What Are Facts and Dimensions in a Data Warehouse?

- Facts and dimensions are the fundamental elements that define a data warehouse.
- They record relevant events of a subject or functional area (facts) and the characteristics that define them (dimensions).
- Data warehouses are data storage and retrieval systems (i.e., databases) specifically designed to support business intelligence (BI) and OLAP (online analytical processing) activities.

The Four Pillars of the Data Warehouse

- 1. Oriented to a single subject or a particular functional area. For example, it is oriented to company sales.
- 2. They unify and create consistency among data from disparate sources.
- 3. Persistent and immutable. Once data enters a data warehouse, it stays there and does not change.
- 4. Structured in time intervals. To provide information from a historical perspective, data warehouses record information over different intervals, such as weekly, monthly, quarterly, etc.

A data warehouse is a database:

- From a technical point of view, a data warehouse is a database.
- Therefore, it is composed of tables, fields, relations, keys – just like any other database.
- And as such, we can model it using a database design tool such as VERTABELO.

The tables of a data warehouse:

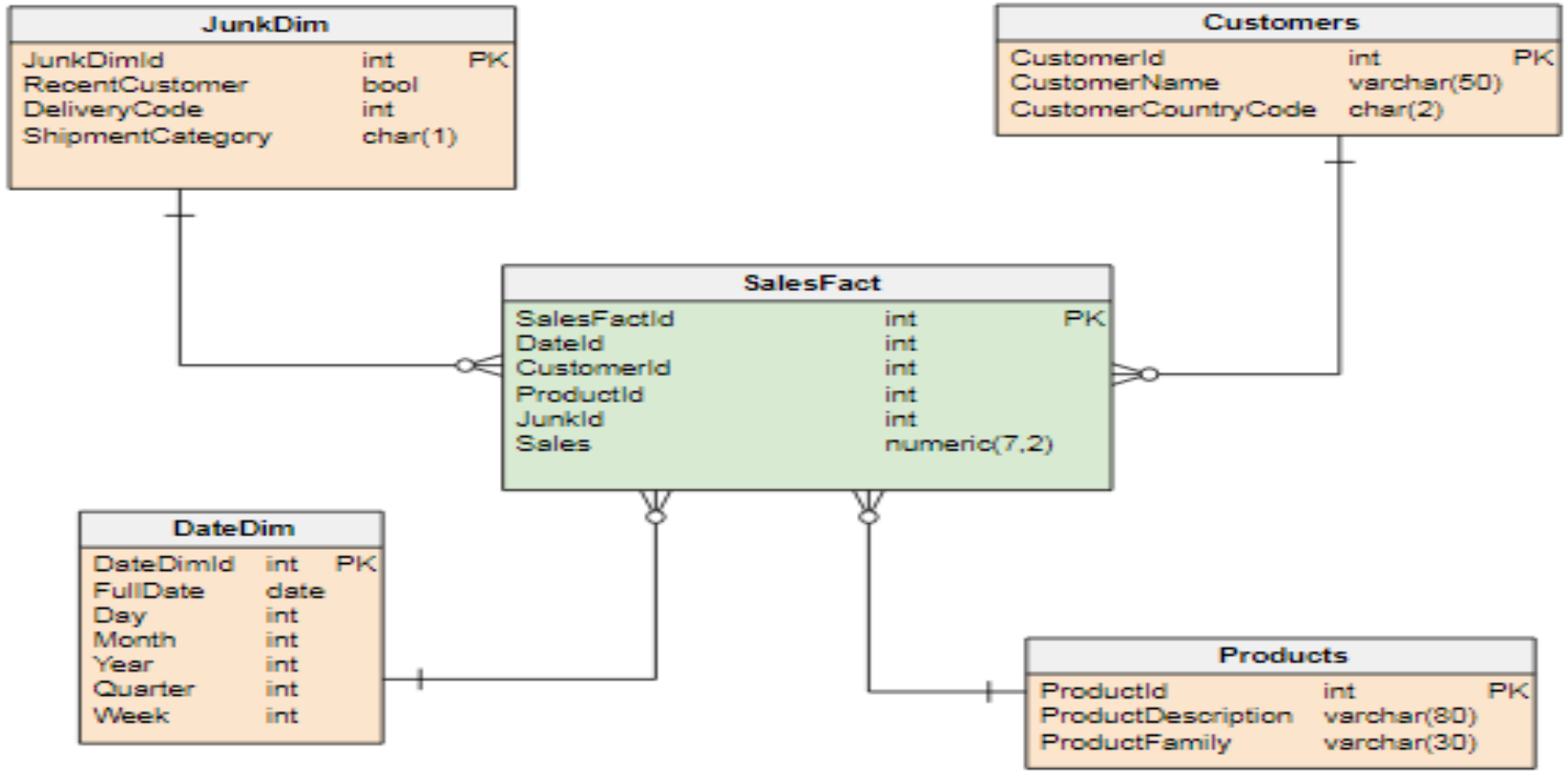
Data warehouse tables are divided into two main categories:

1. fact tables and
2. dimension tables.

