AWS Privacy Reference Architecture

AWS Prescriptive Guidance

AWS Prescriptive Guidance: AWS Privacy Reference Architecture

Table of Contents

AWS Privacy Reference Architecture

Amazon Web Services (contributors)

September 2025 (document history)

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

Notices

This guide is provided for the purposes of information only. It isn't legal advice and shouldn't be relied on as legal advice. AWS encourages its customers to obtain appropriate advice on their implementation of privacy and data protection environments, and more generally, applicable laws relevant to their business.

Customers are responsible for making their own independent assessment of the information in this document. This document: (a) is for informational purposes only, (b) represents current AWS product offerings and practices, which are subject to change without notice, and (c) doesn't create any commitments or assurances from AWS and its affiliates, suppliers, or licensors. AWS products or services are provided "as is" without warranties, representations, or conditions of any kind, whether express or implied.

The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document isn't part of, nor does it modify, any agreement between AWS and its customers.

Introduction

The AWS Privacy Reference Architecture (AWS PRA) provides a set of guidelines specific to the design and configuration of privacy-supporting controls in AWS services. This guide can help you make decisions about people, process, and technology that help support privacy in the AWS Cloud.

The AWS shared responsibility model and privacy

In the AWS Cloud, you share responsibility for security and compliance with AWS. AWS is responsible for security *of* the cloud, which means that AWS is responsible protecting the

infrastructure that runs all of the services offered in the AWS Cloud. You are responsible for security *in* the cloud, which means that you are responsible for configuring and managing AWS services in accordance with security and privacy requirements. For more information, see the AWS shared responsibility model.

AWS services provide capabilities that allow you to implement your own privacy controls in the cloud in order to support your privacy requirements. Your privacy responsibility varies based on many factors, including the AWS services and AWS Regions you choose, the integration of those services into your IT environment, and the laws and regulations applicable to your organization and workload.

When using AWS services, you maintain control over your content. Specifically, *customer content* is defined as software (including machine images), data, text, audio, video, or images that you or any end user transfer to us for processing, storage, or hosting by AWS services in connection with your account. It also includes any computational results that you or an end user derive by using AWS services. You are responsible for managing the following decisions, which are under your control:

- The data you choose to collect, store, or process on AWS
- The AWS services you use with the data
- The AWS Region where you collect, store, or process data
- The format and structure of your data and whether it's masked, anonymized, or encrypted
- How you define, store, rotate, and operate your cryptographic keys for encryption
- Who has access and when they have access to your data, and how those access rights are granted, managed, and revoked

Once you understand the AWS shared responsibility model and how it generally applies to operating in the cloud, you must determine how it applies to your use case. The AWS services that you choose to use determine the amount of configuration you must perform as part of your organization's privacy responsibilities. For example, a service such as Amazon Elastic Compute Cloud (Amazon EC2) is categorized as infrastructure as a service (IaaS). As such, if you use Amazon EC2, you must perform all of the necessary privacy configurations for guest operating systems and for the application software or utilities you install on your EC2 instances. When you use an abstracted service, such as Amazon Simple Storage Service (Amazon S3) and Amazon DynamoDB, AWS is responsible for the infrastructure layer, the operating system, and platforms. Your responsibility is to manage and classify the data (customer content) and to configure the policies

used to access the endpoints in order to store and retrieve data. For more information about how AWS helps you protect data and privacy, see Data protection and privacy at AWS.

Understanding the AWS PRA

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

The section describes the relationship between the AWS Privacy Reference Architecture (AWS PRA) and other AWS guidance. This section also reviews the general layout and structure of the example AWS multi-account environment in the AWS PRA.

This section contains the following topics:

Using the AWS PRA and the AWS SRA

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

The AWS PRA provides patterns that customers have found helpful in planning foundational and application-level privacy controls for their infrastructure and workloads in AWS. The AWS Security Reference Architecture (AWS SRA) provides a set of guidelines for building an architecture that implements and supports the right set of security controls across your AWS landing zone and applications. In order to establish the privacy controls detailed in this guide, the AWS PRA assumes many of the same foundational guidelines and account structure that are described in the AWS SRA. The AWS PRA and AWS SRA detail many of the same key AWS services. This guide includes only brief descriptions of these services. You can learn more about these services and how they're used in a security context in the AWS SRA.

The AWS SRA can help you design, implement, and manage AWS security services so that they align with AWS recommended practices. You can use the AWS SRA as a standalone guide, or you can use the AWS SRA and AWS PRA as companion guides. Many of the security guidelines detailed in the AWS SRA can be followed in tandem with the privacy controls that are detailed in the AWS PRA. Similar to security, there are foundational privacy considerations that can be helpful to make early in your AWS Cloud journey because these decisions can affect the design of the organization's account structure. For example, some questions you might consider include:

- How does my organization define personal data?
- Does my organization support applications that process personal data?
- What about applications that process other types of regulated data?
- What organization-level controls can I implement to keep my developers and cloud engineers as far away from personal data as possible?
- How do I segregate personal data from other types of data?
- What are my organization's cross-border data transfer requirements?

The answers to many of these questions can have implications for the design of your cloud environment, such as your AWS account structure, service control policies, and AWS Identity and Access Management (IAM) roles.

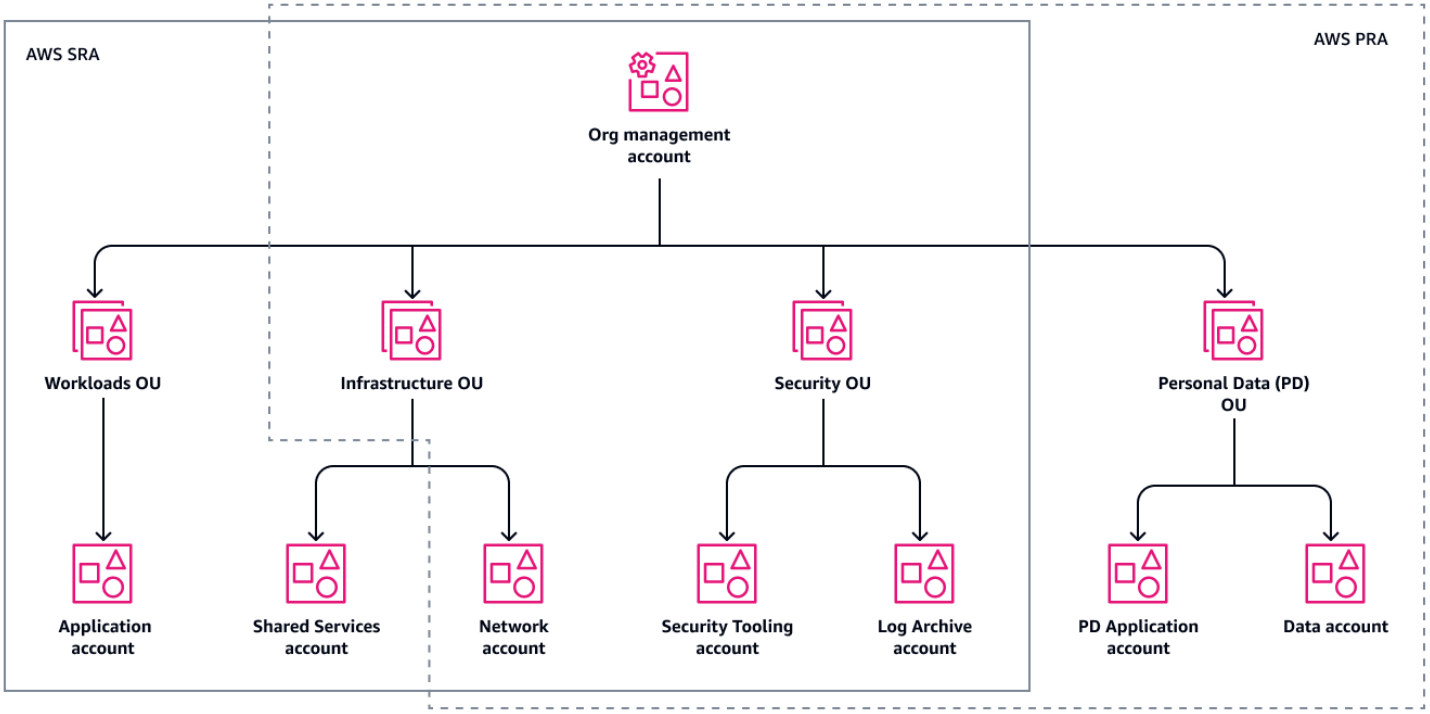
AWS Organizations and the dedicated account structure

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

AWS Organizations is an account management service that helps you centrally manage and govern multiple AWS accounts. Use of AWS Organizations is the basis of a well-architected, multi-account AWS environment. For more information, see Establishing your best practice AWS environment.

The following diagram shows the high-level account and organizational unit (OU) structure of the AWS PRA. For the most part, the organizational structure of the AWS PRA matches the organizational structure of the AWS SRA.



The deviations from the AWS SRA organization include:

- The AWS PRA adds the Personal Data (PD) OU, which is dedicated for collecting, storing, and processing personal data. This structural separation provides flexibility so that you can define specific, fine-grained controls to help protect personal data from unintended disclosure.
- In the Infrastructure OU, the AWS PRA doesn't currently include additional guidance for the Shared Services account that is described in the AWS SRA.
- The AWS PRA doesn't currently include additional guidance for the Workloads OU that is described in the AWS SRA. Applications that collect or process personal data are located in dedicated accounts in the PD OU.

You can use AWS Control Tower for overall foundational governance and automated deployment of security and privacy controls across your organization. If AWS Control Tower isn't in use today in your organization, you can still deploy many of the security and privacy controls in AWS Control Tower, such as service control policies and AWS Config rules, in their respective services.

You might find it helpful to consider the processing of personal data when you plan your account and OU structure, including an account segmentation strategy. You might need to consider the types of data you are processing for their unique use cases and applicable laws and regulations. For example, cardholder data is protected under the Payment Card Industry Data Security Standard

(PCI DSS), and protected health information might be subject to the Health Insurance Portability and Accountability Act (HIPAA). You might want to review which environments contain personal data and plan your segmentation strategy heavily on that. A typical account segmentation strategy can include dedicated AWS accounts that align to the software development lifecycle (SDLC), such as dedicated accounts for development, staging or quality assurance (QA), and production. A segmentation strategy such as this can be a critical component in the overall design discussion, and your OUs might need to align to your specific regulatory requirements.

Some multi-account AWS environments require dedicated application accounts per AWS Region, or they might require multi-account landing zones. In this case, you require additional segmentation to meet the unique data sovereignty requirements for your customers and regulators. For more information, see Strategizing for global expansion in this guide.

Operationalizing AWS privacy services

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

For many, privacy is cross-cutting. Many different teams have a part to play, including regulatory, compliance, and engineering teams. When your organization has started to define the key people and policy components of your privacy program, you can map controls against a privacy compliance framework for consistent operations. A framework can serve as a rubric for implementing foundational and application-specific privacy controls for personal data in your AWS environment.

Regardless of the framework that customers use to categorize their privacy requirements, privacy compliance, privacy engineering, and application teams often need to work together to achieve implementation goals. For example, regulatory and compliance teams might provide the high-level requirements, and engineering and application teams configure AWS services and features to align to these requirements. Starting with a control framework can help you define more prescriptive organizational and technical controls.

When defining the technical controls of AWS services and features, another key decision is whether a control should apply to the entire organization, an OU, an account, or a specific resource. Some services and features are a great fit for implementing controls across your full AWS organization.

For example, blocking public access to Amazon S3 buckets is a specific control that is preferably configured at the organization root rather than individually for each account. However, your retention policies might vary from application to application, which means that you might apply the control at the resource level.

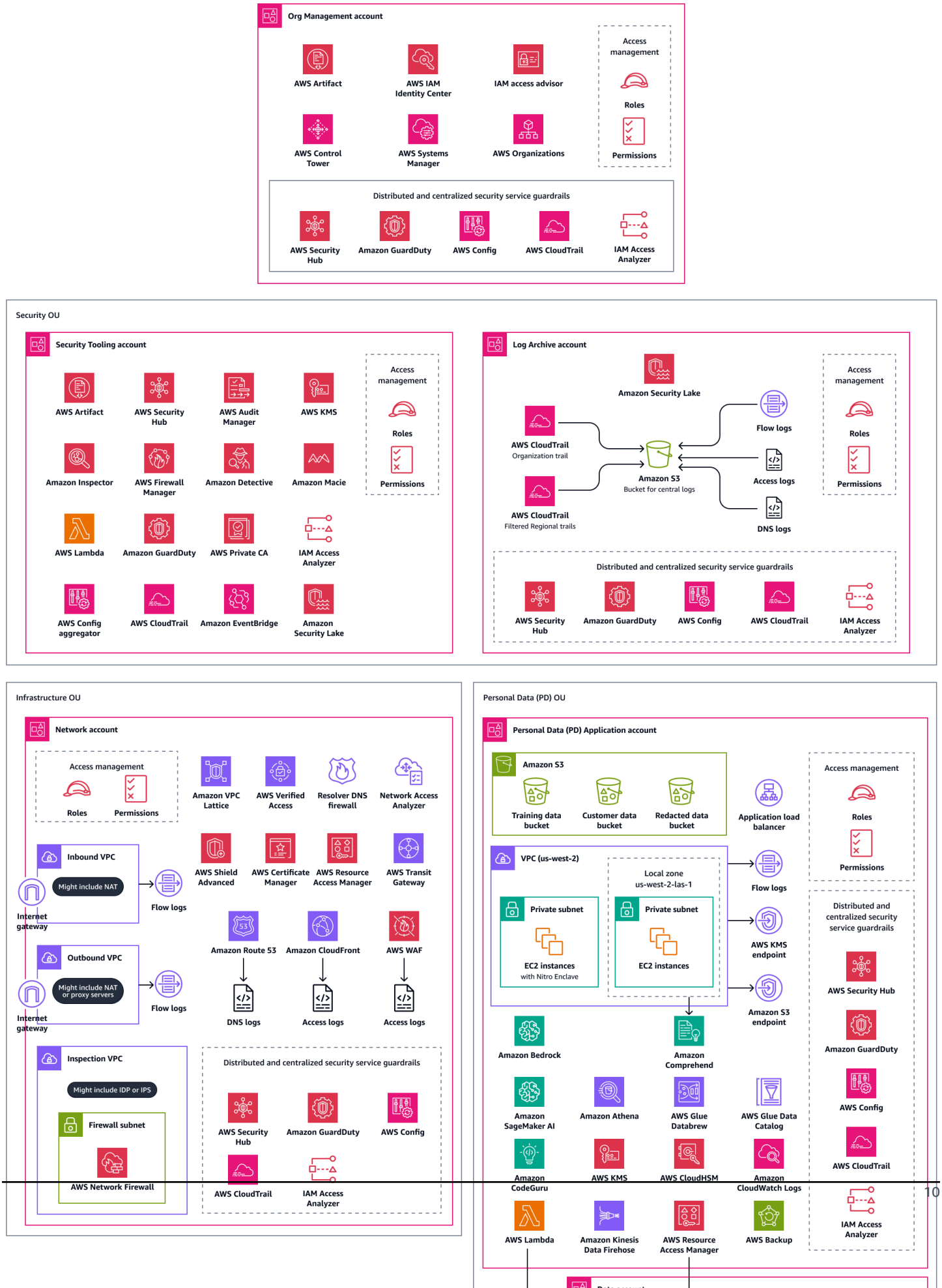
To help you accelerate operationalizing privacy in your organization, AWS offers audit and compliance advisory services for your AWS workloads. For more information, contact AWS SAS.

The AWS Privacy Reference Architecture

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

The following diagram illustrates the AWS Privacy Reference Architecture (AWS PRA). This is an example of an architecture that connects many privacy-related AWS services and features. This architecture is built on a landing zone that is governed by AWS Control Tower.


