AS PER RANI CHANNAMMA UNIVERSITY (NON CBCS), BELAGAVI – JULY 2021

BCA VI SEM

BUSINESS INTELLIGENCE

UNIT – IV
TYPES OF DATA MODELS, TECHNIQUES

PRESENTATION BY :
SANTOSH S.UMADI,
DEPARTMENT OF COMPUTER SCIENCE & BCA,
BLDEA's COMMERCE, BHS ARTS & TGP SCIENCE COLLEGE,
JAMKHANDI – 587301
Email id: santoshumadi@gmail.com
Blog: santoshumadi.blogspot.com

SYLLABUS

RANI CHANNAMMA UNIVERSITY, BELAGAVI

17BCAECOT63: Business Intelligence

Teaching Hours: 4 Hrs/week **Marks: Main Exam: 50**
IA: 20

Unit I: Business View of Information Technology applications: Business Enterprise Organization , Its functions, and core business process, baldrige business excellence frame work (Optional reading) Key purpose of using IT in business, The connected world : Characteristics of Internet _Ready IT Applications, Enterprise applications(ERP/CRM) and bespoke IT applications, information users and their requirements, Types of digital data , structured data , unstructured data, Semi-structured data , Difference between semi structured and structured data. 10 Hrs

Unit II: Introduction to OLTP and OLAP : OLTP(online transaction processing) OLAP(online Analytical Processing) Different OLAP Architectures , OLTP and OLAP, Data models for OLTP and OLAP, Role of OLAP tools in the BI Architecture , should OLAP be performed directly on operational data bases. Business intelligence: Using analytical information of decision support, Information sources before dawn of BI , BI defined , evolution of BI and role of DSS , EIS, MIS and digital dash boards, Need for BI at virtually all levels , BI for past , present and future, The BI value Chain , Introduction to Business analytics. 08 Hrs

Unit III:BI definitions and concepts : BI component Framework , BI Users, Business Intelligence Applications, BI roles and responsibilities, Basics of data integration , Need for data Warehouse ,Definition of data Warehouse, ODS, Ralph Kimball's Approach vs Inmon's Approach , Goals of data warehouse, Constituents of data Warehouse , Data integration, Data integration technologies , Data Quality , Data Profiling, A case Study from the Healthcare Domain. 10 Hrs

SYLLABUS

Unit IV:Types of Data Model: Data Modelling techniques, Fact table, Dimension table, Typical dimensional Models, Dimensional Modelling Life cycle, Understanding Measures and performance measurement System terminology , navigating a Business Enterprise. 10 Hrs

Unit V:Basics of Enterprise Reporting: Reporting perspectives common to all levels of Enterprise, Report Standardization and Presentation practices, Enterprise Reporting characteristics in OLAP World , Balanced score card , Dash boards. 10 Hrs

Text Books:

1. R.N.Prasad, Seema Acharya , Fundamentals of Business analytics, First Edition , 2011, Wiley-India

Reference Books:

1. GaliShmueli,. Nitin R Patel , peter C . Bruce, " Data mining for Business Intelligence" Wiley-India, 2011.
2. Ralph Kimball ,Margy Ross, "Practical tools for Data Warehosuing and Business Intelligence" , second Edition Wiley-India 2011.

DATA MODEL

- Just like a circuit diagram is to an electrical engineer,
- An assembly diagram is to a mechanical engineer, and
- A blueprint of a building is to a civil engineer
- So is the data models/data diagrams for a data architect

DATA MODEL

- Data model is a diagrammatic representation of the data & the relationship between its different entities.
- Although, time consuming, the process of creating data model is extremely important.
- It assists in identifying how the entities are related through a visual representation of their relationships and then helps reduce possible errors in the database design.

DATA MODEL

Basics of data modeling include,

- Entity – real world object
- Attribute – characteristic property of an entity
- Cardinality of a relationship – defines relationship between entities

CATEGORIES OF DATA MODEL

- Conceptual data model
- Logical data model
- Physical data model

DATA MODEL

Conceptual data model

- The conceptual data model is designed by identifying the various entities and the highest level relationships between them
- It identifies the most important entities
- It identifies relationships between different entities
- It does not support the specification of attributes
- It does not support the specification of the primary key

DATA MODEL

Logical data model

- The logical data model is used to describe data in as much detail as possible
- It identifies all entities and the relationships among them
- It identifies all the attributes for each entity
- It specifies the primary key for each entity
- It specifies the foreign keys

DATA MODEL

Physical data model

- It is a representation of how the model will be built in the database
- It will exhibit all the table structures, including column names, column datatypes, column constraints, primary key, foreign key, relationship between tables
- Specification of all tables – columns/ constraints are specified
- Foreign keys are specified

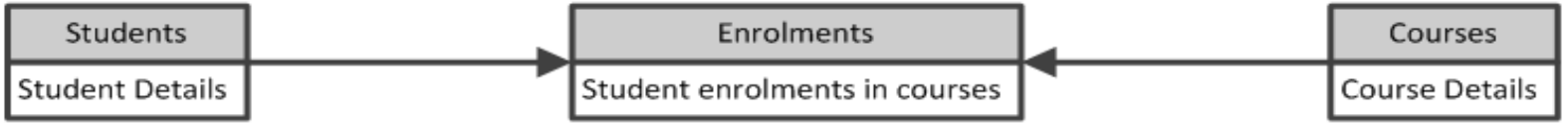
DATA MODEL

Physical data model

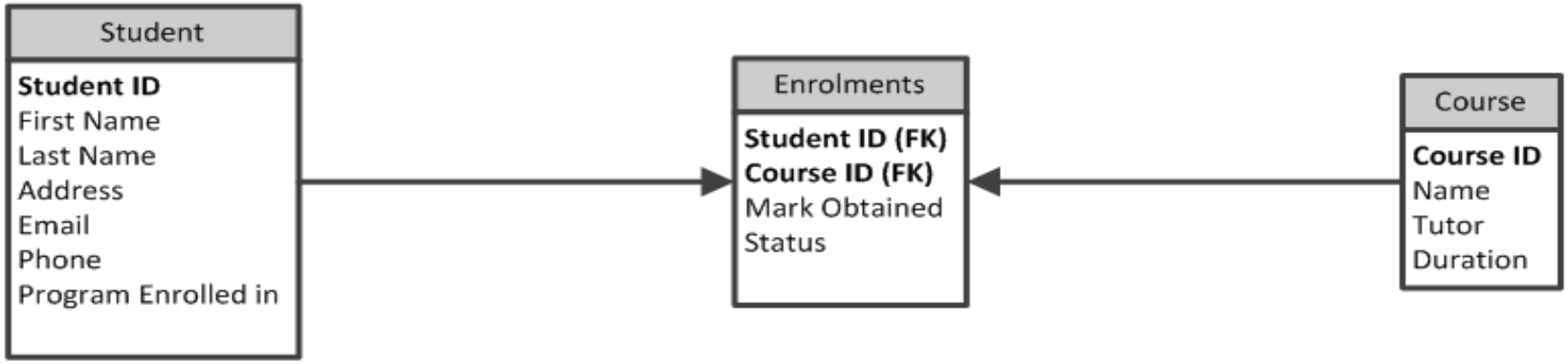
Steps for designing a physical model are

