AWS Builders Online Series

CLOSING KEYNOTE

AWS re:Invent Highlights

Kris Howard
Head of Developer Relations, APJ
AWS

Donnie Prakoso
Principal Developer Advocate
AWS

Who is a Builder?

Why do Builders choose AWS?

Sustainability

Breadth of services

productivity

Pace of innovation

Community

Global infrastructure

Trust

Agility

Resilience

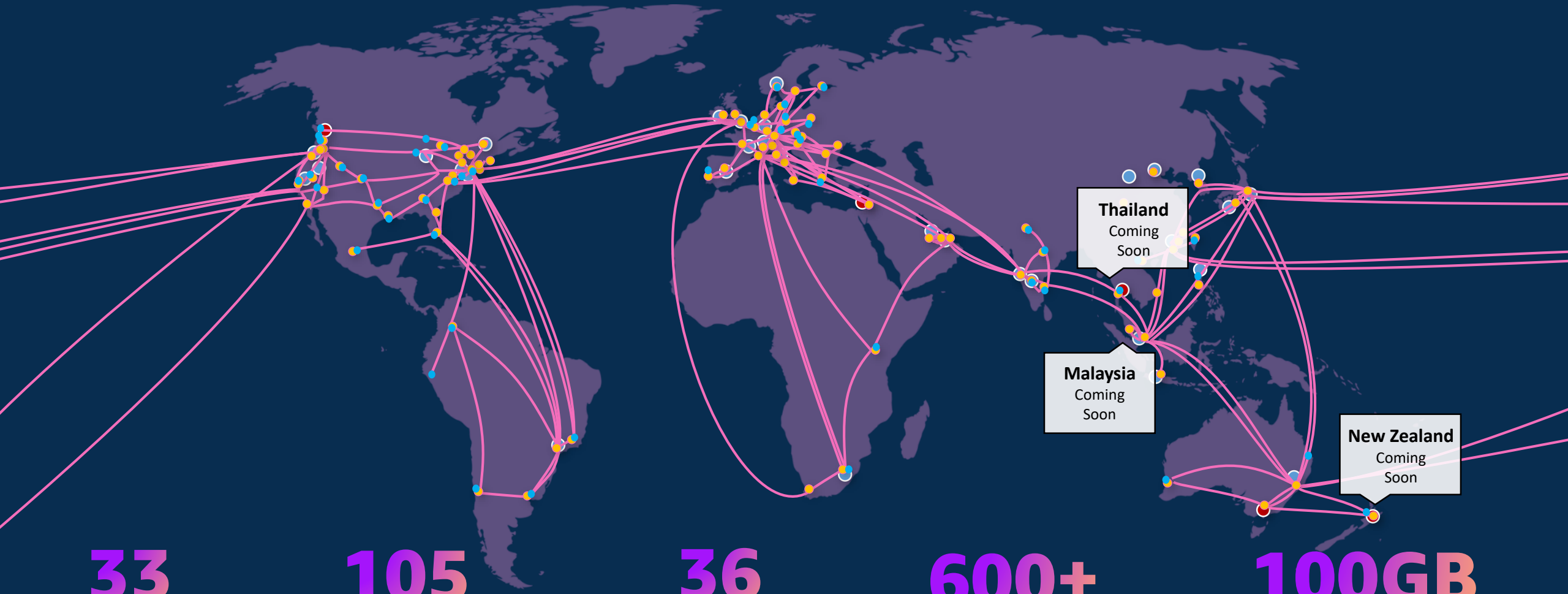
Customer focus

Cost

Security

Performance

Safe and secure global infrastructure



33
Regions
4 Coming Soon

105
Availability Zones (AZ)

36
Local zones

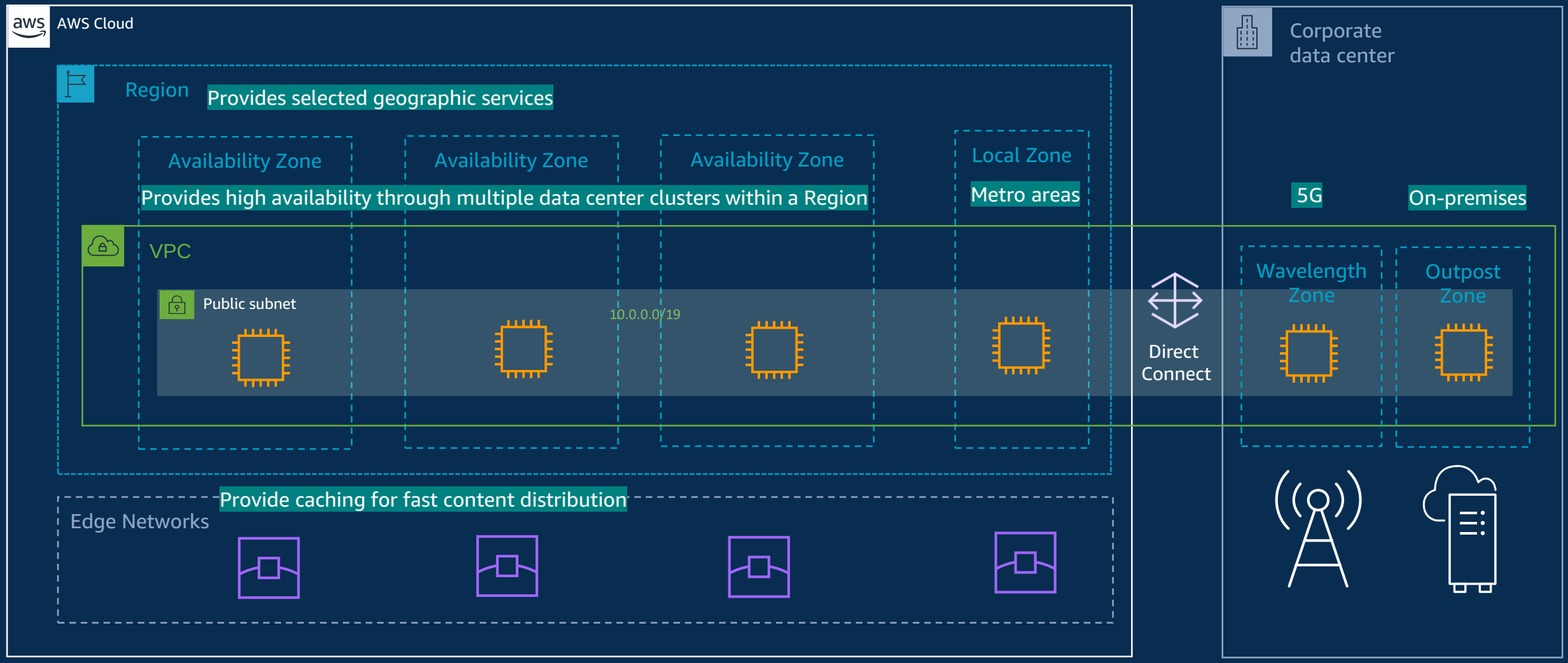
600+
PoPs

100GB
100% encryption
private network

aws

https://aws.amazon.com/ko/about-aws/global-infrastructure/ (Jan. 2024)

Safe and secure global infrastructure



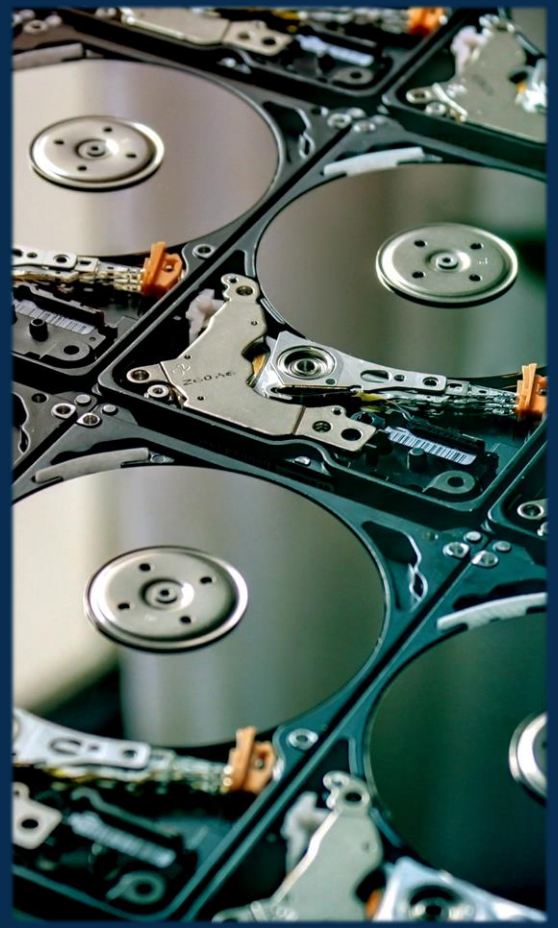
The broadest and deepest cloud

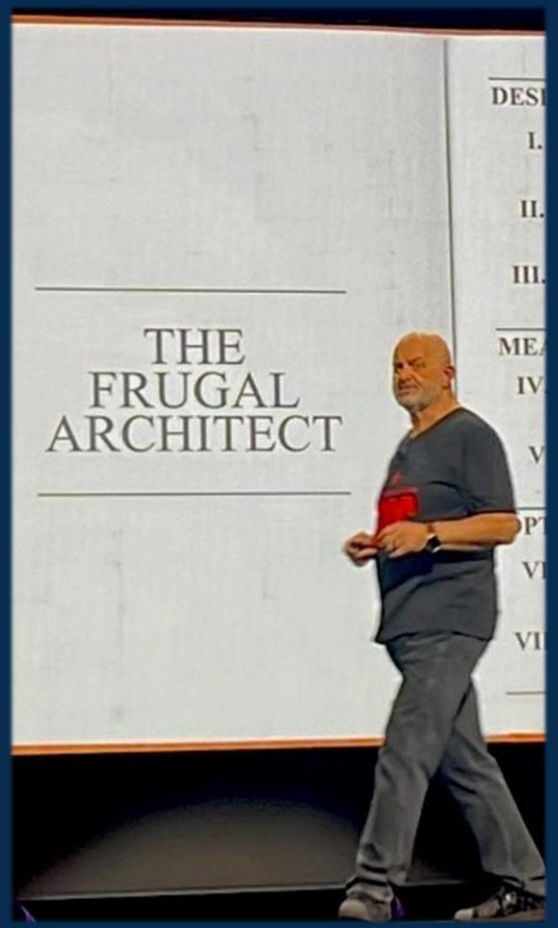
Customer Enablement	Cost Management	End User Computing	Internet of Things	Robotics, Blockchain, AR/VR, Satellite, Quantum Technology
5 SERVICES	4 SERVICES	3 SERVICES	12 SERVICES	5 SERVICES
Machine Learning	Media Services	Developer Tools	Application Integration	Front-end Web & Mobile, Gaming
26 SERVICES	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

