

# Giới thiệu về NoSQL

## Mục tiêu:

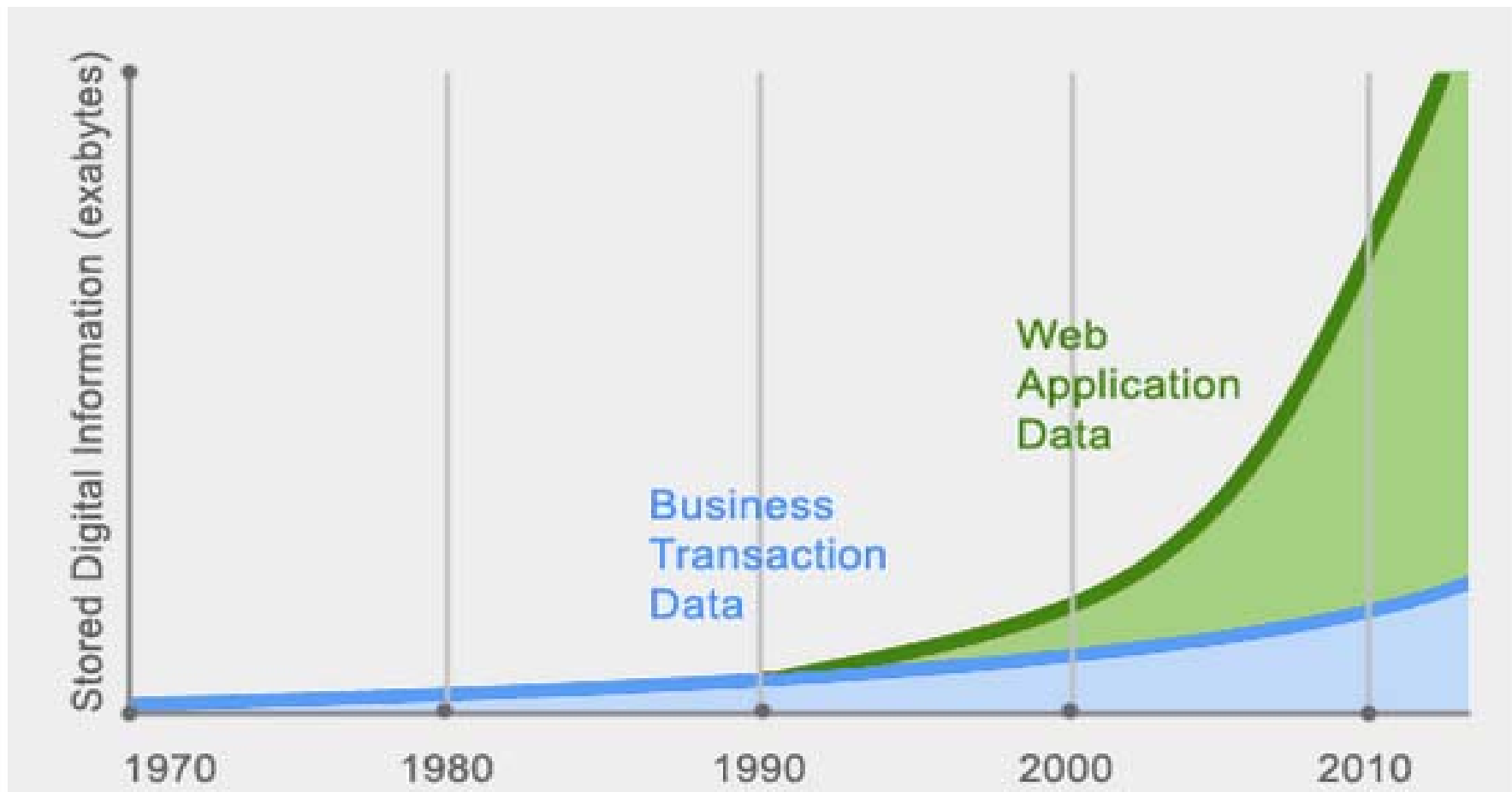
- Hiểu được sự ra đời của NoSQL
- Hiểu được các tiếp cận tiêu biểu của NoSQL để biểu diễn cơ sở dữ liệu



