|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Bộ Giáo dục và Đào tạo**Trường Đại Học Sư Phạm Tp.HCM** | | | | Khoa Công Nghệ Thông Tin *Bộ môn: Nghiên Cứu Khoa Học* | |
| **0B****BÁO CÁO KẾT QUẢ NGHIÊN CỨU** | | | | | |
| Ngày thực hiện báo cáo: | | | | | |
|  | | | | | |
| tÊN ĐỀ TÀI | | | Tìm hiểu về Big Data, ứng dụng và hướng phát triển trong năm năm tới. | | |
| LĨNH VỰC CHUYÊN NGÀNH | | | hệ thống thông tin | | |
| LOẠI HÌNH NGHIÊN CỨU | | | Lí thuyết | | |
| Người hướng dẫn | | | TS Lê Đức Long | | |
| *sinh viên thực hiện 1* | | | Lê Văn Giác | | 43.10.104.037 |
| *sinh viên thực hiện 2* | | | Nguyễn Minh Hiếu | | 43.01.104. |
|  | | | | | |
| **GIỚI THIỆU** | | | | | |
|  | |  | | |  |
| CÂU HỎI GỢI Ý | **Lĩnh vực và nghiên cứu liên quan đã và đang phát triển như thế nào?**  **Các vấn đề, và bài toán đặt ra cần giải quyết là gì?** | | | | |
| hƯỚNG DẪN | Giới thiệu tổng quan về đề tài – những vấn đề và lĩnh vực liên quan đến đề tài. | | | | |
| Lời mở đầu:  Trong thời đại công nghệ thông tin bùng nổ như hiện nay, rất nhiều công nghệ lướt trên những con sóng công nghệ lớn, tác động mạnh lên cuộc sống của nhân loại….Một trong số đó chúng ta không thể không nhắc tới Big Data.Thật vậy trong cuộc sống hiện nay Big Data không ngừng phát triển và cùng với sự phát triển đó là vô số lợi ích mà Big Data đem lại.Do vậy vấn đề đặt ra trong báo cáo này sẽ là tìm hiểu những kiến thức cơ bản về Big Data và hướng phát triển trong thời gian tới. | | | | | |
|  | | | | | |
| Ghi chú | | | | | |
| (1 trang) | | | | | |
|  | | | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **MỤC TIÊU NGHIÊN CỨU CỦA ĐỀ TÀI** | | | |
|  | |  |  |
| CÂU HỎI GỢI Ý | **Mục tiêu nghiên cứu chính của đề tài là gì?**  **Bài toán chính cần phải giải quyết là gì?** | | |
| hƯỚNG DẪN | Đặt bài toán giải quyết và trình bày mục tiêu nghiên cứu chính của đề tài.  Cần phải làm rõ bài toán chính của đề tài – cũng có thể giới thiệu sơ lược các bài toán con (nếu có) | | |
| Với vấn đề được đặt ra thì mục tiêu của nhóm chúng tôi là: -Tim hiểu khái quát về các định nghĩa và lịch sử phát triển cua Big Data.  -Các ứng dụng của BigData trong đời sống.  -Một số mặt tiêu cực của BigData.  -Hướng phát triển của BigData trong năm năm tới(2018 2023). | | | |
|  | | | |
| Ghi chú | | | |
| (1 trang) | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **TÌNH HÌNH NGHIÊN CỨU TRONG VÀ NGOÀI NƯỚC** | | | |
|  | |  |  |
| CÂU HỎI GỢI Ý | **Lĩnh vực và nghiên cứu liên quan đã và đang phát triển như thế nào?**  **Các vấn đề, và bài toán đặt ra cần giải quyết là gì?** | | |
| hƯỚNG DẪN | Tìm hiểu các nghiên cứu đã công bố gần nhất (5 năm trở lại) về lĩnh vực liên quan trong và ngoài nước.  Phân tích các kết quả đóng góp, nhận xét các hạn chế còn tồn tại. | | |
