Recommended security controls for implementing AWS CAF security capabilities

AWS Prescriptive Guidance

AWS Prescriptive Guidance: Recommended security controls for implementing AWS CAF security capabilities

Table of Contents

Recommended security controls for implementing AWS CAF security capabilities

Rishi Singla and Rovan Omar, Amazon Web Services (AWS)

November 2023 (document history)

Security is the top priority at AWS. To help relieve your operational burden, you share responsibility for cloud security and compliance with AWS. AWS is responsible for security *of* the cloud, which means protecting the infrastructure that runs the services offered in the AWS Cloud. You are responsible for security *in* the cloud, such as your data and applications. This guide provides security controls that can help you meet your security responsibilities in the AWS Cloud.

The AWS Cloud Adoption Framework (AWS CAF) provides best practices that are designed to improve your cloud readiness. AWS CAF categorizes those best practices into six perspectives: business, people, governance, platform, security, and operations. This guide focuses on the following capabilities in the security perspective:

- **Identity and access management** – Manage human and machine identities and their permissions at scale.
- **Threat detection** – Configure logging and monitoring to detect and investigate a potential security misconfiguration, threat, or unexpected behavior.
- **Protecting infrastructure** – Protect systems and services from unintended or unauthorized access and potential vulnerabilities.
- **Protecting data** – Categorize data based on levels of sensitivity. Maintain visibility and control over data and how it is accessed and used in your organization.
- **Incident response** – Establish mechanisms to respond to and mitigate the potential impact of security incidents.

Failure to implement preventative, detective, and responsive security controls for these AWS CAF security capabilities can pose a critical risk to your cloud environment, and it can disrupt your business. Implementing the security controls in this guide can help your organization protect its cloud environment.

Note

AWS provides services, tools, and frameworks that can help you operate securely in the AWS Cloud. This guide aligns with and supplements the AWS Well-Architected Framework, AWS Cloud Adoption Framework (AWS CAF), the AWS Security Reference Architecture (AWS SRA), and other security recommendations published by AWS. The controls in this guide aren't comprehensive of all cloud security considerations, and this guide isn't intended to replace these frameworks.

Security control recommendations for managing identity and access

You can create identities in AWS, or you can connect an external identity source. Through AWS Identity and Access Management (IAM) policies, you grant users the necessary permissions so that they can access or manage AWS resources and integrated applications. Effective identity and access management helps validate that the right people and machines have access to the right resources under the right conditions. The AWS Well-Architected Framework provides best practices for managing identities and their permissions. Examples of best practices include relying on a centralized identity provider and using strong sign-in mechanisms, such as multi-factor authentication (MFA). The security controls in this section can help you implement these best practices.

Controls in this section:

- Monitor and configure notifications for root user activity
- Don't create access keys for the root user
- Enable MFA for the root user
- Follow the security best practices for IAM
- Grant least-privilege permissions
- Define permission guardrails at the workload level
- Rotate IAM access keys at a regular interval
- Identify resources that are shared with an external entity

Monitor and configure notifications for root user activity

When you first create an AWS account, you begin with a single sign-in identity called the *root user*. By default, the root user has full access to all AWS services and resources in the account. You should tightly control and monitor the root user, and you should use it only for tasks that require root user credentials.

For more information, see the following resources:

- Grant least-privilege access in the AWS Well-Architected Framework
- Monitor IAM root user activity in AWS Prescriptive Guidance

Don't create access keys for the root user

The root user is the most privileged user in an AWS account. Disabling programmatic access to the root user helps reduce the risk of inadvertent exposure of the user credentials and subsequent compromise of the cloud environment. We recommend that you create and use IAM roles as temporary credentials for accessing your AWS accounts and resources.

For more information, see the following resources:

- IAM root user access key should not exist in the AWS Security Hub CSPM documentation
- Deleting access keys for the root user in the IAM documentation
- IAM roles in the IAM documentation

Enable MFA for the root user

We recommend that you enable multiple multi-factor authentication (MFA) devices for the AWS account root user and IAM users. This raises the security bar in AWS accounts and can simplify access management. Because a root user is a highly privileged user that can perform privileged actions, it's crucial to require MFA for the root user. You can use a hardware MFA device that generates a numeric code based on the time-based one-time password (TOTP) algorithm, a FIDO hardware security key, or a virtual authenticator application.

In 2024, MFA will be required to access the root user of any AWS account. For more information, see Secure by Design: AWS to enhance MFA requirements in 2024 in the AWS Security Blog. We strongly encourage you to extend this security practice and require MFA for all user types in your AWS environments.

If possible, we recommended that you use a hardware MFA device for the root user. Virtual MFA might not provide the same level of security as hardware MFA devices. You can use virtual MFA while waiting for hardware purchase approval or delivery.

In situations where you manage hundreds of accounts in AWS Organizations, depending on your organization's risk tolerance, it might not be scalable to use hardware-based MFA for the root user of each account in an organizational unit (OU). In this case, you can choose one account in the OU that acts as an OU management account, and then disable the root user for the other accounts in that OU. By default, the OU management account doesn't have access to the other accounts. By setting up cross-account access in advance, you can access the other accounts from the OU

management account in an emergency. To set up cross-account access, you create an IAM role in the member account, and you define policies so that only the root user in the OU management account can assume this role. For more information, see Tutorial: Delegate access across AWS accounts using IAM roles in the IAM documentation.

We recommend that you enable multiple MFA devices for your root user credentials. You can register up to eight MFA devices of any combination.

For more information, see the following resources:

- Enabling a hardware TOTP token in the IAM documentation
- Enabling a virtual multi-factor authentication (MFA) device in the IAM documentation
- Enabling a FIDO security key in the IAM documentation
- Secure your root user sign-in with multi-factor authentication (MFA) in the IAM documentation

Follow the security best practices for IAM

The IAM documentation includes a list of best practices that are designed to help you secure your AWS accounts and resources. It includes recommendations for configuring access and permissions according to the principle of least privilege. Examples of IAM security best practices include configuring identity federation, requiring MFA, and using temporary credentials.

For more information, see the following resources:

- Security best practices in IAM in the IAM documentation
- Using temporary credentials with AWS resources in the IAM documentation

Grant least-privilege permissions

Least privilege is the practice of grant only the permissions required to perform a task. You do this by defining the actions that can be taken on specific resources under specific conditions.

Attribute-based access control (ABAC) is an authorization strategy that defines permissions based on attributes, such as their tags. You can use group, identity, and resource attributes to dynamically define permissions at scale, rather than defining permissions for individual users. For example, you can use ABAC to allow a group of developers to access only resources that have a specific tag associated with their project.

For more information, see the following resources:

- Apply least-privilege permissions in the IAM documentation
- What is ABAC for AWS in the IAM documentation

Define permission guardrails at the workload level

It's best practice to use a multi-account strategy because it provides flexibility to define guardrails at the workload level. The AWS Security Reference Architecture offers prescriptive guidance about how to structure your accounts. These accounts are managed as an organization in AWS Organizations, and the accounts are grouped into *organizational units (OUs)*.

AWS services, such as AWS Control Tower, can help you centrally manage controls across an organization. We recommend that you define a clear purpose for each account or OU within the organization, and apply controls according to that purpose. AWS Control Tower implements preventive, detective, and proactive controls that help you govern the resources and monitor compliance. A *preventive control* is designed to prevent an event from occurring. A *detective control* is designed to detect, log, and alert after an event has occurred. A *proactive control* is designed to prevent the deployment of noncompliant resources by scanning resources before they are provisioned.

For more information, see the following resources:

- Separate workloads using accounts in the AWS Well-Architected Framework
- AWS Security Reference Architecture (AWS SRA) in AWS Prescriptive Guidance
- About controls in AWS Control Tower in the AWS Control Tower documentation
- Implementing security controls on AWS in AWS Prescriptive Guidance
- Use service control policies to set permission guardrails across accounts in your AWS Organization in the AWS Security Blog

Rotate IAM access keys at a regular interval

It's a best practice to update access keys for use cases that require long-term credentials. We recommend rotating access keys every 90 days or less. Rotating access keys reduces the risk that an access key that is associated with a compromised or terminated account is used. It also prevents

access by using an old key that might have been lost, compromised, or stolen. Always update applications after rotating the access keys.

For more information, see the following resources:

- Update access keys when needed for use cases that require long-term credentials in the IAM documentation
- Automatically rotate IAM user access keys at scale with AWS Organizations and AWS Secrets Manager in AWS Prescriptive Guidance
- Updating access keys in the IAM documentation

Identify resources that are shared with an external entity

An *external entity* is a resource, application, service, or user that is outside of your AWS organization, such as another AWS accounts, a root user, an IAM user or role, a federated user, an AWS service, or an anonymous (or unauthenticated) user. It is a security best practice to use IAM Access Analyzer to identify the resources in your organization and accounts, such as Amazon Simple Storage Service (Amazon S3) buckets or IAM roles, that are shared with an external entity. This helps you identify unintended access to resources and data, which is a security risk.

For more information, see the following resources:

- Verify public and cross-account access to resources with IAM Access Analyzer in the IAM documentation
- Analyze public and cross-account access in the AWS Well-Architected Framework
- Using AWS Identity and Access Management Access Analyzer in the IAM documentation

Security control recommendations for logging and monitoring

Logging and monitoring are important aspects of threat detection. Threat detection is one of the security perspective capabilities in the AWS Cloud Adoption Framework (AWS CAF). By using log data, your organization can monitor your environment to understand and identify potential security misconfigurations, threats, and unexpected behaviors. Understanding potential threats can help your organization prioritize security controls, and effective threat detection can help you respond to threats more quickly.

Controls in this section:

- Configure at least one multi-Region trail in CloudTrail
- Configure logging at the service and application level
- Establish a centralized location for analyzing logs and responding to security events
- Prevent unauthorized access to S3 buckets that contain CloudTrail log files
- Configure alerts for changes to security groups or network ACLs
- Configure alerts for CloudWatch alarms that enter the ALARM state

Configure at least one multi-Region trail in CloudTrail

AWS CloudTrail helps you audit the governance, compliance, and operational risk of your AWS account. Actions taken by a user, role, or an AWS service are recorded as events in CloudTrail. Events include actions taken in the AWS Management Console, AWS Command Line Interface (AWS CLI), and AWS SDKs and APIs. This event history helps you analyze your security posture, track resource changes, and audit compliance.

For an ongoing record of events in your AWS account, you must create a trail. Each trail should be configured to log events in all AWS Regions. By logging events in all AWS Regions, you ensure that all events that occur in your AWS account are logged, regardless of which AWS Region they occurred in. A multi-Region trail ensures that global service events are logged.

For more information, see the following resources:

- CloudTrail detective security best practices in the CloudTrail documentation

- Converting a trail that applies to one Region to apply to all Regions in the CloudTrail documentation
- Enabling and disabling global service event logging in the CloudTrail documentation

Configure logging at the service and application level

The AWS Well-Architected Framework recommends that you retain security event logs from services and applications. This is a fundamental principle of security for audit, investigations, and operational use cases. Service and application log retention is a common security requirement that is driven by governance, risk, and compliance (GRC) standards, policies, and procedures.

Security operations teams rely on logs and search tools to discover potential events of interest that might indicate unauthorized activity or unintentional change. You can enable logging for different services, depending on the use case. For example, you can log Amazon S3 bucket access, AWS WAF web ACL traffic, Amazon API Gateway traffic at the network layer, or Amazon CloudFront distributions.

For more information, see the following resources:

- Stream Amazon CloudWatch Logs to a centralized account for audit and analysis in the AWS Architecture Blog
- Configure service and application logging in the AWS Well-Architected Framework

Establish a centralized location for analyzing logs and responding to security events

Manually analyzing logs and processing information is insufficient to keep up with the volume of information associated with complex architectures. Analysis and reporting alone don't facilitate event assignment to the correct resource in a timely fashion. The AWS Well-Architected Framework recommends that you integrate AWS security events and findings into a notification and workflow system, such as a ticketing, bug, or security information and event management (SIEM) system. These systems help you assign, route, and manage security events.