- Convert entities into tables/relation
- Convert relationships into foreign keys
- Convert attributes into columns/fields

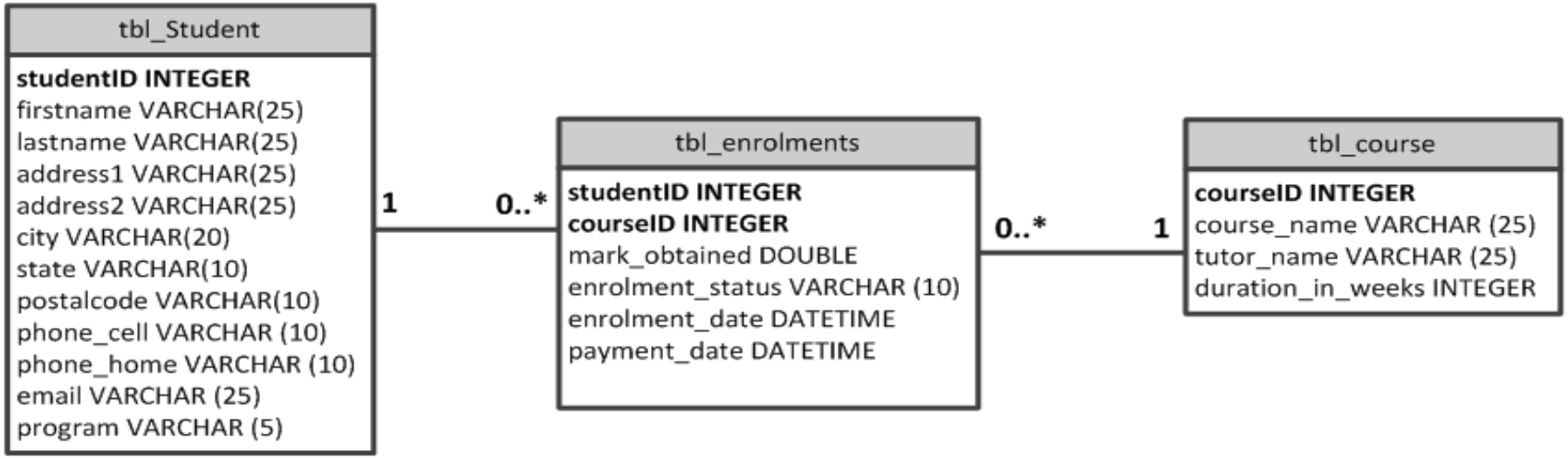
Conceptual Model



Logical Model



Physical Model



DATA MODELLING TECHNIQUES

Dimensional modeling

- Provides multi-dimensional perspectives of the same data

- Same data, multiple perspectives

DIMENSIONAL MODELING

Consider the scenario

- The company "xyz" sells its products in north, north-west & western regions of India.
- They have sales units at mumbai, pune,ahmedabad,delhi & punjab.
- The "president" of the company wants the latest information to measure the sales performance and to take corrective actions if required.
- He has requested this information from his business analysts.

DIMENSIONAL MODELING

SALES REPORT OF "XYZ" COMPANY

REPORT – 1 :

THE NUMBER OF UNITS SOLD : 113

- **Although the above data is correct, but it is not able to convey any useful information to the president as he cannot view the data from any perspective**

DIMENSIONAL MODELING

SALES REPORT OF "XYZ" COMPANY

REPORT – 2 :

THE NUMBER OF UNITS SOLD OVER TIME :

JANUARY	FEBRUARY	MARCH	APRIL
14	41	33	25

- **The above information conveys some information, but does not give a complete picture**

DIMENSIONAL MODELING

SALES REPORT OF "XYZ" COMPANY

REPORT – 3 :

THE NUMBER OF UNITS SOLD FOR EACH PRODUCT OVER TIME :

PRODUCT	JAN	FEB	MAR	APR
DIGITAL CAMERA			6	17
MOBILE PHONES	6	16	6	8
PEN DRIVES	8	25	21	

- The above information gives the information with product and time dimensions but not with the region dimension

DIMENSIONAL MODELING

SALES REPORT OF "XYZ" COMPANY

REPORT – 4 :

THE NUMBER OF UNITS SOLD IN EACH REGION FOR EACH PRODUCT OVER TIME :

REGION	PRODUCT	JAN	FEB	MAR	APR
MUMBAI	DIGITAL CAMERA			3	10
	MOBILE PHONES	3	16	6	
	PEN DRIVES	4	16	6	
PUNE	DIGITAL CAMERA			3	7
	MOBILE PHONES	3			8
	PEN DRIVES	4	9	15	

DIMENSIONAL MODELING

- **This method of analyzing a performance measure (number of units sold) by looking at it through various perspectives, this is known as "dimensional modeling"**

- **It allows the database to become more understandable and simpler**

- **Dimensional database can be viewed as a cube having three or more dimensional /perspectives for analyzing the given data**

DIMENSIONAL MODELING

- **Dimensional modeling divides the database into two parts:**

a) **Measurements : facts -> numeric values based on different dimensions**

b) **Context : dimensions -> describe "who, what, when, where, why & how" aspect of the measurements**

DIMENSIONAL MODELING

a) **MEASUREMENTS : FACTS -> number of tickets booked, amount paid .. etc**

b) **Context : dimensions -> customer details, time of booking, time of travel, mode of payment, origin city, destination city .. etc**

FACT TABLE

- **Fact table consists of various measurements**

- **A fact table consists of facts of a particular business process**

  ex: sales revenue by month by product

- **Facts are also known as measurements or metrics. A fact table record captures a measurement or a metric**

FACT TABLE

- The measures are numeric in nature
- They represent “how much”, “how many” aspect of a question
- For example : price, product sales, product inventory
- For example : mobile phone price is a fact, its descriptions are dimensions

FACT TABLE

- **Measures are the facts, the numbers of an event**
- **For example the total profit is a measure, the total sales too.**
- **The dimensions are the details that explains your fact**
- **The time is a dimension (when?), The location is a dimension (where?) And all attributes of a fact (who?, What? Product?)**

