

The answer to each question can be one or more map/reduce jobs. For each job, explain how the mapper and reducer work (what is their input, output, and what they do). When applicable, explain how the partitioner the grouping comparator work. If you are creating a custom class, explain the attributes and the sorting (when applicable). The only global variable that you can use is a single integer (i.e., the number N read, which is a parameter to the main method). You can assume that N records fit in main memory. No actual source code required, but please include pseudo-code where appropriate.

Q1[10]. Consider two files. One file has information about students (ID, name, address, phone number, courses taken).

1, John, 123 Main, 233 223 5566, (CSC365 CSC369 CSC469)
is an example.

Consider a second files that has information about courses and their difficulty.

(CSC365, 1)
(CSC369, 1)
(CSC469, 2)
is an example.

Your goal is to print the names and addresses of students that have taken all the top N most difficult classes.

Q2[10]. Consider an input file that has information about students (ID, name, address, phone number, (course taken, grade).

1, John, 123 Main, 233 223 5566, ((CSC365 A) (CSC369 A) (CSC469 B))
is an example.

The problem is to print the N students with the highest GPA. You can assume that you get 4 points for A, 3 points for a B, 2 points for a C, and 1 point for a D, and 0 points for an F. The average GPA will be a real number between 0 and 4.

Q3[10]. Consider the following input file.

Enrolled (student id, course name). For example: (1,CSC354) means that John is enrolled in CSC354.

The problem is to print the top N most popular classes (i.e., classes with the highest enrollment).