**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

KHOA HỆ THỐNG THÔNG TIN

🙤🙧🟍🙥🙦



**BÀI TẬP TIỂU LUẬN**

**CƠ SỞ DỮ LIỆU PHÂN TÁN**

**HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU NOSQL**

**COUCHBASE SERVER**

|  |  |  |
| --- | --- | --- |
| Sinh viên thực hiện: | | |
| STT | Họ tên | MSSV |
| 1 | Nguyễn Văn Thịnh | 19522286 |
| 2 | Huỳnh Đắc Nhât Tân | 19522171 |
| 3 | Lê Văn Thanh | 16521116 |
| 4 | Nguyễn Thành Công | 15520070 |

**TP. HỒ CHÍ MINH – 12/2021**

**Mục lục**

[CHƯƠNG 1. GIỚI THIỆU 6](#_Toc91931035)

[1.1 Couchbase Server 6](#_Toc91931036)

[1.2 Lịch sử phát triển 6](#_Toc91931037)

[1.3 Kiến trúc 7](#_Toc91931038)

[1.3.1 Quản lí Cụm 7](#_Toc91931039)

[1.3.2 Quản lý dữ liệu 7](#_Toc91931040)

[1.4 Mô hình lưu trữ 8](#_Toc91931041)

[1.5 Ngôn ngữ thao tác với dữ liệu: N1QL 9](#_Toc91931042)

[1.6 Một số khái niệm cơ bản 12](#_Toc91931043)

[1.6.1 Dashboard: 12](#_Toc91931044)

[1.6.2 Server: 13](#_Toc91931045)

[1.6.3 Bucket: 14](#_Toc91931046)

[1.6.4 XDCR 15](#_Toc91931047)

[1.6.5 Security: 15](#_Toc91931048)

[1.6.6 Settings: 16](#_Toc91931049)

[1.6.7 Logs: 16](#_Toc91931050)

[1.6.8 Documents: 18](#_Toc91931051)

[1.6.9 Query: 19](#_Toc91931052)

[1.6.10 Indexes: 19](#_Toc91931053)

[1.6.11 Search: 20](#_Toc91931054)

[1.6.12 Analytics: 20](#_Toc91931055)

[1.6.13 Eventing: 21](#_Toc91931056)

[1.6.14 Views 21](#_Toc91931057)

[CHƯƠNG 2. CÀI ĐẶT VÀ TRUY VẤN DỮ LIỆU PHÂN TÁN 22](#_Toc91931058)

[2.1 Cài đặt Couchbase server 22](#_Toc91931059)

[2.1.1 Cài đặt Couchbase Server 22](#_Toc91931060)

[2.1.2 Cấu hình Couchbase Server 27](#_Toc91931061)

[2.1.3 Kết nối hai couchbase server 31](#_Toc91931062)

[2.2 Truy vấn dữ liệu phân tán: 33](#_Toc91931063)

[2.2.1 Remote link: 33](#_Toc91931064)

[2.2.2 Thực nghiệm: 33](#_Toc91931065)

[CHƯƠNG 3. THAO TÁC DỮ LIỆU PHÂN TÁN 40](#_Toc91931066)

[3.1 Kỹ thuật nhân bản (Replication) 40](#_Toc91931067)

[3.2 Thêm, xóa, sửa dữ liệu phân tán 41](#_Toc91931068)

[3.2.1 Tạo XDCR để tiến hành thao tác trên máy 2 41](#_Toc91931069)

[3.2.2 Các thao tác lên dữ liệu của bucket máy 2 44](#_Toc91931070)

[TÀI LIỆU THAM KHẢO 47](#_Toc91931071)

*Mục lục hình ảnh*

[Hình 1‑1 Lịch sử phát triển couchbase 7](#_Toc91931072)

[Hình 1‑2 Couchbase dashboard 12](#_Toc91931073)

[Hình 1‑3 Coucbase Server 13](#_Toc91931074)

[Hình 1‑4 Coucbase Buckets 14](#_Toc91931075)

[Hình 1‑5 Couchbase XDCR 15](#_Toc91931076)

[Hình 1‑6 Couchbase Security 16](#_Toc91931077)

[Hình 1‑7 Couchbase Settings 16](#_Toc91931078)

[Hình 1‑8 Couchbase Logs 16](#_Toc91931079)

[Hình 1‑9 Couchbase Documents 18](#_Toc91931080)

[Hình 1‑10 Couchbase Query 19](#_Toc91931081)

[Hình 1‑11 Couchbase Indexes 19](#_Toc91931082)

[Hình 1‑12 Couchbase Search 20](#_Toc91931083)

[Hình 1‑13 Couchbase Analytics 20](#_Toc91931084)

[Hình 1‑14 Couchbase Eventing 21](#_Toc91931085)

[Hình 1‑15 Couchbase Views 21](#_Toc91931086)

[Hình 2‑1 Trang chủ cài đặt Couchbase server 22](#_Toc91931087)

[Hình 2‑2 Giao diện couchbase-server-enterprise\_7.0.3-windows\_amd64.msi 23](#_Toc91931088)

[Hình 2‑3 Các điều khoản và dịch vụ của Couchbase 24](#_Toc91931089)

[Hình 2‑4 Chọn ổ đĩa chứa Couchbase Server 25](#_Toc91931090)

[Hình 2‑5 Couchbase IP vesion selection 26](#_Toc91931091)

[Hình 2‑6 Giao diện Couchbase trên trình duyệt localhost:8091 27](#_Toc91931092)

[Hình 2‑7 Cấu hình các thông tin cluster 28](#_Toc91931093)

[Hình 2‑8 Cluster username và password 29](#_Toc91931094)

[Hình 2‑9 Couchbase Cluster điều khoản và dịch vụ 30](#_Toc91931095)

[Hình 2‑10 Trang Dashboard sau khi tạo xong cluster 30](#_Toc91931096)

[Hình 2‑11 Kết nối mạng ảo thông qua RadminVPN 31](#_Toc91931097)

[Hình 2‑12 Mục XDCR và các chức năng bên trong 31](#_Toc91931098)

[Hình 2‑13 Cửa sổ bật lên khi ấn Add Remote ở Mục XDCRb 32](#_Toc91931099)

[Hình 2‑14 Dữ liệu của bảng BANHMY ở máy 1 33](#_Toc91931100)

[Hình 2‑15 Chọn mục Map From Data Service ở mục Analytics 33](#_Toc91931101)

[Hình 2‑16 Cấu hình Collection bên trong Map From Data Service 34](#_Toc91931102)