| Những năm gần đây , trên thế giới liên tục cập nhập và nghiên cứu về BigData , Talor&Fancis nghiên cứu và viết về dữ liệu lớn trong nghiên cứu khoa học và khám phá,nhóm nghiên cứu Chaowei Yang, Qunying Huang, Zhenlong Li, Kai Liu & Fei Hu nghiên cứu về dữ liệu lớn và điện toán đám mây cơ hội và thách thức,ở Hội nghị quốc tế về tính toán không gian địa lý Nghiên cứu & Ứng dụng không ít những ứng dụng của BigData được công bố ví dụ Dữ liệu lớn và phân tích không gian nâng cao của nhóm tác giả Dr. Xavier Lopez, Director, Spatial and Semantic Technologies, Oracle Corp.U.S.A.Điện toán đám mây và tính toán dữ liệu lớn (Zhiming Xue,Microsoft Corp.,Reston, VA U.S.A.) Berkovich và Liao 2012 Berkovich, S. và D. Liao. 2012. Về Clusterization của Big Data Streams. 9. ACM. Truy cập ngày 25 tháng 11 năm 2015. Dasgupta 2013 Dasgupta, A. 2013.viết về Big Data: Tương lai là trong Analytics, Thế giới không gian địa lý.,..Và rất nhiều những nghiên cứu khác chuyên sâu về một số lĩnh vực của BigData nữa.  Trong nước việc nghiên cứu,ứng dụng cũng đang rất phát triển ,công trình nghiên cứu của tác giả: Trương Xuân Việt2 (chủ nhiệm đề tài), Nguyễn Hùng Dũng2, Lưu Tiến Đạo2, Trương Quốc Định1, Cù Vĩnh Lộc2 đã dựa trên các nền tảng Big data phổ biến, bao gồm: (1) HDFS (Hadoop Distributed File System) của Hadoop dùng trong quản lý tập tin, (2) Lucene để lập chỉ mục nghịch đảo (Inverted index) cho văn bản và (3) Apache Solr hỗ trợ cơ chế quản lý chỉ mục nghịch đảo, tìm kiếm toàn văn và một số chức năng tìm kiếm nâng cao. xây dựng hệ thống truy vấn dữ liệu phi cấu trúc để hỗ trợ tiền kiếm thông tin nội bộ trong truong đại học. Viện công nghiệp phần mềm (2017)và nội dung số (NISCI) đã ứng dụng thành công hệ thống siêu máy tính vào phân tích dữ liệu, chuẩn hóa dữ liệu để xây dựng cơ sở dữ liệu và trang web quản lý thông tin liệt sỹ, mộ liệt sỹ…Ta có thế thấy được từ các trường dến trong chính quyền đã và đang áp dụng BigData nghiên cứu và sử dụng vào trong thực tiễn cuộc sống.Do đó việc tìm hiểu và nâng cao những hiểu biết về BigData rất quan trọng . | | | |
|  | | | |
| Ghi chú | | | |
| (2-3 trang) | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **NỘI DUNG VÀ PHẠM VI NGHIÊN CỨU** | | | |
|  | |  |  |
| CÂU HỎI GỢI Ý | **Nội dung nghiên cứu là gì?**  **Phạm vi nghiên cứu là gì?** | | |
| hƯỚNG DẪN | Nêu nội dung nghiên cứu, bao gồm về mặt lý thuyết, phương pháp luận, công nghệ/kĩ thuật/phương tiện để nghiên cứu.  Nêu phạm vi nghiên cứu của đề tài, bao gồm việc giới hạn phạm vi nghiên cứu và triển khai, các giả định ban đầu đối với nghiên cứu. | | |
| - Phạm vi nghiên cứu: đề tài khóa luận được nghiên cứu giới hạn trong phạm vi Khoa Công nghệ  thông tin – Trường Đại học Sư phạm TPHCM.  -Về nội dung, đề tài tập trung nghiên cứu các vấn đề chính sau: -Định nghĩa, đặc trưng của BigData.  -Lịch sử hình thành.  -Các ứng dụng của BigData.  -Hướng Phát triển trong năm năm tới. | | | | |
|  | | | |
| Ghi chú | | | |
| (1 trang) | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **CÁC GIẢ THUYẾT VÀ CÁCH TIẾP CẬN GIẢI QUYẾT CÁC BÀI TOÁN** | | | |
|  | |  |  |
| CÂU HỎI GỢI Ý | **Các giả thuyết đặt ra để giải quyết bài toán chính?**  **Các cách tiếp cận để giải quyết bài toán đặt ra?**  **Phương pháp, cách thức cụ thể để giải quyết các bài toán đặt ra là gì?** | | |
