Indian Institute of Technology Palakkad
Department of Computer Science and Engineering
Operating Systems - Jul to Nov 2020
18 August 2020

Instructions:

- Run the command `$ make` to generate an executable `input_generator`. This program will generate input files for this assignment.
- Run `input_generator` program with appropriate command line arguments to generate csv files to test your script
- The data contained in the csv file is of the format given below

```
file.csv
============
1118010099,CS3010,23
1118010099,CS3010,39
1118010098,CS3010,12
1118010079,CS3030,40
```

  The first column in each of these files contain names of students, the second column contains the subject, and the third column contains their marks in the corresponding subjects.
- *Read the entire question, brain-storm with your team-mates on a high-level strategy to tackle this problem, break each of it down to smaller sub-problems, and write the outline of these steps as comments inside your shell script. After this, partition the steps identified above between the team members. Upload the completed script.sh (should have roll number of the team members) to moodle.*

1. Write a single BASH script (script.sh) to do the following (script.sh can take multiple command-line options):
    a. Generate the executable `input_generator`
    b. Run the executable `input_generator` with appropriate arguments to generate a master csv file (master.csv) file with 100 million (or more) entries
    c. Split this file into 4 csv files holding branch-wise data (one csv file holding entries of all students of a particular branch). The output csv file should be to the following specifications:
        i. The header row should contain course code of all courses offered by the department.

ii. There should be only one entry per student. The marks obtained by the student in a particular course should be shown in the column corresponding to the course code. If a student has not credited a particular course, then that column should not have any value in it.
iii. There should not be any duplicate entries. In case a roll number has the same course code, but different marks in multiple times, only the entry corresponding to the maximum marks should be retained.

d. Split the master csv file into course-wise files (one file per course code named after the course code such as `CS3010.csv`). The output should have the following specifications:
   i. The course-wise files should be inside a folder names `courses/`.
   ii. Each of these course-wise csv files should not have any duplicates. In case a roll number is repeated multiple times in a course, only the maximum of these marks should be retained.