

22110377_Huỳnh Minh Mẫn

Module 12: Internet of Things

1. Technology terminology

- Chia sẻ dữ liệu: Khi dữ liệu được gửi từ một nơi này đến một nơi khác hoặc khi quyền truy cập được cấp cho nhiều người dùng.
- Internet vạn vật (IoT): Mạng lưới của các thiết bị hàng ngày được tích hợp với cảm biến và công nghệ truyền thông để gửi và nhận dữ liệu qua internet. Nói một cách đơn giản, IoT bao gồm nhiều thiết bị khác nhau kết nối với internet, chẳng hạn như điện thoại thông minh, đồng hồ, thiết bị gia dụng và thậm chí cả xe hơi.
- Cảm biến: Thiết bị điện tử phát hiện các thành phần bên ngoài.

2. Background and misconceptions

- Internet vạn vật (IoT) là thuật ngữ dùng để mô tả mạng lưới các thiết bị kết nối có khả năng phối hợp cùng nhau để cải thiện dịch vụ cho người tiêu dùng và doanh nghiệp. Các dịch vụ đám mây IoT ghi nhận và phân tích dữ liệu từ các cảm biến trên các thiết bị kết nối để cải thiện chức năng, tăng cường hiệu suất, hoặc cung cấp trợ giúp cho người dùng.
- Các loại thiết bị trong IoT rất đa dạng, bao gồm điện thoại, đồng hồ, ô tô, máy móc và thiết bị gia dụng. Bất kỳ thiết bị nào trong số này đều có thể được trang bị cảm biến và kết nối với đám mây. Cảm biến ghi lại thông tin về hoạt động của thiết bị và truyền thông tin đó lên đám mây. Sau đó, các thuật toán phân tích nâng cao được sử dụng để đưa ra quyết định, đề xuất, hoặc thay đổi chức năng của thiết bị.
- Các ví dụ mà bạn có thể quen thuộc bao gồm các thiết bị liên kết, tính năng nhà thông minh và thiết bị đeo. Bạn có thể đeo một thiết bị sử dụng cảm biến để thu thập dữ liệu về sức khỏe cá nhân của bạn. Ngoài ra, bạn cũng có thể quen thuộc với các ứng dụng điện thoại chia sẻ dữ liệu vị trí và cho phép thành viên trong gia đình theo dõi vị trí của nhau. Các ví dụ phức tạp hơn bao gồm hệ thống gia đình kết nối các thiết bị giải trí, đèn, an ninh hoặc thiết bị gia dụng. Các sản phẩm tiêu dùng này chỉ là một phần nhỏ của IoT—còn có rất nhiều ứng dụng trong công nghiệp và thương mại.
- IoT xử lý lượng lớn dữ liệu và thiết bị từ người tiêu dùng và doanh nghiệp. Do đó, các vấn đề về an ninh và quyền riêng tư cần được xem xét cùng với sự gia tăng sử dụng.

dụng IoT. Các công ty an ninh mạng thường chuyên bảo vệ doanh nghiệp khỏi các cuộc tấn công mạng qua IoT và bảo mật dữ liệu khách hàng.

- AWS IoT cung cấp phần mềm thiết bị, dịch vụ kiểm soát và dịch vụ dữ liệu. Phần mềm thiết bị cho phép bạn kết nối an toàn các thiết bị, thu thập dữ liệu và thực hiện các hành động thông minh tại chỗ, ngay cả khi không có kết nối internet. Các dịch vụ kiểm soát cho phép bạn kiểm soát, quản lý và bảo mật các đội thiết bị lớn và đa dạng. Dịch vụ dữ liệu giúp bạn khai thác giá trị từ dữ liệu IoT.
 - Dịch vụ công nghiệp: Khách hàng AWS IoT đang xây dựng các ứng dụng IoT công nghiệp cho chất lượng dự đoán và bảo trì, cũng như để giám sát hoạt động từ xa.
 - Dịch vụ nhà kết nối: Khách hàng AWS IoT đang xây dựng các ứng dụng nhà kết nối cho tự động hóa nhà ở, an ninh và giám sát nhà, và mạng nhà.
 - Thương mại: Khách hàng AWS IoT đang xây dựng các ứng dụng thương mại cho giám sát giao thông, an toàn công cộng và giám sát sức khỏe.
- Một số người có thể nhầm lẫn IoT với lập trình cá nhân hóa hoặc việc cảm nhận hoặc lập trình cho từng thiết bị riêng lẻ. IoT đề cập đến sự tương tác kết nối giữa các thiết bị. IoT không đơn giản chỉ là chức năng thông minh của một thiết bị mà còn bao gồm khả năng của thiết bị trong việc giao tiếp với các nút khác trên đám mây và sử dụng, kết nối dữ liệu trên toàn mạng lưới.

3. Question

1. Do you have or want a smart watch? In what way is a smart watch connected to other devices? Do the devices share any data?

Smart watch kết nối với điện thoại qua Bluetooth hoặc Wi-Fi, chia sẻ dữ liệu như nhịp tim, bước chân, thông báo và vị trí. Dữ liệu này có thể được đồng bộ hóa lên đám mây, giúp người dùng truy cập từ nhiều thiết bị khác nhau. Cloud có thể phân tích dữ liệu để đưa ra khuyến nghị sức khỏe và cải thiện trải nghiệm người dùng.

2. Think about some of the appliances in your home: washer and dryer, refrigerator, heater, security camera, and the like. Some newer appliances are connected to the cloud. What kind of information might the devices sense? What can the cloud analytics do with that information? Be as creative as possible.

Các thiết bị như máy giặt, tủ lạnh, và camera an ninh có thể thu thập dữ liệu về nhiệt độ, độ ẩm, và chuyển động. Dữ liệu này gửi lên đám mây để phân tích, tối

ưu hóa hiệu suất, tiết kiệm năng lượng, và cung cấp cảnh báo khi cần thiết. Cloud analytics giúp dự đoán bảo trì và nâng cao trải nghiệm người dùng.

3. Automobiles are becoming increasingly connected as a part of the IoT. What types of IoT services do you think can be used to make the driving and car ownership experience better, safer, and more enjoyable for the owner?

Ô tô thông minh kết nối với đám mây cung cấp dịch vụ như điều hướng thông minh, cảnh báo an toàn, chẩn đoán từ xa, và giải trí. Dữ liệu từ cảm biến được phân tích trên đám mây để đưa ra cảnh báo kịp thời và khuyến nghị bảo trì. Các dịch vụ này giúp cải thiện an toàn, tiết kiệm nhiên liệu, và nâng cao trải nghiệm lái xe.

Module 13: CloudFormation Templates

1. Technology terminology

- Cơ sở hạ tầng dưới dạng mã (IaC): Quá trình cung cấp và quản lý tài nguyên đám mây bằng cách viết một tệp mẫu có thể đọc được bởi con người và máy móc. Đối với phát triển đám mây Amazon Web Services (AWS), lựa chọn tích hợp sẵn cho IaC là AWS CloudFormation.
- Stack: Một tập hợp các tài nguyên Amazon Web Services (AWS) mà bạn có thể quản lý như một đơn vị duy nhất. Nói cách khác, bạn có thể tạo, cập nhật hoặc xóa một tập hợp tài nguyên bằng cách tạo, cập nhật hoặc xóa các ngăn xếp.

