

re:Invent 2023 re:cap - Dev, DevOps, Compute & Networking

Chan Zaw
Solutions Architect (FSI)

AWS RE:INVENT RE:CAP

re:Invent 2023 Key Stats for Apps and Infra announcements

20 major announcements around Compute, Network and Application Developments

10 / 20 are powered by AI

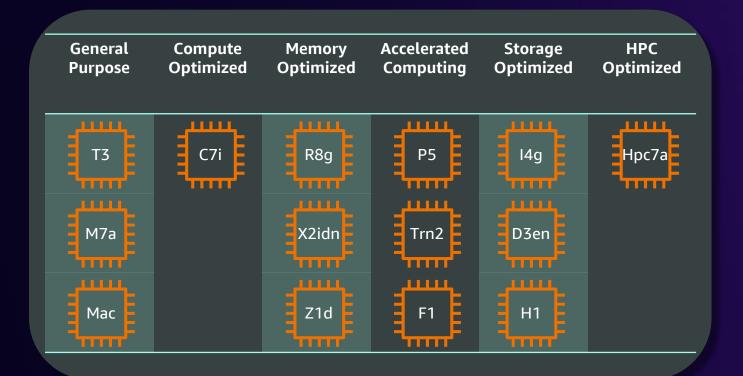


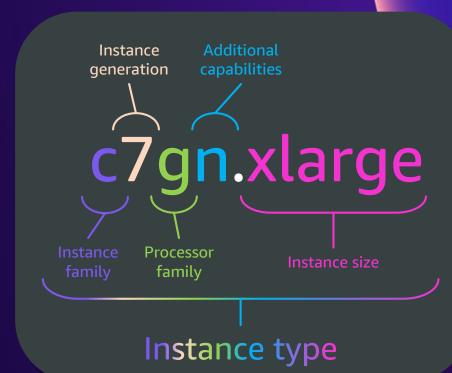


Compute



Quick Reminders on EC2 Naming





Now over 750 instances





Amazon EC2 r8g Instance (with Graviton4)

NEXT GENERATION OF STANDARD RAM FOCUSED INSTANCES WITH NEWEST GRAVITON PROCESSOR

Up to 3x more vCPU and memory than R7g Dedicated caches for every vCPU

Generational Performance Gains

30% faster for web apps 40% better performance for databases 45% faster for large Java applications

Security Capabilities

Always-on memory encryption
Support for pointer authentication
Support for encrypted Amazon Elastic Block Store



Powered by Graviton4 Processors

Who should use? Anyone using r7g and anyone running databases or caches on EC2



Amazon EC2 u7i Instance



MAJOR UPGRADE IN HIGH MEMORY INSTANCES FOR EC2 BASED, ULTRA-LOW LATENCY DATABASES

Designed for very high memory use cases

Up to 125% more compute performance and 120% more memory performance (from u-6/9/12tb1)

2.5x as much EBS bandwidth

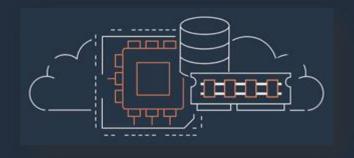
Hydrate in-memory databases at a rate of up to 44 terabytes per hour

Attach up to 128 EBS volumes



io2 Block Express volumes can be up to 64 TiB with 256K IOPS and 32 Gbps, making them a great match for the u7i instance

Available in 16, 24, and 32 TB of DDR5 RAM



8 Sockets, each with a 56 core CPU = 896 vCPU

100Gbps EBS and Network Bandwidth

Who should use it? Anyone running u-6/9/12tb1 instances or running DBs like SAP HANA, Oracle, or SQL Server.





Amazon EC2 Trn2 Instance (Trainium2)

EC2 WITH AN ML-OPTIMIZED PROCESSOR FOR CUSTOMERS WHO BUILD FMS ON AWS HARDWARE

Optimized for training FMs with hundreds of billions to trillions of parameters

4x faster than AWS Trainium

65 exaflops of on-demand supercomputing performance

Purpose-built chip for Generative AI and ML Training



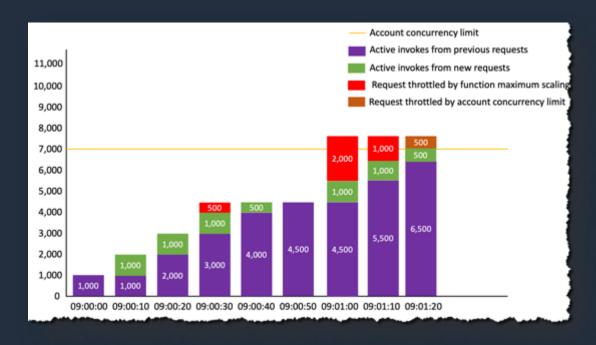
Who should use it? Anyone running Trn1 instances or if you are selecting hardware for large ML workloads





