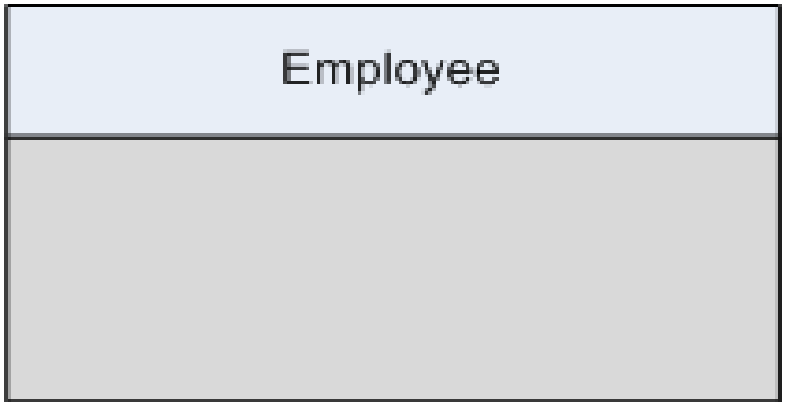
Introduction to On-line Analytical Processing (OLAP)

Scenario:
Internal systems department at Infosys maintains all relevant data in a database. Conceptual schema is as shown below.

OLAP Contd.

- CEO of the company wants the following information from the IS department.
  - Number of employees added in the role of the company during the last quarter/6 months/1 year

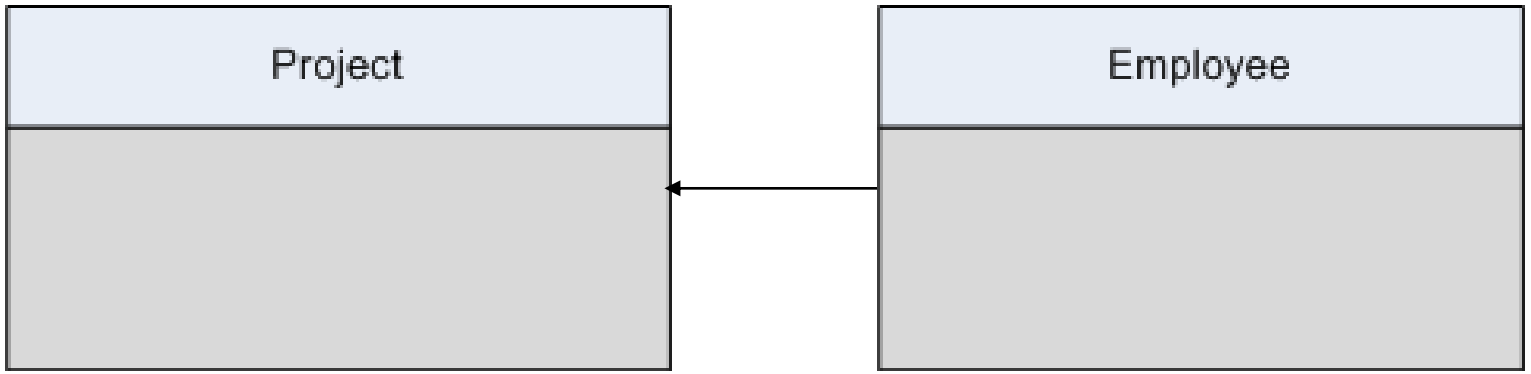
Q1. How many table(s) is/are required ?



OLAP Contd.

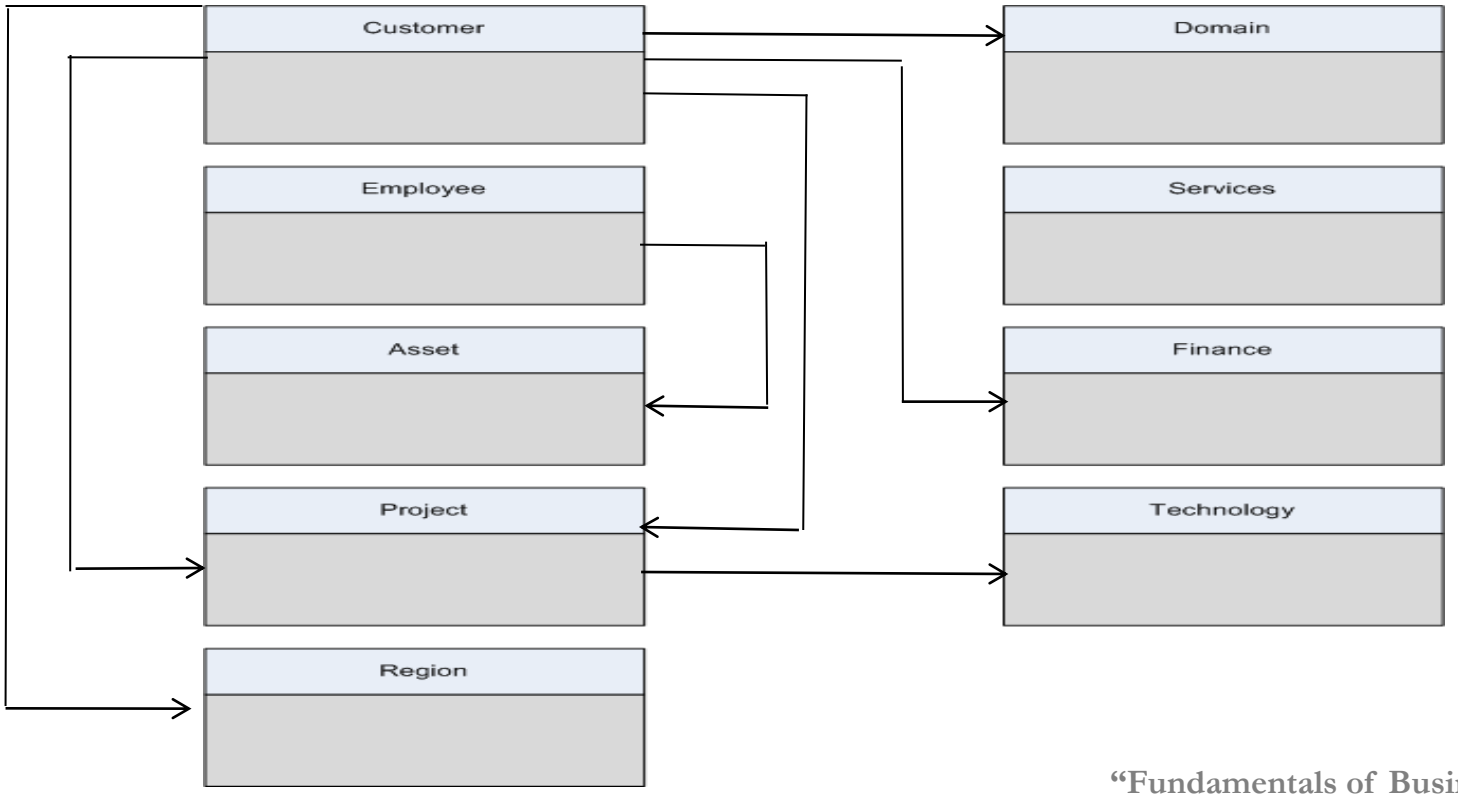
Q2. How many employees are currently in the projects ?

