

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH

KHOA ĐÀO TẠO CHẤT LƯỢNG CAO

NGÀNH CÔNG NGHỆ THÔNG TIN

🙡🙣



BÁO CÁO

GVHD: Cô Phượng

SVTH: Đặng Thanh Thế

Nguyễn Tiến Đạt

Nguyễn Khánh Đức

Đăng Khôi Nguyên

Lữ Phước Hưng

Nguyễn Tấn Sang

TP. Hồ Chí Minh, tháng 04 năm 2019

Mục Lục

# GRAPHDB-NEO4J

## Giới thiệu về Graph Database và Cypher

Graph Database mô tả và lưu trữ dữ liệu dưới dạng đồ thị, một cách trực quan và dễ dàng truy vấn.

Dữ liệu trong graph database được lưu dưới dạng các node. Mỗi node sẽ có 1 label, 1 số properties như một row trong SQL. Các node này được kết nối với nhau bằng các relationship. Graph database tập trung nhiều vào relationship giữa các node, áp dụng nhiều thuật toán duyệt node để tăng tốc độ.

Cypher là ngôn ngữ truy vấn cho Graph Database, có đặc điểm dễ đọc và dễ hiểu đối với cả các nhà phát triển, các chuyên gia CSDL,… Cypher cho phép người dùng tìm kiếm thông tin trên CSDL theo một mô hình cụ thể nào đó. Cypher được xem là ngôn ngữ truy vấn đồ thị dễ tìm hiểu nhất, và là một cơ sở tuyệt vời để học về đồ thị. Khi đã hiểu rõ về Cypher, ta có thể dễ dàng học sang các ngôn ngữ truy vấn đồ thị khác.

Cũng như hầu hết các ngôn ngữ khác, Cypher cũng có các câu lệnh riêng. Câu truy vấn đơn giản nhất bao gồm một mệnh đề  **START**  theo sau bởi **MATCH** và **RETURN**.

* Lệnh **START** mô tả một hoặc nhiều điểm bắt đầu có thể là các nút hoặc các mối quan hệ trong đồ thị. Những điểm bắt đầu sẽ được chọn thông qua tìm kiếm chỉ mục hoặc tìm kiếm trực tiếp dựa trên các nút và các mối quan hệ.
* Lệnh **MATCH** đây là phần đặc tả bằng ví dụ từng phần. Ta sử dụng các kí tự ASCII để đại diện cho các nút và các mối quan hệ, “vẽ” ra các dữ liệu mà chúng ta quan tâm. Chúng ta sử dụng dấu đóng ngoặc và mở ngoặc để vẽ các nút (**Ví dụ**: **(a), (b), (c),…**), sử dụng cặp dấu gạch ngang kết hợp với dấu lớn hơn hoặc nhỏ hơn để vẽ ra các mối quan hệ (**–>, <–**). Các dấu lớn hơn (**>**) và nhỏ hơn (**<**) thể hiện hướng của quan hệ. Giữa cặp dấu gạch ngang có cặp đóng mở ngoặc vuông, nội dung trong cặp đóng mở ngoặc vuông được bắt đầu bằng dấu hai chấm và sau đó là tên của mối quan hệ.
* Lệnh **RETURN** Mệnh đề này quy định việc các nút, các mối quan hệ và các thuộc tính trong dữ liệu được nối nên được trả về như thế nào cho client.

## Ưu Điểm và Nhược Điểm

**Ưu Điểm:**

* Khi cần truy vấn các mối quan hệ, graph database **truy vấn nhanh và dễ hơn nhiều** so với database
* Tính linh động cao, một đặc trưng thấy rõ và khác biệt nhất của Graph Database so với các CSDL khác. Với Graph Database, chúng ta không cần xác định trước cấu trúc hay mô hình của CSDL.Với Graph Database thì khác, chúng ta hoàn toàn có thể thêm vào nút hay các mối quan hệ mới mà không làm ảnh hưởng tới các nút khác, cũng không làm thay đổi kết quả truy vấn đang thực hiện.

**Nhược Điểm:**

* Open Source: đưa ra cho những người phát triển với nhiều lợi ích to lớn, tuy nhiên không thể hỗ trợ tốt như SQL Server.
* Thiếu tính ổn định.
* Do còn khá mới, rất ít người biết về ngôn ngữ này.

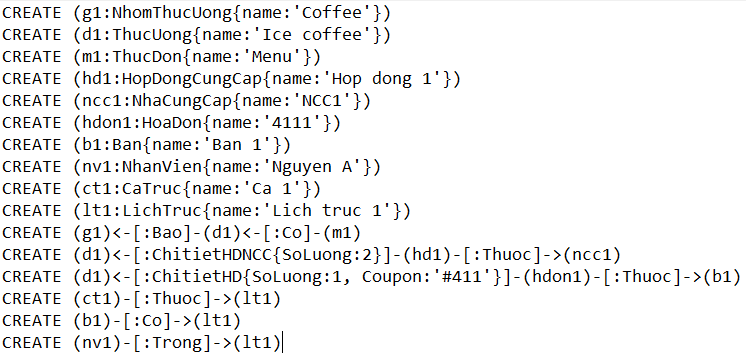
## Demo

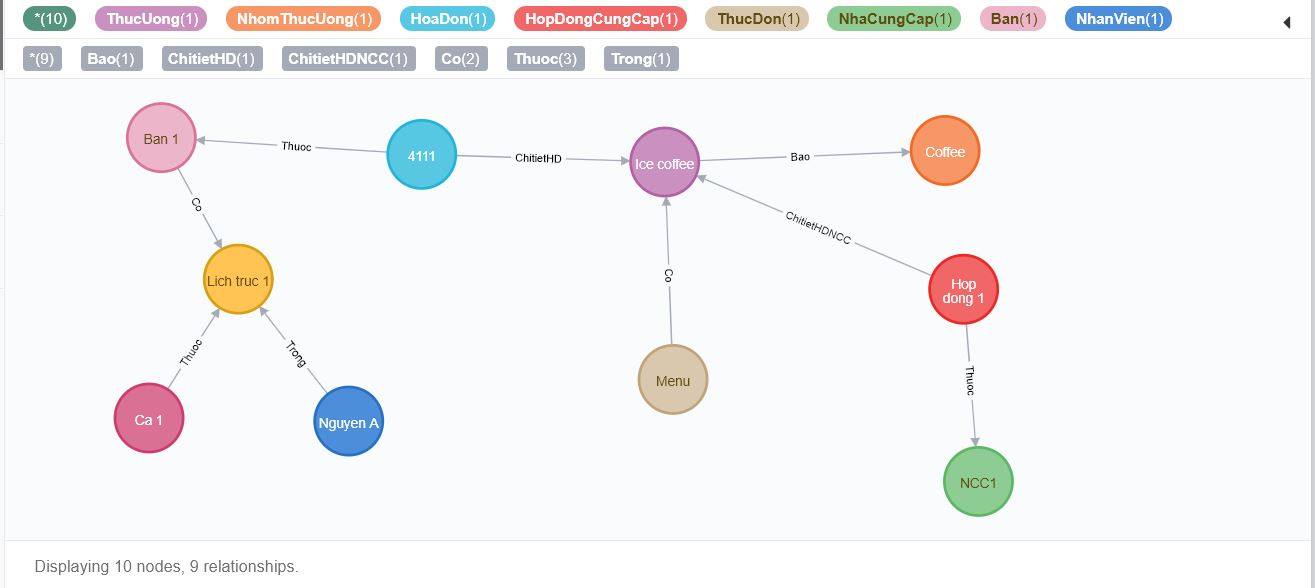
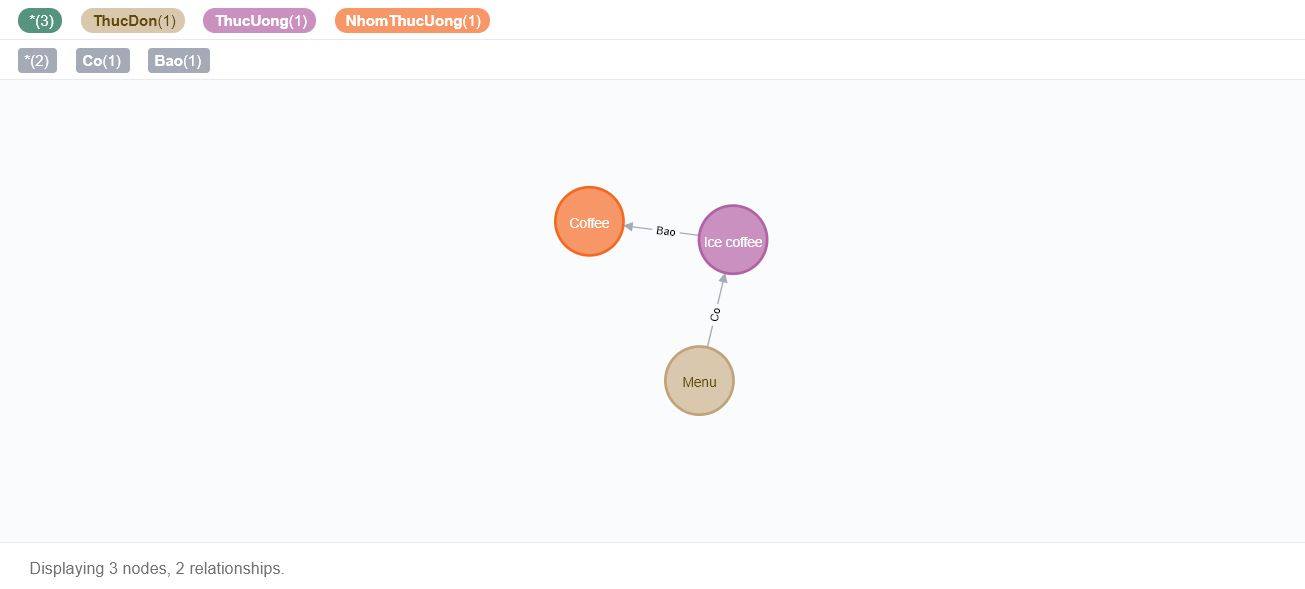
1. Cài Đặt Neo4j

* Download Neo4j tại: <https://neo4j.com/>
* Hướng dẫn cài đặt: <https://www.youtube.com/watch?v=EO57N03U_sIhttps://www.youtube.com/watch?v=EO57N03U_sI>

1. Demo code

* Khởi tạo dữ liệu bằng câu lệnh Create



* Kết quả:
* Xem quan hệ giữa 2 bảng có quan hệ r :
* Thêm thuộc tính cho Object: Nhân Viên

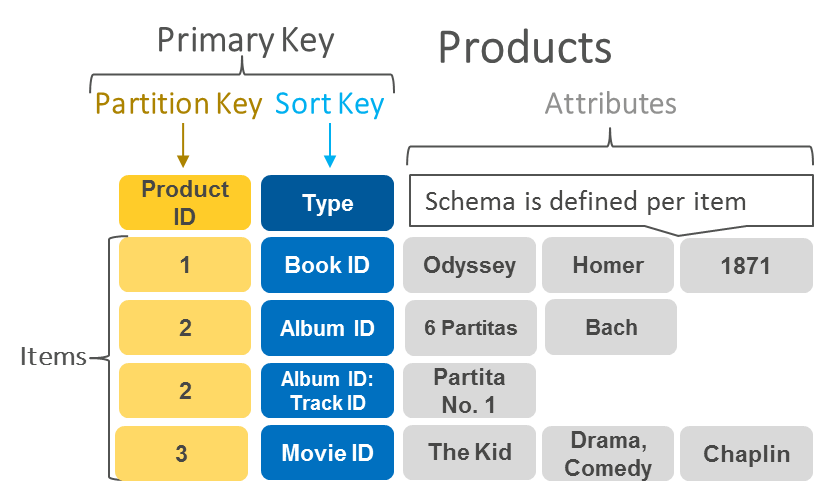
# KEY-VALUE DB: AMAZON DYNAMODB

## Key - value DB: Amazon DynamoDB – dịch vụ cơ sở dữ liệu NOSQL nhanh chóng và linh hoạt cho mọi huy mô :

1. ***The key - value Database :***

A key-value database là một type of nonrelational database sử dụng phương pháp key-value đơn giản để lưu trữ dữ liệu. A key-value database lưu trữ dữ liệu dưới dạng tập hợp các cặp key-value trong đó khóa đóng vai trò là định danh duy nhất. Cả keys and values có thể là bất cứ thứ gì, từ các đối tượng đơn giản đến các đối tượng phức tạp. Key-value databases có khả năng phân vùng cao và cho phép mở rộng theo tỷ lệ mà các loại cơ sở dữ liệu khác không thể đạt được. Ví dụ: [Amazon DynamoDB](https://aws.amazon.com/dynamodb/) phân bổ các phân vùng bổ sung vào bảng nếu phân vùng hiện có lấp đầy dung lượng và cần thêm không gian lưu trữ.

Sơ đồ sau đây cho thấy một ví dụ về dữ liệu được lưu trữ dưới dạng cặp key-value trong DynamoDB.



**Trường hợp sử dụng :**

* Session store .
* Shopping Cart.

**The key-value phổ biến** : Amazon DynamoDB , Apache Cassandra .

1. ***Giới thiệu Amazon DynamoDB –dịch vụ cơ sỡ dữ liệu NoSQL nhanh chóng và linh hoạt cho mọi huy mô :***

Amazon DynamoDB là cơ sở dữ liệu về khóa-giá trị và tài liệu, cung cấp hiệu năng chỉ trong vài mili giây ở mọi quy mô. Đây là cơ sở dữ liệu được quản lý toàn phần, ở nhiều khu vực, nhiều bản chính, có tích hợp tính năng bảo mật, sao lưu và khôi phục, cũng như lưu vào bộ nhớ đệm trong bộ nhớ dành cho các ứng dụng trên quy mô Internet. DynamoDB có thể xử lý hơn 10 nghìn tỷ yêu cầu/ngày và hỗ trợ tối đa hơn 20 triệu yêu cầu/giây.

Nhiều doanh nghiệp phát triển nhanh nhất thế giới như Lyft, Airbnb, Redfin, Samsung, Toyota và Capital One dựa vào quy mô và hiệu năng của DynamoDB để hỗ trợ khối lượng công việc then chốt của họ.

Hơn 100.000 khách hàng AWS đã lựa chọn DynamoDB làm cơ sở dữ liệu về khóa-giá trị và tài liệu cho các ứng dụng di động, web, trò chơi, công nghệ quảng cáo, IoT và các ứng dụng khác cần quyền truy cập dữ liệu với độ trễ thấp ở mọi quy mô. Tạo một bảng mới cho ứng dụng của bạn và để DynamoDB đảm nhiệm phần còn lại.

1. ***Lợi ích :***

Tính năng mới của DynamoDB :

* Chế độ công suất theo yêu cầu.
* Khả năng hỗ trợ tích hợp cho các giao tác ACID.
* Tính năng sao lưu theo yêu cầu và khôi phục về thời điểm trước đó.
* Mã hóa ở trạng thái lưu trữ.

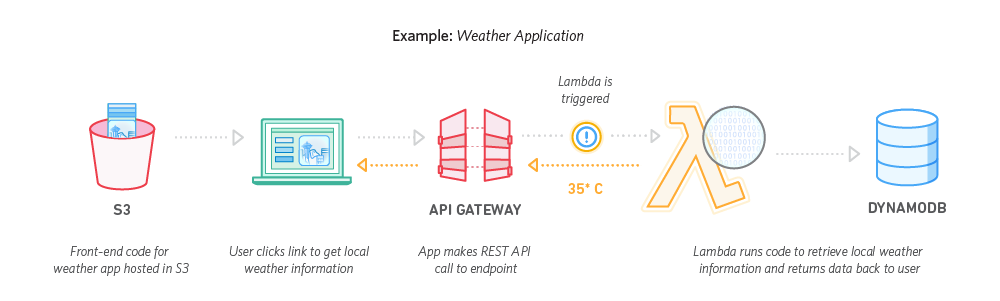
**Hiệu năng trên qui mô lớn :** DynamoDB là cơ sở dữ liệu khóa-giá trị và tài liệu có thể hỗ trợ các bảng ở hầu như mọi kích cỡ bằng khả năng thay đổi quy mô theo chiều ngang. Nhờ đó, DynamoDB có thể mở rộng quy mô lên hơn 10 nghìn tỷ yêu cầu/ngày với mức đỉnh điểm là hơn 20 triệu yêu cầu/giây, trong dung lượng lưu trữ petabyte . DynamoDB cung cấp mô hình dữ liệu khóa-giá trị và tài liệu, độ trễ micrô giây với DynamonDB Accelerator , tự động sao chép các bảng toàn cầu, xử lý dữ liệu theo thời gian thực bằng DynamonDB Streams.

**Serverless :** Với DynamoDB, bạn không cần phải cung cấp, vá hoặc quản lý máy chủ, cũng như không cần phải cài đặt, duy trì hoặc vận hành phần mềm. DynamoDB tự động thay đổi quy mô các bảng để điều chỉnh công suất và duy trì hiệu năng mà không cần quản trị. Tích hợp dung sai cao và độ sẵn sàng hoạt động, nhờ đó ứng dụng của bạn không cần phải thiết kế các tính năng này. DynamoDBcung cấp chế độ công suất đọc/ghi , chế độ theo yêu cầu , tự động thay đổi qui mô , theo dõi sự thay đổi bằng trình kích hoạt .

**Phù hợp dùng trong doanh nghiệp** : DynamoDB được thiết kế cho khối lượng công việc then chốt, bao gồm hỗ trợ các giao tác ACID cho tập hợp ứng dụng khác nhau đòi hỏi logic công việc phức tạp. DynamoDB giúp bảo vệ dữ liệu của bạn bằng tính năng mã hóa và liên tục sao lưu dữ liệu với độ tin cậy được đảm bảo thông qua [thỏa thuận mức dịch vụ](https://aws.amazon.com/dynamodb/sla/). DynamoDB cung cấp giao tác ACID , mã hóa ở trạng thái lưu trữ , khôi phục về thời điểm trước đó , sao lưu và khôi phục theo yêu cầu .

