You can find two input files tweets.csv and users.csv attached

Each line in the user collection (users.csv) contains: login, name and state from a specific user. Each line in the collection of tweets (tweets) has the tweet id, content, and a reference to the user who wrote that tweet.

Assignment 7   (Due 26th Sep, 2019)

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 MapReduce program that computes the natural join between the two collections, using the reduce-side join approach.

4. Write the equivalent join using Pig Latin.

Assignment 8 (continuation from Assignment 7)   (Due 03rd Oct, 2019)

5. 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 following format:

    user_name, number_of_tweets

6. 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 following format:

    user_name, number_of_tweets

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

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