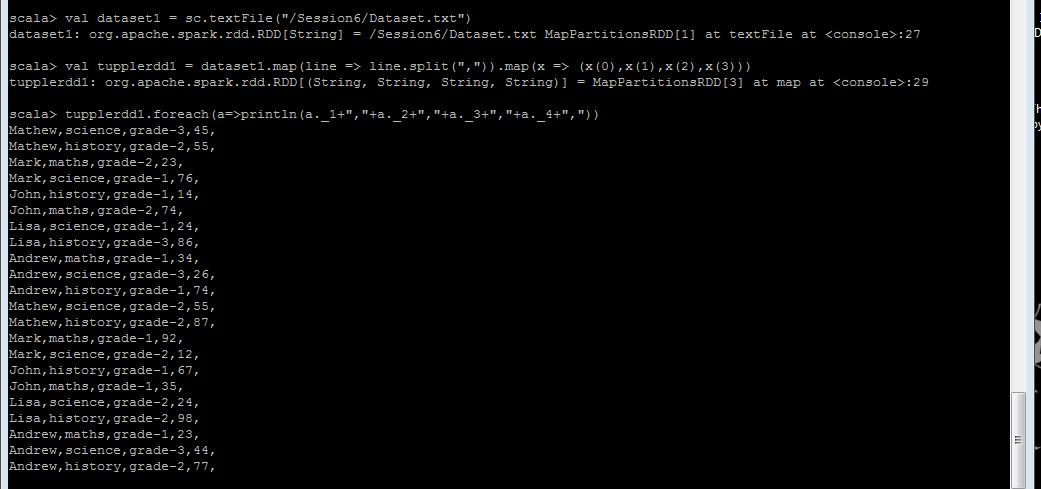
**Session 6 – Assignment 4**

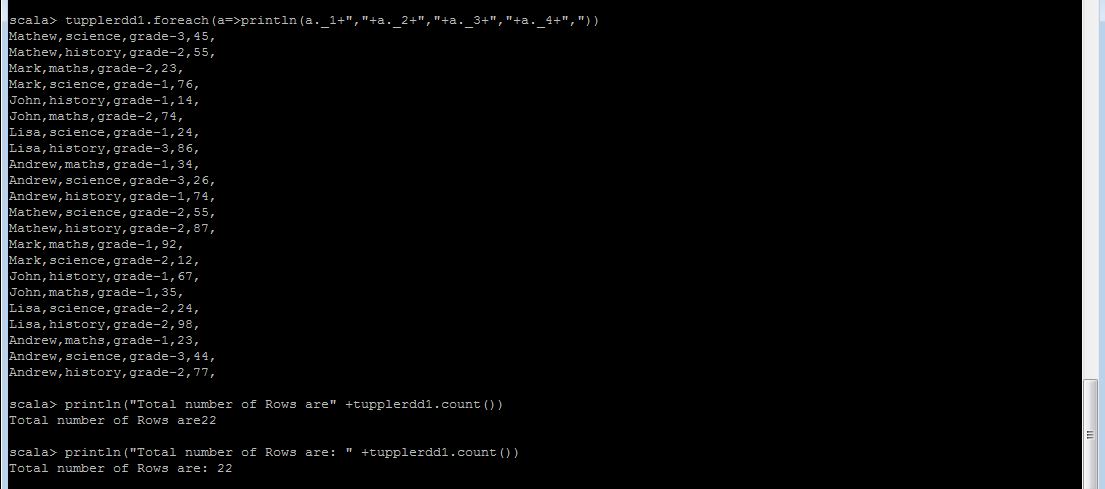
**Problem Statement 1:**

**1. Read the text file, and create a tuppled rdd.**

Read Text file is done with sc.textFile making it an RDD and map applies the function in its param to all element in the RDD and split() create arrays and first 4 arrays are collected to create tuple

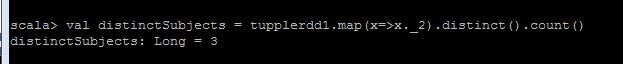


**2. Find the count of total number of rows present.**



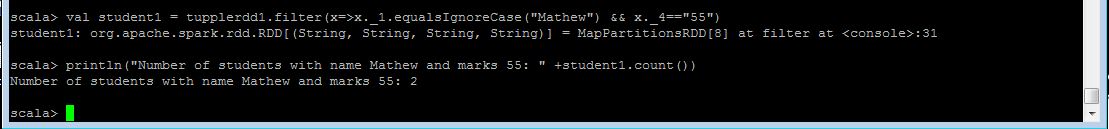
**3. What is the distinct number of subjects present in the entire school**

Use distinct() to get rid of duplicate subject



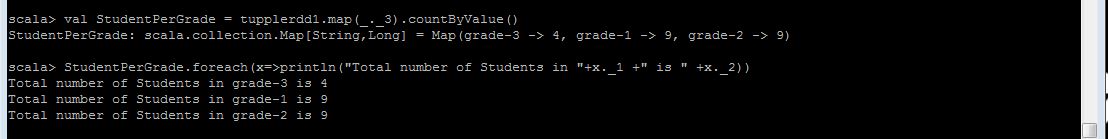
**4. What is the count of the number of students in the school, whose name is Mathew and marks is 55.**

