HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Họ tên học viên: Trần Đức Long

TÍCH HỢP DỮ LIỆU CHO HỆ THỐNG QUẢN LÝ CÁN BỘ, VIÊN CHỨC CỦA VIỆN NĂNG LƯỢNG NGUYÊN TỬ

CHUYÊN NGÀNH: HỆ THỐNG THÔNG TIN

MÃ SỐ: 8.48.01.04 (Hệ thống thông tin)

TÓM TẮT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. Trần Thiện Chính

HÀ NỘI - 2025

Đề án tốt nghiệp được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: TS. Trần Thiện Chính

Phản biện 1: ……

Phản biện 2: ……

Đề án tốt nghiệp sẽ được bảo vệ trước Hội đồng chấm đề án tốt nghiệp thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: ..... giờ ..... ngày ..... tháng .... năm .....

Có thể tìm hiểu đề án tốt nghiệp tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông.

MỞ ĐẦU

1. Lý do chọn đề tài

Hiện nay, trước bối cảnh cuộc Cách mạng công nghiệp lần thứ Tư và xu thế chuyển đổi số đang diễn ra mạnh mẽ, rộng khắp trong mọi mặt của nền kinh tế - xã hội tại Việt Nam nói riêng và trên cả thế giới nói chung. Để tiến hành triển khai nhiệm vụ quan trọng này, Ngày 25/6/2020 Thủ tướng chính phủ đã ra Quyết định 893/QĐ-TTg về việc Phê duyệt Đề án xây dựng cơ sở dữ liệu quốc gia về cán bộ, công chức, viên chức trong các cơ quan nhà nước [1]. Tại Quyết định này nêu rõ quan điểm, mục tiêu, và yêu cầu đối với hệ thống quản lý cơ sở dữ liệu quốc gia do Bộ Nội Vụ khai thác, quản lý và sử dụng.

Viện Năng lượng nguyên tử Việt Nam (tên tiếng Anh: Vietnam Atomic Energy Institute, tên viết tắt là VINATOM là tổ chức sự nghiệp khoa học hạng đặc biệt trực thuộc Bộ Khoa học và Công nghệ (KH&CN). Cơ cấu tổ chức của Viện hiện nay gồm 03 đơn vị chức năng, 03 viện nghiên cứu, 05 trung tâm nghiên cứu tổng cộng hơn 800 cán bộ đang làm việc.

Theo Kế hoạch 141/KH-BCĐTKNQ18 ngày 06/12/2024 của Ban Chỉ đạo về tổng kết việc thực hiện Nghị quyết số 18-NQ/TW về Kế hoạch định hướng sắp xếp, tinh gọn tổ chức bộ máy của Chính phủ có nêu về định hướng cơ cấu, sắp xếp và hợp nhất 14 bộ, cơ quan ngang Bộ. Cụ thể theo Kế hoạch, "Hợp nhất Bộ Thông tin và Truyền thông và Bộ Khoa học và Công nghệ, thực hiện chức năng quản lý nhà nước về các ngành, lĩnh vực hiện đang giao cho Bộ Thông tin và Truyền thông và Bộ Khoa học và Công nghệ. Sau hợp nhất Bộ mới có tên là Bộ Khoa học và Công nghệ". Chính vì thế nên số lượng cơ sở dữ liệu của cả hai Bộ ngành sau sát nhập bị nhân lên, chồng chéo và khiến việc truy xuất và quản lý dữ liệu phức tạp hơn.

Chính vì lý do thực tế cấp thiết này, học viên đề xuất đưa ra giải pháp Tích hợp dữ liệu – Data Integration để kết hợp các dữ liệu không đồng nhất từ nhiều nguồn khác nhau thành một tập hợp dữ liệu thống nhất. Điều này sẽ cho phép công tác khai thác dữ liệu từ nhiều nguồn hiện hữu một cách dễ dàng và hiệu quả, chính xác.

Xuất phát từ những vấn đề nêu trên học viên đã chọn đề tài **"Tích hợp dữ liệu cho hệ thống quản lý cán bộ, viên chức của Viện Năng lượng nguyên tử"**

Bố cục của luận văn này kết cấu gồm 3 chương, đi từ việc trình bày lý thuyết, đặc thù công việc đến thực tiễn cách tiếp cận vấn đề và triển khai công cụ phần mềm, cụ thể như sau:

- Phần mở đầu
- Phần nội dung luận văn gồm:

Chương 1. Tổng quan về vấn đề nghiên cứu

- Giới thiệu về Viện Năng lượng nguyên tử Việt Nam
- Giới thiệu về Hệ thống quản lý công chức, viên chức trong cơ quan nhà nước Việt Nam
- Giới thiệu về các hệ thống quản lý cán bộ đã được triển khai tại đơn vị

Chương 2. Các kỹ thuật tích hợp dữ liệu

- Cách tiếp cận vấn đề xử lý dữ liệu
- Lý thuyết về Tích hợp dữ liệu
- Các kỹ thuật tích hợp dữ liệu
- Đề xuất công nghệ áp dụng

Chương 3. Triển khai áp dụng, thử nghiệm, đánh giá

2. Mục đích nghiên cứu:

Mục đích chính của luận văn là nghiên cứu về cách thức quản lý cơ sở dữ liệu cán bộ, viên chức và người lao động tại Viện Năng lượng nguyên tử Việt Nam trong bối cảnh đất nước đổi mới, đẩy mạnh ứng dụng phát triển ứng dụng công nghệ thông tin và chuyển đổi số. Học viên muốn thông qua luân văn này trình bày cụ thể tình huống thực tế về bài toán quản lý và khai thác cơ sở dữ liệu về cán bộ và người lao động tại đơn vị trước và sau sáp nhập cơ quan chủ quản là Bộ Khoa học và Công nghệ.

Mục tiêu tiếp theo của Luận văn sẽ tìm hiểu lý thuyết và các công trình nghiên cứu về Data Integration và Data Warehouse nhằm giải quyết bài toán được đặt ra. Từ đó tiến tới giới thiệu về các phương pháp tích hợp dữ liệu phổ biến hiện hành.

Cuối cùng, mục tiêu chủ đạo của nghiên cứu là đưa ra một mô hình tích hợp dữ liệu phù hợp để áp dụng vào hệ thống quản lý dữ liệu về viên chức và người lao động hiện hành nhằm thoả mãn và giải quyết các khó khăn và vướng mắc hiện có. Ngoài ra nghiên cứu cũng đưa ra các ví dụ trực quan như hình ảnh, biểu đồ và mô phỏng các thao tác, hướng dẫn, dễ đọc và dễ hiểu cho cán bộ thực hiện công tác chuyên môn.

Nghiên cứu này hướng tới có thể góp phần hỗ trợ xây dựng hệ thống quản lý cán bộ, công chức, viên chức, người lao động hiện hành và được triển khai thử nghiệm thực tế tại Viện Năng lượng nguyên tử Việt Nam, tiến tới trong Bộ KH&CN.

3. Đối tượng và phạm vi nghiên cứu:

- Luận văn tập trung nghiên cứu về lý thuyết Tích hợp dữ liệu, các cách thức triển khai tích hợp dữ liệu và nhu cầu thực tế về việc áp dụng kỹ thuật Data Intergration trong hệ thống dữ liệu về quản lý cán bộ, viên chức và người lao động tại Viện Năng lượng nguyên tử Việt Nam.

- Giới hạn phạm vi nghiên cứu của đề tài: Để đảm bảo tính khả thi và thực tế của luận văn, tác giả tập trung vào các phương pháp và cách triển khai tích hợp dữ liệu ngắn gọn, phù hợp với kinh phí và thân thiện với người dùng nhất đồng thời bám sát vào các nhu cầu thực tế tại đơn vị áp dụng là Viện Năng lượng nguyên tử.

4. Phương pháp nghiên cứu, cách tiếp cận:

- *Nghiên cứu lý thuyết:* Luận văn lấy lý thuyết Tích hợp dữ liệu làm nền tảng và được định hình xuyên suốt trong toàn bộ nghiên cứu. Tiến tới tiến hành việc nghiên cứu hệ thống văn bản quy phạm pháp luật và các văn bản liên quan đến dữ liệu quản lý cán bộ, công chức, viên chức, người lao động trong cơ quan. Nghiên cứu tài liệu, bài báo về Data Integration và Data Warehouse. Các phương pháp Tích hợp dữ liệu, các mô hình triển khai Tích hợp dữ liệu.

