

JON BONSO

---

AWS CERTIFIED  
**CLOUDOPS  
ENGINEER  
ASSOCIATE**  
VERSION 3.0

---



**Tutorials Dojo Study Guide**



## TABLE OF CONTENTS

|  |           |
|--|-----------|
| <b>INTRODUCTION</b>  | <b>8</b>  |
| <b>AWS CERTIFIED CLOUDOPS ENGINEER ASSOCIATE EXAM OVERVIEW</b>               | <b>9</b>  |
| Exam Details   | 9         |
| Exam Scoring   | 11        |
| Exam Benefits  | 12        |
| <b>AWS CERTIFIED CLOUDOPS ENGINEER ASSOCIATE EXAM - STUDY GUIDE AND TIPS</b> | <b>13</b> |
| Study Materials  | 13        |
| AWS Services to Focus On   | 14        |
| Common Exam Scenarios  | 16        |
| Validate Your Knowledge  | 21        |
| Sample Practice Test Questions:  | 21        |
| Question 1   | 21        |
| Question 2   | 24        |
| <b>AWS Deep Dives</b>  | <b>27</b> |
| Amazon EC2   | 27        |
| Components of an EC2 Instance  | 29        |
| Types of EC2 Instances   | 30        |
| Storage with Highest IOPS for EC2 Instance                                   | 31        |
| Instance Purchasing Options  | 32        |
| EC2 Placement Groups   | 34        |
| Cluster Placement Group  | 35        |
| Partition Placement Group  | 36        |
| Spread Placement Group   | 38        |
| EC2 Image Builder  | 40        |
| Image Pipelines  | 40        |
| Image Recipes Configuration  | 40        |
| Source Images  | 40        |
| Build and Test Components  | 42        |
| Storage  | 43        |
| Infrastructure Configuration   | 43        |
| Distribution Settings  | 44        |
| Amazon EC2Rescue   | 45        |
| EC2Rescue for Windows Server   | 45        |
| Diagnose and Rescue an Offline Instance                                      | 45        |



---

|  |    |
|--|----|
| Collecting Logs from an Offline Instance           | 55 |
| Restore Options for an Offline Instance            | 58 |
| Checking the Current Instance                      | 60 |
| EC2Rescue for Windows on Systems Manager           | 62 |
| EC2Rescue for Linux                                | 63 |
| Installing EC2Rescue for Linux                     | 64 |
| Diagnose Issues Using EC2Rescue for Linux          | 64 |
| Creating Instance Backup Using EC2Rescue for Linux | 65 |
| EC2Rescue for Linux on Systems Manager             | 66 |
| AWS Auto Scaling                                   | 67 |
| Auto Scaling Group                                 | 67 |
| Auto Scaling Templates                             | 67 |
| Launch Templates                                   | 67 |
| Auto Scaling Group Configuration                   | 71 |
| Kubernetes Vertical Pod Autoscaler                 | 78 |
| AWS Global Accelerator                             | 78 |
| AWS Compute Optimizer                              | 78 |
| Prerequisites                                      | 79 |
| AWS Compute Optimizer Dashboard                    | 79 |
| Recommendations for EC2 Instances                  | 81 |
| Recommendations for Auto Scaling Group             | 82 |
| Recommendations for EBS Volume Instances           | 83 |
| Elastic Load Balancing                             | 85 |
| Load Balancer Types                                | 85 |
| ELB Features and Components                        | 86 |
| Load Balancer Scheme                               | 86 |
| IP Addresses Type                                  | 86 |
| Listener   | 86 |
| Target Group                                       | 86 |
| Security Groups                                    | 87 |
| Availability Zones                                 | 87 |
| Health Checks                                      | 87 |
| Sticky Sessions                                    | 87 |
| Cross-zone Load Balancing                          | 87 |
| Connection Draining                                | 87 |
| Load Balancer Monitoring                           | 88 |
| Delete Protection                                  | 88 |

---



---

|  |     |
|--|-----|
| Choosing the Right Load Balancer         | 88  |
| Application Load Balancer (ALB)          | 88  |
| Network Load Balancer (NLB)              | 88  |
| Gateway Load Balancer (GWLB)             | 89  |
| S3 Presigned URL                         | 89  |
| Sharing S3 objects using Presigned URL   | 89  |
| Uploading S3 Objects Using Presigned URL | 92  |
| S3 Transfer Acceleration                 | 93  |
| Amazon CloudFront                        | 95  |
| Caching Process                          | 95  |
| CloudFront Policies                      | 95  |
| Cache Policy                             | 96  |
| Origin Request Policy                    | 97  |
| Amazon ElastiCache                       | 99  |
| ElastiCache Memcached and Redis Engine   | 99  |
| Clusters                                 | 99  |
| Sharding                                 | 99  |
| Multithreading                           | 100 |
| High Availability                        | 100 |
| Backup and Restore                       | 101 |
| Key Points:                              | 101 |
| Virtual Private Cloud                    | 102 |
| Network Access Control List (NACL)       | 102 |
| Route Tables                             | 104 |
| VPC Flow logs                            | 106 |
| Traffic Mirroring                        | 107 |
| Amazon Route 53                          | 110 |
| Domain Registration                      | 110 |
| Route 53 Service Integrations            | 111 |
| Hosted Zones                             | 111 |
| Route 53 Health Checks                   | 112 |
| Route 53 Records                         | 112 |
| Routing Policy                           | 114 |
| DNS Record Types                         | 115 |
| Route 53 Resolver                        | 116 |
| Resolver Endpoints                       | 116 |
| Resolver Rules                           | 119 |

---





---

|   |     |
|---|-----|
| Amazon Elastic File System (EFS)                                  | 120 |
| EFS Storage Classes   | 120 |
| Creating a File System  | 121 |
| File System Access Point  | 125 |
| Mounting a File System  | 127 |
| Mount via DNS   | 129 |
| Mount via IP  | 129 |
| Amazon FSx  | 130 |
| Amazon FSx for Windows File Server                                | 130 |
| File System Details   | 130 |
| Optional Details  | 132 |
| Working with Amazon FSx for Windows File Server                   | 132 |
| Amazon FSx for Lustre   | 133 |
| File System Details   | 133 |
| Optional Details  | 134 |
| Working with Amazon FSx for Lustre                                | 134 |
| AWS DataSync  | 136 |
| Supported AWS Storage Service                                     | 136 |
| Working with DataSync   | 136 |
| AWS Transfer Family   | 138 |
| AWS Backup  | 139 |
| Backup Plan   | 139 |
| On-demand Backup  | 141 |
| Backup Vault  | 142 |
| Protected Resources   | 143 |
| Backup Jobs   | 143 |
| Cross-account Management  | 144 |
| Amazon Relational Database Service (RDS)                          | 145 |
| Amazon RDS Features and Components                                | 145 |
| Amazon RDS Database Engines                                       | 145 |
| Choosing Suitable RDS DB Instance Classes                         | 146 |
| Choosing the Right RDS DB Instance Storages                       | 146 |
| Choosing a Region and Availability Zone for RDS Instance          | 148 |
| Increasing Database Availability Using Multi-AZ Deployment        | 148 |
| Improving Database Performance using Read Replica and DB Clusters | 149 |
| Adding an RDS Proxy   | 150 |
| Working with RDS Backup   | 151 |

---



---

|  |     |
|--|-----|
| Monitoring a Database Instance                 | 153 |
| Deleting a Database Instance                   | 156 |
| Amazon OpenSearch Service                      | 158 |
| AWS Config                                     | 158 |
| AWS Config Continuous Configuration Monitoring | 159 |
| Deploying Resources with CloudFormation        | 161 |
| StackSets                                      | 162 |
| Nested Stacks                                  | 162 |
| Deleting a Stack                               | 163 |
| Retain   | 164 |
| Snapshot                                       | 164 |
| AWS Systems Manager Patch and Change Manager   | 166 |
| AWS Systems Manager Patch Manager              | 166 |
| AWS Systems Manager Change Manager             | 167 |
| Encryption on AWS Storage Services             | 170 |
| S3 Encryption                                  | 170 |
| Server-Side Encryption                         | 170 |
| Client-Side Encryption                         | 171 |
| Encrypting Existing S3 Objects                 | 171 |
| EFS Encryption                                 | 171 |
| Data at Rest Encryption                        | 171 |
| Data In Transit Encryption                     | 172 |
| EBS Encryption                                 | 172 |
| Creating Encrypted EBS Volume                  | 172 |
| Snapshots                                      | 173 |
| RDS Encryption                                 | 174 |
| Encrypting RDS Database Instance with AWS KMS  | 174 |
| Securing Database Connection on RDS            | 175 |
| Security on AWS                                | 176 |
| AWS KMS Key Rotation                           | 176 |
| Secrets Manager vs Parameter Store             | 177 |
| IAM Access Analyzer                            | 179 |
| AWS Certificate Manager                        | 181 |
| Amazon Detective                               | 183 |
| Amazon GuardDuty                               | 183 |
| AWS Firewall Manager                           | 183 |
| AWS Directory Service                          | 184 |

---



---

|  |            |
|--|------------|
| AWS Billing and Governance   | 186        |
| AWS Organizations  | 186        |
| Service Control Policies (SCP)   | 189        |
| Cost Explorer  | 190        |
| Cost Allocation Tags   | 191        |
| AWS Cost and Usage Report  | 192        |
| AWS License Manager  | 193        |
| Monitoring and Logging on AWS  | 195        |
| CloudWatch Metrics for EC2   | 195        |
| Creating CloudWatch Alarm  | 198        |
| Working with CloudWatch Logs   | 205        |
| Event-driven Architecture with Amazon EventBridge  | 209        |
| Exploring Events on CloudTrail   | 211        |
| AWS Health Dashboard   | 214        |
| AWS Trusted Advisor  | 214        |
| AWS tools and SDKs   | 214        |
| <b>COMPARISON OF AWS SERVICES</b>  | <b>215</b> |
| S3 vs EBS vs EFS   | 215        |
| Amazon S3 vs Glacier   | 218        |
| S3 Standard vs S3 Standard-IA vs S3 One Zone-IA vs S3 Intelligent Tiering vs S3 Express One Zone | 219        |
| AWS DataSync vs Storage Gateway  | 221        |
| S3 Transfer Acceleration vs Direct Connect vs VPN vs Snowball                                    | 223        |
| Amazon EBS: SSD vs HDD   | 226        |
| Amazon RDS vs Amazon DynamoDB  | 229        |
| Amazon RDS vs Amazon Aurora  | 232        |
| Multi-AZ deployments vs. Multi-Region deployments vs. Read Replicas                              | 237        |
| Amazon Container Services (Amazon ECS) vs AWS Lambda   | 238        |
| Security Group vs NACL   | 240        |
| Application Load Balancer vs Network Load Balancer vs Gateway Load Balancer                      | 242        |
| EC2 Instance Health Check vs ELB Health Check vs Auto Scaling and Custom Health Check            | 245        |
| ELB Health Checks vs Route 53 Health Checks For Target Health Monitoring                         | 248        |
| AWS CloudTrail vs Amazon CloudWatch  | 249        |
| CloudWatch Agent vs SSM Agent vs Custom Daemon Scripts   | 250        |
| Latency Routing vs Geoproximity Routing vs Geolocation Routing                                   | 251        |
| Service Control Policies vs IAM Policies   | 253        |
| S3 Pre-Signed URLs vs CloudFront Signed URLs vs Origin Access Control                            | 255        |
| SNI Custom SSL vs Dedicated IP Custom SSL  | 256        |

---



---

|   |            |
|---|------------|
| Redis (cluster mode enabled vs disabled) vs Memcached | 257        |
| <b>FINAL REMARKS AND TIPS</b>                         | <b>259</b> |
| <b>ABOUT THE AUTHOR</b>                               | <b>260</b> |



## INTRODUCTION

Today, we live in a world of fast-paced innovation and the invention of new technologies, where competitors race to develop the next disruptive product in the market. Companies with on-premises resources are quickly shifting to the cloud, such as AWS, for the many advantages that it brings. Furthermore, AWS has been the leading cloud service provider for the past few years and is continually releasing brand-new offerings. Millions of users and businesses have already adopted the AWS platform for their operations, but not all can capitalize on the benefits that AWS brings to its customers. It takes well-trained individuals to operate on the AWS cloud platform effectively.

AWS is built and managed by highly experienced engineers who offer their expertise to deliver the best products and solutions. That is why you can almost always find a function or service in AWS that would fulfill whatever need or requirement you have. A lot of the heavy lifting is offloaded from you as the customer so that you can dedicate your efforts and resources to your business operations. Another significant benefit of the AWS cloud is that it is extremely cost-effective and offers a way to expedite the launch of your products and services compared with the traditional methods. Resources can be quickly provisioned for a very low price and can be decommissioned quickly once you don't need them anymore. The cloud is an essential infrastructure piece in most companies, as you can quickly spin up or tear down test environments with just a push of a button. It can simplify deployment processes that are usually difficult and expensive to do in traditional data center setups.

The AWS Certified CloudOps Engineer Associate (SOA-C03) is a well-recognized certificate in the IT industry and is a major booster for career opportunities and salary increases. Having this certificate under your belt means that you indeed have the relevant knowledge and skills in deployment, management, and operations on AWS. Once you have gained more experience with AWS, you can also aim for higher-level certifications, such as the AWS Certified DevOps Engineer Professional certificate. The Professional and Specialty level certification exams in AWS are quite difficult and require extensive hands-on experience in order to ensure a pass. So if you are planning to pursue a career in Cloud DevOps, passing the AWS Certified CloudOps Engineer Associate is a great way to start the journey.

**Note:** We took extra care to come up with these concise articles and cheat sheets; however, this is meant to be just a supplementary resource when preparing for the exam. We highly recommend doing these [hands-on lab sessions](#), [video course](#), and [practice exams](#) to expand your knowledge further and improve your test-taking skills.



## AWS CERTIFIED CLOUDOPS ENGINEER ASSOCIATE EXAM OVERVIEW

In 2013, Amazon Web Services (AWS) began the Global Certification Program with the primary purpose of validating the technical skills and knowledge for building secure and reliable cloud-based applications using the AWS platform. By successfully passing the AWS exam, individuals can prove their AWS expertise to their current and future employers. The AWS Certified Solutions Architect - Associate exam was the first AWS certification that was launched, followed by two other role-based certifications: Systems Operations (SysOps) Administrator and Developer Associate later that year.

As of September 2025, the AWS Certified SysOps Administrator - Associate exam has been renamed and updated. The new version is AWS Certified CloudOps Engineer - Associate (SOA-C03), reflecting evolving industry terms and the broader scope of modern cloud operations. This new title applies exclusively to individuals who pass the updated SOA-C03 exam. Those who previously earned the AWS Certified SysOps Administrator - Associate will retain that original designation; the change will not be applied retroactively.

