ĐẠI HỌC QUỐC GIA TP HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

KHOA CÔNG NGHỆ THÔNG TIN

BỘ MÔN HỆ THỐNG THÔNG TIN

**BÁO CÁO ĐỒ ÁN**

**MÔN TRÍ TUỆ KINH DOANH NÂNG CAO**

**Đề tài :**

**XÂY DỰNG KHO DỮ LIỆU CHO NGHIỆP VỤ PHÂN TÍCH CỔ PHIẾU**

**GIẢNG VIÊN HƯỚNG DẪN :** TS. Lê Thị Nhàn

**HỌC VIÊN THỰC HIỆN :**

20C12007 - Trần Đình Lâm

20C12030 – Huỳnh Lâm Phú Sĩ

**KHÓA :** 30

TP.HCM, 07-2021

**Mục Lục**

[**Mục Lục** 1](#_Toc78686470)

[**Bảng danh sách các hình** 2](#_Toc78686471)

[**Bảng thông tin chi tiết nhóm** 3](#_Toc78686472)

[**1.** **Đặt vấn đề** 4](#_Toc78686473)

[**1.1.** **Bài toán phân tích giá cổ phiếu ở thị trường Việt Nam** 4](#_Toc78686474)

[**1.2.** **Đối tượng sử dụng hệ thống** 5](#_Toc78686475)

[**2.** **Mô tả các nghiệp vụ** 5](#_Toc78686476)

[**2.1.** **Nghiệp vụ phân tích cơ bản** 5](#_Toc78686477)

[**2.2.** **Nghiệp vụ phân tích kỹ thuật** 6](#_Toc78686478)

[**2.3.** **Các chức năng của hệ thống** 7](#_Toc78686479)

[**3.** **Dữ liệu đầu vào** 8](#_Toc78686480)

[**3.1.** **Dữ liệu thông tin doanh nghiệp** 8](#_Toc78686481)

[**3.2.** **Dữ liệu lịch sử giá theo ngày** 8](#_Toc78686482)

[**3.3.** **Lưu trữ dữ liệu tại nguồn** 9](#_Toc78686483)

[**4.** **Kiến trúc kho** 9](#_Toc78686484)

[**4.1.** **Sơ đồ tổng quát** 9](#_Toc78686485)

[**4.2.** **Mô hình NDS** 10](#_Toc78686486)

[**4.3.** **Mô hình DDS** 11](#_Toc78686487)

[**5.** **Quy trình ETL** 11](#_Toc78686488)

[**5.1.** **Quy trình ETL từ nguồn ngoài vào NDS** 11](#_Toc78686489)

[**5.2.** **Quy trình ETL từ NDS vào DDS** 12](#_Toc78686490)

[**6.** **Cài đặt** 14](#_Toc78686491)

[**6.1.** **Cài đặt NDS và DDS** 14](#_Toc78686492)

[**6.2.** **Analysis OLAP** 15](#_Toc78686493)

[**7.** **Báo cáo report** 16](#_Toc78686494)

[**8.** **Nhận xét và kết luận** 18](#_Toc78686495)

[References 19](#_Toc78686496)

# **Bảng danh sách các hình**

[Hình A Thị trường chứng khoán Việt Nam ngày 28/05/2021 [1] 4](file:///C:\Users\LAP11777-local\Documents\master_2nd_semester\BusinessIntelligence\ReportBI.docx#_Toc78683094)

[Hình B: Biểu đồ hỗ trợ phân tích kỹ thuật 7](file:///C:\Users\LAP11777-local\Documents\master_2nd_semester\BusinessIntelligence\ReportBI.docx#_Toc78683095)

[Hình C: Sơ đồ chức năng tổng quát của hệ thống 8](file:///C:\Users\LAP11777-local\Documents\master_2nd_semester\BusinessIntelligence\ReportBI.docx#_Toc78683096)

[Hình D: Minh họa dữ liệu thông tin cơ bản của cổ phiếu. Nguồn: cafef [1] 8](file:///C:\Users\LAP11777-local\Documents\master_2nd_semester\BusinessIntelligence\ReportBI.docx#_Toc78683097)

[Hình E: Minh họa lịch sử giá của HPG. Nguồn cophieu68.vn [2] 9](file:///C:\Users\LAP11777-local\Documents\master_2nd_semester\BusinessIntelligence\ReportBI.docx#_Toc78683098)

[Hình F: Kiến trúc kho dữ liệu 10](#_Toc78683099)

[Hình G: Mô hình NDS hoàn chỉnh 10](file:///C:\Users\LAP11777-local\Documents\master_2nd_semester\BusinessIntelligence\ReportBI.docx#_Toc78683100)

