TRƯỜNG ĐẠI HỌC KINH TẾ

**KHOA THỐNG KÊ – TIN HỌC**

–––––––––––––––––––––––––––––––



**BÁO CÁO TỐT NGHIỆP**

**NGÀNH HỆ THỐNG THÔNG TIN QUẢN LÝ**

**CHUYÊN NGÀNH QUẢN TRỊ HỆ THỐNG THÔNG TIN**

**Xây dựng hệ thống quản trị cơ sở dữ liệu bán hàng**

Sinh viên thực hiện : **Nguyễn Hải Nam**

Lớp : **46K21.1**

Đơn vị thực tập : **FPT Software**

Cán bộ hướng dẫn : **Trần Đăng Tín**

Giảng viên hướng dẫn : **ThS. Nguyễn Văn Chức**

**Đà Nẵng, 05/202****4**

NHẬN XÉT CỦA ĐƠN VỊ THỰC TẬP

Họ và tên sinh viên:

Lớp: Khoa: Trường:

Thực tập từ ngày: …./……/ 2024 đến ngày: ........./ .……./ 2024

Tại:

Địa chỉ:

Sau quá trình thực tập tại đơn vị của sinh viên, chúng tôi có một số nhận xét, đánh giá như sau:

| **STT** | **Mục đánh giá** | **Rất tệ** | **Tệ** | **Bình thường** | **Tốt** | **Rất tốt** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | Về thái độ, ý thức, đạo đức, kỷ luật và văn hóa công ty |  |  |  |  |  |
| 2 | Kiến thức chuyên môn |  |  |  |  |  |
| 3 | Khả năng hòa nhập, thích nghi và tác phong nghề nghiệp |  |  |  |  |  |
| 4 | Trách nhiệm, sáng tạo trong công việc |  |  |  |  |  |
| 5 | Sáng tạo trong công việc |  |  |  |  |  |

**Các nhận xét khác** (Khoa chúng tôi mong muốn nhận thêm những ý kiến khác từ quý doanh nghiệp nhằm nâng cao chất lượng đào tạo)

**Điểm:**

……….., ngày .......tháng ......năm 2024

**Xác nhận của đơn vị thực tập**

LỜI CẢM ƠN

Trong thời gian làm bài báo cáo tốt nghiệp, em đã nhận được rất nhiều sự giúp đỡ, sự đóng góp ý kiến và chỉ bảo nhiệt tình của giảng viên, các anh chị đồng nghiệp, gia đình và bạn bè.

Em xin gửi lời cảm ơn chân thành đến ThS. Nguyễn Văn Chức - Giảng viên khoa Thống Kê – Tin Học trường Đại học Kinh Tế - Đại học Đà Nẵng, người đã tận tình hướng dẫn, chỉ bảo trong suốt quá trình em thực tập cũng như làm bài báo cáo này.

Em cũng xin chân thành cảm ơn tới các thầy cô giáo của trường Đại học Kinh Tế - Đại học Đà Nẵng nói chung, các thầy cô giáo trong khoa Thống Kê – Tin Học nói riêng đã dạy dỗ, truyền đạt kiến thức các môn căn bản, các môn chuyên ngành, giúp em có được cơ sở lý thuyết vững vàng và tạo điều kiện cho em trong suốt quá trình học tập và rèn luyện ở trường.

Qua đây em cũng muốn gửi lời cảm ơn đến tập thể ban lãnh đạo, các anh chị trong công ty, đặc biệt là anh Trần Đăng Tín đã tạo cơ hội cho em được thực tập và học hỏi tại công ty. Qua thời gian này, em được học rất nhiều điều mới mẻ và có cơ hội tiếp xúc thực tế với môi trường bên ngoài trường học.

Cuối cùng, em xin gửi lời cảm ơn chân thành đến gia đình và bạn bè đã luôn tạo điều kiện, quan tâm, giúp đỡ và động viên em trong suốt quá trình học tập và hoàn thành bài báo cáo.

Do hạn chế về kiến thức cũng như thời gian thực hiện nên bài báo cáo tốt

nghiệp này chắc chắn sẽ không tránh khỏi những thiếu sót. Em rất mong nhận được ý kiến đóng góp của quý Thầy Cô để có thể hoàn thiện bài báo cáo của mình hơn.

*Em xin chân thành cảm ơn!*

*Sinh viên thực hiện*

***Nguyễn Hải Nam***

LỜI CAM ĐOAN

Em xin cam đoan bài báo cáo thực tập tốt nghiệp: “XÂY DỰNG HỆ THỐNG QUẢN TRỊ CƠ SỞ DỮ LIỆU BÁN HÀNG” là do em thực hiện dưới sự hướng dẫn của ThS. Nguyễn Văn Chức. Ngoài ra không có bất cứ sự sao chép của người khác. Đề tài, nội dung bài báo cáo là sản phẩm mà em đã nỗ lực thực hiện sau quá trình học tập tại trường và trong quá trình thực tập tại doanh nghiệp.

Bài báo cáo tốt nghiệp sử dụng một số tài liệu tham khảo được trích dẫn cụ thể ở mục tài liệu tham khảo. Nếu phát hiện có bất kỳ sự gian lận nào, em xin hoàn toàn chịu trách nhiệm về nội dung bài báo cáo của mình.

*Sinh viên thực hiện*

***Nguyễn Hải Nam***

MỤC LỤC

[NHẬN XÉT CỦA ĐƠN VỊ THỰC TẬP i](#_Toc162213566)

[LỜI CẢM ƠN ii](#_Toc162213567)

[LỜI CAM ĐOAN iii](#_Toc162213568)

[MỤC LỤC iv](#_Toc162213569)

[DANH MỤC HÌNH ẢNH vii](#_Toc162213570)

[DANH MỤC BẢNG BIỂU viii](#_Toc162213571)

[DANH MỤC CÁC TỪ VIẾT TẮT ix](#_Toc162213572)

[LỜI MỞ ĐẦU 1](#_Toc162213573)

[CHƯƠNG 1. TỔNG QUAN VỀ ĐƠN VỊ THỰC TẬP 4](#_Toc162213574)

[1.1. Giới thiệu về công ty: 4](#_Toc162213575)

[1.1.1. Giới thiệu chung: 4](#_Toc162213576)

[1.1.2. Tầm nhìn: 5](#_Toc162213577)

[1.1.3. Sứ mệnh: 5](#_Toc162213578)

[1.1.4. Lĩnh vực hoạt động chính của công ty: 5](#_Toc162213579)

[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 6](#_Toc162213580)

[2.1. Lý thuyết cơ bản phân tích dữ liệu: 6](#_Toc162213581)

[2.1.1. Phân tích dữ liệu là gì? 6](#_Toc162213582)

[*2.1.2.* *Tại sao cần phân tích dữ liệu?* 6](#_Toc162213583)

[*2.1.3.* *Mục tiêu của phân tích dữ liệu:* 6](#_Toc162213584)

[2.2. Tìm hiểu về dữ liệu: 6](#_Toc162213585)

[*2.2.1.* *Khái niệm về dữ liệu:* 6](#_Toc162213586)

[*2.2.2.* *Tại sao lại cần dữ liệu:* 6](#_Toc162213587)

[2.3. Data Transformation: 7](#_Toc162213588)

[*2.3.1.* *Khái niệm:* 7](#_Toc162213589)

[*2.3.2.* *Kỹ thuật:* 7](#_Toc162213590)

[2.4. Mô hình hoá dữ liệu (Data Modeling): 8](#_Toc162213591)

[*2.4.1.* *Khái niệm:* 8](#_Toc162213592)

[*2.4.2.* *Mục đích:* 9](#_Toc162213593)

[*2.4.3.* *Mô hình dữ liệu (Data Model):* 9](#_Toc162213594)

[2.5. Trực quan hóa dữ liệu (Data Visualization): 10](#_Toc162213595)

[*2.5.1.* *Khái niệm:* 10](#_Toc162213596)

[*2.5.2.* *Ưu điểm:* 11](#_Toc162213597)

[*2.5.3.* *Nhược điểm:* 11](#_Toc162213598)

[*2.5.4.* *Lợi ích:* 11](#_Toc162213599)

[2.6. Công cụ Tableau trực quan hóa dữ liệu: 11](#_Toc162213600)