Q3. How many are on bench?



OLAP Contd.

Q4. Which **customer** from a **region** has given maximum **business** during the previous quarter on a **domain** under specific **technology** and who are the **PMs** of the **project** **and** **assets** **owned by them.**



SOLUTION?

MDDM

Answer a Quick Question

So if there were very few updates and more of data-retrieval queries being made on your database, **what do you think would be a better schema to adopt?**

OLTP or OLAP

Introduction to Dimensional Modeling (DM)

- DM is a logical design technique used in Data Warehouses (DW). It is quite directly related to OLAP systems
- DM is a design technique for databases intended to support end-user queries in a DW
- It is oriented around understandability, as opposed to database administration

However, before we actually jump into MDDM…

let's first understand the language of
Dimensional Modeling

MDDM Terminology

- Grain
- Fact
- Dimension
- Cube
- Star
- Snowflake

Of hierarchies and levels…

What Is a Grain?

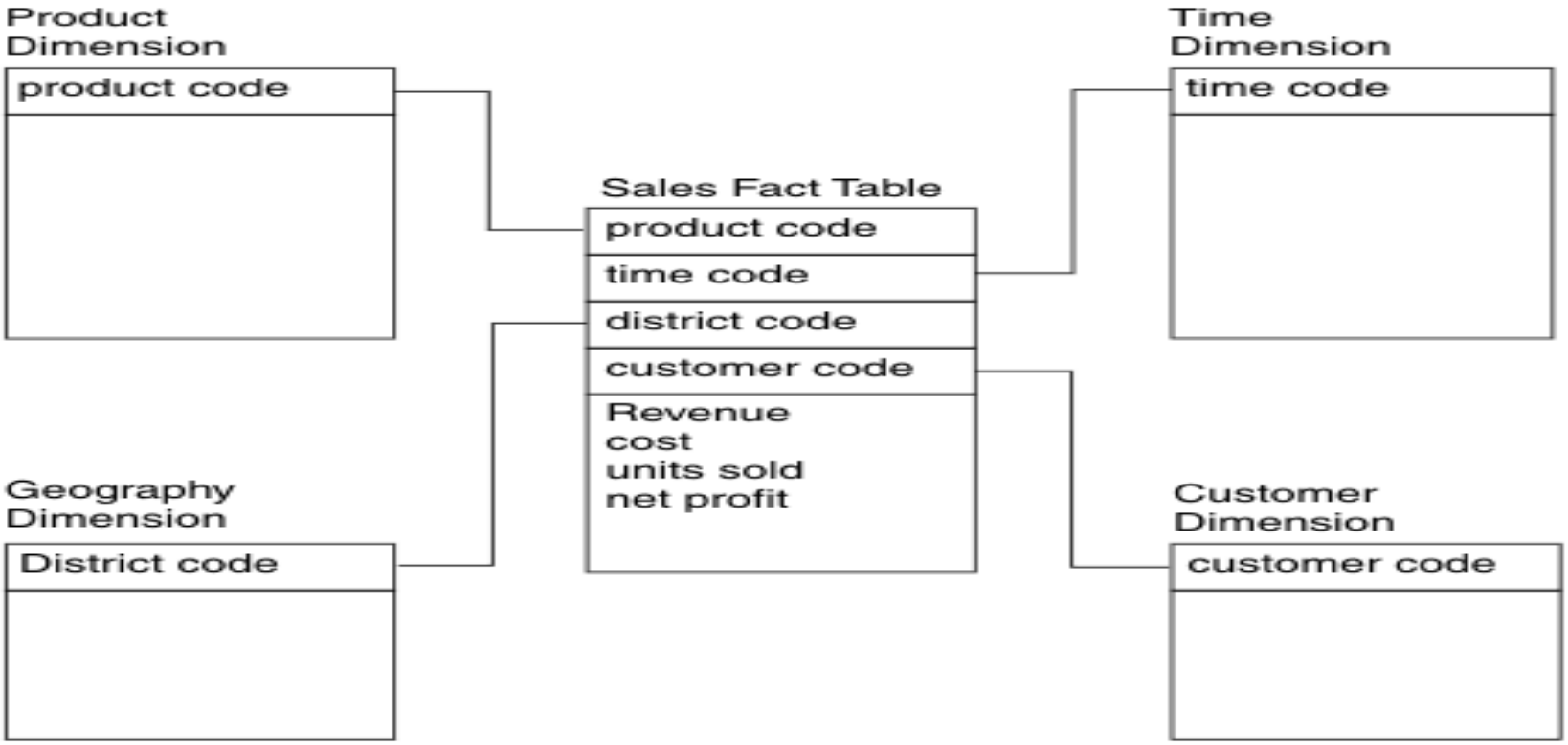
- Identifying the grain also means deciding the level of detail that will be made available in the dimensional model
- *Granularity* is defined as the detailed level of information stored in a table
- The more the detail, the lower is the level of granularity
- The lesser the detail, higher is the level of granularity