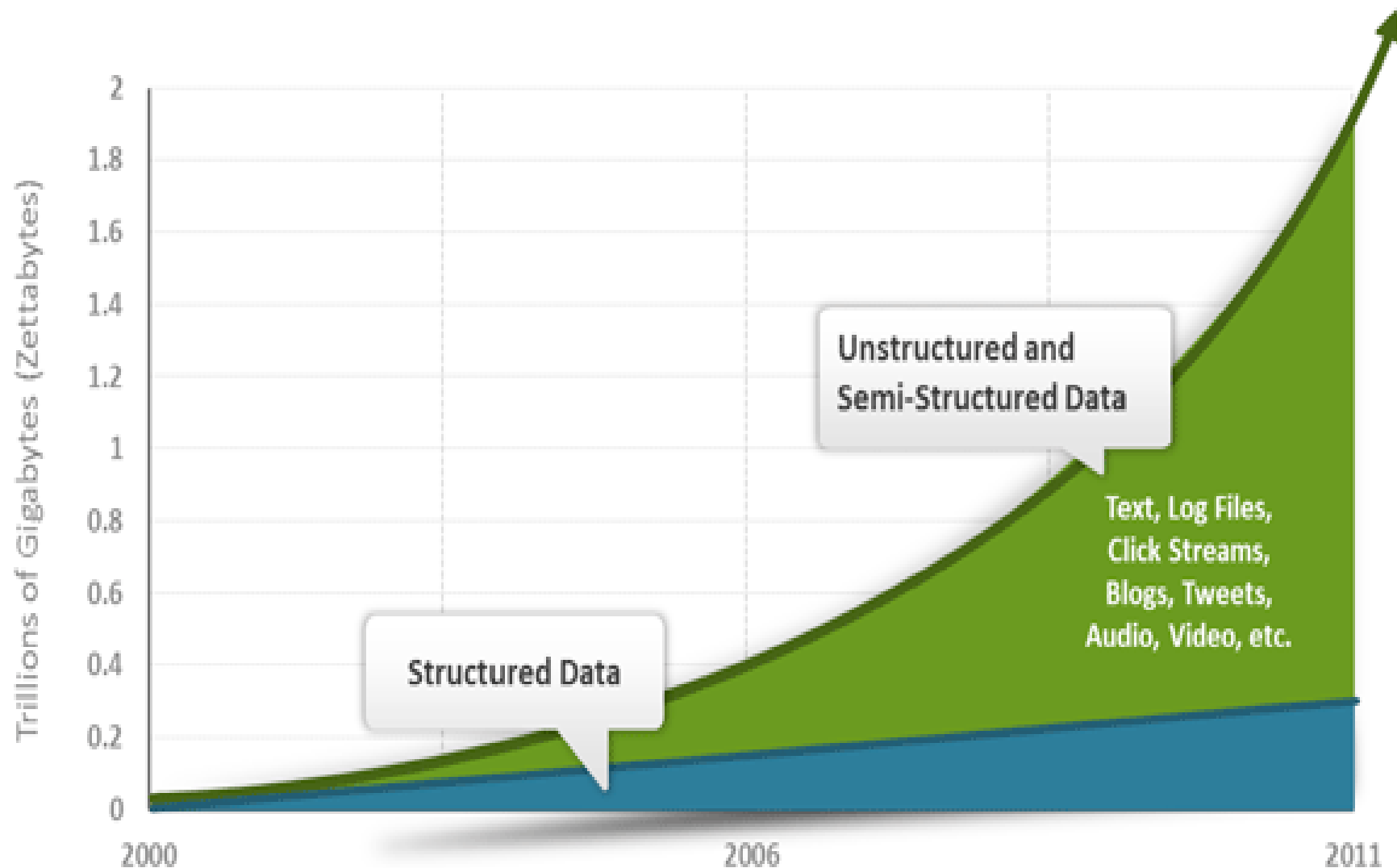
# Nội dung

- Đặt vấn đề
- Thuật ngữ và khái niệm
- Đặc trưng của NoSQL
- Định lý CAP
- Một số tiếp cận tiêu biểu của NoSQL
- Thảo luận

# Đặt vấn đề



# Đặt vấn đề



# Thuật ngữ và khái niệm

- Thuật ngữ NoSQL:
  - No Relational
  - No RDBMS
  - Not Only SQL
- NoSQL là một thuật ngữ chung cho tất cả các cơ sở dữ liệu có cách lưu trữ dữ liệu không tuân theo các nguyên tắc của cơ sở dữ liệu quan hệ

# Đặc trưng của NoSQL

- Lưu trữ khối lượng dữ liệu lớn
- Nhân bản **theo chiều ngang và phân tán**
- Trả lời truy vấn nhanh, chủ yếu là truy vấn, ít cập nhật
- Cập nhật và bổ sung dữ liệu không đồng bộ
- **Không lược đồ**
- ACID không hỗ trợ - chủ yếu là BASE
- Định lý CAP
- Mã nguồn mở

# Định lý CAP



- Khi nói về NoSQL hoặc nói về các hệ phân tán cần nhắc tới định lý CAP
- Định lý CAP: Có 3 yêu cầu quan trọng, liên quan với nhau khi thiết kế các ứng dụng có kiến trúc phân tán, đó là **Consistency, Availability và Partition Tolerance**

# Định lý CAP



- **Consistency** – Dữ liệu trong CSDL luôn nhất quán sau khi thực hiện một thao tác nào đó.  
Ví dụ, Ví dụ, sau thao tác update thì tất cả các trạm đều có giá trị như nhau
- **Availability** – Nghĩa là các dịch vụ của hệ thống luôn sẵn sàng đáp ứng, không có khoảng time gian chết
- **Partition Tolerance** – Nghĩa là hệ thống vẫn tiếp tục hoạt động ngay cả khi kết nối giữa các servers không thực sự tin cậy, tức là server có thể bị phân mảnh thành nhiều nhóm và chúng không giao tiếp được với nhau. (Tính chịu lỗi khi phân mảnh)



