Microsoft giới thiệu cơ sở dự liệu NoSQL cho Microsoft Azure

Microsoft có lẽ đồng nghĩa với nhắc tới cơ sở dữ liệu quan hệ nhờ sản phẩm chủ lực SQL server của họ, nhưng điều đó đã thực sự thay đổi khi họ đưa ra 1 sản phẩm mới NoSQL. Tuần trước, Microsoft đã công bố bản giới thiệu của DocumentDB, một cơ sở dự liệu điện toán đám mây quản lý dữ liệu với sự hỗ trợ của JavaScript sâu và các tính năng như tự động lập chỉ mục và các thao tác.

DocumentDB là một lược đồ thu gọn, cơ sở dữ liệu dùng chung (SSD) này lưu trữ dữ liệu JSON và đưa ra một API REST. Theo Microsoft, dịch vụ Microsoft Azure này đáp ứng một lớp nhu cầu mới trong khi không bỏ đi các khái niệm hữu ích của cơ sở dữ liệu quan hệ.

“Chúng tôi nghe nói từ khách hàng rằng họ cần một cơ sở dữ liệu có thể bắt kịp với sự phát sinh nhanh chóng về dữ liệu trong các ứng dụng của họ - một cái gì đó nhanh, linh hoạt và có khả năng mở rộng. Càng ngày, cơ sở dữ liệu NoSQL đang trở thành công cụ của sự lựa chọn cho nhiều nhà phát triển ứng dụng nhưng chạy và quản lý các cơ sở dữ liệu này có thể tốn kém, đặc biệt là về quy mô. Chúng tôi cũng nghe nói rằng khách hàng muốn thêm vào các khả năng vốn có của các hệ thống cơ sở dữ liệu quan hệ - truy vấn phong phú và xử lý các giao thức quan trọng. Hầu hết các kho lưu trữ dữ liệu cung cấp các lựa chọn để phát triển mạnh mẽ hoặc tính nhất quán, các lược đồ tự do với khả năng truy vấn hạn chế hoặc biểu đồ hóa và các truy vấn phong phú, các giao dịch hay sự mở rộng. Thực tế là rất nhiều trường hợp tồn tại giữa hai điều này và chúng tôi muốn giải quyết chúng”

Làm thế nào để DocumentDB tương tự như cơ sở dữ liệu quan hệ truyền thống? Các dữ liệu được truy vấn bằng cách sử dụng cú pháp SQL đơn giản để nhận các cấu trúc JavaScriptnhw đối tượng và mảng. Khoong giống nhiều cơ sở dữ liệu NoSQL phổ biến mà chỉ dựa vào dữ liệu cuối cùng phù hợp, DocumentDB cho thấy bốn mô hình nhất quán khác nhau.

“DocumentDB cung cấp bốn cấp độ nhất quán riêng biệt cho đọc và truy vấn : Strong, Bounded Staleness, Seasion, và Eventual. Các mức độ cho phép bạn cân bằng giữa tính nhất quán, tính sẵn sàng và độ trễ. Bounded Staleness đảm bảo toàn bộ thứ tự viết cũng như hạn chế tối đa lỗi hỏng, mức độ này hữu ích cho các ứng dụng xử lý thời gian và yêu cầu thứ tự hoạt động. Seasion cung cấp đảm bảo cho đọc viết của riêng bạn và có thể là một lựa chọn tốt cho các ứng dụng trung tâm đa người dùng”

Các chỉ số dữ liệu và giao dịch cũng có khả năng được tùm thấy như trong một cơ sở dữ liệu quan hệ

“Khi bạn thêm các dữ liệu vào bộ dữ liệu, DocumentDB tự động thêm chỉ số cho chúng và chúng có thể sẵn sàng để truy vấn. Tự động lập chỉ mục dữ liệu mà không đòi hỏi chỉ số quan hệ thứ cấp là một khả năng quan trọng của DocumentDB và được kích hoạt bằng cách viết tối ưu hóa.”

Có một loạt các tùy chọn cấu hình trong khi tạo chỉ mục DocumentDB. Trong khi lập chỉ mục tự động được kích hoạt theo mặc định, bạn có thể tắt hành vi này hoặc tinh chỉnh nó. Một số dữ liệu có thể bao gồm hoặc loại trừ chỉ số dựa trên đường dẫn hoặc mô hình, và cập nhật chỉ số bằng chạy đồng bộ hoặc không đồng bộ.