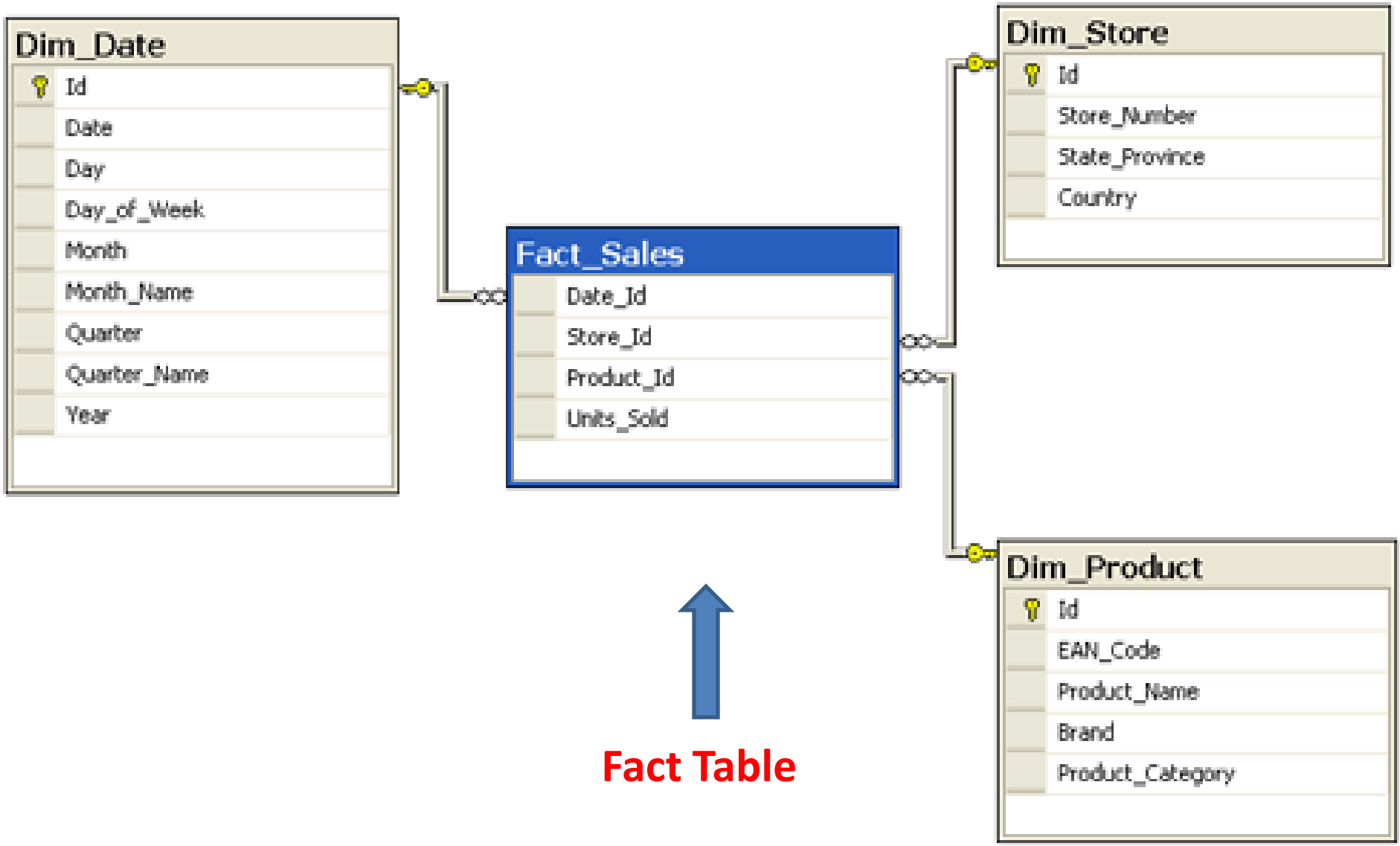
FACT TABLE

EXAMPLE

- We have a fact table fact_sales that has a measure which gives us a number of units sold by date, by store and by product

- All other tables such as dim_date, dim_store and dim_product are dimensions tables.

FACT TABLE


Dim_Date
Id
Date
Day
Day_of_Week
Month
Month_Name
Quarter
Quarter_Name
Year
Fact_Sales
Date_Id
Store_Id
Product_Id
Units_Sold
Dim_Store
Id
Store_Number
State_Province
Country
Dim_Product
Id
EAN_Code
Product_Name
Brand
Product_Category
Fact Table

TYPES OF FACTS

- **Additive facts : as its name implies additive facts are measures which can be added to all dimensions**

- **Semi additive facts : semi additive measures are the measure that can be added to only some dimensions and not across other**

- **Non additive facts : non-additive measures are measures that cannot be added any of the dimensions**

TYPES OF FACTS

ADDITIVE FACTS

TYPES OF FACTS

ADDITIVE FACTS

- The numeric value in a fact table that is more flexible is an additive measure
- For each dimension you can sum up
- If you want to know the total sales of your company you can easily sum up all the sales

TYPES OF FACTS

SEMI ADDITIVE FACTS

Incoming Goods 400

+

Outgoing Goods 100

Total Goods 500 ✗

Irrelevant to ADD

Total Goods 300 ✓

TYPES OF FACTS

SEMI ADDITIVE FACTS

- **Semi-additive facts are facts that can be summed up for some of the dimensions in the fact table, but not the others**

- **For example if you have the number of items in the warehouse for each day, you can sum up the items for each day (total warehouse of the day), but it make no sense to sum up in the year**

TYPES OF FACTS

NON ADDITIVE FACTS

TYPES OF FACTS

NON ADDITIVE FACTS

- WITH THESE FACTS YOU CAN NEVER MAKE A SUM. IF YOU ARE DEALING WITH THE DISCOUNT PERCENTAGE YOU EASILY UNDERSTAND WHAT I'M TALKING ABOUT

- IT'S A LITTLE CRAZY TO SAY "IN MY SHOP THE PERCENTAGE OF TOTAL DISCOUNT IS 1200%", NOBODY WOULD BELIEVE YOU AND WOULD DO WELL!

DIMENSION TABLE

- **Dimension table consists of dimension attributes which describe the dimension elements to enhance detail**