# Định lý CAP



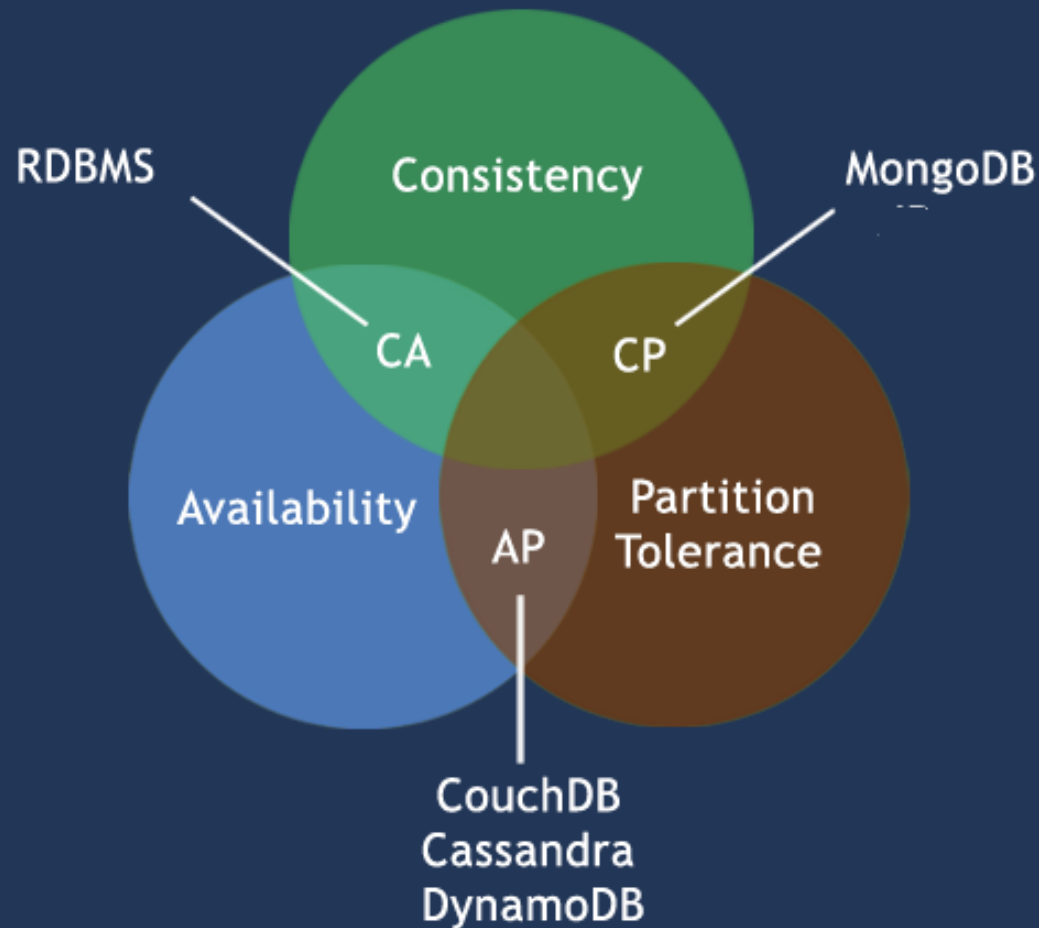
- Nói chung, khó có thể đạt được đồng thời cả 3 yêu cầu trên trong 1 hệ phân tán
- Định lý CAP chỉ ra rằng một hệ phân tán chỉ cần đáp ứng 2 trong 3 yêu cầu trên
- Với hệ phân tán cần phải đáp ứng khả năng chịu lỗi phân mảnh (P), vì vậy hệ thống chỉ cần đáp ứng thêm hoặc là **C** hoặc là **A**.
- Các CSDL NoSQL hiện nay cho phép tạo ra các sự kết hợp khác nhau giữa C và A

# Định lý CAP



- Một số sự kết hợp 2 trong 3 yêu cầu của CAP là CA, CP, AP :
- **CA** – Tất cả các nút của hệ thống luôn liên lạc với nhau, tuy nhiên khi phân mảnh xảy ra thì hệ thống sẽ bị ngưng lại (block)
- **CP** – Một số dữ liệu có thể không truy cập được nhưng phần còn lại vẫn đảm bảo tính nhất quán
- **AP** – Hệ thống luôn có sẵn dữ liệu nhưng một số dữ liệu trả về có thể không đúng đắn

# CAP Theorem



# NoSQL được sử dụng ở đâu?

- **Mạng xã hội: Facebook, Twitter (X)**
  - Công nghệ: Cassandra, MongoDB
  - Sử dụng:
    - Lưu trữ lượng dữ liệu khổng lồ về mối quan hệ người dùng
    - Xử lý nhanh các tương tác như bình luận, like, chia sẻ
    - Mở rộng dễ dàng khi số lượng người dùng tăng

# NoSQL được sử dụng ở đâu?

- **Thương mại điện tử: Amazon, Shopee**
  - Công nghệ: Redis, Couchbase, MongoDB
  - Ứng dụng:
    - Quản lý kho hàng
    - Lưu trữ giỏ hàng và lịch sử mua hàng
    - Hệ thống gợi ý sản phẩm

