Amazon Kinesis Data Firehose

Fernando Medina Corey
CLOUD ARCHITECT

@fmc_sea www.fernandomc.com

Overview

Basic Concepts

- Records, buffers, destinations, retention
- Firehose Limits

Data Transformations with Firehose

- Visualizing transformations
- Lambda transformation examples

Firehose Producers

Firehose Destinations

Demo: Creating a delivery stream

When to use Firehose vs. Data Streams

Kinesis Data Firehose: Basic Concepts

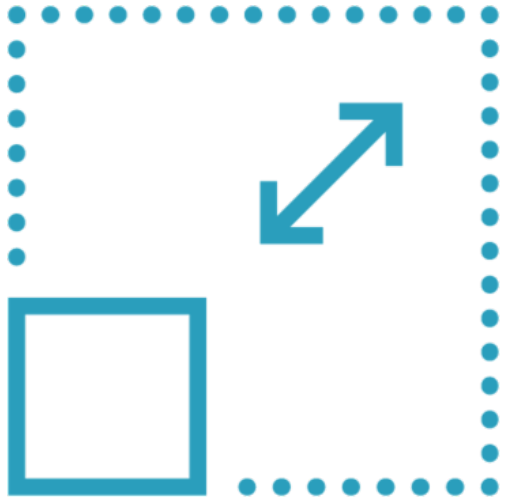
Records, Data Retention, and Destinations

Data Records

Data Retention
(24 hours max)

Destinations
(S3, Splunk, etc.)

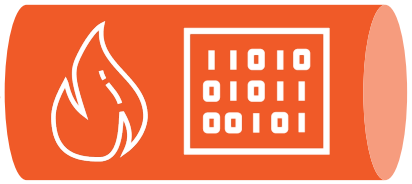
Firehose Buffer Size and Interval



Buffer Size
(1MB – 128MB)

Buffer Interval
(60 – 900 seconds)

