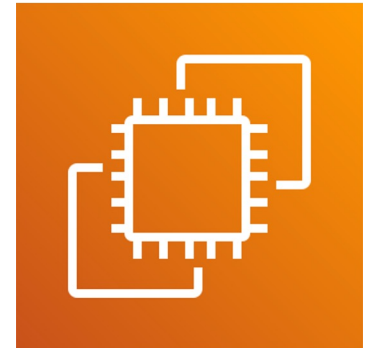


Giới thiệu về AWS EC2

Giới thiệu AWS EC2?

- EC2 viết tắt của Elastic Compute Cloud
- Là dịch vụ lâu đời và quan trọng nhất của AWS
- Cung cấp khả năng tính toán trên môi trường Cloud AWS



EC2 mô hình trả phí (pricing model)



1

On-Demand

Trả phí theo thời gian sử dụng, dùng bao nhiêu trả bấy nhiêu. Không cần trả trước (no upfront cost), không cần cam kết (no commitment)

3

Spot

Bạn đưa giá mức giá thầu (bid price). AWS sẽ đưa ra mức giá Spot Price (mức giá phụ thuộc vào cung cầu và công suất của AWS).

Cách trả phí có thể tiết kiệm lên đến 90% so với On-Demand.

2

Reserved

Đặt trước tài nguyên với AWS. AWS sẽ bảo lưu và đảm bảo tài nguyên.

Mức trả phí tiết kiệm hơn so với On-Demand, tuy nhiên sẽ phải có hợp đồng kéo dài từ 1 ~ 3 năm

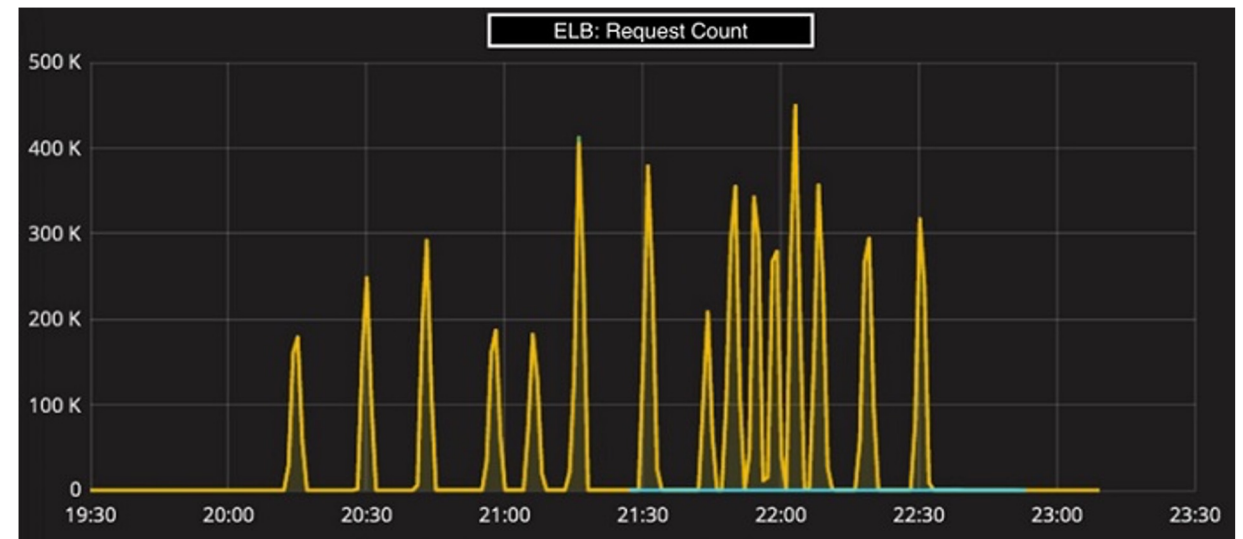
4

Dedicated Host

Đặt riêng Host vật lý của AWS để chạy VM trên đó. Dành cho việc tuân thủ theo licence khi chạy ứng dụng, phần mềm của 3rd party trên Cloud.

On-demand use case

- Dùng cho mục đích thử nghiệm, development trong thời gian ngắn
- Dùng cho ứng dụng chạy trong ngắn hạn, tải không đoán định được
- Không cần cam kết (No commitment), không cần trả trước (no Up-front payment)



Reserved Instance (RI) use case

- Dùng cho các ứng dụng chạy trong thời gian dài, tải chạy ổn định (predictable workload)
- Phải có cam kết (hợp đồng ít nhất 1 năm) và phải thanh toán trước một phần phí.

Standard RI

Convertible RI

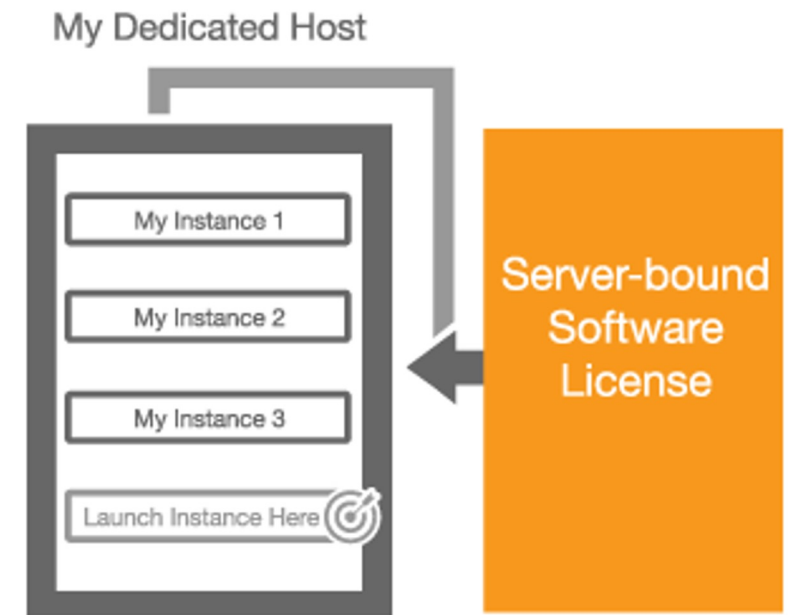
Scheduled RI

Spot Instance use case

- Có thể tiết kiệm tới 90% so với chi phí khi chạy **On-Demand**
- Dành cho ứng dụng có tải chạy trong thời gian ngắn, có thể chấp nhận được việc VM bị tắt đột ngột gây ra Downtime.
- Thường được dùng cho chạy ứng dụng cần cấu hình mạnh với mức phí phù hợp

