TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÀI TẬP LỚN QUẢN TRỊ HỆ THỐNG THÔNG TIN**

**HOW RELIABLE IS BIG DATA**

*Người hướng dẫn*: **GV. HỒ THỊ LINH**

*Người thực hiện*: **TRẦN THỊ VẸN – 52100674**

**TRƯƠNG THÔNG THẾ THÁI - 52000714**

Lớp **: 21050301**

Khoá  **: 25**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÀI TẬP LỚN QUẢN TRỊ HỆ THỐNG THÔNG TIN**

**HOW RELIABLE IS BIG DATA**

*Người hướng dẫn*: **GV. HỒ THỊ LINH**

*Người thực hiện*: **TRẦN THỊ VẸN – 52100674**

**TRƯƠNG THÔNG THẾ THÁI - 52000714**

Lớp **: 21050301**

Khoá  **: 25**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

LỜI CẢM ƠN

Trong suốt quá trình học tập và rèn luyện, chúng em đã nhận được rất nhiều sự giúp đỡ tận tình, sự quan tâm, chăm sóc của GV.Hồ Thị Linh. Ngoài ra, chúng em còn được GV truyền đạt những kiến thức, phương pháp mới về toán hay ho và thú vị, thầy cô còn giúp sinh viên có được nhiều niềm vui trong việc học và cảm thấy thoải mái, … Chúng em xin chân thành cảm ơn các thầy cô rất nhiều trong suốt quá trình học tập này!

Bởi lượng kiến thức của chúng em còn hạn hẹp và gặp nhiều vấn đề trong quá trình học nên báo cáo này sẽ còn nhiều thiếu sót và cần được học hỏi thêm. Chúng em rất mong em sẽ nhận được sự góp ý của quý thầy cô về bài báo cáo này để chúng em rút kinh nghiệm trong những môn học sắp tới. Cuối cùng, chúng em xin chân thành cảm ơn quý thầy cô.

TP Hồ Chí Minh, ngày 01 tháng 11 năm 2023

Sinh viên:

*Trần Thị Vẹn – 52100674*

*Trương Thông Thế Thái - 52100714*

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là sản phẩm đồ án của chúng tôi và được sự hướng dẫn của GV. Phù Trần Tín. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 01 tháng 11 năm 2023*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Trần Thị Vẹn*

*Trương Thông Thế Thái*

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

**Phần xác nhận của GV hướng dẫn**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

**Phần đánh giá của GV chấm bài**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

TÓM TẮT

Big data đặt ra thách thức đối với bảo mật và quyền riêng tư khi lượng lớn thông tin cá nhân được thu thập và phân tích. Big data giúp doanh nghiệp phân tích, thấu hiểu và nhắm đúng khách hàng mục tiêu. Hiểu được hành vi, xu hướng mua hàng của khách hàng, đồng thời dự đoán tình hình thị trường tương lai để đưa ra các quyết định kinh doanh chính xác hơn.

Big data cho phép các doanh nghiệp điều chỉnh sản phẩm phù hợp với thị trường mục tiêu của họ mà không phải chi nhiều tiền cho các chiến dịch quảng cáo không hiệu quả. Bằng cách theo dõi các giao dịch tại điểm bán hàng (POS) và mua hàng trực tuyến, các doanh nghiệp có thể sử dụng Big data để nghiên cứu khách hàng tiềm năng.

Sử dụng những nghiên cứu sâu sắc này, doanh nghiệp có thể đưa ra các chiến lược tiếp thị và mục tiêu đúng hướng, đáp ứng mong đợi của người tiêu dùng, thúc đẩy lòng trung thành với thương hiệu.

MỤC LỤC

TÓM TẮT iv

MỤC LỤC v

CHƯƠNG 1: TỔNG QUAN BIG DATA 1

1.1 Định nghĩa về Big Data: 1

1.2 Phân loại Big data 1

1.3 Đặc trưng của Big Data: 2

1.4 Vai trò của Big Data trong doanh nghiệp 3

1.5 Ứng dụng của big data 4

1.5.1 Ngành tài chính – Ngân hàng 4

1.5.2 Y tế, chăm sóc sức khỏe 4

1.5.3 Thương mại điện tử 4

1.5.4 Ngành bán lẻ 5

1.5.5 Tiếp thị kỹ thuật số 5

1.5.6 Ngành công nghiệp 5

1.5.7 Lĩnh vực giáo dục đào tạo 6

1.5.8 Cơ sở hạ tầng IT 6

CHƯƠNG 2: TÌM HIỂU VỀ CASE STUDY 7

2.1 Tóm tắt các sự việc, sự kiện quan trọng: 7

2.2 Nhận diện và phân tích vấn đề 7

2.2.1 Vấn đề chính của big data ngày nay 7

2.2.2 Dữ liệu lớn cũng có những hạn chế 7

2.2.3 Nắm bắt xu thế của các công ty lớn và việc ứng dụng vào thực tế 8

2.2.4 Những sai lầm khi chưa tìm hiểu và chuẩn bị tốt để áp dụng big data: 9

2.3 Các trường hợp trong thực tế 10

2.4 Một số thách thức cho Big data 13

2.5 Các giải pháp cho big data 13

2.6 Lựa chọn giải pháp và cách thực hiện 15

2.7 Ví dụ thực tế 16

2.8 Bài học rút ra 18

TÀI LIỆU THAM KHẢO I

CHƯƠNG 1: TỔNG QUAN BIG DATA

* 1. Định nghĩa về Big Data

Dữ liệu lớn là thuật ngữ mô tả khối lượng lớn dữ liệu - cả có cấu trúc và không cấu trúc - tràn ngập hoạt động kinh doanh hàng ngày của doanh nghiệp.

Big Data hay “dữ liệu lớn” là thuật ngữ chỉ về các tập dữ liệu khổng lồ và phức tạp, đến mức khó có thể xử lý được bằng các phương pháp truyền thống. Doanh nghiệp sử dụng lượng dữ liệu khổng lồ này để phân tích, chuyển hóa thành thông tin quan trọng để giải quyết các vấn đề liên quan. Dữ liệu lớn có thể được phân tích để có những hiểu biết sâu sắc giúp đưa ra các quyết định và động thái kinh doanh chiến lược tốt hơn.