Lambda Scaling Improvements

HIGH VOLUME AND RAPID SCALING LAMBDA IS EVEN FASTER AT SCALING YOUR INVOCATIONS



This graph shows a function receiving requests and processing them every 10 seconds with an account concurrency limit set to 7,000 concurrent requests. Each function scaling-up rate is fixed to 1,000 concurrent executions every 10 seconds.



Scale up at a 12x faster rate



Scale up to a rate of 1,000 concurrent executions every 10 seconds (to account concurrency limit)



Available in all commercial AWS Regions



Enabled by default for all functions that are invoked synchronously, at no additional cost.





EKS Pod Identity

SIMPLIFIED WORKFLOW FOR AUTHENTICATION OF KUBERNETES APPLICATIONS TO AWS RESOURCES

Scope IAM permissions to a service account, and only Pods that use that service account have access to those permissions

Eliminate the need for third-party solutions such as kiam or kube2iam

A container never has access to credentials that are used by other containers in other Pods.

Simpler method that does not require OIDC identity providers

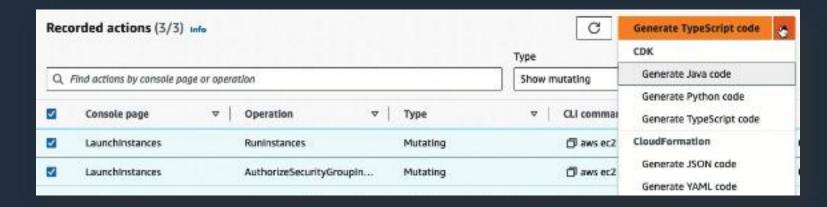
Uses a single IAM principal instead of the separate principals for each cluster that IAM roles for service accounts use.





AWS Console-to-Code (Preview)

A PROMISING PREVIEW OF CONVERTING CONSOLE ACTIONS INTO REUSABLE IAC CODE



Actions taken for EC2 instances will be shown in the Recorded Actions Pane of "Console-to-Code"

Select the actions you want to script

Generate Code in your preferred language





Convert actions performed in the console to reusable code, currently Java CDK, Python CDK, TypeScript CDK, or CloudFormation (JSON or YAML)



Suggested code uses GenAI capabilities following AWS guided best practices to help improve deployment and create reliable production workloads

Customers no longer have to make a choice between Console or Infrastructure-as-Code (IaC)







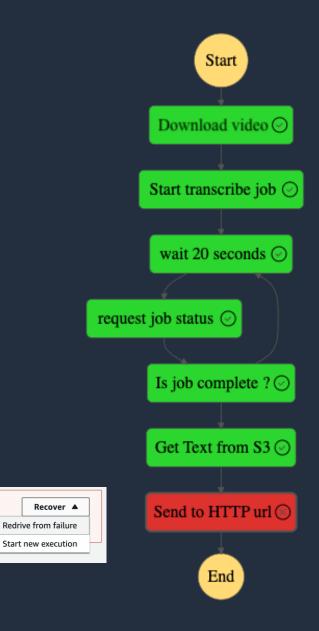
AWS Step Functions - redrive

NEW Feature

GA

Step Functions – Cost Optimisation

- AWS charge for number of state transitions
- Each state transition is a step of your workflow runs
- Some steps are costly
- Redrive help save cost and simplify your workflow error handling









Amazon Q for Developers

Amazon Q to get started building applications on AWS, research best practices, resolve errors, and get assistance in coding new features for your applications

NEW Service

PREVIEW

Overview of Q Offerings and Features

O FALLS INTO TWO PERSONA-BASED OFFERINGS AND IS BEING LAUNCHED FOR INDIVIDUAL SERVICES



Each feature falls under the corresponding service's features

Customers should primarily look for a LLM-powered chat interface, integrated RAG data connections, and process automation from Amazon Q

Some features will have additional costs after Preview.

