

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC BÁCH KHOA



**BÁO CÁO MÔN HỌC**  
**KỸ THUẬT DỮ LIỆU**  
**ĐỀ TÀI: CASSANDRA DATABASE**

Giảng Viên: Võ Thị Ngọc Châu  
Nhóm 11:  
2270194 - Nguyễn Ngọc Thắng

**TP. HỒ CHÍ MINH, 2022**

<b>1. Giới thiệu đề tài</b>	<b>3</b>
<b>2. Cơ sở lý thuyết</b>	<b>4</b>
2.1 Kiến trúc	4
2.2 Mô hình dữ liệu	6
2.3 Kiểu dữ liệu	8
2.4 Tool	10
<b>3. Bài toán áp dụng</b>	<b>11</b>
3.1 Reverse engineering	11
3.2 Di chuyển dữ liệu đến Cassandra	12
3.3. Phân tích những dữ liệu cần thiết đối với ứng dụng	16
3.4. Xây dựng ứng dụng	16
3.4.1 Giao diện Frontend	17
3.4.2 Backend	19
<b>4. Kết quả và đánh giá</b>	<b>21</b>
4.1 Kết quả:	21
4.2 Đánh giá	22
<b>5. Kết luận và hướng phát triển</b>	<b>23</b>
5.1 Kết luận	23
5.2 Hướng phát triển	23
<b>TÀI LIỆU THAM KHẢO</b>	<b>24</b>

## 1. Giới thiệu đề tài

Hiện nay, các dịch vụ trên Internet phải đối mặt với khối lượng thông tin, dữ liệu rất lớn và có mối liên hệ phức tạp với nhau. Hầu hết dữ liệu sẽ được lưu trữ phân tán trên nhiều máy chủ khác nhau để đảm bảo truy cập cho người dùng. Vì vậy, các hệ quản trị cơ sở dữ liệu quan hệ (RDBMS: Relational Database Management System) tỏ ra không còn phù hợp với các dịch vụ như thế này nữa. Người ta bắt đầu nghĩ tới việc phát triển các DBMS (Database Management System) mới phù hợp để quản lý các khối lượng dữ liệu phân tán này và đó chính là NoSQL.

Trong RDBMS, việc lựa chọn hệ quản trị cơ sở dữ liệu là rất dễ dàng bởi vì tất cả các hệ quản trị cơ sở dữ liệu như MySQL, Oracle, MS SQL,... sử dụng cùng loại giải pháp hướng tới các thuộc tính ACID (Atomicity, Consistency, Isolation, và Durability). Khi sử dụng NoSQL, quyết định trở nên khó khăn vì cơ sở dữ liệu NoSQL cung cấp các giải pháp khác nhau và cần hiểu rõ từng giải pháp để chọn ra giải pháp phù hợp nhất với yêu cầu ứng dụng/ hệ thống của mình.

Cassandra database là một cơ sở dữ liệu NoSQL mã nguồn mở và phân tán. Cassandra được phát triển bởi facebook vào năm 2007, và hiện tại thì nó đang được phát triển bởi Apache. Cassandra là sự kết hợp của Google's Bigtable và Amazon's Dynamo. Cassandra có nhiều node trên nhiều cụm, với kết nối ngang hàng. Dữ liệu của cassandra được lưu trong bảng và được truy vấn bởi CQL (Cassandra Query Language)

Để tìm hiểu rõ hơn về các tính năng và khả năng tương thích của Cassandra DB với các công nghệ hiện đại, nhóm sẽ sử dụng Cassandra DB trong việc lưu trữ và truy xuất dữ liệu. Nhóm sử dụng bộ cơ sở dữ liệu về thông tin thu gom (Purchasing) để xây dựng nên một ứng dụng hỗ trợ lựa chọn thu mua và quản lý đơn hàng thu mua dựa trên nền tảng ứng dụng web.

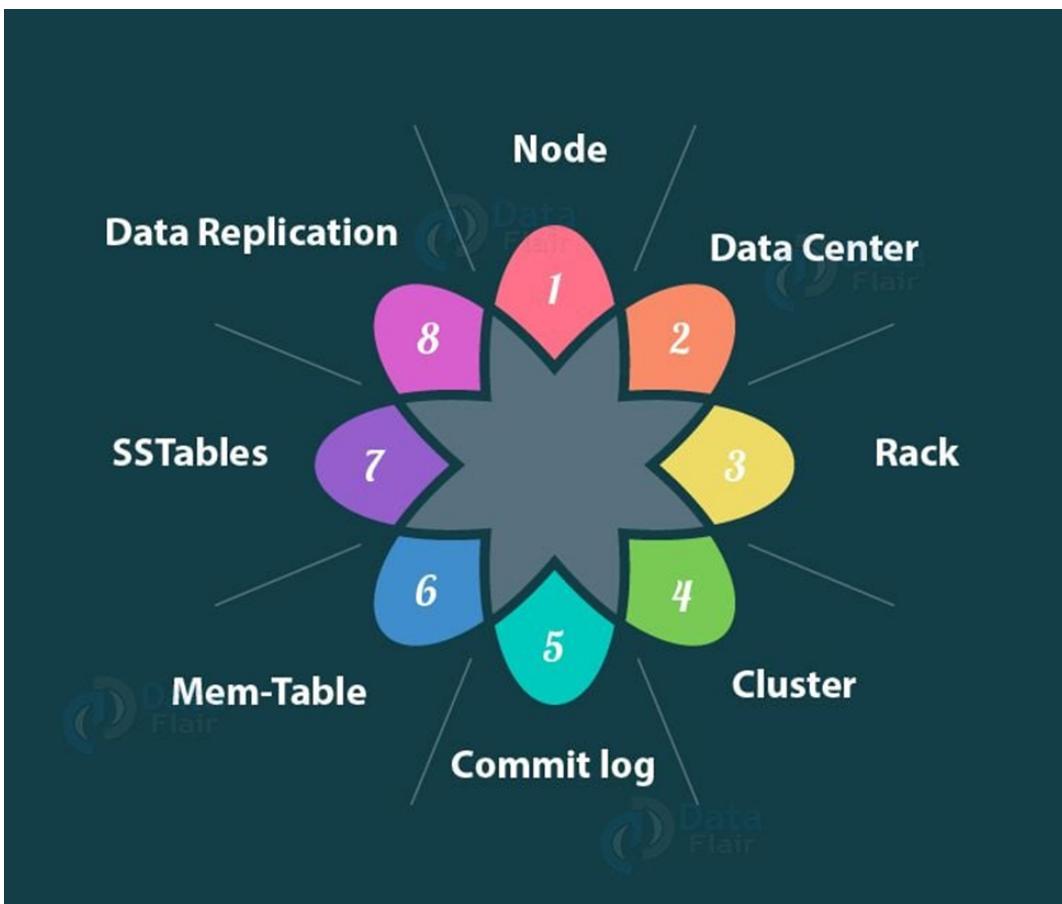
## 2. Cơ sở lý thuyết

### 2.1 Kiến trúc

Về kiến trúc của Cassandra Database thì nó có một số thành phần chính như:

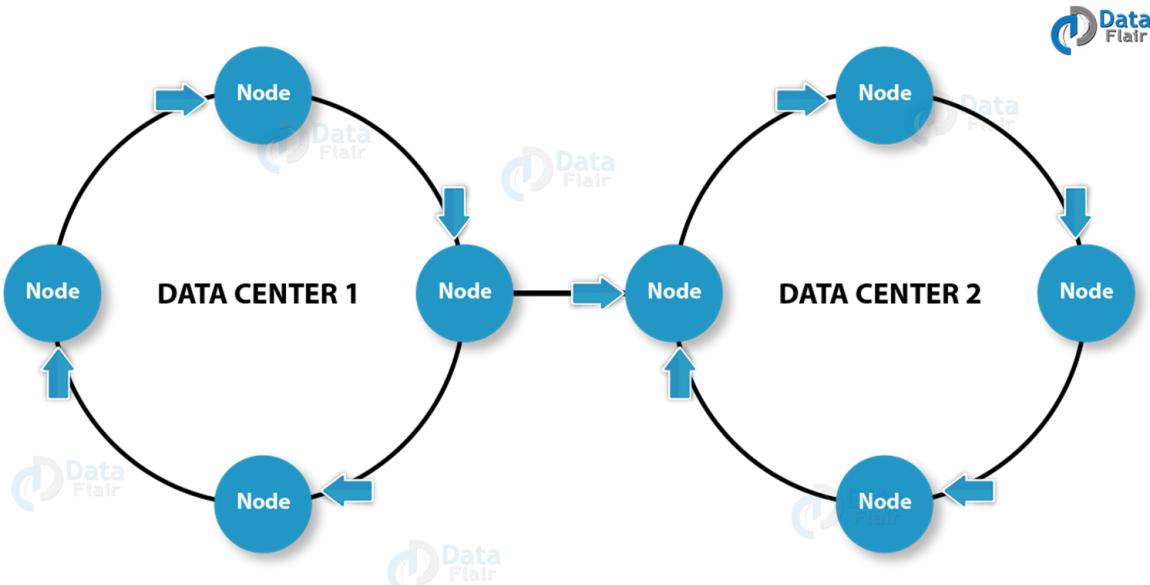
- Node
- DataCenter:
- Rack
- Cluster
- Commit Log
- Mem-Table
- SSTables
- Data Replication

Trong đó, node là thành phần cơ bản trong Apache Cassandra. Nó là nơi thực sự dữ liệu được lưu trữ.



