

Service Name

Amazon Kinesis Data Streams

Service Overview

Amazon Kinesis Data Streams is a fully managed, serverless streaming data service that enables real-time collection, processing, and analysis of large volumes of streaming data. It captures and processes gigabytes of data per second from various sources including website clickstreams, database event streams, financial transactions, social media feeds, IT logs, and location-tracking events. The service provides low-latency data processing with sub-second put-to-get delays, allowing applications to read data from a stream and perform real-time actions such as sending records to dashboards, generating alerts, or forwarding data to other AWS services for further processing.

Key Use Cases

- **Real-time analytics:** Processing streaming data for live dashboards, anomaly detection, and dynamic pricing
- **Log and event data processing:** Accelerated intake and processing of IT infrastructure logs, application logs, and clickstream data
- **Real-time metrics and reporting:** Continuous monitoring and reporting of business metrics and KPIs
- **Complex stream processing:** Building sophisticated data processing pipelines for machine learning and advanced analytics
- **Event-driven applications:** Powering applications that respond to real-time events and triggers

Features

- **Fully Managed Service:** Serverless architecture eliminates infrastructure management overhead
- **Multiple Capacity Modes:** Choose between On-demand Standard, On-demand Advantage, and provisioned modes based on throughput requirements
- **Enhanced Fan-out:** Dedicated throughput of 2 MB/sec per consumer for parallel data processing
- **Server-side Encryption:** Automatic encryption at rest using AWS KMS keys
- **High Availability:** Data synchronously replicated across three Availability Zones
- **Warm Throughput:** Pre-configured throughput capacity up to 10 GiB/s for instant scaling in On-demand Advantage mode
- **Extended Data Retention:** Configurable retention periods from 24 hours up to 365 days

- **Large Record Support:** Support for records up to 10 MiB in size

Value Proposition

Amazon Kinesis Data Streams offers compelling advantages over alternatives through its fully managed serverless architecture that eliminates operational overhead while providing enterprise-grade reliability with 99.95% availability SLA and 99.999% durability. The service delivers sub-second latency for real-time processing with automatic scaling capabilities, making it ideal for unpredictable workloads:

- Its deep integration with the AWS ecosystem enables seamless data flow to analytics services like Lambda, S3, Redshift, and EMR
- The flexible capacity modes allow organizations to optimize costs whether they have consistent or variable traffic patterns, while the enhanced fan-out capability supports multiple concurrent consumers without performance degradation

Service Integration

Kinesis Data Streams integrates extensively with AWS services as both data source and destination. Producer integrations include AWS DMS for change data capture, IoT Core for device data, CloudFront for web analytics, and Aurora/RDS for database streams:

- Consumer integrations encompass Lambda for serverless processing, Kinesis Data Firehose for data delivery to S3/Redshift/OpenSearch, Managed Service for Apache Flink for stream processing, EventBridge for event routing, SQS/SNS for messaging, and AWS Glue for ETL operations
- Analytics integrations include EMR for big data processing, Redshift for data warehousing, and QuickSight for visualization
- The service also supports third-party integrations with Apache Flink, Kafka, Spark, Databricks, and Adobe Experience Platform

Onboarding and Offboarding

Getting started with Kinesis Data Streams involves creating a stream through the AWS Console, CLI, or SDK by specifying a name and choosing between on-demand or provisioned capacity modes. For provisioned mode, determine the required number of shards based on throughput needs (1 MB/s write, 2 MB/s read per shard):

- Configure producers using the Kinesis API, Producer Library (KPL), or Kinesis Agent, and set up consumers using the Client Library (KCL), Lambda functions, or direct SDK calls
- Essential prerequisites include an AWS account, appropriate IAM permissions, and for development environments, Java 7+ with required libraries
- The service provides comprehensive tutorials and code examples for rapid implementation across multiple programming languages and frameworks

Getting Started Guide: <https://docs.aws.amazon.com/streams/latest/dev/examples.html>

Data Removal

Data removal is accomplished by deleting the Kinesis data stream using the DeleteStream API, AWS Console, or CLI, which permanently removes the stream, all shards, and associated data. Applications must be shut down before deletion, and the stream must be in ACTIVE state. The deletion process also removes associated tags and consumer registrations. Data is automatically purged based on the configured retention period (24 hours to 365 days), after which it becomes inaccessible. For immediate data removal, delete the stream entirely.