* 1. Phân loại Big data

**Dữ liệu có cấu trúc:** Dữ liệu có cấu trúc được xem là dữ liệu đơn giản nhất để quản lý và tìm kiếm. Nó là những dữ liệu có thể truy cập, lưu trữ và xử lý ở định dạng cố định. Các thành phần của dữ liệu có cấu trúc được phân loại dễ dàng, cho phép các nhà thiết kế và quản trị viên cơ sở dữ liệu xác định các thuật toán đơn giản để tìm kiếm và phân tích.

**Dữ liệu phi cấu trúc:** Dữ liệu phi cấu trúc là bất kỳ tập hợp dữ liệu nào không được tổ chức hoặc xác định rõ ràng. Loại dữ liệu này hỗn loạn, khó xử lý, khó hiểu và đánh giá. Nó không có cấu trúc cố định và có thể thay đổi vào những thời điểm khác nhau. Dữ liệu phi cấu trúc bao gồm các nhận xét, tweet, lượt chia sẻ, bài đăng trên mạng xã hội, video trên YouTube mà người dùng xem,...

**Dữ liệu bán cấu trúc:** Dữ liệu bán cấu trúc là sự kết hợp giữa dữ liệu có cấu trúc và dữ liệu phi cấu trúc. Email là một ví dụ điển hình vì chúng bao gồm dữ liệu phi cấu trúc trong nội dung thư, cũng như nhiều thuộc tính tổ chức khác như người gửi, người nhận, chủ đề và ngày tháng. Các thiết bị sử dụng gắn thẻ địa lý, thời gian cũng có thể cung cấp dữ liệu có cấu trúc bên cạnh nội dung phi cấu trúc.

* 1. Đặc trưng của Big Data

**Volume (Số lượng lưu trữ)**

Big Data (“dữ liệu lớn”) là tập hợp dữ liệu có dung lượng lưu trữ vượt mức đảm đương của những ứng dụng và công cụ truyền thống. Kích cỡ của Big Data đang từng ngày tăng lên, và tính đến năm 2012 thì nó có thể nằm trong khoảng vài chục terabyte cho đến nhiều petabyte (1 petabyte = 1024 terabyte) chỉ cho một tập hợp dữ liệu mà thôi.

**Velocity (Tốc độ xử lý):** Dung lượng gia tăng của dữ liệu rất nhanh và tốc độ xử lý đang tiến tới real-time. Các ứng dụng phổ biến trên lĩnh vực Internet, Tài chính, Ngân hàng, Hàng không, Quân sự, Y tế – Sức khỏe ngày hôm nay phần lớn dữ liệu lớn đc xử lý real-time. Công nghệ xử lý dữ liệu lớn ngày một tiên tiến cho phép chúng ta xử lý tức thì trước khi chúng được lưu trữ vào cơ sở dữ liệu.

**Variety (Đa dạng chủng loại):** Hình thức lưu trữ và chủng loại dữ liệu ngày một đa dạng hơn. Trước đây chúng ta hay nói đến dữ liệu có cấu trúc thì ngày nay hơn 80% dữ liệu trên thế giới được sinh ra là phi cấu trúc (tài liệu, blog, hình ảnh, video, voice v.v.). Công nghệ Big Data cho phép chúng ta ngày nay liên kết và phân tích đa dạng chủng loại dữ liệu với nhau như comments/post của một nhóm người dùng nào đó trên Facebook với thông tin video được chia sẻ từ Youtube và Twitter.

**Veracity (Độ chính xác):** Một trong những tính chất phức tạp nhất của BigData là độ chính xác của dữ liệu. Với xu hướng Social Media và Social Network ngày nay và sự gia tăng mạnh mẽ tính tương tác và chia sẻ của người dùng Mobile làm cho bức tranh xác định về độ tin cậy & chính xác của dữ liệu ngày một khó khăn hơn. Bài toán phân tích và loại bỏ dữ liệu thiếu chính xác và nhiễu đang là tính chất quan trọng của BigData.

**Value (Giá trị thông tin):** Giá trị thông tin là tính chất quan trọng nhất của xu hướng công nghệ Big Data. Ở đây doanh nghiệp phải hoạch định được những giá trị thông tin hữu ích của BigData cho vấn đề, bài toán hoặc mô hình hoạt động kinh doanh của mình. Có thể nói việc đầu tiên là phải xác định được tính chất “Value” thì mới nên bắt tay vào BigData.

* 1. Vai trò của Big Data trong doanh nghiệp

**Hiểu và nhắm đúng khách hàng mục tiêu:** Dữ liệu của Big Data được thu thập từ nhiều nguồn khác nhau, gồm cả lịch sử trình duyệt, mạng xã hội,... Đây là những kênh người dùng sử dụng rất thường xuyên. Vì vậy, doanh nghiệp phân tích Big Data sẽ hiểu được hành vi, sở thích, nhu cầu của khách hàng đồng thời phân loại và lựa chọn đúng đối tượng khách hàng phù hợp với sản phẩm, dịch vụ của doanh nghiệp.

**Định lượng và tối ưu hóa hiệu suất cá nhân:** Nhờ các thiết bị di động thông minh như smartphone, smartwatch mà thông tin, dữ liệu cá nhân có thể được thu thập dễ dàng. Những dữ liệu cá nhân từ mỗi người dùng sẽ giúp doanh nghiệp có cái nhìn rõ nét về xu hướng hành vi mới nhất của khách hàng. Từ đó, các nhà quản lý có thể định hướng cho chiến lược, kế hoạch mới trong tương lai gần.

**Phòng chống an ninh, giảm thiểu rủi ro:** Big Data được các doanh nghiệp sử dụng để thăm dò, phát hiện và ngăn chặn các nguy cơ, rủi ro về gian lận, xâm nhập hệ thống, đánh cắp thông tin mật. Ví dụ như các ngân hàng và công ty thẻ tín dụng dùng Big Data để ngăn chặn các giao dịch gian lận.

**Tối ưu hóa giá cả:** Việc định giá sản phẩm, dịch vụ luôn là điều rất quan trọng và cũng rất khó khăn đối với các doanh nghiệp. Bởi vì doanh nghiệp cần nghiên cứu rất kỹ về mong muốn khách hàng và mức giá hiện tại của đối thủ cạnh tranh. Nhờ có Big Data mà công cuộc nghiên cứu, thu thập dữ liệu trên trở nên dễ dàng, nhanh chóng, giúp doanh nghiệp định giá một cách chính xác, hiệu quả hơn.