[*2.6.1.* *Khái niệm và các tính năng chính:* 11](#_Toc162213601)

[*2.6.2.* *Cấu tạo của Tableau:* 12](#_Toc162213602)

[*2.6.3.* *So sánh Tableau với Microsoft Power BI:* 12](#_Toc162213603)

[*2.6.4.* *Lợi ích khi sử dụng Tableau:* 13](#_Toc162213604)

[2.7. Hồ dữ liệu (Data Lake): 14](#_Toc162213605)

[*2.7.1.* *Khái niệm:* 14](#_Toc162213606)

[*2.7.2.* *Lợi ích:* 15](#_Toc162213607)

[*2.7.3.* *Kiến trúc hồ dữ liệu:* 15](#_Toc162213608)

[2.8. Kho dữ liệu (Data Warehouse): 16](#_Toc162213609)

[*2.8.1.* *Khái niệm:* 16](#_Toc162213610)

[*2.8.2.* *Các đặc tính của kho dữ liệu:* 16](#_Toc162213611)

[*2.8.3.* *Kiến trúc kho dữ liệu:* 17](#_Toc162213612)

[2.9. Data Lakehouse: 18](#_Toc162213613)

[*2.9.1.* *Khái niệm*: 18](#_Toc162213614)

[*2.9.2.* *Kiến trúc của Data Lakehouse* 18](#_Toc162213615)

[CHƯƠNG 3. BÀI TOÁN PHÂN TÍCH XÂY DỰNG HỆ THỐNG QUẢN TRỊ CƠ SỞ DỮ LIỆU BÁN HÀNG 21](#_Toc162213616)

[3.1. Khảo sát và phân tích yêu cầu: 21](#_Toc162213617)

[CHƯƠNG 4. TRIỂN KHAI THỰC NGHIỆM 22](#_Toc162213618)

[KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 23](#_Toc162213619)

[TÀI LIỆU THAM KHẢO 24](#_Toc162213620)

[PHỤ LỤC 1](#_Toc162213621)

DANH MỤC HÌNH ẢNH

[*Hình 1.1. Logo FPT Software* 4](#_Toc162181317)

[*Hình 2.1. Chuyển đổi định dạng* 7](#_Toc162181318)

[*Hình 2.2. Deduplication* 7](#_Toc162181319)

[*Hình 2.3. Loại bỏ các cột không sử dụng và lặp lại* 8](#_Toc162181320)

[*Hình 2.4. Derivation* 8](#_Toc162181321)

[*Hình 2.5: Lọc* 8](#_Toc162181322)

[*Hình 2.6: Hợp nhất* 8](#_Toc162181323)

[*Hình 2.7. Hồ dữ liệu (Data Lake)* 14](#_Toc162181324)

[*Hình 2.8. Kiến trúc hồ dữ liệu (Data Lake Architecture)* 15](#_Toc162181325)

[*Hình 2.9. Kho dữ liệu (Data Warehouse)* 16](#_Toc162181326)

[*Hình 2.10. Data Lakehouse Architecture* 19](#_Toc162181327)

DANH MỤC BẢNG BIỂU

[*Bảng 1.1. Danh mục các từ viết tắt* ix](#_Toc162181640)

[*Bảng 2.1. So sánh Tableau và Microsoft Power BI* 13](#_Toc162181641)

DANH MỤC CÁC TỪ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| Tên viết tắt | Tên đầy đủ | Ý nghĩa |
| BI | Business Intelligence | Hoạt động kinh doanh thông minh. |
| CNTT | Công nghệ thông tin |  |
| ERP | Enterprise Resource Planning | Hệ thống quản trị nguồn lực doanh nghiệp |
| OLAP | On-Line Analytical Processing | Xử lý phân tích trực tuyến |
| SQL | Structured Query Language | Ngôn ngữ truy vấn |
| KPI | Key Performance Indicator | Chỉ số đánh giá hiệu quả công việc |
| ACID | Atomicity, Consistency, Isolation, Durability | Bốn thuộc tính quan trọng của hệ quản trị cơ sở dữ liệu |

*Bảng 1.1. Danh mục các từ viết tắt*

LỜI MỞ ĐẦU

1. **Lý do chọn đề tài:**

Trong thời đại hiện nay, dữ liệu đang dần trở thành nguồn tài nguyên vô tận để các cá nhân, tổ chức, doanh nghiệp tận dụng khai thác ở mọi lĩnh vực. Nguồn tài nguyên này ngày càng trở nên quan trọng. Đối với các doanh nghiệp, việc khai thác tốt dữ liệu sẽ giúp các nhà quản trị đưa ra những quyết định mang tính chiến lược, đột phá nhằm mở rộng quy mô và nâng cao lợi nhuận trong kinh doanh.

Khi một doanh nghiệp đi vào hoạt động, những nhà quản lý doanh nghiệp sẽ phải đặt các câu hỏi để khai thác những vấn đề đang gặp phải và tình hình khác trong kinh doanh, tốc độ tăng trưởng, lượng giao dịch hàng ngày, hàng tháng, hàng quý, hàng năm, so sánh giữa năm này, năm khác, hoặc phân khúc các khách hàng của doanh nghiệp, hoặc phân tích doanh thu.

Đối với mỗi doanh nghiệp, họ sẽ tự xây dựng một hệ thống quản lý giao dịch hay chính là các ứng dụng, phần mềm, hệ thống vận hành hàng ngày của doanh nghiệp. Tuy nhiên các hệ thống này chỉ được thiết kế cho việc nhập dữ liệu hàng ngày hoặc để vận hành hệ thống. Chúng cũng có khả năng cho phép lấy dữ liệu cho một số báo cáo đơn giản. Tuy nhiên đối với những yêu cầu báo cáo theo nhiều chiều xoay quanh việc: bán hàng, nhập hàng, tồn kho và sản xuất đòi hỏi phải tính toán phức tạp thì hầu như các hệ thống này rất khó thực hiện.

Qua một khoảng thời gian, các công ty dần có những yêu cầu cao hơn và những mong muốn lớn hơn. Trích xuất báo cáo vào những khoảng thời gian nhất định mà không gây sức ép đến cơ sở dữ liệu, một giải pháp được đưa ra đó là nên xây dựng một hệ thống quản trị cơ sở dữ liệu riêng, nhằm mục đích lưu trữ cũng như phục vụ vào mục đích làm các báo cáo khó có nhiều chiều phân tích.

Trước vấn đề trên, cá nhân em nhận thấy được sự cần thiết đối với doanh nghiệp trong việc thiết kế, xây dựng hệ thống quản trị cơ sở dữ liệu và từ đó tạo ra báo cáo theo yêu cầu của doanh nghiệp để từ đó có thể đáp ứng được nhu cầu tăng cao đối với doanh nghiệp. Vì vậy để mang đến một cái nhìn rõ hơn, em đã quyết định thực hiện bài báo cáo tốt nghiệp với đề tài: “XÂY DỰNG HỆ THỐNG QUẢN TRỊ CƠ SỞ DỮ LIỆU BÁN HÀNG.”

1. **Mục tiêu của đề tài:**

* *Mục tiêu nghiên cứu lý thuyết:*
  + Hệ thống hóa các lý thuyết về việc xây dựng và thiết kế hệ thống quản trị cơ sở dữ liệu nhằm đưa ra phương pháp giúp doanh nghiệp hoàn thiện, triển khai tổ chức dữ liệu theo cách hiệu quả tốt nhất để phục vụ cho hoạt động kinh doanh.
  + Nắm được quy trình xây dựng, thiết kế hệ thống quản trị cơ sở dữ liệu thông qua những mong muốn của khách hàng và các công cụ hỗ trợ quá trình xử lý dữ liệu.
* *Mục tiêu nghiên cứu ứng dụng:*
  + Xây dựng hệ thống quản trị cơ sở dữ liệu, xây dựng luồng dữ liệu phục vụ cho mục đích phân tích và đưa ra báo cáo.
  + Xây dựng báo cáo thông qua việc phân tích hệ thống quản trị cơ sở dữ liệu.

