Devanshu Surana
PC-23, Panel C
Batch A BDT
1032210755

BDT Lab Assignment 6

Problem statement:
Install HBase and perform CRUD operations.

Objectives:
1. To learn Hbase concept
2. To perform CRUD operation in HBase

Theory:
Introduction to Hbase with its features

It is an open source, distributed and scalable. NoSQL database that is built on top of the Hadoop DFS. It is designed for handling large volumes of structured and semi-structured data. Here are some of its key features

1. Columnar Storage: Data is stored in a columnar format for efficient column-level access.
2. Scalability: Hbase can scale Horizontally across a cluster for handling massive datasets.
3. Schema Flexibility: It's schema less, allowing storage of data with varying structures in the same table.
4. Strong Consistency: Ensures data consistency for read and write operations within a single row.
5. Automatic Sharding: Data is automatically split and distributed to balanced the workload.
6. Hadoop Integration: Seamlessly integrated with the Hadoop ecosystem for big data applications.

Page No.
Date

Hbase commands:

Hbase provides a set of command-line tools and APIs for interacting with the database. Here are some commonly used Hbase commands:

1. Create table: Use 'create' command, specifying the tablename and column-families.
   eg: create 'mytable', 'cf1', 'cf2'

2. Put data: To insert data in Hbase table, use 'Put' command.
   eg: put 'mytable', 'row1', 'cf1:col1', 'value1'.

3. Get data: To retrieve data in Hbase table.
   eg: get 'mytable', 'row1'.

4. Scan data: We can scan the entire table or the range of rows.
   eg: scan 'mytable'

5. Delete data: To delet data from the table.
   eg: delete 'mytable', 'row1', 'cf1:col1'

6. List tables: To list all the tables in 'Hbase'
   eg: list

7. Disable and delete table.
   eg: disable 'mytable'
   delete : mytable

Platform: 64-bit open source Linux/Windows

Conclusion: Hence, I learned to install Hbase and perform CRUD operations.

FAQ's.

1. State any four use cases of Hbase.

Ans. 1. Real time Analytics: Instant data analysis for applic-ations like social media monitoring and fraud detect-ion.

2. Ef time series data: Efficient storage and retrieval of time-series data like sensor readings and logs.

3. Catalogs and Recommendations: Managing product catalogs and providing personalized product recommendations in e-commerce

4. clickstream Analysis: Analyzing user clickstream data for website optimization and targeted advertising.

2. What are some of the challenges of using Hbase?

Ans. i) Complexity: It can be complex to set up and manage

ii) Consistency vs Scalability.

iii) Data Modelling: Designing Hbase data models demands expertise.

iv) Operational Overhead: Maintainence tasks like data compaction can be resource-intensive.

3. What are the Hierarchy of tables in Hbase?

Ans. i) Name space: It provides a way to organize tables into logical groups.

ii) Table: It is a collection of a rows, each identified by unique row.

iii) Column: Contains actual data with unique column qualifiers

iv) Column, families: Organize data within tables.