Managing stock data and related companies' news using MongoDB Atlas

Group 15:
Anshul Sharma, 110016388
Asmita Bharti, 110006023
Daniel Carneiro Lacher, 110024246
Xinxin Zhao, 105215004

GitHub: https://github.com/MaxiAgrippa/ADTFinalProject

Introduction

- This project focuses on managing massive amount of stock data in a relevant and highly accessible way by using the concepts of cloud database and mongodb.
- This not only store historical and current stock data, but having access to most recent news about company will help to understand market behaviour and use such information to predict behaviour of the stocks.
- Time- series analysis has been used for forecasting the trends.

Related work

Many softwares that related to stock market use relational database to store data for future analyze. The traditional relational database are relational, vertically, table based with predefined schema, and better at multi-row transactions.

However, with the grow of stock market data and the potential relationship founded between stock market and other data, a relational database is more and more difficult to maintain and store various types of data.

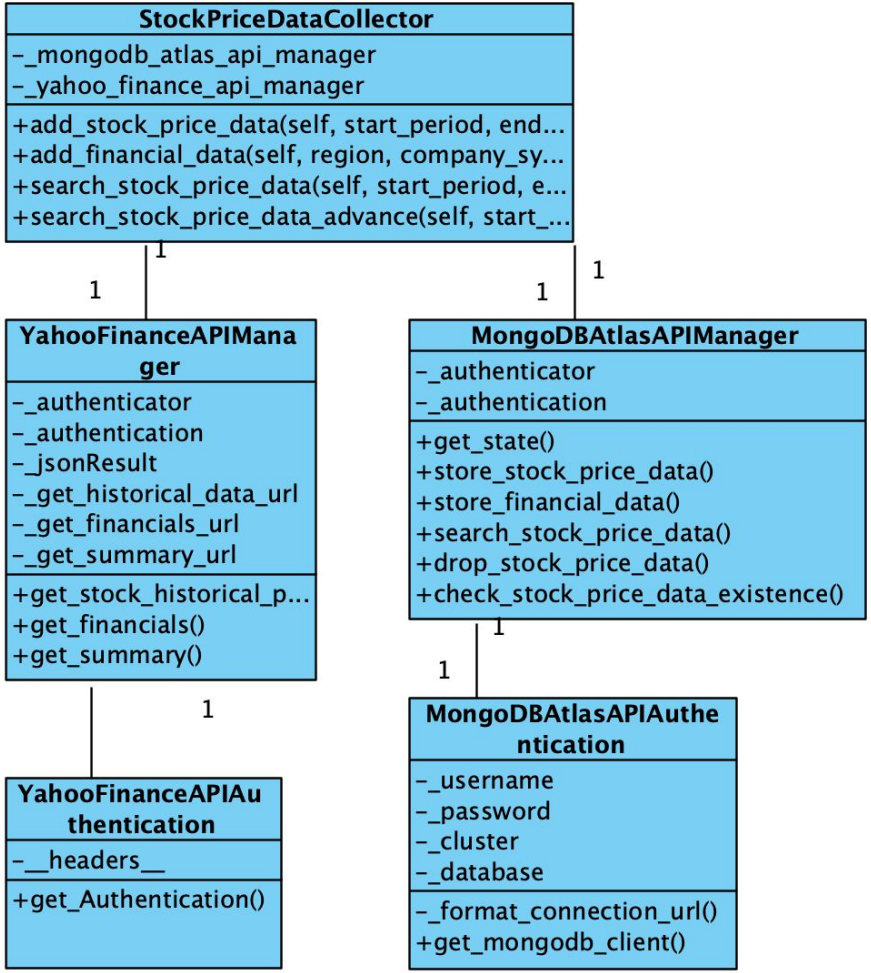
MongoDB, Store to Developers are actively working on analyzing the future market. This introduces a new requirement of storing this massive amount of data in a relevant and highly accessible way. By using the concepts of cloud database and mongodb ensures future scalability.