[Hình 2‑17 Mục analytics Scopes, Links, & Collections sau khi map 34](#_Toc91931103)

[Hình 2‑18 Truy vấn dữ liệu máy 1 đã map lại dữ liệu 35](#_Toc91931104)

[Hình 2‑19 Biểu tượng Remote link 35](#_Toc91931105)

[Hình 2‑20 Cửa sổ tạo remote link 36](#_Toc91931106)

[Hình 2‑21 Mục Analytics Scopes, Links, & Collections sau khi tạo remote link 37](#_Toc91931107)

[Hình 2‑22 Mục Analytics Scopes, Links, & Collections sau khi tạo dataset 38](#_Toc91931108)

[Hình 2‑23 Biểu tượng connect 38](#_Toc91931109)

[Hình 2‑24 Truy vấn dữ liệu máy 2 39](#_Toc91931110)

[Hình 2‑25 Kết quả của câu truy vấn đơn giản 39](#_Toc91931111)

[Hình 3‑1 Mục Replication bên trong mục XDCR 40](#_Toc91931112)

[Hình 3‑2 Các setting để nhân bản dữ liệu 41](#_Toc91931113)

[Hình 3‑3 Tạo bucket mới dùng để thực hiện các thao tác lên dữ liệu 41](#_Toc91931114)

[Hình 3‑4 Mục Add replication trong mục XDCR 42](#_Toc91931115)

[Hình 3‑5 Thiết lập replication trên máy 1 43](#_Toc91931116)

[Hình 3‑6 Thiết lập replication trên máy 2 43](#_Toc91931117)

[Hình 3‑7 Dữ liệu BANHMY1 ở máy 1 44](#_Toc91931118)

[Hình 3‑8 Dữ liệu BANHMY2 ở máy 2 44](#_Toc91931119)

[Hình 3‑9 Thêm dữ liệu vào bucket mới tạo ở máy 1 45](#_Toc91931120)

[Hình 3‑10 Kiểm tra dữ liệu ở bucket máy 2 46](#_Toc91931121)

[Hình 3‑11 Thực hiện các thao tác dữ liệu bằng các chức năng Document 46](#_Toc91931122)

# GIỚI THIỆU

## Couchbase Server

Couchbase Server, ban đầu được gọi là Membase, là gói phần mềm cơ sở dữ liệu định hướng tài liệu đa mô hình NoQuery mã nguồn mở, phân tán (được chia sẻ) tối ưu hóa cho các ứng dụng tương tác. Các ứng dụng này có thể phục vụ nhiều người dùng đồng thời bằng cách tạo, lưu trữ, truy xuất, tổng hợp, thao tác và trình bày dữ liệu. Để hỗ trợ các loại nhu cầu ứng dụng này, Couchbase Server được thiết kế để cung cấp quyền truy cập khóa-giá trị khóa hoặc tài liệu JSON dễ dàng với độ trễ thấp và thông lượng duy trì cao. Phiên bản ban đầu được gọi là Couchbase Lite, sau đó đã được bán trên thị trường với tên Couchbase Mobile kết hợp với các phần mềm khác.

## Lịch sử phát triển

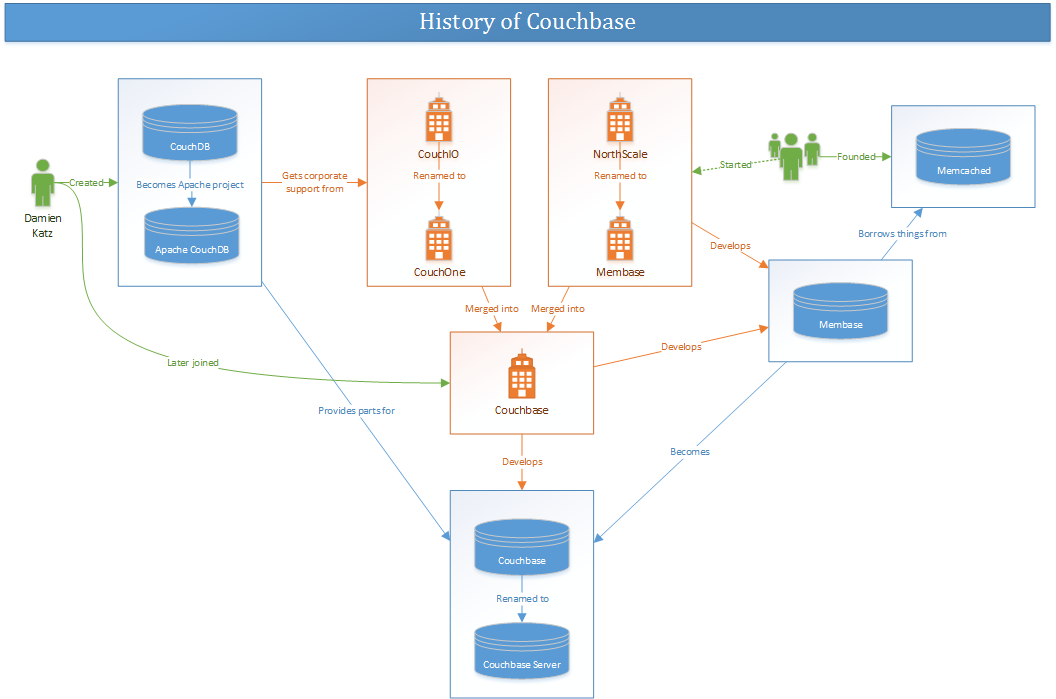
Couchbase được phát triển bởi một số nhà lãnh đạo của dự án memcached, nhóm người đã thành lập công ty NorthScale để phát triển một kho lưu trữ khóa với sự đơn giản, tốc độ và khả năng mở rộng của memcached, nhưng cũng có khả năng lưu trữ và truy vấn cơ sở dữ liệu. Mã nguồn membase ban đầu được đóng góp bởi NorthScale và nhà đồng tài trợ dự án Zynga và Naver Corporation (sau đó gọi là NHN) cho một dự án mới trên membase.org vào tháng 6 năm 2010.

Ngày 8 tháng 2 năm 2011, người sáng lập dự án Membase và Membase, Inc. đã công bố sáp nhập với CouchOne. Công ty được sáp nhập được gọi là Couchbase, Inc.

Tháng 1 năm 2012, Couchbase đã phát hành Couchbase Server 1.8.

Tháng 9 năm 2012, Orbitz cho biết họ đã thay đổi một số hệ thống của mình để sử dụng Couchbase.