Dịch vụ này cung cấp một số ít lựa chọn để áp dụng logic kinh doanh phía máy chủ. DocumentDB hỗ trợ chức năng người dùng tự định nghĩa, thủ tục lưu trữ, và những trigger.

“Nhờ những xử lý sâu cho JavaScript và JSON trực tiếp trong cơ sở dữ liệ, DocumentDB cung cấp một mô hình lập trình trực quan để thực hiện dựa trên JavaScript ứng dụng trực tiếp trên các bộ dữ liệu về thủ tục lưu trữ và các trigger. Điều này cho phép thực hiện hiệu quả kiểm soát đồng thời, phục hồi, tự động lập chỉ mục của đối tượng JSON trực tiếp vào cơ sở dữ liệu cũng như cách kiểm soát dòng dữ liệu, phạm vi biến, phân định, phân tích các ngoại lện khi giao dịch trực tiếp với cơ sở dữ liệu sử dụng JavaScript”

…

“DocumentDB ngầm kết thúc các thủ tục lưu trữ dựa trên JavaScript và kích hoạt một giao thức ACID cô lập với môi trường xung quanh với lưu trữ nhanh dữ liệu trong tập dữ liệu. Trong quá trình thực hiện nó, nếu JavaScript ném một ngoại lệ, toàn bộ giao dịch sẽ được hủy bỏ.”

…

“Lưu ý rằng DocumentDB không yêu cầu bất kỳ quy ước JSON đặc biệt để hệ thống hóa các mối quan hệ giữa các dữ liệu khác nhau, ngôn ngữ truy vấn SQL của DocumentDB cung cấp cho các nhà khai thác truy vấn có thứ tự và quan hệ rất mạnh mẽ để truy vấn dữ liệu mà không có chú thích đặc biệt hoặc phải hệ thống hóa các mối quan hệ giữa các dữ liệu bằng cách sử dụng thuộc tính phân biệt.”

Làm thế nào để thực hiện cơ sở dữ liệu như thế ? GigaOm được báo cáo rằng DocumentDB sử dụng công cụ Hekaton trong bộ nhớ từ SQL Server 2014, phó chủ tịch Microsoft – Scott Guthrie tiết lộ rằng công nghệ này hiện đang được sử dụng rất hiều tại Microsoft.

“Trong năm qua, chúng tôi đã sử dụng DocumentDB trong nội bộ Microsoft cho một số dịch vụ cao cấp. Bây giờ chúng ta có cơ sở dữ liệu DocumentDB hàng TB mỗi 100 giây về kích thước, xử lý hàng triệu các truy vấn phức tạp mỗi ngày, với hiệu suất dự đoán về độ trễ thấp vài mili giây.”

Trong vấn đề này, các dịch vụ NoSQL trên Microsoft Azure. Công ty vẫn cung cấp các kho lưu trữ sử dụng NoSQL lưu trữ theo kiểu key-value có tên là Azure Table Storage, và mới đây họ đã tung ra một dịch vụ bộ nhớ đệm dựa trên Redis. Điều này có nghĩa là MongoDB gần đây đã được bổ sung thêm một dịch vụ quản lý trên trang của Microsoft Azure.

Trong khi DocumentDB chỉ được phát hành bằng một bản preview, Microsoft vẫn cung cấp SDKs cho .NET, Node.js, JavaScript và Python. Giá của sản phẩm được dựa trên “đơn vị dung lượng”. Một đơn vị dung lượng có khả năng đọc 2000 lệnh mỗi giây, 500 lệnh chèn/ thay thế/ xóa các hoạt động mỗi giây, và 10GB dung lượng lưu trữ, Trong thời gian phát hành bản preview, giá cho một đơn vị dung lượng là 0,73$ mỗi ngày, giảm 50% so với giá thương mại dự kiến. DocumentDB hiện đang có sẵn ở Microsoft Tây Mỹ, Microsoft Azure, Bắc Châu Âu và khu vực Tây Châu Âu.

*Nguồn tham khảo: http://www.infoq.com/news/2014/08/microsoft-azure-documentdb*