

Instructions:

See below for a summary of each new service or feature announced at re:Invent 2021, as well as additional resources. Click on a service category in the table of contents below to jump to its respective page.

Please note: You must be registered for [re:Invent 2021 \(including the free virtual pass\)](#) to access the on-demand breakout sessions provided in this document.

Table of contents (by category)

| | | | |
|--------------------------------------|------------------------------------|-----------------------------------|---------------------------------|
| Analytics | Application Integration | Architecture | AI/ML |
| AWS Marketplace | Business Applications | Cloud Operating Model | Compute |
| Containers | Databases | DevOps and Developer Productivity | Enterprise and Migration |
| Front-End Web and Mobile Development | Games | IoT | Networking and Content Delivery |
| Robotics | Security, Identity, and Compliance | Storage | Windows |

Analytics

[Return to
Table of Contents](#)

Amazon EMR Serverless

Run big data applications using open-source frameworks without managing clusters and servers

What is it?

Amazon EMR Serverless is a serverless option on Amazon EMR that makes it easy for data analysts and engineers to run open-source big data analytics frameworks without configuring, managing, and scaling clusters or servers. You get all the features and benefits of Amazon EMR without the need for experts to plan and manage clusters.

Benefits:

- **Run frameworks more easily.** Simply select the open-source framework you want to run for your application, such as Apache Spark, Hive, or Presto, and EMR Serverless automatically provisions and manages the underlying compute and memory resources.
- **Scale on demand.** Run analytics workloads at any scale with automatic on-demand scaling that resizes resources in seconds to meet changing data volumes and processing requirements.
- **Optimize costs.** EMR Serverless automatically scales resources up and down to provide just the right amount of capacity for your application. You pay only for what you use, and you can minimize concerns about over- or under-provisioning.

Use cases:

- **Variable workloads.** As workload demands change, scale application resources seamlessly without having to preconfigure how much compute and memory you need.
- **Interactive data analysis.** Choose the option to pre-initialize application resources and enable sub-second response time for interactive data analysis in EMR Studio.
- **Development and test environments.** Spin up a development and test environment quickly and easily, automatically scale with unpredictable usage, and get products to market faster.

Availability:

Available in the US East (N. Virginia) Region.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon Redshift Serverless

Easily run and scale analytics in seconds without provisioning and managing a data warehouse

What is it?

The world's most widely used cloud data warehouse just got better. Amazon Redshift Serverless makes it easy for you to run petabyte-scale analytics in seconds to gain rapid insights—without the need to configure and manage your data warehouse clusters. Amazon Redshift Serverless automatically provisions and scales the data warehouse capacity to deliver high performance for demanding and unpredictable workloads, and you pay only for the resources you use.

Benefits:

- **Get insights from data more easily.** Amazon Redshift Serverless automatically provisions and manages the underlying infrastructure for running analytics workloads so you can focus on getting insights from data.
- **Deliver consistently high performance.** Automatically scale data warehouse capacity in seconds to deliver fast performance for even the most demanding and unpredictable workloads.
- **Optimize costs.** Save costs by automatically scaling capacity up when busy, and back down when not—you pay only for what you use. Easily manage your spend with granular cost controls.

Use cases:

- **Variable and sporadic workloads.** Scale resources seamlessly as workload demands change or spike with traffic. Amazon Redshift Serverless uses machine learning (ML) techniques to maintain consistent performance.
- **Development and test environments.** Spin up a development and test environment quickly, easily, and economically to get your products to market faster.
- **Impromptu business analytics.** Perform what-if analyses, anomaly detection, and ML-based forecasting while getting fast insights from your data.

Availability:

Available in Preview in the following Regions:

Europe (Frankfurt), Europe (Ireland), US East (N. Virginia), US West (Oregon), US East (Ohio), Asia Pacific (Tokyo)

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Storage API on AWS Lake Formation

A simpler way to ingest, manage, and secure data

What is it?

The new Storage API on AWS Lake Formation provides granular access controls to allow you to specify cell-level security policies on Amazon Simple Storage Service (S3) tables centrally managed in Lake Formation—and the API consistently enforces these policies. The new API also introduces a new type of Amazon S3 table called a “governed table” that makes it simple for you to reliably ingest and manage data at scale. Finally, it provides storage optimization to allow for faster analytics with services like Amazon Athena and Amazon Redshift.

Lake Formation focuses on three pillars as it relates to data on the lakes:

- **Ingestion and identification.** Blueprints simplify ingest. Machine learning (ML) transforms for data cleaning. And the new Storage API introduces the governed table, a new type of S3 table that removes complex extract, transform, and load (ETL) for data updates.
- **Security.** Lake Formation provides a centralized permissions table and column, and the new storage API offers row-level fine-grained permissions. There is also real-time monitoring and integrated auditing.
- **Analytics and ML.** The service features a comprehensive portfolio of integrated tools.

Availability:

Lake Formation governed tables, storage optimization, and row-level security will be initially available in the following five Regions:

US East (N. Virginia), US East (Ohio), US West (Oregon), Europe (Ireland), Asia Pacific (Tokyo)

Resources:

[Webpage](#) | [Blog post](#)

Amazon MSK Serverless

Easily stream data on Amazon Managed Streaming for Apache Kafka (MSK) without managing cluster capacity

What is it?

Amazon MSK Serverless is a new cluster type for Amazon MSK that makes it easy for you to run Apache Kafka without having to manage and scale cluster capacity. MSK Serverless automatically provisions and scales compute and storage resources, so you can use Apache Kafka on demand.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon Kinesis Data Streams On-Demand Mode

Easily stream data at any scale, without having to manage capacity

What is it?

Amazon Kinesis Data Streams On-Demand is a new capacity mode for Kinesis Data Streams that provides automatic provisioning and scaling to handle gigabytes per minute of streaming data. Kinesis Data Streams offers pay-for-use billing mode, starting as low as \$0.015 per hour.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Application integration

[Return to
Table of Contents](#)

Amazon SQS dead-letter queues

Enhance the dead-letter queue (DLQ) management experience for Amazon Simple Queue Service (SQS)

What is it?

Use DLQs to handle messages that a consumer has not successfully processed. When a message's maximum receive count is exceeded, Amazon SQS moves the message to the DLQ associated with the original queue. DLQs must be of the same type as the source queue (standard or First-In-First-Out). You can inspect the messages in DLQs to understand why your consumer has not successfully received them. Once you have remediated the issues, you can move the messages from the DLQ to their respective source queues.

Availability:

Amazon SQS DLQ is available in the following Regions:

US East (Ohio), US East (N. Virginia), US West (N. California), US West (Oregon), Africa (Cape Town), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Seoul), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Tokyo), Asia Pacific (Osaka), Canada (Central), Europe (Frankfurt), Europe (Ireland), Europe (London), Europe (Milan), Europe (Paris), Europe (Stockholm), Middle East (Bahrain), South America (São Paulo), Asia Pacific (Beijing), Asia Pacific (Ningxia)

Resources:

[Webpage](#) | [Blog post](#)

Architecture

[Return to
Table of Contents](#)

Custom Lenses for the AWS Well-Architected Tool

Add specific best practice guidance to the AWS Well-Architected Tool using custom lenses, to create your own pillars, questions, best practices, helpful resources, and improvement plans

What is it?

The AWS WA Tool is designed to help you review the state of your applications and workloads, and provide a central place for architectural best practices and guidance. The AWS WA Tool is based on the AWS Well-Architected Framework, which was developed to help cloud architects build secure, high-performing, resilient, and efficient application infrastructures. AWS solutions architects have used the Framework in tens of thousands of workload reviews, and it provides a consistent approach for evaluating your cloud architecture and implementing designs that will scale with your application needs over time.

You can review workloads against best practices across the architectural pillars of operational excellence, security, reliability, performance efficiency, and cost optimization. The AWS WA Tool delivers a list of issues found in your workloads and step-by-step guidance to make improvements. You can also view the issues found across the portfolio of workloads. We recommend performing a workload review at major milestones in your development cycle.

In addition to the standard guidance provided by the AWS WA Framework and AWS developed lenses, the AWS WA Tool allows you to add specific best practice guidance using custom lenses. By developing your own questions and evaluating your workloads using your organization's best practices, you can perform reviews based on technology or governance requirements specific to your industry.

Benefits:

Access knowledge and best practices used by AWS solutions architects whenever you need it. Answer questions about your application or workload, and the AWS WA Tool delivers an action plan with step-by-step guidance on areas for improvement.

The AWS WA Tool provides a single tool and consistent process to review and measure your cloud architectures. The tool allows you to monitor the status of multiple workloads across your organization and helps you understand potential risks. With the action plan, you can identify next steps for improvement, drive architectural decisions, and build for the cloud with confidence.

The AWS WA Tool supports continuous improvement throughout the workload lifecycle. The tool makes it easy to save point-in-time milestones and track changes to your workload. Initiate new reviews as desired to help ensure your architecture improves over time.

With the AWS WA Tool, you can create custom lenses and share them across your entire organization to measure workloads consistently. Specify rules to help you determine which options can result in a high or medium risk, and provide guidance on resolving those risks.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

The AWS Well-Architected Sustainability Pillar

Learn, measure, and improve your organization's workloads using environmental best practices for cloud computing

What is it?

The AWS Well-Architected Framework has been helping AWS customers improve their cloud workloads since 2015. The Framework consists of design principles, questions, and best practices across multiple pillars: operational excellence, security, reliability, performance efficiency, and cost optimization. We're now introducing the AWS Well-Architected Sustainability Pillar to the AWS WA Framework to help your organization take steps to learn, measure, and improve its workloads using environmental best practices for cloud computing.

Similar to the other pillars, the AWS Well-Architected Sustainability Pillar contains questions aimed at evaluating the design, architecture, and implementation of cloud workloads to reduce energy consumption and improve efficiency. The Sustainability Pillar is designed as a tool to track your progress toward policies and best practices that support a more sustainable future, not just a simple checklist.

The shared responsibility model also applies to sustainability. AWS is responsible for the sustainability of the cloud, while AWS customers are responsible for sustainability in the cloud.

Benefits:

With an increasing number of organizations setting sustainability targets, CTOs, architects, developers, and operations team members are seeking ways they can directly contribute to their organization's sustainability goals. Using the AWS Well-Architected Sustainability Pillar, organizations can make informed decisions as they balance security, cost, performance, reliability, and operational excellence with sustainability outcomes for cloud workloads. Every action taken to reduce resource usage and increase efficiency across all components of a workload contributes to a reduced environmental impact for that workload, and can help contribute to your organization's wider sustainability goals.

Availability:

The AWS Well-Architected Sustainability Pillar is available as a whitepaper on the AWS Well-Architected website.

Resources:

[Webpage](#) | [Blog post](#)

AWS re:Post

Community-driven Q&A service that provides expert technical guidance and best practices

What is it?

AWS re:Post is a new community-driven Q&A service to help AWS customers remove technical roadblocks, accelerate innovation, and enhance operations. re:Post allows you to ask questions about anything related to designing, building, deploying, and operating workloads on AWS, and get answers from community experts, including AWS customers, Partners, and employees.

re:Post replaces AWS Forums and introduces new ways to improve the accuracy of answers provided, as well as the likelihood of receiving an answer from the community. re:Post automatically connects your question with subject matter experts, and it's also integrated with AWS Support. Customers with AWS Premium Support subscriptions receive responses from AWS employees for questions not answered by the community. re:Post is part of the AWS Free Tier and available to anyone with an AWS account at the [AWS re:Post](#) page.

Use cases:

re:Post offers expert-reviewed answers to technical questions related to the design, development, deployment, and operation of applications built on AWS.

- **Remove technical roadblocks.** Work with community experts to find solutions to your design, development, deployment, and operation blockers.
- **Reduce the burden on IT resources.** Tap into self-service knowledge through the broader AWS community, reducing your dependency on in-house IT departments.
- **Build your AWS reputation.** Create a profile, answer questions, and score reputation points that contribute to your community expert status.
- **Strengthen your cloud knowledge.** Use re:Post to learn about AWS services, enhance your general cloud knowledge, or study for AWS certification exams.

Availability:

re:Post is available globally to anyone with an AWS account.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Artificial intelligence and machine learning

[Return to
Table of Contents](#)

Amazon Personalize new features

User segmentation and use case optimized recommenders

What is it?**User segmentation**

Amazon Personalize now offers intelligent user segmentation so you can run more effective prospecting campaigns through your marketing channels. By using two new recipes, you can automatically segment your users based on their interest in different product categories, brands, and more. `aws-item-affinity` identifies users based on their interest in individual items, such as movies, songs, or products. `aws-item-attribute` identifies users based on the attributes they care about, such as genre or price point. This drives higher engagement with marketing campaigns, increases retention through targeted messaging, and improves the return on investment for your marketing spend. For sample cost calculations, see [Amazon Personalize pricing](#).

Use case optimized recommenders for retail, as well as media and entertainment

New recommenders make it faster and easier to deliver high-performing personalized user experiences. You can choose from use cases like "Frequently Bought Together," "Because You Watched X," "Top Picks for You," and more. Map your data to a recommender, and Amazon Personalize chooses the optimal settings for your use case and automates the work of creating and maintaining personalized recommendations. For sample cost calculations, see [Amazon Personalize pricing](#).

Availability:

General availability. For more information, check the [AWS Regional Services](#) page.

Resources:

[Webpage](#) | [Breakout session recording and slides](#)

Amazon Kendra new capabilities

Experience Builder, Search Analytics Dashboard, Custom Document Enrichment

What is it?**Experience Builder**

You can now deploy a fully functional and customizable search experience with Amazon Kendra in a few clicks, without any coding or machine learning (ML) experience. Experience Builder delivers an intuitive visual workflow to quickly build, customize, and launch your Amazon Kendra-powered search application securely on the cloud. You can start with the ready-to-use search experience template in the builder, which can be customized by simply dragging and dropping the components you want, such as filters or sorting. You can invite others to collaborate or test your search application for feedback, and then share the project with all users when you're ready to deploy the experience. Amazon Kendra Experience Builder comes with AWS Single Sign-On (SSO) integration that supports popular identity providers such as Microsoft Azure Active Directory (AD) and Okta, delivering secure end-user SSO authentication while accessing the search experience.

Search Analytics Dashboard

Amazon Kendra Search Analytics Dashboard allows you to better understand quality and usability metrics across your Amazon Kendra-powered search applications. The Search Analytics Dashboard helps administrators and content creators understand how easily end users are finding relevant search results, the quality of the search results, and gaps in the content. The dashboard provides a snapshot of how your users interact with your search application and how effective your search results are. The analytics data can be viewed in a visual dashboard in the console, or you can build your own dashboards by accessing the Search Analytics data through an API. The dashboard empowers you to dive deep into search trends and user behavior to identify insights, as well as bring clarity to potential areas of improvement.

Custom Document Enrichment

With Amazon Kendra Custom Document Enrichment capabilities, you can build a custom ingestion pipeline that pre-processes documents before they get indexed into Amazon Kendra. For example, while ingesting content from a repository like SharePoint using AWS connectors, you can enrich documents with additional metadata, convert scanned documents to text, classify documents, extract entities, and further transform the document using custom extract, transform, and load (ETL) processes. The enrichment is performed by simple rules that can be configured in the console or by invoking functions from AWS Lambda. These functions can optionally call other AWS artificial intelligence (AI) services such as Amazon Comprehend, Amazon Transcribe, or Amazon Textract.

Resources:

[Webpage](#) | [Blog post](#)

Amazon Lex Automated Chatbot Designer (Preview)

Accelerate conversation design

What is it?

Amazon Lex Automated Chatbot Designer helps you design chatbots using existing conversation transcripts in hours rather than weeks. Using machine learning (ML), it can analyze thousands of lines of transcripts in a couple of hours and provide an initial bot design that includes common intents and the information needed to fulfill them. You can iterate on the bot design to deliver effective conversational experiences.

Designing conversational interfaces or chatbots is manual, time consuming, and prone to human errors. Developers often spend hundreds of hours analyzing transcripts and gathering the key information needed to design chatbots. With the Automated Chatbot Designer, you can off-load the analysis of conversation transcripts to Amazon Lex and accelerate the design of your chatbots, reduce errors, and improve your customer experience.

You can try the Automated Chatbot Designer in Preview for free.

Benefits:

- **Reduce manual effort.** Analyze thousands of lines of transcripts to discover intents and compile a list of information required to fulfill them, reducing manual effort in conversation design.
- **Expedite conversation design.** Automatically surface an initial bot design that you can then refine to launch conversational experiences faster.
- **Improve customer experience.** Help your bot understand customers better by minimizing ambiguity between intents. Ensure that intents are well defined and well separated to avoid frustration and improve customer experience.

Availability:

Amazon Lex is available in 12 AWS Regions. For more information, check the [AWS Regional Services](#) page.

Resources:

[Webpage](#) | [Blog post](#)

Amazon SageMaker Canvas

Generate accurate machine learning (ML) predictions—no code required

What is it?

