

Tasks

- Extracting the transactional data from a given MySQL RDS server to HDFS(EC2) instance using Sqoop.
- Transforming the transactional data according to the given target schema using PySpark.
- This transformed data is to be loaded to an S3 bucket.
- Creating the Redshift tables according to the given schema.
- Loading the data from Amazon S3 to Redshift tables.
- Performing the analysis queries.