

Introduction to Amazon Kinesis Firehose

Ray Zhu, Sr. Product Manager, Amazon Kinesis

8/30/2016

Learn more - <http://amzn.to/2c34UfE>

Streaming data applications can deliver compelling, near real-time user experiences, but building the back-end infrastructure to collect and process streaming data is difficult. Amazon Kinesis Firehose makes it easy for you to load streaming data into AWS without having to build custom stream processing applications. In this webinar, we will introduce Amazon Kinesis Firehose and discuss how to ingest streaming data into Amazon S3, Amazon Redshift, and Amazon Elasticsearch Service using Amazon Kinesis Firehose. We will also highlight key use cases based on real-world examples from IoT, AdTech, E-Commerce, and Gaming.

Learning Objectives:

- Get an introduction to streaming data and an overview of Amazon Kinesis Firehose
- Learn about common streaming data use cases from IoT, Ad Tech, E-Commerce, and Gaming
- Understand how to use Amazon Kinesis Firehose to load streaming data into Amazon S3, Amazon Redshift, and Amazon Elasticsearch Service

Agenda

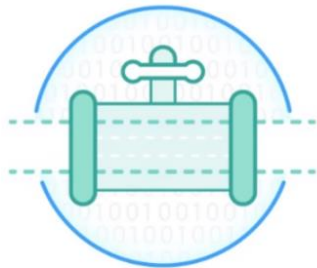
- Streaming Data and Amazon Kinesis
- Kinesis Firehose Overview
- Stream Data to Amazon S3
- Stream Data to Amazon Redshift
- Stream Data to Amazon Elasticsearch Service
- Kinesis Agent
- Kinesis Firehose Monitoring
- Kinesis Firehose Pricing

Streaming Data Scenarios

Scenarios 1 Accelerated Ingest-Transform-Load 2 Continual Metrics Generation 3 Responsive Data Analysis

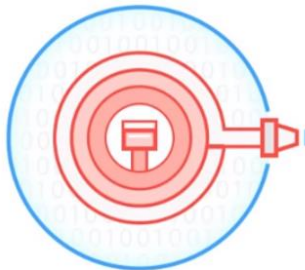
Data Types	IT logs, applications logs, social media / clickstreams, sensor or device data, market data		
Ad/Marketing Tech	Publisher, bidder data aggregation	Advertising metrics like coverage, yield, conversion	Analytics on user engagement with ads, optimized bid / buy engines
IoT	Sensor, device telemetry data ingestion	IT operational metrics dashboards	Sensor operational intelligence, alerts, and notifications
Gaming	Online customer engagement data aggregation	Consumer engagement metrics for level success, transition rates, CTR	Clickstream analytics, leaderboard generation, player-skill match engines
Consumer Engagement	Online customer engagement data aggregation	Consumer engagement metrics like page views, CTR	Clickstream analytics, recommendation engines

Amazon Kinesis



Kinesis Streams

Stores data as a continuous replayable stream for custom applications



Kinesis Firehose

Load streaming data into Amazon S3, Amazon Redshift, and Amazon Elasticsearch Service



Kinesis Analytics

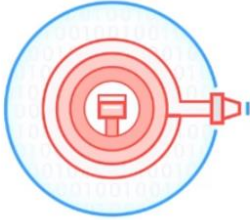
Analyze data streams using standard SQL queries

Match the Services and Use Cases

1 Accelerated Ingest-Transform-Load

2 Continual Metrics Generation

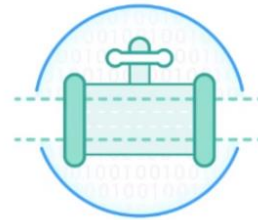
3 Responsive Data Analysis



Kinesis Firehose



Kinesis Analytics



Kinesis Streams

Kinesis Firehose Overview

Key Concepts

1. **Delivery Stream:** The underlying entity of Firehose. Use Firehose by creating a delivery stream to a specified destination and send data to it.
2. **Record:** The data of interest that your data producer sends to a delivery stream. A record can be as large as 1000 KB.
3. **Data Producers:** Producers send records to a Delivery Stream. For example, a web server sends log data to a delivery stream is a data producer.

Data Flow Overview



Capture and submit streaming data to Firehose

Firehose loads streaming data continuously into Amazon S3, Redshift, or Elasticsearch Service

Analyze streaming data using your favorite analytical tools

Zero administration: Capture and deliver streaming data into Amazon S3, Redshift, and Elasticsearch Service without writing an application or managing infrastructure.

Direct-to-data store integration: Batch, compress, and encrypt streaming data for delivery into data destinations in as little as 60 secs using simple configurations.

Seamless elasticity: Seamlessly scale to match data throughput without intervention.