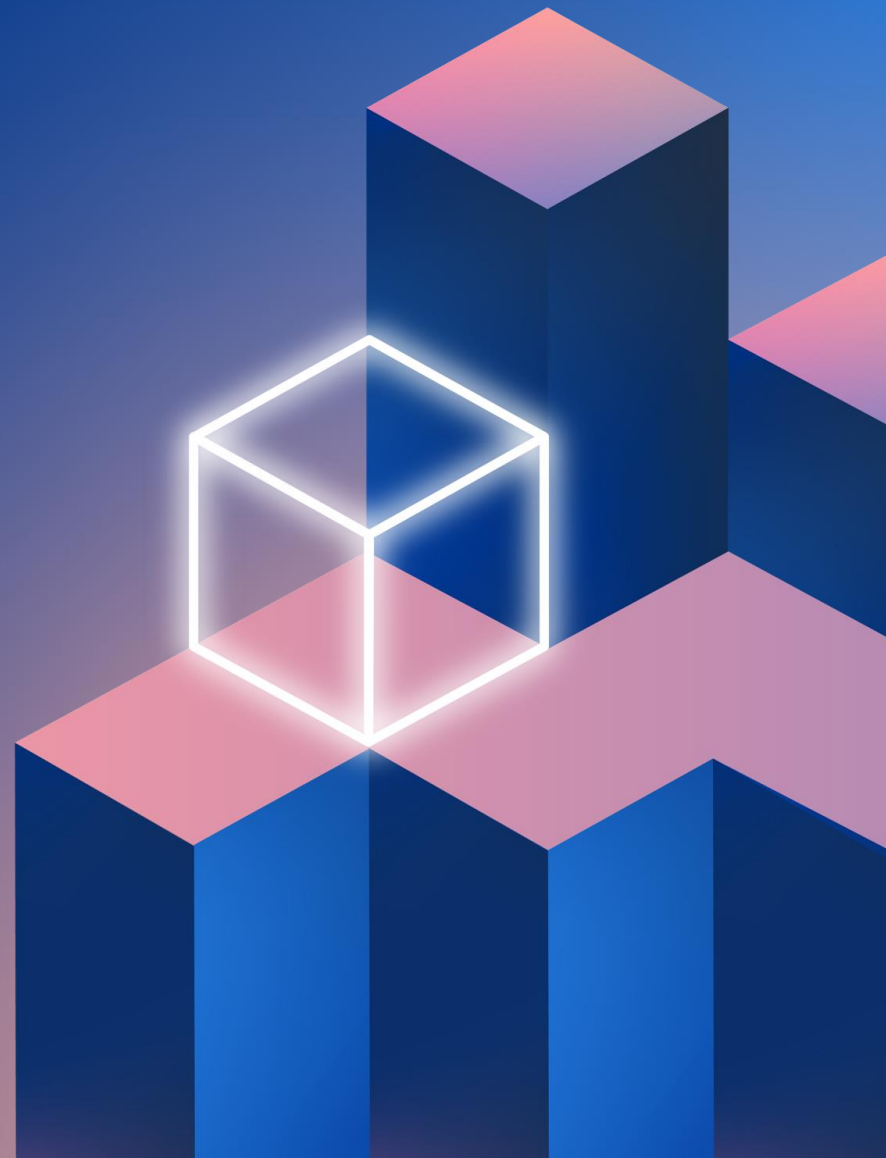
aws

Highlights from AWS re:Invent 2023





Compute and Storage



Your choice of Amazon S3 storage classes

S3 Standard	S3 Intelligent-Tiering	S3 Standard-IA	S3 One Zone-IA	S3 Glacier	S3 Glacier Deep Archive
Frequent			*Access frequency*		**Infrequent**
• Active, frequently accessed data • Milliseconds access • ≥ 3 AZ	• Data with changing access patterns • Milliseconds access • ≥ 3 AZ	• Infrequently accessed data • Milliseconds access • ≥ 3 AZ	• Re-creatable, less accessed data • Milliseconds access • 1 AZ	• Archive data • Minutes or hours access • ≥ 3 AZ	• Archive data • Hours to access • ≥ 3 AZ

6 Storage Classes

NEW

Amazon S3 **Express One Zone**

Highest performance and lowest latency cloud object storage

GENERALLY AVAILABLE

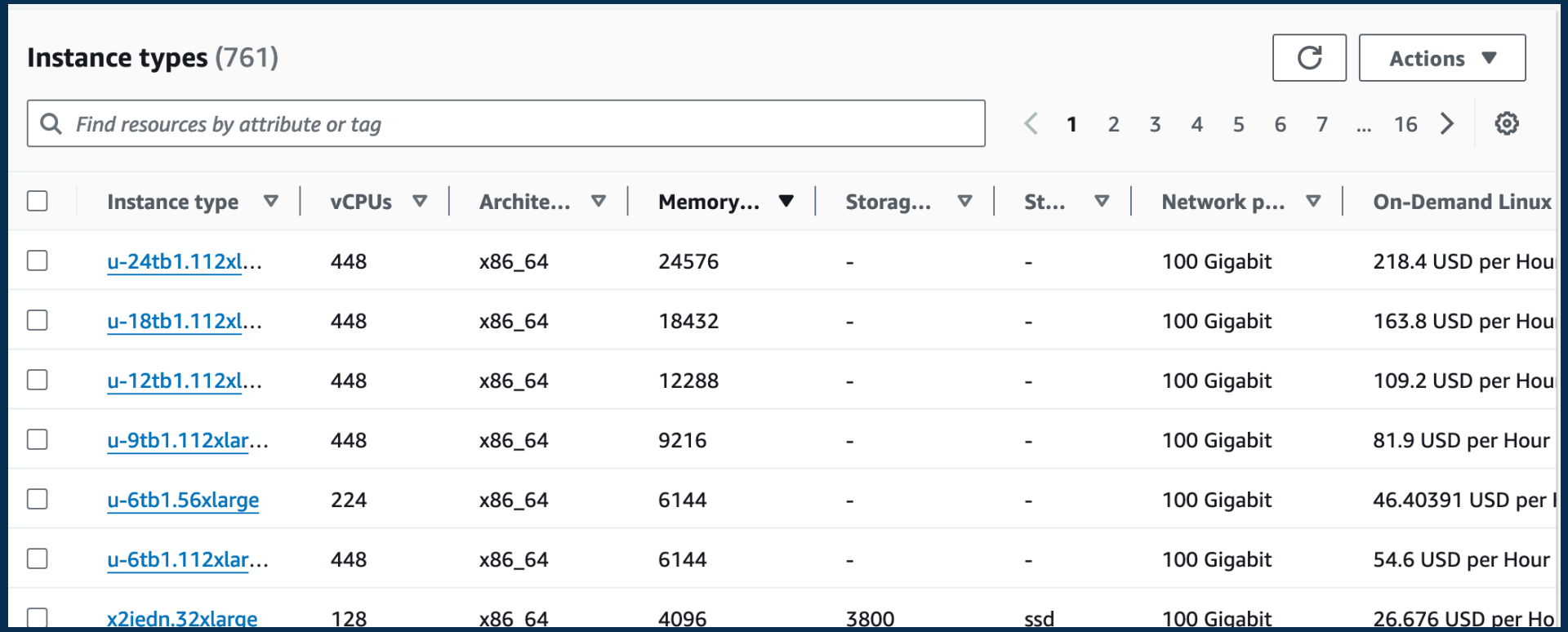
Millions of access requests per minute

Single-digit millisecond latency

50% lower access costs vs S3 Standard

AWS App Runner
AWS Batch
EC2 Image Builder
AWS Elastic Beanstalk
AWS Lambda
Amazon Lightsail
AWS Outposts
AWS Snow Family
AWS Wavelength
...