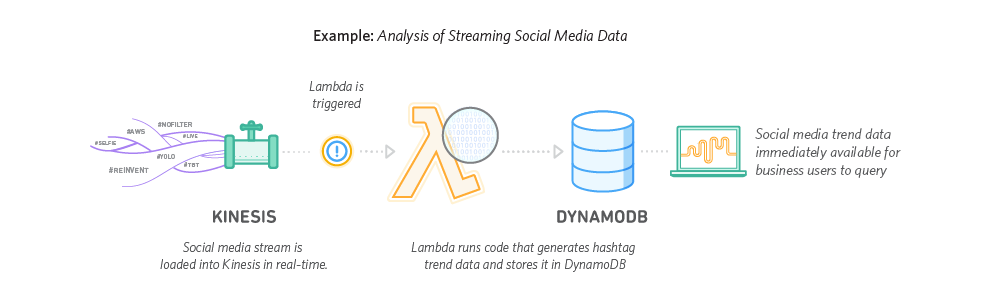
1. ***Trường hợp sử dụng :***

**Ứng dụng web serverless :** Phát triển các ứng dụng web mạnh mẽ mà có thể tự động tăng hoặc giảm quy mô. Bạn sẽ không cần phải duy trì những máy chủ và các ứng dụng của bạn luôn có có độ khả dụng cao.



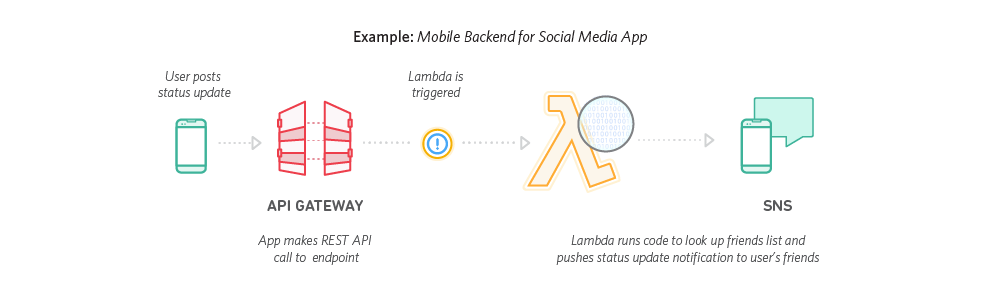
*Kiến trúc tham khảo Use-Case WebApplications*

**Kho dữ liệu vi dịch vụ :** Xây dựng các vi dịch vụ linh hoạt và có thể tái sử dụng nhờ dùng DynamoDB như một kho dữ liệu serverless để đạt hiệu năng nhanh và ổn định.



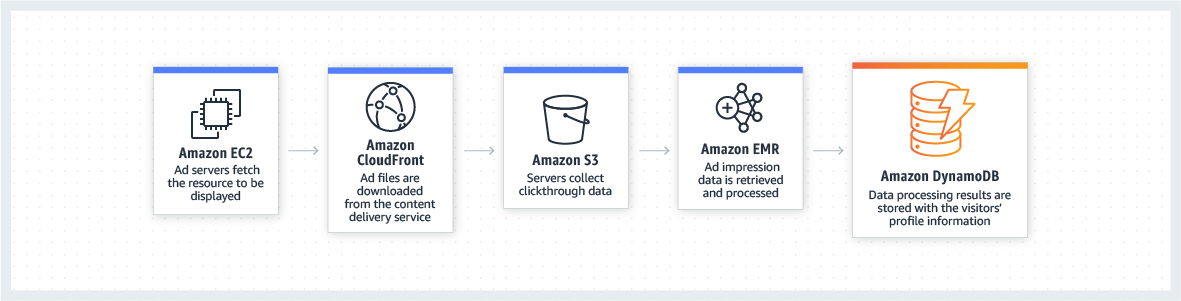
*Kiến trúc tham khảo Use-Case StreamProcessing*

**Backend di động :**Phát triển những ứng dụng di động được cá nhân hóa cùng trải nghiệm mượt mà cho những người dùng của bạn. DynamoDB sẽ lo liệu những tác vụ vận hành, để bạn có thể tập trung vào những ứng dụng của bạn.



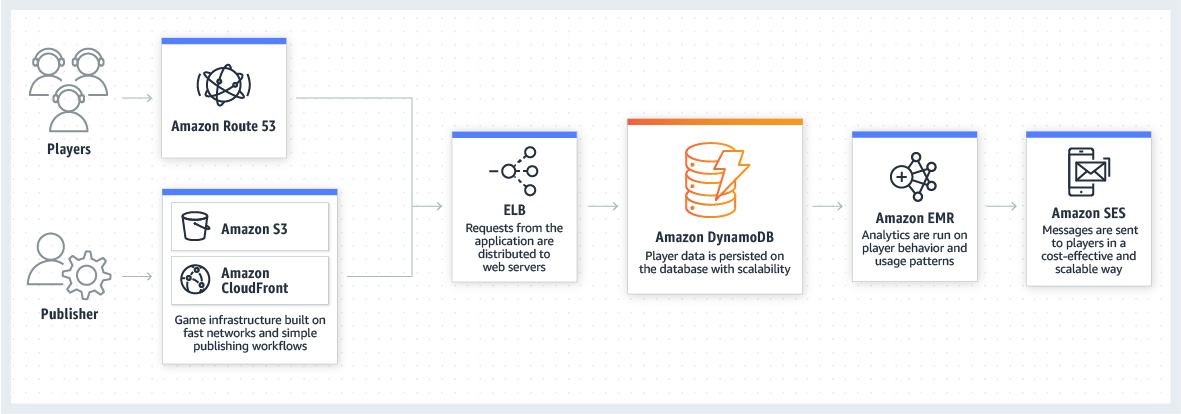
*Kiến trúc tham khảo Use-Case MobileBackends*

**Công nghệ quảng cáo :** Tạo một nền tảng đấu thầu thời gian thực và công cụ khuyến nghị bằng khả năng điều chỉnh quy mô, thông lượng và tính khả dụng của DynamoDB.



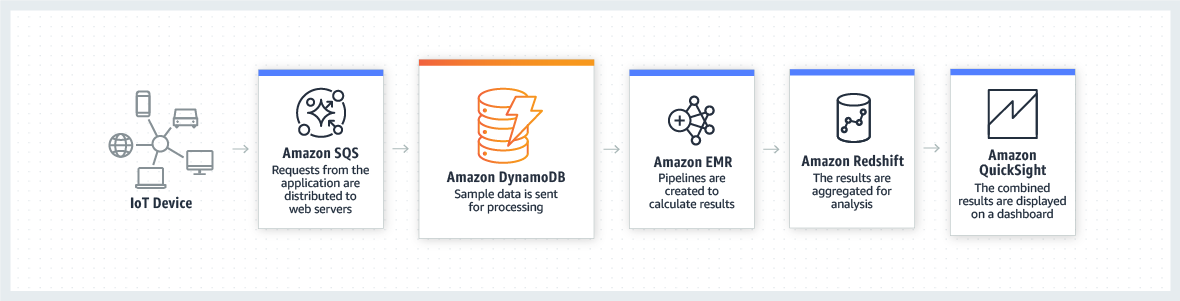
*Kiến trúc tham khảo Product-page-diagram-DynamoDB-Ads*

**Trò chơi :**Tạo trò chơi có thiết kế phản hồi cho di động, máy chơi game và máy tính bàn bằng DynamoDB. Lưu trữ và truy vấn dữ liệu trò chơi như trạng thái người chơi, bảng điểm hoặc nội dung động trong thế giới của trò chơi.



*Kiến trúc tham khảo Product-page-diagram-DynamoDB-Gaming*

**IOT :** Phân tích các thiết bị của bạn bằng cách kết nối dữ liệu IoT dung lượng lớn, tốc độ cao tại DynamoDB tới Amazon Redshift và Amazon QuickSight.



*Kiến trúc tham khảo Product-page-diagram-DynamoDB-IOT*

## Ưu nhược điểm của DynamodDB :

**Ưu điểm :**

* Với DynamoDB, AWS thực sự đã đơn giản hóa các vấn đề cấu hình phức tạp - bạn không phải sử dụng bất kỳ công cụ Database Management Studio, không phải lo lắng về Database Driver và cũng không phải cấu hình các Cluster.
* Chỉ đơn giản là tạo Primary Key, cung cấp RCU/WCU, sử dụng SDK và JSON để làm việc với DynamoDB.
* AWS DynamoDB hứa hẹn về hiệu suất với khả năng mở rộng (scalable) liên tục - chỉ bị giới hạn bởi khả năng vật lý của AWS.
* DynamoDB cung cấp dịch vụ với chi phí thường thấp hơn so với xây dựng kho dữ liệu key-value.

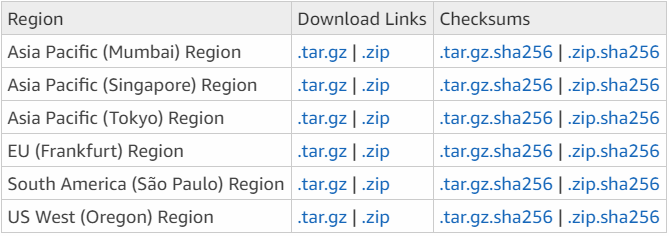
**Nhược điểm :**

* Hơi khó sử dụng cho người mới tiếp cận NoSQL.
* Bị giới hạn lưu lượng lưu thông (lượng write/read mỗi giây được cho phép) theo chi phí phải trả .