1. **Phương pháp nghiên cứu:**

* Phương pháp nghiên cứu lý thuyết:
  + Tìm hiểu lý thuyết về quy trình phân tích xử lý luồng dữ liệu.
  + Xây dựng và thiết kế hệ thống quản trị cơ sở dữ liệu, trực quan hóa dữ liệu.
* Phương pháp nghiên cứu thực nghiệm:
  + Tham gia với vai trò chuyên viên phân tích dữ liệu trong dự án nhằm phân tích các chỉ số đánh giá về đối tượng khách hàng sử dụng dịch vụ, phân tích về doanh số bán được, phân tích biến động theo thời gian.

1. **Đối tượng và phạm vi nghiên cứu:**

* Đối tượng nghiên cứu đề tài: Kho dữ liệu của doanh nghiệp bán hàng.
* Phạm vi đề tài: Nghiên cứu được thực hiện trong vòng 4 tháng kể từ 01/12/2023 – 01/04/2024, thực hiện phân tích và xử lý dữ liệu dựa trên yêu cầu khách hàng.

1. **Nội dung của đề tài:**

Đề tài được tổ chức gồm phần mở đầu, 4 chương nội dung và phần kết luận...

* Mở đầu
* **Chương 1**: Tổng quan về đơn vị thực tập.
* **Chương 2**: Cơ sở lý thuyết
* **Chương 3**: Bài toán phân tích
* **Chương 4**: Triển khai thực nghiệm
* Kết luận và hướng phát triển

# TỔNG QUAN VỀ ĐƠN VỊ THỰC TẬP

## Giới thiệu về công ty:

### Giới thiệu chung:



*Hình 1.1. Logo FPT Software*

FPT Software Đà Nẵng được thành lập vào ngày 13/8/2005 dưới sự chứng kiến của Chủ tịch FPT Trương Gia Bình và Tổng giám đốc FPT Software đầu tiên Nguyễn Thành Nam cùng lãnh đạo thành phố… Trụ sở đầu tiên đặt tại số 51 đường Nguyễn Văn Linh, đánh dấu sự hiện diện của nhà Phần mềm tại mảnh đất miền Trung.

Từ một văn phòng sản xuất phần mềm với quy mô 50 người, đến hết 2017, công ty đã phát triển lên quy mô hơn 2.700 người, mang lại việc làm cho hàng nghìn tri thức trẻ xuất thân miền Trung. Hàng năm, công ty đưa hàng trăm lượt nhân viên sang làm việc tại các nước Nhật Bản, Mỹ, Singapore… Trong 3 năm gần đây, FPT Software Đà Nẵng là công ty có tốc độ tăng trưởng trung bình hơn 70%/năm. Đây là tốc độ tăng trưởng cao nhất Tập đoàn, sự tăng trưởng làm nên cơ hội lớn cho mọi cá nhân đang tham gia vào sự phát triển này.

Hiện tại, FPT Software Đà Nẵng đang có 2 trụ sở làm việc là Tòa nhà FPT An Đồn, P.An Hải Bắc, Q.Sơn Trà và Tòa nhà FPT Complex, đường Nam Kỳ Khởi Nghĩa, P.Hòa Hải, Q.Ngũ Hành Sơn. Nhân viên FPT Software Đà Nẵng tự hào khi làm việc trong khuôn viên xanh, tiện nghi với cảnh quan đẹp mắt, đáp ứng nhu cầu sinh hoạt, thể thao và giải trí… Không có gì ngạc nhiên khi đời sống tinh thần ở đây vô cùng phong phú, với các giải đá bóng nội bộ ngay trên sân nhà.

### Tầm nhìn:

FPT Software hướng đến mục tiêu trở thành công ty có doanh thu tỷ đô và là nhà cung cấp dịch vụ CNTT đẳng cấp thế giới vào năm 2024.

### Sứ mệnh:

Với sứ mệnh “Mang trí tuệ Việt Nam ra toàn cầu”, toàn thể ban lãnh đạo và nhân viên FPT Software cùng đặt ra mục tiêu lớn: đến năm 2020 sẽ đạt 1 tỷ USD doanh thu, 30.000 người. Đây là một mục tiêu đầy thách thức nhưng mang lại cơ hội không giới hạn cho mỗi thành viên của FPT Software và các nhân viên tiềm năng tương lai. Nói đến “1B Mission”, Chủ tịch FPT Software từng khẳng định: “Nhìn lại lịch sử phát triển 15 năm qua, tuy có lúc thịnh lúc suy nhưng FPT Software luôn luôn tăng trưởng trên 30%, do đó chúng ta có thể tin tưởng mục tiêu này là hiện thực”. Đạt được mục tiêu này, FPT Software sẽ trở thành công ty phần mềm toàn cầu thuộc top 10 thế giới.

### Lĩnh vực hoạt động chính của công ty:

FPT Software là công ty chuyên về lĩnh vực gia công phần mềm , chính vì thế nên công ty cung cấp tất cả dịch vụ liên quan như:

* Xuất khẩu phần mềm
* Giải pháp phần mềm
* Phân phối các sản phẩm công nghệ thông tin
* Tích hợp hệ thống
* Cung cấp các giải pháp ,dịch vụ viễn thông và internet
* Triển khai và Tư vấn dịch vụ ERP
* Phát triển hệ thống nhúng
* Các dịch vụ khác

# CƠ SỞ LÝ THUYẾT

## Lý thuyết cơ bản phân tích dữ liệu:

### Phân tích dữ liệu là gì?

Phân tích dữ liệu là quá trình đánh giá dữ liệu bằng các công cụ phân tích hoặc thống kê để khám phá thông tin hữu ích.

Khi dữ liệu được thu thập và sắp xếp bằng các công cụ và các phương pháp, kết quả sẽ được diễn giải để đưa ra quyết định. Kết quả cuối cùng có thể được phân phối dưới dạng tóm tắt hoặc dưới dạng trực quan như biểu đồ hoặc đồ thị.

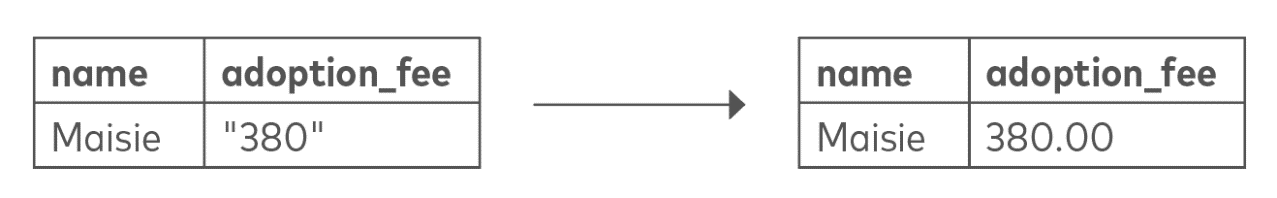
* + 1. ***Tại sao cần phân tích dữ liệu?***
* Giúp cải thiện ra quyết định kinh doanh dựa trên sự hiểu biết. Điều này giúp doanh nghiệp xác định những sản phẩm hoặc dự án đang hoạt động tốt và tập trung đầu tư hơn.
* Giúp doanh nghiệp theo dõi và hiểu rõ thị trường. Điều này giúp doanh nghiệp nhận biết những sở thích, xu hướng và hành vi khách hàng.
* Tối ưu hoạt động kinh doanh bằng cách phân tích dữ liệu và hiểu được cách hoạt động của doanh nghiệp. Qua đó tìm cách để cải thiện hiệu suất, giảm thiểu chi phí và tăng cường lợi nhuận.
  + 1. ***Mục tiêu của phân tích dữ liệu:***
* **Hiểu rõ hơn về dữ liệu.**
* **Phát hiện các mô hình, xu hướng và thông tin hữu ích.**
* **Đưa ra quyết định hoặc dự đoán tương lai.**
  1. **Tìm hiểu về dữ liệu:**
     1. ***Khái niệm về dữ liệu:***

Dữ liệu là tổng hợp các sự thật bao gồm: số liệu, ảnh, video, từ ngữ,... Dữ liệu có thể được tạo ra hằng ngày thông qua những hoạt động, hành vi.

