Task No: Use Case 3 Date:29/10/25	Apache Hive for real-world applications	CO3 K3

AIM: To Develop the Real-time data processing using Apache Hive for analysing the Google Map data

PROCEDURE:

It is often the case that data has to be retrieved in real-time from the source. The processing is done as soon as the data is inputted. For example, Google maps processes traffic data in real-time. As soon as it receives information from the source, the data is output onto its application. Develop the Real-time data processing using Apache Hive for analysing the Google Map data.

Apache Hive is not typically used for real-time data processing. Instead, it's designed for batch processing of large datasets. However, there are other tools and technologies better suited for real-time data processing, such as Apache Kafka, Apache Flink, Apache Storm, Apache Spark Streaming, etc.

That said, if you're determined to use Apache Hive for this task, you might consider a workaround by setting up a near-real-time data processing pipeline with Apache Hive. Here's a high-level overview of how you might achieve this:

1. **Data Collection**: Set up a system to collect real-time data from Google Maps. This could involve using Google Maps APIs to fetch real-time traffic data continuously.
2. **Data Transformation and Ingestion**: Use Apache Kafka or a similar messaging system to ingest the real-time data. You can write a custom producer to fetch data from Google Maps and publish it to Kafka topics.
3. **Stream Processing**: Use a stream processing framework like Apache Storm or Apache Flink to process the incoming data stream in real-time. This step involves data enrichment, filtering, aggregation, or any other processing required for your analysis.
4. **Data Storage**: Store the processed data in a temporary location like HDFS or cloud storage.
5. **Batch Processing with Hive**: Periodically, at certain intervals or when a significant amount of data has accumulated, trigger a batch processing job using Apache Hive. This job will read the processed data from the temporary storage, perform further analysis or aggregation if needed, and store the results in Hive tables.
6. **Data Visualization and Analysis**: Use tools like Apache Superset, Tableau, or custom dashboards to visualize and analyze the data stored in Hive tables.

This approach allows you to leverage the strengths of Apache Hive for batch processing while still achieving near-real-time results by using stream processing frameworks for real-time data processing. However, keep in mind that this is a

complex setup, and using dedicated real-time processing tools might be more efficient and straightforward for this use case.

RESULT: Thus developing the Real-time data processing using Apache Hive for analysing the Google Map data is completed.