Stream Data to S3

Let us see how to configure a Firehose delivery stream through the Firehose console. We are going to see how to use the S3 as our streaming destination

Stream Data to S3

Amazon Kinesis

Create Delivery Stream

Step 1: Destination
Step 2: Configuration
Step 3: Review

Destination

Select the destination where your streaming data will be delivered

Destination* Amazon S3

Delivery stream name* test

S3 Bucket

S3 bucket* firehose-test-bucket

S3 prefix S3 prefix

*Required

Cancel Next

Stream Data to S3

The screenshot shows the 'Create Delivery Stream' wizard in the AWS Management Console for Amazon Kinesis Firehose. The left sidebar shows 'Amazon Kinesis' with 'Kinesis Streams' and 'Kinesis Firehose' options. The main panel is titled 'Create Delivery Stream' and shows 'Step 2: Configuration'. The 'Configuration' section includes:

- S3 Buffer:** 'Buffer size' is set to 5 MB (range 1MB to 128MB in 1MB increments). 'Buffer interval' is set to 300 seconds (range 60s to 900s in 1 second increments).
- S3 Compression and Encryption:** 'Data compression' is set to 'UNCOMPRESSED'. 'Data encryption' is set to 'No Encryption'.
- Error Logging:** 'Enable' is selected.
- IAM Role:** 'Select an IAM role' is shown.

Buttons at the bottom include 'Cancel', 'Previous', and 'Next'. A '*Required' label is present.

Stream Data to Elasticsearch

Stream Data to Elasticsearch

The screenshot shows the 'Create Delivery Stream' wizard in the AWS Management Console for Amazon Kinesis Firehose. The left sidebar shows 'Amazon Kinesis' with 'Kinesis Streams' and 'Kinesis Firehose' options. The main panel is titled 'Create Delivery Stream' and shows 'Step 1: Destination'. The 'Destination' section includes:

- Destination:** 'Amazon Elasticsearch Service' is selected.
- Delivery stream name:** 'test' is entered.
- Elasticsearch domain:** 'Elasticsearch domain' is set to 'raydomain'. 'Index' is set to 'test'. 'Index rotation' is set to 'OneDay'. 'Type' is set to 'test'. 'Retry duration (sec)' is set to 300 (range 0 seconds to 7200 seconds in 1 second increments).
- Backup S3 bucket:** 'Backup mode' is set to 'Failed Documents Only'. 'S3 bucket' is set to 'firehose-test-bucket'. 'S3 prefix' is set to 'S3 prefix'.

Buttons at the bottom include 'Cancel' and 'Next'. A '*Required' label is present.

Stream Data to Elasticsearch

Amazon Kinesis

Create Delivery Stream

Step 1: Destination
Step 2: Configuration
Step 3: Review

Configuration

Configure buffer, compression, logging and IAM role options for your delivery stream.

Elasticsearch Buffer

Firehose buffers incoming data before delivering to your Elasticsearch cluster. You can configure buffer size and buffer interval. The first satisfied condition will trigger the data delivery.

Buffer size* 5
Buffer interval* 300

S3 Buffer

Firehose buffers incoming data before delivering to your backup S3 bucket. You can configure buffer size and buffer interval. The first satisfied condition will trigger the data delivery to your S3 bucket.

Buffer size* 5
Buffer interval* 300

S3 Compression and Encryption

Firehose can compress and encrypt the data before delivering it to your backup S3 bucket.

Data compression UNCOMPRESSED
Data encryption No Encryption

Error Logging

Firehose can log data delivery errors to CloudWatch Logs. If enabled, a CloudWatch Log Group and corresponding Log Stream(s) are created on your behalf. [Learn more](#)

☒ Enable ☐ Disable

IAM Role

Firehose needs an IAM role to access your specified resources, such as the S3 bucket and KMS key. [Learn more](#)

