DATA COLLECTION MODEL

report 1

16:332:568 Software Engineering of Web Applications
Department of Electrical and Computer Engineering

SPRING, 2015
GROUP 13

Edited by

Arunachalam Thirunavukkarasu
Jaiswal Milin
Nayyar Vidur
Pappireddy Bhavya Reddy
Ravuru Rakesh
The State University of New Jersey
Rutgers

March 07, 2015

Contents

1 CONTRIBUTION BREAKDOWN **3**

2 Customer Statement of Requirements(CSR) **4**
2.1 Glossary of Terms 4

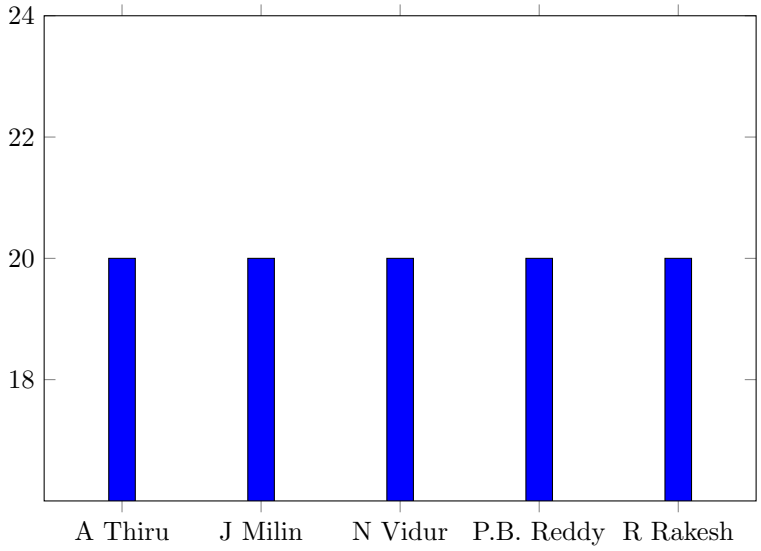
3 Software Components Used **6**
3.1 PostgreSQL database for data storage 6
3.1.1 Disadvantages of MySQL 6
3.1.2 Advantages of MySQL 6
3.2 Java module for Data Collection 7
3.3 Source Control 7

4 System Design **8**
4.1 Database Design 8
4.2 Data Extraction 8
4.2.1 Attributes of the API used 8
4.3 UML Diagram 10
4.4 Note 11

5 Reference **12**

1 CONTRIBUTION BREAKDOWN

Here is the contribution breakdown of the project



Responsibility	Arunachalam Thirunavukkarasu	Jaiswal Milin	Nayyar Vidur	Pappireddy Bhavya Reddy	Ravuru Rakesh
Project Management	20%	20%	20%	20%	20%
Customer Statement	20%	20%	20%	20%	20%
System Design	20%	20%	20%	20%	20%
Database Management	20%	20%	20%	20%	20%
Java programming	20%	20%	20%	20%	20%
Total (100')	20	20	20	20	20

All the members have equally contributed for this project.

Table 1: CONTRIBUTION BREAKDOWN FOR REPORT 1.

2 Customer Statement of Requirements(CSR)

The market in which shares of publicly held companies are issued and traded either through exchanges or over-the-counter markets. Also known as the equity market, the stock market is one of the most vital components of a free-market economy, as it provides companies with access to capital in exchange for giving investors a slice of ownership in the company. The stock market makes it possible to grow small initial sums of money into large ones, and to become wealthy without taking the risk of starting a business or making the sacrifices that often accompany a high-paying career.[1]
The idea of earning a huge sum of money by investing in the stock market attracts a lot of people to invest in the stock market. Most of the people invest in the stock market with little knowledge about the trends in the stock market and often misjudge the trends and end up loosing huge sums of money. Even the stock moguls often make mistake in predicting the market trend due to human error.
It is required to develop such a model that predicts the trend of the stock market with high precision which can aid the investors to bag maximum profits from their investments. The first step towards achieving the goal of developing such a system, it is essential to collect the data of the present and past values of stocks under consideration. We have designed a system that would collect the stock market data from www.yahoo.com and collect it to a database for further manipulations and calculations.

