|  |
| --- |
| **HỌC VIỆN NGÂN HÀNG** |
| **------------------------** |
| **ĐỀ TÀI NGHIÊN CỨU KHOA HỌC SINH VIÊN** |
| **DỰ THI CẤP HỌC VIỆN NGÂN HÀNG** |
| **NĂM HỌC 2022 – 2023** |

**An orange circle with white text and a map

Description automatically generated**

|  |
| --- |
|  |

**NGHIÊN CỨU CÔNG NGHỆ ẢO HÓA DỮ LIỆU ỨNG DỤNG RA QUYẾT ĐỊNH TRONG KINH DOANH THÔNG MINH - THỰC NGHIỆM BÀI TOÁN MỞ RỘNG QUY MÔ CỦA CÔNG TY MAVEN MARKET**

|  |
| --- |
| **Hà Nội - 2024** |

|  |
| --- |
| **HỌC VIỆN NGÂN HÀNG** |
| **------------------------** |
| **ĐỀ TÀI NGHIÊN CỨU KHOA HỌC SINH VIÊN** |
| **DỰ THI CẤP HỌC VIỆN NGÂN HÀNG** |
| **NĂM HỌC 2022 – 2023** |

|  |
| --- |
|  |

**NGHIÊN CỨU CÔNG NGHỆ ẢO HÓA DỮ LIỆU ỨNG DỤNG RA QUYẾT ĐỊNH TRONG KINH DOANH THÔNG MINH - THỰC NGHIỆM BÀI TOÁN MỞ RỘNG QUY MÔ CỦA CÔNG TY MAVEN MARKET**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sinh viên thực hiện:** | Nguyễn Tiến Mạnh | K24HTTTA | 24A4040033 |
|  | Trần Yến Vy | K24HTTTB | 24A4041699 |
| **Giảng viên hướng dẫn:** | TS. Chu Thị Hồng Hải | Khoa Công nghệ thông tin và Kinh tế số | |

**Hà Nội - 2024**

**MỤC LỤC**

[MỤC LỤC 3](#_Toc167531344)

[DANH MỤC HÌNH ẢNH 5](#_Toc167531345)

[PHẦN MỞ ĐẦU 6](#_Toc167531346)

[1. Tính cấp thiết của chủ đề nghiên cứu 6](#_Toc167531347)

[2. Mục tiêu nghiên cứu 6](#_Toc167531348)

[3. Phương pháp nghiên cứu 7](#_Toc167531349)

[4. Đối tượng và phạm vi nghiên cứu 7](#_Toc167531350)

[5. Nội dung của đề tài 7](#_Toc167531351)

[CHƯƠNG 1: CƠ SỞ LÝ LUẬN VỀ ẢO HÓA DỮ LIỆU VÀ KINH DOANH THÔNG MINH 8](#_Toc167531352)

[1.1. Tổng quan về ảo hóa dữ liệu 8](#_Toc167531353)

[1.1.1. Khái niệm 8](#_Toc167531354)

[1.1.2. Kiến trúc và hoạt động của ảo hóa dữ liệu 9](#_Toc167531355)

[1.1.3. Lợi ích của ảo hóa dữ liệu 11](#_Toc167531356)

[1.2. Tổng quan về hệ thống kinh doanh thông minh 12](#_Toc167531357)

[1.2.1. Khái niệm 12](#_Toc167531358)

[1.2.2. Các công cụ Kinh doanh thông minh 13](#_Toc167531359)

[1.2.3. Lợi ích của Hệ thống kinh doanh thông minh 15](#_Toc167531360)

[1.2.4. Vai trò của BI trong tổ chức kinh doanh 16](#_Toc167531361)

[1.3. Kinh doanh thông minh với ảo hóa và kho dữ liệu 17](#_Toc167531362)

[1.3.1. Kinh doanh thông với Kho dữ liệu 17](#_Toc167531363)

[1.3.2. Kinh doanh thông minh với ảo hóa dữ liệu 19](#_Toc167531364)

[1.3.3. So sánh kinh doanh thông minh với kho dữ liệu và ảo hóa dữ liệu 21](#_Toc167531365)

[1.1.4. Xu hướng ứng dụng ảo hóa dữ liệu trong doanh nghiệp 22](#_Toc167531366)

[1.4. Tổng kết Chương 1 23](#_Toc167531367)

[CHƯƠNG 2: CÔNG NGHỆ VÀ QUY TRÌNH TRIỂN KHAI ỨNG DỤNG ẢO HÓA DỮ LIỆU TRONG KINH DOANH THÔNG MINH 24](#_Toc167531368)

[2.1. Các công nghệ ảo hóa dữ liệu 24](#_Toc167531369)

[2.1.1. Các công cụ ảo hóa dữ liệu 24](#_Toc167531370)

[2.1.2. Nhận xét các công nghệ ảo hóa dữ liệu 30](#_Toc167531371)

[2.2. Kiến trúc hệ thống ảo hóa dữ liệu 35](#_Toc167531372)

[2.3. Quy trình triển khai ứng dụng ảo hóa hỗ trợ ra quyết định trong kinh doanh thông minh 38](#_Toc167531373)

[2.3. Tổng kết Chương 2 42](#_Toc167531374)

[CHƯƠNG 3: ỨNG DỤNG CÔNG NGHỆ ẢO HÓA DỮ LIỆU TRONG KINH DOANH THÔNG MINH 44](#_Toc167531375)

[3.1. Mô tả bài toán 44](#_Toc167531376)

[3.2. Phân tích bài toán 44](#_Toc167531377)

[3.2.1. Phân tích nguồn dữ liệu 44](#_Toc167531378)

[3.2.2. Xử lý dữ liệu 45](#_Toc167531379)

[3.3.3. Phân tích kết quả đầu ra 46](#_Toc167531380)

[3.3. Lựa chọn nền tảng công nghệ 46](#_Toc167531381)

[3.3.1. Tổng Quan Về Công Nghệ Ảo Hóa Dữ Liệu Denodo 46](#_Toc167531382)

[3.3.2. Vai Trò của Denodo trong Quản Lý Dữ Liệu 47](#_Toc167531383)

[3.3.3. Mối Quan Hệ Giữa Denodo và Các Hệ Thống Quản Lý Dữ Liệu Khác 47](#_Toc167531384)

[3.3.4. Tính phù hợp của công nghệ Denodo trong việc giải quyết bài toán 47](#_Toc167531385)

[3.4. Xây dựng mô hình ảo hóa dữ liệu 49](#_Toc167531386)

[3.4.1. Xác định các nguồn dữ liệu 50](#_Toc167531387)

[3.4.2. Thiết kế các lớp ảo hóa dữ liệu 50](#_Toc167531388)

[3.4.3. Xây dựng các dịch vụ dữ liệu và thiết lập Query Engine 51](#_Toc167531389)

[3.4.4. Data Catalog và xác định người dùng cuối 51](#_Toc167531390)

[3.5. Triển khai và cấu hình 51](#_Toc167531391)

[3.5.1. Kết nối nguồn dữ liệu 51](#_Toc167531392)

[3.5.2. Tiền xử lý dữ liệu 53](#_Toc167531393)

[3.5.3. Kích hoạt Data Catalog 55](#_Toc167531394)

[3.5.4. Truy vấn dữ liệu 56](#_Toc167531395)

[3.5.5. Kết nối Power BI 57](#_Toc167531396)

[3.6. Thử nghiệm và kiểm tra 58](#_Toc167531397)

[3.7. Tổng kết Chương 3 61](#_Toc167531398)

[KẾT LUẬN 63](#_Toc167531399)

[TÀI LIỆU THAM KHẢO 64](#_Toc167531400)

**DANH MỤC HÌNH ẢNH**

[Hình 1: Kiến trúc và hoạt động của ảo hóa dữ liệu 9](#_Toc167479291)

[Hình 2: Đề xuất hệ thống kiến trúc ảo hóa dữ liệu 37](#_Toc167479292)

[Hình 3: Quy trình thực hiện ảo hóa dữ liệu trong kinh doanh thông minh 39](#_Toc167479293)

[Hình 4: Mô hình ảo hóa dữ liệu 50](#_Toc167479294)

[Hình 5: Kết nối nguồn dữ liệu database 52](#_Toc167479295)

[Hình 6: Kết nối nguồn dữ liệu trên đám mây 52](#_Toc167479296)

[Hình 7: Kết nối nguồn dữ liệu web 53](#_Toc167479297)

[Hình 8: Tạo view 53](#_Toc167479298)

[Hình 9: Google Cloud Dataset 54](#_Toc167479299)

[Hình 10: BEA HTTPS Dataset 55](#_Toc167479300)

[Hình 11: Kích hoạt Data Catalog 56](#_Toc167479301)

[Hình 12: Bảng ảo về doanh thu, chi phí và lợi nhuận của Maven Market 57](#_Toc167479302)

[Hình 13: Kết quả số lượng cửa hàng trên các vùng và bang 57](#_Toc167479303)

[Hình 14: Lựa chọn bảng dữ liệu và phân tích 58](#_Toc167479304)

[Hình 15: Tổng thu nhập người dân Mỹ ở các bang 59](#_Toc167479305)

[Hình 16: Tỷ lệ thu nhập người dân và tỷ lệ số lượng cửa hàng theo các bang 59](#_Toc167479306)

[Hình 17: Biểu đồ thể hiện thu nhập và dân số theo các bang 60](#_Toc167479307)

[Hình 18: Biểu đồ cột thể hiện tổng lợi nhuận, chi phí và doanh thu của cửa hàng theo các bang 60](#_Toc167479308)

# PHẦN MỞ ĐẦU

## Tính cấp thiết của chủ đề nghiên cứu

Sự phức tạp và khối lượng lớn dữ liệu được sinh ra từ các ứng dụng và hệ thống hiện đại đã khiến cho việc khai thác và sử dụng dữ liệu hiệu quả trở thành một vấn đề cấp bách đối với các tổ chức kinh doanh. Trong bối cảnh này, các phương pháp tích hợp dữ liệu truyền thống, thường dựa trên quy trình Trích xuất, Chuyển đổi và Tải (ETL), có thể tiêu tốn nhiều thời gian và tài nguyên, đặc biệt là đối với các doanh nghiệp phải xử lý khối lượng lớn dữ liệu nằm rải rác ở nhiều nơi. Gần đây, ảo hóa dữ liệu, một kỹ thuật tiên tiến, cung cấp giải pháp linh hoạt và tiết kiệm chi phí hơn được nhắc tới, nó cho phép các tổ chức tích hợp các nguồn dữ liệu, tạo nhìn thống nhất về các nguồn dữ liệu và truy cập tòan bộ thông tin chi tiết về dữ liệu trong tổ chức kinh doanh một cách nhanh chóng theo thời gian thực. Điều này không chỉ cải thiện hiệu quả hoạt động mà còn giúp các tổ chức nhanh chóng thích ứng với các yêu cầu kinh doanh luôn thay đổi.

Ảo hóa dữ liệu đang được đặc biệt quan tâm vì nó có nhiều ưu điểm trong việc tích hợp dữ liệu hỗ trợ ra quyết định trong kinh doanh thông minh. Tuy nhiên, đến nay nó vẫn còn là một công nghệ mới trong việc hiểu biết về bản chất, ứng dụng và quy trình triển khai công nghệ này trong các tổ chức kinh doanh. Các nghiên cứu hiện tại, chủ yếu tập trung vào các nguyên tắc chung và lợi ích của ảo hóa dữ liệu, nhưng ít đi sâu vào các thách thức và hạn chế khi triển khai thực tế, đặc biệt là trong các lĩnh vực cụ thể như khai phá dữ liệu. Đến nay, tại Việt Nam, tài liệu nghiên cứu về lĩnh vực này còn thiếu, với hầu hết các nghiên cứu thường tập trung vào các vấn đề rộng lớn hơn như quản lý dữ liệu mà không đi sâu vào các kỹ thuật cụ thể của ảo hóa dữ liệu cũng như ứng dụng của nó cho khai phá dữ liệu phục vụ kinh doanh thông minh trong tổ chức kinh doanh.

Do đó, nghiên cứu này hướng tới mục tiêu giải quyết khoảng trống đã phân tích ở trên bằng cách nghiên cứu các thách thức, cơ hội và phương pháp thực nghiệm tốt nhất liên quan đến việc ảo hóa dữ liệu ứng dụng trong kinh doanh thông minh với kỳ vọng sẽ đóng góp vào cơ sở tri thức hiện có, đồng thời tạo dựng một nền tảng vững chắc cho việc triển khai và sử dụng công nghệ ảo hóa dữ liệu một cách hiệu quả trong bối cảnh kinh doanh hiện đại.

## Mục tiêu nghiên cứu

Mục tiêu tổng quát của nghiên cứu này là:

*Nghiên cứu về khái niệm, công cụ, kiến trúc, quy trình trong ảo hóa dữ liệu và ứng dụng vào phân tích dữ liệu hỗ trợ ra quyết định cho kinh doanh thông minh trong các tổ chức kinh doanh.*

Để thực hiện mục tiêu tổng quát trên, đề tài thực hiện các mục tiêu cụ thể sau:

* Nghiên cứu lý thuyết về ảo hóa dữ liệu và kinh doanh thông minh.
* Nghiên cứu các công nghệ ảo hóa đang được sử dụng tại thị trường hiện nay
* Nghiên cứu các công nghệ tích hợp dữ liệu cho kinh doanh thông minh và so sánh các công nghệ
* Đề xuất kiến trúc chung cho công nghệ ảo hóa và quy trình triển khai công nghệ ảo hóa cho tổ chức kinh doanh.
* Thực nghiệm ảo hóa dữ liệu ứng dụng trong kinh doanh thông minh cho một tổ chức kinh doanh cụ thể với bối cảnh về công nghệ và dữ liệu hiện tại của tổ chức kinh doanh mẫ.

## Phương pháp nghiên cứu

Để đạt được mục tiêu đề ra, đề tài sử dụng các phương pháp nghiên cứu sau:

* *Phương pháp nghiên cứu tài liệu, phỏng vấn chuyên gia* để tìm hiểu về công nghệ mới cũng như quá trình triển khai một công nghệ mới vào hoạt động kinh doanh của tổ chức.
* *Phương pháp mô hình hóa* : Sử dụng các biểu đồ, sơ đồ và đồ thị biểu diễn các vấn đề quản lý và kỹ thuật trong ảo hóa dữ liệu và thực nghiệm.
* *Phương pháp thực nghiệm:* Thực nghiệm lại quy trình, công nghệ đã đề xuất trên thực trạng về dữ liệu cũng như thực trạng công nghệ và mục tiêu kinh doanh hướng tới của công ty, từ đó đưa ra các kết luận, khuyến nghị.

## Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu chính của đề tài là công nghệ ảo hóa và cách triển khai công nghệ ảo hóa trong tổ chức kinh doanh.

## Nội dung của đề tài

Bài nghiên cứu này được cấu trúc thành ba chương chính, mỗi chương nhằm mục tiêu khám phá các khía cạnh khác nhau của ảo hóa dữ liệu và kinh doanh thông minh.

Chương 1: CƠ SỞ LÝ LUẬN VỀ ẢO HÓA DỮ LIỆU VÀ KINH DOANH THÔNG MINH

Chương đầu tiên cung cấp một nền tảng lý thuyết vững chắc về ảo hóa dữ liệu và kinh doanh thông minh. Phần này sẽ trình bày các khái niệm cơ bản, định nghĩa, và tầm quan trọng của ảo hóa dữ liệu trong việc nâng cao hiệu quả kinh doanh. Bên cạnh đó, chương này cũng phân tích các lợi ích, thách thức và hạn chế khi áp dụng ảo hóa dữ liệu trong các tổ chức.

Chương 2: CÔNG NGHỆ VÀ QUY TRÌNH TRIỂN KHAI ỨNG DỤNG ẢO HÓA DỮ LIỆU TRONG KINH DOANH THÔNG MINH

Chương thứ hai đi sâu vào các công nghệ và quy trình triển khai ứng dụng ảo hóa dữ liệu, nhằm hỗ trợ tối ưu cho hoạt động kinh doanh thông minh.

Chương 3: ỨNG DỤNG CÔNG NGHỆ ẢO HÓA DỮ LIỆU TRONG KINH DOANH THÔNG MINH

Chương cuối, tập trung vào việc triển khai công nghệ ảo hóa trong việc ứng dụng vào thực tiễn hoạt động của một daonh nghiệp kinh doanh cụ thể.

# CHƯƠNG 1: CƠ SỞ LÝ LUẬN VỀ ẢO HÓA DỮ LIỆU VÀ KINH DOANH THÔNG MINH

*Trong bối cảnh phát triển không ngừng của công nghệ và dữ liệu, ảo hóa dữ liệu và kinh doanh thông minh đã trở thành những lĩnh vực nghiên cứu và ứng dụng mang tính cách mạng, có khả năng nâng cao hiệu quả hoạt động của các tổ chức kinh doanh. Mục đích của chương này, là cung cấp một nền tảng lý thuyết vững chắc về hai lĩnh vực quan trọng ảo hóa dữ liệu và kinh doanh thông minh. Nội dung chương được chia thành ba phần chính, mỗi phần sẽ tập trung vào một khía cạnh cụ thể của ảo hóa dữ liệu và kinh doanh thông minh.*

## 1.1. Tổng quan về ảo hóa dữ liệu

Trong thế giới dữ liệu ngày càng phát triển và phức tạp hiện nay, ảo hóa dữ liệu đã trở thành một khái niệm không thể thiếu trong các chiến lược quản lý và phân tích dữ liệu của nhiều tổ chức kinh doanh. Việc hiểu rõ về ảo hóa dữ liệu không chỉ giúp các nhà quản lý dữ liệu tăng cường hiệu quả kinh doanh mà còn góp phần vào việc đưa ra các quyết định nhanh chóng và chính xác trong thời đại số.

### 1.1.1. Khái niệm

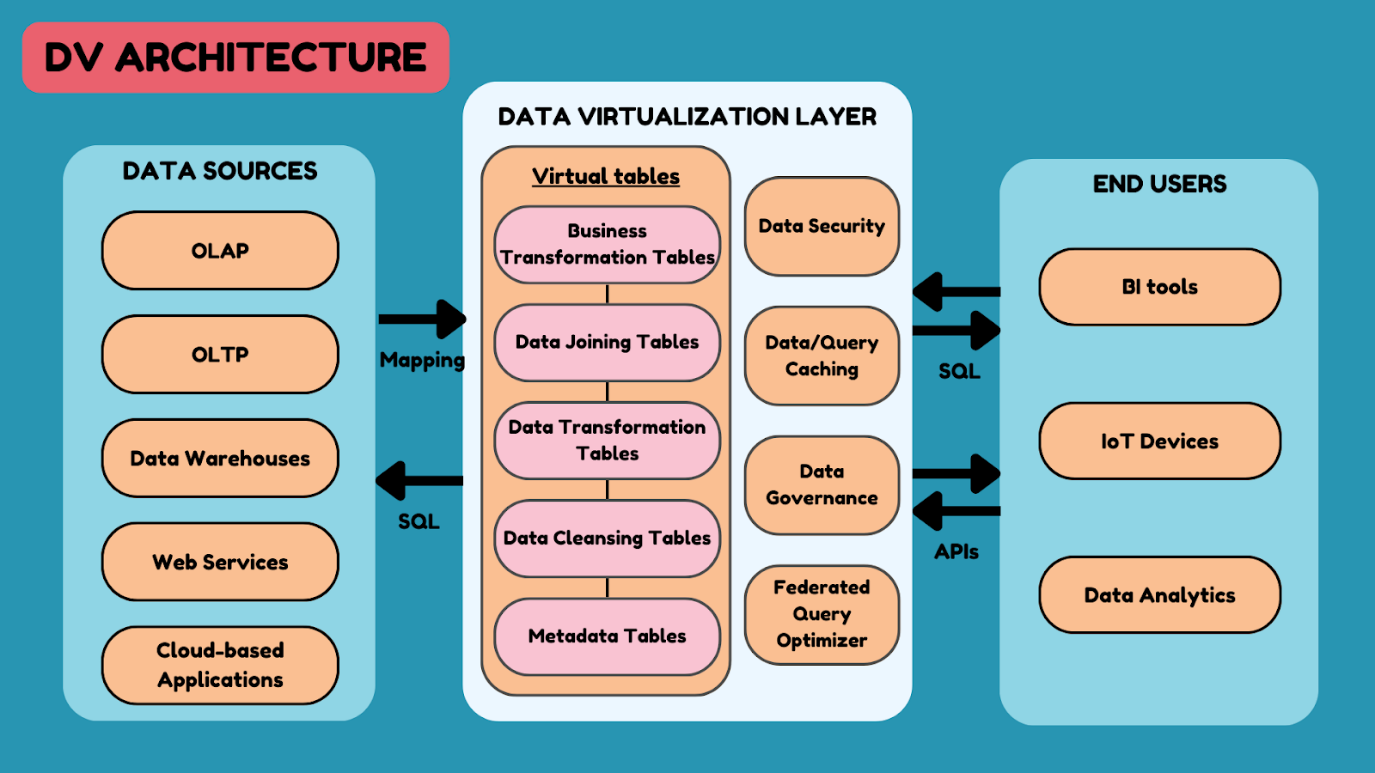
Những năm gần đây, ảo hóa dữ liệu (Data Virtualization) vẫn là một cụm từ đang được thảo luận và đề cập trong các bài báo, tạp chí nghiên cứu trong và ngoài nước. Nó được coi là một nền tảng công nghệ mới, góp phần thúc đẩy sự phát triển trong nghiệp vụ quản lý doanh nghiệp.

* Theo Manoj Muniswamaiah & cộng sự (2019), ảo hóa dữ liệu được các nhà nghiên cứu nhấn mạnh về tính mới nhất trong quy trình tích hợp dữ liệu. Nó tạo ra một lớp trừu tượng duy nhất giữa bên cung cấp và bên sử dụng để truy cập dữ liệu có nguồn gốc từ nhiều ứng dụng và vị trí lưu trữ khác nhau. Mục tiêu chính của nó là cung cấp một nguồn duy nhất cho tất cả người dùng mà không yêu cầu chi tiết kỹ thuật hoặc vị trí cụ thể của dữ liệu cơ bản trong thời gian thực (real-time).
* Theo Seth Earley (2016), ảo hóa dữ liệu cho phép dữ liệu từ các nguồn khác nhau xuất hiện dưới dạng một kho tập trung duy nhất. Thay vì di chuyển dữ liệu đến một nơi mới, dữ liệu tại nơi lưu trữ ban đầu vẫn được giữ nguyên, trong khi chất lượng, quyền sở hữu và chuyển đổi dữ liệu được quản lý ở một lớp "logic" trong hệ thống. Hệ thống các quy tắc quản trị được áp dụng nhất quán và các nguồn dữ liệu được lập danh mục và theo dõi trong một kho lưu trữ tập trung.

Có thể thấy, có nhiều khái niệm khác nhau về ảo hóa dữ liệu, tuy nhiên tất cả chúng đều có điểm chung là coi ảo hóa dữ liệu là nền tảng tích hợp dữ liệu với mục đích chính là tạo ra trung tâm tập hợp và truy xuất dữ liệu. Tuy nhiên, điểm mới và khác biệt của công nghệ này là khả năng kết nối các nguồn dữ liệu khác nhau mà không cần đến các quy trình với nhiều tầng sao chép và tải dữ liệu. Ảo hóa dữ liệu đóng vai trò là một trung gian kết nối trực tiếp người sử dụng với dữ liệu nguồn, cho phép người dùng có thể thay đổi, chỉnh sửa và truy xuất dữ liệu cần thiết mà không làm xáo trộn, thay đổi các thuộc tính của dữ liệu gốc. Đồng thời nó giảm thiểu tính phức tạp từ yêu cầu về kiểu lưu trữ dữ liệu và các yêu cầu xử lý đồng bộ dữ liệu thông qua việc ánh xạ dữ liệu dưới cấu trúc logic và cung cấp các tính năng quản trị nhằm đảm bảo tính an toàn dữ liệu cho người dùng.

### 1.1.2. Kiến trúc và hoạt động của ảo hóa dữ liệu

* Kiến trúc của áo hóa dữ liệu



**Hình 1: Kiến trúc và hoạt động của ảo hóa dữ liệu**

Kiến trúc ảo hóa dữ liệu được chia thành 3 tầng như sau:

* *Tầng kết nối:* Tương tự như trong cấu trúc của kho dữ liệu tập trung, đây là tầng có nhiệm vụ cung cấp dữ liệu bằng cách kết nối và liên kết với các nguồn dữ liệu cần thiết. Các hệ thống lưu trữ dữ liệu ở tầng này có thể được thiết kế dưới dạng có cấu trúc hoặc phi cấu trúc và dưới nhiều dạng lưu trữ khác nhau như cơ sở dữ liệu, kho dữ liệu, các ứng dụng SaaS…
* *Tầng ảo hóa/trừu tượng:* Là tầng quan trọng nhất của mô hình thực hiện các nhiệm vụ truy xuất và xử lý dữ liệu.
* *Tầng sử dụng:* Thường là các công cụ, phần mềm hỗ trợ người sử dụng thực hiện các công việc báo cáo, phân tích đánh giá dữ liệu như Excel, PowerBI, …
* Hoạt động của ảo hóa

Ảo hóa hoạt động theo các giai đoạn sau:

*Giai đoạn 1: Kết nối nguồn dữ liệu*

Ảo hóa dữ liệu sau khi nhận được yêu cầu từ người dùng cần thực hiện tập hợp dữ liệu từ các nguồn dữ liệu khác nhau. Tại các nơi lưu trữ dữ liệu, cơ chế và cách thức truy cập được thiết lập khác nhau đòi hỏi sự linh hoạt của trình liên kết dữ liệu. Dịch vụ sử dụng phương thức kết nối JDBC/ODBC để thực hiện chuyển đổi các yêu cầu thành các câu lệnh truy vấn dưới dạng SQL đối với dữ liệu thuộc cơ sở dữ liệu; hay các phương thức API khác để truy cập vào dữ liệu dịch vụ web và dữ liệu yêu cầu đặc biệt từ các ứng dụng tài chính...

*Giai đoạn 2: Tối ưu và chuẩn hóa câu lệnh truy vấn*

Ở mỗi quá trình, dữ liệu được sử dụng từ các bảng ảo từ quy trình hay bảng ảo trước làm nguồn dữ liệu. Điều này tạo ra một chuỗi các phụ thuộc trong đó mỗi bảng ảo dựa vào dữ liệu và cấu trúc của bảng ảo xuất hiện trước nó. Khi thực hiện một yêu cầu người dùng, ảo hóa sẽ sắp xếp thành câu lệnh SQL hoàn chỉnh và xâu chuỗi truy vấn để xác định các bảng ảo cho từng quá trình xử lý. Giai đoạn này cho phép chuyển đổi và tích hợp dữ liệu phức tạp trên nhiều quy trình và nguồn dữ liệu.

*Giai đoạn 3: Khai thác dữ liệu*

Các nguồn dữ liệu khác nhau có cách sắp xếp và quy tắc dữ liệu khác nhau, các lược đồ lưu trữ dữ liệu và mối quan hệ giữa chúng được thể hiện theo cấu trúc riêng của nơi lưu trữ. Ảo hóa dữ liệu cần đến các thông tin trên để có thể truy xuất dữ liệu một cách chính xác, thích hợp và đầy đủ; và chúng được tập trung lại tạo thành một lớp gọi là lớp siêu dữ liệu (metadata). Lớp siêu dữ liệu này được lưu trữ dưới dạng logic nhằm đáp ứng đặc điểm quan trọng của ảo hóa dữ liệu là hạn chế quá trình sao chép dữ liệu, đồng thời làm thuận tiện cho quá trình chuyển đổi và xử lý dữ liệu.

Tại giai đoạn này, dữ liệu sẽ được sàng lọc và hiển thị tùy theo ủy quyền của người dùng. Điều này đảm bảo rằng người dùng đáp ứng các chính sách quản trị dữ liệu cũng như tăng tỷ lệ truy cập vào dữ liệu thời gian thực (real-time).

*Giai đoạn 4: Xử lý dữ liệu*

Để tích hợp được dữ liệu, ảo hóa dữ liệu phải làm sạch dữ liệu, chuyển đổi dữ liệu và tương quan dữ liệu. Không giống như các kho dữ liệu truyền thống phải thực hiện từng quy trình này một cách riêng biệt và sao chép dữ liệu ở từng quy trình, ảo hóa dữ liệu sẽ thực hiện tất cả các quá trình này chủ yếu trong một bước trước khi đưa ra báo cáo cuối cùng. Mô hình thực hiện các thao tác xử lý dữ liệu chia thành từng bảng hoặc view ảo. Các bảng ảo là tập hợp tập hợp các kết quả truy vấn hoặc thực hiện ánh xạ dữ liệu từ nơi lưu trữ ban đầu và nội dung bảng không được lưu trữ vật lý mà chỉ thể hiện dữ liệu dưới dạng ngôn ngữ tự nhiên dễ hiểu.

