**NAME :** ESHAN SINGH

**ROLL NO :** TEAD21228

**SUBJECT :** SL-III

**CLASS :** TE

**BRANCH :** AI&DS

**EXPERIMENT NO :11**

**TITLE :**

**Create databases and tables, insert small amounts of data, and run simple queries using Impala**

**THEORY:**

**What is Impala?**

In the context of data science, Impala typically refers to Apache Impala, which is an open-source massively parallel processing (MPP) SQL query engine for data stored in Apache Hadoop clusters. It's part of the Hadoop ecosystem and is developed by the Apache Software Foundation.

Apache Impala allows users to run interactive SQL queries on large datasets stored in Hadoop Distributed File System (HDFS) or Hadoop-compatible file systems like Apache HBase or Amazon S3. It provides low-latency SQL queries, making it suitable for real-time analytics and business intelligence applications.

Impala is often used in conjunction with other tools in the Hadoop ecosystem, such as Apache Hive for data warehousing and Apache Spark for big data processing, to create end-to-end data pipelines for data analysis and machine learning tasks.

**APACHE IMPALA:**

**1. Architecture:** Impala follows a distributed architecture, where it leverages the power of multiple nodes in a cluster to process queries in parallel. It employs a combination of a distributed query planner and a distributed execution engine to efficiently process SQL queries across the cluster.

**2. SQL Compatibility:** Impala provides full SQL support, including complex SQL queries, joins, aggregations, and window functions. This makes it easier for users familiar with SQL to interact with large-scale datasets without needing to learn new query languages or paradigms.

**You can execute these SQL statements using the Impala shell or any other SQL client that supports Impala as follows:**

1. Save the above SQL commands in a file, e.g., impala\_commands.sql.
2. Open a terminal and run the Impala shell: impala-shell
3. Once in the Impala shell, execute the commands from the file: source impala\_commands.sql

**Executing a simple query in Impala:**

1. **Connect to Impala:** Use a tool like impala-shell or a SQL client to connect to your Impala instance.
2. **Create a Table:** Create a simple table in Impala. For example:

CREATE TABLE hello\_world\_table (

message STRING);

1. **Insert Data:** Insert a "Hello, World!" message into the table:

INSERT INTO hello\_world\_table VALUES ('Hello, World!');

1. **Run Query:** Execute a SELECT query to retrieve the message:

SELECT \* FROM hello\_world\_table;

1. **View Results:** View the results of the query, which should display "Hello, World!".

**Running Impala program**

To run an Impala program, first, create SQL commands to define databases, tables, and queries, then execute them either through the Impala shell or programmatically using Impala's JDBC or ODBC drivers in your preferred programminglanguage.

**Conclusion:**

In this practical, we utilized Apache Impala to create databases and tables, inserted sample data, and executed basic SQL queries. This hands-on experience demonstrated Impala's effectiveness for interactive data analysis and manipulation on Hadoop clusters.

**CODE & OUTPUT:**





 