Dedicated Hosts use case

- Host vật lý dành riêng cho việc chạy VM
- Phục vụ triển khai phần mềm của 3rd party có giấy phép (licence) yêu cầu chạy hạ tầng ảo hoá riêng biệt, không chung với các người thuê (tenant) khác.
- Có thể trả phí theo On-Demand hoặc Reserved



EC2 instance types

- Kiểu Instance (Instance Type) format: Family.Type (t2.micro, m5.large...)
- Mỗi **Family** sẽ được dành riêng cho các mục đích khác nhau (Cần nhiều RAM hay CPU hay GPU...)
- Mỗi **Type** sẽ cho biết được là công suất của Instance trong **Family** đó
- Ví dụ:

Family	Specialty	Use case
R5	Memory Optimized	Application, DBs
C5	Compute Optimized	High CPU apps, DBs
T2/T3	General purpose, cheap	Web
Mac1	For MacOS	For MacOS user

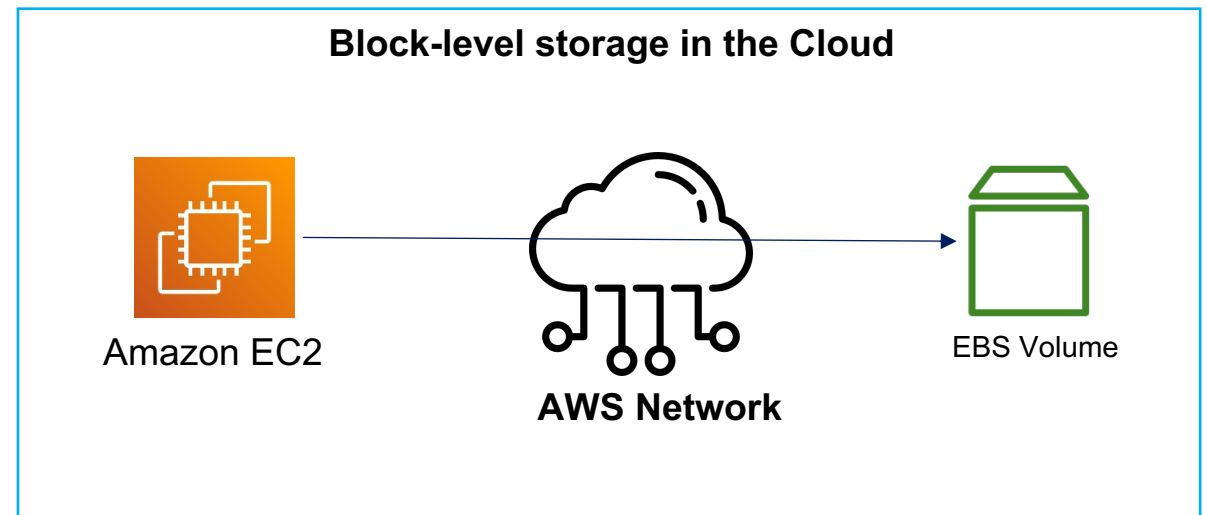
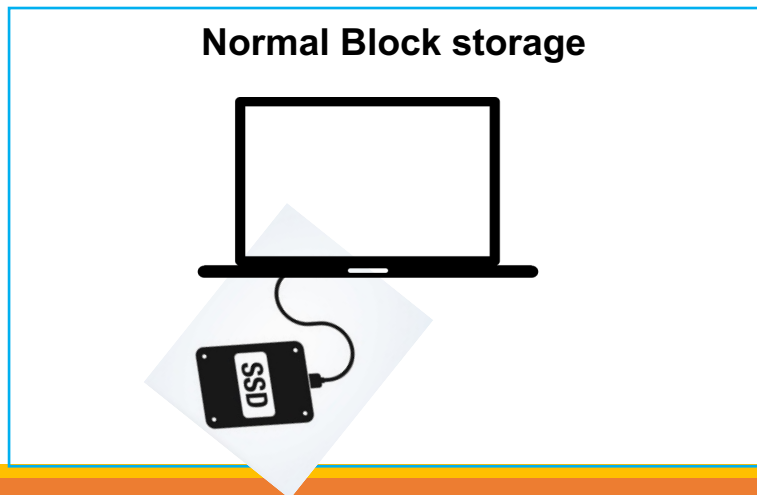
EC2 summary

- AWS EC2 là dịch vụ cung cấp khả năng tính toán trên Cloud AWS
- Mô hình trả phí (Pricing model) cần được tính tới khi triển khai hạ tầng trên AWS
 - On-demand
 - RI
 - Spot
 - Dedicated Hosts
- Mỗi Instance type sẽ dành riêng cho các mục đích khác nhau (Tác vụ cần nhiều RAM, nhiều CPU, nhiều GPU hay cần tối ưu IOPS...)

Giới thiệu về EBS

Khái niệm EBS volume?