Org Management account
AWS Artifact
AWS IAM Identity Center
IAM access advisor
Access management
Roles
Permissions
AWS Control Tower
AWS Systems Manager
AWS Organizations
Distributed and centralized security service guardrails
AWS Security Hub
Amazon GuardDuty
AWS Config
AWS CloudTrail
IAM Access Analyzer
Security OU
Security Tooling account
AWS Artifact
AWS Security Hub
AWS Audit Manager
AWS KMS
Access management
Roles
Permissions
Amazon Inspector
AWS Firewall Manager
Amazon Detective
Amazon Macie
AWS Lambda
Amazon GuardDuty
AWS Private CA
IAM Access Analyzer
AWS Config aggregator
AWS CloudTrail
Amazon EventBridge
Amazon Security Lake
Log Archive account
Amazon Security Lake
AWS CloudTrail
Organization trail
Flow logs
Amazon S3
Bucket for central logs
Access logs
AWS CloudTrail
Filtered Regional trails
DNS logs
Access management
Roles
Permissions
Distributed and centralized security service guardrails
AWS Security Hub
Amazon GuardDuty
AWS Config
AWS CloudTrail
IAM Access Analyzer
Infrastructure OU
Network account
Access management
Roles
Permissions
Amazon VPC Lattice
AWS Verified Access
Resolver DNS firewall
Network Access Analyzer
Inbound VPC
Might include NAT
Internet gateway
Flow logs
AWS Shield Advanced
AWS Certificate Manager
AWS Resource Access Manager
AWS Transit Gateway
Outbound VPC
Might include NAT or proxy servers
Internet gateway
Flow logs
Amazon Route 53
Amazon CloudFront
AWS WAF
DNS logs
Access logs
Access logs
Inspection VPC
Might include IDP or IPS
Firewall subnet
AWS Network Firewall
Distributed and centralized security service guardrails
AWS Security Hub
Amazon GuardDuty
AWS Config
AWS CloudTrail
IAM Access Analyzer
Personal Data (PD) OU
Personal Data (PD) Application account
Amazon S3
Training data bucket
Customer data bucket
Redacted data bucket
Application load balancer
Access management
Roles
Permissions
VPC (us-west-2)
Local zone us-west-2-las-1
Private subnet
EC2 instances with Nitro Enclave
Private subnet
EC2 instances
Flow logs
AWS KMS endpoint
Amazon S3 endpoint
Distributed and centralized security service guardrails
AWS Security Hub
Amazon GuardDuty
AWS Config
AWS CloudTrail
IAM Access Analyzer
Amazon Bedrock
Amazon Comprehend
Amazon SageMaker AI
Amazon Athena
AWS Glue Databrew
AWS Glue Data Catalog
Amazon CodeGuru
AWS KMS
AWS CloudHSM
Amazon CloudWatch Logs
AWS Lambda
Amazon Kinesis Data Firehose
AWS Resource Access Manager
AWS Backup

The AWS PRA includes a serverless web architecture that is hosted in the Personal Data (PD) Application account. The architecture in this account is an example workload that collects personal data directly from consumers. In this workload, users connect through a web tier. The web tier interacts with the application tier. This tier receives inputs from the web tier, processes and stores the data, allows authorized internal teams and third parties to access the data, and eventually archives and deletes the data when it's no longer required. The architecture is purposefully modular and event-driven in order to demonstrate many of the foundational privacy engineering techniques without delving into specific use cases, such as data lakes, containers, compute, or Internet of Things (IoT).

Next, this guide describes each account in the organization in detail. It discusses the privacy-related services and features, considerations and recommendations, and diagrams for each of the following accounts:

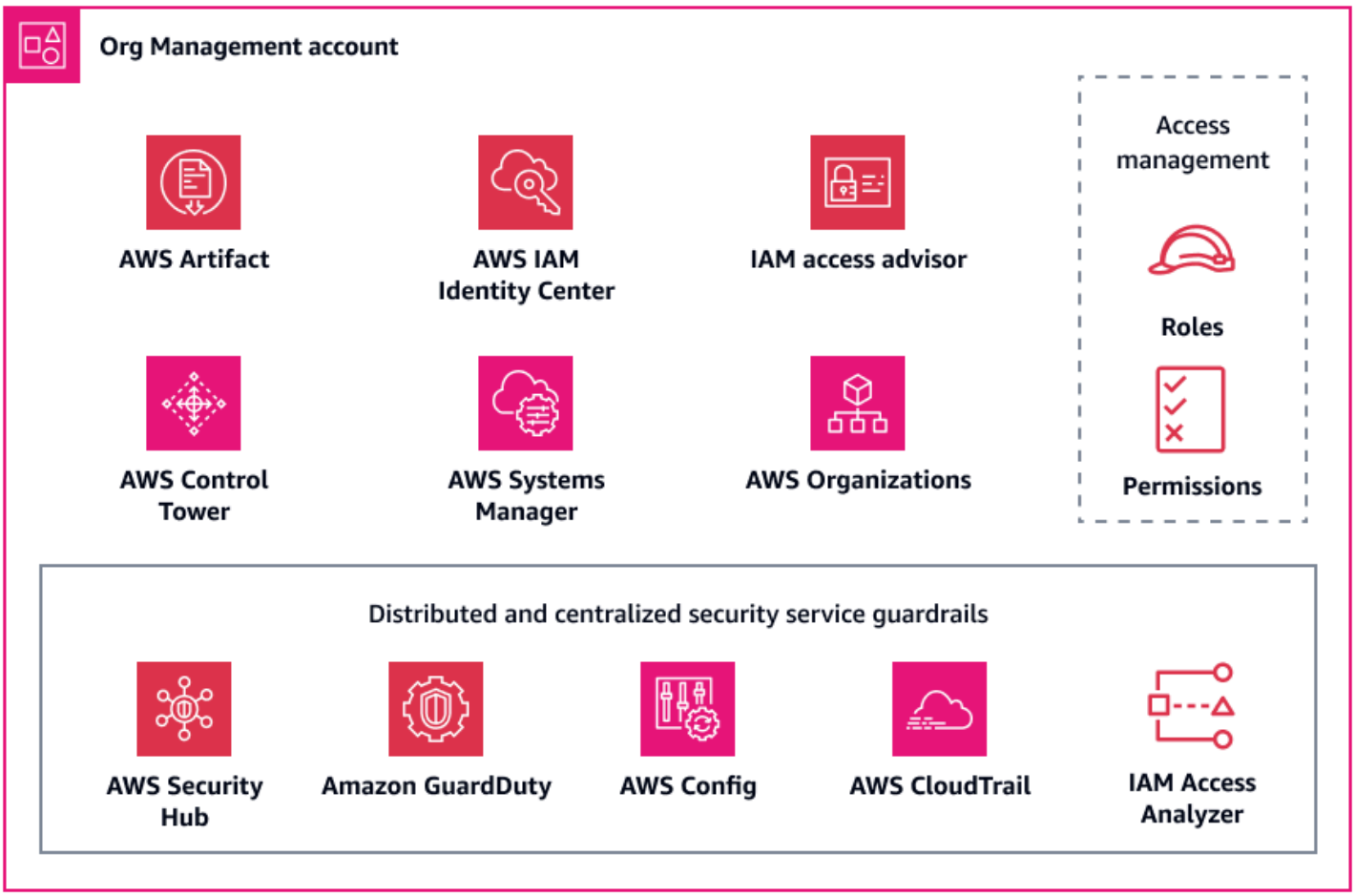
- Org Management account
- Security OU – Security Tooling account
- Security OU – Log Archive account
- Infrastructure OU – Network account
- Personal Data OU – PD Application account

Org Management account

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

The Org Management account is primarily used to manage resource configuration drift for the foundational privacy controls across all of the accounts in your organization, which is managed by AWS Organizations. This account is also where you can deploy new member accounts consistently, with many of the same security and privacy controls. For more information about this account, see the AWS Security Reference Architecture (AWS SRA). The following diagram illustrates the AWS security and privacy services that are configured in the Org Management account.



This section provides more detailed information about the following AWS services that are used in this account:

- AWS Artifact
- AWS Control Tower
- AWS Organizations

AWS Artifact

AWS Artifact can help you with audits by providing on-demand downloads of AWS security and compliance documents. For more information about how this service is used in a security context, see the AWS Security Reference Architecture.

This AWS service helps you understand the controls that you inherit from AWS and determine what controls may be remaining for you to implement in your environment. AWS Artifact provides access to AWS security and compliance reports, such as System and Organization Controls (SOC)

reports and payment card industry (PCI) reports. It also provides access to certifications from accreditation bodies across geographies and compliance verticals that validate the implementation and operating effectiveness of AWS controls. Using AWS Artifact, you can provide the AWS audit artifacts to your auditors or regulators as evidence of AWS security and privacy controls. The following reports might be useful to demonstrate the effectiveness of AWS privacy controls:

- **SOC 2 Type 2 Privacy report** – This report demonstrates the effectiveness of AWS controls for how personal data is collected, used, retained, disclosed, and disposed of. There is also a SOC 3 Privacy report, which is a less detailed description of the SOC 2 privacy controls. For more information, see the SOC FAQ.
- **Cloud Computing Compliance Controls Catalog (C5)** – This report was created by Germany's national cybersecurity authority, Bundesamt für Sicherheit in der Informationstechnik (BSI). It details the security controls that AWS implemented in order to meet the C5 requirements. It also includes additional control requirements for privacy relating to data location, service provisioning, place of jurisdiction, and information disclosure obligations.
- **ISO/IEC 27701:2019 certification report** –-ISO/IEC 27701:2019 describes requirements and guidelines to establish and continuously improve a privacy information management system (PIMS). This report details the scope of this certification and can serve as proof of AWS certification. For more information about this standard, see ISO/IEC 27701:2019 (ISO website).

AWS Control Tower

AWS Control Tower helps you set up and govern an AWS multi-account environment that follows prescriptive security recommended practices. For more information about how this service is used in a security context, see the AWS Security Reference Architecture.

In AWS Control Tower, you can also automate the deployment of many proactive, preventative, and detective controls, also known as *guardrails*, that align to your data privacy requirements, specifically for data residency and sovereignty. For example, you can specify guardrails that limit the transfer of data to only approved AWS Regions. For even more granular control, you can choose from more than 17 guardrails that are designed to control data residency, such as *Disallow Amazon Virtual Private Network (VPN) connections*, *Disallow internet access for an Amazon VPC instance*, and *Deny access to AWS based on the requested AWS Region*. These guardrails consist of a number of AWS CloudFormation hooks, service control policies, and AWS Config rules that can be uniformly deployed across your organization. For more information, see Controls that enhance data residency protection in the AWS Control Tower documentation.

For data sovereignty, AWS Control Tower currently provides preventative controls, such as *Require that an attached Amazon EBS volume is configured to encrypt data at rest* and *Require an AWS KMS key policy to have a statement that limits creation of AWS KMS grants to AWS services*. Sovereignty controls are broader than just data residency controls. They help prevent actions that might violate data residency, granular access restriction, encryption, and resilience requirements. For more information, see Preventive controls that assist with digital sovereignty in the AWS Control Tower documentation.

If you need to deploy privacy guardrails beyond data residency and sovereignty controls, AWS Control Tower includes a number of mandatory controls. These controls are deployed by default across every OU when you set up your landing zone. Many of these are preventative controls that are designed to protect logs, such as *Disallow Deletion of Log Archive* and *Enable Integrity Validation for CloudTrail Log File*.

AWS Control Tower is also integrated with AWS Security Hub CSPM to provide detective controls. These controls are known as Service-Managed Standard: AWS Control Tower. You can use these controls to monitor for configuration drift of privacy-supporting controls, such as encryption at rest for Amazon Relational Database Service (Amazon RDS) database instances.

AWS Organizations

The AWS PRA uses AWS Organizations to centrally manage all accounts within the architecture. For more information, see AWS Organizations and the dedicated account structure in this guide. In AWS Organizations, you can use service control policies (SCPs) and management policies to help protect personal data and privacy.

Service control policies (SCPs)

Service control policies (SCPs) are a type of organization policy that you can use to manage permissions in your organization. They provide centralized control over the maximum available permissions for AWS Identity and Access Management (IAM) roles and users in the target account, organization unit (OU), or entire organization. You can create and apply SCPs from the Org Management account.

You can use AWS Control Tower to deploy SCPs uniformly across your accounts. For more information about the data residency controls you can apply through AWS Control Tower, see AWS Control Tower in this guide. AWS Control Tower includes a full complement of preventative SCPs. If AWS Control Tower isn't currently used in your organization, you can also deploy these controls manually.

Using SCPs to address data residency requirements

It's common to manage personal data residency requirements by storing and processing data within a specific geographic region. In order to verify that a jurisdiction's unique data residency requirements are met, we recommend that you work closely with your regulatory team to confirm your requirements. When these requirements have been determined, there are a number of AWS foundational privacy controls that can help support. For example, you can use SCPs to limit which AWS Regions can be used to process and store data. For a sample policy, see Restrict data transfers across AWS Regions in this guide.

Using SCPs to restrict high-risk API calls

It is important to understand which security and privacy controls AWS is responsible for and which ones you are responsible for. For example, you are responsible for the results of API calls that could be made against the AWS services that you use. You are also responsible for understanding which of those calls could result in changes to your security or privacy posture. If you are concerned about maintaining a certain security and privacy posture, you can enable SCPs that deny certain API calls. These API calls may have implications, such as unintended disclosure of personal data or violations of specific cross-border data transfer. For example, you might want to prohibit the following API calls:

- Enabling public access to Amazon Simple Storage Service (Amazon S3) buckets
- Disabling Amazon GuardDuty or creating suppression rules for data exfiltration findings, such as the Trojan:EC2/DNSDataExfiltration finding
- Deleting AWS WAF data exfiltration rules
- Sharing Amazon Elastic Block Store (Amazon EBS) snapshots publicly
- Removing a member account from the organization
- Disassociating Amazon CodeGuru Reviewer from a repository

Management policies

Management policies in AWS Organizations can help you centrally configure and manage AWS services and their features. The types of management policy you choose determine how policies affect the OUs and accounts that inherit them. Tag policies are an example of a management policy in AWS Organizations that directly relates to privacy.

Using tag policies

Tags are key value pairs that help you manage, identify, organize, search for, and filter AWS resources. It can be useful to apply tags that distinguish the resources in your organization that handle personal data. The use of tags supports many of the privacy solutions in this guide. For example, you might want to apply a tag that indicates the general data classification of the data being processed or stored within the resource. You can write attribute-based access control (ABAC) policies that limit access to resources that have a particular tag or set of tags. For example, your policy might specify that the `SysAdmin` role can't access resources that have the `dataclassification:4` tag. For more information and a tutorial, see Define permissions to access AWS resources based on tags in the IAM documentation. In addition, if your organization uses AWS Backup to apply data retention policies broadly across your backups in many accounts, you can apply a tag that puts that resource within scope for that backup policy.

Tag policies help you maintain consistent tags throughout your organization. In a tag policy, you specify rules that apply to resources when they are tagged. For example, you can require resources to be tagged with specific keys, such as `DataClassification` or `DataSteward`, and you can specify valid case treatments or values for keys. You can also use enforcement to prevent noncompliant tagging requests from completing.

