**Case Study Apache Pig**

1. Hdfs -mkdir -p /casestudy
2. Hadoop fs –put ‘/home/cloudera/Desktop/users.csv’ ‘/casestudy’
3. Hadoop fs –put ‘/home/cloudera/Desktop/tweets.csv’ /casestudy
4. Tweets=LOAD ‘/casestudy/tweets.csv’ USING PigStorage(',') as(tweet\_id:int,tweets:chararray,user\_name :chararray);

1.Write a Pig Latin Query that outputs the login of all users in “NY” state.

2.Write a Pig Latin Query that returns all the tweets that include the word “favorite” ordered by tweet id.

3.Write a Pig Latin Query that computes the natural join between the two collections using the reduce side join approach.

4.Write a Pig Latin Query that returns the number of tweets for each user name (not login) .You should output one user per line, in the format (user\_name, number\_of\_tweets).

5.Write a Pig Latin Query that returns the number of tweets for each user name (not login). Ordered from most active to least active users. You should output one user per line, in the format (user\_name,number\_of\_tweets).

6.Write a Pig Latin Query that returns the name of the users that posted at least two tweets. You should output one user name per line.

7Write a Pig Latin Query that returns the name of users that posted no tweets. You should output one user name per line.