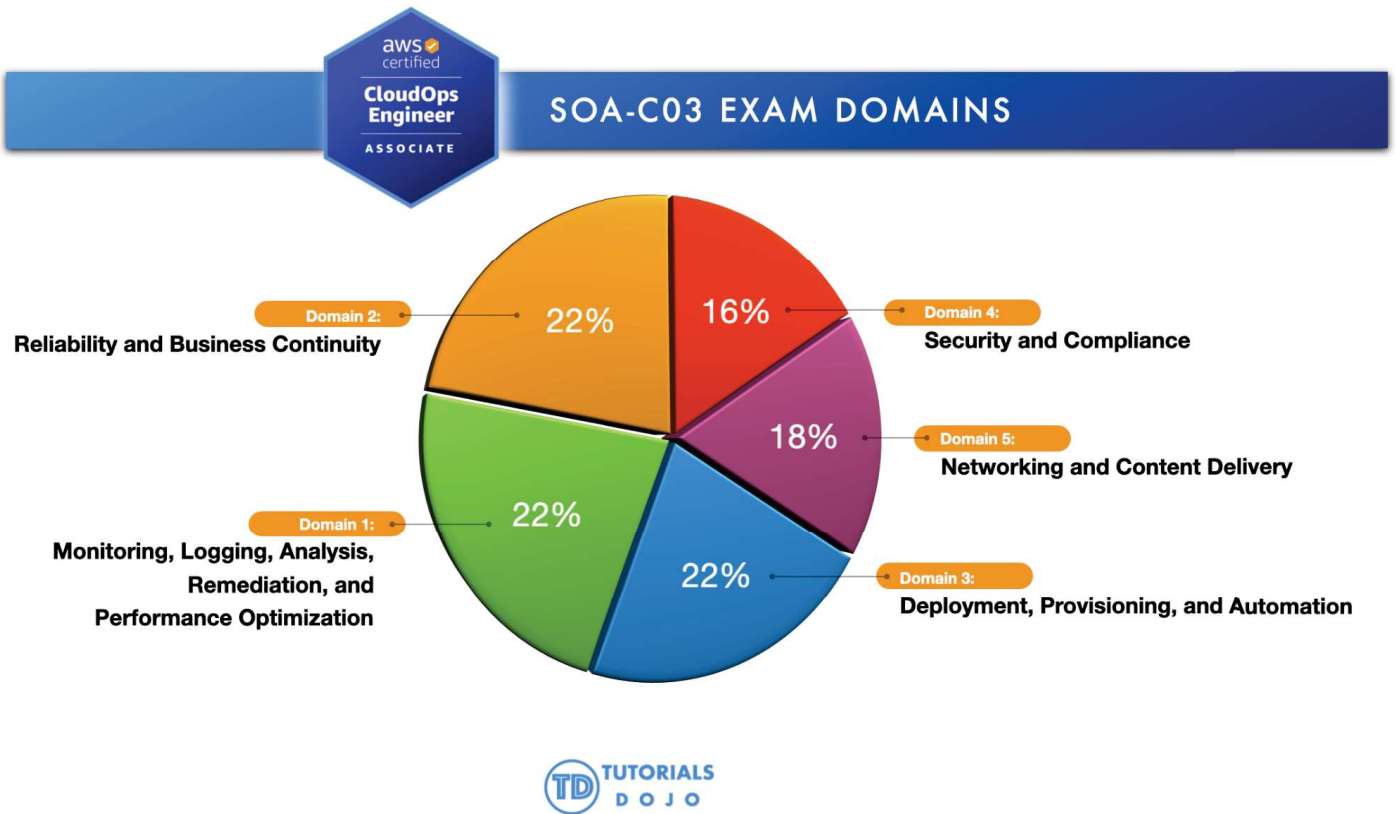
### Exam Details

The AWS Certified CloudOps Engineer - Associate examination is intended for CloudOps engineers with at least one year of experience in deployment, management, troubleshooting, networking, and security on AWS. The number of questions varies, depending on the exam.

|                          |   |
|--------------------------|---|
| <b>Exam Code:</b>        | SOA-C03                                     |
| <b>No. of Questions:</b> | 65  |
| <b>Score Range:</b>      | 100/1000                                    |
| <b>Passing Score:</b>    | 720/1000                                    |
| <b>Time Limit:</b>       | 130 minutes                                 |
| <b>Format:</b>           | Multiple choice/multiple response questions |
| <b>Delivery Method:</b>  | Testing center or online proctored exam     |

### Exam Domains

The AWS Certified CloudOps Engineer - Associate exam has five (5) different domains, each with corresponding weight and topic coverage. The domains are: Monitoring, Logging, Analysis, Remediation, and Performance Optimization (22%), Reliability and Business Continuity (22%), Deployment, Provisioning, and Automation (22%), Security and Compliance (16%), and Networking and Content Delivery (18%).



### Domain 1: Monitoring, Logging, Analysis, Remediation, and Performance Optimization (22%)

- 1.1 Implement metrics, alarms, and filters by using AWS monitoring and logging services
- 1.2 Identify and remediate issues by using monitoring and availability metrics
- 1.3 Implement performance optimization strategies for compute, storage, and database resources

### Domain 2: Reliability and Business Continuity (22%)

- 2.1 Implement scalability and elasticity
- 2.2 Implement highly available and resilient environments
- 2.3 Implement backup and restore strategies

### Domain 3: Deployment, Provisioning, and Automation (22%)

- 3.1 Provision and maintain cloud resources
- 3.2 Automate the management of existing resources





#### **Domain 4: Security and Compliance (16%)**

- 4.1 Implement and manage security and compliance tools and policies
- 4.2 Implement strategies to protect data and infrastructure

#### **Domain 5: Networking and Content Delivery (18%)**

- 5.1 Implement and optimize networking features and connectivity
- 5.2 Configure domains, DNS services, and content delivery
- 5.3 Troubleshoot network connectivity issues

### **Exam Scoring**

You can get a score from 100 to 1,000 with a minimum passing score of **720** when you take the AWS Certified CloudOps Engineer Associate exam. AWS uses a scaled scoring model to associate scores across multiple exam types that may have different levels of difficulty. Your complete score report will be sent to you by email about 3 to 5 business days after your exam.

For individuals who unfortunately did not pass their exams, they must wait for 14 days before they are allowed to retake the exam. There is no hard limit on the number of attempts you can retake an exam. Once you pass, you'll receive various benefits, such as a 50% discount coupon, which you can use for your next AWS exam.

Once you receive your score report via email, the result should also be saved in your AWS Certification account already. The score report contains a table of your performance on each domain, and it will indicate whether you have met the level of competency required for these domains.

Take note that you do not need to achieve competency in all domains for you to pass the exam. At the end of the report, there will be a score performance table that highlights your strengths and weaknesses, which will help you determine the areas you need to improve on.



## Exam Benefits

If you have successfully passed any AWS exam, you will be eligible for the following benefits:

- **Exam Discount** - You'll get a 50% discount voucher that you can apply for your recertification or any other exam you plan to pursue. To access your discount voucher code, go to the "Benefits" section of your AWS Certification account and apply the voucher when you register for your next exam.
- **Certification Digital Badges** - You can showcase your achievements to your colleagues and employers with digital badges on your email signatures, LinkedIn profile, or social media accounts. You can also show your Digital Badge to gain exclusive access to Certification Lounges at AWS re:Invent, regional Appreciation Receptions, and select AWS Summit events. To view your badges, simply go to the "Digital Badges" section of your AWS Certification Account.
- **Event Recognition** - You can receive invitations to regional Appreciation Receptions and use your digital badge for exclusive access to AWS Certification Lounges at AWS re:Invent and select AWS Summit events.

You can visit the official AWS Certification FAQ page to view the frequently asked questions about getting AWS Certified and other information about the AWS Certification: <https://aws.amazon.com/certification/faqs/>.



## AWS CERTIFIED CLOUDOPS ENGINEER ASSOCIATE EXAM - STUDY GUIDE AND TIPS

If you are a Systems Administrator or a DevOps Engineer, then this certification will test your knowledge on various knowledge areas in the AWS Cloud platform. Your experience will come in handy in passing the exam, but this should be complemented by actual experience in working with your cloud workloads. Doing several hands-on labs is quite helpful, too, as it will give you that much-needed experience and insight into how the different AWS services work.

The AWS Certified CloudOps Engineer Associate exam will verify your ability to perform the following:

- Deploy, manage, and operate workloads on AWS
- Support and maintain AWS workloads according to the AWS Well-Architected Framework
- Perform operations by using the AWS Management Console and the AWS CLI
- Implement security controls to meet compliance requirements
- Monitor, log, and troubleshoot systems
- Apply networking concepts (for example, DNS, TCP/IP, firewalls)
- Implement architectural requirements (for example, high availability, performance, capacity)
- Perform business continuity and disaster recovery procedures
- Identify, classify, and remediate incidents

You should also be knowledgeable about the AWS Well-Architected Framework, AWS Deployment Options. Having prior knowledge of fundamental networking and security will also be very valuable. This guide aims to provide you with a straightforward guide when reviewing for this exam.

### Study Materials

There are many study resources that you can use to prepare for the AWS Certified CloudOps Engineer - Associate (SOA-C03) exam; however, your source of truth should always be the [official exam guide](#). You should also know the latest exam code of the CloudOps exam, which is SOA-C03, so you can ensure that the study materials you are using is not for the previous exam version (e.g. SOA-C02). Use the official exam guide to know the relevant exam topics of the test and verify if the resources you are using adequately cover the required knowledge areas.

- [AWS Certified SysOps Administrator - Associate video course](#)
- [AWS Certified CloudOps Engineer - Associate SOA-C03 Practice Exams](#)
- [AWS Cheat Sheets](#)
- [AWS Hands-On Labs](#)



---

## Recommended Whitepapers

The whitepapers listed below are arranged in such a way that you will learn the concepts first, before proceeding to application and best practices. If you need a refresh on your AWS fundamentals, go check out our guide on the [AWS Certified Cloud Practitioner Exam](#) before proceeding below.

1. [Amazon Virtual Private Cloud Connectivity Options](#) - Study how you can connect different VPCs together, your VPCs to your on-premises network, and vice versa.
2. [How AWS Pricing Works](#) - Study the fundamental drivers of cost in AWS, the pricing models of commonly used services in compute, storage, and database, and how to optimize your costs.
3. [AWS Well-Architected Framework](#) - This whitepaper is one of the most important papers that you should study for the SOA-C03 exam. It discusses the different pillars that make up a well-architected cloud environment.
4. [Overview of Deployment Options on AWS](#) - This is an optional whitepaper that you can read to be aware of your deployment options in AWS. There is a chance that this might come up in the exam.
5. [AWS Disaster Recovery Plans](#) - As a CloudOps Engineer, you should be familiar with your DR options when outages occur. Having knowledge of DR will determine how fast you can recover your infrastructure.

## AWS Services to Focus On

AWS offers extensive documentation and well-written FAQs for all of their services. These two will be your primary source of information when studying. Furthermore, as an AWS CloudOps Engineer, you need to be well-versed in a number of AWS products and services since you will almost always be using them in your work. I recommend checking out [Tutorials Dojo's AWS Cheat Sheets](#) which provides a summarized but highly informative set of notes and tips for your review of these services.

Core AWS Services to focus on:

1. [EC2](#) - As the most fundamental computing service offered by AWS, you should know about EC2 inside out.
2. [Elastic Load Balancer](#) - Load balancing is very important for a highly available system. Study the different types of ELBs and the features each of them supports.
3. [Auto Scaling](#) - Study what services in AWS can be auto-scaled, what triggers scaling, and how auto-scaling increases/decreases the number of instances.
4. [Elastic Block Store](#) - As the primary storage solution of EC2, study the types of EBS volumes available. Also study how to secure, backup, and restore EBS volumes.
5. [S3 / Glacier](#) - Study what the S3 storage types are and the differences between them. Also review the capabilities of S3, such as hosting a static website, securing access to objects using policies, lifecycle policies, etc. Learn as much about S3 as you can.



6. VPC - Study every service that is used to create a VPC (subnets, route tables, internet gateways, nat gateways, VPN gateways, etc). Also, review the differences between network access control lists and security groups and during which situations they are applied.
7. Route 53 - Study the different types of records in Route 53 and the various routing policies. Know what hosted zones and domains are.
8. RDS - Know how each RDS database differs from one another, and how they are different from Aurora. Determine what makes Aurora unique, and when it should be preferred from other databases (in terms of function, speed, cost, etc). Learn about parameter groups, option groups, and subnet groups.
9. DynamoDB - Consider how DynamoDB compares to RDS, ElastiCache, and Redshift. This service is also commonly used for serverless applications along with Lambda.
10. ElastiCache - Familiarize yourself with ElastiCache Redis and its functions. Determine the areas/services where you can place a caching mechanism to improve data throughput, such as managing the session state of an ELB, optimizing RDS instances, etc.
11. SQS - Gather info on why SQS is helpful in decoupling systems. Study how messages in the queues are being managed (standard queues, FIFO queues, dead letter queues). Know the differences between SQS, SNS, SES, and Amazon MQ.
12. SNS - Study the function of SNS and what services can be integrated with it. Also, be familiar with the supported recipients of SNS notifications.
13. IAM - Services such as IAM Users, Groups, Policies, and Roles are the most important to learn. Study how IAM integrates with other services and how it secures your application through different policies. Also, read on the best practices when using IAM.
14. CloudWatch - Study how monitoring is done in AWS and what types of metrics are sent to CloudWatch. Also read upon CloudWatch Logs, CloudWatch Alarms, and the custom metrics made available with CloudWatch Agent.
15. CloudTrail - Familiarize yourself with how CloudTrail works, and what kinds of logs it stores as compared to CloudWatch Logs.
16. Config - Be familiar with the situations where AWS Config is useful.
17. CloudFormation - Study how CloudFormation is used to automate infrastructure deployment. Learn the basic makeup of a CloudFormation template, stack, and stack set.
18. KMS - Familiarize how KMS integrates with other services in storing encryption keys.
19. Secrets Manager - Understand how Secrets Manager stores secrets and how you can use them with other AWS services.
20. Parameter Store - Know when to use Parameter store and how compute services like EC2, ECS, and Lambda utilize it.
21. DataSync - Familiarize yourself with which AWS services can be used to migrate data from an on-premises data center.

