



CODESTAR

Virtual Machine & Instance

CodeStar Academy

Nội dung chính

1. Virtual Machine (VM) & Instance
2. AMI
3. EC2 Status
4. ENI & IP
5. Pricing Model
6. Placement Group

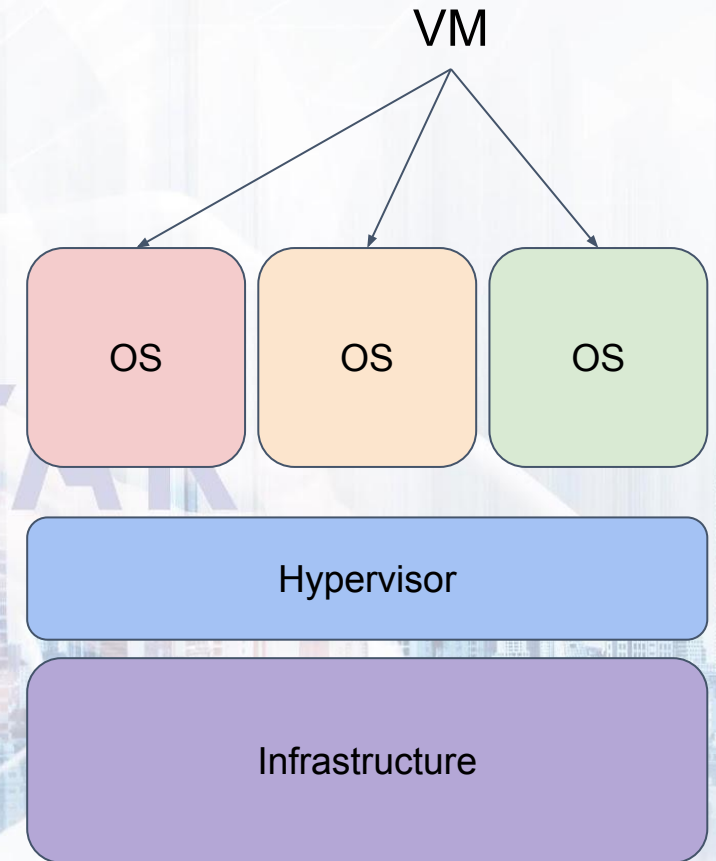


CODESTAR

Virtual Machine (VM) & Instance

Mô hình máy ảo (VM):

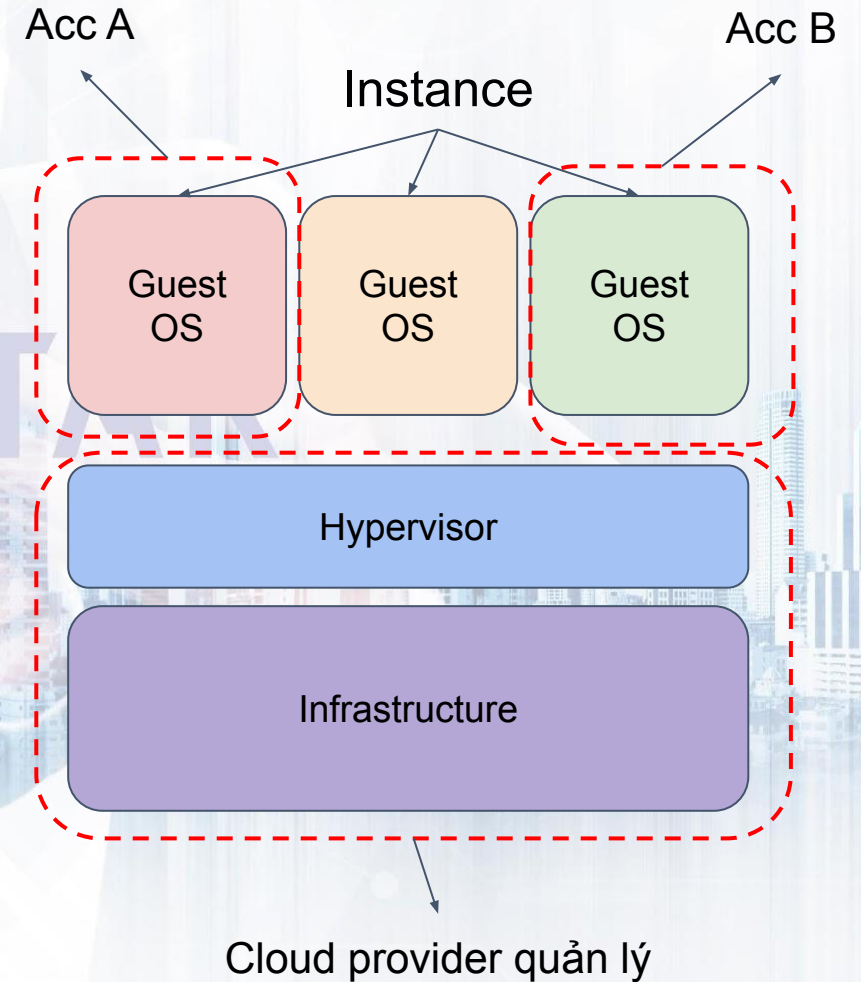
- Mô hình thiết lập nhiều máy ảo, nhiều hệ điều hành trên cùng một thiết bị phần cứng.
- Sử dụng Hypervisor để phân chia phần cứng.
- Mỗi hệ điều hành được phân chia một lượng resource khác nhau từ Infrastructure và sử dụng



Virtual Machine (VM) & Instance

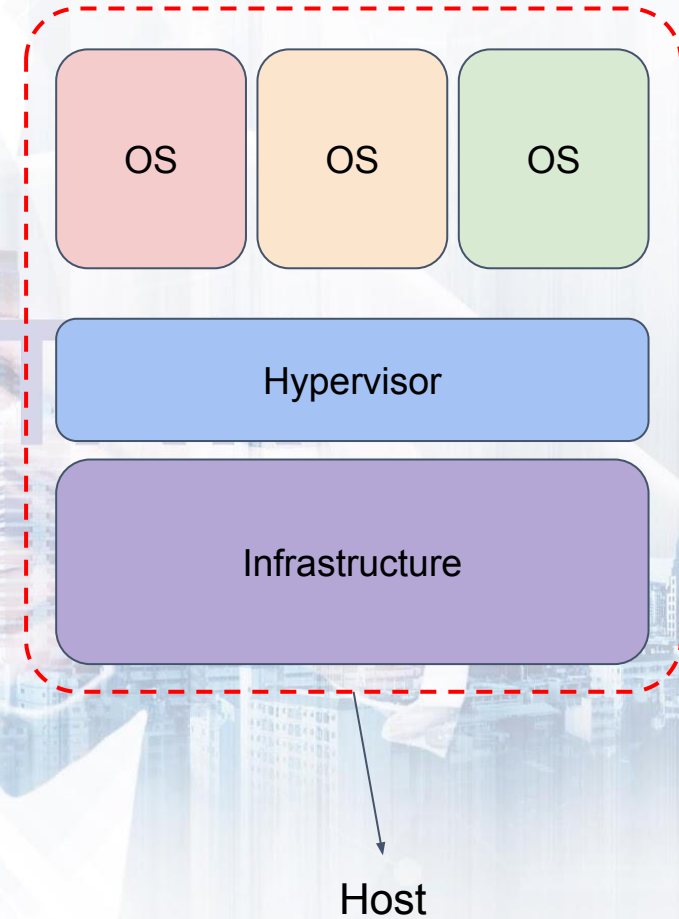
Instance là concept thường gặp trên Cloud Computing

Về cơ bản, Instance chính là một VM, chỉ có điều VM này được đặt trên các cloud provider.



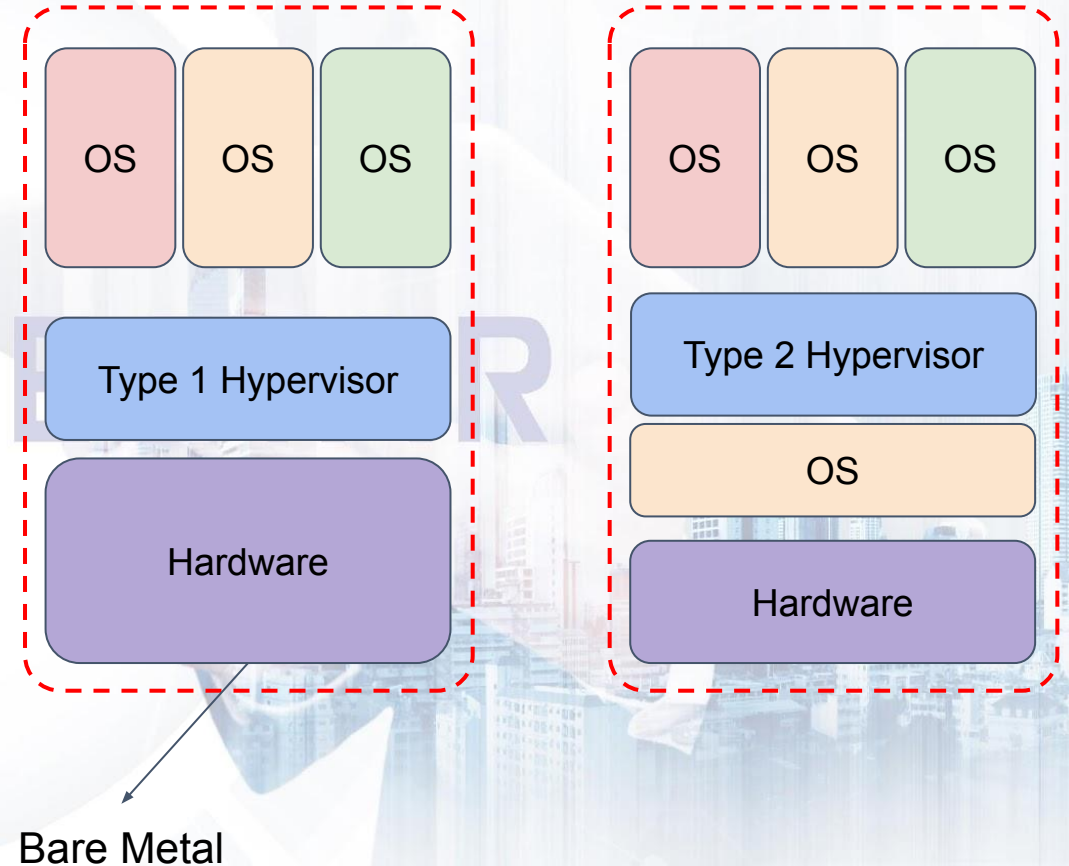
Virtual Machine (VM) & Instance

- Thiết bị phần cứng dùng để chứa các tài nguyên của Instance được gọi là **host**.
- Chúng ta phân biệt host với VM và Instance. Trong một số trường hợp VM và instance có thể được dùng thay thế cho nhau.



Virtual Machine (VM) & Instance

- Hypervisor là thành phần nằm giữa phần cứng (Hardware) và các Instance, giúp khởi tạo và quản lý các instance/VM.
- Các loại Hypervisor gồm:
 - + Type 1 Hypervisor
 - + Type 2 Hypervisor

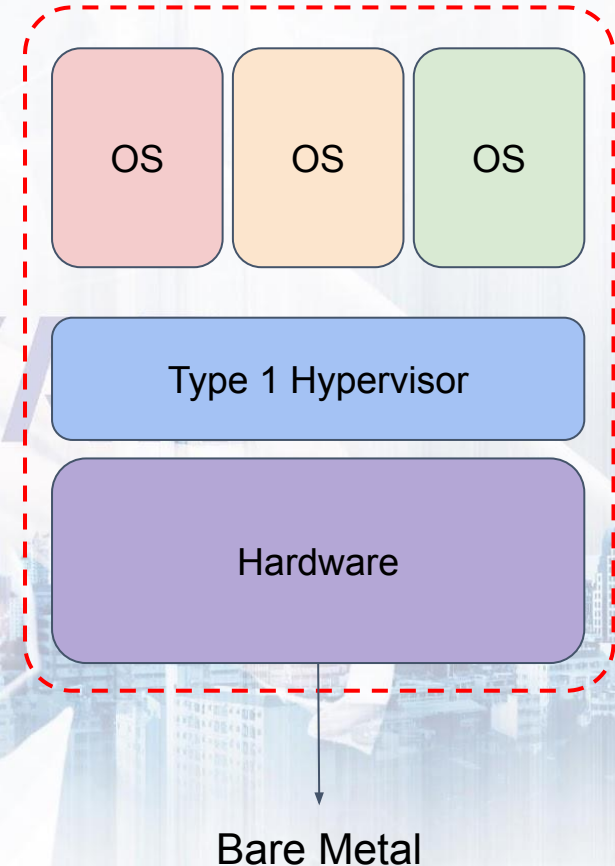


Virtual Machine (VM) & Instance

- Type 1 Hypervisor (Bare metal hypervisor)

Một số loại hypervisor thuộc nhóm này:

- + VMware vSphere / ESXi
- + Hyper-V
- + Xen / Citrix XenServer
- + Red Hat Enterprise Virtualization (RHEV)
- + KVM

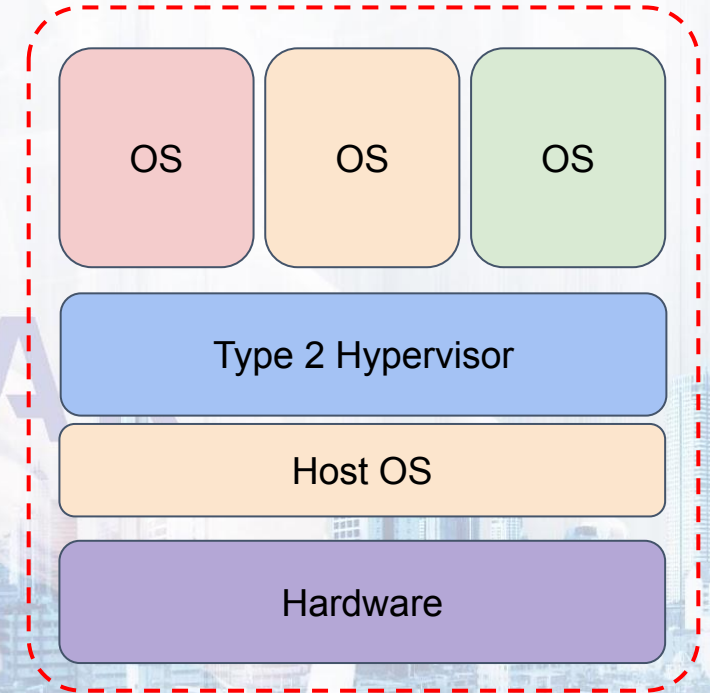


Virtual Machine (VM) & Instance

- Type 2 Hypervisor: (Hosted Hypervisor)

Một số loại hypervisor thuộc nhóm này:

- + Oracle Virtual Box
- + VMware Workstation
- + Oracle Solaris Zones
- + VMware Fusion
- + Oracle VM Server for x86



Virtual Machine (VM) & Instance

vCPU:

vCPU là một đơn vị đo khả năng của một Chip nói chung, được tính bằng:

Số luồng * Số Core xử lý * Số CPU

VD một CPU có 2 nhân, 4 luồng, 4 bộ CPU

-> Chúng ta có $2 * 4 * 4 = 32$ vCPU.

