|  |  |
| --- | --- |
|  |  |
| ĐOÀN DUY THÀNH | **BỘ CÔNG THƯƠNG**  **TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI**  **---------------------------------------** |
| ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC KHOA HỌC MÁY TÍNH  **XÂY DỰNG ỨNG DỤNG LÀM SẠCH, LƯU TRỮ, PHÂN TÍCH VÀ TRỰC QUAN HÓA CÁC YẾU TỐ PHÁT TRIỂN CỦA DOANH NGHIỆP SỬ DỤNG PYTHON, MYSQL VÀ POWER BI** |
| |  |  | | --- | --- | | **CBHD:** | **ThS. Vũ Duy Giang** | | **Sinh viên:** | **Đoàn Duy Thành** | | **Mã sinh viên:** | **2020603903** |   Hà Nội – Năm 2024 |
| KHOA HỌC MÁY TÍNH |

# LỜI CẢM ƠN

Trước tiên với tình cảm sâu sắc và chân thành nhất, cho phép em được bày tỏ lòng biết ơn đến các thầy cô của trường Đại học Công Nghiệp Hà Nội, đặc biệt là các thầy cô khoa Công Nghệ Thông Tin của trường đã tạo điều kiện cho em có một kỳ đồ án tốt nghiệp.

Và em cũng xin chân thành cảm ơn thầy giáo ThS. Vũ Duy Giang người đã tận tâm hướng dẫn em qua từng buổi nói chuyện, hướng dẫn, thảo luận về các lĩnh vực trong đề tài. Em rất biết ơn các góp ý, sự chỉ dẫn và kiến thức chuyên môn mà thầy giáo đã chia sẻ với em trong quá trình thực hiện báo cáo. Nhờ những lời khuyên và hướng dẫn của thầy đã giúp em đã có cơ hội học hỏi và phát triển kỹ năng của mình.

Không chỉ hướng dẫn và hỗ trợ em hoàn thành đồ án tốt nghiệp một cách thành công, sự đóng góp của thầy giáo còn giúp em hiểu rõ hơn về lĩnh vực chuyên môn và củng cố kiến thức hoàn thành đồ án tốt nghiệp. Nhờ thầy giáo mà em đã có cơ hội nắm bắt những khía cạnh quan trọng và phát triển tư duy sáng tạo. Đồ án này không thể hoàn thành một cách xuất sắc như vậy nếu không có sự giúp đỡ của thầy. Em rất cảm ơn sự tâm huyết và nhiệt tình giúp đỡ trong xuyên suốt quá trình thực hiện báo cáo này.

Trong quá trình thực hiện báo cáo, cũng như là trong quá trình thực hiện đề tài đã chọn, em cũng đã rất cố gắng xây dựng và hoàn thiện sản phẩm. Tuy nhiên trong quá trình thực hiện sẽ khó tránh khỏi những thiết hụt và sai sót. Em rất mong nhận được ý kiến đóng góp từ thầy cô để học thêm được nhiều kinh nghiệm và sẽ hoàn thành tốt bài báo cáo tốt nghiệp. Em xin chúc thầy luôn mạnh khỏe, tràn đầy may mắn và thành công trong cuộc sống.

Em xin chân thành cảm ơn!

**Sinh viên thực hiện**

Đoàn Duy Thành

**MỤC LỤC**

[CHƯƠNG 1. LỜI CẢM ƠN 2](#_Toc175469383)

[DANH MỤC HÌNH ẢNH 4](#_Toc175469384)

[DANH MỤC BẢNG BIỂU 5](#_Toc175469385)

[LỜI MỞ ĐẦU 1](#_Toc175469386)

[1. Lý do lựa chọn đề tài 1](#_Toc175469387)

[2. Mục tiêu của đề tài 2](#_Toc175469388)

[3. Phạm vi đề tài 2](#_Toc175469389)

[4. Dự kiến kết quả 2](#_Toc175469390)

[5. Bố cục đề tài 3](#_Toc175469391)

[CHƯƠNG 1. TỔNG QUAN VỀ ĐỀ TÀI 4](#_Toc175469392)

[1.1. Khái quát về thị trường 4](#_Toc175469393)

[1.2. Các kỹ thuật giải quyết trong đề tài 4](#_Toc175469394)

[1.2.1. Tổng quan về ngôn ngữ Python 4](#_Toc175469395)

[1.2.2. Tổng quan về các Framework 5](#_Toc175469396)

[1.2.3. Tổng quan về SQL và MySQL 8](#_Toc175469397)

[1.2.4. Tổng quan về Power BI 9](#_Toc175469398)

[CHƯƠNG 2. PHÂN TÍCH THIẾT KẾ 12](#_Toc175469399)

[2.1. Yêu cầu hệ thống 12](#_Toc175469400)

[2.1.1. Phân tích yêu cầu 12](#_Toc175469401)

[2.1.2. Phân tích yêu cầu phi chức năng 12](#_Toc175469402)

[2.2. Xây dựng biểu đồ usecase 12](#_Toc175469403)