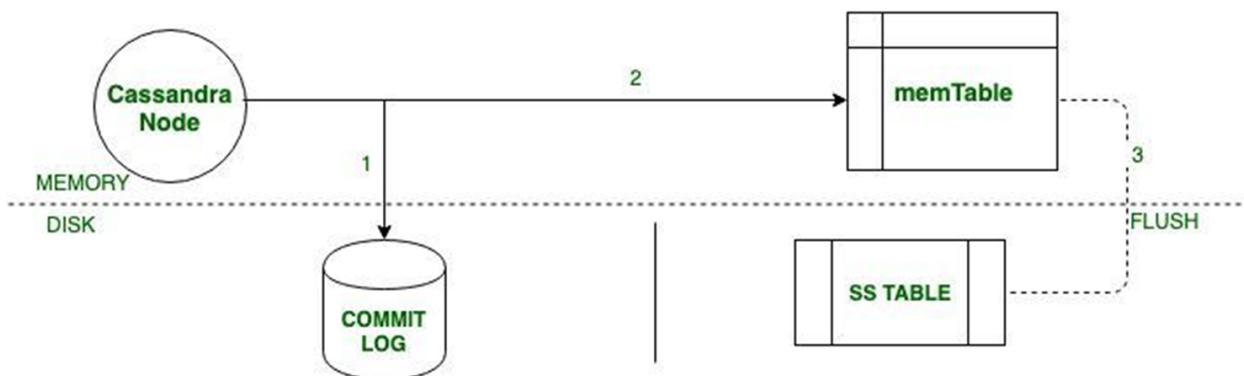
Hình 1: các thành phần chính trong kiến trúc của Cassandra DB

Hình bên dưới minh họa một cụm (Cluster). Chúng ta có thể thấy data center được tạo thành từ nhiều node và nhiều data center thì tạo thành cluster



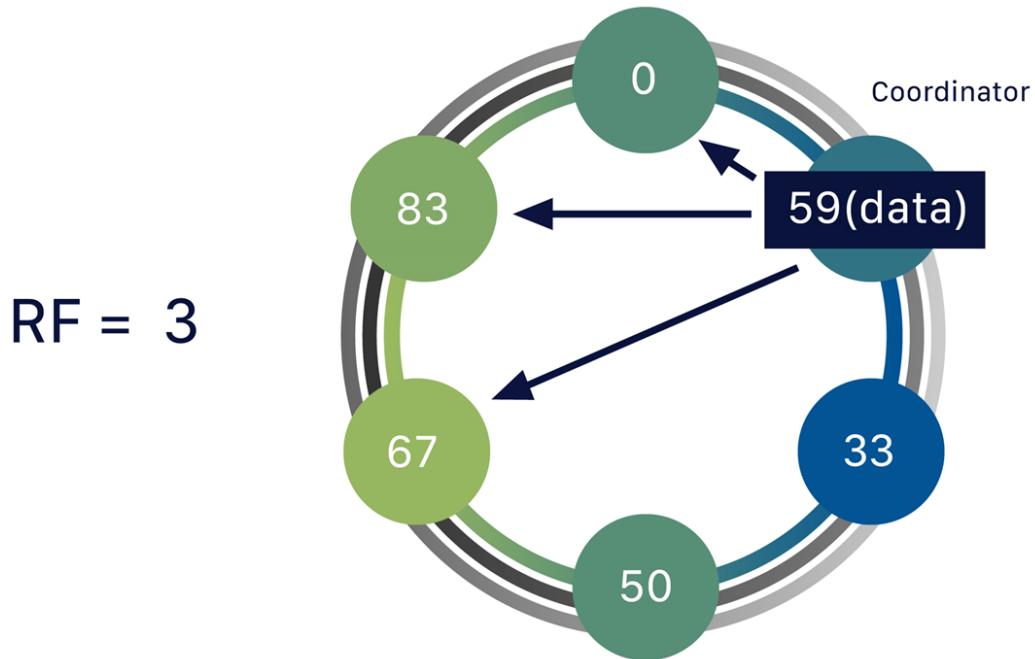
Hình 2: Minh họa Cluster

- Công cụ lưu trữ:
- commit log: là điểm nhập đầu tiên trong khi ghi vào đĩa hoặc memTable. Mục đích của commit log trong apache Cassandra là để giải quyết các vấn đề đồng bộ hóa máy chủ nếu một nút dữ liệu gặp sự cố.
- mem-table: sau khi data được ghi vào commit log, thì sau đó, data sẽ được lưu trong mem-table. dữ liệu được ghi trong mem-table tạm thời.
- sstable: khi mem-table đạt đến một ngưỡng nhất định thì dữ liệu sẽ được chuyển vào tệp đĩa sstable.



Hình 3: Minh họa Storage Engine

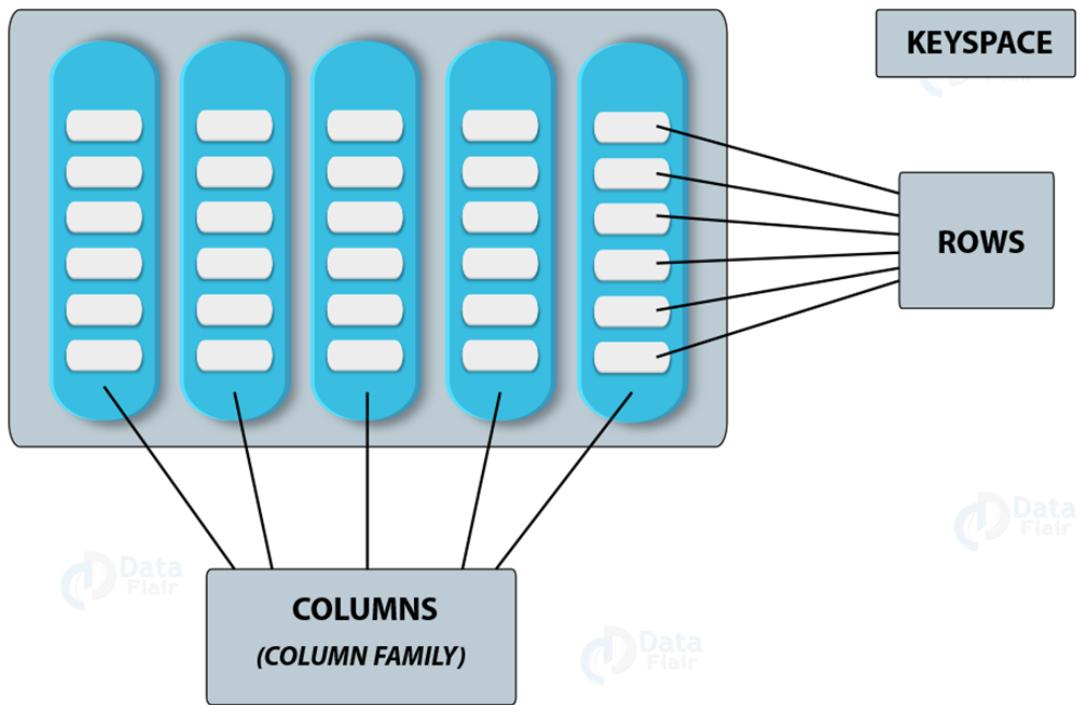
- Chiến lược sao chép dữ liệu:
- Về cơ bản, nó được sử dụng để sao lưu nhằm đảm bảo không có điểm lỗi nào. Trong chiến lược này, cassandra sử dụng tính năng sao chép để đạt được tính khả dụng và độ bền cao.
- Có 2 chiến lược sao chép: là chiến lược đơn giản và chiến lược cấu trúc liên mạng. Hình bên dưới minh họa một chiến lược đơn giản. Trong chiến lược này, thì nó cho phép một số nguyên RF (hệ số nhân bản) được xác định. ví dụ RF là 3, thì sẽ có 3 nút khác nhau sẽ lưu trữ bản sao.



