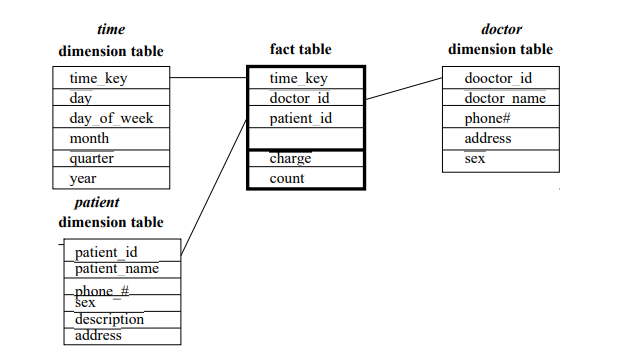
**SET-1**

1. Suppose that a data warehouse consists of the three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit.

* Draw a star schema diagram for the above data warehouse
* Starting with the base cuboid [day, doctor, patient], what specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2004?
* To obtain the same list, write an SQL query assuming the data is stored in a relational database with the schema fee (day, month, year, doctor, hospital, patient, count, charge).

Ans:

* Draw a star schema diagram for the above data warehouse

CREATE TABLE TTIME(TIME\_KEY DATETIME PRIMARY KEY,DAY VARCHAR(10),DAY\_OF\_THE\_WEEK VARCHAR(10),MONTH\_NAME VARCHAR(10),QUARTER INT,TYEAR INT);

CREATE TABLE DOCTOR(DOCTOR\_ID INT PRIMARY KEY,DOCTOR\_NAME VARCHAR(20),PHONE NUMBER(10), ADDRESS VARCHAR(50), SEX VARCHAR(10));

CREATE TABLE PATIENT(PATIENT\_ID INT PRIMARY KEY, PATIENT\_NAME VARCHAR(20), PHONE NUMBER(10), SEX VARCHAR(10), DESCRIPTION VARCHAR(100), ADDRESS VARCHAR(50));

INSERT INTO TIME VALUES('2017-07-23 13:10:11', 22, 'MON', 'JAN','3','2014');

INSERT INTO TIME VALUES('2017-07-23 13:10:11', 22, 'MON', 'JAN','3','2014');

INSERT INTO TIME VALUES('2017-07-23 13:10:11', 22, 'MON', 'JAN','3','2014');

INSERT INTO TIME VALUES('2017-07-23 13:10:11', 22, 'MON', 'JAN','3','2014');

INSERT INTO TIME VALUES('2017-07-23 13:10:11', 22, 'MON', 'JAN','3','2014');

INSERT INTO DOCTOR VALUES (001, 'ABC', '9876654321', 'ELURU', 'MALE');

INSERT INTO DOCTOR VALUES (002, 'BCD', '9876654123', 'GUNTUR', 'FEMALE');

INSERT INTO DOCTOR VALUES (003, 'CDE', '9876612345', 'GUNTUR', 'MALE');

INSERT INTO DOCTOR VALUES (004, 'EFG', '9876654213', 'VIJAYAWADA', 'FEMALE');

INSERT INTO DOCTOR VALUES (005, 'GHI', '9876656451', 'ELURU', 'MALE');

INSERT INTO PATIENT VALUES (001, 'BDG', 9876543210, 'MALE', 'FEVER', 'ELURU');

INSERT INTO PATIENT VALUES (002, 'GHS', 8765432109, 'MALE', 'COLD', 'GUNTUR');

INSERT INTO PATIENT VALUES (003, 'SDF', 7654321098, 'MALE', 'ACCIDENT', 'ELURU');

INSERT INTO PATIENT VALUES (004, 'GSF', 6543210987, 'MALE', 'TYPHOID', 'GUNTUR');

INSERT INTO PATIENT VALUES (005, 'VXD', 5432109876, 'MALE', 'DENGUE', 'GUNTUR');

CREATE VIEW FACTT AS SELECT DOCTOR.DOCTOR\_ID, PATIENT.PATIENT\_ID, TIME.TIME\_KEY, FROM DOCTOR, PATIENT, TIME;

DESC FACTT;

SELECT \* FROM FACTT;

* Starting with the base cuboid [day, doctor, patient], what specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2004?

