

FOR FREE MATERIALS JOIN HERE

JOIN IN TELEGRAM: [CLICK HERE](#)

FOLLOW IN INSTAGRAM: [CLICK HERE](#)

FOLLOW IN TWITTER: [CLICK HERE](#)

FOR FREE MATERIALS JOIN HERE

JOIN IN TELEGRAM: [CLICK HERE](#)

FOLLOW IN INSTAGRAM: [CLICK HERE](#)

FOLLOW IN TWITTER: [CLICK HERE](#)

JON BONSO AND LERVIN JOHN OBANDO



AWS CERTIFIED
**SYSOPS
ADMINISTRATOR
ASSOCIATE**
VERSION 2.0



Tutorials Dojo
Study Guide and Cheat Sheets



TABLE OF CONTENTS

INTRODUCTION	7
AWS CERTIFIED SYSOPS ADMINISTRATOR ASSOCIATE EXAM OVERVIEW	8
AWS CERTIFIED SYSOPS ADMINISTRATOR ASSOCIATE EXAM - STUDY GUIDE AND TIPS	12
Study Materials	12
AWS Services to Focus On	14
Exam Labs	16
Common Exam Scenarios	16
Validate Your Knowledge	21
Sample Practice Test Questions:	22
Question 1	22
Question 2	24
AWS Deep Dives	28
EC2 Placement Groups	28
Cluster Placement Group	28
Partition Placement Group	30
Spread Placement Group	32
EC2 Image Builder	34
Image Pipelines	34
Image Recipes Configuration	34
Source Images	34
Build and Test Components	36
Storage	37
Infrastructure Configuration	37
Distribution Settings	38
Amazon EC2Rescue	39
EC2Rescue for Windows Server	39
Diagnose and Rescue an Offline Instance	39
Collecting Logs from an Offline Instance	49
Restore Options for an Offline Instance	52
Checking the Current Instance	54
EC2Rescue for Windows on Systems Manager	56
EC2Rescue for Linux	57
Installing EC2Rescue for Linux	58



Diagnose Issues Using EC2Rescue for Linux	58
Creating Instance Backup Using EC2Rescue for Linux	59
EC2Rescue for Linux on Systems Manager	59
AWS Auto Scaling	61
Auto Scaling Group	61
Auto Scaling Templates	61
Launch Templates	61
Launch Configurations	65
Auto Scaling Group Configuration	65
AWS Compute Optimizer	72
Prerequisites	72
AWS Compute Optimizer Dashboard	73
Recommendations for EC2 Instances	74
Recommendations for Auto Scaling Group	75
Recommendations for EBS Volume Instances	76
Elastic Load Balancing	78
Load Balancer Types	79
ELB Features and Components	79
Load Balancer Scheme	79
IP Addresses Type	79
Listener	79
Target Group	80
Security Groups	80
Availability Zones	80
Health Checks	80
Sticky Sessions	81
Cross-zone Load Balancing	81
Connection Draining	81
Load Balancer Monitoring	81
Delete Protection	81
Choosing the Right Load Balancer	81
Application Load Balancer (ALB)	81
Network Load Balancer (NLB)	82
Gateway Load Balancer (GWLB)	82
S3 Presigned URL	83
Sharing S3 objects using Presigned URL	83
Uploading S3 Objects Using Presigned URL	86



S3 Transfer Acceleration	87
Amazon CloudFront	89
Caching Process	89
CloudFront Policies	89
Cache Policy	90
Origin Request Policy	91
Amazon ElastiCache	93
ElastiCache Memcached and Redis Engine	93
Clusters	93
Sharding	93
Multithreading	94
High Availability	94
Backup and Restore	94
Key Points	95
Virtual Private Cloud	96
Network Access Control List (NACL)	96
Route Tables	98
VPC Flow logs	100
Traffic Mirroring	101
Amazon Route 53	104
Domain Registration	104
Route 53 Service Integrations	105
Hosted Zones	105
Route 53 Health Checks	106
Route 53 Records	106
Routing Policy	108
DNS Record Types	109
Route 53 Resolver	110
Resolver Endpoints	110
Resolver Rules	113
Amazon Elastic File System (EFS)	115
EFS Storage Classes	115
Creating a File System	116
File System Access Point	120
Mounting a File System	122
Mount via DNS	124
Mount via IP	124



Amazon FSx	124
Amazon FSx for Windows File Server	125
File System Details	125
Optional Details	127
Working with Amazon FSx for Windows File Server	127
Amazon FSx for Lustre	128
File System Details	128
Optional Details	129
Working with Amazon FSx for Lustre	129
AWS DataSync	131
Supported AWS Storage Service	131
Working with DataSync	131
AWS Backup	134
Backup Plan	134
On-demand Backup	136
Backup Vault	137
Protected Resources	138
Backup Jobs	138
Cross-account Management	139
Amazon Relational Database Service (RDS)	140
Amazon RDS Features and Components	140
Amazon RDS Database Engines	140
Choosing Suitable RDS DB Instance Classes	141
Choosing the Right RDS DB Instance Storages	141
Choosing a Region and Availability Zone for RDS Instance	143
Increasing Database Availability Using Multi-AZ Deployment	143
Improving Database Performance using Read Replica and DB Clusters	144
Adding an RDS Proxy	145
Working with RDS Backup	146
Monitoring a Database Instance	148
Deleting a Database Instance	151
AWS Config	152
AWS Config Continuous Configuration Monitoring	153
Deploying Resources with CloudFormation	156
StackSets	156
Nested Stacks	156
Deleting a Stack	157



Retain	157
Snapshot	158
AWS Systems Manager Patch and Change Manager	160
AWS Systems Manager Patch Manager	160
AWS Systems Manager Change Manager	161
Encryption on AWS Storage Services	164
S3 Encryption	164
Server-Side Encryption	164
Client-Side Encryption	165
Encrypting Existing S3 Objects	165
EFS Encryption	165
Data at Rest Encryption	165
Data In Transit Encryption	166
EBS Encryption	166
Creating Encrypted EBS Volume	166
Snapshots	167
RDS Encryption	168
Encrypting RDS Database Instance with AWS KMS	168
Securing Database Connection on RDS	169
Security on AWS	170
AWS KMS Customer Master Key (CMK) Rotation	170
Secrets Manager vs Parameter Store	171
IAM Access Analyzer	173
AWS Certificate Manager	175
AWS Billing and Governance	178
AWS Organizations	178
Service Control Policies (SCP)	181
Cost Explorer	182
Cost Allocation Tags	183
AWS License Manager	184
Monitoring and Logging on AWS	187
CloudWatch Metrics for EC2	187
Creating CloudWatch Alarm	190
Working with CloudWatch Logs	196
Event-driven Architecture with Amazon EventBridge	200
Exploring Events on CloudTrail	202
COMPARISON OF AWS SERVICES	205



S3 vs EBS vs EFS	205
Amazon S3 vs Glacier	207
S3 Standard vs S3 Standard-IA vs S3 One Zone-IA vs S3 Intelligent Tiering	208
AWS DataSync vs Storage Gateway	209
S3 Transfer Acceleration vs Direct Connect vs VPN vs Snowball vs Snowmobile	210
Amazon EBS: SSD vs HDD	213
RDS vs DynamoDB	216
RDS vs Aurora	219
Multi-AZ deployments vs. Multi-Region deployments vs. Read Replicas	224
EC2 Container Services ECS vs Lambda	225
Security Group vs NACL	226
Application Load Balancer vs Network Load Balancer vs Gateway Load Balancer	228
EC2 Instance Health Check vs ELB Health Check vs Auto Scaling and Custom Health Check	231
ELB Health Checks vs Route 53 Health Checks For Target Health Monitoring	232
CloudTrail vs CloudWatch	233
CloudWatch Agent vs SSM Agent vs Custom Daemon Scripts	234
Latency Routing vs Geoproximity Routing vs Geolocation Routing	235
Service Control Policies vs IAM Policies	237
S3 Pre-Signed URLs vs CloudFront Signed URLs vs Origin Access Identity	238
SNI Custom SSL vs Dedicated IP Custom SSL	239
Redis (cluster mode enabled vs disabled) vs Memcached	240
FINAL REMARKS AND TIPS	241
ABOUT THE AUTHORS	242



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 environments 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 maturing. 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 specialized individuals to operate on the AWS cloud successfully.

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 cost-effective. Resources can be quickly provisioned for a very low price and can be decommissioned as quickly once you don't need them anymore. The cloud is an essential piece in DevOps and SysOps since you can quickly spin up test environments and simplify deployment processes that are usually difficult and expensive to do in traditional data center setups.

The AWS Certified SysOps Administrator Associate is a well-recognized certificate in the IT industry and is a major booster for career opportunities and salary increases. Having this certificate in your portfolio means that you have the 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. Previously, it was mandatory that you first pass the SysOps Associate exam since the Professional level certificate is very tough and requires much hands-on experience from the exam taker. So if you are planning to pursue a career in Cloud DevOps, passing the AWS Certified SysOps Administrator 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 working on [hands-on sessions](#) and [practice exams](#) to further expand your knowledge and improve your test-taking skills.



AWS CERTIFIED SYSOPS ADMINISTRATOR 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 other two role-based certifications: Systems Operations (SysOps) Administrator and Developer Associate later that year.

Exam Details

The AWS Certified SysOps Administrator - Associate examination is intended for systems administrators in a systems operations role with at least one year of experience in deployment, management, and operations on AWS. The number of questions varies, depending on the exam.

Exam Code:	SOA-C02
No. of Questions:	50 to 55 multiple choice/multiple answer and 3 exam labs
Score Range:	100/1000
Cost:	150 USD (Practice exam: 20 USD)
Passing Score:	720/1000
Time Limit:	3 hours (180 minutes) - 2 hrs for multiple choice and 1 hr for exam labs
Format:	Multiple choice/multiple answers and Exam Labs
Delivery Method:	Testing center or online proctored exam

Exam Labs

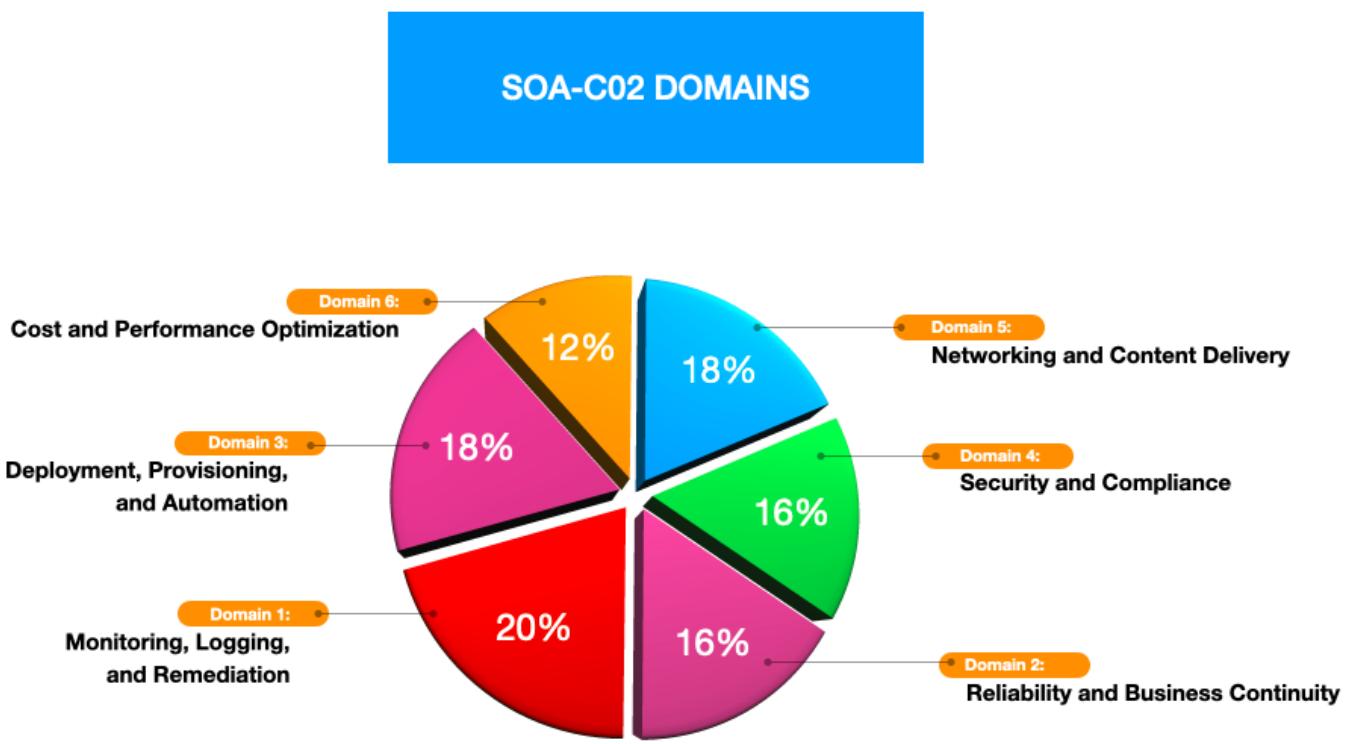
The exam labs add an extra dimension to the exam. This will make sure that you are not only familiar with cloud concepts but also adept in troubleshooting and building solutions using the AWS Management Console or the AWS Command Line Interface.

At the beginning of the exam, you will receive a message regarding the number of questions in the multiple-choice and multiple-response section, including the number of exam labs in the exam lab section.

Make sure to allot at least 20 minutes to complete each exam lab. Finish all instructions on an exam lab before deciding to move to the next one. You will NOT be able to return to a prior exam lab. You are allowed to use the virtual machine notepad or AWS CLI while working on the exam labs.

Exam Domains

The AWS Certified SysOps Administrator - Associate exam has seven (6) different domains, each with corresponding weight and topic coverage. The domains are: Monitoring, Logging, and Remediation (20%), Reliability and Business Continuity (16%), Deployment, Provisioning, and Automation (18%), Security and Compliance (16%), Networking and Content Delivery (18%), and Cost and Performance Optimization (12%).



Domain 1: Monitoring, Logging, and Remediation (20%)

- 1.1 Implement metrics, alarms, and filters by using AWS monitoring and logging services
- 1.2 Remediate issues based on monitoring and availability metrics

Domain 2: Reliability and Business Continuity (16%)

- 2.1 Implement scalability and elasticity
- 2.2 Implement high availability and resilient environments
- 2.3 Implement backup and restore strategies



Domain 3: Deployment, Provisioning, and Automation (18%)

- 3.1 Provision and maintain cloud resources
- 3.2 Automate manual or repeatable processes

Domain 4: Security and Compliance (16%)

- 4.1 Implement and manage security and compliance policies
- 4.2 Implement data and infrastructure protection strategies

Domain 5: Networking and Content Delivery (18%)

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

Domain 6: Cost and Performance Optimization (12%)

- 6.1 Implement cost optimization strategies
- 6.2 Implement performance optimization strategies

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 SysOps Administrator 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 1 - 5 business days after your exam.

For individuals who unfortunately do not pass their exams, you must wait 14 days before you 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 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 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.
- **Free Practice Exam** - To help you prepare for your next exam, AWS provides another voucher that you can use to take any official AWS practice exam for free. You can access your voucher code from the "Benefits" section of your AWS Certification Account.
- **AWS Certified Store** - All AWS certified professionals will be given access to exclusive AWS Certified merchandise. You can get your store access from the "Benefits" section of your AWS Certification Account.
- **Certification Digital Badges** - You can showcase your achievements to your colleagues and employers with digital badges on your email signatures, LinkedIn profile, or on your 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.
- **Eligibility to join AWS IQ** - With the AWS IQ program, you can monetize your AWS skills online by providing hands-on assistance to customers around the globe. AWS IQ will help you stay sharp and be well-versed on various AWS technologies. You can work at the comforts of your home and decide when or where you want to work. Interested individuals must be based in the US, have an Associate, Professional, or Specialty AWS Certification and be over 18 of age.

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 SYSOPS ADMINISTRATOR ASSOCIATE EXAM - STUDY GUIDE AND TIPS

If you are a Systems Administrator or a DevOps Engineer, then this certification will test your knowledge on these areas in AWS. Your experience in these fields will come in handy in passing the exam, but this should be complemented by actual AWS SysOps knowledge. In the AWS Certified SysOps Administrator Associate Exam (or AWS SOA for short), the questions including the [exam labs](#) will test your ability to perform the following:

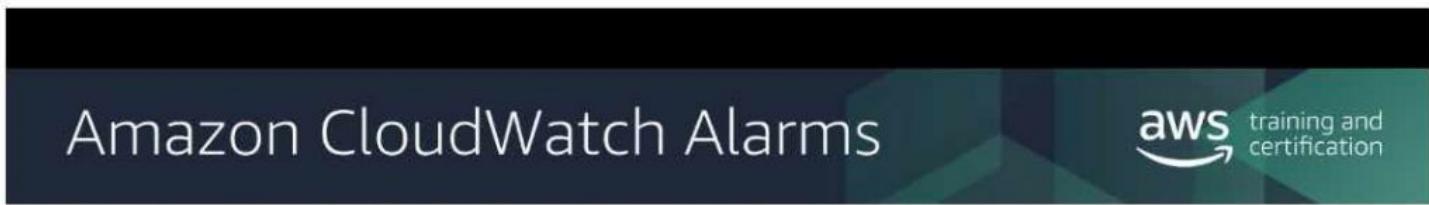
- Deploy, manage, and operate scalable, highly available, and fault-tolerant systems on AWS
- Implement and control the flow of data to and from AWS
- Select the appropriate AWS service based on compute, data, or security requirements
- Identify appropriate use of AWS operational best practices
- Estimate AWS usage costs and identify operational cost control mechanisms
- Migrate on-premises workloads to AWS

Given the scope of the questions and [exam labs](#), you should learn the concepts of the AWS architecture, the AWS Operational Framework, as well as the AWS CLI and AWS SDK/API tools. Having prior knowledge of fundamental networking and security will also be very valuable. This guide aims to provide you a straightforward guide when reviewing for this exam.

Study Materials

The [FREE AWS Exam Readiness video course](#), [official AWS sample questions](#), whitepapers, AWS Documentation, [AWS cheat sheets](#), and [AWS practice exams](#) will be your primary study materials for this exam. There are multiple papers that you should read and familiarize yourself with as a SysOps Administrator.

Having an AWS account you can use will help ingest the different concepts within these whitepapers. Since the exam itself contains multiple scenario questions and exam labs, using the services and applying them in practice yourself will allow you to determine the types of situations they are applied in.



- Tests a selected metric against a specific threshold
- Triggers an action, conditionally



Threshold **not** exceeded



Threshold exceeded



Alarm has started, metric isn't available, or insufficient data



Additional details regarding your AWS SOA exam can be seen in this [AWS exam blueprint](#).

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. [Development and Test on AWS](#) - Study how you can leverage AWS to create development and test environments, implement pipelines and automation, and perform different validation tests for your applications.
3. [Backup and Recovery Approaches Using AWS](#) - Learn which AWS services offer backup and restore features. It is also important to know how these backups are stored and secured, and selecting the correct storage options for them.
4. [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.



5. [Amazon Web Services: Overview of Security Processes](#) - You should study the different security features in AWS - including infrastructure, account, network, application, and data security. Determine which aspects of security are your responsibilities, and which are AWS'.
6. [AWS Security Best Practices](#) - This whitepaper complements the previous. Understand the security best practices and their purpose in your environment. Some services offer more than one form of security feature, such as multiple key management schemes for encryption. It is important that you can determine which form is most suitable to the given scenarios in your exam.
7. [Architecting for the Cloud: AWS Best Practices](#) - Be sure to understand the best practices in AWS since exam questions will focus their scenarios around these best practices. The whitepaper contains a number of design principles with examples for each.
8. [AWS Well-Architected Framework](#) - This whitepaper is one of the most important papers that you should study for the SOA-C02 exam. It discusses the different pillars that make up a well-architected cloud environment.

Optional whitepapers:

1. [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.
2. [AWS Disaster Recovery Plans](#) - As a SysOps Administrator, 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 SysOps Administrator, 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 on these services.

Core services to study:

1. [EC2](#) - As the most fundamental compute 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 the S3 storage types and what differs 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. Study also the different 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 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. [OpsWorks](#)
8. [Direct Connect](#)

For the exam version (SOA-C02), you should also know the following services:

1. Amazon FSx for Windows File Server and Amazon FSx for Lustre
2. AWS Backup
3. EC2 Image Builder
4. S3 Transfer Acceleration
5. AWS Global Accelerator
6. RDS Proxy
7. IAM Access Analyzer

Exam Labs

The SOA-C02 includes an exam labs section where you have to perform SysOps related tasks on the AWS Management Console. To prepare for this, make sure to play around with the different AWS services covered in the exam. You don't need to memorize all the configurations for each service. But you have to be really good at navigating the AWS Management console to understand where you can configure the requirements in each exam lab. Focus on preparing for exam labs on setting up a VPC, CloudWatch, Load Balancer, Auto Scaling, CloudFormation, and S3.

View our sample exam lab [here](#).

Common Exam Scenarios

Scenario	Solution
Monitoring, Reporting, and Remediation	
You need to set up an alert that notifies the IT manager about EC2 instances service limits.	Use Amazon CloudWatch Events to detect and react to changes in the status of Trusted Advisor checks



You need to track the deletion and rotation of CMKs.	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.	http://169.254.169.254/latest/
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 <= 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.

Reliability and Business Continuity

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 SurgeQueueLength 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.
Deployment, Provisioning, and Automation	
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.
A total of 500 TB of data needs to be transferred to Amazon S3 in the fastest way.	Use multiple AWS Snowball devices
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
Security and Compliance	
You have to rotate an existing CMK with imported key material every 6 months	Create a new CMK with imported key material and update the key ID to point to the new CMK
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 SS3-S3, 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 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
Cost and Performance Optimization	
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.

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 portal. [Tutorials Dojo](#) also offers a top-notch set of [AWS Certified SysOps Administrator 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 SysOps Administrator needs to write a template that will automatically copy objects from an existing S3 bucket into the new one.

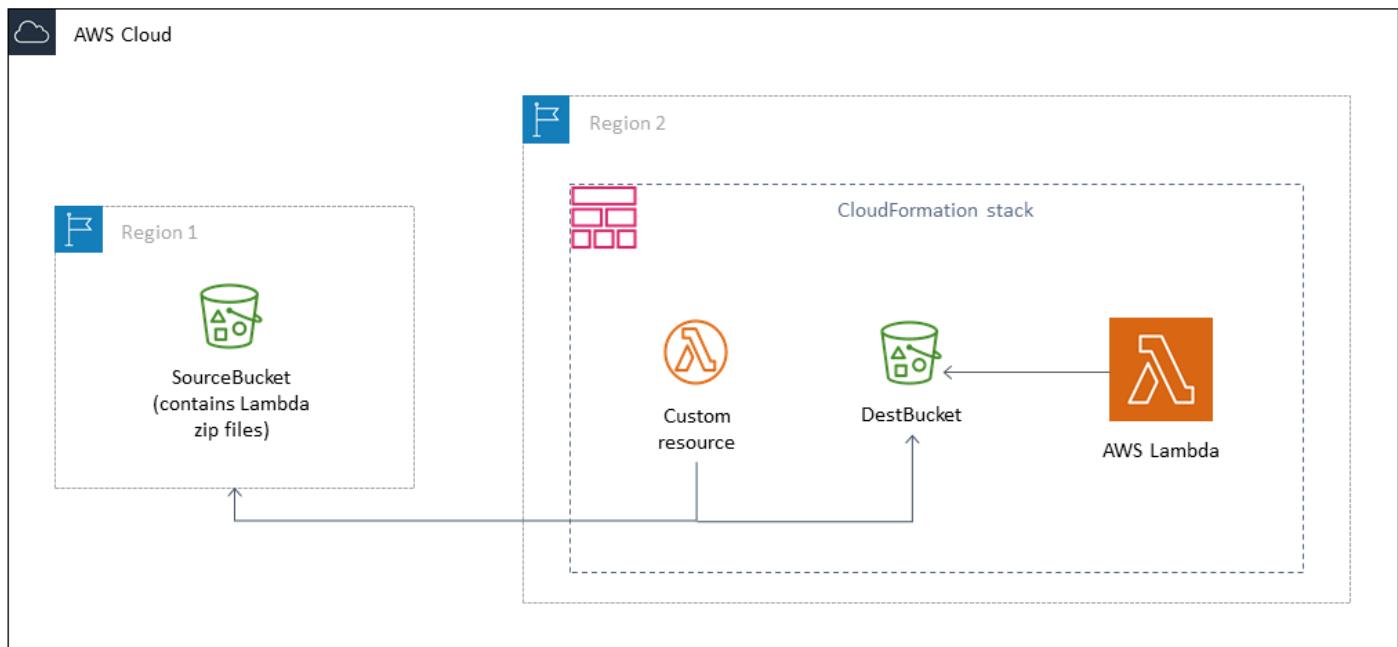
Which of the following is the most suitable configuration for this scenario?

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. Set up the CloudFormation template to use the AWS Data Pipeline CopyActivity object to copy the files from the existing S3 bucket to the new S3 bucket.
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: **Set up the CloudFormation template to CopyActivity object to copy the files from the existing S3 bucket to the new S3 bucket** is incorrect because CopyActivity does not support copying multipart Amazon S3 files. The most suitable configuration to copy the objects from an existing bucket to a new S3 bucket is to use a custom Lambda resource in CloudFormation.

The option that says: **Configure the CloudFormation template to modify the 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

A company plans to expand its use of AWS services across its product portfolios. To ensure separation of business processes for billing, security, and compliance, the SysOps Administrator must provide each department with new AWS accounts having governance guardrails and a defined baseline in place. An efficient and scalable provisioning process is required to optimize the workflow and save time.

Which of the following options can satisfy the given requirement?

1. Use AWS Batch and AWS Organizations to automatically provision new resources and accounts.
2. Use AWS Control Tower to generate templates in the Account Factory and to provision new accounts in AWS Service Catalog.



3. Use AWS OpsWorks for Chef Automate and bootstrapping scripts to handle the configuration management and provisioning tasks.
4. Use AWS Service Catalog and AWS Config to automate account creation and configuration.

Correct Answer: 2

AWS Control Tower provides a single location to easily set up your new well-architected multi-account environment and govern your AWS workloads with rules for security, operations, and internal compliance. You can automate the setup of your AWS environment with best-practices blueprints for multi-account structure, identity, access management, and account provisioning workflow. For ongoing governance, you can select and apply pre-packaged policies enterprise-wide or to specific groups of accounts.

AWS Control Tower > Account factory

Account factory Info

The account factory enables you to create standardized baselines and network configurations for accounts in your organization. Your users can configure and provision these new accounts in AWS Service Catalog.

Network configuration		
 Provision new account <input type="checkbox"/>		Edit
The following VPC configuration options are available to your users when they provision new accounts. You can modify these settings anytime.		
Internet-accessible subnet Disallow	Address range (CIDR) for account VPCs 172.31.0.0/16	Regions for VPC creation EU (Ireland) US East (N. Virginia) US East (Ohio) US West (Oregon)
Maximum number of private subnets 1		
Availability Zone count 3		

AWS Control Tower provides three methods for creating member accounts:

- Through the Account Factory console that is part of AWS Service Catalog.



-
- Through the Enroll account feature within AWS Control Tower.
 - From your AWS Control Tower landing zone's management account, using Lambda code and appropriate IAM roles.

AWS Control Tower offers "guardrails" for ongoing governance of your AWS environment. Guardrails provide governance controls by preventing deployment of resources that don't conform to selected policies or detecting non-conformance of provisioned resources. AWS Control Tower automatically implements guardrails using multiple building blocks such as AWS CloudFormation to establish a baseline, AWS Organizations service control policies (SCPs) to prevent configuration changes, and AWS Config rules to continuously detect non-conformance.

In this scenario, the requirement is to provide each department with AWS accounts that have governance guardrails and defined baseline in place. To save time and resources, you can use AWS Control Tower to automate the account creation. With the appropriate user group permissions, you can specify standardized baselines and network configurations for all accounts in the organization.

Hence, the correct answer is: **Use AWS Control Tower to generate templates in the Account Factory and to provision new accounts in AWS Service Catalog.**

The option that says: **Use AWS Service Catalog and AWS Config to automate account creation and configuration** is incorrect. Although you can use AWS Service Catalog to create and manage catalogs of your IT services, it still does not offer "guardrails" for ongoing governance of your AWS environment. Moreover, AWS Config is mainly used to evaluate the configuration of various AWS services in an AWS environment and not to provision new accounts.

The option that says: **Use AWS Batch and AWS Organizations to automatically provision new resources and accounts** is incorrect because AWS Batch can only provision compute resources. Furthermore, AWS Organizations do not fully provide governance guardrails unlike AWS Control Tower.

The option that says: **Use AWS OpsWorks for Chef Automate and bootstrapping scripts to handle the configuration management and provisioning tasks** is incorrect because AWS OpsWorks is not a suitable service to be used in provisioning new AWS accounts. The common use case for AWS OpsWorks for Chef Automate is to automate operational tasks on Amazon EC2 instances and on-premises servers.

References:

<https://docs.aws.amazon.com/controlltower/latest/userguide/account-factory.html>
<https://aws.amazon.com/blogs/mt/how-to-automate-the-creation-of-multiple-accounts-in-aws-control-tower/>
<https://aws.amazon.com/blogs/aws/aws-control-tower-set-up-govern-a-multi-account-aws-environment/>

Check out these AWS Cheat Sheets:

<https://tutorialsdojo.com/aws-cheat-sheets/>



Click [here](#) for more **AWS Certified SysOps Administrator 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 AWS SOA certification is one of the most sought after certifications in the SysOps Administration field. The exam will not be easy to pass, but it'll be worth it when you do.



AWS Deep Dives

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.

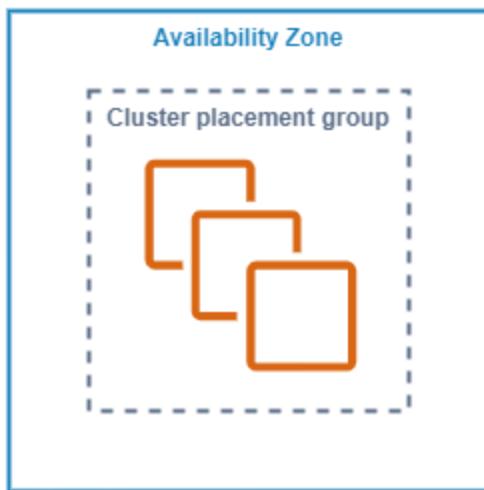
The screenshot shows the 'Create placement group' page in the AWS Management Console. The top navigation bar shows 'EC2 > Placement groups > Create placement group'. The main title is 'Create placement group'. Below it is a section titled 'Placement group settings' containing a 'Name' field with the value 'EC2TestPG'. A 'Placement strategy' dropdown menu is open, showing 'Cluster' as the selected option. Other options include 'Cluster', 'Spread', and 'Partition'. There is also an 'Add tag' button and a note stating 'You can add 50 more tags.' At the bottom right are 'Cancel' and 'Create group' buttons.

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.



Step 3: Configure Instance Details

Configure the instance to suit your requirements. You can launch multiple instances from the same AMI, request Spot instances to take advantage of lower costs, or use Auto Scaling to automatically manage the number of instances based on demand.

The screenshot shows the 'Configure Instance Details' section of the AWS Lambda 'Create Function' wizard. It includes fields for 'Number of instances' (set to 3), 'Purchasing option' (checkbox for Request Spot instances), 'Network' (VPC selected), 'Subnet' (No preference), 'Auto-assign Public IP' (Use subnet setting), and 'Placement group' settings. The 'Placement group' section is highlighted with a red box, showing options for adding to an existing group ('Add to existing placement group') or creating a new one ('Add to a new placement group'). A sub-section for 'Placement group name' shows 'EC2PGCluster (cluster)' selected. A note at the bottom suggests launching instances into an Auto Scaling Group for better management.

Number of instances Launch into Auto Scaling Group

You may want to consider launching these instances into an Auto Scaling Group to help manage them.

Purchasing option Request Spot instances

Network Create new VPC

Subnet

Auto-assign Public IP

Placement group Add instance to placement group
 Add to existing placement group.
 Add to a new placement group.

Placement group name

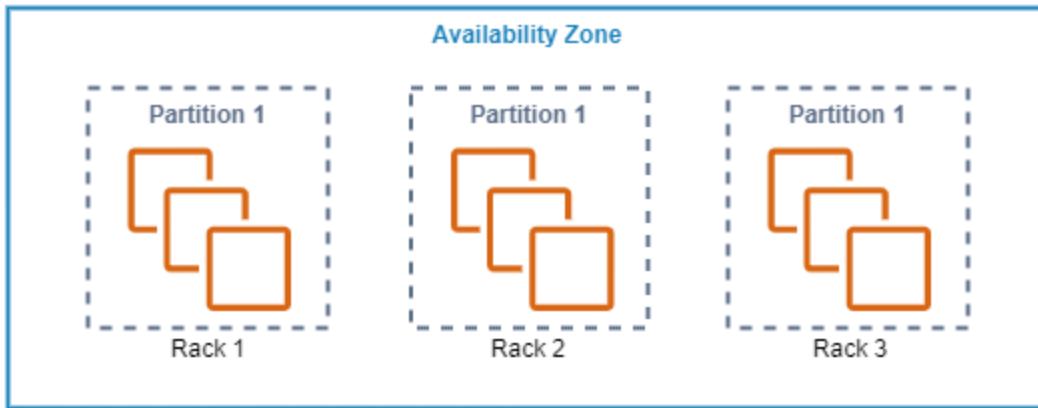
However, if you need to launch an instance to an existing placement group with running instances and encounter an insufficient capacity error, try to stop and start all of the running instances inside the placement group, then relaunch the instance. Doing so may force the instances to boot into new hardware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



When launching instances to the partition placement group, you can either let Amazon EC2 handle the distribution or specify the specific partition.

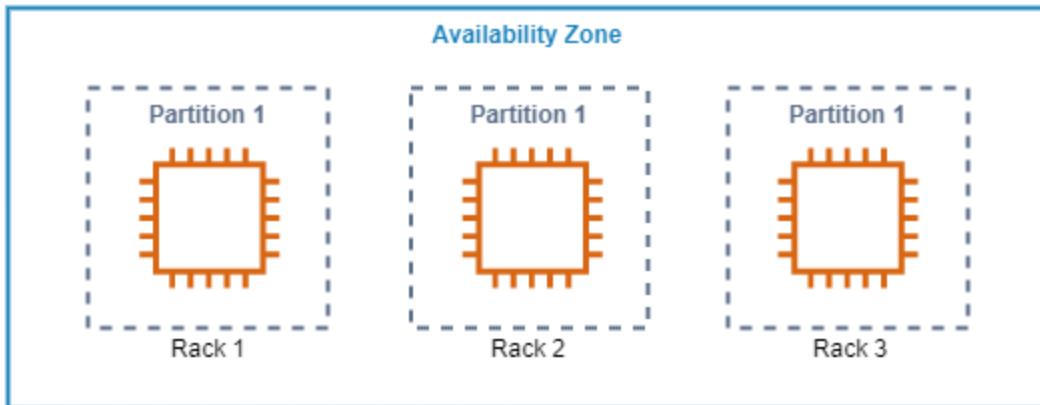
Placement group ⓘ	<input checked="" type="checkbox"/> Add instance to placement group
Placement group name ⓘ	<input checked="" type="radio"/> Add to existing placement group. <input type="radio"/> Add to a new placement group. EC2PartitionPG (partition)
Target partition ⓘ	Auto distribution Auto distribution
Capacity Reservation ⓘ	1 2 3 4 5 6 7
Domain join directory ⓘ	
IAM role ⓘ	

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.



The screenshot shows the 'Create placement group' wizard in the AWS Management Console. The current step is 'Placement group settings'. The 'Name' field contains 'EC2SpreadPG'. The 'Placement strategy' dropdown is set to 'Spread'. There is a section for 'Tags (Optional)' with a note that no tags are associated with the resource, and a button to 'Add tag'. A note at the bottom says 'You can add 50 more tags.' At the bottom right are 'Cancel' and 'Create group' buttons.

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 on 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.

The screenshot shows the 'Build schedule' section of the AWS Lambda configuration interface. It includes three options: 'Schedule builder' (selected), 'CRON expression', and 'Manual'. Below these are fields for 'Frequency' (set to 'Run pipeline every Week'), 'Day' (set to 'Monday'), 'Time' (set to '09:00'), and 'Time zone' (set to 'UTC').

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)	Expand all
1	amazon-cloudwatch-agent-linux ▶ Versioning options	X
2	amazon-corretto-8-jdk ▶ Versioning options	X



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 all
Sequence	Component (drag the component up or down to change the sequence)	Owner	X
1	amazon-cloudwatch-agent-linux ▶ Versioning options	Owner: Amazon	X
2	amazon-corretto-8-jdk ▶ Versioning options	Owner: Amazon	X

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
/dev/xvda	snap-0896bce87dc58384b	General Purpose SSD (gp2)
Size (GiB)	IOPS	Encryption (KMS alias)
8	100	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:

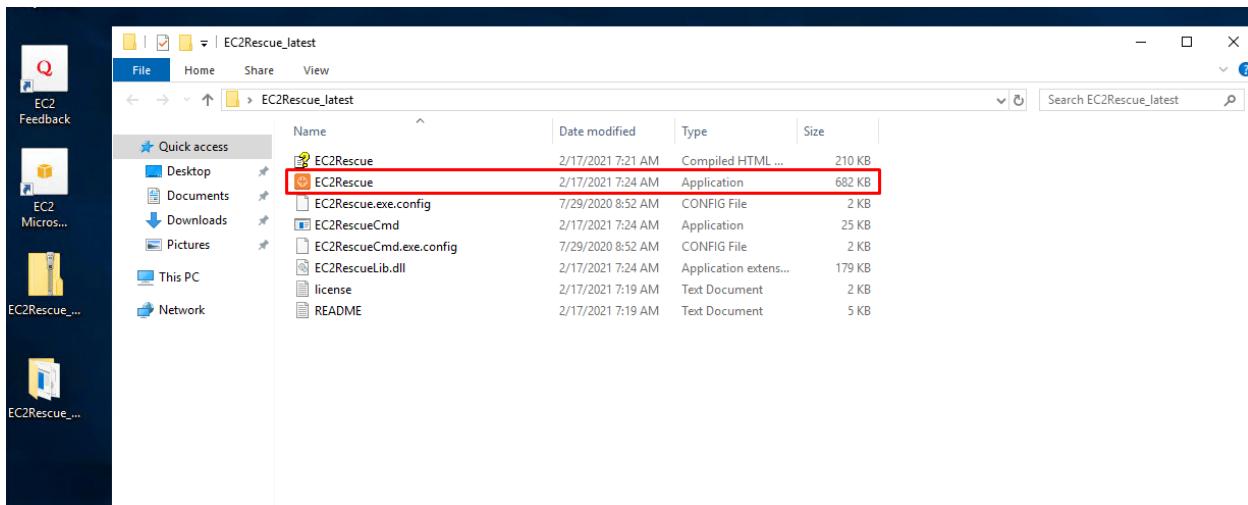
- The EC2Rescue tool only runs on Windows Server 2008 R2 or later and requires .NET Framework 3.5 SP1 or later.
- The EC2Rescue instance host should also be accessible using an RDP connection.
- The instance where EC2Rescue is installed and the instance to be diagnosed should reside on the same Availability Zone.

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

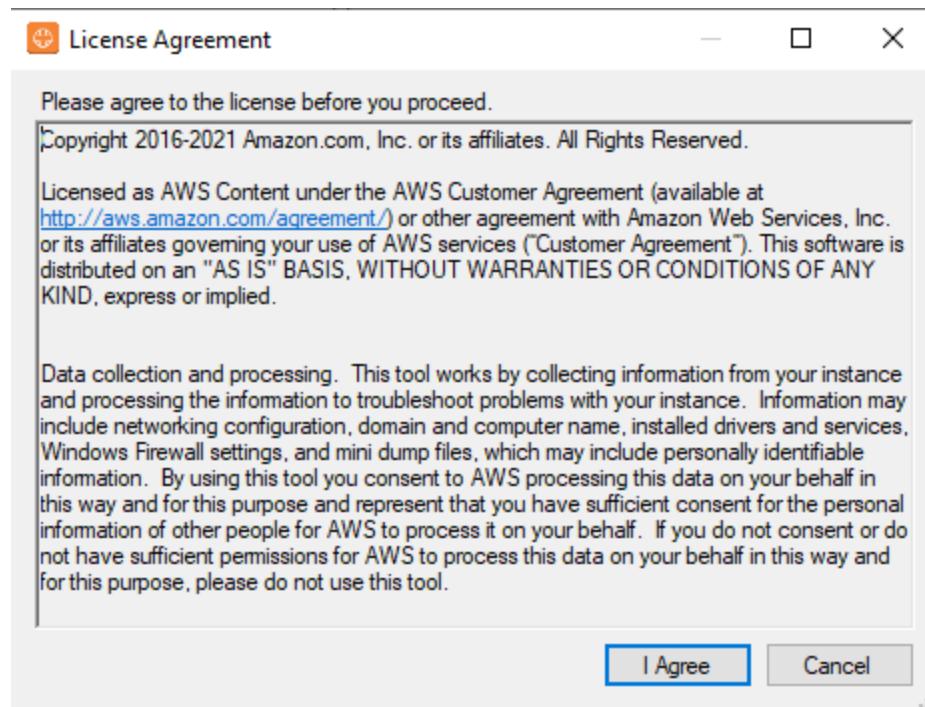
1. Connect to the EC2Rescue host and download the tool [here](#) using a browser or using the PowerShell command below.

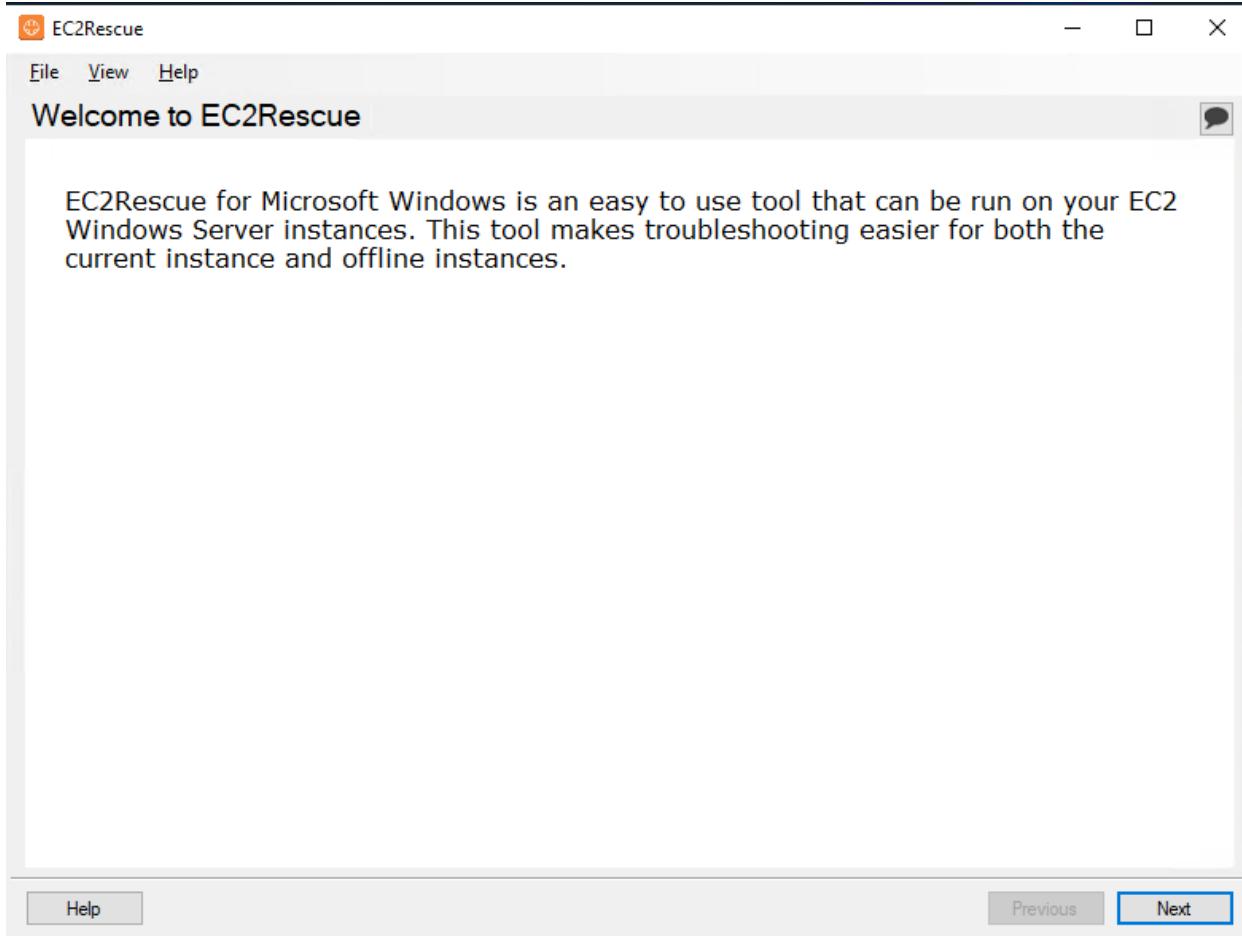
```
Invoke-WebRequest https://s3.amazonaws.com/ec2rescue/windows/EC2Rescue_latest.zip -OutFile $env:USERPROFILE\Desktop\EC2Rescue_latest.zip
```

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

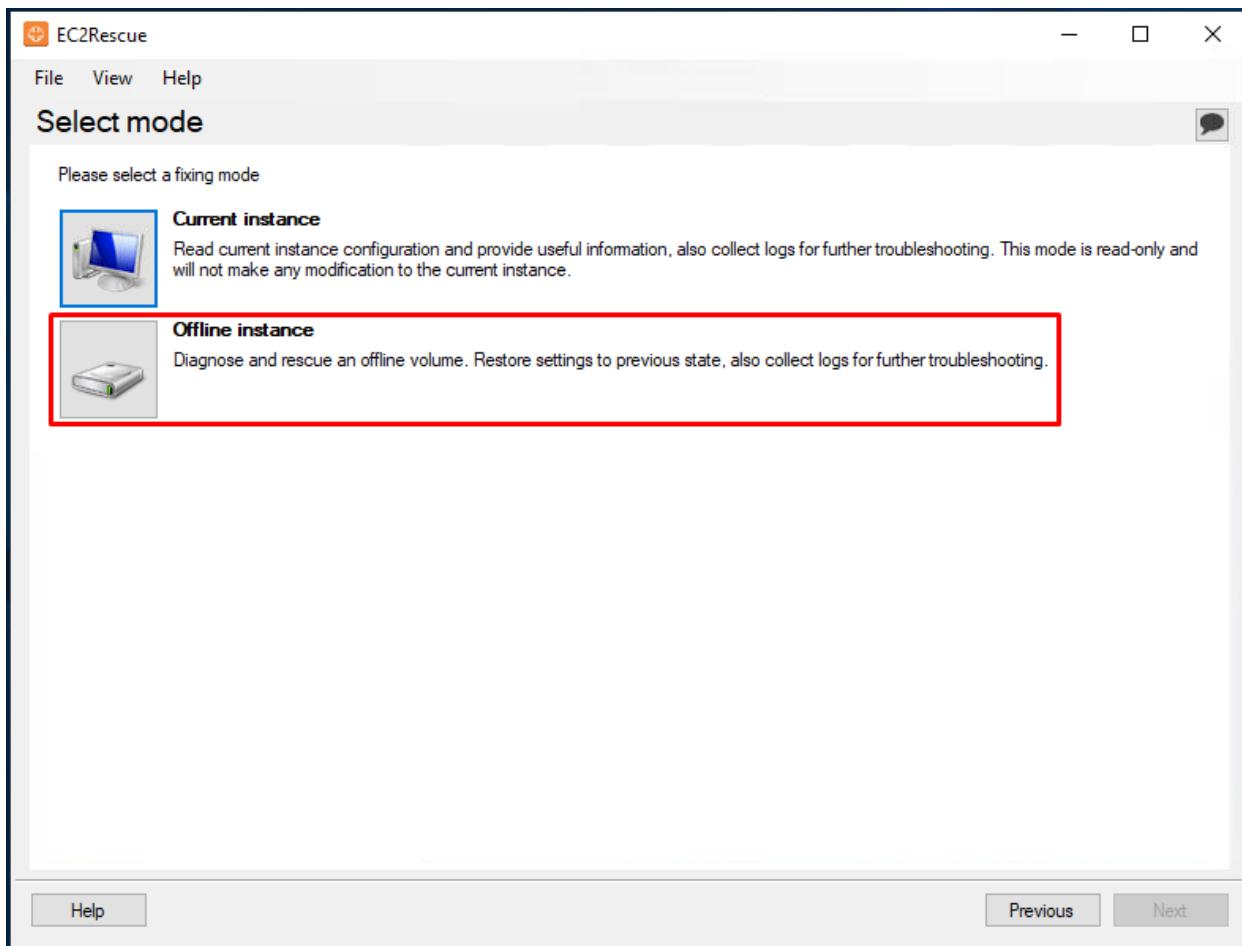


3. Click **I agree** on the license agreement and click **Next** on the Welcome screen.

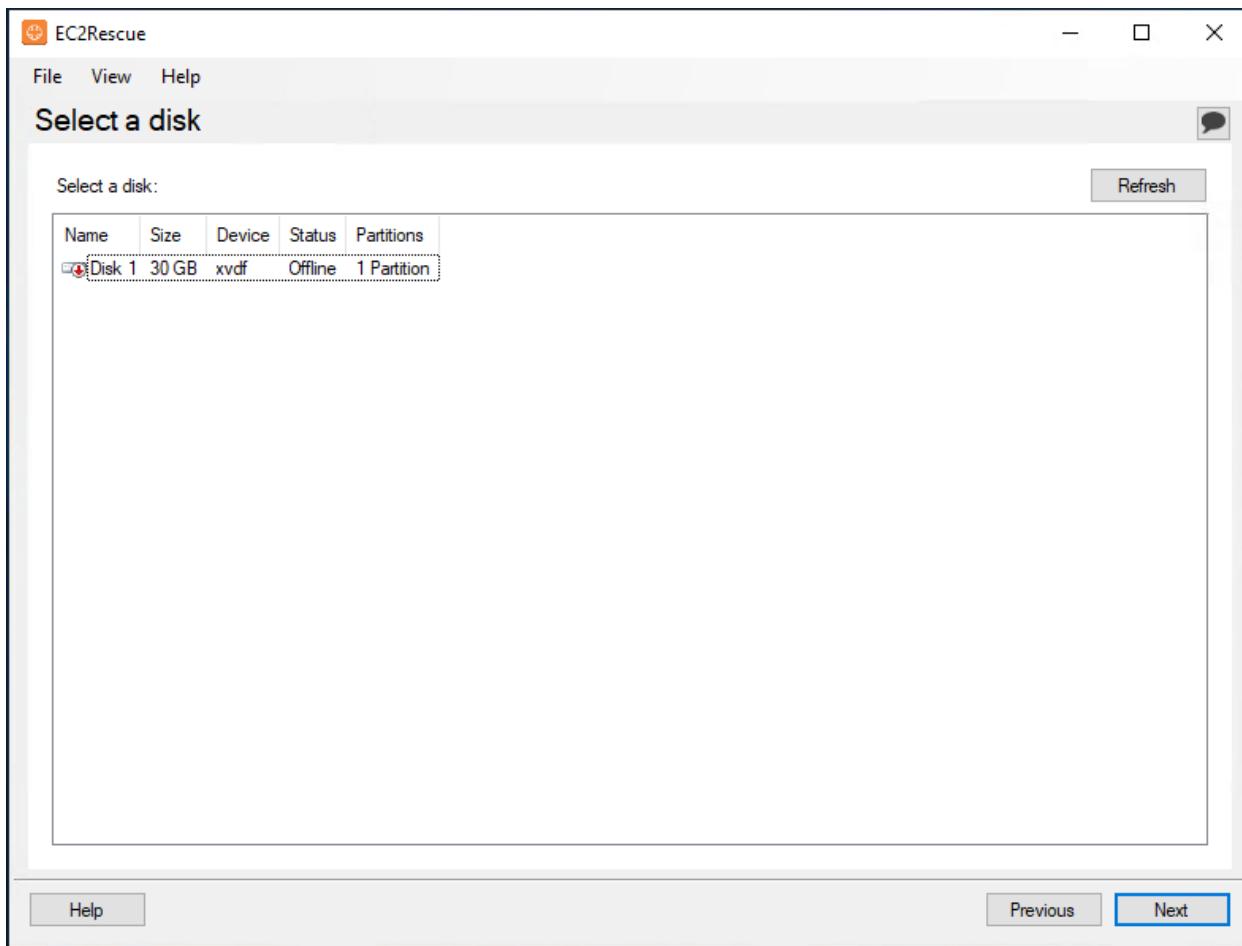




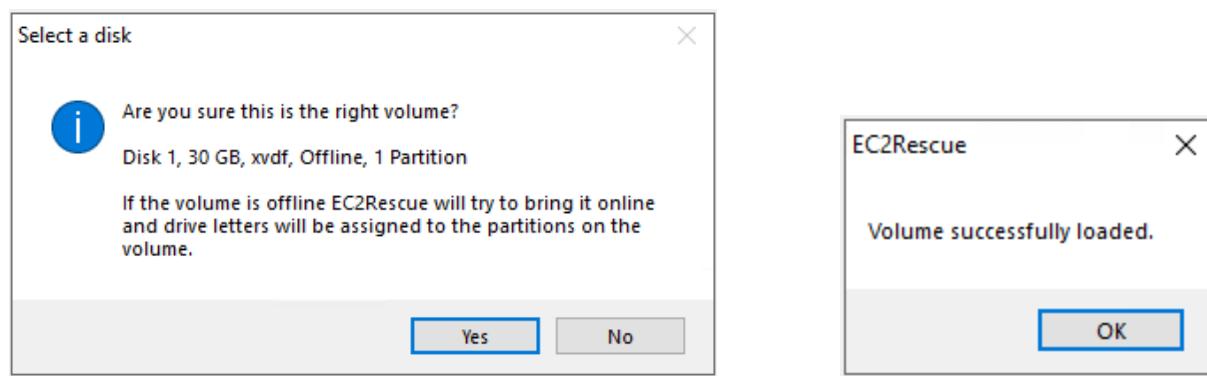
4. Select **Offline instance** and click **Next**.