Một máy host nói chung, có thể được chia thành nhiều vCPU và cung cấp cho các instance khác nhau. Các EC2 instance sử dụng cùng một máy host được gọi là **shared instance** (tenancy)

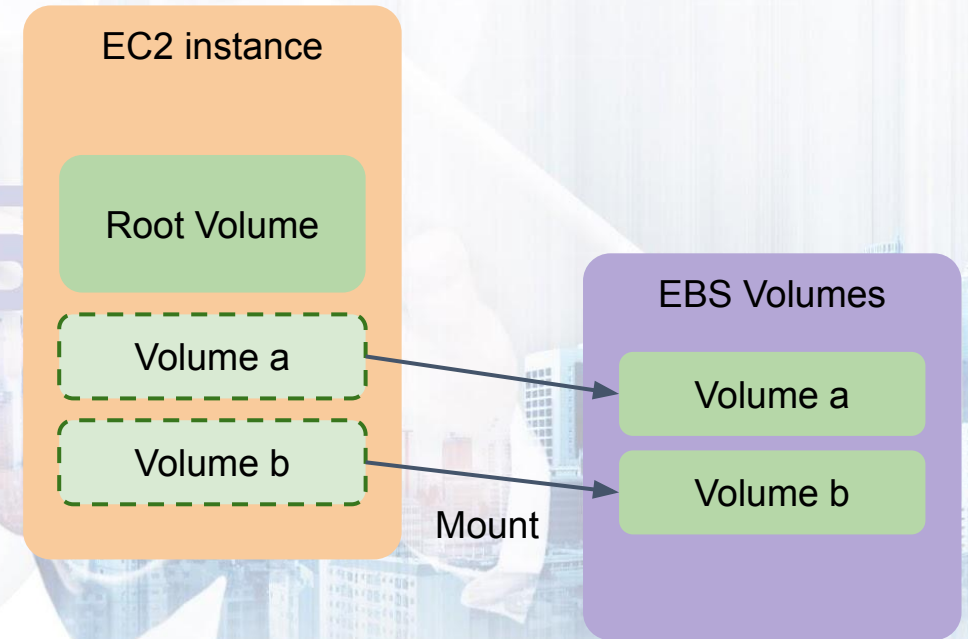
Virtual Machine (VM) & Instance

Root Volume:

EC2 instance sẽ luôn kèm theo tối thiểu một Volume để có thể chứa hệ điều hành (tương tự như Windows sẽ cần ổ C, Linux sẽ cần thư mục gốc / để chứa các file hệ điều hành)

=> Được gọi là Root Volume

Có thể Mount thêm các EBS Volume khác vào instance để sử dụng.



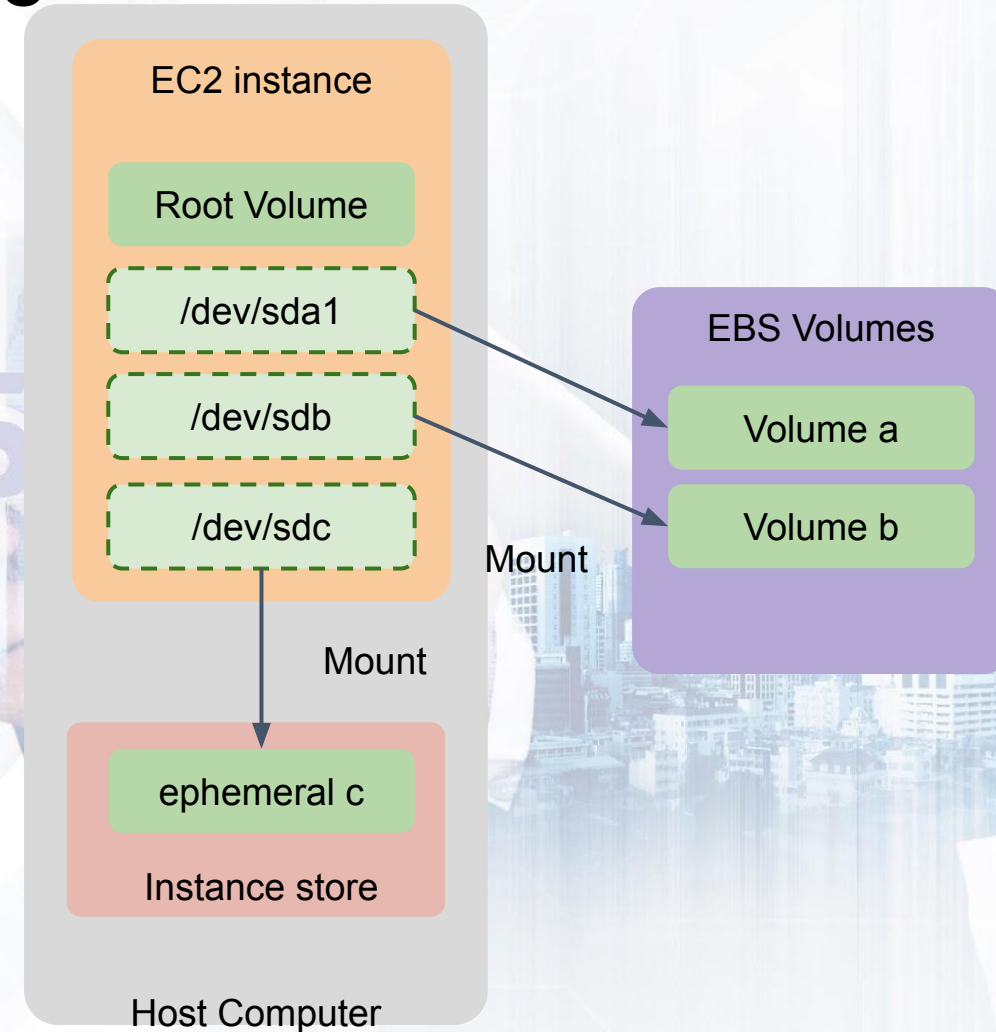
Virtual Machine (VM) & Instance

Instance store vs EBS:

Các instance lưu trữ thông tin dưới dạng các block device mapping.

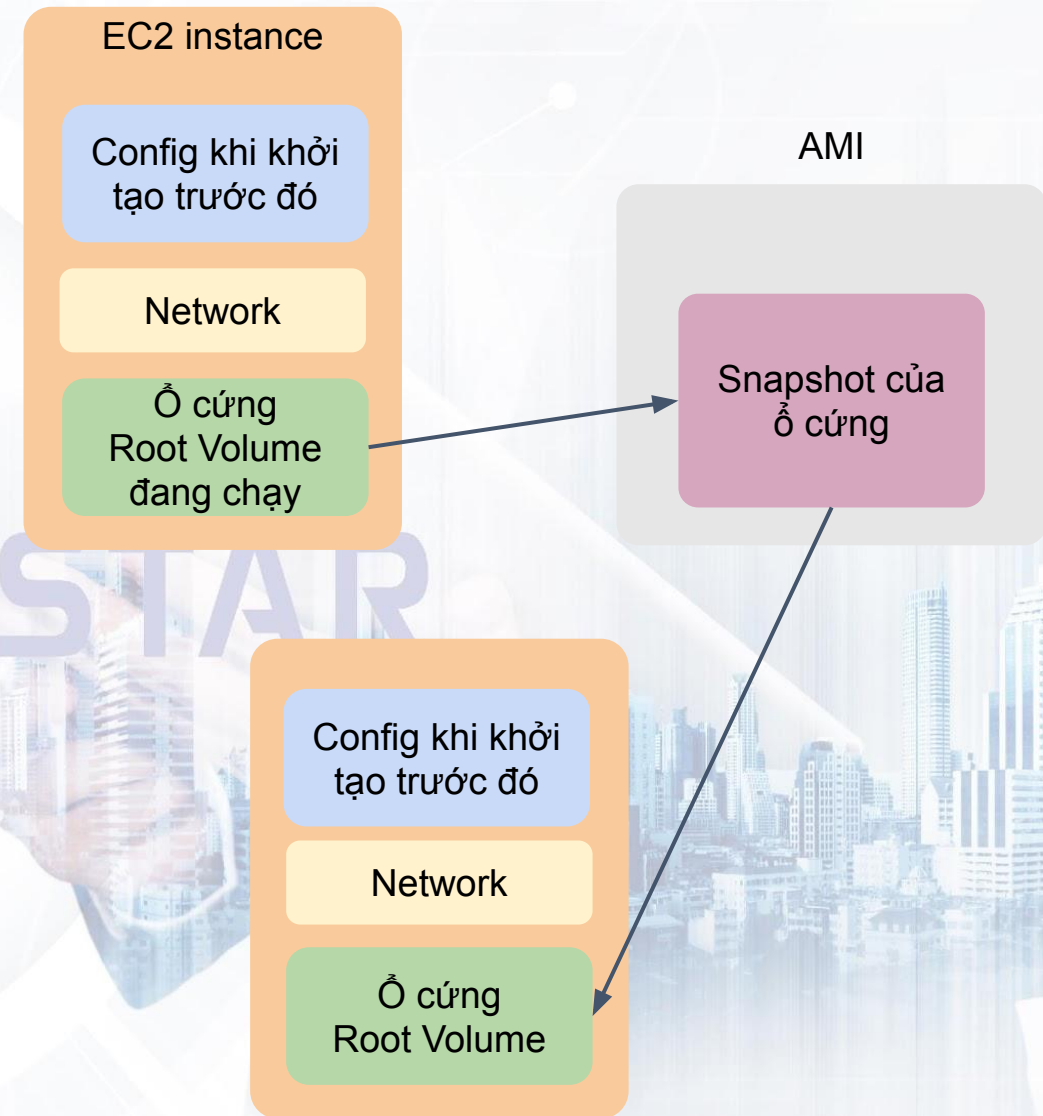
Instance store được lưu trữ trên chính host computer, được gọi là *ephemeral*.

EBS Volumes sẽ được lưu trữ trên các volume phía bên ngoài host computer, định dạng *vol-xxxxxxx*.

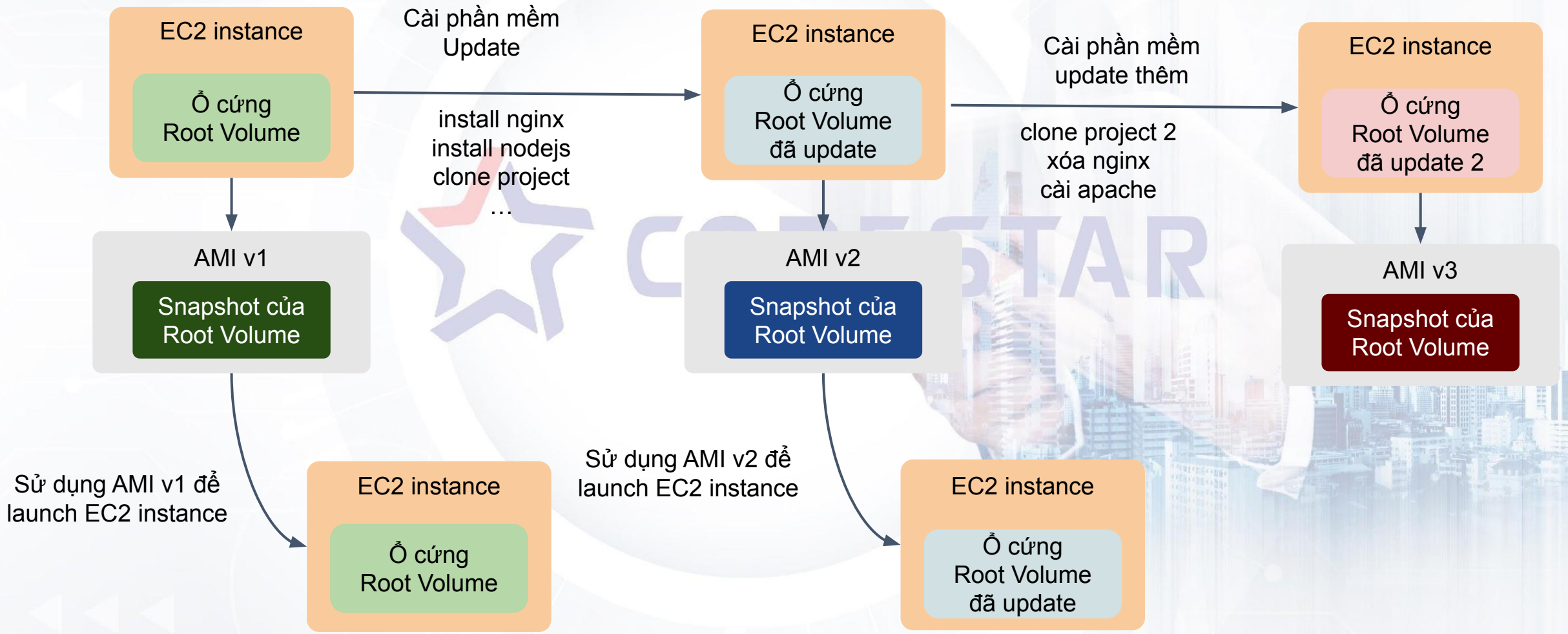


AMI

- **AMI** là một image, lưu trữ trạng thái hiện tại của Root Volume + EBS (máy host).
- **AMI** sẽ tương đương với 1 bản zip toàn bộ hệ thống hiện tại (chủ yếu bao gồm các thông tin về OS) đang chạy trong hệ thống.



AMI



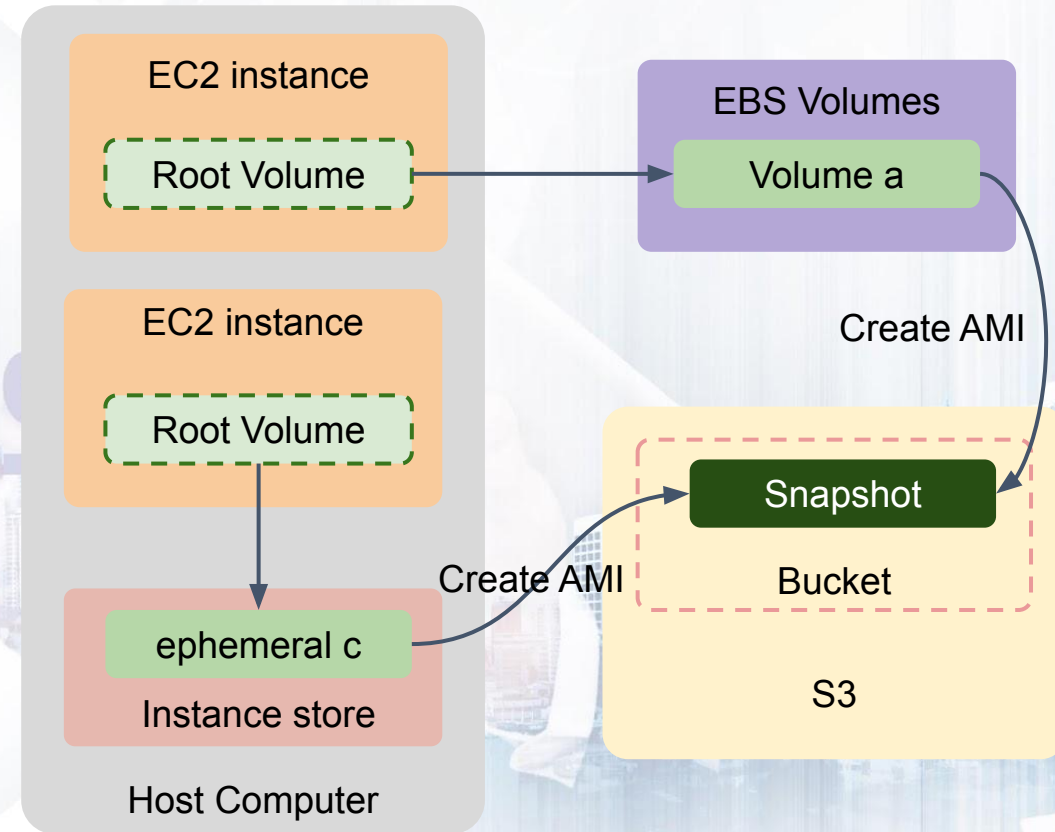
AMI

- Storage cho Root Device:

Phần Storage của AMI được chia thành 2 loại:

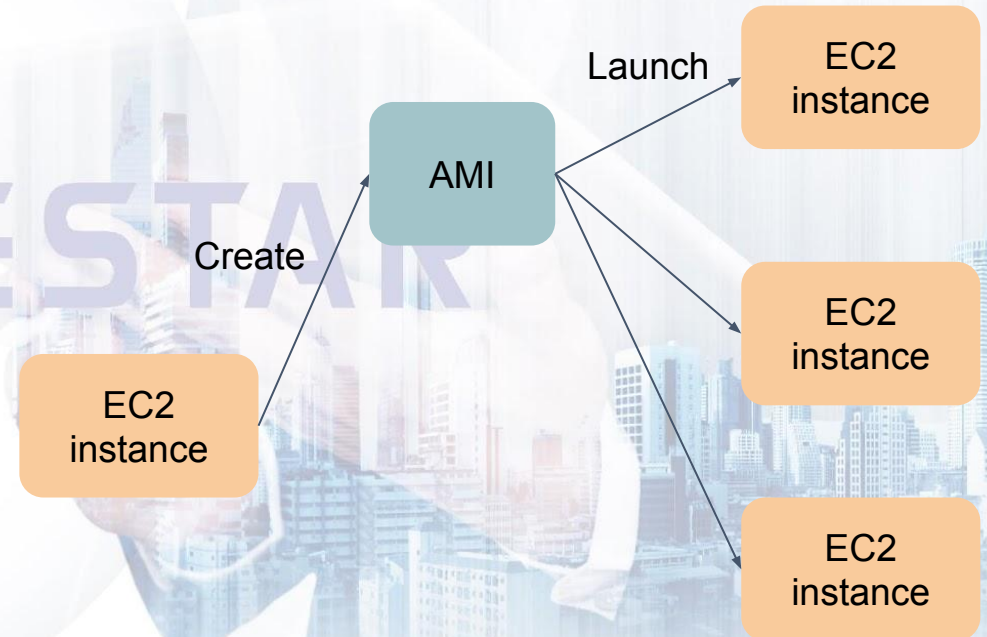
- **EBS-backed AMI:** Lưu trữ bằng EBS Snapshot
- **instance-store backed AMI:** Lưu trữ bằng S3
(Lưu ý tối đa size: 10GB)

2 loại AMI này có thể được chuyển đổi cho nhau.