**Nắm bắt các giao dịch tài chính:** Thương mại điện tử đã và đang phát triển mạnh mẽ trên toàn thế giới, trong đó có cả Việt Nam. Do đó, các giao dịch tài chính trên website hay các app thương mại điện tử ngày càng tăng lên và thể hiện rõ hành vi mua của khách hàng. Các thuật toán Big Data được doanh nghiệp sử dụng để gợi ý, đưa ra quyết định giao dịch cho khách hàng, tăng tỷ lệ chốt đơn cho doanh nghiệp.

**Hỗ trợ đổi mới:** Đổi mới kinh doanh dựa trên những hiểu biết mà các nhà phân tích khám phá thông qua Big data. Việc phát triển sản phẩm có thể được thực hiện bằng cách biết được người tiêu dùng nghĩ gì về sản phẩm/ dịch vụ của doanh nghiệp. Các nhà quản lý phải đưa ra quy trình theo dõi, phản hồi, và các đối thủ trong thị trường hiện tại. Phân tích dữ liệu đồng thời cũng giúp giám sát thị trường theo thời gian thực, hỗ trợ đổi mới kịp thời để bắt kịp xu hướng.

* 1. Ứng dụng của big data
     1. Ngành tài chính – Ngân hàng

Trong lĩnh vực tài chính ngân hàng, Dữ liệu lớn được sử dụng để phân tích và xác định các địa điểm tập tin trung tâm theo nhiều nhu cầu của khách hàng tiềm ẩn khả năng phát triển chi nhánh mới, dự đoán lượng tiền mặt cần có sẵn để cung cấp ứng dụng ở một chi nhánh tại thời điểm cụ thể, nâng cao hệ thống ngân hàng kỹ thuật số, phát hiện các hoạt động gian lận và báo cáo cho các chuyên gia liên quan giúp đảm bảo an ninh cho ngân hàng.

* + 1. Y tế, chăm sóc sức khỏe

Trong ngành y tế, Big Data không chỉ được ứng dụng để xác định phương hướng điều trị mà còn giúp cải thiện quá trình chăm sóc sức khỏe. Big Data giúp dự đoán thời điểm cần có sự hiện diện của bác sĩ, giúp theo dõi tình hình bệnh nhân bằng cách theo dõi hồ sơ sức khỏe điện tử, theo dõi bệnh nhân và gửi báo cáo cho các bác sĩ liên quan. Ngoài ra, Big Data còn có thể đánh giá các triệu chứng và xác định nhiều căn bệnh ở giai đoạn đầu, lưu giữ các hồ sơ nhạy cảm được bảo mật và lưu trữ dữ liệu theo một cách hiệu quả. Đặc biệt, nó cũng có thể dự báo trước các khu vực có nguy cơ bùng phát dịch bệnh nguy hiểm.

* + 1. Thương mại điện tử

Đối với ngành thương mại điện tử, doanh nghiệp nào sở hữu và ứng dụng tốt Dữ liệu lớn sẽ tạo ra lợi thế cạnh tranh lớn trên thị trường. Big Data giúp nhà quản lý xác định sản phẩm nào được xem nhiều nhất để tối ưu thời gian hiển thị, tự động gửi mã ưu đãi cho những sản phẩm khách hàng bỏ sọt vào hàng nhưng không mua. Đặc biệt, Big Data có thể phân tích hành vi, sở thích, sự quan tâm của khách hàng, giúp nhà quản lý lý trí hơn về khách hàng để cung cấp các sản phẩm theo đúng xu hướng, nhu cầu thị trường.

* + 1. Ngành bán lẻ

Dữ liệu lớn trong việc xây dựng mô hình chi tiêu của khách hàng, giúp dự đoán cung – cầu cho các sản phẩm. Kết hợp với dữ liệu về thời điểm giao dịch, dữ liệu truyền thông xã hội, dự báo thời tiết để xác định chính xác sản phẩm phù hợp nhất để luôn sẵn sàng cung ứng cho khách hàng. Các nhà quản lý có thể dựa vào dữ liệu về thói quen mua hàng, sở thích của khách hàng để xác định vị trí, cách bố trí sản phẩm trên kệ hàng và đưa ra các chiến lược kinh doanh mới để cải thiện.

* + 1. Tiếp thị kỹ thuật số

Digital Marketing ngày càng trở thành một công cụ quan trọng đối với các doanh nghiệp hiện nay. Khi ứng dụng Big Data, doanh nghiệp sử dụng Digital Marketing có thể xác định các đối tượng mục tiêu trên các trang mạng xã hội dựa trên thông tin nhân khẩu học, giới tính, tuổi tác, sở thích. Bên cạnh đó, có thể cá nhân hóa các hoạt động tìm kiếm trên Google, Email Marketing, hiển thị quảng cáo phù hợp và tạo báo cáo chi tiết sau mỗi chiến dịch quảng cáo.

* + 1. Ngành công nghiệp

Phát triển phần mềm sản phẩm: Các công ty sử dụng Dữ liệu lớn để xây dựng các mô hình dự đoán cho các sản phẩm và dịch vụ mới. Họ phân loại các thuộc tính chính của sản phẩm/dịch vụ trong quá khứ và hiện tại, sau đó mô hình hóa mối quan hệ giữa các thuộc tính và phát triển thành phần mềm hoàn chỉnh.

Tăng trải nghiệm khách hàng: Big Data cho phép bạn thu thập dữ liệu từ các phương tiện truyền thông xã hội, lịch sử web, nhật ký cuộc gọi và các nguồn khác. Nhờ đó, doanh nghiệp hiểu rõ khách hàng của mình hơn qua chính hành vi thực tế của khách hàng.

Máy học (Machine Learning): Máy học là một bước tiến của công nghệ hiện đại và Dữ liệu lớn đóng góp phần lớn trong sự phát triển của máy học. Hiện máy móc này có thể tự học từ dữ liệu lịch sử thay vì con người phải thiết lập chương trình, lịch sử dữ liệu này được thu thập đầy đủ vào Dữ liệu lớn.

Thúc đẩy sự tiến bộ mới: Dữ liệu lớn cung cấp dữ liệu về thông tin sản phẩm, lịch sử phát triển của mọi lĩnh vực, ngành nghề. Qua đó, nhà quản lý có thể xác định được những điểm cần cải tiến, phát triển để công nghệ mới lại sự sáng tạo, mới mẻ, hữu ích cho các ngành công nghiệp.

* + 1. Lĩnh vực giáo dục đào tạo

Ngành giáo dục sử dụng Dữ liệu lớn để ước tính số lượng học sinh, tuyển sinh viên hàng năm, quản lý hồ sơ học sinh, sinh viên, truy xuất thông tin nhanh cần thiết khi gặp các vấn đề phát sinh. Ngoài ra, Big Data còn giúp ước tính được nhu cầu tuyển dụng cho các ngành nghề hàng năm để đề ra các phương án đào tạo nhằm đáp ứng các nguồn nhân lực cho xã hội.

* + 1. Cơ sở hạ tầng IT

Điểm đặc biệt của Big Data là lượng dữ liệu lớn và tăng cao qua nhiều năm, do đó cần có hệ thống lưu trữ lưu trữ và máy chủ được thiết kế đặc biệt cho Big Data. Ngoài ra còn có phần mềm thu thập, quản lý và phân tích dữ liệu. Doanh nghiệp nếu sử dụng Big Data lâu dài thì nên đầu tư cơ sở hạ tầng IT vàng, chất lượng để đảm bảo nguồn dữ liệu không bị mất đi hoặc không cập nhật. Phần lớn các cơ sở hạ tầng này sẽ tập trung vào một vị trí, bởi vì các công ty muốn tiếp tục tận dụng các khoản đầu tư vào trung tâm dữ liệu của mình. Nhưng ngày càng có nhiều tổ chức dựa trên các dịch vụ điện toán đám mây đám mây để xử lý nhiều yêu cầu dữ liệu lớn của họ.

Cải thiện kết nối giữa các thiết bị điện tử.

Chuyển các gói dữ liệu qua mạng được kết nối tiết kiệm thời gian và tiền bạc.

Tự động hóa các nhiệm vụ giúp cải thiện chất lượng dịch vụ của doanh nghiệp và giảm nhu cầu can thiệp của con người.

CHƯƠNG 2: TÌM HIỂU VỀ CASE STUDY

* 1. Tóm tắt các sự việc, sự kiện quan trọng:
* Vấn đề chính của big data ngày nay.
* Dữ liệu lớn cũng có những hạn chế.
* Nắm bắt xu thế của các công ty lớn và việc ứng dụng vào thực tế.
* Lợi ích to lớn từ việc sử dụng thực tế và có hiệu quả của big data.
* Những sai lầm khi chưa tìm hiểu và chuẩn bị tốt để áp dụng big data.
  1. Nhận diện và phân tích vấn đề
     1. Vấn đề chính của big data ngày nay

Các công ty ngày nay đang phải đối mặt với lượng dữ liệu khổng lồ từ mạng xã hội, công cụ tìm kiếm và cảm biến cũng như từ các nguồn truyền thống.

Thấy có rất nhiều tổ chức như Amazon, Youtube, Spotify, v.v. sử dụng dữ liệu lớn trong các hoạt động vận hành của mình, cũng cho thấy dữ liệu lớn ảnh hưởng như thế nào đến tổ chức và những lợi ích mà nó tạo ra.

Theo một ước lượng, mỗi ngày trên thế giới tạo ra khoảng 2,5 quintillion byte dữ liệu. Việc hiểu rõ về "big data" để cải thiện quyết định và hiệu suất kinh doanh đã trở thành một trong những cơ hội chính cho tổ chức có kích thước và hình dạng khác nhau, nhưng cũng đồng thời mang đến những thách thức lớn.

2.2.2 Dữ liệu lớn cũng có những hạn chế

Giải thích rằng việc sử dụng dữ liệu lớn phải có mục tiêu rõ ràng và nói rõ rằng không phải tất cả dữ liệu chỉ có thể đo lường được không có nghĩa là nó nên được đo lường.

Việc vội vã bắt đầu dự án dữ liệu lớn mà không thiết lập mục tiêu kinh doanh trước sẽ không mang lại lợi ích gì cho công ty. Vì họ không biết loại nào thông tin họ đang tìm kiếm, họ đang tích lũy hàng núi dữ liệu không rõ ràng mục tiêu. Điều này sẽ gây lãng phí thời gian và nguồn lực.

Có khả năng diễn giải sai lệch từ kết quả phân tích dữ liệu lớn. Ví dụ, việc kiểm tra dữ liệu lớn có thể cho thấy rằng từ năm 2006 đến năm 2011 Hoa Kỳ tỷ lệ giết người có mối tương quan chặt chẽ với thị phần của Internet Explorer, vì cả hai đều giảm mạnh, nhưng điều đó không nhất thiết có nghĩa là có bất kỳ ý nghĩa nào mối liên hệ giữa hai hiện tượng.

Các mô hình dự đoán dữ liệu lớn đôi khi không lý tưởng để thể hiện xu hướng trong tương lai. Khi hệ thống sử dụng kỹ thuật lập mô hình lỗi thời, các khuyến nghị từ phần mềm sẽ vô dụng vì chúng không thể dự đoán đúng và chính xác.

Các tập dữ liệu và mô hình dự báo dựa trên dữ liệu có thể bị nhầm lẫn bởi các giả định sai sót và không chú ý đến ngữ cảnh. Ví dụ: Google đã phát triển một mô hình để dự báo xu hướng cúm, nhưng nó dựa trên một giả định sai sót và thuật toán chỉ nhìn vào những con số chứ không nhìn vào bối cảnh của kết quả.

* + 1. Nắm bắt xu thế của các công ty lớn và việc ứng dụng vào thực tế

Các công ty lớn đã dựa vào lượng lớn thông tin khách hàng đã thu thập được để tại ra các đề xuất cá nhân vd như: Các công ty như Amazon, YouTube, và Spotify sử dụng dữ liệu lớn để tạo ra hàng triệu đề xuất cá nhân cho khách hàng. Các dịch vụ trực tuyến giúp người dùng phân tích được các giá sản phẩm giúp người tiêu dùng lựa chọn được giá sản phẩm tốt nhất có thể.

=> Big data cũng mang lại những lợi ích trong lĩnh vực thể thao giáo dục, khoa học, chăm sóc sức khỏe và lực lượng chức năng.

**Sự kiện thực tế cho thấy lợi ích của big data trong case study:**

Việc phân tích hàng tỷ điểm dữ liệu được thu thập về bệnh nhân của Dịch vụ Y tế Quốc gia Anh (NHS) đã tiết kiệm khoảng 581 triệu bảng Anh (784 triệu đô la Mỹ). Dữ liệu được lưu trữ trong một Oracle Exadata Database Machine, có thể nhanh chóng phân tích các lượng dữ liệu rất lớn. NHS đã sử dụng kết quả từ phân tích dữ liệu lớn để tạo ra bảng điều khiển nhận biết bệnh nhân đang dùng 10 hoặc hơn loại thuốc cùng một lúc và những bệnh nhân đang sử dụng quá nhiều loại kháng sinh. Tổng hợp một lượng lớn dữ liệu về các loại thuốc và liệu pháp được đưa ra cho bệnh nhân ung thư và liên kết thông tin đó với kết quả của bệnh nhân đã giúp NHS xác định các giao thức điều trị hiệu quả hơn

Thành phố New York sử dụng dữ liệu để giảm tỷ lệ tội phạm thông qua chương trình CompStat, một hệ thống báo cáo hàng tuần về tình hình tội phạm bằng cách sử dụng một cơ sở dữ liệu toàn thành phố bao gồm tất cả các tội phạm hoặc khiếu nại được báo cáo Dữ liệu CompStat có thể được hiển thị trên bản đồ cho thấy vị trí tội phạm và bắt giữ, các điểm nóng tội phạm và các thông tin khác để giúp các chỉ huy khu phố nhanh chóng xác định mô hình và xu hướng và triển khai nhân viên cảnh sát nơi chúng cần thiết nhất.

2.2.4 Những sai lầm khi chưa tìm hiểu và chuẩn bị tốt để áp dụng big data:

Cùng với nhiều tổ chức khác nhau cho thấy rằng việc sử dụng dữ liệu không chính xác cũng có thể tạo ra kết quả không chính xác, nhấn mạnh rằng khi sử dụng dữ liệu lớn, các hệ thống phải chú ý đến bối cảnh chứ không chỉ con số.

**Sự kiện thực tế cho thấy sai lầm của tổ chức khi dùng sai mục đích của big data trong case study:**

Công ty Điện năng Meridian Ltd., một công ty sản xuất và phân phối điện hoạt động tại New Zealand và Úc, đã chuyển sang không sử dụng hệ thống bảo trì thiết bị dự đoán đang bị xuống cấp. Phần mềm này được thiết kế để dự đoán nhu cầu bảo trì của tất cả các thiết bị lớn mà công ty sở hữu và vận hành, bao gồm máy phát điện, turbine gió, biến áp, cầu dao, và pin công nghiệp. Tuy nhiên, hệ thống sử dụng các kỹ thuật mô hình hóa lỗi thời và không thể thực sự dự đoán sự cố của thiết bị. Nó chạy các mô phỏng về các kịch bản khác nhau và dự đoán khi nào tài sản sẽ bị hỏng theo các kiểm tra mô phỏng. Các đề xuất của phần mềm không hữu ích vì chúng không dự đoán chính xác những thiết bị nào thực sự hỏng trong thế giới thực. Cuối cùng, Meridian đã thay thế hệ thống cũ bằng phần mềm Dự đoán Bảo trì và Chất lượng của IBM, dựa trên dự đoán từ dữ liệu thời gian thực hơn từ thiết bị.

Công ty đã tạo ra hầu hết các khách hàng tiềm năng và doanh thu cuối cùng từ các triển lãm và hội nghị thương mại đã không tận dụng được dữ liệu lớn để đo lường lưu lượng truy cập trang web liên quan đến số lượng đề cập trên Twitter.

Một số năm trước, Google phát triển một thuật toán đầu tiên bằng cách sử dụng dữ liệu thu thập từ tìm kiếm web để xác định chính xác bao nhiêu người mắc cúm và bệnh lan truyền như thế nào. Nó cố gắng tính toán số người mắc cúm ở Hoa Kỳ bằng cách liên kết vị trí của người dùng với các truy vấn tìm kiếm liên quan đến cúm trên Google. Google liên tục đánh giá quá cao tỷ lệ cúm, so với dữ liệu truyền thống được thu thập sau đó bởi Trung tâm Kiểm soát Bệnh tật Hoa Kỳ (CDC). Một số nhà khoa học đề xuất rằng Google đã bị "lừa dối" bởi sự chú ý rộng rãi của truyền thông về mùa cúm nặng nề ở Hoa Kỳ trong năm đó, được tăng cường thêm bởi sự chú ý của truyền thông xã hội. Mô hình được phát triển để dự báo xu hướng cúm dựa trên giả định sai lầm rằng sự xuất hiện của các truy vấn tìm kiếm liên quan đến cúm trên Google là một chỉ số chính xác của số người thực sự mắc cúm.

Để kết thúc, case study này không quên nhắc nhở người đọc rằng dữ liệu lớn đặt ra một số thách thức đối với bảo mật và quyền riêng tư thông tin. Các công ty hiện đang tích cực thu thập và khai thác các tập dữ liệu khổng lồ về con người để dự đoán hành vi của họ và thao túng họ để đạt được lợi nhuận tối đa.

* 1. Các trường hợp trong thực tế

1. LinkedIn

LinkedIn đã thành công trong việc sử dụng Big Data để cải thiện trải nghiệm người dùng và đạt được sự cạnh tranh lớn trên thị trường mạng xã hội. Cách họ tận dụng Dữ liệu lớn bao gồm:

Tăng tương tác hồ sơ: Sử dụng Dữ liệu lớn để theo dõi hành vi người dùng và phát triển tính năng "Public Profiles" để kết nối người dùng.

Tối ưu hóa tìm việc: Phân tích Dữ liệu lớn để hiểu rõ nhu cầu và mong muốn người dùng, tạo tính năng xem ai đã xem hồ sơ, nâng cao trải nghiệm tìm kiếm việc làm.

Tối ưu quảng cáo và tăng doanh thu:

Sử dụng thông tin chi tiết từ Dữ liệu lớn để tạo quảng cáo gợi ý, tăng tương tác và lượt xem trang web.

***Kết quả đạt được:***

LinkedIn thu hút 40 triệu thành viên mới trong nửa đầu năm 2015 và có doanh thu hơn 700 triệu đô la vào năm 2020, chủ yếu nhờ sự tận dụng thông tin người dùng thông qua Big Data.

1. Netflix – Dữ liệu lớn giúp tối ưu trải nghiệm người dùng và hiệu quả quảng cáo

Netflix đã thành công trong việc sử dụng Big Data để tối ưu hóa trải nghiệm người dùng và cải thiện khả năng giữ chân khách hàng. Dưới đây là cách Netflix đã áp dụng Dữ liệu lớn:

*Gia hạn loạt phim gốc:*

* Netflix sử dụng Dữ liệu lớn để phân tích hành vi xem phim của người dùng.
* Gia hạn 93% loạt phim gốc nổi tiếng dựa trên dữ liệu về thời gian xem, thiết bị sử dụng và thời lượng xem.

*Tăng khả năng giữ chân khách hàng trung thành:*

* Phân tích dữ liệu tìm kiếm và xem phim trước đây của khách hàng.
* Netflix sử dụng dữ liệu để đề xuất những bộ phim phù hợp và làm tăng khả năng giữ chân khách hàng.

*Tối ưu quảng cáo và tiếp cận đối tượng chính xác:*

* Sử dụng dữ liệu chi tiết từ Big Data để tạo trailer tùy chỉnh cho từng đối tượng.
* Netflix quảng bá chương trình "House of Cards" thông qua trailer được tối ưu hóa dựa trên thói quen xem phim của khán giả.

*Kết quả đạt được:*

* Sau chỉ 2 năm triển khai Dữ liệu lớn, Netflix thu về hơn 1 tỷ USD và giữ chân 80% khách hàng.
* Số lượng người đăng ký và sử dụng tăng nhanh chóng, đặc biệt trong thời gian giãn cách xã hội do đại dịch Covid-19.
* Netflix xây dựng nền tảng cá nhân hóa dựa trên sở thích của người xem. Phân tích và dự đoán giúp Netflix xác định khách hàng mục tiêu và tối ưu hóa chiến dịch tiếp thị.

1. Walmart – Dữ liệu lớn góp phần đẩy mạnh năng suất của chuỗi siêu thị toàn cầu

Walmart là nhà bán lẻ lớn nhất trên thế giới với hơn 2 triệu nhân viên, 20.000 cửa hàng đặt tại 28 quốc gia. Chuỗi siêu thị Walmart kinh doanh hàng triệu sản phẩm với hàng triệu giao dịch mỗi ngày, vấn đề đặt ra là làm thế nào để khách hàng có thể mua những gì họ cần đúng nơi, đúng thời điểm với giá cả hợp lý, nhất là trong bối cảnh thương mại điện tử đã và đang chiếm dần vị thế của hình thức bán hàng truyền thống.

**Tái định giá sản phẩm:**

Walmart sử dụng dữ liệu lớn để phân tích thói quen chi tiêu và lựa chọn sản phẩm của khách hàng.

* Đội Fast Big Data Team của Walmart nghiên cứu và triển khai sáng kiến dẫn đầu về dữ liệu lớn trên quy mô toàn doanh nghiệp.
* Thông tin về khách hàng và chi tiêu được phân tích để hiểu rõ nguyên nhân giảm doanh số bán hàng, giúp định giá lại sản phẩm nhanh chóng.

*Quản lý hàng tồn kho:*

* Walmart sử dụng dữ liệu lớn để theo dõi bán hàng theo từng khu vực địa lý, tránh tình trạng hàng tồn kho lớn.
* Các cảnh báo thời gian thực được kích hoạt khi nhận thấy địa điểm có doanh số thấp, giúp giải quyết tình trạng hàng tồn kho và đáp ứng nhanh chóng nhu cầu tăng cao trong các dịp lễ hội.

*Lắng nghe nhu cầu khách hàng:*

* Dự án "Walmart’s Social Genome" ghi lại cuộc trò chuyện trên phương tiện truyền thông xã hội để dự đoán sản phẩm mà người dùng có thể muốn mua.
* Dịch vụ Shopycat sử dụng dữ liệu từ Social Media để dự đoán thói quen mua sắm của khách hàng dựa trên ảnh hưởng từ bạn bè.

*Kết quả đạt được:*

* Cải thiện đáng kể về hiệu quả kinh tế và tăng cường vị thế cạnh tranh của Walmart. Điều này thể hiện qua sự vượt trội trong doanh thu.
* Walmart giữ vị trí hàng đầu trong ngành bán lẻ thế giới, vượt qua các đối thủ lớn như Carrefour, và duy trì sự dẫn đầu trước các tập đoàn khác như Amazon và Alibaba.
  1. Một số thách thức cho Big data

Truyền dữ liệu: việc truyền dữ liệu lớn thường phải gánh chịu chi phí cao, đây là “nút cổ chai” của việc tính toán Big data. Tuy nhiên, truyền dữ liệu là không thể tránh khỏi trong các ứng dụng Big data. Nâng cao hiệu quả truyền dữ liệu lớn là một yếu tố quan trọng để nâng cao tính toán Big data.

Tốc độ xử lý trong các yêu cầu thời gian thực: khi dữ liệu số lượng dữ liệu tăng nhanh chóng, gây ra một thách thức rất lớn đối với các ứng dụng thời gian thực, thì việc tìm ra các phương pháp hiệu quả trong suốt luồng dữ liệu là cần thiết để đáp ứng yêu cầu về thời gian thực.

Nền tảng Big data: mặc dù Hadoop đã trở thành một trụ cột trong nền tảng phân tích Big data nhưng nó vẫn còn trong giai đoạn phát triển, so với cơ sở dữ liệu quan hệ. Đầu tiên, Hadoop phải tích hợp với thời gian thực cho việc thu thập và truyền Big data, và cung cấp xử lý nhanh hơn dựa trên các mô hình xử lý hàng loạt. Thứ hai, Hadoop nên cung cấp một giao diện lập trình ngắn gọn, và ẩn những tiến trình xử lý phức tạp bên dưới. Thứ ba, trong những hệ thống Hadoop lớn, số lượng máy chủ lên hàng ngàn, thậm chí hàng trăm ngàn, nghĩa là năng lượng tiêu thụ đáng kể. Vì vậy, Hadoop nên có cơ chế sử dụng năng lượng hiệu quả.

