# Table of Contents

What Is Amazon Kinesis Data Firehose? ................................................................. 1
- Key Concepts ................................................................................................... 1
- Data Flow ...................................................................................................... 1

Setting Up .......................................................................................................... 4
- Sign Up for AWS ........................................................................................ 4
- Optional: Download Libraries and Tools ......................................................... 4

Creating a Kinesis Data Delivery Stream ............................................................ 5
- Name and source .......................................................................................... 5
- Transform records ......................................................................................... 5
- Choose destination ....................................................................................... 6
  - Choose Amazon S3 for Your Destination ................................................... 6
  - Choose Amazon Redshift for Your Destination .......................................... 7
  - Choose Amazon ES for Your Destination .................................................. 8
  - Choose Splunk for Your Destination ........................................................ 9

Configure settings ............................................................................................ 10

Testing Your Delivery Stream .......................................................................... 12
- Prerequisites ................................................................................................. 12

Sending Data to a Kinesis Data Delivery Stream .............................................. 15
- Writing Using Kinesis Data Streams ............................................................. 15
- Writing Using the Agent ................................................................................ 16
  - Prerequisites ............................................................................................ 16
  - Download and Install the Agent ................................................................. 17
  - Configure and Start the Agent .................................................................. 17
  - Agent Configuration Settings .................................................................. 18
  - Monitor Multiple File Directories and Write to Multiple Streams .......... 20
  - Use the Agent to Pre-process Data .............................................................. 21
  - Agent CLI Commands ............................................................................... 24

Writing Using the AWS SDK ........................................................................... 24
- Single Write Operations Using PutRecord ................................................. 25
- Batch Write Operations Using PutRecordBatch ......................................... 25

Writing Using CloudWatch Logs .................................................................... 25

Writing Using CloudWatch Events ................................................................. 25

Writing Using AWS IoT .................................................................................... 26

Data Transformation ....................................................................................... 27
- Data Transformation Flow .......................................................................... 27
- Data Transformation and Status Model ....................................................... 27
- Lambda Blueprints ...................................................................................... 27
- Data Transformation Failure Handling ....................................................... 28
- Source Record Backup ............................................................................... 29

Record Format Conversion ............................................................................ 30
- Record Format Conversion Requirements ................................................... 30
- Choosing the JSON Deserializer ................................................................. 30
- Choosing the Serializer .............................................................................. 31

Converting Input Record Format (Console) ....................................................... 31

Converting Input Record Format (API) ............................................................ 32

Record Format Conversion Error Handling .................................................... 32

Data Delivery ................................................................................................. 33
- Data Delivery Format .................................................................................. 33
- Data Delivery Frequency ............................................................................. 33
- Data Delivery Failure Handling .................................................................. 34
What Is Amazon Kinesis Data Firehose?

Amazon Kinesis Data Firehose is a fully managed service for delivering real-time streaming data to destinations such as Amazon Simple Storage Service (Amazon S3), Amazon Redshift, Amazon Elasticsearch Service (Amazon ES), and Splunk. Kinesis Data Firehose is part of the Kinesis streaming data platform, along with Kinesis Streams and Amazon Kinesis Data Analytics. With Kinesis Data Firehose, you don't need to write applications or manage resources. You configure your data producers to send data to Kinesis Data Firehose, and it automatically delivers the data to the destination that you specified. You can also configure Kinesis Data Firehose to transform your data before delivering it.

For more information about AWS big data solutions, see Big Data. For more information about AWS streaming data solutions, see What is Streaming Data?

Key Concepts

As you get started with Kinesis Data Firehose, you'll benefit from understanding the following concepts:

Kinesis data delivery stream

The underlying entity of Kinesis Data Firehose. You use Kinesis Data Firehose by creating a Kinesis data delivery stream and then sending data to it. For more information, see Creating an Amazon Kinesis Data Firehose Delivery Stream (p. 5) and Sending Data to an Amazon Kinesis Firehose Delivery Stream (p. 15).

record

The data of interest that your data producer sends to a Kinesis data delivery stream. A record can be as large as 1,000 KB.

data producer

Producers send records to Kinesis data delivery streams. For example, a web server that sends log data to a Kinesis data delivery stream is a data producer. You can also configure your Kinesis data delivery stream to automatically read data from an existing Kinesis stream, and load it into destinations. For more information, see Sending Data to an Amazon Kinesis Firehose Delivery Stream (p. 15).

buffer size and buffer interval

Kinesis Data Firehose buffers incoming streaming data to a certain size or for a certain period of time before delivering it to destinations. **Buffer Size** is in MBs and **Buffer Interval** is in seconds.

Data Flow

For Amazon S3 destinations, streaming data is delivered to your S3 bucket. If data transformation is enabled, you can optionally back up source data to another Amazon S3 bucket.
For Amazon Redshift destinations, streaming data is delivered to your S3 bucket first. Kinesis Data Firehose then issues an Amazon Redshift COPY command to load data from your S3 bucket to your Amazon Redshift cluster. If data transformation is enabled, you can optionally back up source data to another Amazon S3 bucket.

For Amazon ES destinations, streaming data is delivered to your Amazon ES cluster, and it can optionally be backed up to your S3 bucket concurrently.
For Splunk destinations, streaming data is delivered to Splunk, and it can optionally be backed up to your S3 bucket concurrently.
Setting Up for Amazon Kinesis Data Firehose

Before you use Kinesis Data Firehose for the first time, complete the following tasks.

Tasks
- Sign Up for AWS (p. 4)
- Optional: Download Libraries and Tools (p. 4)

Sign Up for AWS

When you sign up for Amazon Web Services (AWS), your AWS account is automatically signed up for all services in AWS, including Kinesis Data Firehose. You are charged only for the services that you use.

If you have an AWS account already, skip to the next task. If you don't have an AWS account, use the following procedure to create one.

To sign up for an AWS account
1. Open https://aws.amazon.com/, and then choose Create an AWS Account.
   
   Note
   This might be unavailable in your browser if you previously signed into the AWS Management Console. In that case, choose Sign in to a different account, and then choose Create a new AWS account.

2. Follow the online instructions.
   
   Part of the sign-up procedure involves receiving a phone call and entering a PIN using the phone keypad.

Optional: Download Libraries and Tools

The following libraries and tools will help you work with Kinesis Data Firehose programmatically and from the command line:

- The Amazon Kinesis Firehose API Reference is the basic set of operations that Kinesis Data Firehose supports.
- The AWS SDKs for Java, .NET, Node.js, Python, and Ruby include Kinesis Data Firehose support and samples.

If your version of the AWS SDK for Java does not include samples for Kinesis Data Firehose, you can also download the latest AWS SDK from GitHub.

- The AWS Command Line Interface supports Kinesis Data Firehose. The AWS CLI enables you to control multiple AWS services from the command line and automate them through scripts.
Creating an Amazon Kinesis Data Firehose Delivery Stream

You can use the AWS Management Console or an AWS SDK to create a Kinesis data delivery stream to your chosen destination.

You can update the configuration of your Kinesis data delivery stream at any time after it's created, using the Kinesis Data Firehose console or UpdateDestination. Your Kinesis data delivery stream remains in the ACTIVE state while your configuration is updated, and you can continue to send data. The updated configuration normally takes effect within a few minutes. The version number of a Kinesis data delivery stream is increased by a value of 1 after you update the configuration, and it is reflected in the delivered Amazon S3 object name. For more information, see Amazon S3 Object Name Format (p. 35).

The following topics describe how to create a Kinesis data delivery stream:

Contents
- Name and source (p. 5)
- Transform records (p. 5)
- Choose destination (p. 6)
- Configure settings (p. 10)

Name and source

This topic describes the Name and source page of the Create Delivery Stream wizard.

Name and source

1. Open the Kinesis Firehose console at https://console.aws.amazon.com/firehose/.
2. Choose Create Delivery Stream. On the Name and source page, enter values for the following fields:

   Delivery stream name
   
   The name of your Kinesis data delivery stream.

   Source
   
   • Direct PUT or other sources: Choose this option to create a Kinesis data delivery stream that producer applications write directly to.
   • Kinesis stream: Choose this option to configure a Kinesis data delivery stream that uses a Kinesis stream as a data source. You can then use Amazon Kinesis Data Firehose to read data easily from an existing Kinesis stream and load it into destinations. For more information about using Kinesis Data Streams as your data source, see Writing to Amazon Kinesis Data Firehose Using Kinesis Data Streams.
3. Choose Next to advance to the Transform records (p. 5) page.

Transform records

This topic describes the Transform records page of the Create Delivery Stream wizard.
Transform records

1. On the Transform records with AWS Lambda page, enter values for the following fields:

   Record transformation

   Choose Disabled to create a Kinesis data delivery stream that does not transform incoming data. Choose Enabled to specify a Lambda function that Kinesis Data Firehose can invoke to transform incoming data before delivering it. You can configure a new Lambda function using one of the Lambda blueprints or choose an existing Lambda function. Your Lambda function must contain the status model required by Kinesis Data Firehose. For more information, see Amazon Kinesis Data Firehose Data Transformation (p. 27).

2. Choose Next to advance to the Choose destination (p. 6) page.

Choose destination

This topic describes the Choose destination page of the Create Delivery Stream wizard.

Amazon Kinesis Data Firehose can send records to Amazon S3, Amazon Redshift, or Amazon Elasticsearch Service.

Topics
- Choose Amazon S3 for Your Destination (p. 6)
- Choose Amazon Redshift for Your Destination (p. 7)
- Choose Amazon ES for Your Destination (p. 8)
- Choose Splunk for Your Destination (p. 9)

Choose Amazon S3 for Your Destination

This section describes options for using Amazon S3 for your destination.

To choose Amazon S3 for your destination

- On the Choose destination page, enter values for the following fields:

  Destination

  Choose Amazon S3.

  Destination S3 bucket

  Choose an S3 bucket that you own where the streaming data should be delivered. You can create a new S3 bucket or choose an existing one.

  Destination S3 bucket prefix

  (Optional) To use the default prefix for S3 objects, leave this option blank. Kinesis Data Firehose automatically uses a prefix in "YYYY/MM/DD/HH" UTC time format for delivered S3 objects. You can add to the start of this prefix. For more information, see Amazon S3 Object Name Format (p. 35).

  Source record S3 backup

  Choose Disabled to disable source record backup. If you enable data transformation with Lambda, you can enable source record backup to deliver untransformed incoming data to
Choose Amazon Redshift for Your Destination

This section describes options for using Amazon Redshift for your destination.

To choose Amazon Redshift for your destination

- On the Choose destination page, enter values for the following fields:
  
  **Destination**
  
  Choose Amazon Redshift.

  **Cluster**
  
  The Amazon Redshift cluster to which S3 bucket data is copied. Configure the Amazon Redshift cluster to be publicly accessible and unblock Kinesis Data Firehose IP addresses. For more information, see Grant Kinesis Data Firehose Access to an Amazon Redshift Destination (p. 57).

  **User name**
  
  An Amazon Redshift user with permissions to access the Amazon Redshift cluster. This user needs to have the Amazon Redshift INSERT privilege for copying data from the S3 bucket to the Amazon Redshift cluster.

  **Password**
  
  The password for the user who has permissions to access the cluster.

  **Database**
  
  The Amazon Redshift database to where the data is copied.

  **Table**
  
  The Amazon Redshift table to where the data is copied.

  **Columns**
  
  (Optional) The specific columns of the table to which the data is copied. Use this option if the number of columns defined in your S3 objects is less than the number of columns within the Amazon Redshift table.

  **Intermediate S3 bucket**
  
  Kinesis Data Firehose delivers your data to your S3 bucket first and then issues an Amazon Redshift COPY command to load the data into your Amazon Redshift cluster. Specify an S3 bucket that you own where the streaming data should be delivered. Create a new S3 bucket or choose an existing bucket that you own.

  Kinesis Data Firehose doesn’t delete the data from your S3 bucket after loading it to your Amazon Redshift cluster. You can manage the data in your S3 bucket using a lifecycle configuration. For more information, see Object Lifecycle Management in the Amazon Simple Storage Service Developer Guide.

  **Intermediate S3 bucket prefix**
  
  (Optional) To use the default prefix for S3 objects, leave this option blank. Kinesis Data Firehose automatically uses a prefix in "YYYY/MM/DD/HH" UTC time format for delivered S3 objects. You can add to the start of this prefix. For more information, see Amazon S3 Object Name Format (p. 35).
COPY options

Parameters that you can specify in the Amazon Redshift COPY command. These may be required for your configuration. For example, "GZIP" is required if S3 data compression is enabled; "REGION" is required if your S3 bucket isn’t in the same AWS Region as your Amazon Redshift cluster. For more information, see COPY in the Amazon Redshift Database Developer Guide.

COPY command

The Amazon Redshift COPY command. For more information, see COPY in the Amazon Redshift Database Developer Guide.

Retry duration

Time duration (0–7200 seconds) for Kinesis Data Firehose to retry if data COPY to your Amazon Redshift cluster fails. If the value is 0 (zero) seconds, Kinesis Data Firehose does not retry upon COPY command failure.

Source record S3 backup

If you enable data transformation with Lambda, you can enable source record backup to deliver untransformed incoming data to a separate S3 bucket. You cannot disable source record backup after you enable it.

Backup S3 bucket

The S3 bucket to receive the untransformed data.

Backup S3 bucket prefix

To use the default prefix for source record backup, leave this option blank. Kinesis Data Firehose automatically uses a prefix in "YYYY/MM/DD/HH" UTC time format for delivered S3 objects. You can add to the start of this prefix. For more information, see Amazon S3 Object Name Format (p. 35). This value is optional.

Choose Amazon ES for Your Destination

This section describes options for using Amazon ES for your destination.

To choose Amazon ES for your destination

1. On the Choose destination page, enter values for the following fields:

   Destination

   Choose Amazon Elasticsearch Service.

   Domain

   The Amazon ES domain to which your data is delivered.

   Index

   The Elasticsearch index name to be used when indexing data to your Amazon ES cluster.

   Index rotation

   Choose whether and how often the Elasticsearch index should be rotated. If index rotation is enabled, Kinesis Firehose appends the corresponding time stamp to the specified index name and rotates. For more information, see Index Rotation for the Amazon ES Destination (p. 36).
Choose Splunk for Your Destination

This section describes options for using Splunk for your destination.

To choose Splunk for your destination

- On the Choose destination page, enter values for the following fields:
  
  **Destination**

  Choose Splunk.

  **Splunk cluster endpoint**

  To determine the endpoint, see Configure Amazon Kinesis Firehose to Send Data to the Splunk Platform in the Splunk documentation.

  **Splunk endpoint type**

  Choose Raw in most cases. Choose Event if you have preprocessed your data using AWS Lambda in order to send data to different indexes by event type. For detailed information on what endpoint to use, see Configure Amazon Kinesis Firehose to Send Data to the Splunk Platform in the Splunk documentation.

  **Authentication token**

  To set up a Splunk endpoint that can receive data from Kinesis Firehose, follow the instructions at Installation and Configuration Overview for the Splunk Add-on for Amazon Kinesis Firehose.
Configure settings

Configure settings

1. On the Configure settings page, enter values for the following fields:

   Buffer size, Buffer interval

   Kinesis Data Firehose buffers incoming data before delivering it to Amazon S3. You can choose a buffer size (1–128 MBs) or buffer interval (60–900 seconds); whichever condition is satisfied first triggers data delivery to Amazon S3. If you enable data transformation, the buffer interval applies from the time transformed data is received by Kinesis Data Firehose to the data delivery
to Amazon S3. In circumstances where data delivery to the destination falls behind data writing to the delivery stream, Kinesis Data Firehose raises the buffer size dynamically to catch up and ensure that all data is delivered to the destination.

**Compression**

Choose GZIP, Snappy, or Zip data compression, or no data compression. Snappy or Zip compression are not available for delivery streams with Amazon Redshift as the destination.

**Encryption**

Kinesis Data Firehose supports Amazon S3 server-side encryption with AWS Key Management Service (AWS KMS) for encrypting delivered data in Amazon S3. You can choose to not encrypt the data or to encrypt with a key from the list of AWS KMS keys that you own. For more information, see [Protecting Data Using Server-Side Encryption with AWS KMS–Managed Keys (SSE-KMS)](https://docs.aws.amazon.com/kinesis/latest/dev/index.html).

