Query 9

Which are the employers along with the number of petitions who have the success rate more that 70% in petitions and total petitions filed more that 1000?

Platform used: PIG

Input file:

/home/cloudera/h1b\_data

Explanation:

Step1) Filer data from h1b\_data by case status values of CERTIFIED and CERTIFIED-WITHDRAWN

Step2) Group the bag in step 1 by employer name.

Step3) Count the number of CERTIFIED and CERTIFIED-WITHDRAWN petitions in bag from step 3.

Step4) take data from step 1 and group it by employer name.

Step5) find count of the bag on step 4 to get total petitions for each employer.

Step6) JOIN values in step 3 and step 5 in a bag.

Step7) divide success petitions by total petitions

Step8) Generate selected rows for final out and filter the same by given conditions(succ\_rate > 70 and tot\_petitions > 1000).

Pig file:



Output file:

