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 GIỮA KÌ**

**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, 05-2021

**Mục Lục**

[**Mục Lục** 1](#_Toc73189363)

[**Bảng danh sách các hình** 2](#_Toc73189364)

[**Bảng thông tin chi tiết nhóm** 3](#_Toc73189365)

[**1.** **ĐẶT VẤN ĐỀ:** 4](#_Toc73189366)

[**1.1.** **Bài toán phân tích giá cổ phiếu ở thị trường Việt Nam** 4](#_Toc73189367)

[**1.2.** **Đối tượng sử dụng hệ thống** 5](#_Toc73189371)

[**2.** **Mô tả các nghiệp vụ** 5](#_Toc73189377)

[**2.1.** **Nghiệp vụ phân tích cơ bản** 5](#_Toc73189378)

[**2.2.** **Nghiệp vụ phân tích kỹ thuật** 6](#_Toc73189379)

[**2.3.** **Nghiệp vụ so sánh 2 cổ phiếu** 7](#_Toc73189380)

[**3.** **Dữ liệu đầu vào** 9](#_Toc73189381)

[**3.1.** **Dữ liệu thông tin doanh nghiệp** 9](#_Toc73189382)

[**3.2.** **Dữ liệu lịch sử giá theo ngày** 9](#_Toc73189383)

[**3.3.** **Dữ liệu thống kê giao dịch theo ngày** 9](#_Toc73189384)

[**3.4.** **Dữ liệu các giao dịch của khối ngoại** 9](#_Toc73189385)

[**3.5.** **Dữ liệu các giao dịch của cổ đông lớn, cổ đông nội bộ** 9](#_Toc73189386)

[**4.** **Kiến trúc kho** 10](#_Toc73189387)

[**4.1.** **Sơ đồ tổng quát** 10](#_Toc73189388)

[**4.2.** **Mô hình NDS** 10](#_Toc73189389)

[**4.3.** **Mô hình DDS** 12](#_Toc73189390)

[**5.** **Cài đặt** 12](#_Toc73189391)

[References 13](#_Toc73189392)

# **Bảng danh sách các hình**

[Hình A Thị trường chứng khoán Việt Nam ngày 28/05/2021 [1] 4](file:///C:\Users\LAP11777-local\Desktop\ReportBI.docx#_Toc73191369)

[Hình B: Biểu đồ hỗ trợ phân tích kỹ thuật 7](file:///C:\Users\LAP11777-local\Desktop\ReportBI.docx#_Toc73191370)

[Hình C: Sơ đồ chức năng tổng quát của hệ thống 8](file:///C:\Users\LAP11777-local\Desktop\ReportBI.docx#_Toc73191371)

[Hình D: Minh họa dữ liệu thông tin cơ bản của cổ phiếu. Nguồn: cafef [1] 9](file:///C:\Users\LAP11777-local\Desktop\ReportBI.docx#_Toc73191372)

[Hình E: Minh họa lịch sử giá của HPG. Nguồn cophieu68.vn [2] 9](file:///C:\Users\LAP11777-local\Desktop\ReportBI.docx#_Toc73191373)

[Hình F: Kiến trúc kho dữ liệu 10](#_Toc73191374)

[Hình G: Mô hình NDS sơ khởi 11](file:///C:\Users\LAP11777-local\Desktop\ReportBI.docx#_Toc73191375)

[Hình H: Mô hình DDS sơ khởi 12](#_Toc73191376)

# **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 |  |  |
| Tìm các nguồn dữ liệu cổ phiếu Việt Nam |
| Phân tích các chức năng cần có của hệ thống |
|  | Xây dựng mô hình NDS |  |  |
|  | Viết báo cáo và làm slide |  |  |
| 20C12030  Huỳnh Lâm Phú Sĩ | Tìm hiểu quy trình nghiệp vụ phân tích kỹ thuật của cổ phiếu |  |  |
| Xây dựng kiến trúc kho dữ liệu |
| Xây dựng mô hình DDS |
|  | 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:**