For more information, see the following resources:

- Analyze logs, findings, and metrics centrally in the AWS Well-Architected Framework

- Analyze security, compliance, and operational activity using CloudTrail and Amazon Athena in the AWS Security Blog
- AWS Partners that provide threat detection and response services in the AWS Partners Portfolio

Prevent unauthorized access to S3 buckets that contain CloudTrail log files

By default, CloudTrail log files are stored in Amazon S3 buckets. It is a security best practice to prevent unauthorized access to any Amazon S3 bucket that contains CloudTrail log files. This helps you maintain the integrity, completeness, and availability of these logs, which is crucial for forensic and auditing purposes. If you want to log data events for S3 buckets that contain CloudTrail log files, you can create a CloudTrail trail for this purpose.

For more information, see the following resources:

- Configuring block public access settings for your S3 buckets in the Amazon S3 documentation
- CloudTrail preventative security best practices in the CloudTrail documentation
- Creating a trail in the CloudTrail documentation

Configure alerts for changes to security groups or network ACLs

A *security group* in Amazon Virtual Private Cloud (Amazon VPC) controls the traffic that is allowed to reach and leave the resources that it is associated with. A *network access control list (ACL)* allows or denies specific inbound or outbound traffic at the subnet level of the VPC. These resources are critical to managing access in your AWS environment.

Create and configure an Amazon CloudWatch alarm that notifies you if a security group or network ACL configuration changes. Configure this alarm to alert you every time an AWS API call is performed to update security groups. You can also use services, such as Amazon EventBridge and AWS Config, to automatically respond to these types of security events.

For more information, see the following resources:

- Automatically revert and receive notifications about changes to your Amazon VPC security groups in the AWS Security Blog

- Using Amazon CloudWatch alarms in the CloudWatch documentation
- Implement actionable security events in the AWS Well-Architected Framework
- Automate response to events in the AWS Well-Architected Framework

Configure alerts for CloudWatch alarms that enter the ALARM state

In CloudWatch, you can specify what actions an alarm takes when it changes state between the `OK`, `ALARM`, and `INSUFFICIENT_DATA` states. The most common type of alarm action is to notify one or more people by sending a message to an Amazon Simple Notification Service (Amazon SNS) topic. You can also configure alarms to create OpsItems or incidents in AWS Systems Manager.

We recommend that you activate alarm actions to automatically alert if a monitored metric is outside of the defined threshold. Monitoring alarms helps you identify unusual activities and quickly respond to security and operational issues.

For more information, see the following resources:

- Implement actionable security events in the AWS Well-Architected Framework
- Alarm actions in the CloudWatch documentation

Security control recommendations for protecting infrastructure

Infrastructure protection is a key part of any security program. It includes control methodologies that help you protect your networks and compute resources. Examples of infrastructure protection include trust boundaries, a defense-in-depth approach, security hardening, patch management, and operating system authentication and authorization. For more information, see Infrastructure protection in the AWS Well-Architected Framework. The security controls in this section can help you implement best practices for infrastructure protection.

Controls in this section:

- Specify default root objects for CloudFront distributions
- Scan application code to identify common security issues
- Create network layers by using dedicated VPCs and subnets
- Restrict incoming traffic to only authorized ports
- Block public access to Systems Manager documents
- Block public access to Lambda functions
- Restrict inbound and outbound traffic in the default security group
- Scan for software vulnerabilities and unintended network exposure
- Set up AWS WAF
- Configure advanced protections against DDoS attacks
- Use a defense-in-depth approach to control network traffic

Specify default root objects for CloudFront distributions

Amazon CloudFront speeds up distribution of your web content by delivering it through a worldwide network of data centers, which lowers latency and improves performance. If you don't define a default root object, requests for the root of your distribution pass to your origin server. If you are using an Amazon Simple Storage Service (Amazon S3) origin, the request might return a list of the contents in your S3 bucket or a list of the private contents of your origin. Specifying a default root object helps you avoid exposing the contents of your distribution.

For more information, see the following resources:

- Specifying a default root object in the CloudFront documentation

Scan application code to identify common security issues

The AWS Well-Architected Framework recommends that you scan libraries and dependencies for issues and defects. There are many source code analysis tools that you can use to scan source code. For example, Amazon CodeGuru can scan for common security issues in Java or Python applications and provide recommendations for remediation.

For more information, see the following resources:

- CodeGuru documentation
- Source code analysis tools on the OWASP Foundation website
- Perform vulnerability management in the AWS Well-Architected Framework

Create network layers by using dedicated VPCs and subnets

The AWS Well-Architected Framework recommends that you group components that share sensitivity requirements into layers. This minimizes the potential scope of impact of unauthorized access. For example, a database cluster that doesn't require internet access should be placed in a private subnet of its VPC to make sure that there is no route to or from the internet.

AWS offers many services that can help you test and identify public reachability. For example, Reachability Analyzer is a configuration analysis tool that helps you test connectivity between a source and destination resources in your VPCs. Also, Network Access Analyzer can help you identify unintended network access to resources.

For more information, see the following resources:

- Create network layers in the AWS Well-Architected Framework
- Reachability Analyzer documentation
- Network Access Analyzer documentation
- Create a subnet in the Amazon Virtual Private Cloud (Amazon VPC) documentation

Restrict incoming traffic to only authorized ports

Unrestricted access, such as traffic from the `0.0.0.0/0` source IP address, increases the risk for malicious activity, such as hacking, denial-of-service (DoS) attacks, and loss of data. Security groups provide stateful filtering of ingress and egress network traffic to AWS resources. No security group should allow unrestricted ingress access to well-known ports, such as SSH and Windows remote desktop protocol (RDP). For inbound traffic, in your security groups, allow only TCP or UDP connections on authorized ports. For connecting to Amazon Elastic Compute Cloud (Amazon EC2) instances, use Session Manager or Run Command instead of direct SSH or RDP access.

For more information, see the following resources:

- Work with security groups in the Amazon EC2 documentation
- Control traffic to your AWS resources using security groups in the Amazon VPC documentation

Block public access to Systems Manager documents

Unless your use case requires public sharing to be turned on, the AWS Systems Manager best practices recommend that you block public sharing for Systems Manager documents. Public sharing might provide unintended access to documents. A public Systems Manager document can expose valuable and sensitive information about your account, resources, and internal processes.

For more information, see the following resources:

- Best practices for shared Systems Manager documents in the Systems Manager documentation
- Modify permissions for a shared Systems Manager document in the Systems Manager documentation

Block public access to Lambda functions

AWS Lambda is a compute service that helps you run code without needing to provision or manage servers. Lambda functions should not be publicly accessible because this might allow unintended access to the function code.

We recommend that you configure resource-based policies for Lambda functions to deny access from outside of your account. You can achieve this by removing permissions or by adding the `AWS:SourceAccount` condition to the statement that allows access. You can update resource-

based policies for Lambda functions through the Lambda API or AWS Command Line Interface (AWS CLI).

We also recommend that you enable the **[Lambda.1] Lambda function policies should prohibit public access** control in AWS Security Hub CSPM. This control validates that resource-based policies for Lambda functions prohibit public access.

For more information, see the following resources:

- AWS Lambda controls in the Security Hub CSPM documentation
- Using resource-based policies for Lambda in the Lambda documentation
- Resources and conditions for Lambda actions in the Lambda documentation

Restrict inbound and outbound traffic in the default security group

If you don't associate a custom security group when you provision an AWS resource, then the resource is associated with the VPC's default security group. The default rules for this security group allow all inbound traffic from all resources that are assigned to this security group, and they allow all outbound IPv4 and IPv6 traffic. This might permit unintended traffic to the resource.

AWS recommends that you don't use the default security group. Instead, create custom security groups for specific resources or groups of resources.

Because the default security group can't be deleted, we recommend that you change the default security group rules to restrict inbound and outbound traffic. When configuring security group rules, follow the principle of least privilege.

We also recommend that you enable the **[EC2.2] VPC default security groups should not allow inbound or outbound traffic** control in Security Hub CSPM. This control validates that the default security group of a VPC denies inbound and outbound traffic.

For more information, see the following resources:

- Control traffic to your AWS resources using security groups in the Amazon VPC documentation
- Default security groups for your VPCs in the Amazon VPC documentation
- Amazon EC2 controls in the Security Hub CSPM documentation

Scan for software vulnerabilities and unintended network exposure

We recommend that you enable Amazon Inspector in all of your accounts. Amazon Inspector is a vulnerability management service that continually scans your Amazon EC2 instances, Amazon Elastic Container Registry (Amazon ECR) container images, and Lambda functions for software vulnerabilities and unintended network exposure. It also supports deep inspection of Amazon EC2 instances. When Amazon Inspector identifies a vulnerability or an open network path, it produces a finding that you can investigate. If Amazon Inspector and Security Hub CSPM are both set up in your account, then Amazon Inspector automatically sends security findings to Security Hub CSPM for centralized management.

For more information, see the following resources:

- Scanning resources with Amazon Inspector in the Amazon Inspector documentation
- Amazon Inspector Deep inspection for Amazon EC2 in the Amazon Inspector documentation
- Scan EC2 AMIs using Amazon Inspector in the AWS Security Blog
- Building a scalable vulnerability management program on AWS in AWS Prescriptive Guidance
- Automate network protection in the AWS Well-Architected Framework
- Automate compute protection in the AWS Well-Architected Framework

Set up AWS WAF

AWS WAF is a web application firewall that helps you monitor and block HTTP or HTTPS requests that are forwarded to your protected web application resources, such as Amazon API Gateway APIs, Amazon CloudFront distributions, or Application Load Balancers. Based on criteria that you specify, the service responds to requests either with the requested content, with an HTTP 403 status code (Forbidden), or with a custom response. AWS WAF can help protect web applications or APIs against common web exploits that can affect availability, compromise security, or consume excessive resources. Consider setting up AWS WAF in your AWS accounts and using a combination of AWS managed rules, custom rules, and partner integrations to help protect your applications from application layer (layer 7) attacks.

For more information, see the following resources:

- Getting started with AWS WAF in the AWS WAF documentation

- AWS WAF delivery partners on the AWS website
- Security automations for AWS WAF in the AWS Solutions Library
- Implement inspection and protection in the AWS Well-Architected Framework

Configure advanced protections against DDoS attacks

AWS Shield provides protections against distributed denial of service (DDoS) attacks for AWS resources at the network and transport layers (layer 3 and 4) and the application layer (layer 7). This service is available in two options: AWS Shield Standard and AWS Shield Advanced. Shield Standard automatically protects supported AWS resources, at no additional charge.

We recommend that you subscribe to Shield Advanced, which provides expanded DDoS attack protection for protected resources. The protections that you receive from Shield Advanced vary depending on your architecture and configuration choices. Consider implementing Shield Advanced protections for applications where you need any of the following:

- Guaranteed availability for the users of the application.
- Rapid access to DDoS mitigation experts if the application is affected by a DDoS attack.
- Awareness by AWS that the application might be affected by a DDoS attack and notification of attacks from AWS and escalation to your security or operations teams.
- Predictability in your cloud costs, including when a DDoS attack affects your use of AWS services.

For more information, see the following resources:

- AWS Shield Advanced overview in the Shield documentation
- AWS Shield Advanced protected resources in the Shield documentation
- AWS Shield Advanced capabilities and options in the Shield documentation
- Responding to DDoS events in the Shield documentation
- Implement inspection and protection in the AWS Well-Architected Framework

Use a defense-in-depth approach to control network traffic

AWS Network Firewall is a stateful, managed, network firewall and intrusion detection and prevention service for virtual private clouds (VPCs) in the AWS Cloud. It helps you deploy essential

network protections at the perimeter of the VPC. This includes filtering traffic going to and coming from an internet gateway, NAT gateway, or over VPN or AWS Direct Connect. Network Firewall includes features that help protect against common network threats. The stateful firewall in Network Firewall can incorporate context from traffic flows, such as connections and protocols, to enforce policies.