[Hình H: Mô hình DDS hoàn chỉnh 11](#_Toc78683101)

# **Bảng thông tin chi tiết nhóm**

|  |  |  |  |
| --- | --- | --- | --- |
| **BẢNG PHÂN CÔNG & ĐÁNH GIÁ HOÀN THÀNH CÔNG VIỆC** | | | |
| **Người thực hiện** | **Công việc thực hiện** | **Mức độ hoàn thành** | **Đánh giá của nhóm** |
| 20C12007  Trần Đình Lâm | Tìm hiểu quy trình nghiệp vụ phân tích cơ bản của cổ phiếu | 80% | 80% |
| Tìm các nguồn dữ liệu cổ phiếu Việt Nam |
| Phân tích các chức năng cần có của hệ thống |
| Xây dựng mô hình NDS & ETL |
| Viết báo cáo và làm slide |
| 20C12030  Huỳnh Lâm Phú Sĩ | Tìm hiểu quy trình nghiệp vụ phân tích kỹ thuật của cổ phiếu | 80% | 80% |
| Xây dựng kiến trúc kho dữ liệu |
| Xây dựng mô hình DDS & ETL |
| Viết báo cáo và làm slide |

1. **Đặt vấn đề**
2. **Bài toán phân tích giá cổ phiếu ở thị trường Việt Nam**

Thị trường chứng khoán Việt Nam là một phần rất quan trọng trong thị trường tài chính, và là nơi huy động vốn hiệu quả từ xã hội. Kể từ nửa đầu năm 2020 đến nay thị trường chứng khoán Việt Nam trở nên sôi động và tăng trưởng cực kỳ nóng, trở thành kênh thu hút đầu tư hấp dẫn.

Tuy nhiên giá cổ phiếu thường biến động mạnh và phụ thuộc vào rất nhiều yếu tố cả bên trong lẫn bên ngoài. Việc đầu tư cổ phiếu lại thường bị phụ thuộc vào cảm xúc, cảm tính hay tâm lý của nhiều nhà đầu tư chưa có kinh nghiệm. Từ đó rất dễ nhận thấy đây là kênh đầu tư không ít rủi ro nếu người đầu tư thiếu kiến thức và thông tin cần thiết.

Hình A Thị trường chứng khoán Việt Nam ngày 28/05/2021 [1]

Vì vậy, nhóm đề xuất xây dựng một hệ thống kho dữ liệu để tổng hợp thông tin giao dịch của các cổ phiếu Việt Nam, tính toán và đưa ra các chỉ số định lượng cơ bản, các báo cáo cần thiết để hỗ trợ người tham gia đầu tư cổ phiếu nắm được các chỉ số cơ bản, để lựa chọn cổ phiếu phù hợp và đưa ra chiến lược đầu tư trung và dài hạn.

1. **Đối tượng sử dụng hệ thống**

Các đối tượng có thể sử dụng đầu ra của hệ thống bao gồm:

* Quỹ đầu tư chứng khoán
* Bộ phận phân tích giá cổ phiếu
* Nhân viên tư vấn đầu tư
* Nhà đầu tư cá nhân

1. **Mô tả các nghiệp vụ**
2. **Nghiệp vụ phân tích cơ bản**

Nghiệp vụ Tính toán, Phân tích các chỉ số cơ bản của một cổ phiếu/ nhóm cổ phiếu:

Trong nghiệp vụ phân tích chứng khoán, các chỉ số cơ bản của một cổ phiếu hay một loại cổ phiếu là các thông tin đầu tiên cần phải nắm bắt. Trong phương pháp phân tích cơ bản, các nhóm chỉ số sau đây thường được các phân tích viên sử dụng nhiều nhất:

Nhóm hệ số giá trị: Chỉ số EPS (Earning per Share)

* + EPS – Thu nhập trên một Cổ phiếu: là chỉ số cho biết khả năng sinh lợi của Công ty trên một đơn vị cổ phần
    - EPS = (∑ thu nhập sau thuế - ∑ số cổ tức của Cổ phiếu ưu đãi)/ ∑ số Cổ phiếu đang lưu hành
    - Nếu chỉ số EPS càng cao thì nó càng cho thấy khả năng sinh lời của Công ty càng lớn và ngược lại.
    - So sánh chỉ số EPS qua các thời kỳ sẽ giúp chúng ta biết được tốc độ tăng trưởng của doanh nghiệp đang phân tích.
  + Chỉ số P/E (Price per earning) = Giá hiện tại của Cổ phiếu/ EPS
    - Là tỉ lệ giá trên thu nhập
    - Cho biết mức độ đánh giá, sẵn sàng trả giá cho cổ phiếu đó trên thị trường.
  + Chỉ số P/B Chỉ số Giá/Giá trị sổ sách (Price-to-Book ratio, P/B Ratio). Chỉ số này giúp xác định mức độ đánh giá của thị trường đối với khả năng sinh lời của công ty đó.

