

ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC VÀ KỸ THUẬT MÁY TÍNH

—o0o—



HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU
Document Store (MongoDB, CouchDB, SimpleDB)

Giảng viên hướng dẫn: Thầy Trần Quang

Sinh viên thực hiện:

Họ và tên	MSSV
Trương Ngọc Anh	1410141
Lê Nguyễn Minh Trí	1414207
Nguyễn Quốc Kim Hoàng	141xxxx
Nguyễn Khánh Bình	141xxxx

Mục lục

1	Giới thiệu Document store và MongoDB	5
1.1	5
1.2	5
2	Thiết kế cơ sở dữ liệu với MongoDB	6
2.1	•	6
2.2	•	6
3	Sharding	7
4	Các cơ sở dữ liệu Document store	8
4.1	Tổng quan về CouchDB	8
4.1.1	CouchDB là gì ?	8
4.1.2	Tại sao chúng ta lại cần CouchDB ?	8
4.1.3	Mô hình dữ liệu	9
4.2	Các tính năng chính	9
4.2.1	Lưu trữ dạng document	9
4.2.2	Các thuộc tính ACID	10
4.2.3	Khả năng nén(compaction)	10

4.2.4	Views	11
4.3	Tổng quan về SimpleDB	11
4.3.1	SimpleDB là gì ?	11
4.4	Tại sao chúng ta lại cần SimpleDB ?	12
4.4.1	Lợi ích	12
4.5	Một số chức năng Use Cases	13
4.5.1	Login	13
4.5.2	Online games	13
4.6	So sánh	14

Danh sách hình vẽ

Danh sách bảng

Chương 1

Giới thiệu Document store và MongoDB

1.1

1.2

Chương 2

Thiết kế cơ sở dữ liệu với MongoDB

2.1 •

2.2 •

Chương 3

Sharding

Cái này tự chỉnh lại tiêu đề cho hợp lý

Chương 4

Các cơ sở dữ liệu Document store

4.1 Tổng quan về CouchDB

4.1.1 CouchDB là gì ?

- CouchDB là một cơ sở dữ liệu dạng NoSQL mã nguồn mở database lưu trữ dữ liệu dạng document/JSON.
- CouchDB được thiết kế nhằm tối tính dễ sử dụng và phục vụ cho môi trường web.

4.1.2 Tại sao chúng ta lại cần CouchDB ?

- CouchDB có API dạng RESTFul giúp cho việc giao tiếp với cơ sở dữ liệu được đơn giản.

- Các RESTFul API rất trực quan và dễ thao tác.
- Dữ liệu được lưu dưới cấu trúc document rất mềm dẻo, chúng ta không cần phải lo lắng về cấu trúc dữ liệu.
- Map/reduce giúp việc lọc, tìm, tổng hợp dữ liệu dễ hơn bao giờ hết.
- Nhân bản/đồng bộ là sức mạnh đặc biệt của CouchDB mà hiếm database nào có.

4.1.3 Mô hình dữ liệu

- Database là một cấu trúc dữ liệu lớn của CouchDB.
- Mỗi Database là một danh sách các document độc lập.
- Document bao gồm dữ liệu người dùng thao tác lẫn thông tin về phiên bản của dữ liệu để tiện việc merge dữ liệu.
- CouchDB sử dụng cơ chế phiên bản hoá dữ liệu để tránh tình trạng khoá dữ liệu khi đang ghi.

4.2 Các tính năng chính

4.2.1 Lưu trữ dạng document

CouchDB là một NoSQL database dạng document. Document là một đơn vị dữ liệu (giống như một object của Javascript), mỗi field có một tên riêng không trùng nhau, chứa các loại dữ liệu hay chữ,

số, Boolean, danh sách,... Không có bất kì giới hạn nào về dung lượng hay text, số field trong một document.

CouchDB cung cấp một RESTful API cho việc đọc và ghi (bao gồm: thêm, sửa, xóa) document.

Đây là một ví dụ về document:

```
{ "title": "Macbook", "price": 1500, "SKU": "abcd1234" }
```

4.2.2 Các thuộc tính ACID

Khi dữ liệu được ghi xuống ổ cứng thì nó sẽ không bị ghi đè. Bất kì thay đổi nào (thêm, sửa, xóa) đều theo chuẩn Atomic, có nghĩa là dữ liệu sẽ được lưu lại toàn diện hoặc không được lưu lại. Database không bao giờ thêm hay sửa một phần dữ liệu.

Hầu hết các cập nhật đều được serialized để đảm bảo tất cả người dùng có thể đọc document mà không bị chờ đợi hoặc gián đoạn.

4.2.3 Khả năng nén (compaction)

Nén là một hành động giúp giải phóng dung lượng ổ cứng được sử dụng bằng cách xóa đi các dữ liệu không còn được sử dụng. Khi tiến hành nén dữ liệu ở một file thì một file mới với định dạng .compaction sẽ được tạo ra và dữ liệu sẽ được sao chép vào file mới này. Khi quá trình copy hoàn thành thì file cũ sẽ được xóa bỏ.

Database vẫn online trong quá trình nén và các thao tác thay đổi / đọc dữ liệu vẫn diễn ra bình thường.

