**ASSIGNMENT 17.2**

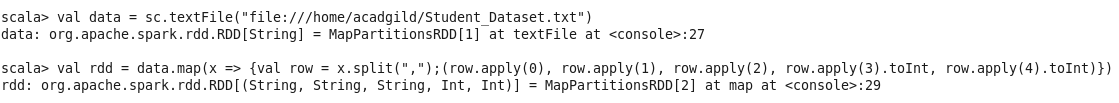
Dataset Link: **https://drive.google.com/file/d/0B\_P3pWagdIrrQnAxZHZKcmxBQW8/view**

PROBLEM STATEMENT 1:

1) Read the text file, and create a tupled rdd.

Solution:

We have read the text file from the above dataset link given and then created a tuple rdd.



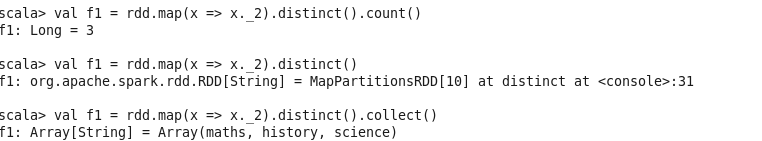
2) Find the count of total number of rows present.

Solution:



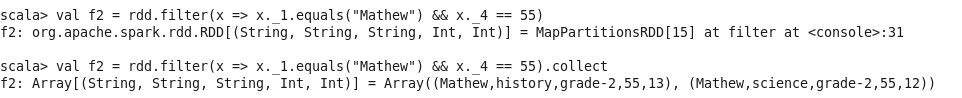
3) What is the distinct number of subjects present in the entire school.

Solution:



4) What is the count of the number of students in the school, whose name is Mathew and marks, is 55

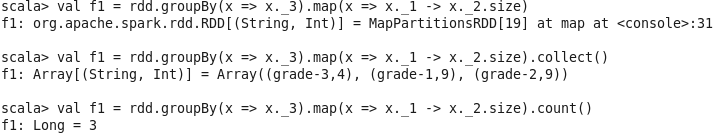
Solution:



PROBLEM STATEMENT 2

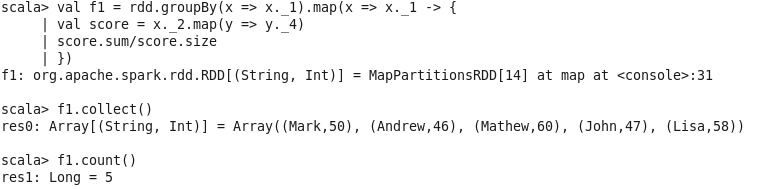
1) What is the count of students per grade in the school?

Solution:



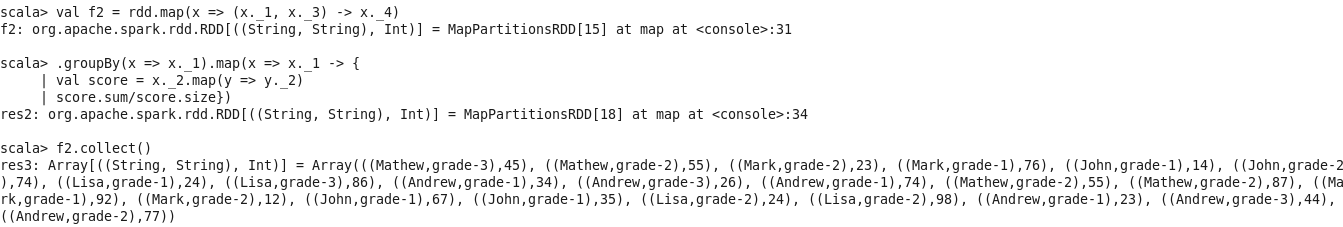
2) Find the average of each student (Note - Mathew is grade-1, is different from Mathew in some other grade!)

Solution:



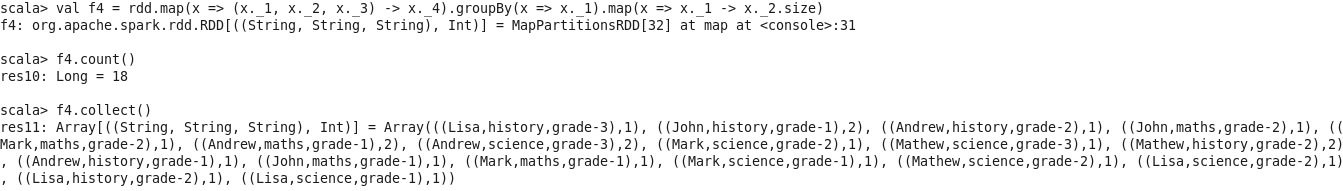
3) What is the average score of students in each subject across all grades?

Solutions:



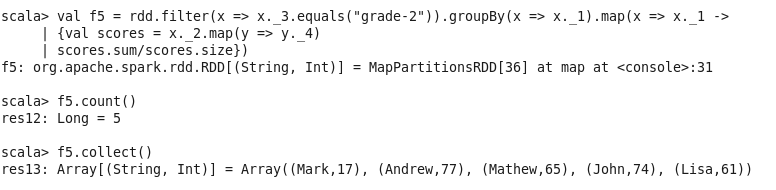
4) What is the average score of students in each subject per grade?

Solutions:



5) For all students in grade-2, how many have average score greater than 50?

Solutions:



PROBLEM STATEMENT 3:

Are there any students in the college that satisfy the below criteria :

1. Average score per student\_name across all grades is same as average score per student\_name per grade

Solutions:

