

AWS re:Invent

DECEMBER 1 - 5, 2025 | LAS VEGAS, NV

Amazon EKS Auto Mode: Evolving Kubernetes ops to enable innovation

Sai Vennam

he/him

Principal Containers Specialist
Amazon Web Services

Alex Kestner

he/him

Principal Product Manager – Technical
Amazon Web Services

Daniel Levine

he/him

Senior Lead Software Engineer
Capital One

From zero to hero in 5 minutes with
EKS Auto Mode

Kubernetes and Amazon EKS

Kubernetes – the production platform



93%

Using in production or
actively evaluating

Why Kubernetes?



Simplicity

195 CNCF projects



Consistency

100s of compatible tools



Extensibility

Unlimited customization

Amazon EKS



The most trusted and secure way to run
Kubernetes



Enables you to build reliable, stable, and
secure applications in any environment



Fully upstream and certified conformant
Kubernetes

8 years of managed Kubernetes on AWS

2018

2019

2020

2021

2022

2023

2024

2025

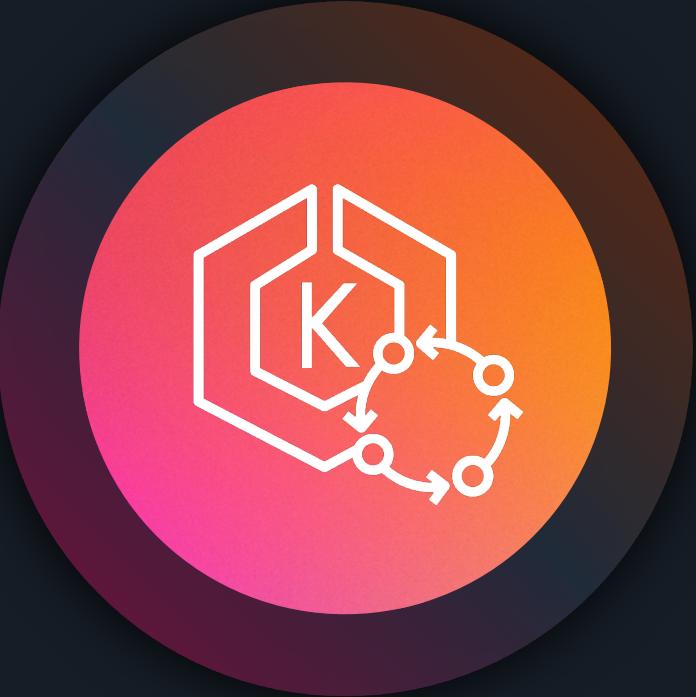
EKS Generally Available		ISO, PCI, and SOC Compliance	Price reduced to \$0.10 per hour	OIDC access authentication	IPv6 clusters available	Expansion to 32 AWS regions	Amazon Linux 2023	100,000 node clusters
Managed Cluster Version Updates		Expansion to 15 AWS regions	Secrets Encryption	Karpenter project	PrivateLink support	Kubernetes Network Policy enforcement	Automatic Version Upgrades	Provisioned Control plane
GPU Support		CSI drivers for EBS, EFS, FSx for Lustre	SLA raised to 99.95%	Cluster creation reduced by 40% control plane scaling	Local clusters on Outposts	Upgrade Insights	Zonal Shift	MCP Servers
HIPAA eligible		Pod security policies	EKS CIS Benchmark	EKS Anywhere GA	Trainium instances	Extended version support	Full IPv6 support	Managed Capabilities
		Managed Node Groups	ACK Project	EKS Connector	Add-ons from AWS marketplace	Version launch acceleration	Multi-cluster dashboard	EKS Dashboard
		EKS Fargate	Load balancer controller	FedRamp High Compliance	Nitro enclaves	Pod Identity	Auto Mode	Default K8s API encryption
		EKS on AWS Local zones	EKS on AWS Local zones		Launch time reductions	S3 Mountpoints CSI Driver	Hybrid Nodes	Auto Mode Gov Cloud
		EKS on AWS Outposts	K8s Resources in EKS console	EFA, P4d instance support			Node health & auto-repair	EKS Community Addons
		EKS Add-ons						





AMAZON EKS

Runs tens
of millions
of clusters every year



AMAZON EKS

Auto Mode

Automate your entire Kubernetes cluster infrastructure

WHAT IS AMAZON EKS **Auto Mode**



Application-ready clusters, with essential Kubernetes capabilities

Dynamic compute scaling, AWS-managed instances, secure and healthy by default

Automatic compute optimization, simplified cluster upgrades

WHY AMAZON EKS **Auto Mode**



Accelerate innovation by offloading cluster operations to AWS



Improve performance, availability, and security of your applications with AWS operational excellence



Optimize compute costs with automatic capacity planning, maintenance and dynamic scaling

Deploying and operating Kubernetes applications



Provision control plane



Install plug-ins



Select and provision compute



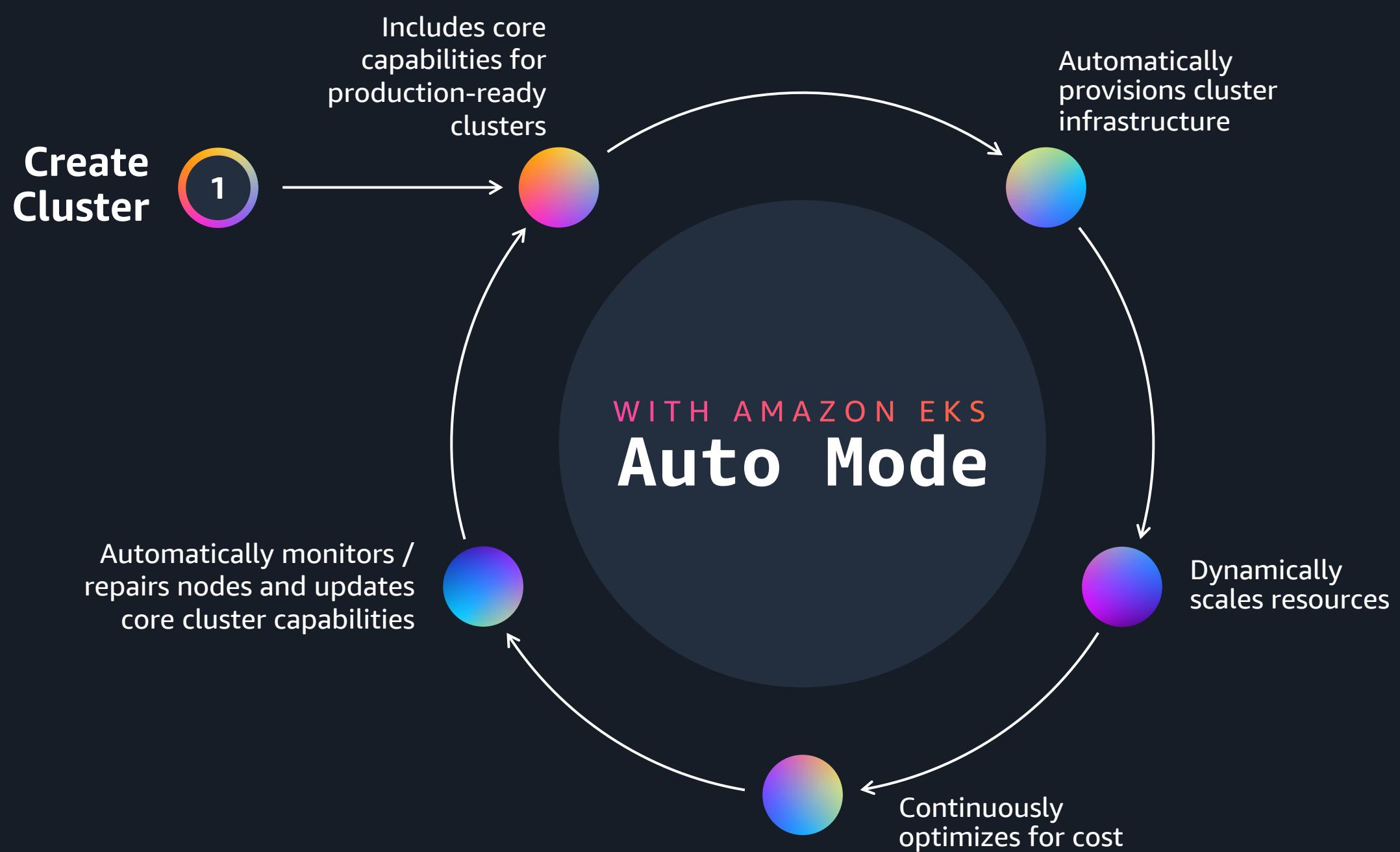
Deploy application



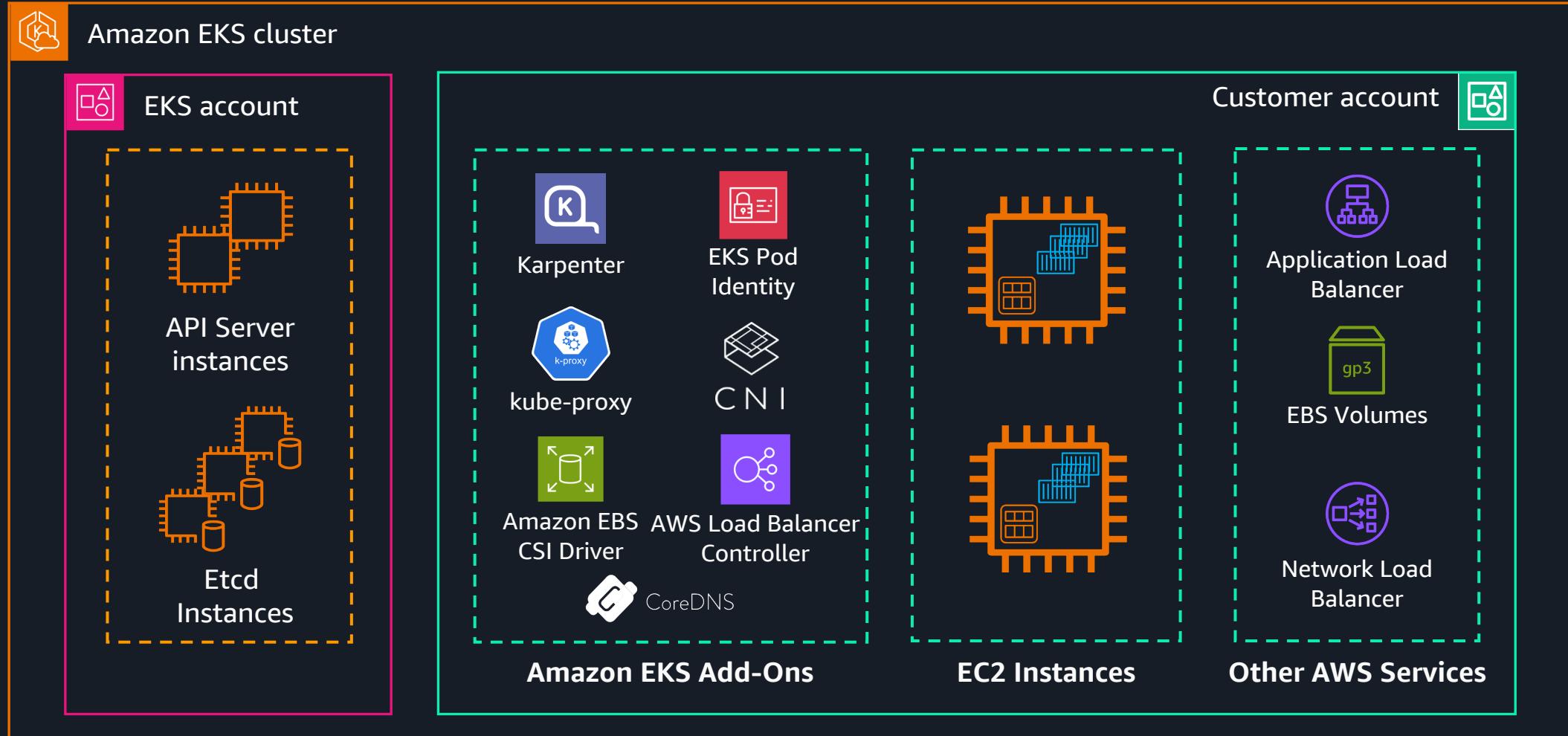