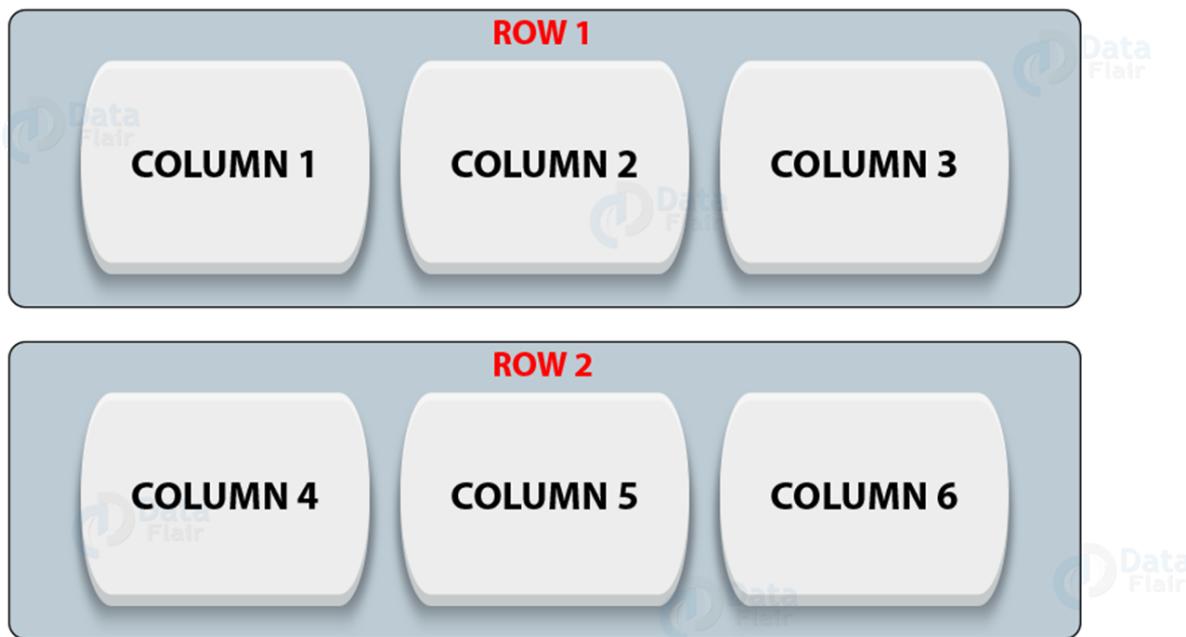
Hình 4: Minh họa chiến lược sao chép dữ liệu

## 2.2 Mô hình dữ liệu

Trong Mô hình dữ liệu Cassandra, Cassandra Keyspace là một vùng chứa dữ liệu. Trong Keyspace sẽ chứa các họ cột (column family). Họ cột trong Cassandra là một tập hợp các hàng chứa các cột được sắp xếp theo thứ tự. Chúng đại diện cho một cấu trúc của dữ liệu được lưu trữ. Tham chiếu với cơ sở dữ liệu quan hệ thì column family tương tự như bảng của cơ sở dữ liệu quan hệ.



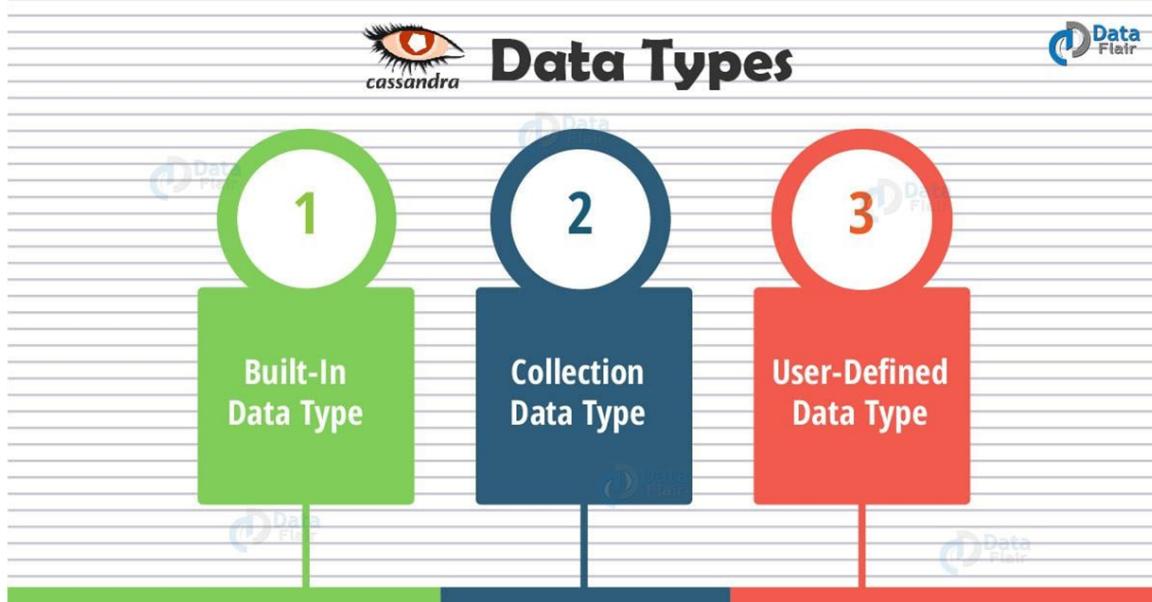
Hình 5: Minh họa Keyspace trong mô hình dữ liệu Cassandra



Hình 6: Minh họa Row trong mô hình dữ liệu Cassandra

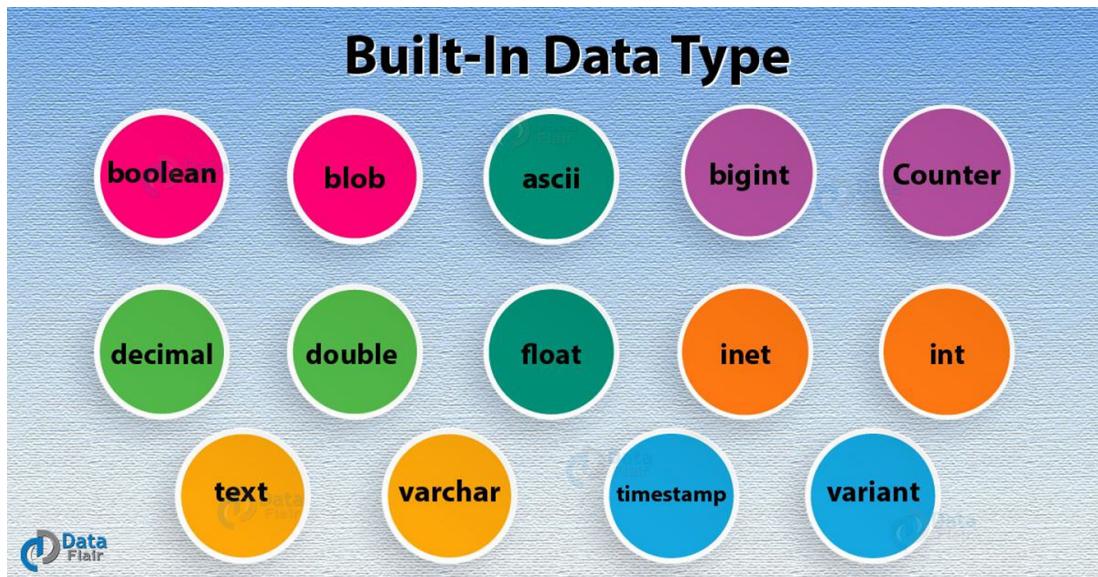
## 2.3 Kiểu dữ liệu

Kiểu dữ liệu của Cassandra được phân thành 3 nhóm, đó là kiểu dữ liệu Built-in, kiểu dữ liệu Collection và kiểu dữ liệu do người dùng định nghĩa.