**Some additional services we recommend to review:**

1. Trusted Advisor
2. Systems Manager



3. [CodeDeploy](#)
4. [CodePipeline](#)
5. [CloudFront](#)
6. [Cost and Billing Management Console](#)
7. [Direct Connect](#)
8. Amazon FSx for Windows File Server and Amazon FSx for Lustre
9. AWS Backup
10. EC2 Image Builder
11. S3 Transfer Acceleration
12. AWS Global Accelerator
13. RDS Proxy
14. IAM Access Analyzer

## Common Exam Scenarios

| Scenario  | Solution  |
|---|---|
| <b>Monitoring, Logging, Analysis, Remediation, and Performance Optimization</b>                                       |   |
| You need to set up an alert that notifies the IT manager about EC2 instances service limits.                          | Use Amazon Eventbridge to detect and react to changes in the status of Trusted Advisor checks |
| You need to track the deletion and rotation of KMS keys.  | Use AWS CloudTrail to log AWS KMS API calls   |
| You need to investigate if the traffic is reaching the EC2 instance.  | Use VPC flow logs   |
| You need to ensure that the SSH protocol is always disabled on private servers.                                       | Use AWS Config Rules  |
| You need to retrieve the instance metadata of an EC2 instance.  | <a href="http://169.254.169.254/latest/">http://169.254.169.254/latest/</a>                   |
| You have to monitor the CPU usage of a single process in your EC2 instance.   | Use the CloudWatch Agent procstat plugin to monitor system utilization.                       |
| You need to generate a report on the replication and encryption status of all of the objects stored in the S3 bucket. | Use S3 Inventory  |
| Metric to use to alarm when all instances behind an ALB becomes unhealthy   | AWS/ApplicationELB HealthyHostCount<br>≤ 0  |



|   |  |
|---|--|
| Monitor restricted CIDR changes on a security group and remove them automatically.  | Use AWS Config to evaluate the security group and AWS Systems Manager Automation document to remove the unwanted CIDR range.   |
| Monitor CreateUser API call via email   | Utilize Amazon EventBridge, declare CloudTrail as a source, and CreateUser as an event pattern. Create an SNS topic and set it as an event target on Amazon EventBridge. |
| You have to automate the process of patching managed instances with security-related updates.   | Use AWS Systems Manager Patch Manager  |
| You need to analyze the data hosted in Amazon S3 using standard SQL.  | Use Amazon Athena  |
| Improving the site speed of a static S3 web hosting with customers around the globe   | Create a CloudFront web distribution and set Amazon S3 as the origin.  |
| You need to implement a solution to enforce the tagging of all instances that will be launched in the VPC.                              | Use AWS Service Catalog TagOption library  |
| You need to get billing alerts once it reaches a certain limit.   | Enable billing alerts in Account Preferences of the AWS Console.   |
| Resize an Amazon ElastiCache for Redis cluster.   | Use online resizing for Amazon ElastiCache Redis cluster   |
| No sharing of Reserved Instance (RI) discounts between AWS accounts in the Organization   | Disable RI discount sharing via management account and provision instances using individual AWS accounts.  |
| <b>Reliability and Business Continuity</b>  |  |
| When the incoming message traffic increases, the EC2 instances fall behind and it takes too long to process the messages.               | Create an Auto Scaling group that can scale out based on the number of messages in the queue.  |
| You need to log the client's IP address, latencies, request paths, and server responses that go through your Application Load Balancer. | Enable access logging in ALB and store the logs on an S3 bucket.   |
| You need to determine which cipher is used for the SSL connection in your ELB.  | Enable Server Order Preference   |
| You need to monitor the total number of requests or connections in your load balancer.  | Monitor the <code>SurgeQueueLength</code> metric   |





|   |   |
|---|---|
| You need to ensure that the backups of an Amazon Redshift cluster are always available.   | Configure the Amazon Redshift cluster to automatically copy snapshots of a cluster to another region.   |
| Highly available File Server that supports SMB and manages file permissions using Windows Access Control List (A).  | Multi-AZ Amazon FSx for Windows File Server   |
| Slow load time when uploading objects to S3   | S3 Transfer Acceleration  |
| PercentIOLimit metric hits 100% on EFS  | Create a new Max I/O performance mode EFS file system and migrate data to the new file system using AWS DataSync.   |
| Must ensure data integrity when performing EBS backups  | Build a Lambda function that uses CreateImage API to generate AMI of the EC2 instance and include a reboot parameter. Create an Amazon EventBridge rule to execute the Lambda function daily. |
| <b>Deployment, Provisioning, and Automation</b>   |   |
| You must remotely execute shell scripts and securely manage the configuration of EC2 instances.   | Use Systems Manager Run Command   |
| You need to identify the configuration changes in the CloudFormation resources.   | Use drift detection   |
| Requires a CloudFormation template that can be reused for multiple environments. If the template has been updated, all the stack that is referencing it will automatically use the updated configuration. | Use Nested Stacks   |
| You need to automate the process of updating the CloudFormation templates to map to the latest AMI IDs.   | Use CloudFormation with Systems Manager Parameter Store   |
| The eviction count in Amazon ElastiCache for Memcached has exceeded its threshold.  | Scale the cluster by increasing the number of nodes.  |
| You need to provide each department a new AWS account with governance guardrails and a defined baseline in place.   | Set up AWS Control Tower  |



|  |   |
|--|---|
| An S3 bucket must be configured to move the objects older than 60 days to the Infrequent Access storage class.   | Set up a lifecycle policy   |
| You need to monitor all the COPY and UNLOAD traffic in the Redshift cluster.   | Enable Enhanced VPC routing on the Redshift cluster.  |
| TLS certificate should be renewed automatically  | Request a public certificate via AWS Certificate Manager (ACM)                                    |
| Get cost expenses of each AWS user account   | Enable the createdBy tag in the Billing and Management console                                    |
| Provisioning instances on ASG takes time because of software dependencies installed via the UserData script.   | EC2 Image Builder   |
| Get cost expenses of each AWS user account   | Enable the createdBy tag in the Billing and Management console                                    |
| <b>Security and Compliance</b>   |   |
| You have to rotate an existing KMS key with imported key material every 6 months   | Create a new KMS key with imported key material and update the key ID to point to the new KMS key |
| A company needs to restrict access to the data in an S3 bucket.  | Use S3 ACL and bucket policy  |
| Mitigate malicious attacks such as SQL injection and DDoS attacks from unknown origins.  | Use AWS WAF and Shield  |
| You need to define an IAM policy to enable the user to pass a role to an AWS service.  | Define iam:PassRole in the IAM policy   |
| You need to create a solution that allows multiple EC2 instances in a private subnet to use AWS KMS and the traffic must not pass through the public Internet. | Configure a VPC endpoint  |
| You need to encrypt all the objects at rest in your S3 bucket  | Use S3-S3, AWS KMS keys (SSE-KMS) or SSE-C  |
| Enable authentication to AWS services using Active Directory Federation Services.  | Amazon Cognito user pool  |



|  |   |
|--|---|
| Create a bucket policy to only allow AWS accounts in the organization to access an S3 bucket.  | Set principal to (*) and create a condition for PrincipalOrgId  |
| Read, update, delete messages from SQS queues from an instance.  | Create a policy with sqs:SendMessage, sqs:ReceiveMessage, sqs:DeleteMessage, and attach the policy to a new role that can perform API calls to AWS. Associate the new role to the instance. |
| RDS credentials should not be hardcoded on Lambda functions  | Use Secrets Manager to store credentials  |
| Networking and Content Delivery  |   |
| You need to allow the EC2 instances in your VPC that support IPv6 to connect to the Internet but block any incoming connection.          | Set up an egress-only Internet gateway  |
| You have to establish a dedicated connection between their on-premises network and their Amazon VPC.                                     | Set up a Direct Connect connection  |
| You need to increase the cache hit ratio for a CloudFront web distribution.  | Add a Cache-Control max-age and increase the TTL by specifying the longest value for max-age  |
| You need to ensure that users are consistently directed to the AWS region nearest to them.   | Set up a Route 53 Geoproximity routing policy   |
| A company plans to implement a hybrid cloud architecture. You need to allow your resources on AWS the connectivity to external networks. | Assign an Internet Gateway to the VPC<br>Create a Virtual Private Gateway   |
| Users being served desktop version on mobile phones  | Add a User-Agent header to the list of origin custom header on CloudFront.  |
| DNS record at the apex domain  | ALIAS record  |

## Validate Your Knowledge

Once you have finished your review and you are more than confident of your knowledge, test yourself with some practice exams available online. AWS offers a practice exam that you can try out at their [aws.training](https://aws.training) portal. [Tutorials Dojo](#) also offers a top-notch set of [AWS Certified CloudOps Engineer Associate practice tests](#). Each test contains unique questions that will surely help verify if you have missed out on anything important that might appear on your exam. You can pair our practice exams with this study guide eBook to further help in your exam preparations.



### Sample Practice Test Questions:

#### Question 1

A company is heavily using AWS CloudFormation templates to automate the deployment of their cloud resources. The CloudOps Engineer needs to write a template that will automatically copy objects from an existing S3 bucket into the new one.

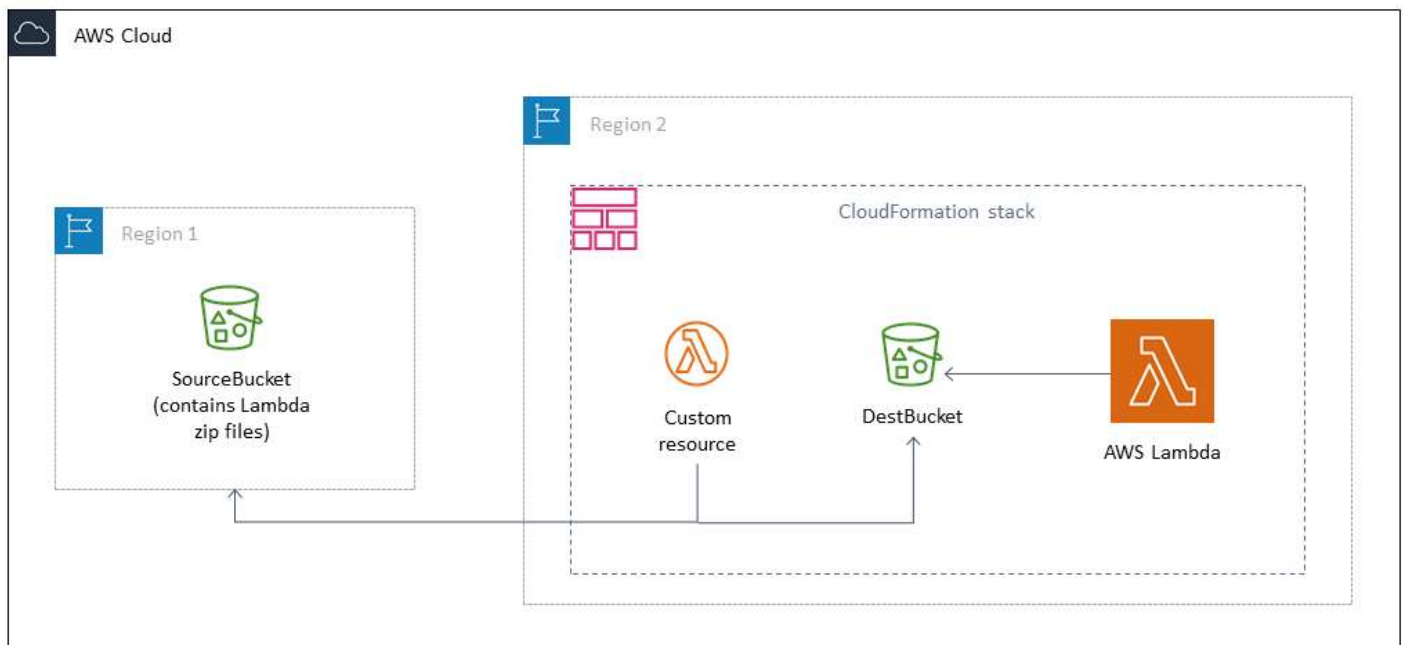
Which of the following is the most suitable configuration for this scenario?



1. Set up an AWS Lambda function and configure it to perform the copy operation. Integrate the Lambda function to the CloudFormation template as a custom resource.
2. Configure the CloudFormation template to modify the existing S3 bucket to allow cross-origin requests.
3. Configure the CloudFormation template to set up an AWS Step Functions state machine to orchestrate the copy process from the existing S3 bucket to the new one.
4. Configure the CloudFormation template to enable cross-region replication on the existing S3 bucket and select the new S3 bucket as the destination.

**Correct Answer: 1**

**AWS CloudFormation** gives you an easy way to model a collection of related AWS and third-party resources, provision them quickly and consistently, and manage them throughout their lifecycles, by treating infrastructure as code. A CloudFormation template describes your desired resources and their dependencies so you can launch and configure them together as a stack. You can use a template to create, update, and delete an entire stack as a single unit, as often as you need to, instead of managing resources individually. You can manage and provision stacks across multiple AWS accounts and AWS Regions.



In an AWS CloudFormation template, you can specify a Lambda function as the target of a custom resource. Use custom resources to process parameters, retrieve configuration values, or call other AWS services during stack lifecycle events. When you associate a Lambda function with a custom resource, the function is invoked whenever the custom resource is created, updated, or deleted. AWS CloudFormation calls a Lambda API to invoke the function and to pass all the request data (such as the request type and resource properties) to the function. The power and customizability of Lambda functions in combination with AWS CloudFormation



enable a wide range of scenarios, such as dynamically looking up AMI IDs during stack creation, or implementing and using utility functions, such as string reversal functions.

The requirement for this scenario is to copy all the objects from an existing S3 bucket to a new S3 bucket created by the CloudFormation template. To accomplish this requirement, you need to create a custom Lambda function that can copy the objects from the source bucket to the new S3 bucket. You can also define the options you want Amazon S3 to apply during replication, such as server-side encryption, replica ownership, and transitioning replicas to another storage class.

Hence, the correct answer is: **Set up an AWS Lambda function and configure it to perform the copy operation. Integrate the Lambda function to the Cloudformation template as a custom resource.**

The option that says: **Configure the Cloudformation template to enable cross-region replication on the existing S3 bucket and select the new S3 bucket as the destination** is incorrect because this option won't be able to copy the existing objects to the new S3 bucket. For this configuration, you need to invoke Lambda first to copy the objects in the S3 bucket.

The option that says: **Configure the CloudFormation template to set up an AWS Step Functions state machine to orchestrate the copy process from the existing S3 bucket to the new one** is incorrect because AWS Step Function is primarily designed for orchestrating complex workflows. Using it for a simple file copy operation adds unnecessary complexity and overhead.

The option that says: **Configure the CloudFormation template to modify the existing S3 bucket to allow cross-origin requests** is incorrect because the scenario did not state anything about allowing cross-origin access to your Amazon S3 resources. Also, this option does not have the capability to copy all the objects from an existing S3 bucket to a new S3 bucket.

#### References:

<https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/template-custom-resources-lambda.html>

<https://aws.amazon.com/blogs/infrastructure-and-automation/deploying-aws-lambda-functions-using-aws-cloudformation-the-portable-way/>

<https://aws.amazon.com/blogs/devops/use-aws-cloudformation-to-automate-the-creation-of-an-s3-bucket-with-cross-region-replication-enabled/>

Check out this AWS CloudFormation Cheat Sheet:

<https://tutorialsdojo.com/aws-cloudformation/>



## Question 2

An eCommerce company has a suite of microservices-based retail applications on a Kubernetes cluster using Amazon Elastic Kubernetes Service (Amazon EKS) in AWS.

The application suite has been running for a few months when the DevOps team notices a surge of application traffic whenever there's a scheduled promotion event. To avoid loss of revenue, the CloudOps Engineer must ensure an uninterrupted service and preempt any potential degradation of the service in the production environment.

Which combination of actions can provide the MOST operationally efficient solution that can meet the above requirement? (Select TWO.)

1. Enable the built-in Kubernetes Horizontal Pod Autoscaler option in Amazon EKS.
2. Set up a Kubernetes Metrics Server first in the Amazon EKS cluster and enable the Kubernetes Vertical Pod Autoscaler.
3. Integrate both Amazon CloudWatch Contributor Insights and Amazon CloudWatch Application Insights in the EKS cluster.
4. Install the Kubernetes Metrics Server in the Amazon EKS cluster and use the Kubernetes Horizontal Pod Autoscaler.
5. Configure the kubectl client to communicate with the Amazon EKS cluster.

