Managing identity and access for VMware Cloud on AWS

AWS Prescriptive Guidance

AWS Prescriptive Guidance: Managing identity and access for VMware Cloud on AWS

Table of Contents

Managing identity and access for VMware Cloud on AWS

Richard Milner-Watts, Abdenour Kansab, and Chris Porter, Amazon Web Services (AWS)

Vern Bolinius, VMware

June 2023 (document history)

Identity and access management is the principle of limiting systems access to only authorized users and applications, including restricting access to only the necessary network resources. In cloud environments, identity and access management controls typically consist of the policies and services that you use to identify, authenticate, and authorize users, groups of users, and applications.

VMware Cloud on AWS supports your VMware vSphere-based workloads in the AWS Cloud. You can use many VMware services and tools to configure, manage, back up, monitor, and analyze this cloud infrastructure. The features and controls you use to manage identity and access vary between services. This document provides best practices and recommendations for managing identity and access for the following VMware services:

- VMware Aria Operations
- VMware Aria Operations for Logs
- VMware Aria Operations for Networks
- VMware Cloud Disaster Recovery
- VMware Cloud on AWS
- VMware Cloud Services Console
- VMware HCX
- VMware NSX
- VMware Site Recovery
- VMware vCenter Server

This guide provides an overview and best practices of identity and access management for VMware Cloud on AWS and related VMware services. It includes a brief description of each service and discusses the identity access and management considerations for that service. We also provide recommendations for configuring the service as part of VMware Cloud on AWS.

> **Important**
>
> Many of the VMware services discussed in this guide are used in other cloud or on-premises VMware solutions. The recommendations and best practices in this guide are specific to VMware Cloud on AWS. These recommendations might not apply to other environments.

Intended audience

This guide is intended for architects and security engineers who are responsible for implementing VMware Cloud on AWS in their cloud or hybrid environment.

Targeted business outcomes

This guide helps you do the following:

1. Understand the various identity and access management controls for VMware Cloud on AWS and related VMware services
2. Become familiar with the recommended best practices that help you securely operate VMware Cloud on AWS
3. Understand the options that are available for federated authentication through an external identity provider

Identity management overview

VMware uses the following industry-standard concepts and identity hierarchy to manage identification, authentication, and authorization:

- *Users* are the individuals who access your environment in some capacity. You can create local users, or you can use federation to authenticate users from an external identity provider. For more information, see Identity federation and SSO.
- *Groups* provide a mechanism to logically group a collection of users together. This helps you grant consistent permissions to those users and reduces administrative overhead. Roles are used to grant permissions to a user or a group. For more information, see Roles and Permissions in the SDDC (VMware documentation).
- *Organizations* in VMware Cloud control access to one or more VMware services. Users and groups must belong to an organization in order to access the services in the organization. You can enable the Identity Governance and Administration feature to allow federated identities to self-service request membership to a VMware organization. For more information, see VMware Cloud Services Console.

Permissions can grant access to a specific object, or they can be inherited from parent objects. If multiple overlapping permissions are assigned to a user or group, the most permissive permission applies. For more information, see Hierarchical Inheritance of Permissions (VMware documentation).

You can use these structural elements to adopt a least-privilege policy and establish logical access boundaries within your infrastructure based on user requirements. *Least-privilege* is the principle of granting users and applications only the minimum access necessary to perform their tasks. In the event of unauthorized access, this industry best practice can help limit an attacker's ability to cause damage or steal sensitive data. And even for authorized users, this principle can prevent users accessing data they shouldn't have. Giving users access to only the necessary resources can also improve productivity and reduce the need for troubleshooting support.

When using VMware Cloud on AWS, there are two primary services and tools for managing identity and access: VMware Cloud Services Console and VMware vCenter Server. Later in this guide, we discuss these services in greater detail.

Identity federation and SSO

Many companies want to set up federation with an external identity provider (IdP). This allows you to provide a single sign-on (SSO) experience to your users. Both VMware Cloud and vCenter Server support enterprise federation:

- VMware Cloud supports Security Assertion Markup Language (SAML) 2.0 based IdPs and supports Lightweight Directory Access Protocol (LDAP). For more information, see What is enterprise federation and how does it work with VMware Cloud Services (VMware documentation).
- When you operate vCenter Server on VMware Cloud on AWS, federation to vCenter Server by using an external IdP isn't currently supported. Only the built-in IdP can be used, which supports using Microsoft Active Directory through LDAP. For more information, see Identity Sources for vCenter Server with vCenter Single Sign-On (VMware documentation).

Some of the other related VMware services discussed in this guide also support direct federation from an IdP. However, configuring federation in every service creates additional points of user management and becomes difficult to manage. Instead, you can use groups and roles in VMware Cloud Services Console to use a common identity source and configure permissions for other VMware Cloud services. Also, you can configure Hybrid Linked Mode in order to use the same identities with an on-premises vCenter Server instance. This reduces the number of points of federation and identity management down to two services. For more information about Hybrid Linked Mode, see Configuring Hybrid Linked Mode (VMware documentation).