Hình 7: Minh họa 3 nhóm kiểu dữ liệu của Cassandra

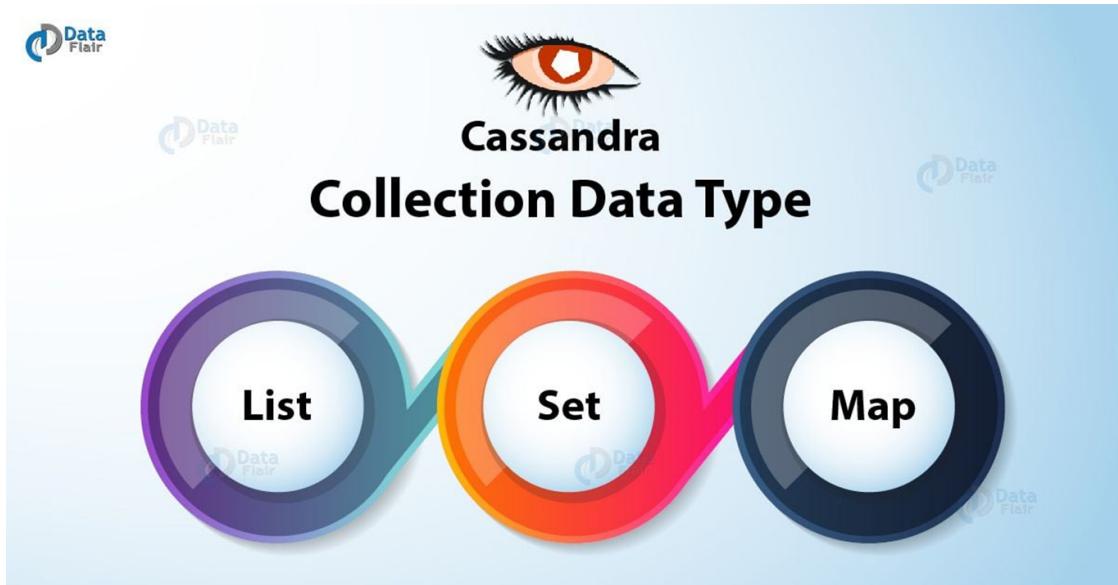
Built-In Data Type: đây là kiểu dữ liệu cơ bản, đã được xác định trước trong Cassandra. Bao gồm các loại như bool, ascii, bigint (số nguyên dài có dấu, 64 bit), inet (địa chỉ ip), text, timestamp,...



Hình 8: Các kiểu dữ liệu trong nhóm Built-In Data Type

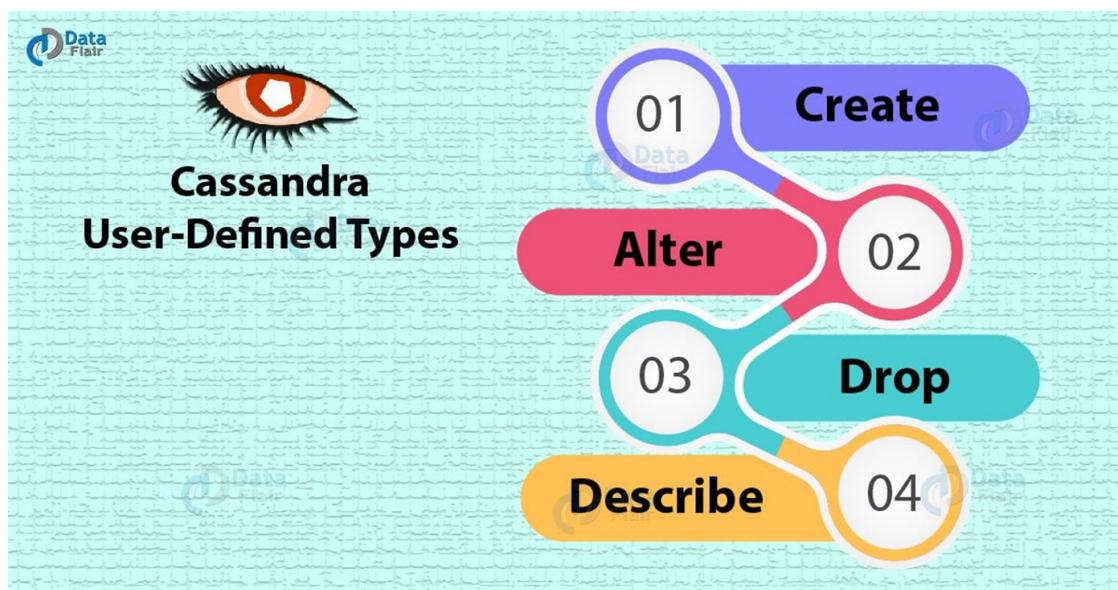
Nhóm kiểu dữ liệu Collection gồm có 3 loại: List, Set, Map.

- List: đại diện cho một tập hợp một hoặc nhiều phần tử trong một bảng
- Set: đại diện cho một tập hợp một hoặc nhiều phần tử được sắp xếp trong một bảng
- Map: đại diện cho tập hợp các cặp khóa-giá trị.



Hình 9: Các kiểu dữ liệu trong nhóm Collection Data Type

Kiểu dữ liệu do người dùng xác định: CQL cho phép người dùng tạo ra kiểu dữ liệu theo yêu cầu của riêng họ. Người dùng có thể tạo, thay thế, xóa, mô tả kiểu dữ liệu đó.



Hình 10: Minh họa kiểu dữ liệu do người dùng tự định nghĩa

## 2.4 Tool

Dưới đây là một số công cụ được cung cấp bởi Cassandra:

- Cqlsh: là giao diện dòng lệnh để tương tác với Cassandra bằng CQL
- Nodetool: cho phép chúng ta giới hạn các vấn đề từ cluster đến các node. Cung cấp các thông tin chi tiết về trạng thái của cassandra
- SSTable Tool: Bộ công cụ để phân tích cú pháp, tạo và thực hiện các nội dung thú vị với Cassandra 3.x SSTables. (đang phát triển && chưa ổn định)
- Cassandra Stress: được sử dụng để đo và kiểm tra tải của cụm Cassandra