## Correct Answers: 4 & 5

Take note that the Kubernetes Horizontal Pod Autoscaler feature automatically scales the number of Pods in a deployment, replication controller, or replica set based on that resource's CPU utilization. The Horizontal Pod Autoscaler can help your applications scale out automatically to meet increased demand or scale in when resources are not needed, thus freeing up your nodes for other applications. When you set a target CPU utilization percentage, the Horizontal Pod Autoscaler scales your application in or out to try to meet that target.



```
apiVersion: autoscaling/v2
kind: HorizontalPodAutoscaler
metadata:
  name: tutorialsdojo-nginx
spec:
  scaleTargetRef:
    apiVersion: apps/v1
    kind: Deployment
    name: nginx
  minReplicas: 1
  maxReplicas: 10
  metrics:
  - type: Resource
    resource:
      name: cpu
      target:
        type: Utilization
        averageUtilization: 50
  - type: Resource
    resource:
      name: memory
      target:
        type: AverageValue
        averageValue: 100Mi
```

Horizontal Pod Autoscaler

Resource metrics for tracking

Hence, the correct answers are:

- Install the **Kubernetes Metrics Server** in the Amazon EKS cluster and use the **Kubernetes Horizontal Pod Autoscaler**.
- Configure the **kubectl** client to communicate with the Amazon EKS cluster.

The option that says: **Set up a Kubernetes Metrics Server first in the Amazon EKS cluster and enable the Kubernetes Vertical Pod Autoscaler** is incorrect. While setting up a Kubernetes Metrics Server is right, the use of a Kubernetes Vertical Pod Autoscaler (VPA) is not required at all. VPA is useful for optimizing resource utilization within individual pods as it involves more fine-tuning and requires a deeper understanding of the application's resource requirements. In contrast, Horizontal Pod Autoscaler (HPA) is more geared toward



handling changes in demand by adjusting the number of pod replicas, making it a simpler and more automated solution for handling increased application traffic.

The option that says: **Enable the built-in Kubernetes Horizontal Pod Autoscaler option in Amazon EKS** is incorrect because there are no built-in Horizontal Pod Autoscaler in Amazon EKS. You have to manually install the Kubernetes Metrics Server and the Horizontal Pod Autoscaler programs first in your Amazon EKS cluster.

The option that says: **Integrate both Amazon CloudWatch Contributor Insights and Amazon CloudWatch Application Insights in the EKS cluster** is incorrect. Even though CloudWatch Contributor Insights and Application Insights are great for monitoring, this combination is not capable of scaling the Amazon EKS cluster. You need to use a Kubernetes Horizontal Pod Autoscaler for this case.

#### References:

<https://docs.aws.amazon.com/eks/latest/userguide/horizontal-pod-autoscaler.html>

<https://kubernetes.io/docs/tasks/run-application/horizontal-pod-autoscale/#how-does-a-horizontal-pod-autoscaler-work>

<https://docs.aws.amazon.com/eks/latest/userguide/eks-workloads.html>

#### Check out this Amazon Elastic Kubernetes Service Cheat Sheet:

<https://tutorialsdodo.com/amazon-elastic-kubernetes-service-eks/>

Click [here](#) for more **AWS Certified CloudOps Engineer - Associate practice exam questions**.

It is best to get some rest before the day of your exam, and review any notes that you have written down. If you have done well in the **practice tests**, go over the questions where you made a mistake and understand why so. If you are not feeling so confident after trying the practice tests, you can just reschedule your exam and take your time preparing. The exam will not be easy to pass, but it'll be worth it when you do.

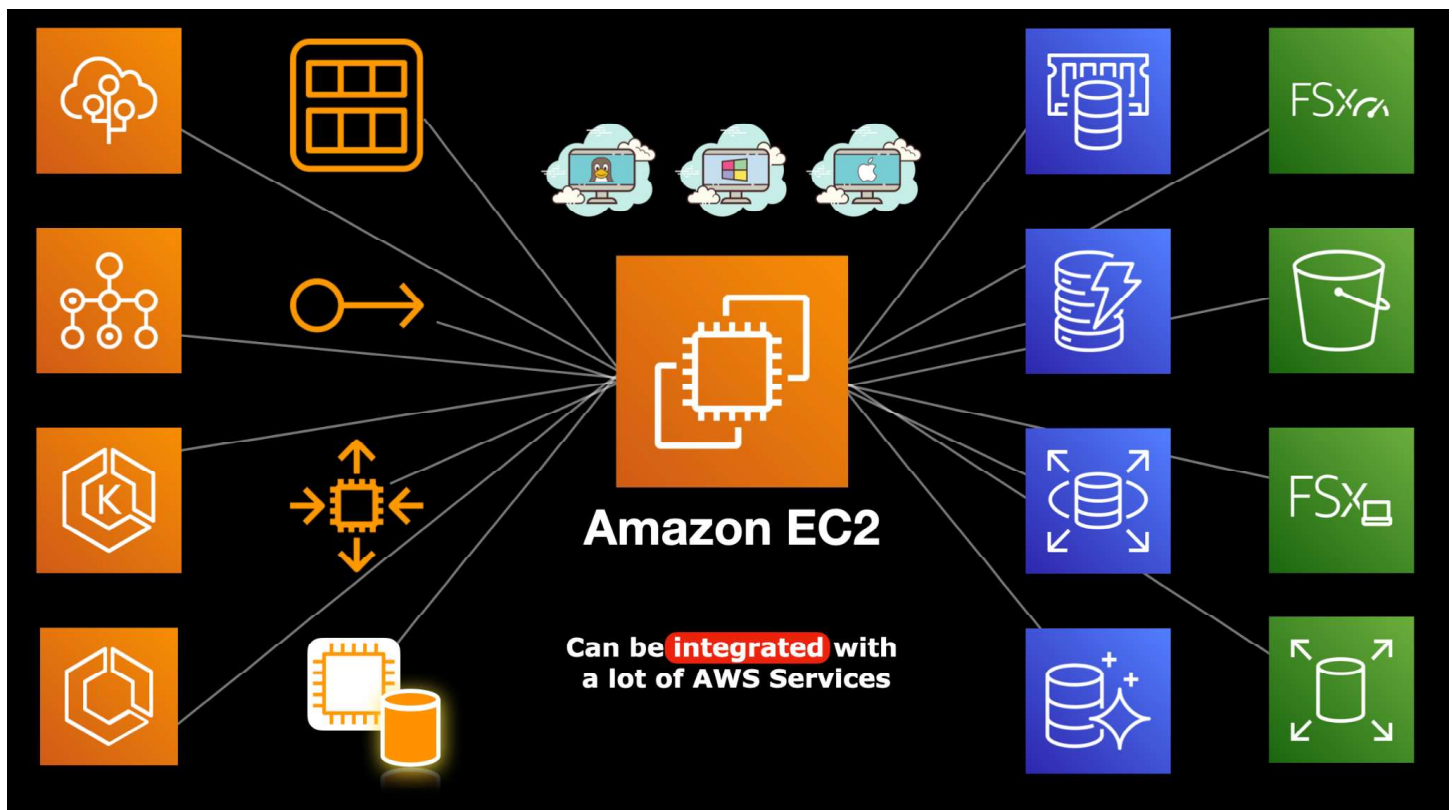
## AWS Deep Dives

### Amazon EC2

Amazon Elastic Compute Cloud (EC2) is a computing service that runs virtual servers in the cloud. It allows you to launch Linux or Windows virtual machines to host your applications and manage them remotely – wherever you are in the globe.

You and AWS have a shared responsibility in managing your Amazon EC2 virtual machines. AWS manages the data centers, physical facilities, the hardware components, the host operating system, and the virtualization layer that powers the entire Amazon EC2 service. On the other hand, you are responsible for your guest operating system, applying OS patches, setting up security access controls, and managing your data.

Amazon EC2 can be integrated into other AWS services to accomplish a certain task or to meet your specifications. It can be used to do a variety of functions – from running applications, hosting a self-managed database, processing batch jobs and so much more.





An Amazon EC2 virtual machine is somewhat similar to your desktop or laptop that you may be using right now. It also has a CPU, a Random Access Memory, a Network Interface, an IP address, and even a system image backup. You can also attach a Solid State Drive, or a Hard Disk Drive (HDD) to your EC2 instance for more storage; You can even connect it to a shared network file system, to allow multiple computers to access the same files.

Just like your computer, you can also integrate a lot of other AWS services with Amazon EC2. You can attach various storage, networking, and security services to an Amazon EC2 instance. There are many options available to purchase your EC2 instance, that can help you lower down your operating costs. Some AWS Services are even using Amazon EC2 as its underlying compute component. These services orchestrate or control a group of EC2 instances to perform a specific function, such as scaling or batch processing. It is also used on AWS-managed databases, containers, serverless computing engines, microservices, and many more! This is why Amazon EC2 is considered as the basic building block in AWS – it is used in almost every service!

For storage, you can use different AWS Storage services with your Amazon EC2 instance to store and process data. You can attach an Instance store for your temporary data or an Amazon EBS volume for persistent storage.

You can also mount a file system to your EC2 instances. You can connect it to Amazon EFS or Amazon FSx. For your static media files or object data, you can store them in Amazon S3 and then retrieve them back to your EC2 instance via an API or through an HTTP and FTP client.

For networking, you launch your EC2 instance on either a public or a private subnet in a Virtual Private Cloud or VPC. You can associate an Elastic IP address to your instance for it to have a static IPv4 address. An elastic network interface can also be used as a virtual network card for your EC2 instance. If you have a group of interdependent instances, you can organize them into a placement group. This placement group can be a cluster, a spread, or a partition type that enables you to minimize correlated failures, lower network latency, and achieve high throughput.

AWS also offers enhanced networking features to provide high-performance networking capabilities by using an Elastic Network Adapter or an Intel 82599 Virtual Function (VF) interface. If you have a High-Performance Computing workload or machine learning applications, you can attach an Elastic Fabric Adapter to your instance to provide a higher network throughput than your regular TCP transport.

For scaling, you can use Amazon EC2 Auto Scaling to automatically add more EC2 instances to process the increasing number of traffic in your application. Auto Scaling can also terminate the underutilized instances if the demand decreases – this can cut down your server expenses in half, or even more!

For system image backup, you can take a snapshot of your EC2 instance by creating an Amazon Machine Image, or AMI.



The AMI is just like a disk image of your Mac, Linux, or Windows computer that contains custom data and system configurations that you have set. It enables you to launch a pre-configured Amazon EC2 instance that can be used for auto-scaling, migration, and backups. If your EC2 instance crashes, you can easily restore your data using an AMI. It is also helpful if you want to move your server to another Available Zone, another Region, or even another AWS account. You can also launch one or more EC2 instances using a single AMI.

There are more AWS services and features that you can integrate with Amazon EC2. We will cover these services in the succeeding chapters of this eBook.

## Components of an EC2 Instance

You must know the components of an EC2 instance since this is one of the core AWS services that you'll be encountering the most in the exam.

- 1) When creating an EC2 instance, you always start off by choosing a **base AMI or Amazon Machine Image**. An AMI contains the OS, settings, and other applications that you will use in your server. AWS has many pre-built AMIs for you to choose from, and there are also custom AMIs created by other users which are sold on the AWS Marketplace for you to use. If you have created your own AMI before, it will also be available for you to select. AMIs cannot be modified after launch.
- 2) After you have chosen your AMI, you select the **instance type and size** of your EC2 instance. The type and size will determine the physical properties of your instance, such as CPU, RAM, network speed, and more. There are many instance types and sizes to choose from and the selection will depend on your workload for the instance. You can freely modify your instance type even after you've launched your instance, which is commonly known as "right-sizing".
- 3) Once you have chosen your AMI and your hardware, you can now configure your instance settings.
  - a) If you are working on the console, the first thing you'll indicate is the **number of instances** you'd like to launch with these specifications you made.
  - b) You specify whether you'd like to launch **spot instances** or use another instance billing type (on-demand or reserved).
  - c) You configure which **VPC and subnet** the instance should be launched in, and whether it should receive a **public IP address** or not.
  - d) You choose whether to include the instance in a **placement group** or not.
  - e) You indicate if the instance will be joined to one of your **domains/directories**.
  - f) Next is the **IAM role** that you'd like to provide to your EC2 instance. The IAM role will provide the instance with permission to interact with other AWS resources indicated in its permission policy.
  - g) **Shutdown behavior** lets you specify if the instance should only be stopped or should be terminated once the instance goes into a stopped state. If the instance supports **hibernation**, you can also enable the hibernation feature.
  - h) You can enable the **termination protection** feature to protect your instance from accidental termination.





- i) If you have **EFS file systems** that you'd like to immediately mount to your EC2 instance, you can specify them during **launch**.
  - j) Lastly, you can specify if you have commands you'd like your EC2 instance to execute once it has launched. These commands are written in the **user data** section and submitted to the system.
- 4) After you have configured your instance settings, you now need to add **storage** to your EC2 instance. A volume is automatically created for you since this volume will contain the OS and other applications of your AMI. You can add more storage as needed and specify the type and size of EBS storage you'd like to allocate. Other settings include specifying which EBS volumes are to be included for termination when the EC2 instance is terminated, as well as encryption.
  - 5) When you have allocated the necessary storage for your instances, next is adding **tags** for easier identification and classification.
  - 6) After adding in the tags, you now create or add **security groups** to your EC2 instance, which will serve as firewalls to your servers. Security groups will moderate the inbound and outbound traffic permissions of your EC2 instance. You can also add, remove, and modify your security group settings later on.
  - 7) Lastly, access to the EC2 instance will need to be secured using one of your **key pairs**. Make sure that you have a copy of this key pair so that you'll be able to connect to your instance when it is launched. There is no way to reassociate another key pair once you've launched the instance. You can also proceed without selecting a key pair, but then you would have no way of directly accessing your instance unless you have enabled some other login method in the AMI or via Systems Manager.
  - 8) Once you are happy with your instance, proceed with the launch. Wait for your EC2 instance to finish preparing itself, and you should be able to connect to it if there aren't any issues.

#### References:

[https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EC2\\_GetStarted.html](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EC2_GetStarted.html)

<https://tutorialsdjo.com/amazon-elastic-compute-cloud-amazon-ec2/>

#### Types of EC2 Instances

1. **General Purpose** — Provides a balance of computing, memory, and networking resources and can be used for a variety of diverse workloads. Instances under the T-family have burstable performance capabilities to provide higher CPU performance when the CPU is under high load in exchange for CPU credits. Once the credits run out, your instance will not be able to burst anymore. More credits can be earned at a certain rate per hour, depending on the instance size.
2. **Compute Optimized** — Ideal for compute-bound applications that benefit from high-performance processors. Instances belonging to this family are well suited for batch processing workloads, media transcoding, high-performance web servers, high-performance computing, scientific modeling, dedicated gaming servers and ad server engines, machine learning inference, and other compute-intensive applications.
3. **Memory Optimized** — Designed to deliver fast performance for workloads that process large data sets in memory.



4. **Accelerated Computing** — Uses hardware accelerators or co-processors to perform functions such as floating point number calculations, graphics processing, or data pattern matching more efficiently than on CPUs.
5. **Storage Optimized** — Designed for workloads that require high, sequential read and write access to very large data sets on local storage. They are optimized to deliver tens of thousands of low-latency, random I/O operations per second (IOPS) to applications.
6. **Nitro-based** — The Nitro System provides bare metal capabilities that eliminate virtualization overhead and support workloads that require full access to host hardware. When you mount EBS Provisioned IOPS volumes on Nitro-based instances, you can provision from 100 IOPS up to 64,000 IOPS per volume compared to just up to 32,000 on other instances.

