Assignment 1 - BDA

* **Task 1: Given the spreadsheet file convert it into a csv.**
* Just simply open Phpmyadmin using xampp.
* Go to import Section.
* Choose spreadsheet file which you want to convert into .csv file.
* Select format as a CSV.
* Click on Go.
* **We can also do this by just changing SAVE AS TYPE to csv**
* **Task 2: Import a csv into MySQL database table.**
* Just simply open phpMyAdmin using xampp.
* Go to import section.
* Choose csv file which you want to convert into .csv file.
* Select format as a SQL
* Click on Go.

* **Task 3: Write a computer program to read records from database and generate data file.**

**A. XML**

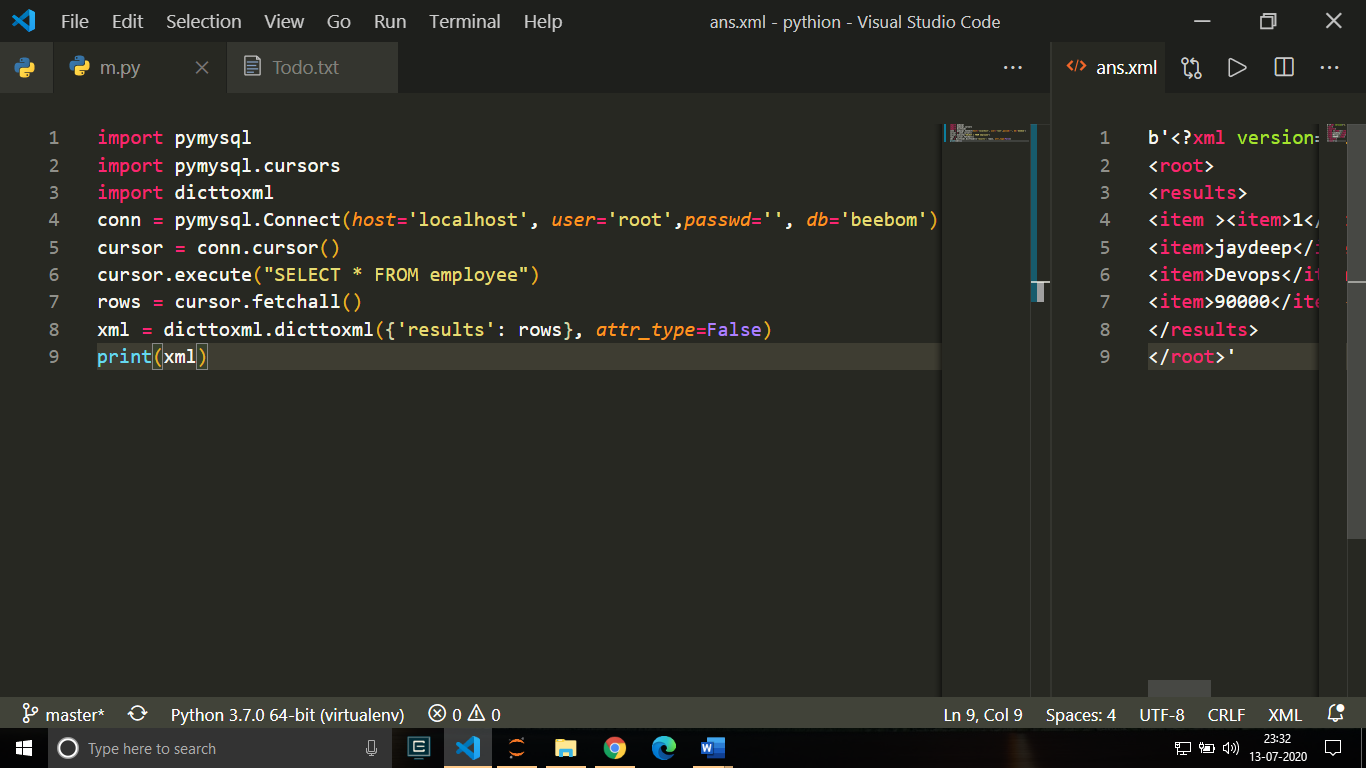
**B. JSON**

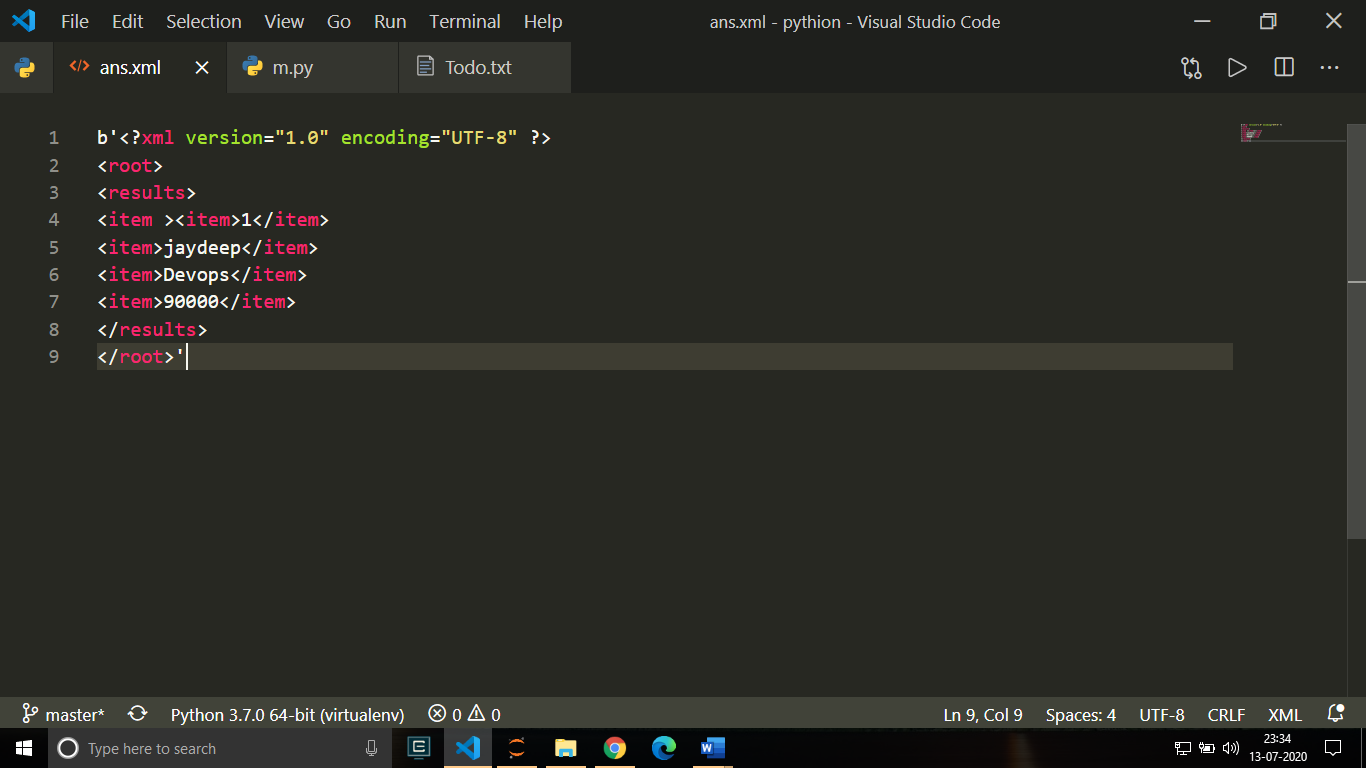
***FOR XML***

As shown in the Above Image we need to write Python code which converts sql file to xml using dicttoxml module of python.

* Make a connection with Database using pymsql module
* Execute the Select Query for Given Sql Table
* And Now Using Dicttoxml module Write data in Xml File