General best practices

> **Important**
>
> Many of the VMware services discussed in this guide are used in other cloud or on-premises VMware solutions. The recommendations and best practices in this guide are specific to VMware Cloud on AWS. These recommendations might not apply to other environments.

Consider the following AWS recommendations for managing identity and access to your VMware cloud infrastructure:

- Apply a policy of least privilege. Use role-based access control (RBAC) to grant the minimum permissions and access required for users to perform their function.
- When possible, grant permissions to groups rather than to individual users.
- Avoid configuring local users. Authenticate users against an external, federated identity provider.
- Configure multi-factor authentication for all users.
- Your password policy should include password strength and rotation requirements.
- Document a break-glass procedure to take full administrative control over the VMware organization and related services. Break glass, which draws its name from breaking the glass to pull a fire alarm, refers to a means for a person to quickly obtain administrative access in exceptional circumstances, by using an approved and audited process.
- If you have on-premises data centers or multiple vCenter Server instances, use Hybrid Linked Mode to connect your cloud vCenter Server instance with on-premises vCenter Single Sign-On domain. This helps you manage your cloud and on-premises resources from a single vSphere Client interface.
- When possible, configure management endpoints, such as vCenter Server, HCX Cloud Manager, and NSX Manager, to be accessible from only internal networks, rather than from the public internet.
- Do not use local credentials, such as the **cloudadmin** account, for administrative purposes. Reserve these accounts for use in your break-glass procedure. Actions performed using administrative local user accounts can't be attributed to a specific individual, so these accounts could be used to make changes without accountability.

- Change the passwords for local accounts, such as root and administrative users, to strong values and securely store these credentials in an audited password store. Establish an approval process for granting access to these passwords.
- If local credentials will persist for long periods, such as for multiple months or longer, establish a process for rotating the credentials (for example, if you're using VMware HCX to stretch a network).

These recommendations apply to all of the VMware service configurations for VMware Cloud on AWS. Additional recommendations for each service are covered later in this guide.

VMware identity management services

When using VMware Cloud on AWS, there are two primary services and tools for managing identity and access: VMware Cloud Services Console and VMware vCenter Server.

VMware Cloud Services Console

VMware Cloud Services Console (VMware documentation) helps you manage your VMware Cloud services portfolio, which includes VMware Cloud on AWS. In this service, you can:

- Manage entities, such as users and groups
- Manage organizations, which control access to other cloud services, such as VMware Cloud Disaster Recovery (VCDR) and the VMware Aria Suite
- Assign roles to resources and services
- View the OAuth applications that have access to your organization
- Configure enterprise federation for the organization
- Enable and deploy VMware Cloud services, such as VMware Aria and VMware Cloud on AWS
- Manage billing and subscriptions
- Get VMware support

Managing identity and access

By properly setting up users, groups, roles, and organizations in VMware Cloud Services Console, you can implement a least-privilege access policy.

Securing access to the VMware Cloud Services Console is critical because administrative users of this service can change permissions throughout your VMware cloud environment and access sensitive information, such as billing information. To access all console features, such as billing and support, users must also be linked with a VMware Customer Connect profile (formally known as *MyVMware*).

In VMware Cloud Services Console, you use the following types of roles to grant permissions to users and groups:

- **Organization roles** – These roles pertain to the VMware Cloud organization directly, granting permissions within the VMware Cloud Services Console. There are two standard roles. The

Organization owner role has full permissions to administer the organization. The **Organization member** role has read access to the VMware Cloud Services Console. For more information, see What organization roles are available in VMware Cloud Services (VMware documentation).

- **Service roles** – These roles allow you to assign permissions to use a specific service. For example, an entity with the **DR Admin** service role can administer VMware Cloud Disaster Recovery (VCDR) in the dedicated service console. Every service available within the organization has one or more associated service roles. For more information about the available service roles, refer to the VMware documentation for the service of interest.

The VMware Cloud Services Console supports authentication policies. These can stipulate that a user must provide a second authentication token when logging in, also known as *multi-factor authentication* (MFA).

For more information about managing identity and access in this service, see Identity and Access Management (VMware documentation).

AWS recommendations

In addition to the General best practices, AWS recommends the following when configuring VMware Cloud Services Console for VMware Cloud on AWS:

- When creating an organization, use a VMware Customer Connect profile and associated corporate email address that does not belong to an individual, such as **vmwarecloudroot@example.com**. This account should be treated as a service, or root, account, and you should audit usage and restrict access to the email account. Immediately configure account federation with your corporate identity provider (IdP) so that users can access the organization without using this account. Reserve this account for use in a break-glass procedure for addressing issues with the federated IdP.
- Use federated identities for the organization to grant access to other cloud services, such as VMware Cloud Disaster Recovery (VCDR). Do not individually manage users or federation in multiple services. This simplifies managing access to multiple services, such as when users join or leave the company.
- Assign the **Organization owner** role sparingly. Entities with this role can grant themselves full access to all aspects of the organization and any associated cloud services.