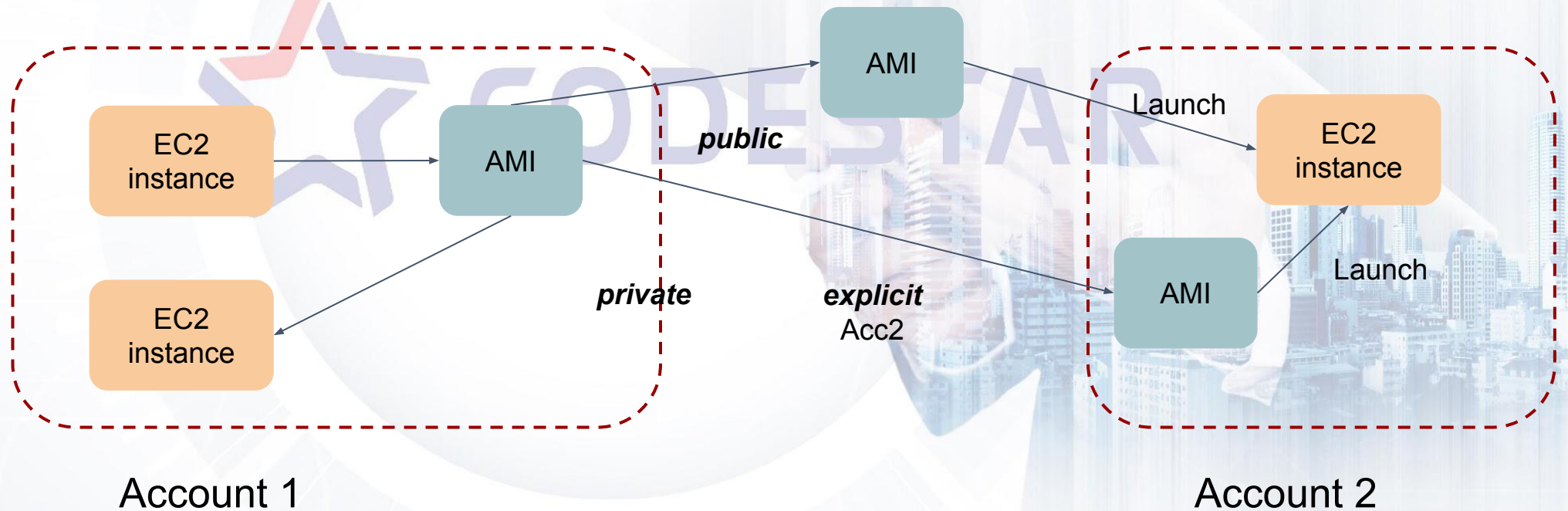
AMI

- AMI giúp chúng ta có thể tạo ra nhiều Instance với nội dung giống nhau bên trong.
- Nếu chúng ta là AMI owner, chúng ta có thể share cho các tài khoản khác với Launch permission.



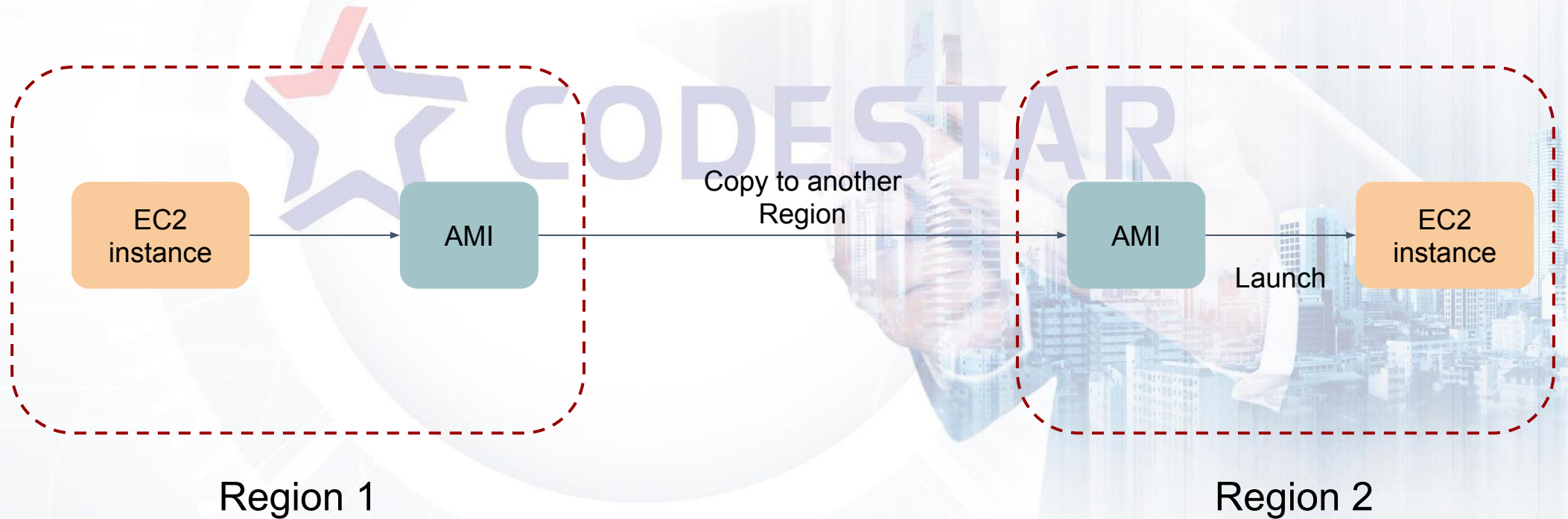
AMI

- Use case sử dụng với launch permission



AMI

- Copy AMI sang Region khác: Do AMI đặc thù cho việc khởi tạo instance cho riêng AZ đó, nên khi chuyển Region, chúng ta cần Copy sang

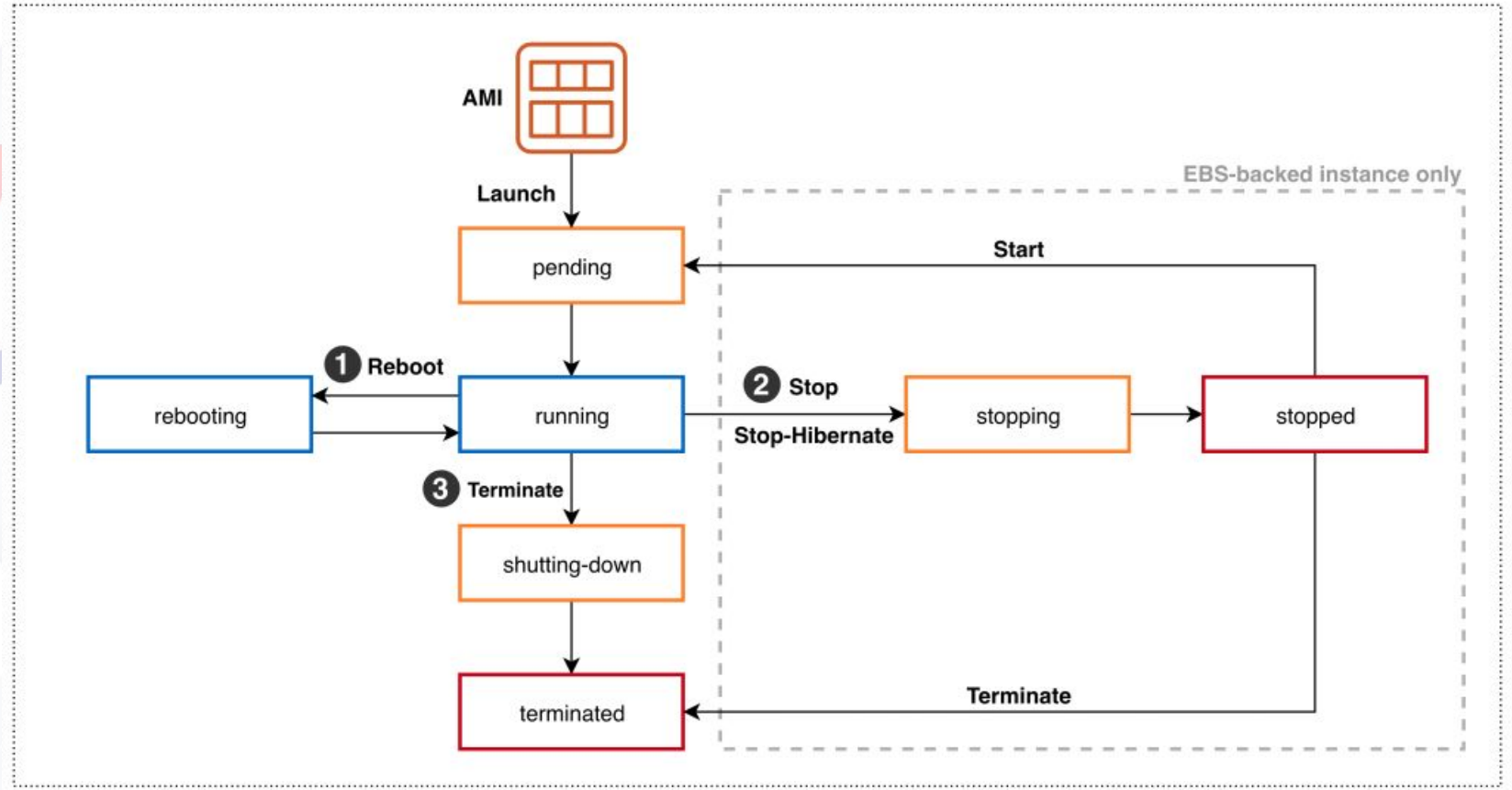


EC2 Instance Lifecycle

EC2 bị tính phí khi ở trạng thái running.

Khi ở trạng thái stopped, EC2 instance sẽ không bị tính phí, nhưng EBS volume sử dụng cho Root volume vẫn bị tính phí.

Amazon EC2 instance

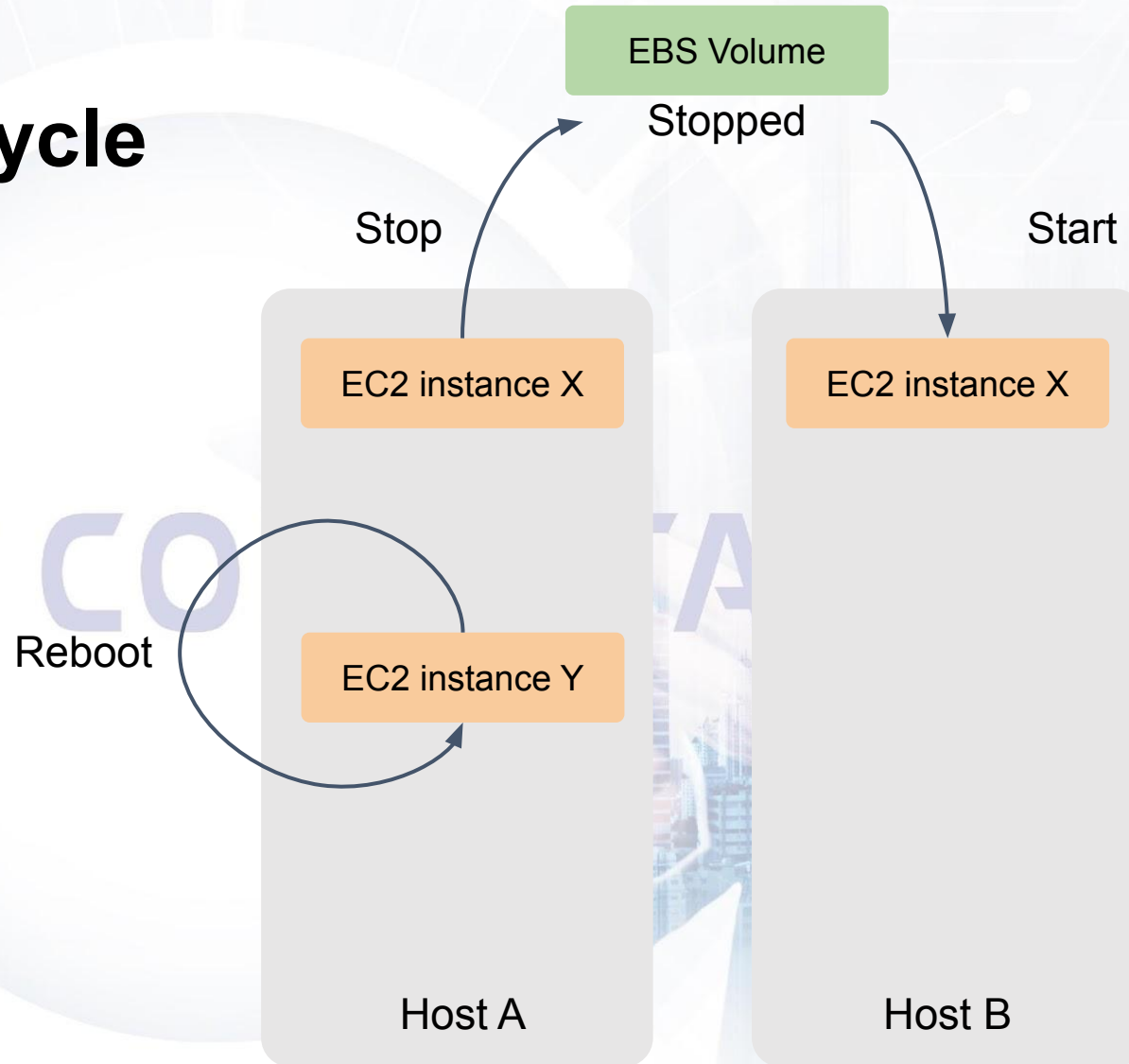


EC2 Instance Lifecycle

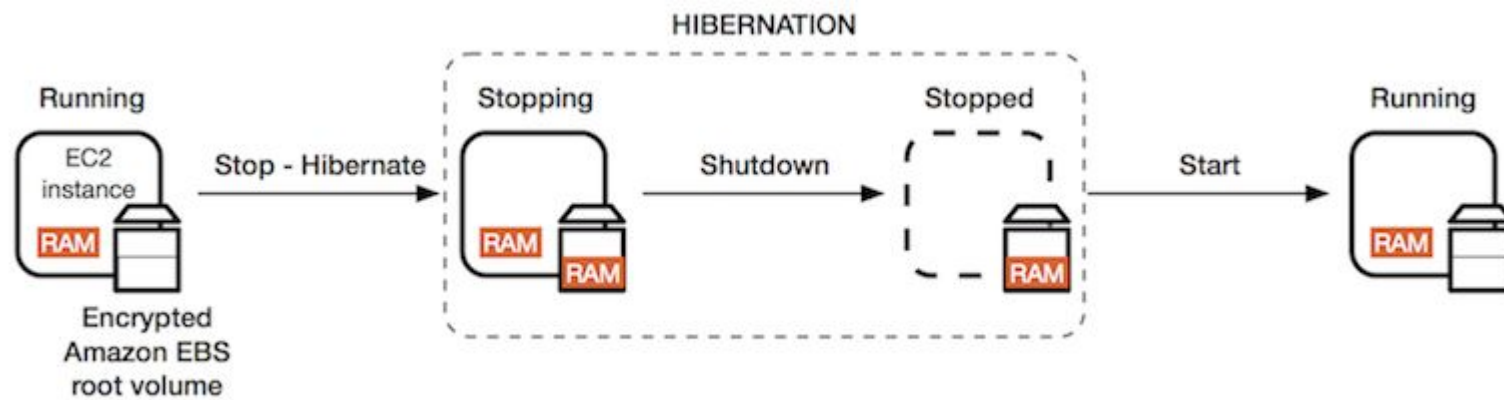
Reboot vs Stop & Start

Khi Reboot, EC2 instance vẫn nằm trên máy host trước đó.