What can be measured, can be controlled…

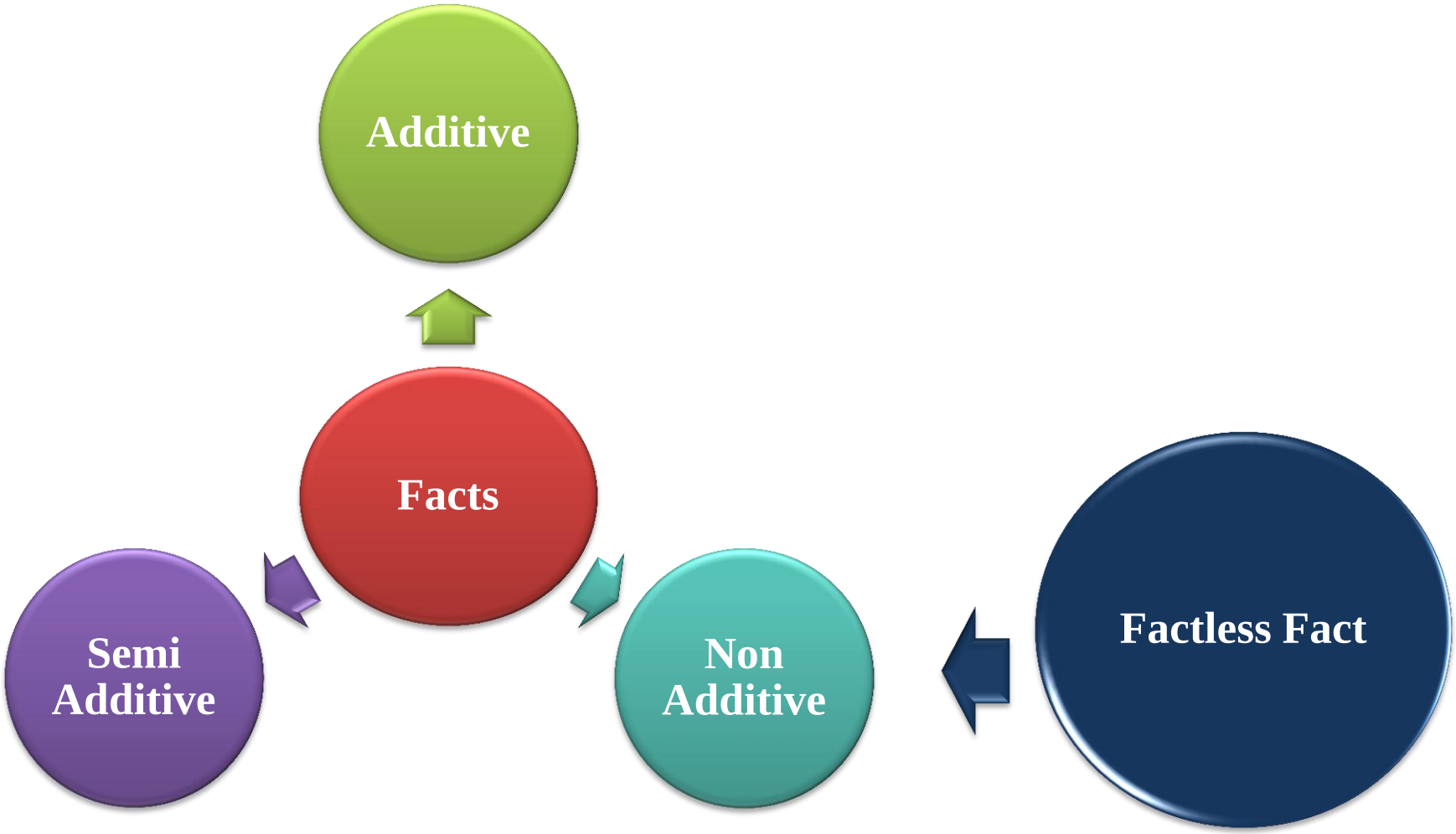
…and do you know how such measurements are stored in a data warehouse?

Facts and Fact Tables

- Consists of at least two or more foreign keys
- Generally has huge numbers of records
- Useful facts tend to be numeric and additive