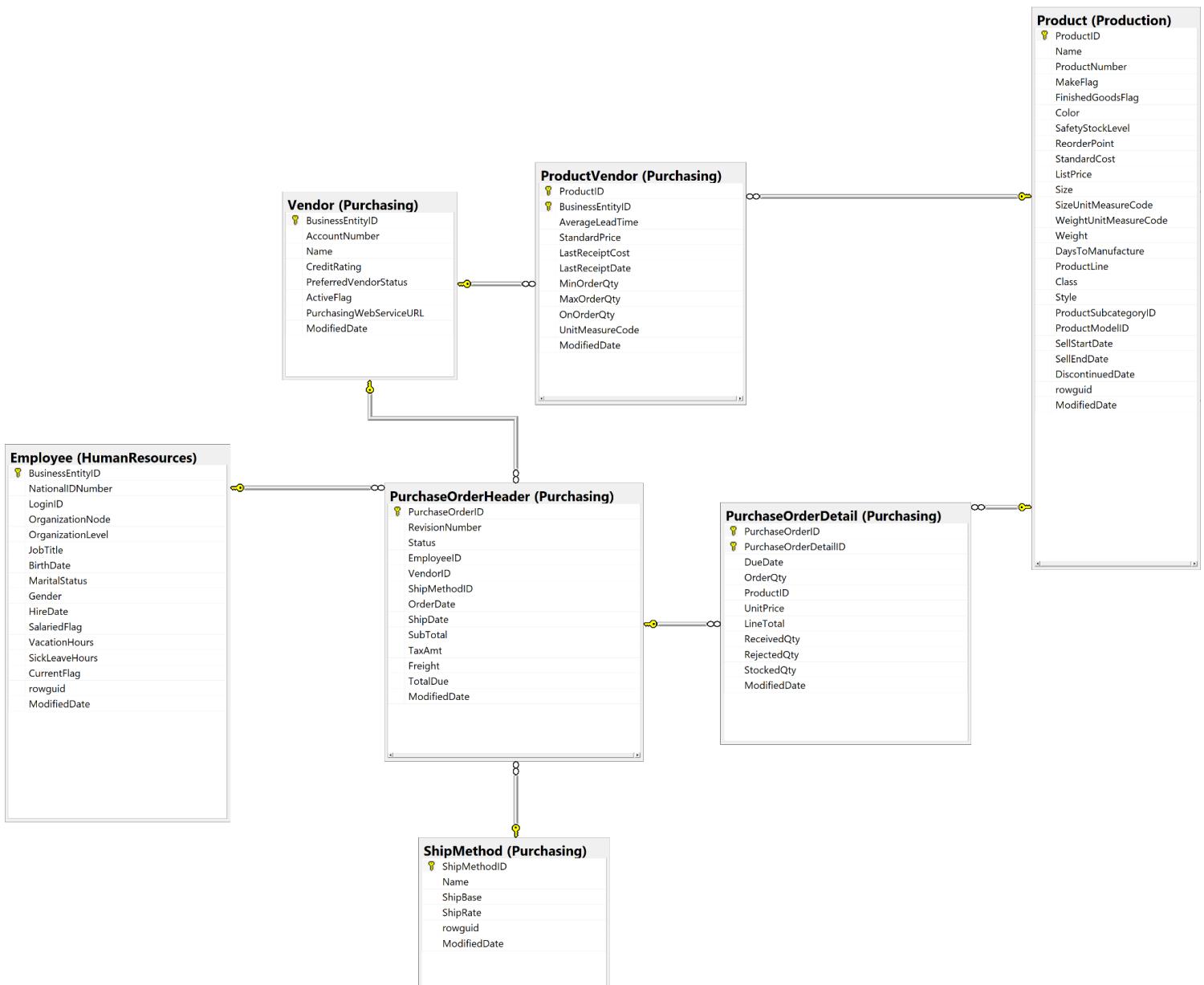
### **3. Bài toán áp dụng**

Để áp dụng Cassandra vào bài toán thực tế, nhóm đã xây dựng mô hình ứng dụng web, hỗ trợ lựa chọn thu mua và quản lý đơn hàng thu mua. Thu mua là một công việc để đảm bảo đủ nguyên vật liệu để phục vụ cho việc duy trì sản xuất và kinh doanh của công ty. Ứng dụng được xây dựng dành riêng cho những nhân viên làm công việc thu mua. Ứng dụng cho phép các nhân viên thu mua truy xuất các thông tin về nhà cung cấp của sản phẩm (vendor), lịch sử các đơn hàng mua sản phẩm để từ đó có thể lựa chọn nhà cung cấp phù hợp. Ngoài ra, ứng dụng cũng hỗ trợ nhân viên thu mua truy xuất, quản lý các đơn hàng thu mua.

Hệ thống mà nhóm xây dựng là một hệ thống có hệ cơ sở dữ liệu là Cassandra database, server backend sử dụng NodeJs, giao diện UI sử dụng ReactJs để làm ứng dụng web.

#### **3.1 Reverse engineering**

Nhóm được giao AdventureWorks, một dữ liệu mẫu được phát hành bởi MS SQL Server. Do đó, để mở được tập dữ liệu này, chúng ta sẽ sử dụng công cụ của Microsoft, đó là Microsoft SQL Server Management Studio. Nhờ công cụ này, ta dễ dàng xem các dữ liệu, các bảng, cột, các khóa, mối quan hệ của dữ liệu, và xem được ER Diagram của chúng.



Hình 11: ER Diagram của database về Purchasing

### 3.2 Di chuyển dữ liệu đến Cassandra

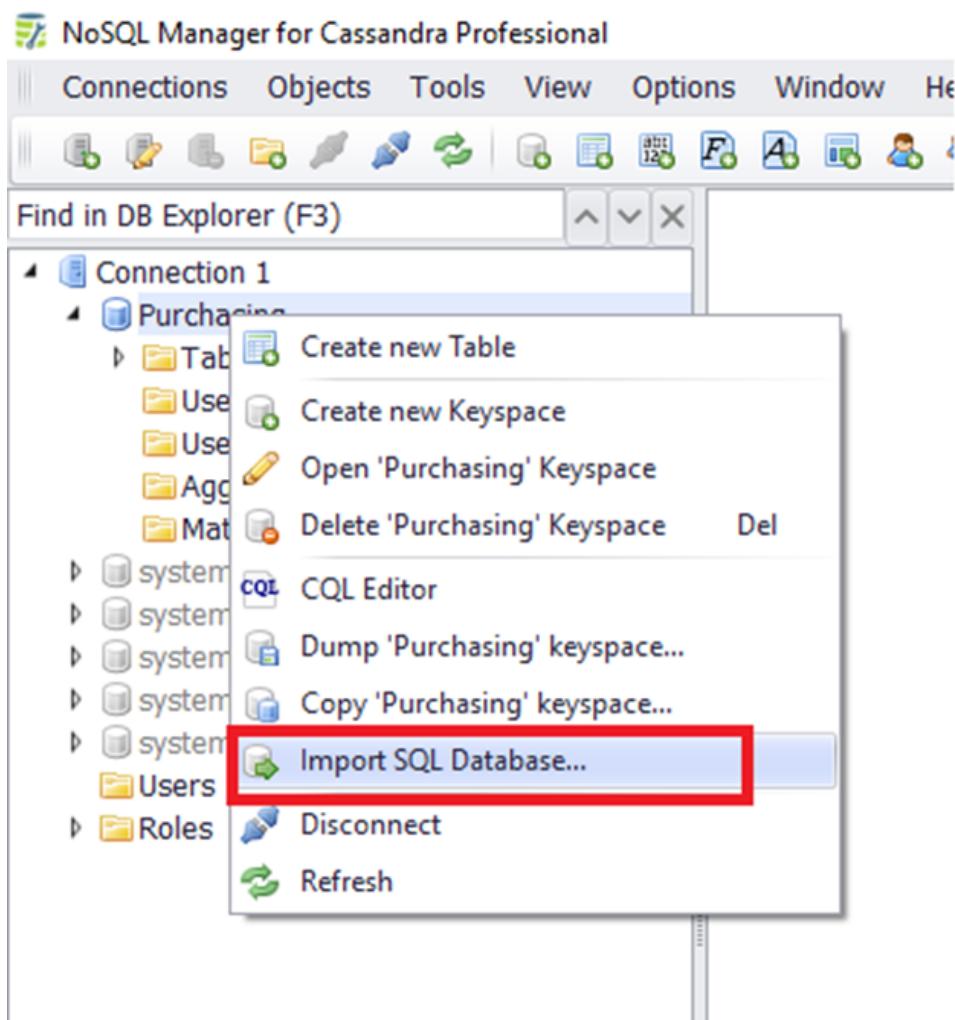
Để di chuyển dữ liệu đến Cassandra, chúng ta có nhiều cách khác nhau:

- Từ MS SQL Server, chúng ta có thể xuất dữ liệu sang CSV. Sau đó từ tool cqlsh, chúng ta có thể đọc dữ liệu từ file csv và ghi vào Cassandra database.
- Có thể viết một script để đọc dữ liệu trực tiếp từ MS SQL Server và ghi vào Cassandra.

- Sử dụng một công cụ GUI quản lý Cassandra, cho phép hỗ trợ import data trực tiếp từ MS SQL Server.

Qua những cách trên, nhóm nhận thấy cách sử dụng một công cụ quản lý Cassandra có hỗ trợ import data trực tiếp từ MS SQL Server sẽ là cách đơn giản, thuận tiện và nhanh chóng nhất. Do đó, nhóm quyết định sử dụng công cụ NoSQL Manager For Cassandra để di chuyển dữ liệu đến Cassandra.