Ngoài ra còn một số chỉ số cơ bản khác phục vụ cho quá trình phân tích nêu trên

#### **Mục đích:**

* + Thông qua các chỉ số cơ bản, xác định được giá cả của một cổ phiếu là đắt hay rẻ so với thị trường

#### **Thông tin đầu vào:**

* + Thông tin tổng số cổ phiếu lưu hành của từng công ty
  + Thông tin thị giá hiện tại (và có thể là quá khứ) của các cổ phiếu
  + Thông tin chỉ số EPS hiện tại của các cổ phiếu

#### **Kết quả đầu ra:**

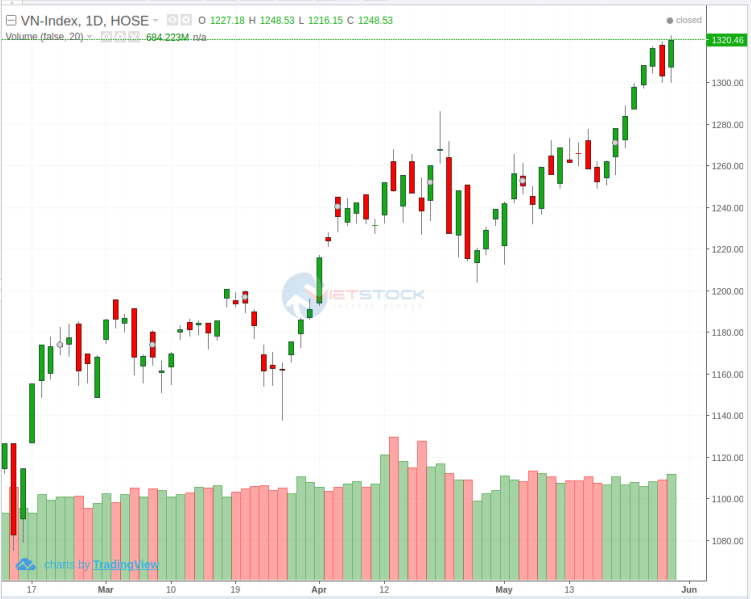
* + Chỉ số P/E của từng cổ phiếu
  + Phân bổ chỉ số P/E của toàn thị trường
  + Xác định một cổ phiếu bất kỳ thuộc mức độ nào trên thị trường
  + Xác định một cổ phiếu bất kỳ thuộc mức độ nào trong nhóm ngành tương ứng

#### **Ý nghĩa rút trích được từ đầu ra:**

* + Nếu chỉ số P/E cao hơn trung bình chung toàn thị trường hoặc trung bình chung toàn ngành, tức cổ phiếu đang được kỳ vọng, nên đang được định giá cao, giá đang đắt
  + Nếu chỉ số P/E tương đối thấp hoặc ngang mức trung bình, nghĩa là cổ phiếu đang ít được kỳ vọng, nên giá đang rẻ
  + Thông qua phân tích này, ta có thể tư vấn việc lựa chọn đầu tư vào cổ phiếu giá rẻ hoặc giá đắt, phụ thuộc vào chiến lược đầu tư của khách hàng/Quỹ đầu tư đang chọn.

1. **Nghiệp vụ phân tích kỹ thuật**

Phân tích kỹ thuật là phương pháp phân tích dựa hoàn toàn vào diễn biến giao dịch của thị trường và lịch sử tăng giảm của giá cổ phiếu, khối lượng giao dịch hay mối tương quan giữa cung và cầu trên thị trường đối với một hoặc một nhóm cổ phiếu nhất định, nhằm tìm ra quy luật chung hoặc chu kỳ của giá cổ phiếu. Nghiệp vụ phân tích này cần dựa trên biểu đồ tăng giảm giá thị trường của các mã cổ phiếu.

Thông tin đầu vào của nghiệp vụ là lịch sử giá của một số hoặc tất cả các cổ phiếu trên thị trường theo từng ngày giao dịch. Đầu ra của nghiệp vụ là biểu đồ đặc thù biến động giá, khối lượng giao dịch của một cổ phiếu hoặc toàn thị trường.