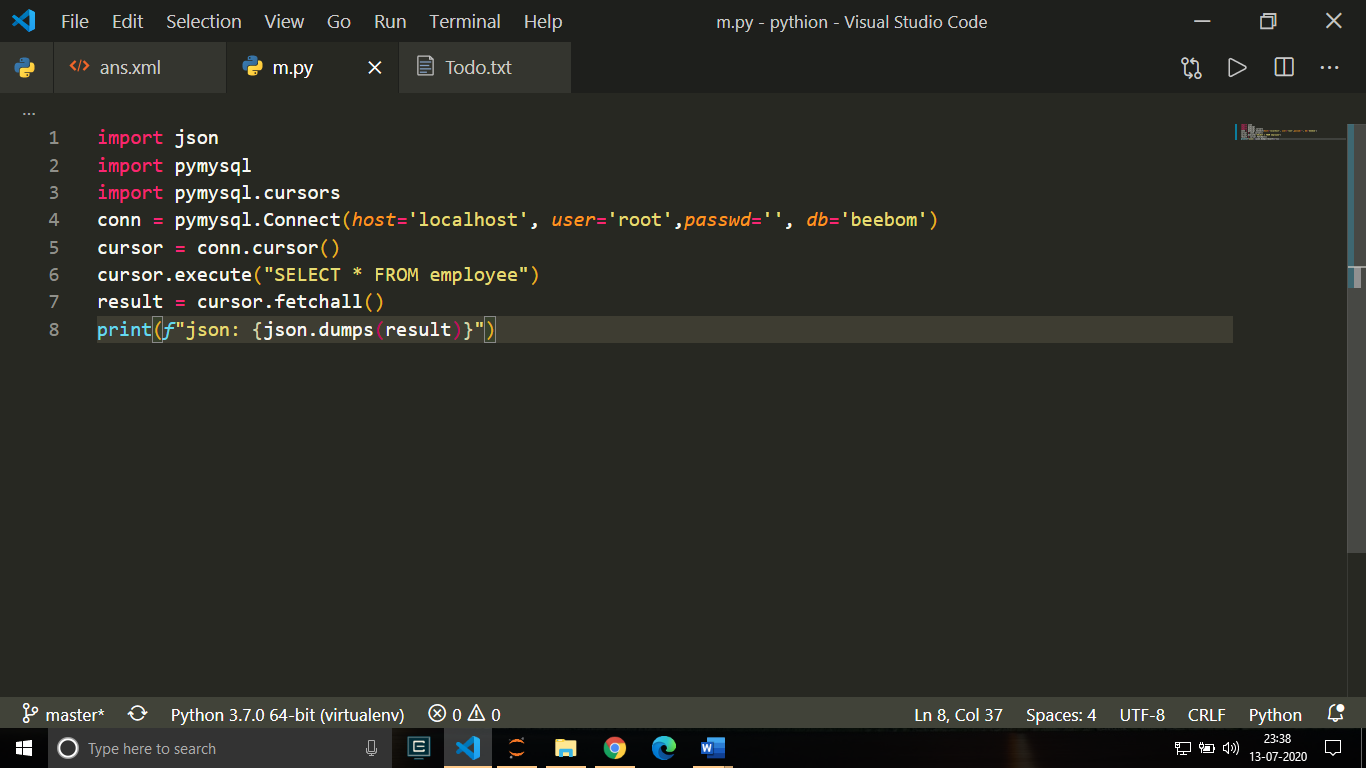
Code:



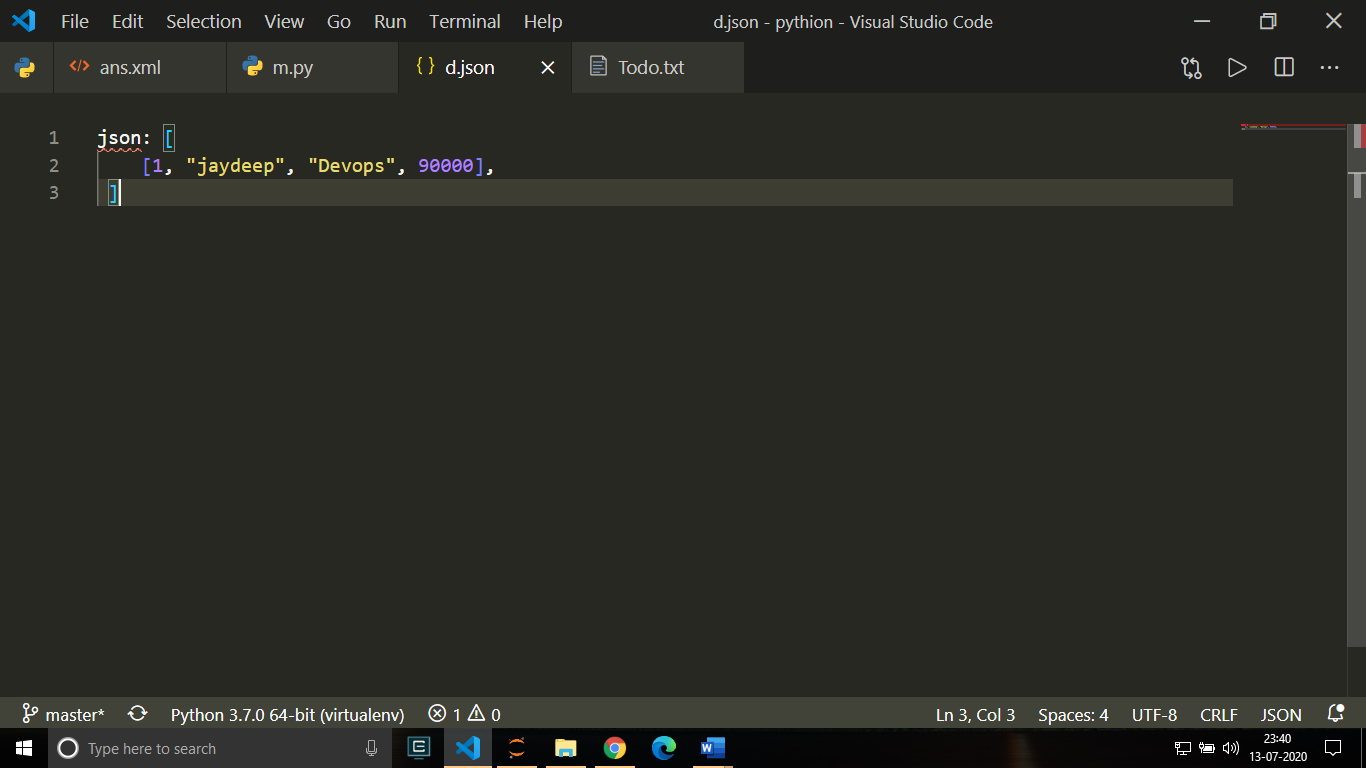
Output:

***For JSON***

Code:



Output:



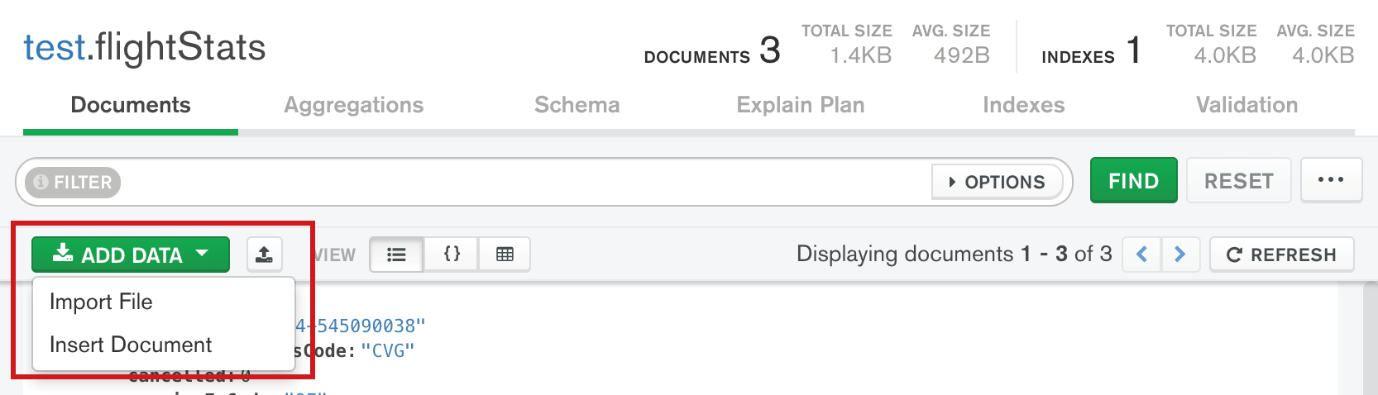
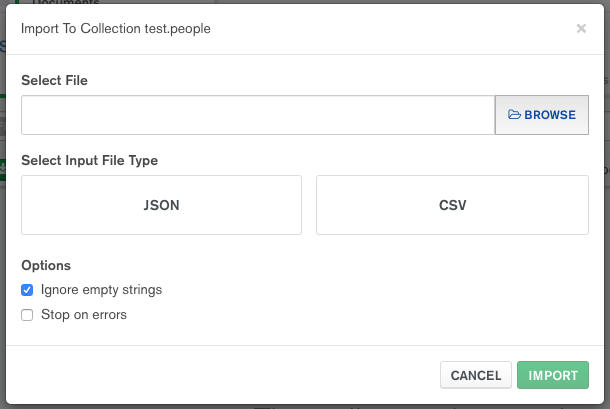
As shown in the Above Image we need to write Python code which converts sql file to JSON using Literal String Interpolation (also known as F-Strings) of python.