Amazon SageMaker Canvas is a visual point-and-click interface that expands ML access to business analysts, allowing them to generate accurate predictions without writing code or having ML expertise. SageMaker Canvas makes it easy to browse and access cloud and on-premises data sources. You can quickly combine, cleanse, and prepare datasets to ensure your data is ready for ML. SageMaker Canvas then makes it easy to build ML models and generate accurate predictions with just a few clicks. You can also easily export results, explain and interpret models, and share models with data scientists within your organization to review.

Benefits:

- **Generate ML predictions without writing code.** SageMaker Canvas provides a visual point-and-click interface for business analysts to build ML models and generate accurate predictions without writing code or having any previous ML experience.
- **Quickly access and prepare data for ML.** With SageMaker Canvas, you can quickly connect and access data from cloud and on-premises data sources, combine datasets, and create unified datasets for training ML models. SageMaker Canvas automatically detects and corrects data errors and analyzes data readiness for ML.
- **Use built-in AutoML to generate predictions.** SageMaker Canvas uses powerful AutoML technology from Amazon SageMaker to automatically create ML models based on your unique use case. This allows SageMaker Canvas to identify the best model based on your dataset so you can generate accurate predictions—whether singular or in bulk.
- **Validate ML models with data scientists.** SageMaker Canvas is integrated with Amazon SageMaker Studio, making it easier for business analysts to share models and datasets with data scientists so they can validate and further refine the ML model.

Use cases:

- **Predict customer churn.** Use product-consumption and purchase-history data to uncover customer churn patterns and predict those at risk of churning in the future.
- **Optimize price and revenue.** Predict the prices of goods and services using historical demand and pricing and seasonal trends to offer the best prices to customers while maximizing revenue.
- **Improve on-time deliveries.** Predict delivery times using order, fulfillment, transit, and holiday data to optimize the supply chain and deliver goods with greater efficiency.
- **Plan inventory efficiently.** Predict inventory needs by combining historical sales and demand data with associated web traffic, pricing, product category, weather, and holiday data.

Availability:

SageMaker Canvas is supported in the largest SageMaker Regions, including the following:
US East (N. Virginia), US East (Ohio), US West (Oregon), Europe (Ireland), Europe (Frankfurt)

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon SageMaker Inference Recommender

Automatic inference instance selection and load testing

What is it?

Amazon SageMaker Inference Recommender helps you choose the best available compute instance and configuration to deploy machine learning (ML) models for optimal inference performance and cost. SageMaker Inference Recommender automatically selects the compute instance type, instance count, container parameters, and model optimizations for inference to maximize performance and minimize cost. You can use SageMaker Inference Recommender from Amazon SageMaker Studio, the AWS Command Line Interface (CLI), or the AWS SDK, and within minutes, get recommendations to deploy your ML model. You can then deploy your model to one of the recommended instances or run a fully managed load test on a set of instance types you choose without worrying about testing infrastructure. You can review the results of the load test in SageMaker Studio and evaluate the trade-offs between latency, throughput, and cost to select the most optimal deployment configuration for your use case.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon SageMaker Serverless Inference (Preview)

Easily deploy machine learning (ML) models for inference without configuring or managing any of the underlying infrastructure

What is it?

For use cases with intermittent and unpredictable usage patterns, Amazon SageMaker Serverless Inference (Preview) allows you to deploy ML models on pay-per-use pricing without worrying about servers or clusters. When deploying your model, simply select the serverless option, and Amazon SageMaker automatically provisions, scales, and turns off compute capacity based on the volume of inference requests. As a result, you don't need to manage complex scaling policies and forecast traffic demand up front. With SageMaker Serverless Inference, you pay only for the compute capacity used to run the inference requests, billed by the millisecond, and the amount of data processed—you're not charged for periods of no traffic.

Availability:

SageMaker Serverless Inference is available in public Preview in the following Regions:

US East (N. Virginia), US East (Ohio), US West (Oregon), EU (Ireland), Asia Pacific (Tokyo), Asia Pacific (Sydney)

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon SageMaker Studio Lab (Preview)

Learn and experiment with machine learning (ML) using a no-setup, free development environment

What is it?

Amazon SageMaker Studio Lab is a free ML development environment that provides compute, storage (up to 15 GB), 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, manage identity and access, or even sign up for an AWS account. SageMaker Studio Lab accelerates model building through GitHub integration, and it comes preconfigured with the most popular ML tools, frameworks, and libraries so can start immediately. SageMaker 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.

Benefits:

- **No AWS account needed.** To get started with SageMaker Studio Lab, use your email address to [register for an account](#). Your SageMaker Studio Lab account is separate from an AWS account and doesn't require a credit card.
- **Choose compute power.** SageMaker Studio Lab offers either CPU or GPU sessions for your project. You can choose to run Jupyter notebooks with a 12-hour CPU session for complex algorithms, or a four-hour GPU session for deep learning (DL) architectures such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). There's no limit to the number of compute sessions you can run—after a session ends you can start a new one.
- **Persistent storage.** SageMaker Studio Lab provides persistent sessions with 15 GB of dedicated storage, so you can save your work and pick up where you left off.
- **Prepackaged ML frameworks.** Choose the best Python package manager for your project, such as Pip, Conda, and Mamba. By default, SageMaker Studio Lab features support for the Terminal and Git command lines, and GitHub integration for collaboration. Setup is fast, and it's easy to run a notebook.

Use cases:

- **ML education.** SageMaker Studio Lab allows students to open Jupyter notebooks instantly and get started with the ML frameworks TensorFlow, PyTorch, and Apache MXNet, as well as libraries such as XGBoost and AutoGluon. Conduct hands-on tutorials, exercises, and labs easily at no cost. ML instructors can spend more time teaching and less time debugging technical issues associated with cloud configuration.
- **Hackathons and challenges.** Data science communities can use SageMaker Studio Lab to engage aspiring data scientists in challenges and hackathons, helping them gain hands-on experience and solve real-world problems.
- **Corporate ML training and workforce development.** Enterprise training programs can use SageMaker Studio Lab in internal training and workforce development activities to meet the growing demand among working professionals seeking to build their skills in data science and ML, conduct experiments, and implement rapid prototypes.

Availability:

Users do not choose a Region. SageMaker Studio Lab removes the complexity of having to choose a Region. Instead, sign in to the URL and start building ML models.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon SageMaker Training Compiler

A deep learning (DL) compiler that accelerates the training of DL models by up to 50%

What is it?

Amazon SageMaker Training Compiler is a DL compiler that accelerates the model training process by up to 50% through graph- and kernel-level optimizations that make more efficient use of GPUs. Moreover, you can add either data parallelism or model parallelism to your training script with a few lines of code, and the Amazon SageMaker distributed training libraries will automatically split models and training datasets across GPU instances to help you complete distributed training faster.

State-of-the-art DL models consist of complex multilayered neural networks with billions of parameters that can take thousands of GPU hours to train. Optimizing such models on training infrastructure requires extensive knowledge of DL and systems engineering; this is challenging even for narrow use cases. Although there are open-source implementations of compilers that optimize the DL training process, they can lack the flexibility to integrate DL frameworks with certain hardware, such as GPU instances.

SageMaker Training Compiler is a capability of SageMaker that makes these hard-to-implement optimizations to reduce training time on GPU instances. The compiler optimizes DL models to accelerate training by more efficiently using SageMaker machine learning (ML) GPU instances. SageMaker Training Compiler is available at no additional charge within SageMaker and can help reduce total billable time as it accelerates training.

Availability:

Available in the following Regions at general availability:

US East (N. Virginia), US East (Ohio), US West (Oregon), and Europe (Ireland)

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon SageMaker Ground Truth Plus

Create high-quality training datasets without having to build labeling applications or manage a labeling workforce

What is it?

With Amazon SageMaker Ground Truth Plus, you can easily create high-quality training datasets without having to build labeling applications or manage labeling workforces on your own. SageMaker Ground Truth Plus helps reduce data labeling costs by up to 40%. SageMaker Ground Truth Plus provides an expert workforce that is trained on machine learning (ML) tasks and can help meet your data security, privacy, and compliance requirements. You simply upload your data, and SageMaker Ground Truth Plus creates data labeling workflows and manages workflows on your behalf.

Availability:

Available in the US East (N. Virginia) Region.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon CodeGuru Reviewer Secrets Detector

Machine learning (ML)–based analysis to help you detect secrets that are hard-coded in your repository or configuration files

What is it?

Amazon CodeGuru Reviewer Secrets Detector uses ML-based analysis to help you detect secrets that are hard-coded in your repository or configuration files, including passwords, API keys, SSH keys, access tokens, database connection strings, and JSON Web Tokens (JWTs).

Benefits:

Part of Amazon CodeGuru Reviewer, CodeGuru Secrets Detector is an automated mechanism that checks code for these secrets and provides point-and-click steps to secure them using AWS Secrets Manager. It can also identify specific keys generated by the most common API providers, including AWS, Atlassian, GitHub, Salesforce, HubSpot, and Stripe.

Availability:

CodeGuru is currently available in 10 AWS Regions:

US East (N. Virginia), US East (Ohio), US West (Oregon), Europe (Ireland), Europe (Frankfurt), Europe (London), Europe (Stockholm), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Tokyo)

For the latest information, see the [AWS Regional Services](#) page.

Resources:

[Webpage](#) | [Blog post](#)

Amazon Textract AnalyzeID

Specialized support for identity documents

What is it?

Amazon Textract AnalyzeID can extract data from identity documents (IDs) such as passports, driver's licenses, and state IDs issued by the U.S., without the need for templates or configuration.

Resources:

[Blog post](#)

Amazon SageMaker Studio Notebooks

New ways to conduct data processing, analytics, and machine learning (ML) workflows in one unified Studio notebook

What is it?

Amazon SageMaker Studio now offers new ways to conduct data processing, analytics, and ML workflows in one unified Studio notebook. Just as you can write different documents from one word processing program, you can use this universal Studio notebook to access a wide range of data sources and write code for any transformation for a variety of data workloads.

Benefits:

- **Wide variety of data sources.** Studio notebooks have built-in integration with Spark, Hive, and Presto running on Amazon EMR clusters and data lakes running on Amazon Simple Storage Service (S3)—with support for additional data sources coming in early 2022—so you can access and manipulate data without switching services.
- **Complete data processing.** Browse and query data sources, explore metadata and schemas, and start processing jobs for analytics or ML workflows.
- **Code in your preferred framework.** Write code using your preferred framework to build, train, and deploy ML models—all without leaving the universal Studio notebook.

Use cases:

- **Data processing.** With Studio notebooks, you can interactively access, transform, and analyze a wide range of data in a highly visual way so it's easier to prepare data for any workload.
- **ML.** You can write code on preferred frameworks such as TensorFlow and PyTorch to build ML applications.

Resources:

[Webpage](#) | [Blog post](#)

AWS AI & ML Scholarship Program

Hands-on learning, scholarships, and mentorship for people underserved or underrepresented in tech

What is it?

The AWS Artificial Intelligence (AI) and Machine Learning (ML) Scholarship program, in collaboration with Intel and Udacity, aims to help underrepresented and underserved high school and college students learn foundational ML concepts to prepare them for careers in AI and ML. Using the all-new AWS DeepRacer Student service, the program lets students from around the world access dozens of hours of free training modules and tutorials on the basics of ML and its real-world applications. Two thousand students are eligible to earn scholarships annually as part of Amazon's commitment to a more diverse future AI and ML workforce. Eligible students must successfully complete an application, complete educational module quizzes, and meet AWS DeepRacer Student League performance targets. The DeepRacer Student League is a global autonomous racing competition exclusively for AWS AI and ML students.

Resources:

[Webpage](#) | [Blog post](#)

AWS Marketplace for Containers Anywhere

Easily find Kubernetes applications and deploy them on AWS or on premises

What is it?

With AWS Marketplace for Containers Anywhere, you can find, subscribe to, and deploy third-party Kubernetes applications from AWS Marketplace on any Kubernetes cluster in any environment, extending the container capabilities of AWS Marketplace. You can deploy third-party Kubernetes applications on premises using Amazon Elastic Kubernetes Service Anywhere (EKS Anywhere). You can also deploy self-managed Kubernetes clusters on premises or on Amazon Elastic Compute Cloud (EC2). Find over 500 vetted container applications on AWS Marketplace, where you can manage upgrades with a few clicks and track all your licenses and bills in one place. Discover commercial and open-source Kubernetes-based applications from popular independent software vendors (ISVs), and deploy these applications in any Kubernetes environment in minutes, including on premises. When using AWS Marketplace for Containers Anywhere, you have the same benefits you get with any other products on AWS Marketplace, including consolidated billing, flexible payment options, and lower pricing for long-term contracts.

Benefits:

- **Ease of deployment.** Deploy applications on Amazon EKS, Amazon EKS Anywhere, or any self-managed Kubernetes cluster on AWS or on premises.
- **Consistent deployment experience.** Scale faster by procuring an application that has been scanned for common vulnerabilities and exposures (CVEs).
- **Extensive catalog.** Save time by purchasing on AWS Marketplace, which offers open-source and commercial Kubernetes applications.

Use cases:

- **Find Kubernetes management applications.** Browse a growing catalog of Kubernetes management applications to find the ones you need.
- **Deploy anywhere.** Subscribe to and deploy third-party Kubernetes applications on any AWS managed or self-managed Kubernetes cluster, on Amazon EC2, or on premises.
- **Manage licenses across your AWS Organization.** Share licenses across your Organization. Licenses are transferable between Kubernetes clusters.
- **Migrate to AWS.** Deploy third-party Kubernetes applications to on-premises Kubernetes clusters and migrate to AWS using the same license when ready.

Availability:

Available in 25 AWS Regions. See details on the [AWS Regional Services](#) page.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Data Exchange for APIs

Find, subscribe to, and use hundreds of third-party APIs in the cloud

What is it?

AWS Data Exchange for APIs allows you to find, subscribe to, and use third-party API products from providers on AWS Data Exchange. With AWS Data Exchange for APIs, use AWS-native authentication and governance, explore consistent API documentation, and use supported AWS SDKs to make API calls. Data providers can now reach millions of AWS customers that consume API-based data by adding their APIs to the AWS Data Exchange catalog, and more easily manage subscriber authentication, entitlement, and billing.

Benefits:

For subscribers:

- **Find the APIs your organization needs—all in one place.** Browse the AWS Data Exchange catalog of products containing APIs, and view relevant information on product detail pages.
- **Subscribe to APIs from multiple vendors.** Easily manage and govern your organization's subscriptions and entitlements through the AWS Data Exchange console, AWS License Manager, and AWS Identity and Access Management (IAM).
- **Easily call third-party APIs.** Use AWS SDKs to make API calls using your AWS credentials.

For data providers:

- Reach AWS customers with your APIs.
- Manage subscriptions and entitlements.
- Make it easy for customers to use your APIs.

Availability:

Available in 25 AWS Regions. See details on the [AWS Regional Services](#) page.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon Connect new capabilities

Automated call summarization in Contact Lens for Amazon Connect and Automated Chatbot Designer in Amazon Lex

What is it?

Automated call summarization in Contact Lens for Amazon Connect

Call summarization (generally available) is a new capability in Contact Lens for Amazon Connect that uses machine learning (ML) to automatically generate a summary of customer service conversations, improving the productivity of agents and managers so they can focus on providing excellent customer experiences. Contact center agents typically spend between five and 10 minutes after each call summarizing notes, and managers spend a significant amount of time listening to call recordings or reading transcripts when they're investigating customer issues. Contact Lens for Amazon Connect with call summarization helps agents and managers become more efficient by automatically generating a summary after customer calls. Using ML, this new call summarization capability automatically identifies key parts of a customer conversation, assigns a tag (such as issue, outcome, or action item), and displays a summary that can be expanded to view the call's full transcript. With call summarization, agents can consistently and accurately capture key parts of the customer interaction, and then revisit the summary when they're following up with a customer to resolve an issue. Managers can quickly understand the context of an interaction by viewing the call summary alongside the call recording and contact details on Amazon Connect—instead of spending time reading the entire transcript. The new Contact Lens for Amazon Connect call summarization capability makes agents more efficient, decreases customer wait times, and makes it easier for managers to understand the context of a call.