Fact Types



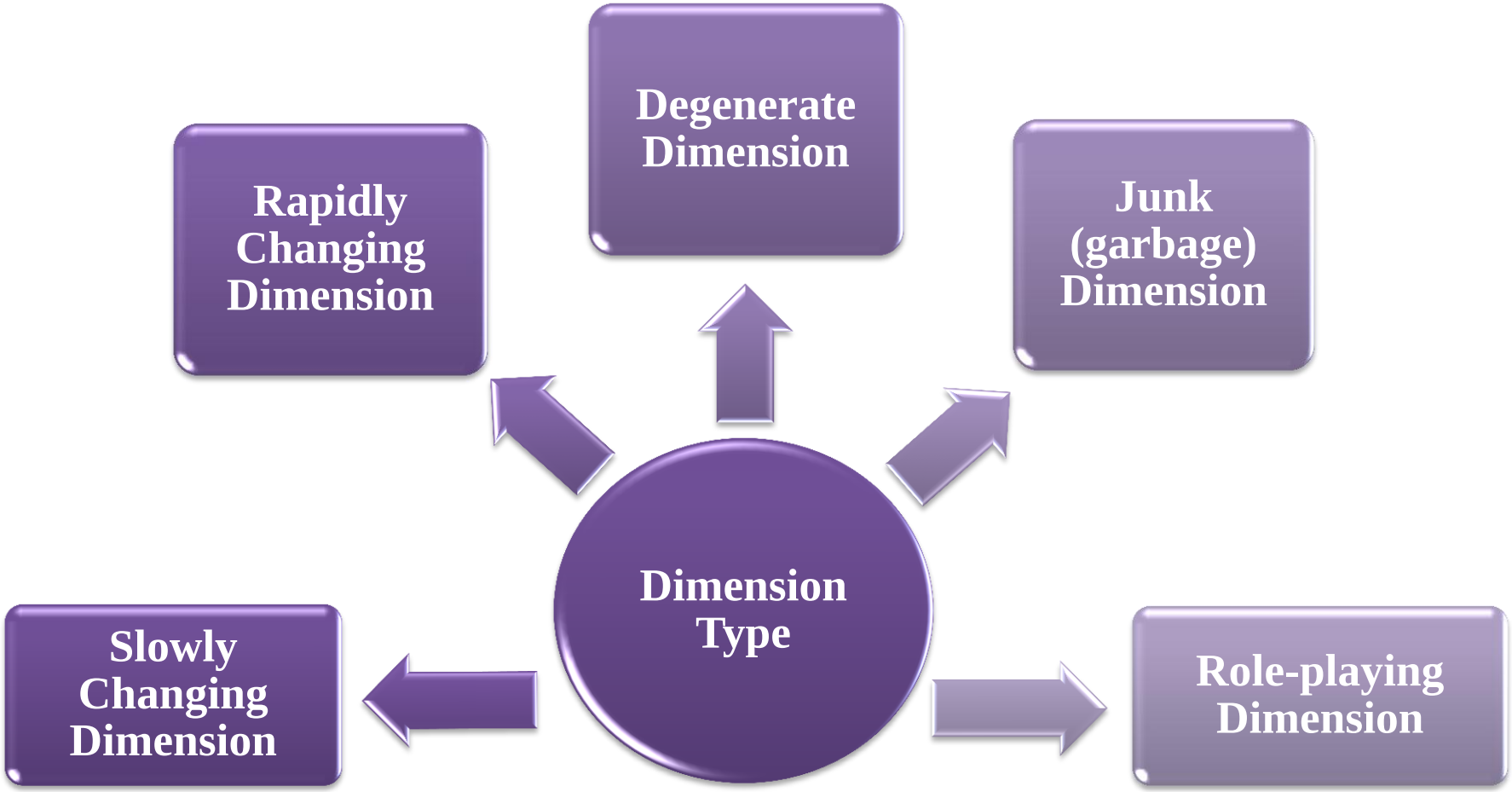
And what about descriptive data?

What Are Dimensions/Dimension Tables?

- The dimensional tables contain attributes (descriptive) which are typically static values containing textual data or discrete numbers which behave as text values.

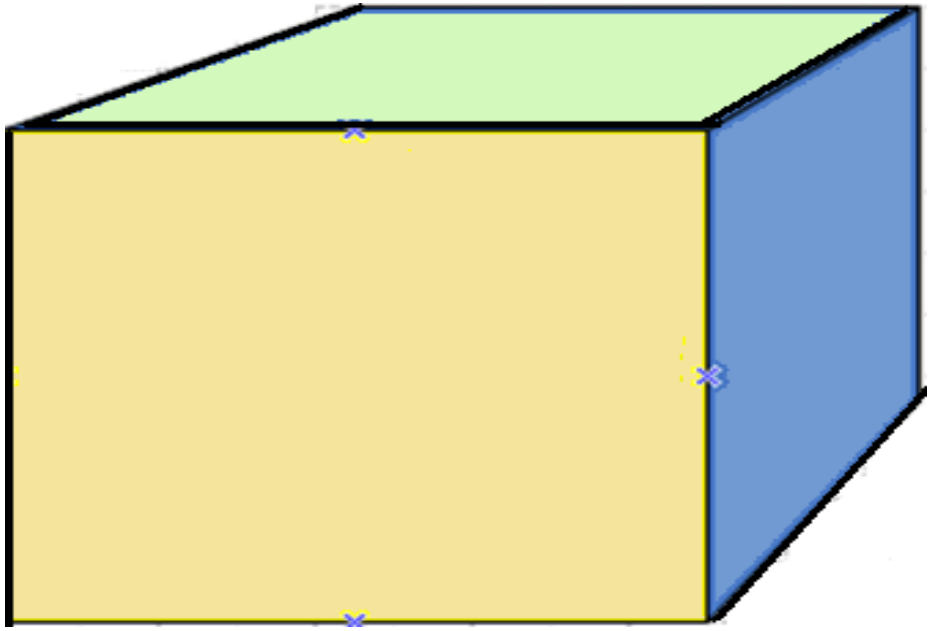
- Main functionalities :
  - Query filtering\constraining
  - Query result set labeling