First, we should use roll-up operation to get the year 2004(rolling-up from day then month to year). After getting that, we need to use slice operation to select (2004). Second, we should use roll-up operation again to get all patients. Then, we need to use slice operation to select (all). Finally, we get list the total fee collected by each doctor in 2004.

So,

1. ROLL UP FROM DAY TO MONTH TO YEAR

2. SLICE FOR YEAR = “2004”

3. ROLL UP ON PATIENT FROM INDIVIDUAL PATIENT TO ALL

4. SLICE FOR PATIENT = “ALL”

4. GET THE LIST OF TOTAL FEE COLLECTED BY EACH DOCTOR IN 2004

c. Ans

SELECT DOCTOR, SUM(CHARGE) FROM FEE WHERE YEAR = 2004 GROUP BY DOCTOR

1. Implementation of Text Mining on the data warehouse using weather dataset.

**SET-2**

1. Suppose that a data warehouse consists of the four dimensions, date, spectator, location, and game, and the two measures, count and charge, where charge is the fare that a spectator pays when watching a game on a given date. Spectators may be students, adults, or seniors, with each category having its own charge rate.

* Draw a star schema diagram for the data warehouse.
* Starting with the base cuboid [date, spectator, location, game], what specific OLAP operations should one perform in order to list the total charge paid by student spectators at GM Place in 2011?

***Create database, and tables in relevance to the above data warehouse***

CREATE TABLE DDATE(DATE\_ID DATE PRIMARY KEY, DAY VARCHAR(10), M0NTH VARCHAR(10), QUARTER VARCHAR(10), YEAR NUMBER(4));

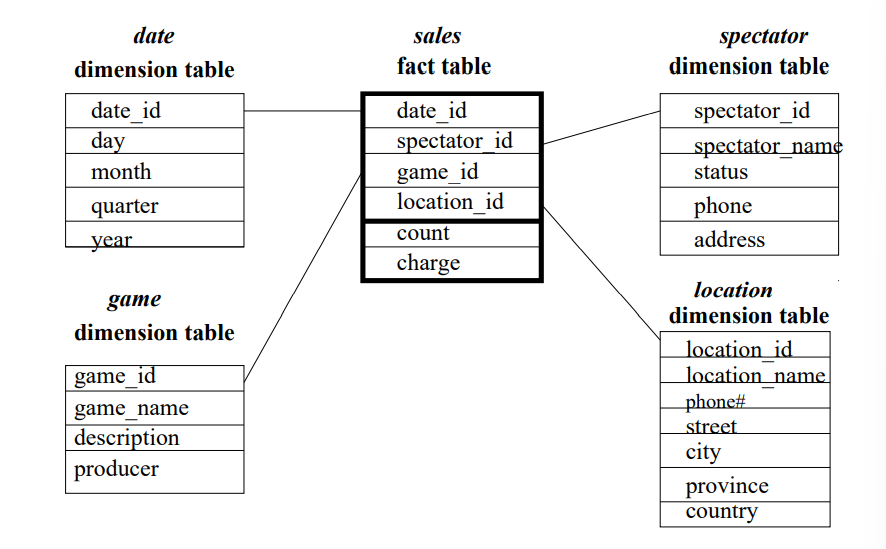
CREATE TABLE GAME(GAME\_ID INT PRIMARY KEY, GAME\_NAME VARCHAR(20),DESCRIPTION VARCHAR(50), PRODUCER VARCHAR(20));

CREATE TABLE SPECTATOR(GAME\_ID INT PRIMARY KEY, SPECTATOR\_NAME VARCHAR(20), STATUS VARCHAR(10),PHONE NUMBER(10),ADDRESS VARCHAR(50));

CREATE TABLE LOCATION(LOCATION\_ID INT PRIMARY KEY, LOCATION\_NAME VARCHAR(20), PHONE NUMBER(10), STREET VARCHAR(20), CITY VARCHAR(20), PROVINCE VARCHAR(20), COUNTRY VARCHAR(20));