Amazon Lex Automated Chatbot Designer

Amazon Lex Automated Chatbot Designer (available in Preview) is a new capability that uses ML to reduce chatbot design times from weeks to hours. Amazon Lex helps you build, test, and deploy voice and text bots (also known as chatbots) for services like Amazon Connect. Many customer service organizations use chatbots to answer customer questions, provide information, and complete tasks like paying a bill. However, designing a chatbot is manual, time consuming, and complex. It requires a deep understanding of spoken language and human interactions. Developers start the chatbot design process by analyzing transcripts to find the reason for the contact and the desired outcome, which are commonly called "intents" by chatbot designers. For example, a person contacting an insurance company to file a claim might say, "My basement is flooded. I need to start a new claim." The intent in this case is to "file a new claim." Chatbot developers can spend weeks listening to call recordings and analyzing thousands of lines of customer service transcripts to find the right intents to design contact center chatbots. However, that process is difficult and prone to error, and it could lead to poor customer experiences. A chatbot with missing, incomplete, or overlapping intents will fail to resolve issues and requests accurately, resulting in frustrated customers. However, the Automated Chatbot Designer uses ML to automatically design chatbots in hours instead of weeks. First, developers access the Automated Chatbot Designer and upload transcripts by going to the Amazon Lex console. Next, the Automated Chatbot Designer uses ML to analyze those transcripts and creates an initial chatbot design. The automated chatbot designer can analyze thousands of lines of transcripts within a few hours, minimizing developer effort and reducing the time required to design a chatbot. In the Amazon Lex console, the Automated Chatbot Designer includes common intents, associated phrases, and a list of information the chatbot will need to capture to resolve issues (such as the customer policy number and claim type). Developers can iterate on that initial design, change chatbot prompts and responses, and then build, test, and deploy the chatbot on Amazon Lex. With the Automated Chatbot Designer, you can streamline the lengthy design process and create chatbots that understand customer requests and improve customer experiences.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Cloud Operating Model

[Return to
Table of Contents](#)

Amazon CloudWatch RUM

Optimize end-user experience by reducing mean time to resolution (MTTR) for client-side performance issues

What is it?

Amazon CloudWatch RUM helps application developers and DevOps engineers optimize end-user experience by reducing MTTR for client-side performance issues. The service provides real user monitoring (RUM) so you can monitor the client-side performance of your web applications (desktop and mobile browser) in real time and quickly resolve any issues. Obtain relevant debugging data—such as error messages, stack traces, and user sessions—to fix performance issues like JavaScript errors, crashes, and latencies. You can also isolate the end-user impact to the number of users, geolocations, or browsers. RUM enhances the end-user observability of the Amazon CloudWatch Application Performance Monitoring (APM) portfolio.

RUM reduces MTTR by identifying client-side performance issues for a quicker resolution. With RUM, Amazon CloudWatch now includes all application monitoring data, including frontend and backend metrics—giving you more visibility into your application's performance. For example, RUM surfaces relevant debugging data for errors by providing user session data to recreate the issue, stack traces, and trace to the backend infrastructure nodes through an integration with Amazon CloudWatch ServiceLens and AWS X-Ray. RUM helps developers prioritize features and bug fixes by aggregating how users are interacting with the application, such as comparing bounced user sessions.

Benefits:

- Reduce MTTR by identifying client-side performance issues for a quicker resolution.
- Surface relevant debugging data—such as error messages, stack traces, and tracing to server-side infrastructure—through an integration with CloudWatch ServiceLens and X-Ray. Consolidate application monitoring data, including client- and server-side metrics and traces, to gain more visibility into your application's performance.
- Understand last mile performance in the end-user experience, including DNS resolution and initial connection time.
- Quantify user impact and prioritize accordingly. Determine how prevalent an issue is, and spot performance anomalies pertaining to a location, browser, or device.

Use cases:

- **Optimize the end-user experience.** RUM allows for day-to-day monitoring of web applications' client-side performance. Spot issues (including errors and slow page loads) or anomalies (interactions with browsers) and fix them to optimize the customer experience.
- **Connect the client side to resources.** Use the X-Ray tracing feature to connect any client-side interaction to server-side infrastructure resources.
- **Understand user interaction with applications.** Gain insights into how users interact with your applications, including aggregate page-to-page navigation.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon CloudWatch Evidently

Safely validate new features across the full software stack before release

What is it?

Amazon CloudWatch Evidently is a new feature for application developers that allows them to safely validate new features across the full software stack before release. Use Evidently to conduct experiments, reduce risk, and identify unintended consequences before rolling out features for general use. When launching new features, you can expose features to certain users, monitor key metrics such as page load times or conversions, and then safely increase to a wider user base. Evidently also allows you to try out different designs, collect user data, and release the best design in production. With Evidently, optimize user experience by using a single solution for complete application monitoring through Amazon CloudWatch.

Evidently is a safe launch monitoring and experimentation feature to manage user traffic allocation, run experiments, and capture user data. Launch a new feature to a small set of users, collect data, get alerted for issues, and then iterate as required. You can also identify unintended consequences, reduce risks, and safely launch new features. Monitor CloudWatch metrics, including those generated by Amazon CloudWatch RUM. RUM is a real user monitoring (RUM) feature that collects client-side application performance and user journeys. With the addition of Evidently, CloudWatch is now the only tool you need to monitor and verify the end-user experience. Evidently removes the guesswork by helping developers conduct experiments, try different designs and configurations, get user feedback, and deploy the chosen design to production. Correctly interpret and act on the results, without the need for advanced statistical knowledge. The Evidently statistical engine runs robust analysis on the results, providing insights such as anytime-valid p-values and confidence intervals—allowing you to make decisions even as the experiment is in progress.

Benefits:

- Reduce the risk of application disruption through controlled launches and experiments to validate new features before rolling them out for general use.
- Launch application features faster by easily correlating user metrics, and help them remain safer through more control and visibility into your launches.
- Remove guesswork by allowing developers to conduct experiments, try different designs and configurations, get user feedback, and deploy the chosen design to production.
- Interpret and act on experimental results without the need for advanced statistical knowledge.
- Validate the user experience by using a single solution for complete application monitoring through CloudWatch.

Use cases:

- **Controlled launches (traffic splitting).** Use Evidently to split traffic between different versions of a new user-facing feature or backend implementation. Separate launches from deployments, and use the optional stop switch to return to the default version.
- **Launch monitoring.** Gain visibility into your metrics based on the version served.
- **Blue/green deployment monitoring.** Monitor your blue/green deployments by tracing the error counts of each version and setting an alarm on those metrics.
- **Experimentation.** Try different designs and configurations, get statistical results for decision-making such as anytime-valid p-values and confidence intervals, and launch the chosen design to production.

Availability:

Generally available in nine AWS Regions:

US East (N. Virginia), US East (Ohio), US West (Oregon), Europe (Ireland), Europe (Frankfurt), Europe (Stockholm), Asia Pacific (Sydney), Asia Pacific (Tokyo), Asia Pacific (Singapore)

For more information, see the [AWS Regional Services](#) page.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Chatbot supports the management of AWS resources in Slack (Preview)

Manage AWS resources and remediate issues in AWS workloads by running AWS Command Line Interface (CLI) commands and AWS System Manager automation runbooks from Slack channels

Benefits:

- **Manage AWS resources in Slack.** Securely run AWS CLI commands to configure AWS resources and remediate issues with AWS Chatbot. Service administrators can tailor AWS Chatbot permissions to their compliance and security needs with features such as AWS Identity and Access Management (IAM) user roles for running mutative AWS CLI commands, as well as IAM channel guardrails for granting the least privileges to channel members.
- **Remediate issues and operate CloudOps workflows by running AWS Systems Manager Automation runbooks.** Run Automation runbooks from Slack channels to resolve issues and run CloudOps workflows. Look up AWS Systems Manager runbooks, supply runbook parameters, and then run the runbook from Slack channels. AWS Chatbot also delivers runbook status notifications in Slack to update you on the runbook progress.
- **Natural language–based task understanding.** AWS Chatbot also accelerates task completion with natural language–aided task understanding. When members in chatrooms type tasks, such as “@aws describe cw alarms,” AWS Chatbot understands the implied AWS CLI service, operation, and parameters from the typed inputs. It guides channel members in correcting syntax errors, completing the AWS CLI command, and prompting required parameters before successfully initiating command execution.

Use cases:

DevOps and engineering teams are increasingly moving their operations, system management, and CI/CD workflows to chat applications to streamline activities and improve team collaboration. Without a secure and integrated ChatOps experience, managing AWS resources and resolving issues requires switching contexts between different tools, correlating data from various sources, and continuously sharing progress in the chat room, leading to communication overhead and long task completion cycles. On receiving notifications in Slack channels, AWS customers were required to switch to the AWS Management Console, AWS CLI, or other tools to remediate the incidents and configure their AWS environments.

However, with this new feature in AWS Chatbot, you can now monitor, operate, and troubleshoot AWS workloads from Slack channels without switching context between Slack and other AWS management tools. Securely run AWS CLI commands from your Slack channels to perform common DevOps tasks, including scaling Amazon Elastic Compute Cloud (EC2) instances, running AWS Systems Manager runbooks, and changing AWS Lambda concurrency limits. Additionally, service administrators can use policy guardrails, as well as account-level and user-role permissions, to meet their security and compliance needs.

Resources:

[Webpage](#) | [Blog post](#)

SSM-IoT Integration with AWS Systems Manager

Manage both system software and applications for edge devices through an integrated experience in the AWS IoT Greengrass console

What is it?

SSM-IoT Integration is a new AWS IoT Greengrass integration with AWS Systems Manager that allows IT and edge-device administrators to securely manage their edge devices, such as industrial equipment and PCs, alongside their IT assets, such as Amazon Elastic Compute Cloud (EC2) instances, AWS Outposts, and on-premises servers. As a result, administrators can manage system software—such as patching, as well as remote access and edge application management—using a single integrated console.

Benefits:

- SSM-IoT Integration increases edge-device administrator efficiency by allowing administrators to manage both system software and applications for edge devices through an integrated experience in the AWS IoT Greengrass console.
- With SSM-IoT Integration, edge-device administrators can provision, patch, and remotely access their fleet of edge devices, such as industrial PCs, using a one-time wizard-based onboarding experience.
- SSM-IoT Integration saves time and money by also allowing IT administrators to view and manage AWS IoT Greengrass devices alongside IT assets, such as Amazon EC2 instances, with a consistent set of tools and operational policies on Systems Manager, without the need to invest in any custom integrations.

Use cases:

- Visualize edge devices alongside IT assets, such as EC2 instances and on-premises instances, using Systems Manager.
- Easily automate the system software management of edge devices using prebuilt Systems Manager runbooks.
- Manage system software and edge-device applications through an integrated experience.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Control Tower new features

AWS Control Tower Account Factory for Terraform (AFT) and data residency guardrails

What is it?

AWS Control Tower offers the easiest way to set up and govern a secure, multi-account AWS environment. It establishes a landing zone based on best-practice blueprints, and allows governance using guardrails you can choose from a prepackaged list. The landing zone is a well-architected, multi-account baseline that follows AWS best practices. Guardrails implement governance rules for security, compliance, and operations.

AWS Control Tower is for customers who want to create or manage their multi-account AWS environment with best practices. It offers prescriptive guidance to govern your AWS environment at scale. It gives you control over your environment without sacrificing the speed and agility AWS provides for builders. You will benefit from AWS Control Tower if you're building a new AWS environment, starting your journey with AWS, beginning a new cloud initiative, completely new to AWS, or have an existing multi-account AWS environment but would prefer a solution with built-in blueprints and guardrails.

AWS Control Tower now supports the Terraform pipeline to provision and customize your accounts through AWS Control Tower Account Factory for Terraform (AFT). New data residency guardrails help keep your customer data in the AWS Regions you specify.

[The Customizations for AWS Control Tower](#) solution combines AWS Control Tower with other highly available, trusted AWS services to help you more quickly set up a secure, multi-account AWS environment using AWS best practices. Before deploying this solution, you must have an AWS Control Tower landing zone deployed in your account.

AWS Marketplace now offers integrated third-party [software solutions for AWS Control Tower](#). Built by independent software vendors (ISVs), these solutions help solve infrastructure and operational use cases, including security for a multi-account environment, centralized networking, operational intelligence, and Security and Information Event Management (SIEM).

Availability:

To see a current list of Regions where AWS Control Tower is available, please visit the [AWS Regional Services](#) page.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon CloudWatch Metrics Insights

Analyze and visualize operational metrics at scale

What is it?

Amazon CloudWatch Metrics Insights is a high-performance query engine that helps you sort and analyze your operational metrics in real time and create aggregations as you go using standard SQL queries. Metrics Insights helps you understand your application's health and performance by allowing you to analyze your metrics at scale. The query engine is integrated with Amazon CloudWatch dashboards so you can save queries in your health and performance dashboards for a better monitoring experience. You can also access Metrics Insights on Amazon Managed Grafana and Grafana open-source consoles to empower your Grafana dashboards with fast, flexible, and powerful query capability. These features are also all available programmatically through the AWS Command Line Interface (CLI), AWS SDK, and AWS CloudFormation.

Metrics Insights is the fastest way to gain real-time operational insights and react to those insights from large-scale time-series data. With grouping and aggregation capabilities, you can query across CloudWatch metrics at scale for a high-level overview of your applications and infrastructure, allowing you to pinpoint issues more quickly. Metrics Insights also allows you to create health and performance dashboards using your query results to monitor whether your applications are failing (due to an issue such as memory leaks), which could impact the end customer experience. You can now query across millions of metrics, sort and limit results, and filter metrics to, for example, see the most CPU-intensive Amazon Elastic Compute Cloud (EC2) instances.

Benefits:

- Get operational insights by sorting your data according to application name, AWS Auto Scaling group, instance type, or any other label you choose to define your metrics and resources and understand your business.
- Define application-level dashboards to monitor service health and when you need to troubleshoot. Easily refine queries to pinpoint issues and resolve them quickly.
- Predict problems earlier, prevent issues from occurring, and troubleshoot to quickly isolate, diagnose, and remediate issues.

Use cases:

- **Infrastructure monitoring and troubleshooting.** Metrics Insights helps you analyze your operational metrics from different perspectives to gain insights about the patterns in your infrastructure's performance in real time. Aggregate your default AWS metrics by AWS resource tag or property. For example, instantly analyze thousands of instances by CPU utilization to troubleshoot a nonresponsive application. You can also group your metrics by Auto Scaling group to identify the problematic group and pinpoint failing instances. Once the instance is isolated, you can now recover the application by quickly rebooting problematic instances. In addition, create dashboards on a Metrics Insights query, identify the problematic instance, and take preventive actions earlier before the application fails.
- **Application monitoring.** With Metrics Insights, you can use your existing custom metrics or publish new metrics at any scale by using labels that represent groupings relevant to your business, such as application name, environment, owner, or version numbers. For example, monitor and compare order-rate performance on retail websites in different countries. To accomplish this, group your order-rate metric by country and sum the metrics to visualize and compare each site's performance. Metrics Insights is a powerful, interactive query engine that will help you sort and analyze your metrics down to the finest detail to help you quickly gain actionable insights on your business.
- **Deployment monitoring.** Metrics Insights uses a range of powerful, out-of-the-box chart types to help you aggregate and group your metrics and visualize results. These chart types will stay up to date as resources are deployed or terminated. For example, when monitoring blue/green deployments, publish metrics with labels representing error type, version, and application name or ID; create a Metrics Insights query to summarize error counts by version; and then create a dashboard for it. Next, you can monitor the sum of error metrics and identify when that number changes due to a new version that was introduced. As a result, you can identify when you need to roll back the new version, reducing the risks and eliminating the downtime.
- **Integration with independent software vendor (ISV) tools.** Use the Metrics Insights flexible query engine not only through the CloudWatch Metrics console, but also through other services using the Amazon CloudWatch GetMetricData API, such as the Amazon CloudWatch plugin for Grafana. This improves the compatibility of Metrics Insights with other services, as well as its integration capability with data outside CloudWatch. With this ability, for instance, you can run your SQL queries on the GetMetricData API and easily plot your graphs and dashboards on Grafana.

Availability:

Available in all AWS commercial Regions except China. For more information, see the [AWS Regional Services](#) page.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Compute

[Return to
Table of Contents](#)

AWS Compute Optimizer enhanced infrastructure metrics

Recommends optimal AWS resources to reduce costs and improve performance for your workloads

What is it?

[AWS Compute Optimizer](#) now offers enhanced infrastructure metrics, a feature that, when activated, enhances your Amazon Elastic Compute Cloud (EC2) instance and AWS Auto Scaling group recommendations by capturing monthly or quarterly utilization patterns.

Benefits:

Enhance your Amazon EC2 instance and Auto Scaling group recommendations by capturing monthly or quarterly utilization patterns.

Use cases:

If you have monthly or quarterly usage patterns, the three-month lookback can make Amazon EC2 and Auto Scaling group recommendations more accurate and representative of workload characteristics.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Outposts servers

Run AWS Outposts in locations with limited space or smaller capacity requirements

What is it?

Outposts servers, which are a part of the Outposts family, provide local compute and networking services for sites with space and capacity constraints, like retail stores, branch offices, healthcare provider locations, or factory floors. They're ideal for low-latency and local data processing needs.

Outposts servers are delivered directly to you, and installed by either your on-site personnel or a third-party vendor. Once connected to your network, AWS will remotely provision compute and storage resources. With Outposts servers, you can run native AWS services locally, such as Amazon Elastic Compute Cloud (EC2), Amazon Elastic Container Service (ECS), and AWS IoT Greengrass. You can use the same AWS APIs, control plane, and tools you're familiar with to manage servers across hundreds or thousands of on-premises locations.

Benefits:

- Deploy fully managed, rack-mountable servers on premises for space-constrained locations or smaller capacity requirements.
- Deliver innovative applications requiring low-latency, local data processing closer to end users and on-premises systems.
- Build and operate with familiar AWS infrastructure, services, APIs, and tools for a truly consistent hybrid experience.