Thông thường, đối với các dữ liệu nguồn có cấu trúc mô hình dữ liệu quan hệ, các kết quả của bảng ảo được thực thi bằng cách sử dụng SQL như một ngôn ngữ truy vấn dữ liệu. Tuy nhiên, cách hoạt động như vậy lại không thể đáp ứng được nhu cầu đối với các nguồn dữ liệu lưu trữ ban đầu không có cấu trúc quan hệ chẳng hạn như phân cấp, đa chiều, kho lưu trữ key -value hoặc mô hình dữ liệu đối tượng… Trong những trường hợp như vậy, DV sẽ hiển thị dữ liệu cơ bản dưới dạng bảng ảo không tường minh (implicit virtual tables). Các bảng này sẽ không có bất kỳ SQL nào xác định lược đồ mà nó sẽ chỉ là ánh xạ tự nhiên của dữ liệu cơ bản trong quan hệ. Mục đích của việc ánh xạ là chuyển đổi các mô hình dữ liệu đa dạng này thành mô hình dữ liệu quan hệ và cung cấp SQL làm ngôn ngữ thống nhất để truy cập dữ liệu.

*Giai đoạn 5: Bộ nhớ đệm lưu trữ tạm thời*

Bộ nhớ đệm dữ liệu là một chức năng quan trọng trong các kho tập trung dữ liệu. Nó cung cấp sự cân bằng với khả năng hiệu suất cao, giảm độ trễ thông tin và giảm chi phí sao chép dữ liệu không cần thiết. Tương tự như vậy, ảo hóa dữ liệu cung cấp bộ nhớ tạm thời sau mỗi thao tác hoạt động và xử lý dữ liệu nhằm kịp thời chỉnh sửa các kết quả truy vấn và xem lại lịch sử làm việc. Bên cạnh đó, bộ nhớ đệm còn giảm tải được thời gian thực hiện lại các quy trình di chuyển dữ liệu dựa trên các câu lệnh truy vấn thường xuyên. (Pullokkaran, 2013)

### 1.1.3. Lợi ích của ảo hóa dữ liệu

Ảo hóa dữ liệu là một cách tiếp cận mang tính biến đổi cho phép doanh nghiệp quản lý dữ liệu hiệu quả hơn. Nó tạo một lớp logic cho phép người dùng truy cập và thao tác nhiều loại dữ liệu mà không cần quan tâm đến các chi tiết kỹ thuật cơ bản, chẳng hạn như định dạng dữ liệu hoặc vị trí lưu trữ (Joshi, 2021). Công nghệ này đặc biệt có lợi cho các doanh nghiệp xử lý các bộ dữ liệu lớn và đa dạng, bao gồm dữ liệu lớn, phương tiện truyền thông xã hội, web và dữ liệu thiết bị IoT. Một số lợi ích hàng đầu của ảo hóa dữ liệu cho doanh nghiệp có thể được nhận ra:

* *Tính linh hoạt trong việc được ra quyết định:* Ảo hóa dữ liệu cung cấp cái nhìn thống nhất cho khách hàng, cho phép người dùng cuối truy cập và sửa đổi dữ liệu được lưu trữ trên các hệ thống khác nhau. Khả năng này rất quan trọng để kết nối dữ liệu cần thiết cho việc phân tích và ra quyết định chiến lược.
* *Truy cập dữ liệu theo thời gian thực:* Không giống như các phương pháp truyền thống thường liên quan đến việc di chuyển các khối dữ liệu lớn, ảo hóa dữ liệu sử dụng con trỏ để truy cập dữ liệu, yêu cầu ít bộ nhớ hơn, cung cấp quyền truy cập hiệu suất và dữ liệu khi cần.
* *Quản lý dữ liệu hiệu quả về mặt chi phí:* Ảo hoá dữ liệu không sao chép hoặc lưu trữ dữ liệu, giúp giảm chi phí lưu trữ. Nó kết nỗi người dùng với dữ liệu cần thiết và cung cấp dữ liệu đó theo thời gian thực, điều này có thể đặc biệt tiết kiệm chi phí cho doanh nghiệp.
* *Quản trị và bảo mật dữ liệu nâng cao:* Bằng cách tập trung bảo mật và quản trị đối với các nguồn dữ liệu khác nhau, ảo hóa dữ liệu đảm bảo rằng dữ liệu được phân phối theo cách an toàn và được quản lý, điều này rất cần thiết cho niềm tin của người tiêu dùng và sự tuân thủ quy định.
* *Khả năng mở rộng và linh hoạt:* Khi các doanh nghiệp ngày càng đi lên và phát triển, ảo hóa dữ liệu cho phép các doanh nghiệp này mở rộng quy mô và điều chỉnh các chiến lược quản lý dữ liệu của mình mà không cần thay đổi cơ sở hạ tầng trên diện rộng.
* *Hỗ trợ nhiều loại phân tích:* Ảo hóa dữ liệu hỗ trợ nhiều loại phân tích khác nhau, bao gồm phân tích dự đoán, phân tích trực quan và phân tích phát trực tuyến trên các bản cập nhật dữ liệu gần đây nhất. Điều này cho phép các doanh nghiệp tận dụng dữ liệu của mình cho nhiều mục đích phân tích khác nhau.

Ảo hóa dữ liệu là yếu tố hỗ trợ chính cho các doanh nghiệp đang tìm cách tối ưu hóa hoạt động quản lý dữ liệu của mình. Nó cung cấp một cách linh hoạt, an toàn và tiết kiệm chi phí để truy cập và phân tích dữ liệu, điều này có thể mang lại kết quả kinh doanh tốt hơn và lợi thế cạnh tranh.

## 1.2. Tổng quan về hệ thống kinh doanh thông minh

Trong bối cảnh kinh tế toàn cầu ngày càng cạnh tranh khốc liệt, các doanh nghiệp luôn tìm kiếm những giải pháp tối ưu để nâng cao hiệu quả hoạt động và khả năng đáp ứng linh hoạt trước các thay đổi của thị trường. Một trong những giải pháp được đánh giá cao và đã chứng minh được hiệu quả thực tiễn đó là xây dựng hệ thống kinh doanh thông minh (Business Intelligence - BI). Hệ thống này không chỉ hỗ trợ các nhà quản lý trong việc ra quyết định nhanh chóng và chính xác hơn mà còn góp phần thúc đẩy sự sáng tạo và đổi mới liên tục trong tổ chức.

Hệ thống kinh doanh thông minh dựa trên nền tảng của dữ liệu để phân tích và cung cấp những cái nhìn sâu sắc về hoạt động kinh doanh. Bằng việc tổng hợp và phân tích lượng lớn dữ liệu từ nhiều nguồn khác nhau, BI cho phép các doanh nghiệp không chỉ hiểu rõ về hiện trạng mà còn dự báo các xu hướng tương lai, từ đó định hình chiến lược phát triển bền vững.

### 1.2.1. Khái niệm

Trong một thế giới hiện đại và phát triển như ngày nay, việc sản sinh và truy cập dữ liệu càng trở nên dễ dàng. Ước tính từ khảo sát của Duarte (2023), có khoảng 330 triệu terabyte dữ liệu được tạo ra mỗi ngày nhưng lại chỉ có tầm 2% dữ liệu được thực sự sử dụng và lưu trữ cho các mục đích khác nhau. Điều này cho thấy rằng mặc dù có rất nhiều dữ liệu được tạo ra, nhưng lượng thông tin hữu ích từ lượng dữ liệu đó lại rất hạn chế. Đây là tình trạng chung của bất kỳ một doanh nghiệp, một công ty trong công việc quản lý và phát triển.

Có thể nói rằng, dữ liệu vừa là nguồn tài nguyên tiềm năng và dồi dào nhất cho một tổ chức nhưng nó cũng vừa là hao sản khi chiếm giữ quá nhiều và không ngừng đòi hỏi không gian lưu trữ. Vậy để có thể sử dụng chúng một cách hợp lý và có hiệu quả, doanh nghiệp cần có sự tập trung dữ liệu lớn, phân loại có chọn lọc và một quy trình cụ thể và hợp lý để đưa ra quyết định cuối cùng chuẩn xác. Khi đó, thuật ngữ kinh doanh thông minh được ra đời.

Trong một bài nghiên cứu khoa học về hệ thống hỗ trợ ra quyết định của mình, Turban và cộng sự (2005) đã định nghĩa kinh doanh thông minh (BI) là “*Một khái niệm rộng bao hàm các ứng dụng và kỹ thuật thực hiện thu thập, lưu trữ, phân tích và cấp quyền truy cập dữ liệu để giúp người dùng doanh nghiệp đưa ra quyết định chiến lược và kinh doanh tốt hơn*.”

Theo Carlos và cộng sự (2021), Kinh doanh thông minh (BI) được định nghĩa là “*một quá trình ra quyết định được hỗ trợ bởi việc tích hợp và phân tích dữ liệu tài nguyên của tổ chức*”, “có vai trò quan trọng trong quản lý doanh nghiệp”.

Về cơ bản, một hệ thống kinh doanh thông minh là tập hợp các quy trình tích hợp và lưu trữ dữ liệu để thực hiện các nhiệm vụ xử lý và phân tích chúng thành các thông tin mang tính quyết định và có ý nghĩa xây dựng trong hoạt động hay nghiệp vụ của tổ chức. Thông thường, BI đảm nhiệm việc thực thi các phương pháp ETL nhằm trích xuất, biến đổi và sao chép dữ liệu từ nhiều nguồn, dưới nhiều dạng khác nhau; sau đó tập trung dữ liệu tại một kho lưu trữ nhất định để thực hiện các công việc phân tích, đánh giá và thống kế khi cần thiết. Các công cụ BI thường được hình dung chủ yếu ở các phần mềm hỗ trợ doanh nghiệp được tiếp xúc nhanh chóng với nguồn đầu vào là dữ liệu và làm việc với báo cáo, khai phá và phân tích dữ liệu để đưa ra kết luận về một xu hướng nào đó và là tiền đề quyết định cho hoạt động doanh nghiệp.

### 1.2.2. Các công cụ Kinh doanh thông minh

Trong môi trường kinh doanh ngày nay, việc sử dụng Công cụ Kinh doanh Thông minh (BI) đã trở thành một phần quan trọng của chiến lược quản lý dữ liệu và ra quyết định. Các công cụ này không chỉ giúp tổ chức tổ chức dữ liệu một cách hiệu quả mà còn cung cấp các khả năng phân tích và báo cáo đa dạng để hỗ trợ quyết định kinh doanh. Một số loại công cụ BI phổ biến được sử dụng rộng rãi trong doanh nghiệp hiện đại:

* *Công cụ BI truyền thống*

Công cụ BI truyền thống là các ứng dụng phần mềm được phát triển để xử lý và phân tích dữ liệu kinh doanh từ các nguồn khác nhau. Chúng thường bao gồm các tính năng như truy vấn dữ liệu, báo cáo tương tác và phân tích đa chiều. Các ví dụ tiêu biểu bao gồm:

*Microsoft Power BI:* Một trong những nền tảng phổ biến nhất cho phân tích dữ liệu, Power BI cung cấp các công cụ mạnh mẽ cho việc tạo báo cáo, trực quan hóa dữ liệu và chia sẻ thông tin.

*Tableau:* Một công cụ trực quan hóa dữ liệu mạnh mẽ, Tableau cho phép người dùng tạo ra các biểu đồ và bảng điều khiển tương tác để khám phá dữ liệu và kể câu chuyện về dữ liệu một cách hấp dẫn.

* *Công cụ BI hiện đại (BI self-service)*

Công cụ BI Hiện đại, hay còn gọi là BI Self-service, đánh dấu một bước tiến quan trọng trong lĩnh vực phân tích dữ liệu. Với sự phát triển của công nghệ, người dùng không cần phải phụ thuộc vào các nhóm IT hoặc các chuyên gia dữ liệu để có thể trích xuất, phân tích và trình bày dữ liệu kinh doanh. Thay vào đó, các công cụ BI self-service cung cấp một giao diện dễ sử dụng và trực quan, cho phép người dùng không chuyên về dữ liệu tạo ra các báo cáo và trực quan hóa dữ liệu một cách độc lập.

Các công cụ này thường đi kèm với các tính năng tự động hóa và trích xuất dữ liệu từ nhiều nguồn khác nhau, bao gồm cả cơ sở dữ liệu nội bộ, dữ liệu từ bên ngoài và dữ liệu đám mây. Điều này giúp người dùng tạo ra các báo cáo và trực quan hóa dữ liệu một cách nhanh chóng và linh hoạt mà không cần sự hỗ trợ nhiều từ các nhóm IT.

Công cụ BI Self-service không chỉ giúp tăng cường sự linh hoạt và tự chủ trong việc quản lý dữ liệu mà còn tạo ra một môi trường làm việc đào tạo và khám phá dữ liệu cho toàn bộ tổ chức. Điều này có thể thúc đẩy sự sáng tạo và khám phá, giúp các nhân viên hiểu rõ hơn về dữ liệu và tạo ra giá trị kinh doanh từ dữ liệu.

Các công cụ BI tự phục vụ (self-service BI) cho phép người dùng không chuyên về dữ liệu có thể tạo và tùy chỉnh báo cáo mà không cần sự hỗ trợ từ các chuyên gia IT. Ví dụ một số công cụ như:

Qlik Sense: Qlik Sense cung cấp một giao diện trực quan và dễ sử dụng cho việc tạo báo cáo và trực quan hóa dữ liệu. Người dùng có thể kết hợp dữ liệu từ nhiều nguồn và tạo các bảng điều khiển tương tác.

Google Data Studio: Dựa trên công nghệ đám mây của Google, Data Studio cho phép người dùng tạo và chia sẻ báo cáo trực tuyến một cách dễ dàng, đồng thời tích hợp dữ liệu từ nhiều nguồn khác nhau.

* *Công cụ BI trên nền tảng đám mây*

Công cụ BI Trên Nền Tảng Đám Mây đã thay đổi cách mà doanh nghiệp tiếp cận và sử dụng dữ liệu kinh doanh. Thay vì phải đầu tư vào cơ sở hạ tầng phức tạp và tốn kém, các công cụ BI trên nền tảng đám mây cung cấp một giải pháp linh hoạt và tiết kiệm chi phí cho việc lưu trữ, xử lý và phân tích dữ liệu.

Các doanh nghiệp có thể truy cập vào các dịch vụ và tài nguyên phân tích dữ liệu từ xa thông qua internet, mà không cần phải lo lắng về việc quản lý và duy trì các máy chủ và hạ tầng phần cứng. Điều này giúp giảm bớt những hạn chế về mặt tài chính và kỹ thuật, đồng thời cung cấp khả năng mở rộng linh hoạt theo nhu cầu của doanh nghiệp.

Công cụ BI Trên Nền Tảng Đám Mây thường đi kèm với các tính năng bảo mật cao cấp, giúp bảo vệ dữ liệu kinh doanh và tuân thủ các quy định về bảo mật thông tin. Đồng thời, việc sử dụng các dịch vụ đám mây cũng giúp giảm thiểu rủi ro liên quan đến mất mát dữ liệu và phục hồi sau thảm họa.

Sự kết hợp giữa tính linh hoạt và tiết kiệm chi phí của công cụ BI Trên Nền Tảng Đám Mây đã làm cho nó trở thành một lựa chọn phổ biến đối với nhiều doanh nghiệp, đặc biệt là các doanh nghiệp nhỏ và vừa. Đồng thời, việc tích hợp các công cụ này vào môi trường kinh doanh có thể mang lại nhiều lợi ích về mặt chiến lược và cạnh tranh. Amazon QuickSight và Microsoft Azure BI Services là những ví dụ cề công cụ cung cấp các dịch vụ BI trên nền tảng đám mây.

Amazon QuickSight: QuickSight là một dịch vụ trực quan hóa dữ liệu trên đám mây của Amazon, cung cấp các công cụ cho việc tạo báo cáo, trực quan hóa dữ liệu và chia sẻ thông tin một cách dễ dàng.

Microsoft Azure BI Services: Azure BI Services cung cấp một loạt các dịch vụ và công cụ cho việc xử lý và phân tích dữ liệu trên nền tảng đám mây của Microsoft, bao gồm Power BI và Azure Synapse Analytics.

* *Công cụ Phân tích nâng cao và AI*

Công cụ Phân Tích Nâng Cao và Trí Tuệ Nhân Tạo (AI) đánh dấu một sự tiến bộ lớn trong lĩnh vực Kinh doanh Thông minh (BI), mang lại khả năng phân tích dữ liệu một cách tự động và thông minh hơn bao giờ hết. Thông qua việc áp dụng các thuật toán máy học và học sâu, các công cụ này có khả năng tự động hóa quá trình phân tích dữ liệu, từ việc nhận diện mẫu và xu hướng đến việc dự báo và đề xuất giải pháp.

Một trong những lợi ích lớn nhất của việc sử dụng công cụ Phân Tích Nâng Cao và Trí Tuệ Nhân Tạo là khả năng dự báo và đưa ra các quyết định dựa trên dữ liệu lịch sử và hiện tại. Nhờ vào việc học từ dữ liệu và tự đào tạo, các công cụ này có thể phát hiện các xu hướng tiềm ẩn, dự báo kịch bản tương lai và đề xuất các biện pháp cần thực hiện để tối ưu hóa hiệu suất kinh doanh.

Ngoài ra, công cụ Phân Tích Nâng Cao và Trí Tuệ Nhân Tạo cũng giúp tối ưu hóa quy trình ra quyết định bằng cách tự động hóa các nhiệm vụ phức tạp và tăng cường sự linh hoạt trong việc quản lý dữ liệu. Chúng có khả năng tự động phát hiện và điều chỉnh các mô hình phân tích dự liệu dựa trên các điều kiện và thay đổi trong môi trường kinh doanh.

Thêm vào đó, sự kết hợp giữa công nghệ Phân Tích Nâng Cao và Trí Tuệ Nhân Tạo và BI giúp doanh nghiệp hiểu rõ hơn về hành vi khách hàng, dự báo nhu cầu thị trường và tối ưu hóa chiến lược kinh doanh. Điều này giúp doanh nghiệp tăng cường cạnh tranh trên thị trường và tạo ra giá trị kinh doanh lớn hơn. Các công cụ này có thể giúp doanh nghiệp tìm ra các cơ hội kinh doanh mới và dự báo rủi ro. Microsoft Azure AI và IBM Watson là các ví dụ điển hình.

* *IBM Cognos Analytics:* Cognos Analytics sử dụng trí tuệ nhân tạo để cung cấp gợi ý phân tích và dự đoán, giúp người dùng tìm ra mô hình và xu hướng tiềm ẩn trong dữ liệu.
* *SAS Visual Analytics:* SAS Visual Analytics cung cấp các công cụ phân tích nâng cao và trí tuệ nhân tạo, giúp người dùng tìm hiểu sâu hơn về dữ liệu và đưa ra quyết định thông minh.

### 1.2.3. Lợi ích của Hệ thống kinh doanh thông minh

Hệ thống Kinh doanh Thông minh (BI) đã trở thành một công cụ quan trọng giúp doanh nghiệp tối ưu hóa quá trình ra quyết định, cải thiện hiệu quả hoạt động và tăng cường khả năng cạnh tranh trên thị trường. Các lợi ích chính của hệ thống BI có thể được phân tích thông qua các khía cạnh sau:

* *Cải thiện Quá trình Ra Quyết định*

Một trong những lợi ích chính của hệ thống BI là khả năng cung cấp thông tin kịp thời và chính xác để hỗ trợ quá trình ra quyết định. Theo nghiên cứu của Davenport và Harris (2007), các doanh nghiệp sử dụng BI có khả năng ra quyết định nhanh chóng và chính xác hơn so với những doanh nghiệp không sử dụng công cụ này. BI giúp tổng hợp dữ liệu từ nhiều nguồn khác nhau, phân tích và trình bày dưới dạng dễ hiểu, giúp người quản lý có cái nhìn tổng quan và sâu sắc về hoạt động kinh doanh.

* *Tăng Hiệu quả Hoạt động*

BI không chỉ giúp cải thiện ra quyết định mà còn nâng cao hiệu quả hoạt động bằng cách tự động hóa các quy trình thu thập và phân tích dữ liệu. Theo một nghiên cứu của Gartner (2020), các doanh nghiệp sử dụng hệ thống BI có thể giảm thời gian và chi phí liên quan đến việc thu thập và xử lý dữ liệu. Điều này giúp doanh nghiệp tiết kiệm nguồn lực và tập trung vào các hoạt động kinh doanh cốt lõi. Bằng cách loại bỏ các quy trình thủ công và tự động hóa các báo cáo, BI giúp tăng cường năng suất và hiệu quả hoạt động tổng thể.

* *Tăng Cường Khả năng Cạnh tranh*

Khả năng phân tích dữ liệu toàn diện và nhanh chóng của hệ thống BI giúp doanh nghiệp nhận diện và nắm bắt các cơ hội kinh doanh mới, cũng như phản ứng nhanh chóng với các thay đổi trên thị trường. Theo McKinsey (2018), các doanh nghiệp ứng dụng BI có khả năng nhận diện xu hướng thị trường sớm hơn và điều chỉnh chiến lược kinh doanh của họ một cách hiệu quả hơn. Điều này mang lại lợi thế cạnh tranh đáng kể, giúp doanh nghiệp duy trì và mở rộng thị phần.

* *Dự đoán và Quản lý Rủi ro*

Hệ thống BI cung cấp các công cụ phân tích dự đoán, giúp doanh nghiệp dự báo các xu hướng và nguy cơ tiềm ẩn. Báo cáo của Forrester (2019) chỉ ra rằng các doanh nghiệp sử dụng BI có khả năng dự đoán các vấn đề rủi ro và xử lý chúng trước khi chúng trở thành các vấn đề nghiêm trọng. Nhờ vào khả năng phân tích dữ liệu lịch sử và dự báo tương lai, doanh nghiệp có thể đưa ra các biện pháp phòng ngừa và quản lý rủi ro hiệu quả hơn.

* *Tối ưu Hóa Quản lý Tài nguyên*

BI giúp doanh nghiệp tối ưu hóa việc quản lý tài nguyên thông qua việc cung cấp cái nhìn toàn diện về các hoạt động kinh doanh. Điều này bao gồm quản lý hàng tồn kho, điều phối chuỗi cung ứng và quản lý nhân sự. Theo Harvard Business Review (2017), doanh nghiệp sử dụng BI có thể tối ưu hóa quy trình kinh doanh và quản lý tài nguyên một cách hiệu quả, giảm lãng phí và tối đa hóa lợi nhuận.

### 1.2.4. Vai trò của BI trong tổ chức kinh doanh

Có thể nói rằng, một hệ thống kinh doanh thông minh đóng vai trò là cơ sở cho mọi tổ chức kinh doanh đưa ra các quyết định và xu hướng phát triển một cách sáng suốt. Hệ thống BI giúp doanh nghiệp tận dụng và sử dụng hiệu quả nguồn tài nguyên dữ liệu của mình. Các lợi ích chính của BI có thể kể đến như sau:

*Thứ nhất*, có thể dễ thấy rằng hệ thống BI cung cấp cho doanh nghiệp khả năng xử lý và phân tích dữ liệu nhanh chóng và chính xác. Dữ liệu từ dạng thô và lộn xộn sẽ được khai phá thành các thông tin giá trị cho các báo cáo doanh nghiệp, từ đó góp phần tăng tỷ lệ chính xác cho các quyết định kinh doanh.

*Thứ hai*, tối ưu hóa quy trình phân tích kinh doanh là một điều không thiếu đối với một hệ thống được xây dựng hợp lý và quy củ. Điều này không chỉ giúp tiết kiệm thời gian hoạt động trong từng giai đoạn mà còn tăng năng suất làm việc của cả máy móc lẫn con người.

*Thứ ba,* từ việc phân tích và khả năng xây dựng báo cáo doanh nghiệp dễ dàng và thường xuyên, các nhà điều hành có thể phát hiện và giảm thiểu các hành vi gian lận trong hoạt động hệ thống tổ chức và hoạt động kinh doanh. BI tạo cơ hội cho tổ chức được theo dõi tiến độ và đo lường hiệu quả công việc, giúp cải thiện nguồn nhân lực và quản trị hệ thống.

Quan điểm về lợi ích và khả năng ứng dụng hệ thống kinh doanh thông có thể đa dạng tùy theo mỗi một doanh nghiệp có nhu cầu và sự phân chia hệ thống khác nhau. Nhưng nhìn chung, sự cần thiết của một hệ thống BI trong tổ chức là không thể bàn cãi khi thế giới đang hướng tới sự nhanh chóng và tối ưu trong mọi khả năng để đáp ứng nhu cầu của con người.

## 1.3. Kinh doanh thông minh với ảo hóa và kho dữ liệu

Trong bối cảnh thị trường ngày càng cạnh tranh, các công ty tìm kiếm những giải pháp để gia tăng hiệu suất và sự linh hoạt của hệ thống quản lý dữ liệu. Kinh doanh thông minh (BI) truyền thống, dù đã đóng góp đáng kể vào khả năng phân tích và ra quyết định dựa trên dữ liệu, vẫn tồn tại những hạn chế không nhỏ. Đặc biệt là trong việc xử lý và tích hợp các nguồn dữ liệu phong phú và đa dạng, đòi hỏi nhiều thời gian và nguồn lực, từ đó làm chậm quá trình ra quyết định và giảm hiệu quả hoạt động.

Trong hoàn cảnh đó, ảo hóa dữ liệu (DV) đã nổi lên như một công nghệ tiềm năng giải quyết các hạn chế của kinh doanh thông minh truyền thống bằng cách cung cấp một tầng trừu tượng hóa giúp truy cập dữ liệu từ nhiều nguồn một cách nhanh chóng và hiệu quả mà không cần đến quá trình di chuyển hay sao chép dữ liệu phức tạp. Sự tích hợp này không những giúp giảm đáng kể chi phí và thời gian cần thiết cho việc quản lý dữ liệu mà còn tăng cường khả năng đáp ứng linh hoạt trước các yêu cầu phân tích dữ liệu ngày càng cao.

Kết hợp kinh doanh thông minh với ảo hóa dữ liệu mang lại cho các tổ chức khả năng phân tích dữ liệu chính xác và kịp thời hơn, từ đó nâng cao khả năng ra quyết định và đổi mới sáng tạo. Điều này không chỉ làm tăng hiệu suất kinh doanh mà còn củng cố lợi thế cạnh tranh trong môi trường kinh doanh số ngày nay. Phần nghiên cứu này sẽ khám phá sâu hơn về cách thức và lý do tại sao ảo hóa dữ liệu trở thành một phần không thể thiếu trong việc phát triển và triển khai các giải pháp kinh doanh thông minh hiện đại. Phần này nhóm cung cấp một cái nhìn toàn diện về bối cảnh và tầm quan trọng của ảo hóa dữ liệu trong việc thúc đẩy phát triển của kinh doanh thông minh, hướng tới việc nêu bật những lợi ích và giải pháp mà ảo hóa dữ liệu mang lại cho kinh doanh thông minh trong thời đại số hiện nay.

### 1.3.1. Kinh doanh thông với Kho dữ liệu

Trong thời đại thông tin ngày nay, việc sử dụng dữ liệu để đưa ra quyết định kinh doanh đã trở thành một yếu tố then chốt không thể thiếu cho sự thành công của các doanh nghiệp. Kinh doanh thông minh (Business Intelligence - BI) là một công cụ mạnh mẽ giúp các nhà quản lý không chỉ hiểu biết sâu sắc về hiện trạng kinh doanh mà còn có khả năng dự báo và phản ứng nhanh chóng với các thay đổi của thị trường. Phần nghiên cứu này tập trung vào việc khám phá cách thức mà BI truyền thống có thể được tích hợp với kho dữ liệu để tăng cường khả năng phân tích và quản lý dữ liệu hiệu quả hơn.

#### a. Tổng quan về Kho Dữ liệu và BI

Kho dữ liệu, một hệ thống hoàn thiện được cấu trúc để nâng cao khả năng phân tích dữ liệu và hỗ trợ quyết định chiến lược trong môi trường doanh nghiệp, đóng vai trò không thể phủ nhận trong cải thiện hiệu suất và hiệu quả hoạt động của tổ chức. Với vai trò là một nơi lưu trữ dữ liệu tích hợp và thống nhất, chúng từng bước khẳng định sức mạnh của mình không chỉ trong việc cung cấp nguồn dữ liệu đa chiều mà còn trong việc tối ưu hóa quá trình truy vấn và phân tích, thay vì chỉ tập trung vào xử lý các giao dịch cụ thể.

Hơn nữa, sự kết hợp thông minh giữa kho dữ liệu và các công cụ Kinh doanh thông minh (BI) đã tạo ra một nền tảng vững chắc, mang lại sức mạnh cho tổ chức trong việc sản xuất báo cáo tự động, phân tích dữ liệu lịch sử một cách toàn diện, và thậm chí, dự đoán các xu hướng tiềm ẩn. Sự kết hợp này không chỉ đơn thuần là một liên kết dữ liệu mà còn là một hành động chiến lược, tạo nên một nguồn tài nguyên thông tin đáng tin cậy và linh hoạt, phục vụ cho quyết định chiến lược của doanh nghiệp.

#### b. Thu thập và Tích hợp Dữ liệu

Trong quá trình phát triển hệ thống thông tin, việc thu thập và tích hợp dữ liệu là một bước cơ bản và quan trọng, đóng vai trò không thể phủ nhận trong việc xây dựng nền tảng thông tin hiệu quả cho tổ chức. Dưới đây là hai khía cạnh quan trọng:

Tập trung dữ liệu: Kho dữ liệu là trái tim của quy trình này, là nơi tập trung và lưu trữ dữ liệu từ nhiều nguồn khác nhau như hệ thống ERP, CRM và các ứng dụng khác. Mục tiêu là lưu trữ dữ liệu dưới dạng đã được làm sạch và chuẩn hóa, tạo ra một nguồn dữ liệu đồng nhất và dễ dàng truy xuất.