VMware vCenter Server

VMware vCenter Server (VMware website) is a management plane for administering VMware vSphere environments. In vCenter Server, you manage the entities that can access vSphere resources, such as virtual machines, and access add-ons, such as VMware HCX and VMware Site Recovery. You manage vCenter Server through the vSphere Client application. In vCenter Server, you can:

- Manage virtual machines, VMware ESXi hosts, and VMware vSAN storage
- Configure and manage vCenter Single Sign-On

If you have on-premises data centers, you can use Hybrid Linked Mode to link your cloud vCenter Server instance to an on-premises vCenter Single Sign-On domain. If the vCenter Single Sign-On domain contains multiple vCenter Server instances that are connected using Enhanced Linked Mode, all of those instances are linked to your cloud SDDC. By using this mode, you can view and manage your on-premises and cloud data centers from a single vSphere Client interface, and you can migrate workloads between your on-premises data center and cloud SDDC. For more information, see Configuring Hybrid Linked Mode (VMware documentation).

Managing identity and access

In software-defined data centers (SDDCs) (VMware website) for VMware Cloud on AWS, the way in which you operate vCenter Server is similar to an on-premises SDDC. The primary difference is that VMware Cloud on AWS is a managed service. Therefore, VMware is responsible for certain administrative tasks, such as managing hosts, clusters, and management virtual machines. For more information, see What's Different in the Cloud? and Global permissions (VMware documentation).

Because VMware performs some administrative tasks for the SDDC, a cloud administrator requires fewer privileges than an administrator of an on-premises data center. When you create a VMware Cloud on AWS SDDC, a **cloudadmin** user is automatically created and assigned the CloudAdmin role (VMware documentation). You can use this privileged, local user account to access vCenter Server and vCenter Single Sign-On. Users who have the VMware Cloud on AWS **Administrator** or **Administrator (Delete Restricted)** service role in VMware Cloud Services Console can obtain the credentials for the **cloudadmin** user. The **CloudAdmin role** has the maximum possible permissions in vCenter Server for a VMware Cloud on AWS SDDC. For more information about this service role, see CloudAdmin Privileges (VMware documentation). The **cloudadmin** user is the only local

user available for vCenter Server in VMware Cloud on AWS. To grant access for other users, use an external identity source.

vCenter Single Sign-On is an authentication broker that provides security token exchange infrastructure. When a user authenticates to vCenter Single Sign-On, that user receives a token that can be used to authenticate with vCenter Server and other add-on services by using API calls. The **cloudadmin** user can configure an external identity source for vCenter Server. For more information, see Identity Sources for vCenter Server with vCenter Single Sign-On (VMware documentation).

In vCenter Server, you use the following three types of roles to grant permissions to users and groups:

- **System roles** – You can't edit or delete these roles.
- **Sample roles** – These roles represent frequently performed combinations of tasks. You can copy, edit, or delete these roles.
- **Custom roles** – If the system and sample roles don't provide the access control you want, you can create custom roles in the vSphere Client. You can duplicate and modify an existing role, or you can create a new role. For more information, see Create a vCenter Server Custom Role (VMware documentation).

For each object in the SDDC inventory, you can assign only one role to a user or group. If, for a single object, a user or group requires a combination of built-in roles, there are two options. The first option is to create a custom role with the required permissions. The other option is to create two groups, assign a built-in role to each, and then add the user to both groups.

AWS recommendations

In addition to the General best practices, AWS recommends the following when configuring vCenter Server for VMware Cloud on AWS:

- Use the **cloudadmin** user account to configure an external identity source in vCenter Single Sign-On. Assign appropriate users from the external identity source to be used for administrative purposes, and then discontinue use of the **cloudadmin** user. For best practices when configuring vCenter Single Sign-On, see Information Security and Access for vCenter Server (VMware documentation).
- In vSphere Client, update the **cloudadmin** credentials for each vCenter Server instance to a new value, and then store them securely. This change isn't reflected in the VMware Cloud Services

Console. For example, viewing the credentials through the Cloud Services Console shows the original value.

> **Note**
>
> If the credentials for this account are lost, VMware support can reset them.

- Do not use the **cloudadmin** account for day-to-day access. Reserve this account for use as part of a break-glass procedure.
- Restrict network access to vCenter Server to only private networks.

Related VMware services

This chapter provides best practices and recommendations for managing identity and access for the following VMware services related to VMware Cloud on AWS:

- Services managed through VMware Cloud Services Console:
  - VMware Cloud on AWS
  - VMware NSX
  - VMware Aria Operations for Logs
  - VMware Aria Operations for Networks
  - VMware Aria Operations
  - VMware Cloud Disaster Recovery
- Services managed through VMware vCenter Server:
  - VMware HCX
  - VMware Site Recovery

This guide provides a brief description of each service, discusses the identity access and management controls for that service, and includes AWS recommendations for configuring that service as part of VMware Cloud on AWS.

