

XÂY DỰNG TRUNG TÂM TÂM DỮ LIỆU LỚN CHO HỆ THỐNG ĐẠI HỌC THÔNG MINH

TS. Lê Việt Thuỷ

Trường Đại học Kinh tế Quốc dân, thuylv@neu.edu.vn

ThS. Nguyễn Văn Hoàng

Trường Đại học Kinh tế Quốc dân, nguyenhoang@neu.edu.vn

TÓM TẮT:

Hệ thống đại học thông minh là một hệ thống thông tin giáo dục hiện đại tiên tiến được hỗ trợ bởi các công nghệ chuyển đổi số như điện toán đám mây, Internet vạn vật, dữ liệu lớn. Ứng dụng dữ liệu lớn là trọng tâm và thách thức của hệ thống đại học thông minh. Dữ liệu lớn là một khía cạnh thiết yếu của quá trình chuyển đổi số mà gần đây đã nhận được sự quan tâm lớn từ cả giới học thuật và các nhà thực hành. Cho đến nay, nhiều nghiên cứu đã được thực hiện nhằm tìm hiểu toàn diện việc ứng dụng dữ liệu lớn trong các lĩnh vực khác nhau cho các mục đích khác nhau. Tuy nhiên, đánh giá một cách toàn diện vẫn còn thiếu dữ liệu lớn trong giáo dục. Hầu hết các nghiên cứu dữ liệu lớn về giáo dục đều tập trung vào hành vi và hoạt động của người học. Bài viết này thực hiện nghiên cứu ứng dụng dữ liệu lớn trong hệ thống đại học thông minh, thúc đẩy chuyển đổi công nghệ thông tin của các trường đại học thành công nghệ thông tin dịch vụ thông minh nhằm nêu bật thực trạng ứng dụng và mở ra hướng đi mới để sử dụng thành công dữ liệu lớn trong giáo dục.

Từ khóa: Đại học thông minh, hệ thống thông tin, dữ liệu lớn, chuyển đổi số.

1. GIỚI THIỆU

Thế giới đang thay đổi nhanh chóng do sự xuất hiện của các công nghệ chuyển đổi số [1]. Hiện tại, một số lượng lớn các thiết bị công nghệ được sử dụng bởi các hành vi cá nhân hàng ngày con người [2]. Trong mỗi khoảnh khắc, một lượng lớn dữ liệu được tạo ra thông qua các thiết bị [3]. Để xử lý lượng dữ liệu khổng lồ này, các công nghệ và ứng dụng hiện tại đang được phát triển mạnh mẽ. Các công nghệ và ứng dụng này rất hữu ích cho việc phân tích và lưu trữ dữ liệu [4]. Nhu cầu dữ liệu lớn cũng đang gia tăng đáng kể trong các lĩnh vực nỗ lực khác nhau như bảo hiểm và xây dựng, y tế, viễn thông và thương mại điện tử. Theo Dresner Advisory Services [5] công nghệ (14%), dịch vụ tài chính (10%), tư vấn (9%), chăm sóc sức khỏe (9%), giáo dục (8%) và viễn thông (7%) là những lĩnh vực hoạt động tích cực nhất trong việc tạo ra nhiều giá trị cho dữ liệu lớn.

Như vậy, có thể thấy ngành giáo dục cũng không phải là ngoại lệ của xu thế này. Trong lĩnh vực giáo dục, một khối lượng lớn dữ liệu được tạo ra thông qua các khóa học trực tuyến, các hoạt động giảng dạy, học tập và nghiên cứu khoa học. Với sự ra đời của dữ liệu lớn, giờ đây, giáo viên có thể truy cập kết quả học tập, cách học của sinh viên và đưa ra các phân tích và phản hồi tức thì giúp nâng cao chất lượng dịch vụ đào tạo trong nhà trường, thay vì phải đợi đến các dịp đánh giá cuối học kỳ hoặc cuối năm như trước đây. Dữ liệu học thuật có thể giúp giáo viên phân tích phương pháp sư phạm giảng dạy, giúp họ có những thay đổi phù hợp

theo nhu cầu và yêu cầu của sinh viên. Nhiều trang web giáo dục trực tuyến đã được thiết kế và nhiều khóa học dựa trên sở thích của từng sinh viên đã được giới thiệu. Sự cải thiện trong lĩnh vực giáo dục phụ thuộc vào sự tiếp thu và ứng dụng công nghệ. Quản trị dữ liệu lớn có thể đóng một vai trò to lớn trong việc quản lý các vấn đề giáo dục [6]. Do đó, các chuyên gia và những người làm giáo dục cần hiểu được hiệu quả của dữ liệu lớn trong giáo dục để giảm thiểu các vấn đề giáo dục.

Ở Việt Nam, tác giả Vũ Hưng Hải [7] đã phân tích vai trò của dữ liệu lớn trong giáo dục đại học và đề xuất một số khuyến nghị nhằm ứng dụng và khai thác dữ liệu lớn trong các tổ chức giáo dục đại học ở Việt Nam. Tác giả Mai Anh Thơ [8] đã chỉ ra phân tích dữ liệu lớn trong giáo dục là một xu thế mà hầu hết các trường đại học chất lượng trên thế giới đã và đang tiếp tục triển khai vì các lợi ích lớn của việc này mang lại cho các trường đại học. Mặc dù nhiều lĩnh vực đã được đề cập trong các nghiên cứu tổng quan trước đây, nhưng đến nay, chưa có những nghiên cứu xây dựng hệ thống thông tin lưu trữ và chia sẻ dữ liệu lớn trong giáo dục tại Việt Nam.