For more information, see the following resources:

- AWS Network Firewall documentation
- Control traffic at all layers in the AWS Well-Architected Framework

Security control recommendations for protecting data

The AWS Well-Architected Framework groups the best practices for protecting data into three categories: data classification, protecting data at rest, and protecting data in transit. The security controls in this section can help you implement best practices for data protection. These foundational best practices should be in place before you architect any workloads in the cloud. They prevent data mishandling, and they help you meet organizational, regulatory, and compliance obligations. Use the security controls in this section to implement best practices for data protection.

Controls in this section:

- Identify and classify data at the workload level
- Establish controls for each data classification level
- Encrypt data at rest
- Encrypt data in transit
- Block public access to Amazon EBS snapshots
- Block public access to Amazon RDS snapshots
- Block public access to Amazon RDS, Amazon Redshift, and AWS DMS resources
- Block public access to Amazon S3 buckets
- Require MFA to delete data in critical Amazon S3 buckets
- Configure Amazon OpenSearch Service domains in a VPC
- Configure alerts for AWS KMS key deletion
- Block public access to AWS KMS keys
- Configure load balancer listeners to use secure protocols

Identify and classify data at the workload level

Data classification is a process for identifying and categorizing the data in your network based on its criticality and sensitivity. It is a critical component of any cybersecurity risk management strategy because it helps you determine the appropriate protection and retention controls for the data. Data classification often reduces the frequency of data duplication. This can reduce storage and backup costs and accelerate searches.

We recommend that you understand the type and classification of data that your workload is processing, the associated business processes, where the data is stored, and who owns the data. Data classification helps workload owners to identify locations that store sensitive data and determine how that data should be accessed and shared. *Tags* are key-value pairs that act as metadata for organizing the AWS resources. Tags can help to manage, identify, organize, search for, and filter resources.

For more information, see the following resources:

- Data classification in AWS Whitepapers
- Identify the data within your workload in the AWS Well-Architected Framework

Establish controls for each data classification level

Define data protection controls for each classification level. For example, use recommended controls to secure data that is classified as public, and protect sensitive data with additional controls. Use mechanisms and tools that reduce or eliminate the need to directly access or manually process data. Automation of data identification and classification reduces the risk of misclassification, mishandling, modification, or human error.

For example, consider using Amazon Macie to scan Amazon Simple Storage Service (Amazon S3) buckets for sensitive data, such as personally identifiable information (PII). Also, you can automate the detection of unintended data access by using VPC Flow Logs in Amazon Virtual Private Cloud (Amazon VPC).

For more information, see the following resources:

- Define data protection controls in the AWS Well-Architected Framework
- Automate identification and classification in the AWS Well-Architected Framework
- AWS Privacy Reference Architecture (AWS PRA) in AWS Prescriptive Guidance
- Discovering sensitive data with Amazon Macie in the Macie documentation
- Logging IP traffic using VPC Flow Logs in the Amazon VPC documentation
- Common techniques to detect PHI and PII data using AWS services in the AWS for Industries blog

Encrypt data at rest

Data at rest is data that is stationary in your network, such as data that is in storage. Implementing encryption and appropriate access controls for data at rest helps reduce the risk of unauthorized access. *Encryption* is a computing process that transforms plaintext data, which Is human-readable, into ciphertext. You need an encryption key in order to decrypt the content back into plaintext so that it can be used. In the AWS Cloud, you can use AWS Key Management Service (AWS KMS) to create and control cryptographic keys that help protect your data.

As discussed in Establish controls for each data classification level, we recommend creating a policy that specifies what type of data requires encryption. Include criteria for how to determine which data should be encrypted and which data should be protected with another technique, such as tokenization or hashing.

For more information, see the following resources:

- Configuring default encryption in the Amazon S3 documentation
- Encryption by default for new EBS volumes and snapshot copies in the Amazon EC2 documentation
- Encrypting Amazon Aurora resources in the Amazon Aurora documentation
- Introduction to the cryptographic details of AWS KMS in the AWS KMS documentation
- Creating an enterprise encryption strategy for data at rest in AWS Prescriptive Guidance
- Enforce encryption at rest in the AWS Well-Architected Framework
- For more information about encryption in specific AWS services, see the AWS documentation for that service

Encrypt data in transit

Data in transit is data that is actively moving through your network, such as between network resources. Encrypt all data in transit by using secure TLS protocols and cipher suites. Network traffic between the resources and the internet must be encrypted in order to help prevent unauthorized access to the data. When possible, use TLS to encrypt network traffic within your internal AWS environment.

For more information, see the following resources:

- Requiring HTTPS for communication between viewers and CloudFront in the Amazon CloudFront documentation
- AWS PrivateLink documentation
- Enforce encryption in transit in the AWS Well-Architected Framework
- For more information about encryption in specific AWS services, see the AWS documentation for that service

Block public access to Amazon EBS snapshots

Amazon Elastic Block Store (Amazon EBS) provides block-level storage volumes for use with Amazon Elastic Compute Cloud (Amazon EC2) instances. You can back up the data on your Amazon EBS volumes to Amazon S3 by taking point-in-time snapshots. You can share snapshots publicly with all other AWS accounts, or you can share them privately with individual AWS accounts that you specify.

We recommend that you don't publicly share Amazon EBS snapshots. This might inadvertently expose sensitive data. When you share a snapshot, you are giving others access to the data in the snapshot. Share snapshots only with people that you trust with all of this data.

For more information, see the following resources:

- Share a snapshot in the Amazon EC2 documentation
- Amazon EBS snapshots should not be publicly restorable in the AWS Security Hub CSPM documentation
- ebs-snapshot-public-restorable-check in the AWS Config documentation

Block public access to Amazon RDS snapshots

Amazon Relational Database Service (Amazon RDS) helps you set up, operate, and scale a relational database in the AWS Cloud. Amazon RDS creates and saves automated backups of your database (DB) instance or Multi-AZ DB cluster during the backup window of your DB instance. Amazon RDS creates a storage volume snapshot of your DB instance, backing up the entire DB instance and not just individual databases. You can share a manual snapshot for the purposes of copying the snapshot or restoring a DB instance from it.

If you share a snapshot as public, make sure that none of the data in the snapshot is private or sensitive. When a snapshot is shared publicly, it gives all AWS accounts permission to access the data. This can result in unintended exposure of the data in your Amazon RDS instance.

For more information, see the following resources:

- Sharing a DB snapshot in the Amazon RDS documentation
- rds-snapshots-public-prohibited in the AWS Config documentation
- RDS snapshot should be private in the Security Hub CSPM documentation

Block public access to Amazon RDS, Amazon Redshift, and AWS DMS resources

You can configure Amazon RDS DB instances, Amazon Redshift clusters, and AWS Database Migration Service (AWS DMS) replication instances to be publicly accessible. If the `publiclyAccessible` field value is `true`, then these resources are publicly accessible. Allowing public access can result in unnecessary traffic, exposure, or data leaks. We recommend that you don't allow public access to these resources.

We recommend that you enable AWS Config rules or Security Hub CSPM controls to detect whether Amazon RDS DB instances, AWS DMS replication instances, or Amazon Redshift clusters allow public access.

> **Note**
>
> The public access settings for AWS DMS replication instances can't be modified after the instance has been provisioned. To change the public access setting, delete the current instance and then recreate it. When you recreate it, don't select the **Publicly accessible** option.

For more information, see the following resources:

- AWS DMS replication instances should not be public in the Security Hub CSPM documentation
- RDS DB Instances should prohibit public access in the Security Hub CSPM documentation
- Amazon Redshift clusters should prohibit public access in the Security Hub CSPM documentation

- rds-instance-public-access-check in the AWS Config documentation
- dms-replication-not-public in the AWS Config documentation
- redshift-cluster-public-access-check in the AWS Config documentation
- Modifying an Amazon RDS DB instance in the Amazon RDS documentation
- Modifying a cluster in the Amazon Redshift documentation

Block public access to Amazon S3 buckets

It's an Amazon S3 security best practice to ensure that your buckets are not publicly accessible. Unless you explicitly require anyone on the internet to be able to read or write to your bucket, make sure that your bucket is not public. This helps protect the integrity and security of the data. You can use AWS Config rules and Security Hub CSPM controls to confirm that your Amazon S3 buckets are compliant with this best practice.

For more information, see the following resources:

- Amazon S3 security best practices in the Amazon S3 documentation
- S3 Block Public Access setting should be enabled in the Security Hub CSPM documentation
- S3 buckets should prohibit public read access in the Security Hub CSPM documentation
- S3 buckets should prohibit public write access in the Security Hub CSPM documentation
- s3-bucket-public-read-prohibited rule in the AWS Config documentation
- s3-bucket-public-write-prohibited in the AWS Config documentation

Require MFA to delete data in critical Amazon S3 buckets

When working with S3 Versioning in Amazon S3 buckets, you can optionally add another layer of security by configuring a bucket to enable MFA (multi-factor authentication) delete. When you do this, the bucket owner must include two forms of authentication in any request to delete a version or change the versioning state of the bucket. We recommend that you enable this feature for buckets that contain data that's critical to your organization. This can prevent accidental bucket and data deletions.

For more information, see the following resources:

- Configuring MFA delete in the Amazon S3 documentation

Configure Amazon OpenSearch Service domains in a VPC

Amazon OpenSearch Service is a managed service that helps you deploy, operate, and scale OpenSearch clusters in the AWS Cloud. Amazon OpenSearch Service supports OpenSearch and legacy Elasticsearch open source software (OSS). Amazon OpenSearch Service domains that are deployed within a VPC can communicate with VPC resources over the private AWS network, without the need to traverse the public internet. This configuration improves your security posture by restricting access to the data in transit. We recommend that you don't attach Amazon OpenSearch Service domains to public subnets and that the VPC is configured according to best practices.

For more information, see the following resources:

- Launching your Amazon OpenSearch Service domains within a VPC in the Amazon OpenSearch Service documentation
- opensearch-in-vpc-only in the AWS Config documentation
- OpenSearch domains should be in a VPC in the Security Hub CSPM documentation

Configure alerts for AWS KMS key deletion

AWS Key Management Service (AWS KMS) keys cannot be recovered after they have been deleted. If a KMS key is delete, data that is still encrypted under that key is permanently unrecoverable. If you need to retain access to the data, before you delete the key, you must decrypt the data or reencrypt it with a new KMS key. You should delete a KMS key only when you are sure that you don't need to use it anymore.

We recommend that you configure an Amazon CloudWatch alarm that notifies you if someone initiates deletion of a KMS key. Because it is destructive and potentially dangerous to delete a KMS key, AWS KMS requires that you set a waiting period and schedule deletion in 7–30 days. This provides an opportunity to review the scheduled deletion and cancel it, if necessary.

For more information, see the following resources:

- Scheduling and canceling key deletion in the AWS KMS documentation
- Creating an alarm that detects use of a KMS key pending deletion in the AWS KMS documentation
- AWS KMS keys should not be deleted unintentionally in the Security Hub CSPM documentation

Block public access to AWS KMS keys

Key policies are the primary way to control access to AWS KMS keys. Every KMS key has exactly one key policy. Allowing anonymous access to KMS keys can lead to a sensitive data leak. We recommend that you identify any publicly accessible KMS keys and update their access policies in order to prevent unsigned requests made to these resources.