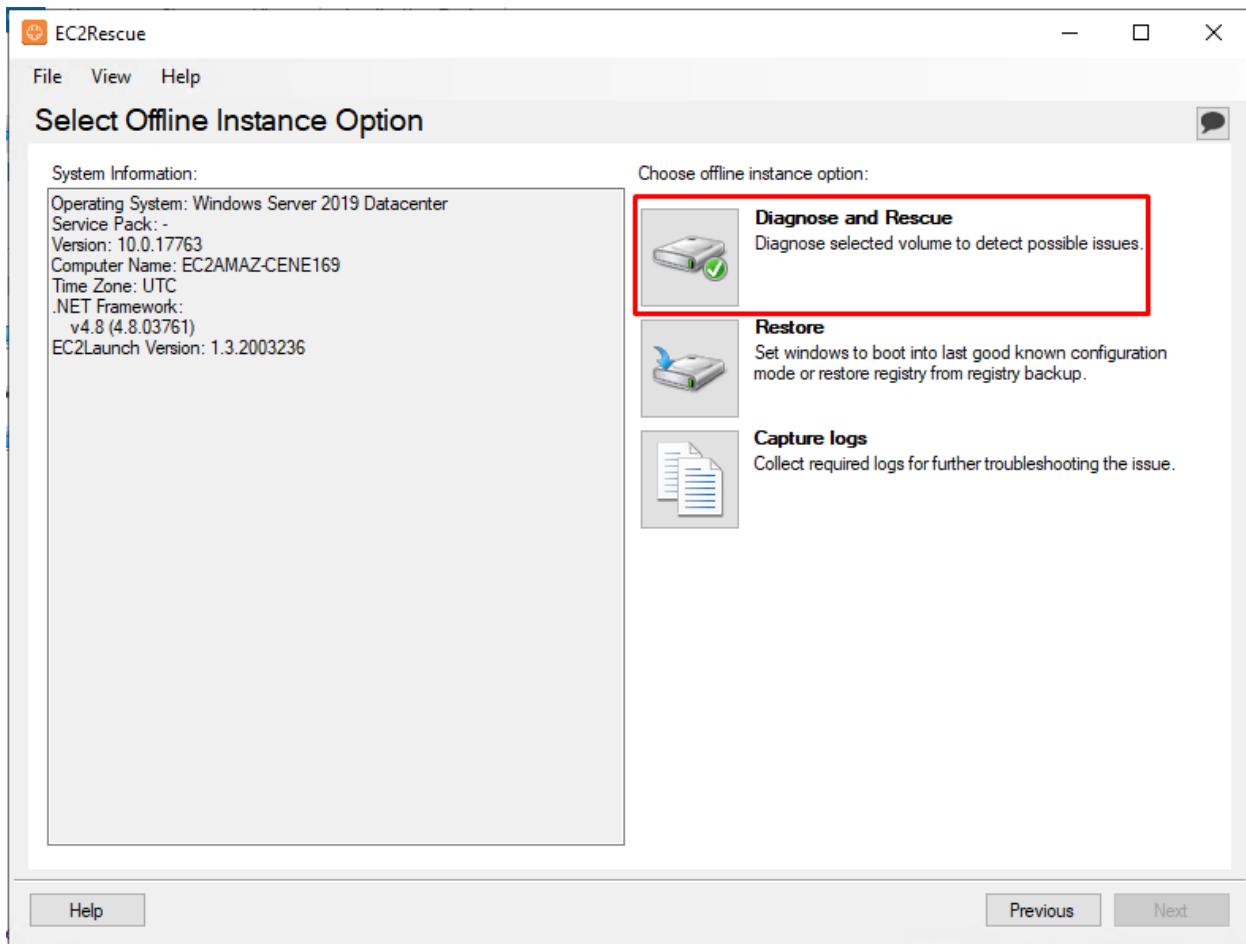
5. Select the **EBS volume** of the problematic instance and click **Next**. If you are checking multiple root volumes, note the device name when attaching the volumes to the EC2Rescue host.



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 necessary configurations, including their status and description. Click **Next** to proceed.

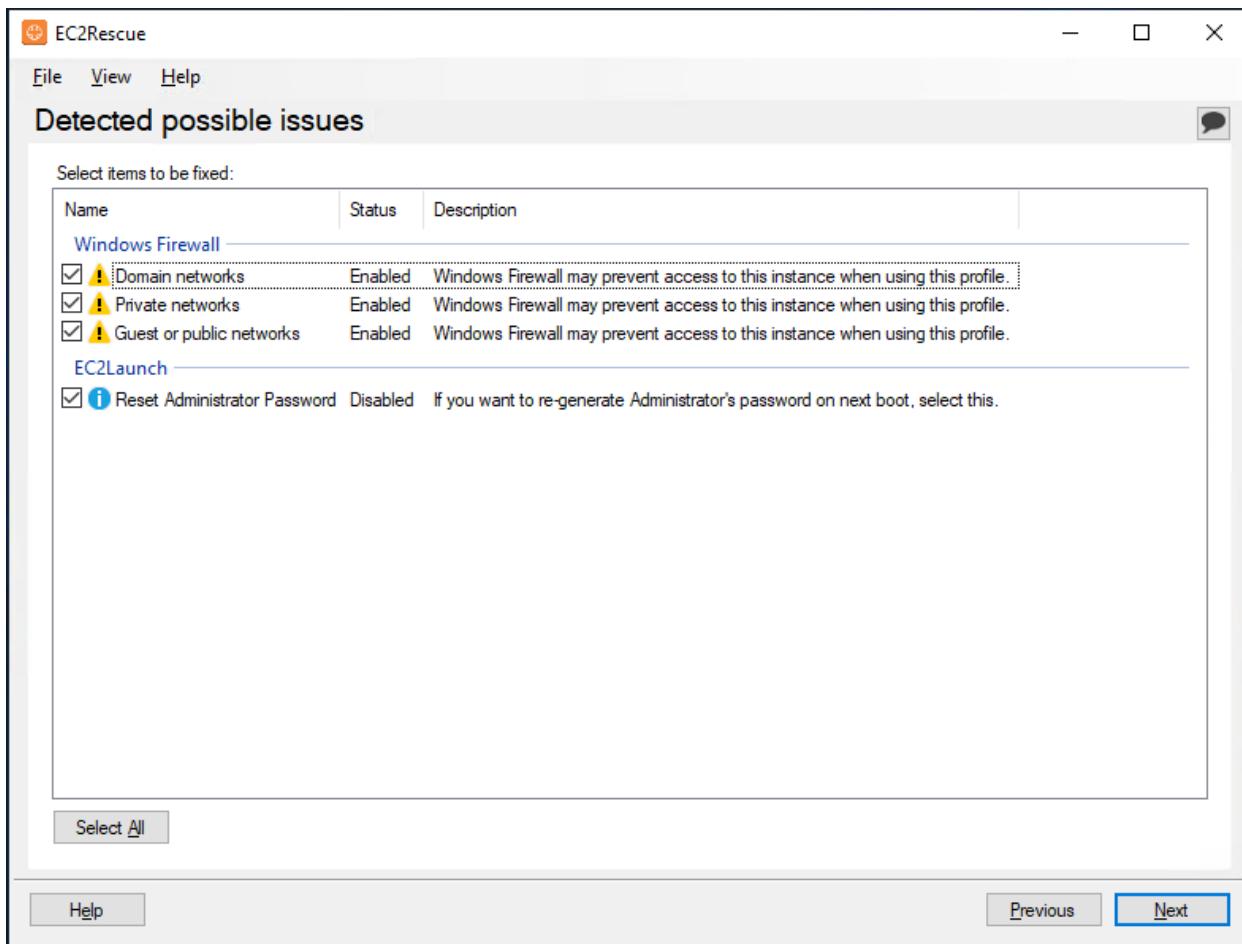


The screenshot shows the EC2Rescue application window titled "EC2Rescue". The menu bar includes "File", "View", and "Help". A "Summary" tab is selected, displaying a list of checked items. The table has columns for "Name", "Status", and "Description". The items are grouped under sections: "System Time", "Windows Firewall", "Remote Desktop", "EC2Launch", and "Network Interface".

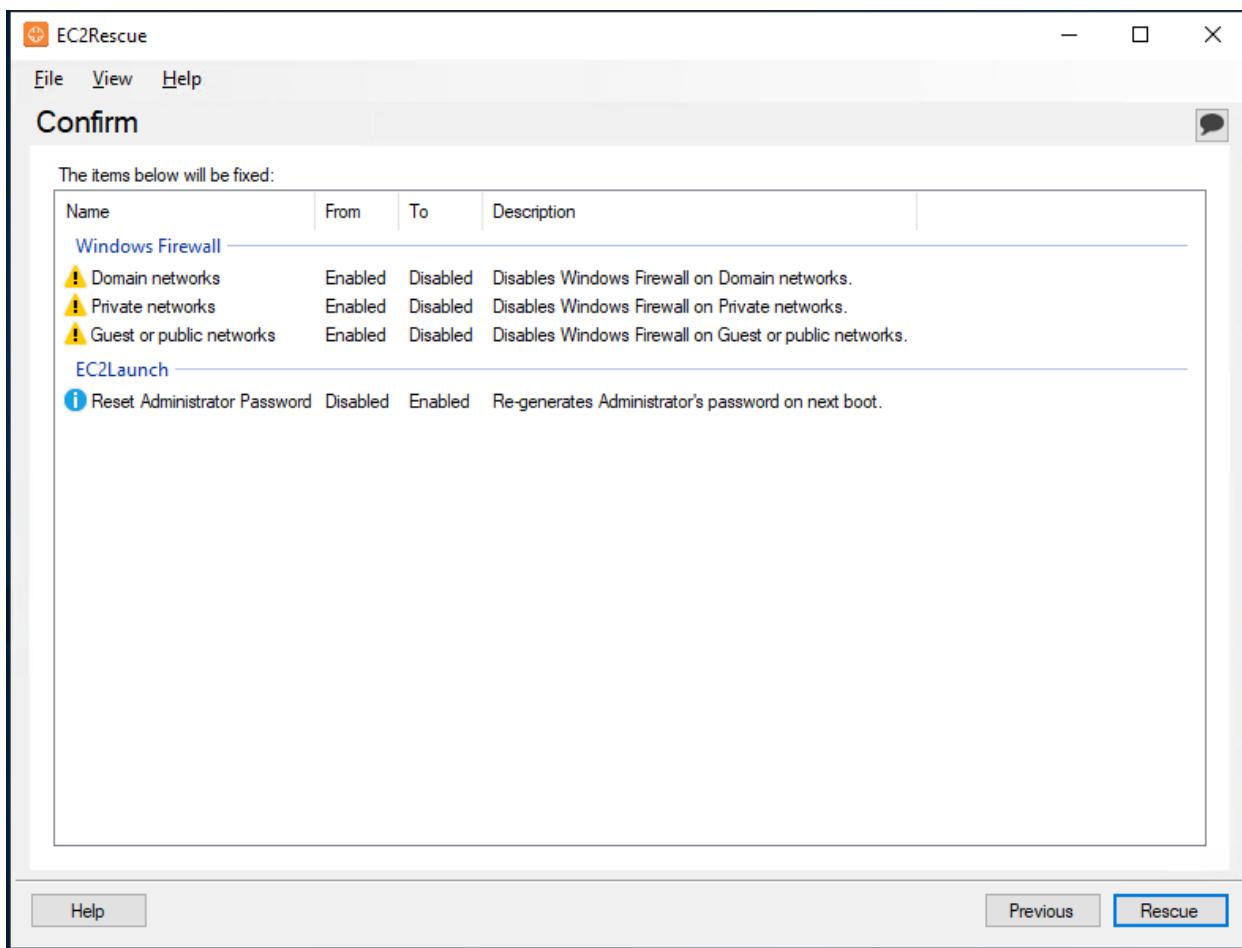
Name	Status	Description
System Time		
RealTimetsUniversal	Enabled	If this is disabled, Windows system time drifts when timezone is set other than UTC.
Windows Firewall		
Domain networks	Enabled	Windows Firewall may prevent access to this instance when using this profile.
Private networks	Enabled	Windows Firewall may prevent access to this instance when using this profile.
Guest or public networks	Enabled	Windows Firewall may prevent access to this instance when using this profile.
Remote Desktop		
Service Start	Manual	Remote Desktop service should start automatically.
Remote Desktop Connections	Enabled	Remote Desktop Connections should be enabled to allow access.
TCP Port	3389	The default listening port is TCP/3389.
EC2Launch		
Installation	Installed	EC2Launch 1.3.2003236 is installed.
Reset Administrator Password	Disabled	If you want to re-generate Administrator's password on next boot, select this.
Network Interface		
DHCP Service Startup	Automatic	The service should be started automatically.
Ethernet detail	N/A	{54B31D7E-36BF-4BBE-9AB2-106A939CD78C}, AWS PV Network Device, 8.2.5.32
DHCP on Ethernet	Enabled	IP address on the network interface must match with its private IP address. Enabling DHCP is recommended.
Ethernet 2 detail	N/A	{7AD95F04-05BD-4CDA-B380-4D44F3C0F356}, Intel(R) 82599 Virtual Function, 2.1.185.0
DHCP on Ethernet 2	Enabled	IP address on the network interface must match with its private IP address. Enabling DHCP is recommended.

At the bottom, there are "Help", "Previous", and "Next" buttons. The "Next" button is highlighted in blue.

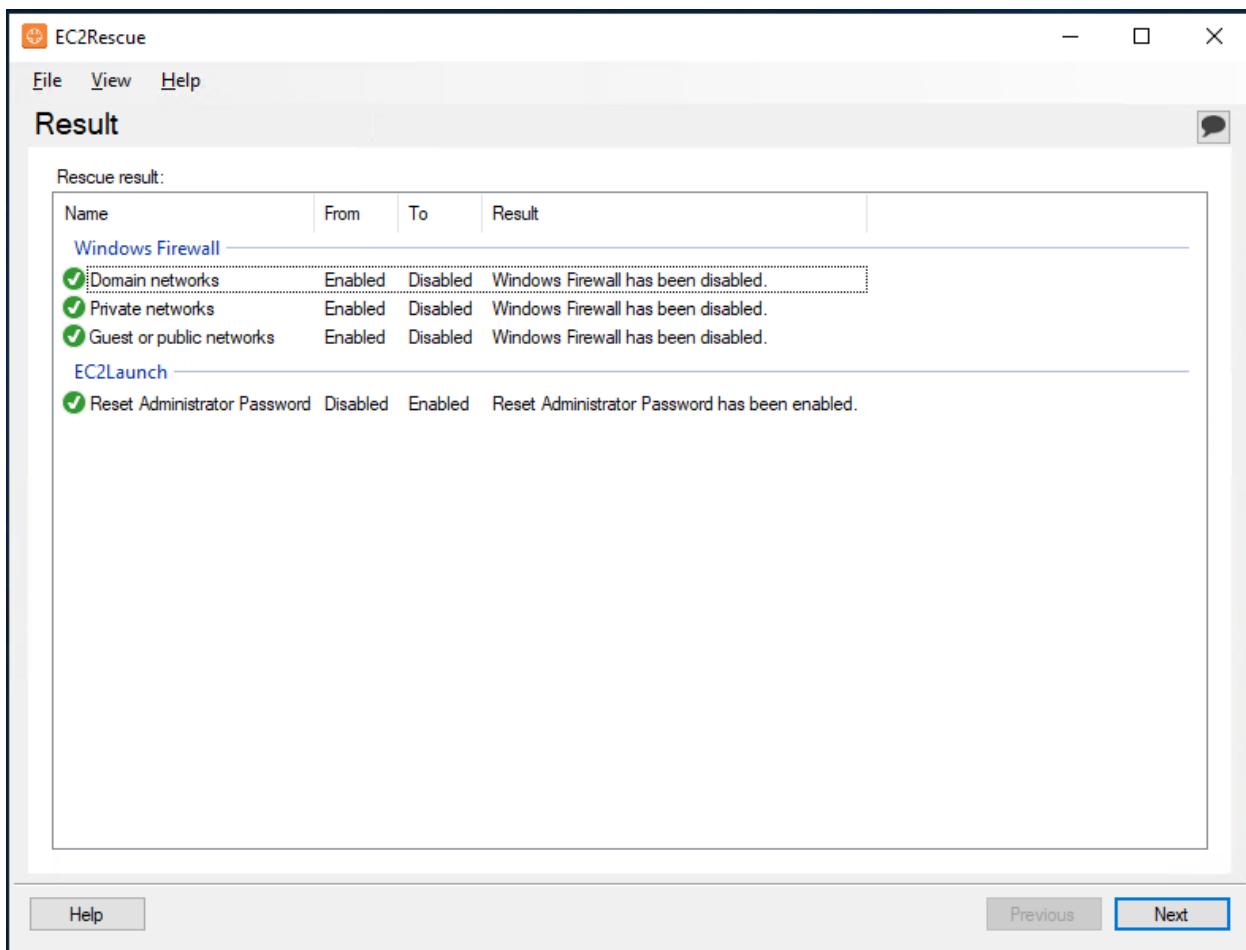
- EC2Rescue will give you a list of potential issues of the instance. From this, you can select the fixes you find necessary for your instance. Click **Next** to proceed.



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



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



12. Click **Finish**.

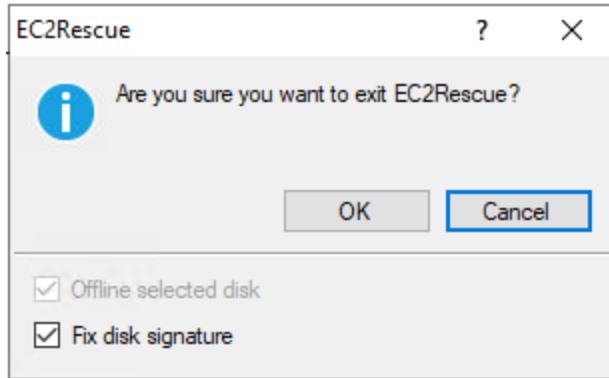


EC2Rescue can take the volume offline by clicking the [Finish] button or exiting the application. The volume can then be attached to the original instance as (/dev/sda1), after which the original instance can be started.

To view the audit log for this session, click on [View] > [Show log].

Help Previous Finish

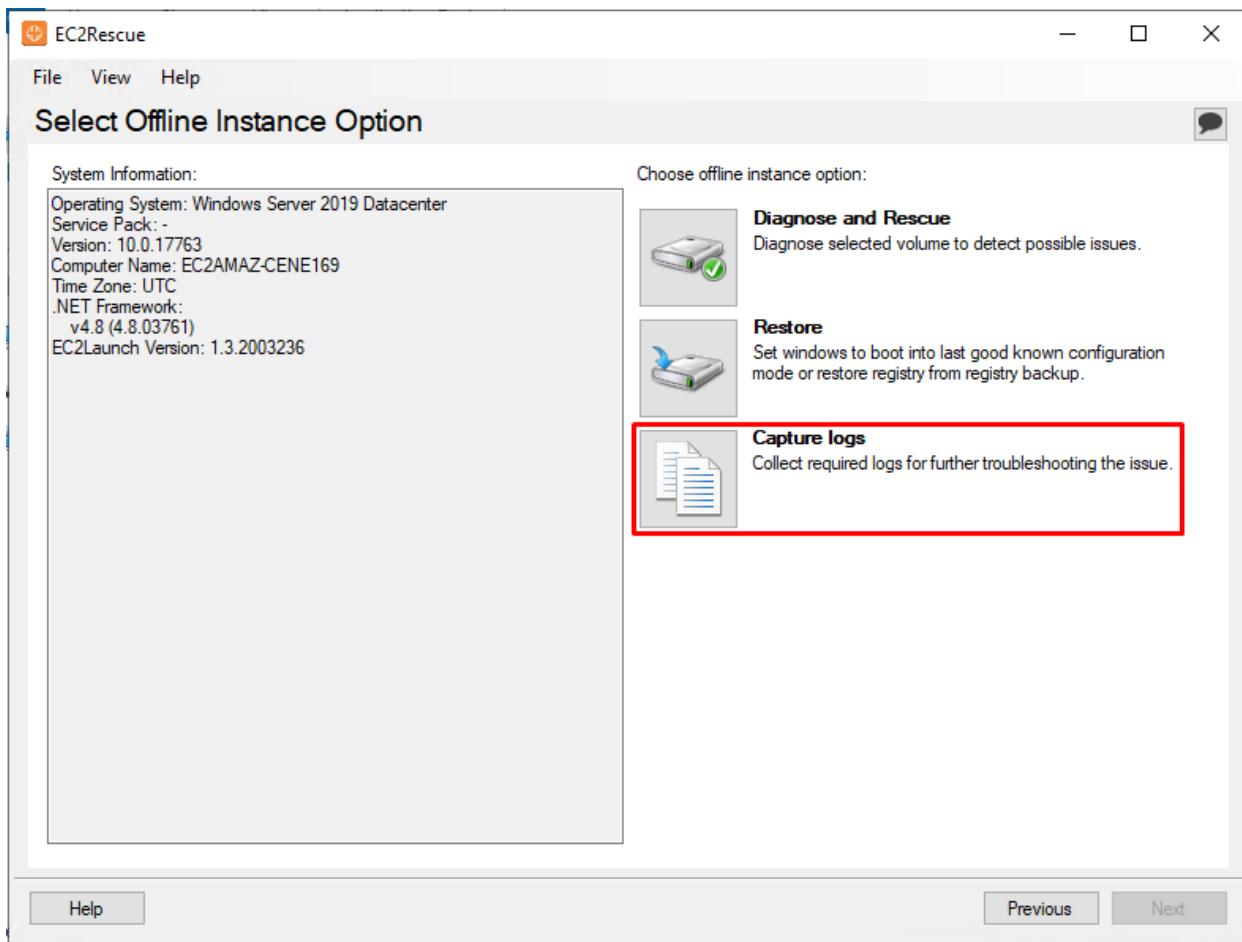
13. When exiting the EC2Rescue tool, check the **Fix disk signature** to resolve the boot issue caused by disk signature mismatch. The offline selected disk is also checked when the Fix disk signature is checked.



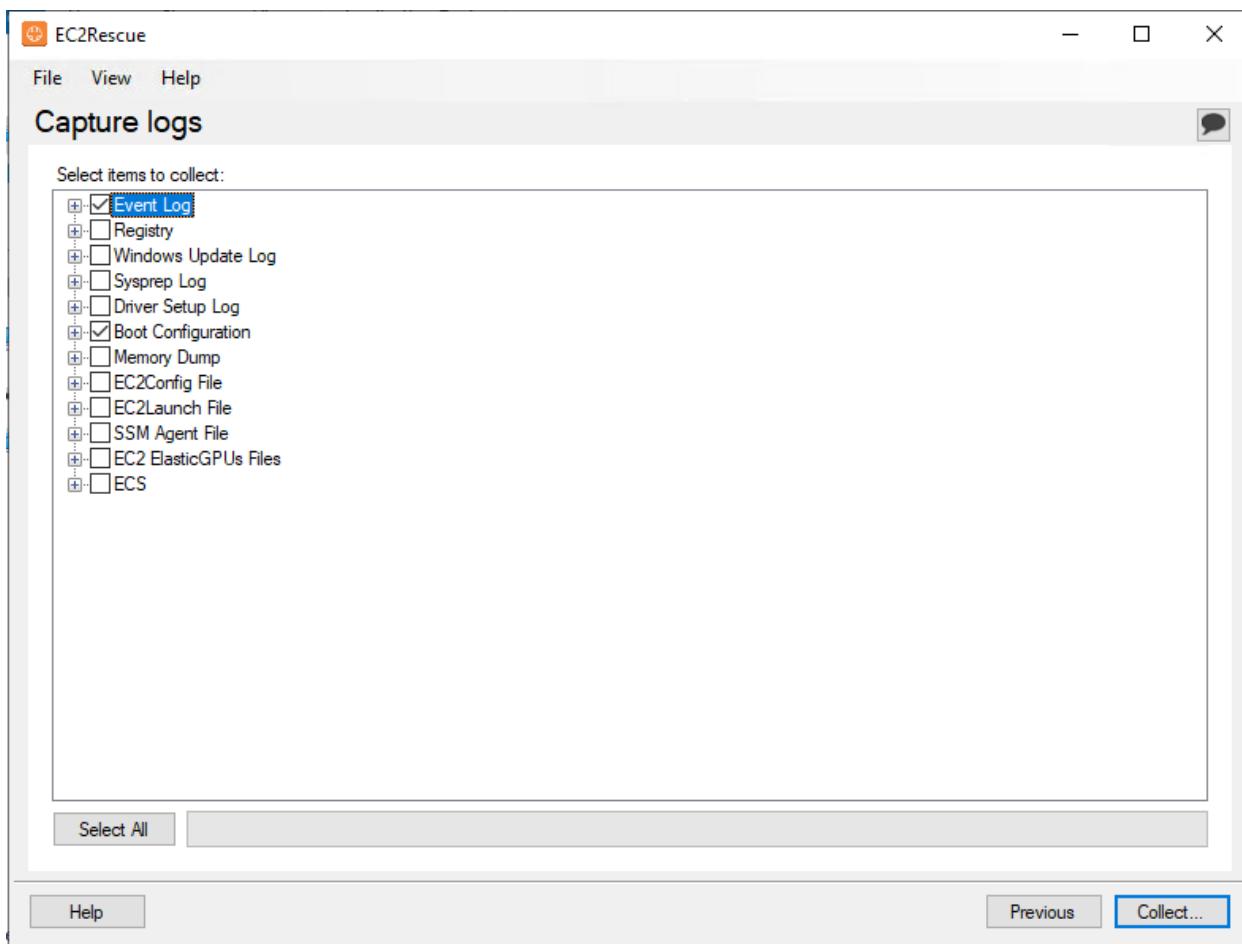
Collecting Logs from an Offline Instance

To collect logs from the instance, select Capture Logs from the offline instance option. The following instruction will guide you on how to collect logs from an offline instance.

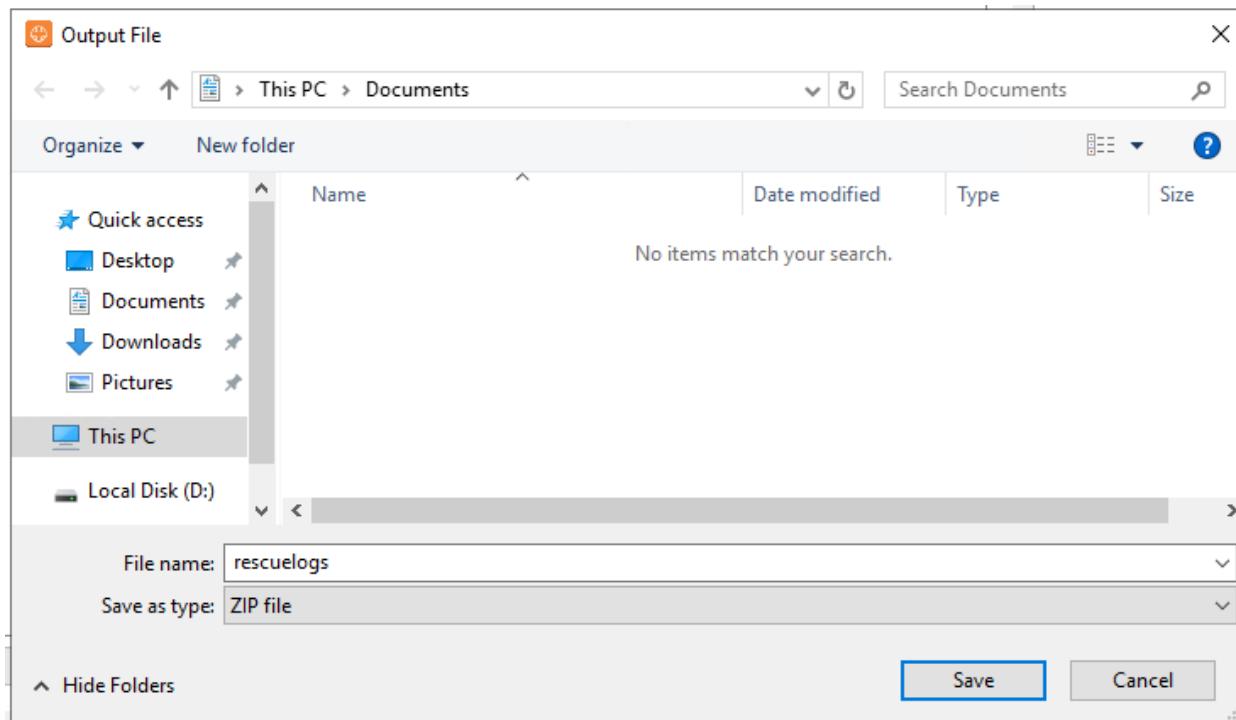
1. Select **Capture logs** from the offline instance option.



2. Select the logs you want to collect. Click **Collect** to proceed.



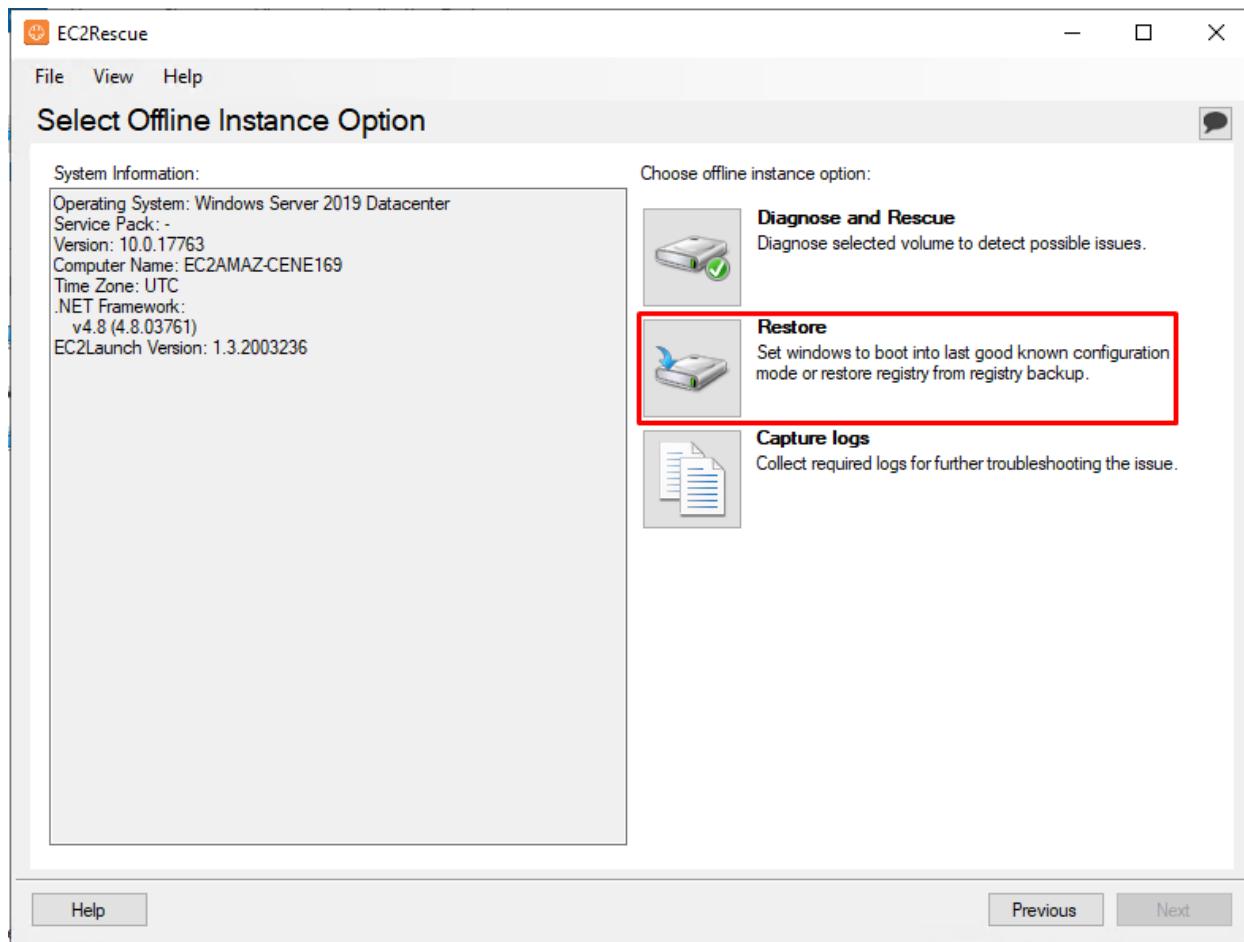
3. Save the logs as a zip file.



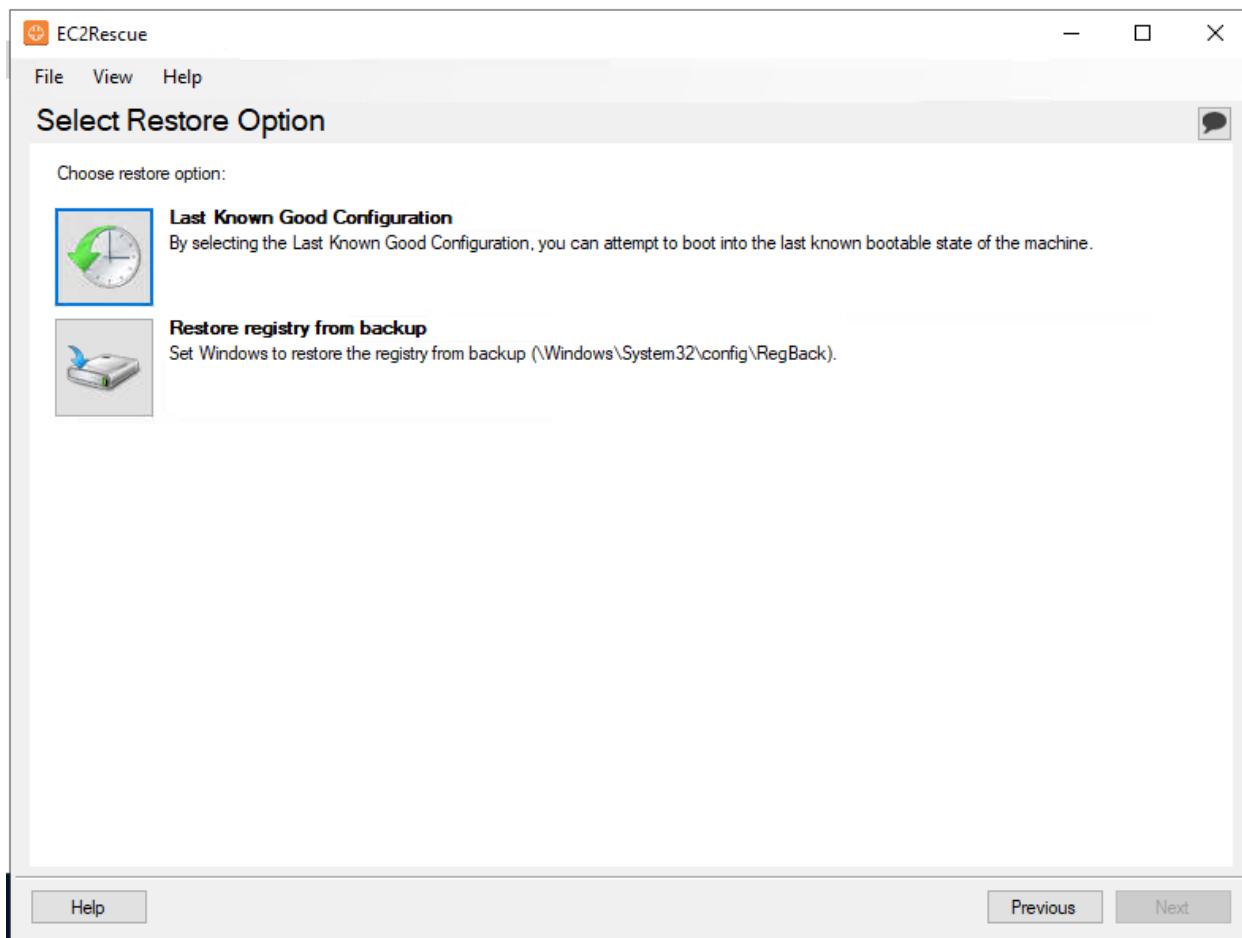
Restore Options for an Offline Instance

Using the EC2Rescue tool, you can also restore an instance. EC2Rescue provides two restore options: **Last Known Good Configuration** and **Restore registry from backup**.

You restore an offline instance by selecting Restore on the offline instance options.

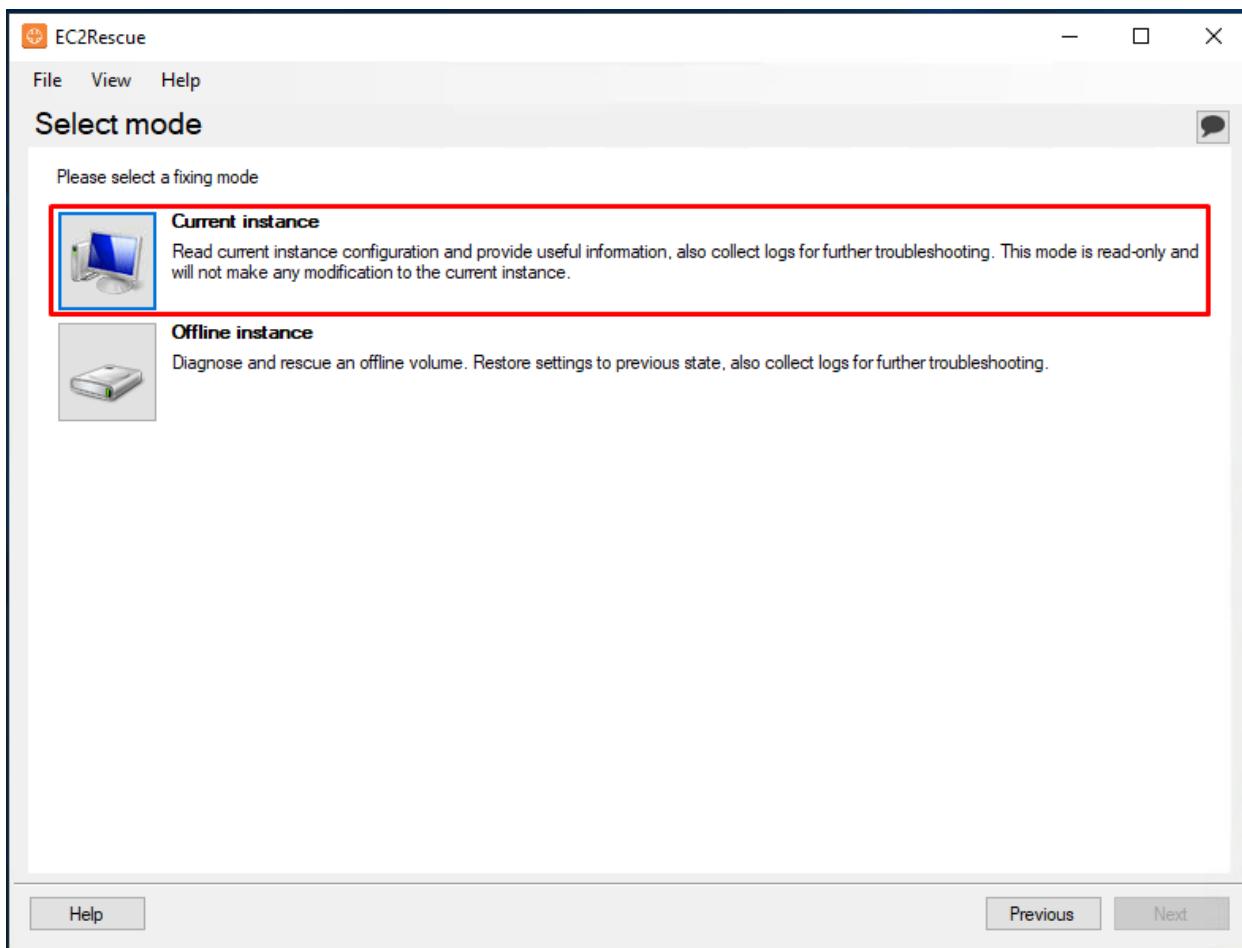


Then select a Restore option to begin the restoration.

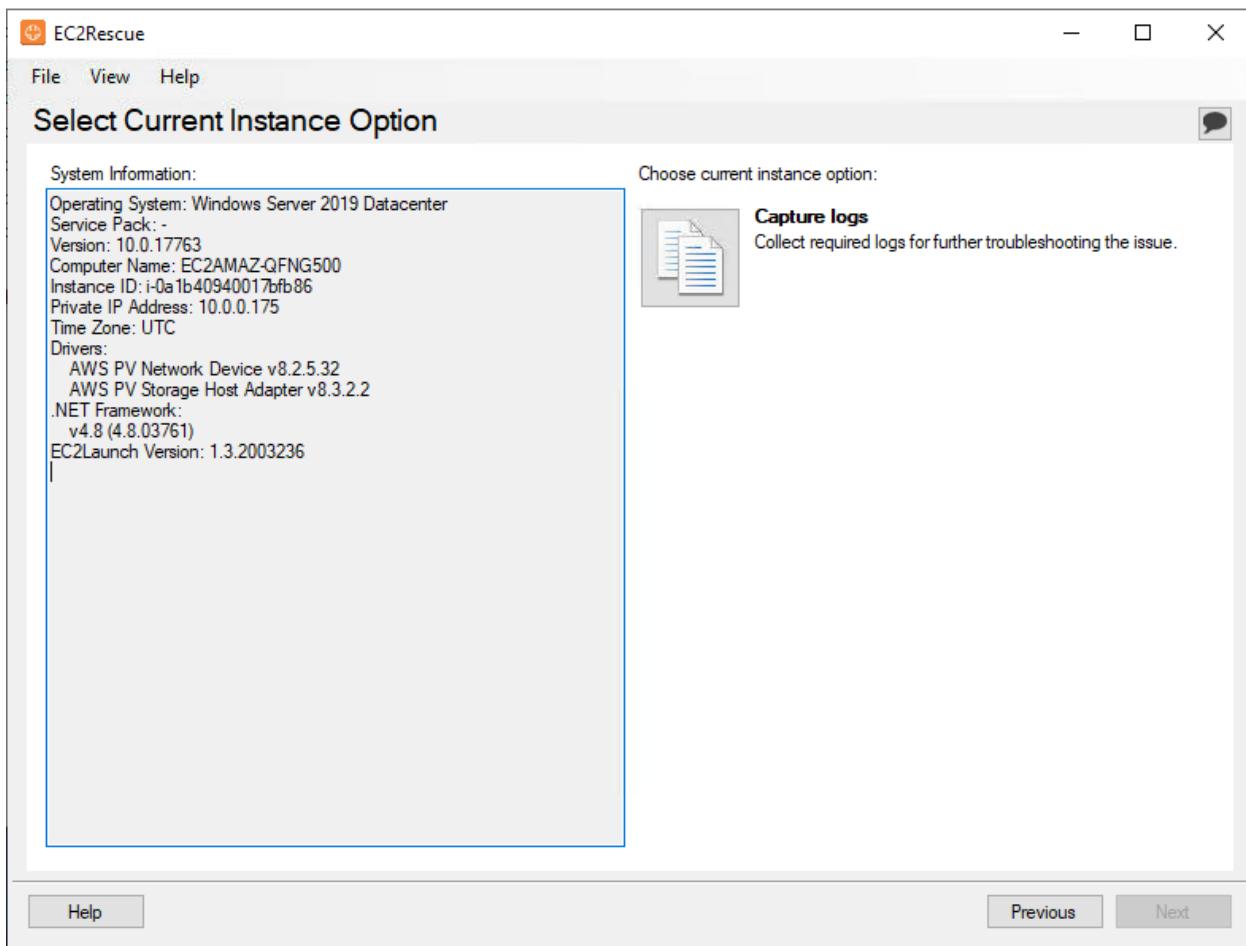


Checking the Current Instance

EC2Rescue also allows you to check the configuration and logs of its host instance. To check the current instance, select **Current Instance** mode.



The EC2Rescue tool only allows you to capture logs of the current instance, unlike the offline instance mode where you can run a diagnostic and restore.



EC2Rescue for Windows on Systems Manager

You can also use EC2Rescue using the Run command in AWS Systems Manager. The command document for EC2Rescue for Windows Server is named [AWSSupport-RunEC2RescueForWindowsTool](#). When you run this command, it will download and verify EC2Rescue for Windows Server and install a PowerShell module that allows you to run EC2RescueCmd. You can find this command document by simply typing EC2Rescue on the command search.



Run a command

Command document

Select the type of command that you want to run.

Search by keyword or filter by tag or attributes

Search: EC2Rescue X Clear filters

Name	Owner	Platform types
AWSsupport-RunEC2RescueForWindowsTool	Amazon	Windows

Given that your windows instance is added as Managed Instances on the Systems Manager, you will be able to run and do the following commands:

1. **ResetAccess** - resets the password of the local windows administrator and creates a new password in the Parameter Store. Requires KMS Key ID to encrypt new administrator password.
2. **CollectLogs** - collects and uploads valuable logs from the operating system to an S3 bucket.
3. **FixAll** - fix an offline root volume attached to the instance.

AWSsupport-RunEC2RescueForWindowsTool

Description Content Versions Details

Run command

▼ Parameters

Document version
14 (Default)

Command String (Required) Choose one of: ResetAccess - Resets the local Administrator password and stores the new password in Parameter Store | CollectLogs: Collects troubleshooting logs from the Operating System, and uploads them to an S3 bucket in your account | FixAll: Attempts to fix an offline Windows root volume attached to the current instance

Parameters String (Required) Parameters for the command: ResetAccess - KMS Key ID (not the alias) to encrypt the new Administrator password | CollectLogs: S3 bucket to upload the logs to | FixAll: Device name for the offline remediation.