ETL (Extract, Transform, Load): Quá trình này đóng vai trò then chốt trong việc đưa dữ liệu từ các nguồn khác nhau vào kho dữ liệu. Đầu tiên, dữ liệu được trích xuất từ các nguồn, sau đó được biến đổi thông qua các bước như làm sạch, chuẩn hóa và kết hợp. Cuối cùng, dữ liệu được tải vào kho dữ liệu. Quá trình ETL này đảm bảo rằng dữ liệu được cập nhật đúng thời điểm, nhất quán và sẵn sàng cho các hoạt động phân tích và sử dụng trong toàn bộ tổ chức.

#### c. Phân Tích và Truy vấn

Trong quá trình khai thác và sử dụng dữ liệu, việc phân tích và truy vấn dữ liệu đóng vai trò then chốt trong việc khai phá giá trị tiềm ẩn của thông tin. Một trong những yếu tố quan trọng nhất của kho dữ liệu là khả năng hỗ trợ các truy vấn phức tạp một cách nhanh chóng và hiệu quả. Điều này không chỉ quan trọng đối với việc cung cấp thông tin thời gian thực cho các quyết định chiến lược mà còn là cơ sở để xây dựng các báo cáo và phân tích dữ liệu phức tạp trong lĩnh vực Business Intelligence (BI).

Để thực hiện việc truy cập và phân tích dữ liệu từ kho dữ liệu một cách hiệu quả, các tổ chức thường sử dụng các công cụ BI như SAP Business Objects, IBM Cognos, hoặc Microsoft BI. Những công cụ này không chỉ cung cấp khả năng truy cập dữ liệu một cách dễ dàng mà còn cho phép tạo ra các báo cáo động, dashboard, và thực hiện phân tích chi tiết một cách trực quan và linh hoạt. Bằng cách này, các nhà quản lý và người dùng cuối có thể dễ dàng tiếp cận và hiểu được thông tin, từ đó đưa ra các quyết định chiến lược dựa trên dữ liệu có căn cứ và chính xác.

#### d. Đưa Ra Quyết Định Dựa Trên Dữ Liệu

Trong bối cảnh môi trường kinh doanh ngày nay, việc đưa ra quyết định dựa trên dữ liệu không chỉ là một tiến bộ mà là một yếu tố cần thiết để tồn tại và phát triển. Một trong những ứng dụng quan trọng của dữ liệu lịch sử từ kho dữ liệu là trong việc phân tích và dự đoán xu hướng tương lai của doanh nghiệp. Bằng cách khai thác và phân tích dữ liệu lịch sử, các nhà quản lý có thể hiểu rõ hơn về biến động và xu hướng của thị trường, từ đó đưa ra quyết định chiến lược dựa trên những dự báo có căn cứ và chính xác.

Các công cụ Business Intelligence (BI) là một phần không thể thiếu trong việc cung cấp thông tin chính xác và cập nhật để hỗ trợ quyết định kinh doanh. Bằng cách truy cập và phân tích dữ liệu từ kho dữ liệu, các nhà quản lý có thể đánh giá hiệu quả của các chiến lược, đo lường và theo dõi hiệu suất doanh nghiệp một cách kỹ lưỡng, và quản lý rủi ro một cách hiệu quả. Điều này không chỉ giúp tối ưu hóa hoạt động kinh doanh mà còn tạo ra lợi thế cạnh tranh cho tổ chức trong môi trường thị trường cạnh tranh ngày nay.

#### e. Bảo mật và Tuân Thủ

Trong môi trường kinh doanh ngày nay, việc bảo vệ dữ liệu và tuân thủ các quy định pháp lý là yếu tố không thể thiếu để đảm bảo sự an toàn và uy tín của tổ chức. Kho dữ liệu thường được trang bị các biện pháp bảo mật mạnh mẽ nhằm ngăn chặn truy cập trái phép vào dữ liệu. Các biện pháp này có thể bao gồm kiểm soát quyền truy cập chi tiết, mã hóa dữ liệu và triển khai các giao thức an ninh tiên tiến. Điều này đảm bảo rằng dữ liệu được bảo vệ một cách toàn diện và không bị tiết lộ cho các bên không có quyền. Kho dữ liệu đóng vai trò quan trọng trong việc giúp tổ chức tuân thủ các yêu cầu pháp lý và quy định về bảo mật dữ liệu, như GDPR và HIPAA. Bằng cách quản lý dữ liệu một cách có hệ thống và bảo mật, kho dữ liệu đảm bảo rằng tổ chức hoạt động trong khuôn khổ của các quy định pháp lý và tránh được các rủi ro liên quan đến việc không tuân thủ.

Kết hợp với các công cụ Business Intelligence (BI), kho dữ liệu không chỉ cung cấp một giải pháp mạnh mẽ cho việc phân tích dữ liệu doanh nghiệp mà còn đóng vai trò quan trọng trong việc tăng cường hiệu suất và khả năng cạnh tranh của tổ chức. Với sự hỗ trợ của công nghệ này, các doanh nghiệp có thể khai thác tối đa lượng dữ liệu lớn để đạt được mục tiêu kinh doanh và phát triển bền vững trong thời đại kỹ thuật số ngày nay.

### 1.3.2. Kinh doanh thông minh với ảo hóa dữ liệu

BI hiện đại là BI sử dụng công nghệ ảo hóa dữ liệu để thực hiện việc thu thập và phân tích dữ liệu từ nhiều nguồn khác nhau. Hệ thống này đề cao giải quyết các vấn đề kinh doanh bằng công cụ tự phục vụ, chuyển trọng tâm phân tích và báo cáo ra khỏi các nhà chuyên gia CNTT, đồng thời cung cấp các công cụ chính thống có quyền truy cập và linh hoạt để người dùng doanh nghiệp có thể tạo báo cáo và phân tích nhanh chóng cũng như chia sẻ dữ liệu với những người dùng khác để đưa ra quyết định và tối ưu hóa kết quả kinh doanh. Các đặc điểm chi tiết được trình bày dưới đây:

#### a. Công nghệ và ứng dụng

Trong hệ thống kinh doanh thông minh, ảo hóa dữ liệu (DV) là nền tảng công nghệ tích hợp dữ liệu được sử dụng với mục đích giúp tối ưu hóa quá trình thu thập, xử lý và phân tích dữ liệu. Nó cho phép các doanh nghiệp truy cập và quản lý dữ liệu từ nhiều nguồn khác nhau mà không cần phải sao chép hoặc di chuyển dữ liệu đó.

DV sử dụng các trình kết nối và các giao thức mạng khác nhau như API REST, SOAP hoặc JDBC, ODBC,... để phục vụ cho việc kết nối với các nguồn dữ liệu và các phần mềm phân tích báo cáo. Các nguồn dữ liệu có thể bao gồm cơ sở dữ liệu quan hệ và phi quan hệ, kho tập trung lưu trữ và các nền tảng đám mây,... đều có thể được kết nối và lấy dữ liệu đến tay người dùng, công cụ và ứng dụng doanh nghiệp như Tableau, Power BI, Cognos,...

Các trình tối ưu và xử lý truy vấn cũng được kết hợp vào nền tảng với mục đích cải thiện khả năng trả về kết quả truy vấn, tiết kiệm thời gian cho quá trình ra quyết định và đem lại những thông tin chính xác nhất với yêu cầu người dùng. Bên cạnh đó, các tùy chọn, thiết bị được triển khai nhanh hơn cũng như giao diện người dùng cuối đơn giản hơn, có thể tùy chỉnh và hợp tác, góp phần tăng mức độ nhận diện và thích ứng của người dùng đối với nền tảng.

DV cũng được xem là một lợi thế cho hệ thống kinh doanh thông minh trong việc tận dụng các nền tảng công nghệ tiên tiến nhất ngày nay như AI và học máy, Internet vạn vật, Blockchain để tăng cường khả năng cung cấp thông tin một cách nhanh chóng và chính xác, tăng cường bảo mật và minh bạch trong quản lý dữ liệu và phân tích dữ liệu từ nhiều nguồn khác nhau, giúp tổ chức đưa ra quyết định kinh doanh thông minh hơn.

Có thể nói, ảo hóa dữ liệu cung cấp một môi trường làm việc và nguồn cung ứng dữ liệu đa dạng, góp phần tăng cường hiệu suất kinh doanh và tạo ra giá trị cho doanh nghiệp bằng cách cung cấp thông tin chính xác và cập nhật trong thời gian thực.

#### b. Dữ liệu và phân tích

Dữ liệu và phân tích là một phần quan trọng của kinh doanh thông minh (BI) hiện đại và ảo hóa dữ liệu đóng vai trò chính trong việc thu thập, lưu trữ, xử lý và phân tích dữ liệu. Về mặt thu thập dữ liệu, DV cho phép tổ chức tiếp cận các dữ liệu từ nhiều nguồn khác nhau, bao gồm cả dữ liệu cấu trúc và không cấu trúc. Điều này giúp tổ chức giảm tải được thời gian và tính phức tạp trong việc đồng bộ và sao lưu dữ liệu, không hạn chế về số lượng dữ liệu thu thập và đồng thời có được một cái nhìn toàn diện về hoạt động kinh doanh của mình. Tiếp đó, xét về mặt tích hợp thì đây là một trong những lợi thế lớn nhất của ảo hóa dữ liệu khi nền tảng công nghệ này cho phép người dùng được đồng nhất dữ liệu mà không cần phải sao chép hay di chuyển dữ liệu nhờ vào tính năng ánh xạ dữ liệu trên bảng ảo. Điều này giúp giảm bớt sự phức tạp và chi phí liên quan đến quản lý dữ liệu.

Một trong các tính năng quan trọng của ảo hóa dữ liệu là khả năng phân tích dữ liệu mạnh mẽ. Tổ chức có thể sử dụng đồng thời nhiều công cụ phân tích để khám phá xu hướng, mô hình và thông tin quan trọng khác từ dữ liệu của họ thông qua nền tảng này. Ảo hóa dữ liệu góp phần che giấu và ẩn đi sự phức tạp của dữ liệu, thay vào đó nó trừu tượng hóa chúng. Các thông tin kỹ thuật chi tiết được ảo hóa dữ liệu làm đơn giản hóa và đồng thời đưa quá trình tiến hành nghiệp vụ trở nên tinh gọn và năng suất hơn. DV cũng cho phép tổ chức truy cập và phân tích dữ liệu theo thời gian thực, giúp tổ chức được nhanh chóng đưa ra quyết định kinh doanh dựa trên thông tin cập nhật mới nhất.

#### c. Kinh doanh có hướng tới khách hàng

Nhờ khả năng tích hợp dữ liệu nhanh chóng và hạn chế việc sao chép dữ liệu, DV là một trong các công cụ quan trọng giúp doanh nghiệp tiếp cận tới dữ liệu lớn (Big Data) để thu thập các thông tin về khách hàng trên mọi nền tảng. Từ đó, doanh nghiệp có thể có cái nhìn toàn diện hơn về nhu cầu và mong muốn của khách hàng, góp phần giúp đưa ra chiến lược kinh doanh thích hợp để tối ưu hóa sản phẩm và dịch vụ của họ.

#### d. Quản lý rủi ro và bảo mật

Trong môi trường kinh doanh thông minh, việc đảm bảo bảo mật dữ liệu và quản lý rủi ro là rất quan trọng trong việc bảo vệ người dùng, tài nguyên và các quyết định doanh nghiệp. Để đáp ứng được nhu cầu đó, ảo hóa dữ liệu được trang bị đầy đủ các cơ chế bảo mật mạnh mẽ bằng cách cung cấp một lớp bảo mật giữa nguồn dữ liệu và người dùng cuối. Các cơ chế được thiết kế với khả năng bảo vệ dữ liệu, bao gồm mã hóa dữ liệu, quản lý quyền truy cập, theo dõi và ghi lại hoạt động truy cập dữ liệu. Đồng thời, nền tảng công nghệ này cũng giúp doanh nghiệp tuân thủ các quy định, đảm bảo tính bảo mật dữ liệu từ nhiều nguồn khác nhau được toàn vẹn và bảo vệ thông tin khách hàng khỏi các mối đe dọa mạng.

#### e. Chuyển đổi kinh doanh và chiến lược

Bằng cách tạo ra một lớp trừu tượng giữa nguồn dữ liệu và người dùng, ảo hóa dữ liệu cho phép doanh nghiệp truy cập và sử dụng dữ liệu mà không đòi hỏi việc sao chép và di chuyển chúng thường xuyên. Điều này góp phần tạo môi trường ra quyết định nói chung và tích hợp nói riêng một cách linh hoạt và hiệu quả; cung cấp khả năng thích ứng trong mọi thay đổi về công nghệ, quy trình hay mục tiêu kinh doanh. Từ đó, doanh nghiệp được tạo điều kiện thuận lợi trong việc phát triển chiến lược kinh doanh phù hợp và triển khai các dự án chuyển đổi một cách nhanh chóng và hiệu quả.

### 1.3.3. So sánh kinh doanh thông minh với kho dữ liệu và ảo hóa dữ liệu

Dù các hệ thống kinh doanh thông minh theo phương pháp truyền thống là nền tảng trong công cuộc xây dựng và phát triển doanh nghiệp, song thời đại công nghệ số khiến cho sự thay đổi và cải tiến là không thể tránh khỏi nhằm tạo vị thế cho tổ chức không bị tụt hậu so với thị trường.

*Thứ nhất*, thay vì phải sao chép và di chuyển dữ liệu đến nhiều nơi, ảo hóa dữ liệu cung cấp môi trường ảo tập trung các nguồn dữ liệu khác nhau. Đây là đặc trưng của một hệ thống kinh doanh thông minh xu thế mới nhằm cho phép các nhà cung cấp dữ liệu và người dùng tương tác trên cùng một nền tảng, đồng thời được tham gia vào quá trình truy xuất và xử lý dữ liệu để cung cấp dữ liệu cần thiết.

*Thứ hai,* giữa hai hệ thống, BI hiện đại có bản chất linh hoạt hơn và hỗ trợ cá nhân hóa. Nó đơn giản hóa sự phức tạp của việc tích hợp và biến đổi trên các nguồn dữ liệu khiến cho thời gian thực hiện quy trình trở nên tinh gọn và nhanh chóng hơn. Bên cạnh đó, ảo hóa dữ liệu hiển thị các quá trình biến đổi và kết quả dưới dạng bảng ảo, giúp người dùng dễ dàng thay đổi các thao tác tùy theo yêu cầu của doanh nghiệp mà không phải bắt đầu lại từ đầu quy trình.

*Thứ ba*, nền tảng này hứa hẹn một phương pháp để truy cập nhanh chóng vào dữ liệu, trên nhiều nguồn dữ liệu không đồng nhất. BI hiện đại là một hệ thống góp phần xây dựng môi trường làm việc nơi mà nguồn tài nguyên của nó - dữ liệu có thể được sử dụng dễ dàng bởi tất cả người dùng trong doanh nghiệp bằng cách giảm sự phụ thuộc của người dùng vào các chuyên gia CNTT và đồng thời thiết lập một môi trường kinh doanh hợp tác và chia sẻ thông tin tự do.

*Thứ tư*, việc xây dựng hệ thống quản trị dữ liệu chặt chẽ cũng là một yếu tố giúp ảo hóa dữ liệu trở nên thích hợp cho hệ thống BI kiểu mới. Nó cung cấp một điểm truy cập trung tâm vào dữ liệu, nơi lưu trữ các thông tin và đặc điểm của dữ liệu và nơi lưu trữ nó. Sau đó nó thực hiện cơ chế bảo mật và quản lý chung trên điểm đó với mục đích kiểm soát và cấp quyền cho những ai có nhu cầu sử dụng dữ liệu.

Bảng so sánh tóm tắt:

|  |  |  |
| --- | --- | --- |
| Đặc điểm | BI với kho dữ liệu | BI với ảo hóa dữ liệu |
| Tích hợp dữ liệu | Yêu cầu sao chép và di chuyển dữ liệu đến nhiều nơi | Tập trung dữ liệu vào một môi trường ảo, giúp tương tác và xử lý dữ liệu trên cùng một nền tảng |
| Linh Hoạt và Hỗ Trợ | Thấp; phức tạp trong việc tích hợp và biến đổi dữ liệu | Cao; đơn giản hóa việc tích hợp và biến đổi dữ liệu |
| Cá Nhân Hóa | Thấp; ít sự linh hoạt trong việc điều chỉnh quy trình | Cao; dễ dàng thay đổi quy trình mà không cần bắt đầu lại từ đầu |
| Truy Cập Dữ Liệu | Chậm và phụ thuộc vào các chuyên gia CNTT | Nhanh chóng và độc lập với các chuyên gia CNTT |
| Quản Trị Dữ Liệu | Tách biệt và phức tạp | Chặt chẽ và hỗ trợ tốt cho môi trường BI hiện đại |

BI truyền thống và BI hiện đại đều có những ưu và nhược điểm riêng. BI truyền thống yêu cầu sao chép và di chuyển dữ liệu đến nhiều nơi, dẫn đến sự phức tạp trong việc quản lý và truy cập dữ liệu. Sự phụ thuộc vào các chuyên gia CNTT cũng cao, gây cản trở trong việc phân tích và ra quyết định nhanh chóng. Tuy nhiên, nó vẫn cung cấp nền tảng vững chắc cho việc lưu trữ và quản lý dữ liệu trong thời gian dài.

Ngược lại, BI hiện đại mang tính linh hoạt và hỗ trợ cá nhân hóa cao hơn. Nó tập trung dữ liệu vào một môi trường ảo, cho phép người dùng dễ dàng tương tác và xử lý dữ liệu mà không cần sự can thiệp nhiều từ các chuyên gia CNTT. Điều này giúp tiết kiệm thời gian và tăng hiệu suất công việc. Hệ thống quản trị dữ liệu chặt chẽ của BI hiện đại cũng đảm bảo an toàn và quản lý dữ liệu hiệu quả, phù hợp với yêu cầu của môi trường kinh doanh hiện đại và năng động.

Nhìn chung, BI hiện đại là một bước tiến vượt bậc, cung cấp nhiều lợi ích hơn về mặt linh hoạt, hiệu suất và quản lý dữ liệu, đáp ứng tốt hơn nhu cầu của các tổ chức trong thời đại số hóa.

### 1.1.4. Xu hướng ứng dụng ảo hóa dữ liệu trong doanh nghiệp

Theo Van der Lans (Deploying Data Virtualization in Business Intelligence Systems, 2012), “Triển khai ảo hóa dữ liệu trong hệ thống thông minh kinh doanh có nghĩa là người tiêu dùng dữ liệu đang truy cập vào các kho lưu trữ dữ liệu, chẳng hạn như kho dữ liệu trung tâm, trung tâm dữ liệu và kho lưu trữ dữ liệu cá nhân, thông qua máy chủ ảo hóa dữ liệu.” Từ khái niệm, cấu trúc của nền tảng, ta có thể thấy rằng ảo hóa dữ liệu đóng vai trò như một kho tập trung dữ liệu với nhiệm vụ chính là tích hợp dữ liệu và cung cấp môi trường truy cập, tiếp cận chúng. Với các tính năng của mình, ảo hóa dữ liệu vượt trội hơn các kho lưu trữ dữ liệu truyền thống ở khả năng hạn chế sao chép dữ liệu nhiều lần và tiết kiệm chi phí lưu trữ cũng như thời gian đáp ứng yêu cầu. Tuy vậy, bài nghiên cứu này khai thác ảo hóa dữ liệu dưới góc độ là một nền tảng bổ sung cho sự phát triển hơn của hệ thống kinh doanh thông minh hiện có, cụ thể hơn là một công cụ hỗ trợ cho những khuyết điểm của kho dữ liệu thông thường. Vai trò của ảo hóa dữ liệu cho hướng phát triển của BI sẽ được đề cập ở mục 1.3. Sự cần thiết của ảo hóa dữ liệu trong sự phát triển của BI.

## 1.4. Tổng kết Chương 1

Trong chương này, đề tài giới thiệu cơ sở lý thuyết về ảo hóa dữ liệu và kinh doanh thông minh cũng như mối quan hệ giữa chúng. Các tổ chức sử dụng ảo hóa dữ liệu có thể thấy rằng họ có thể đạt được một cái nhìn thống nhất về dữ liệu từ nhiều nguồn và định dạng khác nhau, qua đó cung cấp cái nhìn sâu sắc hơn về hoạt động kinh doanh và tăng cường khả năng ra quyết định dựa trên dữ liệu. Chương tiếp theo, Chương 2 nhóm nghiên cứu sẽ đi sâu vào các công nghệ ảo hóa dữ liệu cụ thể và các quy trình ảo hóa dữ liệu được áp dụng trong thực tế, khám phá cách chúng được triển khai để hỗ trợ nghiệp vụ và đáp ứng các mục tiêu kinh doanh thông minh một cách hiệu quả.

# CHƯƠNG 2: CÔNG NGHỆ VÀ QUY TRÌNH TRIỂN KHAI ỨNG DỤNG ẢO HÓA DỮ LIỆU TRONG KINH DOANH THÔNG MINH

*Trong chương này, trước tiên, đề tài tìm hiểu về các công nghệ phổ biến trong lĩnh vực ảo hóa dữ liệu, từ các công cụ phân tích đơn giản đến các hệ thống phức tạp ứng dụng trí tuệ nhân tạo và học máy. Tiếp theo, đề tài sẽ đi sâu vào kiến trúc hệ thống ảo hóa dữ liệu, điều này là cực kỳ quan trọng để hiểu cách các dữ liệu được tổ chức, lưu trữ và truy cập để đảm bảo tính linh hoạt và hiệu quả. Cuối cùng, đề tài xem xét quy trình triển khai ứng dụng ảo hóa và cách nó hỗ trợ quyết định kinh doanh thông minh, từ việc thu thập dữ liệu đến phân tích và ứng dụng kết quả vào các chiến lược kinh doanh*.

## 2.1. Các công nghệ ảo hóa dữ liệu

Các công nghệ ảo hóa được giới thiệu dưới đây, là các công nghệ được đánh giá cao theo kết quả khảo sát và báo cáo tổng hợp về các công nghệ Data Virtualization của Gartner năm 2023.

### 2.1.1. Các công cụ ảo hóa dữ liệu

a*. Denodo Platform*

Denodo là một nền tảng tiên phong trong lĩnh vực ảo hóa dữ liệu, cung cấp giải pháp tích hợp dữ liệu linh hoạt và hiệu quả. Nền tảng Denodo đã đạt giải thưởng là nền tảng hàng đầu về tích hợp, quản lý và cung cấp dữ liệu bằng cách sử dụng một phương pháp hợp lý để cho phép BI tự phục vụ (self-service), khoa học dữ liệu, tích hợp dữ liệu đa đám mây và quản lý dữ liệu doanh nghiệp.

Vào năm 1999, Angel Viña - một giáo sư tại Đại học A Coruña, Tây Ban Nha đã nhận ra một vấn đề rằng việc tích hợp dữ liệu theo cách truyền thống với kho dữ liệu có khả năng khiến các doanh nghiệp phát triển không bền vững. Ông đã tưởng tượng một chiến lược tích hợp dữ liệu dựa trên ảo hóa dữ liệu, một chiến lược hiện đại mà người dùng có thể tích hợp dữ liệu mà không cần sao chép nó, và từ đó, Denodo ra đời.

Nền tảng Denodo giúp tổ chức kết nối với tất cả dữ liệu của họ trong thời gian thực. Điều này giúp giải phóng dữ liệu và đưa nó vào tầm tay của người dùng doanh nghiệp mà không cần lo lắng về nơi lưu trữ, định dạng hoặc tốc độ thay đổi của dữ liệu. Denodo được thiết kế là nơi mà dữ liệu trở nên phổ biến và an toàn cho tất cả người dùng và ứng dụng kinh doanh, nó giải quyết vấn đề về tích hợp dữ liệu và giúp dữ liệu tự do lưu thông.

Denodo cho phép tạo ra các biểu diễn dữ liệu dưới dạng logic, giúp người dùng có thể truy vấn và phân tích dữ liệu từ các hệ thống phân tán một cách nhanh chóng và dễ dàng. Nền tảng này sử dụng công nghệ ảo hóa dữ liệu để thiết lập một lớp truy cập dữ liệu duy nhất, cho phép truy cập thời gian thực đến dữ liệu được lưu trữ trên nhiều nguồn dữ liệu đa dạng. Điều này giúp giảm chi phí tích hợp và cung cấp dữ liệu, đồng thời tăng cường hiệu quả quản lý và linh hoạt trong việc ra quyết định dựa trên dữ liệu.

Denodo cũng cung cấp các khả năng như tăng tốc truy vấn thông minh với sự hỗ trợ của trí tuệ nhân tạo, ngữ nghĩa tiên tiến thông qua danh mục dữ liệu giúp khám phá dữ liệu dễ dàng hơn, và kết nối đa dạng với hơn 200 kết nối, hỗ trợ việc công bố dữ liệu thông qua các giao diện chuẩn như SQL, JSON, REST và GraphQL APIs.

Đối với các nhà nghiên cứu, Denodo mang lại lợi ích lớn trong việc tự động hóa quá trình tích hợp dữ liệu từ các nguồn khác nhau sử dụng AI/ML, giúp họ nhanh chóng trực quan hóa dữ liệu cần thiết để xây dựng mô hình và phát triển phân tích. Đây là một yếu tố quan trọng giúp tăng tốc độ nghiên cứu và phát triển các giải pháp khoa học dựa trên dữ liệu chính xác và cập nhật.

*b. TIBCO Data Virtualization*

Được phát triển bởi TIBCO Software Inc, TIBCO Data Virtualization là một giải pháp định hướng phát triển ảo hóa dữ liệu mạnh mẽ, giúp các doanh nghiệp giải quyết các thách thức về tích hợp dữ liệu và giảm độ phức tạp và chi phí trong khi vẫn có thể truy cập dữ liệu một cách nhanh chóng. TDV cung cấp khả năng kết nối dữ liệu từ nhiều nguồn bao gồm cả dữ liệu doanh nghiệp truyền thống, dữ liệu lớn, dữ liệu đám mây và IoT, sử dụng nó phục vụ cho nhiều ứng dụng phân tích, tự phục vụ, hệ thống kinh doanh thông minh và hệ thống chuyển giao.

TIBCO Data Virtualization được thiết lập với mục tiêu giảm thời gian tìm kiếm dữ liệu và tăng khả năng thực hiện các hoạt động phân tích. Nền tảng này cung cấp người dùng điều kiện để có thể có một cái nhìn thống nhất từ nhiều nguồn dữ liệu trong doanh nghiệp mà không cần di chuyển hoặc sao chép dữ liệu vào kho dữ liệu vật lý. TIBCO mong muốn được giúp doanh nghiệp xây dựng một nền văn hóa nơi mà dữ liệu trở thành tài sản quý báu và được sử dụng trong các quyết định kinh doanh thông minh.

TIBCO Data Virtualization cho phép tổ chức quản lý và kết nối với tất cả dữ liệu trong một góc nhìn duy nhất, với hơn 350 kết nối được xây sẵn và khả năng tạo ra kết nối tùy chỉnh. TDV cung cấp người dùng khả năng truy cập dữ liệu thông qua JDBC, ODBC, REST và SOAP; đồng thời thiết kế 3 chế độ on-premise, cloud và AWS Marketplace và cung cấp cấu trúc mô-đun, cache và các trình tối ưu hỗ trợ tất cả các giai đoạn của việc phát triển, vận hành và quản lý ứng dụng nhằm đáp ứng được nhu cầu quản lý dữ liệu của doanh nghiệp.

Nền tảng này xây dựng môi trường dịch vụ dữ liệu ảo để truy cập, biến đổi và cung cấp dữ liệu mà doanh nghiệp cần để thực hiện các hoạt động kinh doanh của mình. Người dùng cũng được tạo điều kiện được nhanh chóng tận dụng giá trị từ dữ liệu của mình thông qua việc dễ dàng truy cập và giải quyết các thách thức kinh doanh. Bên cạnh đó, khi kết hợp với ứng dụng phân tích, khả năng ảo hóa dữ liệu của TIBCO cho phép doanh nghiệp xem dữ liệu thay đổi theo thời gian thực và quản lý dữ liệu trong một trải nghiệm thống nhất và toàn vẹn.

c*. IBM Data Virtualization*

IBM Data Virtualization là một phần của nền tảng IBM Cloud Pak for Data và là một phần của chiến lược mở rộng dịch vụ đám mây của IBM. Nền tảng này được ra mắt vào năm 2019 với mục tiêu cung cấp một nền tảng linh hoạt và mở rộng cho khách hàng, giúp khách hàng tối ưu hóa việc sử dụng dữ liệu và áp dụng AI trong kinh doanh.

