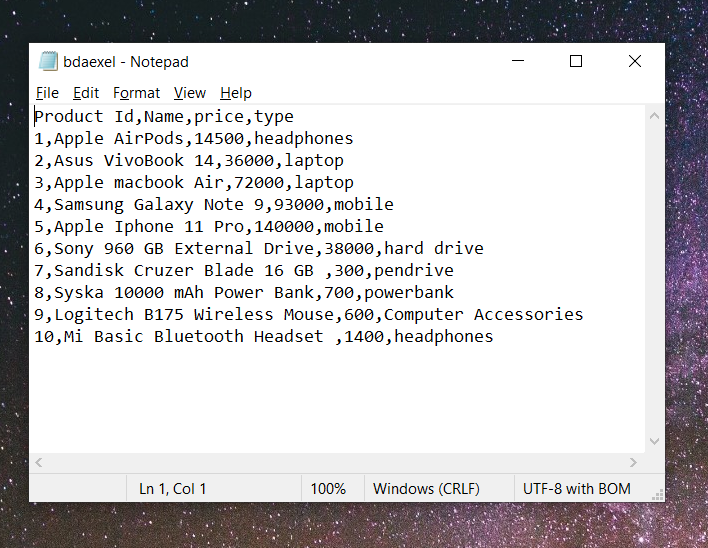
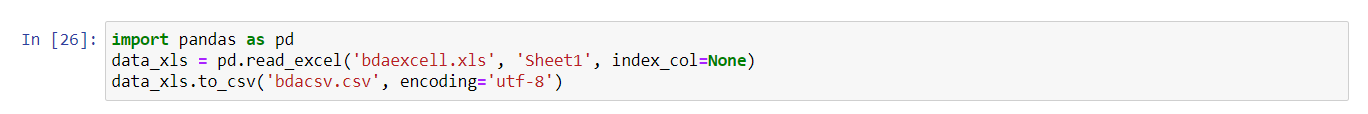
**LAB 1**

**Task 1: Given the spreadsheet file convert it into a csv**

We can convert a spreadsheet into csv (Comma Separated Values) by saving it as a csv file. This option is provided in Microsoft excel files and many other files. After that we can see this file in any text editor and all the data of each row is stored separated by comma in which the first row represents the title of columns. Here the file given below is the csv file of an excel file.



We can convert the spreadsheet into csv file using python code, too.



**Task 2: Import a csv into MySQL database table**

For the given task I have used MySQL tables using phpMyAdmin. In phpMyAdmin if we want to import a csv file then we need to perform the following steps.

1. Select database.
2. Select a table.
3. Then click on import button and choose file.
4. Select ‘CSV using LOAD DATA’ for format.
5. Choose the format and click on Go.