2. TỔNG QUAN VỀ DỮ LIỆU LỚN

Như đã trình bày ở trên về đặc tính, dữ liệu lớn thường bao gồm tập hợp dữ liệu với kích thước vượt xa khả năng của các công cụ phần mềm thông thường để thu thập, hiển thị, quản lý và xử lý dữ liệu trong một thời gian có thể chấp nhận được. Kích thước dữ liệu lớn cũng là một biến số liên tục thay đổi. Như năm 2012 thì phạm vi một vài terabytes tới nhiều petabytes dữ liệu thì ngày nay đã lên đến hàng chục zettabytes. Dữ liệu lớn cũng yêu cầu một tập hợp các kỹ thuật và công nghệ được tích hợp theo hình thức mới để khai phá từ tập dữ liệu đa dạng, phức tạp và có quy mô lớn. [9]

Trong báo cáo nghiên cứu năm 2001, META Group (bây giờ là Gartner) [10] nhà phân tích Doug Laney định nghĩa những thách thức và cơ hội tăng dữ liệu như là 3 chiều, tăng giá trị dữ liệu (value), tốc độ vào ra của dữ liệu (velocity), và khối giới hạn của kiểu dữ liệu (variety). Gartner, và nhiều ngành công nghiệp tiếp tục sử dụng mô hình ‘3Vs’ để mô tả dữ liệu lớn. Trong năm 2012, Gartner đã cập nhật định nghĩa như sau: “Dữ liệu lớn là khối lượng lớn, tốc độ cao hoặc loại hình thông tin rất đa dạng mà yêu cầu phương thức xử lý mới để cho phép tăng cường ra quyết định, khám phá bên trong và xử lý tối ưu”. Định nghĩa ‘3Vs’ của Gartner vẫn được sử dụng rộng rãi, và trong phù hợp với định nghĩa: “Dữ liệu lớn tiêu biểu cho tập thông tin mà đặc điểm như khối lượng lớn (Volume), tốc độ cao (Velocity) và đa dạng (Variety) để yêu cầu phương thức phân tích và công nghệ riêng biệt để biến nó thành có giá trị”. Thêm nữa, vài tổ chức đã thêm vào tính xác thực (Veracity) để mô tả về nó. 3Vs đã được mở rộng để bổ sung đặc tính của dữ liệu lớn. Gần đây, dữ liệu lớn đã trở thành vấn đề thu hút được nhiều các nhà nghiên cứu quan tâm nhằm xác định và mô tả đặc điểm của dữ liệu lớn theo những cách khác nhau [11]. Theo Yassine và cộng sự [12], dữ liệu lớn là một khối lượng lớn dữ liệu. Tuy nhiên, De Mauro, Greco và Grimaldi [13] gọi nó là một tài sản thông tin có giá trị cao được đặc trưng bởi số lượng, tốc độ và tính đa dạng cao. Hơn nữa, Shahat [14] đã mô tả dữ liệu lớn là những tập dữ liệu lớn khó xử lý, kiểm soát hoặc kiểm tra theo cách truyền thống. Tổng hợp lại, chúng ta có thể thấy dữ liệu lớn thường được đặc trưng bởi năm

chữ V là Khối lượng (Volume), Tính đa dạng (Velocity), Vận tốc (Variety), Tính xác thực (Veracity) và Giá trị (Value).

2.1. Công nghệ dữ liệu lớn

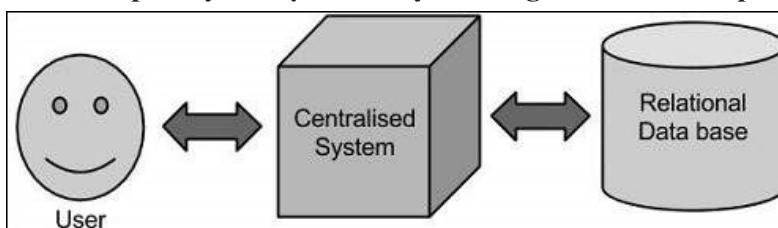
Để xử lý dữ liệu có nhiều đặc tính như trên, cần phải có hệ thống tích hợp lưu trữ dung lượng lớn và các công nghệ có thể xử lý ở định dạng có cấu trúc hoặc định dạng không có cấu trúc (hình ảnh, video, email, ...) kể cả việc hiện thị theo thời gian thực.

