Submitted by KALAISELVAN P (24.06.2023)

EC2 Task 1

1. Explain the steps to create AWS EC2?

2. Types of EC2 Instances?

3. Saving Plans/Purchasing Models of EC2?

4. What is EBS Volumes and its Types?

5. What is AMI?

6. What is Snapshot?

7. Diff btw Snashot & AMI?

8. What is Security group used for?

9. What is the use of NACL?

10. Diff btw Security Group & NACL?

11. What is EIP or Static IP?

12. Difference between EIP & Dynamic IP?

13. What is user data used for?

14. Explain the life cycle of EC2?

15. What is 2/2 Checks in EC2?

16. Diff btw Spot instance & On-Demand Instance?

17. Diff btw Spot instance & Dedicated Instance?

18. Diff btw Dedicated instance & Dedicated Host?

19. Diff btw Stop instance & Terminate instance?

20. Diff btw Stop instance & hibernate instance?

21. What is Placement group? and its types?

22. What is ENI?

23. How many Security group & EIP Can we can in a region?

24. What is the use of key pair? (Public key & Private key?

25. What is Capactiy Reservation ?

1. Explain the steps to create AWS EC2

1. Sign in to AWS Management Console
2. Select a region (for Eg: asia pacific-mumbai (ap-south-1))
3. Navigate to EC2
4. Launch Instance
5. Choose an AMI (An AMI includes the operating system, software applications, libraries, and configurations required to run an instance. AWS provides a wide range of AMIs, including various operating systems (such as Amazon Linux, Ubuntu, Windows Server, etc.) and different software configurations.)
6. Choose an instance type (Each instance type is optimized to support specific workloads and has different combinations of CPU, memory, storage, and networking capacity. t2.micro is an instance type with 1 gb memory available in free tier category)
7. Configure instance details (Here, you can specify additional configuration options like the number of instances you want to launch, network settings, subnet, security groups, IAM roles, and more.
8. Add storage (upto 30gb is available in free tier. Also we can add additional EBS (Elastic Block Store) volumes if needed.)
9. Add tags (optional)
10. Configure security groups (Security groups act as virtual firewalls for our EC2 instances, controlling inbound and outbound traffic.)
11. Review and launch
12. Create or select a key pair (If you don't have an existing key pair (private key), you'll need to create a new key pair to securely connect to your instance. Windows server needs .pem keypair and linux based servers require .ppk typenkey pair. .pem key pair can be converted to .pem by using puttygen software)
13. Launch the instance
14. View instance status (2/2 checks: 1. System checks (host hardware, network connectivity, and other system-level factors) 2. Instance checks (checking the operating system, instance status checks, and other instance-specific factors))

2. Types of EC2 Instances?

Instance Family

Current Generation Instance Types

1. **General purpose**

   It offers a proper mix of compute, memory, and networking resources and can be used for a wide range of workloads. for the applications like web servers and code repositories that use these resources in equal parts

   General-purpose instances:

   Within the General Purpose family there are many EC2 instance types:

   Mac:

   Mac instances - to access to macOS, allowing them to develop, build, test, and sign applications that require the Xcode IDE.

   T2, T3, T3a, T4g (t2.nano, t2.micro, t2.small, t2.medium, t2.large):

   The T instance family includes burstable instances that are appropriate for websites, web applications, development environments, microservices, and line of business applications.

   M4, M5zn, M5n, M5a, M5, M6g (m4.large, m4.xlarge, m4.2xlarge, m4.4xlarge)

   The M instance family is appropriate for small and medium-sized databases, data processing tasks that require more memory, cluster computing, and other enterprise applications.

   A1:

   AWS EC2 A1 instances offer significant cost savings and are ideal for scale-out and Arm-based workloads supported by the Arm ecosystem. They are the first EC2 instances powered by AWS Graviton Processors, which include 64-bit Arm Neoverse cores and AWS-designed custom silicon.