Your choice of Amazon EC2 instance types



Instance types (761)

Actions

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
u-24tb1.112xl...	448	x86_64	24576	-	-	100 Gigabit	218.4 USD per Hou
u-18tb1.112xl...	448	x86_64	18432	-	-	100 Gigabit	163.8 USD per Hou
u-12tb1.112xl...	448	x86_64	12288	-	-	100 Gigabit	109.2 USD per Hou
u-9tb1.112xlar...	448	x86_64	9216	-	-	100 Gigabit	81.9 USD per Hour
u-6tb1.56xlarge	224	x86_64	6144	-	-	100 Gigabit	46.40391 USD per I
u-6tb1.112xlar...	448	x86_64	6144	-	-	100 Gigabit	54.6 USD per Hour
x2iedn.32xlarge	128	x86_64	4096	3800	ssd	100 Gigabit	26.676 USD per Ho

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

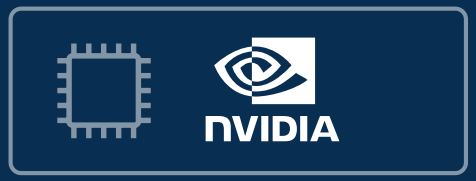
- Processor types (AWS, Intel, AMD)
- Processor speed (up to 4.5GHz)
- High footprint (up to 24TiB)
- Instance storage (HDD and SSD)
- Accelerated computing (GPU, FPGA , and ASIC)
- Networking (up to 800 Gbps)
- Bare metal

OPTIONS

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

aws

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

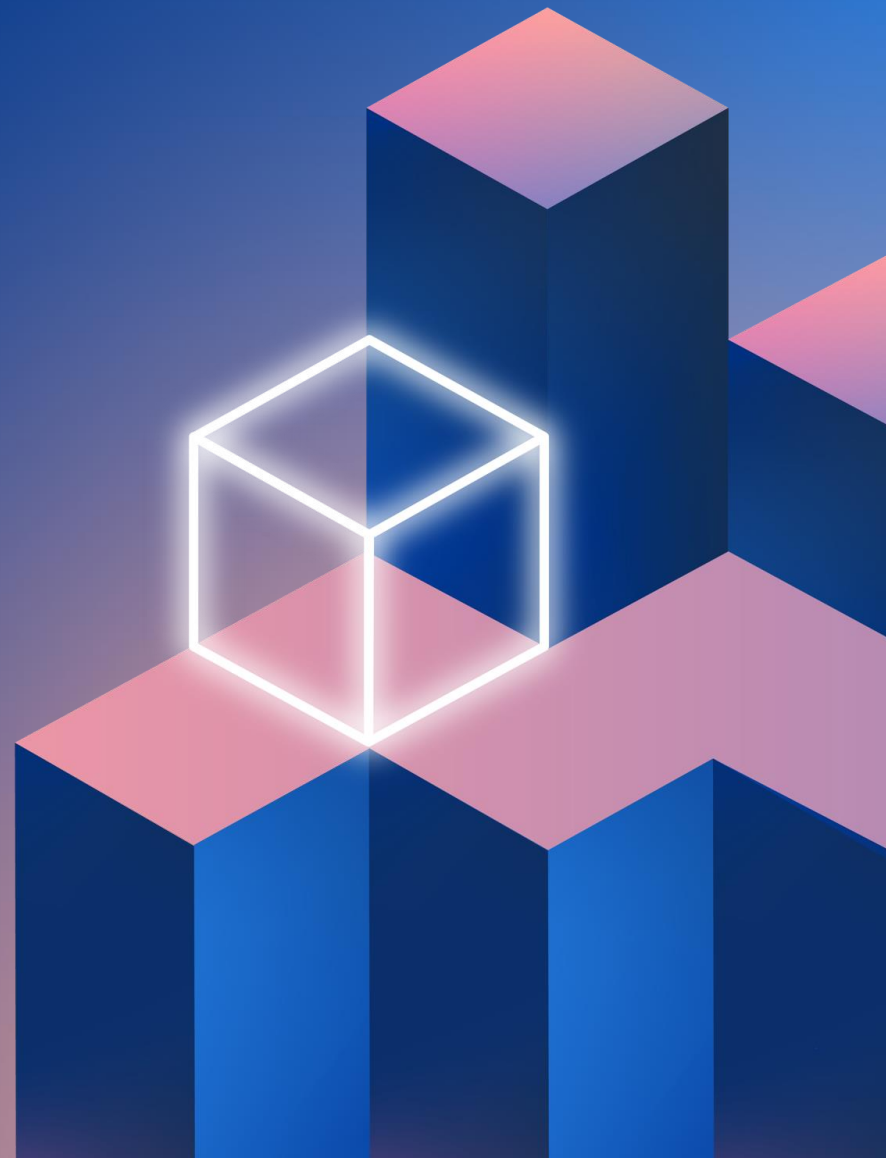
EC2 G6/G6e Instances

NVIDIA L4 GPU and L40S GPUs

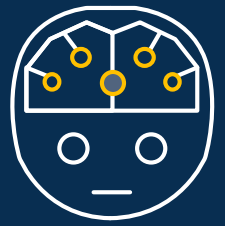
COMING IN 2024

Support for Small Language Models and 3D graphics simulations

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)

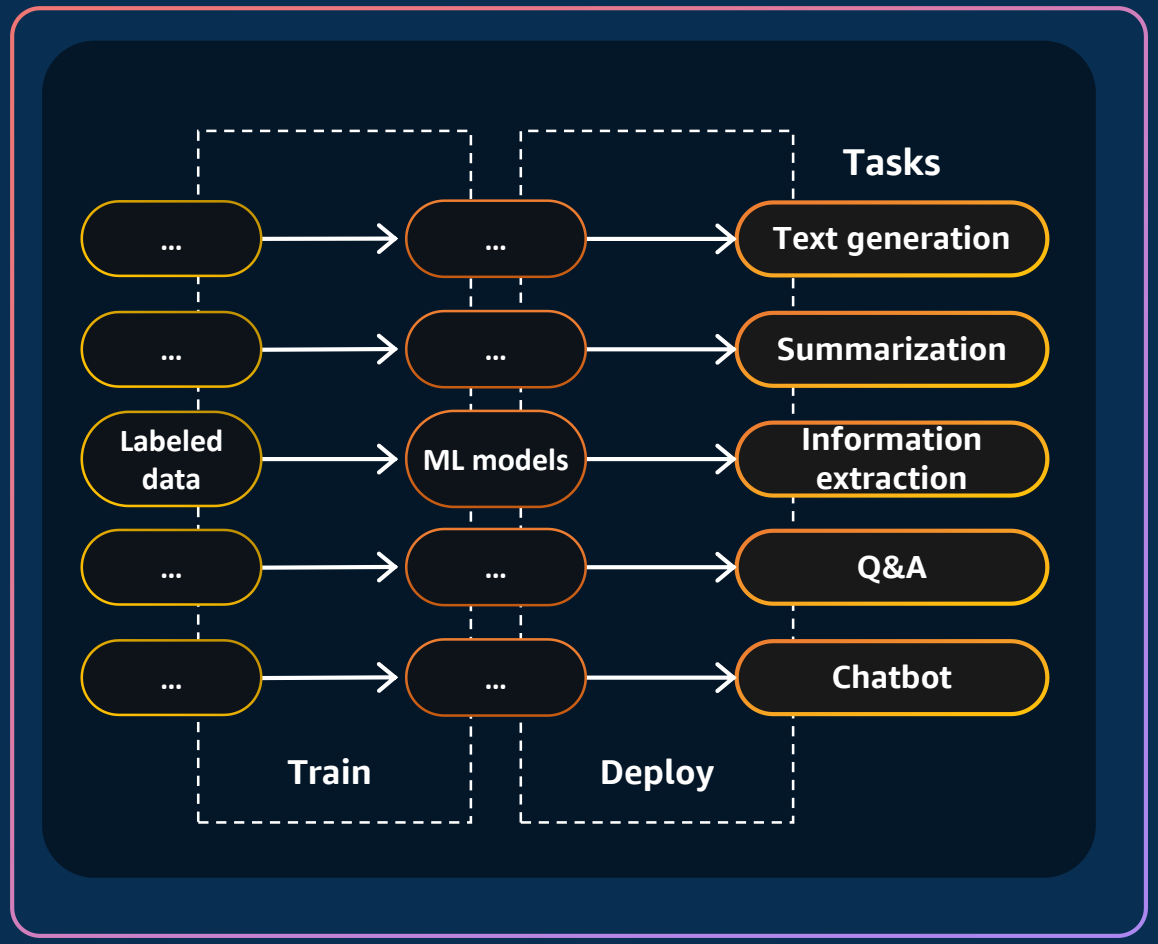
A subset of ML composed of deeply multi-layered neural networks that perform tasks like speech and image recognition

Generative AI

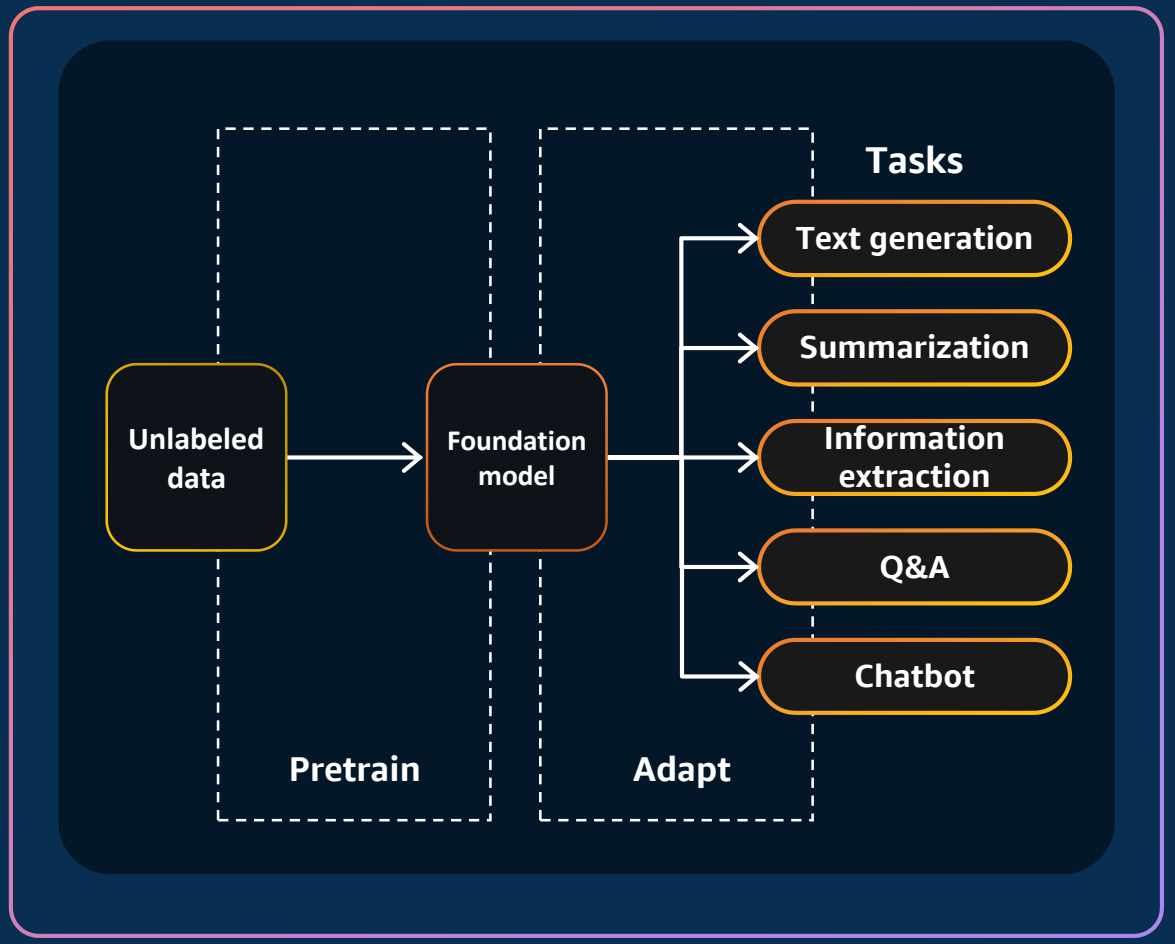
Powered by large models that are pretrained on vast corpora of data and commonly referred to as foundation models (FMs)