2.1 Glossary of Terms

1. Stock Market: A particular market where stocks and bonds are traded.

2. Data Base: A comprehensive collection of related data organized for convenient access,.

3. Singleton function: A singleton design pattern ensures that only one instance of a class is created

4. JDBC: Part of the Java Development Kit which defines an application programming interface for Java for standard SQL access to databases from Java programs. .

5. PostgreSQL: Is an advanced relational database management system with some object oriented approaches. PostgreSQL is developed and distributed as free software, and while retaining its freedom it remains technically and feature-wise a worthy competitor to even the most advanced commercial alternatives. .

6. TimerTask: A task that can be scheduled for one-time or repeated execution by a Timer..

7. API: An abbreviation of application program interface, is a set of routines, protocols, and tools for building software applications. The API specifies how software components should interact and are used when programming graphical user interface (GUI) components..

8. JAR: In software, JAR (Java Archive) is a package file format typically used to aggregate many Java class files and associated metadata and resources (text, images, etc.) into one file to distribute application software or libraries on the Java platform..

9. Equity: On a company's balance sheet, the amount of the funds contributed by the owners (the stockholders) plus the retained earnings (or losses). Also referred to as "shareholders' equity".

10. URL: Is one type of Uniform Resource Identifier (URI); the generic term for all types of names and addresses that refer to objects on the World Wide Web.

3 Software Components Used

To achieve targets of the data collection phase, our project has the following components:

3.1 PostgreSQL database for data storage

We decided to use PostGre over MySQL because of the points given below:

3.1.1 Disadvantages of MySQL

Known limitations: By design, MySQL does not intend to do everything and it comes with functional limitations that some state-of-the-art applications might require.

Reliability issues: The way certain functionality gets handled with MySQL (e.g. references, transactions, auditing etc.) renders it a little-less reliable compared to some other RDBMSs.

Stagnated development: Although MySQL is still technical an open-source product, there are complaints regarding the development process since its acquisition. However, it should be noted that there are some MySQL-based, fully-integrated databases that add value on top of the standard MySQL installations (e.g. MariaDB).

3.1.2 Advantages of MySQL

An open-source SQL standard compliant RDBMS: PostgreSQL is open-source and free, yet a very powerful relational database management system.

Strong community: PostgreSQL is supported by a devoted and experienced community which can be accessed through knowledge-bases and Q&A sites 24/7 for free.

Strong third-party support: Regardless of the extremely advanced features, PostgreSQL is adorned with many great and open-source third-party tools for designing, managing and using the management system.

Extensible: It is possible to extend PostgreSQL programmatically with stored procedures, like an advanced RDBMS should be.

Objective: PostgreSQL is not just a relational database management system but an objective one - with support for nesting, and more.

3.2 Java module for Data Collection

For the data extraction phase we chose java because of the the easily available Yahoo finance API available in JAR format. Java is also widely used with great documentation which has been used for many years now.

3.3 Source Control

We decided to use git with bitbucket for source control because it gives us distributed revision control with access to all the source files from the cloud.
Distributed revision control takes a peer-to-peer approach to version control, as opposed to the client-server approach of centralized systems. Rather than a single, central repository on which clients synchronize, each peer's working copy of the codebase is a complete repository. Distributed revision control synchronizes repositories by exchanging patches (sets of changes) from peer to peer.
In a DVCS (such as Git, Mercurial, Bazaar or Darcs), clients dont just check out the latest snapshot of the files: they fully mirror the repository. Thus if any server dies, and these systems were collaborating via it, any of the client repositories can be copied back up to the server to restore it. Every clone is really a full backup of all the data.

4 System Design

4.1 Database Design

- As mentioned in the earlier parts of the document PostgreSQL is used as the database for our project.
- We have 3 relations existing as of now ie., stocks, historical_data and real_time_data
- The description, schemas and details of the database design are included in the document named Database_Schema.pdf which is part of the deliverables.

4.2 Data Extraction

4.2.1 Attributes of the API used

SYMBOL A unique series of letters assigned to a security for trading purposes. They are also known as "ticker symbols."

