

Streaming Data Analysis Using Multi-Cloud Strategy

Ft. [r/wallstreetbets](#)

What is Multi-Cloud?

Multi-cloud is a strategy where an organization leverages two or more cloud computing platforms to perform various tasks.



Why Multi-Cloud?

- ✓ Choose best service from each provider
- ✓ Safe alternatives for cloud backup
- ✓ Selective pricing

Multi-Cloud Use Case

Following which stocks Reddit's WallStreetBets crowd is talking about was a tactic that paid off well in 2021 because of GameStop and AMC Entertainment.

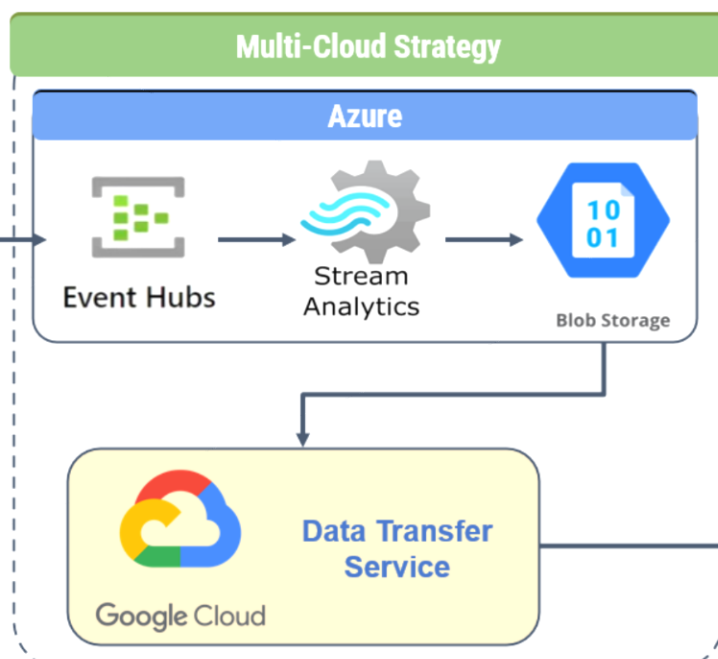
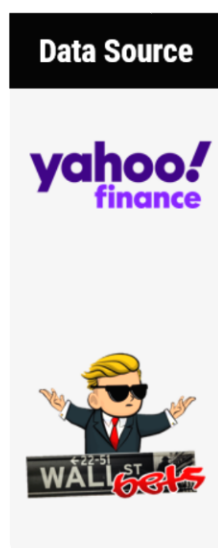
We are interested in visualizing popular stocks from WallStreetBets using the latest multi-cloud strategy.



Data Ingestion

Data Storage

Dashboarding



Process Roadmap





Data Source



Reddit API

Extract stock symbols for top trending posts.



Use python script to pull data from the API's and push it to cloud storage

Reddit API

Get: Stock Volume
50 Day Moving Avg
200 Day Moving Avg
Previous day closing price
Daily price for last 30 days
Live Price



Data Storage



Dashboard Overview



Azure Events Hub

Events hub instance capture data from python



Azure Stream Analytics Job

Scheduled SQL query to convert event data to a table



Blob Storage/ADLS Gen2

Stores the table from Stream Analytics job in a csv file



Google Data Transfer Service

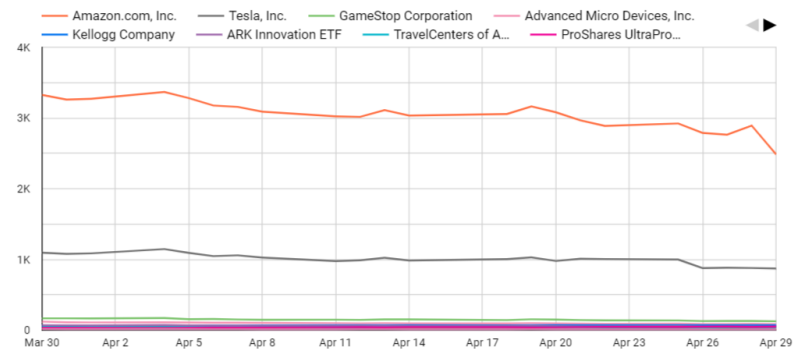
Scheduled data transfer job connects to Azure container



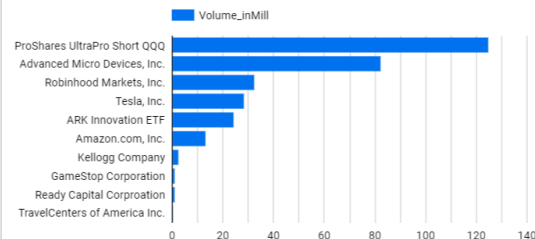
Google Cloud Storage

Files from Azure are stored in Google Cloud storage bucket

Historic Daily Closing Prices of Trending Stocks in Last Month



The Volume of Shares (in Million) of Trending Stocks



%change in liveprice of selected company vs it's previous day closing price

-10%

(livePrice is 10% below target)



Ankur Aggarwal

aggar158@umn.edu



Garrett Carlblom

carlb050@umn.edu



Yan Dong

dong0248@umn.edu

Group 13

Check Our Video on YouTube!

https://youtu.be/iZopl_Rtcw0



Pranitha Peddi

peddi024@umn.edu



Gayathri Ramanathan

raman141@umn.edu