Assignment

1. Perform analysis on Real-time dataset using Tableau.
2. Perform analysis on Real-time dataset using POWER BI.
3. Classifying Agricultural Crop Pest Data Using Hadoop MapReduce Based C5.0 Algorithm(Link: <https://www.researchgate.net/profile/Revathy-Rathinasamy/publication/332216974_Classifying_Agricultural_Crop_PestData_Using_Hadoop_MapReduceBased_C50_Algorithm/links/5ca6f84d299bf118c4b33aa7/Classifying-Agricultural-Crop-PestData-Using-Hadoop-MapReduceBased-C50-Algorithm.pdf)Check> this paper and write observations(Roll Numbers 1 to 20)
4. <https://link.springer.com/article/10.1186/s40537-022-00659-3(Check> this paper and write your observations )(21 to 40)
5. https://link.springer.com/article/10.1007/s11042-023-17381-8(**Efficient and secure**medical big data management system using optimal map-reduce framework and deep learning) this paper and write observations(Roll Numbers 41 to LE6)
6. Consider the following tables.

**User Table**: This table could store information about Twitter users. Each row could represent a unique user, with columns for attributes such as user ID, username, bio, follower count, following count, etc.

RowKey: User\_ID

----------------------------------------------

| User\_ID | Username | Bio | Follower\_Count | ...

----------------------------------------------

| 123456 | user123 | ... | 1000 | ...

| 789012 | user789 | ... | 500 | ...

**Tweet Table**: This table could store individual tweets. Each row could represent a tweet, with columns for attributes such as tweet ID, user ID (foreign key), timestamp, content, retweet count, favorite count, etc.

RowKey: Tweet\_ID

-----------------------------------------------------

| Tweet\_ID | User\_ID | Timestamp | Content | Retweets | ...

-----------------------------------------------------

| 1001 | 123456 | ... | ... | 50 | ...

| 1002 | 789012 | ... | ... | 20 | ...

**Hashtag Table**: This table could store hashtags used in tweets, along with references to the tweets they appear in. Each row could represent a unique hashtag, with columns for the hashtag text and a list of tweet IDs containing that hashtag.

RowKey: Hashtag\_Text

--------------------------------

| Hashtag\_Text | Tweet\_IDs |

--------------------------------

| #BigData | 1001, 1005... |

| #Analytics | 1002, 1008... |

**User Activity Table**: This table could store user activity, such as tweets, retweets, replies, etc. Each row could represent a user's activity, with columns for attributes such as activity ID, user ID, timestamp, activity type, etc.

RowKey: Activity\_ID

--------------------------------------------

| Activity\_ID | User\_ID | Timestamp | Type |

--------------------------------------------

| 5001 | 123456 | ... | Tweet|

| 5002 | 789012 | ... | Retweet|

Queries

1. Create above tables in HBase Shell.
2. Query to retrieve information about a specific user based on their user ID 12345
3. Query to retrieve the content of a specific tweet based on its tweet ID.
4. Query to retrieve all tweets posted by a specific user.
5. Query to retrieve all tweets containing a specific hashtag.
6. Query to count the number of retweets for a specific tweet.
7. Query to retrieve all activities (tweets, retweets, etc.) by a specific user
8. Query to retrieve tweets posted within a specific time range
9. Query to retrieve the top N tweets with the highest retweet count
10. Query to retrieve the top N tweets with the highest retweet count

Note:Assume that all queries are executed in HBase Shell.