**References:**

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-types.html>  
<https://tutorialsdojo.com/amazon-elastic-compute-cloud-amazon-ec2/>

### Storage with Highest IOPS for EC2 Instance

When talking about storage and IOPS in EC2 instances, the first thing that pops into the minds of people is Amazon EBS Provisioned IOPS. Amazon EBS Provisioned IOPS volumes are the highest performing EBS volumes designed for your critical, I/O intensive applications. These volumes are ideal for both IOPS-intensive and throughput-intensive workloads that require extremely low latency. And since they are EBS volumes, your data will also persist even after shutdowns or reboots. You can create snapshots of these volumes and copy them over to your other instances, and much more.

But what if you require both high IOPS and low latency performance, and the data doesn't necessarily have to be stored on the volume? If you have this requirement, then the instance store volumes on specific instance types might be preferable to EBS-provisioned IOPS volumes. EBS volumes are attached to EC2 instances virtually, so there is still some latency in there. Instance store volumes are physically attached to the EC2 instances themselves, so your instances are able to access the data much faster. Instance store volumes can come in HDD, SSD, or NVME SSD, depending on the instance type you choose. Available storage space will depend on the instance type as well.

**Reference:**

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/InstanceStorage.html>





## Instance Purchasing Options

AWS offers multiple options for you to purchase compute capacity that will best suit your needs. Aside from pricing on different instance types and instance sizes, you can also specify how you'd like to pay for the compute capacity. With EC2 instances, you have the following purchase options:

- 1) **On-Demand Instances** – You pay by the hour or the second depending on which instances you run for each running instance. If your instances are in a stopped state, then you do not incur instance charges. No long-term commitments.
- 2) **Savings Plans** – Receive discounts on your EC2 costs by committing to a consistent amount of usage, in USD per hour, for a term of 1 or 3 years. You can achieve higher discount rates by paying a portion of the total bill upfront, or paying full upfront. There are two types of Savings Plans available:
  - a) **Compute Savings Plans** provide the most flexibility since it automatically applies your discount regardless of instance family, size, AZ, region, OS or tenancy, and also applies to Fargate and Lambda usage.
  - b) **EC2 Instance Savings Plans** provide the lowest prices but you are committed to usage of individual instance families in a region only. The plan reduces your cost on the selected instance family in that region regardless of AZ, size, OS, or tenancy. You can freely modify your instance sizes within the instance family in that region without losing your discount.
- 3) **Reserved Instances (RI)** – Similar to Saving Plans but less flexible since you are making a commitment to a consistent instance configuration, including instance type and Region, for a term of 1 or 3 years. You can also pay partial upfront or full upfront for higher discount rates. A Reserved Instance has four instance attributes that determine its price:
  - a) Instance type
  - b) Region
  - c) Tenancy - shared (default) or single-tenant (dedicated) hardware.
  - d) Platform or OS

Reserved Instances are automatically applied to running On-Demand Instances provided that the specifications match. A benefit of Reserved Instances is that you can sell unused Standard Reserved Instances in the AWS Marketplace. There are also different types of RIs for you to choose from:

- a) **Standard RIs** - Provide the most significant discount rates and are best suited for steady-state usage.
- b) **Convertible RIs** - Provide a discount and the capability to change the attributes of the RI as long as the resulting RI is of equal or greater value.
- c) **Scheduled RIs** - These are available to launch within the time windows you reserve. This option allows you to match your capacity reservation to a predictable recurring schedule that only requires a fraction of a day, a week, or a month.



|   | Standard RI | Convertible RI |
|---|-------------|----------------|
| Applies to usage across all Availability Zones in an AWS region               | Yes         | Yes            |
| Can be shared between multiple accounts within a consolidated billing family. | Yes         | Yes            |
| Change Availability Zone, instance size (for Linux OS), networking type       | Yes         | Yes            |
| Change instance families, operating system, tenancy, and payment option       | No          | Yes            |
| Benefit from Price Reductions   | No          | Yes            |
| Can be bought/sold in Marketplace   | Yes         | No             |

- 4) **Spot Instances** – Unused EC2 instances that are available for a cheap price, which can reduce your costs significantly. The hourly price for a Spot Instance is called a Spot price. The Spot price of each instance type in each Availability Zone is set by Amazon EC2, and is adjusted gradually based on the long-term supply of and demand for Spot Instances. Your Spot Instance runs whenever capacity is available and the maximum price per hour that you've placed for your request exceeds the Spot price. When the Spot price goes higher than your specified price, your Spot Instance will be stopped or terminated after a two minute warning. Use Spot Instances only when your workloads can be interrupted
- 5) **Dedicated Hosts** – You pay for a physical host that is fully dedicated to running your instances, and bring your existing per-socket, per-core, or per-VM software licenses to reduce costs. Support for multiple instance sizes on the same Dedicated Host is available for the following instance families: c5, m5, r5, c5n, r5n, and m5n. Dedicated Hosts also offers options for upfront payment for higher discounts.
- 6) **Dedicated Instances** – Pay by the hour for instances that run on single-tenant hardware. Dedicated Instances that belong to different AWS accounts are physically isolated at a hardware level. Only your compute nodes run in single-tenant hardware; EBS volumes do not.

|  | Dedicated Hosts   | Dedicated Instances  |
|--|---|----------------------|
| <b>Billing</b>                                   | Per-host billing  | Per-instance billing |
| <b>Visibility of sockets, cores, and host ID</b> | Provides visibility on the number of sockets and physical cores | No visibility        |
| <b>Host and instance affinity</b>                | Allows you to consistently deploy                               | Not supported        |



|   |   |   |
|---|---|---|
|   | your instances to the same physical server over time  |   |
| <b>Targeted instance placement</b>                | Provides additional visibility and control over how instances are placed on a physical server | Not supported   |
| <b>Automatic instance recovery</b>                | Supported   | Supported   |
| <b>Bring Your Own License (BYOL)</b>              | Supported   | Not supported   |
| <b>Instances must run within a VPC</b>            | Yes   | Yes   |
| <b>Can be combined with other billing options</b> | On-demand Dedicated Hosts, Reserved Dedicated Hosts, Savings Plans                            | On-demand Instances, Reserved Dedicated Instances, Spot Instances |

- 7) **Capacity Reservations** – Allows you to reserve capacity for your EC2 instances in a specific Availability Zone for any duration. No commitment required.

**References:**

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-purchasing-options.html>  
<https://aws.amazon.com/ec2/pricing/>  
<https://tutorialsdojo.com/amazon-elastic-compute-cloud-amazon-ec2/>

## EC2 Placement Groups

Placement Groups is a logical grouping of your interdependent instances in AWS. This logical grouping affects how your instances are placed on the underlying hardware. Having the instances in a placement group has particular benefits in terms of network latency, throughput, and minimizing correlated hardware failure. By default, AWS automatically spreads out your instances across underlying hardware to reduce this correlated hardware failure.

AWS offers different placement strategies which can suit the placement requirements of your application hosted in Amazon EC2.

EC2 > Placement groups > Create placement group
? ↺ 🔍

### Create placement group info

**Placement group settings**

**Name**

**Placement strategy**  
Determines how the instances are placed on the underlying hardware.

Cluster
Cluster
Spread
Partition

You can add up to 50 more tags.

Cancel
Create group

## Cluster Placement Group

A cluster placement group is a logical group of instances within a single Availability Zone and instances from peered VPC in the same region. Through VPC Peering, you can still add instances from different Availability Zones to your cluster placement group.

Instances on the same cluster group have a low network latency and high throughput. A cluster placement group is beneficial to applications with a high volume of network traffic between their instances. To further maximize these network performance benefits, you can choose instance types with enhanced networking for your cluster placement group.



AWS recommends launching the instances for the cluster placement group through a single launch request. They also recommend using the same instance type for all the instances in the placement group to minimize



the chance of getting an insufficient capacity error. This error comes out when there is not enough hardware capacity to launch an instance. For example, when adding more instances to an existing placement group or adding instances with a different instance type. The capacity error can also be encountered when you stop and then start an instance again in a placement group.

The screenshot displays two panels from the AWS Management Console. The left panel, titled 'Advanced details', shows a 'Placement group' section with a dropdown menu set to 'td-demo' (Strategy: cluster, Shared: No) and a text field containing the ID 'pg-0e6e94d7c5db0a11b'. A 'Create new placement group' link is visible. The right panel, titled 'Summary', shows a 'Number of instances' field set to '3'. Below this, a note states: 'When launching more than 1 instance, consider EC2 Auto Scaling'.

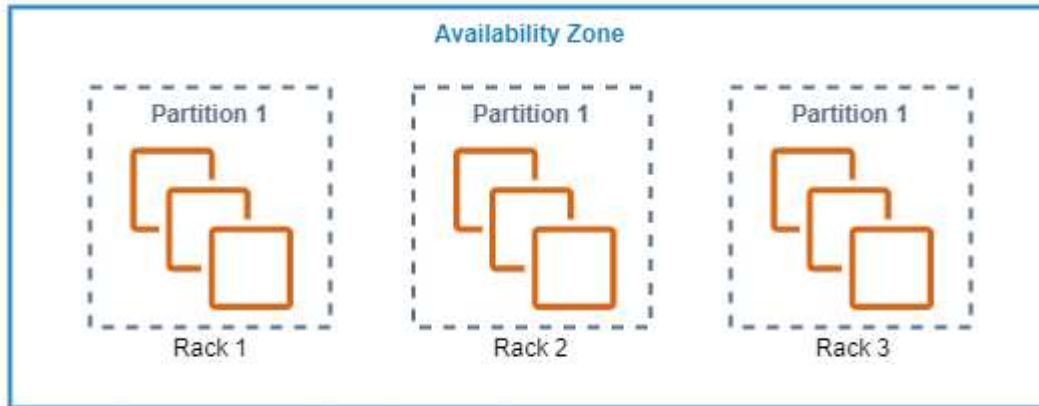
However, if you need to launch an instance to an existing placement group with running instances and encounter an insufficient capacity error, try to stop and start all of the running instances inside the placement group, then relaunch the instance. Doing so may force the instances to boot into new hardware capable of accommodating all the instance requests for the placement group.

Cluster placement groups are commonly used for High-Performance Computing (HPC) applications, like genomics, computational chemistry, financial risk modeling, machine learning, deep learning, etc.

## Partition Placement Group

A partition placement group spreads all instances into logical segments called partitions. Each partition has a dedicated rack with its network and power source. This placement strategy ensures that all partitions are isolated from each other, reducing the risk of correlated hardware failures.

Also, partition placement groups can have partitions from different Availability Zones in the same Region with a limit of seven partitions per AZ. The account limit determines the maximum number of instances. However, a maximum of two partitions is allowed for the partition placement group with Dedicated Instance.



When launching instances to the partition placement group, you can specify the specific partition.

#### Placement group [Info](#)

demo-partition-placement-group  
Strategy: partition Number of partitions: 7 Shared: No

pg-0cb5107ee65696ce6

[Create new placement group](#)

#### Target partition [Info](#)

Select

Select

1

2

3

4

5

6

7

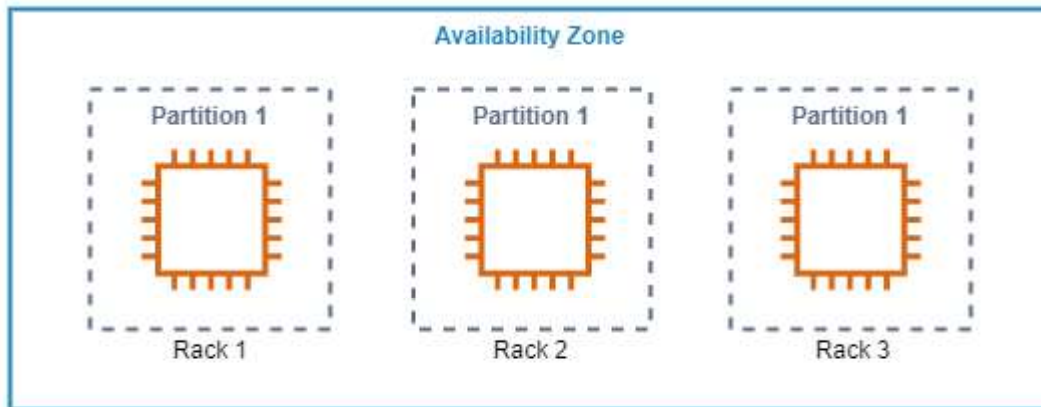
To achieve high availability for the application, we often go to multi-AZ deployment, but some applications are dependent on internode latency, thus making it unavailable for multi-AZ deployment. With a partition placement group, you can deploy this kind of application in a single Availability Zone but with improved performance and less chance for correlated hardware faults.

Applications like HDFS, HBase, and Cassandra are benefiting from this kind of placement strategy. Because they are topology-aware applications, they can use the topology information to make intelligent data storage decisions.



## Spread Placement Group

A Spread placement group is a placement strategy that strictly hosts instances separately on a distinct rack that has an individual network and power source. Since all instances are hosted on distinct racks, you can freely have multiple instance types or add instances over time on your spread placement group.



Since instances on the spread placement group are isolated from each other, the chance of having hardware faults is reduced when compared to instances sharing the same rack.

Like partition placement groups, spread placement groups can also span on different Availability Zones with a maximum of seven running instances per AZ.

EC2 > Placement groups > Create placement group

Create placement group

Placement group settings

Name

Placement strategy

Determines how the instances are placed on the underlying hardware.

Spread

Cluster

Spread

Partition

Tags - optional

No tags associated with the resource.

Add new tag

You can add up to 50 more tags.

Cancel

Create group



For the Partition and Spread placement group, there are times when a unique hardware is unavailable to accommodate all instance requests. When this happens, try to request again later as more hardware becomes available over time.

**Reference:**

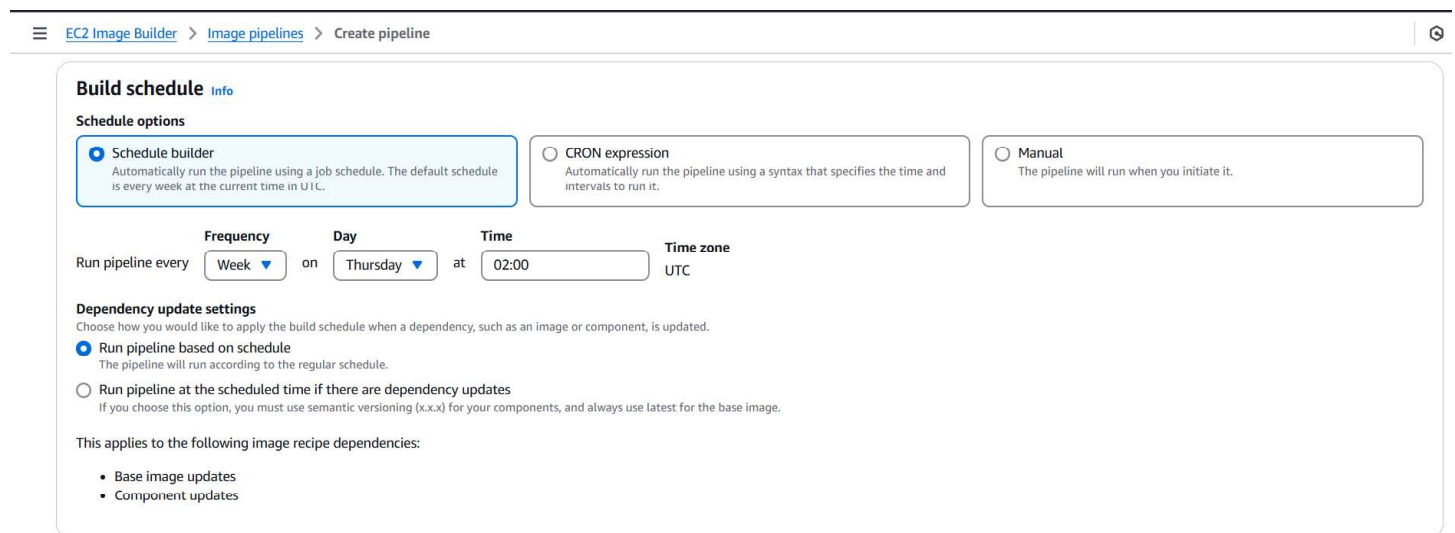
<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html>

## EC2 Image Builder

EC2 Image Builder is an AWS service that automates the process of creating, managing, and deploying machine and docker images both for your AWS environment and on-premises. You can keep your images updated through the image builder, automate image customization, validate image integrity and functionality through testing, and deploy images in different AWS regions. The image builder is pretty straightforward; it lets you create an Image Pipeline, configure an Image recipe, define the infrastructure, and set the image distribution.