What's a Foundation Model (FM)?

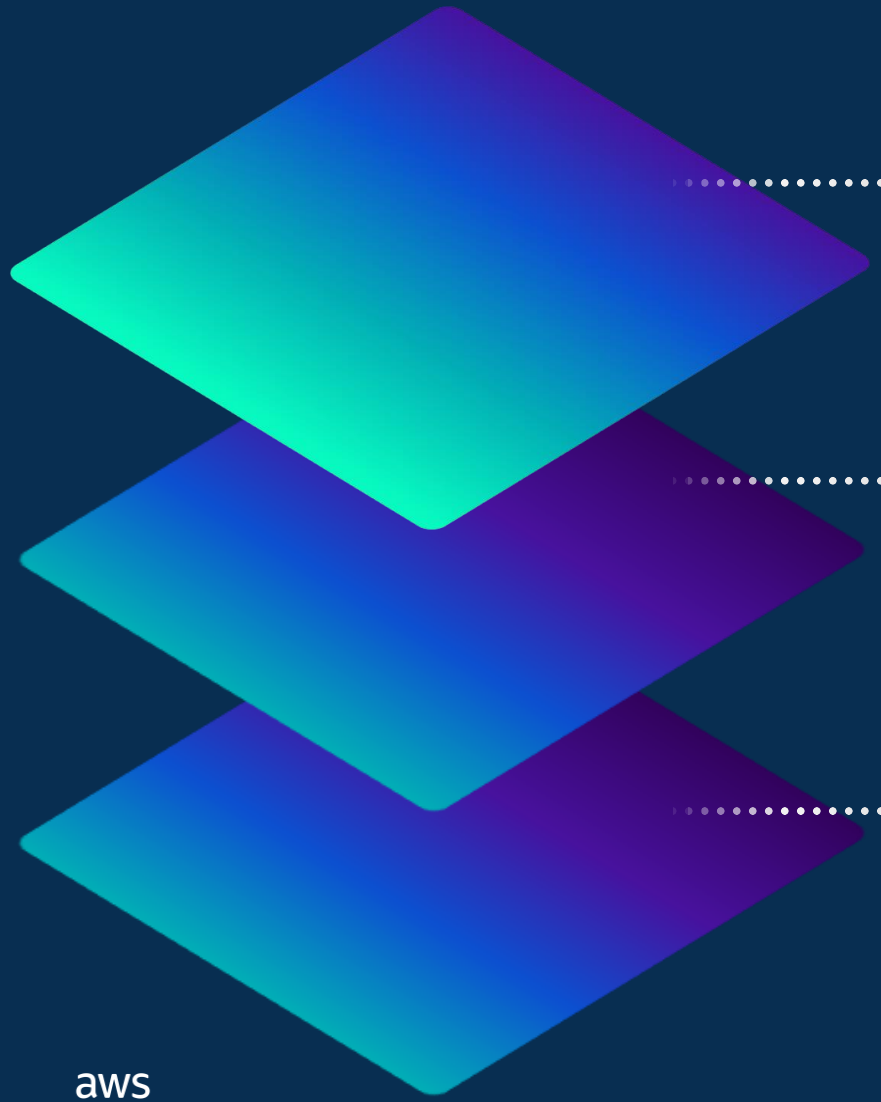
Traditional ML Model



Foundation Model



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

aws

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

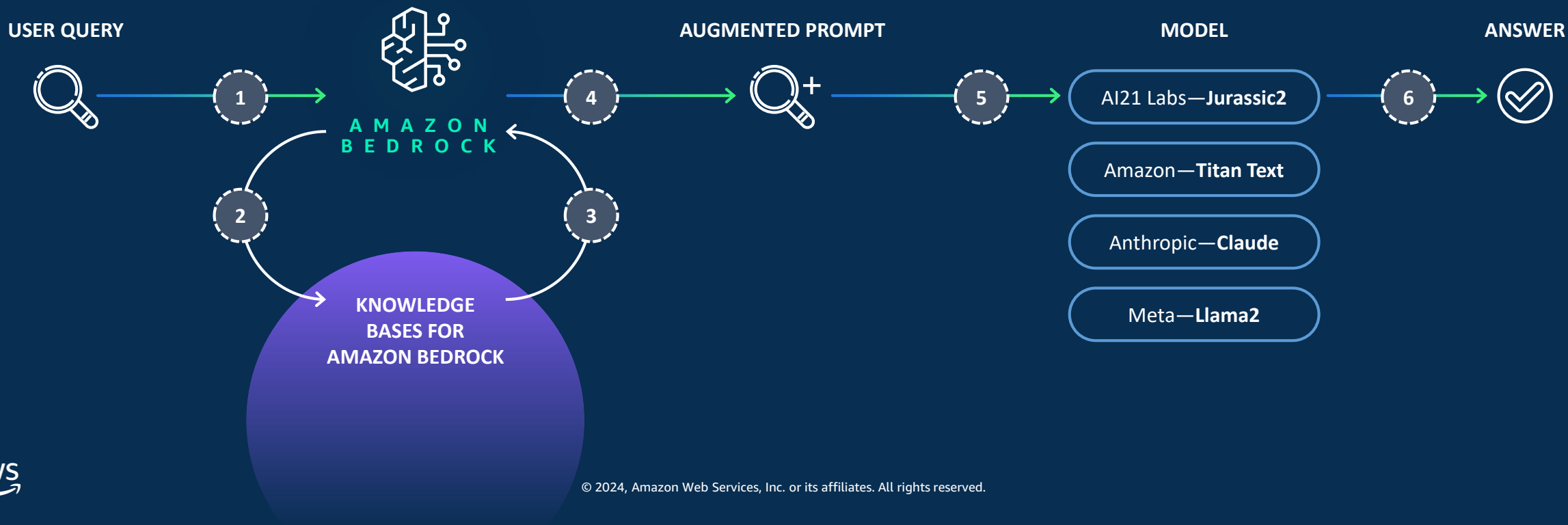
NATIVE SUPPORT FOR RETRIEVAL AUGMENTED GENERATION (RAG)