See aws.amazon.com/q/pricing for up-to-date details!

Amazon Q Offerings for Individual Services



Amazon Q for Amazon Connect



Amazon Q in AWS Supply Chain





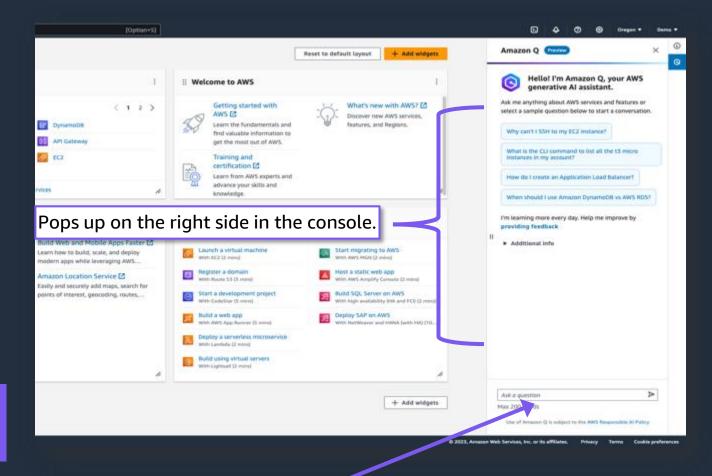


Builder: Amazon Q AWS Expert Interface

AWS FOCUSED EXPERT INTERFACE WITH GENAI CONVERSATIONAL Q&A TRAINED ON AWS SOURCES

- Trained on 17 years of AWS knowledge
- Assists you everywhere you work with AWS
- Converses with you to explore new AWS capabilities, learn unfamiliar technologies, and architect solutions
- Works with you to troubleshoot, build new features, and upgrade languages on AWS

Also present in AWS Documentation, IDE (via CodeWhisperer), and Slack/Teams







GENERALLY AVAILABLE

Amazon CodeWhisperer

Build apps faster and more securely with an AI coding companion



Generate code suggestions in real-time



Scan code for hard-to-find vulnerabilities



Flag code that resembles opensource training data or filter by default

FREE FOR INDIVIDUAL TIER





NEW

Amazon Code Whisperer



IaC Support such as Cloud Formation, CDK and Terraform



Generates organization-specific code recommendations (Preview)



Expanded support(TypeScript, C#) for security scanning (GA)







Builder: CodeWhisperer with Amazon Q

A GENAI INTERFACE WITHIN YOUR IDE CONSOLE THAT CAN PERFORM Q&A AND GENERATE CODE

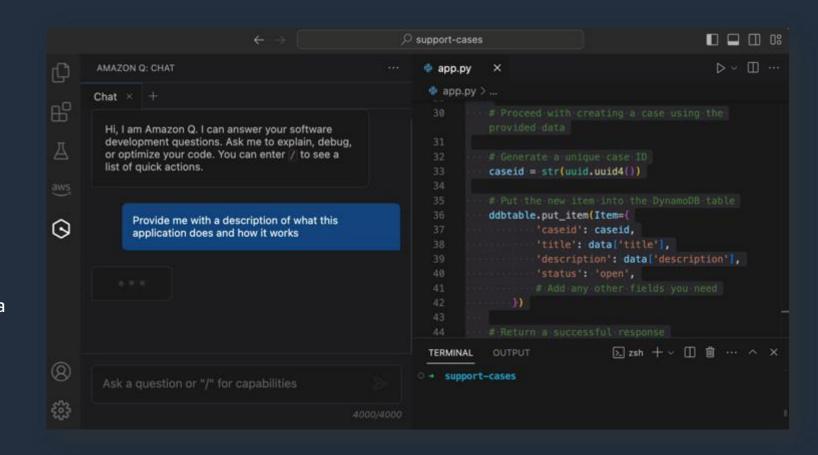
CodeWhisperer Capabilities with Q

Your code is used, contextually, to enhance the **Q&A** conversational interface

Summarize code, explain functions, and provide deeper insights

Assists with debugging, testing, and optimizing your code

Q can generate tests and add them to code with a click



At this time (DURING PREVIEW): Amazon Q in IDE is part of Amazon Q Builder subscriptions and require Amazon CodeWhisperer