## Image Pipelines

To automate the creation of images, AWS allows you to create a pipeline where you can configure the necessary components of your custom images. The image creation will run based on the defined build schedule and frequency or can be manually run.



EC2 Image Builder > Image pipelines > Create pipeline

### Build schedule info

**Schedule options**

- ☒ **Schedule builder**  
Automatically run the pipeline using a job schedule. The default schedule is every week at the current time in UTC.
- ☐ **CRON expression**  
Automatically run the pipeline using a syntax that specifies the time and intervals to run it.
- ☐ **Manual**  
The pipeline will run when you initiate it.

Run pipeline every **Week** on **Thursday** at **02:00** **UTC**

**Dependency update settings**  
Choose how you would like to apply the build schedule when a dependency, such as an image or component, is updated.

- ☒ **Run pipeline based on schedule**  
The pipeline will run according to the regular schedule.
- ☐ **Run pipeline at the scheduled time if there are dependency updates**  
If you choose this option, you must use semantic versioning (x.x.x) for your components, and always use latest for the base image.

This applies to the following image recipe dependencies:

- Base image updates
- Component updates

## Image Recipes Configuration

Image recipes are where you define the customization and testing of your images. Image recipes are reusable and have version control. It consists of the following components.

### Source Images

The source image will be the baseline of your custom image. Image builder supports the customization for Amazon Machine Image (AMI) and Docker image. For AMI, this can be AWS-managed images or a custom AMI. Likewise, for Docker images, it can be AWS-managed images, an ECR image, or a public image from Docker Hub.

## Image type

Choose the image type

### Output type

☒ Amazon Machine Image (AMI)



☐ Docker image



You can select from different operating systems and versions; availability depends on the image type.

### Image Operating System (OS)

Image Builder supports Amazon Linux, Windows, Ubuntu, CentOS, RHEL, and SLES.

☒ Amazon Linux  
Amazon Linux 2



☐ Windows  
Windows Server 2012R2, 2016, 2019,  
2004, and 20H2



☐ Ubuntu  
Ubuntu 16, 18 and 20



☐ CentOS  
CentOS 7 and 8



☐ Red Hat Enterprise Linux (RHEL)  
RHEL 7 and 8



☐ SUSE Linux Enterprise Server  
(SLES)  
SLES 12 and 15



Image builder installs SSM Agent during the build process, but you can remove the agent after the pipeline execution.



### Instance configuration [Info](#)

Choose the instance configuration

#### SSM agent

EC2 Image Builder uses AWS Systems Manager agent as part of the image build process. The agent is installed for you automatically if it was not installed in the source image.

- ☒ Remove SSM agent after pipeline execution  
If you deselect this box, Image Builder keeps the SSM agent in the output image.

Should it be necessary to run a command on the instance launch, you can set it on the User Data. Note that defining User data requires your source image to have the SSM Agent pre-installed or that you include SSM Agent installation on the User Data.

#### User data

You can specify user data to configure an instance or run a configuration script during launch.

**i** When you provide user data, you must also ensure that the SSM agent is already installed on the source image or that you install it with your user data input.

Enter the user data.





- ☐ The user data is already base64

## Build and Test Components

A build component installs software packages to your source image. You can select from Amazon-managed build components, share build components to your AWS account, or create a new one. See the example Amazon-managed build components below.

#### Selected components (2)

Expand the component to view versioning options and input parameters. To sort the build sequence, drag the components up and down.





| Sequence | Component (drag the component up or down to change the sequence)  | <input type="checkbox"/> Expand all |
|----------|---|-------------------------------------|
| 1        | <div> amazon-cloudwatch-agent-linux<br/>▶ Versioning options</div> <div>Owner: Amazon</div> <div></div> |                                     |
| 2        | <div> amazon-corretto-8-jdk<br/>▶ Versioning options</div> <div>Owner: Amazon</div> <div></div>         |                                     |



Test Components are optional, but it's a better option to configure this to validate the integrity and functionality of the output image. You can also use Amazon-managed, shared, or create a new test component. See the example Amazon-manage test components below.

#### Selected components (2)

Expand the component to view versioning options and input parameters. To sort the build sequence, drag the components up and down.

| Sequence | Component (drag the component up or down to change the sequence)  | Expand all               |
|----------|---|--------------------------|
| 1        | <div> amazon-cloudwatch-agent-linux<br/>▶ Versioning options</div> <div>Owner: Amazon</div> <div></div> | <input type="checkbox"/> |
| 2        | <div> amazon-corretto-8-jdk<br/>▶ Versioning options</div> <div>Owner: Amazon</div> <div></div>         | <input type="checkbox"/> |

## Storage

Storage configuration is optional. You can configure this during the instance launch.

### Storage (volumes) - optional

The storage device settings for your pipeline.

#### ▼ EBS volume 1 (AMI root)

|  |   |  |
|--|---|--|
| Device name                            | Snapshot - optional                                 | Volume type  |
| <input type="text" value="/dev/xvda"/> | <input type="text" value="snap-0896bce87dc58384b"/> | <input type="text" value="General Purpose SSD (gp2)"/> |
| Size (GiB)                             | IOPS  | Encryption (KMS alias)                                 |
| <input type="text" value="8"/>         | <input type="text" value="100"/>                    | <input type="text" value="Do not enable"/>             |

☒ Delete on termination

## Infrastructure Configuration

The Infrastructure configuration is an optional configuration on the image pipeline. You can configure the Instance Type, VPC settings, IAM role, and Tags for the output image. A notification can also be published using SNS.





## AWS infrastructure

Service-specific defaults will be applied if you do not select values.

### Instance type [Info](#)

Select one or more instance types to customize your image.

Choose one or more instance types ▼

### SNS topic [Info](#)

Select an SNS topic to receive notifications and alerts from EC2 Image Builder

Choose SNS topic ▼



[Create SNS topic](#)

### ► VPC, subnet and security groups

Specify advanced settings to launch the instance that will customize your image.

### ► Troubleshooting settings [Info](#)

Specify settings to troubleshoot issues with building your image.

Besides the default IAM policies that the image builder uses, the configured IAM role should also have the necessary permissions to execute all the build and test components defined on the image recipe.

Default IAM Policies for Image Builder:

- *EC2InstanceProfileForImageBuilder*
- *EC2InstanceProfileForImageBuilderECRContainerBuilds*
- *AmazonSSMManagedInstanceCore*

## Distribution Settings

You can configure the image deployment on the Distribution settings. You can choose multiple AWS Regions as image destinations. For Amazon Machine Images, you can configure the output image name, AMI sharing, and the license and launch template configuration. For the docker images, you need to specify the Regions and ECR repository name.

### Reference:

<https://docs.aws.amazon.com/imagebuilder/latest/userguide/what-is-image-builder.html>



## Amazon EC2Rescue

While AWS takes care of the underlying infrastructure for EC2, customers are responsible for configuring, maintaining, and troubleshooting their instances.

### EC2Rescue for Windows Server

EC2Rescue for Windows Server is a downloadable tool for Windows Server instances to help you diagnose and troubleshoot issues. You can also use EC2Rescue to detect potential problems in your current instances.

#### Diagnose and Rescue an Offline Instance

EC2Rescue scans and diagnoses the Amazon EBS root volumes of the problematic instances. To do this, EC2Rescue requires a host instance where it will be installed. The EBS root volume should be detached from the problematic instance and attached to the EC2Rescue instance host.

Reminders:

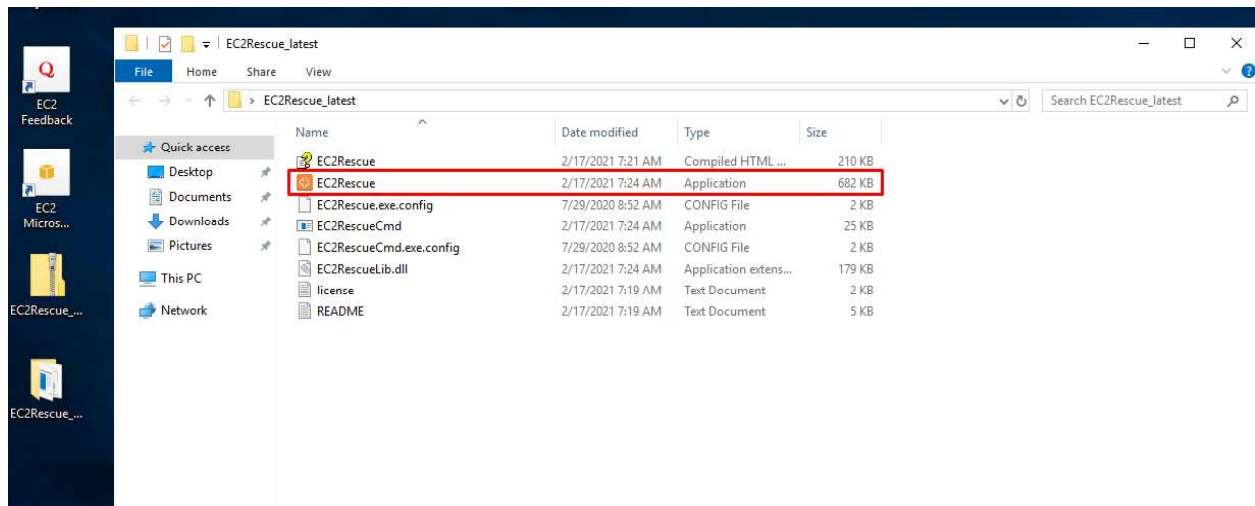
- The EC2Rescue tool only runs on Windows Server 2008 R2 or later and requires .NET Framework 3.5 SPI or later.
- The EC2Rescue instance host should also be accessible using an RDP connection.
- The instance where EC2Rescue is installed and the instance to be diagnosed should reside on the same Availability Zone.

The following instructions will guide you on how to check an instance using EC2Rescue.

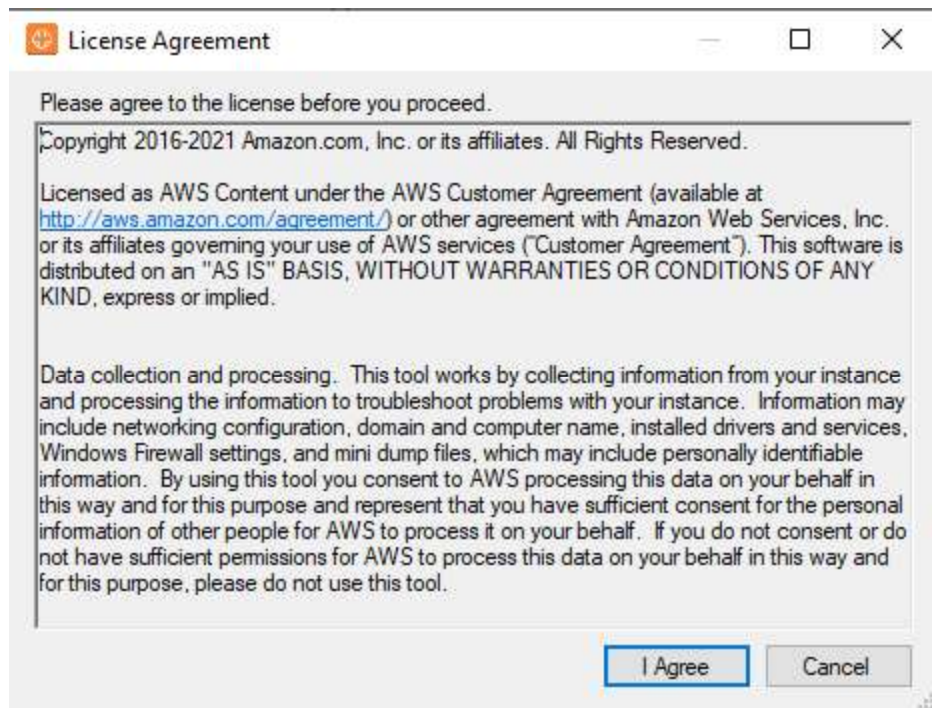
1. Connect to the EC2Rescue host and download the tool [here](#) using a browser or using the PowerShell command below.

