CHƯƠNG 3

Quản lý dữ liệu, kho dữ liệu và quản trị dữ liệu

|  |  |
| --- | --- |
| TỔNG KẾT CHƯƠNG | MỤC TIÊU HỌC TẬP |
| Trường hợp 3.1:  Thang máy Tiết kiệm 1,5 triệu đô la và 900 cây xanh do số hóa tài liệu |  |
| 3.1 Quản lý dữ liệu | 3.1 **Mô tả** các nguyên tắc cơ bản của quản lý dữ liệu và cách các hệ thống quản lý cơ sở dữ liệu, bao gồm cả công nghệ chuỗi khối, giúp các công ty cải thiện hiệu suất. |
| 3.2 Kho dữ liệu và chợ dữ liệu | 3.2 **Xác định** sự khác biệt giữa cơ sở dữ liệu, kho dữ liệu và data mart và lý do tại sao một công ty sẽ chuyển từ cơ sở dữ liệu sang kho dữ liệu hoặc data mart: |
| 3.3 Quản trị dữ liệu và dữ liệu chủ | 3.3 **Mô tả** tầm quan trọng của quản trị dữ liệu trong việc cung cấp dữ liệu đáng tin cậy có sẵn khi nào và ở đâu cần thiết và tại sao quản lý dữ liệu chủ (MDM) là một sáng kiến quản trị dữ liệu quan trọng. |
| 3.4 Quản lý thông tin | **3.4 Giải thích** khái niệm quản lý thông tin và tại sao nó  "quan trọng đối với một tổ chức để tuân theo các thông lệ quản lý thông tin tốt. |
| 3.5 Tài liệu điện tử , Bản ghi và Nội dung quản lý | **3.5 Định nghĩa** các thuật ngữ : tài liệu điện tử, bản ghi điện tử và nội dung điện tử và cách các công nghệ quản lý chúng đang được sử dụng để làm cho các tổ chức hoạt động hiệu quả hơn. |
|
| Trường hợp 3.2 Dữ liệu bẩn  Gây nguy hiểm cho các nỗ lực gây quỹ của trường đại học |  |

Trường hợp 3.1 Trường hợp mở đầu

Thang máy ThyssenKrupp tiết kiệm 1,5 triệu đô la và 900 cây bằng cách số hóa tài liệu

Công ty

Thang máy ThyssenKrupp (<https://www.thyssenkruppelevator.com> là một phần của ThyssenKrupp AG, một tập đoàn công nghiệp đa dạng toàn cầu được thành lập vào năm 1811. Với 155.000 nhân viên tại gần 80 quốc gia, ThyssenKrupp AG phát triển các sản phẩm chất lượng cao, các quy trình và dịch vụ công nghiệp thông minh nhằm thúc đẩy Các kỹ sư của ThyssenKrupp Elevator, cung cấp và bảo trì các thang máy, thang cuốn, lối đi bộ di động và cầu lên máy bay chất lượng cao để mọi người di chuyển an toàn, thoải mái và hiệu quả. các giải pháp tiên tiến để định hình lại ngành thang máy và biến các thành phố thành nơi tốt nhất để sinh sống. Họ thực hiện điều này bằng cách cung cấp năng suất và hiệu quả vận chuyển cao hơn, sử dụng hiệu quả hơn diện tích sàn tòa nhà, giảm mức tiêu thụ năng lượng và tối đa hóa tính khả dụng của các giải pháp vận chuyển hành khách hiện có.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Vấn đề  Xây dựng thang máy là một quá trình sử dụng nhiều dữ liệu. Khoảng 700 tài liệu được tạo, xử lý và lưu trữ trong quá trình lắp đặt một thang máy. Nhân 700 với 2.100 thang máy mà ThyssenKrupp sản xuất hàng năm, thêm thang cuốn sản xuất, lối đi bộ di chuyển, cầu lên máy bay cho hành khách, thang máy và cầu thang bộ và họ phải xử lý gần MỘT TRIỆU tài liệu mỗi năm!  Tại ThyssenKrupp, quy trình thu thập dữ liệu chủ yếu dựa trên giấy tờ. Điều này khiến nhân viên rất khó tìm thấy thông tin quan trọng trong các tài liệu liên tục được chuyển qua lại giữa hơn 100 chi nhánh và địa điểm dịch vụ trong một "thư mục màu đỏ" vật lý. Để thêm vào sự nhầm lẫn, khi dự án tiến triển, các thành viên của nhóm dự án đã thêm các tài liệu giấy của riêng họ theo cách thủ công. Điều này gây ra lãng phí thời gian, thiếu khả năng hiển thị công việc và sự chậm trễ trong việc giao sản phẩm.  Theo Mike Fuller, nhà phát triển/nhà phân tích của ThyssenKrupp, "chúng tôi phải đối mặt với một số thách thức đang diễn ra, vì chi phí và thời gian để quản lý các quy trình nặng về tài liệu của chúng tôi liên tục tăng lên."    Các vấn đề khác xảy ra khi các tài liệu được đính kèm vào email trong các hệ thống nội bộ không được quản lý tốt. Ngoài ra, ThyssenKrupp đã phải chi 25.000 đô la mỗi tháng để vận chuyển các bản cứng của tài liệu giữa hơn 100 chi nhánh của mình trên khắp Hoa Kỳ và Canada  Việc sử dụng các thư mục mạng cũng góp phần gây ra sự cố. Cấu trúc tệp được lưu trữ trên nhiều máy chủ thường có hơn chín cấp độ, tên đường dẫn thư mục thường vượt quá 255 ký tự tối đa của hệ thống và tài liệu được lưu trữ không thống nhất và sử dụng nhiều quy ước đặt tên khác nhau. Phương pháp lưu trữ dữ liệu này dẫn đến các tệp bị mất, \_ các tài liệu trùng lặp và các vấn đề về kiểm soát phiên bản, đồng thời khiến ThyssenKrupp tốn nhiều thời gian và tiền bạc hơn.    Damon Rodkey, Giám đốc CNTT tại ThyssenKrupp, đã tóm tắt các vấn đề về dữ liệu của ThyssenKrupp bằng cách nói rằng anh ấy thường cảm thấy rằng việc tìm kiếm dầu ở sân sau của mình sẽ dễ dàng hơn là cố gắng tìm một tài liệu cụ thể trong các thư mục mạng được chia sẻ của ThyssenKrupp.  Giải pháp  ThyssenKrupp đã chỉ định một nhóm kỹ sư, điều phối viên, nhà quản lý và người soạn thảo đa chức năng để xem xét quy trình quản lý tài liệu và khám phá các giải pháp khác nhau. Họ đã lập bản đồ các quy trình kinh doanh hiện có và tạo ra tầm nhìn cho một hệ thống lý tưởng có thể nhanh chóng tìm và theo dõi tài liệu.  Mục tiêu của nhóm là gấp đôi—số hóa các hệ thống dựa trên giấy tờ của ThyssenKrupp và tự động hóa quá trình khám phá thông tin và tài liệu. Để đạt được điều này, họ cần một ứng dụng phần mềm tùy chỉnh được thiết kế cho các công ty sản xuất. Họ đã tìm thấy câu trả lời của mình trong M-Files (<https://www.m-files.com/en> ), một nền tảng quản lý tài liệu và nội dung thông minh. Nền tảng M-files được thiết kế để cải thiện việc nhóm và quản lý nội dung cũng như thông tin phi cấu trúc, sử dụng siêu dữ liệu để gắn thẻ các thuộc tính chính trong nội dung của tài liệu, chẳng hạn như vị trí, số khách hàng, loại tài liệu, ngày tháng, v.v. . Điều này cung cấp cho người dùng chế độ xem 360 độ trên tất cả các hệ thống để hiển thị' “cái gì” của một tài liệu chứ không chỉ “ở đâu” nó được lưu trữ.    M-Files cũng kích hoạt các quy trình công việc, chẳng hạn như quy trình phê duyệt và gửi bản vẽ của ThyssenKrupp, để đảm bảo các bản vẽ được xem xét đúng cách trước khi phê duyệt. Các tính năng khác của M-Files mà ThyssenKrupp thấy hữu ích trong hoạt động kinh doanh toàn cầu của mình là triển khai trên đám mây hoặc tại chỗ và cập nhật bảo mật theo thời gian thực để đảm bảo dữ liệu của ThyssenKrupp luôn cập nhật và an toàn, các chỉ số cải thiện Hiệu suất đảm bảo rằng ThyssenKmpp có thể theo dõi các cải tiến liên tục về tốc độ và độ chính xác trong quản lý dữ liệu,  Mike Fuller bày tỏ sự tán thành với giải pháp của nhóm bằng cách nói: "sự kết hợp giữa khả năng cấu hình, khả năng mở rộng, tính dễ sử dụng và chi phí triển khai thấp khiến M-Files rõ ràng phù hợp với giai đoạn chuyển đổi kỹ thuật số của chúng tôi” .    Kết quả  Nhờ M-Files, ThyssenKrupp hiện đang quản lý gần 100 triệu tài liệu của mình, trên 500 người dùng trong bảy phòng ban theo cách hiệu quả và hiệu quả hơn nhiều. Nicole Harris, Điều phối viên sửa chữa tại ThyssenKrupp nhận xét rằng: "Giờ đây, chúng tôi sắp xếp các tài liệu được gắn thẻ theo một số tên, mã số công việc và chi nhánh. Điều này giúp việc tìm kiếm tài liệu trở nên đơn giản hơn rất nhiều và quan trọng nhất là nhanh hơn,"  Quy trình làm việc mới đã được giới thiệu. Thư mục vật lý màu đỏ lỗi thời đã được thay thế bằng các quy trình công việc tự động hiệu quả cao mới. ThyssenKrupp cũng đã triển khai quy trình làm việc tự động để tạo và đặt tên cho các bộ chứa tài liệu nhiều tệp tiêu chuẩn cho các bản vẽ và một bộ chứa khác để bắt đầu quá trình | |  |
| tự động chỉ định bản vẽ dựa trên yêu cầu. Bằng cách này, các sản phẩm và lịch biểu có thể được theo dõi dễ dàng và nhanh chóng hơn trong hệ thống. Việc truyền tài liệu đến các nguồn bên trong và bên ngoài cũng được tự động hóa trong M-Files. Những quy trình công việc mới hiệu quả này đã loại bỏ các chi phí giấy tờ không cần thiết, loại bỏ các tài liệu điện tử và giấy bị thất lạc, loại bỏ hơn 13 lần chuyển giao tài liệu vật lý giữa các nhân viên và giữa nhiều phòng ban, đồng thời tăng tính minh bạch của dự án.  Sự chậm trễ cũng đã được giảm đáng kể thông qua việc bổ sung trình theo dõi quy trình làm việc dựa trên phân tích ghi lại các tác vụ và quy trình trong quá trình xây dựng thang máy. Trình theo dõi thông báo cho người quản lý nhóm, người quản lý dự án và ban quản lý về bất kỳ thời hạn nào bị bỏ lỡ để họ có thể hành động theo.    Cuối cùng, để thay thế các email khó quản lý và theo dõi, ThyssenKrupp hiện theo dõi thông tin liên lạc giữa các thành viên trong nhóm liên quan đến các câu hỏi chung của dự án. Các thành viên trong nhóm có thể tạo câu hỏi trong cơ sở dữ liệu và đồng bộ hóa chúng với quy trình làm việc của M-Files để theo dõi và giải quyết các truy vấn dễ dàng hơn và nhanh hơn. Điều này cho phép ThyssenKrupp liên tục cải thiện các quy trình của mình bằng cách loại bỏ các câu hỏi thường gặp (như rào cản) cho quy trình công việc hiệu quả hơn.    Cách tiếp cận sáng tạo của họ để hợp lý hóa quy trình công việc và làm phong phú chúng bằng siêu dữ liệu đã khiến ThyssenKrupp trở thành một ví dụ điển hình về quản lý dữ liệu hiệu quả thông qua chuyển đổi kỹ thuật số.    câu hỏi  1. ThyssenKrupp phải xử lý bao nhiêu tài liệu để xây dựng một thang máy?  2 Việc sử dụng M-Files đã mang lại lợi ích như thế nào cho ThyssenKrupp?  3.ThyssenKrupp phải đối mặt với hậu quả gì vì sự chậm trễ của dự án?    4. Thang máy ThyssenKrupp sử dụng M-Files cho những nhiệm vụ nào?    Nguồn: Tổng hợp từ IQ BG, Inc. ( 2018) , Bitpipe ( 2019) và ThyssenKruppelevator.dom. | |
| BẠN CÓ BIẾT KHÔNG?  Quản lý cơ sở dữ liệu và quản trị dữ liệu chưa bao giờ quan trọng hơn thế. Với sự ra đời của phương tiện truyền thông xã hội, Web, dữ liệu lớn và Đám mây, dữ liệu của tổ chức nằm rải rác ở mọi nơi—tại chỗ và ngoài cơ sở. Nếu không có khả năng hiển thị và kiểm soát dữ liệu của họ một cách thích hợp, các công ty có nguy cơ tổn thất lớn về doanh thu và năng suất và thậm chí có thể phải đối mặt với viễn cảnh phá sản! | |

## Giới thiệu

Thu thập và duy trì dữ liệu đã sử dụng là một khía cạnh quan trọng của bất kỳ doanh nghiệp nào. Với sự ra đời của máy tính, lượng dữ liệu được các tổ chức thu thập và lưu trữ đã tăng theo cấp số nhân trong 20 năm qua cùng với các công nghệ mới thú vị để hỗ trợ nó. Quản lý lượng dữ liệu khổng lồ này là một thách thức quan trọng đối với các tổ chức ngày nay. Biết cách thức và địa điểm tìm dữ liệu, lưu trữ dữ liệu một cách hiệu quả, phân tích dữ liệu theo những cách mới để tăng lợi thế cạnh tranh của tổ chức và cho phép đúng người truy cập dữ liệu vào đúng thời điểm là tất cả các thành phần cơ bản của việc quản lý số lượng dữ liệu ngày càng tăng của công ty. Độ tin cậy của các quyết định kinh doanh phụ thuộc vào việc truy cập dữ liệu chất lượng cao, kịp thời, dễ truy cập và chất lượng của dữ liệu phụ thuộc vào độ tin cậy và hiệu quả của phương pháp quản lý dữ liệu của tổ chức.

Quản lý dữ liệu là một kỹ thuật giúp các công ty cải thiện năng suất bằng cách đảm bảo rằng mọi người có thể tìm thấy dữ liệu họ cần mà không cần tiến hành tìm kiếm lâu và khó khăn. Mục tiêu của quản lý dữ liệu là cung cấp cơ sở hạ tầng và công cụ để chuyển đổi dữ liệu thô của công ty thành thông tin và kiến thức chất lượng cao có thể sử dụng được. Giống như việc bạn cần học cách xác định, kiểm soát, bảo vệ, phân tích và đầu tư tài sản tài chính, bạn cũng cần học cách quản lý tài sản dữ liệu của công ty.

Cơ sở dữ liệu, quản lý dữ liệu và các công nghệ cơ sở dữ liệu cung cấp các công cụ mà doanh nghiệp cần trong hệ sinh thái công nghệ cao mới được đặc trưng bởi sự phụ thuộc lẫn nhau ngày càng tăng giữa doanh nghiệp và CNTT, hợp tác bên trong và bên ngoài doanh nghiệp và các hành vi dựa trên sự tin tưởng.

**Hệ sinh thái Biz-tech** đòi hỏi doanh nghiệp coi CNTT như một đối tác bình đẳng và CNTT *sẽ đạt được thành công.*

Trong chương này, bạn sẽ học các thuật ngữ về cơ sở dữ liệu và các công nghệ hỗ trợ các cơ sở dữ liệu lớn và nhỏ. Bạn cũng sẽ tìm hiểu cách thức quản lý thông tin và kiến thức để tối đa hóa lợi ích của các công nghệ cơ sở dữ liệu hiện có và mới nổi cũng như cách thức quản lý chức năng cơ sở dữ liệu.

3.1 Quản lý dữ liệu

**L03.1** Mô tả các nguyên tắc cơ bản của quản lý dữ liệu và cách các hệ thống quản lý cơ sở dữ liệu, bao gồm cả công nghệ chuỗi khối, giúp các công ty cải thiện hiệu suất.

**Quản lý dữ liệu** giám sát vòng đời từ đầu đến cuối của dữ liệu từ khi tạo và lưu trữ ban đầu cho đến khi dữ liệu trở nên lỗi thời và bị xóa. Do khối lượng dữ liệu đáng kinh ngạc mà một tổ chức điển hình tạo ra, việc quản lý dữ liệu hiệu quả là rất quan trọng để ngăn chi phí lưu trữ vượt khỏi tầm kiểm soát và kiểm soát tốc độ tăng trưởng dữ liệu đồng thời hỗ trợ hiệu suất cao hơn.

**Quản lý dữ liệu** là quản lý luồng dữ liệu từ khi tạo và lưu trữ ban đầu cho đến khi dữ liệu trở nên lỗi thời và bị xóa.

Các mục tiêu của quản lý dữ liệu hiệu quả bao gồm:

1. Giảm thiểu rủi ro và chi phí tuân thủ các quy định,
2. Đảm bảo các yêu cầu pháp lý được đáp ứng.
3. Bảo vệ an toàn dữ liệu.

4. Duy trì tính chính xác và tính khả dụng của dữ liệu.

1. Chứng nhận tính nhất quán trong dữ liệu đến từ hoặc đi đến nhiều địa điểm.
2. Đảm bảo rằng dữ liệu tuân thủ các phương pháp hay nhất của tổ chức đối với truy suất, lưu trữ, sao lưu và xử lý.

* Thông thường, dữ liệu mới hơn và dữ liệu được truy cập thường xuyên hơn được lưu trữ trên phương tiện lưu trữ nhanh hơn nhưng đắt tiền hơn trong khi dữ liệu ít quan trọng hơn được lưu trữ trên phương tiện rẻ hơn, chậm hơn.
* **Lợi ích** của việc quản lý dữ liệu bao gồm;
* tuân thủ tốt hơn
* bảo mật cao hơn
* trách nhiệm pháp lý ít hơn
* cải thiện chiến lược bán hàng và tiếp thị
* phân loại sản phẩm tốt hơn
* cải thiện quản trị dữ liệu để giảm thiểu rủi ro.

### Các công nghệ cơ sở dữ liệu



Các công nghệ quản lý dữ liệu sau đây là nền tảng cho việc quản lý dữ liệu nhằm giúp người dùng luôn được cập nhật thông tin và hỗ trợ các nhu cầu kinh doanh đa dạng của tổ chức:

* **Cơ sở dữ liệu** Tập hợp các bộ dữ liệu hoặc bản ghi được lưu trữ một cách có hệ thống.
* **Hệ quản trị cơ sở dữ liệu** ( DBMS ) tích hợp với các hệ thống thu thập dữ liệu như TPS và các ứng dụng kinh doanh để lưu trữ dữ liệu một cách có tổ chức và cung cấp phương tiện để truy cập và quản lý dữ liệu.
* **Kho dữ liệu** Tập dữ liệu lớn tích hợp dữ liệu từ nhiều cơ sở dữ liệu và silo dữ liệu trong toàn tổ chức, đồng thời sắp xếp chúng để phân tích phức hợp, khám phá tri thức và để hỗ trợ ra quyết định. Ví dụ: dữ liệu được trích xuất từ cơ sở dữ liệu, được xử lý để chuẩn hóa định dạng của chúng , sau đó được tải vào kho dữ liệu vào những thời điểm cụ thể, chẳng hạn như hàng tuần. Như vậy, dữ liệu trong kho dữ liệu là không thay đổi và sẵn sàng để phân tích.
* **Data mart** là một dạng Kho dữ liệu quy mô nhỏ hỗ trợ cho một chức năng hoặc một bộ phận Các doanh nghiệp không đủ khả năng đầu tư vào kho dữ liệu có thể bắt đầu với một hoặc nhiều data mart.

Mỗi công nghệ cơ sở dữ liệu này sẽ được thảo luận trong các phần sau.

### **cơ sở dữ liệu**

Trước khi cơ sở dữ liệu ra đời, các tệp không liên quan được sử dụng để lưu trữ, truy xuất và tìm kiếm dữ liệu theo cách thủ công. Đó là một quá trình tẻ nhạt ảnh hưởng xấu đến hiệu suất, tăng trưởng và tính bền vững của tổ chức. Vào thời điểm mà các tổ chức đang phải vật lộn với việc duy trì tính toàn vẹn của dữ liệu, giảm sự không nhất quán của dữ liệu và đảm bảo an toàn dữ liệu, thì cơ sở dữ liệu là một giải pháp đáng hoan nghênh và là một bước tiến thú vị trong thời đại công nghệ kỹ thuật số. Cơ sở dữ liệu lưu trữ dữ liệu được tạo bởi các ứng dụng kinh doanh, cảm biến, hoạt động và hệ thống xử lý giao dịch (TPS). Trong cơ sở dữ liệu, mọi câu lệnh được thực hiện như một giao dịch . Giao dịch truy cập dữ liệu bằng thao tác đọc/ghi. Để đảm bảo tính toàn vẹn của cơ sở dữ liệu, tất cả dữ liệu phải đáp ứng kiểm tra ACID được mô tả trong Tech Note 3.1 .

**Giao dịch** là một đơn vị logic duy nhất của công việc truy cập và có thể sửa đổi nội dung của cơ sở dữ liệu.

|  |
| --- |
| Ghi chú công nghệ 3.1  Thử nghiệm tính ACID dữ liệu  Tất cả dữ liệu được thu thập, theo dõi và lưu trữ trong cơ sở dữ liệu cũng như được cơ sở dữ liệu xử lý và phân phối phải đáp ứng bốn tiêu chí— Nguyên tử , Tính nhất quán , Cô lập và Bền vững (ACID). Atomicity. Consistency. Isolation , and Durability (ACID).  **Tính nguyên tử** có nghĩa là bất kỳ sửa đổi nào đối với cơ sở dữ liệu (chèn, cập nhật, xóa) đều được thực hiện hoặc loại bỏ hoàn toàn.  **Tính nhất quán** đề cập đến việc để tất cả dữ liệu ở trạng thái nhất quán khi kết thúc một giao dịch để duy trì tính toàn vẹn của cơ sở dữ liệu.  **Sự cô lập** làm cho một giao dịch tách biệt và độc lập với bất kỳ giao dịch nào khác.  **Độ bền** đòi hỏi một giao dịch đã hoàn thành, sau khi được cam kết, sẽ trở thành một phần của cơ sở dữ liệu và không thể đảo ngược,  N  Một ví dụ đơn giản cho thấy các thuộc tính ACID của cơ sở dữ liệu hoạt động như thế nào.  Hãy tưởng tượng rằng bạn muốn chuyển một số tiền từ tài khoản séc của mình và gửi chúng vào tài khoản tiết kiệm của bạn. Để thực hiện giao dịch này, một mục nhật ký phải được thực hiện cho mục đích kiểm toán. Trong trường hợp này Atomicity có nghĩa là không thể rút tiền ra khỏi tài khoản séc của bạn mà sau đó không được gửi vào tài khoản tiết kiệm của bạn. Nếu việc chuyển tiền bị gián đoạn vì bất kỳ lý do gì, số dư tài khoản séc và tài khoản tiết kiệm của bạn sẽ không thay đổi. Tính nhất quán đảm bảo rằng hồ sơ giao dịch của bạn được ghi vào nhật ký. Nếu nhật ký đã đầy hoặc mục nhập không thể được ghi lại vì một số lý do khác, thì chuyển khoản sẽ bị hủy và số dư tài khoản của bạn trở lại số tiền ban đầu. Cô lập đảm bảo chuyển khoản của bạn không thể bị ảnh hưởng bởi bất kỳ giao dịch nào khác. Các giao dịch khác để thay đổi số dư kiểm tra của bạn phải đợi cho đến khi giao dịch hiện tại của bạn hoàn tất. Độ bền đảm bảo rằng giao dịch không bị mất sau khi được lưu hoặc được commit. Vì vậy, sự cố mất điện hoặc sự cố hệ thống sẽ không làm mất bất kỳ dữ liệu nào của bạn. |

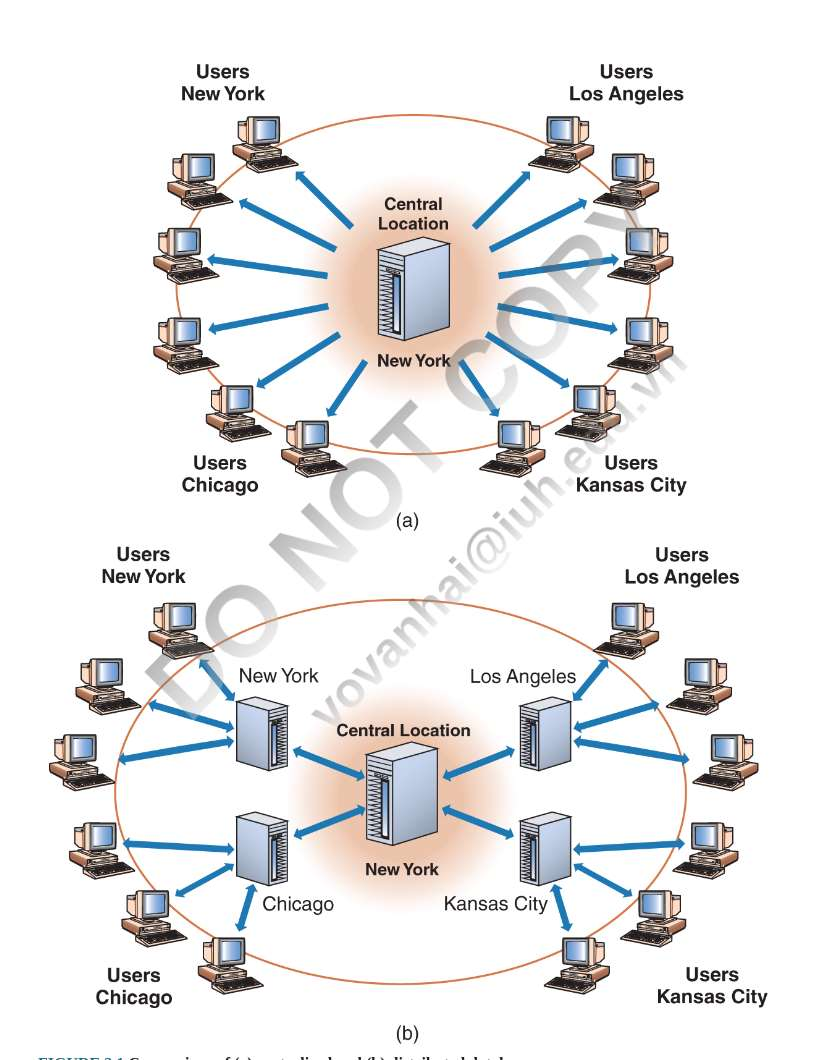
#### Kiến trúc cơ sở dữ liệu tập trung và phân tán

Các doanh nghiệp vừa và lớn thường có cơ sở dữ liệu được phân loại là tập trung hoặc phân tán như trong hình 3.1 .

**Cơ sở dữ liệu tập trung** lưu trữ tất cả dữ liệu dưới dạng một thể thống nhất trong một máy tính trung tâm duy nhất, chẳng hạn như máy tính lớn hoặc máy chủ ở một vị trí địa lý.

**Cơ sở dữ liệu phân tán** lưu trữ các phần của cơ sở dữ liệu trên nhiều máy tính được điều khiển bởi hệ thống quản lý cơ sở dữ liệu (DBMS) trong 1 mạng với cấu hình máy khách-máy chủ.

## 



HÌNH 3.1 So sánh (a) cơ sở dữ liệu tập trung và (b) cơ sở dữ liệu phân tán.

#### Cơ sở dữ liệu tập trung

Cách tiếp cận truyền thống để lưu trữ dữ liệu trong các doanh nghiệp lớn hơn là trên nền tảng cơ sở dữ liệu tập trung. Trong một cơ sở dữ liệu tập trung, dữ liệu được lưu trữ và duy trì trên một máy tính lớn ở một vị trí duy nhất. Người dùng truy cập nguồn dữ liệu duy nhất từ nhiều điểm trên mạng (hình 3.1). Khi khối lượng dữ liệu tăng lên, cấu hình tập trung này trở nên khó sử dụng và gây ra tắc nghẽn khi truy cập dữ liệu.

Lợi ích của kiến trúc cơ sở dữ liệu tập trung bao gồm:

1. **Kiểm soát tốt hơn chất lượng dữ liệu** Tính nhất quán của dữ liệu sẽ dễ dàng hơn khi dữ liệu được lưu giữ ở một vị trí thực vì việc thêm, cập nhật và xóa dữ liệu có thể được thực hiện theo cách được giám sát và có trật tự.

2. **Dễ sử dụng** Dễ bảo trì và sao lưu hơn ở một vị trí thực tế.

1. **Bảo mật CNTT tốt hơn** Dữ liệu được truy cập thông qua máy tính chủ tập trung, chúng có thể được bảo vệ dễ dàng hơn khỏi sự truy cập hoặc sửa đổi trái phép.
2. **Toàn vẹn dữ liệu tốt hơn** Dữ liệu được lưu trữ trong cơ sở dữ liệu tập trung không bị trùng lặp ở những nơi khác.

Nhược điểm của cơ sở dữ liệu tập trung bao gồm:

1. **Độ trễ truyền** Khi người dùng phân tán về mặt địa lý, máy tính trung tâm có thể phản hồi chậm do sự thay đổi về tốc độ truyền dữ liệu.
2. **Bảo mật** Tính toàn vẹn của dữ liệu có thể dễ dàng bị xâm phạm nếu ai đó có quyền truy cập vào máy tính trung tâm và dữ liệu có thể được thêm, thay đổi và xóa.

3. **Độ tin cậy** Một số lượng lớn các yêu cầu đồng thời có thể khiến máy tính trung tâm trở nên quá tải và không còn phản hồi.

4. **Khả năng mở rộng** Rất khó để mở rộng một máy tính tập trung có dung lượng hạn chế và không thể xử lý vô số yêu cầu

**Khả năng mở rộng** là khả năng tăng kích thước của hệ thống để xử lý sự tăng trưởng dữ liệu hoặc tải của số lượng người dùng đồng thời ngày càng tăng, nghĩa là các hệ thống có thể mở rộng đáp ứng hiệu quả nhu cầu điện toán hiệu suất cao.