Use cases:

- **Manufacturing.** Process and analyze data that responds in near real time to factory floor equipment. Reduce error rates and improve your operational outcomes by processing data and images from your on-premises production line with AWS IoT services and machine learning (ML).
- **Telecommunications.** Use cloud services and tools to orchestrate, update, scale, and manage your networks. Deliver the high performance and low latency required for SD-WAN and private 5G applications on Outposts servers.
- **Retail.** Run sales systems at all your retail shop locations for low latency and local network access. Aggregate inventory data and analyze customer behavior using data lakes and ML models running in an AWS Region.
- **Healthcare clinics and hospitals.** Identify trends and conditions by analyzing real-time equipment data and using ML inference. Allow for rapid evaluation of patient monitoring devices to accelerate medical diagnosis.

Availability:

Outposts can be shipped to and installed in the following countries and territories:

- NA: the US and Canada
- EMEA: All EU countries, the United Kingdom (UK), Switzerland, Norway, Bahrain, Gibraltar, and Morocco
- APAC: Australia, New Zealand, Japan, Hong Kong Special Administrative Region, Macau, Singapore, Indonesia, Malaysia, the Philippines, and Brunei
- CA: Puerto Rico

Support for more countries and territories is coming soon.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon EC2 Im4gn and Is4gen instances

Best price performance on Amazon Elastic Compute Cloud (EC2) for storage-intensive applications

What is it?

Amazon EC2 Im4gn and Is4gen instances are next-generation, storage-optimized instances designed for running applications that require high throughput and low-latency access to large amounts of data on local SSD storage, such as SQL databases (MySQL, MariaDB, PostgreSQL), NoSQL databases (Cassandra, ScyllaDB, MongoDB), search engines, analytics, streaming, and large distributed file systems. They're powered by AWS Graviton2 processors and provide up to 30 TB of storage with AWS Nitro SSDs. The AWS Nitro SSDs are AWS-designed SSDs that provide high I/O performance, low latency, minimal latency variability, and security with always-on encryption.

Im4gn instances provide the best price performance for storage-optimized workloads on Amazon EC2 and up to 100 Gbps of networking bandwidth. They're ideal for running applications, such as MySQL databases, NoSQL databases, and file systems, that require dense local SSD storage and higher compute performance compared to I3 instances.

Is4gen instances provide the lowest cost per TB and highest density per vCPU of SSD storage on EC2. They're ideal for running applications such as stream processing and monitoring, real-time databases, log analytics, and distributed file systems.

Benefits:

- **EC2 Im4gn instances provide the best price performance for a large number of storage-intensive applications that use Linux.** They provide up to 40% better price performance compared to EC2 I3 instances. Im4gn instances also feature Nitro SSDs and provide up to 44% lower cost per TB of SSD storage compared to I3 instances.
- Is4gen instances provide the lowest cost per TB of SSD storage on EC2. They provide up to 30 TB of Nitro SSD storage with up to 15% lower cost per TB of storage compared to I3en instances. Im4gn instances also feature Nitro SSDs and provide up to 44% lower cost per TB of SSD storage compared to I3 instances.
- **Im4gn and Is4gen instances are powered by AWS Graviton2 processors, feature Nitro SSDs, and are built on the Nitro System.** AWS Graviton2 processors use always-on 256-bit DRAM encryption. The data on the Nitro SSDs is encrypted using ephemeral keys for maximum security. The Nitro System, a combination of dedicated hardware and lightweight hypervisor, delivers nearly all the compute and memory resources of the host hardware to your instances for better overall performance. The Nitro System also features the Nitro Security Chip, which provides a secure cloud platform with a minimized attack surface and prohibited administrative access to prevent human error and tampering. These instances also support encrypted Amazon Elastic Block Store (EBS) volumes by default.
- **With Im4gn and Is4gen instances, optimize your infrastructure based on your workload needs for the best price performance or for the lowest cost per TB of storage on EC2.** Im4gn and Is4gen instances add to the broadest and deepest portfolio of instances in the cloud and provide six different instance sizes that feature varying amounts of vCPU, memory, networking, and storage.
- **Popular Linux operating systems, such as Amazon Linux 2, Red Hat Enterprise Linux, SUSE, and Ubuntu, support AWS Graviton-based instances.** Many popular applications and services for security, monitoring and management, containers, and CI/CD from AWS and AWS software Partners also support AWS Graviton-based instances. The AWS Graviton Ready program provides certified solutions from Partner software vendors that can be used on AWS Graviton-based instances.

Resources:

[Webpage](#) | [Blog post](#)

Amazon EC2 M1 Mac instances (Preview)

Develop, build, test, and sign Apple apps on Amazon Elastic Compute Cloud (EC2)

What is it?

Amazon EC2 Mac instances allow you to run on-demand macOS workloads in the cloud for the first time, extending the flexibility, scalability, and cost benefits of AWS to all Apple developers. With EC2 Mac instances, developers creating apps for iPhone, iPad, Mac, Apple Watch, Apple TV, and Safari can provision and access macOS environments within minutes, dynamically scale capacity as needed, and benefit from the pay-as-you-go pricing of AWS. x86-based EC2 Mac instances are built on Apple's Mac mini computers featuring Intel Core i7 processors and are powered by the [AWS Nitro System](#). They offer a choice of macOS Mojave (10.14), macOS Catalina (10.15), macOS Big Sur (11), and macOS Monterey (12) as Amazon Machine Images (AMIs). EC2 M1 Mac instances (now in Preview) are built on Apple M1 Mac mini computers and powered by the [Nitro System](#). They deliver up to 60% better price performance over x86-based EC2 Mac instances for iOS and macOS application build workloads. EC2 M1 Mac instances enable ARM64 macOS environments for the first time on AWS, and support macOS Big Sur (11) and macOS Monterey (12) as AMIs.

Benefits:

- **Quickly provision macOS environments.** Time and resources previously spent building and maintaining on-premises macOS environments can now be refocused on building creative and useful apps. Development teams can seamlessly provision and access macOS compute environments on demand to enjoy convenient, distributed testing and fast app builds—bringing additional choice to developers so they can use Mac as their trusted platform, whether on premises or in the cloud. EC2 Mac instances off-load the heavy lifting that comes with managing infrastructure to AWS, which means Apple developers can focus entirely on building their applications.
- **Reduced costs.** EC2 Mac instances allow developers to launch macOS environments within minutes, adjust provisioned capacity as needed, and pay only for actual usage with the pay-as-you-go pricing of AWS. Developers save money since they need to pay only for the systems in use. For example, more capacity can be used when building an app, and less capacity when testing.
- **Simplify rearchitecting, building, and testing for Apple silicon.** EC2 M1 Mac instances (now in Preview) enable Apple silicon and native ARM64 macOS environments for the first time on AWS to develop, build, test, deploy, and run Apple applications. Developers rearchitecting their macOS applications to natively support ARM64 architecture on Apple silicon can now enjoy faster builds with bare-metal performance, as well as convenient distributed testing without having to procure, install, manage, patch, and upgrade physical Apple silicon infrastructure.
- **Extend your toolkits.** EC2 Mac instances provide developers with seamless access to dozens of AWS services so they can more easily and efficiently collaborate with team members, as well as develop, build, test, analyze, and improve their apps. Similar to other EC2 instances, you can easily use EC2 Mac instances together with AWS services and features like Amazon Virtual Private Cloud (VPC) for network security, Amazon Elastic Block Storage (EBS) for expandable storage, Elastic Load Balancing (ELB) for distributing build queues, Amazon FSx for scalable file storage, and AWS Systems Manager for configuring, managing, and patching macOS environments.

Availability:

EC2 x86 Mac instances are available across the following AWS Regions:

US East (N. Virginia, Ohio), US West (Oregon), Europe (Ireland, Frankfurt, London, Stockholm), and Asia Pacific (Singapore, Seoul, Tokyo, Mumbai, and Sydney)

EC2 M1 Mac instances in Preview are available in the US East (N. Virginia) and US West (Oregon) Regions.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Private 5G

Deploy, operate, and scale a private cellular network

What is it?

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 enable 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.

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.

Benefits:

- Connect millions of devices and machines with the low latency and high bandwidth of a private 5G network.
- Get your network up and running in days with no long planning cycles or complex integrations, and with an automated setup.
- Secure your network with granular access controls for all connected devices, integrated with existing IT policies.
- Scale your network capacity on demand or add devices with a few clicks, and pay only for the capacity and throughput you use.

Use cases:

- **Run a smart manufacturing facility.** Connect devices, sensors, machines, and legacy equipment within manufacturing facilities to track assembly-line production in real time with reliable data linkage across connected machines.
- **Enable business-critical applications.** Support a range of low-latency use cases such as AR/VR applications for design engineering, image analysis during medical procedures, and autonomous guided vehicles at fulfillment centers.
- **Deliver reliable campus connectivity.** Provide high-quality, reliable connectivity at event venues, college campuses, and enterprise facilities. Support data and video communications, security, logistics, and more.

Availability:

Currently in Preview.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon EC2 C7g instances

Best price performance for compute-intensive workloads on Amazon Elastic Compute Cloud (EC2)

What is it?

Amazon EC2 C7g instances, powered by the latest-generation AWS Graviton3 processors, provide the best price performance on EC2 for compute-intensive workloads. EC2 C7g instances are ideal for high performance computing (HPC), batch processing, electronic design automation (EDA), gaming, video encoding, scientific modeling, distributed analytics, CPU-based machine learning (ML) inference, and ad serving. They offer up to 25% better performance over the sixth-generation AWS Graviton2-based C6g instances. C7g instances are the first in the cloud to feature DDR5 memory, which provides 50% higher memory bandwidth compared to DDR4 memory—allowing high-speed access to data in memory.

Benefits:

- **EC2 C7g instances deliver up to 25% better performance over AWS Graviton2-based C6g instances.** They're ideal for a large number of compute-intensive applications, such as HPC, video encoding, gaming, and CPU-based ML inference, that are built on Linux.
- **C7g instances are powered by AWS Graviton3 processors and built on the AWS Nitro System.** AWS Graviton3 processors offer enhanced security with always-on memory encryption, dedicated caches for every vCPU, and support for pointer authentication. The Nitro System is a combination of dedicated hardware and a lightweight hypervisor that delivers isolated multi-tenancy, private networking, and fast local storage. C7g instances also support encrypted Amazon Elastic Block Store (EBS) volumes by default.
- **Many Linux operating systems support AWS Graviton-based instances, including Amazon Linux 2, Red Hat Enterprise Linux, SUSE, and Ubuntu.** Many popular applications and services for security, monitoring and management, containers, and CI/CD from AWS and software partners also support AWS Graviton-based instances. The AWS Graviton Ready program offers certified solutions from AWS Partner software vendors that can be used on AWS Graviton-based instances.

Use cases:

EC2 C7g instances are ideal for HPC, batch processing, EDA, gaming, video encoding, scientific modeling, distributed analytics, CPU-based ML inference, and ad serving.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon EC2 G5g instances

Best price performance on Amazon Elastic Compute Cloud (EC2) for Android game streaming

What is it?

Amazon EC2 G5g instances are powered by AWS Graviton2 processors and feature NVIDIA T4G Tensor Core GPUs to provide the best price performance on EC2 for graphics workloads, such as Android game streaming. They're the first Arm-based instances on a major cloud to feature GPU acceleration.

With EC2 G5g instances, game streaming customers can run Android games natively on Arm-based instances, encode the rendered graphics, and stream the game over network to a mobile device. This helps reduce development effort by eliminating the need for cross-compilation or emulation, and lowers the cost per stream per hour by up to 30%.

The G5g instances are also a cost-effective platform for machine learning (ML) inference. They're ideal for deploying deep learning (DL) applications that need access to NVIDIA GPUs and their associated artificial intelligence (AI) libraries.

Benefits:

- **With EC2 G5g instances, game streaming customers can run Android games natively on Arm-based instances, encode the rendered graphics, and stream the game over the network to a mobile device,** without the need for cross-compilation or emulation. This helps simplify development effort and reduce time to market, while lowering the cost per stream per hour by up to 30%.
- **G5g instances are powered by AWS Graviton2 processors and feature NVIDIA T4G Tensor Core GPUs to provide cost-effective ML inference.** This helps reduce the infrastructure cost of running DL models in production and allows developers and businesses to deploy DL capabilities more pervasively in their applications.
- **G5g instances are powered by AWS Graviton2 processors and built on the AWS Nitro System.** AWS Graviton2 processors feature always-on 256-bit DRAM encryption and 50% faster per core encryption performance compared to first-generation AWS Graviton. The Nitro System is a combination of dedicated hardware and lightweight hypervisor that delivers nearly all the compute and memory resources of the host hardware to your instances for better overall performance and security. G5g instances also support encrypted Amazon Elastic Block Store (EBS) storage volumes by default.
- **Popular Linux operating systems, such as Red Hat Enterprise Linux, SUSE, and Ubuntu, support AWS Graviton processors,** which are based on the 64-bit Arm architecture. Many popular applications and services for security, monitoring and management, containers, and CI/CD from AWS and independent software vendors (ISVs) also support AWS Graviton2-based instances.

Use cases:

- Best price performance for Android game streaming
- Cost-effective ML inference

Resources:

[Webpage](#) | [Blog post](#)

Amazon EC2 Trn1 instances

Best price performance for training deep learning (DL) models in the cloud

What is it?

Amazon Elastic Compute Cloud (EC2) Trn1 instances deliver the best price performance for training DL models in the cloud for use cases such as natural language processing (NLP), computer vision, search, recommendation, ranking, and more. Trn1 instances are powered by AWS Trainium, the second machine learning (ML) chip built by AWS that is optimized for high-performance DL training.

Trn1 instances support up to 16 AWS Trainium accelerators, up to 800 Gbps of Elastic Fabric Adapter (EFA) networking bandwidth, and 768 GB/s of ultra-high-speed, NeuronLink connectivity.

Trn1 instances are deployed on Amazon EC2 UltraClusters consisting of tens of thousands of AWS Trainium accelerators to rapidly train even the most complex DL models with trillions of parameters.

Developers can get started quickly on Trn1 instances using the AWS Neuron SDK and easily train models using leading ML frameworks.

Benefits:

- Trn1 instances are powered by AWS Trainium accelerators that are purpose built for ML training to deliver the best price performance for training DL models in the cloud.
- Deploy Trn1 instances on EC2 UltraClusters to scale model training to 10,000+ accelerators interconnected with petabit-scale networking for the fastest ML training on EC2.
- Easily start with Trn1 instances using the AWS Neuron SDK, which is integrated with leading ML frameworks such as PyTorch and TensorFlow, and continue using existing ML workflows with minimal code change.
- Trn1 instances are built on the AWS Nitro System, a combination of dedicated hardware and lightweight hypervisor that provides you with a rich collection of flexible building blocks to assemble the compute, storage, memory, and networking resources you need for better overall performance and security.

Use cases:

Developers can use EC2 Trn1 instances to train ML models across a broad set of applications, such as autonomous vehicles, virtual personal assistants, sentiment analysis, speech recognition, fraud detection, and forecasting.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Compute Optimizer resource efficiency scoring

Quickly identify and prioritize top optimization opportunities

What is it?

[AWS Compute Optimizer](#) now helps you quickly identify and prioritize top optimization opportunities through two new sets of dashboard-level metrics: savings opportunity and performance improvement opportunity.

Savings opportunity metrics

These metrics quantify the Amazon Elastic Compute Cloud (EC2), Amazon Elastic Block Store (EBS), and AWS Lambda monthly savings you can achieve at the account, resource-type, and resource levels by adopting Compute Optimizer recommendations. You can use these metrics to evaluate and prioritize cost-efficiency opportunities, as well as monitor your cost efficiency over time.

Performance improvement opportunity metrics

These metrics quantify the percentage and number of under-provisioned resources at the account and resource-type levels. You can use these metrics to evaluate and prioritize performance improvement opportunities that address resource bottleneck risk.

Benefits:

Quickly identify resources that have cost reduction or performance improvement opportunities through two new dashboard-level metrics: savings opportunity and performance improvement opportunity.

Use cases:

- Central infrastructure manager can see which account(s) have the highest potential savings available or the largest number of under-provisioned resources to prioritize their engagement.
- Engineering owner can see which resource type has the largest savings to help prioritize their recommendation evaluation.
- Engineering owner can see which resource type has the largest number of under-provisioned resources to help prioritize their recommendation evaluation.
- Engineering owner can see which resource type has the most severely under-provisioned (or at-risk) resources and prioritize evaluation of the resource type where those resources are concentrated, or prioritize the individual resources with the highest risk across resources types.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Local Zones

Over 30 new Local Zones in major cities around the world

What is it?

At re:Invent 2021, we announced we will be launching over 30 new Local Zones in major cities around the world. These new Local Zones will be made available starting in 2022 in over 21 countries—including Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, Colombia, Czech Republic, Denmark, Finland, Germany, Greece, India, Kenya, Netherlands, Norway, Philippines, Poland, Portugal, and South Africa—and will join 16 Local Zones across the US, allowing you to serve your end users around the world with even lower latency.

