AWS Whitepaper

Overview of Amazon Web Services

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Overview of Amazon Web Services: AWS Whitepaper

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

Abstract and introduction **1**
Introduction 1
What is cloud computing? **2**
Six advantages of cloud computing **3**
Types of cloud computing **4**
Deployment models 4
Cloud 4
Private cloud (on-premises) 4
Hybrid 4
Global infrastructure **5**
Security and compliance **6**
Security 6
Benefits of AWS security 7
Compliance 7
AWS services **8**
Accessing AWS services 9
Analytics 9
Amazon Athena 11
Amazon CloudSearch 11
Amazon DataZone 11
Amazon EMR 12
Amazon FinSpace 12
Amazon Kinesis 12
Amazon Data Firehose 13
Amazon Managed Service for Apache Flink 13
Amazon Kinesis Data Streams 14
Amazon Kinesis Video Streams 14
Amazon OpenSearch Service 14
Amazon OpenSearch Serverless 14
Amazon Redshift 15
Amazon Redshift Serverless 15
QuickSight 15
AWS Clean Rooms 16
AWS Data Exchange 16

AWS Data Pipeline ... 16
AWS Entity Resolution ... 17
AWS Glue ... 17
AWS Lake Formation ... 18
Amazon Managed Streaming for Apache Kafka (Amazon MSK) ... 18
Application integration ... 19
AWS Step Functions ... 21
Amazon AppFlow ... 21
AWS B2B Data Interchange ... 21
Amazon EventBridge ... 22
Amazon Managed Workflows for Apache Airflow (MWAA) ... 22
Amazon MQ ... 22
Amazon Simple Notification Service ... 23
Amazon Simple Queue Service ... 23
Amazon Simple Workflow Service ... 23
Blockchain ... 23
Business applications ... 24
Alexa for Business ... 25
AWS AppFabric ... 25
Amazon Chime ... 25
Amazon Chime SDK ... 26
Amazon Connect ... 26
Amazon Pinpoint ... 26
Amazon SES ... 27
Amazon WorkDocs ... 27
Amazon WorkMail ... 27
Cloud Financial Management ... 28
AWS Application Cost Profiler ... 29
AWS Billing Conductor ... 29
AWS Cost Explorer ... 30
AWS Budgets ... 30
AWS Cost and Usage Report ... 30
Reserved Instance (RI) reporting ... 30
Savings Plans ... 31
Compute ... 31
Compare AWS compute services ... 33

Amazon EC2 ... 35
Amazon EC2 Auto Scaling ... 37
Amazon EC2 Image Builder ... 38
Amazon Lightsail ... 38
Amazon Linux 2023 ... 38
AWS App Runner ... 39
AWS Batch ... 39
AWS Elastic Beanstalk ... 39
AWS Fargate ... 40
AWS Lambda ... 40
AWS Serverless Application Repository ... 40
AWS Outposts ... 41
AWS Wavelength ... 41
VMware Cloud on AWS ... 42
Customer enablement ... 43
Containers ... 43
Amazon Elastic Container Registry ... 44
Amazon Elastic Container Service ... 45
Amazon Elastic Kubernetes Service ... 45
AWS App2Container ... 45
Red Hat OpenShift Service on AWS ... 46
Databases ... 46
Compare AWS database services ... 48
Amazon Aurora ... 49
Amazon DynamoDB ... 50
Amazon ElastiCache ... 51
Amazon Keyspaces (for Apache Cassandra) ... 51
Amazon MemoryDB ... 52
Amazon Neptune ... 52
Amazon Relational Database Service ... 53
Amazon RDS for Db2 ... 53
Amazon RDS on VMware ... 53
Amazon Quantum Ledger Database (Amazon QLDB) ... 54
Amazon Timestream ... 55
Amazon DocumentDB (with MongoDB compatibility) ... 55
Amazon Lightsail managed databases ... 55

Developer tools ... 56
AWS Infrastructure Composer ... 56
AWS Cloud9 ... 57
AWS CloudShell ... 57
AWS CodeArtifact ... 57
AWS CodeBuild ... 57
Amazon CodeCatalyst ... 58
AWS CodeCommit ... 58
AWS CodeDeploy ... 58
AWS CodePipeline ... 58
Amazon Corretto ... 59
AWS Fault Injection Service ... 59
Amazon Q Developer ... 59
AWS X-Ray ... 60
End user computing ... 60
Frontend web and mobile services ... 62
AWS Amplify ... 63
AWS AppSync ... 63
AWS Device Farm ... 63
Amazon Location Service ... 64
Game tech ... 64
IoT ... 64
AWS IoT Analytics ... 66
AWS IoT Button ... 67
AWS IoT Core ... 67
AWS IoT Device Defender ... 67
AWS IoT Device Management ... 68
AWS IoT Events ... 68
AWS IoT ExpressLink ... 69
AWS IoT FleetWise ... 69
AWS IoT Greengrass ... 70
AWS IoT SiteWise ... 70
AWS IoT TwinMaker ... 71
AWS Partner Device Catalog ... 71
FreeRTOS ... 71
ML and AI ... 72

Amazon Augmented AI ... 74
Amazon Bedrock ... 74
Amazon CodeGuru ... 75
Amazon Comprehend ... 75
Amazon DevOps Guru ... 75
Amazon Forecast ... 76
Amazon Fraud Detector ... 77
Amazon Comprehend Medical ... 77
Amazon Kendra ... 77
Amazon Lex ... 78
Amazon Lookout for Equipment ... 78
Amazon Lookout for Metrics ... 79
Amazon Lookout for Vision ... 79
Amazon Monitron ... 80
Amazon PartyRock ... 80
Amazon Personalize ... 80
Amazon Polly ... 81
Amazon Q ... 82
Amazon Rekognition ... 82
Amazon SageMaker AI ... 83
Amazon Textract ... 89
Amazon Transcribe ... 90
Amazon Translate ... 91
AWS DeepComposer ... 91
AWS DeepRacer ... 91
AWS HealthLake ... 91
AWS HealthScribe ... 92
AWS Panorama ... 92
Management and governance ... 93
AWS Auto Scaling ... 94
AWS CloudFormation ... 94
AWS CloudTrail ... 95
Amazon CloudWatch ... 95
AWS Compute Optimizer ... 95
AWS Console Mobile Application ... 96
AWS Control Tower ... 96

AWS Config ... 97
AWS Health ... 97
AWS Launch Wizard ... 97
AWS License Manager ... 98
Amazon Managed Grafana ... 98
Amazon Managed Service for Prometheus ... 99
AWS Organizations ... 99
AWS OpsWorks ... 99
AWS Proton ... 100
Amazon Q Developer in chat applications (formerly AWS Chatbot) ... 100
AWS Service Catalog ... 100
AWS Systems Manager ... 101
AWS Trusted Advisor ... 103
AWS User Notifications ... 103
AWS Well-Architected Tool ... 103
Media ... 103
Amazon Elastic Transcoder ... 104
Amazon Interactive Video Service ... 104
Amazon Nimble Studio ... 105
AWS Elemental Appliances and Software ... 105
AWS Elemental MediaConnect ... 105
AWS Elemental MediaConvert ... 106
AWS Elemental MediaLive ... 106
AWS Elemental MediaPackage ... 106
AWS Elemental MediaStore ... 106
AWS Elemental MediaTailor ... 107
Migration and transfer ... 107
AWS Application Discovery Service ... 108
AWS Application Migration Service ... 109
AWS Database Migration Service ... 109
AWS Mainframe Modernization Service ... 109
AWS Migration Hub ... 110
AWS Snow Family ... 110
AWS DataSync ... 112
AWS Transfer Family ... 112
Networking and content delivery ... 113

Amazon API Gateway 114
Amazon CloudFront 114
Amazon Route 53 115
AWS Verified Access 115
Amazon VPC 116
Amazon VPC Lattice 116
AWS App Mesh 116
AWS Cloud Map 117
AWS Direct Connect 118
AWS Global Accelerator 118
AWS PrivateLink 119
AWS Private 5G 119
AWS Transit Gateway 119
AWS VPN 120
Elastic Load Balancing 120
Integrated Private Wireless on AWS 121
Quantum technologies 121
Robotics 122
Satellite 123
Security, identity, and compliance 124
Amazon Cognito 126
Amazon Detective 126
Amazon GuardDuty 127
Amazon Inspector 128
Amazon Macie 128
Amazon Security Lake 129
Amazon Verified Permissions 130
AWS Artifact 130
AWS Audit Manager 130
AWS Certificate Manager 131
AWS CloudHSM 131
AWS Directory Service 132
AWS Firewall Manager 132
AWS Identity and Access Management 132
AWS Key Management Service 133
AWS Network Firewall 133

AWS Resource Access Manager 134
AWS Secrets Manager 134
AWS Security Hub 135
AWS Shield 135
AWS IAM Identity Center 136
AWS WAF 137
AWS WAF Captcha 137
Storage 137
AWS Backup 139
Amazon Elastic Block Store 139
AWS Elastic Disaster Recovery 139
Amazon Elastic File System 140
Amazon File Cache 140
Amazon FSx for Lustre 141
Amazon FSx for NetApp ONTAP 141
Amazon FSx for OpenZFS 142
Amazon FSx for Windows File Server 142
Amazon Simple Storage Service 142
AWS Storage Gateway 143
Next steps 145
Are you Well-Architected? 145
Conclusion 147
Resources 148
Document history 149
..... 149
AWS Glossary 154

Overview of Amazon Web Services

Publication date: **August 27, 2024** (Document history)

Amazon Web Services offers a broad set of global cloud-based products including compute, storage, databases, analytics, networking, mobile, developer tools, management tools, IoT, security, and enterprise applications: on-demand, available in seconds, with pay-as-you-go pricing. From data warehousing to deployment tools, directories to content delivery, over 200 AWS services are available.

New services can be provisioned quickly, without the upfront fixed expense. This allows enterprises, start-ups, small and medium-sized businesses, and customers in the public sector to access the building blocks they need to respond quickly to changing business requirements. This whitepaper provides you with an overview of the benefits of the AWS Cloud and introduces you to the services that make up the platform.

Introduction

In 2006, Amazon Web Services (AWS) began offering IT infrastructure services to businesses as web services—now commonly known as cloud computing. One of the key benefits of cloud computing is the opportunity to replace upfront capital infrastructure expenses with low variable costs that scale with your business. With the cloud, businesses no longer need to plan for and procure servers and other IT infrastructure weeks or months in advance. Instead, they can instantly spin up hundreds or thousands of servers in minutes and deliver results faster.

Today, AWS provides a highly reliable, scalable, low-cost infrastructure platform in the cloud that powers hundreds of thousands of businesses in 190 countries around the world.

This video explores how millions of customers use AWS to take advantage of the efficiencies of cloud computing: What is AWS? | Amazon Web Services

What is cloud computing?

Cloud computing is the on-demand delivery of compute power, database, storage, applications, and other IT resources through a cloud services platform via the internet with pay-as-you-go pricing. Whether you are running applications that share photos to millions of mobile users or you're supporting the critical operations of your business, a cloud services platform provides rapid access to flexible and low-cost IT resources. With cloud computing, you don't need to make large upfront investments in hardware and spend a lot of time on the heavy lifting of managing that hardware. Instead, you can provision exactly the right type and size of computing resources you need to power your newest bright idea or operate your IT department. You can access as many resources as you need, almost instantly, and only pay for what you use.

Cloud computing provides a simple way to access servers, storage, databases and a broad set of application services over the internet. A cloud services platform such as Amazon Web Services owns and maintains the network-connected hardware required for these application services, while you provision and use what you need via a web application.

Six advantages of cloud computing

- **Trade fixed expense for variable expense** – Instead of having to invest heavily in data centers and servers before you know how you're going to use them, you can pay only when you consume computing resources, and pay only for how much you consume.
- **Benefit from massive economies of scale** – By using cloud computing, you can achieve a lower variable cost than you can get on your own. Because usage from hundreds of thousands of customers is aggregated in the cloud, providers such as AWS can achieve higher economies of scale, which translates into lower pay as-you-go prices.
- **Stop guessing capacity** – Eliminate guessing on your infrastructure capacity needs. When you make a capacity decision prior to deploying an application, you often end up either sitting on expensive idle resources or dealing with limited capacity. With cloud computing, these problems go away. You can access as much or as little capacity as you need, and scale up and down as required with only a few minutes' notice.
- **Increase speed and agility** – In a cloud computing environment, new IT resources are only a click away, which means that you reduce the time to make those resources available to your developers from weeks to just minutes. This results in a dramatic increase in agility for the organization, since the cost and time it takes to experiment and develop is significantly lower.
- **Stop spending money running and maintaining data centers** – Focus on projects that differentiate your business, not the infrastructure. Cloud computing lets you focus on your own customers, rather than on the heavy lifting of racking, stacking, and powering servers.
- **Go global in minutes** – Easily deploy your application in multiple regions around the world with just a few clicks. This means you can provide lower latency and a better experience for your customers at minimal cost.

Types of cloud computing

Cloud computing provides developers and IT departments with the ability to focus on what matters most and avoid undifferentiated work such as procurement, maintenance, and capacity planning. As cloud computing has grown in popularity, several different models and deployment strategies have emerged to help meet specific needs of different users. Each type provides you with different levels of control, flexibility, and management.

Deployment models

Cloud

A cloud-based application is fully deployed in the cloud and all parts of the application run in the cloud. Applications in the cloud have either been created in the cloud or have been migrated from an existing infrastructure to take advantage of the benefits of cloud computing. Cloud-based applications can be built on low-level infrastructure pieces or can use higher level services that provide abstraction from the management, architecting, and scaling requirements of core infrastructure.

Private cloud (on-premises)

The deployment of resources on-premises, using virtualization and resource management tools, is sometimes called the *private cloud*. On-premises deployment doesn't provide many of the benefits of cloud computing but it is sometimes sought for its ability to provide dedicated resources. In most cases, this deployment model is the same as legacy IT infrastructure while using application management and virtualization technologies to try and increase resource utilization.

Hybrid

A hybrid deployment is a way to connect infrastructure and applications between cloud-based resources and existing resources that are not located in the cloud. The most common method of hybrid deployment is between the cloud and existing on-premises infrastructure to extend, and grow, an organization's infrastructure into the cloud while connecting cloud resources to the internal system. For more information on how AWS can help you with your hybrid deployment, visit our AWS Solutions for Hybrid and Multicloud page.

Global infrastructure

The AWS Cloud infrastructure is built around AWS Regions and Availability Zones. An AWS Region is a physical location in the world where we have multiple Availability Zones. Availability Zones consist of one or more discrete data centers, each with redundant power, networking, and connectivity, housed in separate facilities. These Availability Zones offer you the ability to operate production applications and databases that are more highly available, fault tolerant, and scalable than would be possible from a single data center. For the latest information on the AWS Cloud Availability Zones and AWS Regions, refer to AWS Global Infrastructure.

Security and compliance

Security

Cloud security at AWS is the highest priority. As organizations embrace the scalability and flexibility of the cloud, AWS is helping them evolve security, identity, and compliance into key business enablers. AWS builds security into the core of our cloud infrastructure, and offers foundational services to help organizations meet their unique security requirements in the cloud.

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

An advantage of the AWS Cloud is that it allows you to scale and innovate, while maintaining a secure environment and paying only for the services you use. This means that you can have the security you need at a lower cost than in an on-premises environment.

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

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

AWS provides you with guidance and expertise through online resources, personnel, and partners. AWS provides you with advisories for current issues, plus you have the opportunity to work with AWS when you encounter security issues.

You get access to hundreds of tools and features to help you to meet your security objectives. AWS provides security-specific tools and features across network security, configuration management, access control, and data encryption.

Finally, AWS environments are continuously audited, with certifications from accreditation bodies across geographies and verticals. In the AWS environment, you can take advantage of automated tools for asset inventory and privileged access reporting.

Benefits of AWS security

- **Keep Your data safe** — The AWS infrastructure puts strong safeguards in place to help protect your privacy. All data is stored in highly secure AWS data centers.
- **Meet compliance requirements** — AWS manages dozens of compliance programs in its infrastructure. This means that segments of your compliance have already been completed.
- **Save money** — Cut costs by using AWS data centers. Maintain the highest standard of security without having to manage your own facility
- **Scale quickly** — Security scales with your AWS Cloud usage. No matter the size of your business, the AWS infrastructure is designed to keep your data safe.

Compliance

AWS Cloud Compliance helps you understand the robust controls in place at AWS for security and data protection in the cloud. Compliance is a shared responsibility between AWS and the customer, and you can visit the Shared Responsibility Model to learn more. Customers can feel confident in operating and building on top of the security controls AWS uses on its infrastructure.

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

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

AWS provides customers a wide range of information on its IT control environment in whitepapers, reports, certifications, accreditations, and other third-party attestations. More information is available in the Risk and Compliance whitepaper and the AWS Security Center.

AWS services by category

AWS consists of many cloud services that you can use in combinations tailored to your business or organizational needs. This section introduces the major AWS services by category. Choose a category to explore its services.

To access the services, you can use the AWS Management Console, the AWS Command Line Interface (AWS CLI), or Software Development Kits (SDKs).

Topics

- Accessing AWS services
- Analytics
- Application integration
- Blockchain
- Business applications
- Cloud Financial Management
- Compute
- Customer enablement
- Containers
- Databases
- Developer tools
- End user computing
- Frontend web and mobile services
- Game tech
- Internet of Things (IoT)
- Machine Learning (ML) and Artificial Intelligence (AI)
- Management and governance
- Media
- Migration and transfer
- Networking and content delivery
- Quantum technologies

- Robotics
- Satellite
- Security, identity, and compliance
- Storage

Accessing AWS services

AWS Management Console

Access and manage Amazon Web Services through the AWS Management Console, a simple and intuitive user interface. You can also use the AWS Management Console Application to quickly view resources on the go.

AWS Command Line Interface (AWS CLI)

The AWS Command Line Interface (AWS CLI) is a unified tool to manage your AWS services. With just one tool to download and configure, you can control multiple AWS services from the command line and automate them through scripts.

AWS CloudShell, which can be found next to the search bar in the AWS Management Console, provides a browser-based shell that is pre-authenticated with your console credentials. Using CloudShell, you can quickly run AWS commands and scripts without leaving your web browser.

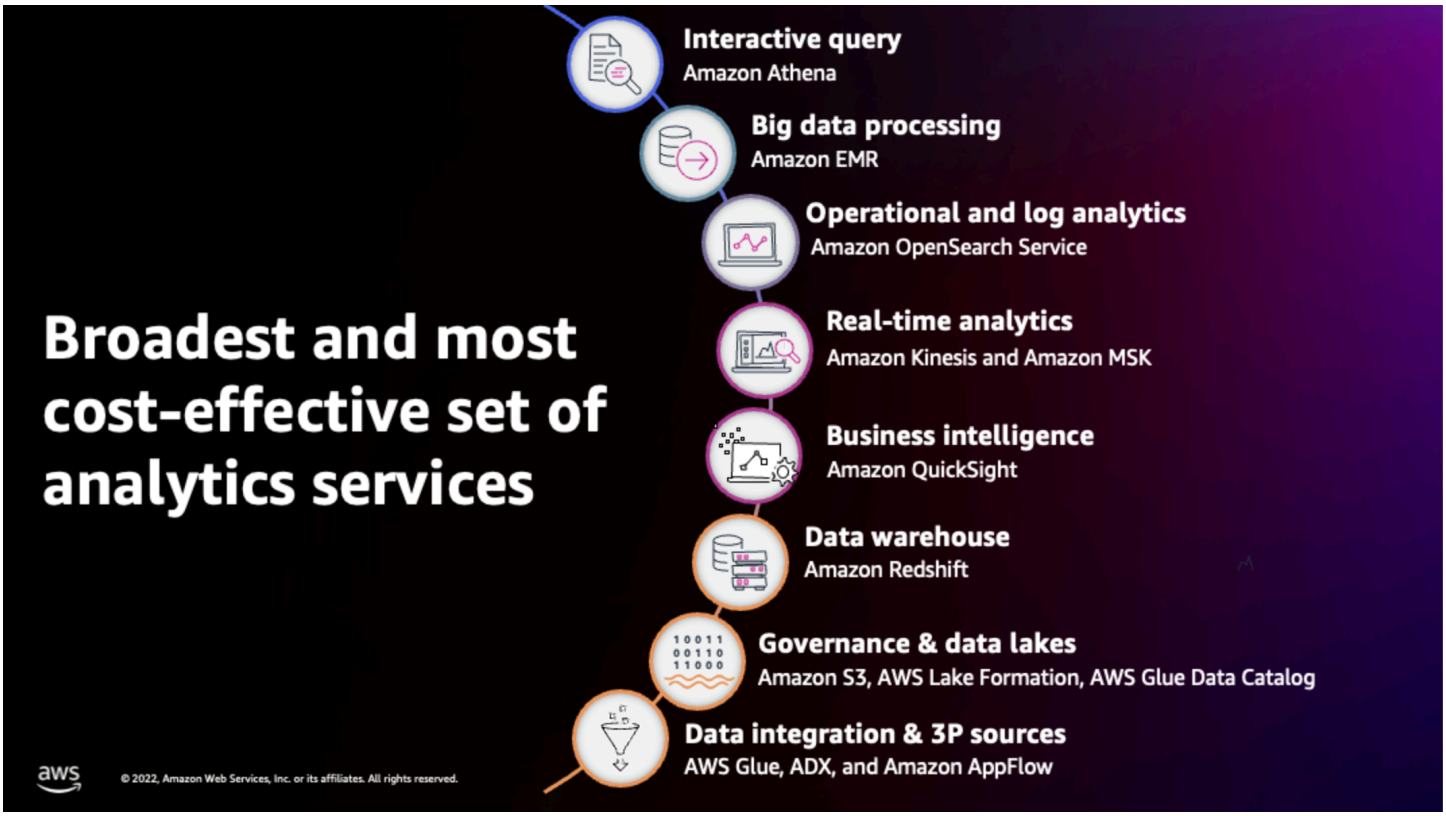
Software Development Kits (SDKs)

Our Software Development Kits (SDKs) simplify using AWS services in your applications with an Application Program Interface (API) tailored to your programming language or platform.

Analytics

AWS provides a comprehensive set of analytics services that fit all your data analytics needs and enables organizations of all sizes and industries to reinvent their business with data. From storage and management, data governance, actions, and experiences, AWS offers purpose-built services that provide the best price-performance, scalability, and lowest cost.

Each service is described after the diagram. To help you decide which service best meets your needs, see Choosing an AWS analytics service. For general information, see Analytics on AWS.



Analytics services

- Amazon Athena
- Amazon CloudSearch
- Amazon DataZone
- Amazon EMR
- Amazon FinSpace
- Amazon Kinesis
- Amazon Data Firehose
- Amazon Managed Service for Apache Flink
- Amazon Kinesis Data Streams
- Amazon Kinesis Video Streams
- Amazon OpenSearch Service
- Amazon OpenSearch Serverless
- Amazon Redshift

- Amazon Redshift Serverless
- QuickSight
- AWS Clean Rooms
- AWS Data Exchange
- AWS Data Pipeline
- AWS Entity Resolution
- AWS Glue
- AWS Lake Formation
- Amazon Managed Streaming for Apache Kafka (Amazon MSK)

Amazon Athena

Amazon Athena is an interactive query service that makes it easy to analyze data in Amazon S3 using standard SQL. Athena is serverless, so there is no infrastructure to manage, and you pay only for the queries that you run.

Athena is easy to use. Simply point to your data in Amazon S3, define the schema, and start querying using standard SQL. Most results are delivered within seconds. With Athena, there's no need for complex extract, transform, and load (ETL) jobs to prepare your data for analysis. This makes it easy for anyone with SQL skills to quickly analyze large-scale datasets.

Athena is out-of-the-box integrated with AWS Glue Data Catalog, allowing you to create a unified metadata repository across various services, crawl data sources to discover schemas and populate your Catalog with new and modified table and partition definitions, and maintain schema versioning.

Amazon CloudSearch

Amazon CloudSearch is a managed service in the AWS Cloud that makes it simple and cost-effective to set up, manage, and scale a search solution for your website or application. Amazon CloudSearch supports 34 languages and popular search features such as highlighting, autocomplete, and geospatial search.

Amazon DataZone

Amazon DataZone is a data management service that you can use to publish data and make it available to the business data catalog through your personalized web application. You can

access your data more securely regardless of where it is stored—on AWS, on premises, or in SaaS applications such as Salesforce. Amazon DataZone simplifies your experience across AWS services such as Amazon Redshift, Amazon Athena, AWS Glue, AWS Lake Formation, and QuickSight.

Amazon EMR

Amazon EMR is the industry-leading cloud big data platform for processing vast amounts of data using open source tools such as Apache Spark, Apache Hive, Apache HBase, Apache Flink, Apache Hudi, and Presto. Amazon EMR makes it easy to set up, operate, and scale your big data environments by automating time-consuming tasks such as provisioning capacity and tuning clusters. With Amazon EMR, you can run petabyte-scale analysis at less than half of the cost of traditional on-premises solutions and over 3x faster than standard Apache Spark. You can run workloads on Amazon EC2 instances, on Amazon Elastic Kubernetes Service (Amazon EKS) clusters, or on-premises using Amazon EMR on AWS Outposts.

Amazon FinSpace

Amazon FinSpace is a data management and analytics service purpose-built for the financial services industry (FSI). FinSpace reduces the time you spend finding and preparing petabytes of financial data to be ready for analysis from months to minutes.

Financial services organizations analyze data from internal data stores such as portfolio, actuarial, and risk management systems as well as petabytes of data from third-party data feeds, such as historical securities prices from stock exchanges. It can take months to find the right data, get permissions to access the data in a compliant way, and prepare it for analysis.

FinSpace removes the heavy lifting of building and maintaining a data management system for financial analytics. With FinSpace, you collect data and catalog it by relevant business concepts such as asset class, risk classification, or geographic region. FinSpace makes it easy to discover and share data across your organization in accordance with your compliance requirements. You define your data access policies in one place and FinSpace enforces them while keeping audit logs to allow for compliance and activity reporting. FinSpace also includes a library of 100+ functions, such as time bars and Bollinger bands, for you to prepare data for analysis.

Amazon Kinesis

Amazon Kinesis makes it easy to collect, process, and analyze real-time, streaming data so you can get timely insights and react quickly to new information. Amazon Kinesis offers key capabilities to cost-effectively process streaming data at any scale, along with the flexibility to choose the

tools that best suit the requirements of your application. With Amazon Kinesis, you can ingest real-time data such as video, audio, application logs, website clickstreams, and IoT telemetry data for machine learning (ML), analytics, and other applications. Amazon Kinesis enables you to process and analyze data as it arrives and respond instantly instead of having to wait until all your data is collected before the processing can begin.

Amazon Kinesis currently offers four services: Firehose, Managed Service for Apache Flink, Kinesis Data Streams, and Kinesis Video Streams.

Amazon Data Firehose

Amazon Data Firehose is the easiest way to reliably load streaming data into data stores and analytics tools. It can capture, transform, and load streaming data into Amazon S3, Amazon Redshift, Amazon OpenSearch Service, and Splunk, enabling near real-time analytics with existing business intelligence tools and dashboards you're already using today. It is a fully managed service that automatically scales to match the throughput of your data and requires no ongoing administration. It can also batch, compress, transform, and encrypt the data before loading it, minimizing the amount of storage used at the destination and increasing security.

You can easily create a Firehose delivery stream from the AWS Management Console, configure it with a few clicks, and start sending data to the stream from hundreds of thousands of data sources to be loaded continuously to AWS—all in just a few minutes. You can also configure your delivery stream to automatically convert the incoming data to columnar formats such as Apache Parquet and Apache ORC, before the data is delivered to Amazon S3, for cost-effective storage and analytics.

Amazon Managed Service for Apache Flink

Amazon Managed Service for Apache Flink is the easiest way to analyze streaming data, gain actionable insights, and respond to your business and customer needs in real time. Amazon Managed Service for Apache Flink reduces the complexity of building, managing, and integrating streaming applications with other AWS services. SQL users can easily query streaming data or build entire streaming applications using templates and an interactive SQL editor. Java developers can quickly build sophisticated streaming applications using open source Java libraries and AWS integrations to transform and analyze data in real-time.

Amazon Managed Service for Apache Flink takes care of everything required to run your queries continuously and scales automatically to match the volume and throughput rate of your incoming data.

Amazon Kinesis Data Streams

Amazon Kinesis Data Streams is a massively scalable and durable real-time data streaming service. Kinesis Data Streams can continuously capture gigabytes of data per second from hundreds of thousands of sources such as website clickstreams, database event streams, financial transactions, social media feeds, IT logs, and location-tracking events. The data collected is available in milliseconds to enable real-time analytics use cases such as real-time dashboards, real-time anomaly detection, dynamic pricing, and more.

Amazon Kinesis Video Streams

Amazon Kinesis Video Streams makes it easy to securely stream video from connected devices to AWS for analytics, ML, playback, and other processing. Kinesis Video Streams automatically provisions and elastically scales all the infrastructure needed to ingest streaming video data from millions of devices. It also durably stores, encrypts, and indexes video data in your streams, and allows you to access your data through easy-to-use APIs. Kinesis Video Streams enables you to playback video for live and on-demand viewing, and quickly build applications that take advantage of computer vision and video analytics through integration with Amazon Rekognition Video, and libraries for ML frameworks such as Apache MxNet, TensorFlow, and OpenCV.

Amazon OpenSearch Service

Amazon OpenSearch Service (OpenSearch Service) makes it easy to deploy, secure, operate, and scale OpenSearch to search, analyze, and visualize data in real-time. With Amazon OpenSearch Service, you get easy-to-use APIs and real-time analytics capabilities to power use-cases such as log analytics, full-text search, application monitoring, and clickstream analytics, with enterprise-grade availability, scalability, and security. The service offers integrations with open-source tools such as OpenSearch Dashboards and Logstash for data ingestion and visualization. It also integrates seamlessly with other AWS services such as Amazon Virtual Private Cloud (Amazon VPC), AWS Key Management Service (AWS KMS), Amazon Data Firehose, AWS Lambda, AWS Identity and Access Management (IAM), Amazon Cognito, and Amazon CloudWatch, so that you can go from raw data to actionable insights quickly.

Amazon OpenSearch Serverless

Amazon OpenSearch Serverless is a serverless option in Amazon OpenSearch Service. As a developer, you can use OpenSearch Serverless to run petabyte-scale workloads without

configuring, managing, and scaling OpenSearch clusters. You get the same interactive millisecond response times as OpenSearch Service with the simplicity of a serverless environment.

The vector engine for Amazon OpenSearch Serverless, adds a simple, scalable, and high-performing vector storage and search capability to help developers build ML-augmented search experiences and generative AI applications without having to manage vector database infrastructure. Use cases for vector search collections include image search, document search, music retrieval, product recommendation, video search, location-based search, fraud detection, and anomaly detection.

Amazon Redshift

Amazon Redshift is the most widely used cloud data warehouse. It makes it fast, simple and cost-effective to analyze all your data using standard SQL and your existing Business Intelligence (BI) tools. It allows you to run complex analytic queries against terabytes to petabytes of structured and semi-structured data, using sophisticated query optimization, columnar storage on high-performance storage, and massively parallel query completion. Most results come back in seconds. You can start small for just $0.25 per hour with no commitments and scale out to petabytes of data for $1,000 per terabyte per year, less than a tenth the cost of traditional on-premises solutions.

Amazon Redshift Serverless

Amazon Redshift Serverless makes it easier to run and scale analytics without having to manage your data warehouse infrastructure. Developers, data scientists, and analysts can work across databases, data warehouses, and data lakes to build reporting and dashboarding applications, perform near real-time analytics, share and collaborate on data, and build and train machine learning (ML) models. Go from large amounts of data to insights in seconds. Amazon Redshift Serverless automatically provisions and intelligently scales data warehouse capacity to deliver fast performance for even the most demanding and unpredictable workloads, and you pay only for what you use. Just load data and start querying right away in Amazon Redshift Query Editor or in your favorite business intelligence (BI) tool and continue to enjoy the best price performance and familiar SQL features in an easy-to-use, zero administration environment.

QuickSight

QuickSight is a fast, cloud-powered business intelligence (BI) service that makes it easy for you to deliver insights to everyone in your organization. QuickSight lets you create and publish interactive dashboards that can be accessed from browsers or mobile devices. You can embed dashboards

into your applications, providing your customers with powerful self-service analytics. QuickSight easily scales to tens of thousands of users without any software to install, servers to deploy, or infrastructure to manage.

AWS Clean Rooms

AWS Clean Rooms helps companies and their partners more easily and securely analyze and collaborate on their collective datasets–without sharing or copying one another's underlying data. With AWS Clean Rooms, customers can create a secure data clean room in minutes, and collaborate with any other company on the AWS Cloud to generate unique insights about advertising campaigns, investment decisions, and research and development.

AWS Data Exchange

AWS Data Exchange makes it easy to find, subscribe to, and use third-party data in the cloud. Qualified data providers include category-leading brands such as Reuters, who curate data from over 2.2 million unique news stories per year in multiple languages; Change Healthcare, who process and anonymize more than 14 billion healthcare transactions and $1 trillion in claims annually; Dun & Bradstreet, who maintain a database of more than 330 million global business records; and Foursquare, whose location data is derived from 220 million unique consumers and includes more than 60 million global commercial venues.

Once subscribed to a data product, you can use the AWS Data Exchange API to load data directly into Amazon S3 and then analyze it with a wide variety of AWS analytics and ML services. For example, property insurers can subscribe to data to analyze historical weather patterns to calibrate insurance coverage requirements in different geographies; restaurants can subscribe to population and location data to identify optimal regions for expansion; academic researchers can conduct studies on climate change by subscribing to data on carbon dioxide emissions; and healthcare professionals can subscribe to aggregated data from historical clinical trials to accelerate their research activities.

For data providers, AWS Data Exchange makes it easy to reach the millions of AWS customers migrating to the cloud by removing the need to build and maintain infrastructure for data storage, delivery, billing, and entitling.

AWS Data Pipeline

AWS Data Pipeline is a web service that helps you reliably process and move data between different AWS compute and storage services, as well as on-premises data sources, at specified

intervals. With AWS Data Pipeline, you can regularly access your data where it's stored, transform and process it at scale, and efficiently transfer the results to AWS services such as Amazon S3, Amazon RDS, Amazon DynamoDB, and Amazon EMR.

AWS Data Pipeline helps you easily create complex data processing workloads that are fault tolerant, repeatable, and highly available. You don't have to worry about ensuring resource availability, managing inter-task dependencies, retrying transient failures or timeouts in individual tasks, or creating a failure notification system. AWS Data Pipeline also allows you to move and process data that was previously locked up in on-premises data silos.

AWS Entity Resolution

AWS Entity Resolution is a service that helps you match and link related records stored across multiple applications, channels, and data stores without building a custom solution. Using flexible, configurable ML and rule-based techniques, AWS Entity Resolution can remove duplicate records, create customer profiles by connecting different customer interactions, and personalize experiences across advertising and marketing campaigns, loyalty programs, and e-commerce. For example, you can create a unified view of customer interactions by linking recent events, such as ad clicks, cart abandonment, and purchases, into a unique match ID.

AWS Glue

AWS Glue is a fully managed extract, transform, and load (ETL) service that makes it easy for customers to prepare and load their data for analytics. You can create and run an ETL job with a few clicks in the AWS Management Console. You simply point AWS Glue to your data stored in AWS, and AWS Glue discovers your data and stores the associated metadata (such as table definition and schema) in the AWS Glue Data Catalog. Once cataloged, your data is immediately searchable, queryable, and available for ETL.

AWS Glue Data Integration Engines provide access to data using Apache Spark, PySpark, and Python. With the addition of AWS Glue for Ray, you can further scale your workloads using Ray, an open-source unified compute framework.

AWS Glue Data Quality can measure and monitor the data quality of Amazon S3 based data lakes, data warehouses, and other data repositories. It automatically computes statistics, recommends quality rules, and can monitor and alert you when it detects missing, stale, or bad data. You can access it in the AWS Glue Data Catalog and in AWS Glue Data Catalog ETL jobs.

AWS Lake Formation

AWS Lake Formation is a service that makes it easy to set up a secure data lake in days. A data lake is a centralized, curated, and secured repository that stores all your data, both in its original form and prepared for analysis. A data lake enables you to break down data silos and combine different types of analytics to gain insights and guide better business decisions.