- EBS viết tắt của Elastic Block Storage
- EBS cung cấp **block-level storage** trên Cloud
- EBS volume có thể dễ dàng attach/detach vào một Instance
- EBS là AZ scope (Nếu AZ của EBS bị sự cố, dữ liệu sẽ bị mất)



Các loại EBS volume

Solid state drives (SSD)			Hard disk drives (HDD)		
Loại Volume	General Purpose SSD	Provisioned IOPS SSD	Throughput Optimized HDD	Cold HDD	Magnetic
Mô tả	Cho các tác vụ cơ bản. Cân đối giữa giá và hiệu năng	Cho các tác vụ cần hiệu năng cao, độ trễ thấp, throughput cao khi đọc ghi vào ổ đĩa (EBS volume)	HDD với mức giá thấp, dữ liệu thường xuyên truy cập. Dành cho các tác vụ không yêu cầu cao về hiệu năng và throughput	HDD với mức giá thấp, dữ liệu ít thường xuyên được truy cập.	HDD thế hệ cũ
Tên	gp2, gp3	io1, io2	st1	sc1	standard
Use Cases	Boot volumes, low-latency interactive applications, dev, test.	I/O intensive NoSQL and relational databases.	Big data Data warehouses Log processing	File Servers	Workloads where data is infrequently accessed
Giới hạn về kích thước volume	1 GiB - 16 TiB	4 GiB - 64 TiB	125 GiB - 16 TiB	125 GiB - 16 TiB	1 GiB-1 TiB
Throughput	250 MiB/s (gp2) 1000 MiB/s (gp3)	1000 MiB/s (io1, io2) 4000 MiB/s (io2 express)	500 MiB/s	250 MiB/s	40–90 MiB/s
Max IOPS/volume	16000	64000	500	250	40 - 200

Exam tips

- EBS volume là **block-level storage** (Khác với **Object storage**)
- EBS là AZ scope. Không thể mount một EBS volume với một EC2 instance ở một AZ khác.
- Lưu ý với Volume dạng GP2 (General Purpose)
 - Với kích thước Volume $< 5,334 \text{ GiB} \Rightarrow \text{IOPS} = \text{Volume size} * 3$
 - Với kích thước Volume $5,334 \text{ GiB} \sim 16\text{TiB} \Rightarrow \text{IOPS} = 16,000$
 - Burst IOPS = 3000 cho các Volume có kích thước $< 1000 \text{ GiB}$

Snapshot

Khái niệm Snapshot?

- Snapshot là bản sao của một EBS volume
- Snapshot là **point in time copy** của một EBS volume
- Snapshot là incremental (Chỉ những **Block** bị thay đổi so với bản Snapshot trước đó mới được upload lên S3)
- Để nhất quán dữ liệu (Consistent Data), Instance nên được stop trước khi tạo Snapshot
- Khi tạo Snapshot, Volume sẽ bị tăng tải nên hạn chế tạo Snapshot khi đang Volume phải xử lý nhiều tác vụ



AMI

Khái niệm AMI?

- **AMI** viết tắt của Amazon Machine Image
- AMI được sử dụng để chạy một **EC2 Instance**
- Để lựa chọn AMI để chạy một EC2 Instance có thể dựa vào các tiêu chí sau:
 - Region chạy EC2 Instance
 - Hệ điều hành
 - Kiến trúc (amd or arm)
 - Ổ đĩa boot hệ điều hành (EBS hay Instance store)

Các kiểu AMI?

EBS-backed Volume

- AMI loại này được tạo từ **EC2 Instance** có ổ đĩa Root (ổ đĩa cài OS) là một **EBS-Backed Volume**.
- EC2 chạy từ AMI loại này có thể **Stopped** được
- Dữ liệu trong ổ đĩa Root được bảo toàn

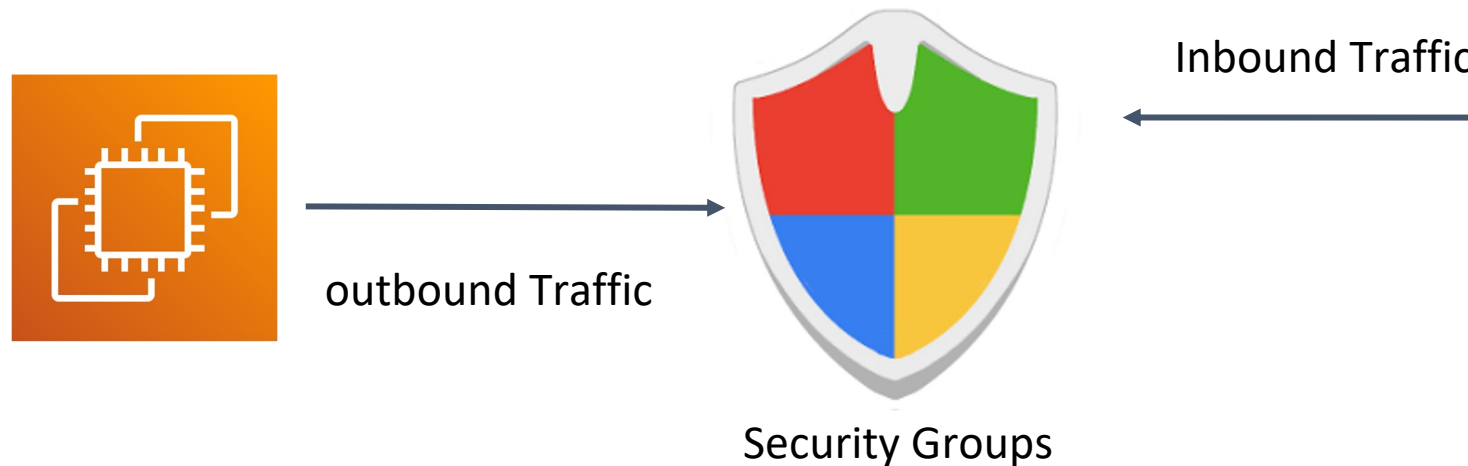
Instance Store Volume

- AMI loại này được tạo từ **EC2 Instance** có ổ đĩa Root (ổ đĩa cài OS) là một **Instance Store Volume**.
- EC2 chạy từ AMI loại này có không thể **Stopped** được
- Khi Host vật lý bị sự cố, **EC2 Instance** chạy từ AMI loại này sẽ bị sự cố mất dữ liệu (mất ổ Root)

Security Groups

Khái niệm Security Groups?