Công nghệ dữ liệu lớn rất quan trọng trong việc cung cấp và phân tích chính xác hơn, có thể dẫn đến việc ra quyết định cụ thể hơn, dẫn đến hiệu quả hoạt động cao hơn, giảm chi phí và giảm rủi ro cho doanh nghiệp. Để khai thác sức mạnh của dữ liệu lớn, sẽ cần một cơ sở hạ tầng có thể quản lý và xử lý khối lượng lớn dữ liệu có cấu trúc và phi cấu trúc trong thời gian thực và có thể bảo vệ quyền riêng tư và bảo mật của dữ liệu. Có nhiều công nghệ xử lý dữ liệu lớn khác nhau trên thị trường được cung cấp bởi nhiều hãng công nghệ lớn nhỏ khác nhau bao gồm Google, Amazon, IBM, Microsoft, Oracle, Hitachi Vantara, Cloudera, v.v.. Có một đặc điểm chung của các hãng cung cấp công nghệ xử lý dữ liệu lớn này là họ đều có một hệ thống thông tin theo cách tiếp cận nền tảng (platform) dựa trên sức mạnh của điện toán đám mây (cloud computing) sẽ được phân tích chi tiết ở các phần sau.

2.2. Cách tiếp cận truyền thống

Theo cách tiếp cận này, một doanh nghiệp sẽ có một máy tính để lưu trữ và xử lý dữ liệu lớn. Đối với mục đích lưu trữ, các lập trình viên sẽ nhận được sự trợ giúp của sự lựa chọn của các nhà cung cấp cơ sở dữ liệu như Oracle, IBM, v.v. Trong cách tiếp cận này, người dùng tương tác với ứng dụng, từ đó xử lý phần lưu trữ và phân tích dữ liệu.

Hình 1: Mô hình tiếp xử lý dữ liệu lớn truyền thống. Theo tutorialspoint.com [15]

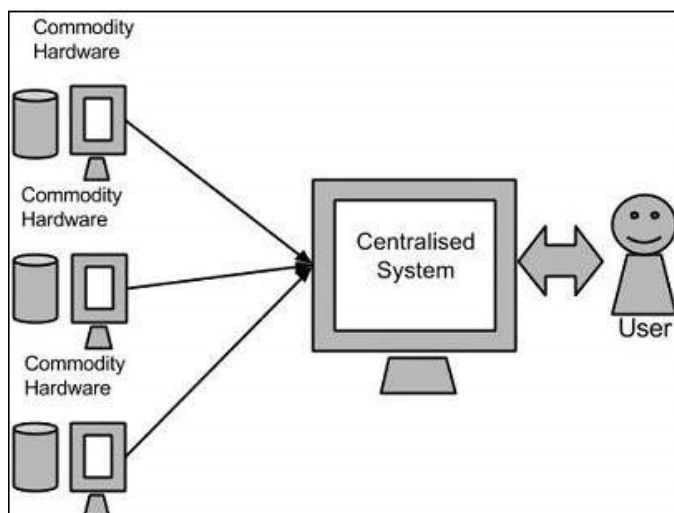


Cách tiếp cận này hoạt động tốt với những ứng dụng xử lý dữ liệu ít không lồ hơn, có thể được cung cấp bởi các máy chủ cơ sở dữ liệu tiêu chuẩn, có cấu trúc. Nhưng khi phải xử lý một lượng lớn dữ liệu mở rộng có đặc tính đa dạng như miêu tả ở phần Tổng quan, thì việc xử lý dữ liệu đó là một nhiệm vụ khó khăn rất nhiều. Chúng ta hình dung việc dữ liệu chảy về điểm xử lý sẽ tạo thành hiện tượng nút cổ chai (bottleneck) làm tắc nghẽn hệ thống thông tin. Do vậy, cần có giải pháp hiệu quả hơn.

2.3. Giải pháp của Google

Google đã giải quyết vấn đề này bằng cách sử dụng một thuật toán có tên là MapReduce. Thuật toán này chia nhiệm vụ thành các phần nhỏ và giao chúng cho nhiều máy tính, đồng thời thu thập kết quả từ chúng mà khi được tích hợp, tạo thành tập dữ liệu kết quả.

Hình 2: Mô hình xử lý dữ liệu lớn của google. Theo tutorialspoint.com

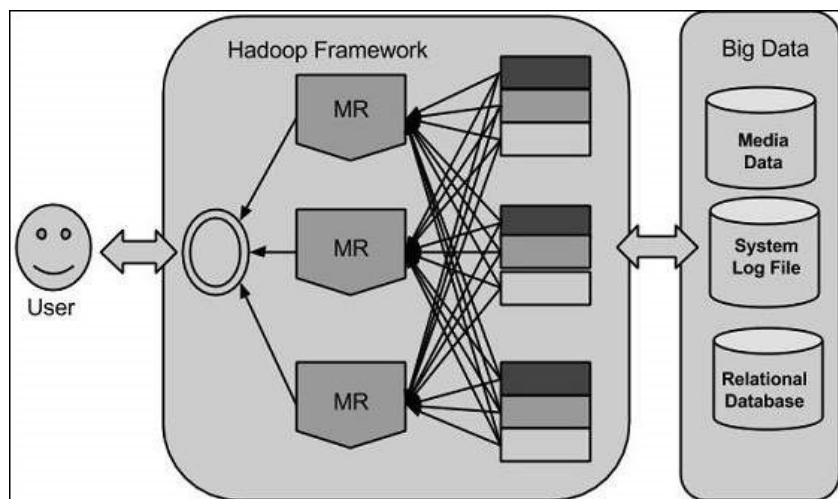


Cách tiếp cận này giúp cho việc xử lý dữ liệu được chia nhỏ ra nhằm tận dụng tối đa tài nguyên và năng lực tính toán của nhiều máy chủ khác nhau. Các máy chủ này tạo thành các cụm tính toán (cluster computing) theo dạng dịch vụ (as a service), một đặc tính cốt lõi của các nền tảng điện toán đám mây. Nhờ đó mà các tác vụ sẽ được chia sẻ thành các dịch vụ khác nhau có thể được đáp ứng song song cùng một lúc giúp giải quyết hiện tượng nút cổ chai của cách tiếp cận truyền thống.