Tháng 12 năm 2012, Couchbase Server 2.0 (được công bố vào tháng 7 năm 2011) đã được phát hành và bao gồm một kho tài liệu JSON mới.



Hình ‑ Lịch sử phát triển couchbase

## Kiến trúc

Mỗi nút Couchbase bao gồm : dịch vụ dữ liệu, dịch vụ chỉ mục, dịch vụ truy vấn và thành phần quản lý cụm. Bắt đầu với phiên bản 4.0, ba dịch vụ có thể được phân phối để chạy trên các nút riêng biệt của cụm nếu cần.

### Quản lí Cụm

Trình quản lý cụm giám sát cấu hình và hành vi của tất cả các máy chủ trong cụm Couchbase. Nó cấu hình và giám sát hành vi giữa các nút như quản lý các luồng nhân rộng và cân bằng lại các hoạt động. Nó cũng cung cấp các hàm tổng hợp và đồng thuận số liệu cho cụm và giao diện quản lý cụm RESTful. Trình quản lý cụm sử dụng ngôn ngữ lập trình Erlang và Nền tảng viễn thông mở.

Sao chép dữ liệu trong các nút của một cụm có thể được kiểm soát bằng nhiều tham số. Vào tháng 12 năm 2012 đã được thêm vào những tính năng để nhân rộng giữa các trung tâm dữ liệu khác nhau.

### Quản lý dữ liệu

Trình quản lý dữ liệu lưu trữ và truy xuất tài liệu để đáp ứng các hoạt động dữ liệu từ các ứng dụng. Nó ghi dữ liệu một cách không đồng bộ vào đĩa sau khi xác nhận với client. Trong phiên bản 1.7 trở lên, các ứng dụng có thể tùy chọn hoặc đảm bảo dữ liệu được ghi vào nhiều máy chủ hoặc vào đĩa trước khi xác nhận ghi vào client. Các tham số xác định thời gian gây ảnh hưởng khi dữ liệu được duy trì, và cách bộ nhớ tối đa và di chuyển từ bộ nhớ chính sang đĩa được xử lý. Nó hỗ trợ các bộ làm việc lớn hơn dung lượng bộ nhớ cho mỗi "nút" hoặc "nhóm". Các hệ thống bên ngoài có thể đăng ký các luồng dữ liệu được lọc, ví dụ như hỗ trợ lập chỉ mục tìm kiếm toàn văn bản, phân tích dữ liệu hoặc lưu trữ.

* Định dạng dữ liệu: Document là đơn vị thao tác dữ liệu cơ bản nhất trong Couchbase Server. Document được lưu trữ ở định dạng JSON không có lược đồ được xác định trước.
* Bộ đệm được quản lý đối tượng **:** Couchbase Server bao gồm bộ đệm được quản lý đối tượng đa luồng tích hợp, thực hiện các API tương thích được ghi nhớ như get, set, xóa, chắp thêm, thêm vào, v.v.
* Công cụ lưu trữ: Couchbase Server có thiết kế lưu trữ nối đuôi giúp xử lý được khi dữ liệu bị hư hỏng, OOM hoặc mất điện đột ngột. Dữ liệu được ghi vào tệp dữ liệu theo cách append-only (chỉ cho bổ sung thêm bản ghi mới sau bản ghi cũ), cho phép Couchbase thực hiện hầu hết việc ghi tuần tự để cập nhật và cung cấp các mẫu truy cập được tối ưu hóa cho I/O của đĩa.

## Mô hình lưu trữ

Khái niệm trung tâm của một kho lưu trữ tài liệu là khái niệm về "tài liệu" (document). Trong khi mỗi cơ sở dữ liệu hướng tài liệu thực hiện khác nhau về chi tiết của định nghĩa này, nói chung, tất cả chúng đều giả định rằng các tài liệu đóng gói và mã hóa dữ liệu (hoặc thông tin) trong một số định dạng hoặc mã hóa tiêu chuẩn. Mã hóa được sử dụng bao gồm XML, [YAML](https://vi.wikipedia.org/w/index.php?title=YAML&action=edit&redlink=1), và [JSON](https://vi.wikipedia.org/wiki/JavaScript_Object_Notation) cũng như các dạng nhị phân như [BSON](https://vi.wikipedia.org/w/index.php?title=BSON&action=edit&redlink=1). Các tài liệu được định địa chỉ trong cơ sở dữ liệu thông qua một từ khóa duy nhất đại diện cho tài liệu đó. Một trong những đặc điểm định nghĩa khác của một cơ sở dữ liệu hướng tài liệu là ngoài việc tra cứu từ khóa được thực hiện bởi một kho lưu trữ khóa-giá trị, cơ sở dữ liệu đó còn cung cấp một API hoặc ngôn ngữ truy vấn để lấy tài liệu dựa trên nội dung của chúng

Những triển khai khác nhau cung cấp nhiều cách khác nhau để tổ chức và / hoặc nhóm các tài liệu:

* Các bộ sưu tập
* Thẻ đánh dấu (tag)
* Siêu dữ liệu không nhìn thấy được
* Phân cấp thư mục

So với cơ sở dữ liệu quan hệ, ví dụ, các bộ sưu tập có thể được coi là tương tự như các bảng biểu và các tài liệu tương tự như các hồ sơ/bản ghi. Nhưng chúng là khác nhau: mỗi bản ghi trong một bảng có cùng một trình tự của các miền, trong khi các tài liệu trong bộ sưu tập có thể có các miền hoàn toàn khác nhau.

## Ngôn ngữ thao tác với dữ liệu: N1QL

N1QL (phát âm là niken), được sử dụng để thao tác dữ liệu JSON trong Couchbase, giống như SQL thao tác dữ liệu trong RDBMS; các câu lệnh SELECT, INSERT, UPDATE, DELETE, MERGE để thao tác trên dữ liệu JSON. Nó được công bố vào tháng 3 năm 2015.

Mô hình dữ liệu N1QL là dạng thông thường không phải đầu tiên (N1NF) với sự hỗ trợ cho các thuộc tính lồng nhau và chuẩn hóa hướng miền. Mô hình dữ liệu N1QL cũng là một siêu dữ liệu phù hợp và khái quát hóa của mô hình quan hệ.