Khi Stop, instance sẽ được release khỏi host. Sau đó khi Start, instance sẽ được khởi tạo tại một host mới.



EC2 Instance Lifecycle



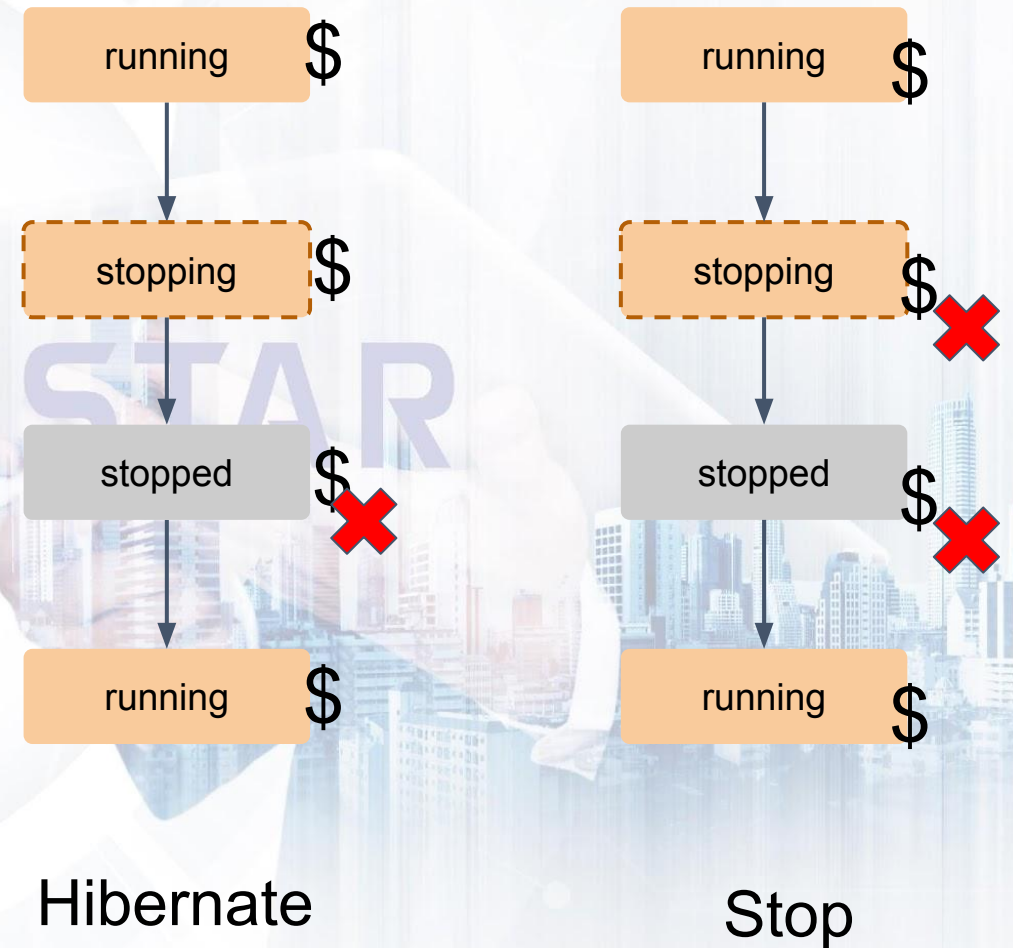
Hibernate vs Stop

- Khi thực hiện Hibernate, các thông tin ở trong RAM được lưu trữ vào EBS root volume (đảm bảo còn đủ dung lượng), sau đó EC2 instance sẽ được release (không tính phí) cho tới khi bật lại.

EC2 Instance Lifecycle

Hibernate vs Stop

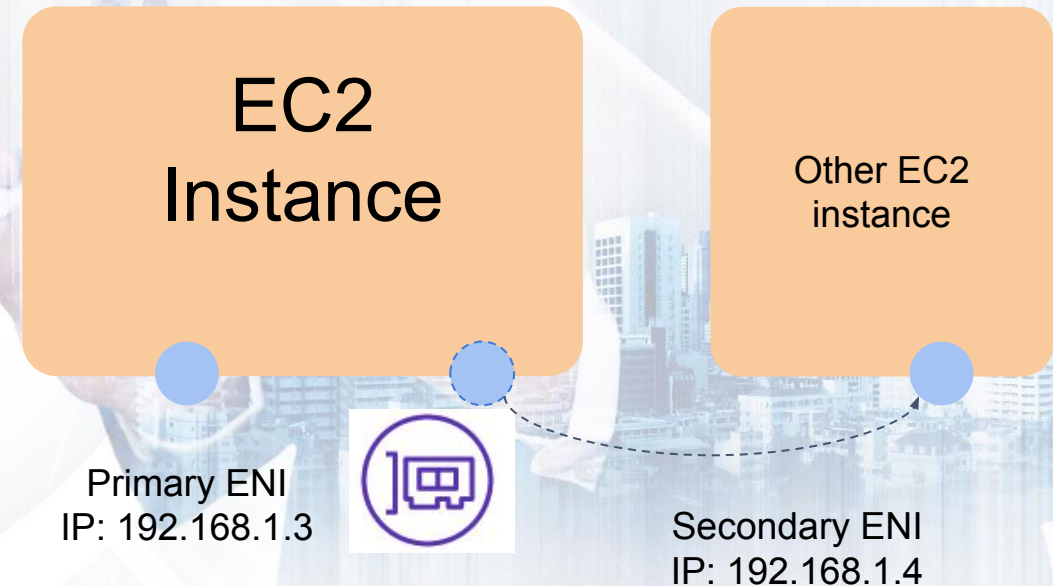
- Khi Hibernate, chúng ta sẽ vẫn bị tính phí khi instance Stopping, và dừng tính phí khi chuyển sang Stopped. Sau khi start lại, phí sẽ được tính kể từ khi instance chuyển sang trạng thái Running.
- Với trường hợp Stop, chúng ta sẽ không bị tính phí khi Instance ở trạng thái Stopping.



ENI & IP

EC2 Instance và các thành phần VM nói chung, được gắn một thành phần tên là Elastic Network Interface (ENI), là một card mạng ảo, được sử dụng để có thể giao tiếp với bên ngoài.

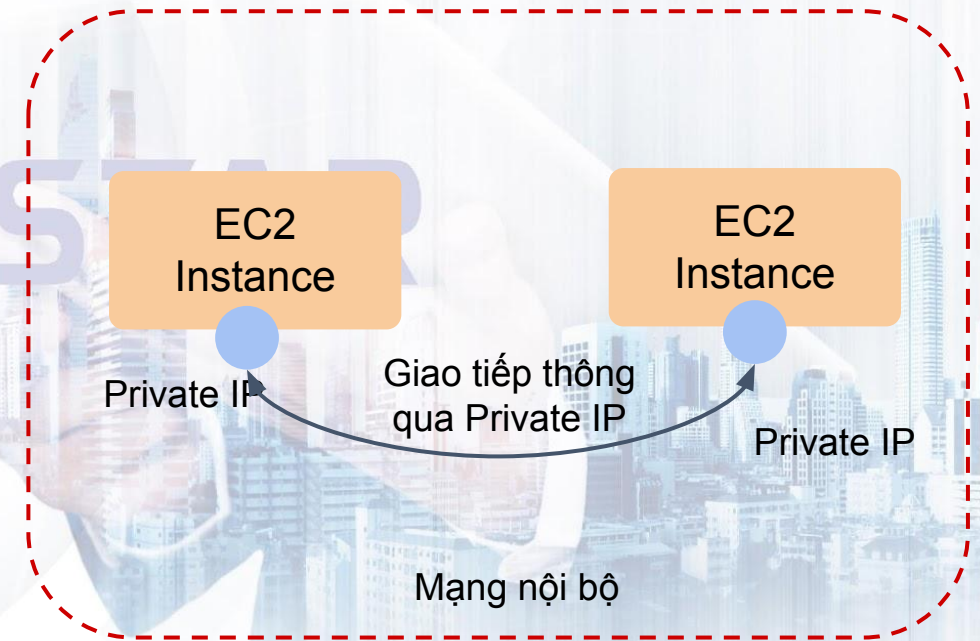
- EC2 có thể gắn nhiều ENI, và sử dụng vào các Network khác nhau.



ENI & IP

ENI có thể được gắn các loại IP sau:

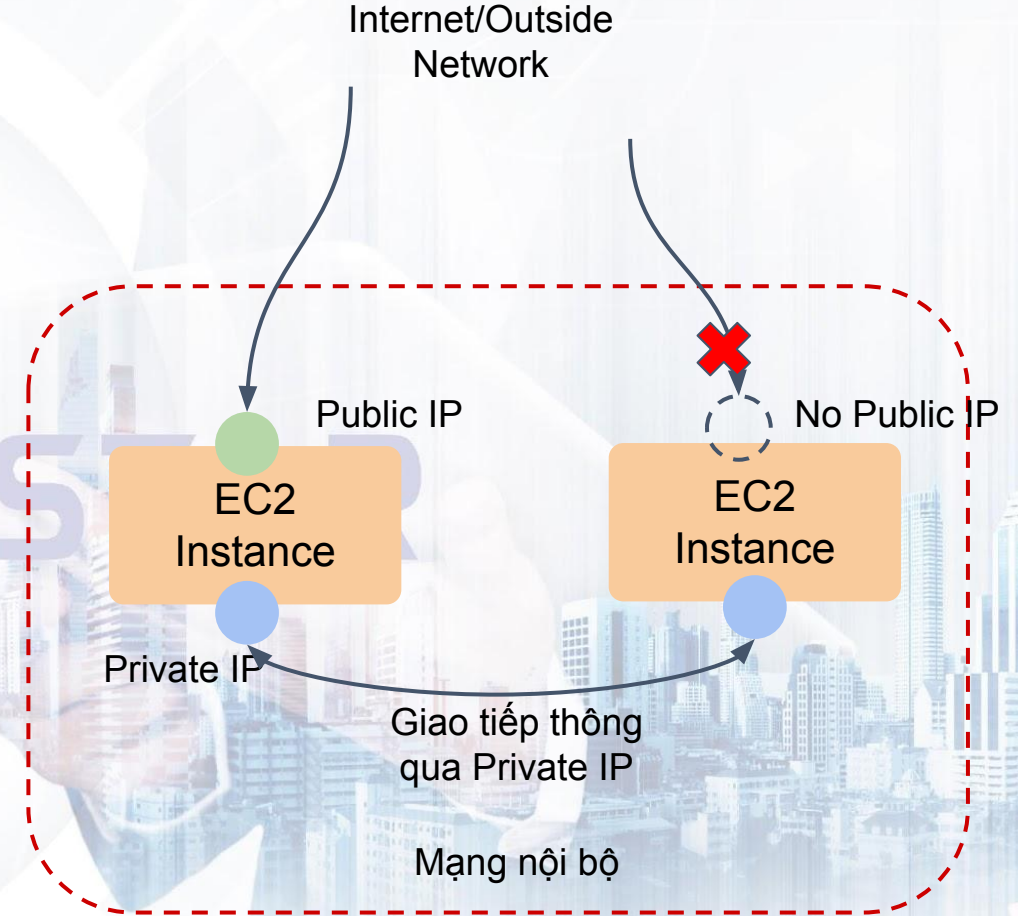
- **Private IP:** được gắn vào EC2 instance trong mọi trường hợp. Đây là IP tuân theo config của Network đặt trong nó.
- Song song với Private IP, Instance cũng sẽ được gắn một private DNS, cho phép một DNS nội bộ trở tới Instance đó.



ENI & IP

ENI có thể được gắn các loại IP sau:

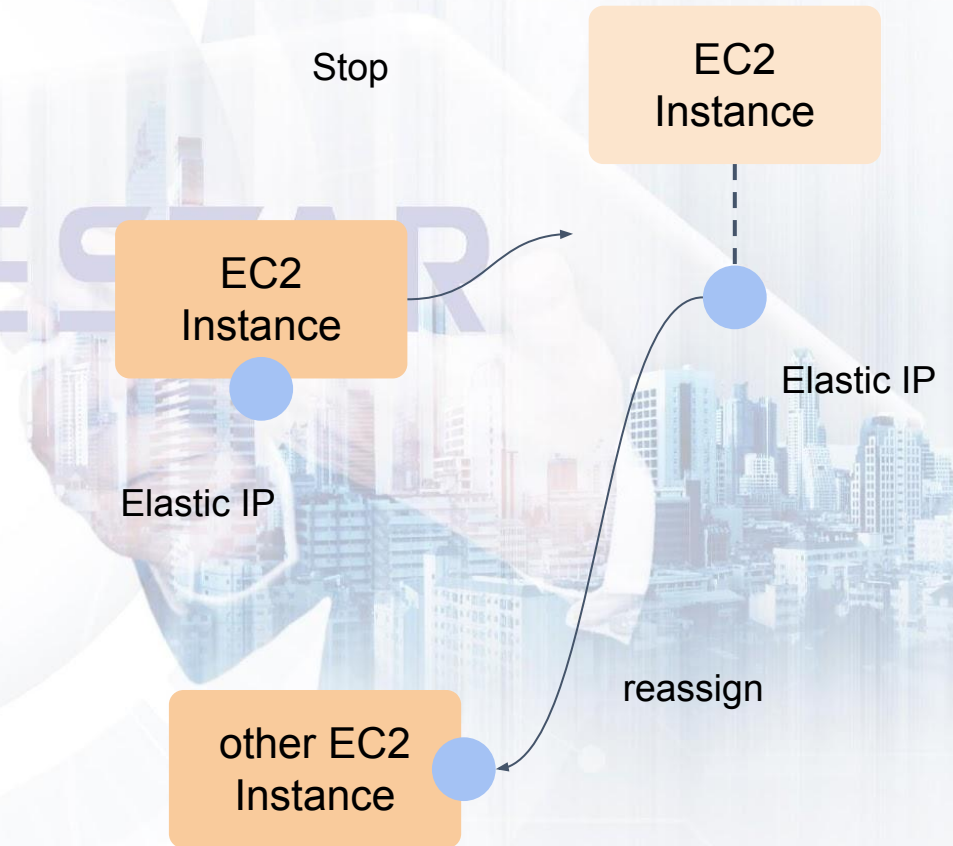
- **Public IP:** được sử dụng khi có thiết lập Public Access và có thể được truy cập từ Internet (nếu được phép từ SG).
- Public IP không nhất thiết phải có khi khởi tạo Instance.
- Bị tính phí khi gắn vào instance.



