

GIẢI PHÁP MỞ RỘNG KHO DỮ LIỆU PHỤC VỤ QUẢN LÝ ĐIỀU HÀNH VÀ RA QUYẾT ĐỊNH TRONG CÁC TỔ CHỨC, DOANH NGHIỆP

Chu Thị Hồng Hải^{1*}, Nguyễn Thị Liệu²

¹Học viện Ngân hàng

²Trường Đại học Công nghệ Đồng Nai

*Tác giả liên hệ: Chu Thị Hồng Hải, email: haict@hvnhv.edu.vn

THÔNG TIN CHUNG

Ngày nhận bài: 02/01/2023

Ngày nhận bài sửa: 05/12/2023

Ngày duyệt đăng: 08/12/2023

TỪ KHOÁ

Kho dữ liệu;

Kết nối dữ liệu;

Chia sẻ dữ liệu;

Mở rộng kho dữ liệu;

Tích hợp dữ liệu.

TÓM TẮT

Kho dữ liệu đóng vai trò then chốt trong việc tập hợp, xử lý và cung cấp dữ liệu cho các quyết định kinh doanh trong các tổ chức, doanh nghiệp. Qua nghiên cứu thực trạng cho thấy nhu cầu mở rộng kho dữ liệu nhằm tăng độ phủ dữ liệu, chiều sâu thông tin,... là cần thiết, thường xuyên và liên tục. Để giúp các tổ chức, doanh nghiệp nhanh chóng thực hiện hoạt động mở rộng kho dữ liệu, bài viết tập trung trình bày các nội dung liên quan tới kho dữ liệu, mở rộng kho dữ liệu và đặc biệt là các nhóm giải pháp giúp mở rộng kho dữ liệu trong các tổ chức, doanh nghiệp.

1. ĐẶT VẤN ĐỀ

Trong kỷ nguyên số hóa, thông tin trở thành một nguồn lực vô cùng quý giá của doanh nghiệp. Mỗi ngày, hàng triệu dữ liệu được tạo ra, lưu trữ, ... và câu hỏi đặt ra: *Làm thế nào để quản lý và khai thác được giá trị từ lượng thông tin khổng lồ này?* Kho dữ liệu là một giải pháp tốt cho tổ chức, doanh nghiệp.

Khái niệm kho dữ liệu (Data Warehouse) không còn xa lạ, năm 1988 kho dữ liệu được nhắc tới bởi hai nhà nghiên cứu của IBM là Barry Devlin và Paul Murphy. Thời gian đó, nhu cầu lưu trữ dữ liệu phát sinh khi các hệ thống máy tính trở nên phức tạp hơn và phải xử lý lượng dữ liệu ngày càng lớn [William H. Inmon, 2008]. Trong những năm đầu, kho dữ liệu chủ yếu được sử dụng trong các doanh nghiệp lớn, có nguồn

lực vì việc triển khai và duy trì hệ thống còn rất phức tạp. Ngày nay, nhiều công nghệ mới ra đời đã hỗ trợ cho việc xây dựng, vận hành và phát triển kho dữ liệu trở lên đơn giản, thuận tiện. Mặt khác, với sự phát triển nhanh chóng của nền kinh tế số, xã hội số, các quốc gia thực hiện Chính phủ điện tử, chính phủ số,... đã đòi hỏi các doanh nghiệp muốn phát triển, đứng vững trên thị trường, tạo đột phá thì chắc chắn phải quan tâm đến dữ liệu, coi dữ liệu là một trong những tài sản quý giá của tổ chức, cần được quản lý, khai thác sao cho hiệu quả, nâng cao giá trị của dữ liệu. Chính vì vậy, gần đây kho dữ liệu đóng vai trò vô cùng quan trọng và then chốt trong các chiến lược, mục tiêu phát triển doanh nghiệp. Nhờ kho dữ liệu các doanh nghiệp có khả năng phân tích chuyên sâu các dữ liệu theo chuỗi thời gian, giúp nhà lãnh đạo hiểu thêm quá trình kinh doanh và

còn là nguồn dữ liệu nghiên cứu vô giá cho các nhà phân tích nhu cầu xu hướng khách hàng, thị trường,.... có thể nói dữ liệu là cơ sở để doanh nghiệp cải thiện hoạt động kinh doanh liên tục của mình.

Song hành với hoạt động xây dựng kho dữ liệu, hoạt động mở rộng kho dữ liệu cũng vô cùng cần thiết. Trong quá trình vận hành hằng ngày, nhu cầu của khách hàng thay đổi, mục tiêu của tổ chức thay đổi, do vậy các dữ liệu thu thập, lưu trữ trong doanh nghiệp cũng thay đổi theo thời gian. Vậy việc mở rộng kho dữ liệu sẽ giúp các tổ chức, doanh nghiệp tăng độ phủ dữ liệu, tăng chiều sâu thông tin hỗ trợ ra quyết định trong quản lý điều hành và kinh doanh tốt hơn, tạo lợi thế cạnh tranh của doanh nghiệp. Để giúp các tổ chức, doanh nghiệp nhận thức tốt hơn về tầm quan trọng của dữ liệu, cơ sở cũng như phương pháp thực hiện xây dựng và mở rộng kho dữ liệu, các nội dung: (1) giới thiệu lý thuyết về kho dữ liệu và mở rộng kho dữ liệu; (2) nhu cầu mở rộng kho dữ liệu của tổ chức doanh nghiệp; (3) các nhóm giải pháp mở rộng kho dữ liệu sẽ được trình bày chi tiết ở các phần dưới đây.