Topologies:

- Facts and dimensions in a data warehouse should form a layout that responds to a particular topology.
- There are two main topologies: the STAR SCHEMA and the SNOWFLAKE SCHEMA.
- In a star schema, individual dimensions surround a single fact table, while a snowflake schema has a hierarchy of dimensions.

A typical star-shaped data warehouse schema: the fact table sits in the middle, surrounded by the dimension tables.



Fact tables:

- Fact tables are the core tables of a data warehouse.
- They contain quantitative information, commonly associated with points in time.
- They are used in trends, comparisons, aggregations, and groupings.
- They feed analysis and visualization tools to allow insights to be discovered about the functional area.

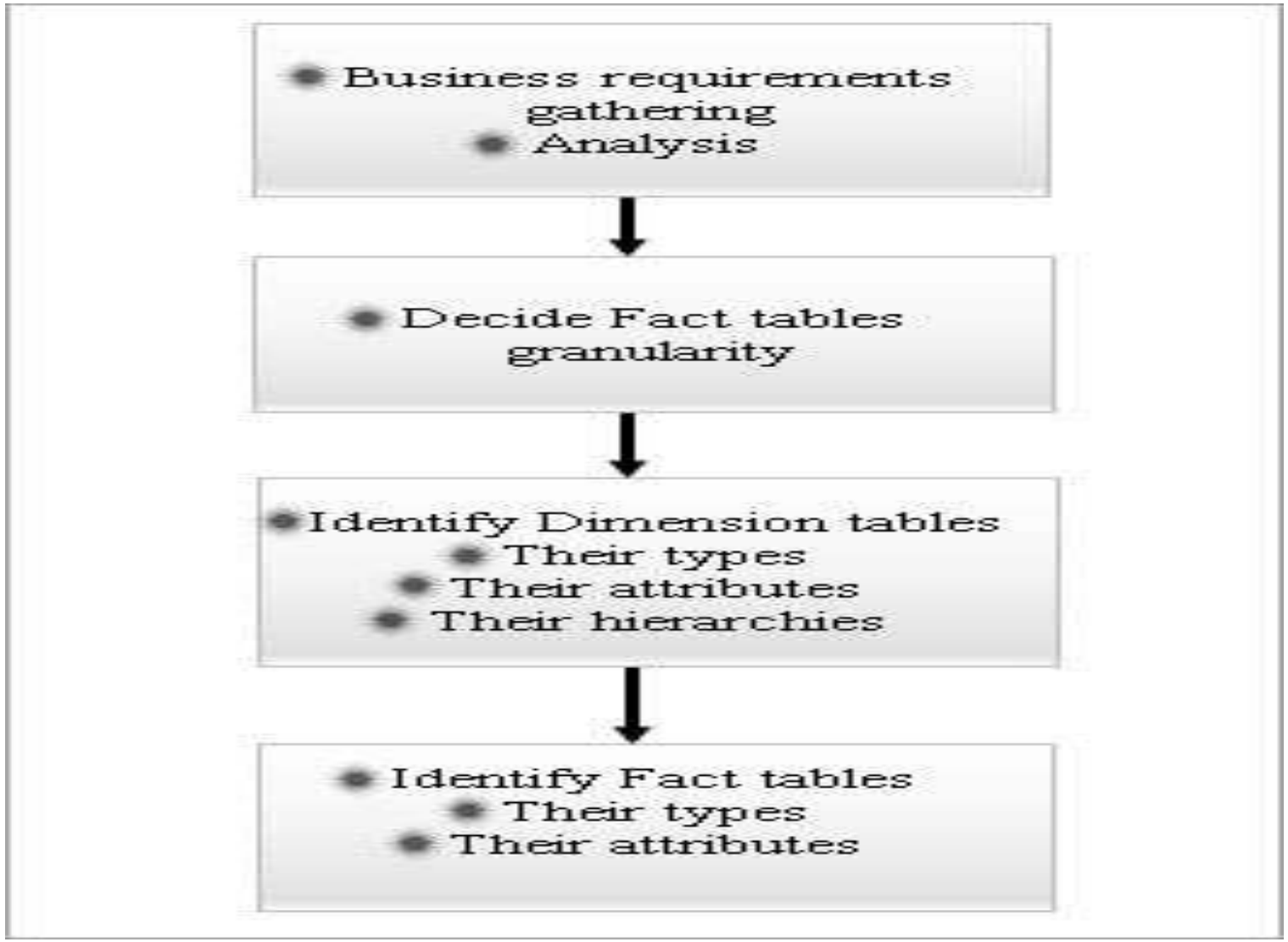
Dimensions:

- Dimensions, on the other hand, are collections of reference information about the facts in a data warehouse.
- Dimensions categorize and describe the facts recorded in a data warehouse to provide meaningful, categorized, and descriptive answers to business questions.