Buffer Size

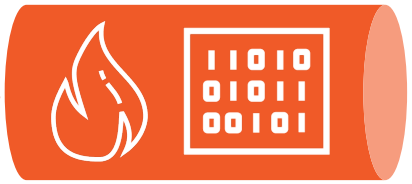


Kinesis Firehose

1 MB Buffer Size

S3 Bucket

Buffer Size

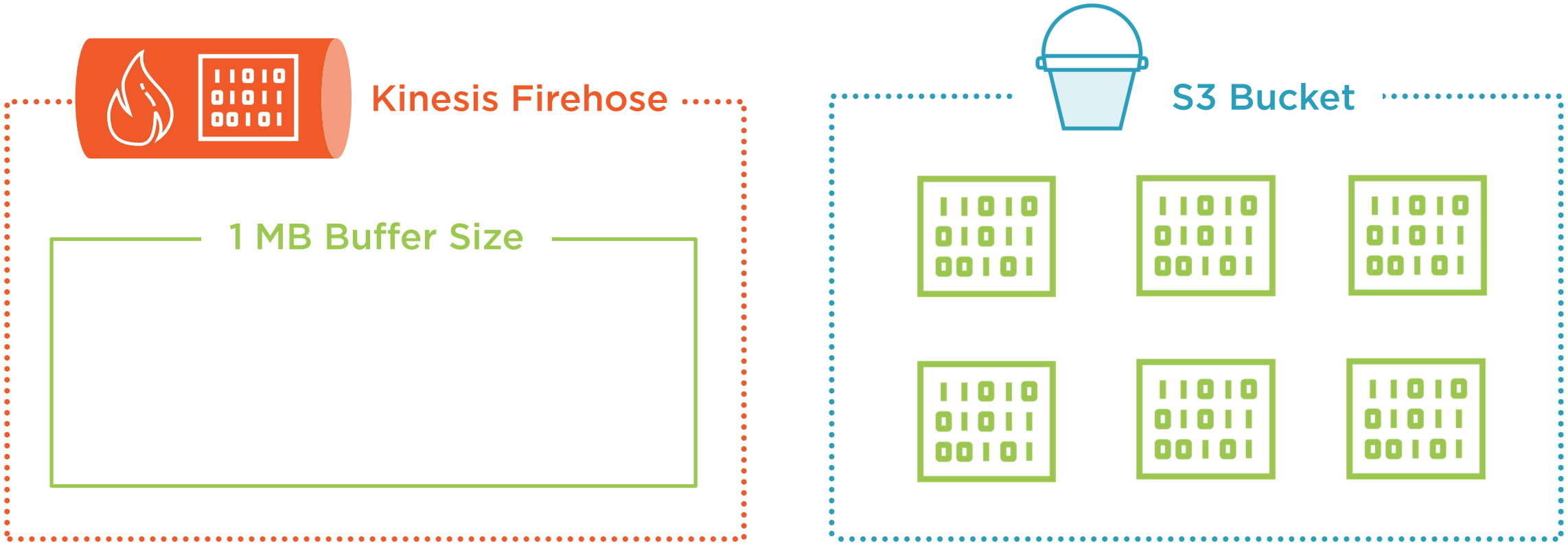


Kinesis Firehose

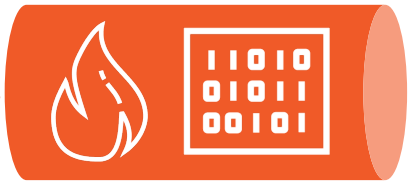
1 MB Buffer Size

S3 Bucket

Buffer Size



Buffer Interval

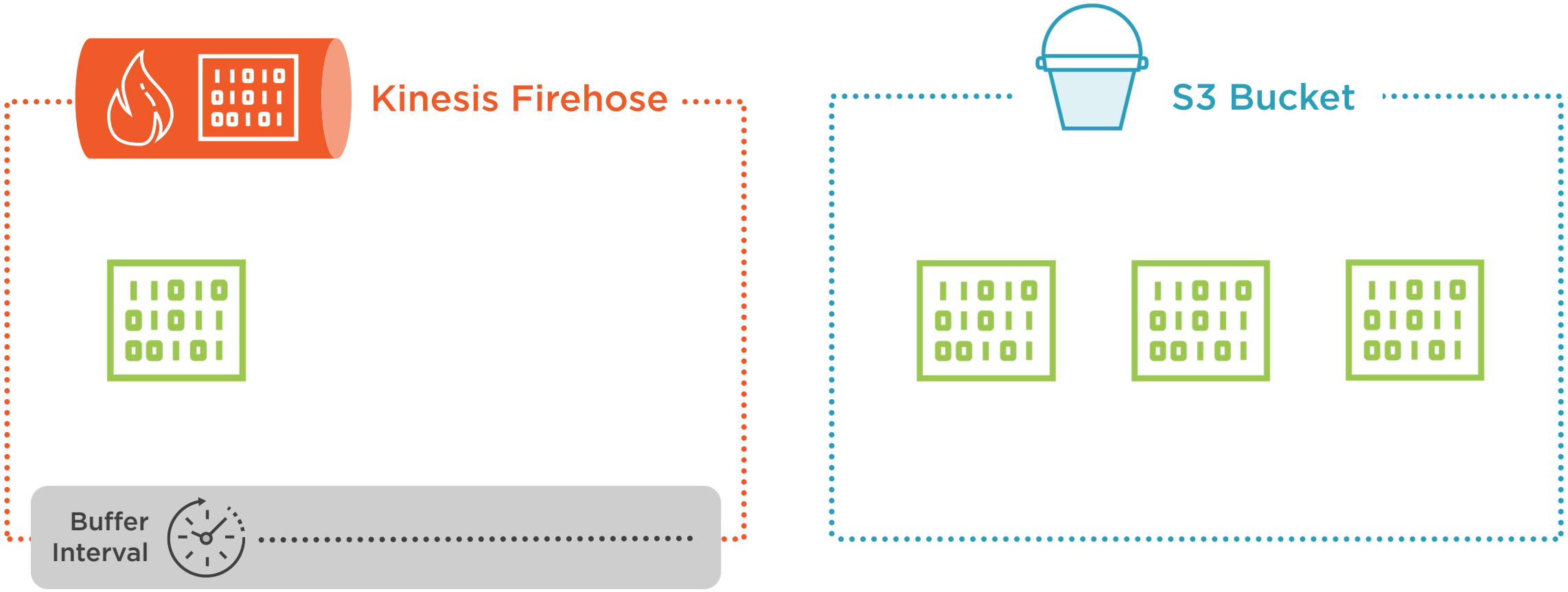


Kinesis Firehose

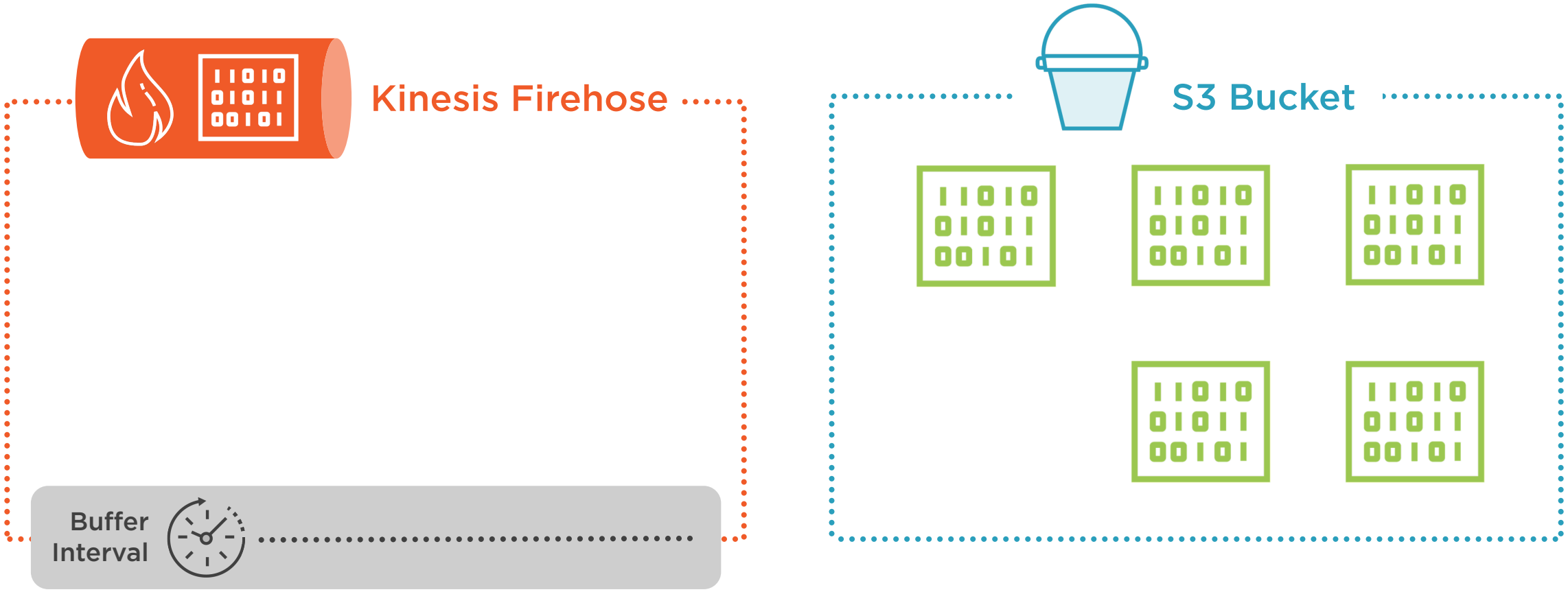
Buffer Interval

S3 Bucket

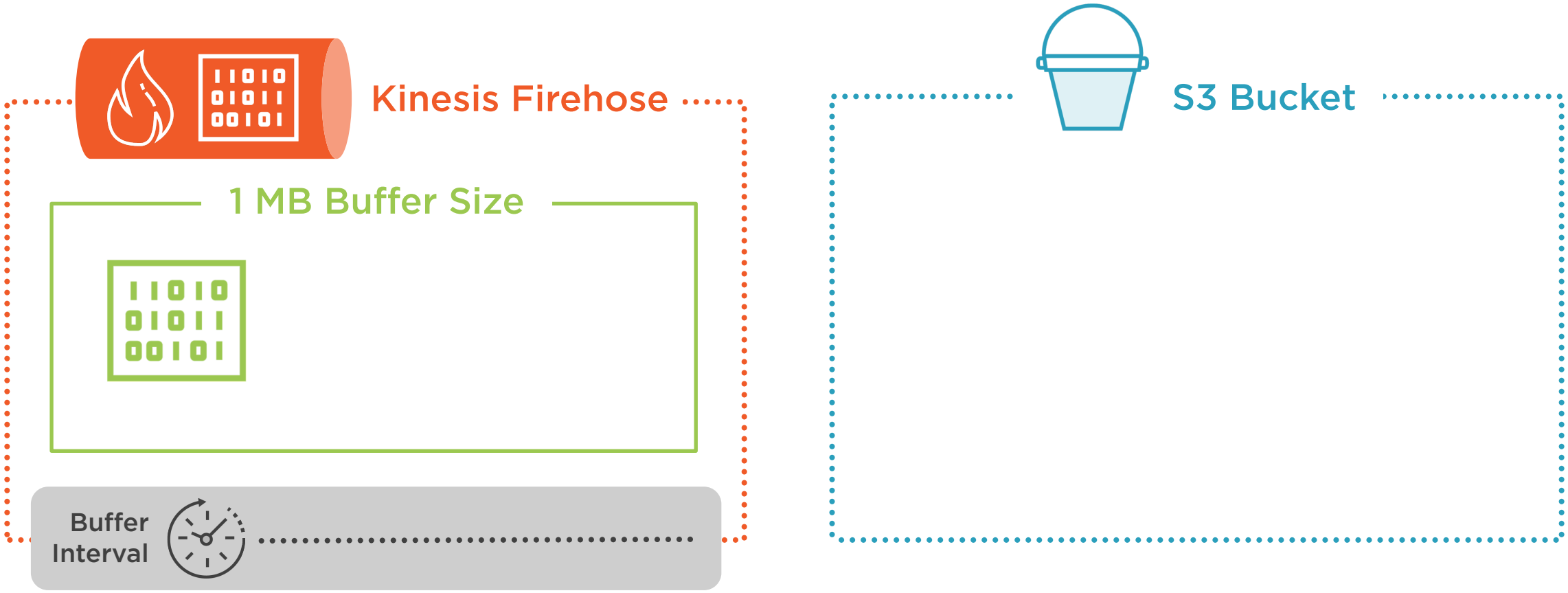
Buffer Interval



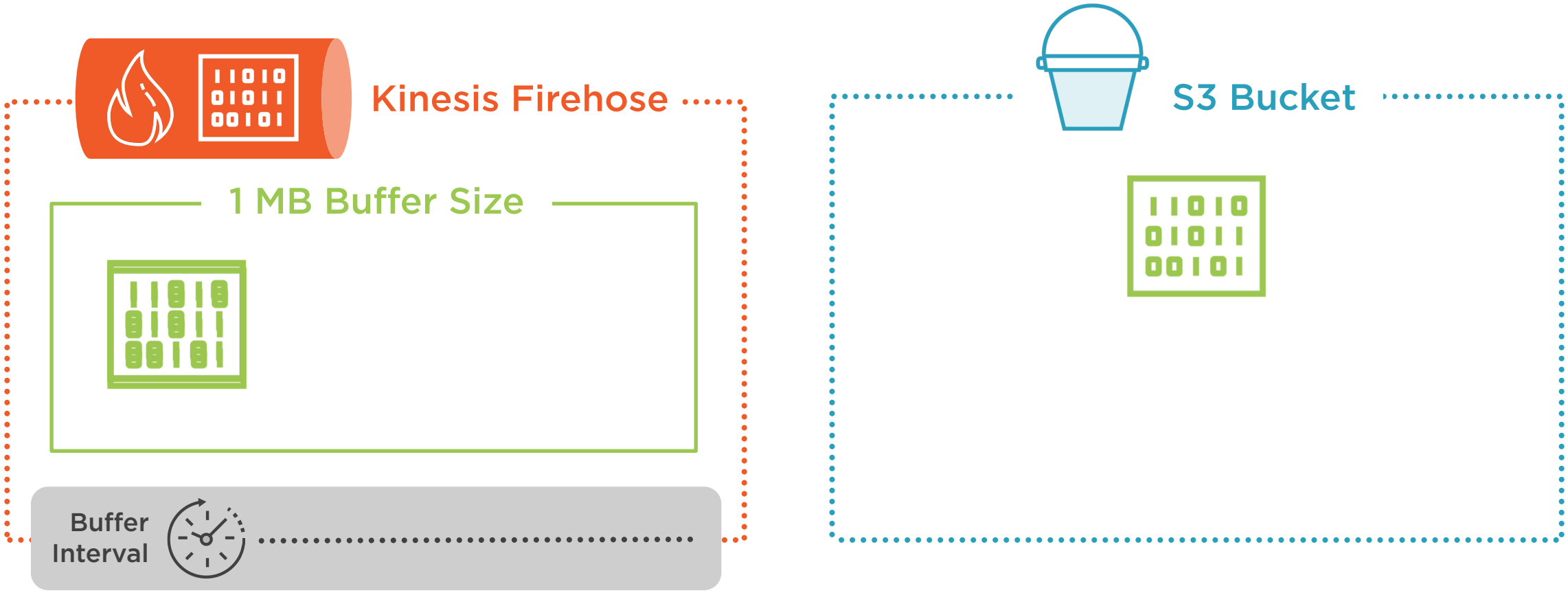
Buffer Interval



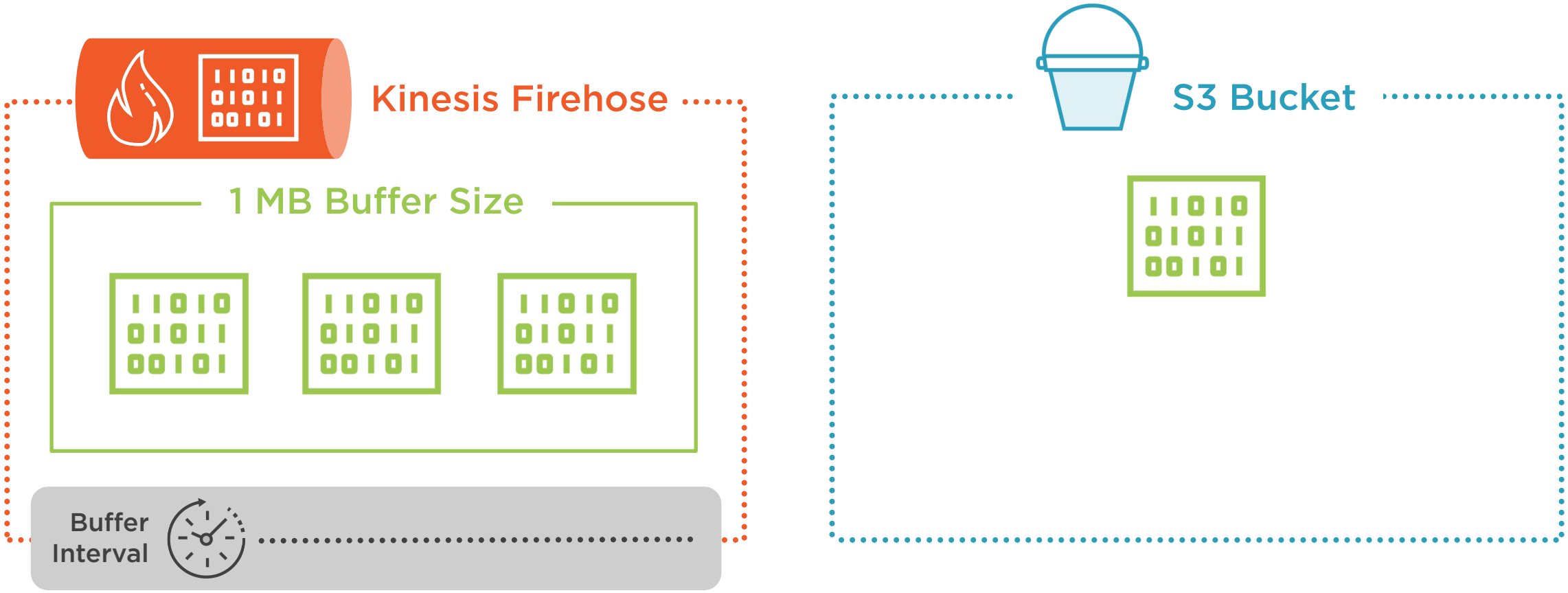
Buffer Size and Interval



Buffer Size and Interval



Buffer Size and Interval