2. **Compute optimized** (high performance computing from high-performance processors) applications: batch processing workloads, media transcoding, high-performance web servers, high-performance computing (HPC), scientific modeling, dedicated gaming servers, and ad server engines, machine learning inference, and other compute-intensive applications.

   c4.large, c4.xlarge, cd.2xlarge, c4.4xlarge, cd.8xlarge

   c3.large, c3.xlarge, c3.2xlarge, c3.4xlarge, c3.8xlarge

3. **Memory optimized** (high performance database)

   Used for running large in-memory databases in the cloud, such as SAP HANA production deployments

   r3.large, r3.xlarge, r3.2xlarge, r3.4xlarge, r3.8xlarge

4. **Storage optimized** (high-speed, sequential read and write access to very large data sets on local storage)

   i2.xlarge | i2.2xlarge | i2.4xlarge | i2.8xlarge |

   d2.xlarge | d2.2xlarge | d2.4xlarge | d2.8xlarge

5. **Accelerated computing instances (**Graphics processing**)**
   Include additional hardware (GPUs, FPGAs) to provide massive amounts of parallel processing for tasks such as graphics processing.

   g2.2xlarge | g2.8xlarge

6. **High-Performance Networking:**

High network performance instances (e.g., instances with "n" in their names): These instances are designed for applications that require high network throughput, low latency, or both, such as high-performance computing (HPC) and media processing.

7. **Specialized Workloads:**

F1 instances: Intended for applications that require hardware acceleration using FPGAs (field-programmable gate arrays), such as genomics research, financial analysis, and video processing.

A1 instances: Powered by ARM-based processors, they are suitable for scale-out workloads, microservices, and web servers.

Ref: https://intellipaat.com/blog/aws-ec2-instance-types/

https://cloudacademy.com/blog/aws-ec2-instance-types-explained/

3. a). Saving Plans/ b). ec2 pricing models?

a. Saving Plans

1.EC2 instance Savings Plans:

EC2 Savings Plans provide the highest discount for Amazon Elastic Compute Cloud (EC2) instances. EC2 is a service that allows you to rent virtual servers in the cloud. With EC2 Savings Plans, you can get a discount similar to what you would get with Reserved Instances, which are long-term commitments to using specific instances. The savings plan is flexible, allowing you to change instance size and operating system. However, it is limited to a specific instance family within a region, so you can switch between similar instances but not different families.

2. Compute Savings Plans

Compute Savings Plans allow you to choose instances from different families, regions, operating systems, and even different compute services. This gives you more options and flexibility in how you use your compute resources, allowing you to save up to 66% on the costs.

Differentiate: Ec2 Savings Plans Vs Compute Savings Plans

In terms of Flexibility:

EC2 Savings Plans: While EC2 Savings Plans offer high discounts, they have limitations on flexibility. They allow you to change instance size and operating system within the same instance family and region. However, you cannot switch between different instance families.

Compute Savings Plans: Compute Savings Plans provide more flexibility. You can choose instances from different instance families, regions, operating systems, and even different compute services like EC2, Lambda, or Fargate. This flexibility allows you to optimize your compute usage across a wide range of options.

In terms of Discount Levels:

EC2 Savings Plans: The discount levels provided by EC2 Savings Plans are the same as Standard Reserved Instances. The exact discount rate depends on factors such as the term length and payment option you choose.

Compute Savings Plans: Compute Savings Plans offer a slightly lower discount level compared to EC2 Savings Plans, but they still provide significant savings. The discount rate for Compute Savings Plans is typically around 66%, while EC2 Savings Plans can reach up to 72%.

b. EC2 pricing models:

1. **On-Demand Instances:**
    - Pay for what you use, typically on a per-second basis.
    - No upfront payment or long-term commitment required.
    - Suitable for short-term and unpredictable workloads.
2. **Reserved Instances:**
    - Up to 75% discount compared to On-Demand pricing.
    - Requires upfront payment for a specified term of 1 or 3 years.
    - Reserved capacity for a specific instance type and region.
    - Recommended for steady-state usage applications with long-term commitments.

3. **Convertible Reserved Instances:**
    - Similar to Reserved Instances but with the flexibility to change the instance type.
    - Offers up to a 54% discount compared to On-Demand pricing.
    - Reservation period can be 1 or 3 years.
    - Recommended for steady-state usage applications that may require instance type changes.
4. **Scheduled Reserved Instances (Scheduled Instances):**
    - Allows you to purchase instances for a specific recurring schedule (e.g., daily, weekly).
    - Pay for the time that the instances are scheduled to run.
    - Suitable for applications that have predictable, periodic usage patterns.
5. **Spot Instances:**
    - Bid on unused EC2 instances, offering potential cost savings of up to 90% compared to On-Demand pricing.
    - The price varies based on supply and demand.
    - Instances can be terminated with a 2-minute notification if the spot price goes above your bid.
    - Recommended for **non-critical workloads**, batch processing, and tasks resilient to interruptions.
6. **Dedicated Hosts:**
    - Provides a physical server with EC2 instance capacity dedicated solely to your use.
    - Offers visibility and control over instance placement on the dedicated hardware.
    - Useful for software with complex licensing models or strong regulatory/compliance requirements.
7. **Dedicated Instances:**
    - Instances running on single-tenant hardware, but may share hardware with other instances in the same AWS account.
    - Billed by the hour.
    - No control over instance placement, as the hardware may change after stop/start.
8. **Savings Plans:**
    - A flexible cost-saving option that provides significant discounts compared to On-Demand pricing.
    - Allows you to change instance types while still enjoying the savings.
    - Recommended over Reserved Instances by AWS for its flexibility and savings potential.