| hƯỚNG DẪN | Đặt ra những giả thuyết, hay vấn đề-bài toán con cần phải giải quyết để đạt được mục tiêu nghiên cứu đề tài.  Các cách tiếp cận (dự kiến) để giải quyết các giả thuyết, bài toán con đã đặt.  Trình bày phương pháp, cách thức giải quyết các bài toán con. | | |
| Về Định nghĩa về BigData:  Theo Gartner thì: “Dữ liệu lớn là các tài sản thông tin có khối lượng lớn, tốc độ cao và thông tin cao đòi hỏi các hình thức xử lý thông tin mang tính sáng tạo, hiệu quả về chi phí để tăng cường cái nhìn sâu sắc và đưa ra quyết định.” (“Gartner IT Glossary, n.d.”)  Còn TechAmerica Foundation thì định nghĩa: “Dữ liệu lớn là thuật ngữ mô tả khối lượng lớn dữ liệu tốc độ cao, phức tạp và biến đổi đòi hỏi kỹ thuật và công nghệ tiên tiến để cho phép nắm bắt, lưu trữ, phân phối, quản lý và phân tích thông tin.” (Ủy ban dữ liệu lớn liên bang của TechAmerica Foundation, 2012). Qua hai định nghĩa trên chúng ta có thể phần nào hình dung được cái gì vầ như thế nào được gọi là Bigđata (dữ liệu lớn). Sự phát triển của công nghệ và sự hiểu biết của con người về dữ liệu đã chuyển dữ liệu xử lý từ chế độ tĩnh truyền thống sang trường dữ liệu tăng tốc đặc trưng bởi thể tích, vận tốc, đa dạng, tính xác thực và giá trị (đó chính là 5V đặc trưng của BigData).  Đầu tiên đề cập đến khối lượng dữ liệu (Volume) đang phát triển bùng nổ và mở rộng ra ngoài khả năng xử lý các tập dữ liệu lớn của chúng tôi; khối lượng là bộ mô tả phổ biến nhất của Big Data. Định nghĩa về khối lượng dữ liệu lớn là tương đối và thay đổi theo các yếu tố, chẳng hạn như thời gian và loại dữ liệu. Những gì có thể được coi là dữ liệu lớn ngày nay có thể không đáp ứng ngưỡng trong tương lai bởi vì dung lượng lưu trữ sẽ tăng lên, cho phép các bộ dữ liệu lớn hơn được ghi lại. Ngoài ra, loại dữ liệu, được thảo luận dưới dạng đa dạng, xác định ý nghĩa của ‘big’. Hai tập dữ liệu có cùng kích thước có thể yêu cầu các công nghệ quản lý dữ liệu khác nhau dựa trên loại dữ liệu của chúng, ví dụ: dữ liệu bảng so với video. Vì vậy, định nghĩa về dữ liệu lớn cũng phụ thuộc vào ngành. Do đó, những cân nhắc này khiến việc xác định ngưỡng cụ thể cho khối lượng dữ liệu lớn là không thực tế.  Tốc độ (Velocity) đề cập đến việc tạo nhanh và truyền dữ liệu trên Internet như được minh họa bằng việc thu thập dữ liệu từ mạng xã hội, hàng loạt cảm biến từ vi mô (nguyên tử) đến mức vĩ mô (toàn cầu) và truyền dữ liệu từ cảm biến tới siêu máy tính và người ra quyết định.  Mức độ đa dạng cao (variety), đặc điểm xác định dữ liệu lớn, không nhất thiết phải là mới. Các tổ chức đã tích trữ dữ liệu không có cấu trúc từ các nguồn nội bộ (ví dụ: dữ liệu cảm biến) và các nguồn bên ngoài (ví dụ: phương tiện truyền thông xã hội). Tuy nhiên, sự nổi lên của công nghệ quản lý dữ liệu mới và phân tích, cho phép các tổ chức tận dụng dữ liệu trong quy trình kinh doanh của họ, là khía cạnh đổi mới. Ví dụ, công nghệ nhận diện khuôn mặt cho phép các nhà bán lẻ truyền thống có được thông tin tình báo về lưu lượng truy cập cửa hàng, thành phần tuổi tác