2. TỔNG QUAN VỀ KHO DỮ LIỆU

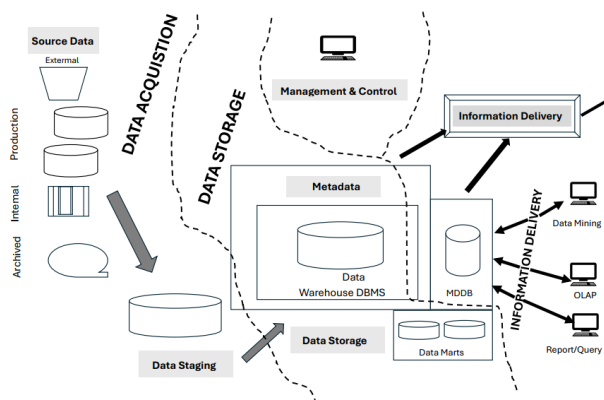
2.1. Khái niệm kho dữ liệu

Kho dữ liệu (Data Warehouse) là nơi lưu trữ dữ liệu trung tâm được trích xuất từ các cơ sở dữ liệu lịch sử hoặc từ cơ sở dữ liệu vận hành. Dữ liệu trong kho dữ liệu là hướng chủ đề, không thay đổi và mang tính lịch sử. Vì vậy kho dữ liệu có xu hướng chứa các tập dữ liệu cực lớn.[Ponniiah, 2012]

Mục tiêu chính của kho dữ liệu là cung cấp một nguồn dữ liệu chất lượng và hiệu quả cho việc phân tích kinh doanh, báo cáo, và hỗ trợ ra quyết định.

2.2. Kiến trúc kho dữ liệu

Kiến trúc của kho dữ liệu có thể được thiết kế với nhiều tầng khác nhau, nhưng một trong những kiến trúc phổ biến nhất là kiến trúc 3 tầng (hoặc 3 lớp) [Ponniiah, 2012]:



Hình 1. Kiến trúc kho dữ liệu

Nguồn: [Ponniiah, 2012]

Tầng Thu thập dữ liệu (Data acquisition Layer): Tại tầng này, dữ liệu được thu thập và chuẩn bị để lưu trữ trong kho dữ liệu. Hệ thống thực hiện thu thập dữ liệu từ các nguồn khác nhau, bao gồm dữ liệu nội bộ (các cơ sở dữ liệu tác nghiệp, dữ liệu từ các ứng dụng, dữ liệu báo cáo tổng hợp, các hệ thống camera giám sát, cảm biến IoT,...), và các nguồn khác từ bên ngoài.

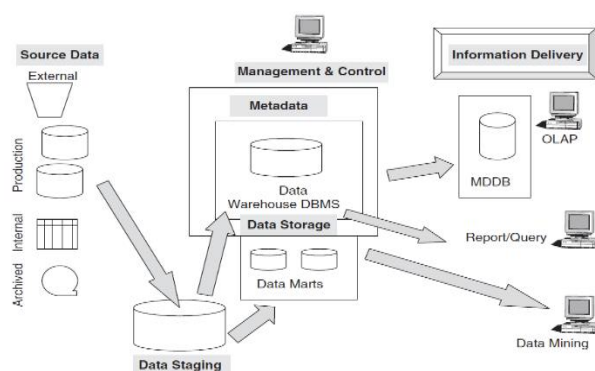
Tầng Lưu trữ dữ liệu (Data Storage Layer): Dữ liệu sau khi đã được thu thập và chuẩn bị ở tầng thu thập dữ liệu được đưa vào lưu trữ. Tầng này là nơi lưu trữ tất cả thông tin thu thập từ các nguồn dữ liệu khác nhau. Bên cạnh những dữ liệu được lưu trữ tập trung trong kho dữ liệu, tùy theo tổ chức, doanh nghiệp dữ liệu cũng có thể được trích xuất, biến đổi, xử lý và lưu trữ trong các kho dữ liệu con (data mart) để phục vụ những yêu cầu nghiệp vụ chuyên biệt (chẳng hạn data mart phục vụ cho kinh doanh, data mart cho tiếp thị, data marts cho quản lý quan hệ khách hàng,...).

Tầng Phân tích và báo cáo (Data Analysis and Reporting Layer): Tầng này cung cấp khả năng truy vấn, phân tích, và trích xuất thông tin

từ kho dữ liệu để hỗ trợ quyết định quản lý, kinh doanh.

2.3. Các hoạt động chính trong kho dữ liệu

Kho dữ liệu (Data Warehouse) thực hiện nhiều hoạt động để cung cấp thông tin hữu ích cho tổ chức và quyết định kinh doanh.



Hình 2. Luồng dữ liệu trong kho

Nguồn: [Ponniah, 2012]

Các hoạt động quan trọng của kho dữ liệu (Ponniah, 2012) bao gồm:

Thu thập dữ liệu: Kho dữ liệu thu thập dữ liệu từ nhiều nguồn khác nhau, bao gồm các hệ thống nội bộ và bên ngoài của tổ chức. Dữ liệu có thể bao gồm dữ liệu từ các hệ thống giao dịch, dữ liệu từ mạng xã hội, và nhiều nguồn dữ liệu khác.

Tích hợp dữ liệu: Kho dữ liệu tích hợp dữ liệu từ các nguồn khác nhau thành một cơ sở dữ liệu đồng nhất. Quá trình này bao gồm làm sạch, chuyển đổi, và kết nối dữ liệu để đảm bảo tính nhất quán và đúng định dạng.

Lưu trữ dữ liệu: Dữ liệu sau khi được tích hợp được lưu trữ trong kho dữ liệu. Kho dữ liệu thường sử dụng các hệ quản trị cơ sở dữ liệu (DBMS) để xử lý, lưu trữ và quản lý dữ liệu.

Tối ưu hóa dữ liệu cho phân tích: Dữ liệu trong kho dữ liệu được tối ưu hóa để phục vụ cho các hoạt động phân tích, bao gồm việc tạo các chỉ số, tổ chức dữ liệu theo cách tối ưu cho tốc độ truy vấn, và đảm bảo rằng dữ liệu được cập nhật định kỳ.

