Service Name

Amazon OpenSearch Service

Service Overview

Amazon OpenSearch Service is a fully managed service that simplifies the deployment, operation, and scaling of OpenSearch clusters in the AWS Cloud. OpenSearch is a fully open-source search and analytics engine used for log analytics, real-time application monitoring, and clickstream analysis. Amazon OpenSearch Service provisions all resources for an OpenSearch cluster and manages its infrastructure, including automatic node replacement and scaling. The service offers built-in integrations with OpenSearch Dashboards, Logstash, and AWS services including Amazon Data Firehose, AWS Lambda, and Amazon CloudWatch, making it ideal for querying and searching large amounts of data such as analyzing activity logs, CloudWatch Logs, product usage data, social media sentiments, and data stream updates from other AWS services.

Key Use Cases

- Log analytics: Processing and analyzing machine-generated time series data to gain insights into operations, security, and user behavior from application logs, system telemetry, and operational data
- Real-time application monitoring: Monitoring application performance, identifying and troubleshooting errors, tracking system health, and analyzing user behavior patterns in real-time
- Full-text search: Building sophisticated search applications with capabilities like prefix match, multi-match, wildcard match, and range filtering for e-commerce, content management, and document search
- Security analytics: Analyzing security logs, detecting anomalies, monitoring threats, and performing security investigations with built-in security rules and custom detection capabilities
- Business intelligence and analytics: Creating interactive dashboards, performing data visualization, analyzing clickstream data, and deriving business insights from large datasets
- Observability and monitoring: Handling time-series data from logs, metrics, and traces for DevOps teams to identify performance bottlenecks and system anomalies

Features

- **Serverless Deployment**: On-demand, auto-scaling configuration that automatically scales compute resources based on application needs with cost-effective pricing model

- **Multi-tier Storage**: UltraWarm for cost-effective storage of infrequently accessed data using S3 and caching, and cold storage for rarely accessed data with significant cost savings
- **Fine-grained Access Control**: Role-based access control, security at the index, document, and field level, multi-tenancy support, and HTTP basic authentication
- **Vector Search Engine**: Rich lexical and vector search capabilities for AI applications with GPU-accelerated and auto-optimized vector indexes supporting billion-scale vector databases
- **Multi-AZ with Standby**: Enhanced availability and performance configuration ensuring resilience to single AZ failures with 99.99% availability SLA
- **Index State Management**: Automated lifecycle management for indexes including transitions between hot, warm, and cold storage tiers based on customizable policies
- **OpenSearch Dashboards**: Native visualization tool for creating interactive dashboards, exploring data, performing analytics, and building custom visualizations without additional fees
- **Cross-cluster Operations**: Cross-cluster search and replication capabilities for disaster recovery, reduced latency, and querying data across multiple domains
- **Automated Snapshots**: Hourly automated snapshots stored in S3 for data durability and manual snapshot capabilities for cross-region disaster recovery
- **Agentic Search**: Intelligent agent-driven system that understands user intent, orchestrates tools, generates OpenSearch DSL queries, and provides natural language interactions

Value Proposition

Amazon OpenSearch Service provides a fully managed infrastructure that eliminates the complexity and operational overhead of managing OpenSearch clusters while delivering enterprise-grade capabilities. Unlike self-managed solutions, it offers automatic hardware provisioning, software installation and patching, 24x7 monitoring and repairs, zero downtime updates, and seamless scaling up to 3 petabytes of data:

- The service provides significant cost advantages through tiered storage options like UltraWarm (up to 40% cost savings) and cold storage, AWS Graviton2 instances offering up to 44% price/performance improvements, and Reserved Instances with up to 52% discounts
- Enterprise-grade security is built-in with encryption at rest and in transit, VPC integration, fine-grained access control, and compliance with standards including HIPAA, SOC 2, PCI DSS, and FedRAMP
- The service offers unique capabilities like serverless deployment with automatic scaling, vector search for AI applications, Multi-AZ with Standby for 99.99% availability, and native integrations with the broader AWS ecosystem including CloudWatch, Lambda, S3, and Kinesis

Service Integration

Amazon OpenSearch Service integrates deeply with core AWS services to provide a comprehensive analytics and search platform. Data ingestion is supported through Amazon Kinesis Data Streams, Amazon Data Firehose, AWS Lambda, and direct integration with Amazon S3 for log analytics:

- Monitoring and observability are enhanced through native Amazon CloudWatch integration for metrics and alerting, plus AWS CloudTrail for API call auditing
- Security integrations include AWS IAM for access control, AWS Key Management Service for encryption key management, and Amazon VPC for network isolation
- For data processing, the service connects with Amazon DynamoDB Streams for real-time data changes, Amazon SNS for notifications, and AWS Glue for ETL operations
- Storage integration with Amazon S3 enables UltraWarm and cold storage tiers, while Amazon Security Lake integration provides advanced security analytics
- The service also supports Amazon Cognito for authentication, AWS Organizations for multi-account management, and AWS Service Catalog for standardized deployments, creating a unified experience across the AWS ecosystem

Onboarding and Offboarding

Getting started with Amazon OpenSearch Service involves creating an AWS account, setting up appropriate IAM permissions, and choosing between managed clusters or serverless collections based on workload requirements. Customers can create their first domain through the OpenSearch Service console, AWS CLI, or AWS SDKs, specifying domain name, OpenSearch version, instance types, storage options, and network configuration:

- The service provides Migration Assistant to facilitate migrations from self-managed Elasticsearch or OpenSearch clusters, automating data transfer and optimizing performance
- Initial setup includes configuring security settings such as fine-grained access control, encryption options, and VPC settings
- Data can be ingested through various methods including direct API calls, Amazon Kinesis, AWS Lambda, or bulk upload operations
- OpenSearch Dashboards is automatically available for visualization and data exploration
- AWS provides comprehensive tutorials, workshops, and documentation through AWS Skill Builder, plus migration support and training for eligible customers to ensure successful adoption

Getting Started Guide: https://docs.aws.amazon.com/opensearch-service/latest/developerguide/gsg.html

Data Removal

When offboarding from Amazon OpenSearch Service, customers must first export or migrate their data using manual snapshots stored in their own S3 buckets for cross-region or cross-account transfers. All domain data, including indexes and metadata, should be backed up before deletion. The offboarding process involves deleting the OpenSearch Service domain through the console, CLI, or API, which permanently removes all associated resources including compute instances, storage volumes, and automated snapshots. Manual snapshots stored in customer-managed S3 buckets remain accessible after domain deletion, enabling data recovery if needed. Customers should also remove associated IAM policies, security groups, and VPC configurations as part of the cleanup process.

Backup/Restore and Disaster Recovery

Amazon OpenSearch Service provides comprehensive backup and disaster recovery through automated and manual snapshot capabilities. The service automatically creates hourly snapshots of all domains, storing them in Amazon S3 for up to 14 days at no additional charge. Manual snapshots can be created and stored in customer-managed S3 buckets for longer retention and cross-region disaster recovery. Snapshots include index metadata, cluster metadata, and index data, enabling complete domain restoration. For disaster recovery, customers can implement active-passive strategies using cross-region snapshot restoration, with Amazon Route 53 handling traffic redirection during failover scenarios. The service also supports cross-cluster replication for real-time data synchronization across geographically distributed domains.

Backup Capabilities

Amazon OpenSearch Service includes built-in automated backup through hourly snapshots stored in Amazon S3, retained for 14 days at no additional cost. Manual snapshots can be created on-demand and stored in customer-managed S3 buckets for extended retention and cross-region access:

- The service supports Index State Management policies for automated snapshot creation of critical indices
- However, Amazon OpenSearch Service does not directly integrate with AWS Backup service, requiring customers to use native OpenSearch snapshot mechanisms for backup and restore operations

Disaster Recovery

Amazon OpenSearch Service supports disaster recovery through multiple mechanisms including cross-region snapshot restoration, cross-cluster replication, and Multi-AZ deployments with standby instances. The service enables active-passive disaster recovery strategies using manual snapshots stored in S3 buckets across regions:

- Cross-cluster replication provides real-time data synchronization between domains in different regions
- However, Amazon OpenSearch Service does not integrate with AWS Elastic Disaster Recovery service, instead relying on OpenSearch-native capabilities and AWS infrastructure resilience for disaster recovery scenarios

Recovery Objectives

Amazon OpenSearch Service helps customers achieve their RPO and RTO objectives through automated hourly snapshots (RPO of 1 hour), Multi-AZ with Standby configuration for near-zero RTO during single AZ failures, and cross-cluster replication for real-time data synchronization. Manual snapshots enable customer-defined RPO based on backup frequency. The service's blue-green deployment model ensures zero downtime for most configuration changes. For critical workloads, the Multi-AZ with Standby feature provides 99.99% availability SLA and rapid failover capabilities, while snapshot-based recovery can restore entire domains within minutes to hours depending on data volume.

Service Constraints

Service Quotas: https://docs.aws.amazon.com/general/latest/gr/opensearch-service.html

FAQ: https://aws.amazon.com/opensearch-service/faqs/

Technical Requirements

Service Documentation: https://docs.aws.amazon.com/opensearch-service/latest/developerguide/what-is.html

User Guide: https://docs.aws.amazon.com/opensearch-service/latest/developerguide/

API Reference: https://docs.aws.amazon.com/opensearch-service/latest/APIReference/

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/opensearch-service/pricing/

Cost Optimization

Amazon OpenSearch Service offers unique cost optimization through tiered storage architecture with UltraWarm providing up to 40% cost savings for infrequently accessed data using S3-based storage, and cold storage for rarely accessed data at the lowest cost. AWS Graviton2 instances deliver up to 44% better price/performance compared to previous generation instances:

- Reserved Instances provide up to 52% discounts for predictable workloads with one or three-year terms
- The serverless deployment option enables pay-per-use pricing, eliminating overprovisioning costs for variable workloads
- OpenSearch Optimized Instances (OR1 and OI2) offer improved durability and performance with cost savings
- Customers can implement Index State Management policies to automatically transition data through storage tiers based on age and access patterns, optimizing costs while maintaining performance for active data

Savings Considerations

Primary cost savings opportunities include implementing storage tier strategies by moving infrequently accessed data to UltraWarm (40% savings) and cold storage for maximum cost reduction. Reserved Instances provide significant discounts of up to 52% for predictable workloads compared to on-demand pricing:

- Choosing appropriate instance types including AWS Graviton2 instances for better price/performance ratios and right-sizing clusters based on actual usage patterns prevents overprovisioning
- Serverless deployment eliminates idle resource costs for variable workloads
- Automated Index State Management reduces manual operational overhead while optimizing storage costs
- Regular monitoring using CloudWatch metrics and AWS Cost Explorer helps identify optimization opportunities
- Leveraging free automated snapshots instead of manual snapshots reduces storage costs, while cross-region data transfer optimization minimizes bandwidth charges for multi-region deployments