2. Background and misconceptions

- CloudFormation là một dịch vụ cho phép người dùng dễ dàng xây dựng và quản lý một ngăn xếp các tài nguyên AWS, chẳng hạn như Amazon Elastic Compute Cloud (Amazon EC2) và Amazon Simple Storage Service (Amazon S3). Đây là một dạng của IaC. Với CloudFormation, người dùng có thể thực hiện quy trình này bằng cách sử dụng các mã và tập lệnh AWS được xây dựng sẵn, giúp quy trình trở nên thân thiện hơn với người dùng. Sinh viên nên hình dung CloudFormation như là cách mà mạng điện toán đám mây có cấu trúc ảo, giống như các hệ thống máy tính thế hệ trước có cấu trúc phần cứng, dây cáp, và công cụ vật lý.
- CloudFormation đơn giản hóa việc cung cấp và quản lý trên AWS. Bạn có thể tạo các mẫu cho kiến trúc dịch vụ hoặc ứng dụng mong muốn, và CloudFormation sẽ sử dụng các mẫu này để cung cấp nhanh chóng và đáng tin cậy các dịch vụ hoặc

ứng dụng. Những tập hợp dịch vụ này được gọi là ngăn xếp. Người dùng cũng có thể dễ dàng cập nhật hoặc sao chép các ngăn xếp khi cần.

- CloudFormation cung cấp một ngôn ngữ chung để bạn mô hình hóa và cung cấp các tài nguyên ứng dụng AWS và bên thứ ba trong môi trường đám mây của mình. Với CloudFormation, bạn có thể sử dụng ngôn ngữ lập trình hoặc một tệp văn bản đơn giản để mô hình hóa và cung cấp tài nguyên theo cách tự động và bảo mật, cho các ứng dụng trên tất cả các khu vực và tài khoản. Điều này cung cấp một nơi duy nhất để lưu giữ thông tin về trạng thái của ứng dụng (nguồn sự thật) cho các tài nguyên AWS và bên thứ ba của bạn.
- Các mẫu CloudFormation là những khối mã cài đặt sẵn lớn. Mã thường được viết bằng ngôn ngữ JSON (JavaScript Object Notation) hoặc YAML (viết tắt của cụm từ vui nhộn “YAML Ain't Markup Language”). Bảng điều khiển AWS cho phép người dùng làm việc với mã hoặc sử dụng dạng đơn giản của các mẫu là các biểu tượng trực quan. Cả hai dịch vụ này đều dễ dàng hơn rất nhiều so với việc có một đội ngũ DevOps thực sự viết mã.
- Bạn có thể nghĩ rằng một mạng đám mây tồn tại thực sự trên bầu trời, chẳng hạn như trong vệ tinh hoặc một khái niệm trừu tượng nào đó về một đám mây. Bạn nên hiểu rằng mạng điện toán đám mây là kết nối với các ứng dụng trên các máy chủ của AWS chứ không phải trên các thiết bị cục bộ.

3. Question

1. Cloud-based computing systems are set up with a network architecture just like an actual physical network would be. A physical network can contain devices with large hard drives full of software, many cables connecting the system, and a large server that controls everything right on your premise. How do you think cloud-based networks might function?

Cloud-based networks hoạt động tương tự như mạng vật lý nhưng thay vì thiết bị phần cứng, nó sử dụng tài nguyên ảo hóa trên đám mây. Các máy chủ, lưu trữ, và mạng được cung cấp bởi nhà cung cấp dịch vụ (như AWS). Người dùng có thể quản lý và mở rộng hệ thống dễ dàng qua giao diện web mà không cần cài đặt hoặc bảo trì phần cứng tại chỗ.

2. AWS provides templates for creating networks. What is a template? Have you ever used a template? How was it helpful? What do you think a template for a network would include?

Template là một bản mẫu được thiết kế sẵn để giúp người dùng tạo ra cấu hình mạng hoặc dịch vụ dễ dàng hơn. Tôi đã sử dụng template khi tạo CloudFormation để triển khai các dịch vụ trên AWS. Template giúp tiết kiệm thời gian, giảm thiểu lỗi và đảm bảo cấu hình nhất quán. Một template cho mạng thường bao gồm cấu hình VPC, subnets, security groups, và các quy tắc định tuyến.

3. Is it better to have your infrastructure of services run through AWS or develop your own network with employees doing the coding? What are some pros and cons of each approach?

AWS:

Ưu điểm: Dễ dàng mở rộng, quản lý, có hỗ trợ dịch vụ và bảo mật tốt, triển khai nhanh.

Nhược điểm: Chi phí có thể cao, phụ thuộc vào nhà cung cấp, ít linh hoạt trong tùy chỉnh đặc biệt.

Mạng tự phát triển:

Ưu điểm: Linh hoạt, tự kiểm soát hoàn toàn, tùy chỉnh theo nhu cầu riêng.

Nhược điểm: Đòi hỏi nhiều nhân lực, chi phí bảo trì cao, thời gian triển khai lâu hơn

Module 14: Big Data

1. Technology terminology

- Dữ liệu: Là bất kỳ mẫu thông tin nào được lưu trữ và/hoặc xử lý bởi máy tính hoặc thiết bị di động.
- Dữ liệu lớn (Big Data): Dữ liệu lớn có thể được mô tả qua các thách thức trong việc quản lý dữ liệu—do khối lượng, tốc độ và sự đa dạng dữ liệu tăng trưởng mạnh mẽ—không thể giải quyết được bằng các cơ sở dữ liệu truyền thống. Mặc dù có rất nhiều định nghĩa về dữ liệu lớn, hầu hết trong số chúng đều bao gồm khái niệm ba V của dữ liệu lớn: khối lượng, sự đa dạng và tốc độ.
- Internet vạn vật (IoT): Mạng lưới các thiết bị thông dụng hàng ngày được tích hợp cảm biến và công nghệ giao tiếp, cho phép chúng gửi và nhận dữ liệu qua

Internet. Về cơ bản, IoT bao gồm nhiều thiết bị khác nhau được kết nối với internet, chẳng hạn như điện thoại thông minh, đồng hồ, thiết bị gia dụng và thậm chí cả ô tô.

- Xử lý dữ liệu: Là bước trong đó dữ liệu được chuyển đổi từ trạng thái thô ban đầu thành định dạng có thể sử dụng được—thường thông qua việc làm sạch, sắp xếp, tổng hợp, kết hợp và thậm chí thực hiện các chức năng và thuật toán nâng cao. Các tập dữ liệu kết quả sau đó được lưu trữ để xử lý thêm hoặc được cung cấp cho việc tiêu thụ qua các công cụ trí tuệ doanh nghiệp và trực quan hóa dữ liệu.
- Trí tuệ nhân tạo (AI): Là hệ thống máy tính có khả năng thực hiện các nhiệm vụ thường yêu cầu trí tuệ con người, như nhận thức thị giác, nhận diện giọng nói, ra quyết định, và dịch thuật giữa các ngôn ngữ.
- Apache Hadoop: Là công nghệ quản lý khối lượng lớn dữ liệu trong môi trường phân tán. Hadoop được phát triển bởi Apache và hiện có sẵn để chạy trên các cụm máy chủ ảo trong AWS Cloud.