VMware Cloud on AWS

VMware Cloud on AWS (VMware documentation) is a service jointly designed by AWS and VMware to help you migrate and extend your on-premises VMware vSphere-based environments to the AWS Cloud.

You can access VMware Cloud on AWS through the VMware Cloud Services Console, if you belong to an organization that grants access to this service. In VMware Cloud on AWS, you can:

- Create and delete SDDCs.
- Administer SDDC groups.
- Administer SDDCs, including networking and cluster parameters.
- Access the **cloudadmin** user credentials for VMware vCenter Server. For more information about this user, see VMware vCenter Server in this guide.

- Access the **cloud_admin** user credentials for VMware NSX. For more information about this user, see VMware NSX in this guide.
- Enable and deploy add-on services within SDDCs, such as VMware Site Recovery and VMware HCX.
- Access consoles for add-on services, including HCX and VMware Site Recovery.

Managing identity and access

You use VMware Cloud Services Console to manage identities and access to VMware Cloud on AWS. For VMware Cloud on AWS, the following service roles are available:

- **Administrator** – This role has full access to VMware Cloud on AWS.
- **Administrator (Delete Restricted)** – This role has full access to VMware Cloud on AWS, excluding SDDC delete operations.
- **NSX Cloud Admin**
- **NSX Cloud Auditor**

> **Note**
>
> **NSX Cloud Admin** and **NSX Cloud Auditor** are related to the use of VMware NSX. For more information, see VMware NSX.

One of the two **Administrator** roles is required to access an SDDC within the Cloud Services Portal. Users without one of the two NSX Cloud roles cannot access the SDDC **Networking and Security** tab within the Cloud Services Portal, additionally they cannot access NSX admin credentials.

AWS recommendations

In addition to the General best practices, AWS recommends the following when configuring VMware Cloud on AWS:

- To grant assess to administrators, use only the **Administrator (Delete Restricted)** role. Reserve the **Administrator** role for break-glass access when you need to delete an SDDC.
- Do not grant the NSX roles to users who do not need access to networking and firewall configurations. For more information, see VMware NSX in this guide.

- Change the passwords for the **cloudadmin** local user account to a strong value and securely store these credentials in an audited password store. You can change this password in VMware vCenter Server by using the vSphere Web Client.

VMware NSX

VMware NSX (VMware documentation) provides a network virtualization layer that reproduces the Open Systems Interconnection (OSI) model from layer 2 through layer 7; with features including switching, routing, and firewalls. There are two versions of NSX. The original version (NSX-V) requires that you also deploy vCenter Server. The newer version (NSX-T) is decoupled from vCenter Server, which enables support for hybrid architectures. VMware Cloud on AWS uses NSX-T.

NSX, along with vSphere and vSAN, is a core component of VMware Cloud on AWS. NSX provides all of the networking functionality within an SDDC and manages the interaction between the overlay networking and the AWS native components that form the network underlay. NSX is tightly coupled with other services, such as vCenter Server and VMware HCX, which call NSX APIs to manage resources.

In NSX, you can:

- Manage switching and routing
- Manage firewalls, including using a distributed firewall for inline inspection between VMs or between the network and the public internet
- Manage virtual private networks (VPNs)
- Configure Dynamic Host Configuration Protocol (DHCP) and Domain Name System (DNS)

You can access NSX from the VMware Cloud Services Console or through the dedicated NSX Manager web user interface (UI). NSX Manager web UI offers some additional features that are not available in the VMware Cloud Services Console. For more information, see SDDC Network Administration with NSX Manager (VMware documentation).

Note the following when accessing NSX in VMware Cloud on AWS:

- To access NSX through the VMware Cloud Services Console, you must be assigned the VMware Cloud on AWS **Administrator** role. You can access NSX on the SDDC **Networking and Security** tab. For more information about this role, see VMware Cloud on AWS in this guide.

- You can open the NSX Manager web UI by choosing the link on the SDDC **Settings** tab or by choosing **Open NSX Manager** on the SDDC **Summary** page. For more information, see Open NSX Manager (VMware documentation).
- If the SDDC is in Payment Card Industry Data Security Standard (PCI DSS) mode, you cannot access NSX through the **Networking and Security** tab in the VMware Cloud Services Console. You must use the NSX Manager web UI.

Managing identity and access

You use VMware Cloud Services Console to manage identities and access to VMware NSX. For NSX in VMware Cloud on AWS, the following service roles are available:

- **NSX Cloud Admin** – This role can administer VMware NSX functionality with VMware Cloud on AWS.
- **NSX Cloud Auditor** – This role can view NSX service settings and events but cannot make any changes.

> **Note**
>
> Despite their names, these roles are not related to the VMware NSX Cloud service.