Bảo mật dữ liệu và quyền riêng tư: là vấn đề rất quan trọng. Một số ví dụ trong thực tế cho thấy, không chỉ thông tin cá nhân người tiêu dùng, thông tin mật của các tổ chức mà ngay cả các bí mật an ninh quốc gia cũng có thể bị xâm phạm. Do vậy, giải quyết các vấn đề an ninh dữ liệu bằng các công cụ kỹ thuật và các chính sách trở nên vô cùng cấp bách. Các nền tảng Big data nên cân bằng tốt giữa việc truy cập dữ liệu và xử lý dữ liệu.

* 1. Các giải pháp cho big data

***Hệ thống lưu trữ phân tán (Distributed Storage):*** Sử dụng các hệ thống như Hadoop Distributed File System (HDFS) hoặc Amazon S3 để lưu trữ dữ liệu trên nhiều nút khác nhau.

***Công nghệ xử lý phân tán (Distributed Processing):*** Sử dụng Hadoop, Apache Spark hoặc Flink để xử lý dữ liệu trên nhiều nút đồng thời, giúp tăng hiệu suất và tốc độ xử lý.

***Cơ sở dữ liệu NoSQL:*** Sử dụng cơ sở dữ liệu không quan hệ như MongoDB, Cassandra, hoặc Couchbase để lưu trữ và truy xuất dữ liệu một cách hiệu quả.

***Công nghệ in-memory:*** Sử dụng công nghệ lưu trữ trong bộ nhớ như Apache Ignite hoặc Redis để tăng tốc độ truy xuất dữ liệu.

Công cụ trực quan hóa và phân tích:

Sử dụng công cụ như Tableau, Power BI, hoặc Apache Superset để phân tích và trực quan hóa dữ liệu một cách dễ dàng.

***Tích hợp Machine Learning***: Kết hợp big data với machine learning để phân tích dữ liệu và đưa ra dự đoán thông minh. Apache Mahout và TensorFlow là một số công cụ phổ biến.

***Bảo mật dữ liệu:*** Áp dụng các biện pháp bảo mật mạnh mẽ để đảm bảo an toàn cho dữ liệu lớn, bao gồm mã hóa dữ liệu, quản lý danh tính và giám sát an ninh.

***Tối ưu hóa hiệu suất:*** Tối ưu hóa cấu trúc dữ liệu và truy vấn để đảm bảo hiệu suất tốt nhất trong quá trình xử lý và truy cập dữ liệu.

***Tự động hóa quy trình:***Sử dụng công cụ tự động hóa như Apache Airflow để quản lý và lên lịch cho các quy trình xử lý dữ liệu.

***Đám mây (Cloud Computing):*** Sử dụng dịch vụ đám mây như Amazon Web Services (AWS), Microsoft Azure, hoặc Google Cloud Platform để mở rộng và quản lý cơ sở hạ tầng dễ dàng.

* 1. Lựa chọn giải pháp và cách thực hiện

Lựa chọn một giải pháp dữ liệu lớn không phải là công việc dễ dàng. Khi các công ty mọi kích cỡ đang cố gắng khai thác nhiều hơn từ cơ sở dữ liệu hiện tại của họ, các nhà cung cấp Dữ liệu lớn đang đổ vào thị trường để cung cấp một loạt các giải pháp Dữ liệu lớn, bao gồm từ công nghệ cơ sở dữ liệu đến các công cụ trực quan hóa. Với sự lựa chọn đa dạng như vậy, người mua phải định rõ mục tiêu của mình để tìm ra các công cụ phù hợp với mục tiêu của họ. Các bước lựa chọn đúng giải pháp cho big data:

**Bước 1: Định rõ vấn đề - đúng cách**

Bước đầu tiên khi chọn một giải pháp cho big data là xác định vấn đề kinh doanh bạn muốn giải quyết, nhưng khi làm điều này, quan trọng là phải đặt ra vấn đề đó một cách sao cho dự án có thể được tài trợ.

**Bước 2: Hỏi xem giải quyết vấn đề này sẽ mang lại gì**

Đặt câu hỏi về những gì sẽ được đạt được bằng cách giải quyết vấn đề cụ thể này. Xác định giá trị kinh doanh, giảm chi phí, tăng hiệu suất, v.v.

**Bước 3: Xếp hạng các lợi ích dự kiến**

Xếp hạng các lợi ích dự kiến (quyết định tốt hơn, giảm chi phí, tăng hiệu suất, v.v.). Điều này sẽ giúp hạn chế tìm kiếm của bạn.

**Bước 4: Tự đánh giá**

Trước khi mua một giải pháp big data mới, quan trọng là đánh giá chính xác tình hình hiện tại của bạn. Xem xét kiến trúc phần mềm bạn đang sử dụng và xác định xem nó có tương thích với một số giải pháp hay không.

**Bước 5: Xác định cách đo tiến triển**

Điều này quan trọng để người dùng của bạn, và cũng giúp bạn theo dõi tiến trình và điều chỉnh hướng đi nếu cần thiết. Hỏi xem giải pháp big data bạn chọn sẽ giúp bạn đo lường tiến triển như thế nào và sửa lỗi nếu bạn không tiến triển.

**Bước 6: Xem xét người sử dụng cuối cùng**

Xác định ai sẽ thực sự sử dụng công cụ. Nếu bạn không có chuyên gia big data trong nhóm, bạn có thể giới hạn mình với big data dưới dạng dịch vụ hoặc tìm các phần mở rộng phân tích cho phần mềm bạn đã sử dụng.

**Bước 7: Xác định cách chuyển từ dự án sang quy trình**

Nhiều dự án big data ban đầu sẽ là thăm dò hơn là gì khác. Tuy nhiên, khi bạn thăm dò dữ liệu của mình để tìm các lĩnh vực có thể mang lại giá trị, hãy theo dõi cách bạn quyết định nào để nghiên cứu dữ liệu, làm thế nào bạn trích xuất giá trị, làm thế nào dữ liệu đưa ra các hành động và làm thế nào bạn đo lường những gì bạn đã làm.