EC2Rescue for Linux

You can diagnose and troubleshoot EC2 instances running in Linux using EC2Rescue for Linux. This tool has over 100 modules on its library ready to use in checking Linux instances. EC2Rescue for Linux supports the following Linux distributions.

- Amazon Linux 2
- Amazon Linux 2016.09+
- SUSE Linux Enterprise Server 12+
- RHEL 7+
- Ubuntu 16.04+

EC2Rescue for Linux also requires Python 2.7.9+ or 3.2+ installed.



Installing EC2Rescue for Linux

To use EC2Rescue for Linux, you need to download and install the tool in a working Linux machine which will serve as its host. You can download the EC2Rescue tool using the command below.

```
curl -O https://s3.amazonaws.com/ec2rescuelinux/ec2rl.tgz
```

You also need to download the sha56 hash file to verify the integrity.

```
curl -O https://s3.amazonaws.com/ec2rescuelinux/ec2rl.tgz.sha256
```

Verify the integrity of the tool tar archive.

```
sha256sum -c ec2rl.tgz.sha256
```

Extract the archive.

```
tar -xvf ec2rl.tgz
```

Run EC2Rescue help to view help and to verify the installation.

```
cd ec2rl-<version_number>
./ec2rl help
```

Diagnose Issues Using EC2Rescue for Linux

The EC2Rescue for Linux contains modules that you can use to diagnose and get valuable information about an instance. You can either run all the modules or any specific module you want. Some modules require root access so you may want to use an account with root access or run the command with *sudo*.

Run all modules.

```
./ec2rl run
```

Run a specific module.

```
./ec2rl run --only-modules=module_name --arguments
```

If you are unsure what module to use, you can list and find information about the module using the following command.



List all modules.

```
./ec2rl list

#get help for a module
./ec2rl help module_name
```

EC2Rescue generates a log file that you can view in the `/var/tmp/ec2rl` directory after running a module. You can either upload the log file to an S3 bucket or AWS Support.

Upload log file to S3 bucket.

```
./ec2rl upload --upload-directory=/var/tmp/ec2rl/2017-05-11T15_39_21.893145
--presigned-url="s3presignedurl"
```

Upload log file to AWS Support

```
./ec2rl upload --upload-directory=/var/tmp/ec2rl/2017-05-11T15_39_21.893145
--support-url="URLfromAWSsupport"
```

Creating Instance Backup Using EC2Rescue for Linux

With the EC2Rescue tool, you can create a backup for the Linux instance. Backup can be done as AMI or volume snapshot.

Backup using AMI

```
./ec2rl run --backup=ami
```

Backup all volume.

```
./ec2rl run --backup=allvolumes
```

Backup specific volume.

```
./ec2rl run --backup=volumeID
```

EC2Rescue for Linux on Systems Manager

Systems Manager also has automation documents that check Linux instances called `AWSsupport-TroubleshootSSH`. When you execute this automation, it will install EC2Rescue for Linux to check and troubleshoot the instance.



You can find this document under the Remediation category from the Automation in Systems Manager.

The screenshot shows the AWS Systems Manager Automation Execute interface. The left sidebar lists 'Document categories' including AWS Documentation, Remediation (selected), Patching, Security, Instance management, Data backup, AMI management, Self service support workflows, Resource management, and Cost management. The main area is titled 'Automation document' and contains a search bar. Below it are several automation document cards:

Document Name	Owner	Platform types
AWS-CreateJiraIssue	Owner Amazon	Platform types Windows, Linux
AWS-CreateServiceNowIncident	Owner Amazon	Platform types Windows, Linux
AWS-EnableCloudTrail	Owner Amazon	Platform types Windows, Linux
AWS-PublishSNSNotification	Owner Amazon	Platform types Windows, Linux
AWS-ReleaseElasticIP	Owner Amazon	Platform types Windows, Linux
AWSSupport-StartEC2RescueWorkflow	Owner Amazon	Platform types Windows, Linux, MacOS
AWSSupport-TroubleshootRDP	Owner Amazon	Platform types Windows, Linux
AWSSupport-TroubleshootSSH	Owner Amazon	Platform types Windows, Linux
AWSConfigRemediation-CancelKeyDeletion	Owner Amazon	Platform types Windows, Linux, MacOS
AWSConfigRemediation-ConfigureCodeBuildProjectWithKMSCK	Owner Amazon	Platform types Windows, Linux, MacOS
AWSConfigRemediation-ConfigureLambdaFunctionXRayTracing	Owner Amazon	Platform types Windows, Linux
AWSConfigRemediation-ConfigureS3BucketPublicAccessBlock	Owner Amazon	Platform types Windows, Linux

The 'AWSSupport-TroubleshootSSH' document is highlighted with a red border and a blue selection dot.

References:

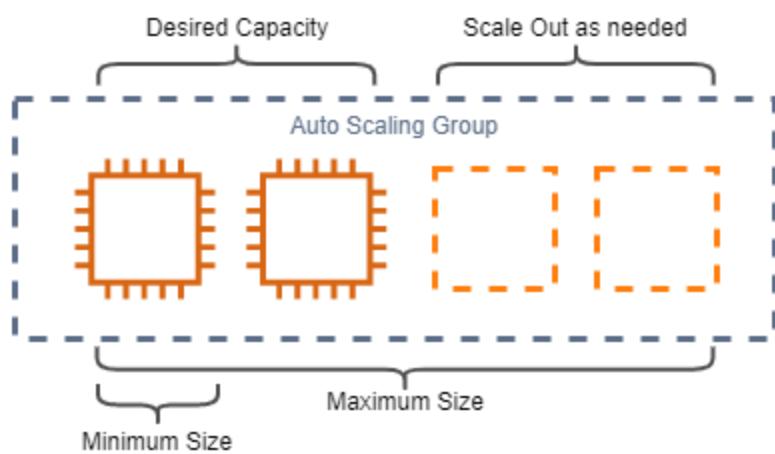
<https://docs.aws.amazon.com/AWSEC2/latest/WindowsGuide/Windows-Server-EC2Rescue.html>

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/Linux-Server-EC2Rescue.html>

AWS Auto Scaling

Auto Scaling Group

Auto Scaling allows AWS resources to scale up and down quickly to handle your application load. On EC2, you can create an Auto Scaling Group to provision a fleet of instances. You can specify its minimum, maximum, and desired number of instances according to your requirements.



EC2 Auto Scaling offers different scaling options that include manual, dynamic, and scheduled scaling. As the word implies, manual scaling allows you to manually add and remove instances from the Auto Scaling Group. Dynamic scaling enables you to set a specific condition to trigger autoscaling. Scheduled scaling scales your Auto Scaling Group at a scheduled time. During scale-out, additional instances spawn while instances are terminated during scale-in.

Auto Scaling Templates

To create an Auto Scaling Group, you need to configure Launch Templates or Launch Configurations. These templates contain the necessary instance parameters for the Auto Scaling Group, like Amazon machine image (AMI) and Instance Type.

Launch Templates

Launch Templates are reusable templates that contain parameters needed to provision EC2 instances. When creating Launch Templates, defining only the common parameters like template name, AMI, and networking platform is recommended.

You may want to enable Auto Scaling guidance to guide you on the template creation. If enabled, the Amazon machine image (AMI) parameter is required.



Launch template name and description

Launch template name - *required*

MyTemplate

Must be unique to this account. Max 128 chars. No spaces or special characters like '&', '*', '@'.

Template version description

A prod webserver for MyApp

Max 255 chars

Auto Scaling guidance [Info](#)

Select this if you intend to use this template with EC2 Auto Scaling

Provide guidance to help me set up a template that I can use with EC2 Auto Scaling

► [Template tags](#)

► [Source template](#)

You can already define the Instance Type, but this can be changed or overridden during the Auto Scaling Group configuration. The key pair is the security credentials that will be used when connecting to the instance.

▼ Amazon machine image (AMI) - *required* [Info](#)

AMI - *required*

amzn2-ami-hvm-2.0.20210326.0-x86_64-gp2
ami-0742b4e67307206f
architecture: 64-bit (x86) virtualization: hvm

▼ Instance type [Info](#)

Instance type

t2.micro	Free tier eligible
Family: t2	1 vCPU
1 GiB Memory	

[Compare instance types](#)

▼ Key pair (login) [Info](#)

You can use a key pair to securely connect to your instance. Ensure that you have access to the selected key pair before you launch the instance.

Key pair name

webapp	Template value
--------	----------------

[Create new key pair](#)



You can select multiple security groups as well. Note that when no security group is assigned, instances will use the default security group. You can only leave the security group empty if you intend to configure this using the network interface.

The screenshot shows the 'Network settings' section of a CloudFormation template. It includes options for 'Networking platform' (Virtual Private Cloud (VPC) selected), 'Security groups' (MyEc2DMZ sg-2914b459 selected), and a 'Select security groups' dropdown.

Volume is automatically created and will depend on the selected AMI; you can optionally create additional volume and add tags.

The screenshot shows the 'Storage (volumes)' section with a single volume entry and an 'Add new volume' button. Below it is the 'Resource tags' section with a note about currently empty tags and an 'Add tag' button.



You can skip adding a network interface if you want to keep the default network interface. If you don't want to use auto-assignment of IP addresses to your instances and assign specific IP addresses, you may want to change the default network interface. See further instructions [here](#).

The screenshot shows the 'Network interfaces' section with a note: 'No network interfaces are currently included in this template. Add a network interface to include it in the launch template.' A 'Add network interface' button is present. Below it is the 'Advanced details' section. At the bottom right are 'Cancel' and 'Create template version' buttons.

Don't forget to review the parameters before you proceed with the creation of the template.

You can keep multiple template versions when modifying an existing launch template. You can also specify the default version.

The screenshot shows the 'Launch template details' page for 'WebappLT'. It displays basic information like Launch template ID, name, version, and owner. Below is the 'Launch template version details' section. The 'Version' dropdown menu is highlighted with a red box, showing options '1 (Default)', '2', and '1 (Default)'. Other fields include Description, Date created, Created by, Instance details (AMI ID, Instance type, Availability Zone, Key pair name), Storage, Resource tags, Network interfaces, and Advanced details.



Launch Configurations

Launch configurations work like Launch Templates. It also holds configuration necessary for instance provisioning. Launch configurations are also reusable and can be used to create multiple Auto Scaling Groups.

AWS does not recommend using Launch Configuration as this doesn't give full functionality for EC2 instances and EC2 Auto Scaling. Launch Configuration also offers limited AMIs compared to Launch Templates. In addition, there is no versioning on Launch Configuration. Once a launch configuration is created, you cannot modify it. Moreover, an Auto Scaling Group can only be associated with a single Launch Configuration. It means that to update a launch configuration for an existing Auto Scaling Group, you need to create a new launch configuration and configure the Auto Scaling Group to use it.

Auto Scaling Group Configuration

As we mentioned above, you can create an Auto Scaling Group using Launch Templates (AWS recommended) and Launch Configuration. Instances in Auto Scaling Group can be On-Demand, Spot instances, or both. You can specify the template version when you use Launch Template.

The screenshot shows the AWS Lambda function configuration interface. At the top, there's a header with 'Launch template' and a 'Info' link, along with a 'Switch to launch configuration' link. Below the header, there's a section for 'Launch template' with a description: 'Choose a launch template that contains the instance-level settings, such as the Amazon Machine Image (AMI), instance type, key pair, and security groups.' A dropdown menu is open, showing 'WebappLT' and a 'Create a launch template' link. A red box highlights the dropdown menu. Below the dropdown, there's a 'Version' section with a dropdown menu set to 'Default (1)' and a 'Create' button. Another 'Create a launch template version' link is present. The main configuration area includes sections for 'Description' (Webapp Launch Template), 'AMI ID' (ami-0742b4e673072066f), 'Key pair name' (webapp), 'Launch template' (WebappLT), 'Security groups' (-), and 'Instance type' (t2.micro). There's also a 'Request Spot Instances' section set to 'No'.

Launch templates allow you to combine purchase options and instance types.



Instance purchase options Info

Use the launch template to create a uniform configuration among all of the instances in the group. Or define options to accommodate a wide variety of requirements, such as launching Spot and On-Demand Instances.

Adhere to launch template

The launch template determines the purchase option (On-Demand or Spot) and instance type.

Combine purchase options and instance types

Specify how much On-Demand and Spot capacity to launch and multiple instance types (optional). This choice is most helpful for optimizing the scale and cost for a fleet of instances.

You can maximize the scaling and cost savings by setting the percentage split for the On-Demand and Spot instances. You can also set the Spot allocation strategy and configure Capacity optimized Spot settings.

Instances distribution

On-Demand base capacity - *optional*

Specify how much On-Demand capacity the Auto Scaling group should have for its base portion. The maximum group size will be increased (but not decreased) to this value.

0

On-Demand Instances

On-Demand percentage above base

Define the percentage split of On-Demand Instances and Spot Instances for your additional capacity beyond the base portion.

70

% On-Demand

30

% Spot

Spot allocation strategy per Availability Zone

Capacity optimized (recommended)

Launch Spot Instances optimally based on the available Spot capacity.

Lowest price

Launch Spot Instances from the lowest priced instance pools.

Capacity optimized Spot settings

Prioritize instance types Info

You set the priority order for your instance types, and EC2 attempts to fulfill Spot capacity based on these priorities while still optimizing for capacity.

Capacity rebalance Info

When you enable capacity rebalancing, and a rebalance notification is sent to an instance, EC2 Auto Scaling automatically attempts to replace the instance before it is interrupted.

You can still select the instance type for your instances. If you do, the instance type on the launch template is overridden. By default, instances are equally weighted. To use instance weighting for your instances, use the



units like vCPU or memory to determine the weight. The weight determines the number of units that the instance type represents towards the desired capacity.

Instance types [Info](#)

Choose the instance types that best suit the needs of your application.

Primary instance type [Weight](#) [Info](#)

1. t2.small 1vCPU 2 Gib Memory	1	^	^	X
-----------------------------------	---	-------------------	-------------------	-------------------

Additional instance types

[Redo recommendations](#)

2. t2.medium 2vCPU 4 Gib Memory	2	^	^	X
------------------------------------	---	-------------------	-------------------	-------------------

[Add instance type](#)

Set the VPC and Subnet for the Auto Scaling Group. Make sure that the security group set on the Launch Template resides within the VPC that you selected.

Network [Info](#)

For most applications, you can use multiple Availability Zones and let EC2 Auto Scaling balance your instances across the zones. The default VPC and default subnets are suitable for getting started quickly.

VPC

vpc-03c57950e44f8783b (vpc_webapp) 10.0.0.0/16	C
---	-------------------

[Create a VPC](#)

Subnets

Select subnets	C
----------------	-------------------

us-east-1c subnet-0af258dc9096781ad (Public subnet) 10.0.0.0/24	X
--	-------------------

[Create a subnet](#)

[Cancel](#) [Previous](#) [Skip to review](#) **Next**



You have an option to add a Load Balancer on top of your Auto Scaling Group. By default, EC2 health checks are enabled, but you can add an ELB health check as well.

Configure advanced options Info

Choose a load balancer to distribute incoming traffic for your application across instances to make it more reliable and easily scalable. You can also set options that give you more control over health check replacements and monitoring.

Load balancing - optional Info

Use the options below to attach your Auto Scaling group to an existing load balancer, or to a new load balancer that you define.

No load balancer

Traffic to your Auto Scaling group will not be fronted by a load balancer.

Attach to an existing load balancer

Choose from your existing load balancers.

Attach to a new load balancer

Quickly create a basic load balancer to attach to your Auto Scaling group.

Health checks - optional

Health check type Info

EC2 Auto Scaling automatically replaces instances that fail health checks. If you enabled load balancing, you can enable ELB health checks in addition to the EC2 health checks that are always enabled.

EC2 ELB

Health check grace period

The amount of time until EC2 Auto Scaling performs the first health check on new instances after they are put into service.

300 seconds

Additional settings - optional

Monitoring Info

Enable group metrics collection within CloudWatch

Cancel

Previous

Skip to review

Next

It is recommended to enable group metrics collection within CloudWatch to monitor your Auto Scaling Group. We will further discuss monitoring in the latter part of this eBook.



You can set the desired, minimum, and maximum number of instances for your Auto Scaling Group. When the Auto Scaling Group is created, it will spawn a number of instances equal to its desired capacity. If there are no scaling policies defined, the desired capacity will be maintained and will go through periodic health checks. Unhealthy instances will be terminated and will be replaced by new instances.

Group size - optional Info

Specify the size of the Auto Scaling group by changing the desired capacity. You can also specify minimum and maximum capacity limits. Your desired capacity must be within the limit range.

Desired capacity

Minimum capacity

Maximum capacity

Scaling policies will determine if an Auto Scaling Group will scale out or scale in. You can use different instance metrics to define the scaling policy. For example, if an instance reached an average CPU utilization of 80%, a new instance will be spawned. And once it drops below 80%, the instance will be terminated. A cooldown will be applied during a scaling process to ensure that the previous scaling has been completely done before starting another scale process. You may want to add a warmup time for your instance to exclude their metric since instances tend to have a high utilization when they are launched.



Scaling policies - optional

Choose whether to use a scaling policy to dynamically resize your Auto Scaling group to meet changes in demand. [Info](#)

Target tracking scaling policy
Choose a desired outcome and leave it to the scaling policy to add and remove capacity as needed to achieve that outcome.

None

Scaling policy name: CPUUtilPolicy

Metric type: Average CPU utilization

Target value: 80

Instances need: 120 seconds warm up before including in metric

Disable scale in to create only a scale-out policy

Lastly, you can also set up an SNS integration if you want to get notified of the launch and termination of instances.

Add notifications [Info](#)

Send notifications to SNS topics whenever Amazon EC2 Auto Scaling launches or terminates the EC2 instances in your Auto Scaling group.

You can optionally add tags as well. Tags will be applied to all of the instances inside your Auto Scaling Group.



Add tags Info

Add tags to help you search, filter, and track your Auto Scaling group across AWS. You can also choose to automatically add these tags to instances when they are launched.

i You can optionally choose to add tags to instances (and their attached EBS volumes) by specifying tags in your launch template. We recommend caution, however, because the tag values for instances from your launch template will be overridden if there are any duplicate keys specified for the Auto Scaling group. X

Tags (0)

Add tag

50 remaining

Cancel

Previous

Next

After reviewing all of the group configurations, you can now proceed with creating the Auto Scaling Group.

Cancel

Create Auto Scaling group

Now you can view vital information like scaling policy, instance management, and monitoring inside your Auto Scaling Group.

The screenshot shows the AWS EC2 Auto Scaling groups page. The navigation bar at the top includes links for EC2, Auto Scaling groups, and WebappASG. Below the navigation, there are tabs for Details, Activity, Automatic scaling, Instance management, Monitoring, and Instance refresh. The Details tab is selected. The main content area displays 'Group details' for the WebappASG group. It lists the following configuration:

Group details	
Desired capacity	1
Minimum capacity	1
Maximum capacity	3
Auto Scaling group name	WebappASG
Date created	Thu Apr 08 2021 01:16:22 GMT+0800 (Philippine Standard Time)
Amazon Resource Name (ARN)	arn:aws:autoscaling:us-east-1:947117271373:autoScalingGroup:43028311-5084-42c8-bb95-71e80395ea9c:autoScalingGroupName/WebappASG

If you expect a significant traffic change towards your Auto Scaling Group, you can set a Scheduled Action under the Automatic scaling tab.

Reference:

<https://docs.aws.amazon.com/autoscaling/ec2/userguide/what-is-amazon-ec2-auto-scaling.html>



AWS Compute Optimizer

AWS Compute Optimizer is a free tool that generates recommendations on cost-savings and performance improvement according to your workloads. It checks the current configuration of your AWS resources, analyzes the historical data from CloudWatch metrics to identify your usage patterns, gives a calculated projection, and provides recommendations. AWS Compute Optimizer leverages the capabilities of AI/ML-based analytics to provide optimal recommendations.

Avoid commonly encountered provisioning issues with AWS Compute Optimizer

- **Over-provisioned Resources** = Unnecessary infrastructure cost
- **Under-provisioned Resources** = Service/Performance degradation

AWS Compute Optimizer supports the following services.

- Amazon Elastic Compute Cloud (Amazon EC2) instances
- Amazon EC2 Auto Scaling groups
- Amazon Elastic Block Store (Amazon EBS) volumes
- AWS Lambda functions

Prerequisites

Only AWS resources with a minimum of **30 consecutive hours of CloudWatch metric** data should receive recommendations from AWS Compute Optimizer, except for Lambda functions which don't need a metric requirement. Once a resource meets the CloudWatch metrics requirement, AWS Compute Optimizer will begin analyzing data which may take up to 12 hours.

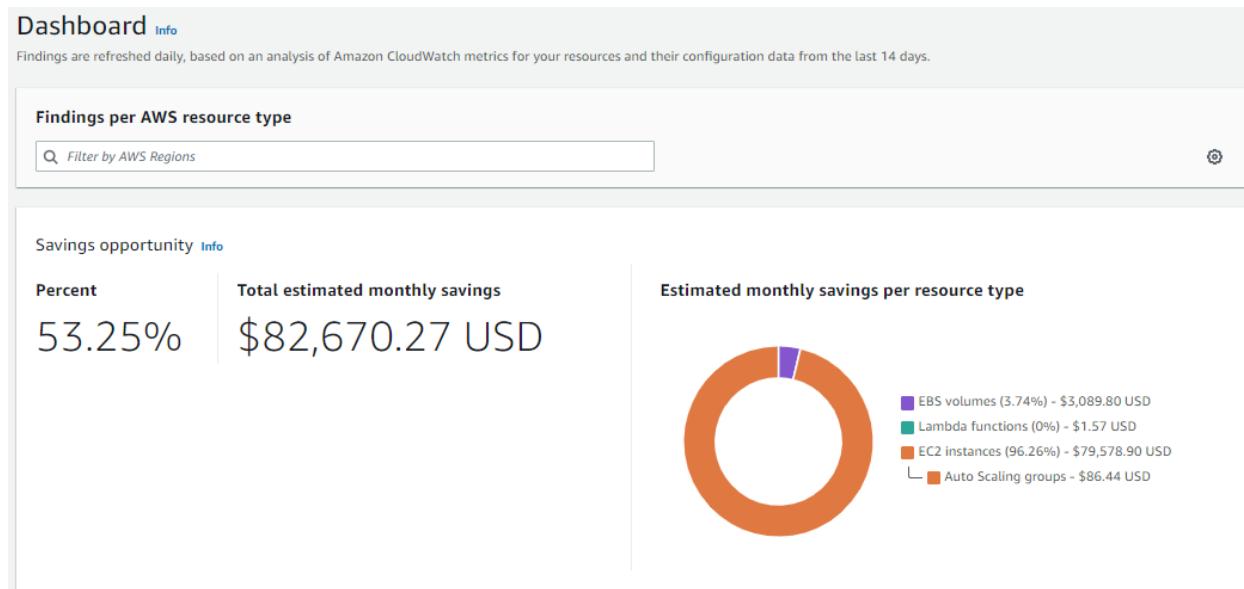
Supported AWS Resource	Requirement
Amazon EC2	<ul style="list-style-type: none">• Supported Instance types: C1,C3,C4,C5 ,C5a,C5ad,C5d,C5n,C6a,C6g,C6gn,C7g,D2,D3e n,H1,Hpc6a,I2,I3,I3en,Im4gn,Is4gen,M1,M3,M 4,M5,M5a,M5ad,M5d,M5dn,M5n,M5zn,M6a,M 6g,M6gd,M6i,R3,R4,R5,R5a,R5ad,R5b,R5d,R5d n,R5n,R6g,R6gd,R6i,T1,T2,T3,T3a,T4g,X1,X1e, X2gd,z1d
Auto Scaling Group	<ul style="list-style-type: none">• All supported EC2 Instance Types

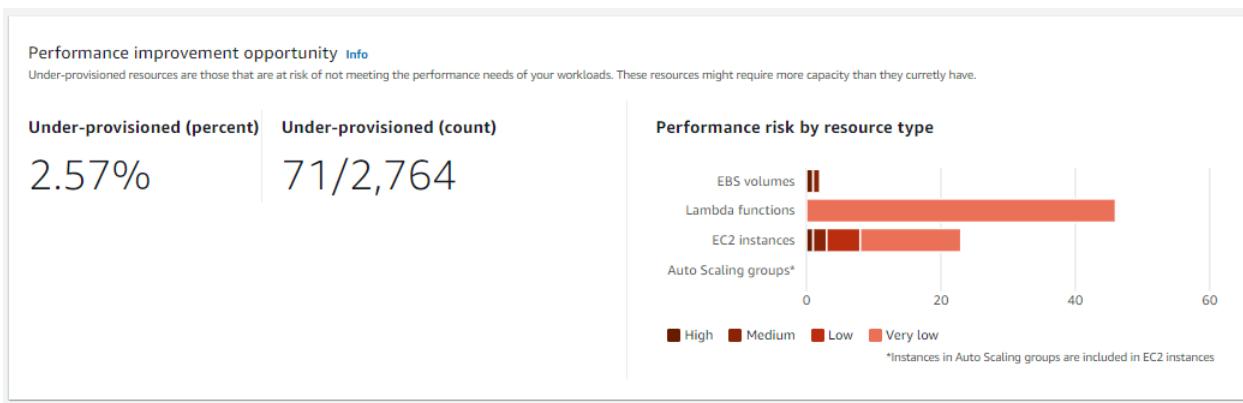


	<ul style="list-style-type: none">• Single instance type within ASG (no mixed instance types).• The same values for desired, minimum, and maximum capacity.• No scaling policy is attached.• No overrides (changes on Launch Configuration).
EBS Volume	<ul style="list-style-type: none">• EBS Volumes must be attached to an EC2 instance.• Supported EBS Volume Types: gp2, gp3, io1, io2
Lambda Function	<ul style="list-style-type: none">• Lambda functions with memory \leq 1,792 MB• Lambda functions with 50 times invocations in the last 14 days

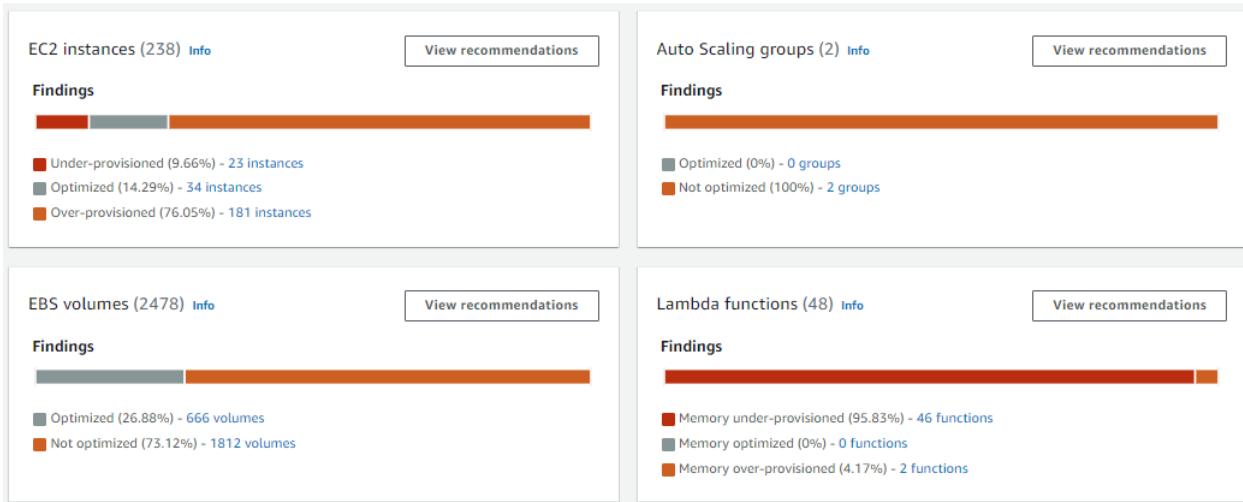
AWS Compute Optimizer Dashboard

AWS Compute Optimizer has a dashboard that helps users visualize the data like Savings opportunities and Under/Over-provisioned percentages.





Sample findings Summary per AWS Resources



Amazon Compute Optimizer will provide vital information to help you strategize accordingly from a performance and cost standpoint. This information includes Estimated monthly savings, Savings opportunity (%), Recommended instance type, Price difference (%), and many others.

Recommendations for EC2 Instances

AWS Compute Optimizer analyzes the metrics collected on the CloudWatch Metrics. For Amazon EC2 instances, these metrics are the following:

- CPUUtilization
- Memory utilization
- NetworkIn/NetworkOut
- NetworkPacketsIn/NetworkPacketsOut
- DiskReadOps/DiskWriteOps



- DiskReadBytes/DiskWriteBytes
- VolumeReadBytes/VolumeWriteOps
- VolumeReadBytes/VolumeWriteByte
- VolumeReadOps/VolumeWriteOps
- VolumeReadOps/VolumeWriteByte

An Amazon EC2 Instance can be classified as:

- Over-provisioned - underutilized infrastructure, unnecessary costs
- Under-provisioned - over utilized infrastructure, performance degradation
- Optimized - instance is correctly configured, and optimized base on the workloads

The finding reasons determine the overall finding classification for the current instance. Below are some examples of finding reasons.

- CPU over-provisioned
- Memory over-provisioned
- EBS IOPS over-provisioned
- EBS throughput over-provisioned
- Network bandwidth over-provisioned

AWS Compute Optimizer may give the following recommendations for EC2 instances.

- Recommended instance type
- Recommended On-Demand price
- Recommended RI coverage (%)
- Recommended RI utilization (%)

Sample EC2 recommendations from taken AWS Console

Recommendations for EC2 instances (20+)						Info	Export	Open in EC2 console	View details		
Recommendations for current resources to improve cost and performance.							<	1	...	>	⟳
CPU architecture preference: Current, Graviton (aws-ar... ▾											
Filter by one or more properties											
Region = Asia Pacific (Singapore)	X	Clear filters									
Instance ID	▼	Instance name	▼	Finding Info	▼	Finding reasons Info	▼	Estimated monthly savings (On-Demand) Info	Savings opportunity (%) Info	▼	
i-00██████████		VMWEBLOGIC24		Over-provisioned		CPU over-provisioned, Memory over-prov...		\$1513.7300	77.09%		
i-01██████████		VMWEBAPP1		Over-provisioned		CPU over-provisioned, Memory over-prov...		\$593.0500	81.73%		
i-01██████████		VMWEBAPP2		Over-provisioned		CPU over-provisioned, Memory over-prov...		\$593.0500	81.73%		
i-01██████████		VMDBPRD3		Over-provisioned		CPU over-provisioned, Memory over-prov...		\$532.6100	60.00%		

Recommendations for Auto Scaling Group



An Auto Scaling group can be classified as Not optimized, or Optimized.

- Optimized - correctly configured and optimized based on the workloads
- Not Optimized - at least one specification/metric is Over-provisioned, or Under-provisioned

AWS Compute Optimizer may give the following recommendations for Auto Scaling Group.

- Recommended instance type
- Recommended On-Demand price
- Recommended RI coverage (%)
- Recommended RI utilization (%)

Sample Auto Scaling Group recommendations taken from AWS Console

Recommendations for Auto Scaling groups (2) <small>Info</small>							Export	Open in ASG console	View details
Recommendations for current resources to improve cost and performance.							<	1	>
<small>CPU architecture preference: Current, Graviton (aws-arm64)</small>									
<small>Filter by one or more properties</small>									
Region = Asia Pacific (Singapore)	X	Clear filters							
Auto Scaling group name	Finding <small>Info</small>	Estimated monthly savings (On-Demand) <small>Info</small>	Savings opportunity (%) <small>Info</small>	Current instance type	Effective enhanced infrastructure metrics <small>Info</small>	Current On-Demand price <small>Info</small>			
asgwebprd01	Not optimized	\$65.4100	42.42%	t3.xlarge	Inactive	\$0.2112 per hour			
asgwebprd02	Not optimized	\$65.4100	42.42%	t3.xlarge	Inactive	\$0.2112 per hour			

Recommendations for EBS Volume Instances

The following EBS Volume metrics are collected.

- VolumeReadBytes
- VolumeWriteBytes
- VolumeReadOps
- VolumeWriteOps

An EBS Volume can be classified as Not optimized, or Optimized.

- Optimized - correctly configured and optimized based on the workloads
- Not Optimized - at least one specification/metric is Over-provisioned, or Under-provisioned

AWS Compute Optimizer may give the following recommendations for EBS Volumes.

- Recommended volume type
- Recommended size



- Recommended IOPS
- Recommended throughput
- Recommended monthly price

Sample EBS Volume recommendations taken from AWS Console

Volume ID	Finding Info	Estimated monthly savings (On-Demand) <small>Info</small>	Savings opportunity (%) <small>Info</small>	Current volume type	Current size	Current IOPS
vol-██████████	Optimized	-	-	General Purpose SSD (gp2)	15 GiB	100
vol-██████████	Not optimized	\$0.144	8.57%	General Purpose SSD (gp3)	16 GiB	3000
vol-██████████	Not optimized	\$6.000	11.11%	General Purpose SSD (gp3)	500 GiB	3000
vol-██████████	Not optimized	\$0.144	1.48%	General Purpose SSD (gp3)	100 GiB	3000
vol-██████████	Not optimized	\$0.528	0.52%	General Purpose SSD (gp3)	1000 GiB	3000
vol-██████████	Optimized	-	-	General Purpose SSD (gp2)	200 GiB	600

Lambda function metrics being used .

- Invocations
- Duration
- Errors
- Throttles

Lambda function can be classified as Not optimized, or Optimized.

- Optimized - correctly configured and optimized based on the workloads
- Not Optimized - at least one specification/metric is Over-provisioned, or Under-provisioned

AWS Compute Optimizer may give the following recommendations for Lambda Volumes.

- Recommended configured memory
- Recommended cost (low)
- Recommended cost (high)

Sample Lambda recommendations taken from AWS Console

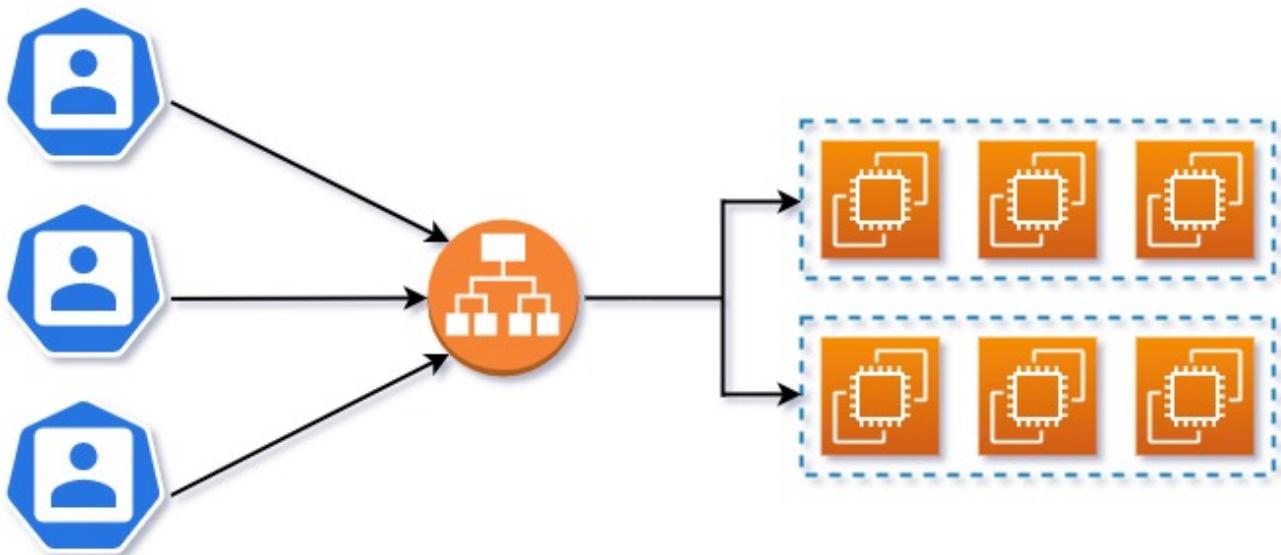


Recommendations for Lambda functions (40+) Info						Export	Open in Lambda Console	View details
Recommendations for current resources to improve cost and performance.								
<input type="text"/> Filter by one or more properties						Clear filters		
Function name	Function version Info	Finding Info	Finding reason Info	Estimated monthly savings (On-Demand) Info	Savings opportunity (%) Info			
FuncRunApplyDaily	\$LATEST	Not optimized	Memory over-provisioned	\$1.57	14.37%			
FuncAuditUserWeekly	\$LATEST	Not optimized	Memory over-provisioned	\$0.00082	10.61%			

Reference: <https://docs.aws.amazon.com/compute-optimizer/latest/ug/what-is-compute-optimizer.html>

Elastic Load Balancing

Elastic Load Balancing (ELB) is a traffic distribution service that offers high availability and elasticity for your application. ELB is a fully managed service. Its load balancer scales as traffic changes allowing it to handle millions of requests per second. A load balancer works by accepting traffic and routing it across multiple targets in one or more Availability Zones (AZ). A load balancer node is created on each enabled Availability Zone. The load balancer automatically handles the distribution of the traffic.





Load Balancer Types

ELB offers different types of load balancers. The type of load balancer depends on the traffic that you want to distribute. Another thing to check when setting up an ELB is whether the load balancer would be internet-facing or internal.

ELB supports the following types of load balancers for your workloads:

- Application Load Balancer - Layer 7 load balancing/HTTP/HTTPS traffic
- Network Load Balancer - Layer 4 load balancing/TCP/UDP traffic
- Gateway Load Balancer - Layer 3 Gateway + Layer 4 Load Balancing

ELB Features and Components

Elastic Load Balancing is placed on top of your application for traffic distribution. A load balancer is designed to increase the availability and security of your application. Aside from the load balancer type, you need to define other load balancer components and features to maximize its capabilities.

Load Balancer Scheme

You can set a load balancer to either internal or internet-facing. An internet-facing load balancer has a Public IP address assigned on its node, and the requests are routed over the internet from the client to a target. On the other hand, the nodes of an internal load balancer have a Private IP address, and requests are routed privately from client to target.

IP Addresses Type

For Application and Network Load Balancer, you can specify the IP Address type that it will support. You can choose between IPv4 or dualstack (both IPv4 and IPv6).

Listener

The listener is a process that identifies the request through ports and protocols. By defining the listener rule, a load balancer will know the traffic it will accept. Here's an Application Load Balancer listener as an example.



Listeners

A listener is a process that checks for connection requests, using the protocol and port that you configured.

Load Balancer Protocol	Load Balancer Port
HTTP	80
<input type="button" value="Choose a protocol"/>	<input type="text"/>
<input checked="" type="button" value="Choose a protocol"/>	<input type="text"/>
HTTP	
HTTPS (Secure HTTP)	

Target Group

The requests are routed to one or more targets through the port and protocol defined on the listener through the target groups. ELB supports various types of targets such as:

- EC2 Instances
- Containers
- Lambda Functions
- IP Addresses
- Virtual Appliances

Security Groups

A security group is an additional layer of security for load balancers. Only the traffic allowed by the security group will reach the load balancer.

Availability Zones

Availability Zones are locations within an AWS Region. You should always enable the AZ where your targets are located. Take note also that you can only specify one subnet per Availability Zone.

Health Checks

Health checks are configured inside a target group. It ensures that the requests are routed to healthy targets by monitoring your registered targets' HTTP and HTTPS endpoints. Health checks are also configurable; you can change the value of the threshold, interval, timeout, and success code.



Health checks

The associated load balancer periodically sends requests, per the settings below, to the registered targets to test their status.

Health check protocol

HTTP ▾

Health check path

Use the default path of "/" to ping the root, or specify a custom path if preferred.

/

Up to 1024 characters allowed.

► Advanced health check settings

Sticky Sessions

The sticky session is an ELB feature that guarantees that traffic from the same client is routed only to the same target.

Cross-zone Load Balancing

Cross-zone Load Balancing is an ELB feature that enables the load balancer node to distribute all traffic across all Availability Zones with healthy targets. When disabled, the load balancer node distributes the traffic only to its AZ.

Connection Draining

Connection draining or deregistration delay is a feature that allows a request to be completed before unregistering a target.

Load Balancer Monitoring

CloudWatch integrates with Elastic Load Balancing to monitor and log its metrics. API calls to ELB are also logged on CloudTrail.

Delete Protection

Deletion protection prevents the accidental deletion of a load balancer.

Choosing the Right Load Balancer

Application Load Balancer (ALB)

For routing HTTP and HTTPS requests, an Application Load Balancer is a good choice. Since this load balancer works on Layer 7, it can distribute your web application's HTTP and HTTPS requests. An ALB can work as an



internet-facing and internal load balancer, and work for IPv4 and dualstack IP addresses. It requires a minimum of two different subnets on different Availability Zones. Note that a server certificate is required if you have a listener configured for HTTPS protocol. You can either use a certificate from AWS Certificate Manager or IAM.

Network Load Balancer (NLB)

A Network Load Balancer is ideal for distributing TCP and UDP traffic from the client to your application. Like an Application Load Balancer, the Network Load Balancer can also be either internal or internet-facing. It also supports both IPv4 and dualstack IP addresses. You can set up a Network Load Balancer using only one subnet, but two subnets are ideal for improving your application's fault tolerance. Make sure as well that the target group has at least one valid target to accept the requests. If multiple AZs are specified, you can enable Cross-Zone Load Balancing to distribute the traffic in all targets equally. The direction of the network traffic will depend on the rules defined by the listener and routing.

The screenshot shows the AWS Network Load Balancer (NLB) configuration interface. A single listener named "Listener TCP:80" is listed. The "Protocol" is set to "TCP" and the "Port" is "80". The "Default action" is "Forward to" a target group named "Web-EC2-Group" (Target type: Instance). There is a "Create target group" button and a "Remove" button. An "Add listener" button is also visible.

Gateway Load Balancer (GWLB)

A Gateway Load Balancer eases the management of multiple virtual appliances. A GWLB can distribute traffic across multiple virtual appliances that support Generic Network Virtualization Encapsulation (GENEVE). You can enable GWLB on a single subnet, but the target group should have at least one target that accepts the GENEVE protocol. Take note that Gateway Load Balancer only supports IPv4 addresses; it doesn't support IPv6 and dualstack.



Summary

Review and confirm your configurations. [Estimate cost](#)

Basic configuration Edit	Network mapping Edit	IP listener routing Edit	Tags Edit
<i>Load balancer name not defined</i> <ul style="list-style-type: none">• IPv4	VPC vpc-03c57950e44f8783b vpc_webapp <ul style="list-style-type: none">• us-east-1a subnet-0c06bf5364a2e0f3f web-subnet1• us-east-1b subnet-060fcfc450d83dcff web-subnet-2	GENEVE:6081 defaults to TD-VA	<i>None</i>

References:

<https://docs.aws.amazon.com/elasticloadbalancing/latest/userguide/what-is-load-balancing.html>

<https://docs.aws.amazon.com/elasticloadbalancing/latest/userguide/how-elastic-load-balancing-works.html>

S3 Presigned URL

All buckets and objects are stored privately by default in S3. This is a much more secure design than having your objects publicly accessible by default. But if you have a storage requirement where you must share an object or bucket with your customers to upload an object quickly, you can use the S3 presigned URL to generate a URL for your customer. The URL generated can be used by a customer to view or upload an object. The action and the duration of the access can also be defined when generating a URL.

Sharing S3 objects using Presigned URL

There are multiple ways of creating a presigned URL to share an object, such as:

- AWS CLI
- AWS SDK for Java
- AWS Explorer for Visual Studio
- REST API
- .NET
- Ruby
- PHP
- Node.js
- Python
- Go

You can use the following security credentials when generating a presigned URL. Note that the user who generates a presigned URL should have the necessary permission to execute the action of the generated URL.



Security Credentials	Validity
IAM instance profile	6 hours max validity
AWS Security Token Service	36 hours max validity for permanent credentials
IAM user (access key and secret access key)	Seven days max validity when using AWS Signature Version 4

In this example, we use AWS Command Line Interface to generate a URL and share an object in S3.

The screenshot shows the AWS S3 console for the 'tutorialsdojo-test' bucket. The 'Permissions' tab is selected. Under 'Access', it says 'Bucket and objects not public'. In the 'Block public access (bucket settings)' section, the 'Edit' button is visible. The 'Block all public access' setting is turned 'On'. Below it, five specific options are also set to 'On': 'Block public access to buckets and objects granted through new access control lists (ACLs)', 'Block public access to buckets and objects granted through any access control lists (ACLs)', 'Block public access to buckets and objects granted through new public bucket or access point policies', 'Block public access to buckets and objects granted through any public bucket or access point policies', and 'Block public and cross-account access to buckets and objects through any public bucket or access point policies'.



Amazon S3 > tutorialsdojo-test

tutorialsdojo-test Info

Objects Properties Permissions Metrics Management Access Points

Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permission.

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified
<input type="checkbox"/>	sample.txt	txt	July 11, 2021, 01:03:24 (UTC+08:00)

To generate a presigned URL, follow the command below.

```
aws s3 presign <s3objectURL>
```

Example:

```
aws s3 presign s3://tutorialsdojo-test/sample.txt
```

Sample Output

```
https://tutorialsdojo-test.s3.ap-southeast-1.amazonaws.com/sample.txt?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIA5ZBEVVFGZ6VYNLZ%2F20210710%2Fap-southeast-1%2Fs3%2Faws4_request&X-Amz-Date=20210710T171224Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=409f641d1b06f2122b17fb26f9ef56a84b475714b4eb7fec29151746d963df5a
```

Use the generated URL to grant temporary access on the object.

The screenshot shows a web browser window with the URL https://tutorialsdojo-test.s3.ap-southeast-1.amazonaws.com/sample.txt?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIA5ZBEVVFGZ6VYNLZ%2F20210710%2Fap-southeast-1%2Fs3%2Faws4_request&X-Amz-Date=20210710T171224Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=409f641d1b06f2122b17fb26f9ef56a84b475714b4eb7fec29151746d963df5a. The page content is "This is a test."

Optionally, you can define an access duration for the URL by adding the `--expires-in` parameter. If you didn't specify the duration, it would get a default value of 3600 seconds.



```
aws s3 presign <s3objectURL> --expires-in <value>
```

Uploading S3 Objects Using Presigned URL

Presigned URL also enables you to upload an object to a specific bucket. Users who upload an object using the presigned URL need no credentials or permissions.

To generate a presigned URL for uploading objects, you can use AWS SDK. You need to provide valid security credentials and specify the following parameters.

- Bucket Name
- Object Key
- HTTP method (PUT method)
- Expiration

In this example, we use AWS SDK for Python. The function `generate_presigned_url` will generate the presigned URL with the defined parameters. You can set the validity by defining the `ExpiresIn` value in seconds.

```
import boto3
url = boto3.client('s3').generate_presigned_url(
    ClientMethod='put_object',
    Params={'Bucket': 'S3BucketName', 'Key': 'ObjectKey'},
    ExpiresIn=3600)
```

Reference:

<https://docs.aws.amazon.com/AmazonS3/latest/userguide/using-presigned-url.html>

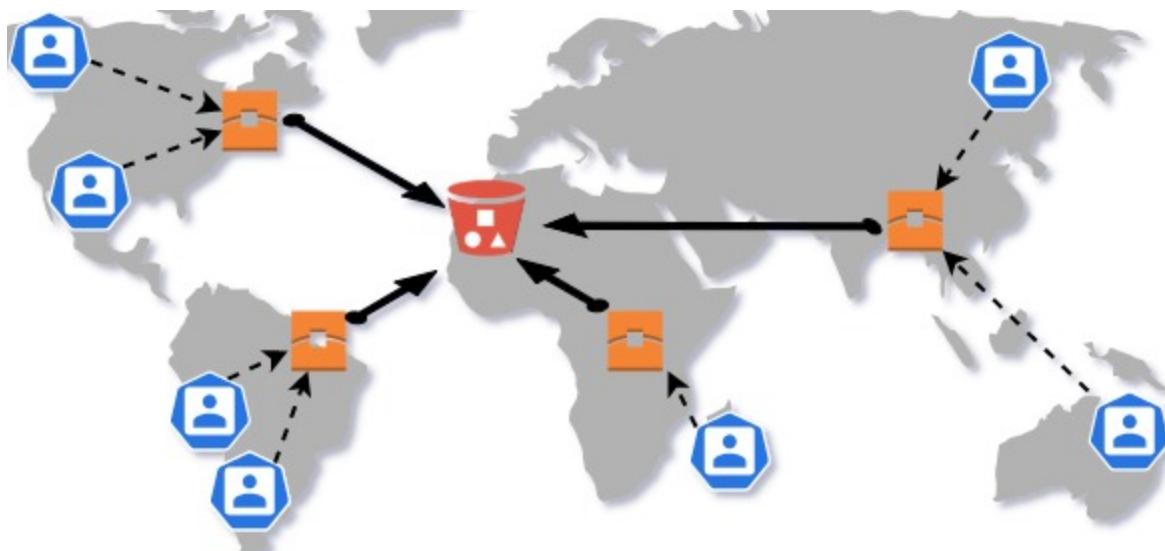


S3 Transfer Acceleration

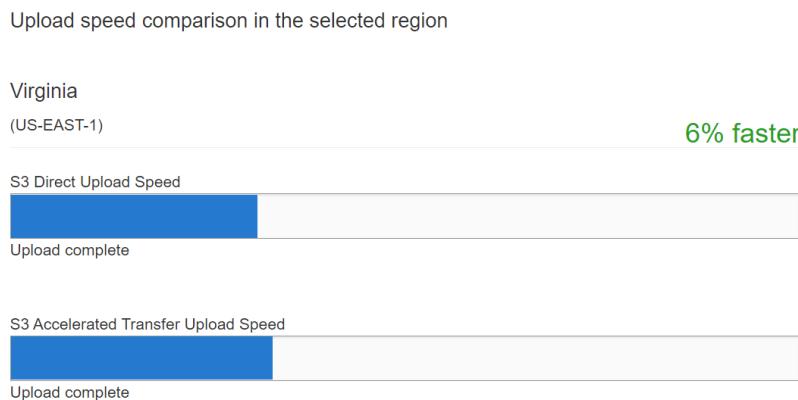
When catering to a global audience, the distance between you and your client is undoubtedly an issue. This is also applicable for the S3 service, especially when uploading large objects across the regions. It will surely take a long time to upload hundreds of gigabytes of things to an S3 bucket located on the other side of the globe. To address this issue, you can enable the S3 Transfer Acceleration feature on S3 buckets.

