Service Name

AWS Compute Optimizer

Service Overview

AWS Compute Optimizer is a machine learning-powered service that analyzes the configuration and utilization metrics of AWS resources to provide optimization recommendations. It evaluates Amazon EC2 instances, Auto Scaling groups, EBS volumes, AWS Lambda functions, Amazon ECS services on Fargate, commercial software licenses, and Amazon Aurora/RDS databases to identify right-sizing opportunities, idle resources, and cost optimization potential. The service analyzes historical utilization patterns and provides actionable recommendations to reduce costs and improve performance by suggesting optimal resource configurations based on actual usage data.

Key Use Cases

- Cost optimization: Identify over-provisioned and idle resources to reduce infrastructure costs by rightsizing or eliminating unnecessary resources
- Performance optimization: Recommend instance types and configurations that better match workload requirements to improve application performance
- Resource rightsizing: Analyze utilization patterns across EC2 instances, Auto Scaling groups, Lambda functions, EBS volumes, and databases to suggest optimal sizing
- License optimization: Optimize commercial software licenses like Microsoft SQL Server on EC2 to reduce licensing costs
- Multi-account optimization: Centrally manage optimization recommendations across AWS Organizations with delegated administration capabilities

Features

- **Resource Analysis**: Analyzes EC2 instances, Auto Scaling groups, EBS volumes, Lambda functions, ECS services, and RDS/Aurora databases using machine learning and utilization metrics
- **Automation Rules**: Creates automation rules to automatically implement optimization recommendations based on criteria like AWS Region and resource tags on recurring schedules
- **Enhanced Infrastructure Metrics**: Extends analysis lookback period from 14 days to 93 days for more accurate quarterly utilization pattern detection
- **Idle Resource Detection**: Identifies unused or underutilized resources like unattached EBS volumes and idle EC2 instances for cleanup
- **Performance Risk Assessment**: Evaluates performance risk of recommended changes with ratings from very low to very high based on utilization patterns

- **Savings Estimation**: Calculates potential cost savings and considers pricing discounts when generating recommendations for cost optimization
- **External Metrics Integration**: Ingests metrics from external observability products to enhance recommendation accuracy
- **Organization Integration**: Integrates with AWS Organizations for centralized management and delegated administration across multiple accounts

Value Proposition

AWS Compute Optimizer provides machine learning-powered optimization recommendations that deliver measurable cost savings and performance improvements without requiring extensive manual analysis. Unlike basic monitoring tools, it offers automated rightsizing recommendations with performance risk assessment, ensuring changes won't negatively impact applications:

- The service provides automation capabilities to implement recommendations at scale across organizations, extended analytics with 93-day lookback periods for seasonal patterns, and integration with existing AWS services
- It's particularly valuable because it considers actual utilization patterns rather than theoretical capacity, offers reversible automated actions for operational safety, and provides detailed savings estimates with support for custom pricing discounts

Service Integration

AWS Compute Optimizer integrates deeply with Amazon CloudWatch for metrics collection and analysis, AWS Organizations for multi-account management, and AWS Cost Explorer for unified cost optimization views. It works with Amazon EC2, Auto Scaling groups, Lambda, ECS on Fargate, EBS, Aurora, and RDS for resource optimization:

- The service integrates with AWS Trusted Advisor for complementary recommendations, AWS Systems Manager for automation workflows, and AWS Application Discovery Service for workload assessment
- It also connects with AWS Backup for data protection during optimization, AWS Resource Groups for organized management, and external observability platforms for enhanced metrics ingestion

Onboarding and Offboarding

Customers get started by opting into AWS Compute Optimizer through the console, CLI, or API, which automatically creates a service-linked role for data access. The service supports standalone accounts, organization member accounts, and management accounts with options to include all organization members:

- After opting in, Compute Optimizer begins analyzing resources with sufficient metrics data (30 hours for EC2/EBS, 50 invocations for Lambda)

- Enhanced infrastructure metrics can be activated for deeper 93-day analysis
- Organizations can designate delegated administrators and configure automation rules for scaled optimization
- The service requires appropriate IAM permissions and supports trusted access for organization-wide management

Getting Started Guide: https://docs.aws.amazon.com/compute-optimizer/latest/ug/getting-started.html

Data Removal

Customers can opt out of AWS Compute Optimizer through the console, CLI, or API, which stops new analysis and recommendation generation. To remove historical data, customers must contact AWS Support to request deletion of stored CloudWatch metrics and configuration data used for recommendations. The service automatically stops accessing resources upon opt-out, and service-linked roles can be deleted if no longer needed.

Backup/Restore and Disaster Recovery

AWS Compute Optimizer does not provide traditional backup and disaster recovery capabilities as it is an optimization service rather than a data storage service. The service itself is highly available across AWS regions and maintains recommendation data, but does not back up customer resources. However, automation features include safety mechanisms like creating EBS snapshots before volume modifications to enable data recovery if needed.

Backup Capabilities

AWS Compute Optimizer does not have inherent backup capabilities and does not integrate with AWS Backup as it is an analysis and recommendation service rather than a data service. The service focuses on optimization recommendations rather than data protection, though automation features include safety mechanisms like snapshot creation before resource modifications.

Disaster Recovery

AWS Compute Optimizer does not directly support disaster recovery scenarios and does not integrate with AWS Elastic Disaster Recovery. As an optimization service, it focuses on cost and performance recommendations rather than business continuity. The service itself is regionally distributed for high availability but does not provide disaster recovery capabilities for customer workloads.

Recovery Objectives

AWS Compute Optimizer does not directly help customers meet RPO and RTO objectives as it is not a backup or disaster recovery service. However, it can indirectly support resilience by recommending properly sized resources for better performance and

identifying idle resources that could be repurposed for disaster recovery scenarios in multi-region architectures.

Service Constraints

Service Quotas: https://docs.aws.amazon.com/general/latest/gr/compute-optimizer.html

FAQ: https://aws.amazon.com/compute-optimizer/faqs/

Technical Requirements

Service Documentation: https://docs.aws.amazon.com/compute-optimizer/latest/ug/what-is-compute-optimizer.html

User Guide: https://docs.aws.amazon.com/compute-optimizer/latest/ug/getting-started.html

API Reference: https://docs.aws.amazon.com/compute-optimizer/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/compute-optimizer/pricing/

Cost Optimization

AWS Compute Optimizer offers unique cost optimization through machine learning-driven rightsizing recommendations that consider actual utilization patterns rather than theoretical capacity. The service provides automated cleanup of idle resources like unattached EBS volumes, license optimization for commercial software, and personalized savings estimation that accounts for customer-specific pricing discounts and Reserved Instances. Enhanced infrastructure metrics enable quarterly pattern analysis for seasonal workloads, while automation rules allow scaled optimization across entire organizations with reversible actions for operational safety and confidence in implementation.

Savings Considerations

Primary cost savings come from rightsizing over-provisioned resources, eliminating idle resources, and optimizing commercial software licenses. The service identifies instances that can be downsized while maintaining performance, recommends modern instance types with better price-performance ratios, and suggests migration to AWS Graviton instances for additional savings:

- Enhanced metrics capture seasonal patterns to prevent under-provisioning during peak periods

- Automation features enable continuous optimization at scale, while performance risk assessment ensures cost reductions don’t compromise application performance, maximizing sustainable long-term savings opportunities