**Hive DDL for Raw and Star Schema Tables**

**Raw Tables**

1. **raw\_user\_logs**

CREATE EXTERNAL TABLE IF NOT EXISTS raw\_user\_logs (

user\_id INT,

content\_id INT,

action STRING,

`timestamp` STRING,

device STRING,

region STRING,

session\_id STRING

)

PARTITIONED BY (year STRING, month STRING, day STRING)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '/raw/logs';

1. **raw\_content\_metadata**

CREATE EXTERNAL TABLE IF NOT EXISTS raw\_content\_metadata (

content\_id INT,

title STRING,

category STRING,

length INT,

artist STRING

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '/raw/metadata';

**Star Schema Tables**

1. **dim\_content**

CREATE TABLE IF NOT EXISTS dim\_content (

content\_id INT,

title STRING,

category STRING,

length INT,

artist STRING

)

STORED AS PARQUET;

1. **fact\_user\_actions**

CREATE TABLE IF NOT EXISTS fact\_user\_actions (

user\_id INT,

content\_id INT,

action STRING,

event\_timestamp TIMESTAMP,

device STRING,

region STRING,

session\_id STRING

)

PARTITIONED BY (year STRING, month STRING, day STRING)

STORED AS PARQUET;

**Data Transformation Commands**

1. **Populating dim\_content**

INSERT OVERWRITE TABLE dim\_content

SELECT

content\_id,

title,

category,

length,

artist

FROM raw\_content\_metadata;

1. **Populating fact\_user\_actions**

SET hive.exec.dynamic.partition=true;

SET hive.exec.dynamic.partition.mode=nonstrict;

SET mapreduce.job.reduces=1;

INSERT OVERWRITE TABLE fact\_user\_actions

PARTITION (year, month, day)

SELECT

user\_id,

content\_id,

action,

CAST(`timestamp` AS TIMESTAMP) AS event\_timestamp,

device,

region,

session\_id,

year,

month,

day

FROM raw\_user\_logs;