2. Background and misconceptions

- Bạn đã biết dữ liệu là gì, nhưng còn dữ liệu lớn thì sao? Bạn có thể đã nghe về nó, nhưng liệu bạn có hiểu nó khác biệt như thế nào so với dữ liệu truyền thống? Dữ liệu lớn không chỉ đơn giản là một nhóm dữ liệu lớn. Dữ liệu lớn có những đặc điểm đặc biệt cần sử dụng các công cụ đặc biệt để xử lý và phân tích. Dữ liệu lớn được phân biệt bởi kích thước khổng lồ (khối lượng), phạm vi đa dạng của các loại dữ liệu (sự đa dạng), và tốc độ cực nhanh mà nó được thu thập và xử lý (tốc độ). Ba đặc điểm này là cơ sở để xác định dữ liệu là dữ liệu lớn. Ngoài ra, độ tin cậy (veracity) và giá trị (value) cũng thường được dùng để mô tả dữ liệu lớn.
 - **Volume:** Đề cập đến lượng dữ liệu.
 - **Variety:** Đề cập đến các loại dữ liệu trong bộ dữ liệu.
 - **Velocity:** Đề cập đến tốc độ thu thập và xử lý dữ liệu.
 - **Veracity:** Đề cập đến độ chính xác và độ tin cậy của dữ liệu.
 - **Value:** Đề cập đến tính hữu ích của dữ liệu.
- Khi những đặc điểm này xuất hiện, dữ liệu được gọi là dữ liệu lớn và được lưu trữ và xử lý khác biệt so với dữ liệu truyền thống. Dữ liệu lớn đang được sử dụng mỗi ngày để mang lại những lợi ích lớn trong hầu hết mọi ngành nghề.
Ví dụ về dữ liệu lớn:

- Một công ty khoa học đời sống thu thập và phân tích thông tin về DNA, protein và nhiều yếu tố khác từ các tế bào đơn của con người. Cơ thể con người có hàng chục nghìn tỷ tế bào, vì vậy có một lượng dữ liệu khổng lồ. Các nhà nghiên cứu sử dụng thông tin này để hiểu rõ hơn về sinh học ung thư, từ đó có thể phát triển các phương pháp điều trị hiệu quả.
- Một công ty dịch vụ tài chính giúp các doanh nghiệp nhỏ và vừa truy cập thông tin báo cáo tín dụng. Các doanh nghiệp sử dụng thông tin này để xác định rủi ro tài chính của khách hàng và đối tác, giúp họ đưa ra quyết định tốt hơn về đối tác kinh doanh và tránh các tình huống rủi ro tài chính lớn.
- Một tổ chức đua xe quốc tế sử dụng dữ liệu lớn để thiết kế và thử nghiệm ô tô. Với dữ liệu này, tổ chức có thể chạy các mô phỏng phức tạp một cách nhanh chóng và do đó tạo ra các thiết kế mới nhanh chóng. Các thiết kế mới giúp giảm sự nhiễu loạn cho những chiếc xe đi phía sau, giúp các xe có thể chạy gần nhau hơn và tạo ra nhiều cơ hội cho các tay đua vượt qua nhau trong các cuộc đua, khiến các cuộc đua trở nên hấp dẫn hơn với người hâm mộ.
- Một công ty game xã hội thu thập một lượng dữ liệu khổng lồ về các trò chơi mỗi ngày. Bằng cách phân tích dữ liệu này, công ty có thể hiểu rõ hơn và tối ưu hóa trải nghiệm người chơi.
- Một công ty năng lượng có hàng chục nghìn tua-bin gió trên toàn thế giới. Các tua-bin này thu thập dữ liệu, thực hiện chẩn đoán, theo dõi sản xuất và tối ưu hóa cách thức vận hành. Điều này giúp tăng hiệu quả và sản lượng năng lượng.
- Một nhà cung cấp khí đốt thiên nhiên sử dụng dữ liệu lớn để chạy các mô phỏng và thực hiện nghiên cứu để xác định nơi khí đốt thiên nhiên nằm dưới mặt đất.
- Một công ty vận tải đường sắt thu thập dữ liệu từ các cảm biến IoT trên các đoàn tàu. Điều này giúp khách hàng biết khi nào họ sẽ nhận được hàng hóa vận chuyển, để có thể lên kế hoạch sắp xếp nhân viên tiếp nhận hàng hóa.
- Một chuỗi nhà hàng quốc tế sử dụng dữ liệu khổng lồ để quản lý sản xuất thực phẩm và vận chuyển để đảm bảo các nhà hàng có nguyên liệu cần thiết khi cần. Khách hàng không phải lo lắng về việc các món ăn yêu thích hết hàng. Công ty cũng sử dụng phân tích dữ liệu để xác định các vị trí tốt nhất để mở nhà hàng mới.

- Thu thập dữ liệu thô—giao dịch, nhật ký, thiết bị di động, hình ảnh, và nhiều hơn nữa—là thử thách đầu tiên mà nhiều tổ chức phải đối mặt khi xử lý dữ liệu lớn. Một nền tảng dữ liệu lớn tốt giúp bước này dễ dàng hơn, cho phép các nhà phát triển thu thập một loạt dữ liệu từ cấu trúc đến phi cấu trúc với bất kỳ tốc độ nào, từ theo lô đến theo thời gian thực.
- Tất nhiên, giống như bất kỳ công nghệ nào, dữ liệu lớn cũng có rủi ro. Những rủi ro này có thể được phóng đại do cách mà các bộ dữ liệu được xử lý và bởi tầm với rộng lớn của các công nghệ này. Một số rủi ro chính liên quan đến dữ liệu lớn bao gồm quyền riêng tư và sự thiên vị. Dữ liệu luôn có rủi ro về quyền riêng tư khi được thu thập và lưu trữ. Tuy nhiên, với dữ liệu truyền thống, các chi tiết nhạy cảm thường đã được biết đến và có thể bảo vệ, hoặc ít nhất là theo dõi. Nhưng với dữ liệu lớn, điều này không phải lúc nào cũng đúng. Dữ liệu lớn thường được tổng hợp hoặc kết hợp từ nhiều nguồn khác nhau, và các bước được thực hiện để loại bỏ các thông tin nhận dạng và làm cho dữ liệu trở nên vô danh. Điều này tưởng chừng như hợp lý, nhưng thực tế không đơn giản như vậy. Mặc dù thông tin về bạn được lưu trữ trong bộ dữ liệu lớn có thể không có tên hoặc địa chỉ của bạn, nhưng nó có thể có những manh mối khác về danh tính của bạn và có thể được truy vết lại với một chút công sức và sử dụng trí tuệ nhân tạo.
- Một rủi ro chính khác của dữ liệu lớn là chúng ta có thể vô tình dựa vào các kết quả có sự thiên vị. Ví dụ, nếu kết quả thiên vị khiến bạn không thể vay tiền để đi học đại học, hay không thể lấy bằng lái xe hoặc lên máy bay? Dữ liệu từ mạng xã hội đang được sử dụng trong một số ngành để phân loại người dùng dựa trên các kết nối, lượt thích, bài viết, và dữ liệu khác. Nếu mô hình phân tích sai lầm giả định rằng một hồ sơ cá nhân cụ thể sẽ không phải là rủi ro tín dụng tốt hoặc có thể có hành vi bạo lực, thì kết quả phân tích sau này sẽ không chính xác. Tuy nhiên, chúng ta có thể không nhận ra sai lầm đó hoặc không thể chứng minh được, vì các thuật toán thường được ẩn và chúng ta không thể thấy logic được sử dụng.
- Dù có rủi ro, dữ liệu lớn vẫn ở khắp nơi và là một chủ đề nóng. Bạn có thể tìm hiểu thêm về các sản phẩm và dịch vụ của AWS cho dữ liệu lớn tại các liên kết sau:
- AWS Certified Big Data – Specialty
(<https://aws.amazon.com/certification/certified-big-data-specialty/>)