Benefits:

- **Low latency to local end users.** Local Zones place compute, storage, database, and other select AWS services closer to end users, allowing you to open up new possibilities and deliver innovative applications and services that require single-digit millisecond latencies for more end users.
- **Consistent AWS experience.** Local Zones allow you to use the same AWS infrastructure, services, APIs, and tools you're familiar with in the cloud. Applications have fast, secure, and seamless access to the full breadth of services in the parent Region.
- **Flexible and scalable.** Local Zones are part of AWS Global Infrastructure, thereby offering the benefits AWS Regions offer today, such as elasticity, availability, selection, and low pay-as-you-go pricing. You can start small and scale as your needs grow, and pay only for the resources you use.

Use cases:

- **Run low-latency applications at the edge.** Build and deploy applications close to end users to allow real-time gaming, live streaming, augmented and virtual reality (AR/VR), virtual workstations, and more.
- **Simplify hybrid cloud migrations.** Migrate your applications to a nearby Local Zone, while still meeting the low-latency requirements of hybrid deployment.
- **Address stringent data residency requirements.** We'll help with your efforts to comply with state and local data residency requirements in sectors such as healthcare, financial services, iGaming, and government.

Availability:

Local Zones are generally available in Boston, Chicago, Dallas, Denver, Houston, Kansas City, Las Vegas, Los Angeles, Miami, Minneapolis, New York City (located in New Jersey), Philadelphia, and Portland. With an additional three Local Zones launching in 2021 in Atlanta, Phoenix, and Seattle, you can deliver ultralow-latency applications to end users in cities across the US.

Furthermore, at re:Invent 2021, we announced we will be launching over 30 new Local Zones in major cities around the world. These new Local Zones will be made available starting in 2022 in over 21 countries—including Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, Colombia, Czech Republic, Denmark, Finland, Germany, Greece, India, Kenya, Netherlands, Norway, Philippines, Poland, Portugal, and South Africa—and will join 16 Local Zones across the US, allowing you to serve your end users around the world with even lower latency.

See the [AWS Local Zones](#) page for more information.

Resources:

[Webpage](#)

Containers

[Return to
Table of Contents](#)

Karpenter

Flexible, high-performance Kubernetes cluster auto-scaler that helps improve application availability and cluster efficiency

What is it?

Karpenter is a flexible, high-performance Kubernetes cluster auto-scaler that helps improve application availability and resource utilization. Karpenter launches right-sized compute resources—such as Amazon Elastic Compute Cloud (EC2) instances—in response to changing application load in under a minute. These Amazon EC2 instances are based on the specific needs of a cluster's workloads, such as compute, storage, acceleration, and scheduling requirements. Today, Amazon Elastic Kubernetes Service (EKS) supports clusters using Karpenter on AWS, although Karpenter is designed to work with any conformant Kubernetes cluster.

Benefits:

- **Improves cluster efficiency.** With Karpenter, the capacity needed to efficiently run cluster workloads can be deployed and ready in as little as 15 seconds. As workloads scale, Karpenter works directly with EC2 to automatically add or remove the nodes required, reducing the need for costly over-provisioning and preventing slow, expensive scale-downs.
- **Built for scale.** Karpenter makes cluster scaling decisions in seconds when demand changes so workloads have the compute resources they need to maintain availability and performance, even in the largest Kubernetes clusters.
- **Reduces operational overhead.** Karpenter provisions and scales the exact compute resources that a cluster's workloads need so they have the right compute capacity, right when they need it. By using Karpenter, you can focus on getting the most out of Kubernetes for your organization, instead of worrying about how to provision, manage, or scale compute.

Use cases:

Karpenter works with all kinds of Kubernetes applications, although it performs particularly well for use cases that require quickly provisioning and deprovisioning large numbers of diverse compute resources.

- **Machine learning (ML).** Batch jobs to train ML models, run simulations, or perform complex financial calculations benefit from Karpenter automatically determining the types of compute resources required and then rapidly provisioning and deprovisioning them.
- **"Soft" multi-tenant clusters.** Karpenter works well for multi-tenant Kubernetes clusters with friendly tenants (for example, different teams in the same organization) that have many different types of applications running side by side. By provisioning compute that satisfies the resource requirements of all the cluster's applications simultaneously, without requiring operators to deeply understand each of them, Karpenter makes it easy to maintain availability and improve cluster utilization.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon ECR pull through cache repositories

Retrieve, store, and sync container artifacts stored in publicly accessible container registries.

What is it?

With Amazon Elastic Container Registry (ECR) pull through cache repositories, you can retrieve, store, and sync container artifacts stored in publicly accessible container registries. Amazon ECR offers the high download rates you need and the availability, security, and scale you've come to depend on. With frequent registry syncs and no additional tools to manage, pull through cache repositories help you keep container images sourced from public registries up to date.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Databases

[Return to
Table of Contents](#)

Amazon Timestream

Fast, scalable, and serverless time-series database

What is it?

Amazon Timestream is a fast, scalable, and serverless time-series database service for IoT and operational applications that makes it easy to store and analyze trillions of events per day up to 1,000 times faster—at as little as 1/10th the cost of relational databases. Timestream saves you time and cost in managing the lifecycle of time-series data by keeping recent data in memory and moving historical data to a cost-optimized storage tier based on user-defined policies. The Timestream purpose-built query engine lets you access and analyze recent and historical data together, without needing to specify explicitly in the query whether the data resides in the in-memory or cost-optimized tier. Timestream has built-in time-series analytic functions, helping you identify trends and patterns in your data in near real time. Timestream is serverless and automatically scales up or down to adjust capacity and performance, so you don't need to manage the underlying infrastructure, freeing you to focus on building your applications.

Benefits:

- **High performance at low cost.** Timestream is designed to provide interactive and affordable real-time analytics, with up to 1,000 times faster query performance at as little as 1/10th the cost of relational databases. With product features such as scheduled queries, multi-measure records, and data storage tiering, you can process, store, and analyze your time-series data at a fraction of the cost of existing time-series solutions. Timestream can help you derive faster and more affordable insights from your data so you can continue to make more data-driven business decisions.
- **Serverless with auto-scaling.** Timestream is serverless—there are no servers to manage and no capacity to provision, so you can focus on building your applications. Timestream gives you the scale to process trillions of events and millions of queries per day. As your application needs change, it automatically scales to adjust capacity.
- **Data lifecycle management.** Timestream simplifies the complex process of data lifecycle management. It offers storage tiering, with a memory store for recent data and a magnetic store for historical data. Timestream automates the transfer of data from the memory store to the magnetic store based on user-configurable policies.
- **Simplified data access.** With Timestream, you no longer need to use disparate tools to access recent and historical data. The Timestream purpose-built query engine transparently accesses and combines data across storage tiers without the need to specify the data location.
- **Purpose built for time-series data.** You can quickly analyze time-series data using SQL, with built-in time-series functions for smoothing, approximation, and interpolation. Timestream also supports advanced aggregates, window functions, and complex data types such as arrays and rows.
- **Always encrypted.** Timestream helps ensure your time-series data is always encrypted, whether at rest or in transit. Timestream also allows you to specify an AWS Key Management Service (AWS KMS) customer-managed key (CMK) for encrypting data in the magnetic store.

Use cases:

- **IoT applications.** Timestream makes it possible for you to quickly analyze time-series data generated by IoT applications using built-in analytic functions such as smoothing, approximation, and interpolation. For example, a smart home device manufacturer can use Timestream to collect motion or temperature data from the device sensors, interpolate to identify the time ranges without motion, and alert consumers to take actions such as turning down the heat to save energy.
- **DevOps applications.** Timestream is ideal for DevOps solutions that monitor health and usage metrics and analyze data in real time to improve performance and availability. For example, with Timestream, you can collect and analyze operational metrics such as CPU/memory utilization, network data, and IOPS to monitor health and optimize instance usage.
- **Analytic applications.** Timestream allows you to easily store and analyze data at scale. For example, with clickstream data, you can use Timestream to store and process the incoming and outgoing web traffic for your applications. Timestream also provides aggregate functions to analyze this data and get insights such as path to purchase and shopping-cart abandonment rate.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon DevOps Guru for RDS

Amazon DevOps Guru for RDS, powered by machine learning (ML) and available for all Amazon Relational Database Service (RDS) engines, helps make it easy to improve an application's operational performance and availability

What is it?

DevOps Guru for RDS is a new ML-powered capability empowering developers and DevOps engineers to quickly detect, diagnose, and remediate a wide variety of database-related issues on Amazon RDS. When DevOps Guru for RDS detects a database-related issue (such as resource overutilization, or misbehavior of certain SQL queries), the service immediately notifies you and provides diagnostic information, details on the extent of the problem, and intelligent recommendations to quickly resolve the issue. With DevOps Guru for RDS, you can monitor your databases for performance bottlenecks and operational issues—no manual setup, ML expertise, or deep database expertise required. To get started, turn on Amazon RDS Performance Insights on the Amazon RDS console and navigate to the DevOps Guru console to enable the service for your Amazon Aurora resources, other supported resources, or your entire account.

DevOps Guru for RDS currently supports Amazon Aurora MySQL-Compatible Edition and Amazon Aurora PostgreSQL-Compatible Edition, with expanded support for additional RDS database engines coming later.

Benefits:

- **Detect and diagnose RDS database performance bottlenecks and operational issues.** DevOps Guru for RDS continuously analyzes database telemetry, such as database load, database counters, and operating system metrics on the database, to automatically detect and correlate related anomalies and help resolve relational database issues in minutes.
- **Natively integrate with AWS services to automatically receive notifications.** When DevOps Guru for RDS detects a performance bottleneck or operational issue, it displays its findings in the Amazon DevOps Guru console and sends notifications through AWS services, such as Amazon EventBridge and Amazon Simple Notification Service (SNS). This allows developers to automatically manage and take real-time action on performance and operational issues before they become customer-impacting outages.
- **Reduce time to resolution from days to minutes.** With DevOps Guru for RDS, you can quickly understand the cause of a performance or operational problem without searching through hundreds of database metrics. With intelligent recommendations and remediation steps, developers and DevOps engineers can resolve issues in minutes without enlisting help from database experts.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon RDS Custom

Managed database service for applications that need to customize the underlying operating system and database environment

What is it?

Amazon Relational Database Service (RDS) Custom is a managed database service for legacy, custom, and packaged applications that require access to the underlying operating system and database environment. Amazon RDS Custom automates the setup, operation, and scaling of databases in the cloud while providing access to the database and underlying operating system to configure settings, install patches, and enable native features to meet the dependent application's requirements. RDS Custom is available for Amazon Relational Database Service (RDS) for Oracle and for SQL Server.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon DynamoDB Standard-Infrequent Access

Reduce costs by up to 60% for infrequently accessed data

What is it?

The Amazon DynamoDB Standard-Infrequent Access (DynamoDB Standard-IA) table class helps you reduce your DynamoDB costs by up to 60% for tables that store infrequently accessed data. The DynamoDB Standard-IA table class is ideal for use cases that require long-term storage of infrequently accessed data, such as application logs, old social media posts, ecommerce order history, and past gaming achievements.

The DynamoDB Standard-IA table class offers lower storage costs than the DynamoDB Standard tables, making it the most cost-effective option for tables with storage as the dominant table cost. The DynamoDB Standard table class offers lower throughput costs than the DynamoDB Standard-IA table class. DynamoDB Standard is your default table class and the most cost-effective option for the vast majority of workloads. To learn more about DynamoDB Standard-IA pricing, see the [Amazon DynamoDB pricing](#) page.

You can switch between DynamoDB Standard and DynamoDB Standard-IA table classes with no impact on table performance, durability, or availability, and without changing your application code. For more information about using table classes, see the [DynamoDB Developer Guide](#).

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS DMS Fleet Advisor

Accelerate database and analytics migration planning by automating inventory and receiving tailored migration advice

What is it?

AWS DMS Fleet Advisor is a fully managed capability of AWS Database Migration Service (AWS DMS) that automates migration planning and helps you migrate database and analytics fleets to the cloud at scale with minimal effort. To accelerate migrations, DMS Fleet Advisor automatically inventories and assesses your on-premises database and analytics server fleet and identifies potential migration paths. AWS DMS helps you confidently migrate your databases and analytics systems to AWS with virtually no downtime.

DMS Fleet Advisor discovers and analyzes the same source databases supported on DMS, including Oracle, Microsoft SQL Server, MySQL, MongoDB, and more. DMS Fleet Advisor delivers results in a few hours, instead of weeks or even months, without using third-party tools or hiring migration experts. DMS Fleet Advisor is a free capability of DMS.

Benefits:

- **Build a migration plan in hours instead of weeks.** Automatically build database and analytics inventories, analyze them, and create a customized migration plan with virtually no downtime in a few hours, instead of weeks or months.
- **Save on the costs associated with migration planning and migrating workloads.** Avoid investing in expensive migration experts and third-party tools by automating your inventory and migration planning. Easily start with just a few clicks in the AWS Management Console.
- **Identify databases to migrate at scale with minimal effort.** Discover and analyze fleets of database and analytics servers to identify potential AWS migration targets and migrate your fleet to the cloud with minimal effort.

Use cases:

- **Achieve operationally excellent cloud migrations.** Ensuring operational excellence during a cloud migration requires a detailed migration plan. With DMS Fleet Advisor, automate the detailed inventorying and analysis of your on-premises workloads and easily understand the effort and organizational changes required to develop a customized migration plan.
- **Modernize your data.** Build a business case to modernize your database and analytics workloads with minimal effort with DMS Fleet Advisor. Gain a simple view of your application's applicable migration targets within AWS and understand the complexity of migrating to each target.
- **Build a proof of concept.** Quickly build a proof of concept and begin testing the critical functionality of your applications by automatically inventorying and analyzing your existing database and analytic servers. Validate your migration and share findings as a report to quickly obtain stakeholder support.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

DevOps and developer productivity

[Return to
Table of Contents](#)

AWS SDK for Swift (Developer Preview)

Develop and deploy applications with the AWS SDK for Swift, which makes it easy to call AWS services using idiomatic Swift APIs

What is it?

The SDK for Swift simplifies the use of AWS services by providing a set of libraries that are consistent and familiar for Swift developers. All AWS SDKs support API lifecycle considerations, such as credential management, retries, data marshaling, and serialization. Visit [aws-sdk-swift](#) on GitHub for AWS-focused open-source Swift libraries.

The SDK for Swift is written with multi-platform support in mind. It supports iOS, macOS, and Linux environments today, with support for other Swift-supported platforms coming in future releases. It takes advantage of modern Swift language features, such as `async/await`, inferred types, and others. In addition, the SDK for Swift has a modular architecture with a separate package published for each service, allowing you to reduce application bundle size and improve application performance by importing only the packages your application requires.

Benefits:

- **Multi-platform support.** The SDK for Swift is written with multi-platform support in mind. It supports iOS, macOS, and Linux environments today, with support for other Swift-supported platforms coming in future releases.
- **Familiar Swift experience.** The SDK for Swift takes advantage of modern Swift language features, such as `async/await`, inferred types, and others. The SDK also integrates with popular Swift libraries, such as Swift Log and Swift Collections, and it uses Swift Package Manager (SPM) for compiling and distribution.
- **Modularity.** The SDK for Swift has a modular architecture with a separate package published for each service. This allows you to reduce application bundle size and improve application performance by importing only the packages your application requires.

Use cases:

- Develop and deploy applications with the SDK for Swift.
- Easily call AWS services using idiomatic Swift APIs.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS SDK for Kotlin (Developer Preview)

Develop and deploy applications with the AWS SDK for Kotlin, which makes it easy to call AWS services using idiomatic Kotlin APIs

What is it?

The SDK for Kotlin simplifies the use of AWS services by providing a set of libraries that are consistent and familiar for Kotlin developers. All AWS SDKs support API lifecycle considerations, such as credential management, retries, data marshaling, and serialization. Visit [aws-sdk-kotlin](#) on GitHub for AWS-focused open-source Kotlin libraries.

The SDK for Kotlin is multi-platform, with support for JVM and Android environments available today and other Kotlin-supported platforms coming in future releases. It takes advantage of modern Kotlin language features, such as domain-specific language (DSL) builders, which allows for concise and readable configuration expressions. SDK for Kotlin also helps you reduce application bundle size and improve application performance by importing only the packages your application requires.

Benefits:

- **Multi-platform support.** The AS SDK for Kotlin is multi-platform, with support for JVM and Android environments available today and other Kotlin-supported platforms coming in future releases.
- **Familiar Kotlin experience.** The SDK for Kotlin takes advantage of modern Kotlin language features, such as DSL builders, which allows for concise and readable configuration expressions. Every API call is asynchronous and runs efficiently in coroutines, yielding to other code while waiting on IO and optimizing the use of compute resources.
- **Modularity.** The SDK for Kotlin has a modular architecture with a separate package published for each service. This allows you to reduce application bundle size and improve application performance by importing only the packages your application requires.

Use cases:

- Develop and deploy applications with the SDK for Kotlin.
- Easily call AWS services using idiomatic Kotlin APIs.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Cloud Development Kit v2

Define cloud infrastructure using familiar programming languages

What is it?

