

# Spar-Nord-Bank\_ETL-Project

The project task was to build a batch ETL pipeline - first, to ingest transactional data from RDS into HDFS (using AWS EC2) via Sqoop; next, to transform the data using PySpark (using AWS EC2) to create relevant dimension and fact tables (Data Mart); next, to upload these tables into AWS S3 buckets; finally, load them from S3 into AWS Redshift tables. Lastly, I performed analytical queries on the loaded data, to answer business questions.