Types of Dimensions



Understanding Dimension – Cube

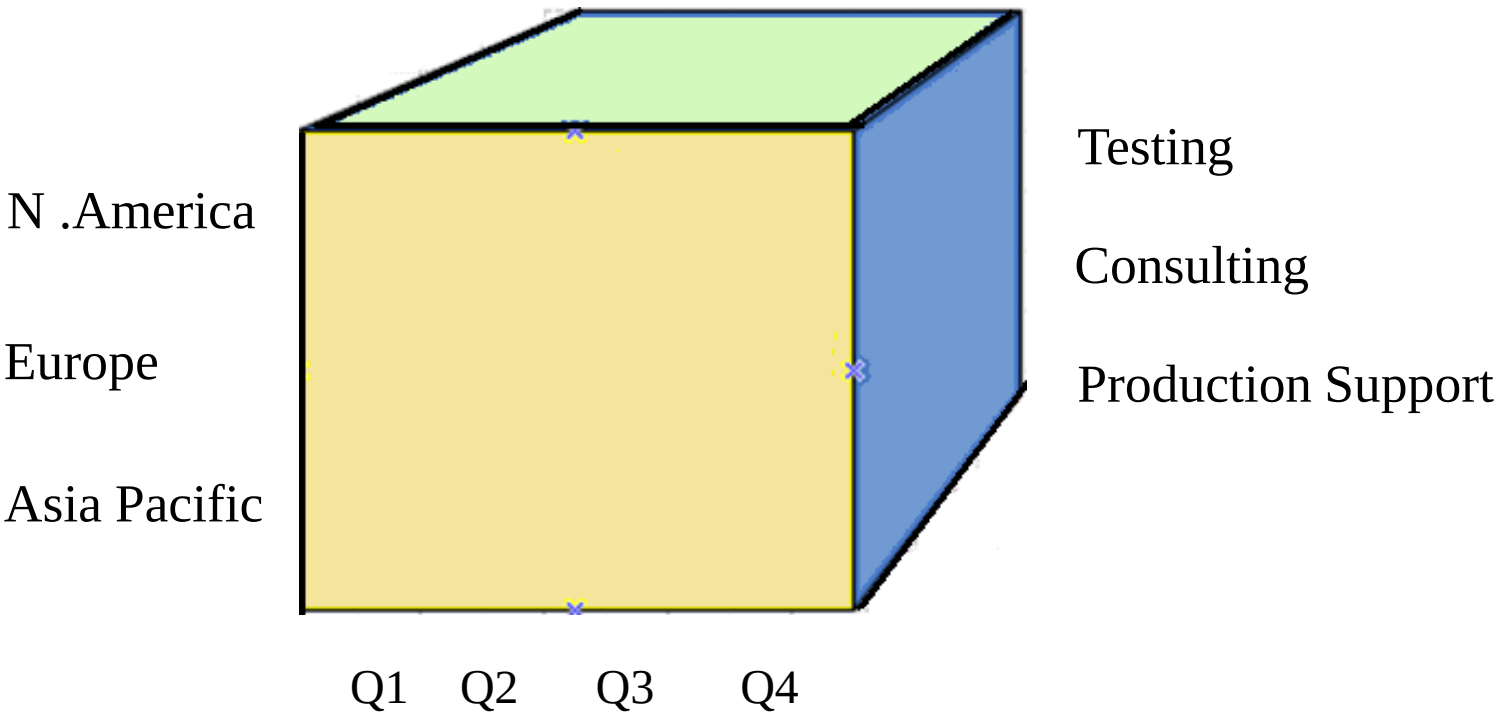
- An extension to the two-dimensional Table.
- For example in the previous scenario CEO wants a report on revenue generated by different services across regions during each quarter



Understanding Dimension – Cube (contd.)

Dimension Hierarchy

Grain
Fact



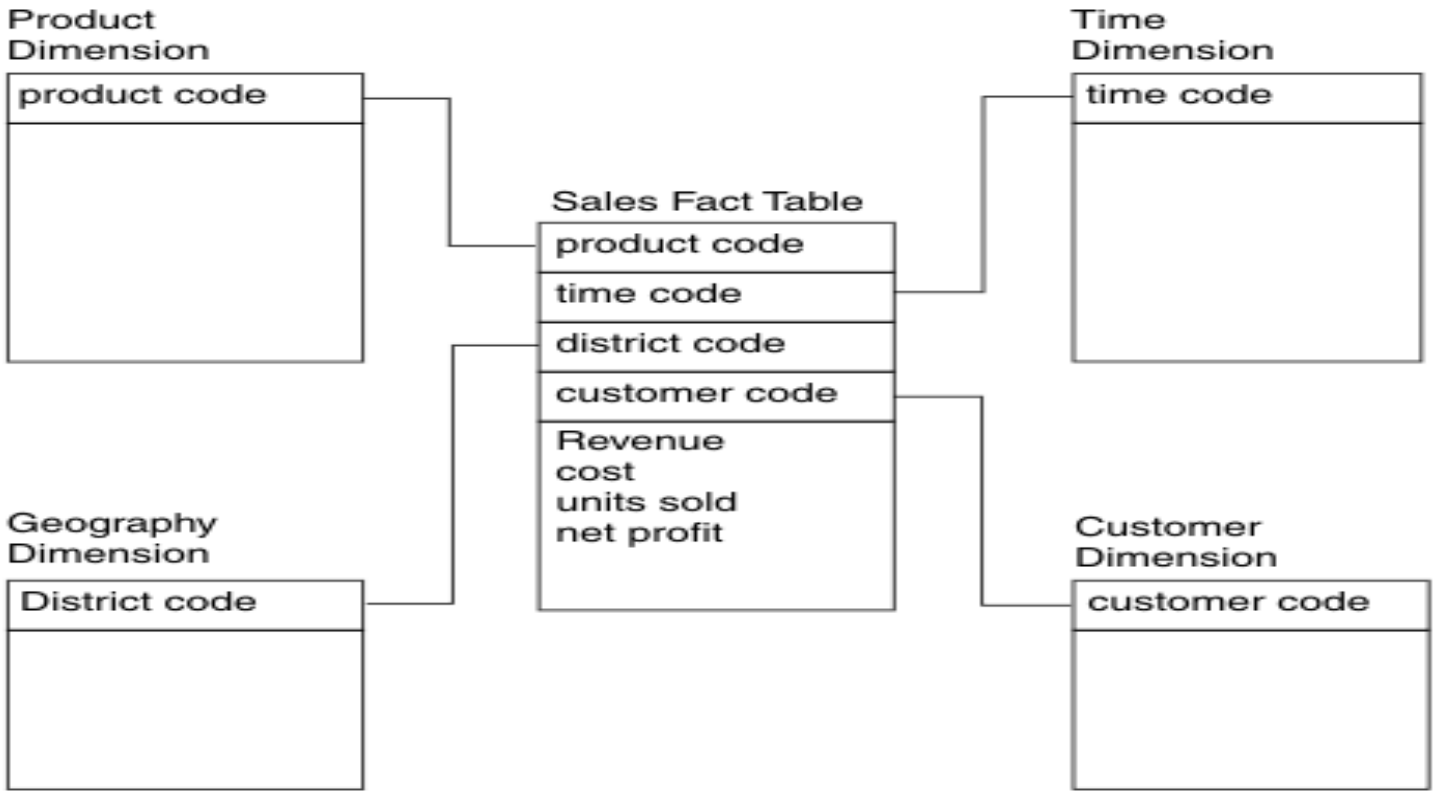
Answer a Quick Question

Can a data warehouse schema take different forms with respect to normalization?