2.4. Hadoop

Sử dụng giải pháp do Google cung cấp như miêu tả ở Hình 2, Doug Cutting và nhóm của mình đã phát triển một Dự án nguồn mở có tên là HADOOP. Hadoop chạy các ứng dụng bằng cách sử dụng thuật toán MapReduce, nơi dữ liệu được xử lý song song với những dữ liệu khác. Tóm lại, Hadoop được sử dụng để phát triển các ứng dụng có thể thực hiện phân tích thống kê hoàn chỉnh về lượng dữ liệu khổng lồ

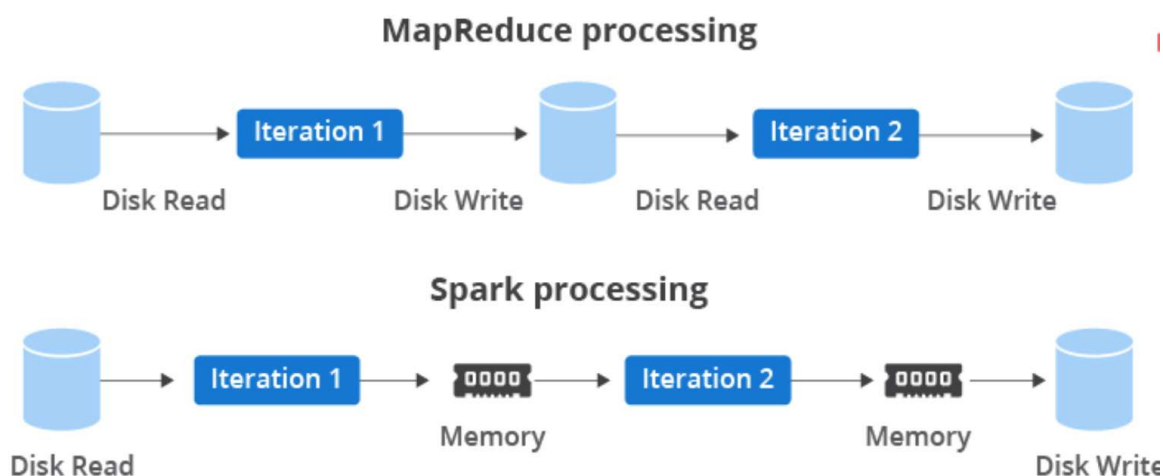
Hình 3: Mô hình xử lý dữ liệu lớn của Hadoop. Theo tutorialspoint.com



2.5. Apache Spark

Apache Spark là một nền tảng tính toán cụm (cluster computing) mã nguồn mở được phát triển sơ khởi vào năm 2009 bởi AMPLab tại đại học California. Sau này, Spark đã được trao cho Apache Software Foundation vào năm 2013 và được phát triển cho đến nay. Nó cho phép xây dựng các mô hình dự đoán nhanh chóng với việc tính toán được thực hiện trên một cụm các máy tính, có thể tính toán cùng lúc trên toàn bộ tập dữ liệu mà không cần phải trích xuất mẫu tính toán thử nghiệm. Tốc độ xử lý của Spark có được do việc tính toán được thực hiện cùng lúc trên nhiều máy khác nhau. Đồng thời việc tính toán được thực hiện tối ưu ngay ở bộ nhớ trong (in-memories) hay thực hiện hoàn toàn trên RAM, mà không phải mất nhiều thời gian truy cập vào đĩa cứng vật lý giúp được tốc độ xử lý được cải thiện rất lớn