The screenshot shows the 'Edit transfer acceleration' configuration page for the 'tutorialsdojo-test' bucket. The 'Transfer acceleration' section is visible, with the 'Enable' option selected (indicated by a red arrow). The 'Accelerated endpoint' field contains the value 'tutorialsdojo-test.s3-accelerate.amazonaws.com', which is also highlighted with a red box. A callout box provides information about using the accelerated endpoint for faster data transfers, mentioning additional fees and linking to the Amazon S3 pricing page. At the bottom right, there are 'Cancel' and 'Save changes' buttons.

S3 Transfer Acceleration integrates with Amazon CloudFront and uses its AWS Edge location around the globe to create an accelerated endpoint for a much faster data transfer. In this way, your S3 bucket is much closer to your users. Data will be uploaded on an AWS Edge Location and routed to your S3 bucket through the AWS internal network.



AWS offers a [Speed Comparison](#) tool to compare the performance of accelerated transfer and standard S3 data transfer on different AWS regions.



Reference:

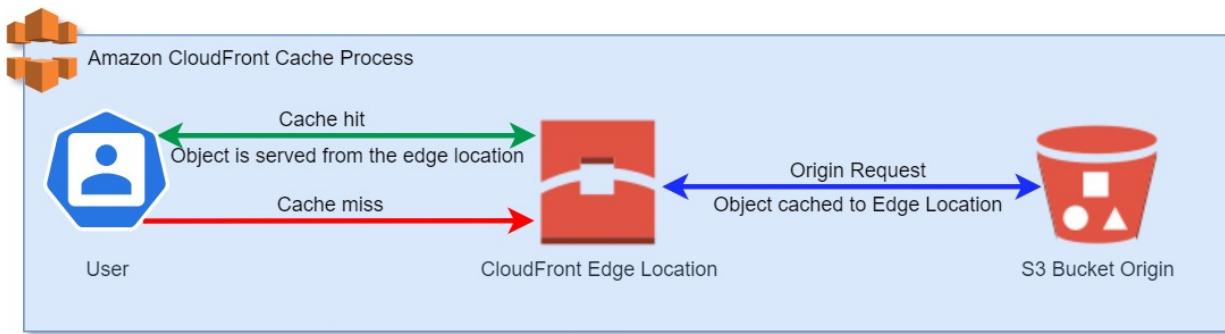
<https://docs.aws.amazon.com/AmazonS3/latest/userguide/transfer-acceleration.html>

Amazon CloudFront

CloudFront enables customers to deliver content throughout the globe by using CloudFront edge location. Content is cached to different edge locations from an origin server allowing viewer requests to be directed to the nearest edge location instead of the origin server.

Caching Process

Every object being cached has a unique identifier called **Cache Key** associated with them. This key may contain values like domain name, URL, HTTP header, query strings, and cookies. When a viewer sends a request to an edge location, CloudFront determines if the request is a cache hit or miss by comparing the values of the request with the values in the cache key.



A cache hit occurs when the request and cache key's value matches, thus confirming that the requested object is cached on the edge location. On the other hand, the request is forwarded to the origin server (origin request) to retrieve the requested object during a cache miss event. The **Origin Request** contains values like URL, request body, and headers. The origin server then sends the object to the user through the edge location and caching simultaneously.

The cache key directly affects the cache hit ratio of an object; the more values defined on the cache key, the lower the number of hits that can affect your application or website performance.

CloudFront Policies

CloudFront Policies help you to granularly define what values you want to include in Cache Key and Origin Request. First, it is essential to determine what values your application/website requires from the viewer's request. You may also want to choose the values being forwarded to the origin server if you plan to collect these data.



Cache Policy

Amazon CloudFront provides a preconfigured **AWS-managed cache policy** that you can use for your distribution. Below is a comparison of the AWS-managed cache policy.

Cache Policy	Cache Key Settings			Compression Support	
	Header	Cookies	Query Strings	Gzip	Brotli
Amplify	Authorization CloudFront-Viewer-Country Host	All	All	Yes	Yes
CachingDisabled	None	None	None	No	No
CachingOptimized	None	None	None	Yes	Yes
CachingOptimizedForUncompressedObjects	None	None	None	No	No
Elemental-MediaPackage	origin	None	aws.manifestfilter start end	Yes	No

If the AWS-managed cache policy doesn't fit your requirements, you can always create a **custom cache policy**. A custom cache policy gives you the flexibility to specify header, and block or accept a particular query string and cookies.



Cache key settings Info

Headers
Choose which headers to include in the cache key.

Include the following headers ▾

Add header
Select an existing header element or create a custom header. (max 10)

Select headers ▾ Add custom

Authorization X Host X CloudFront-Viewer-Country X

Query strings
Choose which query strings to include in the cache key.

Include all query strings except ▾

Block
Block a query string from the cache key.
Add item

Cookies
Choose which cookies to include in the cache key.

All ▾

Origin Request Policy

Origin Request Policy defines the information included in the origin requests. AWS also provides a preconfigured **AWS-managed origin request** policy for origin requests. Below is a comparison of the AWS-managed origin request policy.

Origin Request Policy	Cache Key Settings		
	Header	Cookies	Query Strings
AllViewer	All	All	All
CORS-CustomOrigin	origin	None	None
CORS-S3Origin	origin access-control-request-headers access-control-request-method	None	None



Elemental-MediaTailor-PersonalizedManifests	origin access-control-request-headers x-forwarded-for access-control-request-method user-agent	None	All
UserAgentRefererHeaders	referer user-agent	None	None

If the AWS-managed origin request policy doesn't fit your requirements, you can always create a custom cache policy. A custom origin request policy gives you the flexibility to specify header, query strings, and cookies.

Origin request settings info

Headers
Choose which headers to include in origin requests.

Include the following headers ▾

Add header
Select an existing header element or create a custom header. (max 10)

Select headers ▾ Add custom

Origin X CloudFront-Viewer-Country X

Query strings
Choose which query strings to include in origin requests.

None ▾

Cookies
Choose which cookies to include in origin requests.

Include specified cookies ▾

Allow
Add the name of the cookie to include in origin requests.

session_id Remove

Add item

Reference:

<https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/HowCloudFrontWorks.html>
<https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/working-with-policies.html>



Amazon ElastiCache

ElastiCache Memcached and Redis Engine

AWS offers a fully managed service for in-memory caching for your applications. ElastiCache generally supports both Memcached and Redis engines. Both are identical, but you certainly shouldn't easily choose a specific engine over the other since both of them have specific use cases. Here are the things you might want to consider when choosing an ElastiCache engine.

Create your Amazon ElastiCache cluster



Cluster engine Redis

In-memory data structure store used as database, cache and message broker. ElastiCache for Redis offers Multi-AZ with Auto-Failover and enhanced robustness.

Cluster Mode enabled

Memcached

High-performance, distributed memory object caching system, intended for use in speeding up dynamic web applications.

Clusters

To start using ElastiCache, you create a cluster and specify the engine to be used. A cluster contains one or more cache nodes with their memory and compute resources according to the node type. Each node in a cluster runs a cache engine. For the Redis engine, you can choose to enable cluster mode during cluster creation. Redis (Cluster Mode enabled) offers a limited option for cluster modification, but Memcached and Redis (Cluster Mode disabled) support adding and removing nodes from the cluster.

Sharding

One key difference between Memcached and Redis is the sharding feature. Memcached doesn't support the use of shards, while Redis does. Shard is a hierarchical arrangement of multiple nodes that allows you to have primary and replica nodes. Redis (Cluster Mode enabled) may contain up to 500 shards, while Redis (Cluster Mode disabled) may only have a single shard on a cluster. Both Redis and Memcached can split their data



across all nodes. Still, sharding allows for much efficient processing, which is beneficial when working with a large scale in-memory caching.

Multithreading

Memcache supports multithreading allowing it to process through multiple cores and making it easy to scale up its computing resources. On the other hand, Redis runs on a single thread, which allows it to scale horizontally on a cluster.

High Availability

Suppose high availability is a crucial requirement for your caching workloads. In that case, Redis is the engine to choose from since Memcache doesn't support replication and high availability. Redis supports replication through sharding. For Redis (Cluster Mode disabled), all nodes can contain all the cluster's data in a single shard, while for Redis (Cluster Mode enabled), data are partitioned across all shards. This feature also enables Redis for automatic failover. Below is an example of a node, replica, and Multi-AZ configuration on ElastiCache Redis.

The screenshot shows the configuration options for a new ElastiCache Redis cluster. It includes three main fields: 'Node type' set to 'cache.r6g.large (13.07 GiB)', 'Number of replicas' set to '2', and 'Multi-AZ' checked. Each field has an information icon (i) to its right.

Node type	cache.r6g.large (13.07 GiB)	i
Number of replicas	2	i
Multi-AZ	<input checked="" type="checkbox"/>	i

Backup and Restore

An ElastiCache cluster running on the Redis engine can use S3 to export a backup. You can set up a scheduled automatic backup or trigger a manual backup for your Redis cluster. You can also create a Redis cluster using a backup from an S3 bucket. On the other hand, Memcached doesn't support backup and restore. Here's an example backup configuration of the ElastiCache cluster using the Redis engine.



Import data to cluster

Seed RDB file S3 location

myBucket/myFolder/objectName



Use comma to separate multiple paths in the field

Backup

Enable automatic backups



Backup retention period

4



day(s)

Backup window



No preference



Specify backup window



Key Points

The workloads you'll be processing must be considered when choosing the right engine for ElastiCache. If you'll be working with simple data types like objects and need the capability to scale depending on the number of workloads, it is recommended to use Memcached. If your data includes complex data types (string, sets, sorted sets, lists, hashes) and requires high availability and failover capabilities, Redis is the right choice.

References:

<https://docs.aws.amazon.com/AmazonElastiCache/latest/red-ug>SelectEngine.html>

<https://docs.aws.amazon.com/AmazonElastiCache/latest/red-ug/WhatIs.html>

<https://docs.aws.amazon.com/AmazonElastiCache/latest/mem-ug/WhatIs.html>



Virtual Private Cloud

Network Access Control List (NACL)

AWS recommends applying security on any level of your architecture. For VPC Subnets, it's the **Network Access Control List (NACL)**. The NACL may allow or deny network traffic according to the rules you applied. Unlike a security group, the NACLs are stateless. Accepted inbound traffic is not automatically permitted on outbound; you need to define it explicitly. Changes made to an NACL are also applied immediately. You can't associate multiple NACL to a subnet simultaneously, but you can associate an NACL to various subnets. You create a rule by specifying the following details.

- Rule number.
- Type
- Protocol
- Port range
- Source (Inbound rules only)
- Destination (Outbound rules only)
- Action (Allow/Deny)

Each subnet requires an NACL associated with it. AWS provides a default network ACL that accepts both inbound and outbound traffic.

Default NACL Inbound Rules

Inbound rules	Outbound rules	Subnet associations	Tags																		
Inbound rules (2)																					
<input type="button" value="Edit inbound rules"/>																					
<input type="button" value="Filter inbound rules"/> < 1 > ⚙																					
<table border="1"><thead><tr><th>Rule number</th><th>Type</th><th>Protocol</th><th>Port range</th><th>Source</th><th>Allow/Deny</th></tr></thead><tbody><tr><td>100</td><td>All traffic</td><td>All</td><td>All</td><td>0.0.0.0/0</td><td> Allow</td></tr><tr><td>*</td><td>All traffic</td><td>All</td><td>All</td><td>0.0.0.0/0</td><td> Deny</td></tr></tbody></table>				Rule number	Type	Protocol	Port range	Source	Allow/Deny	100	All traffic	All	All	0.0.0.0/0	Allow	*	All traffic	All	All	0.0.0.0/0	Deny
Rule number	Type	Protocol	Port range	Source	Allow/Deny																
100	All traffic	All	All	0.0.0.0/0	Allow																
*	All traffic	All	All	0.0.0.0/0	Deny																

Default NACL Outbound Rules

Inbound rules	Outbound rules	Subnet associations	Tags																		
Outbound rules (2)																					
<input type="button" value="Edit outbound rules"/>																					
<input type="button" value="Filter outbound rules"/> < 1 > ⚙																					
<table border="1"><thead><tr><th>Rule number</th><th>Type</th><th>Protocol</th><th>Port range</th><th>Destination</th><th>Allow/Deny</th></tr></thead><tbody><tr><td>100</td><td>All traffic</td><td>All</td><td>All</td><td>0.0.0.0/0</td><td> Allow</td></tr><tr><td>*</td><td>All traffic</td><td>All</td><td>All</td><td>0.0.0.0/0</td><td> Deny</td></tr></tbody></table>				Rule number	Type	Protocol	Port range	Destination	Allow/Deny	100	All traffic	All	All	0.0.0.0/0	Allow	*	All traffic	All	All	0.0.0.0/0	Deny
Rule number	Type	Protocol	Port range	Destination	Allow/Deny																
100	All traffic	All	All	0.0.0.0/0	Allow																
*	All traffic	All	All	0.0.0.0/0	Deny																



The default network ACL works fine if you don't have strict security requirements. Still, it's a good practice to leverage this feature to another layer of security in your cloud architecture.

You can create your own fine-grained rules on a custom network ACL and associate them with your subnets. Always consider and evaluate your application's traffic and traffic from your users to avoid any conflicts on the rule. By default, a newly created custom NACL denies all inbound traffic.

Below is an example of a custom NACL with fine-grained rules for inbound and outbound. The custom NACL allows HTTPS from IPv4 addresses, SSH from a specific IP address, and Oracle database connection from a particular CIDR block. The rule with the lowest rule number is evaluated first; in this case, the HTTPS rule is applied first, then followed by the rule for SSH and Oracle database traffic. The rule with an asterisk (*) on rule number guarantees that all network traffic that is not explicitly defined is denied. Note that a rule is immediately applied for specific traffic and doesn't consider conflicts with other rules with a higher rule number.

acl-0d1d563c4a57c265c / td-nacl-app

Details **Inbound rules** Outbound rules Subnet associations Tags

Inbound rules (4)

Edit inbound rules

Rule number	Type	Protocol	Port range	Source	Allow/Deny
100	HTTPS (443)	TCP (6)	443	0.0.0.0/0	Allow
110	SSH (22)	TCP (6)	22	206.1.110.3/32	Allow
120	Oracle (1521)	TCP (6)	1521	10.1.0.0/24	Allow
*	All traffic	All	All	0.0.0.0/0	Deny

acl-0d1d563c4a57c265c / td-nacl-app

Details Inbound rules **Outbound rules** Subnet associations Tags

Outbound rules (4)

Edit outbound rules

Rule number	Type	Protocol	Port range	Destination	Allow/Deny
100	HTTPS (443)	TCP (6)	443	0.0.0.0/0	Allow
110	SSH (22)	TCP (6)	22	206.1.110.3/32	Allow
120	Oracle (1521)	TCP (6)	1521	10.1.0.0/24	Allow
*	All traffic	All	All	0.0.0.0/0	Deny



Route Tables

The Route tables control the traffic routing of your subnets and gateways. The route table evaluates the destination and target defined in the route to direct traffic. All subnets in a VPC require a route table. When a VPC is created, the main route table is also created, which handles the traffic routing within that VPC. The main route table is implicitly associated with a VPC until you explicitly associate a new custom route table to it. Remember that you can associate a route table with multiple subnets. However, a subnet can be associated with only one route table at a time.

On the route table, you define a destination and target. A destination determines where the traffic from an IP address range or a prefix goes. Traffic from the destination is routed to whatever is defined on the route target.

- **Egress Only Internet Gateway** - allows only outbound IPv6 traffic to the public internet. Prohibits IPv6 connection from the internet
- **Gateway Load Balancer Endpoint** - a VPC endpoint that the Gateway Load Balancer uses for distributing traffic between the service provider and service consumer VPC
- **Instance** - EC2 instance within the same instance
- **Internet Gateway** - allows inbound and outbound communication between VPC and public internet
- **Local** - route for the local VPC
- **NAT Gateway** - resides on a public subnet. Allows internet connection for private subnets
- **Network Interface** - virtual network card attached to an instance
- **Outpost Local Gateway** - allows communication between VPC and on-premises network
- **Peering Connection** - the connection between two VPCs
- **Transit Gateway** - interconnects multiple VPC and on-premises networks using a network transit hub
- **Virtual Private Gateway** - virtual router used for VPN tunnel



Edit routes

The screenshot shows a search interface for route targets. On the left, there is a table header with 'Destination' and 'Target'. Below it, a row has '172.31.0.0/16' in the Destination column and a search input field in the Target column containing 'local'. A dropdown menu is open next to the search input, listing various target types: Egress Only Internet Gateway, Gateway Load Balancer Endpoint, Instance, Internet Gateway, local, NAT Gateway, Network Interface, Outpost Local Gateway, Peering Connection, Transit Gateway, and Virtual Private Gateway. The 'local' option is visible in the dropdown.

Here is an example of a custom route table. Creating a route table automatically creates a local route to enable traffic routing within the VPC. You also see a route for all IPv4 and IPv6 addresses routed to an Internet Gateway. You can route traffic from a prefix (Amazon S3) to a specific network interface within a VPC.

The screenshot shows the 'Routes' tab of a Route Table configuration page. At the top, there are tabs for 'Routes', 'Subnet associations', 'Edge associations', 'Route propagation', and 'Tags'. The 'Routes' tab is selected. Below the tabs, there is a table titled 'Routes (4)' with columns: Destination, Target, Status, and Propagated. The table data is as follows:

Destination	Target	Status	Propagated
172.31.0.0/16	local	Active	No
0.0.0.0/0	igw-e8a2a68c	Active	No
::/0	igw-e8a2a68c	Active	No
pl-6fa54006	eni-02d4a726c1fc1dd1a	Active	No



VPC Flow logs

Using VPC Flow logs, you can track IP traffic to and from your network interfaces in a VPC. This feature helps you to monitor traffic or troubleshoot a network-related operational issue.

You create a Flow Log on the Network Interface on the EC2 console. You can set the Flow log to capture accepted and rejected traffic only, or both. You also set the logging interval to either 10 minutes or 1 minute.

Filter

The type of traffic to capture (accepted traffic only, rejected traffic only, or all traffic).

- Accept
- Reject
- All

Maximum aggregation interval Info

The maximum interval of time during which a flow of packets is captured and aggregated into a flow log record.

- 10 minutes
- 1 minute

Destination

The destination to which to publish the flow log data.

- Send to CloudWatch logs
- Send to an S3 bucket

When you create a Flow log, you have an option to publish the data into an Amazon CloudWatch logs or Amazon S3 bucket. When you choose to publish the data to Amazon CloudWatch logs, you need to specify the log and an IAM role; otherwise, set the ARN for the Amazon S3 bucket.

You can also choose the log format. You can use the AWS default format or choose only specific log attributes by selecting a custom format.



Log record format

The fields to include in the flow log record.

- AWS default format
- Custom format

Format preview

```
 ${version} ${account-id} ${interface-id} ${srcaddr} ${dstaddr} ${srcport} ${dstport}  
 ${protocol} ${packets} ${bytes} ${start} ${end} ${action} ${log-status}
```

Here is an example of a flow log published to an Amazon CloudWatch Logs log group. The log filter is helpful for things like viewing traffic from a specific address, checking traffic on particular ports, and others.

Log events		
You can use the filter bar below to search for and match terms, phrases, or values in your log events. Learn more about filter patterns		
<input type="checkbox"/> View as text	Actions	Create Metric Filter
<input type="text"/> Filter events	Clear	1m 30m 1h 12h Custom
▶	Timestamp	Message
		No older events at this moment. Retry
▶	2021-08-09T22:11:34.000+08:00	2 947117271373 eni-02d4a726c1fc1dd1a 193.93.62.46 172.31.15.23 24472 3389 6 8 1486 1628518294 1628518354 ACCEPT OK
▶	2021-08-09T22:11:34.000+08:00	2 947117271373 eni-02d4a726c1fc1dd1a 193.93.62.81 172.31.15.23 42462 3389 6 8 1502 1628518294 1628518354 ACCEPT OK
▶	2021-08-09T22:11:34.000+08:00	2 947117271373 eni-02d4a726c1fc1dd1a 193.27.228.61 172.31.15.23 49891 1564 6 1 40 1628518294 1628518354 REJECT OK
▶	2021-08-09T22:11:34.000+08:00	2 947117271373 eni-02d4a726c1fc1dd1a 172.31.106.84 172.31.15.23 193.93.62.47 3389 1132 6 7 1905 1628518294 1628518354 ACCEPT OK
▶	2021-08-09T22:11:34.000+08:00	2 947117271373 eni-02d4a726c1fc1dd1a 23.2.106.84 172.31.15.23 80 39275 6 1 44 1628518294 1628518354 REJECT OK
▶	2021-08-09T22:11:34.000+08:00	2 947117271373 eni-02d4a726c1fc1dd1a 54.240.227.54 172.31.15.23 443 62003 6 24 7177 1628518294 1628518354 ACCEPT OK
▶	2021-08-09T22:11:34.000+08:00	2 947117271373 eni-02d4a726c1fc1dd1a 193.93.62.47 172.31.15.23 1132 3389 6 8 1502 1628518294 1628518354 ACCEPT OK
▶	2021-08-09T22:11:34.000+08:00	2 947117271373 eni-02d4a726c1fc1dd1a 193.93.62.23 172.31.15.23 52584 3389 6 8 1486 1628518294 1628518354 ACCEPT OK
▶	2021-08-09T22:11:34.000+08:00	2 947117271373 eni-02d4a726c1fc1dd1a 185.156.73.19 172.31.15.23 55308 57418 6 1 40 1628518294 1628518354 REJECT OK
▶	2021-08-09T22:11:34.000+08:00	2 947117271373 eni-02d4a726c1fc1dd1a 172.31.15.23 54.240.227.54 62001 443 6 4 214 1628518294 1628518354 ACCEPT OK

Traffic Mirroring

Traffic Mirroring is a VPC feature that allows you to stream both inbound and outbound traffic from a network interface to multiple targets.

With traffic mirroring, you can analyze your network traffic in real-time by sending the mirrored traffic to many monitoring/security devices. The device can be on the same VPC or a different VPC connected via VPC Peering, AWS Transit Gateway, or the Gateway Load Balancer endpoint. Since the network packets are



captured from the network interface level, running an agent on your instance is unnecessary. With a simple configuration, you can quickly detect threats and identify vulnerabilities.

To use Traffic Mirroring, you need to configure the following.

Mirror targets - the destination of mirrored traffic. A target can be any of the following:

- Network interface
- Network Load Balancer with UDP listener
- Gateway Load Balancer with UDP listener

Choose target
Target type cannot be modified after creation.

Target type
Network Interface

Target
eni-075359fb30a8a9db7

Mirror filter - defines the specific inbound and outbound traffic that is mirrored. No traffic is accepted by default; you need to accept or reject traffic explicitly.

Inbound rules - optional

Number	Rule action	Protocol	Source port range - optional	Destination port range - optional	Source CIDR block	Destination CIDR block	Description
100	reject	TCP (6)	22	22	0.0.0.0/0	0.0.0.0/0	(X)
200	accept	TCP (6)	8080	7832	0.0.0.0/0	10.1.0.0/24	(X)

Add rule

Outbound rules - optional

Number	Rule action	Protocol	Source port range - optional	Destination port range - optional	Source CIDR block	Destination CIDR block	Description
100	reject	TCP (6)	22	22	0.0.0.0/0	0.0.0.0/0	(X)
200	accept	TCP (6)	8080	7832	0.0.0.0/0	10.1.0.0/24	(X)

Add rule

Mirror sessions - set up the traffic mirroring by specifying the Mirror source, target, and Filter.



Session settings
Set description, source, and target.

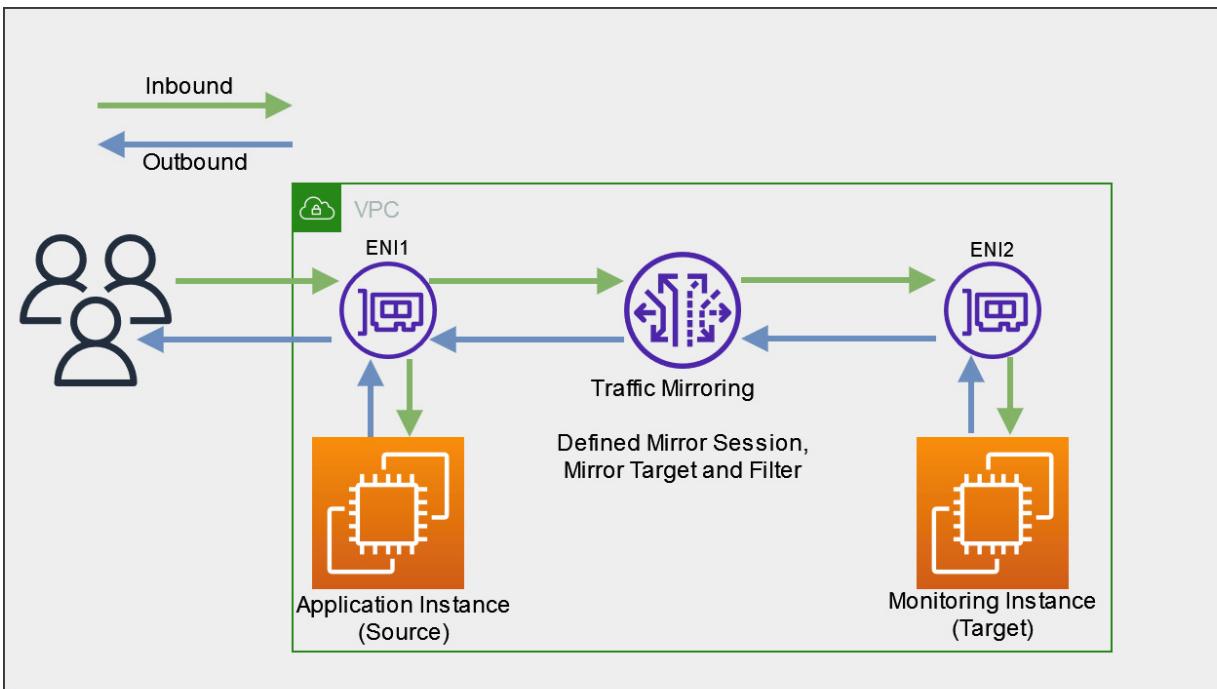
Name tag - optional

Description - optional

Mirror source
The resource that you want to monitor.
 X G
Only network interfaces of type "Interface" are allowed.

Mirror target
A network interface, or a network load balancer, or a gateway load balancer endpoint that is the destination for mirrored traffic.
 X G Create target

In the example below, users access an application hosted in an EC2 instance. On the same VPC, a monitoring instance is spawned (a SaaS or an instance with any network/packet analyzer capabilities). According to the defined Filter, inbound and outbound traffic from the users going to the application instance is mirrored in the monitoring instance. From there, the monitoring instance can see and analyze the same traffic the application has.



References:

- <https://docs.aws.amazon.com/vpc/latest/userguide/vpc-network-acls.html>
- https://docs.aws.amazon.com/vpc/latest/userguide/VPC_Route_Tables.html



<https://docs.aws.amazon.com/vpc/latest/userguide/flow-logs.html>

<https://docs.aws.amazon.com/vpc/latest/mirroring/what-is-traffic-mirroring.html>

Amazon Route 53

Amazon Route 53 is a scalable Domain Name Service (DNS) for your resources inside and outside an AWS environment. Route 53 gives you capabilities to register a domain, use different routing policies for DNS routing, and monitor resources through health checks.

Domain Registration

Domain registration on Route 53 is straightforward. You search by providing a domain name and extension (.com, .net, etc.). Route 53 will tell you the availability of the domain and gives you domain suggestions. Domains are priced per year, and the prices also vary per domain extension.

Choose a domain name

cloudcomputing	.com - \$12.00	Check
----------------	----------------	-------

Availability for 'cloudcomputing.com'

Domain Name	Status	Price /1 Year	Action
cloudcomputing.com	✗ Unavailable		

Related domain suggestions

Domain Name	Status	Price /1 Year	Action
alphacloudcomputing.com	✓ Available	\$12.00	Add to cart
alphacloudcomputing.net	✓ Available	\$11.00	Add to cart
analyticscomputing.com	✓ Available	\$12.00	Add to cart
appscomputing.com	✓ Available	\$12.00	Add to cart
cloudcomputing.systems	✓ Available	\$21.00	Add to cart
cloudcomputingsystems.net	✓ Available	\$11.00	Add to cart
cloudforcecomputing.com	✓ Available	\$12.00	Add to cart
intercloudcomputing.com	✓ Available	\$12.00	Add to cart
saintcloudcomputing.com	✓ Available	\$12.00	Add to cart



If you have an existing domain on another DNS provider, you can also transfer its registration under Route 53.

Transfer Domain to Route 53

You can transfer registration for one or more domains from another registrar to Route 53. Before you continue, do the following:

- Confirm that the domain is transferable. See [Transfer requirements for top-level domains](#).
- For each domain that you want to transfer, perform the first four steps of [Transferring registration for a domain to Route 53](#).

To transfer up to five domains, you can enter each domain name below.

To transfer more than five domains, you can use the [Transfer multiple domains to Route 53](#) page.

A screenshot of a web interface showing a search bar with a placeholder, a dropdown menu set to '.com - \$12.00', and a blue 'Check' button.

Route 53 Service Integrations

Route 53 works with multiple AWS resources and even resources outside AWS. For AWS resources, you can route traffic to the following:

- **Amazon EC2** - route traffic to website and application hosted on EC2 instances using IP address.
- **Amazon VPC** - route traffic to a VPC endpoint using its IPv4 address
- **Elastic Load Balancing** - route traffic to Load Balancer's IPv4 address
- **Amazon API Gateway** - route traffic to an API endpoint using IPv4 address
- **Amazon CloudFront** - route traffic to CloudFront distribution using IPv4/IPv6 address
- **AWS Elastic Beanstalk** - route traffic to Elastic Beanstalk environment using IPv4 address
- **Amazon Lightsail** - route traffic to a Lightsail instance using IPv4 address
- **Amazon WorkMail** - route traffic to WorkMail using MX, TXT, and CNAME record
- **Amazon RDS** - route traffic to DB instance using its domain name
- **Amazon S3** - route traffic to a static website hosted on an S3 bucket using S3 website endpoint

Hosted Zones

For easy routing management, hosted zones are configured to hold routing information. You can create a hosted zone by specifying your domain and the zone type. A hosted zone can be either of the two:

- **Public Hosted Zones** - traffic are routed on the internet
- **Private Hosted Zones** - traffic are routed within Amazon VPC



Route 53 Health Checks

Route 53 can monitor the health status of multiple resources for more efficient traffic distribution. You can set up endpoint monitoring for your resources like web servers by sending HTTP/HTTPS or TCP requests. Aside from monitoring endpoints, Route 53 can also monitor other health checks and CloudWatch Alarm. Monitoring the other health checks can be helpful if you are monitoring the number of resources with the same function. You still have health checks (child) on the individual resources, but instead of receiving multiple alerts, you only receive alerts from the parent health check that monitors the other health checks.

For CloudWatch Alarms, Route 53 health check returns a healthy status for *OK* state and unhealthy status for *ALARM* state. For *INSUFFICIENT* alarm state, you can configure a health check to either return a healthy, unhealthy, or the last known status.

Name	<input type="text" value="webserver"/> i
What to monitor	<input checked="" type="radio"/> Endpoint i
	<input type="radio"/> Status of other health checks (calculated health check) i
	<input type="radio"/> State of CloudWatch alarm

Route 53 health check can also monitor resources listed on a DNS record set for DNS failover. Route 53 can monitor resources specified in a record and do the DNS queries only to the healthy resources.

Route 53 Records

When you create a hosted zone, you can create records where you specify record types of your resources as well as the routing policy.



Route 53 > Hosted zones > test.net > Create record

Quick create record [Info](#) [Switch to wizard](#) [Add another record](#)

▼ Record 1 [Delete](#)

Record name [Info](#) test.net Record type [Info](#) A – Routes traffic to an IPv4 address

Value [Info](#) Alias

Valid characters: a-z, 0-9, ! " # \$ % & ' () * + , - / ; ; < = > ? @ [\] ^ _ { } . ~

Enter multiple values on separate lines.

TTL (seconds) [Info](#) 300 Routing policy [Info](#) Simple routing

Recommended values: 60 to 172800 (two days)

[Cancel](#) [Create records](#)

When routing traffic to other AWS resources, Route 53 supports the use of Aliases.

Route traffic to [Info](#) Alias

[Choose endpoint](#)

Alias to API Gateway API

Alias to CloudFront distribution

Alias to Elastic Beanstalk environment

Alias to Application and Classic Load Balancer

Alias to Network Load Balancer

Alias to Global Accelerator

Alias to S3 website endpoint

Alias to VPC endpoint



Routing Policy

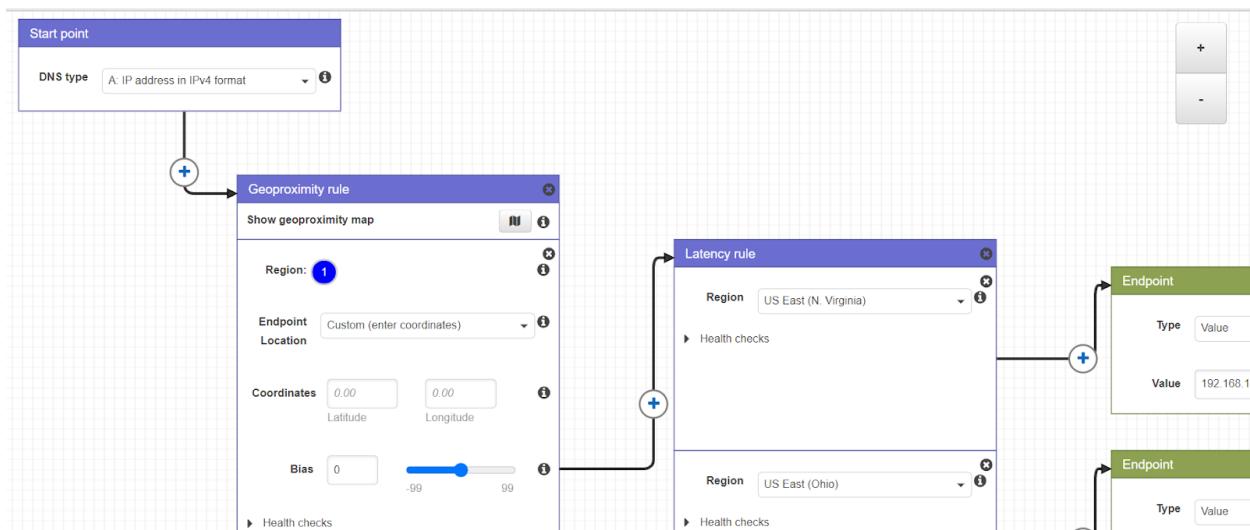
How Route 53 distributes traffic to your resources will depend on the routing policy you specified. Route 53 offers the following routing policies.

- **Simple Routing Policy** - routes traffic to a single resource
- **Failover Routing Policy** - lets you define a primary and secondary resource. Traffic is automatically transferred to secondary resources when the primary resource is unhealthy
- **Geolocation Routing Policy** - route user traffic to specific resources based on their location
- **Latency Routing Policy** - route traffic to your resource's AWS region with the lowest latency
- **Multivalue Answer Routing Policy** - supports multiple value return for DNS queries
- **Weighted Routing Policy** - route traffic by specifying the weight to various resources.
- **Geoproximity Routing Policy** - determines the traffic routing according to the location of the user and resources. It can be used on Traffic Flow only.

Routing policy Switch to quick create

<p><input checked="" type="radio"/> Simple routing Use if you're routing traffic to just one resource, such as a webserver.</p> 	<p><input type="radio"/> Weighted Use when you have multiple resources that do the same job, and you want to specify the proportion of traffic that goes to each resource. For example: two or more EC2 instances.</p> 	<p><input type="radio"/> Geolocation Use when you want to route traffic based on the location of your users.</p> 
<p><input type="radio"/> Latency Use when you have resources in multiple AWS Regions and you want to route traffic to the Region that provides the best latency.</p> 	<p><input type="radio"/> Failover Use to route traffic to a resource when the resource is healthy, or to a different resource when the first resource is unhealthy.</p> 	<p><input type="radio"/> Multivalue answer Use when you want Route 53 to respond to DNS queries with up to eight healthy records selected at random.</p> 

For complex DNS routing of many resources, you can use Route 53 Traffic Flow. It is a visual editor to help you create complex DNS routing like using multiple routing policies for different resources.



DNS Record Types

When adding a resource to a record, you specify the record type and its value. Route 53 supports the following record types.

Record Type	Value/Used for
A	IPv4 address
AAAA	IPv6 address
CAA	Verification of certificates used by a domain/subdomain
CNAME	Map a domain to another domain/subdomain
DS	Establishing the chain of trust for DNS Security Extensions (DNSSEC)
MX	mail server
NAPTR	Dynamic Delegation Discovery System (DDDS) applications
NS	Provides Name Servers for a hosted zone
PTR	Map IP address to Domain
SOA	Provide DNS information about a domain
SPF	Verification of email sender (Not recommended)
SRV	Accessing services by defining priority, weight, and port



TXT

For application-specific values and verification of email sender

Route 53 Resolver

Route 53 Resolver is a DNS resolver that automatically answers DNS queries within a VPC. By default, it resolves queries for

- EC2 instances domain names
- Records in Private Hosted Zones
- Records in Public Name Servers

Resolver Endpoints

If there is a need to set up DNS resolution between a VPC and an on-premise network, you will need to create Resolver Endpoints. You can create an Inbound Endpoint (going to Route 53 resolver), Outbound Endpoint (going to another DNS resolver), or both.

Basic configuration

Direction of DNS queries Info
You can configure endpoints for inbound DNS queries (to your VPC), outbound DNS queries (from your VPC), or both.

Inbound and outbound
Configure endpoints that allow DNS queries both to and from your VPC.

Inbound only
Configure an endpoint that allows DNS queries to your VPC from your network or another VPC.

Outbound only
Configure an endpoint that allows DNS queries from your VPC to your network or another VPC.



An on-premise network must be connected to your VPC via AWS Direct Connect/AWS VPN, or if with a different VPC, it must be connected using VPC Peering.



Inbound Endpoints - DNS Resolvers from on-premise network/peered VPC forward DNS queries to your Route 53 Resolver.

General settings for inbound endpoint

Endpoint name
A friendly name lets you easily find your endpoint on the dashboard.
The endpoint name can have up to 64 characters. Valid characters: a-z, A-Z, 0-9, space, _ (underscore), and - (hyphen)

VPC in the Region: ap-northeast-1 (Tokyo) [Info](#)
All inbound DNS queries will flow through this VPC on the way to Resolver. You can't change this value after you create an endpoint.
▼

Security group for this endpoint [Info](#)
A security group controls access to this VPC. The security group that you choose must include one or more inbound rules. You can't change this value after you create an endpoint.
▼

IP addresses [Info](#)
To improve reliability, Resolver requires that you specify two IP addresses for DNS queries. We recommend that you specify IP addresses in two different Availability Zones. After you add the first two IP addresses, you can optionally add more in the same or different Availability Zones.

▼ IP address #1

Availability Zone [Info](#)
The Availability Zone that you choose for inbound DNS queries must be configured with a subnet.
▼

Subnet [Info](#)
The subnet that you choose must have an available IP address. Only IPv4 addresses are supported.
▼

IP address [Info](#)
For inbound DNS queries, you can either let the service choose an IP address for you from the available IP addresses in the subnet, or you can specify the IP address yourself.

Use an IP address that is selected automatically
 Use an IP address that you specify

▼ IP address #2



Outbound Endpoints - Route 53 Resolver uses Rules to conditionally forward DNS queries to DNS Resolvers from on-premise network/peered VPC.

General settings for outbound endpoint

Endpoint name
A friendly name lets you easily find your endpoint on the dashboard.

The endpoint name can have up to 64 characters. Valid characters: a-z, A-Z, 0-9, space, _ (underscore), and - (hyphen)

VPC in the Region: ap-northeast-1 (Tokyo) [Info](#)
All outbound DNS queries will flow through this VPC on the way from other VPCs. You can't change this value after you create an endpoint.
▼

Security group for this endpoint [Info](#)
A security group controls access to this VPC. The security group that you choose must include one or more outbound rules. You can't change this value after you create an endpoint.
▼

IP addresses [Info](#)

To improve reliability, Resolver requires that you specify two IP addresses for DNS queries. We recommend that you specify IP addresses in two different Availability Zones. After you add the first two IP addresses, you can optionally add more in the same or different Availability Zones.

▼ IP address #1 Remove IP address

Availability Zone [Info](#)
The Availability Zone that you choose for outbound DNS queries must be configured with a subnet.
▼

Subnet [Info](#)
The subnet that you choose must have an available IP address. Only IPv4 addresses are supported.
▼

IP address [Info](#)
For outbound DNS queries, you can either let the service choose an IP address for you from the available IP addresses in the subnet, or you can specify the IP address yourself.

Use an IP address that is selected automatically
 Use an IP address that you specify

▼ IP address #2 Remove IP address



Resolver Rules

Rules are associated with an Outbound Endpoint to define how the DNS queries from your VPC will be forwarded.

- Conditional forwarding rules - forward DNS queries for specified domain names to DNS resolvers on your network. Queries to the specified domain name are forwarded to the defined Target IP Addresses
- System rules - handles the DNS queries for the specified domain names instead of the other DNS resolvers outside your VPC
- Recursive rule (Internet Resolver) - created automatically by the Resolver. Other domain names that are not defined explicitly on the rules go to this recursive resolver

Rule for outbound traffic
For queries that originate in your VPC, you can define how to forward DNS queries out of the VPC.

Name
A friendly name helps you find your rule on the dashboard.

The rule name can have up to 64 characters. Valid characters: a-z, A-Z, 0-9, space, _ (underscore), and - (hyphen)

Rule type Info
Choose **Forward** to forward DNS queries to the IP addresses that you specify in **Target IP addresses** section near the bottom of this page. Choose **System** to have Resolver handle queries for a specified subdomain. You can't change this value after you create a rule.

Domain name Info
DNS queries for this domain name are forwarded to the IP address that you specify in the **Target IP addresses** section near the bottom of the page. If a query matches multiple rules (example.com and www.example.com), outbound DNS queries are routed using the rule that contains the most specific domain name (www.example.com). You can't change this value after you create a rule.

Target IP addresses Info
DNS queries are forwarded to the following IPv4 addresses:

IP address	Port	Action
192.0.8.64	53	<input type="button" value="Remove target"/>
<input type="button" value="Add target"/>		



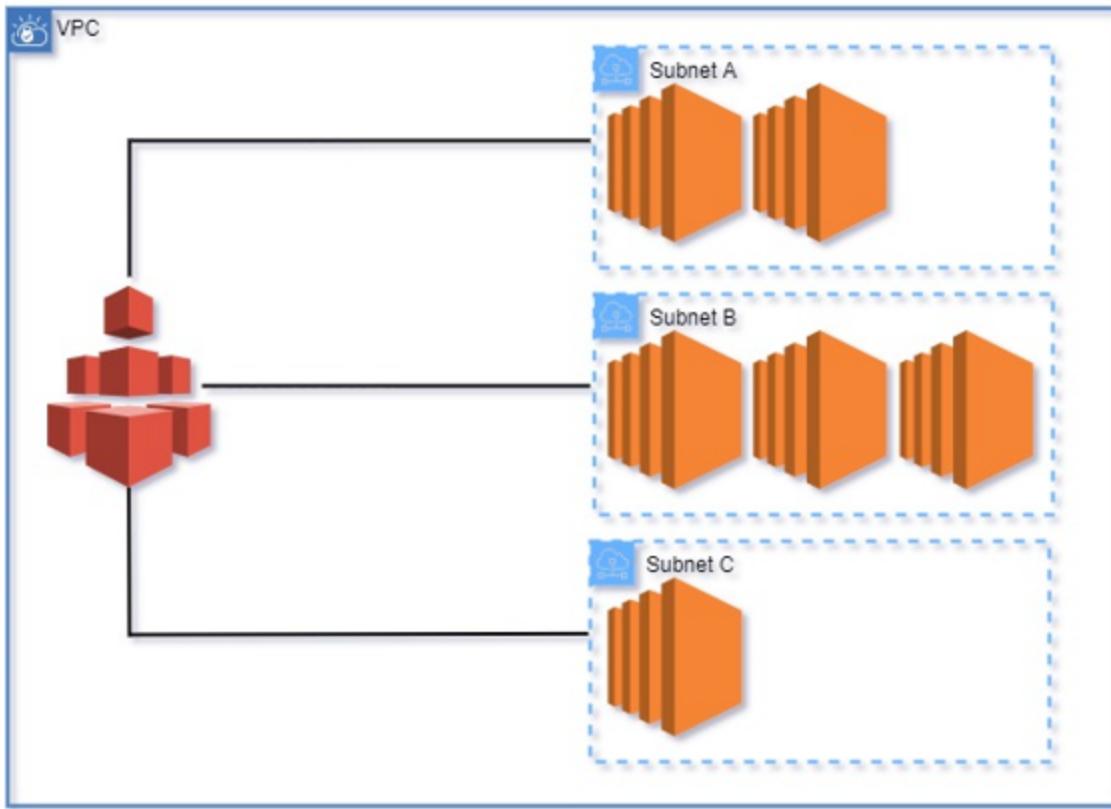
Reference:

<https://docs.aws.amazon.com/Route53/latest/DeveloperGuide>Welcome.html>

<https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/resolver.html>

Amazon Elastic File System (EFS)

Elastic File System (EFS) is a fully managed storage service from AWS that enables you to provision a simple and scalable file system that can be used with AWS Cloud services and on-premises resources. EFS can accommodate multiple connections all at the same time. For example, you can create an EFS on your VPC with various instances and mount it on the instances.



EFS Storage Classes

Like Amazon S3, Amazon EFS also has different storage classes to offer for different file system requirements. The EFS Classes differ on availability and durability.

- **EFS Standard** - frequently accessed data / multi-AZ storage
- **EFS Standard-Infrequent Access (IA)** - long term storage for infrequently accessed data / multi-AZ storage
- **EFS One Zone** - frequently accessed data / stored on single AZ
- **EFS One Zone-IA** - long term storage for infrequently accessed data / stored on single AZ



Creating a File System

To create an EFS, you need to specify a VPC and storage classes.

Create file system X

Create an EFS file system with service recommended settings. [Learn more](#)

Name - *optional*
Name your file system.

Name must not be longer than 256 characters, and must only contain letters, numbers, and these characters: + - = . _ : /

Virtual Private Cloud (VPC)
Choose the VPC where you want EC2 instances to connect to your file system. [Learn more](#)

vpc-5e1bff38
default

Availability and Durability
Choose Regional (recommended) to create a file system using regional storage classes. Choose One Zone to create a file system using One Zone storage classes. [Learn more](#)

Regional
Stores data redundantly across multiple AZs

One Zone
Stores data redundantly within a single AZ

Cancel Customize Create

More settings can be configured when you click on the Customize button.

Enabling automatic backups allows you to backup your file system using AWS Backup.

Automatic backups

Automatically backup your file system data with AWS Backup using recommended settings. Additional pricing applies. [Learn more](#)

Enable automatic backups

Lifecycle management

Automatically save money as access patterns change by moving files into the Standard - Infrequent Access storage class. [Learn more](#)

30 days since last access

Like S3, EFS also has Lifecycle management that moves files to IA storage classes (Standard-IA/One Zone-IA) to save on costs effectively. The default policy set for lifecycle management is 30 days, but this can be changed.



Lifecycle management

Automatically save money as access patterns change by moving files into the Standard - Infrequent Access storage class. [Learn more](#)

30 days since last access	
None	
7 days since last access	
14 days since last access	
30 days since last access	
60 days since last access	
90 days since last access	

You can also set the performance mode of your file system to either **General Purpose** or **Max I/O** with no additional cost. You can select General Purpose mode for latency-sensitive use cases and Max I/O if you want a throughput performance that scales up. Max I/O performance mode is not available for file systems with the One Zone storage class. Performance mode can't also be changed once the file system is created.

Performance mode

Set your file system's performance mode based on IOPS required. [Learn more](#)

General Purpose

Ideal for latency-sensitive use cases, like web serving environments and content management systems

Max I/O

Scale to higher levels of aggregate throughput and operations per second

EFS also has two throughput modes: bursting for throughput scaling and provisioned for fixed throughput.

Throughput mode

Set how your file system's throughput limits are determined. [Learn more](#)

Bursting

Throughput scales with file system size

Provisioned

Throughput fixed at specified amount



EFS uses AWS KMS service to encrypt the data at rest on the file system. You can also use a custom KMS key for the encryption.

Encryption

Choose to enable encryption of your file system's data at rest. Uses the AWS KMS service key (aws/elasticfilesystem) by default. [Learn more](#)

- Enable encryption of data at rest

▼ Customize encryption settings

KMS key

Choose or input a KMS key ID or ARN to use instead of the AWS KMS service key. [Learn more](#)

Choose an AWS KMS key or enter an ARN

[Create an AWS KMS key](#)

Depending on the availability you selected, your file system will have different mount targets per availability zone.

Network access

Network

Virtual Private Cloud (VPC)

Choose the VPC where you want EC2 instances to connect to your file system. [Learn more](#)

vpc-03c57950e44f8783b
vpc_webapp

Mount targets

A mount target provides an NFSv4 endpoint at which you can mount an Amazon EFS file system. We recommend creating one mount target per Availability Zone. [Learn more](#)

Availability zone	Subnet ID	IP address	Security groups
us-east-1c	subnet-0af258dc9096781ad	Automatic	Choose security groups ▾ Remove
			sg-07e64c62e3815f63a X default

[Add mount target](#)

For additional control, EFS also has a file system policy as an optional feature. The following are the current policies available which can also be customized using the policy editor.

- Prevent root access by default
- Enforce read-only access by default
- Prevent anonymous access
- Enforce in-transit encryption for all clients



File system policy - optional

Policy options

Select one or more of these common policy options, or create a custom policy using the editor. [Learn more](#)

Prevent root access by default*

Enforce read-only access by default*

Prevent anonymous access

Enforce in-transit encryption for all clients

* Identity-based policies can override these default permissions.