The current documented plan after GA is that when a user has paid for Amazon Q Builder seat, they also get CodeWhisperer Professional included. RELINVENT RE:CAP



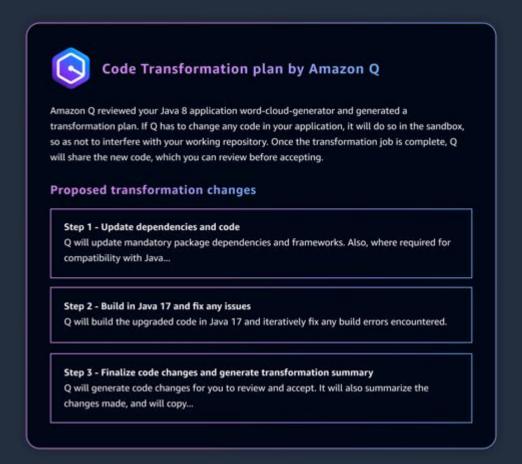


Builder: Amazon Q Code Transformations

AMAZON Q AND CODE WHISPERER ACCELERATE LANGUAGE VERSION UPGRADES AND OTHER MODERNIZATIONS

Amazon Q Code Transform Support

- Current: Java language upgrades from Java 8 to Java 17
 - Tests internally moved 1000 apps in 2 days, about 10 minutes per app (versus typical days per app)
- Coming Soon: .NET Framework to cross-platform .NET upgrades
- Coming Soon: Accelerate the migration from Windows to Linux (coming soon)



At this time: Amazon Q Code Transforms are part of Amazon Q Builder subscriptions and require Amazon CodeWhisperer





Builder: Amazon Q Network Troubleshooting

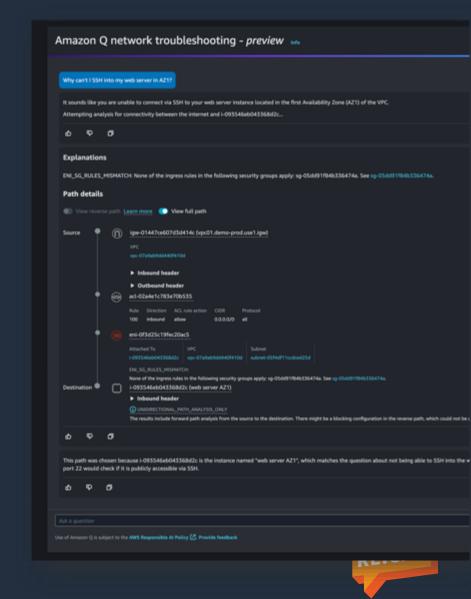
AWS VPC REACHABILITY ANALYZER WORKS WITH AMAZON Q TO HELP TROUBLESHOOT YOUR NETWORK

Identify and resolve connectivity issues via natural language

Ask network troubleshooting questions in a conversational format

- Why can't I SSH into my Linux instance?
- Why can't I access the internet from vpc-1a2b3c4d?
- Are any of my instances publicly accessible?
- Do I have any egress from my VPC to the internet?
- Which of my security groups are preventing access to my RDS database?

Provides a simple, concise explanation and recommendation, as well as a map of the path details.







AWS Config generative Al-powered NLP

NEW Service

PREVIEW

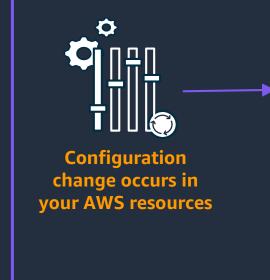
AWS Config

ASSESS, AUDIT, AND EVALUATE



Assess, audit,

and evaluate





AWS Config

Records and normalizes the changes into a consistent format



Rules and conformance packs evaluate compliance



Sample conformance packs to get started: CIS, CMMC, FedRAMP, NIST, PCI, HIPAA, and more



AWS Config APIs and console



Amazon SNS



Amazon CloudWatch



Amazon S3

Access change history and compliance results using the console or APIs; Amazon CloudWatch Events or Amazon SNS alert you when changes occur; deliver change history and snapshot files to your Amazon S3 bucket for analysis