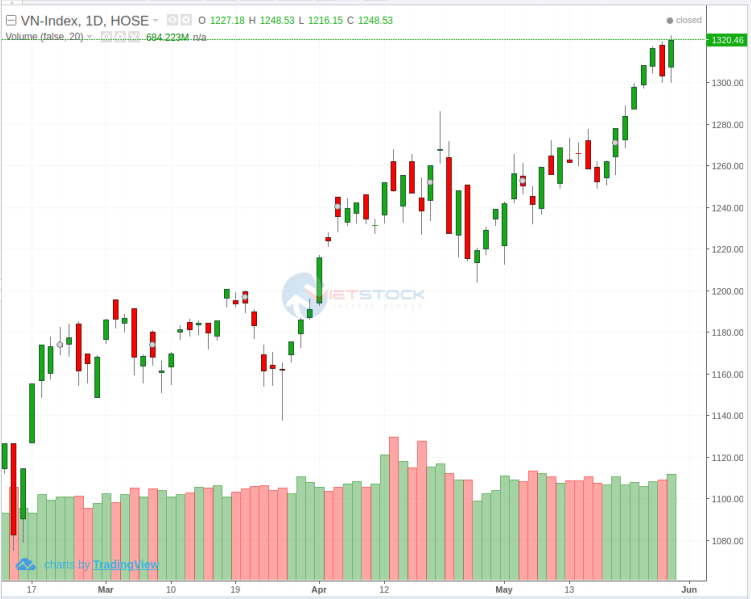
* + Chỉ số P/E của từng cổ phiếu
  + Phân bổ chỉ số P/E của toàn thị trường
  + Xác định một cổ phiếu bất kỳ thuộc mức độ nào trên thị trường
  + Xác định một cổ phiếu bất kỳ thuộc mức độ nào trong nhóm ngành tương ứng

#### **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. **Nghiệp vụ so sánh 2 cổ phiếu**

So sánh giá trị và sự khác nhau, giống nhau giữa 2 hay nhiều cổ phiếu là nghiệp vụ quan trọng giúp nhà đầu tư cân nhắc và đối sánh các cổ phiếu với nhau, từ đó ra quyết định khi có sự phân vân giữa các loại cổ phiếu.

Thông thường, các chỉ số cần so sánh với nhau bao gồm 2 nhóm: chỉ số cơ bản và chỉ số kỹ thuật

**Chỉ số cơ bản:** EPS, P/E, P/B

**Chỉ số kỹ thuật:**

* Tương quan biến động giá trong tuần/tháng/3 tháng/6 tháng gần nhất
* Tương quan tăng trưởng tổng thể sau các năm

#### **Thông tin đầu vào:**

* Thông tin chỉ số EPS, P/E, P/B hiện tại của 2 cổ phiếu
* Lịch sử biến động giá của 2 cổ phiếu từ khi lên sàn

#### **Kết quả đầu ra:**

* Tương quan Chỉ số P/E, EPS,... thể hiện dưới dạng biểu đồ cột hoặc dòng ngang (Bar chart)
* Tương quan tăng trưởng tổng thể sau các năm như ví dụ trên, thể hiện ở dạng bảng hoặc dạng Bar chart

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**

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

1. **Dữ liệu thống kê giao dịch theo ngày**

Dữ liệu giá của từng cổ phiếu trên 2 sàng HSX và HNX theo từng ngày ([cophieu68.vn](http://cophieu68.vn/)), định dạng file txt hoặc csv

1. **Dữ liệu các giao dịch của khối ngoại**

Dữ liệu giao hàng ngày của các cổ đông lớn/nội bộ của công ty ([stockbiz.vn](http://stockbiz.vn/))

1. **Dữ liệu các giao dịch của cổ đông lớn, cổ đông nội bộ**

Dữ liệu hằng ngày của các cổ đông lớn/nội bộ của công ty (stockbiz.vn)

1. **Lưu trữ dữ liệu tại nguồn**

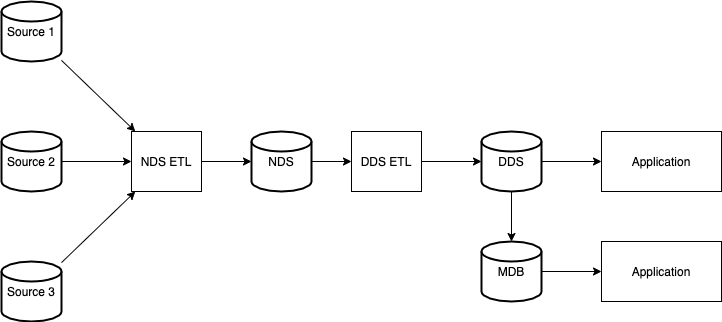
Dữ liệu tại nguồn bao gồm dữ liệu đến từ 3 nguồn như sau:

* 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)

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 dataware house
    - 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 quy trình ETL

Hình F: Kiến trúc kho dữ liệu

1. **Mô hình NDS**

Diagram

Description automatically generated

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 1: 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 2: 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 3: 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 4: 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 5: 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 6: Bước đọc dữ liệu tại bảng nds\_lichsugia và cập nhật vào bảng giatricophieu

1. **Cài đặt**
   1. **Cài đặt dữ liệu nguồn**
   2. **Cài đặt NDS**

NDS được cài đặt trên MySQL.

Graphical user interface, application

Description automatically generated

Hình 7: Cài đặt tại NDS

* 1. **Cài đặt DDS**

DDS được cài đặt trên PostgresQL.

A picture containing graphical user interface

Description automatically generated

Hình 8: Cài đặt tại DDS

* 1. **OLAP**

1. **Báo cáo report**

# 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/. |