**Error logging**

Kinesis Data Firehose can log the Lambda invocation, if data transformation is enabled, and send data delivery errors to CloudWatch Logs so that you can view the specific error logs if the Lambda invocation or data delivery fails. For more information, see [Monitoring with Amazon CloudWatch Logs](https://docs.aws.amazon.com/currie/latest/userguide/cloudwatch-logs-streams.html).

**IAM role**

You can choose to create a new role where required permissions are assigned automatically, or choose an existing role created for Kinesis Data Firehose. The role is used to grant Kinesis Data Firehose access to your S3 bucket, AWS KMS key (if data encryption is enabled), and Lambda function (if data transformation is enabled). The console might create a role with placeholders. You can safely ignore or safely delete lines with `%FIREHOSE_BUCKET_NAME%`, `%FIREHOSE_DEFAULT_FUNCTION%`, or `%FIREHOSE_DEFAULT_VERSION%`. For more information, see [Grant Kinesis Data Firehose Access to an Amazon S3 Destination](https://docs.aws.amazon.com/kinesis/latest/dev/how-to-grant-s3-access.html).

2. Review the settings and choose **Create Delivery Stream**.

The new Kinesis data delivery stream takes a few moments in the **Creating** state before it is available. After your Kinesis data delivery stream is in an **Active** state, you can start sending data to it from your producer.
Testing Your Delivery Stream Using Sample Data

You can use the AWS Management Console to ingest simulated stock ticker data. The console runs a script in your browser to put sample records in your Kinesis Firehose delivery stream. This enables you to test the configuration of your delivery stream without having to generate your own test data.

The following is an example from the simulated data:

```
{"TICKER_SYMBOL":"QXZ","SECTOR":"HEALTHCARE","CHANGE":-0.05,"PRICE":84.51}
```

Note that standard Amazon Kinesis Firehose charges apply when your delivery stream transmits the data, but there is no charge when the data is generated. To stop incurring these charges, you can stop the sample stream from the console at any time.

Contents

- Prerequisites (p. 12)
- Test Using Amazon S3 as the Destination (p. 12)
- Test Using Amazon Redshift as the Destination (p. 12)
- Test Using Amazon ES as the Destination (p. 13)
- Test Using Splunk as the Destination (p. 13)

Prerequisites

Before you begin, create a delivery stream. For more information, see Creating an Amazon Kinesis Data Firehose Delivery Stream (p. 5).

Test Using Amazon S3 as the Destination

Use the following procedure to test your delivery stream using Amazon Simple Storage Service (Amazon S3) as the destination.

To test a delivery stream using Amazon S3

1. Open the Kinesis Firehose console at https://console.aws.amazon.com/firehose/.
2. Choose the delivery stream.
3. Under Test with demo data, choose Start sending demo data to generate sample stock ticker data.
4. Follow the onscreen instructions to verify that data is being delivered to your S3 bucket. Note that it might take a few minutes for new objects to appear in your bucket, based on the buffering configuration of your bucket.
5. When the test is complete, choose Stop sending demo data to stop incurring usage charges.

Test Using Amazon Redshift as the Destination

Use the following procedure to test your delivery stream using Amazon Redshift as the destination.
To test a delivery stream using Amazon Redshift

1. Your delivery stream expects a table to be present in your Amazon Redshift cluster. Connect to Amazon Redshift through a SQL interface and run the following statement to create a table that accepts the sample data.

```
create table firehose_test_table
(
    Ticker.Symbol varchar(4),
    Sector varchar(16),
    Change float,
    Price float
);
```

2. Open the Kinesis Firehose console at https://console.aws.amazon.com/firehose/.
3. Choose the delivery stream.
4. Edit the destination details for your delivery stream to point to the newly created firehose_test_table table.
5. Under Test with demo data, choose Start sending demo data to generate sample stock ticker data.
6. Follow the onscreen instructions to verify that data is being delivered to your table. Note that it might take a few minutes for new rows to appear in your table, based on the buffering configuration.
7. When the test is complete, choose Stop sending demo data to stop incurring usage charges.
8. Edit the destination details for your Kinesis Firehose delivery stream to point to another table.
9. (Optional) Delete the firehose_test_table table.

Test Using Amazon ES as the Destination

Use the following procedure to test your delivery stream using Amazon Elasticsearch Service (Amazon ES) as the destination.

To test a delivery stream using Amazon ES

1. Open the Kinesis Firehose console at https://console.aws.amazon.com/firehose/.
2. Choose the delivery stream.
3. Under Test with demo data, choose Start sending demo data to generate sample stock ticker data.
4. Follow the onscreen instructions to verify that data is being delivered to your Amazon ES domain. For more information, see Searching Documents in an Amazon ES Domain in the Amazon Elasticsearch Service Developer Guide.
5. When the test is complete, choose Stop sending demo data to stop incurring usage charges.

Test Using Splunk as the Destination

Use the following procedure to test your delivery stream using Splunk as the destination.

To test a delivery stream using Splunk

1. Open the Kinesis Firehose console at https://console.aws.amazon.com/firehose/.
2. Choose the delivery stream.
3. Under Test with demo data, choose Start sending demo data to generate sample stock ticker data.
4. Check whether the data is being delivered to your Splunk index. Example search terms in Splunk are `sourcetype="aws:firehose:json"` and `index="name-of-your-splunk-index"`. For more information about how to search for events in Splunk, see Search Manual in the Splunk documentation.

If the test data doesn't appear in your Splunk index, check your Amazon S3 bucket for failed events. Also see Data Not Delivered to Splunk.

5. When you finish testing, choose Stop sending demo data to stop incurring usage charges.
Sending Data to an Amazon Kinesis Firehose Delivery Stream

You can send data to your Kinesis data delivery stream using different types of sources: You can use a Kinesis stream, the Kinesis Agent, or the Kinesis Firehose API using the AWS SDK. You can also use Amazon CloudWatch Logs, CloudWatch Events, or AWS IoT as your data source. If you are new to Kinesis Firehose, take some time to become familiar with the concepts and terminology presented in What Is Amazon Kinesis Data Firehose? (p. 1).

Note
Some AWS services can only send messages and events to a Kinesis Data Firehose delivery stream that is in the same Region. If your Kinesis Data Firehose delivery stream doesn't appear as an option when you're configuring a target for Amazon CloudWatch Logs, CloudWatch Events, or AWS IoT, verify that your Kinesis Data Firehose delivery stream is in the same Region as your other services.

Topics
- Writing to Amazon Kinesis Data Firehose Using Kinesis Data Streams (p. 15)
- Writing to Amazon Kinesis Data Firehose Using Kinesis Agent (p. 16)
- Writing to a Kinesis Data Delivery Stream Using the AWS SDK (p. 24)
- Writing to Amazon Kinesis Data Firehose Using CloudWatch Logs (p. 25)
- Writing to Amazon Kinesis Data Firehose Using CloudWatch Events (p. 25)
- Writing to Amazon Kinesis Data Firehose Using AWS IoT (p. 26)

Writing to Amazon Kinesis Data Firehose Using Kinesis Data Streams

You can configure Kinesis Data Streams to send information to a Kinesis Data Firehose delivery stream.

1. Open the Kinesis Firehose console at https://console.aws.amazon.com/firehose/.
2. Choose Create Delivery Stream. On the Name and source page, enter values for the following fields:

   Delivery stream name
   The name of your Kinesis data delivery stream.

   Source
   Choose the Kinesis stream option to configure a Kinesis data delivery stream that uses a Kinesis stream as a data source. You can then use Amazon Kinesis Data Firehose to read data easily from an existing Kinesis stream and load it into destinations.

   To use a Kinesis stream as a source, choose an existing stream in the Kinesis stream list, or choose Create new to create a new Kinesis stream. After you create a new stream, choose
Refresh to update the Kinesis stream list. If you have a large number of streams, filter the list using Filter by name.

Note
When you configure a Kinesis stream as the source of a Kinesis Data Firehose delivery stream, the Kinesis Data Firehose PutRecord and PutRecordBatch operations are disabled. To add data to your Kinesis Data Firehose delivery stream in this case, use the Kinesis Data Streams PutRecord and PutRecords operations.

Kinesis Data Firehose starts reading data from the LATEST position of your Kinesis stream. For more information about Kinesis Data Streams positions, see GetShardIterator. Kinesis Data Firehose calls the Kinesis Data Streams GetRecords operation once per second for each shard.

More than one Kinesis Data Firehose delivery stream can read from the same Kinesis stream. Other Kinesis applications (consumers) can also read from the same stream. Each call from any Kinesis Data Firehose delivery stream or other consumer application counts against the overall throttling limit for the shard. To avoid getting throttled, plan your applications carefully. For more information about Kinesis Data Streams limits, see Amazon Kinesis Streams Limits.

3. Choose Next to advance to the Transform records (p. 5) page.

Writing to Amazon Kinesis Data Firehose Using Kinesis Agent

Kinesis Agent is a stand-alone Java software application that offers an easy way to collect and send data to Kinesis Data Firehose. The agent continuously monitors a set of files and sends new data to your Kinesis data delivery stream. The agent handles file rotation, checkpointing, and retry upon failures. It delivers all of your data in a reliable, timely, and simple manner. It also emits Amazon CloudWatch metrics to help you better monitor and troubleshoot the streaming process.

By default, records are parsed from each file based on the newline (‘\n’) character. However, the agent can also be configured to parse multi-line records (see Agent Configuration Settings (p. 18)).

You can install the agent on Linux-based server environments such as web servers, log servers, and database servers. After installing the agent, configure it by specifying the files to monitor and the delivery stream for the data. After the agent is configured, it durably collects data from the files and reliably sends it to the delivery stream.

Topics
• Prerequisites (p. 16)
• Download and Install the Agent (p. 17)
• Configure and Start the Agent (p. 17)
• Agent Configuration Settings (p. 18)
• Monitor Multiple File Directories and Write to Multiple Streams (p. 20)
• Use the Agent to Pre-process Data (p. 21)
• Agent CLI Commands (p. 24)

Prerequisites
• Your operating system must be either Amazon Linux AMI with version 2015.09 or later, or Red Hat Enterprise Linux version 7 or later.
Download and Install the Agent

First, connect to your instance. For more information, see Connect to Your Instance in the Amazon EC2 User Guide for Linux Instances. If you have trouble connecting, see Troubleshooting Connecting to Your Instance in the Amazon EC2 User Guide for Linux Instances.

Next, install the agent using one of the following methods.

**To set up the agent using the Amazon Linux AMI**

Use the following command to download and install the agent:

```
sudo yum install -y aws-kinesis-agent
```

**To set up the agent using Red Hat Enterprise Linux**

Use the following command to download and install the agent:

```
```

**To set up the agent using GitHub**

1. Download the agent from awslabs/amazon-kinesis-agent.
2. Install the agent by navigating to the download directory and running the following command:

```
sudo ./setup --install
```

Configure and Start the Agent

**To configure and start the agent**

1. Open and edit the configuration file (as superuser if using default file access permissions): `/etc/aws-kinesis/agent.json`
In this configuration file, specify the files ("filePattern") from which the agent collects data, and the name of the delivery stream ("deliveryStream") to which the agent sends data. Note that the file name is a pattern, and the agent recognizes file rotations. You can rotate files or create new files no more than once per second. The agent uses the file creation timestamp to determine which files to track and tail into your delivery stream; creating new files or rotating files more frequently than once per second does not allow the agent to differentiate properly between them.

```
{
  "flows": [
    {
      "filePattern": "/tmp/app.log*",
      "deliveryStream": "yourdeliverystream"
    }
  ]
}
```

Note that the default region is `us-east-1`. If you are using a different region, add the `firehose.endpoint` setting to the configuration file, specifying the endpoint for your region. For more information, see Agent Configuration Settings (p. 18).

2. Start the agent manually:

```
sudo service aws-kinesis-agent start
```

3. (Optional) Configure the agent to start on system startup:

```
sudo chkconfig aws-kinesis-agent on
```

The agent is now running as a system service in the background. It continuously monitors the specified files and sends data to the specified delivery stream. Agent activity is logged in `/var/log/aws-kinesis-agent/aws-kinesis-agent.log`.

### Agent Configuration Settings

The agent supports two mandatory configuration settings, `filePattern` and `deliveryStream`, plus optional configuration settings for additional features. You can specify both mandatory and optional configuration settings in `/etc/aws-kinesis/agent.json`.

Whenever you change the configuration file, you must stop and start the agent, using the following commands:

```
sudo service aws-kinesis-agent stop
sudo service aws-kinesis-agent start
```

Alternatively, you could use the following command:

```
sudo service aws-kinesis-agent restart
```

The following are the general configuration settings.

<table>
<thead>
<tr>
<th>Configuration Setting</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>assumeRoleARN</td>
<td>The ARN of the role to be assumed by the user. For more information, see <a href="#">Delegate Access Across AWS Accounts Using IAM Roles</a> in the IAM User Guide.</td>
</tr>
</tbody>
</table>
### Agent Configuration Settings

<table>
<thead>
<tr>
<th>Configuration Setting</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>assumeRoleExternalId</td>
<td>An optional identifier that determines who can assume the role. For more information, see <a href="#">How to Use an External ID in the IAM User Guide</a>.</td>
</tr>
<tr>
<td>awsAccessKeyId</td>
<td>AWS access key ID that overrides the default credentials. This setting takes precedence over all other credential providers.</td>
</tr>
<tr>
<td>awsSecretAccessKey</td>
<td>AWS secret key that overrides the default credentials. This setting takes precedence over all other credential providers.</td>
</tr>
<tr>
<td>cloudwatch.emitMetrics</td>
<td>Enables the agent to emit metrics to CloudWatch if set (true). Default: true</td>
</tr>
<tr>
<td>cloudwatch.endpoint</td>
<td>The regional endpoint for CloudWatch. Default: monitoring.us-east-1.amazonaws.com</td>
</tr>
<tr>
<td>firehose.endpoint</td>
<td>The regional endpoint for Kinesis Data Firehose. Default: firehose.us-east-1.amazonaws.com</td>
</tr>
</tbody>
</table>

The following are the flow configuration settings.