However, setting up and managing data lakes today involves a lot of manual, complicated, and time-consuming tasks. This work includes loading data from diverse sources, monitoring those data flows, setting up partitions, turning on encryption and managing keys, defining transformation jobs and monitoring their operation, re-organizing data into a columnar format, configuring access control settings, deduplicating redundant data, matching linked records, granting access to data sets, and auditing access over time.

Creating a data lake with Lake Formation is as simple as defining where your data resides and what data access and security policies you want to apply. Lake Formation then collects and catalogs data from databases and object storage, moves the data into your new Amazon S3 data lake, cleans and classifies data using ML algorithms, and secures access to your sensitive data. Your users can then access a centralized catalog of data which describes available data sets and their appropriate usage. Your users then leverage these data sets with their choice of analytics and ML services, such as Amazon EMR for Apache Spark, Amazon Redshift, Amazon Athena, SageMaker AI, and QuickSight.

Amazon Managed Streaming for Apache Kafka (Amazon MSK)

Amazon Managed Streaming for Apache Kafka (Amazon MSK) is a fully managed service that makes it easy for you to build and run applications that use Apache Kafka to process streaming data. Apache Kafka is an open-source platform for building real-time streaming data pipelines and applications. With Amazon MSK, you can use Apache Kafka APIs to populate data lakes, stream changes to and from databases, and power ML and analytics applications.

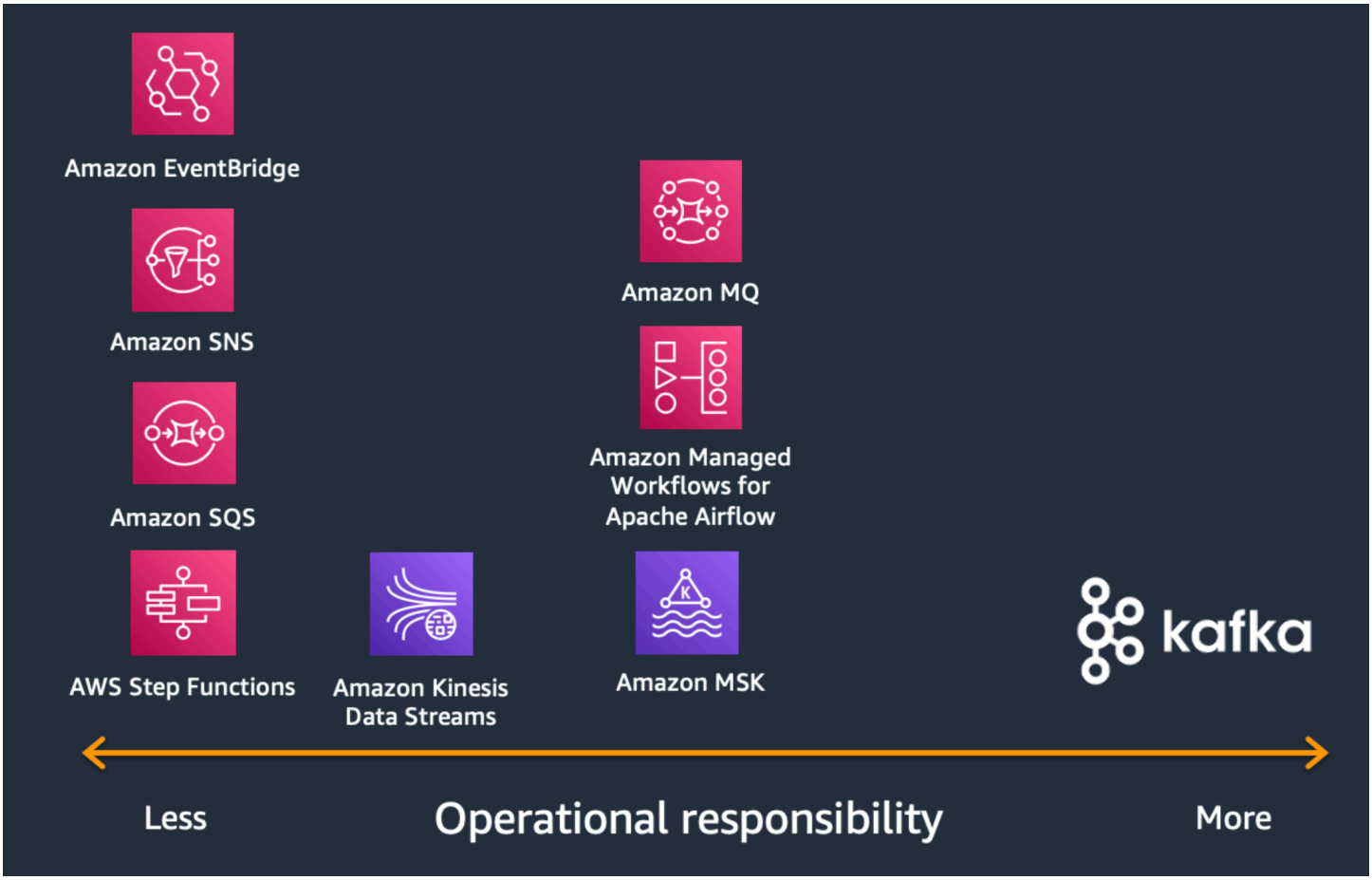
Apache Kafka clusters are challenging to setup, scale, and manage in production. When you run Apache Kafka on your own, you need to provision servers, configure Apache Kafka manually, replace servers when they fail, orchestrate server patches and upgrades, architect the cluster for high availability, ensure data is durably stored and secured, setup monitoring and alarms, and carefully plan scaling events to support load changes. Amazon MSK makes it easy for you to build and run production applications on Apache Kafka without needing Apache Kafka infrastructure management expertise. That means you spend less time managing infrastructure and more time building applications.

With a few clicks in the Amazon MSK console you can create highly available Apache Kafka clusters with settings and configuration based on Apache Kafka's deployment best practices. Amazon MSK automatically provisions and runs your Apache Kafka clusters. Amazon MSK continuously monitors cluster health and automatically replaces unhealthy nodes with no downtime to your application. In addition, Amazon MSK secures your Apache Kafka cluster by encrypting data at rest.

Application integration

Application integration on AWS is a suite of services that enable communication between decoupled components within microservices, distributed systems, and serverless applications. You don't need to refactor your entire architecture to benefit—decoupling applications at any scale can reduce the impact of changes, making it easier to update and faster to release new features.

Each service is described after the diagram. To help you decide which service best meets your needs, see Choosing an AWS application integration service or Amazon SQS, Amazon SNS, or Amazon EventBridge?. For general information, see Application Integration on AWS.



Services

- AWS Step Functions
- Amazon AppFlow
- AWS B2B Data Interchange
- Amazon EventBridge
- Amazon Managed Workflows for Apache Airflow (MWAA)
- Amazon MQ
- Amazon Simple Notification Service
- Amazon Simple Queue Service
- Amazon Simple Workflow Service

AWS Step Functions

AWS Step Functions is a fully managed service that makes it easy to coordinate the components of distributed applications and microservices using visual workflows. Building applications from individual components that each perform a discrete function lets you scale easily and change applications quickly. Step Functions is a reliable way to coordinate components and step through the functions of your application. Step Functions provides a graphical console to arrange and visualize the components of your application as a series of steps. This makes it simple to build and run multi-step applications. Step Functions automatically initiates and tracks each step, and retries when there are errors, so your application runs in order and as expected. Step Functions logs the state of each step, so when things do go wrong, you can diagnose and debug problems quickly. You can change and add steps without even writing code, so you can easily evolve your application and innovate faster.

Amazon AppFlow

Amazon AppFlow is a fully managed integration service that enables you to securely transfer data between Software-as-a-Service (SaaS) applications such as Salesforce, Zendesk, Slack, and ServiceNow, and AWS services such as Amazon S3 and Amazon Redshift, in just a few clicks. With Amazon AppFlow, you can run data flows at enterprise scale at the frequency you choose - on a schedule, in response to a business event, or on demand. You can configure data transformation capabilities such as filtering and validation to generate rich, ready-to-use data as part of the flow itself, without additional steps. Amazon AppFlow; automatically encrypts data in motion, and allows users to restrict data from flowing over the public internet for SaaS applications that are integrated with AWS PrivateLink, reducing exposure to security threats.

AWS B2B Data Interchange

AWS B2B Data Interchange (B2Bi) automates the transformation of Electronic Data Interchange (EDI) documents into JSON and XML formats to simplify your downstream data integrations. Businesses use EDI documents to exchange transactional data with trading partners, such as suppliers and end customers, using standardized formats such as X12.

With B2Bi, you can onboard and manage your trading partners and automate the transformation of EDI documents into common data representations such as JSON and XML using a low-code interface. This approach reduces the time, complexity, and cost associated with preparing and integrating EDI data into their business applications and purpose-built data lakes. As a result,

you can concentrate on using transactional data to drive business insights using the AWS suite of analytics, AI, and ML services.

Amazon EventBridge

Amazon EventBridge is a serverless event bus that makes it easier to build event-driven applications at scale using events generated from your applications, integrated Software-as-a-Service (SaaS) applications, and AWS services. EventBridge delivers a stream of real-time data from event sources such as Zendesk or Shopify to targets such as AWS Lambda and other SaaS applications. You can set up routing rules to determine where to send your data to build application architectures that react in real-time to your data sources with event publisher and consumer completely decoupled.

Amazon Managed Workflows for Apache Airflow (MWAA)

Amazon Managed Workflows for Apache Airflow (MWAA) is a managed orchestration service for Apache Airflow that makes it easier to set up and operate end-to-end data pipelines in the cloud at scale. Apache Airflow is an open-source tool used to programmatically author, schedule, and monitor sequences of processes and tasks referred to as “workflows.” With Managed Workflows, you can use Airflow and Python to create workflows without having to manage the underlying infrastructure for scalability, availability, and security. Managed Workflows automatically scales its workflow capacity to meet your needs, and is integrated with AWS security services to help provide you with fast and secure access to data.

Amazon MQ

Amazon MQ is a managed message broker service for Apache ActiveMQ Classic and RabbitMQ that makes it easy to set up and operate message brokers in the cloud. Message brokers allow different software systems–often using different programming languages, and on different platforms–to communicate and exchange information. Amazon MQ reduces your operational load by managing the provisioning, setup, and maintenance of ActiveMQ and RabbitMQ, popular open-source message brokers. Connecting your current applications to Amazon MQ is easy because it uses industry-standard APIs and protocols for messaging, including JMS, NMS, AMQP, STOMP, MQTT, and WebSocket. Using standards means that in most cases, there’s no need to rewrite any messaging code when you migrate to AWS.

Amazon Simple Notification Service

Amazon Simple Notification Service (Amazon SNS) is a highly available, durable, secure, fully managed pub/sub messaging service that enables you to decouple microservices, distributed systems, and serverless applications. Amazon SNS provides topics for high-throughput, push-based, many-to-many messaging. Using Amazon SNS topics, your publisher systems can fan out messages to a large number of subscriber endpoints for parallel processing, including Amazon SQS queues, AWS Lambda functions, and HTTP/S webhooks. Additionally, SNS can be used to fan out notifications to end users using mobile push, SMS, and email.

Amazon Simple Queue Service

Amazon Simple Queue Service (Amazon SQS) is a fully managed message queuing service that enables you to decouple and scale microservices, distributed systems, and serverless applications. SQS eliminates the complexity and overhead associated with managing and operating message oriented middleware, and empowers developers to focus on differentiating work. Using Amazon SQS, you can send, store, and receive messages between software components at any volume, without losing messages or requiring other services to be available. Get started with Amazon SQS in minutes using the AWS Management Console, AWS CLI, or SDK of your choice, and three simple commands.

Amazon SQS offers two types of message queues. Standard queues offer maximum throughput, best-effort ordering, and at-least-once delivery. Amazon SQS FIFO queues are designed to guarantee that messages are processed exactly once, in the exact order that they are sent.

Amazon Simple Workflow Service

Amazon Simple Workflow Service (Amazon SWF) helps developers build, run, and scale background jobs that have parallel or sequential steps. You can think of Amazon SWF as a fully managed state tracker and task coordinator in the cloud. If your application's steps take more than 500 milliseconds to complete, you need to track the state of processing. If you need to recover or retry if a task fails, Amazon SWF can help you.

Blockchain

Amazon Managed Blockchain

Amazon Managed Blockchain is a fully managed service that makes it easy to create and manage scalable blockchain networks using the popular open source frameworks Hyperledger Fabric and Ethereum.

Blockchain makes it possible to build applications where multiple parties can run transactions without the need for a trusted, central authority. Today, building a scalable blockchain network with existing technologies is complex to set up and hard to manage. To create a blockchain network, each network member needs to manually provision hardware, install software, create and manage certificates for access control, and configure networking components. Once the blockchain network is running, you need to continuously monitor the infrastructure and adapt to changes, such as an increase in transaction requests, or new members joining or leaving the network.

Amazon Managed Blockchain is a fully managed service that allows you to set up and manage a scalable blockchain network with just a few clicks. Amazon Managed Blockchain eliminates the overhead required to create the network, and automatically scales to meet the demands of thousands of applications running millions of transactions. Once your network is up and running, Managed Blockchain makes it easy to manage and maintain your blockchain network. It manages your certificates, lets you easily invite new members to join the network, and tracks operational metrics such as usage of compute, memory, and storage resources. In addition, Managed Blockchain can replicate an immutable copy of your blockchain network activity into Amazon Quantum Ledger Database (Amazon QLDB), a fully managed ledger database. This allows you to easily analyze the network activity outside the network and gain insights into trends.

Business applications

Innovative business applications with the same on-demand scalability, reliability, pay-as-you go pricing, and machine learning that drives AWS cloud infrastructure.

For general information, see AWS Business Applications.

Applications

- Alexa for Business
- AWS AppFabric
- Amazon Chime
- Amazon Chime SDK
- Amazon Connect
- Amazon Pinpoint
- Amazon SES
- Amazon WorkDocs
- Amazon WorkMail

Alexa for Business

Alexa for Business is a service that enables organizations and employees to use Alexa to get more work done. With Alexa for Business, employees can use Alexa as their intelligent assistant to be more productive in meeting rooms, at their desks, and even with the Alexa devices they already have at home.

AWS AppFabric

AWS AppFabric is a fully managed service that aggregates and normalizes security data across software as a service (SaaS) applications. Previously, integrating SaaS applications with existing security tools required teams to build, manage, and maintain their own point-to-point (P2P) integrations so that security teams could monitor event logs and understand activity from each application. With AppFabric, you can quickly connect multiple SaaS applications to increase observability, productivity, and security—with no coding required.

After the SaaS applications are authorized and connected, AppFabric ingests the data and normalizes it using the Open Cybersecurity Schema Framework (OCSF). OCSF allows you to set common policies, standardize security alerts, and quickly manage user access across multiple applications.

Amazon Chime

Amazon Chime is a communications service that transforms online meetings with a secure, easy-to-use application that you can trust. Amazon Chime works seamlessly across your devices so that

you can stay connected. You can use Amazon Chime for online meetings, video conferencing, calls, chat, and to share content, both inside and outside your organization.

Amazon Chime works with Alexa for Business, which means you can use Alexa to start your meetings with your voice. Alexa can start your video meetings in large conference rooms, and automatically dial into online meetings in smaller huddle rooms and from your desk.

Amazon Chime SDK

With the Amazon Chime SDK, builders can easily add real-time voice, video, and messaging powered by ML into their applications.

Amazon Connect

Amazon Connect is a self-service, omnichannel cloud contact center service that makes it easy for any business to deliver better customer service at lower cost. Amazon Connect is based on the same contact center technology used by Amazon customer service associates around the world to power millions of customer conversations. The self-service graphical interface in Amazon Connect makes it easy for non-technical users to design contact flows, manage agents, and track performance metrics – no specialized skills required. There are no up-front payments or long-term commitments and no infrastructure to manage with Amazon Connect; customers pay by the minute for Amazon Connect usage plus any associated telephony services.

Amazon Pinpoint

Amazon Pinpoint makes it easy to send targeted messages to your customers through multiple engagement channels. Examples of targeted campaigns are promotional alerts and customer retention campaigns, and transactional messages are messages such as order confirmations and password reset messages.

You can integrate Amazon Pinpoint into your mobile and web apps to capture usage data to provide you with insight into how customers interact with your apps. Amazon Pinpoint also tracks the ways that your customers respond to the messages you send—for example, by showing you the number of messages that were delivered, opened, or clicked.

You can develop custom audience segments and send them pre-scheduled targeted campaigns via email, SMS, and push notifications. Targeted campaigns are useful for sending promotional or educational content to re-engage and retain your users.

You can send transactional messages using the console or the Amazon Pinpoint REST API. Transactional campaigns can be sent via email, SMS, push notifications, and voice messages. You can also use the API to build custom applications that deliver campaign and transactional messages.

Amazon SES

Amazon Simple Email Service (Amazon SES) is a cost-effective, flexible, and scalable email service that enables developers to send mail from within any application. You can configure Amazon SES quickly to support several email use cases, including transactional, marketing, or mass email communications. The Amazon SES flexible IP deployment and email authentication options help drive higher deliverability and protect sender reputation, while sending analytics measure the impact of each email. With Amazon SES, you can send email securely, globally, and at scale.

Amazon WorkDocs

> **Notice**
>
> New customer sign-ups and account upgrades are no longer available for Amazon WorkDocs. Learn about migration steps here: How to migrate data from Amazon WorkDocs.

Amazon WorkDocs is a fully managed, secure enterprise storage and sharing service with strong administrative controls and feedback capabilities that improve user productivity.

Users can comment on files, send them to others for feedback, and upload new versions without having to resort to emailing multiple versions of their files as attachments. Users can take advantage of these capabilities wherever they are, using the device of their choice, including PCs, Macs, tablets, and phones. Amazon WorkDocs offers IT administrators the option of integrating with existing corporate directories, flexible sharing policies and control of the location where data is stored.

Amazon WorkMail

Amazon WorkMail is a secure, managed business email and calendar service with support for existing desktop and mobile email client applications. Amazon WorkMail gives users the ability to seamlessly access their email, contacts, and calendars using the client application of their choice, including Microsoft Outlook, native iOS and Android email applications, any

client application supporting the IMAP protocol, or directly through a web browser. You can integrate Amazon WorkMail with your existing corporate directory, use email journaling to meet compliance requirements, and control both the keys that encrypt your data and the location in which your data is stored. You can also set up interoperability with Microsoft Exchange Server, and programmatically manage users, groups, and resources using the Amazon WorkMail SDK.

Cloud Financial Management

Whether you were born in the cloud, or you are just starting your migration journey to the cloud, AWS has a set of solutions to help you manage and optimize your spend.

Each service is described after the diagram. To help you decide which service best meets your needs, see Choosing an AWS cost management strategy. For general information, see Cloud Financial Management with AWS.

Services

- AWS Application Cost Profiler

- AWS Billing Conductor
- AWS Cost Explorer
- AWS Budgets
- AWS Cost and Usage Report
- Reserved Instance (RI) reporting
- Savings Plans

AWS Application Cost Profiler

AWS Application Cost Profiler provides you the ability to track the consumption of shared AWS resources used by software applications and report granular cost breakdown across tenant base. You can achieve economies of scale with the shared infrastructure model, while still maintaining a clear line of sight to detailed resource consumption information across multiple dimensions.

With the proportionate cost insights of shared AWS resources, organizations running applications can establish the data foundation for accurate cost allocation model, and ISV selling applications can better understand your profitability and customize pricing strategies for your end customers.

AWS Billing Conductor

AWS Billing Conductor is a fully managed service that can support the showback and chargeback workflows of AWS Solution Providers and Enterprise customers. Using AWS Billing Conductor, you can customize your monthly billing data. The console models the billing relationship between you and your customers or business units. You can also customize a pro forma version of your billing data each month to accurately show or charge back your customers.

AWS Billing Conductor doesn't change the way that you're billed by Amazon Web Services each month. Instead, it provides you with a mechanism to configure, generate, and display rates to certain customers over a given billing period. You can also use it to analyze the difference between the rates you apply to your accounting groupings relative to your actual rates from AWS. As a result of your AWS Billing Conductor configuration, the payer account can also see the custom rate that's applied on the billing details page of the AWS Billing console, or configure a cost and usage report per billing group.

You can configure the billing groups and pricing plans using the AWS Billing Conductor or the AWS Billing Conductor API. For more information about AWS Billing Conductor service quotas, refer to Quotas and restrictions.

AWS Cost Explorer

AWS Cost Explorer has an easy-to-use interface that lets you visualize, understand, and manage your AWS costs and usage over time. Get started quickly by creating custom reports (including charts and tabular data) that analyze cost and usage data, both at a high level (such as total costs and usage across all accounts) and for highly-specific requests (such as `m2.2xlarge` costs within account Y that are tagged "`project: secretProject`").

AWS Budgets

AWS Budgets gives you the ability to set custom budgets that alert you when your costs or usage exceed (or are forecasted to exceed) your budgeted amount. You can also use AWS Budgets to set RI utilization or coverage targets and receive alerts when your utilization drops below the threshold you define. RI alerts support Amazon EC2, Amazon RDS, Amazon Redshift, and Amazon ElastiCache reservations.

Budgets can be tracked at the monthly, quarterly, or yearly level, and you can customize the start and end dates. You can further refine your budget to track costs associated with multiple dimensions, such as AWS service, linked account, tag, and others. Budget alerts can be sent via email and/or Amazon Simple Notification Service (Amazon SNS) topic.

Budgets can be created and tracked from the AWS Budgets dashboard or via the AWS Budgets API.

AWS Cost and Usage Report

The AWS Cost and Usage Report is a single location for accessing comprehensive information about your AWS costs and usage.

The AWS Cost and Usage Report lists AWS usage for each service category used by an account and its IAM users in hourly or daily line items, as well as any tags that you have activated for cost allocation purposes. You can also customize the AWS Cost and Usage Report to aggregate your usage data to the daily or monthly level.

Reserved Instance (RI) reporting

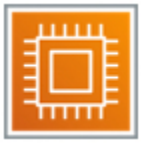
AWS provides a number of RI-specific cost management solutions out-of-the-box to help you better understand and manage your RIs. Using the RI Utilization and Coverage reports available in AWS Cost Explorer, you can visualize your RI data at an aggregate level or inspect a particular

RI subscription. To access the most detailed RI information available, you can leverage the AWS Cost and Usage Report. You can also set a custom RI utilization target via AWS Budgets and receive alerts when your utilization drops below the threshold you define.

Savings Plans

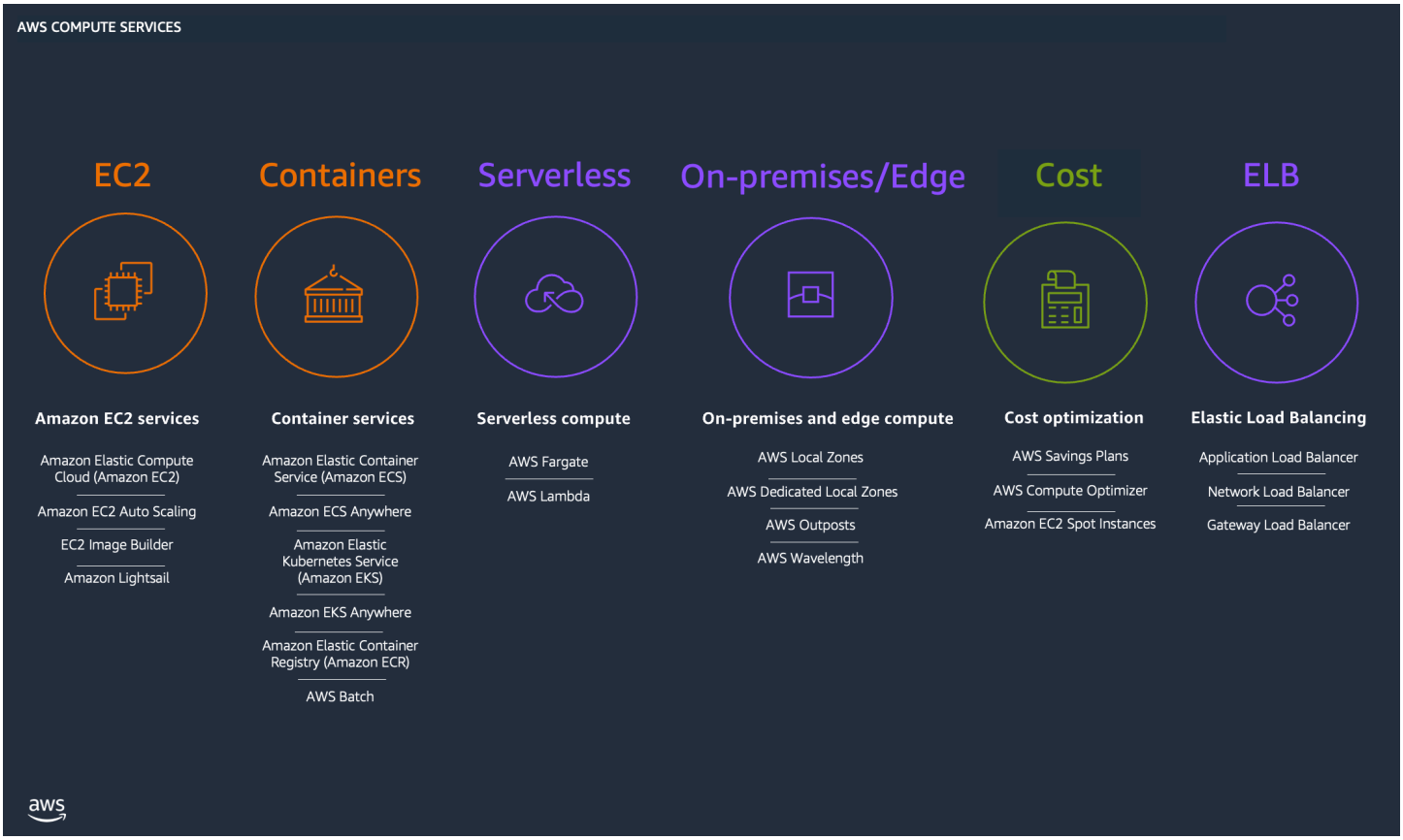
Savings Plans is a flexible pricing model offering lower prices compared to On-Demand pricing, in exchange for a specific usage commitment (measured in $/hour) for a one or three-year period. AWS offers three types of Savings Plans – Compute Savings Plans, Amazon EC2 Instance Savings Plans, and Amazon SageMaker AI Savings Plans. Compute Savings Plans apply to usage across Amazon EC2, AWS Lambda, and AWS Fargate. The Amazon EC2 Instance Savings Plans apply to EC2 usage, and Amazon SageMaker AI Savings Plans apply to Amazon SageMaker AI usage. You can easily sign up a one- or three-year term Savings Plans in AWS Cost Explorer and manage your plans by taking advantage of recommendations, performance reporting, and budget alerts.

Compute



Millions of organizations run diverse workloads using AWS compute services.

Each service is described after the diagram. To help you decide which service best meets your needs, see Choosing an AWS compute service or Amazon Lightsail, AWS Elastic Beanstalk, or Amazon EC2?. For general information, see Compute on AWS.



Topics

- Compare AWS compute services
- Amazon EC2
- Amazon EC2 Auto Scaling
- Amazon EC2 Image Builder
- Amazon Lightsail
- Amazon Linux 2023
- AWS App Runner
- AWS Batch
- AWS Elastic Beanstalk
- AWS Fargate
- AWS Lambda
- AWS Serverless Application Repository
- AWS Outposts
- AWS Wavelength

- VMware Cloud on AWS

Compare AWS compute services

Category	AWS service
Instances (virtual machines)	• Amazon Elastic Compute Cloud (Amazon EC2) — Secure and resizeable compute capacity (virtual servers) in the cloud • Amazon EC2 Spot Instances— Run fault-tol erant workloads for up to 90% off • Amazon EC2 Auto Scaling — Automatically add or remove compute capacity to meet changes in demand • Amazon Lightsail — Easy-to-use cloud platform that offers you everything you need to build an application or website • AWS Batch — Fully managed batch processing at any scale
Containers	• Amazon Elastic Container Service (Amazon ECS) — Highly secure, reliable, and scalable way to run containers • Amazon ECS Anywhere — Run containers on customer managed infrastructure • Amazon Elastic Container Registry (Amazon ECR) — Easily store, manage, and deploy container images • Amazon Elastic Kubernetes Service (Amazon EKS) — Fully managed Kubernetes service • Amazon EKS Anywhere — Create and operate Kubernetes clusters on your own infrastructure

Category	AWS service
	• AWS Fargate — Serverless compute for containers • AWS App Runner — Build and run containerized applications on a fully managed service
Serverless	• AWS Lambda — Run code without thinking about servers. Pay only for the compute time you consume.
Edge and hybrid	• AWS Outposts — Run AWS infrastructure and services on premises for a truly consistent hybrid experience • AWS Snow Family — Collect and process data in rugged or disconnected edge environments • AWS Wavelength — Deliver ultra-low latency application for 5G devices • VMware Cloud on AWS — Preferred service for all vSphere workloads to rapidly extend and migrate to the cloud • AWS Local Zones — Run latency sensitive applications closer to end-users

Category	AWS service
Cost and capacity management	• AWS Savings Plan — Flexible pricing model that provides savings of up to 72% on AWS compute usage • AWS Compute Optimizer — Recommend s optimal AWS compute resources for your workloads to reduce costs and improve performance • AWS Elastic Beanstalk — Easy-to-use service for deploying and scaling web applications and services • EC2 Image Builder — Build and maintain secure Linux or Windows Server images • Elastic Load Balancing (ELB) — Automatic ally distribute incoming application traffic across multiple targets

Amazon EC2

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers.

The simple web interface of Amazon EC2 allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment. Amazon EC2 reduces the time required to obtain and boot new server instances (called Amazon EC2 instances) to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change. Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use. Amazon EC2 provides developers and system administrators the tools to build failure resilient applications and isolate themselves from common failure scenarios.

Instance types

Amazon EC2 passes on to you the financial benefits of Amazon scale. You pay a very low rate for the compute capacity you actually consume. For a more detailed description, refer to Amazon EC2 pricing.

Amazon EC2 instance types are named based on their family, generation, processor family, additional capabilities, and size.

- **On-Demand Instances** — With On-Demand Instances, you pay for compute capacity by the hour or the second depending on which instances you run. No longer-term commitments or upfront payments are needed. You can increase or decrease your compute capacity depending on the demands of your application and only pay the specified per hourly rates for the instance you use. On-Demand Instances are recommended for:
  - Users that prefer the low cost and flexibility of Amazon EC2 without any up-front payment or long-term commitment
  - Applications with short-term, spiky, or unpredictable workloads that cannot be interrupted
  - Applications being developed or tested on Amazon EC2 for the first time
- **Spot Instances** —Spot Instances are available at up to a 90% discount compared to On-Demand prices and let you take advantage of unused Amazon EC2 capacity in the AWS Cloud. You can significantly reduce the cost of running your applications, grow your application's compute capacity and throughput for the same budget, and enable new types of cloud computing applications. Spot Instances are recommended for:
  - Applications that have flexible start and end times
  - Applications that are only feasible at very low compute prices
  - Users with urgent computing needs for large amounts of additional capacity
- **Reserved Instances** — Reserved Instances provide you with a significant discount (up to 72%) compared to On-Demand Instance pricing. You have the flexibility to change families, operating system types, and tenancies while benefiting from Reserved Instance pricing when you use Convertible Reserved Instances.
- **C7g Instances** — C7g Instances, powered by the latest generation AWS Graviton3 processors, provide the best price performance in Amazon EC2 for compute-intensive workloads. C7g instances are ideal for high performance computing (HPC), batch processing, electronic design automation (EDA), gaming, video encoding, scientific modeling, distributed analytics, CPU-based ML inference, and ad serving.

- **Inf2 Instances** — Inf2 Instances are purpose--built for deep learning inference. They deliver high performance at the lowest cost in Amazon EC2 for generative AI models, including large language models (LLMs) and vision transformers. Inf2 instances are powered by AWS Inferentia2, the second-generation AWS Inferentia accelerator.
- **M7g Instances** — M7g instances, powered by the latest generation AWS Graviton3 processors, provide the best price performance in Amazon EC2 for general purpose workloads. M7g instances are ideal for applications built on open-source software such as application servers, microservices, gaming servers, mid-size data stores, and caching fleets.
- **R7g Instances** — R7g Instances, powered by the latest generation AWS Graviton3 processors, provide the best price performance in Amazon EC2 for memory-intensive workloads. R7g instances are ideal for memory-intensive workloads such as open-source databases, in-memory caches, and near real-time big data analytics.
- **Trn1 Instances** — Trn1 Instances, powered by AWS Trainium accelerators, are purpose-built for high-performance deep learning training of generative AI models, including LLMs and latent diffusion models. Trn1 instances offer up to 50% cost-to-train savings over other comparable Amazon EC2 instances.
- **Savings Plans** — Savings Plans are a flexible pricing model that offer low prices on EC2 and Fargate usage, in exchange for a commitment to a consistent amount of usage (measured in $/hour) for a one or three year term.
- **Dedicated Hosts** — A Dedicated Host is a physical EC2 server dedicated for your use. Dedicated Hosts can help you reduce costs by allowing you to use your existing server-bound software licenses, including Windows Server, Microsoft SQL Server, and SUSE Linux Enterprise Server (subject to your license terms), and can also help you meet compliance requirements.

Amazon EC2 Auto Scaling

Amazon EC2 Auto Scaling helps you maintain application availability and allows you to automatically add or remove EC2 instances according to conditions you define. You can use the fleet management features of Amazon EC2 Auto Scaling to maintain the health and availability of your fleet. You can also use the dynamic and predictive scaling features of Amazon EC2 Auto Scaling to add or remove EC2 instances. Dynamic scaling responds to changing demand and predictive scaling automatically schedules the right number of EC2 instances based on predicted demand. Dynamic scaling and predictive scaling can be used together to scale faster.

Amazon EC2 Image Builder

EC2 Image Builder simplifies the building, testing, and deployment of VMs and container images for use on AWS or on-premises.

Keeping virtual machine (VM) and container images up-to-date can be time consuming, resource intensive, and error-prone. Currently, customers either manually update and snapshot VMs or have teams that build automation scripts to maintain images.

EC2 Image Builder significantly reduces the effort of keeping images up-to-date and secure by providing a simple graphical interface, built-in automation, and AWS-provided security settings. With Image Builder, there are no manual steps for updating an image nor do you have to build your own automation pipeline.

Image Builder is offered at no cost, other than the cost of the underlying AWS resources used to create, store, and share the images.

Amazon Lightsail

Amazon Lightsail is designed to be the easiest way to launch and manage a virtual private server with AWS. Lightsail plans include everything you need to jumpstart your project – a VM, SSD-based storage, data transfer, DNS management, and a static IP address – for a low, predictable price.

Amazon Linux 2023

Amazon Linux 2023 (AL2023) is our new Linux-based operating system for AWS that is designed to provide a secure, stable, high-performance environment to develop and run your cloud applications. AL2023 provides seamless integration with various AWS services and development tools, and offers optimized performance for Amazon EC2 Graviton-based instances and Support at no additional licensing cost. Starting with AL2023, a new Amazon Linux major release will be available every two years. This cadence provides you with a more predictable release cycle and up to 5 years of support, making it easier for you to plan your upgrades.

AL2023 offers several improvements over Amazon Linux 2 (AL2). For example, AL2023 takes a security-by-default approach to help improve your security posture with preconfigured security policies, SELinux in permissive mode and IMDSv2 enabled by default, and the availability of kernel live patching. With deterministic upgrades through versioned repositories, you can lock to a specific version of the Amazon Linux package repository, giving you control over how and when you absorb

updates. With this capability, you can adhere to operational best practices more efficiently by ensuring consistency between package versions and updates across your environment. For a full comparison, refer to Comparing Amazon Linux 2 and Amazon Linux 2023.

Amazon Linux 2023 is generally available in all AWS Regions, including the AWS GovCloud (US) and the China Regions.

AWS App Runner

AWS App Runner is a fully managed service that makes it easy for developers to quickly deploy containerized web applications and APIs, at scale and with no prior infrastructure experience required. Start with your source code or a container image. AWS App Runner automatically builds and deploys the web application and load balances traffic with encryption. App Runner also scales up or down automatically to meet your traffic needs. With App Runner, rather than thinking about servers or scaling, you have more time to focus on your applications.