Monitor, repair, and upgrade



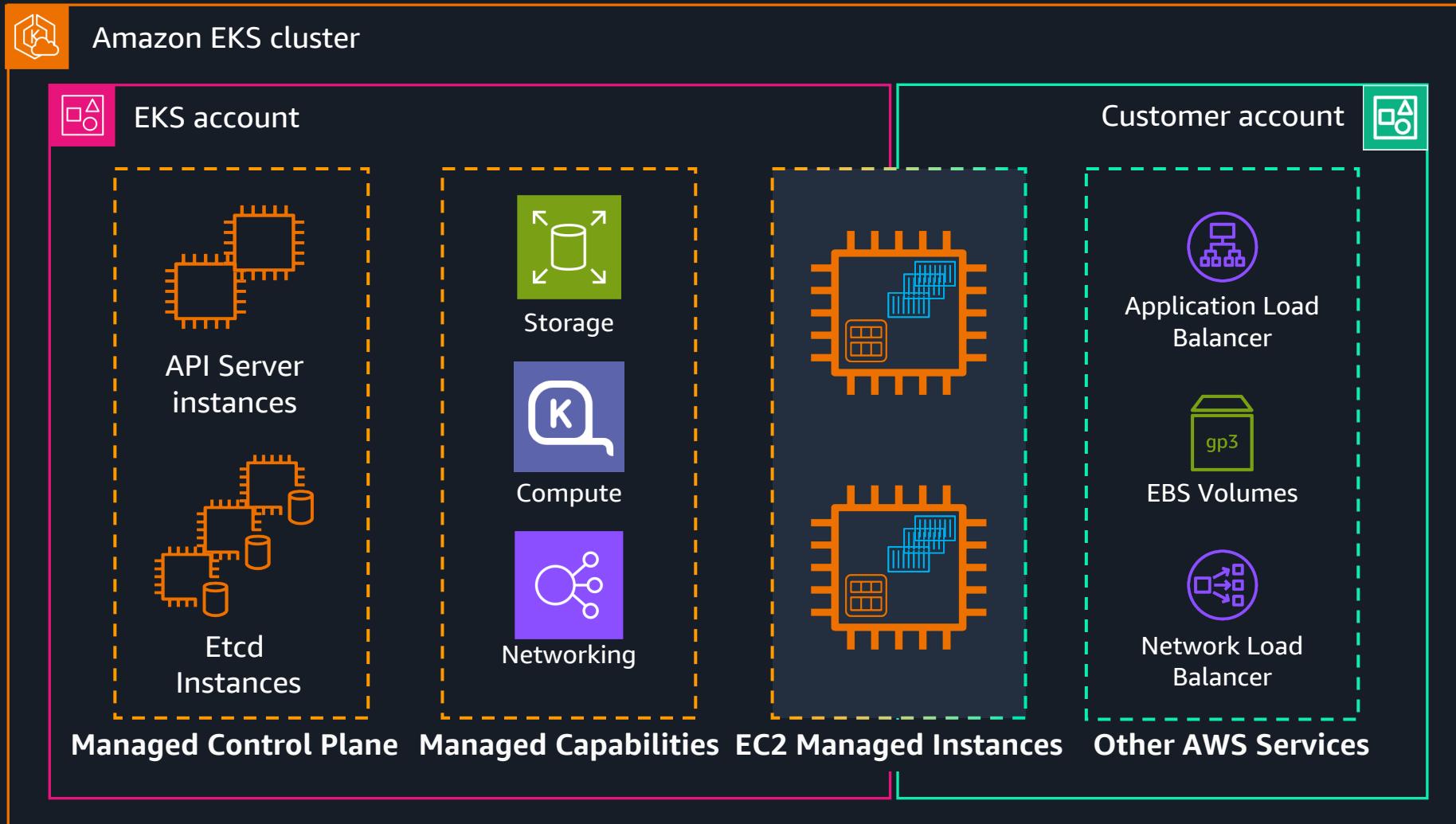
Analyze and optimize



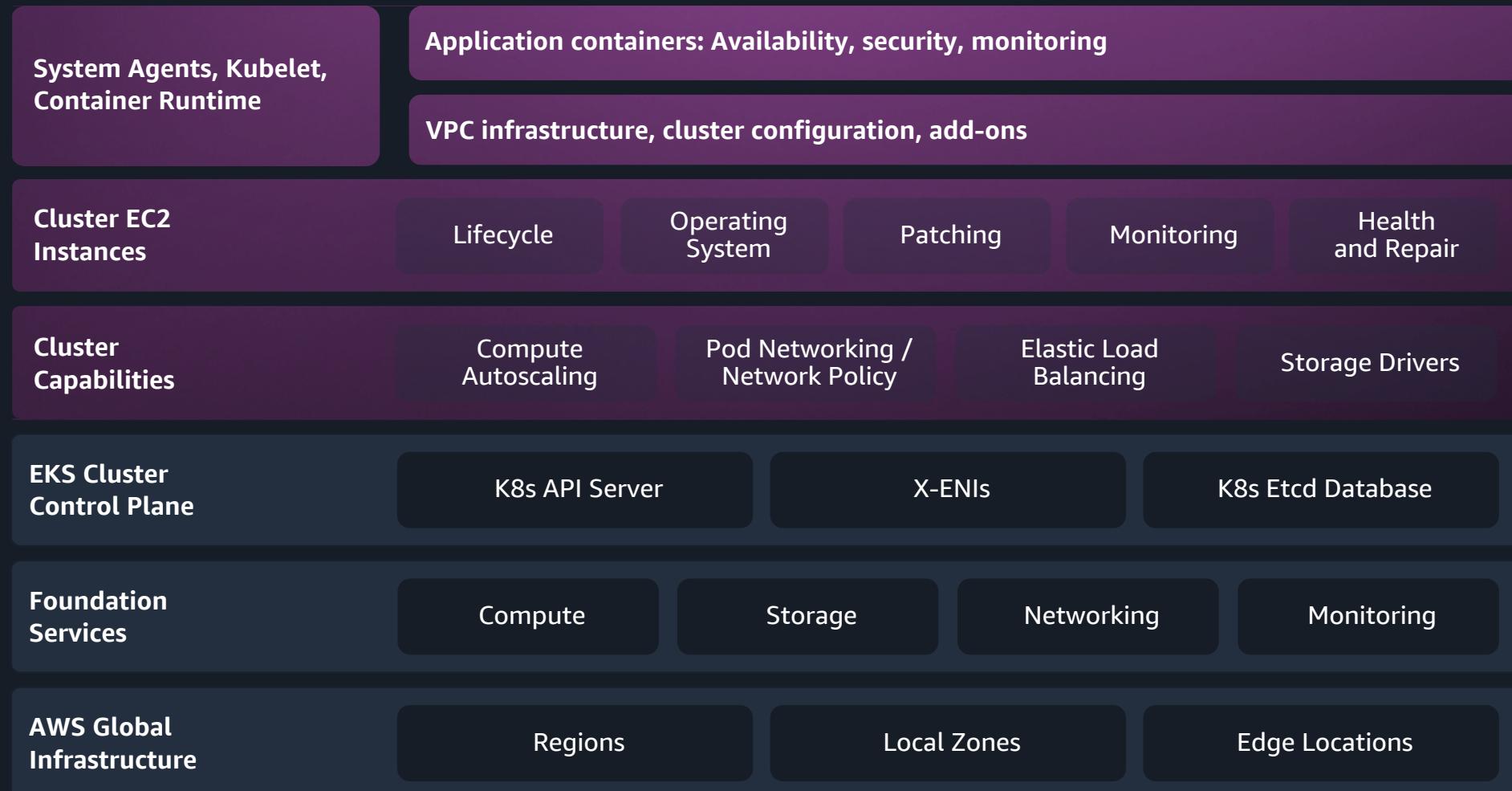
Standard Amazon EKS cluster architecture



