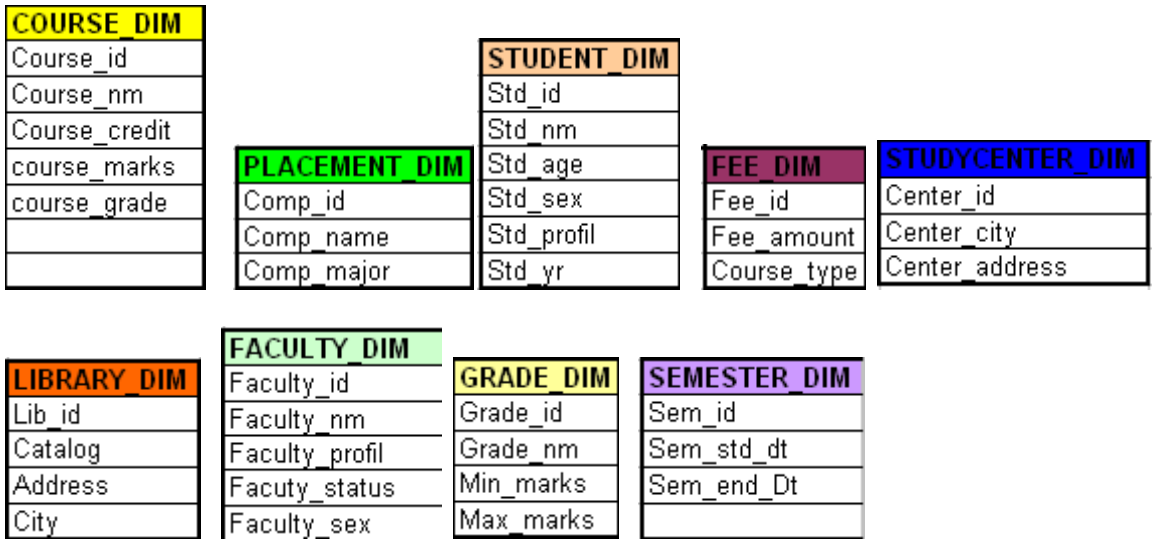
ASSIGNMENT -3

Participants –
Anirban Roy Choudhury
Siddharth Verma
Vikram Sahu
Satyajit Mohanty

1) The following are the different Dimensions tables that can be created to manage the IIMK Data Warehouse:



2) The following are the 3 classes of schema popularly used in Data Warehouse:

- Star Schema
- Snowflake Schema
- Constellation with Conforming Dimension

Mentioned below is the brief description about each schema

Star Schema:

This is the simplest and most commonly used schema diagram in the Data Warehousing domain. The Dimension tables are highly denormalized and each fact table is tightly coupled to the dimension tables depicting a star diagram.

Snowflake Schema:

Snowflake schema is a logical arrangement of the tables in a multidimensional database such that E-R diagram resembles snowflake shape.
The Dimension tables are highly normalized and are coupled with the fact table.

