Data Archiver

Sqoop command to import selected columns from mysql ‘ads’ table into hdfs location:

- **sqoop import --connect jdbc:mysql://ads-database.czdhajbrjw1o.us-east-1.rds.amazonaws.com:3306/ads_db --username admin --password password --table ads --columns "campaign_id,category,budget,cpm,cpc,cpa,target_device" --target-dir "/home/hadoop/sqoop/" --m 1**

Hive queries to create hive tables and load data into the hive tables from hdfs location:

- **create table ads_info(campaign_id varchar(100), category varchar(50), budget float, cpm float, cpc float, cpa float, target_device varchar(50)) row format delimited fields terminated by “,”;**

- **load data inpath '/home/hadoop/sqoop/part-m-00000' into table ads_info;**

- **create table user_feedback(campaign_id varchar(100),user_id varchar(100),request_id varchar(100),click int,view int,acquisition int,auction_cpm float,auction_cpc float,auction_cpa float,target_age_range varchar(10),target_location varchar(50),target_gender varchar(3),target_income_bucket varchar(3),target_device_type varchar(50),campaign_start_time varchar(50),campaign_end_time varchar(50),action varchar(50),expenditure float,feedback_timestamp timestamp) row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde' with serdeproperties ("separatorChar" = ","quoteChar"="\""escapeChar"="\\") stored as textfile;**

- **load data inpath '/home/hadoop/user_feedback.csv' into table user_feedback;**