**Mô tả đặc điểm của một hệ thống phù hợp với DynamoDB**

AWS DynamoDB là service database NoSQL (do chính AWS phát triển) với đặc tính fast và flexible, dành cho tất cả các application đòi hỏi tính consistent & latency thấp với khả năng scaling tốt. DynamoDB chuyện dụng cho mobile, web, gaming, ad-tech, Internet of Things, ...

Những năm gần đây, nhiều doanh nghiệp đang nhu yếu giải quyết và xử lý tài liệu nhanh gọn để phân phối nhu yếu của số lượng lớn người dùng như shopping, quảng cáo trên Internet, game show, shop thương mại điện tử …

Nói cách khác, các công ty có nhu cầu với xử lý dữ liệu lớn tốc độ cao như vậy đang phát triển nhanh chóng. Và một trong những dịch vụ được sử dụng để xử lý dữ liệu hiệu quả trong trường hợp này là **Amazon DynamoDB**.

**DynamoDB** tự động phân tán **dữ liệu và traffic** của một bảng ra một số lượng server vừa đủ để có thể xử lý request capacity đặt ra bởi khách hàng và lượng dữ liệu lưu trữ, và đồng thời đảm bảo hiệu suất nhanh và đồng nhất. Tất cả dự liệu được **lưu trữ trên SSD** và tự động được sao chép ra các vùng sẵn sàng (Availability Zones) trong một khu vực (Region) để cung cấp độ sẵn sàng cao và độ bền của dữ liệu (high availablity and data durability).

Trong **DynamoDB**, bạn thiết kế lược đồ của mình một cách cụ thể để thực hiện các truy vấn phổ biến và quan trọng nhất nhanh nhất và ít tốn kém nhất có thể. Cấu trúc dữ liệu của bạn được điều chỉnh cho phù hợp với các yêu cầu cụ thể của các trường hợp sử dụng kinh doanh của bạn.

**Data** là cơ sở để ra các quyết định cho mọi hoạt động trong kinh doanh. Nhu cầu báo cáo đa dạng được tổng hợp từ mọi nguồn dữ liệu có thể có trong hệ thống. Đối tượng người dùng bao gồm từ các cấp vận hành thấp nhất tới các cấp giám đốc cao nhất. Tính chính xác, sự đa dạng, tốc độ cập nhật của dữ liệu là các đòi hỏi của hệ thống báo cáo.

**Các yếu tố cần xem xét khi chọn cơ sở dữ liệu**

Khi chọn một cơ sở dữ liệu từ nhiều sản phẩm khác nhau, nên đặt ra các câu hỏi:

1. **Kích thước dữ liệu cho cơ sở dữ liệu mới của chúng tôi có vượt quá 1 TB trong 12–18 tháng tới** không?

Nếu có, chúng ta nên xem xét **DynamoDB**. **DynamoDB** vừa có khả năng mở rộng vừa có tính đàn hồi cao. Không có giới hạn về số lượng dữ liệu mà bạn có thể lưu trữ trong bảng **DynamoDB** và dịch vụ tự động phân bổ thêm dung lượng khi bạn lưu trữ nhiều dữ liệu hơn bằng cách sử dụng các hoạt động API ghi **DynamoDB**.

1. **Một bảng MySQL sẽ có hơn 10 triệu hàng hay 10 GB dữ liệu?**

Nếu có, chúng ta có thể sử dụng **DynamoDb** để đạt được khả năng mở rộng theo chiều ngang và tránh tiêu tốn công sức cho sharding. Đó là bởi vì khi một bảng MySQL chứa hơn 10 triệu hàng dữ liệu hoặc một bảng duy nhất chiếm hơn 10 GB dung lượng đĩa, thì cơ sở dữ liệu có thể chậm hơn và khó bảo trì. Ví dụ, một số truy vấn SQL rất khó để tối ưu hóa và ngôn ngữ định nghĩa dữ liệu trực tuyến trở thành một thách thức lớn. Một số người dùng **DynamoDB** có các bảng vượt quá 100 TB nhưng hiệu suất đọc và ghi của các bảng này không khác gì so với bảng có kích thước dưới 1 GB.