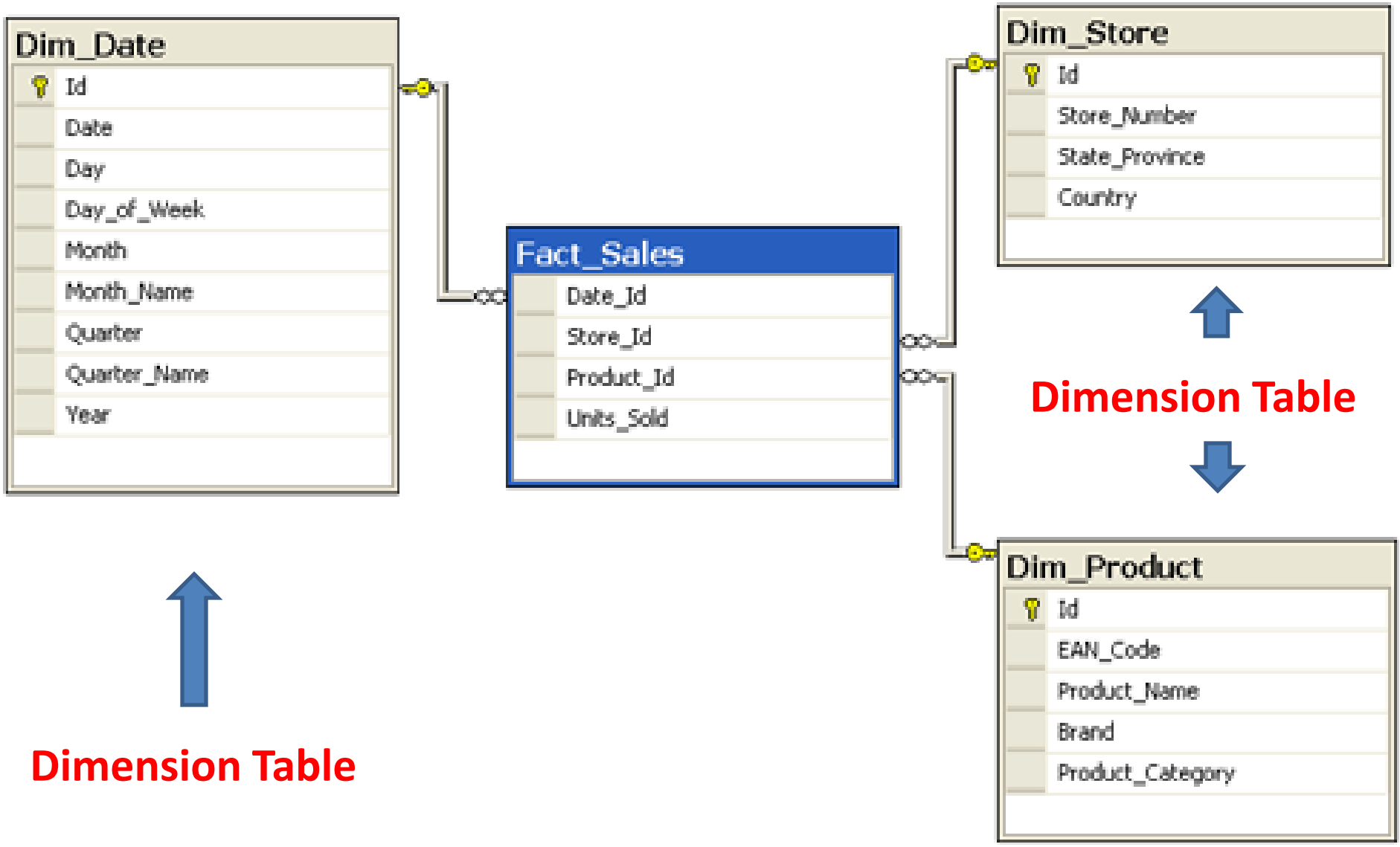
- **Dimension tables are used to describe dimensions; they contain dimension keys, values and attributes**

DIMENSION TABLE

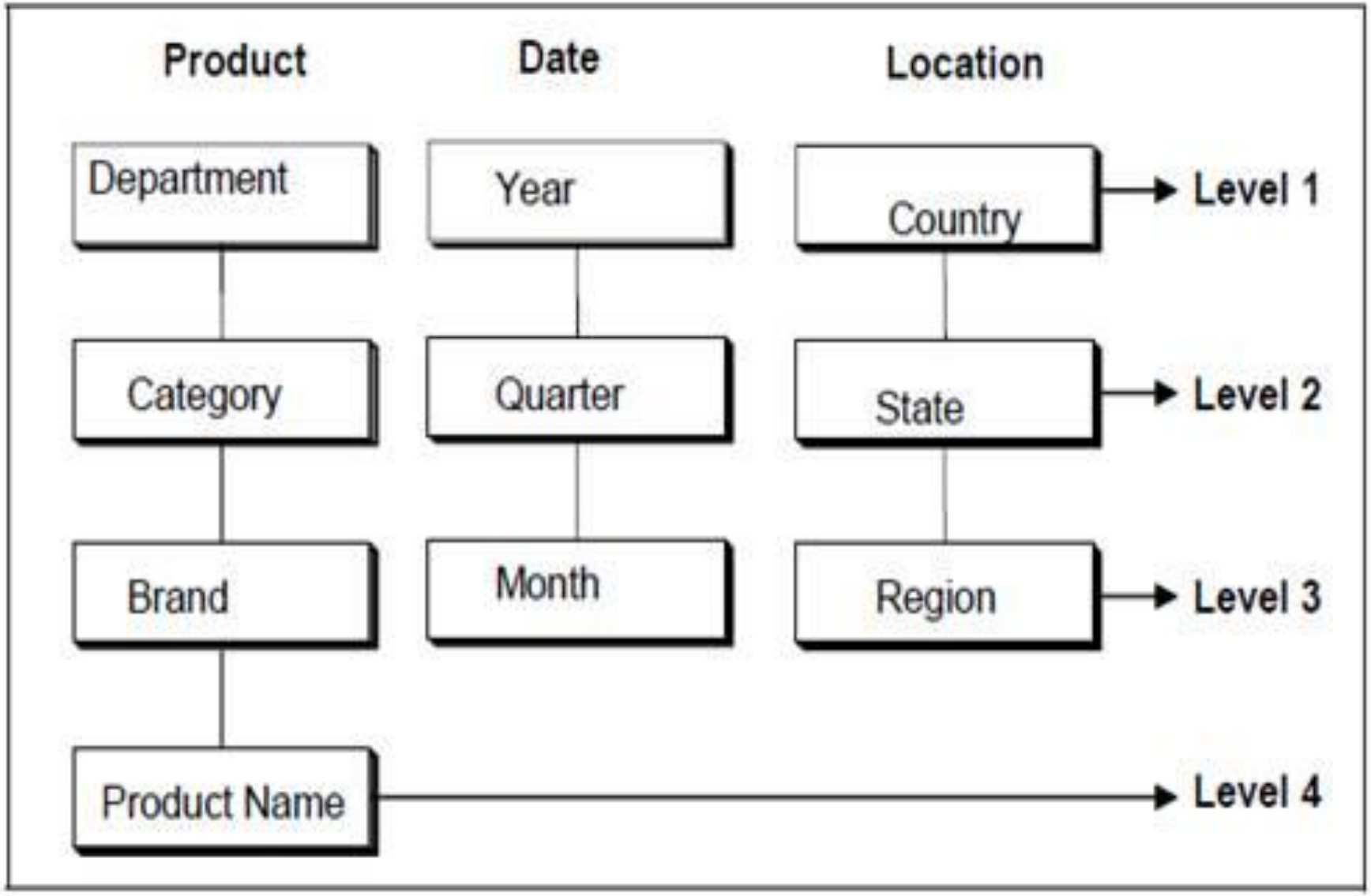
- For example, the time dimension would contain every hour, day, week, month, quarter and year that has occurred since you started your business operations

- Product dimension could contain a name and description of products you sell, their unit price, color, weight

DIMENSION TABLE



DIMENSION HIERARCHY



TYPES OF DIMENSION TABLES

- Slowly changing dimension
- Rapidly changing dimension
- Role playing dimension
- Degenerate dimension

TYPES OF DIMENSION TABLES

SLOWLY CHANGING DIMENSION

- **Dimension attributes that change slowly over a period of time rather than changing regularly is grouped as scd**