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 |

aws

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

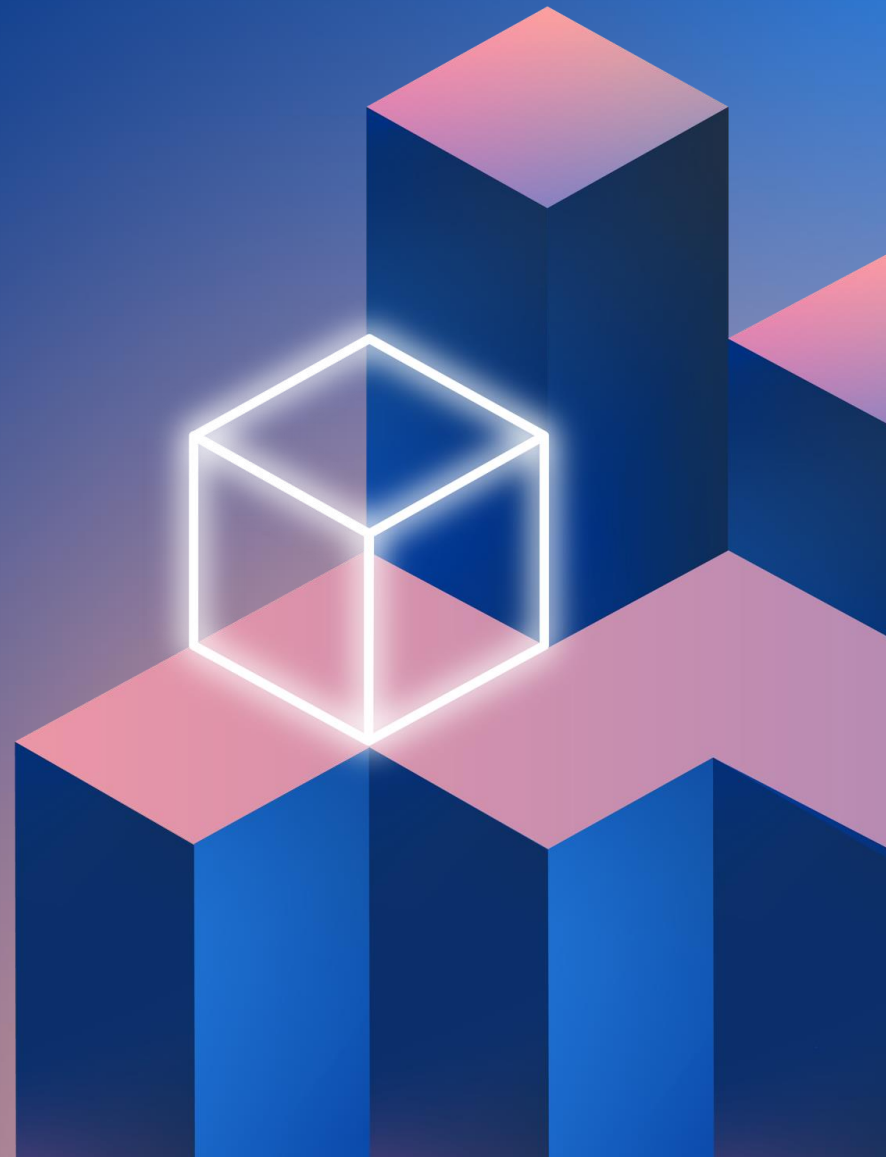
Built to be **secure and private**

aws

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

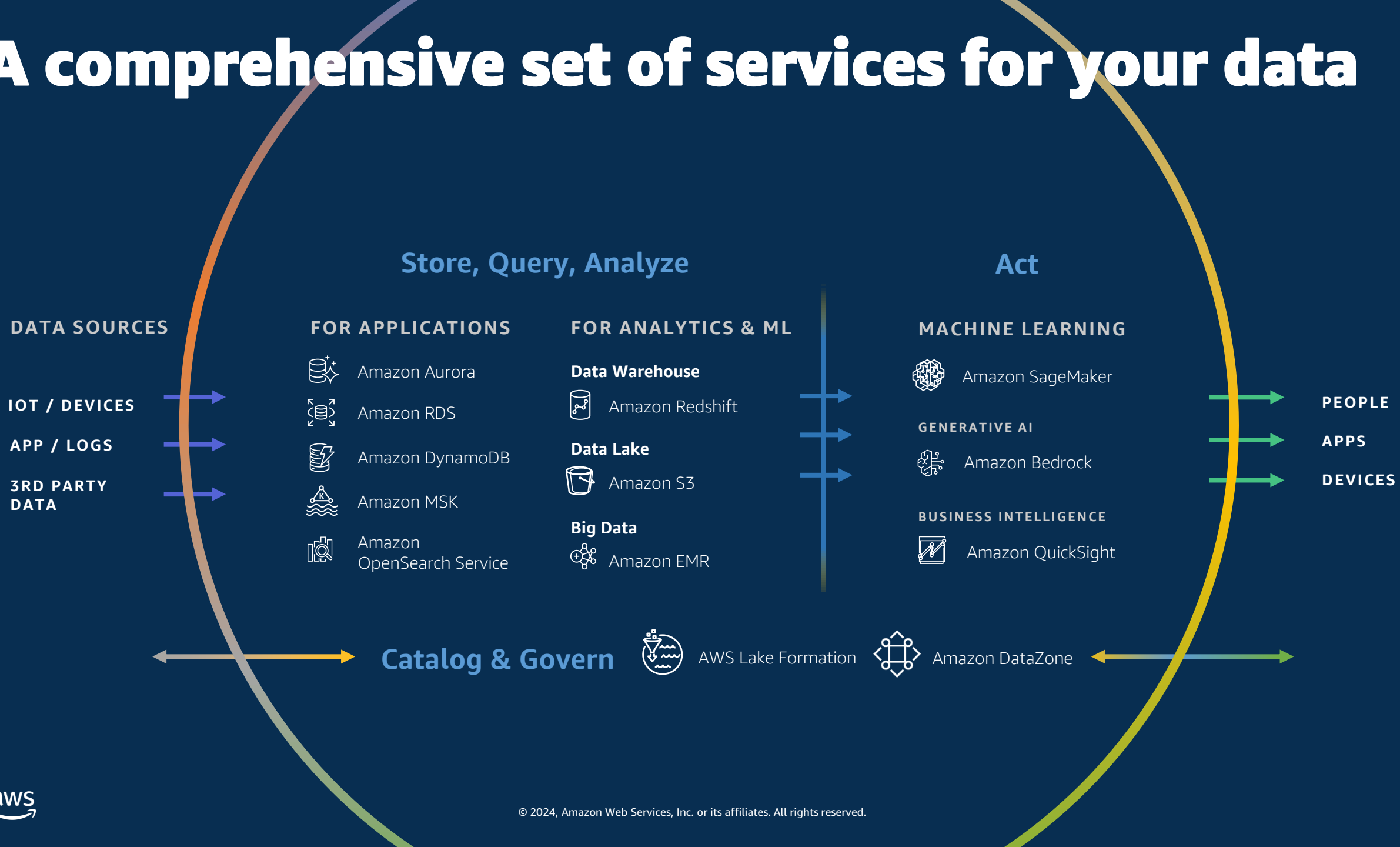
Comprehensive
Set of tools
for any use case

