

AWS Builders Online Series

CLOSING KEYNOTE

AWS re:Invent Highlights

Kris Howard

Head of Developer
Relations, APJ
AWS

Donnie Prakoso

Principal Developer
Advocate
AWS



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.





Who is a Builder?

Developers

InfoSec

DevOps
Engineers

IT
Pros

Data Scientists
and Engineers

Architects

Business
leaders



Sustainability

Breadth of services



productivity

Pace of innovation

Community

Global infrastructure

Trust

Agility

Resilience

Why do Builders choose AWS?

Customer focus

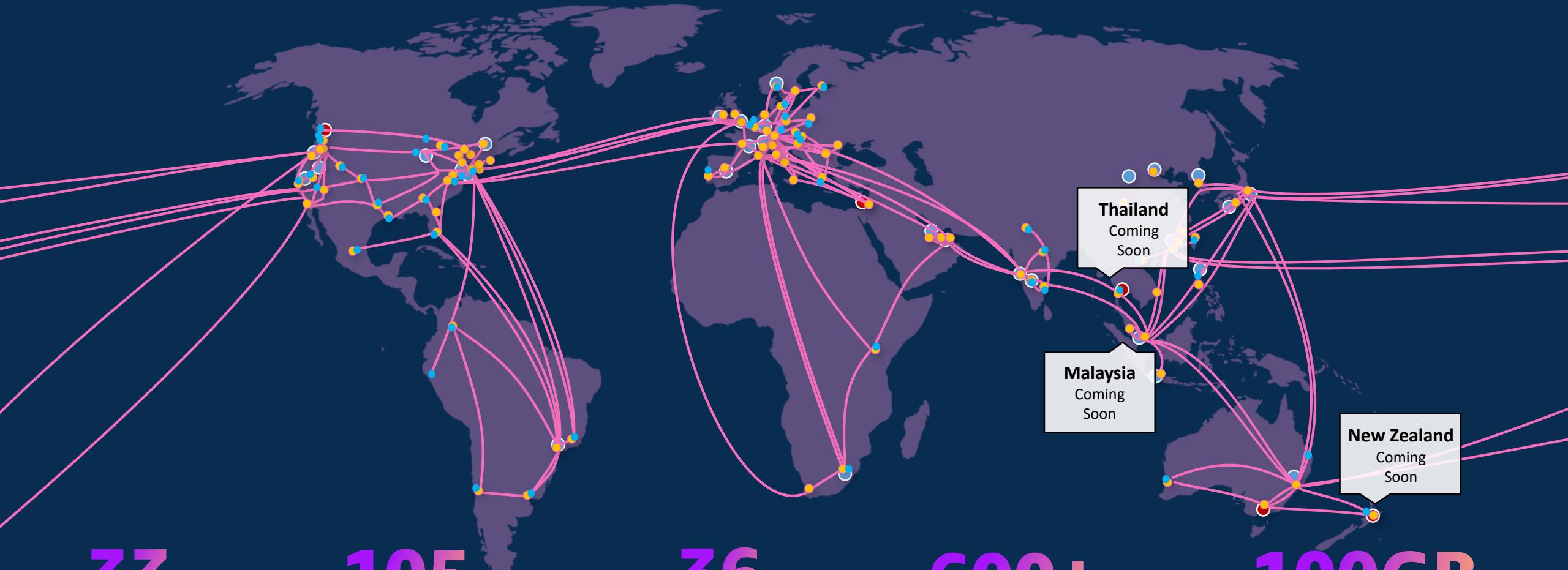
Cost

Security

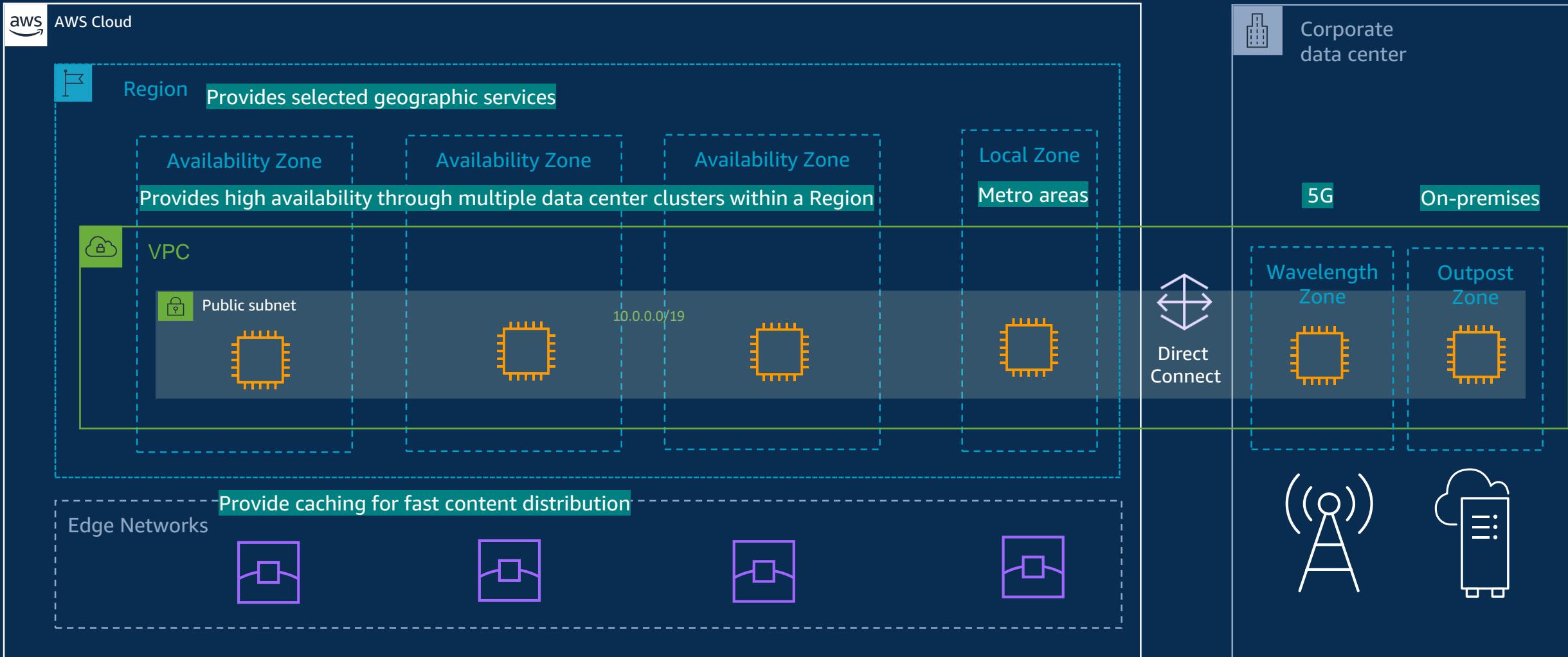
Performance



Safe and secure global infrastructure



Safe and secure global infrastructure



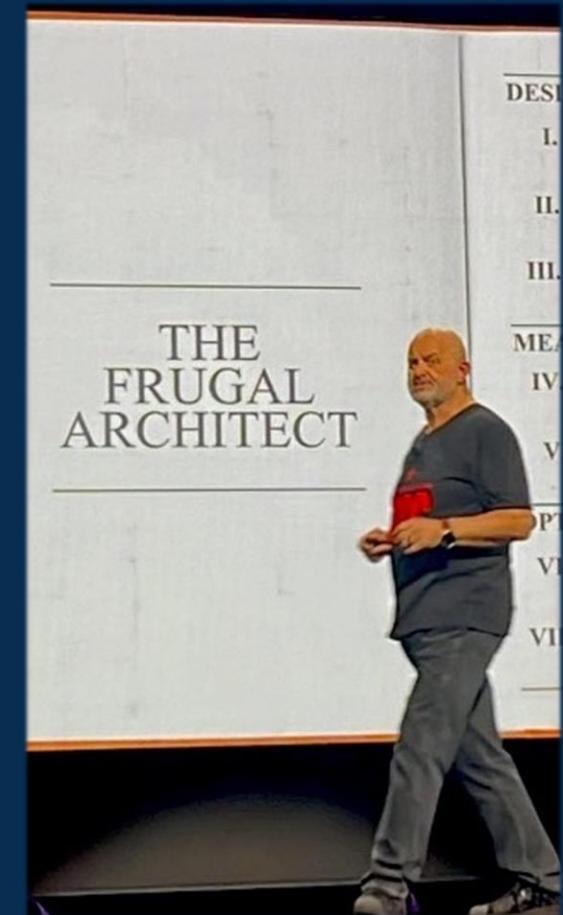
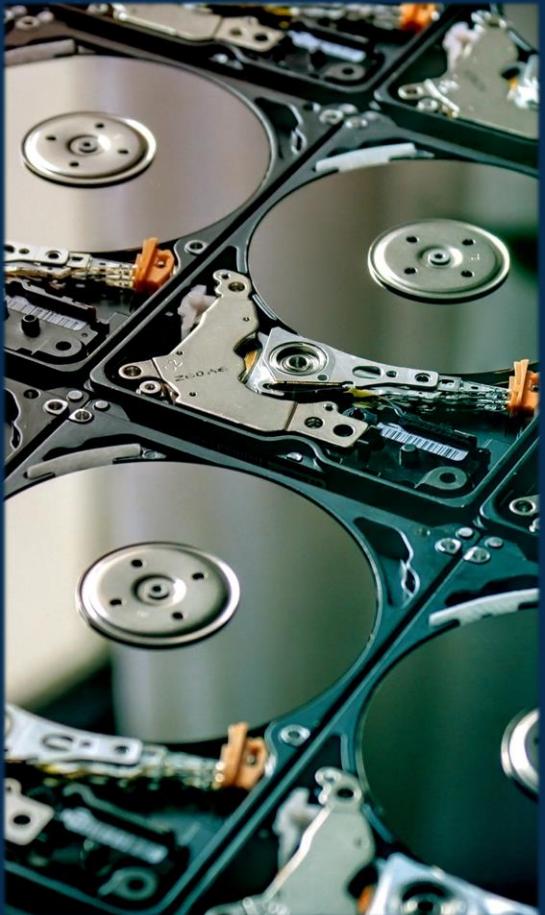
The **broadest** and **deepest** cloud

 Customer Enablement	 Cost Management	 End User Computing	 Internet of Things	 Robotics, Blockchain, AR/VR, Satellite, Quantum Technology
5 SERVICES	4 SERVICES	3 SERVICES	12 SERVICES	5 SERVICES
 Machine Learning	 Media Services	 Developer Tools	 Application Integration	 Front-end Web & Mobile, Gaming
26 SERVICES	11 SERVICES	14 SERVICES	8 SERVICES	6 SERVICES
 Analytics	 Management & Governance	 Security, Identity, & Compliance	 Migration & Transfer	 Business Applications
14 SERVICES	24 SERVICES	22 SERVICES	9 SERVICES	10 SERVICES
 Computing	 Container	 Storage	 Database	 Networking & Content Delivery
10 SERVICES	4 SERVICES	7 SERVICES	9 SERVICES	10 SERVICES

200+ Services



Highlights from AWS re:Invent 2023



Compute and Storage



Amazon S3

AWS Backup
AWS Storage
Amazon EBS
Amazon EFS

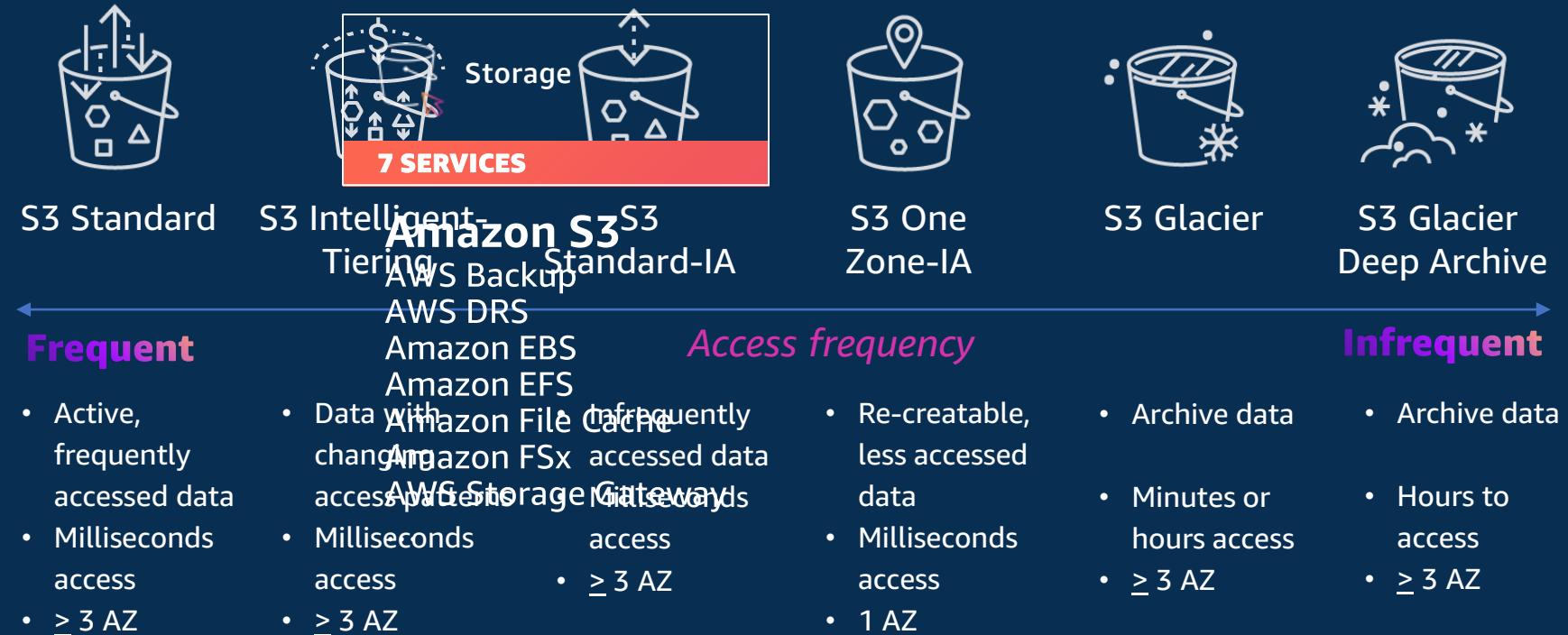
7 SERVICES

Amazon FSx
AWS Storage Gateway

...