Dimensional Data Models

- Dimensional data models are the data structures that are available to the end-users in ETL flow, to query and analyze the data.
- The ETL process ends up with loading data into the target Dimensional Data Models.
- Every dimensional data model is built with a fact table surrounded by multiple dimension tables.

Steps to be followed while designing a Dimensional Data Model:



Dimension Tables

- Dimension tables play a key role in the DW system by storing all the analyzed metric values.
- These values are stored under easily selectable dimensional attributes (columns) in the table.
- The quality of a DW system mostly depends on the depth of dimension attributes.

The structure of dimension table:

1) Dimension Table Key:

- Every Dimension table will have any one of its dimension attributes as a primary key to uniquely identify each row.
- Hence the distinct numeric values of that attribute can act as primary keys.
- If the attribute values are not unique in any case, then you can consider sequentially generated system numbers as the primary keys.
- These are also called as Surrogate keys.

- Dimensional data models must have the **referential integrity** constraint for each key between dimensions and facts.
- Thus Fact tables will have a foreign key reference for each primary/surrogate key in the dimension table to maintain referential integrity.
- If it is failed, then the respective fact table data cannot be retrieved for that dimension key.

2) Table Is Wide:

- We can say that dimension tables are wide as we can add any number of attributes to a dimension table at any point in the DW cycle.
- DW architect will request the ETL team to add respective new attributes to the schema.
- In real-time scenarios, we can see dimension tables with 50 (or) more attributes.

3) Textual Attributes:

- Dimensional attributes can be of any type as preferably text (or) numeric.
- Textual attributes will have real business words rather than codes.
- Dimension tables are not meant for calculations hence numeric values are rarely used for dimensional attributes.

4) Attributes May Not Be Directly Related:

- All the attributes in a dimension table may not be related to one another.

5) Not Normalized:

- Normalizing a dimension table brings more intermediary tables into the picture which is not efficient.
- Thus dimension tables are not normalized.
- Dimensional attributes can act as the source for constraints in queries and can also be displayed as labels in the reports.
- The queries will perform efficiently if we directly pick an attribute from the dimension table and refer directly to the respective fact table without touching any other intermediary tables.

6) Drilling Down and Rolling Up:

- Dimension attributes have the capability to drill down (or) roll up the data whenever needed.