<table>
<thead>
<tr>
<th>Configuration Setting</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dataProcessingOptions</td>
<td>The list of processing options applied to each parsed record before it is sent to the delivery stream. The processing options are performed in the specified order. For more information, see <a href="#">Use the Agent to Pre-process Data (p. 21)</a>.</td>
</tr>
<tr>
<td>deliveryStream</td>
<td>[Required] The name of the delivery stream.</td>
</tr>
<tr>
<td>filePattern</td>
<td>[Required] A glob for the files that need to be monitored by the agent. Any file that matches this pattern is picked up by the agent automatically and monitored. For all files matching this pattern, read permission must be granted to aws-kinesis-agent-user. For the directory containing the files, read and execute permissions must be granted to aws-kinesis-agent-user.</td>
</tr>
<tr>
<td>initialPosition</td>
<td>The initial position from which the file started to be parsed. Valid values are START_OF_FILE and END_OF_FILE. Default: END_OF_FILE</td>
</tr>
<tr>
<td>maxBufferAgeMillis</td>
<td>The maximum time, in milliseconds, for which the agent buffers data before sending it to the delivery stream. Value range: 1,000 to 900,000 (1 second to 15 minutes) Default: 60,000 (1 minute)</td>
</tr>
<tr>
<td>maxBufferSizeBytes</td>
<td>The maximum size, in bytes, for which the agent buffers data before sending it to the delivery stream. Value range: 1 to 4,194,304 (4 MB) Default: 4,194,304 (4 MB)</td>
</tr>
</tbody>
</table>
Monitor Multiple File Directories and Write to Multiple Streams

<table>
<thead>
<tr>
<th>Configuration Setting</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>maxBufferSizeRecords</td>
<td>The maximum number of records for which the agent buffers data before sending it to the delivery stream.</td>
</tr>
<tr>
<td>Value range: 1 to 500</td>
<td>Default: 500</td>
</tr>
<tr>
<td>minTimeBetweenFilePollsMillis</td>
<td>The time interval, in milliseconds, at which the agent polls and parses the monitored files for new data.</td>
</tr>
<tr>
<td>Value range: 1 or more</td>
<td>Default: 100</td>
</tr>
<tr>
<td>multiLineStartPattern</td>
<td>The pattern for identifying the start of a record. A record is made of a line that matches the pattern and any following lines that don't match the pattern. The valid values are regular expressions. By default, each new line in the log files is parsed as one record.</td>
</tr>
<tr>
<td>skipHeaderLines</td>
<td>The number of lines for the agent to skip parsing at the beginning of monitored files.</td>
</tr>
<tr>
<td>Value range: 0 or more</td>
<td>Default: 0 (zero)</td>
</tr>
<tr>
<td>truncatedRecordTerminator</td>
<td>The string that the agent uses to truncate a parsed record when the record size exceeds the Kinesis Firehose record size limit. (1,000 KB)</td>
</tr>
<tr>
<td>Default: '\n' (newline)</td>
<td></td>
</tr>
</tbody>
</table>

Monitor Multiple File Directories and Write to Multiple Streams

By specifying multiple flow configuration settings, you can configure the agent to monitor multiple file directories and send data to multiple streams. In the following configuration example, the agent monitors two file directories and sends data to an Kinesis stream and a Kinesis Firehose delivery stream respectively. Note that you can specify different endpoints for Kinesis Data Streams and Kinesis Firehose so that your Kinesis stream and Kinesis Firehose delivery stream don't need to be in the same region.

```json
{
    "cloudwatch.emitMetrics": true,
    "kinesis.endpoint": "https://your/kinesis/endpoint",
    "firehose.endpoint": "https://your/firehose/endpoint",
    "flows": [
        {
            "filePattern": "/tmp/app1.log*",
            "kinesisStream": "yourKinesisStream"
        },
        {
            "filePattern": "/tmp/app2.log*",
            "deliveryStream": "yourfirehosedeliverystream"
        }
    ]
}
```
For more detailed information about using the agent with Amazon Kinesis Data Streams, see Writing to Amazon Kinesis Data Streams with Kinesis Agent.

Use the Agent to Pre-process Data

The agent can pre-process the records parsed from monitored files before sending them to your delivery stream. You can enable this feature by adding the `dataProcessingOptions` configuration setting to your file flow. One or more processing options can be added and they will be performed in the specified order.

The agent supports the following processing options. Because the agent is open-source, you can further develop and extend its processing options. You can download the agent from Kinesis Agent.

Processing Options

SINGLELINE

Converts a multi-line record to a single line record by removing newline characters, leading spaces, and trailing spaces.

```
{  
  "optionName": "SINGLELINE"
}
```

CSVTOJSON

Converts a record from delimiter separated format to JSON format.

```
{  
  "optionName": "CSVTOJSON",  
  "customFieldNames": [ "field1", "field2", ... ],  
  "delimiter": "yourdelimiter"
}
```

customFieldNames

[Required] The field names used as keys in each JSON key value pair. For example, if you specify ["f1", "f2"], the record "v1, v2" is converted to {"f1":"v1", "f2":"v2"}.

delimiter

The string used as the delimiter in the record. The default is a comma (,).

LOGTOJSON

Converts a record from a log format to JSON format. The supported log formats are Apache Common Log, Apache Combined Log, Apache Error Log, and RFC3164 Syslog.

```
{  
  "optionName": "LOGTOJSON",  
  "logFormat": "logformat",  
  "matchPattern": "yourregexpattern",  
  "customFieldNames": [ "field1", "field2", ... ]
}
```

logFormat

[Required] The log entry format. The following are possible values:

- COMMONAPACHELOG — The Apache Common Log format. Each log entry has the following pattern by default: "%{host} %{ident} %{authuser} [ %{datetime} ] " %{request} " %{response} %{bytes}".
Use the Agent to Pre-process Data

- **COMBINEDAPACHELOG** — The Apache Combined Log format. Each log entry has the following pattern by default: "\%{host} \%{ident} \%{authuser} \%{datetime}\n\"\%{request}\" \%{response} \%{bytes} \%{referrer} \%{agent}".
- **APACHEERRORLOG** — The Apache Error Log format. Each log entry has the following pattern by default: "\[%{timestamp}\] \[%{module}:\%{severity}\] \[pid \%{processid}:tid %{threadid}] \[client: \%{client}\] \%{message}".
- **SYSLOG** — The RFC3164 Syslog format. Each log entry has the following pattern by default: "\%{timestamp} \%{hostname} \%{program}\[%{processid}\]: %\{message\}".

**matchPattern**

Overrides the default pattern for the specified log format. Use this setting to extract values from log entries if they use a custom format. If you specify `matchPattern`, you must also specify `customFieldNames`.

**customFieldNames**

The custom field names used as keys in each JSON key value pair. You can use this setting to define field names for values extracted from `matchPattern`, or override the default field names of predefined log formats.

**Example : LOGTOJSON Configuration**

Here is one example of a `LOGTOJSON` configuration for an Apache Common Log entry converted to JSON format:

```json
{
  "optionName": "LOGTOJSON",
  "logFormat": "COMMONAPACHELOG"
}
```

**Before conversion:**

```
64.242.88.10 - - [07/Mar/2004:16:10:02 -0800] "GET /mailman/listinfo/hsdivision HTTP/1.1" 200 6291
```

**After conversion:**

```
{"host":"64.242.88.10","ident":null,"authuser":null,"datetime":"07/Mar/2004:16:10:02 -0800","request":"GET /mailman/listinfo/hsdivision HTTP/1.1","response":"200","bytes":"6291"}
```

**Example : LOGTOJSON Configuration With Custom Fields**

Here is another example `LOGTOJSON` configuration:

```json
{
  "optionName": "LOGTOJSON",
  "logFormat": "COMMONAPACHELOG",
  "customFieldNames": ["f1", "f2", "f3", "f4", "f5", "f6", "f7"]
}
```

With this configuration setting, the same Apache Common Log entry from the previous example is converted to JSON format as follows:

```
{"f1":"64.242.88.10","f2":null,"f3":null,"f4":"07/Mar/2004:16:10:02 -0800","f5":"GET /mailman/listinfo/hsdivision HTTP/1.1","f6":"200","f7":"6291"}
```
Example : Convert Apache Common Log Entry

The following flow configuration converts an Apache Common Log entry to a single line record in JSON format:

```json
{
    "flows": [
        {
            "filePattern": "/tmp/app.log*",
            "deliveryStream": "my-delivery-stream",
            "dataProcessingOptions": [
                {
                    "optionName": "LOGTOJSON",
                    "logFormat": "COMMONAPACHELOG"
                }
            ]
        }
    ]
}
```

Example : Convert Multi-Line Records

The following flow configuration parses multi-line records whose first line starts with "[ SEQUENCE=". Each record is first converted to a single line record. Then, values are extracted from the record based on a tab delimiter. Extracted values are mapped to specified customFieldNames values to form a single-line record in JSON format.

```json
{
    "flows": [
        {
            "filePattern": "/tmp/app.log*",
            "deliveryStream": "my-delivery-stream",
            "multiLineStartPattern": \"[SEQUENCE=\",
            "dataProcessingOptions": [
                {
                    "optionName": "SINGLELINE"
                },
                {
                    "optionName": "CSVTOJSON",
                    "customFieldNames": [ "field1", "field2", "field3" ],
                    "delimiter": \"\t\"
                }
            ]
        }
    ]
}
```

Example : LOGTOJSON Configuration with Match Pattern

Here is one example of a LOGTOJSON configuration for an Apache Common Log entry converted to JSON format, with the last field (bytes) omitted:

```json
{
    "optionName": "LOGTOJSON",
    "logFormat": "COMMONAPACHELOG",
    "matchPattern": "^(\\d.+) (\S+) (\S+) \[(\[\w:/]+\s[\+\-]\d{4})\] "(.+?)" (\d(3))",
    "customFieldNames": ["host", "ident", "authuser", "datetime", "request", "response"]
}
```

Before conversion:

"23"
Agent CLI Commands

Automatically start the agent on system startup:

```bash
sudo chkconfig aws-kinesis-agent on
```

Check the status of the agent:

```bash
sudo service aws-kinesis-agent status
```

Stop the agent:

```bash
sudo service aws-kinesis-agent stop
```

Read the agent's log file from this location:

```
/var/log/aws-kinesis-agent/aws-kinesis-agent.log
```

Uninstall the agent:

```bash
sudo yum remove aws-kinesis-agent
```

Writing to a Kinesis Data Delivery Stream Using the AWS SDK

You can use the Kinesis Data Firehose API to send data to a Kinesis data delivery stream using the AWS SDK for Java, .NET, Node.js, Python, or Ruby. If you are new to Kinesis Data Firehose, take some time to become familiar with the concepts and terminology presented in What Is Amazon Kinesis Data Firehose? (p. 1). For more information, see Start Developing with Amazon Web Services.

These examples do not represent production-ready code, in that they do not check for all possible exceptions, or account for all possible security or performance considerations.

The Kinesis Data Firehose API offers two operations for sending data to your Kinesis data delivery stream: `PutRecord` and `PutRecordBatch`. `PutRecord()` sends one data record within one call and `PutRecordBatch()` can send multiple data records within one call.

**Topics**

- Single Write Operations Using `PutRecord` (p. 25)
- Batch Write Operations Using `PutRecordBatch` (p. 25)
Single Write Operations Using PutRecord

Putting data requires only the Kinesis data delivery stream name and a byte buffer (\(<=1000\) KB). Because Kinesis Data Firehose batches multiple records before loading the file into Amazon S3, you may want to add a record separator. To put data one record at a time into a Kinesis data delivery stream, use the following code:

```java
PutRecordRequest putRecordRequest = new PutRecordRequest();
putRecordRequest.setDeliveryStreamName(deliveryStreamName);

String data = line + "\n";
Record record = createRecord(data);
putRecordRequest.setRecord(record);

// Put record into the DeliveryStream
firehoseClient.putRecord(putRecordRequest);
```

For more code context, see the sample code included in the AWS SDK. For information about request and response syntax, see the relevant topic in Amazon Kinesis Firehose API Operations.

Batch Write Operations Using PutRecordBatch

Putting data requires only the Kinesis data delivery stream name and a list of records. Because Kinesis Data Firehose batches multiple records before loading the file into Amazon S3, you may want to add a record separator. To put data records in batches into a Kinesis data delivery stream, use the following code:

```java
PutRecordBatchRequest putRecordBatchRequest = new PutRecordBatchRequest();
putRecordBatchRequest.setDeliveryStreamName(deliveryStreamName);
putRecordBatchRequest.setRecords(recordList);

// Put Record Batch records. Max No. Of Records we can put in a
// single put record batch request is 500
firehoseClient.putRecordBatch(putRecordBatchRequest);
recordList.clear();
```

For more code context, see the sample code included in the AWS SDK. For information about request and response syntax, see the relevant topic in Amazon Kinesis Firehose API Operations.

Writing to Amazon Kinesis Data Firehose Using CloudWatch Logs

For information on how to create a CloudWatch Logs subscription that sends log events to Kinesis Data Firehose, see Subscription Filters with Amazon Kinesis Firehose.

Writing to Amazon Kinesis Data Firehose Using CloudWatch Events

You can configure Amazon CloudWatch to send events to a Kinesis Data Firehose delivery stream by adding a target to a CloudWatch Events rule.
To create a target for a CloudWatch Events rule that sends events to an existing Kinesis Data Firehose delivery stream

1. When creating your rule, in the Step 1: Create Rule page, for Targets, choose Add target, and then choose Firehose delivery stream.
2. For Delivery stream, select an existing Kinesis Data Firehose delivery stream.

For more information on creating CloudWatch Events rules, see Getting Started with Amazon CloudWatch Events.

Writing to Amazon Kinesis Data Firehose Using AWS IoT

You can configure AWS IoT to send information to a Kinesis Data Firehose delivery stream by adding an action.

To create an action that sends events to an existing Kinesis Data Firehose delivery stream

1. When creating your rule, in the Set one or more actions page, choose Add action, and then choose Send messages to an Amazon Kinesis Data Firehose stream.
2. Choose Configure action.
3. For Stream name, select an existing Kinesis Data Firehose delivery stream.
4. For Separator, choose a separator character to be inserted between records.
5. For IAM role name, choose an existing IAM role or choose Create a new role.
6. Choose Add action.

For more information on creating AWS IoT rules, see AWS IoT Rule Tutorials.
Amazon Kinesis Data Firehose Data Transformation

Kinesis Data Firehose can invoke your Lambda function to transform incoming source data and deliver the transformed data to destinations. You can enable Kinesis Data Firehose data transformation when you create your delivery stream.

Data Transformation Flow

When you enable Kinesis Data Firehose data transformation, Kinesis Data Firehose buffers incoming data up to 3 MB by default. (To adjust the buffering size, use the ProcessingConfiguration API with the ProcessorParameter called BufferSizeInMBs.) Kinesis Data Firehose then invokes the specified Lambda function asynchronously with each buffered batch using the AWS Lambda synchronous invocation mode. The transformed data is sent from Lambda to Kinesis Data Firehose, which then sends it to the destination when the specified destination buffering size or buffering interval is reached, whichever happens first.

Important
The Lambda synchronous invocation mode has a payload size limit of 6 MB for both the request and the response. Make sure that your buffering size for sending the request to the function is less than or equal to 6 MB and that the response returned by your function doesn't exceed 6 MB, either.

Data Transformation and Status Model

All transformed records from Lambda must contain the following parameters or Kinesis Data Firehose rejects them and treats that as a data transformation failure.

recordId
The record ID is passed from Kinesis Data Firehose to Lambda during the invocation. The transformed record must contain the same record ID. Any mismatch between the ID of the original record and the ID of the transformed record is treated as a data transformation failure.

result
The status of the data transformation of the record. The possible values are: "Ok" (the record was transformed successfully), "Dropped" (the record was dropped intentionally by your processing logic), and "ProcessingFailed" (the record could not be transformed). If a record has a status of "Ok" or "Dropped", Kinesis Data Firehose considers it successfully processed. Otherwise, Kinesis Data Firehose considers it unsuccessfully processed.

data
The transformed data payload, after base64-encoding.

Lambda Blueprints

Kinesis Data Firehose provides the following Lambda blueprints that you can use to create a Lambda function for data transformation.
Data Transformation Failure Handling

If your Lambda function invocation fails because of a network timeout or because you’ve reached the Lambda invocation limit, Kinesis Data Firehose retries the invocation three times by default. If the invocation does not succeed, Kinesis Data Firehose then skips that batch of records. The skipped records are treated as unsuccessfully processed records. You can specify or override the retry options using the CreateDeliveryStream or UpdateDestination API. For this type of failure, you can log invocation errors to Amazon CloudWatch Logs. For more information, see Monitoring with Amazon CloudWatch Logs (p. 43).

If the status of the data transformation of a record is ProcessingFailed, Kinesis Data Firehose treats the record as unsuccessfully processed. For this type of failure, you can emit error logs to Amazon CloudWatch Logs from your Lambda function. For more information, see Accessing Amazon CloudWatch Logs for AWS Lambda in the AWS Lambda Developer Guide.

If data transformation fails, the unsuccessfully processed records are delivered to your S3 bucket in the processing_failed folder. The records have the following format:

```json
{
  "attemptsMade": "count",
  "arrivalTimestamp": "timestamp",
  "errorCode": "code",
  "errorMessage": "message",
  "attemptEndingTimestamp": "timestamp",
  "rawData": "data",
  "lambdaArn": "arn"
}
```

attemptsMade

   The number of invocation requests attempted.

arrivalTimestamp

   The time that the record was received by Kinesis Data Firehose.

errorCode

   The HTTP error code returned by Lambda.
errorMessage

The error message returned by Lambda.

attemptEndingTimestamp

The time that Kinesis Data Firehose stopped attempting Lambda invocations.

rawData

The base64-encoded record data.

lambdaArn

The Amazon Resource Name (ARN) of the Lambda function.