#### Cơ sở dữ liệu phân tán

Cơ sở dữ liệu phân tán đã được giới thiệu để khắc phục các vấn đề về tốc độ và khả năng truy cập liên quan đến cách tiếp cận cơ sở dữ liệu tập trung. Trong cơ sở dữ liệu phân tán, dữ liệu của công ty được lưu trữ trên các máy chủ được điều khiển bởi hệ thống quản lý cơ sở dữ liệu (DBMS), như trong Hình 3.2 . Các máy chủ nằm trong trung tâm dữ liệu của công ty, đám mây riêng và/hoặc đám mây công cộng và các user truy cập vào các máy chủ thông qua Mạng diện rộng (WAN). Các DBMS phổ biến nhất là Oracle (https://www.oracle.com/index.html), IBM Db2 (https://www.ibm.com/analytics/db2) và Microsoft QuickBase (https://www .quickbase.com/).

##### 

HÌNH 3.2 Kiến trúc cơ sở dữ liệu phân tán cho trụ sở chính, sản xuất, bán hàng và tiếp thị. 

**Ưu điểm** của cơ sở dữ liệu phân tán bao gồm:

1**. Bảo mật** Cơ sở dữ liệu phân tán được kiểm soát bởi một cơ quan duy nhất được chỉ định xác thực thông tin đăng nhập của khách hàng trước khi cung cấp quyền truy cập vào cơ sở dữ liệu.

2, **Độ tin cậy** Nếu một trang web gặp sự cố, hệ thống sẽ tiếp tục hoạt động.

1. **Tốc độ** Tìm kiếm một phần cơ sở dữ liệu nhanh hơn so với tìm kiếm một phần dữ liệu thống nhất.
2. **Khả năng mở rộng** Tải được chia sẻ giữa một số máy tính. Do đó, tài nguyên máy chủ lớn hơn nhiều so với cơ sở dữ liệu tập trung.

**Nhược điểm** liên quan đến cơ sở dữ liệu phân tán bao gồm:

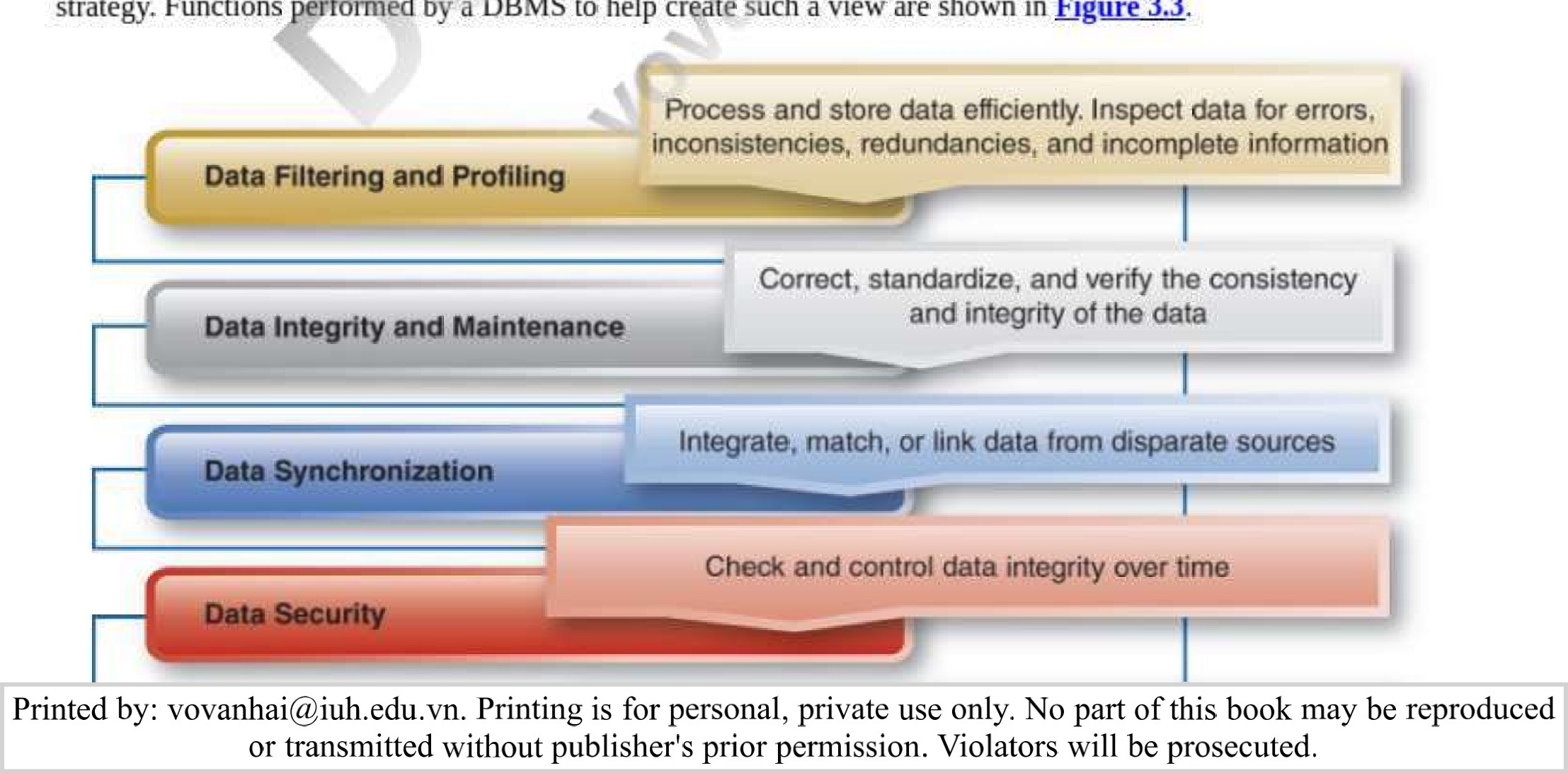
I**. Tính khả dụng** Nếu có sự cố với mạng mà cơ sở dữ liệu phân tán đang sử dụng; có thể khó truy cập dữ liệu khi cần.

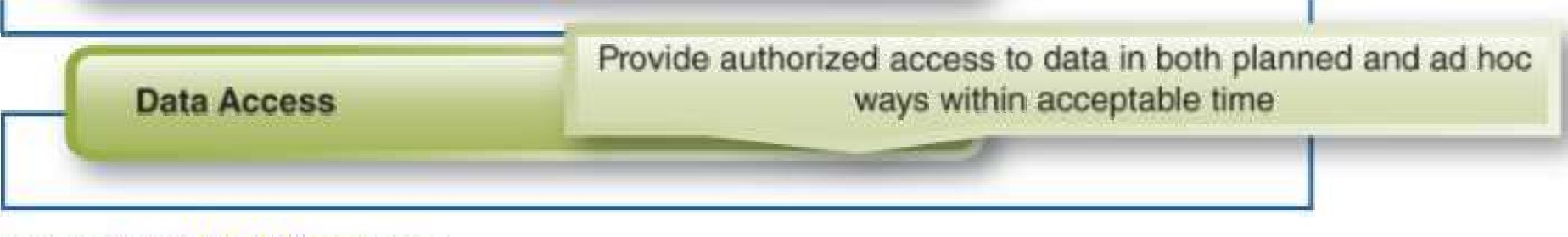
2. **Chi phí** Mua phần cứng và phần mềm phù hợp có thể tốn kém.

3. **Bảo mật** Nếu bảo mật của cơ quan được chỉ định bị xâm phạm, dữ liệu có thể bị sửa đổi hoặc thậm chí bị xóa.

##### Hệ thống quản lý cơ sở dữ liệu (DBMS)

Để tận dụng tối đa cơ sở dữ liệu có thể tự động tải, truy xuất hoặc sửa đổi dữ liệu hiện có từ hệ thống, phải sử dụng gói phần mềm được thiết kế để định nghĩa, thao tác và quản lý dữ liệu. Gói phần mềm này được gọi là hệ thống quản lý cơ sở dữ liệu..-

* **Hệ thống quản lý cơ sở dữ liệu** về cơ bản là một hệ thống lưu giữ dữ liệu được vi tính hóa, cung cấp chế độ xem dữ liệu chính xác và nhất quán trong toàn doanh nghiệp để đưa ra các quyết định sáng suốt, có thể hành động hỗ trợ chiến lược kinh doanh Các chức năng được định dạng bởi DBMS để giúp tạo một dạng xem như vậy được hiển thị trong Hình 3.3 .
* .



HÌNH 3.3 Các chức năng của DBMS.

**Xử lý giao dịch trực tuyến so với xử lý phân tích trực tuyến** Khi hầu hết các giao dịch kinh doanh xảy ra, nghĩa là một mặt hàng được bán hoặc trả lại, đơn đặt hàng được gửi hoặc hủy, thanh toán hoặc đặt cọc được thực hiện, các thay đổi được thực hiện ngay lập tức đối với cơ sở dữ liệu. Người sử dùng DBMS có thể thực hiện bốn thao tác cơ bản để thực hiện những thay đổi này đối với dữ liệu được lưu trữ trong cơ sở dữ liệu hoặc quản lý cấu trúc thực tế của cơ sở dữ liệu. Bốn thao tác này được biết đến với từ viết tắt CRUD—tạo (thêm), đọc, cập nhật và xóa.

Một DBMS ghi lại và xử lý các giao dịch này trong cơ sở dữ liệu và hỗ trợ các truy vấn và báo cáo. Với chức năng của nó, một DBMS được gọi là một **hệ thống xử lý giao dịch trực tuyến ( OLTP** ). Cơ sở dữ liệu OLTP có thể xử lý hàng triệu giao dịch mỗi giây.

**Xử lý giao dịch trực tuyến (OLTP)** **Online transaction processing** là một thiết kế cơ sở dữ liệu chia nhỏ thông tin phức tạp thành các bảng dữ liệu đơn giản hơn để đạt được sự cân bằng giữa hiệu quả xử lý giao dịch và hiệu quả truy vấn.

Cơ sở dữ liệu không phải là công cụ tốt nhất để làm việc với các tập dữ liệu lớn, phức tạp— quy trình phân tích trực tuyến ( OLAP ) online analytics processing và hỗ trợ ra quyết định. Những hạn chế này đã dẫn đến sự ra đời của công nghệ kho dữ liệu. Kho dữ liệu và siêu thị dữ liệu data marts được tối ưu hóa cho OLAP và hỗ trợ ra quyết định.

**Xử lý phân tích trực tuyến (OLAP)** online analytics processing là phân tích dữ liệu phức tạp từ kho dữ liệu.

Tóm lại, cơ sở dữ liệu được sử dụng tốt nhất để xử lý giao dịch và xử lý truy vấn cực nhanh. Mặt khác, kho dữ liệu và siêu thị dữ liệu phù hợp hơn để phân tích dữ liệu phức tạp và sẽ được thảo luận thêm trong Phần 3.2 .

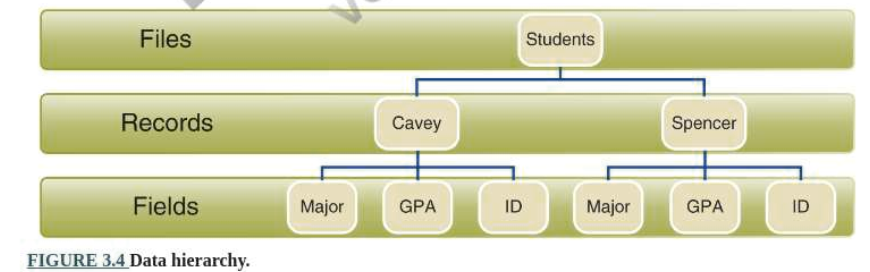
#### Các thành phần của DBMS

Có bốn yếu tố quan trọng trong bất kỳ DBMS nào. Đó là: cấu trúc dữ liệu, ngôn ngữ mô hình hóa dữ liệu, ngôn ngữ truy vấn dữ liệu và cơ chế giao dịch.

##### **Cấu trúc dữ liệu**

**Cấu trúc dữ liệu** là một định dạng chuyên biệt để tổ chức và lưu trữ dữ liệu. Các cấu trúc dữ liệu chung bao gồm tệp, bản ghi, bảng, cây, v.v.

Để thu thập, truy xuất và sửa đổi dữ liệu một cách hiệu quả, dữ liệu phải được cung cấp theo kiểu có tổ chức. Bất kỳ cấu trúc dữ liệu nào cũng được thiết kế để tổ chức dữ liệu phù hợp với mục đích cụ thể để có thể truy cập, sửa đổi và sử dụng dữ liệu theo những cách thích hợp. Trong hình 3.4, tệp (STUDENTS) là tập hợp các bản ghi liên quan (CAVEY, SPENCER) và bản ghi là tập hợp các trường liên quan. Mỗi trường dữ liệu chứa một dữ kiện hoặc thuộc tính duy nhất (MAJOR, GPA, ID) về thực thể được mô tả trong một bản ghi. Nếu có 250 sinh viên, mỗi sinh viên sẽ có một bản ghi và tập hợp tất cả 250 bản ghi sẽ tạo thành tệp STUDENTS.



HÌNH 3.4 Phân cấp dữ liệu.

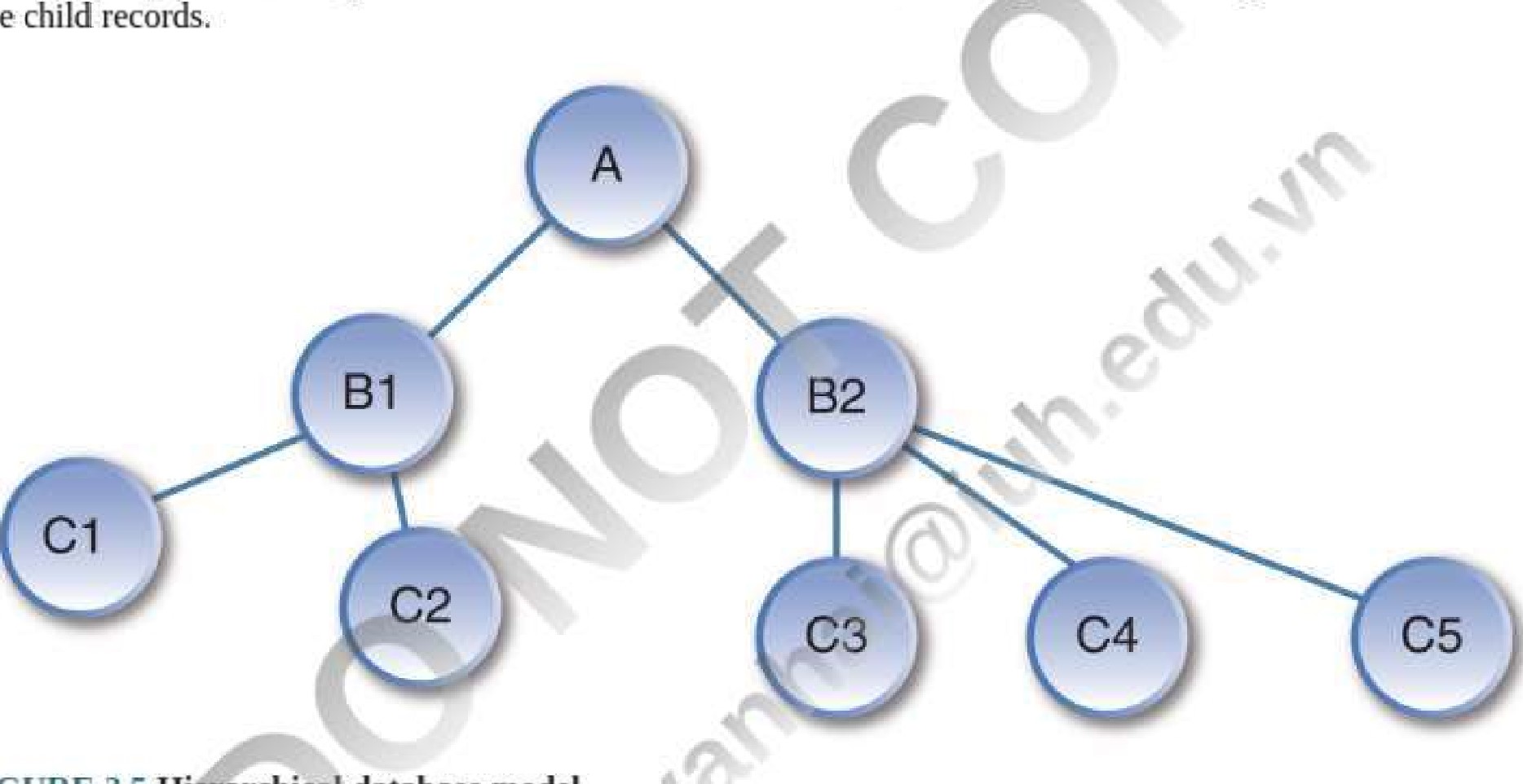
##### Ngôn ngữ mô hình hóa dữ liệu

**Ngôn ngữ truy vấn cơ sở dữ liệu** đề cập đến bất kỳ ngôn ngữ lập trình máy tính nào tìm và truy xuất dữ liệu từ cơ sở dữ liệu và hệ thống thông tin bằng cách xử lý các truy vấn của người dùng.

Trong những năm qua, một số cách tiếp cận phổ biến đối với ngôn ngữ mô hình hóa của DBMS đã được phát triển. Chúng bao gồm phân cấp, mạng, quan hệ và hướng đối tượng. Trong số này, cơ sở dữ liệu quan hệ đã trở thành mô hình cơ sở dữ liệu được sử dụng phổ biến nhất được hầu hết các doanh nghiệp áp dụng.

Lấy ví dụ về một đơn đặt hàng nhiều sản phẩm của 1 khách hàng. Trong trường hợp này, một đơn hàng (gốc) có thể có nhiều sản phẩm (con) và một sản phẩm (con) có thể xuất hiện trên nhiều đơn hàng (gốc). Mối quan hệ nhiều-nhiều này đánh bại các quy tắc phân cấp và do đó gây khó khăn, nếu không muốn nói là không thể mô hình hóa bằng cách sử dụng phương pháp phân cấp. Vì kiểu quan hệ này thường xuất hiện trong các quy trình kinh doanh, điều quan trọng là phải khắc phục hạn chế này của mô hình **cơ sở dữ liệu phân cấp** (hình 3.5 ).

**Cơ sở dữ liệu phân cấp** có cấu trúc dạng cây dựa trên mối quan hệ một-nhiều cha-con giữa các phần tử dữ liệu. Mỗi bản ghi, ngoại trừ nút ban đầu được gọi là gốc, có một bản ghi cha và mỗi bản ghi cha có một hoặc nhiều bản ghi con.



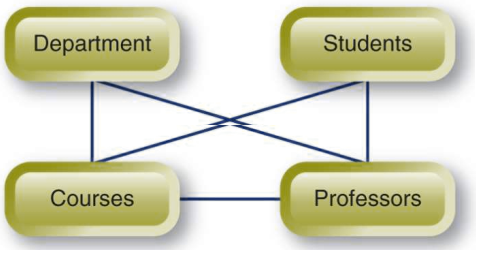
4.

HÌNH 3.5 Mô hình cơ sở dữ liệu phân cấp.

###### **Mạng**

**Mô hình dữ liệu mạng** đã được tạo để khắc phục những hạn chế của mô hình phân cấp bằng cách cho phép mỗi đứa trẻ có nhiều cha mẹ (hình 3.6 ). Tuy nhiên, cách tiếp cận mạng vẫn có vấn đề của nó. Các lập trình viên máy tính sớm phát hiện ra rằng việc đạt được sự hiểu biết có ý nghĩa về cấu trúc dữ liệu để làm cho mô hình mạng hoạt động hiệu quả là cực kỳ tốn thời gian và sự phức tạp của các mối quan hệ nhiều-nhiều rất khó thực hiện và duy trì.

Mô hình dữ liệu mạng là mô hình dữ liệu cho phép nhiều bản ghi được liên kết với cùng một bản ghi mẹ.



HÌNH 3.6 Mô hình cơ sở dữ liệu mạng.

###### **quan hệ**

Sự lặp lại tiếp theo của các mô hình cơ sở dữ liệu đã cố gắng giải quyết các vấn đề liên quan đến các mô hình cơ sở dữ liệu trước đó là mô hình dữ liệu quan hệ . Thay vì phải điều hướng qua cấu trúc cơ sở dữ liệu, cơ sở dữ liệu quan hệ lưu trữ dữ liệu trong bảng (tệp) bao gồm hàng (bản ghi) và cột (trường), tương tự như định dạng của bảng tính. Bằng cách này, nó loại bỏ tình trạng khó xử về mối quan hệ cha-con và cho phép bất kỳ tệp nào có liên quan đến tệp khác thông qua một trường chung xác định duy nhất một bản ghi riêng lẻ.

**Mô hình dữ liệu quan hệ** là một cách tiếp cận để quản lý dữ liệu bằng cấu trúc và ngôn ngữ liên quan đến việc sử dụng các bảng dữ liệu để thu thập các nhóm phần tử thành các quan hệ.

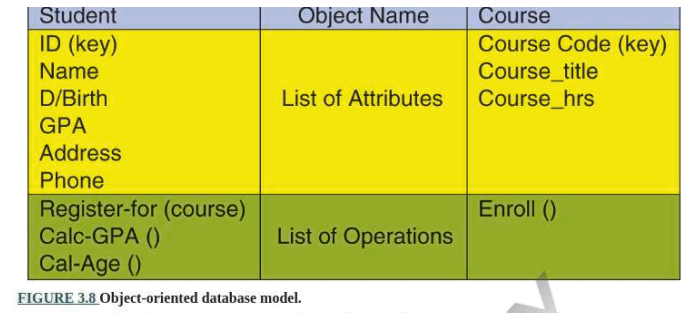
Trong hình 3.7, các bảng được gọi là Product và Invoice có liên quan với nhau thông qua trường chung "product—num". Bằng cách kết nối các tệp thông qua các mối quan hệ trực tiếp này, không cần phải theo dõi đường dẫn vòng quanh, tốn thời gian đến và từ các tệp riêng lẻ và các phiên bản mới có thể dễ dàng được thêm vào bất kỳ tệp nào khi có nhu cầu. 

# 

Hướng đối tượng (OO)

Mô hình cơ sở dữ liệu tiếp theo xuất hiện là Hệ quản lý cơ sở dữ liệu hướng đối tượng ( OODMS) hỗ trợ việc tạo và lập mô hình dữ liệu dưới dạng các đối tượng như được sử dụng trong lập trình hướng đối tượng. Một đối tượng riêng lẻ (bản ghi) kết hợp các thuộc tính (trường) với các thao tác và xử lý chúng cùng nhau (hình 3.8 ).

**Một hệ quản lý cơ sở dữ liệu hướng đối tượng** là một mô hình dữ liệu hỗ trợ việc lập mô hình và tạo ra các thực thể dữ liệu dưới dạng các đối tượng chứa cả dữ liệu và các mối quan hệ của những dữ liệu đó.



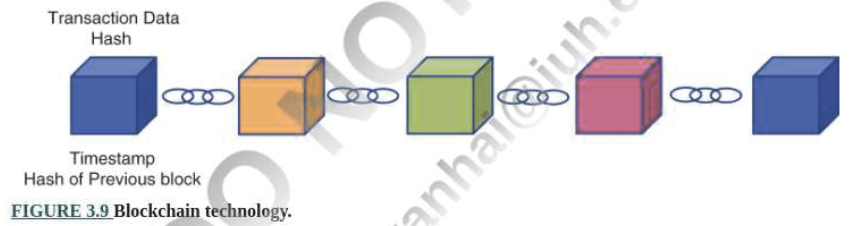
HÌNH 3.8 Mô hình cơ sở dữ liệu hướng đối tượng.

Cơ sở dữ liệu hướng đối tượng là một loại mô hình dữ liệu chuyên biệt hơn và được sử dụng tốt nhất trong các ứng dụng kinh doanh trong các môi trường phức tạp đòi hỏi xử lý hiệu năng cao. Ví dụ về các ngành có nhu cầu cao đối với loại mô hình cơ sở dữ liệu này có thể được tìm thấy trong các lĩnh vực kỹ thuật, viễn thông, dịch vụ tài chính chuyên biệt và nghiên cứu khoa học.

## **chuỗi khối - blockchain**

Chuỗi khối là mô hình cơ sở dữ liệu mới nhất (hình 3.9 ) và có khả năng thay đổi cách các công ty sắp xếp, ghi lại và xác minh các giao dịch. Blockchain ban đầu được thiết kế để đánh dấu thời gian cho các tài liệu kỹ thuật số để chúng không thể bị lùi ngày hoặc thay đổi. Sau đó, nó đã được Satoshi Nakomoto điều chỉnh để tạo ra tiền điện tử kỹ thuật số, Bitcoin. Vậy, blockchain là gì — nó hoạt động như thế nào, nó giải quyết những vấn đề gì và ngày nay nó được sử dụng như thế nào?

**Chuỗi khối** là một sổ cái phân tán được đại diện bởi một chuỗi khối dữ liệu tuần tự ghi lại các giao dịch, thiết lập danh tính của người dùng và thiết lập hợp đồng.

.

* Chuỗi khối có thể thay thế các vai trò truyền thống được thực hiện bởi các chủ ngân hàng, kiểm toán viên, luật sư và cố vấn tài chính. Với blockchain, bạn có thể loại bỏ những "người trung gian" này và tăng đáng kể hiệu quả và bảo mật.
* Mặc dù nó được triển khai dưới dạng chuỗi khối database phân tán nhưng khác với mô hình cơ sở dữ liệu quan hệ và hướng đối tượng  theo nhiều cách. Trong một chuỗi khối, nhiều người dùng có thể ghi các mục nhập vào cơ sở dữ liệu và cộng đồng người dùng kiểm soát cách dữ liệu được sửa đổi và cập nhật, thay vì quản trị viên được chỉ định của bên thứ ba trong mô hình quan hệ.
* Chuỗi khối cũng được thiết kế để bảo mật và tạo niềm tin. Bảo mật được tích hợp vào một hệ thống chuỗi khối thông qua máy chủ đánh dấu thời gian phân tán và mạng ngang hàng. Kết quả là một cơ sở dữ liệu bao gồm một chuỗi các khối dữ liệu tuần tự được quản lý tự động theo cách phi tập trung. Chuỗi các khối tuần tự cho phép mọi người gửi giá trị ở bất kỳ đâu mà tệp chuỗi khối có thể được truy cập. Các khối trong chuỗi khối được liên kết với nhau bằng cách sử dụng các khóa được tạo bằng mật mã (hash).

**Băm (Hash)** là một hàm lấy một giá trị đầu vào và xuất ra một số thập lục phân có kích thước cố định duy nhất \_ là khóa được tạo bằng mật mã cho dữ liệu.

Hàm băm là xương sống của công nghệ blochain. Mỗi block có giá trị hash của riêng nó, một nhãn thời gian, dữ liệu và giá trị hash của block trước đó như trong hình 3.9. Giá trị hash của khối trước đó liên kết chuỗi các khối và chính kỹ thuật này làm cho chuỗi khối trở nên an toàn (bất biến). Nếu một khối bằng cách nào đó bị xóa khỏi chuỗi, thì tất cả các khối tiếp theo sẽ chứa các giá trị hash không chính xác và sự hỏng hóc của chuỗi khối sẽ được biết với tất cả người dùng của nó. Khối đầu tiên không có hàm băm của khối trước đó và được gọi là khối gốc.

Để chỉnh sửa một khối, người dùng phải cung cấp hàm băm của riêng họ và khóa (băm) được tạo bằng mật mã của khối trước đó. Với thông tin đó, người dùng có thể chuyển giá trị của bất kỳ thứ gì được lưu trữ trong các khối mà họ đang chỉnh sửa. Người dùng chỉ có thể chỉnh sửa (các) khối mà họ sở hữu và các chỉnh sửa không được xác minh bằng hai khóa sẽ bị từ chối và giao dịch bị hủy bỏ. Các khối không bao giờ được thay thế. Thay vào đó, khi một khối được chỉnh sửa, một khối mới sẽ được tạo. Theo cách này, một lịch sử đầy đủ được duy trì cho tất cả các giao dịch trong chuỗi khối và tính bảo mật của dữ liệu được đảm bảo. Ví dụ: vi phạm dữ liệu trong một công ty dược phẩm có thể khiến thông tin sức khỏe được bảo vệ của hàng triệu bệnh nhân gặp rủi ro. Mã ngẫu nhiên (băm) của chuỗi khối cần thiết để mã hóa thông tin có nghĩa là không có gì để giải mã vì thông tin không thể bị thay đổi sau khi được nhập vào, tạo ra một môi trường an toàn hơn nhiều so với trước đây.

**Bằng chứng công việc và mạng ngang hàng** **Proof -of-work and peer-to-peer network** là hai cơ chế khác được sử dụng để tăng cường bảo mật cho chuỗi khối. Proof-of-work làm chậm quá trình tạo các khối mới khi dữ liệu được thay đổi trong một khối hiện có để ngăn chặn giả mạo. Mạng ngang hàng cho phép mọi người tham gia chuỗi khối và gửi một bản sao đầy đủ của chuỗi khối cho mỗi người dùng. Khi một khối mới được tạo, khối đó sẽ được gửi tới mọi người trên mạng và mỗi người dùng sẽ kiểm tra khối mới để đảm bảo rằng nó không bị thay đổi. Nếu phần lớn người dùng xác minh khối mới, mỗi người dùng sẽ thêm nó vào chuỗi khối của riêng họ. Đây được gọi là sự đồng thuận và nó điều khiển khối nào sẽ được thêm vào chuỗi khối và khối nào sẽ không. Sự phát triển gần đây nhất trong chuỗi khối là "hợp đồng thông minh". Đây là các hướng dẫn được lưu trữ trên chuỗi khối để tự động thực hiện các giao dịch lặp lại dựa trên các điều kiện nhất định. Hơi giống với “thanh toán tự động" trong các ứng dụng tài chính.