Hình 4: So sánh cách hoạt động của MapReduce & Spark. Theo Nitin Kumar, Medium

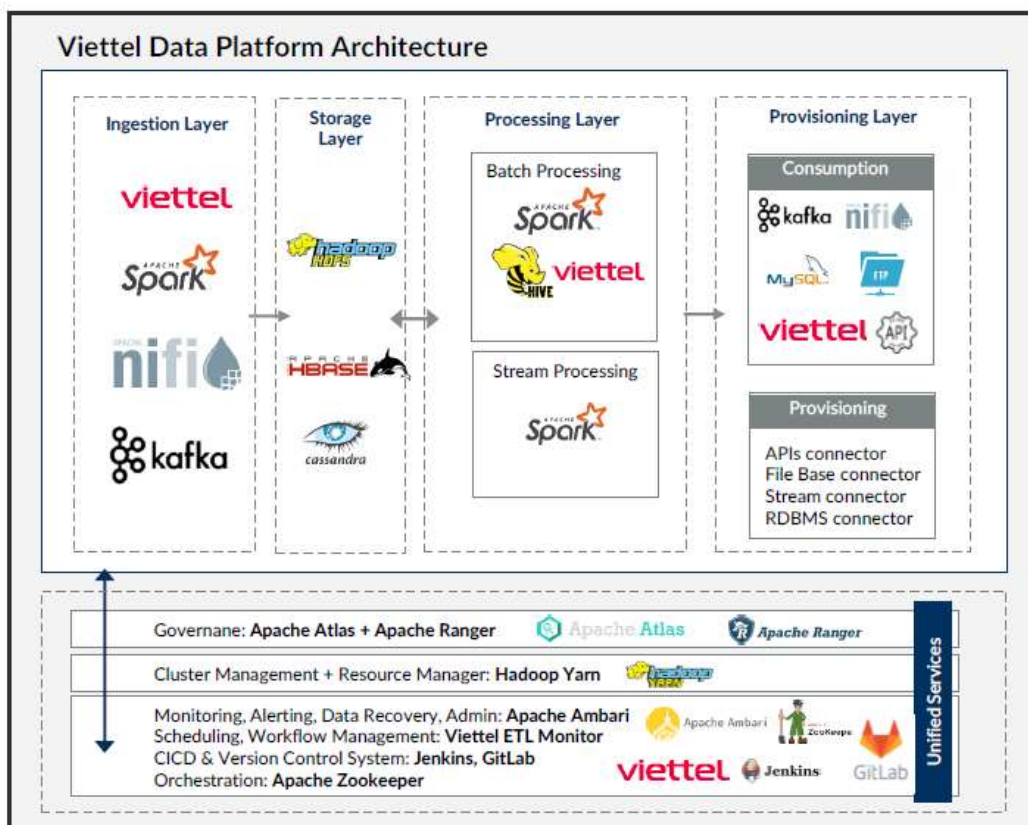


3. KHUNG XỬ LÝ DỮ LIỆU LỚN TẠI VIỆT NAM

Viettel Data Platform

Năm 2012 Viettel đã bắt đầu tiên phong tổ chức nghiên cứu, xây dựng Hệ thống xử lý và khai thác dữ liệu lớn tập trung. Tại đây các hệ thống như Viettel BI, Viettel Data Lake, Viettel Data Platform... được nghiên cứu và triển khai thành công ở cả thị trường Việt Nam và các thị trường nước ngoài mà Viettel đang đầu tư kinh doanh. Các sản phẩm này sử dụng Apache Spark làm công nghệ xử lý dữ liệu chính. Đây là công nghệ xử lý dữ liệu lớn mã nguồn mở hàng đầu trong các hệ thống Dữ liệu lớn thế giới nhờ ưu điểm về tốc độ và sự ổn định vượt trội so với những công nghệ khác. Các ứng dụng sử dụng Spark có hiệu năng gấp từ 10-100 lần việc xử lý dữ liệu sử dụng công nghệ khác. [16] Tháng 12 năm 2020, Bộ Thông tin và Truyền thông (TT&TT) phối hợp cùng Viettel ra mắt nền tảng Khai phá dữ liệu – Viettel Data Mining Platform, nằm trong khuôn khổ sự kiện “Make in Vietnam” nhằm thực hiện Chương trình chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030 được Thủ tướng Chính phủ phê duyệt[17].

Hình 8: Viettel Data Platform



4. ĐỀ XUẤT MÔ HÌNH NEU IO PORTAL – TRUNG TÂM XỬ LÝ DỮ LIỆU LỚN TẠI ĐẠI HỌC KINH TẾ QUỐC DÂN

4.1. Hệ thống đại học thông minh tại Đại học Kinh tế Quốc dân (NEU)

Trong những năm gần đây, đại học thông minh không chỉ là một chủ đề nghiên cứu được quan tâm mà đã trở thành một xu thế phát triển giáo dục đại học tại các nước có trình độ kinh tế cao, tuy nhiên, đây vẫn là một nội dung mới lạ tại Việt Nam. Nghiên cứu và triển khai đại học thông minh nhận được sự quan tâm của chính phủ cũng như các tổ chức trong và ngoài nước.

Tại Việt Nam, theo Nghị quyết số 29-NQ/TW ngày 04/01/2013 của Ban chấp hành Trung ương [18] về đổi mới căn bản, toàn diện giáo dục và đào tạo đã nêu rõ quan điểm chỉ đạo “Giáo dục và Đào tạo là quốc sách hàng đầu, là sự nghiệp của Đảng, Nhà nước và của toàn dân, đầu tư cho giáo dục là đầu tư phát triển, được ưu tiên đi trước trong các chương trình, kế hoạch phát triển kinh tế - xã hội”, do vậy, ngành Giáo dục ở Việt Nam đang rất quan tâm, đẩy mạnh mô hình trường học thông minh (THTM) nhằm đáp ứng yêu cầu xây dựng xã hội thông minh. Theo Nguyễn Hữu Đức, Hà Quang Thụy, Phạm Bảo Sơn, Trần Trọng Hiếu và Tôn Quang Cường [19] đưa ra trong nghiên cứu của mình thì đại học thông minh là một cơ sở giáo dục định hướng đổi mới sáng tạo được chuyển đổi số; sử dụng hạ tầng số (pháp lý

số, nhân lực số, dữ liệu số, công nghệ số và ứng dụng số) để cung cấp dịch vụ đào tạo cá thể hóa cho người học mọi thể hệ ở trong nước và trên khắp thế giới, đáp ứng yêu cầu học tập suốt đời và phát triển bền vững của các cá nhân cũng như các quốc gia.