Ans:

Draw a star schema diagram for the data warehouse



CREATE VIEW FACTT AS SELECT DDATE.DATE\_ID, GAME.GAME\_ID,SPECTATOR.SPECTATOR\_ID, LOCATION.LOCATION\_ID FROM DDATE, GAME, SPECTATOR, LOCATION;

DESC FACTT;

INSERT INTO DDATE VALUES('2019-05-21', 'MONDAY', 'JUNE', 'FIRST', 2017);

INSERT INTO DDATE VALUES('2019-05-21', 'MONDAY', 'JUNE', 'FIRST', 2017);

INSERT INTO DDATE VALUES('2019-05-21', 'MONDAY', 'JUNE', 'FIRST', 2017);

INSERT INTO GAME VALUES(001, 'PUBG', 'ACTION GAME', 'TECHO');

INSERT INTO GAME VALUES(001, 'PUBG', 'ACTION GAME', 'TECHO');

INSERT INTO GAME VALUES(001, 'PUBG', 'ACTION GAME', 'TECHO');

INSERT INTO SPECTATOR VALUES(001, 'ABC', 'ONLINE', 9876543210, 'GUNTUR');

INSERT INTO SPECTATOR VALUES(001, 'ABC', 'ONLINE', 9876543210, 'GUNTUR');

INSERT INTO SPECTATOR VALUES(001, 'ABC', 'ONLINE', 9876543210, 'GUNTUR');

INSERT INTO LOCATION VALUES(001, 'ABC', 9876543210, 'KOTHAPET', 'GUNTUR', 'SFDS', 'INDIA');

INSERT INTO LOCATION VALUES(001, 'ABC', 9876543210, 'KOTHAPET', 'GUNTUR', 'SFDS', 'INDIA');

INSERT INTO LOCATION VALUES(001, 'ABC', 9876543210, 'KOTHAPET', 'GUNTUR', 'SFDS', 'INDIA');

Starting with the base cuboid [date, spectator, location, game], what specific OLAP operations should one perform in order to list the total charge paid by student spectators at GM Place in 2011?

 ROLL-UP ON DATE FROM DATE ID TO YEAR.

2. ROLL-UP ON SPECTATOR FROM SPECTATOR ID TO STATUS.

3. ROLL-UP ON LOCATION FROM LOCATION ID TO LOCATION NAME.

4. ROLL-UP ON THE GAME FROM GAME ID TO ALL.

5. DICE WITH STATUS= “STUDENTS”, LOCATION NAME= “GM PLACE”, AND YEAR=2011

1. Implementation of Text Mining on the data warehouse using labor dataset.

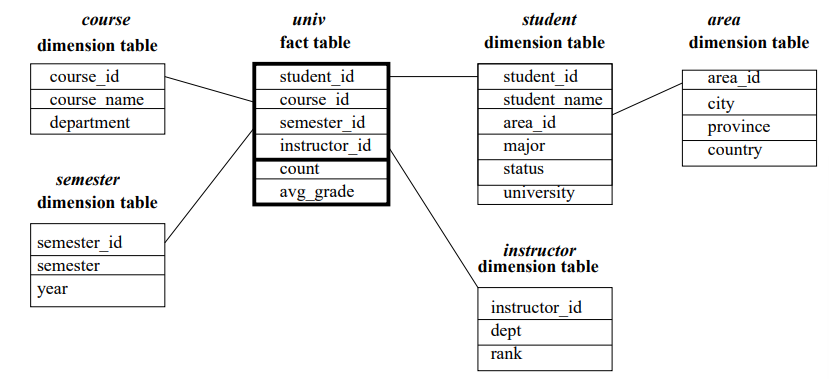
**SET-3**

1. Suppose that a data warehouse for Big-University consists of the following four dimensions: student, course, semester, and instructor, and two measures count and avg grade. When at the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the avg grade measure stores the actual course grade of the student. At higher conceptual levels, avg grade stores the average grade for the given combination.