For more information, see the following resources:

- Security best practices for AWS Key Management Service in the AWS KMS documentation
- Changing a key policy in the AWS KMS documentation
- Determining access to AWS KMS keys in the AWS KMS documentation

Configure load balancer listeners to use secure protocols

Elastic Load Balancing automatically distributes incoming application traffic across multiple targets. You configure your load balancer to accept incoming traffic by specifying one or more *listeners*. A listener is a process that checks for connection requests, using the protocol and port that you configure. Each type of load balancer supports different protocols and ports:

- Application Load Balancers make routing decisions at the application layer and use HTTP or HTTPS protocols.
- Network Load Balancers make routing decisions at the transport layer and use TCP, TLS, UDP, or TCP_UDP protocols.
- Classic Load Balancers make routing decisions at either the transport layer (by using TCP or SSL protocols) or at the application layer (by using HTTP or HTTPS protocols).

We recommend that you always use HTTPS or TLS protocols. These protocols make sure that the load balancer is responsible for encrypting and decrypting the traffic between the client and the target.

For more information, see the following resources:

- Listeners for your Application Load Balancers in the Elastic Load Balancing documentation
- Listeners for your Classic Load Balancer in the Elastic Load Balancing documentation
- Listeners for your Network Load Balancers in the Elastic Load Balancing documentation

- Ensure AWS load balancers use secure listener protocols in AWS Prescriptive Guidance
- elb-tls-https-listeners-only in the AWS Config documentation
- Classic Load Balancer listeners should be configured with HTTPS or TLS termination in the Security Hub CSPM documentation
- Application Load Balancer should be configured to redirect all HTTP requests to HTTPS in the Security Hub CSPM documentation

Security recommendations for responding to incidents

When a security event occurs in your organization, your users must be prepared to respond to the issue. All users should have a basic understanding of your organization's security response processes. Planning, training, and experience are critical to a successful incident response program. Ideally, you prepare your organization before a potential security event occurs. The AWS Well-Architected Framework identifies three foundations that are required for a successful incident response program in the cloud: *preparation*, *operations*, and *post-incident activity*. For more information, see Aspects of AWS incident response in the AWS Well-Architected Framework.

With the exception of security controls that notify you about events or automatically respond to them, there are limited controls that you can establish for incident response. A strong incident response posture is primarily established through the plans, processes, runbooks, playbook, and training programs that you use in your organization. You can use the controls and recommendations in this section to implement best practices for your incident response program. For more information about best practices for incident response and implementation guidance, see Incident response in the AWS Well-Architected Framework.

Recommendations in this section:

- Define an incident response plan
- Create and maintain incident response runbooks and playbooks
- Implement event-driven security automation
- Document how operational teams should engage with Support
- Configure alerts for security events

Define an incident response plan

Establish a well-defined incident response plan (IRP). The incident response plan is designed to be the foundation for your incident response program. This plan must be customized to address the needs of each organization.

For more information, see the following resources:

- Develop and test an incident response plan in the *AWS Security Incident Response Guide*
- Develop incident management plans in the AWS Well-Architected Framework

- Identify key personnel and external resources in the AWS Well-Architected Framework

Create and maintain incident response runbooks and playbooks

A key part of preparing for an incident response processes is developing playbooks. Incident response playbooks provide a series of recommended steps that users follow when a security event occurs. Having a clear structure and steps simplifies the response and reduces the likelihood for human error.

For more information, see the following resources:

- What to create playbooks for in the *AWS Security Incident Response Guide*
- AWS incident response playbook samples on GitHub
- Develop and test security incident response playbooks in the AWS Well-Architected Framework

Implement event-driven security automation

Security response automation is a predefined and programmed action that is designed to automatically respond to or remediate a security event. These automations serve as detective or responsive security controls that help you implement AWS security best practices. Examples of automated response actions include modifying a VPC security group, patching an Amazon EC2 instance, or rotating credentials.

Many AWS services support automated responses. For example, you can configure an Amazon CloudWatch alarm for specific metrics, and the alarm can initiate action when the alarm changes state. Through Amazon EventBridge, you can also configure automated response and remediation for findings in AWS Security Hub CSPM and Amazon Inspector.

For more information please see the below resources:

- Remediate Amazon Inspector security findings automatically in the AWS Security Blog
- Get started with security response automation on AWS in the AWS Security Blog
- Automated security response on AWS in the AWS Solutions Library
- Using Amazon CloudWatch alarms in the CloudWatch documentation
- Automated response and remediation in the Security Hub CSPM documentation

- Creating custom responses to Amazon Inspector findings with Amazon EventBridge in the Amazon Inspector documentation

Document how operational teams should engage with Support

For your AWS account, you can define a primary contact and three alternate contacts. We recommend that you provide a security contact for each AWS account or for your organization.

AWS Support offers a range of plans that provide access to tools and expertise that can support the success and operational health of AWS solutions. Also, consider whether your organization would benefit from using AWS Managed Services instead of an Support plan. AWS Managed Services (AMS) helps you operate more efficiently and securely by providing ongoing management of your AWS infrastructure, including monitoring, incident management, security guidance, patch support, and backup for AWS workloads. The AMS support model can be a better fit for organizations that have limited resources on their cloud operations teams. We recommend that you compare these models and plans to choose the best fit for your organization use case and cloud maturity level.

For more information, see the following resources:

- Understand AWS response teams and support in the *AWS Security Incident Response Guide*
- Update the alternate contacts for your AWS account in the *AWS Account Management Guide*
- Compare Support Plans on the AWS website
- Strategy for using AWS Managed Services to achieve target business outcomes in AWS Prescriptive Guidance

Configure alerts for security events

Detecting an abnormality is as important as the measures implemented to control that abnormality. An alert is the main component of the detection phase. It generates a notification to initiate the incident response process based on AWS account activity of interest. Make sure that alerts include relevant information for the team to take action.

For more information, see the following resources:

- Detection in the *AWS Security Incident Response Guide*

- Prepare forensic capabilities in the AWS Well-Architected Framework
- Implement actionable security events in the AWS Well-Architected Framework

Next steps

As you continue on your cloud journey, it is important to apply these documented controls, guidance, and remediation options. These recommendations help improve your cloud security posture and help you meet your security responsibilities in the AWS Cloud, as defined in the AWS shared responsibility model.

For next steps, we recommend the following:

- For more information about best practices and implementation guidance, review the six pillars of the AWS Well-Architected Framework.
- For the AWS services that your organization uses, review the list of available AWS Security Hub CSPM controls and evaluate whether you should enable any of these controls in your environment.
- For the AWS services that your organization uses, review the list of available AWS Config managed rules and evaluate whether you should enable any of these rules in your environment.

Document history

The following table describes significant changes to this guide. If you want to be notified about future updates, you can subscribe to an RSS feed.

Change	Description	Date
MFA for root user	We updated the recommend ations and provided more information in the MFA for the root user section.	November 9, 2023
Initial publication	—	October 27, 2023

AWS Prescriptive Guidance glossary

The following are commonly used terms in strategies, guides, and patterns provided by AWS Prescriptive Guidance. To suggest entries, please use the **Provide feedback** link at the end of the glossary.

Numbers

7 Rs

Seven common migration strategies for moving applications to the cloud. These strategies build upon the 5 Rs that Gartner identified in 2011 and consist of the following:

- Refactor/re-architect – Move an application and modify its architecture by taking full advantage of cloud-native features to improve agility, performance, and scalability. This typically involves porting the operating system and database. Example: Migrate your on-premises Oracle database to the Amazon Aurora PostgreSQL-Compatible Edition.
- Replatform (lift and reshape) – Move an application to the cloud, and introduce some level of optimization to take advantage of cloud capabilities. Example: Migrate your on-premises Oracle database to Amazon Relational Database Service (Amazon RDS) for Oracle in the AWS Cloud.
- Repurchase (drop and shop) – Switch to a different product, typically by moving from a traditional license to a SaaS model. Example: Migrate your customer relationship management (CRM) system to Salesforce.com.
- Rehost (lift and shift) – Move an application to the cloud without making any changes to take advantage of cloud capabilities. Example: Migrate your on-premises Oracle database to Oracle on an EC2 instance in the AWS Cloud.
- Relocate (hypervisor-level lift and shift) – Move infrastructure to the cloud without purchasing new hardware, rewriting applications, or modifying your existing operations. You migrate servers from an on-premises platform to a cloud service for the same platform. Example: Migrate a Microsoft Hyper-V application to AWS.
- Retain (revisit) – Keep applications in your source environment. These might include applications that require major refactoring, and you want to postpone that work until a later time, and legacy applications that you want to retain, because there's no business justification for migrating them.

- Retire – Decommission or remove applications that are no longer needed in your source environment.

A

ABAC

See attribute-based access control.

abstracted services

See managed services.

ACID

See atomicity, consistency, isolation, durability.

active-active migration

A database migration method in which the source and target databases are kept in sync (by using a bidirectional replication tool or dual write operations), and both databases handle transactions from connecting applications during migration. This method supports migration in small, controlled batches instead of requiring a one-time cutover. It's more flexible but requires more work than active-passive migration.

active-passive migration

A database migration method in which the source and target databases are kept in sync, but only the source database handles transactions from connecting applications while data is replicated to the target database. The target database doesn't accept any transactions during migration.

aggregate function

A SQL function that operates on a group of rows and calculates a single return value for the group. Examples of aggregate functions include `SUM` and `MAX`.

AI

See artificial intelligence.

AIOps

See artificial intelligence operations.

anonymization

The process of permanently deleting personal information in a dataset. Anonymization can help protect personal privacy. Anonymized data is no longer considered to be personal data.

anti-pattern

A frequently used solution for a recurring issue where the solution is counter-productive, ineffective, or less effective than an alternative.

application control

A security approach that allows the use of only approved applications in order to help protect a system from malware.

application portfolio

A collection of detailed information about each application used by an organization, including the cost to build and maintain the application, and its business value. This information is key to the portfolio discovery and analysis process and helps identify and prioritize the applications to be migrated, modernized, and optimized.

artificial intelligence (AI)

The field of computer science that is dedicated to using computing technologies to perform cognitive functions that are typically associated with humans, such as learning, solving problems, and recognizing patterns. For more information, see What is Artificial Intelligence?

artificial intelligence operations (AIOps)

The process of using machine learning techniques to solve operational problems, reduce operational incidents and human intervention, and increase service quality. For more information about how AIOps is used in the AWS migration strategy, see the operations integration guide.

asymmetric encryption

An encryption algorithm that uses a pair of keys, a public key for encryption and a private key for decryption. You can share the public key because it isn't used for decryption, but access to the private key should be highly restricted.

atomicity, consistency, isolation, durability (ACID)

A set of software properties that guarantee the data validity and operational reliability of a database, even in the case of errors, power failures, or other problems.

attribute-based access control (ABAC)

The practice of creating fine-grained permissions based on user attributes, such as department, job role, and team name. For more information, see ABAC for AWS in the AWS Identity and Access Management (IAM) documentation.

authoritative data source

A location where you store the primary version of data, which is considered to be the most reliable source of information. You can copy data from the authoritative data source to other locations for the purposes of processing or modifying the data, such as anonymizing, redacting, or pseudonymizing it.

Availability Zone

A distinct location within an AWS Region that is insulated from failures in other Availability Zones and provides inexpensive, low-latency network connectivity to other Availability Zones in the same Region.

AWS Cloud Adoption Framework (AWS CAF)

A framework of guidelines and best practices from AWS to help organizations develop an efficient and effective plan to move successfully to the cloud. AWS CAF organizes guidance into six focus areas called perspectives: business, people, governance, platform, security, and operations. The business, people, and governance perspectives focus on business skills and processes; the platform, security, and operations perspectives focus on technical skills and processes. For example, the people perspective targets stakeholders who handle human resources (HR), staffing functions, and people management. For this perspective, AWS CAF provides guidance for people development, training, and communications to help ready the organization for successful cloud adoption. For more information, see the AWS CAF website and the AWS CAF whitepaper.

