**😊 Kinesis Data Firehose**

Amazon Kinesis Data Firehose is a fully managed service for real-time data delivery that allows you to reliably load streaming data into data lakes, data stores, and analytics services. It can ingest data from various sources, transform or process it using AWS Lambda, and then automatically deliver it to destinations like Amazon S3, Amazon Redshift, Amazon Elasticsearch Service (now OpenSearch), or even third-party services like Splunk.

Here’s an overview of how Kinesis Data Firehose works:

**Key Features:**

1. **Real-Time Streaming Data Capture**: Kinesis Data Firehose can continuously capture and transform streaming data in real-time from services like Kinesis Data Streams, IoT devices, or custom applications.
2. **Data Transformation**: You can use AWS Lambda to transform or enrich the data before it is delivered. This is useful for cleaning or reformatting data on the fly.
3. **Automatic Scaling**: The service automatically scales to accommodate the incoming data volume without requiring user intervention.
4. **Buffering and Batching**: You can configure Kinesis Data Firehose to buffer data for a specified period (or until a certain data size is reached) before delivering it. This allows optimization of both costs and throughput.
5. **Delivery Destinations**: Firehose supports multiple destinations, including:
   * **Amazon S3**: for durable storage.
   * **Amazon Redshift**: for analytics and querying.
   * **Amazon OpenSearch Service**: for search and visualization.
   * **Splunk**: for data analysis and monitoring.
6. **Monitoring and Security**:
   * Integration with Amazon CloudWatch allows you to monitor data flow, delivery success, and failure rates.
   * Supports encryption (both at rest and in transit), and data compression (like GZIP, ZIP), which can reduce storage and transfer costs.

**Common Use Cases:**

* **Real-Time Analytics**: Capture and analyze log data, sensor data, and other real-time metrics for insights.
* **Data Lakes**: Stream large volumes of data to Amazon S3 to build a data lake.
* **Event-Driven Architectures**: Integrate with Lambda for real-time data transformations and event processing.

**Benefits:**

* **Fully Managed**: No infrastructure management is required. AWS handles provisioning, scaling, and maintenance.
* **Ease of Integration**: Works seamlessly with other AWS services like S3, Redshift, and OpenSearch.
* **Cost-Effective**: Pay only for the volume of data processed and delivered, with no upfront costs.

**In this exercise, the goal is to set up a data stream using Amazon Kinesis Data Firehose and deliver the data to an S3 bucket. You'll first create an S3 bucket and a Kinesis Data Stream using simple AWS CLI commands. After confirming the resources on the AWS Console, you configure a Kinesis Data Firehose to connect the stream (as the source) with the S3 bucket (as the destination).**

**Once the Firehose is created, you test it by sending demo data through the stream, which automatically gets stored in the S3 bucket. After a few minutes, you can download and view the file in your S3 bucket, which contains the streamed data.**

**The end goal is to demonstrate how data from a Kinesis stream can be automatically delivered to an S3 bucket, making it accessible for storage, analysis, or processing.**

**😄 To begin with the Lab:**

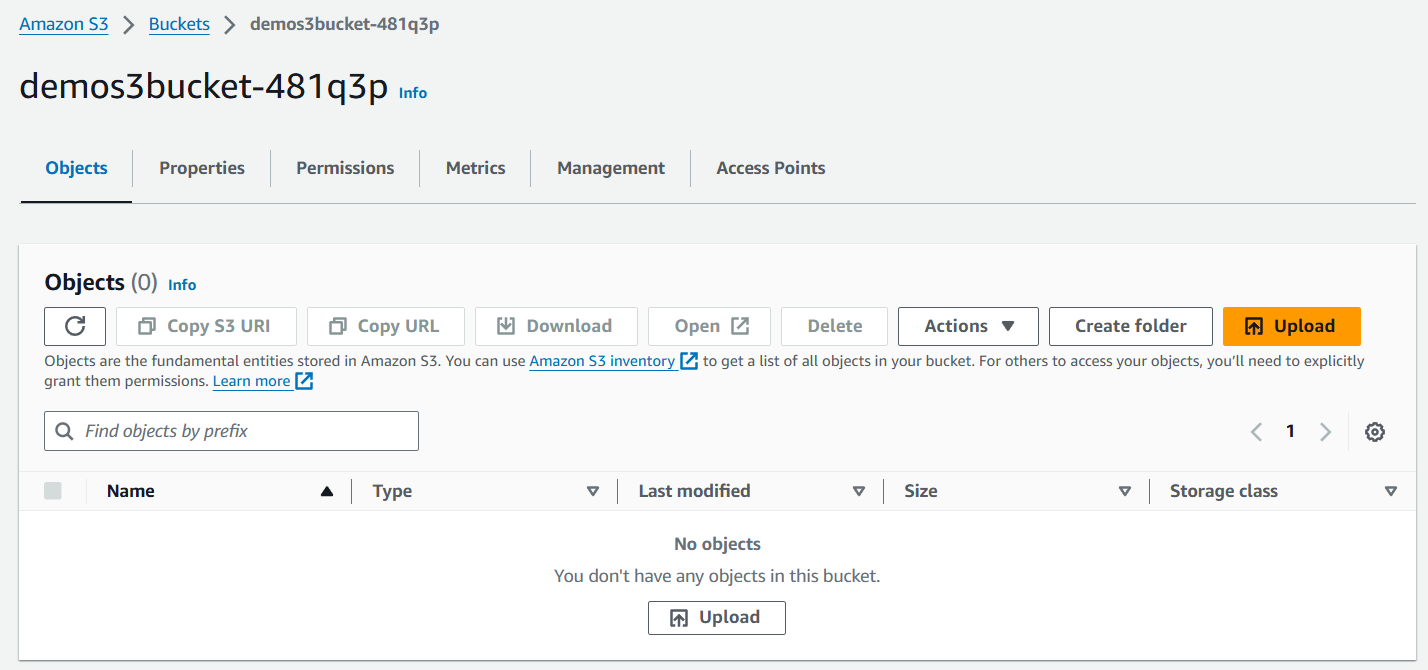
1. In this lab, we will stream data to Amazon S3 using Kinesis Data Firehose.
2. First, we need to create an S3 bucket and Kinesis Data Stream using the commands given below. Also, you will have a separate file for the commands.