AWS Batch

AWS Batch enables developers, scientists, and engineers to easily and efficiently run hundreds of thousands of batch computing jobs on AWS. AWS Batch dynamically provisions the optimal quantity and type of compute resources (such as CPU or memory-optimized instances) based on the volume and specific resource requirements of the batch jobs submitted. With AWS Batch, there is no need to install and manage batch computing software or server clusters that you use to run your jobs, allowing you to focus on analyzing results and solving problems. AWS Batch plans, schedules, and runs your batch computing workloads across the full range of AWS compute services and features, such as Amazon EC2 and Spot Instances.

AWS Elastic Beanstalk

AWS Elastic Beanstalk is an easy-to-use service for deploying and scaling web applications and services developed with Java, .NET, PHP, Node.js, Python, Ruby, Go, and Docker on familiar servers such as Apache, Nginx, Passenger, and Internet Information Services (IIS).

You can simply upload your code, and AWS Elastic Beanstalk automatically handles the deployment, from capacity provisioning, load balancing, and auto scaling to application health monitoring. At the same time, you retain full control over the AWS resources powering your application and can access the underlying resources at any time.

AWS Fargate

AWS Fargate is a compute engine for Amazon ECS that allows you to run containers without having to manage servers or clusters. With AWS Fargate, you no longer have to provision, configure, and scale clusters of VMs to run containers. This removes the need to choose server types, decide when to scale your clusters, or optimize cluster packing. Fargate removes the need for you to interact with or think about servers or clusters. Fargate lets you focus on designing and building your applications instead of managing the infrastructure that runs them.

Amazon ECS has two modes: Fargate launch type and EC2 launch type. With Fargate launch type, all you have to do is package your application in containers, specify the CPU and memory requirements, define networking and IAM policies, and launch the application. EC2 launch type allows you to have server-level, more granular control over the infrastructure that runs your container applications. With EC2 launch type, you can use Amazon ECS to manage a cluster of servers and schedule placement of containers on the servers. Amazon ECS keeps track of all the CPU, memory and other resources in your cluster, and also finds the best server for a container to run on based on your specified resource requirements.

You are responsible for provisioning, patching, and scaling clusters of servers. You can decide which type of server to use, which applications and how many containers to run in a cluster to optimize utilization, and when you should add or remove servers from a cluster. EC2 launch type gives you more control of your server clusters and provides a broader range of customization options, which might be required to support some specific applications or possible compliance and government requirements.

AWS Lambda

AWS Lambda lets you run code without provisioning or managing servers. You pay only for the compute time you consume—there is no charge when your code is not running. With Lambda, you can run code for virtually any type of application or backend service—all with zero administration. Just upload your code, and Lambda takes care of everything required to run and scale your code with high availability. You can set up your code to automatically run from other AWS services, or you can call it directly from any web or mobile app.

AWS Serverless Application Repository

The AWS Serverless Application Repository enables you to quickly deploy code samples, components, and complete applications for common use cases such as web and mobile backends,

event and data processing, logging, monitoring, Internet of Things (IoT), and more. Each application is packaged with an AWS Serverless Application Model (AWS SAM) template that defines the AWS resources used. Publicly shared applications also include a link to the application's source code. There is no additional charge to use the AWS Serverless Application Repository - you only pay for the AWS resources used in the applications you deploy.

You can also use the AWS Serverless Application Repository to publish your own applications and share them within your team, across your organization, or with the community at large. To share an application you've built, publish it to the AWS Serverless Application Repository.

AWS Outposts

AWS Outposts bring native AWS services, infrastructure, and operating models to virtually any data center, co-location space, or on-premises facility. You can use the same APIs, the same tools, the same hardware, and the same functionality across on-premises and the cloud to deliver a truly consistent hybrid experience. Outposts can be used to support workloads that need to remain on-premises due to low latency or local data processing needs.

AWS Outposts come in two variants:

- VMware Cloud on AWS Outposts allows you to use the same VMware control plane and APIs you use to run your infrastructure.
- AWS-native variant of AWS Outposts allows you to use the same exact APIs and control plane you use to run in the AWS Cloud, but on-premises.

AWS Outposts infrastructure is fully managed, maintained, and supported by AWS to deliver access to the latest AWS services. Getting started is easy, you simply log into the AWS Management Console to order your Outposts servers, choosing from a wide range of compute and storage options. You can order one or more servers, or quarter, half, and full rack units.

AWS Wavelength

AWS Wavelength is an AWS Infrastructure offering optimized for mobile edge computing applications. Wavelength Zones are AWS infrastructure deployments that embed AWS compute and storage services within communications service providers' (CSP) datacenters at the edge of the 5G network, so application traffic from 5G devices can reach application servers running in Wavelength Zones without leaving the telecommunications network. This avoids the latency that would result from application traffic having to traverse multiple hops across the Internet to reach

their destination, enabling customers to take full advantage of the latency and bandwidth benefits offered by modern 5G networks.

VMware Cloud on AWS

VMware Cloud on AWS is an integrated cloud offering jointly developed by AWS and VMware delivering a highly scalable, secure and innovative service that allows organizations to seamlessly migrate and extend their on-premises VMware vSphere-based environments to the AWS Cloud running on next-generation Amazon Elastic Compute Cloud (Amazon EC2) bare metal infrastructure. VMware Cloud on AWS is ideal for enterprise IT infrastructure and operations organizations looking to migrate their on-premises vSphere-based workloads to the public cloud, consolidate and extend their data center capacities, and optimize, simplify and modernize their disaster recovery solutions.

VMware Cloud on AWS is delivered, sold, and supported globally by VMware and its partners with availability in the following AWS Regions: AWS Europe (Stockholm), AWS US East (Northern Virginia), AWS US East (Ohio), AWS US West (Northern California), AWS US West (Oregon), AWS Canada (Central), AWS Europe (Frankfurt), AWS Europe (Ireland), AWS Europe (London), AWS Europe (Paris), AWS Europe (Milan), AWS Asia Pacific (Singapore), AWS Asia Pacific (Sydney), AWS Asia Pacific (Tokyo), AWS Asia Pacific (Mumbai) Region, AWS South America (Sao Paulo), AWS Asia Pacific (Seoul), and AWS GovCloud (US West). With each release, VMware Cloud on AWS availability will expand into additional global regions.

VMware Cloud on AWS brings the broad, diverse and rich innovations of AWS services natively to the enterprise applications running on VMware's compute, storage and network virtualization platforms. This allows organizations to easily and rapidly add new innovations to their enterprise applications by natively integrating AWS infrastructure and platform capabilities such as AWS Lambda, Amazon Simple Queue Service (Amazon SQS), Amazon S3, Elastic Load Balancing, Amazon RDS, Amazon DynamoDB, Amazon Kinesis, and Amazon Redshift, among many others.

With VMware Cloud on AWS, organizations can simplify their Hybrid IT operations by using the same VMware Cloud Foundation technologies including vSphere, vSAN, NSX, and vCenter Server across their on-premises data centers and on the AWS Cloud without having to purchase any new or custom hardware, rewrite applications, or modify their operating models. The service automatically provisions infrastructure and provides full VM compatibility and workload portability between your on-premises environments and the AWS Cloud. With VMware Cloud on AWS, you can use a broad range of AWS services, including compute, databases, analytics, IoT, security, mobile, deployment, application services, and more.

Customer enablement

AWS Managed Services

AWS Managed Services provides ongoing management of your AWS infrastructure so you can focus on your applications. By implementing best practices to maintain your infrastructure, AWS Managed Services helps to reduce your operational overhead and risk. AWS Managed Services automates common activities such as change requests, monitoring, patch management, security, and backup services, and provides full lifecycle services to provision, run, and support your infrastructure. Our rigor and controls help to enforce your corporate and security infrastructure policies, and enables you to develop solutions and applications using your preferred development approach. AWS Managed Services improves agility, reduces cost, and unburdens you from infrastructure operations so that you can direct resources toward differentiating your business.

AWS re:Post Private

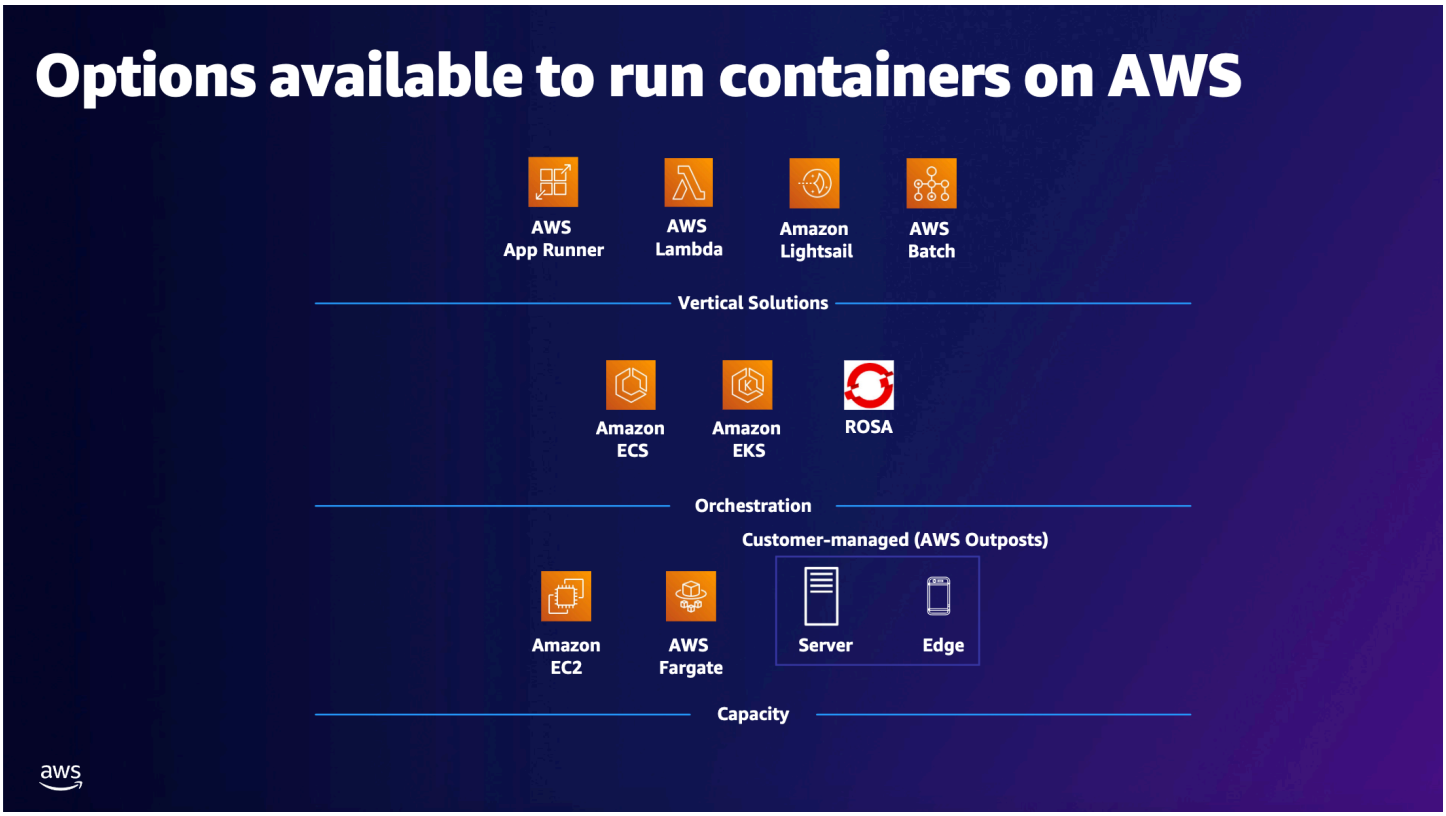
AWS re:Post Private is a private version of AWS re:Post for enterprises with Enterprise Support or Enterprise On-Ramp Support plans. It provides access to knowledge and experts to accelerate cloud adoption and increase developer productivity. With your organization-specific re:Post Private, you can build an organization-specific developer community that drives efficiencies at scale and provides access to valuable knowledge resources. re:Post Private centralizes trusted AWS technical content and offers private discussion forums to improve how your teams collaborate internally and with AWS to remove technical obstacles, accelerate innovation, and scale more efficiently in the cloud.

Containers

AWS offers services that give you a secure place to store and manage your container images, orchestration that manages when and where your containers run, and flexible compute engines to

power your containers. AWS can help manage your containers and their deployments for you, so you don't have to worry about the underlying infrastructure.

Each service is described after the diagram. To help you decide which service best meets your needs, see Choosing an AWS container service or Amazon Lightsail, AWS Elastic Beanstalk, or Amazon EC2?. For general information, see Containers on AWS.



Services

- Amazon Elastic Container Registry
- Amazon Elastic Container Service
- Amazon Elastic Kubernetes Service
- AWS App2Container
- Red Hat OpenShift Service on AWS

Amazon Elastic Container Registry

Amazon Elastic Container Registry (Amazon ECR) is a fully managed Docker container registry that makes it easy for developers to store, manage, and deploy Docker container images. Amazon ECR is integrated with Amazon Elastic Container Service (Amazon ECS), simplifying your development to

production workflow. Amazon ECR eliminates the need to operate your own container repositories or worry about scaling the underlying infrastructure. Amazon ECR hosts your images in a highly available and scalable architecture, allowing you to reliably deploy containers for your applications. Integration with AWS Identity and Access Management (IAM) provides resource-level control of each repository. With Amazon ECR, there are no upfront fees or commitments. You pay only for the amount of data you store in your repositories and data transferred to the internet.

Amazon Elastic Container Service

Amazon Elastic Container Service (Amazon ECS) is a highly scalable, high-performance container orchestration service that supports Docker containers and allows you to easily run and scale containerized applications on AWS. Amazon ECS eliminates the need for you to install and operate your own container orchestration software, manage and scale a cluster of virtual machines (VMs), or schedule containers on those VMs.

With simple API calls, you can launch and stop Docker-enabled applications, query the complete state of your application, and access many familiar features such as IAM roles, security groups, load balancers, Amazon CloudWatch Events, AWS CloudFormation templates, and AWS CloudTrail logs.

Amazon Elastic Kubernetes Service

Amazon Elastic Kubernetes Service (Amazon EKS) makes it easy to deploy, manage, and scale containerized applications using Kubernetes on AWS.

Amazon EKS runs the Kubernetes management infrastructure for you across multiple AWS Availability Zones to eliminate a single point of failure. Amazon EKS is certified Kubernetes conformant so you can use existing tooling and plugins from partners and the Kubernetes community. Applications running on any standard Kubernetes environment are fully compatible and can be easily migrated to Amazon EKS.

AWS App2Container

AWS App2Container (A2C) is a command-line tool for modernizing .NET and Java applications into containerized applications. A2C analyzes and builds an inventory of all applications running in VMs, on-premises or in the cloud. You simply select the application you want to containerize, and A2C packages the application artifact and identified dependencies into container images, configures the network ports, and generates the ECS task and Kubernetes pod definitions. A2C provisions, through AWS CloudFormation, the cloud infrastructure and CI/CD pipelines required to deploy the

containerized .NET or Java application into production. With A2C, you can easily modernize your existing applications and standardize the deployment and operations through containers.

Red Hat OpenShift Service on AWS

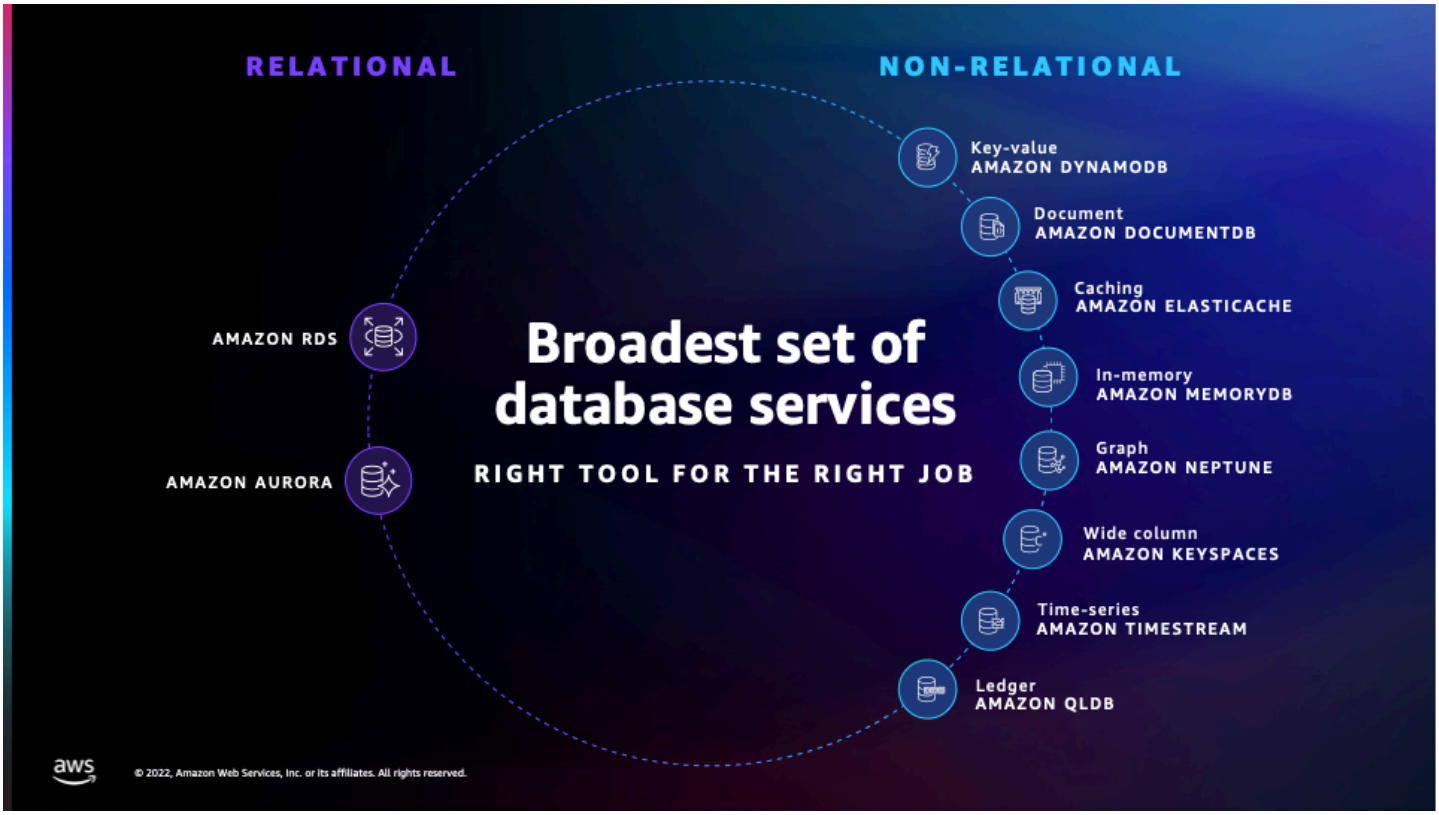
Red Hat OpenShift Service on AWS (ROSA) provides an integrated experience to use OpenShift. If you are already familiar with OpenShift, you can accelerate your application development process by leveraging familiar OpenShift APIs and tools for deployments on AWS. With ROSA, you can use the wide range of AWS compute, database, analytics, machine learning (ML), networking, mobile, and other services to build secure and scalable applications faster. ROSA comes with pay-as-you-go hourly and annual billing, a 99.95% SLA, and joint support from AWS and Red Hat.

ROSA makes it easier for you to focus on deploying applications and accelerating innovation by moving the cluster lifecycle management to Red Hat and AWS. With ROSA, you can run containerized applications with your existing OpenShift workflows and reduce the complexity of management.

Databases

AWS databases offer a high-performance, secure, and reliable foundation to power generative AI solutions and data-driven applications that drive value for your business and customers.

Each service is described after the diagram. To help you decide which service best meets your needs, see Choosing an AWS database service. For general information, see AWS Cloud Databases.



Topics

- Compare AWS database services
- Amazon Aurora
- Amazon DynamoDB
- Amazon ElastiCache
- Amazon Keyspaces (for Apache Cassandra)
- Amazon MemoryDB
- Amazon Neptune
- Amazon Relational Database Service
- Amazon RDS for Db2
- Amazon RDS on VMware
- Amazon Quantum Ledger Database (Amazon QLDB)
- Amazon Timestream
- Amazon DocumentDB (with MongoDB compatibility)
- Amazon Lightsail managed databases

Compare AWS database services

Database	Use cases	AWS services
Relational	Traditional applications, enterprise resource planning (ERP), customer relations hip management (CRM), e-commerce	• **Amazon Aurora** — Designed for unparalle led high performance and availability at global scale with full MySQL and PostgreSQL compatibility • **Amazon RDS** — Set up, operate, and scale a relational database in the cloud with just a few clicks • **Amazon Redshift** — Accelerate your time to insights with fast, easy, and secure cloud data warehousing at scale
Key-value	High-traffic web applications, e-commerce systems, gaming applications	• **Amazon DynamoDB** — Fast, flexible NoSQL database service for single-digit millisecond performan ce at any scale
In-memory	Caching, session managemen t, gaming leaderboards, geospatial applications	• **Amazon ElastiCache** — Unlock microsecond latency and scale with in-memory caching • **Amazon MemoryDB** — Redis-compatible, durable, in-memory database service for ultra-fast performance

Database	Use cases	AWS services
Document	Content management, catalogs, user profiles	• **Amazon DocumentDB (with MongoDB compatibility)** — Scale JSON workloads with ease using a fully managed document database service
Wide column	High-scale industrial apps for equipment maintenance, fleet management, and route optimization	• **Amazon Keyspaces** — A scalable, highly available, and managed Apache Cassandra–compatible database service
Graph	Fraud detection, social networking, recommendation engines	• **Amazon Neptune** — Build and run graph applications with highly connected datasets
Time series	Internet of Things (IoT) applications, DevOps, industrial telemetry	• **Amazon Timestream** — Fast, scalable, serverless time series database
Ledger	Systems of record, supply chain, registrations, banking transactions	• **Amazon Ledger Database Service (QLDB)** — Maintain an immutable, cryptographically verifiable log of data changes

Amazon Aurora

Amazon Aurora is a MySQL and PostgreSQL compatible relational database engine that combines the speed and availability of high-end commercial databases with the simplicity and cost-effectiveness of open source databases.

Amazon Aurora is up to five times faster than standard MySQL databases and three times faster than standard PostgreSQL databases. It provides the security, availability, and reliability of

commercial databases at 1/10th the cost. Amazon Aurora is fully managed by Amazon Relational Database Service (Amazon RDS), which automates time-consuming administration tasks such as hardware provisioning, database setup, patching, and backups.

Amazon Aurora features a distributed, fault-tolerant, self-healing storage system that auto-scales up to 128TB per database instance. It delivers high performance and availability with up to 15 low-latency read replicas, point-in-time recovery, continuous backup to Amazon S3, and replication across three Availability Zones (AZs).

Amazon Aurora I/O-Optimized is a cluster configuration that offers improved price performance and predictable pricing for customers with I/O-intensive applications, such as e-commerce applications, payment processing systems, and financial applications. Aurora-Optimized offers improved performance, increasing throughput and reducing latency to support your most demanding workloads, with up to 40 percent cost savings when your I/O spend exceeds 25 percent of your current Aurora database spend.

Amazon Aurora MySQL zero-ETL integration with Amazon Redshift, now available in public preview, enables near real-time analytics and machine learning of data stored in Aurora MySQL-Compatible Edition. Transactional data written to Aurora is available to you in Amazon Redshift within seconds, without building and maintaining complex data pipelines.

Amazon DynamoDB

Amazon DynamoDB is a key-value and document database that delivers single-digit millisecond performance at any scale. It's a fully managed, multi-Region database with built-in security, backup and restore, and in-memory caching for internet-scale applications. DynamoDB can handle more than 10 trillion requests per day and support peaks of more than 20 million requests per second.

Many of the world's fastest growing businesses such as Lyft, Airbnb, and Redfin, as well as enterprises such as Samsung, Toyota, and Capital One, depend on the scale and performance of DynamoDB to support their mission-critical workloads.

Hundreds of thousands of AWS customers have chosen DynamoDB as their key-value and document database for mobile, web, gaming, ad tech, Internet of Things (IoT), and other applications that need low-latency data access at any scale. Create a new table for your application and let DynamoDB handle the rest.

Amazon ElastiCache

Amazon ElastiCache is a web service that makes it easy to deploy, operate, and scale an in-memory cache in the cloud. The service improves the performance of web applications by allowing you to retrieve information from fast, managed, in-memory caches, instead of relying entirely on slower disk-based databases.

ElastiCache supports two open-source in-memory caching engines:

- Redis — a fast, open-source, in-memory key-value data store for use as a database, cache, message broker, and queue. Amazon ElastiCache (Redis OSS) is a Redis-compatible in-memory service that delivers the ease-of-use and power of Redis along with the availability, reliability, and performance suitable for the most demanding applications. Both single-node and up to 15-shard clusters are available, enabling scalability to up to 3.55 TiB of in-memory data. Amazon ElastiCache (Redis OSS) is fully managed, scalable, and secure. This makes it an ideal candidate to power high-performance use cases such as web, mobile apps, gaming, ad-tech, and IoT.
- Memcached — a widely adopted memory object caching system. Amazon ElastiCache (Memcached) is protocol compliant with Memcached, so popular tools that you use today with existing Memcached environments will work seamlessly with the service.

Amazon ElastiCache Serverless is a serverless option for Amazon ElastiCache that simplifies cache management and instantly scales to support the most demanding applications. With ElastiCache Serverless, you can create a highly available and scalable cache in less than a minute, eliminating the need to plan for, provision, and manage cache cluster capacity. ElastiCache Serverless automatically stores data redundantly across multiple Availability Zones (AZs) and provides a 99.99% availability Service Level Agreement (SLA). With ElastiCache Serverless, you pay for data stored and compute consumed by your workload, with no upfront commitments or additional costs.

Amazon Keyspaces (for Apache Cassandra)

Amazon Keyspaces (for Apache Cassandra) is a scalable, highly available, and managed Apache Cassandra–compatible database service. With Amazon Keyspaces, you can run your Cassandra workloads on AWS using the same Cassandra application code and developer tools that you use today. You don't have to provision, patch, or manage servers, and you don't have to install, maintain, or operate software. Amazon Keyspaces is serverless, so you pay for only the resources you use and the service can automatically scale tables up and down in response to application

traffic. You can build applications that serve thousands of requests per second with virtually unlimited throughput and storage. Data is encrypted by default and Amazon Keyspaces enables you to back up your table data continuously using point-in-time recovery. Amazon Keyspaces gives you the performance, elasticity, and enterprise features you need to operate business-critical Cassandra workloads at scale.

Amazon MemoryDB

Amazon MemoryDB is a Redis-compatible, durable, in-memory database service that delivers ultra-fast performance. It is purpose-built for modern applications with microservices architectures.

MemoryDB is compatible with Redis, a popular open source data store, enabling customers to quickly build applications using the same flexible and friendly Redis data structures, APIs, and commands that they already use today. With MemoryDB, all of your data is stored in memory, which enables you to achieve microsecond read and single-digit millisecond write latency and high throughput. MemoryDB also stores data durably across multiple Availability Zones using a distributed transactional log to allow fast failover, database recovery, and node restarts. Delivering both in-memory performance and Multi-AZ durability, MemoryDB can be used as a high-performance primary database for your microservices applications eliminating the need to separately manage both a cache and durable database.

Amazon Neptune

Amazon Neptune is a fast, reliable, fully managed graph database service that makes it easy to build and run applications that work with highly connected datasets. The core of Amazon Neptune is a purpose-built, high-performance graph database engine optimized for storing billions of relationships and querying the graph with milliseconds latency. Amazon Neptune supports popular graph models Property Graph and W3C's RDF, and their respective query languages Apache TinkerPop Gremlin and SPARQL, allowing you to easily build queries that efficiently navigate highly connected datasets. Neptune powers graph use cases such as recommendation engines, fraud detection, knowledge graphs, drug discovery, and network security.

Amazon Neptune is highly available, with read replicas, point-in-time recovery, continuous backup to Amazon S3, and replication across Availability Zones. Neptune is secure with support for encryption at rest. Neptune is fully managed, so you no longer need to worry about database management tasks such as hardware provisioning, software patching, setup, configuration, or backups.

Amazon Neptune Analytics is an analytics database engine for quickly analyzing large volumes of graph data to get insights and find trends from data stored in Amazon S3 buckets or a Neptune database. Neptune Analytics uses built-in algorithms, vector search, and in-memory computing to run queries on data with tens of billions of relationships in seconds.

Amazon Relational Database Service

Amazon Relational Database Service (Amazon RDS) makes it easy to set up, operate, and scale a relational database in the cloud. It provides cost-efficient and resizable capacity while automating time-consuming administration tasks such as hardware provisioning, database setup, patching and backups. It frees you to focus on your applications so you can give them the fast performance, high availability, security and compatibility they need.

Amazon RDS is available on several database instance types—optimized for memory, performance or I/O—and provides you with six familiar database engines to choose from, including MySQL, MariaDB, PostgreSQL, Oracle Database, Microsoft SQL Server, and Amazon RDS on AWS Outposts. You can use the AWS Database Migration Service to easily migrate or replicate your existing databases to Amazon RDS.

Amazon RDS for Db2

Amazon RDS for Db2 makes it easy to set up, operate, and scale Db2 deployments in the cloud. Amazon RDS automates time-consuming database administration tasks, such as provisioning, backups, software patching, monitoring, and more, to free up time to innovate and drive business value. It also offers high availability with Multi-AZ deployment, disaster recovery solutions with cross-Region backups, and security features to support your business-critical workloads. In addition, you can integrate with other IBM and AWS services to gain new insights and scale your analytics workloads.

Amazon RDS on VMware

Amazon Relational Database Service (Amazon RDS) on VMware lets you deploy managed databases in on-premises VMware environments using the Amazon RDS technology enjoyed by hundreds of thousands of AWS customers. Amazon RDS provides cost-efficient and resizable capacity while automating time-consuming administration tasks including hardware provisioning, database setup, patching, and backups, freeing you to focus on your applications. Amazon RDS on VMware brings these same benefits to your on-premises deployments, making it easy to set up, operate, and scale databases in VMware vSphere private data centers, or to migrate them to AWS.

Amazon RDS on VMware allows you to utilize the same simple interface for managing databases in on-premises VMware environments as you would use in AWS. You can easily replicate Amazon RDS on VMware databases to Amazon RDS instances in AWS, enabling low-cost hybrid deployments for disaster recovery, read replica bursting, and optional long-term backup retention in Amazon Simple Storage Service (Amazon S3).

Amazon Quantum Ledger Database (Amazon QLDB)

Amazon QLDB is a fully managed ledger database that provides a transparent, immutable, and cryptographically verifiable transaction log owned by a central trusted authority. Amazon QLDB tracks each and every application data change and maintains a complete and verifiable history of changes over time.

Ledgers are typically used to record a history of economic and financial activity in an organization. Many organizations build applications with ledger-like functionality because they want to maintain an accurate history of their applications' data, for example, tracking the history of credits and debits in banking transactions, verifying the data lineage of an insurance claim, or tracing movement of an item in a supply chain network. Ledger applications are often implemented using custom audit tables or audit trails created in relational databases. However, building audit functionality with relational databases is time-consuming and prone to human error. It requires custom development, and since relational databases are not inherently immutable, any unintended changes to the data are hard to track and verify. Alternatively, blockchain frameworks, such as Hyperledger Fabric and Ethereum, can also be used as a ledger. However, this adds complexity as you need to set-up an entire blockchain network with multiple nodes, manage its infrastructure, and require the nodes to validate each transaction before it can be added to the ledger.

Amazon QLDB is a new class of database that eliminates the need to engage in the complex development effort of building your own ledger-like applications. With QLDB, your data's change history is immutable – it cannot be altered or deleted – and using cryptography, you can easily verify that there have been no unintended modifications to your application's data. QLDB uses an immutable transactional log, known as a journal, that tracks each application data change and maintains a complete and verifiable history of changes over time. QLDB is easy to use because it provides developers with a familiar SQL-like API, a flexible document data model, and full support for transactions. QLDB is also serverless, so it automatically scales to support the demands of your application. There are no servers to manage and no read or write limits to configure. With QLDB, you only pay for what you use.

Amazon Timestream

Amazon Timestream is a fast, scalable, fully managed time series database service for IoT and operational applications that makes it easy to store and analyze trillions of events per day at 1/10th the cost of relational databases. Driven by the rise of IoT devices, IT systems, and smart industrial machines, time series data — data that measures how things change over time — is one of the fastest growing data types. Time-series data has specific characteristics such as typically arriving in time order form, data is append-only, and queries are always over a time interval. While relational databases can store this data, they are inefficient at processing this data as they lack optimizations such as storing and retrieving data by time intervals.

Timestream is a purpose-built time series database that efficiently stores and processes this data by time intervals. With Timestream, you can easily store and analyze log data for DevOps, sensor data for IoT applications, and industrial telemetry data for equipment maintenance. As your data grows over time, the Timestream adaptive query processing engine understands its location and format, making your data simpler and faster to analyze. Timestream also automates rollups, retention, tiering, and compression of data, so you can manage your data at the lowest possible cost. Timestream is serverless, so there are no servers to manage. It manages time-consuming tasks such as server provisioning, software patching, setup, configuration, or data retention and tiering, freeing you to focus on building your applications.

Amazon DocumentDB (with MongoDB compatibility)

Amazon DocumentDB (with MongoDB compatibility) is a fast, scalable, highly available, and fully managed document database service that supports MongoDB workloads.

Amazon DocumentDB is designed from the ground-up to give you the performance, scalability, and availability you need when operating mission-critical MongoDB workloads at scale. Amazon DocumentDB implements the Apache 2.0 open source MongoDB 3.6 and 4.0 APIs by emulating the responses that a MongoDB client expects from a MongoDB server, allowing you to use your existing MongoDB drivers and tools with Amazon DocumentDB (with MongoDB compatibility).

Amazon Lightsail managed databases

Amazon Lightsail managed databases are separate from compute workloads, so you can build applications and websites on Lightsail instances without interruption. Lightsail supports MySQL and PostgreSQL databases , and you can configure them for standard availability for regular workloads or high availability for critical workloads. Lightsail-managed databases bundle the

underlying compute, SSD-based storage, and data transfer bandwidth into a fixed monthly price. You can manage your Lightsail-managed database by using the Lightsail console, the AWS Command Line Interface (AWS CLI), the Lightsail API, or an AWS SDK.

Developer tools

Topics

- AWS Infrastructure Composer
- AWS Cloud9
- AWS CloudShell
- AWS CodeArtifact
- AWS CodeBuild
- Amazon CodeCatalyst
- AWS CodeCommit
- AWS CodeDeploy
- AWS CodePipeline
- Amazon Corretto
- AWS Fault Injection Service
- Amazon Q Developer
- AWS X-Ray

AWS Infrastructure Composer

AWS Infrastructure Composer helps you visually compose and configure serverless applications from AWS services backed by deployment-ready infrastructure as code (IaC). Infrastructure Composer helps you drag and drop serverless resources onto a visual, browser-based canvas. You can connect them to quickly create your serverless application architecture. The canvas also supports grouping of resources into larger architectural components to simplify editing and configuration. AWS Infrastructure Composer can generate deployment-ready configuration with default settings based on the services that make up your application architecture. Infrastructure

Composer supports generating both AWS CloudFormation and AWS Serverless Application Model (SAM) artifacts.

AWS Cloud9

AWS Cloud9 is a cloud-based integrated development environment (IDE) that lets you write, run, and debug your code with just a browser. It includes a code editor, debugger, and terminal. AWS Cloud9 comes prepackaged with essential tools for popular programming languages, including JavaScript, Python, PHP, and more, so you don't need to install files or configure your development machine to start new projects. Since your AWS Cloud9 IDE is cloud-based, you can work on your projects from your office, home, or anywhere using an internet-connected machine. AWS Cloud9 also provides a seamless experience for developing serverless applications enabling you to easily define resources, debug, and switch between local and remote running of serverless applications. With AWS Cloud9, you can quickly share your development environment with your team, enabling you to pair program and track each other's inputs in real time.