- *Nghiên cứu thực tiễn:* Từ nền tảng lý thuyết được nghiên cứu, Luận văn tiếp tục nghiên cứu các cách thức đưa phương pháp Tích hợp dữ liệu và thử nghiệm trên một số cơ sở dữ liệu được xây dựng nhằm mô phỏng lại hệ thống quản lý cán bộ, công chức, viên chức và người lao động tại Viện Năng lượng nguyên tử Việt Nam nhằm giải quyết các khó khăn, rào cản hiện hữu.

- *Cách tiếp cận nghiên cứu:* Dựa trên cơ sở các kiến thức đã được trang bị trong quá trình học tập, nghiên cứu, đề xuất xây dựng, thiết kế tích hợp dữ liệu quản lý cán bộ, công chức, viên chức, người lao động của Viện Năng lượng nguyên tử Việt Nam phù hợp, thích ứng với yêu cầu thực tiễn và quy định của pháp luật có liên quan. Từ đó tiến tới áp dụng vào thực tế với cơ sở dữ liệu về công chức, viên chức và người lao động tại Viện Năng lượng nguyên tử Việt Nam. Song song với đó là học hỏi, khảo sát và phản hồi lại ý kiến của các cán bộ, đồng nghiệp có liên quan đến công tác khai thác và sử

dụng hệ thống quản lý cán bộ nhằm cải thiện và hiệu chỉnh mô hình Tích hợp dữ liệu sao cho thân thiện, hiệu quả, kinh tế và có tính ứng dụng cao.

Kết quả dự kiến thu được:

- Tóm tắt các nghiên cứu và lý thuyết liên quan đến Tích hợp dữ liệu và các mô hình triển khai Tích hợp dữ liệu hiện hành.

- Xây dựng phương án tích hợp dữ liệu quản lý cán bộ, công chức, viên chức, người lao động tại Viện Năng lượng nguyên tử Việt Nam, có khả năng tổng hợp dữ liệu từ nhiều dạng, nhiều nguồn, lưu trữ dữ liệu tổ chức cán bộ một cách chính xác, hiệu quả, bảo mật quyền riêng tư, hiển thị dữ liệu đúng với yêu cầu truy xuất, dễ dàng sử dụng thông tin. Đảm bảo đáp ứng yêu cầu dữ liệu "đúng, đủ, sạch, sống, thống nhất, dùng chung".

- Khả năng và địa chỉ ứng dụng: Có khả năng ứng dụng trong Viện Năng lượng nguyên tử Việt Nam, giúp giảm thiểu rất nhiều các công tác nhập, xuất dữ liệu đối với từng cá nhân cán bộ, viên chức và người lao động tại đơn vị. Đặt nền móng cho bài toán tích hợp dữ liệu để ứng dụng vào hệ thống cơ sở dữ liệu của Bộ Khoa học và Công nghệ về công chức, viên chức và người lao động

- Dự kiến hiệu quả mang lại: Giúp nâng cao hiệu quả của công tác tổ chức cán bộ, có khả năng kết nối với trục liên thông kết nối, chia sẻ dữ liệu cấp Bộ (của Bộ KH&CN) đảm bảo yêu cầu về bảo mật và an toàn thông tin, giúp tiết kiệm thời gian tra cứu và giảm bớt yêu cầu về lưu trữ.

CHƯƠNG 1: TỔNG QUAN VỀ VẤN ĐỀ NGHIÊN CỨU

Tích hợp dữ liệu hay còn được biết đến là Data Integration là một khái niệm quan trọng trong việc giải quyết bài toán quản lý và khai thác dữ liệu lớn (Big Data) trong thời đại 4.0. Thách thức đặt ra trong việc xử lý dữ liệu Big Data đòi hỏi các cơ sở dữ liệu phải được gom lại để xử lý hiệu quả và Tích hợp dữ liệu là kỹ thuật hoàn hảo để giải quyết vấn đề đó. Hiện tại có ba hướng giải pháp chính cho bài toán quản lý và khai thác dữ liệu hiệu quả hiện nay thường được sử dụng là Tích hợp dữ liệu (Data Integration), Hợp nhất dữ liệu (Data Consolidation) và Tổng hợp dữ liệu (Data Aggregation) tuỳ vào nhu cầu sử dụng và trường hợp sử dụng mà người dùng có thể chọn phương án thích hợp. Mối quan hệ giữa ba quá trình này mang tính tuần tự và bổ trợ lẫn nhau trong một quy trình tổng thể về quản trị liên thông dữ liệu: Tích hợp dữ liệu => Tổng hợp dữ liệu => Hợp nhất dữ liệu. Tuy nhiên tùy thuộc vào mục đích sử dụng và yêu cầu về dữ liệu đầu ra và nguồn kinh phí mà người dùng có thể chọn phương án quản lý dữ liệu hợp lý nhất. Trong ngữ cảnh rộng hơn thì thuật ngữ "Tích hợp dữ liệu" thường được dùng để chỉ toàn bộ các kỹ thuật này.

Nghiên cứu xây dựng và đề xuất phương án Tích hợp dữ liệu nhằm giải quyết bài toán về khai thác và sử dụng cơ sở dữ liệu cán bộ, viên chức, người lao động tại Viện Năng lượng nguyên tử Việt Nam sau sáp nhập.

1.1. Cách tiếp cận vấn đề quản lý và khai thác dữ liệu

1.1.1. Tích hợp dữ liệu (Data Integration)

Tích hợp dữ liệu hay Data Integration thường được sử dụng để đề cập đến quá trình kết hợp dữ liệu từ nhiều nguồn, nhiều cơ sở dữ liệu khác nhau và rời rạc trở thành một thể thống nhất (thường là một cơ sở dữ liệu) nhằm đảm bảo tính toàn vẹn của dữ liệu đồng thời cung cấp khả năng truy xuất dữ liệu đồng bộ. Ứng dụng điển hình của quá trình này là giúp giảm thiểu phân mảnh dữ liệu, hỗ trợ truy vấn và khai thác dữ liệu cụ thể trong các nền tảng Dữ liệu lớn (Big Data) như Kho dữ liệu (Data warehouse), Hồ dữ liệu (Data lake) hoặc các hệ thống Internet vạn vật (Internet of Things)…

1.2. Giới thiệu về Viện Năng lượng nguyên tử Việt Nam

Viện Năng lượng nguyên tử Việt Nam là tổ chức sự nghiệp khoa học và công nghệ hạng đặc biệt trực thuộc Bộ Khoa học và Công nghệ có chức năng giúp Bộ trưởng nghiên cứu cơ bản, nghiên cứu ứng dụng và triển khai các hoạt động ứng dụng kết quả nghiên cứu trong lĩnh vực năng lượng nguyên tử; hỗ trợ kỹ thuật phục vụ công tác quản lý nhà nước về năng lượng nguyên tử, an toàn bức xạ và hạt nhân; tổ chức hoạt động đào tạo, dịch vụ trong lĩnh vực năng lượng nguyên tử theo quy định của pháp luật.

1.2.1. Công tác quản lý cán bộ tại Viện

Công tác tổ chức cán bộ là một phần quan trọng trong công tác quản lý nhân sự của mọi tổ chức, cơ quan, viện nghiên cứu, doanh nghiệp hay các tổ chức chính trị, xã hội và tại Viện NLNTVN cũng không phải là ngoại lệ. Trong các Viện nghiên cứu như Viện Năng lượng nguyên tử, công tác tổ chức cán bộ đóng vai trò thiết yếu trong việc bảo đảm sự vận hành ổn định, hiệu quả, đồng thời tạo ra các điều kiện để đội ngũ cán bộ, viên chức có thể phát huy tối đa năng lực, đóng góp cho sự phát triển chung của tổ chức.

Một hệ thống quản lý cơ sở dữ liệu về cán bộ trong mọi cơ quan, tổ chức là cần thiết để khiến cho người quản lý có được một cái nhìn toàn diện về người làm việc. Yêu cầu cấp thiết với công tác cán bộ tại Viện là có một hệ thống dữ liệu tổng hợp về thông tin chuyên ngành; độ tuổi; sở trường công tác; địa chỉ; chức vụ của toàn bộ cán bộ thuộc Viện trên khắp toàn quốc tập trung tại cùng trên một hệ thống là vô cùng cấp thiết. Hệ thống này sẽ giúp người quản trị con người có cái nhìn tổng quan về nguồn lực nhân sự tại đơn vị, giúp phân bổ, điều chuyển công tác phù hợp với chuyên môn nhân sự và đặc thù công việc yêu cầu, góp phần thực hiện các nhiệm vụ quan trọng của Viện trong thời đại mới của đất nước.

