

AWS Solutions Architect

IAM, EC2, EBS



Н. Ганжигүүр
Fibo Cloud

ICT Guest @ICT2024#

Conclusion

- Cloud computing - Tech + Business
- How do you compute, Not where
- Iceberg - Cloud Native
- Productivity, Agility, Cost
- X as a service
- Private, Public, Hybrid
- Shared responsibility model
- Region, Availability Zone, Edge locations
- AWS access
 - Web Console GUI
 - CLI
 - SDK

How to choose an AWS Region?



- **Compliance with data governance and legal requirements:** data never leaves a region without your explicit permission
- **Proximity to customers:** reduced latency
- **Available services within a Region:** new services and new features aren't available in every Region
- **Pricing:** pricing varies region to region and is transparent in the service pricing page

Services Across Regions

- **AWS has Global Services:**
 - Identity and Access Management (IAM)
 - Route 53 (DNS service)
 - CloudFront (Content Delivery Network)
 - WAF (Web Application Firewall)
- **Most AWS services are Region-scoped:**
 - Amazon EC2 (Infrastructure as a Service)
 - Elastic Beanstalk (Platform as a Service)
 - Lambda (Function as a Service)
 - Rekognition (Software as a Service)
- **Region Table:**
 - <https://aws.amazon.com/about-aws/global-infrastructure/regional-product-services>

Interesting

<https://www.concurrencylabs.com/blog/choose-your-aws-region-wisely/>

AWS Support

AWS Support

1:1 support 24/7 from experience technicians

- Technical help, recommendation
- Request limitation
- Billing or Account issue

Basic, Developer, Business & Enterprise

<https://aws.amazon.com/premiumsupport/plans/>



AWS Support Plans

Basic	Developer	Business	Enterprise
Email Support only For Billing and Account	Tech Support via Email ~24 hours until reply No third party support	Tech Support via Chat, Phone Anytime 24/7	
	General Guidance		< 24 hrs
	System Impaired	Production System Impaired	< 12 hrs
		Production System DOWN!	< 4 hrs
			< 1 hrs
			Business-Critical System DOWN! < 15m
		 Personal Concierge	
		 TAM	
7 Trusted Advisor Checks		All Trusted Advisor Checks	
\$0 USD /month	\$20 USD /month	\$100 USD / month	\$15,000 USD / month

Developer

Minimum spend of \$29.00

- or -

3% of monthly AWS charges

Charges will be at least the minimum charge of \$29.00 or the result of the calculation, whichever is higher

Business

Minimum spend of \$100.00

- or -

10% of monthly AWS charges for the first \$0--\$10K

7% of monthly AWS charges from \$10K--\$80K

5% of monthly AWS charges from \$80K--\$250K

3% of monthly AWS charges over \$250K

Charges will be at least the minimum charge of \$100.00 or the result of the calculation, whichever is higher

*[AWS Countdown Premium](#) available for an additional fee.

Enterprise On-Ramp

Minimum spend of \$5,500.00

- or -

10% of monthly AWS charges

Charges will be at least the minimum charge of \$5,500.00 or the result of the calculation, whichever is higher

*[AWS Countdown Premium](#) available for an additional fee.

*[AWS re:Post Private](#) available for an additional fee.

Enterprise

Minimum spend of \$15,000.00

- or -

10% of monthly AWS charges for the first \$0--\$150K

7% of monthly AWS charges from \$150K--\$500K

5% of monthly AWS charges from \$500K--\$1M

3% of monthly AWS charges over \$1M

Charges will be at least the minimum charge of \$15,000.00 or the result of the calculation, whichever is higher.

*[AWS Incident Detection and Response](#) available for an additional fee.

*[AWS Countdown Premium](#) available for an additional fee.

*[AWS re:Post Private](#) available for an additional fee.

AWS Support

AWS Partner network - <https://aws.amazon.com/partners/>

Forum - <https://forums.aws.amazon.com/index.jspa?cu-additional-resource%2F=>

FAQ

Community

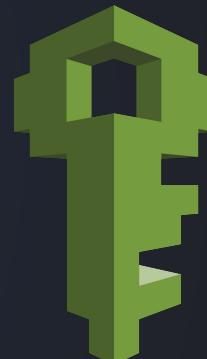
Re:post - <https://repost.aws/>

Identity Access Management

IAM: Users & Groups



- IAM = Identity and Access Management, Global service
- Root account created by default, shouldn't be used or shared
- Users are people within your organization, and can be grouped
- Groups only contain users, not other groups
- Users don't have to belong to a group, and user can belong to multiple groups



