**Write a pig script to find the count of each hash tag in the data set**

**Tweetshashtag.pig**

REGISTER '/home/acadgild/Downloads/elephant-bird-hadoop-compat-4.1.jar';

REGISTER '/home/acadgild/Downloads /elephant-bird-pig-4.1.jar';

REGISTER '/home/acadgild/Downloads /json-simple-1.1.1.jar';

load\_tweets = LOAD '/user/tweets/' USING com.twitter.elephantbird.pig.load.JsonLoader('-nestedLoad') AS myMap;

extract\_details = FOREACH load\_tweets GENERATE FLATTEN(myMap#'entities') as (m:map[]),FLATTEN(myMap#'id') as id;

describe extract\_details;

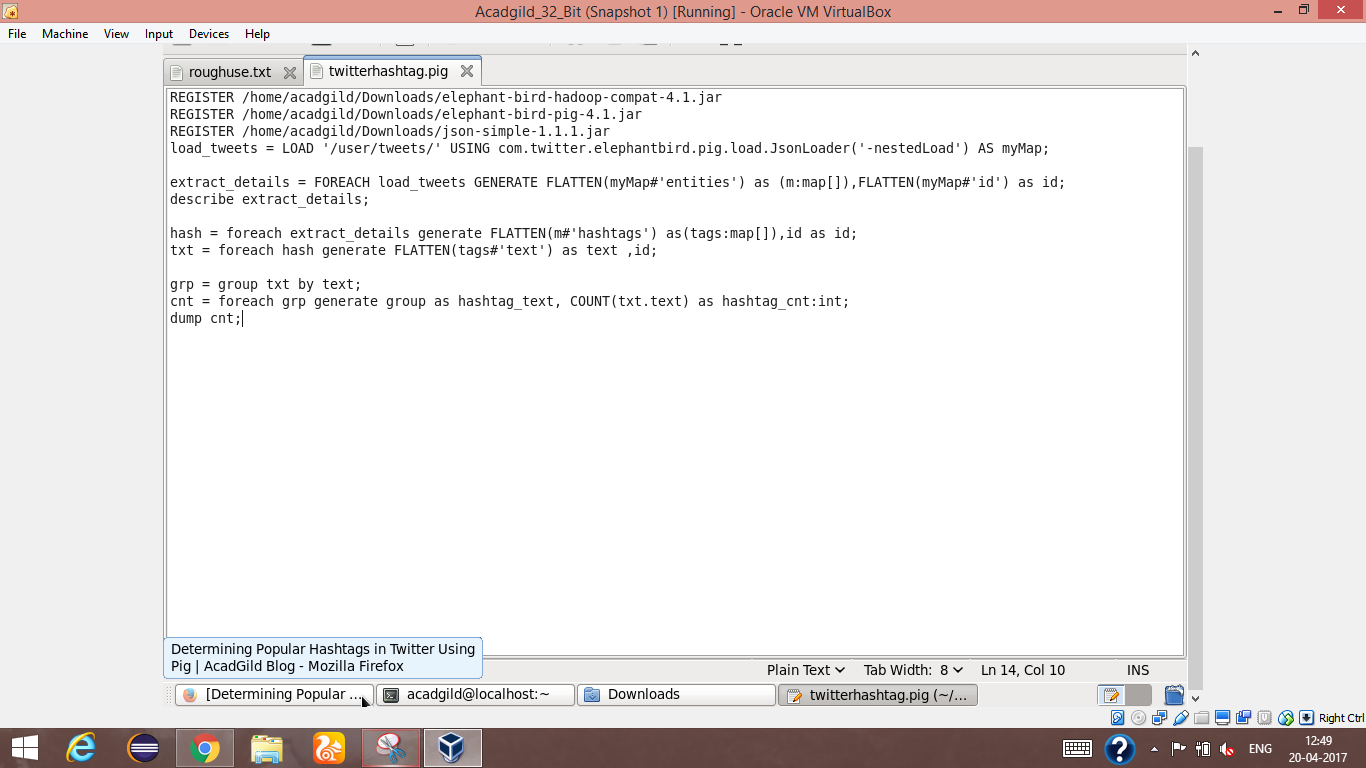
hash = foreach extract\_details generate FLATTEN(m#'hashtags') as(tags:map[]),id as id;

txt = foreach hash generate FLATTEN(tags#'text') as text ,id;