Source Record Backup

Kinesis Data Firehose can back up all untransformed records to your S3 bucket concurrently while delivering transformed records to the destination. You can enable source record backup when you create or update your delivery stream. You cannot disable source record backup after you enable it.
Amazon Kinesis Data Firehose can convert the format of your input data from JSON to Apache Parquet or Apache ORC before storing the data in Amazon S3. Parquet and ORC are columnar data formats that save space and enable faster queries compared to row-oriented formats like JSON. If you want to convert an input format other than JSON, such as comma-separated values (CSV) or structured text, you can use AWS Lambda to transform it to JSON first. For more information, see Data Transformation (p. 27).

Topics
• Record Format Conversion Requirements (p. 30)
• Choosing the JSON Deserializer (p. 30)
• Choosing the Serializer (p. 31)
• Converting Input Record Format (Console) (p. 31)
• Converting Input Record Format (API) (p. 32)
• Record Format Conversion Error Handling (p. 32)

Record Format Conversion Requirements

Kinesis Data Firehose requires the following three elements to convert the format of your record data:

• A deserializer to read the JSON of your input data – You can choose one of two types of deserializers: Apache Hive JSON SerDe or OpenX JSON SerDe.

• A schema to determine how to interpret that data – Use AWS Glue to create a schema in the AWS Glue Data Catalog. Kinesis Data Firehose then references that schema and uses it to interpret your input data. You can use the same schema to configure both Kinesis Data Firehose and your analytics software. For more information, see Populating the AWS Glue Data Catalog in the AWS Glue Developer Guide.

• A serializer to convert the data to the target columnar storage format (Parquet or ORC) – You can choose one of two types of serializers: ORC SerDe or Parquet SerDe.

Important
If you enable record format conversion, you can't set your Kinesis Data Firehose destination to be Amazon Elasticsearch Service (Amazon ES), Amazon Redshift, or Splunk. With format conversion enabled, Amazon S3 is the only destination that you can use for your Kinesis Data Firehose delivery stream.

You can convert the format of your data even if you aggregate your records before sending them to Kinesis Data Firehose.

Choosing the JSON Deserializer

Choose the OpenX JSON SerDe if your input JSON contains time stamps in the following formats:
Choosing the Serializer

The serializer that you choose depends on your business needs. To learn more about the two serializer options, see ORC SerDe and Parquet SerDe.

Converting Input Record Format (Console)

You can enable data format conversion on the console when you create or update a Kinesis delivery stream.

To enable data format conversion for a data delivery stream

1. Sign in to the AWS Management Console, and open the Kinesis Data Firehose console at https://console.aws.amazon.com/firehose/.
2. Choose a Kinesis Data Firehose delivery stream to update, or create a new delivery stream by following the steps in Creating an Amazon Kinesis Data Firehose Delivery Stream (p. 5).
3. On the Process records page, under Convert record format, set Record format conversion to Enabled, and follow the console instructions.
Converting Input Record Format (API)

If you want Kinesis Data Firehose to convert the format of your input data from JSON to Parquet or ORC, specify the optional `DataFormatConversionConfiguration` element in `ExtendedS3DestinationConfiguration` or in `ExtendedS3DestinationUpdate`. If you specify `DataFormatConversionConfiguration`, the following restrictions apply:

- In `BufferingHints`, you can't set `SizeInMBs` to a value less than 64 if you enable record format conversion. Also, when format conversion isn't enabled, the default value is 5, but it becomes 128 when you enable it.
- You must set `CompressionFormat` in `ExtendedS3DestinationConfiguration` or in `ExtendedS3DestinationUpdate` to `UNCOMPRESSED`. The default value for `CompressionFormat` is `UNCOMPRESSED`. Therefore, you can also leave it unspecified in `ExtendedS3DestinationConfiguration`. The data still gets compressed as part of the serialization process, using Snappy compression by default. When you configure the serializer, you can choose other types of compression.

Record Format Conversion Error Handling

When Kinesis Data Firehose can't parse or deserialize a record (for example, when the data doesn't match the schema), it writes it to Amazon S3 with an error prefix. If this write fails, Kinesis Data Firehose retries it forever, blocking further delivery. For each failed record, Kinesis Data Firehose writes a JSON document with the following schema:

```json
{
    "attemptsMade": long,
    "arrivalTimestamp": long,
    "lastErrorCode": string,
    "lastErrorMessage": string,
    "attemptEndingTimestamp": long,
    "rawData": string,
    "sequenceNumber": string,
    "subSequenceNumber": long,
    "dataCatalogTable": {
        "catalogId": string,
        "databaseName": string,
        "tableName": string,
        "region": string,
        "versionId": string,
        "catalogArn": string
    }
}
```
Amazon Kinesis Data Firehose Data Delivery

After data is sent to your delivery stream, it is automatically delivered to the destination you choose.

Topics
- Data Delivery Format (p. 33)
- Data Delivery Frequency (p. 33)
- Data Delivery Failure Handling (p. 34)
- Amazon S3 Object Name Format (p. 35)
- Index Rotation for the Amazon ES Destination (p. 36)

Data Delivery Format

For data delivery to Amazon S3, Kinesis Firehose concatenates multiple incoming records based on buffering configuration of your delivery stream, and then delivers them to Amazon S3 as an S3 object. You may want to add a record separator at the end of each record before you send it to Kinesis Firehose so that you can divide a delivered S3 object to individual records.

For data delivery to Amazon Redshift, Kinesis Firehose first delivers incoming data to your S3 bucket in the format described earlier. Kinesis Firehose then issues an Amazon Redshift COPY command to load the data from your S3 bucket to your Amazon Redshift cluster. You need to make sure that after Kinesis Firehose concatenates multiple incoming records to an S3 object, the S3 object can be copied to your Amazon Redshift cluster. For more information, see Amazon Redshift COPY Command Data Format Parameters.

For data delivery to Amazon ES, Kinesis Firehose buffers incoming records based on buffering configuration of your delivery stream. It then generates an Elasticsearch bulk request to index multiple records to your Elasticsearch cluster. You need to make sure that your record is UTF-8 encoded and flattened to a single-line JSON object before you send it to Kinesis Firehose. Also, the rest.action.multi.allow_explicit_index option for your Elasticsearch cluster must be set to true (default) in order to take bulk requests with an explicit index that is set per record. For more information, see Amazon ES Configure Advanced Options in the Amazon Elasticsearch Service Developer Guide.

For data delivery to Splunk, Kinesis Data Firehose concatenates the bytes that you send. If you want delimiters in your data, such as a new line character, you must insert them yourself. Make sure that Splunk is configured to parse any such delimiters.

Data Delivery Frequency

Each Kinesis Data Firehose destination has its own data delivery frequency.

Amazon S3

The frequency of data delivery to Amazon S3 is determined by the S3 Buffer size and Buffer interval value you configured for your delivery stream. Kinesis Data Firehose buffers incoming data
before delivering it to Amazon S3. You can configure the values for S3 **Buffer size** (1–128 MB) or **Buffer interval** (60–900 seconds), and the condition satisfied first triggers data delivery to Amazon S3. In circumstances where data delivery to the destination is falling behind data writing to the delivery stream, Kinesis Data Firehose raises the buffer size dynamically to catch up and make sure that all data is delivered to the destination.

**Amazon Redshift**

The frequency of data **COPY** operations from Amazon S3 to Amazon Redshift is determined by how fast your Amazon Redshift cluster can finish the **COPY** command. If there is still data to copy, Kinesis Data Firehose issues a new **COPY** command as soon as the previous **COPY** command is successfully finished by Amazon Redshift.

**Amazon Elasticsearch Service**

The frequency of data delivery to Amazon ES is determined by the Elasticsearch **Buffer size** and **Buffer interval** values that you configured for your delivery stream. Kinesis Data Firehose buffers incoming data before delivering it to Amazon ES. You can configure the values for Elasticsearch **Buffer size** (1–100 MB) or **Buffer interval** (60–900 seconds), and the condition satisfied first triggers data delivery to Amazon ES.

**Splunk**

Kinesis Data Firehose buffers incoming data before delivering it to Splunk. The buffer size is 5 MB and the buffer interval is 60 seconds. The condition satisfied first triggers data delivery to Splunk. The buffer size and interval aren't configurable. These numbers are optimal.

## Data Delivery Failure Handling

Each Kinesis Data Firehose destination has its own data delivery failure handling.

**Amazon S3**

Data delivery to your S3 bucket might fail for reasons such as the bucket doesn't exist anymore, the IAM role that Kinesis Firehose assumes doesn't have access to the bucket, the network failed, or similar events. Under these conditions, Kinesis Firehose keeps retrying for up to 24 hours until the delivery succeeds. The maximum data storage time of Kinesis Firehose is 24 hours, and your data is lost if data delivery fails for more than 24 hours.

**Amazon Redshift**

For the Amazon Redshift destination, you can specify a retry duration (0–7200 seconds) when creating a delivery stream.

Data delivery to your Amazon Redshift cluster might fail for reasons such as an incorrect Amazon Redshift cluster configuration of your delivery stream, an Amazon Redshift cluster under maintenance, network failure, or similar events. Under these conditions, Kinesis Data Firehose retries for the specified time duration and skips that particular batch of S3 objects. The skipped objects' information is delivered to your S3 bucket as a manifest file in the `errors/` folder, which you can use for manual backfill. For information about how to COPY data manually with manifest files, see Using a Manifest to Specify Data Files.

**Amazon Elasticsearch Service**

For the Amazon ES destination, you can specify a retry duration (0–7200 seconds) when creating a delivery stream.

Data delivery to your Amazon ES cluster might fail for reasons such as an incorrect Amazon ES cluster configuration of your delivery stream, an Amazon ES cluster under maintenance, network failure, or similar events. Under these conditions, Kinesis Data Firehose retries for the specified time duration and skips that particular batch of S3 objects. The skipped objects' information is delivered to your S3 bucket as a manifest file in the `errors/` folder, which you can use for manual backfill. For information about how to COPY data manually with manifest files, see Using a Manifest to Specify Data Files.
duration and then skips that particular index request. The skipped documents are delivered to your S3 bucket in the `elasticsearch_failed/` folder, which you can use for manual backfill. Each document has the following JSON format:

```json
{
    "attemptsMade": "(number of index requests attempted)",
    "arrivalTimestamp": "(the time when the document was received by Firehose)",
    "errorCode": "(http error code returned by Elasticsearch)",
    "errorMessage": "(error message returned by Elasticsearch)",
    "attemptEndingTimestamp": "(the time when Firehose stopped attempting index request)",
    "esDocumentId": "(intended Elasticsearch document ID)",
    "esIndexName": "(intended Elasticsearch index name)",
    "esTypeName": "(intended Elasticsearch type name)",
    "rawData": "(base64-encoded document data)"
}
```

**Splunk**

When Kinesis Data Firehose sends data to Splunk, it waits for acknowledgment from Splunk. If an error occurs or the acknowledgment doesn't arrive within the acknowledgment timeout period, Kinesis Firehose starts the retry duration counter and keeps retrying until the retry duration expires, after which Kinesis Firehose considers it a data delivery failure and backs up the data to your Amazon S3 bucket.

Every time Kinesis Firehose sends data to Splunk, whether it's the initial attempt or a retry, it restarts the acknowledgement timeout counter and waits for an acknowledgement to arrive from Splunk. Even if the retry duration expires, Kinesis Firehose still waits for the acknowledgment until it receives it or the acknowledgement timeout is reached. If the acknowledgment times out, Kinesis Firehose checks to determine whether there's time left in the retry counter. If there is time left, it retries again and repeats the logic until it receives an acknowledgement or determines that the retry time has expired.

A failure to receive an acknowledgement isn't the only kind of data delivery error that can occur. For information about the other types of data delivery errors that you might encounter, see [Splunk Data Delivery Errors](#). Any data delivery error triggers the retry logic if your retry duration is greater than 0.

The following is an example error record:

```json
{
    "attemptsMade": 0,
    "arrivalTimestamp": 1506035354675,
    "errorCode": "Splunk.AckTimeout",
    "errorMessage": "Did not receive an acknowledgement from HEC before the HEC acknowledgement timeout expired. Despite the acknowledgement timeout, it's possible the data was indexed successfully in Splunk. Kinesis Firehose backs up in Amazon S3 data for which the acknowledgement timeout expired.",
    "attemptEndingTimestamp": 13626284715507,
    "rawData": "MiAyNTE2MjAyNzIyMDkgZW5pLTA1ZjMyMmQ1IDIxOC45Mi4xODguMjE0IDE3Mi4xNi4xLjE2NyAyNTIzMyAxNDMzIDYgMSA0MCAxNTA2MDM0NzM0IDE1MDYwMzQ3OTQgUkVKRUNUIE9LCg==",
    "EventId": "4957719392811417339600778471082492393164139877200035842.0"
}
```

### Amazon S3 Object Name Format

Kinesis Data Firehose adds a UTC time prefix in the format `YYYY/MM/DD/HH` before writing objects to Amazon S3. This prefix creates a logical hierarchy in the bucket, where each forward slash (`/`) creates a level in the hierarchy. You can modify this structure by adding to the start of the prefix when you create...
Index Rotation for the Amazon ES Destination

For the Amazon ES destination, you can specify a time-based index rotation option from one of the following five options: **NoRotation, OneHour, OneDay, OneWeek, or OneMonth**.

Depending on the rotation option you choose, Kinesis Data Firehose appends a portion of the UTC arrival time stamp to your specified index name, and rotates the appended time stamp accordingly. The following example shows the resulting index name in Amazon ES for each index rotation option, where the specified index name is `myindex` and the arrival time stamp is 2016-02-25T13:00:00Z.

<table>
<thead>
<tr>
<th>RotationPeriod</th>
<th>IndexName</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoRotation</td>
<td>myindex</td>
</tr>
<tr>
<td>OneHour</td>
<td>myindex-2016-02-25-13</td>
</tr>
<tr>
<td>OneDay</td>
<td>myindex-2016-02-25</td>
</tr>
<tr>
<td>OneWeek</td>
<td>myindex-2016-w08</td>
</tr>
<tr>
<td>OneMonth</td>
<td>myindex-2016-02</td>
</tr>
</tbody>
</table>
Using Server-Side Encryption with Amazon Kinesis Data Firehose

If you have sensitive data, you can enable server-side data encryption when you use Amazon Kinesis Data Firehose. However, this is only possible if you use a Kinesis stream as your data source. When you configure a Kinesis stream as the data source of a Kinesis Data Firehose delivery stream, Kinesis Data Firehose no longer stores the data at rest. Instead, the data is stored in the Kinesis stream.

When you send data from your data producers to your Kinesis stream, the Kinesis Data Streams service encrypts your data using an AWS KMS key before storing it at rest. When your Kinesis Data Firehose delivery stream reads the data from your Kinesis stream, the Kinesis Data Streams service first decrypts the data and then sends it to Kinesis Data Firehose. Kinesis Data Firehose buffers the data in memory based on the buffering hints that you specify and then delivers it to your destinations without storing the unencrypted data at rest.

For information about how to enable server-side encryption for Kinesis Data Streams, see Using Server-Side Encryption.
Monitoring Amazon Kinesis Data Firehose

You can monitor Amazon Kinesis Data Firehose using the following features:

- **CloudWatch metrics (p. 38)**— Kinesis Data Firehose sends Amazon CloudWatch custom metrics with detailed monitoring for each delivery stream.
- **CloudWatch Logs (p. 43)**— Kinesis Data Firehose sends CloudWatch custom logs with detailed monitoring for each delivery stream.
- **Kinesis Agent (p. 48)**— Kinesis Agent publishes custom CloudWatch metrics to help assess whether the agent is working as expected.
- **API logging and history (p. 49)**— Kinesis Data Firehose uses AWS CloudTrail to log API calls and store the data in an Amazon S3 bucket, and to maintain API call history.

Monitoring with Amazon CloudWatch Metrics

Kinesis Data Firehose integrates with CloudWatch metrics so that you can collect, view, and analyze CloudWatch metrics for your Kinesis data delivery streams. For example, you can monitor the IncomingBytes and IncomingRecords metrics to keep track of data ingested into Kinesis Data Firehose from data producers.

