

AWS Cloud for beginner

Instructor: Linh Nguyen

(Engineering Consultant, AWS Cloud Solution Architect)

Level: Beginner

“Không có việc gì khó, chỉ sợ không biết làm”

NoSQL – DynamoDB

Copyright@Linh Nguyen on Udemy

Target

- Hiểu được NoSQL là gì? Khác biệt gì với SQL (Relational database).
- Hiểu được DynamoDB là gì? Đặc trưng, ưu nhược điểm, usecases của DynamoDB
- DynamoDB Pricing.
- Thao tác cơ bản với DynamoDB
- Mô hình Global Table của DynamoDB
- Indexing trong Dynamo DB là gì? có mấy loại Indexing?
- Thực hành thao tác với Indexing trên DynamoDB.
- Giới thiệu các usecase thành công với DynamoDB của các công ty lớn.

NoSQL là gì?

Non-Relational Database, còn được gọi là NoSQL (Not Only SQL) là một hệ thống cơ sở dữ liệu mà không sử dụng mô hình quan hệ truyền thống dưới dạng các bảng và các quan hệ khóa ngoại. Thay vào đó, nó sử dụng một cấu trúc dữ liệu khác, phù hợp hơn với các ứng dụng có khối lượng dữ liệu lớn, tốc độ truy vấn nhanh và tính mở rộng cao hơn.

Copyright@Linh Nguyen on Udemy

Một số mô hình NoSQL

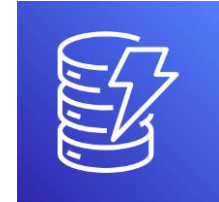
- **Key-value:** Lưu trữ dữ liệu dưới dạng các cặp key-value (khóa-giá trị). Các khóa được sử dụng để truy cập và lấy dữ liệu, trong khi giá trị có thể là bất kỳ kiểu dữ liệu nào.
- **Document:** Lưu trữ dữ liệu dưới dạng tài liệu, thường là định dạng JSON hoặc XML. Các tài liệu được lưu trữ theo dạng phi cấu trúc, cho phép dữ liệu được lưu trữ một cách linh hoạt và thêm vào dễ dàng.
- **Column oriented:** Lưu trữ dữ liệu dưới dạng các bảng với hàng và cột, nhưng khác với cơ sở dữ liệu quan hệ, các cột có thể được thêm và loại bỏ một cách độc lập.
- **Graph:** Lưu trữ dữ liệu dưới dạng các nút và mối quan hệ giữa chúng, cung cấp khả năng xử lý dữ liệu phức tạp.