- Security Groups hoạt động giống như một Firewall ảo trên AWS
- Security Groups kiểm soát truy cập vào và ra của một EC2 Instance hoặc một Load Balancer (LB)



Các tính năng của Security Groups



- Security Group là dạng Firewall có lưu trạng thái (Stateful Firewall)
- Chỉ có thể thêm được các luật cho phép (Allow/Whitelist), không thể thêm các luật từ chối (Deny/Blacklist)
- Không thể sử dụng SG để block một hay một dải IP
- Các luật được thêm vào SG được áp dụng ngay lập tức (Immediately)

Tính năng Security Groups (cont.)

- Một SG có thể attach vào một hay nhiều EC2 Instance, một EC2 Instance có thể gắn nhiều SGs
- Security Groups rule mặc định:
 - Tất cả Inbound Traffic sẽ bị chặn (block)
 - Tất cả Outbound Traffic sẽ được thông (allow)

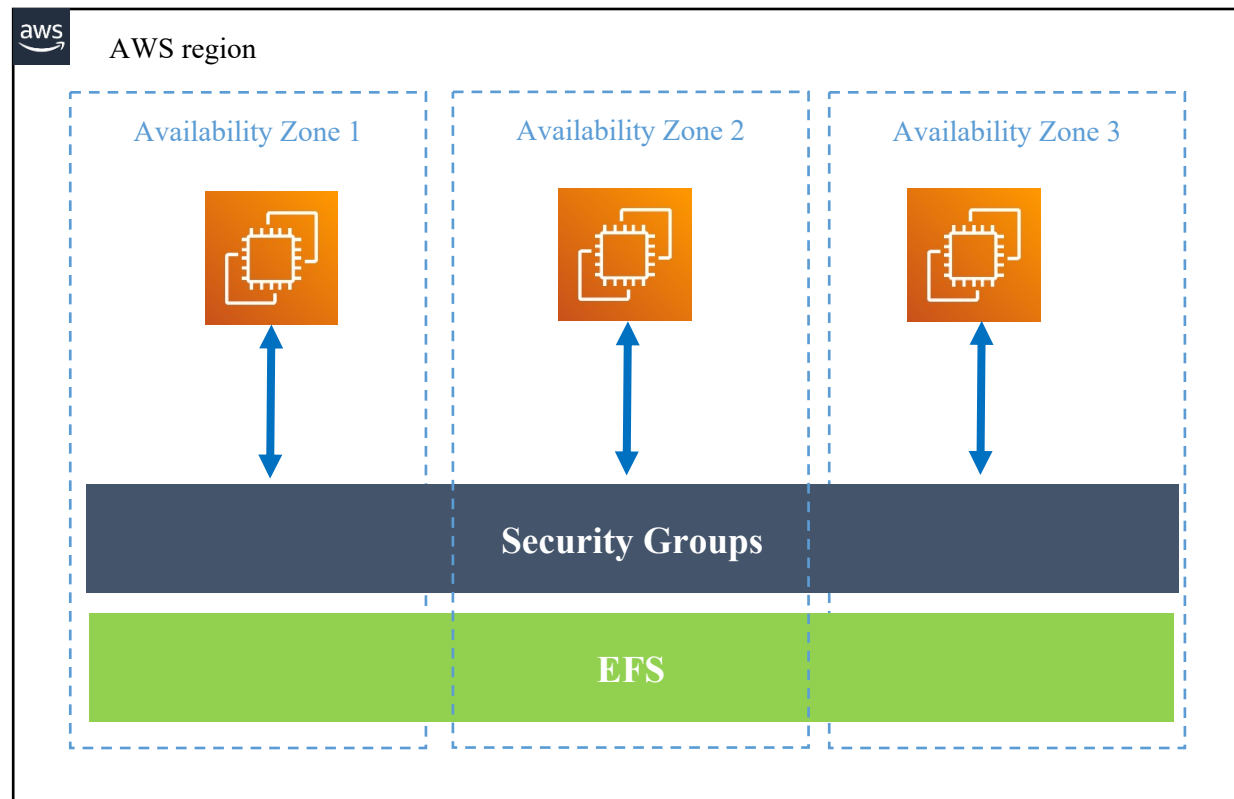
Giới thiệu về EFS

Khái niệm EFS?

- EFS viết tắt của Elastic File System.
- Là dịch vụ cung cấp File Server được quản lý bởi AWS
- EFS có thể trải rộng giữa nhiều AZ, hoặc Region
- Dịch vụ có tính High Availability và Scalability
- EFS có thể gắn với Security Group để kiểm soát truy cập về mặt kết nối



Khái niệm EFS (cont.)

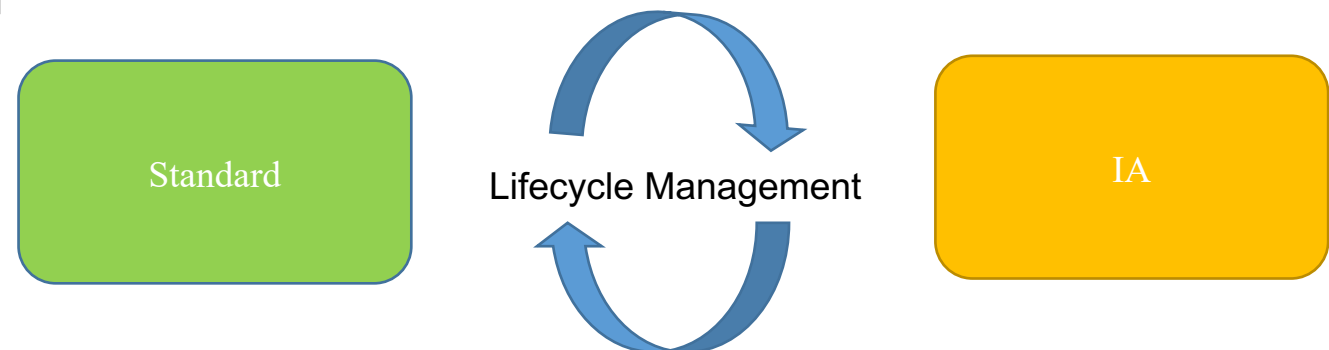


Khái niệm EFS (cont.)