EKS Auto Mode cluster architecture



Shared Responsibility Model with EKS previously

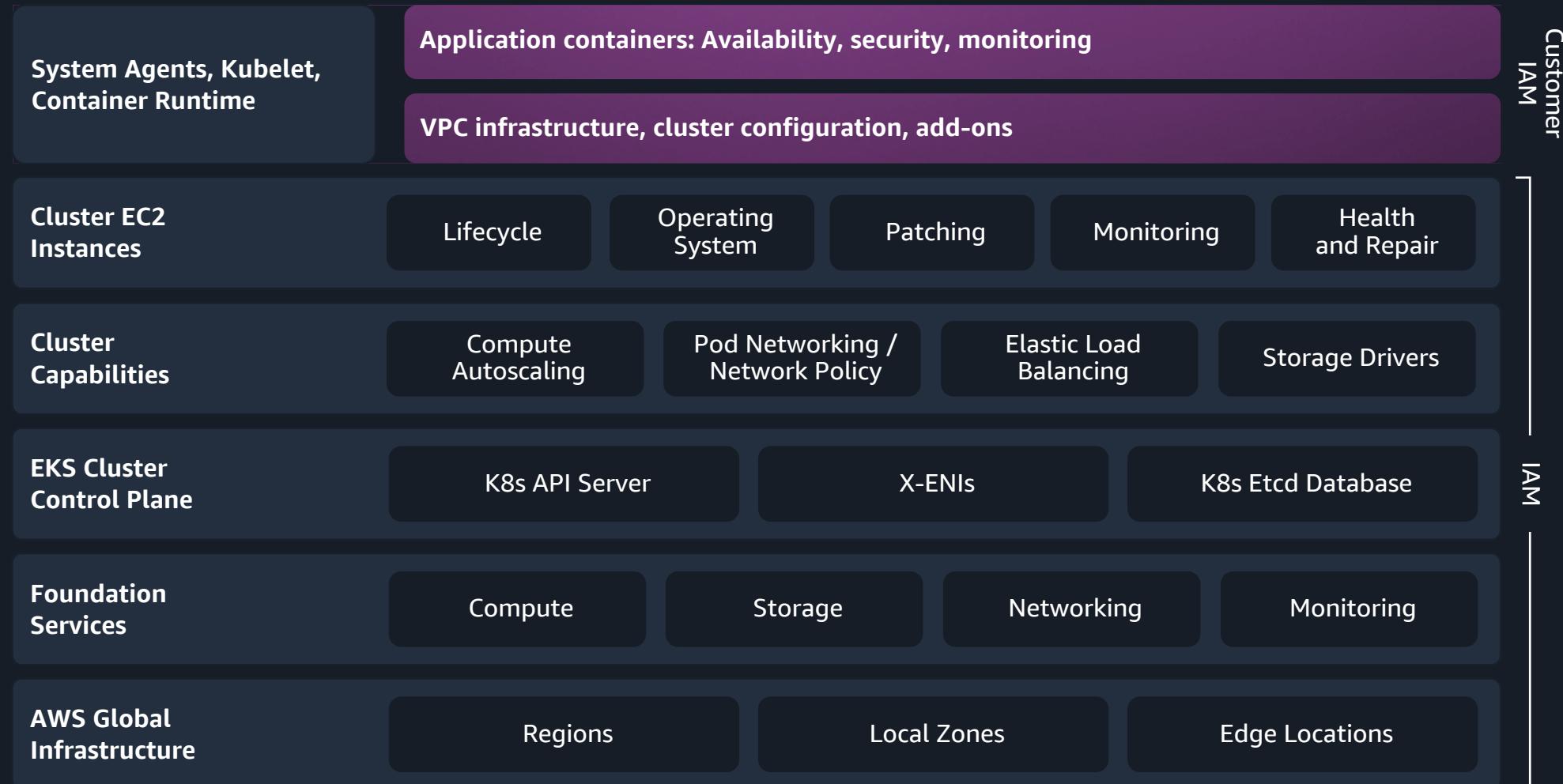


Managed by
Customers



Managed by
Amazon
Web Services

Shared Responsibility Model with EKS Auto Mode