AWS Workload Qualification Framework (AWS WQF)

A tool that evaluates database migration workloads, recommends migration strategies, and provides work estimates. AWS WQF is included with AWS Schema Conversion Tool (AWS SCT). It analyzes database schemas and code objects, application code, dependencies, and performance characteristics, and provides assessment reports.

B

bad bot

A bot that is intended to disrupt or cause harm to individuals or organizations.

BCP

See business continuity planning.

behavior graph

A unified, interactive view of resource behavior and interactions over time. You can use a behavior graph with Amazon Detective to examine failed logon attempts, suspicious API calls, and similar actions. For more information, see Data in a behavior graph in the Detective documentation.

big-endian system

A system that stores the most significant byte first. See also endianness.

binary classification

A process that predicts a binary outcome (one of two possible classes). For example, your ML model might need to predict problems such as “Is this email spam or not spam?" or "Is this product a book or a car?"

bloom filter

A probabilistic, memory-efficient data structure that is used to test whether an element is a member of a set.

blue/green deployment

A deployment strategy where you create two separate but identical environments. You run the current application version in one environment (blue) and the new application version in the other environment (green). This strategy helps you quickly roll back with minimal impact.

bot

A software application that runs automated tasks over the internet and simulates human activity or interaction. Some bots are useful or beneficial, such as web crawlers that index information on the internet. Some other bots, known as *bad bots*, are intended to disrupt or cause harm to individuals or organizations.

botnet

Networks of bots that are infected by malware and are under the control of a single party, known as a *bot herder* or *bot operator*. Botnets are the best-known mechanism to scale bots and their impact.

branch

A contained area of a code repository. The first branch created in a repository is the *main branch*. You can create a new branch from an existing branch, and you can then develop features or fix bugs in the new branch. A branch you create to build a feature is commonly referred to as a *feature branch*. When the feature is ready for release, you merge the feature branch back into the main branch. For more information, see About branches (GitHub documentation).

break-glass access

In exceptional circumstances and through an approved process, a quick means for a user to gain access to an AWS account that they don't typically have permissions to access. For more information, see the Implement break-glass procedures indicator in the AWS Well-Architected guidance.

brownfield strategy

The existing infrastructure in your environment. When adopting a brownfield strategy for a system architecture, you design the architecture around the constraints of the current systems and infrastructure. If you are expanding the existing infrastructure, you might blend brownfield and greenfield strategies.

buffer cache

The memory area where the most frequently accessed data is stored.

business capability

What a business does to generate value (for example, sales, customer service, or marketing). Microservices architectures and development decisions can be driven by business capabilities. For more information, see the Organized around business capabilities section of the Running containerized microservices on AWS whitepaper.

business continuity planning (BCP)

A plan that addresses the potential impact of a disruptive event, such as a large-scale migration, on operations and enables a business to resume operations quickly.

C

CAF

See AWS Cloud Adoption Framework.

canary deployment

The slow and incremental release of a version to end users. When you are confident, you deploy the new version and replace the current version in its entirety.

CCoE

See Cloud Center of Excellence.

CDC

See change data capture.

change data capture (CDC)

The process of tracking changes to a data source, such as a database table, and recording metadata about the change. You can use CDC for various purposes, such as auditing or replicating changes in a target system to maintain synchronization.

chaos engineering

Intentionally introducing failures or disruptive events to test a system's resilience. You can use AWS Fault Injection Service (AWS FIS) to perform experiments that stress your AWS workloads and evaluate their response.

CI/CD

See continuous integration and continuous delivery.

classification

A categorization process that helps generate predictions. ML models for classification problems predict a discrete value. Discrete values are always distinct from one another. For example, a model might need to evaluate whether or not there is a car in an image.

client-side encryption

Encryption of data locally, before the target AWS service receives it.

Cloud Center of Excellence (CCoE)

A multi-disciplinary team that drives cloud adoption efforts across an organization, including developing cloud best practices, mobilizing resources, establishing migration timelines, and leading the organization through large-scale transformations. For more information, see the CCoE posts on the AWS Cloud Enterprise Strategy Blog.

cloud computing

The cloud technology that is typically used for remote data storage and IoT device management. Cloud computing is commonly connected to edge computing technology.

cloud operating model

In an IT organization, the operating model that is used to build, mature, and optimize one or more cloud environments. For more information, see Building your Cloud Operating Model.

cloud stages of adoption

The four phases that organizations typically go through when they migrate to the AWS Cloud:

- Project – Running a few cloud-related projects for proof of concept and learning purposes
- Foundation – Making foundational investments to scale your cloud adoption (e.g., creating a landing zone, defining a CCoE, establishing an operations model)
- Migration – Migrating individual applications
- Re-invention – Optimizing products and services, and innovating in the cloud

These stages were defined by Stephen Orban in the blog post The Journey Toward Cloud-First & the Stages of Adoption on the AWS Cloud Enterprise Strategy blog. For information about how they relate to the AWS migration strategy, see the migration readiness guide.

CMDB

See configuration management database.

code repository

A location where source code and other assets, such as documentation, samples, and scripts, are stored and updated through version control processes. Common cloud repositories include GitHub or Bitbucket Cloud. Each version of the code is called a *branch*. In a microservice structure, each repository is devoted to a single piece of functionality. A single CI/CD pipeline can use multiple repositories.

cold cache

A buffer cache that is empty, not well populated, or contains stale or irrelevant data. This affects performance because the database instance must read from the main memory or disk, which is slower than reading from the buffer cache.

cold data

Data that is rarely accessed and is typically historical. When querying this kind of data, slow queries are typically acceptable. Moving this data to lower-performing and less expensive storage tiers or classes can reduce costs.

computer vision (CV)

A field of AI that uses machine learning to analyze and extract information from visual formats such as digital images and videos. For example, Amazon SageMaker AI provides image processing algorithms for CV.

configuration drift

For a workload, a configuration change from the expected state. It might cause the workload to become noncompliant, and it's typically gradual and unintentional.

configuration management database (CMDB)

A repository that stores and manages information about a database and its IT environment, including both hardware and software components and their configurations. You typically use data from a CMDB in the portfolio discovery and analysis stage of migration.

conformance pack

A collection of AWS Config rules and remediation actions that you can assemble to customize your compliance and security checks. You can deploy a conformance pack as a single entity in an AWS account and Region, or across an organization, by using a YAML template. For more information, see Conformance packs in the AWS Config documentation.

continuous integration and continuous delivery (CI/CD)

The process of automating the source, build, test, staging, and production stages of the software release process. CI/CD is commonly described as a pipeline. CI/CD can help you automate processes, improve productivity, improve code quality, and deliver faster. For more information, see Benefits of continuous delivery. CD can also stand for *continuous deployment*. For more information, see Continuous Delivery vs. Continuous Deployment.

CV

See computer vision.

D

data at rest

Data that is stationary in your network, such as data that is in storage.

data classification

A process for identifying and categorizing the data in your network based on its criticality and sensitivity. It is a critical component of any cybersecurity risk management strategy because it helps you determine the appropriate protection and retention controls for the data. Data classification is a component of the security pillar in the AWS Well-Architected Framework. For more information, see Data classification.

data drift

A meaningful variation between the production data and the data that was used to train an ML model, or a meaningful change in the input data over time. Data drift can reduce the overall quality, accuracy, and fairness in ML model predictions.

data in transit

Data that is actively moving through your network, such as between network resources.

data mesh

An architectural framework that provides distributed, decentralized data ownership with centralized management and governance.

data minimization

The principle of collecting and processing only the data that is strictly necessary. Practicing data minimization in the AWS Cloud can reduce privacy risks, costs, and your analytics carbon footprint.

data perimeter

A set of preventive guardrails in your AWS environment that help make sure that only trusted identities are accessing trusted resources from expected networks. For more information, see Building a data perimeter on AWS.

data preprocessing

To transform raw data into a format that is easily parsed by your ML model. Preprocessing data can mean removing certain columns or rows and addressing missing, inconsistent, or duplicate values.

data provenance

The process of tracking the origin and history of data throughout its lifecycle, such as how the data was generated, transmitted, and stored.

data subject

An individual whose data is being collected and processed.

data warehouse

A data management system that supports business intelligence, such as analytics. Data warehouses commonly contain large amounts of historical data, and they are typically used for queries and analysis.

database definition language (DDL)

Statements or commands for creating or modifying the structure of tables and objects in a database.

database manipulation language (DML)

Statements or commands for modifying (inserting, updating, and deleting) information in a database.

DDL

See database definition language.

deep ensemble

To combine multiple deep learning models for prediction. You can use deep ensembles to obtain a more accurate prediction or for estimating uncertainty in predictions.

deep learning

An ML subfield that uses multiple layers of artificial neural networks to identify mapping between input data and target variables of interest.

defense-in-depth

An information security approach in which a series of security mechanisms and controls are thoughtfully layered throughout a computer network to protect the confidentiality, integrity, and availability of the network and the data within. When you adopt this strategy on AWS, you add multiple controls at different layers of the AWS Organizations structure to help secure resources. For example, a defense-in-depth approach might combine multi-factor authentication, network segmentation, and encryption.

delegated administrator

In AWS Organizations, a compatible service can register an AWS member account to administer the organization's accounts and manage permissions for that service. This account is called the *delegated administrator* for that service. For more information and a list of compatible services, see Services that work with AWS Organizations in the AWS Organizations documentation.

deployment

The process of making an application, new features, or code fixes available in the target environment. Deployment involves implementing changes in a code base and then building and running that code base in the application's environments.

development environment

See environment.

detective control

A security control that is designed to detect, log, and alert after an event has occurred. These controls are a second line of defense, alerting you to security events that bypassed the preventative controls in place. For more information, see Detective controls in *Implementing security controls on AWS*.

development value stream mapping (DVSM)

A process used to identify and prioritize constraints that adversely affect speed and quality in a software development lifecycle. DVSM extends the value stream mapping process originally designed for lean manufacturing practices. It focuses on the steps and teams required to create and move value through the software development process.

digital twin

A virtual representation of a real-world system, such as a building, factory, industrial equipment, or production line. Digital twins support predictive maintenance, remote monitoring, and production optimization.

dimension table

In a star schema, a smaller table that contains data attributes about quantitative data in a fact table. Dimension table attributes are typically text fields or discrete numbers that behave like text. These attributes are commonly used for query constraining, filtering, and result set labeling.

disaster

An event that prevents a workload or system from fulfilling its business objectives in its primary deployed location. These events can be natural disasters, technical failures, or the result of human actions, such as unintentional misconfiguration or a malware attack.

disaster recovery (DR)

The strategy and process you use to minimize downtime and data loss caused by a disaster. For more information, see Disaster Recovery of Workloads on AWS: Recovery in the Cloud in the AWS Well-Architected Framework.

DML

See database manipulation language.

domain-driven design

An approach to developing a complex software system by connecting its components to evolving domains, or core business goals, that each component serves. This concept was introduced by Eric Evans in his book, *Domain-Driven Design: Tackling Complexity in the Heart of Software* (Boston: Addison-Wesley Professional, 2003). For information about how you can use domain-driven design with the strangler fig pattern, see Modernizing legacy Microsoft ASP.NET (ASMX) web services incrementally by using containers and Amazon API Gateway.

DR

See disaster recovery.

drift detection

Tracking deviations from a baselined configuration. For example, you can use AWS CloudFormation to detect drift in system resources, or you can use AWS Control Tower to detect changes in your landing zone that might affect compliance with governance requirements.

DVSM

See development value stream mapping.

E

EDA

See exploratory data analysis.