## Cài đặt DynamoDB :

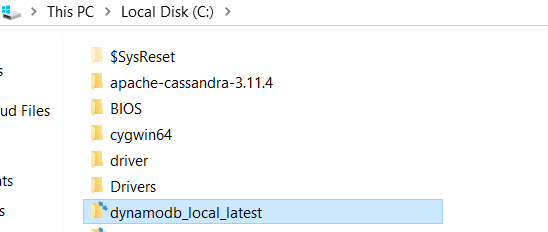
1. ***Tải một trong những phiên bản DynamoDB miễn phí bằng những link sau ở trang:***

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/DynamoDBLocal.DownloadingAndRunning.html>



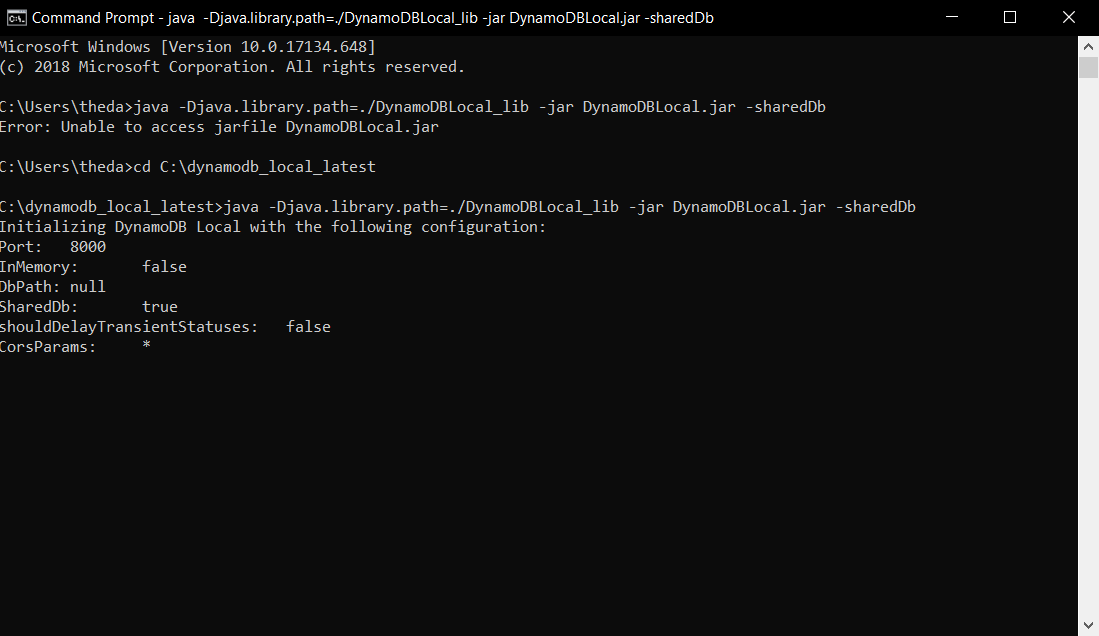
\*Note: để chạy được dynamoDB trên thiết bị, bạn bắt buộc phải có JRE(Java Runtime Enviroment version 6.x trở lên, DynamoDB không chạy được ở các phiên bản thấp hơn

1. ***Sau khi tải bản archive, giải nén và sao chép vào thư mục bạn chọn***



1. ***Để khởi động DynamoDB trên thiết bị, mở command promt window, điều hướng đến folder trên và nhập dòng command sau :***

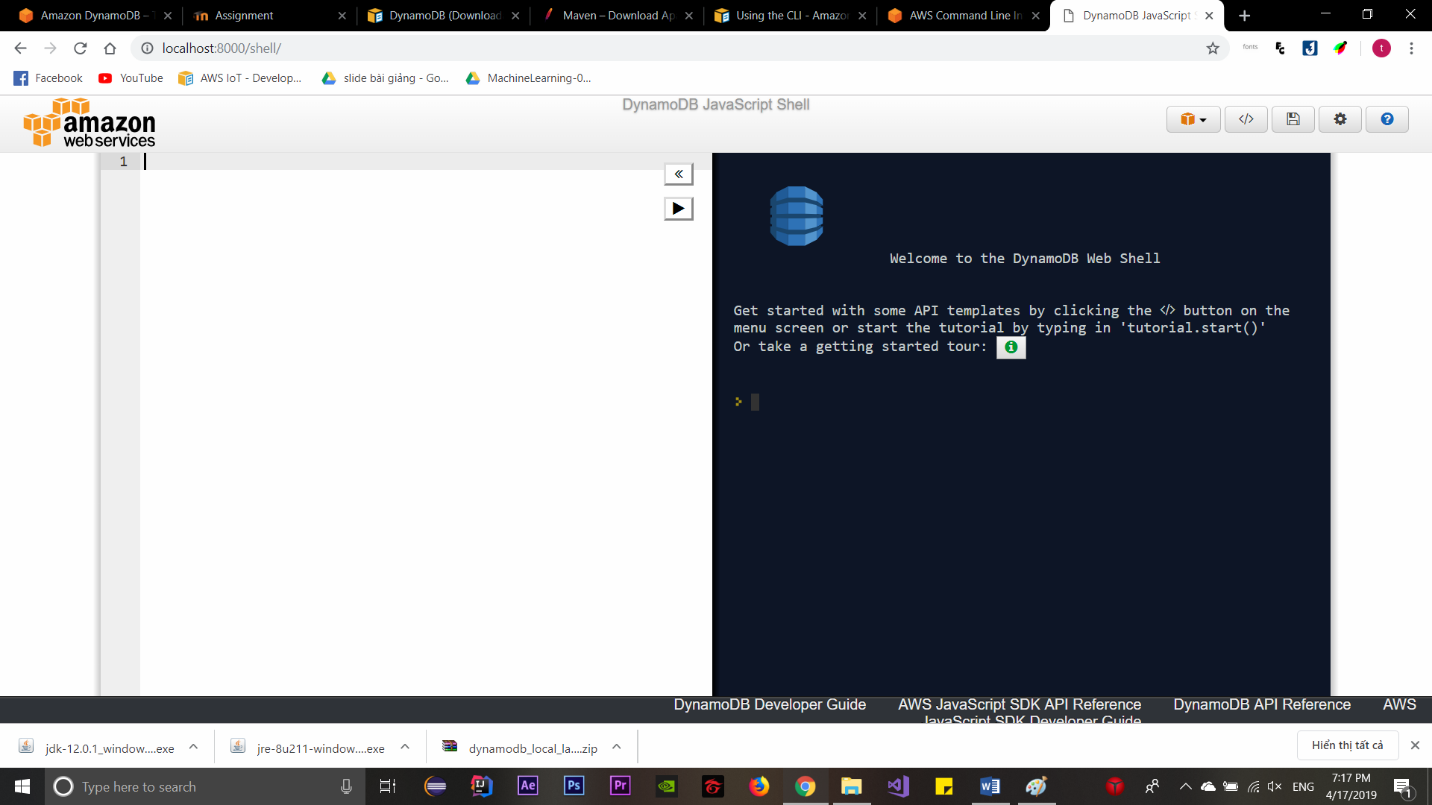
java -Djava.library.path=./DynamoDBLocal\_lib -jar DynamoDBLocal.jar -sharedDb



DynamoDB xử lý các yêu cầu đến cho đến khi bạn dừng nó. Để dừng DynamoDB, nhấn Ctrl + C tại dấu nhắc lệnh.