ENI & IP

ENI có thể được gắn các loại IP sau:

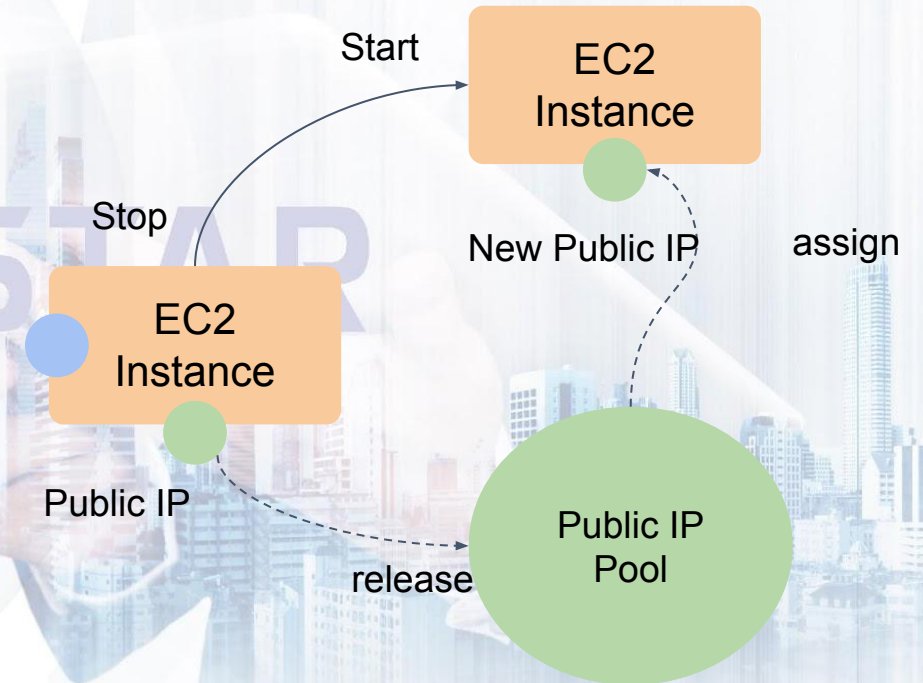
- **Elastic IP:** Tương tự Public IP, nhưng có thể gỡ ra, gắn vào một instance khác.
- Vẫn bị tính phí ngay cả khi không dùng EC2 instance. (phí sử dụng Elastic IP).
- Tài khoản khi mới sử dụng bị giới hạn 5 Elastic IP, nếu muốn tăng thêm, cần Request tăng Quota từ AWS Support.



ENI & IP

Các loại IP trong instance lifecycle:

- Private IP (Primary ENI) sẽ được tồn tại trong suốt vòng đời của EC2 instance.
- Public IPv4 sẽ được release và re-assign trong đại đa số các trường hợp Instance bị Stopped, Hibernate, Terminate, ...
- IPv6 (bao gồm Private và Public IPv6) sẽ không thay đổi trong toàn bộ instance lifecycle.



ENI & IP

Public IP của một instance

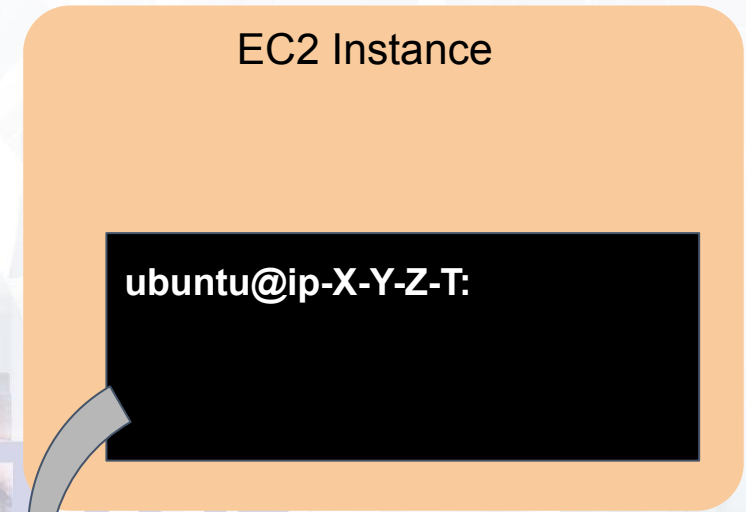
Public IP của một EC2 instance có liên hệ tới trạng thái của EC2 instance đó.

Reboot: Public IP liên kết tới EC2 instance không đổi

Stop: Public IP liên kết tới EC2 instance sẽ bị đổi. Do đó, chúng ta cần lưu ý sử dụng Elastic IP nếu cần sử dụng đến IP của một EC2 instance.

Metadata

- Metadata là những thông tin đặc trưng của riêng instance đó, như private IP, public IP, Startup time, ...
- Khi cần truy xuất những thông tin này trong chính instance hiện tại, chúng ta tạo một request tới các địa chỉ đặc biệt.



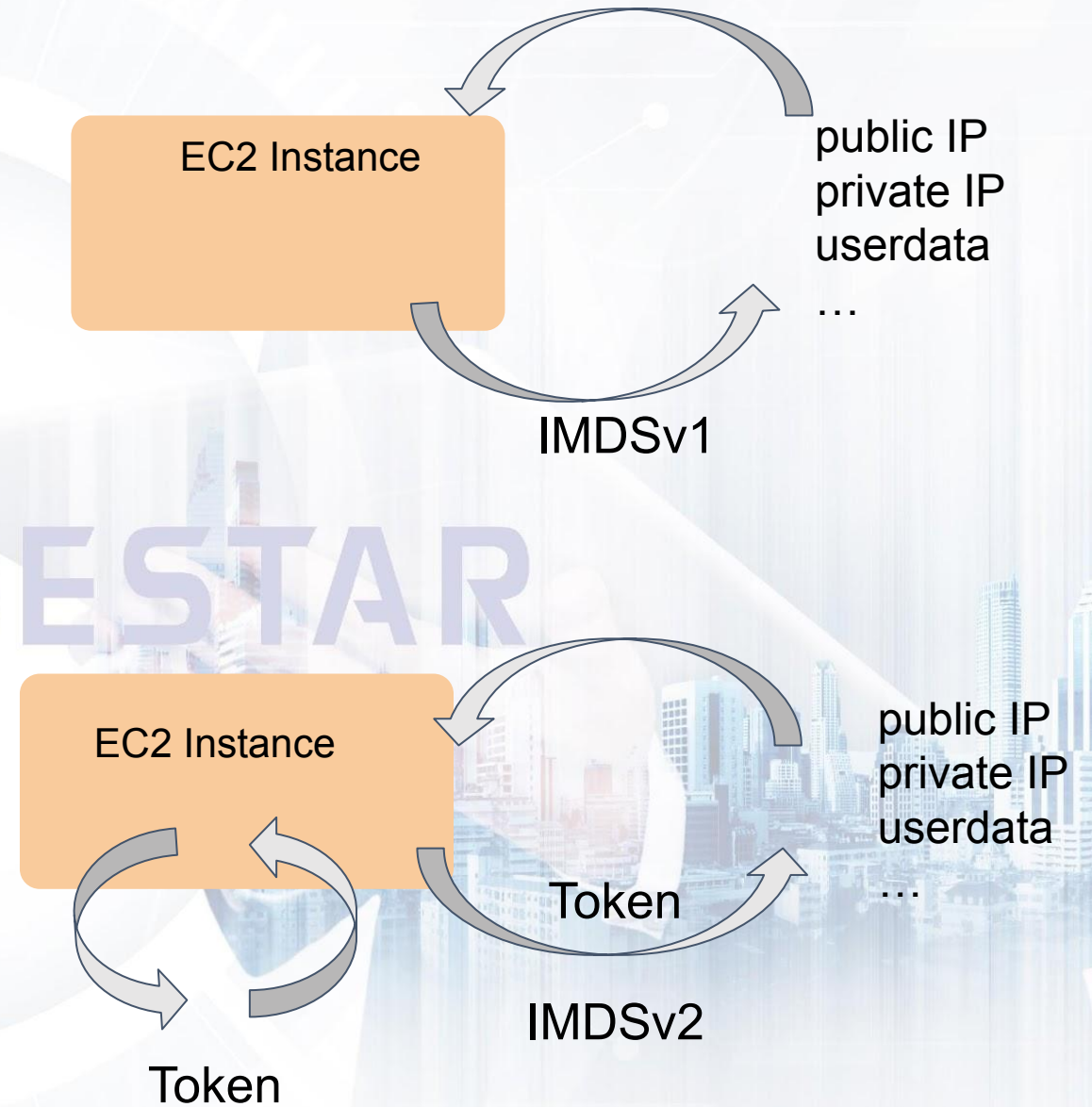
public IP
private IP
userdata
...

IPv4 <http://169.254.169.254/latest/meta-data/>

IPv6 [http://\[fd00:ec2::254\]/latest/meta-data/](http://[fd00:ec2::254]/latest/meta-data/)

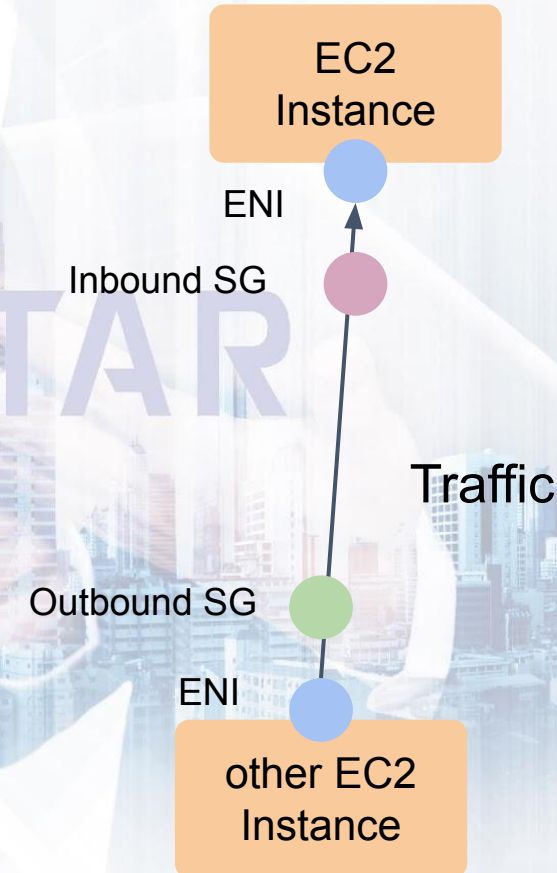
Metadata

- AWS cung cấp 2 phương thức sử dụng metadata, được gọi là IMDSv1 (request/response) và IMDSv2 (sử dụng session)



Security Group

- Security Group có thể được gắn trên ENI gắn vào EC2 instance để tạo một tầng Firewall cho instance
- Security Group (inbound) cho phép một số IP/SG được đi vào và (outbound) có thể cho phép đi tới một số IP/SG khác.
- Cùng một SG có thể gán cho nhiều instance khác nhau.
- Một instance có thể được gán nhiều SG khác nhau.



Security Group

Ở hình bên, chúng ta thấy, traffic từ instance 2 tới instance 1 sẽ phải đi qua Outbound SG của ENI 2 và Inbound SG của ENI 1.

Traffic được đi ra khỏi sg-xxxxxx (allow all nếu đi tới sg-aaaaaa).

Traffic được đi vào sg-aaaaaa chỉ với TCP:80, còn lại không được phép vào.

=> Instance 2 chỉ được gọi TCP:80 tới Instance 1.

Cả 2 instance đều có thể được SSH tới từ IP 172.168.5.6

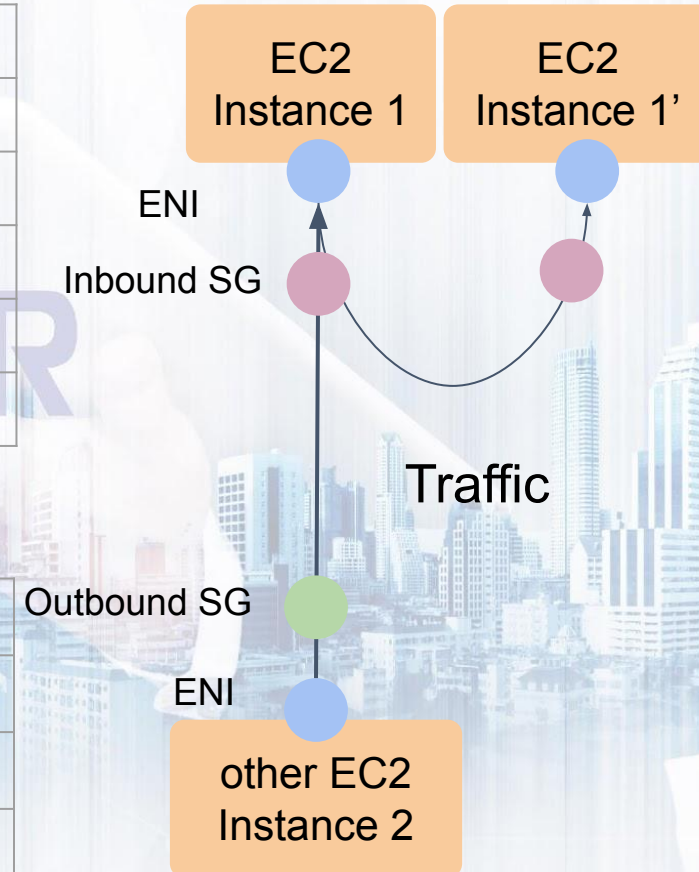
Instance 1 và Instance 1' đều được gán sg-aaaaaa => chúng có thể được trao đổi tất cả traffic lẫn nhau.

sg-aaaaaa

Inbound			
TCP	80	0.0.0.0/0	
SSH	22	172.168.5.6/32	
Allow all	All	sg-aaaaaa	
Outbound			
Allow all	All	0.0.0.0/0	

sg-xxxxxx

Inbound			
TCP	80	sg-aaaaaa	
SSH	22	172.168.5.6/32	
Outbound			
Allow all	All	sg-aaaaaa	



Pricing Model

- EC2 instance tính giá theo một số loại Pricing Model, bao gồm:
 - + On Demand
 - + Reserved Instance
 - + Saving Plan
 - + Spot Instance

Pricing Model

On Demand

Use case: Dùng cho các case sử dụng Instance không cố định, lúc dùng, lúc không, có thể bật và chạy trong ngắn hạn.

- Chi phí đắt nhất khi tính theo giờ.

Number of instances require



Pricing Model

Reserved Instance

- Chi phí thấp hơn On Demand.
- Đặt trước loại Instance type và sử dụng lâu dài

Use case: Dùng cho các case sử dụng Instance ổn định 1 năm, 3 năm hoặc lâu dài hơn.

