**DATAWAREHOUSE CONCEPTS:**

**DATAWAREHOUSE (DWH):**

**DWH IS A RDBMS WHICH IS DESIGNED FOR QUERING AND ANALYZING THE BUSINESS AND ITS NOT FOR TRANSACTIONAL PROCESSING. USUALLY DWH CONTAINS HISTORICAL DATA DERIVED FROM TRANSACTIONAL DATA (FROM DIFFERENT HETEROGENEOUS TRANSACTIONAL SYSTEMS)**

**CHARACTERISTICS OF DWH:**

**SUBJECT ORIENTED**

**INTEGRATED**

**TIME VARIANT**

**NON-VOLATILE**

**SUBJECT ORIENTED:**

**DATAWAREHOUSE IS COLLECTION OF DIFFEERENT INDEPENDENT SUBJECT AREAS.**

**EX: SALES, ACCOUNT, MARKETING, FINANCE (INSURANCE, LOANS)…**

**INTEGRATED:**

**A DWH IS AN INTEGRATED DATABASE WHICH CONTAINS BUSINESS INFO COLLECTED FROM VARIOUS OPERATIONAL DATA SOURCES.**

**TIME VARIANT:**

**A DWH IS A TIME VARIANT DATABASE WHICH WILL ALLOW US TO ANALYZE AND COMPARE THE BUSINESS W.R.T VARIOUS TIME PERIODS (YEARLY, HALF YEARLY, QUARTERLY, MONTHLY, WEEKLY…)**

**NON-VOLATILE: (PERMANENT)**

**WHICH MEANS ONCE THE DATA ENTERED INTO DWH THAT CANNOT BE CHANGED. IT DOESN’T REFLECT TO THE CHANGES TAKEN PLACE IN OPERATIONAL SYSTEM. HERE DATA IS STATIC.**

**DATAMART:**

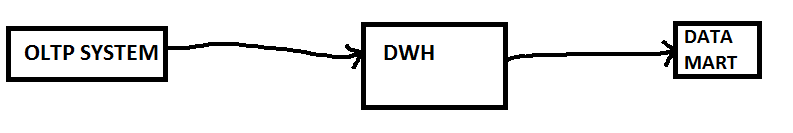
**DATA MART IS A SMALLER VERSION OF DWH.**

**DATA MART DEAL WITH A SINGLE SUBJECT.**

**DATA MART IS SUBSET OF DWH.**

**TYPES OF DATA MARTS:**

**DEPENDENT DATA MARTS – TOP – DOWN APPROACH**

****

**THE DATA IS EXTRACTED FROM OLTP SYSTEM AND THEN POPULATED IN THE CENTRAL DWH AND FROM THE DWH THE DATA IS MOVED TO DATA MART.**

**INDEPENDENT DATA MARTS – BOTTOM – UP APPROACH**

****

**THE DATA IS DIRECTLY LOADED FROM OLTP TO DATAMART. THIS IS SUITABLE FOR SMALL ORGANIZATION OR GROUPS WITHIN THE ORGANIZATION.**

**ODS – OPERATIONAL DATA STORE: THIS CONTAINS OPERATIONAL DATA.**

**DIMENSIONS**

**FACTS**

**SCHEMA**

**DIMENSIONS:**

**MASTER TABLES ARE CALLED AS DIMENSIONS.**

**DIMENSIONS – HAS MASTER DATA.**

**INFORMATION ABOUT THE BUSINESS.**

**DIMENSIONS ARE TABLES WHICH ARE USED TO STORE TEXTUAL INFORMATION ABOUT A BUSINESS WHICH HAS PRIMARY KEY.**

**DIMENSION COLUMNS ARE CALLED AS ATTRIBUTES OR DIMENSION ATTRIBUTES.**

**DIMENSION TABLES ALWAYS DEAL WITH CHAR, VARCHAR(), NCHAR AND NVARCHAR()**

**FACT:**

**FACT TABLE DEAL WITH DATA RELATED TO VALUES WHICH CAN BE CALCULATED.**

**EX: SALES AMOUNT, TAX, PRICE, SALARY, AGE, WEIGHT, HEIGHT, QUANTITY,..**

**FACT TABLE HAS COLUMNS – THOSE ARE CALLED AS MEASURES.**

**FACT TABLE IS ALSO CALLED AS MEASURE GROUP.**

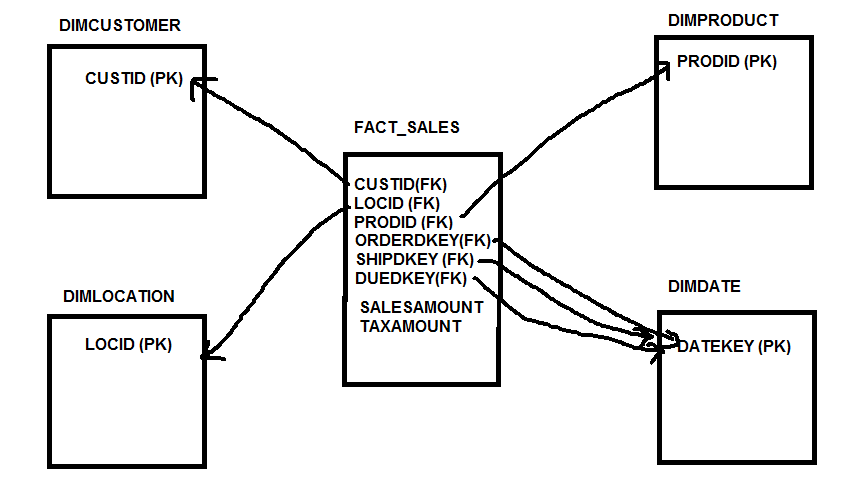
**FACT TABLE HAS FOREIGN KEY RELATED TO THE DIMENSION TABLE.**

**SCHEMAS:**

**PICTORIAL REPRESENTATION OF OUR REQUIREMENTS.**

**HOW DIMENSIONS AND FACTS ARE ARRANGED/RELATED/ORGANIZED.**

1. **STAR SCHEMA**
2. **SNOW FLAKE SCHEMA**
3. **STAR SCHEMA:**

****

**DIMENSIONS HAS PRIMARY KEY AND FACT TABLE HAS FOREIGN KEY.**

**FACT TABLE IS DIRECTLY RELATED TO ALL THE DIMENSION TABLES.**

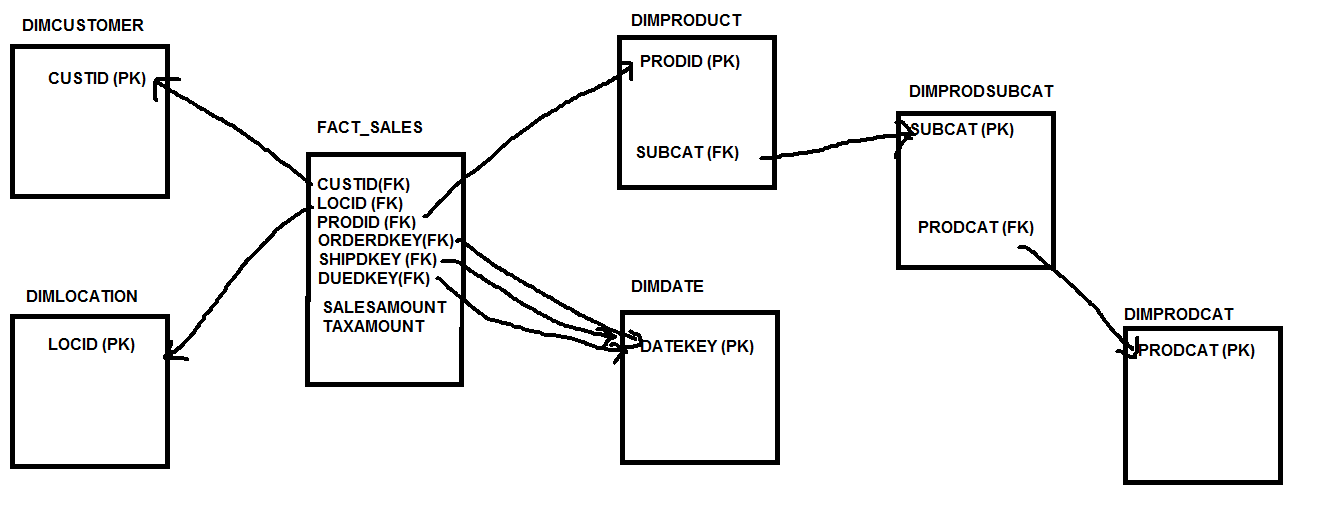
**STAR SCHEMA IS NOTHING BUT FACT TABLE IS AT CENTER AND IS RELATED TO ALL THE DIMENSIONS.**