* Draw a snowflake schema diagram for the data warehouse.
* Starting with the base cuboid [student, course, semester, instructor], what specific OLAP operations (e.g., roll-up from semester to year ) should one perform in order to list the average grade of CS courses for each Big-University student.
* If each dimension has five levels (including all), such as “student < major < status < university < all”, how many cuboids will this cube contain (including the base and apex cuboids)?

Answers:

1. Draw a snowflake schema diagram for the data warehouse.



CREATE TABLE COURSE(COURSE\_ID INT, COURSE\_NAME VARCHAR(20), DEPARTMENT VARCHAR(20));

CREATE TABLE SEMESTER(SEMESTER\_ID INT, YEAR NUMBER(4));

CREATE TABLE STUDENT(STUDENT\_ID INT PRIMARY KEY, AREA\_ID INT REFERENCES AREA, MAJOR VARCHAR(20), STATUS VARCHAR(20), UNIVERSITY VARCHAR(20));

CREATE TABLE AREA(AREA\_ID INT PRIMARY KEY, CITY VARCHAR(20), PROVINCE VARCHAR(20), COUNTRY VARCHAR(20));

CREATE TABLE INSTRUCTOR (INSTRUCTOR\_ID INT, DEPT VARCHAR(20), RANK NUMBER(3));

CREATE VIEW UNIVERSITY AS SELECT STUDENT.STUDENT\_ID, COURSE.COURSE\_ID, SEMESTER.SEMESTER\_ID, INSTRUCTOR.INSTRUCTOR\_ID FROM STUDENT, COURSE, SEMESTER, INSTRUCTOR;

DESC UNIVERSITY;

1. Starting with the base cuboid [student, course, semester, instructor], what specific OLAP operations (e.g., roll-up from semester to year ) should one perform in order to list the average grade of CS courses for each Big-University student.

* ROLL-UP ON COURSE FROM COURSE\_ID TO DEPARTMENT
* ROLL-UP ON SEMESTER FROM SEMESTER\_ID TO ALL
* SLICE FOR COURSE = “CS”

1. If each dimension has five levels (including all), such as “student < major < status < university < all”, how many cuboids will this cube contain (including the base and apex cuboids)?
2. Implementation of Text Mining on the data warehouse using [diabetes](http://storm.cis.fordham.edu/~gweiss/data-mining/weka-data/diabetes.arff) dataset.

**SET-4**

1. Implement star schema using multidimensional data model using SQL.

Fact table:

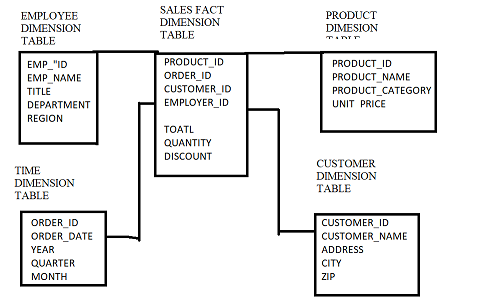
SALES (Product\_ID, Order\_ID, Customer\_ID, Employer\_ID, Total, Quantity, Discount)

Dimension tables:

Employee\_dimension (Emp\_ID, Emp\_Name, Title, Department , Region)

Product\_dimension (Product\_ID, Product\_Name, Product\_Category, Unit\_Price) Customer\_dimension (Customer\_ID, Customer\_Name, Address, City, Zip)

Time\_dimension (Order\_ID, Order\_Date, Year, Quarter, Month)



CREATE TABLE EMPLOYEE(EMP\_ID INT, EMP\_NAME VARCHAR(20), TITLE VARCHAR(20), DEPARTMENT VARCHAR(20), REGION VARCHAR(20));

CREATE TABLE PRODUCT(PRODUCT\_ID INT, PRODUCT\_NAME VARCHAR(20), PRODUCT\_CATEGORY VARCHAR(20), UNIT\_PRICE FLOAT);

CREATE TABLE TIME(ORDER\_ID INT, ORDER\_DATE VARCHAR(15), YEAR NUMBER(4), QUARTER VARCHAR(20), MONTH VARCHAR(10));

CREATE TABLE CUSTOMER(CUSTOMER\_ID INT, CUSTOMER\_NAME VARCHAR(20), ADDRESS VARCHAR(50), CITY VARCHAR(25), ZIP NUMBER(6));