The following users can access NSX:

- The **cloud_admin** local user, which is a built-in and highly privileged local NSX user. Users who have the **NSX Cloud Admin** role can access the credentials for this user account. Despite their similar names, the **cloud_admin** user is distinct from the **cloudadmin@vmc.local** vCenter Single Sign-On local user.
- Users who have been assigned either the **NSX Cloud Admin** service role or the **NSX Cloud Auditor** service role in the VMware Cloud Services Console. These users could be VMware Cloud Services Console users or externally federated users.
- Users who have been directly granted access to NSX from an identity source through LDAP.

AWS recommendations

In addition to the General best practices, AWS recommends the following when configuring NSX for VMware Cloud on AWS:

- If your company has users who are responsible for managing networking and firewalls but are not responsible for managing SDDCs, grant these users one of the NSX roles, but do not grant them the **Administrator** role. These users should access NSX through the NSX Manager web UI.
- Change the passwords for the **cloud_admin** local user account to a strong value and securely store these credentials in an audited password store. To change this password, you must contact VMware support.
- Avoid granting access to external users directly within NSX. Instead, set up enterprise federation in the VMware Cloud Services Console, and then use roles and groups to grant access to this service.

VMware Aria Operations for Logs

VMware Aria Operations for Logs (VMware documentation), formerly *VMware vRealize Log Insight Cloud*, is a log storage and analysis tool that helps you visualize and query the log data produced by your VMware SDDCs. In VMware Aria Operations for Logs, you can:

- Integrate with on-premises instances of vRealize Operations
- Collect and analyze all types of machine-generated log data
- Configure alerts
- Monitor and analyze logs from other VMware services

There are two versions of this centralized log management service. VMware vRealize Log Insight is an on-premises version that can run as an appliance within your SDDC. VMware Aria Operations for Logs is a software-as-a-service (SaaS) version. VMware Cloud on AWS uses the cloud version as the default logging service, and this can't be changed. If you use the on-premises version, you need to forward logs from the cloud instance to your on-premises instance.

VMware Aria Operations for Logs is included with VMware Cloud on AWS. The included version has a limited ingestion capacity and storage retention period. If needed, you can upgrade to a premium subscription to increase these limits. For more information, see Subscriptions and Billing (VMware documentation).

Managing identity and access

You use VMware Cloud Services Console to manage identities and access to VMware Aria Operations for Logs. VMware Aria Operations for Logs uses the same users, including federated identities, and groups that you configured in VMware Cloud Services Console. To grant permissions for this service, you can assign a service role or configure a custom role within VMware Aria Operations for Logs. For more information, see Service Roles (VMware documentation).

VMware vRealize Log Insight has two default roles. The **Administrator** role has full access and control, and the **User** role has read access and can create dashboards. You can use custom roles to grant access to only specific data sets. These data sets contain filters that restrict which log data is available to the user. For more information, see Create a Data Set (VMware documentation).

AWS recommendations

Adhere to the General best practices described previously in this guide. We don't have additional recommendations for managing identity and access in this service.

VMware Aria Operations for Networks

VMware Aria Operations for Networks, formerly *VMware vRealize Network Insight Cloud*, is a SaaS version of vRealize Network Insight. VMware vRealize Network Insight (VMware documentation) helps you understand the traffic flows for your workloads. You can use this service to diagnose networking issues and model firewall rules to support workload segmentation. In VMware Aria Operations for Networks, you can:

- View your hybrid and multi-cloud environments
- Troubleshoot and analyze and traffic flows
- Discovery and analyze applications
- Map dependencies between workloads

There are three versions of this service. *VMware vRealize Network Insight* is an on-premises-only version. *VMware Aria Operations for Networks* is a SaaS version. *vRealize Network Insight Universal* can be deployed as an on-premises solution or as a federated cloud SaaS solution. All versions are compatible with VMware Cloud on AWS.

Managing identity and access

You use VMware Cloud Services Console to manage identities and access to VMware Aria Operations for Networks. VMware Aria Operations for Networks uses the same users, including federated identities, and groups that you configured in VMware Cloud Services Console. For VMware Aria Operations for Networks, the following service roles are available:

- **Administrator** – This role has full access and control.
- **Member** – This role has limited access.
- **Auditor** – This role has read-only access.

AWS recommendations

Adhere to the General best practices described previously in this guide. We don't have additional recommendations for managing identity and access in this service.