Hình B: Biểu đồ hỗ trợ phân tích kỹ thuật

1. **Các chức năng của hệ thống**

Diagram

Description automatically generatedSơ đồ Hình C mô tả tổng quan các chức năng hệ thống cần có để phục vụ nhu cầu phân tích cổ phiếu

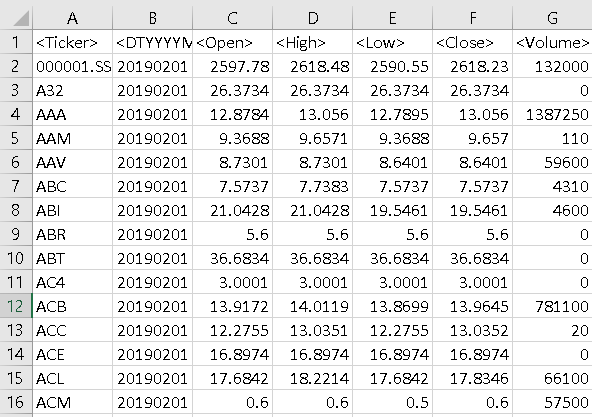
Hình C: Sơ đồ chức năng tổng quát của hệ thống

1. **Dữ liệu đầu vào**
2. **Dữ liệu thông tin doanh nghiệp**

Hình D: Minh họa dữ liệu thông tin cơ bản của cổ phiếu. Nguồn: cafef [1]

1. **Dữ liệu lịch sử giá theo ngày**

Minh họa dữ liệu danh sách lịch sử giá vào cuối ngày của từng sàn, mỗi dòng mô tả giá một cổ phiếu như sau:



Hình 1: Minh họa dữ liệu lịch sử giá

1. **Lưu trữ dữ liệu tại nguồn**

Dữ liệu tại nguồn bao gồm dữ liệu đến từ 3 nguồn như sau:

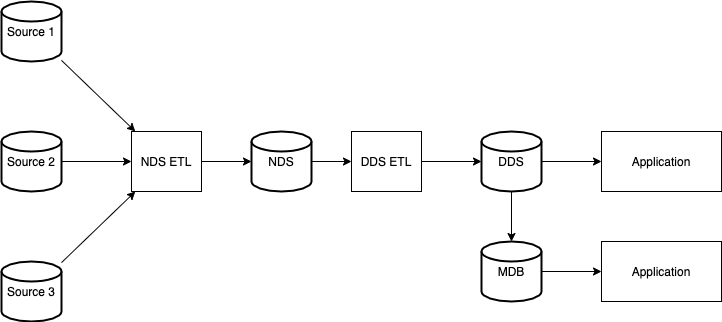
* Nguồn dữ liệu từ các tập tin csv: nguồn dữ liệu gồm thông tin về các chỉ số cổ phiếu, thông tin giá theo ngày của các cổ phiếu
* ****Nguồn dữ liệu từ CSDL: nguồn dữ liệu từ CSDL bao gồm thông tin cổ phiếu từ các sàn HOSE (1 CSDL) và các sàn còn lại (1 CSDL)

Hình E: Minh họa lịch sử giá của HPG. Nguồn cophieu68.vn [2]

1. **Kiến trúc kho**
2. **Sơ đồ tổng quát**

Kho dữ liệu có các thành phần sau:

* NDS và DDS là hai thành phần cơ bản của kho
  + - MDB dùng để chứa dữ liệu theo khối và thực hiện các tác vụ OLAP

Toàn bộ các bước tải dữ liệu giữ các CSDL đều thực hiện qua ETL

Hình F: Kiến trúc kho dữ liệu

1. Diagram

   Description automatically generated**Mô hình NDS**

Hình G: Mô hình NDS hoàn chỉnh

1. **Mô hình DDS**

Diagram

Description automatically generated

Hình H: Mô hình DDS hoàn chỉnh

1. **Quy trình ETL**

Để thực hiện đổ dữ liệu vào NDS và DDS, chúng ta thực hiện quy trình ETL qua 2 bước gồm đổ dữ liệu từ nguồn vào NDS và DDS.

Timeline

Description automatically generated

Hình 2: Quy trình ETL tổng quát

* 1. **Quy trình ETL từ nguồn ngoài vào NDS**

**Quá trình ETL từ nguồn ngoài vào NDS bao gồm các bước:**

* **Đọc dữ liệu từ nguồn dữ liệu cổ phiếu trên sàn HOSE (sàn lớn nhất, nhiều cổ phiếu phổ biến, lớn để đầu tư an toàn)**
* **Đọc dữ liệu từ nguồn dữ liệu cổ phiếu trên các sàn khác (các sàn nhỏ hơn, có các cổ phiếu nhỏ để đầu tư mạo hiểm)**