1.2.2. Hệ thống quản lý công chức, viên chức trong cơ quan nhà nước Việt Nam

Hệ thống giúp các đơn vị có thẩm quyền quản lý cán bộ công chức, viên chức quản lý hồ sơ nhân sự hiệu quả, giảm chi phí, thời gian, công sức, thuận tiện trong việc tra cứu tìm kiếm hồ sơ cán bộ công chức, viên chức. Công tác này yêu cầu cần được thực hiện nhiều tầng, nhiều lớp tuỳ theo cấp độ phân cấp quản lý, mỗi cấp tương ứng với một đơn vị, một cấp quản lý nên yêu cầu phải có hệ thống riêng biệt. Để phục vụ

công tác nhân sự tại Viện nói riêng và trên toàn quốc nói chung thì yêu cầu xuất dữ liệu và nạp dữ liệu giữa các hệ thống là vô cùng cần thiết.

Theo đề nghị của Bộ trưởng Bộ Thông tin và Truyền thông, Thủ tướng Chính phủ đã ra Quyết định số 1819/QĐ-TTg ngày 26 tháng 10 năm 2015 về việc Phê duyệt chương trình quốc gia về ứng dụng công nghệ thông tin trong hoạt động của cơ quan nhà nước giai đoạn 2016 – 2020.

Theo đề nghị của Bộ Nội Vụ, vào ngày 25 tháng 6 năm 2020, Thủ tướng Chính phủ đã ban hành Quyết định số 893/QĐ-TTg về việc phê duyệt đề án xây dựng cơ sở dữ liệu quốc gia về cán bộ, công chức, viên chức trong các cơ quan nhà nước. Quyết định này được đưa ra với quan điểm "Xây dựng và phát triển cơ sở dữ liệu quốc gia về cán bộ, công chức, viên chức trong các cơ quan nhà nước phải phù hợp, thống nhất, đồng bộ với quá trình xây dựng Chính phủ điện tử/Chính quyền điện tử; đảm bảo thống nhất về quản lý cán bộ, công chức, viên chức và tổ chức bộ máy, biên chế; thông tin cơ bản về cán bộ, công chức và tổ chức bộ máy, biên chế của các cơ quan nhà nước được quản lý tập trung, thống nhất trong cơ sở dữ liệu".

Theo Kế hoạch 141/KH-BCĐTKNQ18 năm 2024 định hướng sắp xếp, tinh gọn tổ chức bộ máy của Chính phủ có nêu về định hướng cơ cấu, sắp xếp và hợp nhất 14 bộ, cơ quan ngang Bộ. Hiện tại, cơ cấu tổ chức của bộ máy Chính phủ tại nước ta sau sắp xếp gồm 14 Bộ và 3 Cơ quan ngang Bộ. Chính vì thế nên gây ra tình trạng cơ sở dữ liệu về nhân sự vừa thừa, vừa thiếu, lại chồng chéo về mặt dữ liệu và hệ thống quản lý cán bộ, viên chức và người lao động tại Viện Năng lượng nguyên tử Việt Nam cũng không phải là ngoại lệ.

1.3. Các hệ thống quản lý cán bộ, công chức, viên chức đã được triển khai tại Viện Năng lượng nguyên tử

1.3.1. Hệ thống quản lý cán bộ, công chức, viên chức của Bộ KH&CN xây dựng năm 2020

Để có thể xây dựng được cơ sở dữ liệu đầy đủ, chính xác về đội ngũ cán bộ, công chức, viên chức; sử dụng thông tin cán bộ, công chức, viên chức để phục vụ cho việc tra cứu, xây dựng các báo cáo, thống kê, kiểm soát chặt chẽ số lượng cán bộ, công chức, viên chức, Bộ KH&CN đã ban hành Quyết định 1057/QĐ-BKHCN vào ngày 23 tháng 4 năm 2020 về việc công bố kế hoạch triển khai phần mềm quản lý cán bộ công chức,

viên chức và người lao động tại Bộ KH&CN năm 2020. Hệ thống này được truy cập tại địa chỉ: "https://qlcb.most.gov.vn/Pages/default.aspx".

1.3.2. Hệ thống quản lý cán bộ, công chức, viên chức của Bộ KH&CN năm 2024

Thực hiện Quyết định số 893/QĐ-TTg ngày 25/6/2020 của Thủ tướng Chính phủ, Bộ Nội Vụ đã ban hành Công văn Số 4659/BNV-VP ngày 06 tháng 8 năm 2024 về việc bổ sung và làm sạch dữ liệu CBCCVC. Tuy nhiên Hệ thống đang sử dụng từ 2020 gặp xung đột với định dạng ký tự với hệ thống dữ liệu của Bộ Nội Vụ dẫn đến không thể đồng bộ được dữ liệu cán bộ.

Thực hiện chỉ đạo của Bộ trưởng Bộ KH&CN, Trung tâm Công nghệ thông tin Bộ đã tiến hành triển khai phần mềm mới đồng thời phối hợp với Vụ Tổ chức cán bộ, tiến hành rà soát, cập nhật đầy đủ các trường thông tin theo yêu cầu của Bộ Nội Vụ và đồng bộ với CSDL quốc gia về CBCCVC. Địa chỉ phần mềm mới tại https://qlccvc.most.gov.vn/Pages/default.aspx đối với cán bộ được phân quyền quản lý và https://ccvc.most.gov.vn/Pages/default.aspx đối với cán bộ tự khai.

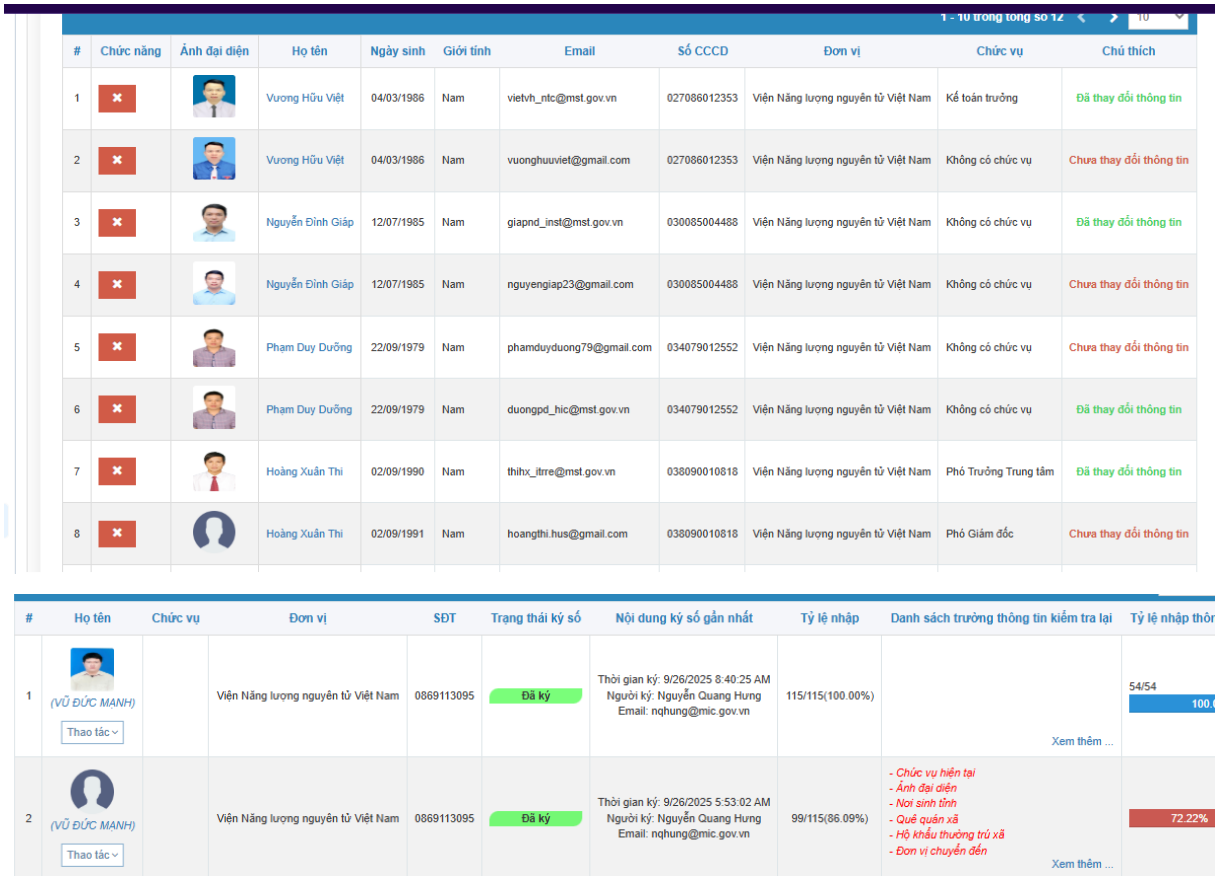
Điểm khác biệt chính là hệ thống này sử dụng một hệ thống cơ sở dữ liệu nguồn hoàn toàn khác, không kế thừa được từ các hệ thống cơ sở dữ liệu cũ do khác định dạng dữ liệu dẫn đến không thể đồng bộ được dữ liệu theo yêu cầu của Bộ Nội Vụ.