Managed by
Customers



Managed by
Amazon
Web Services

EKS Auto Mode in action



WHO AM I?

Dan Levine



Working to enable Kubernetes & EKS in the Enterprise

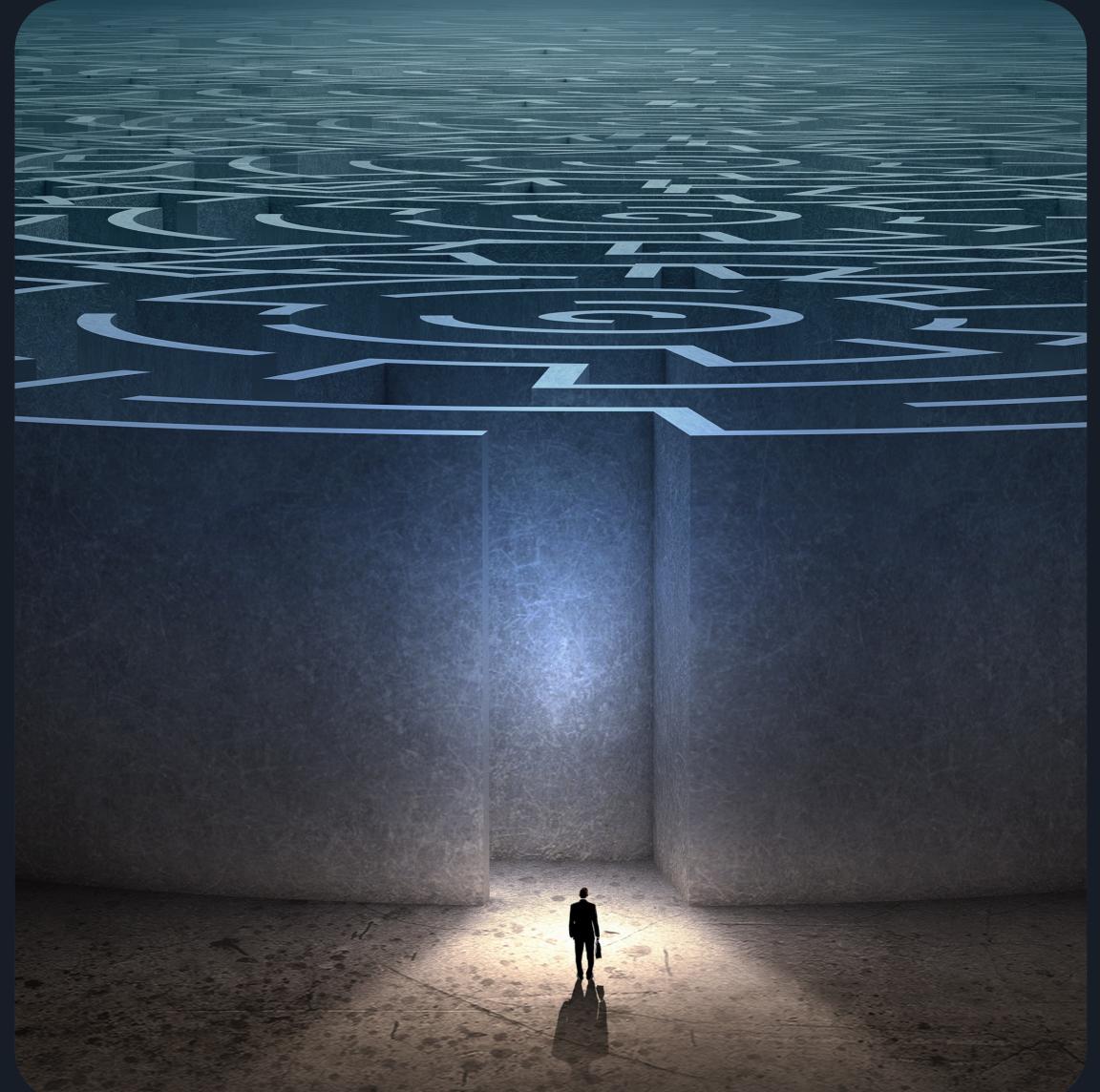
Focused on the intelligent use of compute

