



(12)

Creating HDFS Tables and Joining in Hive

- hdfs_commands.sh

Create directories in HDFS

```
hdfs dfs -mkdir -p /user/hive/warehouse/customers
hdfs dfs -mkdir -p /user/hive/warehouse/orders
```

Put Local files into HDFS

```
hdfs dfs -put customers.csv /user/hive/warehouse/
hdfs dfs -put orders.csv /user/hive/warehouse/orders/
```

- customers.csv

- 1, John Doe
- 2, Jane Smith
- 3, Bob Johnson

- orders.csv

- 1, 2024-01-15, 250.50
- 2, 2024-01-20, 99.99
- 1, 2024-02-10, 180.75
- 3, 2024-03-05, 300.00

hive - create - tables . hql

-- Create external table for customers

```
CREATE EXTERNAL TABLE IF NOT EXISTS customers (  
  customer_id INT,  
  name STRING  
)
```

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

LOCATION '/user/hive/warehouse/customers';

-- Create external table for orders

```
CREATE EXTERNAL TABLE IF NOT EXISTS orders (  
  customer_id INT,  
  order_date STRING,  
  amount Double  
)
```

ROW FORMAT DELIMITED



- Inner Join between customers and orders

SELECT

C.customer_id,

C.name,

O.order_date,

O.amount

FROM customers C

JOIN order O

ON C.customer_id = O.customer_id;