AWS CloudShell

AWS CloudShell is a browser-based shell that makes it easy to securely manage, explore, and interact with your AWS resources. CloudShell is pre-authenticated with your console credentials. Common development and operations tools are pre-installed, so no local installation or configuration is required. With CloudShell, you can quickly run scripts with the AWS Command Line Interface (AWS CLI), experiment with AWS service APIs using the AWS SDKs, or use a range of other tools to be productive. You can use CloudShell right from your browser and at no additional cost.

AWS CodeArtifact

AWS CodeArtifact is a fully managed artifact repository service that makes it easy for organizations of any size to securely store, publish, and share software packages used in their software development process. CodeArtifact can be configured to automatically fetch software packages and dependencies from public artifact repositories so developers have access to the latest versions. CodeArtifact works with commonly used package managers and build tools such as Apache Maven, Gradle, `npm`, `yarn`, `twine`, `pip`, and NuGet making it easy to integrate into existing development workflows.

AWS CodeBuild

AWS CodeBuild is a fully managed build service that compiles source code, runs tests, and produces software packages that are ready to deploy. With CodeBuild, you don't need to provision, manage,

and scale your own build servers. CodeBuild scales continuously and processes multiple builds concurrently, so your builds are not left waiting in a queue. You can get started quickly by using prepackaged build environments, or you can create custom build environments that use your own build tools.

Amazon CodeCatalyst

Amazon CodeCatalyst is an integrated service for software development teams adopting continuous integration/continuous deployment (CI/CD) practices into their software development process. CodeCatalyst is fully managed by AWS and puts the tools you need all in one place. You can plan work, collaborate on code, and build, test, and deploy applications. You can also integrate AWS resources with your projects by connecting your AWS accounts to your CodeCatalyst space. By managing all of the stages and aspects of your application lifecycle in one tool, you can deliver software quickly and confidently.

AWS CodeCommit

AWS CodeCommit is a fully managed source control service that makes it easy for companies to host secure and highly scalable private Git repositories. AWS CodeCommit eliminates the need to operate your own source control system or worry about scaling its infrastructure. You can use AWS CodeCommit to securely store anything from source code to binaries, and it works seamlessly with your existing Git tools.

AWS CodeDeploy

AWS CodeDeploy is a service that automates code deployments to any instance, including EC2 instances and instances running on premises. CodeDeploy makes it easier for you to rapidly release new features, helps you avoid downtime during application deployment, and handles the complexity of updating your applications. You can use CodeDeploy to automate software deployments, eliminating the need for error-prone manual operations. The service scales with your infrastructure so you can easily deploy to one instance or thousands.

AWS CodePipeline

AWS CodePipeline is a fully managed continuous delivery service that helps you automate your release pipelines for fast and reliable application and infrastructure updates. CodePipeline automates the build, test, and deploy phases of your release process every time there is a code change, based on the release model you define. This enables you to rapidly and reliably deliver features and updates. You can easily integrate CodePipeline with third-party services such as

GitHub or with your own custom plugin. With AWS CodePipeline, you only pay for what you use. There are no upfront fees or long-term commitments.

Amazon Corretto

Amazon Corretto is a no-cost, multiplatform, production-ready distribution of the Open Java Development Kit (OpenJDK). Corretto comes with long-term support that will include performance enhancements and security fixes. Amazon runs Corretto internally on thousands of production services, and Corretto is certified as compatible with the Java SE standard. With Corretto, you can develop and run Java applications on popular operating systems, including Amazon Linux 2, Windows, and macOS.

AWS Fault Injection Service

AWS Fault Injection Service is a fully managed service for running fault injection experiments on AWS that makes it easier to improve an application's performance, observability, and resiliency. Fault injection experiments are used in chaos engineering, which is the practice of stressing an application in testing or production environments by creating disruptive events, such as sudden increase in CPU or memory consumption, observing how the system responds, and implementing improvements. Fault injection experiment helps teams create the real-world conditions needed to uncover the hidden bugs, monitoring blind spots, and performance bottlenecks that are difficult to find in distributed systems.

AWS Fault Injection Service simplifies the process of setting up and running controlled fault injection experiments across a range of AWS services so teams can build confidence in their application behavior. With Fault Injection Simulator, teams can quickly set up experiments using pre-built templates that generate the desired disruptions. AWS Fault Injection Service provides the controls and guardrails that teams need to run experiments in production, such as automatically rolling back or stopping the experiment if specific conditions are met. With a few clicks in the console, teams can run complex scenarios with common distributed system failures happening in parallel or building sequentially over time, enabling them to create the real world conditions necessary to find hidden weaknesses.

Amazon Q Developer

Amazon Q Developer (formerly Amazon CodeWhisperer) assists developers and IT professionals with their tasks—from coding, testing, and upgrading applications, to diagnosing errors, performing security scanning and fixes, and optimizing AWS resources. Amazon Q has advanced,

multistep planning and reasoning capabilities that can transform existing code (for example, perform Java version upgrades) and implement new features generated from developer requests.

AWS X-Ray

AWS X-Ray helps developers analyze and debug distributed applications in production or under development, such as those built using a microservices architecture. X-Ray, you can understand how your application and its underlying services are performing so you can identify and troubleshoot the root cause of performance issues and errors. X-Ray provides an end-to-end view of requests as they travel through your application, and shows a map of your application's underlying components. You can use X-Ray to analyze both applications in development and in production, from simple three-tier applications to complex microservices applications consisting of thousands of services.

End user computing

Amazon AppStream 2.0

Amazon AppStream 2.0 is a fully managed application streaming service. You centrally manage your desktop applications on AppStream 2.0 and securely deliver them to any computer. You can easily scale to any number of users across the globe without acquiring, provisioning, and operating hardware or infrastructure. AppStream 2.0 is built on AWS, so you benefit from a data center and network architecture designed for the most security-sensitive organizations. Each user has a fluid and responsive experience with your applications, including GPU-intensive 3D design and engineering ones, because your applications run on virtual machines (VMs) optimized for specific use cases and each streaming session automatically adjusts to network conditions.

Enterprises can use AppStream 2.0 to simplify application delivery and complete their migration to the cloud. Educational institutions can provide every student access to the applications they need for class on any computer. Software vendors can use AppStream 2.0 to deliver trials, demos, and training for their applications with no downloads or installations. They can also develop a full software-as-a-service (SaaS) solution without rewriting their application.

Amazon WorkSpaces

Amazon WorkSpaces is a fully managed, secure cloud desktop service. You can use WorkSpaces to provision either Windows or Linux desktops in just a few minutes and quickly scale to provide thousands of desktops to workers across the globe. You can pay either monthly or hourly, just

for the WorkSpaces you launch, which helps you save money when compared to traditional desktops and on-premises VDI solutions. WorkSpaces helps you eliminate the complexity in managing hardware inventory, OS versions and patches, and Virtual Desktop Infrastructure (VDI), which helps simplify your desktop delivery strategy. With WorkSpaces, your users get a fast, responsive desktop of their choice that they can access anywhere, anytime, from any supported device.

Amazon WorkSpaces Core

Amazon WorkSpaces Core provides cloud-based, fully managed virtual desktop infrastructure (VDI) accessible to third-party VDI management solutions.

- Simplify VDI migration and combine your current VDI software with the security and reliability of AWS.
- Maximize productivity and business continuity with a financially backed 99.9% uptime SLA.
- Scale on demand with fixed-rate hourly billing, no overprovisioning, and no upfront costs.
- Improve user experience and performance with virtual desktops located closer to your global workforce.

Amazon WorkSpaces Thin Client

Amazon WorkSpaces Thin Client is a cost-effective thin client device that is built to work with AWS End User Computing (EUC) virtual desktops to provide users with a complete cloud desktop solution. WorkSpaces Thin Client is a compact device designed to connect two monitors and multiple USB devices such as a keyboard, mouse, headset, and webcam. To maximize endpoint security, WorkSpaces Thin Client devices do not allow local data storage or installation of unapproved applications. The WorkSpaces Thin Client device ships directly to end users or to your company's locations preloaded with device management software.

Amazon Workspaces Web

Amazon WorkSpaces Web is a low-cost, fully managed workspace built specifically to facilitate secure access to internal websites and software-as-a-service (SaaS) applications from existing web browsers, without the administrative burden of appliances or specialized client software. Protect internal content with enterprise controls, while providing access to all the web-based productivity tools users need from any browser.

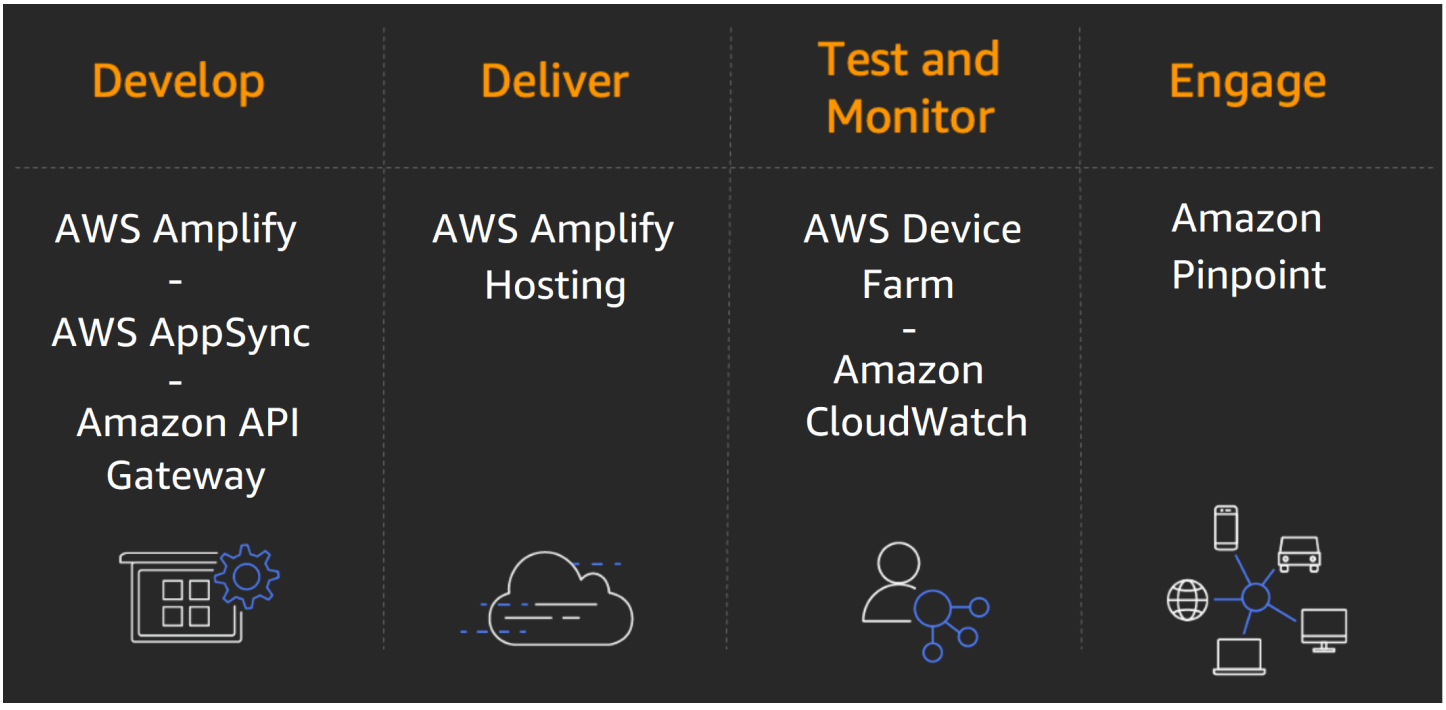
WorkSpaces Web makes it easy for customers to safely provide their employees with access to internal websites and SaaS web applications without the administrative burden of appliances

or specialized client software. WorkSpaces Web provides simple policy tools tailored for user interactions, while offloading common tasks like capacity management, scaling, and maintaining browser images.

Frontend web and mobile services

AWS offers a broad set of tools and services to support development workflows for native iOS, Android, React Native, and JavaScript developers. Discover how easy it is to develop, deploy, and operate your app, even if you are new to AWS.

Each service is described after the diagram. To help you decide which service best meets your needs, see Choosing AWS frontend web and mobile services. For general information, see Frontend Web and Mobile on AWS.



Services

- AWS Amplify
- AWS AppSync
- AWS Device Farm

- Amazon Location Service

AWS Amplify

AWS Amplify makes it easy to create, configure, and implement scalable mobile applications powered by AWS. Amplify seamlessly provisions and manages your mobile backend and provides a simple framework to easily integrate your backend with your iOS, Android, Web, and React Native frontends. Amplify also automates the application release process of both your front-end and back-end allowing you to deliver features faster.

Mobile applications require cloud services for actions that can't be done directly on the device, such as offline data synchronization, storage, or data sharing across multiple users. You often have to configure, set up, and manage multiple services to power the backend. You also have to integrate each of those services into your application by writing multiple lines of code. However, as the number of application features grow, your code and release process becomes more complex and managing the backend requires more time.

Amplify provisions and manages backends for your mobile applications. You just select the capabilities you need such as authentication, analytics, or offline data sync, and Amplify will automatically provision and manage the AWS service that powers each of the capabilities. You can then integrate those capabilities into your application through the Amplify libraries and UI components.

AWS AppSync

AWS AppSync is a serverless back-end for mobile, web, and enterprise applications.

AWS AppSync makes it easy to build data driven mobile and web applications by handling securely all the application data management tasks such as online and offline data access, data synchronization, and data manipulation across multiple data sources. AWS AppSync uses GraphQL, an API query language designed to build client applications by providing an intuitive and flexible syntax for describing their data requirement.

AWS Device Farm

AWS Device Farm is an app testing service that lets you test and interact with your Android, iOS, and web apps on many devices at once, or reproduce issues on a device in real time. View video, screenshots, logs, and performance data to pinpoint and fix issues before shipping your app.

Amazon Location Service

Amazon Location Service makes it easy for developers to add location functionality to applications without compromising data security and user privacy.

Location data is a vital ingredient in today's applications, enabling capabilities ranging from asset tracking to location-based marketing. However, developers face significant barriers when integrating location functionality into their applications. This includes cost, privacy and security compromises, and tedious and slow integration work.

Amazon Location Service provides affordable data, tracking and geofencing capabilities, and native integrations with AWS services, so you can create sophisticated location-enabled applications quickly, without the high cost of custom development. You retain control of your location data with Amazon Location, and you can combine proprietary data with data from the service. Amazon Location provides cost-effective location-based services (LBS) using high-quality data from global, trusted providers Esri and HERE.

Game tech

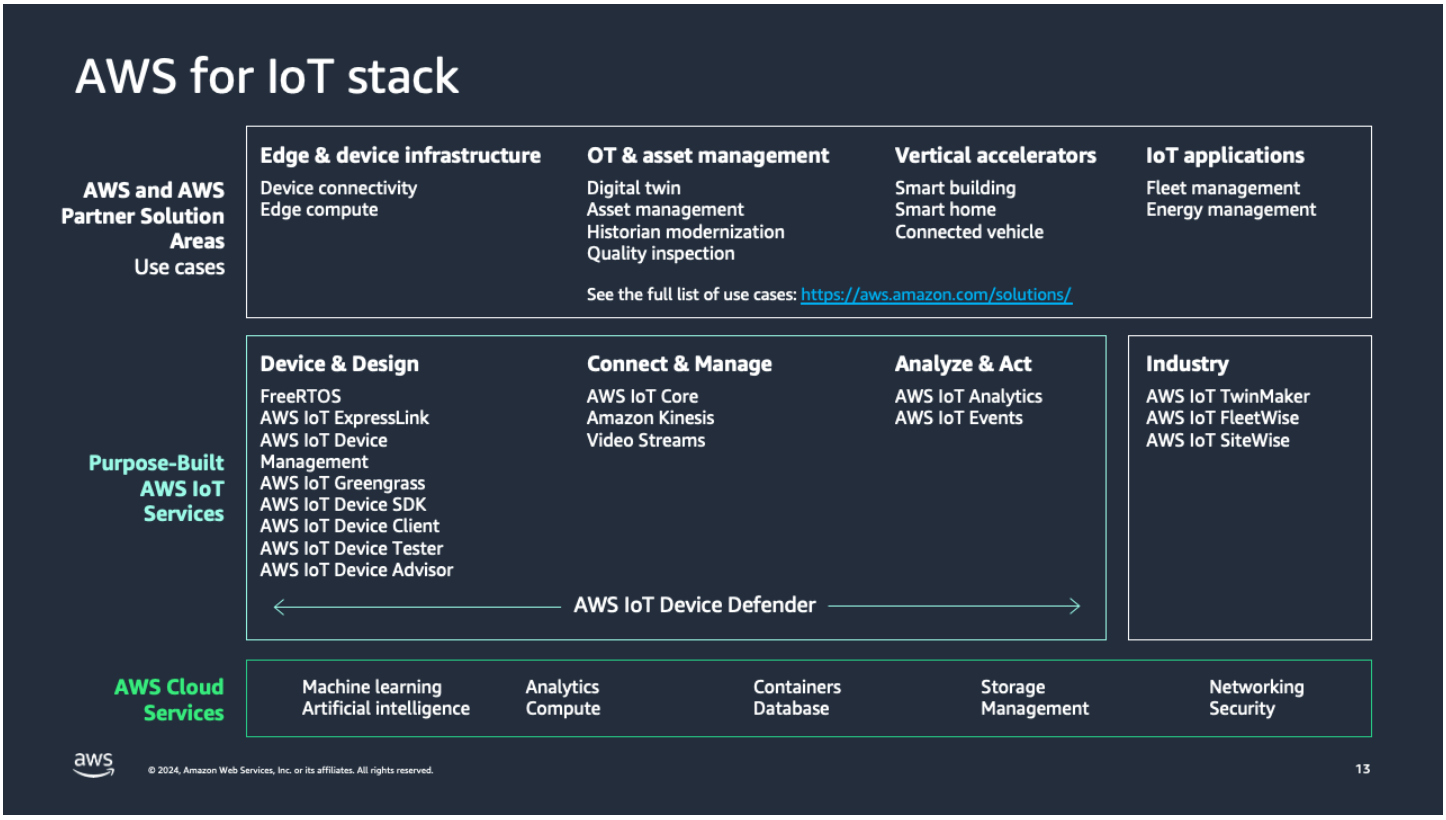
Amazon GameLift Servers

> Amazon GameLift Servers is a managed service for deploying, operating, and scaling dedicated game servers for session-based multiplayer games. Amazon GameLift Servers makes it easy to manage server infrastructure, scale capacity to lower latency and cost, match players into available game sessions, and defend from distributed denial-of-service (DDoS) attacks. You pay for the compute resources and bandwidth your games actually use, without monthly or annual contracts.

Internet of Things (IoT)

AWS offers Internet of Things (IoT) services and solutions to connect and manage billions of devices. Collect, store, and analyze IoT data for industrial, consumer, commercial, and automotive workloads.

Each service is described after the diagram. To help you decide which service best meets your needs, see Choosing an AWS IoT service. For general information, see AWS IoT.



Services

- AWS IoT Analytics
- AWS IoT Button
- AWS IoT Core
- AWS IoT Device Defender
- AWS IoT Device Management
- AWS IoT Events
- AWS IoT ExpressLink
- AWS IoT FleetWise
- AWS IoT Greengrass
- AWS IoT SiteWise

- AWS IoT TwinMaker
- AWS Partner Device Catalog
- FreeRTOS

AWS IoT Analytics

AWS IoT Analytics is a fully managed service that makes it easy to run and operationalize sophisticated analytics on massive volumes of IoT data without having to worry about the cost and complexity typically required to build an IoT analytics platform. It is the easiest way to run analytics on IoT data and get insights to make better and more accurate decisions for IoT applications and machine learning use cases.

IoT data is highly unstructured which makes it difficult to analyze with traditional analytics and business intelligence tools that are designed to process structured data. IoT data comes from devices that often record fairly noisy processes (such as temperature, motion, or sound). The data from these devices can frequently have significant gaps, corrupted messages, and false readings that must be cleaned up before analysis can occur. Also, IoT data is often only meaningful in the context of additional, third party data inputs. For example, to help farmers determine when to water their crops, vineyard irrigation systems often enrich moisture sensor data with rainfall data from the vineyard, allowing for more efficient water usage while maximizing harvest yield.

AWS IoT Analytics automates each of the difficult steps that are required to analyze data from IoT devices. AWS IoT Analytics filters, transforms, and enriches IoT data before storing it in a time series data store for analysis. You can setup the service to collect only the data you need from your devices, apply mathematical transforms to process the data, and enrich the data with device-specific metadata such as device type and location before storing the processed data. Then, you can analyze your data by running ad hoc or scheduled queries using the built-in SQL query engine, or perform more complex analytics and machine learning inference. AWS IoT Analytics makes it easy to get started with machine learning by including pre-built models for common IoT use cases.

You can also use your own custom analysis, packaged in a container, to run AWS IoT Analytics. AWS IoT Analytics automates the running of your custom analyses created in Jupyter Notebook or your own tools (such as Matlab, Octave, and so on) to be run on your schedule.

AWS IoT Analytics is a fully managed service that operationalizes analyses and scales automatically to support up to petabytes of IoT data. With AWS IoT Analytics, you can analyze data from millions of devices and build fast, responsive IoT applications without managing hardware or infrastructure.

AWS IoT Button

The AWS IoT Button is a programmable button based on the Amazon Dash Button hardware. This simple Wi-Fi device is easy to configure, and it's designed for developers to get started with AWS IoT Core, AWS Lambda, Amazon DynamoDB, Amazon SNS, and many other Amazon Web Services without writing device-specific code.

You can code the button's logic in the cloud to configure button clicks to count or track items, call or alert someone, start or stop something, order services, or even provide feedback. For example, you can click the button to unlock or start a car, open your garage door, call a cab, call your spouse or a customer service representative, track the use of common household chores, medications or products, or remotely control your home appliances.

The button can be used as a remote control for Netflix, a switch for your Philips Hue light bulb, a check-in/check-out device for Airbnb guests, or a way to order your favorite pizza for delivery. You can integrate it with third-party APIs such as Twitter, Facebook, Twilio, Slack or even your own company's applications. Connect it to things we haven't even thought of yet.

AWS IoT Core

AWS IoT Core is a managed cloud service that lets connected devices easily and securely interact with cloud applications and other devices. AWS IoT Core can support billions of devices and trillions of messages, and can process and route those messages to AWS endpoints and to other devices reliably and securely. With AWS IoT Core, your applications can keep track of and communicate with all your devices, all the time, even when they aren't connected.

AWS IoT Core makes it easy to use AWS services such as AWS Lambda, Amazon Kinesis, Amazon S3, Amazon SageMaker AI, Amazon DynamoDB, Amazon CloudWatch, AWS CloudTrail, and Amazon QuickSight to build Internet of IoT applications that gather, process, analyze and act on data generated by connected devices, without having to manage any infrastructure.

AWS IoT Device Defender

AWS IoT Device Defender is a fully managed service that helps you secure your fleet of IoT devices. AWS IoT Device Defender continuously audits your IoT configurations to make sure that they aren't deviating from security best practices. A configuration is a set of technical controls you set to help keep information secure when devices are communicating with each other and the cloud. AWS IoT Device Defender makes it easy to maintain and enforce IoT configurations, such as ensuring device identity, authenticating and authorizing devices, and encrypting device data. AWS IoT Device

Defender continuously audits the IoT configurations on your devices against a set of predefined security best practices. AWS IoT Device Defender sends an alert if there are any gaps in your IoT configuration that might create a security risk, such as identity certificates being shared across multiple devices or a device with a revoked identity certificate trying to connect to AWS IoT Core.

AWS IoT Device Defender also lets you continuously monitor security metrics from devices and AWS IoT Core for deviations from what you have defined as appropriate behavior for each device. If something doesn't look right, AWS IoT Device Defender sends out an alert so you can take action to remediate the issue. For example, traffic spikes in outbound traffic might indicate that a device is participating in a DDoS attack. AWS IoT Greengrass and FreeRTOS automatically integrate with AWS IoT Device Defender to provide security metrics from the devices for evaluation.

AWS IoT Device Defender can send alerts to the AWS IoT Console, Amazon CloudWatch, and Amazon SNS. If you determine that you need to take an action based on an alert, you can use AWS IoT Device Management to take mitigating actions such as pushing security fixes.

AWS IoT Device Management

As many IoT deployments consist of hundreds of thousands to millions of devices, it is essential to track, monitor, and manage connected device fleets. You need to ensure your IoT devices work properly and securely after they have been deployed. You also need to secure access to your devices, monitor health, detect and remotely troubleshoot problems, and manage software and firmware updates.

AWS IoT Device Management makes it easy to securely onboard, organize, monitor, and remotely manage IoT devices at scale. With AWS IoT Device Management, you can register your connected devices individually or in bulk, and easily manage permissions so that devices remain secure. You can also organize your devices, monitor and troubleshoot device functionality, query the state of any IoT device in your fleet, and send firmware updates over-the-air (OTA). AWS IoT Device Management is agnostic to device type and OS, so you can manage devices from constrained microcontrollers to connected cars all with the same service. AWS IoT Device Management allows you to scale your fleets and reduce the cost and effort of managing large and diverse IoT device deployments.

AWS IoT Events

AWS IoT Events is a fully managed IoT service that makes it easy to detect and respond to events from IoT sensors and applications. Events are patterns of data identifying more complicated circumstances than expected, such as changes in equipment when a belt is stuck or connected

motion detectors using movement signals to activate lights and security cameras. To detect events before AWS IoT Events, you had to build costly, custom applications to collect data, apply decision logic to detect an event, and then start another application to react to the event. Using AWS IoT Events, it's simple to detect events across thousands of IoT sensors sending different telemetry data, such as temperature from a freezer, humidity from respiratory equipment, and belt speed on a motor, and hundreds of equipment management applications. You simply select the relevant data sources to ingest, define the logic for each event using simple 'if-then-else' statements, and select the alert or custom action to run when an event occurs. AWS IoT Events continuously monitors data from multiple IoT sensors and applications, and it integrates with other services, such as AWS IoT Core and AWS IoT Analytics, to enable early detection and unique insights into events. AWS IoT Events automatically initiates alerts and actions in response to events based on the logic you define. This helps resolve issues quickly, reduce maintenance costs, and increase operational efficiency.

AWS IoT ExpressLink

AWS IoT ExpressLink powers a range of hardware modules developed and offered by AWS Partners, such as Espressif, Infineon, Realtek, and u-blox. Connectivity modules available from the AWS Partner Device Catalog include software implementing AWS mandated security requirements, making it faster and easier for you to securely connect devices to the cloud and seamlessly integrate with a range of AWS services. AWS IoT ExpressLink modules come pre-provisioned with security credentials set by qualified AWS Partners. This enables you to offload the complex work of integrating the networking and cryptography layers to the hardware modules, and develop secure IoT products in a fraction of the time.

Devices with AWS IoT ExpressLink establish a two-way connection with AWS IoT Core through native support of the MQTT (publish/subscribe) communication mechanism, and can create and update AWS IoT Device Shadow documents. With AWS IoT ExpressLink, it's easy to make over-the-air (OTA) updates to both the module and host processor from the AWS IoT Device Management console. You can then remotely deploy security updates, bug fixes, and new firmware updates to add features and keep your device fleet always up to date. Moreover, partner modules with AWS IoT ExpressLink can also connect to the AWS IoT Device Defender to report a number of device metrics that can help detect anomalies and generate alerts.

AWS IoT FleetWise

With AWS IoT FleetWise, you can collect and organize vehicle data and store that data in a standardized way for data analysis in the cloud. AWS IoT FleetWise helps you efficiently transfer

data to the cloud in near real time using intelligent data collection capabilities. These capabilities allow you to reduce the amount of data transferred by defining rules for when to collect and transfer data based on configurable parameters (for instance, vehicle temperature, speed, or make and model). Once the data is in the cloud, you can use it for applications that analyze vehicle fleet health. This analysis can help you to more quickly identify potential maintenance issues or make in-vehicle infotainment systems smarter. You can also feed the data into machine learning (ML) models that improve advanced technologies, such as autonomous driving and advanced driver assistance systems (ADAS).

AWS IoT Greengrass

AWS IoT Greengrass seamlessly extends AWS to devices so they can act locally on the data they generate, while still using the cloud for management, analytics, and durable storage. With AWS IoT Greengrass, connected devices can run AWS Lambda functions, run predictions based on machine learning models, keep device data in sync, and communicate with other devices securely – even when not connected to the internet.

With AWS IoT Greengrass, you can use familiar languages and programming models to create and test your device software in the cloud, and then deploy it to your devices. AWS IoT Greengrass can be programmed to filter device data and only transmit necessary information back to the cloud. You can also connect to third-party applications, on-premises software, and AWS services out-of-the-box with AWS IoT Greengrass Connectors. Connectors also jumpstart device onboarding with pre-built protocol adapter integrations and allow you to streamline authentication via integration with AWS Secrets Manager.

AWS IoT SiteWise

AWS IoT SiteWise is a managed service that makes it easy to collect, store, organize and monitor data from industrial equipment at scale to help you make better, data-driven decisions. You can use AWS IoT SiteWise to monitor operations across facilities, quickly compute common industrial performance metrics, and create applications that analyze industrial equipment data to prevent costly equipment issues and reduce gaps in production. This allows you to collect data consistently across devices, identify issues with remote monitoring more quickly, and improve multi-site processes with centralized data.

Today, getting performance metrics from industrial equipment is challenging because data is often locked into proprietary on-premises data stores and typically requires specialized expertise to retrieve and place in a format that is useful for analysis. AWS IoT SiteWise simplifies this process by

providing software running on a gateway that resides in your facilities and automates the process of collecting and organizing industrial equipment data. This gateway securely connects to your on-premises data servers, collects data, and sends the data to the AWS Cloud. AWS IoT SiteWise also provides interfaces for collecting data from modern industrial applications through MQTT messages or APIs.

You can use AWS IoT SiteWise to model your physical assets, processes and facilities, quickly compute common industrial performance metrics, and create fully managed web applications to help analyze industrial equipment data, reduce costs and make faster decisions. With AWS IoT SiteWise, you can focus on understanding and optimizing your operations, rather than building costly in-house data collection and management applications.

AWS IoT TwinMaker

AWS IoT TwinMaker makes it easier for developers to create digital twins of real-world systems such as buildings, factories, industrial equipment, and production lines. AWS IoT TwinMaker provides the tools you need to build digital twins to help you optimize building operations, increase production output, and improve equipment performance. With the ability to use existing data from multiple sources, create virtual representations of any physical environment, and combine existing 3D models with real-world data, you can now harness digital twins to create a holistic view of your operations faster and with less effort.

AWS Partner Device Catalog

The AWS Partner Device Catalog helps you find devices and hardware to help you explore, build, and go to market with your IoT solutions. Search for and find hardware that works with AWS, including development kits and embedded systems to build new devices, as well as off-the-shelf-devices such as gateways, edge servers, sensors, and cameras for immediate IoT project integration. The choice of AWS enabled hardware from our curated catalog of devices from APN partners can help make the rollout of your IoT projects easier. All devices listed in the AWS Partner Device Catalog are also available for purchase from our partners to get you started quickly.

FreeRTOS

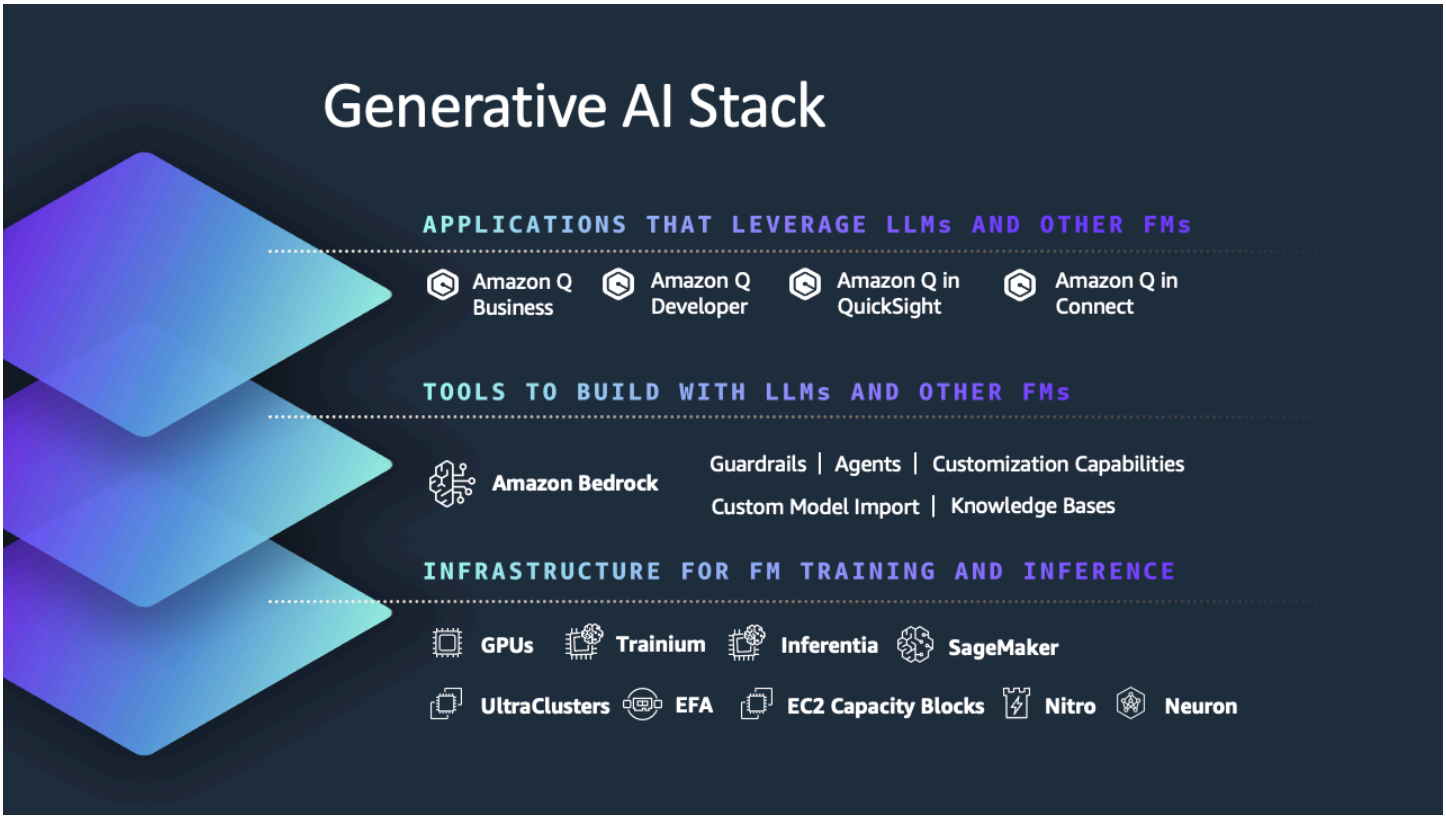
FreeRTOS is an operating system for microcontrollers that makes small, low-power edge devices easy to program, deploy, secure, connect, and manage. FreeRTOS extends the FreeRTOS kernel, a popular open source operating system for microcontrollers, with software libraries that make it easy to securely connect your small, low-power devices to AWS Cloud services such as AWS IoT Core or to more powerful edge devices running AWS IoT Greengrass.

A microcontroller (MCU) is a single chip containing a simple processor that can be found in many devices, including appliances, sensors, fitness trackers, industrial automation, and automobiles. Many of these small devices could benefit from connecting to the cloud or locally to other devices. For example, smart electricity meters need to connect to the cloud to report on usage, and building security systems need to communicate locally so that a door will unlock when you badge in. Microcontrollers have limited compute power and memory capacity and typically perform simple, functional tasks. Microcontrollers frequently run operating systems that do not have built-in functionality to connect to local networks or the cloud, making IoT applications a challenge. FreeRTOS helps solve this problem by providing both the core operating system (to run the edge device) as well as software libraries that make it easy to securely connect to the cloud (or other edge devices) so you can collect data from them for IoT applications and take action.

Machine Learning (ML) and Artificial Intelligence (AI)

AWS helps you at every stage of your ML adoption journey with the most comprehensive set ML services and purpose-built infrastructure. Our pretrained AI services provide ready-made intelligence for your applications and workflows.

