



## **Hadoop: Hive Assignment - 1**

## 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.

## Dataset:

[https://intellipaat-course-attachments.s3.ap-south1.amazonaws.com/Hadoop/hadoop\\_dataset.rar](https://intellipaat-course-attachments.s3.ap-south1.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