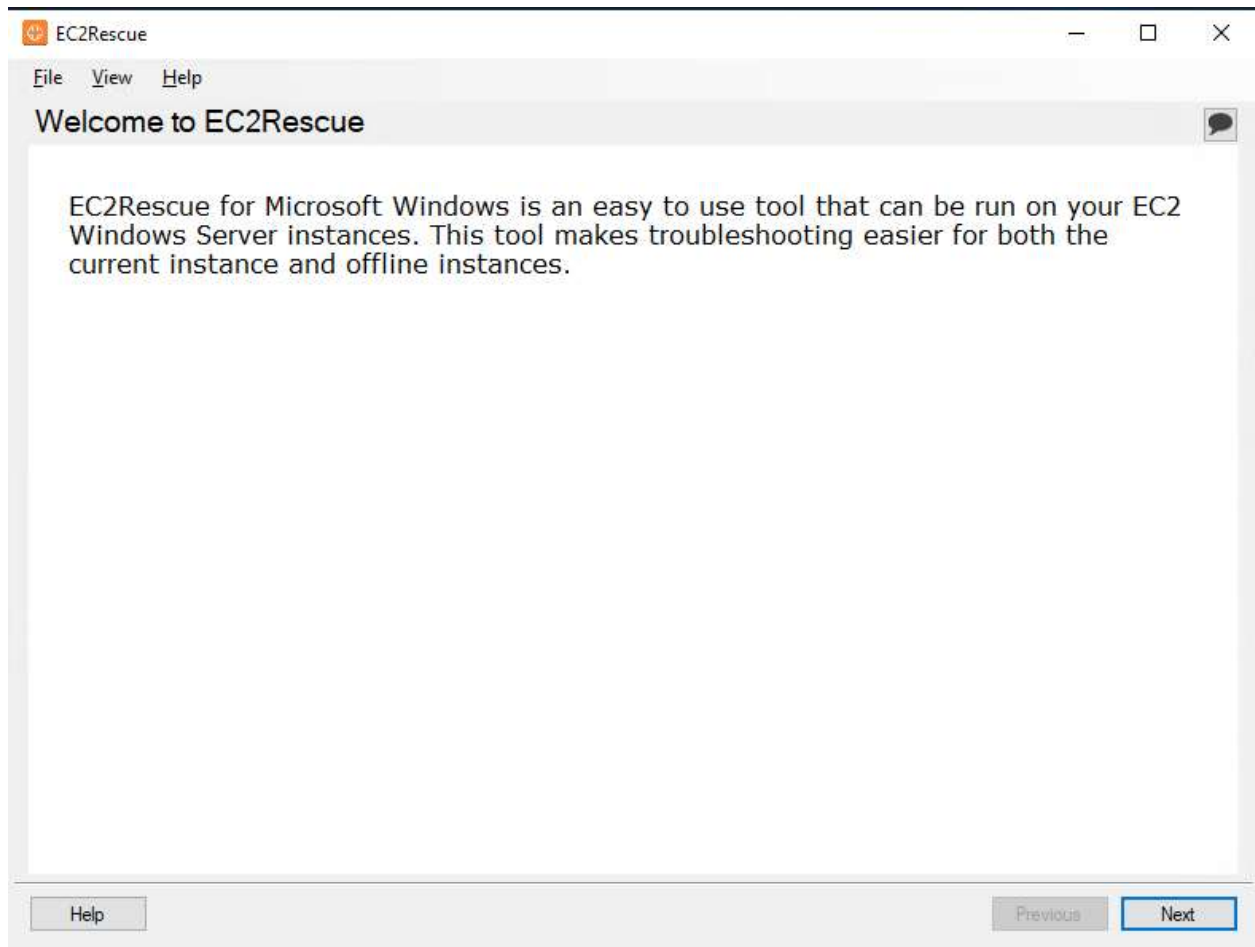
```
Invoke-WebRequest https://s3.amazonaws.com/ec2rescue/windows/EC2Rescue_latest.zip -OutFile  
$env:USERPROFILE\Desktop\EC2Rescue_latest.zip
```

2. Extract the downloaded zip file. Once extracted, run the EC2Rescue application.

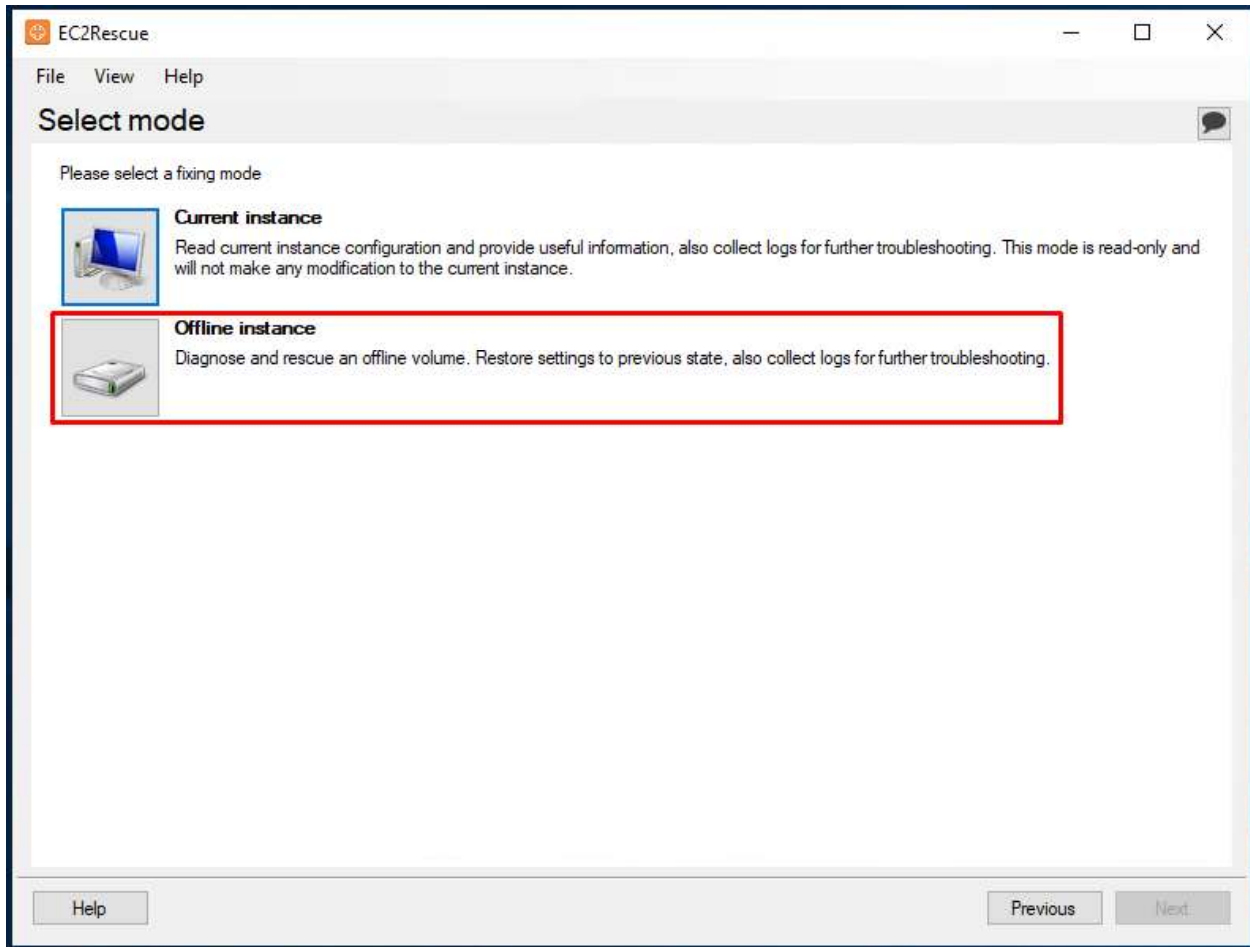


- Click **I agree** on the license agreement and click **Next** on the Welcome screen.

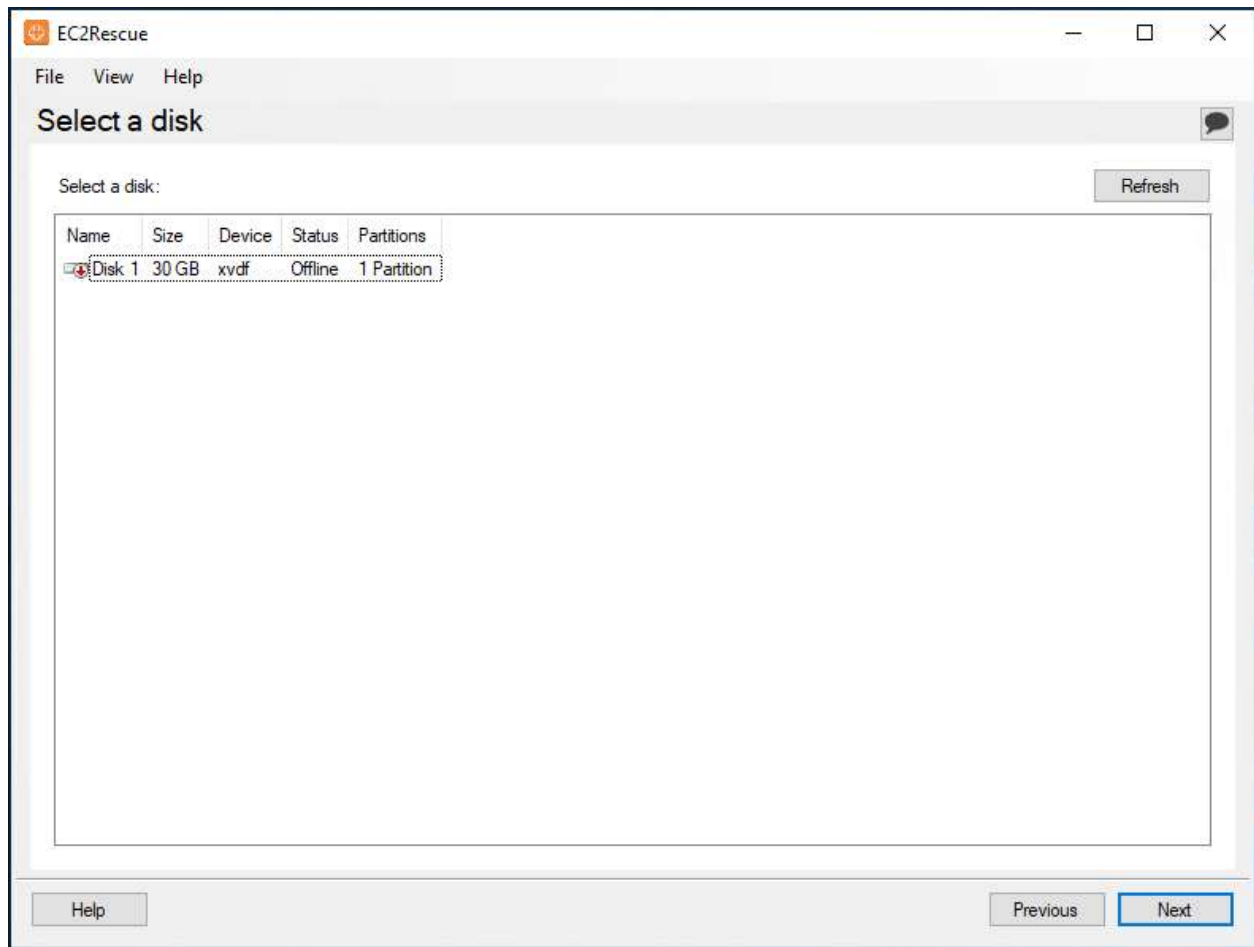




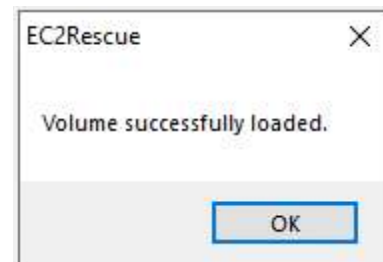
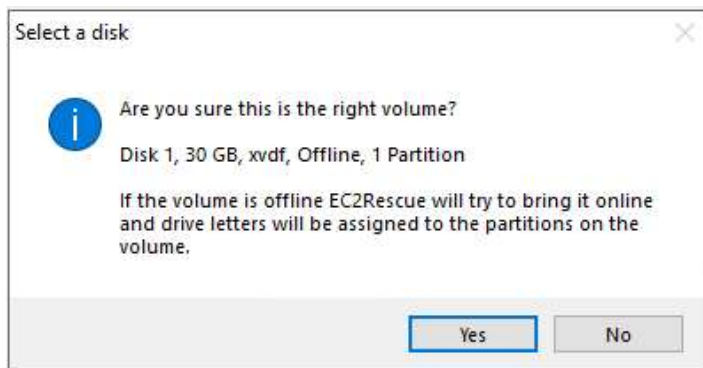
4. Select **Offline instance** and click **Next**.



5. Select the **EBS volume** of the problematic instance and click **Next**. If you are checking multiple root volumes, note the device name when attaching the volumes to the EC2Rescue host.