The name field (x.\_1) must contain Mathew and, marks fields (x.\_4) must contain 55

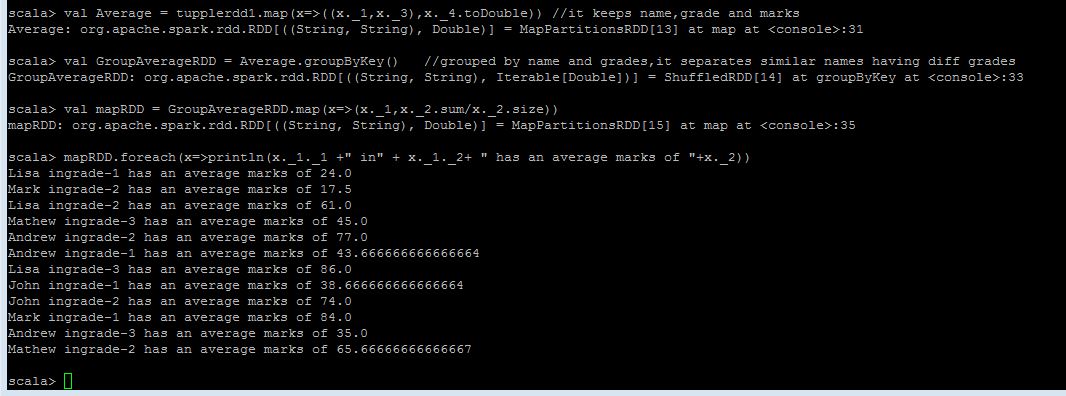


**Problem Statement 2:**

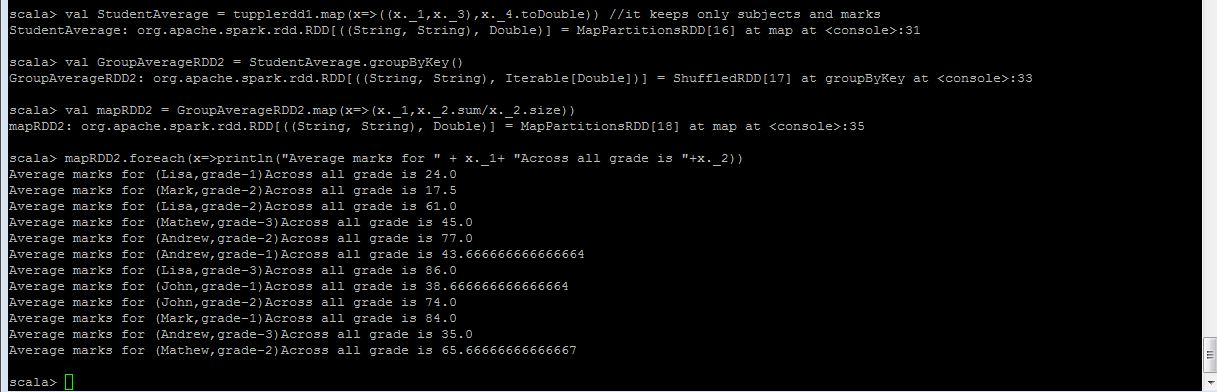
**1. What is the count of students per grade in the school?**



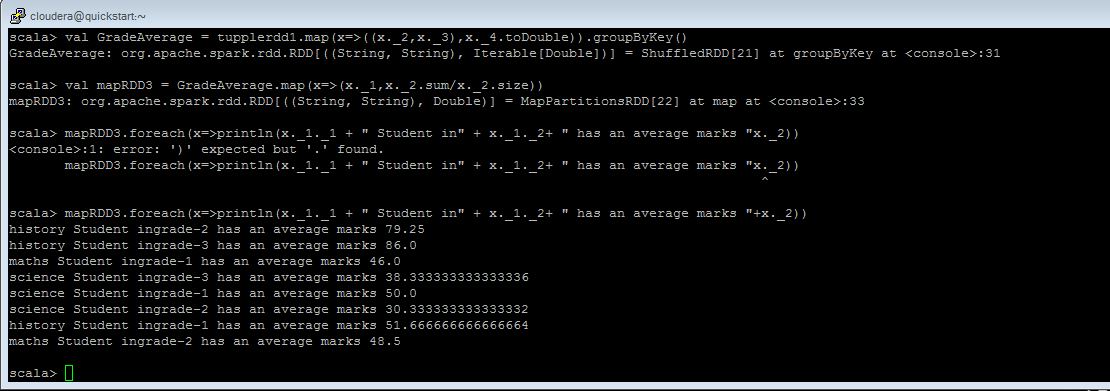
**2. Find the average of each student (Note - Mathew is grade-1, is different from Mathew in some other grade)**



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



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



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