**HERE, DIMDATE DIMENSION IS PLAYING DIFFERENT ROLES AS ORDERDATE, DUEDATE AND SHIPDATE SO THIS IS CALLED AS ROLE PLAYING DIMENSION.**

**SNOWFLAKE SCHEMA:**

**WHICH IS NOT STAR SCHEMA IS CALLED AS SNOWFLAKE SCHEMA.**

**HERE FACT TABLE IS NOT DIRECTLY RELATED TO ALL THE DIMENSION TABLES. IT MAY DIRECTLY RELATE TO FEW DIMENSIONS AND NOT TO FEW DIMENSIONS.**

****

**------**

**TYPES OF DIMENSIONS:**

1. **JUNK DIMENSION**

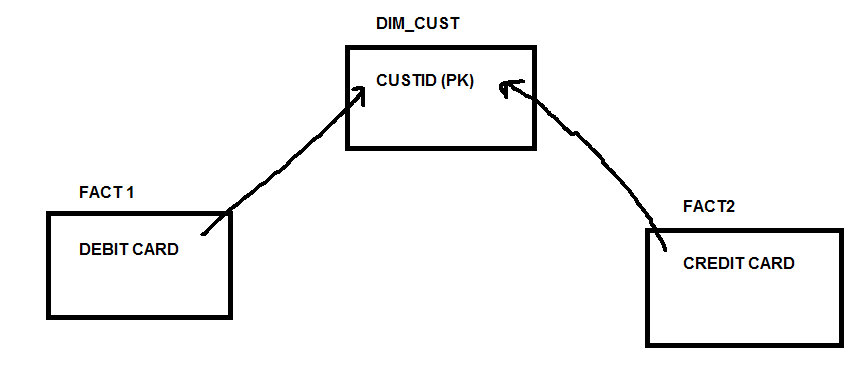
**IT I A GROUP OF FLAGS WHICH GIVE TRUE OR FALSE, YES OR NO, 0 OR 1, ACTIVE OR INACTIVE.**

**ATTRIBUTES IN JUNK DIMENSION DO NOT BELONG TO FACT TABLE.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ATM TRANSACTION** | |  |  |  |  |
| **TRANSID** | **ISTRANSACTION SUCCEFUL** | **ISTRANSACTIONFAILED** | **NO BALANCE** | **INVALID PIN** | **NO MONEY** |
| 1 | 1 | 0 | 0 | 0 | 0 |
| 2 | 0 | 1 | 1 | 0 | 0 |
| 3 | 0 | 1 | 0 | 1 | 0 |
| 4 | 0 | 1 | 0 | 0 | 1 |

1. **CONFORMED DIMENSION:**

**THE DIMENSION WHICH IS SHARED BY TWO OR MORE FACT TABLE IS CALLED AS CONFORMED DIMENSION.**

****

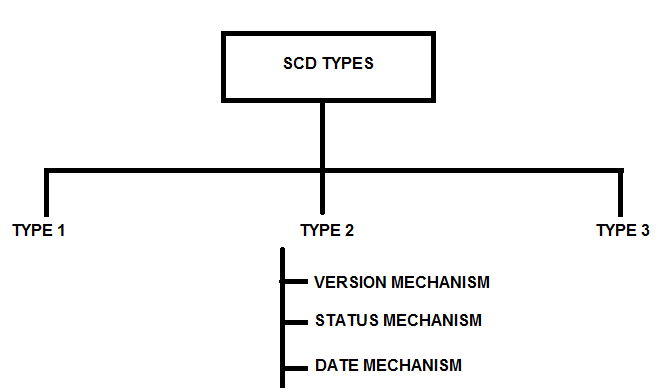
1. **SLOWLY CHANGING DIMENSIONS (SCD):**