Công nghệ chuỗi khối đang tạo ra sự quan tâm đáng kể trong các ngành công nghiệp toàn cầu. Ví dụ: Walmart và thực phẩm của họ sử dụng chuỗi khối để theo dõi rau xanh của họ thông qua toàn bộ quy trình sản xuất và hậu cần để đảm bảo khách hàng có được thứ họ muốn và tin tưởng vào chất lượng của nó. Ngoài bán lẻ, công nghệ chuỗi khối đang được sử dụng cho các ứng dụng sâu rộng trong lĩnh vực tài chính, chăm sóc sức khỏe, truyền thông, chính phủ và các ngành công nghiệp khác. Bemard Marr ( 2018 ) gần đây đã xuất bản một số ví dụ thực tế thú vị về việc sử dụng chuỗi khối trong các ngành khác nhau, bao gồm:

* **Chăm sóc sức khỏe** Trung tâm phòng ngừa và kiểm soát dịch bệnh (<https://www.cdc.gov/> ) đang đưa dữ liệu bùng phát dịch bệnh vào một chuỗi khối để tăng hiệu quả ứng phó và cứu trợ thảm họa
* **Ngân hàng** Finance Barclays (<https://home.barciays/> )đã đưa ra một số sáng kiến blockchain để theo dõi các giao dịch tài chính, tuân thủ và chống gian lận.
* **Sản xuất** Một dự án cộng đồng do doanh nghiệp lãnh đạo có trụ sở tại Brooklyni NY, cho phép các thành viên sản xuất và bán năng lượng tại địa phương để giảm chi phí phân phối năng lượng.
* **Chính phủ** Bộ Lao động và Hưu trí Vương quốc Anh

(<https://www.gov.uk/governmenüorganisations/department-for-work-pensions> ) đang khai thác việc sử dụng công nghệ chuỗi khối để ghi lại và quản lý các khoản thanh toán phúc lợi.

* **Tổ chức phi lợi nhuận** Save the Children (<https://www.savethechildren.org/> ), The Water Project (<https://thewaterproject.org/> ); và Medic Mobile (<https://medicmobile.org/> ) đang khám phá việc sử dụng chuỗi khối để cung cấp tính minh bạch cao hơn và tạo ra các liên kết rõ ràng hơn giữa các khoản đóng góp và kết quả của dự án.
* **bán lẻ** OpenBazaar đang sử dụng chuỗi khối để xây dựng thị trường phi tập trung (https://openbazaar.org/), nơi hàng hóa và dịch vụ có thể được giao dịch mà không cần người trung gian.
* **Bất động sản** Ubiquity (https://www.myubiquity.com/) đang tạo ra một hệ thống dựa trên chuỗi khối để làm trơn tru quy trình pháp lý phức tạp tạo ra ma sát và chi phí trong chuyển nhượng bất động sản.
* **Giao thông vận tải và du lịch** Arcade City (https://www.facebook.com/ArcadeCityHall/) chuyển dịch vụ chia sẻ đi xe và thuê xe lên blockchain để cung cấp cho người lái xe lựa chọn thanh toán bao gồm cả Bitcoin, để cạnh tranh với Uber.
* **Media** Kodak (https://www.kodak.com/) đang phát triển một hệ thống chuỗi khối để theo dõi quyền sở hữu trí tuệ và các khoản thanh toán cho các nhiếp ảnh gia.

Trong tất cả các lĩnh vực công nghiệp, mức độ tin cậy và bảo mật mà hiện tại chỉ blockchain mới có thể cung cấp là rất quan trọng đối với chuỗi cung ứng an toàn . Một vi phạm dữ liệu duy nhất có thể làm mất lòng tin của người tiêu dùng đối với một công ty, đe dọa quyền sở hữu trí tuệ, khiến các thành viên trong nhóm gặp rủi ro, tạo ra các vi phạm tuân thủ tiềm ẩn và góp phần gây ra hậu quả nghiêm trọng trong các chuỗi cung ứng cụ thể, chẳng hạn như những hậu quả gặp phải trong chuỗi cung ứng thực phẩm và chăm sóc sức khỏe của Walmart.

**Chuỗi cung ứng** là một mạng lưới bao gồm một công ty và các nhà cung cấp, người bán hàng, nhà kho, công ty vận tải, trung tâm phân phối và cửa hàng bán lẻ để sản xuất và phân phối sản phẩm hoặc dịch vụ.

**IT at Work 3.1** mô tả cách sử dụng Chuỗi khối đang cho thấy tiềm năng to lớn trong lĩnh vực năng lượng và tiện ích, nơi các lợi ích đang được nhìn thấy trong:

* Giảm giá

• Thúc đẩy sự bền vững môi trường

• Tăng tính minh bạch cho các bên liên quan trong khi không ảnh hưởng đến quyền riêng tư

|  |
| --- |
| CNTT tại nơi làm việc 3.1  Chuỗi khối thúc đẩy các mô hình kinh doanh mới về năng lượng và tính bền vững  Pricewaterhouse Cooper (PwC), một mạng lưới dịch vụ chuyên nghiệp đa quốc gia lớn (https://www.pwc.com/), đã xác định lĩnh vực năng lượng là một trong những ngành mà chuỗi khối có thể có tác động đột phá và biến đổi lớn nhất. Diễn đàn Kinh tế Thế giới, Viện Môi trường Stanford Woods và PwC gần đây đã phát hành một báo cáo chung xác định nhiều cách mà blockchain đã và sẽ mang lại lợi ích cho ngành năng lượng. Một trong những ngành công nghiệp nó có thể giúp ích là dầu khí. Cơ sở hạ tầng yếu kém và các vấn đề về tính minh bạch, hiệu quả và tối ưu hóa đang tác động tiêu cực đến các công ty dầu khí. Ví dụ, xăng dầu là một trong những mặt hàng được giao dịch nhiều nhất và đòi hỏi một mạng lưới phức tạp bao gồm các nhà máy lọc dầu, tàu chở dầu, công nhân, chính phủ và các cơ quan quản lý khác. Thật không may, mạng này bao gồm cơ sở hạ tầng bị cô lập và vô số quy trình không hiệu quả. Để giải quyết những vấn đề này, các công ty dầu khí lớn đang đầu tư vào công nghệ chuỗi khối để giảm chi phí và giảm tác động có hại đến môi trường. Công ty dầu khí mà lo ngại về quyền riêng tư và bí mật thương mại sẽ tạo ra các chuỗi khối riêng tư với các tính năng: như quyền truy cập dữ liệu và hợp đồng thông minh.  Các công ty tiện ích như PG&E (http://www.pgecorp.coml), một công ty tiện ích lớn ở Bắc California, cũng có một số vấn đề mà chuỗi khối có thể giúp giải quyết. Bằng cách sử dụng chuỗi khối, PG&E hy vọng duy trì một sổ cái phân tán để theo dõi chuỗi hành trình của các vật liệu lưới điện. Phục vụ 16 triệu người dân California đòi hỏi phải có một danh mục lớn các cơ sở hạ tầng điện được lắp đặt, điều này dễ dẫn đến sự thiếu hiệu quả. Công ty cũng hỗ trợ 340.000 mái nhà năng lượng mặt trời với 6.000 mái nhà mới tham gia mỗi tháng. có nghĩa là các chương trình thiết bị do khách hàng đặt lên và tương tác với lưới năng lượng của họ. Vì vậy, họ cần các công cụ tốt hơn để kiểm đếm/tính toán lượng sản xuất và tiêu thụ điện. Sổ cái phân phối chuỗi khối có thể giúp thực hiện điều này bằng cách tạo chuỗi hành trình cho các vật liệu lưới để nâng cao khả năng hiển thị của PG&E về tình trạng tài sản của chính họ và giảm chi phí theo dõi chúng theo cách thủ công, Dự án đầu tiên PG&E triển khai là theo dõi các cuộn thép mang dây cáp trên cao, Nếu điều này hiệu quả, họ có kế hoạch mở rộng chuỗi khối để theo dõi máy biến áp, ống dẫn khí đốt và cột điện. Sự tự tin được xây dựng bởi các sổ cái phân tán này có thể cung cấp một liên kết quan trọng giữa hành vi của khách hàng và các giao dịch với bên thứ ba.  Các công ty cũng đang sử dụng các giải pháp chuỗi khối để tăng cường tính bền vững trên toàn cầu. Ví dụ: ở Hà Lan, CirCularise (https://www.circularise.com/) đã tạo một chuỗi khối kết hợp với công nghệ ngang hàng và kỹ thuật mã hóa để tạo ra một nền tảng liên lạc và lưu trữ thông tin phi tập trung cung cấp giá chính xác cho các vật liệu tái chế và theo dõi số lần sản phẩm đã được tái chế. Ví dụ: nó có thể truyền thông tin về vật liệu tái chế giữa những người tham gia trong chuỗi cung ứng đồng thời cho phép họ ẩn danh để tiết lộ cách họ hỗ trợ tính bền vững. Tại Na Uy, Empower (https://wvvv.circularise.com/)—một công ty khởi nghiệp phát triển kinh tế—sử dụng mã thông báo chuỗi khối để khuyến khích tái chế dựa trên đóng góp. Đối với mỗi đô la Euro được quyên góp bởi một tổ chức Empower cam kết thu hồi cùng một lượng rác thải nhựa (tính bằng kg) tính theo trọng lượng để giúp làm sạch môi trường và xóa đói giảm nghèo.  Nguồn: Tổng hợp từ Clawson (201B), PwC (2018 ), Spector ( 2018 ), Diễn đàn Kinh tế Thế giới ( 2019 ), Peshkam (2019) và pgecorp.com. |

## Ngôn ngữ truy vấn dữ liệu

Để có quyền truy cập và sửa đổi dữ liệu trong cơ sở dữ liệu, bạn cần có phần mềm có thể giao tiếp với cơ sở dữ liệu. Đây là nơi các ngôn ngữ truy vấn dữ liệu phát huy tác dụng. Ngôn ngữ truy vấn cơ sở dữ liệu cho phép người dùng đặt câu hỏi về dữ liệu và cho phép duy trì tính bảo mật của cơ sở dữ liệu bằng cách giám sát dữ liệu đăng nhập, gán quyền truy cập cho người dùng và cung cấp các giao thức để thêm dữ liệu vào hệ thống.

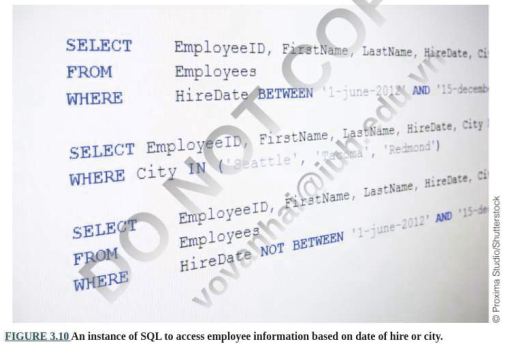
**Truy vấn** (Query) là một yêu cầu đặc biệt (không có kế hoạch) của người dùng đối với dữ liệu cụ thể.

Ngôn ngữ truy vấn có cấu trúc (SQL) là ngôn ngữ khai báo chủ đạo được sử dụng rộng rãi nhất , hoạt động với bất kỳ cơ sở dữ liệu nào để đơn giản hóa việc truy cập dữ liệu bằng cách yêu cầu người dùng chỉ khai báo dữ liệu họ muốn thay vì chỉ cho DBMS cách lấy dữ liệu đó. Có thể bạn đang sử dụng SQL mà không hề biết. Ví dụ: nếu bạn tìm kiếm một danh mục trực tuyến trên một trang web lấy thông tin đầu vào của bạn trên một biểu mẫu hoặc bằng một vài cú nhấp chuột, thì rất có thể khi bạn nhấn nút gửi, nó sẽ soạn một truy vấn SQL để truy xuất bất kỳ bản ghi nào phù hợp với tìm kiếm của bạn, truy xuất thông tin và tạo một trang Web dành riêng cho (các) cụm từ tìm kiếm của bạn. SQL là một ngôn ngữ lập trình tương đối dễ học và việc thành thạo SQL hiện đang được các nhà tuyển dụng yêu cầu cao.

**Ngôn ngữ truy vấn có cấu trúc** (SQL) là ngôn ngữ truy vấn được tiêu chuẩn hóa để truy cập cơ sở dữ liệu.

**Ngôn ngữ khai báo** là ngôn ngữ lập trình cấp cao cho phép người dùng diễn đạt ý họ muốn mà không cần chỉ định cách lấy nó, do đó tách quá trình nêu vấn đề khỏi quá trình Giải quyết vấn đề đó.

hình 3.10 minh họa một yêu cầu SQL cho ba danh sách nhân viên (1) được thuê từ ngày 1 tháng 6 năm 2012 đến ngày 15 tháng 12 năm 2012; (2) sống ở Seattle, Tacoma hoặc Redmond; và (3) KHÔNG được thuê trong khoảng thời gian từ ngày 1 tháng 6 năm 2012 đến ngày 15 tháng 12 năm 2012, sử dụng ba lệnh SQL đơn giản (SELECT, FROM, WHERE).

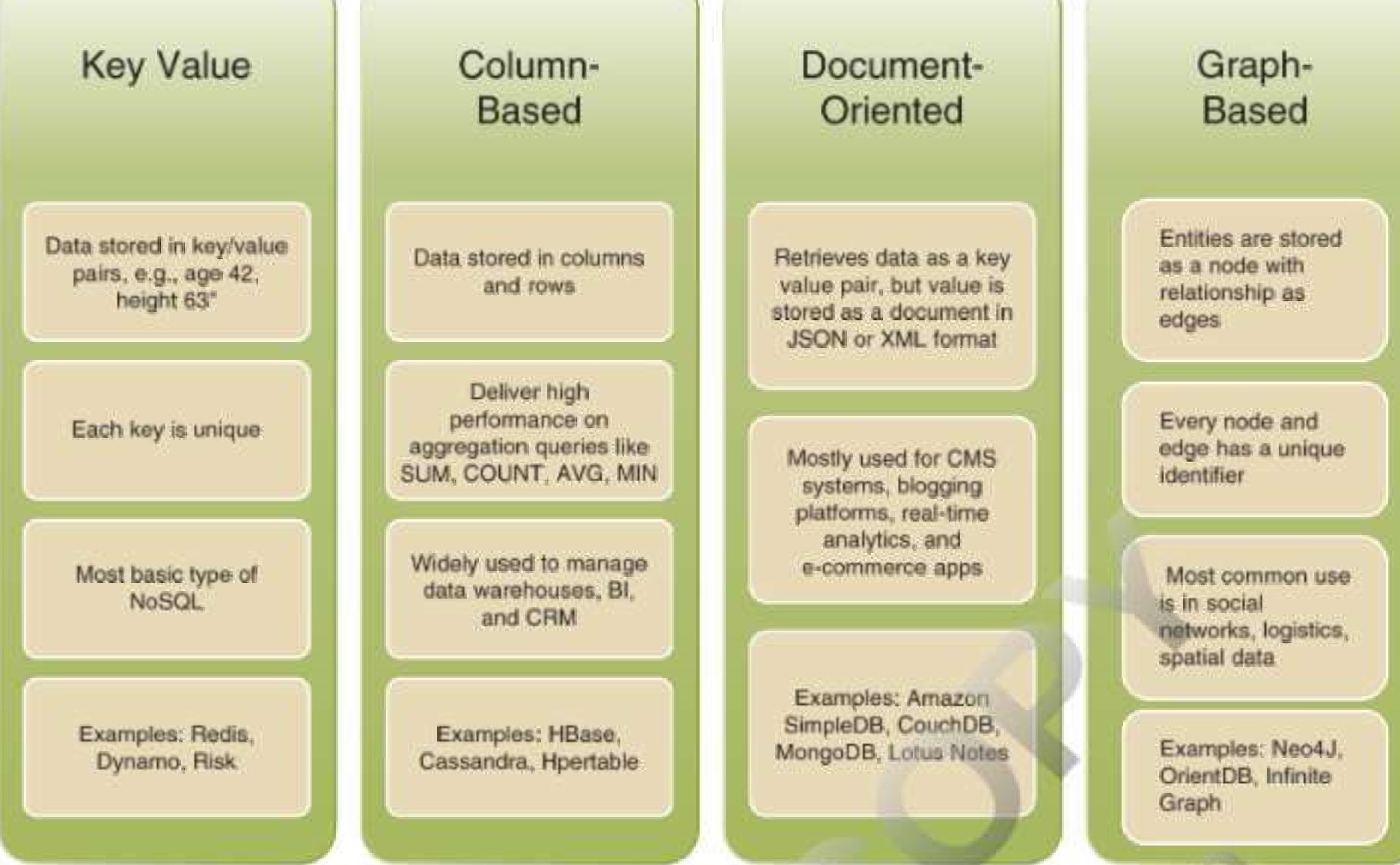


HÌNH 3.10 Một phiên bản SQL để truy cập thông tin nhân viên dựa trên ngày thuê hoặc thành phố.

Để đảm bảo tính nhất quán trong cơ sở dữ liệu SQL, trước và sau khi giao dịch được thực thi, tất cả dữ liệu được kiểm tra dựa trên bốn thuộc tính cơ sở dữ liệu ACID, được mô tả trước đó trong Tech Note 3.1 để đảm bảo các thay đổi đối với cơ sở dữ liệu được lưu một cách nhất quán, an toàn và mạnh mẽ.

## Not Only SQL ( NoSQL)

NoSQL là một ngôn ngữ truy vấn cơ sở dữ liệu phi quan hệ. Không giống như SQL, NoSQL không phải là ngôn ngữ truy vấn khai báo cấp cao. Thay vào đó, nó không có lược đồ xác định trước, để tránh join, dễ mở rộng quy mô và có thể dễ dàng xử lý dữ liệu phi cấu trúcvà không thể đoán trước. NoSQL được thiết kế cho cơ sở dữ liệu phân tán với nhu cầu dữ liệu quy mô rất lớn và được sử dụng bởi các tập đoàn có số lượng giao dịch hàng ngày khổng lồ như Facebook, Twitter, Amazon, Linkedln và Google. Có bốn loại cơ sở dữ liệu NoSQL chính: Lưu trữ khóa-giá trị; hướng cột; đồ thị; và định hướng tài liệu. Mỗi loại trong số này có các thuộc tính và khả năng ứng dụng cụ thể của riêng nó như được trình bày trong hình 3.11 .



HÌNH 3.11 Các loại và đặc điểm của cơ sở dữ liệu phân tán NoSQL.

Không có giải pháp NoSQL đơn lẻ nào tốt hơn bất kỳ giải pháp nào khác, tuy nhiên, một số cơ sở dữ liệu NoSQL đang giải quyết tốt các vấn đề cụ thể, chẳng hạn như: Thay vì tuân theo các yêu cầu ACID của SQL, NoSQL được hướng dẫn bởi **định lý CAP** (**tính nhất quán, tính khả dụng và dung sai phân vùng -** consistency, availability and partition tolerance) thừa nhận rằng cơ sở dữ liệu phân tán chỉ có thể đáp ứng hai trong số ba thành phần của nó. Do đó, một số cơ sở dữ liệu NoSQL tuân thủ AP (tính khả dụng, dung sai phân vùng) trong khi những cơ sở dữ liệu khác đáp ứng CA (tính nhất quán, tính khả dụng) hoặc CP (tính nhất quán, dung sai phân vùng).

Ưu điểm của NoSQL bao gồm:

* hiệu suất cao hơn
* dễ dàng phân phối dữ liệu trên các nút khác nhau, cho phép khả năng mở rộng và khả năng chịu lỗi

**Khả năng chịu lỗi** có nghĩa là không có lỗi đơn lẻ nào dẫn đến mất dịch vụ.

* linh hoạt hơn
* quản trị đơn giản hơn

Với hầu hết các dịch vụ NoSQL, phần lớn chi phí không nằm ở việc mua cơ sở dữ liệu, mà là ở việc triển khai nó.

**NewSQL** là loại cơ sở dữ liệu có thể mở rộng mới nhất hỗ trợ giao diện SQL. Nó kết hợp độ tin cậy của SQL, tốc độ và hiệu suất của NoSQL để cung cấp chức năng và dịch vụ tốt hơn, đồng thời cung cấp cho quản trị viên cơ sở dữ liệu các đảm bảo về hiệu suất ACID. NewSQL hoạt động tốt trong các tổ chức quan tâm đến việc chuyển các ứng dụng hiện có sang nền tảng dữ liệu lớn, phát triển các ứng dụng mới trên các hệ thống xử lý giao dịch trực tuyến có khả năng mở rộng cao và muốn tận dụng kiến thức hiện có của họ về xử lý giao dịch trực tuyến (OLTP).

### **Cơ chế giao dịch**

**Cơ chế giao dịch** là một tập hợp các hoạt động logic được sử dụng để giúp đảm bảo tính toàn vẹn, đồng thời và loại bỏ tính đa dạng.

Có hai cơ chế giao dịch chính. Một giao dịch có thể là cam kết (commit) hoặc khôi phục (rolled back). Giao dịch cam kết khi các thay đổi vĩnh viễn được thực hiện trong giao dịch. Giao dịch khôi phục xảy ra, khi các thay đổi đối với dữ liệu đã được thực hiện trong một giao dịch không được cam kết sẽ được hoàn tác.

### Lợi ích của DBMS

Các hệ thống quản lý cơ sở dữ liệu mang lại lợi ích cho các tổ chức bằng cách giảm độ trễ, tính không ổn định và tính nhất quán của dữ liệu. Chúng cũng hữu ích trong việc giảm thời gian phản hồi truy vấn, tăng khả năng dự đoán của truy vấn và xử lý chúng trong thời gian thực để cho phép đưa ra quyết định nhanh hơn, tự tin hơn. Các yếu tố này cần được đánh giá để xác định tính hiệu quả của DBMS như được giải thích trong Tech Note 3.2 .

|  |
| --- |
| Ghi chú công nghệ 3.2  Các yếu tố ảnh hưởng đến hiệu suất của DBMS  Các yếu tố cần xem xét khi đánh giá hiệu suất của hệ thống quản lý cơ sở dữ liệu bao gồm:   * **Độ trễ dữ liệu** . Thời gian đã trôi qua (hoặc độ trễ) giữa khi dữ liệu được tạo/và khi dữ liệu sẵn có cho một truy vấn hoặc báo cáo có thể ảnh hưởng đáng kể đến hiệu suất của cơ sở dữ liệu. Cơ sở dữ liệu có độ trễ thấp trải qua thời gian trễ nhỏ, trong khi cơ sở dữ liệu có độ trễ cao trải qua độ trễ dài; Các ứng dụng có dung sai khác nhau cho độ trễ. Ví dụ: để hoạt động tốt, cơ sở dữ liệu thị trường chứng khoán cần độ trễ thấp, trong khi độ trễ cao hơn có thể chấp nhận được đối với cơ sở dữ liệu khách hàng của nhà bán lẻ. Hệ thống cơ sở dữ liệu có xu hướng có độ trễ ngắn hơn so với data warehouses.      * **Volatility** . Cơ sở dữ liệu cần có đủ sức mạnh xử lý để xử lý sự biến động của dữ liệu. Tốc độ mà dữ liệu được thêm, cập nhật hoặc xóa xác định khối lượng công việc mà cơ sở dữ liệu phải có khả năng kiểm soát để ngăn ngừa sự cố với tốc độ phản hồi cho các truy vấn,     **Tính nhất quán của dữ liệu**. Dữ liệu phải nhất quán. Có hai loại nhất quán dữ liệu. Tính nhất quán ngay lập tức **Immediate consistency** có nghĩa là ngay khi dữ liệu được cập nhật, phản hồi cho bất kỳ truy vấn mới nào sẽ trả về giá trị được cập nhật. Phương pháp này cung cấp dữ liệu nhất quán cập nhật nhưng yêu cầu độ trễ cao. Tính nhất quán cuối cùng **Eventual consistency** có nghĩa là không phải tất cả các câu trả lời truy vấn đều phản ánh dữ liệu thay đổi đồng nhất và hoạt động tốt ở độ trễ thấp. Trong mô hình này, dữ liệu cuối cùng sẽ trở nên nhất quán (onsistent eventually), nhưng trong thời gian chờ đợi, kết quả truy vấn không nhất quán có thể gây ra sự cố nghiêm trọng cho các phân tích phụ thuộc vào dữ liệu chính xác.   * **Query response time**. Khối lượng dữ liệu ảnh hưởng đến thời gian phản hồi đối với truy vấn và khám phá dữ liệu. Nhiều cơ sở dữ liệu có dữ liệu trước giai đoạn—nghĩa là tóm tắt hoặc tính toán trước kết quả—do đó các truy vấn có tốc độ phản hồi nhanh hơn. * **Query predictability**. Số lượng truy vấn không xác định hoặc không thể đoán trước càng nhiều thì cơ sở dữ liệu càng cần linh hoạt, Quản lý hiệu suất truy vấn hoặc cơ sở dữ liệu càng khó khăn khi khối lượng công việc khó đoán đến mức không thể chuẩn bị trước, Khả năng xử lý khối lượng công việc là tiêu chí quan trọng nhất khi lựa chọn cơ sở dữ liệu. * **Query processing capabilities**. Truy vấn cơ sở dữ liệu được xử lý trong thời gian thực và kết quả được truyền qua mạng có dây hoặc mạng tới màn hình máy tính hoặc thiết bị cầm tay. |

Xếp hạng nhà cung cấp DBMS



Năm hệ thống quản lý cơ sở dữ liệu doanh nghiệp hàng đầu năm 2019 là Oracle's Database l8c, Microsoft SQL Server 2019, IBM DB2, SAP Sybase ASE và PostgreSQL (Stroud 2019 ):

1. **Oracle Database 18c** hợp nhất và quản lý cơ sở dữ liệu dưới dạng dịch vụ đám mây thông qua kiến trúc nhiều bên thuê của Oracle và khả năng xử lý dữ liệu trong bộ nhớ và có thể được cung cấp nhanh chóng.
2. **Microsoft SQL Server** Tính dễ sử dụng, tính khả dụng và khả năng tích hợp hệ điều hành Windows của Microsoft SQL Server khiến nó trở thành lựa chọn dễ dàng cho các hãng mà chọn sản phẩm của Microsoft cho doanh nghiệp của họ.

3.**IBM Db2** được sử dụng rộng rãi trong các trung tâm dữ liệu lớn và chạy trên Linux, UNIX, Windows, IBM iSeries và máy tính lớn.

1. **SAP Sybase ASE** là một lực lượng chính sau 25 năm thành công và cải tiến. Nó hỗ trợ khóa phân vùng, giới hạn truy vấn thoải mái, tối ưu hóa kế hoạch truy vấn và gán luồng động.
2. **PostgreSQL** là cơ sở dữ liệu quan hệ-đối tượng nguồn mở tiên tiến nhất, thường được sử dụng bởi các ứng dụng trò chơi trực tuyến và Skype và Yahoo! Cơ sở dữ liệu này chạy trên nhiều hệ điều hành bao gồm Linux, Windows, FreeBSD và Solaris.

#### câu hỏi

1. Mô tả mục đích và lợi ích của việc quản lý dữ liệu.
2. Định nghĩa cơ sở dữ liệu và hệ quản trị cơ sở dữ liệu (DBMS).
3. Bốn yếu tố của một DBMS là gì?
4. Giải thích chức năng của hệ thống xử lý giao dịch trực tuyến (OLAP) và công nghệ cơ sở dữ liệu nào phù hợp nhất để sử dụng.
5. Mô tả các chức năng của một DBMS.
6. Hệ quản trị cơ sở dữ liệu quan hệ là gì?

7. Các yếu tố chính của một khối trong chuỗi khối là gì?

8. Ba cơ chế giúp giữ an toàn cho chuỗi khối là gì?

## 3.2 Kho dữ liệu và chợ dữ liệu - Data Warehouses and Data Marts

L03.2 Xác định sự khác biệt giữa cơ sở dữ liệu, kho dữ liệu và siêu thị dữ liệu và lý do tại sao một công ty sẽ chuyển từ cơ sở dữ liệu sang kho dữ liệu hoặc siêu thị dữ liệu

Số lượng lớn các giao dịch xảy ra hàng ngày trong một tổ chức đòi hỏi dữ liệu trong cơ sở dữ liệu được sửa đổi hoặc cập nhật liên tục. Tính hay thay đổi của cơ sở dữ liệu khiến chúng ta không thể sử dụng chúng cho các nhiệm vụ ra quyết định và giải quyết vấn đề phức tạp„ Vì lý do này, dữ liệu được trích xuất từ cơ sở dữ liệu, được chuyển đổi (xử lý để chuẩn hóa dữ liệu), sau đó được tải vào kho dữ liệu .

**Kho dữ liệu** là nguồn chính của dữ liệu đã được làm sạch để phân tích, báo cáo và kinh doanh thông minh (BI). Thông thường, dữ liệu được tóm tắt theo cách cho phép phản hồi nhanh các truy vấn. Chẳng hạn, kết quả truy vấn có thể tiết lộ những thay đổi trong hành vi của khách hàng và thúc đẩy quyết định phát triển lại chiến lược quảng cáo. Kho dữ liệu lưu trữ dữ liệu từ các hệ thống nguồn và cơ sở dữ liệu khác nhau trong toàn doanh nghiệp để chạy các truy vấn phân tích đối với các tập dữ liệu khổng lồ được thu thập trong khoảng thời gian dài.

**Kho dữ liệu** là nơi lưu trữ trung tâm dữ liệu tích hợp từ một hoặc nhiều nguồn khác nhau.

Chi phí cao của kho dữ liệu có thể khiến chúng trở nên quá đắt đối với một công ty để triển khai. **Data mart** là phiên bản thu nhỏ, chi phí thấp hơn của kho dữ liệu có thể được triển khai trong thời gian ngắn hơn nhiều, chẳng hạn như trong vòng chưa đầy 90 ngày. Data mart phục vụ một bộ phận hoặc chức năng cụ thể, chẳng hạn như tài chính, tiếp thị hoặc hoạt động. Do chúng chỉ lưu trữ lượng dữ liệu nhỏ hơn nên chúng sẽ nhanh hơn và dễ dàng điều hướng cũng như bảo trì hơn.