Mục tiêu chính của IBM Data Virtualization là giảm bớt sự phức tạp và chi phí liên quan đến việc quản lý dữ liệu lớn, đồng thời cung cấp quyền truy cập dữ liệu nhanh chóng và linh hoạt. Được IBM giới thiệu như một giải pháp ảo hóa dữ liệu tiên tiến, IBM DV cung cấp khả năng truy vấn và phân tích dữ liệu từ nhiều nguồn khác nhau mà không cần sao chép hoặc di chuyển dữ liệu bằng cách hỗ trợ kết nối với nhiều loại cơ sở dữ liệu thông qua JDBC và có các connector tích hợp sẵn, giúp doanh nghiệp tận dụng tối đa giá trị từ dữ liệu của mình. Bên cạnh đó, nền tảng này cho phép doanh nghiệp tạo một cái nhìn thống nhất về dữ liệu, tăng cường khả năng phân tích và ra quyết định dựa trên dữ liệu.

IBM Data Virtualization đóng vai trò là một công cụ quan trọng giúp doanh nghiệp nắm bắt thông tin từ nhiều nguồn dữ liệu và đưa ra quyết định dựa trên dữ liệu chính xác và nhanh chóng. Với khả năng mở rộng và tích hợp cao, nó cho phép doanh nghiệp tùy chỉnh và mở rộng các dịch vụ theo nhu cầu cụ thể, từ đó tạo ra giá trị kinh doanh lớn hơn từ dữ liệu của họ.

Ngoài ra, IBM Cloud Pak for Data còn cung cấp các chức năng quản lý và bảo mật dữ liệu, với việc xác thực và ủy quyền tập trung cho phép người dùng truy cập các nguồn dữ liệu trong một môi trường đáng tin cậy. Các vai trò ảo hóa dữ liệu cung cấp quản lý truy cập chi tiết đến các tài sản ảo hóa, và mọi giao tiếp giữa môi trường và ứng dụng đều được mã hóa an toàn với công nghệ IBM và mã hóa SSL/TLS thông qua các giao thức tiêu chuẩn.

*d. Red Hat JBoss Data Virtualization*

Red Hat JBoss Data Virtualization là một giải pháp ảo hóa dữ liệu mạnh mẽ, giúp các tổ chức tối ưu hóa việc truy cập và quản lý dữ liệu từ nhiều nguồn khác nhau. Đây là một phần của bộ sản phẩm Red Hat JBoss Middleware và được thiết kế để đáp ứng nhu cầu tích hợp dữ liệu ngày càng tăng trong thời đại số.

Red Hat JBoss Data Virtualization được phát triển bởi Red Hat, một công ty nổi tiếng với các giải pháp mã nguồn mở. Nền tảng này được ra mắt vào năm 2014, với mục tiêu giúp doanh nghiệp tối ưu hóa việc sử dụng dữ liệu thông qua việc giải quyết các thách thức liên quan đến việc truy cập và phân tích dữ liệu từ các nguồn phân tán; đồng thời hỗ trợ áp dụng trí tuệ nhân tạo (AI) trong kinh doanh, từ đó tạo ra giá trị kinh doanh lớn hơn từ dữ liệu của họ.

Red Hat JBoss Data Virtualization cung cấp cho người dùng một cái nhìn thống nhất về dữ liệu, giúp người dùng cuối có thể truy cập và phân tích thông tin mà không cần quan tâm đến vị trí hay cấu trúc vật lý của dữ liệu. Điều này giúp giảm thiểu sự phức tạp và chi phí liên quan đến việc di chuyển dữ liệu, đồng thời tăng cường khả năng phân tích và ra quyết định dựa trên dữ liệu.

Red Hat JBoss Data Virtualization hoạt động bằng cách tạo ra một cơ sở dữ liệu ảo (VDB) để ánh xạ nguồn dữ liệu vật lý thành các biểu diễn tích hợp. Khi chạy, các truy vấn gửi đến các biểu diễn này được phối hợp giữa các nguồn dữ liệu vật lý phụ thuộc, theo tiêu chí truy vấn và ánh xạ được xác định bởi VDB. Điều này cho phép truy cập dữ liệu từ nhiều nguồn như cơ sở dữ liệu quan hệ, tệp văn bản, dịch vụ web, và hệ thống ERP/CRM, cũng như nguồn dữ liệu “big data” như Apache Hadoop (Hive) và MongoDB.

JBoss Data Virtualization không chỉ đặt nền tảng cho việc chạy ảo hóa dữ liệu tại on-premise mà còn cung cấp dịch vụ trên đám mây, cho phép người dùng được linh hoạt và tự chủ trong các hoạt động quản lý dữ liệu của mình và tạo nền móng cho sự phát triển của hệ thống kinh doanh thông minh tự phục vụ (self-service BI) và Agile SOA.

*e. Microsoft Data Virtualization*

Tháng 3 năm 2022, Microsoft đã ra mắt tính năng ảo hóa dữ liệu trên nền tảng Azure SQL Managed Instance. Khả năng này mang lại một khái niệm trừu tượng hóa dữ liệu, cho phép truy vấn minh bạch dữ liệu được lưu trữ trong các tệp ở Azure Data Lake Storage và Azure Blob Storage mà không cần sao chép dữ liệu.

Một điểm đặc biệt trong nền tảng ảo hóa dữ liệu của Microsoft là tính năng PolyBase, một tính năng cho phép SQL Server truy vấn dữ liệu trực tiếp từ các nguồn khác nhau bằng T-SQL. Các nguồn này bao gồm SQL Server, Oracle, Teradata, MongoDB, Hadoop clusters, Cosmos DB, và S3-compatible object storage mà không cần cài đặt riêng phần mềm kết nối khách hàng. Ảo hóa dữ liệu với PolyBase cho phép dữ liệu giữ nguyên vị trí gốc và định dạng của nó; người dùng có thể ảo hóa dữ liệu ngoài doanh nghiệp thông qua SQL Server và giảm thiểu việc di chuyển dữ liệu qua các quy trình ETL. PolyBase cũng hỗ trợ kết nối với Azure Databricks thông qua ODBC để tích hợp với dữ liệu được quản lý bởi các bảng Delta Lake.

Ảo hóa dữ liệu của Microsoft cung cấp một giải pháp linh hoạt và hiệu quả để quản lý và phân tích dữ liệu đa dạng từ nhiều nguồn khác nhau, giúp đơn giản hóa quá trình truy cập và phân tích dữ liệu cho các tổ chức.

*f. CData Software - Data Virtuality*

CData Software, một công ty hàng đầu trong lĩnh vực tích hợp dữ liệu, đã phát triển nền tảng Data Virtualization nhằm giải quyết vấn đề phân tán dữ liệu trong thời đại số hóa ngày nay. Được thành lập vào năm 2014, CData đã nhanh chóng trở thành một phần không thể thiếu trong hệ sinh thái dữ liệu của nhiều tổ chức, cung cấp giải pháp cho việc truy cập và phân tích dữ liệu từ nhiều nguồn khác nhau.

Mục tiêu chính của ảo hóa dữ liệu trong CData là tạo ra một lớp dữ liệu logic, cho phép truy cập dữ liệu thời gian thực hoặc gần thời gian thực từ các nguồn dữ liệu khác nhau, dù là trên cơ sở hạ tầng cục bộ hay đám mây. Điều này giúp loại bỏ nhu cầu sao chép hoặc di chuyển dữ liệu, từ đó hỗ trợ các trường hợp sử dụng kinh doanh đa dạng và tải trọng công việc.

Với việc mua lại một nền tảng ảo hóa dữ liệu khác là Data Virtuality vào tháng 4 năm 2024, CData đã mở rộng khả năng của mình trong việc cung cấp kết nối trực tiếp giữa các hệ thống mà không cần di chuyển dữ liệu, cũng như giải pháp ảo hóa dữ liệu cho một lớp ngữ nghĩa thống nhất và được quản lý. Với hơn 200 kết nối được xây dựng sẵn của Data Virtuality, nền tảng này hỗ trợ kiến trúc dữ liệu hiện đại bao gồm đám mây đa dạng, data fabric và data mesh. Điều này không chỉ giúp rút ngắn thời gian phân tích và đưa ra quyết định kinh doanh mà còn tăng cường khả năng quản lý và kiểm soát dữ liệu trong môi trường đám mây ngày càng phức tạp trong các tổ chức doanh nghiệp ngày nay. Data Virtuality đã giúp CData Software đánh dấu một bước phát triển mới, giúp hỗ trợ các doanh nghiệp với các dự án ảo hóa dữ liệu quy mô lớn và đáp ứng nhu cầu ngày càng tăng của thị trường.

*g. Atscale*

AtScale là một công ty được thành lập vào năm 2013 bởi các thành viên của nhóm dữ liệu tại Yahoo. Khi làm việc với dữ liệu lớn, họ đã đối mặt với những thách thức về quy mô dữ liệu, hiệu suất truy vấn, sự phát tán của các chỉ số, các đường ống dữ liệu (data pipelines) phức tạp và shadow BI - hệ thống BI được triển khai và quản lý bởi những người không có chuyên môn; từ đó, các thành viên bắt đầu xây dựng và phát triển một lớp ngữ nghĩa (semantic layer) trên nền tảng AtScale.

AtScale tập trung vào việc giúp khách hàng tạo ra giá trị từ dữ liệu và đầu tư vào công nghệ AI/ML. Công ty cho phép ra quyết định thông minh hơn bằng cách tăng tốc dòng thông tin dựa trên dữ liệu. Nền tảng lớp ngữ nghĩa của công ty đơn giản hóa, tăng tốc và mở rộng khả năng phân tích dữ liệu và khoa học dữ liệu cho khách hàng doanh nghiệp trên tất cả các ngành công nghiệp. Với AtScale, người dùng được trao quyền để tự chủ hóa dữ liệu, triển khai BI tự phục vụ và xây dựng cơ sở hạ tầng phân tích linh hoạt hơn để ra quyết định tốt hơn, có tác động mạnh mẽ hơn.

AtScale là một công cụ ảo hóa dữ liệu với nhiều tính năng và kỹ thuật đặc biệt. AtScale cho phép người dùng truy cập kho dữ liệu bất kể vị trí hoặc nền tảng mà dữ liệu cư trú và tạo điều kiện cho việc tiến hành liên kết các bảng dữ kiện khác nhau để người dùng có thể có một cái nhìn cụ thể và tổng quan hơn về thông tin. AtScale cũng áp dụng học máy để tạo ra các cấu trúc dữ liệu thông minh giúp tăng tốc các truy vấn, trả về kết quả nhanh hơn và giảm chi phí truy vấn. Nền tảng này còn hỗ trợ các phương pháp bảo mật tốt nhất, từ TLS bảo vệ việc truyền dữ liệu, đến LDAPS, Active Directory, IdP và SAML đảm bảo tính xác thực để mọi truy vấn được thực hiện trên kho dữ liệu ảo đều được liên kết với người dùng cuối. AtScale được thiết kế để hoạt động một cách mượt mà với các công cụ kinh doanh thông minh khác nhau và là cầu nối các kho dữ liệu đám mây, kho kết hợp và nội bộ để đảm bảo truy cập dữ liệu liền mạch trên các môi trường khác nhau.

Mục tiêu của AtScale Intelligent Data Virtualization là nhằm đơn giản hóa, tăng tốc và mở rộng khả năng kinh doanh thông minh và khoa học dữ liệu cho khách hàng doanh nghiệp trên tất cả các ngành công nghiệp. Nó cung cấp một giao diện duy nhất, logic của tất cả dữ liệu và cung cấp thời gian phân tích nhanh hơn bằng cách loại bỏ công việc kỹ thuật dữ liệu thủ công cần thiết để di chuyển vật lý dữ liệu đến một nơi.

*h. SAS Federation Server*

SAS Federation Server là một sản phẩm của SAS, một công ty phần mềm thống kê và phân tích dữ liệu hàng đầu thế giới. SAS Federation Server được thiết kế để cung cấp một giải pháp quản lý dữ liệu mạnh mẽ và linh hoạt. SAS Federation Server trước đây được gọi là DataFlux Federation Server. Sự đổi tên này là kết quả của việc tích hợp các sản phẩm DataFlux vào bộ sản phẩm chất lượng dữ liệu, tích hợp dữ liệu, quản lý dữ liệu và giải pháp quản lý dữ liệu của SAS.

SAS Federation Server hoạt động như một trung tâm bằng cách truy cập, quản lý và chia sẻ dữ liệu SAS cũng như một số cơ sở dữ liệu quan hệ phổ biến, bao gồm DB2, Oracle, SAP, SQL Server, Teradata và Greenplum. Nó cho phép khả năng truy vấn mạnh mẽ và cải thiện quản lý nguồn dữ liệu. SAS Federation Server giúp dễ dàng hơn cho người dùng doanh nghiệp truy cập dữ liệu an toàn để báo cáo và phân tích. Nó cung cấp khả năng xây dựng nhanh chóng các chế độ xem ảo của dữ liệu từ nhiều nguồn, cho phép người dùng mức độ kiểm soát thích hợp mà không cần di chuyển dữ liệu một cách vật lý. Điều này giúp cải thiện hiệu suất truy cập dữ liệu bằng cách sử dụng các khả năng đẩy cơ sở dữ liệu, kết nối pooling, caching quyền truy cập đặc quyền và caching dữ liệu trong bộ nhớ.

*i. Actifio Virtual Data Pipeline - Google Cloud*

Nền tảng Actifio Virtual Data Pipeline (VDP) là một phần của Actifio, một công ty công nghệ thông tin tư nhân được thành lập vào tháng 7 năm 2009 bởi Ash Ashutosh tại Waltham, Massachusetts.

Nền tảng Actifio sử dụng công nghệ Virtual Data Pipeline để giải quyết vấn đề dữ liệu sao chép, giúp quản lý dữ liệu một cách hiệu quả và giảm chi phí lưu trữ. Mục tiêu của Actifio là cung cấp một giải pháp quản lý dữ liệu tiên tiến, giúp các chuyên gia CNTT đối phó với sự bùng nổ dữ liệu và chi phí tăng cao. Công nghệ VDP của Actifio cho phép khách hàng kiểm soát tối ưu các mục tiêu quản lý dữ liệu của họ, hỗ trợ độ bền vững, linh hoạt kinh doanh và khả năng di chuyển đám mây. Nền tảng này cho phép tạo ra các bản sao ảo không giới hạn từ một bản sao gốc duy nhất, loại bỏ nhu cầu có nhiều bản sao vật lý của dữ liệu, từ đó giúp các tổ chức tối ưu hóa các hoạt động kinh doanh như sao lưu và phục hồi, kiểm thử và phân tích dữ liệu. Actifio cung cấp một nền tảng quản lý dữ liệu tập trung, giúp giảm phức tạp vận hành và cho phép quản lý dữ liệu dựa trên các SLA, tối ưu hóa việc quản lý vòng đời dữ liệu mà không ảnh hưởng đến hệ thống sản xuất.

Tháng 12 năm 2020, Google Cloud đã công bố mua lại Actifio, việc này đã giúp nền tảng này tăng cường khả năng sao lưu và phục hồi dữ liệu, đồng thời mở rộng khả năng tiếp cận và chuyển giao dữ liệu đến với quy mô lớn hơn, đặc biệt là trên các môi trường đám mây và hybrid. Điều này tạo điều kiện thuận lợi cho người dùng khả năng bảo vệ dữ liệu của doanh nghiệp khi họ triển khai và quản lý các tải trọng công việc quan trọng nhưng vẫn tiết kiệm điện năng và không gian lưu trữ dữ liệu.

*k. Stone Bond Enterprise Enable*

Stone Bond Enterprise Enabler là một nền tảng ảo hóa giúp thống nhất các nguồn thông tin tách biệt, cung cấp khả năng nhìn tổng quan trên nhiều nguồn dữ liệu theo thời gian thực. Điều này cho phép người dùng phân tích dữ liệu kinh doanh ngay khi nó xảy ra để tối đa hóa việc sử dụng tài nguyên doanh nghiệp, giảm chi phí và cải thiện quy trình kinh doanh.

Enterprise Enabler tạo ra các chế độ xem logic của dữ liệu từ các vị trí nguồn gốc, loại bỏ phần lớn sự phức tạp do các kho dữ liệu tách biệt lan rộng trên doanh nghiệp và cung cấp giao diện thân thiện cho cả người dùng có hoặc không có chuyên môn kỹ thuật. Nền tảng này sử dụng công nghệ AppComm độc quyền và kết nối với hơn 100 nguồn lưu trữ để đưa tất cả dữ liệu của tổ chức lại với nhau thông qua ảo hóa dữ liệu, loại bỏ nhu cầu phải đầu tư đắt đỏ và duy trì kho dữ liệu truyền thống.

Stone Bond Enterprise Enabler sử dụng các tính năng Data Catalog và Data Lineage nhằm tìm kiếm dữ liệu và xác định các quy tắc của tập dữ liệu trong giao diện mới. Bên cạnh đó, để bảo đảm tính an toàn dữ liệu, nền tảng phát triển Data Masking và Data Obfuscation để thực hiện nhiệm vụ bảo mật các thông tin nhạy cảm và kiểm soát quyền truy cập. SBEE cũng thiết lập cổng thông tin tự phục vụ (Self Service Web Portal) với mục đích phục vụ tập trung hướng đến các Data Steward và Data Scientist, những người có thể truy cập và phê duyệt các mô hình dữ liệu từ cổng thông tin web bất kể vị trí. Ngoài ra, Enterprise Enabler còn hỗ trợ ảo hóa dữ liệu với SharePoint 2013, cung cấp một nền tảng tích hợp toàn diện được tinh chỉnh cụ thể cho SharePoint. Điều này giúp người dùng và ứng dụng tương tác với dữ liệu trực tiếp như thể nó là một nguồn duy nhất. Nền tảng Stone Bond Enterprise Enabler cũng có hỗ trợ kết nối với Amazon Web Services (AWS) giúp tạo điều kiện cho quá trình tích hợp trên cả môi trường đám mây, hybrid, hay multi-cloud mà không đòi hỏi người dùng phải cài đặt phần mềm hay thiết lập thêm các tính năng.

### 2.1.2. Nhận xét các công nghệ ảo hóa dữ liệu

Phần trước đã giới thiệu và khám phá những nền tảng và công nghệ ảo hóa dữ liệu cụ thể hiện có, cung cấp một bức tranh tổng quát về cách thức mà các công nghệ này đang thay đổi bối cảnh quản lý và phân tích dữ liệu trong kinh doanh thông minh. Để hiểu rõ hơn về hiệu quả và khả năng ứng dụng của từng công nghệ, phần này sẽ thực hiện so sánh các công cụ ảo hóa dữ liệu dựa trên các tính năng chính và điểm khác biệt của công cụ so với các công cụ khác:

|  |  |  |  |
| --- | --- | --- | --- |
| STT | CÔNG NGHỆ/  CÔNG CỤ | CÁC TÍNH NĂNG CHÍNH | ĐIỂM KHÁC BIỆT |
| 1 | Denodo Platform | Khả năng quản trị dữ liệu tích hợp giúp quản trị dữ liệu về tuân thủ và quyền riêng tư  Tính năng Data Catalog và Data Lineage cho khả năng truy vấn tìm kiếm ngữ nghĩa để tìm dữ liệu cũng như cung cấp thông tin chi tiết về cách dữ liệu được sử dụng bởi những người dùng và ứng dụng khác  Xử lý song song với khả năng tối ưu hóa truy vấn giảm thiểu tải lưu lượng mạng và có thể giúp cải thiện thời gian phản hồi cho các tập dữ liệu lớn | Sử dụng AI để tự động đề xuất các tập dữ liệu và tạo truy vấn trong Data Catalog  Tính năng kết hợp trợ lý ảo trong truy vấn để giúp phân tích các truy vấn trước đây và đề xuất các chuẩn lệnh mới để tăng tốc truy vấn |
| 2 | TIBCO Data Virtualization | Sử dụng Business Directory giúp khiến dữ liệu trở nên dễ hiểu hơn, cung cấp khả năng chia sẻ dữ liệu để giúp người dùng phân tích  Cung cấp một công cụ công cụ chuyển đổi tích hợp cho các nguồn dữ liệu phi cấu trúc  Cho phép dữ liệu ảo hóa được cung cấp dưới dạng dịch vụ dữ liệu bằng cách sử dụng các mô tả WSDL (Ngôn ngữ mô tả dịch vụ web) được tạo tự động  Cập nhật dữ liệu liên tục bằng thuật toán tối ưu hóa hiệu suất nâng cao và bảo mật | Khả năng tạo các dịch vụ dữ liệu từ dữ liệu ảo hóa  Cho phép người dùng doanh nghiệp chọn từ Data Catalog để xây dựng chế độ xem cho dữ liệu  Khả năng kết nối với số lượng lớn (350+) các nguồn dữ liệu phân tán, bao gồm cơ sở dữ liệu doanh nghiệp truyền thống, kho dữ liệu lớn, dịch vụ đám mây và thiết bị IoT |
| 3 | IBM Data Virtualization | Kiểm soát quy trình làm việc tích hợp để phân tích và quản trị dữ liệu.  Khả năng triển khai phần mềm cả dưới dạng dịch vụ hoặc on-premise.  Tự động hóa quản trị dữ liệu và Data Catalog để đảm bảo rằng các truy vấn chạy trên dữ liệu hợp lệ, chất lượng cao | Giải pháp hội tụ ảo hóa và phân tích dữ liệu cho môi trường làm việc trên đám mây và AI  Khả năng tích hợp các công cụ và ứng dụng khác trong môi trường của IBM Cloud Pak for Data |
| 4 | Red Hat Jboss Data Virtualization | Lớp trừu tượng cơ sở dữ liệu ảo có thể được truy cập thông qua một giao diện tiêu chuẩn  Tích hợp với JBoss Data Grid để cung cấp bộ nhớ cache cơ sở dữ liệu trong bộ nhớ để tăng tốc tra cứu và hiệu suất tổng thể  Cung cấp công cụ giao diện người dùng đồ họa dựa trên IDE, giúp dễ dàng sử dụng các công cụ và kỹ thuật  Tập trung vào microservices và container, giúp người dùng kết nối microservices của họ với backend lớp dữ liệu ảo đơn giản hơn | Tạo cơ sở dữ liệu ảo (VDB) để ánh xạ các nguồn dữ liệu vật lý thành các góc nhìn tích hợp  Tích hợp với Kubernetes và môi trường dựa trên container  Tích hợp với JBoss Data Grid để tăng tốc truy vấn |
| 5 | Microsoft Data Virtualization | Ảo hóa dữ liệu bằng PolyBase cho phép kết nối dữ liệu từ nhiều nguồn khác nhau và thống nhất ngôn ngữ truy vấn bằng T-SQL  Hỗ trợ cơ sở dữ liệu trong bộ nhớ bởi MS SQL server, giúp tăng hiệu suất truy vấn.  Cung cấp khả năng xử lý truy vấn thông minh, bao gồm gợi ý kho truy vấn và phân tích hoạt động theo thời gian thực | Được tích hợp sẵn trên Microsoft SQL server 2022 và Microsoft Azure Data Studio, liên kết với các sản phẩm khác của MIcrosoft  Thừa hưởng khả năng truy vấn và phân tích dữ liệu từ SQL Server. |
| 6 | CData Software - Data Virtuality | Khả năng sao chép dữ liệu, tốc độ truy vấn nhanh và hiệu suất tổng thể tốt cho các tập dữ liệu lớn  Tất cả dữ liệu trên nhiều nguồn đều có thể được truy vấn theo cách chuẩn hóa bằng SQL.  Tạo mô hình dữ liệu logic từ các nguồn dữ liệu khác nhau bao gồm cơ sở dữ liệu và các nguồn dữ liệu khác như Google Analytics  Hoạt động như một máy chủ ứng dụng dựa trên Java có thể được lưu trữ trên đám mây hoặc on-premise | Kết hợp ảo dữ liệu với giải pháp tích hợp hai phương thức, cho phép người dùng tận dụng cả tiếp cận và truy cập dữ liệu trực tiếp trong một nền tảng kết nối chung |
| 7 | AtScale | Công nghệ bộ nhớ cache thích ứng giúp tối ưu hóa hiệu suất truy vấn và tăng khả năng trả về kết quả truy vấn  Hoạt động tốt với cả kho và kho dữ liệu on-premise và trên đám mây  Nhiều tùy chọn tích hợp với các công cụ Business Intelligence cho phép người dùng truy cập dữ liệu một cách an toàn nhất quán  Áp dụng học máy để tạo ra các cấu trúc dữ liệu thông minh giúp tăng tốc các truy vấn, trả về kết quả nhanh hơn và giảm chi phí truy vấn | Tích hợp với số lượng lớn các ứng dụng Business Intelligence, ngay cả bảng tính Microsoft Excel cũng được hỗ trợ  Tập trung vào việc giúp tạo ra giá trị từ dữ liệu và đầu tư vào công nghệ AI/ML |
| 8 | SaS Federation Server | Cải thiện hiệu suất truy cập dữ liệu bằng cách đẩy dữ liệu xuống cơ sở dữ liệu, pooling kết nối, caching quyền truy cập dữ liệu  Cải thiện hiệu suất bằng công nghệ In-Database thông qua chạy mã SAS DS2 và mô hình điểm số  Cho phép duy trì, cấu hình và giám sát truy cập dữ liệu từ một điểm quản trị duy nhất trong giao diện trình duyệt web | Cải thiện hiệu suất trong cơ sở dữ liệu bằng công nghệ In-Database |
| 9 | Actifio Virtual Data Pipeline | Data Pipeline là nền tảng EDaaS cho phép cung cấp năng lượng cho cả thiết bị phần cứng on-premise hoặc trên đám mây.  Không chỉ ảo hóa dữ liệu mà còn giúp thu thập dữ liệu ứng dụng bằng các phương pháp lấy dữ liệu phù hợp để tạo ra hình ảnh ảo nhất quán tại các thời điểm  Thiết bị phần cứng cung cấp khả năng tăng cường bổ sung, bao gồm kiểm soát truy cập và mã hóa cho các tổ chức có yêu cầu tuân thủ nghiêm ngặt | Tính năng rollback và khôi phục tích hợp có thể giúp khắc phục sự cố với dữ liệu không chính xác, bị mất hoặc bị hỏng |
| 10 | Stone Bond Enterprise Enabler | Hoạt động và lưu trữ giải pháp trên AWS với môi trường hybrid hoặc multi-cloud  Sử dụng công nghệ AppComm độc quyền để kết nối với hơn 100 nguồn lưu trữ dữ liệu  Tính năng Data Catalog và Data Lineage nhằm tìm kiếm dữ liệu và xác định các quy tắc của tập dữ liệu trong giao diện mới.  Nền tảng phát triển Data Masking và Data Obfuscation để bảo đảm tính an toàn dữ liệu  Self Service Web Portal phục vụ riêng cho các Data Steward và Data Scientist | Là công cụ hoàn toàn tập trung vào ảo hóa dữ liệu  Tính năng dành riêng cho các Data Steward và Data Scientist |

Trong bối cảnh kinh doanh hiện đại, ảo hóa dữ liệu đã trở thành một yếu tố trung tâm không chỉ trong việc cải thiện hiệu suất quản lý dữ liệu mà còn trong việc nâng cao khả năng ra quyết định dựa trên dữ liệu. Các công cụ ảo hóa dữ liệu, nhờ khả năng tích hợp và xử lý linh hoạt, đã định hình lại cách thức các tổ chức tiếp cận và sử dụng thông tin. Đánh giá này sẽ cung cấp một cái nhìn tổng quan và phân tích sâu rộng về hiệu suất, tính linh hoạt, khả năng tích hợp và mức độ phức tạp quản lý của một số công cụ ảo hóa dữ liệu hàng đầu.

Hiệu suất của công cụ ảo hóa dữ liệu là yếu tố thiết yếu, đặc biệt khi xử lý và truy xuất các tập dữ liệu lớn trong thời gian thực. Denodo Platform và TIBCO Data Virtualization đều cung cấp khả năng xử lý song song và tối ưu hóa truy vấn, làm giảm đáng kể tải lưu lượng mạng và thời gian chờ cho người dùng cuối. Nổi bật trong số này, Denodo sử dụng các thuật toán tối ưu hóa truy vấn tiên tiến để cải thiện hiệu suất tổng thể và giảm độ trễ.

Tính linh hoạt trong ảo hóa dữ liệu đo lường khả năng của công cụ để thích ứng với các loại hình và nguồn dữ liệu đa dạng. Công cụ như IBM Data Virtualization và Red Hat JBoss Data Virtualization cho thấy khả năng thích ứng cao, hỗ trợ cả dữ liệu có cấu trúc và không cấu trúc từ môi trường on-premises đến cloud. Điểm nổi bật của IBM Data Virtualization là khả năng tích hợp mượt mà với hệ sinh thái đám mây và công nghệ AI, cung cấp khả năng mở rộng và tùy chỉnh cao.

Khả năng tích hợp đánh giá cách các công cụ ảo hóa dữ liệu có thể dễ dàng kết nối và làm việc với các hệ thống IT hiện có. Stone Bond Enterprise Enabler và Microsoft Data Virtualization đều nổi trội với khả năng tích hợp mạnh mẽ, hỗ trợ rộng rãi cho nhiều nguồn dữ liệu và nền tảng. Microsoft Data Virtualization, với sự hỗ trợ của PolyBase, cho phép tích hợp và truy vấn dữ liệu từ nhiều nguồn khác nhau bằng T-SQL, cải thiện khả năng truy cập và tương tác dữ liệu.

Mức độ phức tạp quản lý liên quan đến dễ dàng trong việc triển khai, cấu hình và bảo trì các công cụ ảo hóa dữ liệu. Denodo cung cấp giao diện người dùng trực quan, giảm thiểu độ phức tạp và tăng cường trải nghiệm người dùng, trong khi IBM Data Virtualization tập trung vào giải pháp hội tụ, cung cấp một nền tảng quản lý dữ liệu tổng thể qua môi trường đám mây của IBM.