[2.2.1. Biểu đồ usecase hệ thống 12](#_Toc175469404)

[2.2.2. Đặc tả chi tiết usecase 12](#_Toc175469405)

# DANH MỤC HÌNH ẢNH

# DANH MỤC BẢNG BIỂU

# LỜI MỞ ĐẦU

## Lý do lựa chọn đề tài

Trong bối cảnh doanh nghiệp hiện đại ngày nay, việc quản lý và phân tích dữ liệu trở thành một yếu tố quan trọng quyết định sự thành công và phát triển bền vững. Đề tài "Xây dựng ứng dụng làm sạch, lưu trữ, phân tích và trực quan hóa các yếu tố phát triển của doanh nghiệp sử dụng Python, MySQL và Power BI" được chọn nhằm đáp ứng nhu cầu cấp thiết này.

1. **Tăng cường quản lý dữ liệu**: Doanh nghiệp ngày càng phải đối mặt với khối lượng dữ liệu khổng lồ từ nhiều nguồn khác nhau. Việc làm sạch và lưu trữ dữ liệu hiệu quả giúp đảm bảo chất lượng dữ liệu đầu vào cho các phân tích và quyết định chiến lược. Python, với các thư viện mạnh mẽ như Pandas, là công cụ lý tưởng để xử lý và làm sạch dữ liệu.
2. **Cải thiện phân tích dữ liệu**: Để đưa ra quyết định chính xác, doanh nghiệp cần phân tích dữ liệu một cách chi tiết và toàn diện. MySQL cung cấp một hệ quản trị cơ sở dữ liệu quan hệ hiệu quả để lưu trữ và quản lý dữ liệu.
3. **Trực quan hóa thông tin hiệu quả**: Dữ liệu phân tích cần được trình bày một cách trực quan để dễ dàng tiếp cận và hiểu rõ. Power BI là công cụ mạnh mẽ trong việc tạo báo cáo và bảng điều khiển tương tác, giúp các nhà quản lý và các bên liên quan nhanh chóng nắm bắt thông tin và đưa ra quyết định chiến lược.
4. **Ứng dụng thực tiễn và tiềm năng phát triển**: Đề tài này không chỉ mang lại giá trị học thuật mà còn có ứng dụng thực tiễn cao trong môi trường doanh nghiệp. Việc triển khai một hệ thống tích hợp các công cụ như Python, MySQL, và Power BI không chỉ giúp nâng cao hiệu quả quản lý dữ liệu mà còn mở ra cơ hội cải thiện quy trình làm việc và quyết định dựa trên dữ liệu.
5. **Đáp ứng xu hướng công nghệ hiện đại**: Công nghệ phân tích dữ liệu đang trở thành xu hướng chủ đạo trong ngành công nghệ thông tin. Đề tài này phản ánh sự chuyển mình của ngành công nghiệp, từ việc dựa vào trực giác sang việc sử dụng dữ liệu và phân tích để định hướng chiến lược.

.

## Mục tiêu của đề tài

* Xây dựng ứng dụng có khả năng thu thập và làm sạch dữ liệu: Phát triển một ứng dụng hỗ trợ việc thu thập và làm sạch dữ liệu một cách nhanh chóng trước khi đưa vào cơ sở dữ liệu công ty nhằm đảm bảo tính chính xác và nhất quán của dữ liệu. Sử dụng Python để xử lý dữ liệu thô, loại bỏ lỗi và cảnh báo các lỗi trước khi cập nhật dữ liệu.
* Thiết kế và triển khai cơ sở dữ liệu: Tạo ra một cơ sở dữ liệu MySQL để lưu trữ dữ liệu đã làm sạch. Cơ sở dữ liệu này sẽ bao gồm các bảng và quan hệ cần thiết để tổ chức và quản lý dữ liệu một cách hiệu quả, hỗ trợ việc truy vấn và báo cáo.
* Tạo bảng điều khiển và báo cáo trực quan: Sử dụng Power BI để thiết kế các bảng điều khiển tương tác và báo cáo trực quan. Các báo cáo và bảng điều khiển này sẽ giúp người dùng dễ dàng theo dõi và hiểu các yếu tố phát triển của doanh nghiệp thông qua các biểu đồ, đồ thị, và số liệu trực quan.
* Tích hợp và tự động hóa quy trình: Xây dựng một ứng dụng tích hợp để kết nối các thành phần làm sạch dữ liệu, lưu trữ, phân tích và trực quan hóa.

## Phạm vi đề tài

Về các khía cạnh xử lý trong đề tài, giới hạn và phạm vi sẽ tập trung xử lý các mục sau:

* + Về phạm vi công nghệ: sử dụng Python, MySQL và Power BI
  + Phạm vi chức năng: Thu thập, làm sạch, lưu trữ, phân tích và trực quan hóa dữ liệu
  + Phạm vi dữ liệu: Dữ liệu sẽ lấy từ 4 phòng ban chính của một doanh nghiệp bao gồm Sales, Marketing, Administration, Accounting.