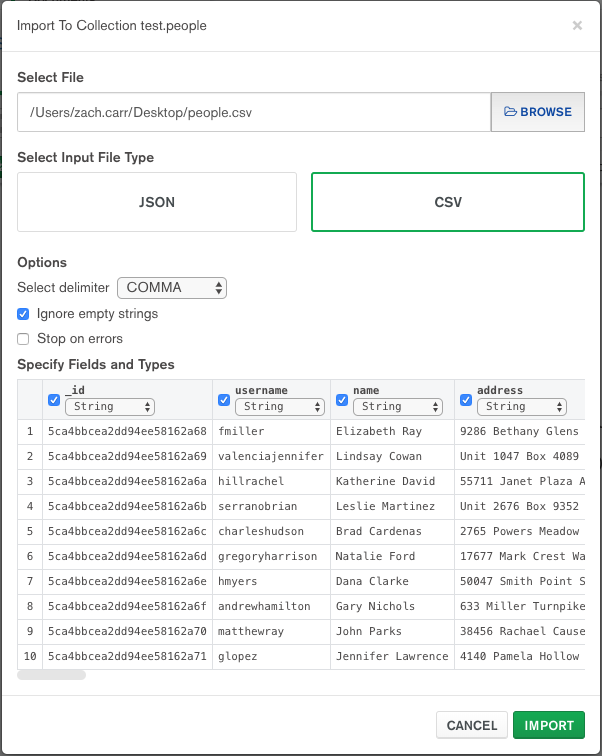
* Make a connection with Database using pymsql module.
* Execute the Select Query for Given Sql Table.
* And Now Using JSON module and F-strings of Python write data in JSON File.
* **Task 4: Import XML/JSON file into another database/table. I.e. MS Access. Oracle, etc.**
* Create New Database in MS Access.
* go to “External Data” Section.
* Go to “New Data source”.
* Go to “From File”.
* Browse the XML File.
* Click on “Ok”.
* **Task 5: Export database dump for data migration/archival.**
* In Phpmyadmin Go into the Database which you want to Export.
* Go to the Export Section
* Click on the “Go”
* Write the new Name of the Exported SQL File.
* **Task 6: Validate/Map data types across different database systems when migrating from one to another.**

***Export SQL File using phpMyAdmin as CSV File.***

* Export whichever Table you want to export from Phpmyadmin
* Select Export type as CSV
* Click on “Go”
* Name the CSV File and Save it.

***Import CSV File using MongoDB Compass and Map it to MongoDB Collection.***

* Navigate to the collection you want to import data into.
* ******Click the Add Data dropdown and Select import file.
* Select the location of source Data file under Select File.

******

* Select the location of the Source Data File.
* Choose File Type As JSON.
* Configure Import options if you want.
* Click on “import”.
* **Task 7: Represent Data Cube and perform operations. OLAP - Data Warehouse.**

Here the data warehouse is going to represent number of admission with average of percentage across branches (Eng-CS, Eng-EC, Med-MMBS, MedDental), across type of students (Male-Open, Male-SC, Female-Open, FemaleSC), across period (2004-1RS, 2004-2RS, 2005-1RS, 2005-2RS), Where RS means reshuffling.

For dimensional modelling we should follow the steps as below.

* Identify the Business Process: -

Identifying the business process is the process of determining business process that the data warehouse represents.

* Identify the Grain: -

Identifying the Grain is the process of identifying the level of detail of the fact table.

* Identify the Dimensions: -

Identifying the dimensions for the business process of interest is the process of representing characteristics such as who, what, where, when, how of a measurement.