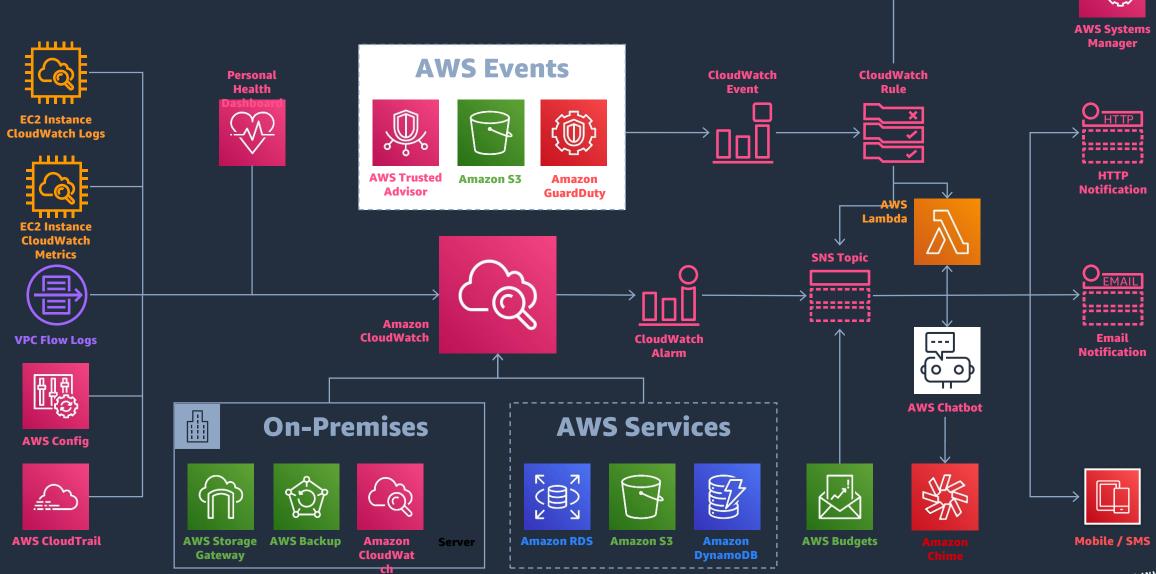


Amazon CloudWatch Generative Al-powered NLP

NEW Service

PREVIEW

Integration







AWS Security Hub

NEW Feature

GA



AWS Security Hub



New Finding Enrichment



Major Dashboard Enhancements



Central Configuration Capabilities



Customize Security Controls in AWS Security Hub







Amazon Inspector agentless vulnerability assessment

NEW Service

PREVIEW

NEW

Amazon Inspector Agentless vulnerability scanning



Agentless Scanning



Vulnerability Assessments from EBS volumes snapshot



Automatic Discovery



Hybrid scanning





DEMO





Questions?







Thank you!



https://pulse.aws/survey/QQT4M40A





Please complete the session survey via received **email**

Analytics

Redshift provisioned concurrency scaling and serverless autoscaling now supports Create Table As Select (CTAS)

Amazon Redshift now supports multi-data warehouse writes through data sharing (preview)

AWS announces support for large language models in Amazon Redshift ML (Preview)

Announcing the Cost and Usage Dashboard powered by Amazon QuickSight

AWS Glue Data Quality announces anomaly detection and dynamic rules

Amazon MSK now supports Graviton3-based M7g instances for new provisioned clusters

Amazon Redshift announces Multidimensional Data Layouts to optimize your query performance (preview)

Amazon Redshift adds support for incremental refresh for materialized views on data lake tables (preview)

AWS Glue Data Catalog supports multi engine views with AWS Analytics Engines

<u>Accelerate data processing and analysis with Amazon EMR and Amazon S3 Express One Zone</u>

Accelerate data lake queries with Amazon Athena and Amazon S3 Express One Zone

AWS Clean Rooms Differential Privacy is now available in preview

Vector engine for Amazon OpenSearch Serverless now generally available

Amazon Redshift announces general availability of support for Apache leeberg

Analytics (cont'd)

Amazon OpenSearch Service zero-ETL integration with Amazon S3 preview now available

AWS announces OR1 for Amazon OpenSearch Service

Announcing enhanced manageability and usability features for Amazon Redshift Serverless

<u>Amazon Redshift announces general availability of row-level security enhancements</u>