## Dự kiến kết quả

Các kết quả dự kiến sau khi hoàn thành đề tài này bao gồm:

* Xây dựng được ứng dụng có thể thu thập, làm sạch và lưu trữ dữ liệu
* Tối ưu hóa cơ sở dữ liệu với chi phí thực thi thấp
* Thiết kế bảng điều khiển và báo cáo trực quan tự động

## Bố cục đề tài

Nội dung quyển báo cáo đồ án tốt nghiệp sẽ bao gồm các chương sau:

**Chương 1: Tổng quan về đề tài**

Trong chương 1, em sẽ nêu ra lý thuyết cơ sở, các kiến thức liên quan tới đề tài.

**Chương 2: Phân tích thiết kế**

Sau khi đưa ra được phát biểu, nêu rõ được công nghệ, cơ sở lý thuyết và xác định được yêu cầu của bài toán, em sẽ đi phân tích và thiết kế các chức năng của bài toán qua các tài liệu.

**Chương 3: Tiến hành thực nghiệm**

Ở chương này, em sẽ tiến hành thực nghiệm thiết kế cơ sở dữ liệu, ứng dụng và bảng điều khiển dựa theo các phân tích từ chương 2.

**Chương 4: Kiểm thử hệ thống**

Ở chương cuối, em sẽ tổng hợp kết quả và kiểm thử sau khi đã hoàn thành thiết kế phát triển ứng dụng ở chương 3.

**Kết luận:**

Cuối cùng là kết luận, tổng hợp các kết quả, các hướng phát triển và mở rộng trong tương lai.

# TỔNG QUAN VỀ ĐỀ TÀI

## Khái quát về thị trường

Trong thời buổi công nghệ hiện nay, chúng ta đã thực hiện được nhiều công việc với tốc độ nhanh hơn và chi phí thấp hơn nhiều so với cách thức truyền thống qua những cách thức mới dễ tiếp cận và tương tác với người dùng hơn. Chính điều này, đã thúc đẩy sự khai sinh và phát triển của nhiều lĩnh vực khiến các công ty mới ra đời ngày càng nhiều điều này giúp nâng cao và biến đổi đáng kể văn hóa, nâng cao chất lượng cuộc sống con người hiện nay.

Trong năm 2023 theo thống kê Cục Quản lý đăng ký kinh doanh cho biết, Năm nay, gần 160.000 doanh nghiệp đăng ký thành lập mới, cao gấp 1,2 lần mức bình quân giai đoạn 2017-2022 và tăng 4,6% so với dự kiến thực hiện cả năm.

Với sự gia tăng nhanh chóng của các doanh nghiệp mới và khối lượng dữ liệu mà họ tạo ra, việc quản lý và phân tích dữ liệu trở thành một thách thức quan trọng. Nhiều doanh nghiệp mới gặp khó khăn trong việc lưu trữ và phân tích dữ liệu để ra quyết định chiến lược. Sự cần thiết phải có các công cụ và giải pháp hiệu quả trong việc quản lý dữ liệu trở nên cấp bách hơn bao giờ hết.

Việc quản lý và phân tích các thông số dữ liệu là điều bắt buộc của các cấp quản lý doanh nghiệp nếu như muốn đưa ra một chiến lược dài hạn và có lợi cho công ty trong việc phát triển. Tuy nhiên việc phải xây dựng báo cáo hàng tháng một cách thủ công sẽ tạo ra một gánh nặng không cần thiết cho các nhân sự làm báo cáo, đặc biệt là việc xây dựng một báo cáo đủ thông tin sẽ rất tốn thời gian và công sức một cách vô nghĩa trong thời đại hiện nay.

Chính vì vậy, đối với một công ty, việc lưu trữ và trực quan hóa dữ liệu của công ty một cách tự động, đầy đủ và chính xác là một việc rất cần thiết. Và đó cũng chính là mục tiêu của đề tài “Xây dựng ứng dụng làm sạch, lưu trữ, phân tích và trực quan hóa các yếu tố phát triển của doanh nghiệp sử dụng Python, MySQL và Power BI”.

## Các kỹ thuật giải quyết trong đề tài

### Tổng quan về ngôn ngữ Python

Python là một ngôn ngữ lập trình được sử dụng rộng rãi trong các ứng dụng web, phát triển phần mềm, khoa học dữ liệu và máy học (ML). Các nhà phát triển sử dụng Python vì nó hiệu quả, dễ học và có thể chạy trên nhiều nền tảng khác nhau. Phần mềm Python được tải xuống miễn phí, tích hợp tốt với tất cả các loại hệ thống và tăng tốc độ phát triển.