* + 1. ***Tại sao lại cần dữ liệu:***
* Định hướng quyết định cho doanh nghiệp bằng cách phân tích và tận dụng dữ liệu.
* Không chỉ trong kinh doanh, thế giới thay đổi từng ngày và dữ liệu là công cụ mạnh mẽ giúp doanh nghiệp duy trì sự cơ động.
* Thông qua dữ liệu, họ có thể xác định các bước quy trình không hiệu quả, tìm kiếm cách cải thiện chất lượng sản phẩm hoặc dịch vụ, và quản lý tài nguyên nhân lực một cách hiệu quả hơn.
  1. **Data Transformation:** 
     1. ***Khái niệm:***

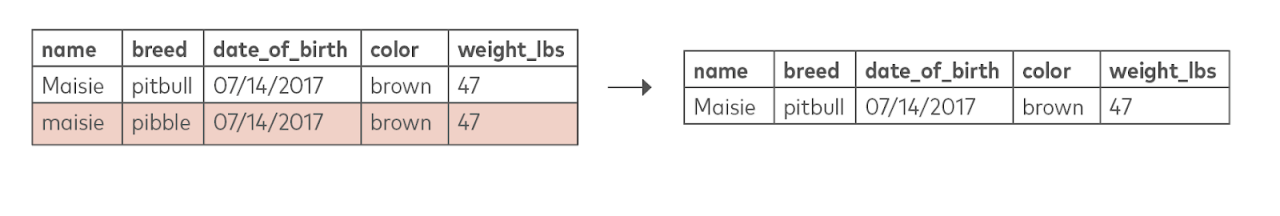
Chuyển đổi dữ liệu (Data Transformation*)* là quá trình sửa đổi, tính toán, phân tách và kết hợp dữ liệu thô thành các mô hình dữ liệu sẵn sàng phân tích. Các doanh nghiệp cần KPI và các thước đo khác để định lượng và hiểu được họ đang làm gì và như thế nào.

* + 1. ***Kỹ thuật:***
* *Chuyển đổi định dạng*: chuyển đổi các ký tự, định dạng ngày tháng và thay đổi kiểu dữ liệu.



*Hình 2.1. Chuyển đổi định dạng*

* *Deduplication:* Xác định và loại bỏ các bản ghi trùng lặp.



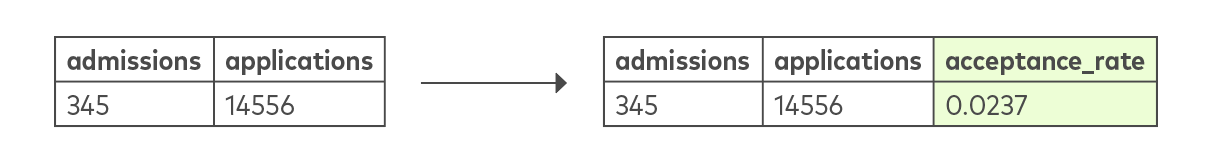
*Hình 2.2. Deduplication*

* *Loại bỏ các cột không sử dụng và lặp lại:* chọn các trường được sử dụng làm tính năng. Điều này giúp cải thiện hiệu suất và tính dễ đọc tổng thể của mô hình.



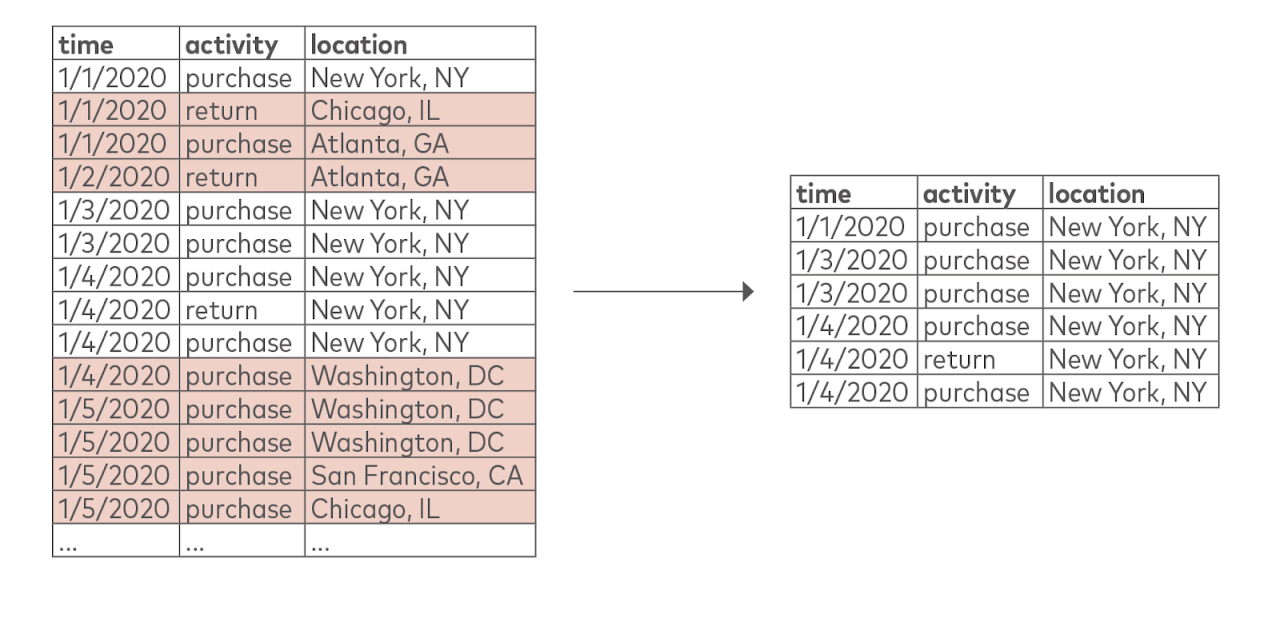
*Hình 2.3. Loại bỏ các cột không sử dụng và lặp lại*

* *Derivation:* bao gồm những phép tính cột chéo đơn giản:



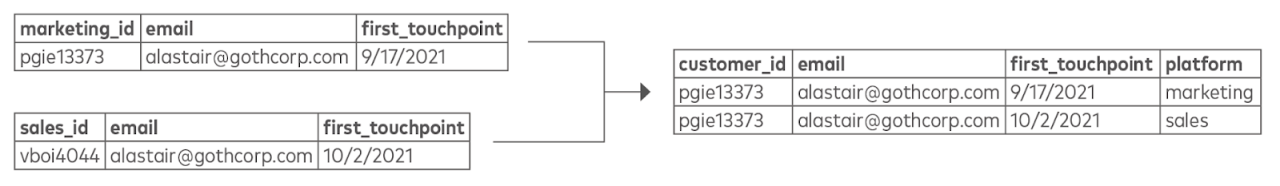
*Hình 2.4. Derivation*

* *Lọc:* loại trừ những dữ liệu trên cơ sở các giá trị hàng hoặc cột nhất định:



*Hình 2.5: Lọc*

* *Hợp nhất:* hay còn gọi là kết hợp, kỹ thuật này sẽ kết hợp các bản ghi từ nhiều bảng.



*Hình 2.6: Hợp nhất*

[1]

* 1. **Mô hình hoá dữ liệu (Data Modeling):**
     1. ***Khái niệm:***

Là quá trình tạo ra một trực quan của toàn bộ hệ thống thông tin hoặc các bộ phận của nó để truyền đạt những kết nối giữa các điểm dữ liệu và cấu trúc.

* + 1. ***Mục đích:***

Nhằm minh họa các loại dữ liệu được lưu trữ trong hệ thống, mối quan hệ giữa các loại dữ liệu đó, cách những dữ liệu đó được nhóm lại cũng như là tổ chức, định dạng và thuộc tính của nó.

* + 1. ***Mô hình dữ liệu (Data Model):***
       1. ***Khái niệm:***

Là những trình bày trực quan của các yếu tố dữ liệu của doanh nghiệp và sự kết nối giữa chúng. Bằng cách định nghĩa và cấu trúc dữ liệu trong bối cảnh của những quy trình có liên quan của doanh nghiệp, những mô hình này hỗ trợ sự phát triển của những hệ thống thông tin hiệu quả.