Theo những gì Heinemann và Uskov đã trình bày trong nghiên cứu mới nhất của các ông gần đây [20] thì dường như Đại học thông minh (ĐHTM) là một cơ sở đào tạo bậc đại học, trong đó hội tụ tất cả các yếu tố được nêu dưới đây. Đó là: Lớp học thông minh, Môi trường học tập thông minh, Giáo viên thông minh, Khuôn viên thông minh, Cộng đồng học tập thông minh, và cả Phương pháp học tập thông minh. SUF [21] là một diễn đàn sáng kiến dựa trên thành viên toàn cầu, tập trung vào “Đại học thông minh” như một khái niệm toàn diện sáng tạo để mở rộng sự hiểu biết và triển khai các công nghệ mới, chuyển đổi bền vững, thực tiễn học thuật tốt nhất và các xu hướng hiện tại trong quản lý và lãnh đạo vì một tương lai tốt hơn của giáo dục đại học. Tổ chức này hiện diện trên 160 quốc gia với hơn 30.000 trường đại học thành viên trên khắp thế giới. Diễn đàn này lại đưa ra một định nghĩa khác, cụ thể, ĐHTM bao gồm 6 thành tố: Khuôn viên thông minh; Con người thông minh; Đào tạo thông minh; Nghiên cứu thông minh; Quản trị thông minh và Ảnh hưởng thông minh. Đại học Thông minh là một trường đại học nâng cao kinh nghiệm giáo dục; nghiên cứu và làm việc của các bên liên quan bằng cách kết hợp các công nghệ kỹ thuật số; sáng tạo và dựa trên internet để cải thiện xã hội nói chung. Đại học thông minh sẽ xác định những cách thức sắc sảo và có thể chấp nhận được, trong đó các trường đại học và sinh viên tốt nghiệp của họ có thể nâng cao những đóng góp của họ cho xã hội, cho di sản văn hóa của trường đại học và cho tương lai của cộng đồng rộng lớn hơn. Đại học thông minh không phải là một điểm đến hay một cột mốc quan trọng.

Nhờ có các thiết bị giáo dục hiện đại và hệ thống máy chủ tiên tiến, các trường đại học ngày nay có nhiều thuận lợi hơn trong việc tiếp cận và xây dựng hệ thống đại học thông minh. Với sự hỗ trợ của Trung tâm Ứng dụng Công nghệ thông tin – Đại học Kinh tế Quốc dân - ngôi trường định hướng phát triển theo mô hình trường đại học thông minh, nhóm nghiên cứu của chúng tôi tập trung thí điểm mô hình này cho hạng mục Nghiên cứu thông minh (Smart Research). Trong các trường đại học, cần có các mục tiêu cụ thể, đội ngũ nhân viên đáng tin cậy và tiêu chuẩn quy trình cụ thể để trở thành trường đại học thông minh. Chúng tôi dùng các công cụ thu thập và phân tích dữ liệu lớn để tập hợp toàn bộ những dữ liệu về công tác nghiên cứu và đổi mới sáng tạo tại nhà trường để đưa ra các phân tích hỗ trợ nhu cầu ra quyết định của cấp quản lý và nghiên cứu. Phần tiếp theo sẽ trình bày cách thức mà chúng tôi đang triển khai.

4.2. Ứng dụng mô hình NEU IO Portal trong Hệ thống đại học thông minh

Công cuộc số hóa, tích hợp, chia sẻ dữ liệu trở thành nhiệm vụ then chốt theo Quyết định số 749/QĐ-TTg của Thủ tướng Chính phủ: Phê duyệt “Chương trình Chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030” [22] và Quyết định số 714/QĐ-TTg của Thủ tướng Chính phủ: Ban hành Danh mục cơ sở dữ liệu quốc gia cần ưu tiên triển khai tạo nền tảng phát triển chính phủ điện tử [23]. NEU IO Portal là cổng thông tin mà tại đây người dùng có thể truy cập để thực hành quan sát, phân tích và nghiên cứu về dữ liệu lớn trong lĩnh vực

kinh tế trên nền tảng đám mây. Tại đây người nghiên cứu về data science có thể chia sẻ kiến thức, dữ liệu về machine learning, data science từ đó có thể áp dụng những kiến thức của mình vào những dự án thực tế. Các dữ liệu được thu thập hàng năm qua các kênh của Nhà Trường, Open data của Quốc gia, ... Hệ sinh thái các công cụ phân tích dữ liệu lớn được tích hợp và cập nhật liên tục trên portal.