hoặc giới tính của khách hàng và mô hình chuyển động trong cửa hàng của họ. Thông tin vô giá này được tận dụng trong các quyết định liên quan đến quảng cáo sản phẩm, vị trí và nhân sự. Dữ liệu Clickstream cung cấp nhiều thông tin về hành vi của khách hàng và các mẫu duyệt web cho các nhà bán lẻ trực tuyến. Clickstream tư vấn về thời gian và chuỗi các trang được khách hàng xem. Sử dụng phân tích dữ liệu lớn, ngay cả các doanh nghiệp vừa và nhỏ (SMEs) có thể khai thác khối lượng lớn dữ liệu bán cấu trúc để cải thiện thiết kế trang web và triển khai các hệ thống giới thiệu sản phẩm bán chéo và cá nhân hiệu quả.  Tính xác thực(Veracity). IBM đã đặt ra tính xác thực là V thứ tư, thể hiện tính không bền vững vốn có trong một số nguồn dữ liệu. Ví dụ, tình cảm của khách hàng trên phương tiện truyền thông xã hội là không chắc chắn trong tự nhiên, vì chúng đòi hỏi sự phán xét của con người. Tuy nhiên, chúng chứa thông tin giá trị. Do đó, cần phải đối phó với dữ liệu không chính xác và không chắc chắn là một khía cạnh khác của dữ liệu lớn, được giải quyết bằng cách sử dụng các công cụ và phân tích được phát triển để quản lý và khai thác dữ liệu không chắc chắn.  Giá trị (value). Oracle đã giới thiệu giá trị là một thuộc tính xác định dữ liệu lớn. Dựa trên định nghĩa của Oracle, dữ liệu lớn thường được đặc trưng bởi “mật độ giá trị thấp” tương đối. Nghĩa là, dữ liệu nhận được trong biểu mẫu ban đầu thường có giá trị thấp so với khối lượng của nó. Tuy nhiên, một giá trị cao có thể thu được bằng cách phân tích khối lượng lớn dữ liệu đó.  (sách International Journal of Information Management)  Theo nhận xét riêng của nhóm thì đặc trưng mang yếu tố quyết định là Giá trị (value) vì khi chúng ta cân nhắc có đầu tư vào nó hay không thì đầu tiên nó phải mang lại cho chúng ta bao nhiêu lợi ích. Ví dụ khi đầu tư phân tích dữ liệu kinh doanh mất 50 tỷ sau đó thu lại rất thấp hoặc rất mất thời gian để thu hồi lại vốn thì có nên đầu tư hay không.  Đã là tiềm hiểu thì không thể bỏ qua lịch sử phát triển của nó. Sau đây là tóm lượt củaBigData (dữ liệu lớn)    1944, Fremont Rider, Thư viện Đại học Wesleyan, “Học giả và Tương lai của Thư viện Nghiên cứu” (sách)Ông Fremont Rider, dựa trên quan sát của ông vào năm 1944 rằng các thư viện của Đại học Mỹ đang tăng gấp đôi sau mỗi 16 năm, được dự đoán vào năm 1944 rằng Thư viện Yale năm 2040 sẽ có“ Khoảng 200.000.000 khối lượng, mà sẽ chiếm hơn 6.000 dặm của kệ ... [yêu cầu] đội ngũ nhân viên biên mục hơn sáu ngàn người.”Rõ ràng, ông Rider không dự đoán số hóa thư viện nhưng ông đã dự đoán chính xác vụ nổ thông tin.Từ năm 1944 đến năm 1980, đã có nhiều bài báo và bài thuyết trình quan sát thấy 'bùng nổ thông tin' và nhu cầu về dung lượng lưu trữ.  1980, nhà xã hội học Charles Tilly, “Lịch sử xã hội mới và lịch sử xã hội cũ mới” Những người phát hiện ra Oxford English Discovery phát hiện ra rằng ông Tilly là người đầu tiên sử dụng thuật ngữ Big Data trong câu này trong bài viết của mình,  "Không ai trong số những câu hỏi lớn đã thực sự mang lại cho bludgeoning của những người lớn dữ liệu."  