ANALYSIS OF CELEBRITIES PROFILE

Description:

This data set contains the data about celebrities' profiles of Twitter.

Parameters:

1. **numeric_id:** A unique id number for every profile.
2. **Verified:** Status of id verified or not.
3. **profile_sidebar_fill_color:** The color for sidebar of profile.
4. **profile_text_color:** The color of Text for every profile.
5. **followers_count:**Total number of followers of each celebrity.
6. **Protected:** Profiles are protected or not.
7. **Location:** Current location mentioned in the profile.
8. **profile_background_color:** Background color of the profile.
9. **utc_offset:** The difference in hours and minutes from Universal Time (UTC) for their current location.
10. **statuses_count:**Total status count of each celebrity.
11. **Description:** The status description of a person.
12. **friends_count:** Total Friends of Celebrities
13. **profile_link_color:** The color of profile link.
14. **profile_image_url:** Image URL of the profile.
15. **Notifications:** The notification of the profile.

16. **profile_background_image_url:** Background image url

17. screen_name

18. profile_background_tile

19. favourites_count

20. **name:** name of the celebrity

21. **url:** Profile URL.

22. **created_at**: when the profile is created

23. time_zone:

24. profile_sidebar_border_color:

25. **following:** To whom they are following.

26. gender

Problem and Solutions:

1. Find the lines that contains http.

2. Find the twitter profile image link for the accounts in which users uses an uploaded image as profile image not the default image.

    a. If uploaded image is used, he will have image link some thing like this: http://s3.amazonaws.com/twitter_production/profile_images/

3. Find the total number of celebrities that have less than 3000 followers.

4. Perform the WordCount in the file and store the output into the HDFS.

5. Find the celebrity profile name whose profile is verified (i.e. true) and location is San Francesco.

6. Find the celebrity name who has the maximum followers.