Kinesis Data Firehose Limits

1k to 5k records/second (by region) – Soft limit

1k to 2k transactions/second (by region) – Soft limit

1MiB to 5MiB data ingested per second (by region) – Soft Limit

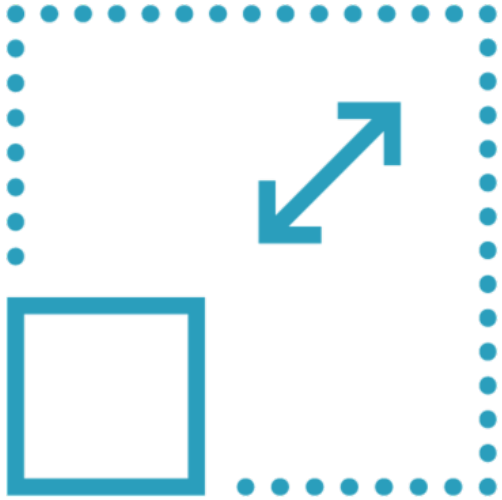
1000 KiB max record size – Hard limit

Kinesis Data Stream Producers – No ingestion limit

Kinesis Data Firehose: Transformations

Transformation Limitations

AWS Lambda only



Buffer size - 3MB

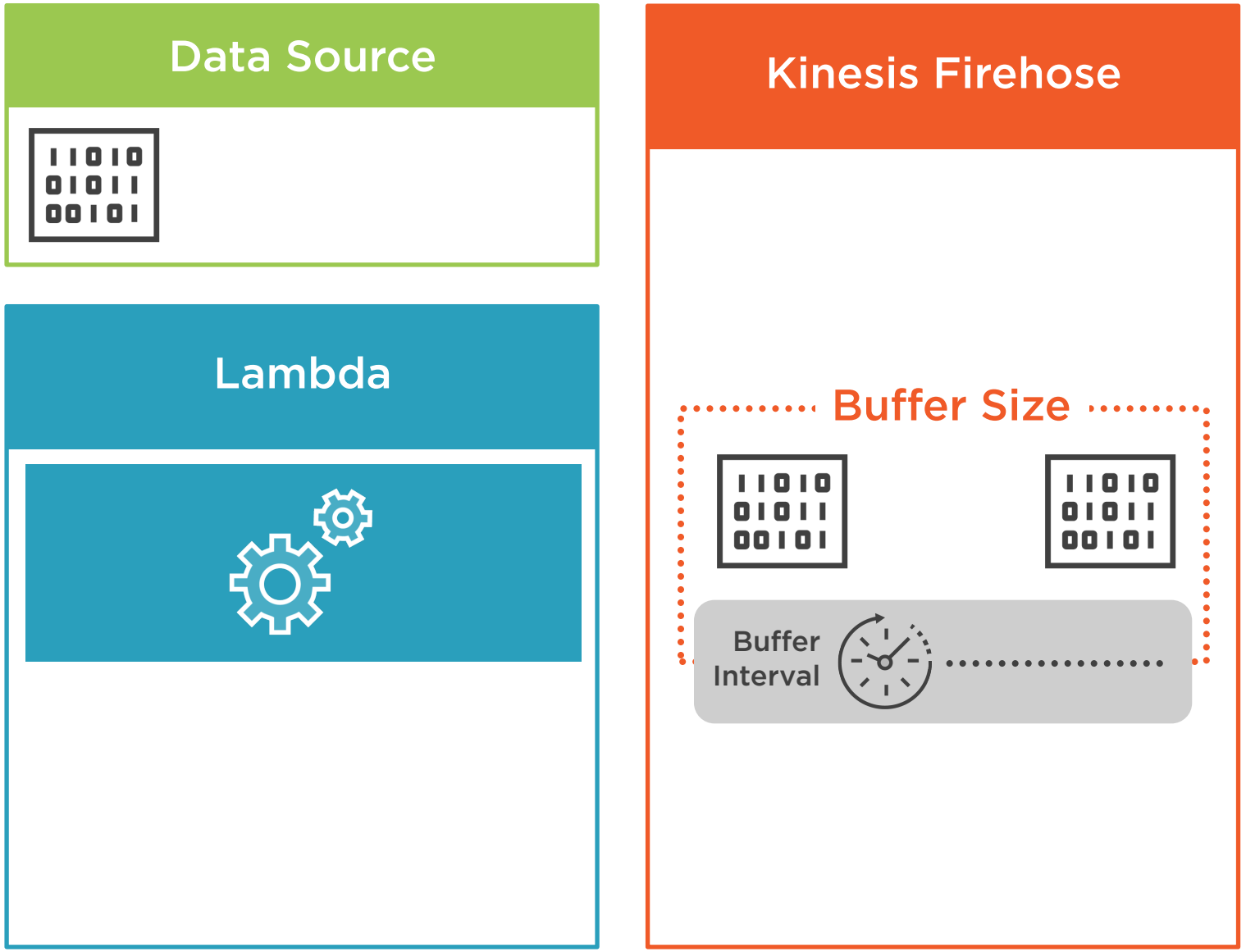
Transformation States

Ok: Processed correctly

Dropped: Intentionally dropped

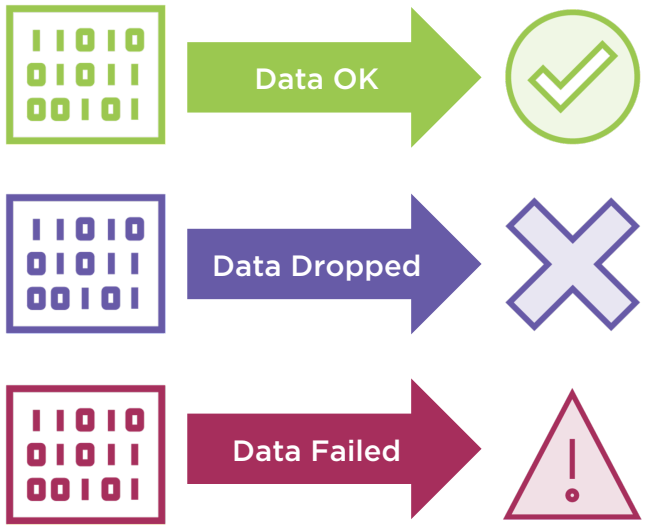
Failed: Can be send to S3 for processing later

