**Big Data Analytics**

**BAD601**

**Experiment 6:**

Use Hive to create, alter, and drop databases, tables, views, functions, and indexes.

**Start Hadoop**

Open Command Prompt as Administrator.

Go to the Hadoop sbin folder:

cd C:\hadoop-3.3.6\sbin and type

start-all.cmd

Confirm Namenode and Datanode started by checking:

http://localhost:9870/ (Namenode UI)

Also in the Command Prompt type

jps

it should show: NameNode, DataNode, SecondaryNameNode, ResourceManager, NodeManager

**Initialize Hive Metastore (First Time Only)**

If not done already, initialize the schema for Hive using Derby (default):

Open Command Prompt

Go to the Hive bin folder:

cd C:\hive\apache-hive-3.1.2-bin\bin and type

hive --service schematool -dbType derby -initSchema

You should see output like: Initialization script completed

**Start Hive CLI**

hive

Wait for the Hive prompt:

hive>

The Main program:

**Create Hive Script File**

Open **Notepad** or any text editor.

Copy and paste the following Hive commands into it:

-- DATABASE OPERATIONS

-- Create a database

CREATE DATABASE IF NOT EXISTS college;

-- Alter a database by setting properties

ALTER DATABASE college SET DBPROPERTIES ('creator' = 'Surbhi', 'created\_on' = '2025-05-02');

-- Describe the database

DESCRIBE DATABASE EXTENDED college;

-- TABLE OPERATIONS

-- Create a table

CREATE TABLE IF NOT EXISTS students (

id INT,

name STRING,

marks FLOAT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE;

-- Alter the table

ALTER TABLE students ADD COLUMNS (gender STRING);

ALTER TABLE students CHANGE name full\_name STRING;

ALTER TABLE students REPLACE COLUMNS (id INT, full\_name STRING, marks FLOAT, gender STRING, grade STRING);

-- Describe the table

DESCRIBE students;

-- VIEW OPERATIONS

-- Insert data into table

INSERT INTO TABLE students VALUES (1, 'John Doe', 92.5, 'M', 'A'), (2, 'Jane Smith', 75.0, 'F', 'B');

-- Create a view

CREATE VIEW IF NOT EXISTS top\_students AS

SELECT id, full\_name, marks FROM students WHERE marks >= 80;

-- Alter a view (by dropping and recreating)

DROP VIEW IF EXISTS top\_students;

CREATE VIEW top\_students AS

SELECT full\_name, marks FROM students WHERE marks >= 80;

-- Drop the view

DROP VIEW IF EXISTS top\_students;

-- FUNCTION OPERATIONS

-- Create a temporary function (UDF)

CREATE TEMPORARY FUNCTION reverse\_udf AS 'org.apache.hadoop.hive.ql.udf.UDFReverse';

-- Use the function

SELECT reverse\_udf('HiveExample') AS reversed;

-- Drop the function

DROP TEMPORARY FUNCTION reverse\_udf;

-- INDEX OPERATIONS

-- Create a new table

CREATE TABLE IF NOT EXISTS marks (

id INT,

name STRING,

marks INT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE;

-- Create an index

CREATE INDEX idx\_marks ON TABLE marks (marks)

AS 'COMPACT'

WITH DEFERRED REBUILD;

-- Rebuild the index

ALTER INDEX idx\_marks ON marks REBUILD;

-- Drop the index

DROP INDEX idx\_marks ON marks;

-- DROP TABLE students;

-- DROP TABLE marks;

-- DROP DATABASE college CASCADE;

Save the file as:

hive\_script.sql

Save it inside: C:\hadoop-3.3.6\apache-hive-3.1.2-bin\bin (or any location you can navigate to from CMD).

**Run Hive Script File**

Run your script using the command below in the same folder:

hive -f hive\_script.sql

This will execute all the Hive commands in the file one by one.