SAP Guides

General SAP Guides

General SAP Guides: SAP Guides

Table of Contents

General SAP Guides

This section covers the following guides.

- Overview
- Amazon EC2 instance types
- Estimation
- AWS Data Provider
- Architecture guidance
- Disaster recovery with AWS Elastic Disaster Recovery
- RISE with SAP on AWS Cloud

Additional SAP on AWS documentation

- SAP HANA on AWS
- SAP NetWeaver on AWS
- Databases for SAP applications on AWS
- AWS Launch Wizard for SAP
- AWS Systems Manager for SAP
- AWS SDK for SAP ABAP
- SAP BusinessObjects on AWS
- AWS Migration Hub Orchestrator

SAP on AWS Overview and Planning

SAP specialists, Amazon Web Services

Last updated: January 2023

This guide provides overview and planning information for SAP customers and partners who are considering implementing or migrating SAP environments or systems to the Amazon Web Services Cloud.

This guide is intended for users who have previous experience installing, migrating, and operating SAP environments and systems on traditional on-premises infrastructure. It consists of three main sections:

- An overview of the AWS Cloud and AWS services, for readers who are new to the cloud.
- An overview of SAP on AWS, including software and licenses, support options, and partner services.
- Technical considerations that will help you plan and get the most out of your SAP environment on AWS.

> **Note**
>
> To access the SAP notes referenced in this guide, you must have an SAP One Support Launchpad user account. For more information, see the SAP Support website.

About this Guide

This guide is part of a content series that provides detailed information about hosting, configuring, and using SAP technologies in the AWS Cloud. For the other guides in the series, ranging from overviews to advanced topics, see https://aws.amazon.com/sap/docs/.

AWS Overview

AWS offers a broad set of global, cloud-based services, including compute, storage, networking, Internet of Things (IoT), and many others. These services help organizations move faster, lower IT costs, and support scalability. AWS is trusted by the largest enterprises and popular start-ups to

power a wide variety of workloads, such as web and mobile applications, game development, data processing and warehousing, storage, and archiving.

AWS Services

AWS provides over 200 cloud services that you can use in combinations tailored to your business or organizational needs. For information about all AWS services, see the Amazon Web Services Cloud Platform documentation.

This section introduces the AWS services that are most relevant for the deployment and operation of SAP solutions. The following list provides a high-level description of each service and its use for SAP systems. To view features, pricing, and documentation for an individual service, follow the *details*link after the description.

Area	Service	Description	SAP uses
Compute	Amazon Elastic Compute Cloud (Amazon EC2)	Secure, resizable compute capacity in the cloud. (details)	Virtual and bare metal servers for the installation and operation of SAP systems.
Storage	Amazon Elastic Block Store (Amazon EBS)	Persistent block storage volumes for use with EC2 instances. (details)	File systems for SAP software (e.g., `/usr/sap`), SAP database log and data files, and SAP local backups.
	Amazon Simple Storage Service (Amazon S3)	Object storage service that offers an extremely durable, highly available, and infinitely scalable data storage infrastru cture. (details)	Highly durable storage for file backups, database backups, archiving data, data lakes, and more.
	Amazon Elastic File System (Amazon EFS)	Simple, scalable, elastic file system	Shared file system for SAP applicati

Area	Service	Description	SAP uses
		for Linux-based workloads for use with AWS Cloud services and on-premises resources. (details)	on servers (e.g., /sapmnt).
	Amazon FSx for Windows File Server	Fully managed, highly durable, and available native Microsoft Windows file system. (details)	Shared file system for SAP applicati on servers (e.g., /sapmnt).
	Amazon FSx for NetApp ONTAP	Fully managed, highly reliable, scalable, high-performing file storage built on NetApp ONTAP file system(details)	Shared file system for SAP applicati on servers (e.g., /sapmnt).
Networking	Amazon Virtual Compute Cloud (Amazon VPC)	Logically isolated section of the AWS Cloud where you can launch AWS resources in a virtual network that you define. (details)	Network for SAP resources. You can control the level of isolation of your EC2 instance from other networks, instances, and on-premises network resources, such as those in production and non-production environments.

Area	Service	Description	SAP uses
	Amazon Site-to-Site VPN	Enables you to securely connect your on-premises network or branch office site to your VPC. (details)	Network connectivity between on-premises systems/users and SAP systems on AWS.
	AWS Directs Connect	Lets you establish private network connectivity between AWS and your data center, office, or co-location environment. (details)	Private network connectivity between on-premises systems/users and the SAP system or environment on AWS.
	Amazon Route 53	Highly available and scalable cloud Domain Name System (DNS) web service. (details)	Name and address resolution for SAP systems running on AWS.
	Amazon Time Sync	Highly accurate and reliable time reference that is natively accessible from EC2 instances. (Linux | Windows)	Time synchronization for your SAP systems on EC2 instances.
Management and operation tools	AWS Management Console	Simple web interface to provision and manage AWS resources. (details)	Provisioning and management of AWS resources for your SAP environment on AWS.

Area	Service	Description	SAP uses
	AWS Command Line Interface (AWS CLI)	Command-line tool set to provision and manage AWS resources. (details)	Creation of scripts to automate the provisioning and management of AWS resources for your SAP environment on AWS.
	AWS CloudFormation	An easy way to create a collection of related AWS resources and provision them in an orderly and predictable fashion. (details)	Automated provisioning of AWS resources for new SAP landscapes, disaster recovery environments, and other use cases.
	Amazon CloudWatch	Monitoring for AWS Cloud resources and the applications you run on AWS: collect and track metrics, collect and monitor log files, and set alarms. (details)	Monitoring SAP systems running on AWS using Amazon CloudWatch Application Insights.
	AWS CloudTrail	Records activity made on your account and delivers log files to your S3 bucket. (details)	Audit capabilities within your AWS account, such as use of the Amazon EC2 API.

Area	Service	Description	SAP uses
	AWS Launch Wizard for SAP	AWS Launch Wizard for SAP is a service that guides you through the sizing, configuration, and deployment of SAP applications on AWS. (details)	Setup and configuration of resources required for your SAP deployment.
	AWS Backint Agent for SAP HANA	SAP certified solution to backup and restore SAP HANA database to and from Amazon S3. (details)	Backup solution to store SAP HANA database backups to Amazon S3.
Security, identity, and compliance	AWS Identity and Access Management (AWS IAM)	Manages access to AWS services and resources securely. Using AWS IAM, you can create and manage AWS users and groups, and use permissions to allow and deny their access to AWS resources. (details)	Fine-grained access control using a least privileged security model to access specific AWS services and actions; e.g., to allow SAP BASIS resources to launch, to stop and start EC2 instances without terminating them.

AWS Global Infrastructure

The AWS Cloud infrastructure is built around Regions and Availability Zones. An AWS Region is a physical location that provides multiple, physically separated and isolated Availability Zones. Each Availability Zone consists of one or more data centers that are connected with low-latency, high-throughput, and highly redundant networking. These Availability Zones offer an easier and more

effective way to design and operate your applications and databases, making them more highly available, fault tolerant, and scalable than traditional single or multiple data center infrastructures.

For a list of the available AWS Regions and to learn more about the AWS global infrastructure, see Global Infrastructure on the AWS website.

AWS Security and Compliance

Security

At AWS, security is our top priority. As an AWS customer, you will benefit from a data center and network architecture built to meet the requirements of the most security-sensitive organizations. Security in the cloud is much like security in your on-premises data centers—only without the costs of maintaining facilities and hardware. In the cloud, you don't have to manage physical servers or storage devices. Instead, you use software-based security tools to monitor and protect the flow of information into and out of your cloud resources.

As an AWS customer you inherit all the best practices of AWS policies, architecture, and operational processes built to satisfy the requirements of our most security-sensitive customers, and get the flexibility and agility you need in security controls.

The AWS Cloud enables a shared responsibility model. While AWS manages security **of** the cloud, you are responsible for security **in** the cloud. This means that you retain control of the security you choose to implement to protect your own data, platform, applications, systems, and networks no differently than you would in an on-site data center.

To learn more about AWS security, see AWS Cloud Security on the AWS website.

Compliance

AWS provides robust controls to help maintain security and data protection in the cloud. As systems are built on top of AWS Cloud infrastructure, compliance responsibilities will be shared. By tying together governance-focused, audit-friendly service features with applicable compliance or audit standards, AWS compliance enablers build on traditional programs and help you operate in an AWS security control environment.

The IT infrastructure that AWS provides to its customers is designed and managed in alignment with best security practices and a variety of IT security standards. The following is a partial list of assurance programs with which AWS complies:

- SOC 1/ISAE 3402, SOC 2, SOC 3
- FISMA, FIPS, DIACAP, and FedRAMP
- PCI DSS Level 1
- ISO 9001, ISO 27001, ISO 27017, ISO 27701, ISO 27018

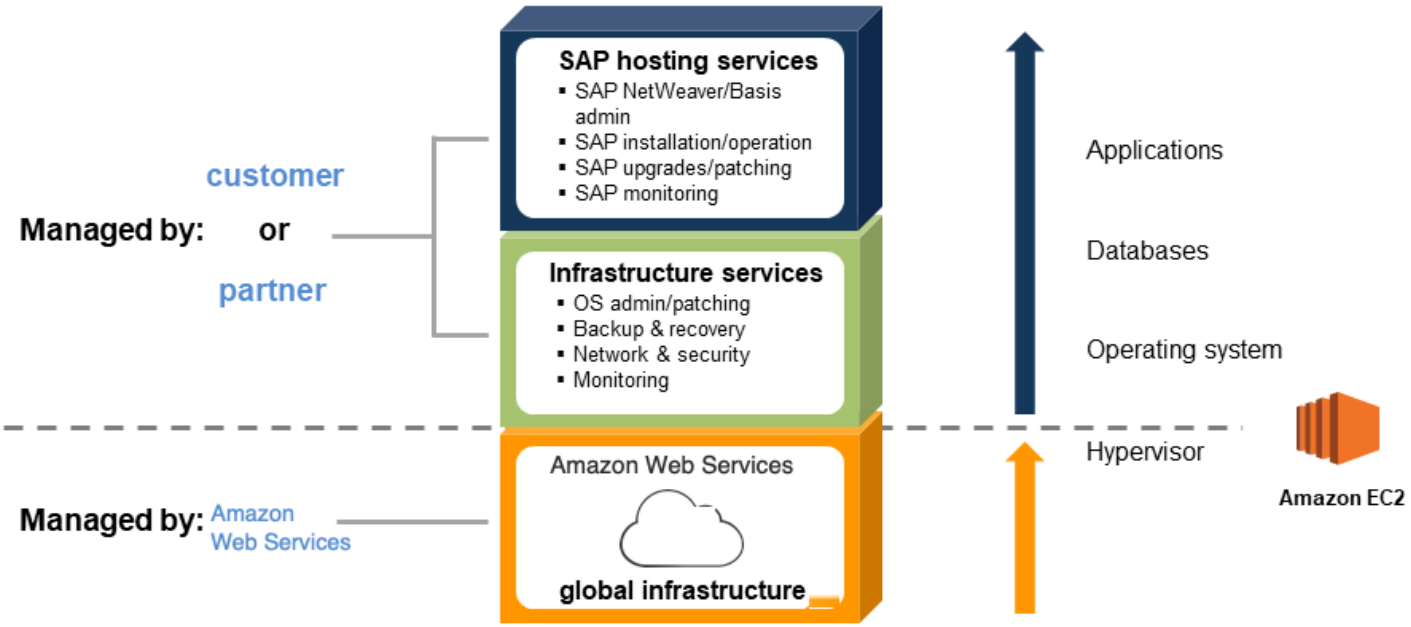
For more information, see AWS Compliance Programs.

AWS Provisioning and Management

The provisioning and management of AWS services and resources use a self-service model managed by the customer or a partner. For an overview of the tools available for provisioning and management, see the management tools in the AWS Services section.

Figure 1 shows the services managed by AWS and the services managed by the customer or partner for SAP.

Figure 1: Managed services for SAP on AWS



SAP on AWS Overview

AWS has been working with SAP since 2011 to help customers deploy and migrate their SAP applications to AWS, and SAP supports running the vast majority of available SAP applications on AWS.

SAP Software and Licenses on AWS

This section describes the options available for SAP software and licenses on AWS.

Bring Your Own Software and License

The majority of SAP solutions that can be run on AWS use a bring-your-own-software and bring-your-own-license (BYOL) model. Running SAP systems on AWS doesn't require special or new SAP licenses. If you're an existing SAP customer, you can use your existing SAP licenses when running SAP on AWS. You are responsible for obtaining a valid SAP license, and you must ensure that you are in compliance with the SAP licensing policies. AWS does not provide or sell SAP licenses.

AWS Marketplace

AWS Marketplace is a digital catalog with thousands of software listings from independent software vendors that makes it easy to find, test, buy, and deploy software that runs on AWS. To view SAP-related offerings available in AWS Marketplace, follow this link: SAP in AWS Marketplace.

SAP Trial and Developer Licenses

The SAP Cloud Appliance Library provides access to an online repository of the latest preconfigured SAP solutions. You can quickly deploy these solutions on AWS by using a launch wizard that automates deployment. Some of the solutions available in the SAP Cloud Appliance Library are provided with free trial or developer edition licenses.

SAP Hardware Key Generation

SAP hardware key generation on EC2 instances uses a specific process that is dependent on the SAP kernel patch level. If a hardware key is generated before patching the SAP kernel to the proper level, and the kernel is updated at a later time, the hardware key may change, making the installed license invalid. For details on how the SAP hardware ID is generated on EC2 instances and the required SAP kernel patch levels see the following SAP notes (SAP One Support Launchpad access required):

- SAP Note 2327159 – SAP NetWeaver License Behavior in Virtual and CLoud Environments
- SAP Note 1178686 – Linux: Alternative method to generate a SAP hardware key
- SAP Note 2327159 – SAP NW License Behavior in Virtual and Cloud Environments
- SAP Note 1697114 – Determination of hardware ID in Amazon clouds

- SAP Note 2113263 – Additional public key for AWS Hardware ID
- SAP Note 2823805 – Additional public keys for AWS Hardware ID
- SAP Note 2319387 – Adjustment of the license check for AWS China

SAP Support on AWS

AWS and SAP have worked together closely to ensure that you receive the same level of support via the same support channels, whether you're running your SAP systems on AWS or on premises.

SAP Solutions Supported on AWS

The majority of SAP solutions that run on traditional on-premises infrastructure are fully supported by SAP on AWS. For the complete list of SAP solutions supported on AWS, see SAP Note 1656099 and the other notes referenced within that note.

SAP Support on AWS

To ensure full support of your SAP on AWS environment from SAP and AWS, you must follow the guidelines and requirements in SAP Note 1656250. Here are the primary requirements you must follow to ensure support of your SAP on AWS environment:

- Enable detailed monitoring for **Amazon CloudWatch** on each EC2 instance to ensure that the required AWS metrics are provided in one-minute intervals. For additional information on Amazon CloudWatch, see https://aws.amazon.com/cloudwatch.
- Install, configure, and run the AWS Data Provider for SAP on each EC2 instance. The AWS Data Provider collects the required performance and configuration data from a variety of sources, including the Amazon EC2 API, Amazon EC2 instance metadata, and Amazon CloudWatch, and shares it with SAP applications, to help monitor and improve the performance of business transations.
- Any AWS account that you use for running SAP systems must have an AWS support plan for either Business Support or Enterprise Support.

Deploying SAP Systems on AWS

The section describes different options available for provisioning AWS infrastructure and installing SAP systems on AWS.

Manual Deployment

The majority of SAP solutions supported on AWS can be installed by manually provisioning the required AWS infrastructure resources and then following the relevant SAP installation document on AWS.

Automated Deployment

AWS Launch Wizard for SAP is a service that guides you through the sizing, configuration, and deployment of SAP applications on AWS. AWS Launch Wizard reduces the time it takes to deploy SAP applications on AWS. You input your application requirements, including SAP HANA settings, SAP landscape settings, and deployment details on the service console, and AWS Launch Wizard identifies the appropriate AWS resources to deploy and run your SAP application.

For more information, see How AWS Launch Wizard for SAP works.

Prebuilt Images

Some SAP solutions are available on AWS as a prebuilt system image that contains a preinstalled and preconfigured SAP system. A prebuilt SAP system image enables you to rapidly provision a new SAP system without spending the time and effort required by a traditional manual SAP installation.

Prebuilt SAP system images are available from the following sources:

- AWS Marketplace
- SAP Cloud Appliance Library

SAP solution	Deployment option(s)
SAP Business Suite (ERP, CRM, etc.)	Manual | SAP CAL
SAP NetWeaver	Manual | AWS Launch Wizard for SAP | SAP CAL
SAP S/4HANA	Manual | AWS Launch Wizard for SAP | SAP CAL
SAP BW/4HANA	Manual | SAP CAL

SAP solution	Deployment option(s)
SAP HANA	Manual | AWS Launch Wizard for SAP | SAP CAL
SAP BusinessObjects BI	Manual | AWS Marketplace | SAP CAL
SAP Commerce (Hybris)	Manual
SAP Business One, version for SAP HANA	Manual | SAP CAL
SAP Business One, version for Microsoft SQL Server	Manual

Getting Assistance from APN Partners

There are AWS Partner Networks (APN) partners who are experienced in deploying and operating SAP solutions, and can help you with your SAP workloads on AWS. For additional information see the following section.

Partner Services for SAP on AWS

The AWS Partner Network (APN) is a community of companies that offer a wide range of services and products on AWS. APN SAP partners can provide SAP-specific services to help you fully maximize the benefits of running SAP solutions on AWS.

Types of Partner Services and Solutions for SAP on AWS

- **Cloud assessment services** – Advisory services to help you develop an efficient and effective plan for your cloud adoption journey. Typical services include financial/TCO (total cost of ownership), technical, security and compliance, and licensing.
- **Proof-of-concept services** – Services to help you test SAP on AWS; for example: SAP ERP/ECC migration to SAP HANA or SAP S/4HANA, SAP Business Warehouse (BW) migration to SAP HANA or SAP BW/4HANA, SAP OS/DB migrations, new SAP solution implementation.
- **Migration services** – Services to migrate existing SAP environments or systems to AWS; for example: all-on-AWS SAP migrations (PRD/QAS/DEV), hybrid SAP migrations (QAS/DEV), single SAP system (e.g., SAP BW) migrations.

- **Managed services** – Managed services for SAP environments on AWS, including: AWS account and resource administration, OS administration/patching, backup and recovery, SAP Basis and SAP NetWeaver.
- **Packaged solutions** – Bundled software and service offerings from SAP Partners that combine SAP software, licenses, implementation, and managed services on AWS, such as SAP S/4HANA, SAP BusinessObjects BI, and many others.
- **ISV software solutions** – Partner software solutions for the migration, integration, and operation of SAP solutions on AWS; for example: system migration, high availability, backup and recovery, data replication, automatic scaling, disaster recovery.

How to Find Partner Solutions for SAP on AWS

The **AWS SAP Partner Solutions** provides a centralized place to search, discover, and connect with trusted APN partners who offer solutions and services to help your business achieve faster time to value and maximize the benefits of running SAP solutions on AWS. For more information, see AWS SAP Competency Partners.

SAP on AWS Planning

If you are an experienced SAP Basis or SAP NetWeaver administrator, there are a number of AWS-specific considerations relating to compute configurations, storage, security, management, and monitoring that will help you get the most out of your SAP environment on AWS. This section provides guidelines for achieving optimal performance, availability, and reliability, and lower total cost of ownership (TCO) while running SAP solutions on AWS.

SAP Notes

Before migrating or implementing an SAP environment on AWS, you should read and follow the relevant SAP notes. Start from SAP Note 1656099 for general information and follow the links to other relevant SAP notes (SAP One Support Launchpad access required).

SAP on AWS Architectures

This section describes the two primary architectural patterns for SAP on AWS: all systems on AWS and hybrid.

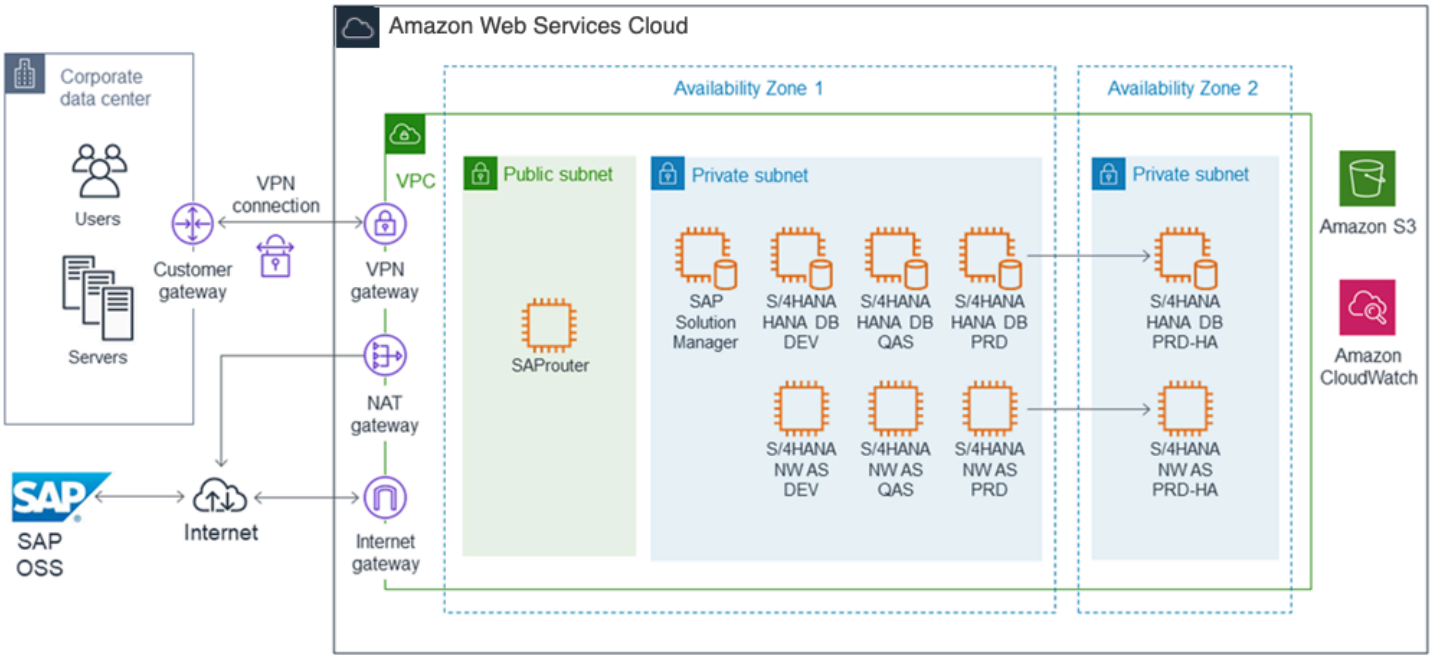
All-on-AWS Architecture

With the SAP All-on-AWS architecture, all systems and components of your SAP environment are hosted on AWS. Example scenarios of such an architecture include:

- Implementation of a complete, new SAP environment on AWS
- Migration of a complete, existing SAP environment to AWS

Figure 3 depicts an SAP all-on-AWS architecture. The SAP environment running on AWS is integrated with on-premises systems and users via a VPN connection or a dedicated network connection via AWS Directs Connect. SAProuter is deployed in a public subnet and assigned a public IP address that is reachable from the internet to enable integration with the SAP OSS network via a secure network communications (SNC) connection. A network address translation (NAT) gateway enables instances in the private subnet to connect to the internet or other AWS services, but prevents instances from receiving inbound traffic that is initiated by someone on the internet. For additional information, see the Configuring Network and Connectivity section.

Figure 3: SAP all-on-AWS architecture



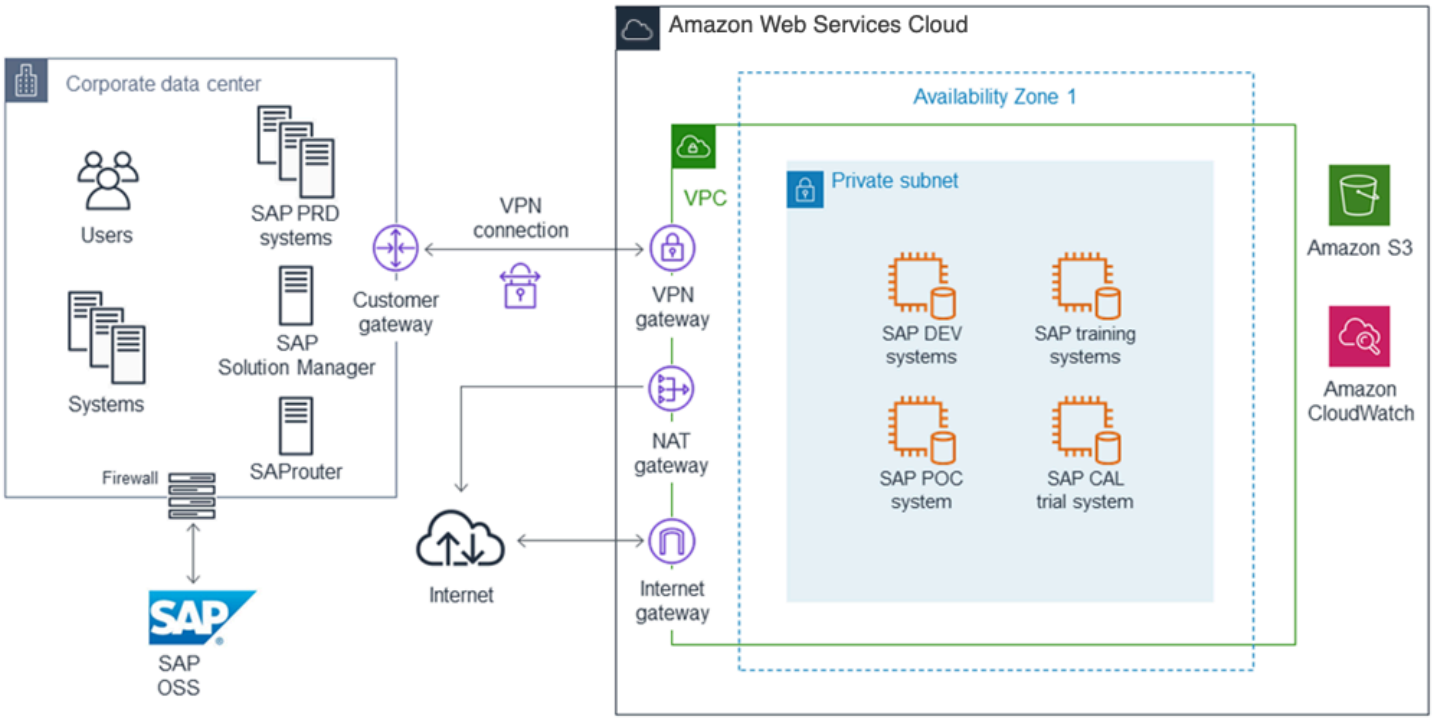
Hybrid AWS Architecture

With an SAP hybrid AWS architecture, some SAP systems and components are hosted on your on-premises infrastructure and others are hosted on the AWS infrastructure. Example scenarios of such an architecture include:

- Running SAP test, trial, training, proof-of-concept (PoC), and similar systems on AWS
- Running non-production SAP landscapes (for example, DEV and QAS) on AWS, integrated with an SAP production landscape running on premises
- Implementing a new SAP application on AWS and integrating it with an existing SAP on-premises environment

Figure 4 depicts an SAP hybrid AWS architecture with SAP DEV and QAS landscapes and SAP test, training, and PoC systems running on AWS. These systems are integrated with SAP systems and users on the corporate network. Connectivity between the VPC and the corporate network is provided with either a VPN connection or an AWS Directs Connect connection. The existing SAProuter and SAP Solution Manager running on the corporate network are used to manage the SAP systems running within the VPC.

Figure 4: SAP hybrid AWS architecture



Choosing an AWS Region and Availability Zone

See the AWS Global Infrastructure section of this guide for information about AWS Regions and Availability Zones.

Choosing a Region

When choosing the AWS Region to deploy your SAP environment in, consider the following factors:

- Proximity to your on-premises data center(s), systems, and end users to minimize network latency.
- Data residency and regulatory compliance requirements.
- Availability of the AWS products and services you plan to use in the region. For a detailed list of AWS products and services by region, see the Region Table on the AWS website.
- Availability of the EC2 instance types you plan to use in the region. To view AWS Region availability for a specific instance type, see the Amazon EC2 Instance Types for SAP webpage.

Choosing an Availability Zone

No special considerations are required when choosing an Availability Zone for your SAP deployment on AWS. All SAP applications (SAP ERP, CRM, SRM, and so on) and systems (SAP database system, SAP Central Services system, and SAP application servers) should be deployed in the same Availability Zone. If high availability (HA) is a requirement, use multiple Availability Zones. For more information, see Architecture guidance for availability and reliability of SAP on AWS.

Network and Connectivity

Amazon VPC

Amazon VPC enables you to define a virtual network in your own, logically isolated area within the AWS Cloud. You can launch your AWS resources, such as instances, into your VPC. Your VPC closely resembles a traditional network that you might operate in your own data center, with the benefits of using the AWS scalable infrastructure. You can configure your VPC; you can select its IP address range, create subnets, and configure route tables, network gateways, and security settings. You can connect instances in your VPC to the internet. You can connect your VPC to your own corporate data center, and make the AWS Cloud an extension of your data center. To protect the resources in each subnet, you can use multiple layers of security, including security groups and network access control lists. For more information, see the Amazon VPC User Guide.

For detailed instructions for setting up and configuring a VPC, and connectivity between your network and VPC, see the Amazon VPC documentation.

Network Connectivity Options

Multiple options are available to provide network connectivity between your on-premises users and systems with your SAP systems running on AWS, including a direct internet connection, hardware VPN, and private network connection.

Private Network Connection

AWS Directs Connect makes it easy to establish a dedicated network connection from your premises to AWS. Using AWS Directs Connect, you can establish private connectivity between AWS and your data center, office, or co-location environment. In many cases, this can reduce your network costs, increase bandwidth throughput, and provide a more consistent network experience than internet-based connections. For additional information, see the AWS Directs Connect User Guide.

Use cases: Recommended for customers who require greater bandwidth and lower latency than possible with a hardware VPN.

For more information, see Amazon Virtual Compute Cloud Connectivity Options.

Direct Internet Connection

The quickest and simplest way to connect to your SAP systems running on AWS involves using a VPC with a single public subnet and an internet gateway to enable communication over the internet. For additional information, see Scenario 1: VPC with a Public Subnet Only in the *Amazon VPC User Guide*.

Use cases: Most suitable for SAP demo, training, and test type systems that do not contain sensitive data.

Site-to-Site / Hardware VPN

AWS Site-to-Site VPN extends your data center or branch office to the cloud via Internet Protocol security (IPsec) tunnels, and supports connecting to both virtual private gateways and AWS Transit Gateway. You can optionally run Border Gateway Protocol (BGP) over the IPsec tunnel for a highly available solution. For additional information, see Adding a Hardware Virtual Private Gateway to your VPC in the *Amazon VPC User Guide*.

Use cases: Recommended for any SAP environments on AWS that require integration with on-premises users and systems.

Client VPN

AWS Client VPN provides a fully-managed VPN solution that can be accessed from anywhere with an internet connection and an OpenVPN-compatible client. It is elastic, automatically scales to meet your demand, and enables your users to connect to both AWS and on-premises networks. AWS Client VPN seamlessly integrates with your existing AWS infrastructure, including Amazon VPC and AWS Directory Service, so you don't have to change your network topology.

Use cases: Provides quick and easy connectivity to your remote workforce and business partners.

Following Security Best Practices

In order to provide end-to-end security and end-to-end privacy, AWS builds services in accordance with security best practices, provides appropriate security features in those services, and documents how to use those features. In addition, AWS customers must use those features and best practices to architect an appropriately secure application environment. Enabling customers to ensure the confidentiality, integrity, and availability of their data is of the utmost importance to AWS, as is maintaining trust and confidence.

Shared Responsibility Environment

There is a shared responsibility model between you as the customer and AWS. AWS operates, manages, and controls the components from the host operating system and virtualization layer down to the physical security of the facilities in which the services operate. In turn, you assume responsibility and management of the guest operating system (including updates and security patches), other associated application software, Amazon VPC setup and configuration, as well as the configuration of the AWS-provided security group firewall. For additional information on AWS security, visit the AWS Cloud Security page and review the various Security Resources available there.

Amazon VPC

The foundation for security of an SAP environment on AWS is the use of Amazon VPC for providing the overall isolation. Amazon VPC includes security details that you must set up to enable proper access and restrictions for your resources. Amazon VPC provides features that you can use to help increase and monitor the security for your VPC:

- **Security groups** act as a firewall for associated EC2 instances, controlling both inbound and outbound traffic at the instance level.
- **Network access control lists (ACLs)** act as a firewall for associated subnets, controlling both inbound and outbound traffic at the subnet level.
- **Route tables** consist of a set of rules, called routes, that determine where network traffic is directed. Each subnet in your VPC must be associated with a route table; the table controls the routing for the subnet.
- **Flow logs** capture information about the IP traffic going to and from network interfaces in your VPC.

For detailed documentation about how to set up and manage security within a VPC, see the Security section of the *Amazon VPC User Guide*.

EC2 Instance Types for SAP

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instance types comprise varying combinations of CPU, memory, storage, and networking capacity and give you the flexibility to choose the appropriate mix of resources for your applications. Each instance type includes one or more instance sizes, allowing you to scale your resources to the requirements of your target workload.

SAP systems deployed on AWS that will require support from SAP must be run on an EC2 instance type that has been certified with SAP. This section describes where you can find details about the EC2 instance types that have been certified with SAP and additional information for specific SAP solutions.

SAP NetWeaver-based Solutions

SAP solutions based on the SAP NetWeaver platform and that use SAP Application Performance Standard (SAPS) for sizing must be run on a specific subset of EC2 instance types in order to receive support from SAP Support. For details, see:

- SAP Note 1656099
- Amazon EC2 Types for SAP

SAP HANA

The SAP HANA platform and SAP solutions that run on top of an SAP HANA database—for example, SAP Suite on HANA, SAP S/4HANA, SAP Business Warehouse (BW) on HANA, SAP BW/4HANA-- require specific EC2 instance types that have been certified for SAP HANA. For more information, see Amazon EC2 instance types for SAP on AWS.

SAP Business One, version for SAP HANA

For information about the EC2 instance types that are certified for SAP Business One, version for SAP HANA, see:

- SAP Note 2058870
- SAP Business One on AWS

Operating Systems

Supported Operating Systems

EC2 instances run on 64-bit virtual processors based on the Intel x86 instruction set. The following 64-bit operating systems and versions are available and supported for SAP solutions on AWS.

- SUSE Linux Enterprise Server (SLES)
- SUSE Linux Enterprise Server for SAP Applications (SLES for SAP)
- Red Hat Enterprise Linux (RHEL)
- Red Hat Enterprise Linux for SAP Solutions (RHEL for SAP)
- Microsoft Windows Server
- Oracle Enterprise Linux

For additional information regarding SAP-supported operating systems on AWS, see SAP Note 1656250.

SLES for SAP and RHEL for SAP

SUSE and Red Hat offer SAP-specific versions of their operating systems that provide the following benefits:

- Configuration and tuning for SAP
- Extended release support
- High availability extension for SAP
- Dedicated support channel

> **Note**
>
> Because of these benefits, we strongly recommend using SLES for SAP or RHEL for SAP with High Availability (HA) and Update Services (US) for your SAP on AWS deployments.

To learn more about SUSE's and Red Hat's operating system versions for SAP, see the following information on the SLES and Red Hat websites.

SLES for SAP

- General information
- SUSE on AWS for SAP Applications

RHEL for SAP

- General information
- Red Hat in the Public Cloud
- Red Hat Cloud Access
- How to Locate Red Hat Cloud Access Gold Images on Amazon EC2
- What is the Difference between Red Hat Cloud Access and Red Hat Enterprise Linux On-Demand Subscriptions in the public cloud?

Operating System Licenses

These operating system licensing options are available for SAP systems on AWS:

- **On-demand** – The operating system software and license are bundled in an Amazon Machine Images (AMI). The fee for the operating system license is included in the On-Demand Instance hourly fee or Reserved Instance fee for the instance type.

- **Bring Your Own License/Subscription (BYOL)** – Bring your existing operating system license or subscription to the AWS Cloud.
- **AWS Marketplace** – Purchase operating system licenses and subscriptions from AWS Marketplace.

The following table lists the licensing options available for each operating system and version. To learn more about each option, follow the link in the table.

Operating system	License/subscription options
SLES	On-demand | BYOL
SLES for SAP	AWS Marketplace | BYOL
RHEL	On-demand | BYOL
RHEL for SAP with HA and US	AWS Marketplace | BYOL
Windows	On-demand | BYOL
Oracle Linux	BYOL

Databases

Supported Databases

All the database platforms and versions supported by SAP for an on-premises infrastructure are also supported by SAP on AWS. For details about the databases supported with specific SAP solutions on AWS, see SAP Note 1656099.

Database Installation and Administration

Customer-Managed Database on Amazon EC2

The majority of SAP solutions use a customer-managed model on Amazon EC2. Installation, configuration, administration, and backup and recovery of the database are done by either the customer or a partner.

The following SAP solutions use a self-managed database model on Amazon EC2:

- SAP Business Suite and SAP NetWeaver-based applications
- SAP HANA
- SAP S/4HANA
- SAP BW/4HANA
- SAP BusinessObjects BI
- SAP Business One

Amazon RDS

Amazon Relational Database Service (Amazon RDS) is a managed service that makes it easy to set up, operate, and scale a relational database in the cloud. It provides cost-efficient and resizable capacity, while managing time-consuming database administration tasks, freeing you up to focus on your applications and business. Amazon RDS is currently supported for the following SAP solutions:

- SAP BusinessObjects BI
- SAP Commerce (previously known as SAP Hybris Commerce)

Amazon Aurora

Amazon Aurora (Aurora) is a MySQL and PostgreSQL-compatible relational database built for the cloud. It combines the performance and availability of traditional enterprise databases with the simplicity and cost-effectiveness of open source databases. Aurora MySQL is currently supported for the following SAP solution:

- SAP Commerce (previously known as SAP Hybris Commerce)

Database Licenses

These database licensing options are available for SAP systems on AWS:

- **On-demand** – The database software and license are bundled in an Amazon Machine Image (AMI). The fee for the database license is included in the On-Demand Instance hourly fee or Reserved Instance fee for the instance type.
- **Bring Your Own License (BYOL)** – Bring your existing database licenses to the AWS Cloud.
- **AWS Marketplace** – Purchase database software and licenses from AWS Marketplace.

The following table lists the licensing options available on AWS for each database. For additional information, follow the links in the *Licensing options* column.

Database	Licensing options
SAP HANA	BYOL
SAP Adaptive Server Enterprise (ASE) (SAP ASE)	BYOL
Microsoft SQL Server	BYOL*
IBM DB2	BYOL
Oracle	BYOL
Amazon Aurora	On-demand

- SQL Server runtime licenses purchased from SAP require either Microsoft Software Assurance or Amazon EC2 Dedicated Hosts to bring these licenses to AWS. For additional information, see:
- SAP Note 2139358 - Effect of changes in licensing terms of SQL Server
- Microsoft Licensing on AWS

SAP Installation Media

The majority of SAP solutions on AWS use a bring-your-own-software model. There are two primary options for copying SAP installation media to AWS:

- **Download from the SAP Software Download Center to Amazon EC2.** From your EC2 instance, connect to the SAP Software Download Center and download the required installation media. This option will most likely be the fastest method for getting SAP installation media to AWS, because EC2 instances have very fast connections to the internet. You can create a dedicated Amazon EBS volume to store installation media, and then attach the volume to different instances as needed. You can also create a snapshot of the Amazon EBS volume and create multiple volumes that you can attach to multiple instances in parallel.

- **Copy from your network to Amazon EC2.** If you already have the required SAP installation media downloaded to a location on your network, you can copy the media from your network directly to an EC2 instance.

SAProuter and SAP Solution Manager

The following sections describe options for SAProuter and SAP Solution Manager when running SAP solutions on AWS.

For SAP All-on-AWS Architecture

When setting up an SAP environment on AWS, you will need to set up an SAP Solution Manager system and SAProuter with a connection to the SAP support network, as you would with any infrastructure. See the all-on-AWS architecture diagram (the section called “Figure 3: SAP all-on-AWS architecture”) for an illustration.

When setting up the SAProuter and SAP support network connection, follow these guidelines:

- Launch the instance that the SAProuter software is installed on into a public subnet of the VPC and assign it an Elastic IP address.
- Create a specific security group for the SAProuter instance with the necessary rules to allow the required inbound and outbound access to the SAP support network.
- Use the Secure Network Communications (SNC) type of internet connection. For more information, see https://service.sap.com/internetconnection.

For SAP Hybrid AWS Architecture

When using AWS as an extension of your IT infrastructure, you can use your existing SAP Solution Manager system and SAProuter that are running in your data center to manage SAP systems running on AWS within a VPC. See the hybrid architecture diagram (Figure 4) for additional information.

Document Revisions

Date	Change	Location
January, 2023	Update	Changes throughout guide

Date	Change	Location
May, 2019	Update	Changes throughout guide
August, 2018	Initial publication	–

Amazon EC2 instance types for SAP on AWS

Amazon Elastic Compute Cloud (Amazon EC2) offers a wide selection of instance types optimized to fit different use cases. The varying combinations of CPU, memory, storage, and networking capacity provide flexibility in selection of resources for your applications. You can choose the instance types that meet the requirements of your workload.

AWS has worked closely with SAP to test and certify Amazon EC2 instance types for SAP on AWS solutions. For more information, see SAP Note 1656099 - SAP Applications on AWS: Supported DB/OS and Amazon EC2 products (requires SAP Portal access) and SAP Certified and Supported SAP HANA Hardware Directory.

Topics

Instance type availability

The availability of an Amazon EC2 instance type is based on your selected Region. For more information about available instance types in your Region, see Amazon EC2 instance types by Region in the *Amazon EC2 Instance Types Guide*.

> **Note**
>
> Certain Amazon EC2 instance families, such as X1, X2idn, X2iedn, and High Memory might not be available across all Availability Zones in a Region. You must confirm while planning that the instance types required for your SAP workloads are available in your target Availability Zone.

You can also determine the availability of an instance type in a Region and its Availability Zone by using the describe-instance-type-offerings command. For examples, see Find an instance type using the AWS CLI in the *Amazon EC2 User Guide*.

SAP NetWeaver supported instances

Previous generation Amazon EC2 instances for SAP NetWeaver are fully supported and these instance types retain the same features and functionality. We recommend using current generation Amazon EC2 instances for new SAP NetWeaver implementations or migrations.

Topics

Current generation instances for SAP NetWeaver

Example

General Purpose

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
m5.large	2	8	2,817	High	Up to 4,750	18,750
m5.xlarge	4	16	5,535	High	Up to 4,750	18,750
m5.2xlarge	8	32	11,269	High	Up to 4,750	18,750
m5.4xlarge	16	64	22,538	High	4,750	18,750
m5.8xlarge	32	128	45,077	10	6,800	30,000
m5.12xlarge	48	192	67,615	10	9,500	40,000

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
m5.16xlarge	64	256	90,153	20	13,600	60,000
m5.24xlarge	96	384	135,230	25	19,000	80,000
m5.metal	96	384	142,000	25	19,000	80,000
m6a.large	2	8	3,023	Up to 12.5	Up to 10,000	40,000
m6a.xlarge	4	16	6,046	Up to 12.5	Up to 10,000	40,000
m6a.2xlarge	8	32	12,093	Up to 12.5	Up to 10,000	40,000
m6a.4xlarge	16	64	24,185	Up to 12.5	Up to 10,000	40,000
m6a.8xlarge	32	128	48,370	12.5	10,000	40,000
m6a.12xlarge	48	192	72,555	18.75	15,000	60,000
m6a.16xlarge	64	256	96,740	25	20,000	80,000
m6a.24xlarge	96	384	145,110	37.5	30,000	120,000
m6a.32xlarge	128	512	193,480	50	40,000	160,000
m6a.48xlarge	192	768	290,220	50	40,000	240,000

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
m6i.large	2	8	3,095	Up to 12.5	Up to 10,000	40,000
m6i.xlarge	4	16	6,190	Up to 12.5	Up to 10,000	40,000
m6i.2xlarge	8	32	12,380	Up to 12.5	Up to 10,000	40,000
m6i.4xlarge	16	64	24,760	Up to 12.5	Up to 10,000	40,000
m6i.8xlarge	32	128	49,520	12.5	10,000	40,000
m6i.12xlarge	48	192	74,280	18.75	15,000	60,000
m6i.16xlarge	64	256	99,040	25	20,600	80,000
m6i.24xlarge	96	384	148,560	37.5	30,000	120,000
m6i.32xlarge	128	512	198,080	50	40,000	160,000
m6id.large	2	8	3,095	Up to 12.5	Up to 10,000	40,000
m6id.xlarge	4	16	6,190	Up to 12.5	Up to 10,000	40,000
m6id.2xlarge	8	32	12,380	Up to 12.5	Up to 10,000	40,000

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
m6id.4xlarge	16	64	24,760	Up to 12.5	Up to 10,000	40,000
m6id.8xlarge	32	128	49,520	12.5	10,000	40,000
m6id.12xlarge	48	192	74,280	18.75	15,000	60,000
m6id.16xlarge	64	256	99,040	25	20,600	80,000
m6id.24xlarge	96	384	148,560	37.5	30,000	120,000
m6id.32xlarge	128	512	198,080	50	40,000	160,000
m6idn.large	2	8	3,095	Up to 25	Up to 25,000	100,000
m6idn.xlarge	4	16	6,190	Up to 30	Up to 25,000	100,000
m6idn.2xlarge	8	32	12,380	Up to 40	Up to 25,000	100,000
m6idn.4xlarge	16	64	24,760	Up to 50	Up to 25,000	100,000
m6idn.8xlarge	32	128	49,520	50	25,000	100,000
m6idn.12xlarge	48	192	74,280	75	37,500	150,000

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
m6idn.16xlarge	64	256	99,040	100	50,000	200,000
m6idn.24xlarge	96	384	148,560	150	75,000	300,000
m6idn.32xlarge	128	512	198,080	200	100,000	400,000
m6in.large	2	8	3,095	Up to 25	Up to 25,000	100,000
m6in.xlarge	4	16	6,190	Up to 30	Up to 25,000	100,000
m6in.2xlarge	8	32	12,380	Up to 40	Up to 25,000	100,000
m6in.4xlarge	16	64	24,760	Up to 50	Up to 25,000	100,000
m6in.8xlarge	32	128	49,520	50	25,000	100,000
m6in.12xlarge	48	192	74,280	75	37,500	150,000
m6in.16xlarge	64	256	99,040	100	50,000	200,000
m6in.24xlarge	96	384	148,560	150	75,000	300,000
m6in.32xlarge	128	512	198,080	200	100,000	400,000

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
m7a.medium	1	4	2,238	Up to 12.5	Up to 10,000	Up to 40,000
m7a.large	2	8	4,476	Up to 12.5	Up to 10,000	Up to 40,000
m7a.xlarge	4	16	8,952	Up to 12.5	Up to 10,000	Up to 40,000
m7a.2xlarge	8	32	17,904	Up to 12.5	Up to 10,000	Up to 40,000
m7a.4xlarge	16	64	35,808	Up to 12.5	Up to 10,000	Up to 40,000
m7a.8xlarge	32	128	71,616	12.5	10,000	40,000
m7a.12xlarge	48	192	107,424	18.75	15,000	60,000
m7a.16xlarge	64	256	143,232	25	20,000	80,000
m7a.24xlarge	96	384	214,848	37.5	30,000	120,000
m7a.32xlarge	128	512	286,464	50	40,000	160,000
m7a.48xlarge	192	768	429,720	50	40,000	240,000
m7i.large	2	8	4,218	Up to 12.5	Up to 10,000	Up to 40,000

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
m7i.xlarge	4	16	8,435	Up to 12.5	Up to 10,000	Up to 40,000
m7i.2xlarge	8	32	16,870	Up to 12.5	Up to 10,000	Up to 40,000
m7i.4xlarge	16	64	33,740	Up to 12.5	Up to 10,000	Up to 40,000
m7i.8xlarge	32	128	67,480	12.5	10,000	40,000
m7i.12xlarge	48	192	101,200	18.75	15,000	60,000
m7i.16xlarge	64	256	123,300	25	20,000	80,000
m7i.24xlarge	96	384	167,470	37.5	30,000	120,000
m7i.48xlarge*	192	768	306,202	50	40,000	240,000
m8i.large	2	8	4,278	Up to 12.5	Up to 10,000	Up to 40,000
m8i.xlarge	4	16	8,556	Up to 12.5	Up to 10,000	Up to 40,000
m8i.2xlarge	8	32	17,112	Up to 15	Up to 10,000	Up to 40,000
m8i.4xlarge	16	64	34,224	Up to 15	Up to 10,000	Up to 40,000

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
m8i.8xlarge	32	128	68,448	15	10,000	40,000
m8i.12xlarge	48	192	102,672	22.5	15,000	60,000
m8i.16xlarge	64	256	136,896	30	20,000	80,000
m8i.24xlarge	96	384	205,334	40	30,000	120,000
m8i.32xlarge	128	512	273,792	50	40,000	160,000
m8i.48xlarge	192	768	410,730	75	60,000	240,000

Compute Optimized

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
c5.large	2	4	2,650	Up to 10	Up to 4,750	20,000
c5.xlarge	4	8	5,300	Up to 10	Up to 4,750	20,000
c5.2xlarge	8	16	10,600	Up to 10	Up to 4,750	20,000
c5.4xlarge	16	32	21,200	Up to 10	4,750	20,000
c5.9xlarge	36	72	47,700	10	9,500	40,000

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
c5.18xlarge	72	144	95,400	25	19,000	80,000
c5a.large	2	4	2,746	Up to 10	Up to 3,170	13,300
c5a.xlarge	4	8	5,493	Up to 10	Up to 3,170	13,300
c5a.2xlarge	8	16	10,986	Up to 10	Up to 3,170	13,300
c5a.4xlarge	16	32	21,973	Up to 10	Up to 3,170	13,300
c5a.8xlarge	32	64	43,943	10	3,170	13,300
c5a.12xlarge	48	96	65,915	12	4,750	20,000
c5a.16xlarge	64	128	87,887	20	6,300	26,700
c5a.24xlarge	96	192	131,830	20	9,500	40,000
c5n.large	2	5	2,650	Up to 25	Up to 4,750	20,000
c5n.xlarge	4	11	5,300	Up to 25	Up to 4,750	20,000
c5n.2xlarge	8	21	10,600	Up to 25	Up to 4,750	20,000

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
c5n.4xlarge	16	42	21,200	Up to 25	4,750	20,000
c5n.9xlarge	36	96	47,700	50	9,500	40,000
c5n.18xlarge	72	192	95,400	100	19,000	80,000
c6a.large	2	4	2,864	Up to 12.5	Up to 10,000	40,000
c6a.xlarge	4	8	5,727	Up to 12.5	Up to 10,000	40,000
c6a.2xlarge	8	16	11,454	Up to 12.5	Up to 10,000	40,000
c6a.4xlarge	16	32	22,908	Up to 12.5	Up to 10,000	40,000
c6a.8xlarge	32	64	45,817	12.5	10,000	40,000
c6a.12xlarge	48	96	68,725	18.75	15,000	60,000
c6a.16xlarge	64	128	91,633	25	20,000	80,000
c6a.24xlarge	96	192	137,450	37.5	30,000	120,000
c6a.32xlarge	128	256	183,267	50	40,000	160,000

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
c6a.48xlarge	192	384	274,900	50	40,000	240,000
c6i.large	2	4	2,950	Up to 12.5	Up to 10,000	40,000
c6i.xlarge	4	8	5,899	Up to 12.5	Up to 10,000	40,000
c6i.2xlarge	8	16	11,799	Up to 12.5	Up to 10,000	40,000
c6i.4xlarge	16	32	23,598	Up to 12.5	Up to 10,000	40,000
c6i.8xlarge	32	64	47,195	12.5	10,000	40,000
c6i.12xlarge	48	96	70,793	18.75	15,000	60,000
c6i.16xlarge	64	128	94,390	25	20,600	80,000
c6i.24xlarge	96	192	141,585	37.5	30,000	120,000
c6i.32xlarge	128	256	188,780	50	40,000	160,000
c6id.large	2	4	2,950	Up to 12.5	Up to 10,000	40,000
c6id.xlarge	4	8	5,899	Up to 12.5	Up to 10,000	40,000
c6id.2xlarge	8	16	11,799	Up to 12.5	Up to 10,000	40,000

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
c6id.4xlarge	16	32	23,598	Up to 12.5	Up to 10,000	40,000
c6id.8xlarge	32	64	47,195	12.5	10,000	40,000
c6id.12xlarge	48	96	70,793	18.75	15,000	60,000
c6id.16xlarge	64	128	94,390	25	20,600	80,000
c6id.24xlarge	96	192	141,585	37.5	30,000	120,000
c6id.32xlarge	128	256	188,780	50	40,000	160,000

Memory Optimized

Instance type	vCPU	Memory (GiB)	SAPS	aSAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
r5.large	2	16	2,891	N/A	Up to 10	Up to 4,750	18,750
r5.xlarge	4	32	5,782	N/A	Up to 10	Up to 4,750	18,750
r5.2xlarge	8	64	11,564	N/A	Up to 10	Up to 4,750	18,750
r5.4xlarge	16	128	23,128	N/A	Up to 10	4,750	18,750

Instance type	vCPU	Memory (GiB)	SAPS	aSAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
r5.8xlarge	32	256	46,257	N/A	10	6,800	30,000
r5.12xlarge	48	384	69,385	N/A	10	9,500	40,000
r5.16xlarge	64	512	92,513	N/A	20	13,600	60,000
r5.24xlarge	96	768	138,770	N/A	25	19,000	80,000
r5.metal	96	768	143,230	N/A	25	19,000	80,000
r5b.large	2	16	2,891	N/A	Up to 10	Up to 10,000	43,333
r5b.xlarge	4	32	5,782	N/A	Up to 10	Up to 10,000	43,333
r5b.2xlarge	8	64	11,564	N/A	Up to 10	Up to 10,000	43,333
r5b.4xlarge	16	128	23,128	N/A	Up to 10	10,000	43,333
r5b.8xlarge	32	256	46,257	N/A	25	20,000	86,667
r5b.12xlarge	48	384	69,385	N/A	50	30,000	130,000
r5b.16xlarge	64	512	92,513	N/A	75	40,000	173,333
r5b.24xlarge	96	768	138,770	N/A	100	60,000	260,000

Instance type	vCPU	Memory (GiB)	SAPS	aSAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
r5b.metal	96	768	143,230	N/A	100	60,000	260,000
r5n.large	2	16	2,891	N/A	Up to 25	Up to 4,750	18,750
r5n.xlarge	4	32	5,782	N/A	Up to 25	Up to 4,750	18,750
r5n.2xlarge	8	64	11,564	N/A	Up to 25	Up to 4,750	18,750
r5n.4xlarge	16	128	23,128	N/A	Up to 25	4,750	18,750
r5n.8xlarge	32	256	46,257	N/A	25	6,800	30,000
r5n.12xlarge	48	384	69,385	N/A	50	9,500	40,000
r5n.16xlarge	64	512	92,513	N/A	75	13,600	60,000
r5n.24xlarge	96	768	138,770	N/A	100	19,000	80,000
r5n.metal	96	768	143,230	N/A	100	19,000	80,000
r6a.large	2	16	2,984	N/A	Up to 12.5	Up to 10,000	40,000
r6a.xlarge	4	32	5,968	N/A	Up to 12.5	Up to 10,000	40,000
r6a.2xlarge	8	64	11,935	N/A	Up to 12.5	Up to 10,000	40,000

Instance type	vCPU	Memory (GiB)	SAPS	aSAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
r6a.4xlarge	16	128	23,871	N/A	Up to 12.5	Up to 10,000	40,000
r6a.8xlarge	32	256	47,742	N/A	12.5	10,000	40,000
r6a.12xlarge	48	384	71,613	N/A	18.75	15,000	60,000
r6a.16xlarge	64	512	95,483	N/A	25	20,000	80,000
r6a.32xlarge	128	1,024	190,967	N/A	50	40,000	160,000
r6a.24xlarge	96	768	143,225	N/A	37.5	30,000	120,000
r6a.48xlarge	192	1,536	286,450	N/A	50	40,000	240,000
r6i.large	2	16	3,063	N/A	Up to 12.5	Up to 10,000	40,000
r6i.xlarge	4	32	6,127	N/A	Up to 12.5	Up to 10,000	40,000
r6i.2xlarge	8	64	12,253	N/A	Up to 12.5	Up to 10,000	40,000
r6i.4xlarge	16	128	24,506	N/A	Up to 12.5	Up to 10,000	40,000
r6i.8xlarge	32	256	49,013	N/A	12.5	10,000	40,000

Instance type	vCPU	Memory (GiB)	SAPS	aSAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
r6i.12xlarge	48	384	73,519	N/A	18.75	15,000	60,000
r6i.16xlarge	64	512	98,025	N/A	25	20,000	80,000
r6i.24xlarge	96	768	147,038	N/A	37.5	30,000	120,000
r6i.32xlarge	128	1,024	196,050	N/A	50	40,000	160,000
r6i.metal	128	1,024	203,600	N/A	50	40,000	160,000
r6id.large	2	16	3,063	N/A	Up to 12.5	Up to 10,000	40,000
r6id.xlarge	4	32	6,127	N/A	Up to 12.5	Up to 10,000	40,000
r6id.2xlarge	8	64	12,253	N/A	Up to 12.5	Up to 10,000	40,000
r6id.4xlarge	16	128	24,506	N/A	Up to 12.5	Up to 10,000	40,000
r6id.8xlarge	32	256	49,013	N/A	12.5	10,000	40,000
r6id.12xlarge	48	384	73,519	N/A	18.75	15,000	60,000
r6id.16xlarge	64	512	98,025	N/A	25	20,000	80,000
r6id.24xlarge	96	768	147,038	N/A	37.5	30,000	120,000

Instance type	vCPU	Memory (GiB)	SAPS	aSAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
r6id.32xlarge	128	1,024	196,050	N/A	50	40,000	160,000
r6idn.large	2	16	3,063	N/A	Up to 25	Up to 25,000	100,000
r6idn.xlarge	4	32	6,127	N/A	Up to 30	Up to 25,000	100,000
r6idn.2xlarge	8	64	12,253	N/A	Up to 40	Up to 25,000	100,000
r6idn.4xlarge	16	128	24,506	N/A	Up to 50	Up to 25,000	100,000
r6idn.8xlarge	32	256	49,013	N/A	50	25,000	100,000
r6idn.12xlarge	48	384	73,519	N/A	75	37,500	150,000
r6idn.16xlarge	64	512	98,025	N/A	100	50,000	200,000
r6idn.24xlarge	96	768	147,038	N/A	150	75,000	300,000
r6idn.32xlarge	128	1,024	196,050	N/A	200	100,000	400,000
r6id.metal	128	1,024	203,600	N/A	50	40,000	160,000
r6in.large	2	16	3,063	N/A	Up to 25	Up to 25,000	100,000

Instance type	vCPU	Memory (GiB)	SAPS	aSAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
r6in.xlarge	4	32	6,127	N/A	Up to 25	Up to 25,000	100,000
r6in.2xlarge	8	64	12,253	N/A	Up to 25	Up to 25,000	100,000
r6in.4xlarge	16	128	24,506	N/A	Up to 25	Up to 25,000	100,000
r6in.8xlarge	32	256	49,013	N/A	12.5	25,000	100,000
r6in.12xlarge	48	384	73,519	N/A	18.75	37,500	150,000
r6in.16xlarge	64	512	98,025	N/A	25	50,000	200,000
r6in.24xlarge	96	768	147,038	N/A	37.5	75,000	300,000
r6in.32xlarge	128	1,024	196,050	N/A	200	100,000	400,000
r7a.medium	1	8	2,255	N/A	Up to 12.5	Up to 10,000	Up to 40,000
r7a.large	2	16	4,510	N/A	Up to 12.5	Up to 10,000	Up to 40,000
r7a.xlarge	4	32	9,020	N/A	Up to 12.5	Up to 10,000	Up to 40,000
r7a.2xlarge	8	64	18,040	N/A	Up to 12.5	Up to 10,000	Up to 40,000

Instance type	vCPU	Memory (GiB)	SAPS	aSAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
r7a.4xlarge	16	128	36,080	N/A	Up to 12.5	Up to 10,000	Up to 40,000
r7a.8xlarge	32	256	72,160	N/A	12.5	10,000	40,000
r7a.12xlarge	48	384	108,240	N/A	18.75	15,000	60,000
r7a.16xlarge	64	512	144,320	N/A	25	20,000	80,000
r7a.24xlarge	96	768	216,480	N/A	37.5	30,000	120,000
r7a.32xlarge	128	1024	288,640	N/A	50	40,000	160,000
r7a.48xlarge	192	1536	432,980	N/A	50	40,000	240,000
r7i.large	2	16	4,155	N/A	Up to 12.5	Up to 10,000	40,000
r7i.xlarge	4	32	8,310	N/A	Up to 12.5	Up to 10,000	40,000
r7i.2xlarge	8	64	16,620	N/A	Up to 12.5	Up to 10,000	40,000
r7i.4xlarge	16	128	33,240	N/A	Up to 12.5	Up to 10,000	40,000
r7i.8xlarge	32	256	66,480	N/A	12.5	10,000	40,000

Instance type	vCPU	Memory (GiB)	SAPS	aSAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
r7i.12xlarge	48	384	99,720	N/A	18.75	15,000	60,000
r7i.16xlarge	64	512	105,500	N/A	25	20,000	80,000
r7i.24xlarge	96	768	158,250	N/A	37.5	30,000	120,000
r7i.48xlarge*	192	1536	296,200	N/A	50	40,000	240,000
r8i.large	2	16	4,802	740	Up to 12.5	Up to 10,000	40,000
r8i.xlarge	4	32	9,604	1,480	Up to 12.5	Up to 10,000	40,000
r8i.2xlarge	8	64	19,208	2,960	Up to 15	Up to 10,000	40,000
r8i.4xlarge	16	128	38,416	5,921	Up to 15	Up to 10,000	40,000
r8i.8xlarge	32	256	76,847	11,842	15	10,000	40,000
r8i.12xlarge	48	384	115,270	17,763	22.5	15,000	60,000
r8i.16xlarge	64	512	138,840	23,683	30	20,000	80,000
r8i.24xlarge	96	768	208,260	35,525	40	30,000	120,000

Instance type	vCPU	Memory (GiB)	SAPS	aSAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
r8i.32xlarge	128	1024	277,680	47,367	50	40,000	160,000
r8i.48xlarge	192	1536	416,520	71,050	75	60,000	240,000
r8i.96xlarge	384	3072	740,050	142,100	100	80,000	480,000
u-3tb1.56xlarge	224	3,072	237,750	N/A	50	19,000	80,000
u-6tb1.56xlarge	224	6,144	380,770	N/A	100	38,000	160,000
u-6tb1.112xlarge	448	6,144	475,500	N/A	100	38,000	160,000
u-6tb1.metal	448	6,144	480,600	N/A	100	38,000	160,000
u-9tb1.112xlarge	448	9,216	475,500	N/A	100	38,000	160,000
u-9tb1.metal	448	9,216	480,600	N/A	100	38,000	160,000
u-12tb1.112xlarge	448	12,288	475,500	N/A	100	38,000	160,000
u-12tb1.metal	448	12,288	480,600	N/A	100	38,000	160,000
u-18tb1.112xlarge	448	18,432	520,330	N/A	100	38,000	160,000

Instance type	vCPU	Memory (GiB)	SAPS	aSAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
u-18tb1.metal	448	18,432	534,130	N/A	100	38,000	160,000
u-24tb1.112xlarge	448	24,576	508,720	N/A	100	38,000	160,000
u-24tb1.metal	448	24,576	517,480	N/A	100	38,000	160,000
u7i-6tb.112xlarge	448	6,144	670,265	N/A	100	60,000	420,000
u7i-8tb.112xlarge	448	8,192	674,950	N/A	100	60,000	420,000
u7i-12tb.224xlarge	896	12,288	1,254,030	N/A	100	60,000	420,000
u7in-16tb.224xlarge	896	16,384	1,281,620	N/A	200	100,000	420,000
u7in-24tb.224xlarge	896	24,576	1,225,250	N/A	200	100,000	420,000
u7inh-32tb.480xlarge	1,920	32,768	N/A	N/A	200	160,000	840,000
x1.16xlarge	64	976	65,750	N/A	10	7,000	40,000
x1.32xlarge	128	1,952	131,500	N/A	25	14,000	80,000

Instance type	vCPU	Memory (GiB)	SAPS	aSAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
x1e.xlarge	4	122	4,109	N/A	Up to 10	500	3,700
x1e.2xlarge	8	244	8,219	N/A	Up to 10	1,000	7,400
x1e.4xlarge	16	488	16,437	N/A	Up to 10	1,750	10,000
x1e.8xlarge	32	976	32,875	N/A	Up to 10	3,500	20,000
x1e.16xlarge	64	1,952	65,750	N/A	10	7,000	40,000
x1e.32xlarge	128	3,904	131,500	N/A	25	14,000	80,000
x2idn.16xlarge	64	1,024	98,025	N/A	50	40,000	150,000
x2idn.24xlarge	96	1,536	147,038	N/A	75	60,000	200,000
x2idn.32xlarge	128	2,048	196,050	N/A	100	80,000	300,000
x2iedn.xlarge	4	128	5,906	N/A	Up to 25	Up to 20,000	12,500
x2iedn.2xlarge	8	256	11,813	N/A	Up to 25	Up to 20,000	25,000
x2iedn.4xlarge	16	512	23,625	N/A	Up to 25	Up to 20,000	50,000

Instance type	vCPU	Memory (GiB)	SAPS	aSAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
x2iedn.8xlarge	32	1,024	47,250	N/A	25	20,000	75,000
x2iedn.16xlarge	64	2,048	94,500	N/A	50	40,000	150,000
x2iedn.24xlarge	96	3,072	141,750	N/A	75	60,000	200,000
x2iedn.32xlarge	128	4,096	189,000	N/A	100	80,000	300,000
x2iezn.2xlarge	8	256	14,170	N/A	Up to 25	3,170	13,333
x2iezn.4xlarge	16	512	28,340	N/A	Up to 25	4,175	20,000
x2iezn.6xlarge	24	768	42,510	N/A	50	9,500	40,000
x2iezn.8xlarge	32	1,024	56,680	N/A	75	12,000	55,000
x2iezn.12xlarge	48	1,536	85,020	N/A	100	19,000	80,000

Storage Optimized

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
i3en.xlarge	4	32	5,782	Up to 25	Up to 4,750	20,000

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
i3en.2xlarge	8	64	11,564	Up to 25	Up to 4,750	20,000
i3en.3xlarge	12	96	17,346	Up to 25	Up to 4,750	20,000
i3en.6xlarge	24	192	34,693	25	4,750	20,000
i3en.12xlarge	48	384	69,385	50	9,500	40,000
i3en.24xlarge	96	768	138,770	100	19,000	80,000
i3en.metal	96	768	143,230	100	19,000	80,000
i4i.large	2	16	3,063	Up to 10	Up to 10,000	40,000
i4i.xlarge	4	32	6,127	Up to 10	Up to 10,000	40,000
i4i.2xlarge	8	64	12,253	Up to 12.5	Up to 10,000	40,000
i4i.4xlarge	16	128	24,506	Up to 25	Up to 10,000	40,000
i4i.8xlarge	32	256	49,013	18.75	10,000	40,000
i4i.12xlarge	48	384	73,519	28.125	15,000	60,000
i4i.16xlarge	64	512	98,025	37.5	20,000	80,000

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	Max IOPS
i4i.24xlarge	96	768	147,038	56.25	30,000	120,000
i4i.32xlarge	128	1,024	196,050	75	40,000	160,000
i4i.metal	128	1,024	203,600	75	40,000	160,000

*m7i.48xlarge and r7i.48xlarge are supported on Windows 2016 and later, SLES 15 SP3 and later, and RHEL 8.6 and later.

*advanced SAP Application Performance Standard (aSAPS) is a hardware-independent unit of measurement that describes the performance of a system configuration in the SAP environment. It is derived from the SAP Quote-to-cash Standard Application benchmark (Q2C)

Previous generation instances for SAP NetWeaver

Example

General Purpose

Instance type	vCPU	Memory (GiB)	SAPS
cc2.8xlarge	32	60.5	90,330
cr1.8xlarge	32	244	30,430
m2.2xlarge	4	32.2	3,700
m2.4xlarge	8	68.4	7,400
m4.large	2	8	2,366
m4.xlarge	4	16	4,732
m4.2xlarge	8	32	9,464

Instance type	vCPU	Memory (GiB)	SAPS
m4.4xlarge	16	64	18,928
m4.10xlarge	40	160	47,320
m4.16xlarge	64	256	75,770

Compute Optimized

Instance type	vCPU	Memory (GiB)	SAPS
c3.large	2	3.75	1,995
c3.xlarge	4	7	3,990
c3.2xlarge	8	15	7,980
c3.4xlarge	16	30	15,915
c3.8xlarge	32	60	31,830
c4.large	2	3.75	2,379
c4.xlarge	4	7.5	4,758
c4.2xlarge	8	15	9,515
c4.4xlarge	16	30	19,030
c4.8xlarge	36	60	37,950

Memory Optimized

Instance type	vCPU	Memory (GiB)	SAPS
r3.large	2	15	1,995
r3.xlarge	4	30.5	3,990

Instance type	vCPU	Memory (GiB)	SAPS
r3.2xlarge	8	61	7,980
r3.4xlarge	16	122	15,960
r3.8xlarge	32	244	31,920
r4.large	2	15.25	2,387
r4.xlarge	4	30.5	4,775
r4.2xlarge	8	61	9,550
r4.4xlarge	16	122	19,100
r4.8xlarge	32	244	38,200
r4.16xlarge	64	488	76,400

SAP HANA certified and non-certified instances

AWS has worked closely with SAP to test and certify Amazon EC2 instance types for SAP on AWS solutions.

Previous generation Amazon EC2 instances for SAP HANA are fully supported and these instance types retain the same features and functionality. We recommend using the current generation Amazon EC2 instance for new SAP HANA implementations or migrations.

All current and previous generation Amazon EC2 instance types for SAP HANA can be used for running non-production workloads. For more information, see SAP Note 2271345.

Contents

- SAP HANA OLTP and OLAP Scale-out
- Non-certified instances

Current generation certified instances

SAP HANA OLTP and OLAP Scale-up

Instanc type	vCPU	Memor (GiB)	SAPS	aSAPS	Netwo (Gbps)	Storag (Mbps)	SAP HANA OLTP prod	SAP HANA OLTP sizing	SAP HANA OLAP prod	SAP HANA OLAP sizing	FSx for ONTAP
r5.8xla e	32	256	46,257	N/A	10	6,800	✓	Standa	✓	Standa	✗
r5.12xl ge	48	384	69,385	N/A	10	9,500	✓	Standa	✓	Standa	✗
r5.16xl ge	64	512	92,513	N/A	20	13,600	✓	Standa	✓	Standa	✗
r5.24xl ge	96	768	138,77	N/A	25	19,000	✓	Standa	✓	Standa	✗
r5.met	96	768	143,23	N/A	25	19,000	✓	Standa	✓	Standa	✗
r5b.8xl ge	32	256	46,257	N/A	25	20,000	✓	Standa	✓	Standa	✗
r5b.12 rge	48	384	69,385	N/A	50	30,000	✓	Standa	✓	Standa	✗
r5b.16 rge	64	512	92,513	N/A	75	40,000	✓	Standa	✓	Standa	✗
r5b.24 rge	96	768	138,77	N/A	100	60,000	✓	Standa	✓	Standa	✗

Instanc type	vCPU	Memor (GiB)	SAPS	aSAPS	Netwo (Gbps)	Storag (Mbps)	SAP HANA OLTP prod	SAP HANA OLTP sizing	SAP HANA OLAP prod	SAP HANA OLAP sizing	FSx for ONTAP
r5b.me	96	768	143,23	N/A	100	60,000	✓	Standa	✓	Standa	✗
r6i.8xla ge	32	256	49,013	N/A	12.5	10,000	✓	Standa	✗	N/A	✗
r6i.12x rge	48	384	73,519	N/A	18.75	15,000	✓	Standa	✓	Standa	✓
r6i.16x rge	64	512	98,025	N/A	25	20,000	✓	Standa	✓	Standa	✓
r6i.24x rge	96	768	147,03	N/A	37.5	30,000	✓	Standa	✓	Standa	✓
r6i.32x rge	128	1,024	196,05	N/A	50	40,000	✓	Standa	✓	Standa	✓
r7i.8xla ge	32	256	66,480	N/A	12.5	10,000	✓	Standa	✓	Standa	✓
r7i.12x rge	48	384	99,720	N/A	18.75	15,000	✓	Standa	✓	Standa	✓
r7i.16x rge	64	512	105,50	N/A	25	20,000	✓	Standa	✓	Standa	✓
r7i.24x rge	96	768	158,25	N/A	37.5	30,000	✓	Standa	✓	Standa	✓
r7i.48x rge	192	1536	296,20	N/A	50	40,000	✓	Standa	✓	Standa	✓
r8i.12x rge	48	384	115,27	17,763	22.5	15,000	✓	Standa	✓	Standa	✓

Instanc type	vCPU	Memor (GiB)	SAPS	aSAPS	Netwo (Gbps)	Storag (Mbps)	SAP HANA OLTP prod	SAP HANA OLTP sizing	SAP HANA OLAP prod	SAP HANA OLAP sizing	FSx for ONTAP
r8i.16x rge	64	512	138,84	23,683	30	20,000	✓	Standa	✓	Standa	✓
r8i.24x rge	96	768	208,26	35,525	40	30,000	✓	Standa	✓	Standa	✓
r8i.32x rge	128	1024	277,68	47,367	50	40,000	✓	Standa	✓	Standa	✓
r8i.48x rge	192	1536	416,52	71,050	75	60,000	✓	Standa	✓	Standa	✓
r8i.96x rge	384	3072	740,05	142,10	100	80,000	✓	Standa	✓	Standa	✓
u-3tb1 xlarge	224	3,072	237,75	N/A	50	19,000	✓	Standa	✓	Worklo	✓
u-6tb1 xlarge	224	6,144	380,77	N/A	100	38,000	✓	Standa	✓	Worklo	✓
u-6tb1 2xlarge	448	6,144	475,50	N/A	100	38,000	✓	Standa	✓	Standa	✓
u-6tb1 tal	448	6,144	480,60	N/A	100	38,000	✓	Standa	✓	Standa	✗
u-9tb1 2xlarge	448	9,216	475,50	N/A	100	38,000	✓	Standa	✓	Worklo	✓
u-9tb1 tal	448	9,216	480,60	N/A	100	38,000	✓	Standa	✓	Worklo	✗

Instanc type	vCPU	Memor (GiB)	SAPS	aSAPS	Netwo (Gbps)	Storag (Mbps)	SAP HANA OLTP prod	SAP HANA OLTP sizing	SAP HANA OLAP prod	SAP HANA OLAP sizing	FSx for ONTAP
u-12tb 12xlarg	448	12,288	475,50	N/A	100	38,000	✓	Standa	✓	Worklo	✓
u-12tb etal	448	12,288	480,60	N/A	100	38,000	✓	Standa	✓	Worklo	✗
u-18tb 12xlarg	448	18,432	520,33	N/A	100	38,000	✓	Worklo	✓	Worklo	✓
u-18tb etal	448	18,432	534,13	N/A	100	38,000	✓	Worklo	✓	Worklo	✗
u-24tb 12xlarg	448	24,576	508,72	N/A	100	38,000	✓	Worklo	✗	N/A	✓
u-24tb etal	448	24,576	517,48	N/A	100	38,000	✓	Worklo	✗	N/A	✗
u7i-6tb 12xlarg	448	6,144	670,26	N/A	100	60,000	✓	Standa	✓	Standa	✓
u7i-8tb 12xlarg	448	8,192	674,95	N/A	100	60,000	✓	Standa	✓	Standa	✓
u7i-12t 224xla	896	12,288	1,254,(	N/A	100	60,000	✓	Standa	✓	Standa	✓
u7in-1(.224xla e	896	16,384	1,281,(	N/A	200	100,00	✓	Standa	✓	Standa	✓
u7in-2 .224xla e	896	24,576	1,225,'	N/A	200	100,00	✓	Worklo	✓	Worklo	✓

Instanc type	vCPU	Memor (GiB)	SAPS	aSAPS	Netwo (Gbps)	Storag (Mbps)	SAP HANA OLTP prod	SAP HANA OLTP sizing	SAP HANA OLAP prod	SAP HANA OLAP sizing	FSx for ONTAP
u7inh-3 b.480x ge	1,920	32,768	N/A	N/A	200	160,00	✓	Standa	✓	Standa	✓
x1.16xl ge	64	976	65,750	N/A	10	7,000	✓	Standa	✓	Standa	✗
x1.32xl ge	128	1,952	131,50	N/A	25	14,000	✓	Standa	✓	Standa	✗
x1e.32 rge	128	3,904	131,50	N/A	25	14,000	✓	Standa	✓	Worklo	✗
x2idn.1 large	64	1,024	98,025	N/A	50	40,000	✓	Standa	✓	Worklo	✓
x2idn.2 large	96	1,536	147,03	N/A	75	60,000	✓	Standa	✓	Worklo	✓
x2idn.3 large	128	2,048	196,05	N/A	100	80,000	✓	Standa	✓	Standa	✓
x2iedn xlarge	96	3,072	141,75	N/A	75	60,000	✓	Standa	✓	Worklo	✓
x2iedn xlarge	128	4,096	189,00	N/A	100	80,000	✓	Standa	✓	Worklo	✓

SAP HANA OLTP and OLAP Scale-out

Instar type	vCPU	Mem (GiB)	SAPS	aSAP	Netw (Gbps	Stora (Mbp	SAP HANA OLTP prod	SAP HANA OLTP sizing	OLTP Max node	SAP HANA OLAP prod	SAP HANA OLAP sizing	OLAP Max node	FSx for ONTAP
r5.24x ge	96	768	138,7	N/A	25	19,00	✗	N/A	N/A	✓	Stand	16	✗
r6i.24 rge	96	768	147,0	N/A	37.5	30,00	✗	N/A	N/A	✓	Stand	16	✓
r6i.32 rge	128	1,024	196,0	N/A	50	40,00	✗	N/A	N/A	✓	Stand	16	✓
u-6tb xlarge	224	6,144	380,7	N/A	100	38,00	✓	Stand	4	✓	Workl	16	✓
u-6tb 2xlarg	448	6,144	475,5	N/A	100	38,00	✓	Stand	4	✓	Stand	16	✓
u-6tb tal	448	6,144	480,6	N/A	100	38,00	✓	Stand	4	✓	Stand	16	✗
u-9tb 2xlarg	448	9,216	475,5	N/A	100	38,00	✓	Stand	4	✓	Workl	16	✓
u-12t 12xla	448	12,28	475,5	N/A	100	38,00	✓	Stand	4	✓	Workl	16	✓
u-12t etal	448	12,28	480,6	N/A	100	38,00	✓	Stand	4	✗	N/A	N/A	✗
u7i-6t 12xla	448	6,144	670,2	N/A	100	60,00	✗	N/A	N/A	✓	Stand	16	✗
u7i-8t 12xla	448	8,192	674,9	N/A	100	60,00	✗	N/A	N/A	✓	Stand	16	✗

Instar type	vCPU	Mem (GiB)	SAPS	aSAP	Netw (Gbps	Stora (Mbp	SAP HANA OLTP prod	SAP HANA OLTP sizing	OLTP Max nodes	SAP HANA OLAP prod	SAP HANA OLAP sizing	OLAP Max nodes	FSx for ONTAP
u7in-1 .224x e	896	12,28	1,254	N/A	100	60,00	✓	Stand	4	✓	Stand	8	✓
u7in-1 .224x e	896	16,38	1,281	N/A	200	100,0	✓	Stand	4	✓	Stand	8	✓
u7in-2 .224x e	896	24,57	1,225	N/A	200	100,0	✓	Workl	4	✓	Workl	8	✓
u7inh b.480 ge	1,920	32,76	N/A	N/A	200	160,0	✓	Stand	4	✓	Workl	8	✗
x1.16 ge	64	976	65,75	N/A	10	7,000	✗	N/A	N/A	✓	Stand	7	✗
x1.32 ge	128	1,952	131,5	N/A	25	14,00	✗	N/A	N/A	✓	Stand	25	✗
x1e.3 rge	128	3,904	131,5	N/A	25	14,00	✗	N/A	N/A	✓	Workl	25	✗
x2idn large	64	1,024	98,02	N/A	50	40,00	✗	N/A	N/A	✓	Stand	16	✓
x2idn large	96	1,536	147,0	N/A	75	60,00	✗	N/A	N/A	✓	Workl	16	✓
x2idn large	128	2,048	196,0	N/A	100	80,00	✗	N/A	N/A	✓	Workl	16	✓

Instar type	vCPU	Mem (GiB)	SAPS	aSAP	Netw (Gbps	Stora (Mbp	SAP HANA OLTP prod	SAP HANA OLTP sizing	OLTP Max node	SAP HANA OLAP prod	SAP HANA OLAP sizing	OLAP Max node	FSx for ONTAP
x2iedn xlarge	96	3,072	141,7	N/A	75	60,00	✗	N/A	N/A	✓	Workl	16	✓
x2iedn xlarge	128	4,096	189,0	N/A	100	80,00	✗	N/A	N/A	✓	Workl	16	✓

Previous generation certified instances

SAP HANA OLTP and OLAP Scale-up

Instance type	vCPU	Memory (GiB)	SAPS	SAP HANA OLTP prod	SAP HANA OLTP sizing	SAP HANA OLAP prod	SAP HANA OLAP sizing
r3.2xlarge	8	61	7,980	✗	Standard	✗	Standard
r3.4xlarge	16	122	15,960	✗	Standard	✗	Standard
r3.8xlarge	32	244	31,920	✓	Standard	✓	Standard
r4.8xlarge	32	244	38,200	10	7,000	✓	Standard
r4.16xlarge	64	488	76,400	25	14,000	✓	Standard

SAP HANA OLTP and OLAP Scale-out

Instance type	vCPU	Memory (GiB)	SAPS	SAP HANA OLTP prod	SAP HANA OLTP sizing	SAP HANA OLAP prod	SAP HANA OLAP sizing
r3.8xlarge	32	244	31,920	✗	N/A	✓	Standard

Non-certified instances

The Amazon EC2 instances in the following table are not certified for production usage. You can use them for running non-production workloads. For more information, see SAP Note 2271345 – Cost-Optimized SAP HANA Hardware for Non-Production Usage (SAP portal access required).

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	FSx for ONTAP
r4.2xlarge	8	61	9,550	Up to 10	1,700	✗
r4.4xlarge	16	122	19,100	Up to 10	3,500	✗
r5.2xlarge	8	64	11,564	Up to 10	Up to 4,750	✗
r5.4xlarge	16	128	23,128	Up to 10	4,750	✗
r5b.2xlarge	8	64	11,564	Up to 10	Up to 10,000	✗
r5b.4xlarge	16	128	23,128	Up to 10	10,000	✗
r6i.2xlarge	8	64	12,253	Up to 12.5	Up to 10,000	✗

Instance type	vCPU	Memory (GiB)	SAPS	Network (Gbps)	Storage (Mbps)	FSx for ONTAP
r6i.4xlarge	16	128	24,506	Up to 12.5	Up to 10,000	✗
x1e.xlarge	4	122	4,109	Up to 10	500	✗
x1e.2xlarge	8	244	8,219	Up to 10	1,000	✗
x1e.4xlarge	16	488	16,437	Up to 10	1,750	✗
x2iedn.xlarge	4	128	5,906	Up to 25	Up to 20,000	✗
x2iedn.2xlarge	8	256	11,813	Up to 25	Up tp 20,000	✗
x2iedn.4xlarge	16	512	23,625	Up to 25	Up to 20,000	✗
x2iedn.8xlarge	32	1,024	47,250	25	20,000	✗
x2iedn.16xlarge	64	2,048	94,500	50	40,000	✗

SAP Business One certified instances, version for SAP HANA

AWS has worked closely with SAP to test and certify Amazon EC2 instance types for SAP on AWS solutions.

Example

Current Generation

Instance type	vCPU	Memory (GiB)	SAPS	Max concurrent users
r5.2xlarge	8	64	11,564	25
r5.4xlarge	16	128	23,128	50
r5.12xlarge	48	384	69,385	150
r5.24xlarge	96	768	138,770	250
r6i.2xlarge	8	64	12,253	25
r6i.4xlarge	16	128	24,506	50
r6i.8xlarge	32	256	49,013	100
r7i.2xlarge	8	64	16,620	25
r7i.4xlarge	16	128	33,240	50
r7i.8xlarge	32	256	66,480	100
r7i.12xlarge	48	384	99,720	150
x1.16xlarge	64	976	65,750	200

Previous Generation

Instance type	vCPU	Memory (GiB)	SAPS	Max concurrent users
c3.8xlarge	32	60	31,830	25
m4.10xlarge	40	160	47,320	50

Instance type	vCPU	Memory (GiB)	SAPS	Max concurrent users
m4.16xlarge	64	256	75,770	100
r3.8xlarge	32	244	31,920	50

Document history for instance types for SAP on AWS

Change	Date
Added Scale-Out (OLAP) support for u7inh-32tb.480xlarge	June 2025
Added U7i to SAP NetWeaver and SAP HANA	May 2024
Added R7i to SAP Business One	January 2024
Added M7i and R7i to SAP NetWeaver	October 2023
Added R7a to SAP NetWeaver	September 2023
Added M7a to SAP NetWeaver	August 2023
Added R6i to SAP Business One	February 2023
Added R6id, R6in, R6idn, M6id, M6in, and M6idn to SAP NetWeaver	January 2023
Added FSx for ONTAP support for u-18tb1.112 and u-24tb1.112	January 2023
Added u-24tb1 and u-18tb1 to SAP HANA and SAP NetWeaver	October 2022
Added I4i to SAP NetWeaver	September 2022
Added FSx for ONTAP support for SAP HANA	September 2022

Change	Date
Added R6a to SAP NetWeaver	July 2022
Added C6i, C6id, and C6a to SAP NetWeaver	July 2022
Added X2idn and R6i to SAP HANA OLAP scale-out	June 2022
Initial publication	April 2022

AWS Data Provider for SAP

AWS Data Provider for SAP is a tool that collects performance-related data from AWS services. It makes this data available to SAP applications to help monitor and improve the performance of business transactions. SAP requires customers to install the agent as described in SAP note 1656250 (login credentials required).

The AWS Data Provider for SAP uses operating system, network, and storage data that is most relevant to the operation of the SAP infrastructure. Its data sources include Amazon Elastic Compute Cloud (Amazon EC2) and Amazon CloudWatch. This guide provides installation, configuration, and troubleshooting information for the AWS Data Provider for SAP on both Linux and Windows.

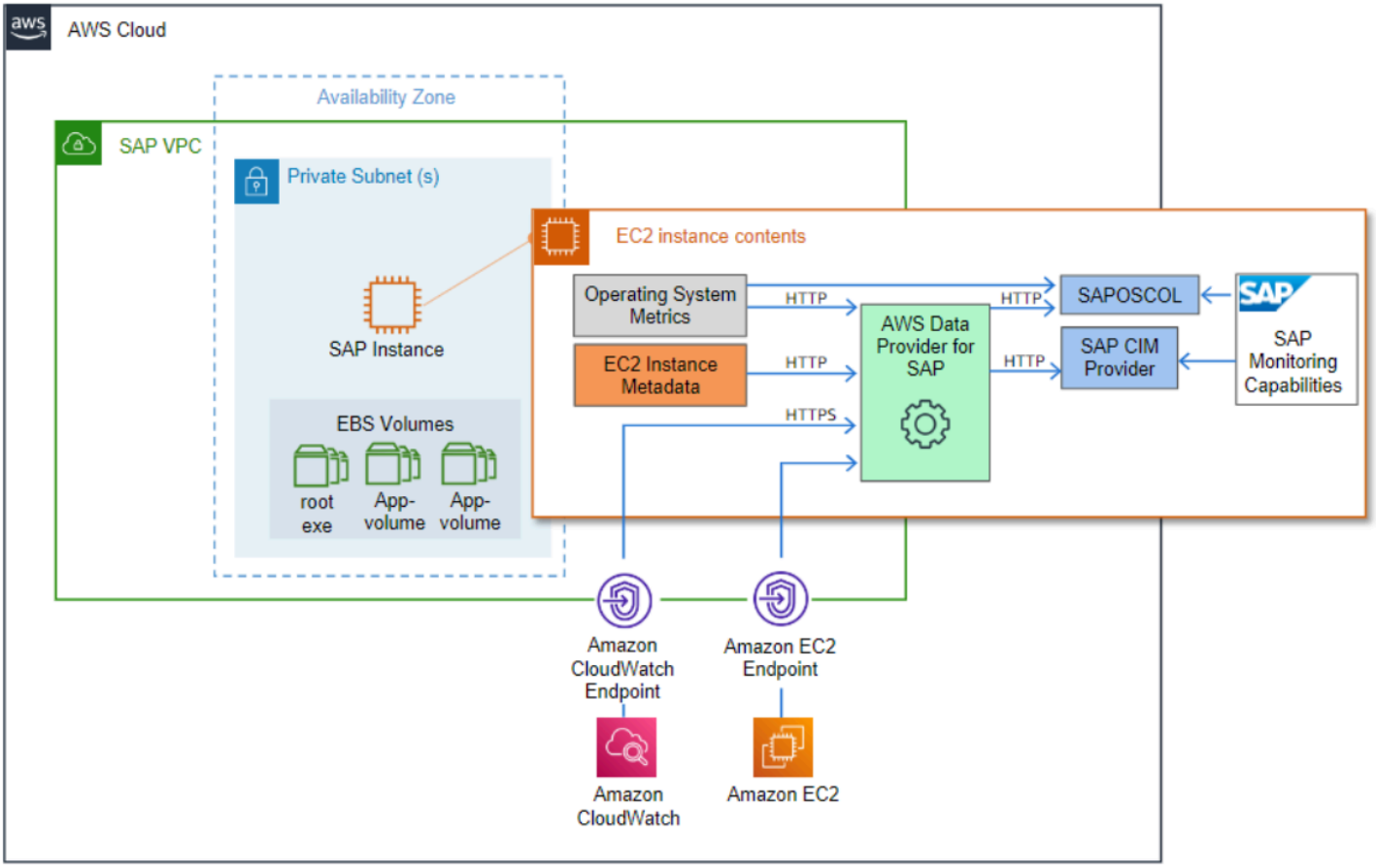
Introduction

Many organizations of all sizes are choosing to host key SAP systems in the Amazon Web Services Cloud. With AWS, you can quickly provision an SAP environment. Additionally, the elastic nature of the AWS Cloud enables you to scale computing resources up and down as needed. As a result, your business can dedicate more resources (both people and funds) to innovation.

Many SAP systems operate daily business transactions and are critical to business functions. As an SAP customer, you need the ability to track and troubleshoot the performance of these transactions. The AWS Data Provider for SAP is a tool that collects key performance data on an Amazon Elastic Compute Cloud (Amazon EC2) instance that SAP applications can use to monitor transactions built by SAP. The data is collected from a variety of sources within your AWS Cloud operating environment, including Amazon EC2 and Amazon CloudWatch. This data includes information about the operating system, network, and storage that is relevant to your SAP infrastructure. Data from the AWS Data Provider for SAP is read by the SAP Operating System Collector (SAPOSCOL) and the SAP CIM Provider.

The diagram provides a high-level illustration of the AWS Data Provider for SAP, its data sources, and its outputs.

Data sources for the AWS Data Provider for SAP



The purpose of this guide is to help you:

- Understand the technical requirements and components necessary to install and operate the AWS Data Provider for SAP.
- Install the AWS Data Provider for SAP.
- Understand the update process for the AWS Data Provider for SAP.
- Troubleshoot installation issues.

Pricing

The DataProvider agent is provided free of charge. However, there are indirect costs associated with running the agent due to SAP requiring monitoring data to be delivered at specific intervals. This causes the DataProvider to do frequent **GetMetric** calls to Amazon CloudWatch and the Amazon EC2 API to retrieve the metric data. The expected costs for these calls ranges approximately from **$20.00 to $40.00** per month per system and will vary based on how many disks are attached to the Amazon EC2 instance.

Example: Costs per month for using the DataProvider agent in the US East (N. Virginia) Region.

Fixed:

- Running the 2 required Amazon VPC endpoints (monitoring, Amazon EC2) is approximately **$14.00 + $0.01** per processed GB of data.