WHO, WHERE, WHAT

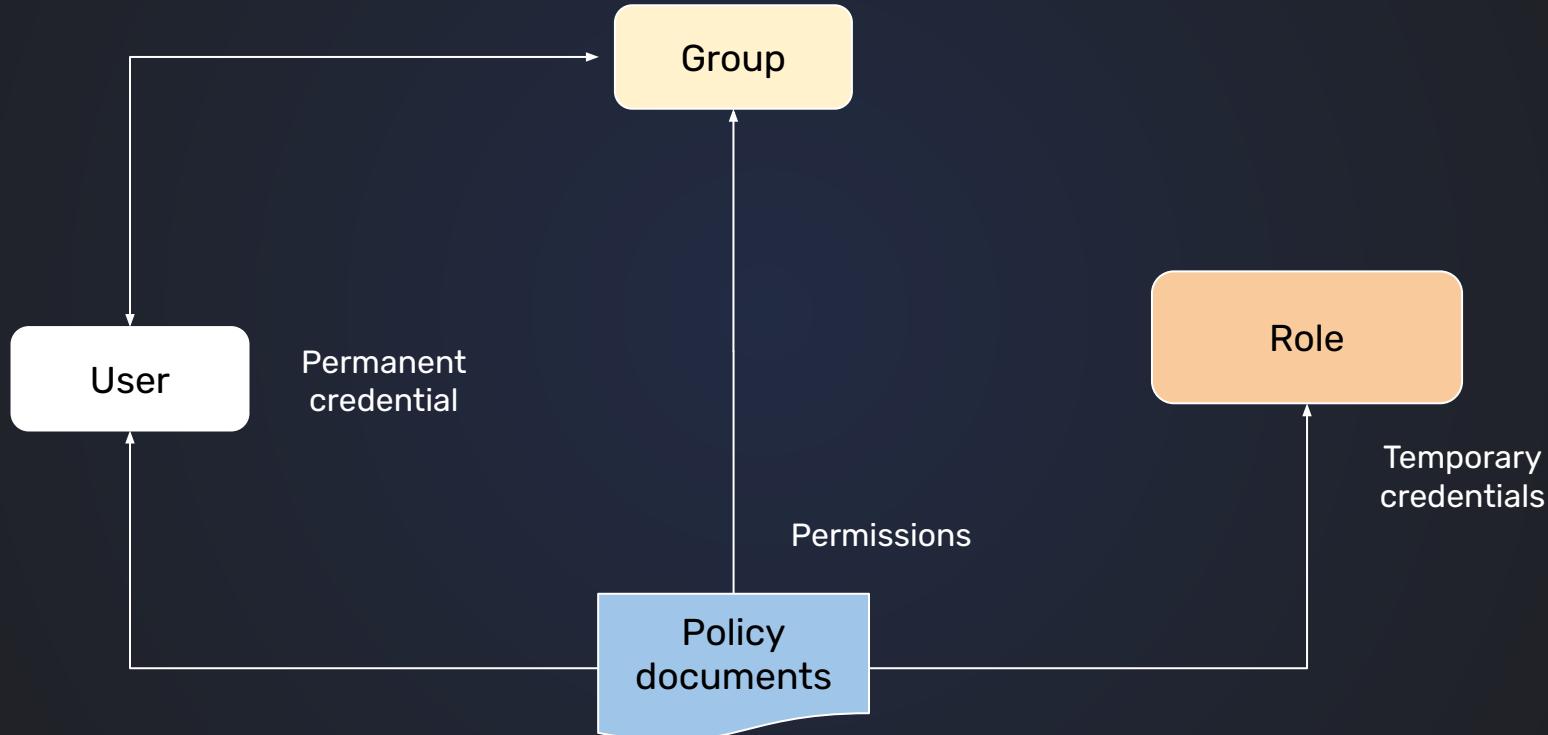
IAM: Permissions

- Users or Groups can be assigned JSON documents called policies
- These policies define the permissions of the users
- In AWS you apply the least privilege principle: don't give more permissions than a user needs

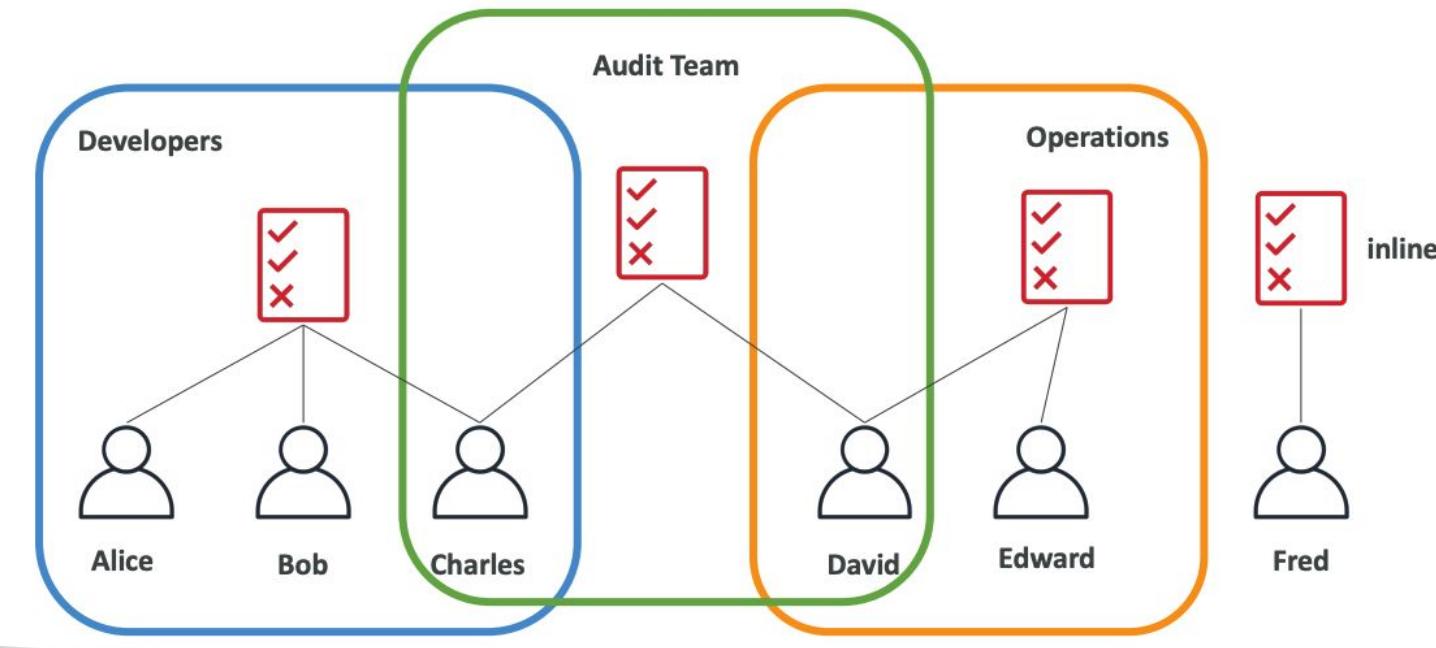
```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Effect": "Allow",  
            "Action": "ec2:Describe*",  
            "Resource": "*"  
        },  
        {  
            "Effect": "Allow",  
            "Action": "elasticloadbalancing:Describe*",  
            "Resource": "*"  
        },  
        {  
            "Effect": "Allow",  
            "Action": [  
                "cloudwatch>ListMetrics",  
                "cloudwatch:GetMetricStatistics",  
                "cloudwatch:Describe*"  
            ],  
            "Resource": "*"  
        }  
    ]  
}
```