Sau khi tạo Keyspace cho Cassandra database, chúng ta nhấn chuột phải trên KeySpace và chọn Import SQL Database:



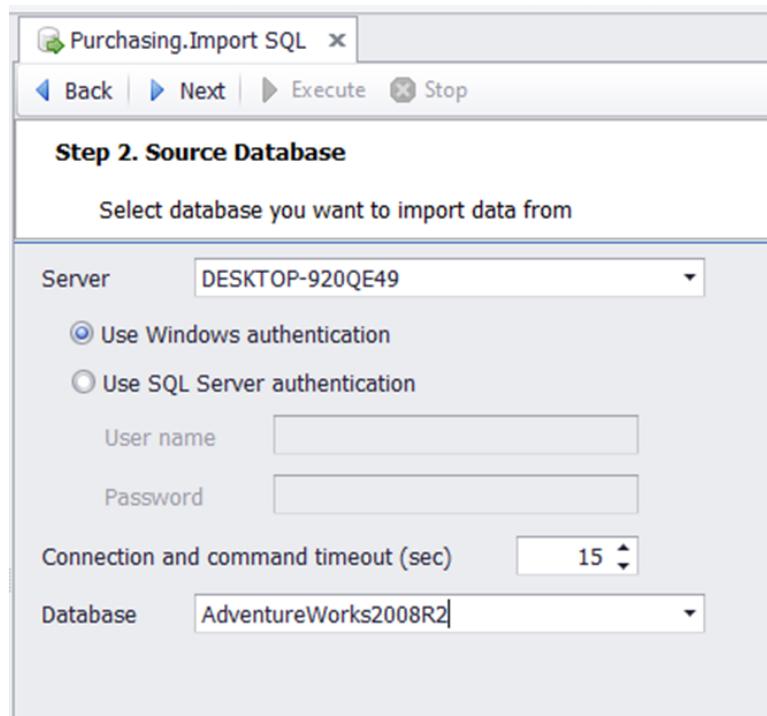
Hình 12: Import SQL Database

Tiếp theo, chúng ta chọn loại cơ sở dữ liệu SQL:



Hình 13: Chọn loại database

Tiếp theo, chúng ta chọn database muốn import, ở đây, chúng ta sẽ chọn AdventureWorks2008R2



Hình 14: Chọn Database

Tiếp theo, chúng ta chọn những bảng muốn di chuyển đến Cassandra:

Select tables you want to import

Available	Selected
Production.vProductAndDescription	Purchasing.ProductVendor
Production.vProductModelCatalogDescription	Purchasing.PurchaseOrderDetail
Production.vProductModelInstructions	Purchasing.PurchaseOrderHeader
Production.WorkOrder	Purchasing.ShipMethod
Production.WorkOrderRouting	Purchasing.Vendor
Sales.CountryRegionCurrency	Purchasing.vVendorWithAddresses
Sales.CreditCard	Purchasing.vVendorWithContacts
Sales.Currency	
Sales.CurrencyRate	
Sales.Customer	
Sales.PersonCreditCard	
Sales.SalesOrderDetail	
Sales.SalesOrderHeader	
Sales.SalesOrderHeaderSalesReason	
Sales.SalesPerson	
Sales.SalesPersonQuotaHistory	
Sales.SalesReason	
Sales.SalesTaxRate	
Sales.SalesTerritory	
Sales.SalesTerritoryHistory	
Sales.ShoppingCartItem	
Sales.SpecialOffer	
Sales.SpecialOfferProduct	
Sales.Store	
Sales.vIndividualCustomer	

Hình 15: chọn Table

Cuối cùng, chúng ta nhấn Execute để bắt đầu import dữ liệu. Hình dưới thể hiện dữ liệu trong Cassandra sau khi đã import.

BusinessEntityID	AccountNumber	ActiveFlag	CreditRating	ModifiedDate	Name	PreferredVendorStatus
1584	TREYRE0001	✓	3	2006-02-25T00:00:00.000+0000	Trey Research	✓
1580	LITWARE0001	✓	1	2005-05-26T00:00:00.000+0000	Litware, Inc.	✓
1548	CONSUMER0001	✓	3	2006-01-24T00:00:00.000+0000	Consumer Cycles	✓
1522	FIRSTNA0001	✓	1	2006-02-25T00:00:00.000+0000	First National Sport Co.	✓
1618	METROSP0001	✓	1	2006-03-01T00:00:00.000+0000	Metro Sport Equipment	✓
1588	SIGNATUR0001	✓	2	2006-03-05T00:00:00.000+0000	Signature Cycles	✓
1638	INLINEA0001	✓	1	2006-02-25T00:00:00.000+0000	Inline Accessories	✓
1672	EXPERTB0001	✓	1	2006-02-25T00:00:00.000+0000	Expert Bike Co	✓
1646	ELECTRON0002	✓	1	2006-02-17T00:00:00.000+0000	Electronic Bike Repair & Supplies	✓
1630	INDIANA0001	✓	1	2006-03-05T00:00:00.000+0000	Indiana Bicycle Center	✓
1678	PROSE0001	✗	4	2005-06-09T00:00:00.000+0000	Proseware, Inc.	✗
1540	BERGERON0001	✓	1	2006-01-23T00:00:00.000+0000	Bergeron Off-Roads	✓
1604	BIKESAT0001	✓	1	2006-01-23T00:00:00.000+0000	Bike Satellite Inc.	✗
1632	SPORTFA0001	✓	1	2006-03-05T00:00:00.000+0000	Sport Fan Co.	✓
1492	AUSTRALI0001	✓	1	2006-01-23T00:00:00.000+0000	Australia Bike Retailer	✓
1674	VARSITY0001	✓	1	2006-03-21T00:00:00.000+0000	Varsity Sport Co.	✓
1640	LEGENDC0001	✓	1	2006-03-05T00:00:00.000+0000	Legend Cycles	✓
1692	CARLSON0001	✓	2	2006-01-24T00:00:00.000+0000	Carlson Specialties	✓
1624	COMPETIT0001	✓	1	2006-01-24T00:00:00.000+0000	Competition Bike Training Systems	✓

Hình 16: Dữ liệu trong Cassandra sau khi import

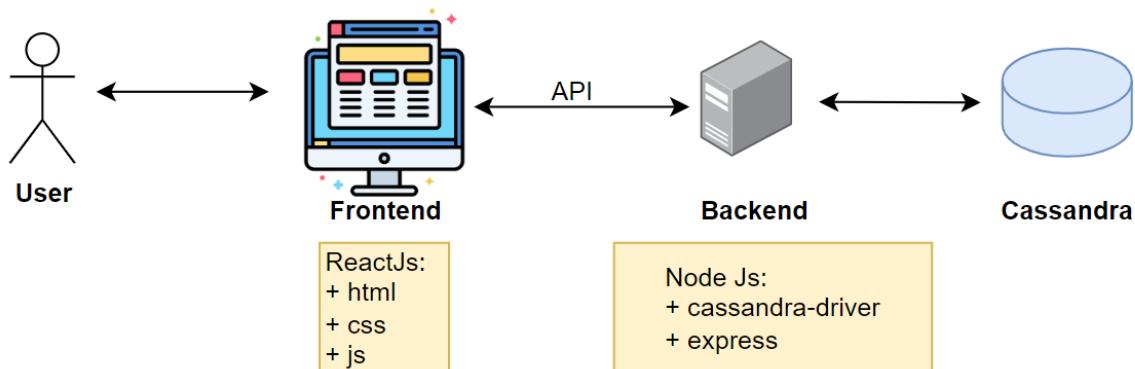
### 3.3. Phân tích những dữ liệu cần thiết đối với ứng dụng

Bởi vì ứng dụng cho phép truy xuất các thông tin về nhà cung cấp của sản phẩm (vendor), lịch sử các đơn hàng mua sản phẩm và các đơn hàng mà nhân viên đã thực hiện. Vì vậy, ứng dụng cần dữ liệu trong những bảng sau để đáp ứng yêu cầu của ứng dụng:

- Production\_Product
- Purchasing\_ProductVendor
- Purchasing\_PurchaseOrderDetail
- Purchasing\_PurchaseOrderHeader
- Purchasing\_ShipMethod
- Purchasing\_Vendor

