**BỘ MÔN HỆ THỐNG THÔNG TIN – KHOA CÔNG NGHỆ THÔNG TIN**

Logo

Description automatically generated**ĐẠI HỌC KHOA HỌC TỰ NHIÊN THÀNH PHỐ HỒ CHÍ MINH, ĐẠI HỌC QUỐC GIA TP HCM**

Sinh viên thực hiện:

|  |  |
| --- | --- |
| 19120464 | Phạm Ngọc Cường |
| 19120583 | Lê Thái Bình Minh |
| 19120590 | Huỳnh Thanh Mỹ |
| 19120529 | Nguyễn Phước Huy |

GV phụ trách: cô Nguyễn Trần Minh Thư

  QUẢN TRỊ CƠ SỞ DỮ LIỆU HIỆN ĐẠI

HỌC KỲ II – NĂM HỌC 2022 - 2023

**qUẢN TRỊ CSDL HIỆN ĐẠI**

**BẢNG THÔNG TIN CHI TIẾT NHÓM**

|  |  |  |
| --- | --- | --- |
| **Mã nhóm:** | 05 | |
| **Số lượng:** | **4** | |
| **MSSV** | **Họ tên** | **Email** | |
| 19120464 | Phạm Ngọc Cường | 19120464@student.hcmus.edu.vn | |
| 19120583 | Lê Thái Bình Minh | 19120583@student.hcmus.edu.vn | |
| 19120590 | Huỳnh Thanh Mỹ | 19120590@student.hcmus.edu.vn | |
| 19120529 | Nguyễn Phước Huy | 19120529@student.hcmus.edu.vn | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Bảng phân công & đánh giá hoàn thành công việc** | | | | |
| **Công việc thực hiện** | **Người thực hiện** | **Mức độ hoàn thành** | **Đánh giá của nhóm** |
|  | 19120464 - Phạm Ngọc Cường | 100% | 10/10 |
|  | 19120590 – Huỳnh Thanh Mỹ | 100% | 10/10 |
|  | 19120529 - Nguyễn Phước Huy | 100% | 10/10 |
|  | 19120583- Lê Thái Bình Minh | 100% | 10/10 |

Mục Lục

[**Kết quả thực hiện** 3](#_Toc138351110)

[I. So sánh Redis, MongoDB, Cassandra và Neo4J 3](#_Toc138351111)

[1. Tổng quan 3](#_Toc138351112)

[2. Chi tiết 0](#_Toc138351113)

[II. Mô tả lại các quy trình nghiệp vụ đặt khách sạn của hệ thống Agoda 0](#_Toc138351114)

[III. Phân tích các yêu cầu lưu trữ, khả năng mở rộng và hiệu suất truy xuất khi sử dụng từng loại NoSQL cho hệ thống đặt phòng khách sạn Agoda 0](#_Toc138351115)

[IV. Lựa chọn và thiết kế dữ liệu phù hợp với các quy trình đã phân tích 0](#_Toc138351116)

[1. Bảng thiết kế cơ sở dữ liệu 0](#_Toc138351117)

[2. Đề xuất về cải thiện hiệu quả truy vấn dựa trên thiết kế đề xuất 0](#_Toc138351118)

[3. Source code 0](#_Toc138351119)

# **Kết quả thực hiện**

# So sánh Redis, MongoDB, Cassandra và Neo4J

## Tổng quan

Redis, Mongo, Cassandra và Neo4j đều là 4 HQT CSDL nổi tiếng hiện nay, sau đây là những ưu điểm và khuyết điểm nổi bật của chúng:

|  | **MongoDB** | **Redis** | **Cassandra** | **Neo4J** |
| --- | --- | --- | --- | --- |
| Ưu điểm | - Tích hợp tốt với các ứng dụng web và mobile. | - Hiệu suất cao trong việc lưu trữ và truy xuất dữ liệu. | - Khả năng mở rộng tuyến tính với số lượng node. | - Cấu trúc dữ liệu đồ thị mạnh mẽ cho việc tìm kiếm và phân tích dữ liệu liên quan đến quan hệ giữa các đối tượng. |
| - Hỗ trợ tính năng sharding cho khả năng mở rộng dữ liệu. | - Hỗ trợ nhiều kiểu dữ liệu và tính năng caching. | - Tính khả dụng cao, có khả năng xử lý các vấn đề về đồng bộ hóa dữ liệu. | - Có tính năng tìm kiếm toàn văn và định tuyến động. |
| - Hỗ trợ các tính năng phức tạp như MapReduce và aggregation. | - Dễ dàng triển khai và sử dụng. | - Có khả năng xử lý các bản ghi hàng tỷ. | - Có khả năng tìm kiếm và truy vấn dữ liệu đồ thị rất nhanh. |
| - Hỗ trợ các tính năng ACID cho các giao dịch trong cùng một document. | - Hỗ trợ tính năng publish/subscribe cho việc xử lý các thông báo realtime. | - Hỗ trợ các tính năng ACID và khả năng xử lý các nút bị lỗi. | - Có khả năng xử lý các truy vấn phức tạp với dữ liệu đồ thị. |
| - Cung cấp tính năng tìm kiếm toàn văn và tìm kiếm theo truy vấn. | - Có khả năng xử lý nhiều truy vấn cùng lúc. | - Cung cấp tính năng tìm kiếm toàn văn và tìm kiếm theo truy vấn. | - Có khả năng tối ưu hóa truy vấn bằng cách sử dụng các chỉ mục đồ thị. |
| Nhược điểm | - Không hỗ trợ transaction qua các document khác nhau. | - Không hỗ trợ tính năng query phức tạp. | - Không hỗ trợ các tính năng phức tạp như MapReduce và aggregation. | - Không phù hợp cho các ứng dụng có dữ liệu lớn và cần khả năng mở rộng tuyến tính. |
| - Có thể dễ bị lỗi đồng bộ hóa khi có nhiều phiên bản của cùng một document. | - Có thể dễ bị mất dữ liệu trong trường hợp xảy ra sự cố. | - Khó khăn trong việc triển khai các hệ thống phân tán. | - Không hỗ trợ tính năng ACID. |
| - Có thể dễ bị tấn công và xâm nhập dữ liệu khi không được cấu hình đúng cách. | - Không phù hợp cho việc lưu trữ các dữ liệu quan trọng. | - Khôngcó tính năng tìm kiếm toàn văn mạnh mẽ và định tuyến động. | - Cần phải có kiến thức về lập trình đồ thị để có thể sử dụng hiệu quả. |

## Chi tiết

Sau khi tìm hiểu, nhóm đã lập bảng chi tiết để so sánh 4 CSDL như sau:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tiêu chí** | **Redis** | **MongoDB** | **Cassandra** | **Neo4j** |
| Kiểu dữ liệu | Nhiều kiểu dữ liệu có kích thước lên đến 512MB | Document JSON | Hỗ trợ nhiều kiểu dữ liệu khác nhau | Hỗ trợ nhiều kiểu dữ liệu |
| Khả năng tương thích | Cao | Cao | Cao | Cao(ngôn ngữ truy vấn Cypher) |
| Hiệu suất | Cao | Cao | Cao | Cao |
| Cộng đồng | Lớn | Lớn | Lớn | Đông đảo |
| Tính nhất quán | Cao | Cao | Cao | Cao |
| Tính bảo mật | Tính bảo mật không cao | Cao | Cao | Cao |
| Tính sẵn sàng | Cao | Cao | Cao | Cao |
| Tính mở rộng | Chiều ngang và dọc | Theo chiều ngang và dọc | cao | Cao |
| Chi phí | Miễn phí | Miễn phí | Miễn phí | Có 2 phiên bản miễn phí và trả phí |