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



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

NEW

GRAPH

Amazon Neptune

SERVERLESS

TIME SERIES

Amazon Timestream

SERVERLESS

LEDGER

Amazon QLDB

SERVERLESS

WIDE COLUMN

Amazon Keyspaces

SERVERLESS

IN-MEMORY

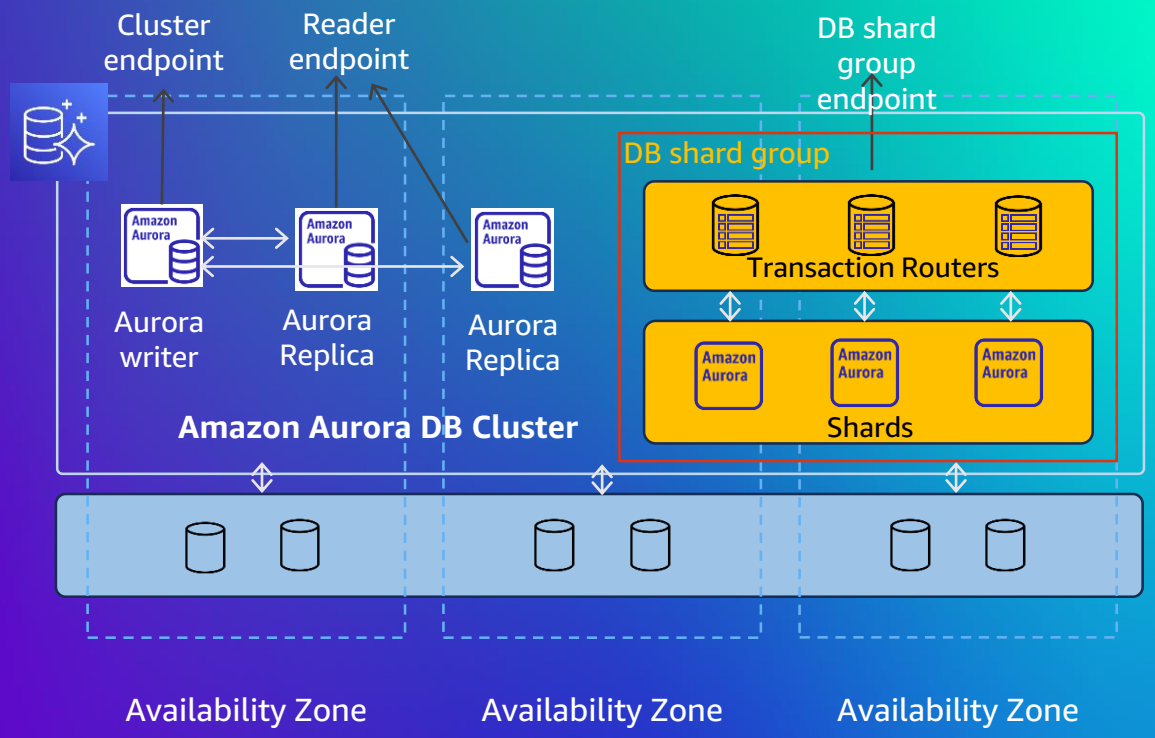
Amazon MemoryDB

NEW

Amazon Aurora **Limitless Database**

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

AVAILABLE IN PREVIEW TODAY



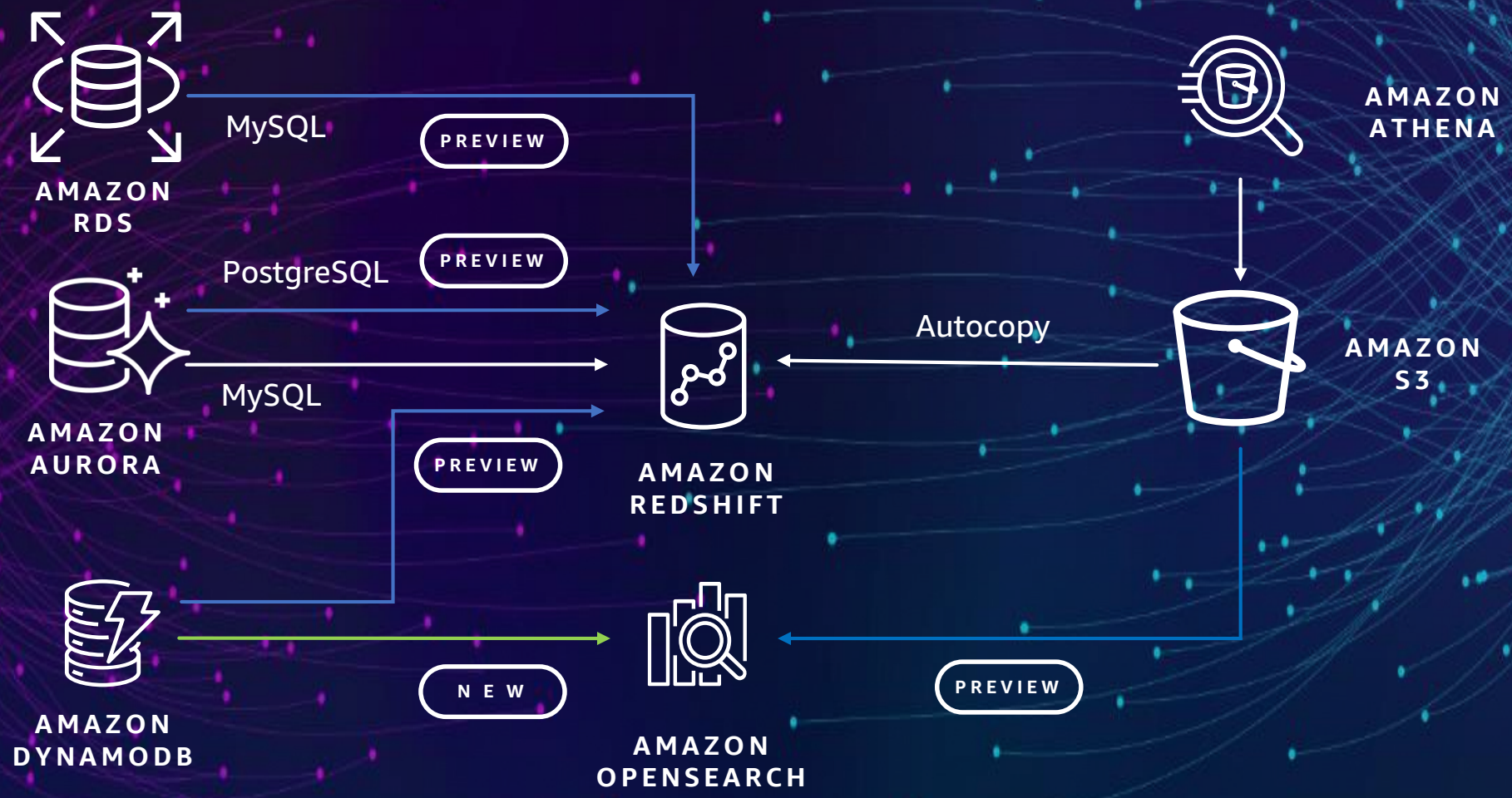
A comprehensive set of AWS analytics services

INTERACTIVE QUERY	BIG DATA	REALTIME	DATA WAREHOUSE	ETL	BUSINESS INTELLIGENCE	OPERATIONAL ANALYTICS
						
Amazon Athena	Amazon EMR	Amazon Kinesis Amazon MSK	Amazon Redshift	AWS Glue	Amazon QuickSight	Amazon OpenSearch Service
SERVERLESS	SERVERLESS	SERVERLESS	SERVERLESS	SERVERLESS	SERVERLESS	SERVERLESS

