17EC10060 - Assignment 1 - Question 1

In this part of the assignment, I first used a case class to form an RDD based on the syntax of each line of the given log file. To obtain the repositories from the log file, I defined a custom mapping function to handle several cases possible, such as the presence of "repos", "Repo", "?", etc. in the rest section of each line of the log file. I mostly used the filter commands, along with the reduceByKey, to print the required statistics.