Những lợi ích của Python bao gồm:

* Các nhà phát triển có thể dễ dàng đọc và hiểu một chương trình Python vì ngôn ngữ này có cú pháp cơ bản giống tiếng Anh.
* Python giúp cải thiện năng suất làm việc của các nhà phát triển vì so với những ngôn ngữ khác, họ có thể sử dụng ít dòng mã hơn để viết một chương trình Python.
* Python có một thư viện tiêu chuẩn lớn, chứa nhiều dòng mã có thể tái sử dụng cho hầu hết mọi tác vụ. Nhờ đó, các nhà phát triển sẽ không cần phải viết mã từ đầu.
* Các nhà phát triển có thể dễ dàng sử dụng Python với các ngôn ngữ lập trình phổ biến khác như Java, C và C++.
* Cộng đồng Python tích cực hoạt động bao gồm hàng triệu nhà phát triển nhiệt tình hỗ trợ trên toàn thế giới. Nếu gặp phải vấn đề, bạn sẽ có thể nhận được sự hỗ trợ nhanh chóng từ cộng đồng.
* Trên Internet có rất nhiều tài nguyên hữu ích nếu bạn muốn học Python. Ví dụ: bạn có thể dễ dàng tìm thấy video, chỉ dẫn, tài liệu và hướng dẫn dành cho nhà phát triển.
* Python có thể được sử dụng trên nhiều hệ điều hành máy tính khác nhau, chẳng hạn như Windows, macOS, Linux và Unix.

### Tổng quan về các Framework

#### Tổng quan về PyQt6

PyQt6 là một bộ công cụ (framework) mạnh mẽ để phát triển ứng dụng đồ họa và giao diện người dùng (GUI) trong Python. Được xây dựng trên nền tảng Qt 6, PyQt6 cung cấp một lớp kết nối giữa Python và Qt, giúp các nhà phát triển dễ dàng tạo ra các ứng dụng desktop với giao diện đồ họa phong phú và hiện đại.

**Qt và PyQt6**

* Qt: Qt là một framework phát triển ứng dụng đa nền tảng mạnh mẽ được phát triển bởi The Qt Company. Nó cung cấp các công cụ và thư viện để tạo giao diện người dùng đồ họa, xử lý sự kiện, và quản lý các thành phần của ứng dụng. Qt hỗ trợ nhiều hệ điều hành như Windows, macOS, Linux, và các thiết bị di động.
* PyQt6: PyQt6 là một binding của Qt 6 cho Python, cho phép các nhà phát triển Python sử dụng các chức năng của Qt trong ứng dụng của mình. PyQt6 mang đến một cách tiếp cận trực quan và dễ dàng để xây dựng giao diện đồ họa, tương tác với các widget và xử lý sự kiện trong ứng dụng Python.

**Các tính năng chính**

* Giao diện người dùng đồ họa: PyQt6 cung cấp một loạt các widget và công cụ giao diện người dùng, bao gồm các cửa sổ, nút bấm, thanh trượt, menu, hộp thoại, và các thành phần giao diện khác. Điều này cho phép tạo ra các giao diện người dùng phong phú và dễ sử dụng.
* Hỗ trợ đa nền tảng: Ứng dụng được phát triển bằng PyQt6 có thể chạy trên nhiều hệ điều hành khác nhau mà không cần phải thay đổi mã nguồn, nhờ vào khả năng hỗ trợ đa nền tảng của Qt.
* Thiết kế giao diện: PyQt6 hỗ trợ thiết kế giao diện đồ họa bằng cách sử dụng Qt Designer, một công cụ kéo-thả giúp tạo giao diện người dùng mà không cần phải viết mã cho từng phần của giao diện.
* Xử lý sự kiện: PyQt6 cung cấp cơ chế xử lý sự kiện mạnh mẽ, giúp quản lý các sự kiện từ người dùng như nhấp chuột, nhập liệu từ bàn phím, và các sự kiện hệ thống khác.
* Hỗ trợ đồ họa và hoạt hình: PyQt6 cung cấp các công cụ để xử lý đồ họa 2D và 3D, cũng như hỗ trợ các hiệu ứng hoạt hình, giúp tạo ra các ứng dụng có giao diện đẹp mắt và sinh động.

**Cài đặt và sử dụng**

* Cài đặt: PyQt6 có thể được cài đặt thông qua trình quản lý gói pip. Câu lệnh cài đặt đơn giản là pip install pyqt6.
* Sử dụng cơ bản: Sau khi cài đặt, các nhà phát triển có thể bắt đầu sử dụng PyQt6 bằng cách nhập các mô-đun của PyQt6 và tạo các lớp giao diện người dùng kế thừa từ các lớp cơ bản của Qt. Ví dụ, tạo một cửa sổ đơn giản có thể được thực hiện bằng cách tạo một lớp kế thừa từ QMainWindow hoặc QWidget.

#### Tổng quan về Pandas

Pandas là một thư viện mã nguồn mở trong Python, được thiết kế để xử lý và phân tích dữ liệu một cách hiệu quả. Được phát triển bởi Wes McKinney vào năm 2008, Pandas đã nhanh chóng trở thành một trong những công cụ quan trọng nhất trong lĩnh vực phân tích dữ liệu và khoa học dữ liệu. Thư viện này cung cấp các cấu trúc dữ liệu linh hoạt và các công cụ mạnh mẽ để làm việc với dữ liệu có cấu trúc và dữ liệu không có cấu trúc.