When using tags as a core component of your privacy control strategy, consider the following:

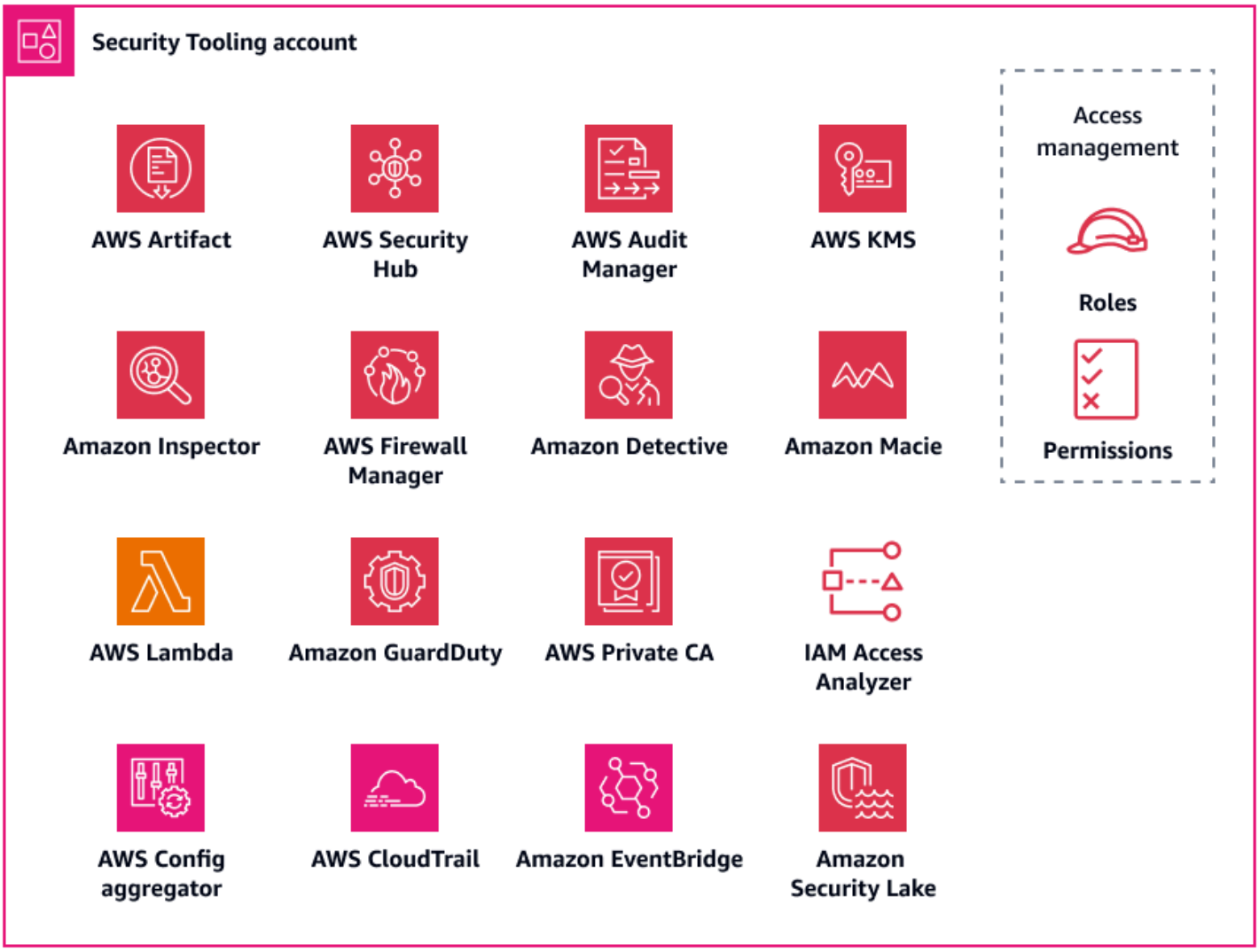
- Consider the implications of placing personal data or other types of sensitive data within tag keys or values. When you contact AWS for technical assistance, AWS might analyze tags and other resource identifiers to help resolve the issue. Tag data is not encrypted, and AWS services, such as AWS Billing and Cost Management, can read them. Therefore, you might want to deidentify tag values and then reidentify them by using a system that you control, such as an IT service management (ITSM) system. AWS recommends not including personally identifiable information in tags.
- Consider that some tag values need to be made immutable (unmodifiable) to prevent circumvention of technical controls, such as ABAC conditions that rely on tags.

Security OU – Security Tooling account

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

The Security Tooling account is dedicated to operating security and privacy foundational services, monitoring AWS accounts, and automating security and privacy alerting and response. For more information about this account, see the AWS Security Reference Architecture (AWS SRA). The following diagram illustrates the AWS security and privacy services that are configured in the Security Tooling account.



This section provides more detailed information about the following in this account:

- AWS CloudTrail
- AWS Config
- Amazon GuardDuty
- IAM Access Analyzer
- Amazon Macie

AWS CloudTrail

AWS CloudTrail helps you audit the overall API activity in your AWS account. Enabling CloudTrail in all AWS accounts and AWS Regions that store, process, or transmit personal data can help you track the use and disclosure of this data. The AWS Security Reference Architecture recommends enabling an organization trail, which is a single trail that logs all events for all accounts in the organization. However, enabling this organization trail aggregates the multi-Region log data into a single Amazon Simple Storage Service (Amazon S3) bucket in the Log Archive account. For accounts that handle personal data, this can bring some additional design considerations. Log records might contain some references to personal data. To meet your data residency and data transfer requirements, you might need to reconsider aggregating cross-Region log data into a single Region where the S3 bucket is located. Your organization might consider which regional workloads should be included or excluded from the organization trail. For workloads that you decide to exclude from the organization trail, you could consider configuring a Region-specific trail that masks personal data. For more information about masking personal data, see the Amazon Data Firehose section of this guide. Ultimately, your organization might have a combination of organization trail and regional trails that aggregate into the centralized Log Archive account.

For more information about configuring a single-Region trail, see the instructions for using the AWS Command Line Interface (AWS CLI) or the console. When you create the organization trail, you can use an opt-in setting in AWS Control Tower, or you can create the trail directly in the CloudTrail console.

For more information on the overall approach and how to manage centralization of logs and data transfer requirements, see the Centralized log storage section in this guide. Whatever configuration you choose, you might want to separate trail management in the Security Tooling account from log storage in the Log Archive account, according to the AWS SRA. This design helps you create least-privilege access policies for those who need to manage logs and those who need to use the log data.

AWS Config

AWS Config provides a detailed view of the resources in your AWS account and how they're configured. It helps you identify how resources relate to one another and how their configurations have changed over time. For more information about how this service is used in a security context, see the AWS Security Reference Architecture.

In AWS Config, you can deploy conformance packs, which are sets of AWS Config rules and remediation actions. Conformance packs provide a general-purpose framework that is designed to enable privacy, security, operational, and cost-optimization governance checks by using managed or custom AWS Config rules. You can use this tool as a part of a larger set of automation tools to track whether your AWS resource configurations comply with your own control framework requirements.

The Operational Best Practices for NIST Privacy Framework v1.0 conformance pack is aligned to a number of privacy-related controls in the NIST Privacy Framework. Each AWS Config rule applies to a specific AWS resource type, and it relates to one or more NIST Privacy Framework controls. You can use this conformance pack to track privacy-related continuous compliance across resources in your accounts. The following are some of the rules included in this conformance pack:

- `no-unrestricted-route-to-igw` – This rule helps prevent data exfiltration on the data plane by continually monitoring VPC route tables for default `0.0.0.0/0` or `::/0` egress routes to an internet gateway. This helps you restrict where internet-bound traffic can be sent, especially if there are CIDR ranges that are known to be malicious.
- `encrypted-volumes` – This rule checks whether Amazon Elastic Block Store (Amazon EBS) volumes that are attached to Amazon Elastic Compute Cloud (Amazon EC2) instances are encrypted. If your organization has specific control requirements that pertain to the usage of AWS Key Management Service (AWS KMS) keys for protection of personal data, you can specify specific key IDs as a part of the rule to check that the volumes are encrypted with a specific AWS KMS key.
- `restricted-common-ports` – This rule checks whether Amazon EC2 security groups allow unrestricted TCP traffic to specified ports. Security groups can help you manage network access by providing stateful filtering of ingress and egress network traffic to AWS resources. Blocking ingress traffic from `0.0.0.0/0` to common ports, such as TCP 3389 and TCP 21, on your resources helps you restrict remote access.

AWS Config can be used for both proactive and reactive compliance checks of your AWS resources. In addition to considering the rules found in the conformance packs, you can incorporate these rules in both detective and proactive evaluation modes. This helps implement privacy checks earlier in your software development lifecycle because the application developers can start incorporating predeployment checks. For example, they can include hooks in their AWS CloudFormation templates that check the declared resource in the template against all privacy-related AWS Config rules that have proactive mode enabled. For more information, see AWS Config Rules Now Support Proactive Compliance (AWS blog post).

Amazon GuardDuty

AWS offers multiple services that might be used to store or process personal data, such as Amazon S3, Amazon Relational Database Service (Amazon RDS), or Amazon EC2 with Kubernetes. Amazon GuardDuty combines intelligent visibility with continuous monitoring to detect indicators that might be related to unintended disclosure of personal data. For more information about how this service is used in a security context, see the AWS Security Reference Architecture.

With GuardDuty, you can identify potentially malicious, privacy-related activity throughout an attack lifecycle. For example, GuardDuty can alert you about connections to blacklisted sites, unusual network port traffic or traffic volumes, DNS exfiltration, unexpected EC2 instance launches, and unusual ISP callers. You can also configure GuardDuty to stop alerts for trusted IP addresses from your own *trusted IP lists* and alert on known malicious IP addresses from your own *threat lists*.

As recommended in the AWS SRA, you can enable GuardDuty for all AWS accounts in your organization and configure the Security Tooling account as the GuardDuty delegated administrator. GuardDuty aggregates findings from across the organization into this single account. For more information, see Managing GuardDuty accounts with AWS Organizations. You can also consider identifying all privacy-related stakeholders in the incident response process, from detection and analysis to containment and eradication, and involving them in any incidents that might involve data exfiltration.

IAM Access Analyzer

Many customers want continual assurance that personal data is being shared appropriately with preapproved and intended third-party processors and no other entities. A data perimeter is a set of preventive guardrails designed to allow only trusted identities from expected networks to access trusted resources in your AWS environment. As you define controls for unintended and intended

disclosure of personal data, you can define trusted identities, trusted resources, and expected networks.

With AWS Identity and Access Management Access Analyzer (IAM Access Analyzer), organizations can define an AWS account zone of trust and configure alerting for violations to that zone of trust. IAM Access Analyzer analyzes IAM policies to help identify and resolve unintended public or cross-account access to potentially sensitive resources. IAM Access Analyzer uses mathematical logic and inference to generate comprehensive findings for resources that can be accessed from outside an AWS account. Finally, for responding to and remediating overly-permissive IAM policies, you can use IAM Access Analyzer to validate existing policies against IAM recommended practices and provide suggestions. IAM Access Analyzer can generate a least-privilege IAM policy that is based on an IAM principal's prior access activity. It analyzes CloudTrail logs and generates a policy that grants only the permissions required to continue performing those tasks.

For more information about how IAM Access Analyzer is used in a security context, see the AWS Security Reference Architecture.

Amazon Macie

Amazon Macie is a service that uses machine learning and pattern matching to discover sensitive data, provides visibility into data security risks, and helps you automate protections against those risks. Macie generates findings when it detects potential policy violations or issues with the security or privacy of your Amazon S3 buckets. Macie is another tool that organizations can use to implement automation in order to support compliance efforts. For more information about how this service is used in a security context, see the AWS Security Reference Architecture.

Macie can detect a large and growing list of sensitive data types, including personally identifiable information (PII), such as names, addresses, and other identifiable attributes. You can even create custom data identifiers in order to define detection criteria that reflects your organization's definition of personal data.

As your organization defines preventative controls for your Amazon S3 buckets that contain personal data, you can use Macie as a validation mechanism to provide continual reassurance of where your personal data lives and how it's protected. To start, enable Macie and configure automated sensitive data discovery. Macie continually analyzes objects in all of your S3 buckets, across accounts and AWS Regions. Macie generates and maintains an interactive heat map that depicts where personal data resides. The automated sensitive data discovery feature is designed to reduce costs and minimize the need to manually configure discovery jobs. You can build on top of

the automated sensitive data discovery feature and use Macie to automatically detect new buckets or new data in existing buckets and then validate the data against the assigned data classification tags. Configure this architecture to notify the appropriate development and privacy teams about misclassified or unclassified buckets in a timely manner.

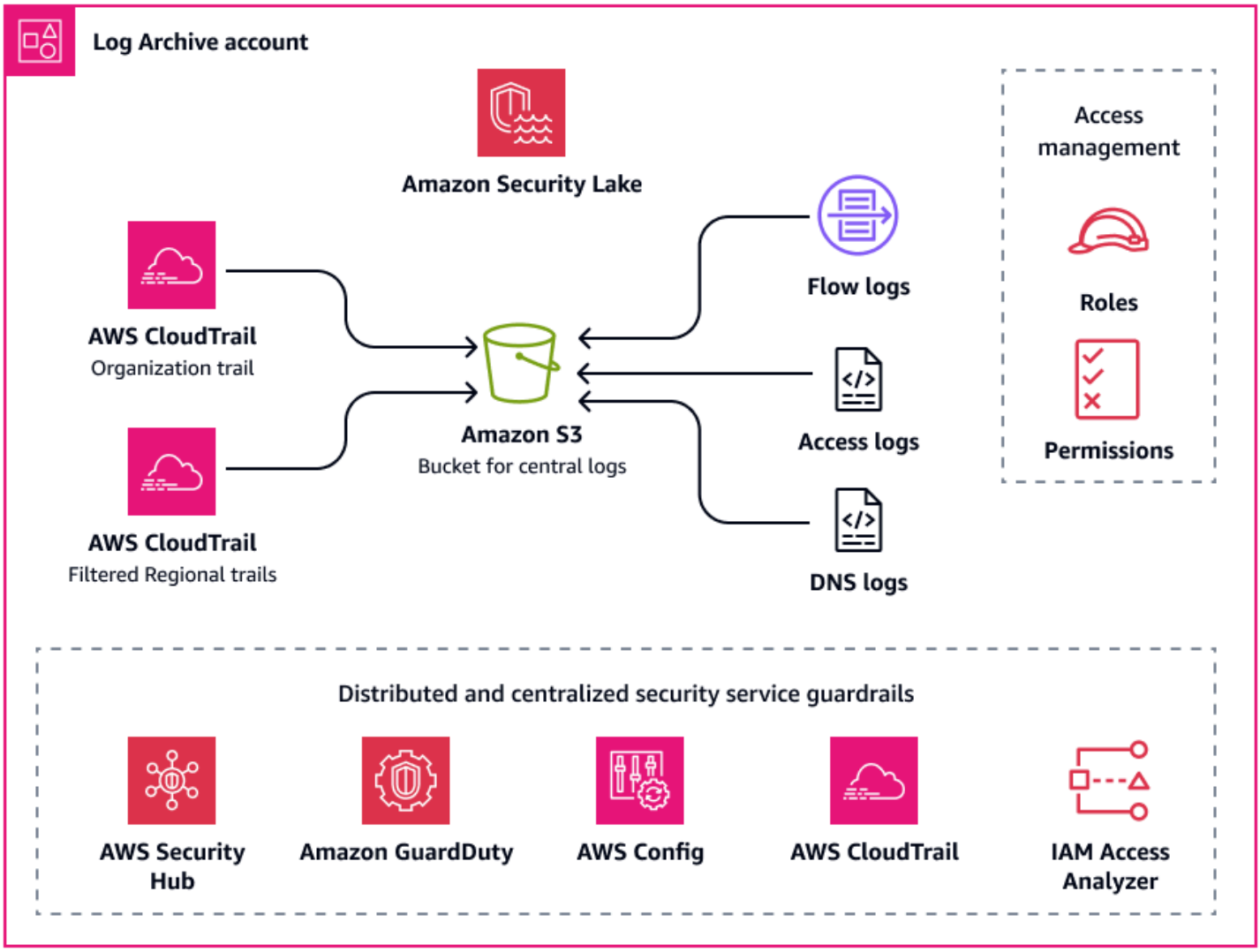
You can enable Macie for every account in your organization by using AWS Organizations. For more information, see Integrating and configuring an organization in Amazon Macie.

Security OU – Log Archive account

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

The Log Archive account is where you centralize infrastructure, service, and application log types. For more information about this account, see the AWS Security Reference Architecture (AWS SRA). With a dedicated account for logs, you can apply consistent alerting across all log types and to confirm that incident responders can access an aggregate of these logs from one place. You can set up security controls and data retention policies all from one place as well, which can simplify the privacy operational overhead. The following diagram illustrates the AWS security and privacy services that are configured in the Log Archive account.