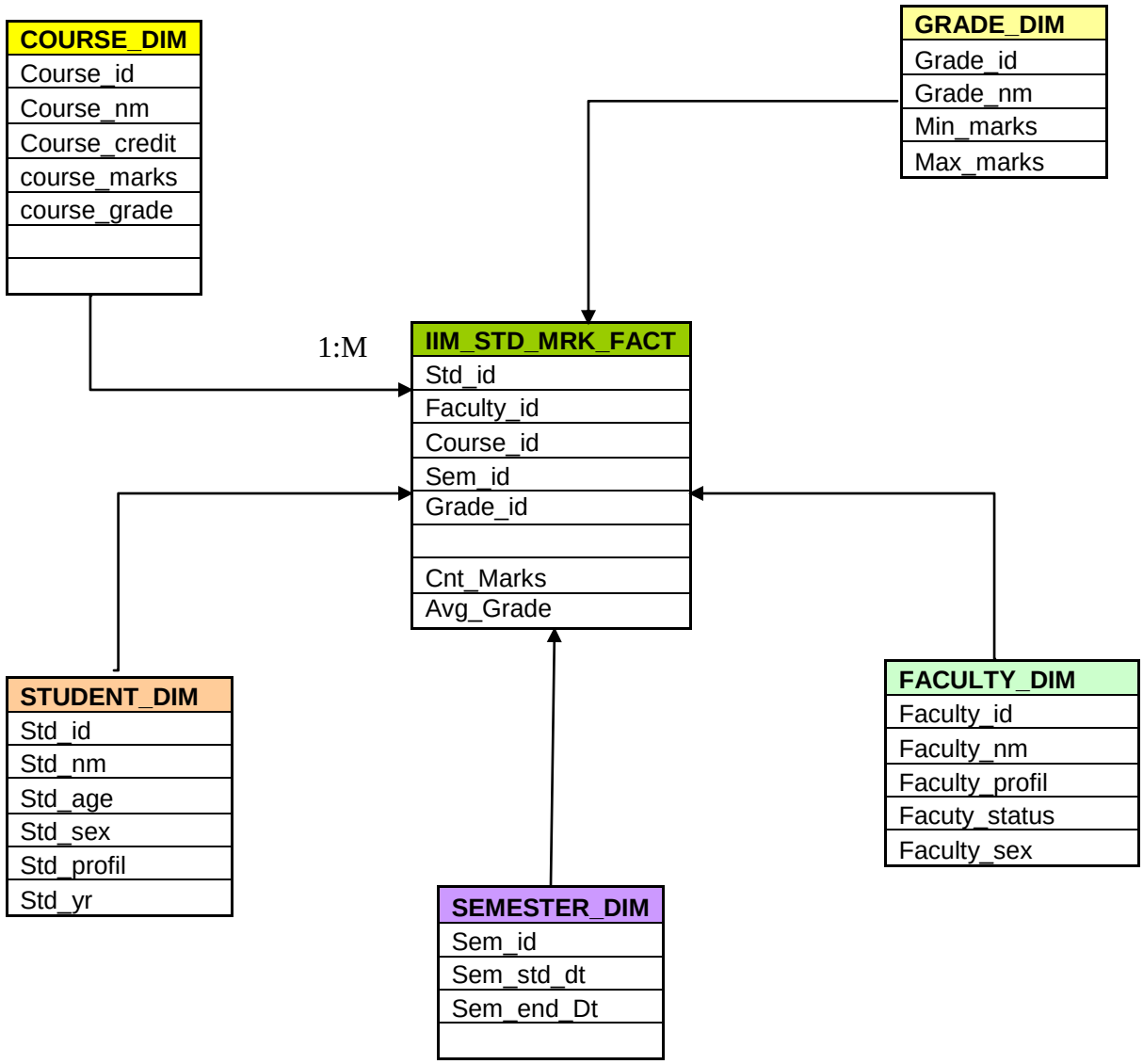
Constellation with Conforming Dimension

This is an extension to the Star schema where conformed dimension is used to link the different fact tables.