> **Note**
>
> These endpoints only need to be created once and are shared by the entire landscape. If you are already using these endpoints, you do not need to create them again.

Per System:

- You should expect around 70,000 API calls a day per instance (with 6 disks attached. At **$0.01** per 1,000 calls, it is approximately **$21.00** per month. The API call number increases or decreases based on the number of disks that are attached.

Technical Requirements

Before creating an SAP instance, ensure that the following technical requirements are met.

Topics

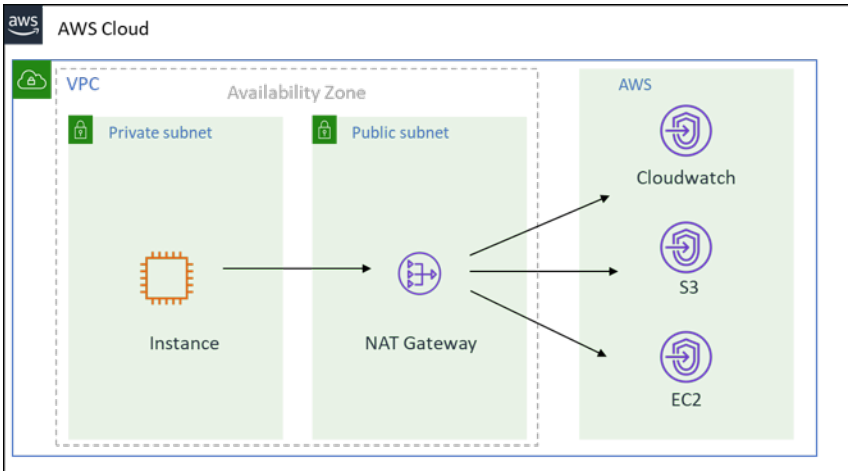
- Amazon VPC Network Topologies
- Amazon VPC Endpoints
- IAM Roles

Amazon VPC Network Topologies

You need to deploy SAP systems that receive information from the AWS Data Provider for SAP within an Amazon Virtual Private Cloud (Amazon VPC). You can use one of the following network topologies to enable routing to internet-based endpoints:

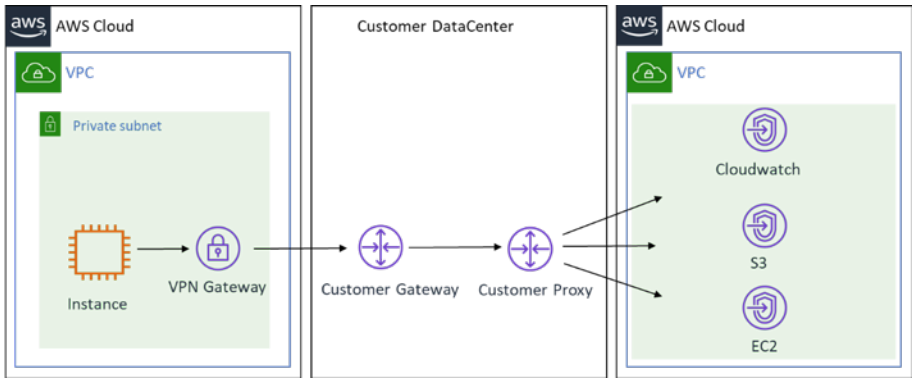
- The first topology configures routes and traffic directly to the AWS Cloud through a NAT gateway within an Amazon VPC. For more information about internet gateways, see the AWS documentation.

Connection to the AWS Cloud via an internet gateway



- A second topology routes traffic from the Amazon VPC, through your organization's on-premises data center, and back to AWS Cloud. For more information about this topology, see the [What is AWS Site-to-Site VPN?]

Connection to the Amazon Web Services Cloud via an on-premises data center



Amazon VPC Endpoints

Create endpoints for the following services that the DataProvider uses:

- Monitoring
- Amazon EC2

To create data endpoints in the AWS console, use the following procedure for each of the two endpoints:

1. Sign in to the [Amazon VPC console], navigate to **Endpoints**, and select **Create Endpoint**.

2. On the next screen, search for the service name, then select the appropriate VPC and route table, and select **Create Endpoint**.
3. After creating all three endpoints you should see them in your list of endpoints as shown below:

IAM Roles

You need to grant the AWS Data Provider for SAP read-only access to the Amazon CloudWatch, Amazon Simple Storage Service (Amazon S3), and Amazon EC2 services so that you can use their APIs. You can do this by creating an AWS Identity and Access Management (IAM) role for your Amazon EC2 instance and attaching a permissions policy.

Use the following procedure to create an IAM role and grant permissions to your Amazon EC2 instance:

1. Sign in to the AWS Management Console and open the IAM console.
2. In the navigation pane, select **Roles**, and select **Create role**.
3. Choose the **AWS service** role type, and select **EC2**.
4. Select **EC2** as the use case, and select **Next Permissions**.
5. Select **Create Policy**, and select **JSON**.
6. Copy and paste the following policy into the input field, replace all existing text, and select **Review Policy**.

> **Note**
>
> If your Amazon EC2 instances are running in Beijing and Ningxia, you must update the **Resource line** with the correct region.

See the following example policies based on your AWS Region.

AWS Regions (except AWS GovCloud (US-East), AWS GovCloud (US-West), Beijing and Ningxia)

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
```

```
            "Sid": "VisualEditor0",
            "Effect": "Allow",
            "Action": [
                "EC2:DescribeInstances",
                "cloudwatch:GetMetricStatistics",
                "EC2:DescribeVolumes"
            ],
            "Resource": "*"
        },
        {
         "Sid": "VisualEditor1",
         "Effect": "Allow",
         "Action": "s3:GetObject",
         "Resource": [
          "arn:aws:s3:::aws-sap-dataprovider-<us-east-1>/config.properties"
           ]
        }
    ]
}
```

Beijing and Ningxia

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "VisualEditor0",
            "Effect": "Allow",
            "Action": [
                "EC2:DescribeInstances",
                "cloudwatch:GetMetricStatistics",
                "EC2:DescribeVolumes"
            ],
            "Resource": "*"
        },
        {
         "Sid": "VisualEditor1",
         "Effect": "Allow",
         "Action": "s3:GetObject",
         "Resource": [
          "arn:aws-cn:s3:::aws-sap-dataprovider-<cn-north-1>/config.properties"
           ]
        }
```

```
    ]
}
```

AWS GovCloud (US-East) and AWS GovCloud (US-West)

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "VisualEditor0",
            "Effect": "Allow",
            "Action": [
                "EC2:DescribeInstances",
                "cloudwatch:GetMetricStatistics",
                "EC2:DescribeVolumes"
            ],
            "Resource": "*"
        },
        {
         "Sid": "VisualEditor1",
         "Effect": "Allow",
         "Action": "s3:GetObject",
         "Resource": [
          "arn:aws-us-gov:s3:::aws-sap-dataprovider-<us-gov-west-1>/
config.properties"
            ]
        }
    ]
}
```

7. Provide a **Name** and **Description** for the role, and select **Create Policy**.
8. Select **Create Policy**. The IAM console confirms the new policy with a message similar to the following.
9. Navigate to the **Create Role** page, refresh the screen, search for the newly created role, and select the policy.
10. Select **Next:Tags**.
11. Add any tags if needed, otherwise select **Next:Review**.
12. Provide a name for the Role and select **Create Role**.

DataProvider 4.3

If you are new to AWS Data Provider for SAP, see Installing DataProvider 4.3.

If you need to update or uninstall DataProvider 4.3, see Updating to DataProvider 4.3.

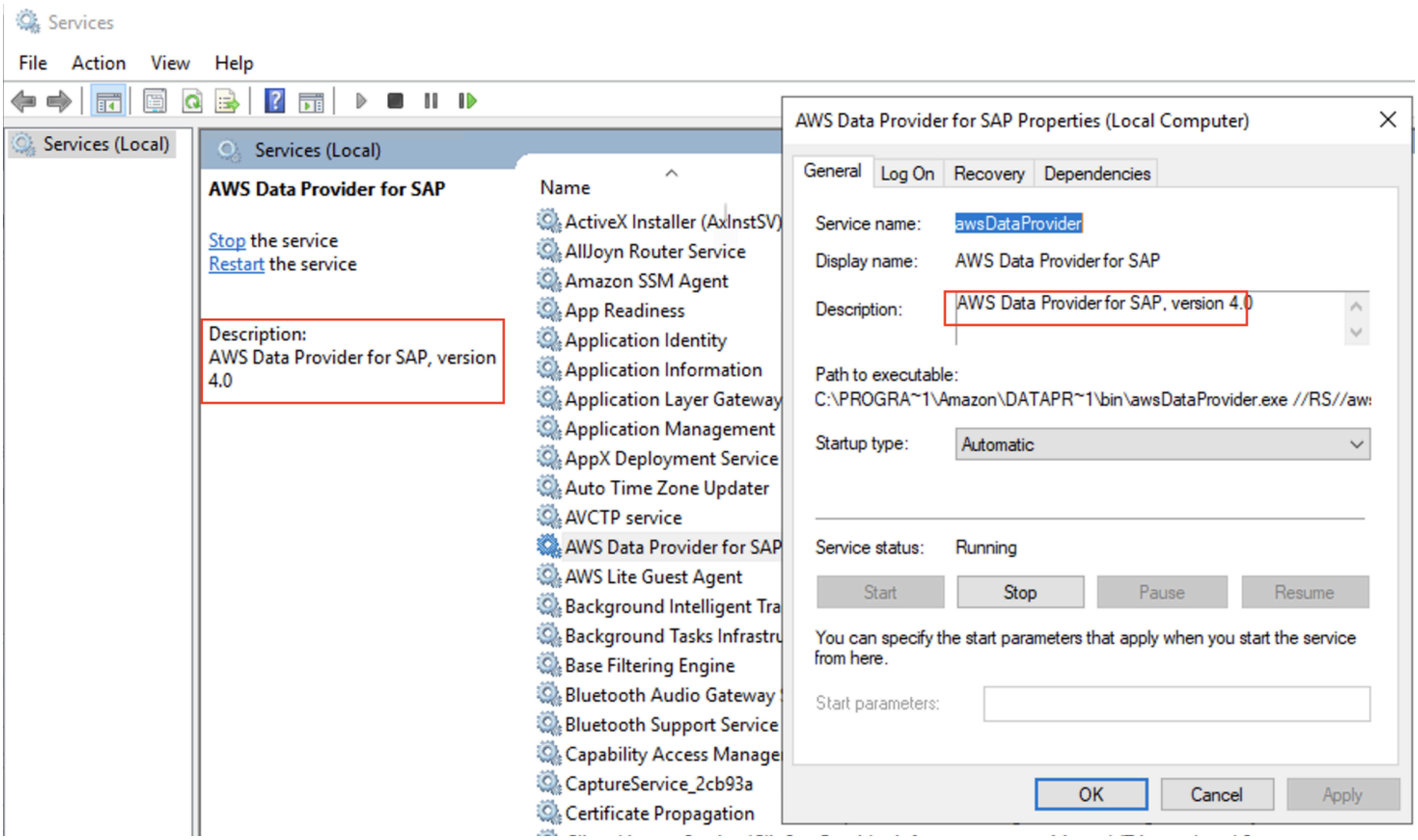
If you have an older version on your system, see Uninstalling older versions.

> **Important**
>
> All the previous versions (v1, v2, v3) of the DataProvider have been deprecated and will no longer receive updates. For new DataProvider installations, you must install DataProvider 4.3 using an SSM distributor.

Run the following command to check the current version of DataProvider on your system.

```
rpm -qa | grep aws-sap
```

To check the current version of DataProvider on **Windows**, go to **Services(Local)**, select **AWS Data Provider for SAP**, and open **Properties**. You can see the current version in the Description field.



Installing DataProvider 4.3

The AWS Data Provider for SAP runs as a service that automatically starts at boot and collects, aggregates, and exposes metrics to the SAP host agent. Metrics are sourced from a variety of providers that pull metrics from the relevant areas of the platform. The AWS Data Provider for SAP is designed to continue operating, regardless of whether its providers have connectivity or permissions to access the AWS service metrics they are requesting. Providers that cannot reach the metrics they are harvesting return blank values.

For example, if your Amazon EC2 instance does not have an IAM role associated with it that grants explicit access to the Amazon CloudWatch **GetMetricStatistics** API, the CloudWatch provider will be unable to perform the **GetMetricStatistics** action on the Amazon EC2 instance and will return blank values.

The provider needs to be installed on each SAP production system in order to be eligible for SAP support. You can only install one instance of the provider at a time on a system.

The AWS Data Provider for SAP is designed to automatically update itself so that it can provide you with the most current metrics. When the AWS Data Provider for SAP starts up, a built-in

update service retrieves the latest versions of its components and metric definitions from an AWS managed Amazon S3 bucket. If the AWS Data Provider for SAP cannot access the update service, it will continue to run as-is.

Installing with an SSM distributor – DataProvider 4.3 (Recommended)

The DataProvider 4.3 version enables you to install the package through a SSM distributor. AWS recommends using this method for installation, you can install the DataProvider using a either a Linux or Windows platform.

Prerequisites for installing the DataProvider using a SSM distributor

SSM-Agent

You must have the `ssm-agent` installed on your instance before you can use the SSM distributor to install the DataProvider Agent. Use the following AWS Systems Manager user guide to install the `ssm-agent` on your instances.

- RHEL: Manually installing SSM Agent on Red Hat Enterprise Linux instances
- SUSE: Manually install SSM Agent on SUSE Linux Enterprise Server instances
- Oracle : Manually installing SSM Agent on Oracle Linux instances
- Windows: Manually installing SSM Agent on Amazon EC2 instances for Windows Server

Java Runtime

The DataProvider is a java application that needs Java Runtime to be installed on the instance to run.

If your instance doesn't already have Java Runtime installed, you can use OpenJDK provided by Amazon Corretto to install the Java Runtime.

With DataProvider 4.3, the following Java Runtime versions are supported:

- Amazon Corretto 8 or OpenJDK 8
- Amazon Corretto 11 or OpenJDK 11
- Amazon Corretto 17 or OpenJDK 17

For more information on how to download and install JDK on your Amazon EC2 instance, see Amazon Corretto Documentation.

In the terminal, run the following commands to verify installation.

```
java -version
```

For example, expected output for Coretto-8.252.09.1 should be as follows:

```
openjdk version "1.8.0_252"OpenJDK Runtime Environment Corretto-8.252.09.1 (build
 1.8.0_252-b09)OpenJDK 64-Bit Server VM Corretto-8.252.09.1 (build 25.252-b09, mixed
 mode)
```

GPG Key

If you are a SUSE user, you must download the DataProvider GPG key and import it before installation.

- GPG Key URL: GPG Key
- Sign in to your SUSE instance and run the following commands to import the key:

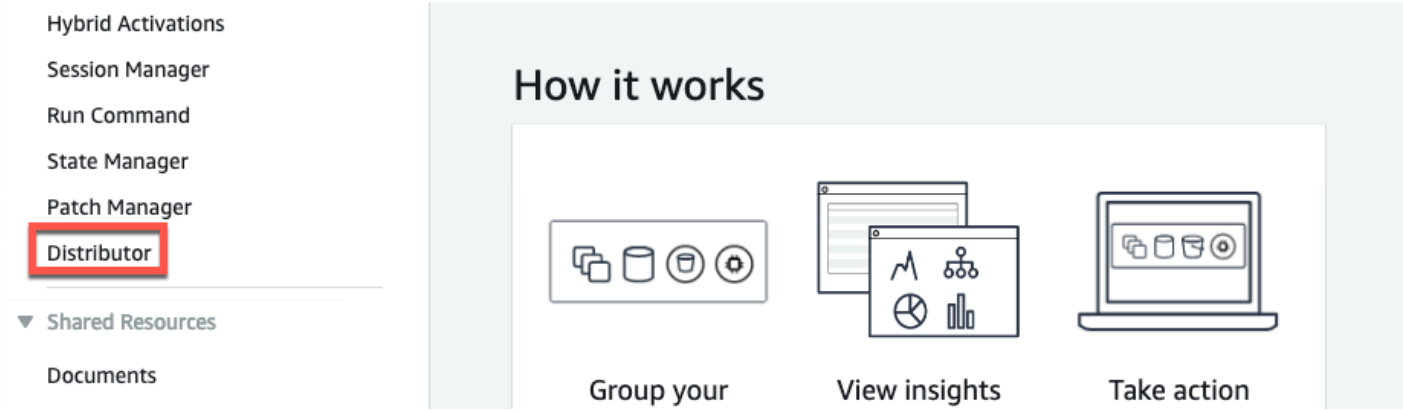
```
wget https//<url to GPG key>
```

```
rpm --import RPM-GPG-KEY-AWS
```

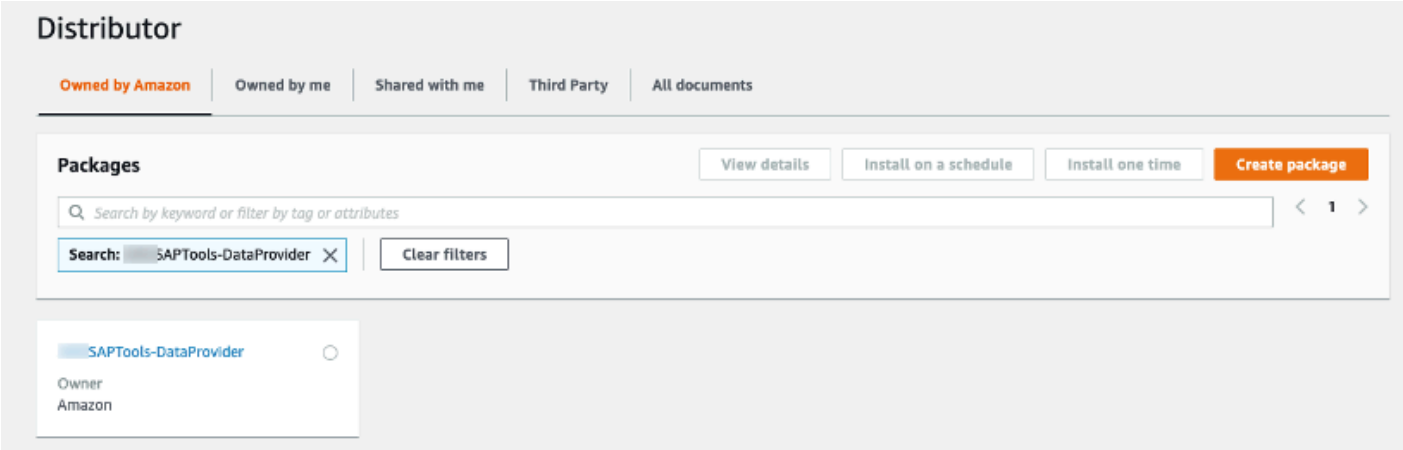
Installing the DataProvider Agent using an SSM distributor

Use the following procedure to install DataProvider 4.3.

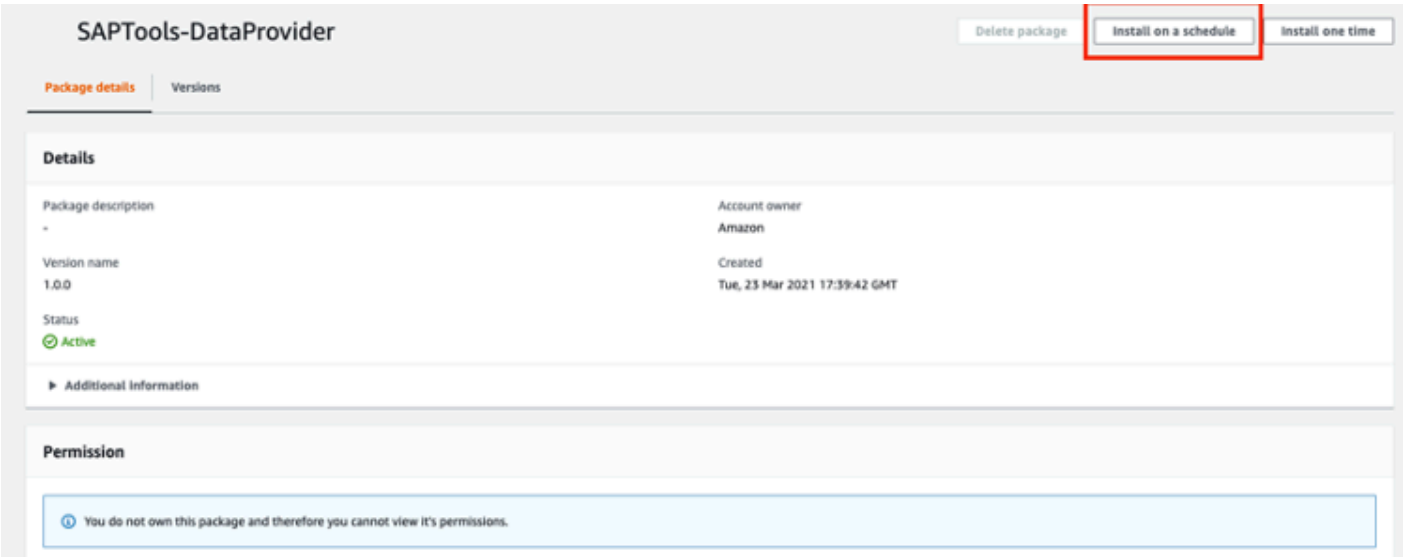
1. Open the Systems Manager console.
2. In the left navigation pane, under the Node Management section, choose **Distributor**.



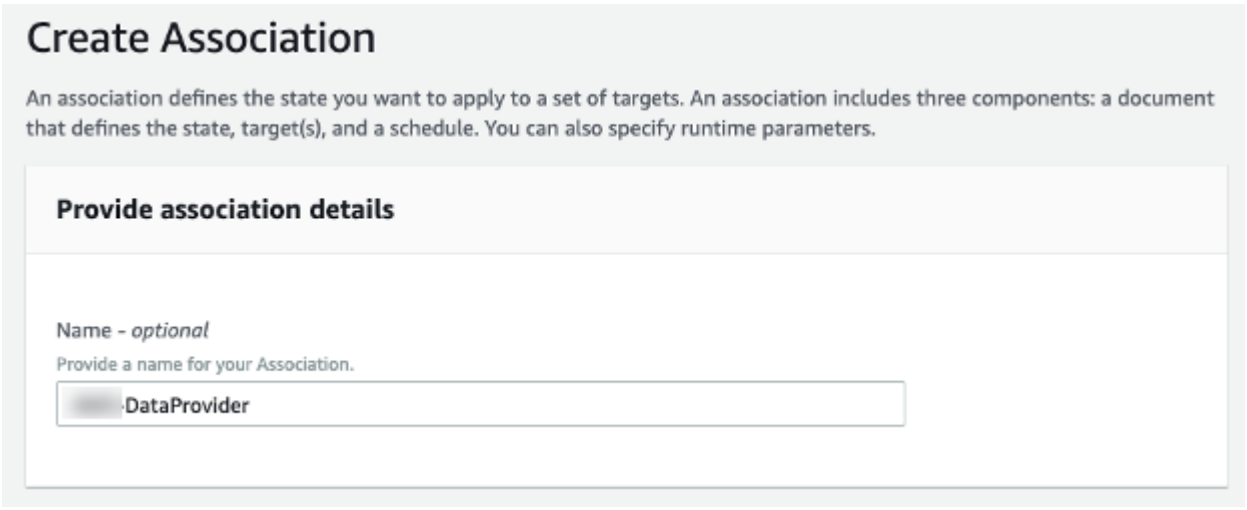
3. In the search bar, type **AWSSAPTools-DataProvider**, and choose the package.



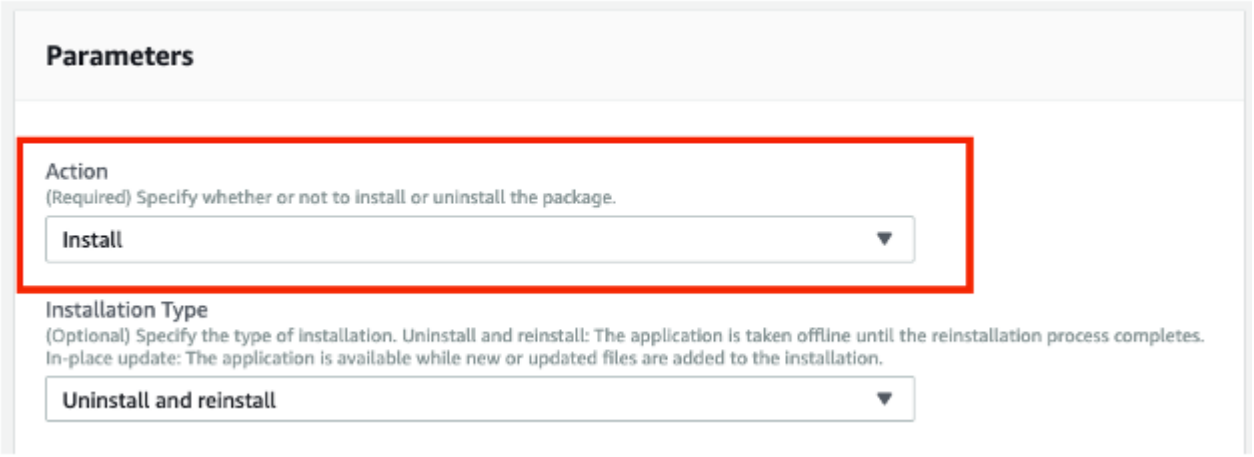
4. To receive auto updates for the DataProvider when there is a new release, choose **Install on a schedule**.



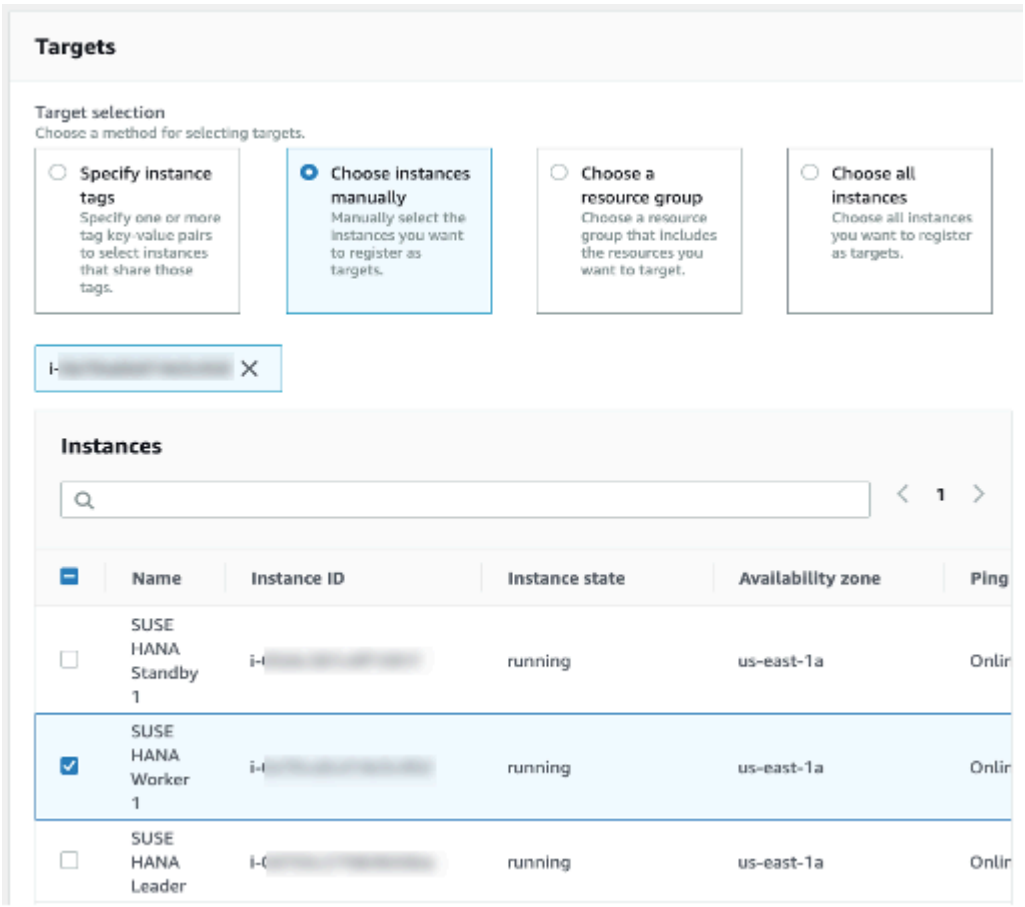
5. On the **Create Association** page, type a **Name** for your association.



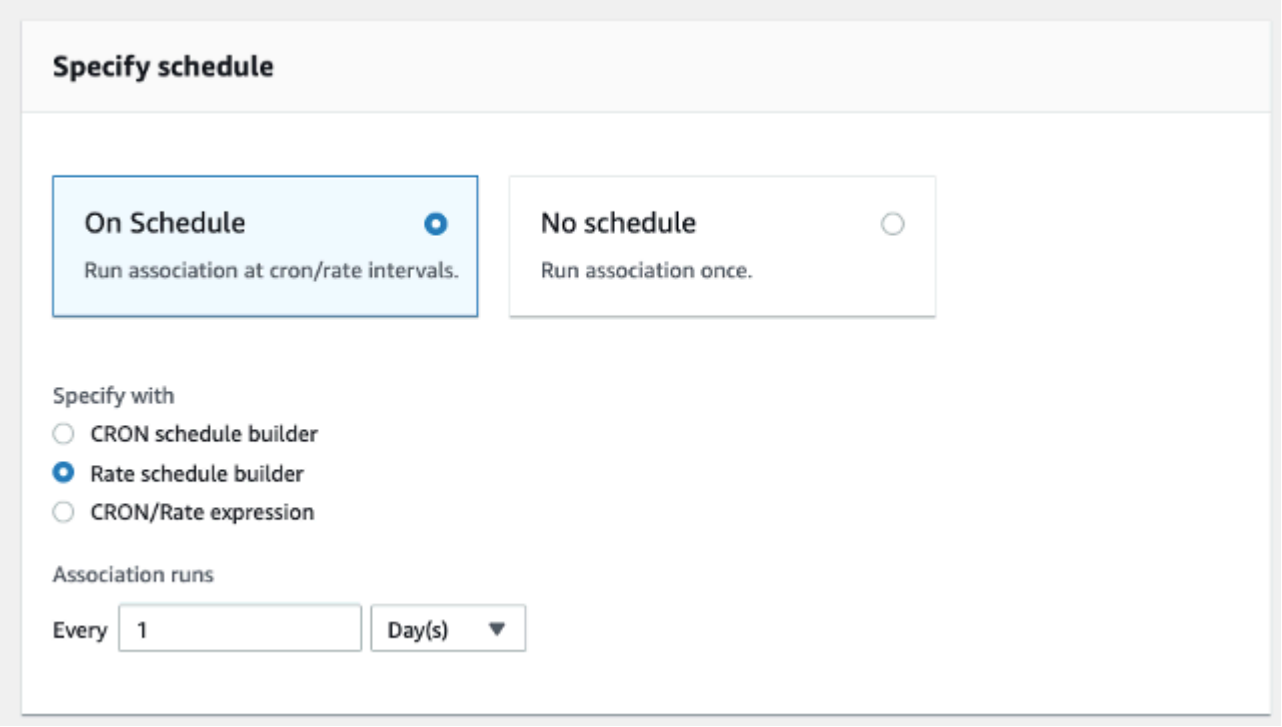
6. In the **Parameters** section, for **Action**, choose **Install**.



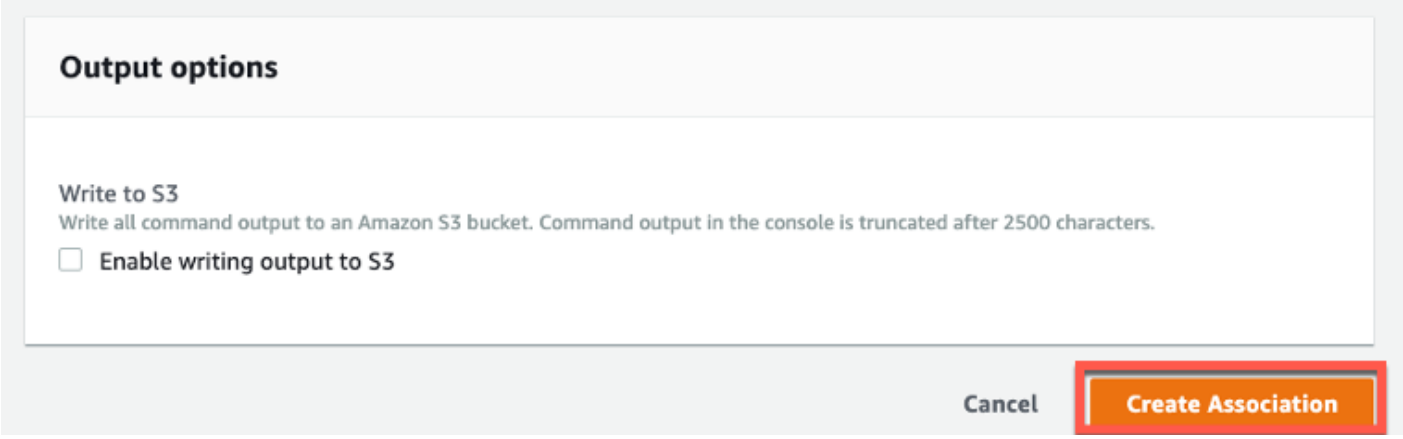
7. In the >**Targets** section, for **Target selection**, select **Choose instances manually**. Then, choose the instances where you want to install the DataProvider.



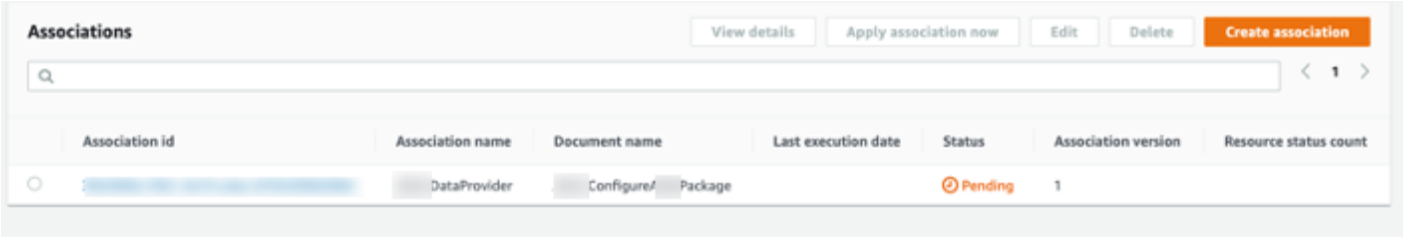
8. In the **Specify schedule** section, make the following selections:
   - Choose **On Schedule**
   - For **Specify with**, choose **Rate schedule builder**.
   - For **Associate runs**, choose **30 days**. (AWS recommends 30 days)



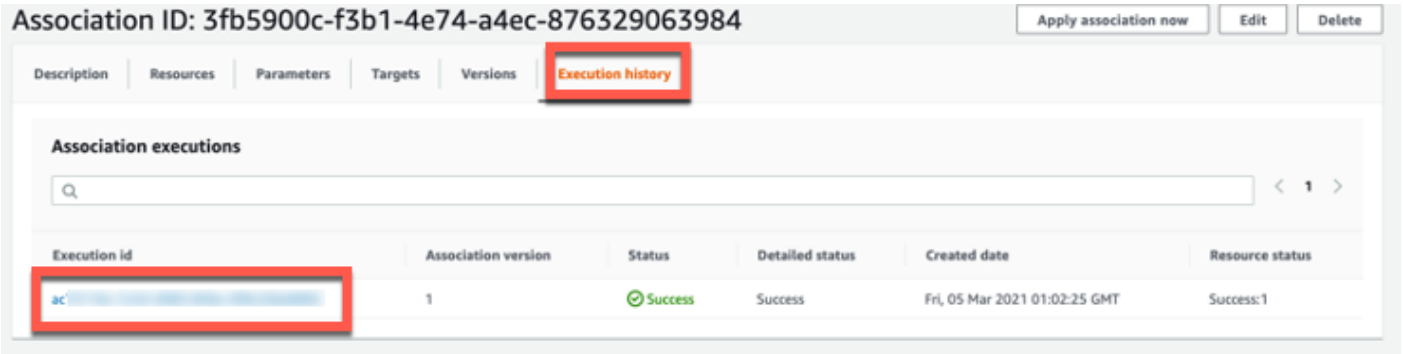
9. In the **Output options** section, choose **Create Association**.



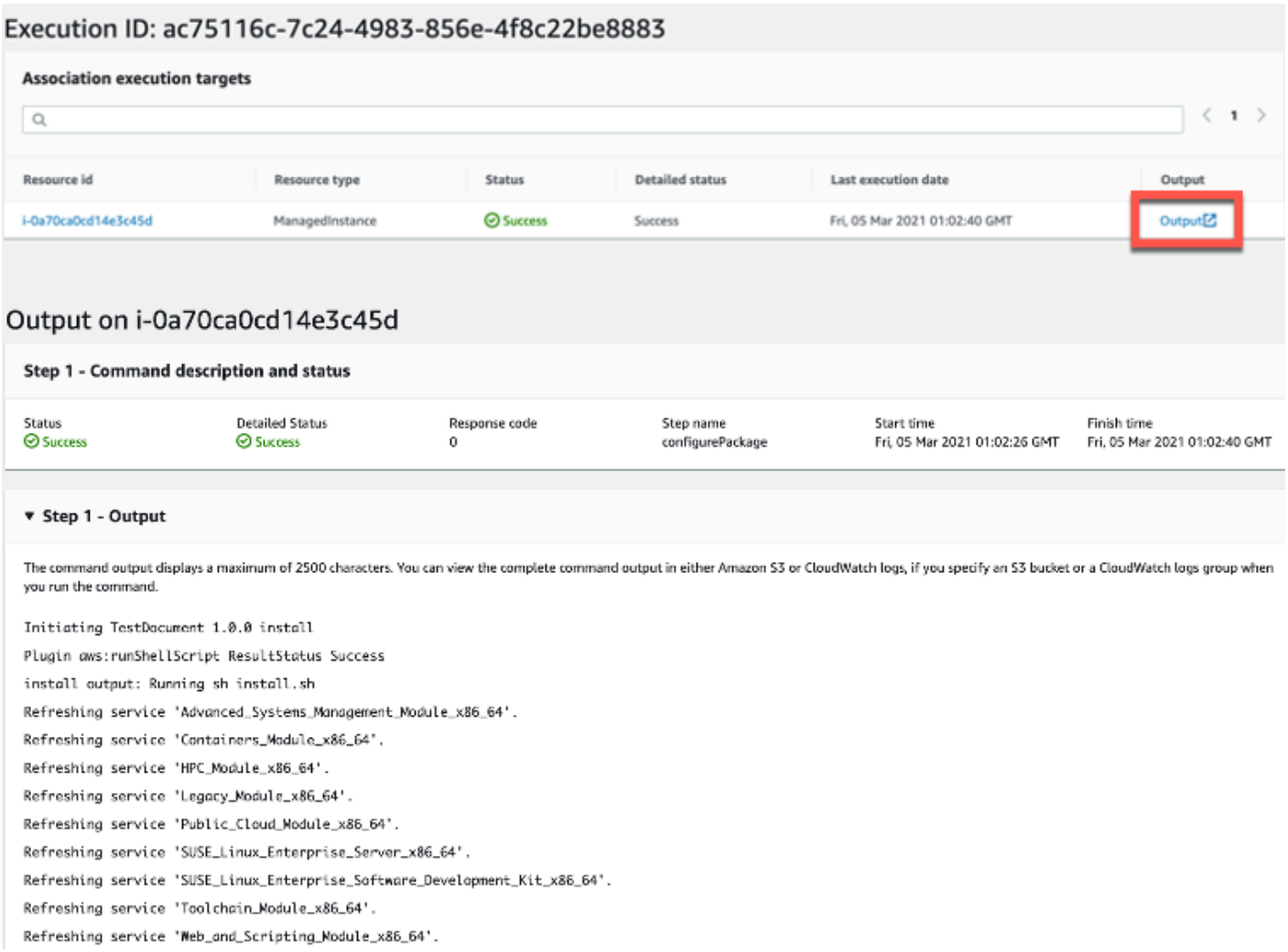
10. Once the association is created, choose the **Association ID**.



11. Choose the **Execution History** tab. Then, choose the Execution id.



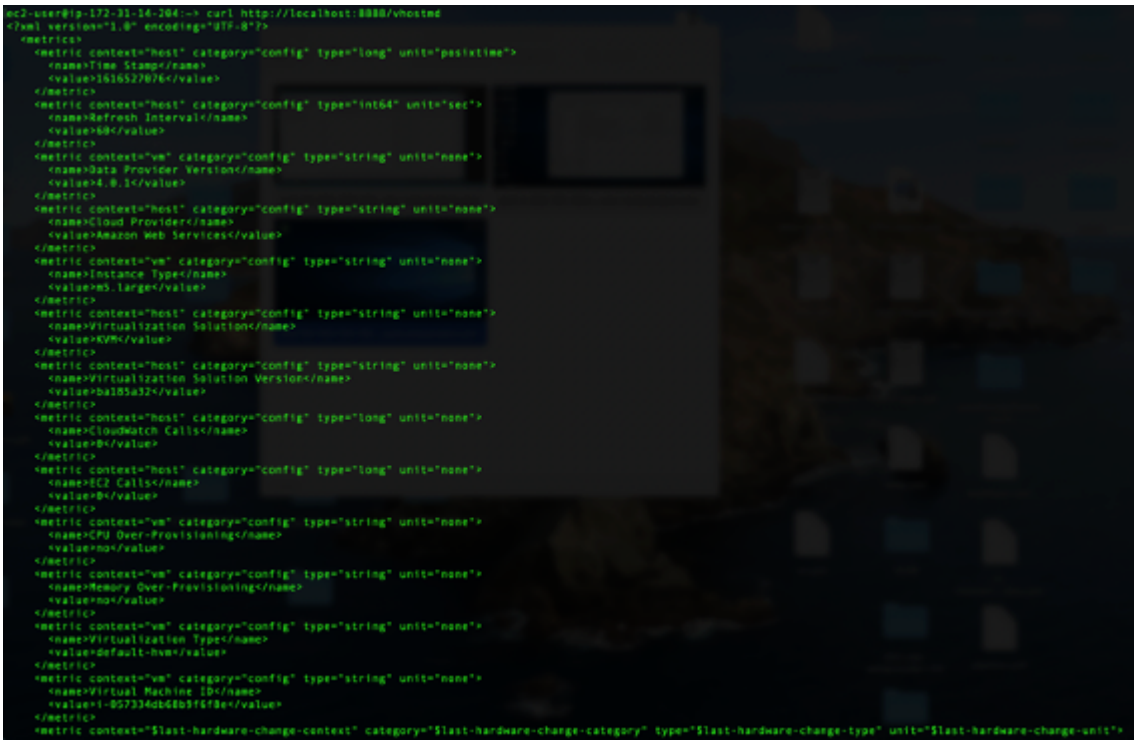
12. On the **Execution ID** page, choose **Output** to see the installation results.



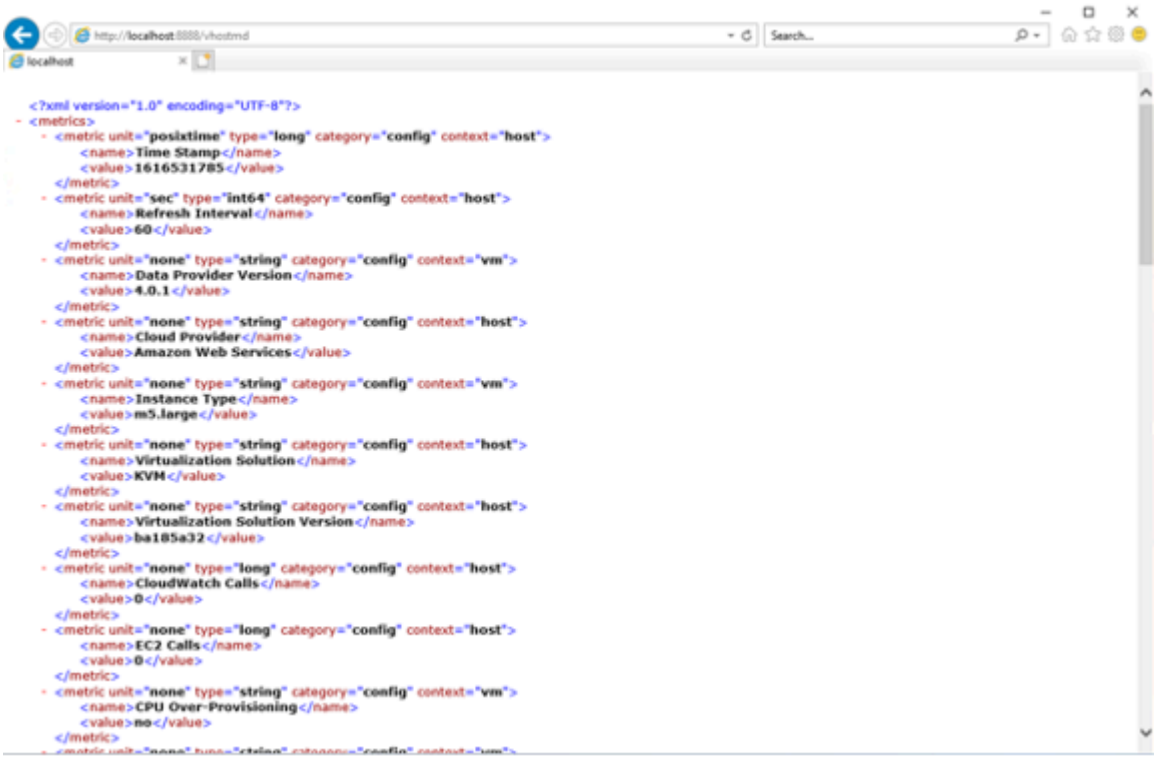

```
The following NEW package is going to be installed:
aws-sap-dataprovider-standalone

The following package has no support information from it's vendor:
aws-sap-dataprovider-standalone

1 new package to install.
Overall download size: 28.0 MiB. Already cached: 0 B. After the operation, additional 28.0 MiB will be used.
Continue? [y/n/...? shows all options] (y): y
Retrieving package aws-sap-dataprovider-standalone-4.0-1.amzn2.x86_64 (1/1), 28.0 MiB ( 28.0 MiB unpacked)

Checking for file conflicts: [...done]
(1/1) Installing: aws-sap-dataprovider-standalone-4.0-1.amzn2.x86_64 [............done]
Additional rpm output:
Will install a SYSTEMD service.

Installing prerequisites.
Done installing prerequisites (SYSTEMD)
*****
Starting the aws-dataprovider service as systemd

*****
Important: Verify log files in /var/log/aws-dataprovider!
*****
Installer completed, exiting.
```

13. Once the installation is completed, log in to the instance, and http://localhost:8888/vhostmd [call the endpoint] to enable the DataProvider to fetch metrics.

- Linux example



- Windows example



Installing with Downloadable Installer – DataProvider 4.3

If you choose to not use SSM to install the DataProvider 4.3, you can manually install the DataProvider using the following procedure.

> **Note**
>
> Before beginning the manual installation you must install the items listed in the Prerequisites section. You don't need to install the `SSM-Agent`. The downloadable DataProvider will not provide auto-updates, to get the latest versions, you must manually check for and download new releases manually.

Download the following file for your environment. By default the files will download in the us-east-1 region, change the default region before downloading if you want the files to download in a different region.

- **Red Hat** https://aws-sap-dataprovider-us-east-1.s3.us-east-1.amazonaws.com/v4/installers/linux/RHEL/aws-sap-dataprovider-rhel-standalone.x86_64.rpm
- **SUSE** https://aws-sap-dataprovider-us-east-1.s3.us-east-1.amazonaws.com/v4/installers/linux/SUSE/aws-sap-dataprovider-sles-standalone.x86_64.rpm
- **Oracle Linux** https://aws-sap-dataprovider-us-east-1.s3.us-east-1.amazonaws.com/v4/installers/linux/ORACLE/aws-sap-dataprovider-oel-standalone.x86_64.rpm
- **Windows** https://aws-sap-dataprovider-us-east-1.s3.us-east-1.amazonaws.com/v4/installers/win/aws-data-provider-installer-win-x64-Standalone.exe
- **GPG Key** GPG Key: https://aws-sap-dataprovider-us-east-1.s3.us-east-1.amazonaws.com/v4/installers/RPM-GPG-KEY-AWS

Installing on Linux

On Linux the data provider is delivered as an RPM package.

SUSE Linux Enterprise Server

To install the AWS Data Provider for SAP on SUSE Linux Enterprise Server (SLES) download the following files:

- **Standard:** aws-sap-dataprovider-sles.x86_64.rpm and GPG Key

- **China:** aws-sap-dataprovider-sles.x86_64.rpm and GPG Key

The files are identical but AWS offers these two location options due to possible connectivity issues when working from China.

To install the data provider run the following commands:

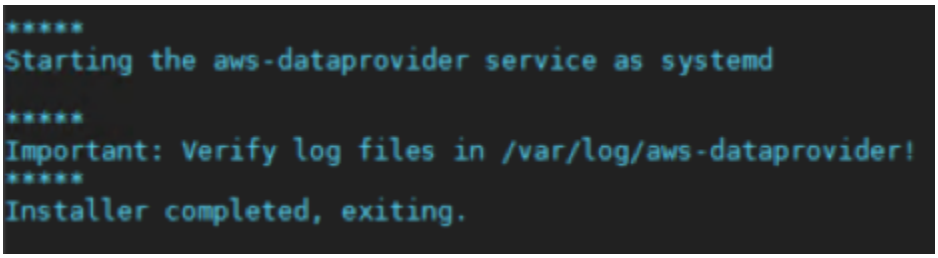
```
wget https://<url to rpm package>
wget https://<url to GPG key>
rpm --import RPM-GPG-KEY-AWS
zypper install -y <rpm package>
```

Example:

```
wget https://aws-sap-dataprovider-us-east-1.s3.us-east-1.amazonaws.com/v4/installers/
linux/SUSE/aws-sap-dataprovider-sles-standalone.x86_64.rpm
wget https://aws-sap-dataprovider-us-east-1.s3.us-east-1.amazonaws.com/v4/installers/
RPM-GPG-KEY-AWS
rpm --import RPM-GPG-KEY-AWS
zypper install -y aws-sap-dataprovider-sles-standalone.x86_64.rpm
```

When the RPM package is installed, the agent starts as a daemon, as seen in the following image.

RPM package installation



Verify that the service is running by calling `netstat -ant` to determine if the listener is running on localhost port 8888.

Verifying installation on Linux

```
p-10-32-59-140:~ # netstat -ant
ctive Internet connections (servers and established)
roto Recv-Q Send-Q Local Address           Foreign Address         State
cp        0      0 [illegible]            0.0.0.0:*               LISTEN
cp        0      0 0.0.0.0:8888           0.0.0.0:*               LISTEN
cp        0      0 127.0.0.1:25           0.0.0.0:*               LISTEN
cp        0      0 0.0.0.0:111            0.0.0.0:*               LISTEN
cp        0     52 10.32.59.140:22        69.120.21.212:49759     ESTABLISHE
p-10-32-59-140:~ #
```

You should also view the log files at `/var/log/aws-dataprovider/messages.0` to ensure the daemon has the appropriate connectivity and authorization to access the required metrics.

Verifying connectivity and authorization on Linux

```
I 0001C Agent has started with version 4.2.0 : Tue Aug 25 18:13:00 UTC 2020

I 0000C Agent log level has been set to INFO

I 10001 Read in jar configuration; read on board configuration

I 10002 Is there a local, optional configuration file at /usr/local/ec2/aws-dataprovider/config.properties ?

I 10003 Read local, optional configuration file from /usr/local/ec2/aws-dataprovider/config.properties

I 03002 vhostmd agent is listening on localhost

I 0000D Agent is starting the vhostmd provider

I 08001  ** Running Diagnostics **

I 08006 Diagnostic : Amazon CloudWatch Connectivity & Access

I 08009 Diagnostic : Passed

I 0800A Diagnostic : EC2 API Connectivity & Access

I 0800D Diagnostic : Passed

I 08010 Diagnostic: Data Collector API

I 08011 Diagnostic : Passed

I 0800E ** Diagnostics Complete **
```

At startup, the monitoring agent runs three sets of diagnostics:

- The AWS connectivity diagnostic ensures network connectivity to Amazon S3 for obtaining automatic updates to the AWS Data Provider for SAP.
- The second diagnostic tests for authorization to access CloudWatch. This authorization requires assigning an IAM role to the Amazon EC2 instance you are running on with an IAM policy that allows access to CloudWatch. For details, see IAM Roles, earlier in this guide.

- The third diagnostic tests for authorization to access Amazon EC2, which also requires an IAM role associated with the Amazon EC2 instance.

The AWS Data Provider for SAP is designed to run with or without connectivity, but you can't obtain updates without connectivity. Amazon CloudWatch and Amazon EC2 will return blank values if you don't have the proper authorizations in place.

You can also call the AWS Data Provider for SAP directly to view the metrics. Calling `wget http://localhost:8888/vhostmd` returns a file of metrics. You can look inside the file to see the metrics that were returned, as shown here.

Viewing metrics on Linux

```
ip-10-32-59-140:~ # wget http://localhost:8888/vhostmd
--2012-10-09 17:10:28--  http://localhost:8888/vhostmd
Resolving localhost... 127.0.0.1, ::1
Connecting to localhost|127.0.0.1|:8888... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/xml]
Saving to: `vhostmd'

    [ <=>                                                   ] 7,58

2012-10-09 17:10:31 (385 MB/s) - `vhostmd' saved [7589]

ip-10-32-59-140:~ # cat vhostmd
<?xml version="1.0" encoding="UTF-8"?>
  <metrics>
    <metric context="host" category="config" type="long" unit="posixtime"
      <name>Time Stamp</name>
      <value>1349802629284</value>
    </metric>
    <metric context="host" category="config" type="int64" unit="sec">
      <name>Refresh Interval</name>
      <value>60</value>
    </metric>
    <metric context="host" category="config" type="string" unit="none">
      <name>Cloud Provider</name>
      <value>Amazon Web Services</value>
    </metric>
    <metric context="host" category="config" type="string" unit="none">
      <name>Instance Type</name>
      <value>m1.large</value>
    </metric>
```

The AWS Data Provider for SAP now starts automatically each time the operating system starts. You can also manually stop and restart the AWS Data Provider for SAP with the following command, which depends on your operating system version:

- SLES 11, Oracle Linux 6, and Red Hat Linux 6:

```
service aws-dataprovider [start|stop]
```

- SLES 12, SLES 15, Oracle Linux 7, Oracle Linux 8, Red Hat Linux 7, and Red Hat Linux 8.

```
systemctl [start|stop] aws-dataprovider
```

You can configure AWS Data Provider to use proxies if you do not have transparent HTTP/HTTPS access to the internet.

1. Stop AWS Data Provider for SAP.
2. Enter proxy information in the file (as seen below) located at `/usr/local/ec2/aws-dataprovider/proxy.properties`.

```
#proxy.properties
#used to set web proxy settings for the {aws} Data Provider for SAP
#Https is the only supported proxy method
#Blank values for everything means no proxy set

https.proxyHost=
https.proxyPort=
https.proxyDomain=
https.proxyUsername=
https.proxyPassword=
```

3. Start AWS Data Provider for SAP.

Installing on Red Hat and Oracle Enterprise Linux

For Red Hat and Oracle Enterprise Linux, the installation steps are the same as described for SLES above but the RPM file and command to install the RPM package differs.

- **Red Hat**

Default: aws-sap-dataprovider-rhel.x86_64.rpm

- **Oracle Enterprise Linux**

  Default: aws-sap-dataprovider-oel.x86_64.rpm

To install the data provider run the following commands:

```
wget https://<url to rpm package>
yum -y install <rpm package>
```

Example:

```
wget https://aws-sap-data-provider.s3.amazonaws.com/Installers/aws-sap-dataprovider-
rhel.x86_64.rpm
yum -y install aws-sap-dataprovider-rhel.x86_64.rpm
```

Installing on Windows

On Windows, the installer is delivered in the form of an NSIS (Nullsoft Scriptable Install System) executable.

1. Open a web browser and download the installer:
   - **Default:** aws-data-provider-installer-win-x64.exe
2. Run the downloaded **exe** file.
3. Verify the installation.
   - Once the installation is complete, you can see the file in `C:\Program Files\Amazon \DataProvider` directory.
   - The installation also creates and starts a Windows service called **AWS Data Provider for SAP**.
   - Verify that the service is running by entering http://localhost:8888/vhostmd in a web browser. The page returns metrics from AWS Data Provider for SAP if your installation is successful.
4. You can configure AWS Data Provider to use proxies if you do not have transparent HTTP/HTTPS access to the internet.

   a. Stop AWS Data Provider for SAP.

b. Enter proxy information in the file (as seen below) located at `C:\Program Files\Amazon \DataProvider\proxy.properties`.

```
#proxy.properties
#used to set web proxy settings for the {aws} Data Provider for SAP
#Https is the only supported proxy method
#Blank values for everything means no proxy set

https.proxyHost=
https.proxyPort=
https.proxyDomain=
https.proxyUsername=
https.proxyPassword=
```

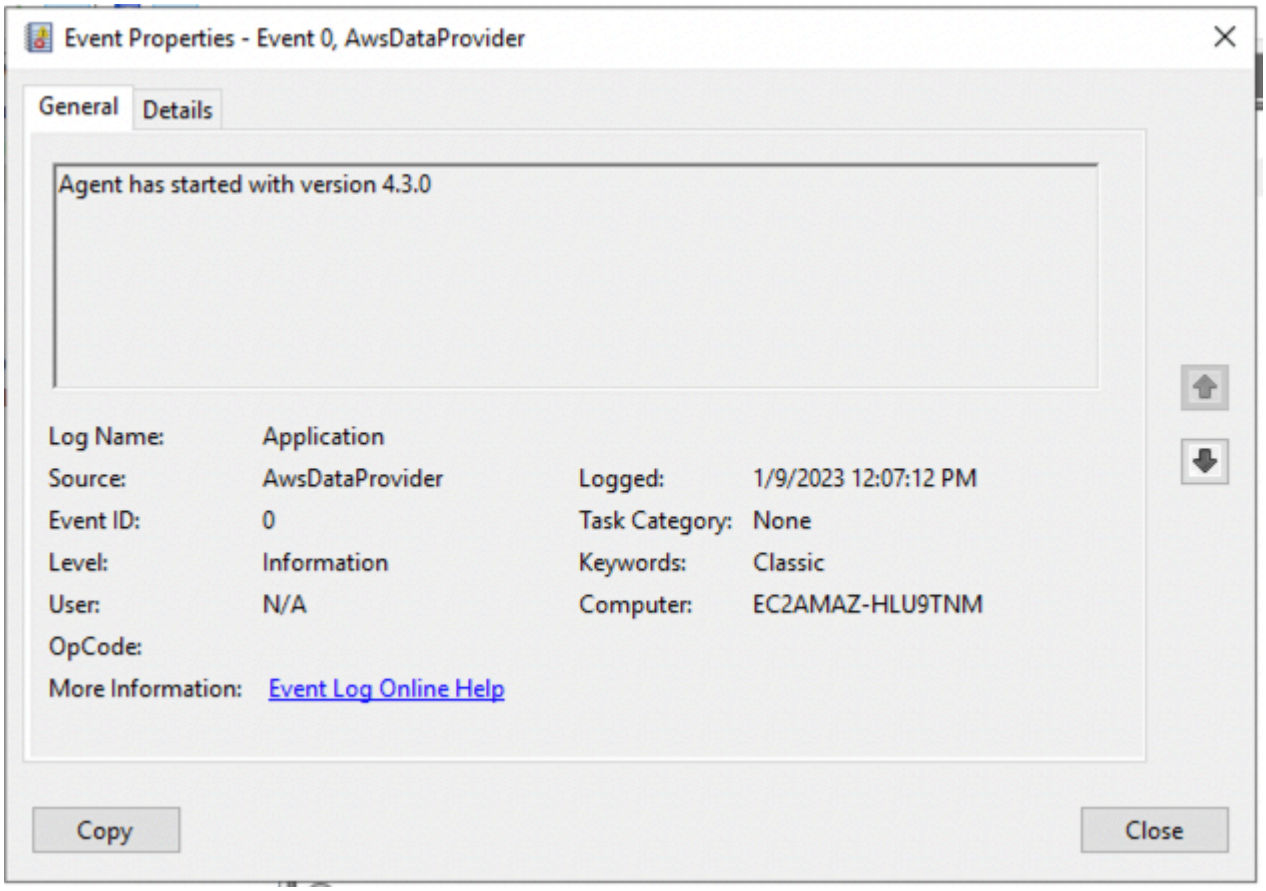
c. Start AWS Data Provider for SAP.

5. Verify that the service is running by calling `netstat -ant` from a command window or from a Windows PowerShell script to determine if the listener is running on localhost port 8888.

```
PS C:\Users\Administrator\Desktop> netstat -ant

Active Connections

  Proto  Local Address          Foreign Address        State           Offload S
  TCP    0.0.0.0:135            0.0.0.0:0              LISTENING       InHost
  TCP    0.0.0.0:445            0.0.0.0:0              LISTENING       InHost
  TCP    0.0.0.0:3389           0.0.0.0:0              LISTENING       InHost
  TCP    0.0.0.0:5985           0.0.0.0:0              LISTENING       InHost
  TCP    0.0.0.0:8888           0.0.0.0:0              LISTENING       InHost
  TCP    0.0.0.0:47001          0.0.0.0:0              LISTENING       InHost
  TCP    0.0.0.0:49152          0.0.0.0:0              LISTENING       InHost
  TCP    0.0.0.0:49153          0.0.0.0:0              LISTENING       InHost
  TCP    0.0.0.0:49154          0.0.0.0:0              LISTENING       InHost
  TCP    0.0.0.0:49155          0.0.0.0:0              LISTENING       InHost
  TCP    0.0.0.0:49156          0.0.0.0:0              LISTENING       InHost
  TCP    10.191.175.27:139      0.0.0.0:0              LISTENING       InHost
  TCP    10.191.175.27:3389     69.120.21.212:50796    ESTABLISHED     InHost
  TCP    10.191.175.27:49511    23.63.240.60:443       CLOSE_WAIT      InHost
  TCP    10.191.175.27:49555    74.125.228.104:80      TIME_WAIT       InHost
  TCP    10.191.175.27:49556    74.125.228.103:80      ESTABLISHED     InHost
  TCP    10.191.175.27:49558    169.254.169.254:80     CLOSE_WAIT      InHost
```

Verifying the installation on Windows

6. Navigate to the Windows event log, and find the application log for startup events from the AWS Data Provider for SAP. Check the diagnostics.



Checking diagnostics on Windows

At startup, the monitoring agent runs three sets of diagnostics:

- The AWS connectivity diagnostic ensures network connectivity to Amazon S3 for obtaining automatic updates to the AWS Data Provider for SAP.
- The second diagnostic tests for authorization to access CloudWatch, which requires assigning an IAM role to the EC2 instance you are running on with an IAM policy that allows access to CloudWatch. For details, see IAM Roles, earlier in this guide.
- The third diagnostic tests for authorization to access Amazon EC2, which also requires an IAM role associated with the Amazon EC2 instance.

The AWS Data Provider for SAP is designed to run with or without connectivity, but you can't obtain updates without connectivity. If you don't have the proper authorizations in place, Amazon CloudWatch and Amazon EC2 return blank values.

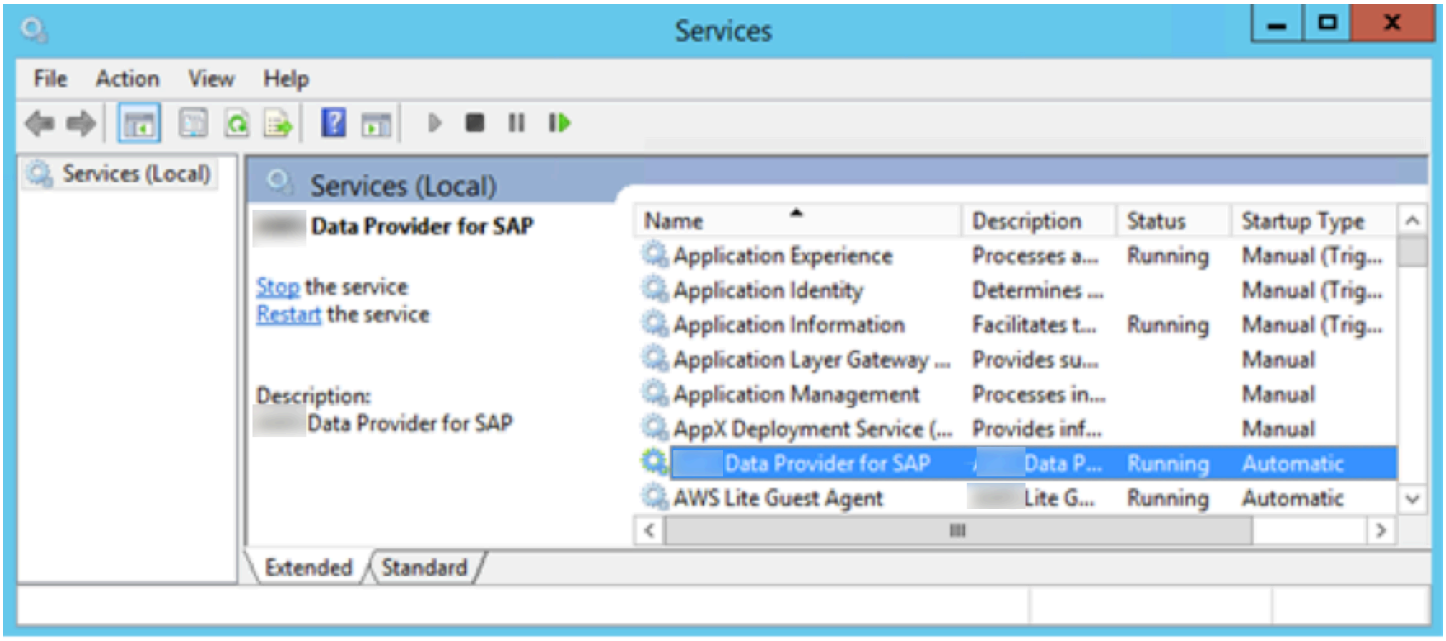
You can also call the AWS Data Provider for SAP directly from your web browser to view metrics, as shown in the figure.

Viewing metrics on Windows


```
– <metrics>
  – <metric context="host" category="config" type="long" unit="posixtime">
      <name>Time Stamp</name>
      <value>1349813196225</value>
    </metric>
  – <metric context="host" category="config" type="int64" unit="sec">
      <name>Refresh Interval</name>
      <value>60</value>
    </metric>
  – <metric context="host" category="config" type="string" unit="none">
      <name>Cloud Provider</name>
      <value>Amazon Web Services</value>
    </metric>
  – <metric context="host" category="config" type="string" unit="none">
      <name>Instance Type</name>
      <value>m1.xlarge</value>
    </metric>
```


AWS Data Provider for SAP now starts automatically each time the operating system starts. You can also manually stop and restart the AWS Data Provider for SAP, just as you would stop and restart any other Windows service.

Stopping and restarting the AWS Data Provider for SAP on Windows



In order to configure proxy settings, you can place a customized `proxy.properties` file in Window's temp directory, which is designated by the windows system variable %TEMP%.

Subscribe to AWS Data Provider Agent for notifications

Amazon Simple Notification Service can notify you when new versions of AWS Data Provider Agent are released. Use the following steps to setup this subscription.

1. Open https://console.aws.amazon.com/sns/v3/home.
2. Ensure that you are in **US East N. Virginia** (us-east-1) Region.
3. In the left navigation pane, select **Subscriptions** > **Create subscription**.
4. Add a **Topic ARN** based on the AWS Region in which you are using AWS Data Provider Agent.

Region	ARN
Default	`arn:aws:sns:us-east-1:804845276281:AWS-DataProvider-SAP-Update`
AWS GovCloud (US-West) and AWS GovCloud (US-East)	`arn:aws-us-gov:sns:us-gov-west-1:140982767562:AWS-DataProvider-SAP-Update`

Region	ARN
China (Beijing) Region and China (Ningxia) Region	`arn:aws-cn:sns:cn-north-1:001645243879:AWS-DataProvider-SAP-Update`

5. **Protocol** – choose Email or SMS.
   - **Email** – enter an email address where you would like to receive the notification in the **Endpoint** field.

     > **Note**
     >
     > To enable email notifications, you must confirm your email subscription by following the instructions you receive on the provided email address.

   - **SMS** – enter a phone number where you would like to receive the notification in the **Endpoint** field.
6. Choose **Create subscription**. You can now receive notifications whenever a new version of AWS Data Provider Agent is released.

To unsubscribe from notifications, use the following steps.

1. Open https://console.aws.amazon.com/sns/v3/home.
2. In the left navigation pane, select **Subscriptions**.
3. Select the subscription from your list of subscriptions and choose **Delete**.

Updating to DataProvider 4.3

If you have previously installed DataProvider 2.0 or 3.0 and want to update to DataProvider 4.3, you need to uninstall the running version first, and then install DataProvider 4.3.

Updating to DataProvider 4.3 using the SSM Distributor package

When you install DataProvider 4.3 through the SSM distributor, it will auto-update the installed package on a new version release. For more information, see Install or update packages.

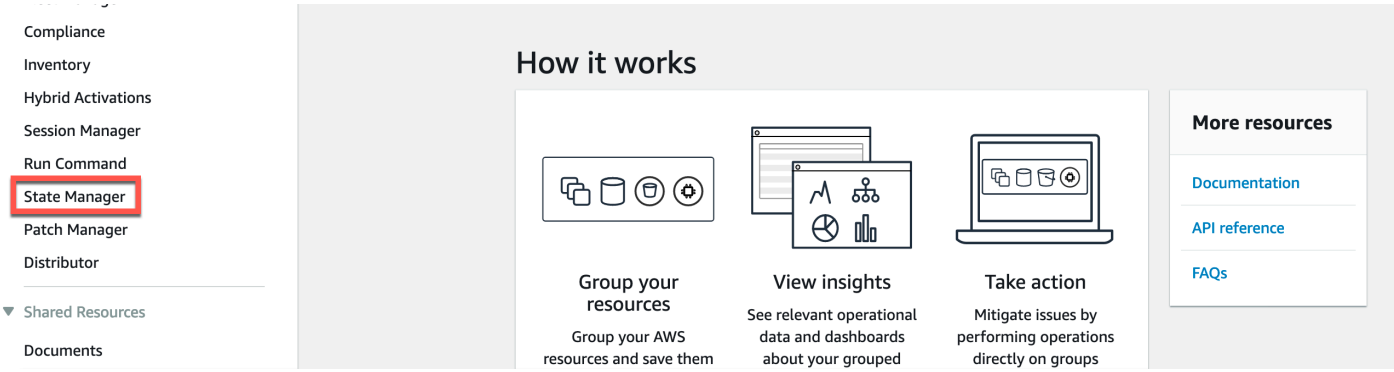
To update the DataProvider 4.3 manually, you must first uninstall the running version and then install the updated version.

DataProvider 4.3 does *not* support the following actions:

- Manual update of the RPM package if the DataProvider has been installed using the SSM distributor.
- Automatic update through the SSM distributor if the DataProvider has been manually installed using RPM.
- Manual update of the RPM package in any scenario.

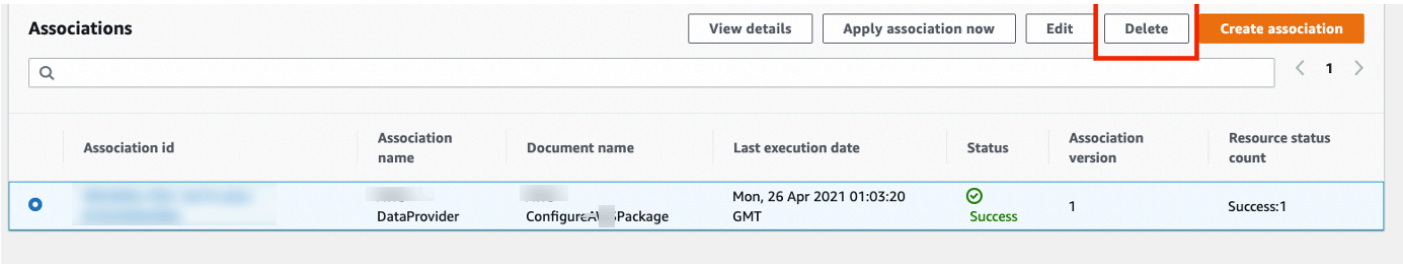
Uninstall DataProvider 4.3 using the SSM distributor

1. Open the AWS Systems Manager console, on the left navigation pane, choose **State Manager**.

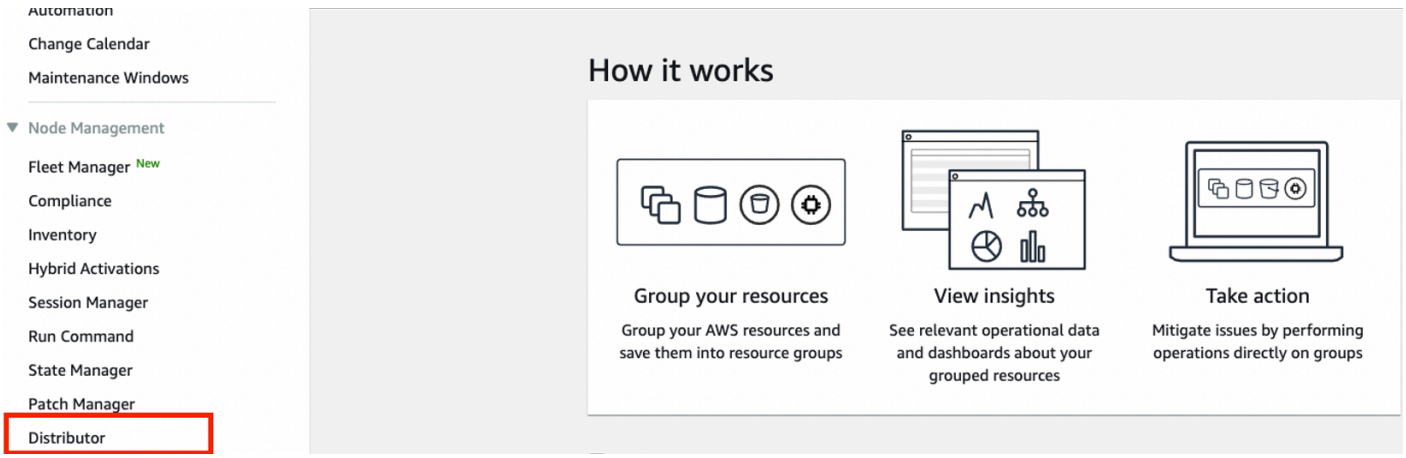


2. On the **Associations** page, and choose the **Association id**. Then, choose **Delete**.

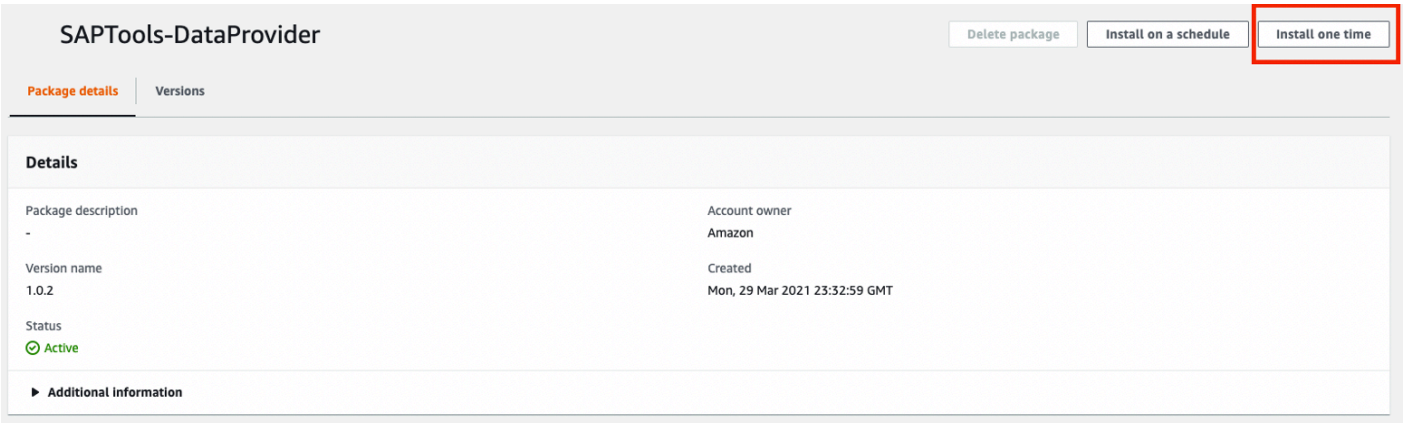
   After the delete is successful, the auto-update stops.



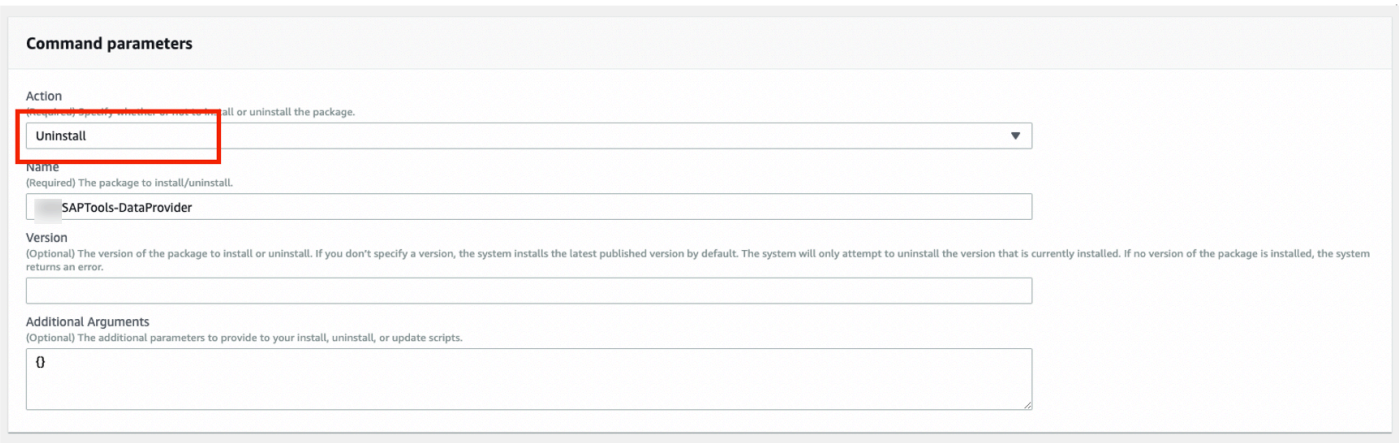
3. On the main page, in the left navigation page, choose **Distributor**.



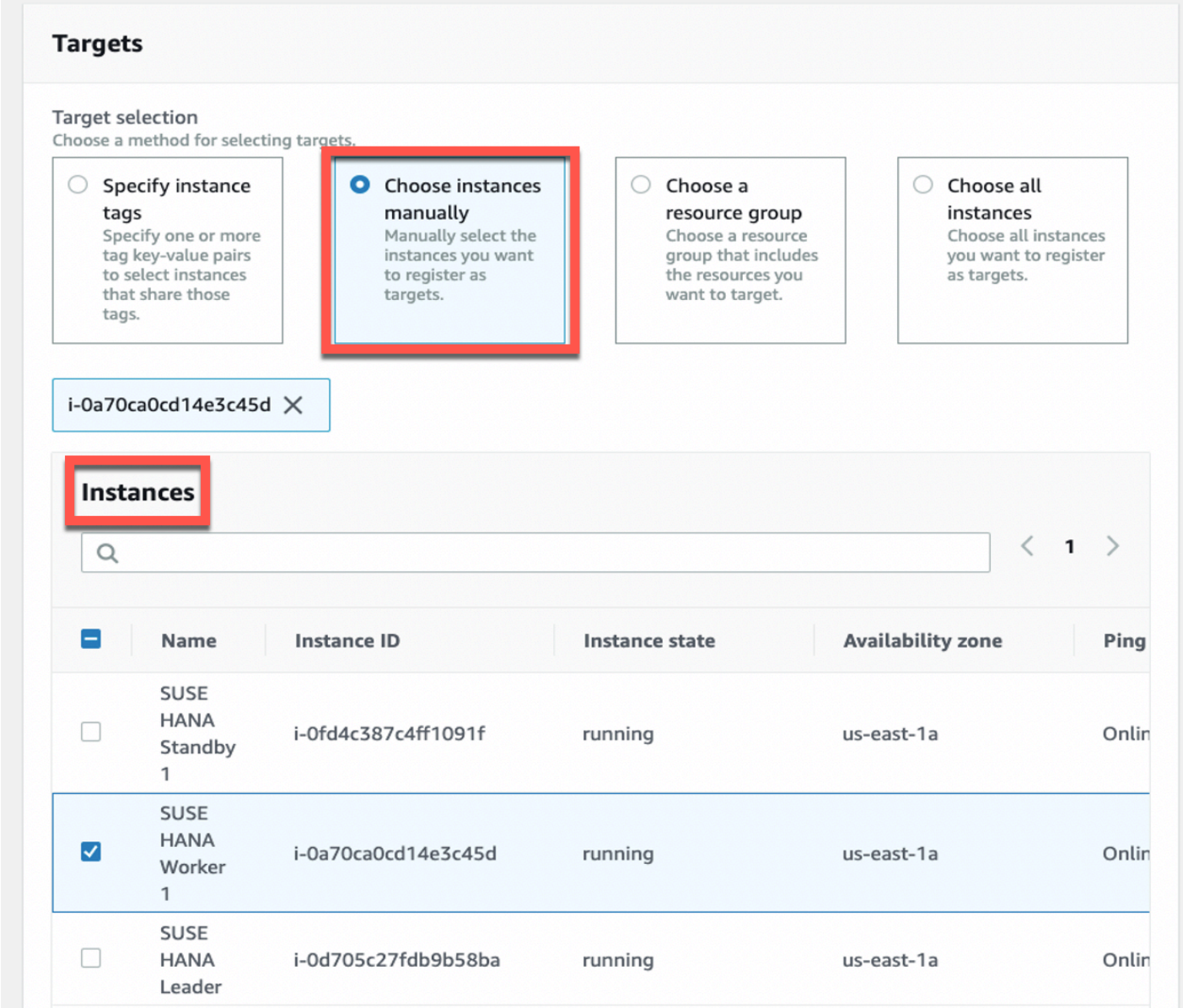
4. Choose the **AWSSAPTools-DataProvider** distributor package, and choose **Install one time**.



5. In the **Command parameters** section, choose **Uninstall**.



6. In the **Targets** section, select **Choose instances manually**. Then choose the **instance** to uninstall the DataProvider.



7. Choose **Run** to begin the uninstall.

Uninstall DataProvider 4.3 manually

RedHat

```
yum erase aws-dataprovider-standalone
```

SUSE

```
zypper rm aws-dataprovider-standalone
```

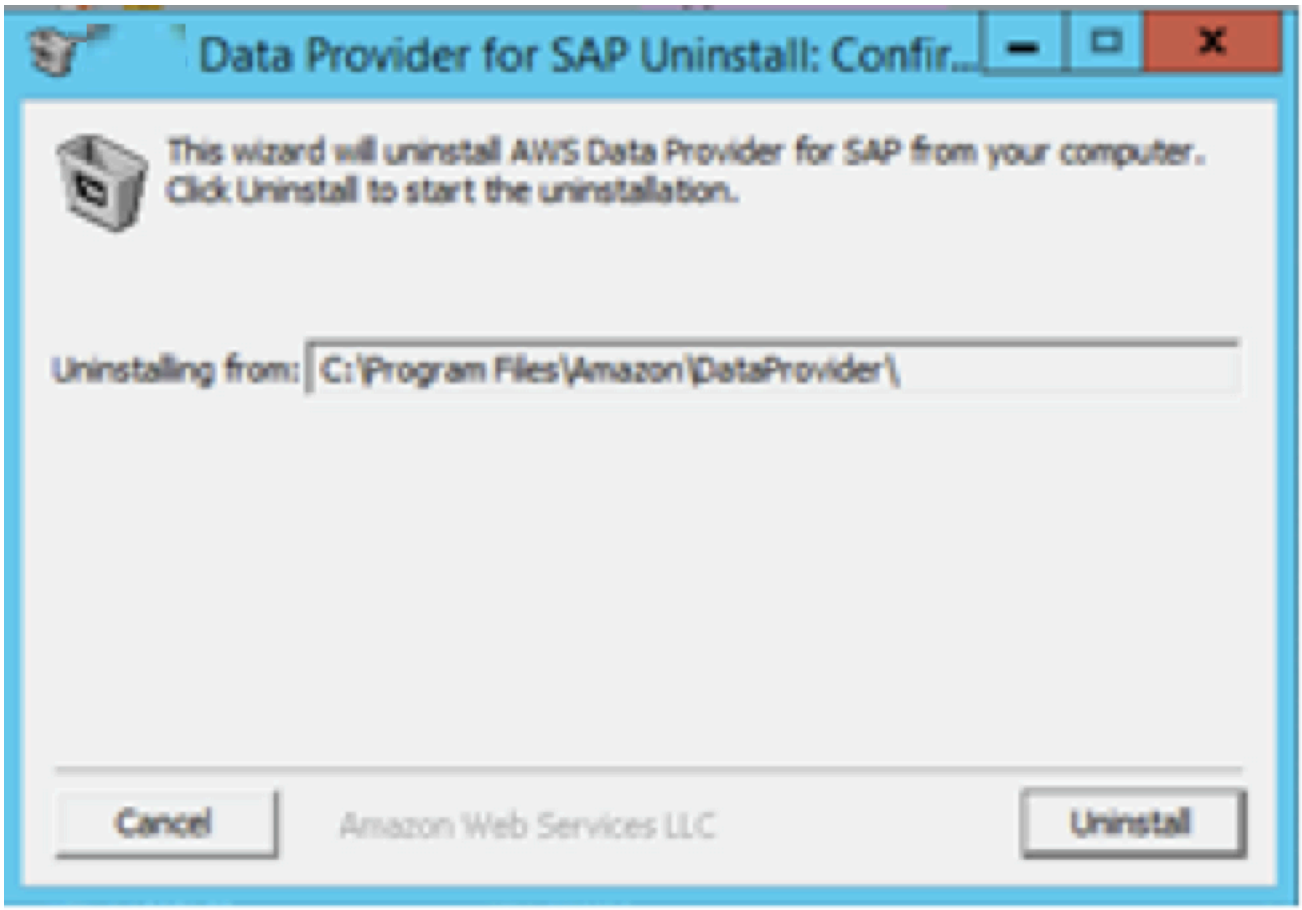
Windows

1. Run the uninstaller.

```
C:\Program Files\AmazonA\DataProvider\uninstall.exe
```

2. When prompted, choose **Uninstall**.

Uninstalling the AWS Data Provider for SAP on Windows



Uninstalling older versions

Uninstalling the AWS Data Provider for SAP does not require SAP downtime and can be done online. The only impact will be a gap in metric monitoring information for the time that the DataProvider was installed on your system.

Uninstall DataProvider 3.0

Linux

1. Log in to Linux as a superuser, like root.
2. Stop and remove the DataProvider using the following command.

SLES

```
zypper remove -y aws-sap-dataprovider
```

RHEL/OEL

```
yum -y erase aws-sap-dataprovider
```

Windows

```
“C:\Program Files\Amazon\DataProvider\uninstall.exe“
```

Uninstall DataProvider 2.0

Linux

1. Log in to Linux as a superuser, like root.
2. Stop and remove the DataProvider using the following command.

```
/usr/local/ec2/aws-agent/bin/aws-agent_uninstall
```

Windows

```
“C:\Program Files\Amazon\DataProvider\uninstall.exe“
```

Troubleshooting

This section provides help to analyze installation problems.

Troubleshooting on Linux

Problem: The installation failed, and I'm not sure if my files are in a consistent state.

Stop and remove the Data Provider with the following command.

SLES:

```
zypper remove -y aws-sap-dataprovider
```

RHEL / OEL:

```
yum -y erase aws-sap-dataprovider
```

Problem: The AWS Data Provider for SAP failed to start at the end of the installation process.

Check the log files in `/var/log/aws-dataprovider` for hints on what is not going as expected. If needed, uninstall and reinstall the Data Provider. If reinstalling the AWS Data Provider for SAP doesn't solve the problem, you can gather debug information about the AWS Data Provider for SAP by editing the `/usr/local/ec2/aws-dataprovider/bin/aws-dataprovider` file.

Debugging the installation on Linux

Now if you run service `aws-dataprovider-start` or `systemctl start aws-dataprovider`, you will get a lot of debugging output that might help you diagnose the root cause of the problem.

Debugging information on Linux

```
Java VM created successfully
Class org/apache/commons/daemon/support/DaemonLoader found
Native methods registered
java_init done
Daemon loading...
Daemon loaded successfully
java_load done
18:05:28.561 I 06002 Starting the data collector engine

18:05:29.230 I 00001 ****** AWS SAP Data Collector Agent is Starting ******

18:05:29.237 I 00002 Software Version: 1.0.47 : Fri Sep 21 22:22:00 UTC 20

18:05:29.239 I 0000C Agent log level has been set to FINE

18:05:29.241 I 08001  ** Running Diagnostics **

18:05:29.241 I 08002 Diagnostic : AWS Connectivity

18:05:30.078 I 08005 Diagnostic : Passed

18:05:30.078 I 08006 Diagnostic : Amazon CloudWatch Connectivity & Access

18:05:30.803 I 08009 Diagnostic : Passed

18:05:30.803 I 0800A Diagnostic : EC2 API Connectivity & Access

18:05:31.082 I 0800D Diagnostic : Passed

18:05:31.083 I 0800E ** Diagnostics Complete **

18:05:31.282 I 03002 vhostmd agent is listening on localhost

18:05:31.282 I 0000D Agent is starting the vhostmd provider

18:05:31.282 I 00003 ************************************************
```

Problem: When I looked at my logs I noticed that my installation failed all diagnostics.

Symptoms of internet connectivity problems on Linux

```
14:32:15.862 I 08001  ** Running Diagnostics **

14:32:15.862 I 08002 Diagnostic : AWS Connectivity

14:33:19.362 W 08003 Diagnostic : Failed

14:33:19.362 I 08006 Diagnostic : Amazon CloudWatch Connectivity & Access

14:33:19.515 W 08007 Diagnostic : Failed

14:33:19.516 I 0800A Diagnostic : EC2 API Connectivity & Access

14:33:19.542 W 0800B Diagnostic : Failed

14:33:19.542 I 0800E ** Diagnostics Complete **
```

Failing *all* diagnostics indicates that there's a problem with your outbound connection to the internet. You can confirm this by pinging a well-known internet location, like www.amazon.com. The most common cause of routing issues is in the VPC network configuration, which needs to

have either an internet gateway in place or a VPN connection to your data center with a route to the internet. For details, see Amazon VPC Network Topologies, earlier in this guide.

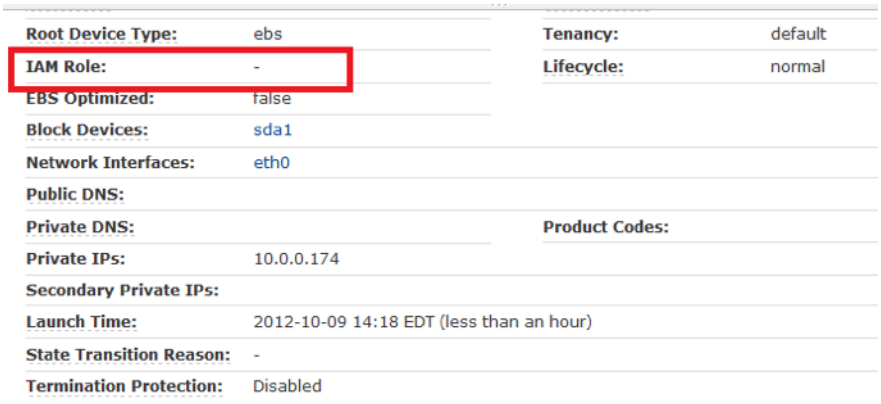
Problem: When I looked at my logs I noticed that I don't have access to CloudWatch and Amazon EC2, but I did pass the first diagnostic for AWS connectivity.

Symptoms of authorization issues on Linux

```
14:38:57.467 I 08001  ** Running Diagnostics **

14:38:57.468 I 08002 Diagnostic : AWS Connectivity

14:38:58.182 I 08005 Diagnostic : Passed

14:38:58.182 I 08006 Diagnostic : Amazon CloudWatch Connectivity & Access

14:38:58.325 W 08007 Diagnostic : Failed

14:38:58.325 I 0800A Diagnostic : EC2 API Connectivity & Access

14:38:58.357 W 0800B Diagnostic : Failed

14:38:58.357 I 0800E ** Diagnostics Complete **
```

This is a clear indicator that you have an authorization issue when trying to access CloudWatch and Amazon EC2. The common cause for this problem is not having an IAM role associated with your instance that contains the IAM policy, as specified in IAM Roles, earlier in this guide. You can quickly diagnose this issue by looking at the Amazon EC2 instance in question in the Amazon EC2 console and verifying the IAM role.

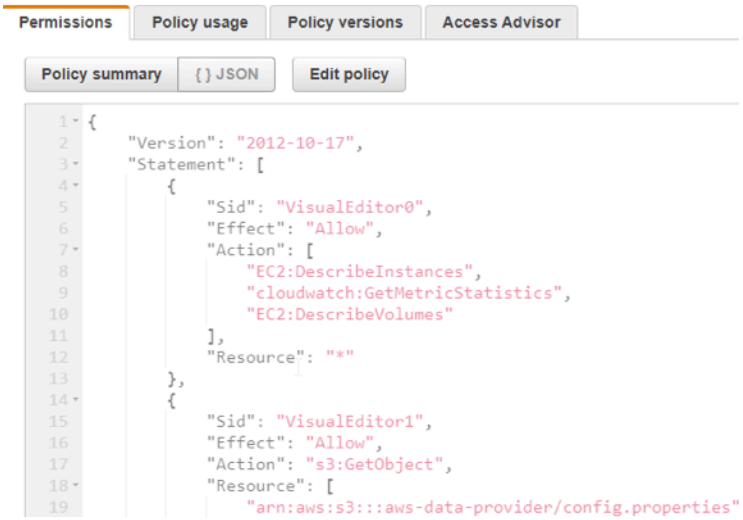
Verifying the IAM role for an EC2 instance



If the IAM role doesn't exist, then create it as specified in IAM Roles, which is described earlier in this guide.

If you do have an IAM role assigned to the instance, go to the IAM console, select the IAM role name, and then expand the policy. Verify that you have the required policy that is specified in IAM Roles, earlier in this guide.

Verifying the policy for the IAM role



Problem: I want to configure/update JAVA_HOME for Data Provider.

Open the `/usr/local/ec2/aws-dataprovider/env` file and update the `JAVA_HOME` variable. Restart Data Provider after the update, with the following command.