So sánh SQL, NoSQL

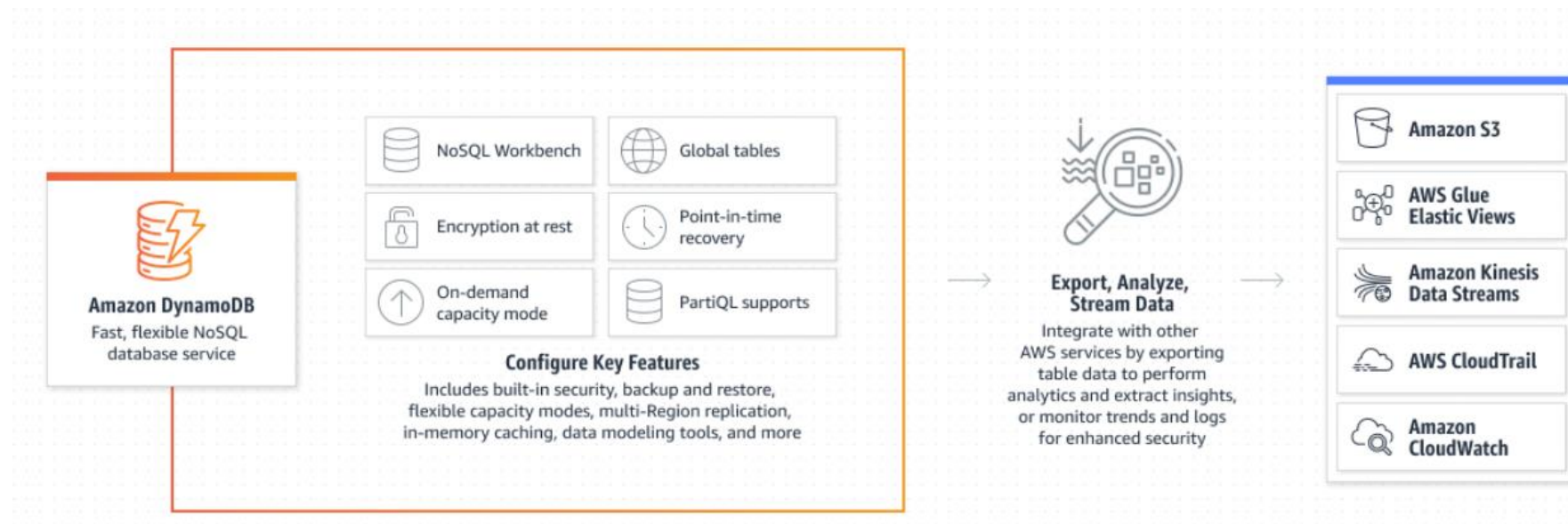
Tiêu chí	Relational Database (SQL)	Non Relational Database (NoSQL)
Cấu trúc dữ liệu	Bảng, Hàng, Cột cần định nghĩa rõ ràng trước khi có thể ghi data.	Không cần định nghĩa trước cấu trúc dữ liệu.
Tính linh hoạt	Cấu trúc của bảng cần được tuân thủ, kém linh hoạt hơn.	Linh hoạt cao. Dễ dàng thêm thuộc tính.
Hiệu suất	Tối ưu cho các thao tác truy vấn phức tạp, join bảng. VD Thống kê data, analytic.	Tối ưu cho xử lý tốc độ cao
Bảo mật	Bảo mật cao do có cấu trúc dữ liệu rõ ràng	Khó khăn trong việc quản lý và bảo mật truy cập.
Tính mở rộng	Khả năng mở rộng kém, thường phải thay đổi cấu hình database.	Khả năng mở rộng dễ dàng. Scale horizontal.

DynamoDB là gì?

“**Amazon DynamoDB** is a fully managed, serverless, key-value NoSQL database designed to run high-performance applications at any scale. DynamoDB offers built-in security, continuous backups, automated multi-Region replication, in-memory caching, and data import and export tools.”



Amazon DynamoDB



Đặc trưng của DynamoDB

- Serverless: hạ tầng được quản lý bởi AWS. User tương tác với DynamoDB thông qua Console, CLI, Các tool client hoặc Software SDK.
- Data được tổ chức thành các đơn vị **table**
- Độ trễ thấp (single digit milisecond)
- SLA: 99.999% availability
- Automatic Scale Up/Down tùy theo workload (WCU, RCU)
- Kết hợp được với nhiều service khác của AWS

Copyright@Linh Nguyen on Udemy

Đặc trưng của DynamoDB

Ưu điểm

- Serverless nên chi phí vận hành thấp, tính khả dụng cao.
- Linh hoạt trong cấu hình (zero idle cost, phù hợp cho startup)
- Khả năng scale không giới hạn (về lý thuyết), độ trễ thấp, hiệu suất cao.
- Strongly consistency
- Hỗ trợ mã hoá

Nhược điểm

- Không phù hợp với data query và analytic (OLAP)
- Thiếu tính năng khi so sánh với SQL (relational DB)

DynamoDB usecases

- Software application: hầu hết các software có nhu cầu về high concurrent cho hàng triệu user đều có thể cân nhắc sử dụng DynamoDB. Vd E-commerce.
- Media metadata store: Lưu trữ metadata cho các media.
- Gaming platform
- Social media: mạng xã hội, bài đăng, bình luận.
- Logistic system.
- Ứng dụng IoT

Copyright@Linh Nguyen on Udemy

DynamoDB usecases

Một số công ty nổi tiếng có sử dụng DynamoDB có thể kể đến:

- Amazon
- Samsung
- Ubisoft
- Netflix
- Airbnb
- ShipStation
- Jira
- Grab

Copyright@Linh Nguyen on Udemy

DynamoDB Pricing

DynamoDB tính tiền trên các thông số

- Write Capacity Unit (WCU) và Read Capacity Unit (RCU). Hai đơn vị này lại có 2 mô hình tính giá là On-demand và Provisioned Capacity.
- Dung lượng lưu trữ (tính theo đơn vị KB). Vd Singapore: \$0.285/GB/Month
- Data backup & restore: Vd Singapore: \$0.228/GB/Month
- Tần suất query data
- Indexing
- Data import/export
- DAX (caching cluster sẽ được trình bày sau)
- Data transfer ra bên ngoài
- ...and more

Tham khảo: <https://aws.amazon.com/dynamodb/pricing/on-demand/>

DynamoDB Pricing – Hiểu rõ về WCU/RCU

RCU: Read capacity unit. Với mỗi đơn vị block 4KB, bạn có thể

- Thực thi 2 request read/s với mode eventually consistency
- Thực thi 1 request read/s với mode strongly consistency

Ví dụ bạn có một data nặng 8KB, khi read với eventually consistency mode bạn cần tiêu tốn 1RCU, khi read với strongly consistency mode bạn cần tiêu tốn 2RCU.

*Lưu ý data vượt qua 4KB sẽ bị làm tròn sang block tiếp theo (vd 5, 6 or 7KB)

Copyright@Linh Nguyen on Udemy

DynamoDB Pricing – Hiểu rõ về WCU/RCU

WCU: Write capacity unit. Với mỗi đơn vị block 1KB, bạn cần

- 1WCU để thực hiện write với standard mode.
- 2WCU để thực hiện write với transactional mode.

Ví dụ bạn có một data nặng 3KB, khi write với normal mode bạn cần tiêu tốn 3WCU, khi write với transactional mode bạn cần tiêu tốn 6WCU.

Copyright@Linh Nguyen on Udemy

DynamoDB Pricing – On-demand vs Provisioned

- Với mode On-demand, DynamoDB sẽ tự động scale WCU, RCU khi có nhu cầu tăng cao, tuy nhiên sẽ có khả năng bị bottle neck nếu work load tăng đột ngột.
- Giá tiền tính như sau:
 - \$1.25 per million write request units
 - \$0.25 per million read request units
- Với mode provisioned, DynamoDB sẽ reserve một số WCU, RCU cố định cho từng bảng hoặc index. Sử dụng cho những table muốn duy trì workload cao.
- Giá tiền tính như sau (giá theo giờ):
 - \$0.00065 per WCU
 - \$0.00013 per RCU

DynamoDB Free tier & Always free

Free tier (first 12 months)

- 25 GB of data storage for tables using the DynamoDB Standard table class
- 2.5 million stream read requests from DynamoDB Streams
- 25 Provisioned WCU
- 25 Provisioned RCU
- 100 GB of data transfer out to the internet, aggregated across all AWS Services and Regions (except China and GovCloud)

Always free

- 25 GB of data storage
- 25 Provisioned WCU
- 25 Provisioned RCU
- *Đủ cho khoảng 200 millions request/month

<https://aws.amazon.com/free/database/?p=ft&z=subnav&loc=3>

DynamoDB Concepts

- Table: đơn vị quản lý cao nhất của DynamoDB. Table không thể trùng tên trên một region.
- Primary key: thông tin bắt buộc khi tạo table, Primary key chia làm 2 loại
 - Simple Primary key: chỉ bao gồm **Partition key**
 - Composite primary key: bao gồm **Partition key** và **Sort key**.
- Global Secondary index (optional): bao gồm một cặp partition key & sort key tùy ý.
- Local Secondary index (Optional): bao gồm một cặp partition key giống với partition key của table và sort key tùy ý.
- Query: thực hiện truy vấn data trên một index.
- Scan: tìm toàn bộ giá trị thoả mãn điều kiện (tất nhiên sẽ chậm hơn query).
- PartiQL: SQL-compatible query language cho phép access data quan hệ, bán cấu trúc hoặc cấu trúc lồng nhau (nested).

DynamoDB Concepts – Data type

DynamoDB hỗ trợ các loại data type sau:

- String
- Number
- Binary
- Boolean
- Null
- Set
- Map
- List

```
{
  "id": "123",
  "title": "Product A",
  "description": "This is the description for Product A",
  "price": 29.99,
  "stock": 50,
  "category": "electronics",
  "tags": ["smartphone", "apple", "ios"],
  "attributes": {
    "color": "black",
    "manufacturer": "Apple",
    "screen_size": 5.8
  },
  "createdAt": "2022-04-30T12:00:00Z",
  "updatedAt": "2022-04-30T14:30:00Z"
}
```

Data sample trong DynamoDB dưới dạng Json view

*Lưu ý không có Date Time. Cần convert sang Unix time hoặc sử dụng format string yyyyymmdd.