Trong khi mỗi công cụ có những điểm mạnh riêng biệt, lựa chọn công cụ phù hợp phải căn cứ vào yêu cầu cụ thể về hiệu suất, tính linh hoạt, khả năng tích hợp và mức độ phức tạp trong quản lý của mỗi tổ chức. Tương lai của ảo hóa dữ liệu sẽ tiếp tục hướng tới cải thiện các tính năng tích hợp AI và máy học, đẩy mạnh tự động hóa và tối ưu hóa quy trình, từ đó mang lại lợi ích cạnh tranh và hiệu quả kinh doanh cao hơn cho các tổ chức. Phát triển này không chỉ giúp tăng cường khả năng phân tích và quản lý dữ liệu mà còn mở rộng khả năng sáng tạo và đổi mới trong kinh doanh thông minh.

## 2.2. Kiến trúc hệ thống ảo hóa dữ liệu

Từ các công cụ đã được trình bày ở trên, có thể thấy rằng mỗi công cụ thuộc một hãng công nghệ có ưu điểm và lợi thế riêng cho việc giải quyết bài toán của doanh nghiệp. Cũng chính vì vậy, các kiến trúc hệ thống ảo hóa dữ liệu giữa các công cụ có những đặc điểm riêng biệt được xây dựng nhằm đáp ứng nhu cầu và nghiệp vụ của tổ chức.

Nhóm nghiên cứu đã tham khảo các kiến trúc từ các công nghệ đã đề cập ở trên như::

* Kiến trúc ảo hóa của Denodo Platform

A screenshot of a computer

Description automatically generated

Hình 1: Kiến trúc ảo hóa dữ liệu của Denodo Platform

*Nguồn:* (Data Virtualization with Denodo)

* Kiến trúc ảo hóa của Tibco

A screenshot of a computer

Description automatically generated

Hình 2: Kiến trúc ảo hóa dữ liệu của Tibco

*Nguồn:* (Tibco, 2021)

* Kiến trúc ảo hóa của RedHat

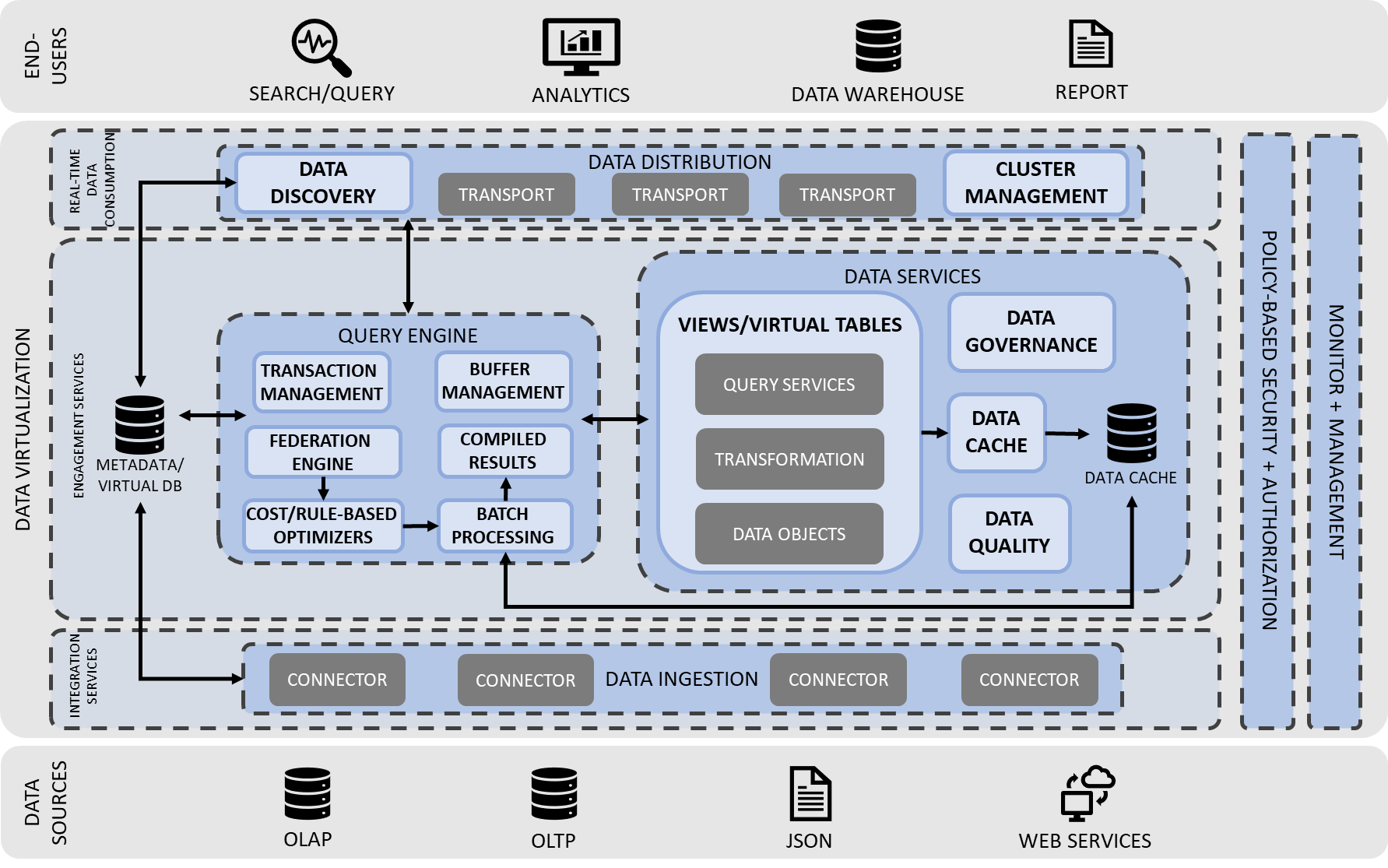
A screenshot of a computer

Description automatically generated

Hình 3: Kiến trúc ảo hóa dữ liệu của RedHat

*Nguồn:* (RedHat, 2020)

Từ việc nghiên cứu các kiến trúc của công nghệ trên, nhóm nghiên cứu đề xuất một kiến trúc tổng thể hệ thống ảo hóa dữ liệu như hình 2 với mục đích làm rõ ràng và chi tiết các thành phần cũng như mối quan hệ giữa các thành phần.



Hình 2: Đề xuất hệ thống kiến trúc ảo hóa dữ liệu

Trong kiến trúc trên, như đối với hầu hết tất cả các hệ thống ảo hóa dữ liệu, hệ thống được chia làm ba tầng với tầng thứ nhất là các nguồn dữ liệu cung cấp nguồn tài nguyên cần thiết cho nhu cầu doanh nghiệp, tầng thứ hai đảm nhiệm công việc tích hợp theo hướng ảo hóa dữ liệu và tầng cuối cùng là các công cụ và người dùng phân tích và xây dựng báo cáo. Chi tiết hơn về kiến trúc tầng ảo hóa, tầng này gồm ba phân vùng chức năng chính là Integration Services, Engagement Services và Real-time Data Consumption. Các phân vụ sẽ được mô tả chi tiết dưới đây:

Tại phân vùng đầu tiên là Integration Servcies, đây là nơi thể hiện rõ nhất đặc trưng của tích hợp dữ liệu trong hệ thống. Phân vùng này thường thực hiện quá trình thu thập và chuẩn bị dữ liệu từ các nguồn khác nhau nằm ở tầng thứ nhất của hệ thống. Integration Servcies cung cấp các connector hỗ trợ cho hệ thống ảo hóa dữ liệu kết nối được với các nguồn dữ liệu cần thiết và đồng thời đưa các dữ liệu vào một hệ thống lưu trữ ảo dưới dạng các lớp logic để xử lý các trừu tượng hoặc các chế độ xem về kết nối và kết hợp dữ liệu.

Tiếp theo, các dữ liệu từ hệ thống lưu trữ ảo sẽ tiếp tục tiến đến phân vùng Engagement Services, nơi mà bắt đầu diễn ra từ quá trình truy vấn dữ liệu và các dữ liệu sẽ được biến đổi và hiện thị theo yêu cầu của người dùng. Khi một câu lệnh được thực thi, Query Engine sẽ tiếp nhận câu lệnh đó, thực hiện phân tích cú pháp và xác định ngữ nghĩa sao cho phù hợp với dữ liệu trong hệ thống lưu trữ ảo. Câu lệnh sau đó có thể sẽ được viết lại đơn giản hơn và được lập thành một kế hoạch truy vấn logic và tối ưu hóa. Sau khi đã có kế hoạch cụ thể, batch processing được tiến hành nhằm xử lý truy vấn và trả về kết quả cho người dùng. Dữ liệu từ các câu truy vấn này có thể đến từ hệ thống lưu trữ ảo hoặc đến từ Data Cache nơi lưu trữ lịch sử các bảng ảo và câu lệnh. Sau đó, người dùng có thể tiếp tục sử dụng nhiều các câu lệnh khác nhau lên bảng ảo đã được tạo trước đó, điều này tạo thành các cây bảng ảo nơi mà các bảng được liên kết mật thiết với nhau. Trong Engagement Services, hệ thống còn cung cấp nhiều chức năng khác như quản lý bộ đệm nhằm kiểm soát xử lý và điều phối câu lệnh hoặc các dịch vụ dữ liệu như Data Governance, Data Cache,...

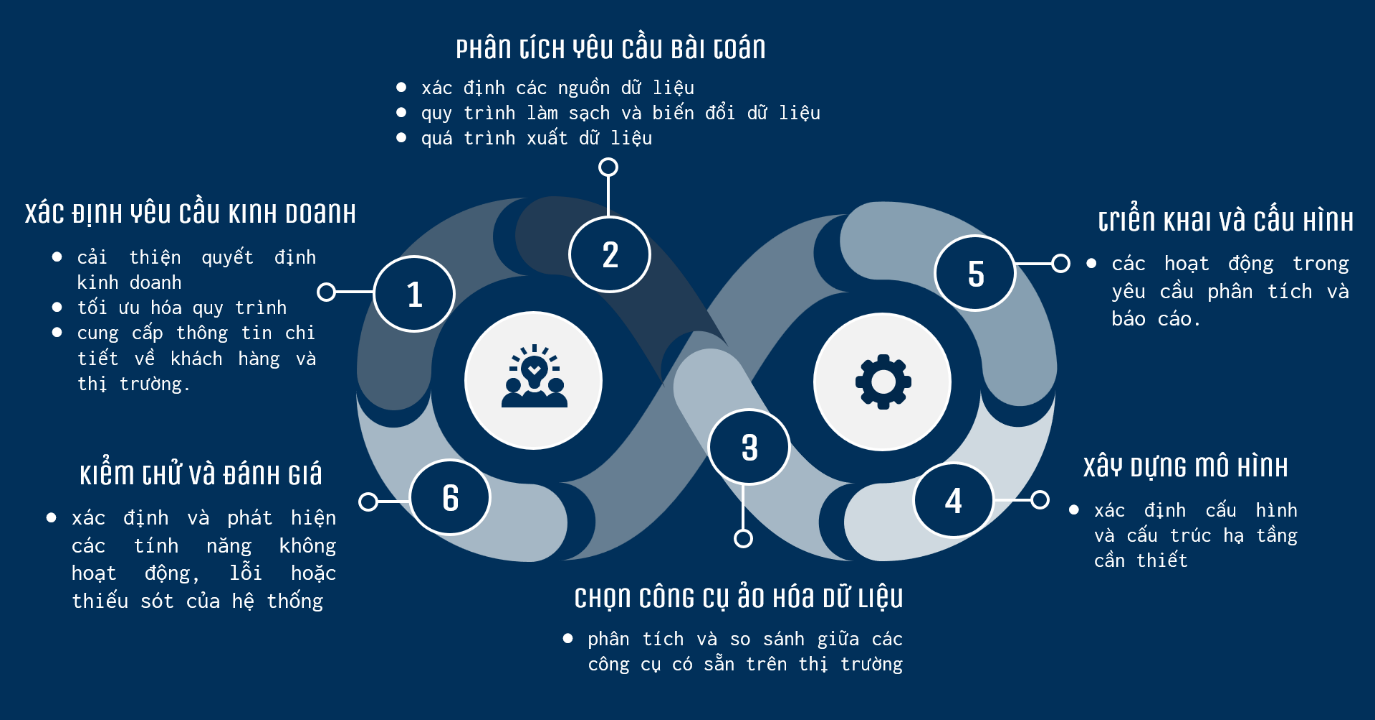
Cuối cùng là phân vùng Real-time Data Consumption với khả năng xử lý cấu hình người dùng và hỗ trợ kết nối người dùng với tầng ảo hóa dữ liệu. Phân vùng này cho phép hệ thống xây dựng tính năng khai thác và khám phá dữ liệu, góp phần giúp người dùng cuối có cơ hội được tiếp cận trực tiếp với dữ liệu theo hướng trừu tượng hóa và dễ sử dụng. Đồng thời, tương tự với phân vùng đầu tiên, nó cũng bao gồm các transport tạo liên kết giữa hai tầng để người dùng có thể truy vấn , tương tác và sử dụng dữ liệu trong thời gian thực.

Bên cạnh đó, để đảm bảo hệ thống luôn hoạt động ổn định và có tính quy củ, các tính năng như quản trị hệ thống, bảo mật và quản lý người dùng được xây dựng trong môi trường hệ thống xuyên suốt quá trình ảo hóa dữ liệu. Các tính năng trên cũng là các tính năng rất cần thiết trong mọi hệ thống tích hợp và quản trị dữ liệu mà doanh nghiệp cần phải cân nhắc khi lựa chọn một hệ thống phục vụ cho nghiệp vụ của mình.

## 2.3. Quy trình triển khai ứng dụng ảo hóa hỗ trợ ra quyết định trong kinh doanh thông minh

Trong môi trường kinh doanh hiện đại, việc nắm bắt và phân tích dữ liệu đã trở thành một phần không thể thiếu để đưa ra quyết định thông minh và hiệu quả. Tuy nhiên, để thực hiện điều này một cách hiệu quả, việc ảo hóa dữ liệu đóng vai trò quan trọng, đặc biệt là trong việc hỗ trợ quyết định kinh doanh thông minh và nhanh chóng.

Để hiểu rõ hơn về cách thức ảo hóa dữ liệu hỗ trợ cho việc đưa ra quyết định kinh doanh thông minh và nhanh chóng, ta cần xác định được quy trình thực hiện ảo hóa dữ liệu cho kinh doanh thông minh trước khi tiến hành các bước triển khai. Quy trình này không chỉ là một chuỗi các bước kỹ thuật, mà còn là một quá trình logic và tổ chức, đảm bảo rằng dữ liệu được thu thập, xử lý và phân tích một cách hợp lý để cung cấp thông tin hữu ích và đáng tin cậy cho quyết định kinh doanh.



Hình 3: Quy trình thực hiện ảo hóa dữ liệu trong kinh doanh thông minh

Quy trình thực hiện ảo hóa dữ liệu cho kinh doanh thông minh gồm các bước sau:

1. *Xác định bài toán kinh doanh*

Trong quá trình phát triển và triển khai ứng dụng, việc xác định mục tiêu và nhu cầu cụ thể đóng vai trò quan trọng trong việc định hình chiến lược và quy trình triển khai. Thu thập và hiểu biết rõ ràng về các yêu cầu của dự án không chỉ là một phần không thể thiếu trong quản lý dự án mà còn là yếu tố chính để đảm bảo thành công của dự án.

Để đạt được điều này, việc xác định các yêu cầu cụ thể mà ứng dụng ảo hóa dữ liệu sẽ hỗ trợ là không thể phủ nhận. Các yêu cầu này có thể bao gồm việc cải thiện quyết định kinh doanh, tối ưu hóa quy trình hoặc cung cấp thông tin chi tiết về khách hàng và thị trường. Để đáp ứng được các yêu cầu này, nhà phát triển cần phải sở hữu kiến thức sâu rộng về hoạt động kinh doanh và khả năng xây dựng chiến lược kinh doanh cũng như thực hiện các biện pháp thích hợp.

Điều này còn đặt ra một thách thức đối với nhà phân tích, đòi hỏi họ phải có khả năng nắm bắt và đánh giá chính xác các yêu cầu từ các bên liên quan, từ đó xây dựng được một kế hoạch triển khai hiệu quả. Đồng thời, sự hiểu biết về cách thức hoạt động của doanh nghiệp và khả năng đánh giá tác động của các giải pháp ảo hóa dữ liệu vào quy trình kinh doanh là không thể phụ nhận.

Trong bối cảnh này, việc tạo ra một môi trường làm việc mà các bộ phận chức năng có thể cùng nhau định hình và đưa ra các yêu cầu cụ thể cho dự án trở nên vô cùng quan trọng. Bằng cách này, những yêu cầu cần thiết có thể được xác định và đảm bảo rằng các giải pháp ảo hóa dữ liệu được xây dựng đáp ứng đúng mục tiêu và nhu cầu của tổ chức.

1. *Phân tích yêu cầu bài toán*

Phân tích yêu cầu bài toán và xác định các yêu cầu cần thiết cho một dự án hệ thống kinh doanh thông minh là một quá trình phức tạp và cần phải được thực hiện một cách cẩn thận. Quá trình này không chỉ tập trung vào việc đặt ra câu hỏi "tại sao" và "như thế nào", mà còn đòi hỏi sự hiểu biết sâu sắc về các yếu tố đầu vào, quy trình xử lý, và đầu ra của bài toán.

Yếu tố đầu tiên trong việc phân tích yêu cầu là hiểu rõ và xác định các nguồn dữ liệu. Trước khi bắt đầu quá trình tích hợp, việc thu thập dữ liệu ban đầu và tiến hành các hoạt động làm quen với dữ liệu là cần thiết. Điều này giúp tạo ra một cơ sở vững chắc cho quá trình tiếp theo. Tiếp theo, việc xác định vấn đề về chất lượng dữ liệu và khám phá thêm về dữ liệu là quan trọng. Mỗi nguồn dữ liệu có thể có định dạng và cấu trúc khác nhau, do đó việc chuyển đổi dữ liệu là bước không thể thiếu để tạo ra một tập dữ liệu đồng nhất và dễ dàng truy vấn.

Yếu tố thứ hai là quy trình làm sạch và biến đổi dữ liệu. Các nhà phân tích cần phải xác định, xây dựng và liệt kê kế hoạch các bước để xử lý dữ liệu một cách hiệu quả. Quá trình này bao gồm việc loại bỏ dữ liệu không chính xác, điền các giá trị thiếu, và biến đổi dữ liệu thành định dạng phù hợp để phục vụ cho mục đích phân tích.

Cuối cùng, sau khi các bước trên đã được hoàn thành, quá trình xuất dữ liệu là bước cuối cùng của giai đoạn này. Xuất dữ liệu không chỉ đơn thuần là việc truyền dữ liệu vào hệ thống ảo hóa dữ liệu, mà còn đòi hỏi rằng tất cả dữ liệu đã được chuyển đổi phải đạt được chuẩn hóa để dễ dàng thực hiện các công việc phân tích và tạo ra báo cáo có ý nghĩa.

1. *Chọn công cụ ảo hóa dữ liệu*

Lựa chọn công cụ ảo hóa dữ liệu là một phần không thể thiếu trong quy trình triển khai hệ thống kinh doanh thông minh. Giai đoạn này đòi hỏi sự cân nhắc kỹ lưỡng để đảm bảo rằng công cụ được chọn không chỉ đáp ứng được nhu cầu cụ thể của doanh nghiệp mà còn có khả năng tích hợp một cách mạch lạc với các công cụ Business Intelligence (BI) hiện có.

Việc lựa chọn một công cụ ảo hóa dữ liệu thích hợp bắt đầu bằng việc phân tích và so sánh giữa các công cụ có sẵn trên thị trường. Cần phải tập trung vào việc tìm ra công cụ có điểm mạnh phù hợp nhất với yêu cầu cụ thể của dự án. Điều này có thể bao gồm tính linh hoạt trong việc tương thích với các nguồn dữ liệu đa dạng, khả năng tích hợp với các hệ thống BI hiện có, và khả năng mở rộng để phục vụ cho quy trình phân tích và báo cáo của doanh nghiệp.

Trong quá trình đánh giá các công cụ ảo hóa dữ liệu, các nhà quản lý dự án cần xem xét kỹ lưỡng các tiêu chí quan trọng như tính linh hoạt, hiệu suất và khả năng mở rộng. Tính linh hoạt là yếu tố quan trọng để đảm bảo rằng công cụ có thể thích nghi và mở rộng theo sự phát triển của doanh nghiệp. Hiệu suất là yếu tố khác cần được xem xét để đảm bảo rằng công cụ có thể xử lý và phân tích dữ liệu một cách hiệu quả, đồng thời đáp ứng được yêu cầu về thời gian thực hiện. Khả năng mở rộng cũng đóng vai trò quan trọng, đặc biệt khi doanh nghiệp phát triển và mở rộng quy mô hoạt động của mình. Việc lựa chọn công cụ ảo hóa dữ liệu là một quyết định chiến lược, đòi hỏi sự cân nhắc kỹ lưỡng và đánh giá tỉ mỉ để đảm bảo rằng doanh nghiệp có thể tận dụng tối đa các lợi ích từ quá trình ảo hóa dữ liệu và phân tích thông tin một cách hiệu quả nhất.

1. *Xây dựng mô hình ảo hóa dữ liệu*

Xây dựng một mô hình ảo hóa dữ liệu là một bước quan trọng trong việc triển khai hệ thống kinh doanh thông minh, đòi hỏi sự chuyên sâu và cẩn thận từ các nhà phát triển và nhà quản lý dự án. Giai đoạn này không chỉ đòi hỏi việc xác định cấu hình và cấu trúc hạ tầng cần thiết, mà còn đòi hỏi sự hiểu biết sâu sắc về yêu cầu cụ thể của dự án.

Để có cái nhìn tổng quát và rõ ràng hơn về bài toán, quan trọng là phải xác định một cấu trúc mô hình ảo hóa dữ liệu hiệu quả. Trong quá trình này, chúng ta tạo ra một lớp trừu tượng ảo hóa dữ liệu, bao gồm các bảng ảo, nhằm định nghĩa các ánh xạ và quy trình biến đổi dữ liệu. Lớp này sẽ trở thành điểm truy cập duy nhất cho các truy vấn từ ứng dụng và người dùng, đảm bảo tính nhất quán và hiệu quả trong việc truy cập và xử lý dữ liệu.

Quá trình xây dựng mô hình ảo hóa dữ liệu đòi hỏi sự tổ chức kỹ lưỡng và kế hoạch hóa chi tiết. Trước hết, cần phải xác định các yêu cầu cụ thể của dự án và định rõ cấu trúc dữ liệu cần thiết để đáp ứng những yêu cầu đó. Sau đó, các nhà phát triển sẽ tiến hành xây dựng các bảng ảo và thiết lập các quy tắc ánh xạ và biến đổi dữ liệu. Quá trình này đòi hỏi sự linh hoạt để đảm bảo rằng mô hình có thể mở rộng và thích nghi với sự phát triển của doanh nghiệp.

Cuối cùng, quan trọng là kiểm tra và đánh giá mô hình để đảm bảo rằng nó đáp ứng được các yêu cầu kỹ thuật và chức năng. Điều này bao gồm việc thử nghiệm tính nhất quán và hiệu suất của mô hình trong các điều kiện thực tế và đảm bảo rằng nó đáp ứng được mọi yêu cầu từ người dùng cuối và các bộ phận liên quan khác trong tổ chức. Quá trình xây dựng một mô hình ảo hóa dữ liệu là một phần không thể thiếu trong việc triển khai hệ thống kinh doanh thông minh và đòi hỏi sự tổ chức kỹ lưỡng và sự chuyên sâu từ các nhà phát triển và nhà quản lý dự án.

1. *Triển khai và cấu hình*

Triển khai và cấu hình nền tảng ảo hóa dữ liệu là một phần quan trọng trong quy trình triển khai hệ thống kinh doanh thông minh, đòi hỏi sự cẩn thận và chuyên nghiệp từ các nhà quản lý dự án và chuyên gia công nghệ.

Sau khi các bước chuẩn bị đã được lên kế hoạch và hoàn thiện, việc triển khai và cấu hình nền tảng ảo hóa là bước quan trọng để đánh giá tính khả thi của ứng dụng. Quy trình này bao gồm một loạt các hoạt động từ việc kết nối với các nguồn dữ liệu, tích hợp dữ liệu từ các nguồn khác nhau, đến việc biến đổi dữ liệu để phù hợp với yêu cầu phân tích và báo cáo.

Đầu tiên, việc kết nối với các nguồn dữ liệu là bước quan trọng để đảm bảo rằng hệ thống có thể truy cập và lấy dữ liệu một cách hiệu quả từ các nguồn khác nhau. Sau đó, việc tích hợp dữ liệu từ các nguồn này vào nền tảng ảo hóa là một quá trình phức tạp, đòi hỏi sự hiểu biết sâu sắc về cấu trúc và định dạng dữ liệu.

Tiếp theo, việc biến đổi dữ liệu là một phần không thể thiếu để đảm bảo rằng dữ liệu được chuẩn hóa và chuẩn bị sẵn sàng cho quá trình phân tích và báo cáo. Quá trình này có thể bao gồm việc loại bỏ dữ liệu không chính xác, điền các giá trị thiếu, hoặc thậm chí biến đổi cấu trúc dữ liệu để đáp ứng các yêu cầu cụ thể.

Cuối cùng, việc thiết lập bảo mật và tạo báo cáo là những bước quan trọng để đảm bảo rằng dữ liệu được bảo vệ và chỉ có những người được ủy quyền mới có thể truy cập và sử dụng. Đồng thời, việc tạo ra các báo cáo phù hợp với yêu cầu bài toán giúp đưa ra những thông tin quan trọng và hữu ích để hỗ trợ quyết định kinh doanh.

Việc triển khai và cấu hình nền tảng ảo hóa dữ liệu là một quy trình phức tạp và đòi hỏi sự chuyên nghiệp và kỹ năng từ các nhà quản lý dự án và chuyên gia công nghệ. Quy trình này không chỉ đảm bảo tính khả thi của ứng dụng mà còn đóng vai trò quan trọng trong việc cung cấp thông tin hữu ích và đáng tin cậy cho quyết định kinh doanh.

1. *Kiểm thử và đánh giá*

Trong giai đoạn cuối cùng của dự án, việc thực hiện các bài kiểm thử đóng vai trò then chốt để đảm bảo tính chính xác và hiệu suất của hệ thống. Mục tiêu chính của quá trình kiểm thử là xác định và phát hiện các tính năng không hoạt động, lỗi hoặc thiếu sót của hệ thống trong quá trình chạy quy trình. Điều này giúp nhóm phát triển hiểu rõ hơn về những điểm yếu và cơ hội cải thiện trước khi hệ thống được triển khai một cách rộng rãi.

Bằng cách thực hiện các bài kiểm thử chi tiết và tỉ mỉ, chúng ta có thể đảm bảo rằng hệ thống hoạt động như mong đợi và đáp ứng được các yêu cầu cụ thể. Đồng thời, việc đánh giá kết quả từ các bài kiểm thử cung cấp cho nhóm phát triển thông tin quan trọng về hiệu quả của hệ thống.

Trong quá trình đánh giá, chúng ta cần phân tích kỹ lưỡng các dữ liệu thu thập được từ các bài kiểm thử và tìm ra những điểm mạnh và điểm yếu của hệ thống. Dựa trên những phân tích này, chúng ta có thể đưa ra các đánh giá chính xác về hiệu quả của hệ thống và đề xuất các điều chỉnh tham số cũng như cải thiện quy trình để tối ưu hóa hiệu suất và độ tin cậy của hệ thống.

Ngoài ra, quá trình kiểm thử và đánh giá cũng đóng vai trò quan trọng trong việc xác nhận rằng hệ thống đáp ứng được các tiêu chí về bảo mật và an toàn dữ liệu. Điều này đặc biệt quan trọng trong môi trường ngày nay, khi mà sự bảo vệ thông tin cá nhân và dữ liệu trở thành ưu tiên hàng đầu.

Kiểm thử và đánh giá không chỉ là bước cuối cùng mà còn là một phần không thể thiếu trong quy trình phát triển hệ thống. Việc thực hiện chúng một cách kỹ lưỡng và chuyên nghiệp không chỉ đảm bảo tính chính xác và hiệu suất của hệ thống mà còn giúp tối ưu hóa quy trình và đảm bảo an toàn thông tin.

## 2.3. Tổng kết Chương 2

Chương này dành để khám phá sâu hơn về ứng dụng của công nghệ ảo hóa dữ liệu trong việc tối ưu hóa quy trình tích hợp dữ liệu, cung cấp quyết định dựa trên thông tin thời gian thực và tăng cường khả năng phân tích dữ liệu. Qua việc tập trung vào các khía cạnh này, có thể hiểu rõ hơn về cách công nghệ này đóng vai trò quan trọng trong việc xây dựng một cơ sở dữ liệu linh hoạt và hiệu quả.