Centralized log storage

Log files (such as AWS CloudTrail logs) might contain information that could be considered personal data. Some organizations choose to use an organization trail in order to aggregate CloudTrail logs across AWS Regions and across accounts into one central location, for visibility purposes. For more information, see AWS CloudTrail in this guide. When implementing centralization of CloudTrail logs, the logs are typically stored in an Amazon Simple Storage Service (Amazon S3) bucket in a single Region.

Depending on your organization's definition of personal data, your contractual obligations to your customers, and applicable regional privacy regulations, you might need to consider cross-border data transfers when it comes to log aggregation. Determine if the personal data within the various log types falls under these restrictions. For example, CloudTrail logs might contain your

organization's employee data, but they might not contain your customers' personal data. If your organization needs to adhere to restricted data transfer requirements, the following options can help support:

- If your organization is providing services in the AWS Cloud to data subjects in multiple countries, you might choose to aggregate all logs in the country that has the most stringent data residency requirements. For example, if you're operating in Germany and it has the most stringent requirements, you might aggregate data in an S3 bucket in the `eu-central-1` AWS Region so that data collected in Germany doesn't leave the borders of Germany. For this option, you can configure a single organization trail in CloudTrail that aggregates logs from across all accounts and AWS Regions into the target Region.
- Redact the personal data that needs to stay in the AWS Region before the data is copied and aggregated to another region. For example, you can mask the personal data in the application's host Region before you transfer the logs to a different Region. For more information about masking personal data, see the Amazon Data Firehose section of this guide.
- If you have stringent data sovereignty concerns, you can maintain a separate multi-account landing zone in the AWS Region that enforces these requirements. This way, you can simplify the landing zone configuration in the Region for centralized logging. It also provides additional infrastructure segregation benefits and helps keep log local to their own Region. Work with your legal counsel to determine which personal data is in scope and which Region-to-Region transfers are allowed. For more information, see Strategizing for global expansion in this guide.

Through service logs, application logs, and operating system (OS) logs, you can use Amazon CloudWatch to monitor AWS services or resources in their corresponding account and Region by default. Many choose to centralize these logs and metrics from multiple accounts and Regions into a single account. By default, these logs persist in their corresponding account and Region where they originate. For centralization, you can use subscription filters and Amazon S3 export tasks to share data into a centralized location. It might be important to include the proper filters and export tasks when aggregating logs from a workload that has cross-border data transfer requirements. If the access logs of a workload contain personal data, you might need to make sure that these are transferred to or retained in specific accounts and Regions.

Amazon Security Lake

As recommended in the AWS SRA, you might want to use the Log Archive account as the delegated administrator account for Amazon Security Lake. When you do this, Security Lake collects

supported logs in dedicated Amazon S3 buckets in the same account as other SRA-recommended security logs.

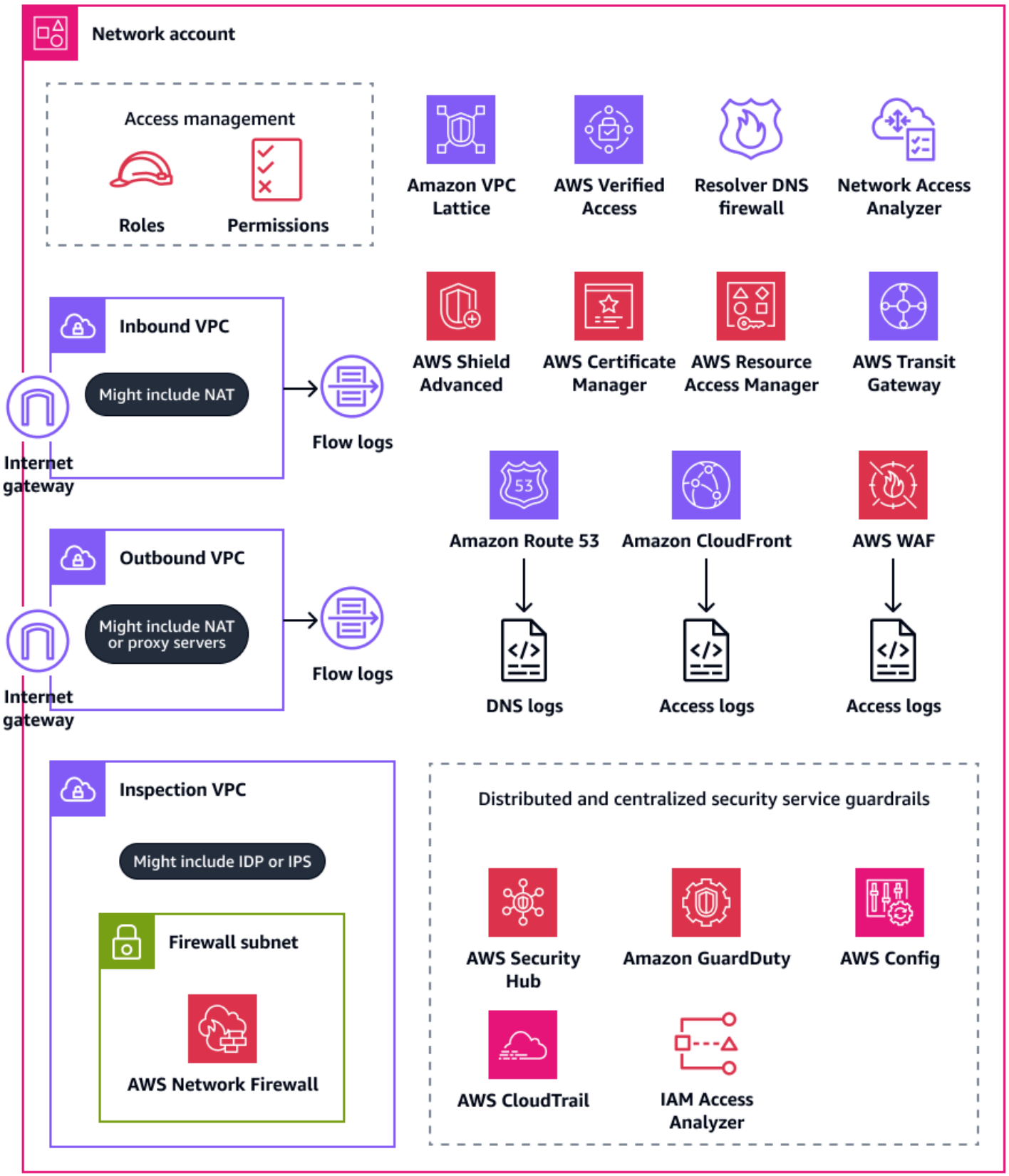
From a privacy perspective, it is important for your incident responders to have access to logs from your AWS environments, SaaS providers, on premises, cloud sources, and third-party sources. This helps them more quickly block and remediate unauthorized access to personal data. The same considerations for log storage most likely apply to log residency and Regional movement within Amazon Security Lake. This is because Security Lake collects security logs and events from the AWS Regions in which you've enabled the service. To comply with data residency requirements, consider your configuration of rollup Regions. A *rollup Region* is a Region where Security Lake consolidates data from one or more contributing Regions, which you select. Your organization might need to align on your Regional compliance requirements for data residency before you can configure Security Lake and rollup Regions.

Infrastructure OU – Network account

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

In the Network account, you manage the networking between your virtual private clouds (VPCs) and the broader internet. In this account, you can implement broad disclosure control mechanisms by using AWS WAF, use AWS Resource Access Manager (AWS RAM) to share VPC subnets and AWS Transit Gateway attachments, and use Amazon CloudFront to support targeted service usage. For more information about this account, see the AWS Security Reference Architecture (AWS SRA). The following diagram illustrates the AWS security and privacy services that are configured in the Network account.


Network account
Access management
Roles
Permissions
Amazon VPC Lattice
AWS Verified Access
Resolver DNS firewall
Network Access Analyzer
AWS Shield Advanced
AWS Certificate Manager
AWS Resource Access Manager
AWS Transit Gateway
Inbound VPC
Might include NAT
Internet gateway
Flow logs
Amazon Route 53
Amazon CloudFront
AWS WAF
Outbound VPC
Might include NAT or proxy servers
Internet gateway
Flow logs
DNS logs
Access logs
Access logs
Inspection VPC
Might include IDP or IPS
Firewall subnet
AWS Network Firewall
Distributed and centralized security service guardrails
AWS Security Hub
Amazon GuardDuty
AWS Config
AWS CloudTrail
IAM Access Analyzer

This section provides more detailed information about the following AWS services that are used in this account:

- Amazon CloudFront
- AWS Resource Access Manager
- AWS Transit Gateway
- AWS WAF

Amazon CloudFront

Amazon CloudFront supports geographic restrictions for frontend applications and file hosting. CloudFront can deliver content through a worldwide network of data centers that are called *edge locations*. When a user requests content that you're serving with CloudFront, the request is routed to the edge location that provides the lowest latency. For more information about how this service is used in a security context, see the AWS Security Reference Architecture.

Your privacy program might currently support compliance with specific regional laws. If your workload is scoped to provide services only to customers who reside only within these regions, you might implement technical measures that prevent usage from other regions. You can use CloudFront geographic restrictions to prevent users in specific geographic locations from accessing content that you are distributing through a CloudFront distribution. For more information and configuration options for geographic restrictions, see Restricting the geographic distribution of your content in the CloudFront documentation.

You can also configure CloudFront to generate access logs that contain detailed information about every user request that CloudFront receives. For more information, see Configuring and using standard logs (access logs) in the CloudFront documentation. Finally, if CloudFront is configured to cache content at a series of edge locations, you might consider where caching occurs. For some organizations, cross-Regional caching might be subject to cross-border data transfer requirements.

AWS Resource Access Manager

AWS Resource Access Manager (AWS RAM) helps you securely share your resources across AWS accounts to reduce operational overhead and provide visibility and auditability. With AWS RAM, organizations can restrict which AWS resources can be shared with other AWS accounts in their organization or with third-party accounts. For more information, see Shareable AWS resources. In the Network account, you can use AWS RAM to share VPC subnets and transit gateway connections. If you use AWS RAM to share a data plane connection with another AWS account, you might

consider establishing processes to check that the connections are made to preapproved AWS Regions and adhere to your data residency requirements.

In addition to sharing VPCs and transit gateway connections, AWS RAM can be used to share resources that don't support IAM resource-based policies. For a workload hosted in the Personal Data OU, you can use AWS RAM to access personal data that is located in a separate AWS account. For more information, see AWS Resource Access Manager in the *Personal Data OU – PD Application account* section.

AWS Transit Gateway

If you want to deploy AWS resources that collect, store, or process personal data in AWS Regions that align with your organizational data residency requirements and you have the appropriate technical safeguards, consider implementing guardrails to prevent unapproved cross-border data flows on the control and data planes. On the control plane, you can limit Region usage and, as a result, cross-Region data flows by using IAM and service control policies.

There are multiple options for controlling cross-Region data flows on the data plane. For example, you can use route tables, VPC peering, and AWS Transit Gateway attachments. AWS Transit Gateway is a central hub that connects virtual private clouds (VPCs) and on-premises networks. As a part of your larger AWS landing zone, you can consider the various ways data can traverse AWS Regions, including out through internet gateways, through direct VPC-to-VPC peering, and through inter-Region peering with AWS Transit Gateway. For example, you can do the following in AWS Transit Gateway:

- Confirm that the east-west and north-south connections between your VPCs and on-premises environments are aligned with your privacy requirements.
- Configure VPC settings according to your privacy requirements.
- Use a service control policy in AWS Organizations and IAM policies to help prevent modifications to your AWS Transit Gateway and Amazon Virtual Private Cloud (Amazon VPC) configurations. For a sample service control policy, see Restrict changes to VPC configurations in this guide.

AWS WAF

To help prevent unintended disclosure of personal data, you can deploy a defense-in-depth approach for your web applications. You can build input validation and rate limiting into your application, but AWS WAF can serve as another line of defense. AWS WAF is a web application

firewall that helps you monitor HTTP and HTTPS requests that are forwarded to your protected web application resources. For more information about how this service is used in a security context, see the AWS Security Reference Architecture.

With AWS WAF, you can define and deploy rules that inspect for specific criteria. The following activities might be associated with unintended disclosure of personal data:

- Traffic from unknown or malicious IP addresses or geographical locations
- Open Worldwide Application Security Project (OWASP) Top 10 attacks, including exfiltration-related attacks such as SQL injection
- High rates of requests
- General bot traffic
- Content scrapers

You can deploy AWS WAF rule groups that are managed by AWS. Some managed rule groups for AWS WAF can be used to detect threats to privacy and personal data, for example:

- SQL database – This rule group contains rules designed to block request patterns associated with exploitation of SQL databases, such as SQL injection attacks. Consider this rule group if your application interfaces with a SQL database.
- Known bad inputs – This rule group contains rules designed to block request patterns that are known to be invalid and are associated with exploitation or discovery of vulnerabilities.
- Bot Control – This rule group contains rules designed to manage requests from bots, which can consume excess resources, skew business metrics, cause downtime, and perform malicious activities.
- Account takeover prevention (ATP) – This rule group contains rules designed to prevent malicious account takeover attempts. This rule group inspects the login attempts sent to your application's login endpoint.

Personal Data OU – PD Application account

Survey

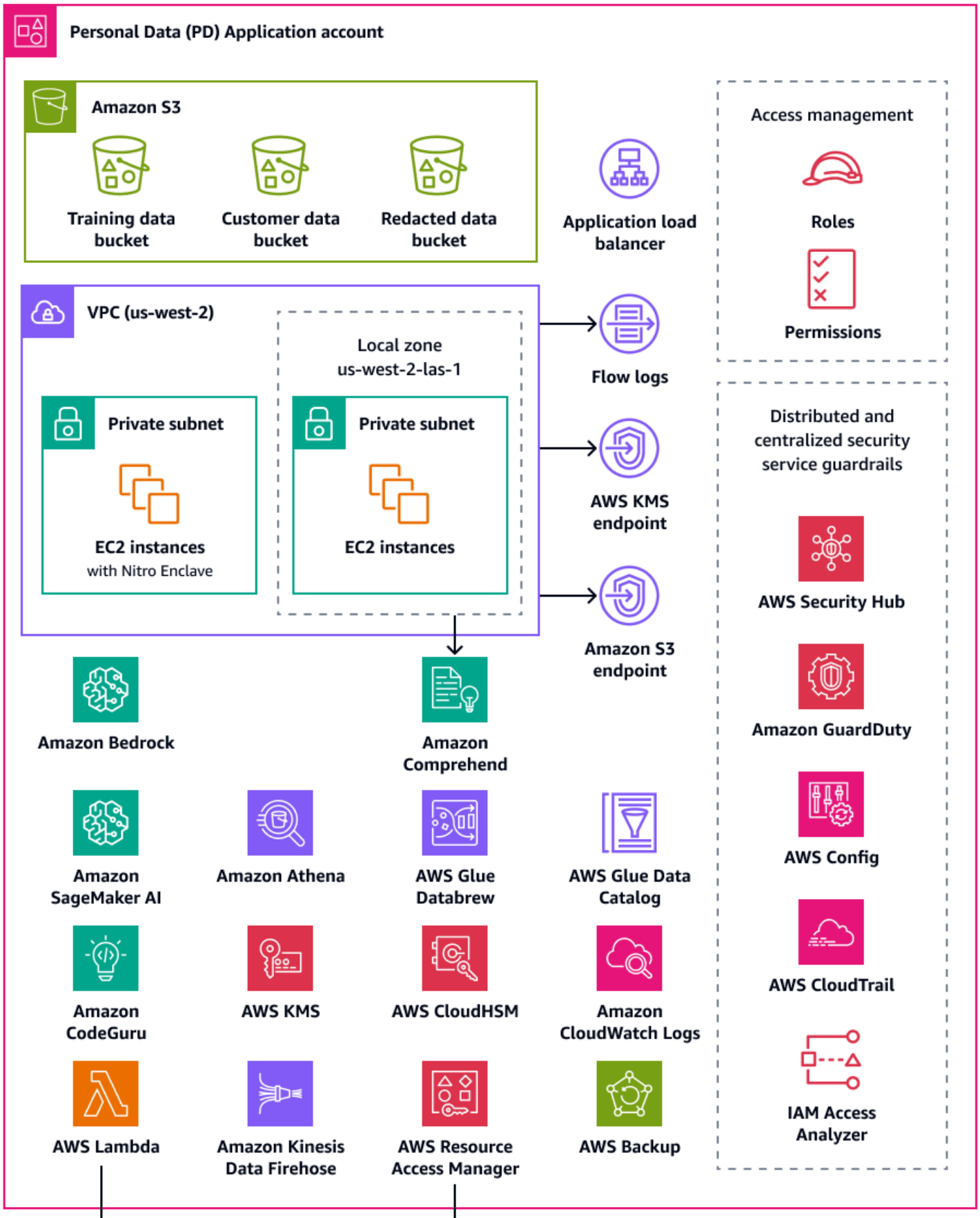
We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

The Personal Data (PD) Application account is where your organization hosts services that collect and process personal data. Specifically, you might store what you define as personal data in this account. The AWS PRA demonstrates a number of example privacy configurations through a multi-tier serverless web architecture. When it comes to operating workloads across an AWS landing zone, privacy configurations should not be considered one-size-fits-all solutions. For example, your goal might be to understand the underlying concepts, how they can enhance privacy, and how your organization can apply solutions to your particular use cases and architectures.