Your choice of Amazon S3 storage classes



6 Storage Classes

NEW

Amazon S3 **Express One Zone**

Highest performance and lowest latency
cloud object storage

GENERALLY AVAILABLE

Millions of access
requests per minute

Single-digit
millisecond latency

50% lower access
costs vs S3 Standard





AWS Lambda
Amazon Lightsail
AWS Outposts
AWS Snow Family
AWS Wavelength

...

Your choice of Amazon EC2 instance types

Instance types (761)									C	Actions ▾										
<input type="text"/> Find resources by attribute or tag									<	1	2	3	4	5	6	7	...	16	>	
	Instance type ▾	vCPUs ▾	Archite... ▾	Memory... ▾	Storag... ▾	St... ▾	Network p... ▾	On-Demand Linux												
<input type="checkbox"/>	u-24tb1.112xl...	448	x86_64	24576	-	-	100 Gigabit	218.4 USD per Hour												
<input type="checkbox"/>	u-18tb1.112xl...	448	x86_64	18432	-	-	100 Gigabit	163.8 USD per Hour												
<input type="checkbox"/>	u-12tb1.112xl...	448	x86_64	12288	-	-	100 Gigabit	109.2 USD per Hour												
<input type="checkbox"/>	u-9tb1.112xlarge	448	x86_64	9216	-	-	100 Gigabit	81.9 USD per Hour												
<input type="checkbox"/>	u-6tb1.56xlarge	224	x86_64	6144	-	-	100 Gigabit	46.40391 USD per Hour												
<input type="checkbox"/>	u-6tb1.112xlarge	448	x86_64	6144	-	-	100 Gigabit	54.6 USD per Hour												
<input type="checkbox"/>	x2iedn.32xlarge	128	x86_64	4096	3800	ssd	100 Gigabit	26.676 USD per Hour												

750+ Instance Types



Your choice of Amazon EC2 instance types

750+
instance types
for virtually any workload

CATEGORIES

General purpose
Burstable
Compute intensive
Memory intensive
Storage (high I/O)
Hardware acceleration
HPC optimized

CAPABILITIES

Processor types
(AWS, Intel, AMD)
Processor speed
(up to 4.5GHz)
High footprint
(up to 24TiB)
Instance storage
(HDD and SSD)
Accelerated computing
(GPU, FPGA , and ASIC)
Networking
(up to 800 Gbps)
Bare metal

OPTIONS

Linux, Unix, Windows,
macOS
Amazon EBS
Elastic Fabric Adapter



Your choice of Amazon EC2 instance types



ARM-based



ML training



ML inference

NEW

AWS Graviton4

The most powerful and energy-efficient chip we have ever built

R8g Instances for EC2 Powered by AWS Graviton4

AVAILABLE IN PREVIEW TODAY

30% faster
than Graviton3

40% faster for
database applications

45% faster for large
Java applications



NEW

AWS Trainium2

Purpose-built chip for generative AI and ML training

COMING IN 2024

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

4x faster than AWS Trainium

65 exaflops of on-demand supercomputing performance





NEW

EC2 P5e Instances

NVIDIA H200 Tensor Core GPUs

COMING IN 2024

Ability to train language models at scale with up to 3200 Gbps of EFA networking

EC2 G6/G6e Instances

NVIDIA L4 GPU and L40S GPUs

COMING IN 2024

Support for Small Language Models and 3D graphics simulations

Trn1

The most cost-efficient, high-performance
deep learning training in the cloud

16

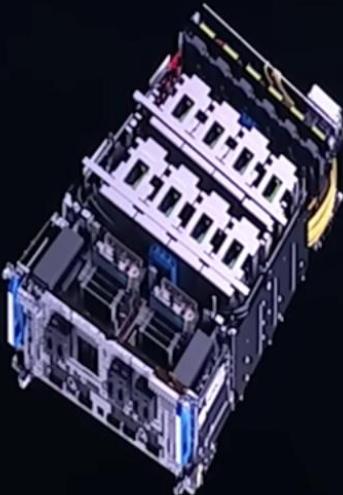
Trainium Processors

512 GB

Memory per Instance

800

GBPS
Network Bandwidth



Trn1

The most cost-efficient
deep learning

16

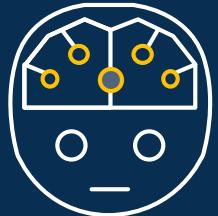
Trainium Processors



Generative AI



Where does Generative AI fit?



Artificial intelligence (AI)

Any technique that allows computers to mimic human intelligence using logic, if-then statements, and machine learning



Machine learning (ML)

A subset of AI that uses machines to search for patterns in data to build logic models automatically



Deep learning (DL)

A subset of ML composed of deeply multi-layered neural networks that perform tasks like speech and image recognition