CREATE VIEW FACTT AS SELECT PRODUCT.PRODUCT\_ID, TTIME.ORDER\_ID, CUSTOMER.CUSTOMER\_ID, EMPLOYEE.EMP\_ID FROM EMPLOYEE, PRODUCT , CUSTOMER, TTIME;

DESC FACTT;

INSERT INTO EMPLOYEE VALUES(001,'ABC','BHS', 'JHU','HUH');

INSERT INTO EMPLOYEE VALUES(002,'GFA','DSF', 'ASF', 'FSG');

INSERT INTO PRODUCT VALUES(001, 'GFSG', 'GGDD', 1254.50);

INSERT INTO PRODUCT VALUES(001, 'GSGS', 'GSRG', 1523.50);

INSERT INTO CUSTOMER VALUES(001, 'ADFA', 'FAES', 'DSFE', 532232);

INSERT INTO CUSTOMER VALUES(001, 'FRFR', 'REWR', 'SRFE', 523411);

INSERT INTO TIME VALUES(001,'23-06-2000', 1999, 'ONE', 'JUNE');

INSERT INTO TIME VALUES(001,'25-07-2000', 2003, 'TWO', 'JULY');

SELECT SUM(UNIT\_PRICE) AS TOATL FROM PRODUCT GROUP BY PRODUCT\_ID;

ALTER VIEW FACTT AS SELECT PRODUCT.SUM(UNIT\_PRICE) AS TOATL FROM PRODUCT GROUP BY PRODUCT\_ID;

1. Explore the correlation-ship analysis on weather data set

**SET-5**

1. Implement snowflake schema using multidimensional data model using SQL.

Fact table:

SALES (Product\_ID, Order\_ID, Customer\_ID, Employer\_ID, Total, Quantity, Discount)

Dimension tables:

Employee\_dimension (Emp\_ID, Emp\_Name, Department\_ID , Region, Territory)

Department\_dimension (Department\_ID, Name, Location)

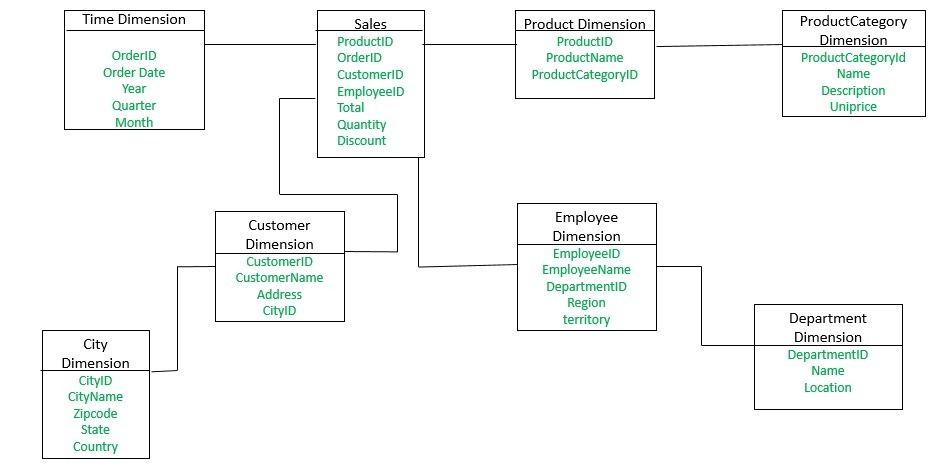
Product\_dimension (Product\_ID, Product\_Name, Product\_Category\_ID)

Product\_Category\_dimenstion (ProductCategoryID, Name, Description, Unit\_Price)

Customer\_dimension (Customer\_ID, Customer\_Name, Address, City\_ID)

City\_Dimension (City\_ID, City\_Name, ZIP\_Code, State, Country)

Time\_dimension (Order\_ID, Order\_Date, Year, Quarter, Month



CREATE TABLE EMPLOYEE(EMPLOYEE\_ID INT PRIMARY KEY, EMP\_NAME VARCHAR(20), DEPARTMENT\_ID INT REFERENCES DEPARTMENT, REGION VARCHAR(20),TERRITORY VARCHAR(20));