* 1. Ví dụ thực tế

Quản lý và Phân tích dữ liệu lớn từ hệ thống phân phối bán lẻ

“Một công ty bán lẻ lớn đang phải đối mặt với một lượng lớn dữ liệu từ hệ thống phân phối của họ. Hệ thống này phát sinh hàng triệu giao dịch mua bán hàng ngày từ hàng trăm cửa hàng. Dữ liệu bao gồm thông tin về doanh số bán hàng, tồn kho, thông tin khách hàng, và các dữ liệu khác từ các điểm bán hàng khác nhau.”

**Bước 1: Định rõ vấn đề - đúng cách**

Tính đến thời điểm hiện tại, công ty này sử dụng hệ thống lưu trữ truyền thống, nhưng họ gặp phải các thách thức lớn:

* Khối lượng lớn dữ liệu: Số lượng dữ liệu đang tăng nhanh chóng do mở rộng kinh doanh và số lượng cửa hàng mới.
* Hiệu suất chậm: Hệ thống hiện tại không đủ mạnh mẽ để xử lý và phân tích một lượng lớn dữ liệu này một cách hiệu quả trong thời gian ngắn.
* Không linh hoạt: Hệ thống truyền thống không đáp ứng linh hoạt đối với các loại dữ liệu đa dạng và không đồng nhất từ các nguồn khác nhau.

**Bước 2: Hỏi xem giải quyết vấn đề này sẽ mang lại gì**

* Tăng hiệu suất
* Giảm chi phí
* Tính linh hoạt
* Tối ưu hóa quy trình quản lý

**Bước 3: Xếp hạng các lợi ích dự kiến**

* Tối ưu hóa quy trình quản lý
* Tính linh hoạt
* Giảm chi phí
* Tăng hiệu suất

**Bước 4: Tự đánh giá**

Sau khi xem xét các thách thức trên, công ty quyết định triển khai Hadoop như một giải pháp cho vấn đề của họ:

* Xử lý phân tán: Hadoop cho phép xử lý phân tán của dữ liệu trên nhiều nút, giúp giải quyết vấn đề về khối lượng lớn dữ liệu.
* Khả năng mở rộng dễ dàng: Với Hadoop, công ty có thể dễ dàng mở rộng hệ thống của mình khi cần thiết để đối mặt với sự tăng trưởng.
* Xử lý dữ liệu đa dạng: Hadoop có thể xử lý dữ liệu từ nhiều nguồn khác nhau và định dạng khác nhau, bao gồm cả dữ liệu cấu trúc và không cấu trúc.

**Bước 5: Xác định cách đo tiến triển**

Công ty cần xác định cách họ sẽ đo lường tiến triển của dự án triển khai giải pháp Hadoop. Điều này bao gồm việc thiết lập các chỉ số hiệu suất và tiêu chí đánh giá để theo dõi hiệu quả của giải pháp. Công ty có thể sử dụng các công cụ phân tích dữ liệu để đo lường các chỉ số quan trọng và đảm bảo rằng giải pháp đang đáp ứng được các mục tiêu đã đặt ra.

**Bước 6: Xem xét người sử dụng cuối cùng**

Điều này bao gồm việc xác định các bộ phận trong tổ chức sẽ tận dụng giải pháp và đảm bảo rằng họ đủ kỹ năng và kiến thức để sử dụng nó hiệu quả. Nếu có sự thiếu hụt về kỹ năng, công ty có thể cân nhắc cung cấp đào tạo và hỗ trợ để đảm bảo người sử dụng có thể tận dụng đầy đủ tiềm năng của giải pháp.

**Bước 7: Xác định cách chuyển từ dự án sang quy trình**

Trong bước này, công ty cần xác định cách họ sẽ chuyển từ mô hình dự án ban đầu sang một quy trình hoạt động liên tục. Điều này bao gồm việc xem xét các bước và quy trình đã thử nghiệm trong dự án ban đầu và định rõ cách chúng có thể được tích hợp vào quy trình tự động để theo dõi và tối ưu hóa chiến lược quản lý dữ liệu lớn liên tục. Việc này giúp công ty duy trì sự linh hoạt và nhanh chóng thích ứng với thay đổi trong môi trường kinh doanh và yêu cầu dữ liệu.

* 1. Bài học rút ra

Big data mang lại cơ hội lớn để cải thiện quyết định và hiệu suất kinh doanh, nhưng đồng thời cũng đối mặt với những thách thức lớn, bao gồm việc xử lý lượng lớn dữ liệu và đảm bảo tính đáng tin cậy của thông tin thu thập được.

Việc phân tích big data có thể cung cấp thông tin hữu ích, nhưng đòi hỏi sự chú ý đến mục tiêu kinh doanh và các chỉ số hiệu suất chính để đảm bảo rằng thông tin được thu thập có ý nghĩa và hỗ trợ quyết định thông minh.

Việc phân tích big data có thể cung cấp thông tin hữu ích, nhưng đòi hỏi sự chú ý đến mục tiêu kinh doanh và các chỉ số hiệu suất chính để đảm bảo rằng thông tin được thu thập có ý nghĩa và hỗ trợ quyết định thông minh.

Sự cảnh báo về việc nhiều công ty bắt đầu dự án big data mà không xác định mục tiêu kinh doanh và các chỉ số hiệu suất. Việc thu thập dữ liệu không có mục tiêu cụ thể có thể dẫn đến việc lãng phí tài nguyên và không đạt được kết quả mong muốn.

Sự tích lũy lượng lớn dữ liệu cá nhân đặt ra những thách thức đối với quyền riêng tư và bảo mật thông tin. Cần có sự quan tâm đặc biệt đối với cách thông tin cá nhân được thu thập và sử dụng.

TÀI LIỆU THAM KHẢO

[1] "Building the Internet of Things: Implement New Business Models, Disrupt Competitors, Transform Your Industry" by Maciej Kranz, 2016.

[2] "Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are" by Seth Stephens-Davidowitz, 2017.

[3] "A Case Study in Big Data Analytics" by Author(s): R.O. Sinnott , H. Duan , Y. Sun, 2016.

[4] "How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results" by Bernard Marr, 2016.