- AWS Customer Success Stories (<https://aws.amazon.com/solutions/case-studies/>)
- Với tất cả những điều này trong đầu, hãy chú ý đến cách công nghệ máy tính hiện đại có thể được áp dụng để đáp ứng các yêu cầu của ba yếu tố V.
 - Cách tiếp cận 1: Một cách tiếp cận là sử dụng một siêu máy tính với lượng bộ nhớ khổng lồ được kết nối với một mạng lưới cực kỳ nhanh.
 - Cách tiếp cận 2: Cách tiếp cận thứ hai là sử dụng nhiều máy tính nhỏ, mỗi máy có một lượng bộ nhớ vừa phải, được kết nối với nhau qua mạng lưới.
 - Cách tiếp cận 3: Cách tiếp cận thứ ba là sử dụng cách tiếp cận thứ hai chạy trên điện toán đám mây thông qua các dịch vụ web, chẳng hạn như AWS, được cung cấp bởi đám mây.
- Nếu chọn Cách tiếp cận 1, thì phần mềm cơ sở dữ liệu quan hệ hoặc phi quan hệ tiêu chuẩn có thể sẽ hoạt động. Tuy nhiên, cần nhớ rằng Cách tiếp cận 1 rất tốn kém về cả chi phí tiền bạc lẫn chi phí duy trì phần cứng và phần mềm.
- Cách tiếp cận 2 sẽ yêu cầu phần mềm có thể xử lý việc phân phối các tác vụ trên một tập hợp các máy vật lý nhỏ hơn với bộ nhớ đi kèm của chúng. Phần mềm này sẽ cần phải chia nhỏ dữ liệu trên các máy và sau đó thực hiện xử lý song song để quản lý dữ liệu một cách có phối hợp nhằm tạo ra kết quả đáng tin cậy. Hadoop là phần mềm có thể thực hiện điều đó.
- Cách tiếp cận 3 sẽ yêu cầu phần mềm giống như vậy, nhưng gánh nặng cung cấp máy và duy trì chúng với các bản nâng cấp hệ điều hành và phần mềm cơ sở dữ liệu giờ đây sẽ thuộc trách nhiệm của nhà cung cấp dịch vụ web, chẳng hạn như AWS. Hơn nữa, cách tiếp cận này có khả năng mở rộng và thu nhỏ, nghĩa là các máy chủ ảo sẽ được thêm vào hoặc loại bỏ khi kích thước dữ liệu yêu cầu, giúp tiết kiệm chi phí khi không cần máy chủ dự phòng.

3. Question

1. What are some examples of data being tracked about your activities? How do you feel about your information being collected? Are you surprised to learn that anonymous data can be de-anonymized to reveal who a person is, especially with the use of AI? How do you feel about that? Do you think you will change your actions based on what you learned in this discussion? How or why?

Dữ liệu về hoạt động của tôi như lịch sử tìm kiếm, vị trí, thời gian sử dụng ứng dụng và các giao dịch trực tuyến đều được theo dõi. Tôi cảm thấy lo ngại về vấn

đề bảo mật và quyền riêng tư khi thông tin này bị thu thập. Việc dữ liệu ẩn danh có thể bị giải mã nhờ AI khiến tôi bất ngờ và không thoải mái. Điều này có thể khiến tôi thay đổi cách sử dụng thiết bị, hạn chế chia sẻ thông tin cá nhân hơn.

2. How much responsibility should governments take when regulating big data? What do you think the biggest challenges will be in writing and enforcing laws concerning big data? How involved do you think regular citizens should be in this process? Why?

Chính phủ nên có trách nhiệm trong việc xây dựng luật bảo vệ dữ liệu cá nhân, nhưng sẽ gặp thách thức khi cân bằng giữa bảo mật và lợi ích kinh tế. Việc thực thi sẽ khó khăn vì công nghệ thay đổi nhanh chóng. Người dân cần tham gia đóng góp ý kiến để đảm bảo luật phản ánh đúng nhu cầu và quyền lợi cá nhân.

3. Briefly, list the research skills needed to work with big data, according to relevant articles and job listings. You can include keywords such as analytics, visualization, and data scientist in your search. Would you be interested in any of these jobs as a profession? Why or why not? If not as a profession, would you enjoy doing some of this work on your own, as a citizen analyst? What industry or type of datasets would you like to work with?

Kỹ năng bao gồm: phân tích dữ liệu, trực quan hóa, lập trình (Python, SQL), học máy, và kiến thức về các công cụ như Tableau, Hadoop. Tôi không muốn theo đuổi nghề này, nhưng có thể hứng thú làm một "citizen analyst" trong các lĩnh vực như chăm sóc sức khỏe hoặc dữ liệu môi trường vì có thể tạo ra tác động tích cực.

4. Data collection has its benefits and can positively impact our world in many ways. For example, medical data can be collected, analyzed, and used to predict future medicinal needs and inform future diagnoses. How else might data collection be of personal benefit to you or your community?

Dữ liệu có thể được dùng để tối ưu hóa giao thông, nâng cao chất lượng giáo dục, hoặc dự đoán xu hướng tiêu dùng để cải thiện dịch vụ công. Cá nhân có thể hưởng lợi từ việc nhận được các khuyến nghị sản phẩm phù hợp, chăm sóc sức khỏe tốt hơn, và thông tin về nguy cơ sức khỏe hoặc ô nhiễm không khí trong khu vực sinh sống.

