

Service Name

Amazon Keyspaces (for Apache Cassandra)

Service Overview

Amazon Keyspaces (for Apache Cassandra) is a scalable, highly available, and fully managed Apache Cassandra-compatible database service. It eliminates the need to provision, patch, or manage servers, and removes the requirement to install, maintain, or operate software. Amazon Keyspaces is serverless, allowing you to pay only for the resources you use while automatically scaling tables up and down in response to application traffic. The service supports building applications that serve thousands of requests per second with virtually unlimited throughput and storage capacity.

Key Use Cases

- **Low latency applications:** Process data at high speeds for applications requiring single-digit-millisecond latency, such as industrial equipment maintenance, trade monitoring, fleet management, and route optimization
- **Open-source technology applications:** Build applications on AWS using open-source Cassandra APIs and drivers available for multiple programming languages including Java, Python, Ruby, .NET, Node.js, PHP, C++, Perl, and Go
- **Migrating Cassandra workloads to cloud:** Move existing Cassandra workloads to AWS Cloud without managing infrastructure, reducing time-consuming and expensive self-management of Cassandra tables

Features

- **Point-in-Time Recovery (PITR):** Continuous backups with restoration to any second within the last 35 days
- **Multi-Region Replication:** Replicate data across multiple AWS Regions for improved availability and resiliency
- **Change Data Capture (CDC) Streams:** Record row-level change events from tables in near-real time
- **On-Demand and Provisioned Capacity:** Flexible billing with on-demand pay-per-request or provisioned capacity modes
- **Time to Live (TTL):** Automatically expire data from tables based on configured values
- **User-Defined Types (UDTs):** Define data structures representing real-world data hierarchies
- **Client-Side Timestamps:** Cassandra-compatible timestamps for conflict resolution and write ordering

- **Encryption at Rest and in Transit:** Built-in security with AWS-managed or customer-managed encryption keys
- **Serverless Architecture:** Automatic scaling without infrastructure management

Value Proposition

Amazon Keyspaces eliminates the operational overhead of managing Apache Cassandra infrastructure while maintaining full compatibility with existing applications and tools. It provides a serverless experience with automatic scaling, high availability across multiple Availability Zones, and enterprise-grade security features including encryption at rest and in transit:

- The service offers significant cost advantages with recent price reductions up to 75% and flexible capacity modes that optimize costs based on usage patterns
- Organizations benefit from reduced administrative burden, improved reliability with 99.999% availability SLA, seamless integration with AWS services, and the ability to focus on application development rather than database management

Service Integration

Amazon Keyspaces integrates natively with core AWS services including Amazon S3 for data archival and analytics, Amazon EMR for big data processing, Amazon Redshift for data warehousing, AWS Lambda for serverless computing, Amazon ElastiCache for caching, AWS CloudWatch for monitoring and alerting, AWS IAM for access management and security, AWS CloudTrail for audit logging, AWS CloudFormation for infrastructure as code, and AWS PrivateLink for secure network connectivity. The service also supports integration with Kinesis Client Library (KCL) for CDC stream processing and works seamlessly with existing Apache Cassandra drivers and developer tools.

Onboarding and Offboarding

Getting started with Amazon Keyspaces requires signing up for AWS, creating an IAM user with appropriate permissions, and optionally installing cqlsh-expansion for command-line access. Customers can create keyspaces and tables using the AWS Management Console, AWS CLI, CQL editor, or AWS SDKs without deploying infrastructure or installing software:

- The service provides comprehensive tutorials, code examples in multiple programming languages (.NET, Java, Python, Kotlin), and sample applications
- AWS CloudShell is recommended for initial setup and testing
- Migration tools and best practices are available for existing Cassandra workloads

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

Data Removal

To offboard from Amazon Keyspaces, customers can delete tables and keyspaces using the AWS Management Console, AWS CLI, or CQL commands. Point-in-time recovery allows restoration within 35 days if needed. Data is permanently removed when tables are deleted, with automatic cleanup of backups and associated resources. Multi-Region tables require deletion from all replicated regions. Customers should review and cancel any ongoing CDC streams, disable auto-scaling policies, and remove associated CloudWatch alarms before final cleanup.

Backup/Restore and Disaster Recovery

Amazon Keyspaces provides continuous backups through Point-in-Time Recovery (PITR) with restoration capabilities to any second within the last 35 days. Data is automatically replicated across multiple Availability Zones within a region for durability. Multi-Region replication enables disaster recovery across geographic regions. The service offers 99.999% availability SLA and built-in resilience without requiring additional disaster recovery services or manual backup management.

Backup Capabilities

Amazon Keyspaces includes built-in Point-in-Time Recovery (PITR) that creates continuous backups of table data automatically. PITR can be enabled per table and allows restoration to any point within the last 35 days. The service does not integrate with AWS Backup but provides native backup capabilities with automated management of backup retention and cleanup.

Disaster Recovery

Amazon Keyspaces provides disaster recovery through Multi-Region replication and automatic data replication across multiple Availability Zones. While it doesn't integrate with AWS Elastic Disaster Recovery, the service offers built-in resilience with 99.999% availability SLA and cross-region failover capabilities for comprehensive disaster recovery without additional services.

Recovery Objectives

Amazon Keyspaces helps meet RPO objectives through continuous PITR backups with point-in-time recovery to any second within 35 days, achieving near-zero RPO. RTO objectives are supported through automatic failover across Availability Zones, multi-region replication for geographic failover, and fast table restoration processes. The serverless architecture ensures rapid scaling and recovery without infrastructure provisioning delays.

Service Constraints

Service Quotas: <https://docs.aws.amazon.com/keyspaces/latest/devguide/quotas.html>

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

Technical Requirements

Service Documentation: <https://docs.aws.amazon.com/keyspaces/latest/devguide/what-is-keyspaces.html>

User Guide: <https://docs.aws.amazon.com/keyspaces/latest/devguide/>

API Reference: <https://docs.aws.amazon.com/keyspaces/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/keyspaces/pricing/>

Cost Optimization

Amazon Keyspaces offers unique cost optimization through dual capacity modes: on-demand for variable workloads with automatic scaling and pay-per-request billing, and provisioned mode for predictable workloads with reserved capacity. Recent price reductions up to 75% make on-demand mode cost-effective for most provisioned workloads:

- Application Auto Scaling automatically adjusts provisioned capacity based on utilization targets
- TTL-based data expiration reduces storage costs by automatically removing expired data
- Right-sizing capacity and identifying unused resources through CloudWatch metrics optimization provides additional savings opportunities

Savings Considerations

Cost savings in Amazon Keyspaces come from eliminating infrastructure management overhead, reducing operational costs through serverless architecture, and leveraging flexible capacity modes. On-demand mode reduces costs by up to 56% for single-region and 65% for multi-region usage compared to previous pricing:

- Provisioned mode offers savings for predictable workloads with reserved capacity planning
- Additional savings include reduced TTL delete costs by 75%, elimination of hardware and software maintenance expenses, and optimized data transfer costs within AWS regions
- The service's automatic scaling prevents over-provisioning while ensuring performance requirements are met