Trong quá trình nghiên cứu, nhóm nghiên cứu đã thấy rằng sự kết hợp giữa các công nghệ ảo hóa dữ liệu và các công cụ phân tích thông minh mang lại lợi ích lớn cho việc đánh giá và tối ưu hóa hệ thống dữ liệu. Không chỉ giúp tạo ra một cái nhìn tổng quan về khả năng và lợi ích của nền tảng công nghệ này mà còn phân tích sâu hơn về các ưu điểm và hạn chế cụ thể của từng công cụ

Nhóm nghiên cứu cũng đã đi sâu vào việc xác định các bước cụ thể trong quy trình xây dựng và triển khai giải pháp ảo hóa dữ liệu. Điều này không chỉ làm bước chuẩn bị cho việc áp dụng mô hình vào các bài toán thực tế mà còn tạo ra tiền đề cho việc nghiên cứu và phát triển tiếp theo của nền tảng này.

# CHƯƠNG 3: ỨNG DỤNG CÔNG NGHỆ ẢO HÓA DỮ LIỆU TRONG KINH DOANH THÔNG MINH

*Chương ba, sẽ tập trung vào việc khám phá chi tiết quá trình ứng dụng công nghệ ảo hóa dữ liệu vào kinh doanh thông minh cho công ty Maven Market với bài toán cụ thể là mở rộng thị trường kinh doanh dựa trên các nguồn dữ liệu được lưu trữ trong các cơ sở dữ liệu khác nhau của công ty. Mục đích của chương này là cung cấp một hướng dẫn toàn diện và chi tiết về việc thực nghiệm công nghệ ảo hóa dữ liệu ứng dụng trong kinh doanh thông minh theo công nghệ và kỹ thuật đã được trình bày trong chương 2.*

## 3.1. Mô tả bài toán

Trong bối cảnh thị trường kinh doanh hiện đại, việc lựa chọn vị trí địa lý để mở rộng kinh doanh không chỉ là một quyết định mang tính hành chính mà còn là một chiến lược cốt lõi, quyết định đến sự phát triển của một doanh nghiệp. Maven Market, một chuỗi cửa hàng bán lẻ, hiện đang đứng trước nhu cầu cần mở rộng quy mô hoạt động tại các bang khác nhau của Mỹ. Để đảm bảo sự thành công của quá trình mở rộng này, việc phân tích kỹ lưỡng các yếu tố như dân số và thu nhập của người dân tại các bang được xem là chìa khóa quan trọng.

Về thực trạng hiện tại của doanh nghiệp, Maven Market đã xây dựng được một vị thế vững chắc tại một số bang chủ chốt với mạng lưới cửa hàng trải rộng ở các khu vực đô thị lớn. Doanh nghiệp này nổi tiếng với việc cung cấp các sản phẩm chất lượng cao và dịch vụ khách hàng xuất sắc, điều này đã giúp thu hút và giữ chân một lượng khách hàng trung thành đáng kể. Tuy nhiên, để duy trì đà tăng trưởng và nắm bắt các cơ hội mới, Maven Market cần phải mở rộng sang các thị trường mới.

Một trong những thách thức chính mà Maven Market đang đối mặt là sự cạnh tranh ngày càng gia tăng từ cả các chuỗi bán lẻ lớn và các cửa hàng nhỏ lẻ địa phương. Đồng thời, sự thay đổi trong xu hướng tiêu dùng và tình hình kinh tế không ổn định cũng đặt ra nhiều áp lực lên kế hoạch mở rộng. Do đó, việc lựa chọn các vị trí mới một cách chiến lược dựa trên phân tích dữ liệu trở nên càng quan trọng hơn bao giờ hết.

## 3.2. Phân tích bài toán

### 3.2.1. Phân tích nguồn dữ liệu

#### a. Dữ liệu lưu trữ trên Microsoft SQL Server của Maven Market

Cơ sở dữ liệu Maven Market - quản lý dữ liệu cho thị trường Maven, được thiết kế để theo dõi và quản lý thông tin liên quan đến các sản phẩm, cửa hàng, khách hàng, và các giao dịch. *Cơ sở dữ liệu Maven Market được lưu trữ dưới dạng cơ sở dữ liệu quan hệ nội bộ trên Microsoft SQL Server* *của Maven Market*, cơ sở dữ liệu này là cơ sở dữ liệu phục vụ cho hoạt động tác nghiệp và điều hành hàng ngày của tổ chức.

#### b. Dữ liệu lưu trữ trên BigQuery của Google Cloud

BigQuery của Google Cloud lưu trữ nhiều bộ dữ liệu mã nguồn mở và cung cấp khả năng truy cập và truy vấn dữ liệu. Đồng thời, vì là một phần của Google Cloud, nó cũng cho phép người dùng lưu trữ dữ liệu của mình trên đám mây của Google và thực hiện các hoạt động tương tác với các bộ dữ liệu khác như truy vấn hoặc kết hợp bảng. Trong BigQuery của Google Cloud lưu có dữ liệu về số lượng người dân Mỹ theo các tiểu bang được cục điều tra dân số của Mỹ báo cáo cập nhật theo định kỳ.

#### c. Dữ liệu lưu trữ trên Cloud của BEA

Dữ liệu bao gồm các chỉ số như GDP, thu nhập cá nhân, chi tiêu tiêu dùng, và đầu tư, …được cung cấp bởi Cục Phân tích Kinh tế Hoa Kỳ (BEA). Các dữ liệu này được công bố công khai và có thể truy cập thông qua API của BEA, giúp dễ dàng lấy dữ liệu theo các tiêu chí cụ thể.

### 3.2.2. Xử lý dữ liệu

Sau khi tích hợp với ba nguồn dữ liệu là Maven Market Database, Google Cloud và BEA Web, việc làm sạch và xử lý dữ liệu là bước quan trọng trong quá trình ra quyết định của một hệ thống kinh doanh thông minh. Tại giai đoạn này, các bước biến đổi và chuẩn hóa dữ liệu sẽ được thực hiện bởi hệ thống ảo hóa dữ liệu trên các bảng ảo nhằm phục vụ cho người dùng nguồn nguyên liệu chính xác và đầy đủ cho yêu cầu báo cáo của họ.

Đối với bộ dữ liệu Maven Market, hệ thống cần đảm bảo sau khi tích hợp, các dữ liệu và các bảng vẫn giữ được tính toàn vẹn. Điều này bao gồm các thuộc tính, điều kiện và mối quan hệ giữa các bảng. Bên cạnh đó, các kiểu dữ liệu của từng bản ghi phải được tương thích với bộ dữ liệu tài nguồn ban đầu nhằm phòng tránh các đánh giá sai và các lỗi truy vấn từ người dùng cuối.

Bên cạnh đó, nhóm nghiên cứu sẽ cần phải thực hiện nhiều bước xử lý hơn với hai bộ dữ liệu còn lại. Các bước loại bỏ dữ liệu trùng lặp và xử lý giá trị khuyết thiếu sẽ được thực hiện trong hai bộ dữ liệu này nhằm tối ưu hóa khả năng đánh giá của người dùng đối với dữ liệu. Các thuộc tính và khóa trong bảng cũng nên cần được xác định trong hai bộ dữ liệu để giúp góp phần cho các hoạt động đối chiếu hoặc kết hợp bảng. Hơn nữa, bộ dữ liệu sau khi được tích hợp cũng phải đảm bảo được các khả năng thao tác trên dữ liệu bao gồm truy vấn, cập nhật, xóa và tạo bảng ảo.

Sau khi thực hiện làm sạch dữ liệu, dữ liệu cũng có thể cần được biến đổi để phù hợp với mục đích phân tích. Điều này có thể bao gồm việc chuyển đổi định dạng dữ liệu, tạo các thuộc tính mới từ dữ liệu hiện có, hoặc thực hiện các phép toán thống kê để tóm tắt dữ liệu. Trong bài toán này, nhóm có thực hiện thiết lập thêm các thuộc tính mới từ ba nguồn dữ liệu bao gồm thuộc tính số lượng cư dân, thuộc tính nguồn thu nhập theo đầu người và thuộc tính số lượng cửa hàng. Các thuộc tính này được xác định theo từng bang ở Hoa Kỳ và lần lượt được biến đổi từ các nguồn Google Cloud, BEA Web và Maven Market Database.

Cuối cùng, dữ liệu sẽ được đưa đến cho người dùng phân tích để rút ra những thông tin có giá trị. Phân tích có thể được thực hiện thông qua các phương pháp thống kê, học máy, hoặc trực quan hóa dữ liệu. Mục tiêu là để xác định các mẫu, xu hướng, và mối quan hệ trong dữ liệu, từ đó hỗ trợ việc đưa ra quyết định và giải quyết bài toán. Vì dựa theo yêu cầu của bài toán và có một vài sự hạn chế nên nhóm nghiên cứu lựa chọn việc phân tích dữ liệu bằng cách hiển thị dữ liệu bằng công cụ Power BI như là một đại diện cho bước xử lý cuối cùng này, từ đó hỗ trợ cho quá trình ra quyết định kinh doanh

Quy trình xử lý dữ liệu áp dụng công nghệ ảo hóa dữ liệu giúp tổ chức tích hợp và quản lý dữ liệu từ nhiều nguồn khác nhau một cách hiệu quả và linh hoạt. Việc này không chỉ giảm bớt gánh nặng quản lý dữ liệu mà còn tối ưu hóa khả năng phân tích và ra quyết định dựa trên dữ liệu. Kết quả là, Maven Market có thể nâng cao hiệu quả kinh doanh và tăng cường lợi thế cạnh tranh trên thị trường.

### 3.3.3. Phân tích kết quả đầu ra

Kết quả đầu ra sẽ tập trung vào những điểm sau:

1. *Phân tích Dân số và Tăng trưởng Dân số*: Kết quả phân tích dân số và tăng trưởng dân số ở từng bang sẽ cung cấp cái nhìn sâu sắc về quy mô và tiềm năng thị trường tiêu dùng. Các bang với dân số đông và tăng trưởng nhanh chóng có thể được xem xét là các mục tiêu tiềm năng cho việc mở rộng cửa hàng.

2. *Phân tích Thu nhập Bình quân*: Đánh giá thu nhập bình quân của người dân tại các bang để hiểu rõ khả năng chi tiêu và xu hướng mua sắm. Những khu vực có thu nhập bình quân cao có thể hướng tới việc tiêu dùng các sản phẩm và dịch vụ cao cấp, cung cấp cơ hội lớn cho Maven Market.

3. *Xác định Vị trí Chiến lược*: Dựa trên phân tích dữ liệu, xác định các vị trí tiềm năng cho việc mở rộng cửa hàng mới. Ưu tiên các bang có dân số đông và thu nhập bình quân cao để tối đa hóa tiềm năng kinh doanh.

4. *Đề xuất Chiến lược Mở rộng*: Dựa trên kết quả phân tích, đề xuất một chiến lược mở rộng chi tiết và linh hoạt, bao gồm các vị trí cụ thể cho các cửa hàng mới và lịch trình triển khai.

## 3.3. Lựa chọn nền tảng công nghệ

Trong chương 2 đã giới thiệu về các công nghệ ảo hóa dữ liệu. Chương3 này, nhóm nghiên cứu sẽ lựa chọn và đi sâu vào nền tảng công nghệ Denodo, bởi tính phù hợp của nó trong việc giải quyết các bài toán phân tích và xử lý dữ liệu phức tạp. Chương này sẽ tập trung vào vai trò của Denodo hỗ trợ trong việc tích hợp, quản lý và phân tích dữ liệu từ nhiều nguồn khác nhau, cung cấp một giải pháp hiệu quả cho các nhu cầu phân tích dữ liệu hiện đại.

### 3.3.1. Tổng Quan Về Công Nghệ Ảo Hóa Dữ Liệu Denodo

Ảo hóa dữ liệu (Data Virtualization) là một công nghệ tiên tiến cho phép truy cập và quản lý dữ liệu từ nhiều nguồn khác nhau mà không cần phải di chuyển hoặc sao chép dữ liệu thực tế. Công nghệ này tạo ra một lớp trừu tượng hóa giữa các nguồn dữ liệu và người dùng, cung cấp một giao diện thống nhất để truy vấn và phân tích dữ liệu từ các nguồn phân tán. Trong số các nền tảng ảo hóa dữ liệu hiện đại, Denodo đã khẳng định vị thế của mình như một giải pháp hàng đầu nhờ khả năng tích hợp mạnh mẽ, hiệu suất cao và tính linh hoạt vượt trội.

Denodo là một nền tảng ảo hóa dữ liệu toàn diện, được thiết kế để cung cấp truy cập thời gian thực và hợp nhất dữ liệu từ nhiều nguồn khác nhau, bao gồm cơ sở dữ liệu truyền thống, kho dữ liệu, hồ dữ liệu, các dịch vụ web và các nguồn dữ liệu đám mây. Denodo cho phép người dùng truy vấn dữ liệu từ các nguồn này thông qua một giao diện duy nhất mà không cần phải thực hiện các quy trình ETL (Extract, Transform, Load) phức tạp và tốn kém.

### 3.3.2. Vai Trò của Denodo trong Quản Lý Dữ Liệu

1. Hợp Nhất Dữ Liệu Thời Gian Thực: Denodo cho phép hợp nhất dữ liệu từ nhiều nguồn khác nhau trong thời gian thực, giúp các tổ chức có cái nhìn toàn diện và cập nhật về dữ liệu của mình mà không cần phải di chuyển dữ liệu vật lý.

2. Tối Ưu Hóa Quy Trình Kinh Doanh: Bằng cách cung cấp truy cập nhanh chóng và liền mạch vào dữ liệu, Denodo giúp tối ưu hóa các quy trình kinh doanh, từ đó nâng cao hiệu suất và hiệu quả hoạt động.

3. Hỗ Trợ Quyết Định Dựa Trên Dữ Liệu: Denodo cung cấp khả năng truy vấn và phân tích dữ liệu mạnh mẽ, hỗ trợ các nhà quản lý và nhà phân tích đưa ra quyết định dựa trên dữ liệu một cách chính xác và kịp thời.

4. Giảm Chi Phí và Tăng Tính Linh Hoạt: Nhờ loại bỏ nhu cầu xây dựng và duy trì các kho dữ liệu phức tạp, Denodo giúp giảm chi phí và tăng tính linh hoạt cho các tổ chức, cho phép họ nhanh chóng thích nghi với các thay đổi trong môi trường kinh doanh.

### 3.3.3. Mối Quan Hệ Giữa Denodo và Các Hệ Thống Quản Lý Dữ Liệu Khác

Denodo có mối quan hệ chặt chẽ với các hệ thống quản lý dữ liệu truyền thống như kho dữ liệu (data warehouse) và hồ dữ liệu (data lake):

1. Kho Dữ Liệu: Trong khi kho dữ liệu tập trung vào việc lưu trữ và quản lý dữ liệu có cấu trúc đã được làm sạch và chuẩn hóa, Denodo cho phép truy cập trực tiếp vào dữ liệu trong kho mà không cần phải di chuyển hoặc sao chép dữ liệu, giúp tối ưu hóa quy trình phân tích.

2. Hồ Dữ Liệu: Hồ dữ liệu lưu trữ mọi loại dữ liệu trong trạng thái nguyên bản, Denodo có thể truy cập và hợp nhất dữ liệu từ hồ dữ liệu, cung cấp một giao diện duy nhất để phân tích dữ liệu phi cấu trúc và bán cấu trúc.

3. Tích Hợp và Tăng Cường: Denodo không thay thế các hệ thống kho và hồ dữ liệu mà bổ sung và tăng cường khả năng của chúng bằng cách cung cấp truy cập thời gian thực và hợp nhất dữ liệu từ nhiều nguồn khác nhau.

### 3.3.4. Tính phù hợp của công nghệ Denodo trong việc giải quyết bài toán

Trong bối cảnh kinh doanh hiện đại, việc lựa chọn vị trí địa lý cho việc mở rộng không chỉ là quyết định mang tính hành chính mà còn là chiến lược cốt lõi quyết định đến sự thành bại của doanh nghiệp. Maven Market, chuỗi cửa hàng bán lẻ đang có nhu cầu mở rộng quy mô hoạt động tại các bang khác nhau của Mỹ, cần một quá trình phân tích dữ liệu sâu rộng để đảm bảo thành công. Việc sử dụng nền tảng Denodo cho phép tích hợp và phân tích dữ liệu từ nhiều nguồn khác nhau một cách hiệu quả và linh hoạt, làm cơ sở cho việc ra quyết định chiến lược này.

1. Tích Hợp Dữ Liệu Liền Mạch từ Nhiều Nguồn

Maven Market sở hữu một cơ sở dữ liệu quan trọng về hoạt động kinh doanh và lịch sử giao dịch của khách hàng. Tuy nhiên, để có cái nhìn toàn diện khi mở rộng, doanh nghiệp cần phải phân tích thêm dữ liệu về dân số và thu nhập từ các nguồn ngoài như Google Cloud và các bộ dữ liệu web. Denodo cho phép truy cập các nguồn dữ liệu này thông qua một lớp ảo hóa, nơi dữ liệu từ cơ sở dữ liệu nội bộ và dữ liệu ngoại vi được kết hợp một cách suôn sẻ mà không cần phải di chuyển dữ liệu.

2. Giảm Chi Phí và Độ Phức Tạp

Việc di chuyển dữ liệu giữa các hệ thống có thể tốn kém và phức tạp, đặc biệt là khi xử lý các bộ dữ liệu lớn từ nhiều nguồn khác nhau. Denodo giảm bớt nhu cầu này bằng cách cung cấp quyền truy cập trực tiếp vào dữ liệu mà không yêu cầu sao chép vật lý, giảm thiểu rủi ro và chi phí liên quan, đồng thời tăng cường bảo mật và tuân thủ dữ liệu.

3. Phân Tích Thời Gian Thực

Trong quá trình mở rộng, việc có thể truy cập và phân tích dữ liệu thời gian thực là rất quan trọng. Denodo cho phép Maven Market truy vấn dữ liệu từ các nguồn khác nhau một cách tức thì, giúp đưa ra quyết định nhanh chóng dựa trên các thông tin mới nhất về tình hình dân số và thu nhập tại các bang cụ thể.

4. Quản Lý và Bảo Mật Dữ Liệu Tập Trung

Denodo cung cấp một nền tảng để quản lý và bảo mật dữ liệu từ tất cả các nguồn một cách tập trung. Điều này giúp Maven Market đảm bảo rằng dữ liệu được sử dụng trong quá trình phân tích là chính xác, cập nhật và tuân thủ các quy định về bảo mật và bảo vệ dữ liệu.

5. Tối Ưu Hóa Việc Ra Quyết Định Dựa Trên Dữ Liệu

Cuối cùng, việc sử dụng Denodo giúp Maven Market tối ưu hóa quyết định của mình bằng cách cung cấp một cái nhìn đa chiều và phân tích sâu rộng về các yếu tố ảnh hưởng đến việc mở rộng. Từ đó, giúp đảm bảo rằng mỗi cửa hàng mới được mở dựa trên cơ sở dữ liệu vững chắc và tiềm năng tối đa để đạt được thành công.

Trong thời đại kinh doanh ngày nay, nền tảng như Denodo không chỉ là công cụ hỗ trợ mà còn là một yếu tố then chốt trong việc đảm bảo sự thành công của các chiến lược mở rộng và tăng trưởng kinh doanh. Với khả năng hợp nhất dữ liệu thời gian thực từ nhiều nguồn khác nhau, Denodo không chỉ giúp các tổ chức tối ưu hóa quy trình kinh doanh và ra quyết định mà còn giảm chi phí và tăng tính linh hoạt trong quản lý dữ liệu. Các tổ chức ngày càng phụ thuộc vào dữ liệu để đạt được lợi thế cạnh tranh, Denodo là một công cụ không thể thiếu trong bộ công cụ quản lý dữ liệu hiện đại, giúp các doanh nghiệp như Maven Market ra quyết định thông minh và phù hợp trong việc lựa chọn địa điểm kinh doanh.

## 3.4. Xây dựng mô hình ảo hóa dữ liệu

Trong bối cảnh kinh doanh ngày càng phụ thuộc vào dữ liệu, việc xây dựng một mô hình ảo hóa dữ liệu hiệu quả trở thành yếu tố then chốt để các tổ chức tối ưu hóa việc quản lý, truy cập và phân tích dữ liệu. Ảo hóa dữ liệu không chỉ cung cấp khả năng tích hợp dữ liệu từ nhiều nguồn khác nhau mà còn tạo ra một lớp trừu tượng hóa giúp đơn giản hóa quá trình truy vấn và phân tích dữ liệu. Điều này mang lại nhiều lợi ích vượt trội, từ việc tiết kiệm chi phí và thời gian đến nâng cao hiệu suất hoạt động và hỗ trợ ra quyết định chiến lược.

Mô hình ảo hóa dữ liệu là một hệ thống phức tạp, đòi hỏi sự hiểu biết sâu sắc về cấu trúc dữ liệu và các phương pháp quản lý dữ liệu hiện đại. Quá trình xây dựng mô hình này bao gồm nhiều bước chi tiết và phức tạp, từ việc xác định các nguồn dữ liệu ban đầu đến thiết kế các lớp ảo hóa và triển khai các dịch vụ dữ liệu cần thiết. Mỗi bước trong quy trình này đều đóng một vai trò quan trọng trong việc đảm bảo tính hiệu quả và thành công của mô hình ảo hóa dữ liệu.

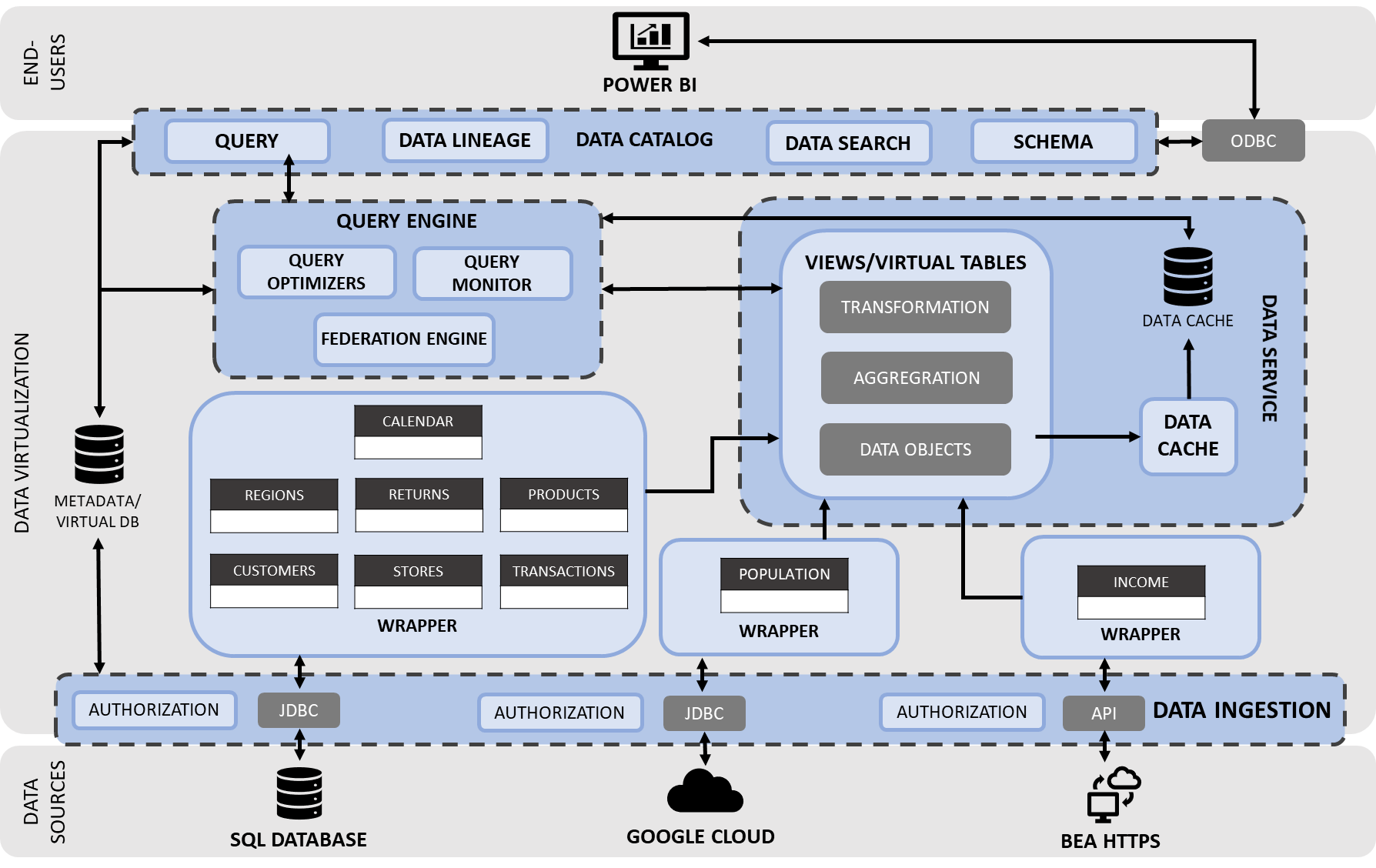
Trong phần này, nhóm nghiên cứu sẽ đi sâu vào các bước cụ thể để xây dựng một mô hình ảo hóa dữ liệu toàn diện. Đầu tiên, chúng ta sẽ bắt đầu với Xác Định Các Nguồn Dữ Liệu, nơi thảo luận về việc nhận diện và đánh giá các nguồn dữ liệu khác nhau, từ cơ sở dữ liệu truyền thống đến các nguồn dữ liệu phi cấu trúc và bán cấu trúc. Việc xác định các nguồn dữ liệu chính xác là bước nền tảng quan trọng, giúp đảm bảo rằng tất cả các dữ liệu cần thiết đều được đưa vào mô hình ảo hóa.

Tiếp theo, Thiết Kế Các Lớp Ảo Hóa Dữ Liệu, trong đó nghiên cứu này sẽ xem xét cách tổ chức và cấu trúc dữ liệu trong các lớp ảo hóa khác nhau. Thiết kế này cần phải đảm bảo tính linh hoạt và khả năng mở rộng, cho phép hệ thống dễ dàng tích hợp với các nguồn dữ liệu mới và đáp ứng các yêu cầu phân tích thay đổi theo thời gian.

Sau đó Xây Dựng Các Dịch Vụ Dữ Liệu và Thiết Lập Query Engine, bao gồm việc triển khai các dịch vụ cần thiết để truy vấn và xử lý dữ liệu một cách hiệu quả. Query Engine đóng vai trò quan trọng trong việc tối ưu hóa hiệu suất truy vấn và đảm bảo rằng các yêu cầu dữ liệu được xử lý một cách nhanh chóng và chính xác.

Cuối cùng, nghiên cứu này sẽ khám phá Data Catalog và Xác Định Người Dùng Cuối, một bước không thể thiếu trong việc quản lý và tổ chức dữ liệu. Data Catalog giúp người dùng dễ dàng tìm kiếm và truy cập dữ liệu, trong khi việc xác định người dùng cuối đảm bảo rằng mô hình ảo hóa dữ liệu được tối ưu hóa để đáp ứng nhu cầu của tất cả các nhóm người dùng trong tổ chức.

Như vậy, bài nghiên cứu có một cái nhìn tổng quan về quá trình xây dựng mô hình ảo hóa dữ liệu. Từng bước trong quy trình này không chỉ đóng góp vào việc tạo ra một hệ thống quản lý dữ liệu hiệu quả mà còn đảm bảo rằng dữ liệu được sử dụng một cách tối ưu nhất để hỗ trợ các quyết định kinh doanh chiến lược. Cụ thể, nhóm nghiên cứu sẽ đi vào từng mục chi tiết để hiểu rõ hơn về cách thức triển khai và những lợi ích cụ thể mà mô hình ảo hóa dữ liệu mang lại.



Hình 4: Mô hình ảo hóa dữ liệu

### 3.4.1. Xác định các nguồn dữ liệu

Sau khi phân tích bài toán cụ thể, ta xác định được ba nguồn dữ liệu chính đảm nhiệm khả năng cung cấp dữ liệu cho mô hình: một cơ sở dữ liệu quan hệ Microsoft SQL Server, một kho lưu trữ đám mây Google Cloud BigQuery, và một tệp dữ liệu từ một nguồn web BEA HTTP. Mỗi nguồn dữ liệu này đều có những đặc điểm và yêu cầu truy cập riêng biệt cần được xem xét khi xây dựng mô hình ảo hóa dữ liệu.

Đối với Microsoft SQL Server, để một máy chủ ảo hóa dữ liệu như Denodo có khả năng kết nối dữ liệu thì đòi hỏi người dùng sử dụng quyền truy cập như Windows authentication hoặc SQL authentication. Đây là một hình thức xác thực mà trong đó người dùng cung cấp tên người dùng và mật khẩu để truy cập vào cơ sở dữ liệu đã được chỉ định và phân quyền.

Tương tự như vậy, Google Cloud BigQuery cũng yêu cầu quá trình xác thực OAuth. OAuth cho phép người dùng chia sẻ tài nguyên đến ảo hóa dữ liệu thông qua tài khoản dịch vụ và khóa truy cập mà không cần phải chia sẻ mật khẩu.

Cuối cùng, ta sử dụng API key hoặc User key để có khả năng truy cập tới dữ liệu từ nguồn web HTTP như BEA. Đây là một chuỗi ký tự duy nhất được cấp cho mỗi người dùng và được sử dụng là một phần trong link truy cập để xác thực yêu cầu của người dùng đó.