AI-based scaling and optimization

PREVIEW

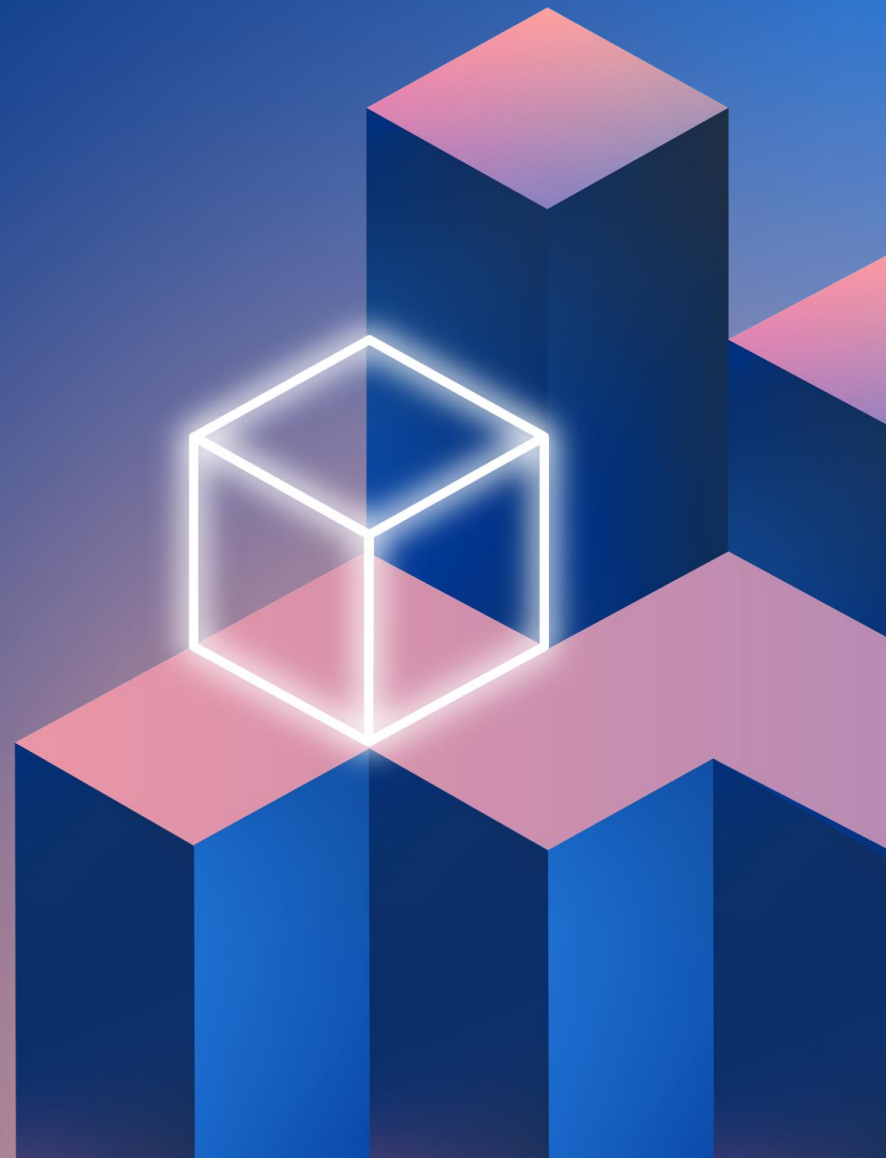
Zero-ETL

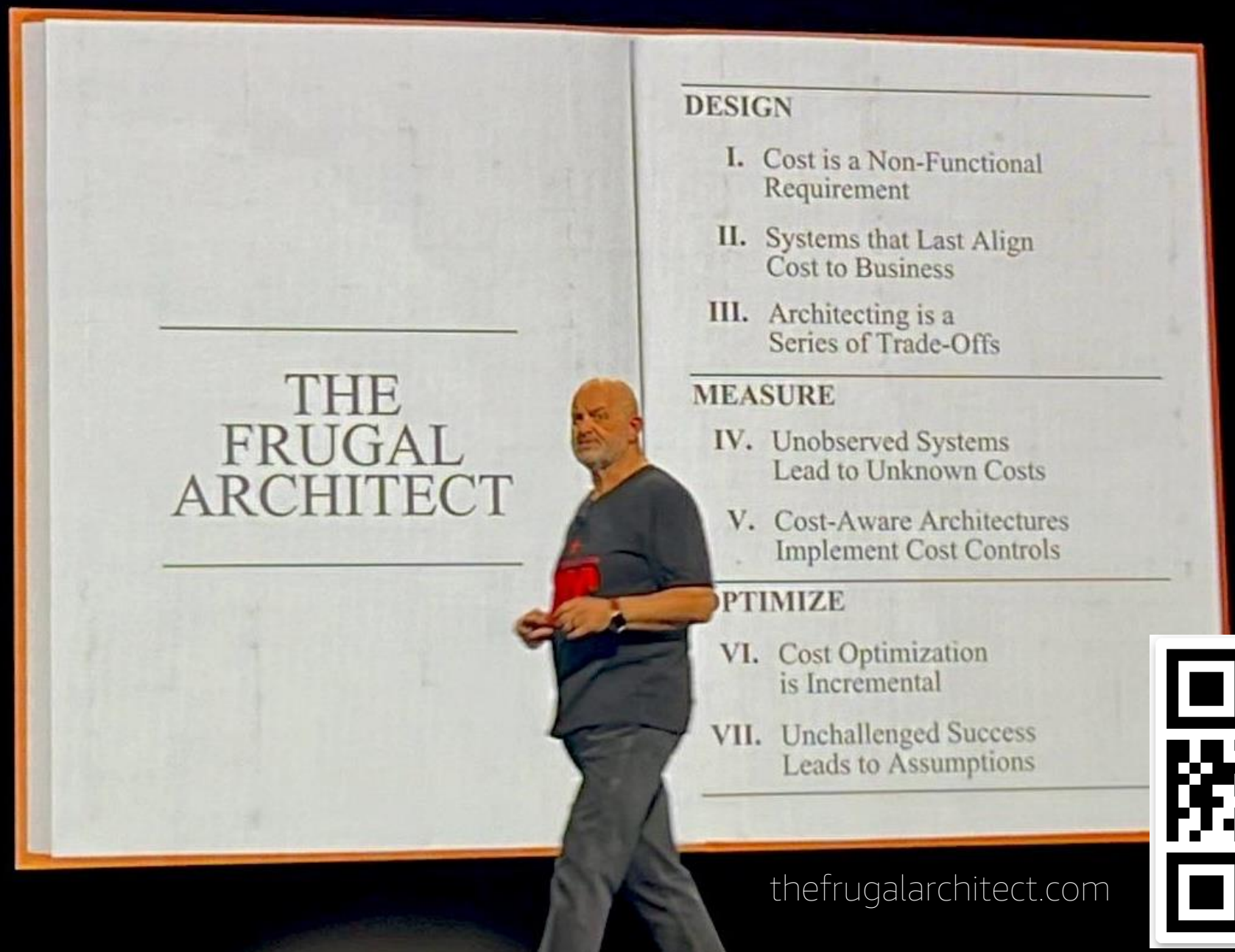


BLU

SMART MOBILITY

The Frugal Architect





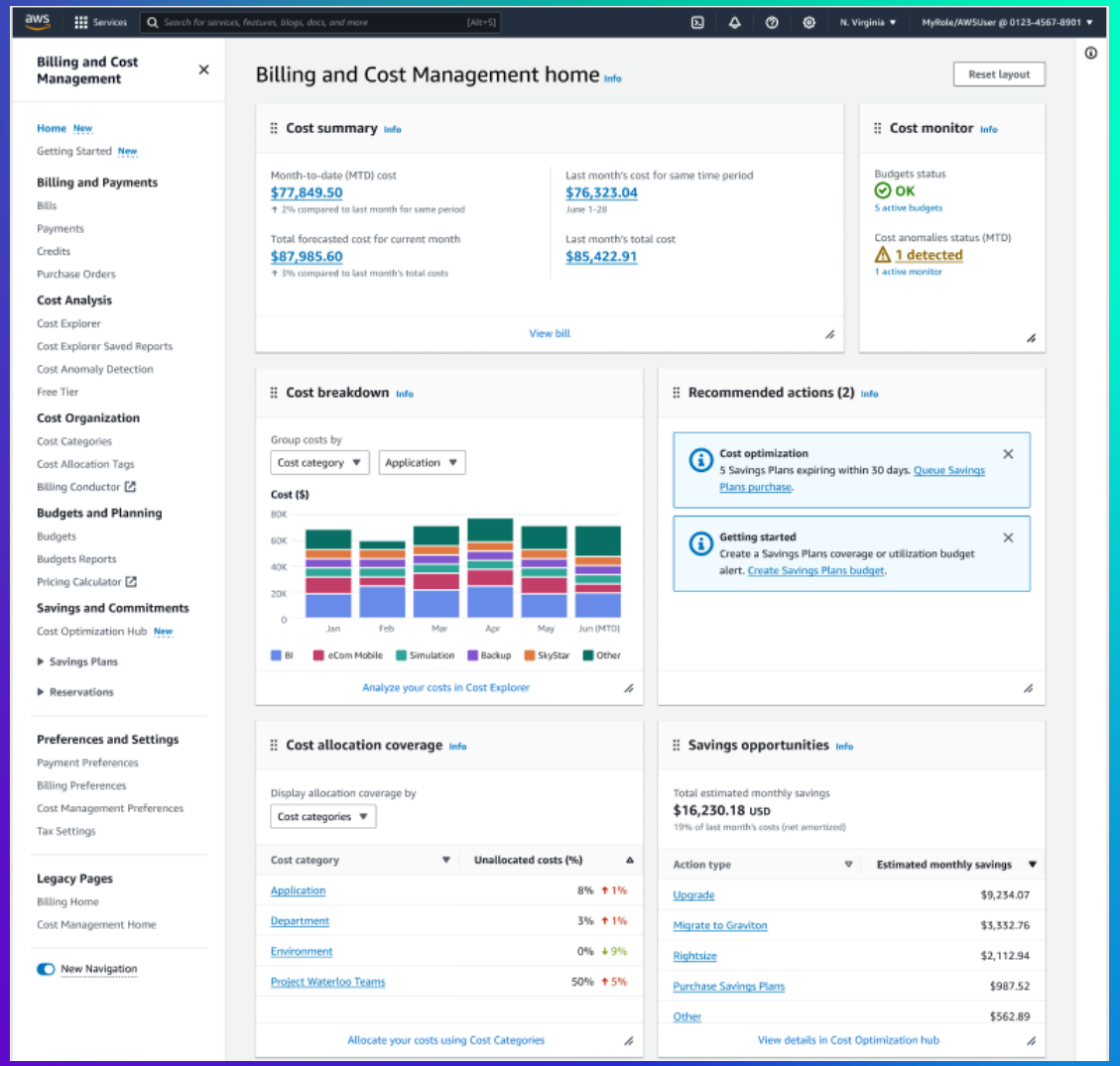
NEW

AWS Billing and Cost Management

Unified Billing and Management Console

Easy deployment from console and
Export granular cost and usage data

GENERALLY AVAILABLE

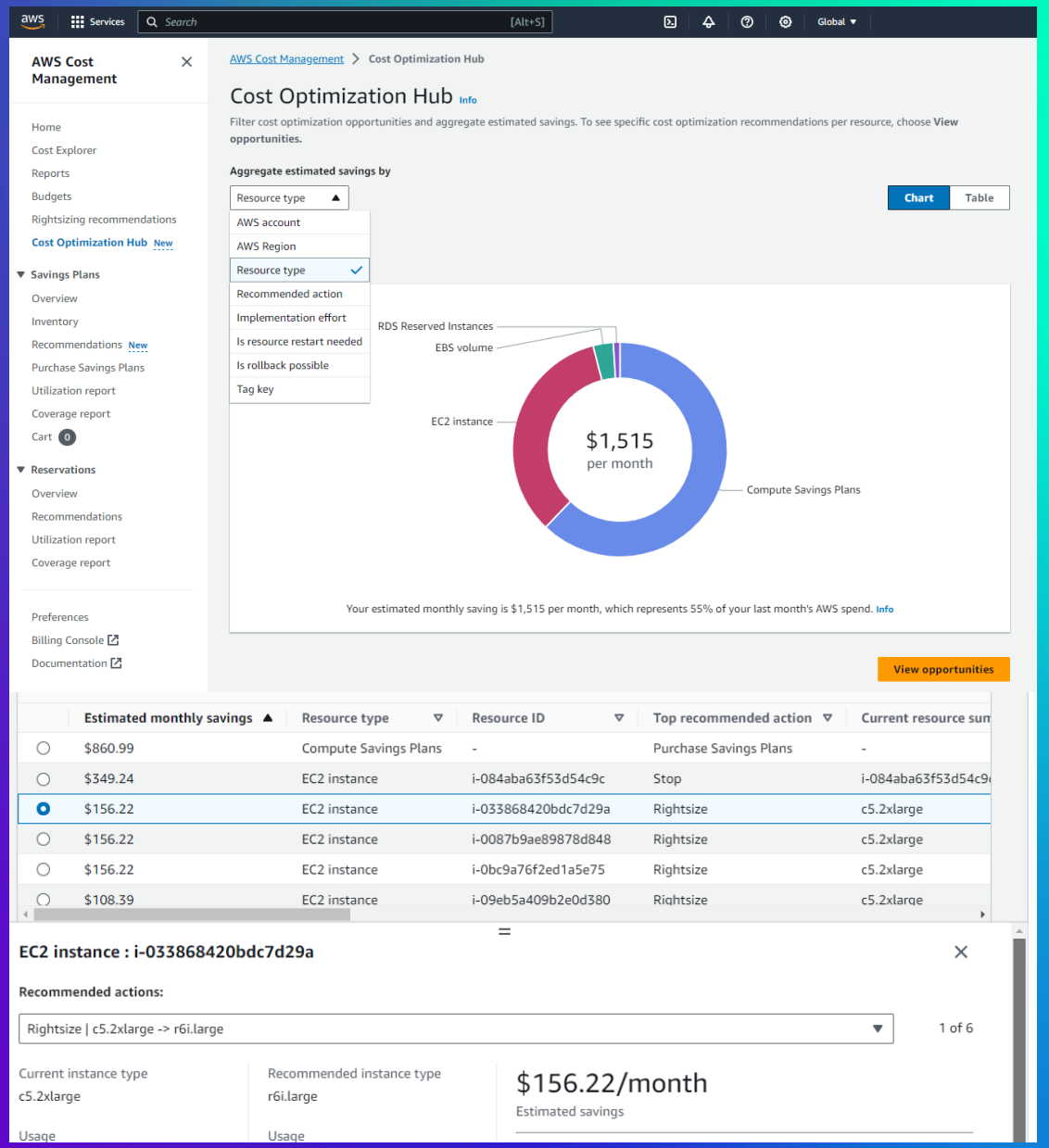


NEW

AWS **Cost Optimization Hub**

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

