DATA INGESTION PIPELINE:
RDS to Amazon Redshift

(.csv)

Spar Nod Bank SRC_ATM_TRANS

Sqoop Ingestion pipeline

MySQL RDS Server

HDFS

Target
schema creation
using dataframes in

PySPark

Data Mart with target Schema

Redshift queries

.csv files

Spark session

Dataframes

Redshift cluster

S3 Bucket