

ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

**Nền tảng dữ liệu chứng khoán Việt Nam
Phân hệ quản lý và lập lịch trình tính toán chỉ báo**

MAI TRƯỜNG SƠN

son.mt183620@sis.hust.edu.vn

**Trường Công nghệ thông tin và truyền thông
Ngành Khoa học Máy tính**

Giảng viên hướng dẫn: TS. Trần Việt Trung

Chữ kí GVHD

Trường:

Công nghệ thông tin và Truyền thông

HÀ NỘI, 03/2023

LỜI CẢM ƠN

Đầu tiên, tôi xin gửi lời cảm ơn đến bố mẹ, em gái và những người thân trong gia đình vì đã luôn ủng hộ và động viên tôi trong suốt gần 05 năm theo học tại Đại học Bách khoa Hà Nội.

Tôi xin gửi lời cảm ơn đến TS. Trần Việt Trung vì đã tận tình hướng dẫn, chỉ bảo để tôi có thể hoàn thiện đề án này; cảm ơn Tuấn, Dương, Tú vì đã đồng hành cùng tôi với đề tài này trong nhiều tháng vừa qua. Tôi cũng xin gửi lời cảm ơn đến các thầy cô giáo ở Bách khoa Hà Nội, đến tập thể IT1-04 K63 vì đã giúp quãng đời sinh viên của tôi trở nên vô cùng đáng nhớ.

Tiếp theo, tôi xin gửi lời cảm ơn đến những đồng nghiệp thuộc phòng Công nghệ số, Ban Công nghệ Thông tin, Tập đoàn Công nghiệp - Viễn thông Quân đội (Viettel) - nơi tôi đang thực tập - vì đã tạo điều kiện công việc tối đa cho tôi trong quá trình thực hiện đề án này.

Cuối cùng, cảm ơn Trang vì đã đến và luôn là chỗ dựa tinh thần, luôn động viên tôi trong những thời điểm khó khăn nhất.

Xin trân trọng cảm ơn!

TÓM TẮT NỘI DUNG ĐỒ ÁN

Thị trường chứng khoán Việt Nam trong những năm trở lại đây chứng kiến một giai đoạn phát triển nhanh chóng với xu hướng đầu tư cá nhân tăng mạnh; từ đó đặt ra yêu cầu từ về những nền tảng ứng dụng theo dõi dữ liệu biến động của thị trường chứng khoán. Hiện nay ở Việt Nam có thể dễ dàng tiếp cận với các ứng dụng như vậy với máy tính và điện thoại thông minh. Những ứng dụng này cung cấp những chức năng cần thiết để người dùng theo dõi thị trường với giao diện dễ sử dụng, bảng biểu, sơ đồ dữ liệu được cập nhật theo thời gian thực.

Bên cạnh những ưu điểm nêu trên, các nền tảng này vẫn còn những hạn chế về dữ liệu của các chỉ báo chứng khoán. Chúng chủ yếu cung cấp những chỉ báo cơ bản, người dùng không thể thêm mới hay tùy biến những chỉ báo theo công thức của riêng mình. Trước thực tế này, nhóm chúng tôi đã quyết định thực hiện xây dựng "Nền tảng dữ liệu thị trường chứng khoán Việt Nam" với chức năng cập nhật dữ liệu giao dịch thị trường, tính toán chỉ báo chứng khoán, cung cấp giao diện website để người dùng có thể tạo mới và chỉnh sửa các chỉ báo theo nhu cầu cá nhân của người sử dụng.

Trong khuôn khổ ĐATN này, mục tiêu của tôi là xây dựng "Phân hệ quản lý và lập lịch trình tính toán chỉ báo", đóng vai trò xử lý và tính toán dữ liệu chỉ báo của thị trường chứng khoán. Sau quá trình nghiên cứu, tìm hiểu và triển khai, tôi đã xây dựng được phân hệ gồm các trình tính toán được quản lý và lập lịch tự động với ngôn ngữ lập trình Python [1] và công nghệ lập lịch Apache Airflow [2], từ đó đóng góp vào hoàn thiện sản phẩm nền tảng dữ liệu của cả nhóm.

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....	1
1.1 Đặt vấn đề.....	1
1.2 Mục tiêu và phạm vi đề tài.....	1
1.3 Định hướng giải pháp.....	2
1.4 Bố cục đồ án	2
CHƯƠNG 2. XÂY DỰNG NỀN TẢNG DỮ LIỆU THỊ TRƯỜNG CHỨNG KHOÁN	4
2.1 Thông tin chung về TTCK Việt Nam.....	4
2.1.1 Các khái niệm cơ bản	4
2.1.2 Các chỉ báo chứng khoán phổ biến.....	5
2.1.3 Khảo sát các ứng dụng nền tảng dữ liệu TTCK tại Việt Nam.....	9
2.2 Tổng quan thiết kế nền tảng dữ liệu chứng khoán.....	10
2.2.1 Tổng quan yêu cầu chức năng.....	10
2.2.2 Yêu cầu phi chức năng.....	10
2.2.3 Thiết kế hệ thống	11
CHƯƠNG 3. CÔNG NGHỆ SỬ DỤNG.....	13
3.1 Công nghệ lưu trữ dữ liệu	13
3.1.1 Apache Druid[3]	13
3.1.2 Apache Kafka[4].....	15
3.2 Công nghệ xử lý dữ liệu - Python[1] và pandas [9].....	17
3.3 Công nghệ lập lịch cho các trình tính toán - Apache Airflow [2]	18
3.4 Công nghệ tự động sinh tệp tin theo mẫu - jinja[11]	20
3.5 Công nghệ đóng gói và triển khai phần mềm - Docker[12]	21

