Stori Tech Challenge

Description

For this challenge, you are tasked with creating a simple data pipeline to copy data from various data sources to a data warehouse. The pipeline should run on a schedule and copy over data that has been either inserted or updated in the data sources since the last pipeline run.

The data warehouse should consist of a small Redshift cluster (you can use a dc2.large for .25c/hr and shut it down when not using it). Both a SQL database and a NoSQL database should be used as data sources. You can choose whichever databases you'd like for these; but you must import the data we've provided.

The guidelines for the pipeline itself are intentionally open-ended and the pipeline can be implemented using whichever tools you choose. We just ask that you use AWS as a cloud provider for any cloud services you leverage. Once you are finished building the pipeline, please commit all code and relevant files to GitHub.

Data

- txns.csv contains fake bank transactions to be imported and used in the SQL database.
- trades.json contains fake trading data to be used in a NoSQL database.

Constraints

- Must use AWS as the cloud provider for any cloud services.
- Must use Redshift as the data warehouse.
- Must use the datasets that we provide. You can add additional data to them if you'd like and/or use additional datasets to demonstrate your pipeline.