* + - 1. ***Phân loại:***
* Mô hình khái niệm (Conceptual Data Model): Là các mô hình miền và cung cấp một cái nhìn toàn cảnh về những gì hệ thống sẽ chứa, cách thức tổ chức và những quy tắc kinh doanh nào có liên quan và thường được tạo ra như một phần của quá trình thu nhập các yêu cầu ban đầu của dự án.
* Mô hình logic (Logical Data Model): Mô hình này ít trừu tượng hơn và cung cấp chi tiết hơn về những khái niệm và mối quan hệ trong lĩnh vực đang được xem xét. Chúng cho biết các thuộc tính dữ liệu, chẳng hạn như kiểu dữ liệu và độ dài tương ứng của chúng.
* Mô hình vật lý (Physical Data Model): Là một lược đồ về cách dữ liệu sẽ được lưu trữ vật lý trong cơ sở dữ liệu. Chúng sẽ cung cấp một thiết kết hoàn thiện có thể được triển khai dưới dạng cơ sở dữ liệu quan hệ, gồm các bảng, kết hợp với việc minh họa mối quan hệ giữa các thực thể cũng như khóa chính, khóa ngoại sẽ được sử dụng để duy trì các mối quan hệ đó.
  + - 1. ***Kỹ thuật:***
* Mô hình dữ liệu phân cấp (Hierarchical database model): Biểu hiện mối quan hệ đơn giản-chi tiết ở định dạng dạng cây. Mỗi bản ghi sẽ có 1 gốc hoặc cha mẹ kết nối tới một hoặc nhiều bảng con.
* Mô hình dữ liệu quan hệ (Relational Data Model): Được triển khai trong nhiều cơ sở dữ liệu quan hệ khác nhau thường được sử dụng trong điện toán doanh nghiệp. Mô hình dữ liệu này không yêu cầu hiểu biết chi tiết về các thuộc tính vật lý của bộ lưu trữ dữ liệu đang được sử dụng. Các phân đoạn dữ liệu được nối một cách rõ ràng thông qua việc sử dụng các bảng, làm giảm độ phức tạp của cơ sở dữ liệu.
* Cơ sở dữ liệu quan hệ (Relational Database): Thường được sử dụng ngôn ngữ truy vấn có cấu trúc (SQL) để quản lý đữ liệu. Các cơ sở dữ liệu này hoạt động tốt để duy trình tính toàn vẹn của dữ liệu và giảm thiểu sự dư thừa.
* Mô hình dữ liệu quan hệ thực thể (Entity-Relationship): Mô hình sử dụng sơ đồ chình thức để biểu thị mối quan hệ giữa các thực thể trong cơ sở dữ liệu.
* Mô hình chiều dữ liệu (Dimensional Data Models): Chúng được thiết kế để tối ưu hóa tốc độ truy xuất dữ liệu cho mục đích phân tích trong data warehouse. Mô hình này thường được sử dụng trên các hệ thống OLAP.
* Lược đồ hình sao (Star Schema): Dữ liệu sẽ được tổ chức thành các sự kiện (Fact) và kích thước (Dimension), trong đó mỗi sự kiện sẽ được bao quanh bởi các kích thước liên kết theo mô hình giống ngôi sao.
* Lược đồ bông tuyết (Snowflake Schema): Tương tự với lược đồ hình sao nhưng sẽ bao gồm các lớp bổ sung có kích thước liên quan, làm mô hình phân nhánh phức tạp hơn. [2]
  1. **Trực quan hóa dữ liệu (Data Visualization):**
     1. ***Khái niệm:***

Là sự thể hiện bằng hình ảnh của thông tin và dữ liệu bằng cách sử

dụng các thành phần trực quan như: biểu đồ, đồ thị, bản đồ,... những công cụ trực quan hóa dữ liệu sẽ cung cấp một cách dễ dàng để thấy và hiểu được những xu hướng, ngoại lai, các mẫu trong dữ liệu.

* + 1. ***Ưu điểm:***
* Việc trực quan hóa dữ liệu sẽ giúp ích cho việc đưa ra quyết định về chiến lược.
* Các bên có liên quan và ban lãnh đạo cấp cao nhất sẽ sử dụng khả năng trực quan hóa dữ liệu để diễn giải dữ liệu sao cho có nghĩa.
* Giúp tiết kiệm thời gian thông qua việc phân tích dữ liệu nhanh hơn và khả năng trực quan hóa góc nhìn toàn cảnh hơn.
  + 1. ***Nhược điểm:***

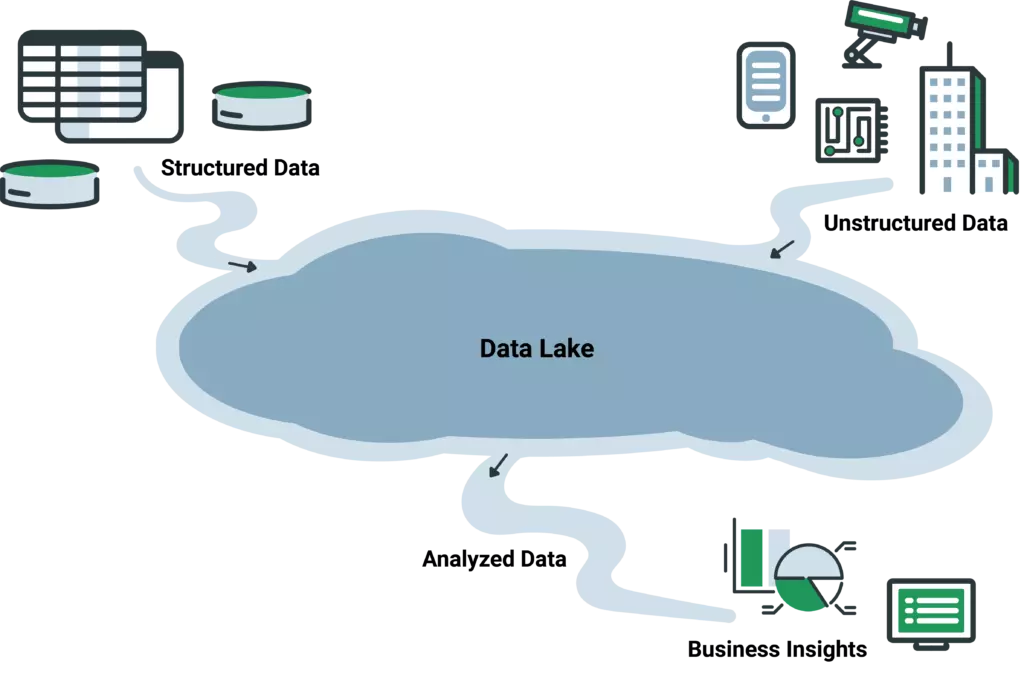
Bên cạnh những ưu điểm, việc trực quan hóa dữ liệu cũng có những nhược điểm. Đôi khi việc người dùng có thể hiểu sai về dữ liệu, người dùng không cẩn thận trong việc xây dựng góc nhìn. Điều này sẽ đưa ra những kết luận và những sự trực quan về dữ liệu một cách sai lệch.