Chart

Description automatically generated with medium confidence

Hình 3: Quy trình Transform lúc đọc dữ liệu từ các sàn nhỏ

* **Đọc dữ liệu giá trị cổ phiểu theo ngày theo từng file tập tin**

Chart

Description automatically generated

Hình 4: Transform lúc đọc dữ liệu từ file để có lịch sử giá cổ phiếu

* **Đọc dữ liệu chỉ số cổ phiếu theo ngày theo các tập tin**

A picture containing diagram

Description automatically generated

Hình 5: Transfrom lúc lấy chỉ số cổ phiếu

* 1. **Quy trình ETL từ NDS vào DDS**

**Quá trình ETL từ NDS vào DDS bao gồm các bước:**

* **Đọc dữ liệu về ngày của các giao dịch để tạo ra 2 dimesion table lag ngày theo giá trị cổ phiếu và ngày theo báo cáo chỉ số cổ phiếu**
* **Đọc dữ liệu từ các bảng nds\_cophieu, nds\_nganh, nds\_phannganh, nds\_loaichiso để tạo ra các bảng dimension cophieu, phanganh, nganhkinhdoanh, chiso**
* **Đọc dữ liệu từ bảng nds\_chisotheongay để tạo ra bảng fact chisotheongay, ngoài ra để phục vụ quá trình tạo report vào OLAP trong MDB, ta thực hiện tạo ra bảng chisoquantrongtheongay**
* **Đọc dữ liệu từ bảng nds\_lichsugia để tạo ra bảng fact giatricophieu**

A picture containing sky, text, map, smoke

Description automatically generated

Hình 6: Quá trình ETL từ NDS vào DDS

* **Để giúp thực hiện cập nhật các dữ liệu trong DDS trong trường hợp NDS có sự thay đổi (cập nhật các chỉ số có tính toán sai hoặc chưa cập nhật đủ dữ liệu), tại bước đổ dữ liệu vào DDS trong mỗi bước ETL tại từng bảng, ta sẽ kiểm tra nếu không đổ được dữ liệu thành công do trùng các khóa chính, ta sẽ tiến hành cập nhật giá trị mới cho các giá trị tại dòng đó.**A picture containing chart

  Description automatically generated

Hình 7: Bước đọc dữ liệu tại bảng nds\_lichsugia và cập nhật vào bảng giatricophieu

1. **Cài đặt**
   1. **Cài đặt NDS và DDS**

Graphical user interface, application

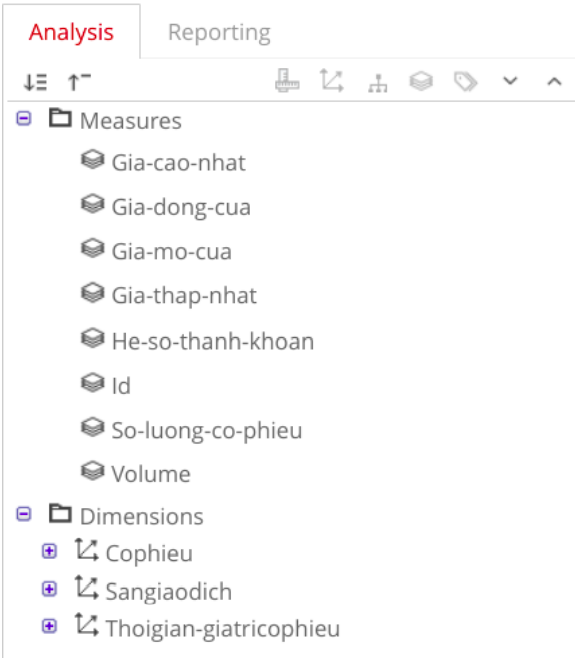
Description automatically generatedNDS được cài đặt trên MySQL, trong khi DDS được cài đặt trên PostgresQL.

A picture containing graphical user interface

Description automatically generatedHình 8: Cài đặt các bảng đã thiết kế tại NDS và DDS

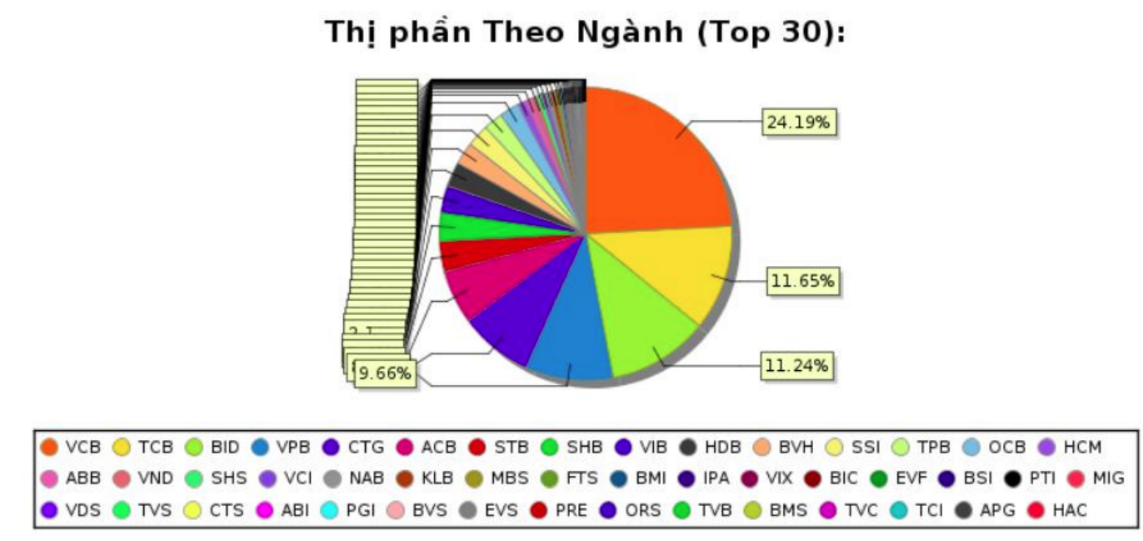
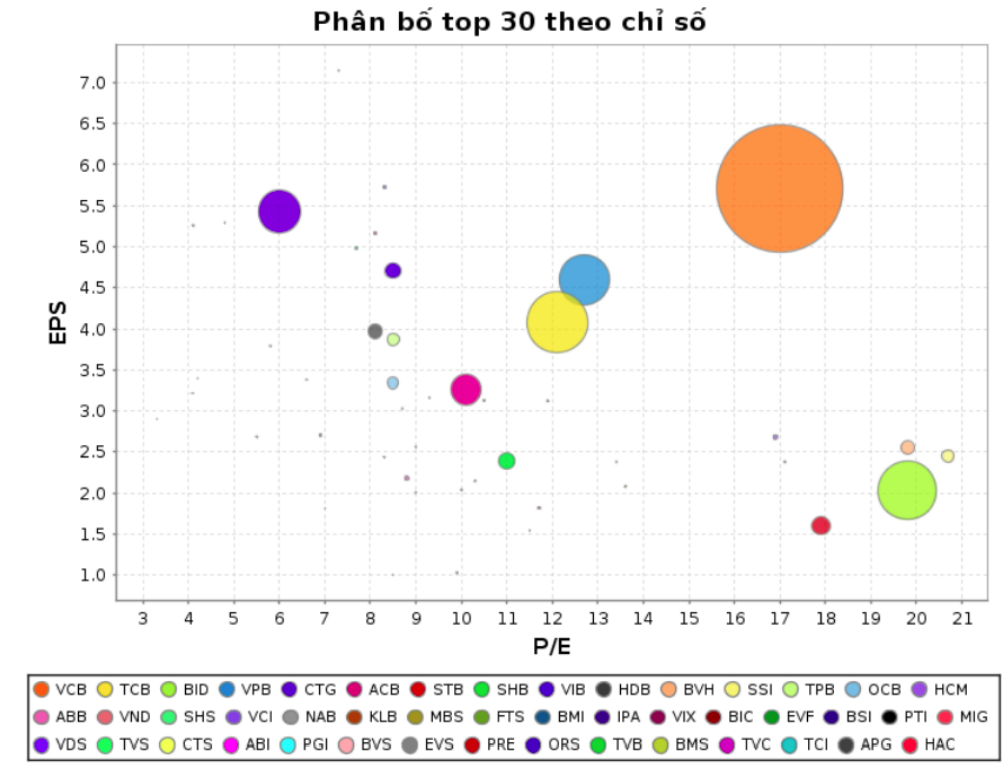
* 1. **Analysis OLAP**

Mô hình Analysis dạng MDB được cài đặt trên Pentaho CDE để thực hiện truy vấn theo chiều các thông tin liên quan đến giao dịch biến động giá của một loại cổ phiếu hoặc của toàn thị trường, trong một khoảng thời gian nhất định (ngày, tuần, tháng, quý hoặc năm). Cụ thể:

* Measure được chọn là dữ liệu giá cao nhất, giá thấp nhất, giá đóng cửa, giá mở cửa, hệ số thanh khoản và volume cổ phiếu ở mỗi phiên giao dịch.
* Các dimension hỗ trợ truy xuất cho nghiệp vụ này bao gồm dữ liệu CoPhieu, dữ liệu SanGiaoDich và danh sách các loại thời gian tương ứng.

Hình 8: Cài đặt mô hình analysis dạng cube

1. **Báo cáo report**

Mục tiêu cuối cùng của hệ thống kho dữ liệu là cung cấp các metadata cho quá trình tạo các report và trực quan hóa thông tin một hay nhiều loại cổ phiếu, và thống kê theo từng nhóm ngành, phân ngành cụ thể. Bằng cách sử dụng công cụ Pentaho Report Designer và Pentaho CDE, nhóm đã dựng được 2 báo cáo cơ bản phục vụ cho 2 nghiệp vụ chính trong phân tích cơ bản và phân tích kỹ thuật đối với một nhóm các cổ phiếu được chỉ định.