4.What is EBS Volumes and its Types

General Purpose SSD (gp2):

- Volume Size: 1 GB to 16 TB
- Baseline Performance: 3 IOPS per GB, up to a maximum of 16,000 IOPS
- Maximum Throughput: 250 MiB/s
- Maximum IOPS: 16,000
- Price per GB-month: $0.10 (US East, N. Virginia)
- Recommended for: Small to medium-sized databases, development and test environments, boot volumes.
- Key features: Balanced price and performance, low-latency, suitable for general-purpose workloads.

Provisioned IOPS SSD (io1):

- Volume Size: 4 GB to 16 TB
- Baseline Performance: **50 IOPS per GB, up to a maximum of 64,000 IOPS**
- Maximum Throughput: 500 MiB/s
- Maximum IOPS: 64,000
- Price per GB-month: $0.125 (US East, N. Virginia)
- Price per provisioned IOPS-month: $0.065 (US East, N. Virginia)
- Recommended for: I/O-intensive applications, large databases, critical workloads.
- Key features: High performance, low-latency, predictable IOPS, suitable for applications requiring specific IOPS requirements.

Throughput Optimized HDD (st1):

- Volume Size: 500 GB to 16 TB
- Baseline Performance: 40 MiB/s per TB, up to a maximum of 500 MiB/s
- Maximum Throughput: 500 MiB/s
- Maximum IOPS: 500
- Price per GB-month: $0.045 (US East, N. Virginia)
- Recommended for: Big data, data warehouses, log processing, streaming workloads.
- Key features: Low-cost storage, high throughput, suitable for large, sequential workloads with frequent access.

Cold HDD (sc1):

- Volume Size: 500 GB to 16 TB
- Baseline Performance: 12 MiB/s per TB, up to a maximum of 250 MiB/s
- Maximum Throughput: 250 MiB/s
- Maximum IOPS: 250
- Price per GB-month: $0.025 (US East, N. Virginia)
- Recommended for: Data archiving, backup, infrequently accessed workloads.
- Key features: Lowest cost per gigabyte, lower throughput compared to other types, suitable for workloads with less frequent access.

Magnetic (standard):

- Volume Size: 1 GB to 1 TB
- Baseline Performance: 3 IOPS per GB, up to a maximum of 200 IOPS
- Maximum Throughput: 100 MiB/s
- Maximum IOPS: 200
- Price per GB-month: $0.05 (US East, N. Virginia)
- Recommended for: Workloads with light I/O requirements, cost optimization.
- Key features: Legacy volume type, lower performance compared to other types, suitable for low-intensity workloads with budget constraints.
- Remember to consider your specific workload requirements, performance needs, and budget when choosing the appropriate EBS volume type.

5.What is AMI?

AWS provides a number of AMIs with pre-loaded operating systems such as Windows, Amazon Linux and Ubuntu. Some of them contain additional software, such as Windows with SQL Server.

An AMI consists of an operating system, pre-installed software, and configuration settings that enable the creation of EC2 instances with specific characteristics. It serves as a template for launching identical instances multiple times, ensuring consistency and easy replication of server environments.

AMI provides several benefits, including:

Easy replication: AMIs allow users to create identical instances quickly. This is useful for scaling applications, creating backups, and setting up development and testing environments.

Customization: Users can customize the configuration settings, install additional software, and make modifications to the AMI to suit their specific requirements.

Versioning: AMIs can have multiple versions, enabling users to track and manage changes over time.

Security and reliability: AWS provides a wide range of AMIs, including hardened and security-focused images, which can help ensure a secure and reliable foundation for EC2 instances.

There are also community AMIs that are created by somebody other than AMI, but shared to all users. For example, a company might load a demo version of their software onto the AMI, so customers can simply launch an Amazon EC2 instance and it will have all software already loaded and configured.

6.What is Snapshot

EBS snapshots in AWS are point-in-time copies of data on Elastic Block Store volumes, providing backup and recovery capabilities. They use incremental backups to store only changed data, reducing time and storage requirements. Snapshots allow for point-in-time restores, enabling data rollback or recovery from corruption. AWS ensures data consistency during snapshot creation. Snapshot lifecycle management features automate regular snapshots and retention policies. Replication to different regions adds an extra layer of data protection.

