17EC10060 - Assignment 1 - Question 3

In this part of the assignment, I first saved the given record file into an RDD. To match the records in this file with the earlier log file, I first keyed both RDDs using the name substring, as given in the instructions. Then I used the join command to join these 2 RDDs into a separate RDD. Finally, I used this joined RDD to find the repositories having most number of failed API calls.