

ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Giải pháp quản lý và chia sẻ dữ liệu dùng chung

NGUYỄN THẾ THIÊN
thien.nt194177@sis.hust.edu.vn

Ngành Khoa học máy tính

Giảng viên hướng dẫn: TS. Vũ Tuyết Trinh

Chữ kí GVHD

Khoa: Khoa học máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 01/2024

LỜI CẢM ƠN

Để hoàn thành đồ án, ngoài sự nỗ lực, học hỏi của bản thân, em đã nhận được nhiều sự quan tâm và giúp đỡ từ các thầy cô, gia đình và đồng nghiệp.

Lời đầu tiên, em xin cảm ơn chân thành đến TS. Vũ Tuyết Trinh, người đã trực tiếp chỉ bảo, hướng dẫn nhiệt tình từng phương pháp để có thể hoàn thiện đồ án, không chỉ vậy, cô còn chia sẻ rất nhiều những kinh nghiệm thực tế cũng như sát sao trong suốt quá trình em thực hiện đồ án. Ngoài ra, em cũng muốn gửi lời cảm ơn đến các thầy cô giảng viên của Đại học Bách Khoa Hà Nội đã giảng dạy và truyền cảm hứng, động lực khi em học tập tại trường. Đồng thời, em cảm ơn anh Nguyễn Minh Đức đã đồng hướng dẫn cùng cô Trinh và các anh/ chị đồng nghiệp trong phòng Quản trị dữ liệu thuộc Trung tâm phát triển phần mềm tại Công ty Viễn thông điện lực và công nghệ thông tin đã có những chia sẻ, góp ý vô cùng thực tế trong quá trình thực hiện đồ án.

Con cũng muốn gửi lời cảm ơn bố, mẹ và những người thân trong gia đình đã luôn ủng hộ, là chỗ dựa tinh thần trong suốt quá trình học tập.

Trong quá trình triển khai và hoàn thiện đồ án, bản thân em nhận thấy không thể tránh được những hạn chế về trình độ kiến thức, thời gian nên đồ án khó tránh khỏi những thiếu sót. Vậy nên, em rất mong sẽ nhận được những ý kiến đóng góp, nhận xét từ các quý thầy cô để đề tài được hoàn thiện hơn.

Sinh viên,

Nguyễn Thế Thiện

TÓM TẮT NỘI DUNG ĐỒ ÁN

Hiện nay Tập đoàn Điện lực Việt Nam đang sử dụng nhiều phần mềm dùng chung, các phần mềm dùng chung này sẽ phục vụ các mảng hoạt động tại doanh nghiệp. Với mô hình triển khai hiện tại, các phần mềm dùng chung sẽ được triển khai tại đơn vị và sẽ lưu trữ dữ liệu vật lý tại đơn vị đó. Trong các PMDC này sử dụng nhiều danh mục dữ liệu dùng chung, dữ liệu dùng cho tham chiếu hoặc dữ liệu chủ. Dữ liệu của các danh mục này có thể được cập nhật tại EVN hoặc tại đơn vị dẫn đến việc chưa có sự nhất quán về quản lý và không đồng nhất về dữ liệu. Để giải quyết vấn đề này cần xây dựng một kho dữ liệu dùng chung để có thể nhất quán về quản lý và chuẩn hóa các danh mục đã có. Về mặt thu thập dữ liệu đối với dữ liệu có sự thay đổi xảy ra tại đơn vị, có thể tiếp cận theo 2 hướng là gọi trực tiếp API của phần mềm hoặc sử dụng các công cụ ETL để truy xuất dữ liệu từ database. Việc sử dụng API sẽ phụ thuộc vào phần mềm nên em lựa chọn hướng sử dụng công cụ ETL để truy xuất dữ liệu vì nó sẽ hoạt động độc lập, không ảnh hưởng tới các ứng dụng, dịch vụ.