7. Difference between Snapshot & AMI?

1. Snapshots are used for backup and recovery of data on EBS volumes, while AMIs are pre-configured templates for launching instances.
2. Snapshots capture the data on an EBS volume, while AMIs capture the entire system state including the operating system, applications, and configurations.
3. Snapshots are stored in Amazon S3 and are incremental, while AMIs are stored in Amazon EBS and can be EBS-backed or instance-store-backed.
4. Snapshots are primarily used for data backup and restoring volumes, while AMIs are used to launch instances with specific configurations for deploying applications.
5. In short, snapshots backup data, while AMIs provide a complete system image for launching instances.

8. What is Security group used for?

AWS Security Groups are virtual firewalls that control inbound and outbound traffic for Amazon Web Services (AWS) resources, such as Amazon EC2 instances, RDS databases, and load balancers. They act as a fundamental security mechanism in the AWS infrastructure, allowing you to define and manage network access rules.

Here are some key points about AWS Security Groups:

Inbound and Outbound Rules: Security Groups are associated with AWS resources and provide a way to define inbound and outbound traffic rules. Inbound rules control incoming traffic to the resource, while outbound rules govern outgoing traffic from the resource.

Stateful Filtering: Security Groups use stateful filtering, meaning that once a connection is allowed, the responses to that connection are automatically allowed as well. This eliminates the need to write separate rules for inbound and outbound traffic.

Port and Protocol Control: You can specify rules based on protocols (e.g., TCP, UDP, ICMP) and ports (e.g., HTTP on port 80, SSH on port 22) to allow or deny traffic. Security Groups support both single ports and port ranges.

Instance-Level Security: Each EC2 instance is associated with one or more Security Groups, acting as a virtual firewall around the instance. Multiple instances can share the same Security Group, allowing them to have similar security configurations.

Dynamic Rule Updates: You can modify Security Group rules dynamically, allowing you to adapt to changing requirements without manually reconfiguring instances.

Zero Trust Model: By default, Security Groups follow a "deny all" approach, meaning that all inbound traffic is blocked unless explicitly allowed by rules. Outbound traffic is allowed by default unless denied by rules.

VPC Bound: Security Groups are specific to a Virtual Private Cloud (VPC) and are primarily used for controlling traffic within the VPC. They cannot be directly shared across VPCs.

AWS Security Groups play a crucial role in securing AWS resources by limiting access to only necessary network traffic. They provide a flexible and scalable way to define network security policies and are an integral part of the overall security architecture in AWS.

9. What is the use of NACL?

Network Access Control Lists (ACLs) are used for:

1. **Filtering traffic:** They permit or deny traffic based on rules, controlling which packets can pass through network devices.
2. **Enforcing security:** They prevent unauthorized access, block specific IP addresses, restrict ports/protocols, and counteract attacks.
3. **Prioritizing traffic:** They allocate resources to critical applications like voice or video data for optimal performance.
4. **Redirection:** They redirect traffic from certain sources to different gateways or servers for management and security purposes.

5. **Network segmentation:** They divide a network into segments, allowing separate security policies and controlling inter-segment communication.

In summary, ACLs provide control, security, performance optimization, traffic management, and network segmentation.

10.Differentiate between Security Group & NACL?

Function:

Security Group: Controls traffic at the instance level.

NACL: Controls traffic at the subnet level.

Scope:

Security Group: Associated with instances, applies to individual instances.

NACL: Associated with subnets, applies to all instances within the subnet.

Statefulness:

Security Group: Stateful - automatically allows return traffic.

NACL: Stateless - separate rules needed for return traffic.

Rule Evaluation:

Security Group: Rules evaluated in order, most specific rule is applied.

NACL: Rules evaluated in order, first matching rule is applied.

Rule Configuration:

Security Group: Uses allow rules for specified protocols, ports, and IP ranges.

NACL: Uses both allow and deny rules for specified protocols, ports, and IP ranges.

Number of Rules:

Security Group: Can have multiple inbound and outbound rules.

NACL: Limited to 20 inbound and 20 outbound rules.

In summary, security groups control traffic at the instance level, while NACLs control traffic at the subnet level. Security groups are stateful and automatically allow return traffic, while NACLs are stateless and require separate rules for return traffic. Security groups have more flexibility and fine-grained control, while NACLs provide broader control within a subnet.