1.3.3. Hệ thống quản lý cán bộ, công chức, viên chức của Bộ KH&CN (sau sắp xếp) năm 2025

Theo Kế hoạch 141/KH-BCĐTKNQ18 năm 2024, "Hợp nhất Bộ Thông tin và Truyền thông và Bộ KH&CN" và điều này trực tiếp tác động tới Viện Năng lượng nguyên tử bởi việc thay đổi cơ quan chủ quản. Thực tế phát sinh ra vấn đề về nhiều cơ sở dữ liệu rời rạc thuộc hai Bộ ngành trước sáp nhập và sau sáp nhập.

Trong quá trình triển khai, cơ quan chủ quản mới là Bộ KH&CN (sau sáp nhập) kết luận hệ thống về cơ sở dữ liệu của hai Bộ được xây dựng khác nhau và không có định hướng kết hợp từ đầu nên xảy ra xung đột dữ liệu trong giai đoạn đồng bộ giữa hai hệ thống. Viện được sự hướng dẫn và hỗ trợ của Vụ Tổ chức cán bộ và Trung tâm Thông tin Bộ đã tiến hành triển khai lại việc kê khai lại hồ sơ cán bộ theo như yêu cầu.

Thực tế hiện nay Viện Năng lượng nguyên tử đã và đang triển khai, hướng dẫn đến từng cán bộ, khởi tạo lại tài khoản, tiến hành kê khai lại thông tin, quá trình công tác như đã từng làm tại Hệ thống QLCBCCVC trước đó. Toàn bộ công việc này đang được Viện tiến hành một cách thủ công, điều đó dẫn đến những bất cập, khó khăn trong việc triển khai.



#	Chức năng	Ảnh đại diện	Họ tên	Ngày sinh	Giới tính	Email	Số CCCD	Đơn vị	Chức vụ	Chú thích
1			Vương Hữu Việt	04/03/1986	Nam	vietvh_ntc@mst.gov.vn	027086012353	Viện Năng lượng nguyên tử Việt Nam	Kế toán trưởng	Đã thay đổi thông tin
2			Vương Hữu Việt	04/03/1986	Nam	vuonghuuviet@gmail.com	027086012353	Viện Năng lượng nguyên tử Việt Nam	Không có chức vụ	Chưa thay đổi thông tin
3			Nguyễn Đình Giáp	12/07/1985	Nam	giapnd_inst@mst.gov.vn	030085004488	Viện Năng lượng nguyên tử Việt Nam	Không có chức vụ	Đã thay đổi thông tin
4			Nguyễn Đình Giáp	12/07/1985	Nam	nguyengiap23@gmail.com	030085004488	Viện Năng lượng nguyên tử Việt Nam	Không có chức vụ	Chưa thay đổi thông tin
5			Phạm Duy Dưỡng	22/09/1979	Nam	phamduyduong79@gmail.com	034079012552	Viện Năng lượng nguyên tử Việt Nam	Không có chức vụ	Chưa thay đổi thông tin
6			Phạm Duy Dưỡng	22/09/1979	Nam	duongpd_hic@mst.gov.vn	034079012552	Viện Năng lượng nguyên tử Việt Nam	Không có chức vụ	Đã thay đổi thông tin
7			Hoàng Xuân Thi	02/09/1990	Nam	thihx_itrre@mst.gov.vn	038090010818	Viện Năng lượng nguyên tử Việt Nam	Phó Trưởng Trung tâm	Đã thay đổi thông tin
8			Hoàng Xuân Thi	02/09/1991	Nam	hoangthi.hus@gmail.com	038090010818	Viện Năng lượng nguyên tử Việt Nam	Phó Giám đốc	Chưa thay đổi thông tin

#	Họ tên	Chức vụ	Đơn vị	SĐT	Trạng thái ký số	Nội dung ký số gần nhất	Tỷ lệ nhập	Danh sách trường thông tin kiểm tra lại	Tỷ lệ nhập thôi
1	(VŨ ĐỨC MẠNH) Thao tác		Viện Năng lượng nguyên tử Việt Nam	0869113095	Đã ký	Thời gian ký: 9/26/2025 8:40:25 AM Người ký: Nguyễn Quang Hưng Email: nqhung@mic.gov.vn	115/115(100.00%)	Xem thêm ...	54/54 100.(
2	(VŨ ĐỨC MẠNH) Thao tác		Viện Năng lượng nguyên tử Việt Nam	0869113095	Đã ký	Thời gian ký: 9/26/2025 5:53:02 AM Người ký: Nguyễn Quang Hưng Email: nqhung@mic.gov.vn	99/115(86.09%)	- Chức vụ hiện tại - Ảnh đại diện - Nơi sinh tỉnh - Quê quán xã - Hộ khẩu thường trú xã - Đơn vị chuyển đến Xem thêm ...	72.22%

Hình 1.1 Lỗi lặp hồ sơ trong quá trình đồng bộ sau sáp nhập Bộ

Qua quá trình nghiên cứu, thảo luận cùng với sự giúp đỡ, hỗ trợ nhiệt tình từ phía các cán bộ chuyên môn ở cả Bộ KH&CN trước và sau sáp nhập, học viên đã phần nào xác định được vấn đề kỹ thuật khiến phát sinh các vấn đề trên trong quy trình đồng bộ dữ liệu đã được triển khai này. Nguyên nhân cụ thể và đề xuất kỹ thuật về hướng giải quyết sẽ được trình bày và giải quyết trong chương 3 của luận văn này.

Xuất phát từ những bất cập cụ thể trong công tác của học viên trong việc quản lý, vận hành và khai thác hệ thống QLCBCCVC của Viện Năng lượng nguyên tử Việt Nam cũng như của cơ quan chủ quản là Bộ KH&CN (sau sáp nhập), tác giả đề xuất giải pháp **"Tích hợp dữ liệu cho hệ thống quản lý cán bộ, viên chức của Viện Năng lượng nguyên tử"**

CHƯƠNG 2: CÁC KỸ THUẬT TÍCH HỢP DỮ LIỆU

2.1. Lý thuyết về Tích hợp dữ liệu - Data Intergration

2.1.1. Tích hợp dữ liệu trên thế giới

Vấn đề tích hợp dữ liệu ngày càng được quan tâm trong cả giới nghiên cứu kỹ thuật và cả giới kinh doanh bởi sự bùng nổ về khối lượng dữ liệu và nhu cầu chia sẻ, khai thác những dữ liệu sẵn có. Để có thể cung cấp các dịch vụ, dữ liệu và khai thác thông tin một cách chính xác và tối ưu nhất đến người sử dụng thì việc ứng dụng tích hợp dữ liệu là cần thiết hơn bao giờ hết.

Lý thuyết về Tích hợp dữ liệu luôn là một chủ đề khiến các nhà nghiên cứu, các tác giả trong và ngoài nước quan tâm sâu sắc và luôn là một đề tài gây tranh cãi. Dựa trên những nghiên cứu tiền nhiệm, thì ta có thể thấy thuật ngữ này thường được dùng để mô tả quy trình liên quan đến việc tổng hợp, kết hợp dữ liệu có thể khai được tại các nguồn dữ liệu khác nhau để cung cấp cho người quản trị dữ liệu có được một cái nhìn thống nhất về chúng. Nhu cầu về Tích hợp dữ liệu luôn tăng cùng với nhu cầu về xử lý, chia sẻ và khai thác Dữ liệu lớn bùng nổ trong bối cảnh hiện tại.

2.1.2. Minh họa sự cần thiết của việc Tích hợp dữ liệu tại Viện Năng lượng nguyên tử

Viện Năng lượng nguyên tử Việt Nam là một tổ chức nghiên cứu khoa học đồng thời vừa đảm nhiệm sản xuất, kinh doanh các sản phẩm, tổ chức hội nghị và đào tạo tiến sĩ trong lĩnh vực nguyên tử. Viện Năng lượng nguyên tử gồm nhiều đơn vị trực thuộc, phụ trách những mảng công việc cụ thể và có trụ sở tại khắp mọi miền tổ quốc từ Bắc vào Nam.