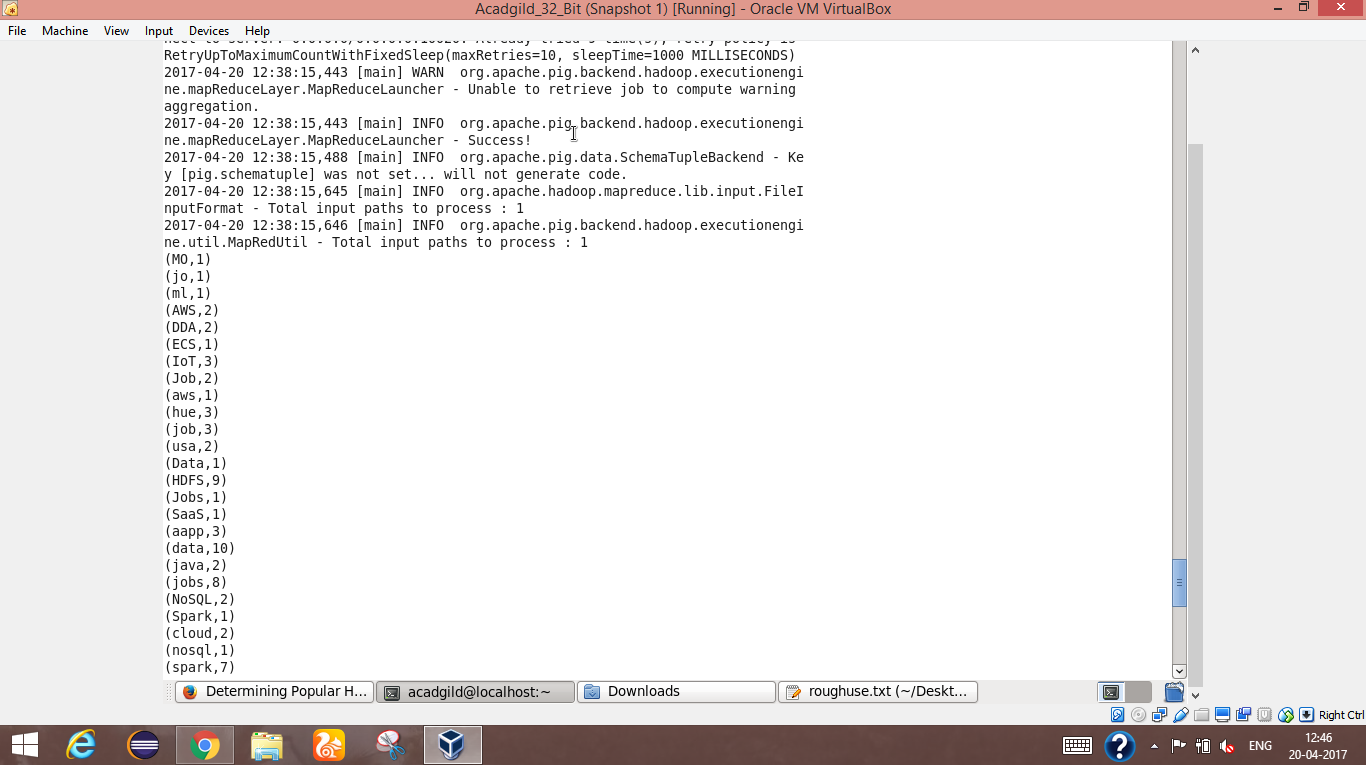
grp = group txt by text;

cnt = foreach grp generate group as hashtag\_text, COUNT(txt.text) as hashtag\_cnt:int;

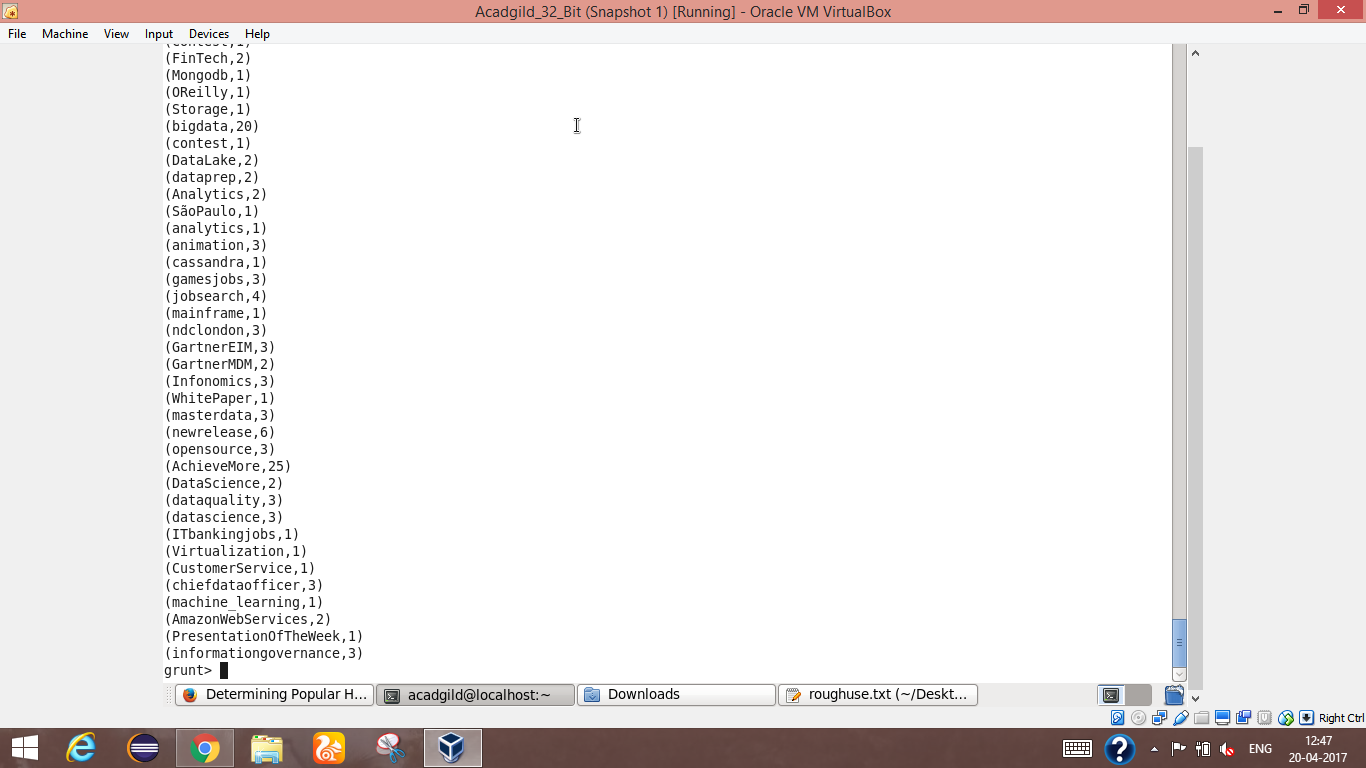
dump cnt ;



Screen which shows the pig script for counting hashtag



Output screen1 which shows hashtag counts



Output screen2 which shows hashtag counts