Hive Commands

start-dfs.sh

start-yarn.sh

which hive

cd /usr/local/hive/bin

show databases;

create database flights_info;

use flights_info;

1) Create, Insert & Select:

CREATE TABLE flights (flight_id INT, flight_date STRING,carrier STRING,origin
> STRING,destination STRING,dep_delay INT,arr_delay INT) row format delimited fields
> terminated by ',' stored as textfile;

INSERT INTO TABLE flights VALUES (1001, '2008-01-01', 'AA', 'JFK', 'LAX', 10, 5);

select * from flights;

2) Alter table:

hive> ALTER TABLE flights ADD COLUMNS (flight_duration INT);

3) External Table:

CREATE EXTERNAL TABLE IF NOT EXISTS external_airports (airport_code
STRING,airport_name STRING, city STRING, state string) row format delimited fields terminated
by ',' stored as textfile;

insert into table external_airports values("CSTM","shivaji maharaj","mumbai", "MH");

select * from external_airports;

LOAD DATA LOCAL INPATH '/home/pl/Desktop/external_airport.txt' INTO TABLE
external_airports;

select * from external_airports;

3) Join:

SELECT f.flight_id, a.airport_name
FROM flights f
JOIN external_airports a
ON f.origin = a.airport_code;

4) Average operation:

```sql
SELECT AVG(dep_delay) AS avg_departure_delay
FROM flight;
```

5) Indexing:

```sql
CREATE INDEX idx_airport_code
ON TABLE external_airports (airport_code)
AS 'COMPACT'
WITH DEFERRED REBUILD;

ALTER INDEX idx_airport_code ON external_airports REBUILD;
```