The metrics that you configure for your Kinesis data delivery streams and agents are automatically collected and pushed to CloudWatch every five minutes. Metrics are archived for two weeks; after that period, the data is discarded.

The metrics collected for Kinesis data delivery streams are free of charge. For information about Kinesis agent metrics, see Monitoring Kinesis Agent Health (p. 48).

Topics

- Service-level CloudWatch Metrics (p. 38)
- API-Level CloudWatch Metrics (p. 40)
- Data Transformation CloudWatch Metrics (p. 41)
- Format Conversion CloudWatch Metrics (p. 42)
- Dimensions for Kinesis Data Firehose (p. 42)
- Accessing CloudWatch Metrics for Kinesis Data Firehose (p. 42)

Service-level CloudWatch Metrics

The AWS/Firehose namespace includes the following service-level metrics.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BackupToS3.Bytes</td>
<td>The number of bytes delivered to Amazon S3 for backup over the specified time period. Kinesis Data Firehose emits this metric when data transformation is enabled for Amazon S3 or Amazon Redshift destinations. Units: Bytes</td>
</tr>
<tr>
<td>Metric</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>BackupToS3.DataFreshness</td>
<td>Age (from getting into Kinesis Data Firehose to now) of the oldest record in Kinesis Data Firehose. Any record older than this age has been delivered to the Amazon S3 bucket for backup. Kinesis Data Firehose emits this metric when data transformation is enabled for Amazon S3 or Amazon Redshift destinations. Units: Seconds</td>
</tr>
<tr>
<td>BackupToS3.Records</td>
<td>The number of records delivered to Amazon S3 for backup over the specified time period. Kinesis Data Firehose emits this metric when data transformation is enabled for Amazon S3 or Amazon Redshift destinations. Units: Count</td>
</tr>
<tr>
<td>BackupToS3.Success</td>
<td>Sum of successful Amazon S3 put commands for backup over sum of all Amazon S3 backup put commands. Kinesis Data Firehose emits this metric when data transformation is enabled for Amazon S3 or Amazon Redshift destinations.</td>
</tr>
<tr>
<td>DeliveryToElasticsearch.Bytes</td>
<td>The number of bytes indexed to Amazon ES over the specified time period. Units: Bytes</td>
</tr>
<tr>
<td>DeliveryToElasticsearch.Records</td>
<td>The number of records indexed to Amazon ES over the specified time period. Units: Count</td>
</tr>
<tr>
<td>DeliveryToElasticsearch.Success</td>
<td>The sum of the successfully indexed records over the sum of records that were attempted.</td>
</tr>
<tr>
<td>DeliveryToRedshift.Bytes</td>
<td>The number of bytes copied to Amazon Redshift over the specified time period. Units: Bytes</td>
</tr>
<tr>
<td>DeliveryToRedshift.Records</td>
<td>The number of records copied to Amazon Redshift over the specified time period. Units: Count</td>
</tr>
<tr>
<td>DeliveryToRedshift.Success</td>
<td>The sum of successful Amazon Redshift COPY commands over the sum of all Amazon Redshift COPY commands.</td>
</tr>
<tr>
<td>DeliveryToS3.Bytes</td>
<td>The number of bytes delivered to Amazon S3 over the specified time period. Units: Bytes</td>
</tr>
<tr>
<td>DeliveryToS3.DataFreshness</td>
<td>The age (from getting into Kinesis Data Firehose to now) of the oldest record in Kinesis Data Firehose. Any record older than this age has been delivered to the S3 bucket. Units: Seconds</td>
</tr>
</tbody>
</table>
### API-Level CloudWatch Metrics

The AWS/Firehose namespace includes the following API-level metrics.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DescribeDeliveryStream.Latency</td>
<td>The time taken per DescribeDeliveryStream operation, measured over the specified time period.</td>
</tr>
<tr>
<td></td>
<td>Units: Milliseconds</td>
</tr>
<tr>
<td>DescribeDeliveryStream.Requests</td>
<td>The total number of DescribeDeliveryStream requests.</td>
</tr>
<tr>
<td></td>
<td>Units: Count</td>
</tr>
<tr>
<td>ListDeliveryStreams.Latency</td>
<td>The time taken per ListDeliveryStream operation, measured over the specified time period.</td>
</tr>
<tr>
<td></td>
<td>Units: Milliseconds</td>
</tr>
<tr>
<td>Metric</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ListDeliveryStreams.Requests</td>
<td>The total number of ListFirehose requests.</td>
</tr>
<tr>
<td></td>
<td>Units: Count</td>
</tr>
<tr>
<td>PutRecord.Bytes</td>
<td>The number of bytes put to the Kinesis Data Firehose delivery stream using</td>
</tr>
<tr>
<td></td>
<td>PutRecord over the specified time period.</td>
</tr>
<tr>
<td></td>
<td>Units: Bytes</td>
</tr>
<tr>
<td>PutRecord.Latency</td>
<td>The time taken per PutRecord operation, measured over the specified time</td>
</tr>
<tr>
<td></td>
<td>period.</td>
</tr>
<tr>
<td></td>
<td>Units: Milliseconds</td>
</tr>
<tr>
<td>PutRecord.Requests</td>
<td>The total number of PutRecord requests, which is equal to total number of</td>
</tr>
<tr>
<td></td>
<td>records from PutRecord operations.</td>
</tr>
<tr>
<td></td>
<td>Units: Count</td>
</tr>
<tr>
<td>PutRecordBatch.Bytes</td>
<td>The number of bytes put to the Kinesis Data Firehose delivery stream using</td>
</tr>
<tr>
<td></td>
<td>PutRecordBatch over the specified time period.</td>
</tr>
<tr>
<td></td>
<td>Units: Bytes</td>
</tr>
<tr>
<td>PutRecordBatch.Latency</td>
<td>The time taken per PutRecordBatch operation, measured over the specified</td>
</tr>
<tr>
<td></td>
<td>time period.</td>
</tr>
<tr>
<td></td>
<td>Units: Milliseconds</td>
</tr>
<tr>
<td>PutRecordBatch.Records</td>
<td>The total number of records from PutRecordBatch operations.</td>
</tr>
<tr>
<td></td>
<td>Units: Count</td>
</tr>
<tr>
<td>PutRecordBatch.Requests</td>
<td>The total number of PutRecordBatch requests.</td>
</tr>
<tr>
<td></td>
<td>Units: Count</td>
</tr>
<tr>
<td>UpdateDeliveryStream.Latency</td>
<td>The time taken per UpdateDeliveryStream operation, measured over the</td>
</tr>
<tr>
<td></td>
<td>specified time period.</td>
</tr>
<tr>
<td></td>
<td>Units: Milliseconds</td>
</tr>
<tr>
<td>UpdateDeliveryStream.Requests</td>
<td>The total number of UpdateDeliveryStream requests.</td>
</tr>
<tr>
<td></td>
<td>Units: Count</td>
</tr>
</tbody>
</table>

**Data Transformation CloudWatch Metrics**

If data transformation with Lambda is enabled, the AWS/Firehose namespace includes the following metrics.
Format Conversion CloudWatch Metrics

If format conversion is enabled, the AWS/Firehose namespace includes the following metrics.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SucceedConversion.Records</td>
<td>The number of successfully converted records.</td>
</tr>
<tr>
<td>Units: Count</td>
<td></td>
</tr>
<tr>
<td>SucceedConversion.Bytes</td>
<td>The size of the successfully converted records.</td>
</tr>
<tr>
<td>Units: Bytes</td>
<td></td>
</tr>
<tr>
<td>FailedConversion.Records</td>
<td>The number of records that could not be converted.</td>
</tr>
<tr>
<td>Units: Count</td>
<td></td>
</tr>
<tr>
<td>FailedConversion.Bytes</td>
<td>The size of the records that could not be converted.</td>
</tr>
<tr>
<td>Units: Bytes</td>
<td></td>
</tr>
</tbody>
</table>

Dimensions for Kinesis Data Firehose

To filter metrics by delivery stream, use the DeliveryStreamName dimension.

Accessing CloudWatch Metrics for Kinesis Data Firehose

You can monitor metrics for Kinesis Data Firehose using the CloudWatch console, command line, or CloudWatch API. The following procedures show you how to access metrics using these different methods.

To access metrics using the CloudWatch console

Monitoring with CloudWatch Logs

Kinesis Data Firehose integrates with CloudWatch Logs so that you can view the specific error logs when the Lambda invocation for data transformation or data delivery fails. You can enable Kinesis Data Firehose error logging when you create your delivery stream.

If you enable Kinesis Data Firehose error logging in the Kinesis Data Firehose console, a log group and corresponding log streams are created for the delivery stream on your behalf. The format of the log group name is /aws/kinesisfirehose/delivery-stream-name, where delivery-stream-name is the name of the corresponding delivery stream. The log stream name is S3Delivery, RedshiftDelivery, or ElasticsearchDelivery, depending on the delivery destination. Lambda invocation errors for data transformation are also logged to the log stream used for data delivery errors.

For example, if you create a delivery stream "MyStream" with Amazon Redshift as the destination and enable Kinesis Data Firehose error logging, the following are created on your behalf: a log group named aws/kinesisfirehose/MyStream and two log streams named S3Delivery and RedshiftDelivery. In this example, the S3Delivery log stream is used for logging errors related to delivery failure to the intermediate S3 bucket, and the RedshiftDelivery log stream is used for logging errors related to Lambda invocation failure and delivery failure to your Amazon Redshift cluster.

If you enable Kinesis Data Firehose error logging through the AWS CLI or an AWS SDK using the CloudWatchLoggingOptions configuration, you must create a log group and a log stream in advance. We recommend reserving that log group and log stream for Kinesis Data Firehose error logging exclusively. Also ensure that the associated IAM policy has "logs:putLogEvents" permission. For more information, see Controlling Access with Amazon Kinesis Data Firehose (p. 55).

Note that Kinesis Data Firehose does not guarantee that all delivery error logs are sent to CloudWatch Logs. In circumstances where delivery failure rate is high, Kinesis Data Firehose samples delivery error logs before sending them to CloudWatch Logs.

There is a nominal charge for error logs sent to CloudWatch Logs. For more information, see Amazon CloudWatch Pricing.

Contents
- Data Delivery Errors (p. 44)
- Lambda Invocation Errors (p. 47)
- Accessing CloudWatch Logs for Kinesis Data Firehose (p. 48)
Data Delivery Errors

The following is a list of data delivery error codes and messages for each Kinesis Data Firehose destination. Each error message also describes the proper action to take to fix the issue.

Errors

- Amazon S3 Data Delivery Errors (p. 44)
- Amazon Redshift Data Delivery Errors (p. 44)
- Splunk Data Delivery Errors (p. 46)
- Amazon Elasticsearch Service Data Delivery Errors (p. 47)

Amazon S3 Data Delivery Errors

Kinesis Data Firehose can send the following Amazon S3-related errors to CloudWatch Logs.

<table>
<thead>
<tr>
<th>Error Code</th>
<th>Error Message and Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3.KMS.NotFoundException</td>
<td>&quot;The provided AWS KMS key was not found. If you are using what you believe to be a valid AWS KMS key with the correct role, check if there is a problem with the account to which the AWS KMS key is attached.&quot;</td>
</tr>
<tr>
<td>S3.KMS.RequestLimitExceeded</td>
<td>&quot;The KMS request per second limit was exceeded while attempting to encrypt S3 objects. Increase the request per second limit.&quot; For more information, see Limits in the AWS Key Management Service Developer Guide.</td>
</tr>
<tr>
<td>S3.AccessDenied</td>
<td>&quot;Access was denied. Ensure that the trust policy for the provided IAM role allows Kinesis Data Firehose to assume the role, and the access policy allows access to the S3 bucket.&quot;</td>
</tr>
<tr>
<td>S3.AccountProblem</td>
<td>&quot;There is a problem with your AWS account that prevents the operation from completing successfully. Contact AWS Support.&quot;</td>
</tr>
<tr>
<td>S3.AllAccessDisabled</td>
<td>&quot;Access to the account provided has been disabled. Contact AWS Support.&quot;</td>
</tr>
<tr>
<td>S3.InvalidPayer</td>
<td>&quot;Access to the account provided has been disabled. Contact AWS Support.&quot;</td>
</tr>
<tr>
<td>S3.NotSignedUp</td>
<td>&quot;The account is not signed up for Amazon S3. Sign the account up or use a different account.&quot;</td>
</tr>
<tr>
<td>S3.NoSuchBucket</td>
<td>&quot;The specified bucket does not exist. Create the bucket or use a different bucket that does exist.&quot;</td>
</tr>
<tr>
<td>S3.MethodNotAllowed</td>
<td>&quot;The specified method is not allowed against this resource. Modify the bucket's policy to allow the correct Amazon S3 operation permissions.&quot;</td>
</tr>
<tr>
<td>InternalError</td>
<td>&quot;An internal error occurred while attempting to deliver data. Delivery will be retried; if the error persists, then it will be reported to AWS for resolution.&quot;</td>
</tr>
</tbody>
</table>

Amazon Redshift Data Delivery Errors

Kinesis Data Firehose can send the following Amazon Redshift-related errors to CloudWatch Logs.
<table>
<thead>
<tr>
<th>Error Code</th>
<th>Error Message and Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redshift.TableNotFound</td>
<td>The destination table in Amazon Redshift to which data should be copied from S3 was not found. Note that Kinesis Data Firehose does not create the Amazon Redshift table if it does not exist.</td>
</tr>
<tr>
<td>Redshift.SyntaxError</td>
<td>The COPY command contains a syntax error. Retry the command.</td>
</tr>
<tr>
<td>Redshift.Authentication</td>
<td>The provided user name and password failed authentication. Provide a valid user name and password.</td>
</tr>
<tr>
<td>Redshift.AccessDenied</td>
<td>Access was denied. Ensure that the trust policy for the provided IAM role allows Kinesis Data Firehose to assume the role.</td>
</tr>
<tr>
<td>Redshift.S3BucketAccessDenied</td>
<td>The COPY command was unable to access the S3 bucket. Ensure that the access policy for the provided IAM role allows access to the S3 bucket.</td>
</tr>
<tr>
<td>Redshift.DataLoadFailed</td>
<td>Loading data into the table failed. Check STL_LOAD_ERRORS system table for details.</td>
</tr>
<tr>
<td>Redshift.ColumnNotFound</td>
<td>A column in the COPY command does not exist in the table. Specify a valid column name.</td>
</tr>
<tr>
<td>Redshift.DatabaseNotFound</td>
<td>The database specified in the Amazon Redshift destination configuration or JDBC URL was not found. Specify a valid database name.</td>
</tr>
<tr>
<td>Redshift.IncorrectCopy</td>
<td>Conflicting or redundant COPY options were provided. Some options are not compatible in certain combinations. Check the COPY command reference for more info.</td>
</tr>
<tr>
<td></td>
<td>For more information, see the Amazon Redshift COPY command in the Amazon Redshift Database Developer Guide.</td>
</tr>
<tr>
<td>Redshift.ConnectionFailed</td>
<td>The connection to the specified Amazon Redshift cluster failed. Ensure that security settings allow Kinesis Data Firehose connections, that the cluster or database specified in the Amazon Redshift destination configuration or JDBC URL is correct, and that the cluster is available.</td>
</tr>
<tr>
<td>Redshift.ColumnMismatch</td>
<td>The number of jsonpaths in the COPY command and the number of columns in the destination table should match. Retry the command.</td>
</tr>
<tr>
<td>Redshift.IncorrectRegion</td>
<td>Amazon Redshift attempted to use the wrong region endpoint for accessing the S3 bucket. Either specify a correct region value in the COPY command options or ensure that the S3 bucket is in the same region as the Amazon Redshift database.</td>
</tr>
<tr>
<td>Redshift.IncorrectJsonPaths</td>
<td>The provided jsonpaths file is not in a supported JSON format. Retry the command.</td>
</tr>
<tr>
<td>Redshift.MissingS3File</td>
<td>One or more S3 files required by Amazon Redshift have been removed from the S3 bucket. Check the S3 bucket policies to remove any automatic deletion of S3 files.</td>
</tr>
</tbody>
</table>
### Amazon Kinesis Firehose Developer Guide

#### Data Delivery Errors

<table>
<thead>
<tr>
<th>Error Code</th>
<th>Error Message and Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redshift.InsufficientPrivilege</td>
<td>&quot;The user does not have permissions to load data into the table. Check the Amazon Redshift user permissions for the INSERT privilege.&quot;</td>
</tr>
<tr>
<td>Redshift.ReadOnlyCluster</td>
<td>&quot;The query cannot be executed because the system is in resize mode. Try the query again later.&quot;</td>
</tr>
<tr>
<td>Redshift.DiskFull</td>
<td>&quot;Data could not be loaded because the disk is full. Increase the capacity of the Amazon Redshift cluster or delete unused data to free disk space.&quot;</td>
</tr>
<tr>
<td>InternalError</td>
<td>&quot;An internal error occurred while attempting to deliver data. Delivery will be retried; if the error persists, then it will be reported to AWS for resolution.&quot;</td>
</tr>
</tbody>
</table>

### Splunk Data Delivery Errors

Kinesis Data Firehose can send the following Splunk-related errors to CloudWatch Logs.