Amazon Redshift now supports metadata security to simplify multitenant applications

Application Integration

AWS Step Functions launches support for HTTPS endpoints and a new TestState API

AWS Step Functions launches optimized integration for Amazon Bedrock

Amazon EventBridge now supports partner integrations with Adobe and Stripe

Amazon SQS announces support for FIFO dead-letter queue redrive

Announcing AWS B2B Data Interchange

Amazon SQS announces increased throughput quota for FIFO High Throughput mode

Blockchain

<u>Announcing preview of AMB Access Polygon, serverless access to Polygon</u> blockchain

Business Applications

Introducing Amazon One Enterprise (Preview)

Amazon Connect now supports two-way SMS

Amazon Connect provides Zero-ETL analytics data lake to access contact center data (preview)

Amazon Connect Contact Lens now provides real-time conversational analytics for chat

Amazon Connect now offers in-app, web, and video calling

Amazon Connect launches no-code UI builder to configure step-by-step quides

Cloud Financial Management

AWS Free Tier usage is now available through the GetFreeTierUsage API

<u>Amazon Web Services announces Unified Billing and Cost Management console</u>

Introducing Cost Optimization Hub

Announcing Data Exports for AWS Billing and Cost Management

Cloud Operations

Amazon Managed Service for Prometheus launches an agentless collector for Prometheus metrics from Amazon EKS

CloudWatch now supports hybrid and multicloud metrics querying and alarming





Cloud Operations (cont'd)

AWS Config now supports periodic recording: Efficiently scale your change tracking

AWS Systems Manager Automation makes it easier to author runbooks with new low-code visual design experience

AWS Compute Optimizer introduces customizable rightsizing recommendations for EC2 Instances

Automate AWS Control Tower landing zone operations using APIs

<u>AWS CloudTrail Lake data now available for zero-ETL analysis in Amazon</u> Athena

Amazon CloudWatch Logs announces Infrequent Access log class

AWS Control Tower announces 65 new controls to help meet digital sovereignty requirements

Observe your applications with Amazon CloudWatch Application Signals (Preview)

myApplications: One place to view and manage your applications on AWS

Compute

Announcing Amazon EC2 High Memory U7i instances (Preview)

ENA Express supports 58 new instances with sizes as small as 16 vCPUs

Announcing new Amazon EC2 R8g instances powered by AWS Graviton4 processors (Preview)

Containers

Amazon EKS introduces EKS Pod Identity

Customer Enablement

Announcing the general availability of AWS re:Post Private

Databases

Announcing the general availability of Amazon RDS for Db2

Announcing Amazon Aurora Limitless Database

AWS announces Amazon ElastiCache Serverless

AWS announces Amazon Aurora PostgreSQL zero-ETL integration with Amazon Redshift (Public Preview)

AWS announces Amazon RDS for MySQL zero-ETL integration with Amazon Redshift (Public Preview)

AWS announces Amazon DynamoDB zero-ETL integration with Amazon Redshift

AWS announces Amazon DynamoDB zero-ETL integration with Amazon OpenSearch Service

Amazon Neptune Analytics is now generally available

AWS announces vector search for Amazon MemoryDB for Redis (Preview)

AWS announces vector search for Amazon DocumentDB

Developer Tools

AWS CloudFormation introduces Git management of stacks

Announcing Multi-Account Experiments for AWS Fault Injection Service

AWS SDK for Kotlin is now generally available

AWS SDK for Rust is now generally available

Introducing the Amazon CodeCatalyst Enterprise Tier

Announcing custom blueprints for Amazon CodeCatalyst

AWS Application Composer announces AWS Step Functions Workflow Studio integration

AWS SDK for Kotlin is now generally available

AWS Fault Injection Service launches two highly requested scenarios

Amazon Inspector enhances container image security by integrating with developer tools

<u>Introducing an Integrated Development Environment (IDE) extension for AWS Application Composer</u>

End User Computing

Amazon WorkSpaces Multi-Region Resilience launches one-way data replication

Amazon WorkSpaces Thin Client is now generally available

Front-End Web & Mobile

AWS AppSync improves support for Amazon Aurora clusters configured with the RDS Data API



Generative Al

AWS Config launches generative AI-powered natural language querying (Preview)

Amazon Inspector expands AWS Lambda code scanning with generative AI powered remediation