```
sudo systemctl daemon-reload
sudo systemctl start aws-dataprovider
```

Troubleshooting on Windows

Problem: The installation failed, and I'm not sure if my files are in a consistent state.

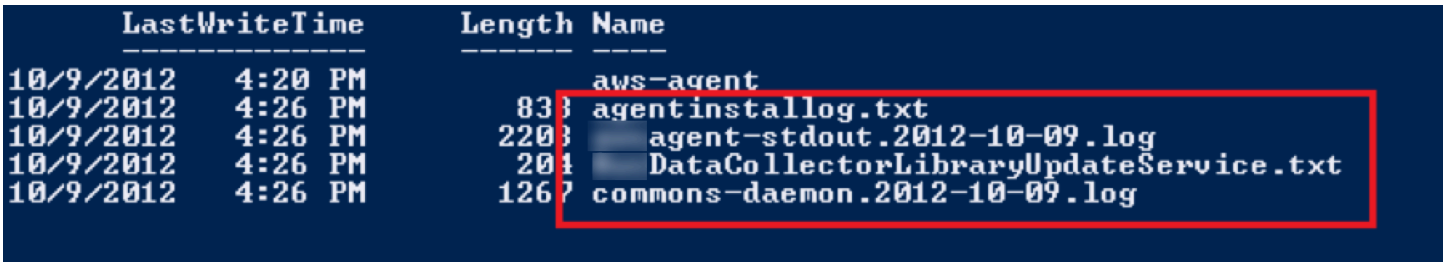
Based on the DataProvider version on your system, follow the procedure in Updating to DataProvider 4.3 or Uninstalling older versions.

Problem: The AWS Data Provider for SAP failed to start at the end of the installation process.

If reinstalling the AWS Data Provider for SAP doesn't solve the problem, you can gather debugging information about the AWS Data Provider for SAP by reviewing the log files in the `C:\Program Files\Amazon\DataProvider` directory.

These log files include an installation log, a log of the service installation, and the output of the AWS Data Provider for SAP itself.

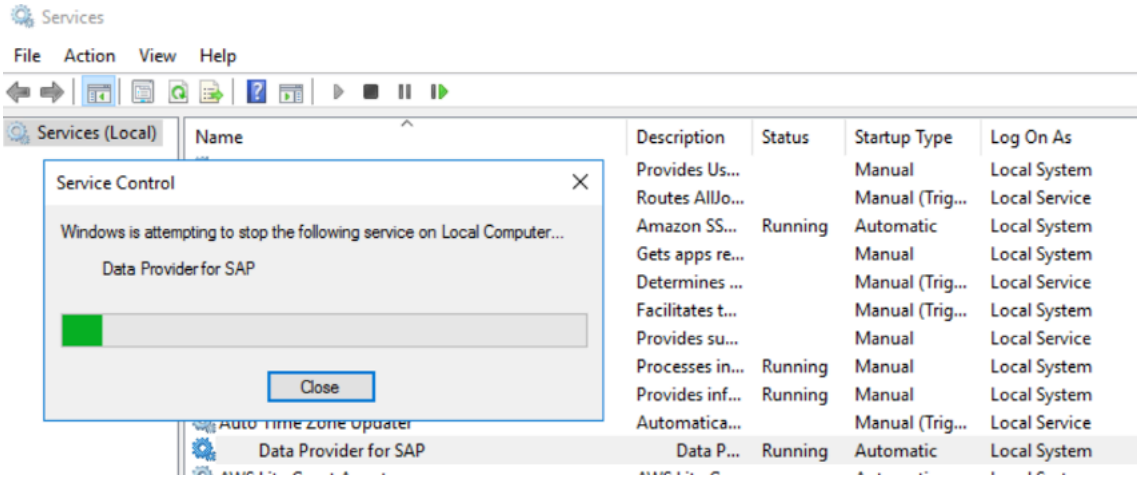
Log files on Windows



Problem: I want to get more detailed log information from the data provider.

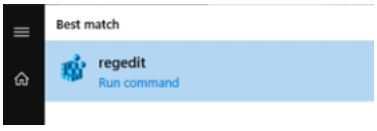
Start by stopping the data provider service.

Stop Service on Windows



Open the registry editor by clicking on the Windows logo in the bottom left and typing `regedit` and then clicking the option that is shown on screen:

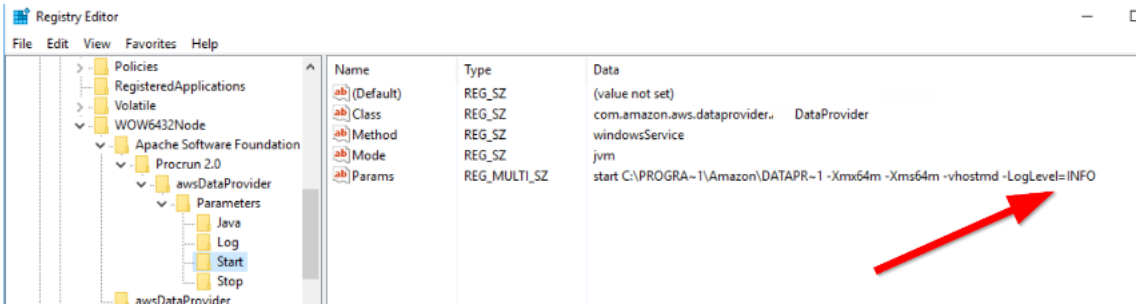
Start `regedit`



In the registry, navigate to the key:

```
HKEY_LOCAL_MACHINE\SOFTWARE\WOW6432Node\Apache Software Foundation\Procrun
 2.0\awsDataProvider\Start
```

Logging setting



The data provider accepts two log levels: INFO and FINE. FINE will generate more detailed logging which can be useful when debugging a problem. The recommendation is to set it back to INFO when you have finished troubleshooting to avoid the logs consuming unnecessary disk space.

Problem: I want to reinstall the AWS Data Provider for SAP from scratch.

Based on the DataProvider version on your system, follow the procedure in Updating to DataProvider 4.3 or Uninstalling older versions.

Problem: When I looked at my logs, I noticed that my installation failed all diagnostics.

Symptoms of internet connectivity problems on Windows

```
14:32:15.862 I 08001  ** Running Diagnostics **

14:32:15.862 I 08002 Diagnostic : [redacted] Connectivity

14:33:19.362 W 08003 Diagnostic : Failed

14:33:19.362 I 08006 Diagnostic : Amazon CloudWatch Connectivity & Access

14:33:19.515 W 08007 Diagnostic : Failed

14:33:19.516 I 0800A Diagnostic : EC2 API Connectivity & Access

14:33:19.542 W 0800B Diagnostic : Failed

14:33:19.542 I 0800E ** Diagnostics Complete **
```

Failing *all* diagnostics indicates that there's a problem with your outbound connection to the internet. You can confirm this by pinging a well-known internet location, like www.amazon.com. The most common cause of routing issues is in the VPC network configuration, which needs to

have either an internet gateway in place or a VPN connection to your data center with a route to the internet.

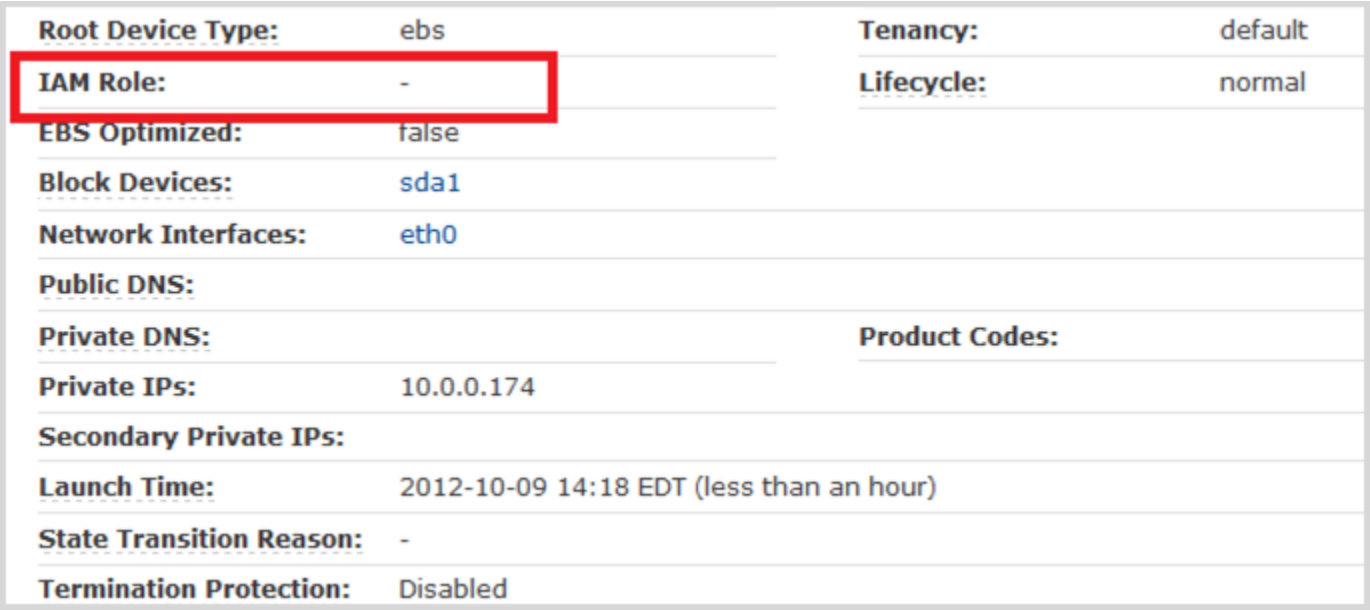
Problem: When I looked at my logs, I noticed that I don't have access to CloudWatch and Amazon EC2, but I did pass the first diagnostic for AWS connectivity.

Symptoms of authorization issues on Windows

```
14:38:57.467 I 08001  ** Running Diagnostics **

14:38:57.468 I 08002 Diagnostic : AWS Connectivity

14:38:58.182 I 08005 Diagnostic : Passed

14:38:58.182 I 08006 Diagnostic : Amazon CloudWatch Connectivity & Access

14:38:58.325 W 08007 Diagnostic : Failed

14:38:58.325 I 0800A Diagnostic : EC2 API Connectivity & Access

14:38:58.357 W 0800B Diagnostic : Failed

14:38:58.357 I 0800E ** Diagnostics Complete **
```

This is a clear indicator that you have an authorization issue when trying to access Amazon CloudWatch and Amazon EC2. The common cause for this problem is not having an IAM role associated with your instance that contains the IAM policy, as specified in IAM Roles earlier in this guide. You can quickly diagnose this issue by looking at the specific EC2 instance in the Amazon EC2 console and verifying the IAM role.

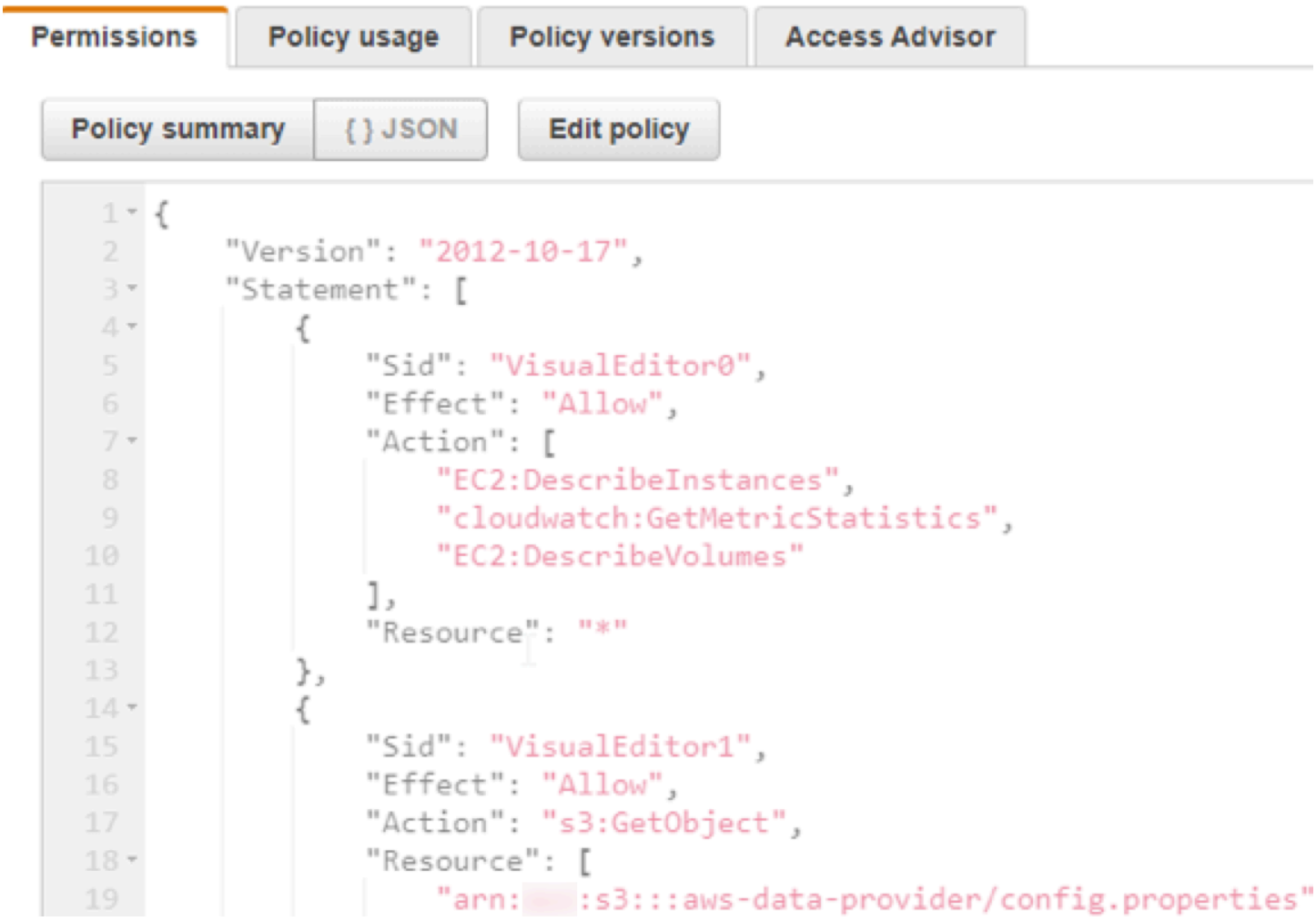
Verifying the IAM role for an EC2 instance



If the IAM role doesn't exist, then create is as specified in IAM Roles, described earlier in this guide.

If you do have an IAM role assigned to the instance, go to the IAM console, select the IAM role name, and then choose **Show**. Verify that you have the required policy that is specified in IAM Roles.

Verifying the policy for the IAM role



Customizing the AWS Data Provider for SAP

Some settings are hard coded in the AWS Data Provider for SAP. You can override existing settings or add new settings. For example, when AWS adds new instance types, you can add these to the AWS Data Provider for SAP configuration.

The AWS Data Provider for SAP creates a database by reading the configuration information from the available `config.properties`, files in this sequence:

- The JAR (Java Archive) file of the data provider application.
- The installation directory. This file is required only if you want to override or extend the current configuration. The default directories are as follows:
  - Linux – `/usr/local/ec2/aws-dataprovider/config.properties`
  - Windows – `C:\Program Files\Amazon\DataProvider\config.properties`

- The Regional S3 bucket. Replace <region> with the Region code for the Region (for example, `us-east-1`).

```
https://aws-sap-dataprovider-<region>.s3.<region>.amazonaws.com/config.properties
```

Syntax Rules for Configuration Files

- The configuration files require a comma after the last value in every row.
- Spaces are not ignored in strings. The entire string between the commas, including any spaces, is accepted as the value.
- If there are multiple rows with the same instance type, the existing value for that type is overwritten.
- Capitalization in strings is case sensitive.

User-Configurable EC2 Instance Types

The AWS Data Provider for SAP maintains a database of all relevant Amazon EC2 instance types for SAP.

Entries for EC2 instance types must be in a comma-separated list, as follows:

ec2type,i-type,cpu,core,threads,t-ecu,ecu,hthread,l-map,w-map,speed,p-ecu,

For example:

```
ec2type,r3.8xlarge,2,16,2,32,1,thread,eth0,lan2,10000,true,
```

where the following applies:

Field name	Content	Example	Type	Description
keyword	ec2type	—	String	A token to identify a record with an EC2 instance description

Field name	Content	Example	Type	Description
i-type (instance -type)	See list	r3.8xlarge	String	Instance type, which must match the EC2 instance metadata string
cpu (CPUs)	1 | 2	2	Integer	Number of sockets
core (Cores)	*integer*	16	Integer	Total number of processor cores
threads (threads per core)	1 | 2	2	Integer	Threads per core
t-ecu (total ECU value)	*integer*	32	Double	ECU value for previous-generation instance types that have ECU ratings; number of cores for post-ECU instance types
ecu (ECU per core)	*double*	1	Double	1 for all post-ECU instance types; total ECU divided by cores for previous-generation instance types that have ECU ratings

Field name	Content	Example	Type	Description
hthread (hyperthr eading)	thread | core	thread	String	**thread** for hyperthreaded instance types; **core** for non-hyperthreaded instance types
l-map (Linux NIC mapping)	eth0	eth0	String	Linux mapping of network interface
w-map (Windows NIC mapping)	eth0	lan2	String	Windows mapping of network interface
speed (network interface speed)	1000 | 2000 | 10000	100000	Integer	Maximum speed of network interface, in KB
p-ecu (post ECU)	true | false	true	Boolean	**true** for modern instances that don't have ECU ratings

User-Configurable EBS Volume Types

The AWS Data Provider for SAP maintains a database of all relevant EBS volume types for SAP.

Entries for EBS volume types must be in a comma-separated list, as follows:

voltype,ebs-type,sample-time,

For example:

```
voltype,io1,60,
```

where the following applies:

Field name	Content	Example	Type	Description
keyword	voltype	—	String	A token to identify a record with an EBS volume descripti on
ebs-type (EBS-type)	io1 | gp2 | sc1 | st1	Io1	String	EBS type, which must match the EBS volume metadata string
sample-time	60 | 300	60	Integer	CloudWatch sample time, in seconds

Important

The sample time is required to calibrate the EBS metrics to the SAP monitoring requirements. Changes in the sample time will lead to incorrect EBS metrics in the SAP monitoring system.

Verification of AWS Data Provider for SAP in SAP system monitoring

The AWS Data Provider for SAP exposes AWS-specific metrics through an XML page at http://localhost:8888/vhostmd of the given system.

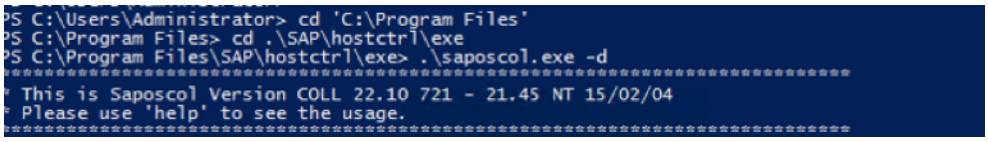
This section explains which metrics get exposed to the SAP system and how you can access them for SAP system monitoring.

Checking Metrics with the SAP Operating System Collector (SAPOSCOL)

The information provided by the AWS Data Provider for SAP is read by the SAP Operating System Collector (SAPOSCOL). You can use the interactive mode of SAPOSCOL to verify that the two tools are working together correctly. The following example shows a lookup under Windows. A lookup under Linux is very similar.

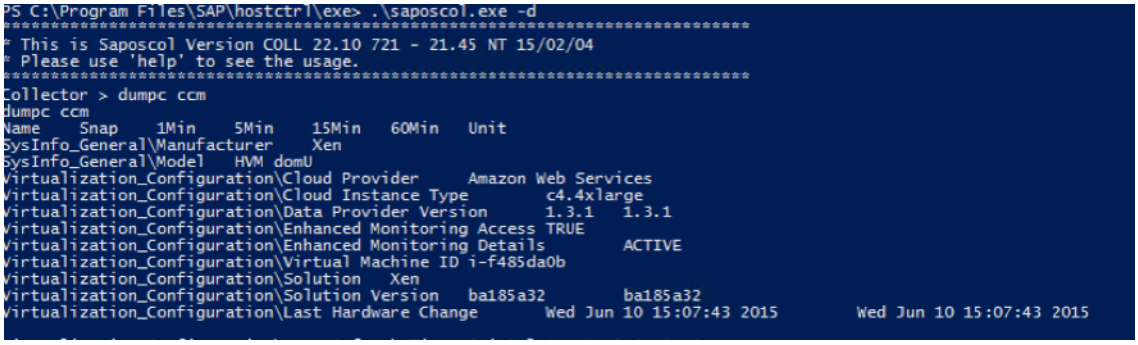
1. Open a Windows command shell and direct the shell to the directory `C:\Program Files\SAP\hostctrl\exe`. Start `saposcol.exe` with the `-d` option.

   Starting SAPOSCOL

2. SAPOSCOL is now in interactive mode. Type `dump ccm` and press **Enter** to list all values gathered. SAPOSCOL will display a lengthy list of metrics, as shown here.

   Metrics from SAPOSCOL

   The following two metrics indicate that SAPOSCOL is collaborating successfully with the AWS Data Provider for SAP:

   - Enhanced Monitoring Access TRUE
   - Enhanced Monitoring Details ACTIVE

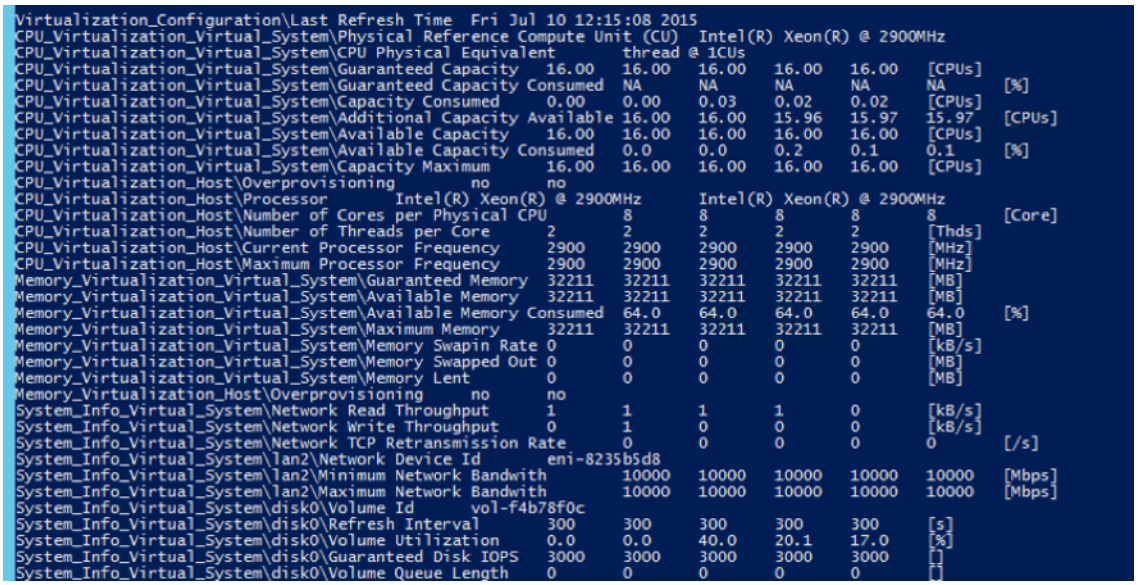
   The AWS-specific metrics start with the following strings:

   +

   - Virtualization_Configuration
   - CPU_Virtualization_Virtual_System
   - Memory_Virtualization_Virtual_System

- System_Info_Virtualization_System

AWS-specific metrics



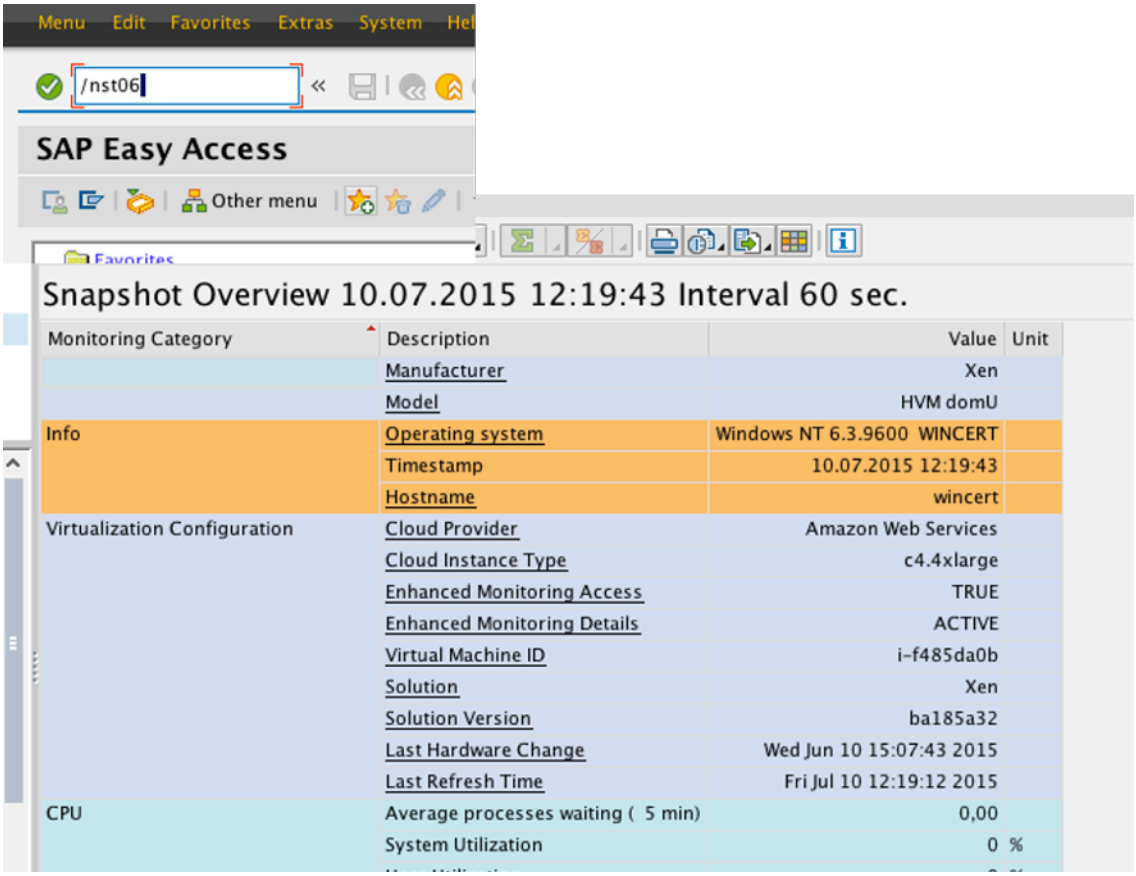
```
Virtualization_Configuration\Last Refresh Time  Fri Jul 10 12:15:08 2015
CPU_Virtualization_Virtual_System\Physical Reference Compute Unit (CU)  Intel(R) Xeon(R) @ 2900MHz
CPU_Virtualization_Virtual_System\CPU Physical Equivalent       thread @ 1CUs
CPU_Virtualization_Virtual_System\Guaranteed Capacity   16.00   16.00   16.00   16.00   16.00   [CPUs]
CPU_Virtualization_Virtual_System\Guaranteed Capacity Consumed  NA      NA      NA      NA      NA      [%]
CPU_Virtualization_Virtual_System\Capacity Consumed     0.00    0.00    0.03    0.02    0.02    [CPUs]
CPU_Virtualization_Virtual_System\Additional Capacity Available 16.00   16.00   15.96   15.97   15.97   [CPUs]
CPU_Virtualization_Virtual_System\Available Capacity    16.00   16.00   16.00   16.00   16.00   [CPUs]
CPU_Virtualization_Virtual_System\Available Capacity Consumed   0.0     0.0     0.2     0.1     0.1     [%]
CPU_Virtualization_Virtual_System\Capacity Maximum      16.00   16.00   16.00   16.00   16.00   [CPUs]
CPU_Virtualization_Host\Overprovisioning        no      no
CPU_Virtualization_Host\Processor       Intel(R) Xeon(R) @ 2900MHz      Intel(R) Xeon(R) @ 2900MHz
CPU_Virtualization_Host\Number of Cores per Physical CPU        8       8       8       8       8       [Core]
CPU_Virtualization_Host\Number of Threads per Core      2       2       2       2       2       [Thds]
CPU_Virtualization_Host\Current Processor Frequency     2900    2900    2900    2900    2900    [MHz]
CPU_Virtualization_Host\Maximum Processor Frequency     2900    2900    2900    2900    2900    [MHz]
Memory_Virtualization_Virtual_System\Guaranteed Memory  32211   32211   32211   32211   32211   [MB]
Memory_Virtualization_Virtual_System\Available Memory   32211   32211   32211   32211   32211   [MB]
Memory_Virtualization_Virtual_System\Available Memory Consumed  64.0    64.0    64.0    64.0    64.0    [%]
Memory_Virtualization_Virtual_System\Maximum Memory     32211   32211   32211   32211   32211   [MB]
Memory_Virtualization_Virtual_System\Memory Swapin Rate 0       0       0       0       0       [kB/s]
Memory_Virtualization_Virtual_System\Memory Swapped Out 0       0       0       0       0       [MB]
Memory_Virtualization_Virtual_System\Memory Lent        0       0       0       0       0       [MB]
Memory_Virtualization_Host\Overprovisioning     no      no
System_Info_Virtual_System\Network Read Throughput      1       1       1       1       0       [kB/s]
System_Info_Virtual_System\Network Write Throughput     0       1       0       0       0       [kB/s]
System_Info_Virtual_System\Network TCP Retransmission Rate      0       0       0       0       0       [/s]
System_Info_Virtual_System\lan2\Network Device Id       eni-8235b5d8
System_Info_Virtual_System\lan2\Minimum Network Bandwith        10000   10000   10000   10000   10000   [Mbps]
System_Info_Virtual_System\lan2\Maximum Network Bandwith        10000   10000   10000   10000   10000   [Mbps]
System_Info_Virtual_System\disk0\Volume Id      vol-f4b78f0c
System_Info_Virtual_System\disk0\Refresh Interval       300     300     300     300     300     [s]
System_Info_Virtual_System\disk0\Volume Utilization     0.0     0.0     40.0    20.1    17.0    [%]
System_Info_Virtual_System\disk0\Guaranteed Disk IOPS   3000    3000    3000    3000    3000    []
System_Info_Virtual_System\disk0\Volume Queue Length    0       0       0       0       0       []
```

Checking Metrics with the SAP CCMS Transactions

SAPOSCOL hands the AWS-enhanced statistics with other operating system-specific metrics to the SAP system. You can also check the AWS-enhanced statistics in the SAP CCMS. You can enter the transaction *st06* (or */nst06*) in the upper-left transaction field of the SAP GUI for quick access to this data.

> **Note**
>
> You will need the appropriate authorizations to look up this information.

Statistics in the SAP CCMS (standard view)



On this screen, you can verify core AWS information such as:

- Cloud provider
- Instance type
- Status of enhanced monitoring access (must be TRUE)
- Status of enhanced monitoring details (must be ACTIVE)
- Virtual machine identifier

> **Important**
>
> The enhanced AWS metrics aren't shown in standard view.

To view enhanced AWS statistics, choose the **Standard View** button in the upper-left corner. It changes to **Expert View** and displays the enhanced AWS statistics. The list that appears is comprehensive. It shows the processor details.

Enhanced AWS statistics (expert view)

CPU Virtualization Host	Overprovisioning	no	
	Processor	Intel(R) Xeon(R) @ 2900MHz	
	Number of Cores per Physical CPU	8	Core
	Number of Threads per Core	2	Thds
	Current Processor Frequency	2.900	MHz
	Maximum Processor Frequency	2.900	MHz
CPU Virtualization Virtual System	Physical Reference Compute Unit (CU)	Intel(R) Xeon(R) @ 2900MHz	
	CPU Physical Equivalent	thread @ 1CUs	
	Guaranteed Capacity	16,00	CPUs
	Capacity Consumed	0,00	CPUs
	Additional Capacity Available	16,00	CPUs
	Available Capacity	16,00	CPUs
	Available Capacity Consumed	0,0	%
	Capacity Maximum	16,00	CPUs

It also shows details about the memory subsystem (main memory and disks) and network interfaces.

Memory and networking statistics (expert view)

Memory Virtualization Host	Overprovisioning	no	
	Network Read Throughput	5	kB/s
	Network Write Throughput	4	kB/s
	Network TCP Retransmission Rate	0	/s
	lan2\Network Device Id	eni-8235b5d8	
	lan2\Minimum Network Bandwith	10.000	Mb...
	lan2\Maximum Network Bandwith	10.000	Mb...
	disk0\Volume Id	vol-f4b78f0c	
	disk0\Refresh Interval	300	s
	disk0\Volume Utilization	0,0	%
	disk0\Guaranteed Disk IOPS	3.000	
	disk0\Volume Queue Length	0	
	disk0\Volume Read Response Time	0	msec
	disk0\Volume Write Response Time	0	msec
	disk0\Volume Read Throughput	1	kB/s
	disk0\Volume Write Throughput	15	kB/s
	disk0\Volume Read Ops	0	/s
	disk0\Volume Write Ops	2	/s
	disk1\Volume Id	vol-89676771	
	disk1\Refresh Interval	300	s
	disk1\Volume Utilization	0,0	%
	disk1\Guaranteed Disk IOPS	300	
	disk1\Volume Queue Length	0	
	disk1\Volume Read Response Time	0	msec
	disk1\Volume Write Response Time	0	msec
	disk1\Volume Read Throughput	0	kB/s
	disk1\Volume Write Throughput	0	kB/s
	disk1\Volume Read Ops	0	/s
	disk1\Volume Write Ops	0	/s

> **Note**
>
> The screen illustrations in Figures 37–39 were taken from SAP NetWeaver 7.4 SP08. This version shows the enhanced AWS statistics in the **Memory Virtualization** section. This problem has been fixed by SAP in later versions of NetWeaver.

Example of captured metrics

This following show example metrics. Your system metrics may slightly differ.

```
<metrics>
<metric context="host" category="config" type="long" unit="posixtime">
<name>Time Stamp</name>
<value>1584376572</value>
</metric>
<metric context="host" category="config" type="int64" unit="sec">
<name>Refresh Interval</name>
<value>60</value>
</metric>
<metric context="vm" category="config" type="string" unit="none">
<name>Data Provider Version</name>
<value>3.0.139</value>
</metric>
<metric context="host" category="config" type="string" unit="none">
<name>Cloud Provider</name>
<value>Amazon Web Services</value>
</metric>
<metric context="vm" category="config" type="string" unit="none">
<name>Instance Type</name>
<value>m5.large</value>
</metric>
<metric context="host" category="config" type="string" unit="none">
<name>Virtualization Solution</name>
<value>KVM</value>
</metric>
<metric context="host" category="config" type="string" unit="none">
<name>Virtualization Solution Version</name>
<value>ba185a32</value>
</metric>
<metric context="host" category="config" type="long" unit="none">
<name>CloudWatch Calls</name>
```

```
<value>12</value>
</metric>
<metric context="host" category="config" type="long" unit="none">
<name>EC2 Calls</name>
<value>4</value>
</metric>
<metric context="vm" category="config" type="string" unit="none">
<name>CPU Over-Provisioning</name>
<value>no</value>
</metric>
<metric context="vm" category="config" type="string" unit="none">
<name>Memory Over-Provisioning</name>
<value>no</value>
</metric>
<metric context="vm" category="config" type="string" unit="none">
<name>Virtualization Type</name>
<value>default-hvm</value>
</metric>
<metric context="vm" category="config" type="string" unit="none">
<name>Virtual Machine ID</name>
<value>i-#################</value>
</metric>
<metric context="vm" category="config" type="long" unit="posixtime">
<name>Last Hardware Change</name>
<value>1572284007</value>
</metric>
<metric context="host" category="cpu" type="string" unit="none">
<name>Processor Type</name>
<value>Intel(R) Xeon(R) @ 2500MHz</value>
</metric>
<metric context="host" category="cpu" type="int64" unit="none">
<name>Number of Cores per CPU</name>
<value>1</value>
</metric>
<metric context="host" category="cpu" type="int64" unit="none">
<name>Number of Threads per Core</name>
<value>2</value>
</metric>
<metric context="host" category="cpu" type="int64" unit="MHz">
<name>Max HW Frequency</name>
<value>2500</value>
</metric>
<metric context="host" category="cpu" type="int64" unit="MHz">
<name>Current HW Frequency</name>
```

```
<value>2500</value>
</metric>
<metric context="vm" category="cpu" type="string" unit="none">
<name>Reference Compute Unit (CU)</name>
<value>Intel(R) Xeon(R) @ 2500MHz</value>
</metric>
<metric context="vm" category="cpu" type="string" unit="none">
<name>vCPU Mapping</name>
<value>thread</value>
</metric>
<metric context="vm" category="cpu" type="long" unit="cu">
<name>Phys. Processing Power per vCPU</name>
<value>1</value>
</metric>
<metric context="vm" category="cpu" type="int64" unit="cu">
<name>Guaranteed VM Processing Power</name>
<value>2</value>
</metric>
<metric context="vm" category="cpu" type="int64" unit="cu">
<name>Current VM Processing Power</name>
<value>2</value>
</metric>
<metric context="vm" category="cpu" type="int64" unit="cu">
<name>Max. VM Processing Power</name>
<value>2</value>
</metric>
<metric context="vm" category="cpu" type="double" unit="percent">
<name>VM Processing Power Consumption</name>
<value>3.00</value>
</metric>
<metric context="vm" category="memory" type="long" unit="MB">
<name>Guaranteed Memory assigned</name>
<value>8274</value>
</metric>
<metric context="vm" category="memory" type="long" unit="MB">
<name>Current Memory assigned</name>
<value>8274</value>
</metric>
<metric context="vm" category="memory" type="long" unit="MB">
<name>Max Memory assigned</name>
<value>8274</value>
</metric>
<metric context="vm" category="memory" type="double" unit="percent">
<name>VM Memory Consumption</name>
```

```
<value>29.00</value>
</metric>
<metric context="vm" category="memory" type="int64" unit="KB/sec">
<name>Memory SwapIn Rate</name>
<value>0</value>
</metric>
<metric context="vm" category="memory" type="int64" unit="MB">
<name>Memory Swapped Out</name>
<value>0</value>
</metric>
<metric context="vm" category="memory" type="int64" unit="MB">
<name>Memory Lent</name>
<value>0</value>
</metric>
<metric context="vm" category="memory" type="int64" unit="MB">
<name>Total Visible Memory</name>
<value>-1</value>
</metric>
<metric context="vm" category="memory" type="int64" unit="percent">
<name>Visible Memory Consumed</name>
<value>-1</value>
</metric>
<metric context="vm" category="memory" type="int64" unit="KB/sec">
<name>Visible Memory SwapIn Rate</name>
<value>0</value>
</metric>
<metric context="vm" category="memory" type="int64" unit="MB">
<name>Visible Memory Swapped Out</name>
<value>0</value>
</metric>
<metric context="vm" category="network" type="int64" unit="bytes">
<name>Network Read Bytes</name>
<value>54110386</value>
</metric>
<metric context="vm" category="network" type="int64" unit="bytes">
<name>Network Write Bytes</name>
<value>1330726</value>
</metric>
<metric context="vm" category="network" type="int64" unit="none">
<name>TCP Packets Retransmitted</name>
<value>396480</value>
</metric>
<metric context="vm" category="network" type="int64" unit="Mbps" device-id="eni--
#################</">
```

```
<name>Minimum Network Bandwidth</name>
<value>10000</value>
</metric>
<metric context="vm" category="network" type="int64" unit="Mbps" device-id="eni--
##################</">
<name>Maximum Network Bandwidth</name>
<value>10000</value>
</metric>
<metric context="vm" category="network" type="string" unit="none" device-id="eni--
##################</">
<name>Mapping</name>
<value>lan2</value>
</metric>
<metric context="vm" category="disk" type="int64" unit="msec" device-id="vol--
##################</">
<name>Volume Idle Time</name>
<value>58489</value>
</metric>
<metric context="vm" category="disk" type="int64" unit="none" device-id="vol--
##################</">
<name>Volume Queue Length</name>
<value>0</value>
</metric>
<metric context="vm" category="disk" type="int64" unit="bytes" device-id="vol--
##################</">
<name>Volume Read Bytes</name>
<value>9878528</value>
</metric>
<metric context="vm" category="disk" type="int64" unit="none" device-id="vol--
##################</">
<name>Volume Read Ops</name>
<value>144</value>
</metric>
<metric context="vm" category="disk" type="int64" unit="msec" device-id="vol--
##################</">
<name>Volume Read Time</name>
<value>246</value>
</metric>
<metric context="vm" category="disk" type="int64" unit="msec" device-id="vol--
##################</">
<name>Volume Write Time</name>
<value>8266</value>
</metric>
```