# Mô tả lại các quy trình nghiệp vụ đặt khách sạn của hệ thống Agoda

Hệ thống Agoda là một trang web đặt phòng khách sạn trực tuyến, cho phép người dùng đặt phòng và thanh toán trực tuyến. Các quy trình nghiệp vụ đặt khách sạn của hệ thống Agoda có thể được mô tả như sau:

* Tìm kiếm khách sạn: Người dùng truy cập vào trang web Agoda và tìm kiếm khách sạn bằng cách nhập địa điểm, ngày nhận phòng và ngày trả phòng.
* Lọc kết quả: Hệ thống Agoda sẽ hiển thị kết quả tìm kiếm và cho phép người dùng lọc kết quả bằng nhiều tiêu chí khác nhau như giá, đánh giá của khách hàng, tiện nghi và vị trí.
* Chọn phòng: Người dùng chọn phòng thích hợp và xác nhận thông tin đặt phòng như số lượng phòng, số lượng khách, thông tin liên lạc và thời gian nhận phòng và trả phòng.
* Thanh toán: Sau khi xác nhận thông tin đặt phòng, người dùng thanh toán bằng các phương thức thanh toán trực tuyến như ví điện tử, thẻ tín dụng hoặc chuyển khoản ngân hàng.
* Xác nhận đặt phòng: Hệ thống Agoda gửi email xác nhận đặt phòng cho người dùng, bao gồm thông tin chi tiết về đặt phòng và số tiền đã thanh toán.
* Nhận phòng: Khi đến khách sạn, người dùng cần xuất trình email xác nhận đặt phòng và giấy tờ tùy thân để nhận phòng.
* Trả phòng: Khi hết thời gian đặt phòng, người dùng trả phòng và thanh toán các dịch vụ sử dụng thêm nếu có.
* Đánh giá: Sau khi trải nghiệm dịch vụ, người dùng có thể đánh giá khách sạn và chia sẻ trải nghiệm của mình trên trang web Agoda, giúp người dùng khác có thêm thông tin khi đặt phòng.

Trên đây là các quy trình nghiệp vụ đặt khách sạn của hệ thống Agoda. Quy trình này đảm bảo tính chính xác và đáng tin cậy của thông tin đặt phòng, giúp người dùng tiết kiệm thời gian và nâng cao trải nghiệm khi sử dụng dịch vụ.

## Mô hình use-case nghiệp vụ

A diagram of a hotel

Description automatically generated with low confidence

* + - * Đặc tả:

|  |  |
| --- | --- |
| Tên Use Case | Search for hotel rooms |
| Mô tả | * UC bắt đầu khi có khách hàng truy cập sử dụng dịch vụ tìm khách sạn. |
| Dòng cơ bản | 1. Khách hàng chọn địa điểm muốn tìm khách sạn, chọn ngày nhận phòng, ngày trả phòng và số lượng phòng cần đặt. 2. Hệ thống ghi nhận thông tin và bắt đầu tìm kiếm khách sạn theo thông tin đã nhập. 3. Hệ thống hiển thị các khách sạn thoả mãn yêu cầu của khách hàng. |
| Dòng thay thế | * A1: Tại bước 1, nếu thông tin khách hàng nhập chưa đầy đủ, hệ thống hiển thị thông báo yêu cầu nhập lại thông tin. * A3: Tại bước 3, nếu hệ thống không tìm thấy bất cứ khách sạn nào thoả mãn, hiển thị thông báo không tìm thấy khách sạn. |

|  |  |
| --- | --- |
| Tên Use Case | Book hotel rooms |
| Mô tả | * UC bắt đầu khi có khách hàng thực hiện thành công UC Search for hotel rooms |
| Dòng cơ bản | 1. Khách hàng chọn khách sạn phù hợp với mong muốn trong danh sách khách sạn hệ thống hiển thị để xem thông tin chi tiết. 2. Hệ thống hiển thị thông tin chi tiết của khách sạn khách hàng đã chọn. 3. Khách hàng click “Đặt phòng” để hoàn tất. 4. Hệ thống ghi nhận thông tin và hiển thị thông báo đặt phòng thành công |