What am I going to talk about?

- Capital One's & Kubernetes
- Operating in a Federated Model
- Convergence on EKS Auto Mode
- Our Users

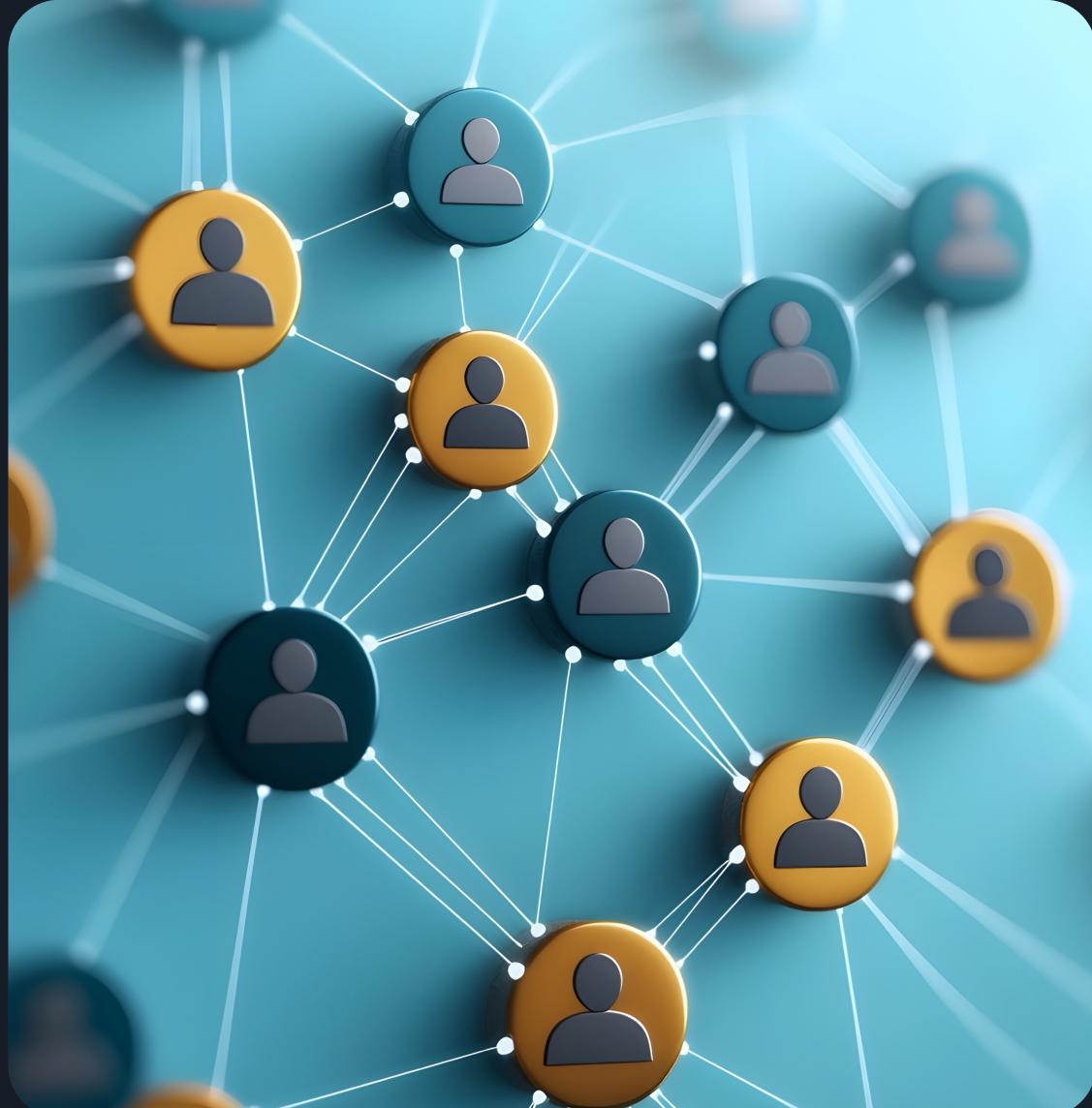
The Challenge: The Scattered Bottleneck

- **High SRE Effort:** Managing Kubernetes cluster and the underlying infrastructure required significant, specialized SRE bandwidth
- **Constant Churn:** A large Number of platform components need to be constantly updated, creating a large backlog
- **Scalability Issues:** Each team cannot effectively support the diverse needs of all the unique challenges they face, which stalls innovation



The Solution: A Federated Model

- **Core Principle:** Shift from a purely centralized "do-it-for-you" model to a "do-it-with-you" collaborative framework.
- **Empowerment:** Designed to "meet platform teams where they're at" by providing them with the tools, standards, and autonomy they need.
- **Shared Ownership:** Enables a large group of people to come together to own, maintain, and evolve different parts of the platform.



How It Works

- **What They Are:** Platform Teams and Special Interest Groups (SIGs)
- **How They Function:** Each SIG has a clear charter and ownership over a specific domain (e.g., SIG-Observability)
- They bring together experts and stakeholders to create, document, and maintain solutions for their domain
- They act as the primary point of contact and expertise for their components
- The central tooling team can continue to operate on evolution of Kubernetes



The Residual Pain

Infrastructure Management

Persistent SRE overhead for provisioning, patching, and securing the underlying Kubernetes clusters and cloud infrastructure

Container Management

Ongoing operational load from image upgrading and container dependency management

Here comes EKS Auto Mode

The next step in our intelligent
usage of highly scalable compute



EKS Auto Mode is the Next Stage In Compute Evolution

Automated Infrastructure Management

EKS Auto removes the operational burden of provisioning, scaling, and patching the cluster's compute

Simplified Container Management

EKS Auto patches and manages the underlying containers that run common business functions

Happy Platform Teams!



EKS Auto Mode 2025 features



FULLY-MANAGED EKS CLUSTERS

EKS Auto Mode 2025 Features

- US GovCloud and Local Zones support
- Advanced configuration
- SOCI parallel pull & unpack
- Security and compliance features