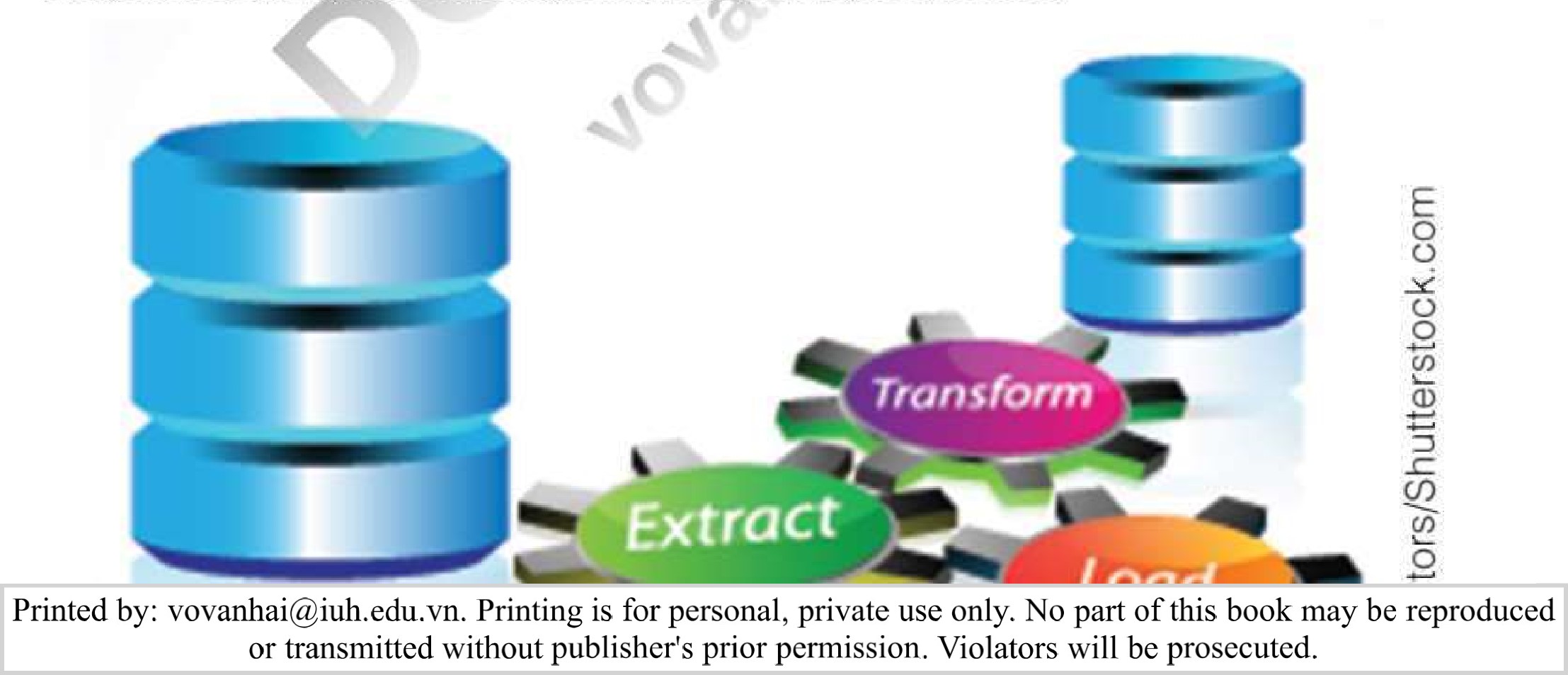
**Data mart** là một kho dữ liệu quy mô nhỏ hỗ trợ một chức năng hoặc một bộ phận. 

### Di chuyển dữ liệu từ Cơ sở dữ liệu sang Kho dữ liệu hoặc Data Mart

Tính không ổn định (volatility) do quá trình xử lý giao dịch liên tục gây ra khiến cho việc phân tích dữ liệu trở nên khó khăn—và nhu cầu xử lý hàng triệu giao dịch mỗi giây tiêu tốn sức mạnh xử lý của cơ sở dữ liệu. Ngược lại, dữ liệu trong kho tương đối ổn định, khi cần để phân tích. Do đó, dữ liệu được chọn sẽ được chuyển từ cơ sở dữ liệu sang kho dữ liệu hoặc siêu thị dữ liệu. Hãy xem xét cơ sở dữ liệu của ngân hàng, Mọi khoản tiền gửi, rút tiền, thanh toán khoản vay hoặc các giao dịch khác được thực hiện bởi hàng nghìn người dùng của ngân hàng đều bổ sung hoặc thay đổi dữ liệu; Cụ thể, dữ liệu được di chuyển theo ba bước chính.

1. Trích xuất từ cơ sở dữ liệu được chỉ định
2. Được chuyển đổi bằng cách chuẩn hóa các định dạng làm sạch dữ liệu và tích hợp chúng.
3. Tải vào kho dữ liệu.

Ba quy trình này— **extract, transform, and load (ETL)** được trình bày trong hình 3.12 . Trong kho dữ liệu hoặc siêu thị dữ liệu, dữ liệu chỉ đọc và không thay đổi cho đến khi chu kỳ ETL tiếp theo được thực hiện.



o

o

HÌNH 3.12 Dữ liệu được đưa vào cơ sở dữ liệu từ hệ thống giao dịch. Dữ liệu quan tâm được trích xuất từ cơ sở dữ liệu, được chuyển đổi để làm sạch và chuẩn hóa chúng, sau đó được tải vào kho dữ liệu. Ba quá trình này được gọi là ETL.

Sau khi dữ liệu được tải vào kho dữ liệu hoặc siêu thị dữ liệu, hai thao tác nữa— ghi nhận dữ liệu thay đổi ( CDC ) và chống trùng lặp dữ liệu **change data capture (CDC)** and **data deduplication** —được thực hiện để chuẩn bị dữ liệu thô cho phân tích:

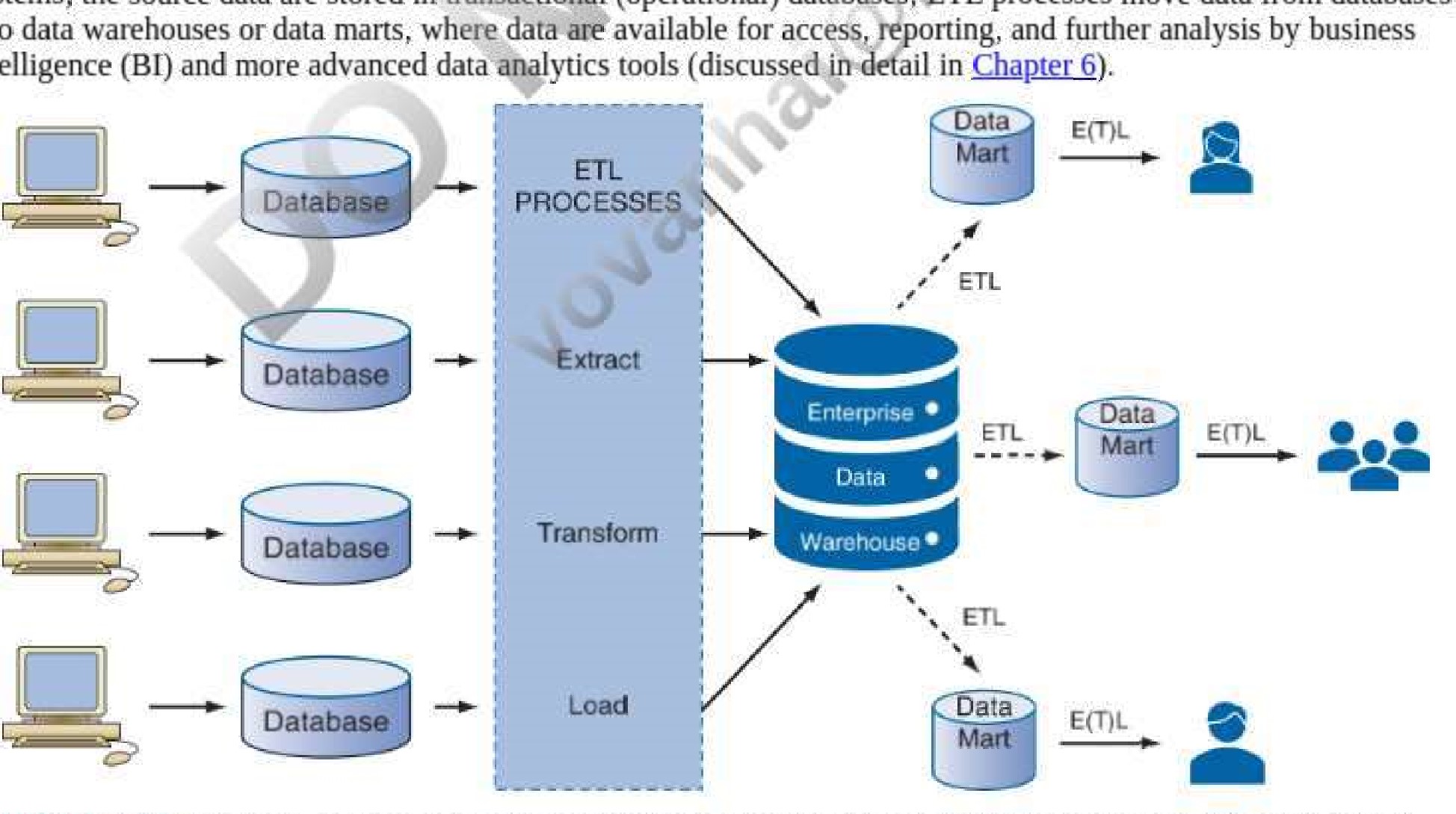
**Các quy trình thu thập dữ liệu thay đổi (CDC)** nắm bắt các thay đổi được thực hiện tại các nguồn dữ liệu, sau đó áp dụng những thay đổi đó trong toàn bộ kho lưu trữ dữ liệu của doanh nghiệp để giữ cho dữ liệu được đồng bộ hóa.

**Các quy trình chống trùng lặp dữ liệu** loại bỏ các bản sao và chuẩn hóa các định dạng dữ liệu.

CDC giảm thiểu các tài nguyên cần thiết cho các quy trình ETL bằng cách chỉ xử lý các thay đổi dữ liệu. Chống trùng lặp dữ liệu giúp giảm thiểu dung lượng lưu trữ và đồng bộ hóa dữ liệu.

### Xây dựng và sử dụng kho dữ liệu

Hình 3.13 Mô hình mô tả một kiến trúc điển hình bao gồm cơ sở dữ liệu, kho dữ liệu và siêu thị dữ liệu tạo dữ liệu sẵn sàng để người dùng truy cập và phân tích. Dữ liệu của tổ chức từ các hệ thống quy trình giao dịch vận hành được lưu trữ trong cơ sở dữ liệu vận hành (phía bên trái của hình). Không phải tất cả dữ liệu đều được chuyển đến kho dữ liệu. Thông thường, chỉ dữ liệu tổng hợp được chuyển. Kho dữ liệu tổ chức dữ liệu theo nhiều cách—theo chủ đề, khu vực chức năng, nhà cung cấp và sản phẩm. Như trong hình, kiến trúc kho dữ liệu xác định luồng dữ liệu bắt đầu khi dữ liệu được thu thập bởi các hệ thống giao dịch; dữ liệu nguồn được lưu trữ trong cơ sở dữ liệu giao dịch (hoạt động); Các quy trình ETL di chuyển dữ liệu từ cơ sở dữ liệu vào kho dữ liệu hoặc siêu thị dữ liệu, nơi dữ liệu có sẵn để doanh nghiệp truy cập, báo cáo và phân tích thêm bằng các công cụ thông minh (BI) và các công cụ phân tích dữ liệu nâng cao hơn (được thảo luận chi tiết trong Chương 6 ).



systems;

the

source

data

are

stored

in

transactional

(operational)

databases;EfL

processes

move

data

from

databases

into

intelligence

HÌNH 3.13 Cơ sở dữ liệu, kho dữ liệu và siêu thị dữ liệu cung cấp dữ liệu cho người dùng truy cập và phân tích.

Một cải tiến kho dữ liệu gần đây là kho dữ liệu đám mây ( CDW ) cloud data warehouse. CDW là kho dữ liệu hiệu suất cao được xây dựng riêng cho đám mây. CDW làm cho việc thu thập dữ liệu trở nên dễ dàng và không giống như các kho dữ liệu truyền thống, chúng cho phép phân tích nhanh chóng và mở ra những lợi thế của việc có một kho lưu trữ dữ liệu dựa trên đám mây. Sự sắp xếp hợp lý các bảng của chúng trong cơ sở dữ liệu đa chiều giống như hình dạng của một bông tuyết—do đó có tên như vậy.

Lưu trữ dữ liệu tập trung trên đám mây giúp loại bỏ sự phức tạp liên quan đến việc tìm kiếm các kho lưu trữ dữ liệu khác nhau để xác định vị trí các bản ghi riêng lẻ và giúp người quản lý tuân thủ các cập nhật, thay đổi và xóa bắt buộc dễ dàng hơn nhiều. CDW đang nhanh chóng thay thế các giải pháp kho dữ liệu tại chỗ và kết hợp vì chúng mang lại khả năng mở rộng nhanh chóng, tăng tính linh hoạt, chi phí thấp hơn, khả năng kết nối tốt hơn cũng như mức độ bảo mật và mã hóa dữ liệu cao. Ví dụ về CDW bao gồm Snowflake, Oracle Autonomous Database và Kintone.

**Kho dữ liệu đám mây (CDW)** là một cơ sở dữ liệu được phân phối trong đám mây công cộng dưới dạng dịch vụ được quản lý, được tối ưu hóa để phân tích, mở rộng quy mô và dễ sử dụng.

#### Hỗ trợ theo thời gian thực từ Kho dữ liệu đang hoạt động

Công nghệ kho dữ liệu ban đầu chủ yếu hỗ trợ các ứng dụng chiến lược không yêu cầu thời gian phản hồi tức thời, tương tác trực tiếp với khách hàng hoặc tích hợp với các hệ thống vận hành. ETL có thể đã được thực hiện một lần mỗi tuần hoặc một lần mỗi tháng. Tuy nhiên, nhu cầu về thông tin để hỗ trợ các hoạt động và tương tác với khách hàng theo thời gian thực dẫn đến việc phân tích và lưu trữ dữ liệu theo thời gian thực—được gọi là kho dữ liệu đang hoạt động ( ADW ) active data warehouse (ADW). Sự gia tăng mạnh mẽ về sức mạnh tính toán, tốc độ xử lý và bộ nhớ đã giúp ADW trở nên khả thi. ADW không được thiết kế để hỗ trợ việc ra quyết định chiến lược của các nhà điều hành, mà là để hỗ trợ các hoạt động.

**Kho dữ liệu hoạt động (ADW)** là khả năng kỹ thuật để nắm bắt các giao dịch khi chúng thay đổi và tích hợp chúng vào kho cùng với việc duy trì bath hoặc làm mới chu kỳ theo lịch trình.

Ví dụ: các công ty vận chuyển như DHL sử dụng đội xe tải khổng lồ để vận chuyển hàng triệu kiện hàng. Mỗi ngày và cả ngày, các nhà quản lý vận hành đưa ra hàng nghìn quyết định ảnh hưởng đến điểm mấu chốt, chẳng hạn như: "Chúng ta có cần bốn xe tải cho lần chạy này không?" "Với hai tài xế bị trì hoãn do thời tiết xấu, chúng tôi có cần gọi thêm trợ giúp không?" Kho dữ liệu truyền thống không phù hợp để hỗ trợ vận hành ngay lập tức, nhưng ADW thì phù hợp. Ví dụ: các công ty có ADW có thể:

* Tương tác với khách hàng để cung cấp dịch vụ khách hàng vượt trội.
* Đáp ứng các sự kiện kinh doanh trong thời gian gần nhất,

Chia sẻ dữ liệu trạng thái cập nhật giữa các thương nhân, nhà cung cấp, khách hàng và cộng sự.

Hai ví dụ về cách các công ty trong ngành bán lẻ, khách sạn và du lịch được hưởng lợi từ việc sử dụng ADW được mô tả trong CNTT tại nơi làm việc 3.2 .

## CNTT tại nơi làm việc 3.2

### Các công ty sử dụng **Active Data Warehousing (ADW)** để cải thiện trải nghiệm của khách hàng



**Bán lẻ: chuỗi siêu thị tạp hóa tư nhân hàng đầu** điều hành hơn 300 cửa hàng ở ba tiểu bang trên khắp miền đông nam Hoa Kỳ muốn có thông tin chi tiết theo thời gian thực bằng cách triển khai tích hợp liên tục giữa các hệ thống điểm bán hàng khác nhau trên các cửa hàng của mình vào một ADW. Trong nhiều năm, chuỗi cửa hàng tạp hóa đã sử dụng quy trình xử lý hàng loạt dữ liệu POS từ 300 cửa hàng của mình vào kho dữ liệu của công ty hàng đêm. Tuy nhiên, khối lượng dữ liệu ngày càng cao của nó từ các giao dịch của 5.000 mặt hàng tại 300 cửa hàng lên tới khoảng 40—50 triệu giao dịch mỗi ngày: Khối lượng giao dịch cao này dẫn đến sự chậm trễ lớn trong báo cáo, phân tích và ra quyết định, thường kéo dài sang ngày tiếp theo. Điều này khiến việc tải lên các giao dịch từ những ngày tiếp theo làm tăng thêm độ phức tạp cho quy trình. Tải dữ liệu thường không thành công và người dùng phải đối mặt với các báo cáo và phân tích không đáng tin cậy. Bất chấp các vấn đề về dữ liệu, chuỗi cửa hàng tạp hóa có lòng trung thành mạnh mẽ của khách hàng cũng như khả năng vận hành và bán hàng hiệu quả của nhân viên bán hàng và quản lý cửa hàng, nhưng điều này đang bị hủy hoại do sự chậm trễ trong báo cáo và phân tích và khiến chuỗi không thể phản ứng kịp thời và hiệu quả trước sự thay đổi của điều kiện thị trường. Giải pháp là một phần mềm của TIBCO và Teradata cùng nhau cung cấp một ADW đáng tin cậy, có khả năng chịu lỗi và cân bằng tải. Giải pháp đưa ra ba cách khác nhau để ngăn chặn lỗi ở các cấp độ khác nhau nhằm cung cấp dữ liệu hiệu quả cho ADW. ADW cung cấp cho người dùng doanh nghiệp một số lợi ích, bao gồm: tạo dữ liệu bán hàng trong vòng 60 phút; tinh chỉnh hàng tồn kho theo thời gian thực tại các cửa hàng và kho hàng; và cung cấp cho người quản lý cửa hàng chế độ xem dữ liệu bán hàng theo thời gian thực và cho phép họ tạo các sự kiện linh hoạt dựa trên dữ liệu thời gian thực.

**Khách sạn và Du lịch:** Travelocity (https:/fwww.travelocity.com/) sử dụng ADW để tìm các giao dịch du lịch tốt nhất dành riêng cho bạn. Mục tiêu là sử dụng "dữ liệu của ngày hôm nay" thay vì "dữ liệu của ngày hôm qua". ADW của công ty du lịch trực tuyến phân tích lịch sử tìm kiếm của bạn và các điểm đến quan tâm; sau đó dự đoán các ưu đãi du lịch bạn rất có thể sẽ mua. Ưu đãi vừa phù hợp vừa kịp thời để nâng cao trải nghiệm của bạn, giúp chốt giao dịch bán hàng trong một thị trường rất cạnh tranh. Ví dụ: khi một khách hàng đang tìm kiếm các chuyến bay và khách sạn ở Las Vegas, Travelocity nhận ra mối quan tâm-—khách hàng muốn đến Vegas. ADW tìm kiếm các chuyến bay có giá tốt nhất từ tất cả các hãng, xây dựng một vài giao dịch trọn gói và giới thiệu chúng theo thời gian thực cho khách hàng.

Khi khách hàng nhìn thấy ưu đãi được cá nhân hóa mà họ thấy thực sự thú vị, ADW sẽ giúp tạo ra trải nghiệm khách hàng tốt hơn. Trải nghiệm dựa trên dữ liệu thời gian thực làm tăng tỷ lệ chuyển đổi và doanh số bán hàng. (nguồn … )

Nội dung kho dữ liệu có thể được gửi tới những người ra quyết định trong toàn doanh nghiệp thông qua đám mây hoặc mạng nội bộ do công ty sở hữu. Người dùng có thể xem, truy vấn và phân tích dữ liệu và tạo báo cáo bằng trình duyệt Web. Đây là những phương pháp cung cấp dữ liệu cực kỳ tiết kiệm và hiệu quả.

#### **Data Warehousing Supports Action as Well as Decisions** Kho dữ liệu hỗ trợ hoạt động và ra quyết định

#### Nhiều tổ chức đã xây dựng kho dữ liệu vì họ thất vọng với dữ liệu không nhất quán nên không thể hỗ trợ các quyết định hoặc hành động. Nhìn từ góc độ này, kho dữ liệu là các khoản đầu tư cơ sở hạ tầng mà các công ty trong nhiều ngành khác nhau thực hiện để hỗ trợ các hoạt động đang diễn ra và trong tương lai, bao gồm:

* **Tiếp thị** Thông báo cho mọi người về tình trạng của sản phẩm, hiệu quả của chương trình tiếp thị và lợi nhuận của dòng sản phẩm; và cho phép họ thực hiện để tối đa hóa lợi nhuận trên mỗi khách hàng
* **Định giá và hợp đồng** Tính toán chi phí chính xác để tối ưu hóa việc định giá hợp đồng. Nếu không có dữ liệu chi phí chính xác, giá có thể thấp hơn hoặc quá gần với chi phí; hoặc giá có thể không cạnh tranh vì chúng quá cao.
* **Dự báo** Ước tính nhu cầu của khách hàng đối với sản phẩm và dịch vụ.

• **Bán hàng** Tính toán lợi nhuận bán hàng và năng suất cho tất cả các lãnh thổ và khu vực; phân tích kết quả theo địa lý, sản phẩm, nhóm bán hàng hoặc cá nhân.

* **Tài chính** Cung cấp dữ liệu thời gian thực cho các điều khoản tín dụng tối ưu, phân tích danh mục đầu tư và các hoạt dộng giúp giảm rủi ro hoặc chi phí nợ khó đòi.

Kho dữ liệu tập hợp dữ liệu từ các nguồn riêng biệt và cơ sở dữ liệu khác nhau trên toàn bộ doanh nghiệp được gọi là kho dữ liệu doanh nghiệp enterprise data warehouses (EDWs)

**Kho dữ liệu doanh nghiệp (EDW)** là kho dữ liệu tích hợp dữ liệu từ nhiều cơ sở dữ liệu khác nhau trong toàn bộ doanh nghiệp.

Một trong những động lực chính của sự thay đổi trong thị trường quản lý dữ liệu là lượng dữ liệu được quản lý ngày càng tăng. Các doanh nghiệp cần các giải pháp kho dữ liệu và quản trị cơ sở dữ liệu mạnh mẽ, phân tích và tạo báo cáo. Các nhà cung cấp thống trị thị trường này đáp ứng nhu cầu quản lý dữ liệu đang phát triển với phần mềm và phần cứng thông minh và tiên tiến hơn. Công nghệ phần cứng tiên tiến cho phép mở rộng khối lượng dữ liệu và khối lượng công việc cao hơn nhiều so với trước đây có thể, hoặc nó có thể xử lý khối lượng công việc cụ thể. Các hệ quản trị cơ sở dữ liệu quan hệ có mục đích chung cũ hơn thiếu khả năng mở rộng hoặc tính linh hoạt đối với các khối lượng công việc chuyên biệt hoặc rất lớn nhưng lại rất giỏi trong những gì chúng làm.

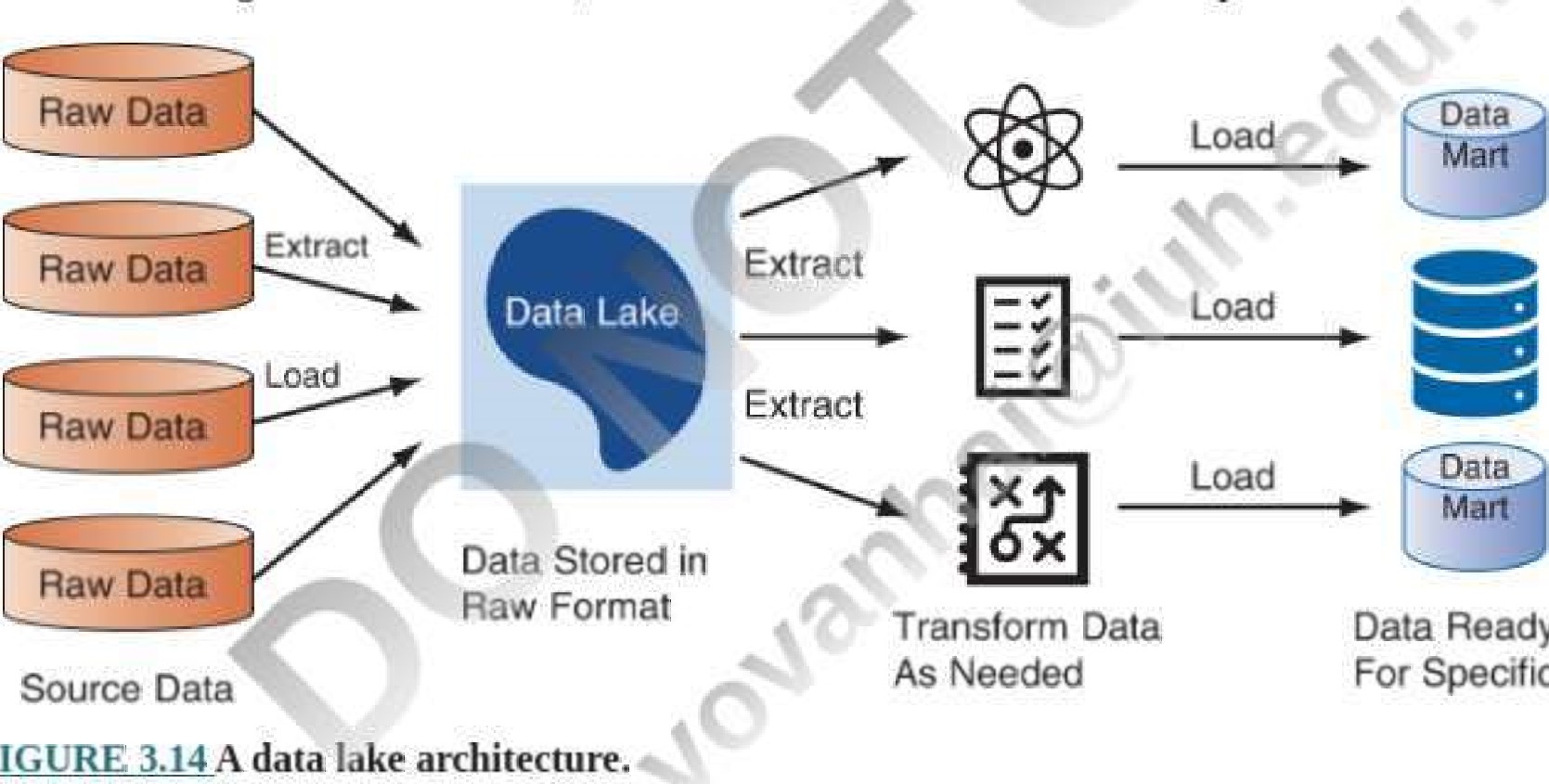
Bảng 3.1 tóm tắt một số ứng dụng thành công của kho dữ liệu.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | BẢNG 3.1 | | | |  |
| Ứng dụng kho dữ liệu theo ngành | | | |
| Ngành công nghiệp | | Các ứng dụng | |
| hãng hàng không | | Phân công phi hành đoàn, triển khai máy bay, phân tích lợi nhuận của tuyến đường và khuyến mãi khách hàng thân thiết | |
| Tài chính ngân hàng | | Dịch vụ khách hàng, phân tích xu hướng, khuyến mãi sản phẩm và dịch vụ, và giảm chi phí IS | |
| Thẻ tín dụng | | Dịch vụ khách hàng, dịch vụ thông tin mới có tính phí, phát hiện gian lận | |
| hợp đồng quốc phòng | | Chuyển giao công nghệ, sản xuất ứng dụng quân sự | |
| E-business | | Kho dữ liệu có năng lực cá nhân hóa, tùy chọn marketing/shopping cho phép bán thêm và bán chéo | |
| Chính phủ | | Báo cáo về lĩnh vực tội phạm, an ninh quốc phòng | |
| Chăm sóc sức khỏe | | Giảm chi phí hoạt động | |
| Đầu tư và bảo hiểm | | Quản lý rủi ro, phân tích biến động thị trường, phân tích xu hướng khách hàng và quản lý danh mục đầu tư | |
| Chuỗi bán lẻ | | Phân tích xu hướng, phân tích mô hình mua hàng, chính sách giá, kiểm soát hàng tồn kho, khuyến mãi bán hàng và quyết định kênh phân phối tối ưu | |

### Hồ dữ liệu Data Lakes

Một nhược điểm lớn của kho dữ liệu là cần phải mô hình hóa dữ liệu trước khi lưu trữ. Sử dụng hồ dữ liệu có thể lưu trữ số lượng lớn dữ liệu thô có cấu trúc, bán cấu trúc và không cấu trúc ở trạng thái tự nhiên của chúng, hoạt động chuyển đổi của quy trình ETL có thể tránh được và không được thực hiện cho đến khi các yêu cầu đối với dữ liệu được xác định là cần thiết (Hình 3.14 ). Hồ dữ liệu cũng có thể bao gồm các bản sao thô của dữ liệu hệ thống nguồn, chẳng hạn như bài đăng trên blog, đánh giá sản phẩm, dữ liệu khách hàng và bán sản phẩm, cũng như dữ liệu được chuyển đổi để dùng cho các tác vụ chẳng hạn như báo cáo, trực quan hóa, phân tích nâng cao và học máy. Tính linh hoạt của các hồ dữ liệu đặc biệt hấp dẫn đối với các nhà phân tích dữ liệu và nhà khoa học dữ liệu, những người có thể dễ dàng định cấu hình và định lại cấu hình cho các mô hình, truy vấn và ứng dụng một cách nhanh chóng. Hai ví dụ về phần mềm hồ dữ liệu bao gồm Microsoft Azure Data Lake và IBM Data Lake.