Biên tập viên Whatsthedata.com cho rằng việc sử dụng của ông Tilly không nằm trong bối cảnh những gì chúng tôi hiểu Big Data hiện nay và tôi đồng ý với họ.  1990, Peter Denning, "Tiết kiệm tất cả các bit", nhà khoa học người Mỹ  Ông Denning, theo ý kiến ​​của tôi, đã giới thiệu khái niệm về những gì có thể với Big Data trong những câu sau đây.  "Sự bắt buộc [đối với các nhà khoa học] để cứu tất cả các bit buộc chúng ta vào một tình huống không thể"và anh ấy tiếp tục nói “Có thể xây dựng các máy có thể nhận ra hoặc dự đoán các mẫu dữ liệu mà không hiểu ý nghĩa của các mẫu. Những máy như vậy cuối cùng có thể đủ nhanh để xử lý các luồng dữ liệu lớn trong thời gian thực”. Ông Denning thực sự đóng đinh những gì có thể với Big Data cách trở lại vào năm 1990. Ấn tượng! Bây giờ chúng tôi nhanh chóng chuyển tiếp đến 1997-1998 khi chúng tôi bắt đầu thấy việc sử dụng thực tế thuật ngữ Big Data trong bối cảnh hiện đại của nó.  1997, Michael Cox và David Ellsworth, “Phân trang nhu cầu được kiểm soát bởi ứng dụng để hiển thị ngoài lõi”. Bài viết của ông Cox và ông Ellsworth là bài viết đầu tiên trong thư viện số ACM sử dụng thuật ngữ “dữ liệu lớn” như sau. “Visualization cung cấp một thách thức thú vị cho các hệ thống máy tính: các bộ dữ liệu nói chung là khá lớn, đánh thuế khả năng của bộ nhớ chính, đĩa cục bộ và thậm chí cả đĩa từ xa. Chúng tôi gọi đây là vấn đề của dữ liệu lớn. Khi bộ dữ liệu không vừa trong bộ nhớ chính (trong lõi), hoặc khi chúng không vừa với đĩa cục bộ, giải pháp phổ biến nhất là thu được nhiều tài nguyên hơn.”  1998, John Mashey, Nhà khoa học trưởng tại SGI “Dữ liệu lớn… và làn sóng hạ tầng tiếp theo”. Đây là nơi cốt truyện dày lên. Bài báo của tờ New York Times ghi nhận ông Mashey lần đầu tiên sử dụng thuật ngữ 'Dữ liệu lớn'. Mặc dù Michael Cox và David Ellsworth dường như đã sử dụng thuật ngữ 'Dữ liệu lớn' trong bản in, ông Mashey được cho là đã sử dụng thuật ngữ này trong các bài phát biểu khác nhau của mình và đó là lý do tại sao ông được cho là đã đến với Big Data. Theo lời của ông Mashey,"Tôi đã sử dụng một nhãn cho một loạt các vấn đề, và tôi muốn cụm từ đơn giản nhất, ngắn nhất để truyền đạt rằng ranh giới của máy tính tiếp tục tiến bộ."  2000, Francis Diebold, “Mô hình nhân tố động” dữ liệu lớn cho đo lường và dự báo kinh tế vĩ mô. Theo bài báo Times, ông Diebold nghĩ (có thể là lưỡi trong má) rằng ông là người đầu tiên sử dụng dữ liệu lớn hạn cho đến khi Times cho tín dụng cho ông Mashey. Tuy nhiên, ông Diebold nên nhận được một số tín dụng cho prescience của mình là nói rằng “Gần đây, nhiều khoa học tốt, dù là vật lý, sinh học hay xã hội, đã bị buộc phải đương đầu — và thường được hưởng lợi từ — hiện tượng “Dữ liệu lớn”. Big Data đề cập đến sự bùng nổ về số lượng (và đôi khi, chất lượng) của dữ liệu có sẵn và có khả năng liên quan, phần lớn là kết quả của những tiến bộ gần đây và chưa từng có trong công nghệ ghi và lưu trữ dữ liệu.”  