Lab-1 DynamoDB basic information

Login vào AWS console, navigate to DynamoDB

1. Tạo table có tên **student** với các thông tin

- id: String – Partition key
- name: String – Sortkey

2. Input một vài record với các thông tin

id	name	birthday	hobbies	skill	is_married
st-01	Linh	1992/05/06	["piano","travel"]	-	false
st-02	Truc	1999/12/03	-	["ielts", "dance"]	true
st-03	Phuong	2000/08/02	["gym", "yoga"]	["guitar"]	true

Lab-1 DynamoDB basic information

Login vào AWS console, navigate to DynamoDB

3. Thực hiện query student có **id: st-01, name: "Linh"**
4. Thực hiện scan student có **name: "Phuong"**
5. Thực hiện scan student có **is_married: false**

Copyright@Linh Nguyen on Udemy

DynamoDB Concepts - Indexing

DynamoDB hỗ trợ 2 loại index là **Global Secondary Index(GSI)** và **Local Secondary Index(LSI)**

- **Local Secondary Index(LSI)**: loại index phụ thuộc vào table và sử dụng partition key của table để xác định vị trí index. LSI giới hạn số lượng index của mỗi table là 5 và chỉ có thể tạo lúc tạo table. Các Attribute của LSI phải nằm trong bảng gốc.
- **Global Secondary Index(GSI)**: một loại index độc lập với table và có thể xác định các attribute khác ngoài partition key để tìm kiếm nhanh dữ liệu. Với GSI, bạn có thể tạo ra các index mới mà không cần phải thay đổi cấu trúc của table gốc. DynamoDB hỗ trợ tối đa 20 GSI trên mỗi table.

Khi tạo index, ta có thể setting attribute projection như sau

- Project all: Secondary index sẽ bao gồm luôn các attribute của bảng chính
- Project key only: Secondary index chỉ bao gồm thông tin primary key.
- Project selected attribute: chỉ những attributes được chỉ định sẽ bao gồm trong index.

Lab 2 - Indexing

Sử dụng lại table của Lab-1, thực hiện tạo Global Secondary Indexing như sau

- Partition key: name
- Sortkey: birthday

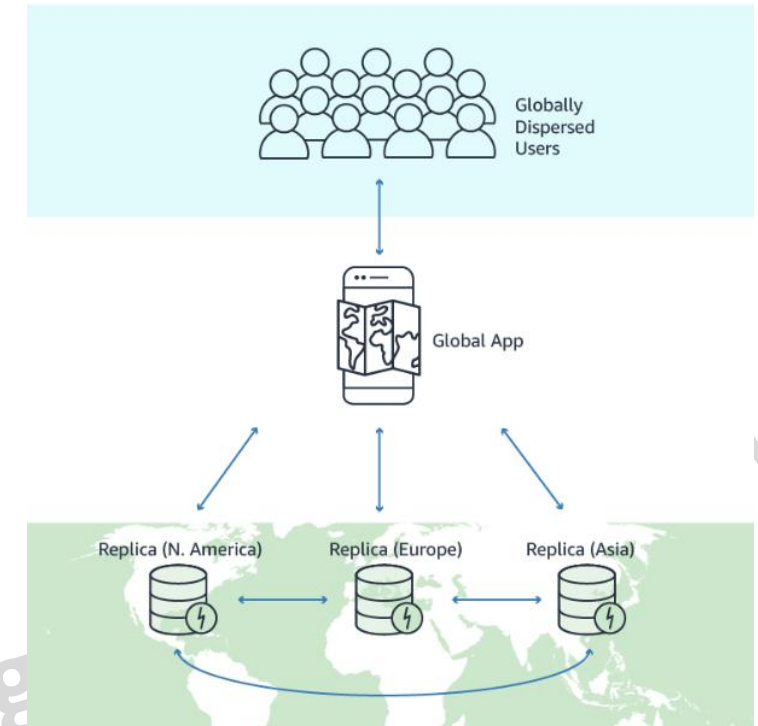
Thực hành query data sử dụng indexing.