Star Schema

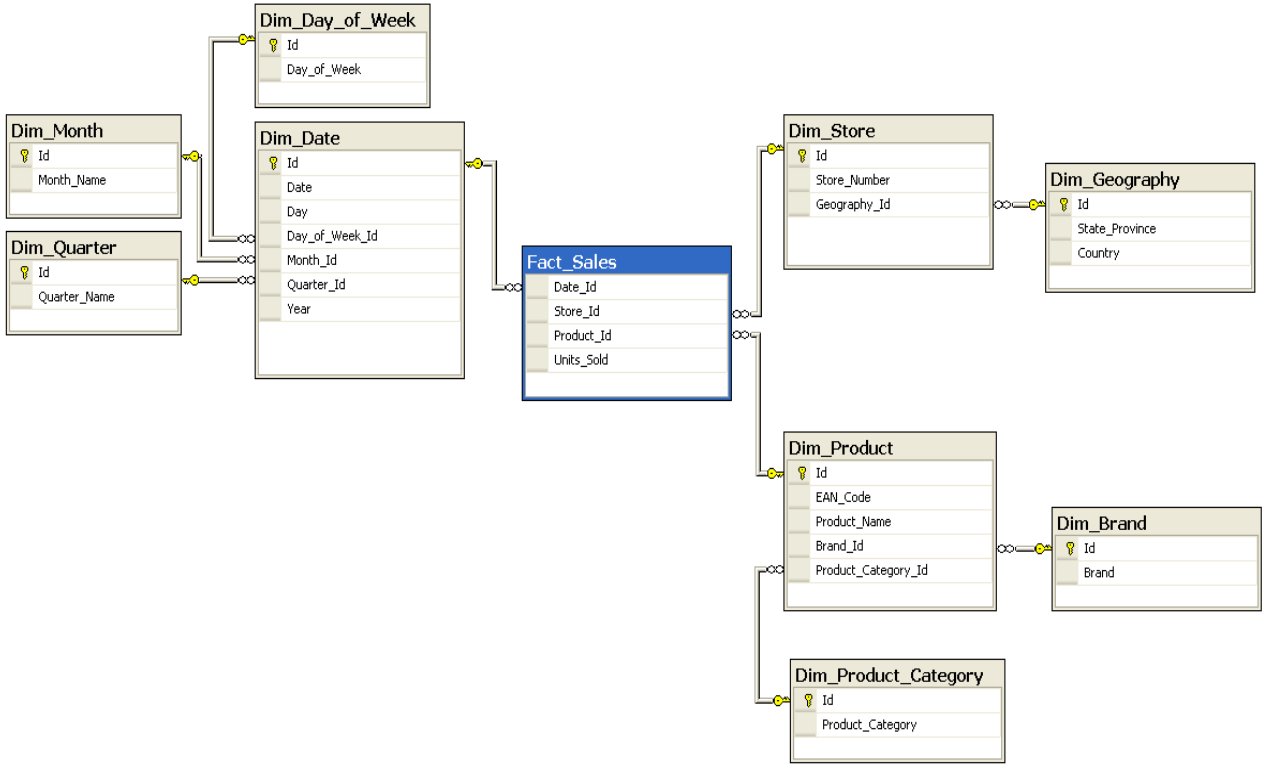
The basic star schema contains four components.
These are:

Fact table, Dimension tables, Attributes and Dimension hierarchies



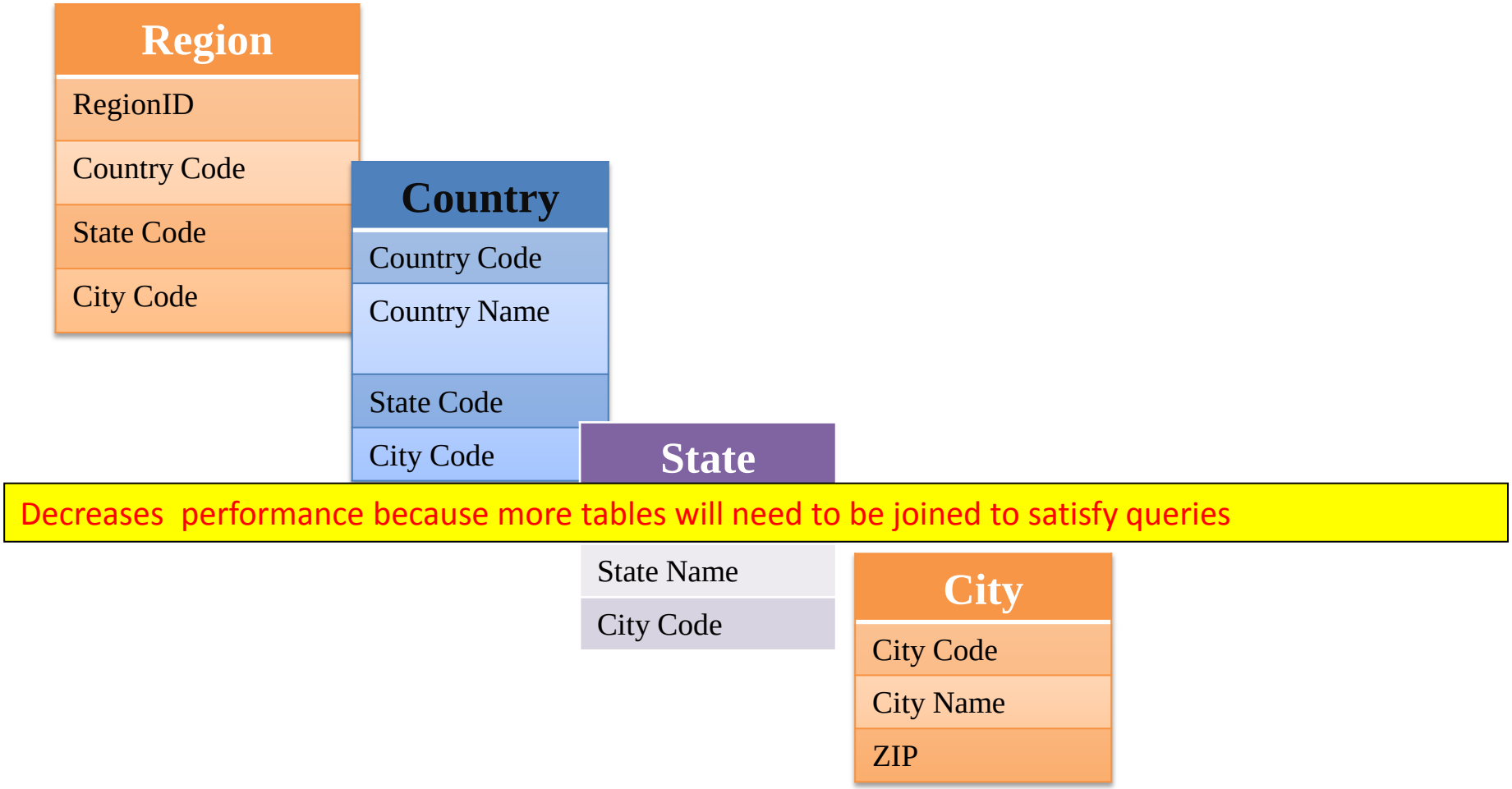
Snow Flake Schema

- Normalization and expansion of the dimension tables in a star schema result in the implementation of a snowflake design.
- A dimension table is said to be snow flaked when the low-cardinality attributes in the dimension have been removed to separate normalized tables and these normalized tables are then joined back into the original dimension table.



