

```
# =====
```

```
# 1 Start Hadoop Daemons
```

```
# =====
```

```
$HADOOP_HOME/sbin/start-dfs.sh
```

```
$HADOOP_HOME/sbin/start-yarn.sh
```

```
jps # Verify NameNode, DataNode, SecondaryNameNode, ResourceManager,  
NodeManager
```

```
# =====
```

```
# 2 Prepare Hive Data
```

```
# =====
```

```
mkdir -p ~/hive_data
```

```
# Main products CSV
```

```
cat > ~/hive_data/products.csv << EOF
```

```
1,Mobile,Electronics,15000
```

```
2,Laptop,Electronics,55000
```

```
3,Shirt,Clothing,1200
```

```
4,Shoes,Clothing,2500
```

```
EOF
```

```
# Partitioned CSVs
```

```
cat > ~/hive_data/products_electronics.csv << EOF
```

```
1,Mobile,Electronics,15000
```

```
2,Laptop,Electronics,55000
```

```
EOF
```

```
cat > ~/hive_data/products_clothing.csv << EOF
```

```
3,Shirt,Clothing,1200
```

```
4,Shoes,Clothing,2500
```

```
EOF
```

```
# =====
```

```
# 3 Start Hive CLI
```

```
# =====
```

```
hive
```

```
# =====
```

```
# 4 Hive: Database & Main Table
```

```
# =====
```

```
CREATE DATABASE IF NOT EXISTS productdb;
```

```
USE productdb;
```

```
SHOW DATABASES;
```

```
CREATE TABLE products(
```

```
    product_id INT,
```

```
    product_name STRING,
```

```
    category STRING,
```

```

    price FLOAT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE;

LOAD DATA LOCAL INPATH '/home/admin1/hive_data/products.csv' INTO
TABLE products;

SELECT * FROM products;

# =====

# 5 Hive: Partitioned Table

# =====

CREATE TABLE products_partitioned(
    product_id INT,
    product_name STRING,
    price FLOAT
)
PARTITIONED BY (category STRING)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE;

# HDFS directories for partitions

```

```
hdfs dfs -mkdir -p
/user/admin1/productdb/products_partitioned/category=Electronics

hdfs dfs -mkdir -p /user/admin1/productdb/products_partitioned/category=Clothing
```

Upload partitioned CSVs

```
hdfs dfs -put ~/hive_data/products_electronics.csv
/user/admin1/productdb/products_partitioned/category=Electronics/

hdfs dfs -put ~/hive_data/products_clothing.csv
/user/admin1/productdb/products_partitioned/category=Clothing/
```

Add partitions in Hive

```
ALTER TABLE products_partitioned ADD PARTITION (category='Electronics')
LOCATION '/user/admin1/productdb/products_partitioned/category=Electronics';

ALTER TABLE products_partitioned ADD PARTITION (category='Clothing')
LOCATION '/user/admin1/productdb/products_partitioned/category=Clothing';
```

```
SHOW PARTITIONS products_partitioned;

SELECT * FROM products_partitioned WHERE category='Electronics';
```

```
# =====
```

 Hive: Queries & Functions

```
# =====
```

Arithmetic

```
SELECT product_name, price, price * 1.1 AS price_with_tax FROM products;
```

String & numeric functions

```
SELECT UPPER(product_name) AS upper_name, ROUND(price,0) AS  
rounded_price FROM products;
```

Create and query a view

```
CREATE VIEW expensive_products AS  
SELECT * FROM products WHERE price > 5000;  
SELECT * FROM expensive_products;
```

Create and rebuild index

```
CREATE INDEX idx_price ON TABLE products (price)  
AS 'COMPACT'  
WITH DEFERRED REBUILD;  
ALTER INDEX idx_price ON products REBUILD;
```

```
SELECT * FROM products WHERE price > 2000;
```

```
# =====
```

7 Hive: Export Table to HDFS

```
# =====
```

```
INSERT OVERWRITE DIRECTORY '/user/admin1/hive_output/products'  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ','  
SELECT * FROM products;
```

```
# Verify exported data

hdfs dfs -ls /user/admin1/hive_output/products

hdfs dfs -cat /user/admin1/hive_output/products/*
```

Commands to Execute in the Linux Terminal

Run these **before opening Hive**:

```
# Start Hadoop daemons

$HADOOP_HOME/sbin/start-dfs.sh

$HADOOP_HOME/sbin/start-yarn.sh

jps # Verify Hadoop daemons
```

```
# Create local Hive data files
```

```
mkdir -p ~/hive_data
```

```
cat > ~/hive_data/products.csv << EOF
```

```
1,Mobile,Electronics,15000
```

```
2,Laptop,Electronics,55000
```

```
3,Shirt,Clothing,1200
```

```
4,Shoes,Clothing,2500
```

```
EOF
```

```
cat > ~/hive_data/products_electronics.csv << EOF
```

```
1,Mobile,Electronics,15000
```

```
2,Laptop,Electronics,55000
```

```
EOF
```

```
cat > ~/hive_data/products_clothing.csv << EOF
```

```
3,Shirt,Clothing,1200
```

```
4,Shoes,Clothing,2500
```

```
EOF
```

```
# Create HDFS directories for Hive partitions
```

```
hdfs dfs -mkdir -p
```

```
/user/admin1/productdb/products_partitioned/category=Electronics
```

```
hdfs dfs -mkdir -p /user/admin1/productdb/products_partitioned/category=Clothing
```

```
# Upload CSVs to HDFS
```

```
hdfs dfs -put ~/hive_data/products_electronics.csv
```

```
/user/admin1/productdb/products_partitioned/category=Electronics/
```

```
hdfs dfs -put ~/hive_data/products_clothing.csv
```

```
/user/admin1/productdb/products_partitioned/category=Clothing/
```

Commands to Execute Inside Hive

After running hive in your terminal:

```
hive
```

Now at the hive> prompt, execute the following:

-- Create and use database

```
CREATE DATABASE IF NOT EXISTS productdb;
```

```
USE productdb;
```

-- Create main table and load data

```
CREATE TABLE products(
```

```
    product_id INT,
```

```
    product_name STRING,
```

```
    category STRING,
```

```
    price FLOAT
```

```
)
```

```
ROW FORMAT DELIMITED
```

```
FIELDS TERMINATED BY ','
```

```
STORED AS TEXTFILE;
```

```
LOAD DATA LOCAL INPATH '/home/admin1/hive_data/products.csv' INTO  
TABLE products;
```

```
SELECT * FROM products;
```

-- Create partitioned table

```
CREATE TABLE products_partitioned(
```

```
    product_id INT,
```

```
    product_name STRING,
```

```
    price FLOAT
```

```
)
```


PARTITIONED BY (category STRING)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE;

-- Add partitions (use HDFS paths)

ALTER TABLE products_partitioned ADD PARTITION (category='Electronics')
LOCATION '/user/admin1/productdb/products_partitioned/category=Electronics';

ALTER TABLE products_partitioned ADD PARTITION (category='Clothing')
LOCATION '/user/admin1/productdb/products_partitioned/category=Clothing';

SHOW PARTITIONS products_partitioned;

SELECT * FROM products_partitioned WHERE category='Electronics';

-- Arithmetic, string and numeric operations

SELECT product_name, price, price * 1.1 AS price_with_tax FROM products;

SELECT UPPER(product_name) AS upper_name, ROUND(price,0) AS
rounded_price FROM products;

-- Create and query a view

CREATE VIEW expensive_products AS

SELECT * FROM products WHERE price > 5000;

SELECT * FROM expensive_products;

-- Create and rebuild index

```
CREATE INDEX idx_price ON TABLE products (price)
```

```
AS 'COMPACT'
```

```
WITH DEFERRED REBUILD;
```

```
ALTER INDEX idx_price ON products REBUILD;
```

```
SELECT * FROM products WHERE price > 2000;
```

```
-- Export results to HDFS
```

```
INSERT OVERWRITE DIRECTORY '/user/admin1/hive_output/products'
```

```
ROW FORMAT DELIMITED
```

```
FIELDS TERMINATED BY ','
```

```
SELECT * FROM products;
```