- Sử dụng EFS cho mục đích chia sẻ dữ liệu (wordpress, Bigdata, Web...)
- Tuân thủ theo POSIX
- Chỉ sử dụng EFS cho Instance Linux, không dùng được cho Instance Window
- Dữ liệu có thể được mã hoá lưu trữ (at rest) sử dụng KMS

EFS storage class

- **Standard**
 - Dành cho lưu trữ dữ liệu thường xuyên truy cập
 - Chi phí được tính cho phần lưu trữ (storage)
- **Infrequent Access (IA)**
 - Dành cho lưu trữ dữ liệu ít thường xuyên truy cập
 - Chi phí được tính cho lưu trữ + lấy dữ liệu



EFS performance mode

- **General Purpose**
 - Dành cho các tác vụ yêu cầu độ trễ thấp
 - Web, CMS, Home Directory
- **Max I/O**
 - Dành cho Workload cần xử lý song song, throughput lớn
 - Bigdata Analytic, Media Processing

ENI, ENA và EFA

Khái niệm ENI, ENA and EFA?

- **ENI**
 - Elastic Network Interface – Cạc mạng ảo (Virtual Network Interface)
- **ENA**
 - Enhanced Networking Adapter.
 - Cung cấp khả năng đáp ứng hiệu năng cao về mặt Networking, dành cho một số dòng Instance type hỗ trợ (VD: r4.2xlarge)
- **EFA**
 - Elastic Fabric Adapter – Network device
 - Dành cho tính toán hiệu năng cao (High Performance Computing - HPC) và ML

ENI

- Một **EC2 Instance** sẽ có thể gắn nhiều **ENI**. Mỗi **ENI** được thiết lập một primary IP và nhiều secondary IP lấy từ dải IP của Subnet
- Mỗi **ENI** có thể gắn nhiều **Security Groups**
- Use case
 - ENI được sử dụng để cấp phát IP cho một EC2 instance
 - Sử dụng nhiều ENI khi instances cần nhiều địa chỉ IPs



Elastic network adapter

ENA

- ENA là một card mạng ảo (Virtual Network Interface)
- Sử dụng SR-IOV (Phương thức ảo hoá thiết bị mạng – Network Device Virtualization Method)
- Cung cấp khả năng đáp ứng hiệu năng cao về mặt Networking (Up to 100 Gbps)
- Use case
 - Dành cho các ứng dụng cần có throughput cao



Elastic network adapter

EFA

- EFA là Network device có thể gắn vào EC2 instance
- Cho phép hỗ trợ tính toán hiệu năng cao (HPC) và ML
- Cho phép OS-bypass và truy cập trực tiếp vào thiết bị EFA
- Use case
 - HPC (High Performance Computing), Machine Learning

Userdata

Userdata

- Cho phép thực thi các câu lệnh, thiết lập config khi Instance start
- Userdata script chỉ thực thi khi Instance Launch lần đầu tiên
- Userdata script sẽ được thực thi với quyền của user root
- Use case:
 - Software update
 - Cài đặt package
 - Cài đặt, thiết lập cấu hình (ssh, user, application)

