Name : Tejas Redkar
Roll NO : BDT2 (22)
Panel - C
PRN : 1032210937

BDT Lab Assignment - 6

* Problem statement :

Install Hbase & perform CRUD operations.

* Objectives :

1) To learn Hbase concept.
2) To perform CRUD operations in HBase.

* Theory :

Introduction to Hbase with its features:
It is an open source, distributed & scalable NoSQL database that is built on top of the Hadoop DFS. It is designed for handling large volumes of structured & 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 ~~shetter~~ cluster for handling massive datasets.

3) Schema flexibility: It's a schemaless, allowing storage of data with varying structures in the same table.

4) Strong Consistency: Ensures data consistency for read & write operations within a single row.

5) Automatic Sharding: Data is automatically split & distributed to balanced the work load.

6) Hadoop Integration: Seamlessly integrated with the Hadoop ecosystem for big data applications.

- Hbase commands:

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

1) Create table: Use "create" command, specifying the table name & column-families

e.g. create 'mytable', 'CF1', 'CF2'

2) Put data: To insert data in Hbase table, use 'Put' command

e.g. put 'mytable', 'row1', 'CF1:col1', 'value1'

3) Get Data : To retrieve data in Hbase table
e.g. get 'mytable', 'row1'.

4) Scan Data : We can scan the entire table or the range of rows.
e.g. scan 'mytable'

5) Delete Data : To delete data from the table
e.g. delete 'mytable', 'row1', 'CF1:col1'

6) List tables : To list all the tables in Hbase
e.g list

7) Disable & Delete tables:
e.g disable 'mytable'
delete 'mytable'

* Platform : 64-bit open source Linux / Windows.

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

* FAQ's

Q1) State any four use cases of Hbase

Ans 1) Real time analytics : Instant data analytics for applications like social media monitoring & fraud detection.

2) EF-time-series data : Efficient storage & retrieval of time-series data like sensor readings & logs

3) Catalogs & Recommendations: Managing product catalogs & providing personalized product

recommendations in e-commerce.

4) Clickstream Analysis: Analyzing user clickstream data for website optimization & targeted advertising.

Q2) What are some of the challenges of using HBase?

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

ii) Consistency & V/S Scalability.

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

iv) operational Overhead: Maintainence tasks like data compaction can be resource intensive

Q3) What are the hierarchy of tables in Hbase?

Ans 1) Namespace: It provides a way to organize tables into logical groups.

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

3) Columns: Contains actual data with unique column qualifiers.

4) Column, families: Organize data within tables.

21/11/23