**Các cấu trúc dữ liệu chính**

* DataFrame: Đây là cấu trúc dữ liệu chính của Pandas, tương tự như một bảng trong cơ sở dữ liệu hoặc một bảng tính Excel. DataFrame là một cấu trúc dữ liệu hai chiều, có thể lưu trữ dữ liệu của nhiều loại khác nhau (số, chuỗi, ngày tháng, v.v.) và cho phép thực hiện các thao tác như lọc, nhóm, và tính toán trên dữ liệu. Mỗi cột trong DataFrame là một đối tượng Series.
* Series: Series là một cấu trúc dữ liệu một chiều, tương tự như một cột trong DataFrame hoặc một mảng một chiều. Series có thể chứa dữ liệu của nhiều loại khác nhau và có chỉ số (index) giúp truy cập và thao tác dữ liệu một cách thuận tiện.

**Các tính năng chính**

* Đọc và ghi dữ liệu: Pandas hỗ trợ đọc và ghi dữ liệu từ nhiều định dạng khác nhau, bao gồm CSV, Excel, SQL, JSON, HDF5, và nhiều định dạng dữ liệu khác. Điều này giúp dễ dàng nhập và xuất dữ liệu từ và đến các nguồn dữ liệu phổ biến.
* Xử lý dữ liệu: Pandas cung cấp nhiều công cụ để xử lý dữ liệu, bao gồm làm sạch dữ liệu (loại bỏ giá trị thiếu, xử lý dữ liệu trùng lặp), chuyển đổi dữ liệu (thay đổi kiểu dữ liệu, thay đổi cấu trúc dữ liệu), và thao tác với dữ liệu (lọc, sắp xếp, nhóm).
* Phân tích dữ liệu: Thư viện này hỗ trợ nhiều công cụ phân tích dữ liệu, bao gồm thống kê mô tả, tính toán các chỉ số, và áp dụng các hàm phân tích phức tạp. Pandas tích hợp với các thư viện phân tích khác như NumPy để cung cấp các phép toán số học và thống kê nâng cao.
* Trực quan hóa dữ liệu: Pandas tích hợp với các công cụ trực quan hóa như Matplotlib và Seaborn, cho phép tạo ra các biểu đồ và đồ thị trực quan từ dữ liệu trong DataFrame hoặc Series.

**Cài đặt và sử dụng**

* Cài đặt: Pandas có thể được cài đặt dễ dàng thông qua trình quản lý gói pip với câu lệnh pip install pandas.
* Sử dụng cơ bản: Sau khi cài đặt, Pandas có thể được sử dụng bằng cách nhập thư viện vào mã nguồn Python. Một số thao tác cơ bản bao gồm tạo DataFrame, đọc dữ liệu từ tệp, và thực hiện các phép toán trên dữ liệu.

### Tổng quan về SQL và MySQL

SQL (Structured Query Language) là ngôn ngữ tiêu chuẩn được sử dụng để quản lý và thao tác dữ liệu trong các hệ quản trị cơ sở dữ liệu quan hệ (RDBMS). Được phát triển vào đầu những năm 1970 tại IBM bởi Donald D. Chamberlin và Raymond F. Boyce, SQL đã trở thành một công cụ không thể thiếu trong lĩnh vực quản lý cơ sở dữ liệu. SQL cung cấp một cách thức nhất quán và chính xác để thực hiện các hoạt động cơ bản như truy vấn, cập nhật, và quản lý cấu trúc cơ sở dữ liệu.

SQL được quy định bởi các tổ chức tiêu chuẩn hóa như ANSI (American National Standards Institute) và ISO (International Organization for Standardization). Chuẩn SQL định nghĩa các cú pháp và chức năng cơ bản mà mọi hệ quản trị cơ sở dữ liệu phải hỗ trợ. SQL cho phép người dùng thực hiện các hoạt động như:

* **Truy vấn dữ liệu**: SQL cung cấp khả năng truy vấn dữ liệu từ cơ sở dữ liệu thông qua các câu lệnh SELECT, giúp lấy thông tin cần thiết từ các bảng cơ sở dữ liệu.
* **Quản lý dữ liệu**: Các câu lệnh INSERT, UPDATE, và DELETE giúp thêm mới, cập nhật, và xóa dữ liệu trong cơ sở dữ liệu.
* **Quản lý cấu trúc cơ sở dữ liệu**: SQL hỗ trợ việc tạo mới, sửa đổi, và xóa các bảng và cấu trúc cơ sở dữ liệu khác bằng các câu lệnh CREATE, ALTER, và DROP.
* **Quản lý quyền truy cập**: Các câu lệnh GRANT và REVOKE cho phép quản lý quyền truy cập của người dùng vào cơ sở dữ liệu.

**MySQL** là một hệ quản trị cơ sở dữ liệu quan hệ mã nguồn mở, được phát triển bởi Oracle Corporation. Được giới thiệu lần đầu vào năm 1995 bởi MySQL AB, MySQL đã nhanh chóng trở thành một trong những hệ quản trị cơ sở dữ liệu phổ biến nhất nhờ vào hiệu suất cao, tính ổn định, và khả năng mở rộng. MySQL sử dụng SQL làm ngôn ngữ truy vấn chính để quản lý và thao tác dữ liệu, đồng thời cung cấp các công cụ và tính năng bổ sung để hỗ trợ quản lý cơ sở dữ liệu.