The AWS Cloud Development Kit (AWS CDK) is an open-source software development framework to define your cloud application resources using familiar programming languages. Provisioning cloud applications can be a challenging process that requires you to perform manual actions, write custom scripts, maintain templates, or learn domain-specific languages. AWS CDK uses the familiarity and expressive power of programming languages for modeling your applications. It provides high-level components called constructs that preconfigure cloud resources with proven defaults, so you can build cloud applications with ease. CDK provisions your resources in a safe, repeatable manner through AWS CloudFormation. It also allows you to compose and share your own custom constructs that incorporate your organization's requirements, helping you expedite new projects. For customers who prefer Terraform, the Cloud Development Kit for Terraform (CDKTF) provides the CDK constructs for defining Terraform HCL state files in TypeScript and Python. For Kubernetes users, the Cloud Development Kit for Kubernetes (CDK8s) project allows you to use CDK constructs for defining Kubernetes configuration in TypeScript, Python, and Java. CDK8s can be used to define Kubernetes infrastructure running anywhere, and it can be used with the CDK Amazon Elastic Kubernetes Service (EKS) construct library.

Benefits:

- **Easier cloud onboarding.** CDK accelerates your onboarding to AWS because there are few new things to learn. CDK allows you to use your existing skills and tools, and apply those to the task of building cloud infrastructure. It also provides high-level components that preconfigure cloud resources with proven defaults, helping you build on AWS without needing to be an expert.
- **Faster development process.** CDK gives you the expressive power of programming languages for defining infrastructure. Familiar features such as objects, loops, and conditions accelerate your development process. You can also use CDK with your integrated development environment (IDE) to take advantage of existing productivity tools and testing frameworks.
- **Customizable and shareable.** With CDK, you can design your own reusable components that meet your organization's security, compliance, and governance requirements. Just like with any other software library, you can easily share components around your organization, allowing you to rapidly bootstrap new projects with best practices by default.
- **No context switching.** CDK allows you to build your cloud application without leaving your IDE. You can write your runtime code and define your AWS resources with the same programming language. You can visualize your CDK application stacks and resources with the AWS Toolkit for Visual Studio Code.

Use cases:

New use cases supported in this update (CDK v2) include:

- **One package for all stable CDK libraries.** In CDK v1, we partitioned the AWS Construct Library into many small packages, roughly one per service, so developers needed to only download the packages for those services they wanted to use. In CDK v2, we consolidated all the stable AWS Construct Libraries into a single package called `aws-cdk-lib`. Get access to all the stable CDK constructs by installing this package, and third-party construct libraries only need to take a dependency on this package as well.
- **Changes to how we handle experimental classes, methods, and properties.** In CDK v2, we strictly follow semantic versioning and no longer make breaking changes to any APIs in minor version releases. We have introduced a new lifecycle in which new, experimental construct libraries go through an incubation period as a library, completely independent from the main `aws-cdk-lib` library.
- **CDK Watch.** Accelerate your inner-loop development iterations by hot swapping your runtime code in CDK projects. When the only change to your CDK project is your AWS Lambda handler code, your Amazon Elastic Container Service (ECS) task, or your AWS Step Functions state machine, the CDK command line interface (CLI) will skip the AWS CloudFormation deployment step and hot swap the code change directly on the running resource.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS SDK for Rust (Developer Preview)

Develop and deploy applications with the AWS SDK for Rust, which makes it easy to call AWS services using idiomatic Rust APIs

What is it?

The SDK for Rust simplifies the use of AWS services by providing a set of libraries that are consistent and familiar for Rust developers. All AWS SDKs support API lifecycle considerations, such as credential management, retries, data marshaling, and serialization. Visit aws-sdk-rust for AWS-focused open-source Rust libraries.

The SDK for Rust uses modern Rust language features, such as `async/await`, non-blocking IO, and builders, and integrates with popular libraries in the Rust ecosystem. It has a modular architecture and a composable middleware stack used to run each service API operation, and it's engineered to be fast with serializers and deserializers that minimize unnecessary copies and allocations.

Benefits:

- **Familiar Rust experience.** The SDK for Rust uses modern Rust language features, such as `async/await`, non-blocking IO, and builders. The SDK also integrates with popular libraries in the Rust ecosystem, such as [Tokio](#), [Tracing](#), and [Hyper](#), and it's distributed on crates.io.
- **Modularity.** The SDK for Rust has a modular architecture with a separate package published for each service. This allows you to minimize your compile times and binary sizes by compiling only code you actually use.
- **Extensibility.** The SDK for Rust has a composable middleware stack used to run each service API operation. The stack allows you to wrap the SDK middleware in your own middleware to support advanced use cases.
- **Performance.** The SDK for Rust is engineered to be fast with serializers and deserializers that minimize unnecessary copies and allocations. This reduces CPU and memory utilization by the SDK, freeing up more of these resources for your application.

Use cases:

- Develop and deploy applications with the SDK for Rust.
- Easily call AWS services using idiomatic Rust APIs.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Construct Hub

Find and use open-source cloud development kit (CDK) libraries

What is it?

Construct Hub is a place to discover CDK construct libraries, which are building blocks for cloud applications published by the open-source community, AWS, and other cloud services. CDK is a software development framework for defining cloud infrastructure using familiar programming languages.

For more information on Construct Hub, visit the [website](#).

Benefits:

- Find libraries for AWS Cloud Development Kit (AWS CDK), which generates AWS CloudFormation templates; Cloud Development Kit for Terraform (CDKTF), which generates Terraform configuration files; and Cloud Development Kit for Kubernetes (CDK8s), which generates Kubernetes manifests.
- Define, test, and deploy cloud infrastructure using high-level programming languages such as TypeScript, Python, Java, and .NET.
- Find documentation, API references, and code samples to quickly build your application.
- Find construct libraries published by the community and cloud service providers such as Datadog, AWS, MongoDB, Aqua Security, and more.

Use cases:

- Find and use open-source CDK libraries for AWS CDK, CDKTF, and CDK8s from the community, cloud technology providers, and AWS.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Enterprise and Migration

[Return to
Table of Contents](#)**AWS Mainframe Modernization (Preview)**

Migrate, modernize, operate, and run mainframe workloads

What is it?

AWS Mainframe Modernization (Preview) is a set of managed tools providing infrastructure and software for migrating, modernizing, and executing mainframe applications. Mainframe Modernization is a unique platform that allows you to migrate your on-premises mainframe workloads to a managed runtime environment on AWS. Mainframe Modernization enables two popular migration patterns: re-platforming and automated refactoring.

Benefits:

- Easily migrate and modernize your applications to eliminate 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 on the Mainframe Modernization environment with no upfront costs.

Use cases:

- **Migrate mainframe workloads.** Migrate faster by refactoring legacy language applications to Java-based services, or re-platforming mainframe applications.
- **Create a migration plan.** Assess your migration readiness and develop a clear strategy for legacy workloads using call graphs, dependency maps, and source searching.
- **Scale DevOps best practices.** Maintain, enhance, run, and operate enterprise applications to follow DevOps best practices and reduce time to market.

Availability:

For more information, see the [AWS Regional Services](#) page.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Migration Hub Refactor Spaces (Preview)

The one destination for migration and modernization

What is it?

AWS Migration Hub Refactor Spaces (Preview) is the new starting point for incremental app refactor that makes it easy to manage the refactor process while operating in production. Using Migration Hub Refactor Spaces, you can focus on refactoring your applications, not creating and managing the underlying infrastructure that makes refactoring possible. This new AWS Migration Hub feature reduces the business risk of evolving applications into microservices or extending existing applications that cannot be modified with new features written in microservices. Migration Hub Refactor Spaces orchestrates AWS services across multiple accounts to create a refactor environment for incrementally evolving an application, helping you realize value earlier.

Benefits:

Migration Hub Refactor Spaces simplifies application refactoring by:

- Reducing the time to set up a refactor environment (cloud anti-corruption layer)
- Reducing the complexity to iteratively extract capabilities as new microservices and reroute traffic from old to new (Strangler Fig pattern)
- Simplifying the management of legacy apps and microservices as a single application with flexible routing control, isolation, and centralized management
- Helping development teams achieve and accelerate tech and deployment independence by simplifying development, management, and operations while apps are changing

Use cases:

Migration Hub Refactor Spaces addresses a common pair of practical problems when transforming applications: setting up an infrastructure for application refactoring and operating evolving applications at scale.

Migration Hub Refactor Spaces helps you combine existing applications and microservices into a single application, while allowing different approaches for architecture and technology, team alignment, and the process between parts.

Using Migration Hub Refactor Spaces, transform legacy applications or extend them with microservices that run on any AWS compute target, such as Amazon Elastic Compute Cloud (EC2), Amazon Elastic Container Service (ECS), Amazon Elastic Kubernetes Service (EKS), AWS Fargate, and AWS Lambda.

Save significant time with Migration Hub Refactor Spaces by creating an infrastructure for application refactoring in minutes.

Availability:

For more information, see the [AWS Regional Services](#) page.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Front-end web and mobile development

[Return to
Table of Contents](#)**AWS Amplify Studio**

Extend AWS Amplify with a visual development environment for creating full-stack apps with minimal coding

What is it?

AWS Amplify Studio is a visual development environment that lets developers easily build and ship complete web and mobile apps in hours instead of weeks. With Amplify Studio, you can quickly build an app backend, create rich user interface (UI) components, and connect a UI to the backend with minimal coding. Amplify Studio exports all UI and infrastructure artifacts as code, so you can maintain full control over your app design and behavior. Ship faster and scale smoothly—with no cloud or AWS expertise.

Benefits:

- **Quickly create feature-rich UIs.** Use a visual interface and library of prebuilt UI components to create and customize a UI for your app with minimal coding.
- **Easily ship a full-stack app.** Set up your frontend UI and corresponding backend on AWS to complete a full-stack app with a few clicks.
- **Increase productivity.** Automate the UI and infrastructure boilerplate code so you can focus on the business logic.
- **Improve designer-developer collaboration.** Streamline the design handoff process and mitigate costly, time-consuming errors.

Availability:

Available in all AWS markets.

Resources:

[Webpage](#) | [Breakout session recording and slides](#)

Games

[Return to
Table of Contents](#)

Open 3D Engine

Scaling 3D development

What is it?

Open 3D Engine (O3DE) is an open-source, real-time 3D development engine for building high-fidelity games and simulations. O3DE provides a multi-threaded renderer, a physics system, networking capabilities, and a suite of content-authoring tools for building interactive 3D experiences with full source control and nonrestrictive licensing. With O3DE, you can provision and integrate with AWS services to provide multiplayer servers, authentication and identity systems, and asset management.

O3DE is well suited to support a broad spectrum of use cases, from AAA-quality games to digital twins and cinema-quality 3D worlds. The engine's modular architecture and ability to be used as an SDK provide complete flexibility to customize core engine capabilities to suit genre- and industry-specific application requirements.

Benefits:

- **O3DE is architected for modularity and flexibility.** Swap out subsystems or easily add integrations to give your games and simulations the capabilities they need without the features they don't.
- **Add multiplayer scaling, analytics, cloud storage, and real-time data sources** to your game directly from the engine with the O3DE AWS Integration Gem.
- **O3DE is free and includes full source.** There are no seat fees, subscription fees, or requirements to share revenue. Since it's licensed under Apache 2.0, you have complete control of your engine technology.

Use cases:

- **Create high-fidelity 3D games and simulations.** Use the O3DE multi-threaded photorealistic renderer, extensible 3D content editor, data-driven character animation system, and real-time physics engine to create interactive applications and experiences.
- **Integrate with AWS services for cloud-connected experiences.** O3DE is a connected engine with built-in AWS integrations that provide cloud services for asset storage, player authentication, application deployment, and multi-player backend services.

Resources:

[Webpage](#) | [Breakout session recording and slides](#)

AWS IoT TwinMaker (Preview)

Optimize operations by easily creating digital twins of real-world systems

What is it?

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.

Benefits:

- Use your existing IoT, video, and enterprise application data where it already lives—without needing to re-ingest or move the data to another location.
- Save time with an automatically generated knowledge graph that binds your data sources to virtual replicas of physical systems to accurately model real-world environments.
- Get an immersive 3D view of your systems and operations to optimize efficiency, increase production, and improve performance.

Use cases:

- **Improve field operations in manufacturing plants.** Quickly pinpoint and address equipment and process anomalies from the plant floor to improve worker productivity and efficiency.
- **Increase equipment uptime in remote facilities.** Remotely diagnose equipment issues with immediate access to all relevant operational data, allowing for faster decision-making.
- **Enhance the tenant experience in commercial buildings.** Monitor live and historical temperature, occupancy, and air-quality data within rooms and open spaces to improve occupant comfort.

Availability:

Available in Preview in the following Regions:
US East (N. Virginia), US West (Oregon), Europe (Ireland), Asia Pacific (Singapore)

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS IoT FleetWise (Preview)

Easily collect, transform, and transfer vehicle data to the cloud in near real time

What is it?

AWS IoT FleetWise is a managed service that makes it easier and more cost effective for automakers to collect, transform, and transfer vehicle data to the cloud in near real time. With AWS IoT FleetWise, automakers can easily collect and organize data from the unique data format present in their vehicles (regardless of make, model, or options) and standardize the data format for easy analysis in the cloud. AWS IoT FleetWise helps you efficiently transfer data to the cloud in near real time using intelligent filtering capabilities. These capabilities allow developers to reduce network traffic by selecting the data to transfer and defining rules for when to transfer it, based on parameters like weather conditions, location, or vehicle type. Once the data is in the cloud, you can use it for use cases like remotely diagnosing issues in individual vehicles, analyzing vehicle fleet health to help prevent potential recalls or safety issues, and using analytics and machine learning (ML) to improve advanced technologies such as autonomous driving and advanced driver assistance systems (ADAS).

Benefits:

- Access standardized, fleet-wide vehicle data without the need to develop custom data collection systems.
- Reduce costs and transfer data more efficiently with intelligent filtering, which transfers only high-value data signals to the cloud.
- Surface vehicle health data in near real time to detect and mitigate issues faster, help avoid large recalls, and maintain customer trust.

Use cases:

- **Train computer vision (CV) models.** Train autonomous vehicles (AVs) and ADAS using camera data collected from a fleet of production vehicles.
- **Help minimize warranty claims and recalls.** Use near real-time data to proactively detect and mitigate fleet-wide quality issues.
- **Improve electric vehicle (EV) range estimates.** Improve EV battery-range estimates with crowdsourced environmental data, such as weather and driving conditions, from nearby vehicles.
- **Notify drivers of changing road conditions.** Collect select data from nearby vehicles and use it to notify drivers of changing road conditions, such as lane closures or construction.

Availability:

Available in the US East (N. Virginia) and Europe (Frankfurt) Regions. For more information, see the [AWS Regional Services](#) page.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS IoT Greengrass Software Catalog

A collection of AWS IoT Greengrass software components developed by the AWS IoT Greengrass community

What is it?

AWS IoT Greengrass Software Catalog is a collection of Greengrass components developed by the Greengrass community. Instead of developing device components for each required capability, you can now easily install, use, and modify components from a list of prebuilt software components on GitHub to kick-start your IoT edge application.

For example, for a security monitoring solution, you can use the Amazon Kinesis Video Streams component to ingest audio and video streams from Real Time Streaming Protocol (RTSP) cameras connected to a Greengrass core device. The data can then be streamed to a local monitoring platform or sent to the cloud. Alternatively, for real-time analytics and local operations monitoring, you can use the InfluxDB and Grafana components to locally process and visualize data from IoT sensors and edge devices. Since these components are a reference implementation of common patterns, appropriately review and test any functionality before deploying it to your production environments. To get started, visit the [Greengrass Software Catalog on GitHub](#).

Resources:

[Webpage](#) | [Blog post](#)

AWS IoT ExpressLink (Preview)

Quickly and easily develop secure IoT devices

What is it?

AWS IoT ExpressLink is connectivity software that powers a range of hardware modules developed and offered by AWS Partners, such as Espressif Systems, Infineon Technologies, and u-blox. These connectivity modules include AWS-validated software, making it faster and easier for you to securely connect almost any product to the cloud, including medical devices, industrial sensors, and consumer products. AWS IoT ExpressLink is preprogrammed to seamlessly integrate with a range of AWS IoT services and comes pre-provisioned with security credentials, allowing you to off-load the complex work of cloud connectivity to the module and develop secure IoT products in a fraction of the time.

Benefits:

- **Easily transform products into IoT devices.** Developers of all skill levels can now transform their products into IoT devices in a fraction of the time and cost, without the need to merge large amounts of code or have a deep understanding of the underlying implementation.
- **Flexibility to connect any type of device.** AWS Partner modules with AWS IoT ExpressLink are compatible even with resource-constrained devices, such as the tiny processor contained in a coffee machine. This allows developers to connect devices to the cloud while preserving the original application software and avoiding costly redesign.
- **Maintain fleet security and health.** AWS Partner modules with AWS IoT ExpressLink provide an out-of-the-box solution to help securely transmit data to and from the cloud through Wi-Fi or cellular protocols. AWS IoT ExpressLink supports over-the-air (OTA) updates to both the module and host processor from the AWS IoT Device Management console so developers can remotely deploy security updates, bug fixes, and new firmware updates to their device fleets, helping to keep the fleets secure over the life of the device.