Amazon CloudWatch announces AI-powered natural language query generation (in preview)

Amazon Personalize now creates themes for recommendations using generative AI

Amazon Transcribe Call Analytics now offers generative call summarization (preview)

New Discover Apps page for PartyRock, an Amazon Bedrock Playground

Announcing new enhancements to Amazon CodeWhisperer

Announcing Conversational FAQ with generative AI for Amazon Lex (Preview)

<u>Announcing AWS Console-to-Code (Preview) to generate code for console actions</u>

Amazon Detective introduces finding group summaries using generative Al

Descriptive Bot Builder with Generative AI

Announcing utterance generation for Amazon Lex

Assisted Slot Resolution with Generative AI

AWS AppFabric's new generative AI feature helps build cross-app experiences (Preview)

AWS Healthscribe is now generally available

Generative AI (cont'd)

Announcing Amazon Q expert capabilities for AWS (Preview)

Boost generative AI application development with Agents for Amazon Bedrock

Amazon Q offers help to optimize EC2 instance type selection (preview)

Amazon Connect announces generative Al powered customer data mapping

AWS Chatbot now supports Amazon Q conversations in Microsoft Teams and Slack

<u>Amazon Connect Contact Lens provides generative AI powered contact</u> summarization (preview)

AWS announces Amazon Q (Preview)

Amazon Q in QuickSight simplifies data exploration with Generative BI capabilities (Preview)

Announcing feature development capability of Amazon Q (Preview) in Amazon CodeCatalyst

AWS Announces Amazon Q is available in preview on the AWS Console Mobile App

Amazon Q in Connect offers generative AI powered agent assistance in real-time

Al recommendations for descriptions in Amazon DataZone (Preview)

Meta Llama 2, Cohere Command Light, and Amazon Titan FMs can now be fine-tuned in Amazon Bedrock

<u>Safeguard generative AI applications with Guardrails for Amazon Bedrock (Preview)</u>

Knowledge Bases for Amazon Bedrock is now generally available

Generative AI (cont'd)

Continued pre-training in Amazon Bedrock now available in preview

Amazon Bedrock now supports batch inference

Amazon Titan Text models—Express and Lite—now generally available in Amazon Bedrock

Amazon Titan Multimodal Embeddings foundation model now generally available in Amazon Bedrock

Stable Diffusion XL 1.0 foundation model from Stability AI is now generally available in Amazon Bedrock

<u>Llama 2 70B foundation model from Meta is now available in Amazon Bedrock</u>

<u>Claude 2.1 foundation model from Anthropic is now generally available in</u> Amazon Bedrock

<u>Evaluate</u>, <u>compare</u>, <u>and select the best FMs for your use case in Amazon Bedrock (Preview)</u>

Amazon Q generative SQL is now available in Amazon Redshift Query Editor (preview)

Announcing Amazon SageMaker HyperPod, a purpose-built infrastructure for distributed training at scale

<u>Leverage FMs for business analysis at scale with Amazon SageMaker</u> <u>Canvas</u>

Amazon SageMaker Canvas now supports natural language instructions for data preparation

<u>Amazon SageMaker Clarify now supports foundation model (FM)</u> <u>evaluations in preview</u>

Announcing new AWS AI Service Cards - to advance responsible AI



Internet of Things

<u>FreeRTOS roadmap and code contribution process now published on</u> freertos.org

AWS IoT SiteWise Edge Now Available on Siemens Industrial Edge Marketplace (Preview)

Announcing vision system data from AWS IoT FleetWise (Preview)

Machine Learning

<u>Powered by foundation model, Amazon Transcribe now supports over</u> 100 languages

Amazon Monitron launches Ex-rated sensors for hazardous locations

Recommend actions that increase brand loyalty with Amazon Personalize Next Best Action

AWS announces CloudWatch Logs Anomaly Detection and Pattern analysis

Amazon SageMaker launches a new version of Large Model Inference DLC with TensorRT-LLM support

AWS Clean Rooms ML is now available in preview

Amazon SageMaker Pipelines now provide a simplified developer experience for AI/ML workflows

<u>Announcing smart sifting of data for Amazon SageMaker Model Training in preview</u>

Amazon SageMaker launches new inference capabilities to reduce costs and latency

<u>Announcing API support for creating Amazon SageMaker Notebook jobs</u>

Machine Learning (cont'd)

