

Service Name

Amazon Elastic Kubernetes Service (EKS)

Service Overview

Amazon Elastic Kubernetes Service (EKS) is a fully managed Kubernetes service that simplifies running Kubernetes on AWS without the need to install, operate, or maintain your own Kubernetes control plane or nodes. EKS automatically manages the availability and scalability of the Kubernetes control plane nodes responsible for scheduling containers, managing application availability, storing cluster data, and other key tasks. The service provisions and scales the control plane across multiple AWS Availability Zones for high availability and fault tolerance, automatically detecting and replacing unhealthy control plane nodes while providing automatic version updates and patching.

Key Use Cases

- **Microservices architecture:** Building and deploying cloud-native applications using microservices patterns with Kubernetes service discovery
- **Machine learning workloads:** Running ML training and inference workloads with popular frameworks and GPU support on scalable infrastructure
- **Batch processing:** Executing cost-effective batch processing and big data workloads using EC2 Spot Instances for compute optimization
- **Hybrid deployments:** Deploying applications across cloud and on-premises environments using EKS Hybrid Nodes and AWS Outposts
- **High-availability applications:** Building resilient applications with Elastic Load Balancing integration and multi-AZ deployment capabilities
- **Platform engineering:** Creating custom Kubernetes APIs and abstractions for standardized application deployment patterns

Features

- **Managed Control Plane:** Fully managed Kubernetes control plane with automatic scaling, patching, and high availability across multiple AZs
- **Amazon EKS Auto Mode:** Serverless compute for containers that automatically manages both control plane and worker nodes with automated patching and scaling
- **EKS Capabilities:** Fully managed cluster features including Argo CD for GitOps, AWS Controllers for Kubernetes (ACK), and Kube Resource Orchestrator (KRO)
- **Multiple Compute Options:** Support for Amazon EC2, AWS Fargate, Nitro and Graviton instances, EC2 Spot Instances, and managed node groups
- **Advanced Networking:** VPC native networking, IPv6 support, load balancing integration, and application networking capabilities

- **Security and Compliance:** EKS Pod Identity, IAM for RBAC, encryption at rest and in transit, compliance certifications, and container image signature verification
- **Hybrid Deployment Support:** Amazon EKS Anywhere, EKS Hybrid Nodes, AWS Outposts integration, and EKS Connector for external cluster management
- **Observability and Monitoring:** Integration with Amazon Managed Service for Prometheus, CloudWatch Container Insights, logging, and cost allocation tagging

Value Proposition

Amazon EKS provides significant operational advantages over self-managed Kubernetes by eliminating the complexity of control plane management while maintaining full Kubernetes compatibility. Organizations benefit from reduced operational overhead as AWS handles control plane provisioning, scaling, patching, and high availability across multiple Availability Zones:

- The service offers enterprise-grade security with deep AWS service integrations for IAM, VPC, and compliance frameworks
- EKS supports diverse deployment patterns from cloud-native to hybrid environments with EKS Anywhere and Hybrid Nodes
- The fully managed approach allows teams to focus on application development rather than infrastructure management while maintaining the flexibility to use existing Kubernetes tools and applications

Service Integration

Amazon EKS integrates deeply with core AWS services to provide a comprehensive container platform. Key integrations include Amazon EC2 and AWS Fargate for compute, Amazon EBS and Amazon EFS for persistent storage, and Amazon ECR for container image management:

- Security integrations span AWS IAM for authentication and authorization, Amazon VPC for network isolation, and AWS CloudTrail for audit logging
- Monitoring and observability are enhanced through Amazon CloudWatch, Amazon Managed Service for Prometheus, and AWS X-Ray
- Load balancing integrates with Elastic Load Balancing (ALB, NLB) while Amazon GuardDuty provides threat detection
- Additional integrations include AWS Controllers for Kubernetes (ACK) for managing AWS resources, AWS Backup for cluster protection, and AWS App Mesh for service mesh capabilities

Onboarding and Offboarding

Customers can get started with Amazon EKS using multiple approaches. The quickest path involves using `eksctl`, AWS's command-line utility that creates clusters with a single command, automatically configuring necessary IAM roles and VPC settings:

- Alternatively, customers can use the AWS Management Console, AWS CLI, or infrastructure-as-code tools like CloudFormation and Terraform
- EKS Auto Mode simplifies the experience further by automatically managing both control plane and worker nodes
- Prerequisites include setting up AWS CLI, kubectl, and appropriate IAM permissions
- AWS provides comprehensive getting started guides, workshops, and sample applications to accelerate adoption
- The service supports both Fargate for serverless containers and EC2-based managed node groups for traditional deployments

Getting Started Guide: <https://docs.aws.amazon.com/eks/latest/userguide/getting-started.html>

Data Removal

To offboard from Amazon EKS, customers must first delete all workloads and persistent volumes to ensure data removal, then delete node groups and the EKS cluster itself. The process involves removing Kubernetes resources, deleting associated AWS resources like Load Balancers and EBS volumes, and finally terminating the cluster through the AWS console, CLI, or eksctl. Data in persistent volumes attached via EBS, EFS, or S3 must be explicitly deleted as these resources persist beyond cluster deletion. AWS provides detailed cleanup procedures to prevent orphaned resources and unexpected charges.

Backup/Restore and Disaster Recovery

Amazon EKS integrates with AWS Backup to provide comprehensive cluster protection including both Kubernetes cluster state and persistent storage. Backups capture Kubernetes manifests, secrets, config maps, and persistent volumes from EBS, EFS, and S3. The service creates composite recovery points enabling full cluster restoration or selective namespace recovery. EKS supports automated backup scheduling, cross-region and cross-account copies, and immutable backup vaults for compliance requirements.

Backup Capabilities

EKS provides native integration with AWS Backup for automated, centralized backup management. The service supports both on-demand and scheduled backups of cluster state and persistent data without requiring agents or add-ons:

- Backup capabilities include incremental backups, lifecycle management, encryption, and cross-region replication
- AWS Backup creates composite recovery points allowing granular restoration options from full clusters to individual persistent volumes

Disaster Recovery

Amazon EKS supports disaster recovery through multiple mechanisms including AWS Backup integration, multi-region deployment capabilities, and AWS Elastic Disaster Recovery integration. The service enables pilot light, warm standby, and active-active disaster recovery strategies. EKS clusters can be restored across regions and accounts, while persistent data protection spans EBS snapshots, EFS replication, and S3 cross-region replication for comprehensive DR coverage.

Recovery Objectives

EKS helps customers meet RPO and RTO objectives through automated backup scheduling aligned with recovery point requirements, cross-region cluster and data replication for faster recovery, and multi-AZ control plane architecture for high availability. The service supports various DR strategies from backup/restore (higher RPO/RTO, lower cost) to active-active deployments (minimal RPO/RTO, higher cost). Integration with AWS Backup enables policy-driven protection with configurable retention periods.

Service Constraints

Service Quotas: <https://docs.aws.amazon.com/eks/latest/userguide/service-quotas.html>

FAQ: <https://aws.amazon.com/eks/faqs/>

Technical Requirements

Service Documentation: <https://docs.aws.amazon.com/eks/latest/userguide/index.html>

User Guide: <https://docs.aws.amazon.com/eks/latest/userguide/index.html>

API Reference: <https://docs.aws.amazon.com/eks/latest/APIReference/Welcome.html>

Pricing Overview

Please see the AWS UK G Cloud 15 Pricing Document accompanying this service in the Digital Marketplace.

Public Pricing: <https://aws.amazon.com/eks/pricing/>

Cost Optimization

Amazon EKS offers multiple cost optimization strategies including EKS Auto Mode for serverless compute with automatic rightsizing, EC2 Spot Instances integration for batch workloads with up to 90% cost savings, and Graviton-based instances for better price-performance. The service supports Compute Savings Plans and Reserved Instances for predictable workloads:

- Cost allocation tagging and integration with AWS Cost Explorer and third-party tools like Kubecost enable granular cost tracking and optimization
- Cluster Autoscaler and Vertical Pod Autoscaler help optimize resource utilization, while topology-aware routing reduces cross-AZ data transfer costs

Savings Considerations

Key cost savings opportunities include using mixed instance types and Spot Instances for non-critical workloads, implementing proper resource requests and limits to avoid over-provisioning, and leveraging EKS Auto Mode to eliminate idle node costs. Customers should optimize networking costs through VPC endpoints and topology-aware routing, use ECR image replication for reduced data transfer, and implement cluster consolidation strategies. Cost monitoring through Kubecost integration and AWS native tools helps identify waste, while appropriate node group sizing and utilizing cheaper storage classes for persistent volumes can significantly reduce operational expenses.