CREATE TABLE DEPARTMENT(DEPARTMENT\_ID INT PRIMARY KEY, NAME VARCHAR(20), LOCATION VARCHAR(20));

CREATE TABLE PRODUCT(PRODUCT\_ID INT PRIMARY KEY, PRODUCT\_NAME VARCHAR(20), PRODUCT\_CATEGORY\_ID INT REFERENCES PRODUCT\_CATEGORY);

CREATE TABLE PRODUCT\_CATEGORY(PRODUCT\_CATEGORY\_ID INT PRIMARY KEY, NAME VARCHAR(20), DESCRIPTION VARCHAR(50), UNIT\_PRICE FLOAT);

CREATE TABLE CUSTOMER(CUSTOMER\_ID INT PRIMARY KEY, CUSTOMER\_NAME VARCHAR(20), ADDRESS VARCHAR(30), CITY\_ID INT REFERENCES CITY);

CREATE TABLE CITY(CITY\_ID INT PRIMARY KEY, CITY\_NAME VARCHAR(20), ZIP\_CODE NUMBER(6), STATE VARCHAR(25), COUNTRY VARCHAR(20));

CREATE TABLE TIME(ORDER\_ID INT, ORDER\_DATE DATE, YEAR NUMBER(4), QUARTER VARCHAR(10), MONTH VARCHAR(10));

CREATE VIEW FACTT AS SELECT PRODUCT.PRODUCT\_ID, TIME.ORDER\_ID, CUSTOMER.CUSTOMER\_ID, EMPLOYEE.EMPLOYEE\_ID FROM PRODUCT, TIME, CUSTOMER, EMPLOYEE;

DESC FACTT;

SELECT \* FROM FACTT;

INSERT INTO EMPLOYEE\_DIMENSION VALUES(1, 'A', 'EAST', 'DELHI');

INSERT INTO EMPLOYEE\_DIMENSION VALUES(2, 'B', 'NORTH', 'LADAKH');

INSERT INTO DEPARTMENT\_DIMENSION VALUES(1, 'AB', 'AGRA');

INSERT INTO DEPARTMENT\_DIMENSION VALUES(2, 'AC', 'TENALI');

INSERT INTO PRODUCT\_DIMENSION VALUES(11,'ABC',101);

INSERT INTO PRODUCT\_DIMENSION VALUES(22,'ABD',102);

INSERT INTO PRODUCT\_CATEGORY\_DIMENSION VALUES('AD', 'FACE CREAME', 10);

INSERT INTO PRODUCT\_CATEGORY\_DIMENSION VALUES('BD', 'MAKEUP KIT', 20);

INSERT INTO CUSTOMER\_DIMENSION VALUES(1001, 'XY', 'BUS STAND');

INSERT INTO CUSTOMER\_DIMENSION VALUES(1002, 'YZ', 'NEAR POLICE STATION');

INSERT INTO CITY\_DIMENSION VALUES(10001, 'GUNTUR', 522201, 'AP', 'INDIA');

INSERT INTO CITY\_DIMENSION VALUES(10002, 'VIJAYAWADA', 522001, 'AP', 'INDIA');

INSERT INTO TIME\_DIMENSION VALUES(01, 25, 2022, 4, 01);

INSERT INTO TIME\_DIMENSION VALUES(02, 13, 2021, 1, 12);

1. Explore the correlation-ship analysis on labor data set

**SET-6**

1. Implement Fact constellation schema using multidimensional data model using SQL.

Fact Table1: Placement (Stud\_roll, Company\_id, TPO\_id, Number of students eligible, Number of students placed)

Fact Table2: Workshop (Stud\_roll, Institute\_id, TPO\_id, Number of students selected, Number of students attended the workshop)