* + 1. ***Lợi ích:***
* Cung cấp cho người đọc nhiều góc nhìn nhằm nhanh chóng tiếp thu thông tin, cải thiện hiểu biết sâu sắc và đưa ra quyết định nhanh hơn.
* Giúp người dùng hiểu biết hơn về mặt kinh doanh của tổ chức để từ đó giúp phát hiện và cải thiện những sai sót.
* Giúp tăng khả năng thu hút và duy trì sự quan tâm của người dùng bằng cách cung cấp thông tin mà họ có thể hiểu được. [3]
  1. **Công cụ Tableau trực quan hóa dữ liệu:** 
     1. ***Khái niệm và các tính năng chính:***
* Khái niệm: Tableau là một công cụ trực quan hóa dữ liệu mạnh mẽ và phát triển nhanh nhất được sử dụng trong ngành Business Intelligence. Công cụ này giúp đơn giản hóa dữ liệu thô ở định dạng rất dễ hiểu, cụ thể là chuyển thành những hình ảnh, biểu đồ trực quan. Tableau không yêu cầu bất kỳ loại kỹ năng lập trình nào để vận hành. Công cụ này thu hút sự quan tâm của mọi người từ mọi lĩnh vực từ kinh doanh, nhà nghiên cứu, các ngành công nghiệp khác nhau.
* Tính năng:
  + Phân tích dữ liệu với BigData.
  + Rà soát, thu nhập, tổng hợp dữ liệu, xây dựng metadata cung cấp cho người dùng nghiệp vụ để chuẩn bị phân tích dữ liệu.
  + Hỗ trợ tạo các truy vấn bằng thao tác đơn giản.
  + Tạo các dashboard tương tác trên biểu đồ, tham biến, tooltip, drill up, drill down, liên kết dashboard, giải thích dữ liệu.
  + Phân tích theo thời gian.
  + Tích hợp được trên cả điện thoại di động với 2 hệ điều hành được sử dụng phổ biến là IOS và Android.
    1. ***Cấu tạo của Tableau:***
* Tableau Prep: Là công cụ dùng để chuẩn bị dữ liệu. Tất cả dữ liệu đầu vào đề được hiển thị và thay đổi tại đây. Giúp người dùng doanh nghiệp và nhà phân tích định hình dữ liệu nhanh chóng.
* Tableau Desktop: Là công cụ trực tiếp phân tích dữ liệu, cung cấp một giao diện khá trực quan cùng cá tính năng đa dạng để mã hóa và phân tích dữ liệu.
* Tableau Online: Có thể xem Tableau Online như một máy chủ miễn phí, không giới hạn lưu trữ, không cần đến máy chủ nhưng vẫn có thể cho phép liên kết hơn 40 nguồn dữ liệu khác nhau.
* Tableau Server: Là nơi chia sẻ các báo cáo phân tích của doanh nghiệp, tuy nhiên những dữ liệu trên đây đã được bảo mật cẩn thận và cấp quyền truy cập.
  + 1. ***So sánh Tableau với Microsoft Power BI:***

|  |  |  |
| --- | --- | --- |
|  | **Power BI** | **Tableau** |
| Hệ sinh thái | Có chức năng kéo và thả dễ dàng, với các tính năng cho phép sao chép tất cả những định dạng trên các biểu đồ trực quan hóa tương tự. | Có tính năng kéo thả, cho phép người dùng tạo visual tương tác nhanh chóng, cũng có thể xây dựng dashboard tương tác chỉ với một vài cái nhấp chuột. |
| Tính dễ sử dụng | Có lợi thế vì dựa trên giao diện quen thuộc với người dùng từ Microsoft Office 365. | Cung cấp một số lợi thế cần thiết để khám phá và trực quan hóa dữ liệu một cách chi tiết. |
| Chi phí | Phiên bản Power BI Professional có giá dưới $10 mỗi tháng cho mỗi người dùng. | Phiên bản Pro của Tableau có giá hơn $35 mỗi tháng cho mỗi người dùng. |
| Hiệu suất | Hoạt động nhanh hơn và tốt hơn với khối lượng dữ liệu không quá lớn. | Xử lý khối lượng dữ liệu lớn một cách nhanh chóng. |
| Giao diện người dùng | Khá trực quan và có thể dễ dàng tích hợp với các sản phẩm của Microsoft. Giao diện dễ học và dễ hiểu. | Giao diện thông minh cho phép tạo và tùy chỉnh dashboard theo yêu cầu của bạn. |
| Nguồn dữ liệu | Hỗ trợ nhiều nguồn dữ liệu khác nhau nhưng có quyền truy cập hạn chế vào các cơ sở dữ liệu và máy chủ. | Có quyền truy cập vào nhiều nguồn dữ liệu và máy chủ. |
| Trực quan hóa dữ liệu | Cung cấp chức năng kéo và thả dễ sử dụng, các tính năng làm cho dữ liệu hấp dẫn trực quan. Cung cấp một loạt các hình ảnh trực quan chi tiết và hấp dẫn để tạo báo cáo và dashboard. | Cho phép người dùng tùy chỉnh dashboard cụ thể cho một thiết bị. Cung cấp các trực quan tương tác hỗ trợ đưa ra những góc nhìn một cách nhanh chóng. Có thể dịch truy vấn thành trực quan hóa. |

*Bảng 2.1. So sánh Tableau và Microsoft Power BI*

* + 1. ***Lợi ích khi sử dụng Tableau:***
* Cung cấp một loạt tính năng tích hợp đa dạng, giúp tạo ra các hình ảnh trực quan độc đáo, ấn tượng.
* Sử dụng câu truy vấn đề khám phá bộ dữ liệu từ nhiều góc độ khác nhau, từ đó hiển thị chúng dưới dạng biểu đồ.
* Được thiết kế vưới giao diện thân thiện với người dùng, không yêu cầu có kiến thức kỹ thuật hoặc lập trình.
* Có thể làm việc với đa dạng các nguồn dữ liệu để tạo ra hình ảnh trực quan và thông tin giá trị.
  1. **Hồ dữ liệu (Data Lake):**



*Hình 2.7. Hồ dữ liệu (Data Lake)*

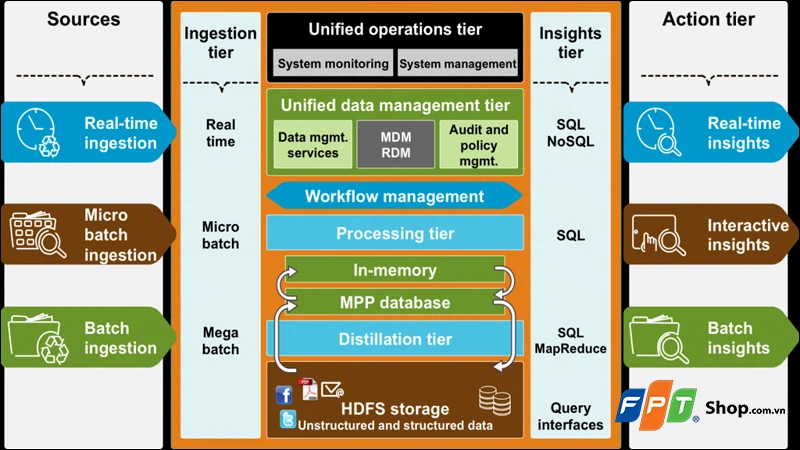
* + 1. ***Khái niệm:***

Là một nơi lưu trữ tập trung được thiết kế để chứa, xử lý và bảo mật lượng lớn dữ liệu đa dạng bao gồm dữ liệu có cấu trúc, bán cấu trúc và phi cấu trúc.

Là nơi cung cấp khả năng lưu trữ lượng lớn dữ liệu, nâng cao hiệu suất phân tích và tích hợp dữ liệu gốc.

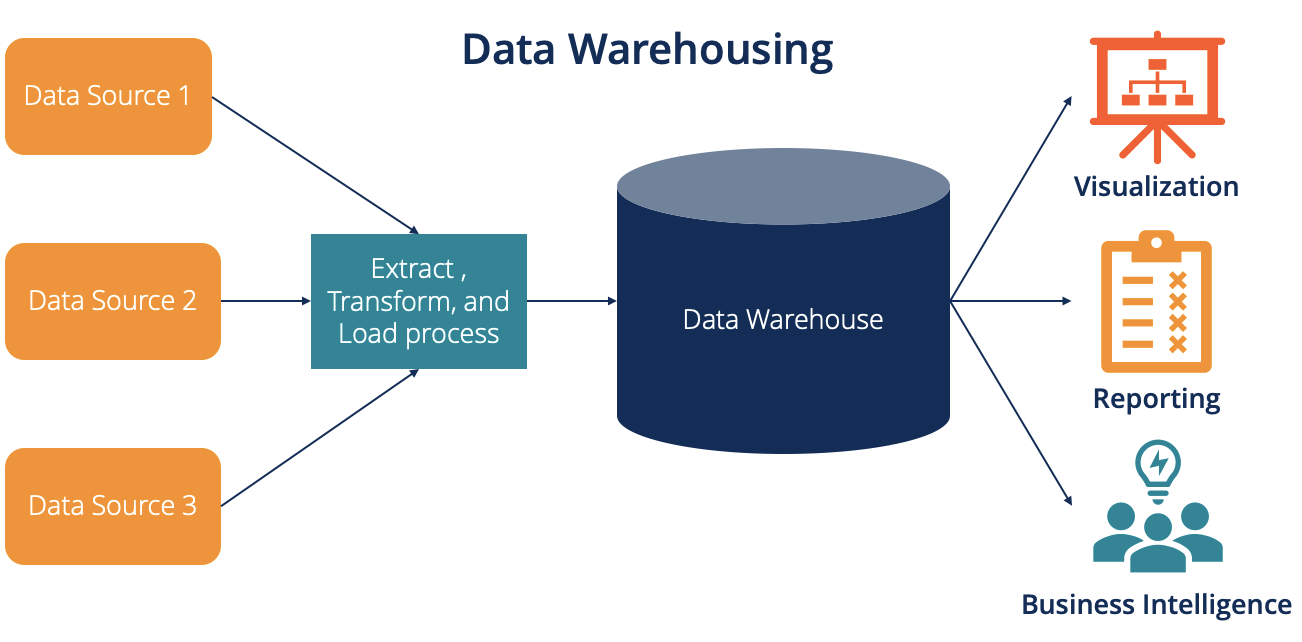
Data Lake cung cấp một nền tảng mở rộng và an toàn cho phép doanh nghiẹp nhập dữ liệu từ mọi nguòn bao gồm cả những dữ liệu từ các hệ thống vật lý, đám mấy hoặc edge computing.

