

---

# CS246: Database Management Systems Lab

Lab # 06 (1 Questions, 50 Points)

Timings: 14:00 to 17:00 Hours Pages: 3

IIT Guwahati

11 Feb 2020 (Tue)

---

## Question 1: (50 points)

Implement the following problem in either C or C++ Programming Language. Use of any other programming language is not allowed and will lead to awarding 0 marks. Use of standard library functions to perform file operations lead to awarding 0 marks.

**Input Data** You are given two input data files containing the following:

**exam-tt-with-duplicates.csv** Each line contains information about one course and its end semester and mid semester exam details as given below

- Course name (59 characters)
- end semester exam date (11 characters)
- end semester exam time (11 characters)
- mid semester exam date (11 characters)
- mid semester exam time (11 characters)
- Each of the above attribute is separated using comma

*Every row of exam-tt-with-duplicates.csv file contains 107 bytes*

**courses04.csv** Each line contains information about one course. The following attributes are given for each course

- Course number (6 characters)
- Course name (59 characters)

*Every row of courses file contains 67 bytes*

**Observations** **exam-tt-with-duplicates.csv** file contains duplicate entries

**Problem Statement** Your task is to:

**Read** The given two files; You must use `open` system call with *read only* mode.

**Count duplicates** You should count the course names having duplicate entries (a course appearing more than once) in the **exam-tt-with-duplicates.csv** file). Print course id and number of times this course name is appearing in the file **exam-tt-with-duplicates.csv**.

**Course having no duplicates** You should also list courses having no duplicate entries in the file **exam-tt-with-duplicates.csv**. That is course appearing only once in this file.

**Write** The above output into two distinct files as described below:

**Having Duplicates** Produce an output file containing only duplicate counts. Its format should be:

```

cid    frequency
BT101  2
BT206  5
BT207  3
BT208  3
BT209  3
...
...
PH704  5
PH706  4
RT514  4
RT515  3

```

Note that the output should be *sorted* as per the *cid*

**Having No Duplicates** Produce an output file containing having no in `exam-tt-with-duplicates` counts.

```

cid    frequency
BT621  1
BT632  1
CE201  1
CE320  1
...
...
PH102  1
PH204  1
PH544  1

```

Note that the output should be *sorted* as per the *cid*

Note the following:

- You should use `open` system call in write only mode.
- If the file already exists, you must over-write the contents of file.
- You should always append data to the file.
- You should provide the following permissions at the time of opening the file.

**Owner** Has Read and write permission but no executable permission.

**Group** Has no read, write and executable permission.

**Other** Has no read, write and executable permission.

**Instructions** Adhere to the following

**File naming** Prepend C/C++ program file names with your roll number. Adhere to the input and output file naming convention as given in the problem description.

**Independent efforts** You should make an honest and independent effort in obtaining the solution to the above problem. You are also encouraged to bring one data structures and algorithms text book and one programming language text book of your choice.

**Discussions** with fellow students are not allowed.

**Internet** Use of internet during lab hours is not allowed.

**Mobile phones** Use of mobile phones in the lab hours is not allowed.

**Evaluation** At the end of 17:00 hours, TAs will come and evaluate your program.

Leave the lab once your evaluation is completed.

**Marking Scheme** The evaluation criteria is as follows:

**30 Marks** For producing for duplicate `cids` and their counts and associated program logic

**10 Marks** For producing for unique `cids` and their counts and associated program logic

**10 Marks** Correct output