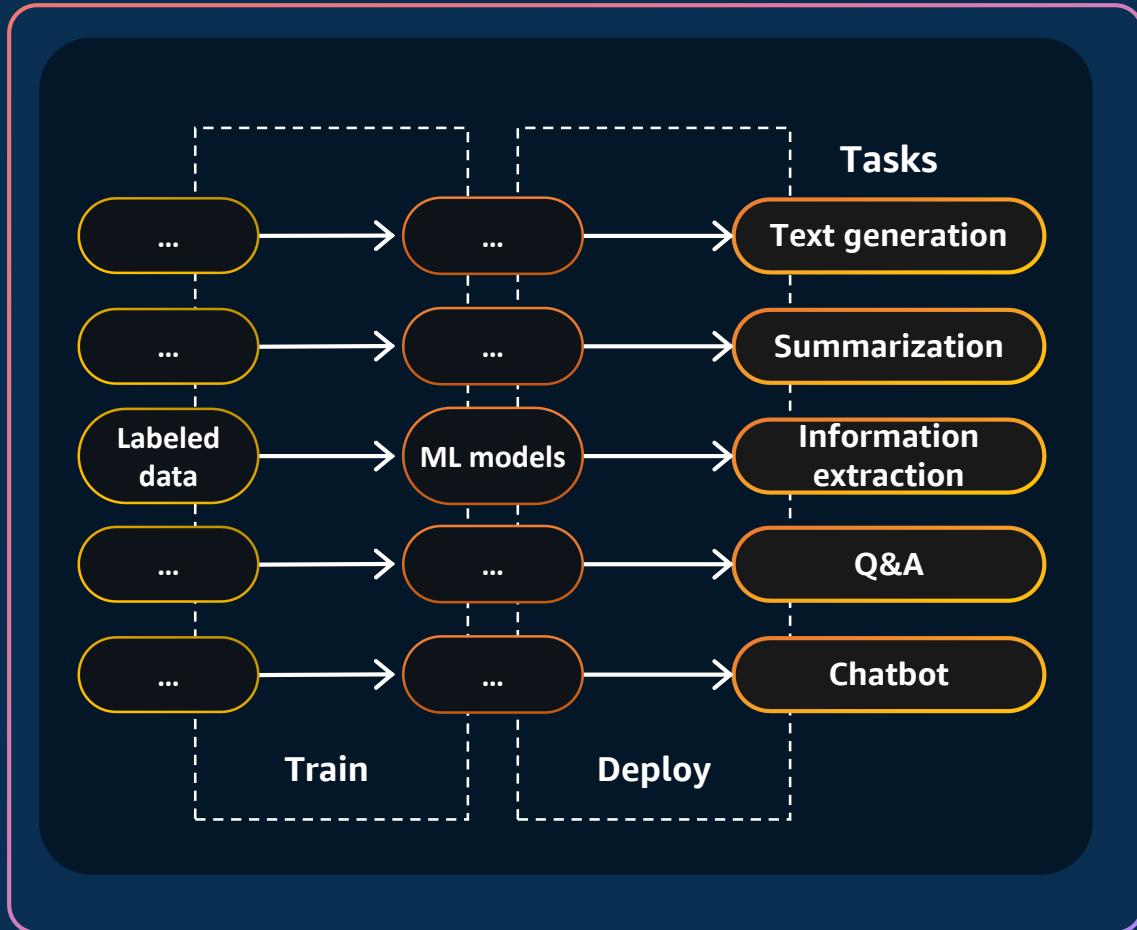


Generative AI

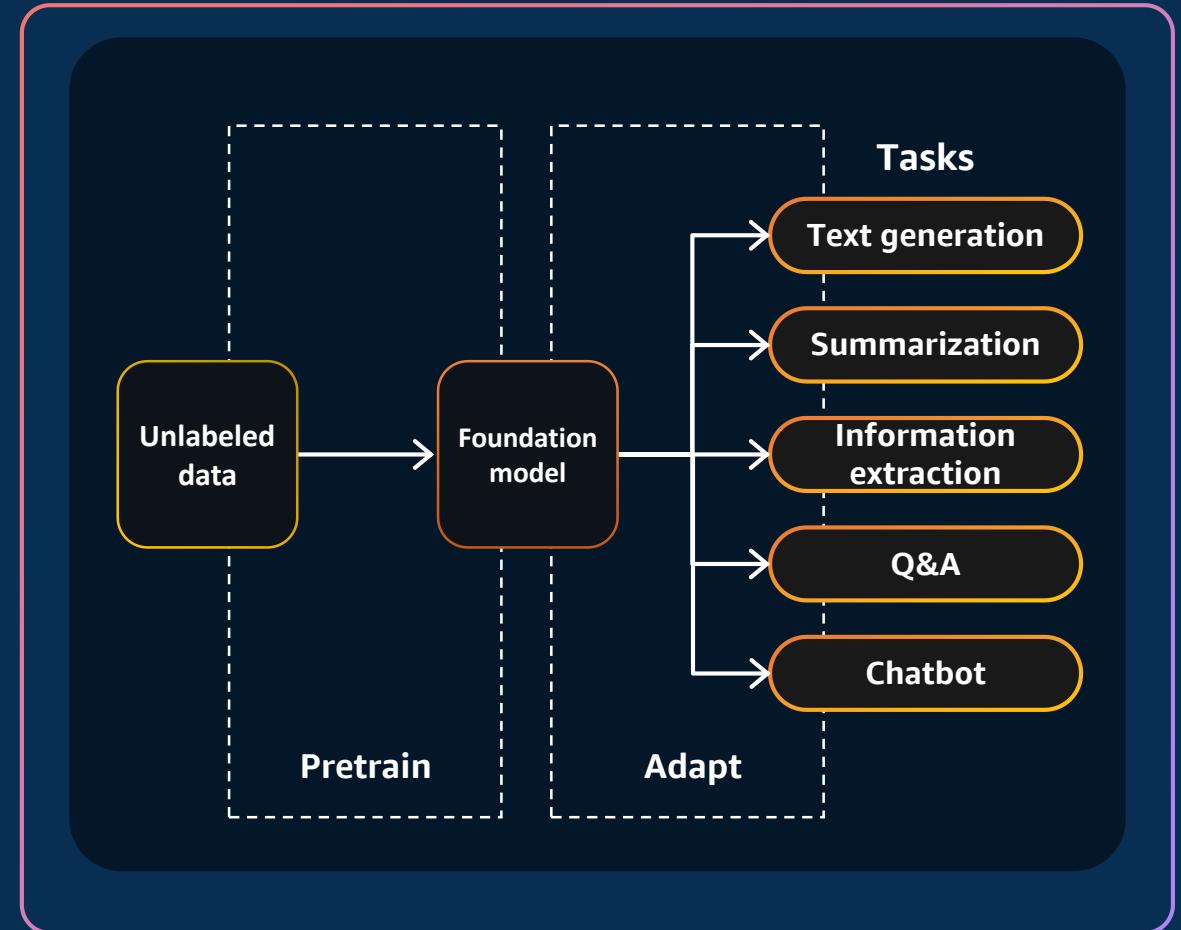
Powered by large models that are pretrained on vast corpora of data and commonly referred to as foundation models (FMs)

What's a Foundation Model (FM)?

Traditional ML Model



Foundation Model



AWS re:Invent



Data



Generative AI



Humans

Generative AI Stack on AWS

APPLICATIONS THAT LEVERAGE FMs



Amazon Q for builders



Amazon Q for business



Amazon CodeWhisperer

TOOLS TO BUILD WITH LLMs & OTHER FMs



Amazon Bedrock

Base models | Agents | Knowledge Base | Guardrails | Fine tuning

INFRASTRUCTURE FOR FM TRAINING & INFERENCE



GPUs



Trainium



Inferentia



SageMaker



UltraClusters



EFA



EC2 Capacity
Blocks



Nitro

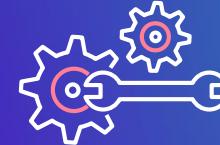


Neuron

NEW

Amazon Bedrock

The easiest way to build and scale generative AI applications with LLMs and other FMs



Choice of industry-leading FMs from AI21 Labs, Amazon, Anthropic, Cohere, Meta, and Stability AI



Customize FMs using your organization's data



Enterprise-grade security and privacy

DEMO

Bedrock demos:

- Playground – text
- Playground – image
- Step function – MAYBE KILL



Knowledge Bases for Amazon Bedrock

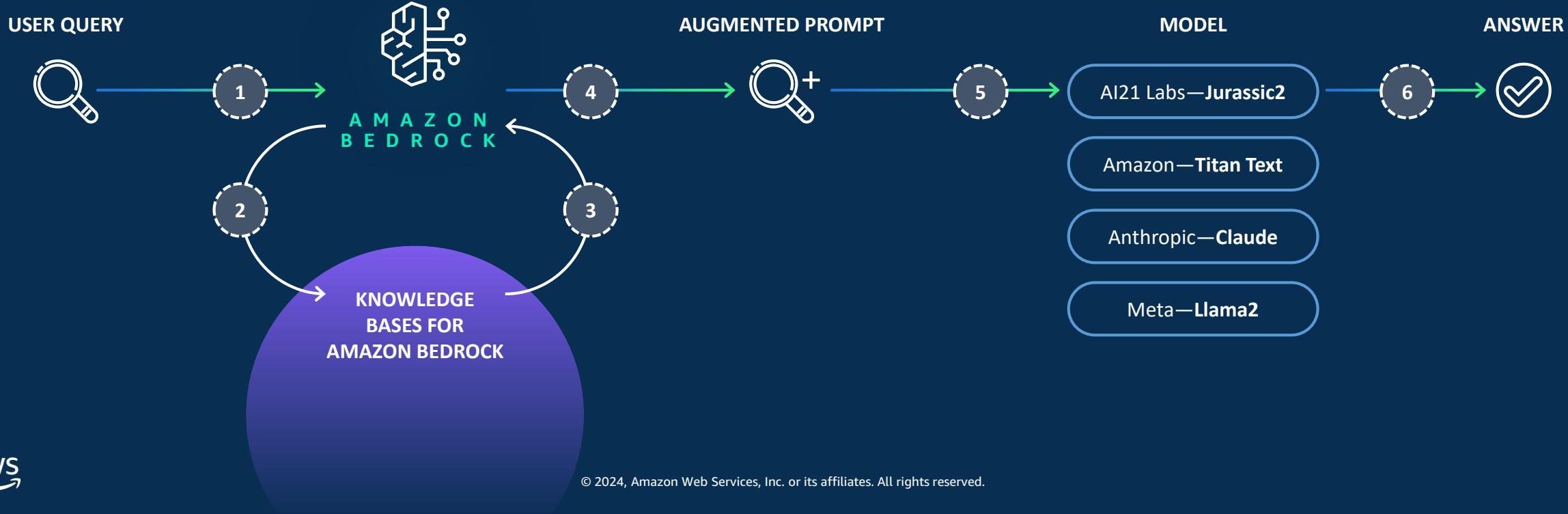
NATIVE SUPPORT FOR RETRIEVAL AUGMENTED GENERATION (RAG)