Để có thể có cái nhìn thống nhất và khách quan nhất về hơn 800 cán bộ đang công tác và làm việc tại Viện, việc Tích hợp dữ liệu cán bộ giữa các đơn vị trực thuộc trên toàn quốc là nhu cầu cấp thiết. Kỹ thuật Tích hợp dữ liệu trong thời đại dữ liệu ngày càng nhiều và càng phân tán giúp giải quyết triệt để vấn đề trùng lặp dữ liệu trong công tác quản lý nhân sự.

Chính vì những lý do trên nên số lượng nguồn dữ liệu liên quan đến nhân sự tại một tập thể, công ty đặc biệt là tại cơ quan quản lý nhà nước là rất lớn và không thống nhất. Mặc dù vậy nhưng đơn vị phụ trách công tác cán bộ luôn cần đến khả năng có thể

truy cập thông tin liên quan đến tất cả cán bộ từ tất cả các cơ sở dữ liệu này. Hiện tại để truy cập một thông tin khi cần thiết, cán bộ tổ chức phải liên hệ tới từng đầu mối quản lý từng cơ sở dữ liệu riêng biệt. Việc tổng hợp dữ liệu từ các cơ sở dữ liệu lại là một vấn đề cấp thiết với công tác quản lý nói chung và công tác cán bộ nói riêng.

Đặc biệt trong bối cảnh hiện tại, sau khi Chính phủ Việt Nam ra quyết định về sắp xếp và tinh gọn bộ máy, nguồn cơ sở dữ liệu về nhân sự tại Viện vì thế mà tăng lên do thay đổi cơ quan chủ quản. Nguồn dữ liệu tăng lên khiến cho các công tác nhân sự cùng với việc tổng hợp và báo cáo số liệu trở nên phức tạp hơn và thiếu chính xác.

2.1.3. Tóm lược về tích hợp dữ liệu

Tổng hợp từ những nghiên cứu tiền nhiệm, chúng ta có thể hiểu khái quát tích hợp dữ liệu là các kỹ thuật, công cụ, phương thức hay một quá trình tổng hợp, kết hợp các dữ liệu không đồng nhất từ các nguồn dữ liệu khác nhau thành một tập hợp dữ liệu thống nhất như đã được định trước. Đặc điểm của một hệ thống tích hợp dữ liệu có tính linh hoạt về nguồn dữ liệu. Nguồn dữ liệu có thể sử dụng bao gồm cả các cơ sở dữ liệu trong các hệ thống khác nhau, cũng có thể là thông tin con người từ các tổ chức khác nhau, cũng có thể là hoá đơn bán hàng từ các cửa hàng khác nhau. Các nguồn dữ liệu có thể sử dụng các ngôn ngữ khác nhau, các định dạng khác nhau. Tích hợp dữ liệu giúp người dùng có có một cái nhìn nhất quán và tổng thể đầy đủ dữ liệu của họ để từ đó có thể khai thác, hiểu biết các thông tin chuyên sâu hơn.

2.2. Các kỹ thuật tích hợp dữ liệu

2.2.1. Kỹ thuật Hợp nhất dữ liệu ETL/ELT (Extract, Transform, Load/Etract, Load, Tranform – Trích xuất, Chuyển đổi, Tải/Trích xuất, Tải, Chuyển đổi)

Kỹ thuật ETL – Tên tiếng anh đầy đủ là Extract, Transform, Load – hay còn được gọi là quy trình Trích xuất, Chuyển đổi, Tải là một kỹ thuật quan trọng trong bài toán quản lý dữ liệu và tích hợp dữ liệu. ETL là cụm từ chỉ kỹ thuật, quy trình xử lý tính toán dữ liệu gồm 3 giai đoạn chính như tên gọi của nó theo thứ tự là Trích xuất, Chuyển đổi và Tải. Ở giai đoạn đầu dữ liệu được trích xuất từ nguồn, sau đó được chuyển đổi và tải lên một kho chứa dữ liệu. Đặc điểm chính của kỹ thuật này là dữ liệu có thể được thu thập từ nhiều nguồn và sau quá trình chuyển đổi cũng có thể xuất ra một hay nhiều đích.

Ở giai đoạn đầu của quy trình, kỹ thuật ETL bắt đầu từ việc trích xuất dữ liệu từ các nguồn dữ liệu hoặc các nguồn hệ thống được cung cấp. Giai đoạn thứ hai của quy trình ETL, sau khi đã hoàn thiện việc trích xuất dữ liệu cần thiết theo yêu cầu thì dữ liệu được đưa vào giai đoạn chuyển đổi. Giai đoạn cuối cùng của quy trình ETL là quy trình Tải dữ liệu hay còn gọi là Load. Tuỳ theo nhu cầu sử dụng và độ phức tạp của mô hình dữ liệu đích mà quy trình này thực hiện các hình thức tải khác nhau.

Song hành với kỹ thuật ETL là kỹ thuật ELT hay còn được biết đến với tên gọi đầy đủ là Extract, Load, Transform (Trích xuất, Tải, Chuyển đổi). Cụ thể về các bước của ELT về cơ bản có nhiều nét tương đồng với ETL về mặt kỹ thuật thực hiện đều gồm ba giai đoạn chính như đã được đề cập. Tuy nhiên, ELT khác biệt với ETL ở thứ tự và quy trình ba bước của nó. Với ETL, người dùng tiến hành trích xuất dữ liệu thô từ các nguồn khác nhau, sau đó tiến hành chuyển đổi dữ liệu và cuối cùng là tải tài nguyên dữ liệu sau khi được chuyển đổi vào đích đến. Còn với ELT, người dùng sau khi tiến hành trích xuất dữ liệu thô từ các nguồn sẽ tiến hành tải luôn các dữ liệu đó ở trạng thái nguyên bản sau khi được trích xuất vào hệ quản trị dữ liệu đích sau đó mới tiến hành chuyển đổi dữ liệu nếu cần ngay tại hệ thống đích.

Một số công cụ đáng chú ý cung cấp giải pháp cho quy trình ETL trong thời đại này là: Informatica PowerCenter – Công cụ tích hợp dữ liệu thường được dùng kèm với trung tâm tích hợp dữ liệu của công ty Informatica; Apache Airflow – Nền tảng mã nguồn mở cho các quy trình ETL do AirBnB phát triển; Pentaho Data Integration (Cũng được biết đến với tên gọi Pentaho Kettle) – Công cụ mã nguồn mở do tập đoàn Pentaho phát triển; Talend Open Studio – Công cụ ETL với mã nguồn mở miễn phí do công ty Talend phát triển. Thư viện Pandas Python – Thư viện mã nguồn mở trong Python.

2.2.2. Kỹ thuật Data Virtualization

Ảo hoá dữ liệu hay còn được gọi là Data Virtualization là một trong những phương pháp tích hợp dữ liệu khá mới so với ETL hay ELT. Trong khoa học máy tính, ảo hoá dữ liệu là công nghệ cung cấp một lớp lưu trữ ảo từ kết quả truy vấn các dữ liệu phân tán với các nguồn dữ liệu khác nhau. Từ đó, các công cụ truy vấn và hiển thị sẽ dựa trên các lớp lưu trữ ảo này mà cung cấp các chức năng truy cập, khai thác và thao tác với dữ liệu trên thời gian thật tại nơi lưu trữ.

Điểm mạnh vượt trội của công nghệ Ảo hoá dữ liệu so với kỹ thuật ETL phổ thông được đề cập trước đó tại luận văn này là nó phép người dùng và các ứng dụng quản trị dữ liệu truy cập và thao tác dữ liệu theo thời gian thực, tuỳ vào tình hình thực tế và nhu cầu hiện tại mà có thể truy vấn dữ liệu hay dạng dữ liệu mong muốn một cách nhanh chóng mà không cần tải dữ liệu vào đích đến là cơ sở dữ liệu hay kho dữ liệu truyền thống. Lớp logic ảo hoá của công nghệ này đóng vai trò như một phiên dịch viên thời gian thực, là cầu nối giữa người dùng dữ liệu và các cơ sở dữ liệu nguồn.