* Một số câu lệnh N1QL:

|  |  |  |
| --- | --- | --- |
|  | **Cú pháp** | **Kết quả** |
| SELECT | SELECT \*  FROM tutorial  WHERE fname = 'Ian' | *{*  *"results": [*  *{*  *"tutorial": {*  *"type": "contact",*  *"title": "Mr.",*  *"fname": "Ian",*  *"lname": "Taylor",*  *"age": 56,*  *"email": "ian@gmail.com",*  *"children": [*  *{*  *"fname": "Abama",*  *"age": 17,*  *"gender": "m"*  *},*  *{*  *"fname": "Bebama",*  *"age": 21,*  *"gender": "m"*  *}*  *],*  *"hobbies": [*  *"golf",*  *"surfing"*  *],*  *"relation": "cousin"*  *}*  *}*  *]*  *}* |
| SELECT (WHERE) | SELECT fname, age  FROM tutorial  WHERE age > 30 | *{*  *"results": [*  *{*  *"age": 46,*  *"fname": "Dave"*  *},*  *{*  *"age": 46,*  *"fname": "Earl"*  *},*  *{*  *"age": 56,*  *"fname": "Ian"*  *},*  *{*  *"age": 40,*  *"fname": "Jane"*  *}*  *]}* |
| SELECT (KEYS) | SELECT fname, email  FROM tutorial  USE KEYS ["dave", "ian"] | *{*  *"results": [*  *{*  *"email": "dave@gmail.com",*  *"fname": "Dave"*  *},*  *{*  *"email": "ian@gmail.com",*  *"fname": "Ian"*  *}*  *]*  *}* |
| SELECT (GROUP HAVING) | SELECT relation, COUNT(\*) AS count  FROM tutorial  GROUP BY relation  HAVING COUNT(\*) > 1 | *{*  *"results": [*  *{*  *"count": 2,*  *"relation": "cousin"*  *},*  *{*  *"count": 2,*  *"relation": "friend"*  *}*  *]*  *}* |
|  | SELECT fname, age, ROUND(age/7) AS age\_dog\_years  FROM tutorial  WHERE fname = 'Dave' | *{*  *"results": [*  *{*  *"age": 46,*  *"age\_dog\_years": 7,*  *"fname": "Dave"*  *}*  *]*  *}* |
| UPDATE | UPDATE tutorial  USE KEYS "baldwin"  SET type = "actor" RETURNING tutorial.type |  |
| DELETE | DELETE FROM tutorial t  USE KEYS "baldwin" RETURNING t |  |
| INSERT | INSERT INTO `bucket` VALUES(KEY,VALUE)  (KEY,{“attr1”:”a”,”attr2”:”b”}) |  |

## Một số khái niệm cơ bản

### Dashboard:

Tổng thể về thông tin cluster hiện tại, bao gồm thông tin về các dịch vụ, các node của server, bộ nhớ tại thời điểm hiện tại

Graphical user interface, application

Description automatically generated

Hình ‑ Couchbase dashboard

### Server:

Quản lí thông tin về server hiện tại. Thêm node vào server hiện tại. Thông tin về các node trong server hiện tại.

Graphical user interface, text, application, email

Description automatically generated

Hình ‑ Coucbase Server

### Bucket:

Quản lí về các bucket hiện tại của Cluster

Tối đa có thể tạo được 30 bucket trên một cluster.

Có ba loại bucket chính:

* Couchbase buckets:

Couchbase buckets sẽ lưu dữ liệu trên cả ổ cứng và RAM. Việc này sẽ giúp sao chép dữ liệu diễn ra nhanh hơn sử dụng DCP và XDCR.

Khi couchbase buckets vượt quá RAM được chia, các dữ liệu khi thêm vào sẽ không được nhận, điều này có nghĩa dữ liệu được lưu trên RAM sẽ bị mất nhưng vẫn được lưu ở ổ cứng. Vì vậy nếu dữ liệu bị xóa ở RAM cần dùng tới thì vẫn tải được dữ liệu ở ổ cứng lên. Việc loại bỏ dữ liệu tràn bộ nhớ được diễn ra theo hai hướng:

* + Value-only: Đây là cách mà couchbase server thường thực hiện, sẽ chỉ xóa các key-value
  + Full: Tất cả các dữ liệu gồm key, key-value và cả dữ liệu dạng JSON đều bị xóa.
* Ephemeral buckets:

Tương tự như Couchbase bucket, tuy nhiên Ephemeral buckets không quá phụ thuộc vào ổ cứng. Tùy theo cách thiết lập ban đầu khi khởi tạo mà khi dữ liệu tràn RAM sẽ có những tình huống khác nhau xảy ra

* Memcached buckets:

Không sử dụng ổ cứng, tất cả các dữ liệu đều lưu trữ trên RAM. Khi dữ liệu vượt quá RAM được cấp, dữ liệu sẽ bị xóa.

Graphical user interface, text, application, email

Description automatically generated

Hình ‑ Coucbase Buckets

### XDCR

XDCR (viết tắt của Cross Data Center Replication) là mục dùng để quản lí nhân bản dữ liệu các bucket giữa những cluster với nhau. Những cluster này có thể ở những địa điểm khác nhau, dữ liệu được đặt tại các máy khác nhau từ đó có thể phòng tránh được sự mất mát dữ liệu khi dữ liệu ở một máy bị mất.

XDCR của bản Community Edition sẽ có một chút giới hạn về chức năng so với bản Enterprise Edition.

XDCR Replication cho phép nhân bản thông qua các nguồn dữ liệu và cluster đích đến theo những cách sau:

* Unidirectionally (nhân bản một chiều): dữ liệu nhân bản một chiều sẽ không liên thông giữa máy trạm và máy chủ, máy trạm không sở hữu dữ liệu của máy chủ, tuy nhiên vẫn có khả năng thực hiện các thao tác thêm, xóa dữ liệu mới được thêm, xóa tại máy chủ.
* Bidirectionally (nhân bản hai chiều): dữ liệu nhân bản hai chiều sẽ liên thông hai dữ liệu tại hai bucket với nhau, điều này có nghĩa hai dữ liệu ở hai máy sẽ là một. Nhân bản hai chiều sẽ giúp người dùng có thể thao tác lên dữ liệu tại hai máy nhanh hơn.

XDCR Replication còn cho phép nhân bản nhiều lần giữa các cluster trên các máy ở các địa điểm khác nhau. Chức năng nhân bản còn hỗ trợ thêm các tính năng filter dùng để tạo thủ công đường dẫn liên kết nhân bản; process để theo dõi quá trình nhân bản; Priority để quản lý các nhân bản khi có từ hai nhân bản trở lên; Security dùng để thiết lập bảo mật trên các nhân bản khi người dùng thao tác,…

Graphical user interface, application

Description automatically generated

Hình ‑ Couchbase XDCR

### Security:

Cài đặt các thông tin bảo mật cho Cluster. Quản lí tài khoản và phân quyền cho các tài khoản tại server

Graphical user interface, text, application, Word, email

Description automatically generated

Hình ‑ Couchbase Security

### Settings:

Các thông tin cài đặt của Cluster hiện tại.