VMware Aria Operations

VMware Aria Operations (VMware documentation), formerly *VMware vRealize Operations Cloud*, is an operations management platform for VMware Cloud on AWS. This service uses artificial intelligence and machine learning (AI/ML) to help you optimize, plan, and scale the applications and infrastructure in your hybrid cloud deployments. In VMware Aria Operations, you can:

- View AI/ML-powered optimization recommendations for performance and capacity
- Management of compliance and resource configurations
- Access tools to help you troubleshoot issues, like resolving customer problems or responding to alerts
- Use management packs (VMware documentation) to expand the monitoring, troubleshooting, and remediation features of this service

There are two versions of this operations management service. *VMware vRealize Operations* is an on-premises version that can run as an appliance within your SDDC. *VMware Aria Operations* is a software-as-a-service (SaaS) version of vRealize Operations. Both versions are compatible with VMware Cloud on AWS. Because VMware Cloud on AWS is a managed service and access to some resources is restricted, not all vRealize Operations features are supported. For more information, see Known Limitations (VMware documentation).

Managing identity and access

You use VMware Cloud Services Console to manage identities and access to VMware Aria Operations. VMware Aria Operations uses the same users, including federated identities, that you configure in VMware Cloud Services Console. To grant permissions for this service, you can assign a service role or configure a custom role within VMware Aria Operations. For more information about the available service roles, see Roles and Privileges (VMware documentation).

There are three built-in roles: **Administrator**, **GeneralUser** and **ReadOnly**, and if needed, you can create custom roles to match specific permissions requirements. You can create groups to minimize the administrative overhead of managing permissions for multiple users.

The on-premises version of VMware vRealize Operations supports local users, and both the cloud and on-premises versions support federated users. However, federation of users to an external identity provider varies between the on-premises and cloud versions of vRealize Operations. For the on-premises version, you can directly federate users from an external IdP through LDAP, or you can use the identities that you federated in vCenter Server. For the cloud version, you use the same users, including federated users, that you configure in VMware Cloud Services Console.

AWS recommendations

In addition to the General best practices, AWS recommends the following when configuring VMware Aria Operations for VMware Cloud on AWS:

- Avoid federating users directly. For the cloud version, federate users in VMware Cloud Services Console, and then use roles and groups to grant access to this service. For the on-premises version of this service, use identities from an authenticated source or enable single sign-on (SSO). For more information, see Authentication sources and Configure a Single Sign-On Source (VMware documentation).

VMware Cloud Disaster Recovery

VMware Cloud Disaster Recovery (VCDR) (VMware documentation) is a disaster recovery as a service (DRaaS) solution that delivers a tiered approach to disaster recovery. You can adjust the costs and timescales for your recovery point objective (RPO) and recovery time objective (RTO) to meet the requirements for a given workload. This helps you balance reliable protection and the efficient use of disaster recovery resources. In VCDR, you can:

- Create backups of virtual machines
- Store backups in durable cloud storage
- Choose between flexible deployment options for restoration targets, from on-demand to hot-standby
- Configure custom RPOs and RTOs

Managing identity and access

You use VMware Cloud Services Console to manage identities and access to VMware Cloud Disaster Recovery. VMware Cloud Disaster Recovery uses the same users, including federated identities, and groups that you configured in VMware Cloud Services Console. To grant permissions for this service, you can assign a VCDR service role or create a custom role within VMware Cloud Disaster Recovery. For more information about the available service roles, see VMware Cloud Disaster Recovery Service Roles (VMware documentation).

VCDR includes several built-in roles that you can use to operate the service:

- **Administrator** – Full control, excluding access to API tokens.
- **Auditor** – Read-only access to the user interface, excluding user management. Access to compliance reports.
- **DR Admin** – Create, test, and run disaster recovery plans.
- **Backup Admin** – Manage protected sites and protection groups. Access to restore VMs.
- **Plan Tester** – Create disaster recovery plans, run test recoveries.
- **SDDC Admin** – Manage SDDCs.

AWS recommendations

Adhere to the General best practices described previously in this guide. We don't have additional recommendations for managing identity and access in this service.

VMware HCX

VMware HCX (VMware documentation) is an application mobility platform that enables workload migrations between SDDCs. VMware HCX is included with VMware Cloud on AWS and can be used to migrate workloads. In VMware HCX, you can:

- Configure multi-site meshes between SDDCs
- Extend networks between HCX sites
- Migrate virtual machines

Managing identity and access

You use VMware vCenter Server to manage identities and access to VMware HCX. VMware HCX requires access to other VMware services to create and manage resources and migrations, including access to vCenter Server and NSX. VMware HCX has two component services:

- **HCX Cloud Manager** – In the VMware Cloud Services Console, you enable VMware HCX for the SDDC. This installs the HCX Cloud Manager appliance within the selected SDDC. For more information, see Deploying the HCX Installer OVA in the vSphere Client (VMware documentation). After deployment, you can use the vCenter Server **cloudadmin** credentials to access the HCX Cloud Manager service.
- **HCX Connector** – You can obtain the HCX Connector Open Virtualization Archive (OVA) file through the HCX Cloud Manager service. You use this file to install an HCX Cloud Manager appliance on any vCenter Server instance, which sets up that instance as a migration source in VMware HCX. Each HCX Connector instance has its own admin and root credentials.