Phân tích dữ liệu: Người dùng cuối và các công cụ phân tích dữ liệu truy cập kho dữ liệu để tạo ra các báo cáo, biểu đồ, và phân tích dữ liệu. Quá trình này giúp tổ chức hiểu rõ hơn về hoạt động của họ, xu hướng, và cơ hội.

Bảo mật và quản lý quyền truy cập: Kho dữ liệu phải đảm bảo rằng dữ liệu được bảo mật và chỉ được truy cập bởi những người có quyền. Nó cũng quản lý quyền truy cập và theo dõi hoạt động của người dùng để đảm bảo tính an toàn.

Sao lưu và phục hồi: Kho dữ liệu thường thực hiện quá trình sao lưu định kỳ để đảm bảo dữ liệu không bị mất. Nếu xảy ra sự cố, nó cũng cung cấp khả năng phục hồi dữ liệu.

Quản lý hiệu suất: Kho dữ liệu cần theo dõi và quản lý hiệu suất để đảm bảo rằng truy vấn và phân tích dữ liệu diễn ra một cách hiệu quả và không gây quá tải cho hệ thống.

2.4. Các phương pháp tiếp cận xây dựng – mở rộng kho dữ liệu

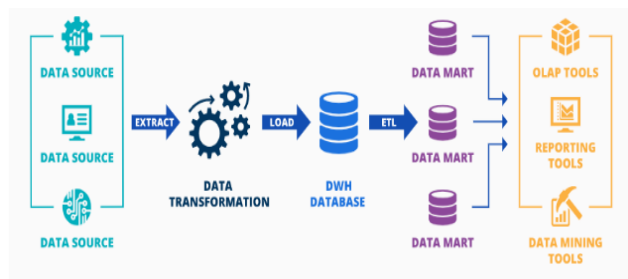
Mỗi dự án kho dữ liệu, với những đặc điểm khác nhau về nghiệp vụ, môi trường, quy mô, nhân sự, ... sẽ có cách tiếp cận riêng. Phần dưới đây trình bày hai phương pháp tiếp cận điển hình trong xây dựng, mở rộng kho dữ liệu.

2.4.1. Phương pháp tiếp cận của Inmon

Đặc điểm cách tiếp cận

Inmon [William H. Inmon, 2008], [Inmon, 2005] sử dụng cách tiếp cận từ trên xuống, với kiến trúc tập trung (Centralized Architecture), trong đó có một kho dữ liệu tập trung Enterprise Data Warehouse (EDW) làm trung tâm. Mọi dữ liệu được từ các nguồn khác nhau tích hợp vào kho dữ liệu tập trung này. Inmon chú trọng đến việc phân tích quy trình nghiệp vụ và xây dựng mô hình dữ liệu dựa trên quy trình này. Dữ liệu trong kho dữ liệu Inmon thường được chuẩn hóa để đảm bảo tính nhất quán.

Xây dựng kho dữ liệu



Hình 3. Luồng dữ liệu trong kho Inmon

Phương pháp Inmon phù hợp với các tổ chức có nhu cầu tích hợp và phân tích dữ liệu phức tạp, và họ thường ưu tiên tính nhất quán và khả năng mở rộng của hệ thống dữ liệu.

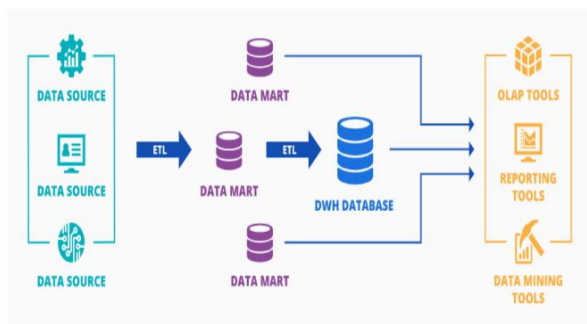
2.4.2. Phương pháp tiếp cận của Kimball

Đặc điểm cách tiếp cận

Kimball tiếp cận từ dưới lên (Bottom-Up Approach), sử dụng kiến trúc phân tán, trong đó dữ liệu được lưu trữ trong các kho dữ liệu con (data mart) riêng biệt cho từng phòng ban hoặc nguồn dữ liệu cụ thể. Bắt đầu từ việc xây dựng các data mart riêng rẽ, các data mart này sau đó được tích hợp để tạo thành kho dữ liệu toàn cục. Dữ liệu trong data mart Kimball thường được thiết kế để tối ưu hóa hiệu suất truy vấn chứ không ưu tiên tính chuẩn hóa của dữ liệu.

Xây dựng kho dữ liệu

Phương pháp xây dựng kho dữ liệu theo tiếp cận của Ralph Kimball, là một tiếp cận phân tán dựa trên việc xây dựng các data mart độc lập và tập trung vào việc cung cấp giá trị nhanh chóng cho người dùng cuối. Dưới đây mô tả cách xây dựng kho dữ liệu theo phương pháp Kimball:



Hình 4. Luồng dữ liệu trong kho Kimball

Phương pháp Kimball thích hợp cho các dự án có sự ưu tiên về việc cung cấp giá trị nhanh chóng cho người dùng cuối và tập trung vào việc tối ưu hóa hiệu suất truy vấn. Nó thường được sử dụng trong các tình huống khi cần cung cấp thông tin phân tích cho các phòng ban cụ thể hoặc mục tiêu kinh doanh cụ thể.