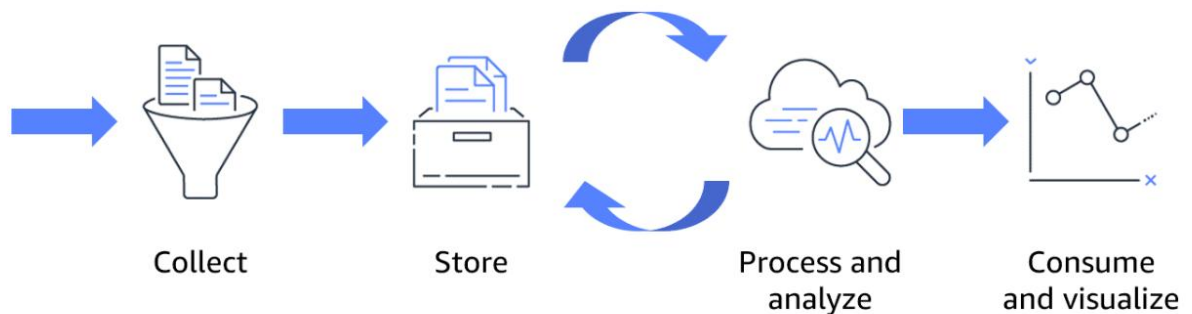
Module 15: Big Data Processing Cycle

1. Technology terminology

- Dữ liệu lớn (Big data): Dữ liệu lớn có thể được mô tả qua những thách thức trong việc quản lý dữ liệu mà—do khối lượng, tốc độ và sự đa dạng của dữ liệu tăng lên đáng kể—không thể giải quyết được bằng cơ sở dữ liệu truyền thống. Mặc dù có rất nhiều định nghĩa về dữ liệu lớn, nhưng hầu hết trong số đó đều bao gồm khái niệm ba yếu tố V của dữ liệu lớn: khối lượng, sự đa dạng và tốc độ. Khái niệm về độ tin cậy và giá trị cũng được đưa vào trong một số thảo luận.
 - Khối lượng (Volume): Đề cập đến lượng dữ liệu.
 - Sự đa dạng (Variety): Đề cập đến phạm vi các loại dữ liệu trong bộ dữ liệu.
 - Tốc độ (Velocity): Đề cập đến tốc độ thu thập và xử lý dữ liệu.
 - Độ tin cậy (Veracity): Đề cập đến độ tin cậy và chính xác của dữ liệu.
 - Giá trị (Value): Đề cập đến tính hữu ích của dữ liệu.
- Phân tích dữ liệu (Data analytics): Công nghệ và quy trình chuyển đổi dữ liệu thô thành các thông tin chi tiết để đưa ra quyết định. Phân tích dữ liệu liên quan đến việc chọn lựa và áp dụng các mô hình thống kê vào bộ dữ liệu để chỉ ra các mẫu hoặc xu hướng, và đặc biệt quan trọng khi cố gắng rút ra kết luận từ dữ liệu lớn.
- Chu trình xử lý dữ liệu lớn (Big data processing cycle): Mô hình về dòng chảy chung của dữ liệu lớn—from việc thu thập dữ liệu thô đến việc tiêu thụ thông tin có thể hành động. Chu trình này thường bao gồm bốn giai đoạn:
 - Thu thập (Collect): Thu thập dữ liệu thô—như giao dịch, nhật ký và thiết bị di động—là thách thức đầu tiên mà nhiều tổ chức phải đối mặt khi làm việc với dữ liệu lớn. Một giải pháp dữ liệu lớn tốt sẽ làm cho bước này dễ dàng hơn, cho phép các nhà phát triển thu thập nhiều loại dữ liệu khác nhau—from có cấu trúc đến không có cấu trúc—with tốc độ từ thời gian thực đến theo lô.
 - Lưu trữ (Store): Bất kỳ giải pháp dữ liệu lớn nào cũng cần có một kho dữ liệu an toàn, có thể mở rộng và bền vững để lưu trữ dữ liệu trước hoặc sau khi xử lý. Tùy thuộc vào yêu cầu cụ thể của bạn, bạn cũng có thể cần các kho tạm thời cho dữ liệu trong quá trình truyền tải.
 - Xử lý và phân tích (Process and analyze): Bước này là nơi dữ liệu được chuyển đổi từ trạng thái thô thành định dạng có thể tiêu thụ—thường là thông qua việc sắp xếp, tổng hợp, kết hợp và thực hiện các chức năng và thuật toán nâng cao hơn. Các bộ dữ liệu kết quả sau đó sẽ được lưu trữ để xử lý tiếp hoặc

được cung cấp để tiêu thụ với các công cụ thông minh doanh nghiệp và trực quan hóa dữ liệu.

- Tiêu thụ và trực quan hóa (Consume and visualize): Dữ liệu lớn là về việc lấy ra các thông tin có giá trị và có thể hành động từ tài sản dữ liệu của bạn. Lý tưởng nhất, dữ liệu được cung cấp cho các bên liên quan thông qua các công cụ thông minh doanh nghiệp tự phục vụ và công cụ trực quan hóa dữ liệu linh hoạt. Những công cụ này cho phép khám phá nhanh chóng và dễ dàng các bộ dữ liệu. Tùy vào loại phân tích, người dùng cũng có thể tiêu thụ dữ liệu kết quả dưới dạng dự đoán thống kê, trong trường hợp phân tích dự đoán, hoặc các hành động được khuyến nghị, trong trường hợp phân tích chỉ dẫn.



Chu trình xử lý dữ liệu lớn bao gồm bốn giai đoạn chính: thu thập, lưu trữ, xử lý và phân tích, và tiêu thụ và trực quan hóa. Dữ liệu có thể di chuyển qua lại giữa giai đoạn lưu trữ và xử lý và phân tích trong suốt chu trình.

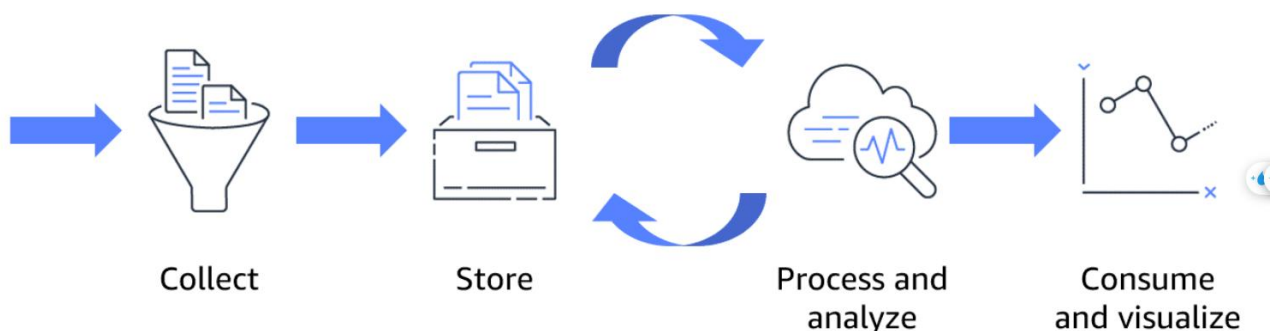
- Bảng điều khiển (Dashboard): Một bản snapshot chỉ đọc của một phân tích mà bạn có thể chia sẻ với các người dùng khác cho mục đích báo cáo. Dữ liệu được sử dụng cho phân tích không được ghi lại như một phần của bảng điều khiển. Khi bạn xem bảng điều khiển, nó phản ánh dữ liệu hiện tại trong các bộ dữ liệu được sử dụng cho phân tích.
- Trí tuệ nhân tạo (AI): Lĩnh vực khoa học máy tính chuyên giải quyết các vấn đề nhận thức thường liên quan đến trí thông minh của con người, như học hỏi, giải quyết vấn đề và nhận diện mẫu. AI trở nên thông minh hơn và học nhanh hơn với nhiều dữ liệu hơn. Mỗi ngày, các doanh nghiệp đang tạo ra nhiều dữ liệu này để vận hành các giải pháp học máy (ML) và học sâu. Với sự ra đời của Internet vạn vật (IoT), công nghệ cảm biến làm gia tăng gấp bội lượng dữ liệu cần phân tích—dữ

liệu từ các nguồn, địa điểm, vật thể và sự kiện trước đây gần như chưa được khai thác.

- Học máy (ML): Tên gọi thường được áp dụng cho một số kỹ thuật Bayes (thống kê) dùng để nhận diện mẫu và học hỏi. ML thường được sử dụng để dự đoán các kết quả trong tương lai dựa trên dữ liệu lịch sử. Cốt lõi của ML là một tập hợp các thuật toán có thể học từ và đưa ra dự đoán dựa trên dữ liệu đã ghi lại. ML sử dụng dữ liệu để tạo mã thống kê (mô hình ML) sẽ đưa ra kết quả đúng dựa trên một mẫu được nhận diện từ các ví dụ đầu vào trước đó. Độ chính xác của mô hình ML chủ yếu phụ thuộc vào chất lượng và số lượng của dữ liệu lịch sử.

2. Background and misconceptions

- Như một cách tổng kết nhanh, đây là một lời nhắc về ý nghĩa của dữ liệu lớn. Phần quan trọng là "lớn", vì lượng dữ liệu trên thế giới ngày nay và tốc độ phát triển của nó là điều đáng kinh ngạc. Hiện tại, chúng ta đang có 1 zettabyte dữ liệu và đang phát triển với tỷ lệ tăng trưởng hàng năm là 42%.
- Zettabyte là gì? Về mặt số mũ, đó là 10^{21} . Nói cách khác, đó là một triệu triệu gigabyte.
- Tuy nhiên, tỷ lệ tăng trưởng thực sự rất ấn tượng. Với 42% hàng năm, có nghĩa là đến năm 2025, sẽ có 175 zettabyte dung lượng lưu trữ trên toàn cầu.
- Vậy tất cả điều này có ý nghĩa gì? Một điều là, giải quyết vấn đề dữ liệu lớn bằng cách sử dụng các thiết bị lưu trữ ngày càng lớn sẽ sớm không khả thi. Do đó, việc sử dụng phương pháp phân tán, trong đó dữ liệu được phân phối qua nhiều thiết bị, sẽ là điều cần thiết, và đây chính là bản chất của dữ liệu lớn. Không chỉ là việc phân tán dữ liệu, mà còn là việc khai thác thông tin từ dữ liệu, điều này sẽ đòi hỏi phần mềm và giải pháp mạng đặc biệt.
- Các dự án dữ liệu lớn cũng có vòng đời, giống như bất kỳ dự án nào khác. Bạn có thể thấy hình ảnh minh họa về chu trình xử lý dữ liệu lớn ở đây.