Let’s visualize
some transformations


Data Source
Lambda
Kinesis Firehose
Buffer Size
Buffer
Interval

Data Source

Lambda



Kinesis Firehose

Buffer Size

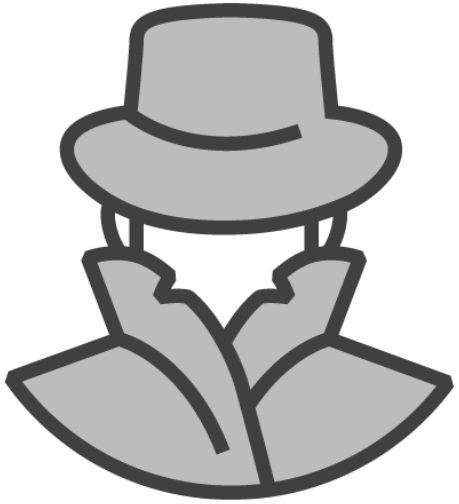
Buffer Interval

S3 Bucket for Failures

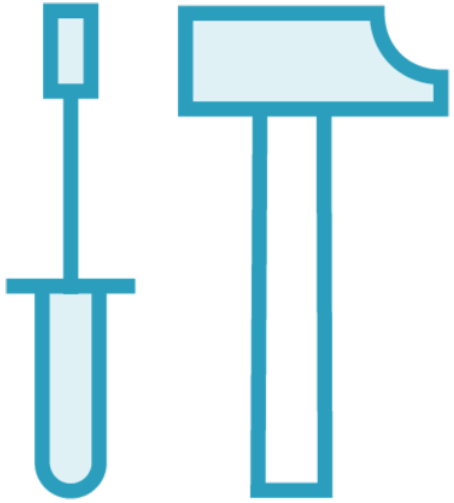
Amazon ES Destination

Kinesis Data Firehose: Producers

Kinesis Data Streams and Firehose Producers



Kinesis Agent

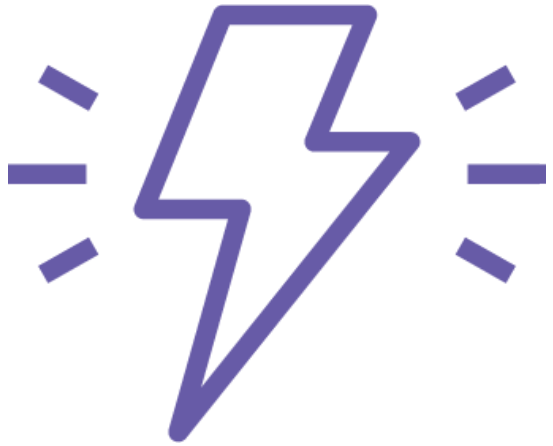


AWS SDKs

Kinesis Data Streams

Kinesis Firehose Producers

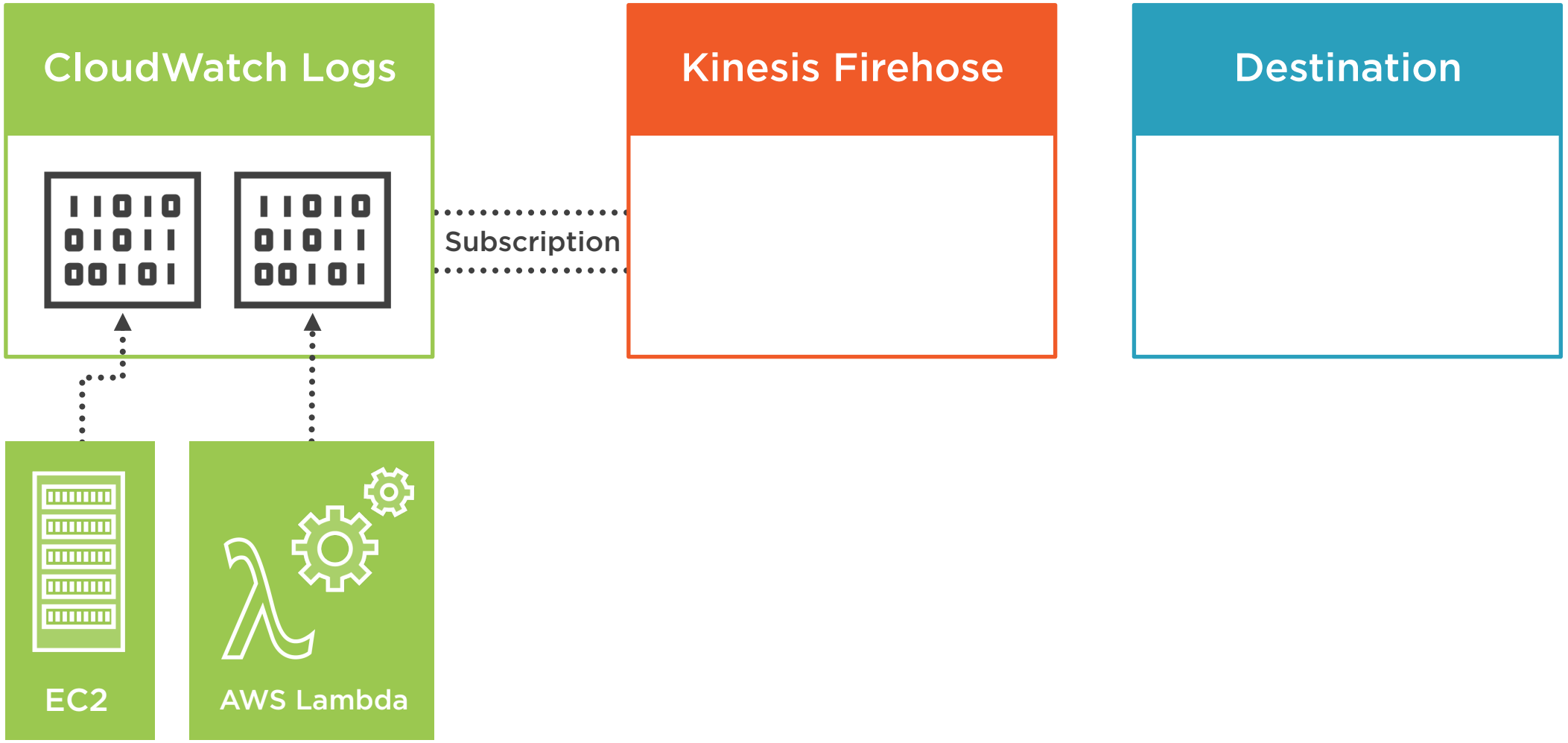
CloudWatch Log Subscriptions



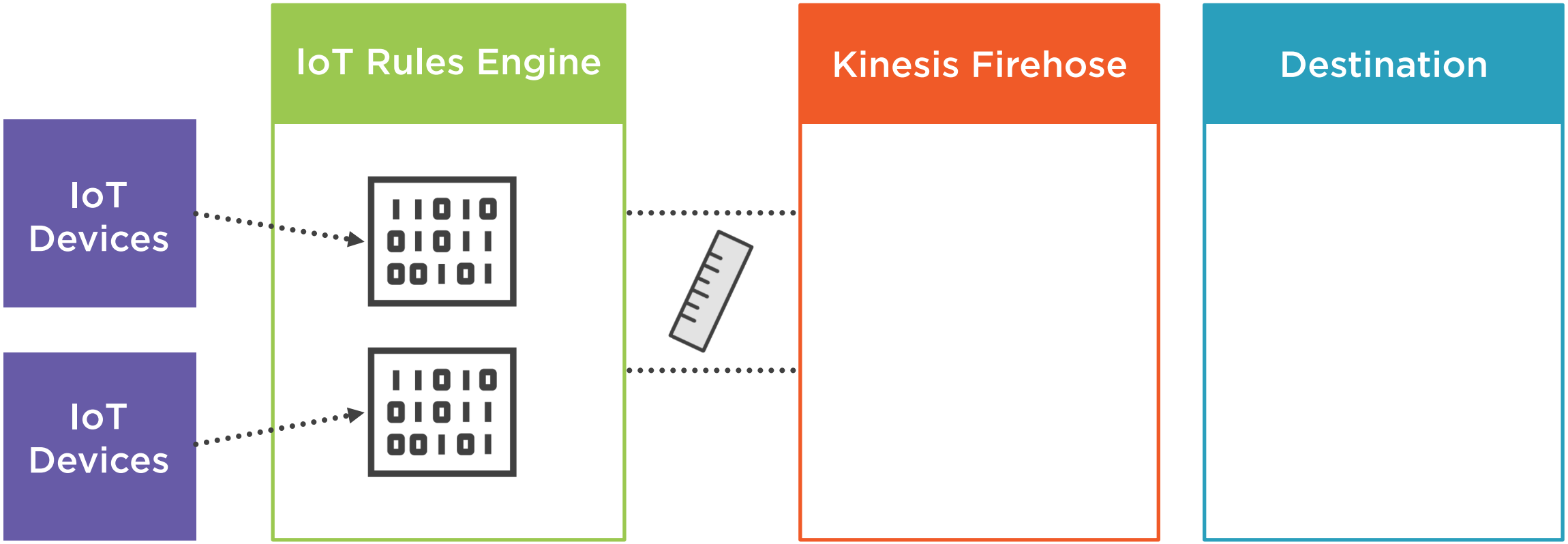
CloudWatch Logs Events Rule

AWS IoT Rules

CloudWatch Log Subscriptions



AWS IoT Rules