Một vài so sánh hai phương pháp tiếp cận:

Bảng 1. Bảng so sánh hai phương pháp Inmon và Kimball

Kimball	Inmon
Kho dữ liệu dựa trên Kimball có thể được thiết lập nhanh chóng	kho dữ liệu dựa trên Inmon thể được thiết lập nhanh chóng lâu hơn
Việc bảo trì kho dữ liệu khó khăn	Việc bảo trì kho dữ liệu khó khăn dễ dàng.
Chi phí ban đầu thấp và chi phí trải đều cho các giai đoạn tiếp theo	Chi phí ban đầu cao, nhưng chi phí phát triển dự án sau đó sẽ thấp hơn.
Yêu cầu nhóm chuyên môn thực hiện.	Yêu cầu nhóm chuyên gia thực hiện
Kho dữ liệu yêu cầu tích hợp dữ liệu tập trung vào lĩnh vực kinh doanh riêng lẻ.	Kho dữ liệu tích hợp dữ liệu trên toàn doanh nghiệp
Tuân theo cách tiếp cận từ dưới lên.	Tuân theo cách tiếp cận từ trên xuống.

Khi xây dựng kho dữ liệu, việc lựa chọn giữa phương pháp Inmon và Kimball phụ thuộc vào nhiều yếu tố như: tổ chức, yêu cầu kinh doanh,

và mục tiêu cụ thể. Phương pháp Inmon thường thích hợp cho các tổ chức lớn với nhu cầu tích hợp và phân tích dữ liệu phức tạp. Phương pháp Kimball thường phù hợp với các dự án tập trung vào việc cung cấp giá trị nhanh chóng cho người dùng cuối thông qua các data mart cụ thể. Trong một số trường hợp, việc kết hợp cả hai phương pháp có thể là lựa chọn tốt.

3. MỞ RỘNG KHO DỮ LIỆU TRONG TỔ CHỨC, DOANH NGHIỆP

3.1. Mở rộng kho dữ liệu

Sau một thời gian được xây dựng, kho dữ liệu được đưa vào vận hành và khai thác. Trong quá trình này, hoạt động quản lý điều hành, các hoạt động nghiệp vụ, các nhu cầu phân tích dữ liệu có thể thay đổi hoặc phát sinh mới. Đến khi kho dữ liệu hiện tại không còn đáp ứng đủ các nhu cầu lưu trữ, phân tích và xử lý dữ liệu của doanh nghiệp nữa thì kho dữ liệu sẽ cần được nâng cấp, mở rộng.

Mở rộng kho dữ liệu là quá trình tăng khả năng của kho dữ liệu để có thể xử lý và lưu trữ thêm dữ liệu hoặc để đáp ứng nhu cầu nghiệp vụ mới hoặc tăng cường hiệu suất xử lý dữ liệu. Nhu cầu mở rộng kho dữ liệu phát sinh khi tổ chức phải đối mặt với dữ liệu ngày càng lớn hoặc nhu cầu người dùng cuối đòi hỏi một hệ thống mạnh mẽ hơn. Có nhiều phương pháp mở rộng kho dữ liệu, dưới đây là một số phương pháp mở rộng kho dữ liệu, tùy theo thực trạng, và bối cảnh cụ thể mà tổ chức, doanh nghiệp sẽ chọn phương án phù hợp để thực hiện:

Mở rộng dung lượng lưu trữ: Đơn giản nhất, ta có thể mở rộng dung lượng lưu trữ của hệ thống kho dữ liệu bằng cách thêm thiết bị phần cứng (ổ cứng hoặc bộ nhớ) vào hệ thống hiện có. Điều này giúp kho dữ liệu có thể lưu trữ thêm dữ liệu mà không cần thay đổi kiến trúc tổng thể.

Tăng hiệu năng xử lý: Nâng cấp các thành phần phần cứng, chẳng hạn như bộ vi xử lý (CPU), bộ nhớ RAM, hoặc hệ thống lưu trữ SSD, có thể tăng hiệu năng xử lý của kho dữ liệu và

giúp nó xử lý các truy vấn và phân tích nhanh hơn.

Mô hình dữ liệu phân mảng (Thêm vùng dữ liệu chuyên đề): Tạo thêm các data mart để phân tách dữ liệu và phân tải công việc xử lý dữ liệu. Mô hình này giúp giảm áp lực trên kho dữ liệu chính và tạo điều kiện thuận lợi cho tối ưu hóa hiệu suất.

Mở rộng tích hợp nguồn dữ liệu: Thêm nhiều nguồn dữ liệu khác nhau để có thêm thông tin mới và đa dạng hóa là giàu dữ liệu trong kho dữ liệu. Điều này có thể bao gồm tích hợp dữ liệu từ các nguồn phi cấu trúc, dữ liệu thời gian thực, hoặc các nguồn dữ liệu bên ngoài tổ chức. Mở rộng nguồn dữ liệu giúp đáp ứng được các nhu cầu phân tích dữ liệu mới (mà hiện tại chưa thực hiện được do chưa có dữ liệu).

Sử dụng các công cụ tối ưu hóa dữ liệu (Data Optimization Tools): Các công cụ tối ưu hóa dữ liệu có thể giúp nén, tổ chức lại, và làm sạch dữ liệu để tiết kiệm không gian lưu trữ và tăng tốc truy vấn.

Sử dụng cơ sở hạ tầng đám mây (Cloud Infrastructure): Chuyển kho dữ liệu lên môi trường đám mây có thể giúp mở rộng một cách dễ dàng dựa trên nhu cầu. Các dịch vụ đám mây cho khả năng mở rộng và có tính linh hoạt cao.

Sử dụng cơ sở hạ tầng như một dịch vụ (Infrastructure as a Service - IaaS): Doanh nghiệp có thể thuê dịch vụ kho dữ liệu và các cơ sở hạ tầng tính toán liên quan (máy chủ, không gian lưu trữ,...) từ nhà cung cấp dịch vụ. Sử dụng IaaS cho phép mở rộng hạ tầng lưu trữ và xử lý mà không cần đầu tư vào phần cứng cụ thể.