<table>
<thead>
<tr>
<th>Error Code</th>
<th>Error Message and Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Splunk.ProxyWithoutStickySessions</td>
<td>&quot;If you have a proxy (ELB or other) between Kinesis Data Firehose and the HEC node, you must enable sticky sessions to support HEC ACKs.&quot;</td>
</tr>
<tr>
<td>Splunk.DisabledToken</td>
<td>&quot;The HEC token is disabled. Enable the token to allow data delivery to Splunk.&quot;</td>
</tr>
<tr>
<td>Splunk.InvalidToken</td>
<td>&quot;The HEC token is invalid. Update Kinesis Data Firehose with a valid HEC token.&quot;</td>
</tr>
<tr>
<td>Splunk.InvalidDataFormat</td>
<td>&quot;The data is not formatted correctly. To see how to properly format data for Raw or Event HEC endpoints, see Splunk Event Data.&quot;</td>
</tr>
<tr>
<td>Splunk.InvalidIndex</td>
<td>&quot;The HEC token or input is configured with an invalid index. Check your index configuration and try again.&quot;</td>
</tr>
<tr>
<td>Splunk.ServerError</td>
<td>&quot;Data delivery to Splunk failed due to a server error from the HEC node. Kinesis Data Firehose will retry sending the data if the retry duration in your Kinesis Data Firehose is greater than 0. If all the retries fail, Kinesis Data Firehose backs up the data to Amazon S3.&quot;</td>
</tr>
<tr>
<td>Splunk.DisabledAck</td>
<td>&quot;Indexer acknowledgement is disabled for the HEC token. Enable indexer acknowledgement and try again. For more info, see Enable indexer acknowledgement.&quot;</td>
</tr>
<tr>
<td>Splunk.AckTimeout</td>
<td>&quot;Did not receive an acknowledgement from HEC before the HEC acknowledgement timeout expired. Despite the acknowledgement timeout, it's possible the data was indexed successfully in Splunk. Kinesis Data Firehose backs up in Amazon S3 data for which the acknowledgement timeout expired.&quot;</td>
</tr>
<tr>
<td>Splunk.MaxRetriesFailed</td>
<td>&quot;Failed to deliver data to Splunk or to receive acknowledgment. Check your HEC health and try again.&quot;</td>
</tr>
<tr>
<td>Splunk.ConnectionTimeout</td>
<td>&quot;The connection to Splunk timed out. This might be a transient error and the request will be retried. Kinesis Data Firehose backs up the data to Amazon S3 if all retries fail.&quot;</td>
</tr>
</tbody>
</table>
Error Code | Error Message and Information
--- | ---
Splunk.InvalidEndpoint | "Could not connect to the HEC endpoint. Make sure that the HEC endpoint URL is valid and reachable from Kinesis Data Firehose."

Lambda.ConnectionClosed | "Unable to send data to Splunk due to a connection failure. This might be a transient error. Increasing the retry duration in your Kinesis Data Firehose configuration might guard against such transient failures."

Splunk.ConnectionClosed | "Could not connect to the HEC endpoint. Make sure that the certificate provided by the peer. Make sure that the certificate and the host are valid."

Splunk.SSLUnverified | "Could not connect to the HEC endpoint. The host does not match the certificate provided by the peer. Make sure that the certificate and the host are valid."

Splunk.SSLHandshake | "Could not connect to the HEC endpoint. Make sure that the certificate and the host are valid."

Amazon Elasticsearch Service Data Delivery Errors

For the Amazon ES destination, Kinesis Firehose sends errors to CloudWatch Logs as they are returned by Elasticsearch.

Lambda Invocation Errors

Kinesis Data Firehose can send the following Lambda invocation errors to CloudWatch Logs.

Error Code | Error Message and Information
--- | ---
Lambda.AssumeRoleAccessDenied | "Access was denied. Ensure that the trust policy for the provided IAM role allows Kinesis Data Firehose to assume the role."

Lambda.InvokeAccessDenied | "Access was denied. Ensure that the access policy allows access to the Lambda function."

Lambda.JsonProcessingException | "There was an error parsing returned records from the Lambda function. Ensure that the returned records follow the status model required by Kinesis Data Firehose."

Lambda.JsonProcessingException | For more information, see Data Transformation and Status Model (p. 27).

Lambda.InvokeLimitExceeded | "The Lambda concurrent execution limit is exceeded. Increase the concurrent execution limit."

Lambda.InvokeLimitExceeded | For more information, see AWS Lambda Limits in the AWS Lambda Developer Guide.

Lambda.DuplicatedRecordId | "Multiple records were returned with the same record ID. Ensure that the Lambda function returns unique record IDs for each record."

Lambda.DuplicatedRecordId | For more information, see Data Transformation and Status Model (p. 27).

Lambda.MissingRecordId | "One or more record IDs were not returned. Ensure that the Lambda function returns all received record IDs."

Lambda.MissingRecordId | For more information, see Data Transformation and Status Model (p. 27).

Lambda.ResourceNotFoundException | "The specified Lambda function does not exist. Use a different function that does exist."
You can view the error logs related to Kinesis Firehose data delivery failure using the Kinesis Firehose console or CloudWatch console. The following procedures show you how to access error logs using these two methods.

**To access error logs using the Kinesis Data Firehose console**

1. Open the Kinesis Firehose console at https://console.aws.amazon.com/firehose/.
2. On the navigation bar, choose a region.
3. Select a delivery stream name to go to the delivery stream details page.
4. Choose Error Log to view a list of error logs related to data delivery failure.

**To access error logs using the CloudWatch console**

2. On the navigation bar, choose a region.
3. In the navigation pane, choose Logs.
4. Select a log group and log stream to view a list of error logs related to data delivery failure.

**Monitoring Kinesis Agent Health**

Kinesis Agent publishes custom CloudWatch metrics with a namespace of AWSKinesisAgent to help assess if the agent is healthy, submitting data into Kinesis Data Firehose as specified, and consuming the appropriate amount of CPU and memory resources on the data producer.

Metrics such as number of records and bytes sent are useful to understand the rate at which the agent is submitting data to the Kinesis data delivery stream. When these metrics fall below expected thresholds
by some percentage or drop to zero, it could indicate configuration issues, network errors, or agent health issues. Metrics such as on-host CPU and memory consumption and agent error counters indicate data producer resource usage, and provide insights into potential configuration or host errors. Finally, the agent also logs service exceptions to help investigate agent issues.

The agent metrics are reported in the region specified in the agent configuration setting `cloudwatch.endpoint`. For more information, see Agent Configuration Settings (p. 18).

There is a nominal charge for metrics emitted from Kinesis Agent, which are enabled by default. For more information, see Amazon CloudWatch Pricing.

### Monitoring with CloudWatch

Kinesis Agent sends the following metrics to CloudWatch.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BytesSent</td>
<td>The number of bytes sent to the Kinesis Data Firehose delivery stream over the specified time period.</td>
</tr>
<tr>
<td></td>
<td>Units: Bytes</td>
</tr>
<tr>
<td>RecordSendAttempts</td>
<td>The number of records attempted (either first time, or as a retry) in a call to <code>PutRecordBatch</code> over the specified time period.</td>
</tr>
<tr>
<td></td>
<td>Units: Count</td>
</tr>
<tr>
<td>RecordSendErrors</td>
<td>The number of records that returned failure status in a call to <code>PutRecordBatch</code>, including retries, over the specified time period.</td>
</tr>
<tr>
<td></td>
<td>Units: Count</td>
</tr>
<tr>
<td>ServiceErrors</td>
<td>The number of calls to <code>PutRecordBatch</code> that resulted in a service error (other than a throttling error) over the specified time period.</td>
</tr>
<tr>
<td></td>
<td>Units: Count</td>
</tr>
</tbody>
</table>

### Monitoring Amazon Kinesis Data Firehose API Calls with AWS CloudTrail

Amazon Kinesis Data Firehose is integrated with AWS CloudTrail, which captures API calls, delivers the log files to an Amazon S3 bucket that you specify, and maintains API call history. CloudTrail captures API calls made from the Kinesis Data Firehose console or from your code to the Kinesis Data Firehose API. With the information collected by CloudTrail, you can determine the request that was made to Kinesis Data Firehose, the IP address from which the request was made, who made the request, when it was made, and so on.

To learn more about CloudTrail, including how to configure and enable it, see the [AWS CloudTrail User Guide](https://docs.aws.amazon.com/lambda/latest/dg/configuration-cloudtrail.html).

#### Topics

- [Kinesis Data Firehose and CloudTrail History (p. 50)](https://docs.aws.amazon.com/lambda/latest/dg/configuration-cloudtrail.html)
- [Kinesis Data Firehose and CloudTrail Logging (p. 50)](https://docs.aws.amazon.com/lambda/latest/dg/configuration-cloudtrail.html)
Kinesis Data Firehose and CloudTrail History

The CloudTrail API activity history feature lets you look up and filter events captured by CloudTrail. You can look up events related to the creation, modification, or deletion of resources in your AWS account on a per-region basis. Events can be looked up by using the CloudTrail console, or programmatically by using the AWS SDKs or AWS CLI.

The following actions are supported:

- CreateDeliveryStream
- DeleteDeliveryStream
- UpdateDestination

Kinesis Data Firehose and CloudTrail Logging

When CloudTrail logging is enabled in your AWS account, API calls made to specific Kinesis Data Firehose actions are tracked in CloudTrail log files. Kinesis Data Firehose actions are written with other AWS service records. CloudTrail determines when to create and write to a new file based on the specified time period and file size.

The following actions are supported:

- CreateDeliveryStream
- DescribeDeliveryStream
- ListDeliveryStreams
- UpdateDestination
- DeleteDeliveryStream

Every log entry contains information about who generated the request. For example, if a request is made to create a delivery stream (CreateDeliveryStream), the identity of the person or service that made the request is logged. The identity information in the log entry helps you determine the following:

- Whether the request was made with root or IAM user credentials
- Whether the request was made with temporary security credentials for a role or federated user
- Whether the request was made by another AWS service

For more information, see the CloudTrail userIdentity Element.

You can store your log files in your bucket for as long as needed but you can also define Amazon S3 lifecycle rules to archive or delete log files automatically. By default, your log files are encrypted by using Amazon S3 server-side encryption (SSE).

You can have CloudTrail publish SNS notifications when new log files are delivered if you want to take quick action upon log file delivery. For information, see Configuring Amazon SNS Notifications in the AWS CloudTrail User Guide.

You can also aggregate Kinesis Data Firehose log files from multiple AWS regions and multiple AWS accounts into a single S3 bucket. For more information, see Receiving CloudTrail Log Files from Multiple Regions and Receiving CloudTrail Log Files from Multiple Accounts in the AWS CloudTrail User Guide.
CloudTrail Log File Entries for Kinesis Data Firehose

CloudTrail log files contain one or more log entries. Each entry lists multiple JSON-formatted events. A log entry represents a single request from any source and includes information about the requested action, the date and time of the action, request parameters, and so on. The log entries are not an ordered stack trace of the public API calls, so they do not appear in any specific order.

The following is an example CloudTrail log entry. Note that for security reasons the user name and password for `RedshiftDestinationConfiguration` will be omitted and returned as an empty string.

```json
{
  "Records": [
    {
      "eventVersion": "1.02",
      "userIdentity": {
        "type": "IAMUser",
        "principalId": "AKIAIOSFODNN7EXAMPLE",
        "arn": "arn:aws:iam::111122223333:user/CloudTrail_Test_User",
        "accountId": "111122223333",
        "accessKeyId": "AKIAI44QH8DH8EXAMPLE",
        "userName": "CloudTrail_Test_User"
      },
      "eventTime": "2016-02-24T18:08:22Z",
      "eventSource": "firehose.amazonaws.com",
      "eventName": "CreateDeliveryStream",
      "awsRegion": "us-east-1",
      "sourceIPAddress": "127.0.0.1",
      "userAgent": "aws-internal/3",
      "requestParameters": {
        "deliveryStreamName": "TestRedshiftStream",
        "redshiftDestinationConfiguration": {
          "s3Configuration": {
            "compressionFormat": "GZIP",
            "prefix": "prefix",
            "bucketARN": "arn:aws:s3:::firehose-cloudtrail-test-bucket",
            "roleARN": "arn:aws:iam::111122223333:role/Firehose",
            "bufferingHints": {
              "sizeInMBs": 3,
              "intervalInSeconds": 900
            },
            "encryptionConfiguration": {
              "kMSEncryptionConfig": {
                "aWSKMSKeyARN": "arn:aws:kms:us-east-1:6ekey"
              }
            }
          }
        },
        "clusterJDBCURL": "jdbc:redshift://example.abc123.us-west-2.redshift.amazonaws.com:5439/dev",
        "copyCommand": {
          "copyOptions": "copyOptions",
          "dataTableName": "dataTable"
        },
        "password": "",
        "username": "",
        "roleARN": "arn:aws:iam::111122223333:role/Firehose"
      },
      "responseElements": {
        "deliveryStreamARN": "arn:aws:firehose:us-east-1:111122223333:deliverystream/TestRedshiftStream"
      }
    }
  ],
  "requestID": "958abf6a-db21-11e5-bb88-91e9617edf5",
  "eventId": "875d2d68-476c-4ad5-bbc6-d02872cfc884"
}
```
"eventType":"AwsApiCall",
"recipientAccountId":"111122223333"
},
{
"eventVersion":"1.02",
"userIdentity":{
  "type":"IAMUser",
  "principalId":"AKIAIOSFODNN7EXAMPLE",
  "arn":"arn:aws:iam::111122223333:user/CloudTrail_Test_User",
  "accountId":"111122223333",
  "accessKeyId":"AKIAI44QH8DHBEXAMPLE",
  "userName":"CloudTrail_Test_User"
},
"eventTime":"2016-02-24T18:08:54Z",
"eventSource":"firehose.amazonaws.com",
"eventName":"DescribeDeliveryStream",
"awsRegion":"us-east-1",
"sourceIPAddress":"127.0.0.1",
"userAgent":"aws-internal/3",
"requestParameters":{
  "deliveryStreamName":"TestRedshiftStream"
},
"responseElements":null,
"requestID":"aa6ea5ed-db21-11e5-bb88-91ae9617edf5",
"eventID":"d9b285d8-d690-4d5c-b9fe-d1ad5ab03f14",
"eventType":"AwsApiCall",
"recipientAccountId":"111122223333"
},
{
"eventVersion":"1.02",
"userIdentity":{
  "type":"IAMUser",
  "principalId":"AKIAIOSFODNN7EXAMPLE",
  "arn":"arn:aws:iam::111122223333:user/CloudTrail_Test_User",
  "accountId":"111122223333",
  "accessKeyId":"AKIAI44QH8DHBEXAMPLE",
  "userName":"CloudTrail_Test_User"
},
"eventTime":"2016-02-24T18:10:00Z",
"eventSource":"firehose.amazonaws.com",
"eventName":"ListDeliveryStreams",
"awsRegion":"us-east-1",
"sourceIPAddress":"127.0.0.1",
"userAgent":"aws-internal/3",
"requestParameters":{
  "limit":10
},
"responseElements":null,
"requestID":"d9b285d8-d690-4d5c-b9fe-d1ad5ab03f14",
"eventID":"d1bf7f86-db21-11e5-bb88-91ae9617edf5",
"eventType":"AwsApiCall",
"recipientAccountId":"111122223333"
},
{
"eventVersion":"1.02",
"userIdentity":{
  "type":"IAMUser",
  "principalId":"AKIAIOSFODNN7EXAMPLE",
  "arn":"arn:aws:iam::111122223333:user/CloudTrail_Test_User",
  "accountId":"111122223333",
  "accessKeyId":"AKIAI44QH8DHBEXAMPLE",
  "userName":"CloudTrail_Test_User"
},
"eventTime":"2016-02-24T18:10:09Z",
"eventSource":"firehose.amazonaws.com",
"eventName":"UpdateDestination",
"awsRegion":"us-east-1",
"sourceIPAddress":"127.0.0.1",
"userAgent":"aws-internal/3",
"requestParameters":null,
"responseElements":null,
"requestID":"d1bf7f86-db21-11e5-bb88-91ae9617edf5",
"eventID":"67f63c74-4335-48c0-9004-4ba35ce00128",
"eventType":"AwsApiCall",
"recipientAccountId":"111122223333"}
"awsRegion": "us-east-1",
"sourceIPAddress": "127.0.0.1",
"userAgent": "aws-internal/3",
"requestParameters": {
    "destinationId": "destinationId-0000000000001",
    "deliveryStreamName": "TestRedshiftStream",
    "currentDeliveryStreamVersionId": "1",
    "redshiftDestinationUpdate": {
        "roleARN": "arn:aws:iam::111122223333:role/Firehose",
        "clusterJDBCURL": "jdbc:redshift://example.abc123.us-west-2.redshift.amazonaws.com:5439/dev",
        "password": "",
        "username": "",
        "copyCommand": {
            "copyOptions": "copyOptions",
            "dataTable": "dataTable"
        },
        "s3Update": {
            "bucketARN": "arn:aws:s3:::firehose-cloudtrail-test-bucket-update",
            "roleARN": "arn:aws:iam::111122223333:role/Firehose",
            "compressionFormat": "GZIP",
            "bufferingHints": {
                "sizeInMBs": 3,
                "intervalInSeconds": 900
            },
            "encryptionConfiguration": {
                "AWSKMSKeyARN": "arn:aws:kms:us-east-1:key"
            },
            "prefix": "arn:aws:s3:::firehose-cloudtrail-test-bucket"
        }
    },
    "responseElements": null,
    "requestID": "d549428d-db21-11e5-bb88-91ae9617edf5",
    "eventID": "1cb21e0b-416a-415d-bbf9-769b152a6585",
    "eventType": "AwsApiCall",
    "recipientAccountId": "111122223333"
},