Relation



IAM Policies inheritance



IAM – Password Policy

- Strong passwords = higher security for your account
- In AWS, you can setup a password policy:
 - Set a minimum password length
 - Require specific character types:
 - including uppercase letters
 - lowercase letters
 - numbers
 - non-alphanumeric characters
 - Allow all IAM users to change their own passwords
 - Require users to change their password after some time (password expiration)
 - Prevent password re-use

Multi Factor Authentication - MFA



- Users have access to your account and can possibly change configurations or delete resources in your AWS account
- You want to protect your Root Accounts and IAM users
- MFA = password you know + security device you own



Alice

Password

+



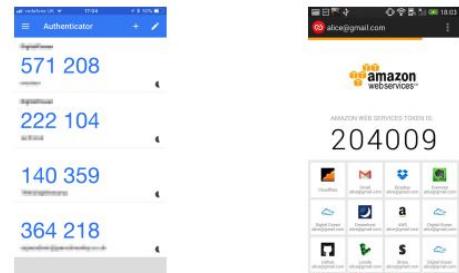
=>

Successful login

- Main benefit of MFA:
if a password is stolen or hacked, the account is not compromised

MFA devices options in AWS

Virtual MFA device



Google Authenticator
(phone only)

Support for multiple tokens on a single device.

Universal 2nd Factor (U2F) Security Key



YubiKey by Yubico (3rd party)

Support for multiple root and IAM users using a single security key

How can users access AWS ?



- To access AWS, you have three options:
 - AWS Management Console (protected by password + MFA)
 - AWS Command Line Interface (CLI): protected by access keys
 - AWS Software Developer Kit (SDK) - for code: protected by access keys
- Access Keys are generated through the AWS Console
- Users manage their own access keys
- Access Keys are secret, just like a password. Don't share them
- Access Key ID ~ = username
- Secret Access Key ~ = password

A tool that enables you to interact with AWS services using commands in your command-line shell

Direct access to the public APIs of AWS services

ACCESS KEY DEMO

```
→ ~ aws s3 cp myfile.txt s3://ccp-mybucket/myfile.txt
upload: ./myfile.txt to s3://ccp-mybucket/myfile.txt
→ ~ aws s3 ls s3://ccp-mybucket
2021-05-14 03:22:52          0 myfile.txt
→ ~ █
```

IAM Security Tools

- **IAM Credentials Report (account-level)**
 - a report that lists all your account's users and the status of their various credentials
- **IAM Access Advisor (user-level)**
 - Access advisor shows the service permissions granted to a user and when those services were last accessed.
 - You can use this information to revise your policies.

IAM Guidelines & Best Practices



- Don't use the root account except for AWS account setup
- One physical user = One AWS user
- Assign users to groups and assign permissions to groups
- Create a **strong password policy**
- Use and enforce the use of Multi Factor Authentication (MFA)
- Create and use **Roles** for giving permissions to AWS services
- Use Access Keys for Programmatic Access (CLI / SDK)
- Audit permissions of your account using IAM Credentials Report & IAM Access Advisor
- Never share IAM users & Access Keys

Homework 1

Account үүсгээд fiboadmin - user үүсгээд над руу явуулаарай.

IAM, User, Group

1. admin, user1, user2, user3 гэсэн **4 IAM User** үүсгэнэ
2. S3-Support, EC2-Support, EC2-Admin гэсэн **3 User group** үүсгэнэ.
3. admin хэрэглэгчид Administrator permission policy өгнө.
4. User Group тус бүрт дараах эрхүүдийг өгнө:
 - a. *S3-Support -> AmazonS3ReadOnlyAccess*
 - b. *EC2-Support -> AmazonEC2ReadOnlyAccess*
 - c. *EC2-Admin -> Шинэ custom policy үүсгэж дараах эрхийг олгоно.*
<https://gist.github.com/Ganjiguur/7d5c95448ba21abb88ca6c3d76bf82bb>
5. Үүсгэсэн хэрэглэгчдийг тус тусын групп лүү оруулна:

User	In Group	Permissions
user-1	S3-Support	Read-Only access to Amazon S3
user-2	EC2-Support	Read-Only access to Amazon EC2
user-3	EC2-Admin	View, Start and Stop Amazon EC2 instances

Homework 1

6. Admin хэрэглэгчээр нэвтэрч S3 bucket болон EC2 micro хэмжээтэй сервер үүсгэнэ.

7. user1-ээр нэвтэрч ороод:

- a. S3 цэс рүү орж үзнэ. Үйлдэл хийнэ (Шинэ bucket үүсгэх г.м.)
- b. EC2 цэс рүү орж үзнэ. Үйлдэл хийнэ (Шинэ server үүсгэх г.м.)

8. user2-р нэвтэрч ороод:

- a. S3 цэс рүү орж үзнэ. Үйлдэл хийнэ (Шинэ bucket үүсгэх г.м.)
- b. EC2 цэс рүү орж үзнэ. Үйлдэл хийнэ (Шинэ server үүсгэх г.м.)
- c. Server унтраах гэж оролд

9. user3-р нэвтэрч ороод:

- a. S3 цэс рүү орж үзнэ. Үйлдэл хийнэ (Шинэ bucket үүсгэх г.м.)
- b. EC2 цэс рүү орж үзнэ. Үйлдэл хийнэ (Шинэ server үүсгэх г.м.)
- c. Server унтраах гэж оролд

10. Admin-aар ороод User group permission-г өөрчлөөд үз.

11. 6, 7, 8 дээр бичигдсэн үйлдлүүдийг давтаж хий.

12. Үүсгэсэн бүх resource-уудаа цэвэрлэх.