► **Grant additional permissions**

Policy editor {JSON}

```
1  {
2   "Version": "2012-10-17",
3   "Id": "efs-policy-wizard-6080e658-bf4d-4742-854e-1ea31f4ac029",
4   "Statement": [
5     {
6       "Sid": "efs-statement-e8834732-63c6-47e3-b405-57d0f3bf2914",
7       "Effect": "Allow",
8       "Principal": {
9         "AWS": "*"
10      },
11      "Action": [
12        "elasticfilesystem:ClientRootAccess",
13        "elasticfilesystem:ClientWrite"
14      ],
15      "Condition": {
16        "Bool": {
17          "elasticfilesystem:AccessedViaMountTarget": "true"
18        }
19      }
20    },
21    {
22      "Sid": "efs-statement-2a1c9146-e388-48a8-9853-93751fce3793",
23      "Effect": "Deny",
24      "Principal": {
25        "AWS": "*"
26      },
27      "Action": "*",
28      "Condition": {
29        "Bool": {
30          "aws:SecureTransport": "false"
31        }
32      }
33    }
34  ]
35 }
```

Manual changes will prevent the use of the policy options on the left until the editor is cleared.

Once the file system is created, you can view a summary of its details with its management and monitoring options.

File systems (2)										<input type="button" value="C"/>	<input type="button" value="View details"/>	<input type="button" value="Delete"/>	<input type="button" value="Create file system"/>
<input type="text"/> Filter by property values										<	1	>	@
Name	File system ID	Encrypted	Total size	Size in Standard / One Zone	Size in Standard-IA / One Zone-IA	Provisioned Throughput (MiB/s)	File system state	Creation time	Availability Zone				
<input type="radio"/> TD-EFS	fs-34df7e74	<input checked="" type="checkbox"/> Encrypted	6.00 KiB	6.00 KiB	0 Bytes	-	<input checked="" type="checkbox"/> Available	Sun, 09 May 2021 02:19:54 GMT	Regional				
<input type="radio"/> TutorialsDojo-EFS	fs-e8dc7da8	<input checked="" type="checkbox"/> Encrypted	6.00 KiB	6.00 KiB	0 Bytes	-	<input checked="" type="checkbox"/> Available	Sun, 09 May 2021 01:57:56 GMT	Regional				

EFS automatically scales as you add or remove files. You only pay for the storage used by the file system.



TutorialsDojo-EFS (fs-ffaf09bf)

Delete Attach Edit

General

Performance mode	Automatic backups
General Purpose	<input checked="" type="checkbox"/> Enabled
Throughput mode	Encrypted
Bursting	8e487659-81bc-48b1-8031-155559d8b330 (aws/elasticfilesystem)
Lifecycle policy	File system state
30 days since last access	<input checked="" type="checkbox"/> Available
Availability zone	
Regional	

Metered size Monitoring Tags File system policy Access points Network

Metered size

Total size	6.00 KiB
Size in Standard / One Zone	6.00 KiB (100%)
Size in Standard-IA / One Zone-IA	0 Bytes (0%)

Size in Standard / One Zone
Size in Standard-IA / One Zone-IA

File System Access Point

Access points are entry points for your applications to connect with your file system. Access points help you efficiently manage your application access. When creating an access point, you need to provide the File System Name/ID, Access Point Name (*Optional*), and Root directory path (*Optional*).



Create access point

An access point is an application-specific entry point into an EFS file system that makes it easier to manage application access to shared datasets. [Learn more](#)

Details

File system

Choose the file system to which your access point is associated.

 [Enter a file system name or ID](#)

Name - optional

 [EFS access point name](#)

Maximum of 256 Unicode letters, whitespace, and numbers, plus + - = . _ : /

Root directory path - optional

Connections use the specified path as the file system's virtual root directory [Learn more](#)

 [Defaults to /](#)

Example: "/foo/bar"

You also have an option to enable operating system (POSIX) identity for your access point by providing the User ID, Group ID, and Secondary group IDs.

POSIX user - optional

The full POSIX identity on the access point that is used for all file operations by NFS clients. [Learn more](#)

User ID

POSIX user ID used for all file system operations using this access point.

 [POSIX UID](#)

Accepts values from 0 to 4294967295

Group ID

POSIX group ID used for all file system operations using this access point.

 [POSIX GID](#)

Accepts values from 0 to 4294967295

Secondary group IDs

Secondary POSIX group IDs used for all file system operations using this access point.

 [Example: "123,456,789"](#)

A comma-separated list of valid POSIX group IDs



Furthermore, you can set permission for your file system root directory by providing the User ID, Group ID, and Permission (defined in octal number).

The screenshot shows a configuration panel for setting root directory creation permissions. It includes fields for Owner user ID, Owner group ID, and POSIX permissions, along with examples and validation notes.

Root directory creation permissions - optional

EFS will automatically create the specified root directory with these permissions if the directory does not already exist. [Learn more](#)

Owner user ID
Owner user ID for the access point's root directory, if the directory does not already exist.

POSIX user ID to apply to path
Accepts values from 0 to 4294967295

Owner group ID
Owner group ID for the access point's root directory, if the directory does not already exist.

POSIX group ID to apply to path
Accepts values from 0 to 4294967295

POSIX permissions to apply to the root directory path
Example: "0755"
An octal number representing the file's mode bits.

File System Security Group

In addition to the security group for your Linux instance to allow incoming ssh traffic on port 22, you also need to create another firewall rule that allows NFS traffic to your instance. The source will depend on your requirement.

The screenshot shows the 'Inbound rules' section of a security group configuration. It lists a single rule for NFS traffic on port 2049 from a specific source.

Inbound rules [Info](#)

Type	Protocol	Port range	Source
NFS	TCP	2049	Custom sg-089681b017da84fc7

[Add rule](#)

Mounting a File System

When launching an EC2 instance, you can select and mount an existing file system that you created. Notice that a script for the file system is automatically generated on the user data. If you haven't made a security



group for your file system, you have an option to allow AWS to create the required security group on your behalf. Please note that EFS only supports Linux instances.

Step 3: Configure Instance Details

Additional charges may apply

File systems (i) fs-34df7e74 | TD-EFS /mnt/efs/fs1 ×

Add file system C Create new file system



Additional security groups required

To enable access to the file system, the required security groups will be automatically created and attached to this instance and the selected file system's mount targets. To manually manage the security groups, clear the check box. [Learn more](#).

Automatically create and attach the required security groups.

Advanced Details

Enclave (i) Enable

Metadata accessible (i) Enabled

Metadata version (i) V1 and V2 (token optional)

Metadata token response hop limit (i) 1

User data (i) As text As file Input is already base64 encoded

```
- yum install -y nfs-utils
- apt-get -y install nfs-common
- file_system_id_1=fs-34df7e74
- efs_mount_point_1=/mnt/efs/fs1
- mkdir -p "${efs_mount_point_1}"
- test -f "/sbin/mount.efs" && printf "\n${file_system_id_1}:/
```

Additionally, you can mount your file system on an existing Linux instance via DNS or IP address. EFS provides both commands when using the EFS mount helper and the NFS client. You can view these commands by selecting the file system then clicking the Attach button.

Amazon EFS > File systems > fs-695a35dd

TD-EFS (fs-695a35dd)

Delete

Attach

General

Edit

Performance mode

General Purpose

Automatic backups

Enabled



Mount via DNS

Mount your Amazon EFS file system on a Linux instance. [Learn more](#)

Mount via DNS

Mount via IP

Using the EFS mount helper:

```
sudo mount -t efs -o tls fs-695a35dd:/ efs
```

Using the NFS client:

```
sudo mount -t nfs4 -o nfsvers=4.1,rsize=1048576,wsize=1048576,hard,timeo=600,retrans=2,noresvport fs-695a35dd.efs.us-east-1.amazonaws.com:/ efs
```

Mount via IP

Mount your Amazon EFS file system on a Linux instance. [Learn more](#)

Mount via DNS

Mount via IP

Availability zone

us-east-1a



Using the NFS client:

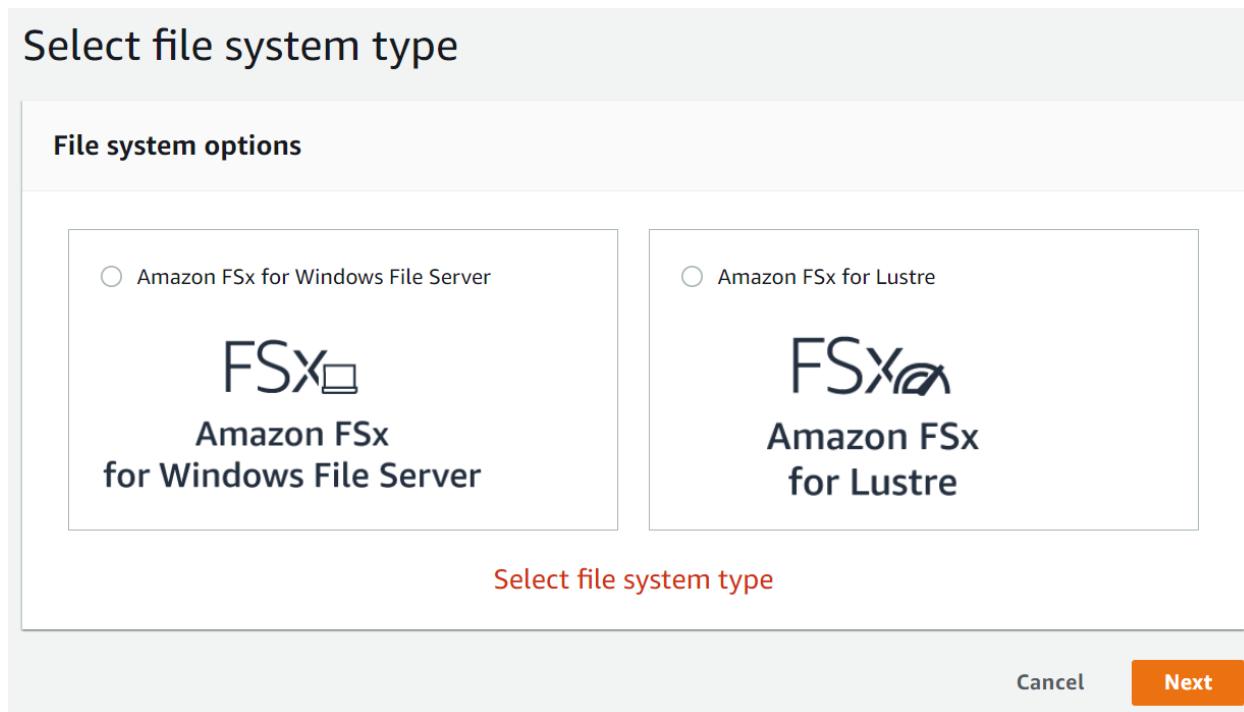
```
sudo mount -t nfs4 -o nfsvers=4.1,rsize=1048576,wsize=1048576,hard,timeo=600,retrans=2,noresvport 172.31.26.238:/ efs
```

Reference:

<https://docs.aws.amazon.com/efs/latest/ug/whatisefs.html>

Amazon FSx

Amazon FSx is a simple and fully managed file system service. It offers a fast and reliable file system performance while maintaining durability and scalability for its data. Underlying hardware and software are AWS managed and are continuously monitored. Amazon FSx also replicates the data on multiple devices within an Availability Zone. Amazon FSx supports two types of file systems: Windows File Server and Lustre File System.



Amazon FSx for Windows File Server

Amazon FSx provides a file system to Windows Operating System accessible through Server Message Block (SMB) Protocol. Like a native Windows file system, Amazon FSx for Windows also has management features like Active Directory integration and user quotas. The file system is also accessible outside AWS Environment through AWS Direct Connect or AWS VPN. Furthermore, it is also accessible by Linux and macOS devices. For faster access to Amazon FSx from an on-premises environment, you can use Amazon FSx File Gateway.

File System Details

When setting up an Amazon FSx for Windows Server, the following details are provided.



Deployment Type

Amazon FSx for Windows Server is available on the following deployment options.

- **Single-AZ** - creates the Windows File Server on a single Availability Zone only.
- **Multi-AZ** - creates the Windows File Server across multiple Availability Zones.

Storage Capacity and Storage Type

The minimum and maximum storage capacity vary depending on the storage type. To obtain much higher storage and throughput, you create multiple file systems and combine them using Microsoft Distributed File System (DFS). The following storage type is supported.

- **Solid State Drives (SSD)** - high performance and low latency workloads.
- **Hard Disk Drives (HDD)** - general workloads, user file sharing

Throughput Capacity

The data transfer rate can be configured on throughput capacity. You can select the AWS recommended throughput capacity (based on storage capacity) or specify the throughput capacity yourself. Throughput bursts for periods of time are also supported.

Virtual Private Cloud

The VPC is where your file system resides. The file system is also accessible from other VPC through peering or Transit Gateway and is accessible outside the AWS environment through AWS Direct Connect or AWS VPN. For Multi-AZ deployment, two subnet/availability zones are required for preferred and standby file systems. Amazon FSx creates an Elastic Network Interface for the instances to connect to the file system.

Windows Authentication

Amazon FSx for Windows Server integrates with Microsoft Active Directory for authentication and folder/file access control. You can select an AWS Managed Microsoft Active Directory or a Self-managed AD if you have it running in an Amazon EC2 instance.

Windows authentication

Choose an Active Directory to provide user authentication and access control for your file system [Info](#)

AWS Managed Microsoft Active Directory

Self-managed Microsoft Active Directory

Choose an AWS Managed Microsoft AD directory to use. [Info](#)

Choose a directory



Create new directory

Encryption

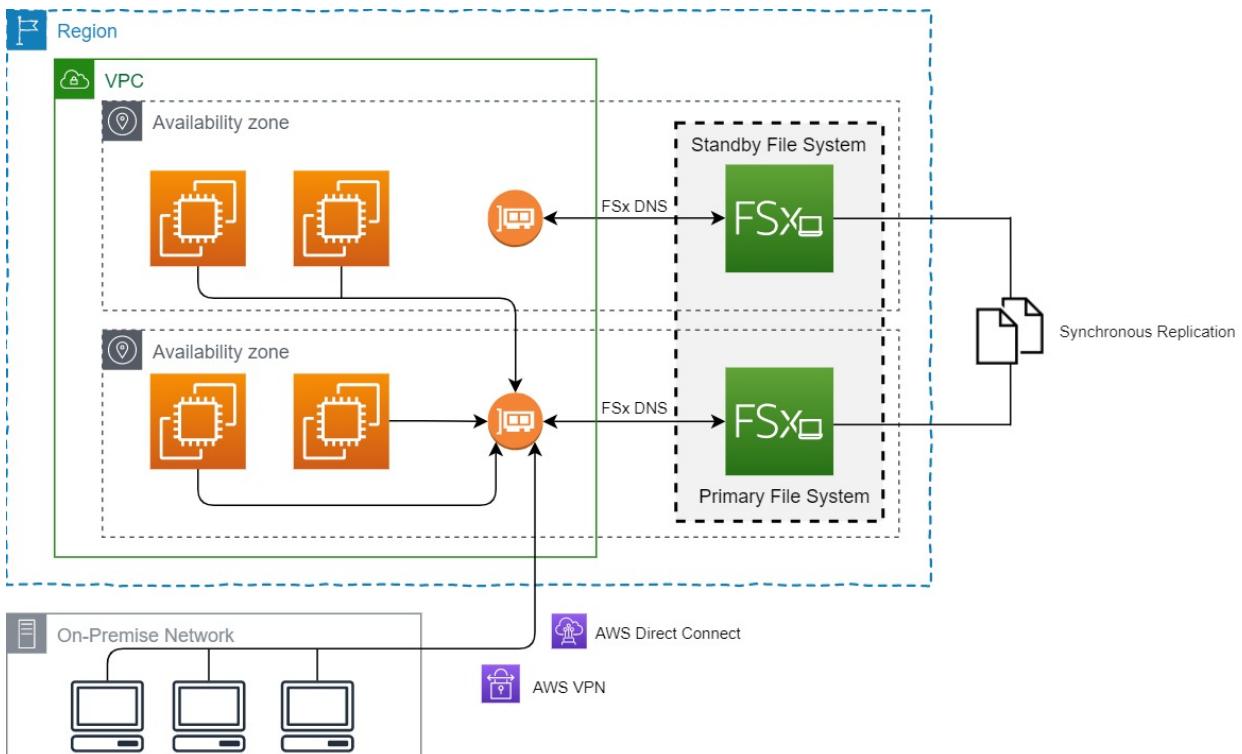
Amazon FSx integrates with AWS KMS for encryption. You can use the default aws/fsx key or provide another key ARN.

Optional Details

- **Auditing** - allow audit logging for files and folders. Log destination can be CloudWatch logs or Kinesis Data Firehouse
- **Access** - define additional DNS names for the file system
- **Backup and Maintenance** - configure the backup schedule and maintenance window
- **Tags** - Metadata/Label for AWS resources

Working with Amazon FSx for Windows File Server

The diagram below shows an Amazon FSx for Windows File Server with Multi-AZ deployment. The file system is created on both preferred and standby availability zones. Any changes on the primary file system are synchronously replicated to the standby file system. The instances access the primary file system through a network interface. In cases when the primary file system becomes unavailable, the Traffic is automatically migrated to the standby file system. The file system is also accessible from an on-premises network using AWS Direct Connect or AWS VPN.





Amazon FSx for Lustre

Lustre is an open-source, parallel file system used for High-Performance Computing (HPC). Through Amazon FSx, you can create a fast, scalable, and cost-effective file system capable of Lustre workloads. Amazon FSx for Lustre offers capabilities like sub-millisecond latencies, hundreds of gigabytes per second of throughput, and millions of IOPS. Amazon FSx for Lustre is POSIX-compliant and can be used for Linux applications. To mount Amazon FSx for Lustre, Linux instances should have Lustre Client installed.

File System Details

When setting up an Amazon FSx for Lustre, the following details are provided.

Deployment and Storage Type

Persistent deployment options have automatic data replication and file server failover capabilities, which is highly recommended for long-term data processing. On the other hand, the scratch deployment option is ideal for temporary, non-critical, and short-term processing. Storage also comes on SSD and HDD storage. The available deployment options and storage types are as follows.

- **Persistent, SSD** - for long term latency-sensitive workloads that require high IOPS/throughput
- **Persistent, HDD** - for long-term workloads that require high throughput but are not latency-sensitive.
Has an optional SSD cache feature.
- **Scratch, SSD** - for short-term workloads and non-critical data

Storage Capacity and Throughput Capacity

The data transfer rate can be configured on throughput capacity. The total throughput capacity depends on the storage capacity. The total throughput increases as the storage capacity increases.

- 50 MB/s/TiB (up to 1.3 GB/s/TiB burst)
- 100 MB/s/TiB (up to 1.3 GB/s/TiB burst)
- 200 MB/s/TiB (up to 1.3 GB/s/TiB burst)

*Throughput capacity = Storage capacity (TiB) * Per unit storage throughput (MB/s)*

Data Compression Type

Amazon FSx for Lustre uses the LZ4 algorithm for data compression. All new files are compressed before being written on the file system and are decompressed when being read. Because of the tremendous LZ4 algorithm, the compression and decompression process has zero to minimal effects on the file system performance.

Virtual Private Cloud

The VPC is where your file system resides. The file system is also accessible from other VPC through peering or Transit Gateway and is accessible outside the AWS environment through AWS Direct Connect or AWS VPN. Amazon FSx for Lustre doesn't come with Multi-AZ deployment, but the data is replicated within the Availability



Zone for persistent deployment. Amazon FSx creates an Elastic Network Interface for the instances to connect to the file system.

Encryption

Amazon FSx integrates with AWS for encryption. You can use the default aws/fsx key or provide another key ARN.

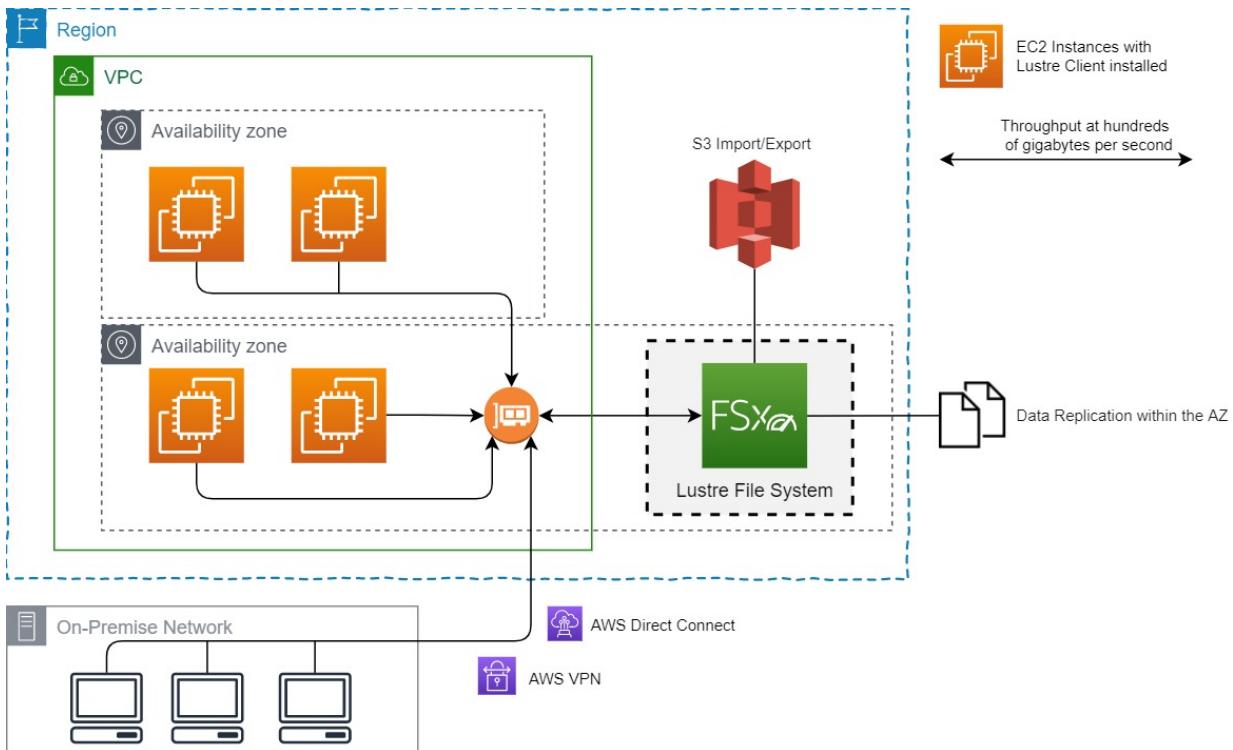
Optional Details

- **Data Repository Import/Export** - enable import data from S3/export data to S3
- **Backup and Maintenance** - configure the backup schedule and maintenance window
- **Tags** - Metadata/Label for AWS resources

Working with Amazon FSx for Lustre

The diagram below shows an Amazon FSx for Lustre in persistent deployment. The file system is being accessed by multiple instances across two availability zones through a network interface. Data is transferred from and to the file system at a throughput rate of hundreds of gigabytes per second. It also shows S3 integration for exporting/importing files in an S3 bucket.

The file system is also accessible from an on-premises network using AWS Direct Connect or AWS VPN.



References:

- <https://docs.aws.amazon.com/fsx/latest/WindowsGuide/what-is.html>
- <https://docs.aws.amazon.com/fsx/latest/LustreGuide/what-is.html>



AWS DataSync

Moving data from on-premises to cloud for data is a bit of a challenge. It could be for a Disaster Recovery (DR) requirement, data archiving, data migration, or any business requirement. AWS DataSync simplifies the process of moving data from on-premises to AWS Storage services and between AWS storage services. AWS DataSync, as the name implies, syncs the data between the data source and its destination. It is an online data transfer service that supports different file-sharing protocols like Network File System (NFS) and Server Message Block Protocol (SMB).

Supported AWS Storage Service

- AWS Snowcone
- Amazon S3
- Amazon EFS
- Amazon FSx for Windows File Server

Working with DataSync

A **DataSync Agent** running on the on-premises environment is required when transferring data to AWS storage service. The agent is not required when transferring data between AWS storage services. A DataSync Agent is a virtual machine that is deployed on the following hypervisors and in Amazon EC2.

- VMware ESxi
- Microsoft Hyper-V
- Kernel-based Virtual Machine (KVM)
- AWS EC2

A **Service endpoint** and **Activation key** is also required. A service endpoint defines where your agent will connect while the activation key associates the agent with your AWS account.

A **Task** is where you configure the source and destination by specifying the location type and its region. You also select the agent here if transferring data from on-premises. You also need to define an IAM role for both the source and destination; you can have AWS auto generate the required IAM role.



Configuration

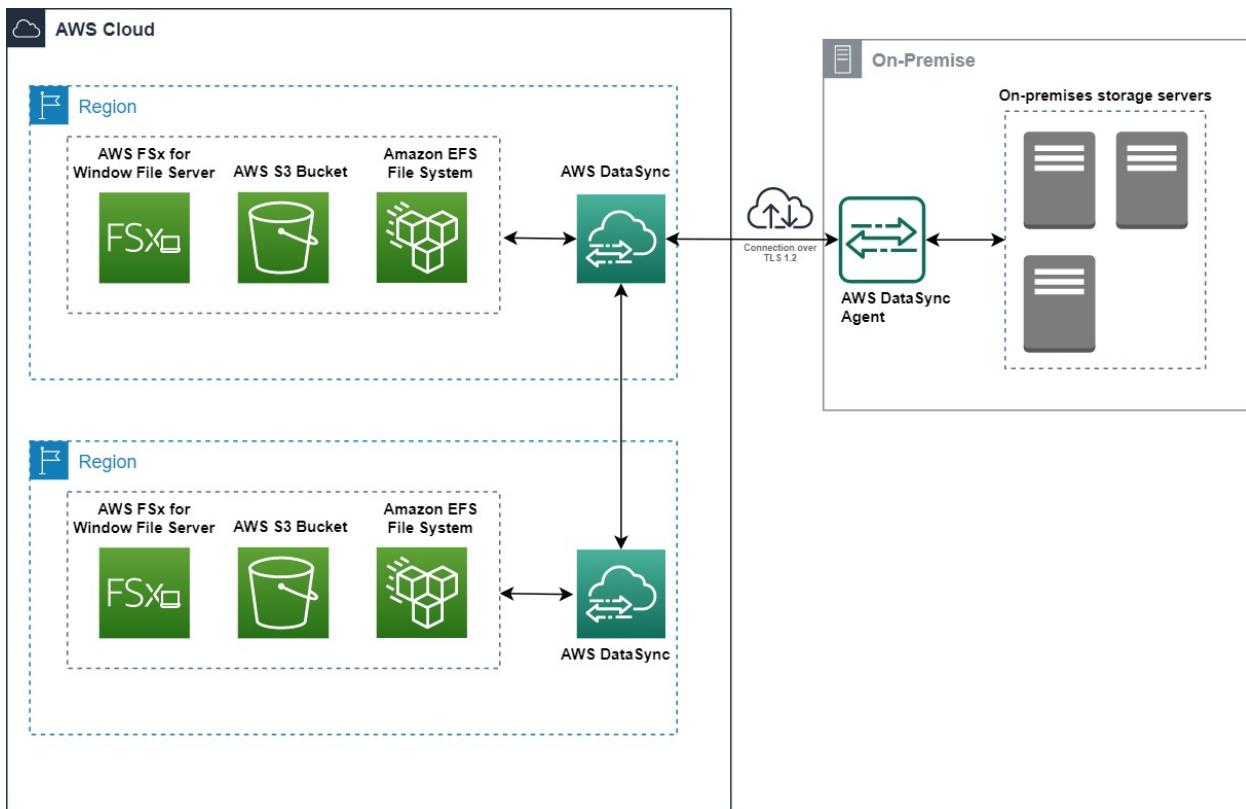
Location type

Amazon S3	▲
Amazon EFS file system	
Amazon FSx for Windows File Server	
Amazon S3	
Network File System (NFS)	
Object storage	
Server Message Block (SMB)	

Moreover, the following settings are configurable for a more refined data transfer task.

- **Data Verification** - AWS DataSync does a data integrity check during the transfer process.
- **Bandwidth limit** - sets the maximum bandwidth (MiB/s) for the task
- **Transfer Mode** - choose whether to copy all the data or only the data that has been changed. Set the behavior for the destination when a deleted file from the source and an existing file is detected.
- **Filter** - specify a folder or file you want to exclude
- **Schedule** - defines the time and frequency of the task
- **Logging** - enable logging of information to a CloudWatch log group

The diagram below shows a typical setup when transferring data from an on-premises storage device to any AWS storage service. The DataSync agent deployed on the on-premises environment enables the transfer process as the data travels to the AWS environment over the Internet using Transport Layer Security (TLS) 1.2 protocol. The diagram also shows the data transfer process between AWS storage (e.g., S3 to EFS, FSx to S3).



References:

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



AWS Backup

AWS Backup is a fully managed backup service that enables you to manage and automate the backup of your AWS resources in one place easily. You can do this by creating a backup plan, assigning an AWS Service to that backup plan, and having it monitored. The AWS Backup service allows you to view all your backup jobs on a single dashboard with details like backup and restoration job status for centralized monitoring.

Supported AWS Resources

- Aurora
- DynamoDB
- EBS
- EC2
- EFS
- FSx
- RDS
- Storage Gateway

You can also back up your Windows VSS-supported applications on Amazon EC2 using AWS Backup.

Backup Plan

This is where you define your backup name and backup rules. The backup rules indicate the schedule, frequency, and lifecycle rules of your backup. You can create a backup plan by either using a template, creating a new plan, or using JSON. Once a backup plan is created, you can then assign AWS resources to it. That AWS resources will be backed up according to its backup plan.



Create Backup plan Info

Start options

Choose how you want to begin. Info

- Start with a template
Create a Backup plan based on a template provided by AWS Backup.
- Build a new plan
Configure a new Backup plan from scratch.
- Define a plan using JSON
Modify the JSON expression of an existing backup plan or create a new expression.

Choose template
Choose a template plan with existing rules.

Choose a template ▾

- Daily-35day-Retention
- Daily-Monthly-1yr-Retention
- Daily-Weekly-Monthly-5yr-Retention
- Daily-Weekly-Monthly-7yr-Retention

[Cancel](#) [Create plan](#)

AWS Backup uses tagging to identify the resources that you want to include in your backup plan. Here in this example, all resources with the tag *EBSBackup:Yes* are included on this backup plan.



Assign resources Info

General

Resource assignment name

EBSVolumes

Resource assignment name is case sensitive. Must contain from 1 to 50 alphanumeric and '-' characters.

IAM role Info

AWS Backup will assume this IAM role when creating and managing recovery points on your behalf.

Default role

If the AWS Backup default role is not present, one will be created for you with the correct permissions.

Choose an IAM role

Assign resources

Assign resources to this Backup plan using tags and resource IDs.

Assign by

Tags

Key

EBSBackup

Value

Yes

Add assignment

Cancel

Assign resources

On-demand Backup

If you want to create a backup of your AWS resource immediately, you can do so through On-demand Backup. You just need to select the AWS resource to backup and define its lifecycle and retention rules.



Create on-demand backup Info

Settings

Resource type

EFS

File system ID

Choose a resource

Backup window

Create backup now

Starts within 1 hour.

Customize backup window

Transition to cold storage Info

Never

Retention period Info

Always

Backup vault Info

Specify the Backup vault this backup is organized in.

aws/efs/automatic-backup-vault

[Create new Backup vault](#)

IAM role Info

Specify the IAM role that AWS Backup will assume when creating and managing backups on your behalf.

Default role

If the AWS Backup default role is not present, one will be created for you with the correct permissions.

Choose an IAM role

► Tags added to recovery points

Tags specified here are added to recovery points when they are created. Tags on the resource will be copied automatically.

[Cancel](#)

[Create on-demand backup](#)

Backup Vault

Backup Vault is where your backups are stored. It uses KMS for encryption to further secure the backups.



Backup vaults (2) Info		
Backup vaults are containers where your backups are stored. You can have one default vault or multiple vaults where backups can be stored.		
Create Backup vault		
Backup vault ▾ <input type="text"/> Filter < 1 > ⚙️		
Backup vault name	Recovery points	KMS encryption key ID
EBSBackupVault	0	d6284805-7764-485b-a0b5-b5afae275c60 🔗
aws/efs/automatic-backup-vault	6	d6284805-7764-485b-a0b5-b5afae275c60 🔗

Protected Resources

This is a list of AWS resources that are being backed up.

Protected resources (3) Info		
Resources backed up by AWS Backup		
Create on-demand backup		
<input type="text"/> Filter < 1 ... > ⚙️		
Resource ID	Resource type	Last backup
file-system/fs-097a1abd	EFS	Jul 6, 2021, 1:09:15 PM UTC+08:00
file-system/fs-b65b3102	EFS	Jul 11, 2021, 1:36:44 AM UTC+08:00
file-system/fs-f15b3145	EFS	Jul 11, 2021, 1:37:51 AM UTC+08:00

Backup Jobs

Jobs hold the records of your scheduled and on-demand backups as well as recovery jobs.



Backup jobs <small>Info</small>				C	Last 24 hours ▾
Backup job ID	Status	Resource ID	Resource type	Creation time	Start by
da081bfb-4ad3-478f-8986-23f38dfc2457	⌚ Completed	file-system/fs-f15b3145	EFS	Jul 11, 2021, 1:37:51 AM UTC+08:00	Jul 11, 2021, 2:37:51 AM UTC+08:00
6dcbff76-8805-44ba-b5fc-377bd720fbca	⌚ Completed	file-system/fs-b65b3102	EFS	Jul 11, 2021, 1:36:44 AM UTC+08:00	Jul 11, 2021, 2:36:44 AM UTC+08:00
0cae2a16-b5fb-4f14-86cb-a32e972fd22b	⌚ Running	volume/vol-0fb58aac145ddf0f3	EBS	Jul 11, 2021, 1:34:27 AM UTC+08:00	Jul 11, 2021, 6:34:27 AM UTC+08:00

Cross-account Management

By integrating AWS Backup to AWS Organizations, you will be able to do a cross-account backup and monitoring. It also allows you to create and use backup policies across different AWS accounts.

Reference:

<https://docs.aws.amazon.com/aws-backup/latest/devguide/whatisbackup.html>



Amazon Relational Database Service (RDS)

AWS introduced Amazon Relational Database Service (RDS) to cater to the most challenging and demanding requirements of setting up a database on the cloud. RDS eases the process of creating and maintaining a relational database on the cloud with notable advantages in availability, flexibility, performance, and cost when compared to databases in a traditional environment.

Amazon RDS Features and Components

Amazon Relational Database Service (RDS) is a managed relational database service. Having a managed relational database service means that maintaining and managing the physical infrastructure for the databases is offloaded from the customer and is being taken care of by AWS allowing customers to focus on the database itself. This includes fault management, software patches, backups, and recoveries. To further understand RDS, let's break down its components and features.

Amazon RDS Database Engines

Amazon RDS supports the following relational database engines.

- **Amazon Aurora** - Amazon's proprietary database engine compatible with MySQL and PostgreSQL but with much higher performance. Amazon Aurora offers both performance and high availability of enterprise databases at a much lower cost.
- **PostgreSQL** - an open-source object-relational database known for being reliable and stable. This database engine supports both SQL (relational) and JSON (non-relational) queries.
- **MySQL** - is an easy-to-use, reliable, and very responsive database, making it the most popular relational database engine.
- **MariaDB** - also a well-known high-performing open-source relational database created by the same people who developed MySQL. MariaDB is compatible with MySQL databases.
- **Oracle** - is a commercial multi-model database management system. Built a superior reputation in terms of performance, flexibility, availability, and security over the years. Widely used for transactional database workloads as well as data warehousing. RDS offers both *Bring-Your-Own-License (BYOL)* and *License Included* licensing models when setting up Oracle Database on RDS.
- **Microsoft SQL Server** - a relational database management system offering from Microsoft. Widely used for transactional workloads, analytics, and business intelligence.



Engine type [Info](#)

Amazon Aurora



MySQL



MariaDB



PostgreSQL



Oracle

ORACLE

Microsoft SQL Server



Choosing Suitable RDS DB Instance Classes

A DB instance is a primary component of RDS. It is a secured, isolated environment for your database. Launching a database engine on RDS requires you to define a DB instance identifier unique within the Region. Every RDS API call and command is associated with a DB instance using its identifier.

RDS offers different DBS Instance classes to match the different processing power and memory requirements. The availability of the instance class options varies per database engine.

- **Standard** - General-purpose instance with balanced performance.
- **Memory-Optimized** - Designed for memory-intensive database workloads.
- **Burstable Performance** - has a baseline performance level that can burst to higher performance; ideal for unpredictable database workloads.

Choosing the Right RDS DB Instance Storage

Amazon RDS utilizes Amazon EBS volumes for its storage. The different storage types have distinct performance and cost; choosing the right storage type will highly depend on the database storage requirement.

- **General Purpose SSD (gp2) Storage** - cost-effective storage with burstable performance that suits most database workloads. Storage performance is relational to volume size i.e. the larger the volume, the better the performance.
- **Provisioned IOPS SSD (io1) Storage** - designed for I/O-intensive workloads that require low latency and consistent I/O throughput. Ideal for OLTP workloads with consistent performance.
- **Magnetic (standard) Storage** - ideal for small workloads and infrequently used data. Magnetic storage uses hard disk drives (HDD). It doesn't offer much performance when compared to gp2 and io1 storage.



Storage

Storage type [Info](#)

Provisioned IOPS (SSD)



Allocated storage

100

GiB

Minimum: 100 GiB, Maximum: 65,536 GiB

Provisioned IOPS [Info](#)

3000

IOPS

Minimum: 1,000 IOPS, Maximum: 80,000 IOPS

Right-sizing database storage is also a crucial part, especially for databases with unpredictable workloads. In these cases, it is ideal to use RDS storage autoscaling. On this feature, RDS will monitor and trigger autoscaling according to the following factors.

- Free space is less than 10% of the total storage.
- The low-storage condition lasts at least 5 minutes.
- At least 6 hours have passed since the last storage modification.

Storage autoscaling [Info](#)

Provides dynamic scaling support for your database's storage based on your application's needs.

Enable storage autoscaling

Enabling this feature will allow the storage to increase once the specified threshold is exceeded.

Maximum storage threshold [Info](#)

Charges will apply when your database autoscales to the specified threshold

1000

GiB

Minimum: 101 GiB, Maximum: 65,536 GiB



Choosing a Region and Availability Zone for RDS Instance

When choosing the Region for your DB instance, it is vital to consider the region of the application connecting to the database and its end users. AWS Regions are isolated and completely independent from other Regions. Likewise, an Availability Zone (AZ) is isolated from different Availability Zones within a Region.

You can also extend an AWS Region to your end-users using Local Zones. A local zone can host services like compute services, databases, and storage. A local zone is set up by creating a VPC Subnet in your Region and assigning it to a Local Zone.

Increasing Database Availability Using Multi-AZ Deployment

A database holds a collection of essential data. Always consider a Multi-AZ deployment, especially for the production databases, as it offers high availability and failover features. Though the odds are minimal, there is always a chance that an Availability Zone will go down. The worst case is an offline production database because of a faulty Availability Zone. Database engines like MySQL, MariaDB, Oracle, and PostgreSQL DB instances leverage Amazon's failover technology for Multi-AZ deployment. At the same time, Microsoft SQL Server uses SQL Server Database Mirroring (DBM) or Always On Availability Groups (AGs) for high availability. Just note that you can't use Local Zones for Multi-AZ deployment.

You can enable Multi-AZ deployment during the database creation or when editing the DB instance for existing databases. When the Multi-AZ deployment is enabled, RDS creates a standby replica in an Availability Zone that is different from the primary database. RDS automatically handles the synchronous replication between the primary database and replicas and the failover process during unexpected disruption on the primary database.

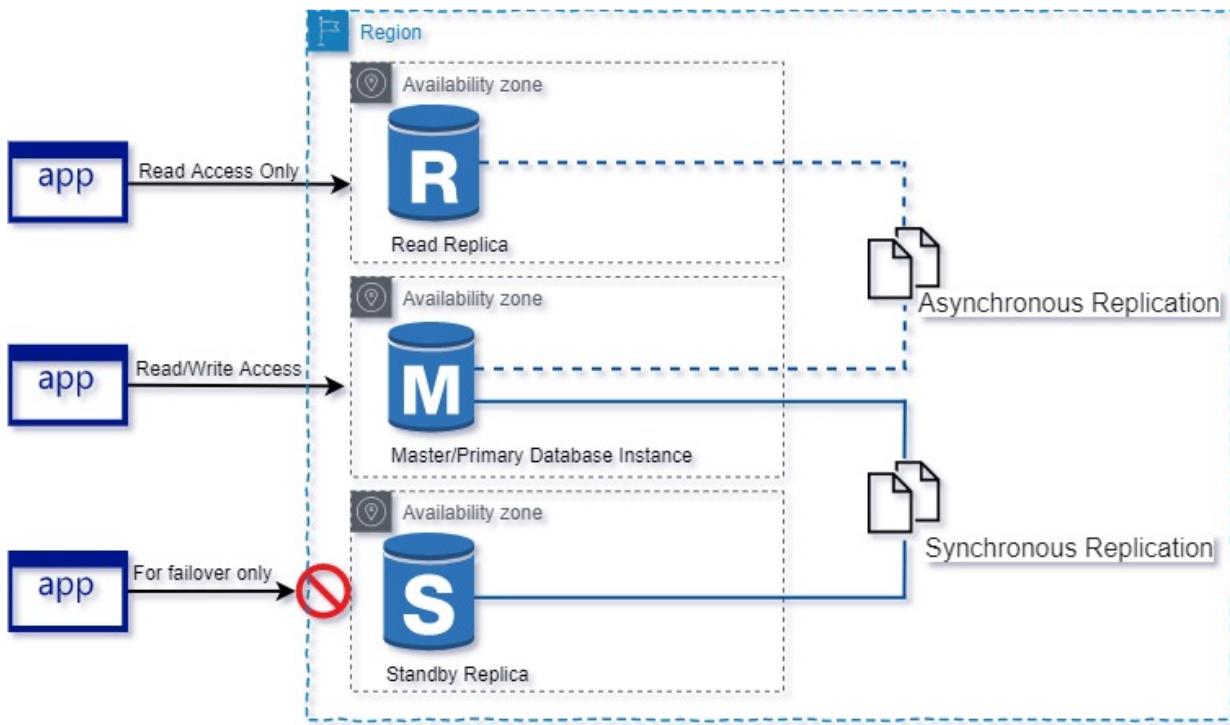
Availability & durability

Multi-AZ deployment [Info](#)

- Create a standby instance (recommended for production usage)
Creates a standby in a different Availability Zone (AZ) to provide data redundancy, eliminate I/O freezes, and minimize latency spikes during system backups.
- Do not create a standby instance

Amazon Aurora has durability and high availability features native to it. Amazon Aurora always keeps a copy of the data in a DB cluster that spans multiple Availability Zones within a Region. Additionally, Amazon Aurora has replication capabilities that are set up to either Single-master or Multi-master. A Single-master replication can

create up to 15 Aurora Replicas or Read Replicas. If a primary database becomes problematic, a replica will be promoted as the primary database.



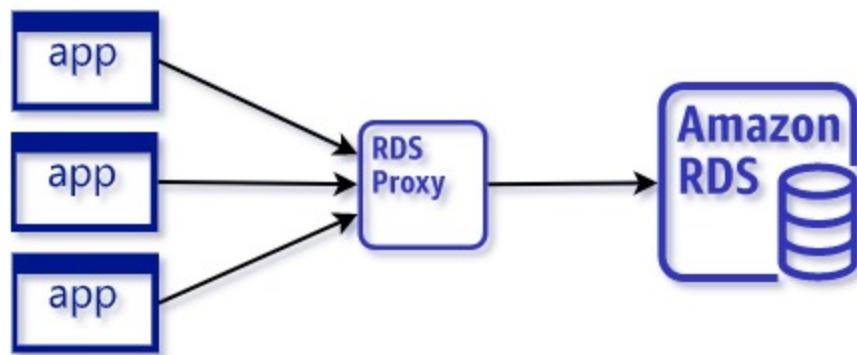
Improving Database Performance using Read Replica and DB Clusters

A Read Replica improves the availability and increases the performance of your database at the same time. You can do this by routing most of the read queries of your application to a read replica instead of the primary database. Changes made on the primary database are asynchronously replicated to its read replica. The database engines MySQL, MariaDB, Oracle, PostgreSQL, Microsoft SQL Server have built-in replication features that RDS uses to create read replicas.

A database cluster is a collection of database servers that shares storage to improve performance and availability. Similarly, an Amazon Aurora DB Cluster consists of one or more database instances with a cluster volume that spreads to multiple Availability Zones. This cluster volume's size increases as needed. RDS replicates the Primary DB Instance to create multiple Aurora Replicas. These replicas can accept read queries to minimize the queries on the primary database, significantly improving the database performance and improving availability, as discussed earlier.

Adding an RDS Proxy

If you expect your application to have a high volume of connections to your database or have an application with unpredictable workloads, you might want to use a proxy. Some applications have a high rate of opening and closing a database connection or otherwise keep a database connection on idle status, leading to high CPU and memory utilization. You can over-provision your database to accommodate all the workloads, or better yet, add an RDS proxy.



RDS proxy operates between your application and database to manage all the connections between these two. You can set a certain percentage of connections for your database according to its number of connection limits; thus, you can expect that your database will consistently process a good percentage of connections avoiding a sudden burst on the number of connections. You can also set a connection timeout for idle connections. RDS proxy enables an application to reuse the pooled connections to minimize the need to establish a new connection. RDS proxy is MySQL and PostgreSQL-compatible and can also require a connection over TLS.



Target group configuration

A target group is a collection of databases that the proxy can connect to. Currently, you can associate each target group with a single RDS DB instance or Aurora DB cluster.

Database

Choose the RDS DB instance or Aurora DB cluster that you want to associate with the proxy.

Choose database ▾

Connection pool maximum connections Info

Specify the maximum allowed connections, as a percentage of the maximum connection limit of your database.

100

Percent

Specify the maximum allowed connections, as a percentage of maximum connection limit of your database. For example, if you have set the maximum connections to 5,000 connections, specifying 50% will allow the proxy to create up to 2,500 connections to the database.

Include reader endpoint Info

Add reader endpoint

Choose whether or not to include a reader endpoint upon creation of this proxy.

Working with RDS Backup

Amazon RDS takes a snapshot of the storage volume used by the DB instance when you enable the automated backup feature. The automated backup allows you to make a Point-in-Time Recovery (PITR) when the necessity occurs.

Backup

Enable automated backups

Creates a point-in-time snapshot of your database

RDS will create a snapshot based on the specified backup window. The retention period can also be set from 0 to 35 days.



Backup retention period [Info](#)

Choose the number of days that RDS should retain automatic backups for this instance.

7 days



Backup window [Info](#)

Select the period for which you want automated backups of the database to be created by Amazon RDS.

- Select window
- No preference

Start time

00



: 00



UTC

Duration

0.5



hours

Copy tags to snapshots

If needed, you can manually take a snapshot at any given time.

RDS > Snapshots > Take snapshot

Take DB Snapshot

Preferences

To take a DB Snapshot, choose a DB Instance and name your DB Snapshot.

DB Instance

DB Instance identifier. This is the unique key that identifies a DB Instance.

td-sample-db-instance2



Snapshot Name

Identifier for the DB Snapshot.

backup-td-sample-db

Snapshot identifier is case insensitive, but stored as all lower-case, as in "mysnapshot". Cannot be null, empty, or blank. Must contain from 1 to 255 alphanumeric characters or hyphens. First character must be a letter. Cannot end with a hyphen or contain two consecutive hyphens.

Cancel

Take snapshot



Amazon Aurora offers a Backtrack option to “rewind” a DB cluster. Backtrack is a way to undo data changes in minutes compared to restoring a DB cluster using a snapshot that could take hours to complete. Backtrack can rewind a DB cluster to its previous state for as long as 72 hours.

Backtrack

Backtrack lets you quickly rewind the DB cluster to a specific point in time, without having to create another DB cluster. [Info](#)

Enable Backtrack

Enabling Backtrack will charge you for storing the changes you make for backtracking.

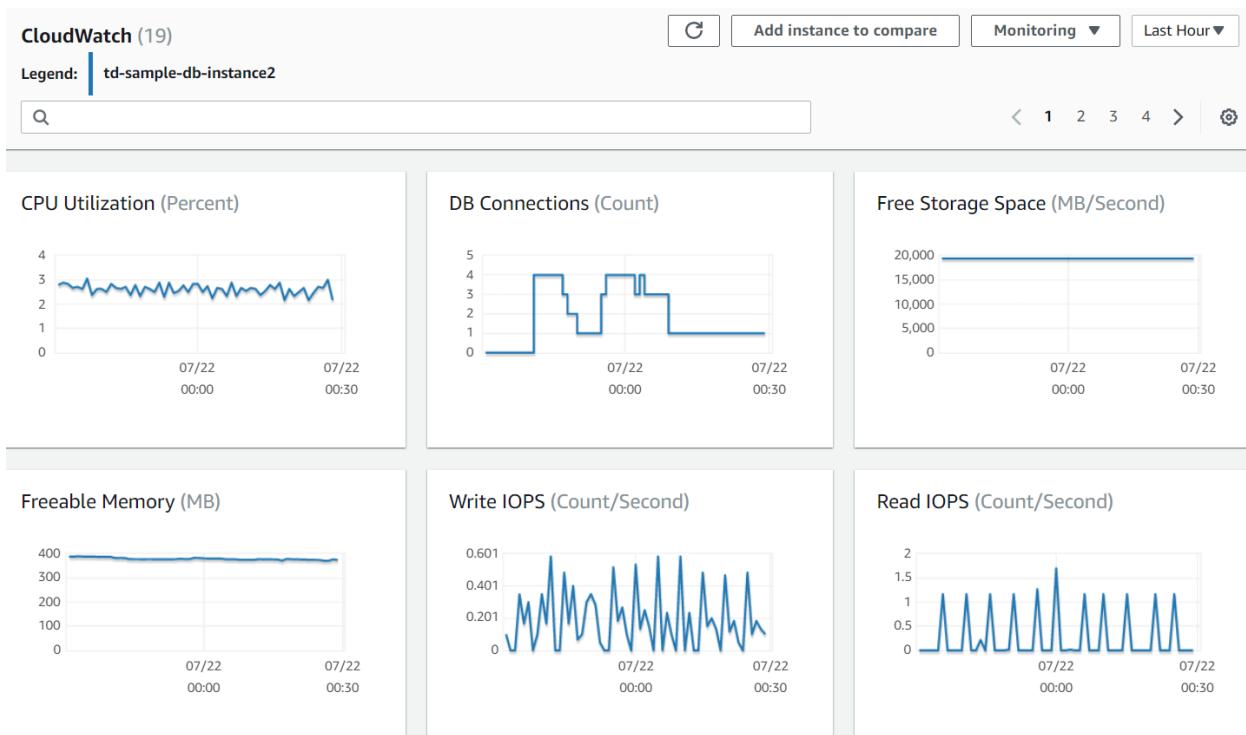
Target Backtrack window

The Backtrack window determines how far back in time you could go. Aurora will try to retain enough log information to support that window of time. [Info](#)

 hours (up to 72)

Monitoring a Database Instance

Amazon RDS integrates with different AWS services to provide efficient monitoring of your databases. Metrics of the active database are sent to Amazon CloudWatch every minute. You can then create Cloudwatch Alarms using these metrics. The default monitoring that CloudWatch provides seems to be enough for testing of non-critical databases.