MySQL cung cấp nhiều tính năng quan trọng bao gồm:

* **Quản lý cơ sở dữ liệu**: MySQL cho phép người dùng tạo, quản lý, và tổ chức cơ sở dữ liệu và các bảng thông qua các công cụ và tính năng mạnh mẽ. Các thao tác CRUD (Create, Read, Update, Delete) cơ bản được thực hiện dễ dàng và hiệu quả.
* **Hiệu suất và tối ưu hóa**: MySQL bao gồm nhiều công cụ và tính năng để tối ưu hóa hiệu suất cơ sở dữ liệu, như chỉ mục, tối ưu hóa truy vấn, và bộ nhớ đệm. Những tính năng này giúp cải thiện thời gian phản hồi và hiệu suất tổng thể của hệ thống.
* **Bảo mật**: MySQL cung cấp các tính năng bảo mật để bảo vệ dữ liệu, bao gồm kiểm soát truy cập, mã hóa dữ liệu, và khả năng sao lưu và phục hồi cơ sở dữ liệu. Điều này đảm bảo tính toàn vẹn và an toàn của dữ liệu trong các ứng dụng.
* **Hỗ trợ đa nền tảng**: MySQL có khả năng chạy trên nhiều hệ điều hành khác nhau, bao gồm Windows, Linux, và macOS, điều này giúp đảm bảo tính linh hoạt và khả năng tích hợp trong môi trường đa dạng.

### Tổng quan về Power BI

Được phát triển vào năm 2013, Power BI là công cụ phân tích và trực quan hoá dữ liệu nổi bật hàng đầu của Microsoft.

Theo định nghĩa của tập đoàn “cha đẻ”, Power BI là một bộ tổng hợp các dịch vụ, ứng dụng và kết nối trong lĩnh vực phân tích dữ liệu giúp bạn kết nối và chuyển đổi dữ liệu từ các nguồn không liên quan – như bảng tính Excel hoặc một tập hợp các kho dữ liệu trên đám mây – thành một tổ hợp thông tin trực quan và dễ tương tác, đồng thời dễ dàng chia sẻ chúng với các đối tượng khác.

Càng ngày, doanh nghiệp càng cần phải làm việc nhiều hơn với dữ liệu (data). Dù quy mô doanh nghiệp của bạn là lớn hay nhỏ, dữ liệu luôn đóng vai trò quan trọng trong quá trình phát triển và định hướng sự thành công. Đặc biệt, trong bối cảnh cuộc cách mạng về công nghệ số đang bùng nổ, data được coi là một nguồn tài nguyên vô giá. Power BI là một trong các công cụ hỗ trợ đắc lực nhất cho việc phân tích và trực quan hoá dữ liệu thành các dashboard. Điều này không chỉ giúp doanh nghiệp tăng cao hiệu suất làm việc, mà còn ra quyết định sáng suốt hơn dựa trên dữ liệu (data-driven), nâng cao vị thế trong môi trường kinh doanh ngày càng cạnh tranh

**Power BI bao gồm ba ứng dụng chính:**

* **Power BI Desktop:**Là ứng dụng cho nền tảng Windows giúp bạn tạo và chỉnh sửa báo cáo, biểu đồ từ nhiều nguồn khác nhau.
* **Power BI Service (hay còn gọi là Power BI Web):** Là dịch vụ phân tích dữ liệu trực tuyến dạng SaaS (Software-as-a-Service),giúp bạn xuất báo cáo, chia sẻ, quản lý và tương tác với các dữ liệu trực tuyến.
* **Power BI Mobile:** Là các ứng dụng di động cho cả hệ điều hành Windows, iOS và Android, cho phép bạn xem và tương tác với các báo cáo Power BI trên điện thoại hoặc máy tính bảng mọi lúc, mọi nơi.

**Các thành phần**

* **Power BI Query:** Là công cụ tập hợp và liên kết các dữ liệu từ nhiều nguồn khác nhau để tiền xử lý và chuyển đổi trước khi nhập vào báo cáo
* **Power BI Pivot:** Là công cụ chọn lọc và cắt giảm dữ liệu trong báo cáo Power BI trước đó để tạo ra các bảng tổng hợp, biểu đồ và báo cáo tương tác phù hợp
* **Power BI Report Builder:** Là công cụ độc lập để tạo báo cáo theo mẫu có định dạng sẵn để in hoặc chia sẻ trong Power BI Service
* **Power BI Report Server**: Là một máy chủ báo cáo tại chỗ cho phép xuất báo cáo Power BI từ Power BI Desktop và chia sẻ trong nội bộ.