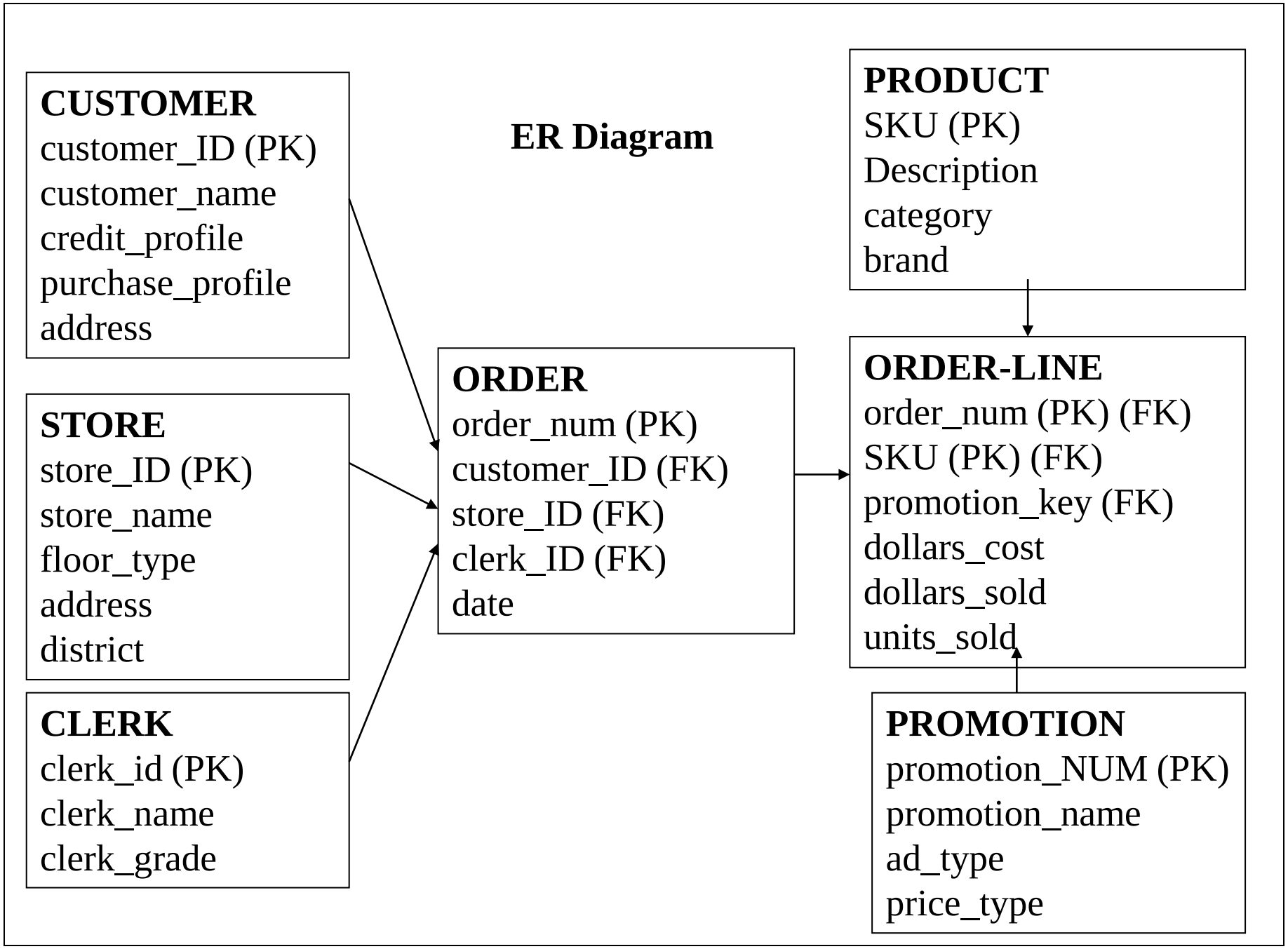
Snow Flaking Example

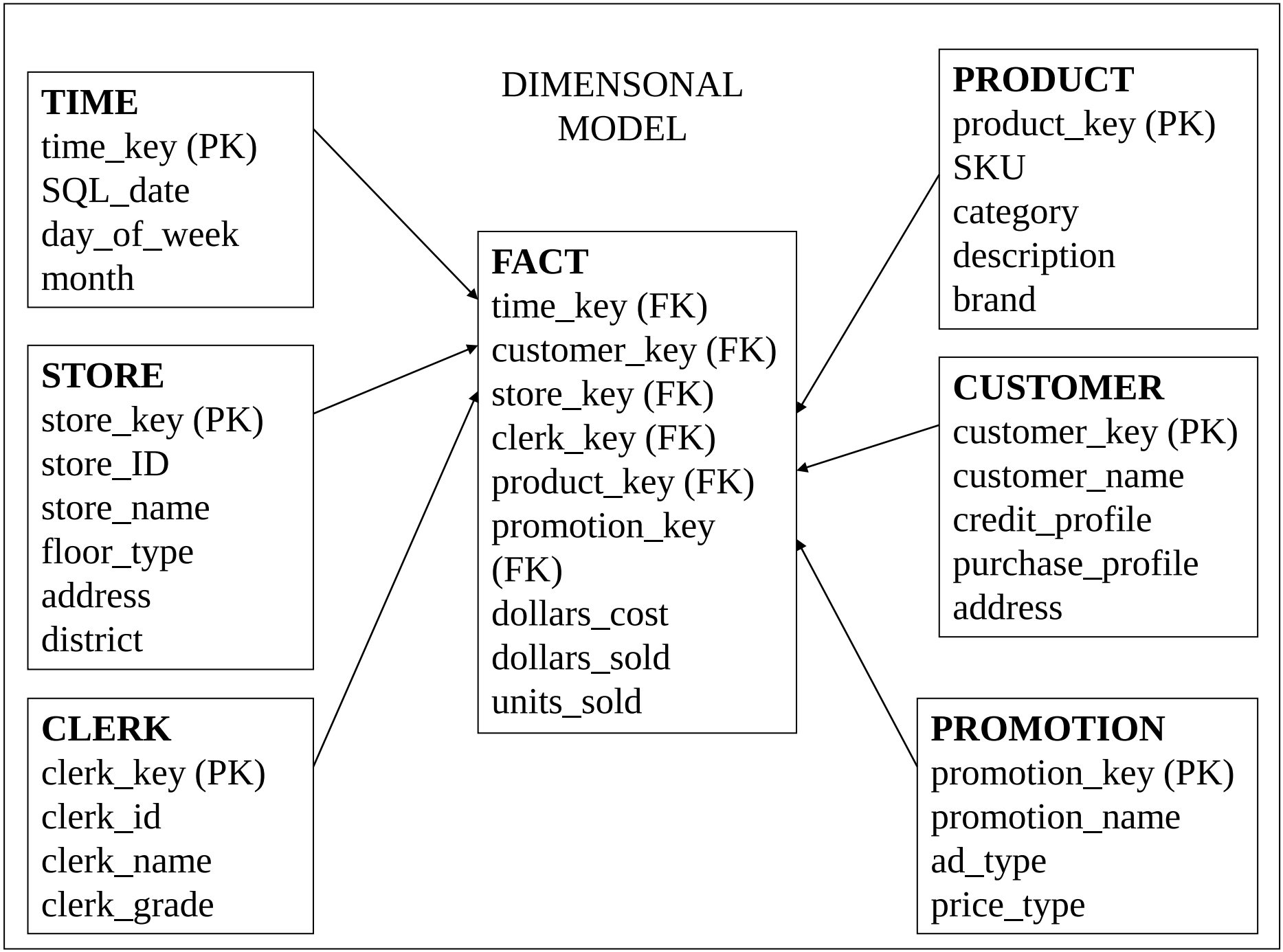
- Consider the Normalized form of Region dimension



Armed with these weapons that we call ‘Concepts’, let’s step into the battlefield!

Case Study
Conversion of a ER Model to a Dimensional Model


ER Diagram
CUSTOMER
customer_ID (PK)
customer_name
credit_profile
purchase_profile
address
STORE
store_ID (PK)
store_name
floor_type
address
district
CLERK
clerk_id (PK)
clerk_name
clerk_grade
ORDER
order_num (PK)
customer_ID (FK)
store_ID (FK)
clerk_ID (FK)
date
PRODUCT
SKU (PK)
Description
category
brand
ORDER-LINE
order_num (PK) (FK)
SKU (PK) (FK)
promotion_key (FK)
dollars_cost
dollars_sold
units_sold
PROMOTION
promotion_NUM (PK)
promotion_name
ad_type
price_type


DIMENSONAL MODEL
TIME
time_key (PK)
SQL_date
day_of_week
month
STORE
store_key (PK)
store_ID
store_name
floor_type
address
district
CLERK
clerk_key (PK)
clerk_id
clerk_name
clerk_grade
FACT
time_key (FK)
customer_key (FK)
store_key (FK)
clerk_key (FK)
product_key (FK)
promotion_key (FK)
dollars_cost
dollars_sold
units_sold
PRODUCT
product_key (PK)
SKU
category
description
brand
CUSTOMER
customer_key (PK)
customer_name
credit_profile
purchase_profile
address
PROMOTION
promotion_key (PK)
promotion_name
ad_type
price_type

Summary

- Basics of Database
- OLTP
- MDDM
- Cube
- Star Schema
- Snowflake schema

Food for Thought!

1. Who according to you would be the user of OLAP?
2. Who would need the Multidimensional perspective of data?