GENERALLY AVAILABLE



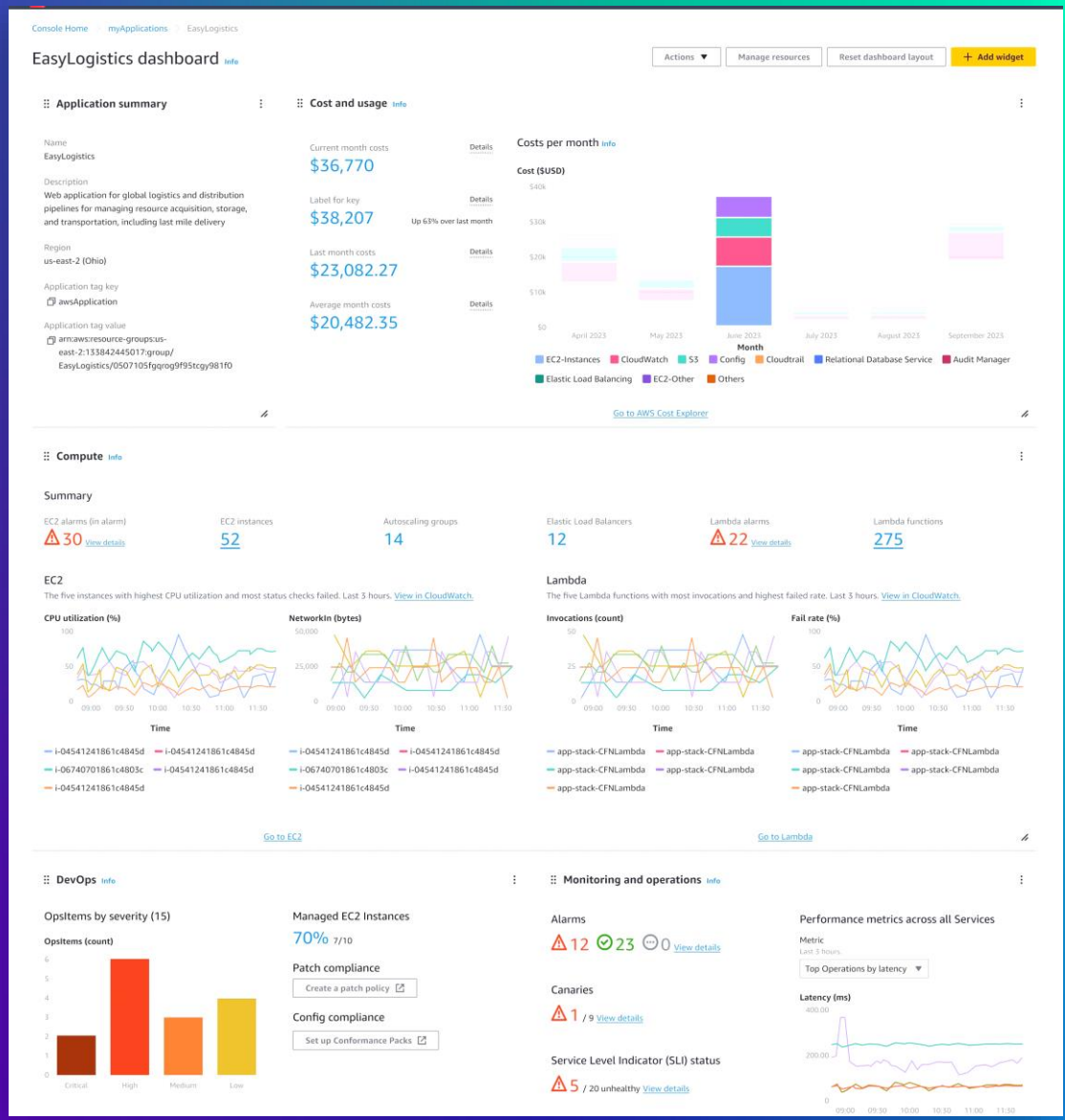
NEW

AWS Management Console

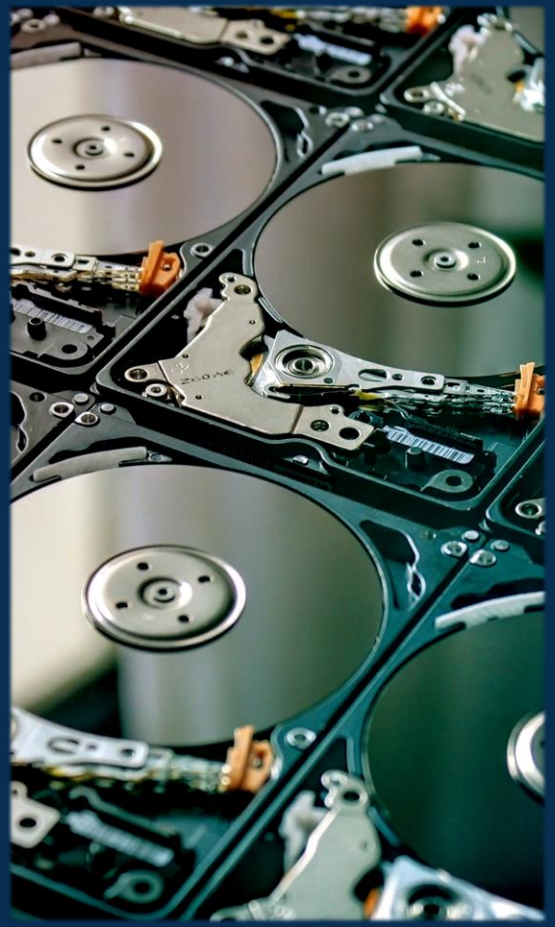
myApplications

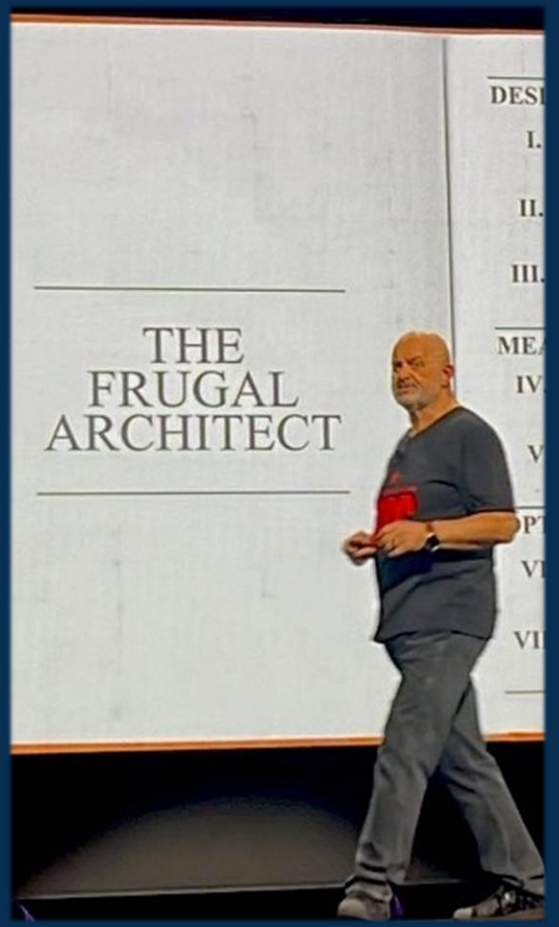
Monitor and manage the cost, health, security posture, and performance of your applications

GENERALLY AVAILABLE



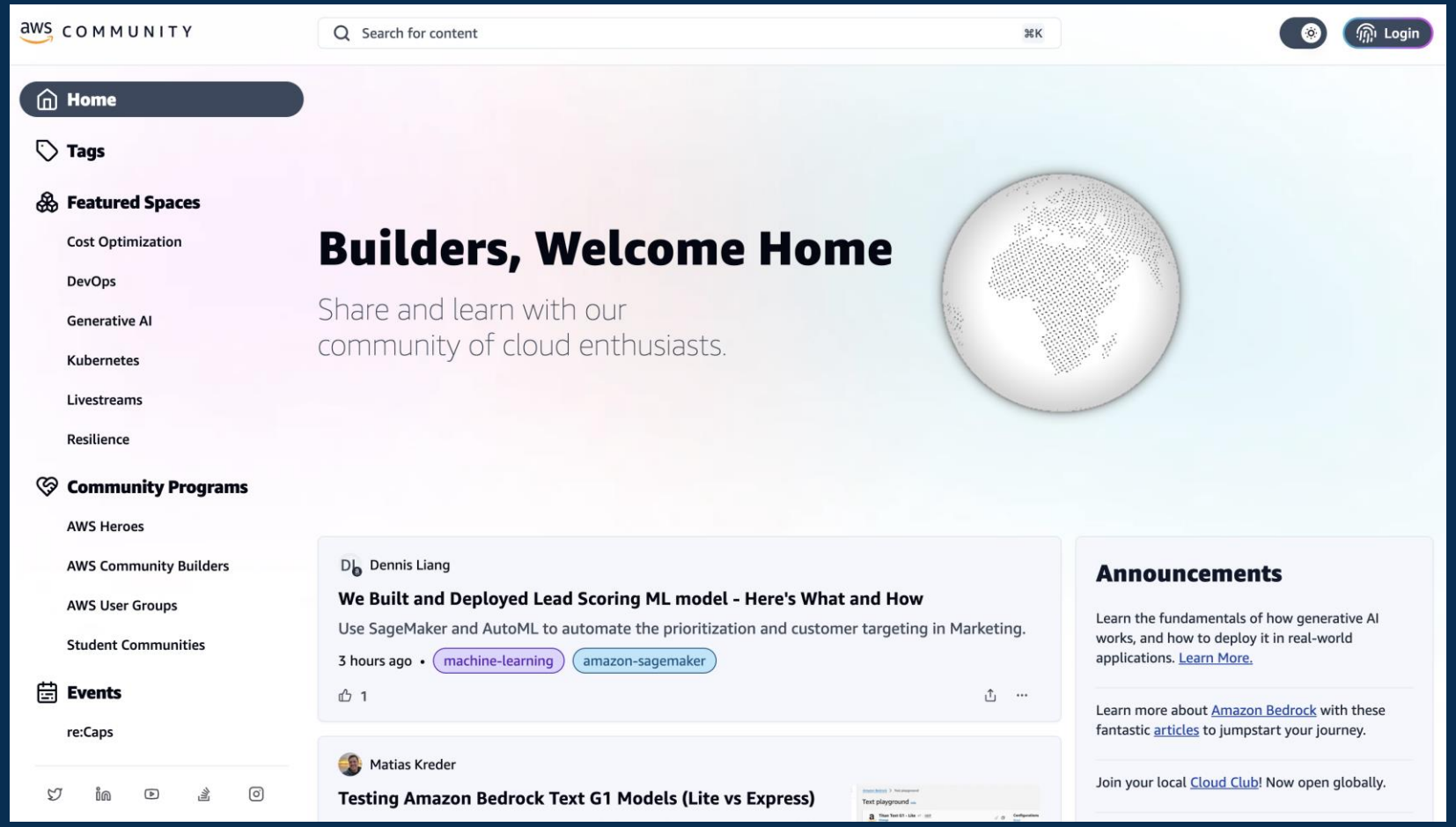
Highlights from AWS re:Invent 2023





Let’s Learn Together!

https://community.aws



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»

aws

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