OPEN PRICE The price at which a security first trades upon the opening of an exchange on a given trading day. A security's opening price is an important marker for that day's trading activity, especially for those interested in measuring short-term results, such as day traders.

CLOSE PRICE The final price at which a security is traded on a given trading day. The closing price represents the most up-to-date valuation of a security until trading commences again on the next trading day.

HIGH PRICE Day's high is the highest price at which a stock traded during the course of the day. Day's high is typically higher than the closing or opening price. More often than not this is higher than the closing price.

LOW PRICE Day's low is the lowest price at which a stock trades over the course of a trading day. Day's low is typically lower than the opening or closing price.

ADJACENT CLOSE A stock's closing price on any given day of trading that has been amended to include any distributions and corporate actions that occurred at any time prior to the next day's open. The adjusted closing price is often used when examining historical returns or performing a detailed analysis on historical returns.

- We used Java as the language for data processing. The yahoo finance jar is included as part of the java class.

Example of Data Extraction

Any available stock can be searched by using the get method of the Yahoo Finance API by passing the stock symbol as an input argument.
Stock stock = YahooFinance.get(YHOO);

- We are extracting the data from 5 stocks in this project. They are

Stock Name - Symbol Google GOOG
Yahoo YHOO
Intel INTC
Apple AAPL
Tesla Tsla
Precisely, these are the records of the stocks relation.

- After fetching the total stock information, the individual elements can be extracted by using specific method for each attribute.

- Like, to find the stocks price, we can can extract it in the following way

BigDecimal price = stock.getQuote().getPrice();

- The java processing layer is integrated with PostgreSQL database by the JDBC component.

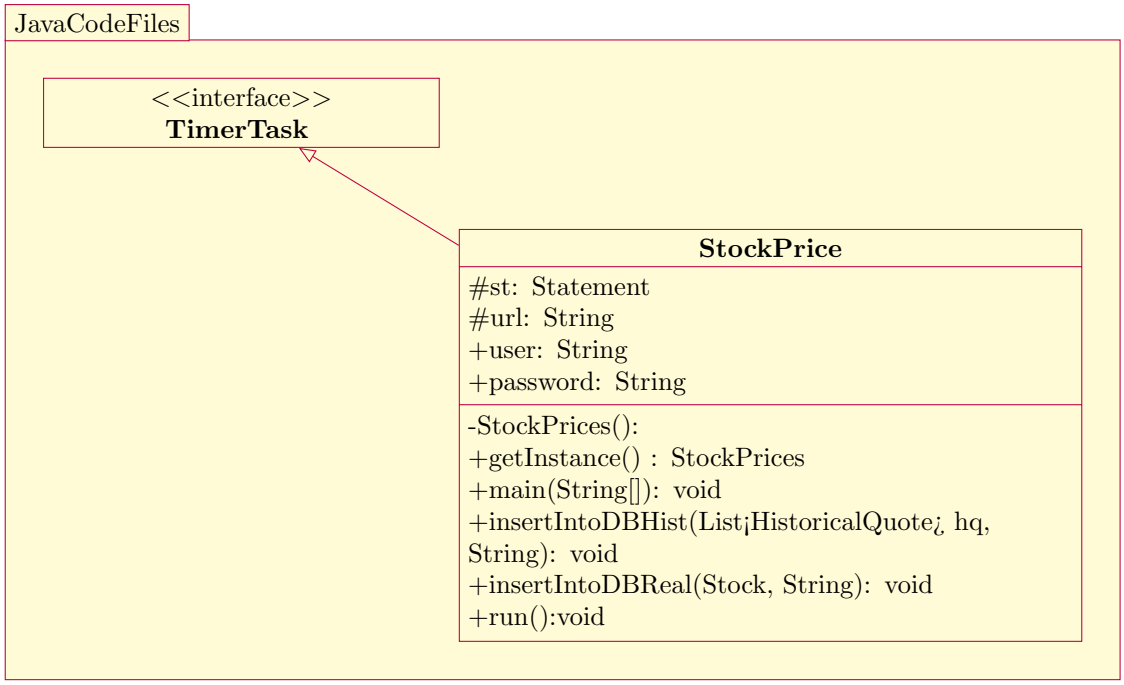
- The extracted data is inserted into the respective historic or real time data tables.