**SCD IS USED TO STORE THE HISTORY IN DIFFERENT TYPES:**

**SCD TYPE1**

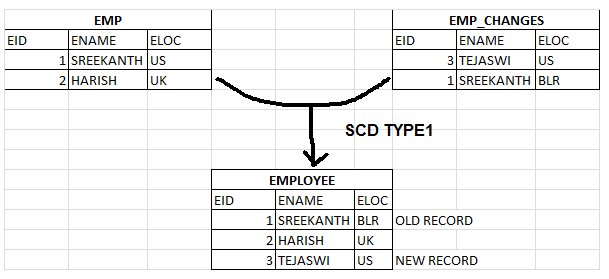
**SCD TYPE2**

**SCD TYPE3**

****

**TYPE1:**

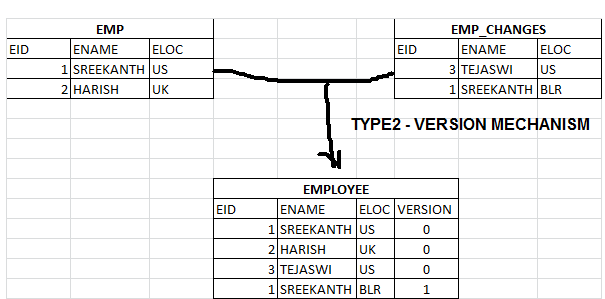
1. **NEW RECORD IS INSERTED.**
2. **OLD RECORD IS UPDATED.**
3. **HERE ONLY ONE RECORD IS THERE FOR EVERY UNIQUE ID – HISTORY IS NOT MAINTAINED.**

****

**TYPE2:**

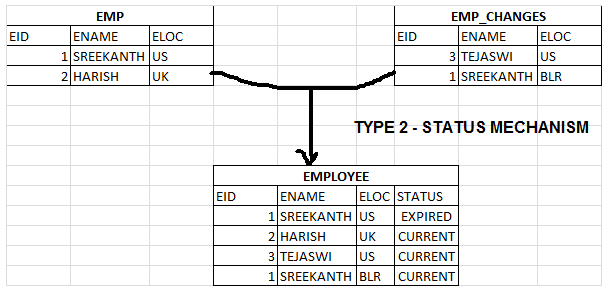
**VERSION MECHANISM:**

1. **NEW RECORD INSERTED WITH VERION ‘0’**
2. **OLD RECORD UPDATED WITH VERSION INCREMENTAL.**
3. **HISTORY IS MAINTAINED WITH VERSIONS, LATEST OR CURRENT IS IDENTIFIED WITH HIGHEST VERSION NUMBER.**

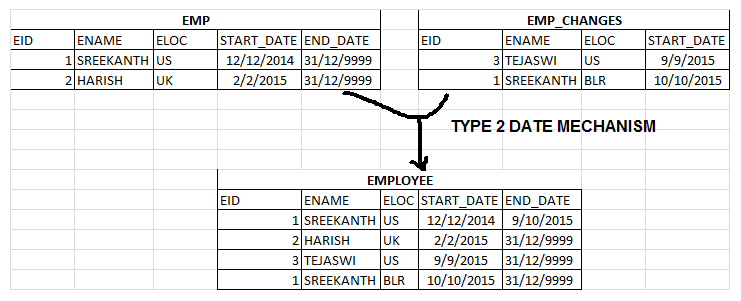
****

**STATUS MECHANISM:**

**NEW RECORD INSERTS WITH STATUS “CURRENT” AND OLD RECORD IS UPDATED WITH STATUS “EXPIRED”.**

****

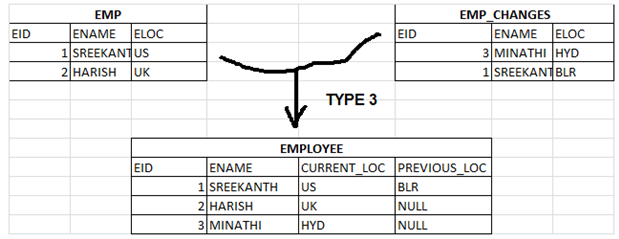
**DATE MECHANISM:**

****

**SCD TYPE 3:**

**PARTIAL HISTORY MAINTENANCE:**

**HERE CURRENT/PREVIOUS RECORDS ARE MAINTAINED.**

****

ROLE PLAYING DIMENSION:

A DIMENSION IS BEING REFFERED BY MANY FACT TABLES THEN THAT DIMENSION IS CALLED AS ROLE PLAYING DIMENSION.

Ex: Dimdate dimension has DateKey primary key column which referred by OrderDateKey, DueDateKey and ShipDateKey as foreign keys , in this case DimDate is a role playing dimension.