- Query tất cả học sinh có name = Linh, birthday = 1992/05/06
- Scan tất cả học sinh có name = Linh, birthday = 1992/05/06
- So sánh tốc độ, consume RCU khi scan và query.

Copyright@Linh Nguyen on Udemy

DynamoDB Global table

DynamoDB **Global table** là một tính năng của Amazon DynamoDB, cho phép bạn sao chép và đồng bộ dữ liệu giữa các khu vực khác nhau trên toàn cầu. Nó cho phép bạn tạo các bảng DynamoDB trên nhiều khu vực và tự động sao chép dữ liệu giữa chúng để đảm bảo dữ liệu được phân phối đồng đều trên toàn cầu và có sẵn cho các ứng dụng trong nhiều vùng địa lý khác nhau.



Copyright@Linh Ng

DynamoDB Accelerator (DAX)

Amazon DynamoDB Accelerator (DAX) is a fully managed, highly available, in-memory [cache](#) for [Amazon DynamoDB](#) that delivers up to a 10 times performance improvement—from milliseconds to **microseconds**—even at **millions of requests per second**.

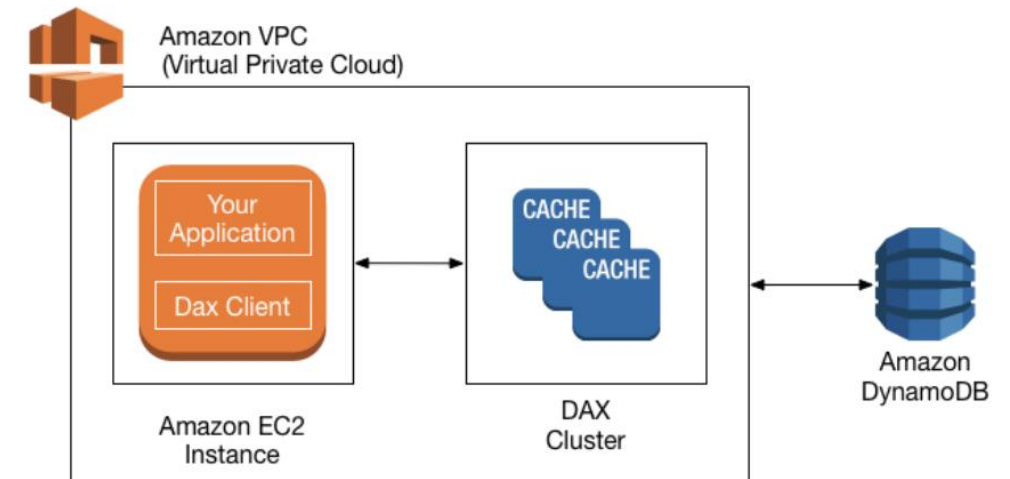


Amazon DynamoDB
Accelerator (DAX)

Copyright@Linh Nguyen on Udemy

DynamoDB Accelerator (DAX)

- Dax hoạt động dưới dạng cluster.
- Để truy cập data thông qua DAX, cần có DAX client.
- Khi chỉ định mode read strongly consistency, request sẽ được forward thẳng tới DynamoDB.