# PHÂN TÍCH THIẾT KẾ

## Yêu cầu hệ thống

### Phân tích yêu cầu chức năng

Ứng dụng được phát triển nhằm mục tiêu hỗ trợ các doanh nghiệp trong việc quản lý và phân tích các yếu tố phát triển, từ đó đưa ra các quyết định chiến lược. Để đạt được mục tiêu này, ứng dụng cần đáp ứng các yêu cầu chức năng sau:

1. **Chức năng làm sạch dữ liệu (Data Cleaning):**
   * **Mô tả:** Ứng dụng cần có khả năng tự động nhận diện và loại bỏ các dữ liệu thiếu sót, dữ liệu trùng lặp, và các lỗi dữ liệu phổ biến khác.
   * **Yêu cầu:**
     + Tích hợp các thuật toán làm sạch dữ liệu hiệu quả.
     + Báo cáo phản hồi trong trường hợp dữ liệu bị lỗi
     + Cung cấp giao diện để người dùng kiểm tra dữ liệu trước khi xác nhận nhập vào
2. **Chức năng lưu trữ dữ liệu (Data Storage):**
   * **Mô tả:** Dữ liệu sau khi được làm sạch cần được lưu trữ an toàn và có cấu trúc trong hệ quản trị cơ sở dữ liệu MySQL.
   * **Yêu cầu:**
     + Cơ sở dữ liệu cần thiết kế để đảm bảo tính toàn vẹn và nhất quán của dữ liệu.
     + Hỗ trợ truy vấn nhanh chóng và hiệu quả để phục vụ quá trình phân tích dữ liệu.
3. **Chức năng trực quan hóa dữ liệu (Data Visualization):**
   * **Mô tả:** Sau khi dữ liệu được phân tích, ứng dụng cần có khả năng trực quan hóa kết quả phân tích thông qua các báo cáo động và biểu đồ.
   * **Yêu cầu:**
     + Tích hợp Power BI để tạo các báo cáo và dashboard trực quan, dễ hiểu.
     + Cung cấp khả năng tương tác với các báo cáo, cho phép người dùng lọc và tùy chỉnh các biểu đồ theo nhu cầu.Phân tích yêu cầu phi chức năng

### Phân tích yêu cầu phi chức năng

Các yêu cầu phi chức năng đề cập đến các tiêu chí mà hệ thống phải tuân thủ để đảm bảo hoạt động hiệu quả, bảo mật, và tương tác tốt với người dùng. Dưới đây là các yêu cầu phi chức năng chính:

1. **Hiệu năng (Performance):**
   * **Mô tả:** Ứng dụng cần đảm bảo khả năng xử lý và phân tích dữ liệu một cách nhanh chóng và hiệu quả.
   * **Yêu cầu:**
     + Thời gian phản hồi cho các truy vấn tính theo giây đối với các tập dữ liệu thông thường.
2. **Khả năng mở rộng (Scalability):**
   * **Mô tả:** Hệ thống cần được thiết kế để có thể dễ dàng mở rộng nếu cần phát triển thêm.
   * **Yêu cầu:**
     + Hệ thống phải có kiến trúc cho phép dễ dàng mở rộng hạ tầng cơ sở và chức năng.
3. **Tính bảo mật (Security):**
   * **Mô tả:** Bảo mật dữ liệu là yếu tố quan trọng đối với ứng dụng, đặc biệt là đối với thông tin nhạy cảm của doanh nghiệp.
   * **Yêu cầu:**
     + Xác thực và phân quyền người dùng chi tiết để hạn chế truy cập trái phép.
4. **Khả năng tương thích (Compatibility):**
   * **Mô tả:** Ứng dụng cần tương thích với nhiều nền tảng và công nghệ khác nhau, cũng như các hệ thống hiện có của doanh nghiệp.
   * **Yêu cầu:**
     + Hỗ trợ nhiều hệ điều hành (Windows, macOS, Linux).
     + Tương thích với các phiên bản khác nhau của MySQL và các hệ thống cơ sở dữ liệu tương tự.
     + Tích hợp dễ dàng với các công cụ BI khác ngoài Power BI nếu cần thiết.
5. **Khả năng sử dụng (Usability):**
   * **Mô tả:** Ứng dụng cần có giao diện thân thiện với người dùng, dễ sử dụng và có trải nghiệm người dùng tốt.
   * **Yêu cầu:**
     + Giao diện trực quan, dễ hiểu, và dễ sử dụng cho cả người dùng không có nền tảng kỹ thuật.
     + Cung cấp tài liệu hướng dẫn sử dụng chi tiết và hỗ trợ người dùng.
6. **Khả năng bảo trì (Maintainability):**
   * **Mô tả:** Ứng dụng cần được thiết kế để dễ dàng bảo trì và cập nhật trong tương lai.
   * **Yêu cầu:**
     + Mã nguồn phải được viết rõ ràng, có chú thích và tuân theo các tiêu chuẩn lập trình.
     + Cung cấp tài liệu kỹ thuật chi tiết để hỗ trợ quá trình bảo trì và phát triển thêm.