Each service is described after the diagram. To help you decide which service best meets your needs, see Choosing an AWS machine learning service, Choosing a generative AI service, and Amazon Bedrock or Amazon SageMaker AI?. For general information, see Build and scale the next wave of AI innovation on AWS.



Services

- Amazon Augmented AI
- Amazon Bedrock
- Amazon CodeGuru
- Amazon Comprehend
- Amazon DevOps Guru
- Amazon Forecast
- Amazon Fraud Detector
- Amazon Comprehend Medical
- Amazon Kendra
- Amazon Lex
- Amazon Lookout for Equipment
- Amazon Lookout for Metrics
- Amazon Lookout for Vision
- Amazon Monitron

- Amazon PartyRock
- Amazon Personalize
- Amazon Polly
- Amazon Q
- Amazon Rekognition
- Amazon SageMaker AI
- Amazon Textract
- Amazon Transcribe
- Amazon Translate
- AWS DeepComposer
- AWS DeepRacer
- AWS HealthLake
- AWS HealthScribe
- AWS Panorama

Amazon Augmented AI

Amazon Augmented AI (Amazon A2I) is a ML service which makes it easy to build the workflows required for human review. Amazon A2I brings human review to all developers, removing the undifferentiated heavy lifting associated with building human review systems or managing large numbers of human reviewers, whether it runs on AWS or not.

Amazon Bedrock

Amazon Bedrock is a fully managed service that makes foundational models (FMs) from Amazon and leading AI companies available through an API. With the Amazon Bedrock serverless experience, you can quickly get started, experiment with FMs, privately customize them with your own data, and seamlessly integrate and deploy FMs into your AWS applications.

You can choose from a variety of foundation models from leading AI companies, such as AI21 Labs, Anthropic, Cohere, DeepSeek, Luma, Meta, Mistral AI, and Stability AI. Or you can use the Amazon Nova foundation models available exclusively in Amazon Bedrock.

Amazon CodeGuru

Amazon CodeGuru is a developer tool that provides intelligent recommendations to improve code quality and identify an application's most expensive lines of code. Integrate CodeGuru into your existing software development workflow to automate code reviews during application development and continuously monitor application's performance in production and provide recommendations and visual clues on how to improve code quality, application performance, and reduce overall cost.

Amazon CodeGuru Reviewer uses ML and automated reasoning to identify critical issues, security vulnerabilities, and hard-to-find bugs during application development and provides recommendations to improve code quality.

Amazon CodeGuru Profiler helps developers find an application's most expensive lines of code by helping them understand the runtime behavior of their applications, identify and remove code inefficiencies, improve performance, and significantly decrease compute costs.

Amazon Comprehend

Amazon Comprehend uses ML and natural language processing (NLP) to help you uncover the insights and relationships in your unstructured data. The service identifies the language of the text; extracts key phrases, places, people, brands, or events; understands how positive or negative the text is; analyzes text using tokenization and parts of speech; and automatically organizes a collection of text files by topic. You can also use AutoML capabilities in Amazon Comprehend to build a custom set of entities or text classification models that are tailored uniquely to your organization's needs.

For extracting complex medical information from unstructured text, you can use Amazon Comprehend Medical. The service can identify medical information, such as medical conditions, medications, dosages, strengths, and frequencies from a variety of sources like doctor's notes, clinical trial reports, and patient health records. Amazon Comprehend Medical also identifies the relationship among the extracted medication and test, treatment and procedure information for easier analysis. For example, the service identifies a particular dosage, strength, and frequency related to a specific medication from unstructured clinical notes.

Amazon DevOps Guru

Amazon DevOps Guru is an ML-powered service that makes it easy to improve an application's operational performance and availability. Amazon DevOps Guru detects behaviors that deviate

from normal operating patterns so you can identify operational issues long before they impact your customers.

Amazon DevOps Guru uses ML models informed by years of Amazon.com and AWS operational excellence to identify anomalous application behavior (such as increased latency, error rates, resource constraints, etc.) and surface critical issues that could cause potential outages or service disruptions. When Amazon DevOps Guru identifies a critical issue, it automatically sends an alert and provides a summary of related anomalies, the likely root cause, and context about when and where the issue occurred. When possible, Amazon DevOps Guru also provides recommendations on how to remediate the issue.

Amazon DevOps Guru automatically ingests operational data from your AWS applications and provides a single dashboard to visualize issues in your operational data. You can get started by enabling Amazon DevOps Guru for all resources in your AWS account, resources in your AWS CloudFormation Stacks, or resources grouped together by AWS tags, with no manual setup or ML expertise required.

Amazon Forecast

Amazon Forecast is a fully managed service that uses ML to deliver highly accurate forecasts.

Companies today use everything from simple spreadsheets to complex financial planning software to attempt to accurately forecast future business outcomes such as product demand, resource needs, or financial performance. These tools build forecasts by looking at a historical series of data, which is called time series data. For example, such tools may try to predict the future sales of a raincoat by looking only at its previous sales data with the underlying assumption that the future is determined by the past. This approach can struggle to produce accurate forecasts for large sets of data that have irregular trends. Also, it fails to easily combine data series that change over time (such as price, discounts, web traffic, and number of employees) with relevant independent variables such as product features and store locations.

Based on the same technology used at Amazon.com, Amazon Forecast uses ML to combine time series data with additional variables to build forecasts. Amazon Forecast requires no ML experience to get started. You only need to provide historical data, plus any additional data that you believe may impact your forecasts. For example, the demand for a particular color of a shirt may change with the seasons and store location. This complex relationship is hard to determine on its own, but ML is ideally suited to recognize it. Once you provide your data, Amazon Forecast will automatically examine it, identify what is meaningful, and produce a forecasting model capable of making predictions that are up to 50% more accurate than looking at time series data alone.

Amazon Forecast is a fully managed service, so there are no servers to provision, and no ML models to build, train, or deploy. You pay only for what you use, and there are no minimum fees and no upfront commitments.

Amazon Fraud Detector

Amazon Fraud Detector is a fully managed service that uses ML and more than 20 years of fraud detection expertise from Amazon, to identify potentially fraudulent activity so customers can catch more online fraud faster. Amazon Fraud Detector automates the time consuming and expensive steps to build, train, and deploy an ML model for fraud detection, making it easier for customers to leverage the technology. Amazon Fraud Detector customizes each model it creates to a customer's own dataset, making the accuracy of models higher than current one-size fits all ML solutions. And, because you pay only for what you use, you avoid large upfront expenses.

Amazon Comprehend Medical

Over the past decade, AWS has witnessed a digital transformation in health, with organizations capturing huge volumes of patient information every day. But this data is often unstructured and the process to extract this information is labor-intensive and error-prone. Amazon Comprehend Medical is a HIPAA-eligible natural language processing (NLP) service that uses machine learning that has been pre-trained to understand and extract health data from medical text, such as prescriptions, procedures, or diagnoses. Amazon Comprehend Medical can help you extract information from unstructured medical text accurately and quickly with medical ontologies like ICD-10-CM, RxNorm, and SNOMED CT and in turn accelerate insurance claim processing, improve population health, and accelerate pharmacovigilance.

Amazon Kendra

Amazon Kendra is an intelligent search service powered by ML. Amazon Kendra reimagines enterprise search for your websites and applications so your employees and customers can easily find the content they are looking for, even when it's scattered across multiple locations and content repositories within your organization.

Using Amazon Kendra, you can stop searching through troves of unstructured data and discover the right answers to your questions, when you need them. Amazon Kendra is a fully managed service, so there are no servers to provision, and no ML models to build, train, or deploy.

Amazon Lex

Amazon Lex is a fully managed artificial intelligence (AI) service to design, build, test, and deploy conversational interfaces into any application using voice and text. Lex provides the advanced deep learning functionalities of automatic speech recognition (ASR) for converting speech to text, and natural language understanding (NLU) to recognize the intent of the text, to enable you to build applications with highly engaging user experiences and lifelike conversational interactions, and create new categories of products. With Amazon Lex, the same deep learning technologies that power Amazon Alexa are now available to any developer, enabling you to quickly and easily build sophisticated, natural language, conversational bots ("chatbots") and voice enabled interactive voice response (IVR) systems.

Amazon Lex enables developers to build conversational chatbots quickly. With Amazon Lex, no deep learning expertise is necessary—to create a bot, you just specify the basic conversation flow in the Amazon Lex console. Amazon Lex manages the dialogue and dynamically adjusts the responses in the conversation. Using the console, you can build, test, and publish your text or voice chatbot. You can then add the conversational interfaces to bots on mobile devices, web applications, and chat platforms (for example, Facebook Messenger). There are no upfront costs or minimum fees to use Amazon Lex - you are charged only for the text or speech requests that are made. The pay-as-you-go pricing and the low cost per request make the service a cost-effective way to build conversational interfaces. With the Amazon Lex free tier, you can easily try Amazon Lex without any initial investment.

Amazon Lookout for Equipment

Amazon Lookout for Equipment analyzes the data from the sensors on your equipment (such as pressure in a generator, flow rate of a compressor, revolutions per minute of fans), to automatically train an ML model based on just your data, for your equipment – with no ML expertise required. Lookout for Equipment uses your unique ML model to analyze incoming sensor data in real-time and accurately identify early warning signs that could lead to machine failures. This means you can detect equipment abnormalities with speed and precision, quickly diagnose issues, take action to reduce expensive downtime, and reduce false alerts.

Amazon Lookout for Metrics

> **Note**
>
> On October 10, 2025, AWS will discontinue support for Amazon Lookout for Metrics. For more information, see Transitioning off Amazon Lookout for Metrics.

Amazon Lookout for Metrics uses ML to automatically detect and diagnose anomalies (outliers from the norm) in business and operational data, such as a sudden dip in sales revenue or customer acquisition rates. In a couple of clicks, you can connect Amazon Lookout for Metrics to popular data stores such as Amazon S3, Amazon Redshift, and Amazon Relational Database Service (Amazon RDS), as well as third-party Software as a Service (SaaS) applications, such as Salesforce, Servicenow, Zendesk, and Marketo, and start monitoring metrics that are important to your business. Lookout for Metrics automatically inspects and prepares the data from these sources to detect anomalies with greater speed and accuracy than traditional methods used for anomaly detection. You can also provide feedback on detected anomalies to tune the results and improve accuracy over time. Lookout for Metrics makes it easy to diagnose detected anomalies by grouping together anomalies that are related to the same event and sending an alert that includes a summary of the potential root cause. It also ranks anomalies in order of severity so that you can prioritize your attention to what matters the most to your business.

Amazon Lookout for Vision

Amazon Lookout for Vision is an ML service that spots defects and anomalies in visual representations using computer vision (CV). With Amazon Lookout for Vision, manufacturing companies can increase quality and reduce operational costs by quickly identifying differences in images of objects at scale. For example, Lookout for Vision can be used to identify missing components in products, damage to vehicles or structures, irregularities in production lines, miniscule defects in silicon wafers, and other similar problems. Amazon Lookout for Vision uses ML to see and understand images from any camera as a person would, but with an even higher degree of accuracy and at a much larger scale. Lookout for Vision allows customers to eliminate the need for costly and inconsistent manual inspection, while improving quality control, defect and damage assessment, and compliance. In minutes, you can begin using Lookout for Vision to automate inspection of images and objects – with no ML expertise required.

Amazon Monitron

Amazon Monitron is an end-to-end system that uses ML to detect abnormal behavior in industrial machinery, enabling you to implement predictive maintenance and reduce unplanned downtime.

Installing sensors and the necessary infrastructure for data connectivity, storage, analytics, and alerting are foundational elements for enabling predictive maintenance. However, to make it work, companies have historically needed skilled technicians and data scientists to piece together a complex solution from scratch. This included identifying and procuring the right type of sensors for their use cases and connecting them together with an IoT gateway (a device that aggregates and transmits data). As a result, few companies have been able to successfully implement predictive maintenance.

Amazon Monitron includes sensors to capture vibration and temperature data from equipment, a gateway device to securely transfer data to AWS, the Amazon Monitron service that analyzes the data for abnormal machine patterns using ML, and a companion mobile app to set up the devices and receive reports on operating behavior and alerts to potential failures in your machinery. You can start monitoring equipment health in minutes without any development work or ML experience required, and enable predictive maintenance with the same technology used to monitor equipment in Amazon Fulfillment Centers.

Amazon PartyRock

Amazon PartyRock makes learning generative AI easy with a hands-on, code-free app builder. Experiment with prompt engineering techniques, review generated responses, and develop intuition for generative AI while creating and exploring fun apps. PartyRock provides access to foundation models (FMs) from Amazon and leading AI companies through Amazon Bedrock, a fully managed serviced service.

Amazon Personalize

Amazon Personalize is an ML service that makes it easy for developers to create individualized recommendations for customers using their applications.

ML is increasingly used to improve customer engagement by powering personalized product and content recommendations, tailored search results, and targeted marketing promotions. However, developing the ML capabilities necessary to produce these sophisticated recommendation systems has been beyond the reach of most organizations today due to the complexity of developing ML

functionality. Amazon Personalize allows developers with no prior ML experience to easily build sophisticated personalization capabilities into their applications, using ML technology perfected from years of use on Amazon.com.

With Amazon Personalize, you provide an activity stream from your application – page views, signups, purchases, and so forth – as well as an inventory of the items you want to recommend, such as articles, products, videos, or music. You can also choose to provide Amazon Personalize with additional demographic information from your users such as age, or geographic location. Amazon Personalize processes and examines the data, identifies what is meaningful, selects the right algorithms, and trains and optimizes a personalization model that is customized for your data.

Amazon Personalize offers optimized recommenders for retail and media and entertainment that make it faster and easier to deliver high-performing personalized user experiences. Amazon Personalize also offers intelligent user segmentation so you can run more effective prospecting campaigns through your marketing channels. With our two new recipes, you can automatically segment your users based on their interest in different product categories, brands, and more.

All data analyzed by Amazon Personalize is kept private and secure, and only used for your customized recommendations. You can start serving your personalized predictions via a simple API call from inside the virtual private cloud that the service maintains. You pay only for what you use, and there are no minimum fees and no upfront commitments.

Amazon Personalize is like having your own Amazon.com ML personalization team at your disposal, 24 hours a day.

Amazon Polly

Amazon Polly is a service that turns text into lifelike speech. Amazon Polly lets you create applications that talk, enabling you to build entirely new categories of speech-enabled products. Amazon Polly is an Amazon artificial intelligence (AI) service that uses advanced deep learning technologies to synthesize speech that sounds like a human voice. Amazon Polly includes a wide selection of lifelike voices spread across dozens of languages, so you can select the ideal voice and build speech-enabled applications that work in many different countries.

Amazon Polly delivers the consistently fast response times required to support real-time, interactive dialog. You can cache and save Amazon Polly speech audio to replay offline or redistribute. And Amazon Polly is easy to use. You simply send the text you want converted into speech to the Amazon Polly API, and Amazon Polly immediately returns the audio stream to your

application so your application can play it directly or store it in a standard audio file format, such as MP3.

In addition to Standard TTS voices, Amazon Polly offers Neural Text-to-Speech (NTTS) voices that deliver advanced improvements in speech quality through a new machine learning approach. Polly's Neural TTS technology also supports a Newscaster speaking style that is tailored to news narration use cases. Finally, Amazon Polly Brand Voice can create a custom voice for your organization. This is a custom engagement where you will work with the Amazon Polly team to build an NTTS voice for the exclusive use of your organization.

With Amazon Polly, you pay only for the number of characters you convert to speech, and you can save and replay Amazon Polly generated speech. The Amazon Polly low cost per character converted, and lack of restrictions on storage and reuse of voice output, make it a cost-effective way to enable Text-to-Speech everywhere.

Amazon Q

Amazon Q is a generative AI-powered assistant for accelerating software development and leveraging your internal data.

Amazon Q Business

> Amazon Q Business can answer questions, provide summaries, generate content, and securely complete tasks based on data and information in your enterprise systems. It empowers employees to be more creative, data-driven, efficient, prepared, and productive.

Amazon Q Developer

> Amazon Q Developer (formerly Amazon CodeWhisperer) assists developers and IT professionals with their tasks—from coding, testing, and upgrading applications, to diagnosing errors, performing security scanning and fixes, and optimizing AWS resources. Amazon Q has advanced, multistep planning and reasoning capabilities that can transform existing code (for example, perform Java version upgrades) and implement new features generated from developer requests.

Amazon Rekognition

Amazon Rekognition makes it easy to add image and video analysis to your applications using proven, highly scalable, deep learning technology that requires no ML expertise to use. With Amazon Rekognition, you can identify objects, people, text, scenes, and activities in images and

videos, as well as detect any inappropriate content. Amazon Rekognition also provides highly accurate facial analysis and facial search capabilities that you can use to detect, analyze, and compare faces for a wide variety of user verification, people counting, and public safety use cases.

With Amazon Rekognition Custom Labels, you can identify the objects and scenes in images that are specific to your business needs. For example, you can build a model to classify specific machine parts on your assembly line or to detect unhealthy plants. Amazon Rekognition Custom Labels takes care of the heavy lifting of model development for you, so no ML experience is required. You simply need to supply images of objects or scenes you want to identify, and the service handles the rest.

Amazon SageMaker AI

With Amazon SageMaker AI, you can build, train, and deploy ML models for any use case with fully managed infrastructure, tools, and workflows. SageMaker AI removes the heavy lifting from each step of the ML process to make it easier to develop high-quality models. SageMaker AI provides all of the components used for ML in a single toolset so models get to production faster with much less effort and at lower cost.

Amazon SageMaker AI Autopilot

Amazon SageMaker AI Autopilot automatically builds, trains, and tunes the best ML models based on your data, while allowing you to maintain full control and visibility. With SageMaker AI Autopilot, you simply provide a tabular dataset and select the target column to predict, which can be a number (such as a house price, called regression), or a category (such as spam/not spam, called classification). SageMaker AI Autopilot will automatically explore different solutions to find the best model. You then can directly deploy the model to production with just one click, or iterate on the recommended solutions with Amazon SageMaker AI Studio to further improve the model quality.

Amazon SageMaker AI Canvas

Amazon SageMaker AI Canvas expands access to ML by providing business analysts with a visual point-and-click interface that allows them to generate accurate ML predictions on their own — without requiring any ML experience or having to write a single line of code.

Amazon SageMaker AI Clarify

Amazon SageMaker AI Clarify provides machine learning developers with greater visibility into their training data and models so they can identify and limit bias and explain predictions. Amazon

SageMaker AI Clarify detects potential bias during data preparation, after model training, and in your deployed model by examining attributes you specify. SageMaker AI Clarify also includes feature importance graphs that help you explain model predictions and produces reports which can be used to support internal presentations or to identify issues with your model that you can take steps to correct.

Amazon SageMaker AI Data Labeling

Amazon SageMaker AI provides data labeling offerings to identify raw data, such as images, text files, and videos, and add informative labels to create high-quality training datasets for your ML models.

Amazon SageMaker AI Data Wrangler

Amazon SageMaker AI Data Wrangler reduces the time it takes to aggregate and prepare data for ML from weeks to minutes. With SageMaker AI Data Wrangler, you can simplify the process of data preparation and feature engineering, and complete each step of the data preparation workflow, including data selection, cleansing, exploration, and visualization from a single visual interface.

Amazon SageMaker AI Edge

Amazon SageMaker AI Edge enables machine learning on edge devices by optimizing, securing, and deploying models to the edge, and then monitoring these models on your fleet of devices, such as smart cameras, robots, and other smart-electronics, to reduce ongoing operational costs. SageMaker AI Edge Compiler optimizes the trained model to be runnable on an edge device. SageMaker AI Edge includes an over-the-air (OTA) deployment mechanism that helps you deploy models on the fleet independent of the application or device firmware. SageMaker AI Edge Agent allows you to run multiple models on the same device. The Agent collects prediction data based on the logic that you control, such as intervals, and uploads it to the cloud so that you can periodically retrain your models over time.

Amazon SageMaker AI Feature Store

Amazon SageMaker AI Feature Store is a purpose-built repository where you can store and access features so it's much easier to name, organize, and reuse them across teams. SageMaker AI Feature Store provides a unified store for features during training and real-time inference without the need to write additional code or create manual processes to keep features consistent. SageMaker AI Feature Store keeps track of the metadata of stored features (such as feature name or version number) so that you can query the features for the right attributes in batches or in real time using

Amazon Athena, an interactive query service. SageMaker AI Feature Store also keeps features updated, because as new data is generated during inference, the single repository is updated so new features are always available for models to use during training and inference.

Amazon SageMaker AI geospatial capabilities

Amazon SageMaker AI geospatial capabilities make it easier for data scientists and machine learning (ML) engineers to build, train, and deploy ML models faster using geospatial data. You have access to data (open-source and third-party), processing, and visualization tools to make it more efficient to prepare geospatial data for ML. You can increase your productivity by using purpose-built algorithms and pre-trained ML models to speed up model building and training, and use built-in visualization tools to explore prediction outputs on an interactive map and then collaborate across teams on insights and results.

Amazon SageMaker AI HyperPod

Amazon SageMaker AI HyperPod removes the undifferentiated heavy lifting involved in building and optimizing machine learning (ML) infrastructure for large language models (LLMs), diffusion models, and foundation models (FMs). SageMaker AI HyperPod is pre-configured with distributed training libraries that enable customers to automatically split training workloads across thousands of accelerators, such as AWS Trainium, and NVIDIA A100 and H100 Graphical Processing Units (GPUs).

SageMaker AI HyperPod also helps ensure that you can continue training uninterrupted by periodically saving checkpoints. When a hardware failure occurs, self-healing clusters automatically detect the failure, repair or replace the faulty instance, and resume the training from the last saved checkpoint, removing the need for you to manually manage this process and helping you train for weeks or months in a distributed setting without disruption. You can customize your computing environment to best suit your needs and configure it with the Amazon SageMaker AI distributed training libraries to achieve optimal performance on AWS.

Amazon SageMaker AI JumpStart

Amazon SageMaker AI JumpStart helps you quickly and easily get started with ML. To make it easier to get started, SageMaker AI JumpStart provides a set of solutions for the most common use cases that can be deployed readily with just a few clicks. The solutions are fully customizable and showcase the use of AWS CloudFormation templates and reference architectures so you can accelerate your ML journey. Amazon SageMaker AI JumpStart also supports one-click

deployment and fine-tuning of more than 150 popular open-source models such as natural language processing, object detection, and image classification models.

Amazon SageMaker AI Model Building

Amazon SageMaker AI provides all the tools and libraries you need to build ML models, the process of iteratively trying different algorithms and evaluating their accuracy to find the best one for your use case. In Amazon SageMaker AI you can pick different algorithms, including over 15 that are built-in and optimized for SageMaker AI, and use over 750 pre-built models from popular model zoos available with a few clicks. SageMaker AI also offers a variety of model building tools, including Amazon SageMaker AI Studio Notebooks, JupyterLab, RStudio, and Code Editor based on Code-OSS (Virtual Studio Code Open Source), where you can run ML models on a small scale to see results and view reports on their performance so you can come up with high-quality working prototypes.

Amazon SageMaker AI Model Training

Amazon SageMaker AI reduces the time and cost to train and tune ML models at scale without the need to manage infrastructure. You can take advantage of the highest-performing ML compute infrastructure currently available, and SageMaker AI can automatically scale infrastructure up or down, from one to thousands of GPUs. Since you pay only for what you use, you can manage your training costs more effectively. To train deep learning models faster, you can use the Amazon SageMaker AI distributed training libraries for better performance or use third-party libraries such as DeepSpeed, Horovod, or Megatron.

Amazon SageMaker AI Model Deployment

Amazon SageMaker AI makes it easy to deploy ML models to make predictions (also known as inference) at the best price-performance for any use case. It provides a broad selection of ML infrastructure and model deployment options to help meet all your ML inference needs. It is a fully managed service and integrates with MLOps tools, so you can scale your model deployment, reduce inference costs, manage models more effectively in production, and reduce operational burden.

Amazon SageMaker AI Pipelines

Amazon SageMaker AI Pipelines is the first purpose-built, easy-to-use continuous integration and continuous delivery (CI/CD) service for ML. With SageMaker AI Pipelines, you can create, automate, and manage end-to-end ML workflows at scale.

Amazon SageMaker AI Studio Lab

Amazon SageMaker AI Studio Lab is a free ML development environment that provides the compute, storage (up to 15GB), and security—all at no cost—for anyone to learn and experiment with ML. All you need to get started is a valid email address—you don't need to configure infrastructure or manage identity and access or even sign up for an AWS account. SageMaker AI Studio Lab accelerates model building through GitHub integration, and it comes preconfigured with the most popular ML tools, frameworks, and libraries to get you started immediately. SageMaker AI Studio Lab automatically saves your work so you don't need to restart in between sessions. It's as easy as closing your laptop and coming back later.

Apache MXNet on AWS

Apache MXNet is a fast and scalable training and inference framework with an easy-to-use, concise API for ML. MXNet includes the Gluon interface that allows developers of all skill levels to get started with deep learning on the cloud, on edge devices, and on mobile apps. In just a few lines of Gluon code, you can build linear regression, convolutional networks and recurrent LSTMs for object detection, speech recognition, recommendation, and personalization. You can get started with MxNet on AWS with a fully managed experience using Amazon SageMaker AI, a platform to build, train, and deploy ML models at scale. Or, you can use the AWS Deep Learning AMIs to build custom environments and workflows with MxNet as well as other frameworks including TensorFlow, PyTorch, Chainer, Keras, Caffe, Caffe2, and Microsoft Cognitive Toolkit.

AWS Deep Learning AMIs

The AWS Deep Learning AMIs provide ML practitioners and researchers with the infrastructure and tools to accelerate deep learning in the cloud, at any scale. You can quickly launch Amazon EC2 instances pre-installed with popular deep learning frameworks and interfaces such as TensorFlow, PyTorch, Apache MXNet, Chainer, Gluon, Horovod, and Keras to train sophisticated, custom AI models, experiment with new algorithms, or to learn new skills and techniques. Whether you need Amazon EC2 GPU or CPU instances, there is no additional charge for the Deep Learning AMIs – you only pay for the AWS resources needed to store and run your applications.

AWS Deep Learning Containers

AWS Deep Learning Containers (AWS DL Containers) are Docker images pre-installed with deep learning frameworks to make it easy to deploy custom machine learning (ML) environments quickly by letting you skip the complicated process of building and optimizing your environments from scratch. AWS DL Containers support TensorFlow, PyTorch, Apache MXNet. You can deploy

AWS DL Containers on Amazon SageMaker AI, Amazon Elastic Kubernetes Service (Amazon EKS), self-managed Kubernetes on Amazon EC2, Amazon Elastic Container Service (Amazon ECS). The containers are available through Amazon Elastic Container Registry (Amazon ECR) and AWS Marketplace at no cost—you pay only for the resources that you use.

Geospatial ML with Amazon SageMaker AI

Amazon SageMaker AI geospatial capabilities allow data scientists and ML engineers to build, train, and deploy ML models using geospatial data faster and at scale. You can access readily available geospatial data sources, efficiently transform or enrich large-scale geospatial datasets with purpose-built operations, and accelerate model building by selecting pretrained ML models. You can also analyze geospatial data and explore model predictions on an interactive map using 3D accelerated graphics with built-in visualization tools. SageMaker Runtime geospatial capabilities can be used for a wide range of use cases, such as maximizing harvest yield and food security, assessing risk and insurance claims, supporting sustainable urban development, and forecasting retail site utilization.

Hugging Face on AWS

With Hugging Face on Amazon SageMaker AI, you can deploy and fine-tune pre-trained models from Hugging Face, an open-source provider of natural language processing (NLP) models known as Transformers, reducing the time it takes to set up and use these NLP models from weeks to minutes. NLP refers to ML algorithms that help computers understand human language. They help with translation, intelligent search, text analysis, and more. However, NLP models can be large and complex (sometimes consisting of hundreds of millions of model parameters), and training and optimizing them requires time, resources, and skill. AWS collaborated with Hugging Face to create Hugging Face AWS Deep Learning Containers (DLCs), which provide data scientists and ML developers a fully managed experience for building, training, and deploying state-of-the-art NLP models on Amazon SageMaker AI.

PyTorch on AWS

PyTorch is an open-source deep learning framework that makes it easy to develop machine learning models and deploy them to production. Using TorchServe, PyTorch's model serving library built and maintained by AWS in partnership with Facebook, PyTorch developers can quickly and easily deploy models to production. PyTorch also provides dynamic computation graphs and libraries for distributed training, which are tuned for high performance on AWS. You can get started with PyTorch on AWS using Amazon SageMaker, a fully managed ML service that makes it

easy and cost-effective to build, train, and deploy PyTorch models at scale. If you prefer to manage the infrastructure yourself, you can use the AWS Deep Learning AMIs or the AWS Deep Learning Containers, which come built from source and optimized for performance with the latest version of PyTorch to quickly deploy custom machine learning environments.

TensorFlow on AWS

TensorFlow is one of many deep learning frameworks available to researchers and developers to enhance their applications with machine learning. AWS provides broad support for TensorFlow, enabling customers to develop and serve their own models across computer vision, natural language processing, speech translation, and more. You can get started with TensorFlow on AWS using Amazon SageMaker AI, a fully managed ML service that makes it easy and cost-effective to build, train, and deploy TensorFlow models at scale. If you prefer to manage the infrastructure yourself, you can use the AWS Deep Learning AMIs or the AWS Deep Learning Containers, which come built from source and optimized for performance with the latest version of TensorFlow to quickly deploy custom ML environments.

Amazon Textract

Amazon Textract is a service that automatically extracts text and data from scanned documents. Amazon Textract goes beyond simple optical character recognition (OCR) to also identify the contents of fields in forms and information stored in tables.

Today, many companies manually extract data from scanned documents such as PDFs, images, tables, and forms, or through simple OCR software that requires manual configuration (which often must be updated when the form changes). To overcome these manual and expensive processes, Amazon Textract uses ML to read and process any type of document, accurately extracting text, handwriting, tables, and other data with no manual effort. Amazon Textract provides you with the flexibility to specify the data you need to extract from documents using queries. You can specify the information you need in the form of natural language questions (such as “What is the customer name”). You do not need to know the data structure in the document (table, form, implied field, nested data) or worry about variations across document versions and formats. Amazon Textract Queries are pre-trained on a large variety of documents including paystubs, bank statements, W-2s, loan application forms, mortgage notes, claims documents, and insurance cards.

With Amazon Textract, you can quickly automate document processing and act on the information extracted, whether you’re automating loans processing or extracting information from invoices and

receipts. Amazon Textract can extract the data in minutes instead of hours or days. Additionally, you can add human reviews with Amazon Augmented AI to provide oversight of your models and check sensitive data.

Amazon Transcribe

Amazon Transcribe is an automatic speech recognition (ASR) service that makes it easy for customers to automatically convert speech to text. The service can transcribe audio files stored in common formats, like WAV and MP3, with time stamps for every word so that you can easily locate the audio in the original source by searching for the text. You can also send a live audio stream to Amazon Transcribe and receive a stream of transcripts in real time. Amazon Transcribe is designed to handle a wide range of speech and acoustic characteristics, including variations in volume, pitch, and speaking rate. The quality and content of the audio signal (including but not limited to factors such as background noise, overlapping speakers, accented speech, or switches between languages within a single audio file) may affect the accuracy of service output. Customers can choose to use Amazon Transcribe for a variety of business applications, including transcription of voice-based customer service calls, generation of subtitles on audio/video content, and conduct (text based) content analysis on audio/video content.

Two very important services derived from Amazon Transcribe include Amazon Transcribe Medical and Amazon Transcribe Call Analytics.

Amazon Transcribe Medical uses advanced ML models to accurately transcribe medical speech into text. Amazon Transcribe Medical can generate text transcripts that can be used to support a variety of use cases, spanning clinical documentation workflow and drug safety monitoring (pharmacovigilance) to subtitling for telemedicine and even contact center analytics in the healthcare and life sciences domains.

Amazon Transcribe Call Analytics is an AI-powered API that provides rich call transcripts and actionable conversation insights that you can add into their call applications to improve customer experience and agent productivity. It combines powerful speech-to-text and custom natural language processing (NLP) models that are trained specifically to understand customer care and outbound sales calls. As a part of AWS Contact Center Intelligence (CCI) solutions, this API is contact center agnostic and makes it easy for customers and ISVs to add call analytics capabilities into their applications.

The easiest way to get started with Amazon Transcribe is to submit a job using the console to transcribe an audio file. You can also call the service directly from the AWS Command Line Interface, or use one of the supported SDKs of your choice to integrate with your applications.

Amazon Translate

Amazon Translate is a neural machine translation service that delivers fast, high-quality, and affordable language translation. Neural machine translation is a form of language translation automation that uses deep learning models to deliver more accurate and more natural sounding translation than traditional statistical and rule-based translation algorithms. Amazon Translate allows you to localize content such as websites and applications for your diverse users, easily translate large volumes of text for analysis, and efficiently enable cross-lingual communication between users.

AWS DeepComposer

AWS DeepComposer is the world's first musical keyboard powered by ML to enable developers of all skill levels to learn Generative AI while creating original music outputs. DeepComposer consists of a USB keyboard that connects to the developer's computer, and the DeepComposer service, accessed through the AWS Management Console. DeepComposer includes tutorials, sample code, and training data that can be used to start building generative models.

AWS DeepRacer

AWS DeepRacer is a 1/18th scale race car which gives you an interesting and fun way to get started with reinforcement learning (RL). RL is an advanced ML technique which takes a very different approach to training models than other ML methods. Its superpower is that it learns very complex behaviors without requiring any labeled training data, and can make short term decisions while optimizing for a longer term goal.

With AWS DeepRacer, you now have a way to get hands-on with RL, experiment, and learn through autonomous driving. You can get started with the virtual car and tracks in the cloud-based 3D racing simulator, and for a real-world experience, you can deploy your trained models onto AWS DeepRacer and race your friends, or take part in the global AWS DeepRacer League. Developers, the race is on.

AWS HealthLake

AWS HealthLake is a HIPAA-eligible service that healthcare providers, health insurance companies, and pharmaceutical companies can use to store, transform, query, and analyze large-scale health data.

Health data is frequently incomplete and inconsistent. It's also often unstructured, with information contained in clinical notes, lab reports, insurance claims, medical images, recorded conversations, and time-series data (for example, heart ECG or brain EEG traces).

Healthcare providers can use HealthLake to store, transform, query, and analyze data in the AWS Cloud. Using the HealthLake integrated medical natural language processing (NLP) capabilities, you can analyze unstructured clinical text from diverse sources. HealthLake transforms unstructured data using natural language processing models, and provides powerful query and search capabilities. You can use HealthLake to organize, index, and structure patient information in a secure, compliant, and auditable manner.

AWS HealthScribe

AWS HealthScribe is a HIPAA-eligible service that allows healthcare software vendors to automatically generate clinical notes by analyzing patient-clinician conversations. AWS HealthScribe combines speech recognition with generative AI to reduce the burden of clinical documentation by transcribing conversations and quickly producing clinical notes. Conversations are segmented to identify the speaker roles for patients and clinicians, extract medical terms, and generate preliminary clinical notes. To protect sensitive patient data, security and privacy are built-in to ensure that the input audio and the output text are not retained in AWS HealthScribe.

AWS Panorama

AWS Panorama is a collection of ML devices and software development kit (SDK) that brings computer vision (CV) to on-premises internet protocol (IP) cameras. With AWS Panorama, you can automate tasks that have traditionally required human inspection to improve visibility into potential issues.

Computer vision can automate visual inspection for tasks such as tracking assets to optimize supply chain operations, monitoring traffic lanes to optimize traffic management, or detecting anomalies to evaluate manufacturing quality. In environments with limited network bandwidth however, or for companies with data governance rules that require on-premises processing and storage of video, computer vision in the cloud can be difficult or impossible to implement. AWS Panorama is an ML service that allows organizations to bring computer vision to on-premises cameras to make predictions locally with high accuracy and low latency.

The AWS Panorama Appliance is a hardware device that adds computer vision to your existing IP cameras and analyzes the video feeds of multiple cameras from a single management interface.

It generates predictions at the edge in milliseconds, meaning you can be notified about potential issues such as when damaged products are detected on a fast-moving production line, or when a vehicle has strayed into a dangerous off-limits zone in a warehouse. And, third-party manufacturers are building new AWS Panorama-enabled cameras and devices to provide even more form factors for your unique use cases. With AWS Panorama you can use ML models from AWS to build your own computer vision applications, or work with a partner from the AWS Partner Network to build CV applications quickly.