1. **Các truy vấn ghi mỗi giây (QPS) cho một nút MySQL có vượt quá 1.000 lần ghi mỗi giây không?**

Nếu có, có thể thử các cách tiếp cận sau:

* Trước khi đưa một ứng dụng vào sản xuất, ước tính có thể không chính xác. Trong điều kiện bình thường, có thể có 1.000 lần ghi mỗi giây. Nhưng khi thực hiện một chiến dịch xúc tiến bán hàng lớn, QPS viết có thể tăng lên 10.000 lần ghi mỗi giây. Tốt hơn là đặt một giá trị tham chiếu thận trọng.
* Lợi ích chính khi sử dụng **DynamoDB** làm kho lưu trữ NoSQL là bạn nhận được thông lượng đọc / ghi được đảm bảo ở bất kỳ mức nào bạn yêu cầu mà không phải lo lắng về việc quản lý một kho dữ liệu được phân nhóm. Vì vậy, nếu ứng dụng của bạn yêu cầu 1000 lần đọc / ghi mỗi giây, bạn chỉ có thể cung cấp bảng **DynamoDB** của mình cho mức thông lượng đó và không thực sự phải lo lắng về cơ sở hạ tầng bên dưới. DynamoDB mang lại hiệu suất lên tới một chữ số mili giây ở bất kỳ quy mô nào.

1. **Khi không thể ước tính chính xác dung lượng của cơ sở dữ liệu**

Ví dụ: kích thước dữ liệu gia tăng cho cơ sở dữ liệu nhật ký trực tuyến trong ba tháng qua vượt quá kích thước trong ba năm qua. Nếu chúng ta triển khai sharding, chúng ta cần phải nâng cấp lại loại cơ sở dữ liệu này nhiều lần. Mỗi quá trình sharding rất phức tạp và tốn nhiều công sức.

**DynamoDB** **Streams** là một tính năng tuyệt vời giúp ghi lại các thay đổi đối với bảng tại thời điểm thay đổi xảy ra, lưu trữ các thay đổi này trong nhật ký được lưu giữ trong 24 giờ.

Bản ghi luồng này được ghi với các thuộc tính khóa chính của các mục đã được thay đổi, đảm bảo rằng tất cả các sửa đổi được ghi lại theo cùng một trình tự mà chúng đã xảy ra. Nó cũng có thể được cấu hình để thu thập dữ liệu bổ sung như trạng thái trước và sau khi thay đổi.

1. **Khi sử dụng cơ sở dữ liệu để chạy các truy vấn phức tạp đa chiều**

Lấy cơ sở dữ liệu đơn hàng thương mại điện tử làm ví dụ. Mỗi hệ thống con cần lọc dữ liệu theo người mua, người bán, trạng thái đơn hàng và phương thức thanh toán. Nếu chúng ta chia nhỏ theo người mua thì sẽ khó truy vấn thông tin của người bán và ngược lại.

Khi thêm một chỉ mục phụ tổng thể vào bảng hiện có, **DynamoDB** sẽ lấp đầy chỉ mục một cách không đồng bộ với các mục hiện có trong bảng. Chỉ mục sẽ có thể truy vấn sau khi tất cả các mục đã được lấp đầy. Thời gian lấp đầy phù thuộc vào kích thước của bảng.

1. **Khi dữ liệu được phân phối không đồng đều trong cơ sở dữ liệu**

Đối với dữ liệu ứng dụng xã hội như "Thích" và "Theo dõi", nếu chia nhỏ cơ sở dữ liệu theo người dùng, dữ liệu có thể được phân phối không đồng đều. Khối lượng dữ liệu cho một số ít phân đoạn có thể cao hơn đáng kể so với các phân đoạn khác. Những mảnh này với lượng lớn dữ liệu cũng là điểm nóng để đọc và ghi. Điều này dễ dẫn đến tắc nghẽn hiệu suất.

Một phương pháp phổ biến là resharding: chia dữ liệu thành nhiều phân vùng hơn để giảm kích thước dữ liệu và lưu lượng đọc-ghi của mỗi phân đoạn.

