

DATA 605 Project

Alpha Vantage and Stocks

By: Aditi, Carlos & Precious



Introduction

Alpha Vantage

- Financial data provider
- Real-time and historical stock, forex, cryptocurrency, and technical indicators' data



Objective

- Big Data Infrastructure for extraction, transformation, loading (ETL), computing, scheduling and container orchestration of stock market data for SP500 index
 - Integrating various tools like Docker, MongoDB, Airflow, Dask and Python libraries
-

APIs

Alpha Vantage & MongoDB

- Stock Market Data
 - Structure and Difficulties
 - API Limitations
 - MongoDB
 - Custom Wrapper
-

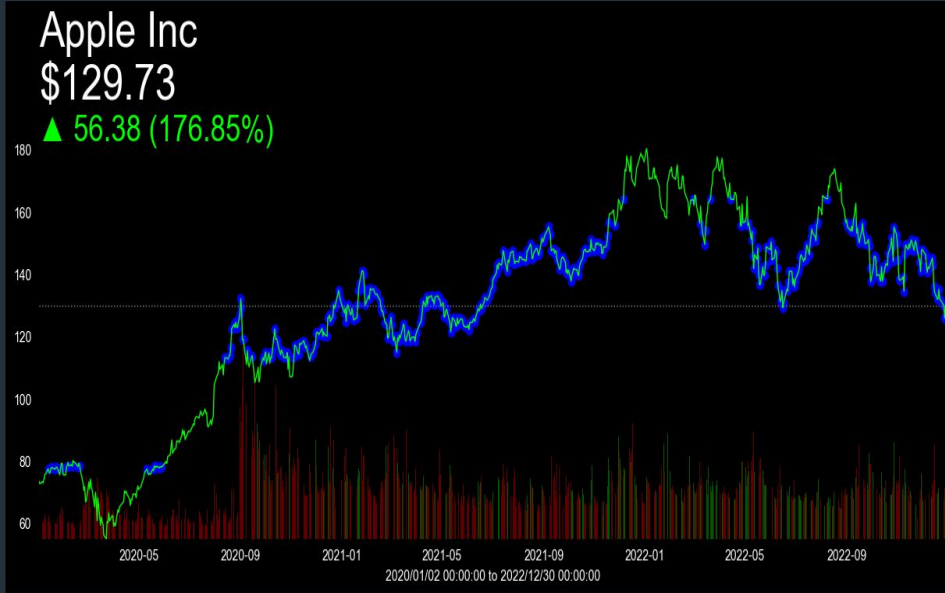
Features

Graphs, Outliers & More

- Outlier detection
 - Isolation Forests & Quantiles
 - Graphs to display data
 - Including outlier detection
 - Additional Computed Properties
 - Moving Averages
 - Relative Strength Indices
-

Example

Anomaly detection using
Isolation forests

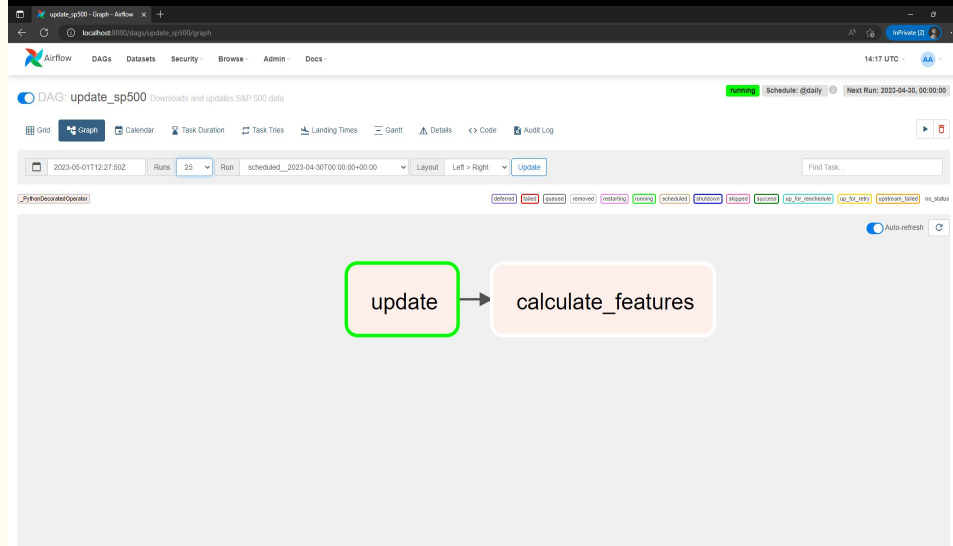


<https://drive.google.com/file/d/14ZUBt-7SxZwwh7zeYQZUZgs4prIcjuGr/view?usp=sharing>

Scheduling

Airflow

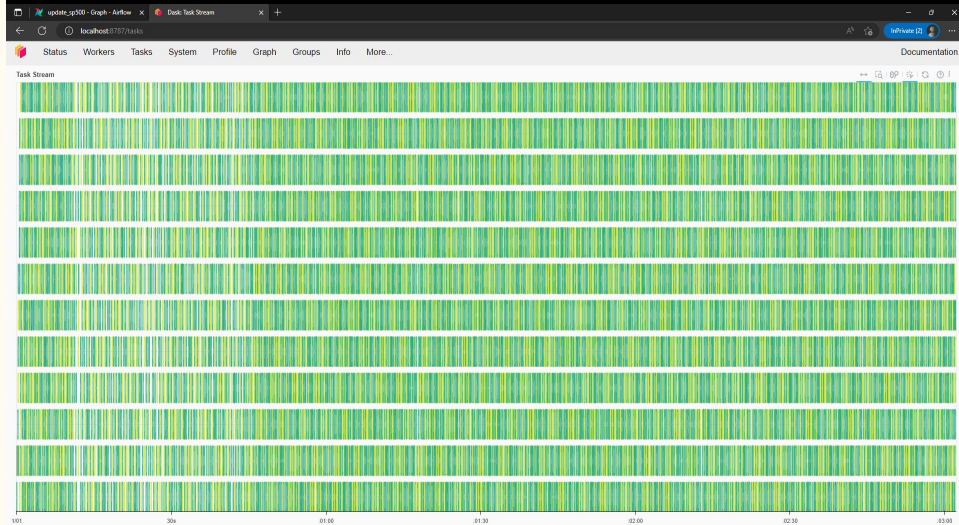
- Two tasks per day
 - Download latest S&P 500 data
 - Compute and save features



Scheduling

Dask

- Parallel processing framework
- Limited by Alpha Vantage API
- Sets up a scalable system



Thank you