Management and governance

With AWS Management and Governance services, you don't have to choose between innovating faster and maintaining control over cost, compliance, and security—you can do both.

For general information, see Management and Governance on AWS.

Services

- AWS Auto Scaling
- AWS CloudFormation
- AWS CloudTrail
- Amazon CloudWatch
- AWS Compute Optimizer
- AWS Console Mobile Application
- AWS Control Tower
- AWS Config
- AWS Health
- AWS Launch Wizard
- AWS License Manager
- Amazon Managed Grafana
- Amazon Managed Service for Prometheus
- AWS Organizations

- AWS OpsWorks
- AWS Proton
- Amazon Q Developer in chat applications (formerly AWS Chatbot)
- AWS Service Catalog
- AWS Systems Manager
- AWS Trusted Advisor
- AWS User Notifications
- AWS Well-Architected Tool

AWS Auto Scaling

AWS Auto Scaling monitors your applications and automatically adjusts capacity to maintain steady, predictable performance at the lowest possible cost. Using AWS Auto Scaling, it's easy to setup application scaling for multiple resources across multiple services in minutes. The service provides a simple, powerful user interface that lets you build scaling plans for resources including Amazon EC2 instances and Spot Fleets, Amazon ECS tasks, Amazon DynamoDB tables and indexes, and Amazon Aurora Replicas. AWS Auto Scaling makes scaling simple with recommendations that allow you to optimize performance, costs, or balance between them. If you're already using Amazon EC2 Auto Scaling to dynamically scale your Amazon EC2 instances, you can now combine it with AWS Auto Scaling to scale additional resources for other AWS services. With AWS Auto Scaling, your applications always have the right resources at the right time.

AWS CloudFormation

AWS CloudFormation gives developers and systems administrators an easy way to create and manage a collection of related AWS resources, provisioning and updating them in an orderly and predictable fashion.

You can use the AWS CloudFormation sample templates or create your own templates to describe your AWS resources, and any associated dependencies or runtime parameters, required to run your application. You don't need to figure out the order for provisioning AWS services or the subtleties of making those dependencies work. CloudFormation takes care of this for you. After the AWS resources are deployed, you can modify and update them in a controlled and predictable way, in effect applying version control to your AWS infrastructure the same way you do with your software. You can also visualize your templates as diagrams and edit them using a drag-and-drop interface with AWS Infrastructure Composer.

AWS CloudTrail

AWS CloudTrail is a web service that records AWS API calls for your account and delivers log files to you. The recorded information includes the identity of the API caller, the time of the API call, the source IP address of the API caller, the request parameters, and the response elements returned by the AWS service.

With CloudTrail, you can get a history of AWS API calls for your account, including API calls made using the AWS Management Console, AWS SDKs, command line tools, and higher-level AWS services (such as AWS CloudFormation). The AWS API call history produced by CloudTrail enables security analysis, resource change tracking, and compliance auditing.

Amazon CloudWatch

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

AWS Compute Optimizer

AWS Compute Optimizer recommends optimal AWS resources for your workloads to reduce costs and improve performance by using machine learning to analyze historical utilization metrics. Over-provisioning resources can lead to unnecessary infrastructure cost, and under-provisioning resources can lead to poor application performance. Compute Optimizer helps you choose optimal configurations for three types of AWS resources: Amazon EC2 instances, Amazon EBS volumes, and AWS Lambda functions, based on your utilization data.

By applying the knowledge drawn from Amazon's own experience running diverse workloads in the cloud, Compute Optimizer identifies workload patterns and recommends optimal AWS resources. Compute Optimizer analyzes the configuration and resource utilization of your workload to identify dozens of defining characteristics, for example, if a workload is CPU-intensive, if it exhibits a daily pattern, or if a workload accesses local storage frequently. The service processes

these characteristics and identifies the hardware resource required by the workload. Compute Optimizer infers how the workload would have performed on various hardware platforms (such as Amazon EC2 instances types) or using different configurations (such as Amazon EBS volume IOPS settings, and AWS Lambda function memory sizes) to offer recommendations.

Compute Optimizer is available to you at no additional charge. To get started, you can opt in to the service in the AWS Compute Optimizer Console.

AWS Console Mobile Application

The AWS Console Mobile Application lets customers view and manage a select set of resources to support incident response while on-the-go.

The AWS Console Mobile Application allows AWS customers to monitor resources through a dedicated dashboard and view configuration details, metrics, and alarms for select AWS services. The Dashboard provides permitted users with a single view a resource's status, with real-time data on Amazon CloudWatch, AWS Health Dashboard, and AWS Billing and Cost Management. Customers can view ongoing issues and follow through to the relevant CloudWatch alarm screen for a detailed view with graphs and configuration options. In addition, customers can check on the status of specific AWS services, view detailed resource screens, and perform select actions.

AWS Control Tower

AWS Control Tower automates the set-up of a baseline environment, or landing zone, that is a secure, well-architected multi-account AWS environment. The configuration of the landing zone is based on best practices that have been established by working with thousands of enterprise customers to create a secure environment that makes it easier to govern AWS workloads with rules for security, operations, and compliance.

As enterprises migrate to AWS, they typically have a large number of applications and distributed teams. They often want to create multiple accounts to allow their teams to work independently, while still maintaining a consistent level of security and compliance. In addition, they use AWS management and security services, such as AWS Organizations, Service Catalog and AWS Config, that provide very granular controls over their workloads. They want to maintain this control, but they also want a way to centrally govern and enforce the best use of AWS services across all the accounts in their environment.

AWS Control Tower automates the set-up of their landing zone and configures AWS management and security services based on established best practices in a secure, compliant, multi-account

environment. Distributed teams are able to provision new AWS accounts quickly, while central teams have the peace of mind knowing that new accounts are aligned with centrally established, company-wide compliance policies. This gives you control over your environment, without sacrificing the speed and agility AWS provides your development teams.

AWS Config

AWS Config is a fully managed service that provides you with an AWS resource inventory, configuration history, and configuration change notifications to enable security and governance. The AWS Config Rules feature enables you to create rules that automatically check the configuration of AWS resources recorded by AWS Config.

With AWS Config, you can discover existing and deleted AWS resources, determine your overall compliance against rules, and dive into configuration details of a resource at any point in time. These capabilities enable compliance auditing, security analysis, resource change tracking, and troubleshooting.

AWS Health

AWS Health provides alerts and remediation guidance when AWS is experiencing events that might affect you. While the Service Health Dashboard displays the general status of AWS services, AWS Health Dashboard gives you a personalized view into the performance and availability of the AWS services underlying your AWS resources. The dashboard displays relevant and timely information to help you manage events in progress, and provides proactive notification to help you plan for scheduled activities. With AWS Health, alerts are automatically initiated by changes in the health of AWS resources, giving you event visibility and guidance to help quickly diagnose and resolve issues.

AWS Launch Wizard

AWS Launch Wizard offers a guided way of sizing, configuring, and deploying AWS resources for third party applications, such as Microsoft SQL Server Always On and HANA based SAP systems, without the need to manually identify and provision individual AWS resources. To start, you input your application requirements, including performance, number of nodes, and connectivity on the service console. Launch Wizard then identifies the right AWS resources, such as EC2 instances and EBS volumes, to deploy and run your application. Launch Wizard provides an estimated cost of deployment, and lets you modify your resources to instantly view an updated cost assessment. Once you approve the AWS resources, Launch Wizard automatically provisions and configures the selected resources to create a fully functioning, production-ready application.

AWS Launch Wizard also creates CloudFormation templates that can serve as a baseline to accelerate subsequent deployments. Launch Wizard is available to you at no additional charge. You only pay for the AWS resources that are provisioned for running your solution.

AWS License Manager

AWS License Manager makes it easier to manage licenses in AWS and on-premises servers from software vendors such as Microsoft, SAP, Oracle, and IBM. AWS License Manager lets administrators create customized licensing rules that emulate the terms of their licensing agreements, and then enforces these rules when an instance of Amazon EC2 gets launched. Administrators can use these rules to limit licensing violations, such as using more licenses than an agreement stipulates or reassigning licenses to different servers on a short-term basis. The rules in AWS License Manager enable you to limit a licensing breach by physically stopping the instance from launching or by notifying administrators about the infringement. Administrators gain control and visibility of all their licenses with the AWS License Manager dashboard and reduce the risk of non-compliance, misreporting, and additional costs due to licensing overages.

AWS License Manager integrates with AWS services to simplify the management of licenses across multiple AWS accounts, IT catalogs, and on-premises, through a single AWS account. License administrators can add rules in Service Catalog, which allows them to create and manage catalogs of IT services that are approved for use on all their AWS accounts. Through seamless integration with AWS Systems Manager and AWS Organizations, administrators can manage licenses across all the AWS accounts in an organization and on-premises environments. AWS Marketplace buyers can also use AWS License Manager to track bring your own license (BYOL) software obtained from the Marketplace and keep a consolidated view of all their licenses.

Amazon Managed Grafana

Amazon Managed Grafana is a fully managed and secure data visualization service that you can use to instantly query, correlate, and visualize operational metrics, logs, and traces from multiple sources. Amazon Managed Grafana makes it easy to deploy, operate, and scale Grafana, a widely deployed open-source data visualization tool that is popular for its extensible data support.

Amazon Managed Grafana provides built-in security features for compliance with corporate governance requirements, including single sign-on, data access control, and audit reporting. Amazon Managed Grafana integrates with AWS data sources, such as Amazon CloudWatch, Amazon OpenSearch Service, AWS X-Ray, AWS IoT SiteWise, Amazon Timestream, and Amazon Managed Service for Prometheus. Amazon Managed Grafana also supports many popular open-source, third party, and other cloud data sources.

Amazon Managed Service for Prometheus

Amazon Managed Service for Prometheus is a serverless, Prometheus-compatible monitoring service for container metrics that makes it easier to securely monitor container environments at scale. With Amazon Managed Service for Prometheus, you can use the same open-source Prometheus data model and query language that you use today to monitor the performance of your containerized workloads, and also enjoy improved scalability, availability, and security without having to manage the underlying infrastructure.

Amazon Managed Service for Prometheus automatically scales the ingestion, storage, and querying of operational metrics as workloads scale up and down. It integrates with AWS security services to enable fast and secure access to data. Designed to be highly available, data ingested into a workspace is replicated across three Availability Zones in the same AWS Region.

AWS Organizations

AWS Organizations helps you centrally manage and govern your environment as you grow and scale your AWS resources. Using AWS Organizations, you can programmatically create new AWS accounts and allocate resources, group accounts to organize your workflows, apply policies to accounts or groups for governance, and simplify billing by using a single payment method for all of your accounts.

In addition, AWS Organizations is integrated with other AWS services so you can define central configurations, security mechanisms, audit requirements, and resource sharing across accounts in your organization. AWS Organizations is available to all AWS customers at no additional charge.

AWS OpsWorks

AWS OpsWorks is a configuration management service that provides managed instances of Chef and Puppet. Chef and Puppet are automation platforms that allow you to use code to automate the configurations of your servers. AWS OpsWorks lets you use Chef and Puppet to automate how servers are configured, deployed, and managed across your Amazon EC2 instances or on-premises compute environments. AWS OpsWorks has three offerings, AWS OpsWorks for Chef Automate, AWS OpsWorks for Puppet Enterprise, and AWS OpsWorks Stacks.

AWS Proton

AWS Proton is the first fully managed delivery service for container and serverless applications. Platform engineering teams can use AWS Proton to connect and coordinate all the different tools needed for infrastructure provisioning, code deployments, monitoring, and updates.

Maintaining hundreds – or sometimes thousands – of microservices with constantly changing infrastructure resources and continuous integration/continuous delivery (CI/CD) configurations is a nearly impossible task for even the most capable platform teams.

AWS Proton solves this by giving platform teams the tools they need to manage this complexity and enforce consistent standards, while making it easy for developers to deploy their code using containers and serverless technologies.

Amazon Q Developer in chat applications (formerly AWS Chatbot)

Amazon Q Developer in chat applications is an interactive agent that makes it easy to monitor and interact with your AWS resources in your Slack, Microsoft Teams, and Amazon Chime chat rooms. With Amazon Q Developer in chat applications, you can receive alerts, run commands to return diagnostic information, invoke AWS Lambda functions, and create AWS support cases.

Amazon Q Developer in chat applications manages the integration between AWS services and your Slack channels, Microsoft Teams, and Amazon Chime chat rooms helping you to get started with ChatOps fast. With just a few clicks, you can start receiving notifications and issuing commands in your chosen channels or chat rooms, so your team doesn't have to switch contexts to collaborate. Amazon Q Developer in chat applications makes it easier for your team to stay updated, collaborate, and respond faster to operational events, security findings, CI/CD workflows, budget, and other alerts for applications running in your AWS accounts.

AWS Service Catalog

AWS Service Catalog allows organizations to create and manage catalogs of IT services that are approved for use on AWS. These IT services can include everything from virtual machine images, servers, software, and databases to complete multi-tier application architectures. Service Catalog allows you to centrally manage commonly deployed IT services and helps you achieve consistent governance and meet your compliance requirements, while enabling users to quickly deploy only the approved IT services they need.

AWS Systems Manager

AWS Systems Manager gives you visibility and control of your infrastructure on AWS. Systems Manager provides a unified user interface so you can view operational data from multiple AWS services and allows you to automate operational tasks across your AWS resources. With Systems Manager, you can group resources, such as Amazon EC2 instances, Amazon S3 buckets, or Amazon RDS instances, by application, view operational data for monitoring and troubleshooting, and take action on your groups of resources. Systems Manager simplifies resource and application management, shortens the time to detect and resolve operational problems, and makes it easy to operate and manage your infrastructure securely at scale.

AWS Systems Manager contains the following tools:

- **Resource groups** — Lets you create a logical group of resources associated with a particular workload such as different layers of an application stack, or production versus development environments. For example, you can group different layers of an application, such as the frontend web layer and the backend data layer. Resource groups can be created, updated, or removed programmatically through the API.
- **Insights dashboard** — Displays operational data that the AWS Systems Manager automatically aggregates for each resource group. Systems Manager eliminates the need for you to navigate across multiple AWS consoles to view your operational data. With Systems Manager you can view API call logs from AWS CloudTrail, resource configuration changes from AWS Config, software inventory, and patch compliance status by resource group. You can also easily integrate your Amazon CloudWatch dashboards, AWS Trusted Advisor notifications, and AWS Health Dashboard performance and availability alerts into your Systems Manager dashboard. Systems Manager centralizes all relevant operational data, so you can have a clear view of your infrastructure compliance and performance.
- **Run command** — Provides a simple way of automating common administrative tasks such as remotely running shell scripts or PowerShell commands, installing software updates, or making changes to the configuration of OS, software, EC2 and instances and servers in your on-premises data center.
- **State Manager** — Helps you define and maintain consistent OS configurations such as firewall settings and anti-malware definitions to comply with your policies. You can monitor the configuration of a large set of instances, specify a configuration policy for the instances, and automatically apply updates or configuration changes.
- **Inventory** — Helps you collect and query configuration and inventory information about your instances and the software installed on them. You can gather details about your instances such

as installed applications, DHCP settings, agent detail, and custom items. You can run queries to track and audit your system configurations.

- **Maintenance Window** — Lets you define a recurring window of time to run administrative and maintenance tasks across your instances. This ensures that installing patches and updates, or making other configuration changes does not disrupt business-critical operations. This helps improve your application availability.
- **Patch Manager** — Helps you select and deploy operating system and software patches automatically across large groups of instances. You can define a maintenance window so that patches are applied only during set times that fit your needs. These capabilities help ensure that your software is always up to date and meets your compliance policies.
- **Automation** — Simplifies common maintenance and deployment tasks, such as updating Amazon Machine Images (AMIs). Use the Automation feature to apply patches, update drivers and agents, or bake applications into your AMI using a streamlined, repeatable, and auditable process.
- **Parameter Store** — Provides an encrypted location to store important administrative information such as passwords and database strings. The Parameter Store integrates with AWS Key Management Service (AWS KMS) to make it easy to encrypt the information you keep in the Parameter Store.
- **Distributor** — Helps you securely distribute and install software packages, such as software agents. Systems Manager Distributor allows you to centrally store and systematically distribute software packages while you maintain control over versioning. You can use Distributor to create and distribute software packages and then install them using Systems Manager Run Command and State Manager. Distributor can also use AWS Identity and Access Management (IAM) policies to control who can create or update packages in your account. You can use the existing IAM policy support for Systems Manager Run Command and State Manager to define who can install packages on your hosts.
- **Session Manager** — Provides a browser-based interactive shell and CLI for managing Windows and Linux EC2 instances, without the need to open inbound ports, manage SSH keys, or use bastion hosts. Administrators can grant and revoke access to instances through a central location by using AWS Identity and Access Management (IAM) policies. This allows you to control which users can access each instance, including the option to provide non-root access to specified users. Once access is provided, you can audit which user accessed an instance and log each command to Amazon S3 or Amazon CloudWatch Logs using AWS CloudTrail.

AWS Trusted Advisor

AWS Trusted Advisor is an online resource to help you reduce cost, increase performance, and improve security by optimizing your AWS environment. Trusted Advisor provides real-time guidance to help you provision your resources following AWS best practices.

AWS User Notifications

AWS User Notifications provides a central location for managing your AWS notifications. You can receive notifications from AWS services, such as AWS Health events, Amazon CloudWatch alarms, or EC2 instance state changes, in a consistent, human-friendly format. These notifications can be delivered in multiple ways including the Console Notification Center (default), email, Amazon Q Developer in chat applications, AWS Console Mobile Application push notifications, or through the User Notifications API.

AWS Well-Architected Tool

The AWS Well-Architected Tool (AWS WA Tool) helps you review the state of your workloads and compares them to the latest AWS architectural best practices. A workload is defined as any set of components that deliver business value, which could be an application or website. The tool is based on the AWS Well-Architected Framework, developed to help cloud architects build secure, high-performing, resilient, efficient, and sustainable application infrastructure.

The Framework provides a consistent approach for customers and partners to evaluate architectures. It has been used in tens of thousands of workload reviews conducted by the AWS Solutions Architecture team and by customers, and provides guidance to help implement designs that scale with application needs over time.

To use the AWS WA Tool, available in the AWS Management Console at no charge, just define your workload and answer a set of questions regarding operational excellence, security, reliability, performance efficiency, cost optimization, and sustainability. The AWS WA Tool then provides a plan on how to architect for the cloud using established best practices.

Media

AWS offers the most purpose-built media services, software, and appliances of any cloud to make creating, transforming, and delivering digital content fast and easy.

For general information, see Media Services on AWS.

Services

- Amazon Elastic Transcoder
- Amazon Interactive Video Service
- Amazon Nimble Studio
- AWS Elemental Appliances and Software
- AWS Elemental MediaConnect
- AWS Elemental MediaConvert
- AWS Elemental MediaLive
- AWS Elemental MediaPackage
- AWS Elemental MediaStore
- AWS Elemental MediaTailor

Amazon Elastic Transcoder

Amazon Elastic Transcoder is media transcoding in the cloud. It is designed to be a highly scalable, easy-to-use, and cost-effective way for developers and businesses to convert (or transcode) media files from their source format into versions that will play back on devices such as smartphones, tablets, and PCs.

Amazon Interactive Video Service

Amazon Interactive Video Service (Amazon IVS) is a managed live streaming solution that is quick and easy to set up, and ideal for creating interactive video experiences. Send your live streams to Amazon IVS using streaming software and the service does everything you need to make low-latency live video available to any viewer around the world, letting you focus on building interactive experiences alongside the live video. You can easily customize and enhance the audience experience through the Amazon IVS player SDK and timed metadata APIs, allowing you to build a more valuable relationship with your viewers on your own websites and applications.

Amazon Nimble Studio

Amazon Nimble Studio empowers creative studios to produce visual effects, animation, and interactive content entirely in the cloud, from storyboard sketch to final deliverable. Rapidly onboard and collaborate with artists globally and create content faster with access to virtual workstations, high-speed storage, and scalable rendering across the AWS global infrastructure.

AWS Elemental Appliances and Software

AWS Elemental Appliances and Software solutions bring advanced video processing and delivery technologies into your data center, co-location space, or on-premises facility. You can deploy AWS Elemental Appliances and Software to encode, package, and deliver video assets on-premises and seamlessly connect with cloud-based video infrastructure. Designed for easy integration with AWS Cloud media solutions, AWS Elemental Appliances and Software support video workloads that need to remain on-premises to accommodate physical camera and router interfaces, managed network delivery, or network bandwidth constraints.

AWS Elemental Live, AWS Elemental Server, and AWS Elemental Conductor come in two variants: ready-to-deploy appliances, or AWS-licensed software that you install on your own hardware. AWS Elemental Link is a compact hardware device that sends live video to the cloud for encoding and delivery to viewers.

AWS Elemental MediaConnect

AWS Elemental MediaConnect is a high-quality transport service for live video. Today, broadcasters and content owners rely on satellite networks or fiber connections to send their high-value content into the cloud or to transmit it to partners for distribution. Both satellite and fiber approaches are expensive, require long lead times to set up, and lack the flexibility to adapt to changing requirements. To be more nimble, some customers have tried to use solutions that transmit live video on top of IP infrastructure, but have struggled with reliability and security.

Now you can get the reliability and security of satellite and fiber combined with the flexibility, agility, and economics of IP-based networks using AWS Elemental MediaConnect. MediaConnect enables you to build mission-critical live video workflows in a fraction of the time and cost of satellite or fiber services. You can use MediaConnect to ingest live video from a remote event site (such as a stadium), share video with a partner (such as a cable TV distributor), or replicate a video stream for processing (such as an over-the-top service). MediaConnect combines reliable video transport, highly secure stream sharing, and real-time network traffic and video monitoring that allow you to focus on your content, not your transport infrastructure.

AWS Elemental MediaConvert

AWS Elemental MediaConvert is a file-based video transcoding service with broadcast-grade features. It allows you to easily create video-on-demand (VOD) content for broadcast and multiscreen delivery at scale. The service combines advanced video and audio capabilities with a simple web services interface and pay-as-you-go pricing. With AWS Elemental MediaConvert, you can focus on delivering compelling media experiences without having to worry about the complexity of building and operating your own video processing infrastructure.

AWS Elemental MediaLive

AWS Elemental MediaLive is a broadcast-grade live video processing service. It lets you create high-quality video streams for delivery to broadcast televisions and internet-connected multiscreen devices, such as connected TVs, tablets, smart phones, and set-top boxes. The service works by encoding your live video streams in real-time, taking a larger-sized live video source and compressing it into smaller versions for distribution to your viewers. With AWS Elemental MediaLive, you can easily set up streams for both live events and 24x7 channels with advanced broadcasting features, high availability, and pay-as-you-go pricing. AWS Elemental MediaLive lets you focus on creating compelling live video experiences for your viewers without the complexity of building and operating broadcast-grade video processing infrastructure.

AWS Elemental MediaPackage

AWS Elemental MediaPackage reliably prepares and protects your video for delivery over the Internet. From a single video input, AWS Elemental MediaPackage creates video streams formatted to play on connected TVs, mobile phones, computers, tablets, and game consoles. It makes it easy to implement popular video features for viewers (start-over, pause, rewind, and so on), such as those commonly found on DVRs. AWS Elemental MediaPackage can also protect your content using Digital Rights Management (DRM). AWS Elemental MediaPackage scales automatically in response to load, so your viewers will always get a great experience without you having to accurately predict in advance the capacity you'll need.

AWS Elemental MediaStore

AWS Elemental MediaStore is an AWS storage service optimized for media. It gives you the performance, consistency, and low latency required to deliver live streaming video content. AWS Elemental MediaStore acts as the origin store in your video workflow. Its high performance

capabilities meet the needs of the most demanding media delivery workloads, combined with long-term, cost-effective storage.

AWS Elemental MediaTailor

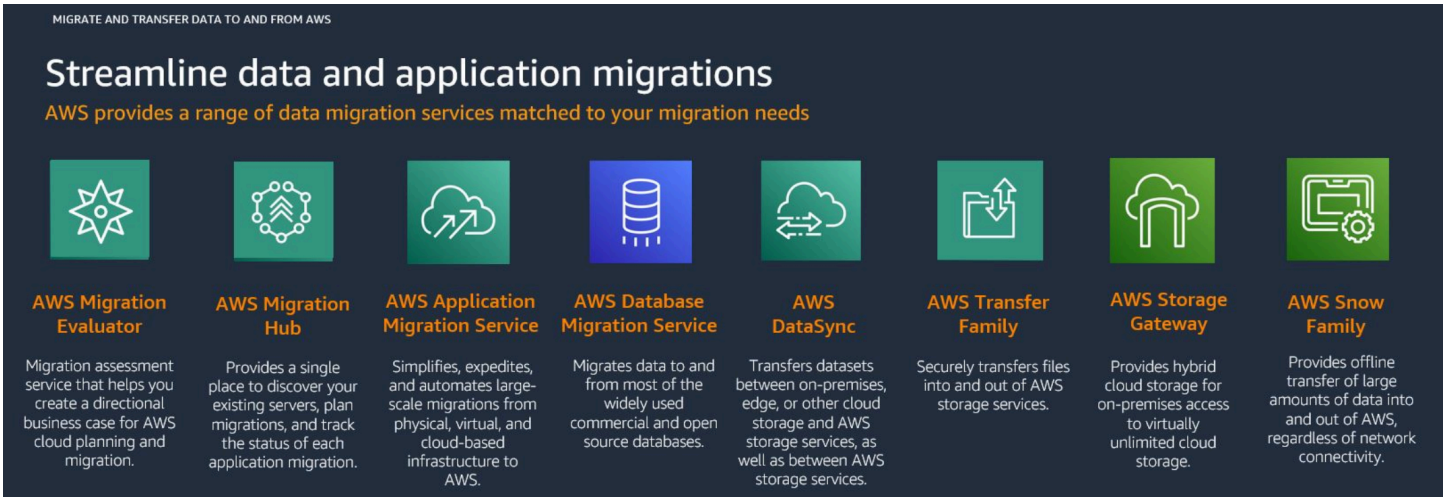
AWS Elemental MediaTailor lets video providers insert individually targeted advertising into their video streams without sacrificing broadcast-level quality-of-service. With AWS Elemental MediaTailor, viewers of your live or on-demand video each receive a stream that combines your content with ads personalized to them. But unlike other personalized ad solutions, with AWS Elemental MediaTailor your entire stream – video and ads – is delivered with broadcast-grade video quality to improve the experience for your viewers. AWS Elemental MediaTailor delivers automated reporting based on both client and server-side ad delivery metrics, making it easy to accurately measure ad impressions and viewer behavior. You can easily monetize unexpected high-demand viewing events with no up-front costs using AWS Elemental MediaTailor. It also improves ad delivery rates, helping you make more money from every video, and it works with a wider variety of content delivery networks, ad decision servers, and client devices.

Also refer to Amazon Kinesis Video Streams

Migration and transfer

AWS offers a wide range of migration tools, guidance, services, and programs to help you assess, migrate and modernize applications and data from building the business case to leveraging AWS services to deliver new experiences.

Each service is described after the diagram. To help you decide which service best meets your needs, see Choosing AWS migration services and tools. For general information, see Migrate and Modernize on AWS.



Services and tools

- AWS Application Discovery Service
- AWS Application Migration Service
- AWS Database Migration Service
- AWS Mainframe Modernization Service
- AWS Migration Hub
- AWS Snow Family
- AWS DataSync
- AWS Transfer Family

AWS Application Discovery Service

AWS Application Discovery Service helps enterprise customers plan migration projects by gathering information about their on-premises data centers.

Planning data center migrations can involve thousands of workloads that are often deeply interdependent. Server utilization data and dependency mapping are important early first steps in the migration process. AWS Application Discovery Service collects and presents configuration, usage, and behavior data from your servers to help you better understand your workloads.

The collected data is retained in encrypted format in an AWS Application Discovery Service data store. You can export this data as a CSV file and use it to estimate the Total Cost of Ownership (TCO) of running on AWS and to plan your migration to AWS. In addition, this data is also available

in AWS Migration Hub, where you can migrate the discovered servers and track their progress as they get migrated to AWS.

AWS Application Migration Service

AWS Application Migration Service (AWS MGN) allows you to quickly realize the benefits of migrating applications to the cloud without changes and with minimal downtime.

AWS Application Migration Service minimizes time-intensive, error-prone manual processes by automatically converting your source servers from physical, virtual, or cloud infrastructure to run natively on AWS. It further simplifies your migration by enabling you to use the same automated process for a wide range of applications.

And by launching non-disruptive tests before migrating, you can be confident that your most critical applications such as SAP, Oracle, and SQL Server will work seamlessly on AWS.

AWS Database Migration Service

AWS Database Migration Service (AWS DMS) helps you migrate databases to AWS easily and securely. The source database remains fully operational during the migration, minimizing downtime to applications that rely on the database. The AWS Database Migration Service can migrate your data to and from most widely used commercial and open-source databases. The service supports homogeneous migrations such as Oracle to Oracle, as well as heterogeneous migrations between different database platforms, such as Oracle to Amazon Aurora or Microsoft SQL Server to MySQL. It also allows you to stream data to Amazon Redshift from any of the supported sources including Amazon Aurora, PostgreSQL, MySQL, MariaDB, Oracle, SAP ASE, and SQL Server, enabling consolidation and easy analysis of data in the petabyte-scale data warehouse. AWS Database Migration Service can also be used for continuous data replication with high availability.

AWS DMS Serverless offers the flexibility to migrate data without needing to provision replication instances, manually monitoring use, and adjusting capacity. AWS DMS Serverless supports popular use cases including continuous data replication, database consolidation, and migrations, even if the source and target database engines differ. For like-to-like or compatible database engines, you can use built-in tools with automatic scaling for a seamless database migration.

AWS Mainframe Modernization Service

AWS Mainframe Modernization Service is a unique service that allows you to migrate your on-premises mainframe workloads to a managed runtime environment on AWS. AWS Mainframe

Modernization Service is a set of managed tools providing infrastructure and software for migrating, modernizing, and running mainframe applications.

- Migrate and modernize your applications to remove the hardware and staffing costs of traditional mainframes.
- Break up and manage your complete migration with infrastructure, software, and tools to refactor and transform legacy applications.
- Deploy, run, and operate migrated applications in the Mainframe Modernization environment with no upfront costs.

AWS Migration Hub

AWS Migration Hub provides a single location to track the progress of application migrations across multiple AWS and partner solutions. Using Migration Hub allows you to choose the AWS and partner migration tools that best fit your needs, while providing visibility into the status of migrations across your portfolio of applications. Migration Hub also provides key metrics and progress for individual applications, regardless of which tools are being used to migrate them. For example, you might use AWS Database Migration Service, AWS Application Migration Service, and partner migration tools such as ATADATA ATAmotion, CloudEndure Live Migration, or RiverMeadow Server Migration Saas to migrate an application comprised of a database, virtualized web servers, and a bare metal server. Using Migration Hub, you can view the migration progress of all the resources in the application. This allows you to quickly get progress updates across all of your migrations, easily identify and troubleshoot any issues, and reduce the overall time and effort spent on your migration projects.

AWS Snow Family

The AWS Snow Family helps customers that need to run operations in austere, non-data center environments, and in locations where there's lack of consistent network connectivity. The Snow Family comprises AWS Snowball and AWS Snowball Edge, and offers a number of physical devices and capacity points, most with built-in computing capabilities. These services help physically transport up to exabytes of data into and out of AWS. Snow Family devices are owned and managed by AWS and integrate with AWS security, monitoring, storage management, and computing capabilities.

AWS Snowball

AWS Snowball is the smallest member of the AWS Snow Family of edge computing, edge storage, and data transfer devices, weighing in at 4.5 pounds (2.1 kg) with 8 terabytes of usable storage. The Snowball appliance is ruggedized, secure, and purpose-built for use outside of a traditional data center. Its small form factor makes it a perfect fit for tight spaces or where portability is a necessity and network connectivity is unreliable. You can use Snowball in backpacks on first responders, or for Internet of Things (IoT), vehicular, and drone use cases. You can run compute applications at the edge, and you can ship the device with data to AWS for offline data transfer, or you can transfer data online with AWS DataSync from edge locations.

Like AWS Snowball Edge, AWS Snowball has multiple layers of security and encryption. You can use either of these services to run edge computing workloads, or to collect, process, and transfer data to AWS. Snowball is designed for data migration needs up to 8 terabytes per device and from space-constrained environments where Snowball Edge devices will not fit.

AWS Snowball Edge

AWS Snowball Edge is an edge computing, data migration, and edge storage device. Snowball Edge can do local processing and run edge-computing workloads in addition to transferring data between your local environment and the AWS Cloud. Each Snowball Edge device can transport data at speeds faster than the internet. This transport is done by shipping the data in the devices through a regional carrier.

Snowball Edge devices have five options for device configurations:

- **Storage-optimized** for data transfer, with up to 80 TB of usable storage capacity. They are well suited for local storage and large scale data transfer.
- **Storage-optimized 210 TB**, with 210 TB of usable storage capacity
- **Storage-optimized with EC2-compatible compute functionality**, with up to 80 TB of usable storage capacity, 40 vCPUs, and 80 GB of memory for compute functionality
- **Compute-optimized**, with the AMD EPYC Gen2 having the most compute functionality with up to 104 vCPUs, 416 GB of memory, and 28 TB of dedicated NVMe SSD for compute instances. The AMD EPYC Gen1 has up to 52 vCPUs, 208 GB of memory, 39.5 TB of usable storage capacity, and 7.68 TB of dedicated NVMe SSD for compute instances.

  You can use these devices for data collection, machine learning (ML) and processing, and storage in environments with intermittent connectivity (such as manufacturing, industrial, and

transportation) or in extremely remote locations (such as military or maritime operations) before shipping them back to AWS.

- **Compute-optimized with GPU** is identical to the compute-optimized AMD EPYC Gen1 option, but also includes an installed graphics processing unit (GPU). The GPU is equivalent to the one available in the P3 Amazon EC2-compatible instance type. You can use these devices for advanced ML workloads and full motion video analysis in disconnected environments.

These devices can also be rack mounted and clustered together to build larger temporary installations.

Snowball supports specific Amazon EC2 instance types and AWS Lambda functions, so you can develop and test in the AWS Cloud, then deploy applications on devices in remote locations to collect, pre-process, and ship the data to AWS. Common use cases include data migration, data transport, image collation, IoT sensor stream capture, and ML.

AWS DataSync

AWS DataSync is a data transfer service that makes it easy for you to automate moving data between on-premises storage and Amazon S3 or Amazon Elastic File System (Amazon EFS). DataSync automatically handles many of the tasks related to data transfers that can slow down migrations or burden your IT operations, including running your own instances, handling encryption, managing scripts, network optimization, and data integrity validation. You can use DataSync to transfer data at speeds up to 10 times faster than open-source tools. DataSync uses an on-premises software agent to connect to your existing storage or file systems using the Network File System (NFS) protocol, so you don't have write scripts or modify your applications to work with AWS APIs. You can use DataSync to copy data over AWS Direct Connect or internet links to AWS. The service enables one-time data migrations, recurring data processing workflows, and automated replication for data protection and recovery. Getting started with DataSync is easy: Deploy the DataSync agent on premises, connect it to a file system or storage array, select Amazon EFS or Amazon S3 as your AWS storage, and start moving data. You pay only for the data you copy.