Spot Instance và Spot Fleet

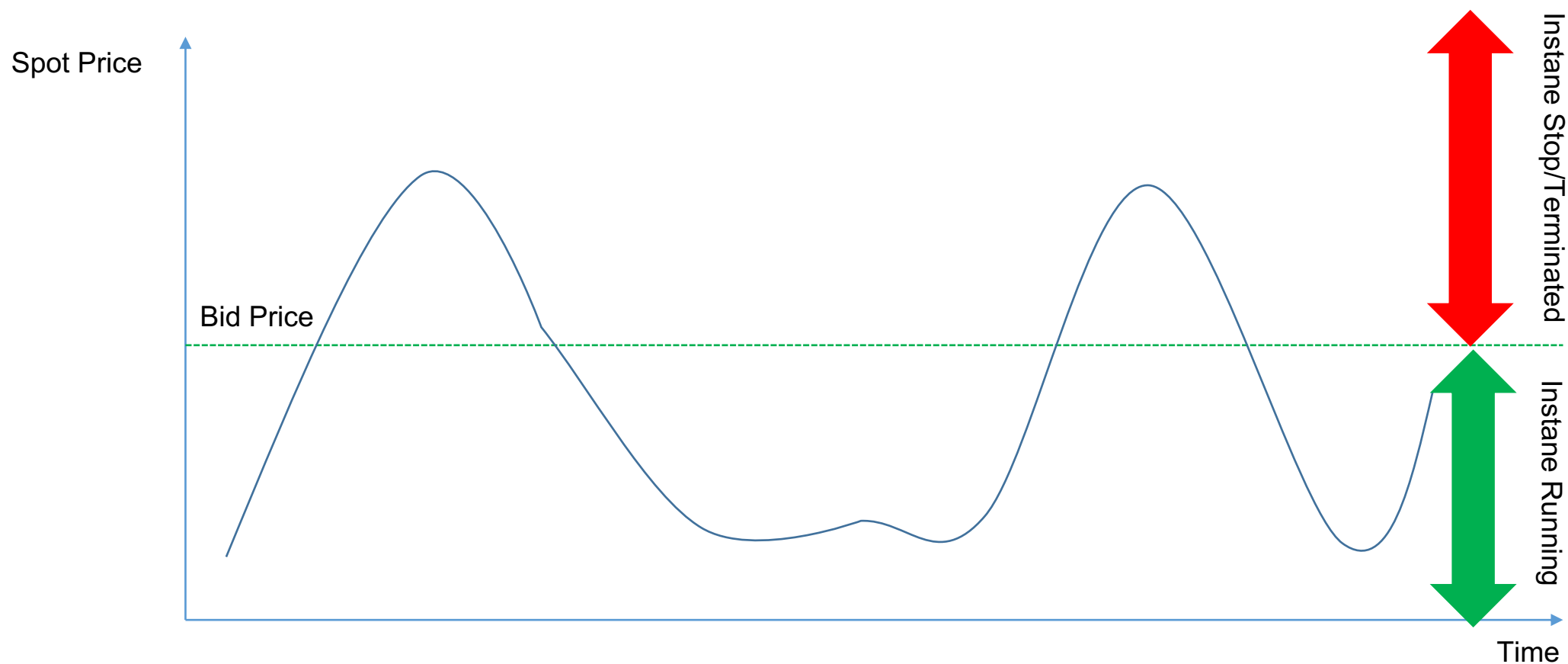
Spot Instance

- Spot Instance cho phép tiết kiệm lên đến 90% so với Instance chạy theo kiểu trả phí On-Demand
- Use cases
 - Dùng cho chạy Batch Job, Phân tích dữ liệu (Data Analysis)
 - Các ứng dụng sử dụng Spot Instance phải xử lý được việc Instance bị dừng đột ngột (Do giá Spot tăng hay do AWS bị hết tài nguyên cho Spot Instance)
 - Không phù hợp cho chạy các ứng dụng quan trọng (uptime 24/7) hay Database

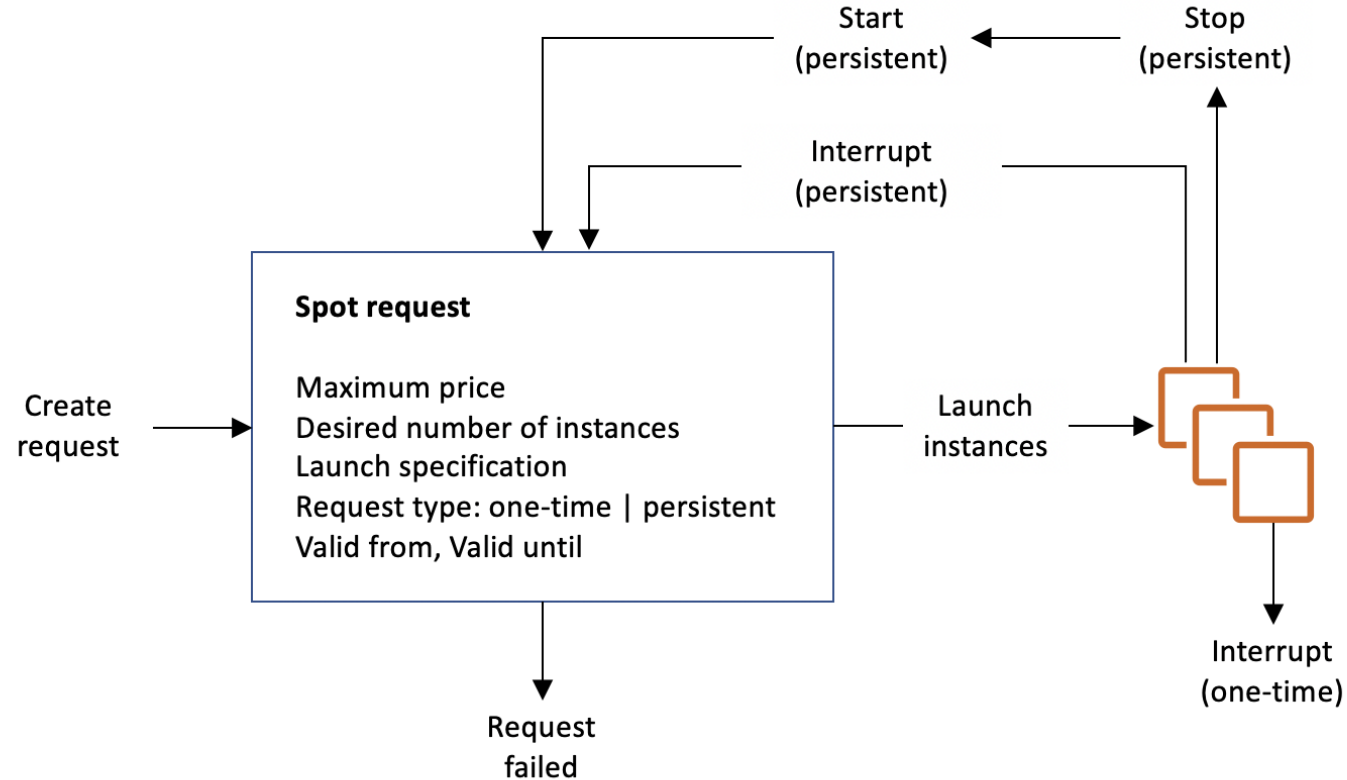
Spot Price

- **Spot Price** là giá cho 1 giờ chạy của một Spot Instance (hourly price for Spot Instance).
- **Spot Price** sẽ thay đổi theo quy luật cung cầu và tài nguyên còn
- **Bid price** (Max spot price) là giá mà bạn sẵn sàng trả cho một Spot Instance.
- Spot Instance sẽ được chạy khi giá **Spot Price** hiện tại < Giá **Bid Price** và tài nguyên của AWS đáp ứng đủ
- Spot Instance sẽ bị terminate hoặc stop khi giá **Bid Price** < giá **Spot Price** hiện tại hoặc tài nguyên của AWS cần thu hồi
- AWS sẽ thông báo 2 phút trước khi interrupt (terminate hoặc stop) Spot Instance

Spot Price (cont.)



Spot Instance Request



Spot Fleet

- Spot Fleets là tập hợp các Spot Instances + On-Demand Instances
- Sử dụng Spot Fleets để đáp ứng được công suất với mức chi phí phù hợp
 - Spot Fleet Request có thể chọn nhiều Spot Capacity Pool (Cùng Instance Type, OS, AZ, Network Platform)
 - Spot Fleet sẽ chọn một Spot Capacity Pool đáp ứng được Spot Fleet Request
 - Có thể set Max Price cho một Spot Fleet. Spot Fleet sẽ không chạy máy mới khi đạt tới mức giá Max Price này ngay cả khi chưa đáp ứng được công suất

Spot Fleet (cont.)