Kinesis Data Firehose: Destinations

Firehose Destinations

Amazon S3

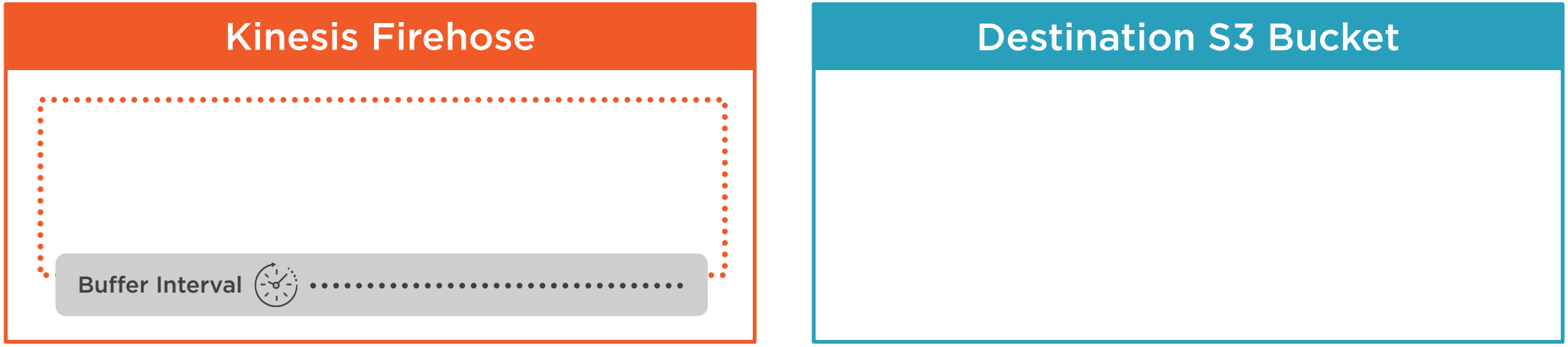
Amazon ES

Redshift

Splunk

Kinesis Firehose to Amazon S3

Kinesis Firehose to Amazon S3



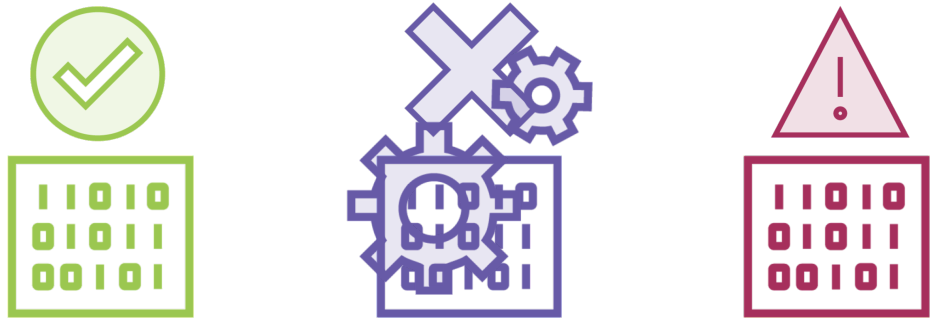
Kinesis Firehose to Amazon S3

Kinesis Firehose

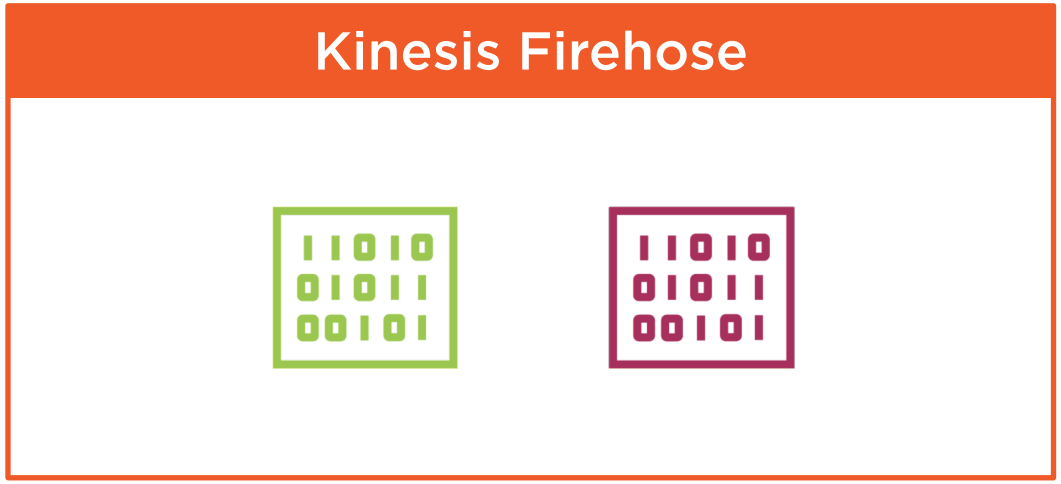
Buffer Interval

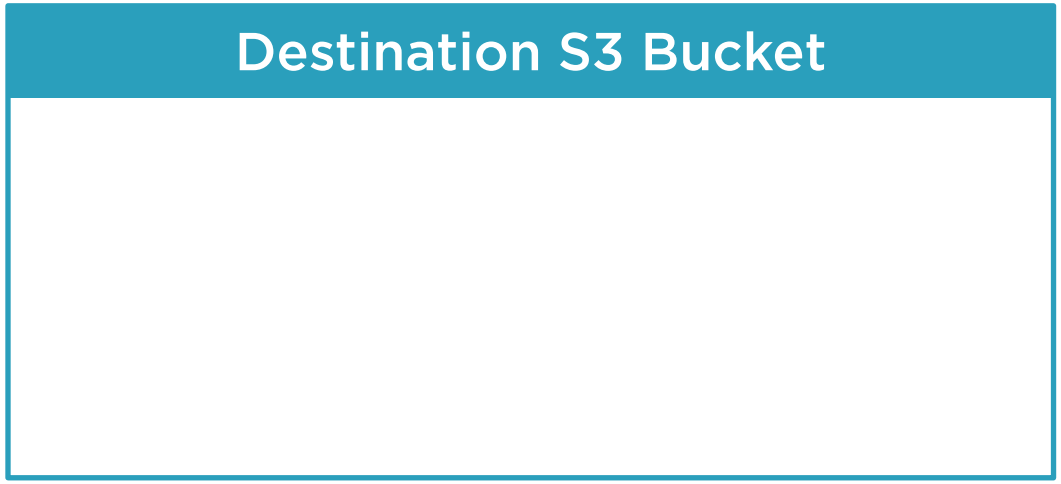
Destination S3 Bucket

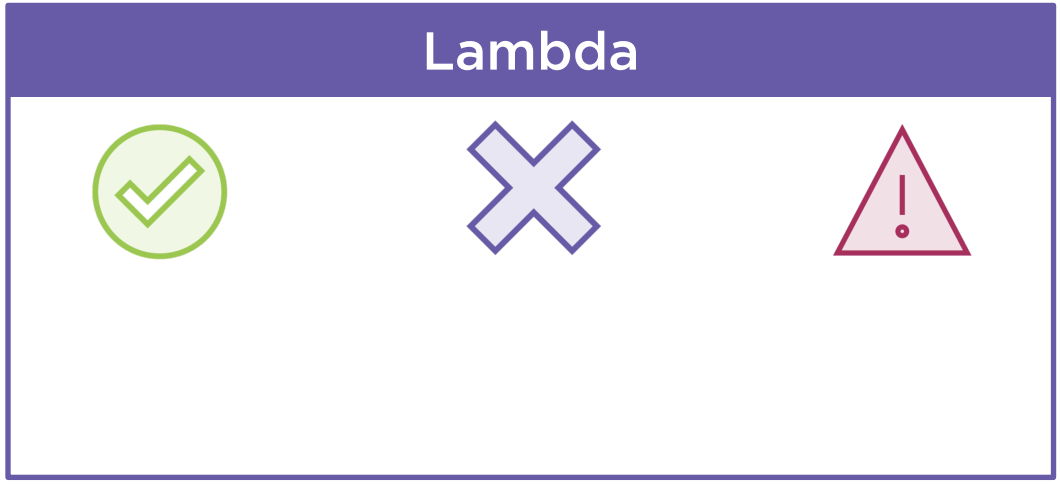
Lambda



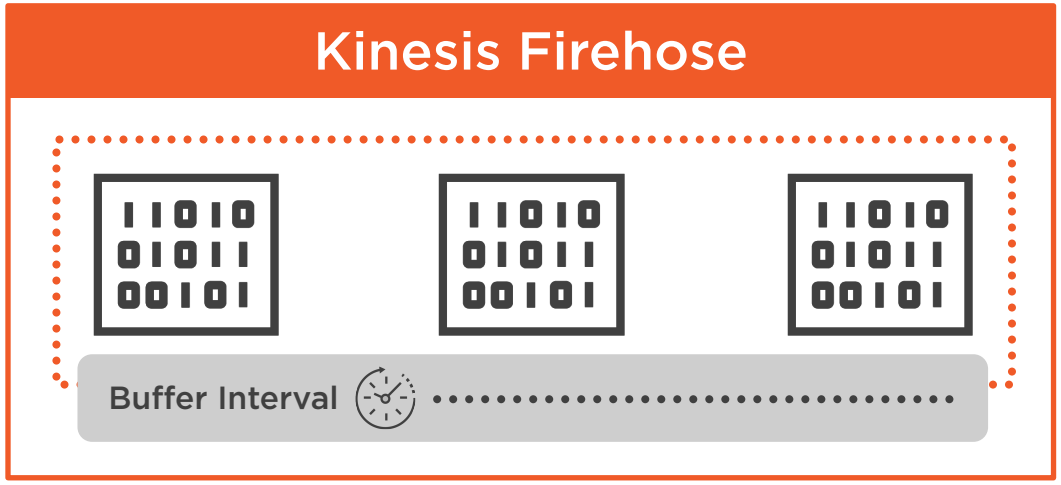
Kinesis Firehose to Amazon S3

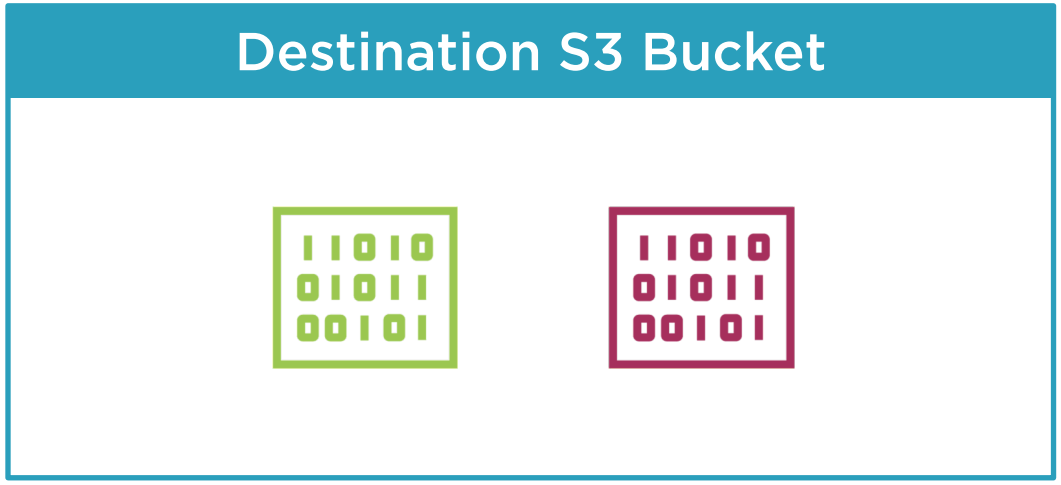


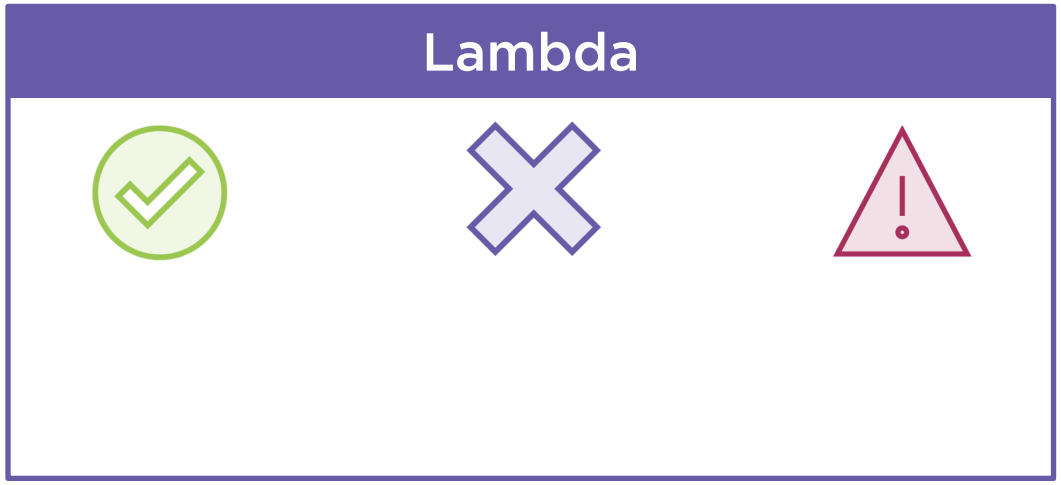


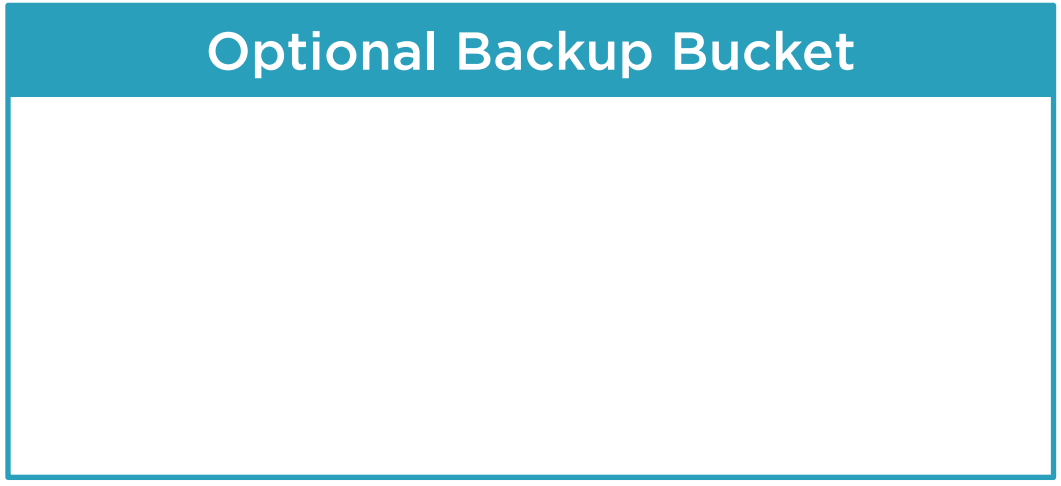


Kinesis Firehose to Amazon S3









