**Objective:**

This document outlines the steps to create a directory in **HDFS**, upload data files, create an **external Hive table**, and query the data.

**Step-by-Step Instructions**

**1. Create HDFS Directory**

To organize and store your data, create the required directory structure in **HDFS**.

hdfs dfs -mkdir -p /tmp/big\_datajan2025/Dharafinal/external

* **hdfs dfs**: HDFS command-line interface.
* **-mkdir**: Command to create a directory.
* **-p**: Ensures that parent directories are created if they don’t exist.
* **Directory Path**: /tmp/big\_datajan2025/Dharafinal/external

**2. Create and Add Initial Data File**

Create a file named **file.txt** with some sample employee data.

vi file.txt

**Add the following content:**

1,abc,234

2,def,456

**Save and exit** (Press Esc, then type :wq and hit Enter).

**3. Upload the File to HDFS**

Use the following command to upload the file.txt file to the **HDFS directory**:

hdfs dfs -put file.txt /tmp/big\_datajan2025/Dharafinal/external

* **-put**: Uploads a local file to the specified HDFS directory.
* **File**: file.txt
* **Destination**: /tmp/big\_datajan2025/Dharafinal/external

**Verify Upload:**

hdfs dfs -ls /tmp/big\_datajan2025/Dharafinal/external

You should see file.txt listed.

**4. Create an External Hive Table**

Launch the **Hive CLI**:

hive

Run the following **DDL statement** to create an **external table**:

CREATE EXTERNAL TABLE employees\_ext (

id INT,

name STRING,

salary DOUBLE

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '/tmp/big\_datajan2025/Dharafinal/external/';

**Explanation:**

* **CREATE EXTERNAL TABLE**: Creates an external table (Hive doesn’t manage the data, only the metadata).
* **ROW FORMAT DELIMITED**: Specifies how the data is formatted.
* **FIELDS TERMINATED BY ','**: Indicates that fields are separated by commas.
* **STORED AS TEXTFILE**: Data is stored in plain text format.
* **LOCATION**: Specifies the existing HDFS directory for the external table.

**Confirm Table Creation:**

SHOW TABLES;

You should see employees\_ext in the list.

**5. Query the Table**

Now, retrieve data from the table:

SELECT \* FROM employees\_ext;

**Expected Output:**

1 abc 234.0

2 def 456.0

**6. Add More Data**

Now, create a new data file:

vi data.txt

**Add the following content:**

1,gjmj,234686

2,defngj,456876

**Save and exit**.

Upload this new file to the same HDFS directory:

hdfs dfs -put data.txt /tmp/big\_datajan2025/Dharafinal/external

**7. Query the Table Again**

Since the table is **external** and points directly to the **HDFS location**, Hive will automatically include the new data.

SELECT \* FROM employees\_ext;

**Expected Output:**

1 abc 234.0

2 def 456.0

1 gjmj 234686.0

2 defngj 456876.0