For production and databases with critical workloads, it is best to enhance monitoring options for a thorough and granular performance monitoring. Moreover, database log files can be stored and accessed using Amazon CloudWatch Logs. Moreover, AWS CloudTrail logs all the activities on RDS by recording all API calls. These features are helpful when troubleshooting or investigating particular database issues.



Monitoring

Enable Enhanced monitoring

Enabling Enhanced monitoring metrics are useful when you want to see how different processes or threads use the CPU

Granularity

60 seconds



Monitoring Role

default



Clicking "Create database" will authorize RDS to create the IAM role rds-monitoring-role

Log exports

Select the log types to publish to Amazon CloudWatch Logs

- Audit log
- Error log
- General log
- Slow query log

Performance Insights provides a dashboard to visualize and analyze your database's performance. This feature allows you to view essential insights like database sessions, queries, database load, and top charts for users, connections, queries, and others.

Performance Insights [Info](#)

Enabling Performance Insights will automatically enable the MySQL Community performance schema.
[Learn more](#)

Enable Performance Insights

Retention period [Info](#)

You can also create a rule on Amazon EventBridge to monitor and log events of your database using an AWS Lambda function.



Event matching pattern
You can use pre-defined pattern provided by a service or create a custom pattern

Pre-defined pattern by service
 Custom pattern

Service provider
AWS services or custom/partner services

AWS

Service name
The name of partner service selected as the event source

Relational Database Service (RDS)

Event type
The type of events as the source of the matching pattern

RDS DB Instance Event

Event pattern

```
1 {  
2   "source": ["aws.rds"],  
3   "detail-type": ["RDS DB Instance Event"]  
4 }
```

Copy Edit

▶ Sample event(s)

Deleting a Database Instance

Like other AWS services, Amazon RDS also has a deletion protection feature to avoid accidental deletion of the database.

Deletion protection

Enable deletion protection

Protects the database from being deleted accidentally. While this option is enabled, you can't delete the database.

This database has deletion protection option enabled



To be able to delete the database, modify the database and disable deletion protection.

Close

You can trigger a database deletion by selecting a database and clicking *Delete* from the Actions.



RDS > Databases > td-sample-db-instance

td-sample-db-instance

Summary

DB identifier	CPU	Status	Class
td-sample-db-instance	3.33%	Available	db.t2.micro
Role	Current activity	Engine	Region
Instance	0 Connections	MySQL Community	us-east-1

Actions ▾

- Stop
- Reboot
- Delete**
- Create read replica
- Promote
- Take snapshot
- Restore to point in time

You will have an option to take a snapshot of your database before deleting a database instance in case you want to keep a backup. Confirmation should also be done by typing “delete me.”

Delete td-sample-db-instance instance? X

Are you sure you want to Delete the **td-sample-db-instance** DB Instance?

Create final snapshot?
Determines whether a final DB Snapshot is created before the DB instance is deleted.

Final snapshot name
The DBSnapshotIdentifier of the new DB Snapshot created.

To confirm deletion, type *delete me* into the field

Cancel **Delete**

Reference:

<https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html>



AWS Config

AWS Config is a service that keeps track of the configuration of your AWS resources. This service allows you to continuously monitor and assess configurations beneficial for audit and compliance requirements. AWS Config keeps the records of resource configuration in an S3 bucket. Additionally, aside from AWS resources, AWS config also supports a wide array of third-party resources to monitor. Resources are tracked according to their record type and category. You also have an option to discover and watch all resources in an entire region.

Resource types to record

- Record all resources supported in this region

- Record specific resource types

To learn more, see [Supported Resource Types](#).

Resource category

- All resource categories
- All resource categories
- Aqua resources
- Atlassian resources
- AWS resources
- AWSQS resources

Resource type

- All resource types

- Choose a role from your account

AWS Config stores the configuration details of its monitored resources to an **Amazon S3** bucket. Each record type has its configuration history. You can also trigger an **Amazon SNS** to notify you of any changes in the configurations.

AWS Config Continuous Configuration Monitoring

Resources are monitored and evaluated according to a set of desired configurations called **Rules**. You can use a predefined but customizable AWS managed rule or create your own Custom Rule using the AWS Lambda function.



Select rule type

Add AWS managed rule
Customize any of the following rules to suit your needs.

Create custom rule
Create custom rules and add them to AWS Config. Associate each custom rule with an AWS Lambda function, which contains the logic that evaluates whether your AWS resources comply with the rule.

Resources evaluation happens when a configuration change occurs, and during the schedule you set. AWS Config marks the resources as *COMPLIANT* or *NON_COMPLIANT* according to its config rules. See below example of an AWS-managed rule for EC2.

<input type="radio"/>	desired-instance-type	EC2	Checks whether your EC2 instances are of the specified instance types.
<input type="radio"/>	ebs-optimized-instance	EC2	Checks whether EBS optimization is enabled for your EC2 instances that can be EBS-optimized.

For reported non-compliant resources, a remediation action can be applied either automatically or manually. AWS also recommends remediation for the non-compliant resource.

▼ Remediation action details

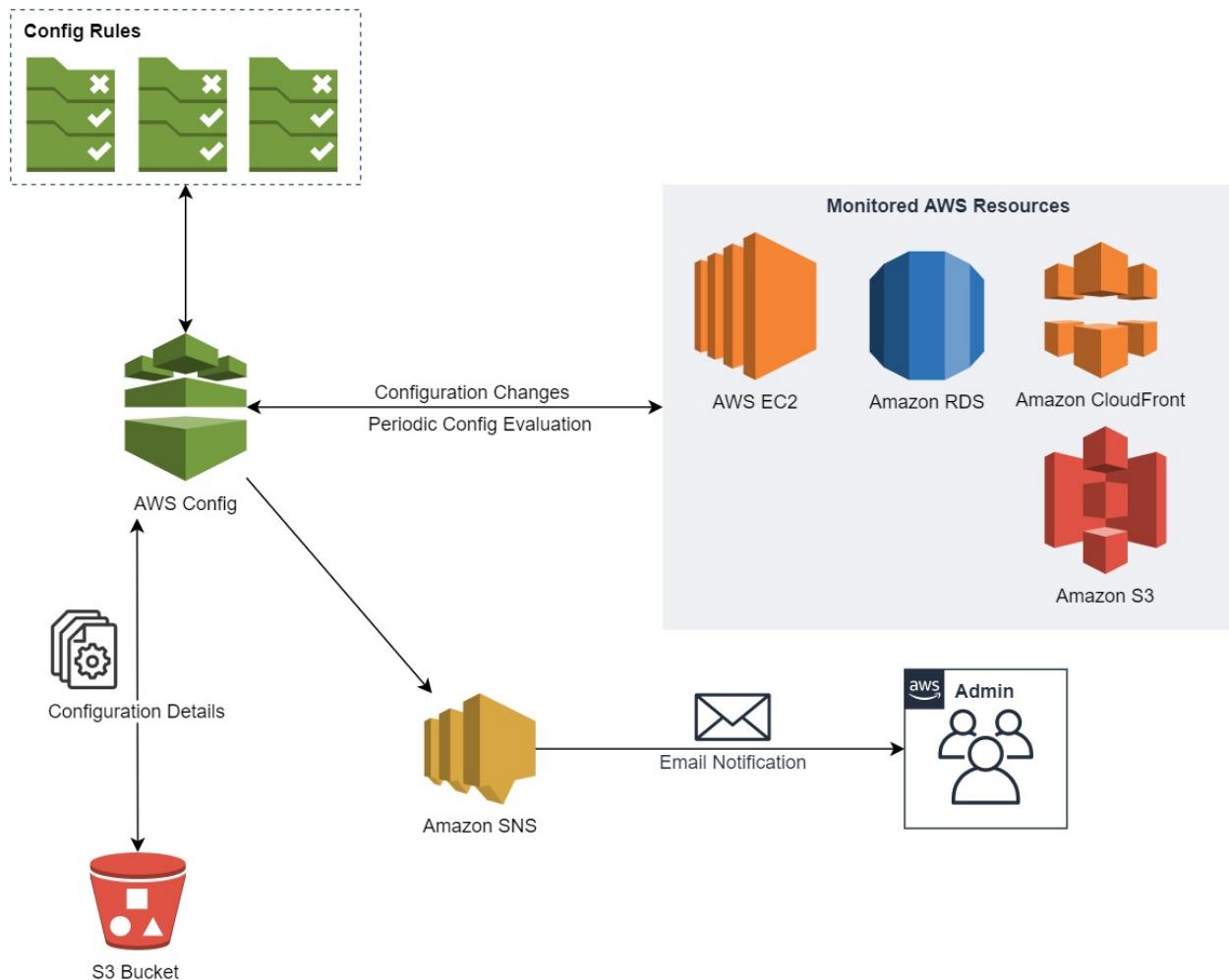
The execution of remediation actions is achieved using AWS Systems Manager Automation

Choose remediation action

Remediation action
<input type="text"/>
<input checked="" type="checkbox"/> AWS-CreateJiralIssue (recommended)
<input checked="" type="checkbox"/> AWS-PublishSNSNotification (recommended)
AWS-ASGEnterStandby
AWS-ASGExitStandby
AWS-AddOpsItemDedupStringToEventBridgeRule
AWS-AttachEBSVolume

To further broaden the scope of your configuration monitoring, you can set up an **AWS Config Aggregator**. Data aggregation will allow you to collect all configuration data across multiple AWS accounts and Regions and consolidate it all in one place.

The diagram below shows a high-level architecture of AWS Config. Multiple AWS resources are being evaluated using the rules periodically or in the event of configuration changes. It also shows the configuration details being stored in an Amazon S3 bucket and email notifications sent via Amazon SNS.



Reference:

<https://docs.aws.amazon.com/config/latest/developerguide/WhatIsConfig.html>



Deploying Resources with CloudFormation

AWS CloudFormation lets you model and create resources for your environment using programming language, leveraging the concept of Infrastructure as Code (IaC). You don't need to make all of your resources one by one manually; these resources are all defined on the AWS CloudFormation template. In this way, an AWS environment can be reliably and quickly reproduced since everything is declared on a template. An example would be creating a new Test environment similar to your current setup or creating a Disaster Recovery environment in another region.

With AWS CloudFormation, you can upload your template, use a sample template, or create a template using Designer. These templates can be in JSON or YAML format. Resources defined from these templates are treated and managed as a single unit called stacks. A template may contain the following sections, with Resources being only the required section.

- **Format Version** - AWS CloudFormation template version
- **Description** - Template description
- **Metadata** - Additional details of the template
- **Parameters** - values used by the template during stack creation
- **Rules** - used to validate the parameters being used by the template
- **Mappings** - matches a key to its corresponding value
- **Conditions** - specifies the condition required for creating resources
- **Transform** - specifies the version of the AWS Serverless Application Model
- **Resources** - identifies the stack resources and their properties
- **Outputs** - defines the output value

StackSets

AWS CloudFormation is used mainly for automating deployments of different applications. For creating resources for multiple regions and multi-accounts deployment, you should consider using StackSets. With StackSets, you can create resources for the Target Accounts using an Administrator Account. The administrator account centrally manages the templates, and you only specify the target accounts to where you want to create your resources.

Nested Stacks

As your infrastructure grows, there will be some cases where you need to declare the same resources to multiple CloudFormation templates. In these cases, it is better to use nested stacks. You can create separate templates for the most used resources and reference them on other templates. This way, you'll avoid copying and pasting the same configuration on your templates, and this also simplifies stack updates.

Below is an example CloudFormation Template main stack in YAML format.



```
AWSTemplateFormatVersion: 2010-09-09
```

Resources:

MyStack:

Type: AWS::CloudFormation::Stack

Properties:

TemplateURL: <https://aws-account.testbucket.amazonaws.com/s3bucketcreate.yaml>

TimeoutInMinutes: 60

Outputs:

StackRef:

Value: !Ref MyStack

OutputFromNestedStack:

Value: !GetAtt MyStack.Outputs.BucketName

The example CloudFormation template below is referenced on **AWS::CloudFormation::Stack** using its S3 object URL, allowing the **GetAtt** function to pull out the output values from the referenced template.

s3bucketcreate.yaml (uploaded on S3)

```
AWSTemplateFormatVersion: 2010-09-09
```

Resources:

SampleBucket:

Type: AWS::S3::Bucket

Outputs:

BucketName:

Value: !Ref 'SampleBucket'

Description: This is a sample bucket

Deleting a Stack

Deleting a stack on CloudFormation also removes all the provisioned resources in it. In some cases, you want some resources to be retained even after deleting its stack. The good thing is that you can do this by defining its *DeletionPolicy*.

To keep the resources when deleting a stack, you need to define its *DeletionPolicy* with *Retain* value on the template. You can set Snapshot as its value for the resources that support snapshot (like RDS databases). With *DeletionPolicy: Snapshot*, a snapshot is created before a resource is deleted. In this way, you will have a backup of the resource deleted from the stack.

Retain

Adding *DeletionPolicy: Retain* on the template will retain the provisioned resources even after deleting its stack.



```
AWSTemplateFormatVersion: 2010-09-09
```

Resources:

 SampleBucket:

 Type: AWS::S3::Bucket

DeletionPolicy: Retain

Outputs:

 BucketName:

 Value: !Ref 'SampleBucket'

 Description: This is a sample bucket

Snapshot

DeletionPolicy: Snapshot: Snapshot can be added on resources that support snapshots like the following:

```
AWS::EC2::Volume
```

```
AWS::ElastiCache::CacheCluster
```

```
AWS::ElastiCache::ReplicationGroup
```

```
AWS::Neptune::DBCluster
```

```
AWS::RDS::DBCluster
```

```
AWS::RDS::DBInstance
```

```
AWS::Redshift::Cluster
```

In this example, we have a Volume created along with an EC2 instance. Because Snapshot is defined as its DeletionPolicy, we expect this to create a snapshot when the stack is deleted.

```
AWSTemplateFormatVersion: 2010-09-09
```

Resources:

 Ec2Instance:

 Type: AWS::EC2::Instance

 Properties:

 ImageId: ami-0615132a0f36d24f4



TestVolume:

Type: AWS::EC2::Volume

DeletionPolicy: Snapshot

Properties:

Size: 10

AvailabilityZone: !GetAtt Ec2Instance.AvailabilityZone

References:

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

<https://tutorialsdojo.com/aws-cloudformation-stacksets-and-nested-stacks/>

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

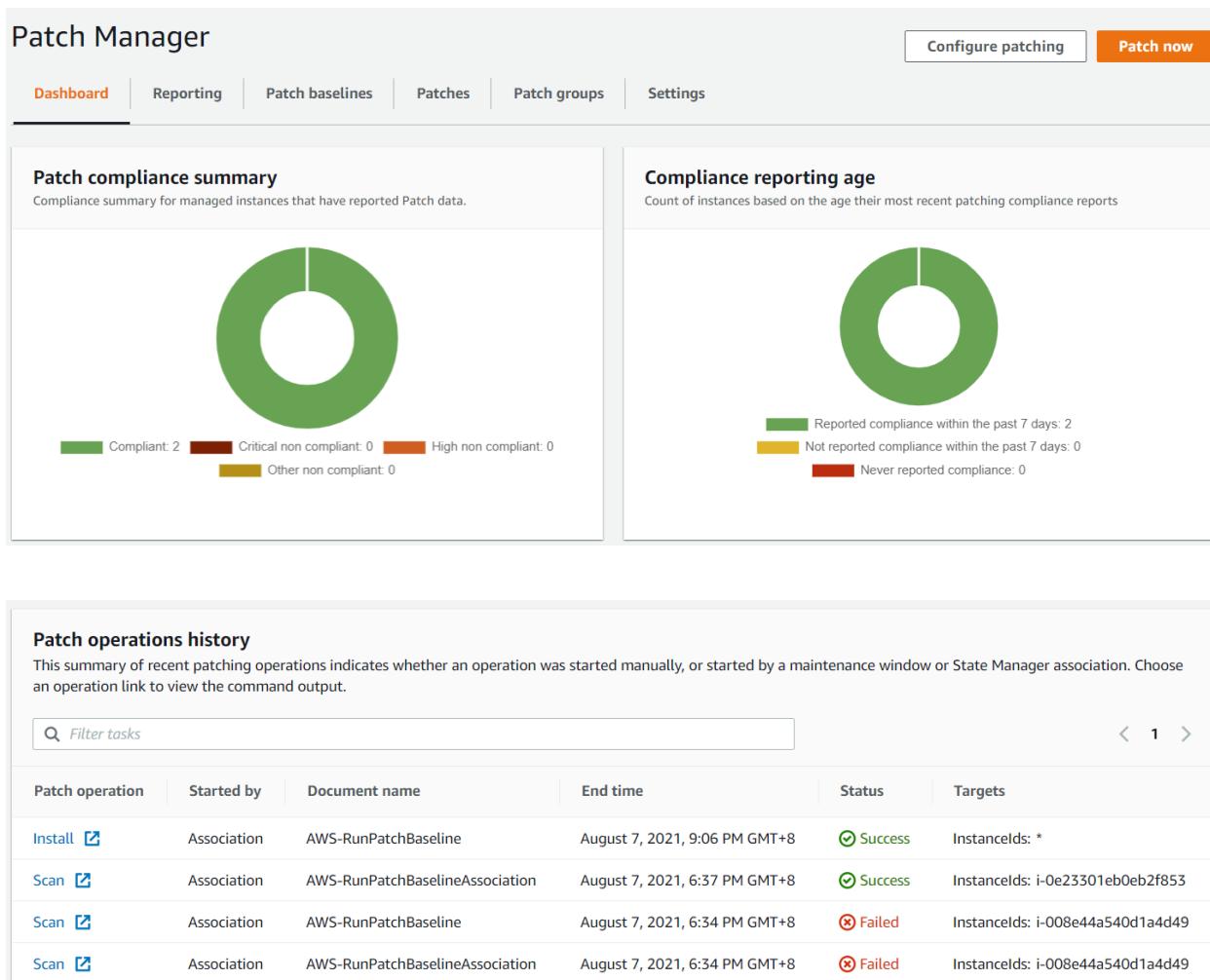


AWS Systems Manager Patch and Change Manager

AWS Systems Manager Patch Manager

Patch Manager is an AWS Systems Manager feature that helps you identify and install the necessary application and operating system updates on your managed instances. Managed instances can be EC2 instances or machines from an on-premises environment with SSM Agent installed.

Patch Manager provides a dashboard that gives essential information like compliance status and patch history.





To review the list of your managed instances, you can check **Systems Manager Fleet Manager**. The Fleet Manager consolidates all managed instances in a single view allowing you to do administration tasks like RDP/SSH, viewing file system, and user/group management.

To further understand how a Patch Manager works, here are the concepts that you need to know:

- **Patch Baseline** defines the patches that are approved for installation on your managed instances. AWS provides a list of Patch Baselines per operating system, but you can configure a custom Patch Baseline.
- When defining instances to patch, you can select instances manually, select instances using tags, or select a Patch Group. A **Patch Group** allows you to group identical instances for a particular patch baseline. You can create a patch by tagging your managed instances with the key: *Patch Group*.
- **Patch Operation** lets you choose between *Scan* and *Scan and Install* operation for your selected instances. Scan operation only identifies and generates a patched list while Install operation downloads and installs all approved patches on the instances.
- **Patch Schedule** defines the maintenance window for the patching. You can also trigger a patch instantly.

The Patch Manager provides a summary after a patch is executed.

Association execution summary	
AWS-PatchNowAssociation	
Association ID	Execution ID
05cf9c45-f891-4571-b4ac-5f602a0566c2	66906b90-1083-4af9-901c-dbcd9c15c345
Status	Operation
Success	Install
Reboot option	Targets
RebootIfNeeded	Instancelds: *
Summary	
Success=2	

AWS Systems Manager Change Manager

Change Manager is a change management capability of AWS Systems Manager. Change management is a practice done by many organizations to control changes in a given system. In an AWS environment, this capability simplifies submitting, reviewing, approving, and implementing operational changes to application and infrastructure configurations. The Change Manager can manage all changes across AWS Regions and AWS Accounts.



To further understand how a Change Manager works, here are the concepts you need to know:

- **Change Templates** contain information like level of approvals, runbook, and notification configuration. You can create a Standard and Emergency change template.
- **Change Request** is a request created from a change template. You can choose the request operation to start at a scheduled time or once the request has been approved.
- A **Runbook** contains the actions that will be performed for a particular request. Below is an example of a runbook for restarting an EC2 instance.

Runbook options

- Select a single runbook
- Define a set of runbooks that can be used
- Any runbook can be used

Runbook

Select an Automation runbook to add to your template.

AWS-RestartEC2Instance

Version 1

View

- **Change Request Approval** defines the needed approval for a particular request. The number of approval depends on the levels defined on the template.

Change request approvals

Specify up to five levels of approvers for change requests created from this change template. Each level can include one or more groups or individual users. All approvals from one level must be received before next-level approvers are notified.

First-level approvals

Approver

Required

Bob

1

[Remove](#)

1 approver to be specified at the request.

- You can add an **Approval Notification** using the SNS topic when configuring a template.



Amazon SNS topic for approval notifications - *optional*

Specify the Amazon SNS topic to notify approvers at this level. Make sure the approvers are subscribed to the topic.

- Enter an SNS Amazon Resource Name (ARN)
- Create an Amazon SNS topic
- Select an existing Amazon SNS topic

Topic ARN

Enter the topic ARN.

Must be a valid Amazon Resource Name (ARN).

Add notification

References:

<https://docs.aws.amazon.com/systems-manager/latest/userguide/systems-manager-patch.html>

<https://docs.aws.amazon.com/systems-manager/latest/userguide/systems-manager-actions-and-change.html>

|



Encryption on AWS Storage Services

S3 Encryption

Encryption is another security feature that S3 has as data protection. Users can opt to use Server-Side Encryption or Client-Side encryption, and both have their use cases depending on your storage requirements. S3 default encryption uses SSE and can be set on your bucket properties tab.

Default encryption
Automatically encrypt new objects stored in this bucket. [Learn more](#)

Default encryption
Disabled

Server-Side Encryption

When using Server-Side Encryption, the new objects uploaded in a bucket are encrypted as it is written on a disk and decrypted when being accessed. There are three different options to use for your encryption keys.

- **SSE-S3** - Encryption keys are managed by the S3 service. AES-256 encryption is used.

Default encryption
Automatically encrypt new objects stored in this bucket. [Learn more](#)

Server-side encryption
 Disable
 Enable

Encryption key type
To upload an object with a customer-provided encryption key (SSE-C), use the AWS CLI, AWS SDK, or Amazon S3 REST API.

Amazon S3 key (SSE-S3)
An encryption key that Amazon S3 creates, manages, and uses for you. [Learn more](#)

AWS Key Management Service key (SSE-KMS)
An encryption key protected by AWS Key Management Service (AWS KMS). [Learn more](#)



- **SSE-KMS** - Leverage AWS KMS capabilities. Users can use CMK (Customer Managed Key) stored on AWS KMS as encryption keys. You can create your own keys or use the default encryption for S3. This comes with additional charges since you are using another AWS service. Enabling the Bucket Key feature will reduce the number of calls to AWS KMS to save cost.

AWS KMS key

- AWS managed key (aws/s3)
arn:aws:kms:ap-southeast-1:947117271373:alias/aws/s3
- Choose from your AWS KMS keys
- Enter AWS KMS key ARN

Bucket Key

Reduce encryption costs by decreasing calls to AWS KMS for new objects in this bucket. To specify a Bucket Key setting for an object, use the AWS CLI, AWS SDK, or Amazon S3 Rest API. [Learn more](#)

- Disable
- Enable

- **SSE-C** - Customer-provided keys are used by S3 to encrypt objects.

Client-Side Encryption

While SSE is a great option to encrypt objects in S3, users may also choose to encrypt the objects before uploading to S3. You can use your own keys for encryption or use a CMK from AWS KMS.

Encrypting Existing S3 Objects

Only new objects uploaded on an S3 bucket are being encrypted when encryption is enabled. To encrypt the existing objects, you can use Amazon S3 Batch operations. This will allow you to identify and copy unencrypted objects, encrypt it, then write it on the same bucket.

EFS Encryption

Like other storage services, EFS also has encryption available for its data. EFS offers encryption both for Data at rest and Data in transit.

Data at Rest Encryption

You can enable encryption on a filesystem when creating it. Note that once a filesystem is created, you can't modify its encryption settings. EFS also uses KMS service to do the encryption; by default it uses `aws/elasticfilesystemkey` but you can also create your own keys.



Encryption

Choose to enable encryption of your file system's data at rest. Uses the AWS KMS service key (aws/elasticfilesystem) by default. [Learn more](#)

Enable encryption of data at rest

▼ Customize encryption settings

KMS key

Choose or input a KMS key ID or ARN to use instead of the AWS KMS service key. [Learn more](#)

Choose an AWS KMS key or enter an ARN

[Create an AWS KMS key](#)

Data In Transit Encryption

Using TLS when mounting your filesystem secures your data in transit. You can use the command provided on the EFS mount helper.

Attach

Mount your Amazon EFS file system on a Linux instance. [Learn more](#)

Mount via DNS

Using the EFS mount helper:

```
 sudo mount -t efs -o tls fs-097a1abd:/ efs
```

EBS Encryption

Elastic Block Storage serves as block storage volumes for EC2 instances. EBS has encryption for both data at rest and data in transit. Like other AWS services, EBS utilizes AWS KMS to handle the encryption.

Creating Encrypted EBS Volume

You can enable the volume encryption when launching an AWS instance. Both root and data volume can be encrypted.

Volume Type <small>(i)</small>	Device <small>(i)</small>	Snapshot <small>(i)</small>	Size (GiB) <small>(i)</small>	Volume Type <small>(i)</small>	IOPS <small>(i)</small>	Throughput (MB/s) <small>(i)</small>	Delete on Termination <small>(i)</small>	Encryption <small>(i)</small>
Root	/dev/xvda	snap-053c42bdb1128764a	8	General Purpose SSD (gp2)	100 / 3000	N/A	<input checked="" type="checkbox"/>	<input style="width: 100%; height: 100%; border: none; background-color: #f0f0f0; padding: 0; margin: 0;" type="button" value="Not Encrypt"/> <input style="width: 100%; height: 100%; border: none; background-color: #f0f0f0; padding: 0; margin: 0;" type="button" value="3f339edf-27e"/> <input style="width: 100%; height: 100%; border: none; background-color: #f0f0f0; padding: 0; margin: 0;" type="button" value="x"/>
EBS	/dev/sdb	Search (case-insensit	7	General Purpose SSD (gp2)	100 / 3000	N/A	<input type="checkbox"/>	



Likewise, you can also enable encryption during volume creation.

Create Volume

Volume Type General Purpose SSD (gp2) 

Size (GiB) 100 (Min: 1 GiB, Max: 16384 GiB) 

IOPS 300 / 3000 (Baseline of 3 IOPS per GiB with a minimum of 100 IOPS, burstable to 3000 IOPS) 

Throughput (MB/s) Not applicable 

Availability Zone* us-east-1a 

Snapshot ID Select a snapshot  

Encryption Encrypt this volume

Note that once a volume is created, you can only modify its volume type and size. You can't modify its encryption.

Snapshots

Snapshots are backups of your EBS volumes. The same keys are used in encrypting a volume and its snapshot. When you take a snapshot from an encrypted volume, the snapshot will automatically be encrypted. Same goes with the snapshots of unencrypted volume - the snapshot will also be unencrypted.

Another thing to note. You can create an encrypted volume from an unencrypted snapshot but you can't create an unencrypted volume from an encrypted snapshot.



Owned By Me		Filter by tags and attributes or search by keyword	?	K	<	1 to 2 of 2	>	>>
Status	Started	Progress	Encryption	KMS Key ID				
completed	July 11, 2021 at 1:26:23 A...	available (100%)	Encrypted	3f339edf-27ac...				
completed	July 11, 2021 at 1:26:29 A...	available (100%)	Not Encrypted					

RDS Encryption

Encryption for storage is a must especially for databases. On AWS Relational Database Services, encryption is available both for data at rest and data in transit for its database instances. It uses AES-256 encryption to secure the data on the RDS database instances.

Encrypting RDS Database Instance with AWS KMS

RDS also leverages AWS KMS for the encryption. During database creation, you have an option to enable the encryption and use your preferred key. For an encrypted database instance, all of its data will be encrypted including read replicas, snapshots, and automated backups.

Encryption using Transparent Data Encryption (TDE) is also supported for Oracle and SQL database instances. Take note that using TDE simultaneously with encryption at rest will lead to some performance issues.

Encryption

Enable encryption
Choose to encrypt the given instance. Master key IDs and aliases appear in the list after they have been created using the AWS Key Management Service console. [Info](#)

AWS KMS Key [Info](#)

(default) aws/rds ▾

Account

947117271373

KMS key ID

alias/aws/rds

Note that encryption is not available for the following database instance classes:



- General Purpose (db.m1.small, db.m1.medium, db.m1.large, db.m1.xlarge)
- Memory Optimized (db.m2.xlarge, db.m2.2xlarge, db.m2.4xlarge)
- Burst Capable (db.t2.micro)

Securing Database Connection on RDS

For data in transit encryption requirements, you can SSL/TLS for the connection between your application and your database instance. A server certificate is used to validate a connection to your database instance; these certificates are also rotated for additional security. The implementation of this method varies depending on the database being used.

References:

- <https://docs.aws.amazon.com/AmazonS3/latest/userguide/bucket-encryption.html>
- <https://docs.aws.amazon.com/efs/latest/ug/encryption.html>
- <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSEncryption.html>
- <https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Overview.Encryption.html>



Security on AWS

AWS KMS Customer Master Key (CMK) Rotation

AWS Key Management Service (AWS KMS) is a managed service that lets you create and manage customer master keys (CMKs). A *customer master key (CMK)* is a logical representation of a master key used to encrypt your data, making it a valuable and sensitive asset to your security. It makes sense, and it is a best practice to do a rotation on these keys.

AWS Managed Key has an automatic rotation feature enabled by default. This feature is optional for Customer Managed Key. However, automatic key rotation is not supported by the following:

- **Asymmetric CMK** where public and private key pairs are used to encrypt/decrypt or sign/verify operations
- CMKs in **custom key stores** where the key material is stored on AWS CloudHSM cluster
- **CMKs that have imported key material which is supported only for symmetric CMKs in AWS KMS key stores**

You may not have an automatic key rotation available on these CMK types, but you can still do a rotation manually.

AWS Owned CMK Rotation

AWS Owned CMK is, as the name implies, AWS owned and managed and is usable by multiple AWS Accounts. The key rotation for AWS Owned CMK varies depending on the AWS service that uses the CMK.

AWS Managed CMK Rotation

AWS Managed CMKs are the ones that are created on your behalf. You can't manage the key rotation for AWS Managed Keys. The automatic key rotation is handled by AWS KMS and is automatically rotated every three years.

AWS Customer Managed CMK Rotation

Unlike AWS Managed CMK, you have full control over Customer Managed CMK, including key rotation. By default, Customer Managed CMK can be rotated every year.



CMK Automatic Key Rotation

Type of CMK	Can view CMK metadata	Can manage CMK	Used only for your AWS account	Automatic Rotation
Customer Managed CMK	Yes	Yes	Yes	Optional. Every 365 days
AWS Managed CMK	Yes	No	Yes	Required. Every 1095 days
AWS Owned CMK	No	No	No	Varies

Tutorials Dojo

Secrets Manager vs Parameter Store

The best practices are always ideal to follow when storing essential and sensitive application information in the cloud. Information like database credentials, application keys, configuration, or any other security information that your applications consume are business-critical, hence should be secured at any given time. AWS recommends securely storing this sensitive information and only retrieving it when needed to avoid embedding credentials on application source code. Moreover, audit and key rotation are also advisable. AWS offers two relevant services for these requirements, each of them with particular use cases.

Secrets Manager – securely store, rotate, and monitor secrets. This service leverages AWS KMS for encryption-at-rest, which is enabled by default. Likewise, by default, Secrets Manager only allows programmatic retrieval of secrets over TLS and Perfect Forward Secrecy (PFS). Secrets can be credentials of databases like RDS, DocumentDB, Redshift, or other secrets like API keys.

Parameter Store – a feature of AWS Systems Manager that manages different configuration data like passwords, strings, and parameter values. Information on Parameter Store can be stored on plain text or as encrypted data using AWS KMS.



Parameter Stores offer two tiers:

Tier

Parameter Store offers standard and advanced parameters.

Standard

Limit of 10,000 parameters. Parameter value size up to 4 KB. Parameter policies are not available. No additional charge.

Advanced

Can create more than 10,000 parameters. Parameter value size up to 8 KB. Parameter policies are available. Charges apply

It also supports the following parameter type:

Type

String

Any string value.

StringList

Separate strings using commas.

SecureString

Encrypt sensitive data using KMS keys from your account or another account.

Here is a side-by-side comparison of Secrets Manager and Parameter Store.

	Parameter Store	Secrets Manager
Key-Value Size	4KB/8KB	10KB
Encryption	Yes	Yes
Reference in CloudFormation	Yes	Yes
Built-in password generation	No	Yes
Key Rotation	No	Yes
Cross-account Access	No	Yes
Cost	Free for standard parameters	Yes
Use Cases	Hierarchical storage for passwords, strings, and parameter	Business-critical secrets like database passwords. Dedicated



	values. Single store for configuration and secrets	secrets store with lifecycle management with rotation capabilities.
--	--	---

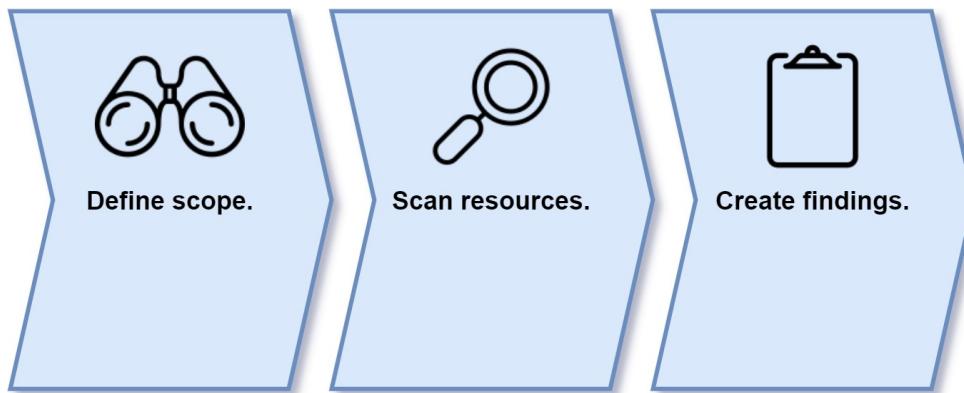
IAM Access Analyzer

The Access Analyzer is a feature of AWS IAM that evaluates the policies associated with AWS resources. It identifies the resources that are shared with an external entity. Moreover, you can preview your policies before proceeding with creation. IAM Access Analyzer will check your policy grammar and validate against the best practices. This feature is a simple yet effective way to point out unintended access, over-provisioned privileges, and other security-related risks. It helps you maintain the principle of least privilege on your resources.

Supported AWS Services

- **Amazon S3** - evaluates the bucket policies, access control list, and access point
- **AWS IAM** - evaluates the trust policies of an IAM role
- **AWS Lambda** - evaluates the policies associated with the functions
- **Amazon SQS** - evaluates the policies associated with a queue
- **AWS KMS** - evaluates the key policies and grants of keys
- **AWS Secrets Manager** - evaluates the policies associated with a secret

The following statements summarize the process of Access Analyzer.



1. Define the scope or zone of trust. The zone of trust will serve as a basis of the Access Analyzer; scope can be either within an AWS Account or Organization. Any entities outside the scope are treated as an external entity; this can be an AWS Account, IAM Account/Role, public entities.
2. Scan all resources policy within the trust zone and determine the resources shared outside the defined trust zone/scope.
3. Create findings to evaluate if access to a particular resource is intended or not.



Active Archived Resolved All

Active findings

Account ID 9471 [REDACTED]

Actions ▾

Filter active findings < 1 >

<input type="checkbox"/>	Finding ID	Resource	External principal	Condition	Shared through	Access level	Updated
<input type="checkbox"/>	d1fadccf-01...	S3 Bucket s3-ap-demo	All Principals	-	Bucket ACL	Read, List	a minute ago

Here is an example of an active finding on an S3 bucket that is publicly exposed. It shows essential details like the ARN, the policy, and the access level.

d1fadccf-0198-410d-a9bd-c85159210564 [Info](#)

⚠ Public: this finding is for a resource that allows public access.

Details

Finding ID d1fadccf-0198-410d-a9bd-c85159210564	Updated 38 minutes ago	Status Active	Shared through Bucket ACL
Resource arn:aws:s3:::s3-ap-demo	External principal All Principals	Condition -	Access level Read <ul style="list-style-type: none">s3:GetBucketAcl List <ul style="list-style-type: none">s3>ListBuckets3>ListBucketMultipartUploadss3>ListBucketVersions
Resource owner account 9471 [REDACTED]			

IAM Access Analyzer allows you to archive (for intended access) the findings or apply the necessary fix (not intended access) to the resource.

Next steps

Intended access	Not intended
If the access is intended, such as access necessary for business processes, you can archive the finding. This lets you focus on findings that are related to potential security risks. When you archive a finding, it's removed from Active findings and the status changes to ⊖ Archived.	If the access isn't intended, it indicates a potential security risk. Use the console for the service associated with the resource to modify or remove the policy that grants the unintended access. To confirm that your change removed the access, choose Rescan . If the access is removed, the status changes to ✓ Resolved.
Archive	Go to S3 console
To automatically archive similar findings, create an archive rule.	arn:aws:s3:::s3-ap-demo

Another popular use of the IAM Access Analyzer is when creating a bucket policy. IAM Access Analyzer will interactively check the syntax/grammar of the policy being created.



Bucket ARN
arn:aws:s3:::tutorialsdojo-test

Policy

```
1 ~ {
2     "Version": "2012-10-17",
3     "Statement": [
4         {
5             "Sid": "AddCannedAcl1",
6             "Effect": "Allow",
7             "Principal": {
8                 "AWS": [
9                     "arn:aws:iam::1331142225388:root",
10                    "arn:aws:iam::980146233025:root"
11                ]
12            },
13            "Action": [
14                "s3:PutObject",
15                "s3:PutObjectAcl",
16                "s3:CreateBucket"
17            ],
18            "Resource": "arn:aws:s3:::tutorialsdojo-test/*"
19        }
20    ]
21 }
```

+ Add new statement

Edit statement
AddCannedAcl Remove

1. Add actions
Choose a service Filter services

Included S3

Available AMP API Gateway API Gateway V2 Access Analyzer Account Activate Alexa for Business Amplify Amplify Admin

2. Add a resource Add

3. Add a condition (optional) Add

Using the Preview feature will trigger the IAM Access Analyzer to scan your policy and provide recommendations according to security best practices. See sample findings below.

JSON Ln 18, Col 50

⌚ Security: 0 ⚡ Errors: 0 ⚠ Warnings: 0 ⓘ Suggestions: 0

Preview external access

Preview external access - optional preview and validate Access Analyzer findings for external access to your resource. Learn more

Analyzer

TD-IAMAnalyzer
Zone of trust: Current account (947117271575)

Preview Last update: 3 minutes ago

All New Resolved Archived Existing

Filter new findings

▶ New An AWS account has write and permissions access.
▶ New An AWS account has write and permissions access.

Access preview ID: 27b4404c-653e-4924-a03e-946c0e61815f

AWS Certificate Manager

The AWS Certificate Manager (ACM) helps you easily provision and manage keys and certificates for your application or websites. With ACM, you can create both public and private SSL/TLS X.509 certificates.



Provision certificates

Provide the name of your site, establish your identity, and let ACM do the rest. ACM manages renewal of SSL/TLS certificates issued by Amazon or by your own private Certificate Authority.

[Get started](#)



Private certificate authority

You or your IT Administrator can establish a secure managed infrastructure for issuing and revoking private digital certificates. Private certificates identify and secure applications, services, devices and users within an organization.

[Get started](#)

AWS Certificate Manager (ACM) allows you to secure your public websites and applications over TLS. You can request new certificates with just a few clicks. You can even request certificates for multiple domains and wildcard domains. Before ACM issues the requested certificates, the request is validated through DNS or Email validation. ACM also handles the auto-renewal of certificates. If you want to use existing certificates, you can do so by importing them to ACM.

The ACM Certificates are free to use and can be integrated into the following AWS services.

- Elastic Load Balancing
- Amazon CloudFront
- AWS Elastic Beanstalk
- AWS App Runner
- Amazon API Gateway
- AWS Nitro Enclaves
- AWS CloudFormation

ACM Private CA allows you to create your certificate authority (CA) hierarchy in AWS. Certificates generated from ACM Private CA can only be privately used and accessible via ACM Private CA API or the AWS CLI. The ACM Private CA isn't free to use; you pay a monthly fee for each CA and certificate that you create and issue.



Select the certificate authority (CA) type



ACM helps you create a private subordinate CA.

- Root CA** Create a root CA. Choose this option if you want to establish a new CA hierarchy.
- Subordinate CA** Create a subordinate CA. Choose this option if you want to make a CA that is subordinate to an existing CA. You can use this option to create issuing CAs as well as intermediate CAs.

Cancel

Next

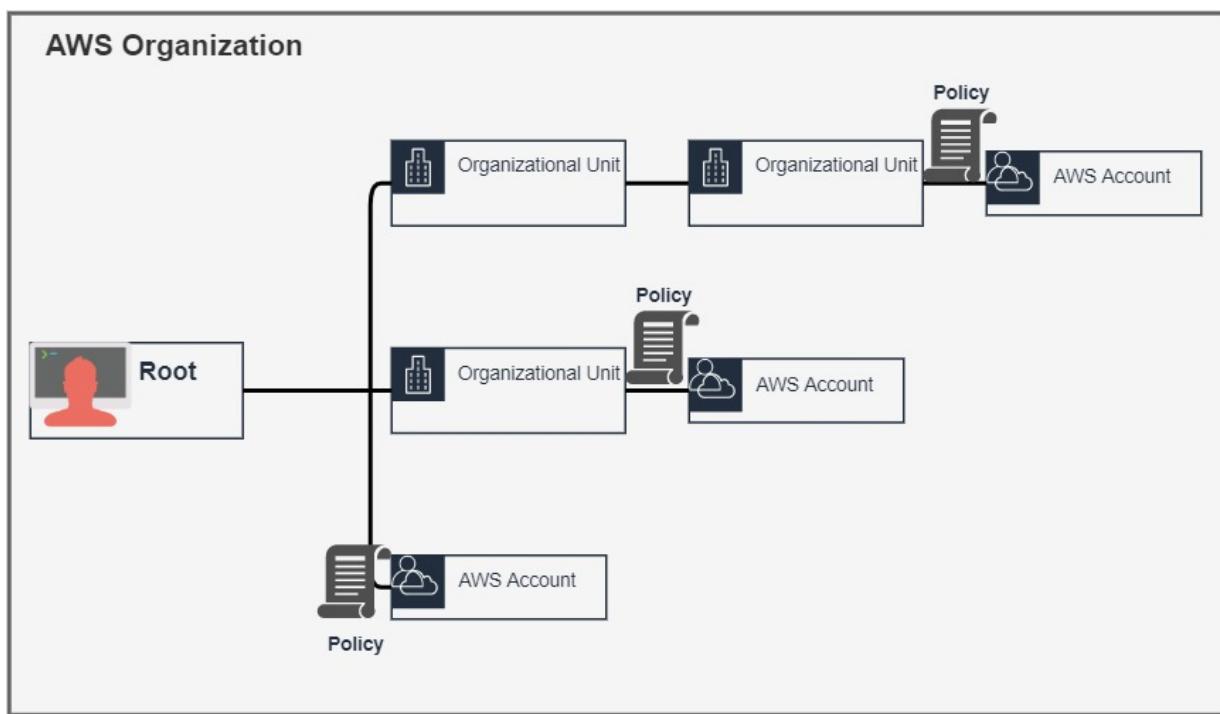
References:

- <https://docs.aws.amazon.com/kms/latest/developerguide/concepts.html>
- <https://docs.aws.amazon.com/secretsmanager/latest/userguide/intro.html>
- <https://docs.aws.amazon.com/systems-manager/latest/userguide/systems-manager-parameter-store.html>
- <https://docs.aws.amazon.com/IAM/latest/UserGuide/what-is-access-analyzer.html>
- <https://docs.aws.amazon.com/acm/latest/userguide/acm-overview.html>
- <https://tutorialsdojo.com/aws-secrets-manager-vs-systems-manager-parameter-store/>

AWS Billing and Governance

AWS Organizations

AWS Organizations' purpose is to ease the management of multiple AWS Accounts by consolidating them into a group called Organization.



Customers mainly benefit from AWS Organizations in terms of billing and account management. Instead of paying bills from each AWS Account, customers can consolidate and pay all bills on their AWS Master Account. Consolidated billing allows the customers to centrally manage their cloud expenditure while saving costs through different pricing discounts.

You can centrally manage all of your AWS Account from your Master Account. When you enable AWS Organizations on your account, it automatically becomes the Master Account; an email verification is also sent to your AWS Account email.



AWS Organizations > AWS accounts

AWS accounts

Add an AWS account

The accounts listed below are members of your organization. The organization's management account is responsible for paying the bills for all accounts in the organization. You can use the tools provided by AWS Organizations to centrally manage these accounts. [Learn more](#)

Organization

Organizational units (OUs) enable you to group several accounts together and administer them as a single unit instead of one at a time.

Find AWS accounts by name, email, or account ID. Find an OU by the exact OU ID.

Hierarchy List

Organizational structure	Account created/joined date
▼ <input type="checkbox"/> <input type="checkbox"/> Root r-v0br	
<input type="checkbox"/> Lervin John Obando <small>management account</small> 9471 [REDACTED] [REDACTED]@outlook.com	Joined 2021/08/09

You can then create an AWS account or invite an existing AWS Account for your Organization; the added AWS accounts are considered member accounts. You can invite an existing AWS Account by specifying the AWS Account ID or the email address.

Add an AWS account

You can add an AWS account to your organization either by creating an account or by inviting an existing AWS account to join your organization.

Create an AWS account

Create an AWS account that is added to your organization.

Invite an existing AWS account

Send an email request to the owner of the account. If they accept, the account joins the organization.

Invite an existing AWS account to join your organization

Email address or account ID of the AWS account to invite

account@example.org or 111122223333

You can also have a hierarchical grouping on your AWS Organizations. You can do this by creating Organizational Units (OU), a logical group, or container for your accounts. OUs may either have an AWS Account or another OU into it. You can also freely move an AWS Account at any level in your Organization.



Organizational structure	Account created/joined date
▼ □ Root r-v0br	
▼ □ Admin Unit ou-v0br-o0pv1ubu	
□ Lervin John Obando management account 9471 [REDACTED] [REDACTED]@outlook.com	Joined 2021/08/09
▼ □ Dev Unit ou-v0br-i1bmmpg8	
□ Iervz 7681 [REDACTED] [REDACTED]@gmail.com	Created 2021/08/09
► □ QA Unit ou-v0br-jmbwbkhb	

You can attach policies at any level on your Organizational hierarchy as another management capability. Organization policies are disabled by default, but you can explicitly enable them. Below are the supported policy types for the AWS Organization.

Policies

Policies in AWS Organizations enable you to manage different features of the AWS accounts in your organization. [Learn more](#)

Supported policy types	
Policy type	Status
AI services opt-out policies Artificial Intelligence (AI) services opt-out policies enable you to control whether AWS AI services can store and use your content. Learn more	Disabled
Backup policies Backup policies enable you to deploy organization-wide backup plans to help ensure compliance across your organization's accounts. Using policies helps ensure consistency in how you implement your backup plans. Learn more	Enabled
Service control policies Service control policies (SCPs) enable central administration over the permissions available within the accounts in your organization. This helps ensure that your accounts stay within your organization's access control guidelines. Learn more	Enabled
Tag policies Tag policies help you standardize tags on all tagged resources across your organization. You can use tag policies to define tag keys (including how they should be capitalized) and their allowed values. Learn more	Enabled

When you attach a policy to one of the hierarchy nodes, all Organizational entities (root, Organizational unit (OU), or account) beneath it will inherit that policy. Here's an example of a Backup Policy being attached to an Organizational Unit (OU).



Attach backup_policy to a target

Select a target that backup_policy policy should be applied to. If you select an organizational unit, the applied policy will affect all AWS accounts that belong to that organizational unit.

AWS Organization

Organizational units (OUs) enable you to group several accounts together and administer them as a single unit instead of one at a time.

Find AWS accounts by name, email, or account ID. Find an OU by the exact OU ID.