```
<metric context="vm" category="disk" type="int64" unit="bytes" device-id="vol--
################</">
<name>Volume Write Bytes</name>
<value>282332160</value>
</metric>
<metric context="vm" category="disk" type="int64" unit="none" device-id="vol--
################</">
<name>Volume Write Ops</name>
<value>3090</value>
</metric>
<metric context="vm" category="disk" type="string" unit="none" device-id="vol--
################</">
<name>Volume Type</name>
<value>gp2</value>
</metric>
<metric context="vm" category="disk" type="int64" unit="none" device-id="vol--
################</">
<name>Guaranteed IOps</name>
<value>180</value>
</metric>
<metric context="vm" category="disk" type="int64" unit="sec" device-id="vol--
################</">
<name>Interval</name>
<value>300</value>
</metric>
<metric context="vm" category="disk" type="string" unit="none" device-id="vol--
################</">
<name>Mapping</name>
<value>disk0</value>
</metric>
</metrics>
```

Version history

Version 4.3.2 (August, 2023)

- Bug fix : Security updates to address CVE-2022-45688.

Version 4.3.1 (June, 2023)

- Bug fix : Data Provider is now setup for successful installation of SAP JVM.

Version 4.3 (January, 2023)

- Added support for JDK 17
- Added reading configuration information from remote Amazon S3 bucket

Version 4.2 (November, 2022)

- Added support for Oracle and Linux
- Added integration with Linux `logrotate` feature
- Updates to the RPM package build.

Version 4.1.1 (September, 2022)

- Added support for new Amazon EC2 instance types.

Version 4.1.0 (January, 2022)

- Added support for JDK 11.
- Added support for new Amazon EC2 instance types.

Version 4.0.3 (December, 2021)

- Bug fixes: Removed Log4j dependency.

Version 4.0.2 (December, 2021)

- Bug fixes: Security updates for Log4j2 issue (CVE-2021-44228).

Version 4.0 (April, 2021)

- Initial release of the 4.0 version.
- Support for SSM package installation.
- Support for IMDSv2.

Version 3.0 (April, 2020)

- Initial release of the 3.0 version.
- Switched the Java Runtime from Oracle to Amazon Corretto.

Version 2.9 (August 30, 2017)

- Added support for China Regions.
- Added Linux uninstaller.
- Linux installer can be customized to install from a custom S3 bucket.
- Silent installer for Windows (does not require any input).
- Improvements in determination of access points.
- Support for X1E instance family.

Version 2.8 (March 1, 2017)

- SLES 12, Red Hat 7, and Oracle Linux 7 will now use SYSTEMD to manage the daemon.
- Support for SLES and SLES for SAP 12 SP2.
- SLES 12 SP1 systems will get migrated from Linux services to SYSTEMD when trying to install the AWS Data Provider without having it de- installed first.
- Minor changes in logging texts.
- Support for R4 and M4 instance types.
- Updated Windows installation verification.

Version 2.7 (December 21, 2016)

- Support for Canada (Central), US East (Ohio), and EU (London) Regions.
- Default access point resolution for common AWS Regions is added.

Version 2.6 (September 1, 2016)

- Bug fixes: Installation script checks for existence of wget
- Support for Oracle Linux.

Version 2.5 (May 2, 2016)

- Bug fixes: Security and stability fixes in versions 2.2-2.4.
- New: Support for new Amazon EBS volume types:
  - Throughput Optimized HDD (st1)
  - Cold HDD (sc1)
- New: Support for the Amazon EC2 X1 instance family.

Version 2.1 (January 20, 2016)

- Support for Asia Pacific (Seoul) Region.
- Bug fix: Version 2.0 pulled files from an incorrect S3 bucket for installation. Version 2.0 needs to be uninstalled before version 2.1 is installed.

Version 2.0 (December 22, 2015)

- New: Windows devices in the range sdb to sdzz get correct SCSI device IDs assigned.
- New: Java VM consumption is now limited to 64 MB maximum heap size.

Version 1.3.1 (July 14, 2015)

- Bug fixes: Security fixes.
- New: Support for C4, D2, and M4 instance types. Users who migrate instances with installed 1.3 agents will automatically receive support for the new instance types through an updated configuration database on the web.

Version 1.3 (February 17, 2015)

- New: Support for new Amazon EC2 C4 instance family.
- Security fix: Upgraded Linux and Windows versions to JRE 8u31.
- Bug fix: Relative performance of c3.8xlarge instances is now reported correctly.
- New: CloudWatch and Amazon EC2 metrics access points:
  - Support for the EU (Frankfurt) Region was added.
  - Access points are user configurable. You can add information about new AWS Regions without having to install a new product version.

  - Access points are now updated from an internet-based database file. You can add new AWS Regions by updating a web-based configuration file and then restarting the daemon/service.
- New: Message log files with fixed disk space consumption are provided on Linux.
- New: User-configurable EC2 instance types are available.
- New: Web update support was added for future EC2 instance types without product updates.
- Bug fix: GP2 volumes now report the correct sample interval time.
- New: User-configurable sample times for new EBS volume types are now available.
- New: The AWS Data Provider for SAP now reports the virtualization type of the EC2 instance.

Version 1.2.2 (October 1, 2014)

- Windows bug fix: Installer executable pulls installation from correct Amazon S3 bucket.
- Windows bug fix: AWS Data Provider for SAP now reports the correct disk mapping for Windows EBS volumes with the following names: xvd[a-z][a-z].

Version 1.2.1 (September 29, 2014)

- Bug fix: EBS volumes now report correct attribute type ("string") for volume type.

Version 1.2 (September 16, 2014)

- New: Support for the T2, R3, and C3 instance families.
- New: Support for post-ECU (EC2 Compute Unit) instance types:
  - New instance types no longer have ECU values.
  - The reference compute power for these instance types is a hardware thread of the given processor. The total CPU power is equal to the number of the vCPUs of a given instance type.
- New: Support for the new EBS GP2 volume type.
  - Every volume is now tagged with the EBS volume type.
- New: Report of EBS one-minute volume statistics.
  - EBS volumes now report their individual sample interval in a separate attribute.
- Bug fix: EBS volume mapping for Windows devices now reports the correct name.
- Bug fix: Installation, update, and operation through HTTP/HTTPS proxies has been fixed.
- New: JRE 8 support has been added on Linux.

SAP on AWS cost estimation

Last updated: May 2023

AWS offers pay-as-you-go pricing. You only pay for the services you use, for as long as you use them. There are no long-term contracts or complex licensing requirements. For more information, see AWS Pricing and https://calculator.aws/#/.

The following is an overview of the pricing characteristics of AWS services that are frequently used for the deployment and operation of SAP systems on AWS.

Topics

AWS Region

AWS service pricing varies between different AWS Regions. You must select the Region in which you want to deploy your SAP system to begin creating an estimate. For more information, see Regions and Availability Zones.

Compute

Amazon Elastic Compute Cloud (Amazon EC2) provides a wide selection of instance types that provide varying combinations of CPU, memory, storage, I/O, and networking capabilities. Each running instance is charged by the hour. For more information, see Amazon EC2 pricing.

Amazon EC2 offers multiple purchasing options that give you flexibility to optimize your costs. For more information, see Instance purchasing options.

Storage

The following AWS services are flexible, cost-effective, and easy-to-use data storage options available for your SAP systems. Each option has a unique combination of performance and durability. For more information, see Cloud storage on AWS.

Topics

- Amazon EBS
- Amazon EFS
- Amazon FSx for Windows File Server
- Amazon FSx for NetApp ONTAP
- Amazon S3

Amazon EBS

Amazon Elastic Block Store (Amazon EBS) provides persistent, block-level storage volumes for Amazon EC2 instances. Each Amazon EC2 instance that runs an SAP environment requires one or more Amazon EBS volumes to store system components, such as operating system, SAP software, SAP database data and log files, and local backup storage.

With Amazon Elastic Block Store, you only pay for what you provision. For more information, see Amazon EBS pricing.

Amazon EBS snapshots

Amazon EBS snapshots are point-in-time copies of your block data stored in Amazon EBS volumes. Amazon EBS snapshots in the Standard tier are stored incrementally, which means you are billed only for the changed blocks stored. Amazon EBS snapshots in the Archive tier are full copies of your block data, which means you are billed for all the blocks stored and not just the changed blocks.

You can also enable a Recycle Bin feature to protect against accidental deletion. Amazon EBS snapshots in the recycle bin are billed at the same rate.

Another feature of Amazon EBS snapshots which is applicable to SAP workloads is the Fast Snapshot Restore (FSR). This feature enables you to promptly restore fully provisioned Amazon EBS volumes from snapshots, regardless of the size of the volume or snapshot. FSR is charged in Data Services Unit-Hours (DSU-Hours) for each snapshot and each Availability Zone in which it is enabled. DSUs mean that you are billed per minute with a one-hour minimum.

Amazon EBS snapshots can be used for both root and binary volumes of your SAP system as well as database volumes.

Amazon EFS

Amazon Elastic File System (Amazon EFS) provides serverless file storage. You can share file data without provisioning or managing storage capacity and performance. Amazon EFS is built to scale on demand to petabytes without disrupting applications, growing and shrinking automatically as you add and remove files. You can create and configure file systems quickly and easily.

Amazon EFS stores every object across multiple Availability Zones, making it highly available. It supports the Network File System version 4 (NFSv4.1 and NFSv4.0) protocol. It is certified for SAP file shares, and it can also be used for storing data files in your SAP landscape.

With Amazon EFS, you pay only for the storage used by your file system, and there is no minimum fee or setup cost. For more information, see Amazon EFS pricing.

Amazon FSx for Windows File Server

Amazon FSx for Windows File Server provides fully managed Microsoft Windows file servers, backed by a fully native Windows file system. It has support for Windows file system features and for the industry-standard Server Message Block (SMB) protocol to access file storage over a network.

FSx for Windows File Server is certified for SAP workloads on AWS, and can also be used for Windows based data file sharing in your SAP landscape.

With FSx for Windows File Server, you only pay for the resources you use, and there is no minimum fee or setup cost. For more information, see Amazon FSx for Windows File Server pricing.

Amazon FSx for NetApp ONTAP

Amazon FSx for NetApp ONTAP is a fully managed service that provides highly reliable, scalable, high-performing, and feature-rich file storage built on NetApp's popular ONTAP file system.

FSx for ONTAP is certified for SAP workloads on AWS.

You are billed for the file systems you use, based on the following categories.

- SSD storage capacity (per gigabtye-month, or GB-month)
- SSD IOPS that you provision above three IOPS/GB (per IOPS-month)
- Throughput capacity (per megabytes per second [MBps]-month)
- Capacity pool storage consumption (per GB-month)
- Capacity pool requests (per read and write)
- Backup storage consumption (per GB-month)

For more information, see Amazon FSx for NetApp ONTAP pricing.

Amazon S3

Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance. You can use Amazon S3 to stage media, store backups, and archive data. Amazon S3 offers a range of storage classes designed for different use cases.

Amazon S3 charges you only for what you actually use, with no hidden fees and no overage charges. This model gives you a variable-cost service that can grow with your business while giving you the cost advantages of AWS infrastructure. For more information, see Amazon S3 pricing.

Network

AWS offers multiple strong and secure networking services.

Topics

- Amazon VPC
- AWS Site-to-Site VPN
- AWS Direct Connect
- Elastic Load Balancing
- Data transfer pricing

Amazon VPC

With Amazon Virtual Private Cloud (Amazon VPC), you can launch AWS resources in a logically isolated virtual network that you've defined. This virtual network closely resembles a traditional network that you'd operate in your own data center, with the benefits of using the scalable infrastructure of AWS.

There's no additional charge for using Amazon VPC. There are charges for some components, such as NAT gateways, IP Address Manager, traffic mirroring, Reachability Analyzer, and Network Access Analyzer. For more information, see Amazon VPC pricing.

Use the following options for a secure connection between your on-premises network and Amazon VPC.

- AWS Transit Gateway is a network transit hub that you can use to interconnect your virtual private clouds (VPCs) and on-premises networks. As your cloud infrastructure expands globally, inter-Region peering connects transit gateways together using the AWS Global Infrastructure. Your data is automatically encrypted and never travels over the public internet.

  You are charged hourly for each attachment on a transit gateway, and you are charged for the amount of traffic processed on the transit gateway. For more information, see AWS Transit Gateway pricing.
- NAT gateway is a Network Address Translation (NAT) service. You can use a NAT gateway so that instances in a private subnet can connect to services outside your VPC but external services cannot initiate a connection with those instances.

  When you provision a NAT gateway, you are charged for each hour that your NAT gateway is available and each Gigabyte of data that it processes. For more information, see Amazon VPC pricing.
- AWS PrivateLink is a highly available, scalable technology that you can use to privately connect your VPC to services as if they were in your VPC. You do not need to use an internet gateway, NAT device, public IP address, AWS Direct Connect connection, or AWS Site-to-Site VPN connection to allow communication with the service from your private subnets. Therefore, you control the specific API endpoints, sites, and services that are reachable from your VPC.

  For information about the pricing for VPC endpoints, see AWS PrivateLink pricing.

AWS Site-to-Site VPN

By default, instances that you launch into an Amazon VPC can't communicate with your own (remote) network. You can enable access to your remote network from your VPC by creating an AWS Site-to-Site VPN (Site-to-Site VPN) connection, and configuring routing to pass traffic through the connection.

You are charged for each VPN connection hour that your VPN connection is provisioned and available. For more information, see AWS Site-to-Site VPN and Accelerated Site-to-Site VPN Connection pricing.

You are charged for data transfer out from Amazon EC2 to the internet. For more information, see Data Transfer.

When you create an accelerated VPN connection, we create and manage two accelerators on your behalf. You are charged an hourly rate and data transfer costs for each accelerator. For more information, see AWS Global Accelerator pricing.

AWS Direct Connect

AWS Direct Connect links your internal network to an AWS Direct Connect location over a standard Ethernet fiber-optic cable. One end of the cable is connected to your router, the other to an AWS Direct Connect router. With this connection, you can create virtual interfaces directly to public AWS services (for example, to Amazon S3 or Amazon VPC), bypassing internet service providers in your network path. An AWS Direct Connect location provides access to AWS in the Region with which it is associated. You can use a single connection in a public Region or AWS GovCloud (US) to access public AWS services in all other public Regions.

AWS Direct Connect has two billing elements: port hours and outbound data transfer. For more information, see AWS Direct Connect pricing.

Elastic Load Balancing

Elastic Load Balancing automatically distributes your incoming traffic across multiple targets, such as Amazon EC2 instances, containers, and IP addresses, in one or more Availability Zones. It monitors the health of its registered targets, and routes traffic only to the healthy targets. Elastic Load Balancing scales your load balancer capacity automatically in response to changes in incoming traffic.

Use Elastic Load Balancing for setting up highly available SAP environments on AWS.

With your load balancer, you pay only for what you use. For more information, see Elastic Load Balancing pricing.

Data transfer pricing

Data transfer pricing varies based on the architecture pattern and use case. It only amounts to a small percentage of the overall cost of SAP workloads on AWS, 1-5%.

The following table summarizes the common data transfer scenarios that can apply to your SAP landscape.

Scenario	Architecture	Data transfer cost	Pricing guidance	Other costs
Inbound to AWS	All	No		
Outbound from AWS to Internet	All	Yes	Based on the services used	
Services in the same AWS Region	Internet gateway	No		
Services in the same AWS Region	NAT gateway	No		Per GB processing charge, see Amazon VPC pricing
Services in the same AWS Region	AWS PrivateLink/ VPC endpoint	No		See AWS PrivateLink pricing
Services in different AWS Regions	All	Yes	Per GB charge for inter-Region data transfer, see Amazon EC2 pricing	

Scenario	Architecture	Data transfer cost	Pricing guidance	Other costs
Components in same Availability Zone	All	No		
Components in different Availability Zones	AWS Transit Gateway	No		AWS Transit Gateway processing charges, see AWS Transit Gateway pricing
Components in different Availability Zones	Amazon VPC peering	Yes	Per GB charge for inter-Region data transfer, see Amazon EC2 pricing	
Components in different AWS Regions	AWS Transit Gateway	Yes	Per GB charge for inter-Region data transfer, see Amazon EC2 pricing	AWS Transit Gateway processing charges
Components in different AWS Regions	Amazon VPC peering	Yes	Per GB charge for inter-Region data transfer, see Amazon EC2 pricing	
Transfer to on-premises or corporate network	AWS VPN	Yes	Per GB charge for data transfer out, see Amazon EC2 pricing	AWS VPN and AWS Transit Gateway charges

Scenario	Architecture	Data transfer cost	Pricing guidance	Other costs
Transfer to on-premises or corporate network	AWS Direct Connect	Yes	Per GB charge for data transfer out based on location and provider, see Amazon EC2 pricing	AWS Direct Connect and AWS Transit Gateway charges

Automation

With AWS Systems Manager for SAP, you can backup and restore SAP HANA databases on Amazon EC2 with AWS Backup.

AWS Systems Manager for SAP is available to you at no additional cost. You only pay for the AWS resources that you provision to manage and operate your SAP environments.

Backup, restore, and recovery

With these services, you can quickly and effectively backup, restore, and recovery your SAP workloads.

Topics

- AWS Backint Agent for SAP HANA
- AWS Elastic Disaster Recovery
- Amazon EBS snapshots

AWS Backint Agent for SAP HANA

AWS Backint Agent for SAP HANA (AWS Backint agent) is an SAP-certified backup and restore application for SAP HANA workloads running on Amazon EC2 instances in the cloud. AWS Backint agent runs as a standalone application that integrates with your existing workflows to back up your SAP HANA database to Amazon S3 and AWS Backup.

AWS Backint agent is a free service. You pay for only the underlying AWS services that you use, for example Amazon S3 or AWS Backup. See the following references for more information.

- Amazon S3 pricing
- AWS Backup pricing

AWS Elastic Disaster Recovery

AWS Elastic Disaster Recovery (Elastic Disaster Recovery) minimizes downtime and data loss with fast, reliable recovery of on-premises and cloud-based applications using affordable storage, minimal compute, and point-in-time recovery.

You only pay for the servers you are actively replicating to AWS. For more information, see AWS Elastic Disaster Recovery pricing.

Amazon EBS snapshots

See the Amazon EBS section.

Migration

The following services enable you to quickly move application and files.

Topics

- Migration Hub Orchestrator
- AWS DataSync

Migration Hub Orchestrator

AWS Migration Hub Orchestrator simplifies and automates the migration of servers and enterprise applications to AWS. It provides a single location to run and track your migrations.

AWS Migration Hub Orchestrator is available to you at no additional cost. You only pay for the AWS resources that you provision for migrations.

AWS DataSync

AWS DataSync is an online data movement and discovery service that simplifies data migration and helps you quickly, easily, and securely move your file or object data to, from, and between AWS storage services.

You pay only for the amount of data that you migrate based on a flat, per-gigabyte fee according to your AWS Region. For more information, see AWS DataSync pricing.

Monitoring

With the use of following services, you can monitor your SAP workloads on AWS.

Topics

- AWS Data Provider
- Amazon CloudWatch Application Insights for SAP HANA
- Amazon CloudWatch
- AWS CloudTrail
- VPC Flow Logs

AWS Data Provider

AWS Data Provider for SAP is a tool that collects performance-related data from AWS services. It makes this data available to SAP applications to help monitor and improve the performance of business transactions.

For information about costs, see AWS Data Provider for SAP pricing.

Amazon CloudWatch Application Insights for SAP HANA

Amazon CloudWatch Application Insights helps you monitor your applications that use Amazon EC2 instances along with other application resources.

CloudWatch Application Insights sets up recommended metrics and logs for selected application resources using CloudWatch metrics, Logs, and Events for notifications on detected problems. These features are charged to your AWS account according to Amazon CloudWatch pricing. For more information, see CloudWatch Application Insights pricing.

Amazon CloudWatch

Amazon CloudWatch monitors your resources and applications running on AWS in real time. You can collect and track metrics for your resources and applications.

For information about CloudWatch pricing, refer the following resources.

- CloudWatch billing and cost
- Amazon CloudWatch pricing

AWS CloudTrail

AWS CloudTrail is an AWS service that helps you enable operational and risk auditing, governance, and compliance of your AWS account. Actions taken by a user, role, or an AWS service are recorded as events in CloudTrail.

CloudTrail is charged pay per usage with no minimum fee. For more information, see AWS CloudTrail pricing.

VPC Flow Logs

VPC Flow Logs is a feature that enables you to capture information about the IP traffic going to and from network interfaces in your VPC.

Data ingestion and archival charges for vended logs apply when you publish flow logs. For more information about pricing when publishing vended logs, open Amazon CloudWatch pricing, select **Logs** > **Vended Logs**.

Operating System licenses

You can bring your own licenses for the operating system of your choice or purchase from AWS Marketplace.

Topics

- Red Hat
- SUSE
- Windows
- Oracle Enterprise Linux

Red Hat

Red Hat offers two Linux distributions to run SAP workloads. For more details, you can check Red Hat documentation – Overview of the Red Hat Enterprise Linux (RHEL) for SAP Solutions subscription.

You can avail these options from AWS Marketplace or Red Hat Cloud Access. When you purchase Red Hat operating systems from AWS, your Support plan includes operating system support.

If you want to launch an Amazon EC2 instance with RHEL for SAP Applications, then you must have a subscription for the Red Hat Cloud Access program. With RHEL 8, RHEL for SAP Solutions or RHEL for SAP Applications is required for running SAP applications in production environments.

RHEL for SAP Solutions on AWS Marketplace is available at an hourly rate or as an annual commitment. RHEL for SAP Solutions is designed specifically to run SAP workloads. It includes a longer lifecycle support with Extended Update Support (E4S) that provides support on specific minor releases for four years from general availability. It also provides all the necessary packages for configuring the Pacemaker based cluster, ensuring reliability and availability of critical production services.

> **Note**
>
> AWS Marketplace pricing displays the software cost for RHEL. The additional cost of RHEL is included in Amazon EC2 pricing.

SUSE

SUSE offers two Linux distributions to run SAP workloads – SUSE Linux Enterprise Server (SLES) and SUSE Linux Enterprise Server for SAP Applications (SLES for SAP)

You can avail these options from AWS Marketplace or from SUSE. When you purchase SUSE operating systems from AWS, your Support plan includes operating system support.

When you bring your own subscription, the support for Amazon EC2 is based on your SUSE purchasing agreement.

SLES for Amazon EC2 is available at an hourly rate or as an annual commitment. RHEL for SAP Solutions is designed specifically to run SAP workloads. It includes a longer lifecycle support with

Extended Update Support (E4S) that provides support on specific minor releases for four years from general availability. It also provides all the necessary packages for configuring the Pacemaker based cluster, ensuring reliability and availability of critical production services.

SLES for SAP on AWS Marketplace is available at an hourly rate or as an annual commitment. The price of the SLES subscription is included with your Amazon EC2 instance cost, and is based on the vCPUs of the Amazon EC2 instance. SLES for SAP is designed specifically to run SAP workloads. It includes a longer lifecycle support with Extended Service Pack Overlap Support that provides 4.5 years of total support. SLES for SAP also offers software components and service offerings like SAP HANA high availability resource agents and cluster connector.

> **Note**
>
> SLES for SAP pricing is the cost of the software and there is not an additional cost for SLES.

Windows

Windows server on Amazon EC2 can be availed at a flat, hourly rate with no commitment (On-Demand) or through a one-time payment (Savings Plan or Reserved Instances). There is no difference in cost in terms of a Windows operating system with either of these options. You can also bring your own licence. For more information, see Microsoft Licensing on AWS.

Oracle Enterprise Linux

SAP requires you to have Oracle Linux Premier Support subscription to use Oracle Enterprise Linux operating system. For additional information, review the following resources from Oracle and SAP.

- Oracle Store
- SAP Note 2069760 - Oracle Linux 7.x SAP Installation and Upgrade (requires SAP portal access)

AWS Marketplace

AWS Marketplace is a curated digital catalog that customers can use to find, buy, deploy, and manage third-party software, data, and services to build solutions and run their businesses.

In AWS Marketplace, products can be free to use or can have associated charges. For more information, see Product pricing.

AWS Support

AWS Support offers different levels of support. For more information, see AWS Support Plan Pricing.

SAP requires you to have at least a Business level of support when running SAP workloads on AWS. To learn more about the SAP prerequisite, see SAP Note 1656250 - SAP on AWS: Support Prerequisites (requires SAP portal access).

Architecture guidance for availability and reliability of SAP on AWS

August 2021

This guide is part of a content series that provides detailed information about hosting, configuring, and using SAP technologies in the Amazon Web Services (AWS) Cloud. For more information, see SAP on AWS Technical Documentation.

Overview

This guide provides a set of architecture guidelines, strategies, and decisions for deploying SAP NetWeaver-based systems with a highly available and reliable configuration on AWS.

In this guide we cover:

- Introduction to SAP high availability and reliability
- Architecture guidelines and decision consideration
- Architecture patterns and recommended usage

This guide is intended for users who have previous experience designing high availability and disaster recovery (HADR) architectures for SAP.

This guide does not cover the business requirements determining the need for HADR and/or the implementation details for a specific partner or customer solution.

Prerequisites

Specialized knowledge

Before following the configuration instructions in this guide, we recommend familiarizing yourself with the following AWS services. (If you are new to AWS, see Getting Started with AWS.)

- Amazon EC2
- Amazon EBS

- Amazon VPC
- Amazon EFS
- Amazon S3

Recommended reading

Before reading this document, we recommend understanding key concepts and best practices from these guides:

- SAP on AWS Overview and Planning
- Getting Started with Architecting SAP on the AWS Cloud

Introduction

For decades, SAP customers protected SAP workloads on premise with two common patterns: high availability and disaster recovery. The advent of cloud computing provided an opportunity to rethink HADR capabilities for SAP, using modern architectures and technologies.

Let's recap the SAP system design and single points of failure that are part of the SAP n-tier architecture.

SAP NetWeaver architecture single points of failure

Figure 1: SAP single points of failure

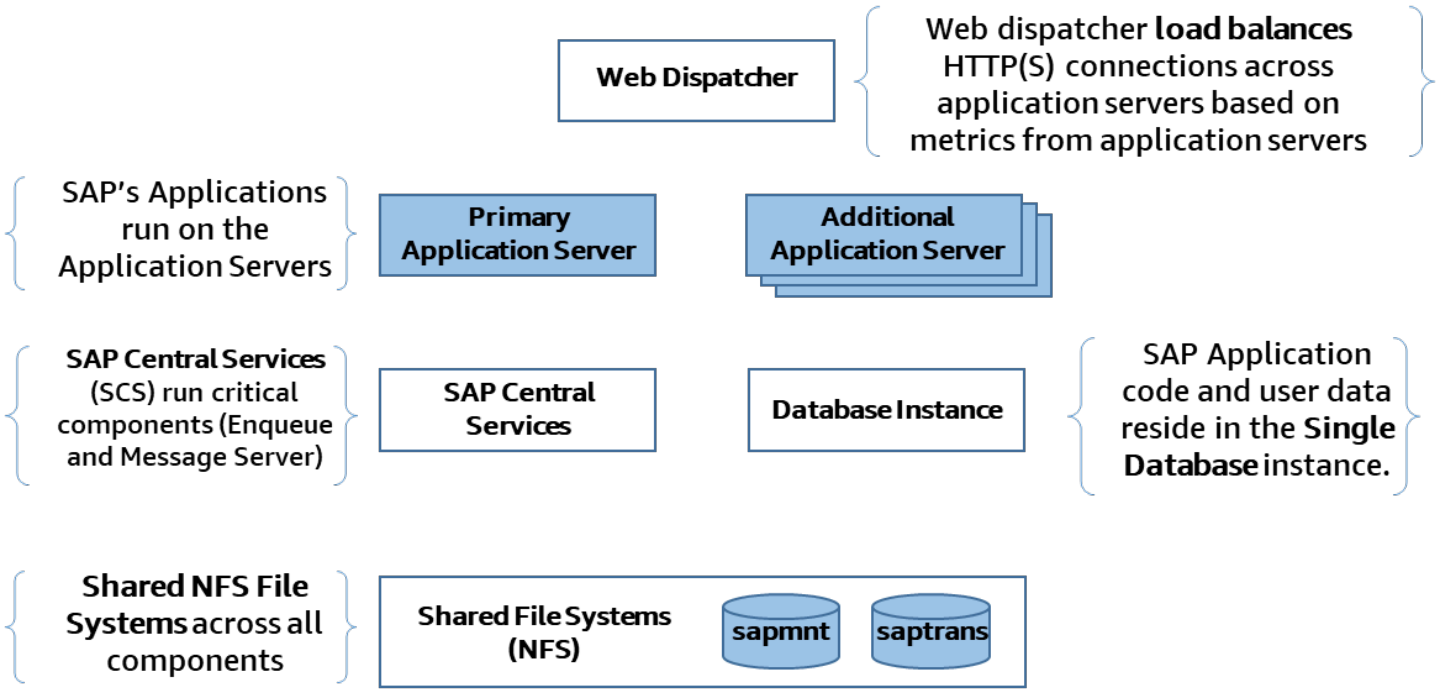


Figure 1 shows the typical SAP NetWeaver architecture, which has several single points of failure which are listed below:

- SAP Central Services (message server and enqueue processes)
- SAP Application Server
- NFS (shared storage)
- Database
- SAP Web Dispatcher

For the SAP Central Services and Database, protection can be added by deploying additional hosts. For example, an additional host running the SAP replicated enqueue can protect the loss of application level locks (enqueue locks) and an additional host running a secondary database instance can protect against data loss.

However, the inherent design of these single points of failure limits the ability to easily take advantage of cloud native features to provide high availability and reliability.

Amazon Elastic File Service (Amazon EFS) is a highly available and durable managed NFS service that runs actively across multiple physical locations (AWS Availability Zones). This service can help protect one of the SAP single points of failure.

High availability and disaster recovery

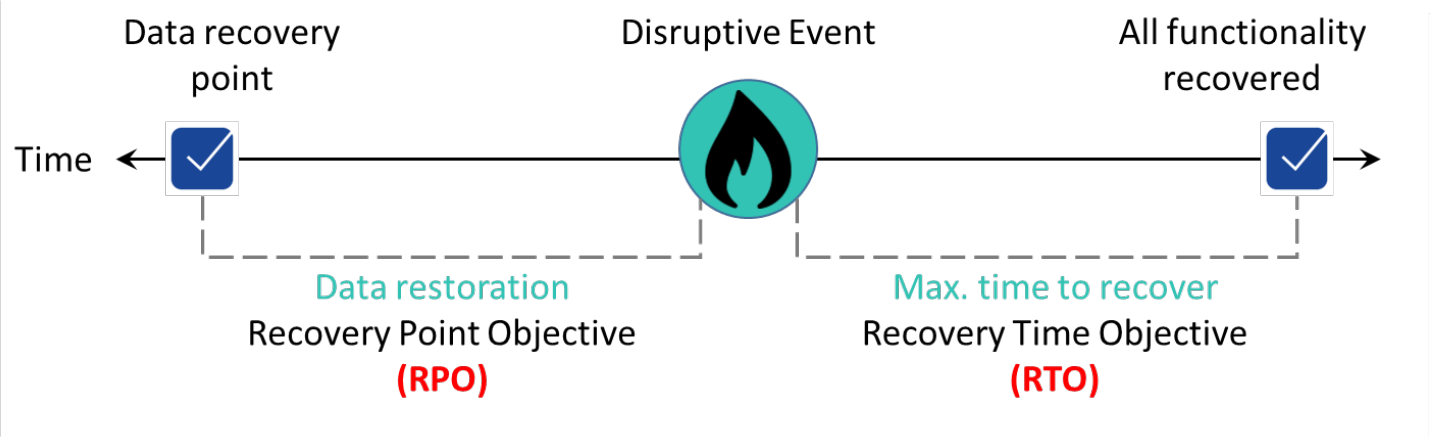
High availability (HA) is the attribute of a system to provide service during defined periods, at acceptable or agreed upon levels and to mask unplanned outages from end users. This is often achieved by using clustered servers. These servers provide automated failure detection, recovery or highly resilient hardware, robust testing, and problem and change management.

Disaster recovery (DR) protects against unplanned major outages, such as site disasters, through reliable and predictable recovery on a different hardware and/or physical location. The loss of data due to corruption or malware is considered a logical disaster event. It is normally resolved in a separate solution, such as recovery from the latest backup or storage snapshot. Logical DR does not necessarily imply a fail over to another facility.

From the perspective of documented and measurable data points, HADR requirements are often defined in terms of the following:

- **Percentage uptime** is the percentage of uptime in a given period (monthly or annual).
- **Mean time to recovery (MTTR)** is the average time required to recover from failure.
- **Return to service (RTS)** is the time it takes to bring the system back to service for the users.
- **Recovery time objective (RTO)** is the maximum length of time that a system or service can be down, how long a solution takes to recover, and the time it takes for a service to be available again.
- **Recovery point objective (RPO)** is how much data a business is willing to lose, expressed in time. It's the maximum time between a failure and the recovery point.

Figure 1: SAP single points of failure



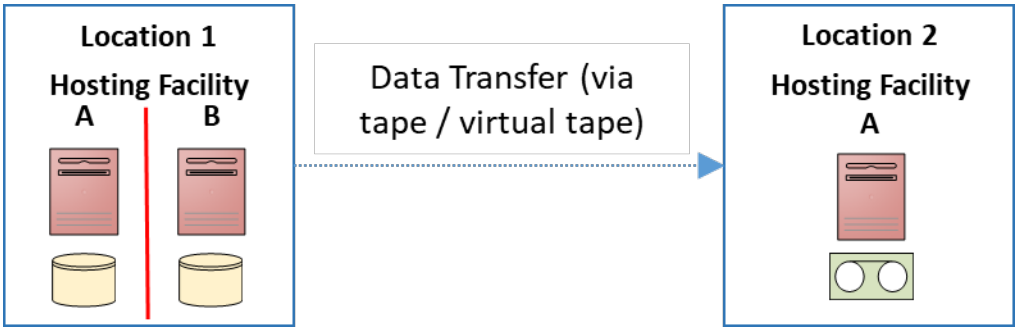
≈

On premises vs. cloud deployment patterns

Traditionally, customers with high availability requirements would deploy their primary compute capabilities in a single data center or hosting facility, often in two separate rooms or data center halls with disparate cooling and power, and high-speed network connectivity. Some customers would run two hosting facilities in close proximity, with a separation of compute capabilities, yet close enough to not be impacted by network latency.

To meet disaster recovery requirements (the preceding scenarios represent an elevated risk to unforeseen location failure), many customers would extend their architecture to include a secondary location where a copy of their data resided, with additional idle compute capacity. The distance between the primary and secondary locations often created the need for asynchronous transfer of data which impacted the recovery point objective. This was the standard and generally accepted architecture pattern for high availability and disaster recovery for many industries and companies running SAP.

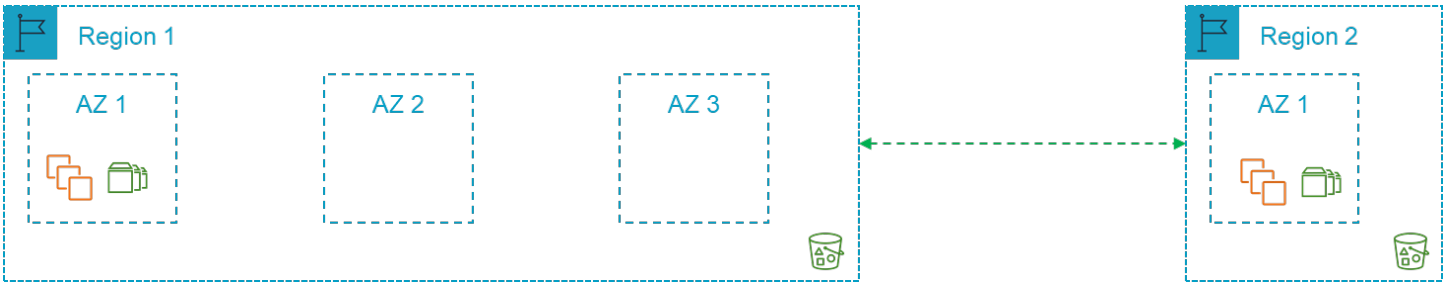
Figure 3: On-premises disaster recovery



In Figure 3, we give an example of an approach that customers often take on premises. In **Location 1**, the customer has two hosting facilities often separate rooms or halls in the same data center where they deploy a high availability architecture for the SAP single point of failure. **Location 2** is the disaster recovery location in which the SAP systems are recovered, in the event of a significant failure of both hosting facilities in **Location 1**.

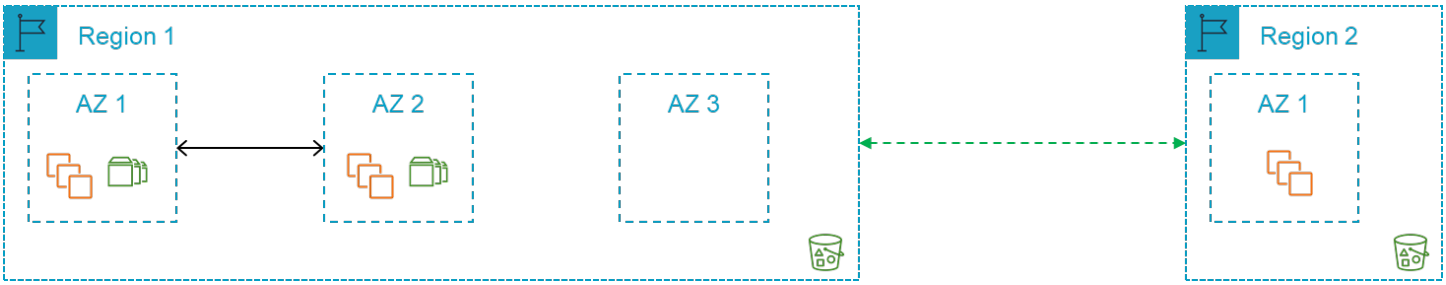
Customers migrating their SAP workloads to cloud providers still revert to this architecture and map it to AWS Regions and Availability Zones (AZs) as depicted in Figure 4. While this architecture can work in your environment, it does not follow the AWS Well-Architected Framework which helps cloud architects build secure, high-performing, resilient, and efficient infrastructure for their applications.

Figure 4: On-premises to AWS region mapping approach



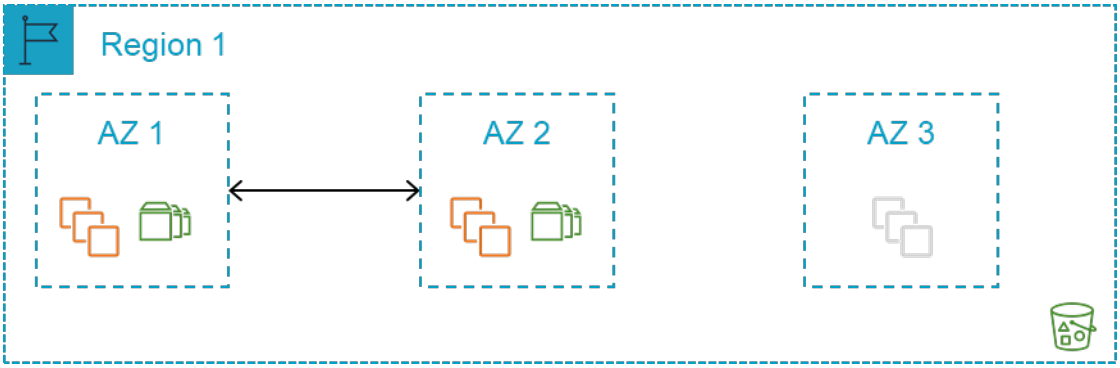
AWS isolates facilities geographically in Regions and Availability Zones. A Multi-AZ approach provides distance while maintaining performance for the primary compute capacity. This approach (Figure 5) greatly reduces the risk of location failure.

Figure 5: Alternative approach for on premises to AWS region mapping



With the risk of location failure significantly reduced for the primary compute capacity, the requirements for a second Region can be evaluated based on business requirements. You can rapidly deploy required capacity in the same or different Region with AWS. Idle hardware is no longer an issue. Data backups can be stored on Amazon Simple Storage Service (Amazon S3) in a single AWS Region or in multiple AWS Regions by leveraging cross-Region replication. This architecture can be simplified and be made readily available (Figure 6).

Figure 6: Single AWS Region approach



In addition to considering the impact of infrastructure or hosting facility failure, another scenario to consider is the loss of business data due to accidental or malicious technical activity.

Loss of business data due to accidental or malicious technical activity is referred to as *logical disaster recovery*. It requires a decision to restore the business data from a good local copy. To enable this, decisions need to be made with regard to the storage location of the data and how it will be used in the event of a *logical disaster recovery*.

Further in this guide, we detail the key architecture guidelines, architecture patterns, and decisions to consider for your availability and reliability requirements.

Architecture guidelines and decisions

This section will provides a brief overview of the AWS services typically used for SAP workloads and some of the key points to understand when designing your architecture for hosting SAP on AWS. If you are already familiar with these AWS services, you can skip this section.

Regions and Availability Zones

The AWS Global Infrastructure consists of AWS Regions and Availability Zones (AZs). For more details on the AWS Global Infrastructure, see Regions and Availability Zones.

Regions

AWS has a global footprint and ensures that customers are served across the world. AWS maintains multiple Regions in North America, South American, Europe, Asia Pacific, and the Middle East.

An AWS Region is a collection of AWS resources in a geographic area. Each Region is isolated and independent. For a list of Region names and codes, see Regional endpoints.

Regions provide fault tolerance, stability, and resilience. They enable you to create redundant resources that remain available and unaffected in the unlikely event of an outage.

AWS Regions consist of multiple Availability Zones (AZs), typically 3. An Availability Zone is a fully isolated partition of the AWS infrastructure. It consists of discrete data centers housed in separate facilities, with redundant power, networking, and connectivity.

You retain complete control and ownership over the AWS Region in which your data is physically located, making it easy to meet Regional compliance and data residency requirements.

Availability Zones

Availability Zones (AZs) enable customers to operate production applications and databases that are more highly available than would be possible from a single data center. Distributing your applications across multiple zones provides the ability to remain resilient in the face of most failure modes, including natural disasters or system failures.

Each Availability Zone can be multiple data centers. At full scale, it can contain hundreds of thousands of servers. They are fully isolated partitions of the AWS global infrastructure. With its own powerful infrastructure, an Availability Zone is physically separated from any other zones. There is a distance of several kilometers, although all are within 100 km (60 miles of each other). This distance provides isolation from the most common disasters that could affect data centers (i.e. floods, fire, severe storms, earthquakes, etc.).

All Availability Zones (AZs) within a Region are interconnected with high-bandwidth and low-latency networking, over fully redundant and dedicated metro fiber. This ensures high-throughput, low-latency networking between zones. The network performance is sufficient to accomplish synchronous replication between zones.

AWS Availability Zones (AZs) enable our customers to run their applications in a highly-available manner. To be highly available, an application needs to run in more than one location simultaneously with the exact same data, thus allowing for a seamless fail over with minimal downtime, in the event of a disaster.

Services

Our general policy is to deliver AWS services, features, and instance types to all AWS Regions within 12 months of general availability, based on customer demand, latency, data sovereignty, and other factors. You can share your interest for local Region delivery, request service roadmap information, or gain insight on service interdependency (under NDA) by contacting your AWS sales representative.

Due to the nature of the service, some AWS services are delivered globally rather than Regionally, such as Route 53, Amazon Chime, Amazon WorkDocs, Amazon WorkMail, WorkSpaces, and Amazon WorkLink.

Other services, such as Amazon Elastic Compute Cloud (Amazon EC2) and Amazon Elastic Block Store (Amazon EBS) are zonal services. When you create an Amazon EC2 or Amazon EBS resource for launch, you need to specify the required Availability Zone within a Region.

Selecting the AWS Regions

When selecting the AWS Region(s) for your SAP environment deployment, you should consider the following:

- Proximity to on-premises data centers, systems, and end users to minimize network latency.
- Data residency and compliance requirements.
- Availability of the AWS products and services that you plan to use in the Region. For more details, see Region Table .
- Availability of the Amazon EC2 instance types that you plan to use in the Region. For more details, see Amazon EC2 Instance Types for SAP.
- Pricing variation between different AWS Regions. For more details, see SAP on AWS Pricing and Optimization guide .

Multi-Region considerations

When deploying across multiple Regions, an important consideration is the associated cost and management effort for core services required in each Region such as networking, security, and audit services.

Network latency

If you decide on a multiple Region approach, you should consider the impact of any increase in the network latency to the secondary Region from your on-premises locations.

Cross-Regional data transfer

AWS provides several methods of data transfer between Regions. These methods are relevant when designing an SAP Architecture for disaster recovery. You should consider any data residency requirements when transferring data to another AWS Region, the costs associated with the data transfer (cross-Region peering and/or Amazon S3 replication), and storage in the secondary Region.

Tier 0 services

When using an AWS Region, there are a number of Tier 0 services that you need before deploying an SAP workload. These include DNS, Active Directory, and/or LDAP as well as any AWS or ISV-provided security and compliance products and services.

AWS accounts

While there is no one-size-fits-all answer for how many AWS accounts a particular customer should have, most organizations want to create more than one AWS account. Multiple accounts provide the highest level of resource and billing isolation.

In the context of SAP workloads, it is common for customers to deploy the Production environment in a separate AWS account. It helps isolate the production environment from the rest of the SAP landscape.

AWS Organizations is an account management service that enables you to consolidate multiple AWS accounts into an *organization* that you create and centrally manage. AWS Organizations includes account management and consolidated billing capabilities. It enables you to better meet the budgetary, security, and compliance needs of your business. As an administrator of an organization, you can create accounts in your organization and invite existing accounts to join the organization.

AWS Landing Zone is a solution that helps customers more quickly set up a secure, multi-account AWS environment based on AWS best practices. You can save time by automating the setup of an environment for running secure and scalable workloads while implementing an initial security baseline through the creation of core accounts and resources. It also provides a baseline environment to get started with a multi-account architecture, AWS Identity and Access Management, governance, data security, network design, and logging.

Note: The AWS Landing Zone solution is delivered by AWS Solutions Architects or Professional Services consultants to create a customized baseline of AWS accounts, networks, and security policies.

Consider using the AWS Landing Zone solution if you are looking to set up a configurable landing zone with rich customization options through custom add-ons such as, Active Directory, and change management through a code deployment and configuration pipeline.

AWS Control Tower provides the easiest way to set up and govern a secure, compliant, multi-account AWS environment based on best practices established by working with thousands of enterprises. With AWS Control Tower, your distributed teams can provision new AWS accounts quickly. Meanwhile your central cloud administrators will know that all accounts are aligned with centrally established, company-wide compliance policies.

Consider using the AWS Control Tower to set up a new AWS environment based on a landing zone with pre-configured blueprints. You can interactively govern your accounts with pre-configured guardrails.

Compute

Amazon Elastic Compute Cloud (Amazon EC2) provides scalable computing capacity in the Amazon Web Services (AWS) cloud. An Amazon EC2 instance is launched in a specific Availability Zone within a specified Amazon Virtual Private Cloud (Amazon VPC).

When the Amazon EC2 instances are deployed across two or more Availability Zones within a single Region then AWS offers an SLA of 99.99%.

Instance types

A range of Amazon EC2 instance types are supported by SAP. When selecting the instance type for your SAP workload, you should consider which tiers allow flexibility on the instance used (application tier). Also consider which tiers will require the use of a specific instance type (database tier) based on compute, memory, storage throughput, and license compliance requirements.

For the tiers with specific instance type requirements and without flexibility to change during a failure scenario, consider having a capacity reservation using Reserved Instances or On-Demand Capacity Reservations within the required Availability Zones and Regions where the instance will run. This approach is called static stability. For more information, see Static stability using Availability Zones.

Reserved Instances

Reserved Instances provide significant savings on your Amazon EC2 costs compared to on-demand instance pricing. Reserved Instances are not physical instances. They are a billing discount applied to the use of On-Demand Instances in your account. To avail the discount benefit, these on-demand instances must match certain attributes, such as instance type and Region.

When you deploy Amazon EC2 across multiple Availability Zones for high availability, we recommend that you use zonal Reserved Instances. In addition to the savings over the on-demand instance pricing, a zonal Reserved Instance provides a capacity reservation in the specified Availability Zone. This ensures that the required capacity is readily available as and when you need it.

For billing purposes, the consolidated billing feature of AWS Organizations treats all of the accounts in the organization as one account. This means that all accounts in the organization can receive the hourly cost benefit of Reserved Instances that are purchased by any other account.

Savings Plans

Savings Plans is a flexible pricing model that provides savings of up to 72% on your AWS compute usage. It offers lower prices on Amazon EC2 instance usage, regardless of instance family, size, tenancy or AWS Region. The Savings Plan model also applies to AWS Fargate and AWS Lambda usage.

Savings Plans offer significant savings over on-demand, just like the Amazon EC2 Reserved Instances, in exchange for a commitment to use a specific amount of compute power (measured in $/hour) for a one- or three-year period.

On-Demand Capacity Reservations

On-Demand Capacity Reservations enable you to reserve capacity for your Amazon EC2 instances in a specific Availability Zone for any duration. This gives you the ability to create and manage Capacity Reservations independently, with the billing discounts offered by Savings Plans or Regional Reserved Instances. You can create Capacity Reservations at any time, you ensure that you always have access to Amazon EC2 capacity when you need it, for as long as you need it. You can create Capacity Reservations at any time, without entering into a one-year or three-year term commitment, and the capacity is available immediately. When you no longer need the reservation, we recommend that you cancel the Capacity Reservation to stop incurring charges for it.

Instance Family Availability across Availability Zones

Certain Amazon EC2 instance families (for example, X1 and High Memory) are not available across all Availability Zones within a Region. You should confirm the instance types required for your SAP workload and check if they are available in your target Availability Zones.

Amazon EC2 auto recovery

Amazon EC2 auto recovery is an Amazon EC2 feature that automatically recovers the instance within the same Availability Zone, if it becomes impaired due to an underlying hardware failure or a problem that requires AWS involvement to repair.

You can enable auto recovery for Amazon EC2 instances by creating an Amazon CloudWatch alarm which monitors the instance status. Examples of problems that cause system status checks to fail include:

- Loss of network connectivity
- Loss of system power
- Software issues on the physical host
- Hardware issues on the physical host that impact network reachability

Though it typically takes under 15 minutes for a failed instance to restart, Amazon EC2 auto recovery does not offer an SLA. Therefore, if the recovery of the application that's running on the failed host is critical (for example, SAP Database or SAP Central Services), you should consider using clustering across two Availability Zones to help ensure high availability.

High Memory Bare Metal Dedicated Hosts

Amazon EC2 High Memory Instances are specifically designed to run large in-memory databases, such as SAP HANA. High Memory Bare Metal instances are available on Amazon EC2 Dedicated Hosts on a one- or three-year reservation.

High Memory instances support Dedicated Host Recovery. Host recovery automatically restarts your instances on a new replacement host if failures are detected on your Dedicated Host. Host recovery reduces the need for manual intervention and lowers the operational burden in case of an unexpected Dedicated Host failure.

We recommend a second-High Memory instance in a different Availability Zone of your chosen Region to protect against zone failure.

Amazon EC2 maintenance

When AWS maintains the underlying host for an instance, it schedules the instance for maintenance. There are two types of maintenance events:

- During network maintenance, scheduled instances lose network connectivity for a brief period of time. Normal network connectivity to your instance is restored after maintenance is complete.
- During power maintenance, scheduled instances are taken offline for a brief period, and then rebooted. When a reboot is performed, all of your instance's configuration settings are retained.

Additionally, we frequently upgrade our Amazon EC2 fleet with many patches and upgrades being applied to instances transparently. However, some updates require a short reboot. Reboots such as these should be infrequent but necessary to apply upgrades that strengthen our security, reliability, and operational performance.

There are two kinds of reboots that can be required as part of Amazon EC2 scheduled maintenance:

- Instance reboots are reboots of your virtual instance and are equivalent to an operating system reboot.
- System reboots require reboots of the underlying physical server hosting an instance.

You can view any upcoming scheduled events for your instances in the AWS Management Console or using the API tools or command line.

If you do not take any action, the impact on your instance is the same in both cases: during your scheduled maintenance window your instance will experience a reboot that in most cases takes a few minutes.

Alternatively, you can migrate your instance to a new host by performing a stop and start on your instance. For more information, see Stop and start your instance. You can automate an immediate stop and start in response to a scheduled maintenance event.

Networking

Amazon Virtual Private Cloud and subnets

An Amazon Virtual Private Cloud (Amazon VPC) is a virtual network dedicated to your AWS account. It is logically isolated from other virtual networks in the AWS Cloud. You can launch your AWS resources, such as Amazon EC2 instances, into your VPC.

When you create a VPC, you must specify a range of IPv4 addresses for the VPC in the form of a Classless Inter-Domain Routing (CIDR) block, for example, 10.0.0.0/16. This is the primary CIDR block for your VPC.

You can create a VPC within your chosen AWS Region and it will be available across all Availability Zones within that Region.

To add a new subnet to your VPC, you must specify an IPv4 CIDR block for the subnet from the range of your VPC. You can specify the Availability Zone in which you want the subnet to reside.

You can have multiple subnets in the same zone but a single subnet cannot span across multiple zones.

To provide future flexibility, we recommend that your subnet and connectivity design support all of the available Availability Zones in your account within the Region, regardless of the number of zones that you initially plan to use within a Region.

Latency across Availability Zones

All Availability Zones (AZs) are interconnected with high-bandwidth, low-latency networking, over fully redundant and dedicated metro fiber. This results in single-digit millisecond latency between resources in different Availability Zones in the same Region.

For high availability, we recommend deploying production SAP workloads across multiple Availability Zones, including the SAP Application Server Layer. If you have SAP transactions or batch jobs that involve significant database calls, we recommend that you run these transactions on SAP Application Servers located in the same Availability Zone as the database and that you use SAP Logon Groups (SMLG) for end users and batch server group (SM61) for background processing jobs. This ensures that latency-sensitive parts of the SAP workload run on the correct application servers.

On premises to AWS connectivity

You can connect to your VPC through a Site-to-Site virtual private network (VPN) or AWS Direct Connect from on premises. AWS Direct Connect provides and SLA of up to 99.99% and Site-to-Site VPN provides an SLA of 99.95%

Site-to-Site VPN connections are to specific Regions. For Direct Connect-based connections, Direct Connect Gateway allows you to connect to multiple Regions.

When establishing connectivity to AWS from on premises, ensure that you have resilient connections either through the use of multiple Direct Connect Links, multiple VPN connections, or a combination of the two.

The AWS Direct Connect Resiliency Toolkit provides a connection wizard with multiple resiliency models. These models help you order dedicated connections to achieve your SLA objective.

VPC endpoints

A VPC endpoint privately connects your VPC to supported AWS services and VPC endpoint services powered by AWS PrivateLink. It doesn't require internet access via an internet gateway, NAT device,

VPN connection, or AWS Direct Connect connection. Instances in your VPC do not require public IP addresses to communicate with resources in the AWS service. Traffic between your VPC and other services does not leave the Amazon network.

VPC endpoints are available for all of the core AWS services that are required to support an SAP-based workload, including Amazon EC2 API, Amazon S3, and Amazon Elastic File System.

Cross-Region peering

Amazon Virtual Private Cloud (Amazon VPC) supports Inter-Region peering between two VPCs in different Regions. This can be used to allow network traffic, such as database replication traffic to flow between two Amazon EC2 instances in different Regions. Inter-Region peering incurs data transfer costs.

AWS Transit Gateway is a network transit hub that you can use to interconnect your virtual private clouds (VPC) within an AWS Region to other VPCs in other AWS Regions and to on premises networks using AWS Direct Connect or VPN. Use of Transit Gateway will incur Transit Gateway costs. AWS Transit Gateway provides an SLA of 99.95% within a Region.

Load balancing

Elastic Load Balancing supports four types of load balancing: Application Load Balancers, Network Load Balancers, Gateway Load Balancers, and Classic Load Balancers.

A Network Load Balancer can be used to support a high-availability deployment of SAP Web Dispatchers and/or SAP Central Services across multiple Availability Zones. For more details, see Overlay IP Routing with Network Load Balancer.

A ***load balancer*** serves as the single point of contact for clients. The load balancer distributes incoming traffic across multiple targets, such as Amazon EC2 instances.

A ***listener*** checks for connection requests from clients, using the protocol and port that you configure, and forwards requests to a target group.

Each ***target group*** routes requests to one or more registered targets, such as Amazon EC2 instances, using the TCP protocol and the specified port number. You can configure health checks on a per target group basis. Health checks are performed on all targets registered to a target group that is specified in a listener rule for your load balancer.

For TCP traffic, the Network Load Balancer selects a target using a flow hash algorithm based on the protocol, source IP address, source port, destination IP address, destination port, and TCP

sequence number. Each individual TCP connection is routed to a single target for the life of the connection.

DNS

Amazon Route 53 is a highly available and scalable Domain Name System (DNS) web service. You can use Route 53 to perform three main functions in any combination: domain registration, DNS routing, and health checking. Route 53 offers an SLA of 100%.

Amazon Route 53 Resolver provides a set of features that enable bi-directional querying between on premises and AWS over private connections.

Storage

Object storage

Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance. Amazon S3 is a Regional service across all Availability Zones within a Region and is designed for 99.999999999% (11 9's) of durability and an SLA of 99.9%.

To protect against data loss, you can perform backups (such as database backups or file backups) to Amazon S3. Additionally, Amazon EBS Snapshots and Amazon Machine Images (AMIs) are stored in Amazon S3.

Amazon S3 Replication enables automatic, asynchronous copying of objects across Amazon S3 buckets. Buckets that are configured for object replication can be owned by the same AWS account or by different accounts.

Amazon S3 replication

You can replicate objects between the same or different AWS Regions.

- Cross-Region replication (CRR) is used to copy objects across Amazon S3 buckets in different AWS Regions.
- Same-Region replication (SRR) is used to copy objects across Amazon S3 buckets in the same AWS Region.

Cross-Region replication incurs the following costs:

- Data Transfer charges for the data transferred between the first and second AWS Regions
- Amazon S3 charges for the data stored in Amazon S3 in the two different AWS Regions

Additionally, you can enable Amazon S3 Replication Time Control with cross-Region replication. Amazon S3 Replication Time Control (Amazon S3 RTC) helps you meet compliance or business requirements for data replication and provides visibility into Amazon S3 replication times. Amazon S3 RTC replicates most objects that you upload to Amazon S3 in seconds, and 99.99 percent of those objects within 15 minutes.

Amazon S3 RTC incurs the following costs in addition to the costs listed above for cross-Region replication:

- Amazon S3 RTC Management Feature - priced per GB
- Amazon CloudWatch Amazon S3 Metrics - priced by number of metrics

Same-Region replication incurs the following costs:

- Charges for the data stored in Amazon S3

Block storage

Amazon Elastic Block Store (Amazon EBS) provides block level storage volumes for use with Amazon EC2 instances. Amazon EBS volumes behave like raw, unformatted block devices. You can mount these volumes as devices on your instances. You can create a file system on top of these volumes or use them in any way that you would use a block device (like a hard drive). You can dynamically change the configuration of a volume that's attached to an instance.

Amazon EBS volumes are placed in a specific Availability Zone where they are automatically replicated to protect you from the failure of a single component. All Amazon EBS volume types offer durable snapshot capabilities and are designed for 99.999% availability per volume and 99.99% service availability with Multi-AZ configuration. The use of a database replication capability, block level replication solution or Amazon EBS Snapshots is required to provide durability of the SAP data stored on Amazon EBS across multiple Availability Zones.

Amazon EBS volumes are designed for an annual failure rate (AFR) of between 0.1% - 0.2%, where failure refers to a complete or partial loss of the volume, depending on the size and performance of the volume. This makes Amazon EBS volumes 20 times more reliable than typical commodity

disk drives, which fail with an AFR of around 4%. For example, if you have 1,000 Amazon EBS volumes running for 1 year, you should expect 1 to 2 will have a failure.

Amazon EBS offers a number of different volume types. It must be used for the SAP database-related data, General Purpose SSD (gp2) or Provisioned IOPS SSD (io1) must be used. The throughput and IOPS requirement will determine if gp2 or io1 is required.

Amazon EBS Multi-Attach enables you to attach a single Provisioned IOPS SSD (io1) volume to up to 16 AWS Nitro-based instances that are in the same Availability Zone. You can attach multiple Multi-Attach enabled volumes to an instance or set of instances. Each instance to which the volume is attached has full read and write permission to the shared volume. Multi-Attach enabled volumes do not support I/O fencing. I/O fencing protocols control write access in a shared storage environment to maintain data consistency. Your applications must provide write ordering for the attached instances to maintain data consistency.

Amazon EBS snapshots

You can back up the data on your Amazon EBS volumes to Amazon S3 by taking point-in-time snapshots. Snapshots are *incremental* backups, which means that only the blocks on the device that have changed after your most recent snapshot are saved. This minimizes the time required to create the snapshot and saves on storage costs by not duplicating data. When you delete a snapshot, only the data unique to that snapshot is removed. Each snapshot contains all of the information that is needed to restore your data (from the moment when the snapshot was taken) to a new Amazon EBS volume.

Amazon EBS Snapshots can be copied (replicated) to a different Region and/or shared with a different AWS Account.

Copying Snapshots across Regions incurs the following costs:

- Data Transfer charges for the data transferred between the first and second AWS Regions
- Amazon EBS Snapshot charges for the data stored in Amazon S3 in the two different AWS Regions

Restoring snapshots

New volumes created from existing Amazon EBS snapshots load lazily in the background. This means that after a volume is created from a snapshot, there is no need to wait for all of the data to transfer from Amazon S3 to your Amazon EBS volume before your attached instance can start accessing the volume and all its data.

This preliminary action takes time and can significantly increase the latency of I/O operations. If your instance accesses data that hasn't yet been loaded, the volume immediately downloads the requested data from Amazon S3, and then continues loading the rest of the volume data in the background.

Fast snapshot restore

Amazon EBS fast snapshot restore enables you to create a volume from a snapshot that is fully-initialized at creation. This eliminates the latency of I/O operations on a block when it is accessed for the first time. Volumes created using fast snapshot restore instantly deliver all of their provisioned performance. To use fast snapshot restore, enable it for specific snapshots in specific Availability Zones. Fast Snapshot Restore is charged in Data Services Unit-Hours (DSUs) for each zone in which it is enabled. DSUs are billed per minute with a 1-hour minimum.

File storage

Amazon EFS

Amazon Elastic File System (Amazon EFS) provides scalable NFS version 4 based file storage for use with Linux-based Amazon EC2 (Windows-based Amazon EC2 instances do not support Amazon EFS). The service is designed to be highly scalable, available, and durable. Amazon EFS file systems store data and metadata across multiple Availability Zones in an AWS Region. Amazon EFS offers an SLA of 99.99%.

Amazon EFS file systems can be shared across accounts and VPCs within the same Region or a different Region, enabling Amazon EFS to be an ideal choice for SAP global file system (/sapmnt) and SAP transport directory (/usr/sap/trans).

AWS DataSync supports Amazon EFS to Amazon EFS transfer between Regions and different AWS Accounts, allowing the replication of key SAP file based data across Regions. AWS Backup can also be used to replicate backups of Amazon EFS file systems across Regions.

Amazon FSx

Amazon FSx for Windows File Server provides fully managed Microsoft Windows file servers, backed by a fully native Windows file system. Amazon FSx offers an SLA of 99.9% and supports both Single-AZ and Multi-AZ File Systems.

With Single-AZ file systems, Amazon FSx automatically replicates your data within an Availability Zone, continuously monitors for hardware failures and automatically replaces infrastructure

components in the event of a failure. Amazon FSx also takes highly durable backups of your file system daily using Windows' Volume Shadow Copy Service that are stored in Amazon S3. You can take additional backups at any point.

Multi-AZ file systems support all the availability and durability features of Single-AZ file systems. In addition, they are designed to provide continuous availability to data, even when an Availability Zone is unavailable. In a Multi-AZ deployment, Amazon FSx automatically provisions and maintains a standby file server in a different zone. Any changes written to disk in your file system are synchronously replicated across Availability Zones to the standby.

Amazon FSx File systems can be shared across Accounts and VPCs within the same Region or a different Region, enabling Amazon FSx to be used not only for the SAP Global File System but also the SAP Transport Directory.

Additionally, Amazon FSx can also be used for providing Continuously Available (CA) File Shares for Microsoft SQL Server.

Monitoring and audit

Amazon CloudWatch