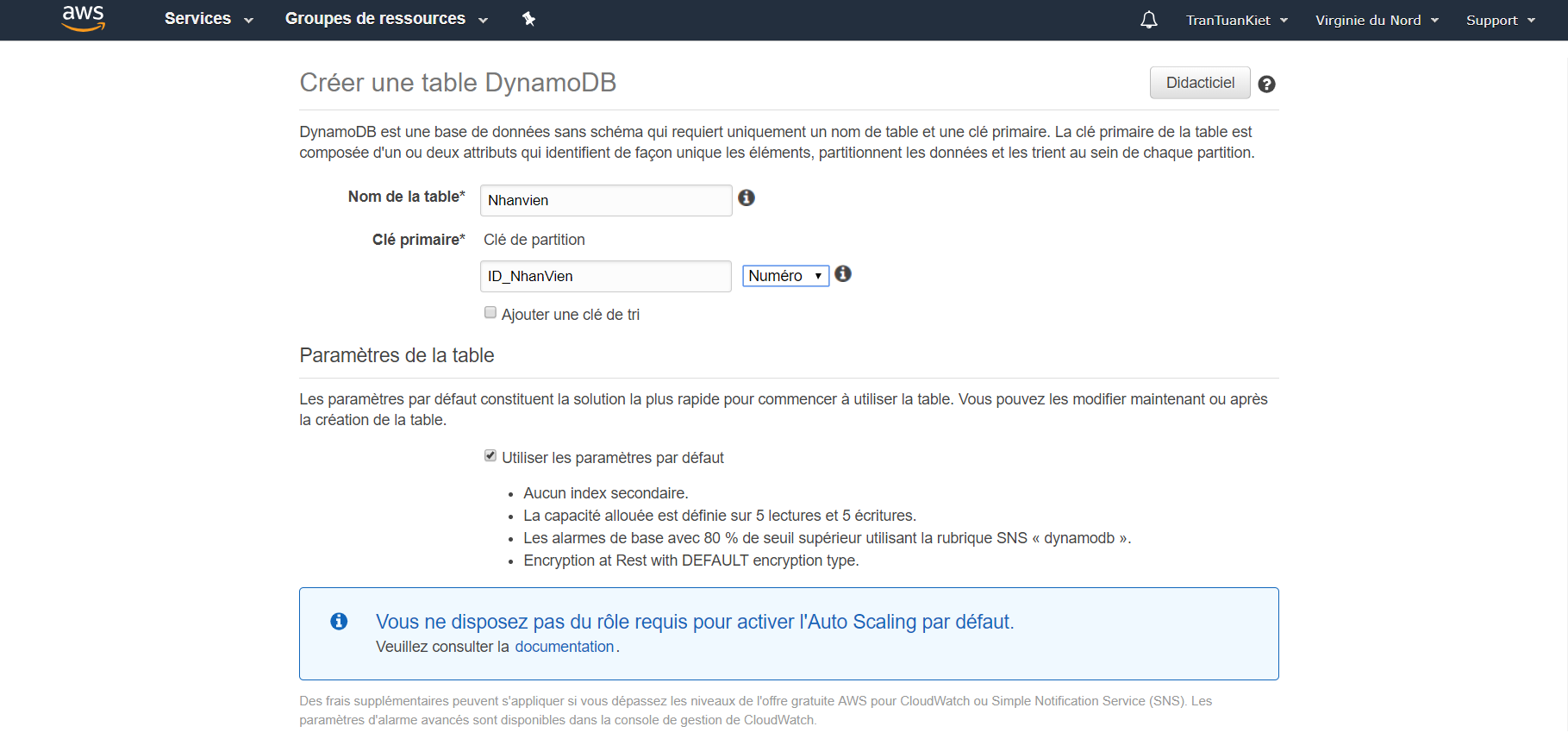
DynamoDB sử dụng cổng 8000 theo mặc định.

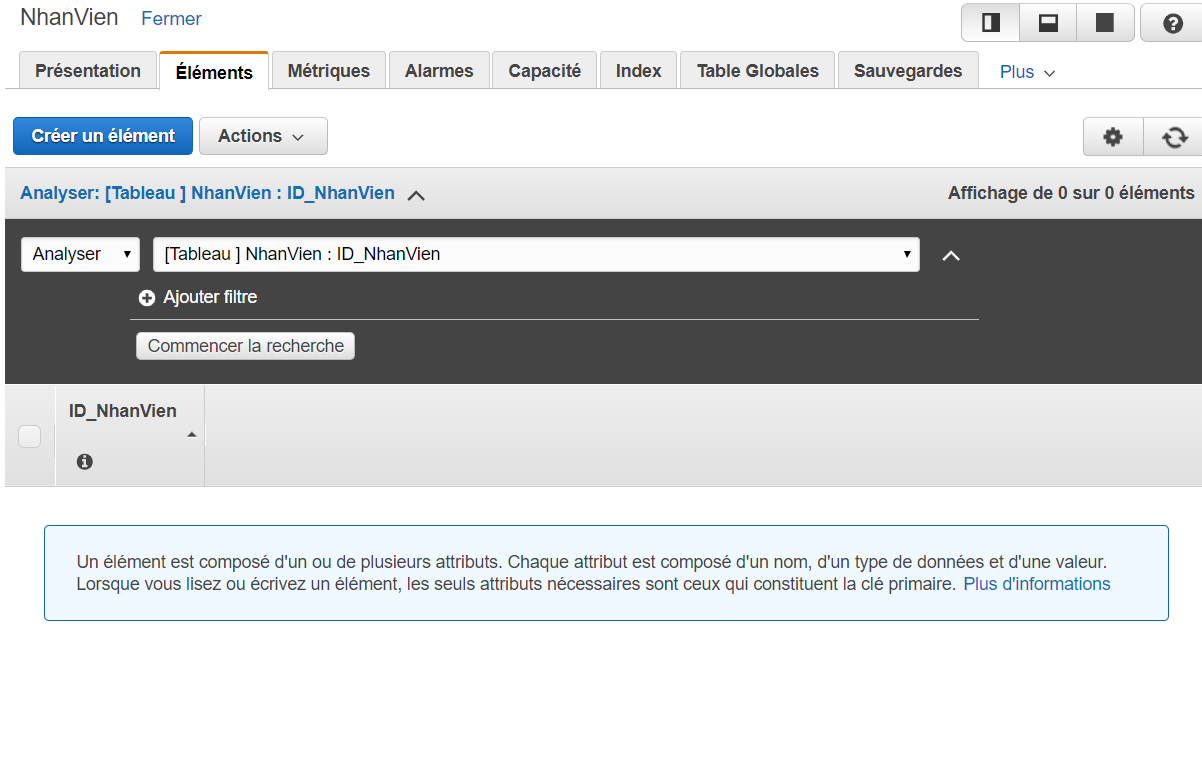
1. ***Để access vào DynamoDB Shell, ta nhập localhost:8000/shell ở thanh tìm kiếm***



1. ***Demo DynamoDB ( table NhanVien)* :**

* Khởi tạo table NhanVien :





* Tạo file NhanVien.json và nhập code để tạo thông tin của một nhân viên:

{

"NhanVien": [

{

"PutRequest": {

"Item": {

"ID\_NhanVien": {

"N": "2019"

},

"TenNhanVien": {

"S": "NguyenTanSang"

},

"NgaySinh": {

"S": "20-09-1998"

},

"SDT": {

"N": "0399677744"

},

"DiaChi": {

"S": "ThuDuc"

},

"Luong": {

"N": "4000000"

}

}

}

}

]

}

* Tiếp theo mở command và nhập các lệnh sau ( lưu ý trỏ vào đường dẫn lưu file NhanVien.json) :

**C:\Users\Asus\Downloads\Test\aws sdk>aws dynamodb batch-write-item --request-items** [**file://NhanVien.json**](file://NhanVien.json)

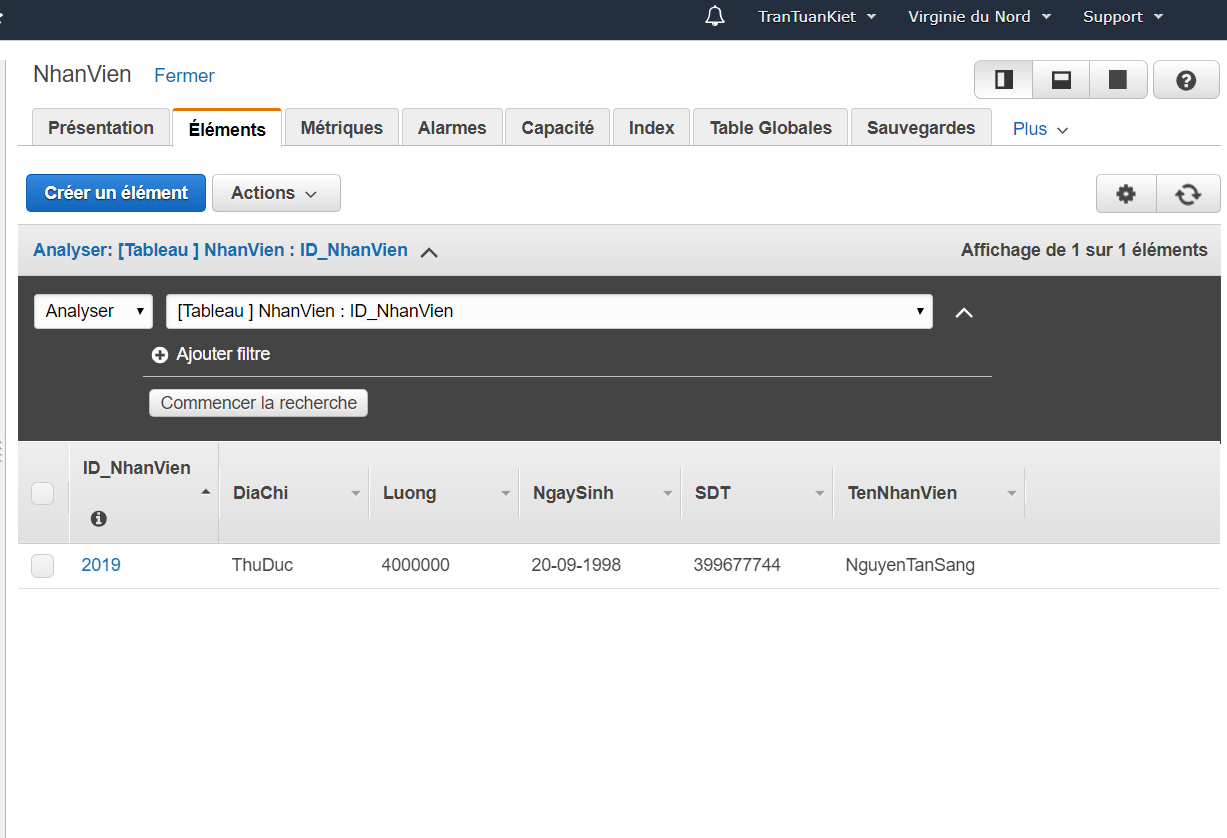
* Thành công sẽ hiện thông báo :