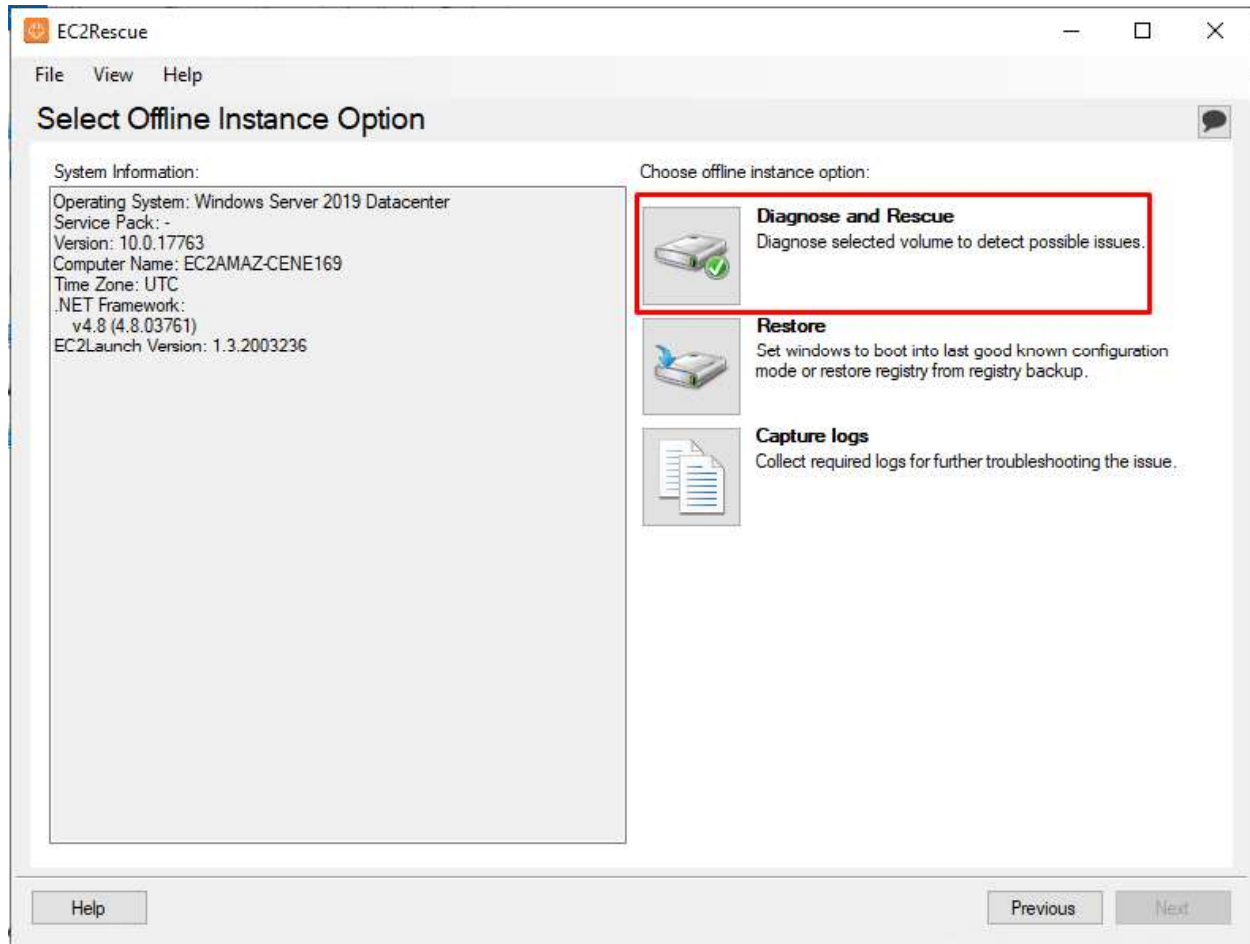


6. Click **Yes** to confirm. A popup window will show once the EBS volume is successfully loaded.

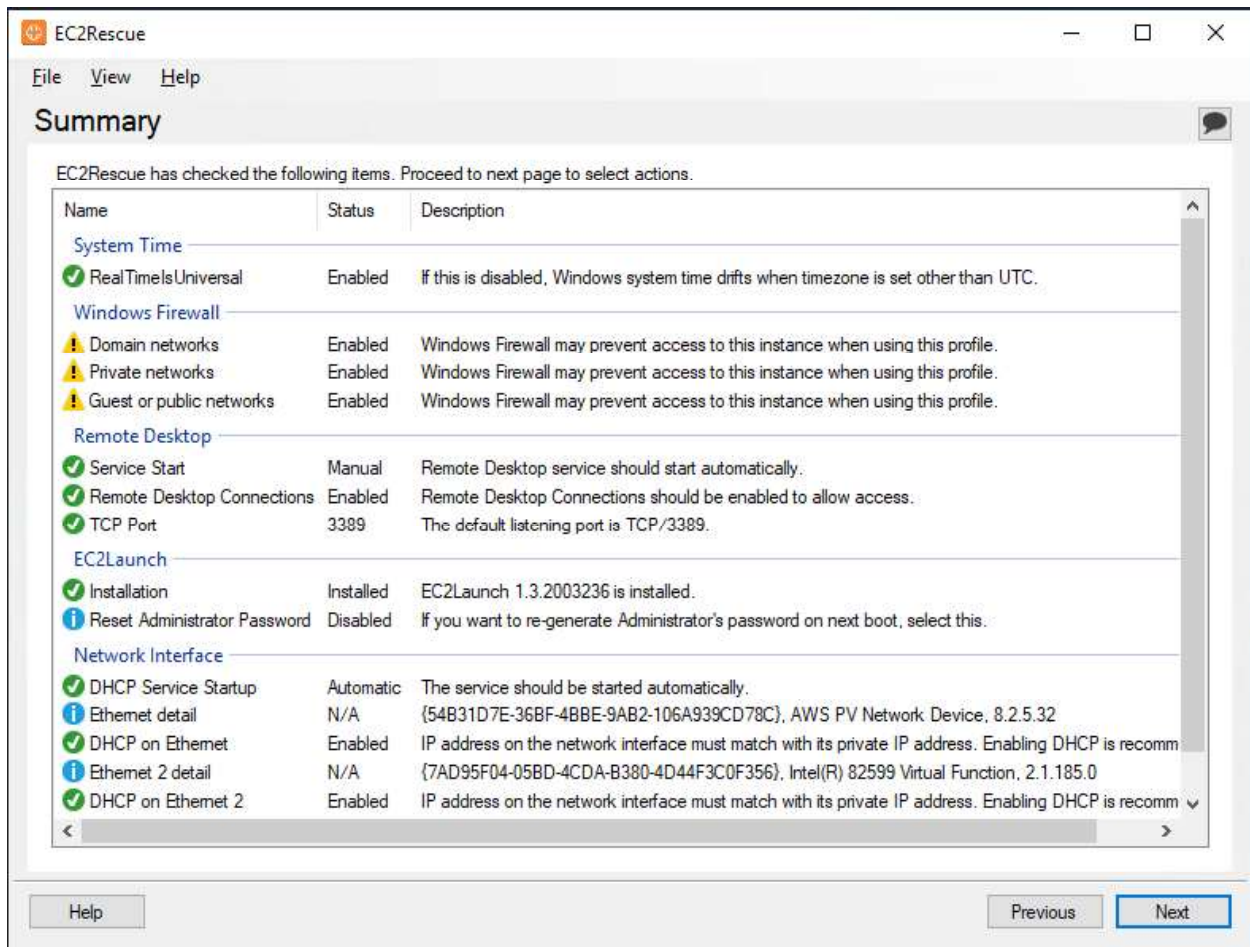


7. Once the volume is loaded, the EC2Rescue tool will display system information of the instance. You will also see different offline instance options. In this case, select **Diagnose and Rescue** to proceed.

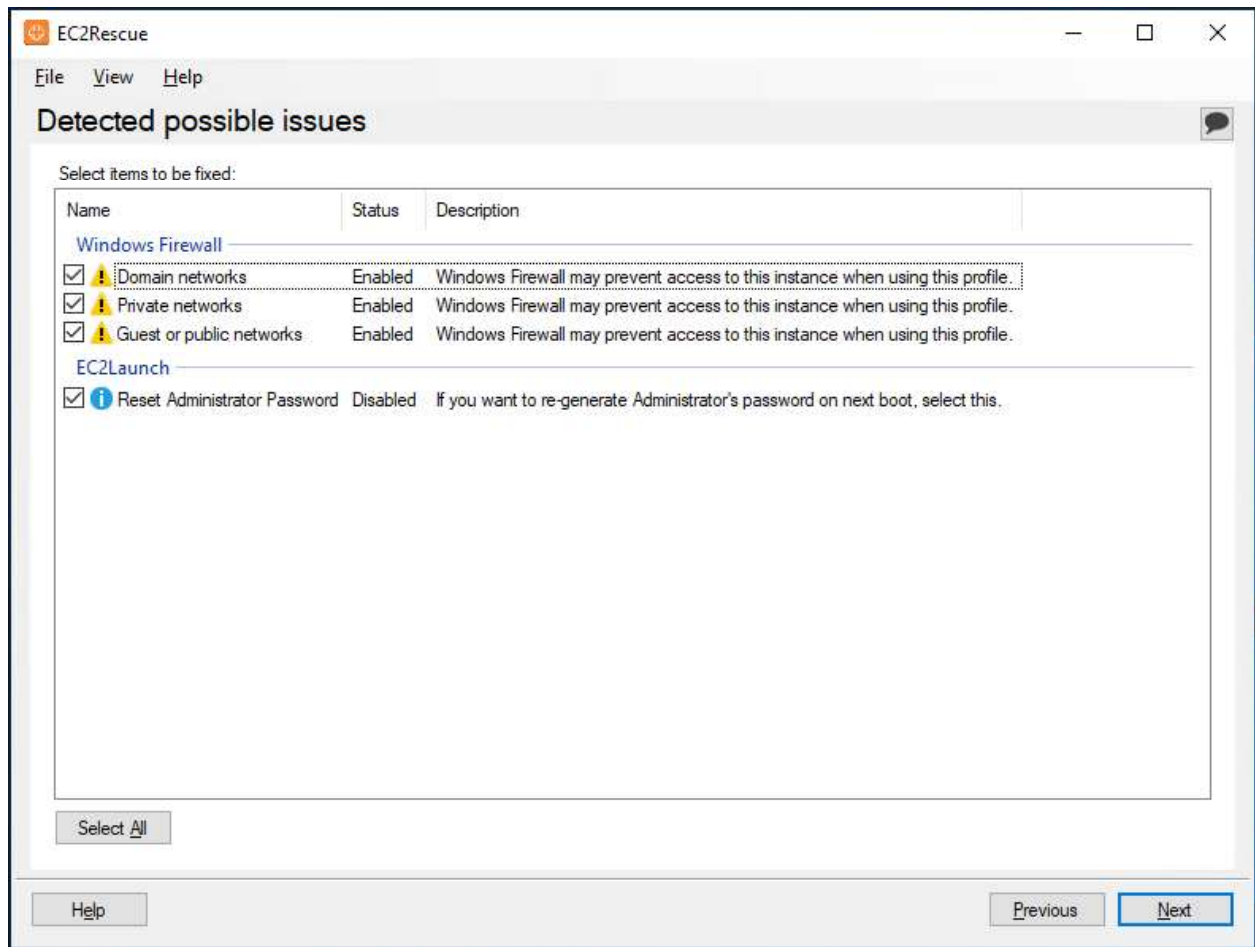




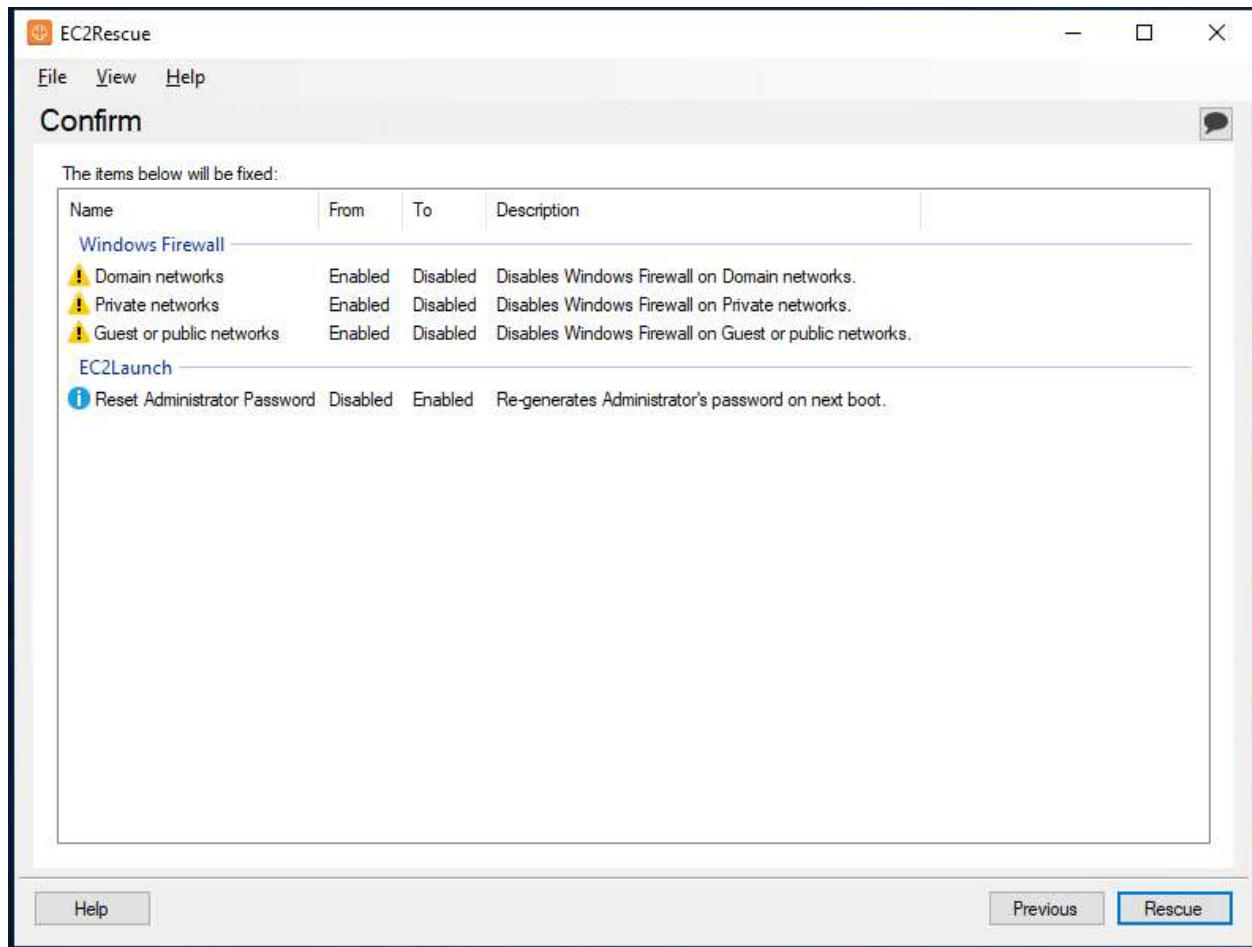
8. The EC2Rescue tool will now start scanning and diagnosing the volume. Once the diagnostic is done, it will summarize the necessary configurations, including their status and description. Click **Next** to proceed.



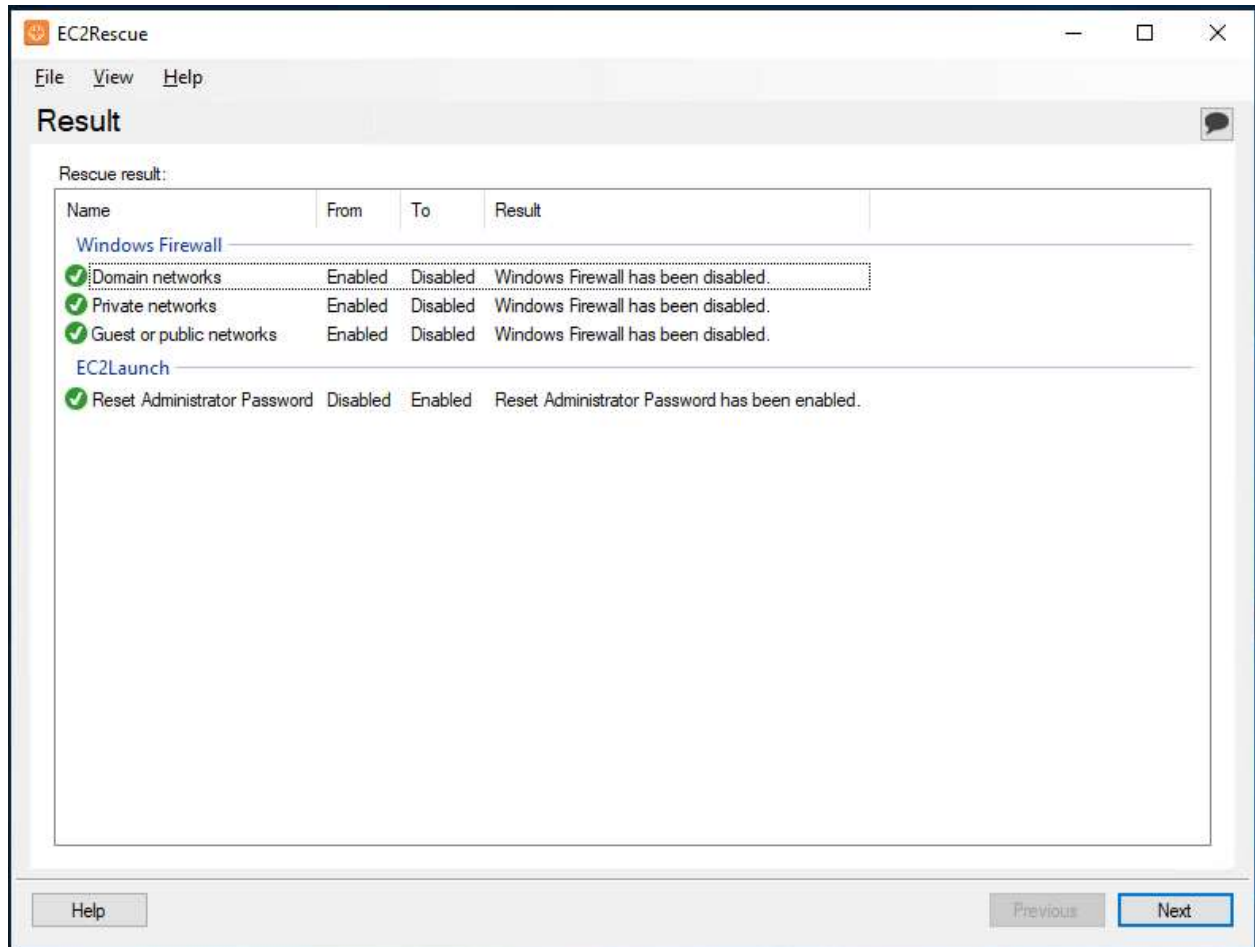
- EC2Rescue will give you a list of potential issues of the instance. From this, you can select the fixes you find necessary for your instance. Click **Next** to proceed.



10. Once confirmed, click **Rescue**.



11. Click **Next** to continue applying the changes.



12. Click **Finish**.