After you have deployed both component services, you can access VMware HCX through vCenter Server. The **Administrators** vCenter Single Sign-On group is automatically granted the **HCX Administrator** role. Installing HCX adds a lot of additional roles and privileges to vCenter Single Sign-On. Use these to create fine-grained access controls for VMware HCX, based on the different types of users.

AWS recommendations

In addition to the General best practices, AWS recommends the following when configuring VMware HCX for VMware Cloud on AWS:

- Use Gateway Firewall rules to restrict network access to the HCX Cloud Manager service.
- Securely store the on-premises HCX Connector **admin** and **root user** credentials. Consider rotating these credentials in accordance with your company policies. VMware manages these credentials on your behalf for HCX Cloud Manager.

- For an on-premises HCX Connector instance, consider creating custom HCX roles that match the needs of your different types of HCX users. For example, create a more permissive role for users who set up and administer HCX, and create a less permissive role for users who manage only migrations.
- When pairing VMware HCX with VMware Cloud on AWS, you must use the **cloudadmin** user. For more information, see the *Resolution* section of HCX - Site Pairing Connectivity Diagnostics (VMware Knowledge Base article 78340).
- When pairing HCX Cloud with VMware Cloud on AWS, authentication is not supported between the VMware Cloud on AWS SDDC and Active Directory. For more information, see [VMC on AWS] AD unsupported for HCX Cloud to Cloud setup (VMware Knowledge Base article 90433).

VMware Site Recovery

VMware Site Recovery (VMware documentation) is an on-demand, disaster recovery as a service (DRaaS) solution that is based on the VMware Site Recovery Manager service for on-premises environments. In VMware Site Recovery, you can:

- Implement replication, orchestration, and automation to help protect workloads in the event of a site failure
- Create an end-to-end disaster recovery solution to help protect SDDCs

Managing identity and access

You use VMware vCenter Server to manage identities and access to VMware Site Recovery. VMware Site Recovery performs operations on behalf of users, such as replicating or powering off a virtual machine. Site Recovery uses roles and privileges to help ensure that only users with the correct permissions can perform recovery operations, such as running all the steps in a recovery plan.

For Site Recovery, the following service roles are available:

- **SrmAdministrator** – This role can perform all Site Recovery configuration and administration operations.
- **HmsCloudAdmin** – This role can list servers, but it can't add or remove them.

When you set up Site Recovery in VMware Cloud on AWS, the following user group updates are automatically configured:

1. A new **SRM Administrators** group is created and assigned the **SrmAdministrator** role.
2. A new **HmsCloudAdministrators** group is created and assigned the **HmsCloudAdmin** role.
3. The **CloudAdminGroup** group is added to both the **SRM Administrators** group and the **HmsCloudAdministrators** group. This provides the **CloudAdminGroup** group transitive permissions to manage Site Recovery Manager and vSphere replication.

For more information, see Learn more about permission configuration for VMware Site Recovery (VMware documentation).

If you use federated identities to access vCenter Server, you must use Hybrid Linked Mode to add entities to these groups. For more information, see Configuring Hybrid Linked Mode (VMware documentation).

AWS recommendations

In addition to the General best practices, AWS recommends the following when configuring Site Recovery for VMware Cloud on AWS:

- Make sure users are assigned the same roles on both the source and target sites. This ensures that protected and recovered objects have identical permissions.
- Use Hybrid Linked Mode to manage Site Recovery role assignments for federated identities within vCenter Server.
- Site Recovery uses private IP addresses only within the SDDC. In keeping with General best practices, ensure that your VMware Cloud on AWS vCenter resolves to a private IP address.

Sample groups and roles

The following table provides an example of an identity and access management strategy for using VMware Cloud on AWS. It outlines the user persona, the VMware services that persona needs to access, the organization and group membership, the roles assigned, and the type of identity used (such as local users or federated identities). Using this table as a starting point, design a strategy for your company that adheres to the best practices recommended in this guide.

User persona	Services accessed	VMware Cloud sample group name	VMware Cloud service roles	vCenter Single Sign-On sample group name	vCenter Single Sign-On role	Identity source
Organization break glass	VMware Cloud Services Console	None	**Organization owner**	None	None	Local user (service account email address)
VMware administrator	VMware Cloud Services Console vCenter Server HCX Site Recovery VCDR vRealize Operations	`vmware_admins`	**Organization owner**	`vmware_admins`	**Administrator**	Federated identity provider

User persona	Services accessed	VMware Cloud sample group name	VMware Cloud service roles	vCenter Single Sign-On sample group name	vCenter Single Sign-On role	Identity source
Backup administrator	vCenter Server	None	None	`vmware_backups`	**Power user**	Federated identity provider
Disaster recovery administrator	vCenter Server VMware Cloud Services Console Site Recovery VCDR	`vmware_dr`	**Organization member** **DR Admin** **DR SDDC Admin**	`vmware_dr`	**SrmAdministrator** **HmsCloudAdmin**	Federated identity provider
VMware operator	VMware Cloud Services Console vCenter Server HCX vRealize Operations	`vmware_ops`	**Organization member** **vROps Administrator**	`vmware_ops`	**Power user**	Federated identity provider