## Xây dựng ứng dụng

### Hệ thống cơ sở dữ liệu

Dựa theo phạm vi đề tài, hệ thống cơ sở dữ liệu được phục vụ để lưu trữ cho 4 phòng ban chính bao gồm Sales, Marketing, Administration và Accounting với các yêu cầu và cấu trúc cơ sở dữ liệu như sau.

#### Sales

Phòng Sales sẽ cần quản lý và trực quan hóa các thông tin quan trọng sau để theo dõi hiệu suất và đưa ra các quyết định kinh doanh:

1. Doanh số bán hàng (Sales Revenue)
2. Số lượng đơn hàng (Number of Orders)
3. Tỷ lệ chuyển đổi (Conversion Rate)
4. Hiệu suất của nhân viên bán hàng (Salesperson Performance)
5. Sản phẩm bán chạy (Top-Selling Products)

Để quản lý và giải quyết các yêu cầu trên, em xây dựng cấu trúc cơ sở dữ liệu như sau.

**Products:**

|  |  |  |
| --- | --- | --- |
| Tên cột | Kiểu dữ liệu | Mô tả |
| ProductId | INT UNSIGNED PRIMARY KEY | Mã sản phẩm |
| ProductName | VARCHAR(255) NOT NULL | Tên sản phẩm |
| Description | TEXT NOT NULL | Mô tả sản phẩm |
| UnitPrice | DECIMAL(12,1) NOT NULL | Giá đơn vị |
| Remaining | INT UNSIGNED NOT NULL | Số lượng còn lại |

**Orders:**

|  |  |  |
| --- | --- | --- |
| Tên cột | Kiểu dữ liệu | Mô tả |
| OrderId | INT UNSIGNED PRIMARY KEY | Mã đơn hàng, khóa chính |
| CustomerId | INT UNSIGNED NOT NULL | Mã khách hàng, liên kết đến bảng Leads |
| OrderDate | DATE NOT NULL | Ngày đặt hàng |
| ProductID | INT UNSIGNED NOT NULL | Mã sản phẩm, liên kết đến bảng Products |
| TotalAmount | DECIMAL(12,1) NOT NULL | Tổng giá trị đơn hàng |

**Payments:**

|  |  |  |
| --- | --- | --- |
| Tên cột | Kiểu dữ liệu | Mô tả |
| PaymentId | INT UNSIGNED PRIMARY KEY | Mã thanh toán, khóa chính |
| OrderId | INT UNSIGNED NOT NULL | Mã đơn hàng, liên kết đến bảng Orders |
| PaymentDate | DATE NOT NULL | Ngày thanh toán |
| Amount | DECIMAL(12,1) NOT NULL | Số tiền thanh toán |

**EmployeePerformanceSales:**

|  |  |  |
| --- | --- | --- |
| Tên cột | Kiểu dữ liệu | Mô tả |
| Employee  Performa  nceSalesID | INT UNSIGNED PRIMARY KEY | Mã hiệu suất nhân viên, khóa chính |
| Employee | INT UNSIGNED NOT NULL | Mã nhân viên, liên kết đến bảng Employee |
| Date | DATE NOT NULL | Ngày đánh giá hiệu suất |
| WorkPerfor  manceScore | INT UNSIGNED NOT NULL | Điểm hiệu suất công việc |
| Conscious  nessScore | INT UNSIGNED NOT NULL | Điểm ý thức |
| Customer  Count | INT UNSIGNED NOT NULL | Số lượng khách hàng phục vụ |
| TotalRevenue | DECIMAL(12,1) NOT NULL | Tổng doanh thu |
| |  | | --- | | DealClosure  Rate | | INT UNSIGNED NOT NULL | Tỷ lệ đóng giao dịch thành công |

**KPISales:**

|  |  |  |
| --- | --- | --- |
| Tên cột | Kiểu dữ liệu | Mô tả |
| KPIId | INT UNSIGNED PRIMARY KEY | Mã KPI, khóa chính |
| Employee | INT UNSIGNED NOT NULL | Mã nhân viên, liên kết đến bảng Employee |
| Date | DATE NOT NULL | Ngày đánh giá KPI |
| WorkPerform  anceScore | INT UNSIGNED NOT NULL | Điểm hiệu suất công việc |
| Consciousn  essScore | INT UNSIGNED NOT NULL | Điểm ý thức |
| Customer  Count | INT UNSIGNED NOT NULL | Số lượng khách hàng phục vụ |
| TotalRevenue | DECIMAL(12,1) NOT NULL | Tổng doanh thu |
| DealClosure  Rate | INT UNSIGNED NOT NULL | Tỷ lệ đóng giao dịch thành công |

#### Marketing

#### Administration

#### Accounting

### Giao diện và chức năng ứng dụng

* Cấu trúc giao diện gồm những button, layout nào, vẽ giao diện cơ bản
* Ứng dụng bao gồm các chức năng gì (làm sạch thế nào, lưu trữ vào đâu, bắt lỗi như nào…)

### DAX và Power BI

* Đâu là insight cần thiết nhất ở từng phòng ban
* Để đưa ra các insight đó dùng biểu đồ và dax như nào.