- a. Users
- b. User-groups
- c. S3 bucket
- d. EC2 server

Homework 2

Create a billing alarm

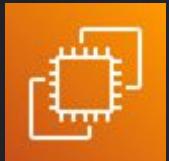
https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/monitor_estimated_charges_with_cloudwatch.html

Usage tips

- Use FAQ
- Keep access token
- CLI is better than all
- Don't use root account
- Billing alarm must have

Compute

Compute services



EC2

Virtual server - CPU, Memory, OS

ECS

Docker as a Service



LightSail

Low price fixed servers



Fargate

Serverless container orchestrator



Lambda

Serverless functions



AWS Batch

planned, scheduled job



Elastic Beanstalk

Orchestrates various AWS services
- PaaS



AWS Outpost

On-prem AWS services

EC2



Elastic Compute Cloud



AWS клауд орчин дах виртуал сервер
үүсгэх үйлчилгээ.
Өөрийн хэрэгцээ шаардлагад нийцсэн
хэмжээ бүхий серверийг хоромын дотор
үүсгэж, удирдах боломжтой байна.



Purchase types

On-demand instances

Least commitment

Launch instances

Default using On-Demand pricing

no up-front payment

no long-term commitment

charged **by the hour or by the minute**

Is for **short-term, test, expirement or unpredictable** workload

Reserved instances (RI)

Best Long-term

Designed for apps that have **steady-state, predictable usage**, or require **reserved capacity**

Team , Class offering , Payment option

1 year contract
3 year contract

Standard - Up to 75% discount
Convertible - Up to 54% discount
Scheduled - Reserve for specific time periods

All upfront, partial
upfront or no upfront

RIs can be share between multiple accounts
Can be sold Reserved Instance Marketplace

Savings plan

	EC2 Instance SP	Compute SP
Savings over On-Demand	<input checked="" type="checkbox"/> Relatively high (Up to 72% off)	<input checked="" type="checkbox"/> Relatively low (Up to 66% off)
Instance family	Fixed	<input checked="" type="checkbox"/> Flexible
Region	Fixed	<input checked="" type="checkbox"/> Flexible
Platform, tenancy, instance size, AZ	Not required	
Applying to Lambda, Fargate	No	<input checked="" type="checkbox"/> Yes
Modifying/Exchanging purchased instances	No	

Spot instances

Biggest savings

Buy unused compute capacity from AWS

90% discount

Conditions

- Can be terminated by AWS at anytime
- If terminated by AWS, we don't get charged hourly
- If we terminate, we will still be charged hourly

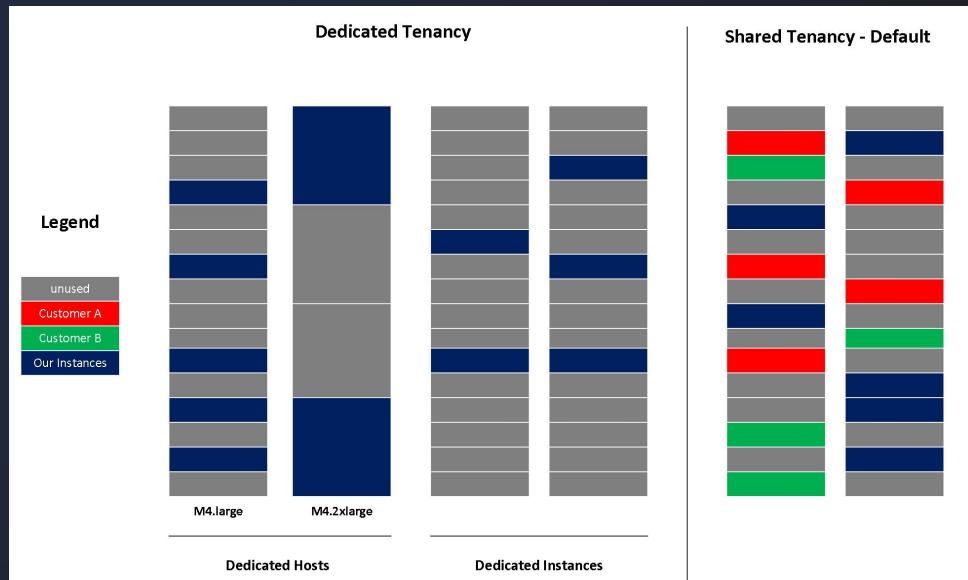
Useful for workloads that are resilient to failure

- Batch jobs
- Data analysis
- Image processing
- Any distributed workloads
- Workloads with a flexible start and end time

Dedicated host

Characteristic	Dedicated Instances	Dedicated Hosts
Enables the use of dedicated physical servers	X	X
Per instance billing (subject to a \$2 per region fee)	X	
Per host billing		X
Visibility of sockets, cores, host ID		X
Affinity between a host and instance		X
Targeted instance placement		X
Automatic instance placement	X	X
Add capacity using an allocation request		X