- Chiến lược cấp phát (Allocation Strategies)
 - Lowest price (default): Chọn Spot Pool nào có giá thấp nhất
 - Diversified: Chọn lựa nhiều Spot Pool khác nhau
 - Capacity Optimized: Chọn Spot Pool cho phép tối đa công suất

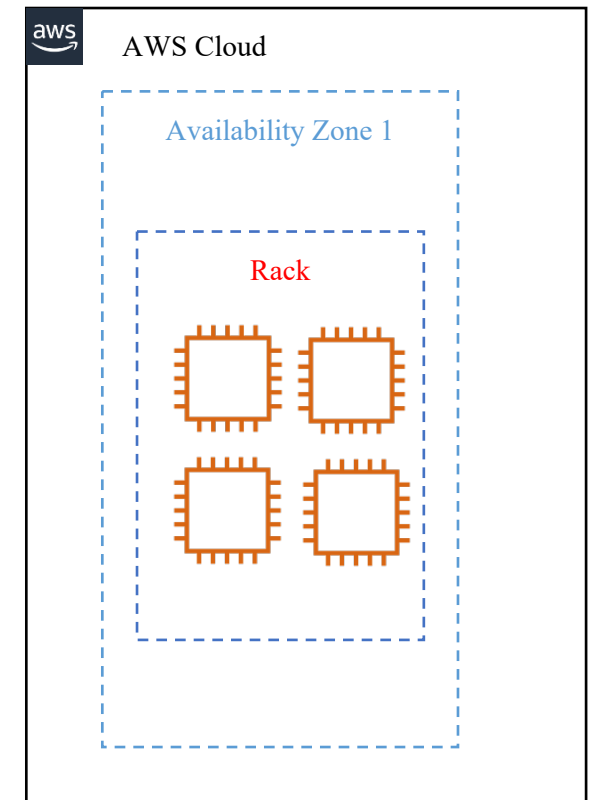
Placement Groups

Placement Groups?

- Mặc định Amazon EC2 sẽ đặt các Instances theo cách để làm giảm thiểu mức độ ảnh hưởng của sự cố
- Placement Groups cho phép kiểm soát cách đặt các Instances vào các cụm Host vật lý
- Sử dụng PG khi các EC2 Instances cần giao tiếp với nhau yêu cầu độ trễ thấp (Low latency) và Throughput cao
- Placement Groups Strategy
 - Cluster
 - Partition
 - Spread

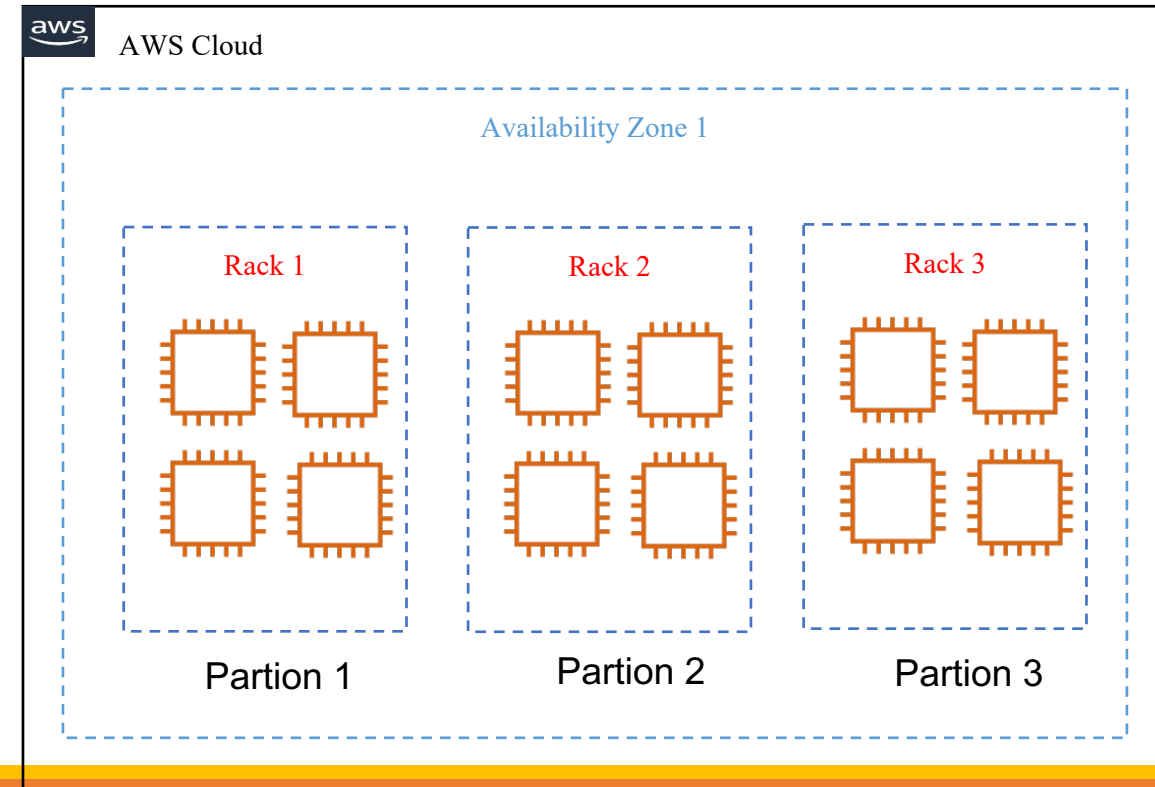
Placement Groups – Cluster

- Đặt tất cả các EC2 Instances ở trên cùng một Rack
- Thông lượng cao (~10Gbps), độ trễ thấp
- Rủi ro mất cả cụm EC2 Instances khi Rack này bị sự cố
- Use case: Bigdata, Low-latency app