AWS Transfer Family

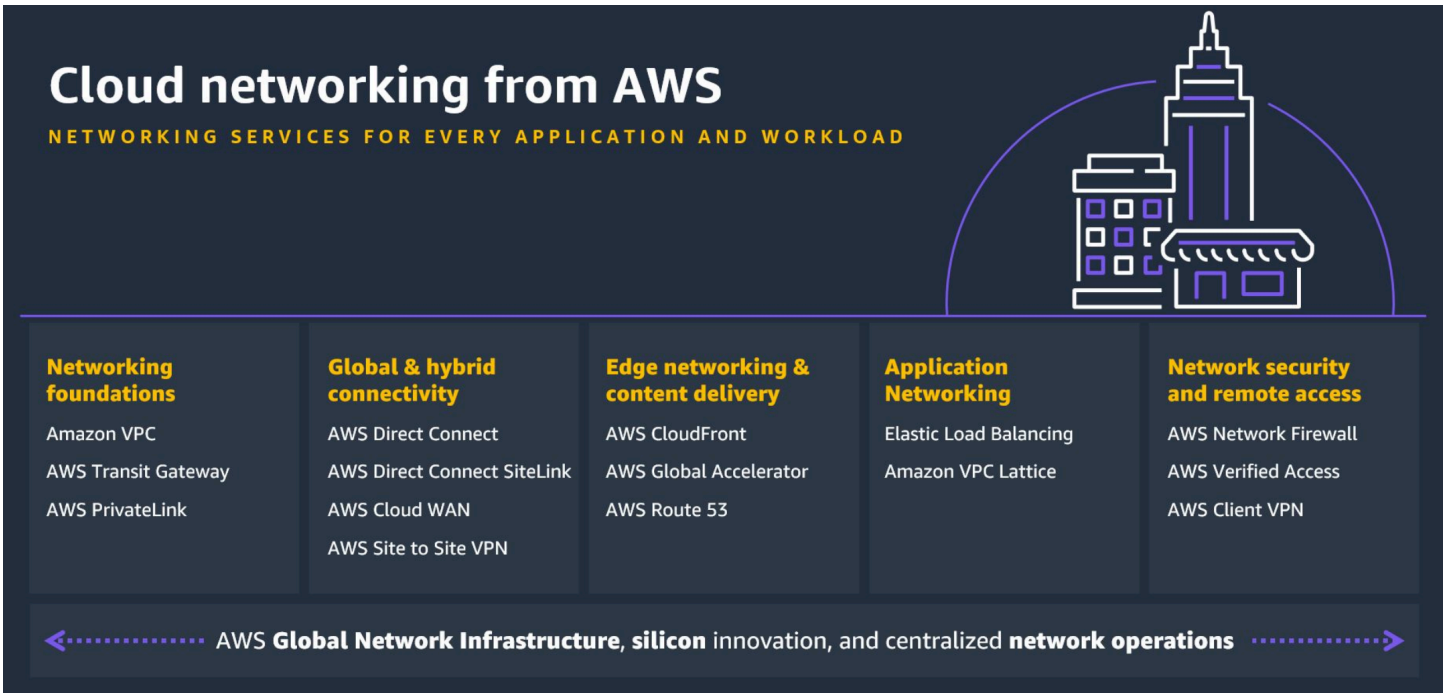
AWS Transfer Family provides fully managed support for file transfers directly into and out of Amazon S3 or Amazon EFS. With support for Secure File Transfer Protocol (SFTP), File Transfer Protocol over SSL (FTPS), and File Transfer Protocol (FTP), the AWS Transfer Family helps you seamlessly migrate your file transfer workflows to AWS by integrating with existing authentication

systems, and providing DNS routing with Amazon Route 53 so nothing changes for your customers and partners, or their applications. With your data in Amazon S3 or Amazon EFS, you can use it with AWS services for processing, analytics, ML, archiving, as well as home directories and developer tools. Getting started with the AWS Transfer Family is easy; there is no infrastructure to buy and set up.

Networking and content delivery

AWS offers a broad set of networking and content delivery services that provide the highest level of reliability, security, and performance in the cloud.

Each service is described after the diagram. To help you decide which service best meets your needs, see Choosing an AWS networking and content delivery service. For general information, see AWS Networking and Content Delivery.



Services

- Amazon API Gateway
- Amazon CloudFront
- Amazon Route 53

- AWS Verified Access
- Amazon VPC
- Amazon VPC Lattice
- AWS App Mesh
- AWS Cloud Map
- AWS Direct Connect
- AWS Global Accelerator
- AWS PrivateLink
- AWS Private 5G
- AWS Transit Gateway
- AWS VPN
- Elastic Load Balancing
- Integrated Private Wireless on AWS

Amazon API Gateway

Amazon API Gateway is a fully managed service that makes it easy for developers to create, publish, maintain, monitor, and secure APIs at any scale. With a few clicks in the AWS Management Console, you can create an API that acts as a "front door" for applications to access data, business logic, or functionality from your back-end services, such as workloads running on Amazon EC2, code running on AWS Lambda, or any web application. Amazon API Gateway handles all the tasks involved in accepting and processing up to hundreds of thousands of concurrent API calls, including traffic management, authorization and access control, monitoring, and API version management.

Amazon CloudFront

Amazon CloudFront is a fast content delivery network (CDN) service that securely delivers data, videos, applications, and APIs to customers globally with low latency, high transfer speeds, all within a developer-friendly environment. CloudFront is integrated with AWS – both physical locations that are directly connected to the AWS global infrastructure, as well as other AWS services. CloudFront works seamlessly with services including AWS Shield for DDoS mitigation, Amazon S3, Elastic Load Balancing or Amazon EC2 as origins for your applications, and Lambda@Edge to run custom code closer to customers' users and to customize the user experience.

You can get started with the content delivery network in minutes, using the same AWS tools that you're already familiar with: APIs, AWS Management Console, AWS CloudFormation, CLIs, and SDKs. Amazon CDN offers a simple, pay-as-you-go pricing model with no upfront fees or required long-term contracts, and support for the CDN is included in your existing Support subscription.

Amazon Route 53

Amazon Route 53 is a highly available and scalable cloud Domain Name System (DNS) web service. It is designed to give developers and businesses an extremely reliable and cost-effective way to route users to internet applications by translating human-readable names, such as `www.example.com`, into the numeric IP addresses, such as `192.0.2.1`, that computers use to connect to each other. Amazon Route 53 is fully compliant with IPv6 as well.

Amazon Route 53 effectively connects user requests to infrastructure running in AWS—such as EC2 instances, elastic load balancers, or Amazon S3 buckets—and can also be used to route users to infrastructure outside of AWS. You can use Amazon Route 53 to configure DNS health checks to route traffic to healthy endpoints or to independently monitor the health of your application and its endpoints.

Amazon Route 53 traffic flow makes it easy for you to manage traffic globally through a variety of routing types, including latency-based routing, Geo DNS, and weighted round robin—all of which can be combined with DNS Failover in order to enable a variety of low-latency, fault-tolerant architectures. Using Amazon Route 53 traffic flow's simple visual editor, you can easily manage how your end users are routed to your application's endpoints—whether in a single AWS Region or distributed around the globe. Amazon Route 53 also offers Domain Name Registration—you can purchase and manage domain names such as `example.com` and Amazon Route 53 will automatically configure DNS settings for your domains.

AWS Verified Access

AWS Verified Access provides corporate users secure access to your applications without using a virtual private network (VPN). Based on AWS Zero Trust principles, Verified Access evaluates each application request in real time to help ensure that users can only access your applications after they meet specified security requirements. You can group applications, or define unique access policies for each application, with conditions based on user identity and device posture data.

Amazon VPC

Amazon Virtual Private Cloud (Amazon VPC) lets you provision a logically isolated section of the AWS Cloud where you can launch AWS resources in a virtual network that you define. You have complete control over your virtual networking environment, including selection of your own IP address range, creation of subnets, and configuration of route tables and network gateways. You can use both IPv4 and IPv6 in your VPC for secure and easy access to resources and applications.

You can easily customize the network configuration for your VPC. For example, you can create a public-facing subnet for your web servers that has access to the Internet, and place your backend systems, such as databases or application servers, in a private-facing subnet with no Internet access. You can leverage multiple layers of security (including security groups and network access control lists) to help control access to EC2 instances in each subnet.

Additionally, you can create a hardware virtual private network (VPN) connection between your corporate data center and your VPC and leverage the AWS Cloud as an extension of your corporate data center.

Amazon VPC Lattice

Amazon VPC Lattice provides fully managed support for service-to-service connectivity and communication. With VPC Lattice, you can use policies to define network traffic management, access, and monitoring to connect compute services in a simplified and secure way across instances, containers, and serverless applications.

AWS App Mesh

AWS App Mesh makes it easy to monitor and control microservices running on AWS. App Mesh standardizes how your microservices communicate, giving you end-to-end visibility and helping to ensure high-availability for your applications.

Modern applications are often composed of multiple microservices that each perform a specific function. This architecture helps to increase the availability and scalability of the application by allowing each component to scale independently based on demand, and automatically degrading functionality when a component fails instead of going offline. Each microservice interacts with all the other microservices through an API. As the number of microservices grows within an application, it becomes increasingly difficult to pinpoint the exact location of errors, re-route traffic after failures, and safely deploy code changes. Previously, this has required you to build monitoring

and control logic directly into your code and redeploy your microservices every time there are changes.

AWS App Mesh makes it easy to run microservices by providing consistent visibility and network traffic controls for every microservice in an application. App Mesh removes the need to update application code to change how monitoring data is collected or traffic is routed between microservices. App Mesh configures each microservice to export monitoring data and implements consistent communications control logic across your application. This makes it easy to quickly pinpoint the exact location of errors and automatically re-route network traffic when there are failures or when code changes need to be deployed.

You can use App Mesh with Amazon ECS and Amazon EKS to better run containerized microservices at scale. App Mesh uses the open source Envoy proxy, making it compatible with a wide range of AWS partner and open source tools for monitoring microservices.

AWS Cloud Map

AWS Cloud Map is a cloud resource discovery service. With AWS Cloud Map, you can define custom names for your application resources, and it maintains the updated location of these dynamically changing resources. This increases your application availability because your web service always discovers the most up-to-date locations of its resources.

Modern applications are typically composed of multiple services that are accessible over an API and perform a specific function. Each service interacts with a variety of other resources such as databases, queues, object stores, and customer-defined microservices, and they also need to be able to find the location of all the infrastructure resources on which it depends, in order to function. You typically manually manage all these resource names and their locations within the application code. However, manual resource management becomes time consuming and error-prone as the number of dependent infrastructure resources increases or the number of microservices dynamically scale up and down based on traffic. You can also use third-party service discovery products, but this requires installing and managing additional software and infrastructure.

AWS Cloud Map allows you to register any application resources such as databases, queues, microservices, and other cloud resources with custom names. AWS Cloud Map then constantly checks the health of resources to make sure the location is up-to-date. The application can then query the registry for the location of the resources needed based on the application version and deployment environment.

AWS Direct Connect

AWS Direct Connect makes it easy to establish a dedicated network connection from your premises to AWS. Using AWS Direct Connect, you can establish private connectivity between AWS and your data center, office, or co-location environment, which in many cases can reduce your network costs, increase bandwidth throughput, and provide a more consistent network experience than Internet-based connections.

AWS Direct Connect lets you establish a dedicated network connection between your network and one of the AWS Direct Connect locations. Using industry standard 802.1Q virtual LANs (VLANs), this dedicated connection can be partitioned into multiple virtual interfaces. This allows you to use the same connection to access public resources, such as objects stored in Amazon S3 using public IP address space, and private resources such as EC2 instances running within a VPC using private IP address space, while maintaining network separation between the public and private environments. Virtual interfaces can be reconfigured at any time to meet your changing needs.

AWS Global Accelerator

AWS Global Accelerator is a networking service that improves the availability and performance of the applications that you offer to your global users.

Today, if you deliver applications to your global users over the public internet, your users might face inconsistent availability and performance as they traverse through multiple public networks to reach your application. These public networks are often congested and each hop can introduce availability and performance risk. AWS Global Accelerator uses the highly available and congestion-free AWS global network to direct internet traffic from your users to your applications on AWS, making your users' experience more consistent.

To improve the availability of your application, you must monitor the health of your application endpoints and route traffic only to healthy endpoints. AWS Global Accelerator improves application availability by continuously monitoring the health of your application endpoints and routing traffic to the closest healthy endpoints.

AWS Global Accelerator also makes it easier to manage your global applications by providing static IP addresses that act as a fixed entry point to your application hosted on AWS which eliminates the complexity of managing specific IP addresses for different AWS Regions and Availability Zones. AWS Global Accelerator is easy to set up, configure and manage.

AWS PrivateLink

AWS PrivateLink simplifies the security of data shared with cloud-based applications by eliminating the exposure of data to the public Internet. AWS PrivateLink provides private connectivity between VPCs, AWS services, and on-premises applications, securely on the Amazon network. AWS PrivateLink makes it easy to connect services across different accounts and VPCs to significantly simplify the network architecture.

AWS Private 5G

AWS Private 5G offers an easy way to use cellular technology to augment your current network. This can help you increase reliability, extend coverage, or allow a new class of workloads, such as factory automation, autonomous robotics, and advanced augmented and virtual reality (AR/VR). You will receive all the Private 5G hardware (including SIM cards) and software you need to deploy your private cellular network and connect devices to your applications.

With a few clicks in the AWS Management Console, deploy a private cellular network that meets your connectivity requirements. Start by specifying the connectivity requirements for the desired location, the number of devices you want to connect, and the geographic area they will cover. AWS will deliver pre-integrated hardware and software components (from both AWS and our AWS Partners) that meet the enterprise connectivity requirements of your private network. AWS delivers and maintains the small cell radio units, servers, 5G core, radio access network (RAN) software, and SIM cards required to set up a private 5G network and connect devices. Once the equipment is powered on, AWS automatically configures and deploys the cellular network. All you need to do is insert the SIM cards into your devices.

AWS Private 5G is also integrated with AWS Identity and Access Management (IAM), which helps you securely access and manage AWS services and resources, including all devices connected to your Private 5G network. Private 5G manages and maintains all the software and hardware components to deliver reliable, predictable network behavior and on-demand scaling to accommodate any number of devices and sensors.

AWS Transit Gateway

AWS Transit Gateway is a service that enables customers to connect their Amazon Virtual Private Clouds (VPCs) and their on-premises networks to a single gateway. As you grow the number of workloads running on AWS, you need to be able to scale your networks across multiple accounts and Amazon VPCs to keep up with the growth. Today, you can connect pairs of Amazon VPCs using

peering. However, managing point-to-point connectivity across many Amazon VPCs, without the ability to centrally manage the connectivity policies, can be operationally costly and cumbersome. For on-premises connectivity, you need to attach your AWS VPN to each individual Amazon VPC. This solution can be time consuming to build and hard to manage when the number of VPCs grows into the hundreds.

With AWS Transit Gateway, you only have to create and manage a single connection from the central gateway in to each Amazon VPC, on-premises data center, or remote office across your network. Transit Gateway acts as a hub that controls how traffic is routed among all the connected networks which act like spokes. This hub and spoke model significantly simplifies management and reduces operational costs because each network only has to connect to the Transit Gateway and not to every other network. Any new VPC is simply connected to the Transit Gateway and is then automatically available to every other network that is connected to the Transit Gateway. This ease of connectivity makes it easy to scale your network as you grow.

AWS VPN

AWS Virtual Private Network (AWS VPN) solutions establish secure connections between your on-premises networks, remote offices, client devices, and the AWS global network. AWS VPN is comprised of two services: AWS Site-to-Site VPN and AWS Client VPN. Each service provides a highly-available, managed, and elastic cloud VPN solution to protect your network traffic.

AWS Site-to-Site VPN creates encrypted tunnels between your network and your Amazon Virtual Private Clouds or AWS Transit Gateways. For managing remote access, AWS Client VPN connects your users to AWS or on-premises resources using a VPN software client.

Elastic Load Balancing

Elastic Load Balancing (ELB) automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, and IP addresses. It can handle the varying load of your application traffic in a single Availability Zone or across multiple Availability Zones. Elastic Load Balancing offers four types of load balancers that all feature the high availability, automatic scaling, and robust security necessary to make your applications fault tolerant.

- Application Load Balancer is best suited for load balancing of HTTP and HTTPS traffic and provides advanced request routing targeted at the delivery of modern application architectures, including microservices and containers. Operating at the individual request level (Layer seven), Application Load Balancer routes traffic to targets within Amazon Virtual Private Cloud (Amazon VPC) based on the content of the request.

- Network Load Balancer is best suited for load balancing of TCP traffic where extreme performance is required. Operating at the connection level (Layer four), Network Load Balancer routes traffic to targets within Amazon Virtual Private Cloud (Amazon VPC) and is capable of handling millions of requests per second while maintaining ultra-low latencies. Network Load Balancer is also optimized to handle sudden and volatile traffic patterns.
- Gateway Load Balancer makes it easy to deploy, scale, and run third-party virtual networking appliances. Providing load balancing and auto scaling for fleets of third-party appliances, Gateway Load Balancer is transparent to the source and destination of traffic. This capability makes it well suited for working with third-party appliances for security, network analytics, and other use cases.
- Classic Load Balancer provides basic load balancing across multiple Amazon EC2 instances and operates at both the request level and connection level. Classic Load Balancer is intended for applications that were built within the EC2-Classic network. EC2-Classic was retired on August 15, 2022.

Integrated Private Wireless on AWS

The Integrated Private Wireless on AWS program is designed to provide enterprises with managed and validated private wireless offerings from leading Communications Service Providers (CSPs). The offerings integrate CSPs' private 5G and 4G LTE wireless networks with AWS services across AWS Regions, AWS Local Zones, AWS Outposts, and AWS Snow Family. AWS Telco Solutions Architects technically validate the offerings for their sound architecture, and adherence to AWS best practices. Telecom companies deliver, operate, and support the offerings.

The program also uses the rich expertise of validated global AWS Independent Software Vendor (ISV) partners to accelerate the time-to-value for private wireless deployment. Integrated Private Wireless on AWS removes the long planning cycles and complex integrations usually required to set up and scale a private wireless network. You can now deploy a secure, reliable, and low-latency private wireless network to power AI/ML and IoT workloads at the edge and at scale.

Quantum technologies

Amazon Braket

Amazon Braket is a fully managed quantum computing service that helps researchers and developers get started with the technology to accelerate research and discovery. Amazon Braket provides a development environment for you to explore and build quantum algorithms, test them on quantum circuit simulators, and run them on different quantum hardware technologies.

Quantum computing has the potential to solve computational problems that are beyond the reach of classical computers by harnessing the laws of quantum mechanics to process information in new ways. This approach to computing could transform areas such as chemical engineering, material science, drug discovery, financial portfolio optimization, and machine learning. But defining those problems and programming quantum computers to solve them requires new skills, which are difficult to acquire without easy access to quantum computing hardware.

Amazon Braket overcomes these challenges so you can explore quantum computing. With Amazon Braket, you can design and build your own quantum algorithms from scratch or choose from a set of pre-built algorithms. Once you have built your algorithm, Amazon Braket provides a choice of simulators to test, troubleshoot and run your algorithms. When you are ready, you can run your algorithm on your choice of different quantum computers, and gate-based computers from Rigetti and IonQ. With Amazon Braket, you can now evaluate the potential of quantum computing for your organization, and build expertise.

Robotics

AWS RoboMaker

AWS RoboMaker is a service that makes it easy to develop, test, and deploy intelligent robotics applications at scale. AWS RoboMaker extends the most widely used open-source robotics software framework, Robot Operating System (ROS), with connectivity to cloud services. This includes AWS machine learning services, monitoring services, and analytics services that enable a robot to stream data, navigate, communicate, comprehend, and learn. AWS RoboMaker provides a robotics development environment for application development, a

robotics simulation service to accelerate application testing, and a robotics fleet management service for remote application deployment, update, and management.

Robots are machines that sense, compute, and take action. Robots need instructions to accomplish tasks, and these instructions come in the form of applications that developers code to determine how the robot will behave. Receiving and processing sensor data, controlling actuators for movement, and performing a specific task are all functions that are typically automated by these intelligent robotics applications. Intelligent robots are being increasingly used in warehouses to distribute inventory, in homes to carry out tedious housework, and in retail stores to provide customer service. Robotics applications use machine learning in order to perform more complex tasks like recognizing an object or face, having a conversation with a person, following a spoken command, or navigating autonomously.

Until now, developing, testing, and deploying intelligent robotics applications was difficult and time consuming. Building intelligent robotics functionality using machine learning is complex and requires specialized skills. Setting up a development environment can take each developer days and building a realistic simulation system to test an application can take months due to the underlying infrastructure needed. Once an application has been developed and tested, a developer needs to build a deployment system to deploy the application into the robot and later update the application while the robot is in use.

AWS RoboMaker provides you with the tools to make building intelligent robotics applications more accessible, a fully managed simulation service for quick and easy testing, and a deployment service for lifecycle management. AWS RoboMaker removes the heavy lifting from each step of robotics development so you can focus on creating innovative robotics applications.

Satellite

AWS Ground Station

AWS Ground Station is a fully managed service that lets you control satellite communications, downlink and process satellite data, and scale your satellite operations quickly, easily and cost-effectively without having to worry about building or managing your own ground station

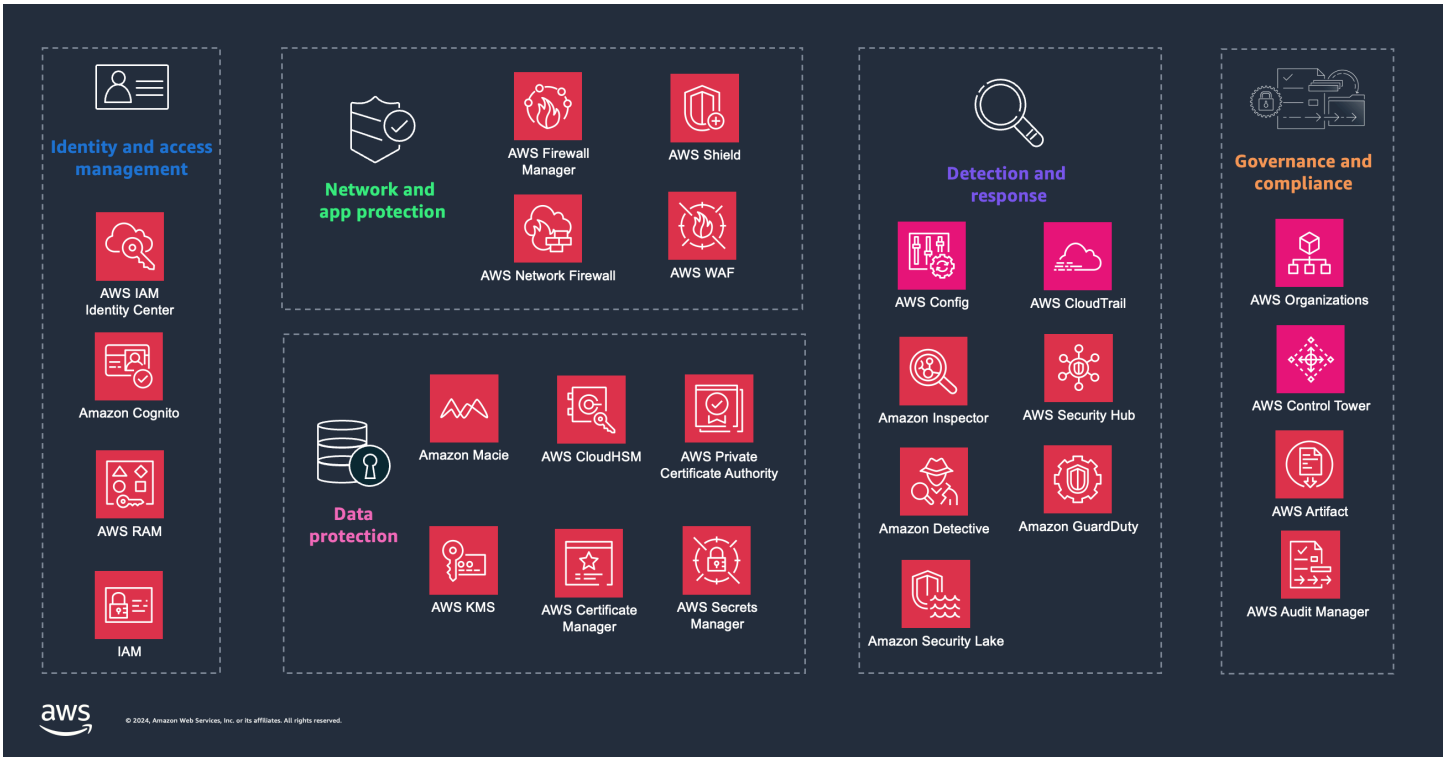
infrastructure. Satellites are used for a wide variety of use cases, including weather forecasting, surface imaging, communications, and video broadcasts. Ground stations are at the core of global satellite networks, which are facilities that provide communications between the ground and the satellites by using antennas to receive data and control systems to send radio signals to command and control the satellite. Today, you must either build your own ground stations and antennas, or obtain long-term leases with ground station providers, often in multiple countries to provide enough opportunities to contact the satellites as they orbit the globe. Once all this data is downloaded, you need servers, storage, and networking in close proximity to the antennas to process, store, and transport the data from the satellites.

AWS Ground Station eliminates these problems by delivering a global ground station as a service. We provide direct access to AWS services and the AWS Global Infrastructure including our low-latency global fiber network right where your data is downloaded into our AWS Ground Station. This enables you to easily control satellite communications, quickly ingest and process your satellite data, and rapidly integrate that data with your applications and other services running in the AWS Cloud. For example, you can use Amazon S3 to store the downloaded data, Amazon Kinesis Data Streams for managing data ingestion from satellites, SageMaker AI for building custom machine learning applications that apply to your data sets, and Amazon EC2 to command and download data from satellites. AWS Ground Station can help you save up to 80% on the cost of your ground station operations by allowing you to pay only for the actual antenna time used, and relying on our global footprint of ground stations to download data when and where you need it, instead of building and operating your own global ground station infrastructure. There are no long-term commitments, and you gain the ability to rapidly scale your satellite communications on-demand when your business needs it.

Security, identity, and compliance

AWS is architected to be the most secure global cloud infrastructure on which to build, migrate, and manage applications and workloads.

Each service is described after the diagram. To help you decide which service best meets your needs, see Choosing AWS security, identity, and governance services. For general information, see Security, Identity, and Compliance on AWS.



Services

- Amazon Cognito
- Amazon Detective
- Amazon GuardDuty
- Amazon Inspector
- Amazon Macie
- Amazon Security Lake
- Amazon Verified Permissions
- AWS Artifact
- AWS Audit Manager
- AWS Certificate Manager
- AWS CloudHSM
- AWS Directory Service
- AWS Firewall Manager
- AWS Identity and Access Management
- AWS Key Management Service

- AWS Network Firewall
- AWS Resource Access Manager
- AWS Secrets Manager
- AWS Security Hub
- AWS Shield
- AWS IAM Identity Center
- AWS WAF
- AWS WAF Captcha

Amazon Cognito

Amazon Cognito lets you add user sign-up, sign-in, and access control to your web and mobile apps quickly and easily. With Amazon Cognito, you can scale to millions of users and supports sign-in with social identity providers such as Apple, Facebook, Twitter, or Amazon, with SAML 2.0 identity solutions, or by using your own identity system.

In addition, Amazon Cognito enables you to save data locally on users' devices, allowing your applications to work even when the devices are offline. You can then synchronize data across users' devices so that their app experience remains consistent regardless of the device they use.

With Amazon Cognito, you can focus on creating great app experiences instead of worrying about building, securing, and scaling a solution to handle user management, authentication, and sync across devices.

Amazon Detective

Amazon Detective makes it easy to analyze, investigate, and quickly identify the root cause of potential security issues or suspicious activities. Amazon Detective automatically collects log data from your AWS resources and uses machine learning, statistical analysis, and graph theory to build a linked set of data that enables you to easily conduct faster and more efficient security investigations. Amazon Detective further simplifies account management for security operations and investigations across all existing and future accounts in an organization using AWS Organizations for up to 1,200 AWS accounts.

AWS security services such as Amazon GuardDuty, Amazon Macie, and AWS Security Hub, as well as partner security products, can be used to identify potential security issues, or findings. These

services are really helpful in alerting you when and where there is possible unauthorized access or suspicious behavior in your AWS deployment. However, sometimes there are security findings that you would like to perform deeper investigations of the events that led to the findings to remediate the root cause. Determining the root cause of security findings can be a complex process for security analysts that often involves collecting and combining logs from many data sources, using extract, transform, and load (ETL) tools, and custom scripting to organize the data.

Amazon Detective simplifies this process by enabling your security teams to easily investigate and quickly get to the root cause of a finding. Detective can analyze trillions of events from multiple data sources such as Amazon Virtual Private Cloud (VPC) Flow Logs, AWS CloudTrail, and Amazon GuardDuty. Detective uses these events to automatically create a unified, interactive view of your resources, users, and the interactions between them over time. With this unified view, you can visualize all the details and context in one place to identify the underlying reasons for the findings, drill down into relevant historical activities, and quickly determine the root cause.

You can get started with Amazon Detective in just a few clicks in the AWS Management Console. There is no software to deploy, or data sources to enable and maintain. You can try Detective at no additional charge with a 30-day free trial that is available to new accounts.

Amazon GuardDuty

Amazon GuardDuty is a threat detection service that continuously monitors for malicious activity and anomalous behavior to protect your AWS accounts, workloads, Kubernetes clusters, and data stored in Amazon Simple Storage Service (Amazon S3). The GuardDuty service monitors for activity such as unusual API calls, unauthorized deployments, and exfiltrated credentials that indicate a possible account reconnaissance or compromise.

Enabled with a few clicks in the AWS Management Console and easily administrated organization-wide with its support of AWS Organizations, Amazon GuardDuty can immediately begin analyzing billions of events across your AWS accounts for signs of unauthorized use. GuardDuty identifies suspected attackers through integrated threat intelligence feeds and machine learning anomaly detection to detect anomalies in account and workload activity. When potential unauthorized use is detected, the service delivers a detailed finding to the GuardDuty console, Amazon CloudWatch Events, and AWS Security Hub. This makes findings actionable and easy to integrate into existing event management and workflow systems. Further investigation to determine the root cause of a finding is easily accomplished by using Amazon Detective directly from the GuardDuty console.

Amazon GuardDuty is cost effective and easy to operate. It does not require you to deploy and maintain software or security infrastructure, meaning it can be enabled quickly with no risk of

negatively impacting existing application and container workloads. There are no upfront costs with GuardDuty, no software to deploy, and no threat intelligence feeds to enable. Furthermore, GuardDuty optimizes costs by applying smart filters and analyzing only a subset of logs relevant to threat detection, and new Amazon GuardDuty accounts are free for 30 days.

Amazon Inspector

Amazon Inspector is a new automated vulnerability management service that continually scans AWS workloads for software vulnerabilities and unintended network exposure. With a few clicks in the AWS Management Console and AWS Organizations, Amazon Inspector can be used across all accounts in your organization. Once started, Amazon Inspector automatically discovers running Amazon Elastic Compute Cloud (Amazon EC2) instances and container images residing in Amazon Elastic Container Registry (Amazon ECR), at any scale, and immediately starts assessing them for known vulnerabilities.

Amazon Inspector has many improvements over Amazon Inspector Classic. For example, the new Amazon Inspector calculates a highly contextualized risk score for each finding by correlating common vulnerabilities and exposures (CVE) information with factors such as network access and exploitability. This score is used to prioritize the most critical vulnerabilities to improve remediation response efficiency. Additionally, Amazon Inspector now uses the widely deployed AWS Systems Manager Agent (SSM Agent) to eliminate the need for you to deploy and maintain a standalone agent to run Amazon EC2 instance assessments. For container workloads, Amazon Inspector is now integrated with Amazon Elastic Container Registry (Amazon ECR) to support intelligent, cost-efficient, and continual vulnerability assessments of container images. All findings are aggregated in the Amazon Inspector console, routed to AWS Security Hub, and pushed through Amazon EventBridge to automate workflows such as ticketing.

All accounts new to Amazon Inspector are eligible for a 15-day free trial to evaluate the service and estimate its cost. During the trial, all eligible Amazon EC2 instances and container images pushed to Amazon ECR are continually scanned at no cost.

Amazon Macie

Amazon Macie is a fully managed data security and data privacy service that uses inventory evaluations, machine learning, and pattern matching to discover sensitive data and accessibility in your Amazon S3 environment. Macie supports scalable on-demand and automated sensitive data discovery jobs that automatically tracks changes to the bucket and only evaluates new or modified objects over time. Using Macie, you can detect a large and growing list of sensitive data

types for many countries and Regions, including multiple types of financial data, personal health information (PHI), and personally identifiable information (PII), as well as custom types. Macie also continually evaluates your Amazon S3 environment to provide an S3 resource summary and security evaluation across all of your accounts. You can search, filter, and sort S3 buckets by metadata variables, such as bucket names, tags, and security controls like encryption status or public accessibility. For any unencrypted buckets, publicly accessible buckets, or buckets shared with AWS accounts outside those you have defined in AWS Organizations, you can be alerted to act.

In the multi-account configuration, a single Macie administrator account can manage all member accounts, including the creation and administration of sensitive data discovery jobs across accounts with AWS Organizations. Security and sensitive data discovery findings are aggregated in the Macie administrator account and sent to Amazon CloudWatch Events and AWS Security Hub. Now using one account, you can integrate with event management, workflow, and ticketing systems or use Macie findings with AWS Step Functions to automate remediation actions. You can quickly get started with Macie using the 30-day trial available to new accounts for S3 bucket inventory and bucket-level evaluation at no charge. Sensitive data discovery is not included in the 30-day trial for bucket evaluation.

Amazon Security Lake

Amazon Security Lake centralizes security data from AWS environments, SaaS providers, on premises, and cloud sources, into a purpose-built data lake that's stored in your AWS account. Security Lake automates the collection and management of security data across accounts and AWS Regions so that you can use your preferred analytics tools while retaining control and ownership over your security data. With Security Lake, you can also improve the protection of your workloads, applications, and data.

Security Lake automates the collection of security-related log and event data from integrated AWS services and third-party services. It also helps you manage the lifecycle of data with customizable retention settings. The data lake is backed by Amazon S3 buckets, and you retain ownership over your data. Security Lake converts ingested data into Apache Parquet format and a standard open-source schema called the Open Cybersecurity Schema Framework (OCSF). With OCSF support, Security Lake normalizes and combines security data from AWS and a broad range of enterprise security data sources.

Other AWS services and third-party services can subscribe to the data that's stored in Security Lake for incident response and security data analytics.

Amazon Verified Permissions

Amazon Verified Permissions is a scalable, fine-grained permissions management and authorization service for custom applications you've built. Verified Permissions enables your developers to build secure applications faster by externalizing authorization and centralizing policy management and administration.

Verified Permissions uses Cedar, an open-source policy language and SDK, to define fine-grained permissions for application users. Your authorization model is defined using principal types, resource types, and valid actions, to control who can take what actions on which resources in a given application context. Policy changes are audited so that you can see who made the changes and when.

AWS Artifact

AWS Artifact is your go-to, central resource for compliance-related information that matters to you. It provides on-demand access to AWS security and compliance reports and select online agreements. Reports available in AWS Artifact include our Service Organization Control (SOC) reports, Payment Card Industry (PCI) reports, and certifications from accreditation bodies across geographies and compliance verticals that validate the implementation and operating effectiveness of AWS security controls. Agreements available in AWS Artifact include the Business Associate Addendum (BAA) and the Nondisclosure Agreement (NDA).

AWS Audit Manager

AWS Audit Manager helps you continuously audit your AWS usage to simplify how you assess risk and compliance with regulations and industry standards. Audit Manager automates evidence collection to reduce the “all hands on deck” manual effort that often happens for audits and enable you to scale your audit capability in the cloud as your business grows. With Audit Manager, it is easy to assess if your policies, procedures, and activities – also known as controls – are operating effectively. When it is time for an audit, AWS Audit Manager helps you manage stakeholder reviews of your controls and enables you to build audit-ready reports with much less manual effort.

The AWS Audit Manager prebuilt frameworks help translate evidence from cloud services into auditor-friendly reports by mapping your AWS resources to the requirements in industry standards or regulations, such as CIS AWS Foundations Benchmark, the General Data Protection Regulation (GDPR), and the Payment Card Industry Data Security Standard (PCI DSS). You can also fully customize a framework and its controls for your unique business requirements. Based on the

framework you select, Audit Manager launches an assessment that continuously collects and organizes relevant evidence from your AWS accounts and resources, such as resource configuration snapshots, user activity, and compliance check results.

You can get started quickly in the AWS Management Console. Just select a prebuilt framework to launch an assessment and begin automatically collecting and organizing evidence.

AWS Certificate Manager

AWS Certificate Manager is a service that lets you easily provision, manage, and deploy Secure Sockets Layer/Transport Layer Security (SSL/TLS) certificates for use with AWS services and your internal connected resources. SSL/TLS certificates are used to secure network communications and establish the identity of websites over the Internet as well as resources on private networks. AWS Certificate Manager removes the time-consuming manual process of purchasing, uploading, and renewing SSL/TLS certificates.