### 3.4.2. Thiết kế các lớp ảo hóa dữ liệu

Đối với mỗi nguồn dữ liệu, chúng ta sẽ thiết kế một lớp ảo hóa dữ liệu tương ứng. Đối với cơ sở dữ liệu Microsoft SQL Server và Google Cloud BigQuery, ta sẽ sử dụng connector JDBC để tạo ra các lớp ảo hóa cho hai nguồn dữ liệu. Còn đối với tệp dữ liệu từ nguồn web BEA HTTP, ta sẽ sử dụng một API dựa trên XML để tạo ra lớp ảo hóa thứ ba đại diện ch nguồn cung cấp dữ liệu. Trong ứng dụng Denodo, các lớp ảo hóa này được gọi là các wrapper và trong bài toán này ta sẽ có ba wrapper tương ứng với ba nguồn dữ liệu. Các wrapper sẽ đảm nhiệm cung cấp một giao diện đơn giản và thống nhất về các bảng, thuộc tính để người dùng có cái nhìn toàn diện về dữ liệu trước khi sử dụng và thực hiện các thao tác biến đổi. Đồng thời, wrapper cũng đóng vai trò trong việc giữ vững khả năng kết nối ổn định và cập nhật mới nhất với các nguồn dữ liệu.

### 3.4.3. Xây dựng các dịch vụ dữ liệu và thiết lập Query Engine

Tiếp theo đó, ta sẽ xây dựng các dịch vụ dữ liệu dựa trên các lớp ảo hóa đã thiết kế. Các dịch vụ dữ liệu trong Denodo bao gồm các bảng ảo hóa định nghĩa và xử lý dữ liệu, data cache, các dịch vụ web và các giao thức hỗ trợ xuất dữ liệu từ Denodo,... Các dịch vụ dữ liệu có thể bao gồm các chức năng như truy vấn, cập nhật, xóa, và thêm dữ liệu.

Bên cạnh đó, Query Engine cũng là một thành phần quan trọng của mô hình ảo hóa dữ liệu bởi khả năng truy vấn trong thời gian thực. Nó đảm nhiệm việc xử lý các truy vấn dữ liệu từ nhiều nguồn, phân tích cú pháp và định dạng lại câu lệnh để có thể tối ưu hóa khả năng truy cập và trả về dữ liệu cho người dùng. Đồng thời, Query Engine cũng hỗ trợ quản lý tốc độ xử lý truy vấn thông qua việc xây dựng kế hoạch truy vấn cụ thể và quản lý bộ đệm để kiểm soát việc tiêu thụ bộ nhớ.

### 3.4.4. Data Catalog và xác định người dùng cuối

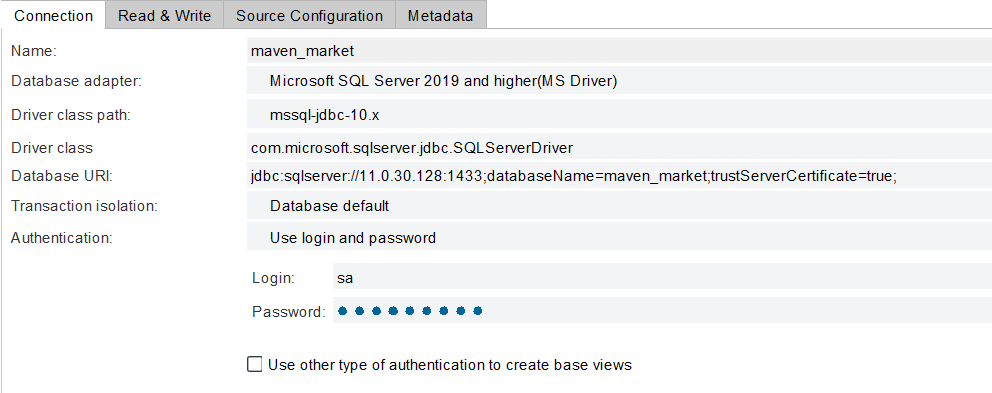
Bên cạnh các dịch vụ dữ liệu đã nêu trên, Data Catalog là một phần không thể thiếu trong các mô hình ảo hóa dữ liệu nói chung và phần mềm Denodo nói riêng. Data Catalog đóng vai trò trong việc tổ chức và quản lý tất cả các nguồn dữ liệu cũng các dịch vụ dữ liệu, đồng thời cung cấp một giao diện trực quan cho người dùng cuối để khám phá, tìm kiếm và truy vấn dữ liệu.

Trong bài toán này, người dùng cuối hay công cụ yêu cầu dữ liệu được xác định là Power BI. Data Catalog có vai trò mật thiết đối với Power BI trong trường hợp này và các công cụ phân tích dữ liệu khác. Nhờ có Data Catalog, Power BI mới có thể truy xuất được dữ liệu chuẩn xác và các mối quan hệ dữ liệu thông qua các dịch vụ dữ liệu chịu sự quản lý của Data Catalog.

## 3.5. Triển khai và cấu hình

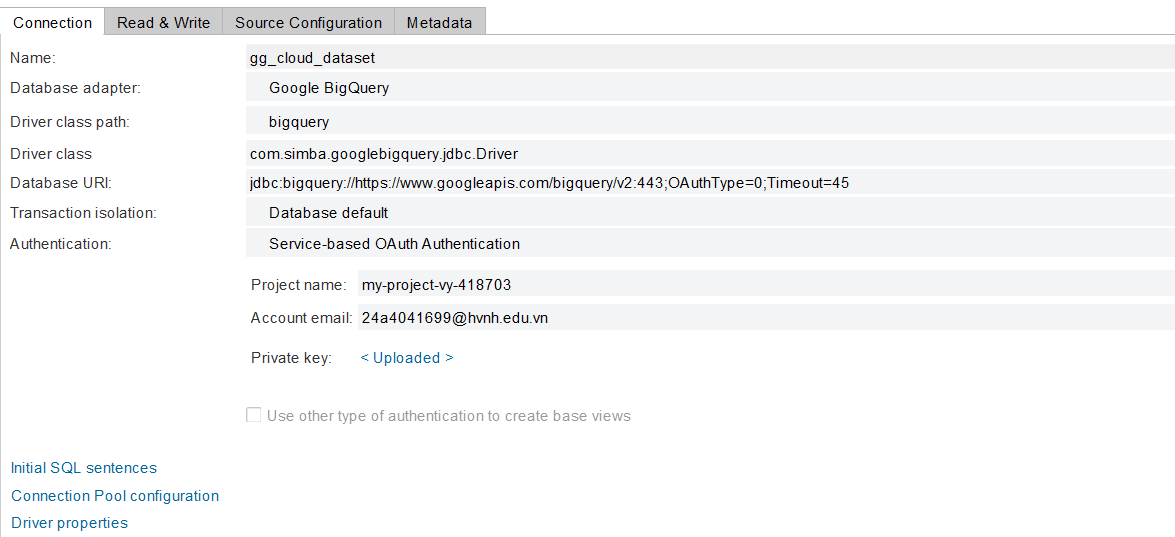
Trong việc xây dựng hệ thống quản lý dữ liệu hiện đại, việc triển khai và cấu hình đóng vai trò quan trọng trong việc đảm bảo rằng dữ liệu được truy cập và sử dụng một cách hiệu quả và an toàn nhất. Quá trình này không chỉ đòi hỏi sự hiểu biết sâu sắc về các công nghệ và công cụ cụ thể mà còn yêu cầu kỹ năng kỹ thuật và phân tích kỹ lưỡng. Trong phần này sẽ đi vào chi tiết về các bước cần thiết để triển khai và cấu hình một hệ thống quản lý dữ liệu, từ việc kết nối nguồn dữ liệu cho đến thiết lập Data Catalog và kích hoạt các dịch vụ truy vấn dữ liệu.

### 3.5.1. Kết nối nguồn dữ liệu



Hình 5: Kết nối nguồn dữ liệu database

Đối với Maven Market database, đặc điểm của nguồn dữ liệu này là cơ sở dữ liệu quan hệ được lưu trữ trong Microsoft SQL Server. Vậy nên để có thể kết nối được, ta sử dụng connector là JDBC và thiết lập Database Adapter là phiên bản Microsoft SQL Server tùy theo phiên bản trên máy chủ và ổ lưu trữ của nền tảng này. Sau đó, ta cần chỉ rõ nguồn dữ liệu mà bài toán hướng tới thông qua đường dẫn URI có cấu trúc sẵn dành cho nguồn dữ liệu bao gồm các thông tin như: tên máy chủ nhằm chỉ máy chủ đang vận hành cơ sở dữ liệu đó; tên cơ sở dữ liệu lưu trữ… Đồng thời, việc cung cấp thông tin và mật khẩu người dùng cũng cần thiết để truy cập vào cơ sở dữ liệu.



Hình 6: Kết nối nguồn dữ liệu trên đám mây

Tương tự như với Maven Market, nguồn dữ liệu thứ hai cần truy cập tới là bộ dữ liệu trên đám mây của Google và công cụ Denodo cũng yêu cầu các thông tin cần thiết để có thể kết nối với nó. Tuy nhiên, vì bài toán chỉ đặt ra yêu cầu dữ liệu từ một số bảng trên đám mây, nên ta thực hiện Create from query để có thể lựa chọn và truy cập các dữ liệu cần thiết cho bài toán:

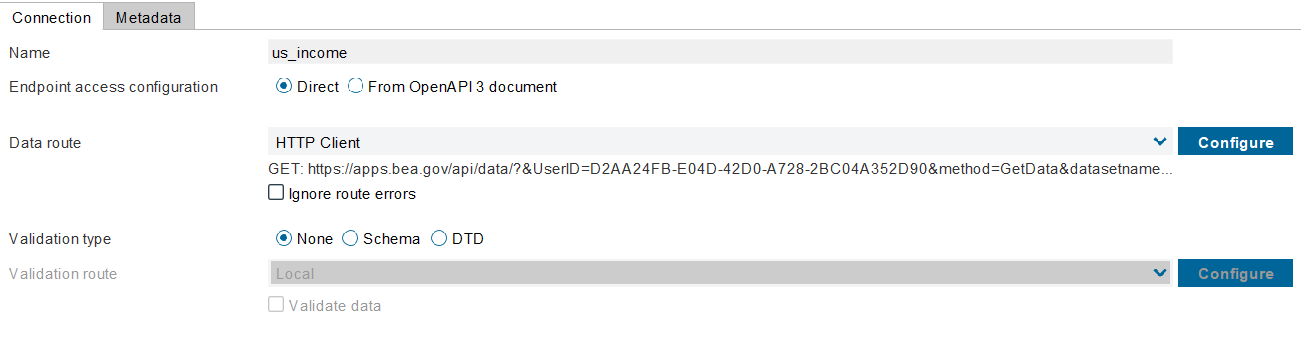
SELECT \* FROM `bigquery-public-data.census\_bureau\_usa.population\_by\_zip\_2000`

INNER JOIN `bigquery-public-data.utility\_us.zipcode\_area`

USING (zipcode)

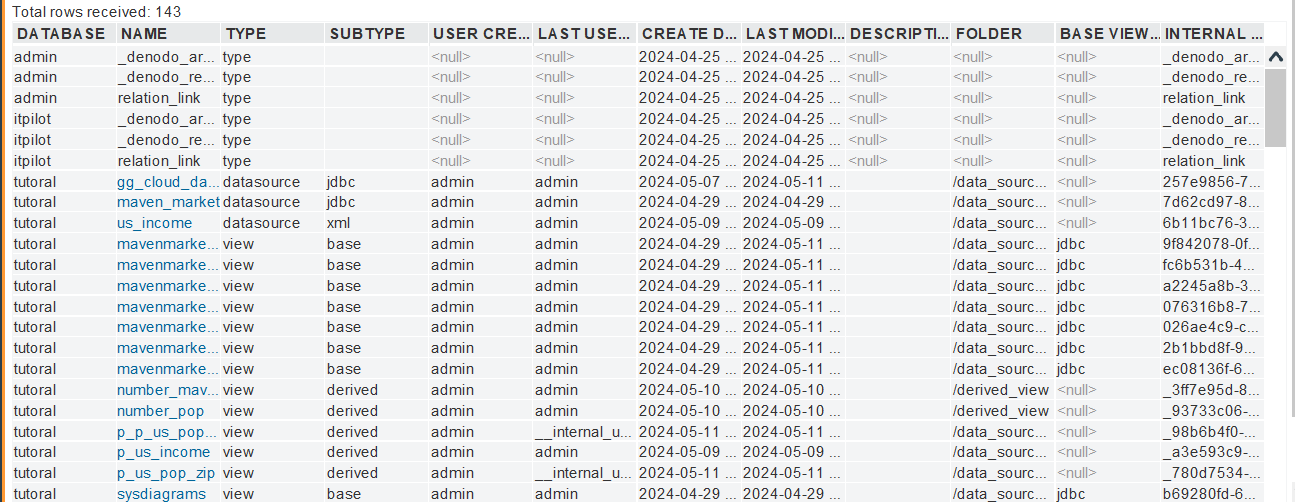
Câu truy vấn trên sẽ được thực hiện trên Google BigQuery và trả về dữ liệu của hai bảng Population by zip 2000 và Zipcode area cho Denodo data virtualization.

Tiếp đó, ở nguồn dữ liệu thứ ba, khác với hai nguồn còn lại được kết nối thông qua JDBC của nguồn dữ liệu, Denodo sử dụng connector là API để kết có thể kết nối với nguồn dữ liệu thông qua đường dẫn có đầu http dưới dạng file XML.



Hình 7: Kết nối nguồn dữ liệu web

Cuối cùng, lựa chọn Create base view để lưu lại kết nối cho mỗi lần truy cập vào máy chủ. Từ đó, Denodo hình thành một schema nơi mà trong mỗi kết nối một wrapper đại diện cho sự kết nối với nguồn dữ liệu và các base view hiển thị cho các bảng ảo được thiết lập từ nguồn dữ liệu gốc được liên kết. Schema này sẽ được lưu trữ trong metadata của hệ thống nhằm phục vụ cho người dùng thông qua Data Catalog.



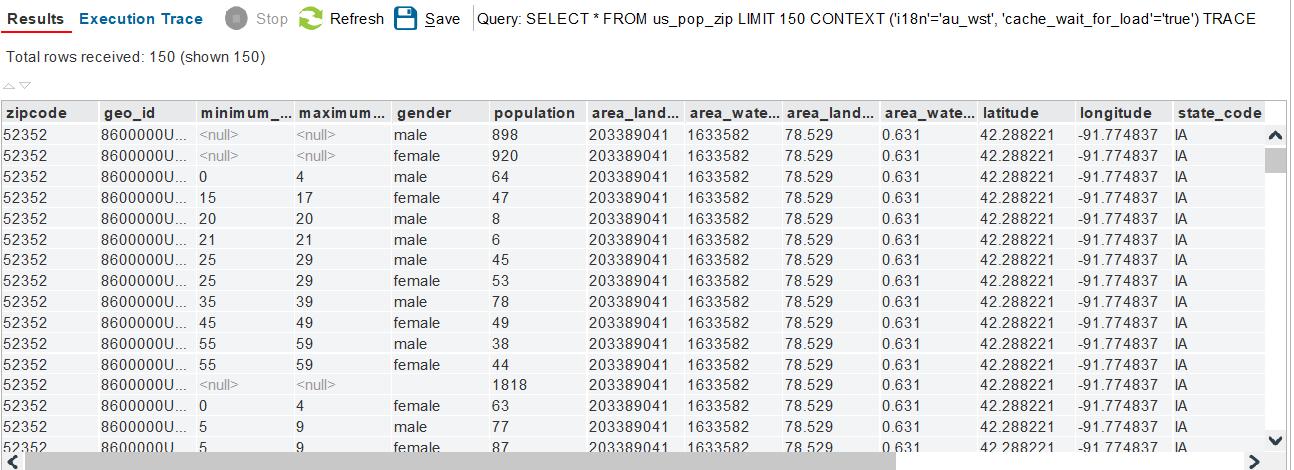
Hình 8: Tạo view

### 3.5.2. Tiền xử lý dữ liệu

Để chuẩn bị cho công cuộc phân tích và đánh giá dữ liệu, tiền xử lý dữ liệu là nhiệm vụ không kém phần quan trọng góp phần giúp kết quả có được độ chính xác cao.

Trong ba nguồn dữ liệu, Maven Market Database là nguồn dữ liệu duy nhất đã được chuẩn hóa sẵn để phục vụ cho mục đích truy cập thường xuyên. Tuy nhiên, hai nguồn còn lại vẫn còn có những định dạng chưa chuẩn và chưa phù hợp với yêu cầu của bài toán, nên ta có các bước thực hiện trên hệ thống như sau:

Đối với Google Cloud Dataset, ta có thể xem trước được bộ dữ liệu như sau:



Hình 9: Google Cloud Dataset

Ta thấy được rằng trong bộ dữ liệu gặp phải các vấn đề của dữ liệu chưa được làm sạch và điều này có thể gây nhiều sai sót trong quá trình phân tích và xây dựng báo cáo cho bài toán. Chính vì vậy, thông qua Denodo, ta tạo các view mới với mục đích tiền xử lý bộ dữ liệu này.

Ở bước đầu tiên, để có thể tạo các Schema thuận tiện cho truy vấn câu lệnh sau này, ta thiết lập khóa chính cho bộ dữ liệu này là [zipcode] và [gender] vì có thể nhận ra rằng mỗi một zipcode được lặp lại hai lần và đưa ra các chỉ số thuộc tính khác nhau qua thuộc tính Gender.

Sau đó, ta thực hiện giải quyết các giá trị khuyết thiếu và trùng lặp trong bộ dữ liệu bằng câu lệnh sau:

SELECT DISTINCT zipcode, geo\_id,

case when minimum\_age is null

then 0 else minimum\_age end as minimum\_age,

case when maximum\_age is null

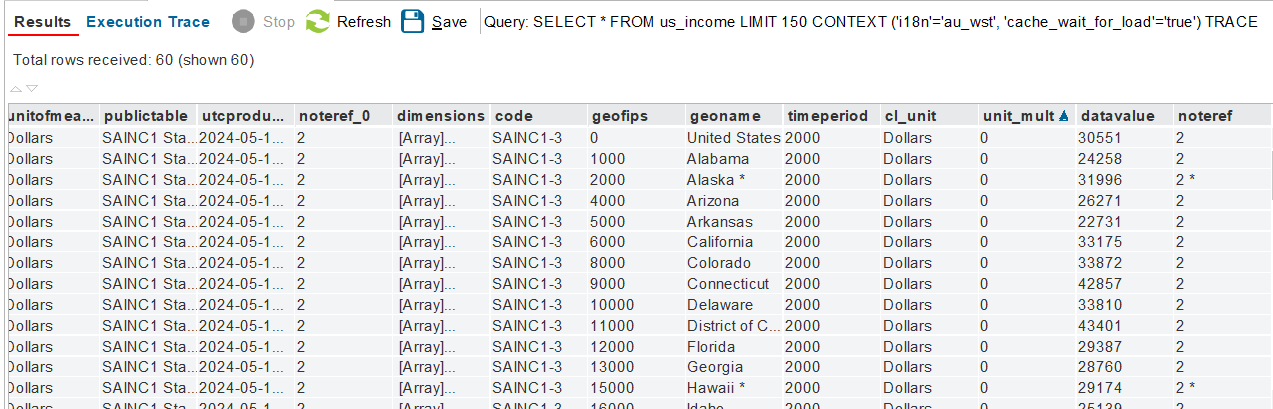
then (select max(maximum\_age) from us\_pop\_zip) else maximum\_age

end as maximum\_age,

gender, population, area\_land\_meters, area\_water\_meters, area\_land\_miles, area\_water\_miles, latitude, longitude, state\_code, state\_name, city, county, state\_fips, zipcode\_geom

FROM us\_pop\_zip

Tiếp đến với nguồn dữ liệu thứ hai là BEA HTTPS Dataset, ta tiếp tục thực hiện các bước tương tự như trên để tiền xử lý dữ liệu.



Hình 10: BEA HTTPS Dataset

Từ đó ta loại bỏ các cột không cần thiết và lựa chọn thuộc tính [geoname] làm khóa chính. Bên cạnh đó, nhận thấy rằng một vài bản ghi trong thuộc tính [geoname] có ký tự khác biệt, ta lập một bảng ảo mới với câu truy vấn để đưa các dòng về định dạng chuẩn:

SELECT CASE WHEN

SUBSTRING (geoname FROM len(geoname) FOR len(geoname)) = '\*'

THEN TRIM(SUBSTRING (geoname FROM 1 FOR len(geoname)-2))

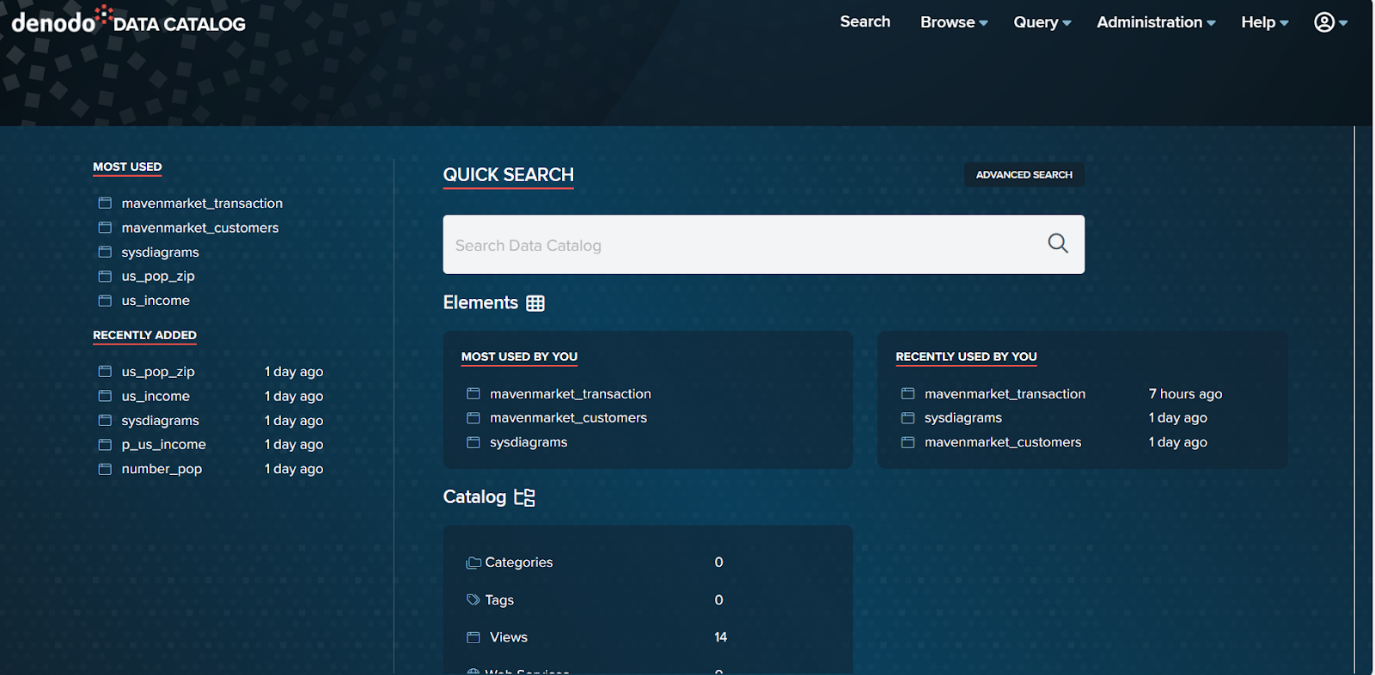
ELSE geoname

END AS geo\_name, geoname, geofips, datavalue FROM us\_income

### 3.5.3. Kích hoạt Data Catalog

Data Catalog đóng vai trò quan trọng trong việc giúp đỡ người dùng doanh nghiệp có được cái nhìn tổng thể mà đơn giản hơn về bộ dữ liệu. Nó lược bỏ các cấu trúc và yêu cầu phức tạp của hệ thống đối với dữ liệu, đồng thời cung cấp giao diện và môi trường truy vấn dữ liệu thân thiện với người dùng.

Để có thể giúp Data Catalog truy cập vào bộ dữ liệu trên cổng Virtual DataPort của Denodo, một cơ sở lưu trữ metadata ảo sẽ được yêu cầu cấp phát và đồng bộ dữ liệu. Từ đó các thông tin về dữ liệu sẽ được cung cấp đầy đủ và cập nhật mới nhất, góp phần giúp người dùng tăng cường khả năng tìm kiếm và khám phá dữ liệu.



Hình 11: Kích hoạt Data Catalog

Tùy theo tài khoản người dùng đã được thiết lập và phân quyền, các bảng ảo hay view sẽ được hiển thị và các yêu cầu truy vấn từ người dùng sẽ được thực hiện trong giao diện này.

### 3.5.4. Truy vấn dữ liệu

Để đáp ứng nhu cầu của bài toán, ta tạo bảng ảo mới truy vấn dữ liệu về các số liệu doanh thu, chi phí và lợi nhuận từ ba bảng [Transaction], [Stores] và [Products] từ bộ dữ liệu Maven Market. Ta có câu lệnh như sau:

SELECT store\_country, store\_state, sum((transaction.quantity\*products.product\_cost)) AS total\_cost, (sum((transaction.quantity\*products.product\_retail\_price))

-sum((transaction.quantity\*products.product\_cost))) AS total\_profit, sum((transaction.quantity\*products.product\_retail\_price)) AS total\_revenue, getyear(ransaction.transaction\_date) AS year

FROM ((transaction JOIN stores USING(store\_id)) JOIN products USING(product\_id)

WHERE getyear(transaction.transaction\_date) = 2000

GROUP BY stores.store\_state, stores.store\_country, getyear(transaction.transaction\_date))

Và được bộ dữ liệu như sau:



Hình 12: Bảng ảo về doanh thu, chi phí và lợi nhuận của Maven Market

Sau đó, sử dụng câu lệnh SELECT store\_state, count(store\_id) GROUP BY store\_state AS num\_store FROM stores để truy vấn số lượng cửa hàng của hệ thống siêu thị này trên các vùng và bang.



Hình 13: Kết quả số lượng cửa hàng trên các vùng và bang

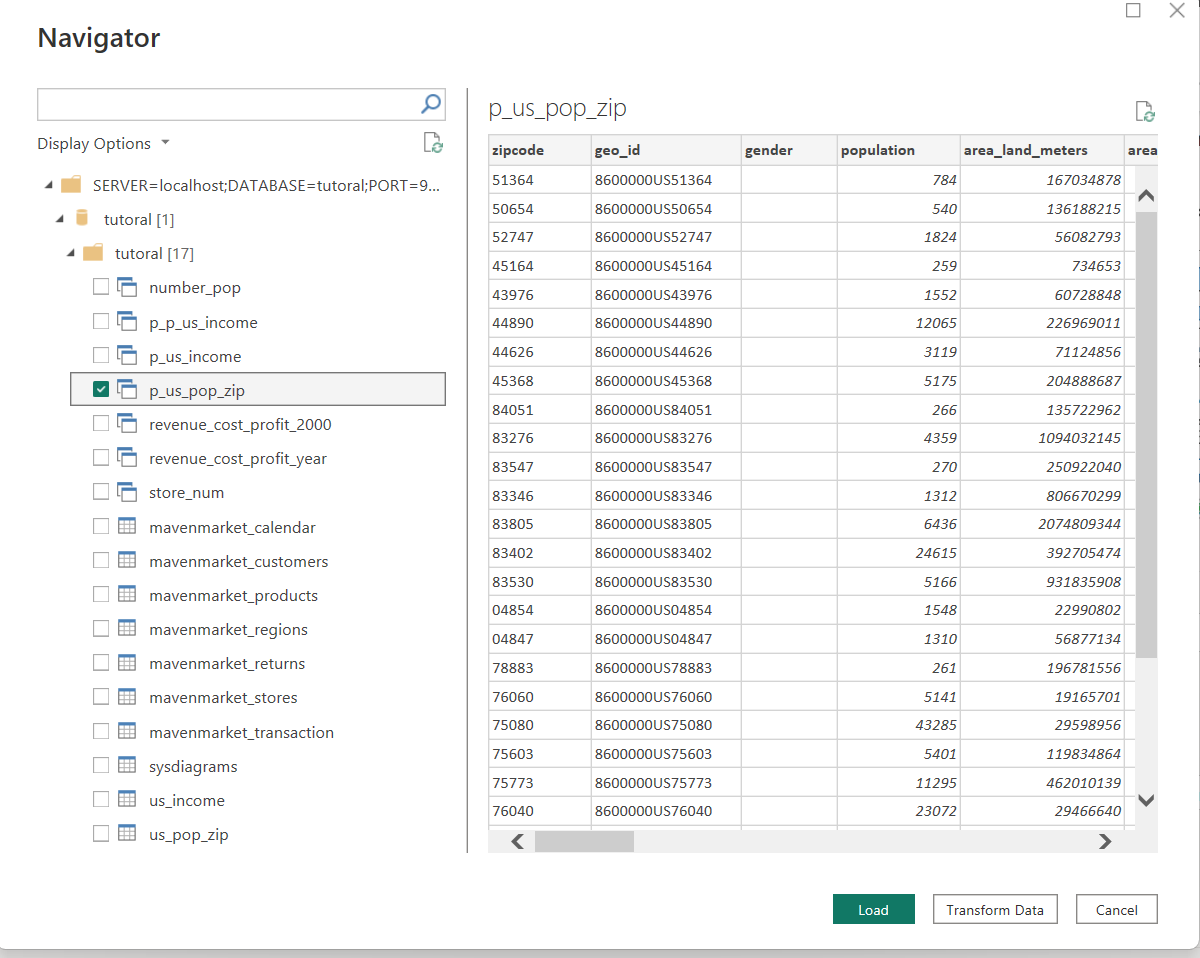
### 3.5.5. Kết nối Power BI

Để kết nối được với Power BI, Denodo sử dụng connector là ODBC đã được thiết lập sẵn trong hệ thống. Ta tải dữ liệu từ Denodo đến Power BI thông qua DSN hoặc Connection String nhằm mục đích chỉ tới máy chủ chứa hệ thống ảo hóa dữ liệu và cơ sở dữ liệu ảo đang chứa các dữ liệu cần thiết cho bài toán.