- The program is made into a background application by converting the java file into an executable jar.

- While running the jar an argument should be passed to specify which information are we storing.

Example: java -jar stockprices.jar hist extracts and stores the historic data.

4.3 UML Diagram



The system has been designed to achieve the objective of acquiring the stock data from an online website and save it into our own database. The system has been implemented on Java SE Environment 8 and the data has been stored on the **PostGreSQL** database. Database and Java components are integrated by Java Data Base Connectivity(JDBC).

StockPrices class provides most of the functionality and is a singleton class which means that only a single instance of this class can be made. SINGLETON CLASS Provide a global point of access to the object . To get real time-data, the StockPrices class extends to the **TimerTask** class. This timer class is a timer for processing the real time data after a fixed interval of time.
The **StockPrice** class has important member functions like *Statement st*, *String url*, *String user* and *String password* . All these member functions are used for manipulating the data in the database.

st member The function is used for running the SQL statements in the database.

url member The is used to point the url of the database being used.

User is the name of the user name who can access the database and

password is linked to a particular user to add security to the system.

The *main()* function lies within the **StockPrices class**, this is the function at which execution begins in Java. The *main()* function takes in an argument in the form of a string. If the argument entered is hist, it represents the historical data and the function to process historic data is executed. Whereas if the entered argument is real, it represents the real time data and the execution is diverted to execute the function that processes real-time data. The *main()* function also checks for erroneous entries and stops executing if the number of arguments provides are less than one. When the correct number of arguments are passed, the *main()* function executes the appropriate function to process either the historical data or real time data.
When the function to process the historical data is called by the *main()* function. A connection with the database to fetch all the historical data for the stocks mentioned is established. The stock data being fetched is for yahoo, google, tesla, apple and intel. The data is fetched using the *get()* function of the yahoo api. Once the data has been collected, the main function calls the *insertIntoDBHist()* function which puts the historical data of each of these stocks into the database. After this the main function will exit.
If the argument passed to the *main()* function requires real data to be processed. Under such circumstances, the *main()* function creates a timer by using the **TimerTask** class. The main fuction will schedule a task to be called every 30 seconds to fetch the real time data using the *get()* function from the yahoo api. Once the data has been collected, the timerTick classs function calls the *insertIntoDBReal()* function to puts the real data of each of the stocks into the database and then returns the control to the man function to exit the *main()* function.

4.4 Note

Please go through the following documents apart from this report:

1. readme.txt for further information about executing the project.

2. Database_Schema.pdf for the database design.

3. csv files for the historic and real time data records collected.

5 Reference

[**1**]. HTTP://WWW.INVESTOPEDIA.COM/TERMS/S/STOCKMARKET.ASP
[**2**]. HTTP://FINANCE.YAHOO.COM
[**3**]**USER STORIES**. HTTP://WWW.ECE.RUTGERS.EDU/~MARSIC/BOOKS/SE/
[**4**]**SOFTWARE ENGINEERING**. HTTP://WWW.ECE.RUTGERS.EDU/~MARSIC/
BOOKS/SE/
[**5**].HTTP://WWW.ENTERPRISEDB.COM/SOLUTIONS/MYSQL-VS-POSTGRESQL
[**6**]. HTTP://YAHOOFINANCE-API.COM/#SINGLESTOCK-EASY
[**7**].HTTP://STACKOVERFLOW.COM/QUESTIONS/5774970/RUN-JAR-FILE-IN-COMMAND-PROMPT
[**8**]. HTTP://EN.WIKIPEDIA.ORG/WIKI/ENTITY\%E2\%80\%93RELATIONSHIP_MODEL
[**9**]. HTTP://WWW.POSTGRESQL.ORG/DOWNLOAD/WINDOWS/
[**10**]. HTTP://DOCS.ORACLE.COM/JAVASE/7/DOCS/API/JAVA/UTIL/TIMER.HTML