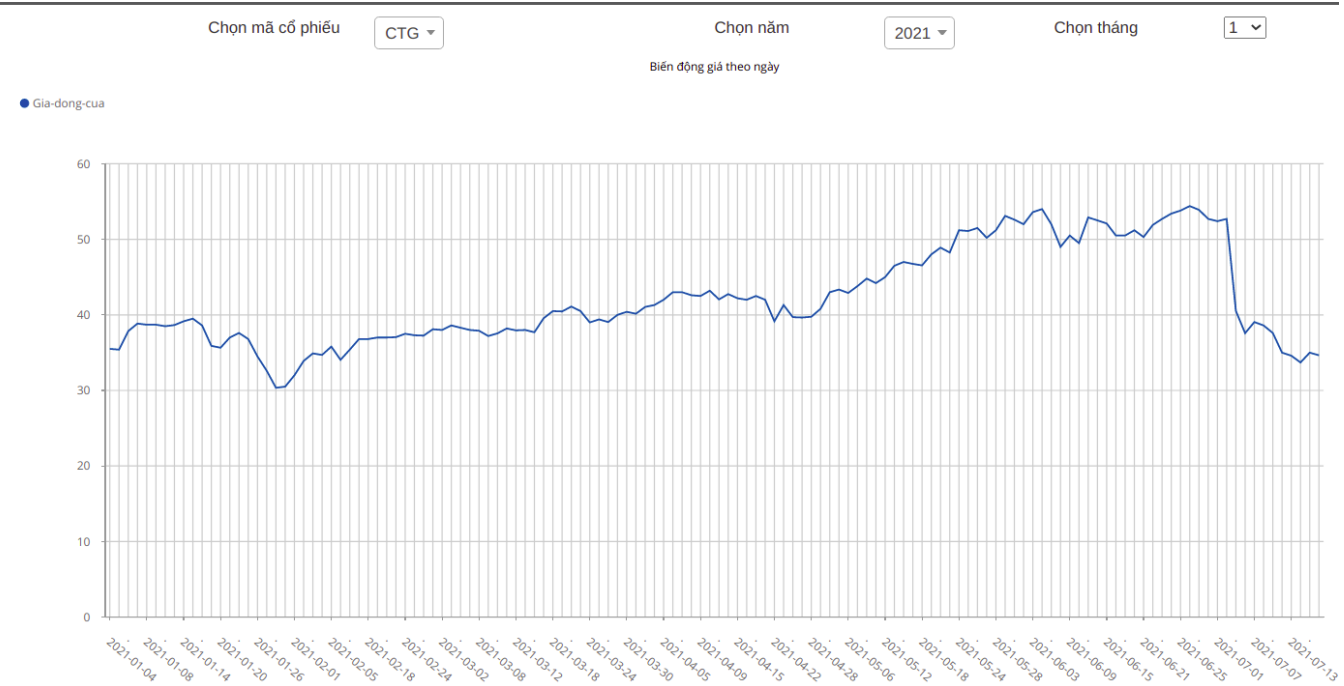
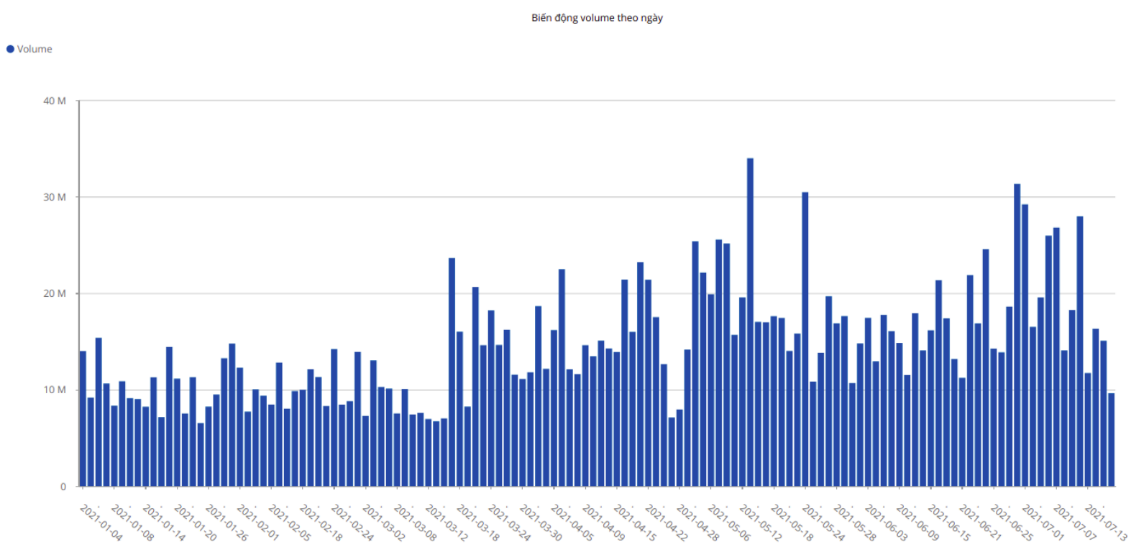
Hình 2: Thị phần các cổ phiếu nhóm ngành tài chính