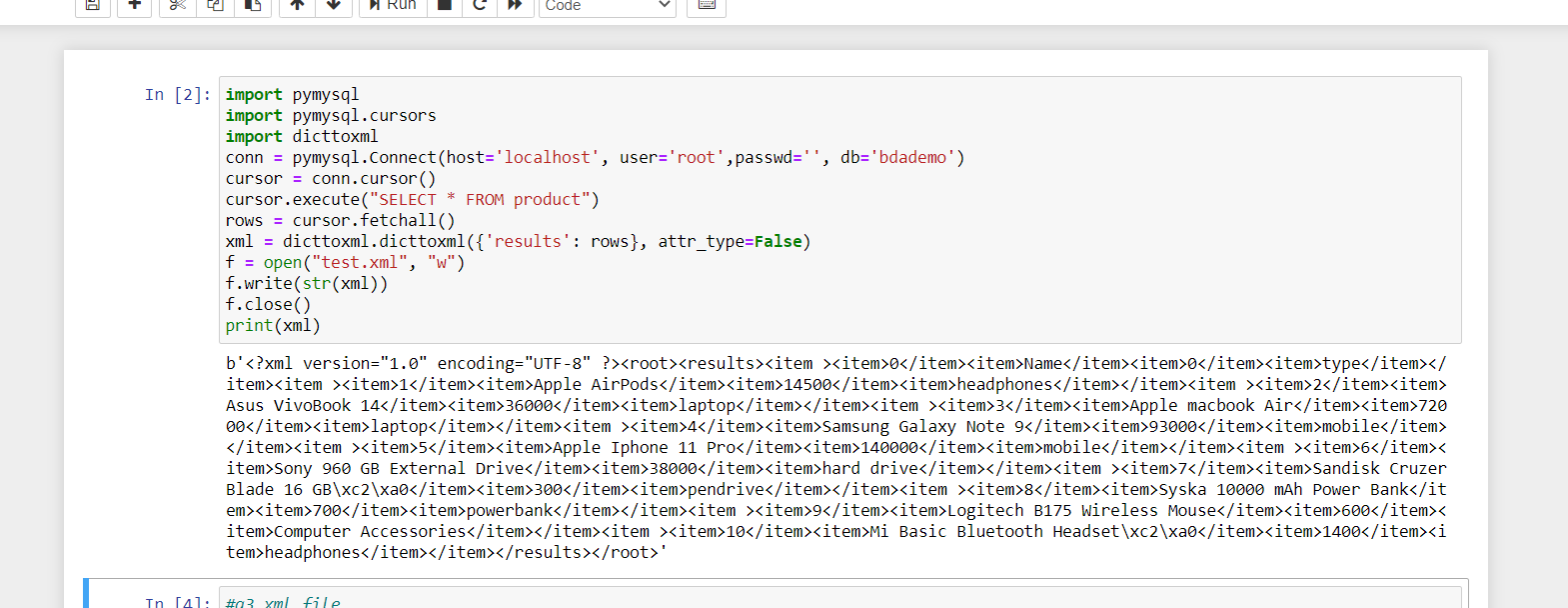
We can do the same by SQL query in MYSQL. To import data from csv file we can run the following query in MySQL.

>>LOAD DATA INFILE 'path/to/file1.csv' INTO TABLE product FIELDS TERMINATED BY ',’ ENCLOSED BY '"' LINES TERMINATED BY '\n' IGNORE 1 ROWS;

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

1. **xml:**

To perform given task, I have created a python program which read data from MySQL and then convert it to the xml format. For doing this task I have used pymysql library to work with phpMyAdmin database and for converting this data to xml I have used dicttoxml library.



1. **json:**

To perform given task, I have created a python program which read data from MySQL and then convert it to the json format. For doing this task I have used pymysql library to work with phpMyAdmin database and for converting this data to xml I have used json library.



Apart from writing the program we can do this task by GUI approach. Many databases provide GUI support for this. We can export the database file in any format like xml, json, csv, yaml etc.

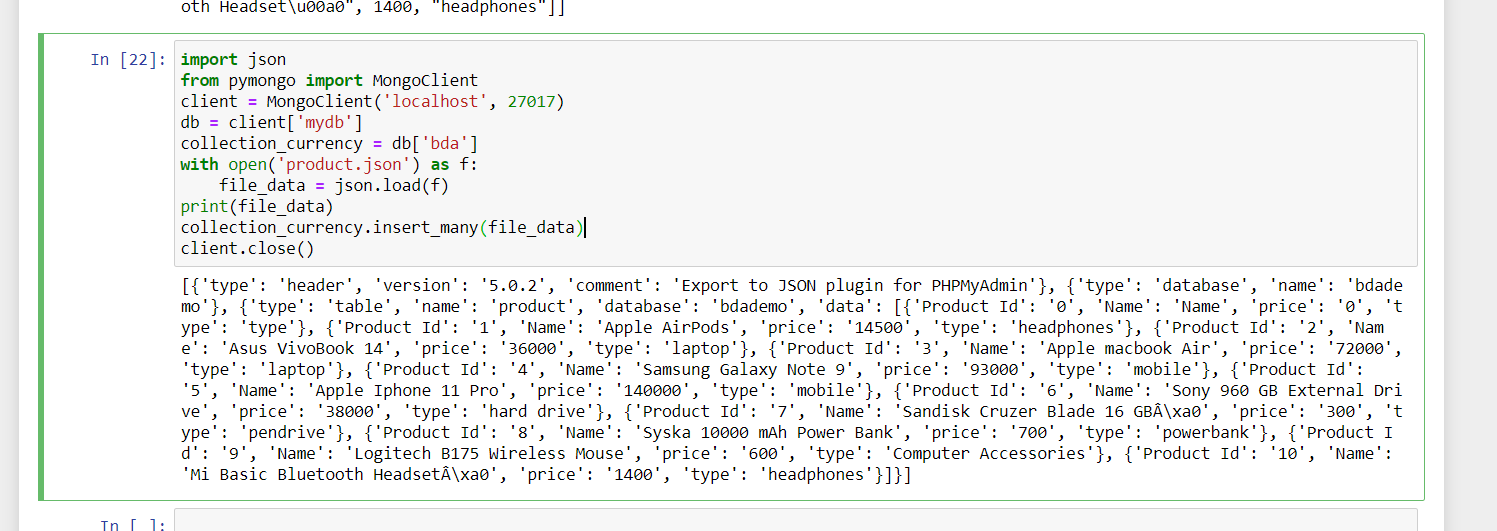
**Task 4: Import XML/JSON file into another database/table. i.e. MS Access. Oracle, etc.**