Most expensive



AWS Marketplace

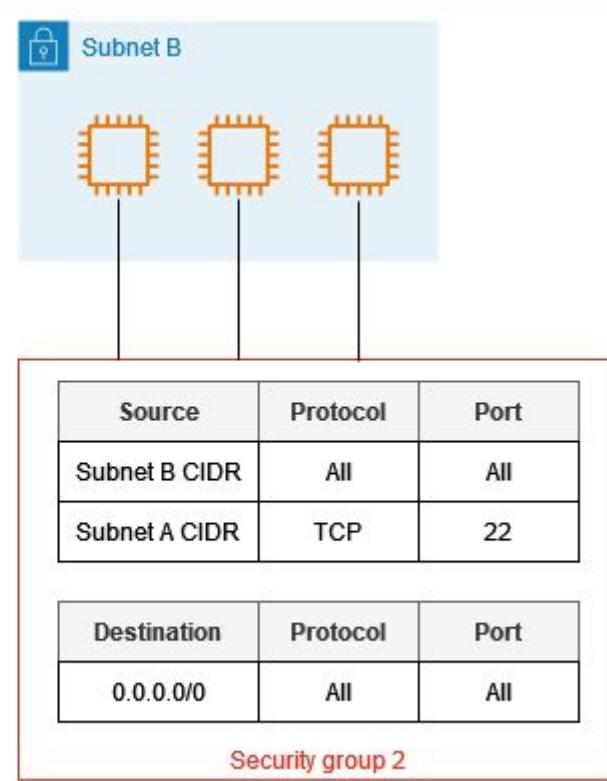
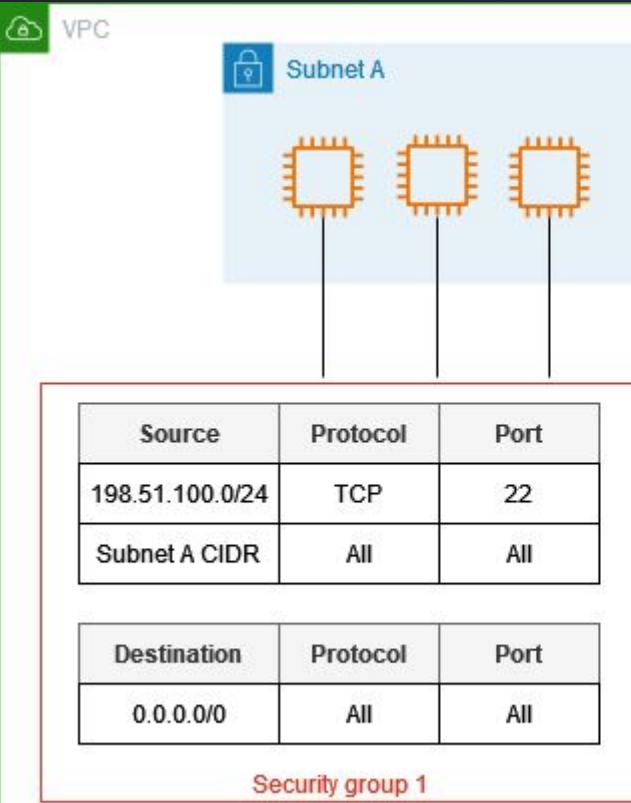
AWS Marketplace

AWS Marketplace is a curated digital catalogue with thousands of software listing from independent software vendors.

- free
- associated charge



Security Groups



Instance types

Instance types

	Type	Description	Mnemonic
General Purpose	a1	Good for scale-out workloads, supported by Arm	a is for Arm processor – or as light as A1 steak sauce
	t-family: t3, t3a, t2	Burstable, good for changing workloads	t is for tiny or turbo
	m-family: m6g, m5, m5a, m5n, m4	Balanced, good for consistent workloads	m is for main or happy medium
Compute Optimized	c-family: c5, c5n, c4	High ratio of compute to memory	c is for compute
Memory Optimized	r-family: r5, r5a, r5n, r4	Good for in-memory databases	r is for RAM
	x1-family: x1e, x1	Good for full in-memory applications	x is for xtreme
	High memory	Good for large in-memory databases	High memory is for... high memory.
	z1d	Both high compute and high memory	z is for zippy
Accelerated Computing	p-family: p3, p2	Good for graphics processing and other GPU uses	p is for pictures
	Inf1	Support machine learning inference applications	Inf is for inference
	g-family: g4, g3	Accelerate machine learning inference and graphics-intensive workloads	g is for graphics
	f1	Customizable hardware acceleration with field programmable gate arrays (FPGAs)	f is for FPGA or feel as in hardware
Storage Optimized	i-family: i3, i3en	SDD-backed, balance of compute and memory	i is for IOPS
	d2	Highest disk ratio	d is for dense
	h1	HDD-backed, balance of compute and memory	H is for HDD

Private vs Public IP (IPv4)

- Networking has two sorts of IPs. IPv4 and IPv6:
 - IPv4: **1.160.10.240**
 - IPv6: **3ffe:1900:4545:3:200:f8ff:fe21:67cf**
- In this course, we will only be using IPv4.
- IPv4 is still the most common format used online.
- IPv6 is newer and solves problems for the Internet of Things (IoT).
- IPv4 allows for **3.7 billion** different addresses in the public space
- IPv4: [0-255].[0-255].[0-255].[0-255].

Elastic IPs

- When you stop and then start an EC2 instance, it can change its public IP.
- If you need to have a fixed public IP for your instance, you need an Elastic IP
- An Elastic IP is a public IPv4 IP you own as long as you don't delete it
- You can attach it to one instance at a time

EBS



EBS is replicated within its AZ
offer component failure, HA,
durability

- General purpose (SSD)
- Provisioned IOPS (SSD)
- Throughput Optimised HDD
- Cold Hard Disk drive
- Magnetic

EBS Volume

- It's a network drive (i.e. not a physical drive)
 - It uses the network to communicate the instance, which means there might be a bit of latency
 - It can be detached from an EC2 instance and attached to another one quickly
- It's locked to an Availability Zone (AZ)
 - An EBS Volume in us-east-1a cannot be attached to us-east-1b
 - To move a volume across, you first need to snapshot it
- Have a provisioned capacity (size in GBs, and IOPS)
 - You get billed for all the provisioned capacity
 - You can increase the capacity of the drive over time