11. What is EIP or Static IP?

Elastic IP (EIP) is a feature offered by cloud computing platforms like AWS. It provides a static, public IP address for virtual machines or instances. Unlike dynamic IP addresses that can change, an EIP remains constant even when the instance is restarted. It ensures consistent accessibility for applications or services that rely on a fixed IP address. EIPs are commonly

used for hosting websites or running servers that require stable IP addresses for DNS configuration or security. However, it's important to note that EIPs can be released and reassigned if not associated with an instance

12. Difference between EIP & Dynamic IP?

S. No	Elastic IP:	Dynamic IP:
1	Static and doesn't change when you restart your instance	Temporary and can change when you restart your device or network.
2	You can associate and disassociate an Elastic IP from instances as needed, allowing you to redirect the IP address to different instances or services within your account.	Dynamic Host Configuration Protocol (DHCP) servers allocate the Dynamic IP addresses to the devices on a network.
3	Suitable for specific use cases: Elastic IPs are commonly used for scenarios where a stable IP address is required, such as hosting a website or running a server that relies on a consistent IP for DNS configuration or security purposes.	Common for residential and consumer internet connections: Dynamic IP addresses are commonly used for home internet connections or consumer-grade network setups where there is no requirement for a fixed IP address.

13. What is user data used for?

EC2 user data is a set of instructions that are executed when an Amazon Elastic Compute Cloud (EC2) instance boots up. It can be used to configure the instance, install software, or run scripts. User data can be specified in a variety of formats, including:

Shell scripts
Cloud-Init directives
Base64-encoded text

When user data is specified, it is executed by the cloud-init software, which is responsible for configuring EC2 instances. Here are some of the common use cases for EC2 user data:

Installing software
Configuring network settings
Creating users and groups
Starting services
Running scripts
Provisioning infrastructure

Bootstrapping containers

EC2 User Data: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/user-data.html

Cloud-Init: https://cloudinit.readthedocs.io/en/latest/

14. Explain the life cycle of EC2?

The life cycle of an EC2 instance with headings:

Launch

- EC2 instance is created with specified parameters.
- Instance enters the "pending" state.

Running

- Instance successfully launched and enters the "running" state.
- Ready to run applications and perform tasks.

Stop

- Instance can be temporarily stopped without termination.
- State changes to "stopped" and retains data and configuration.
- Compute usage is no longer billed.

Hibernate

- Support for hibernation allows saving instance state to root EBS volume.
- Instance is stopped and state, including RAM contents, is preserved.
- Resumes from where it left off when started again.

Reboot

- Instance is restarted while maintaining configuration and data.
- Can be done through the AWS Management Console, CLI, or SDKs.
- Useful for troubleshooting or applying minor updates.

Termination

- Instance is permanently shut down and associated resources are released.
- State changes to "shutting down" and eventually "terminated".
- Data loss occurs if necessary information isn't backed up.

Instance Retirement

- Over time, instances may reach hardware retirement date.
- AWS schedules retirement due to hardware degradation or operational reasons.
- Notifications received, and replacement instance should be launched before retirement date.

Deletion

- In addition to termination, underlying resources like AMIs, snapshots, or EBS volumes can be deleted.

- Cleans up the environment and reduces costs.
- Some resources may have separate retention policies.

Throughout the life cycle, various AWS services and tools can be used for efficient management and monitoring of EC2 instances. These include Elastic Load Balancing (ELB), Auto Scaling, Amazon CloudWatch, and AWS Systems Manager.

15. What is 2/2 Checks in EC2?

2/2 checks: 1. **System checks** (host hardware, network connectivity, and other system-level factors) 2. **Instance checks** (checking the operating system, instance status checks, and other instance-specific factors

16. Diff btw Spot instance & On-Demand Instance?

Spot instances and On-Demand instances are two pricing models in Amazon EC2. Spot instances involve bidding on prices and can be interrupted if the bid price is exceeded, while On-Demand instances have fixed rates and provide uninterrupted access. Spot instances offer cost savings but come with the risk of interruption. On the other hand, On-Demand instances ensure steady performance and reliability. When using Spot instances, you must be prepared for interruptions and receive a two-minute warning before interruption occurs. Termination protection cannot be enabled for Spot instances.

17. Diff btw Spot instance & Dedicated Instance?

Spot instances and Dedicated instances in Amazon EC2 are two distinct types of instances. Spot instances involve bidding on prices and can be interrupted if the Spot price exceeds the bid price. They are cost-effective and suitable for flexible workloads. On the other hand, Dedicated instances run on hardware dedicated to a single AWS account, providing isolation and enhanced privacy. They are ideal for workloads with **strict compliance or security requirements**. Dedicated instances are not shared with other customers and can be launched on-demand. Their pricing is typically higher due to the dedicated hardware.