User persona	Services accessed	VMware Cloud sample group name	VMware Cloud service roles	vCenter Single Sign-On sample group name	vCenter Single Sign-On role	Identity source
Networking team	VMware Cloud Services Console vCenter Server	`vmware_networks`	**Organization member** **NSX Cloud Admin**	`vmware_networks`	**Readonly**	Federated identity provider
Security team	VMware Cloud Services Console vCenter Server HCX (temporary access) Site Recovery VCDR vRealize Operations	`vmware_security`	**Organization member** **vROps ReadOnly**	`vmware_security`	**Readonly**	Federated identity provider

User persona	Services accessed	VMware Cloud sample group name	VMware Cloud service roles	vCenter Single Sign-On sample group name	vCenter Single Sign-On role	Identity source
Auditors	VMware Cloud Services Console vCenter Server	`vmware_audit`	**Organization member**	`vmware_audit`	**Readonly**	Federated identity provider

Next steps

This guide covered the best practices that we recommend for managing identity and access for VMware Cloud on AWS and related VMware services. These recommendations are designed to help you secure your cloud and hybrid cloud infrastructure and prevent unauthorized access, but they are also designed to be scalable and efficient. By assigning users to groups and then assigning roles to groups, you can more quickly grant or restrict permissions and minimize the overhead associated with configuring users individually. Also, by using federation to an external identity provider and vCenter Single Sign-On, you can provide a seamless, single sign-on experience to your users.

Use the Sample groups and roles table to start designing an identity and access management strategy that works for your company. After you have reviewed the recommendations in this guide, we suggest that you review the links provided in the Resources section. These resources will help you learn more about VMware Cloud services and how to configure the best practices described in this guide.

Resources

Related AWS resources

- VMware Cloud on AWS overview and operating model
- Disaster recovery options for workloads on VMware Cloud on AWS
- Configuring storage offload options for VMware Cloud on AWS
- Deploy a VMware SDDC on AWS by using VMware Cloud on AWS
- Migrate VMware SDDC to VMware Cloud on AWS using VMware HCX

VMware documentation

VMware Cloud on AWS

- Setting Up Enterprise Federation for Cloud Services
- VMware Cloud Services Identity and Access Management

VMware vCenter Server and vCenter Single Sign-On

- Understanding Authorization in vSphere
- vSphere Administration in VMware Cloud on AWS
- vSphere Authentication with vCenter Single Sign-On
- Configuring vCenter Single Sign-On Identity Sources
- Hierarchical Inheritance of Permissions
- Information Security and Access for vCenter Server
- vSphere Required Privileges for Common Tasks

VMware NSX

- NSX Administration Guide
- Information Security and Access for NSX-T Data Center

VMware HCX

- VMware HCX User Guide
- VMware HCX User Account and Role Requirements

VMware Aria and vRealize suite

- VMware vRealize Operations Documentation
- Roles and Privileges in vRealize Operations Cloud
- VMware vRealize Log Insight Datasheet
- Getting Started with VMware Aria Operations for Logs
- VMware Cloud Services Guide
- vRealize Network Insight User Management

VMware Site Recovery

- VMware Site Recovery Documentation
- Site Recovery Manager Privileges, Roles and Permissions
- Permission configuration for VMware Site Recovery at VMware Cloud on AWS

VMware Cloud Disaster Recovery

- VMware Cloud Disaster Recovery User Roles

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
VMware HCX access	We updated the AWS recommendations for configuring VMware HCX for VMware Cloud on AWS.	June 5, 2023
Initial publication	—	November 3, 2022

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. This migration scenario is specific to VMware Cloud on AWS, which supports virtual machine (VM) compatibility and workload portability between your on-premises environment and AWS. You can use the VMware Cloud Foundation technologies from your on-premises data centers when you migrate your infrastructure to VMware Cloud on AWS. Example: Relocate the hypervisor hosting your Oracle database to VMware Cloud on 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 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

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.

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.

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 AWS CodeCommit. 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

A field of AI used by machines to identify people, places, and things in images with accuracy at or above human levels. Often built with deep learning models, it automates extraction, analysis, classification, and understanding of useful information from a single image or a sequence of images.

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.

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 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.

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.

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.

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.

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.

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.

G

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.

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, 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.

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.

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 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.

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.

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*.

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.

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.

MPA

See Migration Portfolio Assessment.

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.

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.

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.

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.

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.

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. For more information, see Enabling data persistence in microservices.

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

An approach in system engineering that takes privacy into account throughout the whole engineering 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*.

production environment

See environment.

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.

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).

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 Managing AWS Regions in *AWS General Reference*.

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.

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.

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.

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 Secret in the Secrets Manager documentation.

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.

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.

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.

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.