Graphical user interface, application, Word

Description automatically generated

Hình ‑ Couchbase Settings

### Logs:

Lưu trữ các log tạo bucket, server, tạo các connection, query, index,… Tất cả các thao tác thực hiện trên couchbase server đều được lưu ở đây

Graphical user interface, text, application, email

Description automatically generated

Hình ‑ Couchbase Logs

### Documents:

Lưu giữ các dữ liệu từ các bucket có trên cluster, các thao tác thêm xóa sửa có thể thực hiện nhanh ở đây mà không cần phải thực hiện query.

Ngoài ra chúng ta có thể thêm một cơ sở dữ liệu dưới dạng JSON vào đây.

Graphical user interface, text, application, Word, email

Description automatically generated

Hình ‑ Couchbase Documents

### Query:

Thực hiện các truy vấn cũng như các thao tác thêm, xóa, sửa dữ liệu lên bucket.

Graphical user interface, application, Word

Description automatically generated

Hình ‑ Couchbase Query

### Indexes:

Trước khi thực hiện các truy vấn như select thì cần phải tạo một index để chỉ đích danh dữ liệu.

Cú pháp:

CREATE PRIMARY INDEX ON `table-name`;

Sao khi tạo một index từ couchbase Query, phần Indexes sẽ xuất hiện một index gồm khóa và định danh, từ đó chúng ta có thể thay đổi các thiết lập cũng như drop index.

Graphical user interface, text, application, email

Description automatically generated

Hình ‑ Couchbase Indexes

### Search:

Thêm và tìm kiếm các index khi lượng dữ liệu lớn và có quá nhiều index.

Ngoài ra cũng có khả năng gom nhóm các index thành các alias để dễ dàng quản lý.

Graphical user interface, application

Description automatically generated

Hình ‑ Couchbase Search

### Analytics:

Analytics service giúp phân tích dữ liệu từ các file JSON vào thời gian thực. Bằng cách tạo ra một bản sao lưu dữ liệu gốc, người dùng có thể thực hiện các truy vấn phức tạp với lượng dữ liệu lớn.

Sao khi tạo một bản sao lưu dữ liệu gốc, khi dữ liệu gốc có bất kì thay đổi nào thì dữ liệu được sao lưu sang Analytics service cũng sẽ thay đổi theo.

Vì dữ liệu trên Analytics service là dữ liệu sao lưu nên không hỗ trợ các lệnh thao tác trên dữ liệu gốc như insert, delete, update,…

Graphical user interface, text, application

Description automatically generated

Hình ‑ Couchbase Analytics

### Eventing:

Dùng để tạo các sự kiện thao tác lên cơ sở dữ liệu khi phát triển một ứng dụng sử dụng Couchbase server.

Graphical user interface, text, application, email

Description automatically generated

Hình ‑ Couchbase Eventing

### Views

Tương tự như các hệ quản trị cơ sở dữ liệu khác, couchbase server cũng có view để cho người khác được truy vấn trên một dữ liệu cụ thể. Các thao tác trên dữ liệu không thể được thực hiện trên view và cũng không thể tác động đến dữ liệu gốc.

Graphical user interface, text, application, email

Description automatically generated

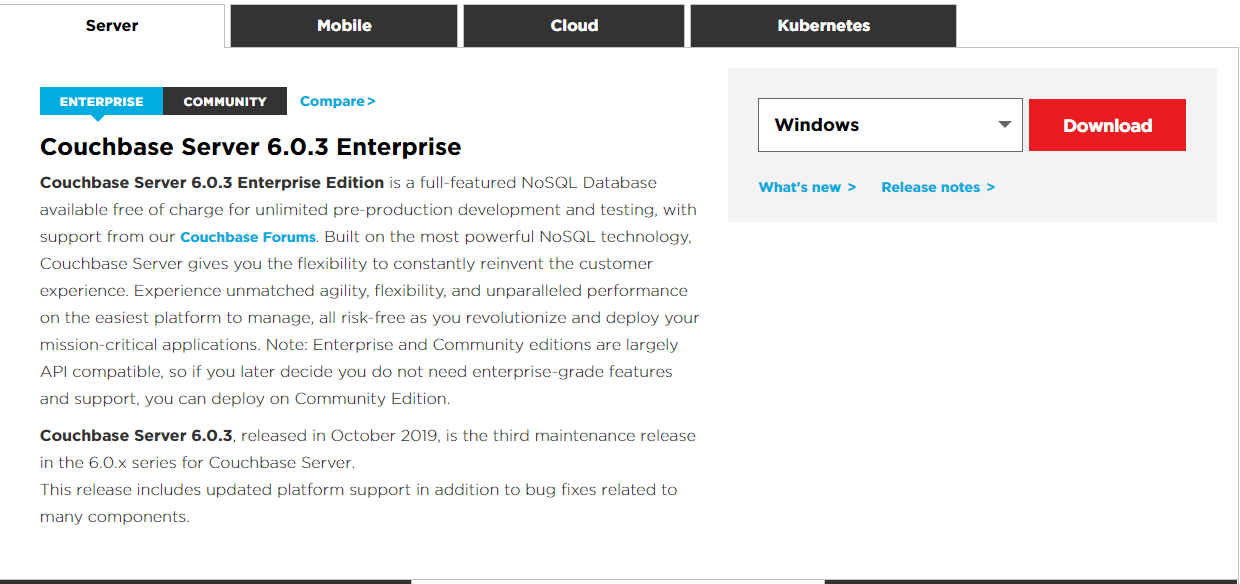
Hình ‑ Couchbase Views

# CÀI ĐẶT VÀ TRUY VẤN DỮ LIỆU PHÂN TÁN

## Cài đặt Couchbase server

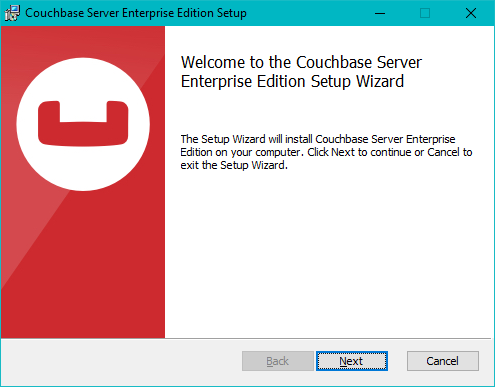
### Cài đặt Couchbase Server

Để tải Couchbase Server, truy cập trang <https://www.couchbase.com/downloads>. Tiểu luận sử dụng phiên bản Couchbase Server Enterprise 7.0.3.