EDI

See electronic data interchange.

edge computing

The technology that increases the computing power for smart devices at the edges of an IoT network. When compared with cloud computing, edge computing can reduce communication latency and improve response time.

electronic data interchange (EDI)

The automated exchange of business documents between organizations. For more information, see What is Electronic Data Interchange.

encryption

A computing process that transforms plaintext data, which is human-readable, into ciphertext.

encryption key

A cryptographic string of randomized bits that is generated by an encryption algorithm. Keys can vary in length, and each key is designed to be unpredictable and unique.

endianness

The order in which bytes are stored in computer memory. Big-endian systems store the most significant byte first. Little-endian systems store the least significant byte first.

endpoint

See service endpoint.

endpoint service

A service that you can host in a virtual private cloud (VPC) to share with other users. You can create an endpoint service with AWS PrivateLink and grant permissions to other AWS accounts or to AWS Identity and Access Management (IAM) principals. These accounts or principals can connect to your endpoint service privately by creating interface VPC endpoints. For more information, see Create an endpoint service in the Amazon Virtual Private Cloud (Amazon VPC) documentation.

enterprise resource planning (ERP)

A system that automates and manages key business processes (such as accounting, MES, and project management) for an enterprise.

envelope encryption

The process of encrypting an encryption key with another encryption key. For more information, see Envelope encryption in the AWS Key Management Service (AWS KMS) documentation.

environment

An instance of a running application. The following are common types of environments in cloud computing:

- development environment – An instance of a running application that is available only to the core team responsible for maintaining the application. Development environments are used to test changes before promoting them to upper environments. This type of environment is sometimes referred to as a *test environment*.
- lower environments – All development environments for an application, such as those used for initial builds and tests.
- production environment – An instance of a running application that end users can access. In a CI/CD pipeline, the production environment is the last deployment environment.
- upper environments – All environments that can be accessed by users other than the core development team. This can include a production environment, preproduction environments, and environments for user acceptance testing.

epic

In agile methodologies, functional categories that help organize and prioritize your work. Epics provide a high-level description of requirements and implementation tasks. For example, AWS CAF security epics include identity and access management, detective controls, infrastructure security, data protection, and incident response. For more information about epics in the AWS migration strategy, see the program implementation guide.

ERP

See enterprise resource planning.

exploratory data analysis (EDA)

The process of analyzing a dataset to understand its main characteristics. You collect or aggregate data and then perform initial investigations to find patterns, detect anomalies, and check assumptions. EDA is performed by calculating summary statistics and creating data visualizations.

F

fact table

The central table in a star schema. It stores quantitative data about business operations. Typically, a fact table contains two types of columns: those that contain measures and those that contain a foreign key to a dimension table.

fail fast

A philosophy that uses frequent and incremental testing to reduce the development lifecycle. It is a critical part of an agile approach.

fault isolation boundary

In the AWS Cloud, a boundary such as an Availability Zone, AWS Region, control plane, or data plane that limits the effect of a failure and helps improve the resilience of workloads. For more information, see AWS Fault Isolation Boundaries.

feature branch

See branch.

features

The input data that you use to make a prediction. For example, in a manufacturing context, features could be images that are periodically captured from the manufacturing line.

feature importance

How significant a feature is for a model's predictions. This is usually expressed as a numerical score that can be calculated through various techniques, such as Shapley Additive Explanations (SHAP) and integrated gradients. For more information, see Machine learning model interpretability with AWS.

feature transformation

To optimize data for the ML process, including enriching data with additional sources, scaling values, or extracting multiple sets of information from a single data field. This enables the ML model to benefit from the data. For example, if you break down the "2021-05-27 00:15:37" date into "2021", "May", "Thu", and "15", you can help the learning algorithm learn nuanced patterns associated with different data components.

few-shot prompting

Providing an LLM with a small number of examples that demonstrate the task and desired output before asking it to perform a similar task. This technique is an application of in-context learning, where models learn from examples (*shots*) that are embedded in prompts. Few-shot prompting can be effective for tasks that require specific formatting, reasoning, or domain knowledge. See also zero-shot prompting.

FGAC

See fine-grained access control.

fine-grained access control (FGAC)

The use of multiple conditions to allow or deny an access request.

flash-cut migration

A database migration method that uses continuous data replication through change data capture to migrate data in the shortest time possible, instead of using a phased approach. The objective is to keep downtime to a minimum.

FM

See foundation model.

foundation model (FM)

A large deep-learning neural network that has been training on massive datasets of generalized and unlabeled data. FMs are capable of performing a wide variety of general tasks, such as understanding language, generating text and images, and conversing in natural language. For more information, see What are Foundation Models.

G

generative AI

A subset of AI models that have been trained on large amounts of data and that can use a simple text prompt to create new content and artifacts, such as images, videos, text, and audio. For more information, see What is Generative AI.

geo blocking

See geographic restrictions.

geographic restrictions (geo blocking)

In Amazon CloudFront, an option to prevent users in specific countries from accessing content distributions. You can use an allow list or block list to specify approved and banned countries. For more information, see Restricting the geographic distribution of your content in the CloudFront documentation.

Gitflow workflow

An approach in which lower and upper environments use different branches in a source code repository. The Gitflow workflow is considered legacy, and the trunk-based workflow is the modern, preferred approach.

golden image

A snapshot of a system or software that is used as a template to deploy new instances of that system or software. For example, in manufacturing, a golden image can be used to provision software on multiple devices and helps improve speed, scalability, and productivity in device manufacturing operations.

greenfield strategy

The absence of existing infrastructure in a new environment. When adopting a greenfield strategy for a system architecture, you can select all new technologies without the restriction

of compatibility with existing infrastructure, also known as brownfield. If you are expanding the existing infrastructure, you might blend brownfield and greenfield strategies.

guardrail

A high-level rule that helps govern resources, policies, and compliance across organizational units (OUs). *Preventive guardrails* enforce policies to ensure alignment to compliance standards. They are implemented by using service control policies and IAM permissions boundaries. *Detective guardrails* detect policy violations and compliance issues, and generate alerts for remediation. They are implemented by using AWS Config, AWS Security Hub CSPM, Amazon GuardDuty, AWS Trusted Advisor, Amazon Inspector, and custom AWS Lambda checks.

H

HA

See high availability.

heterogeneous database migration

Migrating your source database to a target database that uses a different database engine (for example, Oracle to Amazon Aurora). Heterogeneous migration is typically part of a re-architecting effort, and converting the schema can be a complex task. AWS provides AWS SCT that helps with schema conversions.

high availability (HA)

The ability of a workload to operate continuously, without intervention, in the event of challenges or disasters. HA systems are designed to automatically fail over, consistently deliver high-quality performance, and handle different loads and failures with minimal performance impact.

historian modernization

An approach used to modernize and upgrade operational technology (OT) systems to better serve the needs of the manufacturing industry. A *historian* is a type of database that is used to collect and store data from various sources in a factory.

holdout data

A portion of historical, labeled data that is withheld from a dataset that is used to train a machine learning model. You can use holdout data to evaluate the model performance by comparing the model predictions against the holdout data.

homogeneous database migration

Migrating your source database to a target database that shares the same database engine (for example, Microsoft SQL Server to Amazon RDS for SQL Server). Homogeneous migration is typically part of a rehosting or replatforming effort. You can use native database utilities to migrate the schema.

hot data

Data that is frequently accessed, such as real-time data or recent translational data. This data typically requires a high-performance storage tier or class to provide fast query responses.

hotfix

An urgent fix for a critical issue in a production environment. Due to its urgency, a hotfix is usually made outside of the typical DevOps release workflow.

hypercare period

Immediately following cutover, the period of time when a migration team manages and monitors the migrated applications in the cloud in order to address any issues. Typically, this period is 1–4 days in length. At the end of the hypercare period, the migration team typically transfers responsibility for the applications to the cloud operations team.

I

IaC

See infrastructure as code.

identity-based policy

A policy attached to one or more IAM principals that defines their permissions within the AWS Cloud environment.

idle application

An application that has an average CPU and memory usage between 5 and 20 percent over a period of 90 days. In a migration project, it is common to retire these applications or retain them on premises.

IIoT

See industrial Internet of Things.

immutable infrastructure

A model that deploys new infrastructure for production workloads instead of updating, patching, or modifying the existing infrastructure. Immutable infrastructures are inherently more consistent, reliable, and predictable than mutable infrastructure. For more information, see the Deploy using immutable infrastructure best practice in the AWS Well-Architected Framework.

inbound (ingress) VPC

In an AWS multi-account architecture, a VPC that accepts, inspects, and routes network connections from outside an application. The AWS Security Reference Architecture recommends setting up your Network account with inbound, outbound, and inspection VPCs to protect the two-way interface between your application and the broader internet.

incremental migration

A cutover strategy in which you migrate your application in small parts instead of performing a single, full cutover. For example, you might move only a few microservices or users to the new system initially. After you verify that everything is working properly, you can incrementally move additional microservices or users until you can decommission your legacy system. This strategy reduces the risks associated with large migrations.

Industry 4.0

A term that was introduced by Klaus Schwab in 2016 to refer to the modernization of manufacturing processes through advances in connectivity, real-time data, automation, analytics, and AI/ML.

infrastructure

All of the resources and assets contained within an application's environment.

infrastructure as code (IaC)

The process of provisioning and managing an application's infrastructure through a set of configuration files. IaC is designed to help you centralize infrastructure management, standardize resources, and scale quickly so that new environments are repeatable, reliable, and consistent.

industrial Internet of Things (IIoT)

The use of internet-connected sensors and devices in the industrial sectors, such as manufacturing, energy, automotive, healthcare, life sciences, and agriculture. For more information, see Building an industrial Internet of Things (IIoT) digital transformation strategy.

inspection VPC

In an AWS multi-account architecture, a centralized VPC that manages inspections of network traffic between VPCs (in the same or different AWS Regions), the internet, and on-premises networks. The AWS Security Reference Architecture recommends setting up your Network account with inbound, outbound, and inspection VPCs to protect the two-way interface between your application and the broader internet.

Internet of Things (IoT)

The network of connected physical objects with embedded sensors or processors that communicate with other devices and systems through the internet or over a local communication network. For more information, see What is IoT?

interpretability

A characteristic of a machine learning model that describes the degree to which a human can understand how the model's predictions depend on its inputs. For more information, see Machine learning model interpretability with AWS.

IoT

See Internet of Things.

IT information library (ITIL)

A set of best practices for delivering IT services and aligning these services with business requirements. ITIL provides the foundation for ITSM.

IT service management (ITSM)

Activities associated with designing, implementing, managing, and supporting IT services for an organization. For information about integrating cloud operations with ITSM tools, see the operations integration guide.

ITIL

See IT information library.

ITSM

See IT service management.

L

label-based access control (LBAC)

An implementation of mandatory access control (MAC) where the users and the data itself are each explicitly assigned a security label value. The intersection between the user security label and data security label determines which rows and columns can be seen by the user.

landing zone

A landing zone is a well-architected, multi-account AWS environment that is scalable and secure. This is a starting point from which your organizations can quickly launch and deploy workloads and applications with confidence in their security and infrastructure environment. For more information about landing zones, see Setting up a secure and scalable multi-account AWS environment.

large language model (LLM)

A deep learning AI model that is pretrained on a vast amount of data. An LLM can perform multiple tasks, such as answering questions, summarizing documents, translating text into other languages, and completing sentences. For more information, see What are LLMs.

large migration

A migration of 300 or more servers.

LBAC

See label-based access control.

least privilege

The security best practice of granting the minimum permissions required to perform a task. For more information, see Apply least-privilege permissions in the IAM documentation.

lift and shift

See 7 Rs.

little-endian system

A system that stores the least significant byte first. See also endianness.

LLM

See large language model.

lower environments

See environment.

M

machine learning (ML)

A type of artificial intelligence that uses algorithms and techniques for pattern recognition and learning. ML analyzes and learns from recorded data, such as Internet of Things (IoT) data, to generate a statistical model based on patterns. For more information, see Machine Learning.

main branch

See branch.

malware

Software that is designed to compromise computer security or privacy. Malware might disrupt computer systems, leak sensitive information, or gain unauthorized access. Examples of malware include viruses, worms, ransomware, Trojan horses, spyware, and keyloggers.

managed services

AWS services for which AWS operates the infrastructure layer, the operating system, and platforms, and you access the endpoints to store and retrieve data. Amazon Simple Storage Service (Amazon S3) and Amazon DynamoDB are examples of managed services. These are also known as *abstracted services*.

manufacturing execution system (MES)

A software system for tracking, monitoring, documenting, and controlling production processes that convert raw materials to finished products on the shop floor.

