

Solve the following problems using Spark/Scala. Do not create any map data structures. Try to avoid using traditional loops when possible.

Consider three files. One file has information about students (ID, name, address, phone number).

1, John, 123 Main, 233 223 5566
is an example.

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

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

Consider a third file that contains the student ID, course, and grade.

For example:

1, CSC365, A
1, CSC369, A
1, CSC469, B

It contains information about student taking a class and earning a grade.

Q1[10] Write a program that finds the names of the students that have taken at least one of the courses with the greatest difficulty. In our example, John has taken such a course.

Q2[10] Write a program that prints the average course difficulty of the classes that are taken by each student. For our example, the program will print **John, 1.33**. Make sure to print average course difficulty of 0 if the student doesn't take any classes (hint, use left outer or right outer join).

Q3[10]. Write a program that prints the top five most difficult classes.

Q4[10]. Write a program that prints the name of the students ordered by GPA in descending order (starting with the student with the highest GPA). The GPA of a student that has taken no courses should be 0 (use left outer or right outer join).