Dữ liệu đầu vào

- Dữ liệu được thu thập từ các dịch vụ có sẵn của nhà trường như portal, ERP, FB, các sensor, ...
- Dữ liệu nghiên cứu từ các nhà nghiên cứu, cán bộ giáo viên, sinh viên, ...
- Dữ liệu trong trao đổi chính sách của Nhà trường với các tổ chức trong và ngoài nước
- Dữ liệu thu thập từ Open data Quốc gia
- Dữ liệu được cung cấp từ doanh nghiệp đặt hàng các chuyên gia phân tích để giải quyết bài toán kinh doanh của doanh nghiệp.

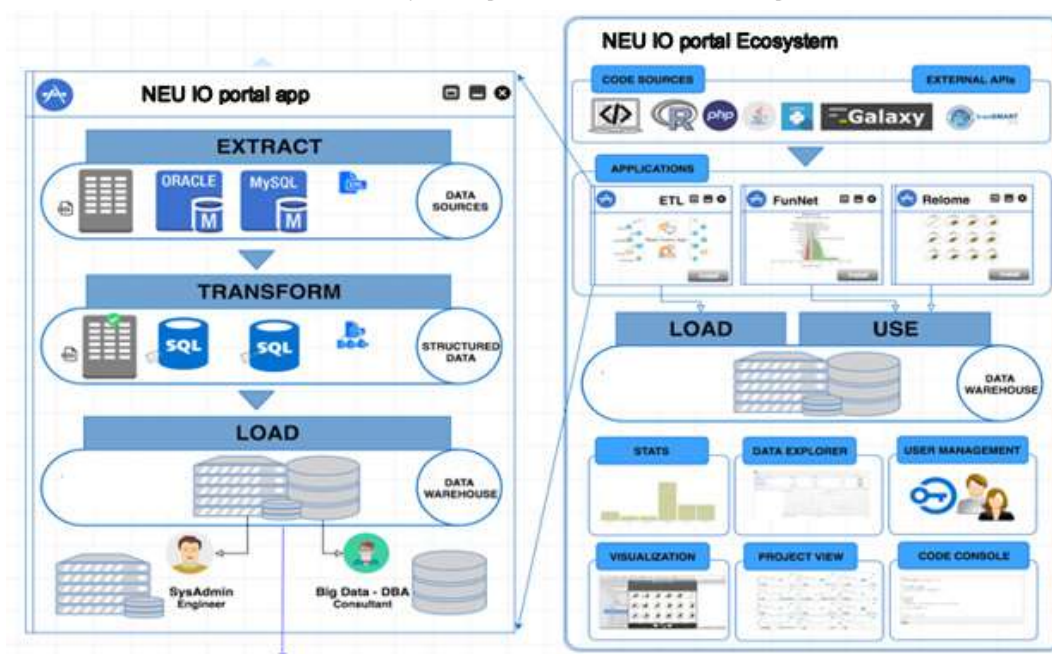
Đối tượng phục vụ

- Các tài khoản truy cập theo phân quyền
- Đối tượng học tập và nghiên cứu trong trường
- Các trường trong khối ngành kinh tế, quản trị kinh doanh
- Các chuyên gia phân tích dữ liệu theo đặt hàng

Các hoạt động

- Tổ chức dữ liệu phục vụ việc thực hành phân tích, nghiên cứu cho sinh viên
- Tổ chức các cuộc thi phân tích dữ liệu cho từng đối tượng

Hình 9: Mô hình hệ thống NEU IO Portal. Tác giả đề xuất



- Thành lập đội chuyên gia phân tích dữ liệu lớn và tư vấn, các chuyên gia này đến từ các giảng viên trong trường, các chuyên gia, nhà khoa học trong cộng đồng.
- Nghiên cứu dữ liệu lớn trong các vấn đề kinh tế để phục vụ tư vấn chính sách cho bộ Giáo dục, Chính phủ.
- Giải quyết bài toán thực tế trên dữ liệu của doanh nghiệp đặt hàng

Mô tả hoạt động của NEU IO Portal

- Phần Back-end: Hệ thống thông tin phân tán cho kho dữ liệu. Sử dụng công nghệ của Hadoop và Spark để lưu trữ và xử lý dữ liệu.
 - ♦ Dữ liệu nhiều nguồn, nhiều định dạng
 - ♦ Kho dữ liệu: rút trích, chuyển đổi và nạp
- Phần Front-end: Giao diện người dùng web cung cấp nhiều chức năng
 - ♦ Cập nhật dữ liệu
 - ♦ Xem dữ liệu
 - ♦ Điều khiển truy cập (phân cấp người dùng)
 - ♦ Trình diễn kết quả một cách sinh động
 - ♦ Kho lưu trữ ứng dụng phân tích dữ liệu
 - ♦ API: phương thức lấy dữ liệu để khai thác hiệu quả

5. KẾT LUẬN

Để tiến tới một Đại học Thông minh, thì một bước đi quan trọng nhất đó chính là việc có được dữ liệu, tổ chức và khai thác dữ liệu. Bài viết đã chỉ ra được sự cần thiết của Trung tâm dữ liệu lớn đối với việc vận hành có hiệu quả Nhà trường, tăng cường cung cấp thông tin đến các đối tượng có lợi ích liên quan (stakeholders). Bên cạnh đó, nghiên cứu đã chỉ ra các thành phần cần có, mô tả chức năng cơ bản của một trung tâm dữ liệu lớn tại Đại học thông minh.