- **Attributes like name, address can change but not too often.**

TYPES OF DIMENSION TABLES

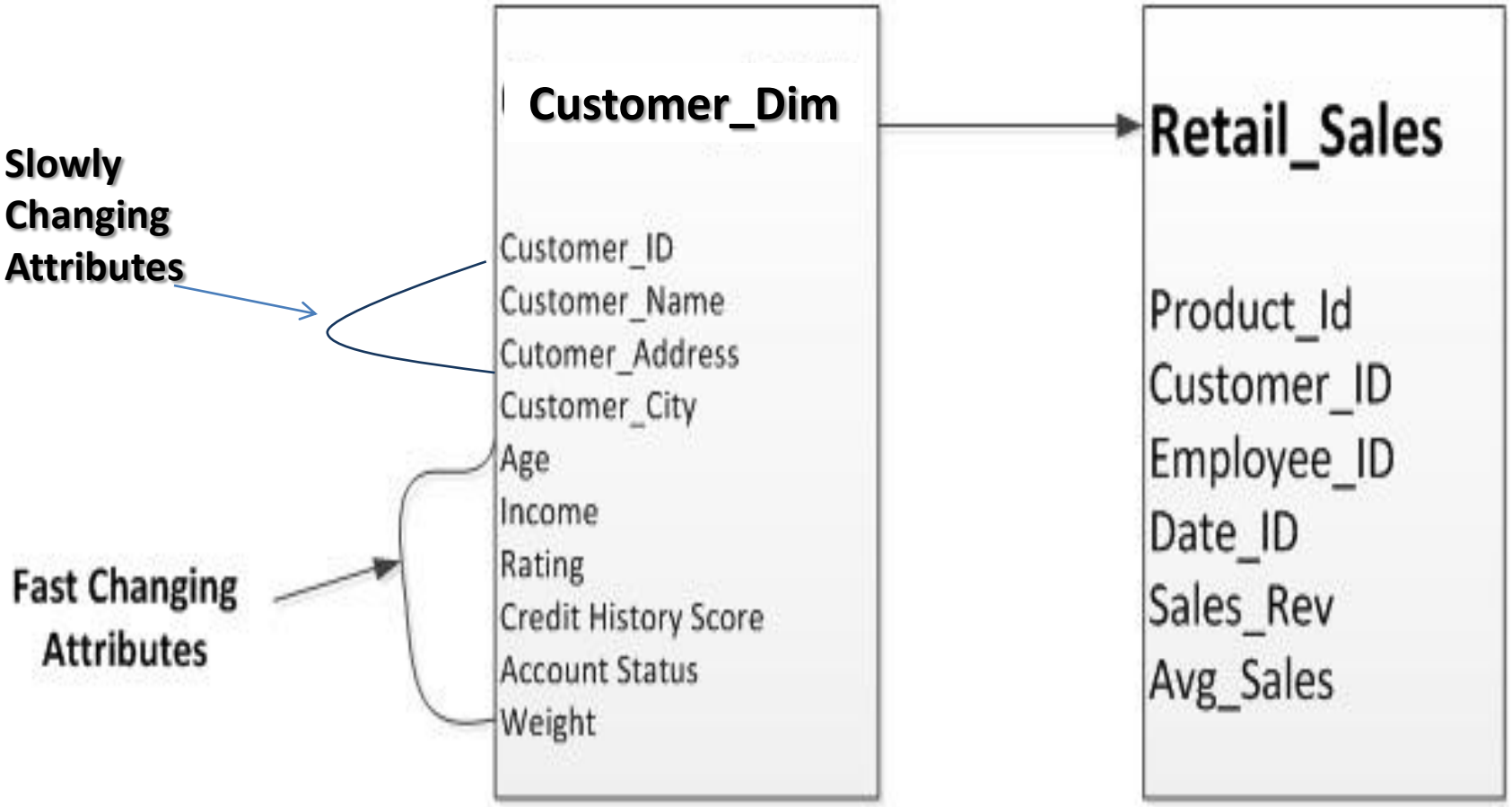
RAPIDLY CHANGING DIMENSION

- **Fast changing dimensions are dimensions that keep changing rapidly.**

- **Examples are customer weight, rating, income, bp .. these are rapidly changing dimension.**

TYPES OF DIMENSION TABLES

RAPIDLY CHANGING DIMENSION



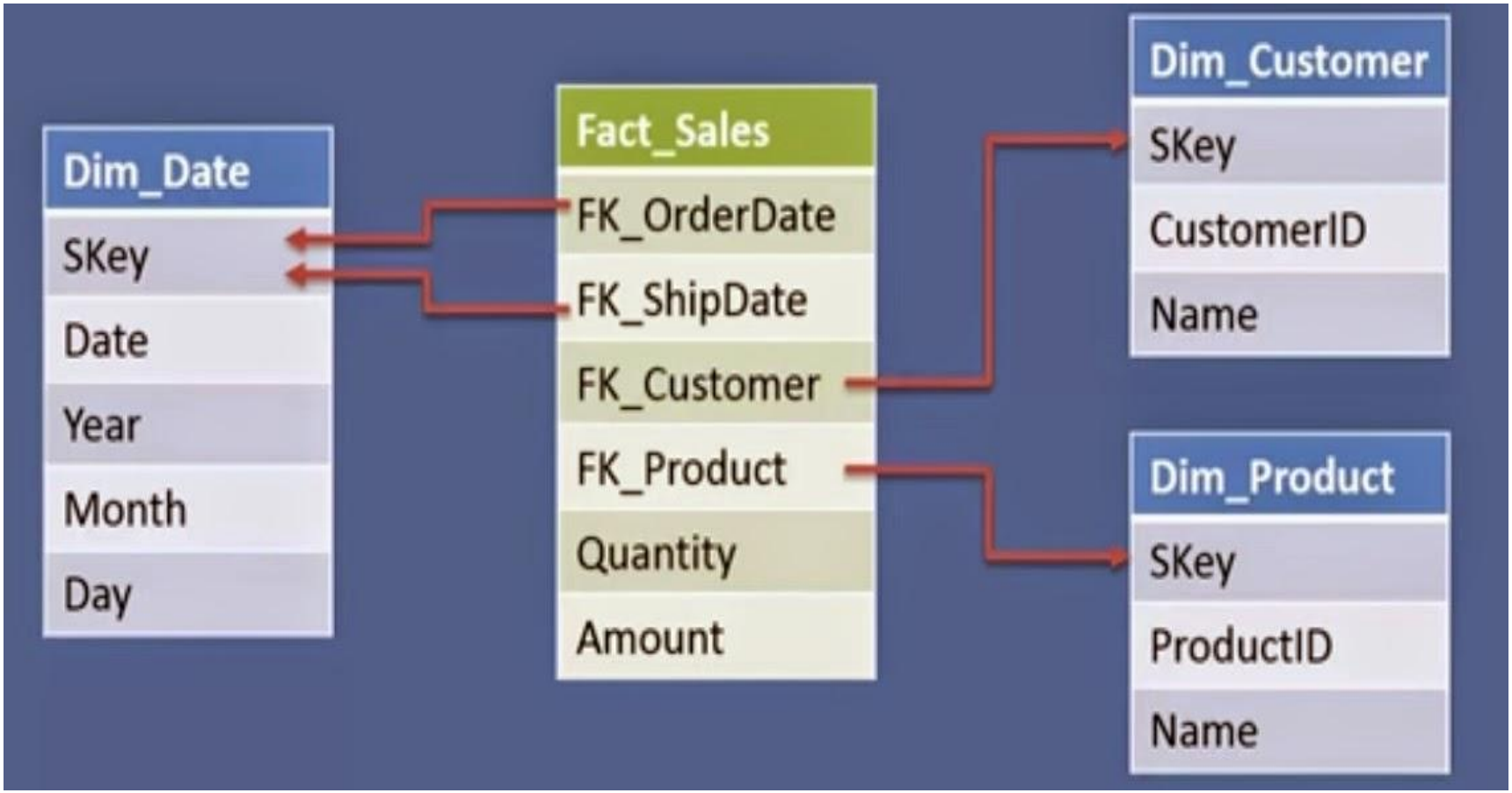
TYPES OF DIMENSION TABLES

ROLE PLAYING DIMENSION

- A single dimension that is expressed differently in a fact table.
- Consider an online transaction involving purchase of a laptop – the moment a order is placed, an order date and a delivery date will be generated. It is observed that both the dates are the attributes of the same time dimension.
- Hence the time dimension is called as role playing dimension

TYPES OF DIMENSION TABLES

ROLE PLAYING DIMENSION



TYPES OF DIMENSION TABLES

DEGENERATE DIMENSION

- A degenerate dimension is a data that is dimension in temperament but is present in a fact table.