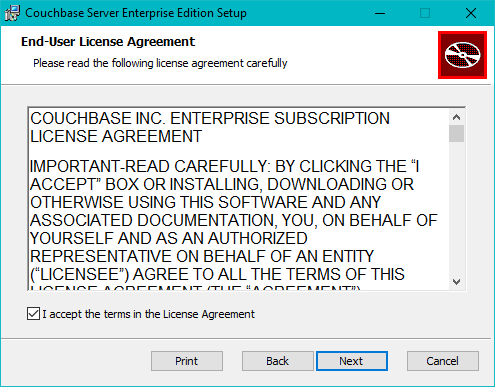
Hình ‑ Trang chủ cài đặt Couchbase server

Khi quá trình tải hoàn tất, khởi chạy file couchbase-server-enterprise\_7.0.3-windows\_amd64.msi để tiến hành cài đặt.



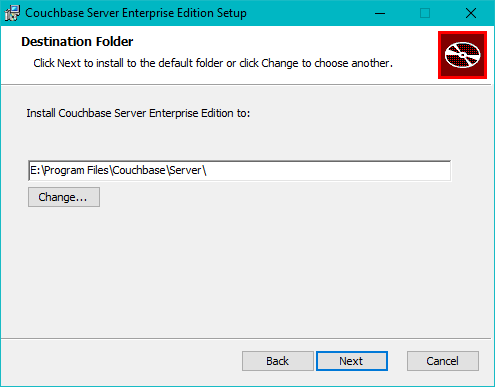
Hình ‑ Giao diện couchbase-server-enterprise\_7.0.3-windows\_amd64.msi

Đồng ý với các điều khoản và dịch vụ của Couchbase Server



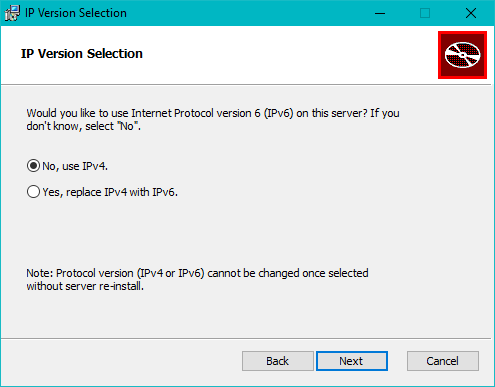
Hình ‑ Các điều khoản và dịch vụ của Couchbase

Chọn vị trí ổ đĩa để lưu Couchbase Server



Hình ‑ Chọn ổ đĩa chứa Couchbase Server

Chọn IP version



Hình ‑ Couchbase IP vesion selection

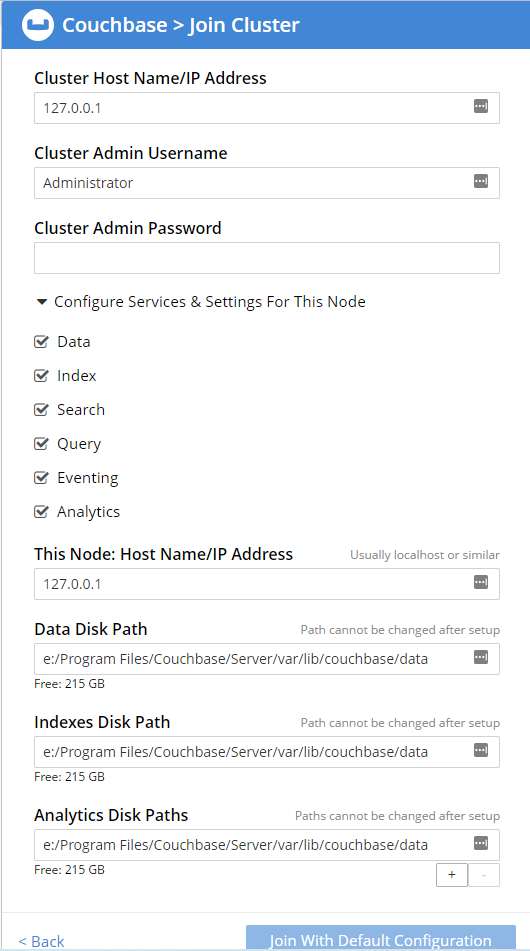
Tiến hành cài đặt Couchbase Server, khi quá trình cài đặt hoàn tất mặc định Couchbase Server sẽ sử dụng cổng 8091. Truy cập vào <http://localhost:8091> để tiếp tục cấu hình thông tin cho Couchbase Server.

### Cấu hình Couchbase Server



Hình ‑ Giao diện Couchbase trên trình duyệt localhost:8091

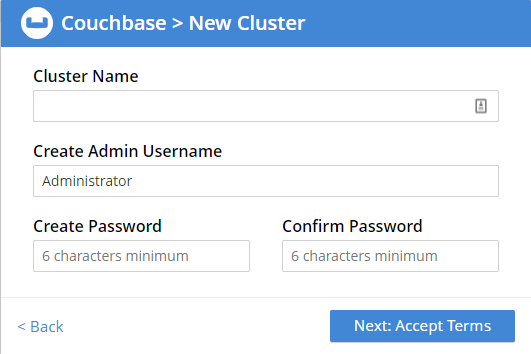
Sẽ có 2 lựa chọn ta có thể thực hiện. Nếu muốn sử dụng máy hiện tại của bạn làm một node khác của server để tăng cường việc xử lí tại một điểm phân tán.



Hình ‑ Cấu hình các thông tin cluster

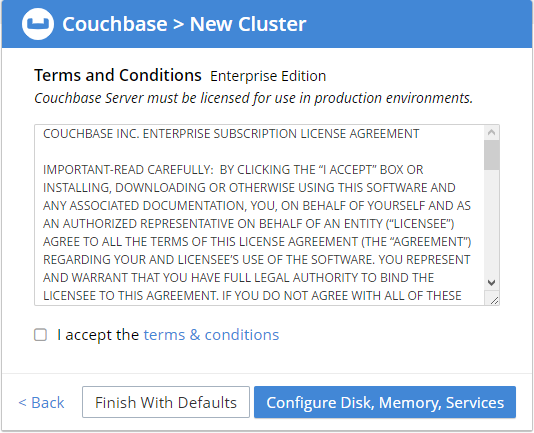
Cần thay đổi giá trị của các field sau để kết nối với một server có sẵn. Đầu tiên tại phần Cluster Host Name/IP Address nhập IP của server ta muốn thêm máy mình làm một node mới. Phần Username và Password nhập thông tin tài khoản quản trị của server muốn kết nối tới. Cuối cùng chọn Join để làm một node mới cho server.