For performing the given task, I have used MS Access database. In MS Access we can easily import a xml file in the database. For doing this we can follow the given steps below.

1. Create database in MS Access.
2. Then from external data tab we can import our external xml file in it.

We can do the same with json file.

In mongo dB, we can perform this by both CLI and GUI. I have imported a json file in mongo dB using python. For this I have used Mongo Client from pymongo library which is used to connect with mongo dB.



**Task 5: Export database dump for data migration/archival**

In phpMyAdmin, this task can be done easily by GUI approach. In GUI approach we just need to select our database and then export it in the sql format.

In MySQL we can perform the task by the command line interface. For this we need to execute the following command.

>>mysqldump -u user -p bdademo > bda1.sql

**Task 6: Validate/Map data types across different database systems when migrating from one to another**

For this task I used 2 different databases – MySQL and oracle. First, I have created database in MySQL and then exported it in the json file format. After that in oracle database we can execute the following query to import the data from the json file.

>>INSERT INTO Product1 (Product\_Id, Name, price, type) SELECT json\_document FROM product WHERE json\_document IS JSON;

By performing the above query, we can import data from json file. And insert this data in the oracle database.

After performing the above we can map the data type difference across these 2 databases. I have observed change in varchar data type. In MySQL we use ‘varchar(len)’ for alphanumeric values but in oracle we use ‘varchar2(len)’ for the same. In MySQL we use int, decimal(x,y) for defining integer and decimal numbers whereas in oracle we use number(x,y) to define any numeric value.

**Task 7: Represent Data Cube and perform operations. OLAP - Data Warehouse**

For this task we should follow the steps given below for a database system. For this task let’s take a database system for admission of students.

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 modeling we should follow the steps as below.

1. Identify the Business Process: -

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

1. Identify the Grain: -

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

1. 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));

1. 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

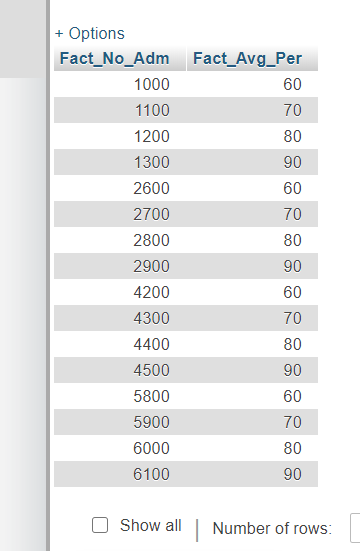
• Drill Down

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');

Output: -



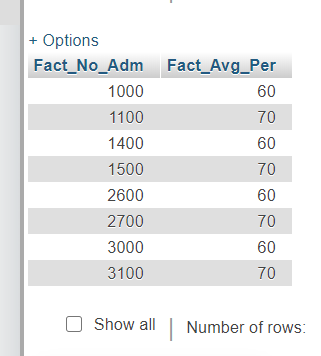
1. DICE

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: -

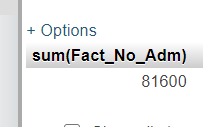


3)Roll UP:

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: -



**Task 8: Generate pdf report.**

For this task we have to generate a pdf report which represents the analysis of given data and can be used to analyse the database easily. For this task, to analyse the data I used a database which represents the data of no. of students admitted in a college in a year. For analysing the data, I have created bar chart which maps year and no. of students admitted. I have done this task using python library matplotlib. And then converted this data to the pdf file which is having the whole data and the bar chart.