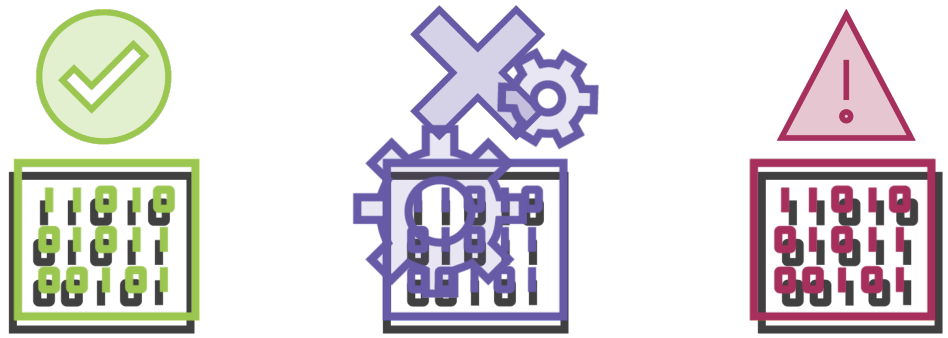
Kinesis Firehose to Amazon S3

Kinesis Firehose

Buffer Interval

Destination S3 Bucket

Lambda



Optional Backup Bucket

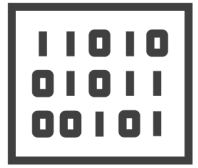
Kinesis Firehose to Amazon S3

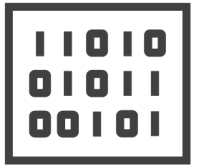
Kinesis Firehose

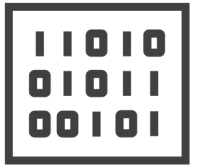
Destination S3 Bucket

Lambda

Optional Backup Bucket



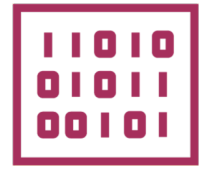




Kinesis Firehose to Amazon S3

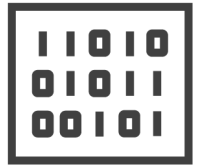
Kinesis Firehose

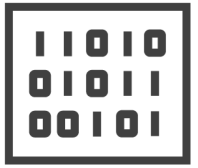
Destination S3 Bucket

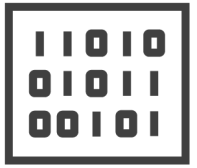


Lambda

Optional Backup Bucket

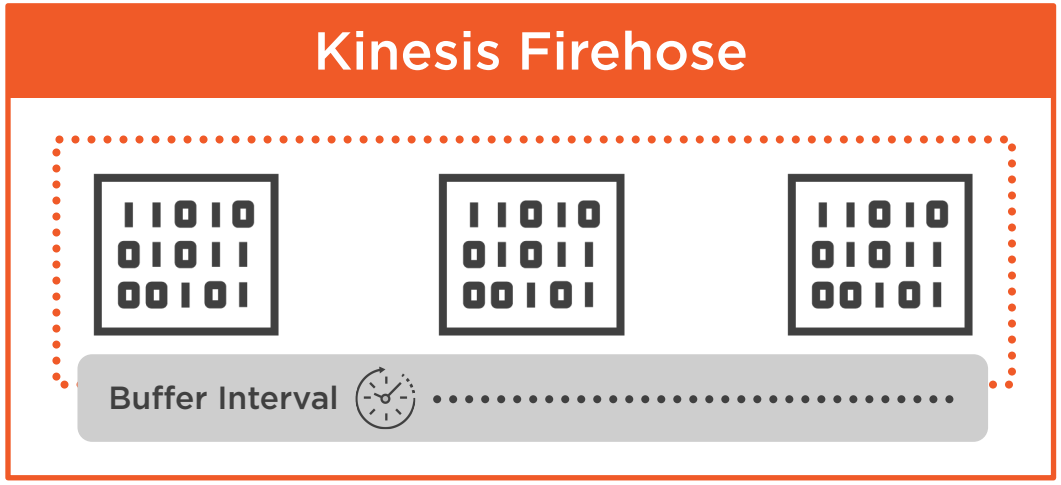