Vì hiện tại chưa có máy nào để ta có thể kết nối, cũng như ta đang thực hiện việc mô phỏng phân tán trên cụm máy, nên cần tạo nhiều server/điểm phân tán. Couchbase Server gọi đây là một cluster. Để tạo mới một cluster làm một điểm phân tán mới, thay vì chọn Join Existing Cluster ta chọn Setup New Cluster.



Hình ‑ Cluster username và password

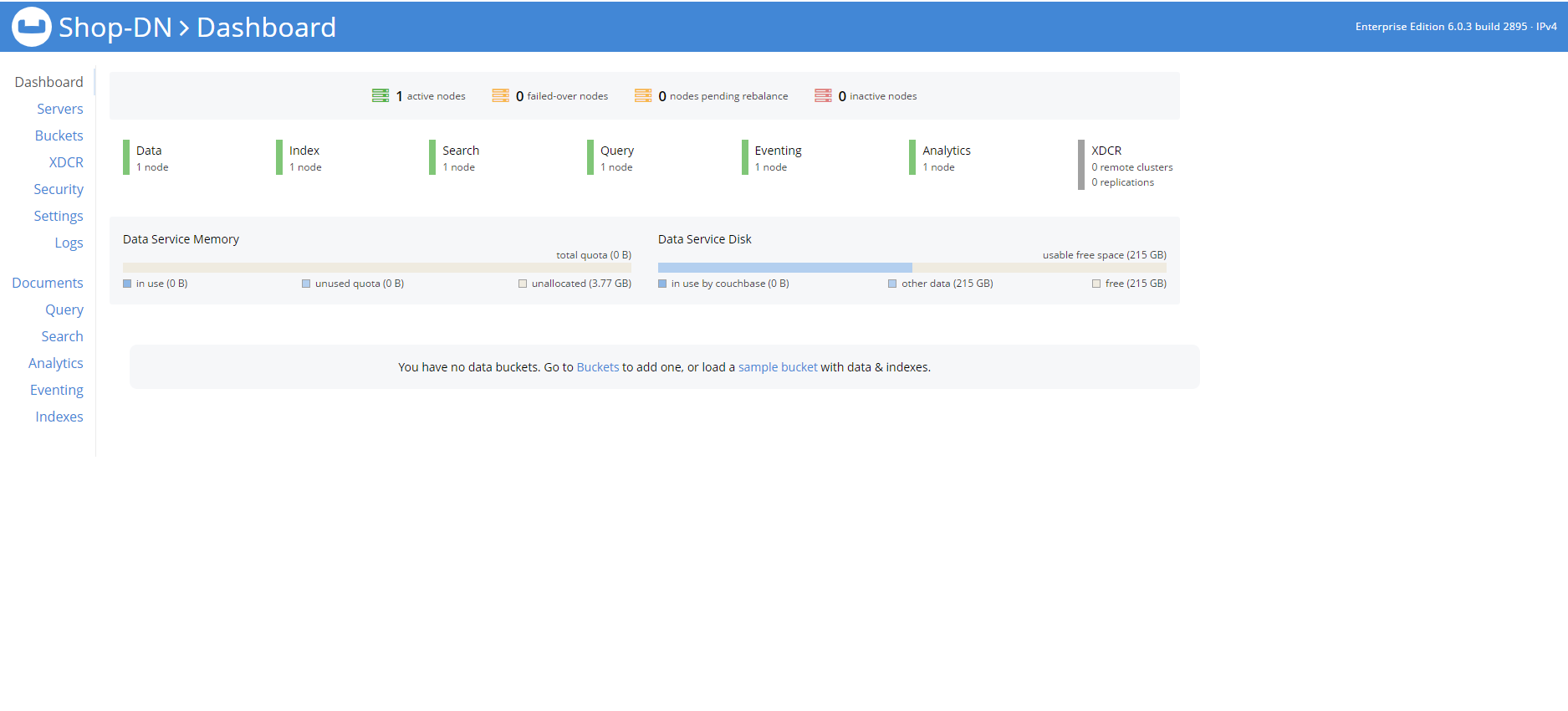
Đặt tên cho Cluster là Shop-DN. Tài khoản quản trị sẽ là Administrator – 123456.



Hình ‑ Couchbase Cluster điều khoản và dịch vụ

Có thể chọn Finish With Defaults để tạo Cluster với những giá trị cấu hình mặc định. Nếu muốn thay đổi thông tin cấu hình, chọn Configure Disk, Memory, Services.

Khi quá trình tạo Cluster mới hoàn tất, một trang mới quản lí Cluster vừa tạo sẽ xuất hiện.



Hình ‑ Trang Dashboard sau khi tạo xong cluster

### Kết nối hai couchbase server

Kết nối Radmin:

Graphical user interface, text, application, chat or text message

Description automatically generated

Hình ‑ Kết nối mạng ảo thông qua RadminVPN

Nhấp Add Remote ở mục XDCR:

Graphical user interface, text, application, email

Description automatically generated

Hình ‑ Mục XDCR và các chức năng bên trong

Nhập các thông tin cần thiết để thực hiện kết nối rồi ấn SAVE.

Graphical user interface, application

Description automatically generated

Hình ‑ Cửa sổ bật lên khi ấn Add Remote ở Mục XDCRb

## Truy vấn dữ liệu phân tán:

### Remote link:

Remote link là tính năng mới nhất của phiên bản Couchbase server 6.6. Remote link cho phép người dùng lấy, phân tích, truy vấn các dữ liệu từ các couchbase data cluster khác nhau.

Remote link gắn liền với Couchbase Analytics service dùng để thực hiện các truy vấn phân tán.

### Thực nghiệm:

**Bước 1:** Tạo bản sao dữ liệu gốc trên máy 1

Dữ liệu gốc ở máy 1:

Graphical user interface, application

Description automatically generated

Hình ‑ Dữ liệu của bảng BANHMY ở máy 1

Ấn vào “Map From Data Service”

Graphical user interface, text, application, chat or text message

Description automatically generated

Hình ‑ Chọn mục Map From Data Service ở mục Analytics

Chọn dữ liệu muốn tạo bản sao truy vấn (ở đây em chọn BANHMY) sau đó ấn Save Mapping.

