CSE 202: Fundamentals of Database Systems
Winter 2018
Home Assignment 1 [10 marks]
Due Date: 13-01-2018 (One week time: No extension will be allowed)

Instructions:

- Write the programs in Python or C++ or Java.
- The naming convention for the files: filename_rollnumber.extension like program1_2016001_2016002.py/program1_2016001_2016002.cpp.
- **The assignment can be done in groups of maximum 2 students.**
- **Compress all the input/output files along with the programs as tar.gz. Your submission (.tar.gz) must contain a text file with your group details (name and roll number of student). Your submission must contains files only with no folder hierarchy created.**

Physical data independence has to do with storage of data in hard disk. Now how this data is stored at disk sector, segment or in file. Database allows independence from these details. This assignment requires to assume that you have data stored in a file where files have been structured at two levels. One that explains the structure of table and the other that has the data. Say table has 3 fields - id, name & address. A program accessing the data needs to know the file structure, which is implemented in any database system.

You need to create a metadata file named as "**metadata_yourrollnumber.txt**" to store the structure of the data file along with the data file.
Metadata information will be stored in text files. Format of the metadata file is as follows:

Field Name	Field Data Type	Field Size

Field Name = Name of the field (Generally, String)
Field DataType = The data type it should support is Character(or String), Integer, Float and Double
Field Size = Number of bytes required to store the value
For example, consider a user file of flight records with the following information:

ID	**Status**	**Price**
360	Landed	5999.99
217	On time	4786.98

239	Arrived	3682.99

Format description of the data file is as follows:

Field Name	Type
ID	Integer(5)
Status	Character(20)
Price	Float(10)

We would require the below metadata file for the above records:

ID	I	5
Status	C	20
Price	P	10

First column indicates field name, second column indicates its datatype and third column indicates its size.

TO DO:

1. Write a Python/C++ program named as "**program1_yourrollnumber.py/.cpp**" that does not use any metadata file and the script already

has information of record structures embedded within it.**[2 marks]**

0.5 mark - print the contents of the data file

1 mark - compute sum of value of a field if it's numeric type.

0.5 mark - throwing an error if non numeric field passed as an argument]

2. Write a Python/C++ program named as "**program2_yourrollnumber.py/.cpp**" to achieve physical data independence using metadata.

a. Read the metadata file and process the structure.

b. Read input file from the user and write it to the user file according to the format given in metadata file.

Both the programs should have following functionality for a user:

I. Print the content of the data file.

II. Given a fieldname as an argument, it should compute the sum of the value of a specific field if the given field is of numeric type. In case of non-numeric field, it

should report an error.

Points to Demonstrate:

1. In case of 2nd program we should not be allowed to change the program even if we change the structure of the data in terms of addition of a new attribute, resizing an attribute or changing the order of the attributes. **[5 marks - Binary]**

2. Show that the 1st program will be efficient as compared to the 2nd program. For this you need to generate a data file with random data values. However, you can use any record Structure.

[3 marks for Viva. Three levels for grading - 0,1,3 marks].