Ông Diebold lần đầu tiên liên kết Big Data một cách rõ ràng với cách chúng ta hiểu thuật ngữ ngày nay.  2001, Trong bài viết quản lý dữ liệu 3D của mình: Kiểm soát khối lượng dữ liệu, vận tốc và đa dạng Doug Laney, nhà phân tích của Gartner, định nghĩa ba điều sẽ trở thành đặc điểm thường được chấp nhận của Big Data.  2005, Các nhà bình luận thông báo rằng chúng ta đang chứng kiến ​​sự ra đời của "Web 2.0" - trang web do người dùng tạo ra, nơi phần lớn nội dung sẽ được cung cấp bởi người dùng dịch vụ, chứ không phải chính các nhà cung cấp dịch vụ. Điều này đạt được thông qua việc tích hợp các trang web kiểu HTML truyền thống với các cơ sở dữ liệu mặt sau rộng lớn được xây dựng trên SQL. 5,5 triệu người đã sử dụng Facebook, được giới thiệu một năm trước đó, để tải lên và chia sẻ dữ liệu của riêng họ với bạn bè.  Năm nay cũng thấy việc tạo ra Hadoop - khung công tác nguồn mở được tạo ra đặc biệt cho việc lưu trữ và phân tích các tập dữ liệu lớn.  2007, Wired mang đến khái niệm Big Data cho công chúng với bài viết của họ The End of Theory: Deluge dữ liệu làm cho mô hình khoa học lỗi thời.  2008, Các máy chủ của thế giới xử lý 9.57 zettabyte (9.57 nghìn tỷ gigabyte) thông tin - tương đương với 12 gigabyte thông tin cho mỗi người, mỗi ngày), theo thông tin bao nhiêu? Báo cáo năm 2010. Trong sản xuất và phổ biến thông tin quốc tế, ước tính có 14,7 exabyte thông tin mới được sản xuất trong năm nay.  2009, Công ty Mỹ trung bình với hơn 1.000 nhân viên đang lưu trữ hơn 200 terabyte dữ liệu theo báo cáo Big Data: Biên giới tiếp theo cho sự đổi mới, cạnh tranh và năng suất của McKinsey Global Institute.  (Bernerd Marr ,2015)  Tất cả nói cho chúng ta hiểu một điều rằng BigData không phải là một hiện tượng đột nhiên xuất hiện mà là một phần trong quá trình phát triển và sử dụng dữ liệu. Giống như một nấc thang cao hơn giúp chúng ta có một cái nhìn khác hơn về kinh tế và xã hội mà thôi.  Có bề dày phát triển như thế thì hiện tại BigData có những ứng dụng/công nghệ như thế nào. Bảng sau sẽ cho thấy được các tính năng (đặc trưng) của BigData trong các miền khác nhau:   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | |  | | Khoa học về Trái Đất | IOT | Khoa học xã hội | Thiên văn học | Kinh doanh | Công nghiệp | | Khối lượng (volume) | | x | X | X | X | X | X | | Tốc độ (Velocity) | | x | X | X | X | X | X | | Đa dạng (variety) | | x |  | X | X |  |  | | Xác thực  (Veracity) | Kích thước cao | x | X | X | X | X | X | | Không thời gian | X | X | X |  | X | X | | Nhiều nguồn | X |  | X | X |  |  | | Giá trị  (value) | | x | X | X | x | X | x |   ([Chaowei Yang](https://www.tandfonline.com/author/Yang%2C+Chaowei), [Qunying Huang](https://www.tandfonline.com/author/Huang%2C+Qunying), [Zhenlong Li](https://www.tandfonline.com/author/Li%2C+Zhenlong), [Kai Liu](https://www.tandfonline.com/author/Liu%2C+Kai) & [Fei Hu](https://www.tandfonline.com/author/Hu%2C+Fei),2016)  Thực tiễn đời sống Bigdata đã có rất nhiều ứng dụng ví dụ như trong chính trị thì Big Data được khai thác thông qua việc phân tích các dữ liệu về xu hướng của người dân điển hình là cựu tổng thống Mỹ Obama đã ứng dựng BigData trong việc phân tích xu hướng người dân vạch kế hoạch và ông đã thành công. Trong thương mại dữ liệu lớn giúp chúng ta thực hiện nhiều công việc phức tạp như: Phân tích hành vi , thói quen tiêu dung, phân tích và quản lý chuổi cung ứng, phân tích tiếp thị chéo kênh.bán hàng theo xu hướng của người tiêu dung.