Hình 3:Tương quan giữa các mã cổ phiếu ngành tài chính

Nghiệp vụ tổng quan đầu tiên là tổng kết thị phần theo một nhóm ngành chung để quan sát tổng thể sức khỏe và tình hình phát triển của nhóm ngành đó. Bên cạnh đó, báo cáo cho thấy một cách trực quan sức ảnh hưởng của các công ty đầu ngành, kèm theo mức độ đánh giá của thị trường đối với cổ phiếu đó đang là cao hay thấp. Chẳng hạn đối với nhóm cổ phiếu ngành tài chính như ví dụ sau:

Các cổ phiếu vốn hóa lớn như VCB, TCB, BID thể hiện sự ổn định hơn trong kinh doanh (EPS cao) so với các tổ chức vốn hóa nhỏ, đồng thời cũng được định giá cao hơn (P/E cao) so với các tổ chức vốn hóa nhỏ. Tuy nhiên, biểu đồ trên có ngoại lệ là CTG(ngân hàng Công Thương) có EPS cao nhưng lại bị thị trường định giá thấp, trong khi vốn hóa chiếm gần 10% thị phần.

Đối với nghiệp vụ phân tích biến động giá, hệ thống cung cấp các biểu đồ cơ bản mô tả biến động giá và khối lượng mua bán của một mã cổ phiếu trong dài hạn (tính theo tuần, tháng hoặc quý). Biểu đồ thể hiện quy luật cung-cầu của thị trường và sự phản ứng của các nhà đầu tư với các sự kiện kinh tế, chính sách tiền tệ và các lần công bố báo cáo tài chính của công ty. Chẳng hạn ta xét biến động giá của cổ phiếu bất thường CTG trên, ta có biểu đồ giá và khối lượng như sau:



Hình 4: Biến động giá và khối lượng của cổ phiếu CTG

1. **Nhận xét và kết luận**

Nghiệp vụ phân tích đầu tư chứng khoán nói chung và cổ phiếu ở thị trường Việt Nam nói riêng là vô cùng phức tạp và chuyên sâu. Trong phạm vi đồ án môn học này, thông qua việc xây dựng kho dữ liệu cổ phiếu, nhóm đã hiện thực hóa được các quy trình cần thiết để tập hợp, tổ chức dữ liệu nhằm cung cấp các thông tin cơ bản làm thước đo tham khảo cho các tổ chức hoặc cá nhân tham gia đầu tư vào thị trường. Tuy nhiên, để thực sự phát huy được vai trò của kho dữ liệu trong nghiệp vụ này, ta cần thêm nhiều dữ liệu cụ thể hơn nữa từ doanh nghiệp và các sàn chứng khoán, bao gồm tình hình kinh doanh của công ty, các báo cáo tài chính của doanh nghiệp, lịch sử trả cổ tức và cổ phiếu thưởng, bảng cân đối kế toán, danh sách các cổ đông lớn và các cổ đông thường xuyên giao dịch. Về phía thị trường ta cần thêm dữ liệu giao dịch của khối ngoại, cũng như thống kê giao dịch tổng thể của toàn thị trường theo mỗi ngày...

Thông qua đồ án, nhóm đã có cơ hội thực hành các phương pháp phân tích, thiết kế kho dữ liệu và cài đặt các mô hình giúp hiện thực hóa kho dữ liệu, bao gồm quy trình ETL, mô hình NDS, DDS, OLAP và kỹ thuật trực quan hóa dữ liệu.

# References

|  |  |
| --- | --- |
| [1] | Cafef, "Kênh thông tin kinh tế - tài chính Việt Nam," Cafef, [Online]. Available: https://cafef.vn. |
| [2] | Cophieu68, "Chứng Khoán, Cổ Phiếu, Tin Chứng Khoán, Thị Trường Chứng Khoán, Chứng Khoán Việt Nam," [Online]. Available: https://www.cophieu68.vn/. |