**{**

**"UnprocessedItems": {}**

**}**

* Kết quả :

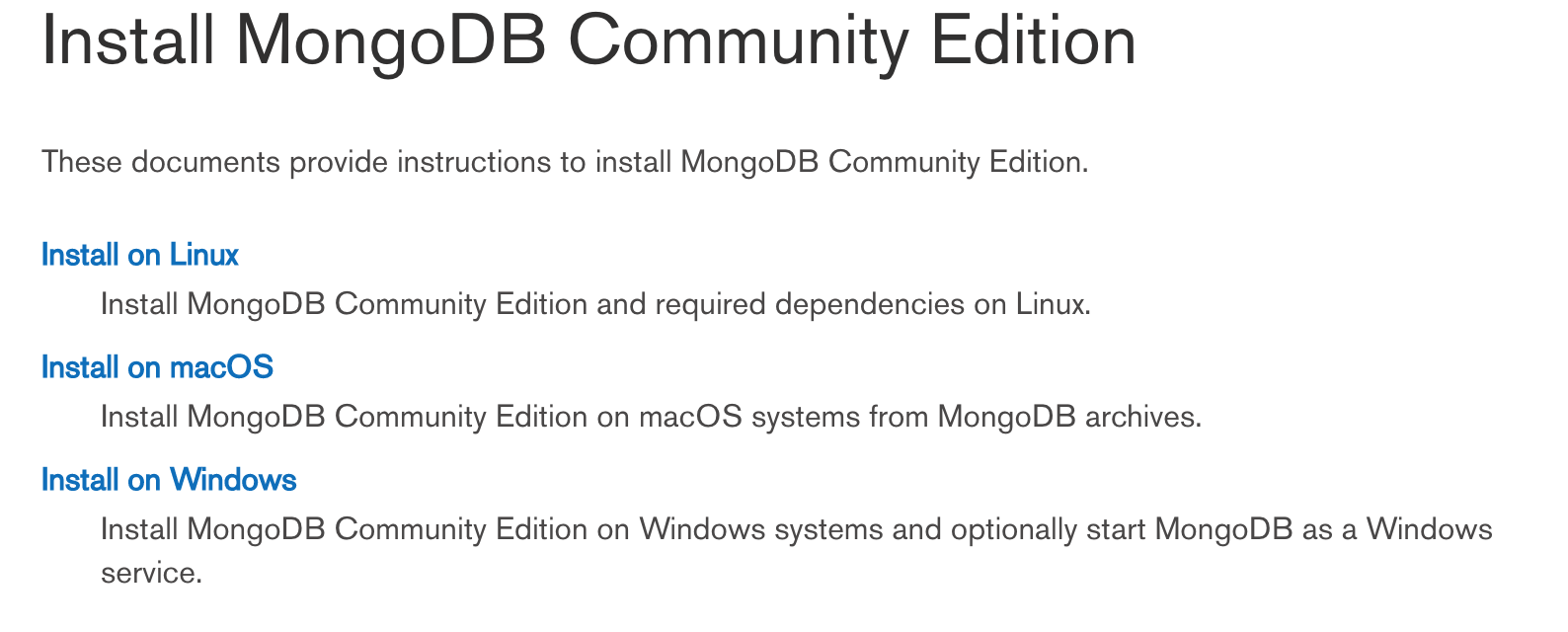


# DOCUMENT-BASED: MONGODB

## Cài đặt

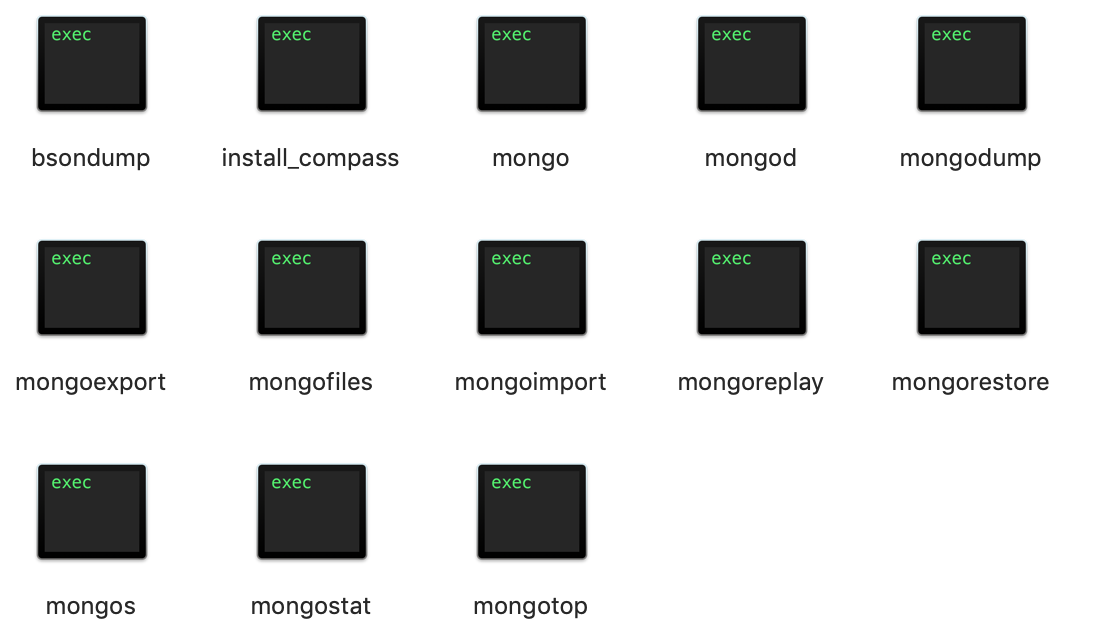
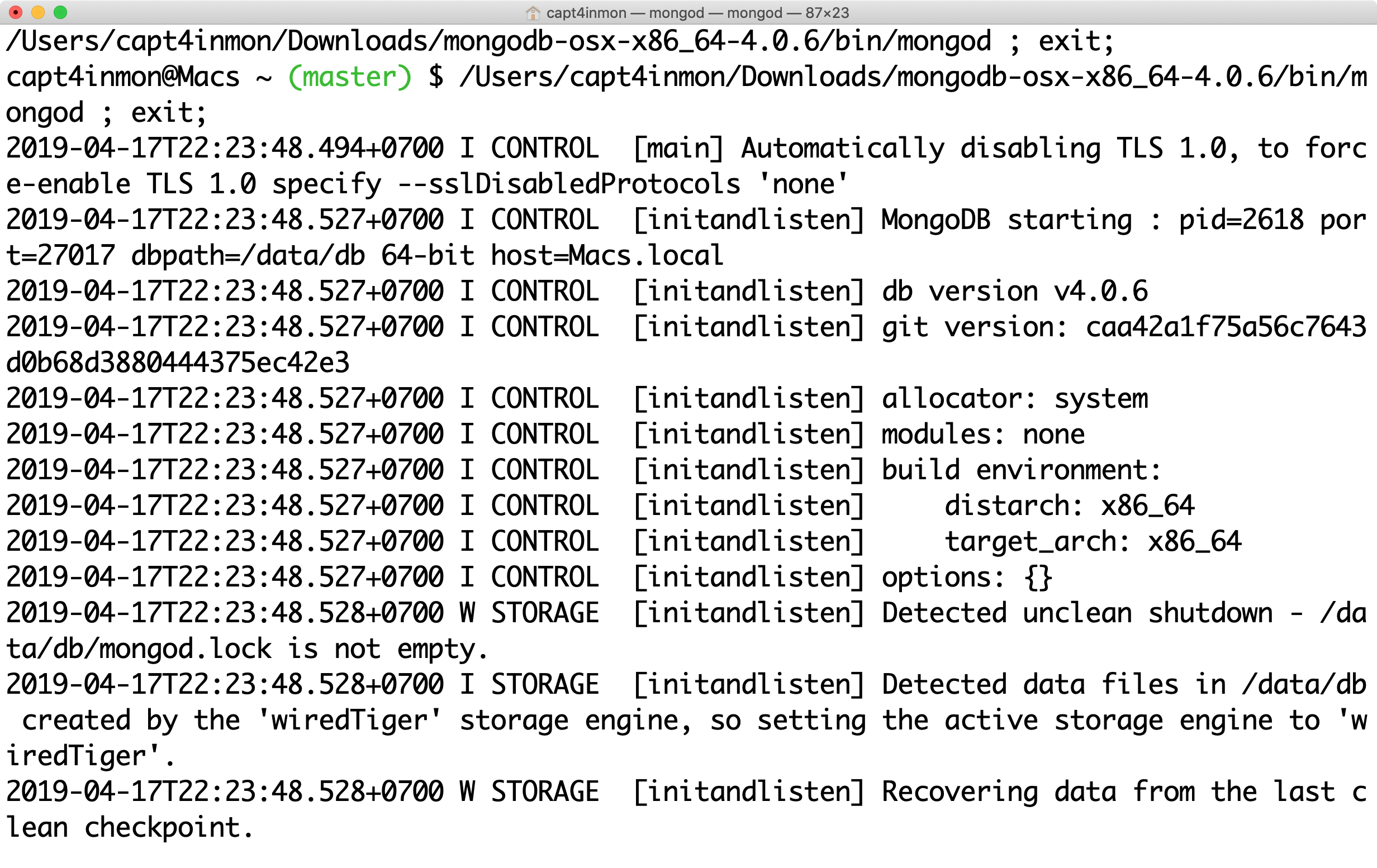
1. Cài đặt MongoDB Community theo từng OS ở trang:

https://docs.mongodb.com/manual/administration/install-community/



\*Note: Ở trên Linux hoặc macOS bạn cần homebrew(cli quản lý các package) sau đó trỏ đường dẫn lệnh tới thư mục mongodb vừa cài.