Sử dụng cơ sở dữ liệu phân tán (Distributed Database): Xây dựng hệ thống kho dữ liệu sử dụng cơ sở dữ liệu phân tán để chia sẻ khối lượng công việc và tối ưu hóa khả năng chịu tải.

Khi mở rộng kho dữ liệu, quan trọng là xem xét các yêu cầu cụ thể của doanh nghiệp và đảm bảo rằng việc mở rộng được thực hiện một cách hiệu quả và bền vững. Điều này bao gồm việc đánh giá kiến trúc, dự đoán tăng trưởng, và lựa

chọn các giải pháp phù hợp để đáp ứng nhu cầu lưu trữ và xử lý dữ liệu ngày càng lớn của doanh nghiệp.

3.2. Nhu cầu mở rộng kho dữ liệu trong các doanh nghiệp

Để xác định nhu cầu mở rộng kho dữ liệu trong các tổ chức, doanh nghiệp Việt nam hiện nay, nhóm tác giả đã tiến hành khảo sát gần 100 doanh nghiệp ở Việt nam đã triển khai xây dựng kho dữ liệu. Qua kết quả khảo sát cho thấy nhu cầu mở rộng kho dữ liệu của các tổ chức, doanh nghiệp là thường xuyên vì họ luôn muốn có nhiều thông tin và thông tin đa chiều để ra quyết định kinh doanh và điều hành. Tuy nhiên, không phải lúc nào các tổ chức, doanh nghiệp có nhu cầu là thực hiện mở rộng được vì để mở rộng kho dữ liệu cần xem xét đến nhiều yếu tố và đảm bảo thực hiện mục tiêu phát triển lâu dài của tổ chức. Dưới đây sẽ phân tích một vài yếu tố về khía cạnh chủ chương, chính sách của nhà nước thúc đẩy các doanh nghiệp cần xây dựng và mở rộng kho dữ liệu, cụ thể:

Ngày 07 tháng 3 năm 2019, Chính phủ đã ban hành Nghị quyết số 17/NQ-CP về một số nhiệm vụ, giải pháp trọng tâm phát triển Chính phủ điện tử giai đoạn 2019 - 2020, định hướng đến năm 2025 [chính phủ, 2019]. Tiếp đó, Thủ tướng Chính phủ đã ban hành Quyết định số 749/QĐ-TTg ngày 03 tháng 6 năm 2020 phê duyệt Chương trình chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030 [chính phủ, 2020]; Quyết định số 942/QĐ-TTg ngày 15 tháng 6 năm 2021 phê duyệt Chiến lược phát triển Chính phủ điện tử hướng tới Chính phủ số giai đoạn 2021 - 2025, định hướng đến năm 2030. Các nghị quyết, quyết định trên có tác động trực tiếp và mạnh mẽ đến việc ứng dụng công nghệ thông tin trong việc quản lý, ra quyết định điều hành tại tổ chức, doanh nghiệp. Thực tế cho thấy, dữ liệu trong các doanh nghiệp còn nằm rải rác ở nhiều nơi, việc lưu trữ chưa được nhất quán, thiếu sự thống nhất, liên thông dữ liệu giữa các đơn vị trong nội bộ cũng như giữa các tổ chức, doanh nghiệp, nhiều dữ liệu đã được thu thập rồi vẫn yêu cầu thu thập lại khi cần ở trong cùng một tổ chức hoặc nhiều tổ chức khác nhau,... điều này

gây mất thời gian, sự thiếu chuyên nghiệp, thiếu sự ủng hộ đồng hành của các bên liên quan,... Để khắc phục những hạn chế từ thực trạng trên, Thủ tướng Chính phủ ban hành nhiều quyết định, nghị định nhằm tạo môi trường pháp lý và định hướng cho các tổ chức, doanh nghiệp nhanh chóng xây dựng, phát triển các sản phẩm liên quan tới lưu trữ, chia sẻ dữ liệu,... các quyết định, nghị định đã được ban hành gần đây sẽ được trình bày theo từng mảng chuyên đề nhằm là nổi bật các quan điểm chỉ đạo và phương hướng thực hiện trong việc khai thác dữ liệu trong các tổ chức, doanh nghiệp.

Lĩnh vực Chuyển đổi số

Chuyển đổi số đã trở thành là xu thế tất yếu, nằm trong ưu tiên phát triển của hầu hết các quốc gia trên thế giới. Tại Việt Nam, chuyển đổi số là một trong những trọng tâm phải thực hiện theo chỉ đạo của Đảng, Nhà nước. Theo đó, tổ chức, doanh nghiệp đồng hành thực hiện các chỉ tiêu liên quan đến chuyển đổi số như: triển khai chức năng kho dữ liệu điện tử của tổ chức, cá nhân trên hệ thống thông tin giải quyết thủ tục hành chính cấp bộ, cấp tỉnh để người dân, doanh nghiệp chỉ phải cung cấp thông tin một lần cho cơ quan nhà nước khi thực hiện dịch vụ công trực tuyến; 100% bộ, ngành, địa phương triển khai các hoạt động nâng cao chất lượng và hiệu quả cung cấp dịch vụ công trực tuyến; triển khai hệ thống thông tin giải quyết thủ tục hành chính cấp bộ, cấp tỉnh trên cơ sở hợp nhất cổng dịch vụ công và hệ thống một cửa điện tử; 100% kết quả giải quyết thủ tục hành chính được số hóa; 50% hồ sơ thủ tục hành chính được người dân, doanh nghiệp thực hiện trực tuyến từ xa.