Trong bài nghiên cứu này, ta sử dụng Connection String như sau:

SERVER=localhost;DATABASE=tutoral;PORT=9996

Sau đó các bảng dữ liệu được hiển thị trong mục Navigator để phục vụ cho mục đích lựa chọn và phân tích dữ liệu. Khi đó, kết nối giữa Denodo và Power BI đã được hoàn tất và các bước tiếp theo của nghiệp vụ sẽ được thực thi theo yêu cầu của người dùng.

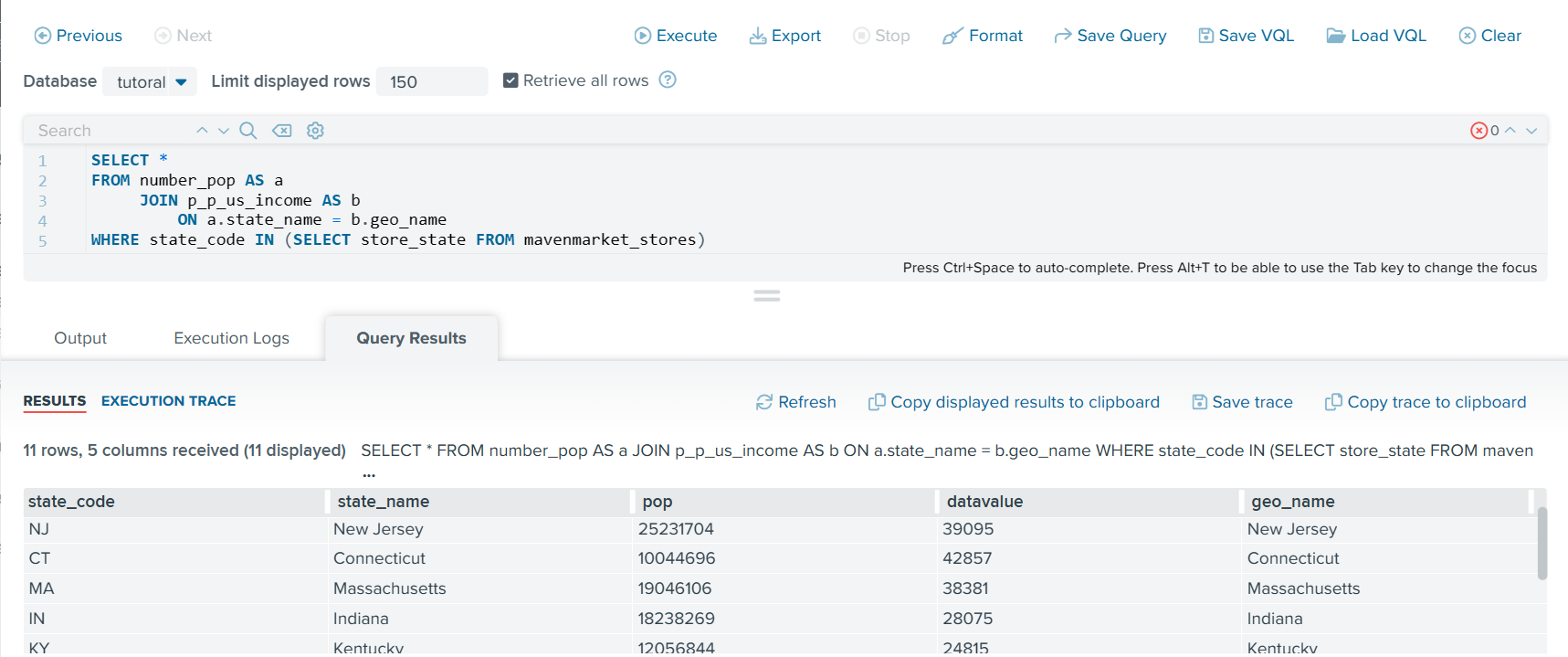


Hình 14: Lựa chọn bảng dữ liệu và phân tích

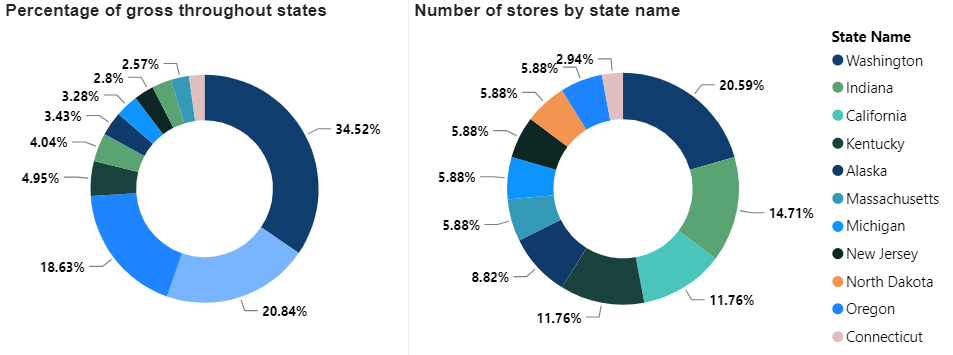
## 3.6. Thử nghiệm và kiểm tra

Như đã đề cập ở chương 2, giai đoạn thử nghiệm và kiểm tra đảm nhiệm việc đảm bảo tính hiểu quả của hệ thống thông qua khả năng hoạt động và truyền tải dữ liệu một cách đầy đủ và chính xác. Trước hết, nhóm nghiên cứu xem xét quy trình tạo báo cáo phân tích qua ứng dụng Power BI và Data Catalog để đánh giá mức độ phản ánh các yêu cầu bài toán và yêu cầu hệ thống đã được phân tích từ trước.

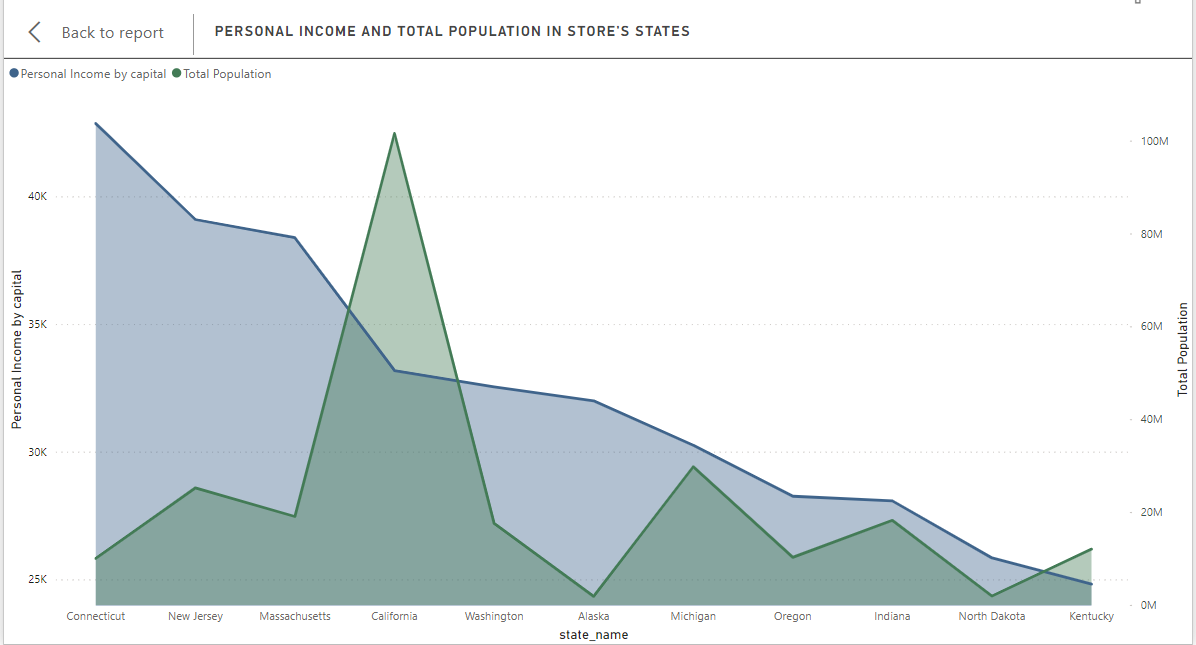
Mục đích chính của nghiên cứu này là hỗ trợ quá trình ra quyết định bằng cách cung cấp dữ liệu gần thời gian thực có thể được sử dụng bởi các bên liên quan BI trong việc hỗ trợ quá trình ra quyết định. Do đó, việc quan trọng nhất của giai đoạn này là đảm bảo đầu ra của bài toán, cũng chính là các biểu đồ giúp trực quan hóa dữ liệu hỗ trợ công cuộc nghiên cứu và phân tích.



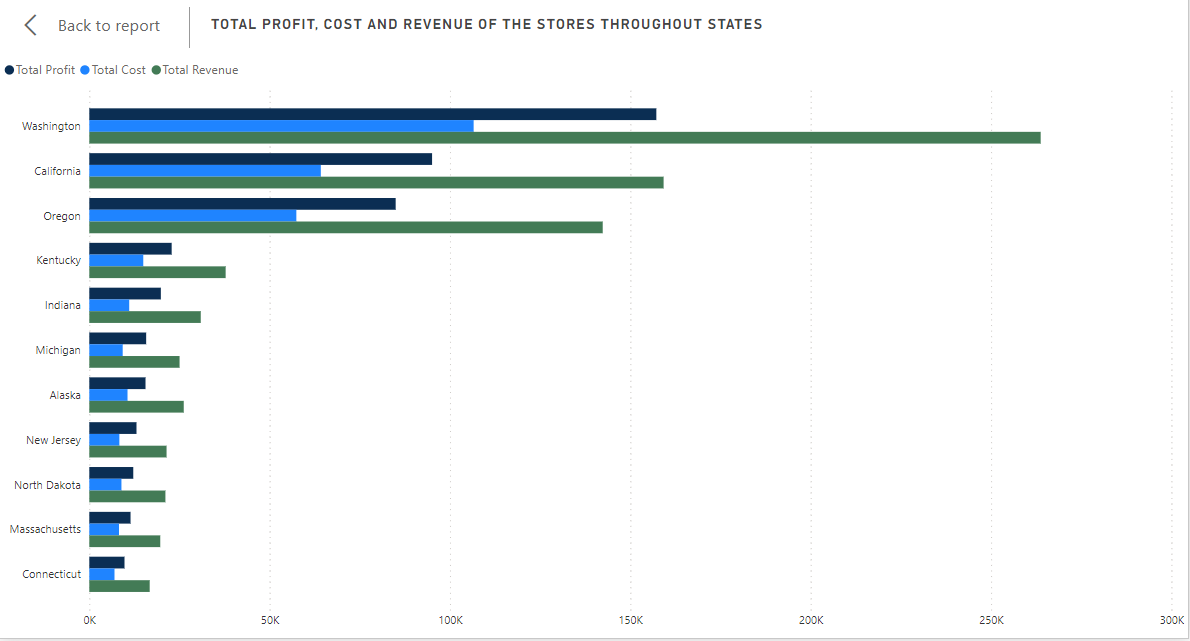
Hình 15: Tổng thu nhập người dân Mỹ ở các bang



Hình 16: Tỷ lệ thu nhập người dân và tỷ lệ số lượng cửa hàng theo các bang



Hình 17: Biểu đồ thể hiện thu nhập và dân số theo các bang



Hình 18: Biểu đồ cột thể hiện tổng lợi nhuận, chi phí và doanh thu của cửa hàng theo các bang

Có thể thấy rằng, về mặt hiển thị, mô hình ảo hóa dữ liệu có khả năng mạnh mẽ trong việc biểu diễn dữ liệu và thông tin bất kể dưới nhiều hình thức như trong một báo cáo BI hoặc dưới dạng một bảng ảo phục vụ cho truy vấn và phân tích thêm. Thực hiện các truy vấn dữ liệu thử nghiệm trên Data Catalog và đặc biệt là với Power BI, mô hình cũng biểu diễn dữ liệu dưới hình thức dễ thao tác và dễ sử dụng đối với mọi người dùng trong tổ chức doanh nghiệp.

Về tiêu chí thứ hai, mô hình trên được nhóm nghiên cứu đánh giá là mang tính linh hoạt và nhạy cảm với dữ liệu. Cụ thể hơn, mô hình đảm bảo khả năng kết nối ổn định và khả năng truy cập dữ liệu từ các báo cáo và bảng điều khiển. Đồng thời, mô hình cũng đảm bảo được tính toàn vẹn dữ liệu trong đó các dữ liệu được hiển thị chính xác và kịp thời trong cả hai công cụ phân tích. Bên cạnh đó, mô hình thể hiện tính năng thích nghi nhanh chóng với mọi sự biến đổi từ nguồn dữ liệu gốc và phản ánh sự thay đổi đó với người dùng trong thời gian ngắn. Điều này có thể coi là một lợi thế của mô hình ảo hóa dữ liệu, đặc biệt đối với các báo cáo dưới dạng dashboard đề cao tính kịp thời và biến đổi của dữ liệu.

Tiếp theo, qua thử nghiệm, mô hình cũng cung cấp người dùng khả năng ngăn ngừa và khôi phục lỗi dữ liệu thông qua tin nhắn thông báo lỗi xuất hiện trên màn hình. Đối với các lỗi truy vấn, mô hình cũng hỗ trợ người dùng sửa lỗi bằng cách xác định các vị trí lỗi trên câu truy vấn. Hơn nữa, mô hình được thiết kế với tính năng hỗ trợ truy vấn bằng các mục phân tách cú pháp truy vấn dễ sử dụng với người dùng để hạn chế nguy cơ xuất hiện lỗi.

Tuy nhiên, bên cạnh các đánh giá tích cực trên, mô hình còn một vài điểm hạn chế khiến cho người dùng gặp nhiều trở ngại trong quy trình của một hệ thống kinh doanh thông minh. Thứ nhất, mặc dù tính năng tích hợp dữ liệu được diễn ra mạnh mẽ và nhanh chóng, mô hình vẫn còn mất nhiều thời gian đối với môi trường đám mây, đặc biệt là với khối lượng dữ liệu quá lớn. Điều này đòi hỏi mô hình cần có bộ đệm có khả năng xử lý mạnh mẽ hơn và đồng thời tốc độ mạng cũng phải đủ lớn để giúp khả năng tích hợp được diễn ra nhanh chóng hơn. Điều thứ hai mà mô hình cần phải cải thiện để nâng cao hiệu suất hoạt động là việc kết nối với các công cụ người dùng cuối. Mô hình cung cấp khả năng ánh xạ với các nguồn dữ liệu để thích nghi với sự thay đổi của dữ liệu nhưng lại không cung cấp khả năng đó với các công cụ phân tích và báo cáo. Chính vì vậy mà trong quá trình thử nghiệm, việc sao chép và tải dữ liệu lên Power BI tốn khá nhiều thời gian và cũng gặp nhiều khó khăn đối với khối lượng dữ liệu lớn và phức tạp.

Như vậy, qua các tiêu chí đánh giá đã được nhóm nghiên cứu nêu ở trên, mô hình được xây dựng có nhiều ưu điểm lớn và bên cạnh đó cũng có các nhược điểm trong quá trình thực hiện phân tích và triển khai bài toán. Từ đó, đánh giá và phát triển mô hình cần là một quá trình liên tục, cần phải luôn được cân nhắc và cải thiện để đảm bảo tính hiệu quả và linh hoạt của mô hình.

## 3.7. Tổng kết Chương 3

Chương này đã mở thêm khám phá sâu rộng về việc áp dụng công nghệ ảo hóa dữ liệu trong bài toán kinh doanh thông minh cụ thể. Qua quá trình phân tích và xây dựng mô hình, chúng ta nhận thấy rằng việc tích hợp dữ liệu từ các nguồn khác nhau là bước quan trọng để đạt được hiệu quả cao và phổ rộng lớn trong việc phân tích dữ liệu. Qua đó, công nghệ Denodo đã chứng minh là một nền tảng ảo hóa dữ liệu mạnh mẽ, cho phép người dùng kết nối, truy vấn và phân tích dữ liệu một cách linh hoạt và nhanh chóng. Trong bài toán này, sự kết hợp giữa Denodo và Power BI đã tạo ra một giải pháp toàn diện, hỗ trợ quyết định kinh doanh dựa trên dữ liệu chính xác và cập nhật mới nhất.

Tuy nhiên, quá trình triển khai và cấu hình hệ thống cũng gặp phải một số thách thức và khó khăn như sau:

Thứ nhất, xác định và hiểu dữ liệu. Việc xác định các nguồn dữ liệu và hiểu rõ chúng trước khi tích hợp là một thách thức, đòi hỏi việc thu thập và làm quen với dữ liệu trước khi có thể tiến hành thiết kế lớp ảo hóa cho nguồn dữ liệu.

Thứ hai, việc kết nối dữ liệu đôi khi cũng là một trở ngại lớn trong việc ứng dụng một mô hình ảo hóa dữ liệu. Tùy vào các nền tảng công nghệ ảo hóa dữ liệu khác nhau mà có các giới hạn về hệ thống kết nối khác nhau. Không phải nền tảng nào cũng bao gồm đầy đủ hết các tính năng kết nối với nguồn dữ liệu hay công cụ phân tích báo cáo của người dùng đang cần tới. Tiếp đó, vấn đề về chất lượng dữ liệu cũng đòi hỏi người dùng phải đánh giá và xem xét kỹ lưỡng. Việc làm sạch và chuẩn hóa dữ liệu vẫn còn là một sự hạn chế khi các câu lệnh truy vấn và xử lý dữ liệu trên nền tảng có nhiều hạn chế và giới hạn. Điều này đòi hỏi người quản lý phải linh hoạt về mặt chuyên môn và am hiểu về hệ thống ảo hóa dữ liệu.

Cuối cùng là khả năng quản lý và cập nhật dữ liệu. Mặc dù ở nền tảng này có ưu điểm lớn là khả năng duy trì cho dữ liệu ảo hóa luôn được cập nhật, góp phần giúp người dùng luôn được tiếp cận với bất kỳ sự thay đổi nào dữ liệu gốc. Tuy nhiên, lợi thế đó cũng đem lại khó khăn nhất định như việc quản lý tốc độ xử lý kết nối khi gặp các trường hợp có khối lượng dữ liệu quá lớn.

Từ những khó khăn như vậy đòi hỏi sự hiểu biết sâu sắc về các công nghệ tham gia vào quy trình và khả năng giải quyết vấn đề linh hoạt của người dùng. Cuối cùng, việc thử nghiệm và kiểm tra đã cung cấp cái nhìn thực tế về hiệu quả của nền tảng, đồng thời mở ra hướng đi mới cho việc ứng dụng công nghệ ảo hóa dữ liệu vào công đoạn tích hợp dữ liệu trong lĩnh vực kinh doanh thông minh.

# KẾT LUẬN

Trải qua quá trình nghiên cứu và phân tích các khía cạnh liên quan đến ảo hóa dữ liệu và kinh doanh thông minh, bài viết nghiên cứu này đã hoàn thành các mục tiêu đặt ra với những điểm nhấn quan trọng trong từng chương. Chương 1 đã thiết lập các cơ sở lý luận vững chắc về ảo hóa dữ liệu và kinh doanh thông minh để có cái nhìn tổng quan quát về khái niệm, lợi ích của hai lĩnh vực này. Việc hiểu rõ các khía cạnh lý thuyết là tiền đề quan trọng để áp dụng vào thực tiễn một cách hiệu quả.

Chương 2 đi sâu vào các công nghệ và quy trình khai thác chức năng ảo hóa dữ liệu trong kinh doanh thông minh. Qua việc phân tích các công nghệ tiên tiến và quy trình triển khai cụ, chương này nhấn mạnh vai trò thiết yếu của ảo hóa dữ liệu trong việc tối ưu hóa các hoạt động kinh doanh thông minh; từ đó, không chỉ nâng cao hiệu suất mà còn cải thiện khả năng ra quyết định dựa trên dữ liệu thật.

Chương 3 trình bày nghiên cứu ứng dụng thực tiễn của công nghệ ảo hóa dữ liệu trong kinh doanh thông minh. Thông qua việc áp dụng công nghệ cụ thể, nghiên cứu cho thấy ảo hóa dữ liệu không chỉ tối ưu hóa hiệu suất mà còn nâng cao năng lực cạnh tranh trên thị trường của doanh nghiệp.

Nghiên cứu này chứng minh rằng ảo hóa dữ liệu là một công cụ mạnh mẽ và cần thiết trong kinh doanh thông minh. Sự kết hợp giữa các lý thuyết sâu rộng và ứng dụng thực tế đã tạo nên nền tảng vững chắc cho những nghiên cứu tiếp theo cũng như mang lại lợi ích thiết thực cho doanh nghiệp. Bằng cách tận dụng dữ liệu và áp dụng phương pháp ảo hóa dữ liệu một cách hiệu quả, doanh nghiệp có thể tối ưu hóa các quyết định kinh doanh, tạo ra sự cạnh tranh và đạt được hiệu quả kinh doanh tối ưu.

# TÀI LIỆU THAM KHẢO

Actifio. (không năm xuất bản). How It Works. In Actifio.

AtScale. (không năm xuất bản). *About AtScale*. Retrieved April 10, 2024, from AtScale: https://www.atscale.com/about/

Blazeclan. (không năm xuất bản). *Top Benefits of Data Virtualization for Your Business.* Retrieved 2024, from Blazeclan: https://blazeclan.com/blog/top-benefits-of-data-virtualization-for-your-business/

Calder, B. (2020). *Google enters agreement to acquire Actifio*. Retrieved April 10, 2024, from Google Cloud: https://cloud.google.com/blog/products/storage-data-transfer/google-enters-agreement-to-acquire-actifio

Carlos Andrés Tavera Romero, Jesús Hamilton Ortiz, Osamah Ibrahim Khalaf, Andrea Ríos Prado. (2021). Business Intelligence: Business Evolution after Industry 4.0. *Sustainability, 13*(18). doi:10.3390/su131810026

cdata. (không năm xuất bản). Retrieved March 10, 2024, from cdata: https://www.cdata.com/solutions/dv/

datanami. (2024). *CData Software Acquires Data Virtuality to Modernize Data Virtualization for the Enterprise.* Retrieved April 10, 2024, from datanami: https://www.datanami.com/this-just-in/cdata-software-acquires-data-virtuality-to-modernize-data-virtualization-for-the-enterprise/

Denodo. (không năm xuất bản). *Data Virtualization with Denodo*. Retrieved from twentylogo-new: https://twenty20sys.com/data-virtualization-with-denodo/

Denodo. (không năm xuất bản). *Denodo Platform - Subscription Bundles*. Retrieved April 10, 2024, from Denodo Community: https://community.denodo.com/docs/html/browse/8.0/en/platform/installation/appendix/denodo\_platform\_feature\_packs/denodo\_platform\_feature\_packs#ai-recommendations-featurepack

Denodo. (không năm xuất bản). *Our Company*. Retrieved March 10, 2024, from Denodo: https://www.denodo.com/en/about-us/our-company

Denodo. (không năm xuất bản). *Who we are*. Retrieved March 10, 2024, from Denodo: https://www.denodo.com/en/about-us/who-we-are

Dilmegani, C. (2024, Feb 14). *What is Data Virtualization? Benefits, Case Studies & Top Tools [2024].* Retrieved 2024, from AIMultiple Research: https://research.aimultiple.com/data-virtualization/

Duarte, F. (2023). *Amount of Data Created Daily (2024).* Retrieved December 22, 2023, from Exploding Topics: https://explodingtopics.com/blog/data-generated-per-day

Earley, S. (2016). Data Virtualization and Digital Agility. *IT Professional, 18*(5), 70-72. doi:10.1109/MITP.2016.95

Efraim Turban, Jay E. Aronson, Ting-Peng Liang. (2005). *Decision support systems and intelligent systems.* Prentice Hall.

Forrester. (không năm xuất bản). Forrester Study: The Total Economic Impact™ of Data Virtualization Using The Denodo Platform.

Gartner. (2022, March 22 ). *Gartner Magic Quadrant for Analytics and Business Intelligence Platforms.* Retrieved 2024, from Gartner: https://www.gartner.com/en/documents/4012759

IBM. (không năm xuất bản). *IBM Data Virtualization*. Retrieved March 10, 2024, from IBM: https://www.ibm.com/products/watson-query

Idemudia, J. (2022, Nov 8). *How Cloud Computing Is Influencing Business Intelligence: Part 1.* Retrieved 2024, from Forbes: https://www.forbes.com/sites/forbestechcouncil/2022/11/08/how-cloud-computing-is-influencing-business-intelligence-part-1/?sh=56a46d201c72

Joshi, S. (2021, August 25). *Data Virtualization 101: Why Businesses Need to Virtualize Data.* Retrieved 2024, from G2: https://www.g2.com/articles/data-virtualization

Kerner, S. M. (2020). *Top Data Virtualization Tools.* Retrieved 2024, from Datamation: https://www.datamation.com/big-data/top-data-virtualization-tools/#chart

Lans, R. F. (2012). Deploying Data Virtualization in Business Intelligence Systems. In *Data Virtualization for Business Intelligence Systems* (pp. 147-176). doi:10.1016/b978-0-12-394425-2.00007-1

Manoj Muniswamaiah, Tilak Agerwala và Charles Tappert. (2019). Data virtualization for decision making in big data. *International Journal of Software Engineering & Applications, 10*(5), 45-53. doi:10.5121/ijsea.2019.10503

Microsoft. (không năm xuất bản). *Data virtualization with PolyBase in SQL Server*. Retrieved March 10, 2024, from Microsoft Learn: https://learn.microsoft.com/en-us/sql/relational-databases/polybase/polybase-guide?view=sql-server-ver16

Pullokkaran, L. J. (2013). *Analysis of data virtualization & enterprise data standardization in business intelligence.* Cambridge: Massachusetts Institute of Technology.

RedHat. (2020). *Chapter 11. Data Virtualization Architecture.* Retrieved from Red Hat Customer Portal: https://access.redhat.com/documentation/en-us/red\_hat\_integration/2020-q2/html/data\_virtualization\_reference/architecture

RedHat. (không năm xuất bản). *Getting Started Guide*. Retrieved March 10, 2024, from RedHat Customer Portal: https://access.redhat.com/documentation/en-us/red\_hat\_jboss\_data\_virtualization/6.4/html-single/getting\_started\_guide/index

Saravanan Prasadh, Sowbhagya. (không năm xuất bản). *Data Virtualization with Denodo.* Retrieved April 10, 2024, from https://twenty20sys.com/data-virtualization-with-denodo/

SaS. (không năm xuất bản). *SAS Federation Server*. Retrieved April 10, 2024, from https://support.sas.com/en/software/federation-server-support.html

SaS. (không năm xuất bản). *SaS Help Center*. Retrieved April 10, 2024, from https://documentation.sas.com/doc/en/pgmsascdc/9.4\_3.5/whatsnew/p0kdf9lkzi1g6jn1xm0kxwx06otz.htm

SCHULWOLF, D. (2019). *Flexible, More Usable Data through Intelligent Data Virtualization.* Retrieved April 10, 2024, from AtScale: https://www.atscale.com/blog/flexible-more-usable-data-through-intelligent-data-virtualization/

stone bond. (không năm xuất bản). Retrieved April 10, 2024, from stone bond: https://stonebond.com/enterprise-enabler/

the Red Hat JBoss Middleware team. (2013). *Red Hat Turns Big Data into Actionable Information with JBoss Data Virtualization.* Retrieved March 10, 2024, from RedHat: https://www.redhat.com/en/blog/red-hat-turns-big-data-into-actionable-information-with-jboss-data-virtualization

Thomas H. Davenport, Rajeev Ronanki. (2018, 1 Jan). Artificial Intelligence for the Real World. *Harvard Business Review Digital Article*.

Tibco. (2021). *TIBCO Data Virtualization Architecture.* Retrieved from Tibco: https://docs.tibco.com/pub/tdv/8.5.0/doc/html/StudioHelp/index.html#page/StudioHelp/chpt1\_Intro\_2.StudioHelp.html

TIBCO. (không năm xuất bản). *Data virtualization*. Retrieved March 10, 2024, from TIBCO: https://www.tibco.com/platform/data-virtualization

W.H. Inmon, R.D. Hackathorn. (2007). *Using the data warehouse.* New York: John Wiley & Sons, Inc.

Wankowski, J. (2019). *Data Virtualization: The Evolution of the Data Lake.* Retrieved March 10, 2024, from IBM: https://www.ibm.com/blog/data-virtualization-the-evolution-of-the-data-lake/