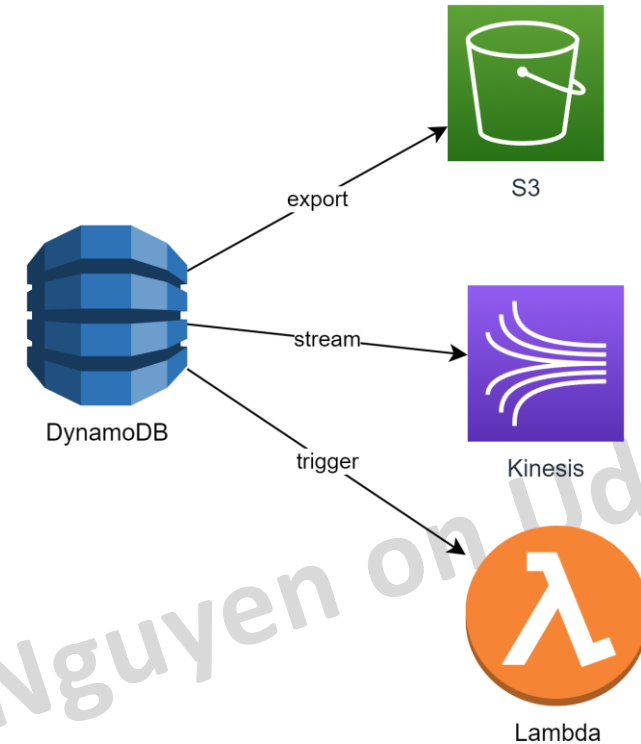


Nguồn: <https://aws.amazon.com/dynamodb/dax/>

Copyright@Linh Nguyen

DynamoDB Export and stream

Dynamodb hỗ trợ export data sang S3 hoặc stream sang các target như Kinesis (một dịch vụ data streaming) hoặc Lambda phục vụ các tác vụ xử lý data tự động.