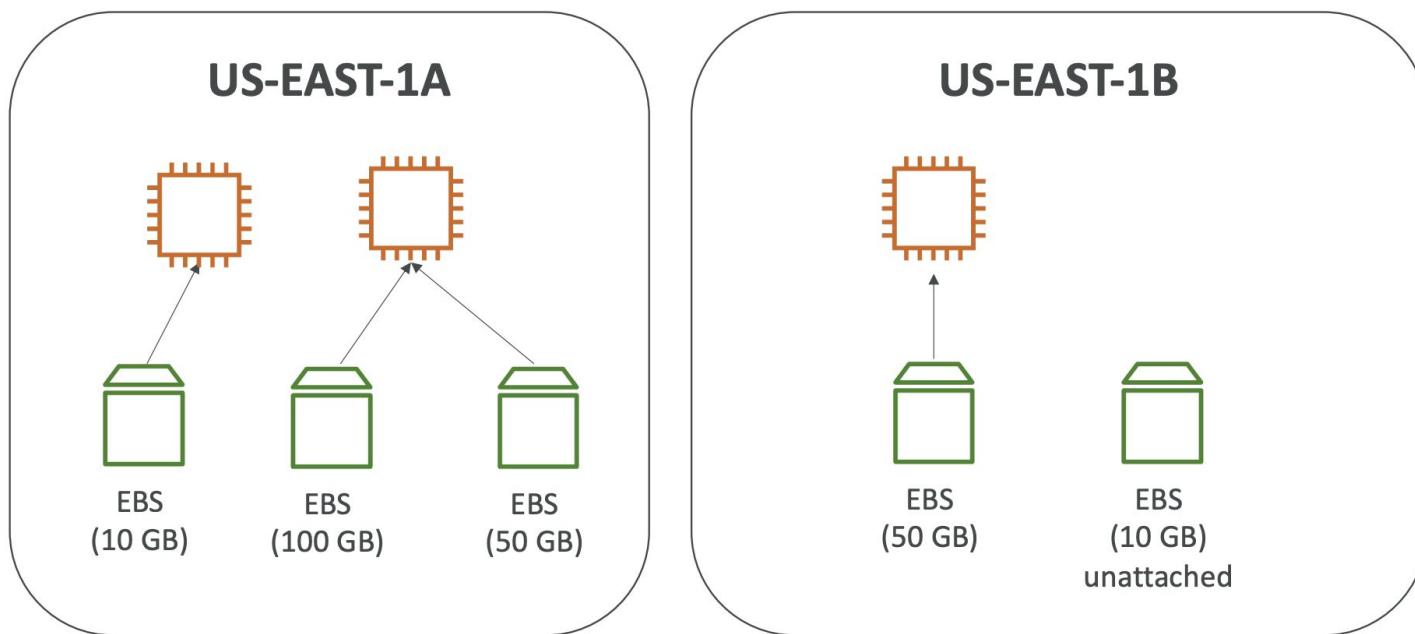
EBS Volume Types

- EBS Volumes come in 6 types
 - [gp2 / gp3 \(SSD\)](#): General purpose SSD volume that balances price and performance for a wide variety of workloads
 - [io1 / io2 Block Express \(SSD\)](#): Highest-performance SSD volume for mission-critical low-latency or high-throughput workloads
 - [st1 \(HDD\)](#): Low cost HDD volume designed for frequently accessed, throughput-intensive workloads
 - [sc1 \(HDD\)](#): Lowest cost HDD volume designed for less frequently accessed workloads
- EBS Volumes are characterized in Size | Throughput | IOPS (I/O Ops Per Sec)
- When in doubt always consult the AWS documentation – it's good!
- Only gp2/gp3 and io1/io2 Block Express can be used as boot volumes

EBS volume type limitations

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-volume-types.html>

EBS Volume - Example



EBS – Delete on Termination attribute

Volume Type	Device	Snapshot	Size (GiB)	Volume Type	IOPS	Throughput (MB/s)	Delete on Termination	Encryption
Root	/dev/xvda	snap-09f18f1862fd23a1b1	8	General Purpose SSD (gp2)	100 / 3000	N/A	<input checked="" type="checkbox"/>	Not Encrypted
EBS	/dev/sdb	Search (case-insensit	8	General Purpose SSD (gp2)	100 / 3000	N/A	<input type="checkbox"/>	Not Encrypted

Add New Volume

- Controls the EBS behaviour when an EC2 instance terminates
 - By default, the root EBS volume is deleted (attribute enabled)
 - By default, any other attached EBS volume is not deleted (attribute disabled)
- This can be controlled by the AWS console / AWS CLI
- Use case: preserve root volume when instance is terminated

EBS Snapshots

- Make a backup (snapshot) of your EBS volume at a point in time
- Not necessary to detach volume to do snapshot, but recommended
- Can copy snapshots across AZ or Region



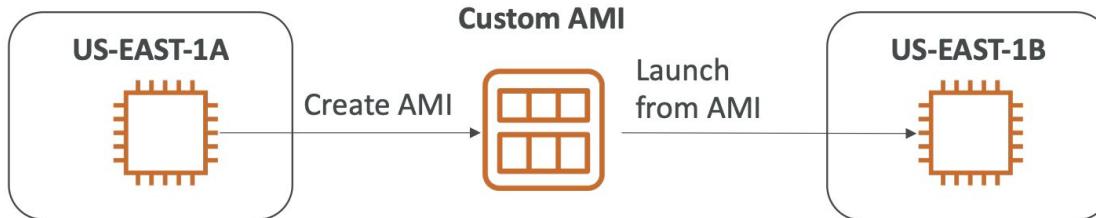
AMI Overview



- AMI = Amazon Machine Image
- AMI are a customization of an EC2 instance
 - You add your own software, configuration, operating system, monitoring...
 - Faster boot / configuration time because all your software is pre-packaged
- AMI are built for a specific region (and can be copied across regions)
- You can launch EC2 instances from:
 - A Public AMI: AWS provided
 - Your own AMI: you make and maintain them yourself
 - An AWS Marketplace AMI: an AMI someone else made (and potentially sells)

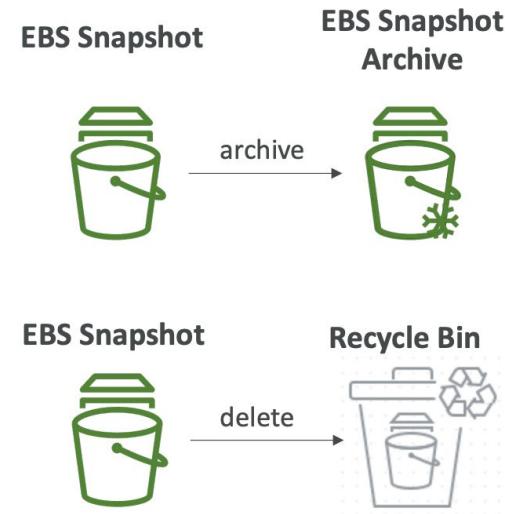
AMI Process (from an EC2 instance)