|  |  |
| --- | --- |
| Tên Use Case | Payment |
| Mô tả | * UC bắt đầu khi khách hàng thực hiện đặt phòng thành công |
| Dòng cơ bản | 1. Hệ thống hiển thị thông tin giá phòng và các phương thức thanh toán. 2. Khách hàng chọn phương thức thanh toán. 3. Sau khi thực hiện đầy đủ các bước thanh toán, khách hàng click “Thanh toán” để hoàn tất 4. Hệ thống ghi nhận thông tin và hiển thị thông báo Thanh toán thành công |
| Dòng thay thế | * A3: Tại bước 3, nếu thông tin khách hàng nhập chưa đầy đủ, hệ thống hiển thị thông báo yêu cầu nhập lại thông tin. * A4: Tại bước 4, nếu hệ thống không ghi nhận được thông tin, hiển thị thông báo lỗi. |

|  |  |
| --- | --- |
| Tên Use Case | Rate hotel & Share experiment |
| Mô tả | * UC bắt đầu khi khách hàng đã thanh toán và trải qua kỳ nghỉ tại khách sạn đó hoặc khi khách hàng bắt đầu xem thông tin chi tiết của từng khách sạn |
| Dòng cơ bản | 1. Khách hàng chọn khách sạn muốn xem đánh giá. 2. Hệ thống hiển thị thông tin các đánh giá thuộc về khách sạn đã chọn. 3. Khách hàng nhập thông tin đánh giá mong muốn của bản thân và click Đánh giá để hoàn tất. 4. Hệ thống ghi nhận thông tin và hiển thị thông báo Đánh giá thành công |
| Dòng thay thế | * A3: Tại bước 3, nếu khách hàng không có nhu cầu đánh giá. Bỏ qua các bước còn lại. * A4: Tại bước 4, nếu hệ thống không ghi nhận được thông tin, hiển thị thông báo lỗi. |

* Sơ đồ hoạt động:

A picture containing text, screenshot, diagram, parallel

Description automatically generated

* asd

## Đặc tả

Minh hoạ:

A picture containing text, screenshot, font, line

Description automatically generated

# Phân tích các yêu cầu lưu trữ, khả năng mở rộng và hiệu suất truy xuất khi sử dụng từng loại NoSQL cho hệ thống đặt phòng khách sạn Agoda