18. Diff btw Dedicated instance & Dedicated Host?

Dedicated instances and Dedicated Hosts are two options for achieving dedicated resources in Amazon EC2. Dedicated instances provide instance-level isolation on dedicated hardware managed by AWS, ensuring privacy and compliance. Dedicated Hosts offer host-level isolation on dedicated physical servers, allowing more control over the underlying infrastructure. Dedicated instances are suitable for enhanced privacy and compliance, while Dedicated Hosts are ideal for specific licensing needs and strict control over physical placement.

19. Diff btw Stop instance & Terminate instance?

Stopping an instance in Amazon EC2 halts it temporarily, saving its state and configuration. The instance can be started again and retains its data, but the RAM contents are not preserved.

On the other hand, terminating an instance permanently shuts it down, removes all associated resources, and cannot be reversed. Terminated instances cannot be restarted, and data loss occurs. It is important to back up necessary data before termination.

20. Diff btw Stop instance & hibernate instance?

Stopping an EC2 instance puts it in a halted state, saving its data and configuration but not preserving the in-memory state. When the instance is started again, it boots up as if it were a new instance. On the other hand, hibernating an instance is a feature available for supported instance types. It saves the current state of the instance, including the contents of RAM, to the root EBS volume. Hibernated instances can be resumed from where they left off, maintaining their exact in-memory state. While stopped instances do not incur ongoing costs, hibernated instances require storage resources and continue to accrue costs for maintaining the preserved state.

21. What is Placement group? and its types?

In Amazon EC2, a placement group is a logical grouping of instances within a single Availability Zone. Placement groups are used to influence the placement of instances to meet specific requirements and optimize network performance. Here are the types of placement groups available in EC2:

Cluster Placement Group(same hardware):

A Cluster placement group is designed for applications that require low-latency, high-bandwidth networking between instances.

Instances within a Cluster placement group are tightly packed within a single rack, which enables high network performance.

It is ideal for applications that require high-performance computing, big data analytics, or tightly coupled workloads.

Spread Placement Group(different hardwares- max:7 instances / group/ AZ):

A Spread placement group spreads instances across underlying hardware to reduce the risk of simultaneous failures.

Instances within a Spread placement group are placed on distinct racks,(different hardwares) which enhances availability and fault tolerance.

It is suitable for applications that have a small number of critical instances or have specific resilience requirements.

Partition Placement Group (may be same or different hardwares):

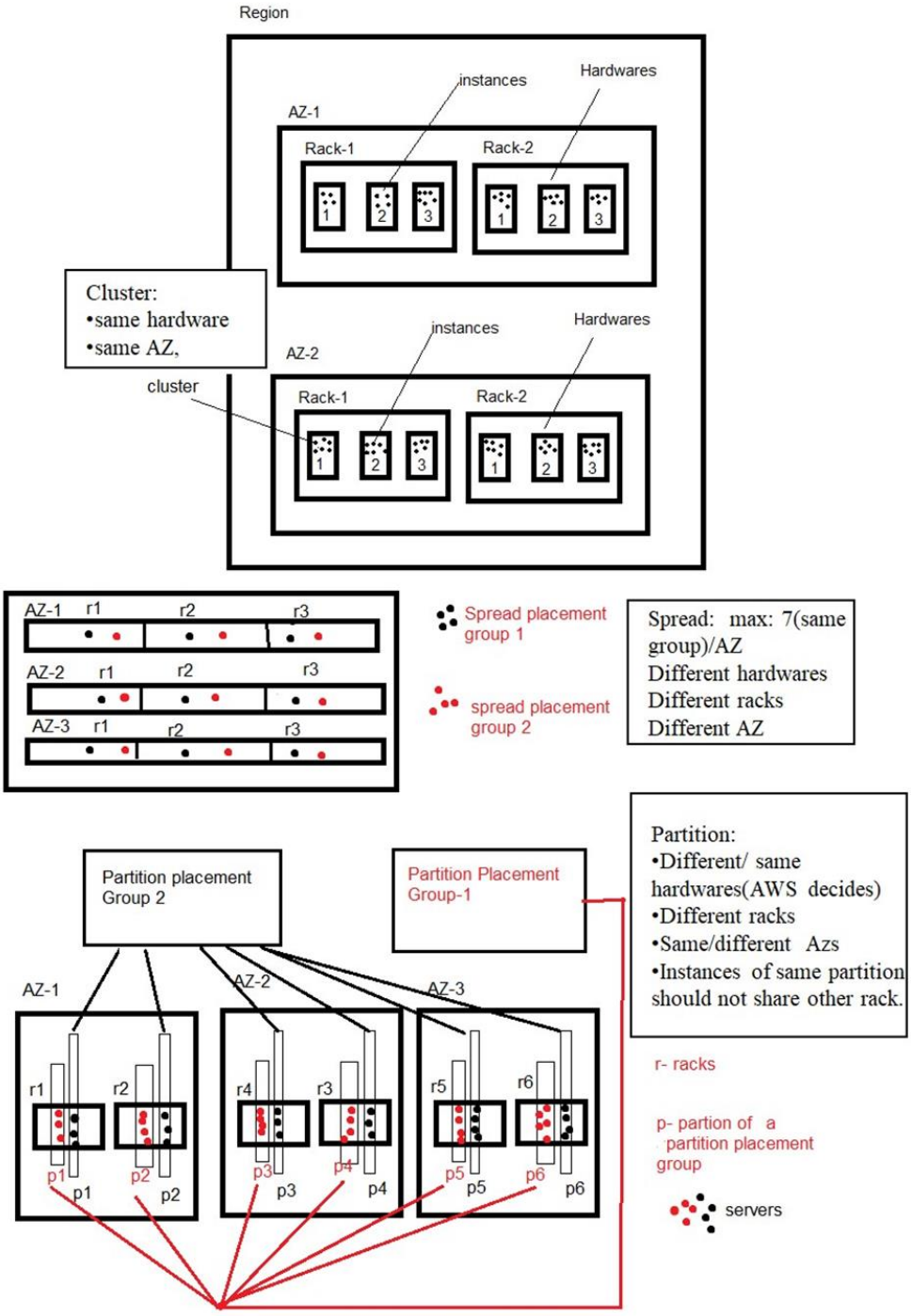
A partition placement group allows you to spread instances across logical partitions called partitions, within an AZ or many AZs in a region.