Backup/Restore and Disaster Recovery

Kinesis Data Streams provides built-in disaster recovery through automatic synchronous replication of data across three Availability Zones within a region. The service achieves high durability (99.999%) and availability (99.95%) through this multi-AZ architecture. Data retention periods from 24 hours to 365 days serve as a form of backup, allowing replay of historical data. For cross-region disaster recovery, customers can implement producer applications that write to streams in multiple regions or use Kinesis Data Firehose to replicate data to S3 across regions.

Backup Capabilities

Kinesis Data Streams does not integrate with AWS Backup service but provides data protection through configurable retention periods and automatic multi-AZ replication. Data is stored durably for the specified retention period (24 hours to 365 days), enabling replay and recovery scenarios:

- Extended retention allows access to historical data for up to one year
- The service's inherent design provides continuous data availability rather than traditional backup mechanisms

Disaster Recovery

The service does not integrate with AWS Elastic Disaster Recovery but provides inherent disaster recovery through multi-AZ data replication and regional resilience. Kinesis Data Streams automatically maintains copies of data across three Availability Zones, ensuring continuity during AZ failures. For broader disaster recovery strategies, customers implement cross-region replication patterns using producers that write to multiple regions or downstream services like Kinesis Data Firehose for data replication.

Recovery Objectives

Kinesis Data Streams supports aggressive RPO and RTO objectives through its real-time replication and sub-second processing capabilities. RPO approaches near-zero due to synchronous replication across Availability Zones, minimizing data loss potential. RTO is

typically minutes for AZ failures due to automatic failover mechanisms. The configurable retention period enables point-in-time recovery scenarios. For cross-region scenarios, RPO depends on replication patterns implemented by customers, while RTO involves redirecting traffic to alternate regions.

Service Constraints

Service Quotas: <https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html>

FAQ: <https://aws.amazon.com/kinesis/data-streams/faqs/>

Technical Requirements

Service Documentation:

<https://docs.aws.amazon.com/streams/latest/dev/introduction.html>

User Guide: <https://docs.aws.amazon.com/streams/latest/dev/introduction.html>

API Reference: <https://docs.aws.amazon.com/kinesis/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/kinesis/data-streams/pricing/>

Cost Optimization

Kinesis Data Streams offers multiple cost optimization strategies through its flexible capacity modes. On-demand Advantage mode provides up to 60% cost savings for workloads processing at least 10 MiB/s with committed usage pricing:

- Optimizing data retention periods reduces storage costs, while efficient shard management in provisioned mode prevents over-provisioning
- Data compression before ingestion reduces throughput costs
- The new warm throughput capability eliminates over-provisioning for predictable traffic spikes
- Enhanced fan-out should be used judiciously as it incurs additional per-consumer costs
- Choosing the appropriate capacity mode based on traffic patterns (on-demand for variable, provisioned for predictable) maximizes cost efficiency

Savings Considerations

Primary cost savings opportunities include selecting the optimal capacity mode for workload patterns, with on-demand modes eliminating over-provisioning costs and provisioned mode offering predictable pricing for steady workloads. On-demand Advantage mode delivers significant savings for high-throughput applications through committed usage pricing:

- Reducing retention periods from default settings can substantially lower storage costs
- Implementing data compression reduces throughput charges
- Efficient partition key strategies prevent hot spotting and reduce resharding costs
- Consolidating multiple small streams can reduce per-stream overhead costs
- Monitoring and right-sizing shard counts in provisioned mode prevents waste, while leveraging AWS Cost Explorer and billing alerts enables ongoing optimization