* + 1. ***Lợi ích:***
* Cung cấp khả năng cho người dùng cộng tác và phân tích dữ liệu theo nhiều phương pháp khác nhau.
* Giúp tăng tốc quá trình ra quyết định với độ chính xác cao hơn.
* Hỗ trợ việc lưu trữ và thực hiện phân tích trên dữ liệu trên Internet dễ dàng hơn.
  + 1. ***Kiến trúc hồ dữ liệu:***



*Hình 2.8. Kiến trúc hồ dữ liệu (Data Lake Architecture)*

* Tầng Ingestion: Là tầng mô tả việc thu nhập dữ liệu, các dữ liệu có thể được thu nhập vào Data Lake theo lô hoặc theo thời gian thực.
* Tầng Insights: Đại diện cho phần nghiên cứu, nơi các thông tin chi tiết được sử dụng. Các truy vấn SQL, NoSQL, thậm chí là Excel có thể được sử dụng để phân tích dữ liệu.
* HDFS: Là nơi lưu trữ cho tất cả dữ liệu trong hệ thống, bao gồm cả dữ liệu có cấu trúc và phi cấu trúc.
* Tầng Distillation: Lấy dữ liệu từ tầng lưu trữ và chuyển đổi thành dữ liệu có cấu trúc để việc phân tích trở nên dễ dàng hơn.
* Tầng Processing: Chạy các thuật toán phân tích và xử lý các truy vấn từ người dùng theo nhiều thời gian khác nhau, tương tác, hoặc hàng loạt để tạo dữ liệu có cấu trúc để phân tích dễ dàng hơn.
* Tầng Unified Operations: Quản lý và giám sát hệ thống, bao gồm kiểm toán, quản lý thành thạo, quản lý dữ liệu, quản lý quy trình làm việc.
  1. **Kho dữ liệu (Data Warehouse):**



*Hình 2.9. Kho dữ liệu (Data Warehouse)*

* + 1. ***Khái niệm:***

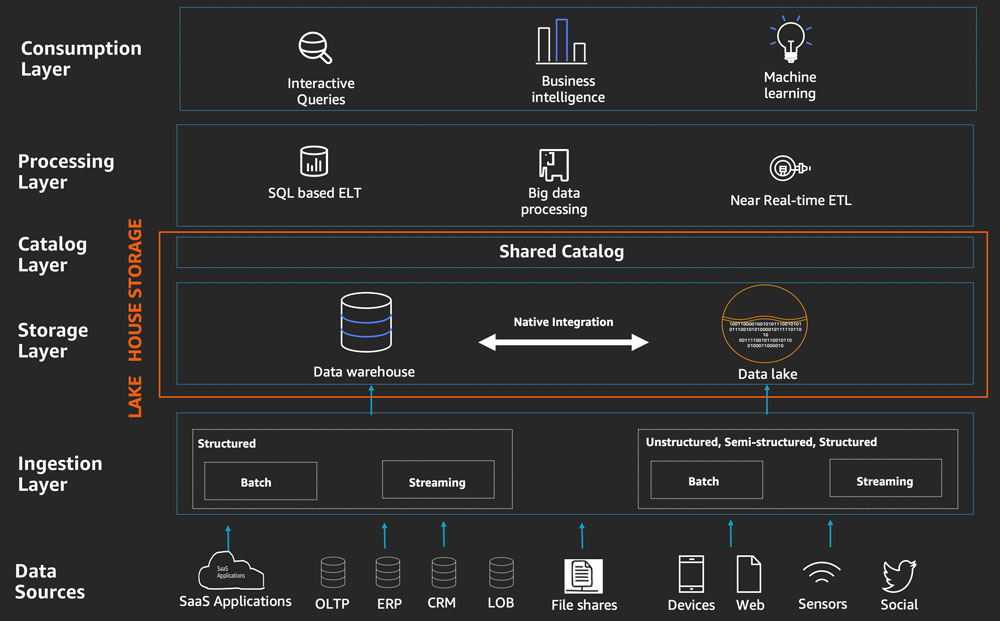
Là một loại hệ thống quản lý lưu trữ dữ liệu được thiết kế để kích hạt và hỗ trợ các hoạt động kinh doanh thông minh (BI), đặc biệt là phân tích. Kho dữ liệu chỉ nhằm mục đích là thực hiện các truy vấn và phân tích tích, thướng chứa một lượng lớn dữ liệu lịch sử.

Dữ liệu trong kho dữ liệu thường được lấy từ nhiều nguồn khác nhau như tệp nhật ký ứng dụng, ứng dụng giao dịch.

* + 1. ***Các đặc tính của kho dữ liệu:***
* Hướng chủ đề (Subject-Orient): Cung cấp thông tin cho một chủ thể cụ thể thay vì các hoạt động liên tục của toàn tổ chức. Các chủ đề đó có thể là bán hàng, khuyến mãi, hàng tồn kho,...
* Được tích hợp (Integrated): Kho dữ liệu được phát triển bằng cách tích hợp dữ liệu từ nhiều nguồn khác nhau thành một định dạng nhất quán. Dữ liệu phải được lưu trữ trong kho dữ liệu một cách nhất quán và được mọi người chấp nhận về cách đặt tên, định dạng và mã hóa.
* Bất biến (Non-volatile): Dữ liệu một khi đã được nhập vào kho dữ liệu phải không thay đổi. Dữ liệu trước đó không bị xóa khi nhập dữ liệu hiện tại.
* Có gán nhãn thời gian (Time-Variant): Dữ liệu được lưu trữ trong kho dữ liệu cung cấp thông tin từ mọt thời điểm lịch sử cụ thể, do đó, dữ liệu được phân loại với một khung thời gian cụ thể.
  + 1. ***Kiến trúc kho dữ liệu:***
* Một hệ thống kho dữ liệu có hai kiến trúc chính: kiến trúc luồng dữ liệu là về cách sắp xếp các kho lưu trữ dữ liệu trong kho dữ liệu và cách truyền dữ liệu từ hệ thống nguồn đến người dùng thông qua các kho lưu trữ dữ liệu và kiến trúc hệ thống là về cấu hình vật lý của máy chủ, mạng, phần mềm, bộ lưu trữ và máy khách.
* Dựa trên khả năng truy cập của người dùng, bạn có thể phân loại kho lưu trữ dữ liệu [kho dữ liệu](https://cafedevcode.com/gioi-thieu-tong-quan-ve-data-warehouse-kho-du-lieu-49.html) thành ba loại:
* Kho lưu trữ dữ liệu hướng tới người dùng: là kho lưu trữ dữ liệu phù hợp cho người dùng cuối và được truy vấn bởi người dùng cuối và các ứng dụng của người dùng cuối.
* Kho lưu trữ dữ liệu nội bộ là kho lưu trữ dữ liệu được sử dụng nội bộ bởi các thành phần [kho dữ liệu](https://cafedevcode.com/gioi-thieu-tong-quan-ve-data-warehouse-kho-du-lieu-49.html) nhằm mục đích tích hợp, làm sạch, ghi nhật ký và chuẩn bị dữ liệu và nó không mở để truy vấn bởi người dùng cuối và ứng dụng của người dùng cuối.
* Kho lưu trữ dữ liệu kết hợp được sử dụng cho cả cơ chế [kho dữ liệu](https://cafedevcode.com/gioi-thieu-tong-quan-ve-data-warehouse-kho-du-lieu-49.html) nội bộ và cho truy vấn của người dùng cuối và ứng dụng của người dung cuối.
* Dựa trên định dạng dữ liệu, có thể phân loại kho lưu trữ dữ liệu thành bốn loại:
  + Stage là kho lưu trữ dữ liệu nội bộ được sử dụng để chuyển đổi và chuẩn bị dữ liệu thu được từ hệ thống nguồn, trước khi dữ liệu được tải đến các kho lưu trữ dữ liệu khác trong [kho dữ liệu](https://cafedevcode.com/gioi-thieu-tong-quan-ve-data-warehouse-kho-du-lieu-49.html).
  + Kho lưu trữ dữ liệu được chuẩn hóa (Normalized Data Store - NDS) là kho lưu trữ dữ liệu chính bên trong dưới dạng một hoặc nhiều cơ sở dữ liệu quan hệ được chuẩn hóa nhằm mục đích tích hợp dữ liệu từ các hệ thống nguồn khác nhau được thu thập trong một giai đoạn, trước khi dữ liệu được tải vào kho lưu trữ dữ liệu dành cho người dùng
  + Kho lưu trữ dữ liệu vận hành (Operational Data Store - ODS) là một kho lưu trữ dữ liệu kết hợp ở dạng một hoặc nhiều cơ sở dữ liệu quan hệ được chuẩn hóa, chứa dữ liệu giao dịch và phiên bản mới nhất của dữ liệu chủ, nhằm mục đích hỗ trợ các ứng dụng vận hành.
  + Kho lưu trữ dữ liệu thứ nguyên (Dimensional Data Store - DDS) là kho lưu trữ dữ liệu hướng tới người dùng, ở dạng một hoặc nhiều cơ sở dữ liệu quan hệ, trong đó dữ liệu được sắp xếp ở định dạng thứ nguyên nhằm mục đích hỗ trợ các truy vấn phân tích.[5]
  1. **Data Lakehouse:**
     1. ***Khái niệm***:

Data Lakehouse là một kiến trúc quản lý dữ liệu mở, mới, kết hợp tính linh hoạt, hiệu quả về chi phí và quy mô của các hồ dữ liệu (Data Lake) với việc quản lý dữ liệu và giao dịch ACID của kho dữ liệu (Data Warehouse), cho phép kích hoạt business intelligence (BI) và machine learning (ML) trên tất cả dữ liệu.

* + 1. ***Kiến trúc của Data Lakehouse***



*Hình 2.10. Data Lakehouse Architecture*

* Nguồn dữ liệu (Data Sources): Cho phép người dùng nhập và phân tích dữ liệu từ nhiều nguồn khác nhau chẳng hạn như ứng dụng dòng doanh nghiệp (LOB), ứng dụng ERP và ứng dụng CRM,... Ngoài các nguồn nội bộ, bạn có thể nhận dữ liệu từ các nguồn hiện đại như ứng dụng web, phương tiện truyền thông xã hội,... Các nguồn hiện đại này thường tạo ra dữ liệu bán cấu trúc và phi cấu trúc, thường là các luồng liên tục.
* Tầng nhập dữ liệu (Data Ingestion Layer): Ingestion layer trong kiến trúc Lakehouse chịu trách nhiệm nhập dữ liệu vào layer lưu trữ Lakehouse. Nó cung cấp khả năng kết nối với các nguồn dữ liệu bên trong và bên ngoài qua nhiều giao thức khác nhau. Nó có thể nhập và cung cấp dữ liệu phát trực tuyến hàng loạt cũng như thời gian thực vào Data Warehouse cũng như các thành phần Data Lake của tầng lưu trữ Lakehouse.
* Tầng lưu trữ dữ liệu (Data Storage Layer): chịu trách nhiệm cung cấp các thành phần bền bỉ, có thể mở rộng và hiệu quả chi phí cao để lưu trữ và quản lý số lượng lớn dữ liệu. Trong Kiến trúc Lakehouse, Data Warehouse và Data Lake được tích hợp nguyên bản để cung cấp một tầng lưu trữ tích hợp hiệu quả về chi phí hỗ trợ dữ liệu không có cấu trúc cũng như dữ liệu có cấu trúc và mô hình cao. Tầng lưu trữ có thể lưu trữ dữ liệu ở các trạng thái sẵn sàng khác nhau.
* Tầng danh mục (Catalog Layer): chịu trách nhiệm lưu trữ metadata kinh doanh và kỹ thuật về các tập dữ liệu được lưu trữ trong layer lưu trữ Lakehouse. Trong Kiến trúc Lakehouse, danh mục được chia sẻ bởi cả Data Lake và Data Storage, đồng thời cho phép viết các truy vấn kết hợp dữ liệu được lưu trữ trong Data Lake cũng như Data Warehouse trong cùng một SQL.
* Tầng xử lý dữ liệu (Data Processing Layer): Các thành phần trong lớp xử lý dữ liệu của Kiến trúc Lakehouse chịu trách nhiệm chuyển đổi dữ liệu sang trạng thái có thể tiêu thụ được thông qua xác thực, dọn dẹp, chuẩn hóa, chuyển đổi dữ liệu.
* Lớp tiêu thụ dữ liệu (Data Consumption Layer): Chịu trách nhiệm cung cấp các thành phần có thể mở rộng và hiệu suất sử dụng giao diện Lakehouse thống nhất để truy cập vào tất cả dữ liệu được lưu trữ trong bộ lưu trữ Lakehouse và tất cả metadata được lưu trữ trong danh mục Lakehouse.

# BÀI TOÁN PHÂN TÍCH XÂY DỰNG HỆ THỐNG QUẢN TRỊ CƠ SỞ DỮ LIỆU BÁN HÀNG

## Khảo sát và phân tích yêu cầu:

Hoạt động của công ty: Xây dựng một hệ thống POS với các đối tác là những cửa hàng bán lẻ. Hiện nay công ty muốn xây dựng một hệ thống quản trị cơ sở dữ liệu nhằm lưu trữ và quản lý dữ liệu về các giao dịch đã thực hiện trên hệ thống POS đã được tải lên hệ thống đám mây (Cloud) của doanh nghiệp. Bên cạnh đó doanh nghiệp mong muốn xây dựng báo cáo dưới dạng dashboard nhằm phục vụ cho mục đích phân tích về khách hàng sử dụng dịch vụ cũng như doanh thu sản phẩm từ những dữ liệu đã thu nhập, xử lý trên hệ thống quản trị cơ sở dữ liệu được xây dựng. Điều này sẽ hỗ trợ cho doanh nghiệp trong việc xác định tình hình kinh doanh hiện tại của doanh nghiệp, từ đó có thể đưa ra những chiến lược kinh doanh phù hợp.

# TRIỂN KHAI THỰC NGHIỆM

4.1. Tổng quan dự án:

4.1.1. Dự án Xây dựng hệ thống quản trị cơ sở dữ liệu bán hàng:

4.1.2. KPI Analysis:

4.1.3. Data Transformation:

4.1.4. Data Modeling:

4.1.5. Data Reporting:

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Đề tài đã thực hiện được …

TÀI LIỆU THAM KHẢO

[1] “Data Tranformation” Fivetran [Online]. Available:

<https://unitrain.edu.vn/data-transformation-chuyen-doi-du-lieu-la-gi/>

[2] “Data Modeling” COLE.VN [Online]. Available:

<https://cole.edu.vn/ky-thuat-data-modeling/>

[3] “Data Visualization” TABLEAU [Online].Available:

<https://www.tableau.com/learn/articles/data-visualization>

[4] “Data Lake” FPT SHOP [Online]. Available:

<https://fptshop.com.vn/tin-tuc/danh-gia/data-lake-la-gi-165101>

[5] “Data Warehouse” BIZFLYCLOUD [Online]. Available:

<https://bizflycloud.vn/tin-tuc/data-warehouse-la-gi-kien-thuc-co-ban-ve-kho-du-lieu-2022051616255507.htm>

PHỤ LỤC