- Start an EC2 instance and customize it
- Stop the instance (for data integrity)
- Build an AMI – this will also create EBS snapshots
- Launch instances from other AMIs



EBS Snapshots Features

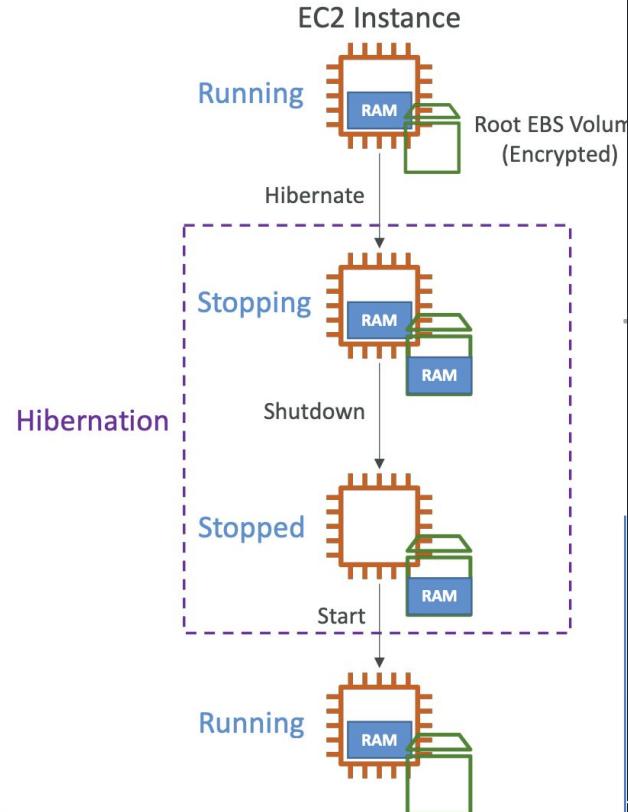
- EBS Snapshot Archive
 - Move a Snapshot to an "archive tier" that is 75% cheaper
 - Takes within 24 to 72 hours for restoring the archive
- Recycle Bin for EBS Snapshots
 - Setup rules to retain deleted snapshots so you can recover them after an accidental deletion
 - Specify retention (from 1 day to 1 year)
- Fast Snapshot Restore (FSR)
 - Force full initialization of snapshot to have no latency on the first use (\$\$\$)



EC2 hibernation

EC2 Hibernate

- Introducing EC2 Hibernate:
 - The in-memory (RAM) state is preserved
 - The instance boot is much faster! (the OS is not stopped / restarted)
 - Under the hood: the RAM state is written to a file in the root EBS volume
 - The root EBS volume must be encrypted
- Use cases:
 - Long-running processing
 - Saving the RAM state
 - Services that take time to initialize



EC2 Hibernate – Good to know

- Supported Instance Families – C3, C4, C5, I3, M3, M4, R3, R4, T2, T3, ...
 - Instance RAM Size – must be less than 150 GB.
 - Instance Size – not supported for bare metal instances.
 - AMI – Amazon Linux 2, Linux AMI, Ubuntu, RHEL, CentOS & Windows...
 - Root Volume – must be EBS, encrypted, not instance store, and large
 - Available for On-Demand, Reserved and Spot Instances
-
- An instance can NOT be hibernated more than 60 days

Userdata

Bootstrap scripts

```
#!/bin/bash
apt-get update -y
apt-get install -y apache2
echo "AWS course #2" > index.html
cp index.html /var/www/html
```

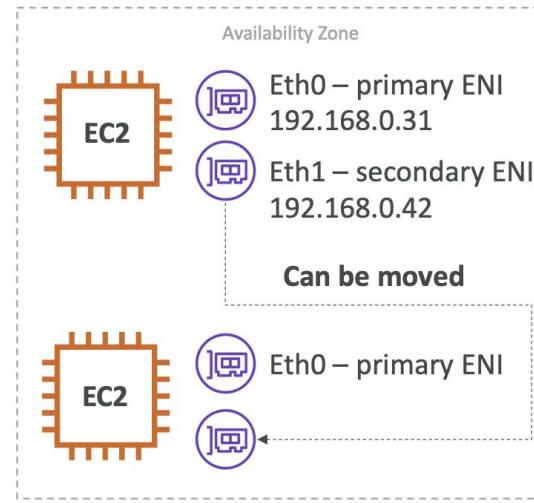
Metadata

Instance informations

<http://169.254.169.254/latest>

Elastic Network Interfaces (ENI)

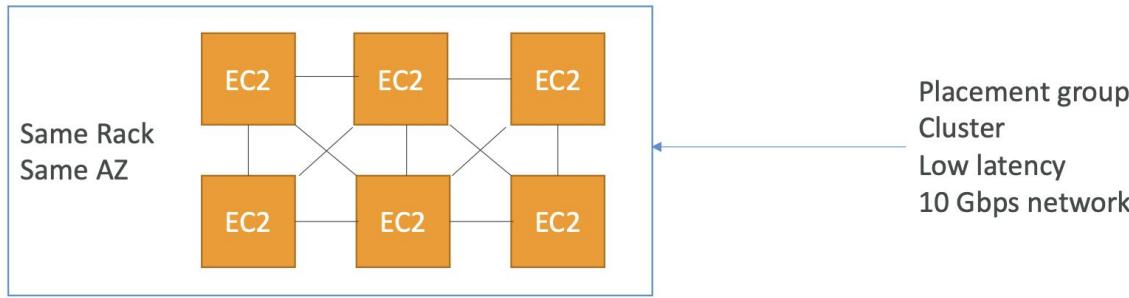
- Logical component in a VPC that represents a virtual network card
- The ENI can have the following attributes:
 - Primary private IPv4, one or more secondary IPv4
 - One Elastic IP (IPv4) per private IPv4
 - One Public IPv4
 - One or more security groups
 - A MAC address
- You can create ENI independently and attach them on the fly (move them) on EC2 instances for failover
- Bound to a specific availability zone (AZ)