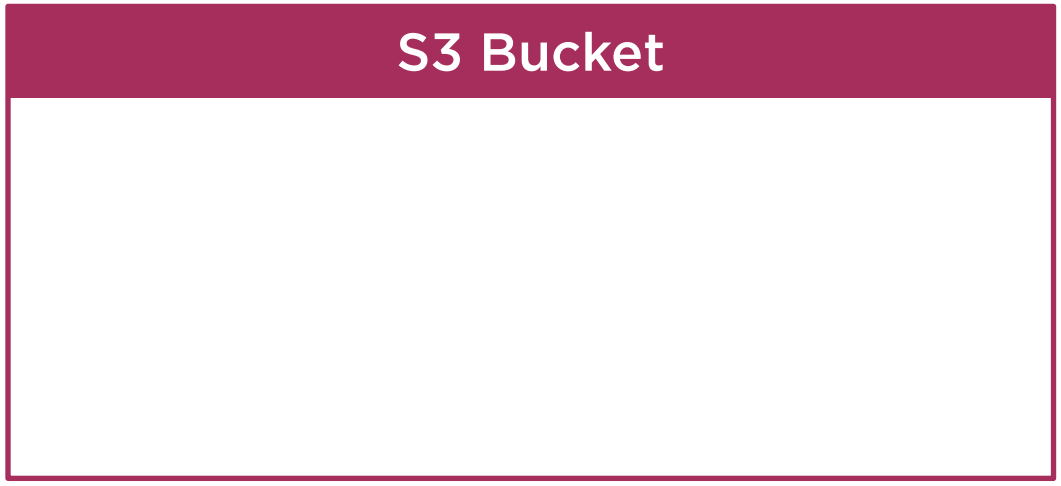




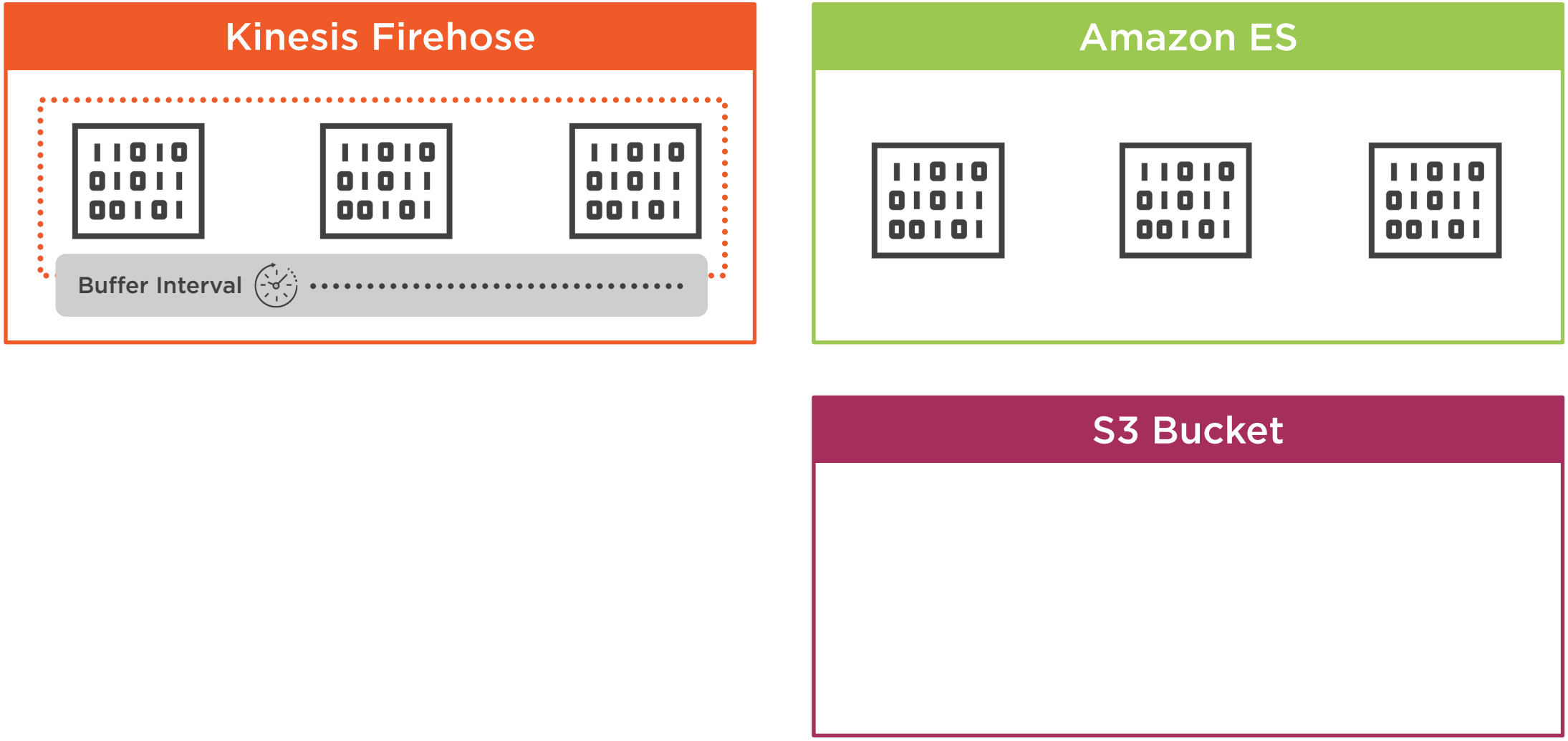
Kinesis Firehose to Amazon ES

Kinesis Firehose to Amazon ES

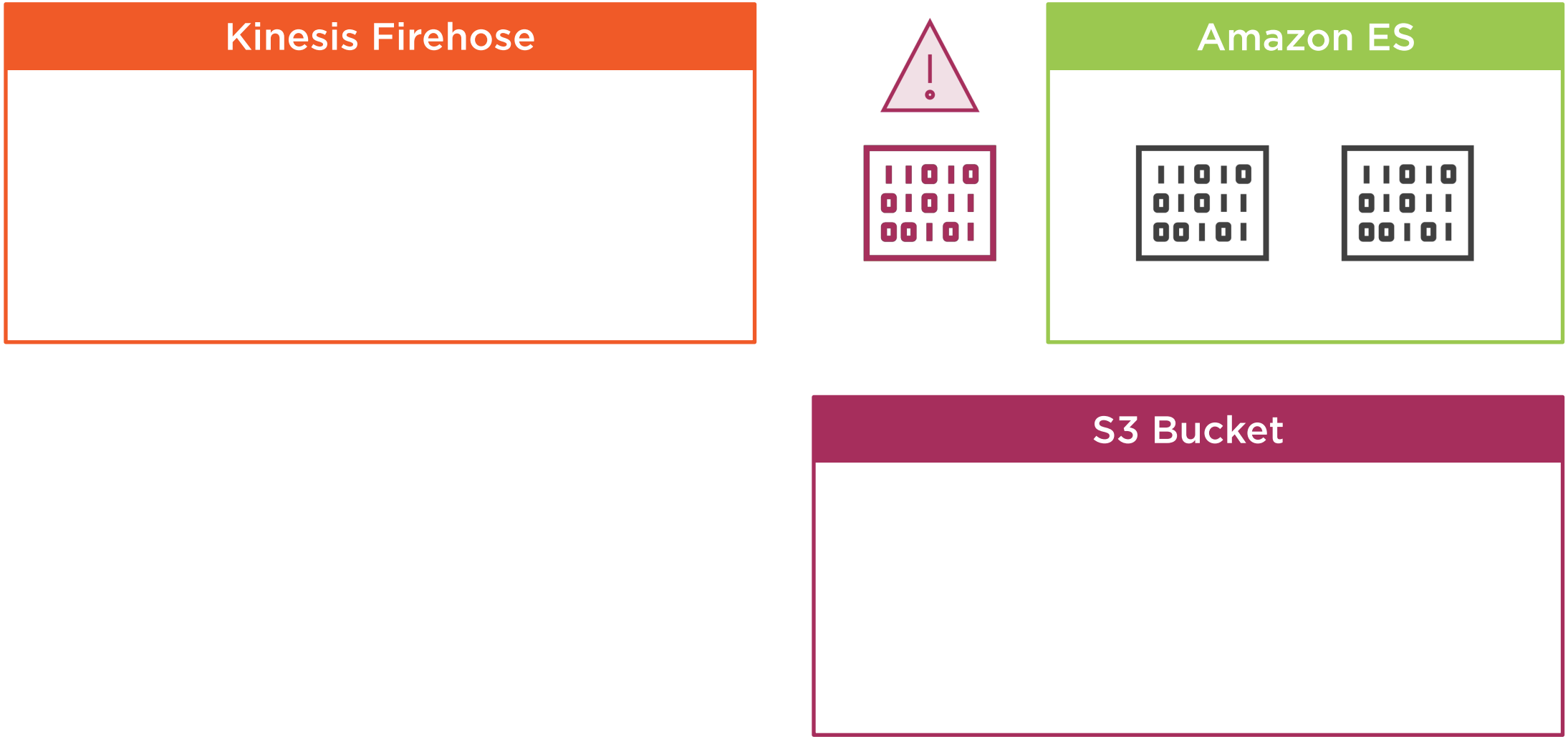




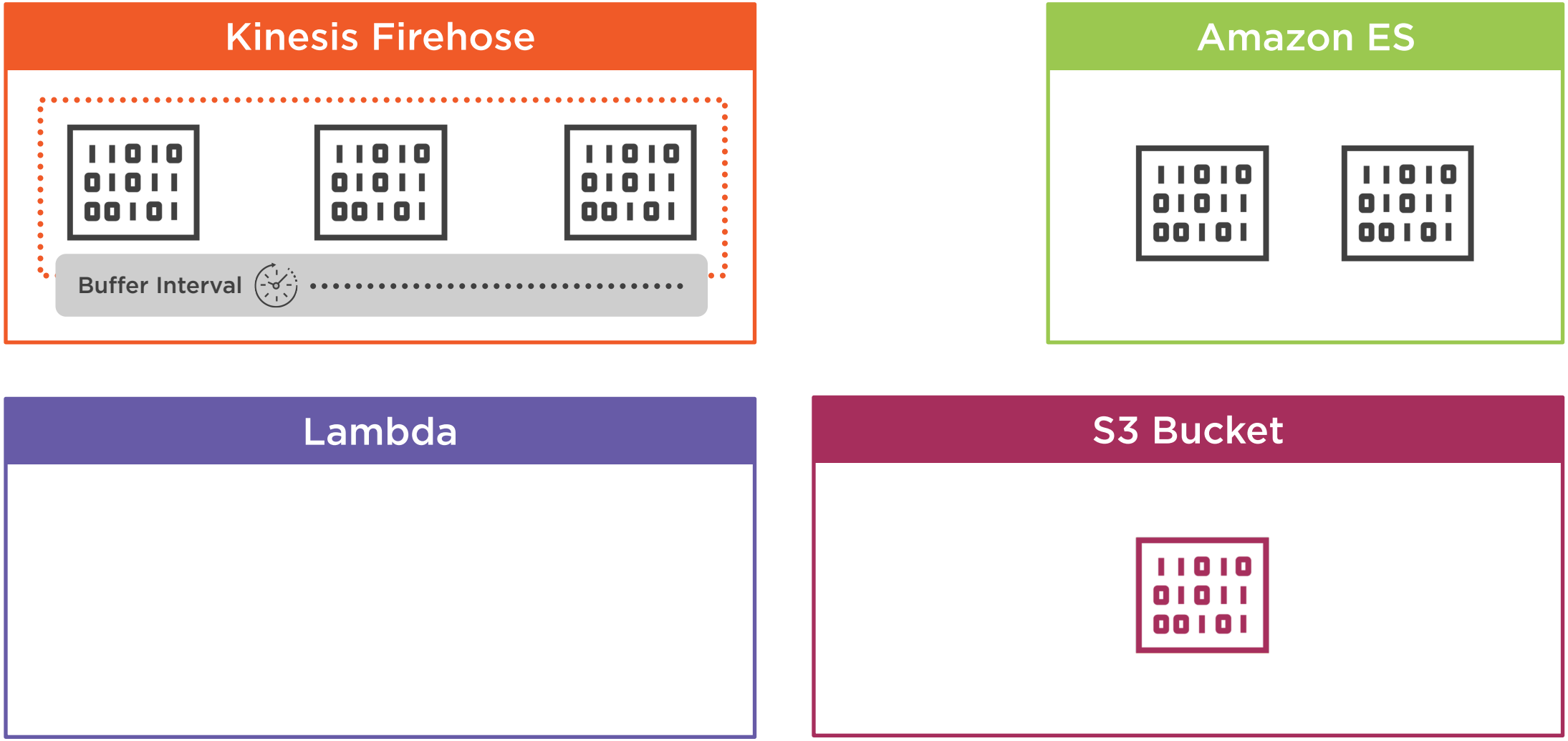
Kinesis Firehose to Amazon ES



Kinesis Firehose to Amazon ES



Kinesis Firehose to Amazon ES



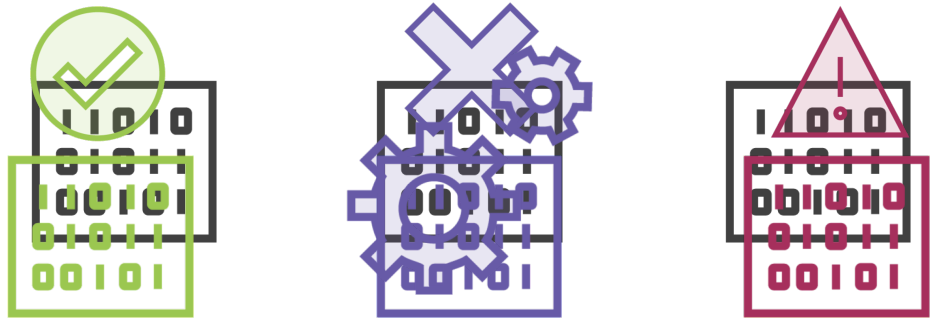
Kinesis Firehose to Amazon ES

Kinesis Firehose

Buffer Interval

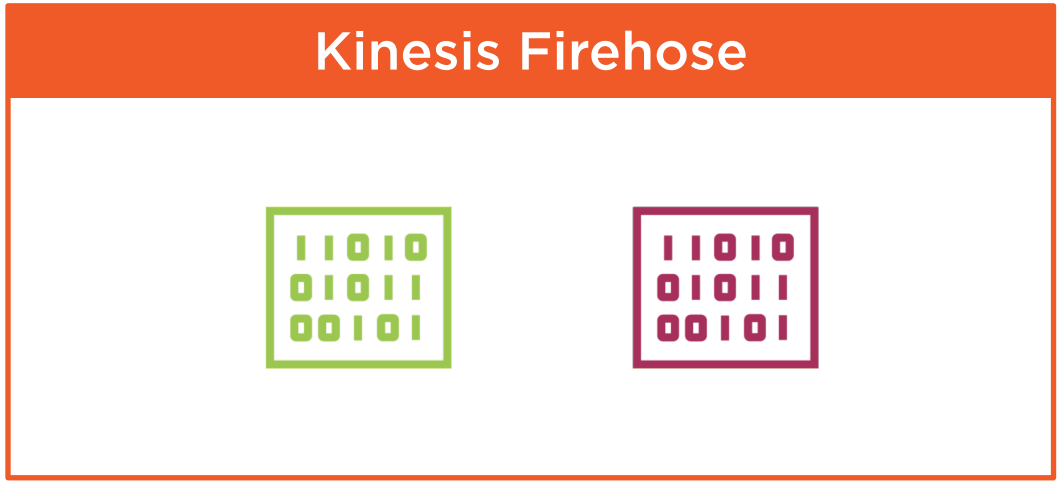
Amazon ES

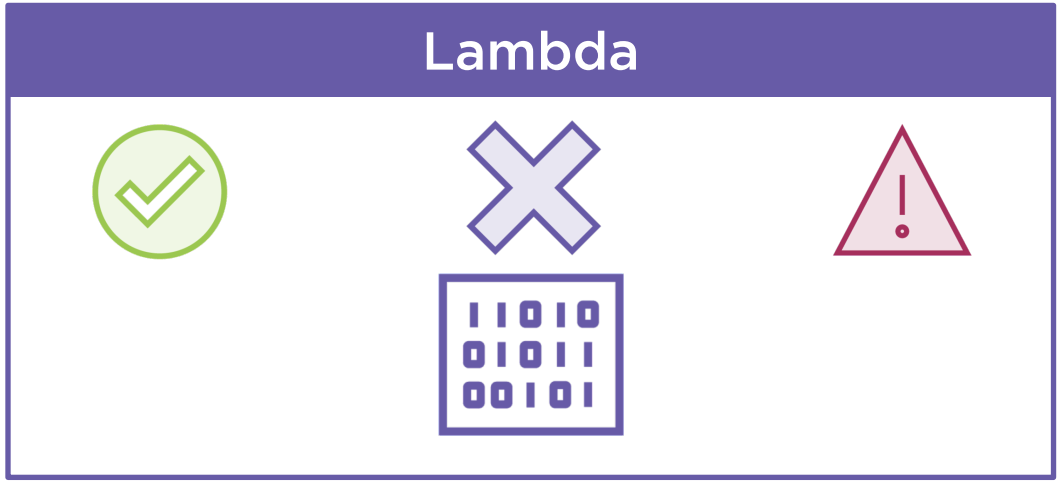
Lambda

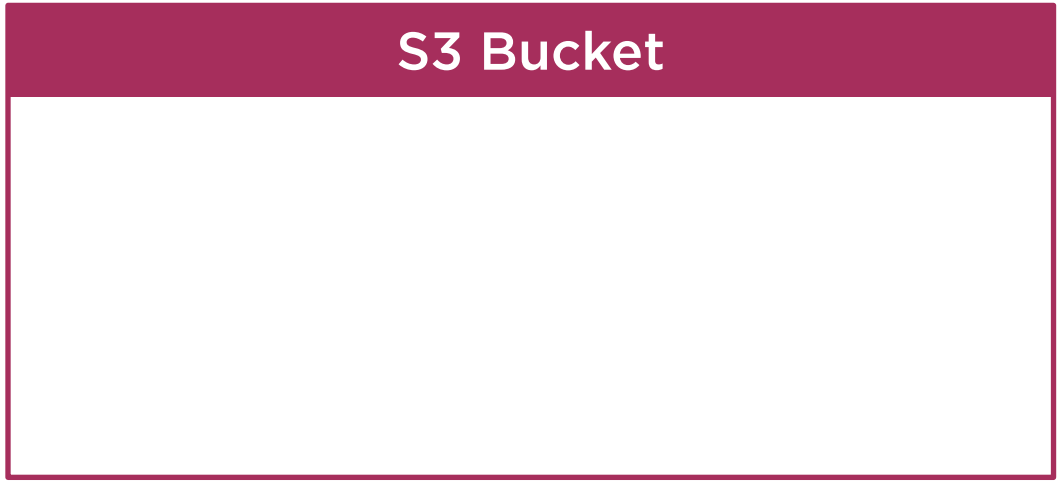


S3 Bucket

Kinesis Firehose to Amazon ES







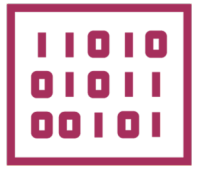
Kinesis Firehose to Amazon ES

