Post-Midsem Lab 1

Date 06/04/2018

Please find the question numbers corresponding to your roll number from the list provided below and attempt only those questions. Also, go through all the instructions carefully.

1. Given the hashtag, **BoxingDay** ,find all popular locations of tweets where the hashtag is being used. Sort the locations in decreasing order of frequency. Your output should include the following:

   Hashtag – Location – Number of times the hashtag is appearing in a tweet from the location

   At the end, display the total number of such rows obtained

2. Given the hashtag, **1YearOfRaees** find all popular user mentions associated with the hashtag. A hashtag and a user mention are said to be associated if they co-occur in same tweet. Sort the user mentions in decreasing order of frequency. Your output should include the following:

   Hashtag – User Mention – Number of times the hashtag and the user mention are co-occurring

   At the end, display the total number of such rows obtained

3. Given the location **Delhi**, find all popular users mentions from that location. Sort the user mentions in decreasing order of frequency. Your output should include the following:

   Location – User mention – Number of times the user mention is appearing from that location

   At the end, display the total number of such rows obtained

4. Given a date **2018-01-15**, find all active userIDs on that date. An user is said to be active on a given data if he has posted a tweet on that date. Sort the userIDs in decreasing order of frequency. Your output should include the following:

   Date – UserID – Number of times the userID has posted a tweet on that date

   At the end, display the total number of such rows obtained

5. Given the hashtag **AUSOpen**, count the number of tweets per day that contains this hashtag between the dates **2018-01-15** to **2018-01-17**. Sort your output in decreasing order of count. Your output should include the following:

   Hashtag – Date – Number of tweets on this date containing this hashtag

   At the end, display the total number of such rows obtained

6. Given the location **Delhi**, count the number of tweets per day from this location between the dates **2018-02-02** and **2018-02-10**. Sort the user in decreasing order of count. Your output should include the following:

   Location – Date – Number of tweets on this date appearing from this location

   At the end, display the total number of such rows obtained

7. Given the userID **923579457767596032**, count the number of tweets the author has posted per day. Sort the output in decreasing order of count. Your output should include the following:

   UserId – Date – Number of times the user has posted on that date

   At the end, display the total number of such rows obtained

8. Given a date, **2018-01-15** find all co-occurring mention-hashtag pairs. A mention and hashtag is said to be a co-occurring pair if they co-appear in the same tweet. Sort your output in decreasing order of frequency. Your output should include the following:

   Date - (mention-hashtag) - Number of times the mention-hashtag pair is co-occurring on this date

   At the end, display the total number of such rows obtained

9. Given a date, **2018-01-15** find all co-occurring hashtag-mention pairs. A hashtag and mention is said to be a co-occurring pair if they co-appear in the same tweet. Sort your output in decreasing order of frequency. Your output should include the following:

   Date - (hashtag-mention) - Number of times the hashtag-mention pair is co-occurring on this date

   At the end, display the total number of such rows obtained

10. Given a date **2017-12-03**, find all co-occurring hashtag-location pairs. A hashtag and location is said to be a co-occurring pair if they co-appear in the same tweet. Sort your output in decreasing order of frequency. Your output should include the following:

    Date - (hashtag-location) - Number of times the hashtag-location pair is co-occurring on this date

    At the end, display the total number of such rows obtained

11. Given a date, **2017-12-30** find all co-occurring mention-location pairs. A mention and location is said to be a co-occurring pair if they co-appear in the same tweet. Sort your output in decreasing order of frequency. Your output should include the following:

    Date - (mention-location) - Number of times the mention-location pair is co-occurring on this date

    At the end, display the total number of such rows obtained

Instructions:

1. Attempt only the questions corresponding to your roll number.
2. It is expected that your computer is installed with the requireddatabase (Cassandra). You are not allowed to access the database from other machines.
3. Time allowed for implementation is 90 minutes (9:00 am to 10:30 am). The remaining time is for evaluation.
4. At the time of evaluation, you have to show the output and submit the following files in the presence of the TA:
    a. Table schema for both questions. The schema should be the one copied from the Cassandra console.
    b. Output files. The output for both the questions in csv format. You may create the csv files in the program itself or manually copy it from the web-interface. Create a separate output file for each question and name it against the question number.
    c. Program files. The python/php/html codes for all the queries should be clubbed in one program file for each filetype. Do not submit any database files or package default files.
    d. Readme file mentioning the type of package and language used as well as any other relevant details.
5. Create a single compressed file (.zip or .tar) that will include all the files mentioned in 4. Name it <roll number>.zip/tar. Send it the the following gmail id: cs3452018@gmail.com
6. Roll number wise allotment of questions is in the following page
7. Marks Distribution : For details on marks distribution, please go through the email sent by the instructor

Roll number wise question distribution:

Roll No.	Questions
130101008	1, 9
130101039	4, 10
140123019	3, 9
150101001	1, 9
150101002	3, 11
150101003	3, 8
150101004	6, 10
150101005	1, 11
150101006	6, 8
150101007	7, 9
150101008	7, 10
150101009	2, 8
150101010	3, 11
150101011	7, 9
150101012	3, 9
150101013	7, 10
150101014	1, 10
150101015	2, 8
150101016	2, 8
150101017	5, 11
150101018	7, 8
150101019	7, 11
150101020	5, 8

150101021	1, 9
150101022	3, 10
150101023	6, 8
150101024	3, 8
150101025	5, 11
150101026	5, 10
150101027	4, 10
150101028	5, 11
150101029	2, 11
150101030	3, 11
150101031	6, 10
150101032	3, 10
150101033	4, 9
150101034	4, 8
150101035	3, 10
150101036	5, 11
150101037	7, 9
150101038	1, 9
150101039	7, 11
150101040	7, 10
150101041	6, 9
150101042	1, 10
150101043	6, 11
150101044	5, 10
150101045	3, 8
150101046	4, 10
150101047	2, 10
150101048	3, 9

150101050	5, 10
150101051	1, 10
150101052	2, 9
150101053	3, 11
150101054	4, 11
150101055	7, 8
150101056	2, 10
150101057	2, 9
150101058	4, 9
150101059	2, 10
150101060	4, 11
150101061	7, 8
150101062	6, 11
150101063	4, 11
150101064	2, 11
150101065	5, 9
150101066	3, 8
150101067	7, 11
150101068	1, 10
150101069	7, 11
150101070	7, 11
150101071	2, 8
150101072	5, 8
150101073	6, 10
150101074	6, 8
150101075	4, 10
150101076	6, 8
150101077	1, 8

150101078	7, 10
150101079	6, 11
150101080	2, 9
150101081	2, 9
150101082	5, 11
150101083	5, 10
150101084	3, 9
150101085	1, 11
150101086	7, 9
150101087	5, 8
150101088	1, 11
150122016	5, 8
150123001	4, 8
150123002	5, 9
150123003	2, 11
150123004	1, 8
150123005	6, 9
150123006	4, 11
150123007	6, 9
150123008	1, 9
150123009	7, 9
150123010	1, 11
150123011	7, 8
150123012	5, 9
150123013	3, 10
150123014	5, 9
150123015	1, 8
150123017	6, 10

150123018	3, 8
150123019	2, 10
150123020	2, 9
150123021	5, 10
150123022	4, 9
150123023	4, 8
150123024	2, 11
150123025	4, 10
150123027	1, 8
150123028	4, 8
150123029	6, 11
150123030	6, 8
150123031	2, 10
150123032	6, 9
150123035	3, 10
150123036	1, 11
150123037	1, 10
150123038	6, 10
150123039	4, 9
150123040	4, 8
150123041	5, 9
150123042	4, 11
150123043	2, 8
150123044	7, 10
150123045	2, 11
150123046	4, 8
150123047	6, 11
150123048	7, 8

150123049	3, 9
150123050	5, 8
150123051	6, 9
150123052	3, 11
150123053	4, 9
150205018	1, 8