Use cases:

- **Optimize industrial applications.** Modernize industrial applications without redesigning the application from scratch. AWS IoT ExpressLink offers out-of-the-box connectivity for all types of devices, regardless of resource constraints.
- **Ship new consumer products faster.** Get consumer products to market in a fraction of the time and at a lower cost by delegating the complex work of securely connecting an embedded device to the module.
- **Reinvent smart buildings and cities.** Manage the security and health of your commercial fleet sensors and actuators at scale. Easily integrate them with AWS IoT services to detect anomalies and take action when metrics deviate from expected behavior.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Networking and Content Delivery

[Return to
Table of Contents](#)

AWS Direct Connect SiteLink

Create private network connections between your on-premises locations

What is it?

AWS Direct Connect SiteLink is a feature of AWS Direct Connect. When you enable it at two or more Direct Connect locations or points of presence (PoPs), you can send data between those locations. SiteLink works with both hosted and dedicated connections.

With the SiteLink feature enabled, you can create private network connections to link the offices, data centers, and colocation facilities in your global network. When using SiteLink, data travels over the fast, secure, and reliable AWS global network following the shortest path between over 100 Direct Connect locations around the globe.

The Direct Connect service helps you create private connections between your on-premises networks and your AWS resources by connecting your network directly to Direct Connect locations. Using the new SiteLink feature, you can link your on-premises locations to Direct Connect and send data between them over the shortest path between Direct Connect locations. With over 100 Direct Connect locations around the world, you can create networks that span multiple continents. Once you've connected your on-premises locations to Direct Connect, you can enable (or disable) SiteLink for that location in minutes. SiteLink uses elastic, pay-as-you-go pricing with no long-term commitments.

Learn more by visiting the [AWS Direct Connect](#) product overview page and API [documentation](#). For a more in-depth technical overview, read the [Introducing AWS Direct Connect SiteLink blog](#) post.

Availability:

Available in all commercial Regions except China.

Resources:

[Blog post](#)

Amazon VPC IP Address Manager

Plan, track, and monitor IP addresses for your AWS workloads

What is it?

Amazon Virtual Private Cloud (VPC) IP Address Manager (IPAM) makes it easier for you to plan, track, and monitor IP addresses for your AWS workloads. IPAM automates IP address assignments to your Amazon VPC, removing the need to use spreadsheet-based planning applications. IPAM also enhances your network observability by showing IP usage across multiple accounts and Amazon VPCs in a unified operational view.

Benefits:

- IPAM automates IP assignments across multiple Regions and accounts based on your application's unique networking and security needs.
- IPAM monitors IP addresses across AWS, helping you detect IP overlaps, track IPs in compliance with security policy, and capture IP utilization trends.
- IPAM provides historical data on your IP addresses, helping to accelerate network troubleshooting and audits.

Use cases:

- **Manage a growing inventory of IP addresses.** AWS customers often have to keep track of information associated with tens or hundreds of thousands of IP addresses. Today, most track this information using spreadsheet-based or homegrown IP planning applications that can be error prone and time consuming. IPAM allows you to easily organize all your IP addresses based on their routing and security needs and set simple business rules to govern IP assignments.
- **Quickly onboard and grow applications.** In typical organizations, developers often wait days to get new IP addresses assigned to them by a central administrator team. Using IPAM, administrators can automate IP address assignment to developers, removing delays in onboarding new applications or growing existing applications.
- **Monitor and troubleshoot your network.** Unexpected or unplanned changes to IP address assignments can lead to IP address conflicts, causing resource connectivity issues and application downtime. In addition, due to a lack of proper monitoring, you may be unable to proactively identify and address these issues. Since IPAM can be the single source of truth for IP usage information, it can help you more efficiently perform routine monitoring activities such as tracking IP utilization, troubleshooting, and auditing.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Cloud WAN (Preview)

Easily build, manage, and monitor global wide area networks (WANs)

What is it?

AWS Cloud WAN helps IT network and cloud architects connect their data centers, branch offices, and cloud resources—reducing the burden of an ever-expanding mix of network technologies required to operate, secure, and manage a global network.

Control and configure your network with a central policy document for defining access policies and traffic routing. Once defined in policy, Cloud WAN deploys and configures services on your behalf. Cloud WAN automates routine network management tasks, such as adding new offices and connecting new attachments and virtual private clouds (VPCs).

Benefits:

- Unify your AWS and on-premises networks to reduce complexity.
- Increase security by segmenting your network to isolate sensitive network traffic from everyday data.
- View your entire network on a single dashboard.
- Use the AWS global network to connect your locations and resources.

Use cases:

- **Build a global network.** Use your choice of local network providers to connect to AWS, then use the AWS global network to connect your locations and VPCs.
- **Easily extend your network.** Save time by automating routine networking tasks, such as adding new connections, branch locations, and Amazon Virtual Private Clouds (VPCs).
- **Visualize your network using a single dashboard.** Track network traffic, view the health of your network, improve performance, and minimize downtime.

Availability:

Currently launching in Preview.

For more information on available Regions, please visit the [AWS Regional Services](#) page.

Resources:

[Webpage](#) | [Breakout session recording and slides](#)

Robotics

[Return to
Table of Contents](#)**AWS IoT RoboRunner**

Build applications that help robot fleets work together seamlessly

What is it?

AWS IoT RoboRunner makes it easy to build applications for optimizing fleets of diverse robots. Today, companies manage robots independently in silos, making it difficult to orchestrate and visualize robots from a single system view. Robots from different systems are difficult to integrate into a central management system. As a result, developers find it challenging to build applications that function seamlessly across a fleet of robots. AWS IoT RoboRunner provides a central data repository for storing and using data from different robot management systems and enterprise management systems. Once robots are connected, you can use sample applications and software development libraries to build management applications on top of the centralized data repository. Use AWS IoT RoboRunner to build complex management applications that require robot interoperability, such as task orchestration, and view information in a single, unified display.

Benefits:

- Increase output and reduce operating costs with seamless collaboration between robots from different vendors.
- Use a common application architecture to quickly and securely integrate different robot types and vendors with work management systems.
- Reduce the complexity of building applications for task orchestration, shared space management, and robot collaboration.

Use cases:**Common use cases:**

- **Enable robot interoperability.** Integrate different robot types using a scalable and flexible robotics automation infrastructure that helps you improve operational efficiency.
- **Optimize robot collaboration.** Accelerate material-handling processes and operational flow by optimizing robot collaboration and complete process automation.
- **Manage robots through a single system.** Improve automation efficiency by optimizing task allocation across a diverse fleet of robots, adjusting for real-time issues through a single system view.

Industrial use cases:

- **Improve manufacturing and logistics processes.** Optimize task allocation across a diverse fleet of robots, adjusting for real-time issues, time constraints, and assignments.
- **Improve material-handling processes.** Optimize robot-to-robot collaboration at a system level.
- **Simplify robot modernization efforts.** Reduce complexity with a scalable and flexible robotics automation infrastructure.

Availability:

Available in the US East (N. Virginia) and Europe (Frankfurt) Regions at launch.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon Inspector

Automated and continual vulnerability management at scale

What is it?

The new Amazon Inspector is a vulnerability management solution that's native, easy to deploy, and highly scalable. The new Amazon Inspector takes a unique approach to vulnerability management by offering continual vulnerability monitoring compared to the traditional periodic assessment model. Unlike Amazon Inspector Classic, the solution now covers both Amazon Elastic Compute Cloud (EC2) and container-based workloads, and uses the widely deployed AWS Systems Manager Agent (SSM Agent) to eliminate the need 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 (ECR) to introduce intelligent, cost-efficient, and continual vulnerability assessments of container images. The new service has the low operational overhead and one-click enablement model that has been successful with other AWS security services, such as Amazon GuardDuty and AWS Security Hub.

Amazon Inspector also includes multi-account management by integrating with AWS Organizations, allowing you to get started and maintain Amazon Inspector organization-wide with little to no configuration. In addition, findings now include more actionable detail compared to Amazon Inspector Classic, including improved contextual risk scoring that clarifies how to prioritize and respond to the most critical findings. Suppression filters were also added that allow you to tune the service based on your tolerance. Findings are pushed to Security Hub and Amazon EventBridge to allow integration with security event management and workflow systems, or to trigger automated remediation actions using AWS Step Functions. As findings are addressed or changes are made to resources, systems and images are automatically, intelligently, and cost-efficiently reevaluated to provide you with a continually updated view into the vulnerability state of your resources organization-wide.

Benefits:

- Immediately discover and scan AWS workloads for software vulnerabilities and unintended network exposure with a single click.
- Consolidate your vulnerability management solutions for both Amazon EC2 and Amazon ECR into one fully managed service.
- Use the highly accurate Amazon Inspector risk score to efficiently prioritize your remediation.
- Reduce mean time to remediate (MTTR) vulnerabilities and streamline workflow with EventBridge and Security Hub integrations.

Use cases:

- **Quickly discover vulnerabilities.** Automatically discover and quickly route vulnerability findings in near real time to the appropriate teams so they can take immediate action.
- **Prioritize patch remediation.** Use up-to-date common vulnerabilities and exposures (CVE) information, combined with factors such as network accessibility, to create context-based risk scores that help you prioritize and address vulnerable resources.
- **Meet compliance requirements.** Support compliance requirements and best practices for the National Institute of Standards and Technology Cybersecurity Framework (NIST CSF), PCI DSS, and others with Amazon Inspector scans.
- **Identify zero-day vulnerabilities sooner.** Accelerate MTTR by using over 50 sources for vulnerability intelligence to help identify zero-day vulnerabilities quickly.

Availability:

Generally available globally across 19 commercial Regions:

Asia Pacific (Singapore), Asia Pacific (Sydney), Europe (Ireland), US East (N. Virginia), US East (Ohio), US West (Oregon), Asia Pacific (Hong Kong), Asia Pacific (Tokyo), Asia Pacific (Seoul), Asia Pacific (Mumbai), Canada (Central), Europe (Frankfurt), Europe (Stockholm), Europe (Milan), Europe (London), Europe (Paris), Middle East (Bahrain), South America (Sao Paulo), US West (N. California)

Visit the [AWS Regional Services](#) page for details.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Shield automatic application-layer DDoS mitigation

Automatic recommendations, testing, and deployment of custom web application firewall (WAF) rules for each attack

What is it?

Automatic recommendation, testing, and deployment of custom WAF rules for each application-layer attack to reduce the time to mitigate and respond to application availability threats.

Benefits:

- AWS WAF rules are automatically created to mitigate detected DDoS events
- AWS WAF rules are tested against normal traffic to minimize false positives
- AWS WAF rules can be created in Count mode to observe effectiveness before deploying in Block mode
- AWS WAF rules are automatically removed after an event subsides
- No manual intervention required

Resources:

[Blog post](#) | [Breakout session recording and slides](#)

Storage

[Return to
Table of Contents](#)

Amazon FSx for OpenZFS

Fully managed shared storage built on the popular OpenZFS file system

What is it?

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. 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. 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.

Benefits:

- Easily migrate Linux file servers to AWS with file storage that matches or exceeds the capabilities you rely on today and is accessible through the NFS protocol.
- Simplify building and testing applications using OpenZFS instant data snapshots and data cloning.
- Accelerate workloads with high-performance storage, delivering up to 1 million IOPS with just a few hundred microseconds of latency.
- Optimize costs with a few clicks by scaling throughput levels and enabling storage efficiency capabilities such as data compression.

Use cases:

- **Migrate your workloads to AWS seamlessly.** Move workloads running on ZFS or other Linux-based file servers to AWS without modifying application code or how you manage data.
- **Deliver insights faster for data analytics workloads.** Power machine learning (ML), financial analytics, and other data-intensive applications with high-IOPS storage.
- **Accelerate content management.** Deliver the low latency needed to scale file-based web serving and content management applications, including WordPress, Drupal, and Magento.
- **Increase development/test velocity.** Test changes efficiently by cloning application data in seconds, and reduce build times with fast storage for repositories and DevOps solutions, such as Git, Bitbucket, and Jenkins.

Availability:

Available in seven AWS Regions. Visit the [AWS Regional Services](#) page for details.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon S3 Event Notifications with Amazon EventBridge

Build, scale, and deploy event-driven applications based on changes to the data you store on Amazon Simple Storage Service (S3)

What is it?

You can now use Amazon S3 Event Notifications with Amazon EventBridge to build, scale, and deploy event-driven applications based on changes to the data you store on Amazon S3. This makes it easier to act on new data in Amazon S3, build multiple applications that react to object changes simultaneously, and replay past events, all without creating additional copies of objects or developing new software. With increased flexibility to process events and send them to multiple targets, you can now more confidently create new serverless applications with advanced analytics and machine learning (ML) at scale without writing single-use custom code.

You can start sending S3 Event Notifications to EventBridge with a few clicks in the [AWS Management Console](#) or with a single API request. To learn more, visit the [S3 User Guide](#). For pricing, visit the [EventBridge pricing](#) page.

Benefits:

- Use advanced filtering and routing capabilities and send events to 18 targets, including AWS Lambda, Amazon Kinesis, AWS Step Functions, and Amazon Simple Queue Service (SQS).
- Simplify your architecture by matching any attribute, or a combination of attributes, for objects in an S3 event. This allows you to filter events by object size, time range, or other event metadata fields before invoking a target Lambda function or other destination. For example, if millions of audio files are uploaded to an S3 bucket, you can filter for specific files and send an event notification to multiple workflows. Through these multiple workflows, the same event can be used to transcribe an audio file, change its media format for streaming, and apply ML to generate a sentiment score.
- Archive and replay S3 events to reprocess an event in case of an error or if a new application module is added.

Use cases:

- Easily build and deploy event-driven applications without custom software.
- Act on new data in S3 faster, make changes to event-driven applications without interruptions, and debug by replaying past events.
- Use the largest set of destinations on the internet with API destinations.

Availability:

Available in [all commercial AWS Regions](#).

Resources:

[Webpage](#) | [Blog post](#)

Recycle Bin for Amazon EBS Snapshots

Quick recovery of deleted snapshots

What is it?

Recycle Bin for Amazon Elastic Block Store (EBS) snapshots is a new capability that allows you to recover accidentally deleted snapshots. Amazon EBS is an easy-to-use, high-performance block storage service for both throughput- and transaction-intensive workloads at any scale. EBS volumes are durable, block-level storage devices that you attach to Amazon Elastic Compute Cloud (EC2) instances running their workloads. You can back up the data on your EBS volumes to Amazon Simple Storage Service (S3) by taking point-in-time EBS snapshots. With Recycle Bin, deleted EBS snapshots move into a recoverable state where they're available for immediate restore for a period of time that you specify, before being permanently deleted. Activate Recycle Bin with a single request or single click in the AWS Management Console.

Benefits:

- Automatically keep deleted snapshots for a retention period that you specify.
- Use predefined rules to specify retention periods for all snapshots, or specific tagged snapshots. Deleted EBS snapshots in your Recycle Bin must be recovered before your user-specified retention period expires.
- Restrict access to Recycle Bin rules and resources using AWS Identity and Access Management (IAM) roles.

Use cases:

- Recover accidentally deleted snapshots for a user-specified period of time.
- Protect against data loss due to human error or accidental snapshot deletion.

Resources:

[Webpage](#) | [Blog post](#)

Amazon EBS Snapshots Archive

Lower your cost of snapshot storage by up to 75%

What is it?

Amazon Elastic Block Store (EBS) Snapshots Archive is a new storage tier on Amazon EBS that you can use to archive EBS snapshots and save up to 75% in storage costs compared to the standard EBS tier for rarely accessed snapshots. EBS is an easy-to-use, high-performance block storage service for both throughput- and transaction-intensive workloads at any scale. EBS volumes are durable, block-level storage devices that you attach to Amazon Elastic Compute Cloud (EC2) instances running their workloads. Back up the data on your EBS volumes to Amazon Simple Storage Service (S3) by taking point-in-time EBS snapshots. Use EBS Snapshots Archive to archive full, point-in-time copies of EBS snapshots that need to be retained for months or years and do not require frequent or fast retrievals.

Benefits:

- Back up the data on your EBS volumes to Amazon S3 by taking point-in-time EBS snapshots. The EBS Snapshots Archive tier provides up to 75% lower costs compared to storing snapshots in the standard tier. There is no additional cost for EBS users to use the EBS Snapshots Archive.
- The EBS Snapshots Archive provides secure, long-term retention of seldom-accessed snapshots that need to be stored for 90+ days.
- With EBS Snapshots Archive, easily archive and manage EBS snapshots, eliminating the need for custom scripts.

Use cases:

- EBS Snapshots Archive provides a low-cost storage tier for you to archive full, point-in-time copies of EBS snapshots that you need to retain for 90 days or more, for regulatory and compliance reasons or future project releases.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Snowball with Tape Gateway

Migrate petabytes of data stored on physical tapes to AWS using AWS Snowball

What is it?