### 3.4. Xây dựng ứng dụng

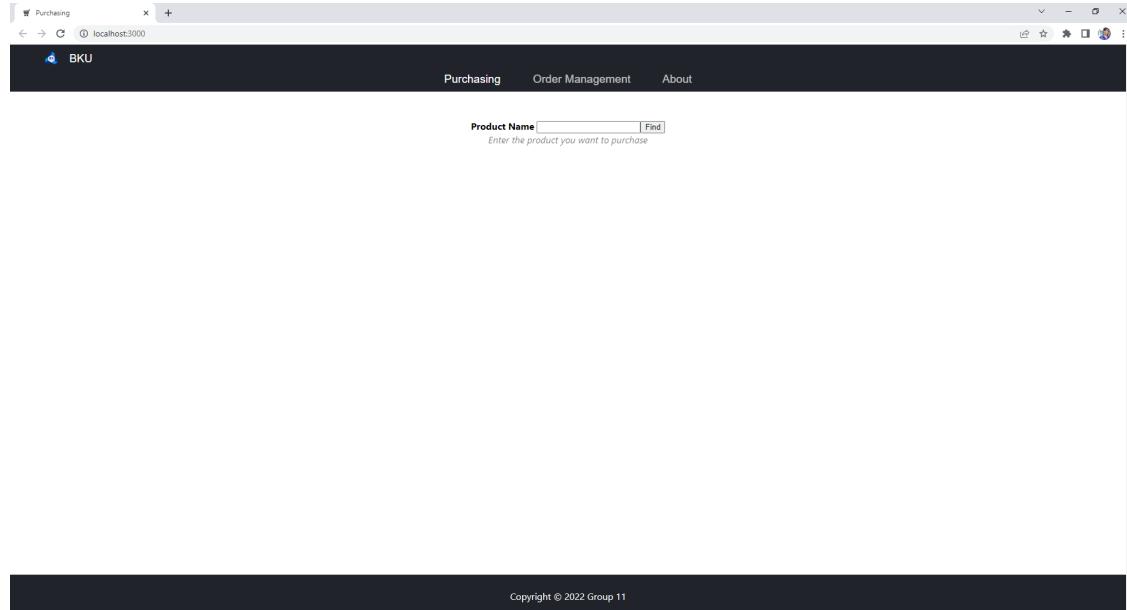
Ứng dụng nhóm xây dựng là một ứng dụng web có thể truy xuất các thông tin về thu mua. Hình dưới mô tả mô hình của ứng dụng. Backend phụ trách việc truy xuất dữ liệu và làm việc trực tiếp với Cassandra DB thông qua thư viện cassandra-driver của Node Js. Bên cạnh đó, Backend sẽ cung cấp các API để giao tiếp với phía Frontend thông qua thư viện express. Phía frontend sẽ tương tác với người dùng, và lấy data dưới dạng json thông qua các API do backend cung cấp.



Hình 17: Mô hình ứng dụng

### 3.4.1 Giao diện Frontend

Giao diện của ứng dụng sẽ có 3 page chính: Purchasing, Order Management, About



Hình 18: Giao diện ứng dụng

Ở page Purchasing sẽ cung cấp cho người thu mua danh sách những nhà cung cấp sản phẩm mà họ cần thu mua, cung cấp lịch sử đơn hàng của sản phẩm đó.

The screenshot shows the "Purchasing" page with a search bar at the top. Below the search bar is a section titled "Vendor List" containing a table with columns: Vendor Name, preferred, Average Lead Time, MaxOrderQty, MinOrderQty, StandardPrice, and Last Receipt Cost. The table data is as follows:

Vendor Name	preferred	Average Lead Time	MaxOrderQty	MinOrderQty	StandardPrice	Last Receipt Cost
Ready Rentals	OK	19	5	1	39.4100	41.3805
Leaf River Terrain	NOK	16	5	1	39.4100	41.3805
Australia Bike Retailer	OK	17	5	1	39.2600	41.2230

Below the vendor list is a section titled "Product Purchase History" containing a table with columns: Vendor Name, OrderQty, ReceivedQty, RejectedQty, StockedQty, and UnitPrice. The table data is as follows:

Vendor Name	OrderQty	ReceivedQty	RejectedQty	StockedQty	UnitPrice
Leaf River Terrain	3	3.00	0.00	3.00	41.3805
Australia Bike Retailer	3	3.00	0.00	3.00	41.2230
Leaf River Terrain	3	3.00	0.00	3.00	41.3805
Ready Rentals	3	3.00	0.00	3.00	41.3805
Ready Rentals	3	3.00	3.00	0.00	41.3805
Australia Bike Retailer	3	3.00	0.00	3.00	41.2230
Australia Bike Retailer	3	3.00	0.00	3.00	41.2230
Australia Bike Retailer	3	3.00	0.00	3.00	41.2230
Ready Rentals	3	3.00	0.00	3.00	41.3805
Australia Bike Retailer	3	3.00	0.00	3.00	41.2230
Ready Rentals	3	3.00	0.00	3.00	41.3805

Hình 19: Giao diện trang Purchasing

Ở page Order Management, sẽ cung cấp danh sách những đơn hàng của người nhân viên, giúp cho nhân viên có thể kiểm tra tình trạng, quản lý đơn hàng của mình.

The screenshot shows a web browser window titled "Purchasing" with the URL "localhost:3000/OrderManagement". The page has a dark header with the logo "BKU" and navigation links for "Purchasing", "Order Management", and "About". Below the header is a search bar with the placeholder "Enter the employee Id you want to find". The main content area is titled "Order List" and displays a table of purchase orders. The table columns are: Purchase Order ID, Product Name, Vendor Name, Ship Method, Ship Date, and Status. The data in the table is as follows:

Purchase Order ID	Product Name	Vendor Name	Ship Method	Ship Date	Status
3372	HL Road Tire	Victory Bikes	OVERNIGHT J-FAST	Tue Jul 29 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
2302	Keyed Washer	Wide World Importers	CARGO TRANSPORT 5	Wed Apr 23 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
2062	ML Mountain Pedal	Crowley Sport	CARGO TRANSPORT 5	Wed Apr 02 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
660	LL Road Tire	Signature Cycles	CARGO TRANSPORT 5	Tue Oct 02 2007 07:00:00 GMT+0700 (Indochina Time)	Complete
878	LL Mountain Pedal	Greenwood Athletic Company	CARGO TRANSPORT 5	Sun Oct 21 2007 07:00:00 GMT+0700 (Indochina Time)	Complete
4010	Racing Socks, L	Jeff's Sporting Goods	OVERSEAS - DELUXE	Sat Jan 05 2008 07:00:00 GMT+0700 (Indochina Time)	Approved
1052	Metal Tread Plate	Custom Frames, Inc.	XRQ - TRUCK GROUND	Sat Nov 24 2007 07:00:00 GMT+0700 (Indochina Time)	Complete
2492	Thin-Jam Lock Nut 5	Ready Rentals	CARGO TRANSPORT 5	Tue May 13 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
2932	Hex Nut 9	Cruger Bike Company	OVERNIGHT J-FAST	Mon Jun 23 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
2442	Chainring Nut	Bike Satellite Inc.	OVERNIGHT J-FAST	Wed May 07 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
3472	Crown Race	Business Equipment Center	CARGO TRANSPORT 5	Tue Aug 05 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
3712	LL Touring Seat/Saddle	Chicago City Saddles	OVERNIGHT J-FAST	Sun Aug 24 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
3662	Paint - Red	Trey Research	CARGO TRANSPORT 5	Wed Aug 20 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
2702	LL Mountain Seat/Saddle	First Rate Bicycles	ZY - EXPRESS	Mon Jun 02 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
120	HL Mountain Seat/Saddle	Hybrid Bicycle Center	OVERNIGHT J-FAST	Wed Apr 19 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
1902	HL Nipple	Consumer Cycles	ZY - EXPRESS	Tue Mar 18 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
1942	Flat Washer 8	Speed Corporation	CARGO TRANSPORT 5	Sat Mar 22 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
3092	Keyed Washer	Wide World Importers	CARGO TRANSPORT 5	Sun Jul 06 2008 07:00:00 GMT+0700 (Indochina Time)	Complete
140			OVERNIGHT J-FAST	Mon May 22 2006 07:00:00 GMT+0700 (Indochina Time)	Complete

Hình 20: Giao diện trang Order Management

Cuối cùng là page About, nó mô tả mục đích của trang web.

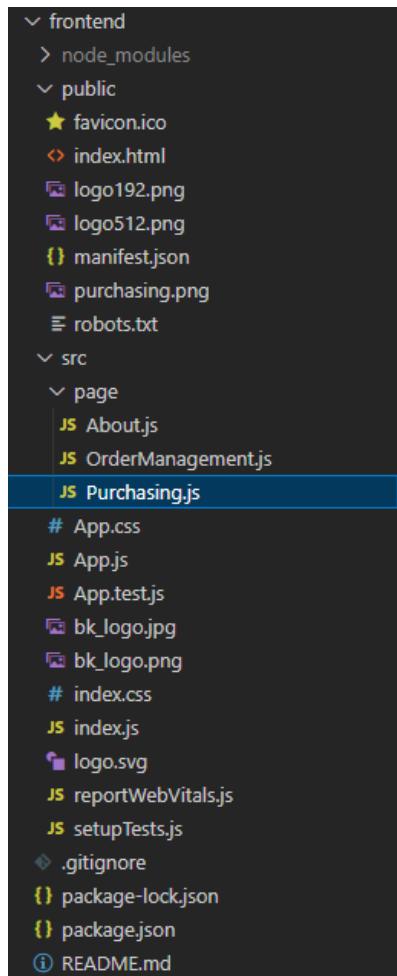
The screenshot shows a web browser window titled "Purchasing" with the URL "localhost:3000/About". The page has a dark header with the logo "BKU" and navigation links for "Purchasing", "Order Management", and "About". Below the header is a section titled "About Us" with the following text:

The Application to support purchasing:  
supplier selection  
purchase contract management

At the bottom of the page, there is a copyright notice: "Copyright © 2022 Group 11".

