



AWS Kinesis Data Stream

AWS Kinesis Data Streams (KDS) is a real-time, scalable, and fully managed service designed to continuously collect, process, and analyze streaming data at a massive scale. It's part of the broader **Amazon Kinesis** suite, which enables real-time data streaming for various use cases, such as real-time analytics, machine learning, and data transformation.

Key Features of AWS Kinesis Data Streams:

1. **Real-time Data Streaming:** Kinesis Data Streams allows ingestion of data in real-time from multiple sources (IoT devices, application logs, social media feeds, etc.) and processes it continuously.
2. **Massive Throughput:** The service supports high volumes of data, where streams can handle terabytes of data per hour from hundreds of thousands of data producers.
3. **Data Durability and Retention:** Data is stored across multiple availability zones, ensuring high availability and durability. Streams can retain data for up to 7 days (default is 24 hours).
4. **Scalable Architecture:** Kinesis Data Streams is highly scalable. You can add or remove shards (units of capacity) to adjust the throughput capacity of your stream.
5. **Custom Processing:** You can consume the stream in real-time using consumer applications such as **AWS Lambda**, **Amazon EC2**, or **Amazon Kinesis Data Analytics**.
6. **Cost-Effective:** Pricing is based on the volume of data you stream and the resources used for processing it.

Common Use Cases:

- **Real-time analytics:** Analyze logs, events, or sensor data as it arrives.
- **Application monitoring:** Collect and process logs and metrics from applications in real-time.
- **Machine learning:** Stream data to a machine learning model for real-time predictions.
- **Data lakes and warehousing:** Continuously stream data into data lakes (e.g., Amazon S3) or warehouses for further storage and analysis.

How It Works:

1. **Data Producers:** Devices or applications send data to the stream in real time. Each data point sent is referred to as a **record**.
2. **Data Stream:** Consists of **shards**, each capable of ingesting data and enabling consumers to process it.
3. **Data Consumers:** Applications, services, or analytics tools read the data from the stream for further processing, storage, or real-time analysis.

This real-time streaming service helps businesses react to data instantly, making it valuable for time-sensitive applications.

Benefits of AWS Kinesis Data Streams:

1. **Instant Insights:** Kinesis lets you see and act on data the moment it arrives, which is useful if you want to stay ahead of problems or make real-time decisions.
2. **Scalable:** Whether you have a small or massive amount of data, Kinesis can handle it without slowing down. It grows as your needs grow.
3. **Cost-Effective:** You only pay for what you use, which makes it affordable for businesses of all sizes. You don't need to invest in expensive infrastructure to handle real-time data.
4. **No Need to Manage Servers:** Kinesis is fully managed by AWS, meaning you don't have to worry about setting up or maintaining servers to process your data. This saves you time and effort.
5. **Works with Other Tools:** You can easily connect Kinesis to other AWS services or external tools to store, analyze, or visualize your data in real-time.



To begin with the Lab:

1. In this lab, we are going to create a data stream in Kinesis and then we are going to put some records in it. We will use CLI, so you need to configure your AWS Credentials for it.
2. So, in your PowerShell you need to write these commands. The command below is used to create a data stream in Kinesis.

```
`aws kinesis create-stream --stream-name demoawsml --shard-count 1`
```

3. Now run this command to check the availability of your data stream.

```
`aws kinesis describe-stream-summary --stream-name demoawsml`
```

4. In the snapshot below, you can see that our stream has been created and it is currently active. You can also read the rest of the details.

```
Windows PowerShell
PS C:\> aws kinesis create-stream --stream-name demoawssl --shard-count 1
PS C:\> aws kinesis describe-stream-summary --stream-name demoawssl
{
  "StreamDescriptionSummary": {
    "StreamName": "demoawssl",
    "StreamARN": "arn:aws:kinesis:us-east-1:878893308172:stream/demoawssl",
    "StreamStatus": "ACTIVE",
    "StreamModeDetails": {
      "StreamMode": "PROVISIONED"
    },
    "RetentionPeriodHours": 24,
    "StreamCreationTimestamp": "2024-10-04T16:38:50+05:30",
    "EnhancedMonitoring": [
      {
        "ShardLevelMetrics": []
      }
    ],
    "EncryptionType": "NONE",
    "OpenShardCount": 1,
    "ConsumerCount": 0
  }
}
PS C:\> |
```

5. Also, you can check your stream in the AWS Console.

Amazon Kinesis > Data streams

Data streams (1) [info](#) Process data in real time Create a Firehose stream Actions Create data stream

<input type="checkbox"/>	Name	Status	Capacity mode	Provisioned shards	Sharing policy	Data retention period	Encryption	Consumers with enhanced fan-out
<input type="checkbox"/>	demoawssl	Active	Provisioned	1	No	1 day	Disabled	0

6. Now we will send the data into our Kinesis data stream using the command given below.

```
aws kinesis put-record --stream-name demoawssl `
  --partition-key 111 `
  --cli-binary-format raw-in-base64-out `
  --data '{"Data\":"1\"}'
```

7. Below you can see that you get the shard ID and the Sequence number which you can use to view your data.

```
PS C:\> aws kinesis put-record --stream-name demoawssl `
>> --partition-key 111 `
>> --cli-binary-format raw-in-base64-out `
>> --data '{"Data": "1"}'
{
  "ShardId": "shardId-000000000000",
  "SequenceNumber": "49656442628749459439097321226804756650621138775260004354"
}

PS C:\> |
```

8. So, now go inside of your Data stream choose Data Viewer, and then in the Shard choose your Shard ID. Then in the Starting Position choose **at sequence number**, copy the sequence number from CLI, and paste it here. Then click on get records, you'll be able to see your data record.

demoawssl Info Delete

Data stream summary

Status Active	Capacity mode Provisioned Data retention period 1 day	ARN arn:aws:kinesis:us-east-1:878893308172:stream/demoawssl	Creation time October 04, 2024 at 16:38 GMT+5:30
------------------	--	--	---

Applications Monitoring Configuration Enhanced fan-out (0) **Data viewer** Data analytics - new Data stream sharing EventBridge Pipes

Shard: shardId-000000000000 Starting position: At sequence number Sequence number: 49656442628749459439097321226804756650621138775260004354 Get records

Records (1) Next records

Shard: shardId-000000000000 Starting position: At sequence number Sequence number: 49656442628749459439097321226804756650621138775260004354

Find records

Partition key	Data	Approximate arrival timestamp	Sequence number
111	{"Data": "1"}	October 04, 2024 at 16:43:12 GMT+5:30	496564426287494594390973212268047566506211...

9. Similarly, you can send new data to your stream using the command mentioned above.

Applications Monitoring Configuration Enhanced fan-out (0) **Data viewer** Data analytics - new Data stream sharing EventBridge Pipes

Shard: shardId-000000000000 Starting position: Trim horizon Get records

Records (2) Next records

Shard: shardId-000000000000 Starting position: Trim horizon

Find records

Partition key	Data	Approximate arrival timestamp	Sequence number
112	{"Data": "2"}	October 04, 2024 at 16:51:01 GMT+5:30	49656442628749459439097321232186894399545...
111	{"Data": "1"}	October 04, 2024 at 16:43:12 GMT+5:30	49656442628749459439097321226804756650621...

10. Once you are done run the below command to delete your Data stream.

```
`aws kinesis delete-stream --stream-name demoawssl`
```