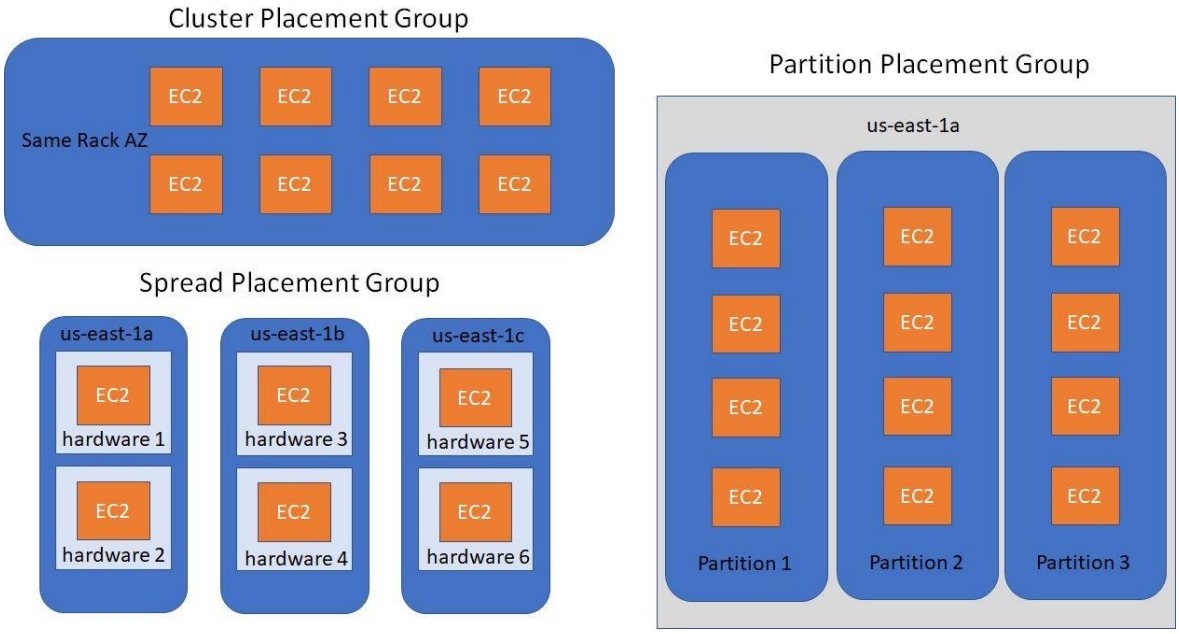
Each partition within the group has its own set of racks, power sources, and network switches.