- Chu trình xử lý dữ liệu lớn bao gồm bốn giai đoạn chính: thu thập, lưu trữ, xử lý và phân tích, và tiêu thụ và trực quan hóa. Dữ liệu có thể di chuyển qua lại giữa các giai đoạn lưu trữ và xử lý, phân tích trong suốt chu trình.
- Các mô hình như thế này có thể giúp chúng ta hiểu các bước trong một quy trình, các loại đầu vào và đầu ra kỳ vọng cho mỗi giai đoạn, và cách giao tiếp về dự án với các thành viên trong nhóm. Chúng cung cấp các hướng dẫn tổng quan về cách dự án nên tiến hành. Mô hình này được cung cấp bởi AWS.
- Các giai đoạn của chu trình có vẻ đơn giản: thu thập, lưu trữ, xử lý và tiêu thụ. Tuy nhiên, các hoạt động trong mỗi giai đoạn giúp đảm bảo chất lượng và tính hữu ích của dữ liệu có thể rất phức tạp.
- Một loạt các dữ liệu thô có thể được thu thập. Các giải pháp bạn chọn cho giai đoạn thu thập phải có khả năng xử lý các loại dữ liệu bạn đang sử dụng, dù đó là dữ liệu có cấu trúc, bán cấu trúc, không cấu trúc hay kết hợp của ba loại này. Dữ liệu có thể được nhận và xử lý theo thời gian thực hoặc qua dòng dữ liệu, điều này phải được hỗ trợ bởi các công nghệ được chọn cho dự án. Sau khi dữ liệu được thu thập, nó phải được lưu trữ trong một cấu trúc hỗ trợ các loại dữ liệu và các xử lý cần thiết. Các tổ chức có nhiều lựa chọn cho lưu trữ dữ liệu lớn, bao gồm cơ sở dữ liệu phi quan hệ, kho dữ liệu và hồ dữ liệu.
 - Lưu trữ dữ liệu lớn: Kiến trúc thu thập và quản lý các bộ dữ liệu lớn và tạo điều kiện cho phân tích dữ liệu theo thời gian thực.
 - Cơ sở dữ liệu phi quan hệ: Một cơ sở dữ liệu không sử dụng các hàng và cột như trong các hệ thống cơ sở dữ liệu truyền thống. Thay vào đó, chúng sử dụng một mô hình lưu trữ tối ưu hóa cho loại dữ liệu được lưu trữ.
 - Kho dữ liệu: Các kho lưu trữ dữ liệu tập trung từ một hoặc nhiều nguồn.
 - Hồ dữ liệu: Kho lưu trữ dữ liệu ở dạng thô.
- Chất lượng của dữ liệu lớn có thể thấp, tùy thuộc vào các loại và nguồn gốc, vì vậy nó thường phải được làm sạch và chuyển đổi để cải thiện độ chính xác của dữ liệu và phân tích kết quả. Khi dữ liệu đã được chuẩn bị và lưu trữ trong kho dữ liệu có thể truy cập, nó sẽ sẵn sàng cho việc phân tích.
- Các tổ chức sử dụng các chỉ số hiệu suất chính, hay các thước đo thành công cụ thể, để điều hướng dữ liệu họ thu thập và những câu hỏi họ đặt ra với dữ liệu sau khi nó đã sẵn sàng để phân tích. Một số doanh nghiệp sẽ theo dõi các chỉ số liên

quan đến an toàn và khuyết tật trong sản xuất, trong khi những doanh nghiệp khác sẽ theo dõi các điểm dữ liệu liên quan đến hoạt động của trang web và hiệu quả của quảng cáo cụ thể. Phân tích có thể giúp các tổ chức trả lời các câu hỏi sau về doanh nghiệp của họ:

- Điều gì đã xảy ra?
 - Tại sao điều đó lại xảy ra?
 - Điều gì sẽ xảy ra?
 - Tôi nên làm gì?
- Kết quả của phân tích cần được trình bày theo cách truyền tải thông tin và kết quả một cách nhanh chóng và rõ ràng. Bảng điều khiển và trực quan hóa thường được sử dụng cho mục đích này.
 - Các giải pháp dữ liệu lớn có thể liên quan đến việc sử dụng AI hoặc ML. Cả hai công nghệ đều dựa vào lượng dữ liệu chất lượng cao lớn để học và tạo ra dự đoán.

3. Question

1. In this module, we will define the big data processing cycle, which is a standard model. Another example of a standard model is the conventional process by which all cars are manufactured, for example, the car being assembled in the same order repeatedly. Why do you think it is important to define a standard model for any process? Should it always be followed? Why or why not? Identify a circumstance where you might deviate from the specific process.

Việc định nghĩa mô hình chuẩn giúp đảm bảo tính nhất quán, hiệu quả và dễ quản lý trong quy trình. Tuy nhiên, không phải lúc nào cũng cần tuân theo. Trong trường hợp xử lý dữ liệu bất thường hoặc khi yêu cầu của dự án thay đổi, ta có thể phải linh hoạt điều chỉnh quy trình để đáp ứng nhu cầu đặc biệt hoặc tiết kiệm thời gian.

2. Big data refers to the massive amount of data available to different people. One example of big data is the number of likes or upvotes that social media posts receive, or the number of views a video has on a streaming service. Ensuring big data quality can be a very costly process, especially with large amounts of time-sensitive data. Organizations must consider the trade-off between cost and quality on each project. Sometimes the extra cost to improve data quality is just not worth it, and sometimes the data needs to be of the highest quality, no matter the cost. When would a business

want a high level of data quality? What types of big data use cases might have justifiably lower standards for the quality of the data? Why? Remember, when considering data quality, we want to think about accuracy, consistency, and timeliness.

Doanh nghiệp cần chất lượng dữ liệu cao khi đưa ra quyết định chiến lược, ví dụ như dự đoán xu hướng khách hàng hoặc phân tích rủi ro tài chính. Tuy nhiên, với các trường hợp như phân tích xu hướng mạng xã hội hoặc nghiên cứu ý kiến khách hàng, dữ liệu có thể có chất lượng thấp hơn vì tính chính xác tuyệt đối không quan trọng bằng việc nhận diện các xu hướng chung.

3. When might you analyze big data in your personal or academic interests? How could this data analysis inform you?

Tôi có thể phân tích dữ liệu lớn trong các dự án nghiên cứu học thuật, như phân tích xu hướng thị trường hoặc đánh giá phản hồi của sinh viên. Việc phân tích này giúp tôi hiểu sâu hơn về vấn đề, phát hiện xu hướng hoặc đưa ra những quyết định chính xác hơn trong quá trình học tập hoặc phát triển dự án cá nhân.