# NoSQL được sử dụng ở đâu?

- **Quản lý chuỗi cung ứng**
- Ví dụ: Theo dõi hàng hóa, logistics
- Công nghệ: MongoDB, Neo4j
- Ứng dụng:
  - Theo dõi trạng thái vận chuyển
  - Quản lý kho
  - Phân tích hiệu quả chuỗi cung ứng

# NoSQL được sử dụng ở đâu?

- **Ứng dụng Game và giải trí trực tuyến**
  - Ví dụ: Fortnite, World of Warcraft
  - Công nghệ: MongoDB, Cassandra
  - Ứng dụng:
    - Lưu trữ trạng thái người chơi
    - Quản lý điểm số, vật phẩm
    - Xử lý các tương tác thời gian thực
    - Hỗ trợ nhiều người chơi đồng thời

# NoSQL được sử dụng ở đâu?

- **Chăm sóc sức khỏe và y tế**
  - Ví dụ: Hệ thống hồ sơ bệnh án điện tử
  - Công nghệ: MongoDB, Couchbase
  - Ứng dụng:
    - Lưu trữ hồ sơ bệnh nhân phi cấu trúc
    - Tích hợp dữ liệu từ nhiều nguồn
    - Hỗ trợ chia sẻ thông tin y tế

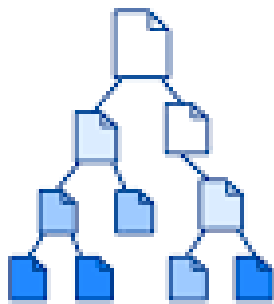


# Một số tiếp cận tiêu biểu của NoSQL

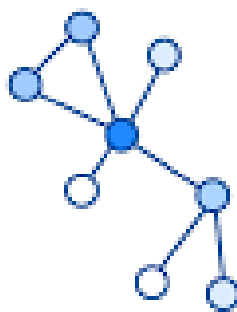


- Key-Value: Lưu trữ dữ liệu theo cặp khóa – giá trị
- Column Family Stores: Lưu trữ dữ liệu theo cột và khóa (theo dòng)
- Document Databases: Lưu trữ dữ liệu theo tài liệu
- Graph Databases: Lưu trữ dữ liệu theo đồ thị

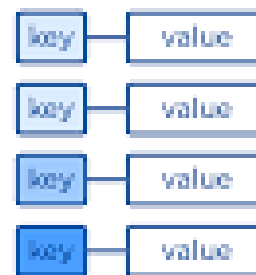
Document



Graph



Key-Value



Wide-column



# Một số tiếp cận tiêu biểu của NoSQL



- Lưu trữ KHÓA-GIÁ TRỊ :
  - Mỗi phần tử dữ liệu được lưu theo cặp Key-Values, trong đó values là một giá trị bất kỳ, có thể là một đối tượng
  - Truy xuất dữ liệu dựa vào key, sau đó dựa vào key trở đến dữ liệu, tốc độ nhanh
  - Kỹ thuật truy xuất: B-tree, B+tree, Hash,...
  - Dữ liệu có thể lưu trữ trên RAM hoặc bộ nhớ ngoài
  - Sản phẩm tiêu biểu: DynamoDB, Azure Table Storage, Riak, Redis,...
  - Giỏ mua hàng của Amazon sử dụng key value (DynamoDB).
  - Lưu trữ key-value store có khả năng mở rộng cao

# Một số tiếp cận tiêu biểu của NoSQL



- Một ví dụ về CSDL Key-Value

Car	
Key	Attributes
1	Make: Nissan Model: Pathfinder Color: Green Year: 2003
2	Make: Nissan Model: Pathfinder Color: Blue Color: Green Year: 2005 Transmission: Auto

# Một số tiếp cận tiêu biểu của NoSQL



- Lưu trữ dạng CỘT:

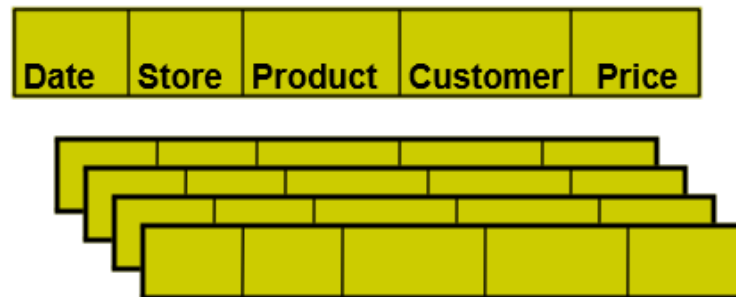
- Mô hình này biểu diễn dữ liệu qua dòng và cột, có thể gọi là mô hình dữ liệu dựa trên bản ghi mở rộng, cho phép tách các cột, các dòng trên nhiều nút
  - Dòng chia theo chiều ngang dựa vào khóa chính
  - Cột chia theo chiều dọc và gom thành các nhóm cột
- Column gồm bộ ba: **1 tên, 1 giá trị và 1 timestamp**
- Tên cột phải có thể có sẵn hoặc phải được đặt bởi ứng dụng, có thể đánh chỉ mục cho cột.
- Timestamp là thời điểm thay đổi dữ liệu cuối, nó chấp nhận phần dữ liệu thay đổi cuối cùng.

# Một số tiếp cận tiêu biểu của NoSQL

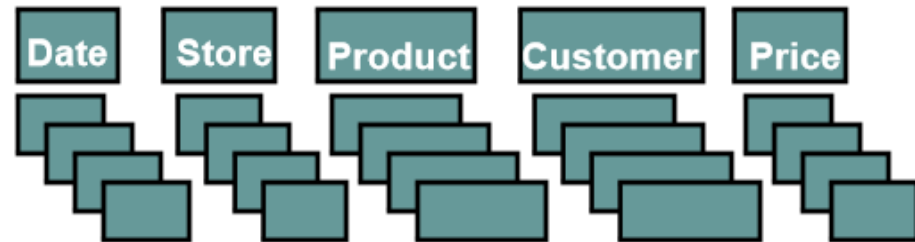


- Row store vs column store:

row-store



column-store



# Một số tiếp cận tiêu biểu của NoSQL



- Ví dụ về CSDL dạng CỘT:

row key	columns		
	ten	Tuoi	email
001	Tú	20	<a href="mailto:tuth@vd.com">tuth@vd.com</a>
	Trang	10	
002	ten	Tuoi	
	Trang	10	
003	ten	Tuoi	diachi
	Văn	21	Việt Nam

# Một số tiếp cận tiêu biểu của NoSQL



- CSDL dạng tài liệu (document database):
  - Thực chất là các cơ sở dữ liệu hướng tài liệu, một thiết kế riêng biệt cho việc lưu trữ tài liệu với định dạng JSON, BSON
  - Các cài đặt có thể là giả lập tương tác trên các cơ sở dữ liệu quan hệ, cơ sở dữ liệu đối tượng hay key-value store.
  - Một số sản phẩm tiêu biểu: MongoDB, Elasticsearch, Couchbase Server, CouchDB, RethinkDB,

# Một số tiếp cận tiêu biểu của NoSQL



- Ví dụ về CSDL document database - MongoDB:

```
{  
  "_id" : ObjectId("4fccbf281168a6aa3c215443"),  
  "first_name" : "Thomas",  
  "last_name" : "Jefferson",  
  "address" : {  
    "street" : "1600 Pennsylvania Ave NW",  
    "city" : "Washington",  
    "state" : "DC"  
  }  
}
```



# Một số tiếp cận tiêu biểu của NoSQL



- Graph Database: Thực thể: nodes; Mỗi liên kết (mối quan hệ): edges