**Hồ dữ liệu** **Data lake** là một nơi lưu trữ duy nhất toàn bộ dữ liệu của doanh nghiệp ở định dạng tự nhiên ( dữ liệu có cấu trúc, bán cấu trúc và phi cấu trúc).



Ready

Specific

Purpose

FIGURE

Questions

Warehouse

I. Sự khác biệt giữa cơ sở dữ liệu và kho dữ liệu là gì?

1. Sự khác biệt giữa kho dữ liệu và siêu thị dữ liệu là gì?
2. Giải thích ETL.

4, Giải thích CDC.

5, Lợi thế của kho dữ liệu doanh nghiệp (EDW) là gì?

1. Tại sao một công ty có thể đầu tư vào một siêu thị dữ liệu thay vì kho dữ liệu?
2. Những cấp độ nào của tổ chức được hưởng lợi nhiều nhất từ kho dữ liệu?
3. Hồ dữ liệu khác với kho dữ liệu như thế nào?

### 3.3 Quản trị dữ liệu và Quản trị dữ liệu Master (MDM)

LO 3.3 Mô tả tầm quan trọng của quản trị dữ liệu trong việc cung cấp dữ liệu đáng tin cậy có sẵn khi nào và ở đâu cần thiết và tại sao quản lý dữ liệu chủ (MDM) là một sáng kiến quản trị dữ liệu quan trọng.

Quản trị dữ liệu mạnh mẽ và MDM là cần thiết để đảm bảo dữ liệu có đủ chất lượng để đáp ứng nhu cầu kinh doanh.

#### Quản trị dữ liệu Data Governance

Một chương trình quản trị dữ liệu hợp lý bao gồm bộ phận quản lý, một bộ các thủ tục và và kế hoạch thực hiện các thủ tục đó.

Sự thành công của mọi chiến lược hoặc nỗ lực tiếp thị dựa trên dữ liệu phụ thuộc vào việc quản trị dữ liệu hiệu quả để có được các chính sách và thủ tục chính thức được phát triển cao để kiểm soát dữ liệu doanh nghiệp. Các chính sách quản trị dữ liệu ( Data governance policies) phải giải quyết dữ liệu có cấu trúc, bán cấu trúc và phi cấu trúc để đảm bảo rằng nhân viên và đối tác kinh doanh được cung cấp dữ liệu chất lượng cao mà họ có thể tin cậy và truy cập theo yêu cầu.

**Quản trị dữ liệu** là quản lý tổng thể tính khả dụng, khả năng sử dụng, toàn vẹn và bảo mật của dữ liệu được sử dụng trong doanh nghiệp,

Điều này cực kỳ quan trọng trong các ngành dịch vụ. Ví dụ, quản trị dữ liệu mạnh mẽ là cần thiết để tuân thủ các yêu cầu nghiêm ngặt về trách nhiệm giải trình của chính phủ đối với lĩnh vực chăm sóc sức khỏe. Các chương trình quản trị dữ liệu, đôi khi được gọi là quản trị dữ liệu doanh nghiệp, xác minh rằng dữ liệu nhập vào hồ sơ sức khỏe điện tử và hệ thống lâm sàng, tài chính và vận hành là chính xác và đầy đủ, đồng thời chỉ những chỉnh sửa được ủy quyền mới có thể được thực hiện và ghi lại.

Với một chương trình quản trị dữ liệu hiệu quả, các nhà quản lý có thể xác định dữ liệu của họ đến từ đâu, ai sở hữu chúng và ai chịu trách nhiệm về cái gì. Quản trị dữ liệu là một dự án toàn doanh nghiệp vì dữ liệu vượt qua các ranh giới và được mọi người trong toàn doanh nghiệp sử dụng. Các quy định mới và áp lực giảm chi phí đã làm tăng tầm quan trọng của việc quản trị dữ liệu hiệu quả. Quản trị loại bỏ chi phí duy trì và lưu trữ dữ liệu xấu, không cần thiết hoặc không chính xác. Những chi phí này tăng lên khi khối lượng dữ liệu tăng lên. Quản trị cũng làm giảm rủi ro pháp lý liên quan đến thông tin không được quản lý hoặc quản lý không nhất quán.

Ba ngành phụ thuộc nhiều vào quản trị dữ liệu để tuân thủ các yêu cầu hoặc các yêu cầu báo cáo là:



• **Ngành thực phẩm** Trong ngành thực phẩm, việc quản trị dữ liệu bắt buộc phải tuân thủ các quy định về an toàn thực phẩm. Các nhà sản xuất và bán lẻ thực phẩm có các hệ thống kiểm soát phức tạp để nếu phát hiện ra một sản phẩm thực phẩm bị ô nhiễm, chẳng hạn như rau bina hoặc bơ đậu phộng, họ có thể truy nguyên vấn đề trở lại một nhà máy chế biến cụ thể hoặc thậm chí là hộ gia đình ở đầu chuỗi thực phẩm.

• **Ngành dịch vụ tài chính** Trong lĩnh vực dịch vụ tài chính, các yêu cầu báo cáo nghiêm ngặt của Đạo luật Dodd-Frank đang dẫn đến việc sử dụng quản trị dữ liệu nhiều hơn. Đạo luật Dodd-Frank là một phần quan trọng của quy định bảo vệ người tiêu dùng đã được thông qua để đảm bảo tính minh bạch và trách nhiệm giải trình nhằm đối phó với cuộc khủng hoảng tài chính năm 2008. Mặc dù một số quy định của nó đã được rút lại, nhưng ngày nay người tiêu dùng vẫn được hưởng lợi từ nhiều biện pháp bảo vệ mà Đạo luật đưa ra, bao gồm cả việc thành lập Cục Bảo vệ Tài chính Người tiêu dùng (CFPB).

• **Ngành chăm sóc sức khỏe** Dữ liệu là tài sản quý giá nhất của ngành chăm sóc sức khỏe. Các bệnh viện có hàng núi dữ liệu bệnh nhân điện tử. Nghĩa vụ báo cáo và trách nhiệm giải trình chăm sóc sức khỏe mới yêu cầu các mô hình quản trị dữ liệu minh bạch để chống gian lận và bảo vệ thông tin của bệnh nhân.

Khi phát triển một chương trình quản trị dữ liệu, điều quan trọng là phải trình bày một chức năng (business case) mạnh mẽ để nhận được sự ủng hộ từ các giám đốc điều hành hàng đầu và các bên liên quan. Một phần quan trọng của trường hợp kinh doanh là ước tính lợi tức đầu tư (ROI) của chương trình quản trị dữ liệu để chỉ ra sẽ gia tăng giá trị cho công ty. Bạn sẽ cần chứng minh ROI dựa trên cả chiến lược kinh doanh và CNTT để đảm bảo rằng các quỹ khả dụng được sử dụng để đáp ứng tốt nhất các mục tiêu kinh doanh.

Để làm được điều này, bạn sẽ cần phân tích kỹ hạ tầng CNTT về cách các thành phần khác nhau của cơ sở hạ tầng CNTT phối hợp với nhau để hỗ trợ các quy trình kinh doanh, cách hệ thống này có thể nhận và sử dụng dữ liệu cần thiết cũng như dữ liệu có thể được truyền đạt và/hoặc dễ dàng như thế nào tái sử dụng. Bạn cũng sẽ cần tính đến rủi ro và các sự kiện bất lợi, chẳng hạn như chi phí liên quan đến việc thu thập lại dữ liệu, chi phí liên quan đến dữ liệu không đáng tin cậy hoặc không phù hợp,và sự chậm trễ liên quan đến dữ liệu không kịp thời hoặc không có sẵn. Tiếp theo, các chi phí này phải được định lượng và mức độ tin cậy của bạn đối với dữ liệu của công ty phải được tính toán để đảm bảo trường hợp kinh doanh của bạn phản ánh chính xác giá trị của chương trình quản trị dữ liệu. Tech Note 3.3 giải thích cách đo lường giá trị của quản trị dữ liệu để đánh giá sức mạnh của nó.

|  |
| --- |
| Ghi chú công nghệ 3.3  Đo lường chính xác giá trị của quản trị dữ liệu  Một chỉ số được sử dụng để thực hiện phép tính này là *độ tin cậy đối với số liệu giả định phụ thuộc vào dữ liệu* (CIDDA) (Reeves & Bowen, 2013 ). CIDDA xác định các lĩnh vực thiếu hụt cụ thể.  Vì vậy, tóm lại, khi xây dựng mô hình quản trị dữ liệu, cần phải:     1. Thành lập đội ngũ lãnh đạo 2. Xác định phạm vi của chương trình 3. Tính toán ROI bằng cách sử dụng CIDDA.   CIDDA được tính bằng cách nhân ba ước tính độ tin cậy bằng công thức sau:  CIDDA = G x M x TS  Với  G : độ Tin tưởng rằng dữ liệu đủ tốt cho mục đích dự định của họ  M : độ Tin tưởng rằng dữ liệu có ý nghĩa như bạn nghĩ họ làm  TS : đột tin tưởng rằng bạn biết dữ liệu đến từ đâu và tin tưởng vào nguồn.  CIDDA là một số liệu chủ quan không có tiêu chuẩn ngành, nhưng nó có thể được đánh giá theo thời gian để đánh giá sự cải thiện về độ tin cậy của chất lượng dữ liệu,  Để đảm bảo bạn hiểu rõ về mục Hộp công cụ CNTT này, hãy tính toán CIDDA của Công ty A theo thời gian , sử dụng các mức độ tin cậy đã nêu trong các khía cạnh khác nhau của dữ liệu công ty trong thời gian QI—Q4 2020: |

Tầm quan trọng của một chương trình quản trị dữ liệu mạnh có thể được đánh giá cao nhất bằng cách xem xét các đặc điểm và hậu quả của việc quản trị dữ liệu yếu hoặc không tồn tại được liệt kê trong Bảng 3.2 .

|  |
| --- |
|  |
| Bảng 3.2 |
| Đặc điểm và hậu quả của quản trị dữ liệu yếu hoặc không tồn tại   * Dữ liệu trùng lặp khiến cho các silo dữ liệu bị cô lập. * Tồn tại sự không nhất quán về ý nghĩa và mức độ chi tiết của các yếu tố dữ liệu, * Người dùng không tin tưởng vào dữ liệu và lãng phí thời gian để xác minh dữ liệu thay vì phân tích chúng để đưa ra quyết định phù hợp. * Dẫn đến việc phân tích dữ liệu không chính xác.   Những quyết định sai lầm được đưa ra dựa trên nhận thức hơn là thực tế, điều này có thể ảnh hưởng tiêu cực đến công ty và khách hàng.   * Dẫn đến tăng khối lượng công việc và thời gian xử lý, |

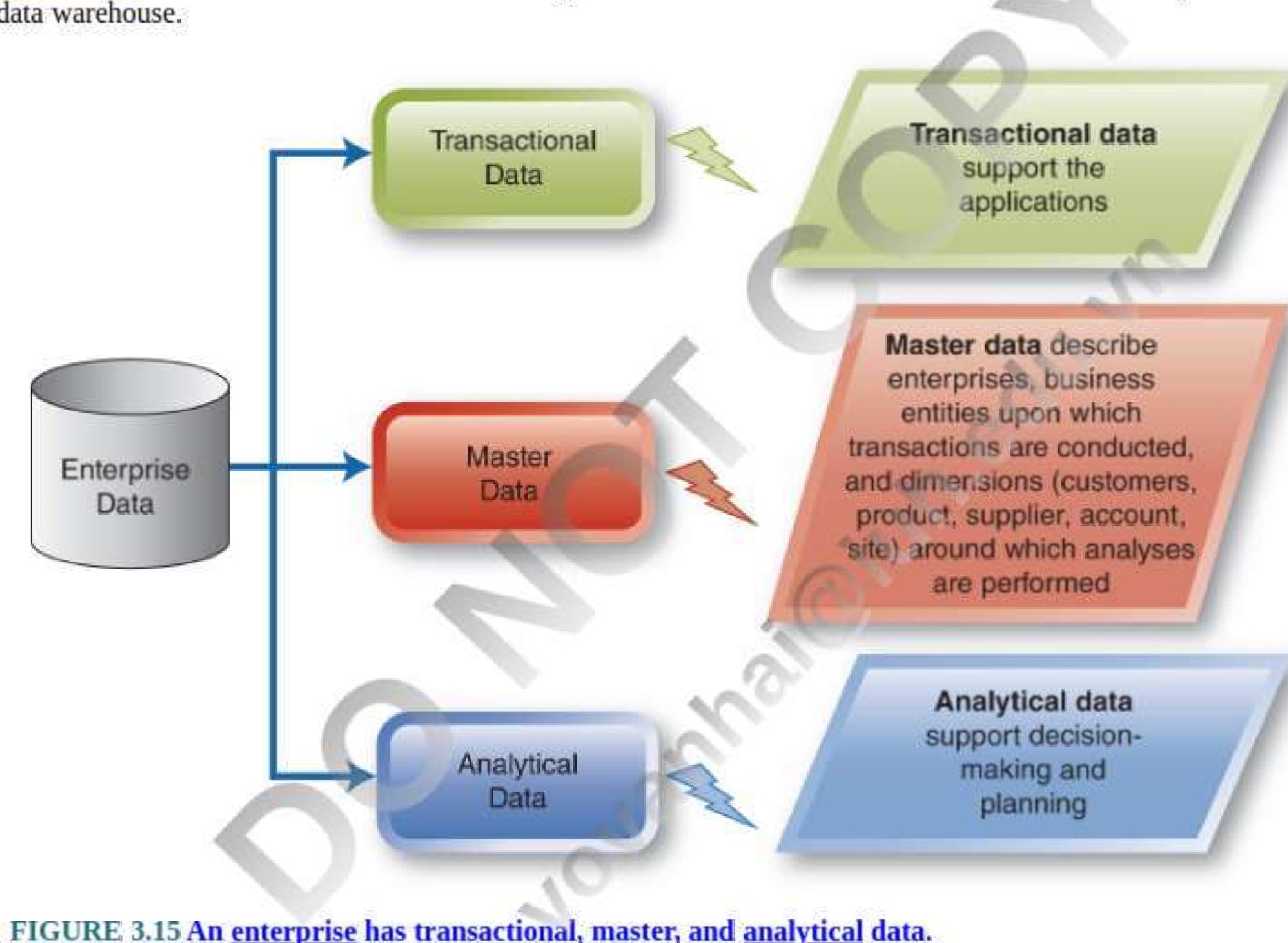
#### Dữ liệu chính và quản lý dữ liệu chính (MDM)

#### Master Data and Master Data Management

Khi dữ liệu trở nên phức tạp hơn và khối lượng của chúng bùng nổ, hiệu suất cơ sở dữ liệu giảm xuống và các tổ chức đang điều chỉnh để tối ưu hóa dữ liệu chính **master data**. Dữ liệu chính về cơ bản khác với khối lượng lớn, tốc độ và sự đa dạng của **dữ liệu lớn** và dữ liệu truyền thống. Ví dụ: khi khách hàng đăng ký bảo hiểm ô tô, dữ liệu được cung cấp trên ứng dụng sẽ trở thành dữ liệu chính cho khách hàng đó. Ngược lại, nếu phương tiện của khách hàng có một thiết bị gửi dữ liệu về hành vi lái xe của họ cho công ty bảo hiểm, thì những dữ liệu do máy tạo ra đó là dữ liệu giao dịch hoặc hoạt động. Các thực thể dữ liệu chính là các thực thể chính của một công ty, chẳng hạn như khách hàng, sản phẩm, nhà cung cấp, nhân viên và tài sản.

**Dữ liệu chính** Master data là thuật ngữ được sử dụng để mô tả thông tin kinh doanh quan trọng về khách hàng, sản phẩm và dịch vụ, nhà cung cấp, địa điểm, nhân viên và những thứ khác cần thiết cho hoạt động và giao dịch kinh doanh.

Ví dụ 3.15 chỉ ra cách dữ liệu chính đóng vai trò như một lớp giữa dữ liệu giao dịch trong cơ sở dữ liệu và dữ liệu phân tích trong cơ sở dữ liệu.



data

HÌNH 3.15 Một doanh nghiệp có dữ liệu giao dịch , dữ liệu chính , và dữ liệu phân tích

Nâng cao trải nghiệm của khách hàng là một ví dụ về ưu tiên kinh doanh cấp bách dựa trên việc tận dụng dữ liệu chính (O'Kane và Moran, 2016 ). Một chiến lược quản trị dữ liệu hiệu quả có thể giúp các tổ chức đạt được điều này được gọi là **quản lý dữ liệu chính** .

**Quản lý dữ liệu chính (MDM)** tích hợp dữ liệu từ nhiều nguồn khác nhau hoặc các ứng dụng của doanh nghiệp để tạo ra một cái nhìn (thống nhất) đầy đủ hơn về khách hàng, sản phẩm hoặc thực thể khác.

##### Tệp tham chiếu chính và các thực thể dữ liệu

##### Master Reference File and Data Entities

**Thực thể dữ liệu** là bất kỳ thứ gì có thực hoặc trừu tượng mà một công ty muốn thu thập và lưu trữ dữ liệu.

Dữ liệu chính thường khá ổn định và thường được lưu trữ trong một số hệ thống khác nhau trải khắp doanh nghiệp. Quản lý dữ liệu chính (MDM) liên kết và đồng bộ hóa tất cả dữ liệu quan trọng từ các hệ thống khác nhau đó vào một tệp được gọi là tệp chính ( a **master file)** để cung cấp một điểm tham chiếu chung. Khi MDM hợp nhất dữ liệu từ nhiều nguồn dữ liệu khác nhau vào tệp hoặc chỉ mục tham chiếu chính này, thì MDM sẽ cung cấp dữ liệu trở lại ứng dụng, do đó tạo ra dữ liệu chính xác và nhất quán trong toàn doanh nghiệp.

**Tệp chính** Master file : là tập hợp các bản ghi mô tả một trong các thực thể chính trong cơ sở dữ liệu, chẳng hạn như khách hàng, sản phẩm, nhân viên và nhà cung cấp. Nó thường được cập nhật định kỳ.

Các giải pháp MDM có thể phức tạp và tốn kém. Với sư phức tạp và chi phí của chúng, hầu hết các giải pháp MDM đều nằm ngoài tầm với của các công ty vừa và nhỏ. Các nhà cung cấp đã giải quyết thách thức này bằng cách cung cấp các dịch vụ MDM do đám mây quản lý. Ví dụ: Phần mềm Dell cung cấp Dell Boomi MDM thế hệ tiếp theo. Boomi cung cấp các dịch vụ MDM, quản lý dữ liệu và các dịch vụ chất lượng dữ liệu (DQS) 100% dựa trên đám mây với khả năng đồng bộ hóa gần như theo thời gian thực (<https://boomi.com/content/video/demo/l0min-demo/> ).

Mỗi bộ phận có nhu cầu dữ liệu chủ riêng biệt. Ví dụ, tiếp thị liên quan đến giá cả sản phẩm, thương hiệu và bao bì sản phẩm, trong khi sản xuất liên quan đến chi phí và lịch trình sản phẩm. Tệp tham chiếu chính của khách hàng có thể cung cấp dữ liệu cho tất cả các hệ thống doanh nghiệp có thành phần quan hệ khách hàng, do đó cung cấp một bức tranh thống nhất hơn về khách hàng. Tương tự, tệp tham chiếu chính của sản phẩm có thể cung cấp dữ liệu cho tất cả các hệ thống sản xuất trong doanh nghiệp.

MDM bao gồm các công cụ để làm sạch và kiểm tra các thành phần dữ liệu chính cũng như các công cụ để tích hợp và đồng bộ hóa dữ liệu để giúp chúng dễ truy cập hơn. MDM cung cấp một giải pháp cho các nhà quản lý đang thất vọng bởi các nguồn dữ liệu bị phân tán và phân tán. CNTT tại nơi làm việc 3.3 minh họa những thách thức quản lý dữ liệu chính mà những người tham gia chuỗi cung ứng chăm sóc sức khỏe phải đối mặt trong việc phát triển các tiêu chuẩn để tạo **danh sách chính** chính xác và đáng tin cậy của các thực thể dữ liệu chính.

|  |
| --- |
| CNTT tại nơi làm việc 3.3  Chuẩn hóa dữ liệu chính là một thách thức trong chăm sóc sức khỏe  Hậu cần chăm sóc sức khỏe rất phức tạp và phụ thuộc vào dữ liệu chất lượng cao, đáng tin cậy. Thật không may, dữ liệu chính được tiêu chuẩn hóa, chính xác là một vấn đề. Tệp tham chiếu chính chính xác của tất cả các sản phẩm và dịch vụ mà hệ thống chăm sóc sức khỏe mua (**item master**) là chìa khóa để cho phép theo dõi và truy tìm đơn đặt hàng, các sáng kiến tìm nguồn cung ứng chiến lược. Vật tư, dịch vụ đã mua và thiết bị vốn chiếm khoảng một phần ba chi phí hoạt động của bệnh viện. Người ta ước tính rằng các **item master** của một bệnh viện dao động từ 18.000 đến 200.000 vật phẩm. Đáng buồn thay, người ta ước tính rằng khoảng 30% đến 40% dữ liệu về item master là không chính xác, tạo ra sự bất đồng trong chuỗi cung ứng và hệ thống y tế.  Ví dụ về sự lãng phí do item master không chính xác bao gồm đặt hàng ngoài hợp đồng, đặt hàng không chính xác, thời gian xử lý đơn hàng kéo dài, không thể nắm bắt các mức cam kết để định giá hợp đồng và chậm hoàn trả do thiếu dữ liệu. Một số chuyên gia Ước tính Tiết kiệm 0,5 % có thể đạt được bằng cách dọn dẹp tổng thể item master. Ngoài ra, đây là khoản đầu tư cao vào việc duy trì item master của họ để cố gắng giữ cho nó chính xác nhất có thể. Việc quản lý dữ liệu không chính xác được ước tính sẽ lãng phí 20% thời gian của người mua dịch vụ chăm sóc sức khỏe hiện tại để khắc phục sự cố và khắc phục sự cố thay vì hỗ trợ chiến lược cho chức năng mua hàng. Một ví dụ nhỏ là sáng kiến tiết kiệm chi phí do Kurt Knothy, VP, SupplyChain tại Specttrum Health bắt đầu, theo đó ông đã giảm 30 loại nước rửa tay khô khác nhau được sử dụng trong suốt quá trình chăm sóc sức khỏe. hệ thống xuống còn ba. Điều này dẫn đến tiết kiệm được 30.000 đô la mỗi năm.  Điều cần thiết là các tiêu chuẩn dữ liệu toàn ngành sẽ đảm bảo dữ liệu giá cả và sản phẩm chính xác \_ được phân loại chính xác cùng với mã vạch/thẻ RFID mà hệ thống có thể sử dụng để gắn thẻ các mặt hàng cho đến đơn vị sử dụng nhỏ nhất. Hiện tại, FDA đang hỗ trợ các tiêu chuẩn UDI thông qua Hiệp hội Quản lý Tài nguyên và Vật liệu Y tế (AHRMM), nhóm thành viên hàng đầu dành cho các chuyên gia chuỗi cung ứng chăm sóc sức khỏe và GSI, một tổ chức thiết kế và thực hiện các tiêu chuẩn toàn cầu cho giao tiếp kinh doanh, chẳng hạn , mã vạch, đang giúp ngành chăm sóc sức khỏe tập trung vào việc phát triển các tiêu chuẩn dữ liệu của ngành.    Mục tiêu là có một chiếc xe tải quay trở lại bến nhận hàng tại một cơ sở chăm sóc sức khỏe và người nhận hàng có thông tin chính xác về sản phẩm, bao bì và thông tin giao hàng dễ sử dụng cho nhóm theo dõi và theo dõi giúp họ giải phóng khỏi tập trung vào các sáng kiến quản lý dữ liệu master chiến lược để cung cấp dịch vụ chăm sóc bệnh nhân tốt nhất có thể và cứu mạng sống.    Nguồn: Tổng hợp từ Krieger (2018 ), Bagley ( 2019 ), Michigan State University ( 2019 ), ahrmm.org và GS1.org. |

Quản trị dữ liệu kết hợp với MDM một sự kết hợp mạnh mẽ. Khi nguồn và khối lượng dữ liệu tiếp tục tăng, do đó, nhu cầu quản lý dữ liệu như một tài sản chiến lược để trích xuất toàn bộ giá trị của nó. Làm cho dữ liệu kinh doanh nhất quán, đáng tin cậy và có thể truy cập trên toàn doanh nghiệp là bước quan trọng đầu tiên trong các mô hình kinh doanh lấy khách hàng làm trung tâm . Với quản trị dữ liệu, các công ty có thể trích xuất giá trị tối đa từ dữ liệu của họ bằng cách tận dụng tốt hơn các cơ hội ẩn chứa trong dữ liệu hành vi.

**Lấy khách hàng làm trung tâm** là một phương pháp kinh doanh tập trung vào việc cung cấp trải nghiệm tích cực cho khách hàng tại và sau điểm bán hàng để thúc đẩy lợi nhuận và đạt được lợi thế cạnh tranh.

#### Lợi ích của Quản trị dữ liệu và Quản lý dữ liệu chủ

Việc có các chính sách và quy trình phát triển cao và được cân nhắc kỹ lưỡng về quản lý dữ liệu và quản lý dữ liệu chủ có thể dẫn đến những lợi ích đáng kể trong một tổ chức, bao gồm:

1. **Cải thiện việc tuân thủ quy định** Các quy định của chính phủ và các yêu cầu tuân thủ đã tăng lên đáng kể trong thập kỷ qua. Các công ty không tuân thủ luật về quyền riêng tư, gian lận, chống rửa tiền, an ninh mạng, an toàn lao động, v.v. sẽ phải đối mặt với các hình phạt nghiêm khắc.
2. **Ra quyết định hiệu quả hơn** Tối đa hóa việc sử dụng dữ liệu để đưa ra quyết định nhanh hơn với độ chắc chắn cao hơn.
3. **Cải thiện hiểu biết về dữ liệu** Khi được triển khai tốt, quản trị dữ liệu sẽ cung cấp một cái nhìn toàn diện về tất cả các tài sản dữ liệu cùng với trách nhiệm giải trình cao hơn đối với dữ liệu cụ thể.

4, **Tăng doanh thu** Các quyết định tốt hơn, nhanh hơn được đưa ra với độ chắc chắn cao hơn dẫn đến các lỗi ít tốn kém hơn như bắt đầu sai và vi phạm dữ liệu, do đó, chi tiêu ít tiền hơn cho việc quản lý PR và thất thoát tài chính.

#### câu hỏi

1. Giải thích tại sao điều quan trọng là phải phát triển một chương trình quản trị dữ liệu hiệu quả?

* 2. Giải thích mục đích quản lý dữ liệu chủ.
* 3. Tại sao mối quan tâm về quản lý dữ liệu và MDM tăng lên ?
* 4. Hai cách mà dữ liệu được sử dụng trong kinh doanh là gì?
* 5. Ba lợi ích của quản trị dữ liệu là gì.

1. Tại sao sự quan tâm đến quản trị dữ liệu và MDM tăng lên?
2. Hai cách mà dữ liệu được sử dụng trong kinh doanh là gì?

3.4 Quản lý thông tin

LO 3.4 Giải thích khái niệm quản lý thông tin và lý do tại sao điều quan trọng đối với một tổ chức là tuân theo các thực hành quản lý thông tin tốt.

Hầu hết mọi người đều quản lý thông tin. Bạn quản lý các tài khoản tài chính, xã hội và đám mây của mình trên nhiều thiết bị di động và máy tính. Bạn cập nhật hoặc đồng bộ hóa ("đồng bộ") lịch, cuộc hẹn, danh sách liên hệ, các tệp media, tài liệu và báo cáo của mình. Các nhiệm vụ lập kế hoạch và ra quyết định hiệu quả như những nhiệm vụ này phụ thuộc vào việc các hệ thống có thể cung cấp dữ liệu ở các định dạng có thể sử dụng được một cách kịp thời. Năng suất của bạn cũng phụ thuộc vào khả năng tương thích của các thiết bị và ứng dụng cũng như khả năng chia sẻ dữ liệu của chúng. Việc không thể chuyển và đồng bộ hóa bất cứ khi nào bạn thêm thiết bị hoặc ứng dụng sẽ gây khó chịu và lãng phí thời gian. Ví dụ: khi chuyển sang thiết bị di động mới nhất, bạn có thể cần tổ chức lại nội dung để xử lý dữ liệu và thiết bị dễ dàng hơn. Để đơn giản hóa các tiện ích bổ sung, nâng cấp, chia sẻ và truy cập, ở cấp độ cá nhân, bạn có thể tận dụng các dịch vụ đám mây như iTunes (<https://www.apple.com/itunes/> ), trình phát đa phương tiện, thư viện đa phương tiện, đài phát thanh intemet và ứng dụng quản lý thiết bị di động hoặc Instagram (https://www.instagram.com), một dịch vụ mạng xã hội chia sẻ ảnh và video. Tại nơi làm việc, bạn có thể sử dụng Diigo (https://www.diigo.com/), một cộng đồng chia sẻ kiến thức và công cụ nghiên cứu mạnh mẽ và Box (https://www.box.com/), một công cụ quản lý nội dung và tệp trên đám mây. dịch vụ chia sẻ cho doanh nghiệp.