Một kỹ thuật tích hợp dữ liệu quan trọng trong môi trường tích hợp dữ liệu cùng với các kỹ thuật trên mà luận văn này mong muốn trình bày là Data Federation. Nó là kỹ thuật tích hợp dữ liệu cho phép người dùng truy cập và truy vấn dữ liệu từ nhiều nguồn dữ liệu khác nhau cùng lúc mà không cần phải trích xuất, sao chép về một nơi duy nhất. Cũng tương tự như kỹ thuật Data Virtualization được đề cập ở trên thì Data Federation cũng tạo ra các cơ sở dữ liệu ảo. Tuy nhiên điểm khác biệt chính là Data Federation không tạo ra một lớp ảo hóa chuyên biệt mà chỉ tập trung vào việc truy vấn dữ liệu qua một giao diện. Đây là cấp độ tích hợp căn bản so với Data Virtualization vì nó không có các lớp logic dữ liệu hay tối ưu hóa như công nghệ Ảo hóa. Kỹ thuật này đóng vai trò nền tảng cho kỹ thuật Ảo hóa dữ liệu.

Công nghệ này còn mới nên các nghiên cứu cũng như tài liệu hỗ trợ cũng như cộng đồng chưa thực sự phát triển. Một số công cụ Ảo hóa dữ liệu đáng chú hiện nay gồm có: Microsft Hyper-V do Microsoft phát triển và duy trì, tích hợp sâu với hệ sinh thái Microsoft và Windows; Xen – Nền tảng ảo hóa mã nguồn mở do Linux Foundation phát triển, được sử dụng rộng rãi với tính linh hoạt cao; Vmware vSphere do Vmware cung cấp dịch vụ và triển khai, có tính bảo mật cao và được sử dụng tại các doanh nghiệp lớn.

2.3. Khảo sát công nghệ để giải quyết bài toán Tích hợp dữ liệu

Nhiệm vụ chính của luận văn này là giải quyết bài toán “Tích hợp dữ liệu cho hệ thống quản lý cán bộ, viên chức của Viện Năng lượng nguyên tử”. Các kỹ thuật và công cụ được nghiên cứu, so sánh để đưa ra lựa chọn đều dựa trên nhu cầu và thực tế tại đơn vị áp dụng. Kỹ thuật và công cụ được chọn phải giải quyết được vấn đề đã nêu ra về xung đột dữ liệu trong bối cảnh sáp nhập Bộ, Ngành. Kỹ thuật được được ra cần phải dễ

dàng áp dụng, có khả năng tùy chỉnh dễ dàng sử dụng, dễ dàng tùy chỉnh theo nhu cầu và mục đích người dùng và tối ưu hóa về chi phí đầu tư.

2.3.1. Kỹ thuật Tích hợp lựa chọn

Dựa trên các yếu tố đề ra và từ nhu cầu thực tế, nghiên cứu nhận định áp dụng kỹ thuật tích hợp dữ liệu theo quy trình ETL là hợp lý nhất với bối cảnh thực trạng tại Viện Năng lượng nguyên tử Việt Nam. Toàn bộ các kỹ thuật tích hợp dữ liệu phổ biến được khảo sát ở mục trước đều có ưu điểm và nhược điểm đặc thù. Tổng quan về góc độ kỹ thuật thì các công nghệ này khác nhau chủ yếu ở giai đoạn thực hiện chuyển đổi dữ liệu và

2.3.2. Công cụ Tích hợp lựa chọn

Pandas là công cụ thư viện hoàn toàn miễn phí cho ngôn ngữ python được nhà phát triển Wes McKinney xây dựng từ năm 2008. Thư viện này sử dụng một cấu trúc dữ liệu riêng là Dataframe để phục vụ các công tác chuyển đổi dữ liệu. Chính vì cấu trúc này nó đem lại cho người dùng nhiều chức năng xử lý tùy theo nhu cầu và bối cảnh thực tế rất linh hoạt mà ít công cụ đem lại được.

Pandas được phân phối hoàn toàn miễn phí, linh hoạt, dễ dàng sử dụng và có sự hỗ trợ cộng đồng lớn, không bị phụ thuộc vào công ty hay tập đoàn nào.

CHƯƠNG 3: XÂY DỰNG CÔNG CỤ PHẦN MỀM VÀ THỬ NGHIỆM ÁP DỤNG, ĐÁNH GIÁ

Căn cứ vào hiện trạng và nhu cầu cho việc giải bài toán Tích hợp dữ liệu cán bộ, viên chức và người lao động tại Viện Năng lượng nguyên tử Việt Nam, học viên tiến hành nghiên cứu và xây dựng mô hình tích hợp dữ liệu theo quy trình ETL tại chính trụ sở Viện. Mô hình này được xây dựng và phát triển để phục vụ cho công tác quản lý cán bộ, viên chức và người lao động; giải quyết được các vấn đề và nhu cầu phát sinh do ảnh hưởng của việc sáp nhập Cơ quan, Bộ, Ban, Ngành. Qua quá trình tích hợp theo quy trình ETL, hệ thống này sẽ đạt được những kết quả sau:

- Trích xuất được dữ liệu từ nhiều nguồn;

- Trích xuất được dữ liệu cụ thể từ các nguồn trên;

- Trích xuất dữ liệu được từ nhiều hệ thống lưu trữ dữ liệu cùng lúc;

- Làm sạch dữ liệu tại các trường thông tin;

- Loại bỏ các trường, cột dữ liệu trùng lặp;

- Chuẩn hoá yêu cầu về trường dữ liệu;

- Chuẩn hoá định dạng, lỗi chính tả viết hoa về các thông tin định danh;

- Chuẩn hoá thông tin về mã số định danh (yêu cầu mã số định danh điện tử theo quy định mới, chuẩn 12 số);

- Thay thế thông tin hòm thư điện tử cá nhân (gmail, yahoo, hotmail) và hòm thư điện tử thuộc Bộ KH&CN cũ (most.gov) thành tài khoản điện tử mới do Bộ KH&CN mới cung cấp và quản lý (mst.gov.vn);

- Tải và xuất dữ liệu sau khi được trích xuất, tổng hợp, chuyển đổi thành một nguồn dữ liệu duy nhất;

- Tuỳ chỉnh môi trường, định dạng, hệ thống đích cho dữ liệu đầu ra, phục vụ công tác đồng bộ với kho dữ liệu của hệ thống phần mềm quản lý cán bộ, công chức, viên chức và người lao động của Bộ KH&CN.

3.1. Xây dựng và cấu hình công cụ phần mềm

Qua khảo sát các kỹ thuật tích hợp dữ liệu hiện hành tại chương 2, để phục vụ cho công tác thực hiện những nhiệm vụ được đề ra, học viên tiến hành nghiên cứu và xây dựng quy trình ETL với sự hỗ trợ của thư viện Pandas trong môi trường Python. Công cụ sử dụng để thực hiện là Pycharm 251.25410.159; Python phiên bản 3.13 và thư viện Pandas qua cú pháp cài đặt Pip [pip install --pre --extra-index https://pypi.anaconda.org/scientific-python-nightly-wheels/simple pandas].

Để có thể xử lý được bài toán đề ra, nghiên cứu cần thiết phải làm rõ hơn về nguyên nhân phát sinh bài toán 'tích hợp dữ liệu cho hệ thống quản lý cán bộ, viên chức của Viện Năng lượng nguyên tử". Học viên đã tiến hành khảo sát cụ thể về quy trình quản lý và khai thác hồ sơ cán bộ của Vụ Tổ chức cán bộ, Bộ KH&CN sau sáp nhập.

Quy trình đồng bộ sau sáp nhập giữa các hệ thống dữ liệu tại Bộ KH&CN trước và sau sáp nhập được áp dụng và triển khai đối với các đơn vị cơ sở trực thuộc nói chung và tại Viện Năng lượng nguyên tử nói riêng được thực hiện theo quy trình: Dữ liệu cán bộ được trích xuất từ các nguồn thông tin tại nhà kho dữ liệu là hệ thống quản lý quản lý cán bộ của Bộ KH&CN (2024) dưới dạng dữ liệu thô và sau đó được tải lại vào hệ thống phần mềm mới của Bộ KH&CN (2025) mà không được chuyển đổi hay làm sạch. Quy trình này được thực hiện một cách nhanh chóng trong bối cảnh sắp xếp và tinh gọn bộ máy hành chính nhà nước. Hình 3.3 là ví dụ về một số dữ liệu về cán bộ, viên chức và người lao động tại Viện Năng lượng nguyên tử Việt Nam được học viên trích xuất ra từ hệ thống mới tại thời gian thực.