* Dimension 1: -

BRANCH Table(WHERE) Represents different branch types available for admission.

|  |  |  |
| --- | --- | --- |
| Field Name | Type | Primary Key |
| Branch\_Id | Integer | Yes |
| Branch Category | Varchar |  |
| Branch\_Stream | varchar |  |

***create table Branch (Branch ID integer not null,Branch\_CATEGORY varchar(15),Branch STREAM varchar(20), PRIMARY KEY(Branch\_ID));***

* Dimension 2:

FELLOW Table (WHO) Represents different fellow types admitted.

|  |  |  |
| --- | --- | --- |
| Field Name | Type | Primary Key |
| Fellow\_Id | Integer | Yes |
| Fellow\_Gender | char |  |
| Fellow\_Category | varchar |  |

* ***create table Fellow (Fellow\_ID integer not null,Fellow\_GENDER char,Fellow\_CATEGORY varchar(5), PRIMARY KEY(Fellow\_ID))***
* Dimension 3:

PERIOD Table (WHEN) Represents different period of admission

|  |  |  |
| --- | --- | --- |
| Field Name | Type | Primary Key |
| Period\_Id | Integer | Yes |
| Period\_TYPE | char |  |
| Period\_YEAR | integer |  |

* ***create table Period (Period\_ID integer not null, Period\_TYPE. integer,Period\_YEAR varchar(5), PRIMARY KEY(Period\_ID));***
* Identify the Facts: -

Identifying the facts is the process of determining measurements for the business process of interest. (what)

FACT TABLE SCHEMA: FACT

|  |  |  |
| --- | --- | --- |
| Field Name | Type | Primary Key |
| Branch\_Id | Integer | Yes |
| Fellow\_Id | Integer | Yes |
| Period\_Id | integer | Yes |
| Fact\_No\_Adm | Integer |  |
| Fact\_Avg\_Per | Integer |  |

* ***create table Fact (Branch\_ID integer references Branch, Fellow\_ID integer references Fellow, Period\_ID integer references Period, Fact\_No\_Adm integer, Fact\_Avg\_Per integer,primary key (Branch\_Id,Fellow\_Id,Period\_Id));***

Basic OLAP Operations:

• Slice

• Dice

• Roll-Up

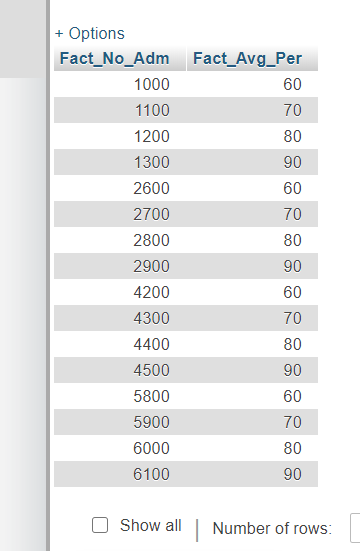
• Drill Down

**1) SLICE**

Find and display number of students only those Male & Open type of fellow with their percentage average for each period and for each type of branch separately.

* ***Select Fact\_No\_Adm, Fact\_Avg\_Per from Branch B,Fellow F,Period D,FACT where ( B.Branch\_Id = FACT.Branch\_Id and F.Fellow\_Id = FACT.Fellow\_Id and D.Period\_Id = FACT.Period\_Id and F.Fellow\_Gender = 'M' and F.Fellow\_Category = 'OPEN');***

Ouput: -



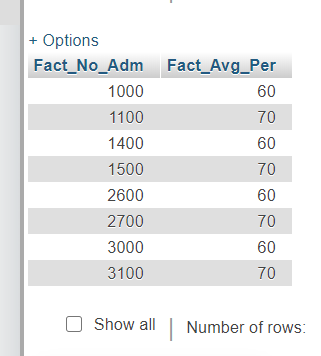
**2)DICE**

Find number of students with their average percentage which satisfies following:

* Belong to Branch Category of Engineering
* Male
* Have reshuffled during year 2004

***Select Fact\_No\_Adm, Fact\_Avg\_Per from Branch B,Fellow F, Period D, FACT where ( B.Branch\_Id = FACT.Branch\_Id and F.Fellow\_Id = FACT.Fellow\_Id and D.Period\_Id = FACT.Period\_Id and (F.Fellow\_Id = 1 or F.Fellow\_Id = 2) and (B.Branch\_Id = 1 or B.Branch\_Id = 2) and (D.Period\_Id = 1 or D.Period\_Id = 2));***

Output: -



**3)Roll UP**:

Find number of total students admitted to branch category of Engineering.

***Select sum(Fact\_No\_Adm) from Branch B,Fellow F,Period D,FACT where B.Branch\_Id = FACT.Branch\_Id and F.Fellow\_Id = FACT.Fellow\_Id and D.Period\_Id = FACT.Period\_Id and B.Branch\_Category = 'ENG' group by B.Branch\_Category;***

Output: -