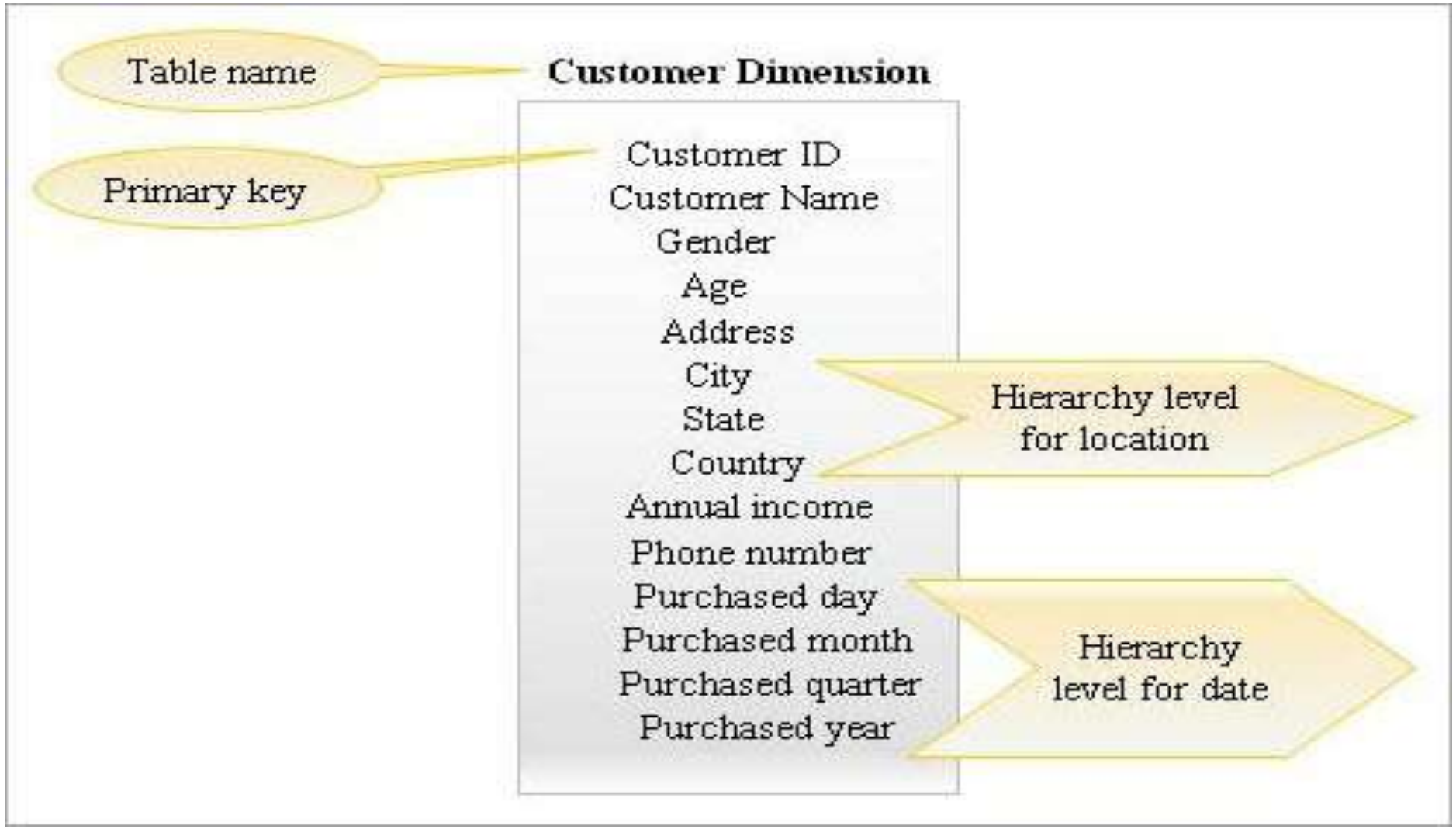
7) Multiple Hierarchies:

- A single dimension table having multiple hierarchies is very common.
- A dimension table will have a simple hierarchy if only one path exists from the bottom level to the top.
- Similarly, it will have multiple hierarchies if there are multiple paths present to reach from the bottom level to the top.

8) Few Records:

- Dimension tables will have less number of records (in hundreds) than the fact tables (in millions).
- Though they are smaller than the facts, they provide all the inputs to the fact tables.

Example of a Customer Dimension Table:



The Basic Load Plan For A Dimension:

- Dimensions can be created in two ways i.e.

1. By extracting the dimension data from external source systems (or)
2. The ETL system can build the dimensions from staging without involving any external sources.

- However, an ETL system without any external processing is more suitable to create dimension tables.

The steps involved in the process are:

1. **Data Cleaning:** Data is cleaned, validated and business rules are applied before loading into the dimension table to maintain consistency.
2. **Data Conforming:** Data from other parts of the data warehouse should be properly aggregated as a single value, with respect to each field of the dimension table.
3. **Share the same Domains:** Once the data is confirmed it is stored again in staging tables.
4. **Data Delivery:** Finally all the dimensional attribute values get loaded with primary / surrogate keys assigned.

Types Of Dimensions:

1) Small Dimensions
2) Conformed Dimension
3) Junk Dimension
4) Role-Playing Dimension
5) Degenerate Dimensions
6) Slowly Changing Dimensions

Fact Tables:

- Fact tables store a set of quantitatively measured values that are used for calculations.
- The fact table's values get displayed in the business reports.
- In contrast to the dimension tables textual data type, fact tables data type is significantly Numeric.
- Fact tables are deep whereas dimension tables are wide as fact tables will have a higher number of rows and a lesser number of columns.
- A primary key defined in the fact table is primarily to identify each row separately.
- The primary key is also called a Composite key in fact table.

- A single fact table can be surrounded by multiple dimension tables. With the help of the **foreign keys** that exist in fact tables, the respective context (verbose data) of the measured values can be referred to in the dimension tables.
- The lowest level of data that can be stored in a fact table is known as **Granularity**.
- The number of dimension tables associated with a fact table is inversely proportional to the granularity of that fact table data. i.e.
- The smallest measurement value needs more dimension tables to be referred.
- In a dimensional model, the fact tables maintain many-to-many relation with dimension tables.