A	B	C	D
9	Nguyễn Thị Thu Hà	034179010151	hantt_itrre@mst.gov.vn
10	Nguyễn Thị Thu Hà	001178038310	hantt_inst@mst.gov.vn
11	Nguyễn Thị Thu Hà	001176047943	nguyenthithuha55@mst.gov.vn
12	Nguyễn Tiến Phong	024078000210	phongnt_nde@mst.gov.vn
13	NGUYỄN TIẾN PHONG	001087000895	qlkh.nde@gmail.com
14	Nguyễn Trung Kiên	035083011272	kiennt_nde@mst.gov.vn
15	Nguyễn Trung Kiên	001089029911	trungkien211013@gmail.com
16	Phạm Bảo Ngọc	068194000848	ngocpb_nri@mst.gov.vn
17	Phạm Bảo Ngọc	001189019141	ngocpb_inst@mst.gov.vn
18	PHẠM THỊ THU TRANG	036189001823	trangptt@mst.gov.vn
19	PHẠM THỊ THU TRANG	NULL	cherrypham.hanoi@gmail.com
20	Phạm Văn Dũng	001067017076	dungpv_nri@mst.gov.vn
21	PHẠM VĂN DŨNG	001067067076	phamdungatbx@gmail.com
22	VŨ ĐỨC MẠNH	034096017423	vuducmanh05021996@gmail.com
23	VŨ ĐỨC MẠNH	034096017424	manhvd_inst@mst.gov.vn
24	VŨ ĐỨC MẠNH	34096017424	manhvd_inst@mst.gov.vn

Hình 3.1 Dữ liệu về cán bộ hiện tại được trích xuất ra từ hệ thống mới 2025

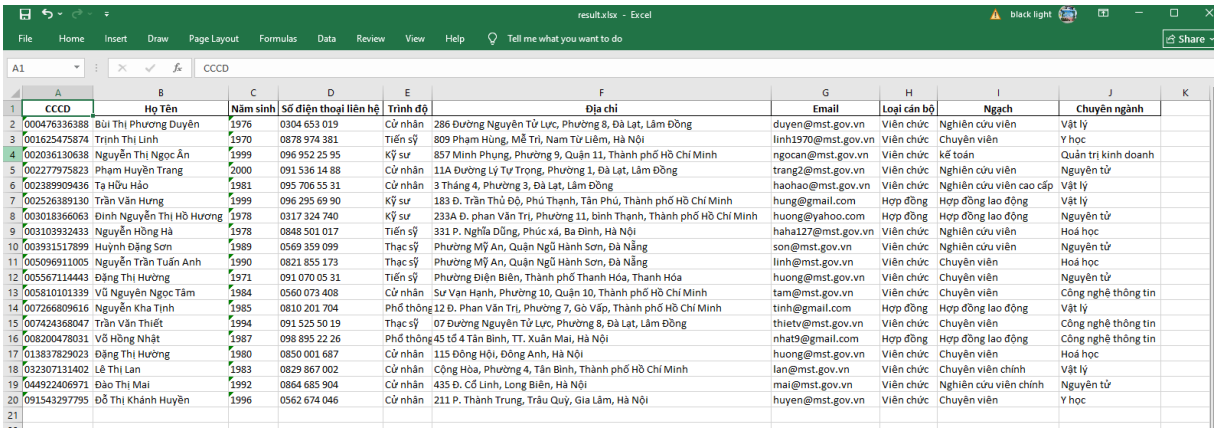
Bảng thống kê được minh họa tại Hình 3.3 góp phần là một minh chứng cụ thể cho sự cấp thiết của đề tài "Tích hợp dữ liệu cho hệ thống quản lý cán bộ, viên chức của Viện Năng lượng nguyên tử". Có thể thấy rõ nhiều cán bộ bị trùng lặp dữ liệu dẫn đến việc khi tạo ra các báo cáo thống kê nhân sự, số liệu và thông tin bị sai lệch, không có giá trị sử dụng.

3.2. Kết quả thử nghiệm

Để phục vụ công tác nghiên cứu xây dựng thử nghiệm vận hành hệ thống tích hợp dữ liệu được đề xuất, học viên tiến hành khởi tạo 3 tệp chứa dữ liệu về cán bộ, người lao động dựa trên các yếu tố và hiện trạng tại các bản ghi hồ sơ nhân sự có sẵn tại hệ thống quản lý cán bộ, viên chức, công chức và người lao động của Bộ KH&CN (sau sáp nhập) nhằm mô phỏng lại hiện trạng thực tế về bài toán tích hợp dữ liệu tại Viện Năng lượng nguyên tử.

Tệp dữ liệu thứ 1 được khởi tạo là tệp dữ liệu về thông tin cán bộ tại Viện được trích xuất ra từ Hệ thống quản lý CCVCNLĐ của Bộ KH&CN (trước sáp nhập). Tệp dữ liệu thứ 2 được khởi tạo là tệp dữ liệu được khởi tạo bởi Bộ KH&CN (mới) nhằm phục công tác đồng bộ. Tệp dữ liệu thứ 3 là tệp dữ liệu cán bộ được yêu cầu bổ sung thông tin, gồm các trường thông tin bổ sung mà tại hệ thống quản lý CCVCCNLĐ của Bộ KH&CN (cũ) nhưng cần thiết phải cập nhật tại hệ thống phần mềm quản lý CCVCNLĐ tại Bộ&CN (mới) nhằm phục vụ các công tác cán bộ.

Với 3 File dữ liệu cán bộ như giả định được đưa vào quy trình tích hợp dữ liệu, hệ thống được đề xuất cho người dùng kết quả là dữ liệu được ghi vào định dạng tuỳ chọn cụ thể như hình ảnh kết quả tại Hình 3.13.


result.xlsx - Excel

CCCD	Họ Tên	Năm sinh	Số điện thoại liên hệ	Trình độ	Địa chỉ	Email	Loại cán bộ	Ngạch	Chuyên ngành
000476336388	Bùi Thị Phương Duyên	1976	0304 653 019	Cử nhân	286 Đường Nguyễn Tử Lực, Phường 8, Đà Lạt, Lâm Đồng	duyen@mst.gov.vn	Viên chức	Nghiên cứu viên	Vật lý
001625475874	Trịnh Thị Linh	1970	0878 974 381	Tiến sỹ	809 Phạm Hùng, Mễ Trì, Nam Từ Liêm, Hà Nội	linh1970@mst.gov.vn	Viên chức	Chuyên viên	Y học
002036130638	Nguyễn Thị Ngọc Ân	1999	096 952 25 95	Kỹ sư	857 Minh Phụng, Phường 9, Quận 11, Thành phố Hồ Chí Minh	ngocan@mst.gov.vn	Viên chức	kế toán	Quản trị kinh doanh
002277975823	Phạm Huyền Trang	2000	091 536 14 88	Cử nhân	11A Đường Lý Tự Trọng, Phường 1, Đà Lạt, Lâm Đồng	trang2@mst.gov.vn	Viên chức	Nghiên cứu viên	Nguyên tử
002389909436	Tạ Hữu Hảo	1981	095 706 55 31	Cử nhân	3 Tháng 4, Phường 3, Đà Lạt, Lâm Đồng	haohao@mst.gov.vn	Viên chức	Nghiên cứu viên cao cấp	Vật lý
002526389130	Trần Văn Hưng	1999	096 295 69 90	Kỹ sư	183 Đ. Trần Thủ Độ, Phú Thạnh, Tân Phú, Thành phố Hồ Chí Minh	hung@gmail.com	Hợp đồng	Hợp đồng lao động	Vật lý
003018366063	Đinh Nguyễn Thị Hồ Hương	1978	0317 324 740	Kỹ sư	233A Đ. phan Văn Trị, Phường 11, bình Thạnh, Thành phố Hồ Chí Minh	huong@yahoo.com	Hợp đồng	Hợp đồng lao động	Nguyên tử
003103932433	Nguyễn Hồng Hà	1978	0848 501 017	Tiến sỹ	331 P. Nghĩa Dũng, Phúc xá, Ba Đình, Hà Nội	haha127@mst.gov.vn	Viên chức	Nghiên cứu viên	Hoá học
003931517899	Huỳnh Đặng Sơn	1989	0569 359 099	Thạc sỹ	Phường Mỹ An, Quận Ngũ Hành Sơn, Đà Nẵng	son@mst.gov.vn	Viên chức	Nghiên cứu viên	Nguyên tử
005096911005	Nguyễn Trần Tuấn Anh	1990	0821 855 173	Thạc sỹ	Phường Mỹ An, Quận Ngũ Hành Sơn, Đà Nẵng	linh@mst.gov.vn	Viên chức	Chuyên viên	Hoá học
005567114443	Đặng Thị Hường	1971	091 070 05 31	Tiến sỹ	Phường Điện Biên, Thành phố Thanh Hóa, Thanh Hóa	huong@mst.gov.vn	Viên chức	Chuyên viên	Nguyên tử
005810101339	Vũ Nguyễn Ngọc Tâm	1984	0560 073 408	Cử nhân	Sư Vạn Hạnh, Phường 10, Quận 10, Thành phố Hồ Chí Minh	tam@mst.gov.vn	Viên chức	Chuyên viên	Công nghệ thông tin
007266809616	Nguyễn Kha Tịnh	1985	0810 201 704	Phổ thông	12 Đ. Phan Văn Trị, Phường 7, Gò Vấp, Thành phố Hồ Chí Minh	tinh@gmail.com	Hợp đồng	Hợp đồng lao động	Vật lý
007424368047	Trần Văn Thiết	1994	091 525 50 19	Thạc sỹ	07 Đường Nguyễn Tử Lực, Phường 8, Đà Lạt, Lâm Đồng	thietv@mst.gov.vn	Viên chức	Chuyên viên	Công nghệ thông tin
008200478031	Võ Hồng Nhật	1987	098 895 22 26	Phổ thông	45 tổ 4 Tân Bình, TT. Xuân Mai, Hà Nội	nhat9@gmail.com	Hợp đồng	Hợp đồng lao động	Công nghệ thông tin
013837829023	Đặng Thị Hường	1980	0850 001 687	Cử nhân	115 Đông Hội, Đông Anh, Hà Nội	huong@mst.gov.vn	Viên chức	Chuyên viên	Hoá học
032307131402	Lê Thị Lan	1983	0829 867 002	Cử nhân	Cộng Hòa, Phường 4, Tân Bình, Thành phố Hồ Chí Minh	lan@mst.gov.vn	Viên chức	Chuyên viên chính	Vật lý
044922406971	Đào Thị Mai	1992	0864 685 904	Cử nhân	435 Đ. Cổ Linh, Long Biên, Hà Nội	mai@mst.gov.vn	Viên chức	Nghiên cứu viên chính	Nguyên tử
091543297795	Đỗ Thị Khánh Huyền	1996	0562 674 046	Cử nhân	211 P. Thành Trung, Trâu Quỳ, Gia Lâm, Hà Nội	huyen@mst.gov.vn	Viên chức	Chuyên viên	Y học