Module 16: Blockchain and Cryptocurrency

1. Technology terminology

- Blockchain: Công nghệ cơ sở dữ liệu phân tán duy trì một tập hợp các giao dịch và hợp đồng thông minh ngày càng phát triển, được bảo vệ khỏi việc giả mạo và sửa đổi bằng cách sử dụng mã hóa.
- Cryptocurrency: Các token kỹ thuật số được triển khai bằng công nghệ blockchain, chia sẻ một số đặc tính của tiền tệ vật chất và có thể được mua, trao đổi và chi tiêu độc lập với cơ quan ngân hàng trung ương.
- Cryptocurrency mining (hoặc cryptomining): Việc thêm giao dịch vào sổ cái blockchain hiện có của các giao dịch phân tán giữa tất cả người dùng của blockchain.
- Decentralized database: Một cơ sở dữ liệu không có vị trí duy nhất; thay vào đó, các mảnh thông tin được lưu trữ ở các vị trí khác nhau và tất cả đều kết nối với nhau.
- Hardening (dữ liệu hoặc hệ thống): Hardening có nghĩa là giảm thiểu các lỗ hổng trong công nghệ. Mục tiêu là giảm thiểu rủi ro bảo mật.

- Hash: Một hàm toán học được sử dụng trong mã hóa, nhận vào một đầu vào dữ liệu, chẳng hạn như danh sách tất cả các giao dịch, và trả về một chuỗi mới có độ dài cố định.
- Immutable transactions: Tính không thể thay đổi có nghĩa là điều gì đó không thể bị thay đổi. Do đó, một giao dịch không thể thay đổi là giao dịch không thể thay đổi. Dữ liệu blockchain không thể bị sửa đổi.
- Smart contract: Một hợp đồng thông minh có các điều khoản của thỏa thuận được viết trong mã. Mã này tồn tại trên mạng blockchain. Mã điều khiển việc thực hiện hợp đồng. Các giao dịch có thể được theo dõi và không thể bị đảo ngược. Hợp đồng thông minh cho phép các giao dịch giữa các bên ẩn danh mà không cần cơ quan trung ương, hệ thống pháp lý hoặc cơ chế thực thi bên ngoài.

2. Background and misconceptions

- Blockchain là một công nghệ cơ sở dữ liệu phi tập trung, duy trì một tập hợp các giao dịch đang phát triển không ngừng và được bảo vệ khỏi sự sửa đổi và chỉnh sửa bằng cách sử dụng mã hóa. Các tài nguyên này cung cấp những điểm khởi đầu tốt để học về blockchain.
- Blockchain lưu trữ các giao dịch trong một chuỗi dữ liệu được tạo thành từ các khối. Mỗi khối có hai phần: tiêu đề khối và thân khối. Thân khối chứa tất cả các giao dịch (dữ liệu) được gán cho khối. Tiêu đề khối chứa thông tin như dấu thời gian, băm của khối trước, băm của tất cả các giao dịch trong thân khối, và kích thước băm cần thiết cho khối mới.
- Hãy tưởng tượng blockchain như thế này: Bạn là bác sĩ. Bạn có một bệnh nhân mắc bệnh mãn tính và mỗi năm họ đến làm một thủ tục. Mỗi thủ tục, bạn có một thư mục riêng (thân khối) chứa các ghi chú, thuốc, thông tin thanh toán và kết quả xét nghiệm. Mục đầu tiên trong mỗi thư mục là một danh sách (tiêu đề khối) như một mục lục, chứa số và vị trí của thư mục các thủ tục trước. Mỗi danh sách có một mã vạch (băm) khi quét sẽ xác nhận tính xác thực của thư mục đó.
- Băm mật mã (được gọi là băm) là một hàm toán học nhận vào một dữ liệu đầu vào, như danh sách tất cả các giao dịch, và trả về một đầu ra có độ dài cố định. Đầu ra có thể trông giống như một tập hợp ngẫu nhiên các chữ cái và số, nhưng kết quả của hàm băm có một số đặc điểm rất quan trọng:

- Hàm băm sẽ luôn tạo ra một chuỗi giống hệt khi nhận được đầu vào giống nhau. Ngay cả một sự thay đổi nhỏ trong đầu vào cũng sẽ tạo ra một kết quả rất khác biệt.
- Không thể giải mã một băm để có được thông điệp ban đầu. Các thông điệp được xác minh với băm bằng cách thực hiện thao tác băm đối với thông điệp và so sánh kết quả với băm xác minh. Đây là lý do tại sao băm được sử dụng để bảo mật dữ liệu.
- Trong một blockchain, tiêu đề của một khối luôn chứa băm của tiêu đề khối trước đó. Theo cách này, mỗi khối trong blockchain được xác nhận bởi khối tiếp theo trong chuỗi.
- Mỗi khối trong blockchain được bảo vệ khỏi việc sửa đổi hay chỉnh sửa vì thay đổi ngay cả một ký tự trong đầu vào sẽ tạo ra sự khác biệt rất lớn trong băm kết quả. Nếu một kẻ tấn công cố gắng thay đổi dữ liệu trong một khối, các băm liên kết trong tất cả các khối tiếp theo sẽ bị sai. Cấu trúc liên kết của blockchain là một trong những cách giúp blockchain được bảo vệ khỏi sự thay đổi và sửa đổi. Tuy nhiên, băm không phải là cách duy nhất để xác nhận các khối.
- Blockchain cũng sử dụng các hệ thống sổ cái phân tán để đảm bảo rằng blockchain được bảo vệ khỏi sự sửa đổi. Các thành viên trong mạng blockchain lưu giữ một bản ghi về các khối hợp lệ. Trong trường hợp một khối bị thay đổi và có câu hỏi về tính hợp lệ, nhiều bản sao của sổ cái có thể được so sánh. Trong một mạng blockchain sử dụng hệ thống sổ cái phân tán, một cuộc bỏ phiếu đa số sẽ được tổ chức và tất cả các thành viên bỏ phiếu dựa trên bản ghi của chính họ về blockchain. Phương pháp này làm cho việc thay đổi một khối trong blockchain trở nên khó khăn ngay cả khi không có một cơ quan trung ương duy nhất kiểm soát sổ cái. Để thực hiện thành công, kẻ tấn công sẽ phải thay đổi 51% các sổ cái của các thành viên.

Lợi thế của blockchain

- Blockchain có lợi thế so với các loại lưu trữ dữ liệu khác vì nó nhanh chóng và bảo mật và không yêu cầu một nguồn quyền lực duy nhất để xác thực các giao dịch mà nó chứa. Các phương pháp mã hóa và xác thực của nó làm cho việc thao túng hoặc làm mất giá trị dữ liệu trở nên khó khăn hơn rất nhiều cho các kẻ tấn

công. Điều này giúp blockchain thay thế cơ quan trung ương mà nhiều hệ thống giao dịch kinh doanh yêu cầu.

Các ngành công nghiệp nào sẽ thay đổi nhờ blockchain?

- Blockchain có thể thay đổi mạnh mẽ các ngành sau: ngân hàng, chăm sóc sức khỏe, chính trị, bất động sản, ngành luật, an ninh, chính phủ, cho thuê và chia sẻ xe, từ thiện và các tổ chức viện trợ, và giáo dục.

Cryptocurrency là gì?

- Cryptocurrency là các loại tiền tệ thay thế được triển khai bằng công nghệ blockchain. Bitcoin là loại cryptocurrency đầu tiên. Cryptocurrency thường xuất hiện trên các phương tiện truyền thông vì chúng có tác động lớn nhất đến công chúng. Nhiều người nhầm lẫn bitcoin (hoặc cryptocurrency nói chung) với blockchain. Quan trọng là phải nhớ rằng cryptocurrency chỉ là một ứng dụng có thể có của blockchain.

Bitcoin là gì?