Lĩnh vực dữ liệu mở

Dữ liệu mở được định nghĩa theo nhiều quan điểm khác nhau như định nghĩa dữ liệu mở là dữ liệu được cấp phép theo giấy phép mở, cho phép chúng ta tự do 3 việc: sử dụng và sử dụng lại, phân phối và phân phối lại (có thể tùy biến và pha trộn). Còn theo Nghị định số 47/2020/NĐ-CP ngày 09/4/2020 đã đưa ra định nghĩa “Dữ liệu mở của cơ quan nhà nước: là dữ liệu được cơ quan nhà nước có thẩm quyền công bố rộng rãi cho cơ

quan, tổ chức, cá nhân tự do sử dụng, tái sử dụng, chia sẻ”.

Đến nay, Việt nam đã công bố 06 CSDL quốc gia đã được xây dựng, hoàn thiện và đưa vào sử dụng là nguồn dữ liệu mở. Các nguồn dữ liệu này, hầu như chưa được tích hợp vào kho dữ liệu của các tổ chức, doanh nghiệp có nhu cầu.

STT	CƠ SỞ DỮ LIỆU QUỐC GIA	HỆ THỐNG	DỮ LIỆU	PHÁP LÝ	CHUẨN TRAO ĐỔI	KẾT NỐI NGSP
1	CSDLQG về Dân cư					
2	CSDLQG về đăng ký doanh nghiệp					
3	CSDL về đất đai quốc gia					
4	CSDLQG về tài chính					
5	CSDLQG về Bảo hiểm					
6	CSDLQG về thông kê tổng hợp dân số					

Hình 5. Các cơ sở dữ liệu mở Quốc gia

Nguồn: Cổng dữ liệu Quốc gia

3.3. Đánh giá nhu cầu và các vấn đề gặp phải

Qua kết quả điều tra, khảo sát tổ chức, doanh nghiệp đã triển khai kho dữ liệu cho thấy:

Đến nay, có 06 CSDL quốc gia đã được xây dựng, hoàn thiện và đưa vào sử dụng, tuy nhiên hầu hết các tổ chức, doanh nghiệp chưa tích hợp các nguồn này vào kho dữ liệu để khai thác.

Kiến trúc kho dữ liệu của các tổ chức, doanh nghiệp hiện nay là kiến trúc mở dễ dàng phát triển theo nhu cầu của tổ chức khi cần có sự thay đổi, phát triển mở rộng về nguồn dữ liệu, vùng dữ liệu chuyên đề (Data Mart) và các công cụ khai thác dữ liệu trong kho.

Lưu trữ dữ liệu trong kho dữ liệu không cho phép lưu trữ được các file dữ liệu có dung lượng lớn, các bản báo cáo bằng video, voice,...

Việc cập nhật các dữ liệu vào khó khăn kiểm soát dữ liệu đã có trong kho hay chưa.

Với thực trạng kho dữ liệu của hầu hết các tổ chức, doanh nghiệp như hiện nay và với nhu cầu về dữ liệu trong giai đoạn thực hiện chủ trương của chính phủ về chính phủ điện tử

từ, chuyển đổi số, phát triển nền kinh tế số,... nhu cầu về dữ liệu của doanh nghiệp sẽ tăng nhanh cả về số lượng, chất lượng và chủng loại việc phát triển mở rộng kho dữ liệu là một nhu cầu tất yếu trong các tổ chức, doanh nghiệp.

4. GIẢI PHÁP MỞ RỘNG KHO DỮ LIỆU

Như đã phân tích trong các nội dung ở trên, kho dữ liệu đóng vai trò quan trọng nhằm đảm bảo nguồn dữ liệu phục vụ quản lý điều hành và ra quyết định kinh doanh trong các doanh nghiệp. Theo thời gian, lượng dữ liệu gia tăng ngày càng nhiều, cả về dung lượng lẫn đa dạng giúp hỗ trợ hoạt động phân tích dữ liệu để tối ưu hóa quy trình vận hành, tăng hiệu quả hoạt động, quản trị rủi ro,... Việc mở rộng kho sẽ có rất nhiều phương án, các phương án đã được đưa ra trong nội dung ở phần 3.1. Tuy nhiên, với thực trạng về kho dữ liệu trong các tổ chức, doanh nghiệp hiện nay và với bối cảnh của nền kinh tế số, chính phủ điện tử, các phương án mở rộng kho dữ liệu được đề xuất phân thành hai nhóm giải pháp: (i) nhóm giải pháp chính nhằm mở rộng kho dữ liệu phục vụ quản lý điều hành và ra quyết định kinh doanh (ii) nhóm giải pháp tăng cường là các giải pháp nhằm hỗ trợ việc khai thác, vận hành nhanh chóng, hiệu quả trên kho dữ liệu, cụ thể:

(i) Nhóm giải pháp chính sẽ giải quyết các bài toán: (1) *Thêm vùng dữ liệu chuyên đề*; (2) *Tích hợp các nguồn dữ liệu từ bên ngoài*; (3) *Tích hợp dữ liệu từ các nguồn phi cấu trúc, dữ liệu thời gian thực*,...