Hình 3.2 Dữ liệu cán bộ được tích hợp từ 3 file dữ liệu có sẵn

Sau tích hợp, dữ liệu đã được tổng hợp ở cả 3 File và toàn bộ các dữ liệu trùng lặp đã bị loại bỏ. Nhân sự được tích hợp ở kết quả cuối bao gồm cả 15 nhân sự là viên chức thuộc biên chế nhà nước, có tài khoản @mst.gov.vn ban đầu tại File 1 và 2 cộng với 4 nhân sự thuộc diện ký hợp đồng lao động với đơn vị tại File 3 thông tin bổ sung. Tất cả các số CCCD được chuẩn hoá thành dạng 12 số, không bị mất số 0 ở đầu dãy. Họ và Tên cán bộ tại đơn vị sau tích hợp cũng được chuẩn hoá về cùng một định dạng viết hoa các chữ cái đầu tiên. Cột Email đã được chuẩn hoá theo tình trạng thực tế, ưu tiên địa chỉ thư điện thử @mst.gov.vn với các cán bộ có tài khoản là viên chức thuộc biên chế nhà nước. Với những cán bộ còn lại thuộc diện lao động hợp đồng, sử dụng dữ liệu email cá nhân. Các trường thông tin được yêu cầu bổ sung tại File 3 cũng được tổng hợp vào dữ liệu nhân sự cuối.

Bài luận văn này không sử dụng hoàn toàn các dữ liệu nhân sự sẵn có tại Viện Năng lượng nguyên tử Việt Nam. Dữ liệu và thông tin chi tiết về cán bộ là thông tin nhân sự thuộc quyền sở hữu của nhà nước. Để đảm bảo tính bảo mật về thông tin cán bộ nên luận văn này chỉ sử dụng một phần để mô phỏng giả định lại thực trạng đã được chứng minh trước đó. Mọi trùng lặp về họ tên, địa chỉ, mã số định danh và các thông tin khác chỉ là sự trùng hợp.

3.3. Định hướng phát triển

Về tính khả thi, các vấn đề được đưa ra để giải quyết chỉ là những vấn đề hiện hữu trong quá trình quản trị dữ liệu cán bộ tại Viện Năng lượng nguyên tử. Hệ thống cần được vận hành một thời gian với vai trò là cầu nối giữa các hệ thống dữ liệu của Viện và của Bộ để có thể hiệu chỉnh cụ thể hơn về các vấn đề chưa hiện hữu.

Về mặt người dùng, hệ thống cần có một giao diện cụ thể để người dùng có thể tương tác và hiệu chỉnh một cách dễ dàng hơn. Hiện tại các hiệu chỉnh vẫn được triển khai một cách thủ công bằng cách tác động trực tiếp vào code nguồn, điều này dễ gây ra sai sót khi thao tác xử lý khối công việc lớn hơn.

Để có thể tiếp tục hỗ trợ và ứng dụng cho việc giải quyết bài toán tích hợp dữ liệu cho hệ thống quản lý cán bộ, viên chức và người lao động tại Viện NLNTVN và hướng tới áp dụng cho hệ thống của Bộ KH&CN, học viên cần phải nghiên cứu chuyên sâu thêm một cách cụ thể về hệ thống đang được triển khai và sử dụng.

KẾT LUẬN

Trong bối cảnh đất nước đẩy mạnh công cuộc chuyển đổi số quốc gia trên cả 3 trụ cột chính là: Chính phủ số, kinh tế số và xã hội số, việc triển khai và hoàn thiện số hóa các hồ sơ cán bộ là vô cùng cần thiết. Với chủ đề của năm 2025 là "Chuyển đổi số toàn dân, toàn diện, toàn trình để tăng tốc, bứt phá phát triển kinh tế số", công tác chuyển đổi số và khai thác thông tin, dữ liệu nhân sự là vô cùng cấp thiết với toàn bộ các cơ quan, công ty, tập thể quản lý nhân sự. Công tác này đóng vai trò là nền tảng để thực hiện tất cả các công tác quản trị con người trên nền tảng số hiện nay.

Khởi điểm về bài toán của nghiên cứu này dựa trên nhu cầu tích hợp dữ liệu về cán bộ, công chức, viên chức và người lao động trên toàn bộ hệ thống quản lý cán bộ quốc gia. Nhiệm vụ này phát sinh từ công tác sáp nhập Tỉnh, Thành cũng như các cơ quan Bộ, Ngành và nó có tính chất nền tảng, dễ dàng hiệu chỉnh cũng như nhân rộng, có khả năng áp dụng rộng rãi tại thời điểm này. Nghiên cứu này đóng vai trò quan trọng trong công tác đồng bộ, chuyển đổi, chuẩn hóa và làm sạch dữ liệu từ các nguồn hệ thống dữ liệu khác nhau.

Qua quá trình thử nghiệm với những nguồn dữ liệu giả định tại Việt NLNT, hệ thống được đề xuất đã giải quyết được phần lớn các vấn đề về nhu cầu tích hợp dữ liệu phát sinh trong bối cảnh đất nước đổi mới, tinh gọn bộ máy quản lý nhà nước. Kết quả quả này không chỉ tạo nền móng mà nó còn có khả năng áp dụng trực tiếp cho các nguồn dữ liệu cán bộ hiện hữu tại Viện NLNT cũng như các đơn vị trực thuộc khác sau khi sáp nhập Bộ ngành.

Đây là một nghiên cứu mang tính chất ứng dụng với nhu cầu phát sinh từ định hướng phát triển quốc gia, do còn nhiều tiềm năng chưa được khai phá, các trường hợp cụ thể chưa được tìm ra dẫn đến các ý tưởng chuyển đổi dữ liệu mới chỉ ở mức tương đối. Do giới hạn về thời gian, kinh phí và điều kiện công tác nên khuyết điểm và hạn chế là không thể tránh khỏi. Học viên mong muốn nhận được sự góp ý, nhận xét và phản hồi quý báu của các quý thầy cô, nhà khoa học, đồng nghiệp để luận văn được hoàn thiện hơn.