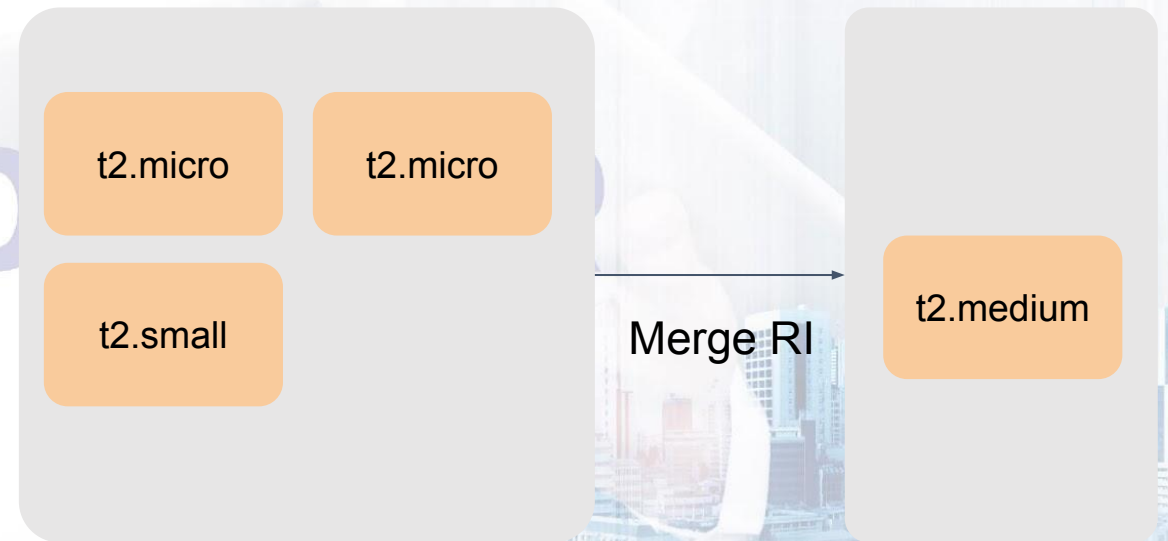
Number of instances require



Pricing Model

Reserved Instance

- Đối với reserve instance, chúng ta có thể gộp các Reserve Instance nhỏ hơn để tạo thành Reserve Instance lớn hơn:
- Requirement:
 - + Cùng family
 - + Cùng platform Linux/Unix

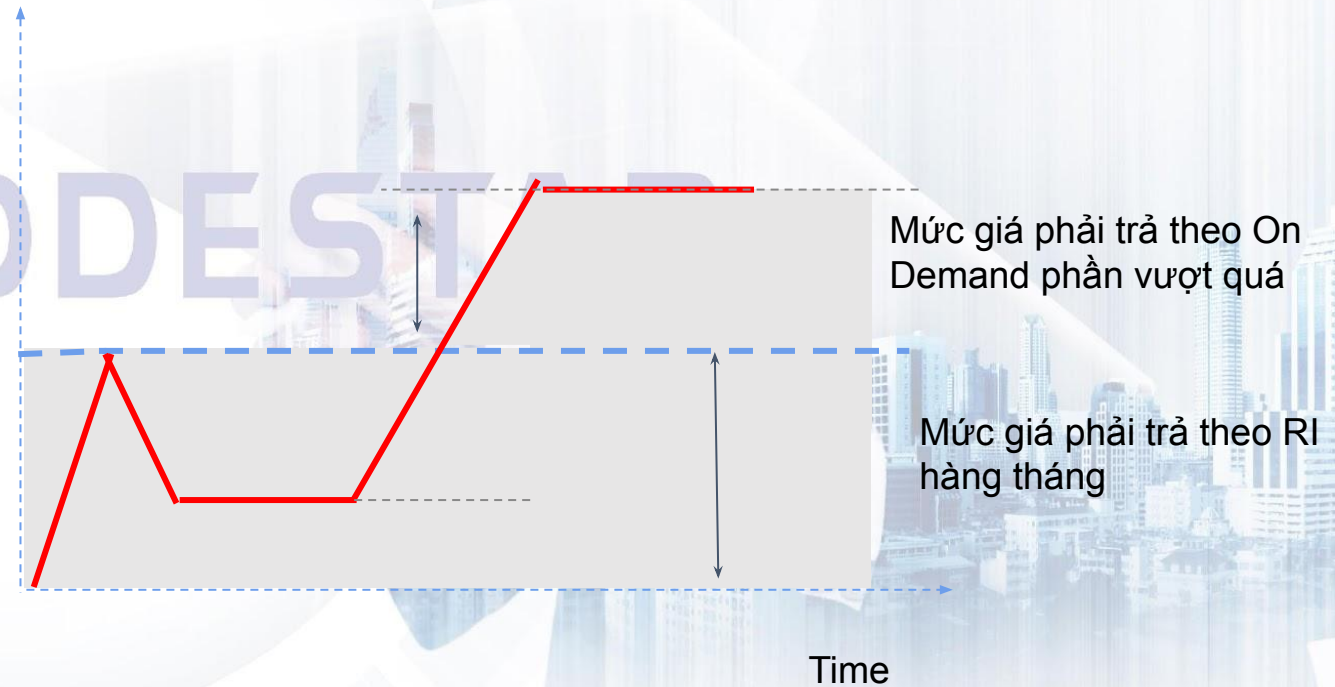


Pricing Model

Reserved Instance

RI thường sử dụng cho các hệ thống ổn định, không còn thay đổi theo thời gian ngắn, cho hệ thống Production hoặc cần chạy thường xuyên.

Number of instances require



Pricing Model

Saving Plans

- Chi phí thấp hơn On Demand, tương tự với Reserve Instance
- Trả phí theo mức rate định kỳ hàng ngày/hàng tháng

Use case: Dùng cho các case sử dụng computing resource ổn định 1 năm, 3 năm hoặc lâu dài hơn tương tự Reserve Instance

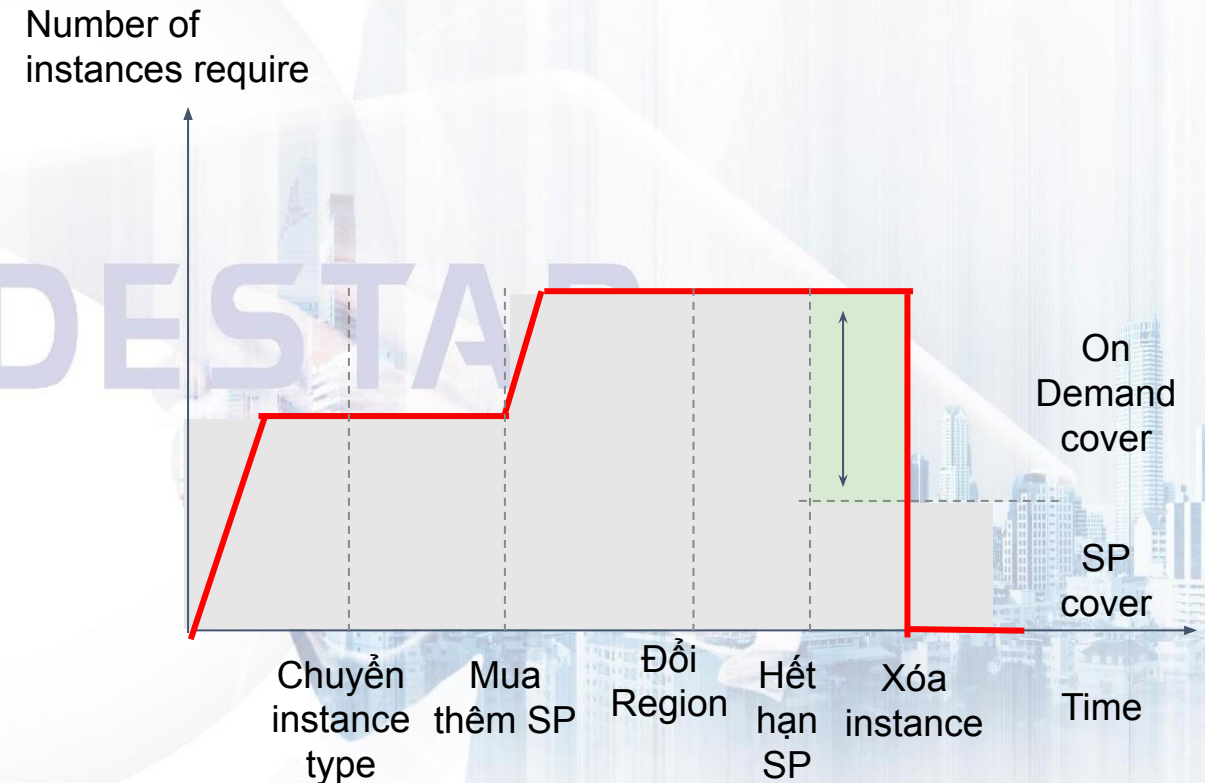
Number of instances require



Pricing Model

Saving Plans

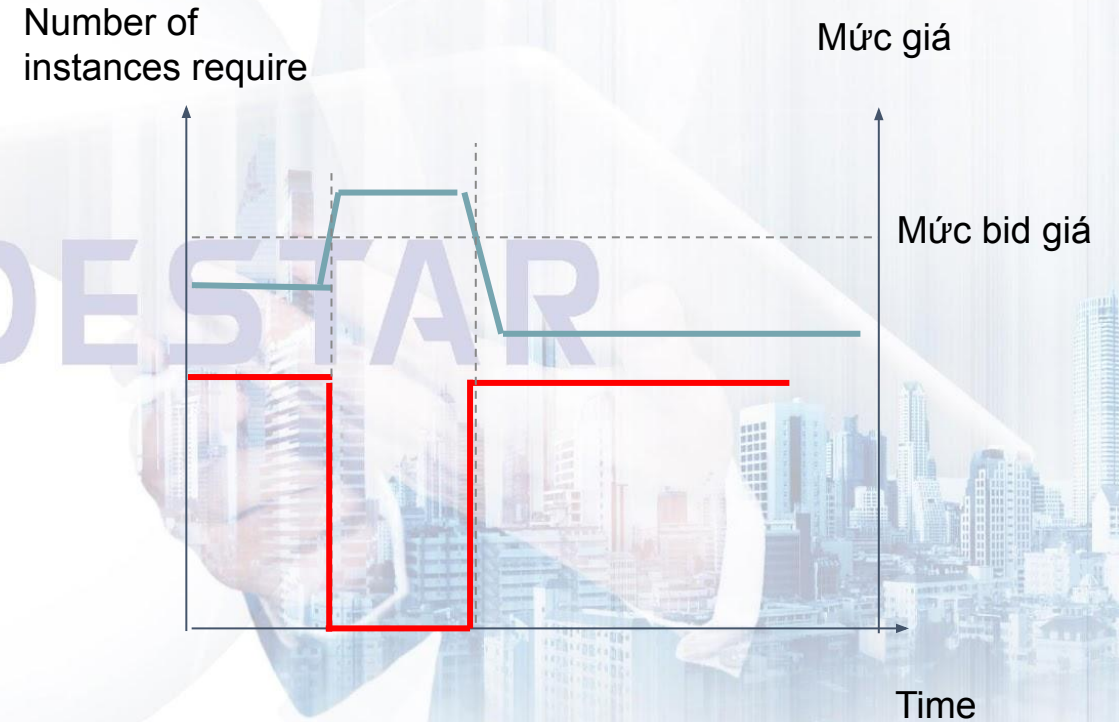
- Saving Plans có thể dễ dàng chuyển đổi giữa các loại Computing Resource như EC2 Instance, Lambda, Fargate.
- Saving Plans có thể sử dụng cho toàn bộ các Region (Compute SP).



Pricing Model

Spot Instance

- Spot Instance là những vị trí thừa ở trong host, mà bên phía AWS chưa dùng đến.
- Chi phí của Spot Instance biến đổi theo từng thời điểm
- Chúng ta sẽ đưa ra một mức giá có thể chấp nhận để trả. Nếu giá của SI nằm trong khoảng có thể chi trả, chúng sẽ được sử dụng với mức giá tại thời điểm đó.



Pricing Model

Spot Instance

- Spot Instance **có thể** tiết kiệm chi phí lên đến 90%. Tuy nhiên việc này không chắc chắn, đồng thời cũng không biết chính xác instance nào sẽ có thể được sử dụng và trong bao lâu

Use case: Dùng cho các case sử dụng computing resource lớn bất thường mà có thể chấp nhận việc bị dừng đột ngột.

Number of instances require



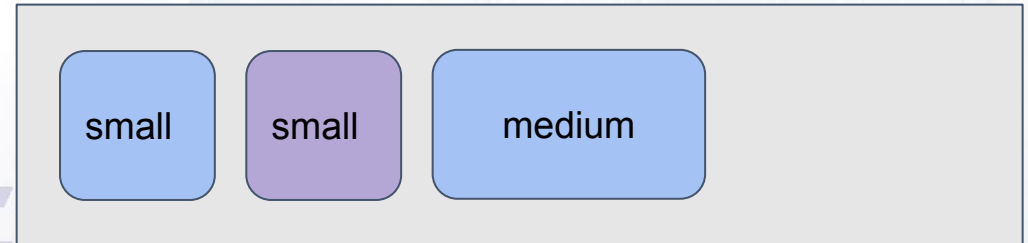
Pricing Model

Dedicated Host - Dedicated instance

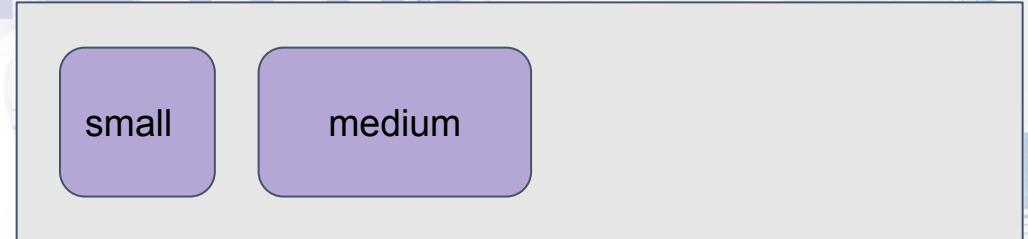
Dedicated Host và Dedicated Instance cho phép chúng ta sử dụng 1 host chuyên dụng cho các instance trên 1 account.

- Dedicated Host cho phép “nhìn thấy” và thao tác nhiều hơn về phía phần cứng. Có thể sử dụng đối với một số BYOL model như Window Server, SUSE, SQL Server, ...
- Dedicated Instance giống như Managed Dedicated Host, nhưng flexible về cách chia instance và ít cần quản lý hơn.

EC2 host



EC2 host



EC2 host

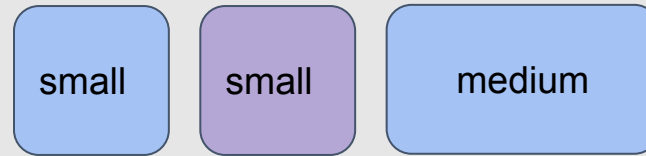


Pricing Model

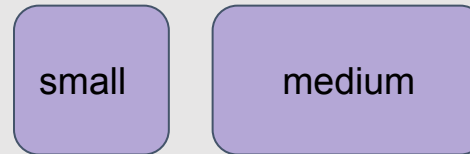
Dedicated Host - Dedicated instance

- Dedicated Host chia thành các instance đồng nhất về type và size.
- Dedicated Host tính tiền cho cả Host
- Dedicated Instances tính tiền instance và thêm một khoản phí dedicated.

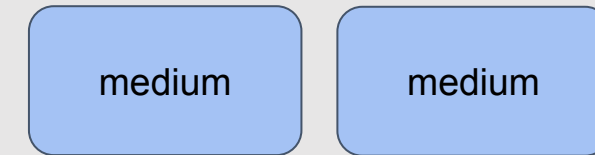
EC2 host



EC2 host



EC2 host



Pricing Model

Capacity Reservation

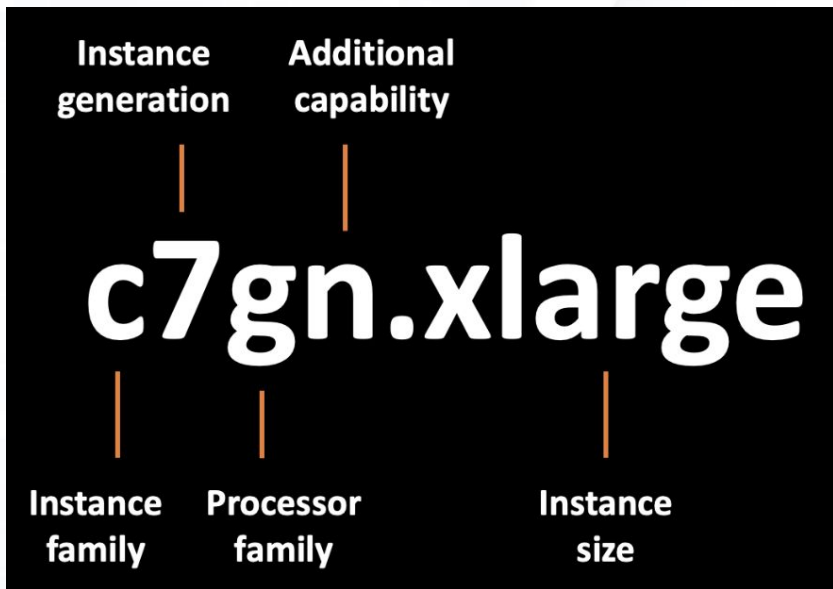
Capacity Reservation là một tính năng đặc biệt của AWS, cho phép chúng ta “đặt trước” một lượng EC2 instances, để có thể sử dụng bất cứ khi nào mà để tránh gặp vấn đề “cháy hàng” một số loại EC2 instances nào đó.