Mục tiêu chính của đồ án là xây dựng dịch vụ nhằm quản lý và chia sẻ các danh mục dữ liệu chung và xây dựng công cụ thu thập dữ liệu từ đơn vị đối với các danh mục này. Từ đó thay thế việc mỗi khi cập nhật các danh mục dùng chung, thông qua mail các đơn vị lại nhập tay vào phần mềm. Dữ liệu dùng chung thu thập về cũng được sử dụng cho việc xây dựng các báo cáo quản trị của EVN.

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....	1
1.1 Bối cảnh	1
1.2 Mục tiêu và phạm vi đề tài.....	2
1.3 Bố cục đồ án	2
CHƯƠNG 2. KHẢO SÁT VÀ PHÂN TÍCH YÊU CẦU.....	3
2.1 Phân tích yêu cầu quản lý dữ liệu dùng chung.....	3
2.2 Phân tích yêu cầu thu thập dữ liệu dùng chung	6
2.3 Phân tích yêu cầu chia sẻ dữ liệu	7
2.4 Yêu cầu báo cáo tổng quan cơ cấu lao động.....	7
2.5 Yêu cầu phi chức năng	7
CHƯƠNG 3. XÂY DỰNG GIẢI PHÁP	9
3.1 Mô hình kiến trúc tổng thể	9
3.2 Thiết kế quá trình thu thập dữ liệu	9
3.3 Thiết kế các bảng dữ liệu dùng chung	12
3.4 Xây dựng API chia sẻ dữ liệu.....	14
CHƯƠNG 4. TRIỂN KHAI CÀI ĐẶT VÀ ĐÁNH GIÁ	16
4.1 Các công nghệ sử dụng	16
4.1.1 Master Data Services.....	16
4.1.2 Apache Kafka.....	16
4.1.3 Kafka Connect.....	18
4.1.4 Debezium.....	18
4.1.5 Pentaho Data Intergration	19
4.2 Cài đặt.....	20
4.3 Kết quả.....	21

4.4 Minh họa các chức năng chính	22
4.5 Một số quy trình nghiệp vụ mới để vận hành MDS	24
4.5.1 Quy trình xây dựng phân tích, thiết kế, xây dựng yêu cầu quản trị miền dữ liệu.....	24
4.5.2 Quy trình thu thập chuẩn hóa dữ liệu dùng chung đã có	25
4.5.3 Quy trình khai báo dữ liệu dùng chung mới	25
4.5.4 Quy trình chia sẻ dữ liệu dùng chung cho các phần mềm	26
4.5.5 Quy trình công bố dữ liệu	26
CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	27
5.1 Kết luận	27
5.2 Hướng phát triển.....	27
TÀI LIỆU THAM KHẢO.....	29
PHỤ LỤC.....	30
A. DANH SÁCH ĐỐI TƯỢNG DỮ LIỆU THEO VĂN BẢN 234/NQ- HĐTV	30

DANH MỤC HÌNH VẼ

Hình 1.1	Cơ cấu tổ chức của EVN	1
Hình 2.1	Mô hình lấy dữ liệu từ nguồn	6
Hình 2.2	Phân cấp tổ chức tại EVN	7
Hình 3.1	Mô hình kiến trúc	9
Hình 3.2	Mô hình thu thập từ nguồn về stagging	11
Hình 3.3	Luồng ETL từ stagging vào trong MDS	11
Hình 3.4	Mô hình liên kết giữa các danh mục	12
Hình 3.5	Cấu trúc thực thể Chức danh	12
Hình 3.6	Cấu trúc thực thể Đơn vị	13
Hình 3.7	Cấu trúc thực thể Khối đơn vị	13
Hình 3.8	Cấu trúc thực thể Loại hình đơn vị	13
Hình 3.9	Cấu trúc thực thể Nhân sự	13
Hình 3.10	Cấu trúc thực thể phòng ban	14
Hình 3.11	Cấu trúc thực thể Tiểu ban	14
Hình 4.1	Kiến trúc của Master Data Services	16
Hình 4.2	Mô hình Kafka tổng quát	17
Hình 4.3	Mô hình dữ liệu di chuyển từ nguồn tới Kafka	18
Hình 4.4	Mô hình kết nối Debezium với Kafka Connect	19
Hình 4.5	Giao diện trang chủ	21
Hình 4.6	Mẫu báo cáo tổng quan cơ cấu lao động	22
Hình 4.7	Giao diện trang quản lý model	22
Hình 4.8	Giao diện trang quản lý người dùng và phân quyền	23
Hình 4.9	Giao diện trang quản lý thực thể	24