4.2.4 Views

Dữ liệu trong CouchDB được lưu trữ trong các document. Bạn có thể tưởng tượng như một database là một table và một document là một row. Khi chúng ta muốn trình bày dữ liệu bằng nhiều góc khác nhau thì chúng ta cần một phương pháp để filter, tổ chức để hiển thị kết quả cuối cùng.

Để giải quyết vấn đề này, CouchDB sử dụng mô hình View. View là một phương pháp tổng hợp dữ liệu trong các document ở một database. Các View được build động và không ảnh hưởng đến dữ liệu đã ghi ở document nên chúng ta có thể có bao nhiêu View tùy ý và tùy vào nhu cầu trình bày dữ liệu.

4.3 Tổng quan về SimpleDB

4.3.1 SimpleDB là gì ?

- SimpleDB (hay Amazon SimpleDB) là một kho lưu trữ dữ liệu noSQL giảm thiểu về công việc quản lý dữ liệu.
- Các nhà phát triển chỉ lưu trữ và truy vấn dữ liệu thông qua các yêu cầu Web Services và Amazon SimpleDB sẽ thực hiện việc đó.

- Với Amazon SimpleDB, bạn có thể tập trung phát triển ứng dụng mà không cần phải quan tâm về vấn đề kết cấu, duy trì phần mềm, quản lý schema và index.

4.4 Tại sao chúng ta lại cần SimpleDB ?

4.4.1 Lợi ích

Dịch vụ cho phép bạn tập trung tất cả vào việc bổ sung giá trị cho phát triển phần mềm, hơn là việc bạn khó khăn và tốn thời gian vào việc quản lý database. Amazon SimpleDB quản lý một cách tự động cơ sở hạ tầng được cung cấp, duy trì phần cứng và phần mềm, tái tạo và index của dữ liệu.

Nếu vấn đề nghiệp vụ của bạn bị thay đổi hoặc cải tạo ứng dụng, bạn có thể dễ dàng đưa sự thay đổi vào trong Amazon SimpleDB mà không cần phải quan tâm về việc làm "vỡ" kết cấu schema - đơn giản chỉ cần thêm vào thuộc tính khác vào trong tập dữ liệu Amazon SimpleDB của bạn khi cần đến.

Amazon SimpleDB cung cấp tổ chức truy cập để lưu trữ và hàm truy vấn một cách truyền thống được sử dụng hoàn toàn bằng nhóm quan hệ dữ liệu (relational database cluster). Dịch vụ cho phép bạn có thể nhanh chóng thêm dữ liệu và dễ dàng truy vấn, sửa dữ liệu thông qua lệnh gọi API dễ dàng.

Amazon SimpleDB cung cấp một điểm cuối https để chắc chắn việc bảo mật, thông báo được mã hoá giữa ứng dụng của bạn hoặc khách hàng hoặc domain của bạn. Thêm vào đó, thông qua việc thông nhất với AWS Identity và quản lý truy cập, bạn có thể thiết lập user hoặc điều khiển group-level truy cập để đặc tả miền SimpleDB và toán tử.

4.5 Một số chức năng Use Cases

4.5.1 Login

- Monitoring or tracking
- Metering
- Trend of business analysis
- Auditing
- Archival or regulation compliance

4.5.2 Online games

- User scores and achievements
- User settings or preferences
- Information about player's items or user-generated content
- Game session state

4.5.3 Indexing Amazon S3 Object Metadata

Nhiều lập trình viên dùng Amazon SimpleDB kết hợp với Amazon Simple Storage Service (Amazon S3). Amazon SimpleDB có thể được dùng để lưu trữ con trỏ tới nơi đối tượng Amazon S3 và thông tin chi tiết về đối tượng đó (metadata), bằng cách bổ sung Amazon S3 với nhiều hàm truy vấn của database. Đối với lập trình viên đang lưu trữ số lượng lớn đối tượng trong Amazon S3, Amazon SimpleDB sẽ linh động, biến đổi, không tốn nhiều cách để lưu trữ đối tượng metadata trong việc "offloading" trên việc quản lý có liên quan đến việc vận hành cơ sở dữ liệu. Một số câu lệnh ví dụ phổ biến của đối tượng metadata có thể dễ dàng lưu trữ, indexed, và truy vấn trong Amazon SimpleDB bao gồm:

- Data type or format (image, video, document)
- User associations or access designations
- Dates the object was created, accessed, or modified
- Name or location of related objects
- User ratings and comments
- Subject or category tags
- Geolocation tag

4.6 So sánh

Name	Amazon SimpleDB	CouchDB	MongoDB
Mô tả	Dịch vụ dữ liệu đơn giản tạo bởi Amazone, với dữ liệu được lưu trên Amazon Cloud .	Là một native JSON - document, sự rộng rãi từ phân phối tổng thể nhóm server cho đến điện thoại.	Một trong những lưu trữ document nổi tiếng
Primary database model	Key-value store	Document store	Document store
Secondary database model	Key-value store
Cloud-based	yes	no	no
User concepts	Access rights for users and roles can be defined via the AWS Identity and Access Management	Access right for users can be defined per database	Access rights for user and roles
Triggers	no	yes	no
Server-side scripts	no	View functions in Javascript	Javascript
Partitioning methods	none	Sharding	Sharding