Hadoop 5: Assignment

Problem statement:

You work for Spring Analytics Pvt. Ltd., a company that specializes in Big Data Analytics. You have subscribed to Gnip data feed for Twitter and have received the feeds in the JSON format (**not in real time**). Using Hive JSON Serde, you have to extract the following fields.

Link for the dataset and resources: https://intellipaat-course-attachments.s3.ap-south-1.amazonaws.com/Hadoop/hadoop_dataset.rar

Instruction:

You have to use 'Hive-Assignment.rar' to get the Flume JSON feed dataset.

Fields to be extracted:

- preferredUsername –> User ID
- displayName –> Display name
- friendsCount –> Friends count
- followersCount –> Followers count
- body –> Content of the tweet
- generator.displayname –> Application that generated the tweet

Note: Use the CDHSerDe jar in the above link: - hive-serdes-1.0-SNAPSHOT.jar