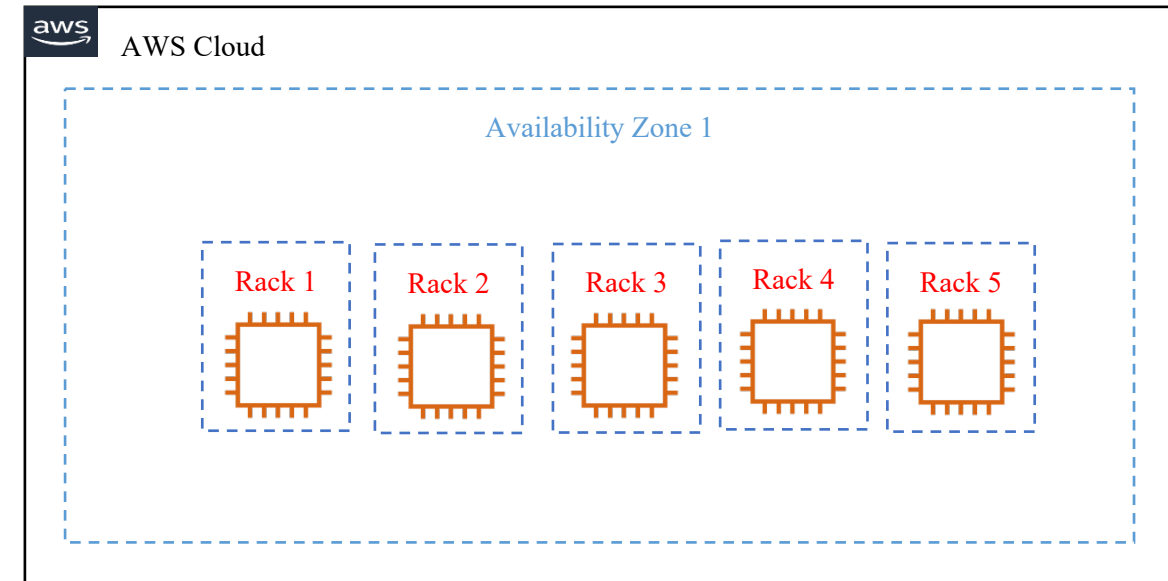
Placement Groups – Partition

- Instances trong cùng một Partition thì sẽ nằm ở các Rack khác nhau
- Các Partition thì có thể nằm trên nhiều AZ trong cùng 1 Region
- Up to 7 Partitions trong một AZ
- Use case: HDFS, Cassandra...



Placement Groups – Spread

- Mỗi Instance được đặt mỗi Rack riêng
- Có thể trải trên nhiều Azs trong cùng 1 Region
- Tối đa 7 Running Instances trên mỗi AZ mỗi Group
- Use case
 - Ứng dụng cần tối tính HA cao
 - Ứng dụng quan trọng, sự cố của một Instance không được ảnh hưởng tới Instance khác



EC2 hibernate

EC2 hibernate?

Khi start Instances, những điều sau đây được thực hiện

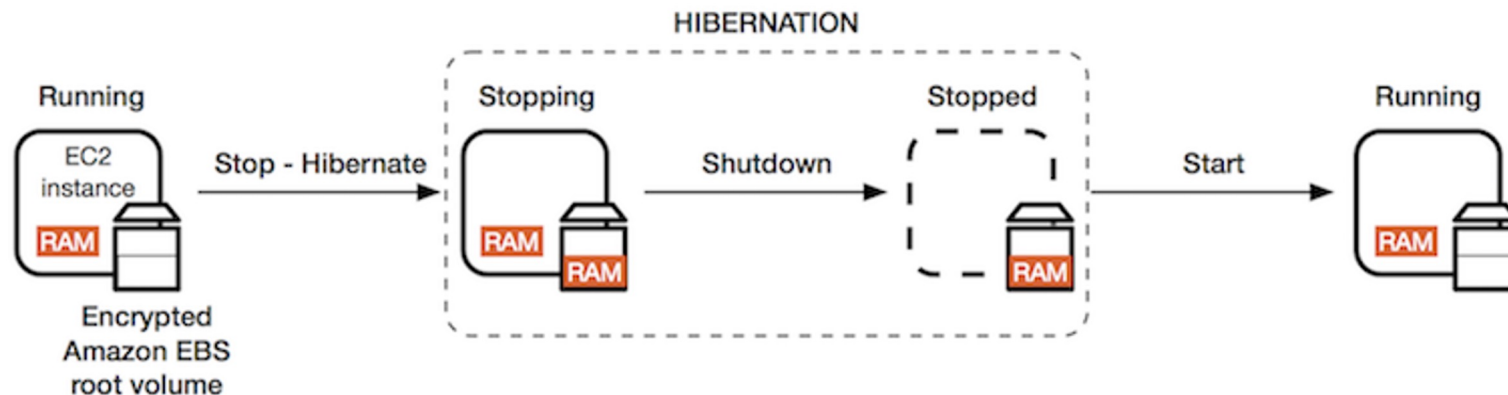
- OS boots up
- Run userdata script (Trong trường hợp chạy lần đầu – First start)
- Start và Load applications vào RAM.

EC2 hibernate?

- Nội dung bên trong RAM được lưu trữ tại EBS Root Volume
- EBS Root Volume phải được mã hoá
- Giúp tăng tốc việc Boot up Instances

Start Instances với EC2 hibernate

- Dữ liệu được load vào RAM. Các ứng dụng được Resume
- EBS Root Volume được khôi phục lại tại trạng thái trước khi xảy ra Hibernate



EC2 Hibernate use case

- Dành cho các tiến trình chạy dài hạn (Long-running process)
- Sử dụng để tăng tốc độ Boot Up dịch vụ với các dịch vụ cần nhiều thời gian để khởi tạo

Exam tips

- Instance RAM size < 150 GB
- Hỗ trợ các Instance families: C3, C4, M3, M4, R3, R4, R5
- Hỗ trợ AMI: Amazon Linux 2, Linux AMI, Ubuntu and Window
- Root Volume cần được mã hoá
- Không thể ở chế độ Hibernate trong vòng > 60 ngày