Techniques

- Collecting a huge amount of diverse data:
  - Big Data
  - The need for a Document database like MongoDB ➔ scalability
- Having data structures (schemas) that can evolve over time:
  - Flexibility of the schema/data model
  - Changes are quickly absorbed ➔ resiliency
- Accessing data with high performance and high availability:
  - Data replication/data fragmentation (transparent in MongoDB)
  - Cloud database
  - Scalability

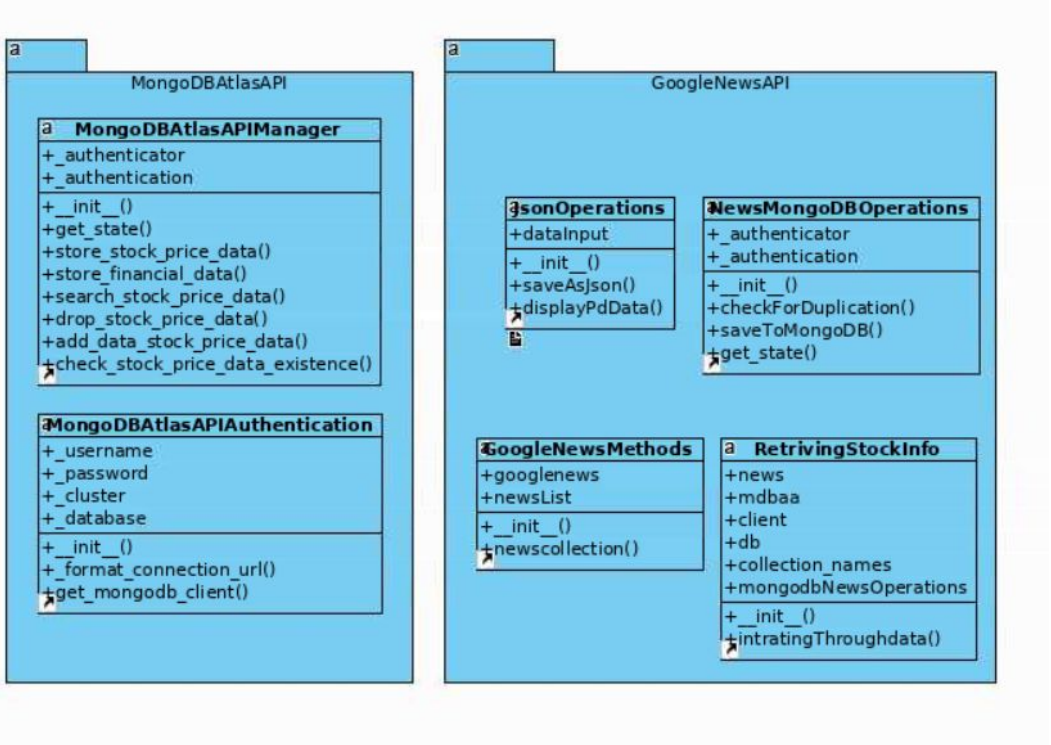
Demo

```
{
    "type": "object",
    "properties": {
        "date": {
            "type": "integer"
        },
        "open": {
            "type": "number"
        },
        "close": {
            "type": "number"
        },
        "volume": {
            "type": "integer"
        },
        "companySymbol": {
            "type": "string"
        },
        "timeZone": {
            "type": "object",
            "properties": {
                "gmtOffset": {
                    "type": "integer"
                }
            }
        }
    }
}
```



Demo


```
{
    "type": "object",
    "properties": {
      "_id": {
        "type": "object",
        "properties": {
          "$oid": {
            "type": "string"
          }
        },
        "required": [
          "$oid"
        ]
      },
      "title": {
        "type": "string"
      },
      "media": {
        "type": "string"
      },
      "date": {
        "type": "string"
      },
      "desc": {
        "type": "string"
      },
      "link": {
        "type": "string"
      },
      "img": {
        "type": "string"
      },
      "companyName": {
        "type": "string"
      }
    },
    "required": [
      "_id",
      "title",
      "media",
      "date",
      "desc",
      "link",
      "img",
      "companyName"
    ]
}
```

Reference

- https://university.mongodb.com/mercury/M220P/2020_July_28/overview
- https://pypi.org/project/GoogleNews/
- https://rapidapi.com/apidojo/api/yahoo-finance1
- https://www.mongodb.com/blog/post/new-to-mongodb-atlas-cross-region-replication-new-instance-sizes
- http://eprints.covenantuniversity.edu.ng/4112/1/Emerging_Trend.pdf
- https://ieeexplore.ieee.org/abstract/document/8489208


THANK
YOU