Migrate physical tapes to virtual tapes. Store your tape backups and archives in the cloud without changing your workflows. Migrate petabytes of physical tape data to the cloud with the AWS Snowball Edge Storage Optimized device. Snowball is available with Tape Gateway built in, so migrating physical tapes to your AWS Storage Gateway virtual tape library (VTL) is quick and easy. Your tape data is migrated to Amazon Simple Storage Service (S3) Glacier Flexible Retrieval (formerly S3 Glacier) or Amazon S3 Glacier Deep Archive to be archived. More details can be found on the [Tape Gateway](#) page.

Availability:

Available in 23 AWS Regions.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon S3 Object Ownership “Bucket owner enforced” setting

Simplifies access management for data stored on Amazon Simple Storage Service (S3)

What is it?

[Amazon S3](#) introduces a new S3 Object Ownership setting, "Bucket owner enforced," that disables access control lists (ACLs) and therefore simplifies access management for data stored on S3. When you apply this bucket-level setting, every object in an S3 bucket is owned by the bucket owner, and ACLs are no longer used to grant permissions. As a result, access to your data is based on policies, including AWS Identity and Access Management (IAM) policies applied to [IAM identities](#), session policies, S3 bucket and access point policies, and virtual private cloud (VPC) endpoint policies. This setting applies to both new and existing objects in a bucket, and you can control access to this setting using IAM policies.

Benefits:

- With the new S3 Object Ownership setting, "Bucket owner enforced," you can easily review, manage, and modify access to your shared datasets on Amazon S3 using only policies.
- Own all objects in an S3 bucket regardless of who uploaded them there.
- The new setting disables ACLs for your bucket and the objects in it, and updates every object so that each object is owned by the bucket owner.
- When you apply this setting, ownership change happens automatically, and applications that write data to the bucket no longer need to specify any ACL. You can enable this setting for existing buckets or when you create a new bucket.

Use cases:

- **Simplify access control.** Forget ACLs. Use policies to manage permission to your bucket.
- **Shared datasets.** Maintain ownership of objects, making it easier to review, manage, and modify access to your shared data based on specific access needs.
- **Data protection.** Maintain ownership over every object in a bucket.

Resources:

[Webpage](#) | [Blog post](#)

Amazon S3 Glacier Instant Retrieval

The lowest-cost archive storage with milliseconds retrieval for rarely accessed data

What is it?

Amazon Simple Storage Service (S3) Glacier Instant Retrieval is an archive storage class that delivers the lowest-cost storage for long-lived data that is rarely accessed and requires retrieval in milliseconds. With Amazon S3 Glacier Instant Retrieval, you can save up to 68% on storage costs compared to using the S3 Standard-Infrequent Access (S3 Standard-IA) storage class, when your data is accessed once per quarter. S3 Glacier Instant Retrieval delivers the fastest access to archive storage, with the same throughput and milliseconds access as the S3 Standard and S3 Standard-IA storage classes. S3 Glacier Instant Retrieval is designed for 99.999999999% (11 nines) data durability and 99.9% availability by redundantly storing data across multiple physically separated Availability Zones (AZs). It's designed for rarely accessed data that still needs immediate access in performance-sensitive use cases like image hosting, online file-sharing applications, medical imaging and health records, news media assets, and genomics.

Benefits:

- **Retrievals in milliseconds.** S3 Glacier Instant Retrieval delivers milliseconds retrieval for archives that need immediate access, such as medical images, news media assets, and user-generated content. For applications that need fast access to your archive data, S3 Glacier Instant Retrieval delivers low-latency and high-throughput performance.
- **Unmatched durability and scalability.** S3 Glacier Instant Retrieval runs on the world's largest global cloud infrastructure with virtually unlimited scalability and is designed for 99.999999999% (11 nines) durability. Data is automatically distributed across multiple AZs that are geographically separated within a Region.
- **Most comprehensive security and compliance capabilities.** S3 Glacier Instant Retrieval offers integration with AWS CloudTrail to log, monitor, and retain storage API call activities for auditing, and it supports three different forms of encryption. S3 also supports security standards and compliance certifications, including SEC Rule 17a-4, PCI-DSS, HIPAA/HITECH eligibility, FedRAMP, EU GDPR, and FISMA. S3 Object Lock enables WORM storage capabilities, helping satisfy compliance requirements for virtually every regulatory agency around the globe.
- **Easily manage your data lifecycle.** Cost-effectively manage data through its lifecycle across the S3 storage classes, which share the S3 APIs. You can use S3 Lifecycle to easily transition your data to lower-cost storage classes, as the data becomes less frequently accessed—without performing any data migration.
- **Low cost.** S3 Glacier Retrieval is the lowest-cost archive storage with milliseconds retrieval for rarely accessed data. It's ideal for data that's accessed once or twice per quarter, and requires immediate access.

Use cases:

- **Media asset workflows.** [Media and entertainment](#) assets, such as video and news footage, require durable storage and can grow to many petabytes over time. Much of this data needs to be available immediately for breaking news events, video rendering, or content development. The S3 Glacier Instant Retrieval storage class allows you to archive older media content affordably while still making it available in milliseconds when it's needed. To save even more on storage costs for media archives that don't require milliseconds access, you can use S3 Glacier Flexible Retrieval (formerly S3 Glacier) or S3 Glacier Deep Archive.
- **Healthcare information archiving.** [Hospital systems](#) need to retain petabytes of patient records (such as LIS, PACS, and EHR) for decades to meet regulatory requirements. The S3 Glacier Instant Retrieval storage class is ideal for medical images or genomics, where milliseconds retrieval is required.
- **Scientific data analytics.** Research organizations generate, analyze, and archive vast amounts of data, such as for genomics or to train machine learning (ML) models. With the S3 Glacier Instant Retrieval storage class, you avoid the complexities of hardware and facility management and capacity planning, with the lowest-cost storage and milliseconds retrieval.

Availability:

The new S3 Glacier Instant Retrieval storage class, the Archive Instant Access tier in S3 Intelligent-Tiering, and the S3 Glacier Flexible Retrieval (formerly S3 Glacier) storage class price reduction and free bulk retrievals are available today in all [AWS Regions](#), including the [AWS GovCloud \(US\) Regions](#), the AWS China (Beijing) Region, operated by Sinnet, and the AWS China (Ningxia) Region, operated by NWCD.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Amazon S3 Intelligent-Tiering: Archive Instant Access tier

Automatic storage cost savings

What is it?

The Amazon Simple Storage Service (S3) Intelligent-Tiering storage class now automatically includes a new Archive Instant Access tier with cost savings of up to 68% for rarely accessed data that needs milliseconds retrieval and high-throughput performance. Amazon S3 Intelligent-Tiering is the first cloud storage that automatically reduces your storage costs on a granular object level by automatically moving data to the most cost-effective access tier based on access frequency—without performance impact, retrieval fees, or operational overhead.

S3 Intelligent-Tiering delivers milliseconds latency and high-throughput performance for frequently, infrequently, and now rarely accessed data in the Frequent, Infrequent, and new Archive Instant Access tiers. You can now use S3 Intelligent-Tiering as the default storage class for virtually any workload, especially data lakes, data analytics, new applications, and user-generated content.

Benefits:

- **Automatic storage cost savings.** S3 Intelligent-Tiering delivers automatic storage costs savings by moving data to lower-cost storage access tiers based on access patterns. The longer data is not accessed, the more you save—up to 68% in the automatic tiers.
- **Retrieval in milliseconds.** In all three of the automatic Access tiers, your data is delivered in milliseconds.
- **Lowest storage cost in the cloud.** By opting in to asynchronous, archive capabilities for objects that are rarely accessed, you can realize storage cost savings of up to 95%, with the lowest storage cost in the cloud.
- **Simple.** There is no operational overhead, no lifecycle charges, no retrieval charges, and no minimum storage duration.
- **Unmatched durability and scalability.** S3 Intelligent-Tiering runs on the world's largest global cloud infrastructure with virtually unlimited scalability and is designed for 99.999999999% (11 nines) durability. Data is automatically distributed across multiple Availability Zones (AZs) that are geographically separated within an Region.

Use cases:

S3 Intelligent-Tiering is the default storage class for virtually any workload, especially data lakes, data analytics, new applications, and user-generated content. Any use case or workload with unknown, unpredictable, or changing access patterns is ideal for S3 Intelligent-Tiering, allowing you to automatically save on storage costs.

Availability:

Amazon S3 Intelligent-Tiering and the new Archive Instant Access tier are available today in all AWS Regions, including the AWS GovCloud (US) Regions, the AWS China (Beijing) Region, operated by Sinnet, and the AWS China (Ningxia) Region, operated by NWCD.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Next generation of Amazon FSx for Lustre file systems

Next-generation Amazon FSx for Lustre provides three improvements to performance and price

What is it?

FSx for Lustre now offers the following three improvements:

- The new file systems provide up to 5x higher throughput per TB (up to 1 GB/s per TB).
- With support for client instances with multiple network interfaces, you can now drive up to 400 Gbps of network bandwidth on Amazon Elastic Compute Cloud (EC2) instances such as P4d and DL1.
- Reduce your cost of throughput by up to 60% compared to previous-generation file systems.

Resources:

[Webpage](#) | [Blog post](#)

Enhanced Amazon S3 integration for Amazon FSx for Lustre

Three new enhancements in the Amazon Simple Storage Service (S3) integration for Amazon FSx for Lustre

What is it?

The Amazon S3 integration for FSx for Lustre now includes the following three enhancements:

- [Supports linking multiple S3 buckets to a file system](#)
- [Automatically exports file updates to S3](#)
- [Automatically updates file system contents as data is deleted and moved on S3](#)

Resources:

[Webpage](#)

Amazon S3 Glacier Flexible Retrieval

Amazon Simple Storage Service (S3) Glacier storage class is now named Amazon S3 Glacier Flexible Retrieval

What is it?

The Amazon S3 Glacier storage class is now named Amazon S3 Glacier Flexible Retrieval. This storage class now also includes free bulk retrievals and a 10% price reduction—making it optimized for use cases such as backup and disaster recovery. As a result, it's even more cost effective, and free bulk retrievals make it ideal for when you need to retrieve large datasets once or twice a year and don't want to worry about the retrieval cost.

Benefits:

S3 Glacier Flexible Retrieval is one of three [S3 Glacier storage classes](#) optimized for archive data with a variety of uses cases, from user-generated content and media archives to analytics, genomics, and compliance archives.

- For archive data that needs immediate access, such as medical images, news media assets, or genomics data, choose the S3 Glacier Instant Retrieval storage class, an archive storage class that delivers the lowest-cost storage with milliseconds retrieval.
- For archive data that does not require immediate access, such as backups and disaster recovery, choose S3 Glacier Flexible Retrieval (formerly S3 Glacier), with flexible data retrieval options from minutes to hours. You can choose from expedited retrievals in 1–5 minutes, standard retrievals in 3–5 hours, and free bulk retrievals in 5–12 hours.
- To save even more on long-lived archive storage, such as for regulatory and compliance data and digital media preservation, choose the S3 Glacier Deep Archive storage class, the lowest-cost storage in the cloud (less than \$1 per TB-month) with data retrieval from 12–48 hours.

Resources:

[Webpage](#) | [Blog post](#)

Amazon EFS Replication (coming soon)

Replicate your file systems within and across AWS Regions

What is it?

Amazon Elastic File System (EFS) Replication (coming soon) allows you to replicate your file system data to another Region or within the same Region in a few clicks, without requiring additional infrastructure or a custom process to monitor and synchronize data changes. Organizations in regulated industries are often subject to compliance requirements that mandate storing secondary data copies several hundred miles away from the original. Amazon EFS Replication automatically and transparently replicates your data to a second file system in a Region or Availability Zone (AZ) of your choice. You can use the EFS console, AWS Command Line Interface (CLI), and APIs to enable replication on an existing file system. EFS Replication is continuous and designed to provide a recovery point objective (RPO) and recovery time objective (RTO) of minutes, allowing you to meet your compliance and business continuity goals.

You can also set up your destination file system independent of your source file system. You can select the destination file system's lifecycle management policy, backup policies, provisioned throughput, mount targets, and access points independent of the source file system. For example, you can optimize destination file system storage costs by enabling EFS lifecycle management with a shorter age-off policy (such as 7 days) than that of the source file system (such as 7, 14, 30, 60, or 90 days). You can also replicate from a source file system created using EFS Standard storage classes to a destination file system created using EFS One Zone storage classes, and the other way around.

Resources:

[Webpage](#) | [Breakout session recording and slides](#)

AWS Backup support for Amazon S3 (Preview)

Use a single policy on AWS Backup to automate the protection of application data stored on Amazon Simple Storage Service (S3) alone or alongside 11 other AWS services

What is it?

Get started with AWS Backup for Amazon S3 (Preview) by creating a backup policy on AWS Backup and assigning S3 buckets to it using tags or resource IDs. AWS Backup allows you to create periodic snapshots and continuous backups of your S3 buckets, and restore your S3 buckets and objects to your specified point in time with a single click in the AWS Backup management console. Additionally, you can use AWS Backup to maintain and demonstrate compliance of your organizational data protection policies to auditors.

Benefits:

- **Centralized data protection.** A single, centralized data protection solution for application data stored on S3 alongside other AWS services. Restore to the source S3 bucket, another existing bucket, or a new bucket. Restore object data, tags, access control lists (ACLs), and user-defined metadata.
- **Point-in-time restore.** Easily restore application data stored on S3 to a specified point in time with a single click.
- **Simplify compliance auditing.** Monitor and audit the compliance status of your data protection policies using AWS Backup Audit Manager.

Use cases:

- Centrally manage data protection across AWS Organizations.
- Create immutable backups of AWS supported services.
- Restore application data stored on S3 to a specified point in time.
- Use lifecycle policies to cold tier your backups.
- Create separable, protected cross-Region and cross-account backups to meet compliance needs

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Backup support for VMware

Use a single policy on AWS Backup to centrally protect your hybrid VMware environments alongside the 12 AWS services (spanning compute, storage, and databases) already supported by AWS Backup. AWS Backup allows you to demonstrate the compliance status of your organizational data protection policies by monitoring backup, copy, and restore operations, and generating unified auditor-ready reports to help satisfy your data governance and regulatory requirements.

What is it?

Using AWS Backup, you can centrally configure backup policies across your AWS applications comprising native AWS services, on-premises VMware, and VMware Cloud on AWS, helping you simplify data protection and automate lifecycle management. You can restore your VMware backups to your on-premises data centers and on VMware Cloud on AWS to meet your data recovery needs. AWS Backup Audit Manager provides built-in and customizable compliance controls to define your data protection policies, automatically detects violations of your defined policies, and prompts you to take corrective actions, allowing you to demonstrate compliance with regulatory requirements.

Benefits:

- **Centralized data protection.** Use a single, centralized data protection solution for hybrid VMware workloads alongside AWS services.
- **Flexible restore options.** With AWS Backup support for VMware, you can restore to on premises and/or VMware Cloud on AWS.
- **Normalized backup policy.** Use the same backup policy across AWS services and VMware.

Use cases:

- Centrally manage data protection across AWS Organizations.
- Create immutable backups of ESXi 6.7.x and 7.x from on-premises VMware and VMware Cloud on AWS.
- Restore to on premises or VMware Cloud on AWS.
- Use lifecycle policies to cold tier your backups.
- Create separable, protected cross-Region and cross-account backups to meet compliance needs

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

AWS Microservice Extractor for .NET

Simplify refactoring .NET applications

What is it?

AWS Microservice Extractor for .NET simplifies the process of refactoring applications into independent services. Modernize and transform your applications with an assistive tool that analyzes source code and runtime metrics to create a visual representation of your application and its dependencies. Microservice Extractor for .NET delivers a holistic visualization of applications, and assists in code refactoring and extracting the codebase into separate code projects that teams can develop, build, and operate independently—improving agility, uptime, and scalability.

Benefits:

- **Faster identification of applications components to refactor into microservices.** Quickly identify application components to refactor into smaller services using Microservice Extractor for .NET, which combines data from code analysis and runtime profiling to produce a single visualization showing each component's metrics and dependencies. As a result, there is no need to manually correlate the outputs of various tools for code and runtime analysis.
- **Guides refactoring based on domain-driven design principles.** Microservice Extractor for .NET helps you adopt industry best practices, such as domain-driven design, by allowing you to label the visualized graph to associate with business processes. The service highlights dependencies that need to be removed so you can extract parts of the application into separate code projects.
- **Assists with refactoring monolith codebases into smaller code projects.** After refactoring your monolithic application to prepare it for extraction as independent code projects, you can use Microservice Extractor for .NET to partition the source code into units that teams can develop, build, deploy, and operate as independent services.

Use cases:

- **Modernize your .NET applications.** Refactor your .NET applications to improve agility, uptime, and scalability.
- **Visualize your application.** Analyze and create a visual representation of your application and its dependencies.
- **Plan .NET modernization.** Visualize your application's source code and extract it into smaller codebases.

Resources:

[Webpage](#) | [Blog post](#) | [Breakout session recording and slides](#)

Please rate this Cheat Sheet



Ratings are anonymous and help us improve the quality of AWS content.