- Data dimension that is stored in the fact table rather than a separate dimension table.

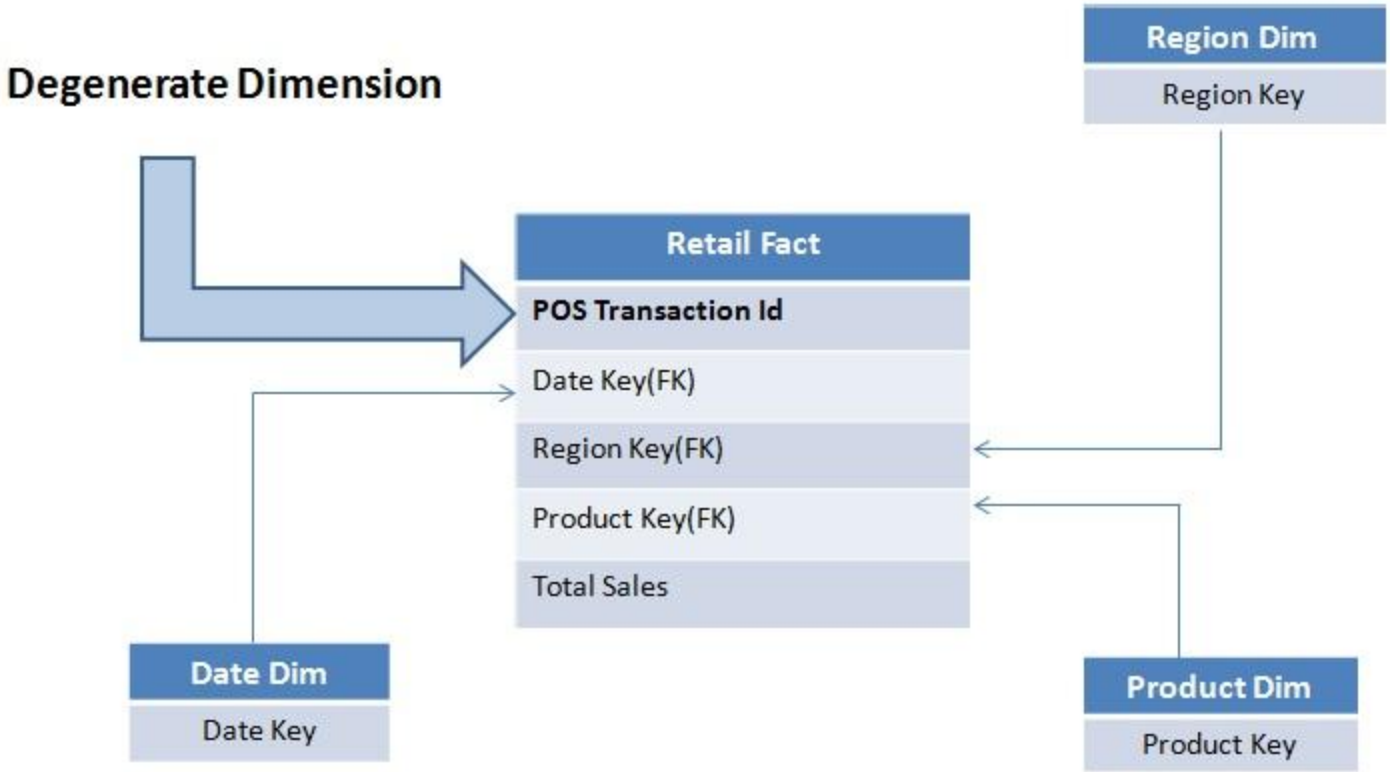
TYPES OF DIMENSION TABLES

DEGENERATE DIMENSION

- Examples of a degenerate dimension are -- customer number, sim card number, telephone number.
- These are fields, which are directly related to an event the fact table stores (eg. Retail store purchase, telecom connection sale, telecom call data record) - but is not eligible for being stored in a separate dimension table.

TYPES OF DIMENSION TABLES

DEGENERATE DIMENSION



MULTI DIMENSION MODELS (OLAP MODELS)

- **ER model is for OLTP, like that, models for OLAP are**
- **Star schema(model)**
- **Snowflake schema(model)**
- **Fact constellation schema(model)**

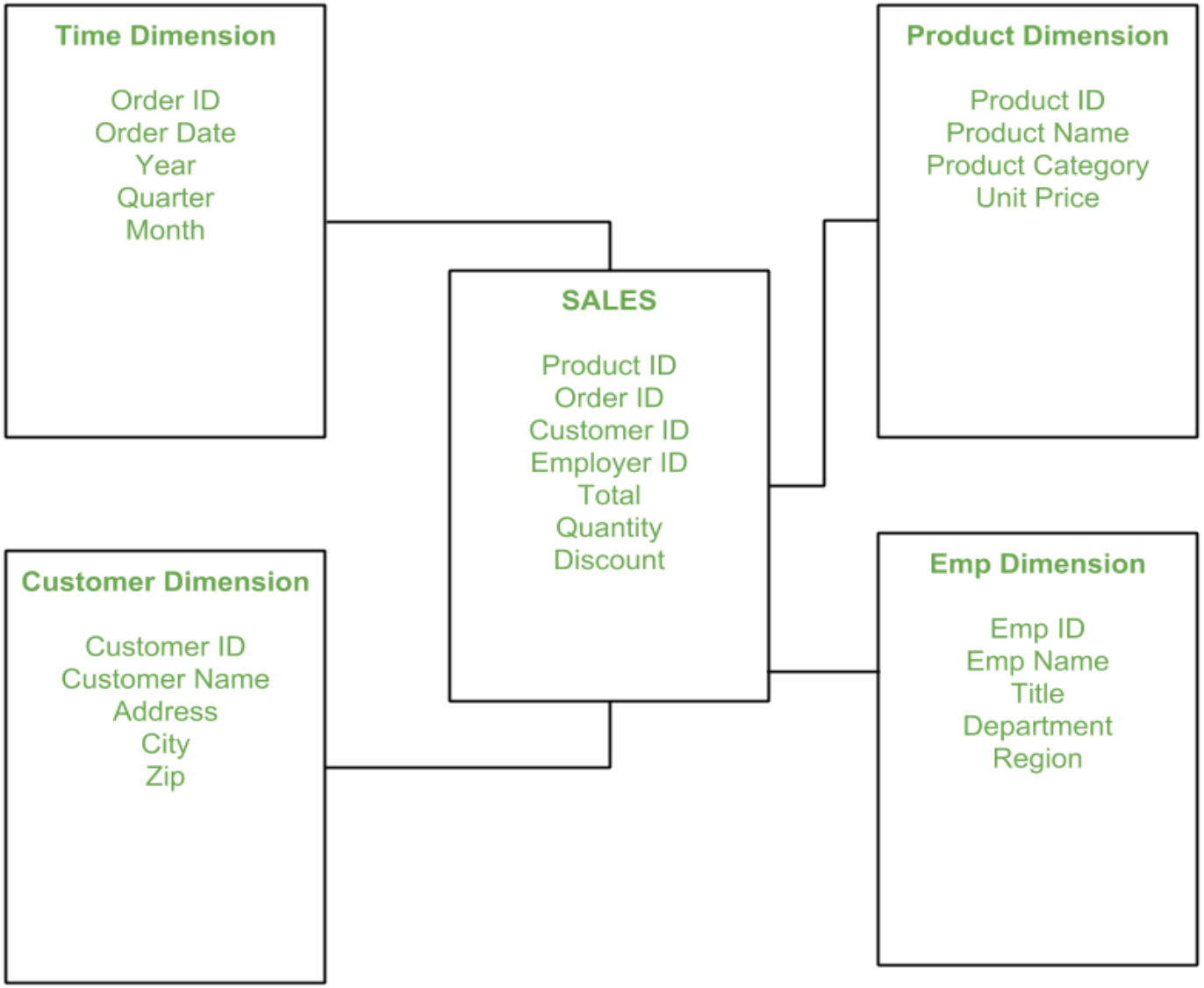
STAR SCHEMA (MODEL)