Điển hình là trang Amazon tự xây dựng cho mình cơ sở lưu là phân tích dữ liệu của người tương tác trên web,FaceBook,ZaLo cũng tương tự như vậy họ lưu trữ thông tin và phân tích trích xuất chúng tùy vào mục đích mà họ sử dụng.Cùng với đó thì bảo mật thông tin cũng phát triển lên để đảm bảo cho sự an toàn của thông tin mà họ đang lưu trữ.Trong giao thông thì Sử dụng công nghệ CDR (call duration recond) trong quá khứ để ước lượng mật độ lưu thông trong thành phố vào cac giờ cao điểm để lên kế hoạch phân luồn giao thông hoặc cử cảnh sát giao thông trực tiếp tới điểm nóng để điều phối giao thông. Tổng đài giao thông đưa ra những lời khuyên cho người đang lưu thông bằng việc thu thập và phân tích thông tin từ những thiết bị di động, các camera giao thông, người lưu thông được định hướng những điểm ùn tắc qua đó giảm thiểu tình trạng kẹt xe ùn tắc rất đáng kể. Trong y tế Ứng dụng NIH: Flu Information dựa trên số lượng tìm kiếm ở một khu vực nào đó sau đó phân tích đối chiếu các kết quả sau cùng đưa ra dự báo về hướng phát triển dịch bệnh tại khu vực đó. Qua đó cho biết tình hình dịch bệnh ở khu vực đó diễn ra như thế nào và đưa ra những biện pháp phòng tránh…  Sau cùng là nhận định về hướng phát triển sắp tới của BigData.  Big Data sử dụng điện toán đám mây. Điện toán đám mây đã có mặt khắp mọi nơi. Dữ liệu lớn góp phần vào tốc độ tăng trưởng của điện toán đám mây. Amazon Redshift (hệ thống kho dữ liệu của Amazon Web Services) và Google BigQuery (hệ dịch vụ phân tích dữ liệu cỡ lớn của Google) đang tăng trưởng nhanh chóng. Các công ty dữ liệu hàng đầu như Teradata cũng đang mở rộng thị trường.  Hệ sinh thái dữ liệu lớn. Khi các cách thức thực hiện mới xuất hiện, một hệ sinh thái mới cũng xuất hiện theo. Dữ liệu lớn cũng vậy, đã xuất hiện các cách thức mới về lưu trữ, sàng lọc, bổ sung dữ liệu, ứng dụng mạng xã hội, phân tích dữ liệu ở cấp độ cao nhất và hiển thị dữ liệu. Chúng ta sẽ còn được chứng kiến nhiều sự thay đổi kinh ngạc khác từ hệ sinh thái hiện hữu. Đặc biệt là các hệ dữ liệu theo phương thức xử lý song song hàng loạt MPP sẽ thay đổi vai trò và không còn chiếm ưu thế như trước nữa.  Sự phát triển của vạn vật kết nối IoT tạo ra những giải pháp dữ liệu mới:  Khái niệm data lake. Một khối lượng dữ liệu khổng lồ tồn tại ở dạng thô một cách tự nhiên được ví với một cái hồ đầy dữ liệu (data lake). Data lake là một khái niệm mới nảy sinh và chắc chắn sẽ còn có thêm nhiều bước tiến mới trong việc quản trị và bảo mật big data lake với những phương thức quản trị mới. (Trinh Minh Giang). | | | |
|  | | | |
| Ghi chú | | | |
| (3-5 trang) | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **KẾT QUẢ VÀ ĐÓNG GÓP CỦA ĐỀ TÀI** | | | |