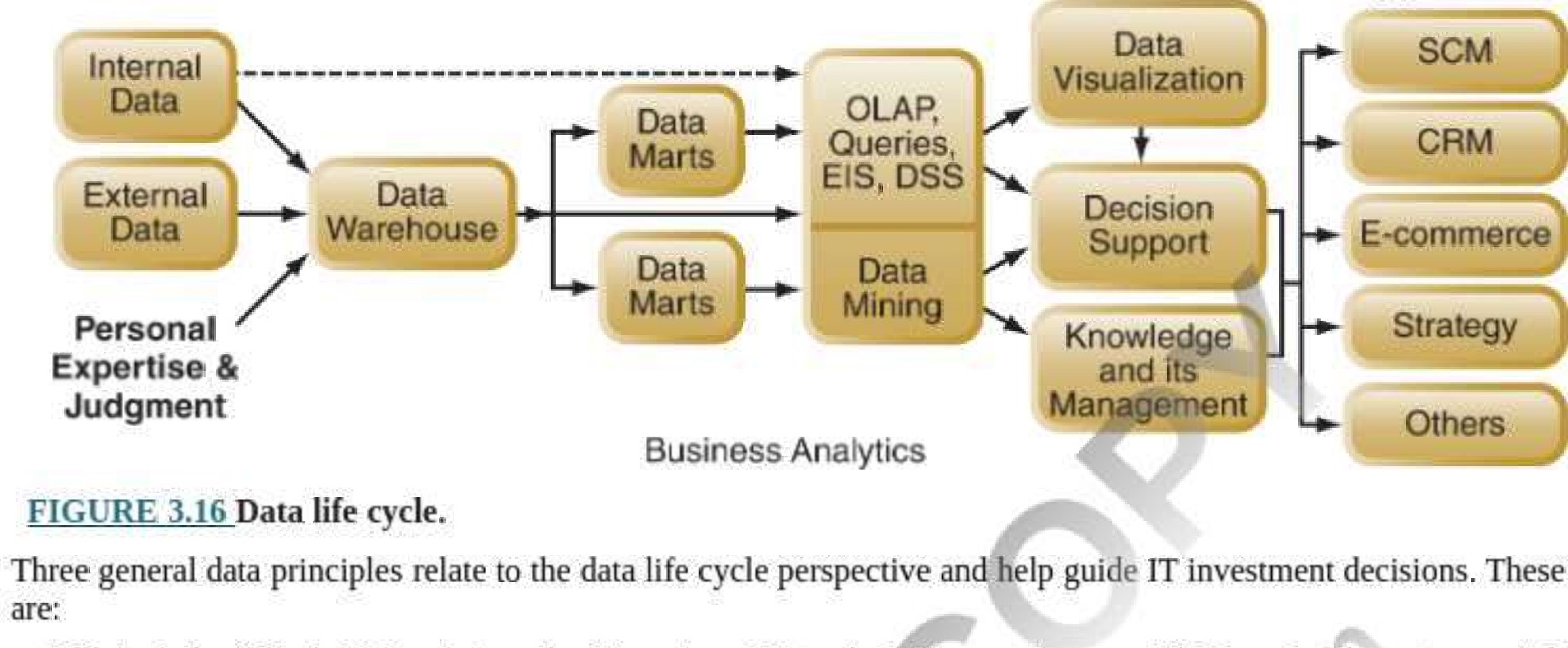
Trong các tổ chức, các tình huống **quản lý thông tin** phức tạp phát sinh hàng ngày và yêu cầu một kế hoạch liên tục để hướng dẫn và kiểm soát các khoản đầu tư vào CNTT. Kế hoạch này được gọi là quản lý thông tin và nó được hướng dẫn bởi chu kỳ dữ liệu và các nguyên tắc dữ liệu.

**Quản lý thông tin** là quá trình thu thập , lưu trữ , quản lý và duy trì dữ liệu chính xác,

kịp thời, đáng tin cậy, hợp lệ, sẵn có, duy nhất và phù hợp với một tổ chức.

#### Vòng đời dữ liệu và nguyên tắc dữ liệu

Vòng đời dữ liệu là một mô hình minh họa cách dữ liệu di chuyển trong một tổ chức như trong Hình 3.16 . Vòng đời dữ liệu bắt đầu với việc lưu trữ trong cơ sở dữ liệu, được tải vào kho dữ liệu để phân tích, sau đó được báo cáo cho nhân viên tri thức hoặc được sử dụng trong các ứng dụng kinh doanh. Quản lý chuỗi cung ứng (SCM), quản lý quan hệ khách hàng (CRM), quản lý tri thức ( KM) và các ứng dụng doanh nghiệp khác yêu cầu dữ liệu cập nhật, dễ truy cập để hoạt động bình thường và được thảo luận chi tiết trong Chương 10 .



and

Databases

Applications

Data

Sources

Data

Storage

Data

Analysis

Results

Business

Ba nguyên tắc chung về dữ liệu liên quan đến vòng đời dữ liệu và giúp hướng dẫn các quyết định đầu tư CNTT :

1. **Nguyên tắc giảm dần giá trị dữ liệu** Giá trị của dữ liệu giảm dần khi chúng già đi. Đây là một nguyên tắc đơn giản nhưng mạnh mẽ . Hầu hết các tổ chức không thể hoạt động ở hiệu suất cao nhất với các điểm mù (thiếu dữ liệu sẵn có) trong 30 ngày hoặc lâu hơn. Các tổ chức dịch vụ tài chính toàn cầu dựa vào dữ liệu gần thời gian thực để đạt hiệu suất cao nhất.
2. **Nguyên tắc sử dụng dữ liệu 90/90** Theo nguyên tắc sử dụng dữ liệu 90/90, phần lớn dữ liệu được lưu trữ, cao tới 90%, hiếm khi được truy cập sau 90 ngày (ngoại trừ mục đích kiểm toán). Nghĩa là, khoảng 90% dữ liệu mất phần lớn giá trị sau ba tháng.
3. **Nguyên tắc của dữ liệu trong bối cảnh** Khả năng nắm bắt, xử lý, định dạng và phân phối dữ liệu trong thời gian gần thực hoặc nhanh hơn đòi hỏi một khoản đầu tư lớn vào kiến trúc dữ liệu ( Chương 2 ) và cơ sở hạ tầng để liên kết hệ thống POS từ xa với hệ thống lưu trữ dữ liệu, phân tích dữ liệu, và các ứng dụng báo cáo. Việc đầu tư có thể được chứng minh trên nguyên tắc rằng dữ liệu phải được tích hợp, xử lý, phân tích và định dạng thành “thông tin có thể thực hiện được ."

Cung cấp quyền truy cập dễ dàng vào khối lượng lớn thông tin chỉ là một trong những thách thức mà các tổ chức phải đối mặt. Thời của việc quản lý dữ liệu có cấu trúc đơn giản đã qua. Giờ đây, các tổ chức phải quản lý nội dung bán cấu trúc và phi cấu trúc từ các nguồn xã hội và di động mặc dù dữ liệu đó có thể có chất lượng đáng ngờ. Các vấn đề xung quanh việc truy cập, quản lý và bảo mật thông tin đòi hỏi phải giải quyết tình trạng xuống cấp và rối loạn thông tin.

#### Khai thác dữ liệu phân tán

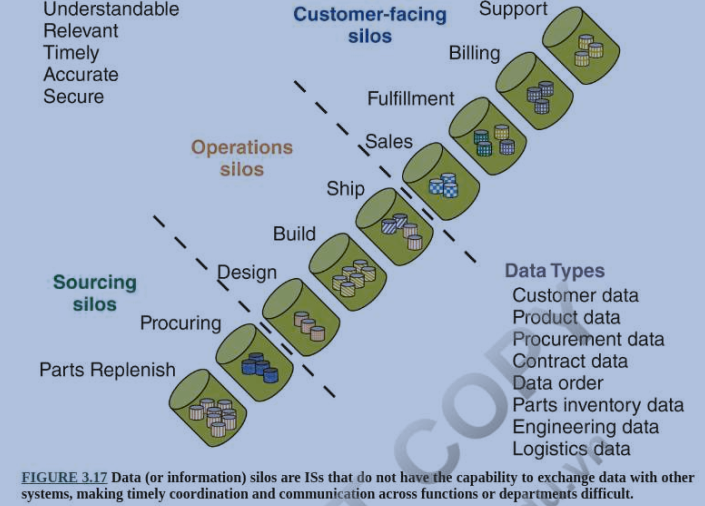
Thông tin kinh doanh thường nằm rải rác khắp doanh nghiệp, được lưu trữ trong các hệ thống riêng biệt dành riêng cho các mục đích cụ thể, chẳng hạn như hoạt động, quản lý chuỗi cung ứng hoặc quản lý quan hệ khách hàng. Các tổ chức lớn có hơn 100 kho dữ liệu (khu vực lưu trữ). Ở nhiều công ty, việc tích hợp các hệ thống khác nhau này bị hạn chế—cũng như khả năng truy cập tất cả thông tin họ cần của người dùng. Kết quả là, bất chấp tất cả thông tin chảy qua các công ty, giám đốc điều hành, quản lý và công nhân thường loay hoay tìm thông tin họ cần để đưa ra quyết định đúng đắn hoặc thực hiện công việc của họ.

#### Phá vỡ silo dữ liệu

Các công ty trên toàn thế giới đang gặp khó khăn trong việc tích hợp hàng nghìn ứng dụng toàn cầu được bảo mật và sắp xếp chúng phù hợp với chiến lược kinh doanh của họ: **Các silo chứa dữ liệu** tồn tại tại nơi làm việc khi các phòng ban và đơn vị kinh doanh tự làm việc với nhau và hiếm khi hợp tác với các bộ phận hoặc đơn vị kinh doanh khác (hình 3.17 ). Hiệu ứng silo xảy ra khi các bộ phận hoặc nhóm riêng biệt không có hệ thống để giao tiếp hiệu quả với nhau, dữ liệu bị mắc kẹt trong các silo dữ liệu của bộ phận và năng suất bị ảnh hưởng. Ví dụ, khi hai bộ phận đang làm việc trên cùng một sáng kiến và không bộ phận nào biết rằng bộ phận kia đang làm việc với nó.

Silo dữ liệu là các kho lưu trữ dữ liệu độc lập Dữ liệu của họ không thể truy cập được bởi các hệ thống thông tin khác mà có nhu cầu cần đến nó hoặc bên ngoài bộ phận đó.





HÌNH 3.17 Các silo dữ liệu (hoặc thông tin) là những hệ thống thôn tin mà không có khả năng trao đổi dữ liệu với các hệ thống khác, gây khó khăn cho việc phối hợp và liên lạc kịp thời giữa các bộ phận chức năng.

Các silo dữ liệu thường được tìm thấy trong các tổ chức dịch vụ, chẳng hạn như health-care và chính phủ. Ví dụ, ở nhiều bệnh viện, từng ngành, khoa, phòng triển khai những app IT riêng mà thường không phân tích kỹ mối quan hệ với các hệ thống khoa, phòng khác. Sự sắp xếp này dẫn đến việc bệnh viện có các nhóm CNTT quản lý một bộ ứng dụng cụ thể Hoặc kho chứa dữ liệu cho một khoa hoặc bộ phận cụ thể.

Vì các silo không thể chia sẻ hoặc trao đổi dữ liệu nên chúng không thể cập nhật một cách nhất quán. Khi dữ liệu không nhất quán trên nhiều ứng dụng doanh nghiệp, chất lượng dữ liệu không thể (và không nên) được tin cậy nếu không có xác minh rộng rãi. Các silo dữ liệu tồn tại khi không có kiến trúc CNTT tổng thể để hướng dẫn các khoản đầu tư CNTT, điều phối dữ liệu và liên lạc. Các silo dữ liệu hỗ trợ một chức năng duy nhất và do đó, không hỗ trợ các nhu cầu đa chức năng của một tổ chức.

Mặc dù công nghệ kết nối nhiều nhân viên hơn bao giờ hết, nhưng hiệu ứng silo vẫn tiếp tục là mối quan tâm trong các tổ chức thuộc mọi loại hình và quy mô: Dưới đây là một số thống kê thú vị được thu thập trong một cuộc khảo sát gần đây (Marchese, 2019 ) về cộng tác tại nơi làm việc.

* 39% nhân viên được khảo sát tin rằng mọi người trong tổ chức của họ KHÔNG hợp tác đủ

75% đánh giá công việc nhóm và cộng tác là "rất quan trọng", NHƯNG chỉ 18% nhận được đánh giá giao tiếp như một phần của đánh giá hiệu suất của họ

• 49% thế hệ thiên niên kỷ ủng hộ các công cụ xã hội để cộng tác tại nơi làm việc

* 97% nhân viên và giám đốc điều hành tin rằng sự thiếu liên kết trong nhóm ảnh hưởng đến kết quả của một nhiệm vụ hoặc dự án

• 86% nhân viên và giám đốc điều hành cho rằng thiếu hợp tác hoặc giao tiếp không hiệu quả là nguyên nhân dẫn đến thất bại tại nơi làm việc

90% nhân viên tin rằng người ra quyết định nên tham khảo ý kiến của người khác trước khi đưa ra quyết định cuối cùng, NHƯNG 40% nhân viên tin rằng những người ra quyết định luôn thất bại trong việc tìm kiếm ý kiến khác

* Dưới 50% cho rằng tổ chức của họ thảo luận các vấn đề một cách trung thực và hiệu quả

##### Văn Hóa Phải Thay Đổi

Nếu các công ty muốn giết chết sự đổi mới, năng suất và sự hợp tác nội bộ lành mạnh thì tất cả những gì họ phải làm là không khuyến khích sự hợp tác và tiếp tục thúc đẩy văn hóa silo thông tin (Christman, 2018 ). Vì vậy, làm thế nào để bạn phá vỡ các silo và thúc đẩy chia sẻ và cộng tác để tăng cường giao tiếp giữa các chức năng? Điều đầu tiên cần nhận ra là việc chuyển từ silo sang hợp tác đòi hỏi phải thay đổi văn hóa. Bạn đang yêu cầu mọi người thay đổi cách làm việc và suy nghĩ khác đi. KAI Partners, Inc. (https://kaipartners.com/), một công ty tư vấn, có một vài gợi ý:

I. **Công khai thừa nhận các mục tiêu được chia sẻ** Tạo một tầm nhìn thống nhất và thành lập các nhóm liên đơn vị nơi các mục tiêu chung được công bố công khai và các chiến thắng được chia sẻ.

2. **Nắm bắt câu hỏi "tại sao"** Khuyến khích mọi người đặt câu hỏi. Mọi người cần thông tin để hoàn thành công việc. Thúc đẩy sự sáng tạo và trí tưởng tượng để khuyến khích sự đổi mới.

3, **Văn hóa xuất phát từ cấp trên** Dẫn đầu bằng tấm gương là cách tốt nhất để mô hình hóa sự thay đổi mà bạn muốn người khác thực hiện. Đừng chỉ khuyến khích nhân viên trở thành "những kẻ phá bĩnh" Đảm bảo rằng quản lý cấp trên, bao gồm cả các chuyên gia cấp C, dẫn dắt các nỗ lực thay đổi.

Bằng cách chuyển từ hệ thống silo sang hệ thống hợp tác, các tổ chức có thể vượt qua các rào cản đối với sự đổi mới mà việc thiếu giao tiếp có thể thúc đẩy.

#### Rác Vào, Rác Ra

Trong kinh doanh, dữ liệu được sử dụng vì hai lý do: để điều hành một doanh nghiệp (sử dụng trong giao dịch và vận hành) và để cải thiện doanh nghiệp (sử dụng cho phân tích). Cả hai đều phụ thuộc vào dữ liệu đáng tin cậy chất lượng cao. Bất kể dữ liệu được thu thập như thế nào, chúng cần được xác thực để người dùng biết rằng họ có thể tin tưởng chúng. Thật không may, dữ liệu về số lượng không được xác thực và cũng không đáng tin cậy. Có lẽ bạn đã quen với cách diễn đạt cổ điển tổng kết tình huống này—"rác vào, rác ra" (GIGO). Để đề phòng điều này xảy ra, các công ty phải áp dụng các biện pháp bảo vệ, chẳng hạn như kiểm tra tính toàn vẹn của dữ liệu. Ví dụ: khi bạn điền vào biểu mẫu trực tuyến, nó sẽ không chấp nhận địa chỉ e-mail hoặc số thẻ tín dụng không đúng định dạng. Loại dữ liệu này được gọi là dữ liệu bẩn dirty data

**Dữ liệu bẩn** là dữ liệu có chất lượng kém đến mức không thể tin cậy hoặc dựa vào chúng để đưa ra quyết định.

Khi các ứng dụng không được quản lý tốt, chúng có thể tạo ra hàng terabyte dữ liệu không liên quan, khiến các công ty chìm trong dữ liệu xấu. Sự hỗn loạn dữ liệu này có thể dẫn đến các lỗi có thể khiến công ty gặp rủi ro. Trong nỗ lực quản lý lượng dữ liệu khổng lồ và quá mức, sẽ tăng nguy cơ thông tin liên quan bị mất (thiếu) hoặc không chính xác.

##### **Dữ liệu không chính xác**

Các nhà quản lý và nhân viên thông tin thường xuyên bị hạn chế bởi các dữ liệu không thể kiểm tra được vì chúng không đầy đủ, không phù hợp với bối cảnh, lỗi thời, không chính xác, không thể truy cập hoặc quá nhiều đến mức họ phải mất hàng tuần để phân tích. Trong những tình huống như vậy, người ra quyết định phải đối mặt với quá nhiều sự không chắc chắn để đưa ra quyết định kinh doanh thông minh.

Các quyết định tồi từ dữ liệu không chính xác dẫn đến tăng chi phí, giảm doanh thu và các hậu quả liên quan đến pháp lý, uy tín và hiệu suất. Ví dụ: nếu dữ liệu được thu thập và phân tích dựa trên thông tin không chính xác do quảng cáo được thực hiện ở sai vị trí cho sai đối tượng, các chiến dịch tiếp thị có thể trở nên sai lệch và không hiệu quả. Sau đó, các công ty phải bắt đầu sửa chữa tốn kém bộ dữ liệu của họ để khắc phục các sự cố do dữ liệu bẩn gây ra. Điều này tạo ra sự sụt giảm trong sự hài lòng của khách hàng và lạm dụng các nguồn lực trong một công ty.

Sự chấp nhận rộng rãi rằng các tổ chức dịch vụ; như các hệ thống chăm sóc sức khỏe, đang chìm trong dữ liệu không chính xác. Ví dụ, một hậu quả ngoài ý muốn của *Hồ sơ sức khỏe điện tử* là dữ liệu có thể không chính xác và đầy đủ như mong đợi. Các giá trị phòng xét nghiệm, kết quả hình ảnh hoặc tài liệu bác sĩ không chính xác dẫn đến sai sót y tế, làm hại bệnh nhân và làm tổn hại uy tín danh tiếng của tổ chức. Điều này đặc biệt đáng lo ngại vì dữ liệu được thu thập từ ghi chú của bác sĩ, mẫu đăng ký, tóm tắt xuất viện, tài liệu, v.v. đang tăng gấp đôi sau mỗi năm năm. Không giống như dữ liệu có cấu trúc, đây là những dữ liệu lộn xộn khiến các nhà cung cấp dịch vụ chăm sóc sức khỏe phải mất rất nhiều thời gian và công sức để đưa vào hệ thống để dùng vào phân tích kinh doanh của họ. Vì vậy, dữ liệu có giá trị nhưng lộn xộn thường bị bỏ qua. Do đó, hàng triệu ghi chú và hồ sơ chi tiết của bệnh nhân không thể truy cập hoặc không có sẵn vì trước đây không có cách nào dễ dàng để nhập và phân tích thông tin chứa trong đó.

##### **Dữ liệu bị mất**

Dữ liệu có thể bị mất khi chuyển từ hệ thống này sang hệ thống khác. Hoặc, dữ liệu có thể không bao giờ được thu thập do hệ thống thu thập dữ liệu được điều chỉnh không phù hợp, chẳng hạn như những hệ thống dựa vào cảm biến hoặc máy quét. Hoặc, dữ liệu có thể không được thu thập đầy đủ chi tiết, như được mô tả.

Bảng 3.3 liệt kê các đặc điểm thường liên quan đến dữ liệu bẩn hoặc kém chất lượng.

|  |  |
| --- | --- |
| BẢNG 3.3 | |
|  | |
| Đặc điểm của dữ liệu bẩn | Sự miêu tả |
| Chưa hoàn thiện | Dữ liệu bị mất |
| Lỗi thời hoặc không hợp lệ | Quá cũ để có giá trị hoặc hữu ích |
| Không đúng | Quá nhiều lỗi |
| Trùng lặp hoặc xung đột | Quá nhiều bản sao hoặc phiên bản của cùng một dữ liệu—và các phiên bản không nhất quán hoặc xung đột với nhau |
| Không chuẩn hóa | Dữ liệu được lưu trữ ở các định dạng không tương thích—và không thể so sánh hoặc tóm tắt |
| Không sử dụng được | Dữ liệu không nằm trong ngữ cảnh để có thể hiểu hoặc diễn giải chính xác tại thời điểm truy cập |

##### 

##### **Giao diện được thiết kế kém**

Bất chấp mọi lời bàn tán về giao diện thân thiện với người dùng, một số hệ thống thông tin thật tệ khi xử lý. Các giao diện hoặc định dạng được thiết kế kém đòi hỏi thêm thời gian và công sức để tìm ra sẽ làm tăng nguy cơ xảy ra lỗi do hiểu sai dữ liệu hoặc bỏ qua chúng.

###### **Định dạng dữ liệu không chuẩn hóa**

Khi người dùng trình bày dữ liệu ở định dạng không nhất quán hoặc không chuẩn hóa, lỗi sẽ tăng lên. Nỗ lực so sánh hoặc phân tích dữ liệu khó khăn hơn và mất nhiều thời gian hơn. Ví dụ: nếu bộ phận Đông Bắc báo cáo tổng doanh thu bán hàng hàng tuần trên mỗi dòng sản phẩm và bộ phận Tây Nam báo cáo doanh thu thuần hàng tháng trên mỗi sản phẩm, bạn không thể so sánh hiệu suất của chúng mà không chuyển đổi dữ liệu về định dạng chung. Hãy xem xét nỗ lực nhiều hơn cần thiết để so sánh doanh số bán hàng liên quan đến nhiệt độ, chẳng hạn như máy điều hòa không khí, khi một số nhiệt độ được biểu thị bằng độ F và những thứ khác được biểu thị bằng C.

###### **Dữ liệu lỗi thời**

Thông tin mà những người ra quyết định muốn liên tục thay đổi và nó thay đổi nhanh hơn khả năng phản hồi của các bộ phận IS vì bốn lý do cuối cùng trong danh sách này. Theo dõi các tweet, lượt truy cập YouTube và nội dung phi cấu trúc khác yêu cầu các khoản đầu tư tốn kém, mà các nhà quản lý thấy rủi ro trong thời kỳ suy thoái kinh tế. 

##### Cái giá của dữ liệu bẩn

Trung bình, một tổ chức có mức tăng trưởng dữ liệu 40% hàng năm và 20% dữ liệu đó bị phát hiện là bẩn. Chi phí của dữ liệu kém chất lượng lan rộng khắp công ty và ảnh hưởng đến các hệ thống từ vận chuyển và tiếp nhận đến kế toán và dịch vụ khách hàng. Lỗi dữ liệu thường phát sinh từ các chức năng hoặc bộ phận phát sinh hoặc tạo dữ liệu chứ không phải trong bộ phận CNTT. Khi tất cả các chi phí được xem xét, giá trị của việc tìm và khắc phục nguyên nhân gây ra lỗi dữ liệu trở nên rõ ràng. Trong thời kỳ ngân sách bị cắt giảm, một số tổ chức có thể không có nguồn lực cho các dự án như vậy và thậm chí có thể không nhận thức được vấn đề. Những người khác có thể dành phần lớn thời gian của họ để khắc phục sự cố, do đó khiến họ không có thời gian để ngăn chặn chúng. Lợi ích của việc đảm bảo rằng dữ liệu chính xác, kịp thời, đầy đủ, nhất quán và duy nhất rất đáng để nỗ lực. Việc sửa dữ liệu bẩn đắt gấp 10 lần so với việc ngăn chặn nó.

Dữ liệu bẩn đang khiến các doanh nghiệp Mỹ thiệt hại hàng trăm tỷ đô la mỗi năm và ảnh hưởng đến khả năng vượt qua môi trường kinh tế khó khăn của họ. Các giá trị lạc hậu và lỗi thời, dữ liệu bị thiếu và các định dạng dữ liệu không nhất quán có thể làm mất khách hàng, doanh thu và doanh thu; phân bổ sai nguồn lực; và chiến lược định giá thiếu sót. Ví dụ về các chi phí này bao gồm:

* **Mất cơ hội** **kinh doanh** Việc kinh doanh bị mất khi cơ hội bán hàng bị bỏ lỡ, đơn hàng bị trả lại do giao sai mặt hàng hoặc sai sót khiến khách hàng bực bội và bỏ đi
* **Tốn Thời gian dành cho việc ngăn ngừa lỗi** Nếu dữ liệu không đáng tin cậy thì nhân viên cần dành nhiều thời gian và công sức hơn để kiểm tra thông tin tránh nhầm lẫn.
* **Tốn Thời gian sửa lỗi** Nhân viên cơ sở dữ liệu cần xử lý sửa lỗi cho cơ sở dữ liệu. 

Chi phí của dữ liệu chất lượng kém có thể được biểu thị theo công thức sau:

Sử dụng công thức này, hãy xem xét ví dụ về chi phí sửa lỗi trong một tổ chức:

* 1. Hai nhân viên cơ sở dữ liệu member dành 25% thời gian trong ngày làm việc của họ để xử lý và xác minh các chỉnh sửa dữ liệu mỗi ngày:
  2. Lương theo giờ là $50 mỗi giờ dựa trên mức lương và lợi ích:
  3. 250 ngày làm việc mỗi năm:

Rất khó để tính toán toàn bộ chi phí của dữ liệu kém chất lượng và những ảnh hưởng lâu dài của nó. Một phần của khó khăn là thời gian trễ giữa lỗi và khi nó được phát hiện. Các lỗi có thể rất khó sửa/ đặc biệt là khi các hệ thống mở rộng khắp doanh nghiệp. Một mối quan tâm khác là tác động của các lỗi có thể không thể đoán trước, ảnh hưởng sâu rộng và nghiêm trọng. CNTT tại nơi làm việc 3.4 xem xét việc kho dữ liệu bẩn có thể ảnh hưởng đến sự phân biệt đối xử trong hệ thống tư pháp hình sự như thế nào.

## CNTT tại nơi làm việc 3.4

### Cẩn thận với dữ liệu bẩn trong các hệ thống chính sách dự đoán

Với việc các cơ quan thực thi pháp luật đang bị giám sát chặt chẽ hơn trong những năm gần đây vì hành vi gây hấn không cân xứng đối với các nghi phạm thiểu số, người ta đã suy đoán rằng phần mềm cảnh sát dự đoán có thể giúp kiểm soát những loại hành vi này. Gần đây, các nhà nghiên cứu tại Đại học New York đã bắt đầu thử nghiệm lý thuyết này và đưa ra một số kết quả đáng lo ngại.

Đã có báo cáo rằng ít nhất 13 khu vực pháp lý ở Hoa Kỳ, hiện đang sử dụng các hệ thống cảnh sát dự báo để dự đoán hoạt động tội phạm và phân bổ nguồn lực của cảnh sát, thật không may, ở nhiều khu vực pháp lý này, các hệ thống này được xây dựng dựa trên dữ liệu dựa trên cáo buộc sai sót, phân biệt chủng tộc và đôi khi các hành vi hoàn toàn trái pháp luật. Điều này có thể bao gồm thao túng dữ liệu trên toàn hệ thống , báo cáo giả mạo của cảnh sát, sử dụng vũ lực trái pháp luật, đưa ra bằng chứng và khám xét vi hiến. Các nhà nghiên cứu tại Viện Al Now và Đại học New York gần đây đã tiến hành một nghiên cứu về các hệ thống kiểm soát dự đoán. Trong đó, họ so sánh bằng chứng đã được chứng minh và các phát hiện khác về các hoạt động bất hợp pháp hoặc thiên vị của cảnh sát được thu thập trong quá trình điều tra của Bộ Tư pháp hoặc xét xử của tòa án liên bang với thông tin có sẵn công khai về thời điểm khu vực tài phán đã sử dụng hệ thống kiểm soát dự đoán. Sau đó, kết quả so sánh được sử dụng để xác định xem dữ liệu của cảnh sát được sử dụng để đào tạo hoặc triển khai các hệ thống kiểm soát dự đoán có được tạo ra trong các giai đoạn có hành vi sai trái và thiên vị được ghi nhận hay không.



Bằng chứng chắc chắn cho thấy rằng Chicago, New Orleans và Quận Maricopa, AZ sử dụng các hệ thống kiểm soát dự đoán dựa trên dữ liệu bẩn, mặc dù ở New Orleans, việc thiếu minh bạch và các cơ chế chịu trách nhiệm khiến việc phát hiện ra mức độ nghiêm trọng của vấn đề và đánh giá đúng trở nên khó khăn. rủi ro liên quan đến dữ liệu thiếu sót trong các hệ thống chính sách dự đoán.