For AWS accounts in your organization that collect, store, or process personal data, you can use AWS Organizations and AWS Control Tower to deploy foundational and repeatable guardrails. Establishing a dedicated organizational unit (OU) for these accounts is critical. For example, you might want to apply data residency guardrails to only a subset of accounts where data residency is a core design consideration. For many organizations, these are the accounts that store and process personal data.

Your organization might consider supporting a dedicated Data account, which is where you store the authoritative source of your personal datasets. An authoritative data source is a location where you store the primary version of data, which might be considered the most reliable and accurate version of the data. For example, you might copy the data from the authoritative data source to other locations, such as Amazon Simple Storage Service (Amazon S3) buckets in the PD Application account that are used to store training data, a subset of customer data, and redacted data. By taking this multi-account approach to separate complete and definitive personal datasets in the Data account from the downstream consumer workloads in the PD Application account, you can reduce the scope of impact in the event of unauthorized access to your accounts.

The following diagram illustrates the AWS security and privacy services that are configured in the PD Application and Data accounts.


Personal Data (PD) Application account
Amazon S3
Training data bucket
Customer data bucket
Redacted data bucket
Application load balancer
Access management
Roles
Permissions
VPC (us-west-2)
Local zone us-west-2-las-1
Private subnet
EC2 instances with Nitro Enclave
Private subnet
EC2 instances
Flow logs
AWS KMS endpoint
Amazon S3 endpoint
Distributed and centralized security service guardrails
AWS Security Hub
Amazon GuardDuty
AWS Config
AWS CloudTrail
IAM Access Analyzer
Amazon Bedrock
Amazon Comprehend
Amazon SageMaker AI
Amazon Athena
AWS Glue Databrew
AWS Glue Data Catalog
Amazon CodeGuru
AWS KMS
AWS CloudHSM
Amazon CloudWatch Logs
AWS Lambda
Amazon Kinesis Data Firehose
AWS Resource Access Manager
AWS Backup
Data account
Access management
Log archive

This section provides more detailed information about the following AWS services that are used in these accounts:

- Amazon Athena
- Amazon Bedrock
- AWS Clean Rooms
- Amazon CloudWatch Logs
- Amazon CodeGuru Reviewer
- Amazon Comprehend
- Amazon Data Firehose
- Amazon DataZone
- AWS Glue
- AWS Key Management Service
- AWS Lake Formation
- AWS Local Zones
- AWS Nitro Enclaves
- AWS PrivateLink
- AWS Resource Access Manager
- Amazon SageMaker AI
- AWS features that help manage the data lifecycle
- AWS services and features that help segment data
- AWS services and features that help discover, classify, or catalog data

Amazon Athena

You can consider data query limitation controls to meet your privacy goals. Amazon Athena is an interactive query service that helps you analyze data directly in Amazon S3 by using standard SQL. You don't have to load the data into Athena; it works directly with the data stored in S3 buckets.

A common use case for Athena is providing data analytics teams with tailored and sanitized datasets. If the datasets contain personal data, you can sanitize the dataset by masking entire columns of personal data that provide little value to the data analytics teams. For more

information, see Anonymize and manage data in your data lake with Amazon Athena and AWS Lake Formation (AWS blog post).

If your data transformation approach requires additional flexibility outside of the supported functions in Athena, you can define custom functions, called user-defined functions (UDF). You can invoke UDFs in a SQL query submitted to Athena, and they run on AWS Lambda. You can use UDFs in `SELECT` and `FILTER` SQL queries, and you can invoke multiple UDFs in the same query. For privacy, you can create UDFs that perform specific types of data masking, such as showing only the last four characters of every value in a column.

Amazon Bedrock

Amazon Bedrock is a fully managed service that provides access to foundation models from leading AI companies like AI21 Labs, Anthropic, Meta, Mistral AI, and Amazon. It helps organizations to build and scale generative AI applications. No matter what platform is used, when using generative AI, organizations could face privacy risks, including the potential exposure of personal data, unauthorized data access, and other compliance violations.

Amazon Bedrock Guardrails is designed to help mitigate these risks by enforcing security and compliance best practices across your generative AI workloads in Amazon Bedrock. The deployment and use of AI resources might not always align with an organization's privacy and compliance requirements. Organizations can struggle with maintaining data privacy when using generative AI models because these models can potentially memorize or reproduce sensitive information. Amazon Bedrock Guardrails helps protect privacy by evaluating user inputs and model responses. Overall, if the input data contains personal data, there can be a risk of this information being exposed in the model's output.

Amazon Bedrock Guardrails provides mechanisms to enforce data protection policies and help prevent unauthorized data exposure. It offers content-filtering capabilities to detect and block personal data in inputs, topic restrictions to help prevent access to inappropriate or risky subject matter, and word filters to mask or redact sensitive terms in model prompts and responses. These capabilities help prevent events that could lead to privacy violations, such as biased responses, or erosion of customer trust. These features can help you make sure that personal data is not inadvertently processed or disclosed by your AI models. Amazon Bedrock Guardrails supports the evaluation of inputs and responses outside of Amazon Bedrock as well. For more information, see Implement model-independent safety measures with Amazon Bedrock Guardrails (AWS blog post).

With Amazon Bedrock Guardrails, you can limit the risk of model hallucinations by using contextual grounding checks, which evaluate factual grounding and the relevance of responses. An example

is deploying a generative AI customer-facing application that is uses third-party data sources in a Retrieval Augmented Generation (RAG) application. The contextual grounding checks can be used to validate model responses against these data sources and filter out inaccurate responses. In the context of the AWS PRA, you can implement Amazon Bedrock Guardrails across the workload accounts, where it enforces specific privacy guardrails that are tailored to each workload's requirements.

AWS Clean Rooms

As organizations look for ways to collaborate with one another through analysis of intersecting or overlapping sensitive datasets, maintaining the security and privacy of that shared data is a concern. AWS Clean Rooms helps you deploy *data clean rooms*, which are secure, neutral environments where organizations can analyze combined datasets without sharing the raw data itself. It also can generate unique insights by providing access to other organizations on AWS without moving or copying data out of their own accounts and without revealing the underlying dataset. All data remains in the source location. Built-in analysis rules constrain the output and restrict the SQL queries. All queries are logged, and collaboration members can view how their data is being queried.

You can create an AWS Clean Rooms collaboration and invite other AWS customers to be members of that collaboration. You grant one member the ability to query the member datasets, and you can choose additional members to receive the results of those queries. If more than one member needs to query the datasets, you can create additional collaborations with the same data sources and different member settings. Each member can filter the data that is shared with the collaboration members, and you can use custom analysis rules to set limitations on how the data they provide to the collaboration can be analyzed.

In addition to restricting the data presented to the collaboration and how it can be used by other members, AWS Clean Rooms provides the following capabilities that can help you protect privacy:

- *Differential privacy* is a mathematic technique that enhances user privacy through adding a carefully calibrated amount of noise to the data. This helps reduce the risk of individual user reidentification within the dataset without obscuring the values of interest. Using AWS Clean Rooms Differential Privacy doesn't require differential privacy expertise.
- AWS Clean Rooms ML allows two or more parties to identify similar users in their data without directly sharing the data with each other. This reduces the risk of membership inference attacks, where a member of the collaboration can identify individuals in the other member's dataset. By creating a lookalike model and generating a lookalike segment, AWS Clean Rooms ML helps you

compare datasets without exposing the original data. This does not require either member to have ML expertise or perform any work outside of AWS Clean Rooms. You retain full control and ownership of the trained model.

- Cryptographic Computing for Clean Rooms (C3R) can be used with analysis rules to derive insights from sensitive data. It cryptographically limits what any other party to the collaboration can learn. Using the C3R encryption client, the data is encrypted at the client before being provided to AWS Clean Rooms. Because the data tables are encrypted using a client-side encryption tool before being uploaded to Amazon S3, the data stays encrypted and persists through processing.

In the AWS PRA, we recommend that you create AWS Clean Rooms collaborations in the Data account. You can use them to share encrypted customer data with third parties. Use them only when there is an overlap in the provided datasets. For more information about how to determine overlap, see List analysis rule in the AWS Clean Rooms documentation.

Amazon CloudWatch Logs

Amazon CloudWatch Logs helps you centralize the logs from all your systems, applications, and AWS services so you can monitor them and archive them securely. In CloudWatch Logs, you can use a data protection policy for new or existing log groups to help minimize the risk of disclosure of personal data. Data protection policies can detect sensitive data, such as personal data, in your logs. The data protection policy can mask that data when users access the logs through the AWS Management Console. When users require direct access to the personal data, according to the overall purpose specification for your workload, you can assign `logs:Unmask` permissions for those users. You can also create an account-wide data protection policy and apply this policy consistently across all accounts in your organization. This configures masking by default for all current and future log groups in CloudWatch Logs. We also recommend that you enable audit reports and send them to another log group, an Amazon S3 bucket, or Amazon Data Firehose. These reports contain a detailed record of data protection findings across each log group.

Amazon CodeGuru Reviewer

For both privacy and security, it's vital to many organizations that they support continuous compliance during both deployment and post-deployment phases. The AWS PRA includes proactive controls in deployment pipelines for applications that process personal data. Amazon CodeGuru Reviewer can detect potential defects that might expose personal data in Java, JavaScript, and Python code. It offers suggestions to developers for improving the code. CodeGuru

Reviewer can identify defects across a wide range of security, privacy, and general recommended practices. It's designed to work with multiple source providers, including AWS CodeCommit, Bitbucket, GitHub, and Amazon S3. Some of the privacy-related defects that CodeGuru Reviewer can detect include:

- SQL injection
- Unsecured cookies
- Missing authorization
- Client-side AWS KMS re-encryption

For a complete list of what CodeGuru Reviewer can detect, see the Amazon CodeGuru Detector Library.

Amazon Comprehend

Amazon Comprehend is a natural-language processing (NLP) service that uses machine learning to uncover valuable insights and connections in English text documents. Amazon Comprehend can detect and redact personal data in structured, semi-structured, or unstructured text documents. For more information, see Personally identifiable information (PII) in the Amazon Comprehend documentation.

Because Amazon Comprehend has many options for application integration through AWS SDKs, you can use Amazon Comprehend to identify personal data in many different places where you collect, store, and process data. You can use Amazon Comprehend ML capabilities to detect and redact personal data in application logs (AWS blog post), customer emails, support tickets, and more. The architecture diagram for the PD Application account shows how you can perform this function for application logs on Amazon EC2. Amazon Comprehend offers two redaction modes:

- `REPLACE_WITH_PII_ENTITY_TYPE` replaces each PII entity with its types. For example, **Jane Doe** would be replaced with **NAME**.
- `MASK` replaces the characters in PII entities with a character of your choice (!, #, $, %, &, , *or @). For example, **Jane Doe** could be replaced with* **** ***.

Amazon Data Firehose

Amazon Data Firehose can be used to capture, transform, and load streaming data into downstream services, such as Amazon Managed Service for Apache Flink or Amazon S3. Firehose is

often used to transport large quantities of streaming data, such as application logs, without having to build processing pipelines from the ground up.

You can use Lambda functions to perform customized or built-in processing before the data is sent downstream. For privacy, this capability supports data minimization and cross-border data transfer requirements. For example, you can use Lambda and Firehose to transform multi-Region log data before it's centralized in the Log Archive account. For more information, see Biogen: Centralized Logging Solution for Multi Accounts (YouTube video). In the PD Application account, you configure Amazon CloudWatch and AWS CloudTrail to push logs to a Firehose delivery stream. A Lambda function transforms the logs and sends them to a central S3 bucket in the Log Archive account. You can configure the Lambda function to mask specific fields that contain personal data. This helps prevent the transfer of personal data across AWS Regions. By using this approach, the personal data is masked before the transfer and centralization, rather than after. For applications in jurisdictions that aren't subject to cross-border transfer requirements, it is typically more operationally efficient and cost effective to aggregate logs through the organizational trail in CloudTrail. For more information, see AWS CloudTrail in the *Security OU – Security Tooling account* section of this guide.

Amazon DataZone

As organizations scale their approach to sharing data through AWS services such as AWS Lake Formation, they want to make sure that differential access is controlled by those who are most familiar with the data: the data owners. However, these data owners might be aware of privacy requirements, such as consent or cross-border data transfer considerations. Amazon DataZone helps the data owners and the data governance team share and consume data across an organization according to your data governance policies. In Amazon DataZone, lines of business (LOBs) manage their own data, and a catalog tracks this ownership. Interested parties can find and request access to data as part of their business tasks. As long as it adheres to the policies established by the data publishers, the data owner can grant access to the underlying tables, without an administrator or moving the data.

In a privacy context, Amazon DataZone can be helpful in the following example use cases:

- A customer-facing application generates usage data that can be shared with a separate marketing LOB. You need to make sure that only data for customers that have opted in to marketing is published to the catalog.

- European customer data is published but may only be subscribed to by LOBs local to the European Economic Area (EEA). For more information, see Enhance data security with fine-grained access controls in Amazon DataZone.

In the AWS PRA, you can connect the data in the shared Amazon S3 bucket to Amazon DataZone as a data producer.

AWS Glue

Maintaining datasets that contain personal data is a key component of Privacy by Design. An organization's data might exist in structured, semi-structured, or unstructured forms. Personal datasets without structure can make it difficult to perform a number of privacy-enhancing operations, including data minimization, tracking down data attributed to a single data subject as a part of a data subject request, ensuring consistent data quality, and overall segmentation of datasets. AWS Glue is a fully managed extract, transform, and load (ETL) service. It can help you categorize, clean, enrich, and move data between data stores and data streams. AWS Glue features are designed to help you discover, prepare, structure, and combine datasets for analytics, machine learning, and application development. You can use AWS Glue to create a predictable and common structure on top of your existing datasets. AWS Glue Data Catalog, AWS Glue DataBrew, and AWS Glue Data Quality are AWS Glue features that can help support your organization's privacy requirements.

AWS Glue Data Catalog

AWS Glue Data Catalog helps you establish maintainable datasets. The Data Catalog contains references to data that is used as sources and targets for extract, transform, and load (ETL) jobs in AWS Glue. Information in the Data Catalog is stored as metadata tables, and each table specifies a single data store. You run an AWS Glue crawler to take inventory of the data in a variety of data store types. You add built-in and custom classifiers to the crawler, and these classifiers infer the data format and schema of the personal data. The crawler then writes the metadata to the Data Catalog. A centralized metadata table can make it easier to respond to data subject requests (such as right to erasure) because it adds structure and predictability across disparate sources of personal data in your AWS environment. For a comprehensive example of how to use Data Catalog to automatically respond to these requests, see Handling data erasure requests in your data lake with Amazon S3 Find and Forget (AWS blog post). Finally, if your organization is using AWS Lake Formation to administer and provide fine-grained access across databases, tables, rows, and cells, Data Catalog is a key component. Data Catalog provides cross-account data sharing and helps

you use tag-based access control to manage your data lake at scale (AWS blog post). For more information, see AWS Lake Formation in this section.