|  | |  |  |
| CÂU HỎI GỢI Ý | **Kết quả và đóng góp của đề tài là gì?** | | |
| hƯỚNG DẪN | Nêu kết quả và đóng góp của đề tài – cụ thể kết quả về mặt lý thuyết, kết quả về thực nghiệm. Việc cài đặt và thử nghiệm như thế nào? Triển khai trong thực tế như thế nào?  So sánh với các công trình, kết quả đã công bố có gì khác biệt? Tính mới và tính cấp thiết của đề tài đã thể hiện như thế nào? | | |
| Với mục tiêu và nội dung nghiên cứu như trên, đề tài tìm hiểu BigData, ứng dụng và hướng phát triển do nhóm Skyler đạt được kết quả sau: - Hệ thống khái niệm, lịch sử của BigData và những đặc trưng có liên quan.  - Ứng dụng vào các miền trong thực tiễn.  - Các hướng phát triển của BigData trong thời gian tới.  Đóng góp chính: củng cố về mặt lý thuyết những hiểu biết cơ bản về BigData, giúp mọi người có cái nhìn cụ thể hơn và làm nền tảng để nghiên cứu vận dụng vào trong thực tiễn những đặc tính nào. | | | |
|  | | | |
| Ghi chú | | | |
| (1 trang) | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN** | | | |
|  | |  |  |
| CÂU HỎI GỢI Ý | **Đề tài đã đạt được các mục tiêu đề ra hay không?**  **Đề tài nghiên cứu đã làm sáng tỏ vấn đề gì mới? Mang tính cấp thiết như thế nào?**  **Phân tích và bàn luận những kết quả đạt được? Những hạn chế còn tồn tại là gì?**  **Hướng phát triển của đề tài là gì?** | | |
| hƯỚNG DẪN | Nêu những kết quả và đóng góp chính của đề tài, cũng như việc phát triển đề tài trong tương lai. | | |
| Báo cáo đã thực hiện mục tiêu đề ra của nhóm nghiên cứu là: -Hệ thống khái niệm, lịch sử của BigData và những đặc trưng có liên quan.  - Ứng dụng vào các miền trong thực tiễn.  - Các hướng phát triển của BigData trong thời gian tới.  Các mặt hạn chế và hướng khắc phục của đề tài: - Thời gian timg hiểu, nghiên cứu bị hạn chế khá nhiều.  - Thiếu sót một số kỹ năng trong nghiên cứu.  - Cần hoàn thiện hơn trong mặt tìm hiểu về hướng phát triển. | | | |
|  | | | |
| Ghi chú | | | |
| (1 trang) | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **TÀI LIỆU THAM KHẢO** | | | |
|  | |  |  |
| CÂU HỎI GỢI Ý | **Các tài liệu tham khảo thuộc lĩnh vực nghiên cứu liên quan là gì?** | | |
| hƯỚNG DẪN | 1. Sách International Journal of Information 2. https://www.tandfonline.com/doi/full/10.1080/17538947.2016.1239771?src=recsys 3. Tổng quan về dữ liệu lớn (Trung tâm thông tin khoa học thống kê Viện Nghiên Cứu khoa Học. 4. <https://www.kdnuggets.com/2017/02/origins-big-data.html> 5. https://portalparts.acm.org/2350000/2345316/fm/frontmatter.pdf?ip=100.36.182.180&CFID=791206509&CFTOKEN=86783680 | | |
|  | | | |
|  | | | |
| Ghi chú | | | |
|  | | | |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | |  |
|  |  |  | | |

|  |  |  |
| --- | --- | --- |
| Người thực hiện đề cương | | |
| sinh viên thực hiện 1 | Họ và tên | Mã số sv | |
| sinh viên thực hiện 2 | Họ và tên | Mã số sv | |

|  |  |  |
| --- | --- | --- |
| ĐÁNH GIÁ của giáo viên hướng dẫn | | |
| Người hướng dẫn | Họ và tên | | |
| Ý kiến |  | | |
|  | ❒ đồng ý cho bảo vệ | ❒ không đồng ý cho bảo vệ | |