MAP

See Migration Acceleration Program.

mechanism

A complete process in which you create a tool, drive adoption of the tool, and then inspect the results in order to make adjustments. A mechanism is a cycle that reinforces and improves itself as it operates. For more information, see Building mechanisms in the AWS Well-Architected Framework.

member account

All AWS accounts other than the management account that are part of an organization in AWS Organizations. An account can be a member of only one organization at a time.

MES

See manufacturing execution system.

Message Queuing Telemetry Transport (MQTT)

A lightweight, machine-to-machine (M2M) communication protocol, based on the publish/subscribe pattern, for resource-constrained IoT devices.

microservice

A small, independent service that communicates over well-defined APIs and is typically owned by small, self-contained teams. For example, an insurance system might include microservices that map to business capabilities, such as sales or marketing, or subdomains, such as purchasing, claims, or analytics. The benefits of microservices include agility, flexible scaling, easy deployment, reusable code, and resilience. For more information, see Integrating microservices by using AWS serverless services.

microservices architecture

An approach to building an application with independent components that run each application process as a microservice. These microservices communicate through a well-defined interface by using lightweight APIs. Each microservice in this architecture can be updated, deployed,

and scaled to meet demand for specific functions of an application. For more information, see Implementing microservices on AWS.

Migration Acceleration Program (MAP)

An AWS program that provides consulting support, training, and services to help organizations build a strong operational foundation for moving to the cloud, and to help offset the initial cost of migrations. MAP includes a migration methodology for executing legacy migrations in a methodical way and a set of tools to automate and accelerate common migration scenarios.

migration at scale

The process of moving the majority of the application portfolio to the cloud in waves, with more applications moved at a faster rate in each wave. This phase uses the best practices and lessons learned from the earlier phases to implement a *migration factory* of teams, tools, and processes to streamline the migration of workloads through automation and agile delivery. This is the third phase of the AWS migration strategy.

migration factory

Cross-functional teams that streamline the migration of workloads through automated, agile approaches. Migration factory teams typically include operations, business analysts and owners, migration engineers, developers, and DevOps professionals working in sprints. Between 20 and 50 percent of an enterprise application portfolio consists of repeated patterns that can be optimized by a factory approach. For more information, see the discussion of migration factories and the Cloud Migration Factory guide in this content set.

migration metadata

The information about the application and server that is needed to complete the migration. Each migration pattern requires a different set of migration metadata. Examples of migration metadata include the target subnet, security group, and AWS account.

migration pattern

A repeatable migration task that details the migration strategy, the migration destination, and the migration application or service used. Example: Rehost migration to Amazon EC2 with AWS Application Migration Service.

Migration Portfolio Assessment (MPA)

An online tool that provides information for validating the business case for migrating to the AWS Cloud. MPA provides detailed portfolio assessment (server right-sizing, pricing, TCO

comparisons, migration cost analysis) as well as migration planning (application data analysis and data collection, application grouping, migration prioritization, and wave planning). The MPA tool (requires login) is available free of charge to all AWS consultants and APN Partner consultants.

Migration Readiness Assessment (MRA)

The process of gaining insights about an organization's cloud readiness status, identifying strengths and weaknesses, and building an action plan to close identified gaps, using the AWS CAF. For more information, see the migration readiness guide. MRA is the first phase of the AWS migration strategy.

migration strategy

The approach used to migrate a workload to the AWS Cloud. For more information, see the 7 Rs entry in this glossary and see Mobilize your organization to accelerate large-scale migrations.

ML

See machine learning.

modernization

Transforming an outdated (legacy or monolithic) application and its infrastructure into an agile, elastic, and highly available system in the cloud to reduce costs, gain efficiencies, and take advantage of innovations. For more information, see Strategy for modernizing applications in the AWS Cloud.

modernization readiness assessment

An evaluation that helps determine the modernization readiness of an organization's applications; identifies benefits, risks, and dependencies; and determines how well the organization can support the future state of those applications. The outcome of the assessment is a blueprint of the target architecture, a roadmap that details development phases and milestones for the modernization process, and an action plan for addressing identified gaps. For more information, see Evaluating modernization readiness for applications in the AWS Cloud.

monolithic applications (monoliths)

Applications that run as a single service with tightly coupled processes. Monolithic applications have several drawbacks. If one application feature experiences a spike in demand, the entire architecture must be scaled. Adding or improving a monolithic application's features also becomes more complex when the code base grows. To address these issues, you can

use a microservices architecture. For more information, see Decomposing monoliths into microservices.

MPA

See Migration Portfolio Assessment.

MQTT

See Message Queuing Telemetry Transport.

multiclass classification

A process that helps generate predictions for multiple classes (predicting one of more than two outcomes). For example, an ML model might ask "Is this product a book, car, or phone?" or "Which product category is most interesting to this customer?"

mutable infrastructure

A model that updates and modifies the existing infrastructure for production workloads. For improved consistency, reliability, and predictability, the AWS Well-Architected Framework recommends the use of immutable infrastructure as a best practice.

O

OAC

See origin access control.

OAI

See origin access identity.

OCM

See organizational change management.

offline migration

A migration method in which the source workload is taken down during the migration process. This method involves extended downtime and is typically used for small, non-critical workloads.

OI

See operations integration.

OLA

See operational-level agreement.

online migration

A migration method in which the source workload is copied to the target system without being taken offline. Applications that are connected to the workload can continue to function during the migration. This method involves zero to minimal downtime and is typically used for critical production workloads.

OPC-UA

See Open Process Communications - Unified Architecture.

Open Process Communications - Unified Architecture (OPC-UA)

A machine-to-machine (M2M) communication protocol for industrial automation. OPC-UA provides an interoperability standard with data encryption, authentication, and authorization schemes.

operational-level agreement (OLA)

An agreement that clarifies what functional IT groups promise to deliver to each other, to support a service-level agreement (SLA).

operational readiness review (ORR)

A checklist of questions and associated best practices that help you understand, evaluate, prevent, or reduce the scope of incidents and possible failures. For more information, see Operational Readiness Reviews (ORR) in the AWS Well-Architected Framework.

operational technology (OT)

Hardware and software systems that work with the physical environment to control industrial operations, equipment, and infrastructure. In manufacturing, the integration of OT and information technology (IT) systems is a key focus for Industry 4.0 transformations.

operations integration (OI)

The process of modernizing operations in the cloud, which involves readiness planning, automation, and integration. For more information, see the operations integration guide.

organization trail

A trail that's created by AWS CloudTrail that logs all events for all AWS accounts in an organization in AWS Organizations. This trail is created in each AWS account that's part of the

organization and tracks the activity in each account. For more information, see Creating a trail for an organization in the CloudTrail documentation.

organizational change management (OCM)

A framework for managing major, disruptive business transformations from a people, culture, and leadership perspective. OCM helps organizations prepare for, and transition to, new systems and strategies by accelerating change adoption, addressing transitional issues, and driving cultural and organizational changes. In the AWS migration strategy, this framework is called *people acceleration*, because of the speed of change required in cloud adoption projects. For more information, see the OCM guide.

origin access control (OAC)

In CloudFront, an enhanced option for restricting access to secure your Amazon Simple Storage Service (Amazon S3) content. OAC supports all S3 buckets in all AWS Regions, server-side encryption with AWS KMS (SSE-KMS), and dynamic PUT and DELETE requests to the S3 bucket.

origin access identity (OAI)

In CloudFront, an option for restricting access to secure your Amazon S3 content. When you use OAI, CloudFront creates a principal that Amazon S3 can authenticate with. Authenticated principals can access content in an S3 bucket only through a specific CloudFront distribution. See also OAC, which provides more granular and enhanced access control.

ORR

See operational readiness review.

OT

See operational technology.

outbound (egress) VPC

In an AWS multi-account architecture, a VPC that handles network connections that are initiated from within an application. The AWS Security Reference Architecture recommends setting up your Network account with inbound, outbound, and inspection VPCs to protect the two-way interface between your application and the broader internet.

P

permissions boundary

An IAM management policy that is attached to IAM principals to set the maximum permissions that the user or role can have. For more information, see Permissions boundaries in the IAM documentation.

personally identifiable information (PII)

Information that, when viewed directly or paired with other related data, can be used to reasonably infer the identity of an individual. Examples of PII include names, addresses, and contact information.

PII

See personally identifiable information.

playbook

A set of predefined steps that capture the work associated with migrations, such as delivering core operations functions in the cloud. A playbook can take the form of scripts, automated runbooks, or a summary of processes or steps required to operate your modernized environment.

PLC

See programmable logic controller.

PLM

See product lifecycle management.

policy

An object that can define permissions (see identity-based policy), specify access conditions (see resource-based policy), or define the maximum permissions for all accounts in an organization in AWS Organizations (see service control policy).

polyglot persistence

Independently choosing a microservice's data storage technology based on data access patterns and other requirements. If your microservices have the same data storage technology, they can encounter implementation challenges or experience poor performance. Microservices are more

easily implemented and achieve better performance and scalability if they use the data store best adapted to their requirements.

portfolio assessment

A process of discovering, analyzing, and prioritizing the application portfolio in order to plan the migration. For more information, see Evaluating migration readiness.

predicate

A query condition that returns `true` or `false`, commonly located in a `WHERE` clause.

predicate pushdown

A database query optimization technique that filters the data in the query before transfer. This reduces the amount of data that must be retrieved and processed from the relational database, and it improves query performance.

preventative control

A security control that is designed to prevent an event from occurring. These controls are a first line of defense to help prevent unauthorized access or unwanted changes to your network. For more information, see Preventative controls in *Implementing security controls on AWS*.

principal

An entity in AWS that can perform actions and access resources. This entity is typically a root user for an AWS account, an IAM role, or a user. For more information, see *Principal* in Roles terms and concepts in the IAM documentation.

privacy by design

A system engineering approach that takes privacy into account through the whole development process.

private hosted zones

A container that holds information about how you want Amazon Route 53 to respond to DNS queries for a domain and its subdomains within one or more VPCs. For more information, see Working with private hosted zones in the Route 53 documentation.

proactive control

A security control designed to prevent the deployment of noncompliant resources. These controls scan resources before they are provisioned. If the resource is not compliant with the control, then it isn't provisioned. For more information, see the Controls reference guide in the

AWS Control Tower documentation and see Proactive controls in *Implementing security controls on AWS*.

product lifecycle management (PLM)

The management of data and processes for a product throughout its entire lifecycle, from design, development, and launch, through growth and maturity, to decline and removal.

production environment

See environment.

programmable logic controller (PLC)

In manufacturing, a highly reliable, adaptable computer that monitors machines and automates manufacturing processes.

prompt chaining

Using the output of one LLM prompt as the input for the next prompt to generate better responses. This technique is used to break down a complex task into subtasks, or to iteratively refine or expand a preliminary response. It helps improve the accuracy and relevance of a model's responses and allows for more granular, personalized results.

pseudonymization

The process of replacing personal identifiers in a dataset with placeholder values. Pseudonymization can help protect personal privacy. Pseudonymized data is still considered to be personal data.

publish/subscribe (pub/sub)

A pattern that enables asynchronous communications among microservices to improve scalability and responsiveness. For example, in a microservices-based MES, a microservice can publish event messages to a channel that other microservices can subscribe to. The system can add new microservices without changing the publishing service.

Q

query plan

A series of steps, like instructions, that are used to access the data in a SQL relational database system.

query plan regression

When a database service optimizer chooses a less optimal plan than it did before a given change to the database environment. This can be caused by changes to statistics, constraints, environment settings, query parameter bindings, and updates to the database engine.