Amazon CloudWatch is a monitoring and observability service built for DevOps engineers, developers, site reliability engineers (SREs), and IT managers. CloudWatch provides you with data and actionable insights to monitor your applications, respond to system-wide performance changes, optimize resource utilization, and get a unified view of operational health. CloudWatch collects monitoring and operational data in the form of logs, metrics, and events, providing you with a unified view of AWS resources, applications, and services that run on AWS and on-premises servers. You can use CloudWatch to detect anomalous behavior in your environments, set alarms, visualize logs and metrics side by side, take automated actions, troubleshoot issues, and discover insights to keep your applications running smoothly.

AWS CloudTrail

AWS CloudTrail is a service that enables governance, compliance, operational auditing, and risk auditing of your AWS account. With CloudTrail, you can log, continuously monitor, and retain account activity related to actions across your AWS infrastructure. CloudTrail provides event history of your AWS account activity, including actions taken through the AWS Management Console, AWS SDKs, command line tools, and other AWS services. This event history simplifies security analysis, resource change tracking, and troubleshooting. In addition, you can use CloudTrail to

detect unusual activity in your AWS accounts. These capabilities help simplify operational analysis and troubleshooting.

Architecture patterns

In this section, we elaborate on the architecture patterns that you can select based on your availability and recovery requirements. We also analyze failure scenarios that can help you select the right pattern for your SAP system(s).

Failure scenarios

For the failure scenarios below, the primary consideration is the physical unavailability of the compute and/or storage capacity within the Availability Zones.

Availability Zone failure

An Availability Zone failure can be caused by a significant availability degradation of one or more AWS services utilized by your resources within that Availability Zone. For example:

- Several Amazon EC2 instances have failed with System Status Check errors or are unreachable and cannot be restarted.
- Several Amazon Elastic Block Store (Amazon EBS) volumes with Volume Status Check errors have failed.

Amazon Elastic Block Store failure

Loss of one or more Amazon EBS volumes attached to a single Amazon EC2 instance may result in the unavailability of a critical component (i.e. the database) of the SAP system.

Amazon EC2 failure

Loss of a single Amazon EC2 instance may result in the unavailability of a critical component (i.e. the database or SAP Central Services) of the SAP system.

Logical data loss

You should also consider the potential for logical data loss where the underlying hardware capacity still exists but the primary copies of the data have been corrupted or lost. This data loss could be due to malicious activity within your AWS account or due to human error.

To protect against logical data loss, it is recommended that regular copies of the data are backed up to an Amazon S3 bucket. This bucket is replicated (using Single-Region or Cross-Region replication) to another Amazon S3 bucket owned by a separate AWS account. With the appropriate AWS Identity and Access Management (IAM) controls between the two AWS accounts, this strategy ensures that not all copies of the data are lost due to malicious activity or human error.

Patterns

In this section, we examine the architecture patterns available to handle the failure scenarios detailed above.

There are two key parameters to consider when selecting a pattern to meet your organization's specific business requirements:

- Availability of compute for the SAP single points of failure
- Availability of the SAP data persisted on Amazon EBS

These parameters determine the time taken to recover from a failure scenario, that is, the time taken by your SAP system to return to service.

Types of architecture patterns

The architecture patterns are grouped into single Region and multi-Region patterns. The distinguishing factor would be:

1. If you require the data to only reside in a specific geographical location (AWS Region) at all times (for example, data residency requirements).

   or

2. If you require the data to reside in two specific geographical locations (AWS Regions) at all times (for example, two copies of SAP data must reside at least 500 miles apart for compliance).

If your production systems are critical to your business and you require minimal downtime in the event of failure, you should select a Multi-Region pattern to ensure that your production systems are highly available at all times. When deploying a Multi-Region pattern, you can benefit from using an automated approach (such as, Cluster solution) for fail over between Availability Zones to minimize the overall downtime and remove the need for human intervention. Multi-Region

patterns not only provide high availability but also disaster recovery, thereby lowering overall costs.

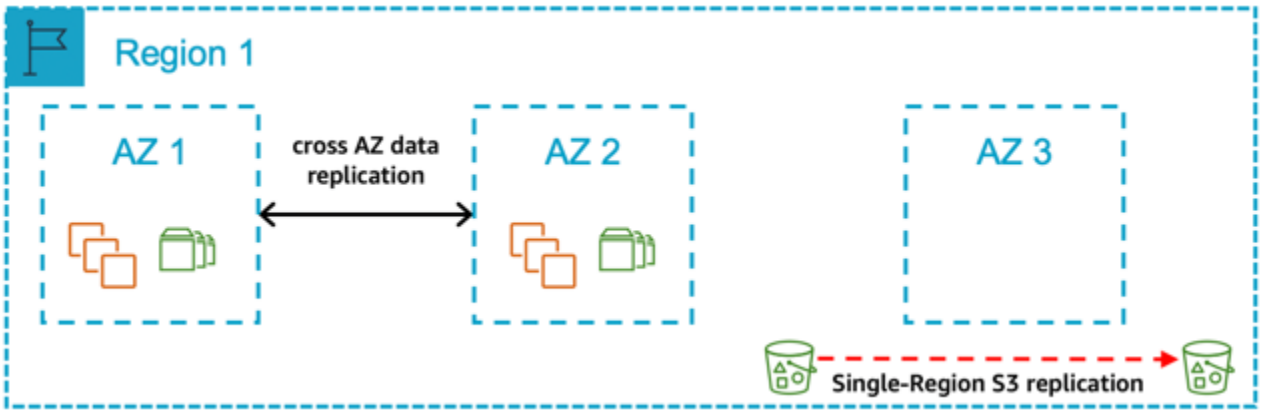
Single Region architecture patterns

Select a single Region pattern if:

- You require the data to reside only in a specific geographical Region (AWS Region) at all times
- You want to avoid the potential network latency considerations associated with a Multi-Region approach
- You want to avoid the cost implications or differences associated with a Multi-Region approach including:
  - AWS service pricing in different AWS Regions
  - Cross-Region data transfer costs

Pattern 1: A single Region with two AZs for production

Figure 7: A single Region with two Availability Zones for production



In this pattern, you deploy all your production systems across two Availability Zones. The compute deployed for the production SAP database and central services tiers are the same size in both Availability Zones, with automated fail over in the event of a zone failure. The compute required for the SAP application tier is split 50/50 between two zones. Your non-production systems are **not** an equivalent size to your production and are deployed in the same zones or a different Availability Zone within the Region.

Select this pattern if:

- You require a defined time window to complete recovery of production as well as assurance of the availability of compute capacity in another Availability Zone for the production SAP database and central services tiers.
- You can accept the additional cost of deploying the required compute and storage for production SAP database and central services tiers across two Availability Zones.
- Your non-production environment is not of equivalent size as production and therefore cannot be used as sacrificial capacity for production in the event of an Availability Zone failure or significant Amazon EC2 service degradation.
- You can accept data replication across Availability Zones (database replication capability or a block level replication solution required) and the associated cost.
- You can accept that automated fail over between Availability Zones requires a third-party cluster solution.
- You can accept the variable time duration required (including any delay in availability of the required compute capacity in the remaining Availability Zones) to return the application tier to 100% capacity in the event of a zone failure.

Key design principles

- 100% compute capacity deployed in Availability Zone 1 and Availability Zone 2 for production SAP database and central services tiers.
- Compute capacity is deployed in Availability Zone 1 and Availability Zone 2 for production application tier (Active/Active). In the event of an Availability Zone failure, the application tier needs to be scaled to return to 100% capacity within the remaining zone.
- The SAP Database is persisted on Amazon EBS in two Availability Zones using either a database replication capability or a block level replication solution.
- Amazon EC2 auto recovery is configured for all instances to protect against underlying hardware failure, with the exception of instances protected by a third-party cluster solution.
- Amazon EFS is used for the SAP Global File Systems.
- SAP Database is backed up regularly to Amazon S3.
- Amazon S3 single-Region replication is configured to protect logical data loss.
- Amazon Machine Image/Amazon EBS Snapshots are taken for all servers on a regular basis.

Benefits

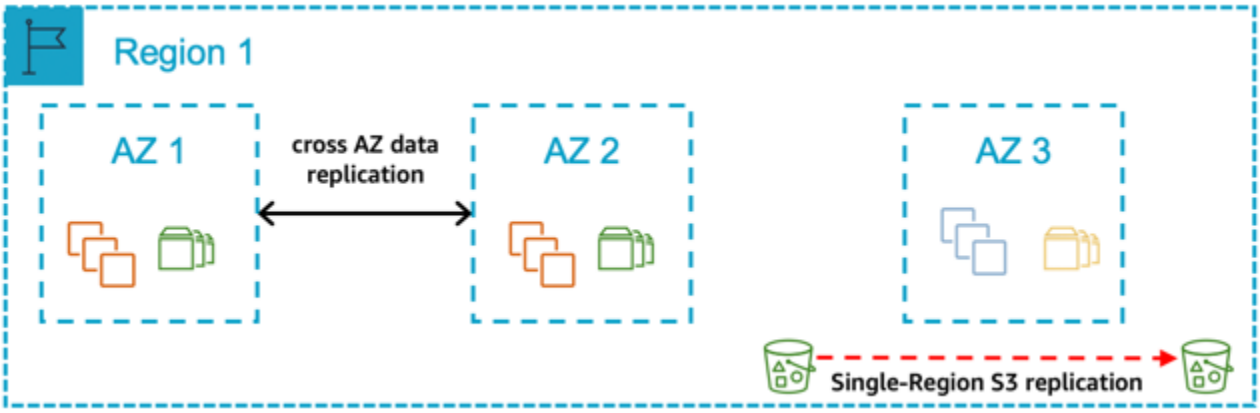
- Low Mean Time to Recovery (MTTR)
- Predictable Return to Service (RTS)
- Ability to protect against significant degradation or total Availability Zone failure through fail over of database and central services tiers to Availability Zone 2
- No requirement to restore data from Amazon S3 in the event of an Availability Zone or Amazon EBS failure

Considerations

- Well documented and tested processes are required for the automated fail over between Availability Zones.
- Well documented and tested processes are required for maintaining the automated fail over solution.
- Well documented and tested processes are required for scaling the AWS resources to return the application tier to required capacity in the event of an Availability Zone failure or significant Amazon EC2 service degradation.

Pattern 2: A single Region with two AZs for production and production sized non-production in a third AZ

Figure 8: A single Region with two Availability Zones for production and production sized non-production in a third Availability Zone



In this pattern, you deploy all your production systems across two Availability Zones. The compute deployed for the production SAP database and central services tiers are the same size in both Availability Zones, with automated fail over in the event of a zone failure. The compute required for the SAP application tier is split 50/50 between two Availability Zones. Your non-production

systems are an equivalent size to your production and deployed in a third Availability Zone. In the event of an Availability Zone failure where your production systems are deployed, the non-production capacity is reallocated to enable production to be returned to a Multi-AZ pattern.

Select this pattern if:

- You require the ability to continue to have a Multi-AZ configuration for production in the event of an Availability Zone failure within the Region.
- You require a defined time window to complete recovery of production and assurance of the availability of the compute capacity in another Availability Zone for the production SAP database and central services tiers.
- You can accept the additional cost of deploying the required compute and storage for production SAP database and central services tiers across two Availability Zones.
- You can accept data replication across Availability Zones (database replication capability or a block level replication solution required) and the associated cost.
- You can accept that automated fail over between Availability Zones requires a third-party cluster solution.
- You can accept the variable time duration required (including any delay in availability of the required compute capacity in the remaining Availability Zones) to return the application tier to 100% capacity in the event of an Availability Zone failure.

Key design principles

- 100% compute capacity is deployed in Availability Zone 1 and Availability Zone 2 for production SAP database and central services tiers.
- 100% production compute capacity (database and central services) is deployed in the third Availability Zone for use by non-production in normal operations.
- Compute capacity is deployed in Availability Zone 1 and Availability Zone 2 for production application tier (Active/Active). In the event of an Availability Zone failure, the application tier needs to be scaled to return to 100% capacity within the remaining zone.
- Amazon EC2 auto recovery is configured for all instances to protect against underlying hardware failure, with the exception of instances protected by a third-party cluster solution.
- The SAP Database is persisted on Amazon EBS in two Availability Zones using either a database replication capability or a block level replication solution.
- Amazon EFS is used for the SAP Global File Systems.

- SAP Database is backed up regularly to Amazon S3.
- Amazon S3 single-Region replication is configured to protect against logical data loss.
- Amazon Machine Image/Amazon EBS Snapshots for all servers are taken on a regular basis.

Benefits

- Low Mean Time to Recovery (MTTR)
- Predictable Return to Service (RTS)
- Ability to protect against significant degradation or total Availability Zone failure through fail over of database and central services tiers to Availability Zone 2
- No requirement to restore data from Amazon S3 in the event of an Availability Zone failure or Amazon EBS failure
- Option for data to be persisted on Amazon EBS in three different Availability Zones, dependent on capabilities of database or block level replication solution
- Use of non-production compute capacity to return to production run across two Availability Zones in the event of a significant degradation or total Availability Zone failure

Considerations

- Well documented and tested processes are required for the automated fail over between Availability Zones.
- Well documented and tested processes are required for maintaining the automated fail over solution.
- Well documented and tested processes are required for scaling the AWS resources to return the application tier to required capacity in the event of an Availability Zone failure or significant Amazon EC2 service degradation.
- Well documented and tested processes are required for re-allocating the compute capacity from non-production to return production to run across two Availability Zones in the event of an Availability Zone failure impacting production.

Pattern 3: A single Region with one AZ for production and another AZ for non-production

Figure 9: A single Region with one Availability Zone for production and another Availability Zone for non-production

In this pattern, you deploy all your production systems in one Availability Zone and all your non-production systems in another Availability Zone. Your non-production systems are an equivalent size to your production.

Select this pattern if:

- You require a defined time window to complete recovery of production and assurance of the availability of compute capacity in another Availability Zone for the SAP database and central services tiers.
- You can accept the additional time required to re-allocate compute capacity from non-production to production as part of the overall time window to recover production.
- You can accept the time required to restore data to Amazon EBS from Amazon S3 in another Availability Zone as part of the overall time window to recover production.
- You can accept the variable time duration required to return the application tier to 100% capacity following an Availability Zone failure (including any delay in availability of the required compute capacity in the remaining Availability Zones).
- You can accept a period of time where there is only one set of computes deployed for the production SAP database and central services tiers in the event of an Availability Zone failure or significant Amazon EC2 service degradation.

Key design principles

- 100% compute capacity is deployed in Availability Zone 1 for production SAP database and central services tiers.
- 100% compute capacity is deployed in Availability Zone 1 for production SAP application tier.

- 100% of production compute capacity (SAP database and central services) is deployed in Availability Zone 2 for use by non-production in normal operations.
- Amazon EC2 auto recovery is configured for all instances to protect against underlying hardware failure.
- The SAP database is persisted on Amazon EBS in a single Availability Zone only and not replicated on another Availability Zone.
- Amazon EFS is used for the SAP Global File Systems.
- SAP Database Data is backed up regularly to Amazon S3.
- Amazon S3 single-Region replication is configured to protect against logical data loss.
- Amazon Machine Image/Amazon EBS Snapshots are taken for all servers on a regular basis.

Benefits

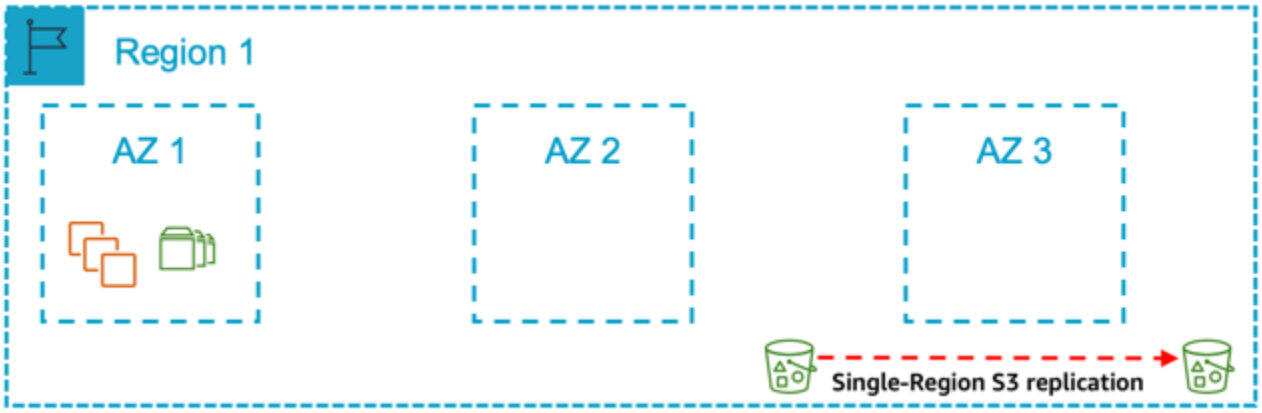
- Cost optimized through use of non-production capacity in the event of production Availability Zone failure
- Required compute capacity deployed in two Availability Zones to allow a more predictable recovery time duration

Considerations

- Well documented and tested processes for re-allocating the required compute capacity from non-production to production and restoring the data in a different Availability Zone are required to ensure recoverability.
- There may be loss of non-production environments in the event of an Availability Zone failure impacting production.
- Due to the lack of high availability across two Availability Zones, the time required to recover production in the event of compute failure or Availability Zone failure increases.

Pattern 4: A single Region with a single AZ for production

Figure 10: A single Region with a single Availability Zone for production



In this pattern, you deploy all your production systems in one Availability Zone and all your non-production systems in either the same Availability Zone or another Availability Zone. Your non-production systems are **not** a similar size to your production.**

Select this pattern if:

- In the event of an Availability Zone failure or significant Amazon EC2 service degradation, you can accept the risks related to the variable time duration required (including any delay in availability of the required compute capacity in the remaining Availability Zones) to re-create the AWS resources in a different Availability Zone and restore the persistent data to Amazon EBS.
- You want to avoid the cost implications with a Multi-AZ approach and accept the related risks of downtime of your production SAP systems.

Key design principles

- 100% compute capacity is deployed in Availability Zone 1 for production SAP database and central services tiers.
- 100% compute capacity is deployed in Availability Zone 1 for production SAP application tier.
- Amazon EC2 is configured for all instances to protect against underlying hardware failure.
- Deployed non-production compute capacity is less than 100% the compute capacity deployed for production SAP database and central services tiers.
- The SAP database is persisted on Amazon EBS in a single Availability Zone only and not replicated on another Availability Zone.
- Amazon EFS is used for the SAP Global File Systems.
- SAP Database is backed up regularly to Amazon S3.

- Amazon S3 single-Region replication is configured to protect against logical data loss.
- Amazon Machine Image/Amazon EBS Snapshots for all servers are taken on a regular basis.

Benefits

- Lowest cost
- Simplest design
- Simplest operation

Considerations

- Well documented and tested processes for scaling the AWS resources and restoring data in a different Availability Zone are required to ensure recoverability.

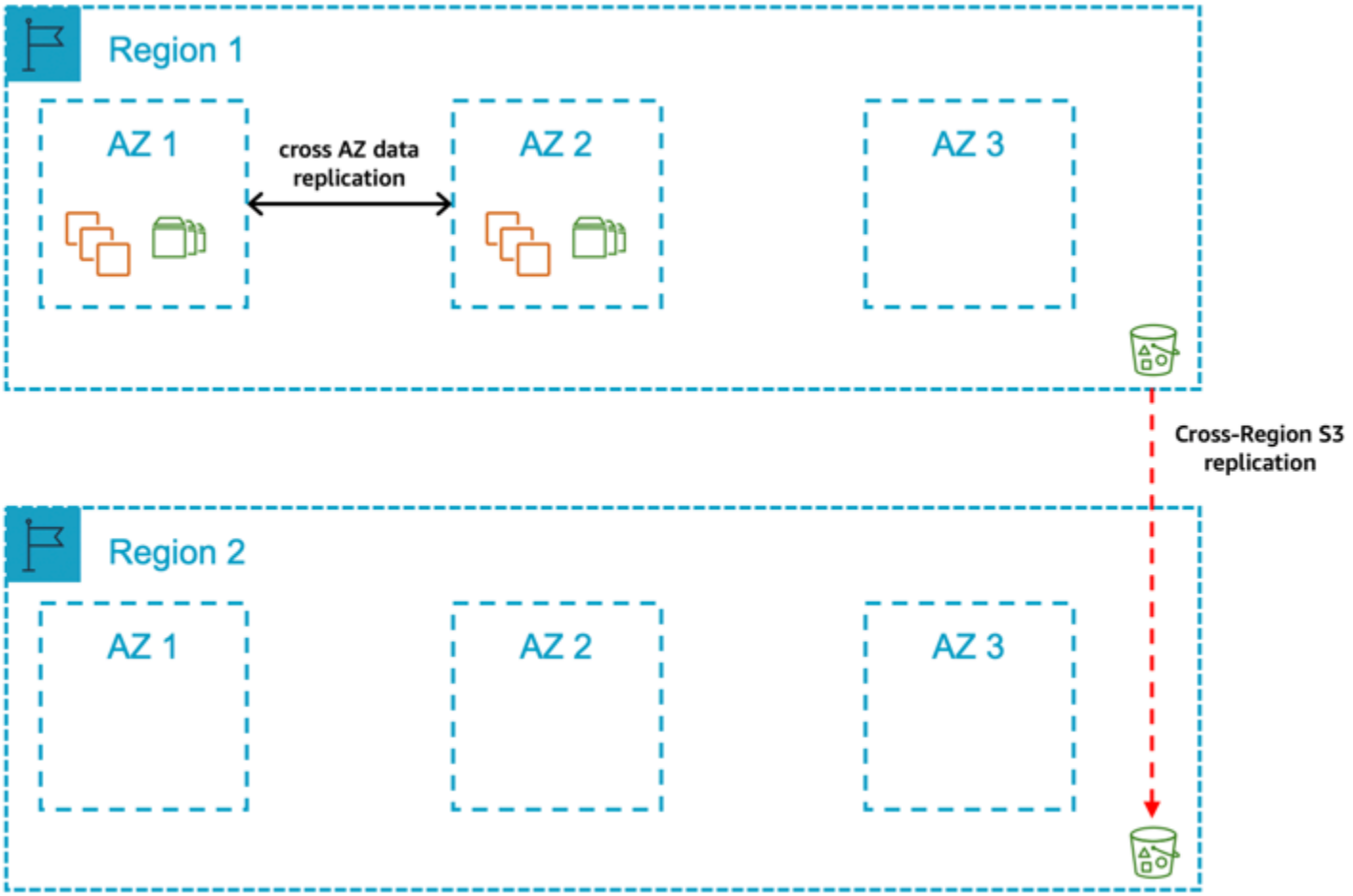
Multi-Region Architecture Patterns

You should select a multi-Region architecture if you require the following:

- You require the data to reside in two specific geographical AWS Regions at all times.
- You can accept the potential network latency considerations associated with a multi-Region approach.
- You can accept the increased complexity associated with multi-Region approach.
- You can accept the cost implications / differences associated with a multi-Region approach including:
  - AWS service pricing (e.g. Amazon EC2) in different AWS Regions
  - Cross-Region data transfer costs
- Additional compute and/or storage costs in the second Region.

Pattern 5: A primary Region with two AZs for production and a secondary Region containing a replica of backups/AMIs

Figure 11: A primary Region with two Availability Zones for production and a secondary Region containing a replica of backups/AMIs



In this pattern, you deploy your production system across two Availability Zones in the primary Region. The compute deployed for the production SAP database and central services tiers are the same size in both Availability Zones with automated fail over in the event of an Availability Zone failure. The compute required for the SAP application tier is split 50/50 between two Availability Zones. Additionally, the production database backups stored in Amazon S3, Amazon EBS Snapshots, and Amazon Machine Images are replicated on the secondary Region. In the event of a complete Region failure, the production systems would be restored from the last set of backups in the second Region.

Select this pattern if:

- You require a defined time window to complete recovery of production and assurance of the availability of compute capacity in another Availability Zone within the primary Region for the production SAP database and central services tiers.

- You can can accept the additional cost of deploying the required compute and storage for production SAP database and central services tiers across two Availability Zones within the primary Region.
- You can accept the cross-Availability Zone related data transfer costs for data replication.
- You can accept that automated fail over between Availability Zones requires a third-party cluster solution.
- You can allow for a period of time where there is only one set of computes deployed for the SAP database and central services in the event of an Availability Zone failure or significant Amazon EC2 failure.
- You can accept that data replication across Availability Zones requires either a database replication capability or a block level replication solution.
- You can accept the variable time duration required (including any delay in availability of the required compute capacity in the remaining Availability Zones) to return the application tier to 100% capacity.
- You can accept the variable time duration required to complete recovery of production in the event of a Region failure.
- You can accept the increased complexity and costs associated with multi-Region approach.
- You can accept that manual actions are required to restore production in the second Region.

Key design principles

- 100% compute capacity is deployed in Availability Zone 1 and Availability Zone 2 for production SAP database and central services tiers.
- Compute capacity is deployed in Availability Zone 1 and Availability Zone 2 for production SAP application tier (Active/Active) and needs to be scaled in the event of an Availability Zone failure or significant Amazon EC2 service degradation.
- Amazon EC2 auto recovery is configured for all instances to protect against underlying hardware failure with the exception of instances protected by a third-party cluster solution.
- The SAP database-related data on Amazon EBS is replicated between Availability Zones using either a database replication capability or a block level replication solution.
- Amazon EFS is used for the SAP Global File Systems and is replicated on the secondary Region.
- SAP Database data is backed up regularly to Amazon S3.
- Amazon Machine Image/Amazon EBS Snapshots are taken for all servers on a regular basis

- Amazon S3 Data (database backups), Amazon EBS Snapshots and Amazon Machine Images are replicated/copied to a secondary Region to protect logical data loss.

Benefits

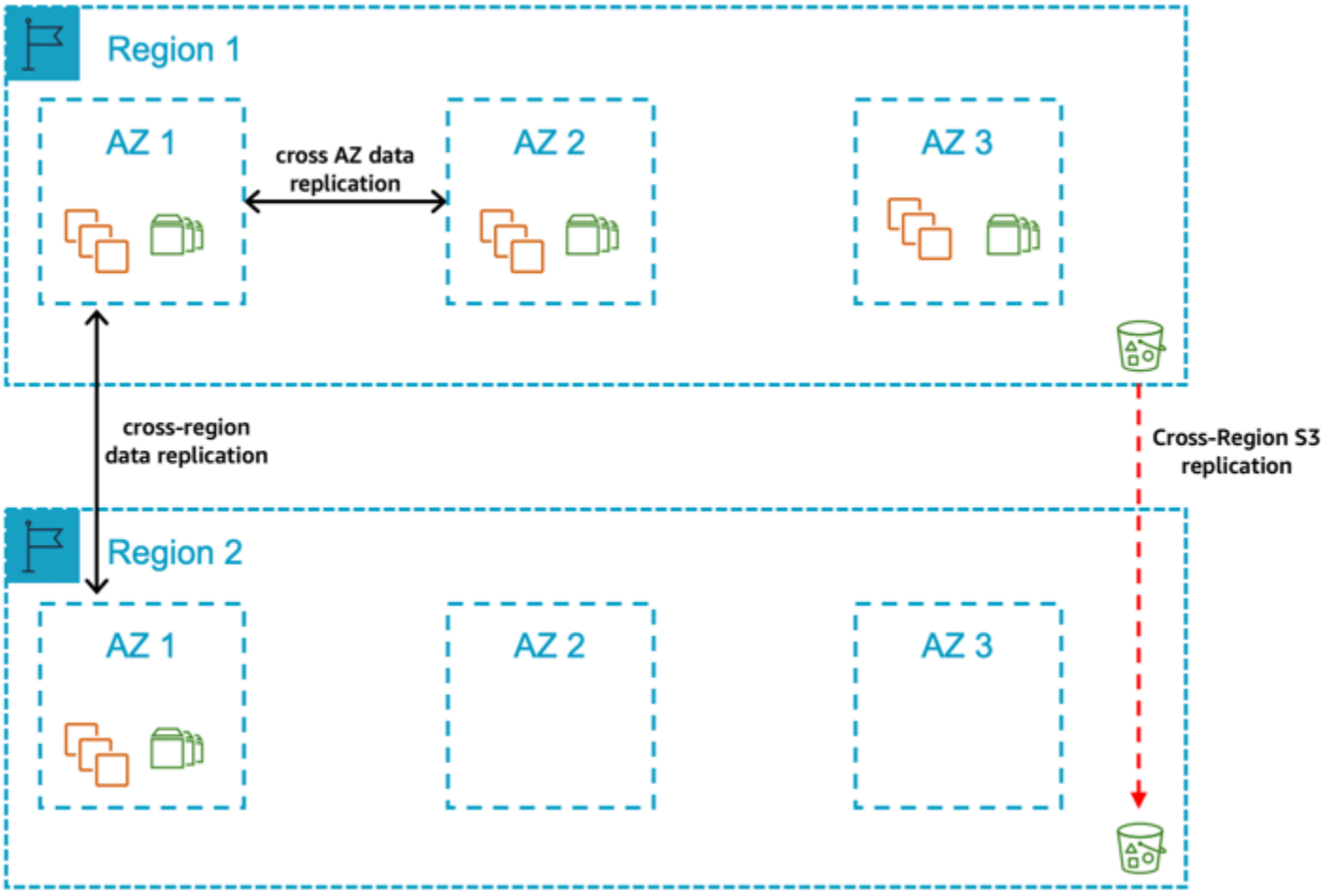
- Low Mean Time to Recovery (MTTR) in the event of Amazon EC2 or Availability Zone failure
- Predictable Return to Service (RTS) in the event of Amazon EC2 or Availability Zone failure
- Database-related data persisted on different sets of Amazon EBS volumes in two Availability Zones via database replication capability or a block level replication solution
- Required compute capacity deployed in two Availability Zones in primary Region
- No dependency on restoring data from Amazon S3 in the event of an Availability Zone failure in the primary Region
- Ability to protect against significant degradation or total Availability Zone failure through fail over to Availability Zone 2 of database and central services tiers
- Ability to protect against significant degradation or total Region failure through fail over to secondary Region

Considerations

- Well documented and tested processes are required for the automated fail over between Availability Zones.
- Well documented and tested processes are required for maintaining the automated fail over solution.
- Well documented and tested processes are required for scaling the AWS resources to return the application tier to full capacity in the event of an Availability Zone failure or significant Amazon EC2 service degradation.
- Well documented and tested processes are required for scaling the AWS resources, restoring the data, and moving production to the secondary Region.
- Higher network latency from your on-premises locations to the secondary AWS Region may impact end user performance.

Pattern 6: A primary Region with two AZs for production and a secondary Region with compute and storage capacity deployed in a single AZ

Figure 12: A primary Region with two Availability Zones for production and a secondary Region with compute and storage capacity deployed in a single Availability Zone



In this pattern, you deploy all of your production systems across two Availability Zones in the primary Region. The compute deployed for the production SAP database and central services tiers are the same size in both Availability Zones with automated fail over in the event of Availability Zone failure. The compute required for the SAP application tier is split 50/50 between two Availability Zones. Your non-production systems are **not** an equivalent size to your production and are deployed in a different Availability Zone within the Region. Additionally, compute capacity is deployed in Availability Zone 1 in secondary Region for production SAP database and central services tiers. The production database is replicated to the secondary Region using a database replication capability or a block level replication solution.

The Production database backups stored in Amazon S3, Amazon EBS Snapshots, and Amazon Machine Images are replicated to the secondary Region. In the event of a complete Region failure, the production systems would be restored in the secondary Region using the replicated data for the database tier and the last set of backups for the SAP central services and application tiers.

Select this pattern if:

- You require a defined time window to complete recovery of production and assurance of the availability of compute capacity in another Availability Zone within the primary Region for the production SAP database and central services tiers.
- You can accept the additional cost of deploying the required compute and storage for production SAP database and central services tiers across two Availability Zones within the primary Region.
- You can accept the increased cost of deploying the required compute and storage for production SAP database and central services tiers across two Availability Zones in the primary Region.
- You can accept the cross-Availability Zone related data transfer costs for data replication.
- You can accept that automated fail over between Availability Zones requires a third-party cluster solution.
- You can allow for a period of time where there is only one set of computes deployed for the SAP database and central services in the event of an Availability Zone failure or significant Amazon EC2 failure.
- You can accept that data replication across Availability Zones of the database-related data requires either a database replication capability or a block level replication solution.
- You can accept the variable time duration required (including any delay in availability of the required compute capacity in the remaining Availability Zones) to return the application tier to 100% capacity.
- You require a defined time window to complete recovery of production in the event of a Region failure.
- You can accept the increased complexity and costs associated with multi-Region approach.
- You require assurance of availability of compute capacity in a single Availability Zone in the secondary Region for the production SAP database and central services tiers.
- You can accept the increased cost of deploying the required compute and storage for production SAP database and central services tiers across two Availability Zones in one Availability Zone in the secondary Region.
- You can accept that manual actions are required to fail over between Regions.

Key design principles

- 100% compute capacity is deployed in Availability Zone 1 and Availability Zone 2 for production SAP database and central services tiers.
- 100% compute capacity is deployed in Availability Zone 1 in the secondary Region for production SAP database and central services tiers.
- Compute capacity is deployed in Availability Zone 1 and Availability Zone 2 for production SAP application tier (Active/Active) and needs to be scaled in the event of an Availability Zone failure or significant Amazon EC2 service degradation.
- Amazon EC2 auto recovery is configured for all instances to protect against underlying hardware failure with the exception of those instances protected by a third-party cluster solution.
- The database-related data on Amazon EBS is replicated between Availability Zones using either a database replication capability or a block level replication solution.
- The SAP database-related data on Amazon EBS is replicated between Regions using either a database replication capability or a block level replication solution.
- Amazon EFS is used for the SAP Global File Systems and replicated to the secondary Region.
- SAP database data is backed up regularly to Amazon S3.
- Amazon Machine Image/Amazon EBS Snapshots are taken for all servers on a regular basis
- Amazon S3 data (database backups), Amazon EBS Snapshots, and Amazon Machine Images are replicated/copied to a secondary Region to protect logical data loss.

Benefits

- Low Mean Time to Recovery (MTTR) in the event of an Amazon EC2, Availability Zone or Region failure
- Predictable Return to Service (RTS)
- Database-related data persisted on different sets of Amazon EBS volumes in two Availability Zones in primary Region and one set of volumes in an Availability Zone in secondary Region via database replication capability or a block level replication solution
- Required compute capacity deployed in two Availability Zones in primary Region and one Availability Zone in secondary Region
- No dependency on restoring data from Amazon S3 in the event of an Availability Zone failure or Region failure

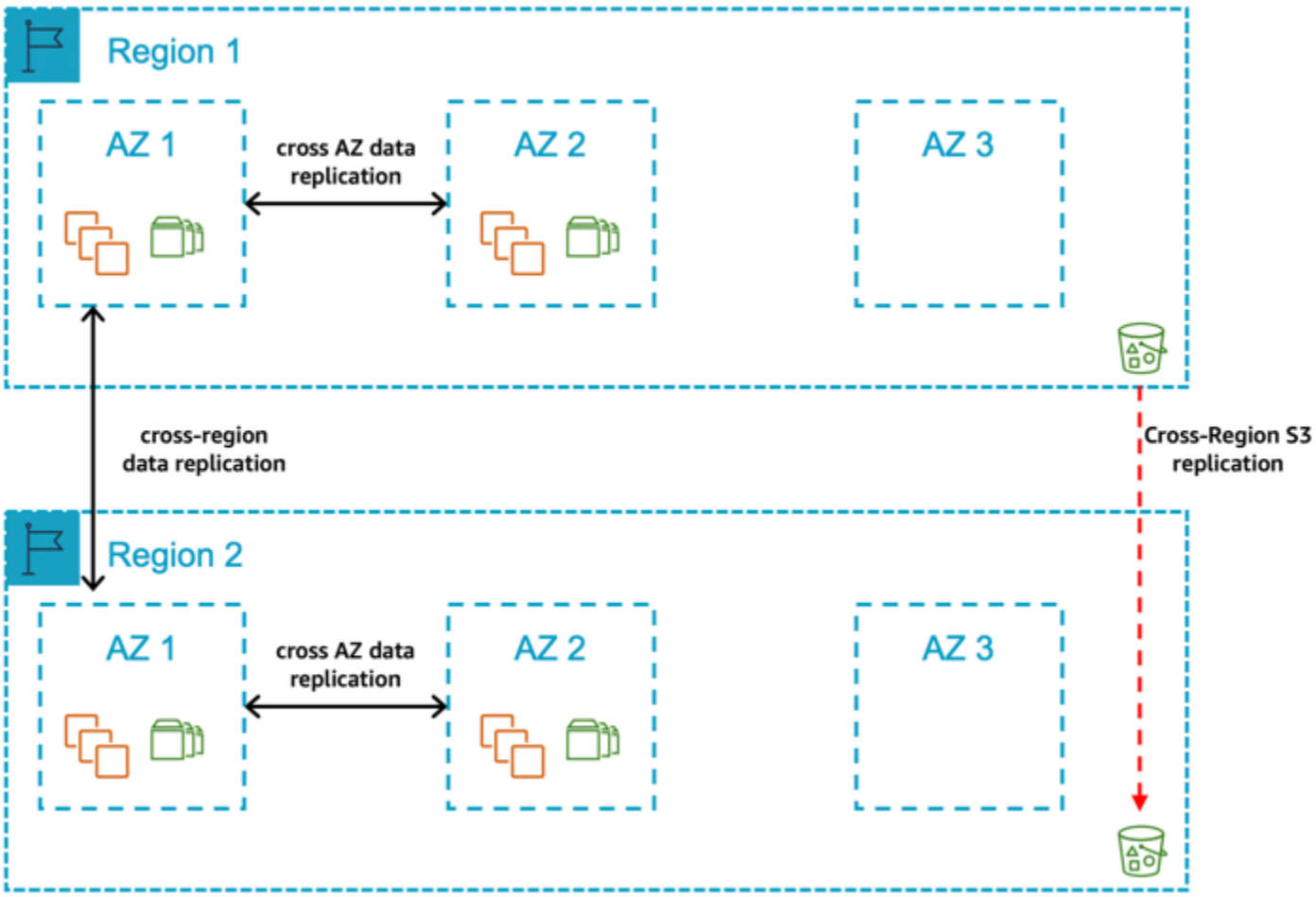
- Ability to protect against significant degradation or total Availability Zone failure through fail over to Availability Zone 2 of database and central services tiers
- Ability to protect against significant degradation or total Region failure through fail over to secondary Region

Considerations

- Well documented and tested processes are required for the automated fail over between Availability Zones.
- Well documented and tested processes are required for maintaining the automated fail over solution.
- Well documented and tested processes are required for scaling the AWS resources to return the application tier to full capacity in the event of an Availability Zone failure or significant Amazon EC2 service degradation.
- Well documented and tested processes are required for moving production to the secondary Region.
- Higher network latency from your on-premises locations to the secondary AWS Region may impact end user performance.
- There is an overhead of maintaining the same software version and patch levels (OS, Database, SAP) across two different Regions.

Pattern 7: A primary Region with two AZs for production and a secondary Region with compute and storage capacity deployed and data replication across two AZs

Figure 13: A primary Region with two Availability Zones for production and a secondary Region with compute and storage capacity deployed and data replication across two Availability Zones



In this pattern, you deploy all of your production systems across two Availability Zones in the primary Region. The compute deployed for the production SAP database and central services tiers are the same size in both Availability Zones with automated fail over in the event of Availability Zone failure. The compute required for the SAP application tier is split 50/50 between two Availability Zone. Additionally, you have compute capacity deployed in Availability Zone 1 and Availability Zone 2 in secondary Region for production SAP database and central services tiers and the production database is replicated to the secondary Region using either a database replication capability or a block level replication solution. The production database backups stored in Amazon S3, Amazon EBS Snapshots, and Amazon Machine Images are replicated on a secondary Region. In the event of a complete Region failure, the production systems would be moved over to the secondary Region manually.

Select this pattern if:

- You require a defined time window to complete recovery of production and assurance of the availability of compute capacity in another Availability Zone within the primary Region for the production SAP database and central services tiers.
- You can accept the additional cost of deploying the required compute and storage for production SAP database and central services tiers across two Availability Zones within the primary Region.
- You can allow for a period of time where there is only one set of computes deployed for the SAP database and central services in the event of an Availability Zone failure or significant Amazon EC2 failure.
- You can accept that data replication across Availability Zones of the database-related data requires either a database replication capability or a block level replication solution.
- You can accept the cross-Availability Zone related data transfer costs for data replication.
- You can accept that automated fail over between Availability Zones requires a third-party cluster solution.
- You can accept the variable time duration required (including any delay in availability of the required compute capacity in the remaining Availability Zones) to return the application tier to 100% capacity.
- You require a defined time window to complete recovery of production in the event of a Region failure.
- You require assurance of availability of compute capacity in two Availability Zones in the secondary Region for the production SAP database and central services tiers.
- You can accept the additional cost of deploying the required compute and storage for production SAP database and central services tiers across two Availability Zones in the secondary Region.
- You can accept the increased complexity and costs associated with multi-Region approach.
- You can accept that manual actions are required to fail over between Regions.

Key design principles

- 100% compute capacity is deployed in Availability Zone 1 and Availability Zone 2 in the primary Region for production SAP database and central services tiers.
- 100% compute capacity is deployed in Availability Zone 1 and Availability Zone 2 in the secondary Region for production SAP database and central services tiers.
- Compute capacity is deployed in Availability Zone 1 and Availability Zone 2 in the primary Region for production SAP application tier (Active/Active) and needs to be scaled in the event of an Availability Zone failure or significant Amazon EC2 service degradation.

- Amazon EC2 auto recovery is configured for all instances to protect against underlying hardware failure with the exception of instances protected by a third-party cluster solution.
- The SAP database-related data on Amazon EBS is replicated between Availability Zones using either a database replication capability or a block level replication solution.
- The SAP database-related data on Amazon EBS is replicated between Regions using either a database replication capability or a block level replication solution.
- Amazon EFS is used for the SAP Global File Systems and replicated on the secondary Region.
- SAP database data is backed up regularly on Amazon S3.
- Amazon Machine Image/Amazon EBS Snapshots for all servers are taken on a regular basis.
- Amazon S3 data (database backups), Amazon EBS Snapshots, and Amazon Machine Images are replicated/copied to a secondary Region to protect logical data loss.

Benefits

- Low Mean Time to Recovery (MTTR) in the event of Amazon EC2, Availability Zone or Region failure
- Predictable Return to Service (RTS)
- Database-related data persisted on different sets of Amazon EBS volumes in two Availability Zones in the primary Region and different sets of Amazon EBS volumes in two Availability Zones in the secondary Region via database replication capability or a block level replication solution
- Required compute capacity deployed in two Availability Zones in primary Region and two Availability Zones in secondary Region
- No dependency on restoring data from Amazon S3 in the event of an Availability Zone or Region failure
- Ability to protect against significant degradation or total Availability Zone failure through fail over to Availability Zone 2 of database and central services tiers
- Ability to protect against significant degradation or total Region failure through fail over to secondary Region

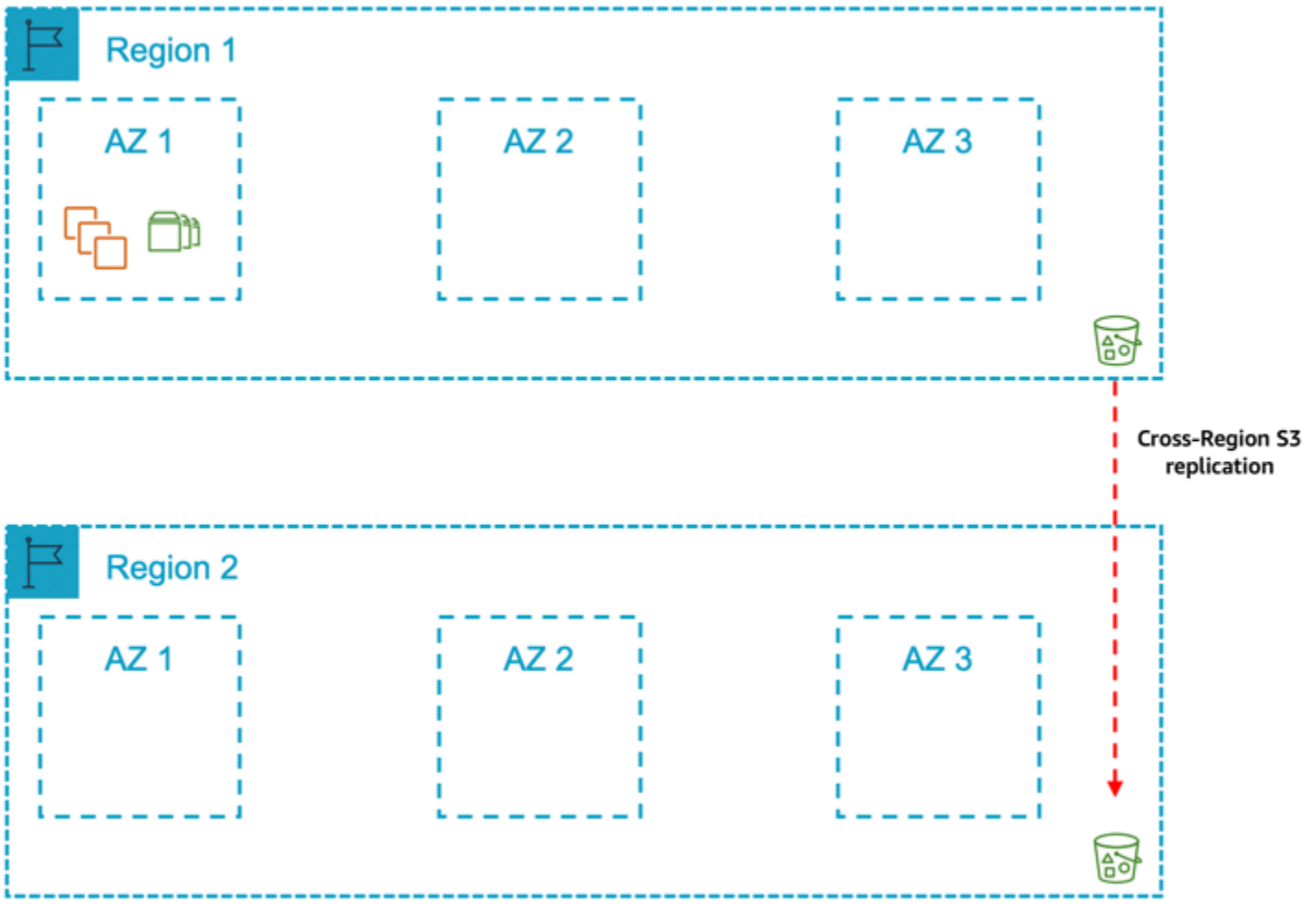
Considerations

- Well documented and tested processes are required for the automated fail over between Availability Zones.

- Well documented and tested processes are required for maintaining the automated fail over solution.
- Well documented and tested processes are required for scaling the AWS resources to return the application tier to full capacity in the event of an Availability Zone failure or significant Amazon EC2 service degradation.
- Well documented and tested processes are required for moving production to the secondary Region.
- Higher network latency from your on-premises locations to the secondary AWS Region may impact end user performance.
- There is an overhead of maintaining the same software version and patch levels (OS, Database, SAP) across two different Regions.

Pattern 8: A primary Region with one AZ for production and a secondary Region containing a replica of backups/AMIs

Figure 14: A primary Region with one Availability Zone for production and a secondary Region containing a replica of backups/AMIs



In this pattern, you deploy your production systems in the primary Region in one Availability Zone. Your non-production systems are **not** an equivalent size to your production and are deployed in the same Availability Zones or a different Availability Zone within the Region.

Additionally, the production database backups stored in Amazon S3, Amazon EBS Snapshots, and Amazon Machine Images are replicated to a secondary Region. In the event of a complete Region failure, the production systems would be restored from the last set of backups in the second Region.

Select this pattern if:

- In the event of an Availability Zone failure or significant Amazon EC2 service degradation, you can accept the risks related to the variable time duration required (including any delay in availability of the required compute capacity in the remaining Availability Zones) to re-create the AWS resources in a different Availability Zone and restore the persistent data to Amazon EBS.

- You can accept the risks related to variable time duration required to complete recovery of production in the event of a Region failure.
- You want to avoid the cost implications with a Multi-AZ approach and accept the related risks of downtime of your production SAP systems.
- You can accept the increased complexity and costs associated with multi-Region approach.
- You can accept that manual actions are required to restore production in the secondary Region.

Key design principles

- 100% compute capacity is deployed in Availability Zone 1 for production SAP database and central services tiers.
- 100% compute capacity is deployed in Availability Zone 1 for production SAP application tier.
- Amazon EC2 auto recovery is configured for all instances to protect against underlying hardware failure.
- Deployed non-production compute capacity is less than 100% of the compute capacity deployed for production SAP database and central services tiers.
- The SAP database is persisted on Amazon EBS in a single Availability Zone only and not replicated to another Availability Zone using either a database replication capability or a block level replication solution.
- Amazon EFS is used for the SAP global file systems.
- SAP database is backed up regularly to Amazon S3.
- Amazon S3 is configured to protect against logical data loss.
- Amazon Machine Image/Amazon EBS Snapshots are taken for all servers on a regular basis.
- Amazon S3 data (database backups), Amazon EBS Snapshots, and Amazon Machine Images are replicated/copied to a secondary Region to protect logical data loss.

Benefits

- Reduced cost compared to Multi-AZ
- Ability to protect against significant degradation or total Region failure through fail over to secondary Region

Considerations

- Well documented and tested processes are required for scaling the AWS resources to return the SAP application tier to full capacity in the event of an Availability Zone failure or significant Amazon EC2 service degradation.
- Well documented and tested processes are required for scaling the AWS resources, restoring the data, and moving production to the secondary Region.
- Higher network latency from your on-premises locations to the secondary AWS Region may impact end user performance.
- In the event of compute, Availability Zone or Region failure due to lack of high availability across two Availability Zones, there is an increased time required to recover production.

Summary

The table below summarizes the patterns and their key characteristics.

#	Single Region	Multi Region	Single AZ Primary	Multi AZ Primary	Single AZ Second Region	Multi AZ Second Region	Prod Capacity in 2nd AZ	Use of non-prod capacity	Cross-Region data replication
1	Yes	No	No	Yes	No	No	Yes	No	No
2	Yes	No	No	Yes	No	No	Yes	Yes	No
3	Yes	No	Yes	No	No	No	No	Yes	No
4	Yes	No	Yes	No	No	No	No	No	No
5	No	Yes	No	Yes	Yes	No	Yes	No	Yes
6	No	Yes	No	Yes	Yes	No	Yes	No	Yes
7	No	Yes	No	Yes	No	Yes	Yes	No	Yes
8	No	Yes	Yes	No	Yes	No	No	No	Yes

Table 1: Summary of patterns

With the flexibility and agility of the AWS Cloud, you have the ability to select any of the patterns described in this guide. You can select the pattern that best meets the business requirements for your SAP systems. It saves you from the trouble of selecting the highest requirement and applying it to all production systems.

For example, if you require highly-available compute capacity in another Availability Zone for the production SAP database and central services tiers of your core ERP system and for your BW system, you can accept the variable time duration required to re-create the AWS resources in a different Availability Zone and restore the persistent data. In this case, you would select Pattern 3 for ERP and Pattern 1 for BW to reduce the overall TCO.

If your requirements change over time, it is possible to move to a different pattern without significant re-design. For example, during the earlier phases of an implementation project, you may not require highly-available compute capacity in another Availability Zone but you can deploy the capacity into a second Availability Zone a few weeks before go-live.

You should consider the following when selecting an architecture pattern to run your SAP system in AWS:

- The geographical residency of the data
- The impact of your production SAP systems' downtime on your organization
- The recovery time objective
- The recovery point objective
- The cost profile

SAP on AWS architecture patterns for Microsoft SQL server

This document provides information about architecture patterns for deploying SAP workloads in AWS Cloud on Microsoft SQL servers. These patterns offer highly available and resilient implementation options while considering your recovery time and point objectives.

Work backwards from your business requirements to define an approach that meets the availability goals of your SAP systems and data. For each failure scenario, the resiliency requirements, acceptable data loss, and mean time to recover need to be proportionate to the criticality of the component and the supported business applications.

You can customize these patterns for your specific business criteria. You should consider the risk and impact of each failure type, and the cost of mitigation when choosing a pattern.

Topics

Patterns

The architecture patterns are divided into two categories.

- Single Region patterns
- Multi-Region patterns

Comparison matrix

The following table provides a comparison of all the architecture patterns discussed further.

Patterns	Business requirements			Solution characteristics		Implementation details	
	Resilience type	Recovery point objective	Recovery time objective	Cost	Complexity	SQL AlwaysOn	Amazon S3 replication
Pattern 1	Single Region disaster recovery	Near zero*	Low	Medium	Medium	2-tier	N/A
Pattern 2		Medium	High	Very low	Very low	N/A	N/A
Pattern 3	Multi-Region disaster recovery	Medium	High	Medium	Medium	2-tier	Cross Region

Pattern 4	Near zero*	Low	High	High	3-tier	Cross Region
Pattern 5	Medium	High	Low	Low	N/A	Cross Region
Pattern 6	Low	Low	Medium	Medium	N/A	N/A

**To achieve near zero recovery point objective, database replication must be setup in synchronous data commit mode within the same AWS Region.*

Single Region architecture patterns for Microsoft SQL server

Single Region architecture patterns help you avoid network latency as your SAP workload components are located in a close proximity within the same Region. Every AWS Region generally has three Availability Zones. For more information, see AWS Global Infrastructure Map.

You can choose these patterns when you need to ensure that your SAP data resides within regional boundaries stipulated by the data sovereignty laws.

The following are the two single Region architecture patterns.

Patterns

- Pattern 1: Single Region with two Availability Zones for production
- Pattern 2: Single Region with one Availability Zone for production

Pattern 1: Single Region with two Availability Zones for production

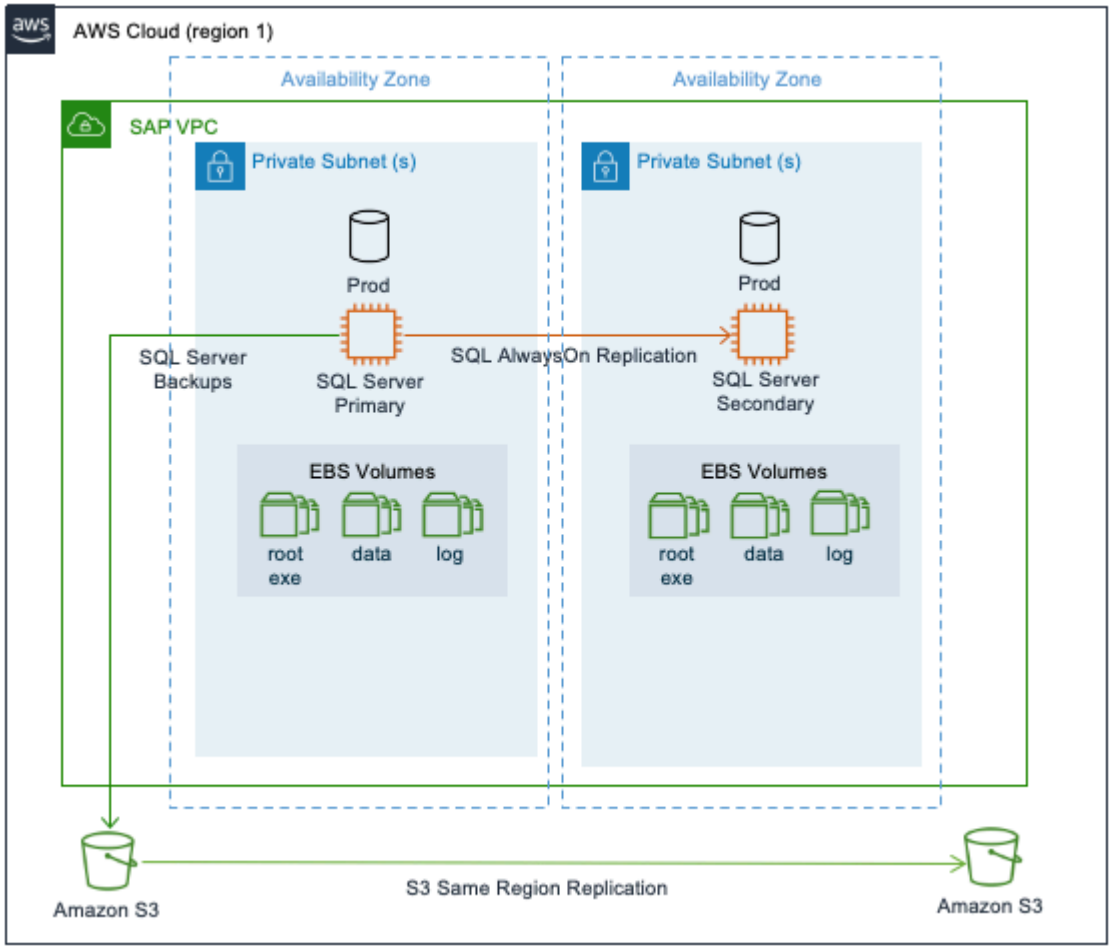
In this pattern, your Microsoft SQL server is deployed across two Availability Zones with AlwaysOn configured on both the instances. The primary and secondary instances are of the same instance type. The secondary instance can be deployed in active/passive or active/active mode. We recommend using the sync mode of replication for the low-latency connectivity between the two Availability Zones.

This pattern is foundational if you are looking for high availability cluster solutions for automated failover to fulfill near-zero recovery point and time objectives. SQL AlwaysOn with Windows cluster for automatic failover provides resiliency against failure scenarios, including the rare occurrence of loss of Availability Zone.

You need to consider the additional cost of licensing for AlwaysOn configuration. Also, provisioning production equivalent instance type as standby adds to the total cost of ownership.

Microsoft SQL server backups can be stored in Amazon S3 buckets. Amazon S3 objects are automatically stored across multiple devices, spanning a minimum of three Availability Zones across a Region. To protect against logical data loss, you can use the Same-Region Replication feature of Amazon S3.

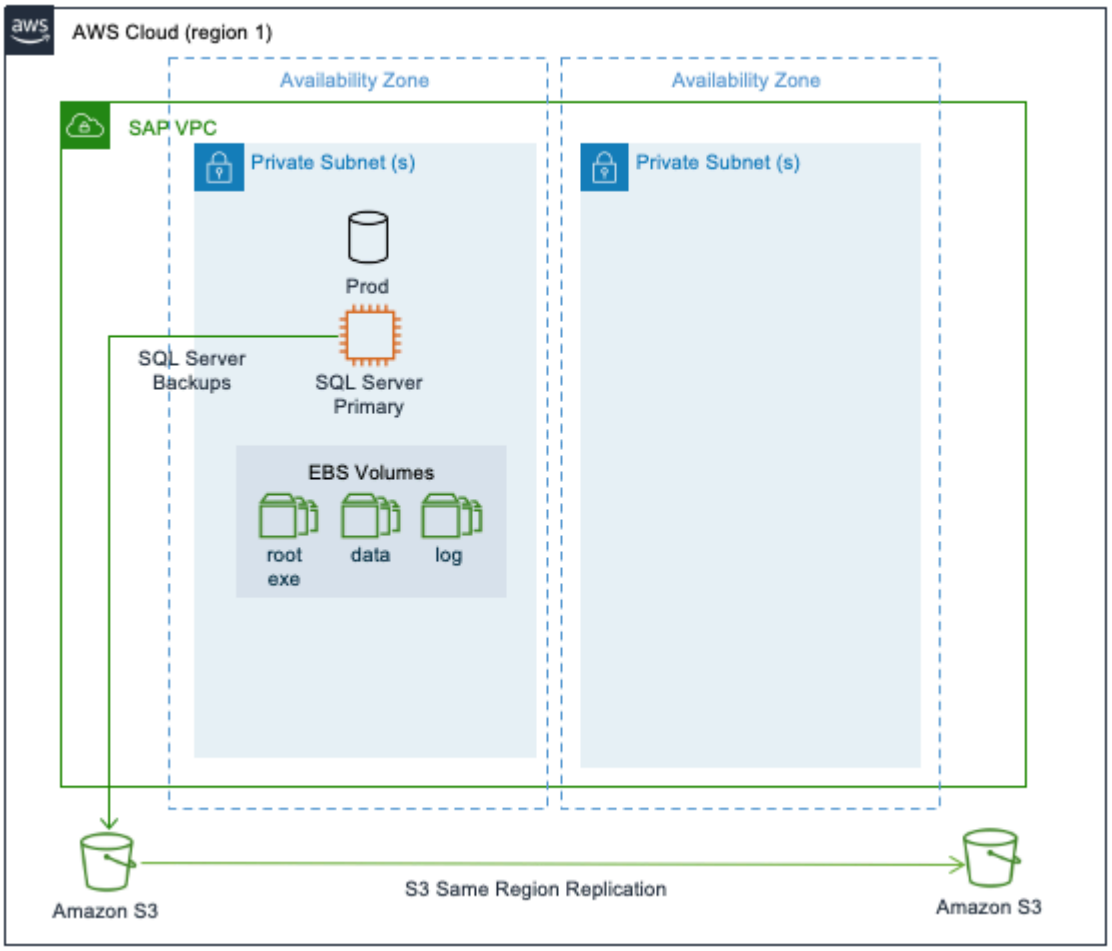
With Same-Region replication, you can setup automatic replication of an Amazon S3 bucket in a separate AWS account. This strategy ensures that not all copies of data are lost due to malicious activity or human error. To setup Same-Region replication, see Setting up replication.



Pattern 2: Single Region with one Availability Zone for production

In this pattern, Microsoft SQL server is deployed as a standalone installation with no target systems to replicate data. This is the most basic and cost-efficient deployment option. The options available to restore business operations during a failure scenario are by Amazon EC2 auto recovery,

in the event of an instance failure or by restoration and recovery from most recent and valid backups, in the event of a significant issue impacting the Availability Zone.



Multi-Region patterns for Microsoft SQL server

AWS Global Infrastructure spans across multiple Regions around the world and this footprint is constantly increasing. For the latest updates, see AWS Global Infrastructure. If you are looking for your SAP data to reside in multiple regions at any given point to ensure increased availability and minimal downtime in the event of failure, you should opt for multi-Region architecture patterns.

When deploying a multi-Region pattern, you can benefit from using an automated approach such as, cluster solution, for fail over between Availability Zones to minimize the overall downtime and remove the need for human intervention. Multi-Region patterns not only provide high availability but also disaster recovery, thereby lowering overall costs. Distance between the chosen regions have direct impact on latency and hence, in a multi-Region pattern, this has to be considered into the overall design.

There are additional cost implications from cross-Region replication or data transfer that also need to be factored into the overall solution pricing. The pricing varies between Regions.

The following are the four multi-Region architecture patterns.

Patterns

- Pattern 3: Primary Region with two Availability Zones for production and secondary Region with a replica of backups/AMIs
- Pattern 4: Primary Region with two Availability Zones for production and secondary Region with compute and storage capacity deployed in a single Availability Zone
- Pattern 5: Primary Region with one Availability Zone for production and a secondary Region with a replica of backups/AMIs
- Pattern 6: Primary Region with one Availability Zone for production and a secondary Region replicated at block level using AWS Elastic Disaster Recovery

Pattern 3: Primary Region with two Availability Zones for production and secondary Region with a replica of backups/AMIs

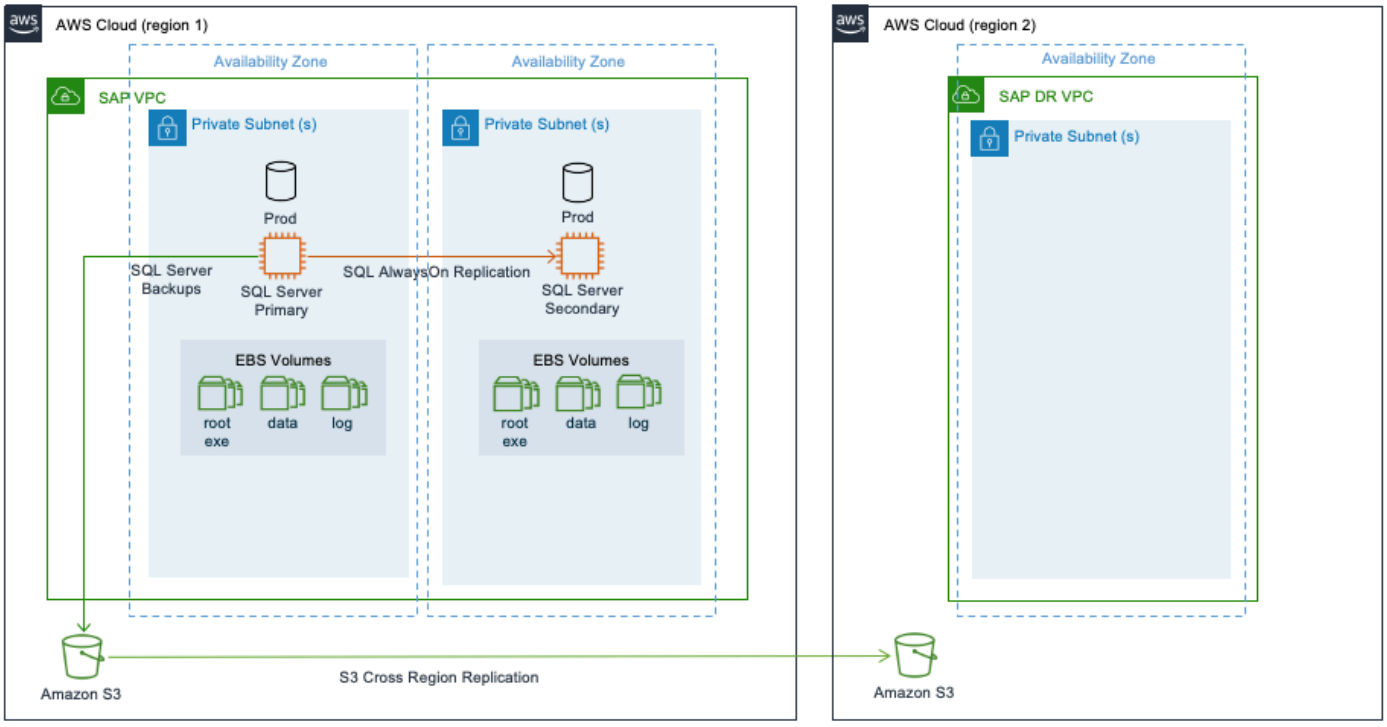
This pattern is similar to pattern 1 where your Microsoft SQL server is highly available. You deploy your production instance across two Availability Zones in the primary Region using AlwaysOn. You can restore your SQL database in a secondary Region with a replica of backups stores in Amazon S3, Amazon EBS, and Amazon Machine Images (AMIs).

With cross-Region replication of files stored in Amazon S3, the data stored in a bucket is automatically (asynchronously) copied to the target Region. Amazon EBS snapshots can be copied between Regions. For more information, see Copy an Amazon EBS snapshot. You can copy an AMI within or across Regions using AWS CLI, AWS Management Console, AWS SDKs or Amazon EC2 APIs. For more information, see Copy an AMI. You can also use AWS Backup to schedule and run snapshots and replications across Regions.

In the event of a complete Region failure, the production SQL server needs to be built in the secondary Region using AMI. You can use AWS CloudFormation templates to automate the launch of a new SQL server. Once your instance is launched, you can then download the last set of backup from Amazon S3 to restore your SQL server to a point-in-time before the disaster event. After restoring and recovering your SQL server in the secondary Region, you can redirect your client traffic to the new instance using DNS.

This architecture provides you with the advantage of implementing your SQL server across multiple Availability Zones with the ability to failover instantly in the event of a failure. For disaster recovery that is outside the primary Region, recovery point objective is constrained by how often you store your SQL backup files in your Amazon S3 bucket, and the time it takes to replicate your Amazon S3 bucket to the target Region. You can use Amazon S3 replication time control for a time-bound replication. For more information, see Enabling Amazon S3 Replication Time Control.

Your recovery time objective depends on the time it takes to build the system in the secondary Region and restore operations from backup files. The amount of time will vary depending on the size of the database. Also, the time required to get the compute capacity for restore procedures may be more in the absence of a reserved instance capacity. This pattern is suitable when you need the lowest possible recovery time and point objectives within a Region and high recovery point and time objectives for disaster recovery outside the primary Region.



Pattern 4: Primary Region with two Availability Zones for production and secondary Region with compute and storage capacity deployed in a single Availability Zone

In addition to the architecture of pattern 3, this pattern has SQL AlwaysOn setup between the SQL server in the primary Region and an identical third instance in one of the Availability Zones in the

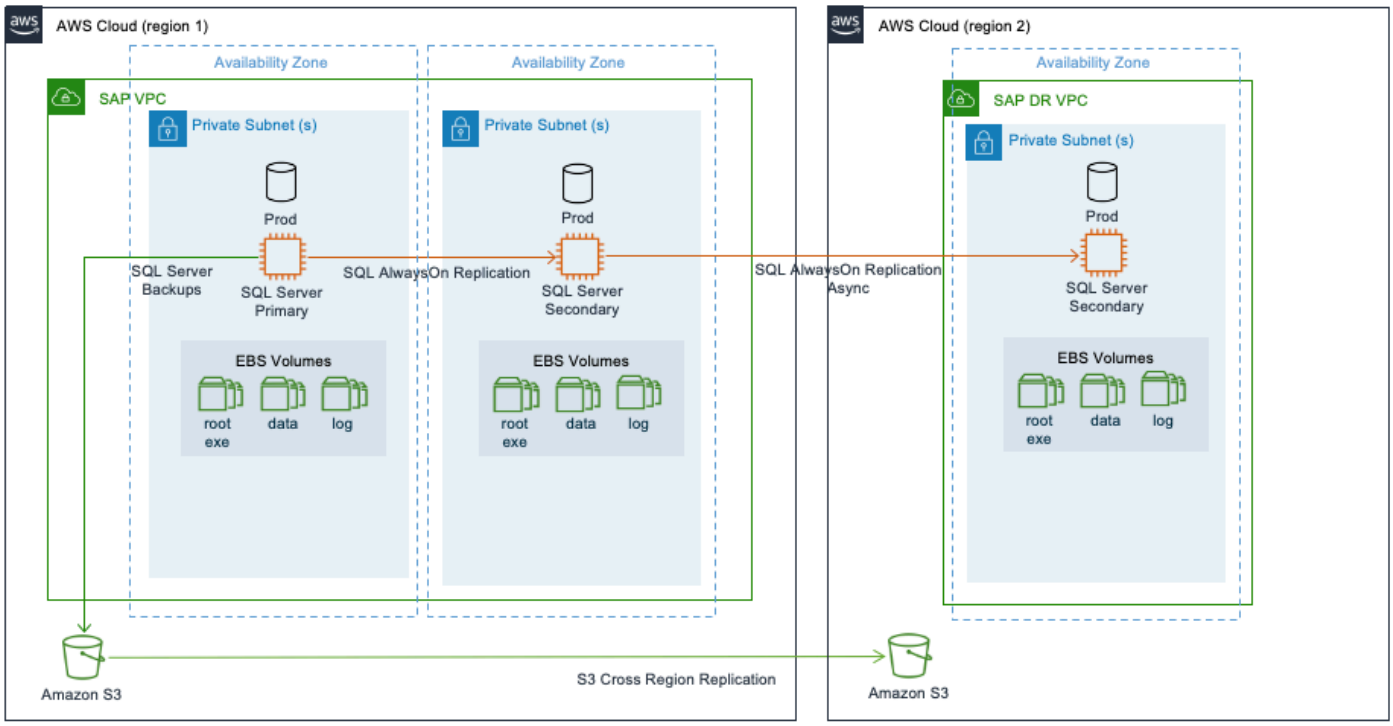
secondary Region. We recommend using the asynchronous (async) mode for SQL AlwaysOn when replicating between AWS Regions due to increased latency.

In the event of a failure in the primary Region, the production workloads are failed over to the secondary Region manually. This pattern ensures that your SAP systems are highly available and are disaster-tolerant. This pattern provides a quicker failover and continuity of business operations with continuous data replication.

There is an increased cost of deploying the required compute and storage for the production SQL server in the secondary Region and of data transfers between Regions. This pattern is suitable when you require disaster recovery outside of the primary Region with low recovery point and time objectives.

This pattern can be deployed in a multi-tier as well as multi-target replication configuration.

The following diagram shows a multi-tier replication where the replication is configured in a chained fashion.



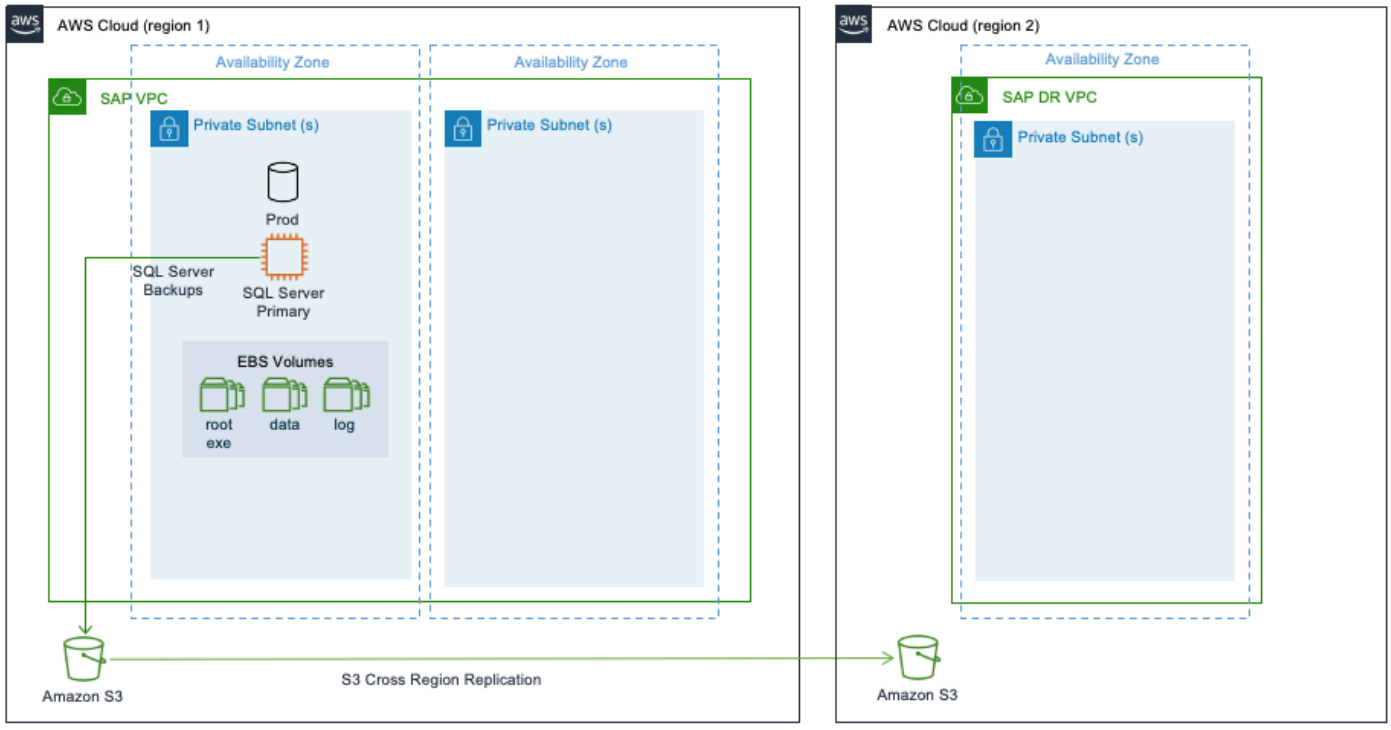
Pattern 5: Primary Region with one Availability Zone for production and a secondary Region with a replica of backups/AMIs

This pattern is similar to pattern 2 with additional disaster recovery in a secondary Region containing replicas of the SQL server backups stored in Amazon S3, Amazon EBS snapshots, and

AMIs. In this pattern, the SQL server is deployed as a standalone installation in the primary Region in one Availability Zone with no target SQL systems to replicate data.

With this pattern, your SQL server is not highly available. In the event of a complete Region failure, the production SQL server needs to be built in the secondary Region using AMI. You can use AWS CloudFormation templates to automate the launch of a new SQL server. Once your instance is launched, you can then download the last set of backup from Amazon S3 to restore your SQL server to a point-in-time before the disaster event. You can then redirect your client traffic to the new instance in the secondary Region using DNS.

For disaster recovery that is outside the primary Region, recovery point objective is constrained by how often you store your SQL backup files in your Amazon S3 bucket and the time it takes to replicate your Amazon S3 bucket to the target Region. Your recovery time objective depends on the time it takes to build the system in the secondary Region and restore operations from backup files. The amount of time will vary depending on the size of the database. This pattern is suitable for non-production or non-critical production systems that can tolerate a downtime required to restore normal operations.



Pattern 6: Primary Region with one Availability Zone for production and a secondary Region replicated at block level using AWS Elastic Disaster Recovery

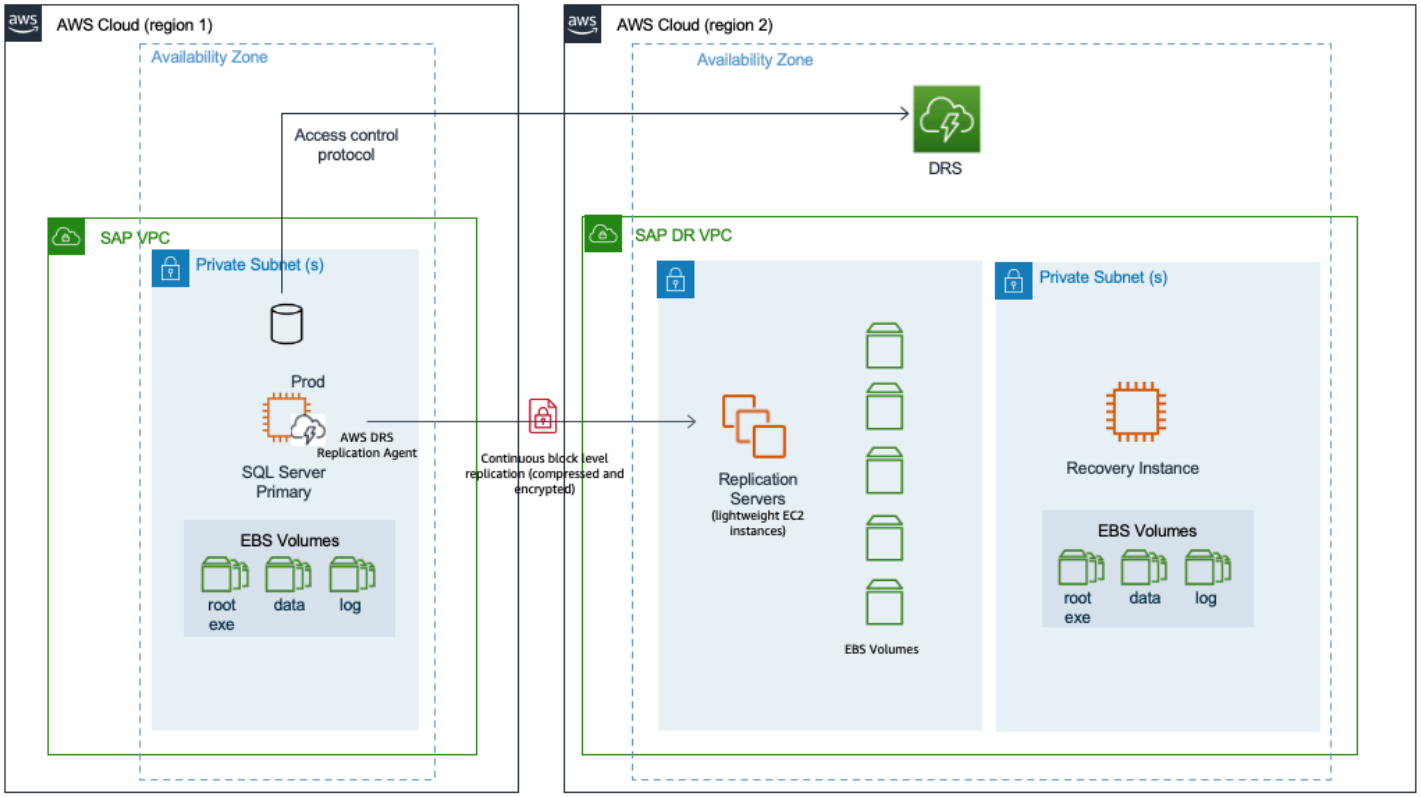
AWS Elastic Disaster Recovery provides organizations with a modern approach to protecting Microsoft SQL server environments by enabling cloud-based disaster recovery on AWS Cloud. For more information, see What is Elastic Disaster Recovery?

Elastic Disaster Recovery uses block level replication and replicates the operating system, databases, application, and system files for supported Windows and Linux operating system versions. To learn more, see Supported operating systems. An initial setup of the AWS Replication Agent is required on the source systems for Elastic Disaster Recovery to initiate secure data replication. The agent runs in memory and recognizes write operations to locally attached disks. These writes are captured and asynchronously replicated into a staging area in your AWS account. During this ongoing replication process, Elastic Disaster Recovery maintains the write order among all disks in the same source server. The replicated Amazon EC2 instances can be run in a *test mode* to perform drills in a segregated environment.

Elastic Disaster Recovery allows you to monitor the data replication status of your recovery instances, view recovery instance details, add recovery instances to Elastic Disaster Recovery, edit recovery instance failback settings, and terminate recovery instances.

With Elastic Disaster Recovery, you can perform a failover by launching recovery instances on AWS Cloud. Once the recovery instance is launched, you must redirect the traffic from your primary site to the recovery site.

AWS Elastic Disaster Recovery uses Amazon EBS snapshots to take point-in-time snapshots of data held within the staging area. To learn more, see Amazon EBS snapshots. It then provides crash consistent point-in-time recovery options that can be used in the event of a disaster or drill. Elastic Disaster Recovery can protect individual nodes of the SQL Server Always On availability group. During disaster recovery, the group is launched as individual SQL server instances on AWS. This solution works for both the SQL Server Standard edition and SQL Server Enterprise edition for any supported version of the SQL server.


AWS Cloud (region 1)
Availability Zone
Access control protocol
SAP VPC
Private Subnet (s)
Prod
AWS DRS Replication Agent
SQL Server Primary
EBS Volumes
root exe
data
log
Continuous block level replication (compressed and encrypted)
AWS Cloud (region 2)
Availability Zone
DRS
SAP DR VPC
Replication Servers
(lightweight EC2 instances)
EBS Volumes
Private Subnet (s)
Recovery Instance
EBS Volumes
root exe
data
log

Disaster recovery for SAP workloads on AWS using AWS Elastic Disaster Recovery

Disasters due to natural events (earthquakes, hurricanes, or floods), application failures, technical failures or human actions cause application downtime and potential data loss, impacting revenue. To mitigate such scenarios, you can create a business continuity plan with the key element of disaster recovery. Designing, implementing, and maintaining a disaster recovery plan is critical for organizations running mission-critical applications, such as SAP. For more information, see Business Continuity Plan (BCP).

AWS Elastic Disaster Recovery enables organizations to quickly and easily implement a new or migrate an existing disaster recovery plan to AWS. The source servers can be hosted on AWS, existing physical or virtual data centers, private cloud or with other cloud providers. We recommend using Elastic Disaster Recovery to implement a disaster recovery plan for your SAP workloads, where AWS is the disaster recovery environment, and the source environment may or may not be on AWS. You can access Elastic Disaster Recovery from the Elastic Disaster Recovery console.

An initial setup of the AWS Replication Agent is required on the source systems for Elastic Disaster Recovery to initiate secure data replication. Your data is replicated using secure protocols, either directly over the internet, or via an encrypted and/or dedicated network connection, to any AWS Region supported by Elastic Disaster Recovery. By replicating the source systems to replication servers in a staging area, the cost of disaster recovery is optimized by using affordable storage, shared servers, and minimal compute resources to maintain ongoing replication.

You can perform non-disruptive tests, known as drills, to confirm that your Elastic Disaster Recovery implementation is ready for a disaster recovery scenario. Elastic Disaster Recovery automatically converts your servers to boot and run natively on AWS when you launch instances for drills or recovery. The service also automatically creates point in time (PIT) snapshots of your server state as it replicates. If you need to recover applications, you can launch recovery instances on AWS within minutes, using the latest snapshot or an earlier PIT snapshot. Once your applications are running on AWS, you can choose to keep them there or initiate data replication back to your primary site when the issue is resolved. You can fail back to your primary site with Elastic Disaster Recovery tools, such as Failback Client.

For more information, see What is Elastic Disaster Recovery?

Topics

Scenarios

The following disaster recovery scenarios are covered in this document.

- in-region – source workload is running on AWS cloud and disaster recovery implementation uses a second Availability Zone in the same AWS Region.
- cross-region – source workload is running on AWS cloud and disaster recovery implementation uses a different AWS Region. The choice of another Region can be for compliance reasons.
- outside of AWS – source workload is running outside of AWS (on-premises, public or private cloud) and disaster recovery is implemented with AWS.

References

This document does not provide detailed steps for setting up and using AWS Elastic Disaster Recovery. For more information, see What is DRSlong;? in the AWS Elastic Disaster Recovery User Guide.

It is important to understand the key business requirements that guide a disaster recovery solution design and implementation, including recovery point objectives, recovery time objectives, along with the disaster recovery plan and disaster recovery drill. Check the following resources for concepts related to a disaster recovery implementation on AWS.

- AWS Elastic Disaster Recovery Core concepts
- AWS Well-Architected Framework : Best Practice 10.1
- Architecture guidance for availability and reliability of SAP on AWS

If you are new to AWS, see the following documents.

- Getting started with AWS
- What is Amazon EC2?
- What is Amazon VPC?
- Amazon Elastic Block Store (Amazon EBS)

To use this information provided here effectively, you must have previous experience installing, migrating, and operating SAP environments and systems on AWS, along with high availability and disaster recovery solution implementation.

Service-level agreements and SAP licenses

For a disaster recovery implementation, Service Level Agreements (SLA) are used to define how resilient your system is at avoiding loss of data and reducing downtime when your workload becomes unavailable due to a disaster event.

An SAP system disaster recovery approach requires replication of the application tier, database tier, and any file shares, such as NFS mounts. The following are some of the factors to consider for your disaster recovery implementation.

Topics

- Recovery time objective (RTO)
- Recovery point objective (RPO)
- Recovery consistency objective (RCO)
- SAP licenses

Recovery time objective (RTO)

Recovery Time Objective (RTO) refers to how quickly can your application recover after an outage. In the event of a disaster, Elastic Disaster Recovery enables you to launch your replicated servers to a fully provisioned state at the target Region within minutes and continue operations. This automated approach supports a low RTO. It can be faster and more effective than a manual approach.

Consideration

As RTO is usually evaluated in the impact to business processes, other factors such as Domain Name System (DNS) propagation, environmental factors, including your disaster recovery team's reaction time, your target environment's storage architecture, operating system boot, and application startup times, influence this target value.

Recovery point objective (RPO)

Elastic Disaster Recovery continuously replicates the changes to the disk at the block level asynchronously, to the target site. The RPO of Elastic Disaster Recovery is typically in the sub-second range. RPO can be influenced by external factors such as, the time taken by the source system to send changes to the staging area. This is further impacted by the volume of transactions on the source system. Other factors include network throughput and latency, source and replication server performance, etc. These factors should be measured to calculate the potential amount of data loss during a disaster recovery event.

Consideration

SAP workloads may from time to time may observe longer amounts of data loss than what is seen with a sub-second RPO due to how Elastic Disaster Recovery manages certain scenarios.

In the event of hard reboots, disk changes, and crashes, Elastic Disaster Recovery triggers a rescan of the disks. During the rescan, the Replication Agent does not replicate the changes of the source server to the target. This creates a lag between the two servers. If the primary system fails during this time, customers may experience a longer amount of data loss (measured in RPO) than expected.

The rescan time depends on multiple factors, and cannot be predicted without testing. A rescan may occur after a reboot of the source server. The rescan time will vary depending on the size of the source disks. The time depends on the performance of the disks (linear read), staging area disk performance, and the rate of write operations on the source server (which are sent in parallel with the rescan). The rescan is functioning normally as long as its moving forward, and is not "stuck".

SAP databases can have large disk sizes and high change rates. We recommend conducting tests to ensure that your SLA requirements are met in such events. Additionally, you must ensure that the primary and target databases are in sync during peak activity cycles.

Recovery consistency objective (RCO)

Many disaster recovery solutions consider only RTO and RPO as SLAs for resiliency. You must also consider Recovery Consistency Objective (RCO) for your SAP workloads. RCO is a measurement