Ví dụ, Sở cảnh sát Chicago đang bị điều tra liên bang về các hành vi phi pháp của cảnh sát khi họ triển khai một hệ thống máy tính xác định những người có nguy cơ trở thành nạn nhân hoặc kẻ phạm tội trong một vụ xả súng hoặc giết người. Nghiên cứu của NYU tiết lộ rằng cư dân có cùng nhân khẩu học với những người được DOJ xác định là mục tiêu của xu hướng kiểm soát của Chicago phù hợp với những người được xác định trong hệ thống kiểm soát dự đoán,

Những phát hiện này gợi ý mạnh mẽ rằng việc triển khai các hệ thống kiểm soát dự đoán trong các khu vực pháp lý với phạm vi rộng lớn

|  |
| --- |
| ——crv"  lịch sử của các hoạt động bất hợp pháp của cảnh sát làm tăng rủi ro rằng dữ liệu bẩn sẽ dẫn đến những dự đoán sai sót hoặc bất hợp pháp về hành vi tội phạm. Đổi lại, điều này có thể gây thêm tác hại cho công dân thông qua phản hồi trong toàn bộ hệ thống tư pháp hình sự.    Trưởng nhóm nghiên cứu Rashida Richardson đã nhanh chóng chỉ ra rằng "Mặc dù nghiên cứu này chỉ giới hạn ở các khu vực pháp lý có lịch sử lâu đời về hành vi sai trái của cảnh sát và các hoạt động phân biệt đối xử của cảnh sát, chúng tôi biết rằng những lo ngại này về các chính sách và hoạt động trị an không chỉ giới hạn ở các khu vực pháp lý này, vì vậy sự xem xét kỹ lưỡng hơn về dữ liệu được sử dụng trong các công nghệ lập chính sách dự đoán là cần thiết trên toàn cầu."  Phát hiện của họ cho thấy rằng các hệ thống kiểm soát dự đoán phải được xử lý thận trọng và điều quan trọng là phải đưa ra các cơ chế để cho phép công chúng trên toàn cầu nhận thức được các hệ thống đó cũng như đánh giá và từ chối chúng.  Nguồn: Tổng hợp từ Richardson, Schultz và Crawford (2019 ), Đạo đức dữ liệu (2019 ); và McClain ( 2019 ). |

Mục tiêu tổng thể của quản lý thông tin là loại bỏ cuộc đấu tranh đó thông qua thiết kế và triển khai chương trình quản trị dữ liệu hợp lý và kiến trúc doanh nghiệp (EA) được lên kế hoạch tốt. Một mục tiêu khác của quản lý thông tin là bảo mật và tuân thủ dữ liệu. Ví dụ: các tổ chức phải không tuân thủ các yêu cầu pháp lý liên tục phát triển, chẳng hạn như Đạo luật Sarbanes-Oxley, Basel Ill, Đạo luật Lạm dụng và Lừa đảo Máy tính (CFAAb Đạo luật USE PATRIOT và Đạo luật Trách nhiệm Giải trình và Di chuyển Bảo hiểm Y tế (HIPAA) . có thể bắt nguồn từ việc sử dụng quản lý thông tin bao gồm:

1, **Nâng cao chất lượng quyết định** Chất lượng quyết định phụ thuộc vào dữ liệu chính xác và đầy đủ.

1. **Cải thiện độ chính xác và độ tin cậy của các dự đoán quản lý** Điều cần thiết đối với các nhà quản lý là có thể dự đoán doanh số bán hàng, nhu cầu sản phẩm, cơ hội và các mối đe dọa cạnh tranh. Dự đoán quản lý tập trung vào "điều gì sẽ xảy ra" trái ngược với báo cáo tài chính về "'điều gì đã xảy ra".
2. **Giảm thời gian và chi phí** định vị và tích hợp thông tin liên quan.
3. **Chất lượng dữ liệu tốt hơn** Chuẩn hóa dữ liệu dẫn đến dữ liệu nhất quán và đáng tin cậy hơn

#### câu hỏi

1. Các giai đoạn của vòng đời dữ liệu là gì?
2. Chức năng quản lý dữ liệu chủ (MDM) là gì?
3. Hậu quả của việc không làm sạch "dữ liệu bẩn" là gì?
4. '"Hiệu ứng silo" là gì và nó ảnh hưởng như thế nào đến hiệu quả kinh doanh?
5. Ba yếu tố nào đang thúc đẩy sự hợp tác và chia sẻ thông tin?
6. Lợi ích kinh doanh của việc quản lý thông tin là gì?

7. Tại sao việc quản lý thông tin tốt lại quan trọng?

### 3.5 Quản lý nội dung, hồ sơ và tài liệu điện tử

L03.5 Định nghĩa các thuật ngữ tài liệu điện tử, hồ sơ điện tử, và nội dung điện tử và cách các công nghệ quản lý chúng đang được sử dụng để làm cho các tổ chức hiệu quả và hiệu quả hơn.

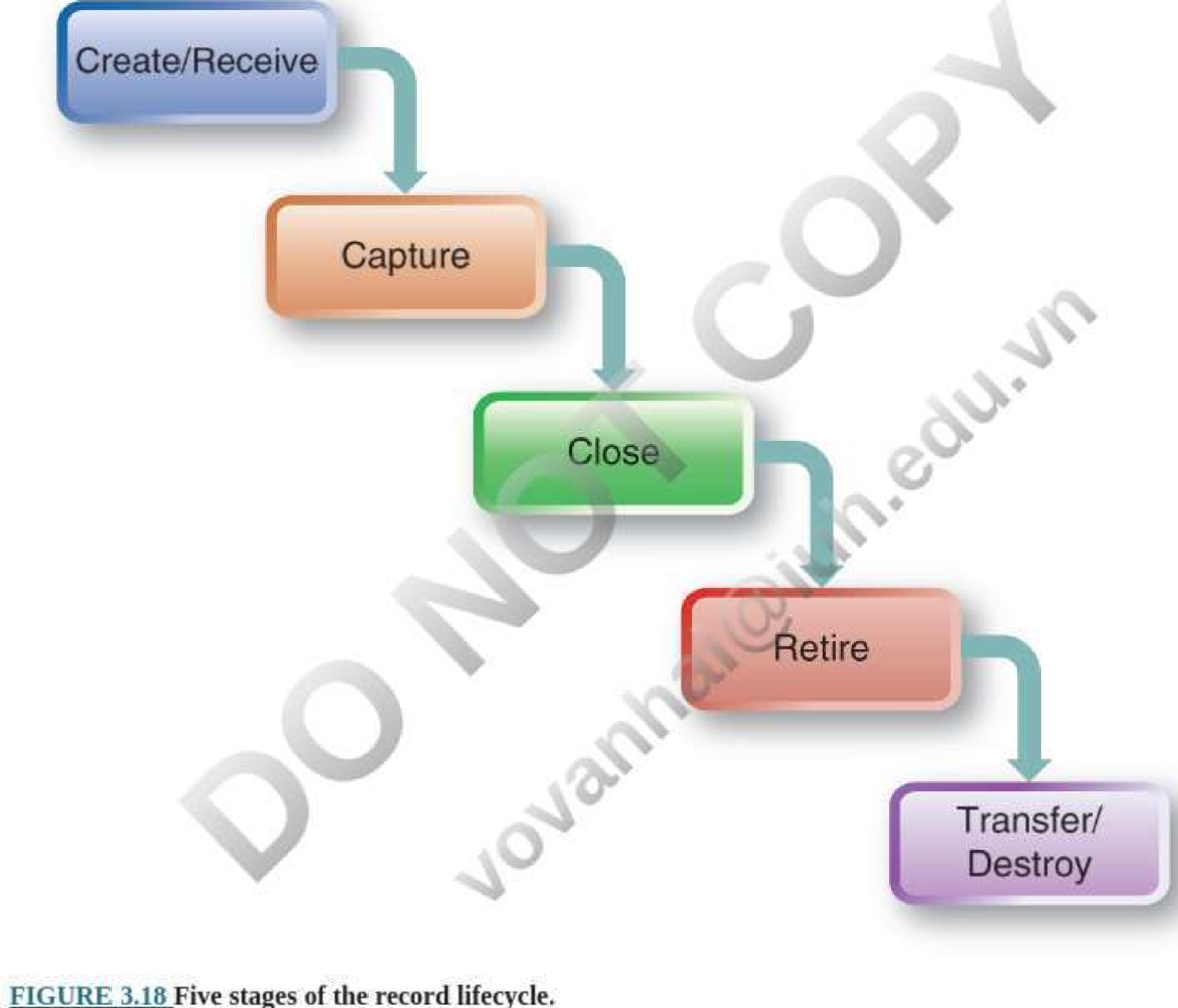
Thông tin và tri thức được lưu trữ trong cơ sở dữ liệu dưới dạng văn bản, hồ sơ và nội dung điện tử. Pricewaterhouse Coopers (PwC), một mạng lưới dịch vụ chuyên nghiệp lớn, gần đây đã báo cáo rằng hơn bốn nghìn tỷ tài liệu giấy được sản xuất chỉ riêng tại Hoa Kỳ và con số này đang tăng lên ở mức đáng báo động 22% mỗi năm! Họ cũng phát hiện ra rằng 7,5% tài liệu bị mất, 3% số còn lại bị lưu sai và tài liệu trung bình được sao chép 19 lần. Họ cũng ước tính rằng các công ty chi 20 đô la lao động để nộp một tài liệu; $120 để tìm tài liệu bị gửi nhầm và $220 để tạo lại tài liệu bị mất. Điều này đồng nghĩa với việc hàng triệu cây xanh bị mất đi và hàng tỷ đô la đã được chi tiêu về thời gian và tiền bạc. Do đó, các công ty ngày càng chuyển sang sử dụng tài liệu điện tử, hồ sơ và quản lý nội dung như một giải pháp.

Đầu tiên, thật hữu ích khi hiểu được sự khác biệt giữa tài liệu điện tử, hồ sơ điện tử và nội dung điện tử. Tiếp theo, chúng tôi sẽ giải thích cách quản lý từng loại và mô tả phần mềm hỗ trợ chúng. Sự khác biệt quan trọng giữa **tài liệu điện tử** và **hồ sơ điện tử** là tài liệu đang được xử lý tích cực và có thể thay đổi. Ngược lại, hồ sơ là công việc đã hoàn thành và không có ý định thay đổi. Các giai đoạn khác nhau trong vòng đời của hồ sơ điện tử được trình bày trong hình 3.18 . **Nội dung điện tử** không chỉ bao gồm các tài liệu và bản ghi bằng văn bản mà còn bao gồm dữ liệu được duy trì ở các định dạng khác nhau bao gồm các tệp âm thanh, video, flash và đa phương tiện.

**Tài liệu điện tử** là bất kỳ giấy tờ, biểu mẫu điện tử, tệp, email, fax, hợp đồng, hợp đồng thuê, v.v. đang được xử lý tích cực.

**Hồ sơ điện tử** là bất kỳ tài liệu nào đã được tạo thành cuối cùng và không còn có nghĩa là bị thay đổi.

**Nội dung điện tử** là tập hợp các tài liệu, bản ghi và dữ liệu phi cấu trúc có sẵn dưới dạng nhiều loại tài sản kỹ thuật số, chẳng hạn như âm thanh, video, flash, tệp đa phương tiện, v.v.



FIGURE

3.18

Five

stages

the

record

lifecycle.

Electronic

Document

Management

Quản lý tài liệu điện tử

Mục đích của **quản lý tài liệu điện tử** là tạo, chia sẻ, tổ chức và truy xuất hiệu quả các tài liệu đang xử lý hoặc tài liệu "sống" một cách có hệ thống. Thông tin trong quản lý tài liệu là nội dung tạm thời, chẳng hạn như hóa đơn được ký và gửi cho người ký tiếp theo hoặc bản nháp cũ hơn của tài liệu bị loại bỏ để sửa đổi. Nó được định hướng theo nội dung trong đó các kho lưu trữ tài liệu thường được tổ chức để tìm tài liệu một cách hiệu quả theo từ khóa hoặc tiêu đề và nhóm các tài liệu theo tên khách hàng hoặc sản phẩm. Mục tiêu của quản lý tài liệu là hiệu quả: giảm giấy tờ, giảm thất lạc và sắp xếp sai tài liệu, cung cấp khả năng truy cập và truy xuất nhanh hơn, dễ dàng hơn; giảm chi phí và không gian lưu trữ, đồng thời hợp lý hóa thông tin và quy trình làm việc,

**Quản lý tài liệu điện tử** là lưu trữ điện tử, duy trì và theo dõi các tài liệu điện tử và hình ảnh điện tử của thông tin trên giấy.

#### Các hệ thống quản lý tài liệu điện tử

Phần mềm hỗ trợ quản lý văn bản điện tử được gọi là **phần mềm quản lý văn bản điện tử ( EDMS )**. EDMS được nhiều người dùng trong toàn tổ chức sử dụng để thực hiện các chức năng công việc hàng ngày của họ. Do EDMS xử lý các tài liệu "sống" được tạo phiên bản và yêu cầu phê duyệt thường xuyên nên việc xem xét và phê duyệt điện tử là các tính năng chính của EDMS. Mặc dù việc quản lý EDMS tạo ra chi phí chung cho người dùng nhập và tạo tài liệu, sức mạnh của nó nhanh chóng được nhận ra khi người dùng trong toàn tổ chức có thể dễ dàng tìm thấy tài liệu.

**Hệ thống quản lý tài liệu điện tử** là một hệ thống phần mềm để tạo, sắp xếp, lưu trữ và truy xuất các loại tài liệu điện tử khác nhau.

#### Quản lý hồ sơ điện tử

Mặt khác, **quản lý hồ sơ điện tử** tập trung vào việc lưu trữ hồ sơ để giúp một tổ chức tránh được các hình phạt do cơ quan quản lý, kiểm toán viên và các cơ quan quản lý khác áp đặt. Lưu trữ bao gồm phân loại, lưu trữ và bảo mật hồ sơ và có kế hoạch xác định hồ sơ nào, khi nào và như thế nào nên bị hủy hoặc bảo quản (trong trường hợp hồ sơ lịch sử quan trọng). Quản lý hồ sơ bao gồm nội dung lịch sử và được điều khiển theo ngữ cảnh, như trong loại tài liệu và lịch trình lưu giữ. Các cơ quan chính phủ dự kiến sẽ quản lý tất cả các hồ sơ vĩnh viễn bằng điện tử để chuẩn bị cho hạn chót ngày 31 tháng 12 năm Z 2022 sắp đến khi Cơ quan Quản lý Hồ sơ và Lưu trữ Quốc gia (NARA) sẽ không còn chấp nhận tài liệu giấy từ các cơ quan chính phủ. David Miller, Giám đốc Dịch vụ Quản lý Hồ sơ cho các Chương trình Trung tâm Hồ sơ Liên bang của NARN, cảnh báo rằng các cơ quan chính phủ có bộ sưu tập hồ sơ cố định "cần nghĩ đến việc đưa những hồ sơ đó vào các trung tâm hồ sơ liên bang trước thời hạn hoặc tìm cách nào đó để thực hiện số hóa quy mô lớn nội dung" (Goldstein, 2012 ). Mục tiêu của quản lý hồ sơ điện tử là tính liên tục và tuân thủ. Để đảm bảo điều này, nó phải kiểm soát số lượng và chất lượng hồ sơ; đơn giản hóa việc lưu trữ/bảo trì và truy cập hồ sơ; đơn giản hóa việc định vị một bản ghi chính xác khi cần thiết; đảm bảo an ninh; xác định thời gian lưu giữ cần thiết và xây dựng hệ thống tiêu hủy hồ sơ khi hết thời gian lưu trữ.

**Quản lý hồ sơ điện tử** thiết lập các chính sách và tiêu chuẩn để duy trì các loại hồ sơ khác nhau.

Hệ thống quản lý hồ sơ điện tử (ERMS) hiệu quả thu thập tất cả dữ liệu và tài liệu kinh doanh tại điểm tiếp xúc đầu tiên của họ—Trung tâm dữ liệu, máy tính xách tay, phòng thư tín, tại địa điểm của khách hàng hoặc văn phòng từ xa. Hồ sơ xâm nhập doanh nghiệp theo nhiều cách—từ biểu mẫu trực tuyến, mã vạch, cảm biến, trang web, trang xã hội, máy sao chép, e-mail, v.v. Ngoài việc chụp toàn bộ tài liệu, dữ liệu quan trọng từ bên trong tài liệu có thể được chụp và lưu trữ trong kho lưu trữ trung tâm, có thể tìm kiếm được. Bằng cách này, dữ liệu có thể truy cập được để hỗ trợ các quyết định kinh doanh sáng suốt và kịp thời.

#### Nghĩa vụ pháp lý để lưu giữ hồ sơ kinh doanh

Các công ty cần chuẩn bị sẵn sàng để ứng phó với cuộc kiểm toán, cuộc điều tra liên bang, vụ kiện hoặc bất kỳ hành động pháp lý nào khác chống lại họ. Chỉ riêng ở Hoa Kỳ, hơn 10 luật và quy định quản lý hồ sơ liên bang phải được tuân thủ khi quản lý hồ sơ của chính phủ,

Thêm vào các quy định này trong lĩnh vực tài chính do Ủy ban Chứng khoán và Giao dịch (SEC) áp đặt, Đạo luật về Trách nhiệm Giải trình và Cung cấp Thông tin Bảo hiểm Y tế (HIPAA) trong ngành chăm sóc sức khỏe và Đạo luật về Quyền riêng tư và Quyền Giáo dục Gia đình (FERPA) trong lĩnh vực giáo dục đã được thảo luận trong CNTT tại nơi làm việc 3.5 và các tổ chức có rất nhiều tiêu chuẩn tuân thủ phải tuân theo. Các loại vụ kiện bao gồm bảo mật của bác sĩ—bệnh nhân, quyền của học sinh, quyền của phụ huynh, vi phạm bằng sáng chế, sơ suất về an toàn sản phẩm, trộm cắp tài sản trí tuệ, vi phạm hợp đồng, chấm dứt sai trái, quấy rối, phân biệt đối xử, v.v.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | CNTT tại nơi làm việc 3.5  Hệ thống quản lý hồ sơ điện tử (ERMS) đang hoạt động  Dưới đây là một số ví dụ về cách các công ty sử dụng ERMS trong lĩnh vực chăm sóc sức khỏe, tài chính và giáo dục: | |  |
| * Chăm sóc sức khỏe: Trung tâm Y tế Anne Arundel (AAMC) là một hệ thống y tế khu vực có trụ sở tại Annapolis, Maryland và phục vụ khu vực có một triệu người. Giống như nhiều hệ thống bệnh viện, AAMC đã cung cấp thông tin cho nhân viên để giúp bệnh nhân và AAMC đưa ra quyết định tốt hơn. Để làm điều này, họ đã triển khai một hệ thống hồ sơ y tế điện tử để chăm sóc bệnh nhân và thanh toán. Họ cũng sử dụng nhiều công cụ để báo cáo và tận dụng thông tin được thu thập như một phần của việc chăm sóc bệnh nhân và bệnh viện thông thường hoạt động. Sử dụng những công nghệ mới này, các nhà phân tích CNTT tại AAMC đã tích cực tạo ra các báo cáo mới. Thật không may, nhân viên thường gặp khó khăn trong việc tìm kiếm dữ liệu họ cần, * Tài chính: American Express (AMEX) sử dụng TELEform, được phát triển bởi Alchemy và Cardiff Software, để thu thập và xử lý hơn một triệu khảo sát về mức độ hài lòng của khách hàng mỗi năm. Îdữ liệu được thu thập trong các mẫu bao gồm hơn 600 biểu mẫu khảo sát khác nhau bằng 12 ngôn ngữ và 11 quốc gia. AMEX tích hợp TELEform với hệ thống cũ của AMEX, cho phép nó phân phối kết quả đã xử lý cho nhiều nhà quản lý . Do các biểu mẫu khảo sát hiện có thể truy cập dễ dàng, AMEX đã giảm số lượng nhân viên xử lý các biểu mẫu này từ 17 xuống còn 1, nhờ đó tiết kiệm cho công ty hơn 500.000 đô la một năm.      * Giáo dục: Nhiều tổ chức giáo dục đại học bao gồm Princeton, Đại học Cincinnati và Đại học Texas ở Austin lưu giữ hồ sơ sinh viên của họ, theo định nghĩa của FERPA, trong ERMS "Để tận dụng tốt nhất' nó, họ trên phương tiện truyền thông bằng văn bản điện tử, chẳng hạn như email, nhắn tin nhanh, bảng thảo luận và các phiên trò chuyện để giao tiếp với sinh viên và với nhau. Theo FERPA, sinh viên không hài lòng với các dịch vụ mà họ nhận được có thể nộp đơn khiếu nại về việc thiếu dịch vụ, thông tin không chính xác và sự thiên vị của giảng viên, sau đó yêu cầu xem hồ sơ của họ. Để giảm rủi ro về thể chế, các giảng viên, nhân viên và quản trị viên của trường đại học cố gắng tối đa hóa việc giao tiếp hiệu quả với sinh viên của họ bằng cách ghi chú điện tử cẩn thận về cách họ cung cấp dịch vụ cũng như ghi lại kiến thức và hỗ trợ mà họ đã cung cấp cho sinh viên.   Nguồn: Tổng hợp từ Green ( 2018 ), ECM. Kết nối ( 2019 ), Michalowicz ( 2019 )s Steinberg ( 2019 ), Princeton  Quản lý Hồ sơ Đại học ( 2012 ), Đại học Cincinnati ( 2019 ) và'Đại học Texas tại Austin ( 2019 )' | |

Bởi vì quản lý cấp cao phải đảm bảo rằng công ty của họ tuân thủ các nhiệm vụ pháp lý và quy định, việc quản lý hồ sơ điện tử (e-records) là một vấn đề chiến lược đối với các tổ chức cả khu vực công và tư nhân. Sự thành công của ERM phụ thuộc rất nhiều vào sự hợp tác của nhiều người chơi chính, cụ thể là quản lý cấp cao, người dùng, người quản lý hồ sơ, nhà lưu trữ, quản trị viên và quan trọng nhất là nhân viên CNTT. Được quản lý đúng cách, hồ sơ là tài sản chiến lược. Quản lý hoặc phá hủy không đúng cách, chúng trở thành nợ phải trả.

Trong những năm gần đây, các cơ quan quản lý như Hiệp hội Quản lý Thông tin và Hình ảnh (AIIM), Cơ quan Quản lý Hồ sơ và Lưu trữ Quốc gia (NARA) và ARMA International (trước đây là Hiệp hội Hồ sơ Quản lý và Quản trị viên) đã tạo và xuất bản các tiêu chuẩn ngành để quản lý tài liệu và hồ sơ. Nhiều bài báo về các phương pháp hay nhất và liên kết đến các nguồn thông tin có giá trị về quản lý hồ sơ và tài liệu có sẵn trên trang web của họ. Hộp công cụ CNTT mô tả tám khung nguyên tắc lưu giữ hồ sơ được chấp nhận rộng rãi của ARMA .

##### Hệ thống quản lý hồ sơ điện tử

Phần mềm hỗ trợ ERM được gọi là **hệ thống quản lý hồ sơ điện tử ( ERMS )**. ERMS hiện đang được triển khai rộng rãi nhất trong ba hệ thống và có thêm chức năng khắc phục thảm họa, tính liên tục trong kinh doanh và tuân thủ. Do đó, các ứng dụng của ERMS thường được tìm thấy trong các công ty được pháp luật yêu cầu lưu giữ các tài liệu, chẳng hạn như hồ sơ sức khỏe phải được lưu giữ trong một khoảng thời gian xác định, tài liệu tài chính trong khoảng thời gian 7 năm, thiết kế sản phẩm trong nhiều thập kỷ và thư điện tử về các chương trình khuyến mãi tiếp thị trong một năm. Các khả năng truy cập nhiều lớp của ERMS, giúp kiểm tra người dùng về thông tin quyền được biết của họ, đảm bảo rằng nhân viên chỉ có thể truy cập và thay đổi các tài liệu mà họ được phép xử lý.

**Hệ thống quản lý hồ sơ điện tử (ERMS)** là công cụ công nghệ được sử dụng để quản lý điện tử việc tạo và duy trì hồ sơ trong các sơ đồ phân loại, áp dụng lịch trình lưu giữ và hủy bỏ, đồng thời kiểm soát truy cập và sử dụng.

Khi các công ty chọn một ERMS để đáp ứng các yêu cầu tuân thủ, họ nên hỏi những câu hỏi sau:

Phần mềm có đáp ứng nhu cầu của tổ chức không? Ví dụ: ERMS có thể được cài đặt trên mạng hiện tại không? Nó có thể được mua như một dịch vụ?

1. Phần mềm có dễ sử dụng và có thể truy cập từ trình duyệt Web, ứng dụng văn phòng và ứng dụng e-mail không? Nếu không, mọi người Sẽ không sử dụng nó.
2. Phần mềm có giao diện người dùng đồ họa và Web nhẹ, hiện đại hỗ trợ hiệu quả cho người dùng từ xa không?
3. Trước khi chọn nhà cung cấp, điều quan trọng là phải kiểm tra quy trình công việc và cách dữ liệu, tài liệu và thông tin liên lạc lưu chuyển trong toàn công ty. Ví dụ: biết thông tin nào trên tài liệu được sử dụng trong các quyết định kinh doanh. Khi các nhu cầu và yêu cầu đó được xác định, chúng sẽ hướng dẫn lựa chọn công nghệ có thể hỗ trợ các loại đầu vào—tức là nắm bắt và lập chỉ mục chúng để chúng có thể được lưu trữ nhất quán và truy xuất theo yêu cầu.

Các phòng ban hoặc công ty có nhân viên dành phần lớn thời gian trong ngày để nộp hoặc truy xuất tài liệu hoặc lưu trữ hồ sơ giấy tờ có thể giảm chi phí đáng kể bằng ERMS. Các hệ thống này giảm thiểu sự thiếu hiệu quả và sự chậm trễ liên quan đến việc quản lý tài liệu giấy và quy trình công việc. Tuy nhiên, chúng không tạo ra một văn phòng không giấy tờ như đã được dự đoán.

Một ERMS có thể giúp doanh nghiệp trở nên hiệu quả và năng suất hơn theo những cách sau:

* Cho phép công ty tiếp cận và sử dụng nội dung trong tài liệu.



* Cắt giảm chi phí lao động bằng cách tự động hóa các quy trình kinh doanh.

• Giảm thời gian và công sức cần thiết để xác định thông tin mà doanh nghiệp cần để hỗ trợ việc ra quyết định.

* Cải thiện tính bảo mật của nội dung, từ đó giảm nguy cơ bị đánh cắp tài sản trí tuệ.
* Giảm thiểu chi phí liên quan đến in ấn, lưu trữ và tìm kiếm nội dung.

Khi quy trình công việc là kỹ thuật số, năng suất tăng, chi phí giảm, nghĩa vụ tuân thủ dễ xác minh hơn và điện toán xanh trở nên khả thi. Điện toán xanh là một sáng kiến nhằm bảo tồn các nguồn tài nguyên thiên nhiên quý giá của chúng ta bằng cách giảm tác động của việc sử dụng máy tính đối với môi trường. Bạn có thể đọc về điện toán xanh và các chủ đề liên quan đến việc giảm lượng khí thải carbon, tính bền vững và trách nhiệm đạo đức và xã hội của một tổ chức trong Chương này

###### 

###### Quản lý nội dung doanh nghiệp

Quản lý nội dung doanh nghiệp là chiếc ô theo đó quản lý tài liệu điện tử và quản lý hồ sơ điện tử tồn tại và có thể là quản lý tài liệu trên steroids. Theo một số cách, nó giống như quản lý tài liệu điện tử, là một tập hợp con của quản lý nội dung doanh nghiệp, nhưng cũng bao gồm việc xử lý nhiều loại tài sản kỹ thuật số khác nhau. Việc quản lý lượng nội dung khổng lồ được thu thập bằng phương pháp điện tử thậm chí còn mang lại nhiều giá trị gia tăng hơn khi được tích hợp và lưu trữ trong kho dữ liệu, nơi có thể dễ dàng phân tích.

**Quản lý nội dung doanh nghiệp** là việc thu thập, lưu trữ, truy vấn và quản lý một bộ tài sản kỹ thuật số đa dạng bao gồm tài liệu, hồ sơ, email, thông tin liên lạc điện tử; hình ảnh, video, flash, âm thanh và đa phương tiện.

###### Hệ thống quản lý nội dung doanh nghiệp

Phần mềm hỗ trợ quản lý nội dung doanh nghiệp được gọi là enterprise content management systems (ECMS). ECMS không chỉ là phương tiện lưu trữ và quản lý tài liệu của tổ chức. Giống như EDMS có thể được coi là tủ hồ sơ kỹ thuật số, ECMS giống một trợ lý kỹ thuật số hơn có thể tự động nhận dạng nội dung trong tài sản kỹ thuật số và biết ai, địa điểm và thời điểm gửi nội dung đó. ECMS đặc biệt hữu ích cho các công ty đang tìm cách quản lý và định cấu hình khối lượng lớn dữ liệu có cấu trúc và phi cấu trúc cũng như xử lý phương tiện thay thế. Mục tiêu của nó là kinh doanh liên tục.

**Hệ thống quản lý nội dung doanh nghiệp (ECMS)** nắm bắt, lưu giữ và quản lý nhiều loại tài sản kỹ thuật số có cấu trúc và không có cấu trúc, đồng thời bảo mật chúng bằng kỹ thuật số tuân thủ các chính sách.

##### Chọn EDMS, ERMS hoặc ECMS

Khi chọn EDMS, ERMS hoặc ECMS, điều quan trọng là phải phù hợp với nhu cầu chiến lược của công ty với khả năng của hệ thống. Để đưa ra lựa chọn sáng suốt, bạn phải hiểu những điểm tương đồng và khác biệt về chức năng giữa ba loại.

Như bạn đã đọc, mục tiêu chung của quản lý tài liệu điện tử, quản lý hồ sơ điện tử và quản lý nội dung doanh nghiệp là **hoạt động kinh doanh liên tục** . Nếu một trong số chúng không thành công, toàn bộ tổ chức có thể gặp rủi ro. Tuy nhiên„khi tất cả đều đạt được mục tiêu của mình, tương lai của tổ chức được đảm bảo.

**Tính liên tục trong kinh doanh** là khả năng của một công ty duy trì các chức năng thiết yếu trong cũng như sau khi thảm họa xảy ra.

Sự khác biệt quan trọng giữa EDMS và ERMS tồn tại bởi vì người dùng EDMS cần đăng ký và kiểm tra các tài liệu được lưu trữ một cách nhanh chóng và dễ dàng, đồng thời mở khóa chúng để sửa đổi trong tương lai, đồng thời duy trì theo dõi phiên bản và lịch sử truy cập.

Mặt khác, người dùng ERMS yêu cầu hồ sơ phải được lưu giữ ở định dạng ban đầu để truy xuất vì lý do tuân thủ hoặc pháp lý. Chỉ lưu trữ tệp trên ổ cứng được chia sẽ của một tổ chức là không đủ để đáp ứng các tiêu chuẩn tuân thủ ngành này.

Một điểm khác biệt nữa là trong một ERMS„ các bản ghi cần được đặt trong các chỉ mục được xác định bởi các Rile bên ngoài, thường được áp đặt bởi các cơ quan chính phủ hoặc cơ quan quản lý. Do đó, việc lưu trữ này thường được sao chép ở cả trung tâm hồ sơ tại chỗ và bên ngoài vì lý do an toàn vì bảo mật phần mềm là điều cần thiết trong ERMS, nhưng chỉ mong muốn trong EDMS.