Securely connect FMs to data sources for RAG to deliver more relevant responses

Fully managed RAG workflow including ingestion, retrieval, and augmentation

Built-in session context management for multi-turn conversations

Automatic citations with retrievals to improve transparency



DEMO

Bedrock – Knowledge Base



Agents for Amazon Bedrock

ENABLE GENERATIVE AI APPLICATIONS TO EXECUTE MULTISTEP TASKS
USING COMPANY SYSTEMS AND DATA SOURCES



1

SELECT YOUR
FOUNDATION MODEL

2

PROVIDE BASIC
INSTRUCTIONS

3

SELECT RELEVANT
DATA SOURCES

4

SPECIFY AVAILABLE ACTIONS

| Breaks down and orchestrates tasks |

| Securely accesses and retrieves company data for RAG |

| Takes action by invoking API calls on your behalf |

| Chain-of-thought trace and ability to modify agent prompts |

NEW



Amazon Q

Your generative AI assistant designed for work that can be tailored to your business, data, code, and operations

AVAILABLE IN PREVIEW TODAY

Engages in conversations to solve problems, generate content, and take action

Understands your company information, code, and systems

Personalizes interactions based on your role and permissions

Built to be secure and private



DEMO

Amazon Q console

Amazon Q troubleshooting

Amazon Q in VSCode



Data is the differentiator



Components of end-to-end data foundation

NEW AND RECENT INNOVATIONS



Comprehensive
Set of tools
for any use case



Integrated
Easily connect
all your data



Governance
From end-to-end

ENHANCING AWS SERVICES WITH GENERATIVE AI

A comprehensive set of services for your data

DATA SOURCES

IOT / DEVICES

APP / LOGS

3RD PARTY DATA

FOR APPLICATIONS

-  Amazon Aurora
-  Amazon RDS
-  Amazon DynamoDB
-  Amazon MSK
-  Amazon OpenSearch Service

FOR ANALYTICS & ML

-  Data Warehouse
-  Amazon Redshift
-  Data Lake
-  Amazon S3
-  Big Data
-  Amazon EMR