- In the star schema, the center of the star can have one fact table and a number of associated dimension tables
- It is known as star schema as its structure resembles a star
- The star schema is the simplest type of data warehouse schema

CHARACTERISTICS OF STAR SCHEMA (MODEL)

- Every dimension in a star schema is represented with the only one-dimension table.
- The dimension table should contain the set of attributes.
- The dimension table is joined to the fact table using a foreign key
- The dimension table are not joined to each other
- Fact table would contain key and measure
- The star schema is easy to understand.
- The dimension tables are not normalized.
- The schema is widely supported by bi tools.

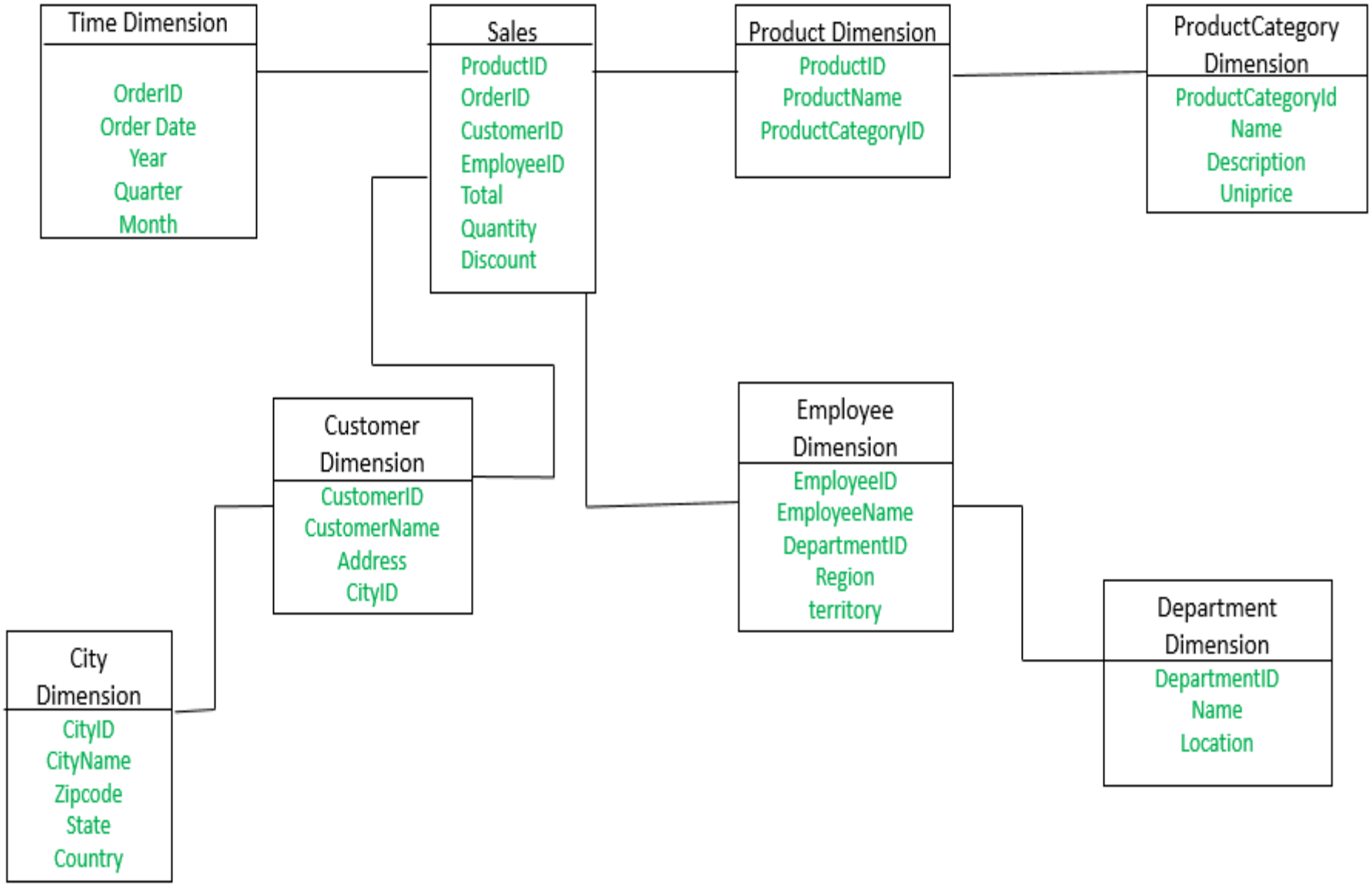
STAR SCHEMA (MODEL)



SNOWFLAKE SCHEMA (MODEL)

- **Snowflake schema is a logical arrangement of tables in a multidimensional database.**
- **A snowflake schema is an extension of a star schema, and it adds additional dimensions.**
- **The dimension tables are normalized which splits data into additional tables.**

SNOWFLAKE SCHEMA (MODEL)



