Bảo mật dữ liệu từ góc nhìn quản trị dữ liệu tại Việt Nam

**Dữ liệu đóng vai trò hết sức quan trọng và đang thay đổi cách thức kinh doanh của nhiều tổ chức, doanh nghiệp (TC/DN). Vậy, các TC/DN Việt Nam đã và đang quản lý, khai thác dữ liệu thế nào? Dữ liệu có được bảo mật hay không? Bài viết này chia sẻ về góc nhìn trong việc bảo mật thông tin dưới khía cạnh của người làm quản trị dữ liệu trong các ngân hàng.**

Năm 2006, Clive Humby - nhà toán học và kiến trúc người Anh làm việc cho Câu lạc bộ thẻ Tesco đã lần đầu tiên đưa ra quan điểm: “Dữ liệu là nguồn dầu mỏ mới. Dữ liệu rất giá trị, nhưng nếu không được xử lý thì gần như không sử dụng được. Giống như dầu mỏ cần được chuyển hóa thành khí gas, nhựa plastic, xăng... để tạo ra các thực thể có giá trị, đem lại lợi nhuận cho tổ chức; dữ liệu cần được chia nhỏ, phân tích để tạo ra giá trị”.

Dữ liệu có ở xung quanh và đang từng giờ thay đổi cuộc sống của chúng ta. Lấy một ví dụ đơn giản, nền tảng chia sẻ video Youtube của Google trả tiền cho các Youtuber tạo ra nội dung chia sẻ dựa trên lượt view. Điều này đặt ra câu hỏi: Chi phí nào để Youtube trả tiền cho các Youtuber? Thực tế, Youtube chèn các quảng cáo vào trong nội dung video. Điều này đồng nghĩa với việc, khi người dùng xem video là sẽ phải xem các nội dung quảng cáo. Việc nội dung quảng cáo phản ánh sản phẩm nào sẽ phụ thuộc hoàn toàn vào việc thu thập, phân tích hành vi xem video của người dùng. Nói đến đây, chắc hẳn độc giả đã thấy tại sao dữ liệu lại là "nguồn dầu mỏ mới".

Nếu bạn vẫn chưa thực sự bị thuyết phục rằng dữ liệu là “nguồn dầu mỏ mới”, thì hãy đọc cuốn sách “Dữ liệu lớn – Big Data” của các tác giả Viktor Mayer-Schonberger và Kenneth Cukier. Ngay từ chương đầu tiên của cuốn sách đã mô tả việc các nhà khoa học dữ liệu của Google chỉ ra phương pháp làm sao để biết mức độ di chuyển và phạm vi của một dịch cúm. Dựa trên việc nghiên cứu hàng tỉ câu hỏi trên [trang tìm kiếm của Google](https://www.google.com/) từ khắp nơi trên thế giới, thông qua các từ khóa tìm kiếm liên quan đến dịch cúm. Từ đó, các nhà khoa học dữ liệu Google có thể bản đồ hóa khu vực có các từ khóa tìm kiếm liên quan để biết mức độ phát triển và phạm vi của dịch cúm. Việc thu thập, lưu trữ, xử lý hàng tỉ câu hỏi từ khắp nơi trên thế giới không phải là công việc dễ dàng. Tuy nhiên, kết quả mà nó đem lại có giá trị to lớn vào thời điểm dịch cúm xảy ra.

**Dữ liệu bị phát tán từ đâu?**

Phần lớn chúng ta đều nghĩ rằng dữ liệu của các tổ chức tài chính ngân hàng bị rò rỉ là do hacker thực hiện các cuộc tấn công vào hệ thống cơ sở dữ liệu của ngân hàng để truy cập trái phép dữ liệu nhạy cảm của khách hàng. Tuy nhiên, đó chỉ là một phần của vấn đề, có rất nhiều trường hợp dữ liệu bị phát tán từ chính những người trong tổ chức, một cách vô tình hay cố ý.

Điển hình, năm 2013 một cán bộ tín dụng của một ngân hàng thương mại cổ phần (TMCP) đã mượn tài khoản của đồng nghiệp để truy cập trái phép và sao lưu dữ liệu của hơn 1.000 khách hàng có số dư tiết kiệm trên 500 triệu VNĐ. Dữ liệu này được sử dụng cho mục đích cá nhân và bán cho các nhân viên của ngân hàng khác. Khi bị phát hiện, cán bộ này đã bị sa thải và phải chịu án tù treo 2 năm. Tuy nhiên, chưa có ai kiểm chứng được dữ liệu đã được bán cho những ai và bao nhiêu lần?

Bên cạnh đó, hình thức công khai bán dữ liệu khách hàng cũng rất phổ biến trên Internet. Lợi dụng tính ẩn danh của google, [một website](https://sites.google.com/site/danhsachkhachhangdata/) đã công khai rao bán dữ liệu của hơn 4 triệu khách hàng ở mọi lĩnh vực, trong đó có lĩnh vực ngân hàng. Một điều ngạc nhiên là dữ liệu được bán có thể phân loại như khách hàng có dùng thẻ và loại thẻ. Những dữ liệu này chỉ có được khi nguồn phát tán từ chính nội bộ của ngân hàng đó.

Vậy các ngân hàng quản lý dữ liệu của mình thế nào? Vấn đề này được giải quyết trên hai khía cạnh: nhận thức về giá trị dữ liệu của ngân hàng và hiện trạng quản lý dữ liệu tại các ngân hàng.