Hierarchy List

Organizational structure	Account created/joined date
▼ ○ <input type="checkbox"/> Root r-v0br	
▶ ○ <input type="checkbox"/> Admin Unit ou-v0br-0opv1ubu	
▶ <input checked="" type="radio"/> Dev Unit ou-v0br-i1bmmpg8	
▶ ○ <input type="checkbox"/> QA Unit ou-v0br-jmbwbkhb	

Cancel Attach policy

Service Control Policies (SCP)

Service control policies (SCPs) are a type of organization policy that you can use to manage permissions in your organization. SCPs are similar to IAM permissions policies except that they don't grant any permissions. Instead, SCPs specify the maximum permissions for an AWS Organizations entity. Attaching an SCP to your organization root or an OU limits the permissions of all the entities beneath it; this includes the root user of any member account.

Since SCPs don't grant any permissions, you still need to attach identity-based or resource-based policies to IAM users or roles or the resources in your organization's accounts to grant permissions.

When creating a service control policy, you make a statement that identifies the maximum permission of an entity to a particular AWS Service. Below is an example of an SCP that limits an account's action to simple EC2 instance management.



The screenshot shows the AWS IAM Policy Editor interface. On the left, a JSON policy document is displayed:

```
1 {  
2     "Version": "2012-10-17",  
3     "Statement": [  
4         {  
5             "Sid": "Statement1",  
6             "Effect": "Deny",  
7             "Action": [  
8                 "ec2:StopInstances",  
9                 "ec2:StartInstances",  
10                "ec2:RebootInstances",  
11                "ec2:RunInstances",  
12                "ec2:DescribeInstanceAttribute",  
13                "ec2:DescribeInstances",  
14                "ec2:DescribeInstanceStatus",  
15                "ec2:DescribeInstanceTypes"  
16            ],  
17            "Resource": [  
18                "*"  
19            ]  
20        }  
21    ]  
22 }
```

On the right, the "Edit statement" panel is open, titled "Statement1". It shows the selected actions for the deny statement:

1. Add actions

All services > EC2

Q instance X

Action
<input type="checkbox"/> RegisterInstanceEventNotificationAttributes
<input type="checkbox"/> ReplaceElbInstanceProfileAssociation
<input type="checkbox"/> ReportInstanceStatus
<input type="checkbox"/> RequestSpotInstances
<input type="checkbox"/> ResetInstanceAttribute
<input checked="" type="checkbox"/> RunInstances
<input type="checkbox"/> RunScheduledInstances
<input checked="" type="checkbox"/> StartInstances
<input checked="" type="checkbox"/> StopInstances
<input type="checkbox"/> TerminateInstances

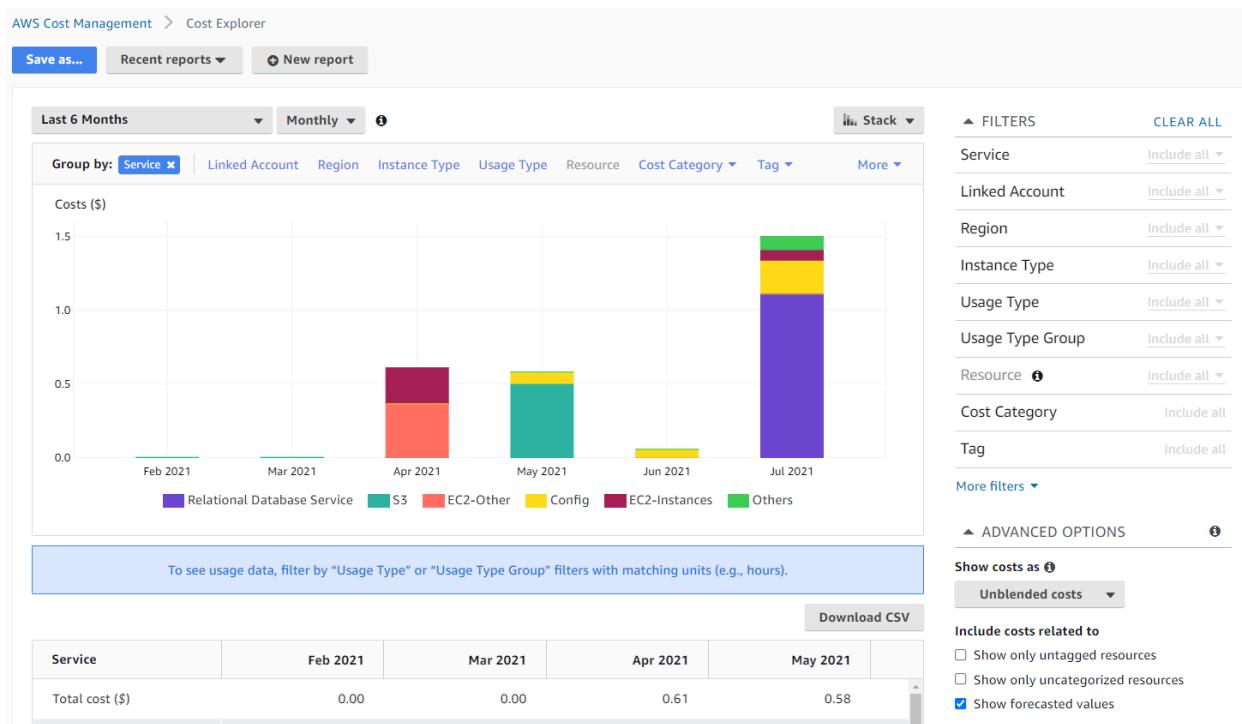
Cost Explorer

The Cost Explorer allows customers to view and analyze their consumption on AWS. This service can help you identify usage patterns like most used services and unusual high service usage, which are crucial for analyzing and maintaining a reasonable cloud expenditure. You can create reports from the Cost Explorer. Below are the types of reports that are supported.

- Cost and usage (recommended)
- Savings Plans reports (Savings Plans utilization, Savings Plans coverage)
- Reservation reports (Reservation utilization, Reservation coverage)

The Cost Explorer presents the costs data into a dashboard graphical report, making it easy to visualize your AWS usage. The dashboard is pre-configured but is customizable if you want to modify the view; it also provides a data filter to help you navigate your data. Data can also be visualized according to groups like Service, Regions, and Allocation Tags.

The Cost Explorer can present up to 12 months' worth of data and give a billing forecast according to your usage and billing patterns for the past months. Furthermore, it can also give a Reserve Instance purchase recommendation. Reports are downloadable as a CSV file.



Cost Allocation Tags

The Cost Allocation Tags is another effective way to organize and track your costs in your AWS cloud environment. You can enable Cost Allocation Tags on your AWS Billing console. AWS associates your resources to cost allocation tags to organize your cost allocation report.

There are two types of Cost Allocation Tags that you can simultaneously use.

- **User-defined cost allocation tags** - tags you created for your resources
- **AWS-generated cost allocation tags** - AWS-created and managed cost allocations tags



Cost allocation tags Info

User-defined cost allocation tags		AWS-generated cost allocation tags
<small>Search for a tag key</small>		
<input type="checkbox"/> Tag key		Status
<input type="checkbox"/> aws:createdBy		Active
<input type="checkbox"/> aws:ec2launchtemplate:id		Active
<input type="checkbox"/> aws:ec2launchtemplate:version		Active
<input type="checkbox"/> aws:autoscaling:groupName		Active
<input type="checkbox"/> aws:backup:source-resource		Inactive
<input type="checkbox"/> aws:elasticfilesystem:default-backup		Inactive
<input type="checkbox"/> aws:ssmmessages:session-id		Inactive
<input type="checkbox"/> aws:ec:fleet-id		Inactive
<input type="checkbox"/> aws:ssmmessages:target-id		Inactive

AWS License Manager

When you have multiple software licenses, tracking and managing them can be challenging, primarily if numerous vendors have issued them. With AWS License Manager, you can have your licenses centralized into one single plane, giving you a better view and control. Its scope can be extended across all AWS Organization Accounts, enabling cross-account license inventory.

Here are the following notable features of the AWS License Manager.

Manage License from multiple sources

Self-managed licenses - create/manage your software license (BYOL). When creating a self-managed license, you define the software license type (vCPUs, Cores, Sockets, Instances). Optionally, you can create Automated discovery rules to automatically discover and track the license of the software you defined from your instances. Once created, you can associate it with an AMI.



The screenshot shows the AWS License Manager console. At the top, it displays the path: AWS License Manager > Self-managed licenses > WindowsServerSTD2019. Below this, the title "WindowsServerSTD2019" is shown, along with "Actions" and "Associate AMI" buttons. A "Summary" section provides basic information: Status (Active), License type (vCPU), and vCPUs in use (0 of 4). Below the summary are tabs for "Resources", "Associated AMIs", "Details", "Automated discovery rules" (which is selected), and "Tags". Under "Automated discovery rules", there is a note stating "Rules cannot be edited or removed. To edit or remove rules, delete the self-managed license and create a new one." An "Add automated discovery rules" button is available. The "Product information" section indicates that this self-managed license does not track license included instances. It lists two instances: "Windows Server Standard" (2016) and "Windows Server Standard" (2019), both associated with "Amazon EC2 and on-premises".

Granted License - These licenses are acquired from the AWS Marketplace or an independent software vendor (ISV).

The screenshot shows the AWS License Manager console under the "Granted licenses" section. The path is AWS License Manager > Granted licenses. It displays three granted licenses:

License ID	Product name	Issuer	Seller of record	Status	Grant status	License start date	License end date	Time updated
L-0ee9f455a53940b1863528f80f54184a	Elastic Cloud (Elasticsearch Service)	AWS Marketplace	Elastic	Available	Active	October 28, 2022	-	October 28, 2022 New
L-720e365128d840b58c2baaa1f223d11f	Snowflake Data Cloud	AWS Marketplace	Snowflake	Available	Active	October 28, 2022	-	October 28, 2022 New
L-936edcab428c4cebb2d94f9401069165	Trend Micro Cloud One	AWS Marketplace	Trend Micro	Available	Active	October 28, 2022	-	October 28, 2022

AWS License Manager Dashboard - consolidated view of all licenses across the account/s



The screenshot shows the AWS License Manager dashboard. At the top, there are three summary boxes: 'Granted licenses' (3), 'Self-managed licenses' (1), and 'Seller issued licenses' (0). Below this, a section titled 'Granted license entitlements (3)' lists three entries: 'Elastic Cloud (Elasticsearch Service)', 'Snowflake Data Cloud', and 'Trend Micro Cloud One', each associated with 'AWS::Marketplace::Usage' and 'Enabled'. A 'View all granted licenses' link is at the bottom.

Hosted Resource Group - tracks the licenses from a collection of Dedicated hosts using the defined Self-managed license associated with an AMI.

The screenshot shows a 'Get started' guide for host resource groups. It consists of three numbered steps: 1. 'Create a self-managed license' (button: 'Create self-managed license'), 2. 'Associate an AMI' (button: 'Associate AMI'), and 3. 'Create a host resource group' (button: 'Create host resource group'). A note above step 1 states: 'A host resource group is a collection of EC2 Dedicated Hosts. After you create a host resource group, License Manager manages the hosts for you and automates tasks such as tracking licenses.' A 'Learn more' link is also present.

License type conversions - allows a customer to switch License type from BYOL to AWS-provided license and vice versa.

References:

- https://docs.aws.amazon.com/organizations/latest/userguide/orgs_introduction.html
- https://docs.aws.amazon.com/organizations/latest/userguide/orgs_manage_policies_scps.html
- <https://docs.aws.amazon.com/awsaccountbilling/latest/aboutv2/ce-what-is.html>
- <https://docs.aws.amazon.com/awsaccountbilling/latest/aboutv2/cost-alloc-tags.html>
- <https://docs.aws.amazon.com/license-manager/latest/userguide/license-manager.html>



Monitoring and Logging on AWS

Amazon CloudWatch is a monitoring tool that you can use for various AWS resources. With CloudWatch, you can collect metrics, logs, and apply event-driven actions using alarms and events.

CloudWatch Metrics for EC2

Amazon EC2 integrates with CloudWatch to monitor and collect performance data from EC2 instances called metrics. It can also collect metrics of the application hosted in the EC2 instance and on-premises server using **CloudWatch Agent**. Amazon CloudWatch consumes the default metrics available for your EC2 instances, but you also have the capabilities to publish your own custom metric. Here's a comparison of the default metrics (AWS-provided) and a custom metric.

	Monitoring Interval	Supported Metrics
Default EC2 Metrics	5 minutes - basic monitoring 1-minute - detailed monitoring	CPU, Network, Disk, Status check
Custom Metrics	1 minute down to a second	Memory, Application Metrics

Metric resolution is simply the granularity of the data being collected. Below are the two types of metric resolution.

- Standard resolution - 1-minute granularity
- High resolution - 1-second granularity

While AWS services provide standard resolution metrics, you can configure your custom metric in high resolution.

By default, Amazon CloudWatch does basic monitoring of all of your instances. CloudWatch collects the metrics at a 5-minute interval. Optionally, you can enable Amazon CloudWatch Detailed Monitoring to collect metrics in 1-minute intervals for an additional cost.



Detailed monitoring [Info](#)

By default, your instance is enabled for basic monitoring. You can optionally enable detailed monitoring.

Instance ID
 i-0e23301eb0eb2f853 (Windows Instance)

Detailed monitoring
 Enable

ⓘ After you enable detailed monitoring, the Amazon EC2 console displays monitoring graphs with a 1-minute period for the instance. Additional charges apply [↗](#)

[Cancel](#) [Save](#)

Instance Status Checks

Amazon EC2 publishes status check metrics to CloudWatch. EC2 performs status checks every minute to evaluate and identify potential failures on the instance and its underlying hardware and software. AWS does the following types of status checks to your instances.

- **System Status Check** - checks the underlying infrastructure to where an instance is hosted
- **Instance Status Check** - checks if the operating system of instance accepts traffic from its network interface

Status checks [Info](#)

Status checks detect problems that may impair i-0e23301eb0eb2f853 (Windows Instance) from running your applications.

Actions ▾

System status checks	Instance status checks
<input checked="" type="checkbox"/> System reachability check passed	<input checked="" type="checkbox"/> Instance reachability check passed

Report the instance status if our checks do not reflect your experience with this instance or if they do not detect issues you are having.

[Report instance status](#)



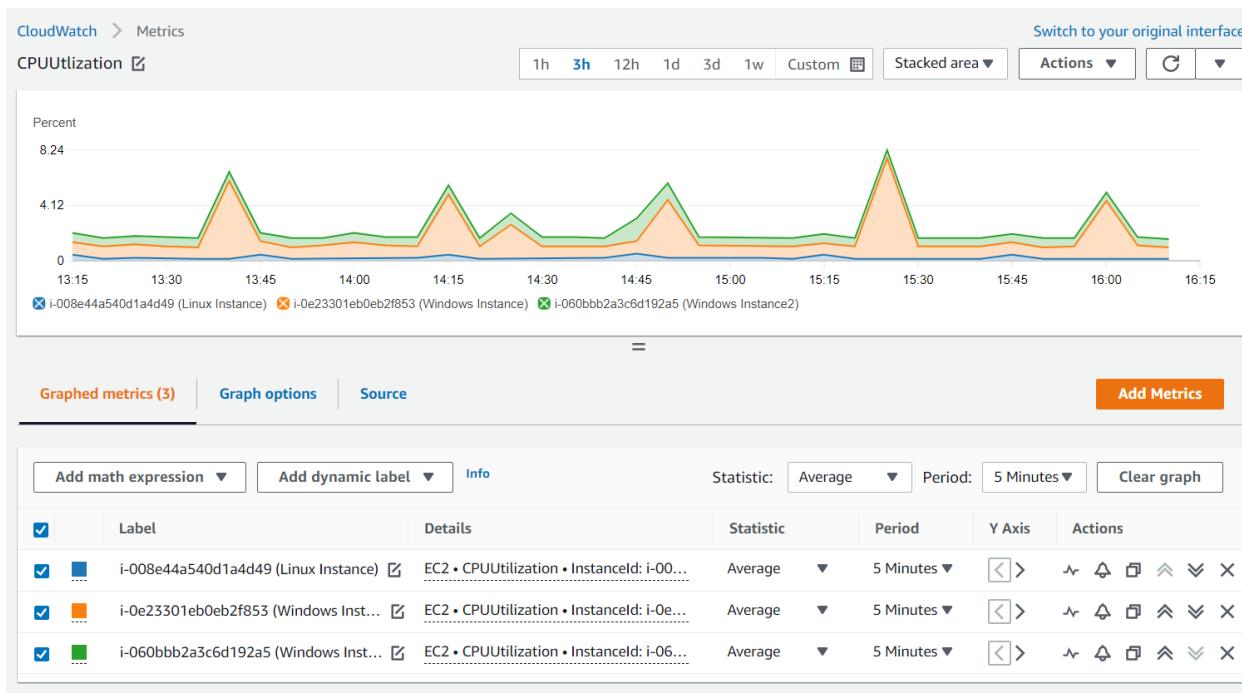
CloudWatch categorizes metrics using Namespaces. AWS provides the default Namespaces per AWS service, but you can also create custom Namespaces for your custom metrics.

All metrics Graphed metrics Graph options Source

▼ – AWS Namespaces

Backup 4 Metrics	EBS 36 Metrics
EC2 85 Metrics	Firehose 2 Metrics
Lambda 16 Metrics	Logs 20 Metrics
S3 6 Metrics	SNS 6 Metrics

Below is an example of how CloudWatch presents the metrics collected from your instances. The graph illustrates the CPUUtilizaton metrics of the instances for the past 3 hrs. You can configure a graph to change the time scope, widget type, or modify the current metrics.



Creating CloudWatch Alarm

One of the valuable tools that you can use for monitoring is the CloudWatch Alarm. You can use CloudWatch Alarm to create alarms from the metrics collected from your AWS resources. The alarm has three states on its lifecycle.

- **OK** - metric is within the threshold/alarm condition isn't met
- **ALARM** - metric is outside the threshold/alarm condition is met
- **INSUFFICIENT_DATA** - no or insufficient data collected/alarm is just created

Example of CloudWatch Alarm with In Alarm state:

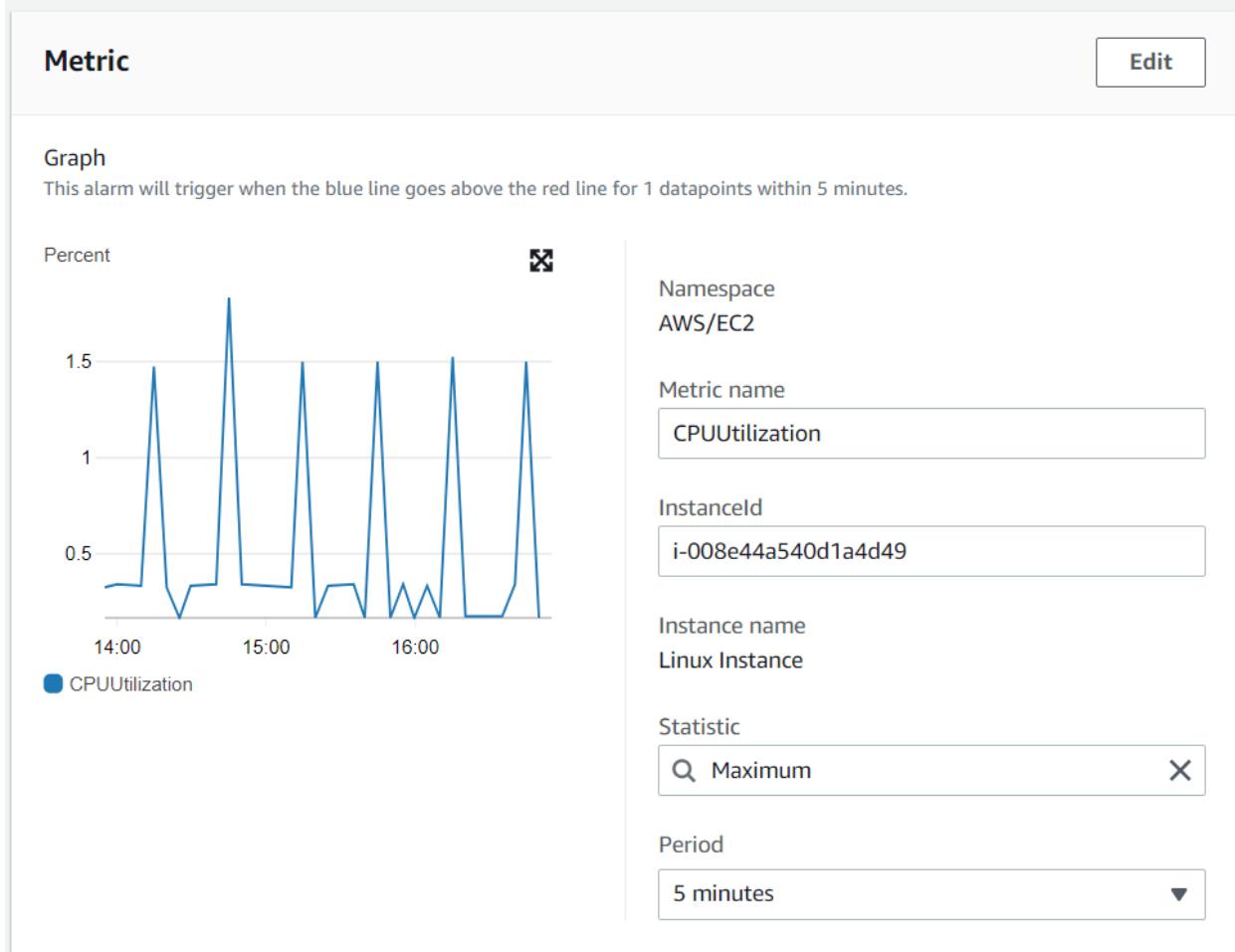
Alarms (1)		<input type="checkbox"/> Hide Auto Scaling alarms	<input type="button" value="Clear selection"/>	<input type="button" value="C"/>	<input type="button" value="Create composite alarm"/>	<input type="button" value="Actions"/>	<input type="button" value="Create alarm"/>
		<input type="text"/> Search			<input type="button" value="Any state"/>	<input type="button" value="Any type"/>	<input type="button" value="< 1 >"/>
<input type="checkbox"/>	Name	<input type="button" value="State"/>	Last state update	<input type="button" value="Conditions"/>			<input type="button" value="Actions"/>
<input type="checkbox"/>	CPUUtilAlarm	⚠ In alarm	2021-08-11 01:26:32	CPUUtilization > 80 for 1 datapoints within 5 minutes	✓	2 action(s) enabled	

Example of CloudWatch Alarm with OK state:



Alarms (1)		<input type="checkbox"/> Hide Auto Scaling alarms	<input type="button" value="Clear selection"/>	<input type="button" value="Create composite alarm"/>	<input type="button" value="Actions"/>	<input type="button" value="Create alarm"/>
		<input type="button" value="Search"/>	<input type="button" value="Any state"/>	<input type="button" value="Any type"/>	<input type="button" value="< 1 >"/>	<input type="button" value=""/>
Name	State	Last state update	Conditions	Actions		
CPUUtilAlarm	<input checked="" type="button"/> OK	2021-08-11 01:36:32	CPUUtilization > 80 for 1 datapoints within 5 minutes	<input checked="" type="button"/> 2 action(s) enabled		

When creating an alarm, you create an alarm trigger by specifying a metric and a condition. You first select an AWS service where you want to create an alarm. Looking at the example below, the CloudWatch Alarm will check the maximum CPUUtilization metric of the EC2 instance within 5 minutes.



You can select Static or Anomaly detection for the threshold type. While Static uses the value from the defined threshold, the Anomaly detection evaluates past data to determine a potential anomaly.



Conditions

Threshold type

Static
Use a value as a threshold

Anomaly detection
Use a band as a threshold

Whenever CPUUtilization is...

Define the alarm condition.

Greater
> threshold

Greater/Equal
>= threshold

Lower/Equal
<= threshold

Lower
< threshold

than...

Define the threshold value.

80

Must be a number

► Additional configuration

Once the metrics and the condition is set, you can define the action you want to do when the alarm is triggered. You can choose any action below.

Trigger an SNS Notification

Amazon CloudWatch will publish a message in the SNS Topic you selected, sending an email notification.



Notification

Alarm state trigger
Define the alarm state that will trigger this action.

In alarm
The metric or expression is outside of the defined threshold.

OK
The metric or expression is within the defined threshold.

Insufficient data
The alarm has just started or not enough data is available.

Select an SNS topic
Define the SNS (Simple Notification Service) topic that will receive the notification.

Select an existing SNS topic

Create new topic

Use topic ARN

Send a notification to...

EC2CWAlarmsTopic X

Only email lists for this account are available.

Email (endpoints)

[\[REDACTED\]@gmail.com](#) - [View in SNS Console](#)

Add notification

Trigger an Auto Scaling Action

You can trigger an auto scaling action only to your account's existing EC2 Auto Scaling group that uses a simple and step scaling policy. You can also configure trigger action for an ECS service.



Auto Scaling action

Alarm state trigger
Define the alarm state that will trigger this action.

In alarm
The metric or expression is outside of the defined threshold.

OK
The metric or expression is within the defined threshold.

Insufficient data
The alarm has just started or not enough data is available.

Remove

Resource type
Select a resource type.

EC2 Auto Scaling group

ECS Service

Select a group

Select a group ▾

Only Auto Scaling groups with a simple scaling or step scaling policy in this account are available.

Take the following action...

Select an action ▾

Only actions for the selected Auto Scaling group are available.

Add Auto Scaling action

Trigger an EC2 Action

You can configure Amazon CloudWatch to trigger an EC2 action for every alarm state that triggers. You can stop, terminate, or reboot the instance.



EC2 action

Alarm state trigger
Define the alarm state that will trigger this action.

In alarm
The metric or expression is outside of the defined threshold.

OK
The metric or expression is within the defined threshold.

Insufficient data
The alarm has just started or not enough data is available.

Remove

Take the following action...

Define what will happen to the EC2 instance with the Instance ID i-008e44a540d1a4d49 when this alarm is triggered.

Recover this instance
You can only recover certain EC2 instance types. [See documentation](#)

Stop this instance
You can only stop an instance if it is backed by an EBS volume. AWS will use the existing Service Linked Role (AWSServiceRoleForCloudWatchEvents) to perform this action. [Show IAM policy document](#)

Terminate this instance
You will not be able to terminate this instance if termination protection is enabled. AWS will use the existing Service Linked Role (AWSServiceRoleForCloudWatchEvents) to perform this action. [Show IAM policy document](#)

Reboot this instance
An instance reboot is equivalent to an operating system reboot. AWS will use the existing Service Linked Role (AWSServiceRoleForCloudWatchEvents) to perform this action. [Show IAM policy document](#)

Add EC2 action

Trigger a Systems Manager Action

Configure a Systems Manager action to create an OpsItem or Incident every time the alarm goes in In Alarm state.



Systems Manager action [Info](#)

Remove

This action will trigger only when the alarm is **In Alarm** state.

Create OpsItem
This will create an OpsItem within OpsCenter with the specified severity and category.

Create incident
This will start an incident using the response plan as a template.

Severity
Define the severity of OpsItem

3 - Medium

Category (optional)
Define the category of OpsItem

Select category

Working with CloudWatch Logs

CloudWatch Logs is a component of CloudWatch that allows you to collect various logs from your application and different AWS services. If you are looking for an AWS native solution to centrally consolidate all of your logs, CloudWatch Logs is the right way to go.

Log groups

CloudWatch collects data called log events from various resources. **Log group** serves as a container for the log events. You define the log group name, log retention, and an optional KMS key when you create a log group.



Create log group

Log group details

Log group name

Retention setting

KMS key ARN - optional

Log events are collected and are dumped to log groups via log streams. **Log streams** represent the sequence of data being collected from multiple sources. A log group can contain multiple log streams. Here is an example of a log group that has log streams to collect log events from the VPC Flow Logs.

Log streams	Metric filters	Subscription filters	Contributor Insights	Tags
Log streams (3)				
<input type="button" value="C"/>	<input type="button" value="Delete"/>	<input type="button" value="Create log stream"/>	<input type="button" value="Search all"/>	
<input type="checkbox"/> Log stream		Last event time		
<input type="checkbox"/> eni-02d4a726c1fc1dd1a-all		2021-08-11 22:11:03 (UTC+08:00)		
<input type="checkbox"/> eni-02d4a726c1fc1dd1a-reject		2021-08-11 22:07:59 (UTC+08:00)		
<input type="checkbox"/> eni-02d4a726c1fc1dd1a-accept		2021-08-11 21:48:58 (UTC+08:00)		

Utilizing Metric Filter

Metric Filter allows you to drill down into your logs. You can monitor specific values or events from your logs and push them into CloudWatch metrics as custom Metrics. By defining a filter pattern, you can filter the log events and choose only the needed data.



The example below filters the log events to display only events that have "172.31.15.23" and "ACCEPT" values. You can test the pattern on custom log data, or better yet, on the log events from your existing log stream.

Create filter pattern

You can use metric filters to monitor events in a log group as they are sent to CloudWatch Logs. You can monitor and count specific terms or extract values from log events and associate the results with a metric. [Learn more about pattern syntax.](#)

Filter pattern

Specify the terms or pattern to match in your log events to create metrics.

X

Test pattern

Select log data to test

eni-02d4a726c1fc1dd1a-all

Log event messages

Type log data to test with your Filter Pattern. Please use line breaks to separate log events.

```
2 947117271373 eni-02d4a726c1fc1dd1a 210.12.108.167 172.31.15.23 62796 6
2 947117271373 eni-02d4a726c1fc1dd1a 52.119.186.176 172.31.15.23 443 568
2 947117271373 eni-02d4a726c1fc1dd1a 172.31.15.23 52.119.186.176 56857 4
2 947117271373 eni-02d4a726c1fc1dd1a 52.119.186.176 172.31.15.23 443 568
2 947117271373 eni-02d4a726c1fc1dd1a 172.31.15.23 52.119.186.176 56854 4
2 947117271373 eni-02d4a726c1fc1dd1a 172.31.15.23 52.119.186.176 56854 4
2 947117271373 eni-02d4a726c1fc1dd1a 185.156.72.10 172.31.15.23 55308 53
```

Test pattern

Results

Please select log event messages above and click "Test pattern" to see results.



Once the filter pattern is defined, you now set details for your metrics. Once the log events are pushed into the CloudWatch Metrics, it will appear as a custom Namespace. You can now also create CloudWatch Alarm using the custom metrics.

Metric details

Metric namespace
Namespaces let you group similar metrics. [Learn more](#)

Create new

Namespaces can be up to 255 characters long; all characters are valid except for colon(:), asterisk(*), dollar(\$), and space().

Metric name
Metric name identifies this metric, and must be unique within the namespace. [Learn more](#)

Metric name can be up to 255 characters long; all characters are valid except for colon(:), asterisk(*), dollar(\$), and space().

Metric value
Metric value is the value published to the metric name when a Filter Pattern match occurs.

Valid metric values are: floating point number (1, 99.9, etc.), numeric field identifiers (\$1, \$2, etc.), or named field identifiers (e.g. \$requestSize for delimited filter pattern or \$.status for JSON-based filter pattern - dollar (\$) or dollar dot (\$.) followed by alphanumeric and/or underscore (_) characters).

Default value – optional
The default value is published to the metric when the pattern does not match. If you leave this blank, no value is published when there is no match. [Learn more](#)

Unit – optional



Event-driven Architecture with Amazon EventBridge

Amazon EventBridge is a serverless event bus service that enables you to build event-driven applications using events from different sources. Amazon EventBridge, as the name implies, serves as a bridge between your application and various event sources. The Amazon EventBridge service includes CloudWatch Events as its component.

To further understand how Amazon EventBridge works, let's break down its components.

Event Bus

Event Bus serves as a receiver of events from different sources. AWS has a default event bus used by AWS services to send events; this is the event bus used by CloudWatch Events. AWS extends the capability of EventBridge to accept events from other sources, including different SaaS providers, through a Custom event bus.

The screenshot shows the AWS EventBridge console interface. It displays two main sections: 'Default event bus' and 'Custom event bus (1)'.

Default event bus:

Name	Amazon Resource Name (ARN)	Schema discovery
default	arn:aws:events:ap-southeast-1:9471:...:event-bus/default	Not Initiated

Custom event bus (1):

Name	Amazon Resource Name (ARN)	Schema discovery
td-event-bus	arn:aws:events:ap-southeast-1:9471:...:event-bus/td-event-bus	Not Initiated

Rules

Rules are used for filtering events from the event bus or creating scheduled invokes for your targets. When you create a rule to match events, you specify a custom pattern or use AWS' predefined pattern. The predefined patterns include patterns for AWS services and SaaS providers like what you see from the example below. AWS supports tons of SaaS applications, including Atlassian, Datadog, New Relic, PagerDuty, and Zendesk, to name a few.



Event matching pattern
You can use pre-defined pattern provided by a service or create a custom pattern

Pre-defined pattern by service
 Custom pattern

Service provider
AWS services or custom/partner services

Service partners

Service name
The name of partner service selected as the event source

Select a service

- CloudAMQP
- Datadog
- Epsagon
- Freshworks
- Game Server Services Co., Ltd.
- Genesys

Event pattern

1

Pre-defined event pattern

Copy Edit

Target

Amazon EventBridge evaluates all the events from the event bus according to the defined rule. Once an event matches a rule, Amazon EventBridge will invoke the defined target. Amazon EventBridge supports multiple AWS Services targets like Amazon EC2 actions, AWS Lambda function, SNS topic, and others.



Select targets

Select target(s) to invoke when an event matches your event pattern or when schedule is triggered (limit of 5 targets per rule).

Target Remove

Select target(s) to invoke when an event matches your event pattern or when schedule is triggered (limit of 5 targets per rule).

Lambda function ▾

Function ▾

▶ Configure version/alias
▶ Configure input
▶ Retry policy and dead-letter queue

Add target

Exploring Events on CloudTrail

AWS CloudTrail records all the API calls made by a user, role, or service in AWS. You can use CloudTrail to review all the activity within your AWS environment, which helps in auditing, compliance requirements, or troubleshooting.

CloudTrail Event History

CloudTrail Event History keeps a record of events for the past 90 days. You use the provided filter to navigate through the data or create an Athena Table for complex queries. Below are the following attributes that you can use to filter CloudTrail events.

- AWS access key
- Event ID
- Event name
- Event sources
- Read-only
- Resource name
- Resource type
- User name



The screenshot shows the AWS CloudTrail Event history interface. At the top, there's a search bar with placeholder text "Select a read-only value" and a dropdown menu set to "Read-only". To the right are buttons for "Download events" and "Create Athena table". Below the search bar is a time range selector with options for 30m, 1h, 3h, 12h, and Custom, with "30m" selected. A pagination control shows page 1 of 2. The main table has columns for "Event name", "Event time", "User name", "Event source", and "Resource type". The table lists several events, all from August 12, 2021, at various times between 01:39:01 and 01:45:13 UTC+08:00. Most events are from the user "ssm.amazonaws.com" and are associated with EC2 instances.

Below is an example of a “ConsoleLogin” event record. It shows essential data like User Identity, Account ID, and Login status.

The screenshot shows the AWS CloudTrail Event record details page for a "ConsoleLogin" event. The event occurred on 2021-08-11T12:49:22Z in the "us-east-1" region. The user identity is a root account with ARN arn:aws:iam::947117271373:root and account ID 947117271373. The event source is signin.amazonaws.com, and the user agent is Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36. The response elements indicate a successful login.

```
{  
    "eventVersion": "1.08",  
    "userIdentity": {  
        "type": "Root",  
        "principalId": "947117271373",  
        "arn": "arn:aws:iam::947117271373:root",  
        "accountId": "947117271373"  
    },  
    "eventTime": "2021-08-11T12:49:22Z",  
    "eventSource": "signin.amazonaws.com",  
    "eventName": "ConsoleLogin",  
    "awsRegion": "us-east-1",  
    "sourceIPAddress": "112.198.123.183",  
    "userAgent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)  
Chrome/92.0.4515.131 Safari/537.36",  
    "requestParameters": null,  
    "responseElements": {  
        "ConsoleLogin": "Success"  
    }  
}
```

Storing CloudTrail Events to S3 Bucket

CloudTrail allows you to create **Trails** for you to export specific CloudTrail events to S3. You can set up KMS encryption, log file validation, and SNS notification when creating trails. Once the logs are stored in an S3 bucket, you can also run queries on the CloudTrail logs using Amazon Athena.



Enable for all accounts in my organization

To review accounts in your organization, open AWS Organizations. [See all accounts](#)

Storage location [Info](#)

Create new S3 bucket

Create a bucket to store logs for the trail.

Use existing S3 bucket

Choose an existing bucket to store logs for this trail.

Trail log bucket and folder

Enter a new S3 bucket name and folder (prefix) to store your logs. Bucket names must be globally unique.

aws-cloudtrail-logs-947117271373-bc6799b6

Logs will be stored in aws-cloudtrail-logs-947117271373-bc6799b6/AWSLogs/947117271373

Log file SSE-KMS encryption [Info](#)

Enabled

Moreover, you have an option to push the logs in an Amazon CloudWatch log group as well.

CloudWatch Logs - optional

Configure CloudWatch Logs to monitor your trail logs and notify you when specific activity occurs. Standard CloudWatch and CloudWatch Logs charges apply. [Learn more](#)

CloudWatch Logs [Info](#)

Enabled

Log group [Info](#)

New

Existing

Log group name

aws-cloudtrail-logs-947117271373-c8379a9f

1-512 characters. Only letters, numbers, dashes, underscores, forward slashes, and periods are allowed.

References:

- https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/monitoring_automated_manual.html
- https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/working_with_metrics.html
- <https://docs.aws.amazon.com/AmazonCloudWatch/latest/logs/WhatIsCloudWatchLogs.html>
- <https://docs.aws.amazon.com/awscloudtrail/latest/userguide/cloudtrail-user-guide.html>



COMPARISON OF AWS SERVICES

S3 vs EBS vs EFS

TD Tutorials Dojo	S3	EBS	EFS
Type of storage	Object storage. You can store virtually any kind of data in any format.	Persistent block level storage for EC2 instances.	POSIX-compliant file storage for EC2 instances
Features	Accessible to anyone or any service with the right permissions	Deliver performance for workloads that require the lowest-latency access to data from a single EC2 instance	Has a file system interface, file system access semantics (such as strong consistency and file locking), and concurrently-accessible storage for multiple EC2 instances
Max Storage Size	Virtually unlimited	16 TiB for one volume	Unlimited system size
Max File Size	Individual Amazon S3 objects can range in size to a maximum of 5 terabytes.	Equivalent to the maximum size of your volumes	47.9 TiB for a single file
Performance (Latency)	Low, for mixed request types, and integration with CloudFront	Lowest, consistent; SSD-backed storages include the highest performance Provisioned OPS SSD and General Purpose SSD that balance price and performance.	Low, consistent; use Max I/O mode for higher performance
Performance (Throughput)	Multiple GBs per second; supports multi-part upload	Up to 2 GB per second. HDD-backed volumes include Throughput Optimized HDD for frequently accessed, throughput intensive workloads and Cold HDD for less frequently accessed data.	10+ GB per second. Bursting Throughput mode scales with the size of the file system. Provisioned Throughput mode offers higher dedicated throughput than bursting throughput.
Durability	Stored redundantly across multiple AZs; has 99.99999999% durability	Stored redundantly in a single AZ	Stored redundantly across multiple AZs
Availability	S3 Standard - 99.99% availability S3 Standard-IA - 99.9% availability S3 One Zone-IA - 99.5% availability. S3 Intelligent Tiering - 99.9%	Has 99.999% availability	99.9% SLA. Runs in multi-AZ



TD Tutorials Dojo	S3	EBS	EFS
Scalability	Highly scalable	Manually increase/decrease your memory size. Attach and detach additional volumes to and from your EC2 instance to scale.	EFS file systems are elastic, and automatically grow and shrink as you add and remove files.
Data Accessing	One to millions of connections over the web; S3 provides a REST web services interface	Single EC2 instance in a single AZ Amazon EBS Multi-Attach enables you to attach a single Provisioned IOPS SSD (io1 or io2) volume to up to 16 Nitro-based instances that are in the same Availability Zone.	One to thousands of EC2 instances or on-premises servers, from multiple AZs, regions, VPCs, and accounts concurrently
Access Control	Uses bucket policies and IAM user policies. Has <i>Block Public Access</i> settings to help manage public access to resources.	IAM Policies, Roles, and Security Groups	Only resources that can access endpoints in your VPC, called a <i>mount target</i> , can access your file system; POSIX-compliant user and group-level permissions
Encryption Methods	Supports SSL endpoints using the HTTPS protocol, Client-Side and Server-Side Encryption (SSE-S3, SSE-C, SSE-KMS)	Encrypts both data-at-rest and data-in-transit through EBS encryption that uses AWS KMS CMKs.	Encrypt data at rest and in transit. Data at rest encryption uses AWS KMS. Data in-transit uses TLS.
Backup and Restoration	Use versioning or cross-region replication	All EBS volume types offer durable snapshot capabilities.	EFS to EFS replication through third party tools or AWS DataSync
Pricing	Billing prices are based on the location of your bucket. Lower costs equals lower prices. You get cheaper prices the more you use S3 storage.	You pay GB-month of provisioned storage, provisioned IOPS-month, GB-month of snapshot data stored in S3	You pay for the amount of file system storage used per month. When using the Provisioned Throughput mode you pay for the throughput you provision per month.
Use Cases	Web serving and content management, media and entertainment, backups, big data analytics, data lake	Boot volumes, transactional and NoSQL databases, data warehousing & ETL	Web serving and content management, enterprise applications, media and entertainment, home directories, database backups, developer tools, container storage, big data analytics
Service endpoint	Can be accessed within and outside a VPC (via S3 bucket URL)	Accessed within one's VPC	Accessed within one's VPC



Amazon S3 vs Glacier

- Amazon S3 is a durable, secure, simple, and fast storage service, while Amazon S3 Glacier is used for archiving solutions.
- Use S3 if you need low latency or frequent access to your data. Use S3 Glacier for low storage cost, and you do not require millisecond access to your data.
- You have three retrieval options when it comes to Glacier, each varying in the cost and speed it retrieves an object for you. You retrieve data in milliseconds from S3.
- Both S3 and Glacier are designed for durability of 99.99999999% of objects across multiple Availability Zones.
- S3 and Glacier are designed for availability of 99.99%.
- S3 can be used to host static web content, while Glacier cannot.
- In S3, users create buckets. In Glacier, users create archives and vaults.
- You can store a virtually unlimited amount of data in both S3 and Glacier.
- A single Glacier archive can contain 40TB of data.
- S3 supports Versioning.
- You can run analytics and querying on S3.
- You can configure a lifecycle policy for your S3 objects to automatically transfer them to Glacier. You can also upload objects directly to either S3 or Glacier.
- S3 Standard-IA and One Zone-IA have a minimum capacity charge per object of 128KB. Glacier's minimum is 40KB.
- Objects stored in S3 have a minimum storage duration of 30 days (except for S3 Standard). Objects that are archived to Glacier have a minimum 90 days of storage. Objects that are deleted, overwritten, or transitioned to a different storage class before the minimum duration will incur the normal usage charge plus a pro-rated request charge for the remainder of the minimum storage duration.
- Glacier has a per GB retrieval fee.
- You can transition objects from some S3 storage classes to another. Glacier objects can only be transitioned to the Glacier Deep Archive storage class.
- S3 (standard, intelligent-tiering, standard-IA, and one zone-IA) and Glacier are backed by an SLA.



S3 Standard vs S3 Standard-IA vs S3 One Zone-IA vs S3 Intelligent Tiering

	S3 Standard	S3 Standard-Infrequent Access (IA)	S3 One Zone-Infrequent Access (IA)	S3 Intelligent Tiering
Features	General-purpose storage of frequently accessed data	For long-lived, rapid but less frequently accessed data; data is stored redundantly in multiple AZs	For long-lived, rapid but less frequently accessed data; data is stored redundantly in only one AZ of your choice	For long-lived data that have unpredictable access patterns
Durability	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)
Availability	99.99%	99.9%	99.5%	99.9%
Availability SLA	99.9%	99%	99%	99%
Number of Availability Zones	At least 3	At least 3	Only 1	At least 3
Minimum capacity charge per object	N/A	128KB	128KB	N/A
Minimum storage duration charge	N/A	30 days	30 days	30 days
Inserting data	Directly PUT into S3 Standard	Directly PUT into S3 Standard-IA or set Lifecycle policies to transition objects from the S3 Standard to the S3 Standard-IA storage class.	Directly PUT into S3 One Zone-IA or set Lifecycle policies to transition objects from the S3 Standard to the S3 One Zone-IA storage class.	Directly PUT into S3 Intelligent-Tiering or set Lifecycle policies to transition objects from the S3 Standard to the S3 Intelligent-Tiering storage class.
Retrieval fee	N/A	per GB retrieved	per GB retrieved	N/A
First byte latency	milliseconds	milliseconds	milliseconds	milliseconds
Storage transition	S3 Standard to all other S3 storage types including Glacier	S3 Standard-IA to S3 One Zone-IA or S3 Glacier	S3 One Zone-IA to S3 Glacier	S3 Intelligent to S3 One Zone-IA or S3 Glacier
Use Cases	Cloud applications, dynamic websites, content distribution, mobile and gaming applications, and big data analytics.	Ideally suited for long-term file storage, older sync and share storage, and other aging data.	For infrequently-accessed storage, like backup copies, disaster recovery copies, or other easily recreatable data.	Data with unknown or changing access patterns, optimize storage costs automatically, and unpredictable workloads



Additional Notes:

- Data stored in the S3 One Zone-IA storage class will be lost in the event of AZ destruction.
- S3 Standard-IA costs less than S3 Standard in terms of storage price, while still providing the same high durability, throughput, and low latency of S3 Standard.
- S3 One Zone-IA has 20% less cost than Standard-IA.
- It is recommended to use multipart upload for objects larger than 100MB.



AWS DataSync vs Storage Gateway

	Data Sync	Storage Gateway
Description	AWS DataSync is an online data transfer service that simplifies, automates, and accelerates the process of copying large amounts of data to and from AWS storage services over the Internet or over AWS Direct Connect.	AWS Storage Gateway is a hybrid cloud storage service that gives you on-premises access to virtually unlimited cloud storage by linking it to S3. Storage Gateway provides 3 types of storage interfaces for your on-premises applications: file, volume, and tape.
How it Works	Uses an agent which is a virtual machine (VM) that is owned by the user and is used to read or write data from your storage systems. You can activate the agent from the Management Console. The agent will then read from a source location, and sync your data to Amazon S3, Amazon EFS, or Amazon FSx for Windows File Server.	Uses a Storage Gateway Appliance - a VM from Amazon - which is installed and hosted on your data center. After the setup, you can use the AWS console to provision your storage options: File Gateway, Cached Volumes, or Stored Volumes, in which data will be saved to Amazon S3. You can also purchase the hardware appliance to facilitate the transfer instead of installing the VM.
Protocol	DataSync connects to existing storage systems and data sources with standard storage protocols (NFS, SMB), or using the Amazon S3 API.	Storage Gateway provides a standard set of storage protocols such as iSCSI, SMB, and NFS.
Storage	AWS DataSync can copy data between Network File Systems (NFS), SMB file servers or self-managed object storages. It can also move data between your on-premises storage and AWS Snowcone, Amazon S3, Amazon EFS, or Amazon FSx.	File Gateway enables you to store and retrieve objects in Amazon S3 using file protocols such as NFS and SMB. Volume Gateway stores your data locally in the gateway and syncs them to Amazon S3. It also allows you to take point-in-time copies of your volumes with EBS snapshots which you can restore and mount to your appliances as iSCSI devices. Tape Gateway data is immediately stored in Amazon S3 and can be archived to Amazon S3 Glacier or Amazon S3 Deep Archive.
Pricing	You are charged standard request, storage, and data transfer rates to read from and write to AWS services, such as Amazon S3, Amazon EFS, Amazon FSx for Windows File Server, and AWS KMS.	You are charged based on the type and amount of storage you use, the requests you make, and the amount of data transferred out of AWS.
Combination	You can use a combination of DataSync and File Gateway to minimize your on-premises' operational costs while seamlessly connecting on-premises applications to your cloud storage. AWS DataSync enables you to automate and accelerate online data transfers to AWS storage services. File Gateway then provides your on-premises applications with low latency access to the migrated data.	



S3 Transfer Acceleration vs Direct Connect vs VPN vs Snowball vs Snowmobile

S3 Transfer Acceleration (TA)

- Amazon S3 Transfer Acceleration makes public Internet transfers to S3 faster, as it leverages Amazon CloudFront's globally distributed AWS Edge Locations.
- There is no guarantee that you will experience increased transfer speeds. If S3 Transfer Acceleration is not likely to be faster than a regular S3 transfer of the same object to the same destination AWS Region, AWS will not charge for the use of S3 TA for that transfer.
- This is not the best transfer service to use if transfer disruption is not tolerable.
- S3 TA provides the same security benefits as regular transfers to Amazon S3. This service also supports multi-part upload.
- **S3 TA vs AWS Snow***
 - The AWS Snow* Migration Services are ideal for moving large batches of data at once. In general, if it will take more than a week to transfer over the Internet, or there are recurring transfer jobs and there is more than 25Mbps of available bandwidth, S3 Transfer Acceleration is a good option.
 - Another option is to use AWS Snowball or Snowmobile to perform initial heavy lift moves and then transfer incremental ongoing changes with S3 Transfer Acceleration.
- **S3 TA vs Direct Connect**
 - AWS Direct Connect is a good choice for customers who have a private networking requirement or who have access to AWS Direct Connect exchanges. S3 Transfer Acceleration is best for submitting data from distributed client locations over the public Internet, or where variable network conditions make throughput poor.
- **S3 TA vs VPN**
 - You typically use (IPsec) VPN if you want your resources contained in a private network. VPN tools such as OpenVPN allow you to setup stricter access controls if you have a private S3 bucket. You can complement this further with the increased speeds from S3 TA.