Placement Groups

- Sometimes you want control over the EC2 Instance placement strategy
- That strategy can be defined using placement groups
- When you create a placement group, you specify one of the following strategies for the group:
 - *Cluster*—clusters instances into a low-latency group in a single Availability Zone
 - *Spread*—spreads instances across underlying hardware (max 7 instances per group per AZ)
 - *Partition*—spreads instances across many different partitions (which rely on different sets of racks) within an AZ. Scales to 100s of EC2 instances per group (Hadoop, Cassandra, Kafka)

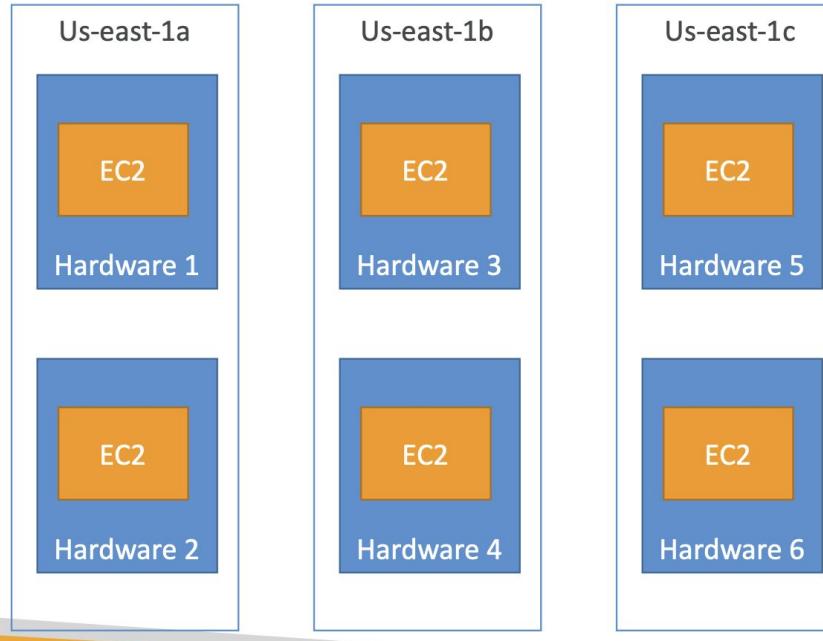
Placement Groups Cluster



- Pros: Great network (10 Gbps bandwidth between instances with Enhanced Networking enabled - recommended)
- Cons: If the rack fails, all instances fail at the same time
- Use case:
 - Big Data job that needs to complete fast
 - Application that needs extremely low latency and high network throughput

Placement Groups

Spread



- Pros:

- Can span across Availability Zones (AZ)
- Reduced risk of simultaneous failure
- EC2 Instances are on different physical hardware

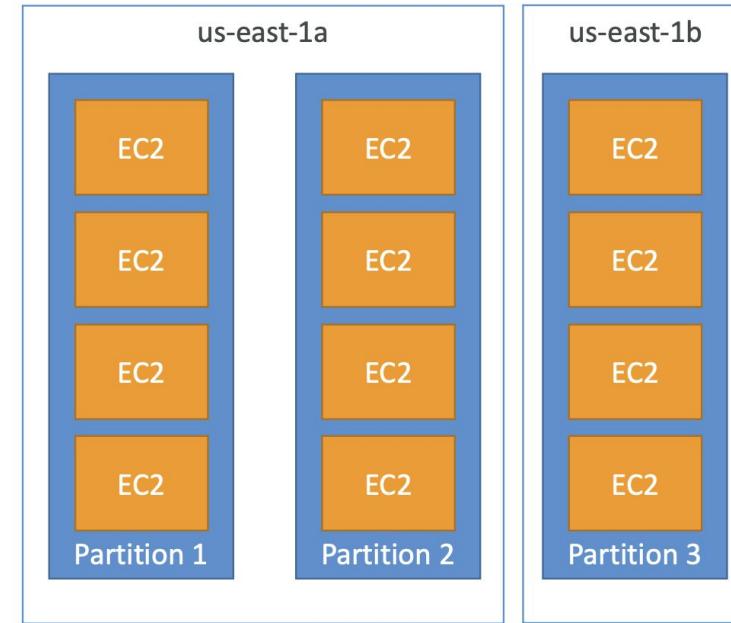
- Cons:

- Limited to 7 instances per AZ per placement group

- Use case:

- Application that needs to maximize high availability
- Critical Applications where each instance must be isolated from failure from each other

Placements Groups Partition



- Up to 7 partitions per AZ
- Can span across multiple AZs in the same region
- Up to 100s of EC2 instances
- The instances in a partition do not share racks with the instances in the other partitions
- A partition failure can affect many EC2 but won't affect other partitions
- EC2 instances get access to the partition information as metadata
- Use cases: HDFS, HBase, Cassandra, Kafka



Free tier

<https://aws.amazon.com/free/>

Homework 3

ДААЛГАВАР 1:

Fixed IP бүхий 2 сервер үүсгэнэ. (micro хэмжээтэй сервер үүсгээрэй)

1. Windows, Linux
2. Сервер лүүгээ файл хуулах эсрэг үйлдлийг бас хийх.
3. Тус бүрт нь 80-р порт дээр ямар нэг веб асаа
4. Үүсгэсэн серверээ өөр AZ, Region дээр clone-дож асаа

ДААЛГАВАР 2:

2 сервер дээрээ үүсгэсэн

- Hard дискний хэмжээг нэмэгдүүлэх (Систем дээрээ нэмэх)
- Шинэ Volume үүсгэж түүнийгээ систем дээрээ оруулж ирэх

БОНУС ДААЛГАВАР:

1. Үүсгэсэн линукс серверийн private key алга болгосон тохиолдолд EC2 серверийн хандалтаа хэрхэн эргүүлж олж авах вэ?
- 2.