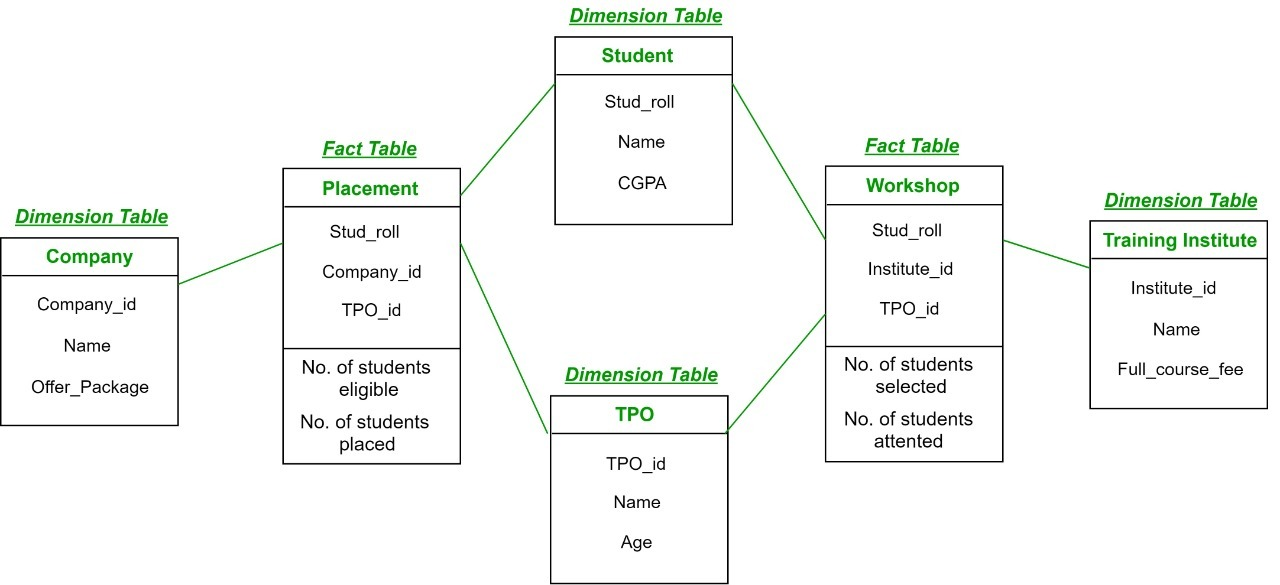
Dimension Tables:

Company (Company\_id, Name, Offer\_package)

Student (Student\_roll, Name, CGPA)

TPO (TPO\_id, Name, Age)

Training\_Institute (Institute\_id, Name, Full\_course\_fee)



CREATE TABLE COMPANY(COMPANY\_ID INTEGER PRIMARY KEY, OFFER\_PACKAGE INTEGER);

CREATE TABLE STUDENT(STUDENT\_ROLL INTEGER PRIMARY KEY, STUDENT\_NAME VARCHAR(20), CGPA FLOAT);

CREATE TABLE TPO(TPO\_ID INTEGER PRIMARY KEY, TPO\_NAME VARCHAR(20), TPO\_AGE INTEGER);

CREATE TABLE TRAINING\_INSTITUTE(INSTITUTE\_ID INTEGER PRIMARY KEY, NAME VARCHAR(20), FULL\_COURSE\_FEE INTEGER);

INSERT INTO COMPANY VALUES(1, 500000);

INSERT INTO COMPANY VALUES(2, 250000);

INSERT INTO STUDENT VALUES(1, 'A', 6.2);

INSERT INTO STUDENT VALUES(2, 'B', 7.6);

INSERT INTO TPO VALUES('a', 'AB', 50);

INSERT INTO TPO VALUES('b', 'BC', 60);

INSERT INTO TRAINING\_INSTITUTE VALUES('101', 200000);

INSERT INTO TRAINING\_INSTITUTE VALUES('102', 150000);

CREATE VIEW PLACEMENT STUDENT.STUDENT\_ROLL, COMPANY.COMPANY\_ID, TPO.TPO\_ID FROM STUDENT, COMPANY, TPO;

CREATE VIEW WORKSHOP STUDENT.STD\_ROLL, TRAINING\_INSTITUTE.INSTITUTE\_ID, TPO.TPO\_ID FROM STUDENT, TRAINING\_INSTITUTE, TPO;