AWS Glue DataBrew

AWS Glue DataBrew helps you clean and normalize data, and it can perform transformations on the data, such as removing or masking personally identifiable information and encrypting sensitive data fields in data pipelines. You can also visually map the lineage of your data to understand the various data sources and transformation steps that the data has been through. This feature becomes increasingly important as your organization works to better understand and track personal data provenance. DataBrew helps you mask personal data during data preparation. You can detect personal data as part of a data profiling job and gather statistics, such as the number of columns that might contain personal data and potential categories. You can then use built-in reversible or irreversible data transformation techniques, including substitution, hashing, encryption, and decryption, all without writing any code. You can then use the cleaned and masked datasets downstream for analytics, reporting, and machine learning tasks. Some of the data masking techniques available in DataBrew include:

- **Hashing** – Apply hash functions to the column values.
- **Substitution** – Replace personal data with other, authentic-looking values.
- **Nulling out or deletion** – Replace a particular field with a null value, or delete the column.
- **Masking out** – Use character scrambling, or mask certain portions in the columns.

The following are the available encryption techniques:

- **Deterministic encryption** – Apply deterministic encryption algorithms to the column values. Deterministic encryption always produces the same ciphertext for a value.
- **Probabilistic encryption** – Apply probabilistic encryption algorithms to the column values. Probabilistic encryption produces different ciphertext each time that it's applied.

For a complete list of provided personal data transformation recipes in DataBrew, see Personally identifiable information (PII) recipe steps.

AWS Glue Data Quality

AWS Glue Data Quality helps you automate and operationalize the delivery of high-quality data across data pipelines, proactively, before they are delivered to your data consumers. AWS Glue Data

Quality provides statistical analysis of data quality issues across your data pipelines, can trigger alerts in Amazon EventBridge, and can make quality rule recommendations for remediation. AWS Glue Data Quality also supports rules creation with a domain-specific language so that you can create custom data quality rules.

AWS Key Management Service

AWS Key Management Service (AWS KMS) helps you create and control cryptographic keys to help protect your data. AWS KMS uses hardware security modules to protect and validate AWS KMS keys under the FIPS 140-2 Cryptographic Module Validation Program. For more information about how this service is used in a security context, see the AWS Security Reference Architecture.

AWS KMS integrates with most AWS services that offer encryption, and you can use KMS keys in your applications that process and store personal data. You can use AWS KMS to help support a variety of your privacy requirements and safeguard personal data, including:

- Using customer managed keys for greater control over strength, rotation, expiration, and other options.
- Using dedicated customer managed keys to protect personal data and secrets that allow access to personal data.
- Defining data classification levels and designating at least one dedicated customer managed key per level. For example, you might have one key to encrypt operational data and another to encrypt personal data.
- Preventing unintended cross-account access to KMS keys.
- Storing KMS keys within the same AWS account as the resource to be encrypted.
- Implementing separation of duties for KMS key administration and usage. For more information, see How to use KMS and IAM to enable independent security controls for encrypted data in S3 (AWS blog post).
- Enforcing automatic key rotation through preventative and reactive guardrails.

By default, KMS keys are stored and can be used only in the Region where they were created. If your organization has specific requirements for data residency and sovereignty, consider whether multi-Region KMS keys are appropriate for your use case. Multi-Region keys are special-purpose KMS keys in different AWS Regions that can be used interchangeably. The process of creating a multi-Region key moves your key material across AWS Region boundaries within AWS KMS, so this lack of regional isolation might not be compatible with your organization's sovereignty and

residency goals. One way to solve for this is to use a different type of KMS key, such as a Region-specific customer managed key.

External key stores

For many organizations, the default AWS KMS key store in the AWS Cloud can fulfill their data sovereignty and general regulatory requirements. But a few might require that encryption keys are created and maintained outside of a cloud environment and that you have independent authorization and audit paths. With external key stores in AWS KMS, you can encrypt personal data with key material that your organization owns and controls outside of the AWS Cloud. You still interact with the AWS KMS API as usual, but AWS KMS interacts only with external key store proxy (XKS proxy) software that you provide. Your external key store proxy then mediates all communication between AWS KMS and your external key manager.

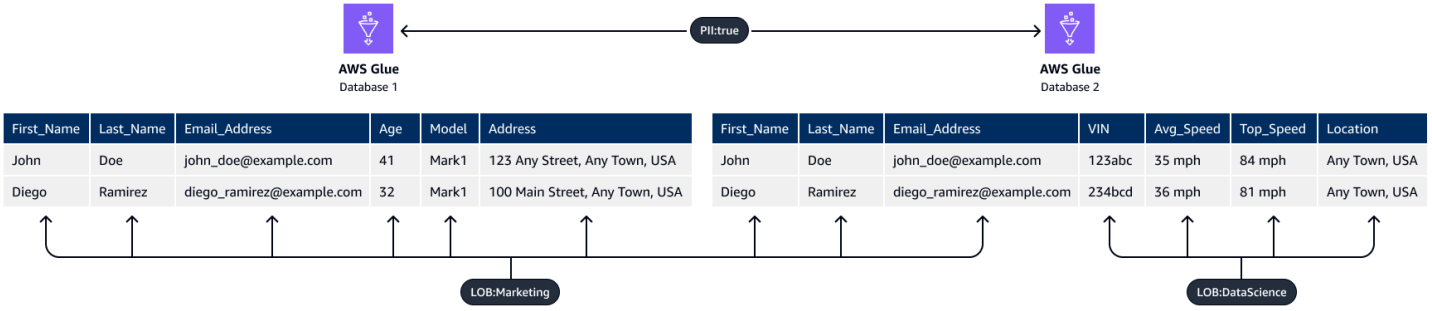
When using an external key store for data encryption, it is important that you consider the additional operational overhead compared to maintaining keys in AWS KMS. With an external key store, you must create, configure, and maintain the external key store. Also, if there are errors in the additional infrastructure you must maintain, such as the XKS proxy, and connectivity is lost, users might be temporarily unable to decrypt and access the data. Work closely with your compliance and regulatory stakeholders to understand the legal and contractual obligations for personal data encryption and your service level agreements for availability and resiliency.

AWS Lake Formation

Many organizations that catalog and categorize their datasets though structured metadata catalogues want to share those datasets across their organization. You can use AWS Identity and Access Management (IAM) permission policies to control access to entire datasets, but more granular control is often required for datasets that contain personal data of varying sensitivity. For example, the purpose specification and use limitation (FPC website) might indicate that a marketing team needs access to customer addresses, but a data science team does not.

There are also privacy challenges associated with data lakes, which centralize access to large amounts of sensitive data in their original format. Most of an organization's data can be centrally accessed in one place, so logical separation of datasets, especially those that contain personal data, can be paramount. AWS Lake Formation can help you set up governance and monitoring when sharing data, whether it be from a single source or many sources contained in a data lake. In the AWS PRA, you can use Lake Formation to provide fine-grained access control to the data in the shared data bucket in the Data account.

You can use the tag-based access control feature in Lake Formation. *Tag-based access control* is an authorization strategy that defines permissions based on attributes. In Lake Formation, these attributes are called *LF-Tags*. Using an LF-Tag, you can attach these tags to Data Catalog databases, tables, and columns and grant the same tags to IAM principals. Lake Formation allows operations on those resources when the principal has been granted access to a tag value that matches the resource tag value. The following image shows how you can assign LF-Tags and permissions to provide differentiated access to personal data.



This example uses the hierarchal nature of tags. Both databases contain personally identifiable information (`PII:true`), but tags at the columnar level limits specific columns to different teams. In this example, IAM principals who have the `PII:true` LF-Tag can access the AWS Glue database resources that have this tag. Principals with the `LOB:DataScience` LF-Tag can access specific columns that have this tag, and principals with the `LOB:Marketing` LF-Tag can access only columns that have this tag. The marketing can access only PII that is relevant to marketing use cases, and the data science team can access only PII that is relevant to their use cases.

AWS Local Zones

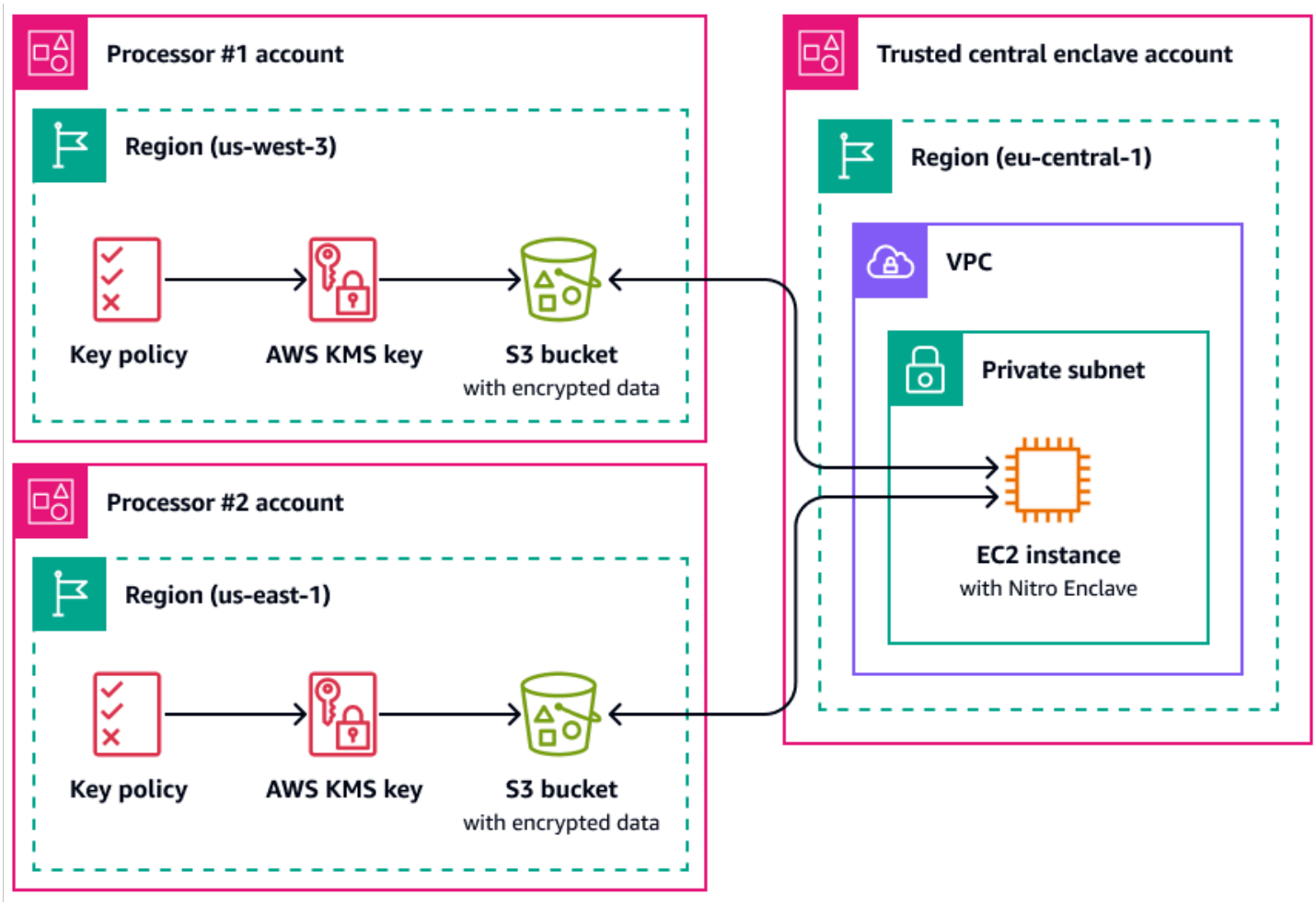
If you need to comply with data residency requirements, you can deploy resources that store and process personal data in specific AWS Regions to support these requirements. You can also use AWS Local Zones, which helps you place compute, storage, database, and other select AWS resources close to large population and industry centers. A Local Zone is an extension of an AWS Region that is in geographic proximity to a large metropolitan area. You can place specific types of resources within a Local Zone, near the Region to which the Local Zone corresponds. Local Zones can help you meet data residency requirements when a Region is unavailable within the same legal jurisdiction. When you use Local Zones, consider the data residency controls that are deployed within your organization. For example, you might need a control to prevent data transfers from a specific Local Zone to another Region. For more information about how to use SCPs to maintain cross-border data transfer guardrails, see Best Practices for managing data residency in AWS Local Zones using landing zone controls (AWS blog post).

AWS Nitro Enclaves

Consider your data segmentation strategy from a processing perspective, such as processing personal data with a compute service such as Amazon Elastic Compute Cloud (Amazon EC2). Confidential computing as a part of a larger architecture strategy can help you isolate personal data processing in an isolated, protected, and trusted CPU enclave. Enclaves are separate, hardened, and highly-constrained virtual machines. AWS Nitro Enclaves is an Amazon EC2 feature that can help you create these isolated compute environments. For more information, see The Security Design of the AWS Nitro System (AWS whitepaper).

Nitro Enclaves deploy a kernel that is separated from the parent instance's kernel. The parent instance's kernel doesn't have access to the enclave. Users can't SSH or remotely access the data and applications in the enclave. Applications that process personal data can be embedded in the enclave and configured to use the enclave's Vsock, the socket that facilitates communication between the enclave and the parent instance.

One use case where Nitro Enclaves can be useful is joint processing between two data processors that are in separate AWS Regions and that might not trust each other. The following image shows how you can use an enclave for central processing, a KMS key for encrypting the personal data before it's sent to the enclave, and an AWS KMS key policy that verifies that the enclave requesting decryption has the unique measurements in its attestation document. For more information and instructions, see Using cryptographic attestation with AWS KMS. For a sample key policy, see Require attestation to use an AWS KMS key in this guide.



With this implementation, only the respective data processors and the underlying enclave have access to the plaintext personal data. The only place the data is exposed, outside of the respective data processors' environments, is in the enclave itself, which is designed to prevent access and tampering.

AWS PrivateLink

Many organizations want to limit the exposure of personal data to untrusted networks. For example, if you want to enhance the privacy of your overall application architecture design, you can segment networks based on data sensitivity (similar to the logical and physical separation of datasets that is discussed in the AWS services and features that help segment data section). AWS PrivateLink helps you create unidirectional, private connections from your virtual private clouds (VPCs) to services outside of the VPC. Using AWS PrivateLink, you can set up dedicated private connections to the services that store or process personal data in your environment; there is no need to connect to public endpoints and transfer this data over untrusted public networks. When you enable AWS PrivateLink service endpoints for the in-scope services, there is no need for an

internet gateway, NAT device, public IP address, AWS Direct Connect connection, or AWS Site-to-Site VPN connection in order to communicate. When you use AWS PrivateLink to connect to a service that provides access to personal data, you can use VPC endpoint policies and security groups to control access, according to your organization's data perimeter definition. For a sample VPC endpoint policy that allows only IAM principles and AWS resources in a trusted organization to access a service endpoint, see Require organization membership to access VPC resources in this guide.

AWS Resource Access Manager

AWS Resource Access Manager (AWS RAM) helps you securely share your resources across AWS accounts to reduce operational overhead and provide visibility and auditability. As you plan your multi-account segmentation strategy, consider using AWS RAM to share the personal data stores that you store in a separate, isolated account. You can share that personal data with other, trusted accounts for the purposes of processing. In AWS RAM, you can manage permissions that define what actions can be performed on shared resources. All API calls to AWS RAM are logged in CloudTrail. Also, you can configure Amazon CloudWatch Events to automatically notify you for specific events in AWS RAM, such as when changes are made to a resource share.