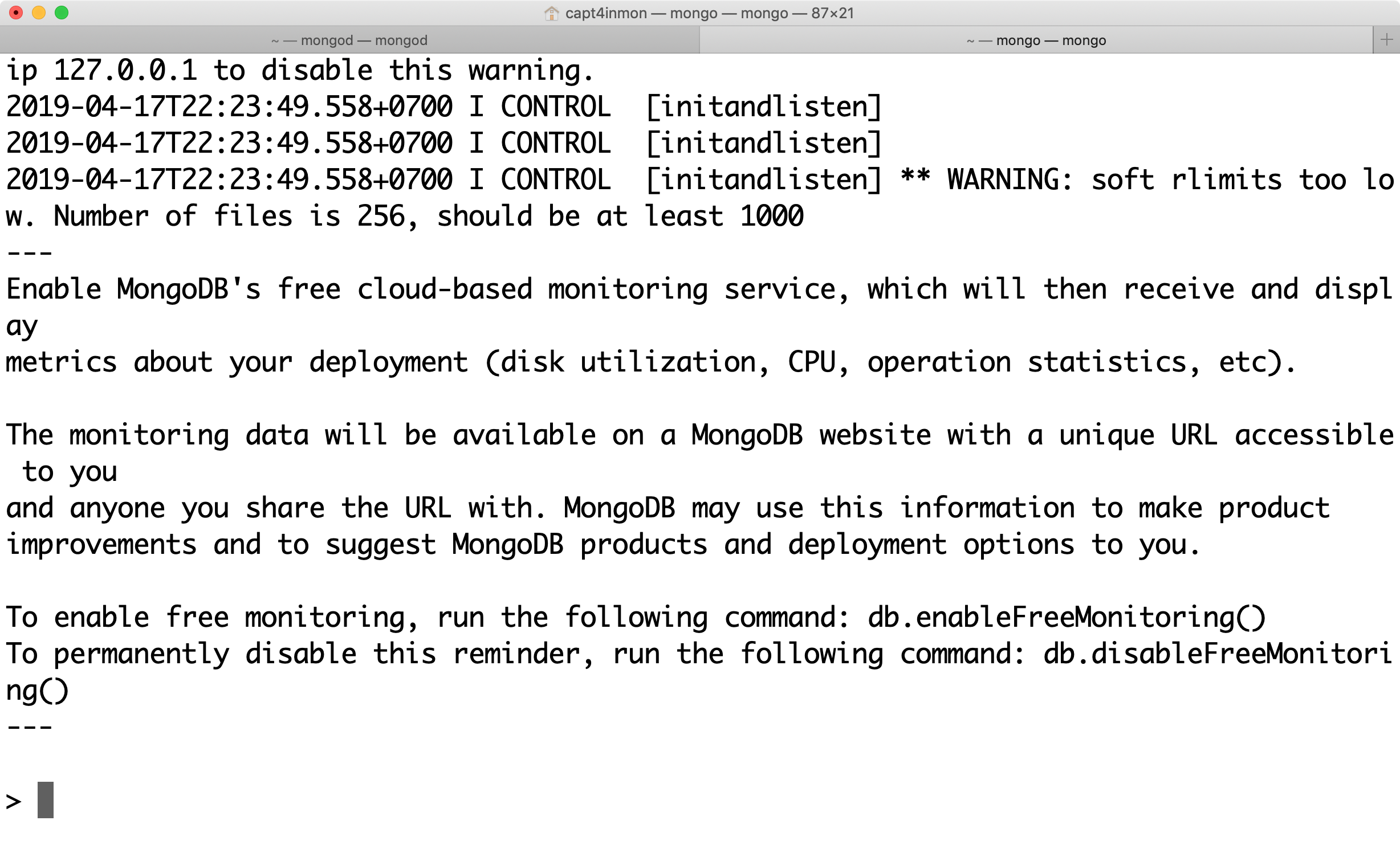
Hoặc có thể tải trực tiếp mongodb.

1. ***Sau khi tải, giải nén và sao chép vào thư mục bạn chọn***
   1. Để khởi động MongoDB trên thiết bị, mở terminal trên mac và gõ (mongod) để start server trên thiết bị

MongoDB xử lý các yêu cầu đến cho đến khi bạn dừng nó. Để dừng

Ctrl + C tại dấu nhắc lệnh.

DynamoDB sử dụng cổng 27017 theo mặc định.

* 1. Để access vào Mongo Shell, ta nhập mongo ở terminal 

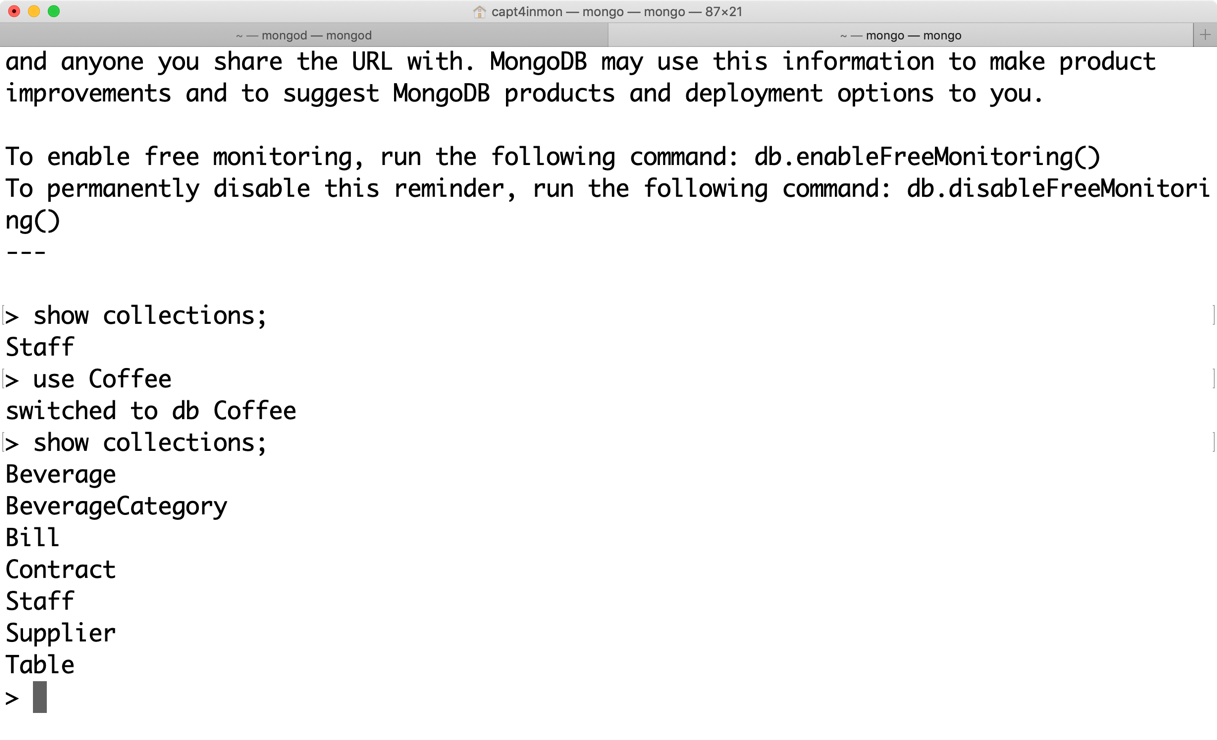
## Deploy mô hình database cho ứng dụng quản lý café

