Database architecture

1. The database(s) you plan to use for storage.

The database used for this final project is PostgreSQL, which is located in (student's) local machine. However, all the information and tables were created on the Amazon Linux shared server..

2. Where the data is loaded from.

This data was loaded from *yfinance* (0.1.55), a library to access the financial data available on Yahoo Finance. The collection of the data was done in Python. Please, see the python file setup. For more information read README file. All the information was saved into the database and created a csv file as previously. The name of this CSV file is *daily_dataPYPL*.

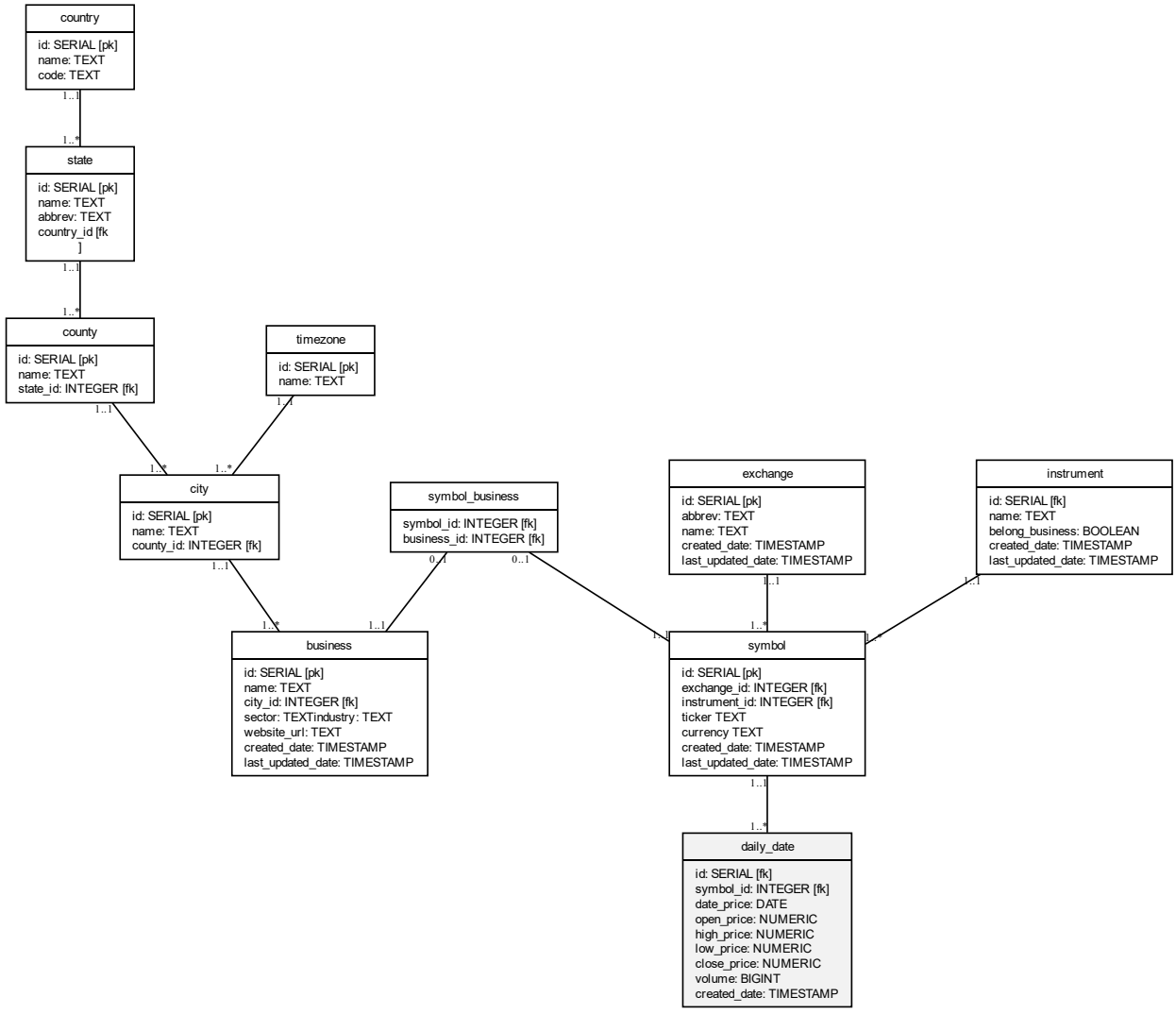
3. Any transformations or scripts to load the data to the database

Data Collection and Data Preparation has been carried out on Python. The libraries used were: pandas (1.0.5), numpy (1.18.5), psycopg2 (2.8.6), and sqlalchemy (1.3.18).
The python files are:

- createTables.py ➔ create all tables into the database.
- setCatalogs.py ➔ import data into tables country, state, city, timezone, exchange type and instrument type.
- setBusinessData.py ➔ import data related to the business. In this case, this data was extracted from *yfinance*. Additionally, it's included information about the symbol (ticker).
- setHistoricalStockData.py ➔ import historical market data. The number of observations depends on how long the business has been on the market.
- setup.py ➔ run all code below.

In the cleaning stage, after setting the frequency of the data (business days) in Python, I found out missing values. After looking into it carefully, these were identified as American holidays. Roughly every year has 11 holidays, which makes sense after I got 53 missing dates (the data collected is made up of 7 years, however, 2015 and 2021 are not completed). To deal with these missing values, I just filled these with the previous value recorded. As an example, the 7th of September of 2015 was Labour Day, therefore, the market was closed by that time.

4. The database schema itself.



** Created using Graphviz. See graphviz.txt to see the structure.*

5. Connections to/from the database for Time Series Analysis module

After creating the database, this one was used to pull the information on the time series analysis project. This also was useful as the data was already cleaned.