"eventVersion": "1.02",
"userIdentity": {
    "type": "IAMUser",
    "principalId": "AKIAIOSFODNN7EXAMPLE",
    "arn": "arn:aws:iam::111122223333:user/CloudTrail_Test_User",
    "accountId": "111122223333",
    "accessKeyId": "AKIAI44QH3DNSHEXAMPLE",
    "username": "CloudTrail_Test_User"
},
"eventTime": "2016-02-24T18:10:12Z",
"eventSource": "firehose.amazonaws.com",
"eventName": "DeleteDeliveryStream",
"awsRegion": "us-east-1",
"sourceIPAddress": "127.0.0.1",
"userAgent": "aws-internal/3",
"requestParameters": {
    "deliveryStreamName": "TestRedshiftStream"
},
"responseElements": null,
"requestID": "d85968c1-db21-11e5-bb88-91ae9617edf5",
"eventID": "dd46bb98-b4e9-42ff-a6af-32d57e636ad1",
"eventType": "AwsApiCall",
"recipientAccountId": "111122223333"}
Controlling Access with Amazon Kinesis Data Firehose

The following sections cover how to control access to and from your Kinesis Data Firehose resources. The information they cover includes how to grant your application access so it can send data to your Kinesis Data Firehose delivery stream. They also describe how you can grant Kinesis Data Firehose access to your Amazon Simple Storage Service (Amazon S3) bucket, Amazon Redshift cluster, or Amazon Elasticsearch Service cluster, as well as the access permissions you need if you use Splunk as your destination. Finally, you'll find in this topic guidance on how to configure Kinesis Data Firehose so it can deliver data to a destination that belongs to a different AWS account. The technology for managing all these forms of access is AWS Identity and Access Management (IAM). For more information about IAM, see What is IAM?

Contents

- Grant Your Application Access to Your Kinesis Data Firehose Resources (p. 55)
- Grant Kinesis Data Firehose Access to an Amazon S3 Destination (p. 56)
- Grant Kinesis Data Firehose Access to an Amazon Redshift Destination (p. 57)
- Grant Kinesis Data Firehose Access to an Amazon ES Destination (p. 60)
- Grant Kinesis Data Firehose Access to a Splunk Destination (p. 62)
- VPC Access to Splunk (p. 64)
- Cross-Account Delivery (p. 64)

Grant Your Application Access to Your Kinesis Data Firehose Resources

To give your application access to your Kinesis Data Firehose delivery stream, use a policy similar to this example. You can adjust the individual API operations to which you grant access by modifying the Action section, or grant access to all operations with "firehose:*".

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "firehose:DeleteDeliveryStream",
        "firehose:PutRecord",
        "firehose:PutRecordBatch",
        "firehose:UpdateDestination"
      ],
      "Resource": [
        "arn:aws:firehose:region:account-id:deliverystream/delivery-stream-name"
      ]
    }
  ]
}
```
Grant Kinesis Data Firehose Access to an Amazon S3 Destination

When you're using an Amazon S3 destination, Kinesis Data Firehose delivers data to your S3 bucket and can optionally use an AWS KMS key that you own for data encryption. If error logging is enabled, Kinesis Data Firehose also sends data delivery errors to your CloudWatch log group and streams. You are required to have an IAM role when creating a delivery stream. Kinesis Data Firehose assumes that IAM role and gains access to the specified bucket, key, and CloudWatch log group and streams.

Use the following trust policy to enable Kinesis Data Firehose to assume the role. Edit the policy to replace \texttt{account-id} with your AWS account ID. This ensures that only you can request Kinesis Data Firehose to assume the IAM role.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "firehose.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "sts:ExternalId": "account-id"
                }
            }
        }
    ]
}
```

Use the following access policy to enable Kinesis Data Firehose to access your S3 bucket and AWS KMS key. If you don't own the S3 bucket, add \texttt{s3:PutObjectAcl} to the list of Amazon S3 actions. This grants the bucket owner full access to the objects delivered by Kinesis Data Firehose. This policy also has a statement that allows access to Amazon Kinesis Data Streams. If you don't use Kinesis Data Streams as your data source, you can remove that statement.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:AbortMultipartUpload",
                "s3:GetBucketLocation",
                "s3:GetObject",
                "s3:ListBucket",
                "s3:ListBucketMultipartUploads",
                "s3:PutObject"
            ],
            "Resource": [
                "arn:aws:s3:::bucket-name",
                "arn:aws:s3:::bucket-name/*"
            ],
            "Principal": {
                "Service": "firehose.amazonaws.com"
            }
        },
        {
            "Effect": "Allow",
            "Action": ["s3:PutObjectAcl"]
        }
    ]
}
```
Grant Kinesis Data Firehose Access to an Amazon Redshift Destination

Refer to the following when you are granting access to Kinesis Data Firehose when using an Amazon Redshift destination.

**Topics**
- IAM Role and Access Policy (p. 58)
- VPC Access to an Amazon Redshift Cluster (p. 59)

For more information about allowing other AWS services to access your AWS resources, see Creating a Role to Delegate Permissions to an AWS Service in the IAM User Guide.
IAM Role and Access Policy

When you're using an Amazon Redshift destination, Kinesis Data Firehose delivers data to your S3 bucket as an intermediate location. It can optionally use an AWS KMS key you own for data encryption. Kinesis Data Firehose then loads the data from the S3 bucket to your Amazon Redshift cluster. If error logging is enabled, Kinesis Data Firehose also sends data delivery errors to your CloudWatch log group and streams. Kinesis Data Firehose uses the specified Amazon Redshift user name and password to access your cluster, and uses an IAM role to access the specified bucket, key, CloudWatch log group, and streams. You are required to have an IAM role when creating a delivery stream.

Use the following trust policy to enable Kinesis Data Firehose to assume the role. Edit the policy to replace `account-id` with your AWS account ID. This is so that only you can request Kinesis Data Firehose to assume the IAM role.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "firehose.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEquals": {
          "sts:ExternalId": "account-id"
        }
      }
    }
  ]
}
```

Use the following access policy to enable Kinesis Data Firehose to access your S3 bucket and AWS KMS key. If you don't own the S3 bucket, add `s3:PutObjectAcl` to the list of Amazon S3 actions, which grants the bucket owner full access to the objects delivered by Kinesis Data Firehose. This policy also has a statement that allows access to Amazon Kinesis Data Streams. If you don't use Kinesis Data Streams as your data source, you can remove that statement.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Resource": ["arn:aws:s3:::bucket-name", "arn:aws:s3:::bucket-name/*"]
    },
    {
      "Effect": "Allow",
      "Action": ["kms:Decrypt"],
      "Resource": "arn:aws:kms:us-east-1:123456789012:key/1234abcd-1234-1234-1234-1234567890ab"
    }
  ]
}
```
"kms:GenerateDataKey"
],
"Resource": [
  "arn:aws:kms:region:account-id:key/key-id"
],
"Condition": {
  "StringEquals": {
    "kms:ViaService": "s3.region.amazonaws.com"
  },
  "StringLike": {
    "kms:EncryptionContext:aws:s3:arn": "arn:aws:s3:::bucket-name/prefix*"
  }
},

"Effect": "Allow",
"Action": [
  "kinesis:DescribeStream",
  "kinesis:GetShardIterator",
  "kinesis:GetRecords"
],
},

"Effect": "Allow",
"Action": [
  "logs:PutLogEvents"
],
"Resource": [
}
}

"Effect": "Allow",
"Action": [
  "lambda:InvokeFunction",
  "lambda:GetFunctionConfiguration"
],
"Resource": [
]}
}

For more information about allowing other AWS services to access your AWS resources, see Creating a Role to Delegate Permissions to an AWS Service in the IAM User Guide.

VPC Access to an Amazon Redshift Cluster

If your Amazon Redshift cluster is in a virtual private cloud (VPC), it must be publicly accessible with a public IP address. Also, grant Kinesis Data Firehose access to your Amazon Redshift cluster by unblocking the Kinesis Data Firehose IP addresses. Kinesis Data Firehose currently uses one CIDR block for each available Region:

- 13.58.135.96/27 for US East (Ohio)
- 52.70.63.192/27 for US East (N. Virginia)
- 13.57.135.192/27 for US West (N. California)
- 52.89.255.224/27 for US West (Oregon)
- 13.228.64.192/27 for Asia Pacific (Singapore)
For more information about how to unblock IP addresses, see the step Authorize Access to the Cluster in the Amazon Redshift Getting Started guide.

Grant Kinesis Data Firehose Access to an Amazon ES Destination

When you're using an Amazon ES destination, Kinesis Data Firehose delivers data to your Amazon ES cluster, and concurrently backs up failed or all documents to your S3 bucket. If error logging is enabled, Kinesis Data Firehose also sends data delivery errors to your CloudWatch log group and streams. Kinesis Data Firehose uses an IAM role to access the specified Elasticsearch domain, S3 bucket, AWS KMS key, and CloudWatch log group and streams. You are required to have an IAM role when creating a delivery stream.

Use the following trust policy to enable Kinesis Data Firehose to assume the role. Edit the following policy to replace account-id with your AWS account ID. This is so that only you can request Kinesis Data Firehose to assume the IAM role.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "firehose.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEquals": {
          "sts:ExternalId": "account-id"
        }
      }
    }
  ]
}
```

Use the following access policy to enable Kinesis Data Firehose to access your S3 bucket, Amazon ES domain, and AWS KMS key. If you do not own the S3 bucket, add s3:PutObjectAcl to the list of Amazon S3 actions, which grants the bucket owner full access to the objects delivered by Kinesis Data Firehose. This policy also has a statement that allows access to Amazon Kinesis Data Streams. If you don't use Kinesis Data Streams as your data source, you can remove that statement.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:AbortMultipartUpload",
        "s3:GetBucketLocation",
        "s3:GetObject",
        "s3:GetObjectAcl",
        "s3:PutObject",
        "s3:PutObjectAcl"
      ],
      "Resource": "arn:aws:s3:::backupbucket"
    }
  ]
}
```
Grant Kinesis Data Firehose Access to an Amazon ES Destination

```
"s3:ListBucket",
"s3:ListBucketMultipartUploads",
"s3:PutObject"
],
"Resource": [
  "arn:aws:s3:::bucket-name",
  "arn:aws:s3:::bucket-name/*"
]
}
{
  "Effect": "Allow",
  "Action": [
    "kms:Decrypt",
    "kms:GenerateDataKey"
  ],
  "Resource": [
    "arn:aws:kms::region:account-id:key/key-id"
  ],
  "Condition": {
    "StringEquals": {
      "kms:ViaService": "s3.region.amazonaws.com"
    },
    "StringLike": {
      "kms:EncryptionContext:aws:s3:arn": "arn:aws:s3:::bucket-name/prefix*"
    }
  }
}
{
  "Effect": "Allow",
  "Action": [
    "es:DescribeElasticsearchDomain",
    "es:DescribeElasticsearchDomains",
    "es:DescribeElasticsearchDomainConfig",
    "es:ESHttpPost",
    "es:ESHttpPut"
  ],
  "Resource": [
    "arn:aws:es::region:account-id:domain/domain-name",
    "arn:aws:es::region:account-id:domain/domain-name/*"
  ]
}
{
  "Effect": "Allow",
  "Action": [
    "es:ESHttpGet"
  ],
  "Resource": [
    "arn:aws:es::region:account-id:domain/domain-name/_all/_settings",
    "arn:aws:es:region:account-id:domain/domain-name/index-name/_mapping/type-name",
    "arn:aws:es:region:account-id:domain/domain-name/index-name/*/_stats"
  ]
}
{
  "Effect": "Allow",
  "Action": [
    "kinesis:DescribeStream",
    "kinesis:GetShardIterator",
    "kinesis:GetRecords"
  ],
  "Resource": "arn:aws:kinesis::region:account-id:stream/stream-name"
```
Grant Kinesis Data Firehose Access to a Splunk Destination

When you're using a Splunk destination, Kinesis Data Firehose delivers data to your Splunk HTTP Event Collector (HEC) endpoint. It also backs up that data to the Amazon S3 bucket that you specify, and you can optionally use an AWS KMS key that you own for Amazon S3 server-side encryption. If error logging is enabled, Kinesis Data Firehose sends data delivery errors to your CloudWatch log streams. You can also use AWS Lambda for data transformation.

You are required to have an IAM role when creating a delivery stream. Kinesis Data Firehose assumes that IAM role and gains access to the specified bucket, key, and CloudWatch log group and streams.

Use the following trust policy to enable Kinesis Data Firehose to assume the role. Edit the policy to replace `account-id` with your AWS account ID. This ensures that only you can request Kinesis Data Firehose to assume the IAM role.

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "firehose.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEquals": {
          "sts:ExternalId": "account-id"
        }
      }
    }
  ]
}
```

For more information about allowing other AWS services to access your AWS resources, see Creating a Role to Delegate Permissions to an AWS Service in the IAM User Guide.
Use the following access policy to enable Kinesis Data Firehose to access your S3 bucket. If you don't own the S3 bucket, add `s3:PutObjectAcl` to the list of Amazon S3 actions, which grants the bucket owner full access to the objects delivered by Kinesis Data Firehose. This policy also grants Kinesis Data Firehose access to CloudWatch for error logging and to AWS Lambda for data transformation. The policy also has a statement that allows access to Amazon Kinesis Data Streams. If you don't use Kinesis Data Streams as your data source, you can remove that statement. Kinesis Firehose doesn't use IAM to access Splunk. For accessing Splunk, it uses your HEC token.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:AbortMultipartUpload",
                "s3:GetBucketLocation",
                "s3:GetObject",
                "s3:ListBucket",
                "s3:ListBucketMultipartUploads",
                "s3:PutObject"
            ],
            "Resource": [
                "arn:aws:s3:::bucket-name",
                "arn:aws:s3:::bucket-name/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:GenerateDataKey"
            ],
            "Resource": [
                "arn:aws:kms:region:account-id:key/key-id"
            ],
            "Condition": {
                "StringEquals": {
                    "kms:ViaService": "s3.region.amazonaws.com"
                },
                "StringLike": {
                    "kms:EncryptionContext:aws:s3:arn": "arn:aws:s3:::bucket-name/prefix*"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "kinesis:DescribeStream",
                "kinesis:GetShardIterator",
                "kinesis:GetRecords"
            ],
        },
        {
            "Effect": "Allow",
            "Action": [
                "logs:PutLogEvents"
            ],
            "Resource": [
            ]
        }
    ]
}`
VPC Access to Splunk

If your Splunk platform is in a VPC, it must be publicly accessible with a public IP address. Also, grant Kinesis Data Firehose access to your Splunk platform by unblocking the Kinesis Data Firehose IP addresses. Kinesis Data Firehose currently uses the following CIDR blocks.