AWS Direct Connect

- Using AWS Direct Connect, data that would have previously been transported over the Internet can now be delivered through a **private physical network connection** between AWS and your datacenter or corporate network. Customers' traffic will remain in AWS global network backbone, after it enters AWS global network backbone.
- Benefits of Direct Connect vs internet-based connections
 - reduced costs
 - increased bandwidth
 - a more consistent network experience
- Each AWS Direct Connect connection can be configured with one or more **virtual interfaces**. Virtual interfaces may be configured to access AWS services such as Amazon EC2 and Amazon S3 using public IP space, or resources in a VPC using private IP space.
- You can run IPv4 and IPv6 on the same virtual interface.
- Direct Connect does not support multicast.



- A Direct Connect connection is **not redundant**. Therefore, a second line needs to be established if redundancy is required. Enable *Bidirectional Forwarding Detection* (BFD) when configuring your connections to ensure fast detection and failover.
- AWS Direct Connect offers SLA.
- Direct Connect vs IPsec VPN
 - A VPC VPN Connection utilizes IPSec to establish **encrypted network connectivity** between your intranet and Amazon VPC **over the Internet**. VPN Connections can be configured in minutes and are a good solution if you have an immediate need, have low to modest bandwidth requirements, and can tolerate the inherent variability in Internet-based connectivity. AWS Direct Connect **does not involve the Internet**; instead, it uses **dedicated, private network connections** between your intranet and Amazon VPC.
- You can combine one or more Direct Connect dedicated network connections with the Amazon VPC VPN. This combination provides an IPsec-encrypted private connection that also includes the benefits of Direct Connect.

AWS VPN

- AWS VPN is comprised of two services:
 - AWS Site-to-Site VPN enables you to securely connect your on-premises network or branch office site to your Amazon VPC.
 - AWS Client VPN enables you to securely connect users to AWS or on-premises networks.
- Data transferred between your VPC and datacenter routes over an encrypted VPN connection to help maintain the confidentiality and integrity of data in transit.
- If data that passes through Direct Connect moves in a dedicated private network line, AWS VPN instead encrypts the data before passing it through the Internet.
- VPN connection throughput can depend on multiple factors, such as the capability of your customer gateway, the capacity of your connection, average packet size, the protocol being used, TCP vs. UDP, and the network latency between your customer gateway and the virtual private gateway.
- All the VPN sessions are **full-tunnel VPN**. (cannot split tunnel)
- AWS Site-to-Site VPN enable you to create **failover** and CloudHub solutions **with AWS Direct Connect**.
- AWS Client VPN is designed to connect devices to your applications. It allows you to choose from **OpenVPN-based client**.

Snowball

- Snowball is a **petabyte-scale data transport** solution that uses secure appliances to transfer large amounts of data into and out of AWS.
- Benefits of Snowball include:
 - lower network costs,
 - Shorter transfer times,
 - and security using 256-bit encryption keys you manage through AWS Key Management Service (KMS)..
- Similar to Direct Connect, AWS Snowball is **physical hardware**. It includes a 10GBaseT network connection. You can order a device with either **50TB** or an **80TB** storage capacity.
- Data transported via Snowball are stored in Amazon S3 once the device arrives at AWS centers.
- AWS Snowball is not only for shipping data into AWS, but also out of AWS.



- AWS Snowball can be used as a quick order for additional temporary petabyte storage.
- For security purposes, data transfers must be completed **within 90 days of a Snowball's preparation**.
- When the transfer is complete and the device is ready to be returned, the E Ink shipping label will automatically update to indicate the correct AWS facility to ship to, and you can track the job status by using Amazon Simple Notification Service (SNS), text messages, or directly in the console.
- Snowball is the best choice if you need to more securely and quickly transfer terabytes to many petabytes of data to AWS. Snowball can also be the right choice if you don't want to make expensive upgrades to your network infrastructure, if you frequently experience large backlogs of data, if you're located in a physically isolated environment, or if you're in an area where high-bandwidth Internet connections are not available or cost-prohibitive.
- If you will be transferring data to AWS on an ongoing basis, it is better to use AWS Direct Connect.
- If multiple users located in different locations are interacting with S3 continuously, it is better to use S3 TA.
- You **cannot** export data directly from S3 Glacier. It should be first restored to S3.

Snowmobile

- Snowmobile is Snowball with larger storage capacity. Snowmobile is literally a mobile truck.
- Snowmobile is an **Exabyte-scale data transfer** service.
- You can transfer up to **100PB** per Snowmobile.
- Snowmobile uses multiple layers of security to help protect your data including dedicated security personnel, GPS tracking, alarm monitoring, 24/7 video surveillance, and an optional escort security vehicle while in transit. All data is encrypted with 256-bit encryption keys you manage through the AWS Key Management Service (KMS).
- After the data transfer is complete, the Snowmobile will be returned to your designated AWS region where your data will be uploaded into the AWS storage services such as S3 or Glacier.
- Snowball vs Snowmobile
 - To migrate large datasets of 10PB or more in a single location, you should use Snowmobile. For datasets less than 10PB or distributed in multiple locations, you should use Snowball.
 - If you have a high speed backbone with hundreds of Gb/s of spare throughput, then you can use Snowmobile to migrate the large datasets all at once. If you have limited bandwidth on your backbone, you should consider using multiple Snowballs to migrate the data incrementally.
 - Snowmobile **does not** support data export. Use Snowball/Snowball Edge for this cause.
- When the data import has been processed and verified, AWS performs a software erasure based on NIST guidelines.



Amazon EBS: SSD vs HDD

On a given volume configuration, certain I/O characteristics drive the performance behavior for your EBS volumes. SSD-backed volumes, such as General Purpose SSD (gp2) and Provisioned IOPS SSD (io1, io2), deliver consistent performance whether an I/O operation is random or sequential. HDD-backed volumes like Throughput Optimized HDD (st1) and Cold HDD (sc1) deliver optimal performance only when I/O operations are large and sequential.

In the exam, always consider the difference between SSD and HDD as shown on the table below. This will allow you to easily eliminate specific EBS-types in the options which are not SSD or not HDD, depending on whether the question asks for a storage type which has *small, random* I/O operations or *large, sequential* I/O operations.

FEATURES	SSD Solid State Drive	HDD Hard Disk Drive
Best for workloads with:	<i>small, random</i> I/O operations	<i>large, sequential</i> I/O operations
Can be used as a bootable volume?	Yes	No
Suitable Use Cases	<ul style="list-style-type: none">- Best for transactional workloads- Critical business applications that require sustained IOPS performance- Large database workloads such as MongoDB, Oracle, Microsoft SQL Server and many others...	<ul style="list-style-type: none">- Best for large streaming workloads requiring consistent, fast throughput at a low price- Big data, Data warehouses, Log processing- Throughput-oriented storage for large volumes of data that is infrequently accessed
Cost	moderate / high 	low 
Dominant Performance Attribute	IOPS	Throughput (MiB/s)





Provisioned IOPS SSD (io1, io2) volumes are designed to meet the needs of I/O-intensive workloads, particularly database workloads, that are sensitive to storage performance and consistency. Unlike gp2, which uses a bucket and credit model to calculate performance, an io1 volume allows you to specify a consistent IOPS rate when you create the volume, and Amazon EBS delivers within 10 percent of the provisioned IOPS performance 99.9 percent of the time over a given year. Provisioned IOPS SSD io2 is an upgrade of Provisioned IOPS SSD io1. It offers higher 99.999% durability and higher IOPS per GiB ratio with 500 IOPS per GiB, all at the same cost as io1 volumes.

Volume Name	General Purpose SSD		Provisioned IOPS SSD	
Volume type	gp3	gp2	io2	io1
Description	General Purpose SSD volume that balances price performance for a wide variety of transactional workloads	General Purpose SSD volume that balances price performance for a wide variety of transactional workloads	High performance SSD volume designed for business-critical latency-sensitive applications	High performance SSD volume designed for latency-sensitive transactional workloads
Use Cases	virtual desktops, medium sized single instance databases such as MSFT SQL Server and Oracle DB, low-latency interactive apps, dev & test, boot volumes	Boot volumes, low-latency interactive apps, dev & test	Workloads that require sub-millisecond latency, and sustained IOPS performance or more than 64,000 IOPS or 1,000 MiB/s of throughput	Workloads that require sustained IOPS performance or more than 16,000 IOPS and I/O-intensive database workloads
Volume Size	1 GB – 16 TB	1 GB – 16 TB	4 GB – 16 TB	4 GB – 16 TB
Durability	99.8% - 99.9% durability	99.8% - 99.9% durability	99.999%	99.8% - 99.9%
Max IOPS / Volume	16,000	16,000	64,000	64,000
Max Throughput / Volume	1000 MB/s	250 MB/s	1,000 MB/s	1,000 MB/s
Max IOPS / Instance	260,000	260,000	160,000	260,000
Max IOPS / GB	N/A	N/A	500 IOPS/GB	50 IOPS/GB
Max Throughput / Instance	7,500 MB/s	7,500 MB/s	4,750 MB/s	7,500 MB/s
Latency	single digit millisecond	single digit millisecond	single digit millisecond	single digit millisecond



Multi-Attach	No	No	Yes	Yes
--------------	----	----	-----	-----

Volume Name	Throughput Optimized HDD	Cold HDD
Volume type	st1	sc1
Description	Low cost HDD volume designed for frequently accessed, throughput-intensive workloads	Throughput-oriented storage for data that is infrequently accessed Scenarios where the lowest storage cost is important
Use Cases	Big data, data warehouses, log processing	Colder data requiring fewer scans per day
Volume Size	125 GB – 16 TB	125 GB – 16 TB
Durability	99.8% - 99.9% durability	99.8% - 99.9% durability
Max IOPS / Volume	500	250
Max Throughput / Volume	500 MB/s	250 MB/s
Max IOPS / Instance	260,000	260,000
Max IOPS / GB	N/A	N/A
Max Throughput / Instance	7,500 MB/s	7,500 MB/s
Multi-Attach	No	No



RDS vs DynamoDB

TD Tutorials Dojo	RDS	DynamoDB
Type of database	Managed relational (SQL) database	Fully managed key-value and document (NoSQL) database
Features	Has several database instance types for different kinds of workloads and supports six database engines - Amazon Aurora, PostgreSQL, MySQL, MariaDB, Oracle Database, and SQL Server.	Delivers single-digit millisecond performance at any scale.
Storage Size	- 128 TB for Aurora engine. - 64 TB for MySQL, MariaDB, Oracle, and PostgreSQL engines. - 16 TB for SQL Server engine.	Supports tables of virtually any size.
Number of tables per unit	Depends on the database engine	256
Performance	General Purpose Storage is an SSD-backed storage option that delivers a consistent baseline of 3 IOPS per provisioned GB with the ability to burst up to 3,000 IOPS. Provisioned IOPS Storage is an SSD-backed storage option designed to deliver a consistent IOPS rate that you specify when creating a database instance, up to 40,000 IOPS per database instance. Amazon RDS provisions that IOPS rate for the lifetime of the database instance. Optimized for OLTP database workloads. Magnetic – Amazon RDS also supports magnetic storage for backward compatibility.	Single-digit millisecond read and write performance. Can handle more than 10 trillion requests per day with peaks greater than 20 million requests per second, over petabytes of storage. DynamoDB Accelerator (DAX) is an in-memory cache that can improve the read performance of your DynamoDB tables by up to 10 times—taking the time required for reads from milliseconds to microseconds, even at millions of requests per second. You specify the read and write throughput for each of your tables.
Availability and durability	Amazon RDS Multi-AZ deployments synchronously replicates your data to a standby instance in a different Availability Zone Amazon RDS will automatically replace the compute instance powering your deployment in the event of a hardware failure..	DynamoDB global tables replicate your data automatically across 3 Availability Zones of your choice of AWS Regions and automatically scale capacity to accommodate your workloads.
Backups	The automated backup feature enables point-in-time recovery for your database instance. Database snapshots are user-initiated backups of your instance stored in Amazon S3 that are kept until you explicitly delete them.	Point-in-time recovery (PITR) provides continuous backups of your DynamoDB table data, and you can restore that table to any point in time up to the second during the preceding 35 days. On-demand backup and restore allows you to create full backups of your DynamoDB tables' data for data archiving.



	RDS	DynamoDB
Scalability	<p>The Amazon Aurora engine will automatically grow the size of your database volume. The MySQL, MariaDB, SQL Server, Oracle, and PostgreSQL engines allow you to scale on-the-fly with zero downtime.</p> <p>RDS also supports storage auto scaling</p> <p>Read replicas are available in Amazon RDS for MySQL, MariaDB, and PostgreSQL as well as Amazon Aurora.</p>	<p>Support tables of virtually any size with horizontal scaling.</p> <p>For tables using on-demand capacity mode, DynamoDB instantly accommodates your workloads as they ramp up or down to any previously reached traffic level.</p> <p>For tables using provisioned capacity, DynamoDB delivers automatic scaling of throughput and storage based on your previously set capacity.</p>
Security	<p>Isolate your database in your own virtual network.</p> <p>Connect to your on-premises IT infrastructure using industry-standard encrypted IPsec VPNs.</p> <p>You can configure firewall settings and control network access to your database instances.</p> <p>Integrates with IAM.</p>	Integrates with IAM.
Encryption	<p>Encrypt your databases using keys you manage through AWS KMS. With encryption enabled, data stored at rest is encrypted, as are its automated backups, read replicas, and snapshots.</p> <p>Supports Transparent Data Encryption in SQL Server and Oracle.</p> <p>Supports the use of SSL to secure data in transit.</p>	DynamoDB encrypts data at rest by default using encryption keys stored in AWS KMS.
Maintenance	Amazon RDS will update databases with the latest patches. You can exert optional control over when and if your database instance is patched.	No maintenance since DynamoDB is serverless.
Pricing	<p>A monthly charge for each database instance that you launch.</p> <p>Option to reserve a DB instance for a one or three year term and receive discounts in pricing, compared to On-Demand instance pricing.</p>	<p>Charges for reading, writing, and storing data in your DynamoDB tables, along with any optional features you choose to enable.</p> <p>There are specific billing options for each of DynamoDB's capacity modes.</p>
Use Cases	Traditional applications, ERP, CRM, and e-commerce.	Internet-scale applications, real-time bidding, shopping carts, and customer preferences, content management, personalization, and mobile applications.



Additional notes:

- DynamoDB has built-in support for ACID transactions.
- DynamoDB uses filter expressions because it does not support complex queries.
- Multi-AZ deployments for the MySQL, MariaDB, Oracle, and PostgreSQL engines utilize synchronous physical replication. Multi-AZ deployments for the SQL Server engine use synchronous logical replication.



RDS vs Aurora

	Aurora	RDS
Type of database	Relational database	
Features	<ul style="list-style-type: none">MySQL and PostgreSQL compatible.5x faster than standard MySQL databases and 3x faster than standard PostgreSQL databases.Use Parallel Query to run transactional and analytical workloads in the same Aurora database, while maintaining high performance.You can distribute and load balance your unique workloads across different sets of Aurora DB instances using custom endpoints.Aurora Serverless allows for on-demand, autoscaling of your Aurora DB instance capacity.	<ul style="list-style-type: none">Has several database instance types for different kinds of workloads and supports five database engines - MySQL, PostgreSQL, MariaDB, Oracle, and SQL Server.Can use either General Purpose Storage and Provisioned IOPS storage to deliver a consistent IOPS performance
Maximum storage capacity	<ul style="list-style-type: none">128 TB	<ul style="list-style-type: none">64 TB for MySQL, MariaDB, Oracle, and PostgreSQL engines16 TB for SQL Server engine
DB instance classes	<ul style="list-style-type: none">Memory Optimized classes - for workloads that need to process large data sets in memory.Burstable classes - provides the instance the ability to burst to a higher level of CPU performance when required by the workload.	<ul style="list-style-type: none">Standard classes - for a wide range of workloads, you can use general purpose instance. It offers a balance of compute, memory, and networking resources.Memory Optimized classes - for workloads that need to process large data sets in memory.Burstable classes - provides the instance the ability to burst to a



		higher level of CPU performance when required by the workload.
Availability and durability	<ul style="list-style-type: none">Amazon Aurora uses RDS Multi-AZ technology to automate failover to one of up to 15 Amazon Aurora Replicas across three Availability ZonesAmazon Aurora Global Database uses storage-based replication to replicate a database across multiple AWS Regions, with typical latency of less than 1 second.Self-healing: data blocks and disks are continuously scanned for errors and replaced automatically.	<ul style="list-style-type: none">Amazon RDS Multi-AZ deployments synchronously replicates your data to a standby instance in a different Availability Zone.Amazon RDS will automatically replace the compute instance powering your deployment in the event of a hardware failure.
Backups	<ul style="list-style-type: none">Point-in-time recovery to restore your database to any second during your retention period, up to the last five minutes.Automatic backup retention period up to thirty-five days.Backtrack to the original database state without needing to restore data from a backup.	<ul style="list-style-type: none">The automated backup feature enables point-in-time recovery for your database instance.Database snapshots are user-initiated backups of your instance stored in Amazon S3 that are kept until you explicitly delete them.



Scalability	<ul style="list-style-type: none">Aurora automatically increases the size of your volumes as your database grows larger (increments of 10 GB).Aurora also supports replica auto-scaling, where it automatically adds and removes DB replicas in response to changes in performance metrics.Cross-region replicas provide fast local reads to your users, and each region can have an additional 15 Aurora replicas to further scale local reads.	<ul style="list-style-type: none">The MySQL, MariaDB, SQL Server, Oracle, and PostgreSQL engines scale your storage automatically as your database workload grows with zero downtime.Read replicas are available for Amazon RDS for MySQL, MariaDB, PostgreSQL, Oracle, and SQL Server. Amazon RDS creates a second DB instance using a snapshot of the source DB instance and uses the engines' native asynchronous replication to update the read replica whenever there is a change to the source.Can scale compute and memory resources (vertically) of up to a maximum of 32 vCPUs and 244 GiB of RAM.
Security	<ul style="list-style-type: none">Isolate the database in your own virtual network via VPC.Connect to your on-premises IT infrastructure using encrypted IPsec VPNs or Direct Connect and VPC Endpoints.Configure security group firewall and network access rules to your database instances.Integrates with IAM.	
Encryption	<ul style="list-style-type: none">Encrypt your databases using keys you manage through AWS KMS. With Amazon Aurora encryption, data stored at rest is encrypted, as are its automated backups, snapshots, and replicas in the same cluster.Supports the use of SSL (AES-256) to secure data in transit.	<ul style="list-style-type: none">Encrypt your databases using keys you manage through AWS KMS. With Amazon RDS encryption, data stored at rest is encrypted, as are its automated backups, read replicas, and snapshots.Supports Transparent Data Encryption in SQL Server and Oracle.Supports the use of SSL to secure data in transit



DB Authentication	<ul style="list-style-type: none">• Password authentication• Password and IAM database authentication	<ul style="list-style-type: none">• Password authentication• Password and IAM database authentication• Password and Kerberos authentication
Maintenance	<ul style="list-style-type: none">• Amazon Aurora automatically updates the database with the latest patches.• Amazon Aurora Serverless enables you to run your database in the cloud without managing/maintaining any database infrastructure.	<ul style="list-style-type: none">• Amazon RDS will update databases with the latest major and minor patches on scheduled maintenance windows. You can exert optional control over when and if your database instance is patched.
Monitoring	<ul style="list-style-type: none">• Use Enhanced Monitoring to collect metrics from the operating system instance.• Use Performance Insights to detect database performance problems and take corrective action.• Uses Amazon SNS to receive a notification on database events.	
Pricing	<ul style="list-style-type: none">• A monthly charge for each database instance that you launch if you use on-demand. This includes both the instance compute capacity and the amount of storage being used.• Option to reserve a DB instance for a one or three-year term (reserve instances) and receive discounts in pricing.	



Use Cases	<ul style="list-style-type: none">• Enterprise applications - a great option for any enterprise application that uses relational database since it handles provisioning, patching, backup, recovery, failure detection, and repair.• SaaS applications - without worrying about the underlying database that powers the application, you can concentrate on building high-quality applications.• Web and mobile gaming - since games need a database with high throughput, storage scalability, and must be highly available. Aurora suits the variable use pattern of these apps perfectly.	<ul style="list-style-type: none">• Web and mobile applications - since the application needs a database with high throughput, storage scalability, and must be highly available. RDS also fulfills the needs of such highly demanding apps.• E-commerce applications - a managed database service that offers PCI compliance. You can just focus on building high-quality customer experiences without thinking of the underlying database.• Mobile and online games - game developers don't need to worry about provisioning, scaling, and monitoring of database servers since RDS manages the database infrastructure.
------------------	--	--



Multi-AZ deployments vs. Multi-Region deployments vs. Read Replicas

Multi-AZ deployments	Multi-Region deployments	Read Replicas
Main purpose is high availability	Main purpose is disaster recovery and local performance	Main purpose is scalability
Non-Aurora: synchronous replication; Aurora: asynchronous replication	Asynchronous replication	Asynchronous replication
Non-Aurora: only the primary instance is active; Aurora: all instances are active	All regions are accessible and can be used for reads	All read replicas are accessible and can be used for readscaling
Non-Aurora: automated backups are taken from standby; Aurora: automated backups are taken from shared storage layer	Automated backups can be taken in each region	No backups configured by default
Always span at least two Availability Zones within a single region	Each region can have a Multi-AZ deployment	Can be within an Availability Zone, Cross-AZ, or Cross-Region
Non-Aurora: database engine version upgrades happen on primary; Aurora: all instances are updated together	Non-Aurora: database engine version upgrade is independent in each region; Aurora: all instances are updated together	Non-Aurora: database engine version upgrade is independent from source instance; Aurora: all instances are updated together
Automatic failover to standby (non-Aurora) or read replica (Aurora) when a problem is detected	Aurora allows promotion of a secondary region to be the master	Can be manually promoted to a standalone database instance (non-Aurora) or to be the primary instance (Aurora)



EC2 Container Services ECS vs Lambda

Amazon EC2 Container Service (ECS)

- Amazon ECS is a highly scalable, high performance container management service that supports Docker containers and allows you to easily run applications on a managed cluster of Amazon EC2 instances. ECS eliminates the need for you to install, operate, and scale your own cluster management infrastructure.
- With ECS, deploying containerized applications is easily accomplished. This service fits well in running batch jobs or in a microservice architecture. You have a central repository where you can upload your Docker Images from ECS container for safekeeping called Amazon ECR.
- Applications in ECS can be written in a stateful or stateless manner.
- The Amazon ECS CLI supports Docker Compose, which allows you to simplify your local development experience as well as easily set up and run your containers on Amazon ECS.
- Since your applications still run on EC2 instances, server management is your responsibility. This gives you more granular control over your system.
- It is up to you to manage scaling and load balancing of your EC2 instances as well, unlike in AWS Lambda where functions scale automatically.
- You are charged for the costs incurred by your EC2 instances in your clusters. Most of the time, Amazon ECS costs more than using AWS Lambda since your active EC2 instances will be charged by the hour.
- One version of Amazon ECS, known as AWS Fargate, will fully manage your infrastructure so you can just focus on deploying containers. AWS Fargate has a different pricing model from the standard EC2 cluster.
- ECS will automatically recover unhealthy containers to ensure that you have the desired number of containers supporting your application.



AWS Lambda

- AWS Lambda is a function-as-a-service offering that runs your code in response to events and automatically manages the compute resources for you, since Lambda is a serverless compute service. With Lambda, you do not have to worry about managing servers, and directly focus on your application code.
- Lambda automatically scales your function to meet demands. It is noteworthy, however, that Lambda has a maximum execution duration per request of 900 seconds or 15 minutes.
- To allow your Lambda function to access other services such as Cloudwatch Logs, you would need to create an execution role that has the necessary permissions to do so.
- You can easily integrate your function with different services such as API Gateway, DynamoDB, CloudFront, etc. using the Lambda console.
- You can test your function code locally in the Lambda console before launching it into production. Currently, Lambda supports only a number of programming languages such as Java, Go, PowerShell, Node.js, C#, Python, and Ruby. ECS is not limited by programming languages since it mainly caters to Docker.
- Lambda functions must be stateless since you do not have volumes for data storage.
- You are charged based on the number of requests for your functions and the duration, the time it takes for your code to execute. To minimize costs, you can throttle the number of concurrent executions running at a time, and the execution time limit of the function.
- With Lambda@Edge, AWS Lambda can run your code across AWS locations globally in response to Amazon CloudFront events, such as requests for content to or from origin servers and viewers. This makes it easier to deliver content to end users with lower latency.



Security Group vs NACL

Security Group	Network Access Control List
Acts as a firewall for associated Amazon EC2 instances	Acts as a firewall for associated subnets
Controls both inbound and outbound traffic at the instance level	Controls both inbound and outbound traffic at the subnet level
You can secure your VPC instances using only security groups	Network ACLs are an additional layer of defense.
Supports allow rules only	Supports allow rules and deny rules
Stateful (Return traffic is automatically allowed, regardless of any rules)	Stateless (Return traffic must be explicitly allowed by rules)
Evaluates all rules before deciding whether to allow traffic	Evaluates rules in number order when deciding whether to allow traffic, starting with the lowest numbered rule.
Applies only to the instance that is associated to it	Applies to all instances in the subnet it is associated with
Has separate rules for inbound and outbound traffic	Has separate rules for inbound and outbound traffic
A newly created security group denies all inbound traffic by default	A newly created nACL denies all inbound traffic by default
A newly created security group has an outbound rule that allows all outbound traffic by default	A newly created nACL denies all outbound traffic default
Instances associated with a security group can't talk to each other unless you add rules allowing it	Each subnet in your VPC must be associated with a network ACL. If none is associated, the default nACL is selected.
Security groups are associated with network interfaces	You can associate a network ACL with multiple subnets; however, a subnet can be associated with only one network ACL at a time.





Your VPC has a default security group with the following rules:

1. Allow inbound traffic from instances assigned to the same security group.
2. Allow all outbound IPv4 traffic and IPv6 traffic if you have allocated an IPv6 CIDR block.

Your VPC has a default network ACL with the following rules:

1. Allows all inbound and outbound IPv4 traffic and, if applicable, IPv6 traffic.
2. Each network ACL also includes a non modifiable and non removable rule whose rule number is an asterisk. This rule ensures that if a packet doesn't match any of the other numbered rules, it's denied.



Application Load Balancer vs Network Load Balancer vs Gateway Load Balancer

Feature	Application Load Balancer	Network Load Balancer	Gateway Load Balancer
Protocols	HTTP, HTTPS, gRPC	TCP, UDP, TLS	IP
Platforms	VPC	VPC	VPC
Health checks	HTTP, HTTPS, gRPC	TCP, HTTP, HTTPS	TCP, HTTP, HTTPS
Cloudwatch Metrics	✓	✓	✓
Logging	✓	✓	✓
Zonal Failover	✓	✓	✓
Connection Draining (deregistration delay)	✓	✓	✓
Load Balancing to multiple ports on the same instance	✓	✓	✓
IP addresses as targets	✓	✓ (TCP, TLS)	✓
Load balancer deletion protection	✓	✓	✓
Configuration idle connection timeout	✓		
Cross-zone load balancing	✓	✓	✓
Sticky sessions	✓	✓	✓
Static IP		✓	
Elastic IP address		✓	
Preserve Source IP address	✓	✓	✓
Resource-based IAM permissions	✓	✓	✓
Tag-based IAM permissions	✓	✓	✓
Slow start	✓		
Web sockets	✓	✓	✓
PrivateLink Support		✓ (TCP, TLS)	✓ (GWLBE)
Source IP address CIDR-based routing	✓		



Feature	Application Load Balancer	Network Load Balancer	Gateway Load Balancer
Layer 7			
Path-based routing	✓		
Host-based routing	✓		
Native HTTP/2	✓		
Redirects	✓		
Fixed response	✓		
Lambda functions as targets	✓		
HTTP header-based routing	✓		
HTTP method-based routing	✓		
Query string parameter-based routing	✓		
Security			
SSL offloading	✓	✓	
Server Name Indication (SNI)	✓	✓	
Back-end server encryption	✓	✓	
User authentication	✓		
Session Resumption	✓	✓	
Terminates flow/proxy behavior	✓	✓	✓

Common features between the load balancers:

- Has instance health check features
- Has built-in CloudWatch monitoring
- Logging features



- Support zonal failover
- Supports connection draining
- Support cross-zone load balancing (evenly distributes traffic across registered instances in enabled AZs)
- Resource-based IAM permission policies
- Tag-based IAM permissions
- Flow stickiness - all packets are sent to one target and return the traffic that comes from the same target.



EC2 Instance Health Check vs ELB Health Check vs Auto Scaling and Custom Health Check

EC2 instance health check

- Amazon EC2 performs automated checks on every running EC2 instance to identify hardware and software issues.
- Status checks are performed every minute and each returns a pass or a fail status.
 - If all checks pass, the overall status of the instance is OK.
 - If one or more checks fail, the overall status is impaired.
- Status checks are built into EC2, so they cannot be disabled or deleted.
- You can create or delete alarms that are triggered based on the result of the status checks.
- There are two types of status checks

System Status Checks

- These checks detect underlying problems with your instance that require AWS involvement to repair. When a system status check fails, you can choose to wait for AWS to fix the issue, or you can resolve it yourself.

Instance Status Checks

- Monitor the software and network configuration of your individual instance. Amazon EC2 checks the health of an instance by sending an address resolution protocol (ARP) request to the ENI. These checks detect problems that require your involvement to repair.

Elastic Load Balancer (ELB) health check

- To discover the availability of your registered EC2 instances, a load balancer periodically sends pings, attempts connections, or sends requests to test the EC2 instances.
- The status of the instances that are healthy at the time of the health check is InService. The status of any instances that are unhealthy at the time of the health check is OutOfService.
- When configuring a health check, you would need to provide the following:
 - a specific port
 - protocol to use
 - HTTP/HTTPS health check succeeds if the instance returns a 200 response code within the health check interval.
 - A TCP health check succeeds if the TCP connection succeeds.
 - An SSL health check succeeds if the SSL handshake succeeds.
 - ping path
- ELB health checks do not support WebSockets.
- The load balancer routes requests only to the healthy instances. When an instance becomes impaired, the load balancer resumes routing requests to the instance only when it has been restored to a healthy state.
- The load balancer checks the health of the registered instances using either
 - the default health check configuration provided by Elastic Load Balancing or
 - a health check configuration that you configure (auto scaling or custom health checks for example).
- Network Load Balancers use active and passive health checks to determine whether a target is available to handle requests.
 - With active health checks, the load balancer periodically sends a request to each registered target to check its status. After each health check is completed, the load balancer node closes the connection that was established.
 - With passive health checks, the load balancer observes how targets respond to connections, which enables it to detect an unhealthy target before it is reported as unhealthy by active health checks. You cannot disable, configure, or monitor passive health checks.

Auto Scaling and Custom health checks

- All instances in your Auto Scaling group start in the healthy state. Instances are assumed to be healthy unless EC2 Auto Scaling receives notification that they are unhealthy. This notification can come from one or more of the following sources:
 - Amazon EC2 (default)
 - Elastic Load Balancing
 - A custom health check.
- After Amazon EC2 Auto Scaling marks an instance as unhealthy, it is scheduled for replacement. If you do not want instances to be replaced, you can suspend the health check process for any individual Auto Scaling group.
- If an instance is in any state other than running or if the system status is impaired, Amazon EC2 Auto Scaling considers the instance to be unhealthy and launches a replacement instance.
- If you attached a load balancer or target group to your Auto Scaling group, Amazon EC2 Auto Scaling determines the health status of the instances by checking both the EC2 status checks and the Elastic Load Balancing health checks.
- Amazon EC2 Auto Scaling waits until the health check grace period ends before checking the health status of the instance. Ensure that the health check grace period covers the expected startup time for your application.
- Health check grace period does not start until lifecycle hook actions are completed and the instance enters the InService state.
- With custom health checks, you can send an instance's health information directly from your system to Amazon EC2 Auto Scaling.



ELB Health Checks vs Route 53 Health Checks For Target Health Monitoring

Health Check Service	AWS Elastic Load Balancing	Amazon Route 53
What is it for?	This health check periodically sends a request to a target instance, server or function to verify its status i.e. available to accept traffic requests.	This health check monitors the state of a record's target, which can be an EC2 instance, a server, or an AWS service that has an endpoint.
Target health check settings	You enter the port and common path of your targets that the load balancer will send the health check request to. Ping Protocol: HTTP Ping Port: 80 Ping Path: /index.html	You enter the domain name or the IP address, port and path that Route 53 will use to send the health check request to if the record is a non-alias record, Specify endpoint by: IP address (selected) Domain name Protocol: HTTP IP address: 192.0.2.44 or 2001:DB8::1 Host name: www.example.com Port*: 80 Path: /images or by setting Evaluate target health to Yes if the record is an alias record. Record type: Alias Route traffic to: Alias A - Routes traffic to an IPv4 address and so... Alias to Application and Classic Load Balancer Evaluate target health: Yes Choose Region: US East (N. Virginia)
Area span	Load balancers can monitor targets that span multiple availability zones but not multiple regions.	Route 53 monitors your targets regardless of their location, as long as they are reachable by Route 53.
Health check frequency	You specify a value between 5 seconds and 300 seconds.	Choose either every 10 seconds or every 30 seconds.
Response timeout	You can enter a value between 2 seconds and 60 seconds.	Cannot be configured.
Criteria to pass health check	You specify a threshold that a target should pass/fail a health check to determine its status. Advanced Details: Response Timeout: 5 seconds Interval: 30 seconds Unhealthy threshold: 2 Healthy threshold: 10	If more than 18% of health checkers report that an endpoint is healthy, Route 53 considers it healthy. If 18% of health checkers or fewer report that an endpoint is healthy, Route 53 considers it unhealthy. Route 53 health check servers are located in different locations worldwide.
Accessibility	Make sure targets are reachable by the load balancer. New targets can be easily added and removed from the load balancer.	Make sure endpoints are reachable and resolvable when users hit your URL. Due to DNS caching, it may take a while for new target endpoints to reflect to end users.
Primary purpose	High availability and fault tolerance for your services	DNS failover routing



CloudTrail vs CloudWatch

- CloudWatch is a monitoring service for AWS resources and applications. CloudTrail is a web service that records API activity in your AWS account. They are both useful monitoring tools in AWS.
- By default, CloudWatch offers free basic monitoring for your resources, such as EC2 instances, EBS volumes, and RDS DB instances. CloudTrail is also enabled by default when you create your AWS account.
- With CloudWatch, you can collect and track metrics, collect and monitor log files, and set alarms. CloudTrail, on the other hand, logs information on who made a request, the services used, the actions performed, parameters for the actions, and the response elements returned by the AWS service. CloudTrail Logs are then stored in an S3 bucket or a CloudWatch Logs log group that you specify.
- You can enable detailed monitoring from your AWS resources to send metric data to CloudWatch more frequently, with an additional cost.
- CloudTrail delivers one free copy of management event logs for each AWS region. Management events include management operations performed on resources in your AWS account, such as when a user logs in to your account. Logging data events are charged. Data events include resource operations performed on or within the resource itself, such as S3 object-level API activity or Lambda function execution activity.
- CloudTrail helps you ensure compliance and regulatory standards.
- CloudWatch Logs reports on application logs, while CloudTrail Logs provide you specific information on what occurred in your AWS account.
- CloudWatch Events is a near real time stream of system events describing changes to your AWS resources. CloudTrail focuses more on AWS API calls made in your AWS account.
- Typically, CloudTrail delivers an event within 15 minutes of the API call. CloudWatch delivers metric data in 5 minutes periods for basic monitoring and 1 minute periods for detailed monitoring. The CloudWatch Logs Agent will send log data every five seconds by default.



CloudWatch Agent vs SSM Agent vs Custom Daemon Scripts

CloudWatch Agent	SSM Agent (AWS Systems Manager)	Custom Daemon Scripts
<p>CloudWatch agent allows you to collect more system-level metrics from your EC2 and on-premises servers than just the standard CloudWatch metrics.</p> <p>It also enables you to retrieve custom metrics from your applications or services using the <i>StatsD</i> and <i>collectd</i> protocols. <i>StatsD</i> is supported on both Linux servers and servers running Windows Server. <i>collectd</i> is supported only on Linux servers.</p> <p>You can use CloudWatch agent to collect logs from your servers and send them to CloudWatch Logs.</p> <p>Metrics collected by the CloudWatch agent are billed as custom metrics.</p> <p>You can install CloudWatch Agent using three ways:</p> <ul style="list-style-type: none">• via Command Line• via SSM Agent• via AWS CloudFormation	<p>SSM Agent is Amazon software that runs on your EC2 instances and your hybrid instances that are configured for Systems Manager.</p> <p>SSM Agent processes requests from the Systems Manager service in the cloud and configures your machine as specified in the request. You can manage servers without having to log in to them using automation.</p> <p>SSM Agent sends status and execution information back to the Systems Manager service by using the <i>EC2 Messaging</i> service.</p> <p>SSM Agent runs on Amazon EC2 instances using root permissions (Linux) or SYSTEM permissions (Windows).</p> <p>CloudWatch agent replaces SSM agent in sending metric logs to CloudWatch Logs.</p>	<p>You use custom scripts (such as cron or bash scripts) if the two previously mentioned agents do not fit your needs.</p> <p>CloudWatch agent is useful for collecting system-level metrics and logs. You can create custom scripts that perform some modifications before the metrics are sent out.</p> <p>SSM Agent is also useful for automation purposes, though Systems Manager does not have a document for every case scenario. You may also have some compliance requirements that would require SSM Agent to be disabled (recall that SSM agent runs at root level permissions).</p>



Latency Routing vs Geoproximity Routing vs Geolocation Routing

Latency Routing

Definition

Lets Route 53 serve user requests from the AWS Region that provides the lowest latency. It does not, however, guarantee that users in the same geographic region will be served from the same location.

Latency-based routing is based on latency measurements performed over a period of time, and the measurements reflect changes in network connectivity and routing.

How it works

To use latency-based routing, you create **latency records** for your resources in multiple AWS Regions. When Route 53 receives a DNS query for your domain or subdomain, it determines which AWS Regions you've created latency records for, determines which region gives the user the lowest latency, and then selects a latency record for that region. Route 53 responds with the value from the selected record, such as the IP address for a web server.

Record sets can be created using any record type supported by Route 53, except NS or SOA records.

Use Case

Use when you have **resources in multiple AWS Regions** and you want to route traffic to the **region that provides the best latency**

Geoproximity Routing

Definition

Lets Amazon Route 53 route traffic to your resources based on the geographic location of your users and your resources.

You can also optionally choose to route more traffic or less to a given resource by specifying a value, known as a bias. A bias expands or shrinks the size of the geographic region from which traffic is routed to a resource.

How it works

To use geoproximity routing, you must use **Route 53 traffic flow**.

You create traffic flow policies for your resources and specify one of the following values for each policy:

- If you're using AWS resources, you can set the AWS Region where your resource is created
- If you're using non-AWS resources, you can enter the latitude and longitude of the resource

Use Case

Use when you want to route traffic **based on the location of** your resources and, optionally, shift traffic from resources in one location to resources in another.



Geolocation Routing

Definition

Resources serve traffic based on the geographic location of your users, meaning the location that DNS queries originate from.

How it works

Geolocation works by **mapping IP addresses to locations**. Some IP addresses aren't mapped to geographic locations, so Amazon Route 53 will receive some DNS queries from locations that it can't identify.

You can create a default record that handles both queries from IP addresses that aren't mapped to any location and queries that come from locations that you haven't created geolocation records for. If you don't create a **default record**, Route 53 returns a "no answer" response for queries from those locations.

No two records should specify the same geographic location.

Use Case

Use when you want to route traffic **based on the location of your users**.

- You can localize your content and present some or all of your website in the language of your users.
- You can restrict distribution of content to only the locations in which you have distribution rights.
- Useful for balancing load across endpoints in a predictable, easy-to-manage way, so that each user location is consistently routed to the same endpoint.





Service Control Policies vs IAM Policies

Service Control Policies (SCP)	IAM Policies
<ul style="list-style-type: none">• SCPs are mainly used along with AWS Organizations organizational units (OUs).• SCPs do not replace IAM Policies such that they do not provide actual permissions. To perform an action, you would still need to grant appropriate IAM Policy permissions.• Even if a Principal is allowed to perform a certain action (granted through IAM Policies), an attached SCP will override that capability if it enforces a Deny on that action.• SCP takes precedence over IAM Policies.• SCPs can be applied to the root of an organization or to individual accounts in an OU.• When you apply an SCP to an OU or an individual AWS account, you choose to either enable (whitelist), or disable (blacklist) the specified AWS service. Access to any service that isn't explicitly allowed by the SCPs associated with an account, its parent OUs, or the management account is denied to the AWS accounts or OUs associated with the SCP.• Any account has only those permissions permitted by every parent above it. If a permission is blocked at any level above the account, either implicitly (by not being included in an Allow policy statement) or explicitly (by being included in a Deny policy statement), a user or role in the affected account can't use that permission, even if there is an attached IAM policy granting Administrator permissions to the user.• SCPs affect only principals that are managed by accounts that are part of the organization.	<ul style="list-style-type: none">• IAM Policies operate at the Principal level. There are two types of IAM policies<ul style="list-style-type: none">- Identity-based policies - attached to an IAM user, group, or role.- Resource-based policies - attached to an AWS resource such as an S3 bucket.• IAM Policies can grant/deny a Principal permissions to perform certain actions to certain resources. This can be used together with SCP to ensure stricter controls in AWS Organizations. An IAM policy can be applied only to IAM users, groups, or roles, and it can never restrict the root identity of the AWS account.• IAM Policies cannot be attached to OUs.• An IAM Policy can allow or deny actions. An explicit allow overrides an implicit deny. An explicit deny overrides an explicit allow.



S3 Pre-Signed URLs vs CloudFront Signed URLs vs Origin Access Identity

S3 Pre-signed URLs	CloudFront Signed URLs	Origin Access Identity (OAI)
All S3 buckets and objects by default are private . Only the object owner has permission to access these objects. Pre-signed URLs use the owner's security credentials to grant others time-limited permission to download or upload objects.	You can control user access to your private content in two ways <ul style="list-style-type: none">• Restrict access to files in CloudFront edge caches• Restrict access to files in your Amazon S3 bucket (unless you've configured it as a website endpoint)	You can configure an S3 bucket as the origin of a CloudFront distribution. OAI prevents users from viewing your S3 files by simply using the direct URL for the file. Instead, they would need to access it through a CloudFront URL.
When creating a pre-signed URL, you (as the owner) need to provide the following: <ul style="list-style-type: none">• Your security credentials• An S3 bucket name• An object key• Specify the HTTP method (GET to download the object or PUT to upload an object)• Expiration date and time of the URL.	You can configure CloudFront to require that users access your files using either signed URLs or signed cookies . You then develop your application either to create and distribute signed URLs to authenticated users or to send Set-Cookie headers that set signed cookies on the viewers for authenticated users. When you create signed URLs or signed cookies to control access to your files, you can specify the following restrictions: <ul style="list-style-type: none">• An expiration date and time for the URL• (Optional) The date and time the URL becomes valid• (Optional) The IP address or range of addresses of the computers that can be used to access your content You can use signed URLs or signed cookies for any CloudFront distribution, regardless of whether the origin is an Amazon S3 bucket or an HTTP server.	To require that users access your content through CloudFront URLs, you perform the following tasks: <ul style="list-style-type: none">• Create a special CloudFront user called an origin access identity.• Give the origin access identity permission to read the files in your bucket.• Remove permission for anyone else to use Amazon S3 URLs to read the files (through bucket policies or ACLs). You cannot set OAI if your S3 bucket is configured as a website endpoint.



SNI Custom SSL vs Dedicated IP Custom SSL

Server Name Indication (SNI) Custom SSL	Dedicated IP Custom SSL
<ul style="list-style-type: none">◆ Relies on the SNI extension of the TLS protocol, which allows multiple domains to serve SSL traffic over the same IP address.◆ Offers the same level of security when using Dedicated IP Custom SSL.◆ If you configure CloudFront to serve HTTPS requests using SNI, CloudFront associates your alternate domain name with an IP address for each edge location. The IP address to your domain name is determined during the SSL/TLS handshake negotiation, and isn't dedicated to your distribution.◆ Some older browsers do not support SNI and will not be able to establish a connection with CloudFront to load the HTTPS version of your content.◆ You can use SNI Custom SSL with no upfront or monthly fees for certificate management.	<ul style="list-style-type: none">◆ Mainly useful for browsers that do not support SNI.◆ For this feature, the Amazon content delivery network allocates dedicated IP addresses to serve your SSL content at each Edge location.◆ You will need to upload a SSL certificate and associate it with your CloudFront distributions.◆ You can associate more than two custom SSL certificate with your AWS Account by submitting a CloudFront Limit Increase Form.◆ This method works for every HTTPS request, regardless of the browser or other viewer that the user is using.◆ Because of the added cost associated with dedicating IP addresses per SSL certificate, AWS charges a fixed monthly fee of \$600 for each custom SSL certificate you associate with your content delivery network distributions, pro-rated by the hour.◆ You can switch to using a custom SSL/TLS certificate with SNI instead and eliminate the charge that is associated with dedicated IP addresses.



Redis (cluster mode enabled vs disabled) vs Memcached

	Redis (cluster mode enabled)	Redis (cluster mode disabled)	Memcached
Data Types	string, sets, sorted sets, lists, hashes, bitmaps, hyperloglog, geospatial indexes	string, sets, sorted sets, lists, hashes, bitmaps, hyperloglog, geospatial indexes	string, objects (like databases)
Data Partitioning (distribute your data among multiple nodes)	Supported	Unsupported	Supported
Modifiable cluster	Only versions 3.2.10 and later	Yes	Yes
Online resharding	Only versions 3.2.10 and later	No	No
Encryption	3.2.6, 4.0.10 and later	3.2.6, 4.0.10 and later	Unsupported
Sub-millisecond latency	Yes	Yes	Yes
FedRAMP, PCI DSS and HIPAA compliant	3.2.6, 4.0.10 and later	3.2.6, 4.0.10 and later	No
Multi-threaded (make use of multiple processing cores)	No	No	Yes
Node type upgrading	No	Yes	No
Engine upgrading	Yes		
Cluster replication (create multiple copies of a primary cluster)	Supported	Supported	Unsupported
Multi-AZ for automatic failover	Required	Optional	Unsupported
Transactions (execute a group of commands as an isolated and atomic operation)	Supported	Supported	Unsupported
Pub/Sub capability	Yes	Yes	No
Backup and restore (keep your data on disk with a point in time snapshot)	Supported	Supported	Unsupported
Lua Scripting (execute transactional Lua scripts)	Supported	Supported	Unsupported
Use Case	<ul style="list-style-type: none">• You need to partition your data across two to 250 or 500 nodes if the Redis engine version is 5.0.6 or higher. (clustered mode only).• You need geospatial indexing (clustered mode or non-clustered mode).• You don't need to support multiple databases.• Plus features of non-clustered mode	<ul style="list-style-type: none">• You need complex data types, such as strings, hashes, lists, sets, sorted sets, and bitmaps.• You need to sort or rank in-memory datasets.• You need persistence of your key store.• You need to replicate your data from the primary to one or more read replicas for read intensive applications.• You need automatic failover if your primary node fails.• You need pub/sub capabilities.• You need backup and restore capabilities• You need to support multiple databases.	<ul style="list-style-type: none">• You need the simplest model possible.• You need to run large nodes with multiple cores or threads.• You need the ability to scale out and in, adding and removing nodes as demand on your system increases and decreases.• You need to cache objects, such as a database.• Needs Auto Discovery to simplify the way an application connects to a cluster.





FINAL REMARKS AND TIPS

Systems Management and Operations are never easy. There is a lot of care and perseverance needed to make sure your development team has a working environment to launch their applications. Beyond that, you also need to ensure that they continue working correctly as expected through continuous monitoring. And in mission-critical enterprise systems, many underlying components such as networks and access controls cannot afford failures. SysOps is often disregarded compared to other fields, but it is undeniable that SysOps administration is the backbone of IT.

Since you are at the end of our eBook, we hope that our AWS deep dives and cheat sheets are able to help you prepare for your AWS SysOps Administrator certification exam. We crafted this eBook as a way for you to review the important concepts that might appear in the actual exam, but it also serves to train you for the actual SysOps administration role. As more and more customers are moving to the cloud to use it for their products and services, management has been a very daunting task that only a few people like you can face head-on. As evidently shown in our exam overview section, there are many domains that you should study for and even more services to familiarize yourself with. We know that these challenges are not going to be easy, but we also know that you can succeed in your endeavors and provide yourself with more opportunities for career advancement. We are with you in every step of your AWS journey.

And with that, we at Tutorials Dojo thank you for supporting us through this eBook. If you wish to validate what you have learned so far, now is a great time to check out our [AWS Certified SysOps Administrator Associate Practice Exams](#). You can also try the free sampler version of our full practice test course [here](#). It will fill in the gaps in your knowledge that you are not aware of, and will give you a sense of the actual exam environment. That way, you'll know what to expect in the actual exam and you can pace yourself through the questions better. If you have any issues, concerns, or constructive feedback on our eBook, feel free to contact us at support@tutorialsdojo.com.

Goodluck on your exam, and we'd love to hear back from you soon.

Your Learning Partners,
Jon Bonso, Lervin John Obando, and the Tutorials Dojo Team



ABOUT THE AUTHORS



Jon Bonso (10x AWS Certified)

Born and raised in the Philippines, Jon is the Co-Founder of [Tutorials Dojo](#). Now based in Sydney, Australia, he has over a decade of diversified experience in Banking, Financial Services, and Telecommunications. He's 10x AWS Certified, an AWS Community Builder, and has worked with various cloud services such as Google Cloud, and Microsoft Azure. Jon is passionate about what he does and dedicates a lot of time creating educational courses. He has given IT seminars to different universities in the Philippines for free and has launched educational websites using his own money and without any external funding.



Lervin John Obando (2x AWS Certified)

Lervin is an experienced Cloud Systems Engineer with a strong background in building solutions in Amazon Web Services (AWS) using services like EC2, S3, VPC, IAM, etc. He has experience in multi-cloud infrastructures like Google Cloud Platform (GCP) and data-center operations across Linux and Windows platforms. As a Systems Engineer, stabilizing systems and implementing strategic IT solutions are his primary commitment and value-added contributions to his organization. His core functions include resolving technical and cloud-related issues, and running multiple command-line scripts, troubleshooting and networking.