Hiện trạng bảo mật, chia sẻ dữ liệu nội bộ trong ngân hàng

Nhận thức về giá trị dữ liệu của ngân hàng

Năm 2018, một khảo sát về mức độ trưởng thành quản trị dữ liệu trong phạm vi nhỏ đã được thực hiện khi triển khai Khung Quản trị dữ liệu tại Việt Nam. Nội dung khảo sát được xây dựng dựa trên hai mô hình nổi tiếng trên thế giới: Mô hình đánh giá năng lực quản lý dữ liệu (Data Management Capabilities Assessment Model - DCAM) do EDM Council phát triển và Đánh giá trưởng thành quản lý dữ liệu (Data Management Maturity Assessment - DMMA) do DAMA International.

Mô hình trưởng thành quản trị dữ liệu được sử dụng trong khảo sát bao gồm 6 mức trưởng thành (Chưa bắt đầu, Có ý tưởng, Đang phát triển, Đã xác định, Hoàn thành và Tối ưu) xét trên 3 khía cạnh (Quy trình, mức độ gắn kết và tổ chức). Kết quả khảo sát cho thấy, các ngân hàng Việt Nam đang ở mức đầu tiên trong 6 mức trưởng thành quản trị dữ liệu. Điều này cho thấy các ngân hàng Việt Nam chưa thực sự chú trọng vào việc quản trị dữ liệu và chưa có mức đầu tư xứng đáng vào công tác này. Tại thị trường Mỹ hay các nước phát triển ở Châu Âu, các ngân hàng đang xếp ở mức 4 (Đã xác định) và đang tiến tới mức 5 (Hoàn thành). Còn khối thị trường đang phát triển ở Châu Á và Châu Phi, thì các ngân hàng đang ở mức 3 là mức Đang phát triển. Như vậy, có một khoảng cách lớn giữa thị trường Việt Nam và các nước trong khu vực và trên thế giới.

Thực tế, thuật ngữ “Quản trị dữ liệu” mới thực sự được các ngân hàng quan tâm từ năm 2014. Hiện tại, việc triển khai quản trị dữ liệu tại các ngân hàng Việt Nam còn rất hạn chế, thậm chí còn rất nhiều các ngân hàng mới có ý tưởng manh nha về vấn đề này. Một tín hiệu tích cực là một số lãnh đạo ngân hàng đã nhìn nhận được tầm quan trọng của dữ liệu cho định hướng kinh doanh của ngân hàng. Tuy nhiên, thông điệp này vẫn chưa được truyền tải tới các cán bộ nhân viên của ngân hàng, những người đang hàng ngày tiếp xúc với dữ liệu.

Về mặt tổ chức đơn vị quản trị dữ liệu, phần lớn các ngân hàng đang sử dụng mô hình phân tán, nghĩa là tồn tại nhiều đơn vị thu thập và khai thác dữ liệu độc lập.

Với mô hình phân tán này, mỗi khối nghiệp vụ sẽ có một phòng, ban thu thập và xử lý dữ liệu độc lập, việc này dẫn đến dữ liệu sẽ bị quản lý phân tán, các đơn vị khai thác cùng một loại dữ liệu sẽ không tối ưu được nguồn lực, dễ dẫn đến thông tin cung cấp cho các cấp lãnh đạo không đồng nhất làm cho việc ra quyết định kinh doanh gặp nhiều khó khăn. Với mô hình phân tán như vậy, việc chia sẻ hay truy cập vào dữ liệu không phù hợp với vị trí công việc rất dễ xảy ra và gây khó khăn trong việc kiểm soát. Đây là nguyên nhân cho việc dữ liệu khách hàng bị phát tán ra bên ngoài mà không kiểm soát.

***Hiện trạng quản lý dữ liệu tại các ngân hàng***

Việc quản lý dữ liệu ngân hàng hiện nay cơ bản chưa được thống nhất. Từ năm 2009 đến nay, phần lớn các ngân hàng đều đã triển khai hệ thống Kho Dữ liệu (Data Warehouse). Tuy nhiên, giá trị đem lại của Kho dữ liệu này rất khó kiểm chứng, khi mà Kho dữ liệu này không phải là duy nhất. Các ngân hàng đang tồn tại song song nhiều kho dữ liệu khác nhau, bản thân các đơn vị nghiệp vụ cũng có kho dữ liệu lưu trữ của riêng mình và gần như không sử dụng kho dữ liệu chung của toàn ngân hàng.

Một điểm đáng lo khác là luồng dữ liệu luân chuyển giữa các hệ thống hiện tại chưa được quy hoạch và cũng chưa được ghi nhận trên tài liệu, dẫn đến rất nhiều khó khăn cho việc quản lý và khai thác dữ liệu. Việc tồn tại song song nhiều kho dữ liệu, luồng dữ liệu chưa được chuẩn hóa và ghi nhận trên tài liệu là một trong những nguyên nhân dẫn đến nguy cơ rò rỉ thông tin.

Như vậy, ở cả hai góc độ là nhận thức về giá trị của dữ liệu và hiện trạng quản lý dữ liệu của ngân hàng đều tồn tại những vấn đề cần phải khắc phục. Làm sao để ngân hàng cải thiện việc quản trị dữ liệu của mình, qua đó nâng cao bảo mật thông tin? Câu trả lời này sẽ được giải đáp tại phần tiếp theo của bài báo. Kính mời quý độc giả đón đọc.

*Nguyễn Minh Đức - Ngân hàng TMCP Quân đội (MB Bank)*