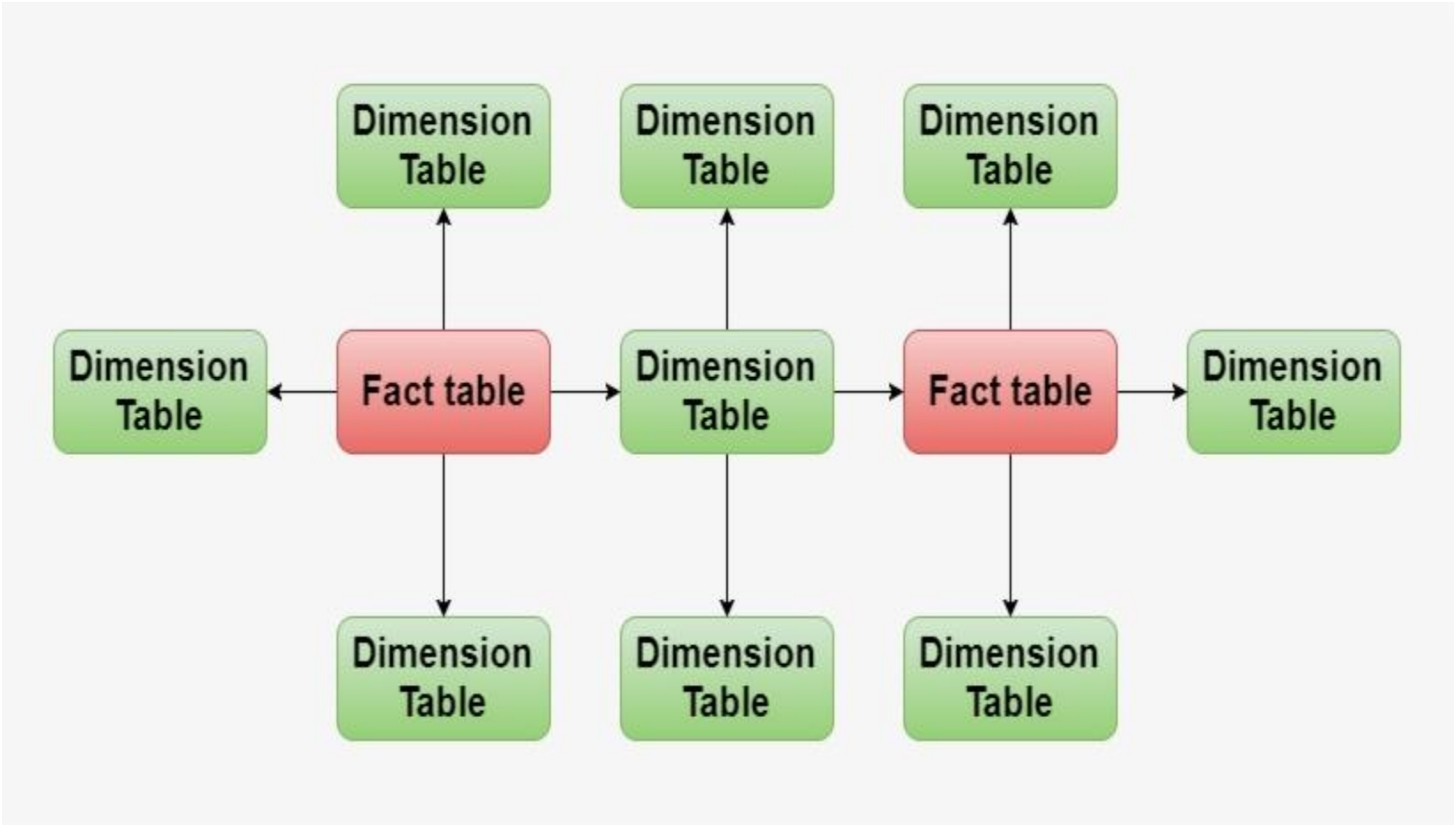
CHARACTERISTICS OF SNOWFLAKE SCHEMA (MODEL)

- The main benefit of the snowflake schema it uses smaller disk space.
- Easier to implement a dimension is added to the schema
- Due to multiple tables query performance is reduced
- The primary challenge that you will face while using the snowflake schema is that you need to perform more maintenance efforts because of the more lookup tables.

FACT CONSTELLATION SCHEMA (MODEL)

- **Fact constellation is a schema for representing multidimensional model.**
- **It is a collection of multiple fact tables having some common dimension tables.**
- **It can be viewed as a collection of several star schemas and hence, also known as galaxy schema.**
- **It is one of the widely used schema for data warehouse designing and it is much more complex than star and snowflake schema.**

FACT CONSTELLATION SCHEMA (MODEL)



DIMENSIONAL MODELING LIFE CYCLE

- **Requirements gathering – process of selecting the business process for which the dimensional model will be designed**
- **Identify the grain – grain refers to the level of detail to which the data can be analyzed. For ex (time : year, quarter, month, day, hour, minute, second)**
- **Identify the dimensions**
- **Identify the facts**
- **Design the dimensional model**

UNDERSTANDING MEASURES & PERFORMANCE

CONSIDER EXAMPLE OF HIGH SCHOOL SYSTEM

- How your report card is defined – Subjects, Minimum marks, Maximum marks, Score in each subject?
- How top 10 students are declared in every grade?
- How your school compares last year's results with current year to show progress to your parents?
- How does your school fare in comparison to other high schools in the neighborhood?
- How does your high school plan for increasing the intake capacity by adding more sections and new teachers?

MEASUREMENT SYSTEM TERMINOLOGY

- **Data : it is a collection of facts which have similar attributes or characteristics.**

  "phone number", "email id" .. etc.

- **Measure : data with associated unit of measure is termed as measure.**

  "lab hours per month" – has a numeric data associated with time duration
  "average wait time for bill payment" – is a measure derived out of multiple data points

MEASUREMENT SYSTEM TERMINOLOGY

- **Metric : it is a system of measures based on standard unit of measure with a business context.**

  "Product defect rate" – by city is an example of measuring "what percentage of goods was returned by customers in different cities"
  "employee attrition rate" – by quarter measures the percentage of employees leaving the company within each three month period.

MEASUREMENT SYSTEM TERMINOLOGY

- **Indicator : It is a business metric used to track business results or success / performance.**
  "call drop frequency" – for mobile users is an indicator of user dissatisfaction.
  "cost per unit shipped" – may indicate the rising cost of delivery of products.

- **Index : it consists of a composite set of indicators used to address the overall health of business operations.**

  "customer satisfaction index" – measured on a scale of 1 to 5

ATTRIBUTES OF GOOD METRIC

- **Name : simple,short & easy to remember**

  Ex : "Internal Assessment"
- **Abbreviation : short form used in the organization, ex : "IA"**
- **Description : complete explanation.**
- **Unit : measurement unit, ex: marks.**
- **Frequency : how often this metric will be measured, ex: semester wise.**
- **Priority : value of the metric in the system.**

ATTRIBUTES OF GOOD METRIC

- **Formula : evaluation system method.**
- **Weight : total weight of the metric.**
- **Perspective : depicts the area of use of metric.**
- **Revision history : list of all changes made to the metric definition.**
- **Date of approval : first implementation date.**
- **Indicators : use indicators for faster analysis**

SMART TEST FOR ENSURING METRIC RELEVANCE TO BUSINESS

- **Specific : metric is clearly defined & understood by all.**
- **Measurable : accurately, easily measurable someone.**
- **Attainable : must have a reachable target.**
- **Result oriented : metric must motivate team members performing the work.**
- **Time bound : should be traceable to the date/time when the actual value measurement was taken.**

END