Hình 21: Giao diện trang About

Cấu trúc thư mục của Frontend:



Hình 22: Cấu trúc thư mục frontend

### 3.4.2 Backend

Backend cung cấp 3 API để frontend lấy dữ liệu, đó là:

- **api/order/product** : trả về thông tin danh sách các sản phẩm, gồm product Name, productId, kết quả trả về dưới dạng json.
- **api/purchasing/:productId** : trả về thông tin danh sách các nhà cung cấp (vendor) và lịch sử đơn hàng của sản phẩm có id là productId , kết quả trả về dưới dạng json.
- **api/order/:employeeId** : trả về thông tin danh sách các đơn hàng được tạo bởi nhân viên với id là employeeId

Để truy xuất dữ liệu từ Cassandra DB, chúng ta sử dụng thư viện cassandra-driver của NodeJs. Trước khi query data, chúng ta cần connect với Cassandra DB:

```
3 const cassandra = require('cassandra-driver');
4 const client = new cassandra.Client({ contactPoints: ['127.0.0.1'], localDataCenter: 'datacenter1' });
5 client.connect();
6
```

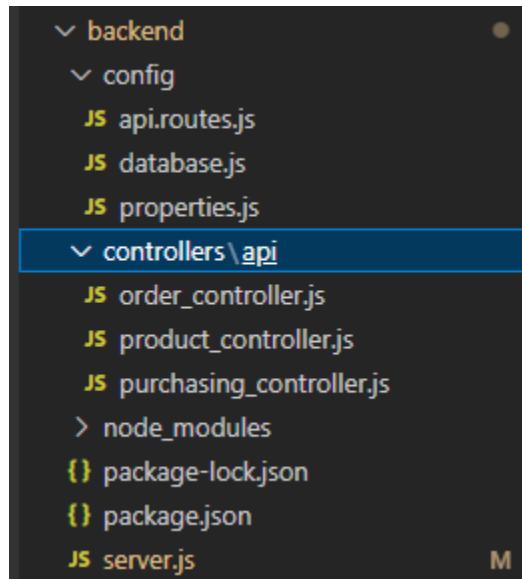
Hình 23: connect backend với Cassandra DB

Để query data, chúng ta chỉ cần truyền một chuỗi có định dạng CQL (Cassandra Query Language) vào hàm truy vấn. Hình bên dưới là một ví dụ truy vấn dữ liệu từ bảng “*Purchasing\_ProductVendor*” ở hàng có “*ProductId*” là “*productId*”

```
let productVendorQuery = "SELECT * FROM \"Purchasing\".\"Purchasing_ProductVendor\" WHERE \"ProductId\" = " + productId + " limit 1000 ALLOW FILTERING";
let productVendor = await client.execute(productVendorQuery);
```

Hình 24: minh họa một query Cassandra DB từ backend

Cấu trúc thư mục của backend



Hình 25: cấu trúc thư mục backend

## 4. Kết quả và đánh giá

### 4.1 Kết quả:

Nhóm đã hoàn thành ứng dụng, và đây là hình ảnh khi triển khai:

The screenshot shows a web browser window with three tabs: 'Purchasing', 'localhost:4000/api/purchasing/4', and 'Purchasing'. The main content area has a dark header with the text 'BKU' and navigation links for 'Purchasing', 'Order Management', and 'About'. Below the header is a search bar with the placeholder 'Enter the product you want to purchase' and a 'Find' button. The first section, 'Vendor List', displays a table with columns: Vendor Name, preferred, Average Lead Time, MaxOrderQty, MinOrderQty, StandardPrice, and Last Receipt Cost. It contains two rows: one for 'Advanced Bicycles' (OK, 17, 5, 1, 41.4100, 43.4805) and one for 'WestAmerica Bicycle Co.' (NOK, 19, 5, 1, 41.2600, 43.3230). The second section, 'Product Purchase History', displays a table with columns: Vendor Name, OrderQty, ReceivedQty, RejectedQty, StockedQty, and UnitPrice. This table lists multiple entries for Advanced Bicycles and WestAmerica Bicycle Co., all showing 3.00 for OrderQty, ReceivedQty, and StockedQty, and 43.4805 for UnitPrice.

Vendor Name	preferred	Average Lead Time	MaxOrderQty	MinOrderQty	StandardPrice	Last Receipt Cost
Advanced Bicycles	OK	17	5	1	41.4100	43.4805
WestAmerica Bicycle Co.	NOK	19	5	1	41.2600	43.3230

Vendor Name	OrderQty	ReceivedQty	RejectedQty	StockedQty	UnitPrice
Advanced Bicycles	3	3.00	0.00	3.00	43.4805
WestAmerica Bicycle Co.	3	3.00	0.00	3.00	43.3230
Advanced Bicycles	3	3.00	0.00	3.00	43.4805
Advanced Bicycles	3	3.00	0.00	3.00	43.4805
WestAmerica Bicycle Co.	3	3.00	0.00	3.00	43.3230
Advanced Bicycles	3	3.00	0.00	3.00	43.4805
Advanced Bicycles	3	3.00	0.00	3.00	43.4805
Advanced Bicycles	3	3.00	0.00	3.00	43.4805
Advanced Bicycles	3	3.00	0.00	3.00	43.4805
WestAmerica Bicycle Co.	3	3.00	0.00	3.00	43.3230
Advanced Bicycles	3	3.00	0.00	3.00	43.4805

Hình 26: Kết quả ứng dụng

## 4.2 Đánh giá

Thông qua quá trình tìm hiểu Cassandra và thực hiện ứng dụng sử dụng Cassandra DB, nhóm nhận thấy:

- **Ưu điểm:**
  - + Mã nguồn mở
  - + Kiến trúc ngang hàng
  - + Khả năng mở rộng, đàn hồi
  - + Khả năng chịu lỗi cao
  - + Hiệu suất cao
  - + Cassandra dễ học và sử dụng
- **Nhược điểm:**
  - + không hỗ trợ nhiều cho việc tính toán trên storage
  - + Vì là dữ liệu phân tán, dữ liệu được lan truyền trên nhiều máy, nên khi có 1 lỗi trong cơ sở dữ liệu thì lỗi này sẽ lan truyền ra toàn bộ các máy trên hệ thống

## **5. Kết luận và hướng phát triển**

### **5.1 Kết luận**

Kết quả đạt được cơ bản các yêu cầu của bài tập lớn:

- reverse engineering
- Di chuyển dữ liệu
- Xây dựng ứng dụng sử dụng Cassandra

Thông qua bài tập lớn này, nhóm có cơ hội tìm hiểu cassandra, và thực hành thực tế, giúp hiểu hơn các lý thuyết được giảng trên lớp.

Ngoài những kết quả cơ bản đã đạt được, nhóm vẫn còn một số mặt hạn chế:

- Chưa mở rộng thêm dữ liệu để đáp ứng với các vấn đề về big data.
- Ứng dụng chỉ mới thực hiện đọc dữ liệu mà chưa thực hiện việc ghi dữ liệu
- Chỉ sử dụng 1 node để chạy cassandra, nên chưa phát huy được hết khả năng của cassandra.

### **5.2 Hướng phát triển**

- Xây dựng thêm nhiều node để có thể phát huy được khả năng của cassandra
- Phát triển thêm nhiều chức năng ứng dụng hơn như là có thể tạo đơn hàng thu mua,...
- Mở rộng thêm dữ liệu

## TÀI LIỆU THAM KHẢO

<https://cassandra.apache.org/doc/latest/>

[https://en.wikipedia.org/wiki/Apache\\_Cassandra](https://en.wikipedia.org/wiki/Apache_Cassandra)

<https://docs.datastax.com/en/cassandra-oss/3.x/index.html>

<https://data-flair.training/blogs/apache-cassandra-tutorial/>

<https://www.geeksforgeeks.org/architecture-of-apache-cassandra>