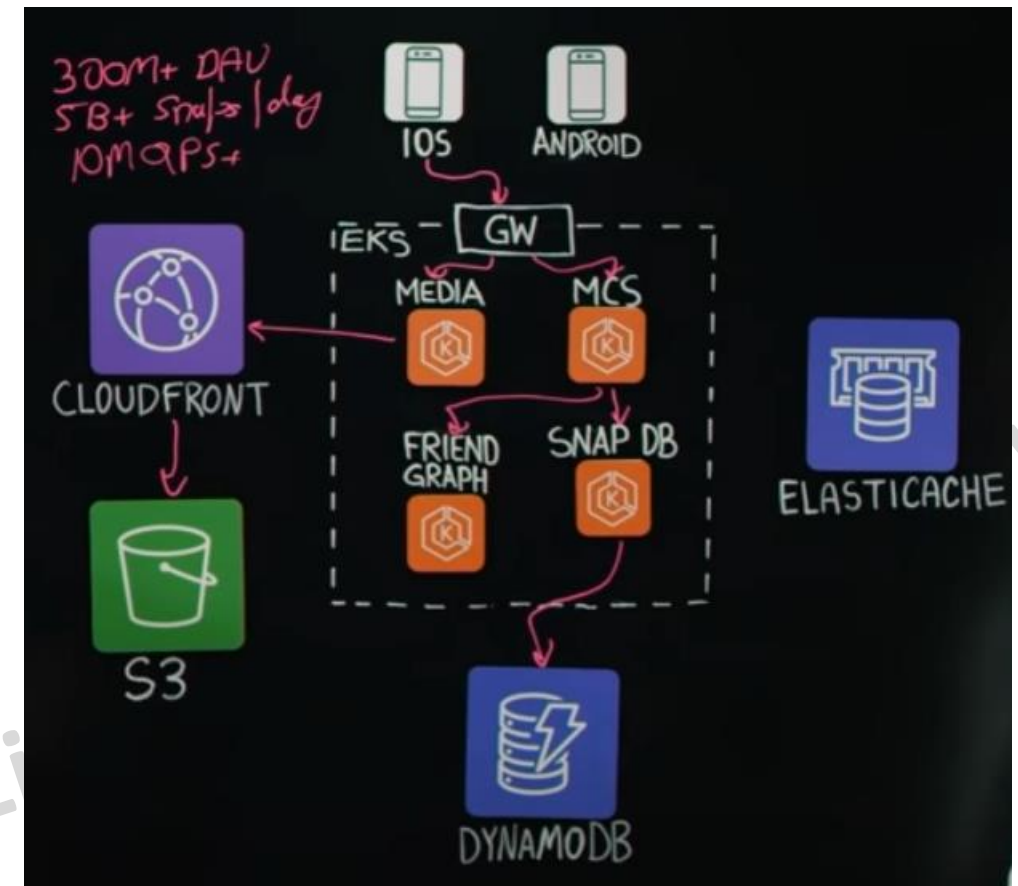


Companies success usecase



Snapchat

- 300M Daily Active User
- 400TB data on DynamoDB



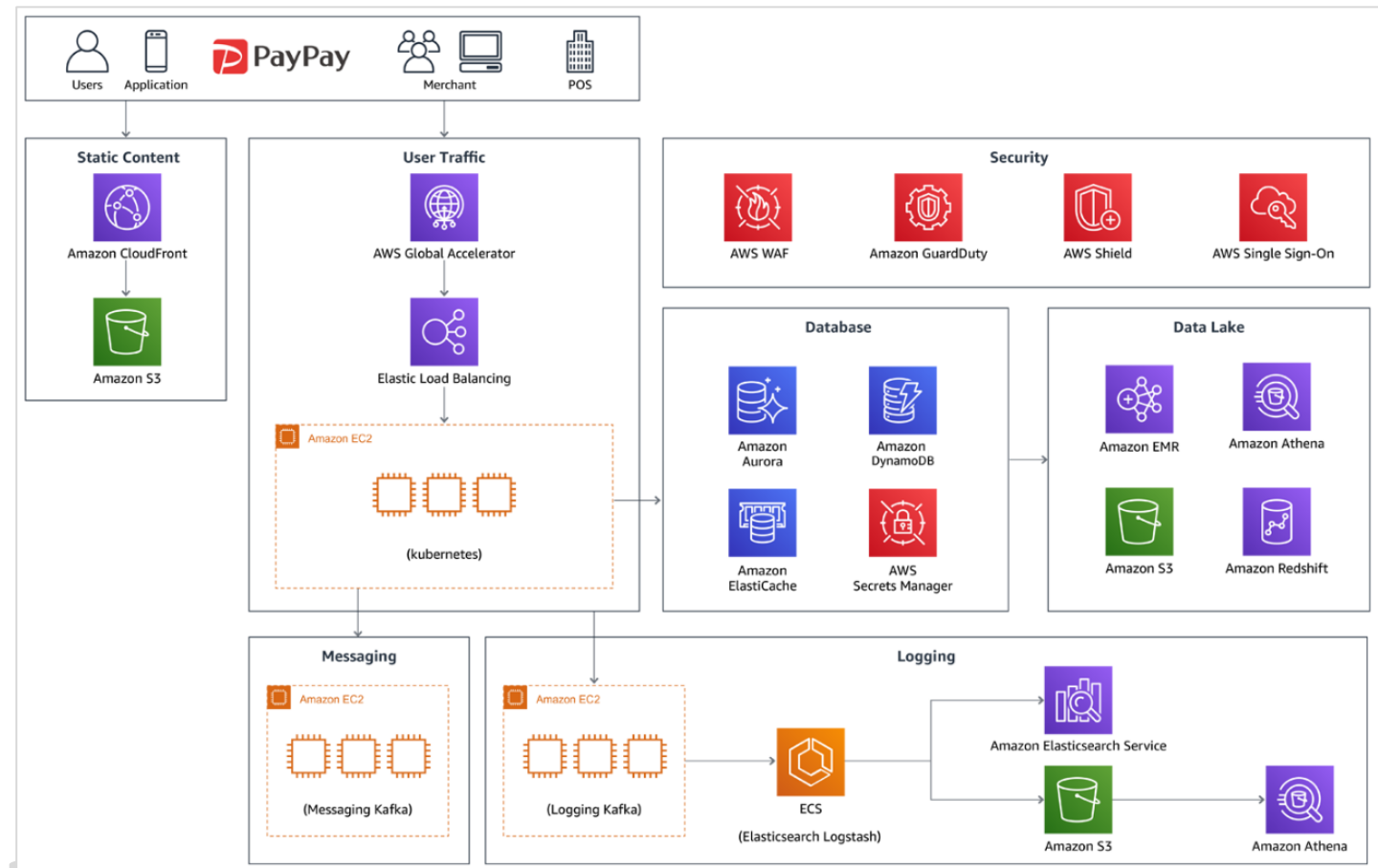
Companies success usecase

Paypay

- one of Japan's largest cashless payment services
- 50 million registered users as of August 2022

Key points

- Microservice
- Cost optimize
- Reduce operational cost
- Active-Active system using two regions
- Superapp and castless platform



Companies success usecase

Read more at: <https://aws.amazon.com/dynamodb/customers/>

Copyright@Linh Nguyen on Udemy

Lab: Clear resources

Login to AWS console, thực hiện nội dung sau:

1. Xoá DynamoDB table

Copyright@Linh Nguyen on Udemy