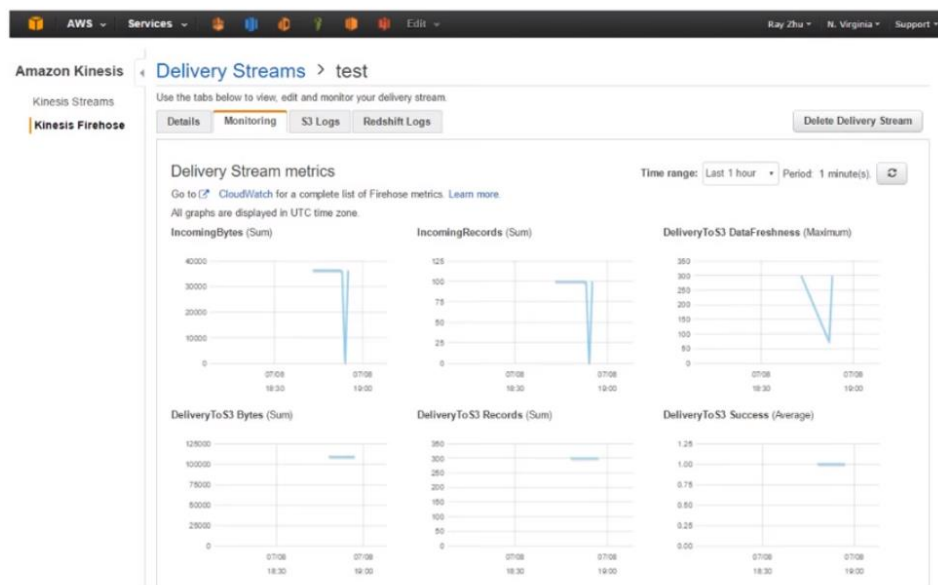
IAM role* Select an IAM role

*Required

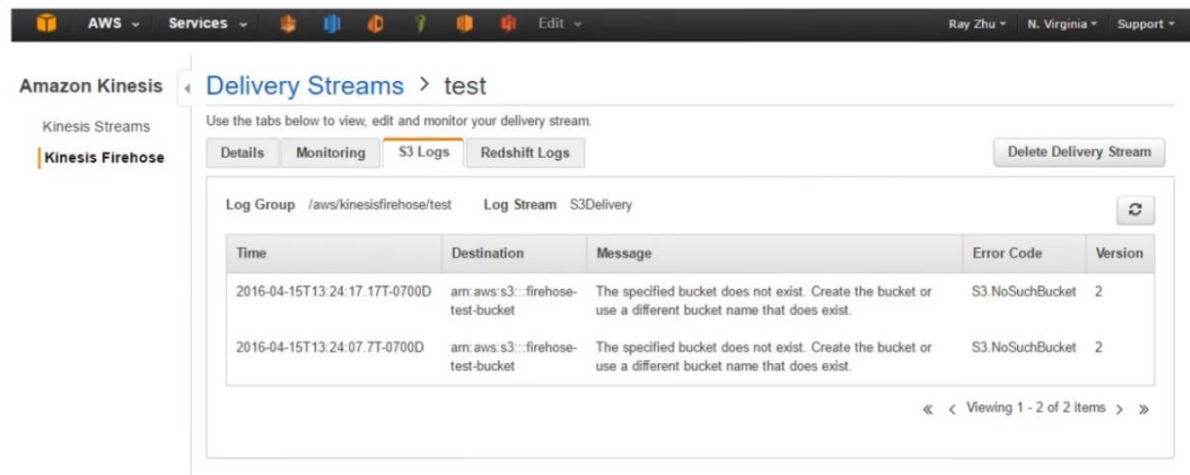
Cancel Previous Next

Kinesis Firehose Monitoring

Monitor with CloudWatch Metrics



Monitor with CloudWatch Logs



The screenshot shows the Amazon Kinesis console interface. The top navigation bar includes the AWS logo, 'Services' dropdown, and user information 'Ray Zhu' and 'N. Virginia'. The left sidebar shows 'Amazon Kinesis' with 'Kinesis Streams' and 'Kinesis Firehose' options. The main content area is titled 'Delivery Streams > test'. Below this, there's a instruction: 'Use the tabs below to view, edit and monitor your delivery stream.' There are four tabs: 'Details', 'Monitoring', 'S3 Logs' (which is selected), and 'Redshift Logs'. A 'Delete Delivery Stream' button is in the top right. The 'S3 Logs' tab displays a table with columns: 'Time', 'Destination', 'Message', 'Error Code', and 'Version'. The table shows two identical error records from 2016-04-15T13:24:17.7T-0700D to 2016-04-15T13:24:07.7T-0700D, both pointing to 'arn:aws:s3:::firehose-test-bucket' with the message 'The specified bucket does not exist. Create the bucket or use a different bucket name that does exist.' and error code 'S3.NoSuchBucket'. A pagination bar at the bottom indicates 'Viewing 1 - 2 of 2 items'.

Time	Destination	Message	Error Code	Version
2016-04-15T13:24:17.7T-0700D	arn:aws:s3:::firehose-test-bucket	The specified bucket does not exist. Create the bucket or use a different bucket name that does exist.	S3.NoSuchBucket	2
2016-04-15T13:24:07.7T-0700D	arn:aws:s3:::firehose-test-bucket	The specified bucket does not exist. Create the bucket or use a different bucket name that does exist.	S3.NoSuchBucket	2

Kinesis Firehose Pricing

Firehose Pricing

Dimension

Per 1 GB of data ingested

Value

\$0.035

* Records that are smaller than 5KB are rounded up to 5KB for billing

Thank you!

Q&A