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Yêu cầu lưu trữ** | **Khả năng mở rộng** | **Hiệu suất truy xuất** | **Ứng dụng** |
| **Redis** | Redis là một cơ sở dữ liệu key-value, phù hợp cho các ứng dụng yêu cầu lưu trữ dữ liệu nhanh và đơn giản. Redis hỗ trợ lưu trữ dữ liệu dạng key-value, list, set, sorted set và hash. | Redis hỗ trợ horizontal scaling với tính năng Redis Cluster, cho phép phân tán dữ liệu trên nhiều node để tăng khả năng mở rộng | Redis có hiệu suất truy xuất nhanh, hỗ trợ caching dữ liệu trong bộ nhớ và có khả năng xử lý hàng triệu truy vấn mỗi giây. | Khi khách hàng tìm kiếm phòng khách sạn trên Agoda, hệ thống sử dụng Redis để lưu trữ cache kết quả tìm kiếm của khách hàng. Khi khách hàng tiếp tục tìm kiếm hoặc truy cập lại trang web Agoda, hệ thống sẽ sử dụng cache được lưu trữ trên Redis thay vì thực hiện lại truy vấn tìm kiếm, giúp tăng tốc độ truy xuất và giảm tải cho cơ sở dữ liệu chính. |
| **MongoDB** | MongoDB là một cơ sở dữ liệu phi cấu trúc (document-oriented), phù hợp cho các ứng dụng yêu cầu lưu trữ dữ liệu phức tạp và đa dạng. MongoDB hỗ trợ lưu trữ dữ liệu dạng BSON (Binary JSON) | MongoDB có tính mở rộng ngang (horizontal scaling) tốt, cho phép phân tán dữ liệu trên nhiều node để tăng khả năng mở rộng | MongoDB có hiệu suất truy xuất nhanh và có thể xử lý hàng triệu truy vấn mỗi giây. | Khi khách hàng đặt phòng khách sạn trên Agoda, hệ thống sử dụng MongoDB để lưu trữ các thông tin về thanh toán và các dịch vụ sử dụng thêm của khách hàng sau khi kết thúc thời gian đặt phòng. MongoDB cho phép lưu trữ dữ liệu phi cấu trúc (document-oriented) và tính mở rộng ngang (horizontal scaling) tốt, giúp hệ thống xử lý các truy vấn đọc/ghi trên nhiều node đồng thời và đảm bảo tính sẵn sàng và độ tin cậy của hệ thống. |
| **Cassandra** | Cassandra là một cơ sở dữ liệu phân tán, phù hợp cho các ứng dụng yêu cầu lưu trữ dữ liệu lớn và có tính mở rộng cao. Cassandra hỗ trợ lưu trữ dữ liệu dạng column family. | Cassandra có tính mở rộng ngang (horizontal scaling) rất tốt, cho phép phân tán dữ liệu trên nhiều node và tự động điều chỉnh khi có thêm node mới hoặc node cũ bị lỗi. | Cassandra có khả năng xử lý hàng triệu truy vấn mỗi giây và có thể xử lý các truy vấn đọc/ghi trên nhiều node đồng thời | Khi khách hàng đặt phòng khách sạn trên Agoda, hệ thống sử dụng Cassandra để lưu trữ thông tin đặt phòng và thông tin liên quan như thông tin khách hàng, thông tin phòng, thời gian đến/đi. Cassandra cho phép lưu trữ dữ liệu phân tán trên nhiều node, đảm bảo tính sẵn sàng và độ tin cậy cao của hệ thống. Hệ thống cũng sử dụng Cassandra để lưu trữ lịch sử đặt phòng của khách hàng và phân tích dữ liệu đặt phòng để cải thiện trải nghiệm khách hàng. |
| **Neo4j** | Neo4j là một cơ sở dữ liệu đồ thị, phù hợp cho các ứng dụng yêu cầu truy vấn dữ liệu phức tạp và phân tích mối quan hệ giữa các đối tượng. | Neo4j có tính mở rộng ngang (horizontal scaling) hạn chế, vì vậy không phù hợp cho các ứng dụng yêu cầu mở rộng quy mô lớn. | Neo4j có hiệu suất truy xuất tốt cho các truy vấn đồ thị phức tạp, nhưng không phù hợp cho các ứng dụng yêu cầu xử lý lượng dữ liệu lớn. | Khi khách hàng tìm kiếm phòng khách sạn trên Agoda, hệ thống sử dụng Neo4j để xây dựng đồ thị mối quan hệ giữa các khách sạn, các phòng và các tiện nghi để giúp khách hàng tìm kiếm và so sánh thông tin phòng khách sạn dễ dàng hơn. Neo4j cũng được sử dụng để lưu trữ các đánh giá và phản hồi của khách hàng về các khách sạn, giúp cải thiện trải nghiệm khách hàng và tăng tính tương tác của hệ thống. |

# Lựa chọn và thiết kế dữ liệu phù hợp với các quy trình đã phân tích

## Bảng thiết kế cơ sở dữ liệu

Asdasd

## Đề xuất về cải thiện hiệu quả truy vấn dựa trên thiết kế đề xuất

Asdasd

## Source code

Asdasd