Technical Documentation

Stocks -ETL Project

Overview: This documentation is part of the ETL-Stocks Projects. The goal is to extract data from a web site through an API call and ingest data into a database using sqlalchemy and analyze it and provide the data to other prospective analysts through an API interface.

We extracted the Stocks data for 20 years for the S&N 500 companies and analyzed it. We also extracted the currency exchange rates for the same period.

The goal is to provide the following stocks information in the different currencies.

1. Highest trading companies by year

2. Highest trading company by volume

3. Highest change in value

4. Lowest change in value

5. Average price by company

6. Address all the above in a foreign currency

Project details:

The project is split into the following major categories:

* 1. Identify the subject area
  2. Identify the data the data sources
  3. Data extraction and cleansing
  4. Design and build database
  5. Design and build the website
  6. Analysis and presentation

1. Identify the Subject area: The team reviewed the requirements of the project and considered the following aspects in selecting the project:
   1. Multiple data sources: The data is extracted from two websites as mentioned above in the “Data Sources” section
   2. Calendar table is built in a Spreadsheet

Based on the above data requirements, the team identified Stocks Data analysis with an option to present the information in different currencies.

1. Data Sources:

Source of S&P 500 tickers:

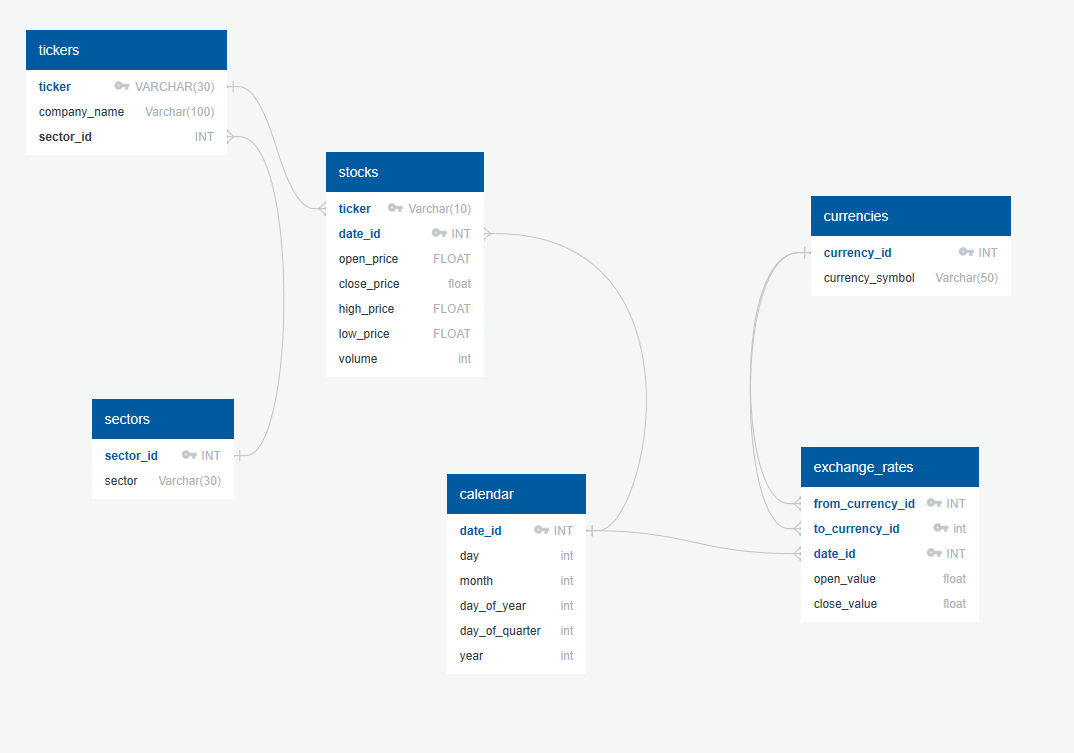
https://datahub.io/core/s-and-p-500-companies#resource-constituents

Stocks data and Exchange Rate:

https://www.alphavantage.co/documentation/# (Stock data & Exchange rate)

1. Data Extraction and Cleansing:
   1. S&P500 Stock Tikers is extracted and loaded in the “Tickers” table
   2. The Stocks data is extracted for the last 20 years and imported in the “Stocks” table. The data is loaded up to the last business day
   3. While the stocks data is extracted the industry to which the Ticker belongs to is added to the “Sectors” table
   4. The calendar data is loaded from the Spreadsheet into the “Calendar” Table
   5. At the end of the process, the code looks for the currencies for which the exchange rates are requested. The exchange rates for the specified currencies is extracted and loaded for the last 20 years.
2. Design and Build of the data base:

The tables were designed and built as per the ERD diagram shown below



5.

1. Query for the stocks trend:

<http://127.0.0.1:5000/endpoints/stock_data/USD/AMZN/2018pwd>