- Static and reserved capacity provisioning



GovCloud (US) and Local Zones

Regional availability

EKS Auto Mode is available in all commercial regions where EKS is available, *except for AWS (China) Beijing and Ningxia regions*

In 2025, we expanded this to include the **AWS GovCloud (US) regions and AWS Local Zones**

Advanced Configuration

Flexibility without operational overhead

Use separate subnets and security groups for pods with **podSubnetSelectorTerms** and **podSecurityGroupSelectorTerms**

Advanced networking configurations like **associatePublicIpAddress**, and **httpsProxy / noProxy**

Provide private certificate material to Auto Mode-managed EC2 instances with **certificateBundles**

EKS AUTO MODE 2025 FEATURES

SOCI Parallel Image Pull & Unpack

Accelerate AI/ML workload startup

SOCI parallel container pull & unpack reduces Time-to-First-Token for AI inference workloads

Automatically enabled for GPU instances in EKS Auto Mode - G, P, Trn and other EC2 instances with local NVME storage

Zero configuration required optimized settings applied automatically

EKS AUTO MODE 2025 FEATURES

Security and Compliance Features

Delegate security and compliance to AWS

Encrypt both root and data storage volumes with `ephemeralStorage.kmsKeyID`

Enable FIPS-validated cryptographic modules in both EKS Auto Mode controllers and instances `advancedSecurity.fips`

Enable teams on Auto Mode without granting additional IAM permissions, just an `instanceProfile`



Static and Reserved Capacity

AWS-Managed EC2 Reserved Capacity

Auto Mode can prioritize reserved capacity, increasing cost-efficiency using **capacityReservationSelectorTerms**.

Supports both On Demand Capacity Reservations and Capacity Blocks for ML

For workloads that don't need Auto Mode's default dynamic and flexible compute, use Static Capacity to pre-provision capacity independently using **spec.replicas**

Wrapping up



EKS Auto Mode

Key Takeaways

Operational Transformation

Complex manual management to automated AWS-managed infrastructure

Kubernetes Power + Simplicity

Full ecosystem access, less operational overhead

Focus on Innovation

Free teams to build business value, not manage infrastructure

Enterprise Ready

Security, compliance, and performance without complexity



Actions and resources



EKS Best Practices Guide



Getting Started with EKS



Hands-on Workshop Series





Thank You

Please complete the session
survey in the mobile app