**Project Phase 1 - Data Collection Module**

**Design Report**

**Group 5**

Song Yang

Xin Yang

Ke Xia

Yiyuan Fu

Zhuohang Li

ECE 16:332:568 -

Software Engineering Web Application

Mar. 1st, 2018

**Individual Contributions**

* Song Yang:
* Xin Yang:
* Ke Xia:
* Yiyuan Fu:
* Zhuohang Li:

1. Introduction

This phase is part of the ECE 568 project to develop a stock-forecasting website. For this module, we are aiming to develop a program that can run continuously to retrieve stock information from financial website and stores the extracted data into a local database. We’ll be using python to develop the crawler code and MySQL for the relational database. The information of following 10 stocks will be used for demonstration: GOOG, AABA, CSCO, T, WMT, NOK, NFLX, APA, NKE, GE.

1. Implementation

The module can be divided into two main stages: data fetching and data storing.

Data Fetching

We’ve developed two separate sets of python code to collect both real-time and historical stock information. One direct way of doing involves two python packages: Selenium and BeautifulSoup4. The main idea is to 1. open financial website (Yahoo-finance is chosen for this project) 2. use Selenium to control browser driver and mimic mouse operation to scroll down and show the full page. 3. use BeautifulSoup4 to resolve HTML to soup package to get data. Another way is to call functions in alpha vantage API to get real-time and historical data. The program is designed to run continuously to collect data every minute and store into local database.

Data Storing

MySQL is used to build a local relational database to store data collected in the previous stage. The entity-relationship and UML are as shown in the following diagrams:

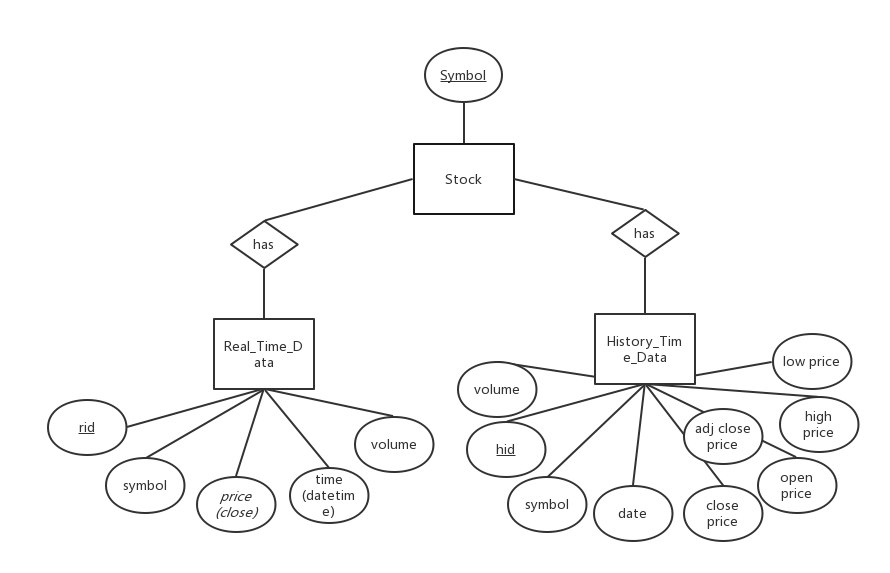


Figure 1: E – R Diagram

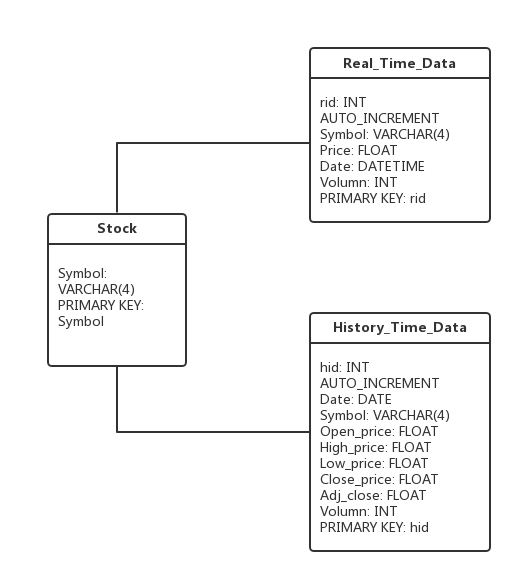


Figure 2: UML Diagram