for the consistency of distributed business data within interlinked systems. In a typical customer environment, SAP systems are tightly integrated and data is frequently exchanged between these systems, like SAP ECC or SAP S/4HANA, SAP BW or SAP BW/4HANA, SAP CRM, SAP SRM, SAP GTS etc. This group of tightly integrated systems is called a *system group*. In case of disaster recovery failover, you may have zero RCO requirement within the system group. This means that in case of disaster recovery failover, all of the databases within the SAP system group must be recovered to the same point-in-time.

Consideration

Elastic Disaster Recovery does not guarantee consistency across multiple source instances. If you have zero RCO requirement, you can use database native replication technology with point-in-time recovery or backtrack with secondary time travel.

For more information, see SAP Note 434647 - Point-in-time recovery in an SAP system group (requires SAP portal access).

SAP licenses

The SAP system is secured by a license using a hardware key. On AWS, the hardware key is based on your Amazon EC2 instance ID. Your Amazon EC2 instance must be launched before you can generate your SAP license. When you recover your SAP system in the disaster recovery site, the SAP license becomes invalid as the disaster recovery site is a new Amazon EC2 instance. The hardware key will no longer match. A temporary SAP license is created when the recovery instance is launched, and it is valid for 28 days. You do not need to create a new SAP license. If you need the disaster recovery instance to continue running after 28 days, you can request a new SAP license with the recovery Amazon EC2 instance ID.

Network, storage, and compute

This section provides information about configuring network, storage, and compute for staging and target environments to achieve disaster recovery goals for your SAP workloads on AWS with Elastic Disaster Recovery.

Topics

Network

Your network architecture and configuration used for disaster recovery can play a significant role in supporting an effective RTO and RPO SLA. You must consider network design and redirecting traffic to recovery instance when disaster recovery is triggered.

The following are the four steps to design network for disaster recovery.

- Connecting the source and target network
- Defining the staging and recovery subnets
- Configuring the network security settings
- SAP end user and integration traffic

Connecting the source and target network

The first step is to choose and configure the network connection method from the source network to the replication servers. You can choose between private or public. For more information, see Data routing and throttling.

Regardless of the method, transferred data is always encrypted in transit. The default method is public, where data is routed over the internet to a public network interface on the replication servers. In the private method, the data is replicated over a private network. A private network selection depends on the disaster recovery scenario in use.

- AWS In-Region disaster recovery – Private networks are generally between VPCs, using either Amazon VPC peering or AWS Transit Gateway for connectivity. We recommend using a different AWS account, and separate Amazon VPC for disaster recovery. For more information, see What is Amazon VPC peering? and What is a transit gateway?.
- AWS Cross-Region disaster recovery – We recommend using the fully redundant AWS network backbone that connects different AWS Regions together. Amazon VPC peering and AWS Transit Gateway enable connectivity between Regions. For more details, see Introduction to Network Transformation on AWS.
- Outside of AWS to AWS disaster recovery – In this scenario, your physical network between your source network and AWS are provided through various telecommunications or internet services providers (ISP). The following solutions are available on AWS.
  - AWS Direct Connect
  - AWS Site-to-Site VPN

- SD-WAN available on AWS marketplace

AWS Direct Connect is commonly used by SAP on AWS customers. It provides more predictable performance against service level agreement (SLA) based targets such as throughput, jitter, and latency, versus VPN or SD-WAN based solutions. You can work with AWS Direct Connect Delivery Partners for guidance on which options are the best fit for your environment.

Defining the staging and recovery subnets

One subnet is recommended to host the replication servers, called the *staging area subnet*. Additional subnets, called the recovery subnets, are necessary as the target of your disaster recovery action. For scenarios where the source network is on AWS, consider how your subnets should be allocated based on your selected AWS account strategy and landing zone. Often this may mean that the staging area subnets should be in a different Amazon VPC than your source servers. For a simplified environment, this may just use different subnets in the same Amazon VPC. This would mean reduced isolation between your production and non-production disaster recovery environments. For more information, see AWS Well-Architected Framework : Best Practice 5.3.

Ultimately, the number and design of these subnets should follow similar concepts as your source environment. For more information, see Network diagrams.

For AWS In-Region disaster recovery scenario, we recommend hosting the staging area subnet in a different Availability Zone than the recovery subnets. This design enables an additional redundancy for disaster recovery. The launched recovery instances are protected by a staging area in a separate Availability Zone. This follows the design principle of using multiple Availability Zones to maintain resiliency.

Configuring the network security settings

Ensure that the required network security settings are configured. This includes enabling access through a number of ports in your on-premises firewall, network security devices, security groups, or network access control lists (network ACL), and possibly other tasks depending on the location of your source environment. For more information, see Replication network requirements.

SAP end user and integration traffic

The following are some of the factors that affect how the end user and integration related network traffic can affect your RTO and RPO.

- DNS propagation time for clients to identify and resolve to new IP
- Delays in network components (if any used) to reroute traffic, such as global or local load balancers, including AWS Application Load Balancers, AWS Global Accelerator, or Amazon Route 53 Public Data Plane

For more information, see Disaster recovery options in the cloud.

Storage

AWS Elastic Disaster Recovery is designed to evaluate and define the optimal Amazon EBS volume settings for your staging environment based on the source server performance. A default performance setting is used for drill and recovery servers. These volumes are sized to match the capacity needs of the source systems. You must review these settings with the specific requirements of your SAP workloads. This ensures an efficient and disaster recovery SLA compliant environment. These different server types have different requirements, and methods of managing storage.

Topics

- Replication servers
- Drill and recovery instances
- Point in time recovery

Replication servers

The staging area requires storage to support ongoing replication from source machines. These Amazon EBS volumes are usually low-cost, hard disk drive (HDD) type storage volumes. However, if the replicated disk write throughput is high, the default Replication server settings dynamically change to a higher performance, solid state drive (SSD) storage type. The default Amazon EBS volume type setting – **Auto volume type selection** for replication servers, is the recommended setting for SAP workloads. It automatically chooses the high-performing, cost-efficient Amazon EBS volumes for your workload requirements.

You have the option to increase the performance of the staging area by selecting solid state drives (SSD). This can help SAP workloads, such as bursty or consistently high transaction rate databases which have a high rate of create, update, and/or delete operations that must be applied to its storage. For such workloads, we recommend monitoring Amazon CloudWatch metrics and check

for any persistent or increasing delays. You can use the following CloudWatch metrics for Elastic Disaster Recovery.

- **LagDuration** – the age of the latest consistent snapshot, in seconds
- **Backlog** – the amount of data yet to be synced, in bytes

If Amazon EBS metrics on the replication server also indicate performance issues, you can change Amazon EBS volume type. See the following resources to learn more.

- Amazon EBS volume performance on Linux instances
- Volumes
- Disk settings

Drill and recovery instances

SAP workloads require at least the gp3 volume type for 90% or more of the use cases, including SAP applications and databases (SAP HANA and any other). If you have a higher per-volume IOPS requirement of more than 16,000 IOPS, or per-volume throughput requirement greater than 1,000 MiB/s, consider `io2` or `io2 Block Express` volumes.

When you launch drill or recovery instances, Elastic Disaster Recovery creates Amazon EBS storage volumes based on the types defined in the launch template. For more information, see Amazon EC2 launch template. The launch template is automatically generated by Elastic Disaster Recovery, with default values for storage performance, using general purpose SSD (volumes sized to match the source system capacity requirements). Review the launch template to confirm that your workload's storage requirements are being met by the default allocations of the launch template.

You can modify the launch template for a different volume type or performance setting. Before modifying, confirm that your target Amazon EC2 instance type supports higher storage. For more details, see Supported instance types. For SAP HANA databases, see Storage configuration. Define the modified version as the default launch template for your server once your changes are applied to the template. We do not recommend adding or removing Amazon EBS volumes in the template when using it with Elastic Disaster Recovery.

For servers that require loading larger amounts of data before they become active, such as database servers, you can configure higher performance settings and types of storage in the launch template. For example, if your server is configured with gp3 storage, then defining more provisioned throughput and IOPS for your storage, and/or using a higher performance scaling

storage such as `io2 Block Express` (with a supported Amazon EC2 instance type), can reduce the time it takes for your drill or recovery instance to handle the expected workload quantity. Once your drill or recovery instance is fully online, you can change revert your storage settings. For more information, see Amazon EBS Elastic Volumes. You can increase the volume size, change the volume type, or adjust the performance of your Amazon EBS volumes, without detaching the volume or restarting the instance.

Point in time recovery

AWS Elastic Disaster Recovery uses Amazon EBS snapshots to give Point in Time (PiT) recovery options that can be used during a drill or recovery. Amazon EBS snapshots of the staging are volumes are continuously taken to provide recovery points of latest (sub-second RPO), 10-minute increments for the first hour, in one hour increments for 24 hours. A daily PiT is retained for the amount of days specified in your Point in Time (PiT) policy. You can specify between 1 to 365 days, with 7 days being the default. For more information, see Understanding Point In Time states.

Compute

You must choose an Amazon EC2 instance type for both the replication server and the recovery server.

Topics

- Replication servers
- Drill and recovery instances
- Source server

Replication servers

The replication server is normally smaller than the source system. `t3.small` is the default instance type, and it can replicate up to 15 volumes. You can use a shared replication server between SAP application servers, or other servers with low change rates.

If you have a workload that is bursty or has consistently high transaction rate databases, with a high rate of create, update, and/or delete operations that must be applied to its storage, you may require different configurations for the staging area. If you see lag in the replication for your workload, change the default replication server to a different instance family. For example, General Purpose Amazon EC2 instance family or use a dedicated replication server. This change can impact cost. For more information, see Replication server configuration.

Drill and recovery instances

For recovery instances, configure the Amazon EC2 launch template settings to match AWS target instances with source. See the following resources for a list of SAP certified instances.

- SAP NetWeaver certified instances
- SAP HANA certified instances

The following are some of the compute-related factors impacting the RTO of your disaster recovery solutions.

- Server startup time
- SAP running on Microsoft Windows Server operating system
- Large SAP HANA database that takes more than 10 minutes to start up
- SAP application(s) installed on the server, and their startup times
- Mismatch in the source and target server and storage configurations – configuring a lesser compute power or storage performance at the target side increases the RTO

You must consider application startup times as a factor in recovery. We recommend choosing an Amazon EC2 instance type and storage configuration that provides an effective startup time. This helps your optimize the RTO for your disaster recovery solutions. Also, performing a disaster recovery test or drill enables you to measure the RTO based on your operating system and database.

SAP systems can run on a variety of operating systems, infrastructure platforms, and processor instruction sets. If your source servers is on-premises or with another cloud provider, it must be compatible with Amazon EC2 and Elastic Disaster Recovery. The source server must have a 64-bit based operating system built for the x86 system architecture. Various x86 based CPUs are available on AWS, being used on source servers, especially if the servers are old models. Using an SAP sizing-based approach to map the source system to an Amazon EC2 instance type is recommended. To learn more, see SAP's Sizing information.

Source server

While the system requirements for the Replication Agent are relatively low, consider the constraints on the source server for CPU, memory, network, storage, and other resources that can impact the

performance of your disaster recovery solution. Size the source server based on these factors. For more information, see Source server requirements.

Disaster recovery scenarios

The following are the three disaster recovery scenarios.

Topics

AWS In-Region disaster recovery

In AWS cloud, Availability Zones are separated by a meaningful distance, within 100 km (60 miles away from each other). This distance provides isolation from the most common disasters that could affect data centers, for instance, floods, fire, severe storms, earthquakes, etc. It is used by many AWS customers today to support their resiliency requirements for SAP workloads. Based on your business continuity requirements, in-Region disaster recovery maybe suitable for you. For more information, see Single Region architecture patterns.

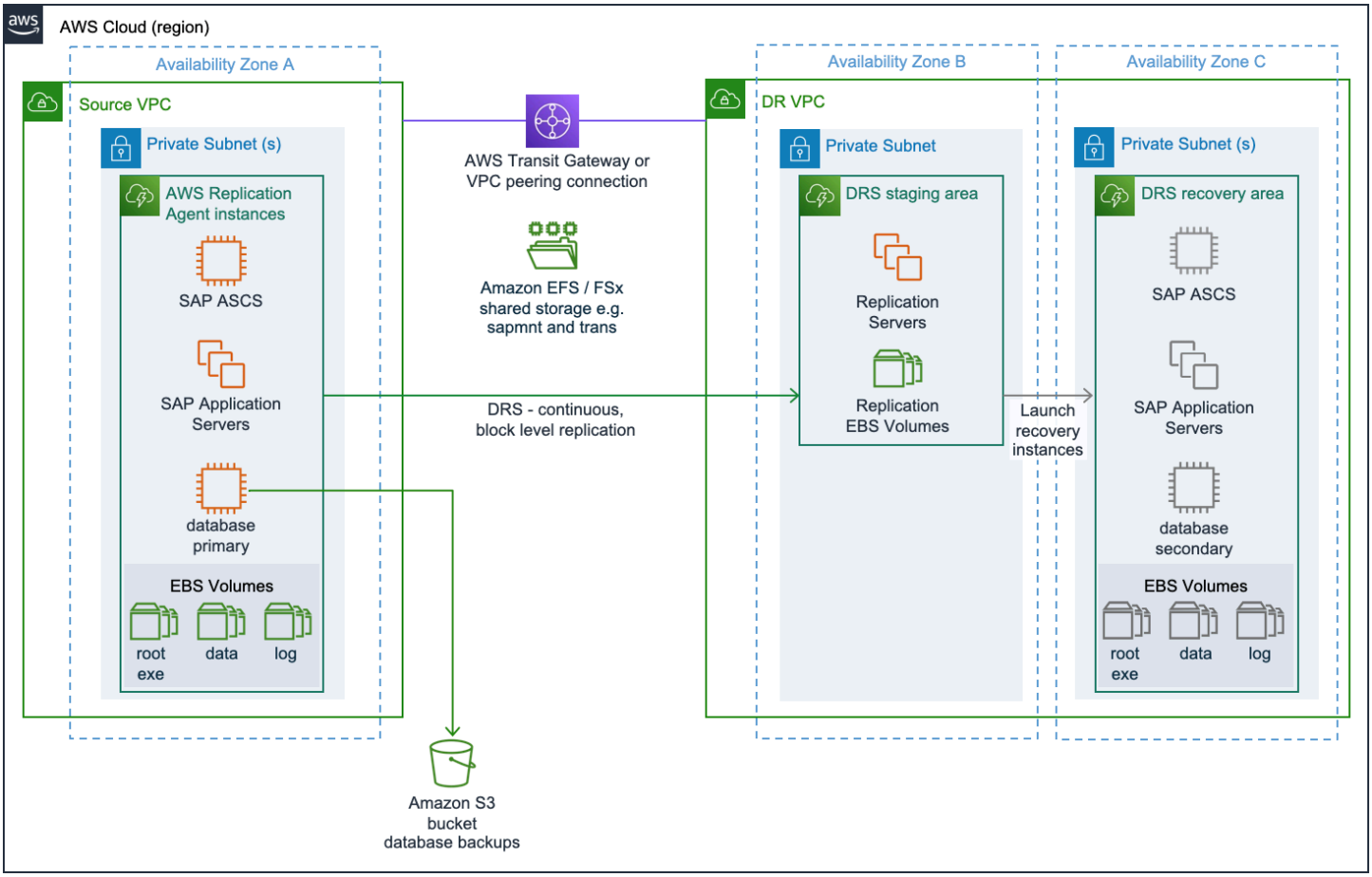
Best practices

- Separate Amazon VPCs for the source and recovery areas in the same Region
- Separate AWS accounts for source and recovery areas
- AWS Transit Gateway or Amazon VPC peering for supporting replication traffic and end user connectivity

  For more information, see Network.
- Multiple Availability Zones resiliency of Amazon S3 buckets and Amazon EFS for data protection
- Separate Availability Zones for source, staging, and recovery areas

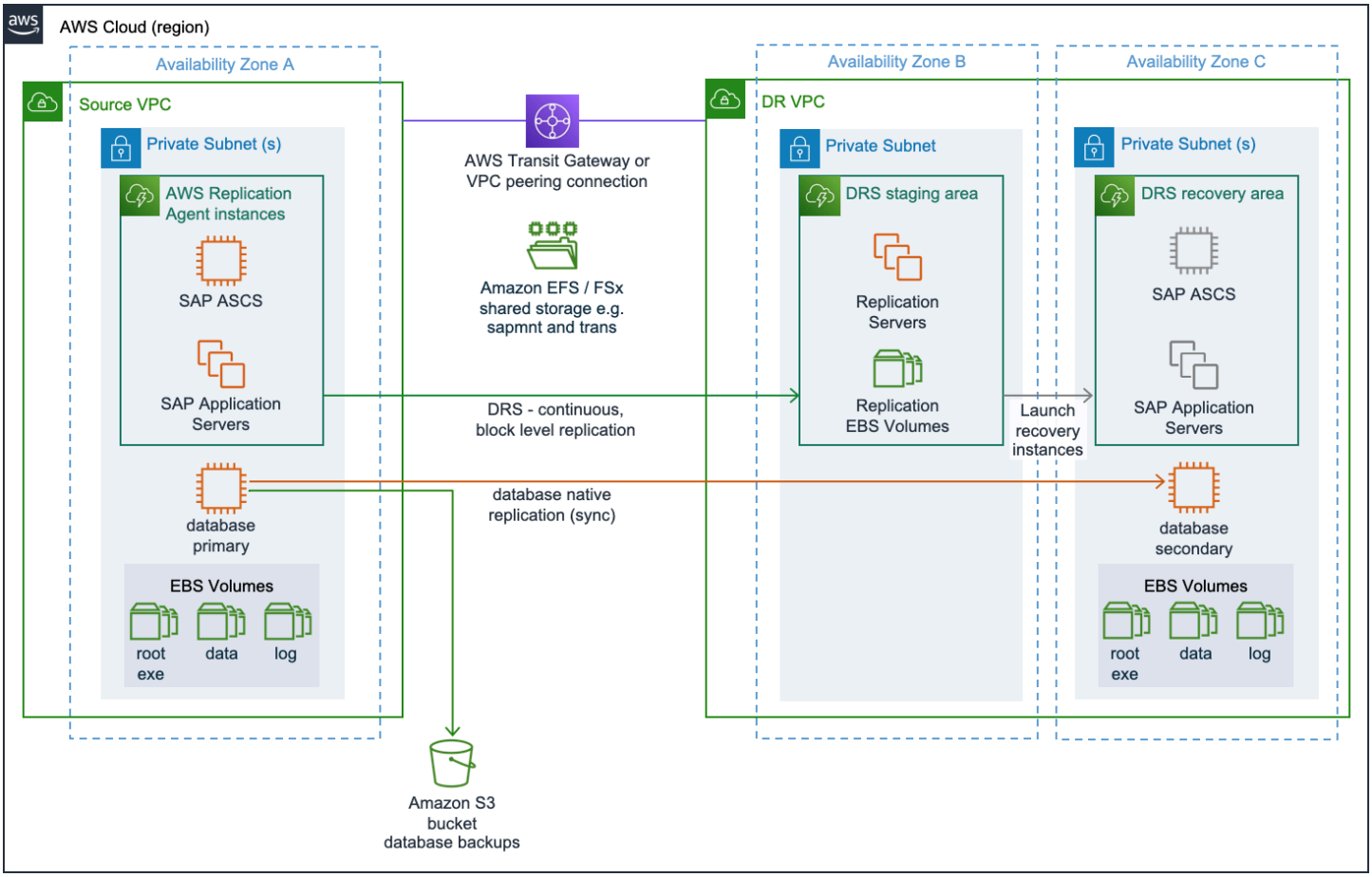
The following two sections cover the reference architectures for this scenario.

Full in-Region disaster recovery implementation



In full in-Region disaster recovery implementation, the source servers running SAP application components, such as central services instance ((A)SCS), primary application server (PAS), additional application server (AAS), and the database, are replicated using Elastic Disaster Recovery.

Hybrid in-Region disaster recovery implementation



In hybrid in-Region disaster recovery implementation, the source servers running SAP application components, such as central services instance [(A)SCS], primary application server (PAS), and additional application server (AAS) are replicated using Elastic Disaster Recovery. The database is replicated using a database native replication method.

AWS Cross-Region disaster recovery

A disaster recovery scenario with multiple AWS Regions enables business continuity with your data storage in two separate geographical locations. For more information, see Multi-Region architecture patterns.

Best practices

- Separate Amazon VPCs for the source and recovery areas in the same Region
- A shared AWS account for source and recovery areas
- AWS Transit Gateway or Amazon VPC peering for supporting replication traffic and end user connectivity

For more information, see Network.

- Replication via Amazon EFS or other file systems to protect shared storage between Regions

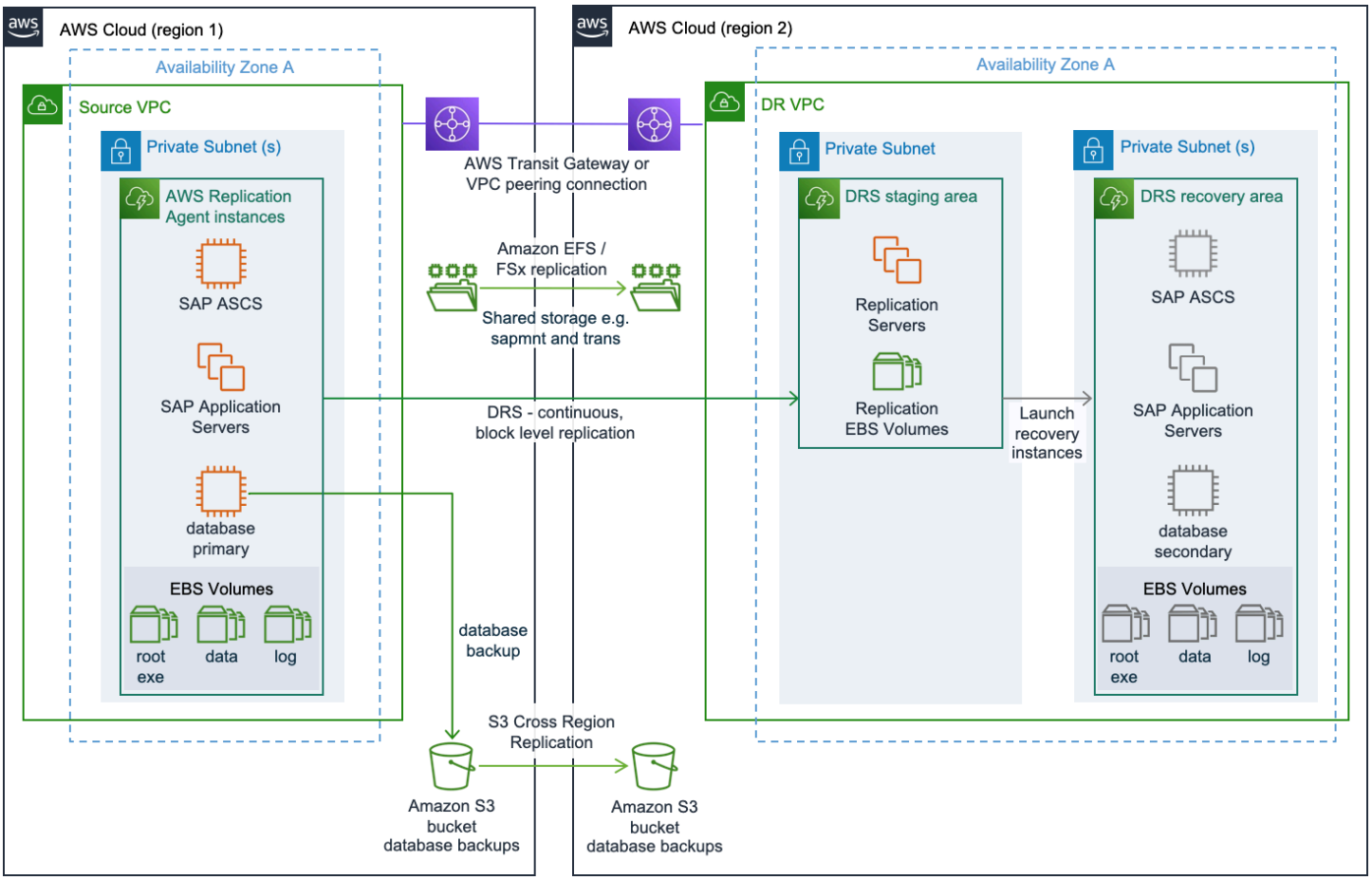
  For more information, see AWS Cross-Region disaster recovery.

> **Note**
>
> You can only replicate in the same AWS account with Amazon EFS.

- Amazon S3 cross-Region replication to provide a copy of your database backups and other Amazon S3 bucket data to disaster recovery Amazon VPC
- Separate subnets for the source, staging, and recovery areas

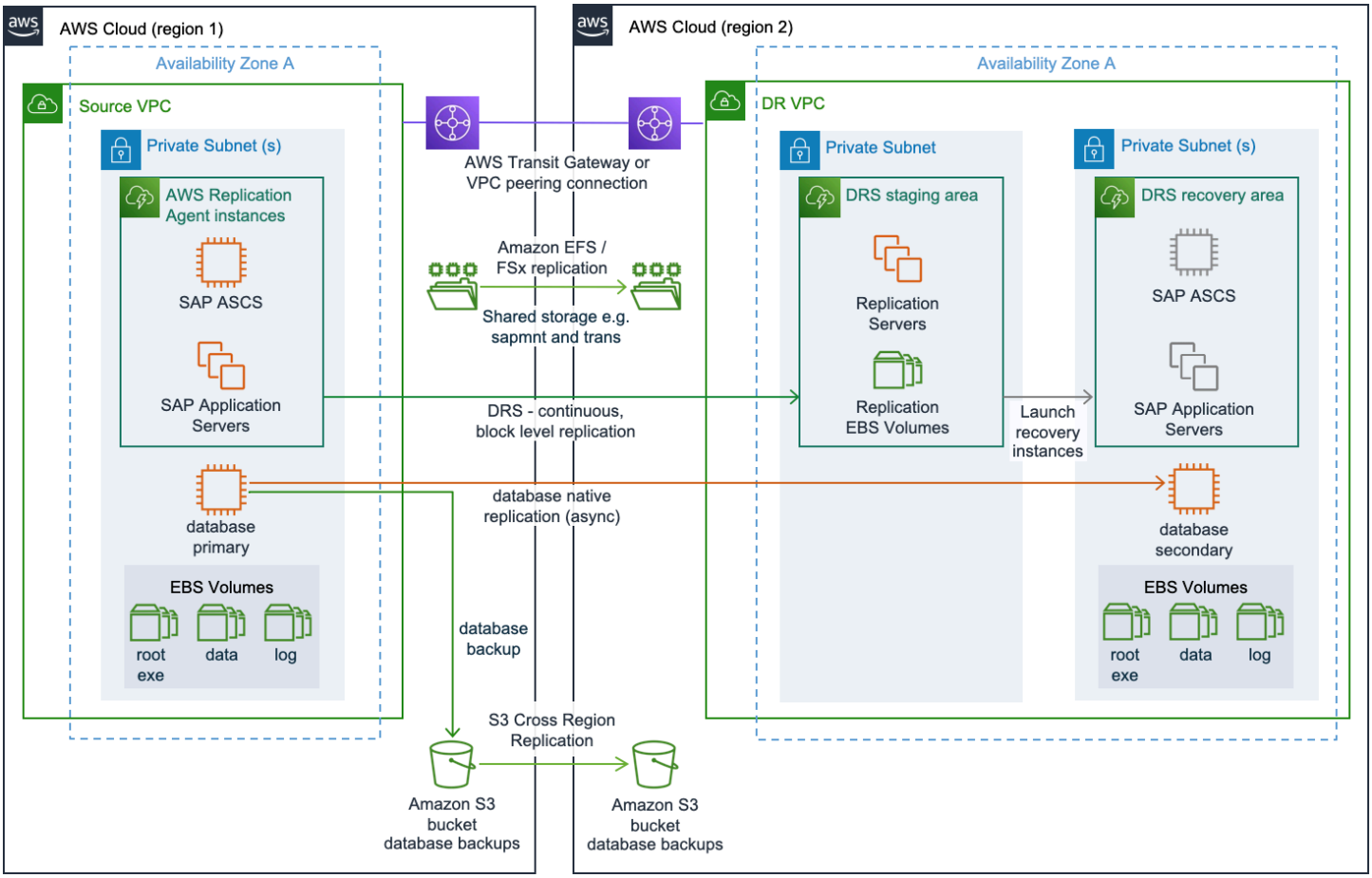
The following two sections cover the reference architectures for this scenario.

Full cross-Region disaster recovery implementation



In full cross-Region disaster recovery implementation, the source servers running SAP application components, such as central services instance ((A)SCS), primary application server (PAS), additional application server (AAS), and the database, are replicated using Elastic Disaster Recovery.

Hybrid cross-Region disaster recovery implementation



In hybrid cross-Region disaster recovery implementation, the source servers running SAP application components, such as central services instance [(A)SCS], primary application server (PAS), and additional application server (AAS) are replicated using Elastic Disaster Recovery. The database is replicated using a database native replication method.

Outside of AWS to AWS disaster recovery

In this scenario, the source systems are running in a non-AWS environment. A hybrid disaster recovery solution like this is can be implemented to quickly add resiliency to your existing production environments on other platforms.

Best practices

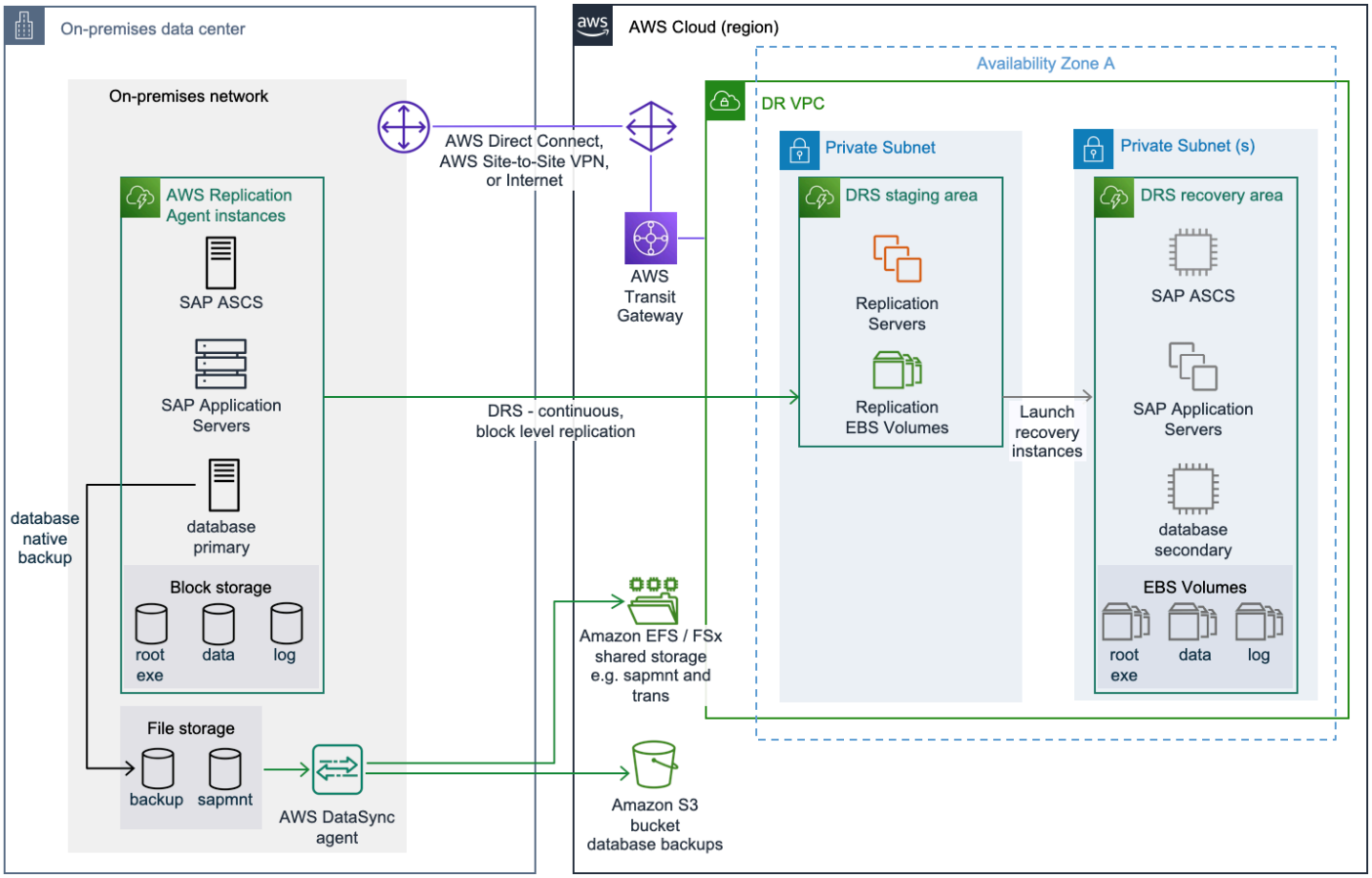
- AWS Direct Connect for supporting replication traffic and end user connectivity

  For more information, see Network.
- AWS DataSync to protect shared storage

  For more information, see Outside of AWS to AWS disaster recovery.
- Separate subnets for the staging and recovery areas

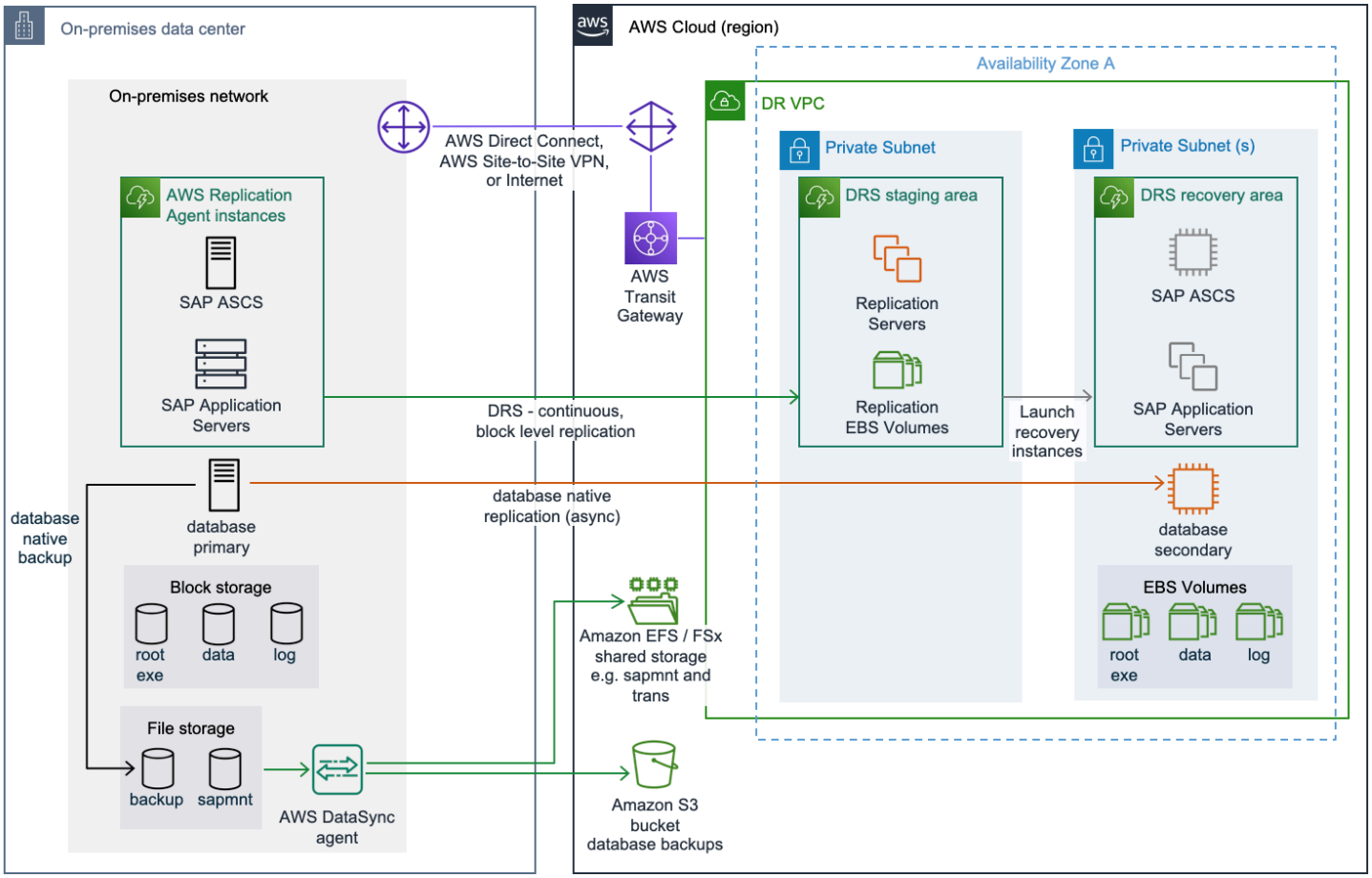
The following two sections cover the reference architectures for this scenario.

Full non-AWS to AWS disaster recovery implementation



In full non-AWS to AWS disaster recovery implementation, the source servers running SAP application components, such as central services instance ((A)SCS), primary application server (PAS), additional application server (AAS), and the database, are replicated using Elastic Disaster Recovery.

Hybrid non-AWS to AWS disaster recovery implementation



In hybrid non-AWS to AWS disaster recovery implementation, the source servers running SAP application components, such as central services instance [(A)SCS], primary application server (PAS), and additional application server (AAS) are replicated using Elastic Disaster Recovery. The database is replicated using a database native replication method.

For more disaster recovery options and information, you can reach out to AWS Support.

Shared storage resiliency

File systems on an SAP server can be created on block type storage, for instance, on locally attached disks or Enterprise Storage Area Network (SAN) devices, or may be based on shared file systems such as SMB or NFS shared volumes from servers or Network Attached Storage (NAS) devices.

As Elastic Disaster Recovery is a block level replication service, it only replicates the disks if they are represented as block storage devices. Other tools and processes must be used to provide resiliency for shared file systems. To address these requirements, we recommend using the fully managed

shared storage services of AWS that are easy and cost-effective to launch, run, and scale feature-rich, high performance, and resilient file systems in the cloud. The choice of your file system depends on the operating system of your disaster recovery scenario.

- Linux – Amazon Elastic File System (Amazon EFS)
- Microsoft Windows Server – Amazon FSx for Windows File Server (Amazon FSx)
- Mixed – Amazon FSx for Windows File Server or Amazon FSx for NetApp ONTAP (FSx for ONTAP)

The following sections provide guidance on file systems based on your disaster recovery scenario.

Topics

AWS In-Region disaster recovery

When using managed services such as Amazon EFS, FSx for ONTAP or FSx for Windows File Server to host your shared file systems, the built-in resiliency offered through their multi-Availability Zone design means that your shared storage is already disaster recovery ready. For further resiliency, ensure that your shared storage is backed up regularly, to protect against potential data corruption.

If you are sharing a file system using NFS or SMB protocols directly from one of your Amazon EC2 instances, you may not need additional steps if it is on Amazon EBS and attached to a server with the Replication Agent. This ensures replication via Elastic Disaster Recovery. If the shared file system is hosted on another Amazon EC2 instance along with other content that is not part of your SAP workload, use OS native tools like `rsync` to manage the replication of this file system to the recovery area.

You can also use AWS DataSync to provide selective replication. It can be scheduled to run once an hour at minimum, and replicate these files to the target storage in the recovery area. You must have an additional agent installed on an Amazon EC2 instance that has access to the file system. For more information, see How AWS DataSync works.

AWS Cross-Region disaster recovery

To support cross-Region disaster recovery, another shared file system must be available in the second Region. The data from the primary shared file system must be replicated on the shared file system in the second Region. Your implementation will differ based on your choice of AWS service.

- Amazon Elastic File System – Amazon EFS native replication can support cross-Region replication within a single AWS account.
- Amazon FSx for Windows File Server – You can also use AWS DataSync to replicate data between your primary to secondary shared storage. For more information, see How AWS DataSync works.
- Amazon FSx for NetApp ONTAP – You can use NetApp SnapMirror to copy your files between FSx for ONTAP file systems on your source and target instances, as frequently as every 5 minutes, to maintain a current copy of your shared file systems. For more information, see Scheduled replication using NetApp SnapMirror.

Outside of AWS to AWS disaster recovery

Depending on your source area design for shared storage, you must consider replicating these files on your disaster recovery instance in AWS. We recommend using AWS DataSync. It can copy data to and from services, such as NFS and SMB shares, along with file systems using Amazon EFS, FSx for Windows File Server, and FSx for ONTAP.

In certain scenarios, you can consider using other options to protect your source area SAP shared file systems, such as if the following is being used on your source environment.

- FSx for ONTAP – You can use NetApp SnapMirror to copy your files between FSx for ONTAP file systems on your source and target instances, as frequently as every 5 minutes, to maintain a current copy of your shared file systems. For more information, see Scheduled replication using NetApp SnapMirror.
- Local storage – Elastic Disaster Recovery will replicate it to your disaster recovery environment on AWS, if Replication Agent can be configured on the source server hosting the local storage.

Implementing disaster recovery on AWS cloud for SAP workloads

Using Elastic Disaster Recovery to implement a disaster recovery solution for SAP workloads on AWS follows different considerations for different parts of a typical SAP workload, such as S/4HANA deployment. The following sections provide guidance on the differences in how to design, implement, and manage Elastic Disaster Recovery when used for the application and the database layers.

Topics

SAP application layer

We recommend using AWS Elastic Disaster Recovery to protect your SAP application servers, such as SAP ASCS/SCS, PAS, AAS etc. Elastic Disaster Recovery supports the SAP application layer based on SAP NetWeaver, ABAP foundation, and stand-alone applications like TREX, content servers, and so on. You can use Elastic Disaster Recovery for Amazon EBS backed storages, such as SAP instance binaries, local files stored on an Amazon EBS volume.

The application layer also contains shared file systems, such as SAP mount, transport, and interface directories. These file systems usually need to be managed separately. For more information, see Shared storage resiliency.

To setup, install Elastic Disaster Recovery agent on the application servers. Create an IAM user with required permissions. Provide Elastic Disaster Recovery agent with the user information to establish a connection with Elastic Disaster Recovery APIs. Once the agent is configured, it engages in an authentication handshake with the TLS 1.3-encrypted Elastic Disaster Recovery API endpoint. The service produces identically sized Amazon EBS volumes in the staging area subnet, for each source volume that is duplicated, for data synchronization. The type of Amazon EBS volumes can be configured in the replication server settings. Replication starts after the staging area subnet resources are generated and the agent is installed. The data is transported with encryption from the source servers to the replication server directly. The service automatically manages the subnet resources for the staging area, scaling them up or down based on the concurrent replication of the source servers and disks.

SAP database layer

AWS Elastic Disaster Recovery is fully supported as disaster recovery solution for SAP applications running on any database, and for SAP applications running on SAP HANA database in scale-up configuration. It is not supported for replication of multi-node SAP databases, such as SAP HANA scale-out cluster.

The data in an SAP system is stored in a database. This data includes master data, transactional data, and ABAP artifacts. You must consider your business RPO and RTO requirements when evaluation Elastic Disaster Recovery for a disaster recovery solution. The service is not application aware but works at the operating system layer by replicating the attached storage to the target staging environment. Based on your RTO and RPO requirements, you can select Elastic Disaster Recovery or database native replication methods, such as SAP HANA System Replication (HSR) for SAP HANA.

The following are the important considerations to choose your database replication method.

Topics

Network bandwidth

AWS Elastic Disaster Recovery works at the operating system layer, with block level replication of attached storage devices. Depending on the change rate at the source, you may need higher network bandwidth to stay current with replication. Database aware technologies such as SAP HSR require lesser network bandwidth, enabling faster replications for systems with high rate of change.

RPO

Elastic Disaster Recovery supports sub-second RPO. For SAP workloads, ensure that your network can support peaks in change rate. If your RPO is very small, we recommend testing database native replication methods along with Elastic Disaster Recovery.

Actions that lead to significant changes to the data of your database cause delays data replication on staging area. It can include a partial or full recovery of a backup to protected volumes for a database on source server. The changes made to your storage volumes are much higher than your usual change rates on source server. Data restored from backup to protected volumes on the source server is treated as changed blocks and is replicated by Elastic Disaster Recovery. The replication servers need additional time to receive and write this larger amount of changed data from the source system. This can impact your business RPO.

It is recommended to manage actions, such as recovery from backups, at less critical workload times. This way, longer RPO values won't impact your workload. You can track the amount of changed data still waiting to be replicated with Elastic Disaster Recovery. For more information, see Recovery dashboard.

Change rate

For databases with high change rates, you can meet the performance requirements with sufficiently performing networks, along with the storage and compute configuration of the replication server. If these changes are insufficient to meet the business performance requirements, you can choose database native replication methods to optimize your RPO.

RTO

With Elastic Disaster Recovery, the target disaster recovery environment is provisioned once the disaster recovery event is triggered. The total time depends on the size of your database, and the chosen Point-in-Time (PiT). You must test your disaster recovery scenario before implementation on production environments.

Cost

As Elastic Disaster Recovery is not using a warm or hot standby approach, the compute costs are minimized for your disaster recovery environment as compared to many other disaster recovery options. For more information, see AWS Elastic Disaster Recovery pricing. With database native replication methods, costs can increase with the compute resources in the disaster recovery area.

RCO

If you have multiple tightly coupled systems, then you need to use database native replication methods.

Storage limits

In most cases, the available Amazon EBS volume types are sufficient to address any storage capacity and performance needs. Depending on the source environment architecture, in some cases, the storage volume on the recovery instance exceeds the capacity and/or performance limits of individual Amazon EBS volumes. This can happen in a non-AWS to AWS disaster recovery implementation with `data` and `log` volumes attached to high workload database servers. For more information, see Amazon EBS volume types.

When migrating servers to AWS, such storage volumes must be refactored to a new storage architecture, for instance, creating striped volume sets. Striped volume sets are defined and maintained using logical volume manager tools in your recovery instance's operating system. For more information, see RAID configuration on Linux. These volume sets will span two or more Amazon EBS volumes, up to the total needed to meet the required volume size and performance. The storage volume data is then copied to the new striped volume set. While it may be possible to automate this process through Elastic Disaster Recovery post-launch scripts or alarm events which trigger code through Amazon EventBridge event rules, the additional steps can cause longer recovery time.

In these cases, implementing a hybrid disaster recovery solution is suitable. Most of the servers are managed by Elastic Disaster Recovery and select servers (with storage performance considerations) use alternative disaster recovery approaches, such as native database replication technologies. The storage architecture refactoring is done when the standby replication server is set up during the initial disaster recovery environment implementation. As the replication now happens at an application level, the disaster recovery server is able to write to a storage architecture that is different from what is on the source server.

RISE with SAP on AWS Cloud

RISE with SAP S/4HANA Cloud, private edition is a cloud ERP offering from SAP. Along with ERP, it includes Business Process Intelligence, Business Platform and Analytics, and Business Networks. SAP maintains responsibility for the holistic service level agreement, cloud operations, and technical support for RISE. You can choose your own cloud service provider in RISE with SAP.

SAP S/4 HANA Cloud, private edition is a single-tenant setup where different customer environments are isolated by AWS accounts and a dedicated Virtual Private Cloud (VPC).

> **Important**
>
> SAP owns and manages the AWS account where RISE with SAP is deployed, and is responsible for the AWS services used to serve your SAP landscape on AWS.

SAP is responsible for security in the cloud in RISE with SAP. For more information, see AWS Cloud Security – Shared Responsibility Model and SAP and Hyperscalers: Clarifying Security in the Cloud. In addition to the security provided by SAP, you can also implement additional security for your SAP landscape. See the Security section for more details.

In your AWS account managed by SAP, SAP manages the AWS services required to run your SAP landscape on AWS. You can still utilize AWS services to extend RISE with SAP in your own AWS account that is not managed by SAP. For example, you can create a data lake with Amazon AppFlow or AWS Glue. See the Extensions section for more details.

> **Note**
>
> You must create a separate AWS account or use your existing AWS account that is not managed by SAP for creating extensions with AWS services.

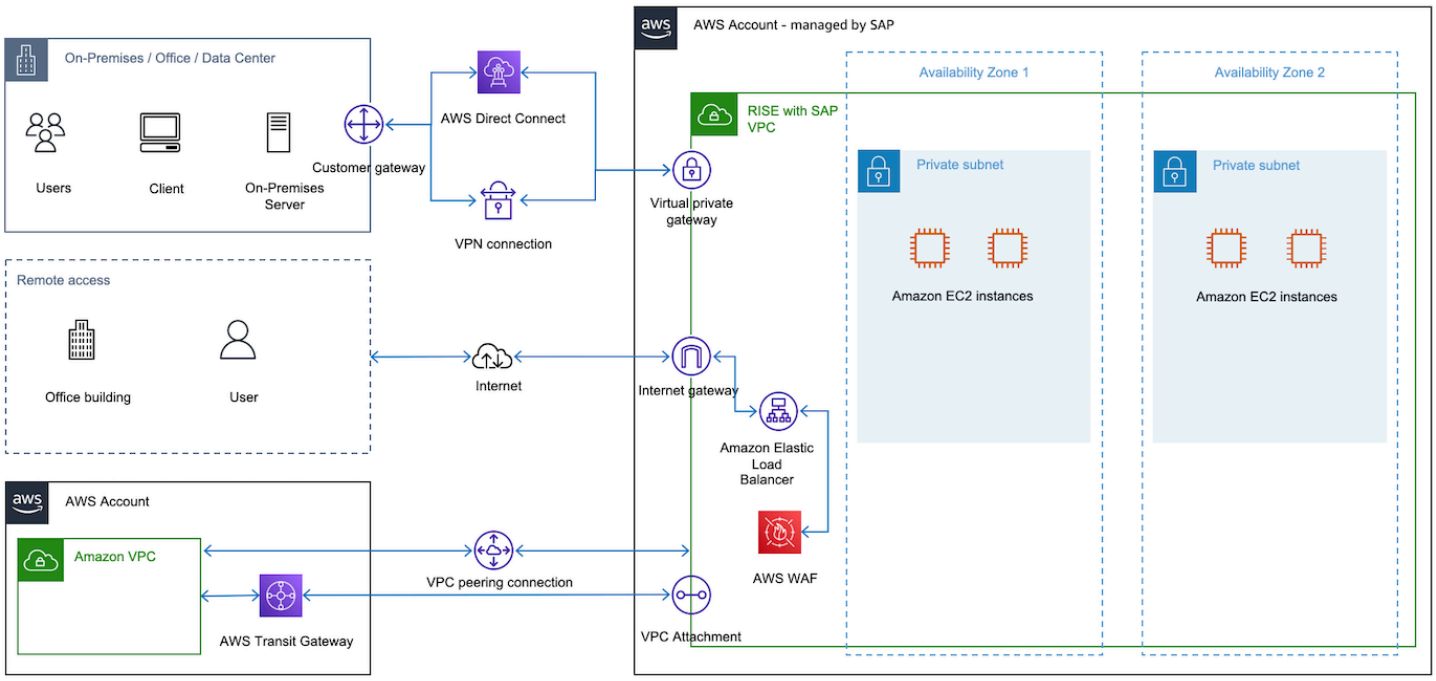
SAP avails Support for AWS account that is managed by SAP. You are not required to establish additional Support for the AWS account managed by SAP.

This documentation is focused on RISE with SAP S/4HANA Cloud, private edition and SAP S/4HANA Cloud, private edition, tailored option. The following topics are covered in this document.

Topics

Connectivity

You must establish connectivity between AWS cloud where your RISE with SAP solution is running and on-premises data centers. You also need a connection for direct data transfer (to avoid routing data via your on-premises locations) and communication between SAP systems and your applications running on AWS cloud. The following image provides an example overview of connectivity to RISE with SAP VPC.



See the following topics for further details:

Topics

- Connect to RISE through nearest AWS Direct Connect POP (including AWS Local Zone)
- Decision tree on connectivity to RISE
- Other considerations

Roles and responsibility for establishing connectivity to RISE

Under RISE with SAP, the SAP Enterprise Cloud Services (ECS) team manages the SAP S/4HANA Private Cloud Environment. The *Supplemental Terms and Conditions* provided by SAP has a section on Excluded Tasks. You are responsible for running such tasks. You can also use a third-party service provider to manage the excluded tasks for you. For further details, see SAP Product Policies.

The primary task required for deploying RISE with SAP is to establish network connectivity to RISE with SAP VPC on AWS. As per the RISE with SAP agreement, you are responsible for establishing a connection to RISE.

We recommend that you spend time understanding the available options on how to connect your on-premises network and/or existing AWS accounts to RISE with SAP VPC on AWS. Review the subsequent sections for more information.

Connecting to RISE from on-premises networks

Connectivity to RISE with SAP on AWS from on-premises is supported using AWS VPN or AWS Direct Connect or a combination of the two.

Topics

- Connecting to RISE with SAP VPC using AWS VPN
- Connecting to RISE with SAP VPC using AWS Direct Connect
- Connecting to RISE with SAP VPC using SD-WAN
- Implementation steps for connectivity between RISE and your on-premises networks

Connecting to RISE with SAP VPC using AWS VPN

Enable access to your remote network from RISE with SAP VPC using AWS Site-to-Site VPN. Traffic between AWS cloud and your on-premises location is encrypted via Internet Protocol security (IPsec) and transferred through a secure tunnel on internet. This option is efficient, and faster to

implement when compared to AWS Direct Connect. For more information, see Connect your VPC to remote networks using AWS Virtual Private Network.

You can get a maximum bandwidth of up to 1.25 Gbps per VPN tunnel. For more information, see Site-to-Site VPN quotas.

To scale beyond the default maximum limit of 1.25 Gbps throughput of a single VPN tunnel, see How can I achieve ECMP routing with multiple Site-to-Site VPN tunnels that are associated with a transit gateway?

When using this option, SAP requires the following details:

- BGP ASN
- IP address of your device

You can obtain these details from your AWS VPN device on-premises.

When connecting your remote network directly to RISE using AWS Site-to-Site AWS VPN, the cost for the AWS VPN Connection and the cost for data transfer out are included in the RISE subscription.

For more information see: AWS Site-to-Site AWS VPN Pricing.

Note: Because the cost associated with the lifecycle and operation of a "Customer gateway device" (a physical device or software application on your side of the Site-to-Site AWS VPN connection) varies, this is not taken into consideration in this document.

Connecting to RISE with SAP VPC using AWS Direct Connect

Use AWS Direct Connect if you require a higher throughput or more consistent network experience than an internet-based connection. AWS Direct Connect links your internal network to an AWS Direct Connect location over a standard Ethernet fiber-optic cable. You can create different types of virtual interfaces (VIFs) to connect with various AWS services. For example, you can create a Public VIF to communicate with public services like Amazon S3 or a Private/Transit VIF for private resources such as Amazon VPC, while bypassing the internet service providers in your network path. For more information, see AWS Direct Connect connections.

You can choose from a dedicated connection of 1 Gbps, 10 Gbps, 100 or 400 Gbps or an AWS Direct Connect Partner's hosted connection where the Partner has an established network link

with AWS cloud. Hosted connections are available from 50 Mbps. 100 Mbps, 200 Mbps, 300 Mbps, 400 Mbps, 500 Mbps, 1 Gbps, 2 Gbps, 5 Gbps, 10 Gbps, and 25 Gbps. You can order hosted connections from an AWS Direct Connect Delivery Partner approved to support this model. For more information, see AWS Direct Connect Delivery Partners.

To connect, use a virtual private gateway in AWS account managed by SAP or a Direct Connect gateway in your AWS account associated with a virtual private gateway in AWS account managed by SAP. For more information, see Direct Connect gateways. Direct Connect gateway can also connect to a AWS Transit Gateway. For more information, see Connecting to RISE using your single AWS account.

You must acquire a *Letter of Authorization* from SAP to setup a AWS Direct Connect dedicated connection in the AWS account managed by SAP.

When connecting your remote network directly to RISE using AWS Direct Connect, the cost for data transfer out (egress) is included in the RISE subscription. Costs associated to the capacity (the maximum rate that data can be transferred through a network connection) and the port hours (the time that a port is provisioned for your use with AWS or an AWS Direct Connect Delivery Partners) are not included in the RISE subscription. AWS Direct Connect does not have setup charges, and you may cancel at any time, however, services provided by your AWS Direct Connect Delivery Partners or other local service provider may have other terms and conditions that apply.

For more information, see: AWS Direct Connect Pricing

Connecting to RISE with SAP VPC using SD-WAN

What is SD-WAN

Software-Defined Wide Area Networking (SD-WAN) is a networking technology that uses software to manage and route traffic across different networks such as Multi-Path Label Switching (MPLS), public internet, or the AWS backbone focusing on improving connectivity and application performance. SD-WAN primarily operates at layer 3 (Network Layer) of the network OSI model offering centralized control, routing, path selection, IP-based policies, and the ability to prioritize specific mission critical applications, such as SAP, making it well-suited for cloud-based RISE with SAP environments.

Although SD-WAN primarily operates at Layer 3, using an overlay network such as broadband internet, it can utilize Layer 2 (Data Link) technologies such as AWS Direct Connect as the underlay network for transport, and Layer 3 (Network) technologies such as AWS Site-to-Site VPN.

In SD-WAN architecture, an SD-WAN headend acts as a hub or centralized network component, while SD-WAN edge devices deployed at branch offices, remote sites or data centers which serves as the entry and exit points for WAN Traffic.

You can refer to more detailed information in the Reference Architectures for Implementing SD-WAN Solutions on AWS.

Scenario A: SD-WAN appliances (edge and/or headend/hub) on-premises

AWS Transit Gateway Connect allows you to extend your SD-WAN network to AWS using GRE (Generic Routing Encapsulation) tunnels without needing additional AWS infrastructure. Through Transit Gateway Connect Peer, you can establish GRE tunnels between your transit gateway in your AWS account and the SD-WAN appliance on-premises which are connected via AWS Direct Connect connection as underlying transport.

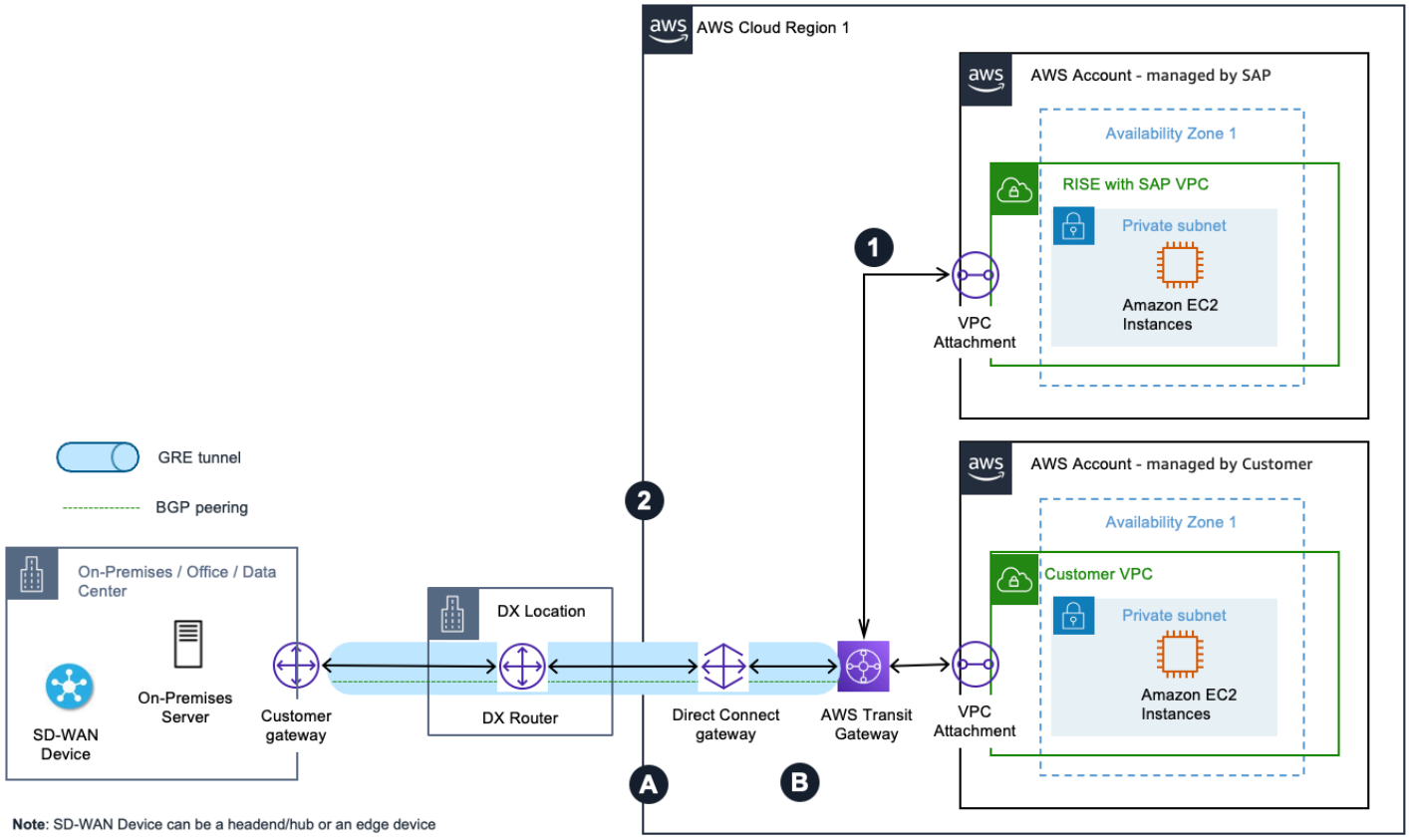
The appliance must be configured to send and receive traffic over a GRE tunnel to and from the transit gateway using the Connect attachment. The appliance must be configured to use BGP (Border Gateway Protocol) for dynamic route updates and health checks.

Each connection can be configured with its own route table and BGP peer, enabling you to extend your on-premises network segmentation via Virtual routing and forwarding (VRF) to aws. The RISE with SAP VPC is attached to the AWS Transit Gateway.

This setup provides a streamlined way to connect your SD-WAN environment with RISE with SAP on AWS using AWS Direct Connect, maintaining network separation while simplifying the overall architecture.

In this scenario, the overlay network is SD-WAN (with GRE Tunnels) with the headend/hub or edge devices deployed on on-premises, and the underlay transport is AWS Direct Connect

Pattern A-1: SD-WAN devices integration with AWS Transit Gateway and AWS Direct Connect with your AWS landing zone



The preceding diagram illustrates a pattern of how you can extend and segment your SD-WAN traffic to AWS without adding extra infrastructure. You can create Transit Gateway connect attachments using an AWS Direct Connect connection as underlying transport in your AWS account.

Outbound from RISE with SAP VPC:

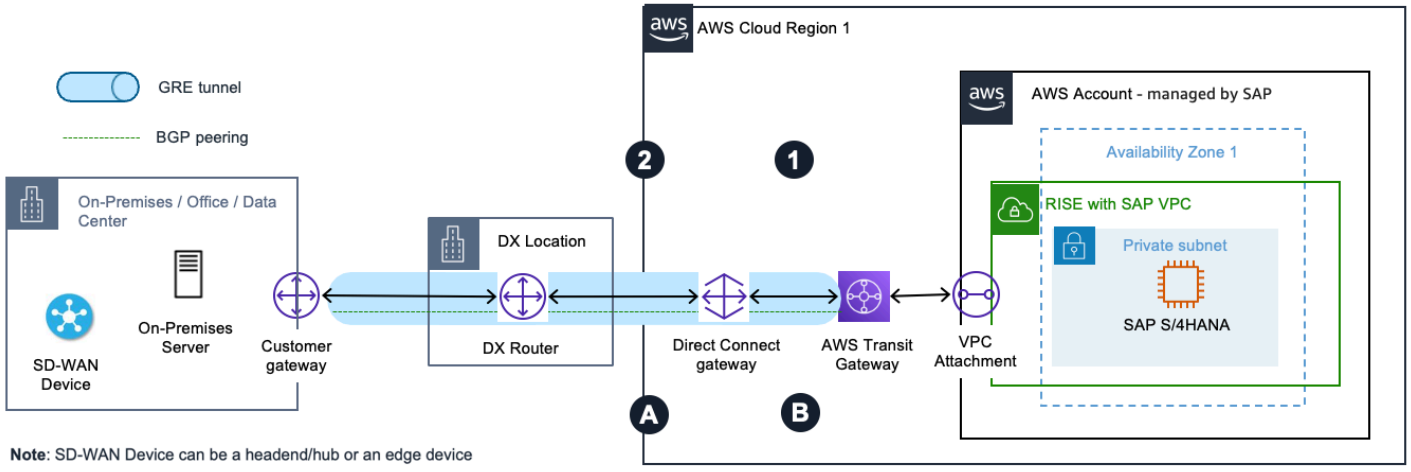
1. Traffic initiated from the RISE VPC to the corporate data center is routed to the Transit Gateway.
2. The Transit Gateway connect attachment uses the Direct Connect connection as the underlay transport and connects the Transit Gateway to the corporate data center SD-WAN device with GRE tunneling and BGP.

Inbound to RISE with SAP VPC:

A. Traffic from the corporate data center SD-WAN device to the RISE VPC is forwarded to the Transit Gateway via the GRE tunnel of the Transit Gateway attachment over the Direct Connect link.

B. Transit Gateway forwards the traffic to the destination RISE with SAP VPC.

Pattern A-2: SD-WAN devices integration with AWS Transit Gateway and AWS Direct Connect with no AWS landing zone



The preceding diagram illustrates a pattern of how you can extend and segment your SD-WAN traffic to AWS without adding extra infrastructure. In RISE with SAP, you can request SAP to create Transit Gateway connect attachments using a Direct Connect connection as underlying transport. Customers can leverage SAP-managed Direct Connect gateway (DXGW) if required.

Outbound from RISE with SAP VPC:

1. Traffic initiated from RISE VPC to the corporate data center is routed to the Transit Gateway.
2. The Transit Gateway connect attachment uses the Direct Connect connection as transport and connects the Transit Gateway to the corporate data center SD-WAN device using GRE tunneling and BGP.

Inbound to RISE with SAP VPC:

A. Traffic from the corporate data center SD-WAN device to the RISE VPC is forwarded to the Transit Gateway via the GRE tunnel of the Transit Gateway attachment over the Direct Connect link.

B. Transit Gateway forwards the traffic to the destination RISE with SAP VPC.

Scenario B: SD-WAN appliances (edge and/or headend/hub devices) in AWS

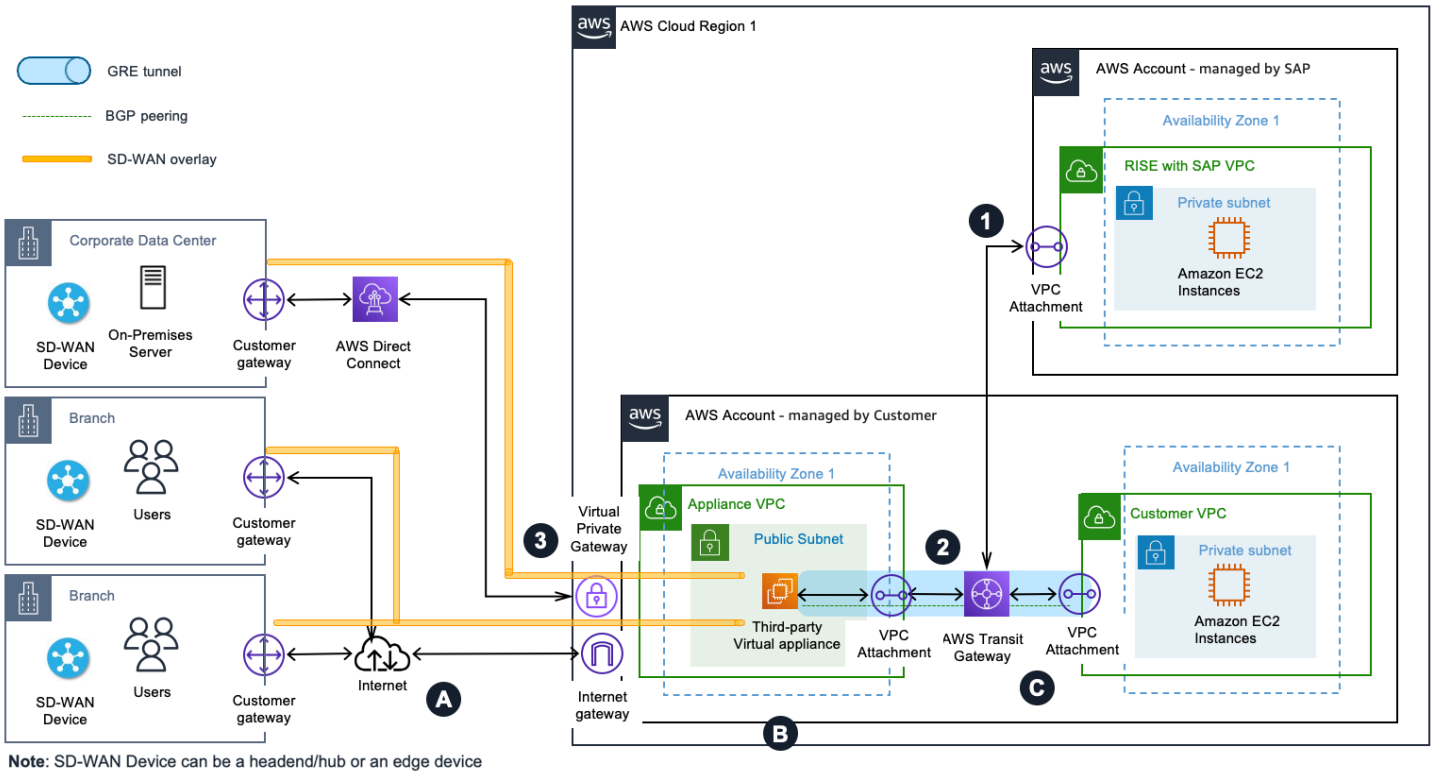
In this scenario, the virtual appliances of the SD-WAN network are deployed in a VPC within aws. Then, you use a VPC attachment as underlying transport for the Transit Gateway connect attachment between the SD-WAN virtual appliances and the Transit Gateway in your AWS

account(s). Similar to Scenario A, Transit Gateway connect attachments support GRE for higher bandwidth performance compared to a VPN connection. It supports BGP for dynamic routing and removes the need to configure static routes. In addition, its integration with Transit Gateway Network Manager provides advanced visibility through global network topology, attachment level performance metrics, and telemetry data.

Between on-premises and AWS, the overlay network is SD-WAN with GRE or IPSec tunnels with the headend/hub deployed within AWS, and the underlay transport could be Internet, MLPS, or Direct Connect. Following are the architecture patterns under this scenario:

Note: Network patterns covered in the following sections are applicable only with your existing or a new landing zone setup on aws. For SD-WAN appliances deployment and connectivity directly with AWS Account – managed by SAP, refer to Pattern A-2.

Pattern B-1: SD-WAN appliances in AWS integrated with AWS Transit Gateway Connect with your AWS landing zone



The preceding diagram illustrates a pattern of integrating your SD-WAN network with Transit Gateway using connect attachments and placing (third-party) virtual appliances of the SD-WAN network in an Appliance VPC within aws. It's common to have SD-WAN edge appliances deployed at branch locations, and on-premises data center to create a full mesh topology.

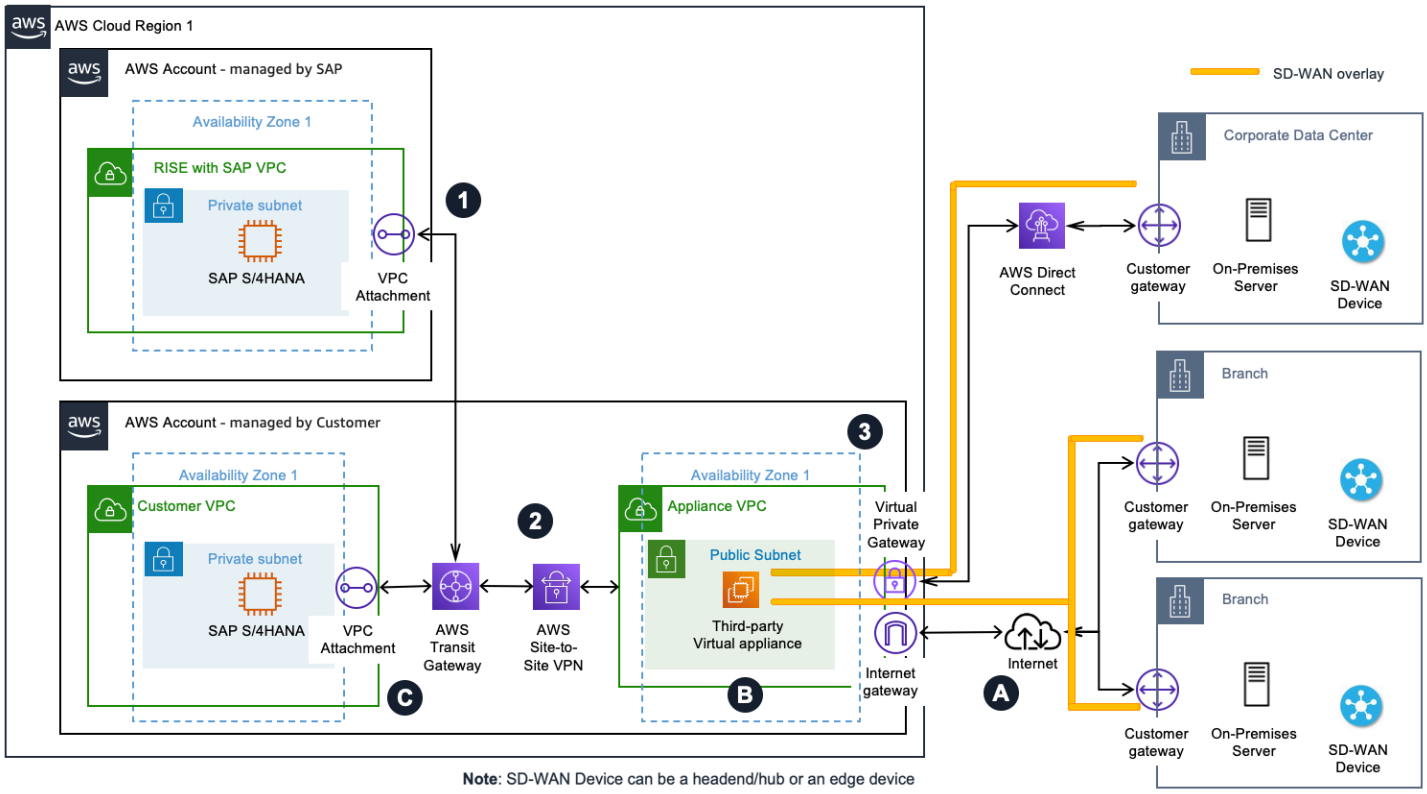
Outbound from RISE with SAP:

1. Traffic initiated from the RISE VPC to the corporate data center is routed to the Transit Gateway.
2. The Transit Gateway connect attachment uses the VPC attachment as transport and connects Transit Gateway to the third-party appliance in the Appliance VPC using GRE tunneling and BGP.
3. The third-party virtual appliance encapsulates the traffic, which uses the SD-WAN overlay – on top of the Direct Connect link – to reach the corporate data center.

Inbound to RISE with SAP:

A. Traffic from branches outside AWS to the RISE VPC reaches the internet gateway of the appliance VPC via the SD-WAN overlay over the internet. Similarly, traffic from the corporate data center to the RISE VPC reaches the virtual private gateway of the Appliance VPC via the SD-WAN overlay over the Direct Connect link.
B. The third-party virtual appliance in the appliance VPC forwards the traffic to the Transit Gateway via the connect attachment.
C. Transit Gateway forwards the traffic to the destination RISE VPC.

Pattern B-2: SD-WAN appliances in AWS integrated with AWS Site-to-Site VPN



The diagram above illustrates a pattern of integrating your SD-WAN network with Transit Gateway using an AWS Site-Site VPN connection and placing (third party) virtual appliances of the SD-WAN network in an Appliance VPC within aws. You may use this option when your third-party virtual appliance does not support GRE. It's common to have SD-WAN edge appliances deployed at branch locations, and on-premises data center to create a full mesh topology.

Outbound from RISE with SAP:

1. Traffic initiated from the RISE VPC to the corporate data center is routed to the Transit Gateway Elastic Network Interface (TGW ENI).
2. The traffic is routed between the Transit Gateway and the third-party virtual appliance using the Site-to-Site VPN connection.
3. The third-party virtual appliance encapsulates the traffic, which uses the SD-WAN overlay – on top of the Direct Connect link – to reach the corporate data center.

Inbound to RISE WITH SAP:

A. Traffic from branches outside AWS to the RISE VPC reaches the internet gateway of the appliance VPC via the SD-WAN overlay over the internet. Similarly, traffic from the corporate

data center to the RISE VPC reaches the virtual private gateway of the appliance VPC via the SD-WAN overlay over the AWS Direct Connect link.

B. The third-party virtual appliance in the appliance VPC forwards the traffic to the Transit Gateway via Site-to-Site VPN connection.

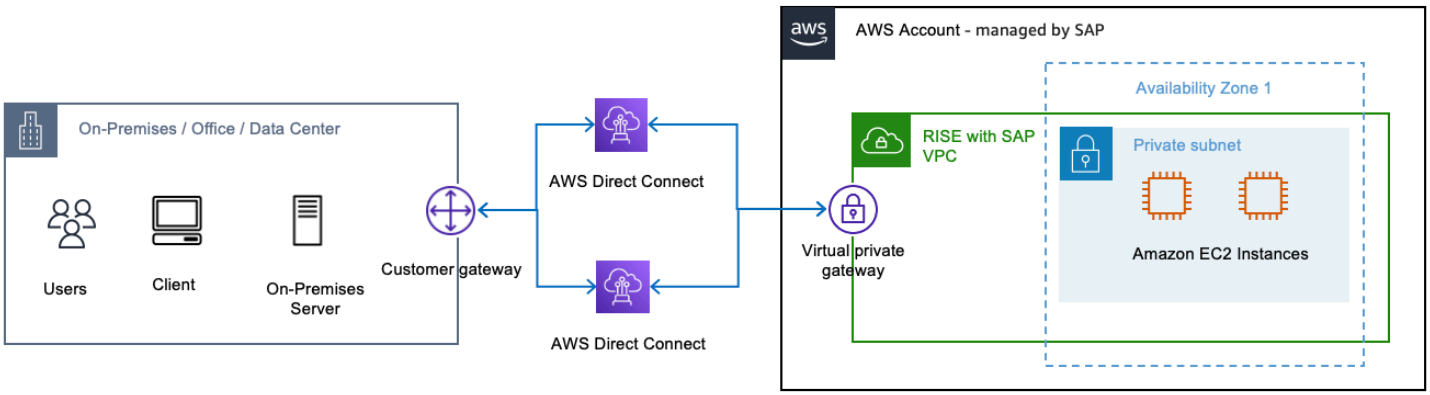
C. Transit Gateway forwards the traffic to TGW ENI of the destination RISE VPC.

Implementation steps for connectivity between RISE and your on-premises networks

This section provides a deeper dive into the implementation steps for connectivity between RISE with SAP and your on-premise environments (without any Customer managed AWS Account usage). The two options we will step into are: first, creating highly resilient deployment for critical workloads, and second, creating cost effective alternative for non-critical workloads.

For each option we'll provide clarity on the details SAP needs, the steps you will take in your on-premise environment.

Option 1: Resilient Deployment for Critical Workloads



AWS Direct Connect (DX) comes in two connection types, namely Dedicated and Hosted. A Dedicated DX is a physical Ethernet connection associated with a single customer, between the customer's private network and aws. Hosted DX is a physical Ethernet connection that an AWS Direct Connect Partner provisions on behalf of a customer. Learn about AWS Direct Connect to familiarize yourself with the service.

To set up a resilient Direct Connect solution for your RISE with SAP deployment, follow these implementation steps:

Prerequisites

Before configuring the Direct Connect connection, ensure your on-premises network is ready. This includes:

- Reviewing the AWS documentation on BGP with AWS Direct Connect for detailed guidance on router configuration.
- Configuring Border Gateway Protocol (BGP) on your routers with MD5 authentication. BGP is a requirement for using Direct Connect.
- Verifying that your network can support multiple BGP connections for redundancy.

Initiate the Setup Process

Start by contacting your SAP ECS (Enterprise Cloud Services) representative and request the "AWS Connectivity Questionnaire" for RISE with SAP on AWS Direct Connect setup. This questionnaire will help gather the necessary information to provision the Direct Connect connection.

We advise you to set up redundant connections for high availability by completing the questionnaire for each Direct Connect connection you plan to establish. Review the Direct Connect Resiliency Recommendations to understand best practices.

Complete the SAP Questionnaire

When filling out the AWS Connectivity Questionnaire, specify that you want to set up a resilient AWS Direct Connect configuration.

In the questionnaire, provide the following details about your Direct Connect connection:

- Whether it’s a new or dedicated Direct Connect connection
- The Direct Connect provider or partner you’ll be using
- The specific Direct Connect region/location
- The minimum number of Direct Connect links required
- The subnet CIDR blocks for the primary and secondary Direct Connect links (in /30 CIDR format)
- The VLAN ID
- The Autonomous System Number (ASN) of your on-premises router
- The IP address ranges of your on-premises network (to allow for proper firewall configuration) Additionally, include information about your on-premises router, such as the make, model, and interface details.

Submit the completed questionnaire to your SAP ECS representative. SAP will then use this information to provision the necessary Direct Connect resources in your RISE with SAP environment on aws.

SAP's Responsibilities

After you submit the completed questionnaire, SAP will handle the following tasks (the list below is illustrative only for this context):

- Create a virtual interface (depending on your DX type: hosted or dedicated)
- Create the Direct Connect Gateway
- If you need SAP to provision Transit Gateway in RISE VPC,
  - Setup the Transit Gateway (including the ASN you provided)
  - Create the Transit Gateway attachment for your VPC
  - Update the route tables to allow the Transit Gateway to communicate with the RISE with SAP network VPC
  - Associate the Transit Gateway with the Direct Connect Gateway, including the CIDR of the RISE with SAP network that will be advertised to your network

Complete the Setup Process

Once you receive the necessary information from SAP, such as the VLAN ID, BGP peer IPs, and optional BGP authentication key, configure your on-premises routers accordingly. This includes setting up the VLAN interface and BGP for the Direct Connect connection. Consult the AWS documentation on router configuration for Direct Connect for detailed instructions.

Configure for active/active topology: Implement routing policies to balance traffic across the redundant Direct Connect connections, leveraging BGP communities or more-specific subnet advertisements to influence path selection from AWS to your on-premises network.

Establish and Test the Connections

Coordinate with SAP to enable the BGP sessions for both Direct Connect connections. Verify the BGP paths and test failover scenarios by simulating the failure of one connection to ensure traffic properly fails over to the other.

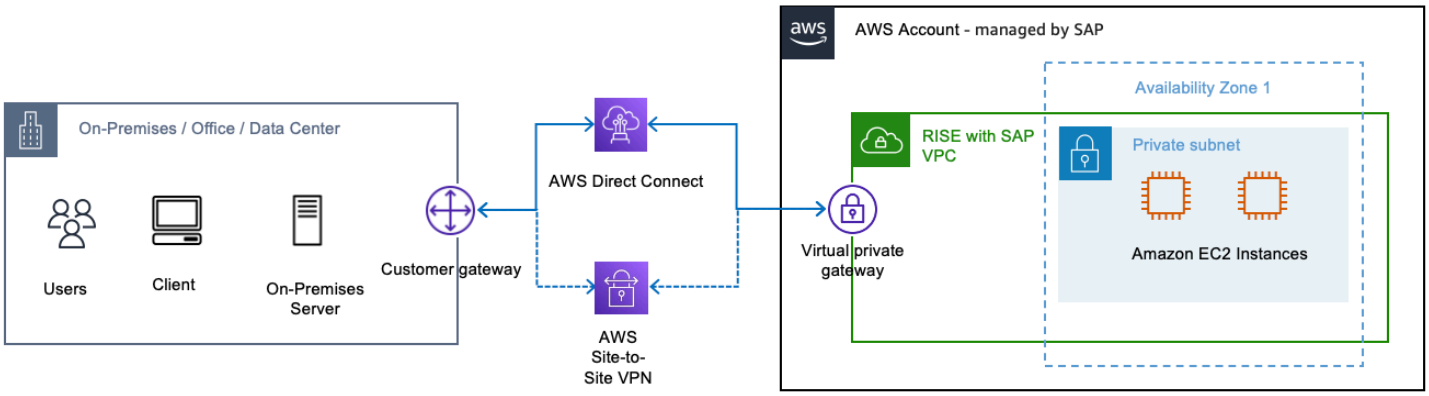
Confirm end-to-end connectivity with SAP for both paths. You can also leverage the AWS Direct Connect Resiliency Toolkit to perform scheduled failover tests and verify the resiliency of your connections. and validate the resiliency of your connections.

Maintain the Connections

Regularly review and update the Direct Connect configurations as needed. Coordinate any changes with SAP. Monitor the performance and availability of both connections, and refer to the AWS documentation on Monitoring Direct Connect for best practices.

By following these steps, you can establish a resilient AWS Direct Connect solution to securely connect your on-premises infrastructure with the RISE with SAP environment on AWS, ensuring high availability and reliable network performance.

Option 2: Cost Effective Alternative for Non-Critical Workloads



Some AWS customers prefer the benefits of one or more AWS Direct Connect connections as their primary connectivity to AWS, coupled with a lower-cost backup solution. Additionally, they may want an agile and adaptable connection that can be quickly established or decommissioned between network locations globally. To achieve these objectives, they can implement AWS Direct Connect connections with an AWS Site-to-Site VPN backup.

The Site-to-Site VPN connection consists of three key components:

1. Virtual Private Gateway (VGW) - The router on the AWS side
2. Customer Gateway (CGW) - The router on the customer side
3. The S2S VPN connection that binds the VGW and CGW together over two secure IPSec tunnels in an active/passive configuration For in-depth documentation on establishing the AWS Site-to-Site VPN connection, you can refer to the AWS documentation at https://docs.aws.amazon.com/vpn/latest/s2svpn/SetUpVPNConnections.html.

Prerequisites

This approach builds on the steps outlined in the previous Option 1 for setting up a Resilient AWS Direct Connect solution. After completing those Direct Connect implementation steps, you can add an Site-to-Site VPN connection as a failover option.

While your Direct Connect connections are being provisioned, you can begin preparing your on-premises infrastructure for the VPN setup: * Review the AWS documentation on Site-to-Site VPN to understand the requirements and best practices. * Ensure your firewalls allow the necessary traffic for the VPN tunnels. * Confirm you have two customer gateway devices or a single device capable of managing multiple VPN tunnels.

The addition of an Site-to-Site VPN connection provides a faster and more agile backup to your primary Direct Connect links. It's a similar process to setting up the Direct Connect, but with a few key differences.

Initiate the Setup Process

Start by contacting your SAP ECS representative again and request the "AWS Connectivity Questionnaire" for adding an AWS Site-to-Site VPN connection to your RISE on AWS setup. Inform SAP of your intent to implement the VPN as a failover to your Direct Connect links.

Complete the SAP Questionnaire

When filling out the AWS Connectivity Questionnaire this time, specify that you want to set up an AWS Site-to-Site VPN in addition to the Direct Connect connections.

In the AWS Connectivity Questionnaire, you'll need to provide the following information about the VPN connection in addition to the details filled out for the DX:

- Customer VPN Gateway details such as the make and model of your customer gateway device(s)
- Customer VPN Gateway Internet facing public IP Address
- Type of Routing (static / dynamic)
- BGP ASN for Dynamic Routing (Customer gateway ASN for BGP. Only 16 bit ASN is supported.)
- ASN for the AWS side of the BGP session (16- or 32-bit ASN)
- Customer Side BGP Peer IP-address (if different from VPN peer IP provided)
- Second Public IP Address (OPTIONAL: only if active-active mode is used)
- Customer On-Premises Network IP ranges Submit the completed questionnaire to SAP. They will then create the VPN connection and provide you with the configuration details.

SAP's Responsibilities

After you submit the completed questionnaire, SAP will handle the following tasks (the list below is illustrative only for this context): * Create the customer gateway (with your provided information like BGP ASN, IP address, and optional private certificate) * Create the AWS Site-to-Site VPN and attach it to the RISE with SAP Transit Gateway and your customer gateway * Provide the VPN configuration file for you to set up on your on-premises router * If you need SAP to provision Transit Gateway in RISE VPC, SAP will add the necessary route to the Transit Gateway route table and update the security groups

Using the information received from SAP, configure the VPN tunnels on your on-premises router. Implement routing policies to prefer the Direct Connect connection over the VPN as the primary path.

Refer to the AWS documentation on router configuration for Direct Connect for guidance on the necessary settings.

Test and Verify Connections

Coordinate with SAP to enable the VPN connection and verify end-to-end connectivity. Test failover scenarios by simulating a Direct Connect failure and ensure traffic properly fails over to the VPN.

Confirm with SAP that the failover is working as expected for both the Direct Connect and VPN paths.

Maintain the Connections

Regularly review and update the configurations for both the Direct Connect and VPN connections. Coordinate any changes with SAP.

Monitor the performance and availability of both connections, and refer to the AWS documentation on monitoring Direct Connect and VPN for best practices.

By implementing this Direct Connect with Site-to-Site VPN failover solution, you can achieve a highly resilient connectivity setup for your RISE with SAP deployment on AWS, ensuring seamless failover and reliable network performance.

Connecting to RISE from your AWS account

You can connect to RISE from your AWS account in the following ways.

Topics

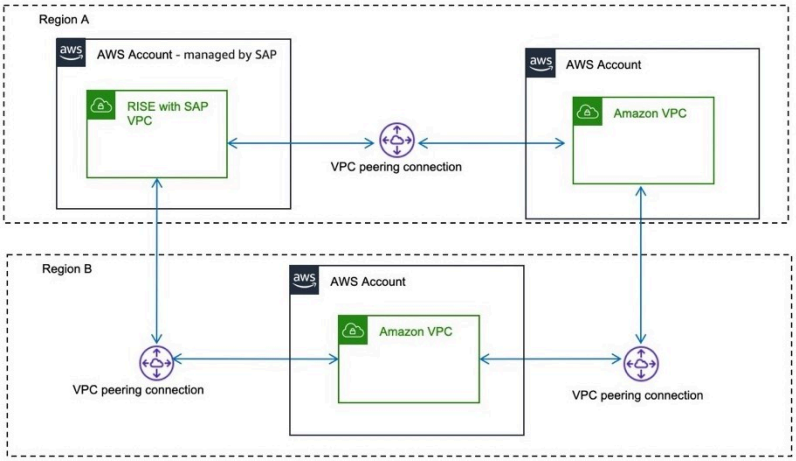
Amazon VPC peering

VPC peering enables network connection between two AWS VPCs using private IPv4 and IPv6 addresses. Instances can communicate over the same network. For more information, see What is VPC peering?

Before setting up a VPC peering connection, you need to create a request for SAP's approval. For a successful VPC peering, the defined IPv4 Classless Inter-Domain Routing (CIDR) block must not overlap. Check with SAP for the CIDR ranges that can be used in RISE with SAP VPC.

VPC peering is one-on-one connection between VPCs, and is not transitive. Traffic cannot transit from one VPC to another via an intermediary VPC. You must setup multiple peering connections to establish direct communication between RISE with SAP VPC and multiple VPCs.

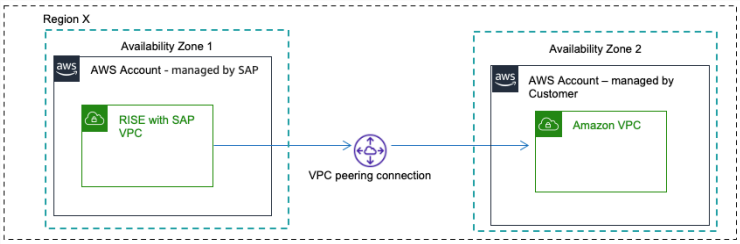
VPC peering works across AWS Regions. All inter-Region traffic is encrypted with no single point of failure or bandwidth bottleneck. Traffic stays on AWS Global Network and never traverses the public internet, reducing threats of common exploits and DDoS attacks.



Data transfer for VPC peering within an Availability Zone is free, and for across Availability Zones is charged per-GB for "data in" to and "data out". Data transfer for VPC peering for across regions

is charged for "out" per-GB. For more information, see Amazon EC2 pricing. In your AWS account, use the Availability Zone ID of AWS account managed by SAP to avoid cross-Availability Zone data transfer charges. You can ask for the Availability Zone ID from SAP. For more information, see Availability Zone IDs for your AWS resources.

Pricing example - VPC peering across Availability Zones



100GB of data sent from the AWS account – managed by SAP via VPC Peering toward the AWS account – managed by Customer across AZs:

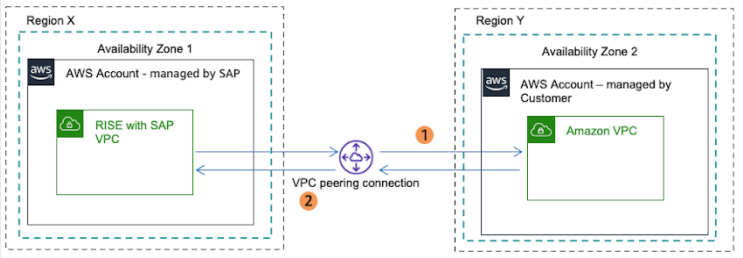
100GB * $0.01per-GB = $1 (out - billed to AWS account – managed by SAP) and 100GB * $0.01per-GB = $1 (IN - billed to AWS account – managed by Customer)

As the cost for data transfer is included In the RISE subscription, the AWS account – managed by Customer will only incur the cost for traffic IN e.g. $0.01 per-GB.

[note: the cost example also applies when Sender is AWS account – managed by Customer and Receiver is AWS account – managed by SAP]

Pricing example - VPC peering across Regions

[note: cost between AWS Regions vary. For more information see: Amazon EC2 pricing Data Transfer]



1). 100GB of data sent from the AWS account – managed by SAP via VPC Peering toward the AWS account – managed by Customer across Regions.

100GB * ($0.01-$0.138per-GB) = $1-$13.8 (out - billed to AWS account – managed by SAP)

As the cost for data transfer is included In the RISE subscription the AWS account – managed by Customer will not incur cost for this example.

2). 100GB of data sent from the AWS account – managed by Customer via VPC Peering toward the AWS account – managed by SAP across Regions.

100GB * ($0.01-$0.138per-GB) = $1-$13.8 (out - billed to AWS account – managed by Customer)

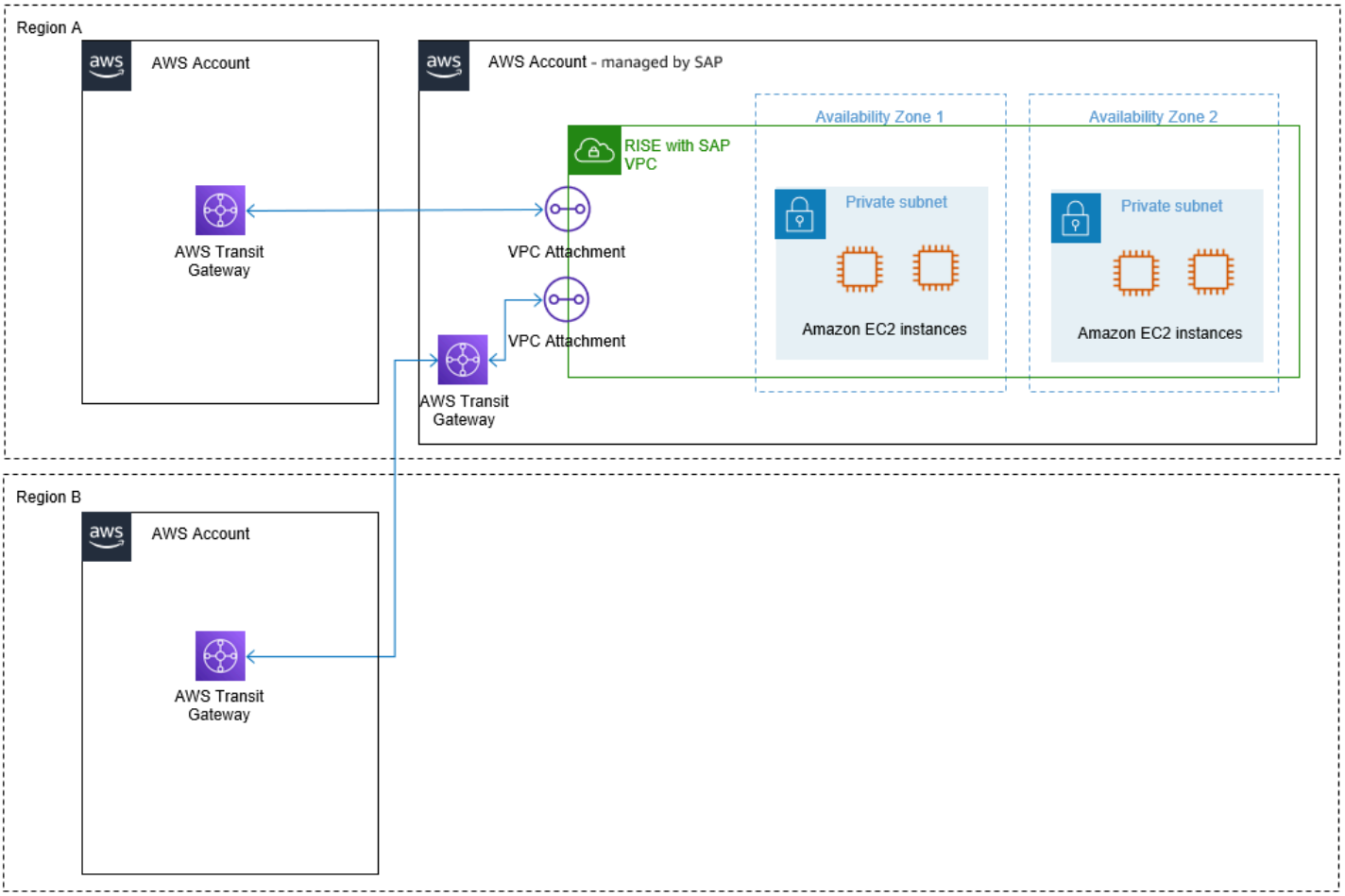
As the cost for data transfer is calculated for "data out" the AWS account – managed by Customer will incur the cost for this example.

AWS Transit Gateway

AWS Transit Gateway is a network transit hub to interconnect Amazon VPCs. It acts as a cloud router, resolving complex peering setup issues by acting as the central communication hub. You need to establish this connection with AWS account managed by SAP only once.

Transit Gateway in your own AWS account

To establish connection with AWS account managed by SAP, create and share AWS Transit Gateway via AWS Resource Access Manager (RAM) in your AWS account. SAP then creates an attachment to enable traffic flow through an entry in route table. As AWS Transit Gateway resides in your AWS account, you can retain control over traffic routing. For more information, see Transit gateway peering attachments.



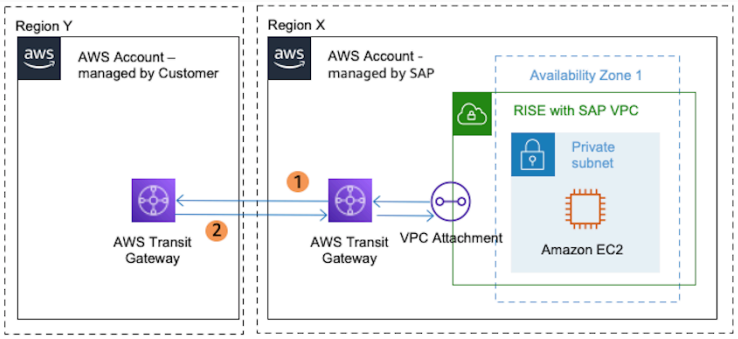
Transit Gateway in AWS account managed by SAP

When you already have an Transit Gateway in another AWS Region, and cannot create another AWS account with Transit Gateway in the Region that has RISE with SAP account, then SAP can provide the Transit Gateway in the RISE with SAP account that will be managed by SAP. You can enable communication between your Transit Gateway and SAP managed Transit Gateway through Transit Gateway Peering. You cannot connect VPC attachments of VPCs outside of the RISE environment to the SAP-managed Transit Gateway.

For peering attachments, each Transit Gateway owner is billed hourly for the peering attachment with the other Transit Gateway, thus the hourly cost for the peering attachment of the Transit Gateway in the SAP account - managed by SAP (for the purpose of Inter Region Transit Gateway Peering) is part of the RISE subscription. However the hourly cost for the peering attachment of the Transit Gateway in the Customer account – Customer managed is billed to the Customer. For more information, see: Transit Gateway pricing

Pricing example - Transit Gateway across VPCs in different Regions

[note: cost between AWS Regions vary. For more information see: Amazon EC2 pricing Data Transfer]



1). 100GB of data sent from a VPC in Region X in the AWS account – managed by SAP via the Transit Gateway that resided in the AWS account – managed by SAP, towards a peered Transit Gateway, in a different Region Y, that resided in the AWS account – managed by Customer ending at a VPC in the AWS account – managed by Customer:

100GB * $0.02per-GB = $2 (Transit Gateway data processing) + 100GB * ($0.01-$0.138per-GB) = $1-$13.8 (Region out) = $3-$15.8 (Total - billed to AWS account – managed by SAP)

Data processing is charged to the VPC owner who sends the traffic to Transit Gateway. As the sending VPC is residing in the AWS account – managed by SAP and the cost for data transfer is included in the RISE Subscription, thus the AWS account – managed by Customer will not incur data transfer cost for this example. As data processing charges do not apply for data sent from a peering attachment to a Transit Gateway and inbound inter-Region data transfer charges are free, no further Data Transfer charges apply to the AWS account – managed by Customer. The AWS account – managed by Customer will only be billed for the price per Transit Gateway peering attachment per hour. Data out of an AZ will always go via Transit Gateway endpoint in that AZ to reach other VPC, so there is no cross AZ Data Transfer costs.