TÀI LIỆU THAM KHẢO

1. Chae and BK, *Khuôn khổ chung để nghiên cứu sự phát triển của hệ sinh thái đổi mới kỹ thuật số: Trường hợp dữ liệu lớn. tạp chí Quản lý Thông tin Quốc tế*, 45, 83–94, 2019.
2. MohammadShorfuzzaman, et al., *Khai thác sức mạnh của phân tích dữ liệu lớn trên đám mây để hỗ trợ phân tích học tập trong môi trường học tập trên thiết bị di động*. Máy tính trong Hành vi Con người, 92 (1), 578–588, 2019.
3. Rehman, M.H., et al., *Vai trò của phân tích dữ liệu lớn trong công nghiệp internet vạn vật*. Hệ thống Máy tính Thế hệ Tương lai, 92, 578–588, 2019.
4. Kalaian, S.A., R.M. Kasim, and N.R. Kasim, *Phương pháp phân tích mô tả và dự đoán cho dữ liệu lớn*. IGI Global, 2019.
5. Dresner, D.v.T.v., *Áp dụng dữ liệu lớn: Trạng thái của thị trường*. ZoomData, 2017.
6. Sorensen, L.C., *Dữ liệu lớn trong quản trị giáo dục: Ứng dụng dự đoán nguy cơ học sinh bỏ học*. Sage, 2018.
7. Hải, V.H., *Dữ liệu lớn và vai trò của nó trong giáo dục đại học*. KT&PT, Số Đặc biệt, tháng 11

- năm 2014, tr. 36-43, 2014.
8. Thor, M.A., *Tổng quan về phân tích dữ liệu lớn trong giáo dục đại học*. Tạp chí Khoa học Giáo dục Kỹ thuật số 50, 2018.
9. Wikipedia. Định nghĩa dữ liệu lớn. Truy cập ngày 19/10/2021; Available from: https://vi.wikipedia.org/wiki/D%E1%BB%AF_li%E1%BB%87u_l%E1%BB%9Bn#%C4%90%E1%B-B%8Bnh_ng%E1%BB%9Bn#%C4%A9a.
10. Glossary, G. *Big Data define*. Accessed: 14 January 2020]; Available from: <https://www.gartner.com/en/information-technology/glossary/big-data>.
11. Mikalef, P., et al., *Khả năng phân tích dữ liệu lớn: Một chương trình nghiên cứu và đánh giá tài liệu có hệ thống*. Information Systems and e-Business Management 16 (3), 547–578, 2018.
12. ShamimHossain GhulamMuhammad, A.S.M., *Phân tích dữ liệu lớn IoT cho ngôi nhà thông minh với sương mù và điện toán đám mây*. Future Generation Computer Systems, 91 (2), 563–573, 2019.
13. Mauro, A.D., et al., *Một định nghĩa chính thức về dữ liệu lớn dựa trên các tính năng thiết yếu của nó*. Emerald insight, 65 (3), 122–135, 2016.
14. ShahatOsman, A.M., *Một khung phân tích dữ liệu lớn mới cho các thành phố thông minh*. Future Generation Computer Systems, 91 (1), 620–633, 2019.
15. tutorialspoint.com. *Hadoop Tutorial*. Access date: 19/10/2021]; Available from: https://www.tutorialspoint.com/hadoop/hadoop_big_data_solutions.htm.
16. myvt.net. *Big Data Viettel vào TOP15 nhóm đóng góp mã nguồn Apache Spark*. [cited Truy cập ngày 19/10/2021; Available from: <https://myvt.net/tin-tuc/big-data-viettel-vao-top15-nhom-dong-gop-ma-nguon-apache-spark/>].
17. baochinhphe.vn. *Ra mắt nền tảng khai phá dữ liệu - Viettel Data Mining Platform*. 2020 Truy cập ngày 31/10/2021]; Available from: <http://baochinhphe.vn/Khoa-hoc-Cong-nghe/Ra-mat-nen-tang-khai-pha-du-lieu-Viettel-Data-Mining-Platform/417348.vgp>.
18. *Nghị quyết số 29-NQ/TW ngày 4/11/2013 Hội nghị Trung ương 8 khóa XI về đổi mới căn bản, toàn diện giáo dục và đào tạo*. 2013.
19. Đức, N.H., et al., *Mô hình khái niệm và xếp hạng đối sánh đại học thông minh V-SMARTH*. 2020.
20. Colleen, H. and V.L. Uskov, *SMART University: Literature Review and Creative Analysis*. 2018.
21. *Diễn đàn ĐHTM (SUF)*. Truy cập ngày 20/9/2021; Available from: <https://s-u-f.org/>.
22. *Quyết định số 749/QĐ-TTg của Thủ tướng Chính phủ : Phê duyệt “Chương trình Chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030”*.
23. *Quyết định số 714/QĐ-TTg của Thủ tướng Chính phủ : Ban hành Danh mục cơ sở dữ liệu quốc gia cần ưu tiên triển khai tạo nền tảng phát triển chính phủ điện tử*.