Một ECMS có một nhiệm vụ rộng lớn hơn. Nó khác với EDMS và ERMS ở chỗ nó được thiết kế để hỗ trợ quản lý vòng đời bản ghi, quản trị thông tin và cộng tác. Một ECMS cũng cung cấp một nguồn thông tin đáng tin cậy duy nhất. Nó cũng cung cấp khả năng kiểm soát và đồng bộ hóa phiên bản, cùng với các chức năng tìm kiếm và khám phá trực quan. Một EÇMS cũng có thể tích hợp với các ứng dụng khác để cung cấp nội dung có thể xem được trên nhiều nền tảng, cùng với các quy trình kinh doanh được tối ưu hóa và bảo mật trong hệ thống lưu trữ và truy xuất nội dung của nó (Trust Radius, 2019).

Các công cụ EDMS, ERMS và ECMS chính là phần mềm luồng công việc, công cụ soạn thảo, máy quét và cơ sở dữ liệu: Các công cụ này có khả năng truy vấn và tìm kiếm để tài liệu có thể được xác định và truy cập giống như dữ liệu trong cơ sở dữ liệu, các bản ghi có thể được lưu trữ và truy xuất dễ dàng, và nội dung của tất cả các loại có thể được thực hiện dễ dàng truy cập theo yêu cầu. Các hệ thống bao gồm từ những hệ thống được thiết kế để hỗ trợ một nhóm làm việc nhỏ đến các hệ thống toàn doanh nghiệp hỗ trợ Web, đầy đủ tính năng.

Ví dụ về các hệ thống quản lý nội dung, hồ sơ và tài liệu điện tử được sử dụng rộng rãi bao gồm IBM e-File

Laserfiche (https•J/www.laserfiche.com/), eFile Cabinet (https://www.efilecabinet.com/), M-Files(liitps://www.mfiles.com/en) và NAVEX Global ( https://www.navexglobal.com/em.us)

##### câu hỏi

1. Định nghĩa thuật ngữ "nội dung điện tử,"

Mô tả sự khác biệt chính giữa tài liệu điện tử và hồ sơ điện tử.

Tại sao các công ty sử dụng hệ thống quản lý hồ sơ điện tử?

Kể tên ba loại vụ kiện có thể được đưa ra chống lại một công ty liên quan đến hồ sơ điện tử của họ.

Đặc điểm nào của ERMS cho phép nhân viên chỉ truy cập và thay đổi các tài liệu mà họ được phép xử lý?

Sự khác biệt giữa hệ thống quản lý tài liệu điện tử (EDMS) và hệ thống quản lý hồ sơ điện tử (ERMS) là gì?

Mục tiêu chung của EDMS, ERMS và ECMS là gì?

### Tóm tắt chương

LO 3.1 Mô tả các nguyên tắc cơ bản của quản lý dữ liệu và cách các hệ thống quản lý cơ sở dữ liệu, bao gồm cả công nghệ chuỗi khối, giúp các công ty cải thiện hiệu suất.

Quản lý dữ liệu giám sát vòng đời của dữ liệu từ đầu đến cuối từ khi tạo và lưu trữ ban đầu cho đến khi dữ liệu trở nên lỗi thời và bị xóa. Do khối lượng dữ liệu đáng kinh ngạc mà tổ chức điển hình tạo ra, việc quản lý dữ liệu hiệu quả là rất quan trọng để ngăn chi phí lưu trữ vượt khỏi tầm kiểm soát và kiểm soát tốc độ tăng trưởng dữ liệu đồng thời hỗ trợ hiệu suất cao hơn. Một số công nghệ quản lý dữ liệu cập nhật thông tin cho người dùng và hỗ trợ các nhu cầu kinh doanh khác nhau. Chúng bao gồm một cơ sở dữ liệu, hệ thống quản lý cơ sở dữ liệu (DBMS), kho dữ liệu¶ancl data marts. Cơ sở dữ liệu lưu trữ dữ liệu được tạo bởi các ứng dụng kinh doanh, cảm biến, hoạt động và TPS. Các doanh nghiệp vừa và lớn thường có cơ sở dữ liệu tập trung hoặc phân tán. Để tận dụng tối đa cơ sở dữ liệu, phải sử dụng gói phần mềm được thiết kế để xác định, thao tác và quản lý dữ liệu. Đây được gọi là hệ thống quản lý cơ sở dữ liệu (DBMS). Các hệ thống quản lý cơ sở dữ liệu mang lại lợi ích cho các tổ chức bằng cách giảm độ trễ, tính không ổn định và tính nhất quán của dữ liệu. Một DBMS được thiết kế tốt cũng có thể giảm thời gian phản hồi truy vấn, quản lý các ot4ueries có khả năng dự đoán và xử lý chúng trong thời gian thực để cho phép ra quyết định nhanh hơn, tự tin hơn. Mục nhập mới nhất vào DBMS là công nghệ chuỗi khối.

Chuỗi khối bổ sung tính bảo mật và độ tin cậy cao hơn nhiều cho cơ sở dữ liệu thông qua cấu trúc, quyền và hợp đồng thông minh bất biến của nó. Công nghệ chuỗi khối đang được sử dụng theo những cách sáng tạo trong nhiều lĩnh vực công nghiệp.



LO 3.2Xác định sự khác biệt giữa cơ sở dữ liệu, kho dữ liệu/•và siêu thị dữ liệu và lý do tại sao một công ty sẽ chuyển từ cơ sở dữ liệu sang kho dữ liệu hoặc siêu thị dữ liệu.

Cơ sở dữ liệu hoạt động tốt với các bộ åata nhỏ hơn nhưng thiếu khả năng mở rộng hoặc tính linh hoạt đối với khối lượng công việc chuyên biệt hoặc rất lớn . Khi một tổ chức01là khối lượng dữ liệu rất lớn@phức tạp, nó cần xem xét kho dữ liệu hoặc data marti Kho dữ liệu và data mart được sử dụng khi tập dữ liệu không thể xử lý dữ liệu một cách hiệu quả và sự biến động của cơ sở dữ liệu khiến nó không thể sử dụng nó cho phức hợp ra quyết định và giải quyết vấn đề. Nhiều tổ chức đã xây dựng kho dữ liệu vì họA,rất thất vọng với dữ liệu không nhất quán không thể hỗ trợ các quyết định của công ty và tương tác của khách hàng Kho dữ liệu hoặc siêu thị dữ liệu là nguồn chính của dữ liệu đã được làm sạch để phân tích, báo cáo và ra quyết định. Chi phí cao của kho dữ liệu có thể khiến chúng trở nên quá đắt đối với các công ty nhỏ hơn để triển khai.

Data mart là phiên bản thu nhỏ, chi phí thấp hơn của kho dữ liệu có thể được triển khai trong khung thời gian ngắn hơn. Kho dữ liệu và kho dữ liệu ngày cải thiện trải nghiệm chung của khách hàng và có thể được gửi đến những người ra quyết định thông qua doanh nghiệp thông qua đám mây hoặc mạng nội bộ do công ty sở hữu để đưa ra quyết định nhanh hơn, chính xác hơn.



LO 3.3 Mô tả tầm quan trọng của quản trị dữ liệu trong việc cung cấp dữ liệu mờ có sẵn khi nào và ở đâu cần thiết và tại sao quản lý dữ liệu chủ (MDM) là một sáng kiến quản trị dữ liệu quan trọng.

Quản trị dữ liệu là một dự án toàn doanh nghiệp vì dữ liệu vượt qua các ranh giới và được mọi người sử dụng trong suốt ntarr-.r\co đến Kt•cinac

quản trị dữ liệu mạnh mẽ và MDM là cần thiết để đảm bảo dữ liệu có đủ chất lượng nhằm đáp ứng nhu cầu kinh doanh Các ngành công nghiệp phụ thuộc rất nhiều vào quản trị dữ liệu để giúp họ tuân thủ các quy định hoặc báo cáo MDM là một phần quan trọng của quản trị dữ liệu.

MDM cung cấp một giải pháp để quản lý bị thất vọng bởi các nguồn dữ liệu bị phân tán và phân tán, Quản trị dữ liệu kết hợp với MDM là một sự kết hợp mạnh mẽ. Nhu cầu quản lý dữ liệu như một tài sản chiến lược để mở khóa toàn bộ giá trị của nó ngày càng trở nên nhiều hơn và khối lượng dữ liệu tiếp tục tăng. Lợi ích của quản trị dữ liệu và MDM bao gồm cải thiện việc tuân thủ quy định, ra quyết định hiệu quả hơn, hiểu rõ hơn về dữ liệu và tăng doanh thu,

LO 3.4 Giải thích khái niệm quản lý thông tin và lý do tại sao điều quan trọng đối với một tổ chức là tuân theo các thông lệ quản lý thông tin tốt.

Quản lý thông tin là quá trình thu thập, lưu trữ, quản lý và duy trì dữ liệu chính xác, kịp thời, đáng tin cậy, hợp lệ, sẵn có, duy nhất và phù hợp với một tổ chức.

Trong kinh doanh, dữ liệu được sử dụng vì hai lý do chính: để điều hành một doanh nghiệp và để cải thiện một doanh nghiệp. Cả hai điều này đều phụ thuộc vào khả năng truy cập dễ dàng và nhanh chóng vào dữ liệu đáng tin cậy chất lượng cao, chính xác, kịp thời, đáng tin cậy, duy nhất và có liên quan.

Do đó, hai sáng kiến quan trọng của quản lý thông tin là loại bỏ các silo dữ liệu và giảm "dữ liệu bẩn" là dữ liệu không chính xác, bị thiếu hoặc lỗi thời được lưu trữ ở định dạng dữ liệu không chuẩn và được truy cập thông qua các giao diện được thiết kế kém. Dữ liệu bẩn có thể khiến công ty gặp rủi ro và rất tốn kém để duy trì dẫn đến mất việc kinh doanh, tốn thời gian cho lỗi và thời gian sửa lỗi. Các silo dữ liệu tạo ra các rào cản đối với sự cộng tác và giao tiếp và có thể dẫn đến việc ra quyết định kém dựa trên thông tin không đầy đủ. Quản lý thông tin tốt có thể cải thiện chất lượng quyết định, độ chính xác và độ tin cậy của các dự đoán quản lý. Nó cũng có thể giảm thời gian và chi phí định vị và tích hợp thông tin liên quan và dẫn đến chất lượng dữ liệu tốt hơn.

LO 3.5Xác định các thuật ngữ tài liệu điện tử, hồ sơ điện tử và nội dung điện tử và cách các công nghệ quản lý chúng đang được sử dụng để làm cho các tổ chức hiệu quả và hiệu quả hơn.

Một tài liệu điện tử đề cập đến một tài liệu đang được làm việc tích cực và có thể được thay đổi. Mặt khác, một bản ghi điện tử là một tài liệu hoàn chỉnh không có ý định thay đổi và nội dung điện tử có bản chất tương tự, nhưng cũng bao gồm các tệp video, hình ảnh, âm thanh, flash và đa phương tiện. Tài liệu điện tử được quản lý bởi hệ thống quản lý tài liệu điện tử (EDMS) để tạo, chia sẻ, sắp xếp và truy xuất hiệu quả các tài liệu sống một cách có hệ thống„ EDMS được sử dụng bởi nhiều người dùng trong một tổ chức để thực hiện công việc hàng ngày và có thể được coi là tủ hồ sơ kỹ thuật số giúp người dùng dễ dàng tìm kiếm tài liệu mọi lúc, mọi nơi. Hệ thống quản lý hồ sơ điện tử (ERMS) thì khác. Nó tập trung vào việc lưu trữ hồ sơ để giúp đáp ứng các yêu cầu tuân thủ và quy định do kiểm toán viên và các cơ quan quản lý khác áp đặt. Ngoài ra, nó: đơn giản hóa việc lưu trữ, bảo trì và truy cập vào các hồ sơ quan trọng, xác định lịch trình thời gian lưu giữ và phát triển một hệ thống tiêu hủy để thu hồi vào cuối vòng đời của chúng. Quản lý nội dung là cái ô dưới đó quản lý tài liệu điện tử và quản lý hồ sơ điện tử. Về bản chất, nó tương tự như quản lý tài liệu điện tử nhưng bao gồm việc xử lý nhiều loại tài sản kỹ thuật số khác nhau. Hệ thống quản lý nội dung doanh nghiệp (ECMS) cung cấp giá trị ngoài việc lưu trữ và quản lý tài liệu của tổ chức. Nó hoạt động như một trợ lý kỹ thuật số tự động nhận dạng nội dung trong tài sản kỹ thuật số và biết gửi cho ai, ở đâu và khi nào. ECMS đặc biệt có giá trị đối với các tổ chức có khối lượng lớn dữ liệu có cấu trúc và phi cấu trúc trong một số phương tiện kỹ thuật số thay thế phải được quản lý và định cấu hình.

### Thuật ngữ ….

### Đảm bảo việc học của bạn

#### Thảo luận: Câu hỏi Tư duy Phản biện

1, Nêu chức năng của cơ sở dữ liệu và kho dữ liệu?

1. Chất lượng dữ liệu ảnh hưởng đến hiệu quả kinh doanh như thế nào?
2. Liệt kê ba loại lãng phí hoặc thiệt hại mà lỗi dữ liệuQan gây ra.
3. Vai trò của một tập tin chính là gì?
4. Đưa ra ba ví dụ về các hoạt động của quy trình kinh doanh sẽ được hưởng lợi đáng kể từ việc có dữ liệu chi tiết theo thời gian thực hoặc gần thời gian thực và xác định các lợi ích.
5. Kể tên hai ngành phụ thuộc vào quản trị dữ liệu để tuân thủ các quy định hoặc yêu cầu báo cáo. Đưa ra một ví dụ về mỗi.

7. Chọn một cách cẩn thận. Giải thích cách một tổ chức trong ngành đó có thể cải thiện sự hài lòng của người tiêu dùng thông qua việc sử dụng kho dữ liệu.

1. Giải thích nguyên tắc sử dụng dữ liệu 90/90.
2. Tại sao quản lý dữ liệu chủ (MDM) lại quan trọng trong các công ty có nhiều nguồn dữ liệu?
3. Kể tên một thách thức mà các tổ chức phải đối mặt về quản lý thông tin
4. Lợi ích hoạt động và lợi thế cạnh tranh của việc sử dụng quản lý hồ sơ và tài liệu điện tử là gì?
5. Làm thế nào một hệ thống quản lý tri thức nhận thức có thể giảm chi phí hoạt động và doanh thu?

#### Khám phá: Bài tập trực tuyến và tương tác

1. Nhiều tổ chức bắt đầu các chương trình quản trị dữ liệu vì các vấn đề cấp bách về tuân thủ ảnh hưởng đến việc sử dụng dữ liệu e. Hoặc các tổ chức cần kiểm soát dữ liệu để tuân thủ một hoặc nhiều quy định, chẳng hạn như

Đạo luật Gramm—Leach Bliley (GLB); HIPAA, Đạo luật chống côn trùng ở nước ngoài (FCPA), Đạo luật Sarbanes—Oxley và một số luật về quyền riêng tư của tiểu bang và liên bang.

1. Nghiên cứu và chọn hai quy định hoặc luật riêng tư của Hoa Kỳ.
2. Mô tả cách quản trị dữ liệu sẽ giúp doanh nghiệp tuân thủ các quy định hoặc luật này



2. Nghiên cứu hai nhà cung cấp dịch vụ quản lý hồ sơ điện tử, chẳng hạn như Iron Mountain.

1. Các khuyến nghị duy trì được thực hiện bởi các nhà cung cấp là gì? Tại sao?
2. Những dịch vụ hoặc giải pháp nào mà mỗi nhà cung cấp cung cấp?

3. Xem video Microsoft QuickBase tại https://www.quickbase.conv/video-demo-watch. Giải thích các lợi ích của ứng dụng thay thế Microsoft Access này và giải thích liệu bạn có nghĩ rằng QuickBase là một cải tiến so với ứng dụng trước đó hay không.

#### Phân tích & Quyết định: Áp dụng các khái niệm CNTT vào các quyết định kinh doanh



1. Công ty Spring Street (SSC) muốn giảm "chi phí ẩn" liên quan đến các quy trình sử dụng nhiều giấy tờ của mình. Các nhân viên dự đoán đùa rằng nếu cửa sổ mở vào một ngày nhiều gió, thì sự hỗn loạn hoàn toàn sẽ xảy ra khi hàng nghìn tờ giấy bắt đầu bay lên. Nếu một ngày lũ lụt, hỏa hoạn hoặc gió lớn xảy ra, công việc kinh doanh sẽ bị đình trệ theo đúng nghĩa đen . Nhân viên kế toán của công ty , Sam Spring, đã quyết định tính toán chi phí. Ông nhận ra rằng một số nhân viên đã dành phần lớn thời gian trong ngày của họ để nộp hoặc lấy tài liệu. Ngoài ra, còn có chi phí hàng tháng để lưu trữ hồ sơ giấy cũßam đã đo lường các hoạt động liên quan đến việc xử lý các báo cáo in và tệp giấy. ì Ước tính trung bình của anh ấy như sau;
   1. Xử lý hồ sơ: Nhân viên phải mất 12 phút để đi bộ đến phòng hồ sơ, xác định vị trí hồ sơ, xử lý hồ sơ, lọc lại hồ sơ và quay trở lại bàn làm việc của mình. Nhân viên làm việc này 4 lần mỗi ngày (năm ngày mỗi tuần).
   2. Số lượng nhân viên: 10 nhân viên chính thức thực hiện các chức năng.
   3. Thay thế tài liệu bị mất: Mỗi ngày một lần, một tài liệu bị "mất" (bị mất, bị thất lạc hoặc bị dính nhiều vết cà phê) và phải được tạo lại„'Tổng chi phí để thay thế mỗi tài liệu bị mất là 200 đô la.
   4. Chi phí kho bãi: Hiện tại, chi phí lưu trữ tài liệu là $ tháng.

sam thích một hệ thống cho phép nhân viên tìm và làm việc với tài liệu mà không cần rời khỏi bàn làm việc. Anh ấy quan tâm nhất đến bộ phận nhân sự và kế toán. Những nhân viên này là những người sử dụng nhiều tệp giấy theo truyền thống và sẽ được hưởng lợi rất nhiều từ hệ thống quản lý tài liệu hiện đại. Tuy nhiên, đồng thời, Sam cũng không thích rủi ro, Anh ấy muốn đầu tư vào các giải pháp giúp giảm rủi ro về chi phí cao hơn trong tương lai. công nhận rằng các yêu cầu của PATRIOT Aces của Hoa Kỳ rằng các tổ chức cung cấp cho chính phủ quyền truy cập ngay lập tức vào hồ sơ áp dụng cho SSC. Anh ấy đã đọc được rằng các tổ chức sản xuất và chính phủ dựa vào việc quản lý tài liệu hiệu quả để đáp ứng 'những yêu cầu pháp lý rộng hơn này'. Cuối cùng, Sam muốn thực hiện một hệ thống khắc phục thảm họa



Chuẩn bị một bản báo cáo cung cấp cho Sam dữ liệu mà anh ấy cần để đánh giá phương pháp quản lý tài liệu tốn nhiều giấy tờ của công ty. Bạn sẽ cần tiến hành nghiên cứu để cung cấp dữ liệu nhằm chuẩn bị báo cáo này. Báo cáo của bạn nên bao gồm các thông tin sau:

* 1. UBCKNN nên chuẩn bị như thế nào cho một ERMS nếu họ quyết định triển khai?
  2. Sử dụng dữ liệu do Sam thu thập, hãy tạo một bảng tính tính toán chi phí xử lý giấy tờ tại SSC dựa trên mức lương trung bình mỗi giờ cho mỗi nhân viên của S28. Thêm chi phí của các tài liệu bị mất vào đây. Sau đó, thêm chi phí nhập kho giấy, tăng 10% mỗi tháng do khối lượng tăng. Trình bày kết quả hiển thị cả tổng số hàng tháng và tổng số hàng năm. Chuẩn bị các biểu đồ để Sam có thể dễ dàng xác định mức tăng dự kiến về chi phí lưu kho trong ba năm tới.
  3. Làm thế nào một ERMS cũng có thể phục vụ như một hệ thống khắc phục thảm họa trong trường hợp hỏa hoạn, lũ lụt hoặc đột nhập?
  4. Gửi đề xuất của bạn cho một giải pháp ERM. Xác định hai nhà cung cấp trong

1. Bạn đang làm việc cho chuỗi cửa hàng quần áo quốc gia. Một số người mua đã gặp rất nhiều vấn đề trong việc lấy các báo cáo được thiết kế tốt từ bộ phận bán hàng và tiếp thị của tổ chức bạn về các hoạt động của họ bằng cách



nhà nước trong năm qua. Họ muốn có thể tạo các báo cáo khi cần và truy cập chúng mọi lúc, mọi nơi. Người mua cần kiến thức này để giúp họ lập kế hoạch mua hàng trong tương lai và nghĩ rằng hệ thống quản lý kiến thức sẽ giúp cải thiện việc truy xuất thông tin từ kho dữ liệu của công ty. Bạn đã được yêu cầu nghiên cứu các gói phần mềm KM và báo cáo lại cho đồng nghiệp của mình. Xem danh sách 10 KM hàng đầu gói phần mềm tại trang web Captena (https://www.capterra.com/sem-compare/knowledge- management-software?utm medium—cpc). Chọn ít nhất ba công cụ phần mềm KM từ danh sách và truy cập trang web của họ và thực hiện bản demo. Viết báo cáo mô tả và so sánh các công cụ KM mà bạn đã demo. Thảo luận về những ưu và nhược điểm của từng người trong số họ và xác định yêu thích của bạn. Giải thích sự lựa chọn của bạn.

Củng cố: Thể hiện sự hiểu biết về các thuật ngữ chính

Giải ô chữ trực tuyến để kiểm tra sự hiểu biết của bạn về các Thuật ngữ chính trong chương này.

#### Tài nguyên web

Nhiều tài nguyên và công cụ học tập có trên trang web của sinh viên. Bạn sẽ tìm thấy các liên kết web hữu ích và các câu đố tự kiểm tra cung cấp phản hồi cho từng cá nhân.

## Trường hợp 3.2

### Trường hợp kinh doanh: Dữ liệu bẩn gây nguy hiểm cho nỗ lực nâng cao niềm vui của trường đại học

Được thành lập vào năm 1861, Đại học Washington (UW) là một trường đại học nghiên cứu công lập có nhiều cơ sở tọa lạc tại Seattle, Tacoma và Bothell, đồng thời là một trung tâm y tế học thuật đẳng cấp thế giới. Trường có tổng số hơn 47.000 sinh viên theo học tại 16 trường cao đẳng và trung học. Nó cung cấp 1.800 khóa học đại học mỗi quý cho hơn 32.000 sinh viên và hàng năm trao hơn 12.000 bằng cử nhân, cao học, tiến sĩ và chuyên môn thông qua 140 khoa học thuật.

#### Cảnh giác với dữ liệu bẩn

Các trường đại học như UW biết rằng việc gây quỹ có nghĩa là số tiền LỚN: Tuy nhiên , dữ liệu bẩn đã cản trở nỗ lực gây quỹ của họ và thay vì là một cơ hội quan trọng, họ lại có một công thức dẫn đến thảm họa: UW thu thập thông tin liên hệ từ nhiều nguồn và điều này đã dẫn đến nhiều lỗi dữ liệu và các bản ghi trùng lặp trong cơ sở dữ liệu tài trợ khổng lồ của họ chứa tên, địa chỉ và thông tin liên quan khác về hơn 900.000 sinh viên, giảng viên, nhân viên, cựu sinh viên và những người tham dự sự kiện thể thao. Họ cũng biết rằng thông tin liên hệ trong hồ sơ nhà tài trợ của UW nhanh chóng trở nên lỗi thời, đặc biệt là đối với sinh viên"và cựu sinh viên trẻ tuổi, những người có xu hướng nhất thời (chuyển công việc mới, kết hôn—đổi tên). dữ liệu dẫn đến một khối lượng lớn không gửi được mai sau lãng phí bưu phí và chi phí sản xuất quá mức, và tệ nhất là mất cơ hội gây quỹ. Mike Visaya, Phó Giám đốc Sáng kiến Công nghệ Chiến lược Quản lý Thông tin (1M), than thở rằng vì lượng dữ liệu khổng lồ mà họ phải xử lý UW "biết rằng điều quan trọng là phải thuê một chuyên gia về chất lượng dữ liệu. "

#### Cải thiện chất lượng dữ liệu

Để cải thiện chất lượng cơ sở dữ liệu của nhà tài trợ, trường Đại học đã làm việc với Melissa Data

(https://www.melissa.com/), một nhóm thu thập và xác minh dữ liệu. Họ đã cùng nhau thực hiện một chương trình chất lượng dữ liệu chủ động mở rộng để giải quyết các vấn đề về dữ liệu không chính xác và đảm bảo tính nhất quán của hồ sơ sinh viên và cựu sinh viên trong cơ sở dữ liệu nhà tài trợ chính của họ. Sử dụng Bộ chất lượng dữ liệu của Melissa, UW đã cập nhật cách xử lý xác minh địa chỉ, điện thoại và email cũng như phân tích tên theo cách khiến Mike Visaya tuyên bố rằng "bây giờ dữ liệu của chúng tôi có ý nghĩa." Để tìm và ngăn trùng lặp dữ liệu, UW đã sử dụng các giải pháp sao chép và khớp bản ghi nâng cao của Melissa Data.



Melissa Data Ý cũng giúp UW phân tích các mối quan hệ theo vị trí cụ thể trong dữ liệu của mình. Ví dụ, UW muốn biết liệu những người sở hữu vé mùa bóng đá của họ có phải là những người đóng góp lớn cho các chương trình học thuật của trường đại học hay không và đã rất ngạc nhiên khi phát hiện ra rằng họ là những người đóng góp lớn. Shawn Drew, Giám đốc 1M của Văn phòng

Phát triển, đã rất ấn tượng. Anh ấy nghĩ rằng các giải pháp cơ sở dữ liệu cho phép UW "kết nối các dấu chấm, tìm các mối quan hệ mới và tận dụng nhiều cách mà những người ủng hộ chúng tôi muốn đóng góp."

#### Xác định Cơ hội Mới

Những cải tiến về chất lượng dữ liệu cũng cho phép Trường có cơ hội theo đuổi các nhà tài trợ quốc tế.

Trước khi triển khai giải pháp, Trường rất ít gửi thư quốc tế. Giờ đây, với công nghệ mới của họ, UW có thể dễ dàng chuẩn hóa và thu thập các liên hệ quốc tế để tăng cường nỗ lực gây quỹ toàn cầu của họ.

|  |
| --- |
| Nhờ giải pháp chất lượng dữ liệu mới của họ, UW đã vượt qua cơn ác mộng về chất lượng dữ liệu bẩn và mất cơ hội gây quỹ để thực hiện ước mơ của mình và huy động được số tiền khổng lồ 2,7 tỷ đô la trong một nỗ lực gây quỹ.  câu hỏi   1. Tại sao lại có dữ liệu bẩn trong cơ sở dữ liệu của UW? 2. Hậu quả đối với UW của dữ liệu bẩn là gì? 3. UW đã giải quyết vấn đề dữ liệu bẩn như thế nào? 4. Những lợi ích mà họ đã trải nghiệm từ việc cải thiện chất lượng cơ sở dữ liệu nhà tài trợ của họ là gì? 5. Cơ hội mới nào đã được xác định bởi hệ thống mới?   Nguồn: Dữ liệu Melissa ( 2019 ) và Đại học Washington ( 2019 ). |
| Hộp công cụ CNTT  Khuôn khổ cho các nguyên tắc lưu trữ hồ sơ được chấp nhận chung  Khuôn khổ về các nguyên tắc lưu trữ hồ sơ được chấp nhận rộng rãi là một công cụ hữu ích để quản lý hồ sơ kinh doanh nhằm đảm bảo rằng chúng hỗ trợ quy định hiện tại và tương lai của doanh nghiệp; các yêu cầu pháp lý, giảm thiểu rủi ro, môi trường và vận hành. Khung này bao gồm tám nguyên tắc hoặc phương pháp hay nhất, cũng hỗ trợ quản trị dữ liệu. Các nguyên tắc này được tạo ra bởi ARMAIntemationaI (trước đây là Hiệp hội các nhà quản lý và quản lý hồ sơ) và các chuyên gia CNTT và pháp lý.   1. Nguyên tắc trách nhiệm Chỉ định một giám đốc điều hành cấp cao giám sát các chính sách và thủ tục lưu giữ hồ sơ để hướng dẫn nhân viên; và đảm bảo khả năng kiểm toán chương trình. 2. Nguyên tắc minh bạch Ghi lại các quy trình và hoạt động của chương trình lưu trữ hồ sơ của tổ chức theo cách dễ hiểu và có sẵn cho tất cả nhân viên và các bên thích hợp   Nguyên tắc toàn vẹn Đảm bảo chương trình lưu trữ hồ sơ có thể đảm bảo hợp lý tính xác thực và độ tin cậy của hồ sơ và dữ liệu.   1. Nguyên tắc bảo vệ Xây dựng chương trình lưu trữ hồ sơ để đảm bảo mức độ bảo vệ hợp lý đối với hồ sơ và thông tin mang tính riêng tư, bí mật, bí mật đặc quyền hoặc cần thiết cho hoạt động kinh doanh liên tục.      1. Nguyên tắc tuân thủ Đảm bảo 'chương trình lưu giữ hồ sơ tuân thủ luật pháp, cơ quan có thẩm quyền hiện hành và chính sách của tổ chức. 2. Nguyên tắc về tính khả dụng Duy trì hồ sơ một cách linh hoạt, đảm bảo truy xuất thông tin cần thiết kịp thời, hiệu quả và chính xác.     Nguyên tắc lưu giữ Duy trì hồ sơ và dữ liệu trong thời gian thích hợp dựa trên các      yêu cầu pháp lý, quy định, tài chính và lịch sử,      8. Nguyên tắc xử lý Xử lý hồ sơ một cách an toàn khi luật pháp hoặc chính sách của tổ chức không còn yêu cầu lưu giữ chúng nữa. |