R

RACI matrix

See responsible, accountable, consulted, informed (RACI).

RAG

See Retrieval Augmented Generation.

ransomware

A malicious software that is designed to block access to a computer system or data until a payment is made.

RASCI matrix

See responsible, accountable, consulted, informed (RACI).

RCAC

See row and column access control.

read replica

A copy of a database that's used for read-only purposes. You can route queries to the read replica to reduce the load on your primary database.

re-architect

See 7 Rs.

recovery point objective (RPO)

The maximum acceptable amount of time since the last data recovery point. This determines what is considered an acceptable loss of data between the last recovery point and the interruption of service.

recovery time objective (RTO)

The maximum acceptable delay between the interruption of service and restoration of service.

refactor

See 7 Rs.

Region

A collection of AWS resources in a geographic area. Each AWS Region is isolated and independent of the others to provide fault tolerance, stability, and resilience. For more information, see Specify which AWS Regions your account can use.

regression

An ML technique that predicts a numeric value. For example, to solve the problem of "What price will this house sell for?" an ML model could use a linear regression model to predict a house's sale price based on known facts about the house (for example, the square footage).

rehost

See 7 Rs.

release

In a deployment process, the act of promoting changes to a production environment.

relocate

See 7 Rs.

replatform

See 7 Rs.

repurchase

See 7 Rs.

resiliency

An application's ability to resist or recover from disruptions. High availability and disaster recovery are common considerations when planning for resiliency in the AWS Cloud. For more information, see AWS Cloud Resilience.

resource-based policy

A policy attached to a resource, such as an Amazon S3 bucket, an endpoint, or an encryption key. This type of policy specifies which principals are allowed access, supported actions, and any other conditions that must be met.

responsible, accountable, consulted, informed (RACI) matrix

A matrix that defines the roles and responsibilities for all parties involved in migration activities and cloud operations. The matrix name is derived from the responsibility types defined in the matrix: responsible (R), accountable (A), consulted (C), and informed (I). The support (S) type is optional. If you include support, the matrix is called a *RASCI matrix*, and if you exclude it, it's called a *RACI matrix*.

responsive control

A security control that is designed to drive remediation of adverse events or deviations from your security baseline. For more information, see Responsive controls in *Implementing security controls on AWS*.

retain

See 7 Rs.

retire

See 7 Rs.

Retrieval Augmented Generation (RAG)

A generative AI technology in which an LLM references an authoritative data source that is outside of its training data sources before generating a response. For example, a RAG model might perform a semantic search of an organization's knowledge base or custom data. For more information, see What is RAG.

rotation

The process of periodically updating a secret to make it more difficult for an attacker to access the credentials.

row and column access control (RCAC)

The use of basic, flexible SQL expressions that have defined access rules. RCAC consists of row permissions and column masks.

RPO

See recovery point objective.

RTO

See recovery time objective.

runbook

A set of manual or automated procedures required to perform a specific task. These are typically built to streamline repetitive operations or procedures with high error rates.

S

SAML 2.0

An open standard that many identity providers (IdPs) use. This feature enables federated single sign-on (SSO), so users can log into the AWS Management Console or call the AWS API operations without you having to create user in IAM for everyone in your organization. For more information about SAML 2.0-based federation, see About SAML 2.0-based federation in the IAM documentation.

SCADA

See supervisory control and data acquisition.

SCP

See service control policy.

secret

In AWS Secrets Manager, confidential or restricted information, such as a password or user credentials, that you store in encrypted form. It consists of the secret value and its metadata. The secret value can be binary, a single string, or multiple strings. For more information, see What's in a Secrets Manager secret? in the Secrets Manager documentation.

security by design

A system engineering approach that takes security into account through the whole development process.

security control

A technical or administrative guardrail that prevents, detects, or reduces the ability of a threat actor to exploit a security vulnerability. There are four primary types of security controls: preventative, detective, responsive, and proactive.

security hardening

The process of reducing the attack surface to make it more resistant to attacks. This can include actions such as removing resources that are no longer needed, implementing the security best practice of granting least privilege, or deactivating unnecessary features in configuration files.

security information and event management (SIEM) system

Tools and services that combine security information management (SIM) and security event management (SEM) systems. A SIEM system collects, monitors, and analyzes data from servers, networks, devices, and other sources to detect threats and security breaches, and to generate alerts.

security response automation

A predefined and programmed action that is designed to automatically respond to or remediate a security event. These automations serve as detective or responsive security controls that help you implement AWS security best practices. Examples of automated response actions include modifying a VPC security group, patching an Amazon EC2 instance, or rotating credentials.

server-side encryption

Encryption of data at its destination, by the AWS service that receives it.

service control policy (SCP)

A policy that provides centralized control over permissions for all accounts in an organization in AWS Organizations. SCPs define guardrails or set limits on actions that an administrator can delegate to users or roles. You can use SCPs as allow lists or deny lists, to specify which services or actions are permitted or prohibited. For more information, see Service control policies in the AWS Organizations documentation.

service endpoint

The URL of the entry point for an AWS service. You can use the endpoint to connect programmatically to the target service. For more information, see AWS service endpoints in *AWS General Reference*.

service-level agreement (SLA)

An agreement that clarifies what an IT team promises to deliver to their customers, such as service uptime and performance.

service-level indicator (SLI)

A measurement of a performance aspect of a service, such as its error rate, availability, or throughput.

service-level objective (SLO)

A target metric that represents the health of a service, as measured by a service-level indicator.

shared responsibility model

A model describing the responsibility you share with AWS for cloud security and compliance. AWS is responsible for security *of* the cloud, whereas you are responsible for security *in* the cloud. For more information, see Shared responsibility model.

SIEM

See security information and event management system.

single point of failure (SPOF)

A failure in a single, critical component of an application that can disrupt the system.

SLA

See service-level agreement.

SLI

See service-level indicator.

SLO

See service-level objective.

split-and-seed model

A pattern for scaling and accelerating modernization projects. As new features and product releases are defined, the core team splits up to create new product teams. This helps scale your organization's capabilities and services, improves developer productivity, and supports rapid

innovation. For more information, see Phased approach to modernizing applications in the AWS Cloud.

SPOF

See single point of failure.

star schema

A database organizational structure that uses one large fact table to store transactional or measured data and uses one or more smaller dimensional tables to store data attributes. This structure is designed for use in a data warehouse or for business intelligence purposes.

strangler fig pattern

An approach to modernizing monolithic systems by incrementally rewriting and replacing system functionality until the legacy system can be decommissioned. This pattern uses the analogy of a fig vine that grows into an established tree and eventually overcomes and replaces its host. The pattern was introduced by Martin Fowler as a way to manage risk when rewriting monolithic systems. For an example of how to apply this pattern, see Modernizing legacy Microsoft ASP.NET (ASMX) web services incrementally by using containers and Amazon API Gateway.

subnet

A range of IP addresses in your VPC. A subnet must reside in a single Availability Zone.

supervisory control and data acquisition (SCADA)

In manufacturing, a system that uses hardware and software to monitor physical assets and production operations.

symmetric encryption

An encryption algorithm that uses the same key to encrypt and decrypt the data.

synthetic testing

Testing a system in a way that simulates user interactions to detect potential issues or to monitor performance. You can use Amazon CloudWatch Synthetics to create these tests.

system prompt

A technique for providing context, instructions, or guidelines to an LLM to direct its behavior. System prompts help set context and establish rules for interactions with users.

T

tags

Key-value pairs that act as metadata for organizing your AWS resources. Tags can help you manage, identify, organize, search for, and filter resources. For more information, see Tagging your AWS resources.

target variable

The value that you are trying to predict in supervised ML. This is also referred to as an *outcome variable*. For example, in a manufacturing setting the target variable could be a product defect.

task list

A tool that is used to track progress through a runbook. A task list contains an overview of the runbook and a list of general tasks to be completed. For each general task, it includes the estimated amount of time required, the owner, and the progress.

test environment

See environment.

training

To provide data for your ML model to learn from. The training data must contain the correct answer. The learning algorithm finds patterns in the training data that map the input data attributes to the target (the answer that you want to predict). It outputs an ML model that captures these patterns. You can then use the ML model to make predictions on new data for which you don't know the target.

transit gateway

A network transit hub that you can use to interconnect your VPCs and on-premises networks. For more information, see What is a transit gateway in the AWS Transit Gateway documentation.

trunk-based workflow

An approach in which developers build and test features locally in a feature branch and then merge those changes into the main branch. The main branch is then built to the development, preproduction, and production environments, sequentially.

trusted access

Granting permissions to a service that you specify to perform tasks in your organization in AWS Organizations and in its accounts on your behalf. The trusted service creates a service-linked role in each account, when that role is needed, to perform management tasks for you. For more information, see Using AWS Organizations with other AWS services in the AWS Organizations documentation.

tuning

To change aspects of your training process to improve the ML model's accuracy. For example, you can train the ML model by generating a labeling set, adding labels, and then repeating these steps several times under different settings to optimize the model.

two-pizza team

A small DevOps team that you can feed with two pizzas. A two-pizza team size ensures the best possible opportunity for collaboration in software development.

U

uncertainty

A concept that refers to imprecise, incomplete, or unknown information that can undermine the reliability of predictive ML models. There are two types of uncertainty: *Epistemic uncertainty* is caused by limited, incomplete data, whereas *aleatoric uncertainty* is caused by the noise and randomness inherent in the data. For more information, see the Quantifying uncertainty in deep learning systems guide.

undifferentiated tasks

Also known as *heavy lifting*, work that is necessary to create and operate an application but that doesn't provide direct value to the end user or provide competitive advantage. Examples of undifferentiated tasks include procurement, maintenance, and capacity planning.

upper environments

See environment.

V

vacuuming

A database maintenance operation that involves cleaning up after incremental updates to reclaim storage and improve performance.

version control

Processes and tools that track changes, such as changes to source code in a repository.

VPC peering

A connection between two VPCs that allows you to route traffic by using private IP addresses. For more information, see What is VPC peering in the Amazon VPC documentation.

vulnerability

A software or hardware flaw that compromises the security of the system.

W

warm cache

A buffer cache that contains current, relevant data that is frequently accessed. The database instance can read from the buffer cache, which is faster than reading from the main memory or disk.

warm data

Data that is infrequently accessed. When querying this kind of data, moderately slow queries are typically acceptable.

window function

A SQL function that performs a calculation on a group of rows that relate in some way to the current record. Window functions are useful for processing tasks, such as calculating a moving average or accessing the value of rows based on the relative position of the current row.

workload

A collection of resources and code that delivers business value, such as a customer-facing application or backend process.

workstream

Functional groups in a migration project that are responsible for a specific set of tasks. Each workstream is independent but supports the other workstreams in the project. For example, the portfolio workstream is responsible for prioritizing applications, wave planning, and collecting migration metadata. The portfolio workstream delivers these assets to the migration workstream, which then migrates the servers and applications.

WORM

See write once, read many.

WQF

See AWS Workload Qualification Framework.

write once, read many (WORM)

A storage model that writes data a single time and prevents the data from being deleted or modified. Authorized users can read the data as many times as needed, but they cannot change it. This data storage infrastructure is considered immutable.

Z

zero-day exploit

An attack, typically malware, that takes advantage of a zero-day vulnerability.

zero-day vulnerability

An unmitigated flaw or vulnerability in a production system. Threat actors can use this type of vulnerability to attack the system. Developers frequently become aware of the vulnerability as a result of the attack.

zero-shot prompting

Providing an LLM with instructions for performing a task but no examples (*shots*) that can help guide it. The LLM must use its pre-trained knowledge to handle the task. The effectiveness of zero-shot prompting depends on the complexity of the task and the quality of the prompt. See also few-shot prompting.

zombie application

An application that has an average CPU and memory usage below 5 percent. In a migration project, it is common to retire these applications.