DANH MỤC BẢNG BIỂU

Bảng 2.1	Bảng mô tả bảng dữ liệu chứa thông tin nhân sự	4
Bảng 2.2	Bảng mô tả bảng dữ liệu cần thu thập của nhân sự	4
Bảng 2.3	Bảng Đơn vị	5
Bảng 2.4	Bảng Khối đơn vị/Loại hình đơn vị	5
Bảng 4.1	Danh sách công cụ sử dụng	20
Bảng A.1	Danh sách đối tượng dùng chung theo quy định 234/NQ-HĐTV	30
Bảng A.2	Danh sách đối tượng dùng chung theo quy định 234/NQ-HĐTV	31
Bảng A.3	Danh sách đối tượng dùng chung theo quy định 234/NQ-HĐTV	32
Bảng A.4	Danh sách đối tượng dùng chung theo quy định 234/NQ-HĐTV	33
Bảng A.5	Danh sách đối tượng dùng chung theo quy định 234/NQ-HĐTV	34
Bảng A.6	Danh sách đối tượng dùng chung theo quy định 234/NQ-HĐTV	35
Bảng A.7	Danh sách đối tượng dùng chung theo quy định 234/NQ-HĐTV	36

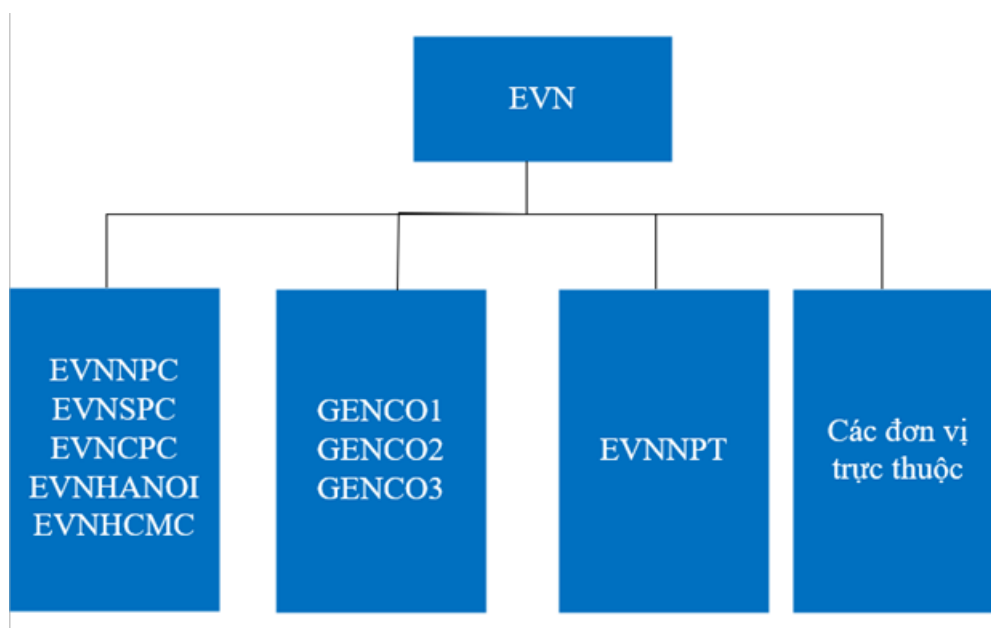
DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT

Thuật ngữ	Ý nghĩa
API	Giao diện lập trình ứng dụng (Application Programming Interface)
CMIS	Hệ thống Thông tin quản lý khách hàng
CSDL	Cơ sở dữ liệu
Digital Office	Hệ thống Văn phòng số
ETL	Extract Transform Load - cung cấp các dịch vụ tích hợp (luân chuyển, làm sạch, chuyển đổi) giữa các CSDL
EVN	Tập đoàn Điện lực Việt Nam
EVNICT	Công ty Viễn thông Điện lực và Công nghệ thông tin
HRMS	Hệ thống Quản lý nguồn nhân lực
HTML	Ngôn ngữ đánh dấu siêu văn bản (HyperText Markup Language)
IMIS	Hệ thống Quản lý đầu tư xây dựng
MDS	Master Data Services – Dịch vụ quản lý dữ liệu dùng chung
PMDC	Phần mềm dùng chung

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI

1.1 Bối cảnh

Hiện nay tại Tập đoàn Điện lực Việt Nam đang sử dụng nhiều phần mềm dung chung (Các phần mềm dùng tại EVN và tất cả các đơn vị). Các phần mềm này được triển khai tại các đơn vị và dữ liệu của mỗi đơn vị được lưu trữ tại chính đơn vị đó.



Hình 1.1: Cơ cấu tổ chức của EVN

Do việc triển khai tại nhiều vị trí khác nhau nên việc có rất nhiều danh mục dữ liệu có thể sử dụng chung nhưng lại chưa có sự nhất quán về quản lý, dẫn đến việc thông tin còn chưa đồng nhất về một đối tượng, hiểu nhầm cùng một đối tượng là 2 hay nhiều đối tượng khác nhau. Ví dụ tên cùng tên phòng ban là Tài chính kế toán nhưng vai trò chức năng tại hai đơn vị khác nhau là khác nhau. Việc đánh mã các đối tượng tại các đơn vị đã có nhưng chưa có sự thống nhất ví dụ 2 nhân sự tại 2 đơn vị khác nhau lại cùng một mã nhân sự. Hiện tại số lượng đối tượng dữ liệu theo danh mục các dữ liệu dùng chung và dữ liệu chủ (QĐ 234 của EVN) đang được lưu trữ, quản lý rất lớn (70 triệu đối tượng) bao gồm 59 danh mục và chia thành 2 loại: tham chiếu hoặc dữ liệu chủ đã được xác định trong phần phụ lục 1 của đề án này. Vì vậy EVN muốn thu thập các dữ liệu danh mục này về một nơi để thực hiện quản lý tập chung, xây dựng báo cáo và phục vụ công tác đánh mã đối tượng trong các danh mục này theo một quy tắc thống nhất dựa theo văn bản 1188/QĐ-EVN.

1.2 Mục tiêu và phạm vi đề tài

Đề tài này với mục tiêu xây dựng một kho dữ liệu dùng chung thử nghiệm để EVN quản lý nhằm mục đích chia sẻ các danh mục đã được chuẩn hóa. Các nội dung công việc chính được thực hiện gồm:

- Quản lý danh mục đối với các đối tượng dữ liệu dùng chung.
- Xây dựng công cụ thu thập dữ liệu từ đơn vị đối với các danh mục này.
- Chia sẻ danh mục dữ liệu dùng chung cho các hệ thống khác.

Nội dung công việc sẽ được triển khai tại EVN đối với dữ liệu là các dữ liệu gốc, dữ liệu tham chiếu, bao gồm 7 danh mục:

- Nhân sự
- Đơn vị
- Chức danh
- Phòng/ban
- Khối đơn vị
- Loại hình đơn vị
- Tiểu ban/hội đồng

1.3 Bố cục đồ án

Phần còn lại của báo cáo đồ án tốt nghiệp được tổ chức như sau:

- Chương 2: Khảo sát và phân tích yêu cầu. Chương này em sẽ tập trung làm rõ, phân tích các yêu cầu, từ đó xác định ra định hướng để giải quyết bài toán
- Chương 3: Xây dựng giải pháp. Trong chương này, em sẽ trình bày về kiến trúc hệ thống em xây dựng, cách giải quyết các vấn đề đã phân tích ở chương 2.
- Chương 4: Triển khai cài đặt và đánh giá thực nghiệm: Chương này em sẽ trình bày các công nghệ em sử dụng, kết quả cài đặt hệ thống và các quy trình nghiệp vụ mới
- Chương 5: Kết luận và hướng phát triển.