

Kinesis Data Streams

Kinesis Data Streams

- Amazon Kinesis Data Streams (KDS) is a massively scalable and durable real-time data streaming service. KDS can continuously capture gigabytes of data per second from hundreds of thousands of sources such as website clickstreams, database event streams, financial transactions, social media feeds, IT logs, and location-tracking events.
- The data collected is available in milliseconds to enable real-time analytics use cases such as real-time dashboards, real-time anomaly detection, dynamic pricing, and more.

Kinesis Data Streams

Benefits :

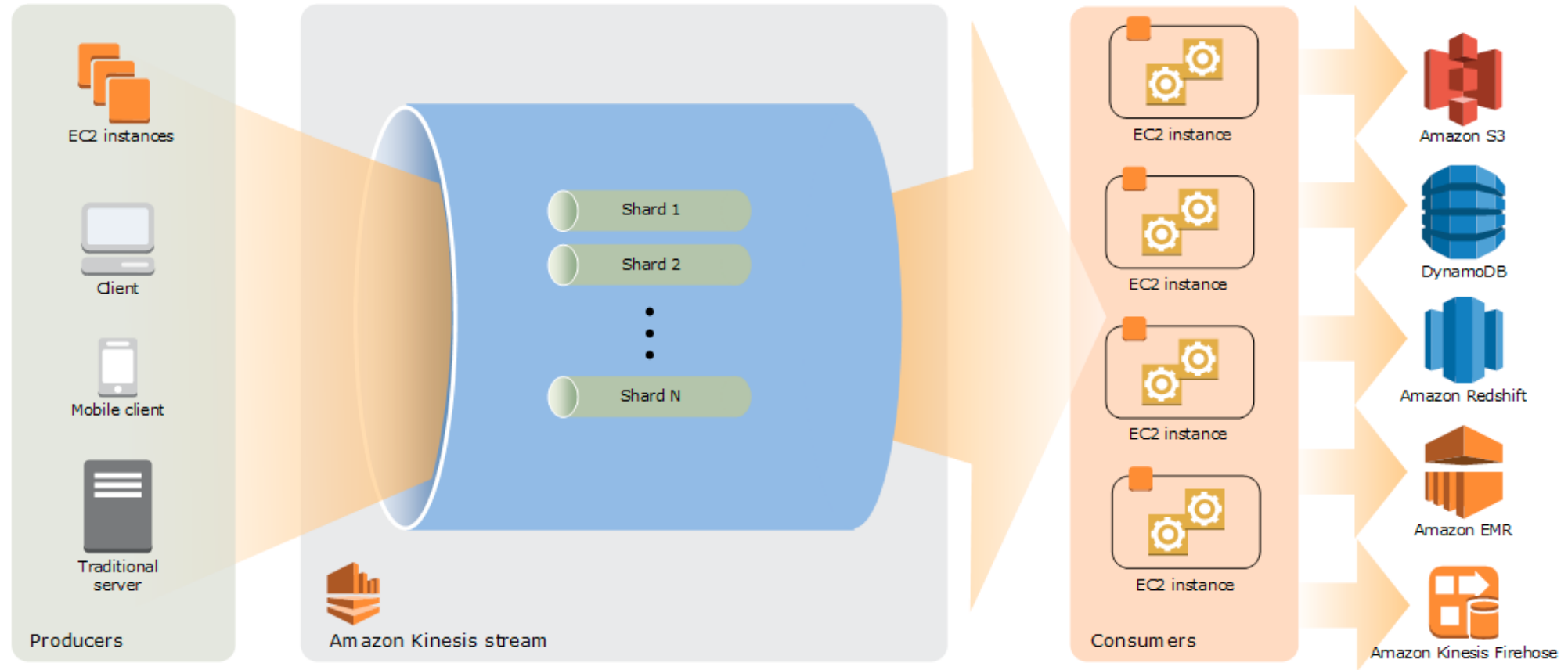
- Make your streaming data available to multiple real-time analytics applications, to Amazon S3, or to AWS Lambda within 70 milliseconds of the data being collected.
- Reduce the probability of data loss. Synchronous replication of your streaming data across three Availability Zones in an AWS Region, and the storage of that data for up to seven days, provide multiple layers of protection from data loss.
- Meet your regulatory and compliance needs by encrypting sensitive data within KDS, and privately accessing your data via your Amazon Virtual Private Cloud (VPC). Data can be secured at-rest by using server-side encryption and AWS KMS master keys.

Kinesis Data Streams

Benefits :

- Build your streaming applications quickly using the AWS SDK, the Kinesis Client Library (KCL), connectors, and agents. Easily process data with built-in integrations to AWS Lambda, Amazon Kinesis Data Analytics, and Amazon Kinesis Data Firehose.
- Kinesis data streams scale from megabytes to terabytes per hour, and scale from thousands to millions of PUT records per second.
- For as little as \$0.015 per hour, you can have a Kinesis data stream with 1MB/second ingest and 2MB/second egress capacity.