2). 100GB of data sent from a VPC in region Y in the AWS account – managed by Customer via the Transit Gateway that resided in the AWS account – managed by Customer, towards a peered Transit Gateway, in a different region X, that resided in the AWS account – managed by SAP ending at a VPC in the AWS account – managed by SAP:

100GB * $0.02per-GB = $2 (Transit Gateway data processing) + 100GB * ($0.01-$0.138per-GB) = $1-$13.8 (Region out) = $3-$15.8 (Total - billed to AWS account – managed by Customer)

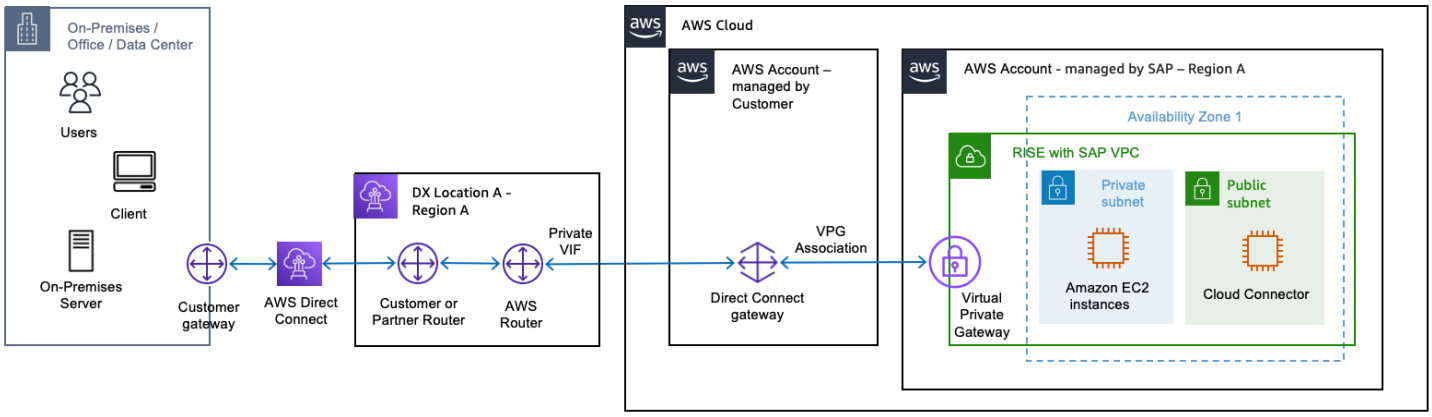
Data processing is charged to the VPC owner who sends the traffic to Transit Gateway. As the sending VPC is residing in the AWS account – managed by Customer all data transfer cost for this example are billed to the AWS account – managed by Customer. In addition, the AWS account – managed by Customer will be billed for the price per Transit Gateway peering attachment per hour.

AWS Direct Connect gateway

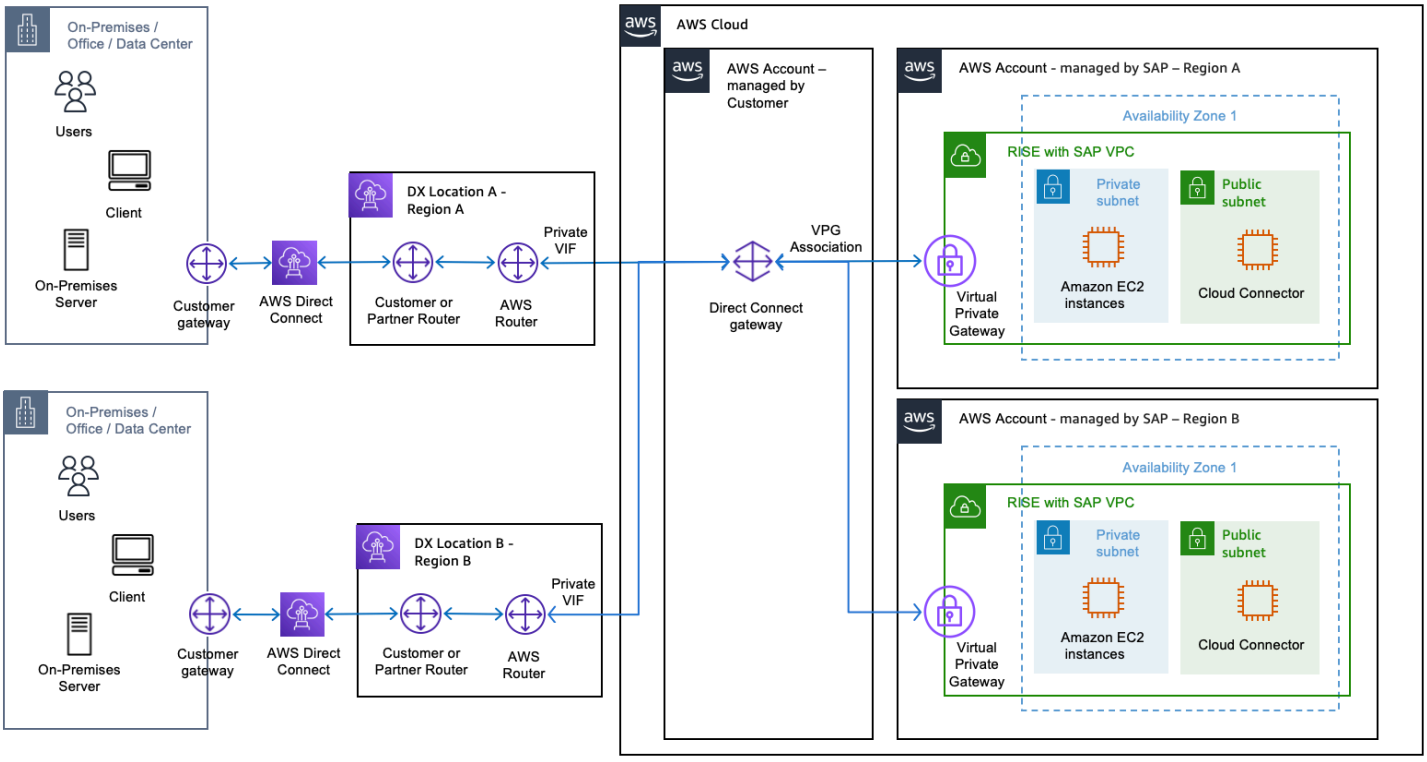
AWS Direct Connect gateway is a global service that enables you to establish private connectivity between your on-premises networks and multiple Amazon VPCs across different AWS regions. This centralized connection hub allows you to consolidate your network architecture, reduce complexity, and maintain secure, high-bandwidth connections while avoiding public internet for your mission-critical workloads.

AWS Direct Connect gateway in your own AWS account

To establish connection with AWS account managed by SAP, create AWS Direct Connect gateway that routes traffic from Private VIF to VPC Private Gateway. As AWS Direct Connect gateway resides in your AWS account, you can retain control over traffic routing.

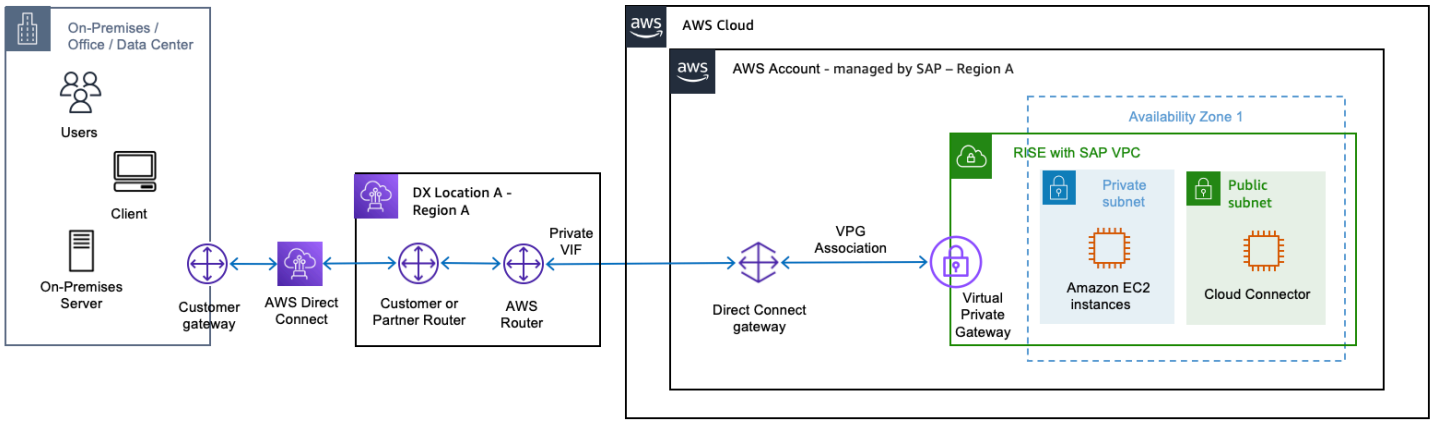


When you have a requirement for connectivity from multiple on-premise sites and/or are using multiple AWS regions for RISE with SAP (i.e. for long range DR), you can simplify the connectivity utilizing Direct Connect Gateway



AWS Direct Connect gateway in AWS account managed by SAP

If you do not have any requirement to own and manage an AWS account, you can request for SAP to provide the AWS Direct Connect gateway that is part of AWS Account which is managed by SAP.



There is no additional charges for AWS Direct Connect gateway itself. You can find out more from the AWS Direct Connect FAQs.

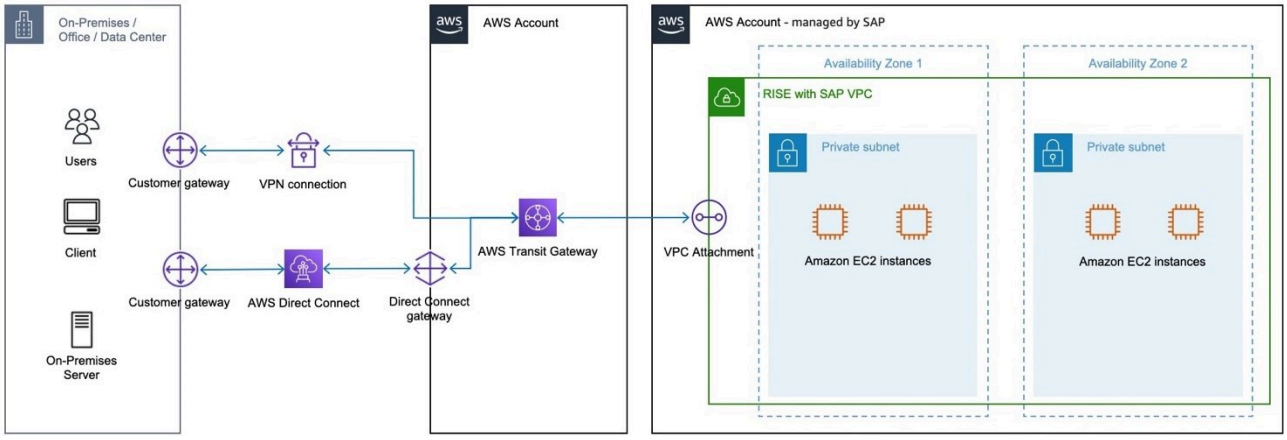
Connecting to RISE using your single AWS account

You can establish connectivity between on-premises and RISE with SAP VPC using your AWS account. This method provides you with more control but also requires managing AWS services in your AWS account. You can use any one of the following options.

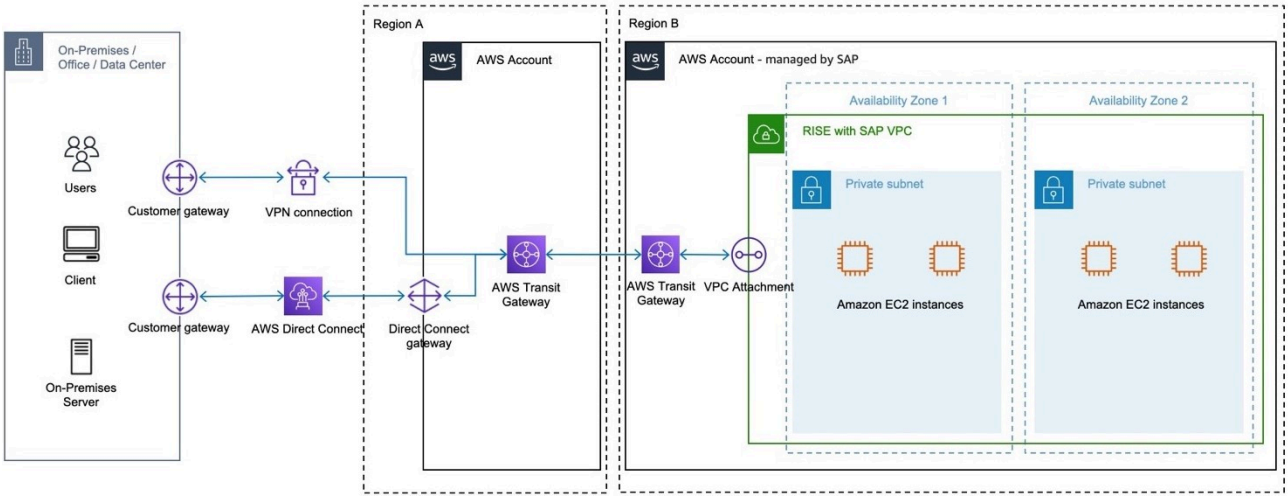
- AWS Transit Gateway – Share AWS Transit Gateway resource in you AWS account with AWS account managed by SAP.
- AWS VPN with AWS Transit Gateway – Create an IPsec VPN connection between your remote network and transit gateway over the internet. For more information, see How AWS Site-to-Site VPN works and Transit gateway VPN attachments.
- Direct Connect gateway – Create a Direct Connect gateway with a transit virtual interface. For more information, see Transit gateway attachments to a Direct Connect gateway.

  To strengthen the security, see How do I establish an AWS VPN over an AWS Direct Connect connection?

The following image shows this option within the same AWS Region.s



The following image shows this option across different AWS Regions.



When you choose AWS Site-to-Site VPN and/or AWS Direct Connect to establish connectivity between on-premises and RISE with SAP VPC using a Transit Gateway in the AWS account - managed by the Customer, either in the same AWS Region or a different AWS Region than the RISE with SAP VPC, the following applies.

Hourly cost:

As the AWS Site-to-Site VPN is residing in the AWS account – managed by Customer and is attached to the Transit Gateway that resides in the AWS account – managed by Customer, the cost for the VPN connection and the cost for the Transit Gateway attachment are billed to the AWS account – managed by Customer

As the Direct Connect and Direct Connect Gateway is residing in the AWS account – managed by Customer and is attached to the Transit Gateway that resides in the AWS account – managed by Customer the cost for the AWS Direct Connect ports hours and the cost for the Transit Gateway attachment are billed to the AWS account – managed by Customer.

For peering attachments, each Transit Gateway owner is billed hourly for the peering attachment with the other Transit Gateway.

Data processing charges:

Data processing charges apply for each gigabyte sent from a VPC, Direct Connect or VPN to/via the Transit Gateway.

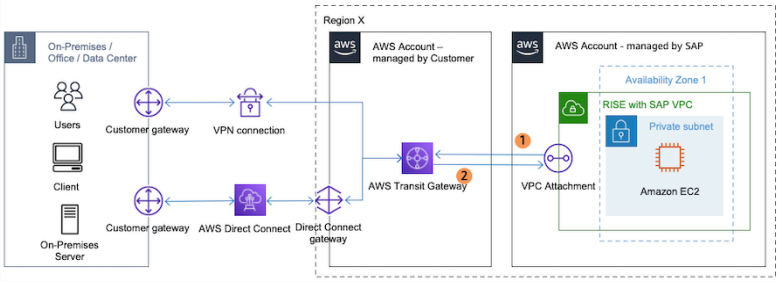
Depending on the source and destination the data processing charges vary and will be billed to the AWS account – managed by Customer, or are already included in the RISE subscription (For a cost estimation example: see below)

For more information see:

- AWS Site-to-Site VPN Pricing
- AWS Direct Connect Pricing
- Transit Gateway pricing

Pricing example – Transit Gateway in VPCs in the same region via VPN or Direct Connect

[note: cost between AWS Regions vary. For more information see: Amazon EC2 pricing Data Transfer]



1). 200GB of data sent from a VPC in the AWS account – managed by SAP via the Transit Gateway that resided in the AWS account – managed by Customer via a VPN or Direct Connect in the AWS account – managed by SAP towards On-Premises:

200GB * $0.02per-GB = $4 (Transit Gateway data processing) + 100 GB * $0.09per-GB = $9 (VPN data transfer out, with the first 100 GB are free, then $ 0.09 per-GB) = $13 (Total data transfer out billed to AWS account – managed by SAP)

or

200GB * $0.02per-GB = $4 (Transit Gateway data processing) + 200GB * ($0.02-$0.19per-GB) = $4-$38 (Direct Connect data transfer out) = $8-$42 (Total data transfer out billed to AWS account – managed by SAP)

Data processing is charged to the VPC owner who sends the traffic to Transit Gateway. As the sending VPC is residing in the AWS account – managed by SAP and the cost for data transfer is included in the RISE Subscription, therefore the AWS account – managed by Customer will not incur Data Transfer cost in this example.

2). 200GB of data sent from On-Premises via a VPN or Direct Connect in the AWS account – managed by Customer via the Transit Gateway that resided in the AWS account – managed by Customer towards VPC in the AWS account – managed by SAP:

200GB * $0.00per-GB = $0 (VPN data transfer in) + 200GB * $0.02per-GB = $4 (Transit Gateway data processing) + $0 (VPN data transfer in) = $4 (Total data transfer in billed to AWS account – managed by Customer)

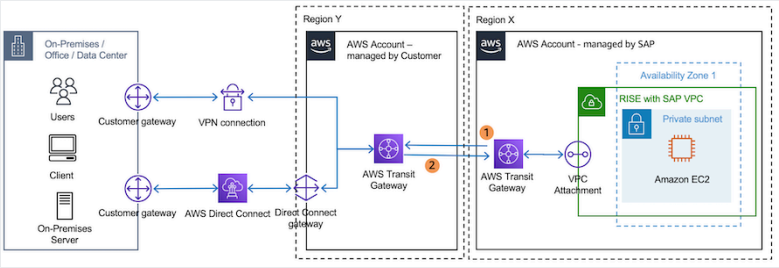
or

200GB * $0.00per-GB = $0 (Direct Connect data transfer in) + 200GB * $0.02per-GB = $4 (Transit Gateway data processing) = $4 (Total data transfer in billed to AWS account – managed by Customer)

Data transfer into AWS is free and this also applies to VPN and Direct Connect therefore the only data processing charge is the data processing of the Transit Gateway. As Transit Gateway resides in the AWS account – managed by Customer the cost for data transfer is billed to the AWS account – managed by Customer

Pricing example – Transit Gateway in VPCs in the different regions via VPN or Direct Connect

[note: cost between AWS Regions vary. For more information see: Amazon EC2 pricing Data Transfer]



1). 200GB of data sent from a VPC in the AWS account – managed by SAP via the Transit Gateway that resided in the AWS account – managed by SAP that is peered with an Transit Gateway in a different Region in the AWS account – managed by Customer via a VPN OR Direct Connect in the AWS account – managed by Customer towards On-Premises:

200GB * $0.02per-GB = $4 (Transit Gateway data processing) + 200GB * ($0.01-$0.138per-GB) = $2-$27.6 (Region out) + 100GB * $0.09per-GB = $9 (VPN data transfer out, with the first 100 GB are free, then $ 0.09 per-GB) = $15-$40.6 (Total data transfer out billed to AWS account – managed by SAP)

or

200GB * $0.02per-GB = $4 (Transit Gateway data processing) + 200GB * ($0.01-$0.138per-GB) = $2-$27.6 (Region out) + 200GB * ($0.02-$0.19per-GB) = $4-$38 (Direct Connect data transfer out) = $10-$69.6 (Total data transfer out billed to AWS account – managed by SAP)

Data processing is charged to the VPC owner who sends the traffic to Transit Gateway. As the sending VPC is residing in the AWS account – managed by SAP and the cost for Data Transfer is included in the RISE subscription, therefore the AWS account – managed by Customer will not incur Data Transfer cost in this example.

2). 200GB of data sent from On-Premises via a VPN or Direct Connect in the AWS account – managed by Customer via the Transit Gateway that resided in the AWS account – managed by Customer via a peered Transit Gateway in a different region in the AWS account – managed by SAP towards a VPC in the AWS account – managed by SAP:

200GB * $0.02per-GB = $4 (Transit Gateway data processing) + 200GB * $0.00per-GB = $0 (VPN data transfer in) + 200GB * ($0.01-$0.138per-GB) = $2-$27.6 (Region out) = $6-$31.6 (Total data transfer in billed to AWS account – managed by Customer)

or

200GB * $0.02per-GB = $4 (Transit Gateway data processing) + 200GB * $0.00per-GB = $0 (Direct Connect data transfer in) + 200GB * ($0.01-$0.138per-GB) = $2-$27.6 (Region out) = $6-$31.6 (Total data transfer in billed to AWS account – managed by Customer)

Data transfer into AWS in is free and this also applies to VPN and Direct Connect therefore the data processing charge is the data processing of the Transit Gateway and the inter-region data transfer charges. As Transit Gateway resides in the AWS account – managed by Customer, the cost for data transfer is billed to the AWS account – managed by Customer.

Connecting to RISE with a shared AWS Landing Zone

Modern SAP landscapes have several connectivity requirements. Services are accessed across on-premises and AWS Cloud as well as across a variety of SaaS solutions and other cloud service providers.

Creating an AWS Landing Zone facilitates secure and scalable connectivity for RISE with SAP. It provides the following benefits:

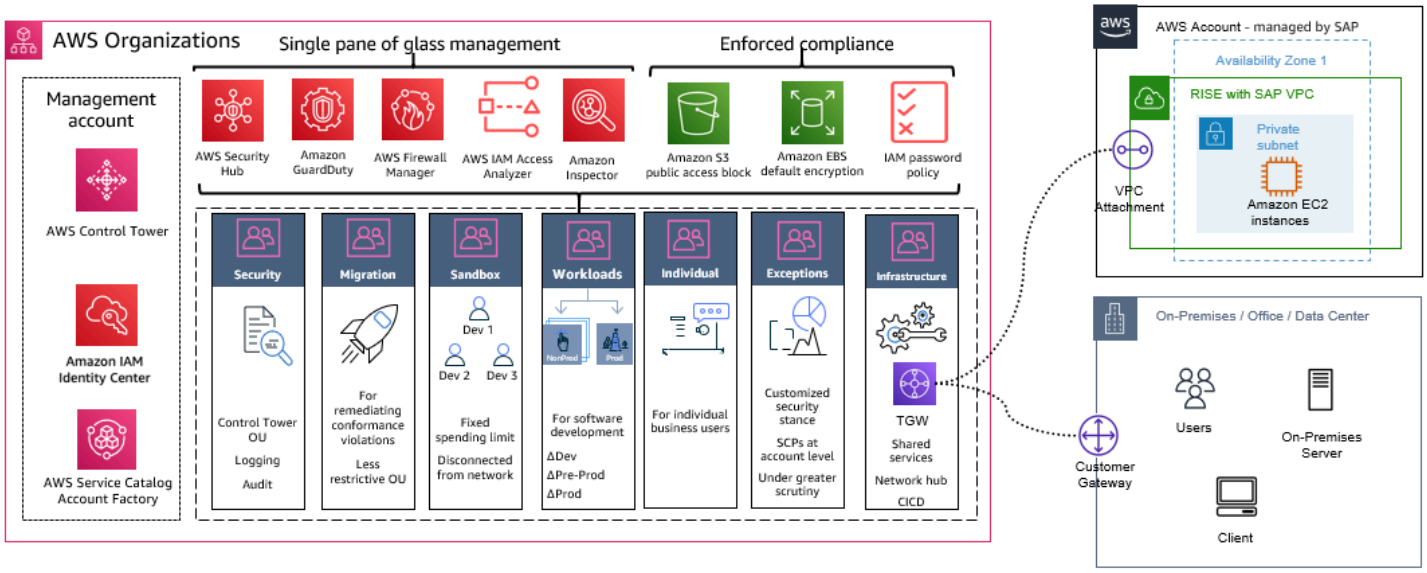
- Control over networking configuration
- Ability to reuse AWS Direct Connect connections across your broader AWS solutions
- Reduced network hops and latency for connectivity to other SaaS solutions and cloud service providers as they are not routed via on-premises
- Ability for additional governance and control through use of AWS services

A Landing Zone is designed to help organizations achieve their cloud initiatives by automating the set-up of an AWS environment that follows AWS Well Architected framework. It provides scalability to cater to all scenarios, from the simplest connectivity, where only RISE with SAP connectivity to on-premises environments is required, to complex requirements with connectivity to multiple SaaS solutions, multiple CSPs and on-premises connectivity.

The key components and benefits of a Landing Zone include:

- **Multi-account structure** – it sets a baseline environment across multiple AWS accounts using an organization unit (OU) structure for different workloads. For instance, production, development, shared services, etc.
- **AWS Identity and Access Management** – it configures AWS Identity and Access Management (IAM) roles and policies for secure access and management of permissions.
- **Networking** – it sets up a Amazon Virtual Private Cloud (Amazon VPC) with subnets, routing tables, and security groups, following the best practices for network isolation and security.
- **Logging and monitoring** – it configures AWS services, such as AWS Config, AWS CloudTrail, Amazon GuardDuty for centralized logging, monitoring, and auditing of resource changes and security events.
- **Security** – it implements AWS security best practices, such as like enabling AWS Config Rules, setting up AWS CloudTrail trails, and creating AWS Security Hub standards.
- **Automation** – it uses AWS CloudFormation templates and AWS Service Catalog to automate the deployment and management of the Landing Zone environment.
- **Customization** – it allows for customization and extension based on specific organizational requirements, such as adding additional AWS services or integrating with existing on-premises infrastructure.

We recommend using an AWS Landing Zone for RISE with SAP connectivity.



Building an AWS Landing Zone

You can implement AWS Landing Zones using AWS Control Tower. It provides an automated process for building the Landing Zone, including management and governance services.

In a simple scenario, a Landing Zone contains a minimal footprint focused on connectivity that is typically centred around AWS Transit Gateway. For more information, see Landing zone.

The following is a general overview of the process:

1. **Define requirement** – understand your organization's security, compliance, and operational requirements. This will help determine the appropriate guardrails, controls, and services to be included in the Landing Zone.
2. **Design architecture** – plan the overall architecture, including the number of accounts (management, shared services, workload accounts), network design (VPCs, subnets, routing), shared services (logging, monitoring, identity management), and security controls (IAM, service control policies, guardrails).
3. **Setup AWS Control Tower** – AWS Control Tower helps in setting up and governing a multi-account AWS environment based on best practices. It allows you to create and provision new AWS accounts and deploy baseline security configurations across those accounts.
4. **Configure AWS Organizations** – Organizations enables you to centrally manage and govern your AWS accounts. Configure Organizations in AWS Control Tower by creating the necessary organizational units (OUs) and service control policies (SCPs).
5. **Deploy core accounts and services** – create and configure the core accounts, such as the management account, shared services accounts (for logging, security tooling), and any other required shared accounts. Deploy shared services, such as CloudTrail, Config, and Security Hub.
6. **Deploy network architecture** – set up the network architecture, including VPCs, subnets, route tables, and any necessary network appliances or services (for example, Transit Gateway for a hub-and-spoke model).
7. **Configure IAM** – establish IAM roles, policies, and groups for controlling access and permissions across the Landing Zone accounts.
8. **Implement security controls** – deploy security services and guardrails, such as Security Hub, Firewall Manager, AWS WAF, and AWS Config Rules, to enforce security best practices and compliance requirements.
9. **Configure logging and monitoring** – set up centralized logging and monitoring solutions, such as CloudWatch, CloudTrail, and Config, to capture and analyze logs and events across the Landing Zone accounts.

10 **Deploy workload accounts** – deploy workload accounts with your Landing Zone. You can create an AWS account to connect to RISE with SAP VPC. We recommended connecting using Transit Gateway for flexibility and ease of management.

11 **Automate and maintain** – use AWS CloudFormation templates or other Infrastructure as Code tools to automate the deployment and maintenance of the Landing Zone resources. Establish processes for ongoing maintenance, updates, and compliance checks.

AWS Professional Services or AWS Partners provide assistance for building and maintaining a landing zone for RISE with SAP.

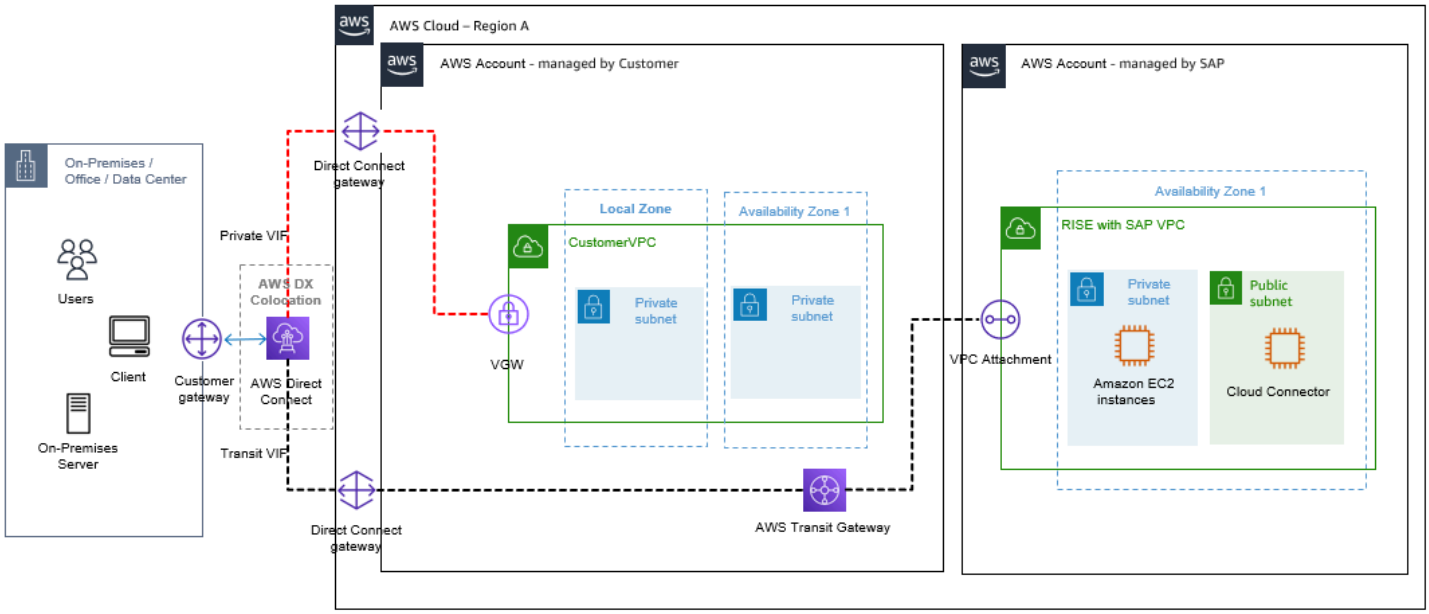
Cost associated to a Customer Managed AWS Landing Zone vary depending on the AWS Services that are used. The AWS Services as described in this paragraph have their own pricing model. For more information on price, see the dedicated pricing pages of the listed AWS Services.

Connect to RISE through nearest AWS Direct Connect POP (including AWS Local Zone)

AWS Direct Connect point-of-presence (POP) is a physical cross-connect that allows users to establish a network connection from their premises to an AWS Region or AWS Local Zone. You can use the nearest Direct Connect POP (for example in AWS Local Zone) to benefit from lower setup and running cost, with the same or lower network latency to your RISE with SAP VPC that runs on parent AWS Region. For more information, see AWS Direct Connect Traffic Flow with AWS Local Zone.

Here is an example scenario - You are based in Philippines, and you would like to deploy RISE with SAP in AWS Singapore Region. You can use Direct Connect POP in Manila to setup Direct Connect from your on-premise data centre or offices. This strategy provides a lower network latency compared to a connecting directly to the AWS Region in Singapore.

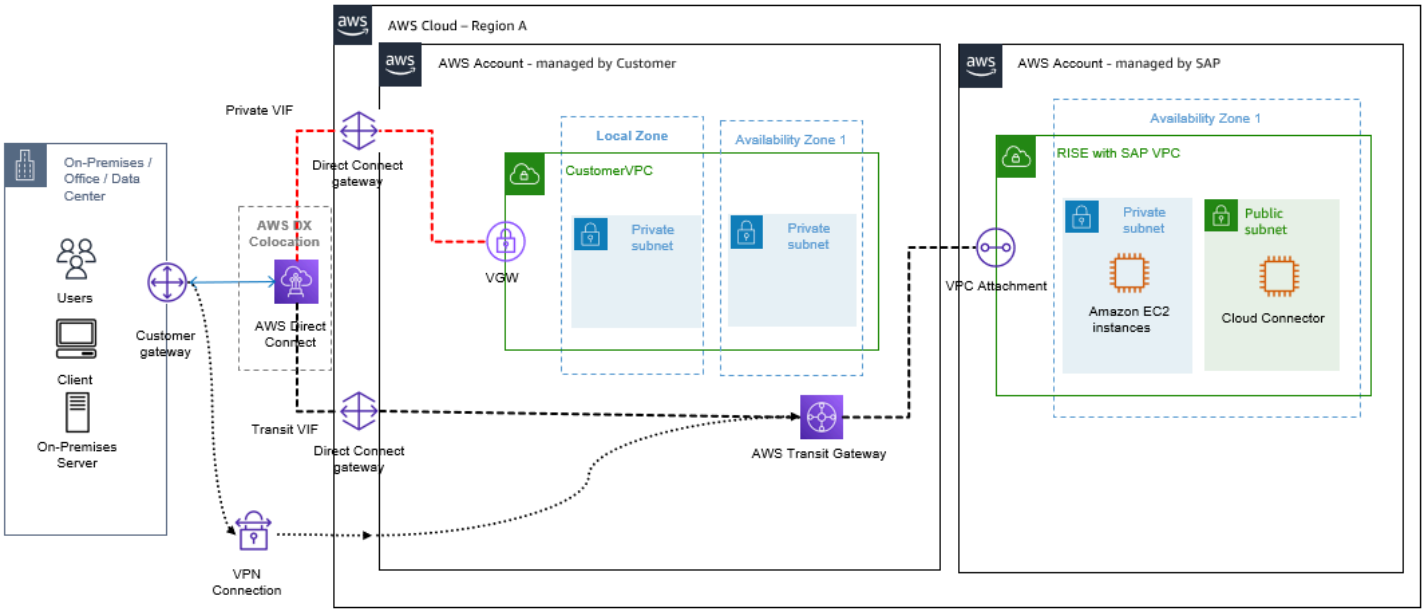
The following diagram displays RISE connectivity through nearest AWS Direct Connect POP.



The following are some considerations when using AWS Direct Connect POP:

- Use separate VPCs for Region (RISE with SAP VPC) and Local Zones based non-SAP workloads
- Use Direct Connect Gateway in AWS Direct Connect POP and Private VIF connectivity
- Use Direct Connect Gateway in AWS Direct Connect POP and Transit VIF connectivity for Region VPCs (RISE with SAP VPC).This is done because Direct Connect Gateway does not exists in AWS Direct Connect POP, and AWS Transit Gateway exists only in AWS Regions.

If resilience is critical, setup a secondary Direct Connect to the AWS Region running RISE with SAP VPC or use AWS Site-to-Site VPN to the AWS Region connectivity option. These services operate within the parent AWS Region, serving as a failover connectivity option ensuring uninterrupted connectivity in the event of disruptions or failures.



Cost of data transferred between a Local Zone and an Availability Zone within the same AWS Region, "in" to and "out" from Amazon EC2 in the Local Zone varies. For more information see: EC2 - On-Demand Pricing - Data Transfer within the same AWS Region

Decision tree on connectivity to RISE

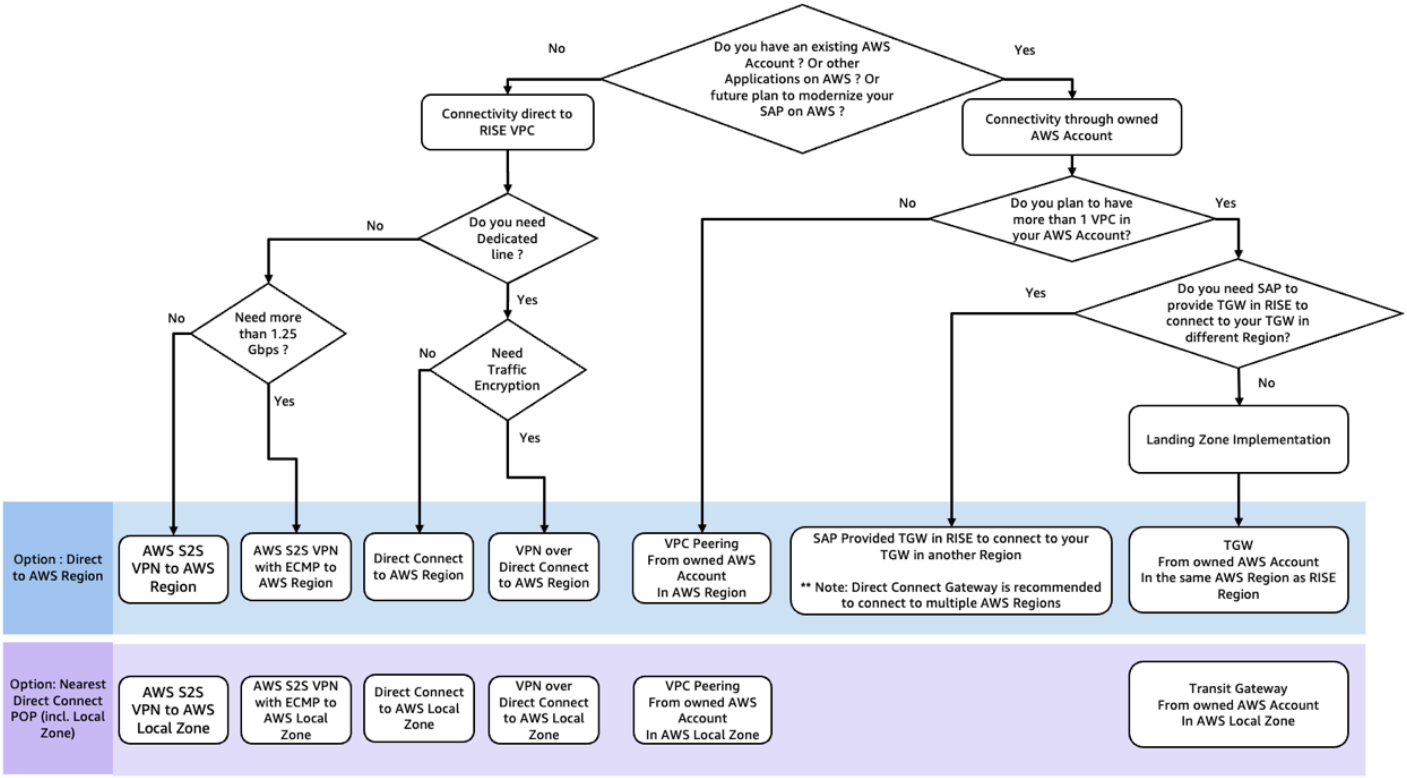
You must establish required connectivity to proceed with RISE with SAP on AWS. The following are a few connectivity patterns described in the preceding sections:

- direct to RISE VPC, supported with Site-to-Site VPN
- direct to RISE VPC, supported with Direct Connect
- connectivity through your AWS account via VPC Peering
- connectivity through Transit Gateway, supporting multi-account deployments
- connectivity through SAP-managed Transit Gateway supporting multi-account deployments

You must also consider if you want to connect:

- directly to an AWS Region where the RISE with SAP VPC is going to be deployed
- or through AWS Local Zone (nearest AWS Direct Connect POP) to benefit from lower setup and running costs, with the same or lower network latency to connect to your RISE with SAP VPC

The decision tree displayed in the following diagram helps you decide which connectivity is suitable based on your requirements, such as future plan of additional AWS or RISE accounts, dedicated line (security, performance), and bandwidth needs.



Note:

1. ECMP requires Transit Gateway for S2S VPN.
2. Direct Connect Gateway is recommended to connect to multiple AWS regions. This simplifies the connectivity setup and avoids TGW peering between AWS regions.

Other considerations

This sections provides information about other considerations when connecting to RISE.

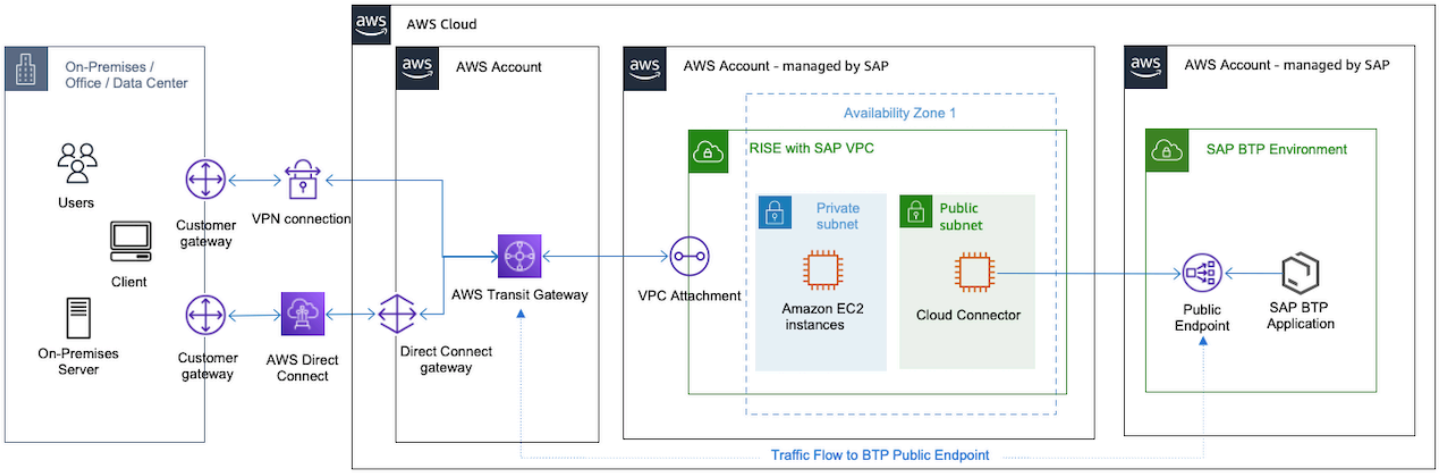
Topics

- SAP Business Technology Platform (BTP) with RISE on AWS
- Connecting to cloud solutions or SaaS from RISE
- Connectivity patterns for multi-cloud to RISE
- How to implement chargeback capability for connectivity to RISE

- Integrating Domain Name System to RISE and Route 53 consideration

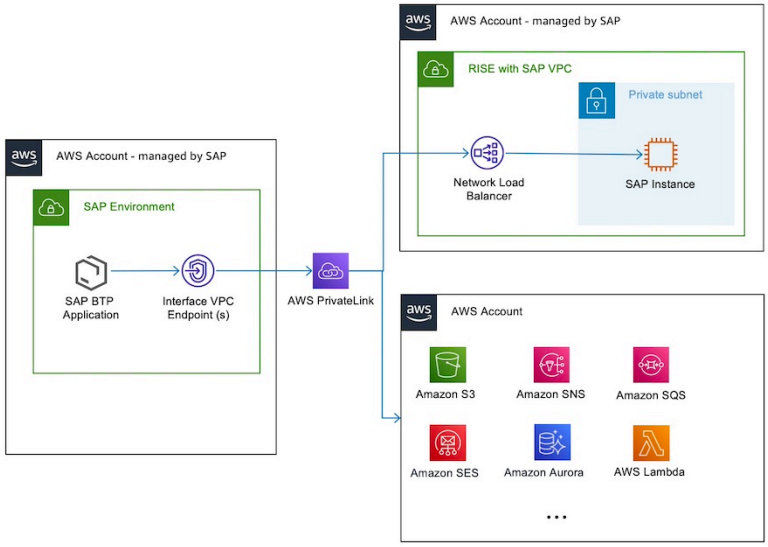
SAP Business Technology Platform (BTP) with RISE on AWS

You can use SAP Business Technology Platform BTP services on AWS to extend the functionality of the RISE with SAP. SAP recommends SAP Cloud Connector to connect RISE with SAP VPC with SAP BTP via internet. When both RISE with SAP and SAP BTP run on AWS (in the same AWS region or different AWS regions), the network traffic is encrypted and contained within AWS Global Network, without going through the internet (see the following diagram). This provides better security and performance for any integration use-cases between RISE with SAP and SAP BTP. For more information, see Amazon VPC FAQs - Does traffic go over the internet when two instances communicate using public IP addresses or when instances communicate with a public AWS service endpoint ?.



As displayed in the preceding diagram, you can configure Transit Gateway to handle both RISE and BTP network traffic. For more information, see How to route internet traffic from on-premise via Amazon VPC?

SAP also offers SAP Private Link Service for SAP BTP on AWS. SAP Private Link connects SAP BTP on AWS with a secure connection without using public IPs in your AWS account.

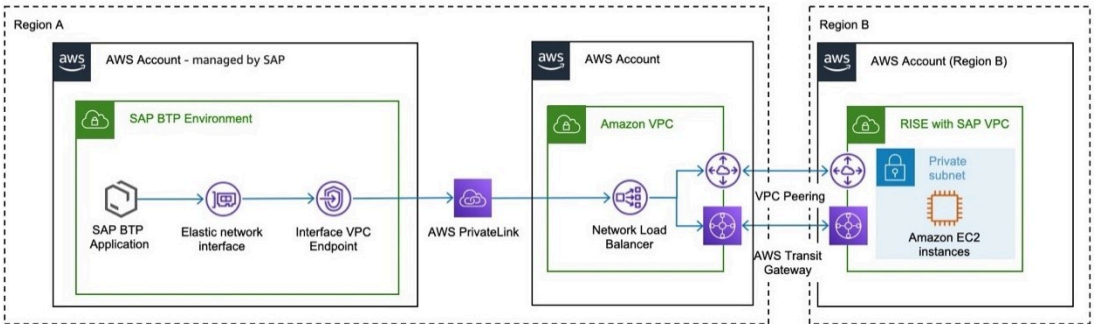


You can connect to an AWS endpoint service from an SAP BTP application running on Cloud Foundry. By establishing this connection, you can directly connect to AWS services, or for example, to an S/4HANA system. For a complete list of supported AWS services, see Consume Amazon Web Services in SAP BTP.

You can establish a secure and private communication between SAP BTP and AWS services with SAP Private Link Service. By using private IP address ranges (RFC 1918), you reduce the attack surface of the application. The connection does not require an internet gateway. If you do not require this extra layer of security, you can still connect via the public APIs of SAP BTP without SAP Private Link, and benefit from AWS global network. For more information, see Amazon VPC FAQs.

SAP Private Link for AWS currently supports connections initiated from SAP BTP Cloud Foundry to AWS.

For AWS services across AWS Regions, you can create a VPC in the same AWS Region as your SAP BTP Cloud Foundry Runtime, and connect these VPCs via VPC peering or AWS Transit Gateway. For a list of supported Regions, see Regions and API Endpoints Available for the Cloud Foundry Environment.



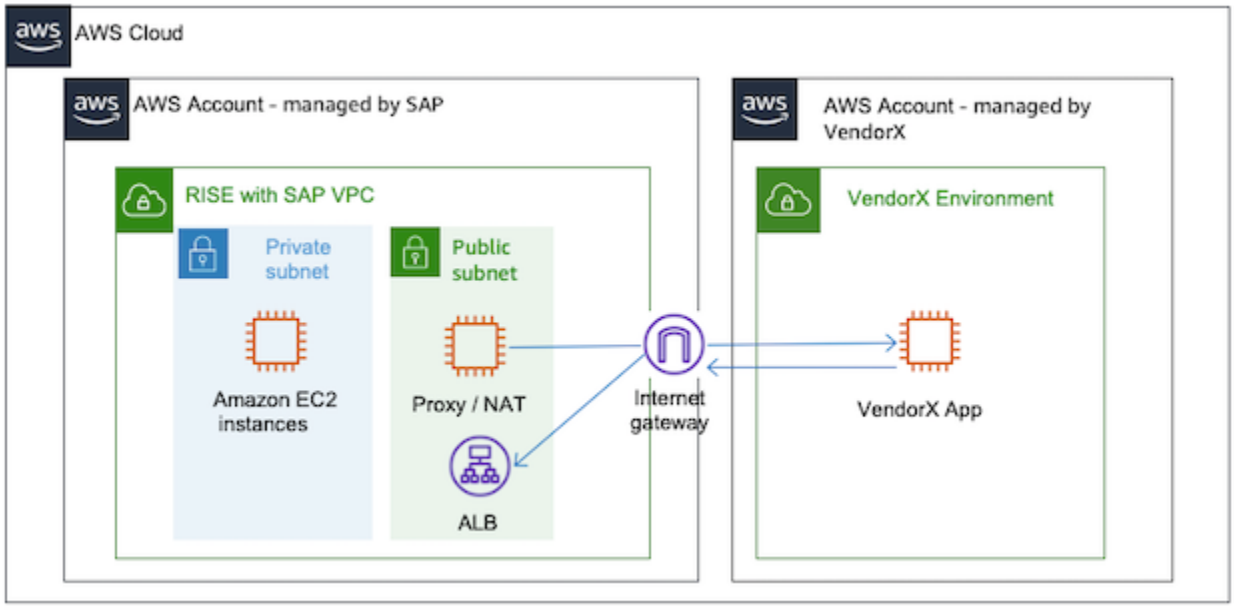
SAP Private Link Service is a paid service offered by SAP on SAP BTP. For more information see: SAP Discovery Center – Services – SAP Private Link Service.

Cost associated to AWS Services in the AWS account - managed by the Customer to facilitate cross region connectivity for example the AWS Network Load Balancer, or Transit Gateway vary. For more information on price, see the dedicated pricing pages of the listed AWS Services.

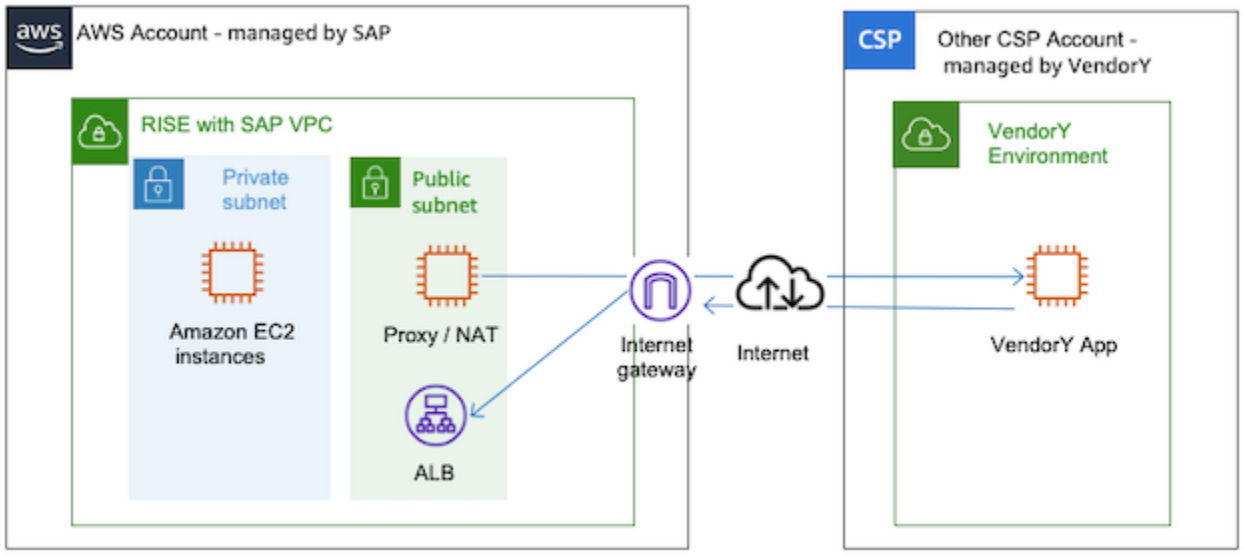
Connecting to cloud solutions or SaaS from RISE

When modernizing the SAP landscape, you may subscribe to several SAP cloud solutions or SaaS from independent software vendors to complement RISE with SAP solution.

When the cloud solutions are running on AWS (in the same AWS region or different AWS regions), the connectivity from RISE with SAP is kept within the AWS global network without requiring internet connectivity. The connectivity is retained through the provided squid proxy server within RISE with SAP VPC.For more information, see Amazon VPC FAQs - Does traffic go over the internet when two instances communicate using public IP addresses or when instances communicate with a public AWS service endpoint ?.



If your cloud is running on other data centre or with another cloud service provider, then you need internet connectivity.



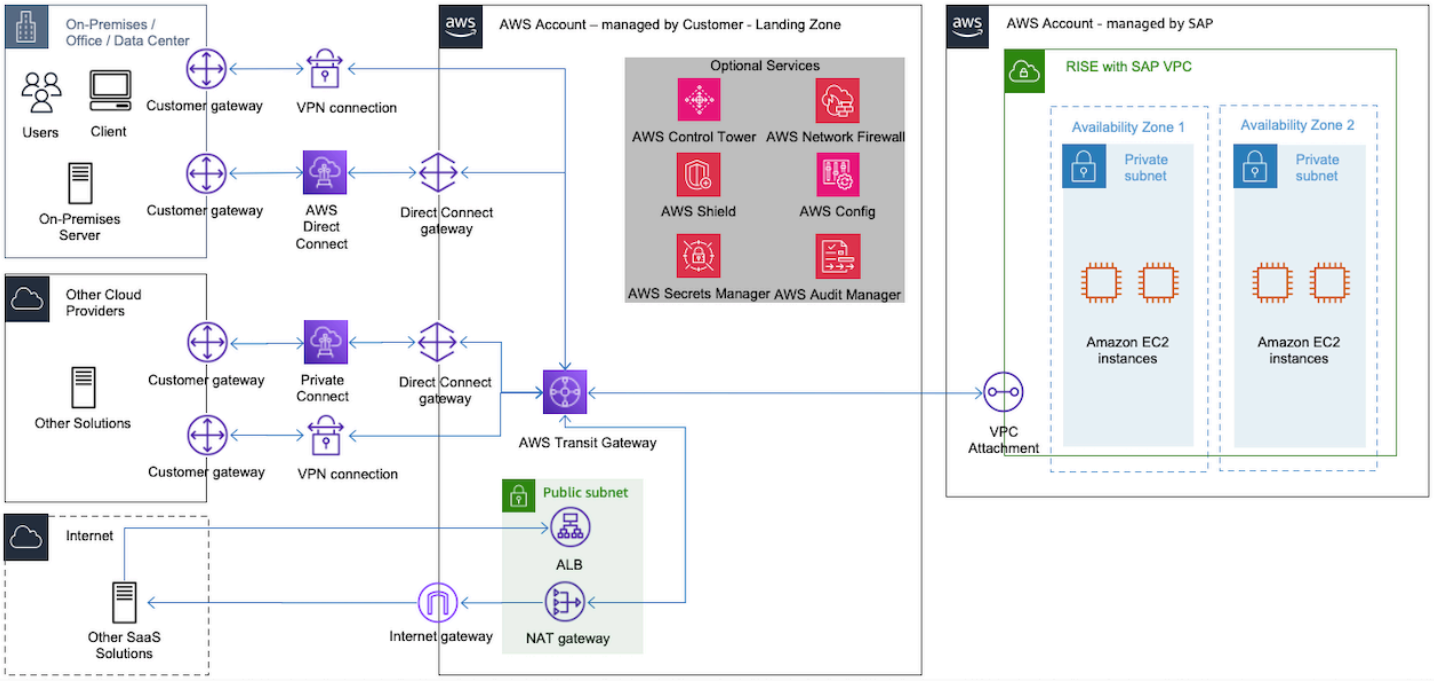
SaaS cloud solutions do not offer connectivity via VPN, Direct Connect or any other means of private connectivity. You can implement a centralized egress to internet architecture to manage this connectivity. For more information, see Centralized egress to internet.

Connectivity patterns for multi-cloud to RISE

In a complex connectivity scenario, you may need to integrate RISE with SAP setup with on-premises, AWS-hosted systems, and a variety of SaaS solutions and other cloud service providers.

Managing connectivity directly from the AWS environment decouples dependencies with on-premises networking infrastructure, improving availability and resiliency of the overall landscape.

You can use public or private connectivity to connect multi-cloud with RISE.



Public connectivity

Connectivity is routed over the public internet. This pattern is typically used for connectivity from RISE with SAP to SaaS solutions that runs across multiple clouds. When building connectivity routed over the public internet, consider the following:

- ensure that all communication is encrypted
- protect end-points by using AWS services, such as Elastic Load Balancers and AWS Shield
- monitor endpoints using Amazon CloudWatch
- ensure that traffic between two public IP addresses hosted on AWS is routed over the AWS network

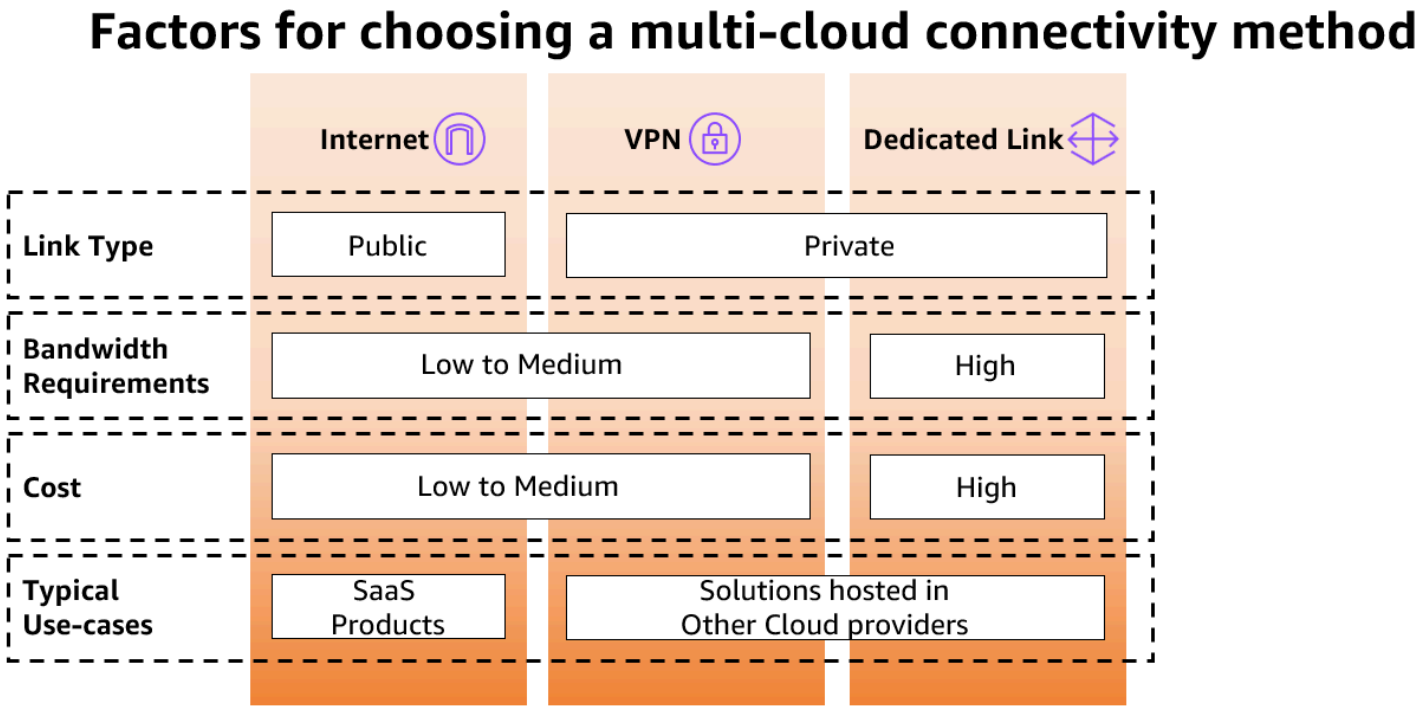
Private connectivity

The following three are the options to establish private connectivity between different cloud service providers:

- Site-to-site VPN encrypted tunnel routed over public internet

- private interconnect using AWS Direct Connect in a managed infrastructure (use Azure ExpressRoute for Azure and Google Dedicated Interconnect for Google Cloud Platform)
- private interconnect using an AWS Direct Connect in a facility with a multi-cloud connectivity provider

The following diagram describes the factors to choose a multi-cloud connectivity method.

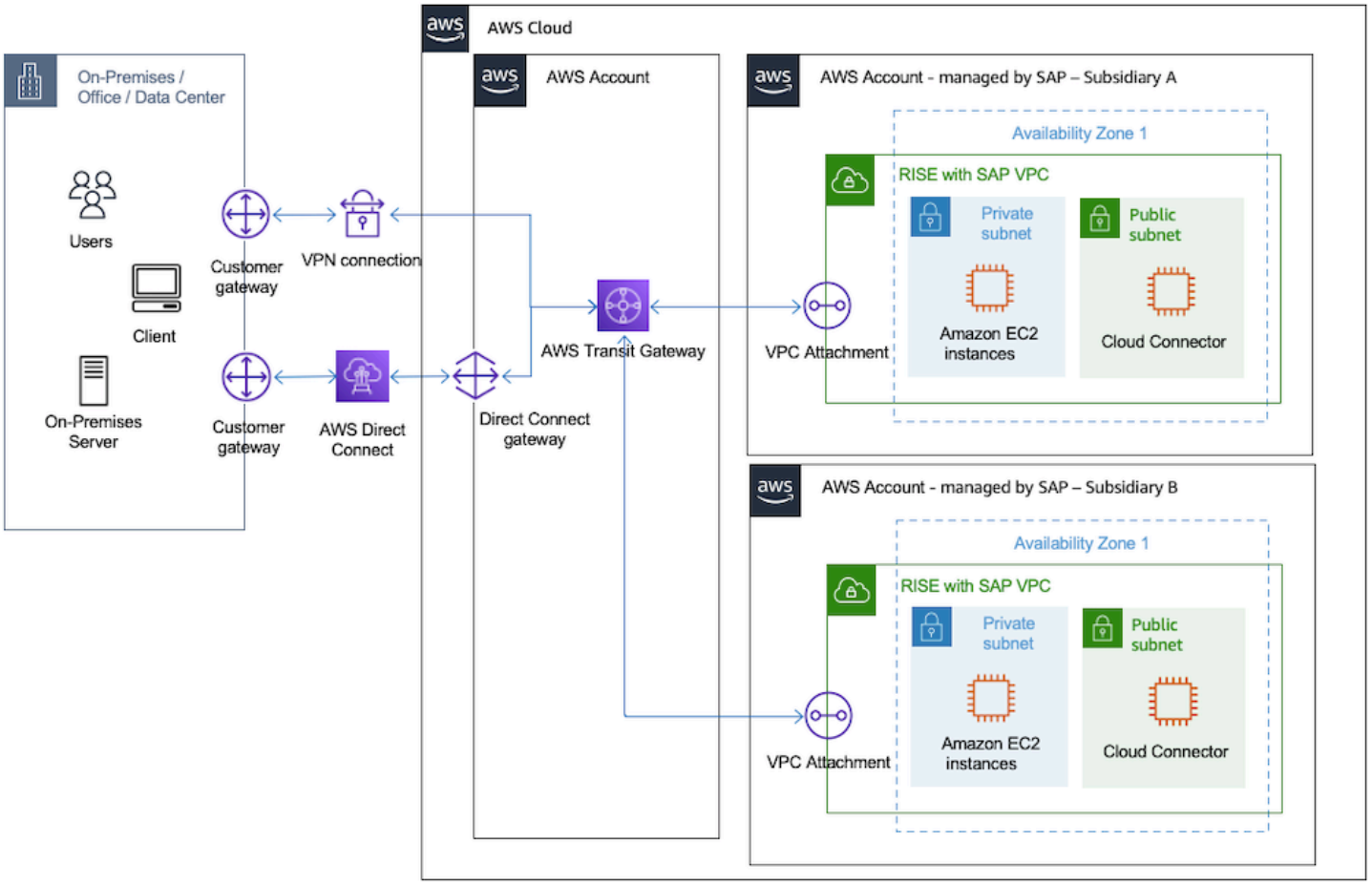


For more information, see Designing private network connectivity between AWS and Microsoft Azure.

How to implement chargeback capability for connectivity to RISE

If you are a company with subsidiaries, you may have different RISE contracts, leading to deployments in separate AWS accounts while requiring an interlinked network connectivity. In this instance, you must deploy Transit Gateway connection in a Landing Zone (multi-account) setup. It can scale your RISE deployment and integrate with multiple RISE with SAP VPCs.

Transit Gateway Flow Logs enables effective cost management. Transit Gateway Flow Logs can be integrated with Cost and Usage Report (CUR) that can be attributed as chargeback to the business units. For more information, see Logging network traffic using Transit Gateway Flow Logs.



The preceding diagram displays how Transit Gateway can be used to connect multiple RISE with SAP VPCs and provide chargeback capability through the Flow Logs.

For more information, see the following blogs:

- Using AWS Transit Gateway Flow Logs to chargeback data processing costs in a multi-account environment
- How-to chargeback shared services: An AWS Transit Gateway example

Use the following steps to enable this setup:

1. Enable Transit Gateway Flow Logs. For more information, see Create a flow log that publishes to Amazon S3.
2. Setup Cost and Usage Reporting and setup Athena to utilize the reporting. For more information, see Creating Cost and Usage Reports and Querying Cost and Usage Reports using Amazon Athena.

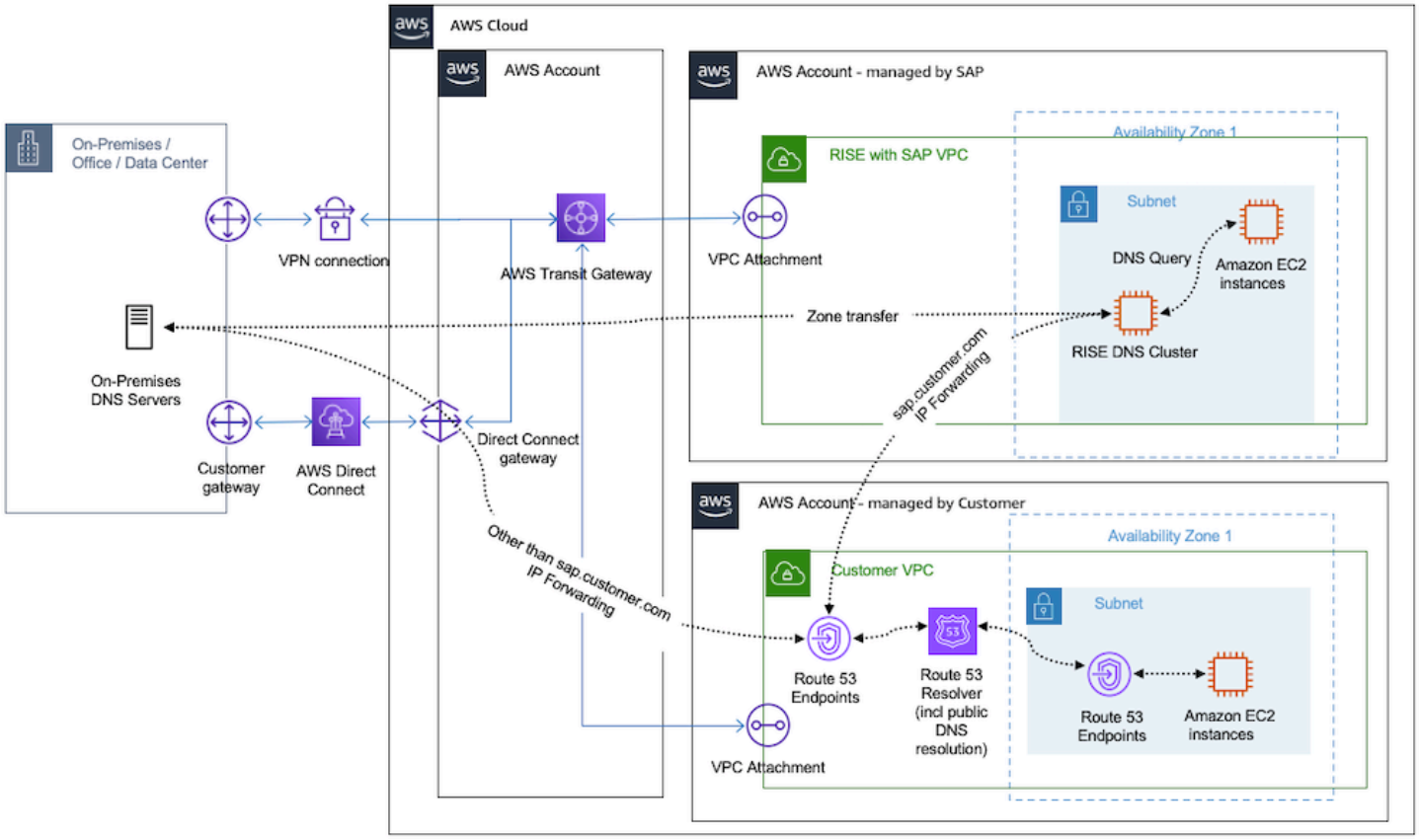
3. Obtain the Transit Gateway data processing charge per-account.
    a. Decide a cost allocation strategy - distribute costs evenly across all accounts or distribute proportionally across all accounts.
    b. Calculate the total network traffic and percentage allocation per account using AWS Transit Gateway query.
    c. Estimate cost per account, by collecting from CloudWatch that collects Network In(Upload) and NetworkOut(Download).
        i. NetworkIn(Upload) + NetworkOut(Download) per usage account/ total data processed in network account
        ii. % of usage x total cost = chargeback cost per usage account

Integrating Domain Name System to RISE and Route 53 consideration

Domain Name System (DNS) translates domain names and IP addresses in both ways such as (www.amazon.com as 1.2.3.4 in IPv4 or 1200:ab00:1024:1::b410:e6b1 in IPv6). There are two methods available to integrate DNS to RISE:

- **DNS Zone Transfer**: This method replicates the DNS databases across a set of DNS Servers. It can provide higher availability, redundancy and higher SLA. In a typical RISE environment, you delegate a sub-domain (example. *.sap.customer.com) to the RISE DNS Servers. If one server becomes unavailable, other servers can continue to provide DNS services without interruption. If a customer is using Route 53, which does not support Zone Transfer, customer need to set up Route 53 inbound resolver together with a Route 53 private hosted zone and provide the resolver IPs to SAP for routing purposes, you can find out more in the DNS Conditional Forwarding as below.
- **DNS Conditional Forwarding (also called IP Forward)**: The DNS servers only forward queries for specific domain names that is not part of its own name space. (example: Amazon Route 53 forwards DNS query of *.sap.customer.com to RISE DNS Servers). This allows for more precise control over how DNS queries are handled, especially in complex network environments. Amazon Route 53 Resolver supports both public and private DNS resolution, facilitates hybrid cloud setups, and enhances availability, performance, and security for AWS resources and on-premises networks.

Here is a reference architecture on how to integrate Amazon Route 53 with RISE.



For more details, refer to the following AWS documentation:

- Forwarding outbound DNS queries to your network
- Forwarding inbound DNS queries to your VPCs
- Set up DNS resolution for hybrid networks in a single-account AWS Environment
- Set up DNS resolution for hybrid networks in a multi-account AWS Environment
- What is Amazon Route 53 Resolver ?

Security

SAP manages the security in AWS account managed by SAP. You can implement additional security mechanisms in your own AWS account.

Topics

- Single Sign-On – SAP Cloud Identity Services and AWS IAM Identity Center
- Single Sign-On – SAP Cloud Identity Services and Microsoft Entra (previously Azure AD)

- Single Sign-On – SAPGUI Front-End
- Advanced security using AWS Services for RISE with SAP
- Integration SAP Key Management Service (SAP KMS) with AWS Key Management Service (AWS KMS)
- How AWS Nitro helps secure RISE with SAP?
- Amazon WorkSpaces as remote access solution for RISE with SAP

Single Sign-On – SAP Cloud Identity Services and AWS IAM Identity Center

One of the security best practices for RISE with SAP is to centralize the user access control through the integration with a corporate Identity Provider (IdP). This makes it easier for you to provision, de-provision and manage your user access across the company including RISE with SAP, AWS services, and others.

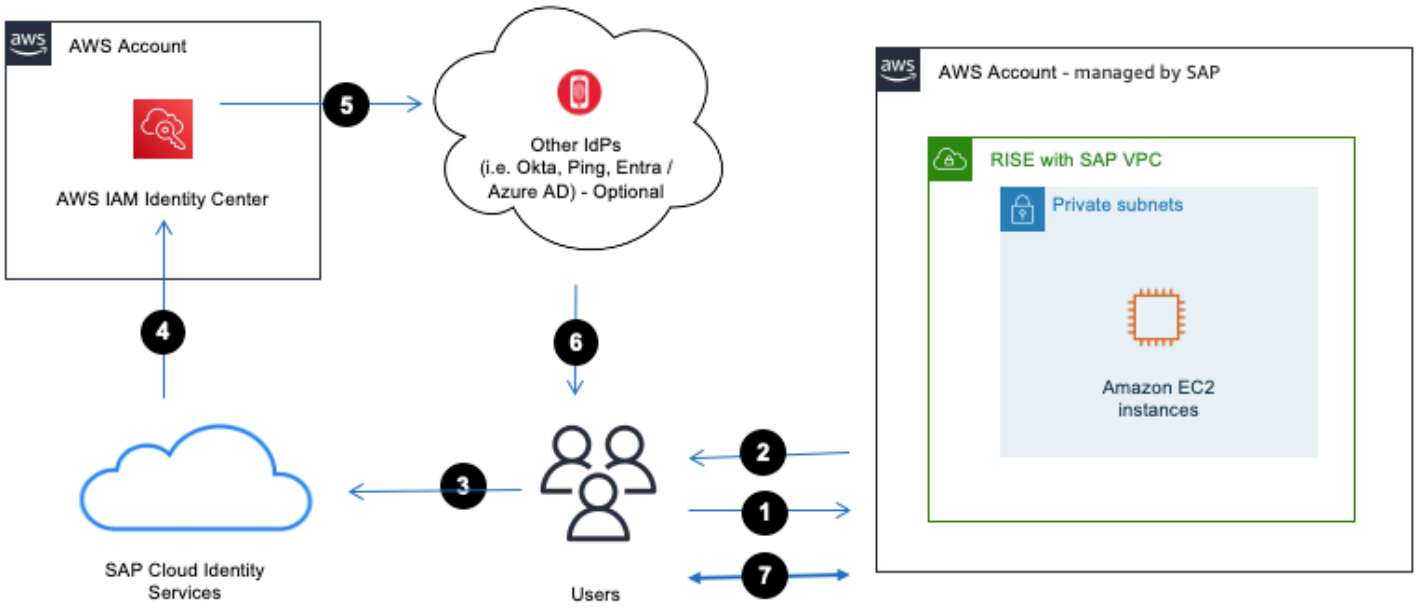
AWS IAM Identity Center is one of the IdP that you can integrate with RISE. IAM Identity Center provides a centralized access points for users to manage AWS account and applications consistently within the AWS Organizations (example in multi accounts setup).

If you already have an existing identity source such as Okta, Ping, Microsoft Windows Active Directory, Microsoft Entra (previously known as Azure Active Directory), or others, you can integrate the identity source to IAM Identity Center through Security Assertion Markup Language (SAML) and System for Cross-Domain Identity Management (SCIM) protocols.

For more information, you can refer to the following references:

- What is IAM Identity Center?
- Integration of IAM Identity Center with other identity source, see Getting started tutorials.
- SAP Cloud Identity Services - Identity Authentication.

The following image shows the integration between Identity Authentication from SAP BTP and AWS IAM Identity Center in the context of RISE with SAP



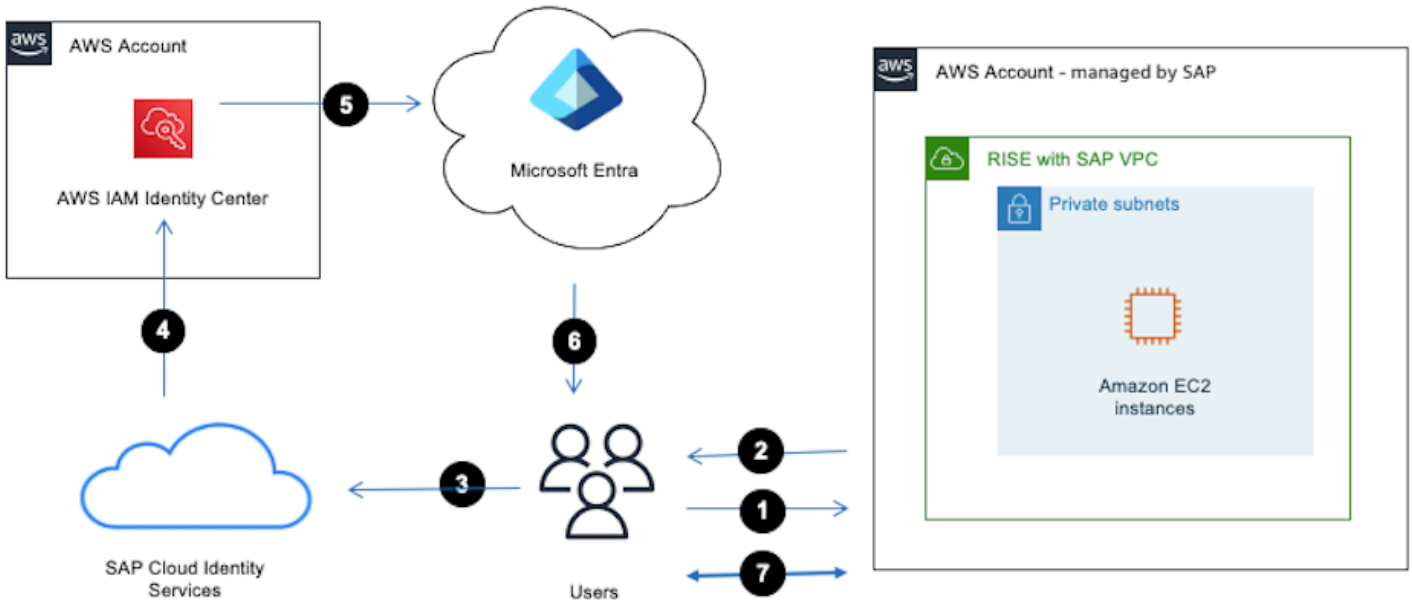
Authentication flow

1. User accesses SAP Fiori via an Internet browser.
2. SAP Fiori will redirect SAML request back to the internet browser.
3. Internet Browser relays the SAML request to SAP Cloud Identity Services.
4. SAP Cloud Identity Service delegate authentication request to IAM Identity Center.
5. If IAM Identity Center integrates with existing identity source such as Okta, Ping, Entra, then IdP will authenticate the user.
6. User is authenticated by IdP and SAML response is provided to the internet browser with user identity information.
7. User can access RISE with SAP systems.

For more information on how to do this, you can refer to AWS IAM Identity Center (successor to AWS SSO) Integration Guide for SAP Cloud Platform Cloud Foundry.

Single Sign-On – SAP Cloud Identity Services and Microsoft Entra (previously Azure AD)

Microsoft Entra (previously Azure AD) or other IdPs can be integrated to SAP Cloud Identity Services directly. This support a direct authentication, when you do not need AWS IAM Identity Center (i.e. no requirement to run a multi account strategy that utilizes AWS Organizations).



Authentication flow

1. User accesses SAP Fiori via an Internet browser.
2. SAP Fiori will redirect SAML request back to the internet browser.
3. Internet Browser relays the SAML request to SAP Cloud Identity Services.
4. SAP Cloud Identity Service delegate authentication request to IdPs.
5. User is authenticated by IdP and SAML response is provided to the internet browser with user identity information.
6. User can access to SAP S/4HANA in RISE with SAP VPC.

For more information on how to do this, you can refer to Enable SSO between Azure AD and SAP Cloud Platform using Identity Authentication Service.

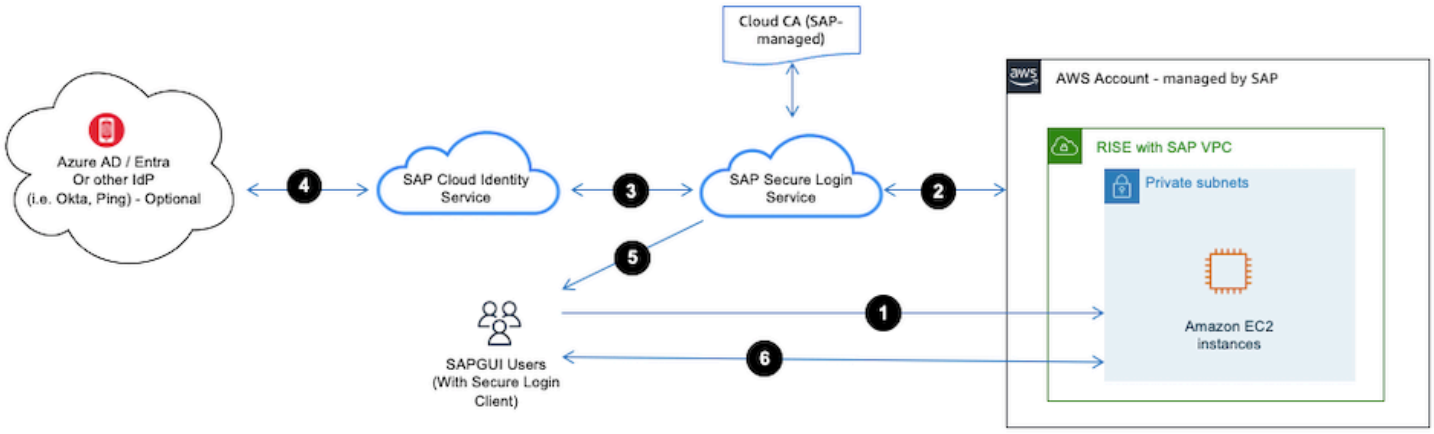
Single Sign-On – SAPGUI Front-End

SAPGUI is a graphical user interface client in the SAP ERP's three-tier architecture of database, application servers and clients. It requires installation in a local desktop that run on Windows or macOS or Linux.

In order to achieve Single-Sign-On for SAPGUI in RISE with SAP, we must use either Kerberos or X.509 method. Kerberos is not recommended by AWS, because it requires user to always be

connected to the corporate network and authenticated against a Microsoft Active Directory which reduce their mobility. Due to this, X509 is recommended.

SAPGUI Single-Sign-On with X509 can be achieved with SAP Secure Login Service on BTP, the image below describes how the integration works.



Authentication flow

1. User accesses SAPGUI on their desktop.
2. SAP S/4HANA will redirect authentication request to SAP Secure Login Service.
3. SAP Secure Login Service will delegate the authentication to SAP Cloud Identity Service.
4. When SAP Cloud Identity Service is integrated to IdP (i.e. Azure AD, Okta, Ping, etc.), then IdP will authenticate the user.
5. User is authenticated by IdP and X509 is provided by SAP Secure Login Service to the SAPGUI.
6. User can access to SAP S/4HANA in RISE with SAP VPC.

For more information on how to do this, you can refer to Securing SAP GUI with SAP Secure Login Service.

Advanced security using AWS Services for RISE with SAP

AWS offers a comprehensive suite of security services that can act as a multi-layered security envelope around RISE with SAP deployments on AWS. These services act as an additional security barrier, intercepting and mitigating potential threats before they can reach the RISE account, providing robust protection and assisting with compliance with industry-standard security best practices.

AWS Network Firewall

AWS Network Firewall is a managed firewall service that provides essential network protection for Amazon Virtual Private Cloud (VPC) environments. AWS Network Firewall acts as a first line of defence, filtering and inspecting all network traffic to and from RISE resources, effectively creating a protective perimeter around a RISE environment.

Key features of AWS Network Firewall include:

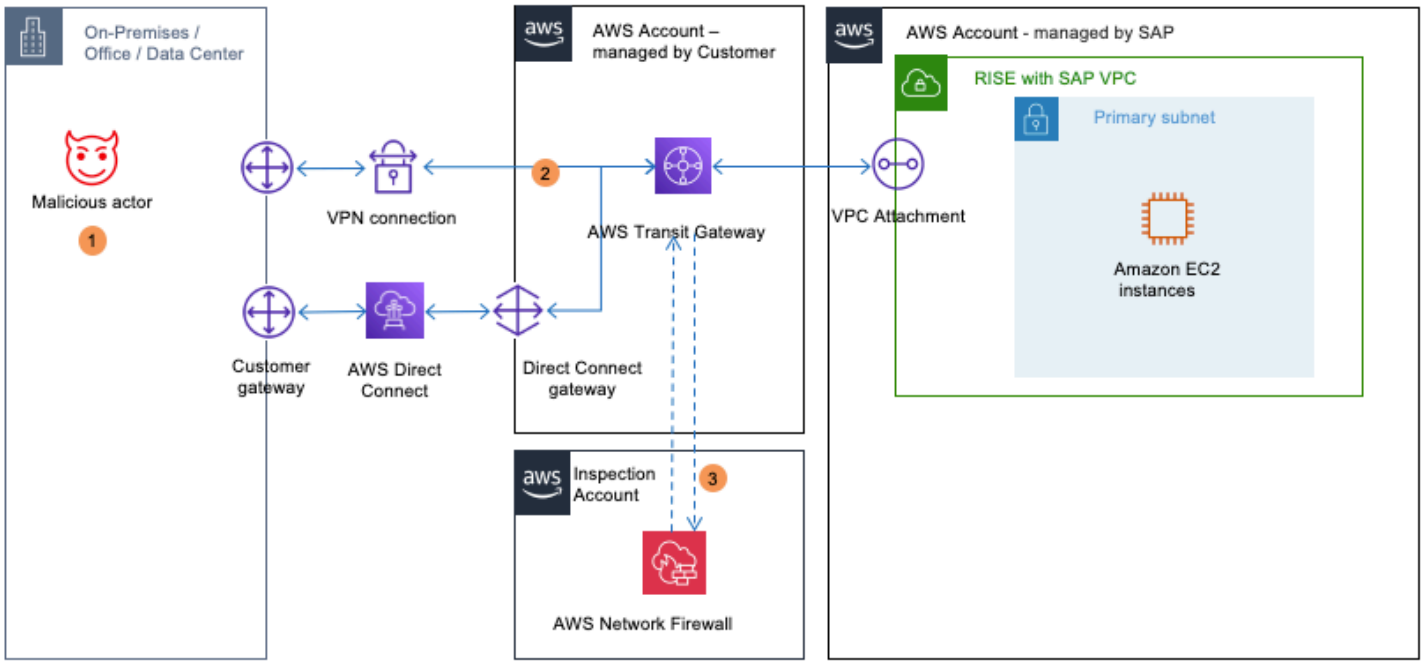
- Stateful Firewall Capabilities. AWS Network Firewall offers advanced stateful firewall features to monitor and control network traffic. It can inspect the complete context of a network connection, including source, destination, ports, and protocols, to detect and block malicious or unauthorized traffic.
- Threat Signature Matching. AWS Network Firewall comes pre-loaded with a comprehensive set of threat detection rules and signatures, continuously updated by AWS, to identify and mitigate known threats, malware, and other malicious activity targeting RISE deployments.
- Custom Rule Definition. In addition to the pre-defined threat signatures, customers can create and deploy custom firewall rules to address specific security requirements or policies unique to connections hitting SAP systems in the RISE environment.
- Centralized Policy Management. AWS Network Firewall allows to define and manage firewall policies centrally, which can then be easily deployed across multiple VPCs including non-SAP VPCs and VPCs associated with the SAP-managed RISE VPC, ensuring consistent security enforcement.
- Scalability and High Availability. As a fully managed service, AWS Network Firewall automatically scales to handle changes in network traffic volume and patterns, ensuring RISE environment remains protected without the need for complex infrastructure management.

In the context of RISE with SAP, AWS Network Firewall can be leveraged for the following:

- Centralized Firewall Management. AWS Network Firewall provides a centralized, managed firewall service to control and monitor network traffic travelling to and from the SAP-managed RISE VPC.
- Stateful Packet Inspection. AWS Network Firewall performs stateful packet inspection, allowing it to detect and mitigate advanced threats by analysing the context of network connections to/from SAP systems within the RISE VPC.

- Regulatory Compliance. AWS Network Firewall helps organizations meet compliance requirements by enforcing security policies and providing logging/auditing capabilities for the RISE with SAP landscape.

Below is example architecture of AWS Network Firewall inspecting network traffic before it reaches RISE with SAP



In the diagram above

1. A malicious actor exploits network misconfiguration to get access to SAP system on RISE.
2. Traffic is first routed through AWS Transit Gateway.
3. Packet inspection by AWS Network Firewall catches abnormal connection attempts..

It is worth noting that AWS Network Firewall can be also used by customers who want to consume SAP BTP services hosted by AWS connecting first to an AWS Transit Gateway with AWS Direct Connect, so that their end-to-end stay on the AWS backbone.

For instructions to configure AWS Network Firewall, see Getting started with AWS Network Firewall.

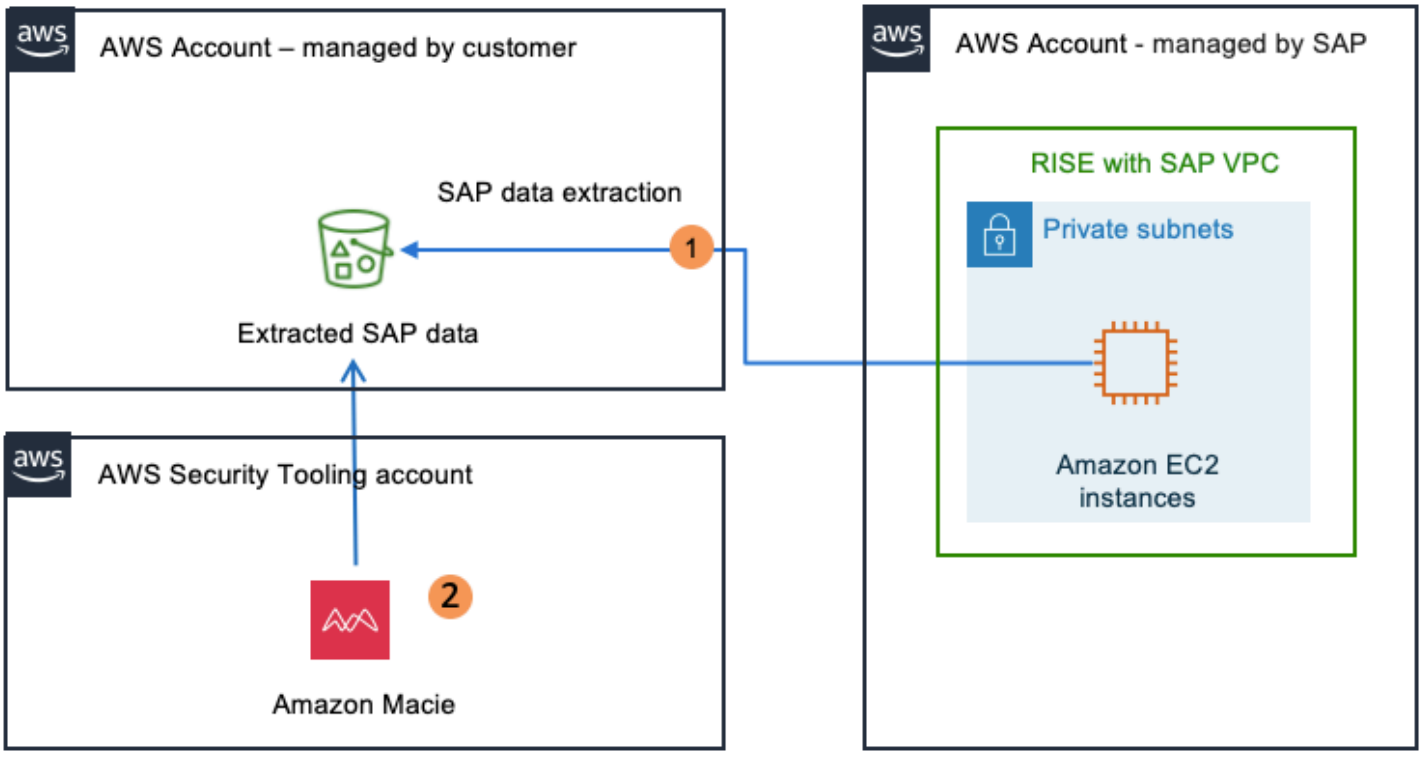
Amazon Macie

Amazon Macie is a data security service that helps customers discover, classify, and protect sensitive data stored in Amazon S3 buckets by continuously monitoring and alerting on potential data risks and unauthorized access attempts.

In the context of RISE with SAP, Amazon Macie can protect Amazon S3 buckets in customer-managed AWS account fed by a RISE with SAP environment, for instance:

- as a RISE customer, backups can be copied from the SAP-managed AWS account to a customer-managed environment and S3 bucket.
- SAP data can be extracted from or a RISE environment (see Architecture Options for extracting SAP Data with AWS Services) to a customer-managed S3 bucket, to enable advanced analytics, machine learning, and business intelligence using other AWS services like Amazon Athena, AWS Glue, and Amazon Sagemaker;
- Certain industries and regulations, such as GDPR, HIPAA, or PCI-DSS, may require long-term storage and preservation of sensitive data. Exporting this data to a customer-managed S3 can help meet these compliance requirements, as S3 provides robust security and durability features.
- Centralized Policy Management. AWS Network Firewall allows to define and manage firewall policies centrally, which can then be easily deployed across multiple VPCs including non-SAP VPCs and VPCs associated with the SAP-managed RISE VPC, ensuring consistent security enforcement.
- Customers can also consume security event logs out of their RISE environment, so ingest in their own S3 buckets or SIEM systems.

Below is example architecture of Amazon Macie continuously scanning an S3 bucket with SAP data extracted from RISE



In the diagram above

1. Data is written to S3 bucket for data lake/compliance reporting purposes.
2. Amazon Macie continuously analyzes bucket to detect Privately Indentifiable Information.

For instructions to configure Amazon Macie, see What is Macie ?.

Amazon GuardDuty

Amazon GuardDuty is a threat detection service that continuously monitors for malicious activity and unauthorized behaviour within an AWS environment. It combines machine learning, anomaly detection, and integrated threat intelligence to identify potential threats and protect AWS account linked to RISE with SAP environments, workloads, and data.

Amazon GuardDuty monitors the following:

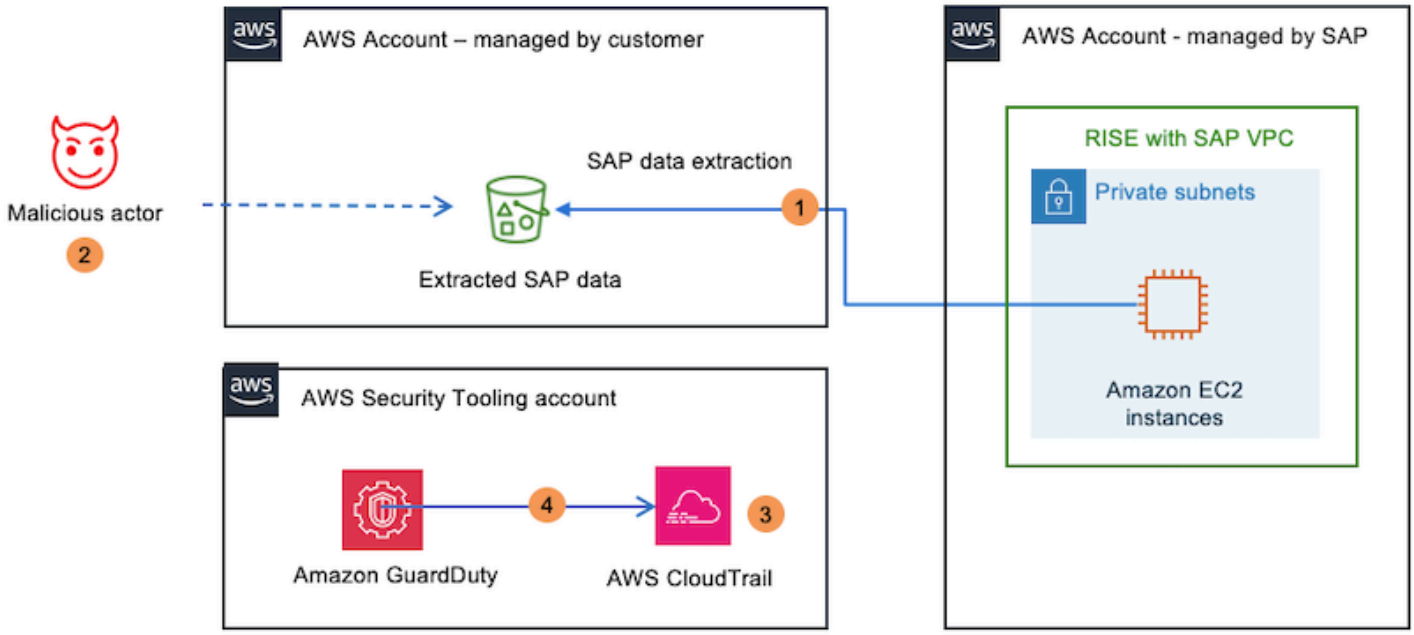
- AWS CloudTrail Logs: Amazon GuardDuty monitors API activity across AWS account to detect suspicious API calls, unauthorized deployments, and unauthorized access attempts to resources. Amazon GuardDuty identifies attempts to access AWS services from unauthorized IP addresses or regions. Amazon GuardDuty detects unusual behaviour in Identity and Access Management (IAM) users, roles, and policies, such as privilege escalation.

- VPC Flow Logs. Amazon GuardDuty analyses network traffic within a Virtual Private Cloud (VPC) to detect unexpected traffic patterns, data exfiltration attempts, or unauthorized access alongside identifying communications between AWS resources and known malicious IP addresses or domains. In the context of RISE with SAP on AWS, the inspection takes places on a VPC fronting the RISE SAP-managed account;
- DNS Logs. Amazon GuardDuty monitors DNS queries made by an AWS resource to detect attempts to connect to malicious domains or unusual DNS request patterns. Amazon GuardDuty also detects the use of Domain Generation Algorithms (DGA) for generating domain names associated with Command and Control servers.

In the context of RISE with SAP, Amazon GuardDuty can be leveraged for the following:

- Intrusion Detection: GuardDuty enables early detection of intrusion attempts into an RISE environment fronted by a customer-managed AWS account by identifying malicious activities such as unauthorized API calls, network reconnaissance, and access attempts from known malicious IP addresses;
- Compliance Validation: For organizations with stringent compliance requirements, GuardDuty helps ensure adherence by continuously monitoring for policy violations and unauthorized access attempts, providing detailed logs and reports for audit purposes. This can be achieved when the SAP RISE environment is accessed from a customer-managed AWS account. See Compliance Validation for more details
- Automated Incident Response. GuardDuty can be integrated with AWS Lambda and AWS Security Hub to automate incident response workflows. Upon detecting a threat, these services can trigger automated remediation actions, such as isolating compromised resources or notifying security teams.

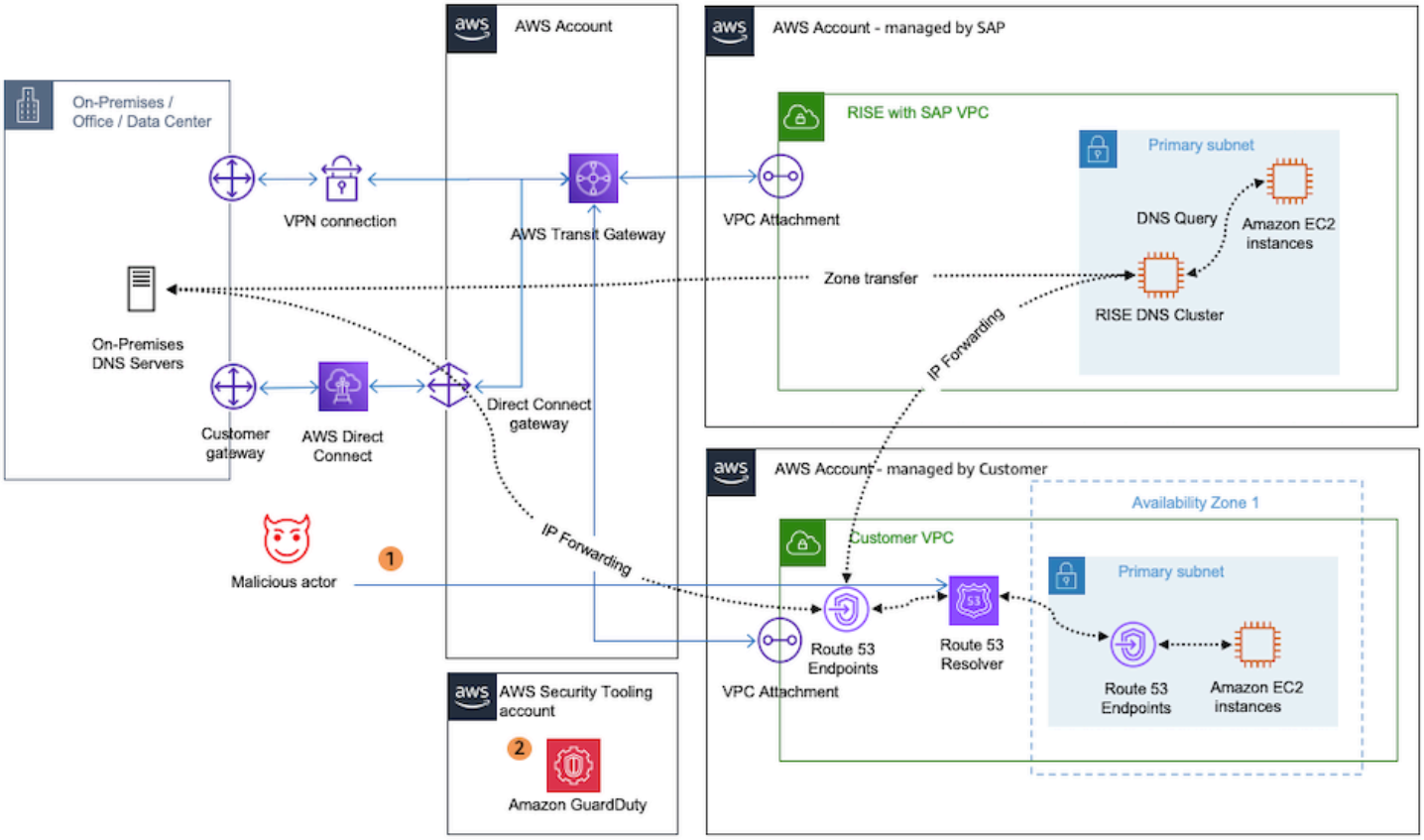
Below is example architecture of GuardDuty monitoring CloudTrail trails of a RISE with SAP deployment on AWS



In the diagram above

1. Data is written to S3 bucket for data lake/compliance reporting purposes.
2. A malicious actor changes IAM rules and IAM permissions on S3 bucket to obtain access.
3. IAM changes are intercepted by AWS CloudTrail.
4. GuardDuty detects suspicious activity and alerts administrators.

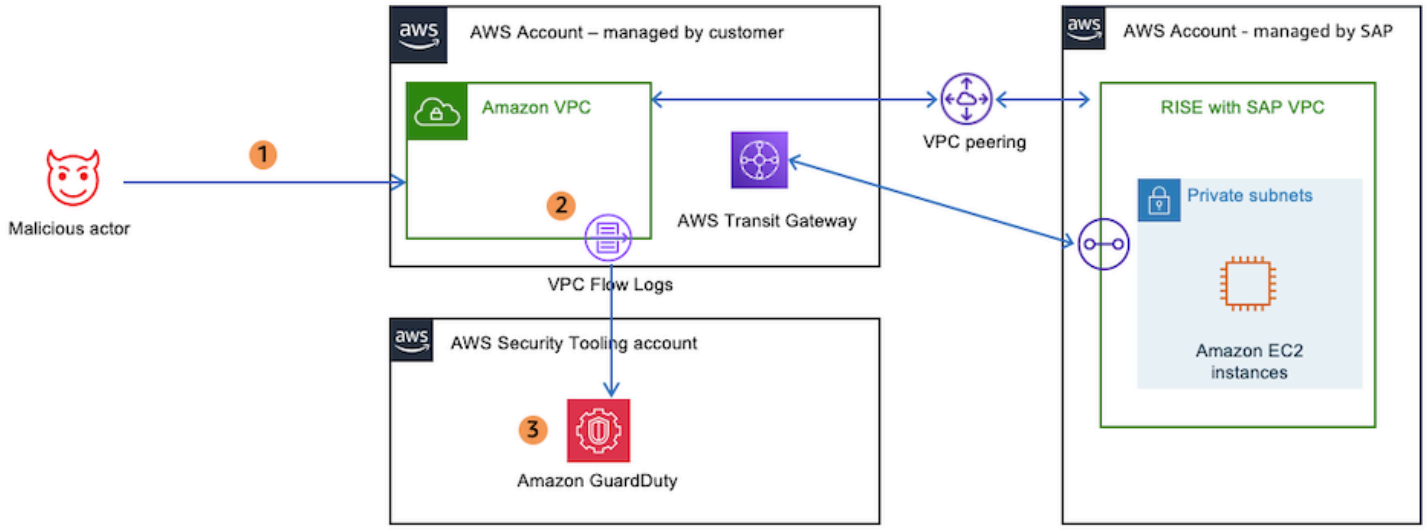
Below is example architecture of GuardDuty monitoring DNS logs of a RISE with SAP deployment on AWS



In the diagram above

1. A malicious actor introduces rogue DNS redirecting users to makeshift SAP systems.
2. The rogue DNS entries are detected by GuardDuty and reported to administrators.

Below is example architecture of GuardDuty monitoring VPC Flow Logs of RISE with SAP VPC



In the diagram above

1. A malicious actor attempts to access SAP systems from VPC managed by customer peered to RISE VPC or scan ports.
2. The connection attempt from malicious actor IP logged in VPC Flow Logs.
3. The suspicious connection attempt is detected by Amazon GuardDuty and reported to administrators.

For instructions to configure Amazon GuardDuty, see Getting Started.

Using security services with AWS Security Hub, Amazon Detective, AWS Audit Manager and Amazon EventBridge

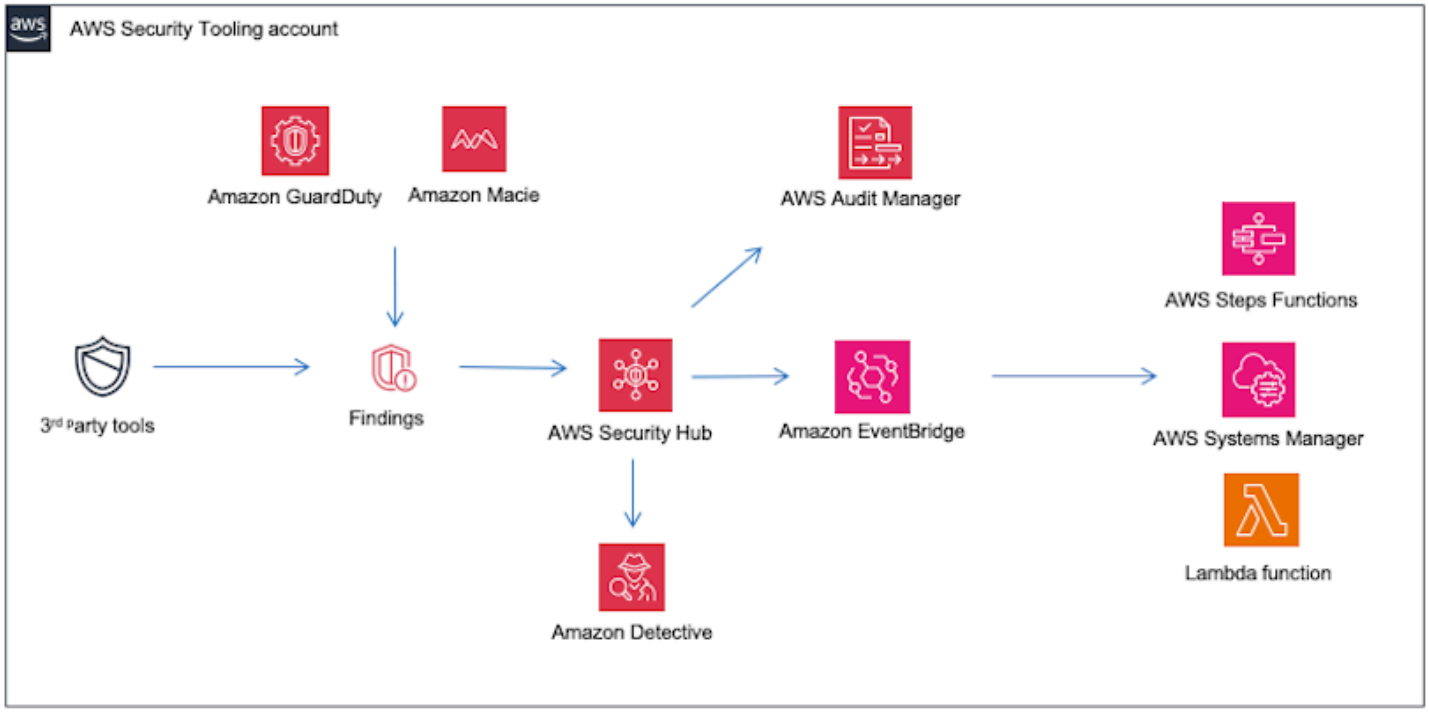
Building on implementation of GuardDuty and Amazon Macie, AWS Security Hub acts as a central hub, consolidating and prioritizing security findings AWS security services. AWS Security Hub provides a unified view of the security posture across services surrounding a RISE with SAP deployment, allowing too quickly identify and address any security issues.

To further investigation and incident response capabilities, Amazon Detective analyses security incidents by gathering and processing relevant log data from AWS resources. This service helps quickly identify the root cause of issues, enabling to take appropriate actions to mitigate the impact.

Maintaining compliance is also a critical aspect of securing a RISE with SAP environment. AWS Audit Manager automates the assessment of AWS resources against industry standards and regulations, helping demonstrate compliance and reduce the risk of non-compliance.

Finally, Amazon EventBridge enables real-time response to security events by triggering custom automated workflows and remediation actions. This service allows to quickly and efficiently address security incidents, minimizing the potential impact on RISE with SAP deployment

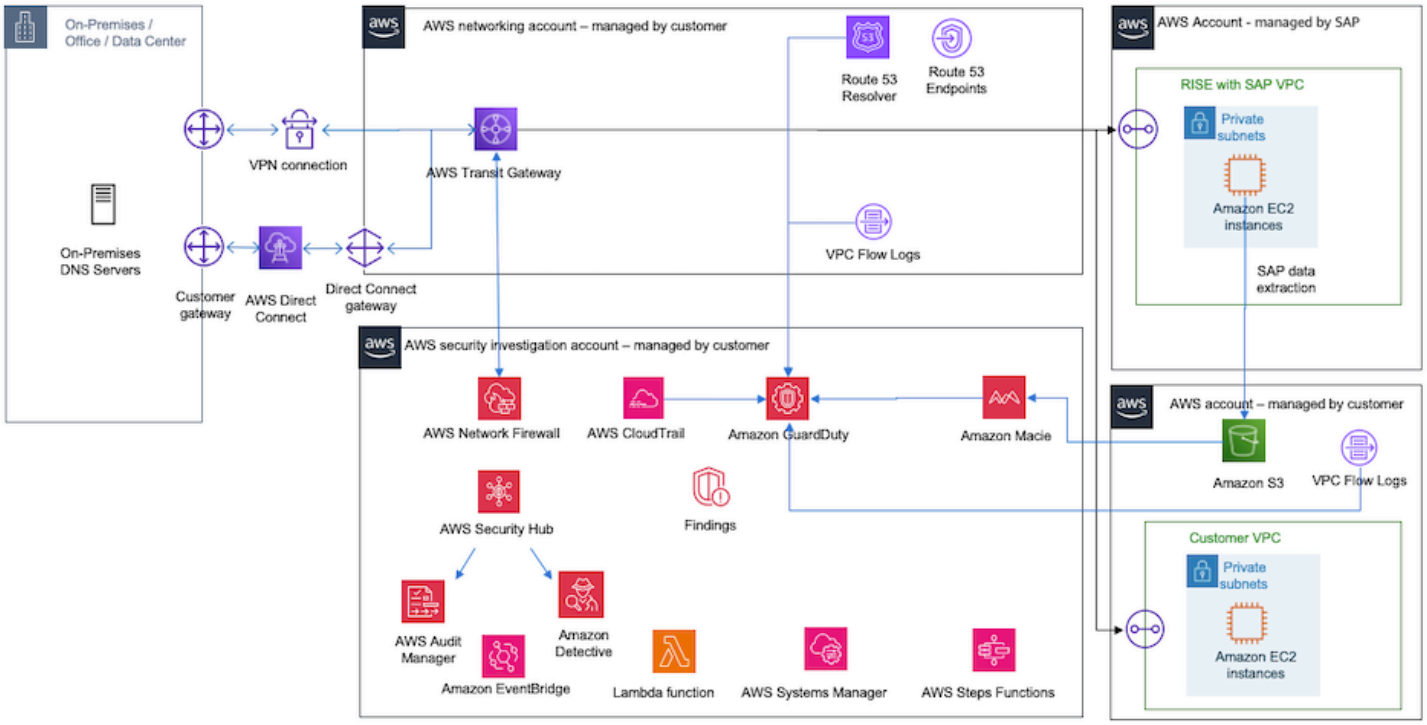
Below is example architecture of AWS Security Hub, Amazon Detective, AWS Audit Manager and Amazon EventBridge paired to RISE with SAP



Using All AWS Security Services

Combining together all services described above allow for an architecture monitoring multiple areas of a RISE on AWS deployment: network traffic, DNS logs, CloudTrail API activity, sensitive information extracted SAP data. Amazon GuardDuty and AWS Security Hub are fed from multiple services and uses AIML intelligence to detect malicious activities and anomalies. Findings are passed to Amazon Detective for a deeper RCA analysis or sent to Amazon EventBridge for custom reporting and alerting.

Below is example architecture of GuardDuty, AWS Network Firewall, Amazon Macie, AWS Security Hub and Amazon Detective combined together to improve security posture of RISE with SAP on AWS deployment



Integration SAP Key Management Service (SAP KMS) with AWS Key Management Service (AWS KMS)

SAP KMS is a multi-cloud SaaS application that helps organizations maintain visibility, control, and encryption of their data in the cloud. It consists of two main services:

- Transparency and Control Service, which provides data governance capabilities, including data lineage, auditing, and compliance monitoring.
- Key Management Service, which enables customer-managed encryption keys for data stored in various cloud services, including AWS. Please note that SAP KMS is not the same as AWS KMS.

SAP KMS is typically used by SAP customers for the following:

- Data Governance and Compliance. SAP KMS helps organizations comply with global data protection regulations, such as GDPR and CCPA, by providing visibility, control, and auditing over data storage and processing.
- Data Transparency. SAP KMS offers real-time insights, data lineage tracking, and auditing capabilities for monitoring data usage and compliance status across cloud environments.

- Security and Access Control. SAP KMS implements advanced security measures, including encryption, access controls, and anomaly detection, to protect sensitive data from unauthorized access.
- Automation and Policy Management. SAP KMS automates the enforcement of data policies and allows for customizable policies tailored to specific business needs and regulatory requirements.

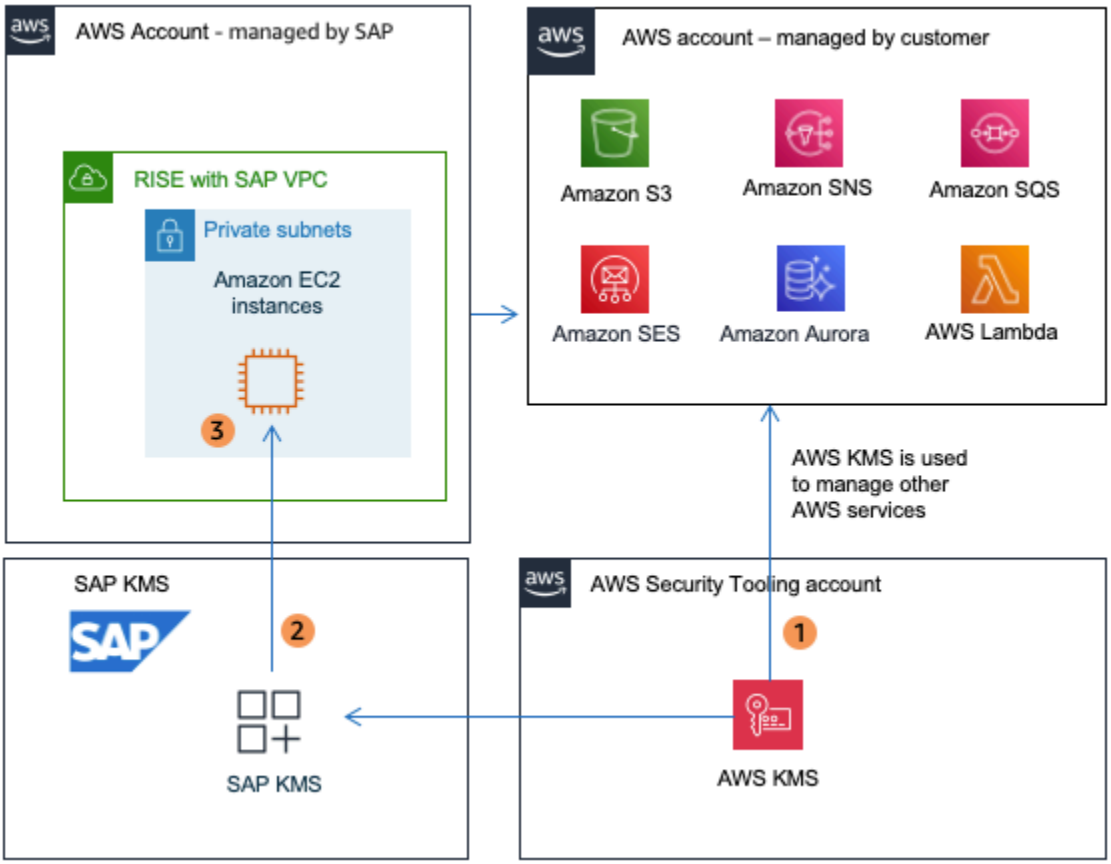
As an alternative to using SAP KMS' own Key Management Service, SAP KMS can be integrated with AWS KMS. Using AWS KMS as the keystore for SAP KMS provides a consistent and centralized approach to key management, especially if AWS KMS is already employed for other AWS workloads, enabling seamless integration, streamlined key lifecycle management, and enhanced security through AWS robust encryption and access control mechanisms.

This integration allows customers to manage and control the encryption keys used to protect their sensitive data, ensuring greater security and compliance. SAP KMS can be interfaced with AWS KMS either in BYOK (Bring Your Own Key) or HYOK (Hold Your Own Key) scenarios:

Area	AWS KMS (BYOK Scenario)	AWS KMS (HYOK Scenario)
Supported Key Types	AES, RSA	RSA
Supported Key Sizes	3072, 4096	3072, 4096
Key Management	Key is created in AWS KMS keystore and imported into the SAP KMS-provided tenant	Key is created and stored in AWS KMS keystore
Key Revocation	Key can be disabled or deleted at any time	Key can be disabled or unregistered at any time

Note that SAP recommends that keystores be enabled in the same AWS Region as the consuming SAP service and the SAP KMS tenant (see AWS BYOK Scenarios). SAP KMS only supports Single-Region Keys, based on Keystore Region Availability.

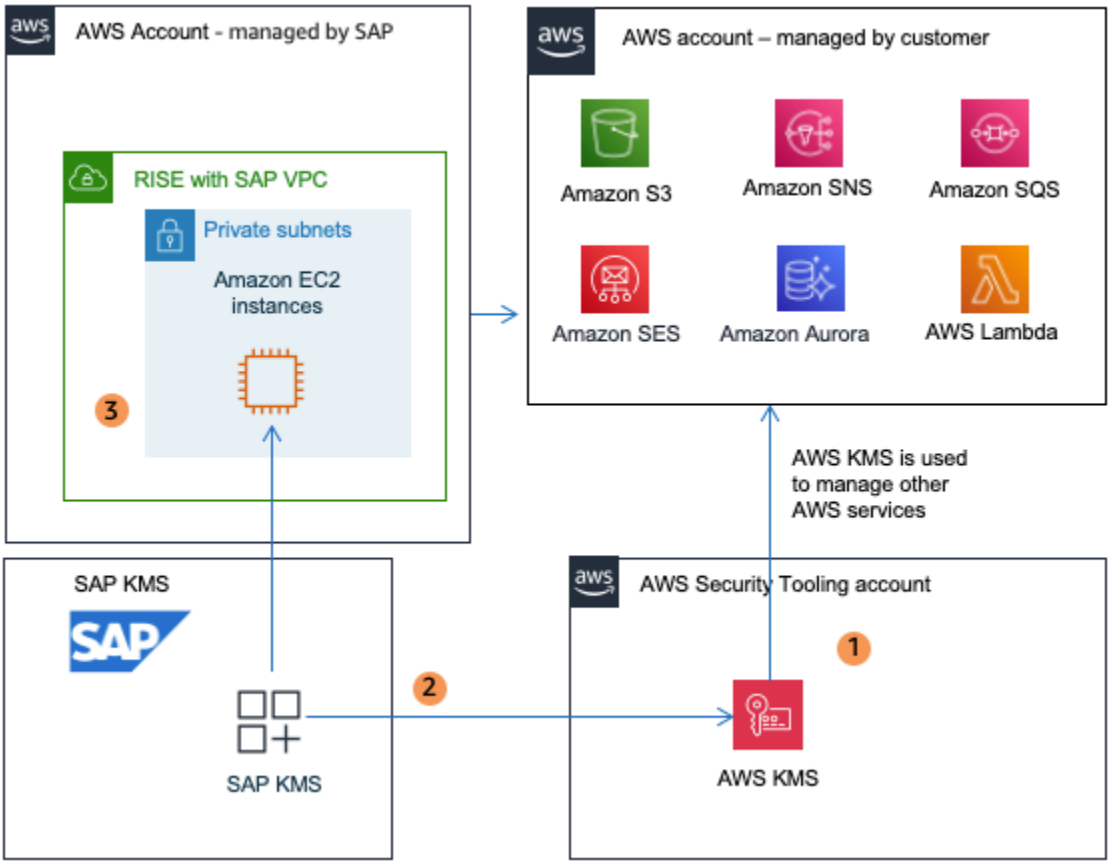
Below is the SAP KMS integration with AWS KMS - BYOK



In the diagram above:

- Key is created in AWS KMS keystore
- Key is imported into SAP KMS tenant
- SAP KMS encrypts SAP data at application level

Below is the SAP KMS integration with AWS KMS - HYOK



In the diagram above:

- Key is created in AWS KMS keystore
- Key is stored in AWS KMS and retrieved by SAP KMS when required
- SAP KMS encrypts SAP data at application level

How AWS Nitro helps secure RISE with SAP?

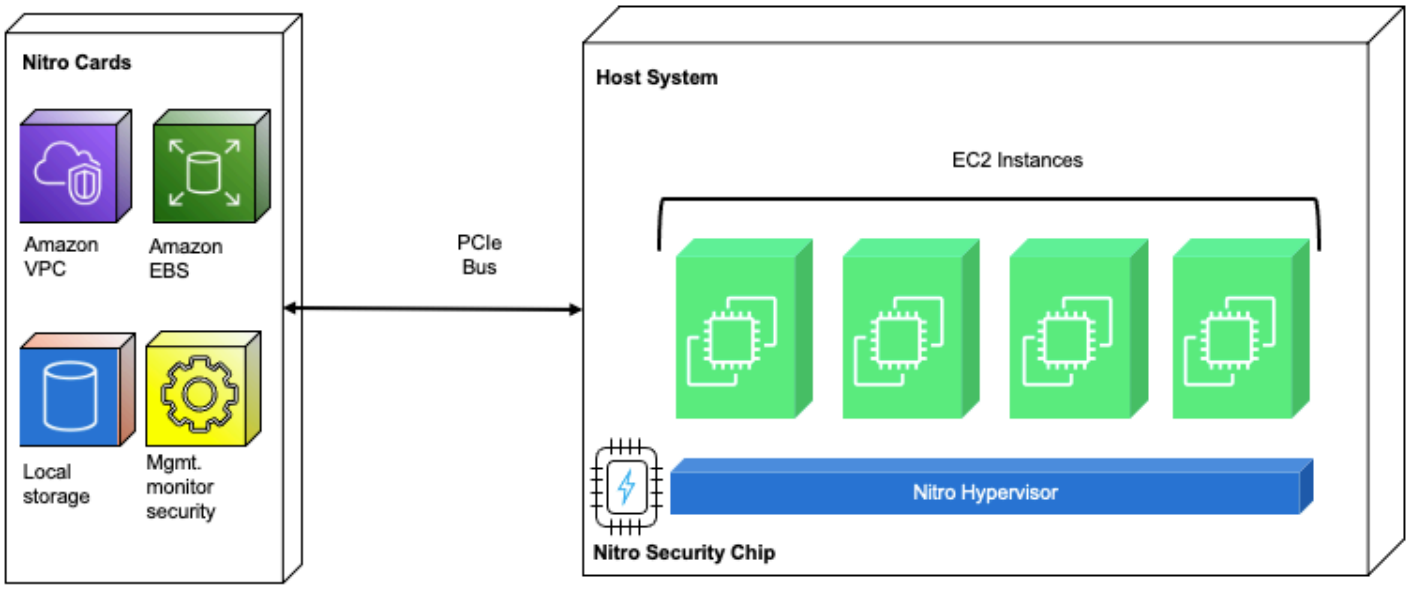
AWS Nitro System is the underlying technology used for Amazon Elastic Compute Cloud (Amazon EC2) instances in RISE with SAP. AWS Nitro System offers a unique set of capabilities that support the most sensitive workloads in a multi-tenanted, hyper-scale cloud environment.

A traditional virtualization architecture consists of "hypervisor" or "Virtual Machine Monitor (VMM)" and what is commonly known as 'Dom0' in Xen project or "parent partition" in Hyper-V. More details on traditional virtualization architecture is available here.

In Nitro System virtualization architecture, the management or control domain components (with privileged access to the hardware and device drivers) are fragmented into independent purpose-

built service processor units (SoC - System on Chip) which are known as Nitro cards. While the "hypervisor" layer remains, the design has been minimized to include only those services and features which are strictly necessary for its task. Additionally, there is also a "Nitro Security Chip" introduced to enhance the security while ensuring there is no overhead on performance.

Below is the Nitro High Level Architecture



The resulting Nitro System has been divided into the following components:

Nitro Cards

Nitro Controller - This is the sole outward facing management interface between the physical server and the control planes for EC2, Amazon EBS, and Amazon VPC. It is implemented as passive API endpoints where each action is logged and all attempts to call an API are cryptographically authenticated and authorized using a fine-grained access control model.Nitro Controller also provides the hardware root of trust for the overall system and is responsible for managing all other components of the server system including the firmware loaded in the system. Firmware for the system is stored on an encrypted SSD that is attached directly to the Nitro Controller. The encryption key for the SSD is designed to be protected by the combination of a Trusted Platform Module (TPM) and the secure boot features of the SoC.Nitro Cards purpose built for specific functions Nitro Cards purpose built for specific functions:

Networking - The newer generation of Nitro cards for VPC transparently encrypt all VPC traffic to other EC2 instances running on hosts also equipped with encryption compatible Nitro Cards. It uses Authenticated Encryption with Associated Data (AEAD) algorithms, with 256-bit encryption. In

RISE with SAP, depending on customer's requirements, different families of compute instances are selected. While AWS provides secure and private connectivity between EC2 instances of all types, in-transit traffic encryption is available between the later generation instances only. Please check whether your RISE with SAP instances are supported for this feature here.

EBS (SSD) storage - The Nitro Card for EBS provide encryption of remote EBS volumes without any practical impact on their performance.

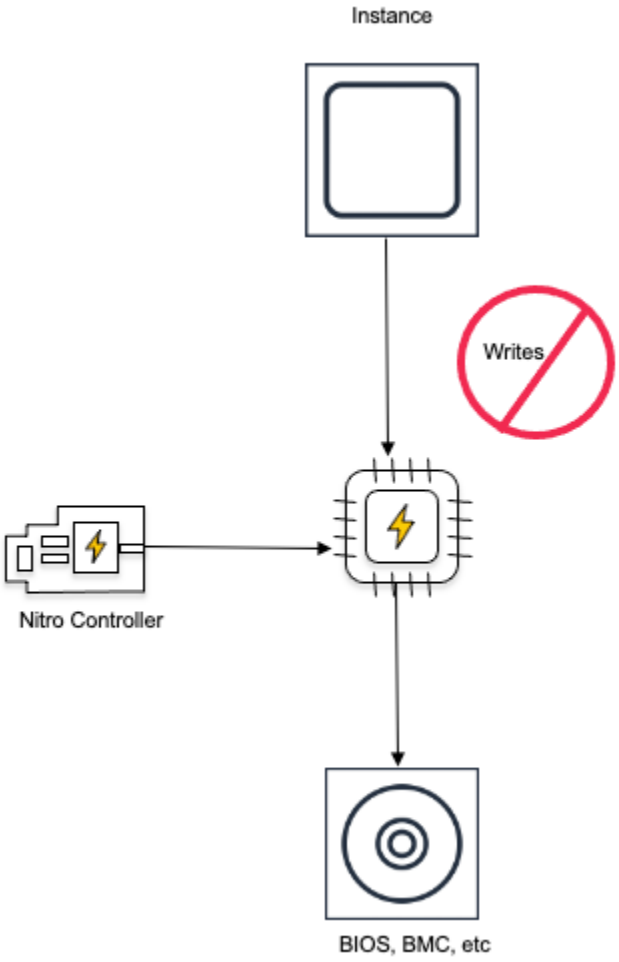
Local instance storage (ephemeral) – Similar to Nitro Card for EBS, the Nitro Card for instance storage provides encryption to local instance storage. All EC2 instances do not have local instance storage and this would depend on the instance types chosen for your RISE with SAP workloads. Details can be found here.

The encryption keys used for VPC, EBS and Instance Storage are only ever present on the system in plaintext within the protected memory of a Nitro Card.

Nitro Security Chip

While the Nitro Controller and other Nitro Cards operate as one domain, the system main board on which SAP workloads runs make up the second domain. While the Nitro Controller and its secure boot process provide the hardware root of trust between the Nitro System components, Nitro Security chip is used to extend that trust and control over the system main board. The Nitro Security Chip is the link between those two domains that extends the control of the Nitro Controller to the system main board, making it a subordinate component of the system, thus extending the Nitro Controller chain of trust to cover it. To maintain the root of trust, all write access to non-volatile storage is blocked in hardware.

Below is when Nitro blocked write access to non-volatile storage



Nitro Hypervisor

Unlike traditional hypervisors, Nitro Hypervisor is not a general-purpose system and does not have a shell nor any type of interactive access mode. Some of the key exclusions in the Nitro Hypervisor which enhances its security posture are networking stack, general purpose file system implementations, peripheral driver support, ssh server, shell etc. Primary functions of Nitro Hypervisor are restricted to:

1. Receive virtual machine management commands (start, stop and so on) sent from Nitro Controller
2. Partition memory and CPU resources by utilizing hardware virtualization features of the server processor
3. Assign SR-IOV virtual functions provided by Nitro hardware interfaces (NVMe block storage for EBS and instance storage, Elastic Network Adapter [ENA] for network, and so on) through PCIe to the appropriate VM

This simplicity of the Nitro Hypervisor is a significant security benefit compared to conventional hypervisors.

Key Benefits of AWS Nitro System

- Nitro chips offload virtualization tasks from the main CPUs, reducing the attack surface and improving overall system security.
- AWS personnel do not have access to Your Content on AWS Nitro System EC2 instances. There are no technical means or APIs available to AWS personnel to access you content on an AWS Nitro System EC2 instance or encrypted-EBS volume attached to an AWS Nitro System EC2 instance. Access to AWS Nitro System EC2 instance APIs – which enable AWS personnel to operate the system without access to your content - is always logged, and requires authentication and authorization. Please find more information here.
- Tenancy protection and prevention of side channel attacks - The Nitro Hypervisor is directed by the Nitro Controller to allocate the full complement of physical cores and memory for the instance. These hardware resources are "pinned" to that particular instance. The CPU cores are not used to run other customer workloads, nor are any instance memory pages shared in any fashion across instances. No sharing of CPU cores means that instances never share CPU core-specific resources, including Level 1 or Level 2 caches thereby providing strong mitigation against side channel attacks. Please find more information here.
- The Nitro architecture allows for secure boot and runtime integrity verification, ensuring the AWS infrastructure is running in a trusted and verified state.
- Both the Nitro Card firmware and the hypervisor are designed to be live-updatable (zero downtime for customer instances). This eliminates the need for carefully balanced tradeoffs around updates yielding improved security posture. Please find more information here.
- Data encryption for both data at rest and in transit using hardware offload engines with secure key storage integrated in the SoC.

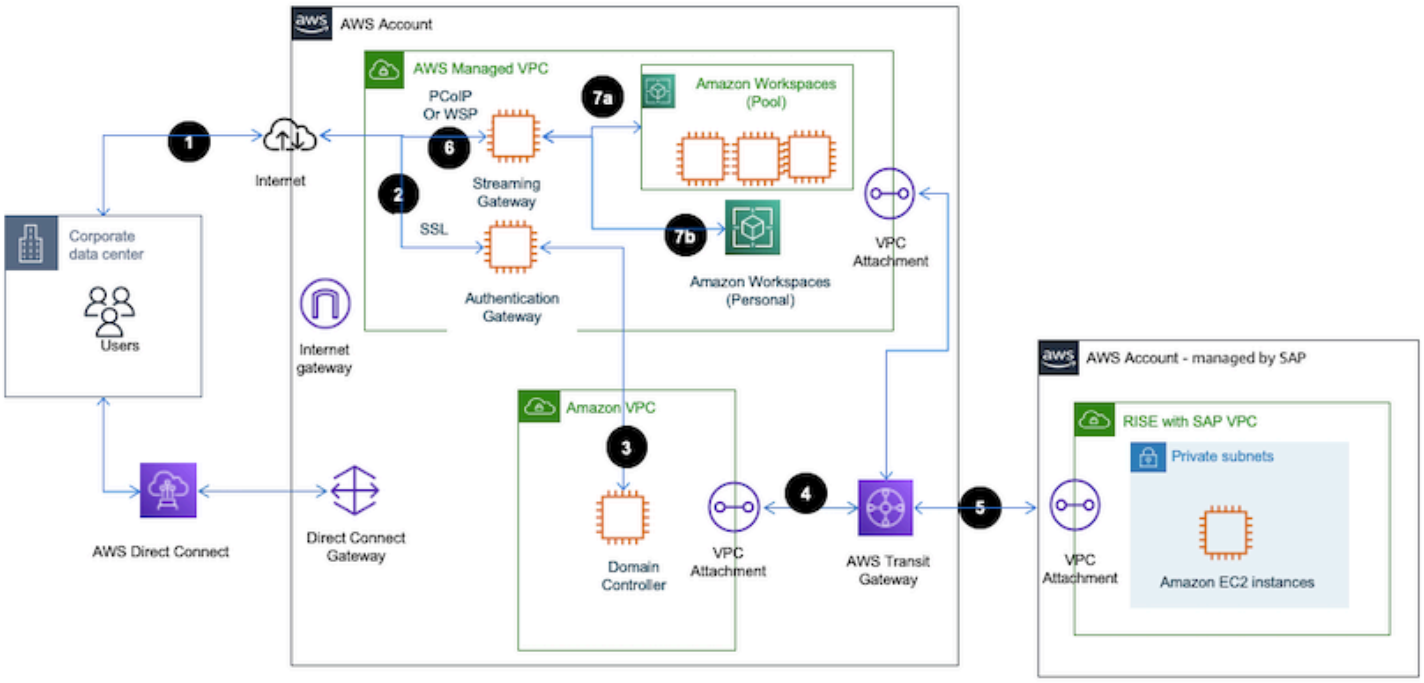
Amazon WorkSpaces as remote access solution for RISE with SAP

Using Amazon WorkSpaces provides a secure, scalable, and managed virtual desktop environment for accessing SAP systems. This virtual desktop can be used as a centrally managed hosting platform for SAP end user software such as SAPGUI and be connected to your SAP S/4HANA environment in RISE with SAP.

Amazon WorkSpaces Personal offers persistent virtual desktops, tailored for users who need a highly-personalized desktop provisioned for their exclusive use, similar to a physical desktop computer assigned to an individual.

Amazon WorkSpaces Pool offers non-persistent virtual desktops, tailored for users who need access to highly-curated desktop environments hosted on ephemeral infrastructure.

The following image shows the use of Amazon WorkSpaces as remote access solution for RISE with SAP.



Traffic flow

1. User initiates a connection to the AWS WorkSpaces URL via a Web browser or WorkSpaces Client.
2. User authenticated through the authentication gateway within the AWS Managed VPC. When an end-user logs in, the Authentication Gateway verifies user against Directory Services and once the user is authenticated, the gateway establishes a secure session for the user to access their virtual desktops. This session management ensures that the user's WorkSpaces remains accessible during their active session and helps maintain session integrity and security. This part of architecture uses Secure Socket Layer (SSL) with TCP protocol on port 443.
3. The connection is routed through another VPC Attachment to reach the Domain Controller in a separate Amazon VPC. The Domain Controller manages permissions and access control policies for users. It ensures that users have the appropriate access to resources based on their roles and

group memberships. This is typically done through integration (such as AWS Managed Microsoft AD or an on-premises AD connected via AWS Directory Service)

4. Transit Gateway manages the routing between VPCs and Direct Connect or VPN. AWS Direct Connect or VPN provides a secure connection from AWS to the SAP RISE environment.
5. A secure session is established between the user's device and the SAP managed RISE VPC.
6. The streaming service gateway within the AWS managed VPC begins to stream the virtual desktop environment to the user's device. This streaming is secured and managed within AWS infrastructure. The streaming gateway securely transmits the desktop stream over the internet to the user's device. The user's device now can access SAP applications like SAP S/4 hosted in the RISE environment through SAP end user software such as SAPGUI.
7. Amazon WorkSpaces allows you to access the following 2 types of WorkSpaces, depending on your organization and user needs

   WorkSpaces Pool, in a pooled configuration, WorkSpaces are dynamically assigned to users from a shared pool. When a user logs in, they may not always connect to the same machine, and changes such as installed applications or user configurations are generally not persistent between sessions

   WorkSpaces Personal, in this configuration, each user is assigned their own dedicated virtual desktop, where they can install applications, save files, and have their settings and data persist between sessions.

Set up Amazon WorkSpaces for SAP RISE Access

1. To use or setup Amazon WorkSpaces to connect to SAP RISE, follow the Get started with WorkSpaces.
2. For more information about integrating Amazon WorkSpaces with SAP Single-sign-on, see How to integrate Amazon WorkSpaces with SAP Single Sign-On
3. Install SAPGUI on your WorkSpaces from SAP Software download
4. Connect to SAP system via the SAPGUI client in WorkSpaces using your SAP System details

Amazon Workspaces Operational Best Practices

1. Monitoring: Use AWS CloudWatch to monitor the performance and health of your WorkSpaces.
2. Backup and Recovery: Ensure that critical data on your WorkSpaces is backed up and that you have a recovery plan in place.

3. Updates and Maintenance: Regularly update the software and systems on your WorkSpaces to ensure security and compliance. By default, Windows WorkSpaces will automatically update weekly.
4. Optimizing Performance

   Scaling and Performance Tuning: You can switch a WorkSpaces between the Standard, Power, Performance, and compute types dependent on user needs.
5. Cost Management

   WorkSpaces Bundles: Consider purchasing virtual desktop bundles inclusive of your end user software needs. Generally, for simple SAPGUI access a "Value" user will save on costs. See the AWS WorkSpaces Pricing page for further details

   Monitoring Usage: Use AWS Cost Explorer and budgets to monitor and manage costs effectively.

   For non-persistent, secure desktop access consider WorkSpaces Pools as a highly cost-effective option.

By following these steps, you can set up Amazon WorkSpaces as an effective remote access solution for RISE with SAP systems, ensuring secure, scalable, and efficient operations.

WorkSpaces Benefits to RISE

Using Amazon WorkSpaces as a remote access solution in a RISE with SAP deployment offers several benefits, particularly around security, access control, and operational efficiency. Here are the key benefits of this approach:

1. **Enhanced Security and Controlled Access**

   Isolated Environment: WorkSpaces provide an isolated environment where access to SAP systems in a RISE deployment can be tightly controlled. This helps prevent unauthorized direct access to critical systems

   No Direct Internet Exposure: By using WorkSpaces as a remote access solution, you can restrict internet access to the SAP environment. External users or administrators must first connect to a secure WorkSpaces, limiting exposure to SAP systems.

   Secure Protocols (PCoIP/WSP): WorkSpaces use secure streaming protocols like PCoIP or WSP, ensuring that data is encrypted during transmission.

Reduced Attack Surface: By utilizing WorkSpaces as the only point of access to SAP systems, you can reduce the attack surface by isolating SAP environments from direct access over the internet or corporate networks.

VPC Integration: WorkSpaces can be deployed in private subnets within an Amazon Virtual Private Cloud (VPC), ensuring secure and direct connectivity to the RISE with SAP infrastructure.

AWS Direct Connect or VPN: You can use AWS Direct Connect or VPN connections to provide a secure network path between the WorkSpaces and SAP environments, further enhancing security.

2. **Centralized Management**

   Unified Access Point: Amazon WorkSpaces serve as a single point of access to manage and operate the RISE with SAP environments, simplifying monitoring and control.

   Audit and Logging: AWS services such as AWS CloudTrail and Amazon CloudWatch can log user actions and monitor activities on the WorkSpaces. This helps with security audits and tracking access to SAP systems.

   Integration with AWS IAM: Role-based access control (RBAC) through AWS Identity and Access Management (IAM) ensures fine-grained access to WorkSpaces and SAP resources. This minimizes the risk of unauthorized access and supports compliance requirements.

3. **Improved Operational Efficiency:**

   On-Demand Scalability: WorkSpaces can be provisioned quickly and scaled on-demand, making it easy to provide access to administrators or developers needing to access the SAP environment without lengthy setup processes.

   Minimal Maintenance: Amazon WorkSpaces are fully managed, which reduces the overhead of maintaining physical servers or traditional remote desktop infrastructure. Updates and patches are handled by AWS, freeing up time for more critical operations.

   Cost Efficiency: WorkSpaces can be configured to charge only when in use (hourly pricing), making it a cost-effective solution for temporary or infrequent access, especially when not in continuous operation.

Remote Access: With WorkSpaces, administrators and users can access the SAP environment securely from any location with an internet connection. This is particularly useful for distributed teams or remote workers supporting SAP environments.

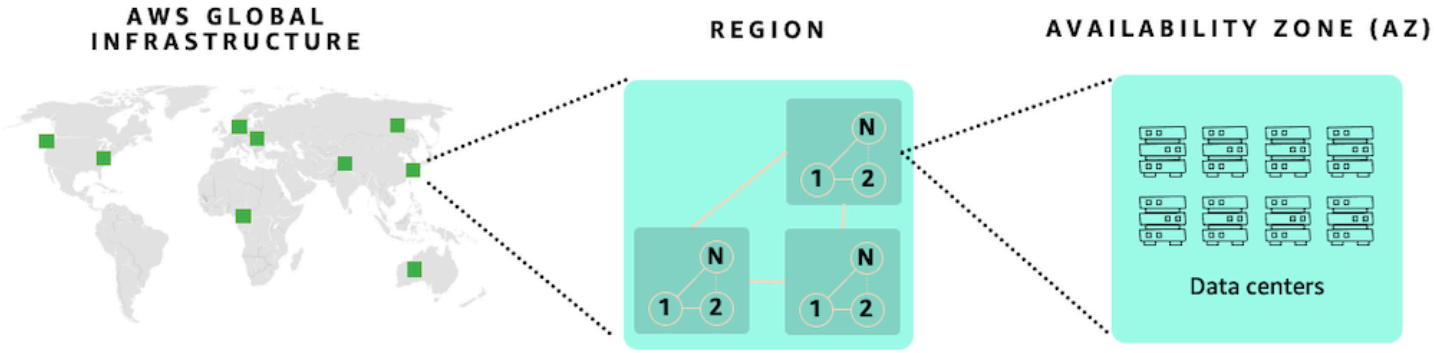
Resilience and Availability: WorkSpaces can be integrated with AWS backup solutions and spread across multiple AWS Availability Zones (AZs), ensuring redundancy and high availability.

Quick Recovery: In case of failure or disaster in the SAP environment, WorkSpaces provide a quick and scalable way to reconnect to alternative environments or backup systems.

Reliability

Reliability is one of the six pillars of SAP Lens - AWS Well-Architected Framework. For more information, see Reliability.

AWS cloud offers reliability with multiple Availability Zones within an AWS Region. This enables your SAP applications on AWS to be more resilient. Each Region is further isolated from other Regions, providing the greatest possible fault tolerance and stability. Within each AWS Region, there are a minimum of three, isolated, physically separate Availability Zones. For more information, see Regions and Availability Zones.

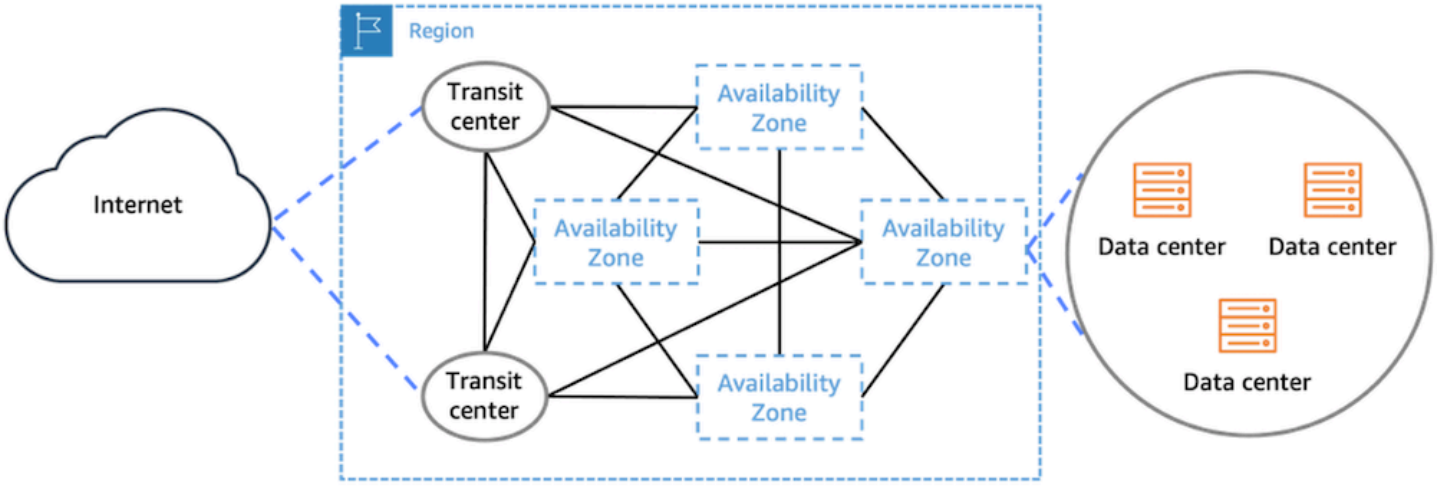