Catalog & Govern



AWS Lake Formation

Act

MACHINE LEARNING

-  Amazon SageMaker

GENERATIVE AI

-  Amazon Bedrock

BUSINESS INTELLIGENCE

-  Amazon QuickSight



A comprehensive set of AWS database services

RELATIONAL



Amazon
RDS



Amazon
Aurora

SERVERLESS

KEY-VALUE



Amazon
DynamoDB

SERVERLESS

DOCUMENT



Amazon
DocumentDB

ELASTIC CLUSTERS

CACHE



Amazon
ElastiCache

SERVERLESS

GRAPH



Amazon
Neptune

SERVERLESS

TIME SERIES



Amazon
Timestream

SERVERLESS

LEDGER



Amazon
QLDB

SERVERLESS

WIDE COLUMN



Amazon
Keyspaces

SERVERLESS

IN-MEMORY



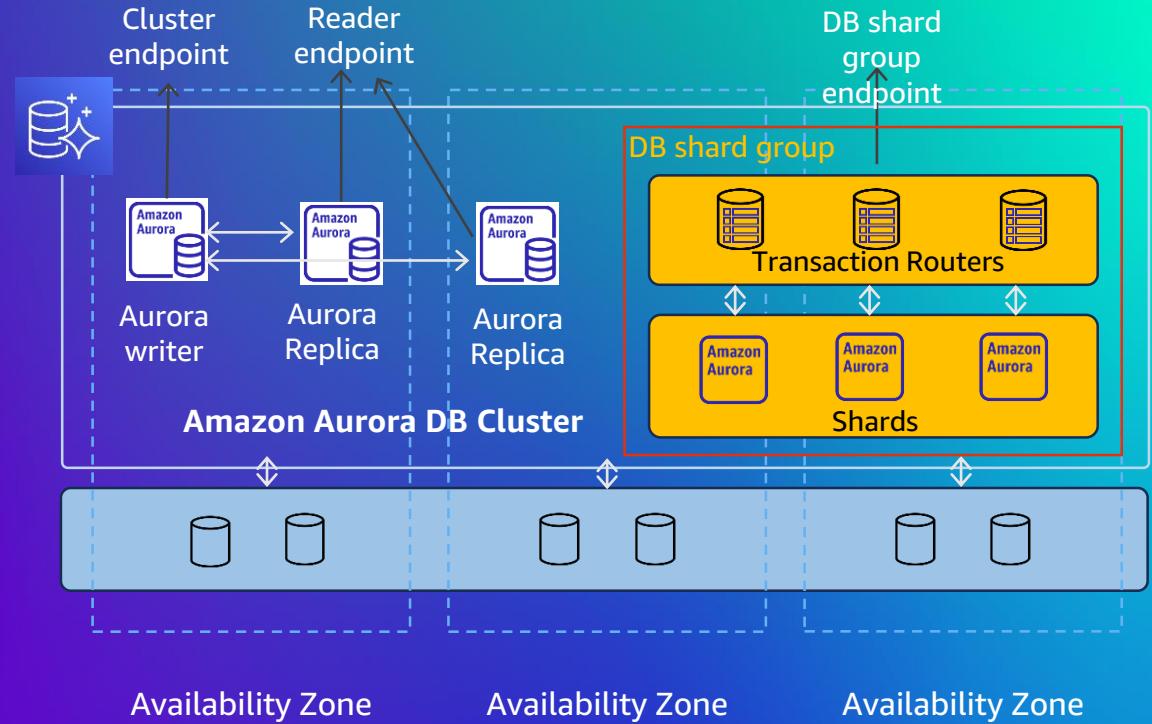
Amazon
MemoryDB

NEW

Amazon Aurora Limitless Database

Managed horizontal scale-out beyond
the limits of a single instance

AVAILABLE IN PREVIEW TODAY



A comprehensive set of AWS analytics services

INTERACTIVE QUERY



Amazon
Athena

SERVERLESS

BIG DATA



Amazon
EMR

SERVERLESS

REALTIME



Amazon Kinesis
Amazon MSK

SERVERLESS

DATA WAREHOUSE



Amazon
Redshift

SERVERLESS

ETL



AWS
Glue

SERVERLESS

BUSINESS INTELLIGENCE



Amazon
QuickSight

SERVERLESS

OPERATIONAL ANALYTICS



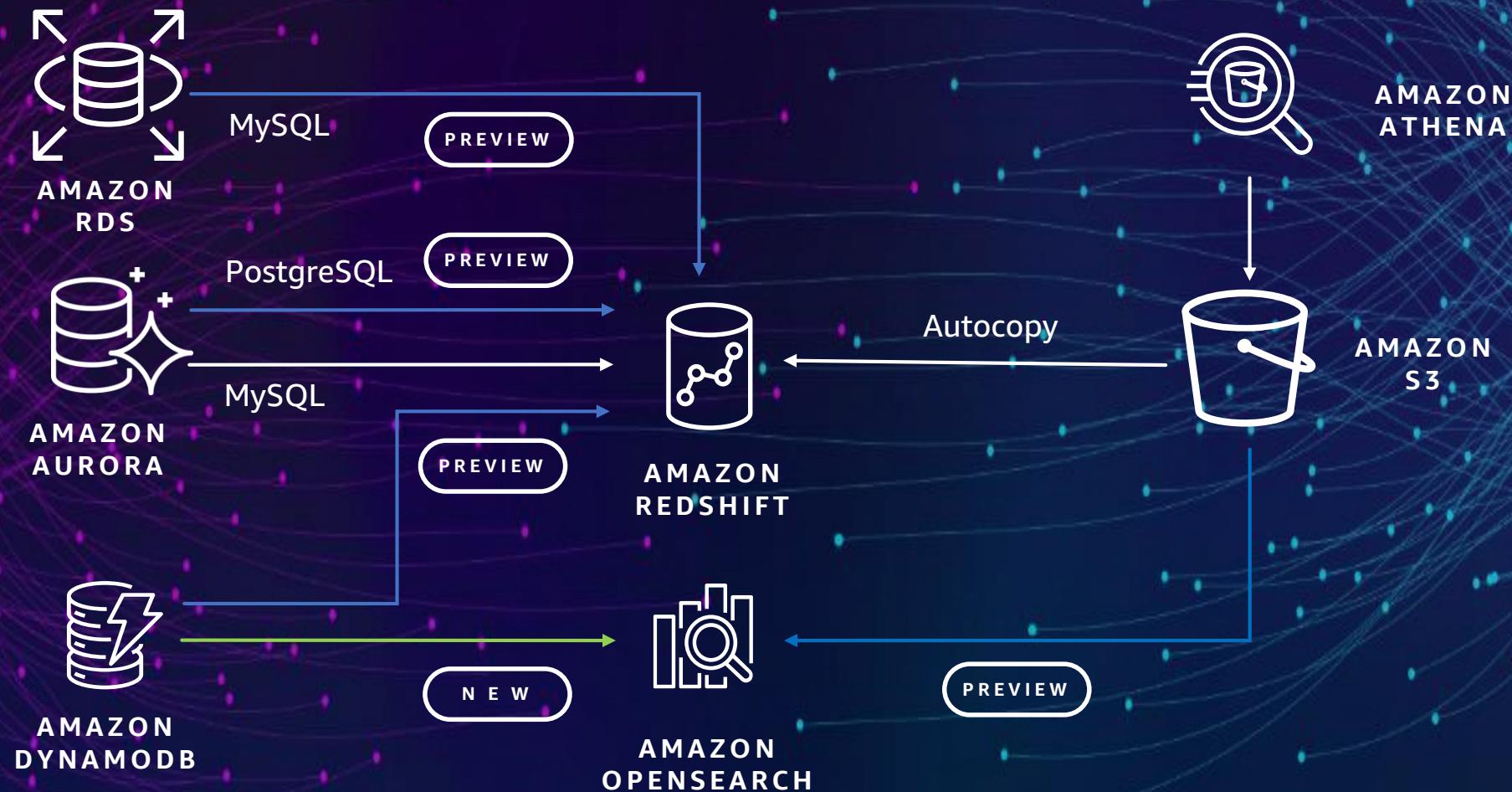
Amazon
OpenSearch Service

SERVERLESS

AI-based scaling and
optimization

PREVIEW

Zero-ETL





The Frugal Architect



THE FRUGAL ARCHITECT



DESIGN

- I. Cost is a Non-Functional Requirement
- II. Systems that Last Align Cost to Business
- III. Architecting is a Series of Trade-Offs

MEASURE

- IV. Unobserved Systems Lead to Unknown Costs
- V. Cost-Aware Architectures Implement Cost Controls

OPTIMIZE

- VI. Cost Optimization is Incremental
- VII. Unchallenged Success Leads to Assumptions

