Stock Market ETL Project

Group Five

Chris Haub, Ryan Permenter, and Jim Worlein

21 December 2020

Overview

Project to extract, transform, and load stock data into a PostgreSQL database and perform queries on this database.

Data & Modeling Approach

Data was scraped from three tables found on Yahoo Finance and combined with a CSV file from NASDAQ. The three Yahoo tables contain daily stock updates on the most active, highest gainers, and biggest losers. The CSV file contains basic information on each listed company by stock symbol.

These four tables were uploaded to a PostgreSQL database with serial primary IDs. The stock symbol acted as a primary key for all the tables.

NASDAQ CSV File:

(https://www.nasdaq.com/market-activity/stocks/screener?exchange=NYSE&render=download)

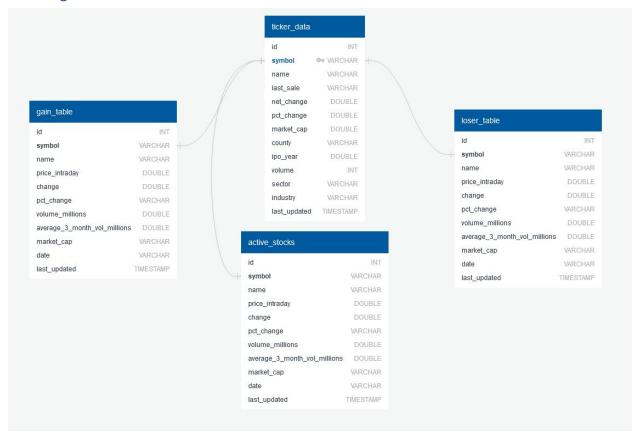
Yahoo Finance:

url = "https://finance.yahoo.com/most-active"

url2 = "https://finance.yahoo.com/gainers"

url3 = "https://finance.yahoo.com/losers"

ERD diagram



SQL Queries

SELECT * FROM active_stocks a INNER JOIN gain_table g ON g.symbol = a.symbol

Data Output			Explain	Messages	Notifications				
4	index bigint	٥	Symbol text	Name text	Price_Intraday double precision	Change double precision	PCT_Change text	Volume_Millions double precision	Avg_3_Month_Vol_Millions double precision
1		16	LAZR	Lumina	35.15	7.53	+27.26%	36.438	7.903
2		19	FUBO	fuboTV	49.48	10.23	+26.06%	32.168	4.175
3		24	QS	Quantu	94.9	21.17	+28.71%	26.559	5.771

Group 5ETL Project

SELECT *

FROM active_stocks a INNER JOIN loser_table I ON I.symbol = a.symbol

Data Output Explain Messages Notifications PCT_Change text Avg_3_Month_Vol_Millions double precision Volume_Millions double precision 7 TSLA 1 Tesla, I... 649.86 -45.14 -6.49% 55.956 48.316 2 23 OXY Occide... 17.47 -1.05 -5.67% 29.116 27.413

SELECT * FROM ticker_data t INNER JOIN active_stocks a ON a.symbol = t.symbol;

4	id integer	symbol text	name text	last_sale text	net_change double precision	pct_change double precision	market_cap double precision	country text
1	4588	NIO	NIO Inc	\$48.95	2.23	4.77	76308459469	[null]
2	17	AAPL	Apple I	\$128.23	1.575	1.24	2223170000000	United States
3	716	BAC	Bank of	\$29.74	1.07	3.73	257275000000	United States
4	10	AAL	Americ	\$16.10	-0.41	-2.48	9833458341	United States
5	2746	GE	General	\$10.77	-0.04	-0.37	94343832210	United States
6	5224	PLTR	Palanti	\$28.51	2.54	9.78	53539891754	[null]
7	2369	F	Ford M	\$8.93	-0.02	-0.22	35527356325	United States
8	6583	TSLA	Tesla I	\$649.86	-45.14	-6.5	616003000000	United States
9	3456	INTC	Intel Co	\$46.36	-1.1	-2.32	189983000000	United States
10	6999	WFC	Wells F	\$29.55	0.54	1.86	122174000000	United States
11	1229	CCL	Carniva	\$21.06	-0.4	-1.86	23250552109	United States
12	2448	FEYE	FireEye	\$20.29	1.055	5.49	4757482553	United States
12	260	AMD	Advans	602.22	2.60	2.0	11225600000	United States

SELECT * FROM active_stocks a LEFT JOIN ticker_data t on t.symbol = a.symbol WHERE t.ipo_year = 1999;

Data Output		Explain		Messages		Notifications					
4	id integer	1	symbol text	<u></u>	name text	price_intraday double precision	change double precision	pct_change text	volume_millions double precision		
1	1	9	PLUG		Plug P	32.37	0.74	+2.36%	35.16		