Use cases: Hệ thống cần đảm bảo có số lượng 200 EC2 instances chạy vào thời điểm 2 tuần nữa. Tuy nhiên trước đây hệ thống đã gặp vấn đề hết EC2 instances để sử dụng khi tới đúng thời điểm cần scale. Để không gặp lại tình trạng đó, chúng ta có thể sử dụng Capacity Reservation.

EC2 instance info: Instance type

General Purpose	Các dòng M, T	Phục vụ cho nhu cầu chung
Compute Optimized	Các dòng C	Các dòng chip tối ưu khả năng tính toán
Memory Optimized	Các dòng R, X	Các dòng chip tối ưu bộ nhớ/sử dụng RAM
Storage Optimized	Các dòng I, D, H	Các dòng chip tối ưu thao tác I/O trên ổ đĩa
Accelerated Computing	Các dòng P, G, Inf, F	Các dòng chip tích hợp GPU, phục vụ cho training ML model, xử lý đồ họa, ...

EC2 instance info: Instance type



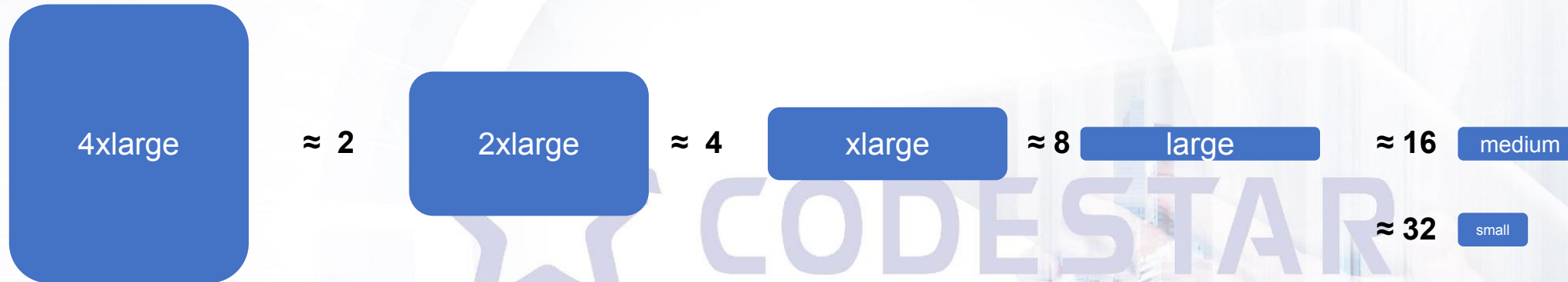
- **Instance family**: là một trong các ký hiệu ở trang trước đó (instance type)
- **Instance generation**: thế hệ instance, thường được đánh số, số càng lớn thì càng ra đời sau.
- **Processor family**: dòng xử lý có 3 loại là a (AMD), i (intel), g (graviton)
- **Additional capability**: Các tính năng bổ sung có:
b (ebs-opt), d (instance store), n (net & ebs opt), e (extra storage/mem), z (High perf), q (qualcomm), flex

VD:

R6gd: Chip dòng R (tối ưu RAM), thế hệ thứ 6, bộ xử lý graviton, có kèm instance store.

C6in: Chip dòng C (tối ưu computing), thế hệ thứ 6, chip intel, tối ưu network và ebs

EC2 instance info: Instance size



vCPU: Đơn vị tương đối đo số core/luồng của chip

Mỗi vCPU, có thể tương đương với ~ Intel Xeon processor.

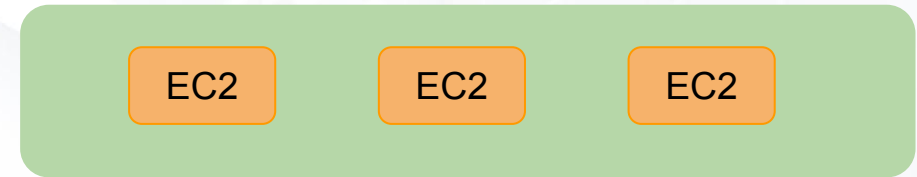
1 vCPU = 4 ECU (EC2 Computing Unit) = 4 x 1 - 1.2 GHz Intel Xeon processor.

Placement Group

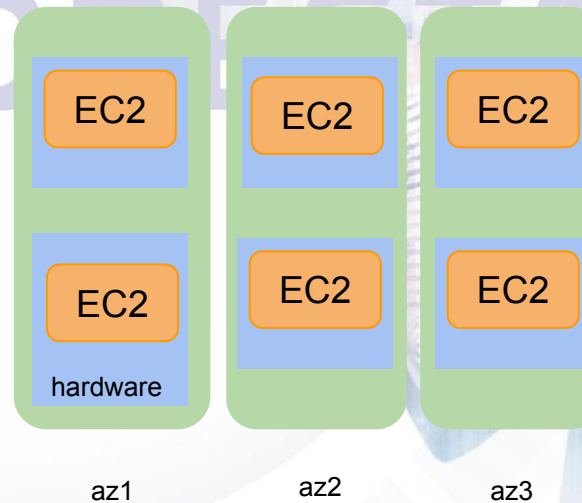
Placement Group là một tùy chọn khi tạo EC2 instance, cho phép chúng ta config vị trí đặt EC2 instance.

EC2 instance thực chất là máy ảo, có thể được đặt trên cùng một thiết bị phần cứng với máy ảo khác.

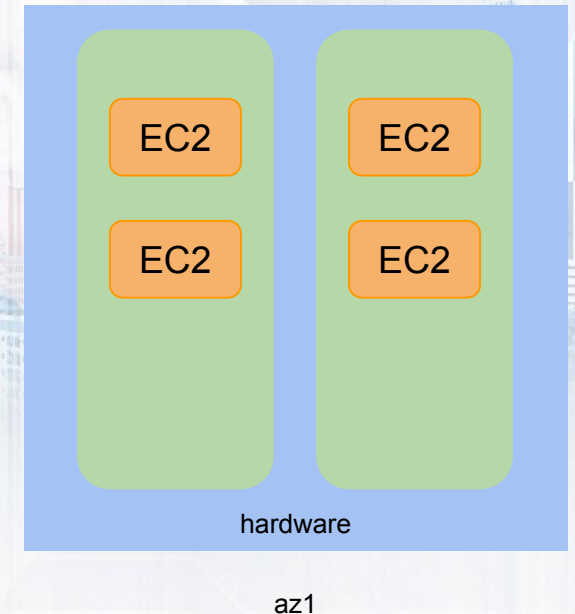
Cluster Placement Group



Spread Placement Group



Partition Placement Group

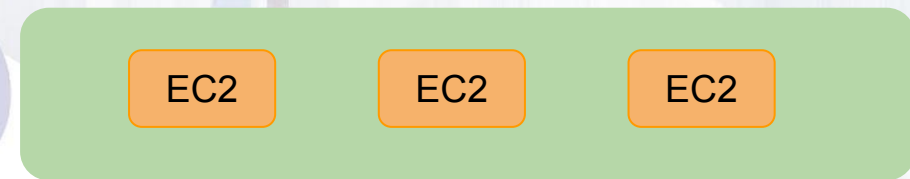


Placement Group

Có 3 kiểu đặt Placement group:

- **Cluster Placement Group:** Đặt trong một Logical Group, giúp các instance trong cùng Cluster có thể tăng tốc độ truyền tải dữ liệu.
- Không thể dùng cho multi AZ

Cluster Placement Group

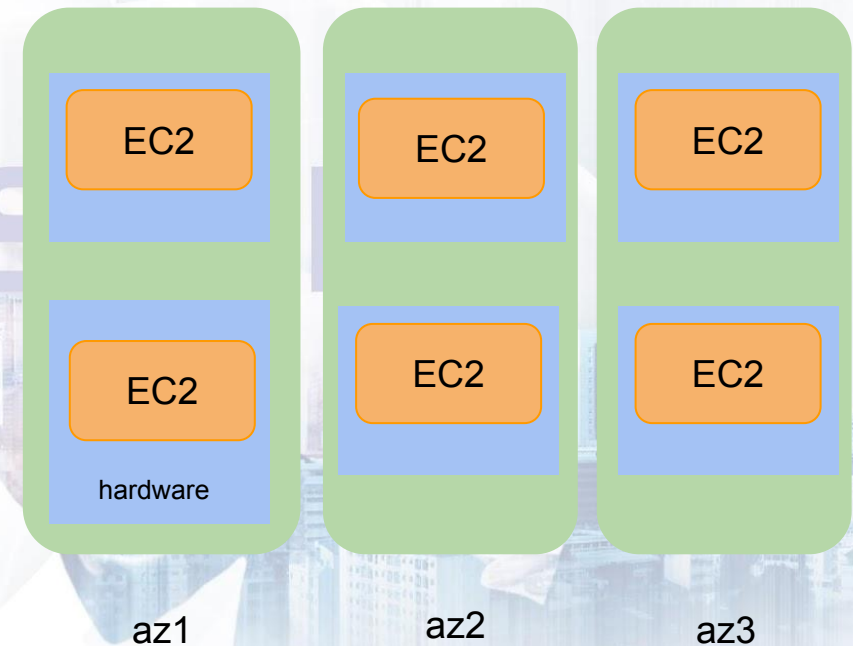


Placement Group

Có 3 kiểu đặt Placement group:

- **Spread Placement Group:** Đây là kiểu placement group mặc định.
- Các EC2 instances sẽ cố gắng đặt trên những thiết bị phần cứng khác nhau.
- Giảm thiểu rủi ro lỗi từ phần cứng
- Có thể chia thành: **Rack level Spread Placement Group** và **Host level Spread Placement Group** (chỉ dùng cho AWS Outposts)

Spread Placement Group

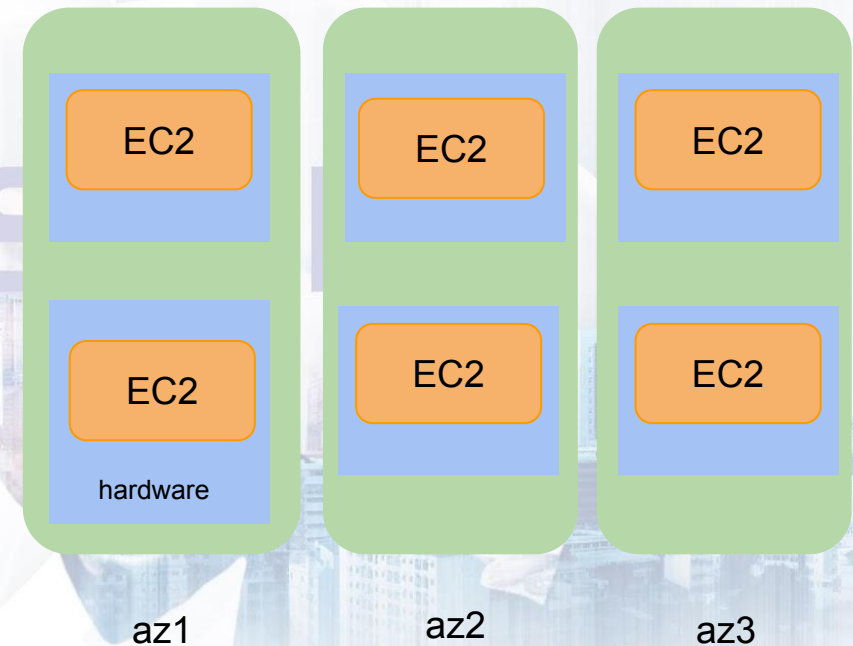


Placement Group

Có 3 kiểu đặt Placement group:

- **Rack level Spread Placement Group** có thể sử dụng tối đa 7 hardware (rack) trên 1 AZ. EC2 instance sử dụng rack thứ 8 sẽ không thể launch -> Có thể sử dụng một Spread Placement Group khác.

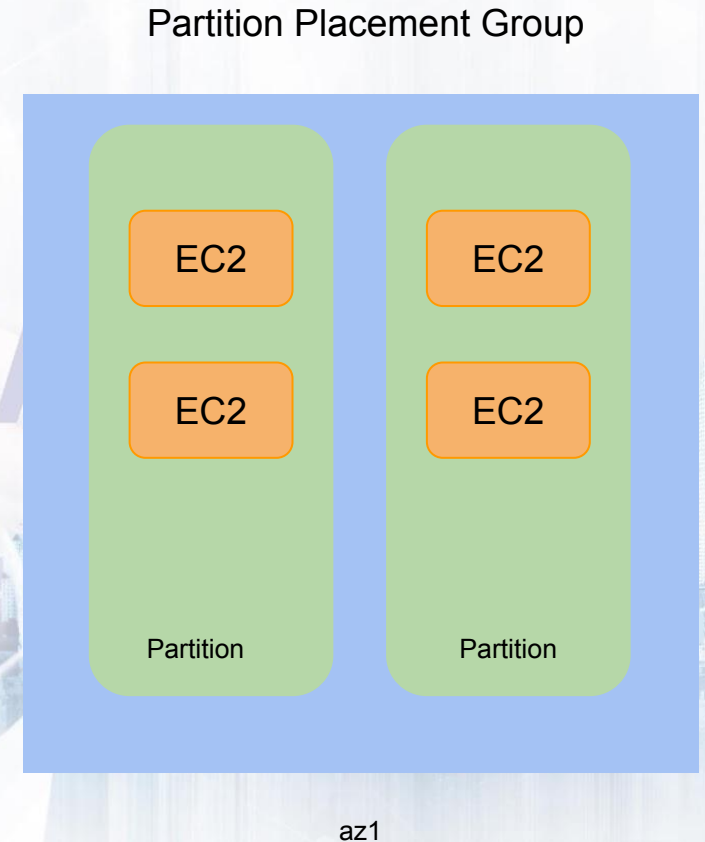
Spread Placement Group



Placement Group

Có 3 kiểu đặt Placement group:

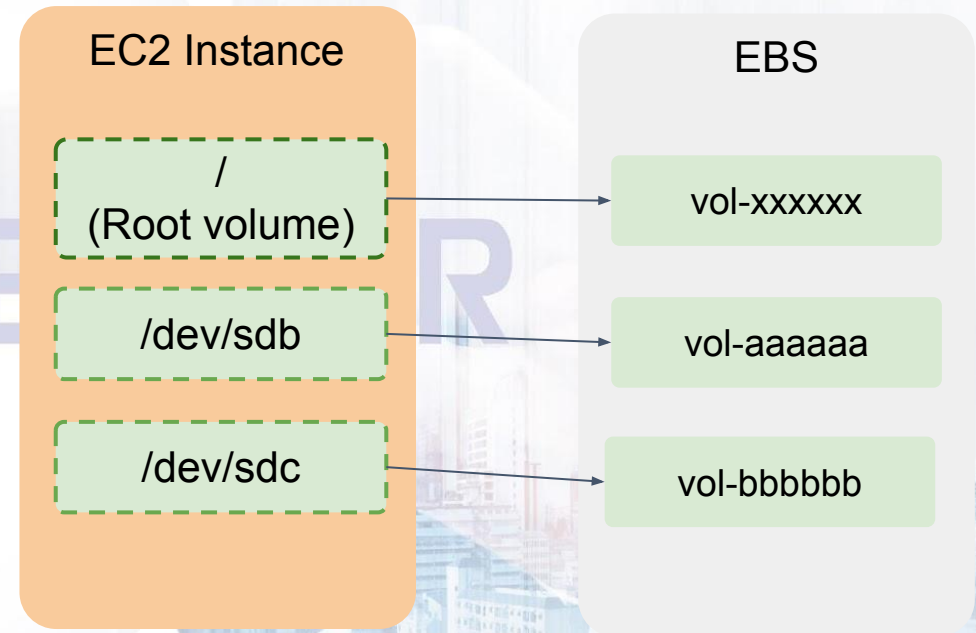
- **Partition Placement Group** có thể sử dụng tối đa 7 partition với mỗi AZ.
- Partition Placement Group cho phép tối ưu network trong Partition, đồng thời cho phép fail tolerance giữa các partition.
- Thích hợp cho các mô hình sử dụng HDFS, HBase, và Cassandra.