Availability Zones enable you to operate production applications and databases that are more highly available than would be possible from a single data center. Distributing your applications across multiple Availability Zones provides the ability to remain resilient in the face of most failure modes, including natural disasters or system failures.

Each Availability Zone can be multiple data centers. At full scale, it can contain hundreds of thousands of servers. They are fully isolated partitions of AWS Global Infrastructure. An Availability Zone is physically separated from any other zones with its own separate power and networking

resources. There is a distance of several kilometers, although all are within 100 km (60 miles of each other). This distance provides isolation from the most common disasters that could affect data centers, such as floods, fire, severe storms, earthquakes, etc.

All Availability Zones within a Region are interconnected with high-bandwidth and low-latency networking, over fully redundant and dedicated metro fiber. This ensures high-throughput, low-latency networking between Availability Zones. The network performance is sufficient to accomplish synchronous replication.



Availability Zones enable you to run your applications in a highly-available manner, with synchronous data replication and automated failover between Availability Zones. RISE with SAP can offer such high available designs for your workload in every AWS Region.

Resiliency and Cost Considerations

SAP has options available for RISE to meet different resiliency requirements. The following key requirements are adjustable for RISE via option packages available from SAP.

- Service Level Agreement (SLA) – describes the targeted availability of the solution.
- Recovery Time Objective (RTO) – describes the targeted duration within which a recovery from a disaster event should be completed.
- Recovery Point Objective (RPO) – describes the targeted level of data loss that may occur during recovery from a disaster event.

For more details, refer to the definitions provided by SAP as part of RISE agreement for specific definitions, clauses, impacts, and penalties in the event of a breach.

The impact of an outage on your organisation and loss of data can cause loss of productivity, loss of income, and can damage reputation. Weighing the trade-off between cost and resiliency can help assess the risk to your organisation.

Resiliency and Performance Considerations

When you opt for short distance disaster recovery option in RISE, the SAP application servers and database servers will be installed across multi Availability Zones. This architecture supports highly available design for your SAP workload.

While using the application servers in multiple Availability Zones in an active-active configuration, it increases the resiliency. In parallel, a higher latency cross Availability Zones from application server to database server is introduced. You can refer to SAP Note 3496343 (Network Latency on AWS) that address in details the increased latency due to the distance between application servers and database servers in multi Availability Zones deployment. This will be discussed in details in the subsequent section.

- Network latency between the SAP application server and database server should be less than 0.7 milliseconds as per SAP Note 1100926
- Network latency for HANA system replication with synchronous data replication (which is required to achieve zero data loss) to be less than 1 millisecond

You can use the AWS Network Manager – Infrastructure Performance tool to automatically measure Inter-AZ, Intra-AZ and Inter-Region network latency. Alternatively, you can use SAP's NIPING tool as per SAP Note 2986631.

When SAP application servers and database servers distributed across multiple Availability Zones (AZs), it significantly enhances system reliability and availability, outweighing the impact of increased network latency.

Cross Availability Zones traffic may increase the time required to perform certain transactions or batch jobs that make frequent calls to the database. In case the impact is high, we recommend keeping this traffic within the same Availability Zone using SAP Logon Groups, RFC Server Groups and Batch Server Groupsulink>. This ensure that the impacted transactions or batch jobs only use application servers in the same availability zone as the database servers.

To automate and optimise the running of such performance-critical batch jobs and transactions on application servers located in the same Availability Zone as the database server, AWS provides example ABAP code which customers can test and implement in their SAP systems.

You may implement further optimization through C-State parameters by referring to AWS re:Post article Inter-AZ Latency for SAP to lower the network latency.

When it is not feasible to run application servers in active-active mode across multi Availability Zones, you can run in active-passive mode by utilizing ABAPSetServerInactive (SAP Note 3075829)

In rare cases, where you observe performance impacts due to latency within one Availability Zone, you may use Cluster Placement Groups to achieve lowest possible latency. You can refer to the Placement Strategies Guide from AWS.

In summary, these are the architecture patterns in multi Availability Zones deployment:

App Servers in AZ1	App Servers in AZ2	Failover Mechanism from AZ1 to AZ2
Active	Active	Automated script (i.e. pacemaker)
Active	Active	Manual adjustment of Logon Groups, RFC and Batch Server Groups
Active	Active	Automatic script to adjust Logon Groups, RFC and Batch Server Groups
Active	Passive	Manual activation of the passive application servers
Active	Passive	Automatic script to activate the passive application servers

To achieve high reliability of SAP workloads, We recommend the following tasks:

1. Discuss with SAP on the Availability SLA requirement for RISE deployment. This will drive the components (i.e. database and application servers) that will be deployed across multiple Availability Zones to maximise reliability and availability of RISE.

2. If you have business scenarios involving batch jobs and/or transactions that makes frequent calls to the database servers, it may be adversely impacted by inter-AZ network latency, you can consider using SAP's workload distribution mechanism (SAP Logon Groups, RFC Server Groups and Batch Server Groups) to ensure these jobs and transactions run on the application servers located in the same Availability Zone as the database server
3. You may implement further optimization of network latency by referring to AWS re:Post article Inter-AZ Latency for SAP.
4. When active-active mode is not feasible, you can run in active–passive mode of application servers utilizing ABAPSetServerInactive (SAP Note 3075829).
5. You can consider putting other workloads, that are outside of RISE, within the same Availability Zone in order to achieve better network latency and lower data transfer cost.

Disaster recovery options

You can implement a disaster recovery solution by replicating data into a second AWS Region. Your SAP workloads are protected in the event of rare occurrence of local or regional failures.

RISE with SAP S/4HANA Cloud, private edition offers the following two options.

- **Short distance disaster recovery** or Metro disaster recovery – RISE with SAP uses multiple Availability Zones in an AWS Region. Unique AWS region with three or more Availability Zones provide the option of short distance disaster recovery in every AWS regions.
- **Long distance disaster recovery** or Regional disaster recovery – RISE with SAP uses a secondary AWS Region as standby for failover systems. Owing to the physical distance between two AWS Regions, data is replicated asynchronously between two AWS Regions.

For more details, see SAP documentation SAP Service Description: Disaster Recovery and Customer Invoked Failover.

Extensions

You can extend RISE with SAP by using AWS services to improve performance, security, agility, and reduce costs. The following table provides recommended AWS services based on use case.

Category	Use case	AWS services
Performance	SAP Fiori and SAP GUI access with proactive observability	Amazon CloudFront, Accelerated Site-to-Site VPN, AWS Internet Monitor
Data Integration	Data Extraction and Data Federation	Amazon Appflow, AWS Glue for SAP OData, AWS SDK for SAP ABAP, Amazon Athena
Data Analytics	Data Lakes and Data Warehouse	AWS Lake Formation, Amazon S3, Amazon QuickSight, Amazon Redshift
Application integration	Application Integration	AWS Lambda and Amazon API Gateway
Archiving and Document Management	Archiving and Document Management	Amazon S3, AWS S3 File Gateway, Amazon EFS
Development and Extension	Development, Compatibility packs and alternatives	AWS SDK for SAP ABAP, AWS Marketplace
Security Extension	Single Sign On, Zero Trust Access	mTLS Authentication through Amazon ALB, AWS Verified Access for SAP
Artificial Intelligence	Generative AI	Amazon Q for Business, Amazon Q in QuickSight, Amazon Bedrock

Performance

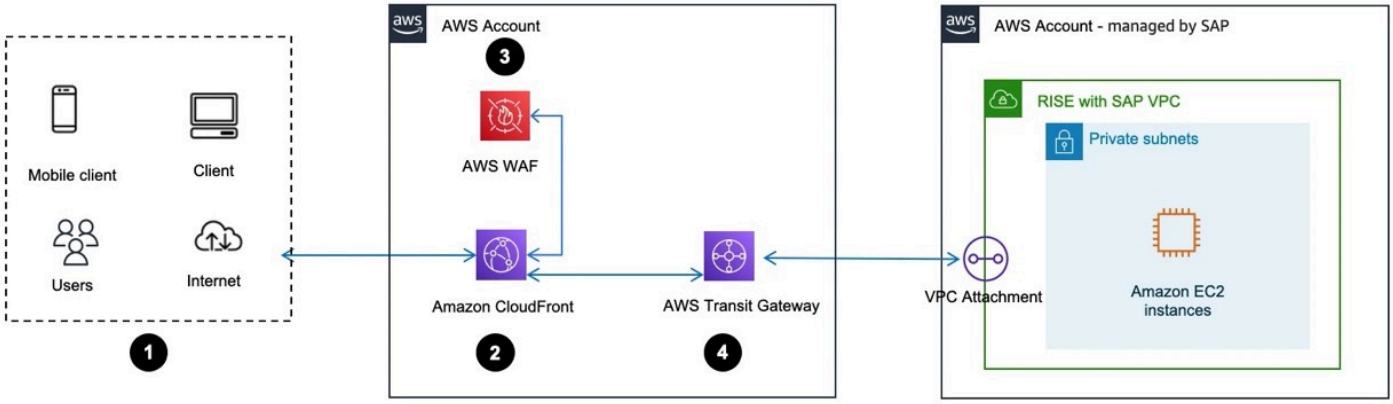
Enhance SAP Fiori performance with Amazon CloudFront

Amazon CloudFront is a Content Delivery Network service to increase performance and reduce latency of SAP Fiori launchpad in RISE with SAP. CloudFront creates a cache for the static content and accelerates dynamic content through edge computing.

Global SAP systems accessed by users from across multiple geographical regions, can use Amazon CloudFront VPC (Virtual Private Cloud) Origins to reduce network latency and improve the SAP end-user experience.

CloudFront VPC Origins is a feature that enhances security and streamlines operations for web applications such as SAP Fiori, hosted in private subnets within the Amazon VPC. This architecture allows CloudFront to serve as the single entry point for SAP Fiori, eliminating the need for public exposure of the SAP servers.

CloudFront VPC Origins is deployed in the customer-managed AWS account, directing SAP users coming through the CloudFront to an internal, AWS Application Load Balancer (ALB). The ALB routes Fiori traffic directly to the SAP systems hosted in the SAP RISE AWS account through the AWS Transit Gateway. The AWS Web Application Firewall (WAF) is optional but recommended to improve security posture.



Data flow

1. User accesses SAP Fiori launchpad via Internet browser or mobile device
2. The request is routed to Amazon CloudFront to the closest edge compute of the user location
3. Optionally, AWS Web Application Firewall (WAF) evaluates the request based on the customer's configured rules to block malicious traffic. Additionally, Distributed Denial of Service (DDOS) protection is also provided by AWS Shield Standard which is automatically included at no extra cost when you use CloudFront with AWS WAF
4. The request is then parsed to the AWS ALB which forwards the traffic to the SAP system hosted in the SAP managed RISE account.

This improves the security posture of SAP systems by:

- Eliminating direct exposure of SAP servers to the public internet
- Reducing the attack surface as CloudFront becomes the only ingress point
- Simplified security management with centralized control through CloudFront
- Easy integration with AWS WAF & AWS Shield Standard for additional protection

Integrating CloudFront VPC Origins with SAP can lead to performance improvements:

- Global users benefit from CloudFront's worldwide edge locations
- Traffic is optimized using the AWS global network backbone. CloudFront traffic stays on the high-throughput AWS global network backbone all the way to your SAP servers, providing optimized performance and low latency
- Static SAP Fiori content is cached at CloudFront edge locations and dynamic SAP Fiori content is accelerated through CloudFront's global edge network

To implement CloudFront VPC Origins for SAP:

1. The applications in RISE with SAP are by default hosted in private VPC subnets, in an AWS account – managed by SAP
2. In the AWS account – managed by customer, create an AWS ALB pointing to the SAP system in the RISE account
3. Create a CloudFront distribution with VPC Origins pointing to the AWS ALB
4. Update the security group for your VPC private origin (AWS ALB in this case) to explicitly allow the CloudFront managed prefix list. This restricts traffic coming to the VPC origin
5. Ensure the same fully qualified domain name is used by CloudFront, ALB, and SAP
6. Configure CloudFront to handle both static and dynamic content from SAP systems
7. Optionally, implement AWS WAF for additional security at the edge

Refer to AWS documentation Restrict access with VPC origins for more information.

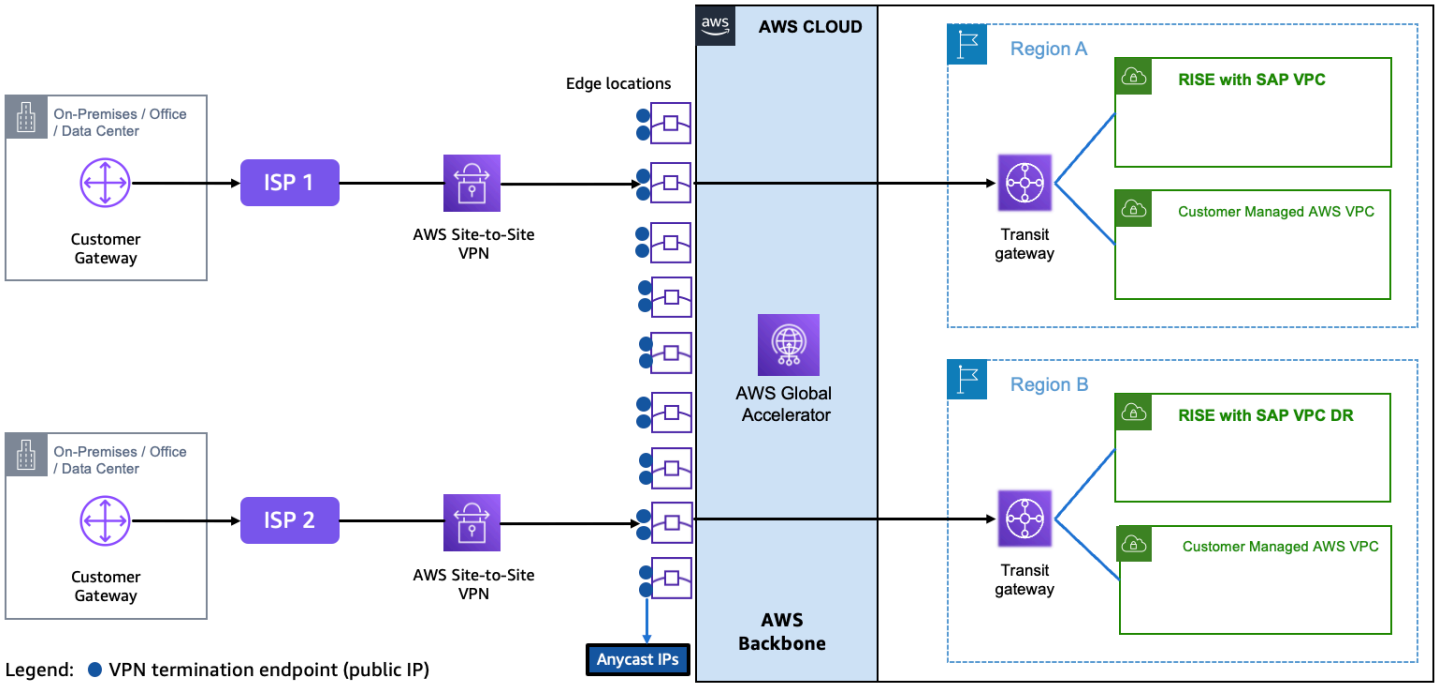
Optimize performance with Accelerated Site-to-Site VPN connections

When you deploy RISE with SAP on AWS for a global roll-out, you can reduce the network latency by leveraging AWS Global Accelerator based Accelerated Site-to-Site VPN. This service complements the foundational Transit Gateway and Direct Connect to address performance challenges for geographically dispersed users while ensuring efficient and secure access to mission-

critical RISE with SAP. It supports both SAP Fiori (HTTPs based) traffic and SAP GUI (TCP based) traffic.

AWS Global Accelerator is a service which create accelerators to improve the performance of applications for local and global users. It operates as a Layer 4 TCP/UDP proxy, optimizing traffic routing through AWS's global network infrastructure. It terminates client TCP connections at AWS edge locations and establishes new TCP connections to backend endpoints over AWS's private backbone. Thus, reduces latency (up to 75% varying by locations) by bypassing public internet hops and ensures congestion-free routing for globally distributed users.

Accelerated Site-to-Site VPN connections combines traditional AWS Site-to-Site VPN with AWS Global Accelerator to optimize traffic routing. It routes the traffic from on-premises network to an AWS edge location that is closest to customer gateway device, leveraging the AWS backbone. This will reduce latency by up to ~30%-60% compared to standard VPNs.



Enhancing observability of RISE with SAP using AWS Internet Monitor

AWS Internet Monitor continuously analyses internet traffic between end users and AWS-hosted applications, detecting network anomalies that may impact RISE with SAP performance. It provides insights into issues like increased latency, packet loss, or regional connectivity disruptions, allowing organizations to proactively address potential outages before they affect SAP workloads.

RISE with SAP relies on stable and predictable network performance, AWS Internet Monitor helps by:

- Identifying ISP or regional network disruptions that impact SAP response times.
- Providing early warnings and actionable recommendations to mitigate network-related service degradation.
- Distinguishing between AWS infrastructure issues and external internet disruptions and streamlining troubleshooting.
- Improving observability of Internet routing, which is dynamic and lacks predictable service-level agreements (SLAs).
- Proactive management of external ISPs and transit providers which may introduce unpredictable latency, packet loss, and congestion issues.

To implement you can refer to the Getting started with Internet Monitor.

Data integration

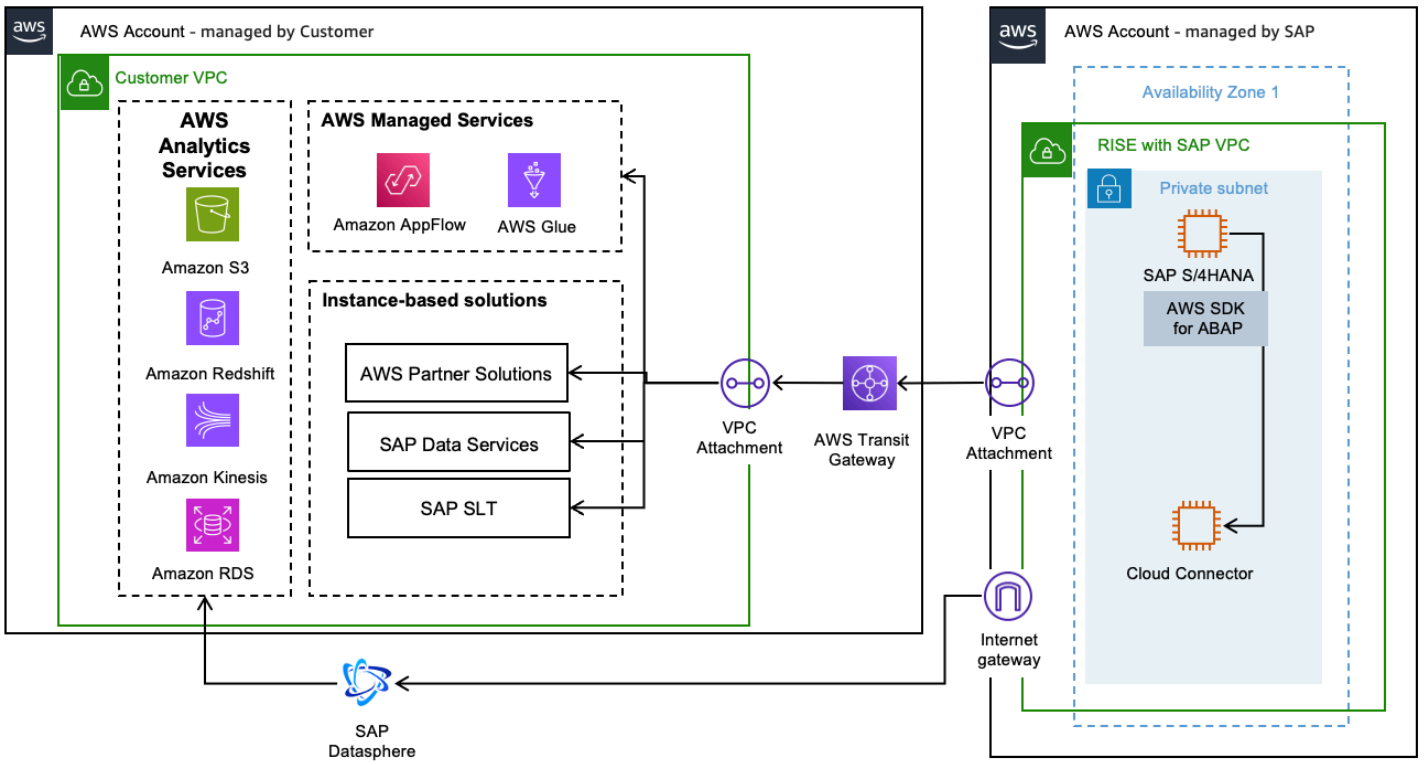
RISE with SAP Extensibility for Data Integration with AWS is a technical framework that enables data flow between SAP systems, AWS services, and third-party solutions. This integration architecture provides standardized APIs, connectors, and protocols to establish secure communication channels, addressing the critical need for seamless enterprise data integration in modern cloud environments

The RISE with SAP Extensibility for Data Integration outlines two primary data handling and integration mechanisms:

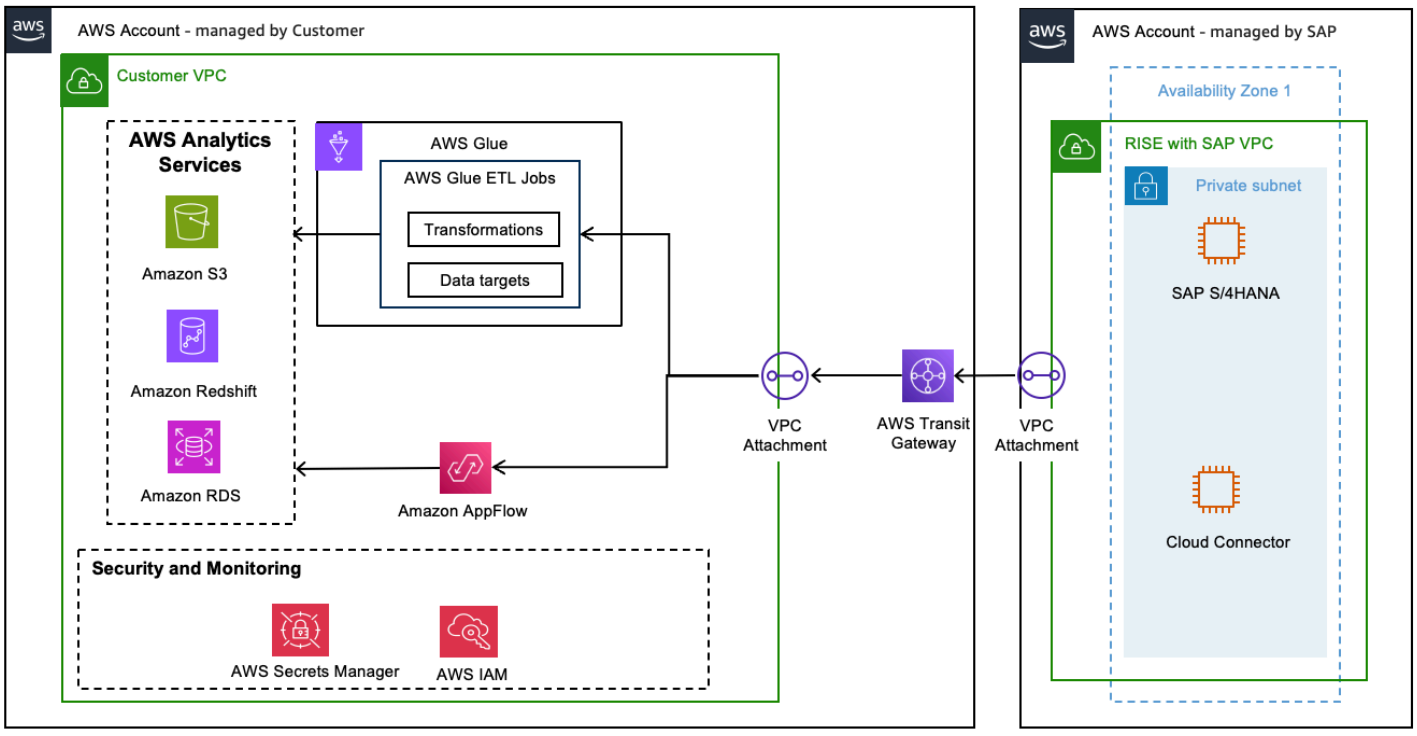
1. Data Extraction
2. Data Federation

Data Extraction

Data extraction from SAP is a crucial step in making the data usable for reporting, analysis, and integration with other systems.



Extracting data from SAP using AWS Managed Services



Amazon AppFlow

The Amazon AppFlow SAP OData connector provides the ability to fetch, create, and update records exposed by SAP S/4HANA and SAP on premises systems through OData APIs. When you connect Amazon AppFlow to ODP providers, you can create flows that run full data transfers or incremental updates. Incremental updates for ODP data are efficient because they transfer only those records that changed since the prior flow run.

The SAP OData connector for Amazon AppFlow user guide provides detailed instructions for setting up AppFlow, covering everything from establishing connectivity to transferring data from SAP.

AWS Glue

AWS Glue is a serverless data integration service that makes it easy for analytics users to discover, prepare, move, and integrate data from multiple sources. With AWS Glue, you can discover and connect to SAP using OData and manage your data in a centralized data catalog. You can visually create, run, and monitor extract, transform, and load (ETL) pipelines to load SAP data into your data lakes and data warehouses.

The Connecting to SAP OData using Glue user guide offers comprehensive instructions for setting up Glue ETL jobs, configuring SAP OData connections, and reading data from SAP, including handling incremental transfers. Note: Key differences of Amazon AppFlow vs AWS Glue for SAP OData

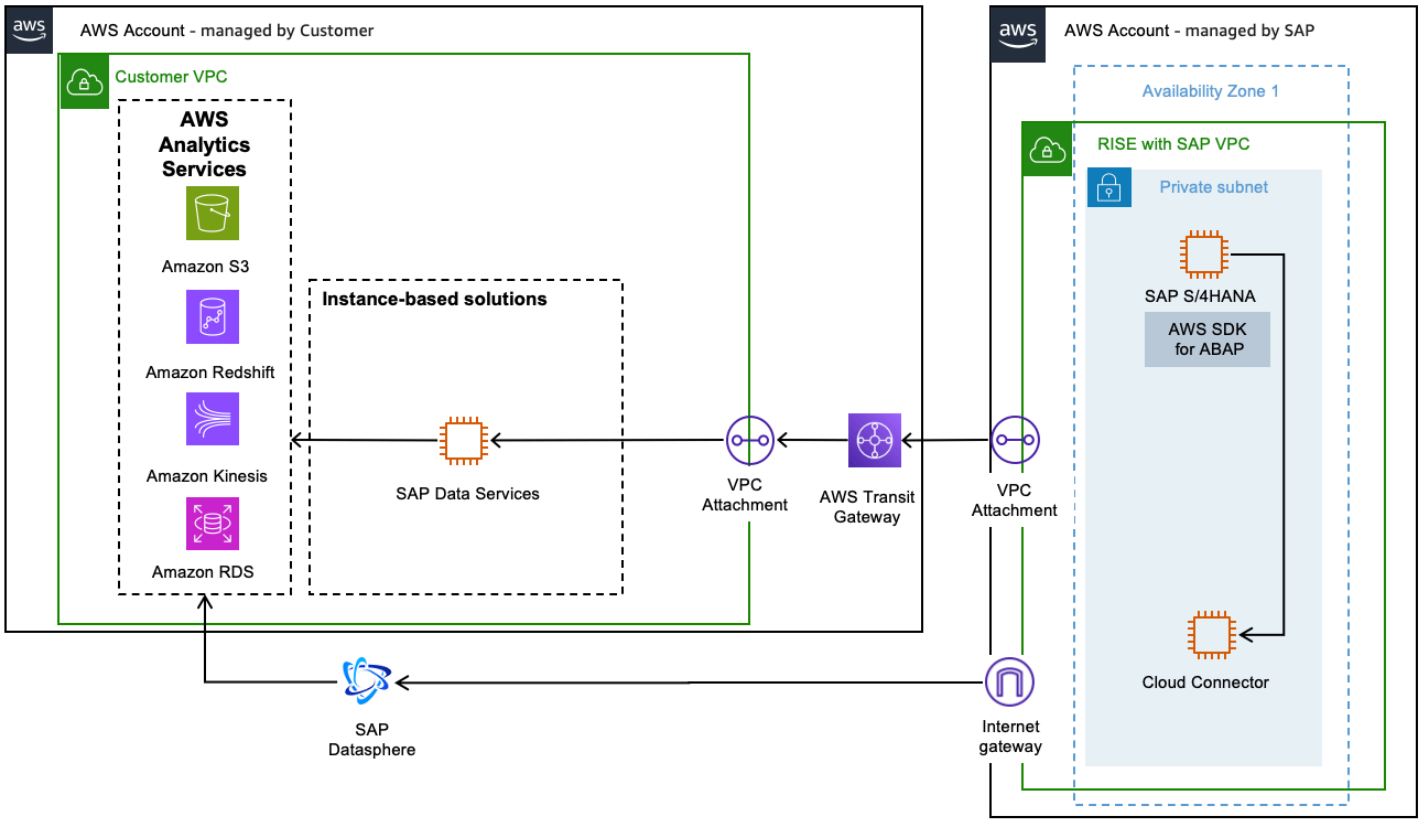
Amazon AppFlow and AWS Glue serve distinct roles in data integration, with each offering unique advantages for different use cases. While AWS Glue excels in complex ETL operations, data discovery, preparation, and extraction, particularly for specialized scenarios like SAP ODP-based extraction. AppFlow is designed as a more streamlined, no-code solution for API-based data transfers.

AWS Glue requires more hands-on management, including code deployment and maintenance, but offers greater flexibility and control over data transformation processes. In contrast, AppFlow provides a user-friendly, point-and-click interface with pre-built connectors for popular SaaS applications, enabling bidirectional data transfers without requiring complex API implementations.

AWS SDK for ABAP

AWS SDK for ABAP simplifies the use of AWS services alongside SAP applications with a client library of modules that are consistent and familiar to ABAP developers. With its capability to consume AWS Services through ABAP code, it allows you to build ETL data pipelines and replication from SAP ABAP Stack to AWS destinations such as S3 buckets, Redshifts, and others.

Extracting data from SAP using SAP services like SAP Datasphere and SAP Data Services



SAP Datasphere

SAP Datasphere offers various connection types such as SAP ABAP Connections, SAP ECC Connections, SAP S/4HANA Cloud Connections supporting RFC and ODP protocols. Refer to SAP Datasphere documentation to choose most appropriate connectivity to extract SAP data. Using premium outbound integration for [Amazon Simple Storage Connection (Amazon S3), configure SAP Datasphere replication flow to ingest data to Amazon S3.

SAP Data Services

SAP Data Services offer various connections to extract data from SAP ECC data. Refer to SAP Data Services documentation to choose most appropriate connectivity. SAP Data Services offers Amazon Redshift Datastore and Amazon S3 datastore to ingest data to AWS. It also offers options for Amazon S3 file location protocol such as encryption type, compression type, batch-size, number of threads, Amazon S3 storage class, etc.

Extracting data from SAP using AWS Partner Solutions

AWS Partner Solutions offer ready to deploy solutions with enhanced features, such as pre-built connectors, specialized data pipelines, and advanced optimization techniques that reduce complexity and improve the speed of deployment.

To find and deploy a solution that fits your specific needs, you can explore the AWS Partner Solutions Finder or browse through the AWS Marketplace, where you can search for and quickly deploy partner solutions tailored to your unique SAP use case.

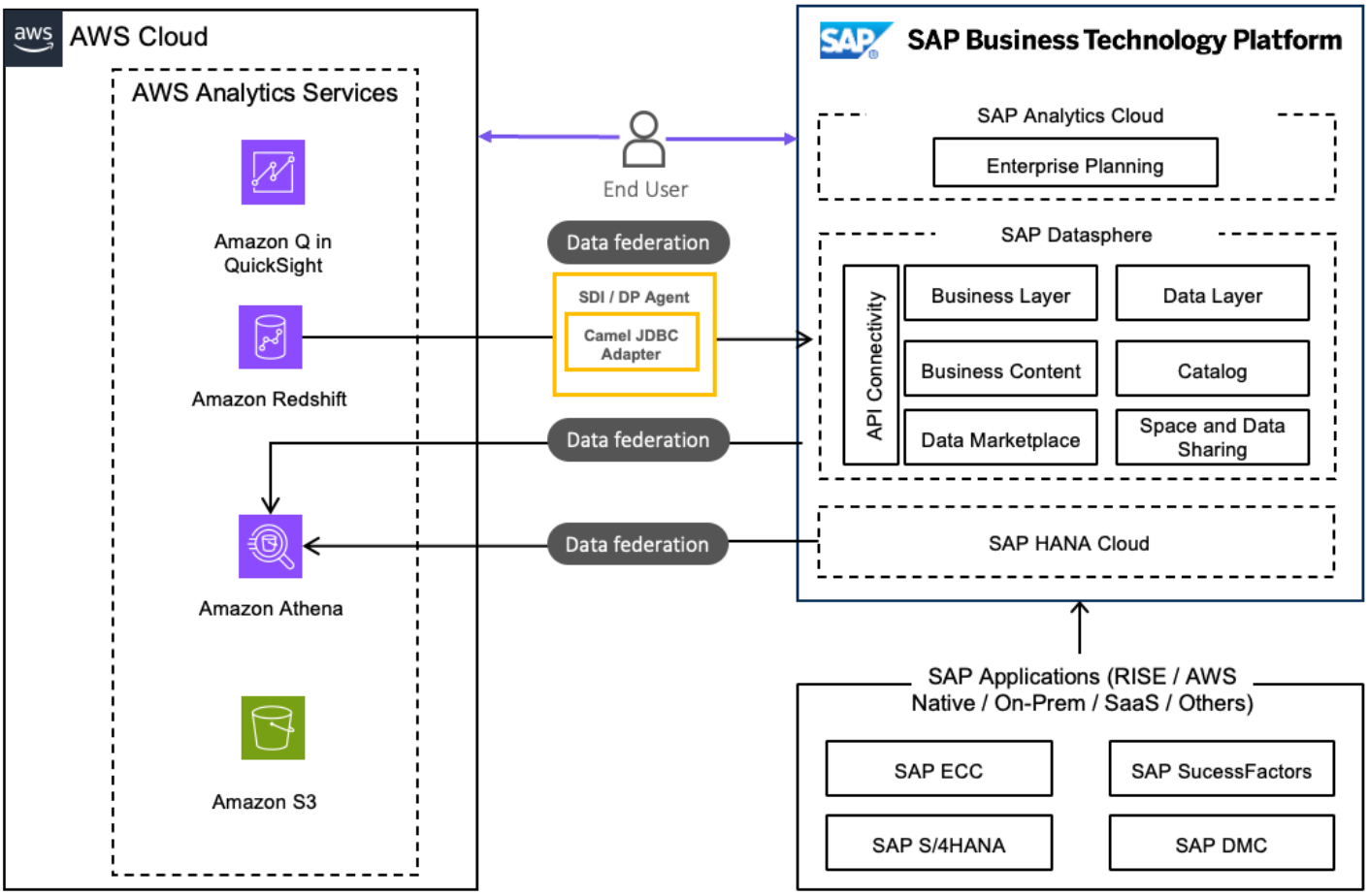
Further Resources

The Guidance for SAP Data Integration and Management on AWS provides the essential data foundation to build data and analytics solutions. It shows how to integrate data from SAP ERP source systems and AWS in real-time or batch mode, with change data capture, using AWS services, SAP products, and AWS Partner Solutions. This Guidance includes an overview reference architecture showing how to ingest SAP systems to AWS in addition to five detailed architectural patterns that complement SAP-supported mechanisms (such as OData, ODP, SLT, and BTP) using AWS services that are highlighted above, SAP products, and AWS Partner Solutions.

Data Federation

Data federation is a data management strategy that enables seamless integration and effective utilization of data from disparate data sources. When there is business requirement to have a consolidated data for transactional, analytics, machine learning, it is preferred for the data to be accessed from the source rather than replicated to avoid latency, inconsistency and extra storage cost.

In the context of SAP and AWS services, it allows organizations to access, combine, and analyze data from both SAP systems and AWS cloud services seamlessly.



Data federation using AWS Managed Services

Amazon Athena

Amazon Athena is a serverless, scalable and flexible interactive query service by AWS that allows to analyze data directly in Amazon S3. The data stored in Amazon S3 from multiple sources can be further transformed into tables and views using Amazon Athena and queried to extract meaningful information in a structured way.

Data in Athena can be accessed from SAP Datasphere through data federation from SAP Datasphere connections. Users can also access SAP Datasphere tables and views from Athena by querying SAP HANA using an Athena Federated Query.

Data can also be federated to the SAP HANA Cloud by configuring Athena as a remote source using the Smart Data Access – Athena adapter. The Athena Federated Query connection can also be used to read data from a stand-alone SAP HANA Cloud environment.

Amazon Redshift

Amazon Redshift is a fully managed, peta-byte scale data warehouse service from AWS. Customers have built their data warehouses and build data models for analytics and reporting.

Data federation from Amazon Redshift into SAP Datasphere is possible with SAP HANA Smart Data Integration (SDI) or the SAP Data Provisioning Agent. Amazon Redshift data can also be federated through the Athena Federated Query data source connector.

Further resources

The Guidance for Data Federation between SAP and AWS outlines the process of federating data between SAP and AWS cloud analytics services, enabling you to establish a data mesh architecture. By federating data between SAP and AWS. you can easily transform and visualize your data in a scalable, secure, and cost-effective way, helping you inform your decision-making.

Data analytics

SAP customers need business insights in real-time to react to business changes and leverage untapped business opportunities. This needs to be realized with modern, cloud native solutions to shift from overnight data processing to real-time analytics. Leveraging AWS and SAP solutions to build analytics solutions, customers can leverage built for purpose analytics services to gain competitive advantage in their respective industries.

Modern data architectures, such as Data Lakes and Data Warehouses, provide a combination of patterns and services that enable organisations to handle large volumes of structured and unstructured data for analysis and reporting, providing also a solid foundation for Artificial Intelligence (AI) and Machine Learning (ML) applications, including Generative AI. These architectures provide building blocks that can be implemented independently as well as complimenting each other, based on requirements and preferences.

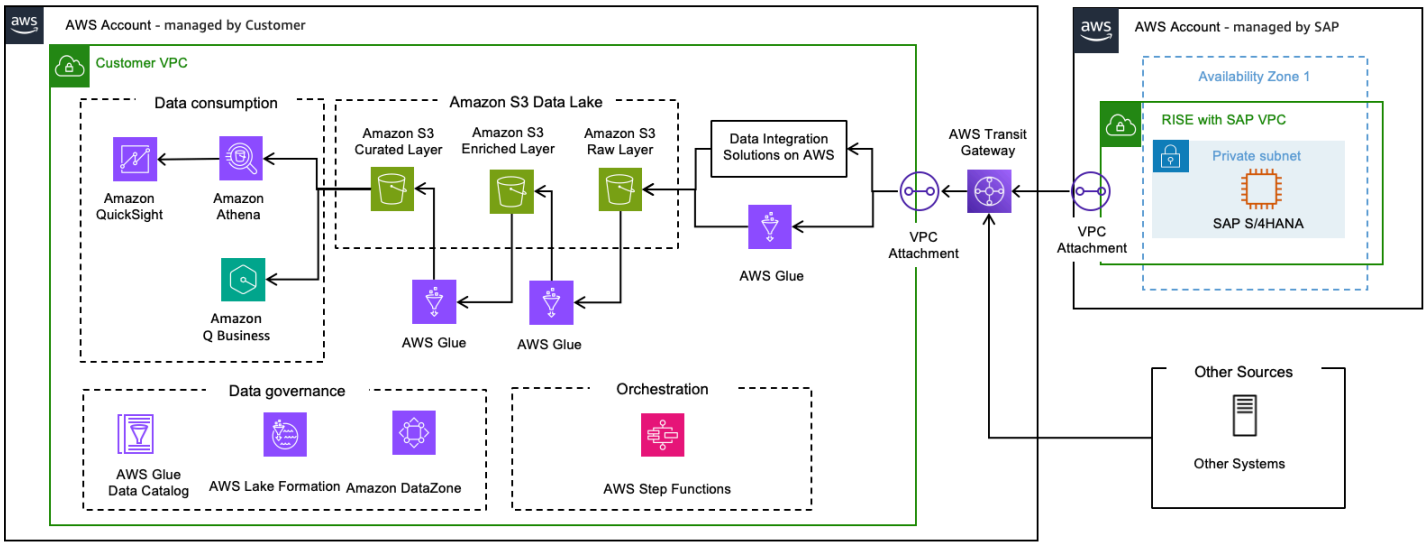
Data Lake Architecture

The Data lake architecture provides building blocks that demonstrate how to combine and consolidate SAP and non-SAP data from disparate sources using analytics and machine learning services on AWS.

Data lake enable customers to handle structured and unstructured data. They are designed based on a “schema-on-read” approach, meaning data can be stored in raw form and, only applies schema or structure upon consumption (i.e.: to create a Financial Report). The structure is defined when reading the data from the source, defining data types and lengths at that point. Due to this,

storage and compute is decoupled, leveraging low cost storage that can scale to petabyte sizes at a fragment of cost compared to traditional databases.

Data lake enables organizations to perform various analytical tasks like creating interactive dashboards, generating visual insights, processing large-scale data, conducting real-time analysis, and implementing machine learning algorithms across diverse data sources.



The Data Lake reference architecture provides three distinct layers to transform raw data into valuable insights:

Raw Layer

The raw layer is the initial layer in a data lake, built on Amazon S3, where data arrives in its original format directly from source systems without any transformation. The data in this layer is used to determine changes and data to consolidate in the next layer since it will contain multiple versions of the same data (changes, full loads, etc).

Data extracted from SAP (via SAP ODP OData or other mechanisms) needs to be prepared for further processing. The extracted data will be packaged in several files (defined by the package or page size in the extraction tool) hence multiple files for a given extraction run can be generated.

Enriched Layer

The Enriched Layer is built on Amazon S3 and it contains a true representation of the data in the source SAP system along with logical deletions and is stored in Iceberg format. The Iceberg Table file format allows the creation of Glue or Athena Tables within the Glue Data Catalog, supporting Database type operations such as Insert, Update and Deletion, with the Iceberg file

format handling the file operations (deletion of records, etc). Iceberg tables also supports the concept of Time Travel, which enables querying data for a specific point in time.

Data from the Raw Layer is inserted or updated in the Enriched layer in the right order based on the table key and is persisted in its original format (no transformation or changes). Each records needs to be enriched with certain attributes such as time of extraction and record number, this can be achieved with the AWS Glue jobs.

Curated Layer

The Curated Layer is the layer where data is stored for data consumption. Records deleted on the source are deleted physically. Any calculations (averages, time between dates, etc) or data manipulation (format changes, lookup from another table) can be stored in this layer, ready to be consumed. Data is updated in this layer using the AWS Glue jobs. Amazon Athena views are created on top of these tables for downstream consumption through Amazon QuickSight or similar tools.

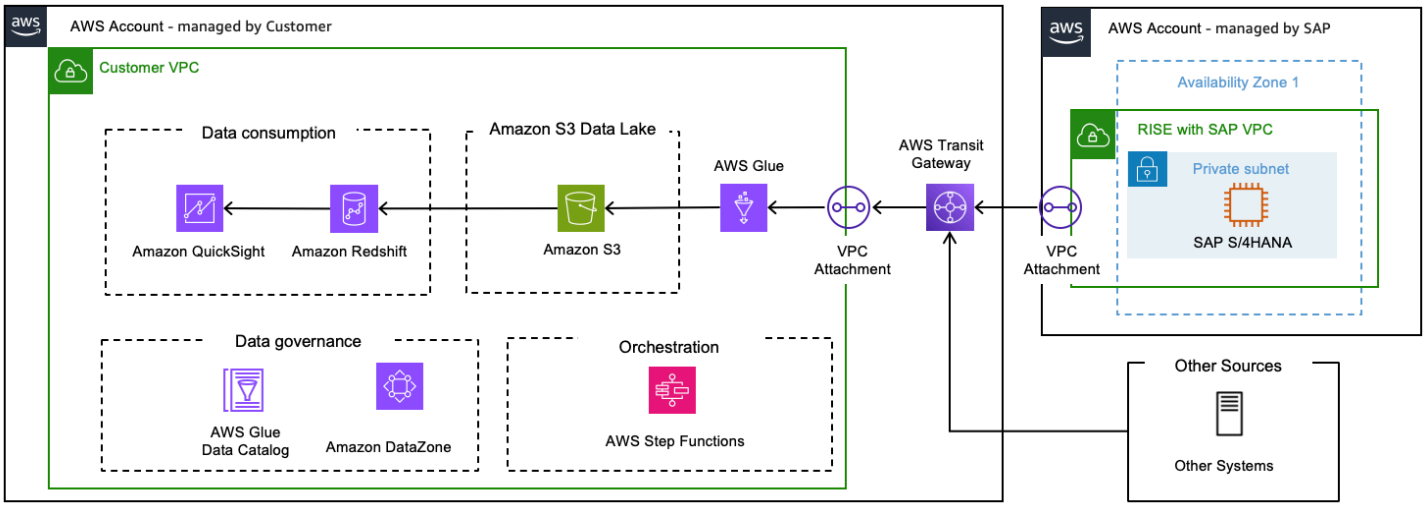
The Data Lakes with SAP and Non-SAP Data on AWS Solution Guidance provides a detailed architecture, steps to implement and accelerators to fast track the implementation of a Data Lake for SAP and non-SAP data. You can refer to the different available options to extract data from SAP to the Data Lake in the prior Data Integration section.

Data Warehouse Architecture

A Data Warehouse is a centralized repository based on “schema-on-write” approach that aggregates structured, historical data from multiple sources (both SAP and non-SAP) to enable advanced analytics, reporting, and business intelligence (BI). It enables organizations to analyze vast amounts of integrated data for informed decision-making, using optimized architectures for complex queries rather than transactional processing.

Business analysts, data engineers, data scientists, and decision-makers utilize business intelligence (BI) tools, SQL clients, and other analytics applications to access data warehouse. The architecture comprises tiers: a front-end client for presenting results, an analytics engine for data access and analysis, and a database server for data loading and storage.

Data is stored in tables and columns within databases, organized by schemas. Data warehouses consolidate data from multiple sources, enabling historical data analysis and ensuring data quality, consistency, and accuracy. Separating analytics processing from transactional databases enhances the performance of both systems, supporting reports, dashboards, and analytics tools by efficiently storing data to minimize I/O and deliver rapid query results to numerous concurrent users.

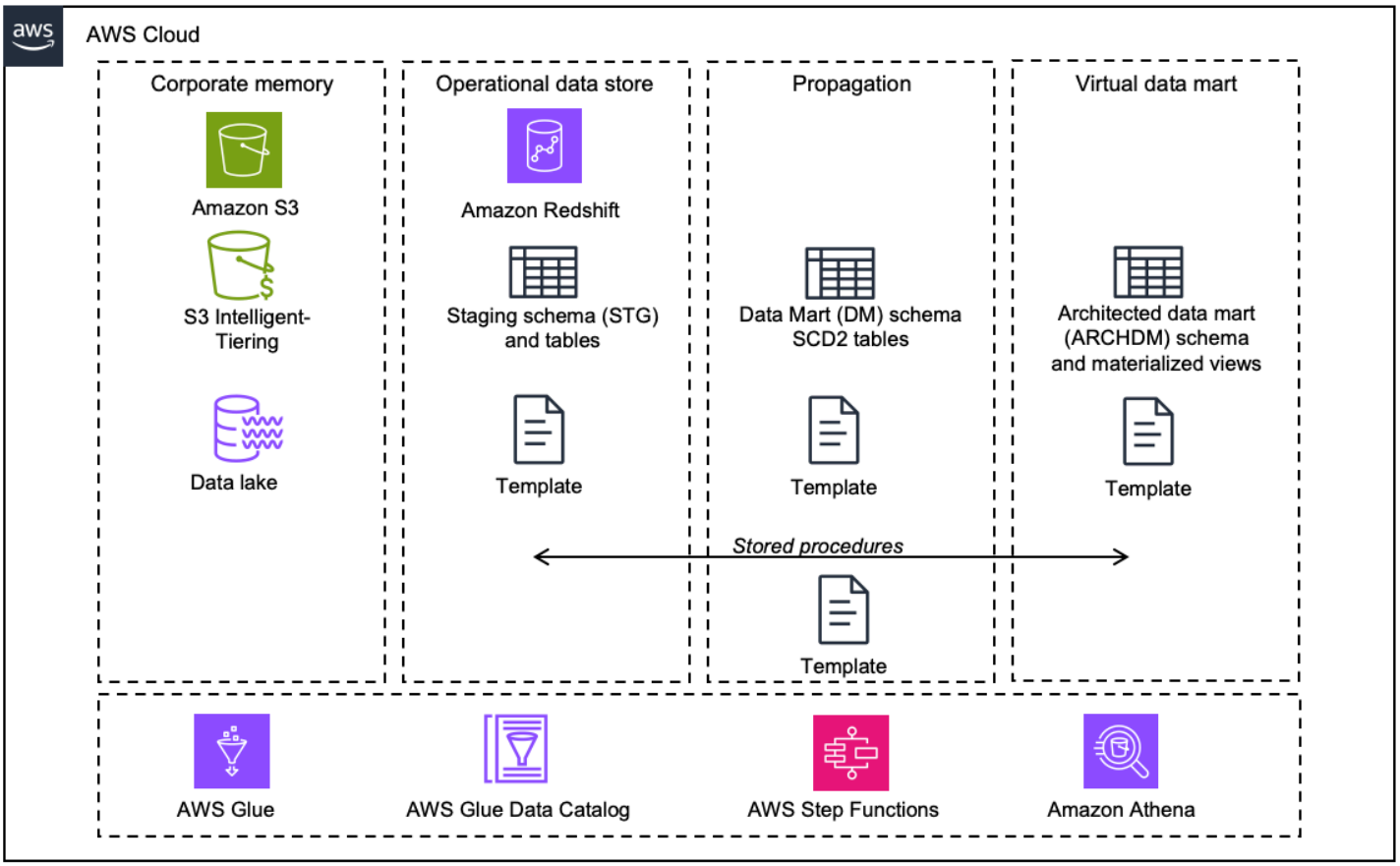


Key Characteristics

- Integrated: Consolidates data from disparate sources (e.g., CRM, ERP) into a unified schema, resolving inconsistencies in formats or naming conventions.
- Time-variant: Tracks historical data, allowing trend analysis over months or years.
- Subject-oriented: Organized around business domains like sales or inventory, rather than operational processes.
- Non-volatile: Data remains static once stored; updates occur via scheduled Extract, Transform, Load (ETL) processes rather than real-time changes.

Architecture Components

- ETL Tools: Automate data extraction from sources, transformation (cleaning and standardizing), and loading into the warehouse.
- Storage Layer:
  - Relational databases for structured data
  - OLAP (Online Analytical Processing) cubes for multidimensional analysis
- Metadata: Describes data origins, transformations, and relationships.
- Access Tools: SQL clients, BI platforms, and machine learning interfaces.



Data warehouses utilize a layered architecture to organize data at different levels of granularity, which helps ensure consistency and flexibility. The most common data warehouse architecture layers are the source, staging, warehouse, and consumption layers. SAP systems also employ a layer-based architecture for data warehouses. In the context of building a SAP cloud data warehouse on AWS. the architecture involves several key layers and components for data acquisition, storage, transformation, and consumption.

Corporate Memory

Amazon S3 Intelligent-Tiering is a storage class that automatically optimizes storage costs by moving data between access tiers based on changing access patterns. This ensures that frequently accessed data is readily available, while less frequently accessed or "colder" data is stored at a lower cost tier.

Operational Data Storage Layer

Amazon Redshift is utilized for operational data storage, propagation, and data mart functionalities. Scripts are provided to create schemas and deploy Data Definition Language (DDL)

with the necessary structures to load SAP source data. These DDLs can be customized to include SAP-specific fields.

Data Propagation Layer

Delta data loaded into S3 via Glue / AppFlow flows is used to generate Slowly Changing Dimension Type 2 (SCD2) tables, which maintain a complete history of changes.

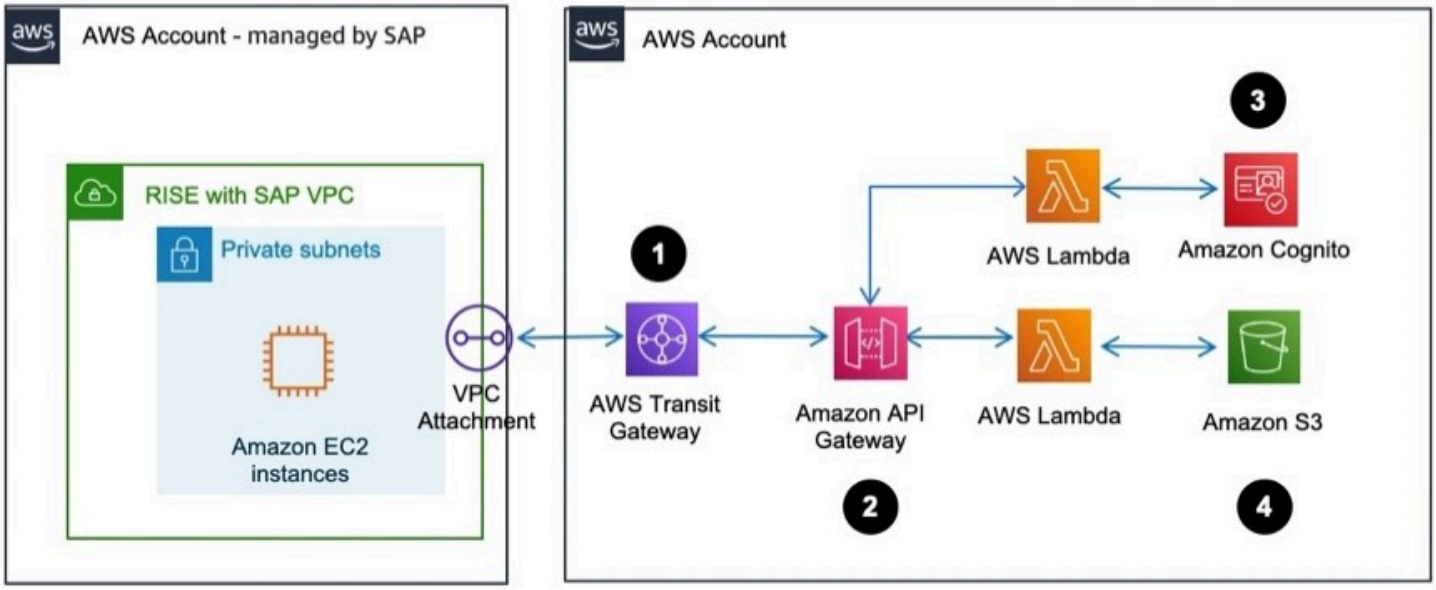
Data Mart Layer

Architected data mart models are created using Materialized Views in Redshift. Transactional data is enriched with master data (attributes and text), building data models that are ready for data consumption.

The Building SAP Data Warehouse on AWS Solution Guidance provides a detailed architecture, steps to implement and accelerators to fast track the implementation of a Data Warehouse for SAP.

Application integration

Deploy Amazon API Gateway to extract data out of SAP S/4HANA via HTTP API. API Gateway can consume data from IDOC, BAPI, and RFC. These need to be translated to a web service call. For more information, see AWS blogs. The following image shows this scenario.



Data flow

1. RISE with SAP VPC is connected to your AWS account not managed by SAP, via AWS Transit Gateway.

2. Amazon API Gateway is configured to route the authentication to AWS Lambda and Amazon Cognito
3. Amazon Cognito authenticates the session.
4. Once authenticated, Amazon API Gateway routes the package to AWS Lambda.
5. AWS Lambda stores the data in an Amazon S3 bucket.

Archiving and Document Management

SAP Data Archiving and Document Management System (DMS) plays a crucial role both before and after migrating to RISE with SAP. It helps businesses effectively manage database growth and optimize overall costs. Before migrating to S/4HANA, archiving reduces migration expenses, minimizes downtime, and lowers risk by decreasing data volume. After moving to S/4HANA, it helps control operational costs and ensures optimal system performance. Additionally, businesses can decommission legacy SAP ECC systems, eliminating unnecessary expenses while retaining access to historical data

Data archiving for structured data. Data archiving is about moving closed business transactions data from a live SAP systems to an offline or secondary storage. The key aspect of data archiving is to set a process and strategy to reduce manual efforts while ensuring compliance with legal data retention requirements.

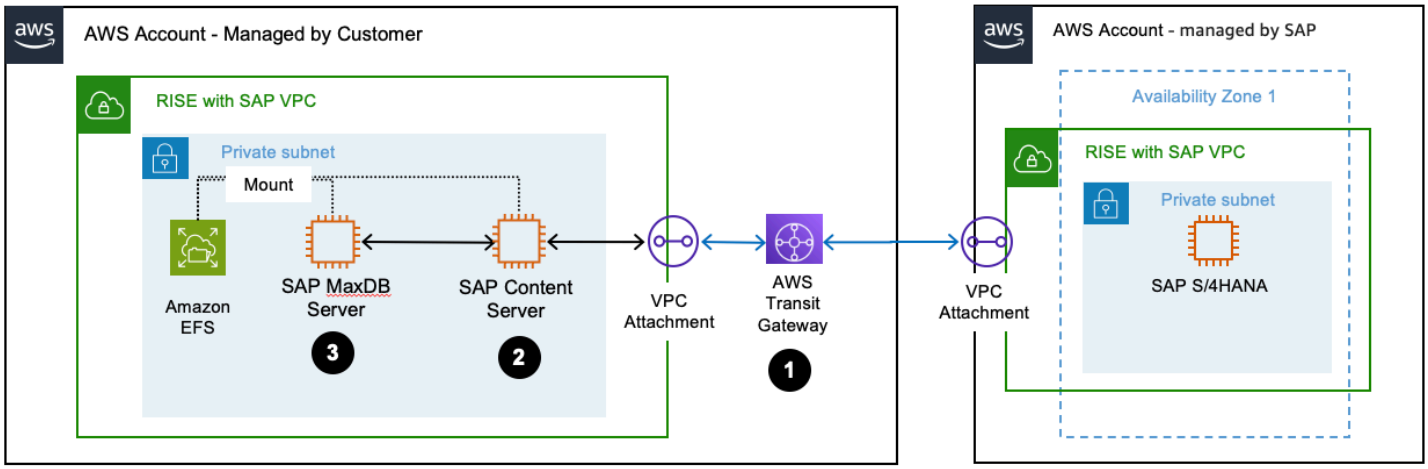
Document management for unstructured data. The difference between data and document archiving is the type of data that you are archiving. Document archiving relates to unstructured data likes invoices, sales orders, delivery notes, and others, which usually come in the format such as pdf, words, excels. This archiving occurs in real-time and it can be stored on any content server and linked to the related SAP transactions.

We shall discuss on the available options for your data archiving and document management systems within SAP.

Option 1 : SAP Content Server running on MaxDB

Many customers migrating to RISE with SAP choose to keep their SAP Content Server on AWS until they transition to SAP BTP Document Management System or OpenText Archiving solution. SAP Content Server is a standalone component designed to store large volumes of electronic documents in various formats. These documents can be securely saved in one or more SAP MaxDB instances or within the file system. Common examples of documents stored in SAP Content

Server include sales invoices, purchase orders, salary slips, emails, agreements, and others. This approach ensures seamless document management integrated into SAP business processes while maintaining accessibility and compliance.

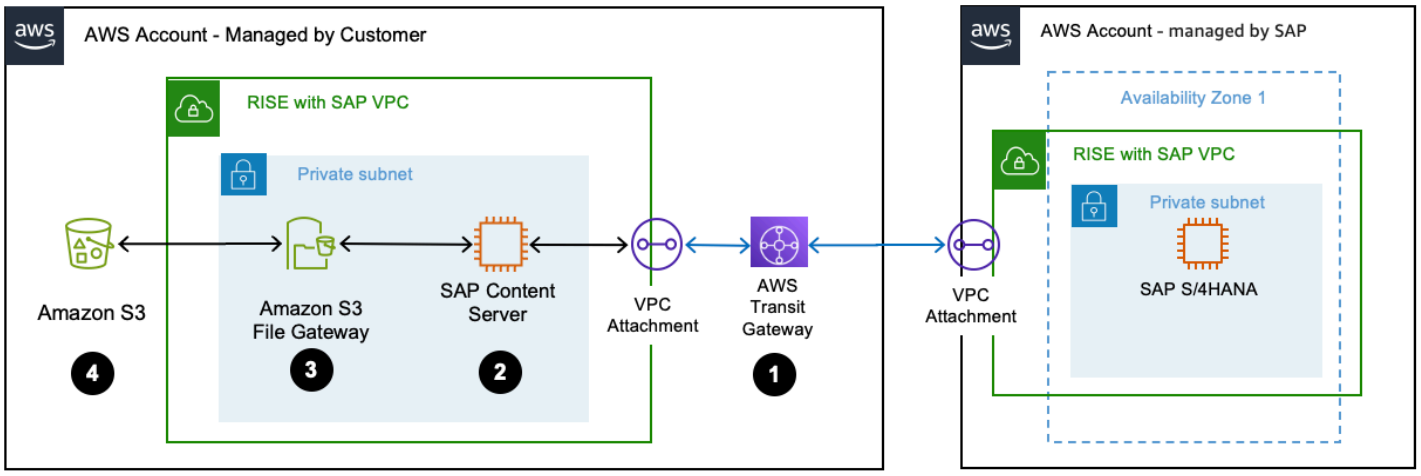


Architecture Description

1. RISE with SAP VPC is connected to an AWS account which you managed via AWS Transit Gateway.
2. SAP Content Server is setup in your AWS account and configured to serve as the destination for data archiving.
3. SAP MaxDB is setup in your AWS account and configured to run on AWS EC2 instance.
4. SAP Content Server High Availability using Amazon EFS. You can consider EFS Infrequent Access for documents which are not frequently accessed.

Option 2: SAP Content Server on Amazon S3 SAP Content Server, along with Amazon S3 can both meet SAP Data Archiving needs by providing scalable and secure storage for archived data. They offer features like versioning, access control, immutability, and integration with SAP systems. This section is relevant for customers experiencing SAP database growth, seeking performance improvements, aiming to reduce storage costs, or needing to meet compliance requirements for long-term data retention in their SAP environment.

The following image shows an SAP Content Server integrated with Amazon S3.



Architecture Description

1. RISE with SAP VPC is connected to an AWS account which you managed via AWS Transit Gateway.
2. SAP Content Server is setup in your AWS account and configured to serve as the destination for data archiving.
3. The SAP Content Server integrates with Amazon S3 File Gateway, which acts as a storage gateway to facilitate file-based storage. S3 File Gateway enables mounting of Amazon S3 as Network File System (NFS).
4. An Amazon S3 bucket stores the necessary archive files. You can use S3 Lifecycle configuration to manage lifecycle of the objects. For enhanced data protection or regulatory compliance, you can implement retention policies using S3 Object Lock. You can move files to different S3 storage classes using automated Lifecycle Management. For more information, see Using Amazon S3 storage classes.

SAP Content Server, in conjunction with Amazon S3, provides a mechanism for transferring archived data to long-term S3 storage such as Amazon S3 Glacier. This archived data can then be accessed using SAP's standard archive read programs.

However, if you require more extensive integration with SAP, third-party solutions like Syntax CxLink or OpenText offer additional libraries. These enhance the integration capabilities, providing more advanced functionalities for managing and accessing archived data directly within the SAP environment. For organizations employing SAP Information Lifecycle Management (ILM) to manage data retention and governance, see how Syntax Cxlink for ILM can enhance your ILM strategy by using Amazon S3 as a secondary storage solution for SAP ILM. This approach

leverages the scalability and cost-effectiveness of cloud storage while maintaining the robust data management capabilities of SAP ILM.

Option 3: SAP OpenText Archiving in RISE

SAP OpenText Archiving is enabling secure document storage, compliance, and cost-efficient data management for RISE with SAP. SAP OpenText Archiving is a cloud-based document management and archiving solution that integrates with SAP to store, retrieve, and manage unstructured content (e.g., invoices, contracts, purchase orders). It ensures compliance with regulatory requirements, reduces database footprint, and optimizes SAP S/4HANA performance. Within RISE with SAP, OpenText is included as an optional component in the RISE BOM.

Option 4: OpenText infoArchive for RISE

OpenText InfoArchive is a modern archive solution and cloud-based service for compliant archiving of both structured and unstructured information that is highly-accessible, scalable, and economical. It's a centralized platform which enables flexible storage options for unstructured content, including storage on Amazon Simple Storage Service (Amazon S3). InfoArchive Cloud Edition on AWS is offered as customer-deployed or as a managed solution by OpenText running on AWS.

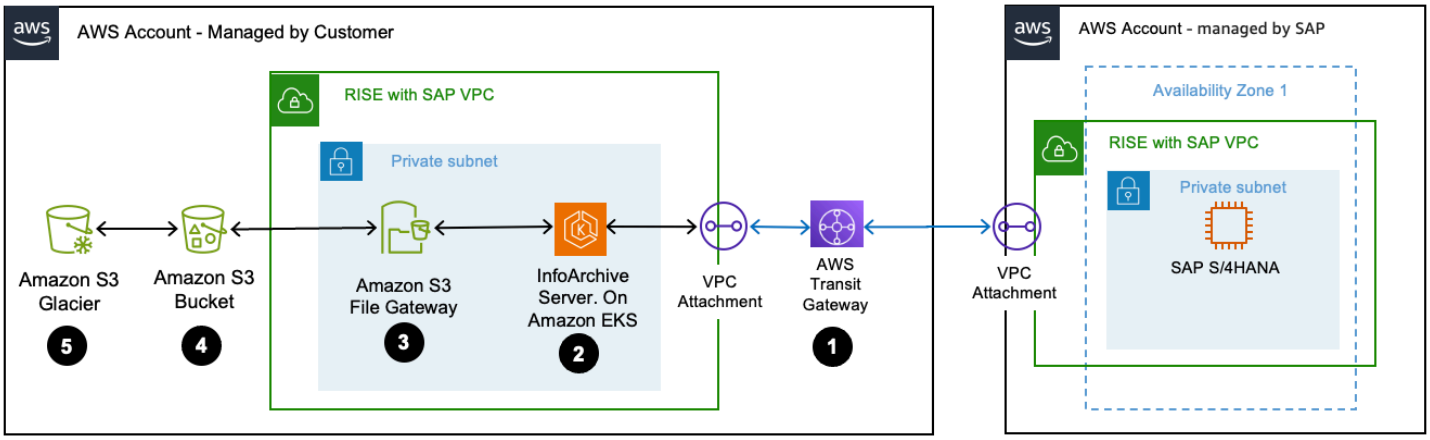
OpenText infoArchive is a general-purpose archiving platform designed to retire legacy SAP applications and store structured and unstructured data from multiple systems. This beyond supports SAP ECC, CRM, HR, and industry-specific systems (Healthcare, Banking, etc.) OpenText infoArchive can be used to Archive inactive data and decommission retired SAP legacy applications. This comes with pre-built SAP views.

Key Features

1. Application Decommissioning – Retires legacy applications while keeping data accessible.
2. Structured and Unstructured Data Archiving – Stores documents, emails, records, and databases.
3. Multi-System Support – Works with SAP, Oracle, Salesforce, Microsoft, and custom applications.
4. Advanced Search & Analytics – Uses AI/ML for insights into archived data.
5. Regulatory Compliance – HIPAA, GDPR, SEC 17a-4, etc.

You can deploy an OpenText infoArchive Server integrated with Amazon S3 for SAP data decommissioning. The following image shows this scenario with AWS services. OpenText InfoArchive on AWS is deployed on Amazon Elastic Kubernetes Service (EKS) for hosting its web

application, OpenText Directory Service for authentication and authorization, and the InfoArchive server. Customers can also procure it through AWS marketplace.



Architecture Description

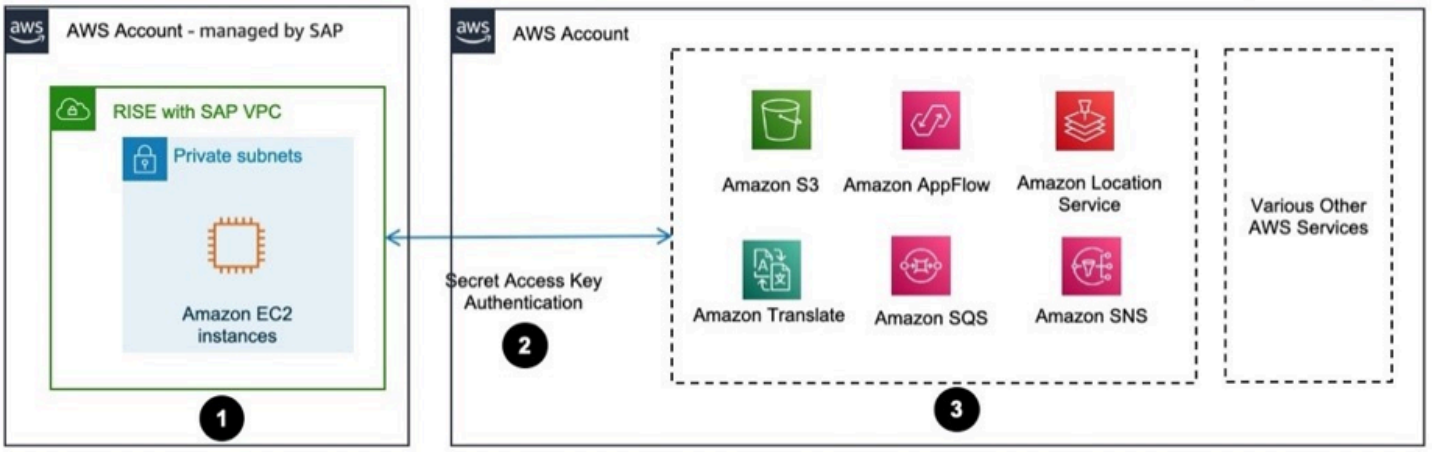
1. RISE with SAP VPC is connected to your AWS account via AWS Transit Gateway.
2. OpenText InfoArchive on AWS is deployed on Amazon Elastic Kubernetes Service (Amazon EKS) in your AWS account and configured to serve as the destination for data archiving.
3. OpenText InfoArchive integrates with Amazon S3 File Gateway, which acts as a storage gateway to facilitate file-based storage. S3 File Gateway enables mounting of Amazon S3 as Network File System (NFS).
4. An Amazon S3 bucket stores the necessary archive files. You can use S3 Lifecycle configuration to manage lifecycle of the objects. For enhanced data protection or regulatory compliance, you can implement retention policies using S3 Object Lock.
5. Older documents can be moved to Amazon S3 Glacier for long-term archival.
6. You can move files to different Amazon S3 storage classes using automated Lifecycle Management. For more information, see Using Amazon S3 storage classes.

Development and extension

AWS SDK for SAP ABAP

Deploy AWS SDK for SAP ABAP on RISE with SAP VPC to avail AWS services using the ABAP language. For more information, see What is AWS SDK for SAP ABAP?

You can authenticate AWS SDK for SAP ABAP with IAM access key. The following image shows this scenario.



Data flow

1. AWS SDK for SAP ABAP is installed via a set of transports in SAP S/4HANA within RISE with SAP VPC.
2. SAP S/4HANA is configured with IAM access key for authenticating access to AWS services. For more information, see Managing access keys for IAM users.
3. Access to AWS services with AWS SDK for SAP ABAP has been established.

Compatibility packs and alternatives

Compatibility packs (CP) are temporary use rights to classic functionality within S/4HANA, created in 2016. It is part of every SAP S/4HANA contract either on-premise and private cloud. This was with goal smooth transition for SAP installed-base customers and gain time to finalize the new simplified application architecture.

During the transition from SAP Business Suite to S/4HANA, business functions moved through these paths in the process. You can find out more from presentation by Michael Deller (SAP) and Roland Hamm (SAP).

In SAP Note 2269324, SAP defines categories to help organizations plan their strategy for compatibility packs. These categories guide decisions for transitioning away from SAP business suite to SAP S/4HANA.

- Alternative Exists
- Alternative Exists with Roadmap - Alternative exists providing core functionality; comprehensive coverage is on roadmap
- Alternative Planned - Planning of development scope and timeline is work in progress

- No Alternative Planned - No intention or plan to provide an alternative beyond 2025
- Clarification - Clarification of strategy in progress

How can AWS helps customers to find alternatives ?

Organizations should evaluate their current SAP landscape and plan their transition strategy considering both SAP compatibility pack expiration dates and available alternatives. When compatibility packs lack alternatives, you can leverage combined AWS and SAP services. This approach aligns with the AWS Refactor and re-architect migration strategy, which focuses on reimagining applications and processes. Here are the details

- SAP and AWS joint reference architecture was developed to address common questions from joint customers and partners on how to utilize SAP BTP and/or AWS services for different business solution scenarios. Refer also to this blog for more details.
- The AWS SDK for SAP ABAP simplifies the use of 200 plus AWS services alongside SAP applications with a client library of modules that are consistent and familiar to ABAP developers.
- SAP Products and AWS Partner Solutions on AWS Marketplace
- You can contact our SAP on AWS expert team to help you guide if needed.

One example “SAP Tax Classification and Reporting” has been tagged as “No Alternative Planned” in the SAP Note 2269324 (refer to S4HANA CompScope – Way Forward – Info – 06032025.xlsx), in this case, you can explore alternative such as the Thomson Reuters ONESource Indirect Tax Determination at AWS Marketplace.

Security Extension

mTLS Authentication

Mutual Transport Layer Security (mTLS) Authentication establishes a secure, two-way encrypted connection between client and server. Unlike standard TLS, where only the server provides a certificate, mTLS requires both parties to present digital certificates.

The mTLS authentication process works in four steps: . The client requests a connection to the server . The server presents its certificate . The client verifies the server’s certificate . The client presents its certificate for server verification and authentication

Why is mTLS Authentication for SAP

The implementation of mutual TLS (mTLS) authentication for SAP systems will enhance security, improve user experience, and reduce operational overhead. It will modernize user authentication infrastructure to support digital transformation while ensuring compliance with security standards. mTLS address below security requirements in SAP environments:

1. Enhanced Security: mTLS provides two-way authentication, ensuring both the client and server verify each other's identity. This significantly reduces the risk of unauthorized access and man-in-the-middle attacks.
2. Seamless User Experience with Single Sign On (SSO): mTLS can be integrated with SSO solutions, allowing users to access multiple SAP applications and services without repeatedly entering credentials. This creates a smoother, more efficient user experience across the SAP ecosystem.
3. Automated Certificate Rotation: mTLS allows for automated rotation of certificates, enhancing security by regularly updating authentication credentials without manual intervention. This reduces the risk of using expired or compromised certificates and minimizes administrative overhead.
4. Principal Propagation for Interfaces: mTLS enables secure principal propagation across different SAP interfaces and systems. This eliminates the need for generic and privileged accounts (like SAP user with SAP_ALL authorization) for system-to-system communication, significantly improving security and auditability.
5. Scalability and Performance: mTLS can be implemented at the network level, offloading authentication processes from application servers. This can lead to improved performance and scalability of SAP systems.
6. Support for Zero Trust Architecture: mTLS aligns well with zero trust security models, where trust is never assumed and always verified.

mTLS Client Authentication with Application Load Balancer

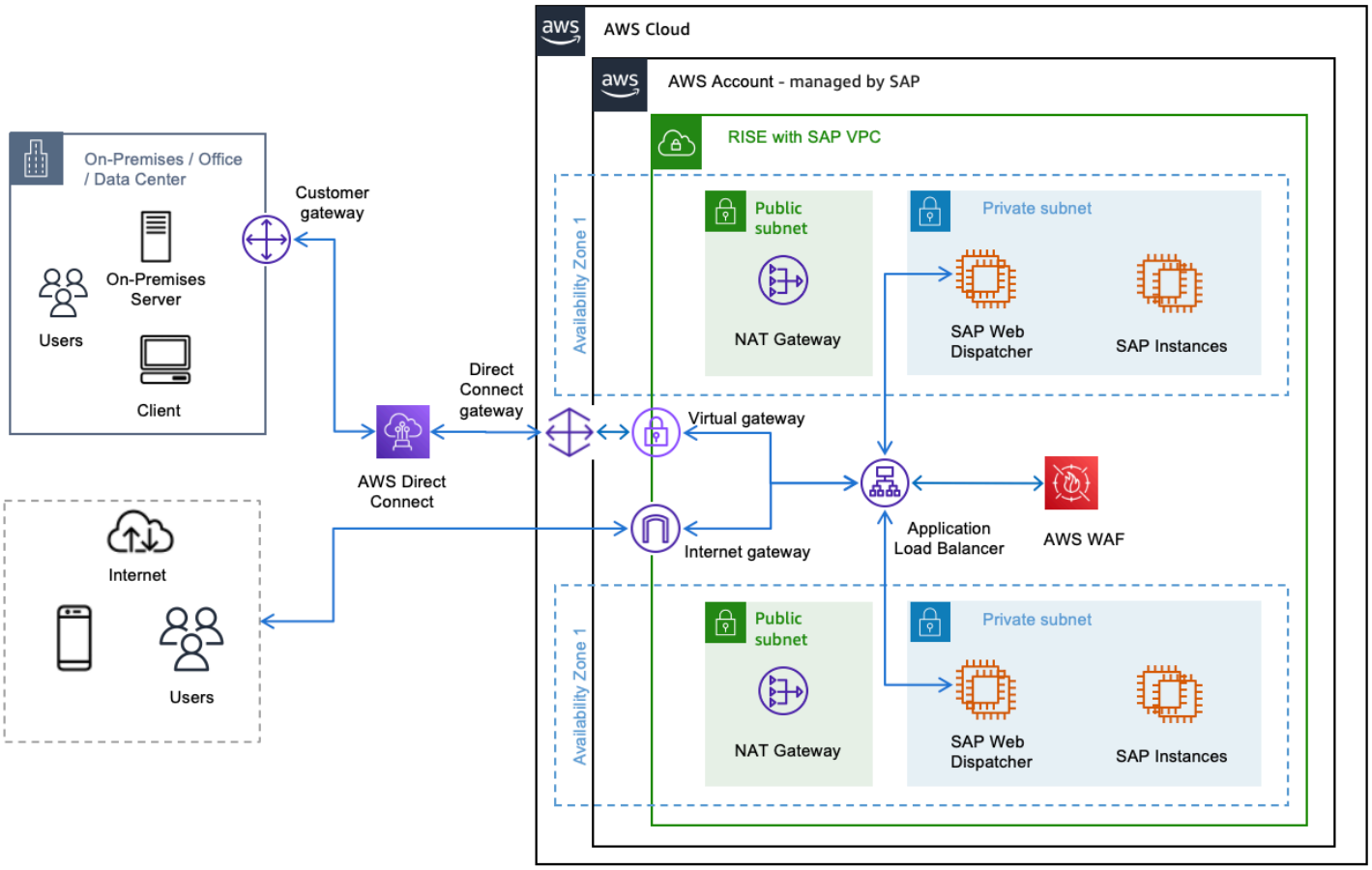
Application Load Balancer (ALB) supports mTLS authentication. It offers two modes: verify and passthrough mode.

Prerequisite

To ensure seamless communication, all SSL (Secure Socket Layer) or TLS certificates used across the infrastructure, including those at the ALB, SAP Web Dispatcher, and S/4HANA systems should originate from a single and trusted root certificate authority to ease the implementation and maintenance of these certificates.

mTLS Architecture Diagram

The diagram below describes a basic SAP on AWS architecture that is adapted to align with the RISE with SAP SKU offering.



mTLS Verify Mode

To enable mTLS verify mode, create a trust store containing a CA certificate bundle. This can be accomplished using AWS Certificate Manager (ACM), AWS Private CA, or by importing your own certificates. Manage revoked certificates using Certificate Revocation Lists (CRLs) stored in Amazon S3 and linked to the trust store.

ALB handles client certificate verification against the trust store, effectively blocking unauthorized requests. This approach offloads mTLS processing from backend targets, improving overall system efficiency. ALB imports CRLs from S3 and performs checks without repeated S3 fetches, minimizing latency.

Beyond client authentication, ALB transmits client certificate metadata through HTTP Headers (e.g., X-Amzn-Mtls-Clientcert-Leaf) to the backend SAP Web Dispatcher via HTTP headers. This

allows for additional logic implementation on backend targets based on certificate details, to meet the requirement for SAP Servers to preserve original “Host Header” information.

This enables the server to process client certificate metadata consistently, even when originating from non-SAP sources like an AWS load balancer terminating the SSL connection. In the event that you are implementing end-to-end encryption through ALB – SAP Web Dispatcher – SAP Servers, you must configure SAP Web Dispatcher profile parameters such as icm/HTTPS/client_certificate_header_name for more details you can refer to this link.

mTLS Passthrough Mode

In mTLS passthrough mode, ALB forwards the client’s entire certificate chain to backend targets. This is done via an HTTP header named X-Amzn-Mtls-Clientcert. The chain, including the leaf certificate, is sent in URL-encoded PEM format with +, =, and / as safe characters. Below are the consideration while using mTLS Passthrough Mode:

- ALB adds no headers if client certificates are absent; backends must handle this.
- Backend targets are responsible for client authentication and error handling.
- For HTTPS listeners, ALB terminates client-ALB TLS and initiates new ALB-backend TLS using target-installed certificates.
- ALB’s TLS termination allows use of any ALB routing algorithm for load balancing.

NLB Passthrough

When you have stringent security compliance rules requiring server-side termination of client TLS connections, you can utilize a Network Load Balancer (NLB).

Key points to note:

1. NLB operates at the transport layer (Layer 4 of the OSI model).
2. It provides low-latency load balancing for TCP/UDP connections.
3. NLB allows the backend servers to handle TLS termination, which can be crucial for certain security compliance scenarios.

This approach ensures that sensitive decryption processes occur on your controlled server environment, potentially meeting specific security mandates while maintaining efficient traffic distribution.

Comparison of mTLS verify mode vs mTLS passthrough mode vs NLB passthrough.

Considerations	ALB with mTLS Verify mode	ALB with mTLS passthrough mode	NLB
OSI Layer	Layer 7 (Application)	Layer 7 (Application)	Layer 4 (Transport)
Integration with AWS WAF	Supported	Supported	Not Supported
Client Authentication	Done by ALB (AWS managed)	Done by backend (Customer managed)	Done by backend (Customer managed)
Client SSL/TLS Termination	At ALB (AWS managed)	At ALB (AWS managed)	At backend target (Customer managed)
Header Based Routing	Supported	Supported	Not Supported
Trust Store	Required at ALB	Not required at ALB	Not required at NLB
Certification Revocation List	Managed at ALB	Managed by backend (if required)	Managed by backend (if required)
Backend Processing Load	Lower	Lower	Higher
Error Handling	Managed by ALB	Managed by backend	Managed by backend

Note: RISE with SAP on AWS supports ALB with mTLS Verify Mode.

Zero Trust Access

AWS Verified Access is a Zero Trust security solution that replaces traditional VPNs for corporate application security. It validates each access request by checking user identity, device health, and location. The service integrates with Okta, Azure Active Directory, and IAM Identity Center while providing detailed access logging and monitoring. See AWS Verified Access for more information.

Key Features and Benefits of AWS Verified Access for SAP

This solution secures SAP landscapes through Zero Trust security, managing both SAPGUI and web-based (HTTPs) access through a unified framework. It encrypts SAPGUI TCP connections and HTTPs access for Fiori applications, eliminating Traditional VPN while maintaining security standards.

Users can access RISE with SAP systems faster (before the VPN connectivity is setup). It allows you to grant secure access to remote users and external consultants, which do not have a VPN access to your corporate network

1. Identity-Centric Security Verified access integrates with existing identity providers (IdP), such as Microsoft Azure AD (Entra), Okta, Ping, and others. It provides real-time user authentication and authorization that support for SAML 2.0 and AWS IAM Identity Center
2. Contextual Access Control Verified Access is able to implement device security posture assessment, location-based access policies, role-based access management and dynamic policy evaluation.
3. Enhanced Performance Verified Access provides a direct and optimized connection paths to SAP systems, thus reducing network latency, improve performance and provide more consistent user experience to SAP systems.
4. Simplified Administration Verified Access provides centralized policy management through AWS Cedar Policy Language and authorization engine. It provides automated compliance reporting, real-time access monitoring and reduced infrastructure maintenance

Implementation Guide

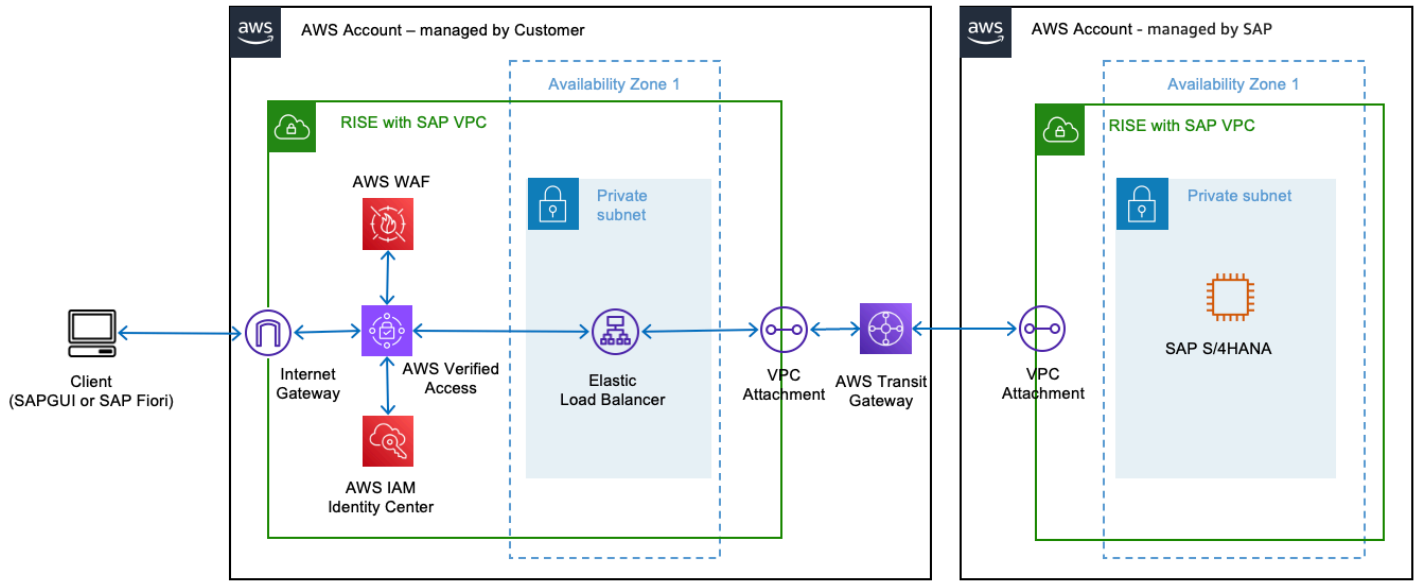
Prerequisites

- AWS IAM Identity Center enabled in the AWS Region that you prefer. For more information, see Enable AWS IAM Identity Center.
- A security group to allow network access to SAP applications.
- SAP application running behind an internal AWS Elastic Load Balancer. Associate your security group with the load balancer. (you can use a Network Load Balancer for both SAP GUI and SAP Fiori access, or Application Load Balancer for SAP Fiori access only).
- A public TLS certificate in AWS Certificate Manager when configuring AWS Verified Access for HTTP(s) based access (i.e. SAP Fiori). Use an RSA certificate with a key length of 1,024 or 2,048.
- A public hosted domain and the permissions required to update DNS records for the domain. (example: Amazon Route 53)

- An IAM policy with the permissions required to create an AWS Verified Access instance. For more information, see Policy for creating Verified Access instances.
- Set the system environment variable **SAP_IPV6_ACTIVE=1** as per SAP note 1346768 (requires a SAP S-user ID to access), this is needed when accessing SAP application using Verified Access endpoint from SAP GUI.

How to Implement AWS Verified Access for SAP

1. Create a Verified Access Trust Provider. After IAM Identity Center is enabled on your AWS account, you can use the following procedure to set up IAM Identity Center as your trust provider for Verified Access.
2. Create a Verified Access instance. You use a Verified Access instance to organize your trust providers and Verified Access groups. Use the following procedures to create a Verified Access instance, and then attach or detach a trust provider from Verified Access.
3. Create a Verified Access group. You use Verified Access groups to organize endpoints by their security requirements. When you create a Verified Access endpoint, you associate the endpoint with a group. Use the following procedure to create a Verified Access group
4. Create a load balancer endpoint for Verified Access. Verified Access endpoint represents an application. Each endpoint is associated witha Verified Access group and inherits the access policy for the group. Use the following procedure to create a load balancer endpoint for Verified Access for SAP application.
5. Configure DNS for the Verified Access endpoint. For this step, you map your SAP application's domain name (for example, www.myapp.example.com) to the domain name of your Verified Access endpoint. To complete the DNS mapping, create a Canonical Name Record (CNAME) with your DNS provider.
6. Add a Verified Access group-level access policy. AWS Verified Access policies allow you to define rules for accessing your SAP applications hosted in AWS. Refer to the following sample statements to derive one for your application as per your requirements.
7. Test the connectivity to your application. You can now test connectivity to your application by entering your SAP application's domain name into your web browser, for HTTP(S) based access such as SAP Fiori.



The preceding diagram describes on how AWS verified Access deployed and integrated with RISE with SAP

Artificial Intelligence

Generative AI for SAP on AWS

Generative AI refers to intelligent systems capable of creating new content like text, images, audio, or code based on the data they have been trained on. These systems employ machine learning techniques, particularly deep learning and neural networks, to identify patterns and relationships within the training data, and then generate novel outputs that resemble the learned information.

As organizations embrace generative AI for their employees and customers, cybersecurity practitioners must rapidly assess the risks, governance, and controls associated with this evolving technology. As security leaders working with the largest, most complex customers at Amazon Web Services (AWS), we're regularly consulted on trends, best practices, and the rapidly evolving landscape of generative AI and the associated security and privacy implications. Generative AI solutions cover multiple use cases that affect your security scope. To better understand the scope and corresponding key security disciplines, see the AWS blog post Securing generative AI: An introduction to the Generative AI Security Scoping Matrix.

SAP and AWS have co-innovated services which help customers to combine SAP's AI innovations and enterprise expertise with Amazon's cutting-edge AI capabilities and technological solutions, thereby unlocking significant opportunities for business enhancement. RISE customers can

accelerate their AI adoption through SAP Business Technology Platform (BTP) AI services like Generative AI Hub and AWS enterprise GenAI services including Amazon Bedrock, and Amazon Q enabling secure, scalable AI solutions.

SAP Data Integration and Management on AWS

Data serves as the cornerstone for the success of any generative AI solution. The quality, quantity, and diversity of data are critical factors that directly influence the performance and efficacy of AI models. We recommend reviewing our Guidance for SAP Data Integration and Management on AWS, which provides the essential data foundation for empowering customers to build AI solutions. It shows how to integrate data from SAP ERP source systems and AWS in real-time or batch mode, with change data capture, using AWS services, SAP products, and AWS Partner Solutions. This includes an overview reference architecture showing how to ingest SAP systems to AWS in addition to detailed architectural patterns that complement SAP-supported mechanisms using AWS services, SAP products, and AWS Partner Solutions.

Ways to implement Generative AI Solutions for RISE on AWS

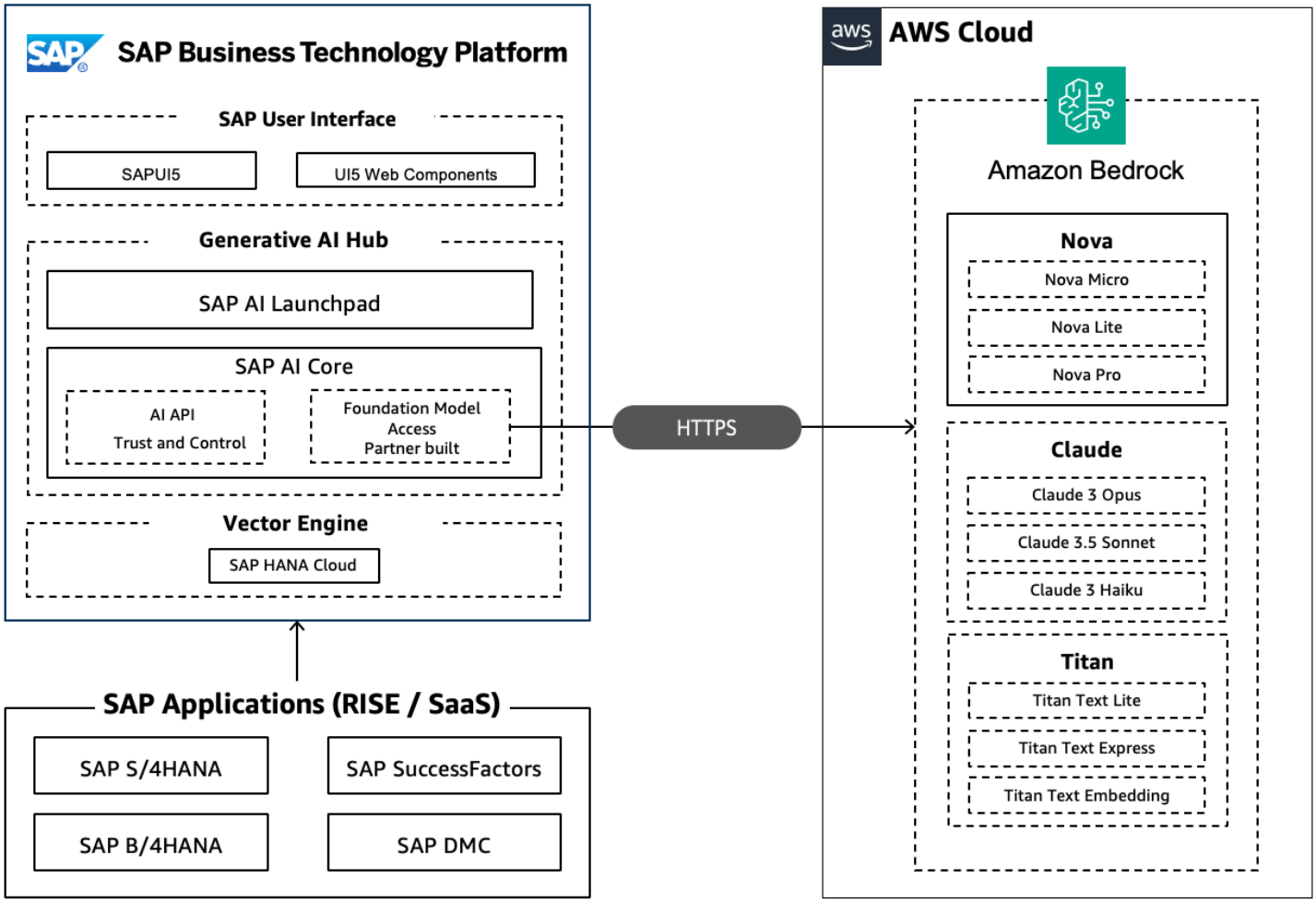
This architectural guidance helps you build advanced AI solutions. It shows you how to effectively combine RISE with SAP, SAP Business Technology Platform (BTP), and AWS's AI services to create powerful and innovative systems.

SAP Generative AI Hub

The SAP Generative AI Hub facilitates enterprise customers to seamlessly access a comprehensive array of commercial and open-source Large Language Models (LLMs) within a secure and controlled environment. SAP's Generative AI Hub further streamlines the process by which organizations running on RISE can access and implement sophisticated AI models hosted by Amazon. For further details on the SAP's Generative AI Hub supported models through Bedrock, please refer to SAP Note 3437766.

With these models now available through the Generative AI Hub, developers can effectively harness them in conjunction with the robust capabilities of SAP BTP to construct and scale AI-driven solutions that integrate seamlessly with SAP applications.

RISE customers can leverage Bedrock LLM models through SAP BTP services to create tailored AI solutions including enhanced capabilities for SAP's AI copilot Joule. This integration enables automated workflows and optimized enterprise data processing, ensuring RISE customers achieve precise, actionable business outcomes in their cloud transformation journey.



The diagram illustrates a design framework for a Generative AI solution or an S/4 HANA extension, created and implemented on the BTP. This design illustrates smooth integration with RISE and SAP Solutions that operate on AWS. utilizing the connectivity service and leveraging Bedrock LLM models through the Generative AI Hub.

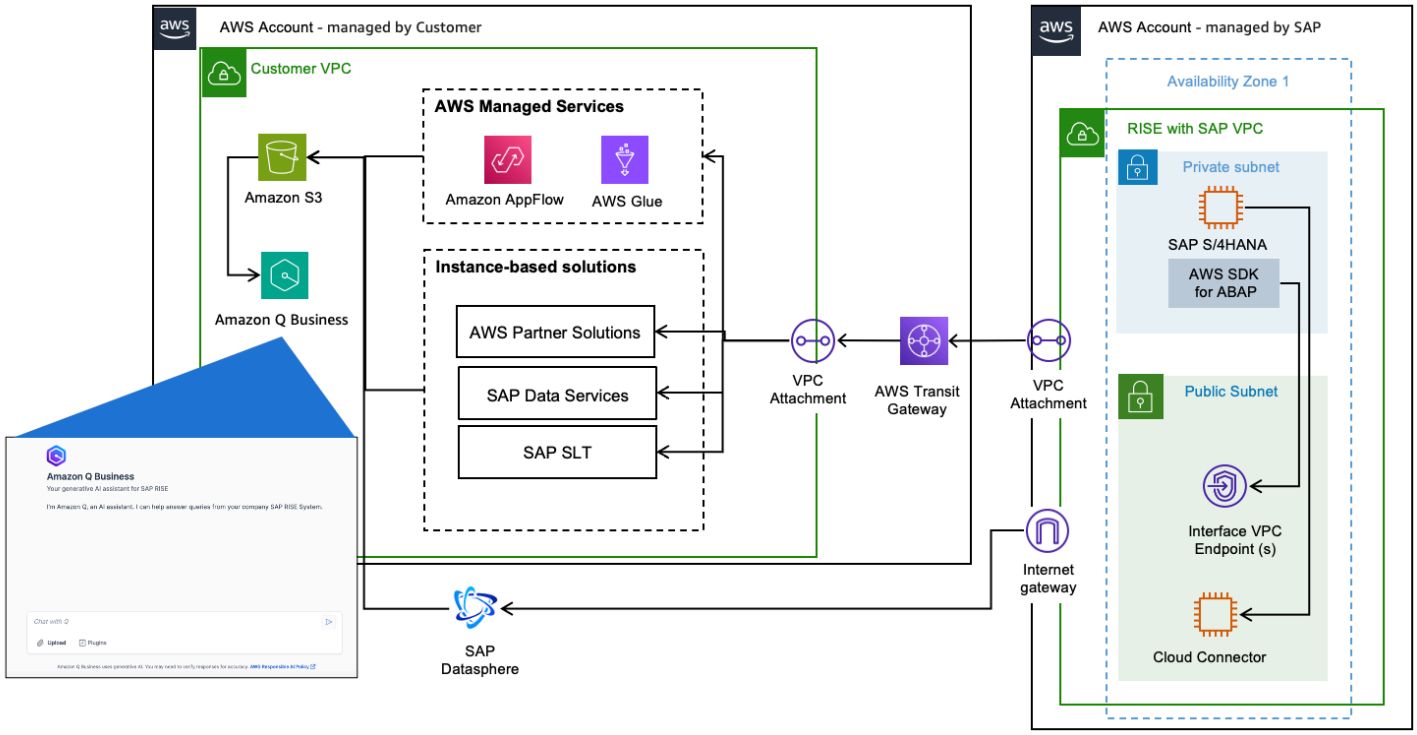
Amazon Q for Business

RISE customers can leverage Amazon Q Business to answer questions, provide summaries, generate content, and complete tasks based on enterprise data. End users receive immediate, permission-aware responses from enterprise data sources with citations. Q Business is a fully managed generative-AI powered assistant with 40+ pre-built connectors to various enterprise applications and data sources.

Customers who choose to break data silos by creating data warehouse or data lake solutions can use SAP and other enterprise data as source for Q Business to :

- Create a unified search experience across systems and data thereby extracting key insights

- Create and share lightweight applications either to select users or add them to an organization's application library
- Perform actions across popular business applications and platforms
- Create and automate complex business workflows



The diagram above illustrates a design framework for Q Business based search for RISE customers. It illustrates how SAP data can be extracted utilizing AWS services and using pre-built connectors from Q Business organizations can create a unified search experience.

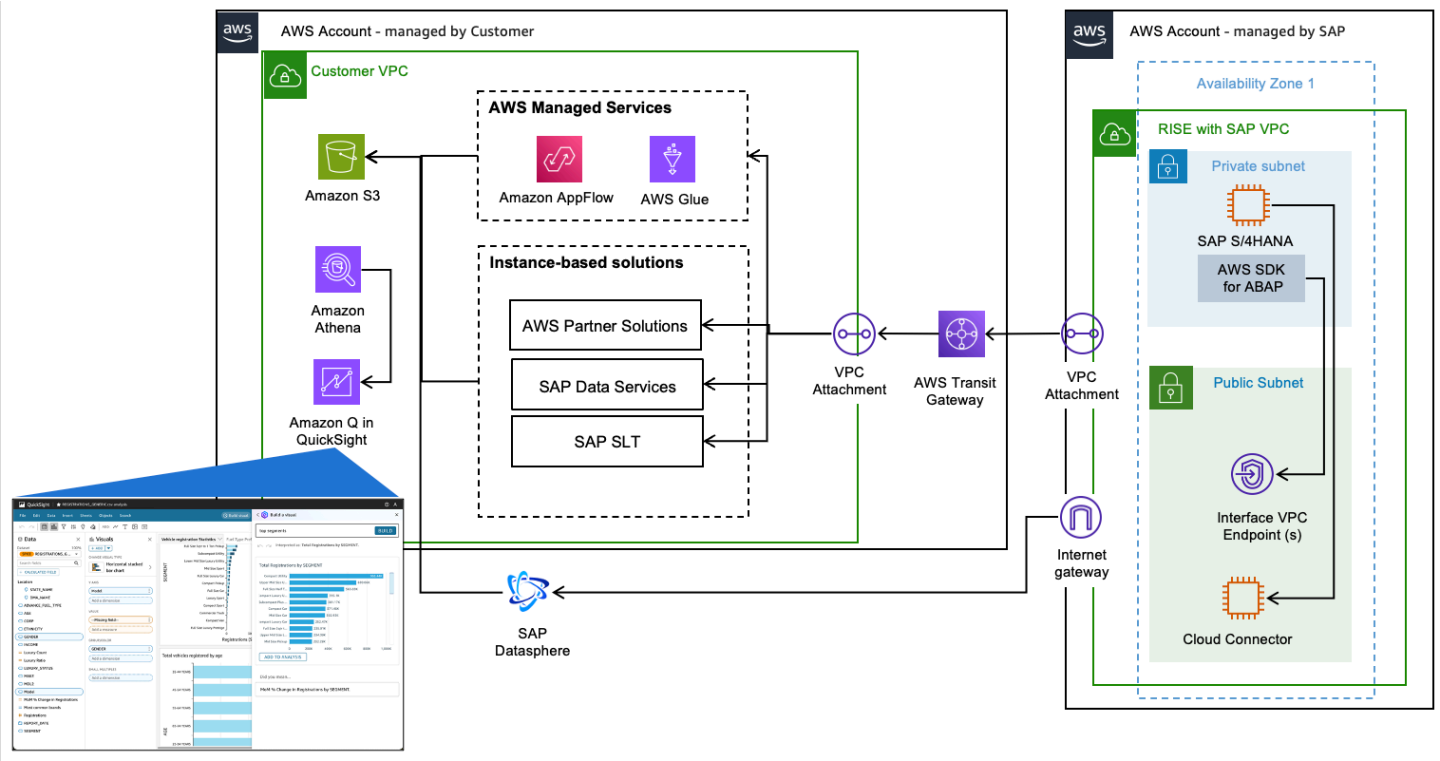
Solution Flow:

1. Establish connectivity with RISE environment by creating AWS Glue connection for SAP OData
2. Ingest relevant SAP data by creating ETL jobs
3. Utilize pre-built connectors to various data sources and applications to connect with Q Business. Ingest the relevant content while inheriting the existing identities, roles and permissions.
4. End users can interact in natural language to derive business insights from data across multiple applications

Amazon Q in QuickSight

Amazon Q in QuickSight revolutionizes SAP data analysis through its advanced 'Generative business intelligence' capabilities, empowering business users with intuitive self-service reporting tools. Using natural language prompts, RISE customers can effortlessly create sophisticated visual dashboards and data narratives without requiring SQL or programming expertise.

This democratization of data analysis dramatically reduces report generation time from days to hours, eliminating dependencies on specialized ABAP developers and/or analytics teams. The system's AI-driven automation intelligently generates contextual titles, organized sections, coherent story flows, and actionable insights with specific recommendations. For RISE customers, this translates into accelerated decision-making processes, with deeper more accessible insights from their enterprise data.



The diagram illustrates a framework of Amazon Q in Quicksight with SAP Data.

Solution Flow:

1. SAP report to process business logic and upload data to Amazon S3.
2. With AWS SDK for SAP ABAP, it will create an Amazon Athena query linked to the SAP report data on S3.
3. Create an QuickSight dataset and topic based on the Athena query.

4. Now using Q in Quicksight, you can interact with the data generated by SAP reports using natural language and get insights of data, to build dashboard and generate stories.

Amazon Bedrock Agents

Amazon Bedrock Agents offers the ability to build and configure autonomous agents in your application. An agent helps end-users complete actions based on organization data and user input. Agents orchestrate interactions between foundation models (FMs), data sources, software applications, and user conversations. In addition, agents automatically call APIs to take actions and invoke Amazon Bedrock Knowledge Bases to supplement information for these actions. By integrating Bedrock Agents, you can accelerate development effort to deliver generative artificial intelligence (generative AI) applications.

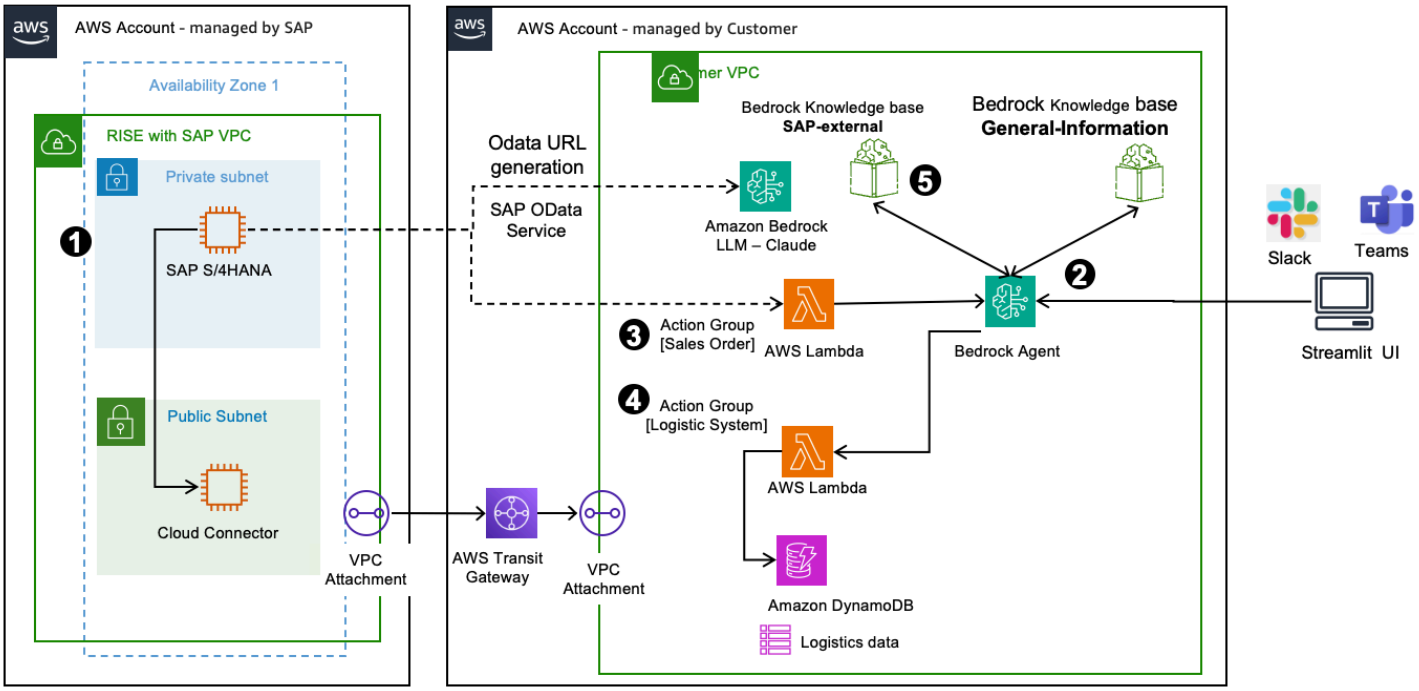
Agents perform the following tasks:

- Extend foundation models to understand user requests and break down the tasks that the agent must perform into smaller steps.
- Collect additional information from a user through natural conversation.
- Take actions to fulfill a customer's request by making API calls to your company systems.
- Augment performance and accuracy by querying data sources.

RISE customers can significantly optimize their business processes by leveraging Bedrock Agents. These agents seamlessly integrate with SAP systems, other enterprise systems and organisation's enterprise knowledge base, enabling natural language interactions to automate complex business processes and workflows.

By combining SAP and other enterprise data with AWS's advanced AI capabilities, customers can create customized, secure and comprehensive solutions that streamline operations, improve decision-making, and drive innovations.

Bedrock Agents can be tailored to specific business requirements, allowing organizations to automate routine tasks, extract valuable insights from SAP and other enterprise data, and scale their AI implementations efficiently across the enterprise. This integration empowers RISE customers to accelerate their digital transformation while maintaining data security and compliance within their SAP ecosystem.



The diagram above illustrates for a Generative AI architecture that automates tasks for your customers and answers questions from their SAP and other enterprise landscape.

Solution Flow:

1. Users access the application through a lightweight application such as Streamlit User Interface, Slack or Microsoft Teams.
2. Based on user inputs, Bedrock Agent orchestrates interactions between different applications, data sources or knowledge bases.
3. If the user query is related to SAP sales order, then the Bedrock Agent invokes the Sales Order action group which calls SAP API to get the related details.
4. If the user query is related to SAP shipping information, Bedrock Agent invokes the Non-SAP Logistic System action group.
5. If the user queries general information from the organisation's enterprise data, the Bedrock Agent invokes the General-Information knowledge base.
6. If the user query requires data from multiple systems, the Bedrock Agent fetches it from the different tools and collates it in the response.

The vision of having an autonomous ERP system may come to a reality where workflow are automated with Bedrock Agents, reducing manual effort and improving productivity of users for a more important business tasks and processes.