- Bitcoin là một cryptocurrency được tạo ra vào năm 2008. Phần mềm mã nguồn mở của Bitcoin được phát hành vào năm 2009 và nhanh chóng phổ biến ý tưởng về blockchain.
- Một tính năng quan trọng của mạng Bitcoin là khai thác. Những người khai thác cryptocurrency là những người tạo ra các khối mới được thêm vào blockchain. Các thợ đào cryptocurrency cạnh tranh với những người khai thác khác để giải quyết một vấn đề toán học phức tạp cho mỗi khối mới, điều này được gọi là chứng minh công việc. Quá trình này được gọi là khai thác vì người sáng tạo Bitcoin so sánh những người khai thác cryptocurrency với những người khai thác vàng.
- Khai thác cryptocurrency liên quan đến việc tạo ra một băm nhỏ hơn hoặc bằng kích thước băm yêu cầu trong khối mới nhất trong blockchain. Điều này có thể nghe có vẻ đơn giản, nhưng quá trình này có một điều kiện. Mức độ khó khăn trong việc tìm kiếm một băm chấp nhận được sẽ được tăng lên theo các khoảng thời gian đều đặn. Vào thời điểm viết bài này, mức độ khó khăn hiện tại là 12.9 nghìn tỷ. Điều này có nghĩa là khả năng một máy tính xử lý được một băm dưới mục tiêu đã chỉ định là 1 trong 12.9 nghìn tỷ. Điều này đặt ra câu hỏi: Tại sao những người khai thác cryptocurrency lại dành thời gian, năng lượng và công suất

tính toán để khai thác? Các thợ đào được thưởng bằng bitcoin của chính họ khi khai thác thành công một khối.

Sản phẩm blockchain của AWS

- AWS cung cấp ba sản phẩm blockchain để tạo các giải pháp blockchain dựa trên đám mây.
 - Amazon Quantum Ledger Database (Amazon QLDB): Một cơ sở dữ liệu sổ cái được quản lý hoàn toàn, cung cấp một nhật ký giao dịch minh bạch, bất biến và có thể xác minh bằng mã hóa, do một cơ quan trung gian đáng tin cậy sở hữu. Amazon QLDB theo dõi mọi thay đổi dữ liệu ứng dụng và duy trì lịch sử đầy đủ và có thể xác minh của các thay đổi theo thời gian.
 - Amazon Managed Blockchain: Một dịch vụ được quản lý hoàn toàn cho phép bạn thiết lập và quản lý một mạng blockchain có thể mở rộng chỉ với vài cú nhấp chuột bằng các framework mã nguồn mở phổ biến như Hyperledger Fabric và Ethereum.
- AWS Blockchain Templates: Cho phép bạn triển khai framework blockchain bạn chọn dưới dạng các container trên Amazon Elastic Container Service (Amazon ECS) hoặc trực tiếp trên một phiên bản Amazon Elastic Compute Cloud (Amazon EC2) chạy Docker.
- Câu chuyện thành công của khách hàng AWS
- Nestlé là một công ty thực phẩm và đồ uống toàn cầu với hơn 2.000 thương hiệu và phân phối toàn cầu tại 189 quốc gia trên thế giới.
- “Minh bạch trong chuỗi cung ứng ngày càng trở nên quan trọng đối với người tiêu dùng, những người muốn biết những gì có trong thực phẩm của họ và nguồn gốc của chúng. Mặc dù Nestlé đã bắt đầu công bố thông tin về chuỗi cung ứng cho 15 mặt hàng chủ yếu của mình, việc sử dụng công nghệ blockchain cho phép theo dõi chính xác hơn,” Armin Nehzat, Giám đốc Công nghệ Kỹ thuật số của Nestlé Oceania cho biết. “Với Amazon Managed Blockchain, chúng tôi có thể thiết lập mạng lưới Hyperledger Fabric của mình và dễ dàng mời các đối tác tham gia hợp tác trong nỗ lực minh bạch chuỗi cung ứng. Amazon Managed Blockchain sẽ cho phép khách hàng của chúng tôi theo dõi sản phẩm của họ trên blockchain từ nông trại đến tiêu dùng.”
- Armin Nehzat, Giám đốc Công nghệ Kỹ thuật số—Nestlé Oceania

Electroneum

- Tại Bắc Mỹ, khoảng 95% người lớn trên 25 tuổi có tài khoản ngân hàng. Trong khi đó, ở các nước đang phát triển, con số này chỉ khoảng 52%. Electroneum là một công ty tiền điện tử có trụ sở tại Anh, xây dựng nền tảng di động tiền điện tử của mình trên AWS và đang sử dụng sức mạnh của blockchain để mở khóa nền kinh tế kỹ thuật số toàn cầu cho hàng triệu người ở các quốc gia đang phát triển. Tiền điện tử có thể cung cấp nền tảng cho hàng triệu người không có ngân hàng trên thế giới đạt được tự do tài chính trên sân chơi tài chính bình đẳng hơn.

3. Question

1. Consider an activity that happens every day: buying a used car from another person. Two people meet, agree on the price of the car, exchange money, and then the owner of the car transfers the title to the person buying the car. Transferring the title involves signing papers and registering the sale of the car with the correct government department. What is the purpose of the title? Why does the government department need to be involved? What would happen if the government department didn't keep track of motor vehicle sales?

Giấy tờ xe (title) chứng minh quyền sở hữu hợp pháp của người mua xe. Chính phủ cần quản lý thông tin này để xác nhận quyền sở hữu, theo dõi thuế, và ngăn chặn gian lận. Nếu không có sự quản lý này, sẽ dễ xảy ra các tranh chấp về quyền sở hữu và khó kiểm soát tình trạng xe bị đánh cắp hoặc gian lận giao dịch.

2. What are some examples of everyday activities that people engage in that require registering a record of the activity with a neutral third party, like the government, to record the activity or verify the information provided by the parties involved?

Một số ví dụ bao gồm:

Mua bán bất động sản (đăng ký quyền sở hữu đất).

Đăng ký kết hôn (cần có chứng nhận hợp pháp).

Mở tài khoản ngân hàng (xác minh danh tính). Việc đăng ký giúp đảm bảo thông tin chính xác, tránh tranh chấp, và bảo vệ quyền lợi của các bên liên quan.

3. Choose an example of a transaction that people engage in every day that is made more complicated and slow because of the need for registering with a third party. List all parties involved in the transaction. Describe the opportunities that arise or the

problems that can be averted if the transaction can be verified and is immutable—that is, able to remain unchanged—and does not need a neutral third party.

Giao dịch bất động sản thường chậm và phức tạp do phải qua quá trình đăng ký quyền sở hữu với chính phủ. Các bên liên quan gồm: người mua, người bán, công chứng viên, ngân hàng, và cơ quan nhà nước. Nếu giao dịch được xác thực tự động và không cần bên thứ ba, sẽ tiết kiệm thời gian, giảm chi phí, và hạn chế rủi ro sai sót hoặc gian lận.

4. Cryptocurrencies are a form of digital money that use technology to replace banks as neutral third parties to record and verify transfers of money. How familiar are you with cryptocurrency? Have you used or invested in cryptocurrency? Would you? Why or why not?

Tôi có kiến thức cơ bản về cryptocurrency nhưng chưa từng sử dụng hoặc đầu tư. Tôi vẫn chưa chắc chắn về việc tham gia vì rủi ro biến động giá cao, chưa có quy định rõ ràng, và nguy cơ bị lừa đảo. Tuy nhiên, tôi thấy tiềm năng của nó trong việc thay thế ngân hàng, giảm chi phí giao dịch, và tăng cường bảo mật.