<u>SageMaker now provides improved SDK tooling and UX for model</u> deployment

Amazon SageMaker now provides a new setup and onboarding experience on AWS SageMaker console

New and improved Amazon SageMaker Studio

Amazon SageMaker Distribution is now available on Code Editor based on Code-OSS and JupyterLab

Bring your own Amazon EFS (Elastic File System) volume to JupyterLab and CodeEditor in Amazon SageMaker Studio

Announcing Code Editor, based on Code-OSS (VS Code – Open Source), in Amazon SageMaker Studio

Amazon SageMaker Studio now provides a faster fully-managed notebooks in JupyterLab

Migration & Transfer

AWS Mainframe Modernization File Transfer with BMC is now generally available

AWS Mainframe Modernization Data Replication with Precisely is now available for IBM i

AWS Mainframe Modernization Replatform with NTT DATA is now available

<u>Virtualization for SPARC on AWS with Stromasys is now available</u>

Introducing AWS Mainframe Modernization Application Testing (Preview)

Networking & Content Delivery

<u>Application Load Balancer can authenticate X.509 certificate based</u> <u>identities with Mutual TLS support</u>

Application Load Balancer increases application availability with Automatic Target Weights

Amazon Route 53 Application Recovery Controller launches zonal autoshift

Partner

Announcing enhanced capabilities for Partner-Led Support

AWS Partner Network launches new Amazon EKS Ready Specialization

Introducing AWS Resilience Competency Partners

Announcing AWS Marketplace APIs for sellers

AWS Partner CRM Connector now supports AWS Marketplace

<u>Announcing Solution Building Enablement for Partners</u>

Announcing SaaS Quick Launch for AWS Marketplace

Quantum Technologies

Amazon Braket launches Braket Direct, a program to dive deeper into quantum computing



Security, Identity, & Compliance

Amazon GuardDuty now supports runtime monitoring for Amazon EC2 (Preview)

IAM Access Analyzer now simplifies inspecting unused access to guide you toward least privilege

New from AWS: You can now customize security controls in AWS Security $\underline{\mathsf{Hub}}$

<u>IAM Access Analyzer introduces custom policy checks powered by</u> automated reasoning

AWS Secrets Manager now supports batch retrieval of secrets

Amazon Detective supports security investigations for Amazon GuardDuty ECS Runtime Monitoring

Amazon Detective announces investigations for IAM

Introducing Amazon GuardDuty ECS Runtime Monitoring, including AWS Fargate

AWS Analytics simplify users' data access across services with IAM Identity Center

Amazon Detective now supports log retrieval from Amazon Security Lake

Amazon Inspector agentless vulnerability assessments for Amazon EC2 now in preview

Announcing new central configuration capabilities in AWS Security Hub

Announcing major dashboard enhancements in AWS Security Hub

Announcing new finding enrichment in AWS Security Hub

Request a Cyber Insurance Quote from an AWS Cyber Insurance Competency Partner

AWS Built-in Competency Partner software automates Installation for customers

Storage

Amazon S3 Access Grants integrate with identity providers to simplify data lake permissions

Announcing Amazon FSx for NetApp ONTAP scale-out file systems

You can now manage and back up FSx for ONTAP FlexGroup volumes using the AWS Console and API

Announcing the new Amazon EFS Archive storage class

<u>Amazon FSx for OpenZFS now supports on-demand data replication across file systems</u>

Amazon EFS now supports up to 250,000 IOPS per file system

FSx for ONTAP now supports creating Multi-AZ file systems in Shared VPC participant accounts

AWS Backup now supports Amazon Elastic Block Store (EBS) Snapshots Archive

Announcing AWS Elastic Disaster Recovery in the AWS GovCloud (US) Regions

Amazon EFS Replication now supports failback

AWS Backup launches support for restore testing

New updates automatically accelerate Amazon S3 data transfer for ML training

Mountpoint for Amazon S3 CSI driver is now generally available

<u>Mountpoint for Amazon S3 now supports the Amazon S3 Express One Zone storage class</u>

Announcing the Amazon S3 Express One Zone storage class

Support

AWS Support launches AWS Countdown with a premium tier

Well Architected

Announcing Lens Catalog for the AWS Well-Architected Tool