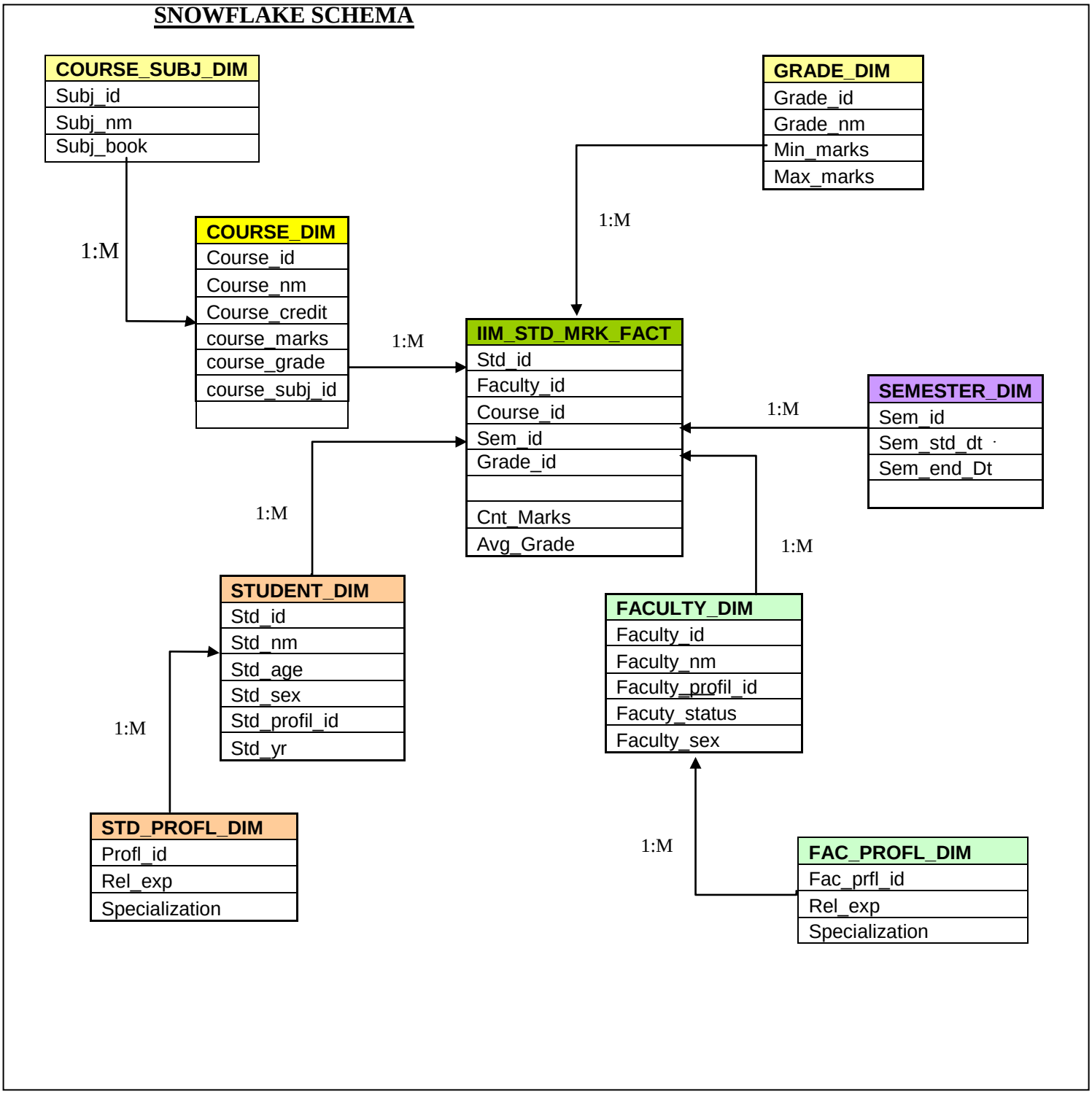
3) ***Mentioned below are the different schema diagrams possible for the IIMK-DW***

Star Schema:

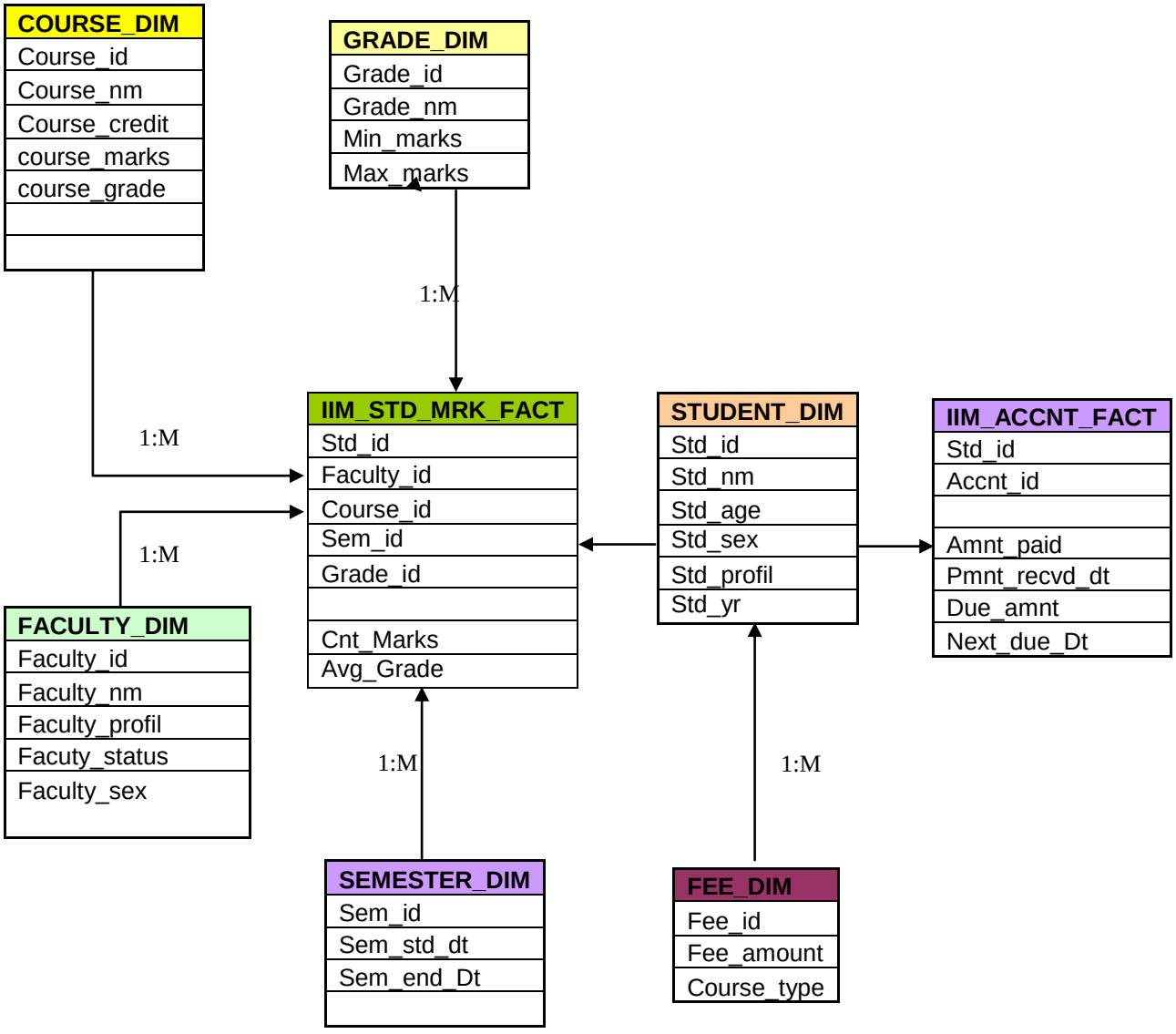
1:M (One to Many)