* 1. Đầu tiên trong shell mongodb ta sử dụng lệnh
* use Coffee

Tạo database Coffee

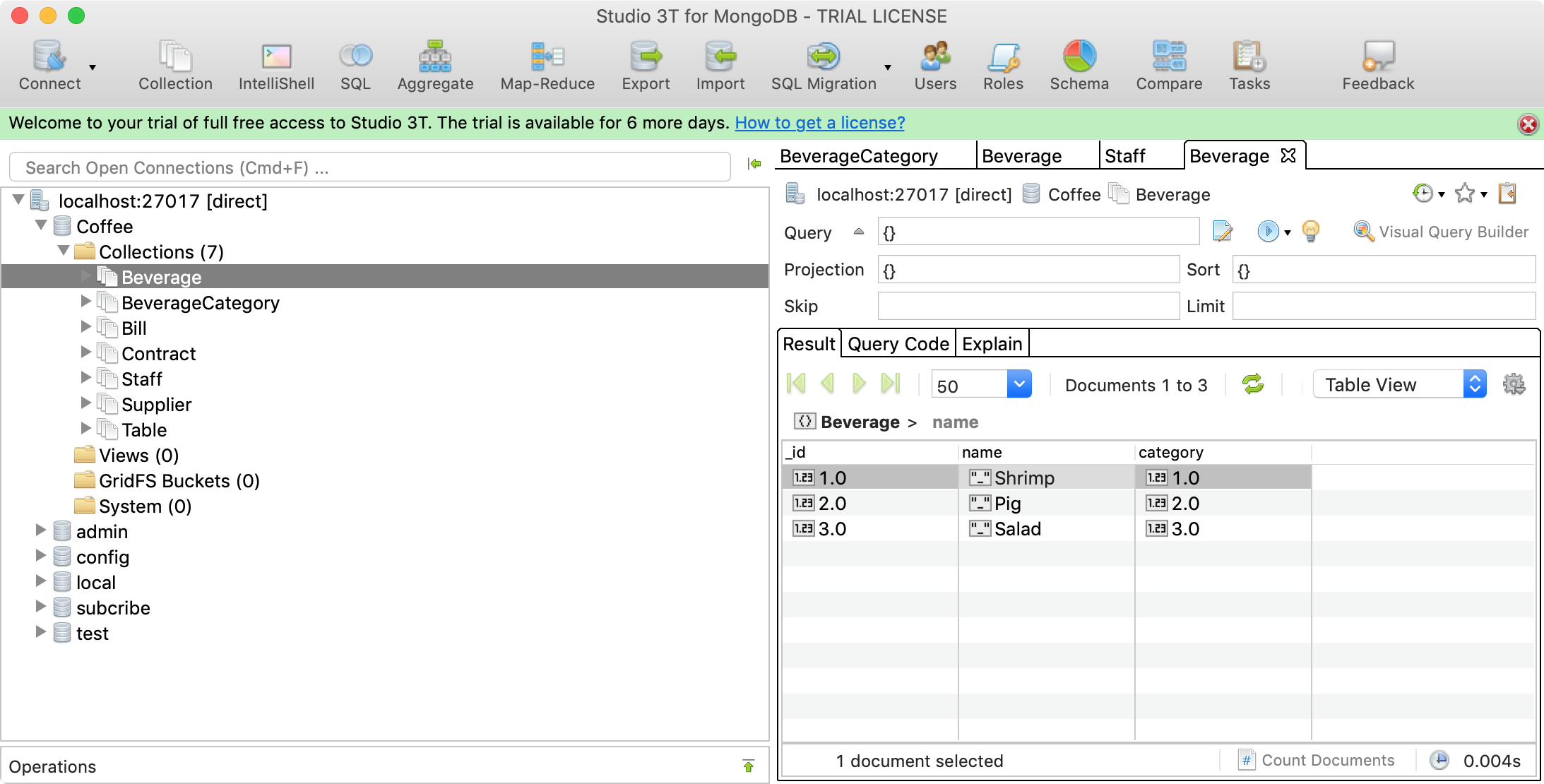
* 1. Tiếp theo tạo các collection
* db.createCollection(name, options)

Tạo Collection



* 1. Thêm dữ liệu vào các collection bằng
* db.Staff.insert({\_id: 1, name: “Duc”, sex: “male”, age: “20”, role: “quan li”})

## Kết Quả



## Đánh giá ưu nhược

* 1. Lược đồ ít hơn - MongoDB là một cơ sở dữ liệu tài liệu trong đó một bộ sưu tập chứa các document khác nhau. Số lượng trường, nội dung và kích thước của tài liệu có thể khác nhau từ document này sang tài liệu khác.
  2. Cấu trúc của một đối tượng là rõ ràng.
  3. Không có liên kết bảng.
  4. Khả năng truy vấn sâu.
  5. MongoDB hỗ trợ các truy vấn động trên các tài liệu bằng ngôn ngữ truy vấn dựa trên tài liệu gần như mạnh mẽ như SQL.
  6. Dễ mở rộng quy mô - MongoDB dễ dàng mở rộng quy mô. Chuyển đổi / ánh xạ các đối tượng ứng dụng thành các đối tượng cơ sở dữ liệu không cần thiết.
  7. Cho phép truy cập dữ liệu nhanh hơn SQL.

## Vì sao nên chọn MongoDB

* 1. Lưu trữ hướng tài liệu - Dữ liệu được lưu trữ dưới dạng tài liệu kiểu JSON.
  2. Index trên bất kỳ thuộc tính nào.
  3. Nhân rộng và tính sẵn sàng cao.
  4. Truy vấn phong phú
  5. Cập nhật nhanh

# COLUMN-BASED: CASSANDRA

## Giới thiệu

Nói về phần mềm trung gian NoSQL dẫn đầu là Google BigTable và Amazon Dynamo, nhưng nhiều thứ khác cũng xuất hiện trong thế giới open source. Trong số đó, Apache Cassandra (Cassandra) đang thu hút sự chú ý đặc biệt gần đây.

Cassandra ban đầu được tạo ra bởi Facebook. Sau đó nó đã được tặng cho Quỹ Apache và tháng 2 năm 2010 và được nâng cấp lên thành dự án hàng đầu của Apache.

<http://cassandra.apache.org/>

Cassandra là một cơ sở dữ liệu phân tán kết hợp mô hình dữ liệu của Google Bigtable với thiết kế hệ thống phân tán như bản sao của Amazon Dynamo. Tất cả được thực hiện bởi Java và được cung cấp trong ASL 2 (Giấy phép Phần mềm Apache Phiên bản 2).

1. ***Cài đặt môi trường***

* Step 1: cài đặt OpenJDK JRE 8.0

Apache Cassandra yêu cầu bản phát hành Java 8. OpenJDK JRE 1.8 mới nhất như sau:

sudo apt install openjdk-8-jre –y

Đã cài đặt OpenJDK JRE 1.8, bạn có thể xác nhận kết quả cài đặt:

java -version

Đầu ra sẽ tương tự như sau:

openjdk version "1.8.0\_151"

OpenJDK Runtime Environment (build 1.8.0\_151-8u151-b12-0ubuntu0.16.04.-b12)

OpenJDK 64-Bit Server VM (build 25.151-b12, mixed mode)

Tùy chọn, bạn có thể tạo biến môi trường JAVA\_HOME như sau:

echo "JAVA\_HOME=$(readlink -f /usr/bin/java | sed "s:bin/java::")" | sudo tee -a /etc/profile

source /etc/profile

echo $JAVA\_HOME

* Step 2: cài đặt Python 2.7, nếu nó thiếu trên hệ thống của bạn

Trước tiên, hãy xác định sự tồn tại và phiên bản của Python trên máy của bạn:

python -V

Trên Ubuntu 16.04 LTS, đầu ra có thể hơi khó hiểu:

The program 'python' can be found in the following packages:

\* python-minimal

\* python3

Ask your administrator to install one of them

Điều đó thực sự có nghĩa là bạn cần phải tự cài đặt Python 2.7:

sudo apt install python -y

Chạy lại lệnh python -V và đầu ra sẽ trở thành:

Python 2.7.12

* Step 3: Cài đặt bản phát hành ổn định mới nhất của Apache Cassandra

Tạo apt Apache Cassandra 3.11.x

echo "deb http://www.apache.org/dist/cassandra/debian 311x main" | sudo tee -a /etc/apt/sources.list.d/cassandra.sources.list

curl https://www.apache.org/dist/cassandra/KEYS | sudo apt-key add -

sudo apt-get update

Nếu bạn gặp lỗi khóa công khai GPG, hãy chạy các lệnh sau để thêm khóa công khai Apache Cassandra được đề cập, đó là A278B781FE4B2BDA trong trường hợp này:

sudo apt-key adv --keyserver pool.sks-keyservers.net --recv-key A278B781FE4B2BDA

sudo apt-get update

Sử dụng repo apt mới được thêm vào để cài đặt Apache Cassandra:

sudo apt-get install cassandra

* Step 4: Kiểm tra cài đặt Apache Cassandra

Bắt đầu daemon Apache Cassandra

sudo service cassandra start

Nếu bạn muốn làm cho Apache Cassandra tự động khởi động tại hệ thống khởi động, hãy chạy lệnh sau:

sudo update-rc.d cassandra defaults

Tiếp theo, sử dụng chương trình gật đầu để hiển thị trạng thái của Apache Cassandra trên nút hiện tại:

nodetool status

Đầu ra sẽ giống như sau:

Datacenter: datacenter1

=======================

Status=Up/Down

|/ State=Normal/Leaving/Joining/Moving

-- Address Load Tokens Owns (effective) Host ID Rack

UN 127.0.0.1 102.66 KiB 256 100.0% 23916cfd-892d-4898-857c-aff9efe2354a rack1

Bạn có thể sử dụng shell để tương tác với Apache Cassandra:

cqlsh localhost

Đầu ra sẽ tương tự như sau:

Connected to Test Cluster at localhost:9042.

[cqlsh 5.0.1 | Cassandra 3.11.2 | CQL spec 3.4.4 | Native protocol v4]

Use HELP for help.

cqlsh>