thefrugalarchitect.com



NEW

AWS Billing and Cost Management

Unified Billing and Management Console

Easy deployment from console and
Export granular cost and usage data

GENERALLY AVAILABLE

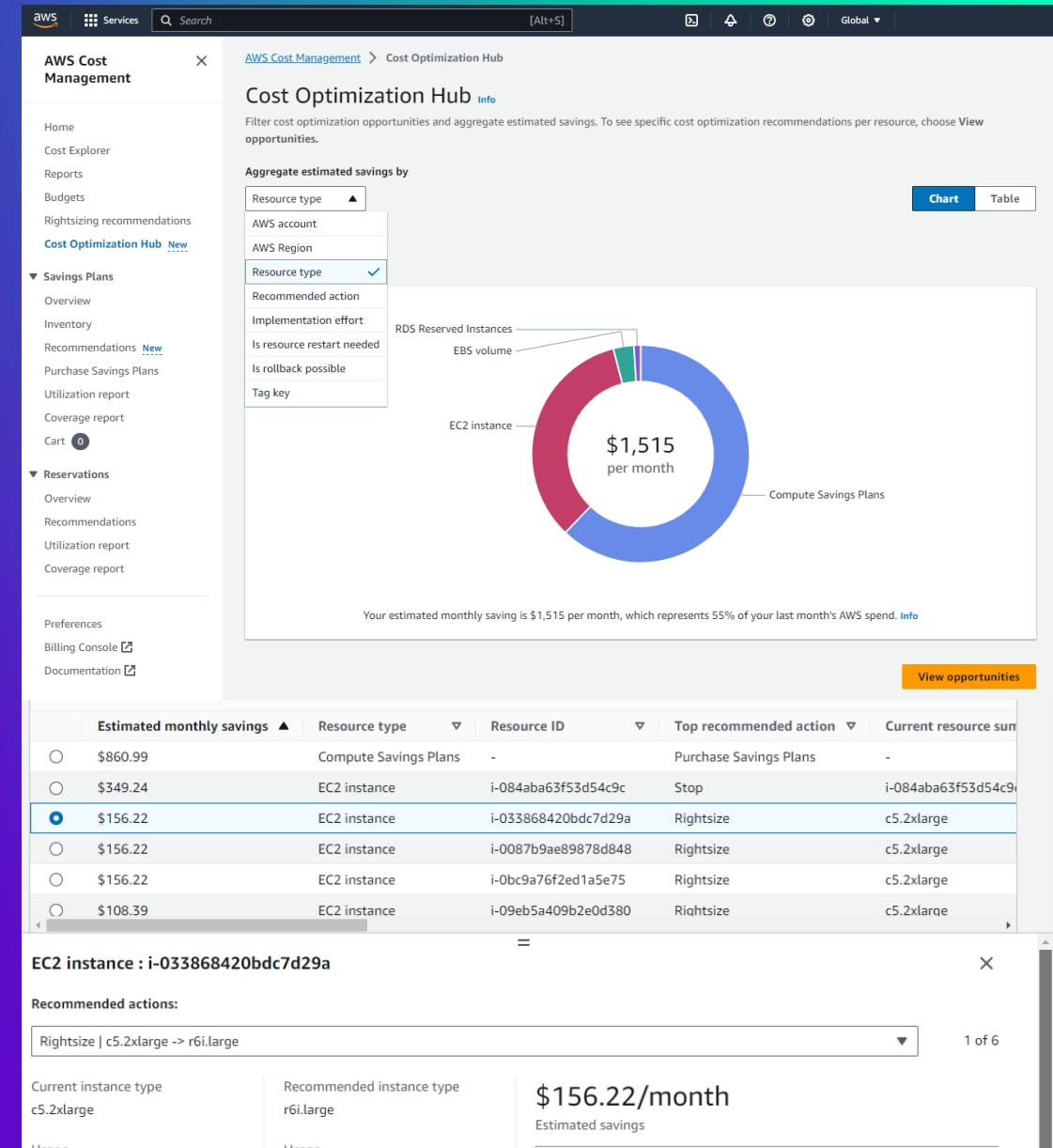
The screenshot shows the AWS Billing and Cost Management home page. On the left, a sidebar menu includes: Home, Getting Started, Billing and Payments (Bills, Payments, Credits, Purchase Orders), Cost Analysis (Cost Explorer, Cost Explorer Saved Reports, Cost Anomaly Detection, Free Tier), Cost Organization (Cost Categories, Cost Allocation Tags, Billing Conductor), Budgets and Planning (Budgets, Budgets Reports, Pricing Calculator), Savings and Commitments (Cost Optimization Hub, Savings Plans, Reservations), Preferences and Settings (Payment Preferences, Billing Preferences, Cost Management Preferences, Tax Settings), Legacy Pages (Billing Home, Cost Management Home), and New Navigation. The main content area displays: Cost summary (Month-to-date (MTD) cost: \$77,849.50, Last month's cost for same time period: \$76,323.04, Total forecasted cost for current month: \$87,985.60, Last month's total cost: \$85,422.91); Cost breakdown (Group costs by Cost category and Application, showing a stacked bar chart for Jan-Jun MTD); Recommended actions (Cost optimization: 5 Savings Plans expiring within 30 days, Getting started: Create a Savings Plans coverage or utilization budget alert); Cost allocation coverage (Display allocation coverage by Cost categories, showing a table for Application, Department, Environment, Project Waterloo Teams); and Savings opportunities (Total estimated monthly savings: \$16,230.18 USD, showing upgrade, migrate to Graviton, rightsize, purchase Savings Plans, other options). A top navigation bar shows the AWS logo, services, search bar, and user information (N. Virginia, MyRole/AWSUser @ 0123-4567-8901).

NEW

AWS Cost Optimization Hub

Centralizes recommended actions to save you money across multiple regions and AWS Accounts