With AWS Certificate Manager, you can quickly request a certificate, deploy it on ACM-integrated AWS resources, such as Elastic Load Balancing, Amazon CloudFront distributions, and APIs on API Gateway, and let AWS Certificate Manager handle certificate renewals. It also enables you to create private certificates for your internal resources and manage the certificate lifecycle centrally. Public and private certificates provisioned through AWS Certificate Manager for use with ACM-integrated services are free. You pay only for the AWS resources you create to run your application.

With AWS Private Certificate Authority, you pay monthly for the operation of the private certificate authority (CA) and for the private certificates you issue. you have a highly available private CA service without the upfront investment and ongoing maintenance costs of operating your own private CA.

AWS CloudHSM

The AWS CloudHSM is a cloud-based hardware security module (HSM) that enables you to easily generate and use your own encryption keys on the AWS Cloud. With AWS CloudHSM, you can manage your own encryption keys using dedicated FIPS 140-2 Level 3 validated HSMs. AWS CloudHSM offers you the flexibility to integrate with your applications using industry-standard APIs, such as PKCS#11, Java Cryptography Extensions (JCE), and Microsoft CryptoNG (CNG) libraries.

AWS CloudHSM is standards-compliant and enables you to export all of your keys to most other commercially-available HSMs, subject to your configurations. It is a fully managed service that

automates time-consuming administrative tasks for you, such as hardware provisioning, software patching, high-availability, and backups. AWS CloudHSM also enables you to scale quickly by adding and removing HSM capacity on-demand, with no up-front costs.

AWS Directory Service

AWS Directory Service for Microsoft Active Directory, also known as AWS Managed Microsoft AD, enables your directory-aware workloads and AWS resources to use managed Active Directory in the AWS Cloud. AWS Managed Microsoft AD is built on actual Microsoft Active Directory and does not require you to synchronize or replicate data from your existing Active Directory to the cloud. You can use standard Active Directory administration tools and take advantage of built-in Active Directory features such as Group Policy and single sign-on (SSO). With AWS Managed Microsoft AD, you can easily join Amazon EC2 and Amazon RDS for SQL Server instances to a domain, and use AWS Enterprise IT applications such as Amazon WorkSpaces with Active Directory users and groups.

AWS Firewall Manager

AWS Firewall Manager is a security management service which allows you to centrally configure and manage firewall rules across your accounts and applications in AWS Organizations. As new applications are created, Firewall Manager makes it easy to bring new applications and resources into compliance by enforcing a common set of security rules. Now you have a single service to build firewall rules, create security policies, and enforce them in a consistent, hierarchical manner across your entire infrastructure, from a central administrator account.

AWS Identity and Access Management

AWS Identity and Access Management (IAM) enables you to securely control access to AWS services and resources for your AWS users, groups, and roles. Using IAM, you can create and manage fine-grained access controls with permissions, specify who can access which services and resources, and under which conditions. IAM allows you to do the following:

- You manage AWS permissions for your workforce users and workloads in AWS IAM Identity Center (IAM Identity Center). IAM Identity Center allows you to manage user access across multiple AWS accounts. With just a few clicks, you can enable a highly available service, easily manage multi-account access and the permissions to all of your accounts in AWS Organizations centrally. IAM Identity Center includes built-in SAML integrations to many business applications, such as Salesforce, Box, and Microsoft Office 365. Further, you can create Security Assertion Markup Language (SAML) 2.0 integrations and extend single sign-on access to any of your SAML-

enabled applications. Your users simply sign in to a user portal with credentials they configure or using their existing corporate credentials to access all their assigned accounts and applications from one place.

- Manage single-account IAM permissions: You can specify access to AWS resources using permissions. Your IAM entities (users, groups, and roles) by default start with no permissions. These identities can be granted permissions by attaching an IAM policy that specifies the type of access, the actions that can be performed, and the resources on which actions can be performed. You can also specify conditions that must be set for access to be allowed or denied.
- Manage single-account IAM roles: IAM roles allows you to delegate access to users or services that normally don't have access to your organization's AWS resources. IAM users or AWS services can assume a role to obtain a temporary security credential that be used to make AWS API calls. You don't have to share long-term credentials or define permissions for each identity.

AWS Key Management Service

AWS Key Management Service (AWS KMS) makes it easy for you to create and manage cryptographic keys and control their use across a wide range of AWS services and in your applications. AWS KMS uses hardware security modules (HSM) to protect and validate your AWS KMS keys under the FIPS 140-2 Cryptographic Module Validation Program. AWS KMS is integrated with AWS CloudTrail to provide you with logs of all key usage to help meet your regulatory and compliance needs.

AWS Network Firewall

AWS Network Firewall is a managed service that makes it easy to deploy essential network protections for all of your Amazon Virtual Private Clouds (VPCs). The service can be setup with just a few clicks and scales automatically with your network traffic, so you don't have to worry about deploying and managing any infrastructure. The AWS Network Firewall flexible rules engine lets you define firewall rules that give you fine-grained control over network traffic, such as blocking outbound Server Message Block (SMB) requests to prevent the spread of malicious activity. You can also import rules you've already written in common open source rule formats as well as enable integrations with managed intelligence feeds sourced by AWS Partners. AWS Network Firewall works together with AWS Firewall Manager so you can build policies based on AWS Network Firewall rules and then centrally apply those policies across your VPCs and accounts.

AWS Network Firewall includes features that provide protections from common network threats. The AWS Network Firewall stateful firewall can incorporate context from traffic flows, such as

tracking connections and protocol identification, to enforce policies such as preventing your VPCs from accessing domains using an unauthorized protocol. The AWS Network Firewall intrusion prevention system (IPS) provides active traffic flow inspection so you can identify and block vulnerability exploits using signature-based detection. AWS Network Firewall also offers web filtering that can stop traffic to known bad URLs and monitor fully qualified domain names.

It's easy to get started with AWS Network Firewall by visiting the Amazon VPC Console to create or import your firewall rules, group them into policies, and apply them to the VPCs you want to protect. AWS Network Firewall pricing is based on the number of firewalls deployed and the amount of traffic inspected. There are no upfront commitments and you pay only for what you use.

AWS Resource Access Manager

AWS Resource Access Manager (AWS RAM) helps you securely share your resources across AWS accounts, within your organization or organizational units (OUs) in AWS Organizations, and with IAM roles and IAM users for supported resource types. You can use AWS RAM to share transit gateways, subnets, AWS License Manager license configurations, Amazon Route 53 Resolver rules, and more resource types.

Many organizations use multiple accounts to create administrative or billing isolation, and to limit the impact of errors. With AWS RAM, you don't need to create duplicate resources in multiple AWS accounts. This reduces the operational overhead of managing resources in every account that you own. Instead, in your multi-account environment, you can create a resource once, and use AWS RAM to share that resource across accounts by creating a resource share. When you create a resource share, you select the resources to share, choose an AWS RAM managed permission per resource type, and specify whom you want to have access to the resources. AWS RAM is available to you at no additional charge.

AWS Secrets Manager

AWS Secrets Manager helps you protect secrets needed to access your applications, services, and IT resources. The service enables you to easily rotate, manage, and retrieve database credentials, API keys, and other secrets throughout their lifecycle. Users and applications retrieve secrets with a call toSecrets Manager APIs, eliminating the need to hardcode sensitive information in plain text. Secrets Manager offers secret rotation with built-in integration for Amazon RDS, Amazon Redshift, and Amazon DocumentDB. The service is also extensible to other types of secrets, including API keys and OAuth tokens. In addition, Secrets Manager enables you to control access to secrets using fine-grained permissions and audit secret rotation centrally for resources in the AWS Cloud, third-party services, and on-premises.

AWS Security Hub

AWS Security Hub is a cloud security posture management service that performs automated, continuous security best practice checks against your AWS resources. Security Hub aggregates your security alerts (i.e. findings) from various AWS services and partner products in a standardized format so that you can more easily take action on them. To maintain a complete view of your security posture in AWS, you need to integrate multiple tools and services including threat detections from Amazon GuardDuty, vulnerabilities from Amazon Inspector, sensitive data classifications from Amazon Macie, resource configuration issues from AWS Config, and AWS Partner Network products. Security Hub simplifies how you understand and improve your security posture with automated security best practice checks powered by AWS Config rules and automated integrations with dozens of AWS services and partner products.

Security Hub enables you to understand your overall security posture via a consolidated security score across all of your AWS accounts, automatically assesses the security of your AWS accounts resources via the AWS Foundational Security Best Practices (FSBP) standard and other compliance frameworks. It also aggregates all of your security findings from dozens of AWS security services and APN products in a single place and format via the AWS Security Finding Format (ASFF), and reduces your Mean Time To Remediation (MTTR) with automated response and remediation support. Security Hub has out-of-the-box integrations with ticketing, chat, Security Information and Event Management (SIEM), Security Orchestration Automation and Response (SOAR), threat investigation, Governance Risk and Compliance (GRC), and incident management tools to provide your users with a complete security operations workflow.

Getting started with Security Hub requires just a few clicks from the AWS Management Console to begin aggregating findings and conducting security checks using our 30-day free trial. You can integrate Security Hub with AWS Organizations to automatically enable the service in all accounts in your organization.

AWS Shield

AWS Shield is a managed Distributed Denial of Service (DDoS) protection service that safeguards web applications running on AWS. AWS Shield provides you with always-on detection and automatic inline mitigations that minimize application downtime and latency, so there is no need to engage Support to benefit from DDoS protection. There are two tiers of AWS Shield: Standard and Advanced.

All AWS customers benefit from the automatic protections of AWS Shield Standard, at no additional charge. AWS Shield Standard defends against most common, frequently occurring

network and transport layer DDoS attacks that target your website or applications. When you use AWS Shield Standard with Amazon CloudFront and Amazon Route 53, you receive comprehensive availability protection against all known infrastructure (Layer 3 and 4) attacks.

For higher levels of protection against attacks targeting your applications running on Amazon Elastic Compute Cloud (Amazon EC2), Elastic Load Balancing (ELB), Amazon CloudFront, and Amazon Route 53 resources, you can subscribe to AWS Shield Advanced. In addition to the network and transport layer protections that come with Standard, AWS Shield Advanced provides additional detection and mitigation against large and sophisticated DDoS attacks, near real-time visibility into attacks, and integration with AWS WAF, a web application firewall. AWS Shield Advanced also gives you 24x7 access to the AWS DDoS Response Team (DRT) and protection against DDoS related spikes in your Amazon Elastic Compute Cloud (Amazon EC2), Elastic Load Balancing (ELB), Amazon CloudFront, and Amazon Route 53 charges.

AWS Shield Advanced is available globally on all Amazon CloudFront and Amazon Route 53 edge locations. You can protect your web applications hosted anywhere in the world by deploying Amazon CloudFront in front of your application. Your origin servers can be Amazon S3, Amazon Elastic Compute Cloud (Amazon EC2), Elastic Load Balancing (ELB), or a custom server outside of AWS. You can also enable AWS Shield Advanced directly on an Elastic IP or Elastic Load Balancing (ELB) in the following AWS Regions: Northern Virginia, Ohio, Oregon, Northern California, Montreal, São Paulo, Ireland, Frankfurt, London, Paris, Stockholm, Singapore, Tokyo, Sydney, Seoul, Mumbai, Milan, and Cape Town.

AWS IAM Identity Center

AWS IAM Identity Center (SSO) is a cloud SSO service that makes it easy to centrally manage SSO access to multiple AWS accounts and business applications. With just a few clicks, you can enable a highly available SSO service without the upfront investment and on-going maintenance costs of operating your own SSO infrastructure. With IAM Identity Center, you can easily manage SSO access and user permissions to all of your accounts in AWS Organizations centrally. IAM Identity Center also includes built-in SAML integrations to many business applications, such as Salesforce, Box, and Microsoft Office 365. Further, by using the IAM Identity Center application configuration wizard, you can create Security Assertion Markup Language (SAML) 2.0 integrations and extend SSO access to any of your SAML-enabled applications. Your users simply sign in to a user portal with credentials they configure in IAM Identity Center or using their existing corporate credentials to access all their assigned accounts and applications from one place.

AWS WAF

AWS WAF is a web application firewall that helps protect your web applications or APIs against common web exploits and bots that may affect availability, compromise security, or consume excessive resources. AWS WAF gives you control over how traffic reaches your applications by enabling you to create security rules that control bot traffic and block common attack patterns, such as SQL injection or cross-site scripting. You can also customize rules that filter out specific traffic patterns. You can get started quickly using Managed Rules for AWS WAF, a pre-configured set of rules managed by AWS or AWS Marketplace sellers to address issues like the OWASP Top 10 security risks and automated bots that consume excess resources, skew metrics, or can cause downtime. These rules are regularly updated as new issues emerge. AWS WAF includes a full-featured API that you can use to automate the creation, deployment, and maintenance of security rules.

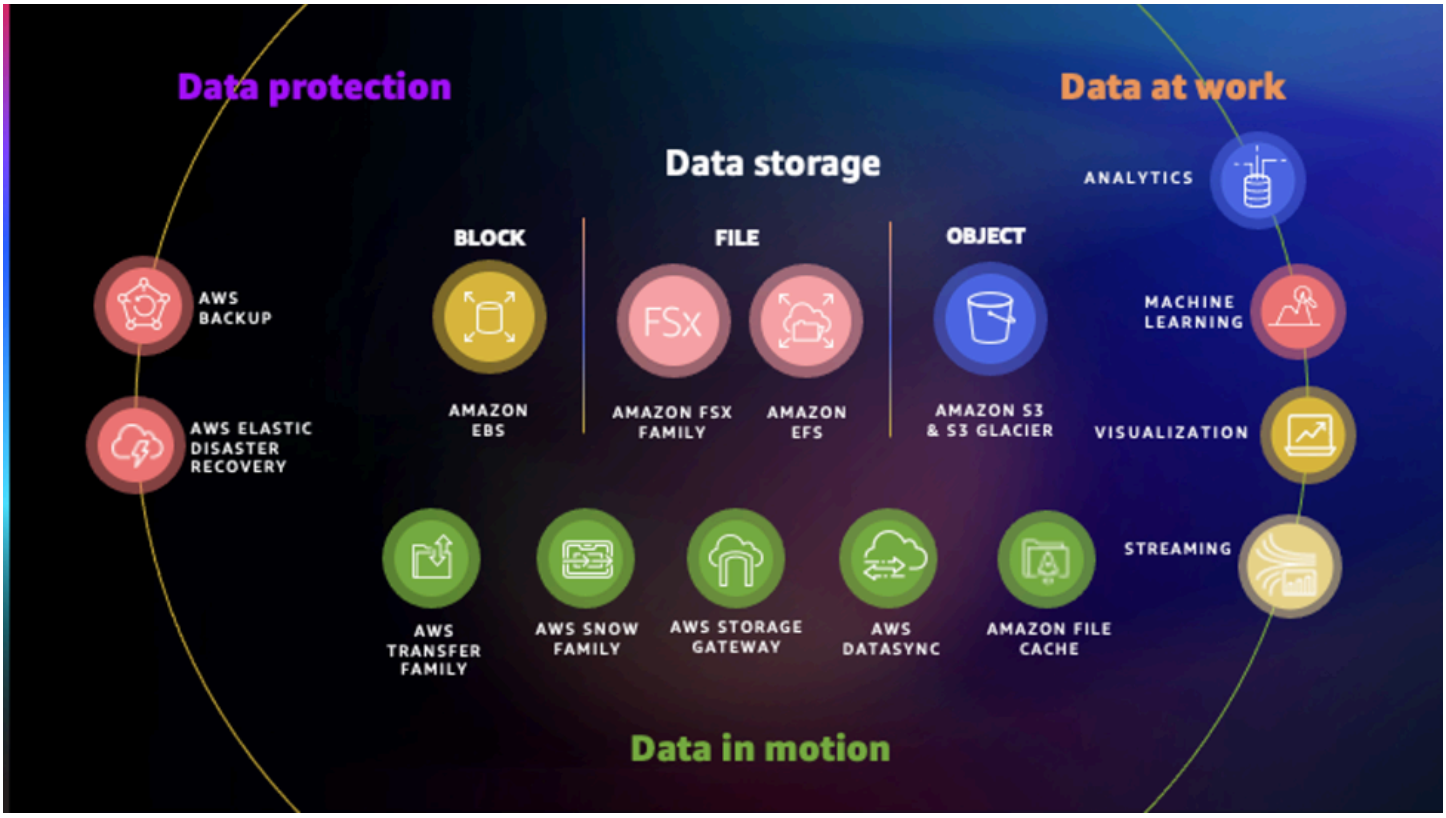
AWS WAF Captcha

AWS WAF Captcha helps block unwanted bot traffic by requiring users to successfully complete challenges before their web request are allowed to reach AWS WAF protected resources. You can configure AWS WAF rules to require WAF Captcha challenges to be solved for specific resources that are frequently targeted by bots such as login, search, and form submissions. You can also require WAF Captcha challenges for suspicious requests based on the rate, attributes, or labels generated from AWS Managed Rules, such as AWS WAF Bot Control or the Amazon IP Reputation list. WAF Captcha challenges are simple for humans while remaining effective against bots. WAF Captcha includes an audio version and is designed to meet Web Content Accessability Guidelines (WCAG) accessibility requirements.

Storage

AWS provides a broad portfolio of storage services with deep functionality for storing, accessing, protecting, and analyzing your data.

Each service is described after the diagram. To help you decide which service best meets your needs, see Choosing an AWS storage service. For general information, see Cloud Storage on AWS.



Services

- AWS Backup
- Amazon Elastic Block Store
- AWS Elastic Disaster Recovery
- Amazon Elastic File System
- Amazon File Cache
- Amazon FSx for Lustre
- Amazon FSx for NetApp ONTAP
- Amazon FSx for OpenZFS
- Amazon FSx for Windows File Server
- Amazon Simple Storage Service
- AWS Storage Gateway

AWS Backup

AWS Backup enables you to centralize and automate data protection across AWS services. AWS Backup offers a cost-effective, fully managed, policy-based service that further simplifies data protection at scale. AWS Backup also helps you support your regulatory compliance or business policies for data protection. Together with AWS Organizations, AWS Backup enables you to centrally deploy data protection policies to configure, manage, and govern your backup activity across your organization's AWS accounts and resources, including Amazon Elastic Compute Cloud (Amazon EC2) instances, Amazon Elastic Block Store (Amazon EBS) volumes, Amazon Relational Database Service (Amazon RDS) databases (including Amazon Aurora clusters), Amazon DynamoDB tables, Amazon Elastic File System (Amazon EFS) file systems, Amazon FSx for Lustre file systems, Amazon FSx for Windows File Server file systems, and AWS Storage Gateway volumes.

Amazon Elastic Block Store

Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure, offering high availability and durability. Amazon EBS volumes offer the consistent and low-latency performance needed to run your workloads. With Amazon EBS, you can scale your usage up or down within minutes—all while paying a low price for only what you provision.

AWS Elastic Disaster Recovery

AWS Elastic Disaster Recovery (Elastic Disaster Recovery) minimizes downtime and data loss with fast, reliable recovery of on-premises and cloud-based applications using affordable storage, minimal compute, and point-in-time recovery. You can configure replication and launch settings, monitor data replication, and launch instances for drills or recovery.

Set up Elastic Disaster Recovery on your source servers to initiate secure data replication. Your data is replicated to a staging area subnet in your AWS account, in the AWS Region that you select. You can perform non-disruptive tests to confirm that implementation is complete. During normal operation, maintain readiness by monitoring replication and periodically performing non-disruptive recovery and failback drills.

If you must replicate to the AWS China Regions or perform replication and recovery into AWS Outposts, use CloudEndure Disaster Recovery available in the AWS Marketplace.

Amazon Elastic File System

Amazon Elastic File System (Amazon EFS) provides a simple, scalable, elastic file system for Linux-based workloads for use with AWS Cloud services and on-premises resources. It is built to scale on demand to petabytes without disrupting applications, growing and shrinking automatically as you add and remove files, so your applications have the storage they need – when they need it. It is designed to provide massively parallel shared access to thousands of Amazon EC2 instances, enabling your applications to achieve high levels of aggregate throughput and IOPS with consistent low latencies. Amazon EFS is a fully managed service that requires no changes to your existing applications and tools, providing access through a standard file system interface for seamless integration. Amazon EFS is a regional service storing data within and across multiple Availability Zones (AZs) for high availability and durability. You can access your file systems across Availability Zones and AWS Regions and share files between thousands of Amazon EC2 instances and on-premises servers via AWS Direct Connect or AWS VPN.

Amazon EFS is well suited to support a broad spectrum of use cases from highly parallelized, scale-out workloads that require the highest possible throughput to single-threaded, latency-sensitive workloads. Use cases such as lift-and-shift enterprise applications, big data analytics, web serving and content management, application development and testing, media and entertainment workflows, database backups, and container storage.

For long-lived data that is accessed only a few times a year or less, consider Amazon EFS Archive, a cost-effective way to retain even your coldest data so that it's always available to power new business insights. Amazon EFS Archive supports the same intelligent tiering experience as existing EFS storage classes. This means that you can combine the sub-millisecond SSD latencies of Amazon EFS Standard for your active frequently-accessed data with the lower costs of Amazon EFS IA and Amazon EFS Archive for your colder data.

Amazon File Cache

Amazon File Cache is a fully managed high-speed cache on AWS that makes it easier to process file data, regardless of where the data is stored. Amazon File Cache serves as temporary, high-performance storage for data in on-premises file systems, or in file systems or object stores on AWS. The service allows you to make dispersed datasets available to file-based applications on AWS with a unified view and high speeds. You can link the cache to multiple NFS—including on-premises and in-cloud—or Amazon Simple Storage Service (Amazon S3) buckets, providing a unified view of and fast access to your data spanning on-premises and multiple AWS Regions. The

cache provides read and write data access to compute workloads on AWS with sub-millisecond latencies, up to hundreds of GB/s of throughput, and up to millions of IOPS.

Amazon FSx for Lustre

Amazon FSx for Lustre is a fully managed file system that is optimized for compute-intensive workloads, such as high performance computing, machine learning, and media data processing workflows. Many of these applications require the high-performance and low latencies of scale-out, parallel file systems. Operating these file systems typically requires specialized expertise and administrative overhead, requiring you to provision storage servers and tune complex performance parameters. With Amazon FSx, you can launch and run a Lustre file system that can process massive data sets at up to hundreds of gigabytes per second of throughput, millions of IOPS, and sub-millisecond latencies.

Amazon FSx for Lustre is seamlessly integrated with Amazon S3, making it easy to link your long-term data sets with your high performance file systems to run compute-intensive workloads. You can automatically copy data from S3 to Amazon FSx for Lustre, run your workloads, and then write results back to S3. Amazon FSx for Lustre also enables you to burst your compute-intensive workloads from on-premises to AWS by allowing you to access your FSx file system over Amazon Direct Connect or VPN. Amazon FSx for Lustre helps you cost-optimize your storage for compute-intensive workloads: It provides cheap and performant non-replicated storage for processing data, with your long-term data stored durably in Amazon S3 or other low-cost data stores. With Amazon FSx, you pay for only the resources you use. There are no minimum commitments, upfront hardware or software costs, or additional fees.

Amazon FSx for NetApp ONTAP

Amazon FSx for NetApp ONTAP offers the first complete, fully managed NetApp file system available in the cloud making it easy for you to migrate or extend existing applications to AWS without changing code or how you manage your data . Built on NetApp ONTAP, Amazon FSx for NetApp ONTAP provides the familiar features, performance, capabilities, and APIs of NetApp file systems with the agility, scalability, and simplicity of a fully managed AWS service.

Amazon FSx for NetApp ONTAP offers high-performance file storage that is broadly accessible from Linux, Windows, and macOS compute instances via the industry-standard NFS, SMB, and iSCSI protocols. With Amazon FSx for NetApp ONTAP, you get low-cost, fully elastic storage capacity with support for compression and deduplication to help you further reduce storage costs. Amazon FSx for NetApp ONTAP file systems can be deployed and managed using the AWS Management Console or NetApp Cloud Manager for seamless set up and administration.

Amazon FSx for OpenZFS

Amazon FSx for OpenZFS is a fully managed file storage service that lets you launch, run, and scale fully managed file systems built on the open-source OpenZFS file system. Amazon FSx for OpenZFS makes it easy to migrate your on-premises file servers—without changing your applications or how you manage data—and build new high-performance, data-driven applications in the cloud.

Amazon FSx for OpenZFS offers the familiar features, performance, and capabilities of OpenZFS file systems with the agility, scalability, and simplicity of a fully managed AWS service.

Amazon FSx for Windows File Server

Amazon FSx for Windows File Server provides a fully managed native Microsoft Windows file system so you can easily move your Windows-based applications that require file storage to AWS. Built on Windows Server, Amazon FSx provides shared file storage with the compatibility and features that your Windows-based applications rely on, including full support for the SMB protocol and Windows NTFS, Active Directory (AD) integration, and Distributed File System (DFS). Amazon FSx uses SSD storage to provide the fast performance your Windows applications and users expect, with high levels of throughput and IOPS, and consistent sub-millisecond latencies. This compatibility and performance is particularly important when moving workloads that require Windows shared file storage, such as CRM, ERP, and .NET applications, as well as home directories.

With Amazon FSx, you can launch highly durable and available Windows file systems that can be accessed from up to thousands of compute instances using the industry-standard SMB protocol. Amazon FSx eliminates the typical administrative overhead of managing Windows file servers. You pay for only the resources used, with no upfront costs, minimum commitments, or additional fees.

Amazon Simple Storage Service

Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance. This means customers of all sizes and industries can use it to store and protect any amount of data for a range of use cases, such as websites, mobile applications, backup and restore, archive, enterprise applications, IoT devices, and big data analytics. Amazon S3 provides easy-to-use management features so you can organize your data and configure finely-tuned access controls to meet your specific business, organizational, and compliance requirements. Amazon S3 is designed for 99.999999999% (11 9s) of durability, and stores data for millions of applications for companies all around the world.

Amazon S3 storage classes are a range of storage classes that you can choose from based on the data access, resiliency, and cost requirements of your workloads. S3 storage classes are purpose-built to provide the lowest cost storage for different access patterns. S3 storage classes are ideal for virtually any use case, including those with demanding performance needs, data residency requirements, unknown or changing access patterns, or archival storage.

The S3 storage classes include:

- **S3 Intelligent-Tiering** for automatic cost savings for data with unknown or changing access patterns
- **S3 Standard** for frequently accessed data
- **S3 Express One Zone** for your most frequently accessed data
- **S3 Standard-Infrequent Access (S3 Standard-IA)** and **S3 One Zone-Infrequent Access (S3 One Zone-IA)** for less frequently accessed data
- **S3 Glacier Instant Retrieval** for archive data that needs immediate access
- **S3 Glacier Flexible Retrieval** (formerly S3 Glacier) for rarely accessed long-term data that does not require immediate access
- **Amazon S3 Glacier Deep Archive (S3 Glacier Deep Archive)** for long-term archive and digital preservation with retrieval in hours at the lowest cost storage in the cloud

If you have data residency requirements that can't be met by an existing AWS Region, you can use the **S3 Outposts** storage class to store your S3 data on premises. Amazon S3 also offers capabilities to manage your data throughout its lifecycle. Once an S3 Lifecycle policy is set, your data will automatically transfer to a different storage class without any changes to your application. For more information, refer to the Amazon S3 storage classes overview info graphic.

You can use S3 Object Lock to help prevent S3 objects from being deleted or overwritten for a fixed amount of time, or indefinitely. Object Lock can help you to meet regulatory requirements that require WORM (write-once-read-many) storage, or to simply add another layer of protection against object changes or deletion.

AWS Storage Gateway

The AWS Storage Gateway is a hybrid storage service that allows your on-premises applications to seamlessly use AWS cloud storage. You can use the service for backup and archiving, disaster recovery, cloud data processing, storage tiering, and migration. Your applications connect to the service through a virtual machine or hardware gateway appliance using standard storage

protocols, such as NFS, SMB and iSCSI. The gateway connects to AWS storage services, such as Amazon S3, S3 Glacier, and Amazon EBS, and Amazon FSx for Windows File Server, providing storage for files, volumes, and virtual tapes in AWS. The service includes a highly-optimized data transfer mechanism, with bandwidth management, automated network resilience, and efficient data transfer, along with a local cache for low-latency on-premises access to your most active data.

Next steps

Reinvent how you work with IT by signing up for the AWS Free Tier, which allows you to gain hands-on experience with a broad selection of AWS products and services. Within the AWS Free Tier, you can test workloads and run applications to learn more and build the right solution for your organization. You can also contact AWS Sales and Business Development.

By signing up for AWS, you have access to Amazon cloud computing services.

> **Note**
>
> The sign-up process requires a credit card, which will not be charged until you start using services. There are no long-term commitments and you can stop using AWS at any time.

To help familiarize yourself with AWS, check out AWS Skill Builder to explore free, on-demand courses developed by the experts at AWS.

Learn about the breadth and depth of AWS on our general AWS Channel and AWS Online Tech Talks.

Get hands-on experience from our self-paced labs.

Are you Well-Architected?

Explore the AWS Well-Architected Framework, which helps you understand the pros and cons of the decisions you make when building systems on AWS. Using the six pillars of the AWS Well-Architected Framework, you can learn architectural best practices for designing and operating reliable, secure, efficient, cost-effective, and sustainable systems in the cloud.

You can use the AWS Well-Architected Tool, available at no charge in the AWS Management Console, to review your workloads against these best practices by answering a set of questions for each pillar. In addition to the Framework and the AWS WA Tool, specialized guidance is provided for various types of applications.

- In the Serverless Application Lens, we focus on best practices for architecting your serverless applications on AWS.

- In the Container Build Lens, we provide cloud-agnostic best practices for building and managing containers and container images. In addition, implementation guidance and examples are provided specific to the AWS Cloud.
- In the Machine Learning Lens, we focus on how to design, deploy, and architect your machine learning workloads in the AWS Cloud.
- In the Data Analytics Lens, we describe a collection of customer-proven best practices for designing well-architected analytics workloads.
- In the Hybrid Networking Lens, we focus on how to design, deploy, and architect hybrid networking for workloads in the AWS Cloud.
- In the IoT Lens and IoT Lens Checklist, we focus on best practices for architecting your IoT applications on AWS.
- In the SAP Lens, we describe a collection of customer-proven design principles and best practices for ensuring SAP workloads on AWS are well-architected.
- In the Games Industry Lens, we focus on designing, architecting, and deploying your games workloads on AWS.
- In the Streaming Media Lens, we focus on the best practices for architecting and improving your streaming media workloads on AWS.
- In the Healthcare Industry Lens, we focus on how to design, deploy, and manage your healthcare workloads.
- In the Financial Services Industry Lens, we focus on best practices for architecting your Financial Services Industry workloads on AWS.
- In the HPC Lens, we focus on best practices for architecting your High Performance Computing (HPC) workloads on AWS.
- In the SaaS Lens, we focus on best practices for architecting your software as a service (SaaS) workloads on AWS.
- In the Government Lens, we focus on best practices for designing and delivering government services on AWS.
- In the Connected Mobility Lens, we focus on best practices for integrating technology into transportation systems and enhancing the overall mobility experience.
- In the Migration Lens, we provide best practices for how to migrate to the AWS Cloud.

For more expert guidance and best practices for your cloud architecture—reference architecture deployments, diagrams, and whitepapers—refer to the AWS Architecture Center.

Conclusion

AWS provides building blocks that you can assemble quickly to support virtually any workload. With AWS, you'll find a complete set of highly available services that are designed to work together to build sophisticated scalable applications.

You have access to highly durable storage, low-cost compute, high-performance databases, management tools, and more. All this is available without up-front cost, and you pay for only what you use. These services help organizations move faster, lower IT costs, and scale. AWS is trusted by the largest enterprises and the hottest start-ups to power a wide variety of workloads, including web and mobile applications, game development, data processing and warehousing, storage, archive, and many others.

Resources

- AWS Decision Guides
- AWS Architecture Center
- This Is My Architecture videos
- AWS Documentation
- AWS Blog
- AWS Well-Architected Framework
- AWS Whitepapers & Guides

Document history

To be notified about updates to this whitepaper, subscribe to the RSS feed.

Change	Description	Date
Whitepaper updated	Added AWS User Notifications and updated AWS Service Catalog.	June 9, 2025
Whitepaper updated	Added links to decision guides where appropriate.	August 27, 2024
Whitepaper updated	Amazon Q added. Amazon CodeWhisperer is now Amazon Q Developer. Amazon WorkDocs notice added.	May 3, 2024
Whitepaper updated	AWS B2B Data Interchange, AWS re:Post Private, Amazon ElastiCache Serverless, Amazon Neptune Analytics, Amazon RDS for Db2, Amazon PartyRock, Amazon SageMaker AI HyperPod, and Amazon WorkSpaces Thin Client added.	March 1, 2024
Whitepaper updated	AWS Snowball Edge information updated.	February 22, 2024
Whitepaper updated	AWS Elastic Disaster Recovery added, other minor updates.	February 15, 2024

Whitepaper updated	Amazon Managed Grafana and Amazon Managed Service for Prometheus added.	February 5, 2024
Whitepaper updated	New Connected Mobility Lens and Migration Lens added to the Well-Architected section.	February 2, 2024
Whitepaper updated	Amazon Lumberyard is no longer offered. Use Open 3D Engine (O3DE), the Apache-licensed successor to Lumberyard.	December 1, 2023
Whitepaper updated	New services added: Amazon CodeCatalyst, AWS Verified Access, Amazon Aurora I/O-Optimized, Amazon SageMaker AI geospatial capabilities, Amazon Security Lake, AWS DMS Serverless, AWS Glue for Ray, AWS Glue Data Quality, Amazon Verified Permissions, AWS AppFabric , AWS Bedrock, vector engine for Amazon OpenSearch Serverless, AWS HealthScr ibe, AWS Entity Resolutio n, and Amazon VPC Lattice. Removed Amazon Sumerian. Numerous editorial changes throughout.	September 28, 2023

Whitepaper updated	New services added: Amazon CodeWhisperer, Amazon DataZone, Amazon Linux 2023, AWS Infrastructure Composer, AWS Clean Rooms, AWS Modular Data Center. New subservices added: Amazon OpenSearch Serverless, Geospacial ML with Amazon Sagemaker, Amazon EC2 C7g Instances, Amazon EC2 Inf2 Instances, Amazon EC2 M7g instances, Amazon EC2 R7g Instances, Amazon EC2 Trn1 Instances. New program added: Integrated Private Wireless on AWS.	April 15, 2023
Whitepaper updated	New services added: Amazon File Cache, AWS IoT ExpressLink, AWS Mainframe Modernization Service. New subservices added: Amazon Connect Cases, Amazon Redshift Serverless, Amazon WorkSpaces Core, AWS WAF Captcha.	December 30, 2022
Whitepaper updated	New Container Build Lens and Healthcare Industry Lens added to the Well-Architected section.	December 23, 2022

Whitepaper updated	New service AWS Billing Conductor added, Global Infrastructure section updated, category icons added, and minor corrections throughout.	June 3, 2022
Whitepaper updated	Added note that EC2-Classic is being retired on August 15, 2022	February 17, 2022
Whitepaper updated	Added new services and compute services comparison table.	January 12, 2022
Whitepaper updated	Amazon Elasticsearch Service renamed Amazon OpenSearch Service.	September 8, 2021
Whitepaper updated	Added new services and updated information throughout.	August 5, 2021
Minor update	Minor text updates to improve accuracy and fix links.	April 12, 2021
Minor update	Minor text updates to improve accuracy.	November 20, 2020
Minor update	Fixed incorrect link.	November 19, 2020
Minor update	Fixed incorrect link.	August 11, 2020
Minor update	Fixed incorrect link.	July 17, 2020
Minor updates	Minor text updates to improve accuracy.	January 1, 2020

Minor updates	Minor text updates to improve accuracy.	October 1, 2019
Whitepaper updated	Added new services and updated information throughout.	December 1, 2018
Whitepaper updated	Added new services and updated information throughout.	April 1, 2017
Initial publication	Overview of Amazon Web Services published.	January 1, 2014

Note

To subscribe to RSS updates, you must have an RSS plug-in enabled for the browser you are using.

AWS Glossary

For the latest AWS terminology, see the AWS glossary in the *AWS Glossary Reference*.