Though you can share many types of AWS resources with other AWS accounts by using resource-based policies in IAM or bucket policies in Amazon S3, AWS RAM provides several additional benefits for privacy. AWS provides data owners with additional visibility over how and with whom the data is shared across your AWS accounts, including:

- Being able to share a resource with an entire OU instead of manually updating lists of account IDs
- Enforcement of the invitation process for share initiation if the consumer account isn't part of your organization
- Visibility into which specific IAM principals have access to each individual resource

If you've previously used a resource-based policy to manage a resource share and want to use AWS RAM instead, use the PromoteResourceShareCreatedFromPolicy API operation.

Amazon SageMaker AI

Amazon SageMaker AI is a managed machine learning (ML) service that helps you build and train ML models and then deploy them into a production-ready hosted environment. SageMaker AI is designed to make it easier to prepare training data and create model features.

Amazon SageMaker Model Monitor

Many organizations consider data drift when training ML models. Data drift is a meaningful variation between the production data and the data that was used to train an ML model, or a meaningful change in the input data over time. Data drift can reduce the overall quality, accuracy, and fairness in ML model predictions. If the statistical nature of the data that a ML model receives in production drifts away from the nature of the baseline data it was trained on, the accuracy of the predictions might decline. Amazon SageMaker Model Monitor can continuously monitor the quality of Amazon SageMaker AI machine learning models in production and monitor data quality. Early and proactive detection of data drift can help you implement corrective actions, such as retraining models, auditing upstream systems, or fixing data quality issues. Model Monitor can alleviate the need to manually monitor models or build additional tooling.

Amazon SageMaker Clarify

Amazon SageMaker Clarify provides insight into model bias and explainability. SageMaker Clarify is commonly used during ML model data preparation and the overall development phase. Developers can specify attributes of interest, such as gender or age, and SageMaker Clarify runs a set of algorithms to detect any presence of bias in those attributes. After the algorithm runs, SageMaker Clarify provides a visual report with a description of the sources and measurements of possible bias so that you can identify steps to remediate the bias. For example, in a financial dataset that contains only a few examples of business loans to one age group as compared to others, SageMaker could flag imbalances so that you can avoid a model that disfavors that age group. You can also check already trained models for bias by reviewing its predictions and by continuously monitoring those ML models for bias. Finally, SageMaker Clarify is integrated with Amazon SageMaker AI Experiments to provide a graph that explains which features contributed most to a model's overall prediction-making process. This information could be useful to meet explainability outcomes, and it could help you determine if a particular model input has more influence than it should on the overall model behavior.

Amazon SageMaker Model Card

Amazon SageMaker Model Card can help you document critical details about your ML models for governance and reporting purposes. These details can include the model owner, general purpose, intended use cases, assumptions made, risk rating of a model, training details and metrics, and evaluation results. For more information, see Model Explainability with AWS Artificial Intelligence and Machine Learning Solutions (AWS whitepaper).

Amazon SageMaker Data Wrangler

Amazon SageMaker Data Wrangler is a machine learning tool that helps streamline the data preparation and feature engineering process. It provides a visual interface that helps data scientists and machine learning engineers to quickly and easily prepare and transform data for use in machine learning models. With Data Wrangler, you can import data from various sources, such as Amazon S3, Amazon Redshift, and Amazon Athena. Then, you can use more than 300 built-in data transformations to clean, normalize, and combine features without having to write any code.

Data Wrangler can be used as part of the data preparation and feature engineering process in the AWS PRA. It supports data encryption at rest and in transit by using AWS KMS, and it uses IAM roles and policies to control access to data and resources. It supports data masking through AWS Glue or Amazon SageMaker Feature Store. If you integrate Data Wrangler with AWS Lake Formation, you can enforce fine-grained data access controls and permissions. You can even use Data Wrangler with Amazon Comprehend to automatically redact personal data from tabular data as a part of your broader ML Ops workflow. For more information, see Automatically redact PII for machine learning using Amazon SageMaker Data Wrangler (AWS blog post).

The versatility of Data Wrangler helps you mask sensitive data for many industries, such as account numbers, credit card numbers, social security numbers, patient names, and medical and military records. You can limit access to any sensitive data or choose to redact it.

AWS features that help manage the data lifecycle

When personal data is no longer required, you can use lifecycle and time-to-live polices for data in many different data stores. When configuring data retention policies, consider the following locations that might contain personal data:

- Databases, such as Amazon DynamoDB and Amazon Relational Database Service (Amazon RDS)
- Amazon S3 buckets
- Logs from CloudWatch and CloudTrail
- Cached data from migrations in AWS Database Migration Service (AWS DMS) and AWS Glue DataBrew projects
- Backups and snapshots

The following AWS services and features can help you configure data retention policies in your AWS environments:

- Amazon S3 Lifecycle – A set of rules that define actions that Amazon S3 applies to a group of objects. In the Amazon S3 Lifecyle configuration, you can create expiration actions, which define when Amazon S3 deletes expired objects on your behalf. For more information, see Managing your storage lifecycle.
- Amazon Data Lifecycle Manager – In Amazon EC2, create a policy that automates the creation, retention, and deletion of Amazon Elastic Block Store (Amazon EBS) snapshots and EBS-backed Amazon Machine Images (AMIs).
- DynamoDB Time to Live (TTL) – Define a per-item timestamp that determines when an item is no longer needed. Shortly after the date and time of the specified timestamp, DynamoDB deletes the item from your table.
- Log retention settings in CloudWatch Logs – You can adjust the retention policy for each log group to a value between 1 day and 10 years.
- AWS Backup – Centrally deploy data protection policies to configure, manage, and govern your backup activity across a variety of AWS resources, including S3 buckets, RDS database instances, DynamoDB tables, EBS volumes, and many more. Apply backup policies to your AWS resources by either specifying resource types or provide additional granularity by applying based on existing resource tags. Audit and report on backup activity from a centralized console to help meet backup compliance requirements.

AWS services and features that help segment data

Data segmentation is the process by which you store data in separate containers. This can help you to provide differentiated security and authentication measures to each dataset and to reduce the scope of impact of exposure for your overall dataset. For example, instead of storing all customer data in one large database, you may segment this data into smaller, more manageable groups.

You can use physical and logical separation to segment personal data:

- **Physical separation** – The act of storing data in separate data stores or distributing your data into separate AWS resources. Though the data is physically separated, both resources might be accessible to the same principals. This is why we recommend combining physical separation with logical separation.
- **Logical separation** – The act of isolating data by using access controls. Different job functions require different levels of access to subsets of personal data. For a sample policy that implements logical separation, see Grant access to specific Amazon DynamoDB attributes in this guide.

The combination of a logical and physical separation provides flexibility, simplicity, and granularity when writing identity-based and resource-based policies to support differentiated access across job functions. For example, it can be operationally complex to create the policies that logically separate different data classifications in a single S3 bucket. Using dedicated S3 buckets for each data classification simplifies policy configuration and management.

AWS services and features that help discover, classify, or catalog data

Some organizations have not started to use extract, load, and transform (ELT) tools in their environment to proactively catalog their data. These customers might be at an early data discovery stage, where they want to better understand the data that they store and process in AWS and how it's structured and classified. You can use Amazon Macie to better understand your PII data in Amazon S3. However, Amazon Macie cannot help you analyze other data sources, such as Amazon Relational Database Service (Amazon RDS) and Amazon Redshift. You can use two approaches to accelerate the initial discovery at the beginning of a larger data mapping exercise:

- **Manual approach** – Make a table with two columns and as many rows as you need. In the first column, write a data characterization (such as user name, address, or gender) that might be in the header or body of a network packet or in any service that you provide. Ask your compliance team to complete the second column. In the second column, enter a "yes" if the data is considered personal and "no" if it isn't. Indicate any type of personal data that is deemed particularly sensitive, such as religious denomination or health data.
- **Automated approach** – Use tooling provided through AWS Marketplace. One such tool is Securiti. These solutions offer integrations that allow them to scan and discover data across multiple AWS resource types, as well as assets in other cloud service platforms. Many of these same solutions can continually collect and maintain an inventory of data assets and data processing activities in a centralized data catalog. If you rely on a tool to perform automated classification, it might require tuning discovery and classification rules in order to align to your organization's definition of personal data.

Sample privacy-related policies

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

Many organizations that handle sensitive data take a preventative-forward approach, with layers of detective and reactive controls implemented throughout. This section provides examples of privacy-related policies for AWS Identity and Access Management (IAM), AWS Organizations, and AWS Key Management Service (AWS KMS). These policies can help your organization meet various use, disclosure limitation, and cross-border data transfer privacy goals by using a preventative approach. Many of these policies are referenced in previous sections in this guide.

This section contains the following sample policies:

- Require access from specific IP addresses
- Require organization membership to access VPC resources
- Restrict data transfers across AWS Regions
- Grant access to specific Amazon DynamoDB attributes
- Restrict changes to VPC configurations
- Require attestation to use an AWS KMS key

Require access from specific IP addresses

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

This policy allows the `john_stiles` user to assume IAM roles only if the call is coming from an IP address in the ranges `192.0.2.0/24` or `203.0.113.0/24`. This policy can help prevent unintended disclosure of personal data and unwanted cross-border data transfers. For example,

if your organization has customer support staff that require access to personal data, you might want that support staff to access that data only from offices that are located in a subset of specific AWS Regions. Also, verify your organization's definition of PII because some policies might require `Condition` or `Principal` sections that restrict access to a specific user or IP address.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::123456789012:user/john_stiles"
      },
      "Action": "sts:AssumeRole"
    },
    {
      "Effect": "Deny",
      "Principal": {
        "AWS": "arn:aws:iam::123456789012:user/john_stiles"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "NotIpAddress": {
          "aws:SourceIp": [
            "192.0.2.0/24",
            "203.0.113.0/24"
          ]
        }
      }
    }
  ]
}
```

Require organization membership to access VPC resources

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

This VPC endpoint policy allows only AWS Identity and Access Management (IAM) principals and resources from the `o-1abcde123` organization to access Amazon Personalize (Amazon S3) endpoints. This preventative control helps establish a zone of trust and define the personal data perimeter. For more information about how this policy can help protect privacy and personal data in your organization, see AWS PrivateLink in this guide.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowOnlyIntendedResourcesAndPrincipals",
            "Effect": "Allow",
            "Principal": "*",
            "Action": "s3:*",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "aws:PrincipalOrgID": "o-1abcde123",
                    "aws:ResourceOrgID": "o-1abcde123"
                }
            }
        }
    ]
}
```

Restrict data transfers across AWS Regions

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

With the exception of two AWS Identity and Access Management (IAM) roles, this service control policy denies API calls to regional AWS services in AWS Regions other than `eu-west-1` and `eu-central-1`. This SCP can help prevent the creation of AWS storage and processing services in unapproved Regions. This can help prevent personal data from being handled by AWS services in those Regions altogether. This policy uses a `NotAction` parameter because it accounts for global AWS services, such as IAM, and services that integrate with global services, such as AWS Key

Management Service (AWS KMS) and Amazon CloudFront. In the parameter values, you can specify those global and other non-applicable services as exceptions. For more information about how this policy can help protect privacy and personal data in your organization, see AWS Organizations in this guide.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "DenyAllOutsideEU",
            "Effect": "Deny",
            "NotAction": [
                "a4b:*",
                "acm:*",
                "aws-marketplace-management:*",
                "aws-marketplace:*",
                "aws-portal:*",
                "budgets:*",
                "ce:*",
                "cloudfront:*",
                "config:*",
                "cur:*",
                "directconnect:*",
                "ec2:DescribeRegions",
                "ec2:DescribeTransitGateways",
                "ec2:DescribeVpnGateways",
                "fms:*",
                "globalaccelerator:*",
                "health:*",
                "iam:*",
                "importexport:*",
                "kms:*",
                "mobileanalytics:*",
                "networkmanager:*",
                "organizations:*",
                "pricing:*",
                "route53:*",
                "route53domains:*",
                "route53-recovery-cluster:*",
                "route53-recovery-control-config:*",
                "route53-recovery-readiness:*",
                "s3:GetAccountPublic*",
                "s3:ListAllMyBuckets",
```

```
                "s3:ListMultiRegionAccessPoints",
                "s3:PutAccountPublic*",
                "shield:*",
                "sts:*",
                "support:*",
                "trustedadvisor:*",
                "waf-regional:*",
                "waf:*",
                "wafv2:*",
                "wellarchitected:*"
            ],
            "Resource": "*",
            "Condition": {
                "StringNotEquals": {
                    "aws:RequestedRegion": [
                        "eu-central-1",
                        "eu-west-1"
                    ]
                },
                "ArnNotLike": {
                    "aws:PrincipalARN": [
                        "arn:aws:iam::*:role/Role1AllowedToBypassThisSCP",
                        "arn:aws:iam::*:role/Role2AllowedToBypassThisSCP"
                    ]
                }
            }
        }
    ]
}
```

Grant access to specific Amazon DynamoDB attributes

> **Survey**
>
> We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

As your organization discusses strategies to physically and logically separate personal data, consider which AWS storage services support fine-grained access control policies in AWS Identity and Access Management (IAM). The following identity-based policy allows retrieval of only the

`UserID`, `SignUpTime`, and `LastLoggedIn` attributes from an Amazon DynamoDB table named `Users`. For example, you might attach this policy to a customer support role instead of giving this role access to the full personal dataset. For more information about how this policy can help protect privacy and personal data in your organization, see AWS services and features that help segment data in this guide.

```
{
   "Version": "2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Action":[
            "dynamodb:GetItem",
            "dynamodb:BatchGetItem",
            "dynamodb:Query",
            "dynamodb:Scan"
         ],
         "Resource":[
            "arn:aws:dynamodb:us-west-2:123456789012:dynamodb:table/Users"
         ],
         "Condition":{
            "ForAllValues:StringEquals":{
               "dynamodb:Attributes":[
                  "UserID",
                  "SignUpTime",
                  "LastLoggedIn"
               ]
            },
            "StringEquals":{
               "dynamodb:Select":[
                  "SPECIFIC_ATTRIBUTES"
               ]
            }
         }
      }
   ]
}
```

Restrict changes to VPC configurations

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

After you have designed and deployed the AWS infrastructure that supports your cross-border data transfer requirements, which includes network data flows, you might want to prevent modifications. The following service control policy helps prevent VPC configuration drift or unintentional modification. It denies new internet gateway attachments, VPC peering connections, transit gateway attachments, and new VPN connections. For more information about how this policy can help protect privacy and personal data in your organization, see AWS Transit Gateway in this guide.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "ec2:AttachInternetGateway",
                "ec2:CreateInternetGateway",
                "ec2:CreateVpcPeeringConnection",
                "ec2:AcceptVpcPeeringConnection",
                "ec2:CreateVpc",
                "ec2:CreateSubnet",
                "ec2:CreateRouteTable",
                "ec2:CreateRoute",
                "ec2:AssociateRouteTable",
                "ec2:ModifyVpcAttribute",
                "ec2:*TransitGateway",
                "ec2:*TransitGateway*",
                "globalaccelerator:Create*",
                "globalaccelerator:Update*"

            ],
            "Resource": "*",
            "Effect": "Deny",
            "Condition": {
                "ArnNotLike": {
```

```
                    "aws:PrincipalARN": [
                        "arn:aws:iam::*:role/Role1AllowedToBypassThisSCP",
                        "arn:aws:iam::*:role/Role2AllowedToBypassThisSCP"
                    ]
                }
            }
        }
    ]
}
```