Instances within a partition placement group can benefit from higher aggregate network bandwidth compared to instances in a single instance placement group.

This type of placement group is useful for large distributed and replicated workloads that can benefit from increased network performance and reduced network contention.

7 partitions/AZ, 100 EC2s/group,





It's important to note that the availability of placement groups may vary by region and instance type. Placement groups cannot span multiple Availability Zones, and the choice of placement group cannot be changed after instances are launched.

Placement groups can significantly impact network performance and fault tolerance, so it's crucial to carefully consider the requirements of your application and select the appropriate placement group type accordingly.

22. What is ENI? (Amazon EC2)

ENI stands for **Elastic Network Interface**. An Elastic Network Interface is a virtual network interface that can be attached to an EC2 instance in a Virtual Private Cloud (VPC). ENIs enable communication between instances within a VPC and with other AWS services and resources.
Key features of ENIs include:
Networking Capabilities: ENIs provide networking capabilities, such as assigning private IP addresses, MAC addresses, and security groups to EC2 instances.

Elasticity: ENIs can be attached and detached from EC2 instances as needed, allowing for flexibility in network configuration and instance management.
Multiple ENIs per Instance: EC2 instances can have multiple ENIs attached, allowing for different network configurations and facilitating scenarios like hosting multiple network-facing applications on a single instance.

High Performance: ENIs support high-performance networking features, such as enhanced networking and placement in placement groups, enabling low-latency, high-bandwidth communication between instances.

Integration with AWS Services: ENIs can be associated with other AWS services, such as Elastic Load Balancer (ELB), Network Load Balancer (NLB), and Elastic IP addresses, to enable network traffic distribution and public IP connectivity.

ENI Attributes: ENIs have various attributes, including a primary private IP address, secondary private IP addresses, a MAC address, a source/destination check flag, and security group associations.

ENIs play a critical role in connecting EC2 instances to networks and other resources, facilitating secure and scalable networking within the AWS infrastructure.

(https://www.youtube.com/watch?v=lRDUzkvoC8A)

23. How many Security group & EIP Can we have in a region?

The maximum number of security groups and Elastic IP (EIP) addresses that you can have in a region depends on the AWS service limits. The specific limits can vary and are subject to change over time.

The following are the quotas for security groups and EIPs in a region:

Security groups:

Default: 2,500

Adjustable: Yes (more than 5000)

Inbound or outbound rules per security group:

Default: 60

Adjustable: Yes

Security groups per network interface:

Default: 5

Adjustable: Up to 16

Elastic IP addresses per region:

Default: 5

Elastic IP addresses per public NAT gateway:

Default: 2

Adjustable: Yes

The following are some additional limitations to keep in mind:

- The total number of rules per network interface cannot exceed 1,000.
- The total number of security groups per instance cannot exceed 5.
- The total number of EIPs per account cannot exceed 50.

https://docs.aws.amazon.com/vpc/latest/userguide/amazon-vpc-limits.html

You can increase this limit by submitting a request through the AWS Support Center.

It's important to note that these limits are subject to change, and AWS may have different limits for specific regions or for different account types, such as new accounts or accounts with higher resource requirements. To get the most up-to-date and accurate information about your specific account's limits, it is recommended to consult the AWS documentation or contact AWS Support.

To view and manage your current resource limits, you can access the AWS Management Console, navigate to the EC2 service, and check the corresponding sections for security groups and Elastic IP addresses.

24. What is the use of key pair? (Public key & Private key)?

In Amazon EC2, a key pair consists of a public key and a private key. The public key is used for encrypting data and is associated with an EC2 instance during launch. It allows secure remote access to the instance. The private key, on the other hand, is securely stored and is used to decrypt data encrypted with the public key. It is used for authentication when connecting to an EC2 instance, proving ownership of the associated public key. The key pair provides a secure method for establishing encrypted communication and ensuring only authorized individuals can access EC2 instances.

25. What is Capacity Reservation?

Capacity Reservation in Amazon EC2 allows users to reserve capacity for specific EC2 instances in a particular Availability Zone. It guarantees dedicated capacity for your account, ensuring availability even during peak times or when capacity is limited. With Capacity Reservations, users have control over instance placement within an Availability Zone, making it useful for licensing requirements or specific hardware needs. It offers flexibility by supporting On-Demand instances, Reserved Instances, and Spot Instances. Capacity Reservations are tied to a specific Availability Zone, and usage can be tracked using AWS CloudWatch or the EC2 API. This feature is particularly beneficial for predictable workloads, compliance requirements, and software licensing restrictions, enabling users to ensure resource availability and have control over instance placement.