- 18.216.68.160/27, 18.216.170.64/27, 18.216.170.96/27 for US East (Ohio)
- 34.238.188.128/26, 34.238.188.192/26, 34.238.195.0/26 for US East (N. Virginia)
- 13.57.180.0/26 for US West (N. California)
- 34.216.24.32/27, 34.216.24.192/27, 34.216.24.224/27 for US West (Oregon)
- 13.229.187.128/26 for Asia Pacific (Singapore)
- 13.211.12.0/26 for Asia Pacific (Sydney)
- 18.194.95.192/27, 18.194.95.224/27, 18.195.48.0/27 for EU (Frankfurt)
- 34.241.197.32/27, 34.241.197.64/27, 34.241.197.96/27 for EU (Ireland)

Cross-Account Delivery

You can configure your Kinesis Data Firehose delivery stream to deliver data to a destination that belongs to a different AWS account, as long as the destination supports resource-based permissions. For more information about resource-based permissions, see Identity-Based (IAM) Permissions and Resource-Based Permissions. For a list of AWS services that support resource-based permissions, see AWS Services That Work with IAM.

The following procedure shows an example of configuring a Kinesis Data Firehose delivery stream owned by account A to deliver data to an Amazon S3 bucket owned by account B.

1. Create an IAM role under account A using steps described in Grant Kinesis Firehose Access to an Amazon S3 Destination.

   **Note**
   The Amazon S3 bucket specified in the access policy is owned by account B in this case. Make sure you add s3:PutObjectAcl to the list of Amazon S3 actions in the access policy, which grants account B full access to the objects delivered by Amazon Kinesis Data Firehose.

2. To allow access from the IAM role previously created, create an S3 bucket policy under account B. The following code is an example of the bucket policy. For more information, see Using Bucket Policies and User Policies.
3. Create a Kinesis Data Firehose delivery stream under account A using the IAM role that you created in step 1.
Troubleshooting Amazon Kinesis Data Firehose

You may not see data delivered to your specified destinations. Use the troubleshooting steps in this topic to solve common issues you might encounter while using Kinesis Data Firehose. For further troubleshooting, see Monitoring with CloudWatch Logs (p. 43).

Issues

- Data Not Delivered to Amazon S3 (p. 66)
- Data Not Delivered to Amazon Redshift (p. 66)
- Data Not Delivered to Amazon Elasticsearch Service (p. 67)
- Data Not Delivered to Splunk (p. 68)
- Delivery Stream Not Available as a Target for CloudWatch Logs, CloudWatch Events, or AWS IoT Action (p. 69)

Data Not Delivered to Amazon S3

Check the following if data is not delivered to your S3 bucket.

- Check the Kinesis Data Firehose `IncomingBytes` and `IncomingRecords` metrics to make sure that data is sent to your Kinesis data delivery stream successfully. For more information, see Monitoring with Amazon CloudWatch Metrics (p. 38).
- If data transformation with Lambda is enabled, check the Kinesis Data Firehose `ExecuteProcessingSuccess` metric to make sure that Kinesis Data Firehose has attempted to invoke your Lambda function. For more information, see Monitoring with Amazon CloudWatch Metrics (p. 38).
- Check the Kinesis Data Firehose `DeliveryToS3.Success` metric to make sure that Kinesis Data Firehose has attempted putting data to your S3 bucket. For more information, see Monitoring with Amazon CloudWatch Metrics (p. 38).
- Enable error logging if it is not already enabled, and check error logs for delivery failure. For more information, see Monitoring with Amazon CloudWatch Logs (p. 43).
- Make sure that the S3 bucket specified in your Kinesis data delivery stream still exists.
- If data transformation with Lambda is enabled, make sure that the Lambda function specified in your delivery stream still exists.
- Make sure that the IAM role specified in your Kinesis data delivery stream has access to your S3 bucket and your Lambda function (if data transformation is enabled). For more information, see Grant Kinesis Data Firehose Access to an Amazon S3 Destination (p. 56).
- If you're using data transformation, make sure that your Lambda function never returns responses whose payload size exceeds 6 MB. For more information, see Amazon Kinesis Data Firehose Data Transformation.

Data Not Delivered to Amazon Redshift

Check the following if data is not delivered to your Amazon Redshift cluster.
Data Not Delivered to Amazon Elasticsearch Service

Check the following if data is not delivered to your Elasticsearch domain.

Note that data can be backed up to your S3 bucket concurrently. If data was not delivered to your S3 bucket, see Data Not Delivered to Amazon S3 (p. 66).

- Check the Kinesis Data Firehose DeliveryToRedshift.Success metric to make sure that Kinesis Data Firehose has attempted to copy data from your S3 bucket to the Amazon Redshift cluster. For more information, see Monitoring with Amazon CloudWatch Metrics (p. 38).
- Enable error logging if it is not already enabled, and check error logs for delivery failure. For more information, see Monitoring with Amazon CloudWatch Logs (p. 43).
- Check the Amazon Redshift STL_CONNECTION_LOG table to see if Kinesis Data Firehose is able to make successful connections. In this table you should be able to see connections and their status based on a user name. For more information, see STL_CONNECTION_LOG in the Amazon Redshift Database Developer Guide.
- If the previous check shows that connections are being established, check the Amazon Redshift STL_LOAD_ERRORS table to verify the reason of the COPY failure. For more information, see STL_LOAD_ERRORS in the Amazon Redshift Database Developer Guide.
- Make sure that the Amazon Redshift configuration in your Kinesis data delivery stream is accurate and valid.
- Make sure that the IAM role specified in your Kinesis data delivery stream has access to the S3 bucket from which Amazon Redshift copies data and the Lambda function for data transformation (if data transformation is enabled). For more information, see Grant Kinesis Data Firehose Access to an Amazon S3 Destination (p. 56).
- If your Amazon Redshift cluster is in a VPC, make sure that the cluster allows access from Kinesis Data Firehose IP addresses. For more information, see Grant Kinesis Data Firehose Access to an Amazon Redshift Destination (p. 57).
- Make sure that the Amazon Redshift cluster is publicly accessible.
- If you're using data transformation, make sure that your Lambda function never returns responses whose payload size exceeds 6 MB. For more information, see Amazon Kinesis Data Firehose Data Transformation.
• Make sure that the IAM role specified in your delivery stream has access to your Amazon ES cluster and Lambda function (if data transformation is enabled). For more information, see Grant Kinesis Data Firehose Access to an Amazon ES Destination (p. 60).

• If you’re using data transformation, make sure that your Lambda function never returns responses whose payload size exceeds 6 MB. For more information, see Amazon Kinesis Data Firehose Data Transformation.

Data Not Delivered to Splunk

Check the following if data is not delivered to your Splunk endpoint.

• Review the Splunk platform requirements. The Splunk Add-on for Kinesis Data Firehose requires Splunk platform version 6.6.X or later. For more information, see Splunk Add-on for Amazon Kinesis Firehose.

• If you have a proxy (Elastic Load Balancing or other) between Kinesis Data Firehose and the HTTP Event Collector (HEC) node, be sure to enable sticky sessions in order to support HEC acknowledgements (ACKs).

• Make sure that you are using a valid HEC token.

• Ensure that the HEC token is enabled. See Enable and disable Event Collector tokens.

• Check to see whether the data that you’re sending to Splunk is formatted correctly. See Format events for HTTP Event Collector.

• Make sure that the HEC token and input event are configured with a valid index.

• When an upload to Splunk fails due to a server error from the HEC node, the request is automatically retried. If all retries fail, the data gets backed up to Amazon S3. Check to see if your data appears in Amazon S3, which is an indication of such a failure.

• Make sure that you’ve enabled indexer acknowledgment on your HEC token. See Enable indexer acknowledgment.

• Increase the value of HECAcknowledgmentTimeoutInSeconds in the Splunk destination configuration of your Kinesis Data Firehose delivery stream.

• Increase the value of DurationInSeconds under RetryOptionsin the Splunk destination configuration of your Kinesis Data Firehose delivery stream.

• Check your HEC health.

• If you’re using data transformation, make sure that your Lambda function never returns responses whose payload size exceeds 6 MB. For more information, see Amazon Kinesis Data Firehose Data Transformation.

• Make sure that the Splunk parameter named ackIdleCleanup is set to true. It’s false by default. To set this parameter to true, do the following:
  • For a managed Splunk Cloud deployment, submit a case using the Splunk Support Portal. In the case, ask Splunk Support to enable the HTTP event collector, set ackIdleCleanup to true in inputs.conf, and create or modify an elastic load balancer to use with this add-on.
  • For a distributed Splunk Enterprise deployment, set the ackIdleCleanup parameter to true in the inputs.conf file, which is located under $SPLUNK_HOME/etc/apps/splunk_httpinput/local/ for *nix users and under %SPLUNK_HOME%/etc/apps/splunk_httpinput/local/ for Windows users.
  • For a single-instance Splunk Enterprise deployment, set the ackIdleCleanup parameter to true in the inputs.conf file, which is located under $SPLUNK_HOME/etc/apps/splunk_httpinput/local/ for *nix users and under %SPLUNK_HOME%/etc/apps/splunk_httpinput/local/ for Windows users.

• See Troubleshoot the Splunk Add-on for Amazon Kinesis Firehose.
Delivery Stream Not Available as a Target for CloudWatch Logs, CloudWatch Events, or AWS IoT Action

Some AWS services can only send messages and events to a Kinesis Data Firehose delivery stream that is in the same Region. Verify that your Kinesis Data Firehose delivery stream is located in the same Region as your other services.
Amazon Kinesis Data Firehose has the following limits.

- By default, each account can have up to 50 Kinesis data delivery streams per Region. If you exceed this limit, a call to CreateDeliveryStream results in a LimitExceededException exception. This limit can be increased using the Amazon Kinesis Data Firehose Limits form.

- When Direct PUT is configured as the data source, each Kinesis data delivery stream is subject to the following limits:
  - For US East (N. Virginia), US West (Oregon), and EU (Ireland): 5,000 records/second, 2,000 transactions/second, and 5 MB/second.
  - For US East (Ohio), US West (N. California), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Tokyo), and EU (Frankfurt): 1,000 records/second, 1,000 transactions/second, and 1 MB/second.

You can submit a limit increase request using the Amazon Kinesis Data Firehose Limits form. The three limits scale proportionally. For example, if you increase the throughput limit in US East (N. Virginia), US West (Oregon), or EU (Ireland) to 10 MB/second, the other two limits increase to 4,000 transactions/second and 10,000 records/second.

**Important**
If the increased limit is much higher than the running traffic, it causes very small delivery batches to destinations, which is inefficient and can result in higher costs at the destination services. Be sure to increase the limit only to match current running traffic, and increase the limit further if traffic increases.

**Note**
When Kinesis Data Streams is configured as the data source, this limit doesn't apply, and Kinesis Data Firehose scales up and down with no limit.

- Each Kinesis data delivery stream stores data records for up to 24 hours in case the delivery destination is unavailable.
- The maximum size of a record sent to Kinesis Data Firehose, before base64-encoding, is 1,000 KB.
- The PutRecordBatch operation can take up to 500 records per call or 4 MB per call, whichever is smaller. This limit cannot be changed.
- The following operations can provide up to five transactions per second: CreateDeliveryStream, DeleteDeliveryStream, DescribeDeliveryStream, ListDeliveryStreams, and UpdateDestination.
- The buffer sizes hints range from 1 MB to 128 MB for Amazon S3 delivery and from 1 MB to 100 MB for Amazon Elasticsearch Service (Amazon ES) delivery. For AWS Lambda processing, you can set a buffering hint between 1 MB and 3 MB using the BufferSizeInMBs processor parameter. The size threshold is applied to the buffer before compression. These options are treated as hints, and therefore Kinesis Data Firehose might choose to use different values when it is optimal.
- The buffer interval hints range from 60 seconds to 900 seconds.
- For Kinesis Data Firehose to Amazon Redshift delivery, only publicly accessible Amazon Redshift clusters are supported.
- The retry duration range is from 0 seconds to 7,200 seconds for Amazon Redshift and Amazon ES delivery.
Document History

The following table describes the important changes to the Amazon Kinesis Firehose documentation.

<table>
<thead>
<tr>
<th>Change</th>
<th>Description</th>
<th>Date Changed</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Kinesis Streams as Source feature</td>
<td>Added Kinesis Streams as a potential source for records for a Firehose Delivery Stream. For more information, see Name and source (p. 5).</td>
<td>August 18, 2017</td>
</tr>
<tr>
<td>Update to console documentation</td>
<td>The delivery stream creation wizard was updated. For more information, see Creating an Amazon Kinesis Data Firehose Delivery Stream (p. 5).</td>
<td>July 19, 2017</td>
</tr>
<tr>
<td>New data transformation</td>
<td>You can configure Kinesis Data Firehose to transform your data before data delivery. For more information, see Amazon Kinesis Data Firehose Data Transformation (p. 27).</td>
<td>December 19, 2016</td>
</tr>
<tr>
<td>New Amazon Redshift COPY retry</td>
<td>You can configure Kinesis Firehose to retry a COPY command to your Amazon Redshift cluster if it fails. For more information, see Creating an Amazon Kinesis Data Firehose Delivery Stream (p. 5), Amazon Kinesis Data Firehose Data Delivery (p. 33), and Amazon Kinesis Data Firehose Limits (p. 70).</td>
<td>May 18, 2016</td>
</tr>
<tr>
<td>New Kinesis Firehose destination, Amazon Elasticsearch Service</td>
<td>You can create a delivery stream with Amazon Elasticsearch Service as the destination. For more information, see Creating an Amazon Kinesis Data Firehose Delivery Stream (p. 5), Amazon Kinesis Data Firehose Data Delivery (p. 33), and Grant Kinesis Data Firehose Access to an Amazon ES Destination (p. 60).</td>
<td>April 19, 2016</td>
</tr>
<tr>
<td>New enhanced CloudWatch metrics and troubleshooting features</td>
<td>Updated Monitoring Amazon Kinesis Data Firehose (p. 38) and Troubleshooting Amazon Kinesis Data Firehose (p. 66).</td>
<td>April 19, 2016</td>
</tr>
<tr>
<td>New enhanced Kinesis agent</td>
<td>Updated Writing to Amazon Kinesis Data Firehose Using Kinesis Agent (p. 16).</td>
<td>April 11, 2016</td>
</tr>
<tr>
<td>New Kinesis agents</td>
<td>Added Writing to Amazon Kinesis Data Firehose Using Kinesis Agent (p. 16).</td>
<td>October 2, 2015</td>
</tr>
<tr>
<td>Initial release</td>
<td>Initial release of the Amazon Kinesis Firehose Developer Guide.</td>
<td>October 4, 2015</td>
</tr>
</tbody>
</table>
AWS Glossary

For the latest AWS terminology, see the AWS Glossary in the AWS General Reference.