Term Project Update #1

Wenliang Zhang   9/19/2021

Proposal 1

1. What is the advanced database area you are focusing on related to this course?

- noSQL database using MongoDB/Cassandra
- distributed big data analytic using pySpark

2. What is the proof of concept component in your term project?

- Being able to save multiple complicated genomic datasets in the json format into MongoDB/Cassandra
- Being able to query data from MongoDB/Cassandra
- Working with pySpark and MongoDB together to do data analysis

3. What are some of the goals that you plan to learn in this project?

- Learn how to design an efficient noSQL database for storing and querying complicated genomic data
- Learn how to working with pySpark and MongoDB/Cassandra together to do data analysis

4. What skills are you bringing from other courses, and what is the new element that you are learning in this class that's related to advanced database management?

- I will bring skills like Python programming, pySpark, and maybe some web application development
- noSQL database like MongoDB/Cassandra and Spark that I am learning from this class will be useful

5. What data are you looking to use specifically?

- Multiple Open Target genetics datasets in json format

Proposal 2

1. What is the advanced database area you are focusing on related to this course?

- Database normalization
- time and temporal data
- Live online database transaction system

2. What is the proof of concept component in your term project?

- Being able to save live data from websocket and save it into database
- Being able to query historical stock price of different aggregation period, such as 5min, 30min, 1 hour and 1day
- Using stored procedure for tracking WSB and ARK ETF stocks (Possibly)
- Being able to track user orders

3. What are some of the goals that you plan to learn in this project?

- Learn how to efficiently design a relational transaction database for stocks (security, company, sector, prices)
- Learn how to handle time and temporal data in a database
- Learn how to efficiently handle live streaming data

4. What skills are you bringing from other courses, and what is the new element that you are learning in this class that's related to advanced database management?

- Skills like Python programming, database API and web application development might be needed for this project
- Database normalization, performance tuning and stored procedure etc. that I am learning from this class will be useful

5. What data are you looking to use specifically?

- Stock data live streaming from the websocket every minute.