1. Explore the correlation-ship analysis on [diabetes](http://storm.cis.fordham.edu/~gweiss/data-mining/weka-data/diabetes.arff) data set

**SET-7**

1. Implement star schema for the following data warehouse.

**orders\_fact** The orders fact table tracks orders of spare parts. The grain of the fact table is one row per order. An order is placed by a single customer, for a single part, from a single supplier. **There are four dimensions**: **customer, part, supplier, and date.** The fact table has four measurement columns: order\_quantity, order\_extendedprice, order\_discount, order\_tax.

**customer\_dim** The customer dimension includes the following attributes: name, address, nation, region, phone, account balance, market segment, and comment.

**part\_dim** The part dimension includes the following attributes: name, manufacturer, brand, type, size, container, retail price, and comment.

**supplier\_dim** The supplier dimension includes the following attributes: name, address, nation, region, phone, account balance, and comment.

**date\_dim** The date dimension includes the following attributes: date value (a SQL date column), month name, month number, year, day of week, and day number in month.

CREATE TABLE CUSTOMER(CUSTOMER\_NAME VARCHAR(20), ADDRESS VARCHAR(50), NATION VARCHAR(20), REGION VARCHAR(20), PHONE NUMBER(10), ACCOUNT\_BALANCE FLOAT, MARKET\_SEGMENT VARCHAR(20), CUSTOMER\_COMMENT VARCHAR(10));

CREATE TABLE PART(PART\_NAME VARCHAR(20), MANFACTURE VARCHAR(20), BRAND VARCHAR(20), PART\_TYPE VARCHAR(20), PART\_SIZE VARCHAR(10), PART\_CONTAINER VARCHAR(20), RETAIL\_PRICE FLOAT, PART\_COMMENT VARCHAR(20));

CREATE TABLE SUPPLIER(SUPPLIER\_NAME VARCHAR(20), ADDRESS VARCHAR(50), NATION VARCHAR(20), REGION VARCHAR(20), PHONE NUMBER(10), ACCOUNT\_BALANCE FLOAT, SUPPLIER\_COMMENT VARCHAR(10));

CREATE TABLE DDATE(DATE\_VALUE DATE, MONTH\_NAME VARCHAR(20), MONTH\_NUMBER INTEGER, DATE\_YEAR NUMBER(4), DAY\_OF\_WEEK VARCHAR(20), DAY\_NUMBER\_IN\_MONTH NUMBER(2));

CREATE VIEW FACTT AS SELECT CUSTOMER.CUSTOMER\_NAME, PART.PART\_NAME, SUPPLIER.SUPPLIER\_NAME, DDATE.DATE\_VALUE FROM CUSTOMER, PART, SUPPLIER, DDATE;

DESC FACTT;

INSERT INTO CUSTOMER VALUES('A', 'GUNTUR', 'INDIA', 'EAST', 9876543210, 1000, '1', 'GOOD');

INSERT INTO CUSTOMER VALUES('B', 'VIJAYAWADA', 'INDIA', 'WEST', 8796543210, 1500, '2', 'NICE');

INSERT INTO PART\_DIM VALUES('A', 'BIKE', 'HONDA', 'MIRROR', 10, 'ABC', 10, 'GOOD');

INSERT INTO PART\_DIM VALUES('B', 'CAR', 'SUZUKI', 'DOOR', 50, 'BCA', 1000, 'BETTER');

INSERT INTO SUPPLIER VALUES('AB', 'TENALI', 'INDIA', 'EAST', 7894561230, 2000, 'OK');

INSERT INTO SUPPLIER VALUES('BC', 'PONNUR', 'INDIA', 'SOUTH', 8794561230, 2500, 'GOOD');

INSERT INTO DATE\_DIM VALUES(2021-06-01, 'JUNE', 06, 2021, 'SUNDAY', 01);

INSERT INTO DATE\_DIM VALUES(2021-05-03, MAY, 07, 2021, 'SATURDAY', 03);

1. Explore the correlation-ship analysis on supermarket dataset