SNOWFLAKE SCHEMA

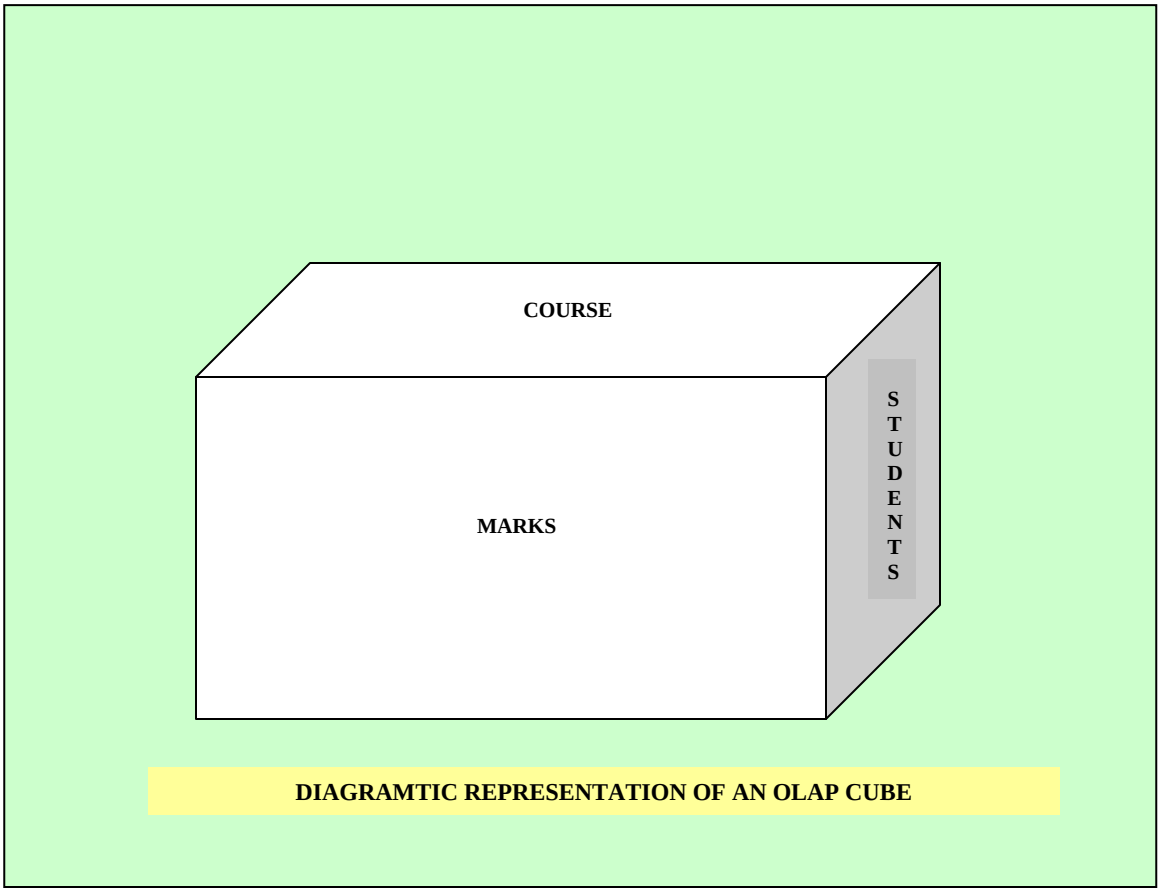


CONSTELLATION SCHEMA

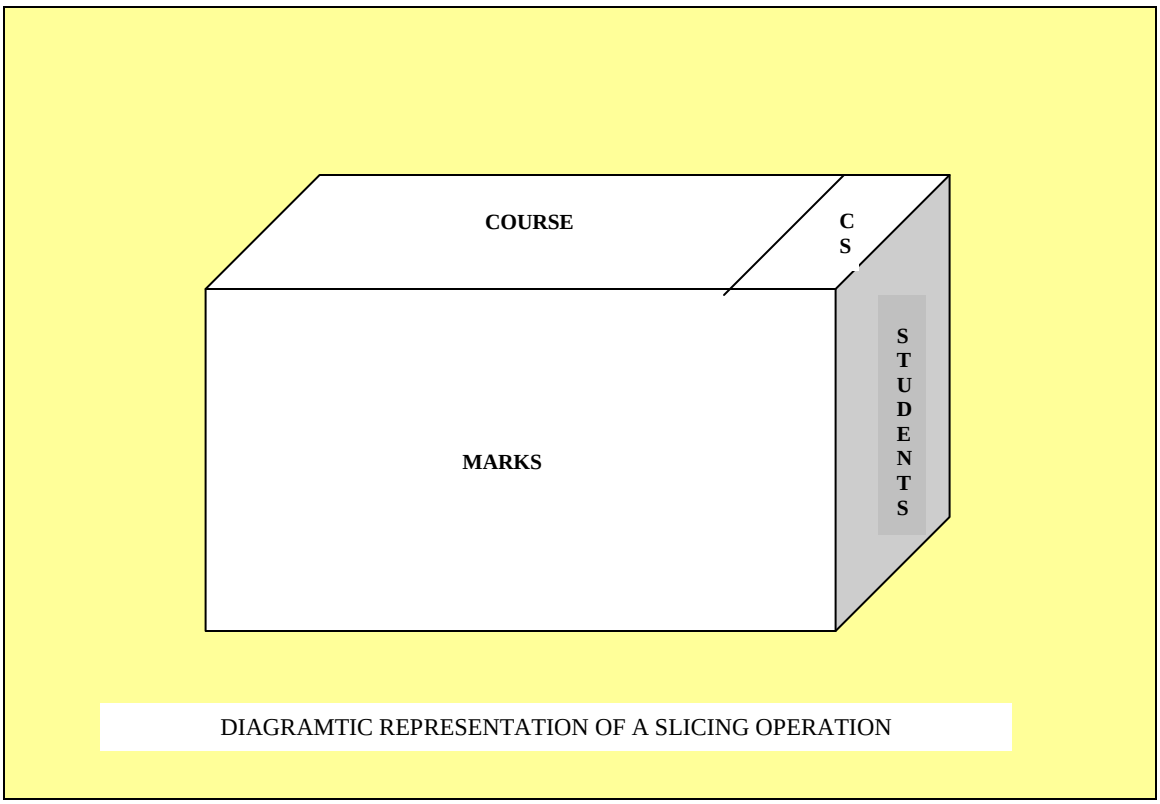


4) ***Finding the average grades scored by each student in CS courses***

The OLAP cube is depicted pictorially in the diagram mentioned below for the operation to take place:



DIAGRAMTIC REPRESENTATION OF AN OLAP CUBE



DIAGRAMTIC REPRESENTATION OF A SLICING OPERATION

A) The above diagram represents a slicing operation over the cube on the COURSE=’CS’ and performing the aggregation of the marks over the student to get the desired result.

Corresponding sql when viewed in BOXI with Teradata DB:

```
SELECT
        B.Std_nm,
        SUM(A.CNT_MRKS)
FROM
        IIM_FACT A
INNER JOIN
        STUDENT_DIM B
ON
        A.Std_id=B.Std_id
INNER JOIN
        COURSE_DIM C
ON
        A.Course_id=C.Course_id
WHERE
        C.Course_nm=’CS’
GROUP BY B.Std_nm
;
```

B)

Apart from this the various rollup functions can be used to find the result:

i) Roll-up on course from course id to program

Corresponding sql when viewed in BOXI with Teradata DB:

```
SEL
        COURSE_ID
       ,SUM(CNT_MRKS)
FROM
       IIM_FACT
GROUP BY RollUp(COURSE_ID)
;
```

ii) Roll-up on semester from semester id to all over the years.

Corresponding sql when viewed in BOXI with Teradata DB:

```
SEL
        SEMESTER_ID
       ,SUM(CNT_MRKS)
FROM
       IIM_FACT
GROUP BY RollUp(SEMESTER_ID)
;
```

iii) Roll up on student.

Corresponding sql when viewed in BOXI with Teradata DB:

```
SEL
        STD_ID
       ,SUM(CNT_MRKS)
FROM
       IIM_FACT
GROUP BY RollUp(STD_ID)
;
```

C)

BO measure objects can be defined over the cube , which is used to aggregate the marks and based on it the different objects like courses and students can be pulled to get a comprehensive report.

5) ***If each dimension has 5 levels,how many cuboids will this cube contain?***

In this problem there are 9 dimensions identified, therefore if each dimension has 5 levels then ‘total number of cuboids’ possible in cube is 5^n

Where n is the number of dimension.

Considering 9 dimensions, with reference to the above problem statement.

The number of cuboids in the cube = (5*5*5*5*5*5*5*5*5)=1953125.