Partition key trong **DynamoDB** của một bản ghi còn được gọi là hash attribute (thuộc tính hàm băm). Thuật ngữ hast attribute được bắt nguồn từ việc sử dụng một hàm băm nội bộ trong **DynamoDB** để phân phối đồng đều các mục dữ liệu trên các phân vùng, dựa trên các giá trị partition key của chúng.

**Nền tảng phù hợp với DynamoDB**

Shopee là nền tảng thương mại điện tử hàng đầu Đông Nam Á và Đài Loan. Đây là một nền tảng được thiết kế riêng cho khu vực, mang đến cho khách hàng trải nghiệm mua sắm trực tuyến dễ dàng, an toàn và nhanh chóng thông qua hỗ trợ hậu cần và thanh toán mạnh mẽ.

Năm nay, shopee tạo ra số lượng cơ sở dữ liệu mỗi tuần nhiều gấp mười lần so với cách đây một hoặc hai năm. Mỗi năm, số lượng cơ sở dữ liệu logic trong môi trường sản xuất tăng lên khoảng ba đến năm lần. Năm 2019, con số này tăng hơn năm lần.

Theo thống kê, 80% cơ sở dữ liệu sản xuất của Shopee có ít hơn 50 GB dữ liệu. Trong số tất cả các cơ sở dữ liệu sản xuất, 2,5% có nhiều hơn 1 TB dữ liệu. Kích thước dữ liệu trung bình của các cơ sở dữ liệu quy mô terabyte này là 2 TB. Lớn nhất có hơn 4 TB dữ liệu.

Để giải quyết kích thước dữ liệu tăng vọt cho cơ sở dữ liệu, chúng ta có thể thử dùng DynaDB. Chúng ta cũng có thể lưu trữ dữ liệu cũ và nâng cấp tài nguyên phần cứng.

DynamoDB là một lựa chọn tốt để lưu trữ dữ liệu. Về mặt lý thuyết, một cụm DynamoDB có thể mở rộng vô hạn và người dùng không cần phải lo lắng về dung lượng ổ đĩa bị giới hạn. DynamoDB có khả năng mở rộng theo chiều ngang đàn hồi trong cả lớp tính toán và lớp lưu trữ. Do đó, chúng ta có thể dần dần thêm các máy chủ theo tốc độ tăng trưởng dữ liệu và lưu lượng đọc ghi ứng dụng. Điều này giúp đảm bảo việc sử dụng phần cứng hiệu quả và ngăn nhiều tài nguyên không hoạt động trong giai đoạn sản xuất ban đầu của cơ sở dữ liệu lưu trữ.

Tóm lại, nếu kích thước dữ liệu của bạn nhỏ và ứng dụng của bạn đang ở giai đoạn đầu, MySQL là một lựa chọn tốt cho bạn. Bạn cũng không cần phải ảnh hưởng đến thiết kế ứng dụng của mình cho sharding. Đó là bởi vì khi doanh nghiệp của bạn phát triển và khối lượng dữ liệu tăng lên, bạn có thể di chuyển dễ dàng từ MySQL sang DynamoDB. Ứng dụng của bạn sẽ có thể mở rộng theo chiều ngang trong khi việc phát triển ứng dụng vẫn linh hoạt.

Shopee là nền tảng thương mại điện tử hàng đầu Đông Nam Á và Đài Loan. Nó đã được ra mắt tại 7 thị trường vào năm 2015 để kết nối người tiêu dùng, người bán và doanh nghiệp trong khu vực. Shopee mang đến trải nghiệm dễ dàng, an toàn và hấp dẫn được hàng triệu người yêu thích hàng ngày. Nó cung cấp nhiều loại sản phẩm, được hỗ trợ bởi thanh toán và hậu cần tích hợp, cũng như các tính năng giải trí phổ biến phù hợp với từng thị trường. Shopee cũng là công ty đóng góp quan trọng vào nền kinh tế kỹ thuật số của khu vực với cam kết vững chắc trong việc giúp các thương hiệu và doanh nhân thành công trong lĩnh vực thương mại điện tử.