**`aws s3api create-bucket --bucket demos3bucket-$RANDOM\_ID `**

**`aws kinesis create-stream --stream-name demodatastream --shard-count 1 `**

**`aws kinesis describe-stream-summary --stream-name demodatastream`**

1. After running these commands, you can go to the AWS Console and check your resources. Below you can see that our S3 bucket and Kinesis Data Stream have been created.



A screenshot of a computer

Description automatically generated

1. Now search for Kinesis Data Firehose and from its dashboard click on Create.

A screenshot of a computer

Description automatically generated

1. First, you need to choose the Source as Kinesis Data Stream and the destination as Amazon S3.

A screenshot of a computer

Description automatically generated

1. Then in the source settings choose your data Stream and leave the default name as it is.

A screenshot of a computer

Description automatically generated

1. After that for the destination choose your S3 bucket which you created earlier and then create your data Firehose.

A screenshot of a computer

Description automatically generated

1. Once your Firehose is created then you can expand the test with demo data feature.

A screenshot of a computer

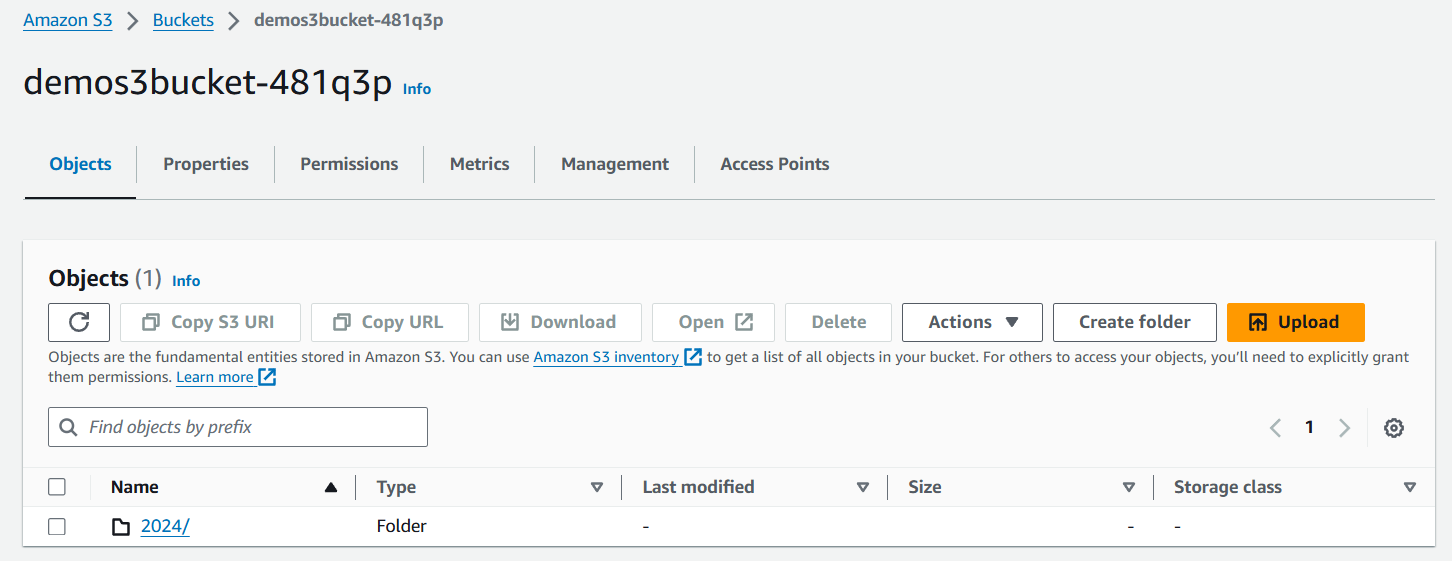
Description automatically generated

1. Here you must click on Start sending demo data and this will start sending some data to your S3 bucket and after a few minutes you must stop sending the data.

A screenshot of a computer

Description automatically generated

1. Then go to S3 and wait for a few minutes, so that your data can pop up. After few minutes you can see that we have the data in our S3 bucket.
2. Now go inside the folders and download the file then open it in notepad.



1. This type of data you will be able to see in the file.

A close-up of a screen

Description automatically generated

1. Once you are done then delete all the resources start by deleting your S3 bucket then your Kinesis data Firehose and then the Data Stream.