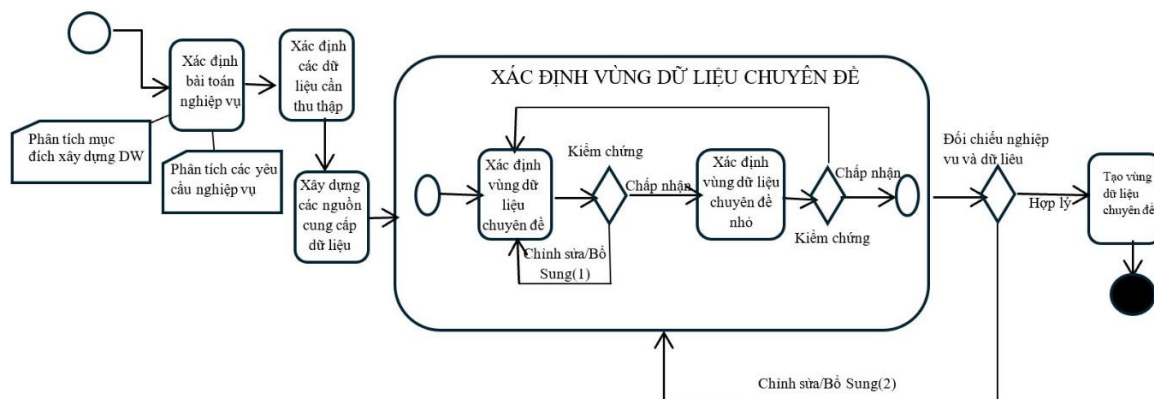
(ii) Nhóm giải pháp tăng cường là các giải pháp khác hỗ trợ hoạt động lưu trữ, khai thác, sử dụng dữ liệu tại kho một cách nhanh chóng, an toàn, hiệu quả giải quyết các bài toán: (1) *Xây dựng công cụ tra cứu dữ liệu tránh trùng lặp các dữ liệu đã thu thập* (2) *Tự động hóa hoạt động nghiệp vụ sử dụng dịch vụ công dựa trên khai phá dữ liệu và kỹ thuật gợi ý thông tin*; (3) *Xây dựng giải pháp tra cứu trên các loại dữ liệu khác nhau (văn bản, hình ảnh, âm thanh...) sử dụng trí tuệ nhân tạo, ...*

Trong các giải pháp được đề xuất ở trên, giải pháp thêm vùng dữ liệu chuyên đề được đánh giá giải pháp sử dụng thường xuyên và có nhu cầu nhiều nhất trong các tổ chức, doanh nghiệp. Do vậy, nhóm tác giả sẽ trình bày chi tiết giải pháp này.

Mô tả giải pháp

Việc mở rộng, thêm vùng dữ liệu chuyên đề thường xuất phát từ nhu cầu nghiệp vụ mới, mục tiêu mới đòi hỏi dữ liệu mới chưa có trong kho. Để các doanh nghiệp dễ dàng thực hiện nhiệm vụ này, nhóm tác giả đề xuất quy trình thực hiện như Hình 6 dưới đây:

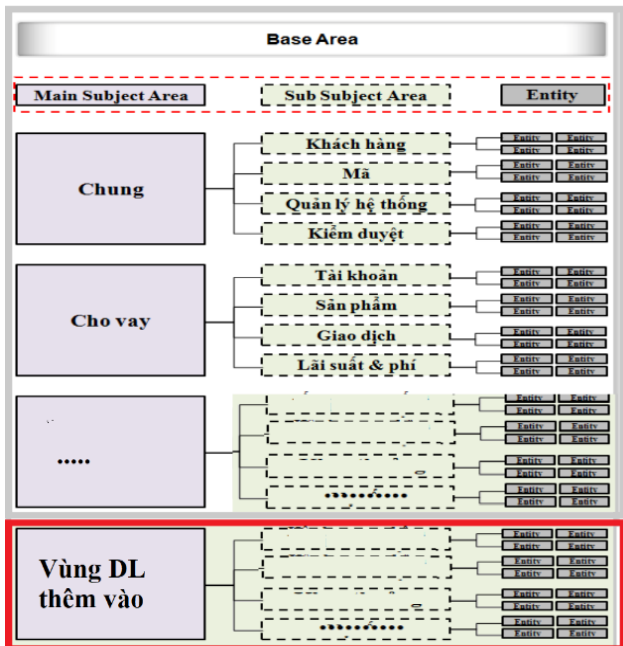
Dữ liệu được thêm vào kho bắt đầu xuất phát từ các nhu cầu nghiệp vụ của các đơn vị trong tổ chức, doanh nghiệp. Khi các đơn vị có nhu cầu dữ liệu các đơn vị gửi yêu cầu tới bộ phận chịu trách nhiệm đưa dữ liệu vào kho. Bộ phận chịu trách nhiệm sẽ kiểm soát dữ liệu được đưa vào trong kho và quyết định đưa dữ liệu vào các chuyên đề dữ liệu đã có hay chưa đưa vào được vì không tương xứng với các vùng dữ liệu đã có trong kho. Khi dữ liệu yêu cầu chưa có trong kho từ các đơn vị gia tăng và đến một ngưỡng nào đó để thuận tiện cho việc thiết lập và lưu trữ và báo cáo thì nhu cầu xây dựng vùng dữ liệu chuyên đề mới xuất hiện. Tiếp theo, đơn vị chức năng phải tiến hành phân tích kỹ các yêu cầu về dữ liệu phục vụ cho mục đích gì? Các nguồn dữ liệu mới lấy ở đâu? Và dựa vào kinh nghiệm triển khai,... để quyết định có thêm vùng dữ liệu mới hay không? Nếu thêm vùng dữ liệu đó thuộc kho lưu trữ nào, nó là một vùng nhỏ của vùng dữ liệu chuyên đề cũ hay tạo vùng dữ liệu chuyên đề mới,... tất cả đó được kiểm chứng rất rõ ràng minh bạch trước khi đưa ra quyết định cuối cùng



Hình 6. Quy trình thêm vùng dữ liệu

Nguồn: Nhóm tác giả đề xuất

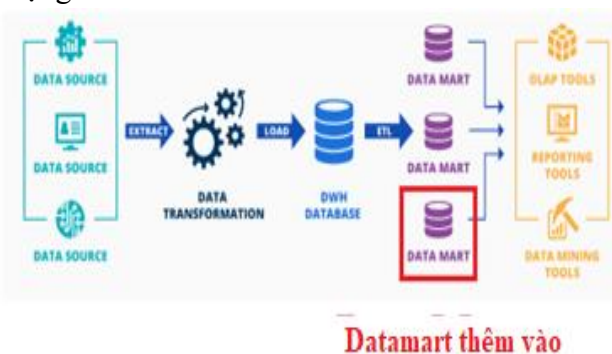
Vùng dữ liệu chuyên đề mới được xác định bằng cách lặp lại quy trình phân chia các Vùng dữ liệu chuyên đề chính thành các Vùng dữ liệu chuyên đề nhỏ (1) trong hình 6 và quy trình phản ánh kết quả đối chiếu giữa nghiệp vụ và dữ liệu vào Vùng dữ liệu chuyên đề (2). Vùng dữ liệu chuyên đề trong kho sau khi thêm vùng dữ liệu mới có dạng như hình dưới đây:



Hình 7. Thêm một vùng dữ liệu mới vào kho

Nguồn: Nhóm tác giả đề xuất

Cấu trúc của vùng dữ liệu thêm vào sẽ phải phù hợp với cấu trúc của các vùng dữ liệu cũ, có dạng như hình dưới:



Hình 8. Cấu trúc vùng dữ liệu thêm vào

Nguồn: Nhóm tác giả đề xuất

5. KẾT LUẬN

Các quyết định, chiến lược kinh doanh cũng như cách thức vận hành của tổ chức, doanh nghiệp trong nền kinh tế số hiện nay đang rất cần vào sự trợ giúp của dữ liệu. Việc mở rộng các nguồn dữ liệu để khai thác phục vụ mục tiêu quản lý điều hành và ra quyết định là một điều tất yếu đối với các doanh nghiệp. Bài viết đã phân tích nhu cầu mở rộng kho dữ liệu trong các tổ chức, doanh nghiệp từ đó đề xuất 03 giải pháp chính, 02 giải pháp tăng cường giúp khai thác và vận hành kho dữ liệu hiệu quả hơn. Đặc biệt, bài viết còn đề xuất 01 quy trình mở rộng kho dữ liệu nhờ thêm vùng dữ liệu chuyên đề phục vụ cho việc phát sinh các nhu cầu mới trong quá trình hoạt động, kinh doanh của tổ chức, doanh nghiệp, đây là hoạt động được cho là mở rộng thường xuyên nhất so với các hoạt động khác trong tổ chức.

Vì tính chất phức tạp của việc mở rộng khai thác các nguồn dữ liệu, cùng với những lo ngại về sự minh bạch và bảo mật, việc mở rộng khai thác các nguồn dữ liệu đã gặp rất nhiều khó khăn do có dữ liệu có tính chất riêng tư, nhạy cảm,...Đã có nhiều nghiên cứu khẳng định rằng, việc mở rộng thường xuyên kho dữ liệu là điều kiện tiên quyết để phát triển các dịch vụ cung cấp, cải thiện năng suất, khả năng cạnh tranh của doanh nghiệp. Tuy nhiên, để việc mở rộng kho dữ liệu trong các tổ chức, doanh nghiệp thực sự hiệu quả thì các doanh nghiệp cũng cần nghiên cứu, phân tích kỹ các nhu cầu về dữ liệu của tổ chức doanh nghiệp mình và lựa chọn một phương pháp, thời điểm phù hợp.

TÀI LIỆU THAM KHẢO

Bộ Thông tin và Truyền thông, 2019. “Đề án giải pháp kết nối, chia sẻ dữ liệu và tái cấu trúc hạ tầng Công nghệ thông tin tại các bộ, ngành, địa phương.

Bộ Thông tin và Truyền thông, 2019. “Đề án giải pháp kết nối, chia sẻ dữ liệu và tái cấu trúc hạ tầng công nghệ thông tin tại các bộ, ngành, địa phương.”

Bộ thông tin và truyền thông, 2017. Thông tư Quy định các yêu cầu kỹ thuật về kết nối

- các hệ thống thông tin, cơ sở dữ liệu với cơ sở dữ liệu quốc gia.
- Chính phủ, 2020. Nghị định quản lý, kết nối và chia sẻ dữ liệu số của cơ quan nhà nước.
- Ponniah, P. (2012). Data Warehousing Fundamentals for IT Professionals (2nd Ed). Wiley.
- Viện công nghệ phần mềm và Nội dung số Việt Nam, 2017. “Một số vấn đề về sự cần thiết ban hành các quy định, chính sách về dữ liệu mở”.
- Granite, B. (2018). Data lakes in a modern data architecture. Inmon, W. H. (2005). Building the Data Warehouse, 4th edition. Wiley.
- Ponniah, P. (2012). Data Warehousing Fundamentals for IT Professionals (2nd Ed).
- Wiley. Ralph Kimball, M. R. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Wiley.
- Saurabh Gupta, V. G. (2018). Practical Enterprise Data Lake Insights: Handle Data-Driven Challenges in an Enterprise Big Data Lake. Apress.
- William H. Inmon, D. S. (2008). DW 2.0: The Architecture for the Next Generation of Data Warehousing. Morgan Kaufmann.

SOLUTIONS TO EXPAND DATA WAREHOUSES FOR MANAGEMENT AND DECISION MAKING IN ORGANIZATIONS AND ENTERPRISES

Chu Thi Hong Hai^{1*}, Nguyen Thi Lieu²

¹Banking Academy of Vietnam

²Dong Nai Technology University

*Corresponding author: Chu Thi Hong Hai, email: haict@hvnhv.edu.vn

GENERAL INFORMATION

Received date: 02/01/2023

Revised date: 05/12/2023

Published date: 08/12/2023

KEYWORD

Text to speech;

Fastpitch;

Transformer;

According to newspaper;

Recognize speech form.

ABSTRACT

The data repository plays a pivotal role in aggregating, processing, and providing data for business decisions within organizations and enterprises. Through empirical research, it is evident that expanding data repositories to increase data coverage, information depth is consistently essential. This article focuses on presenting relevant content regarding data repositories and expansion; particularly, solution groups to facilitate the expansion of data repositories in organizations and businesses.