GENERALLY AVAILABLE

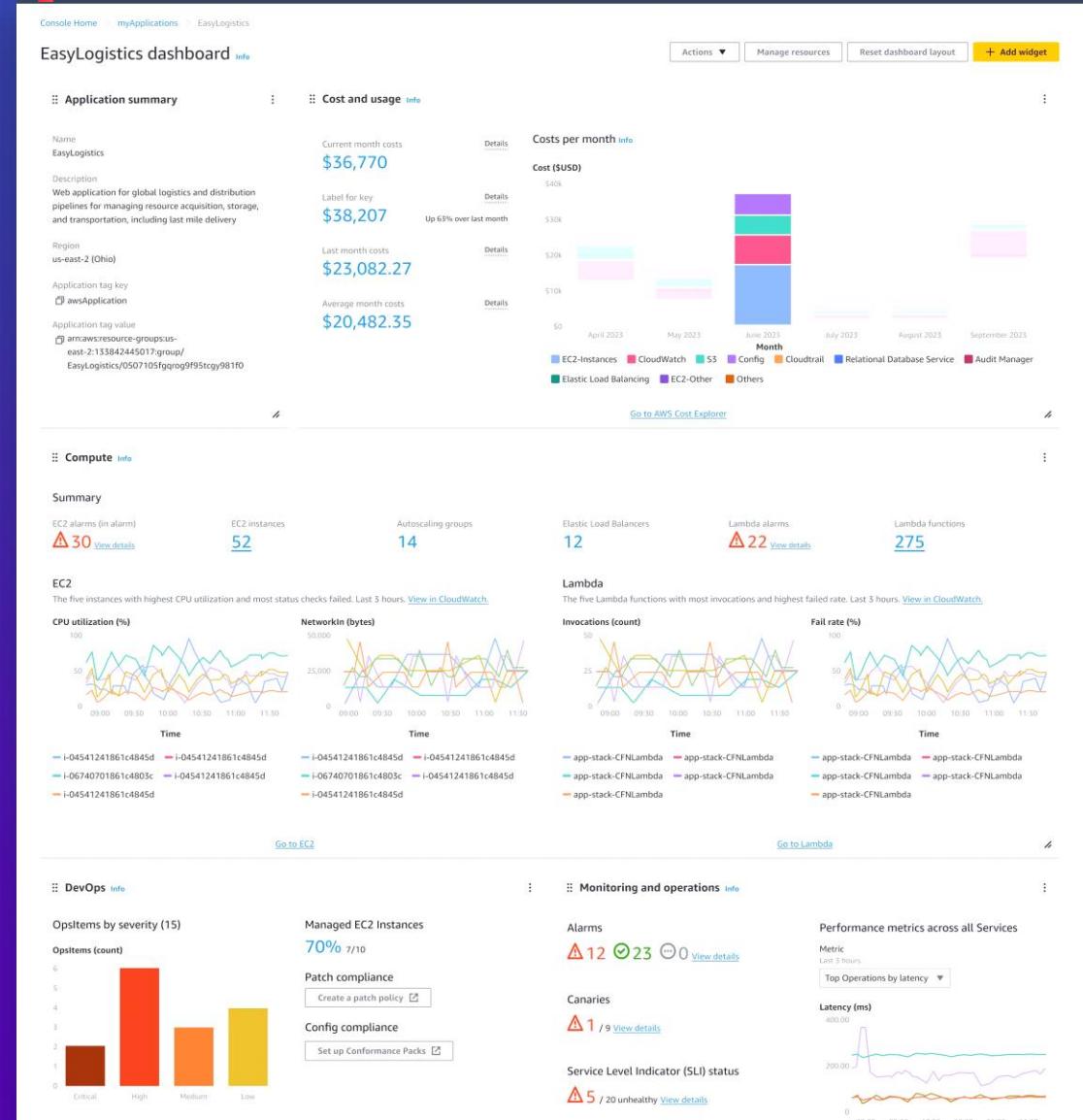


NEW

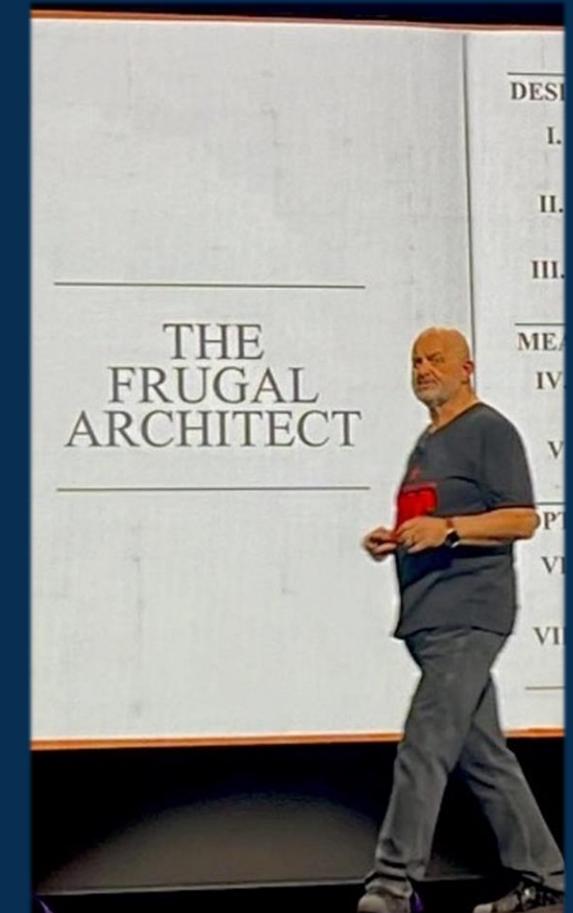
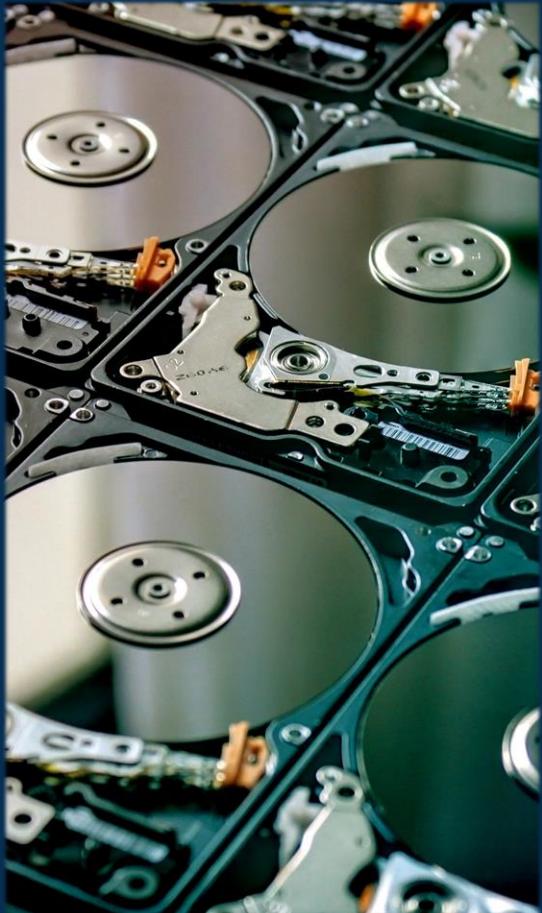
AWS Management Console myApplications

Monitor and manage the cost, health, security posture, and performance of your applications

GENERALLY AVAILABLE



Highlights from AWS re:Invent 2023



Let's Learn Together!

<https://community.aws>

The screenshot shows the AWS Community website homepage. The header features the AWS logo and navigation links for Home, Tags, Featured Spaces, Community Programs, Events, and a search bar. The main banner has a globe graphic and the text "Builders, Welcome Home". Below the banner, there are sections for featured spaces like Cost Optimization, DevOps, Generative AI, Kubernetes, Livestreams, and Resilience. A "Community Programs" section lists AWS Heroes, AWS Community Builders, AWS User Groups, and Student Communities. An "Events" section shows a "re:Caps" event. The main content area displays a post by Dennis Liang titled "We Built and Deployed Lead Scoring ML model - Here's What and How", which uses SageMaker and AutoML for marketing prioritization. Another post by Matias Kreder is also visible. On the right, there are "Announcements" about generative AI and Amazon Bedrock, and a call to join local Cloud Clubs.

aws COMMUNITY

Search for content

Home

Tags

Featured Spaces

- Cost Optimization
- DevOps
- Generative AI
- Kubernetes
- Livestreams
- Resilience

Community Programs

- AWS Heroes
- AWS Community Builders
- AWS User Groups
- Student Communities

Events

re:Caps

Dennis Liang

We Built and Deployed Lead Scoring ML model - Here's What and How

Use SageMaker and AutoML to automate the prioritization and customer targeting in Marketing.

3 hours ago • machine-learning, amazon-sagemaker

1 ...

Matias Kreder

Testing Amazon Bedrock Text G1 Models (Lite vs Express)

Announcements

Learn the fundamentals of how generative AI works, and how to deploy it in real-world applications. [Learn More](#).

Learn more about [Amazon Bedrock](#) with these fantastic [articles](#) to jumpstart your journey.

Join your local [Cloud Club](#)! Now open globally.

AWS TRAINING & CERTIFICATION

Access 600+ free digital courses with AWS Skill Builder

Focus on the cloud skills and services that are most relevant to you across 30+ AWS solutions, including digital self-paced learning plans and ramp-up guides.

LEARN YOUR WAY, EXPLORE SKILLBUILDER.AWS »



Validate your cloud expertise with an AWS Certification

Take the step towards earning an industry-recognised credential. Learn more about how to become an AWS Certified, and AWS resources that can help you prepare.

ACCESS RESOURCES TO PREPARE FOR YOUR EXAM »

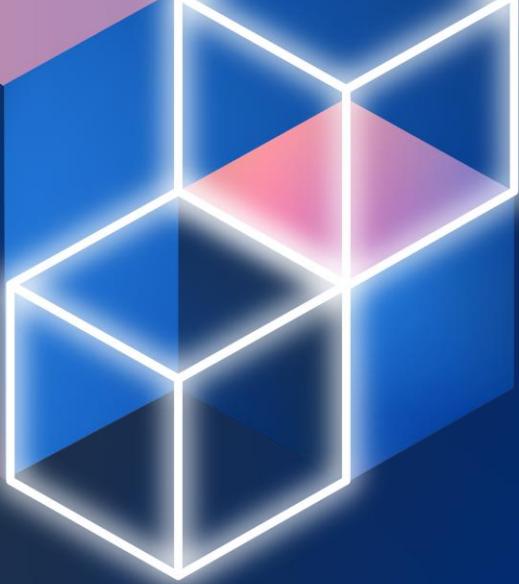


Thank you for attending AWS Builders Online Series

We hope you found it interesting! A kind reminder to **complete the survey**. Let us know what you thought of today's event and how we can improve the event experience for you in the future.

-  aws-apj-marketing@amazon.com
-  twitter.com/AWSCloud
-  facebook.com/AmazonWebServices
-  youtube.com/user/AmazonWebServices
-  linkedin.com/company/amazon-web-services
-  twitch.tv/aws





Thank you!

Kris Howard

Head of Developer
Relations, APJ
AWS

Donnie Prakoso

Principal Developer
Advocate
AWS