EBS

Amazon Elastic Block Store (EBS) là một loại dịch vụ lưu trữ khối (Block storage service), có khả năng tương tự một ổ cứng, có đặc điểm:

- Tốc độ đọc ghi cao, độ trễ thấp, phù hợp các tác vụ yêu cầu workload lớn.
- Thường dùng để lưu trữ những hệ thống thường xuyên truy xuất như database, enterprise application, containerized application, big data analytic engine, file system, ...

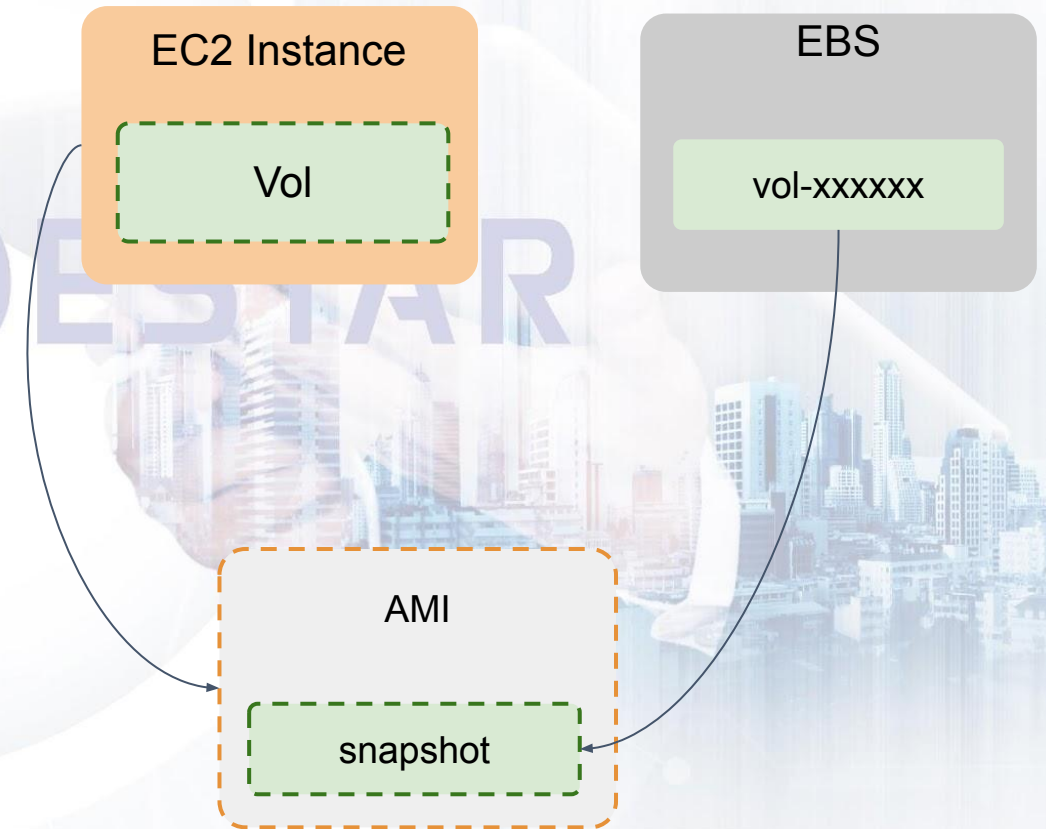


EBS type

	General Purpose SSD		Provisioned IOPS SSD		
Volume type	gp3	gp2	io2 Block Express ‡	io2	io1
Durability	99.8% - 99.9%	99.8% - 99.9%	99.999%	99.999%	99.8% - 99.9%
Use cases	<ul style="list-style-type: none">Apps cần tương tác low-latencyDevelopment and test environments		<ul style="list-style-type: none">Cần độ trễ cực thấp (vài mili giây)IOPS liên tục> 64,000 IOPS hoặc tốc độ 1,000 MiB/s of throughput	<ul style="list-style-type: none">Hệ thống có hiệu năng IOPS ổn định > 16,000 IOPSCác loại DB thường xuyên truy xuất I/O nhiều	
Volume size	1 GiB - 16 TiB		4 GiB - 64 TiB	4 GiB - 16 TiB	
Max IOPS per volume (16 KiB I/O)	16,000		256,000	64,000 †	
Max throughput per volume	1,000 MiB/s	250 MiB/s *	4,000 MiB/s	1,000 MiB/s †	
Amazon EBS Multi-attach	Không hỗ trợ		Có hỗ trợ		

EBS - Snapshot

- EBS Snapshot là 1 bản nén của 1 ổ cứng.
- Khi tạo Snapshot, thông tin Snapshot sẽ được lưu trữ trên S3 và tính tiền theo phí lưu trữ S3.
- Có thể tạo nhiều EBS từ 1 EBS Snapshot trên S3.

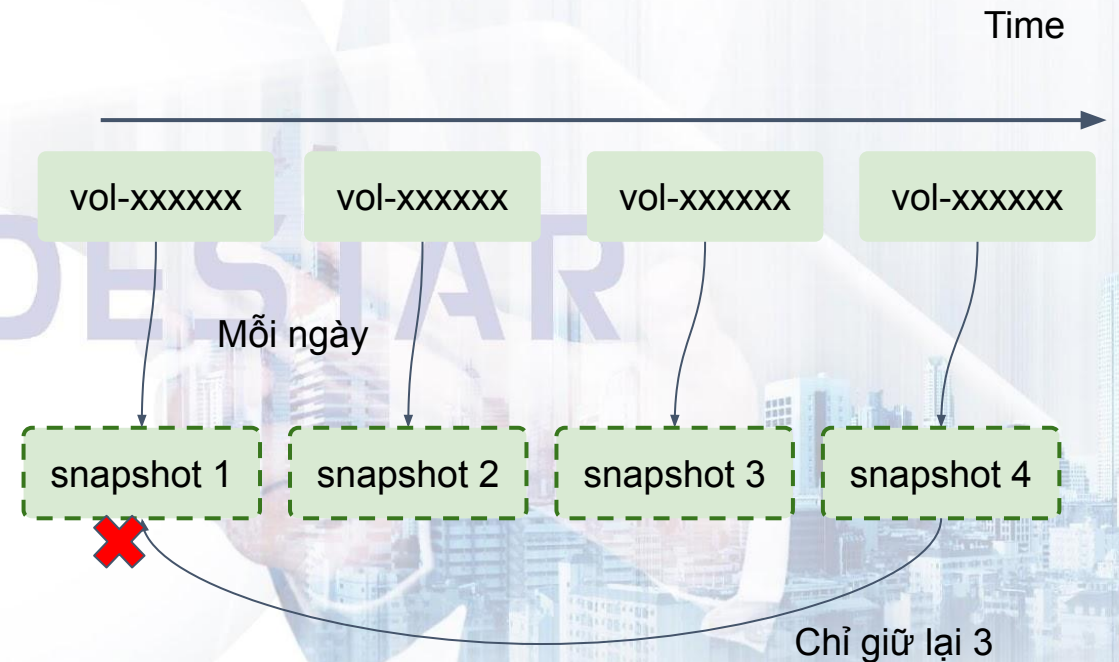


EBS - Snapshot

Lifecycle Manager

Một tính năng miễn phí dành cho EC2 instance và EBS Volume, giúp:

- Tự động đặt lịch backup
- Tự động thiết lập các policy xóa bớt các bản backup cũ
- Tự động phân quyền, đánh tag và quản lý các AMI/snapshot của EBS volume

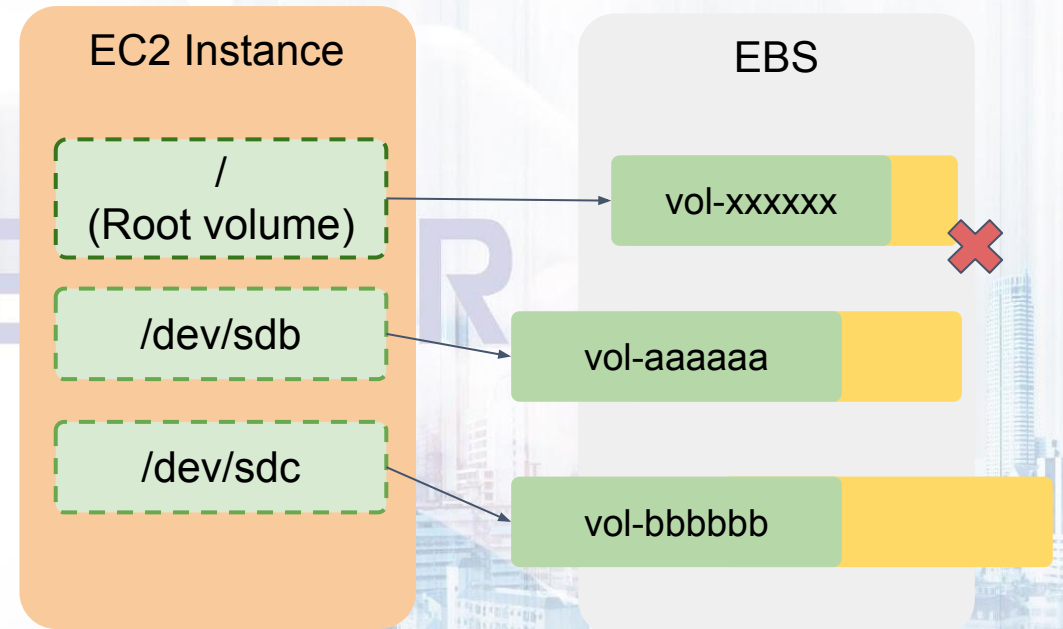


EBS

Amazon Elastic Block Store (EBS) có thể expand, shrink dung lượng của mình

Việc expand/shrink dung lượng ổ Root volume sẽ yêu cầu EC2 instance dừng hoạt động.

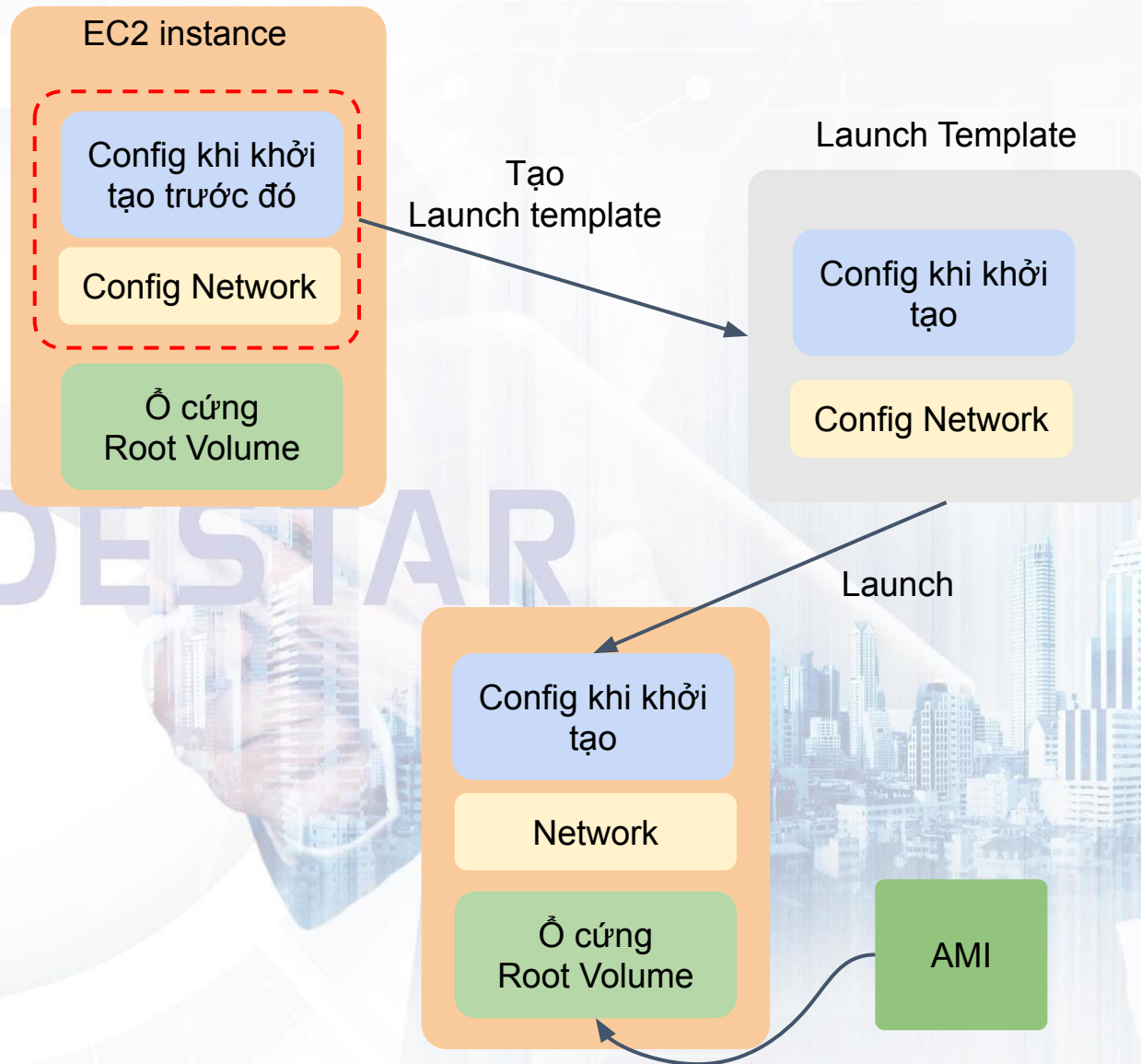
Các ổ khác có thể được expand ngay trong khi EC2 instance vẫn đang chạy.



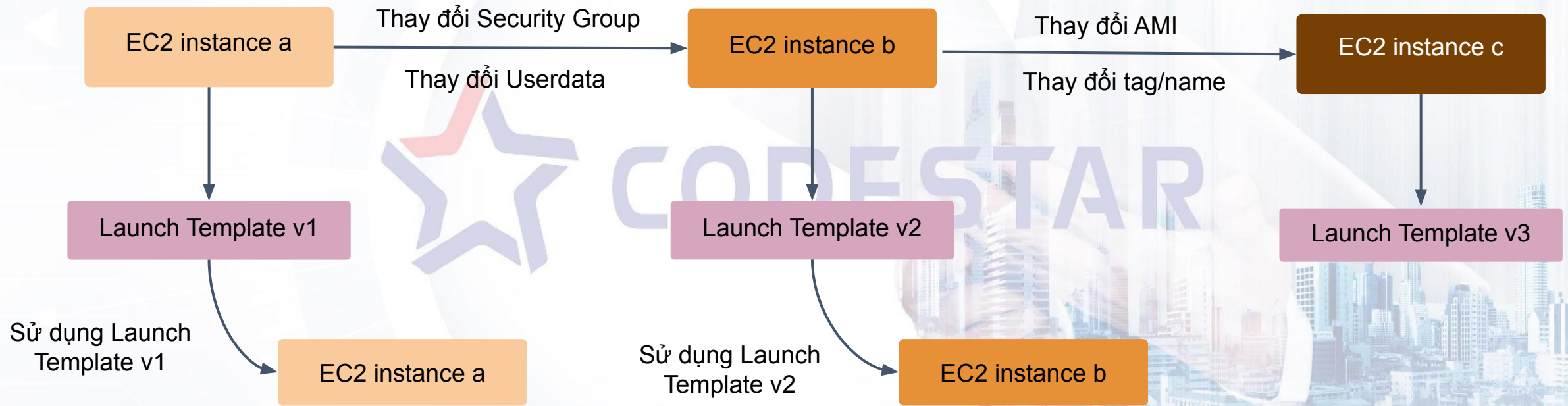
Launch Template

Launch Template là một thành phần lưu lại các config đã sử dụng để khởi tạo EC2 instance -> Có thể khởi tạo thêm EC2 instance một cách nhanh chóng.

Launch Template thường dùng cho các thao tác **tự động** khởi tạo



Launch Template





THANK YOU