CHƯƠNG 4. XÂY DỰNG PHÂN HỆ QUẢN LÝ VÀ LẬP LỊCH TRÌNH TÍNH TOÁN CHỈ BÁO	22
4.1 Tính toán chỉ báo theo công thức.....	22
4.1.1 Thiết kế cơ sở dữ liệu	22
4.1.2 Thiết kế luồng tính toán.....	23
4.1.3 Đặc tả lớp Miner	25
4.1.4 Sơ đồ lớp các trình tính toán.....	29
4.2 Lập lịch các trình tính toán	30
4.2.1 Nhu cầu lập lịch các trình tính toán chỉ báo	30
4.2.2 Lập lịch sử dụng Apache Airflow	31
4.3 Chức năng tự động tạo mới chỉ báo cá nhân	32
4.3.1 Biểu đồ hoạt động của chức năng tạo mới chỉ báo	33
4.3.2 Dữ liệu đầu vào của người dùng	34
4.3.3 Vai trò của phân hệ quản lý và lập lịch các trình tính toán chỉ báo...	35
4.4 Yêu cầu phi chức năng	39
CHƯƠNG 5. KẾT QUẢ THỰC NGHIỆM VÀ ĐÁNH GIÁ.....	41
5.1 Xây dựng ứng dụng.....	41
5.1.1 Thư viện và công cụ sử dụng.....	41
5.1.2 Cấu hình thiết bị	42
5.2 Triển khai	42
5.2.1 Triển khai trình tính toán chỉ báo	42
5.2.2 Triển khai lập lịch trình tính toán chỉ báo	46
5.2.3 Kết quả đạt được	47
5.2.4 Minh họa các chức năng chính	47
5.3 Kiểm thử.....	50

CHƯƠNG 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	53
6.1 Kết luận.....	53
6.2 Hướng phát triển.....	53
TÀI LIỆU THAM KHẢO.....	54

DANH MỤC HÌNH VẼ

Hình 2.1	Mô hình hệ thống cho nền tảng dữ liệu	12
Hình 3.1	Kiến trúc và các thành phần của Apache Druid [7]	13
Hình 3.2	Kiến trúc và các thành phần của Apache Kafka	16
Hình 3.3	Mối tương quan giữa <i>Series</i> và <i>DataFrame</i> trong pandas	18
Hình 3.4	Kiến trúc và các thành phần của Apache Airflow [10]	19
Hình 3.5	Các bước tạo file theo template của jinja	20
Hình 4.1	Apache Druid data model	23
Hình 4.2	Luồng thực thi của một trình tính toán chỉ báo	24
Hình 4.3	Biểu đồ các lớp Miner, DruidConnection và KafkaConnection	24
Hình 4.4	Biểu đồ lớp Miner	25
Hình 4.5	Lớp DruidConnection	26
Hình 4.6	Lớp KafkaConnection	26
Hình 4.7	Sơ đồ lớp gồm các trình tính toán riêng biệt	30
Hình 4.8	Khai báo một Airflow DAG trong file Python	31
Hình 4.9	Khai báo một Task trong file Python	31
Hình 4.10	Khai báo thứ tự các Tasks trong một DAG	32
Hình 4.11	DAG trình tự các trình tính toán chỉ báo	32
Hình 4.12	Sơ đồ hoạt động tính năng tạo mới chỉ báo	33
Hình 4.13	Template tạo Druid spec file	36
Hình 4.14	Template tạo class cho chỉ báo mới	37
Hình 4.15	Các bước cập nhật DAG trên Airflow	38
Hình 4.16	Template cho script khai báo Task cho chỉ báo mới	39
Hình 5.1	Mã nguồn lớp DruidConnection	43
Hình 5.2	Mã nguồn lớp KafkaConnection	44
Hình 5.3	Mã nguồn lớp Miner	45
Hình 5.4	Mã nguồn lớp OBVMiner	46
Hình 5.5	Các container thành phần của Airflow	47
Hình 5.6	Các datasources trên Apache Druid	48
Hình 5.7	Kết quả chỉ báo OBV mã HPG trong tháng 1/2023	48
Hình 5.8	Giao diện web của Airflow	49
Hình 5.9	Giao diện thông tin 1 DAG trên Airflow	49

DANH MỤC BẢNG BIỂU

Bảng 5.1	Danh sách thư viện và công cụ sử dụng	41
Bảng 5.2	Cấu hình máy tính cá nhân	42
Bảng 5.3	Các thống kê khác của phân hệ khi triển khai	47

DANH MỤC THUẬT NGỮ

Tên thuật ngữ	Ý nghĩa
Ticker	Mã cổ phiếu, xác định duy nhất một cổ phiếu cụ thể trên một sàn giao dịch cụ thể
Apache Druid	Hệ thống lưu trữ dữ liệu time-series
Druid datasource	Thành phần lưu trữ dữ liệu trong Apache Druid
Apache Kafka	Nền tảng lưu trữ dữ liệu streaming phân tán
Kafka topic	Một đơn vị tổ chức lưu trữ dữ liệu trong Apache Kafka
File	Tệp tin
Class	Lớp (trong Lập trình hướng đối tượng)
Miner	Trình tính toán chỉ báo
Template	Bản mẫu
Docker	Công nghệ đóng gói và triển khai ứng dụng
MySQL Server	Hệ thống quản trị cơ sở dữ liệu quan hệ mã nguồn mở phát triển bởi Oracle
PostgreSQL	Một thống quản trị cơ sở dữ liệu mã nguồn mở
MsSQL	Hệ thống quản trị cơ sở dữ liệu phát triển bởi Microsoft