Graphical user interface, text, application, chat or text message, email

Description automatically generated

Hình ‑ Cấu hình Collection bên trong Map From Data Service

Lúc này mục Analytics Scopes, Links, & Collections sẽ xuất hiện database ảo dùng để truy vấn của BANHMY trên máy 1

Graphical user interface, application

Description automatically generated

Hình ‑ Mục analytics Scopes, Links, & Collections sau khi map

Thực hiện truy vấn thử ở Analytics service máy 1

Graphical user interface, application, Word

Description automatically generated

Hình ‑ Truy vấn dữ liệu máy 1 đã map lại dữ liệu

**Bước 2:** Tạo remote link liên kết đến cơ sở dữ liệu của máy 2

Ấn vào “+ remote link” ở phần Default

Graphical user interface, application

Description automatically generated

Hình ‑ Biểu tượng Remote link

Điền các thông tin cần thiết và ấn SAVE

Graphical user interface, application

Description automatically generated

Hình ‑ Cửa sổ tạo remote link

Lúc này sẽ xuất hiện một database link

Graphical user interface, application

Description automatically generated

Hình ‑ Mục Analytics Scopes, Links, & Collections sau khi tạo remote link

**Bước 3:** Tạo dataset cho remote link lấy dữ liệu của máy 2

Tạo dataset cho database link vừa tạo để thực hiện lấy dữ liệu của máy 2

A picture containing Word

Description automatically generated

CREATE DATASET <tên bucket mới> ON `<bucket của máy 2>`

AT <tên database link>

Graphical user interface, application

Description automatically generated

Hình ‑ Mục Analytics Scopes, Links, & Collections sau khi tạo dataset

Sau đó chúng ta sẽ ấn vào biểu tượng connect để tiến hành kết nối đến dữ liệu trong bucket của máy 2

Graphical user interface, application, website

Description automatically generated

Hình ‑ Biểu tượng connect

Thực hiện truy vấn dữ liệu bucket BANHMY ở máy 2

Graphical user interface, text, application, Word

Description automatically generated

Hình ‑ Truy vấn dữ liệu máy 2

**Bước 4:** Truy vấn phân tán cơ bản

Tìm tên bánh mì có cả ở máy 1 và máy 2

A screenshot of a computer

Description automatically generated with medium confidence

Kết quả truy vấn:

Graphical user interface, application, Word

Description automatically generated

Hình ‑ Kết quả của câu truy vấn đơn giản

# THAO TÁC DỮ LIỆU PHÂN TÁN

## Kỹ thuật nhân bản (Replication)

Kỹ thuật nhân bản trong couchbase giống với kỹ thuật nhân bản trong các hệ quản trị cơ sở dữ liệu khác.

Để tạo một nhân bản của dữ liệu gốc, chúng ta sẽ làm việc với XDCR. Sau khi đã tạo một connection với cluster máy 2, cả hai máy sẽ cùng tiến hành tạo replication.

Các bước tạo replication trên couchbase server như sau:

**Bước 1:** Nhấp vào “ADD REPLICATION”

Graphical user interface, text, application, email

Description automatically generated

Hình ‑ Mục Replication bên trong mục XDCR

**Bước 2:** Nhập các thông tin liên quan như hình dưới

Trong đó Replication From Bucket là nhân bản từ bucket ở máy chủ

Remote Bucket là chọn bucket được dùng để nhân bản ở máy trạm

Remote Cluster là Cluster ở máy trạm

Graphical user interface, text, application, Word

Description automatically generated

Hình ‑ Các setting để nhân bản dữ liệu

**Bước 3:** Tạo một replication tương tự ở máy trạm. Sau đó đợi hệ thống cập nhật dữ liệu từ bucket có dữ liệu là tạo nhân bản thành công. Nhân bản sau khi tạo ra sẽ có sự liên thông, nếu có bất kì thao tác nào lên dữ liệu trên máy trạm hoặc máy chủ, nhân bản và bucket gốc đều sẽ tự động cập nhật.

## Thêm, xóa, sửa dữ liệu phân tán

Sử dụng kỹ thuật nhân bản để thao tác dữ liệu trên máy trạm.

### Tạo XDCR để tiến hành thao tác trên máy 2

Tạo bucket ở máy 1 dùng để thao tác dữ liệu

Graphical user interface, application

Description automatically generated

Hình ‑ Tạo bucket mới dùng để thực hiện các thao tác lên dữ liệu

Nhấp vào “Add replication”

Graphical user interface, text, application, email

Description automatically generated

Hình ‑ Mục Add replication trong mục XDCR

Nhập các thông tin cần thiết để tiến hành replicate bảng BANHMY2 của máy 2. Nhấn SAVE

Graphical user interface, text, application, Word

Description automatically generated

Hình ‑ Thiết lập replication trên máy 1

Làm tương tự trên máy 2, cũng tiến hành tạo một replication đến bảng BANHMY2 ở máy 1.

Graphical user interface, text, application, email, website

Description automatically generated

Hình ‑ Thiết lập replication trên máy 2

### Các thao tác lên dữ liệu của bucket máy 2

Dữ liệu ban đầu máy 1:

Graphical user interface, application

Description automatically generated

Hình ‑ Dữ liệu BANHMY1 ở máy 1

Dữ liệu ban đầu máy 2:

Graphical user interface, text, application

Description automatically generated

Hình ‑ Dữ liệu BANHMY2 ở máy 2

Tiến hành insert dữ liệu vào bucket BANHMY2 ở máy 1

Graphical user interface, application, Word

Description automatically generated

Hình ‑ Thêm dữ liệu vào bucket mới tạo ở máy 1

Kiểm tra lại dữ liệu bảng BANHMY2 ở máy 2

Graphical user interface, text, application, Word

Description automatically generated

Hình ‑ Kiểm tra dữ liệu ở bucket máy 2

Các thao tác xóa và sửa làm tương tự như thao tác thêm. Có thể sử dụng query hoặc sử dụng các công cụ được cung cấp ở mục document để làm các thao tác xóa, sửa, sao chép một cách nhanh chóng.

Graphical user interface, application

Description automatically generated

Hình 3‑11 Thực hiện các thao tác dữ liệu bằng các chức năng Document

# TÀI LIỆU THAM KHẢO

Inc. Couchbase, Couchbase Document <https://docs.couchbase.com/home/index.html> (truy cập lần cuối: 1/1/2022)

Till Westmann, August 12th, 2020, Remote Links: Analyze Trends with Couchbase Analytics <https://blog.couchbase.com/remote-links-analyze-your-enterprise-with-couchbase-analytics/> (truy cập lần cuối: 1/1/2022)