Require attestation to use an AWS KMS key

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

The following AWS Key Management Service (AWS KMS) key policy allows AWS Nitro Enclave instances to use a KMS key only if the enclave's attestation document in the request matches the measurements in the condition statement. This policy allows only trusted enclaves to decrypt the data. For more information about how this policy can help protect privacy and personal data in your organization, see AWS Nitro Enclaves in this guide. For a complete list of AWS KMS condition keys that can be used in key policies and in AWS Identity and Access Management (IAM) policies, see Condition keys for AWS KMS.

```
{
   "Version": "2012-10-17",
   "Statement": [
      {
         "Sid": "Enable enclave data processing",
         "Effect": "Allow",
         "Principal": {
            "AWS": "arn:aws:iam::123456789012:role/data-processing"
         },
         "Action": [
            "kms:Decrypt",
            "kms:GenerateDataKey",
            "kms:GenerateRandom"
         ],
```

```
         "Resource": "*",
         "Condition": {
            "StringEqualsIgnoreCase": {
               "kms:RecipientAttestation:ImageSha384":
 "EXAMPLE8abcdef7abcdef6abcdef5abcdef4abcdef3abcdef2abcdef1abcdef1abcdef0abcdef1abcdEXAMPLE",
               "kms:RecipientAttestation:PCR0":
 "EXAMPLEbc2ecbb68ed99a13d7122abfc0666b926a79d5379bc58b9445c84217f59cfdd36c08b2c79552928702EXAM
               "kms:RecipientAttestation:PCR1":
 "EXAMPLE050abf6b993c915505f3220e2d82b51aff830ad14cbecc2eec1bf0b4ae749d311c663f464cde9f718aEXAM
               "kms:RecipientAttestation:PCR2":
 "EXAMPLEc300289e872e6ac4d19b0b5ac4a9b020c98295643ff3978610750ce6a86f7edff24e3c0a4a445f2ff8EXAM
               "kms:RecipientAttestation:PCR3":
 "EXAMPLE11de9baee597508183477f097ae385d4a2c885aa655432365b53b812694e230bbe8e1bb1b8de748fe1EXAM
               "kms:RecipientAttestation:PCR4":
 "EXAMPLE6b9b3d89a53b13f5dfd14a1049ec0b80a9ae4b159adde479e9f7f512f33e835a0b9023ca51ada02160EXAM
               "kms:RecipientAttestation:PCR8":
 "EXAMPLE34a884328944cd806127c7784677ab60a154249fd21546a217299ccfa1ebfe4fa96a163bf41d3bcfaeEXAM
            }
         }
      }
   ]
}
```

Strategizing for global expansion

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

AWS Security Assurance Services frequently receives questions regarding architecting for privacy on AWS when expanding globally. Questions revolve around concerns with maintaining compliance with unique privacy requirements, such as data sovereignty obligations or customer contracts, all while avoiding additional costs and operational overhead. Design considerations often include data residency, operator access restriction, resiliency and survivability, and overall independence. For more information, see Meeting digital sovereignty requirements on AWS (AWS re:Invent 2022 presentation).

The following questions are common, and only you can answer them for your use case:

- Where does my customers' personal data need to reside?
- Where is my customer data stored?
- How and where can personal data cross borders?
- Does human or service access to data across regions constitute a transfer?
- How can I be sure that no foreign governments access my customers' personal data?
- Where can I store my backups and hot or cold sites?
- To keep data local, should I maintain an AWS landing zone in every region where I provide services? Or can I use an existing AWS Control Tower landing zone?

For data residency requirements, different architecture deployments might work better for different organizations. Some organizations might have requirements that their customers' personal data stay within a specific region. If so, you might be concerned with how to generally comply with regulations while upholding these obligations. No matter the situation, there are multiple considerations when choosing a multi-account deployment strategy.

To define key architecture design components, work closely with your compliance and contract teams to confirm requirements for where, when, and how personal data can cross AWS Regions.

Determine what qualifies as data transfer, such as moving, copying, or viewing. In addition, understand whether there are specific resiliency and data protection controls that must be implemented. Do backup and disaster recovery strategies require cross-Region failover? If so, determine which Regions you can use to store your backup data. Determine if there are any requirements for data encryption, such as a specific encryption algorithm or dedicated hardware security modules for key generation. After you align with compliance stakeholders on these topics, start to consider design approaches for your multi-account environment.

The following are three approaches you can use to plan your AWS multi-account strategy, in ascending order of infrastructure segregation:

- Central landing zone with managed Regions
- Regional landing zones
- AWS European Sovereign Cloud

It is also important to remember that privacy compliance might not stop at just data sovereignty. Review the rest of this guide to understand possible solutions for many other challenges, such as consent management, data subjects' requests, data governance, and AI bias.

Central landing zone with managed Regions

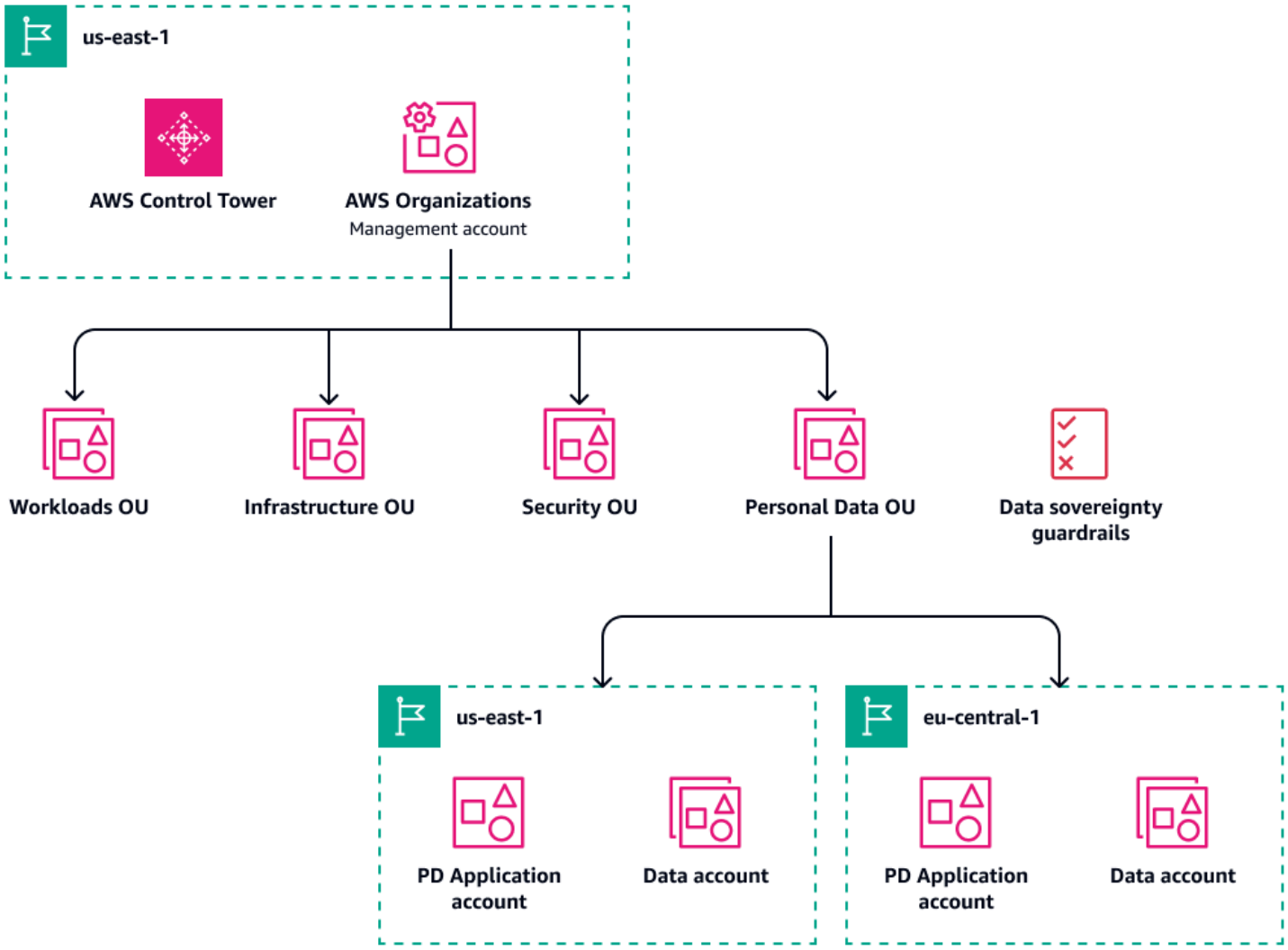
If you want to expand globally but have already established a multi-account architecture in AWS, it's common to want to use the same multi-account landing zone (MALZ) to manage additional AWS Regions. In this configuration, you would continue to operate infrastructure services such as logging, account factory, and general administration from your existing AWS Control Tower landing zone, in the Region where you created it.

For production workloads, you can operate Regional deployments by extending the AWS Control Tower landing zone into a new Region. By doing this, you can extend the AWS Control Tower governance into the new Region. This way, you can keep personal data stores within a specific, managed Region, the data still resides in accounts that benefit from the infrastructure services and AWS Control Tower governance. In AWS Organizations, accounts that contain personal data still roll up under a dedicated Personal Data OU, where all of the data sovereignty guardrails in AWS Control Tower are implemented. In addition, Region-specific workloads are contained in dedicated accounts, rather than establishing production accounts that may contain the same workload in multiple Regions.

This deployment can be the most cost effective, but additional consideration is needed for controlling the flow of personal data across AWS account and Regional boundaries. Consider the following:

- Logs might contain personal data, so some additional configuration may be required to contain or redact sensitive fields to prevent cross-Region transfer during aggregation. For more information and recommended practices to control log aggregation across Regions, see Centralized log storage in this guide.
- Account for isolation of VPCs and the appropriate bidirectional network traffic flow in the AWS Transit Gateway design. You can limit which Transit Gateway attachments are allowed and approved, and you can limit who or what can change the VPC route tables.
- You might need to prevent members of your cloud operations team from accessing personal data. For example, application logs that contain customer transaction data may be deemed of higher sensitivity than other log sources. Additional approvals and technical guardrails might be required, such as role-based access control and attribute-based access control. Also, data might be subject to residency limitations when it comes to access. For example, data in one Region A can only be accessed from within that Region.

The following diagram shows a centralized landing zone with Regional deployments.



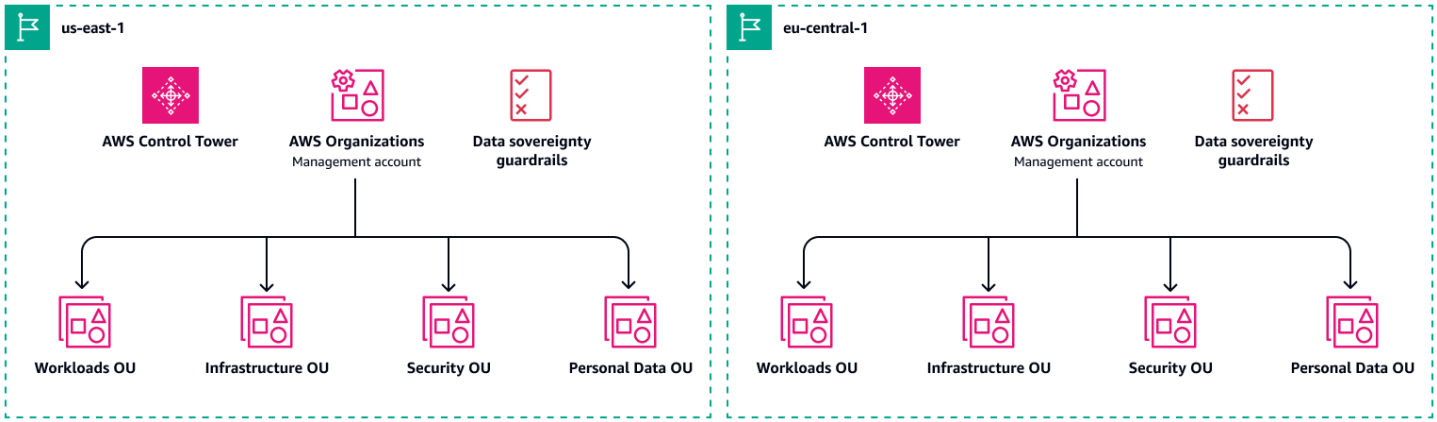
Regional landing zones

Having more than one MALZ can help you achieve stricter compliance requirements by completely isolating workloads that process personal data compared to non-material workloads. AWS Control Tower centralized logging aggregation could be configured by default and therefore simplified. With this approach, you don't need to maintain exceptions for logging with separate streams of logs that require redaction. You could even have a local and dedicated cloud operations team for each MALZ, which limits operator access to local residency.

Many organizations have separate US and EU-based landing zone deployments. Each Regional landing zone has a single, blanket security posture and associated governance for accounts in the Region. For example, encryption of personal data using dedicated HSMs may not be required in workloads in one MALZ, but it might be required in another MALZ.

Although this strategy can scale to meet many current and future requirements, it is important to understand the additional costs and operational overhead associated with maintaining multiple MALZs. For more information, see AWS Control Tower pricing.

The following diagram shows separate landing zones in two Regions.



AWS European Sovereign Cloud

Some organizations require thorough separation of their workloads operating in the European Economic Area (EEA) and workloads operating elsewhere. In this situation, consider the AWS European Sovereign Cloud. The AWS European Sovereign Cloud is a new, independent cloud for Europe, designed to help customers meet the region's evolving sovereignty needs, including stringent data residency, operational autonomy, and resiliency requirements.

The AWS European Sovereign Cloud is physically and logically separate from existing AWS Regions, all while offering the same security, availability, and performance. Only AWS employees who are located in the EU have control of the operations and support for the AWS European Sovereign Cloud. If you have stringent data residency requirements, the AWS European Sovereign Cloud keeps all metadata that you create (such as the roles, permissions, resource labels, and configurations they use to run AWS) in the EU. The AWS European Sovereign Cloud also features its own billing and usage metering systems.

For this approach, you would use a similar pattern as in the previous section, Regional landing zones. However, for services that you provide to European customers, you could deploy a dedicated MALZ in the AWS European Sovereign Cloud.

Resources

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

AWS Prescriptive Guidance

- AWS Security Reference Architecture (AWS SRA)

AWS documentation

- Data protection (AWS Well-Architected Framework)
- Data classification (AWS whitepaper)
- Amazon Web Services: Risk and Compliance (AWS whitepaper)
- Hybrid architectures to address personal data processing requirements (AWS whitepaper)
- Navigating GDPR Compliance on AWS (AWS whitepaper)
- Building a data perimeter on AWS (AWS whitepaper)
- AWS Security Documentation

Other AWS resources

- AWS Compliance Programs
- AWS Shared Responsibility Model
- Data Privacy FAQ
- AWS Security Assurance Services
- AWS Digital Sovereignty Pledge: Control without compromise (AWS blog post)
- AWS Security Learning

Contributors

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

This guide was authored by the AWS Security Assurance Services team. For support implementing the recommendations in this guide and operationalizing your workloads, contact the AWS Security Assurance Services team.

Primary authors

- Amber Welch, AWS Senior Privacy Consultant
- Daniel Nieters, AWS Principal Privacy Consultant
- Robert Carter, AWS Technical Program Manager

Contributors

- Avik Mukherjee, AWS Senior Security Consultant
- David Bounds, AWS Senior Solutions Architect
- Jeff Lombardo, AWS Senior Security Solutions Architect
- Ram Ramani, AWS Principal Security Solutions Architect
- Vanessa Jacobs, AWS Senior Security Consultant
- Thomas Nicholson, AWS Senior Privacy Consultant
- Jose DeJesus, AWS Senior Assurance Consultant
- Doug Pardue, AWS Solutions Architect Manager

Technical writers

- Lilly AbouHarb, AWS Senior Technical Writer

Document history

Survey

We would love to hear from you. Please provide feedback on the AWS PRA by taking a short survey.

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

Change	Description	Date
Signficant updates	We added Cloud Computing Compliance Controls Catalog (C5) to the AWS Artifact section. We added Amazon Security Lake to the Log Archive account. We added Amazon Bedrock, AWS Clean Rooms, Amazon DataZone, AWS Lake Formation, Amazon SageMaker AI, and *AWS services and features that help discover, classify, or catalog data* to the PD Applicati on account. We added the Strategizing for global expansion section.	September 16, 2025
Signficant updates	We made significant updates throughout.	March 26, 2024
Initial publication	—	October 2, 2023