Kinesis Data Streams



Kinesis Data Streams

- **Data Producer** : A data producer is an application that typically emits data records as they are generated to a Kinesis data stream. Data producers assign partition keys to records. Partition keys ultimately determine which shard ingests the data record for a data stream.
- **Data Consumer** : A data consumer is a distributed Kinesis application or AWS service retrieving data from all shards in a stream as it is generated. Most data consumers are retrieving the most recent data in a shard, enabling real-time analytics or handling of data.

Kinesis Data Streams

- **Data Stream** : A data stream is a logical grouping of shards. There are no bounds on the number of shards within a data stream (request a limit increase if you need more). A data stream will retain data for 24 hours, or up to 7 days when extended retention is enabled.
- **Shard** : A shard is the base throughput unit of an Amazon Kinesis data stream. A shard contains an ordered sequence of records ordered by arrival time. One shard can ingest up to 1000 data records per second, or 1MB/sec. You can monitor shard-level metrics in Amazon Kinesis Data Streams.

Kinesis Data Streams

- **Partition Key** : A partition key is typically a meaningful identifier, such as a user ID or timestamp. It is specified by your data producer while putting data into an Amazon Kinesis data stream, and useful for consumers as they can use the partition key to replay or build a history associated with the partition key. The partition key is also used to segregate and route data records to different shards of a stream. For example, assuming you have an Amazon Kinesis data stream with two shards (Shard 1 and Shard 2). You can configure your data producer to use two partition keys (Key A and Key B) so that all data records with Key A are added to Shard 1 and all data records with Key B are added to Shard 2.

Kinesis Data Streams

- **Sequence Number** : A sequence number is a unique identifier for each data record. Sequence number is assigned by Amazon Kinesis Data Streams when a data producer calls PutRecord or PutRecords API to add data to an Amazon Kinesis data stream

Writing data to a stream

- **Amazon Kinesis Data Generator**
- **Amazon Kinesis Data Streams API** : Amazon Kinesis Data Streams provides two APIs for putting data into an Amazon Kinesis stream: PutRecord and PutRecords. PutRecord allows a single data record within an API call and PutRecords allows multiple data records within an API call.
- **Amazon Kinesis Producer Library** : (KPL) is an easy to use and highly configurable library that helps you put data into an Amazon Kinesis data stream.
- **Amazon Kinesis Agent** : Amazon Kinesis Agent is a pre-built Java application that offers an easy way to collect and send data to your Amazon Kinesis stream. You can install the agent on Linux-based server environments such as web servers, log servers, and database servers. The agent monitors certain files and continuously sends data to your stream.

Consuming data from a stream

- **Amazon Kinesis Data Firehose** : Amazon Kinesis Data Firehose is the easiest way to reliably transform and load streaming data into data stores and analytics tools. You can use a Kinesis data stream as a source for a Kinesis data firehose.
- **Amazon Kinesis Data Analytics** : Amazon Kinesis Data Analytics enables you to query streaming data or build entire streaming applications using SQL, so that you can gain actionable insights and respond to your business and customer needs promptly.

Consuming data from a stream

- **AWS Lambda** : You can subscribe Lambda functions to automatically read records off your Kinesis data stream. AWS Lambda is typically used for record-by-record (also known as event-based) stream processing.
- **Amazon Kinesis Client Library** : It is a pre-built library that helps you easily build Amazon Kinesis applications for reading and processing data from an Amazon Kinesis data stream. KCL handles complex issues such as adapting to changes in stream volume, load-balancing streaming data, coordinating distributed services, and processing data with fault-tolerance. KCL 2.x works with Java and Python whereas KCL 1.x works with Java, Python and NodeJS.

Consuming data from a stream

- **Amazon Kinesis Connector Library** : Amazon Kinesis Connector Library is a pre-built library that helps you easily integrate Amazon Kinesis with other AWS services and third-party tools. Amazon Kinesis Client Library (KCL) is required for using Amazon Kinesis Connector Library. The current version of this library provides connectors to Amazon DynamoDB, Amazon Redshift, Amazon S3, and Amazon Elasticsearch Service.

Management Features

- You can privately access Kinesis Data Streams APIs from your Amazon Virtual Private Cloud (VPC) by creating VPC Endpoints. With VPC Endpoints, the routing between the VPC and Kinesis Data Streams is handled by the AWS network without the need for an Internet gateway, NAT gateway, or VPN connection.
- You can encrypt the data you put into Kinesis Data Streams using Server-side encryption or client-side encryption.
- Amazon Kinesis Data Streams integrates with Amazon CloudWatch so that you can easily collect, view, and analyze CloudWatch metrics for your Amazon Kinesis data streams and the shards within those data streams