Kinesis Firehose

Amazon ES

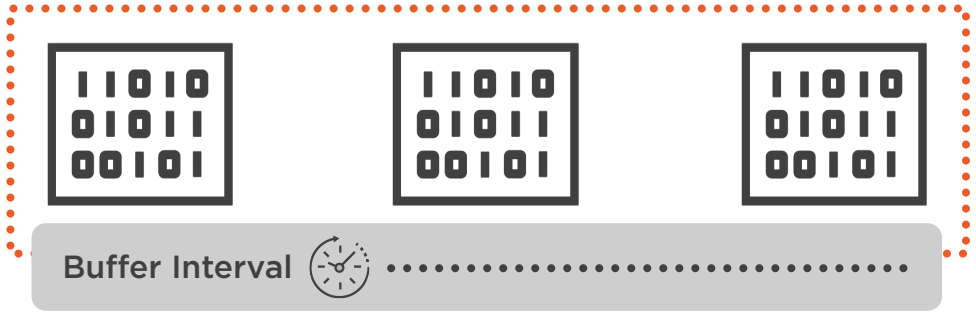
Lambda

S3 Bucket



Kinesis Firehose to Amazon ES

Kinesis Firehose

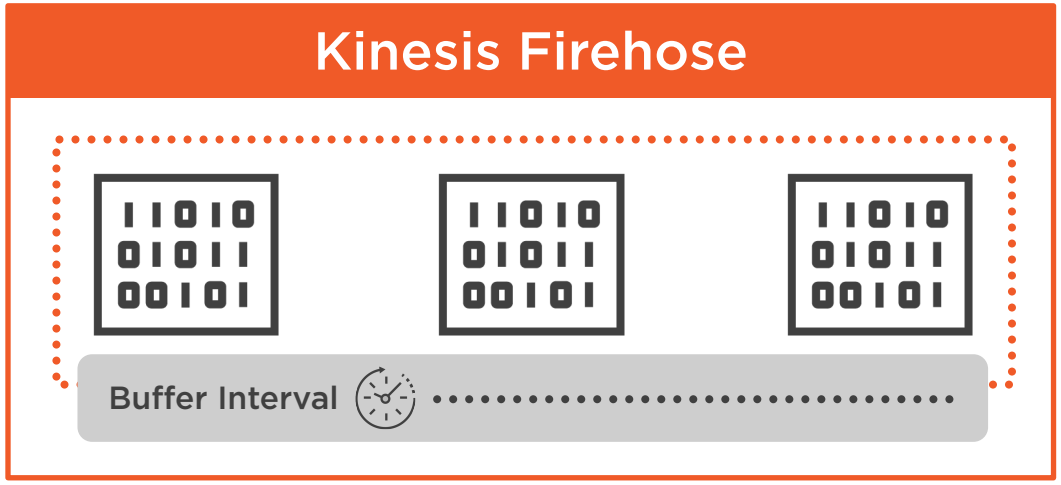


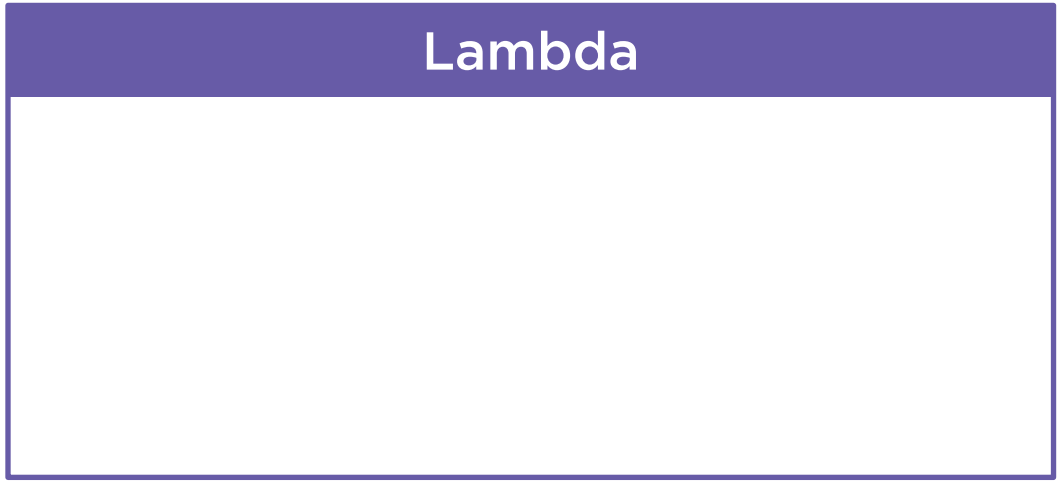
Amazon ES

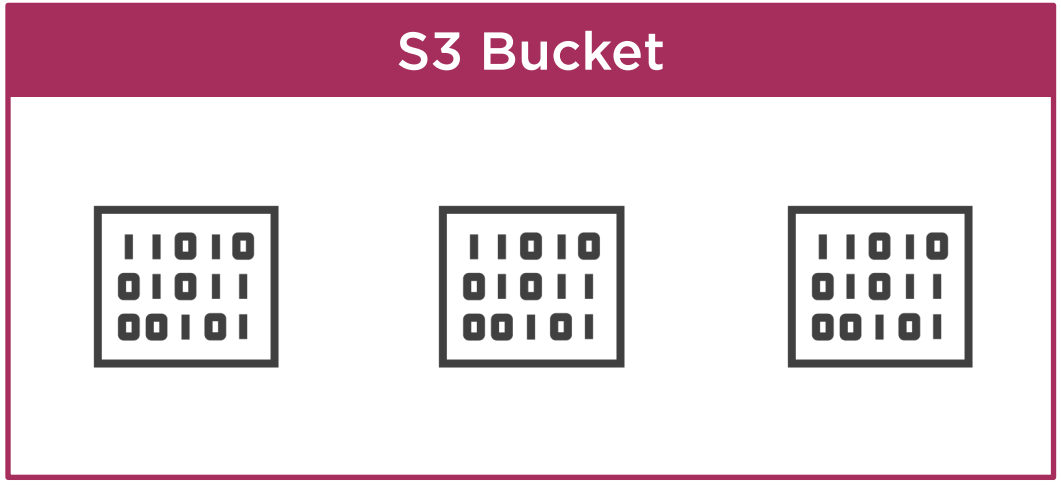
Lambda

S3 Bucket

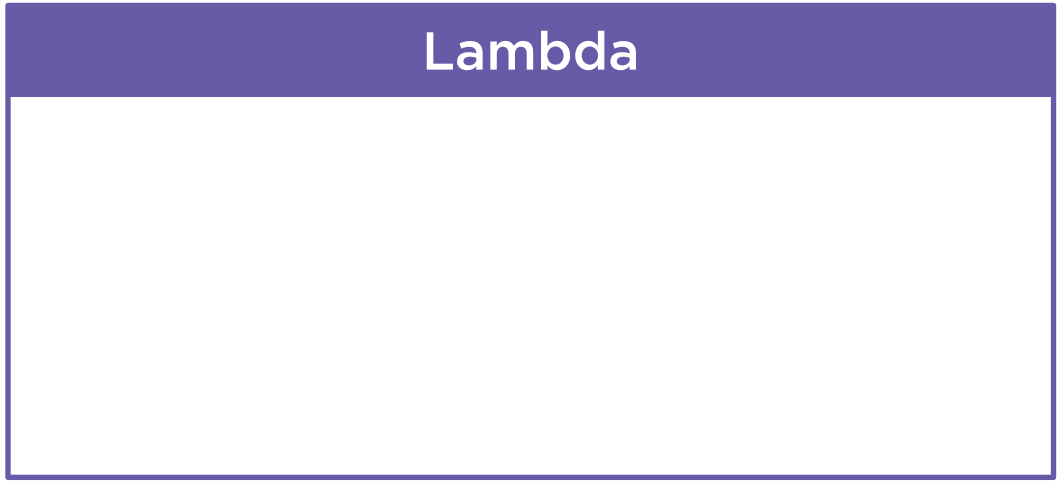
Kinesis Firehose to Amazon ES

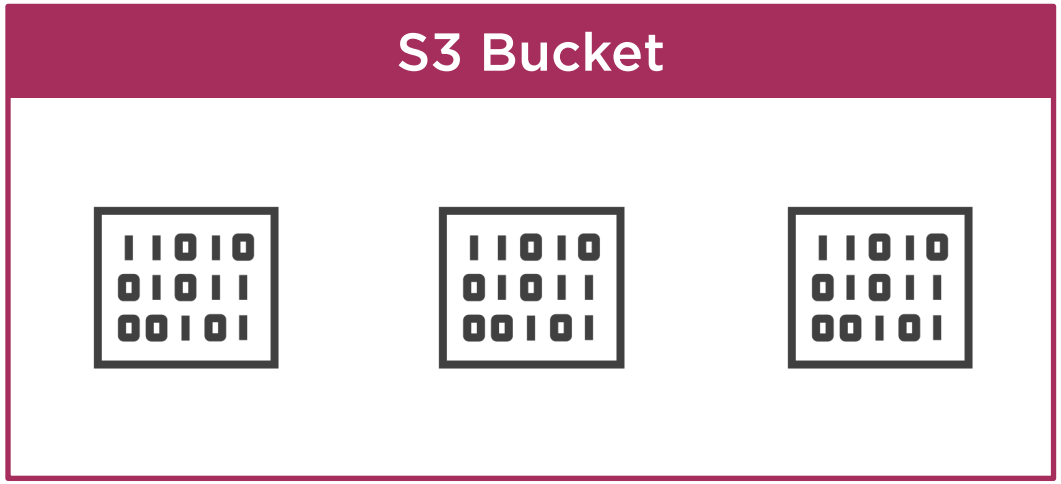






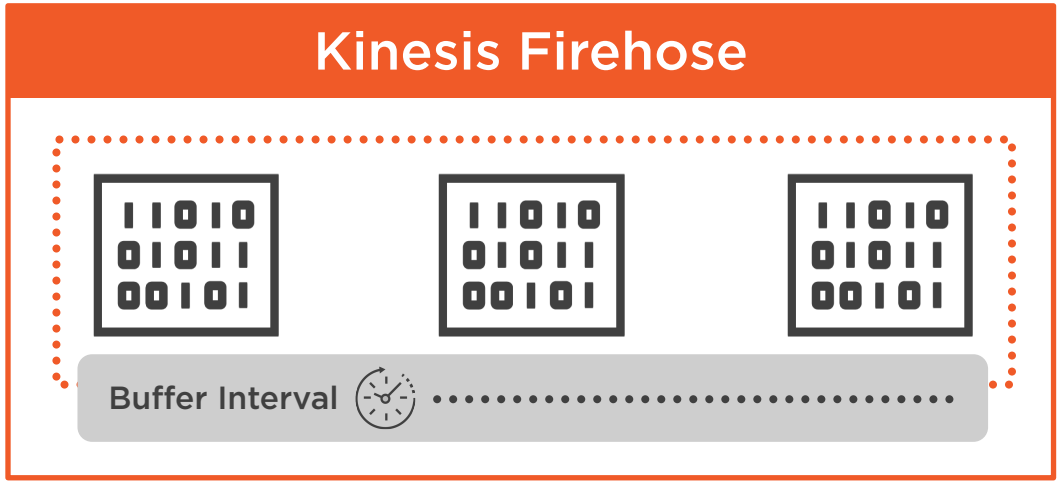
Kinesis Firehose to Amazon ES

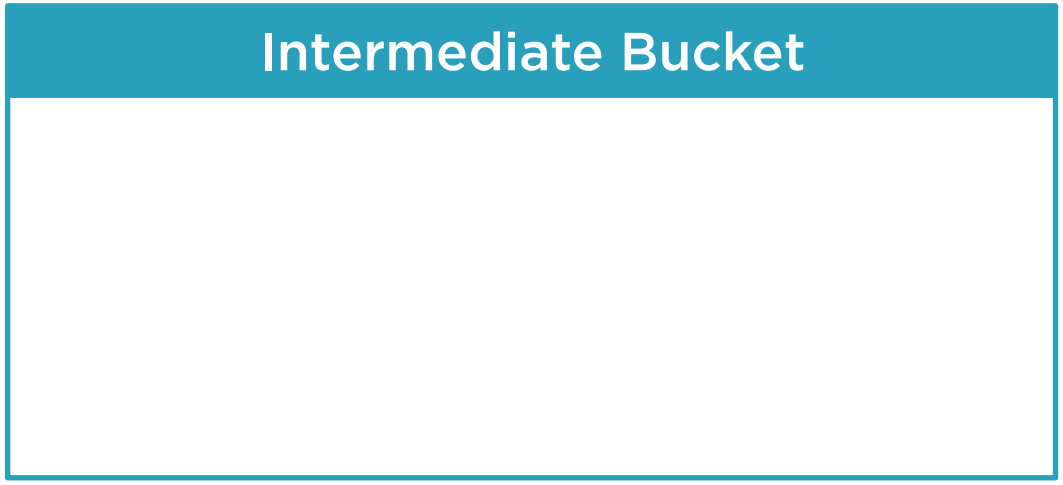


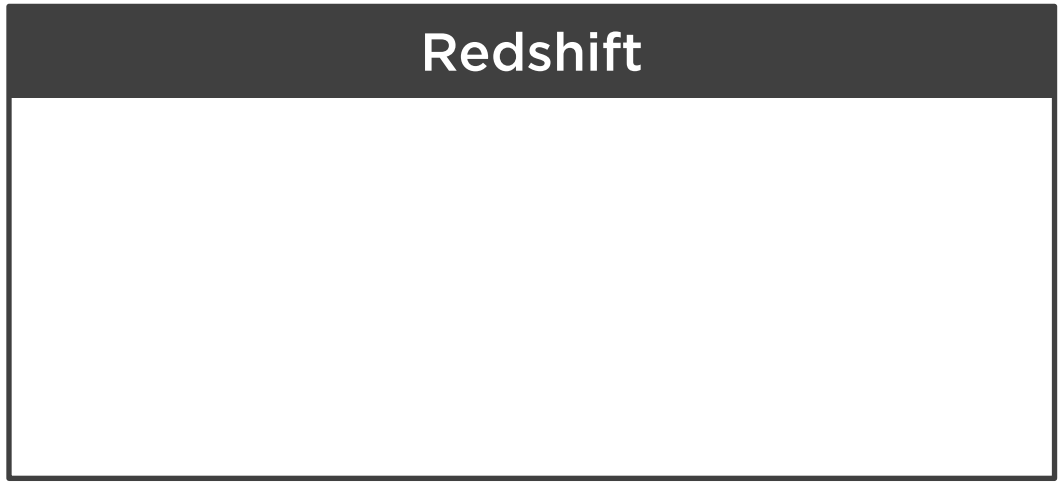


Kinesis Firehose to Amazon Redshift

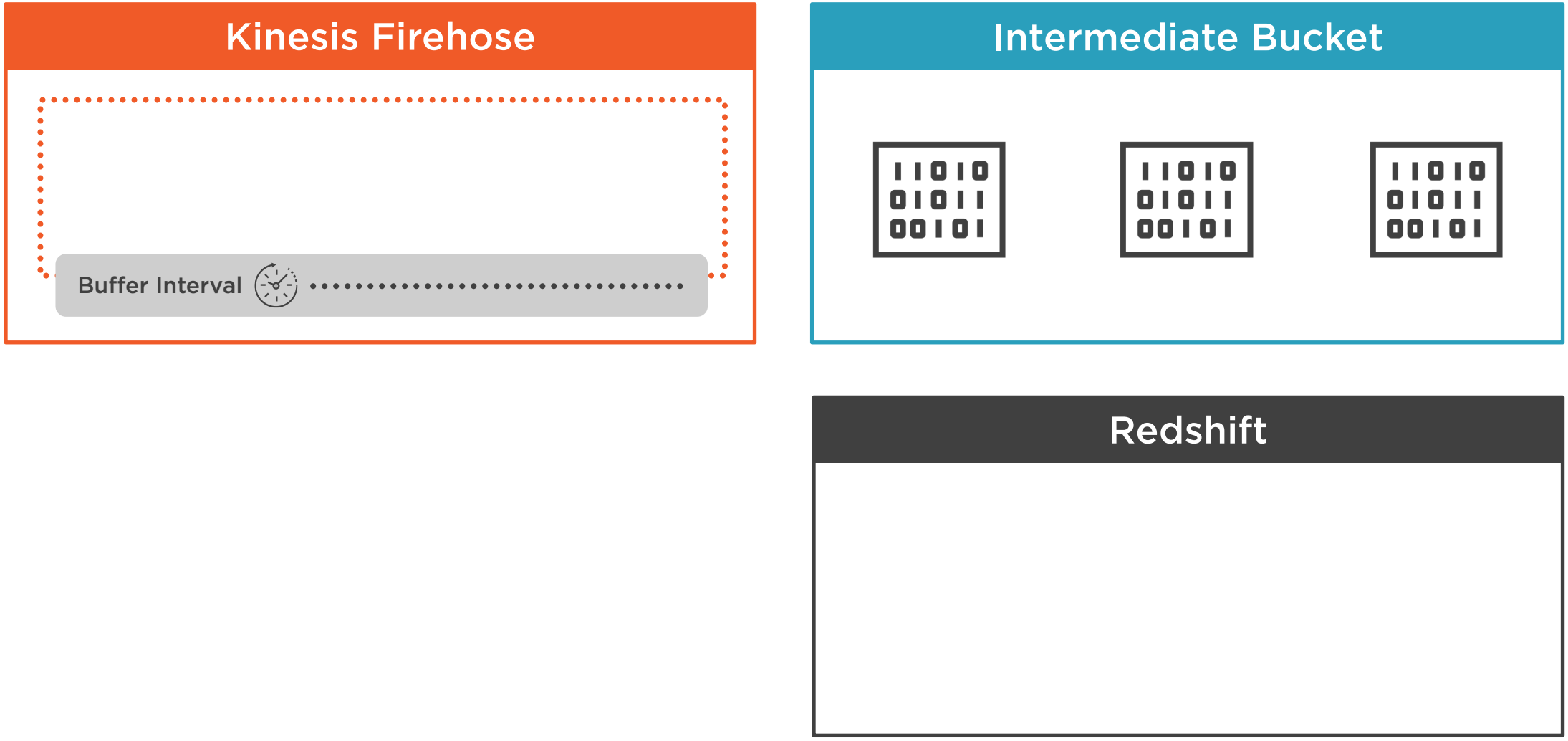
Kinesis Firehose to Amazon Redshift



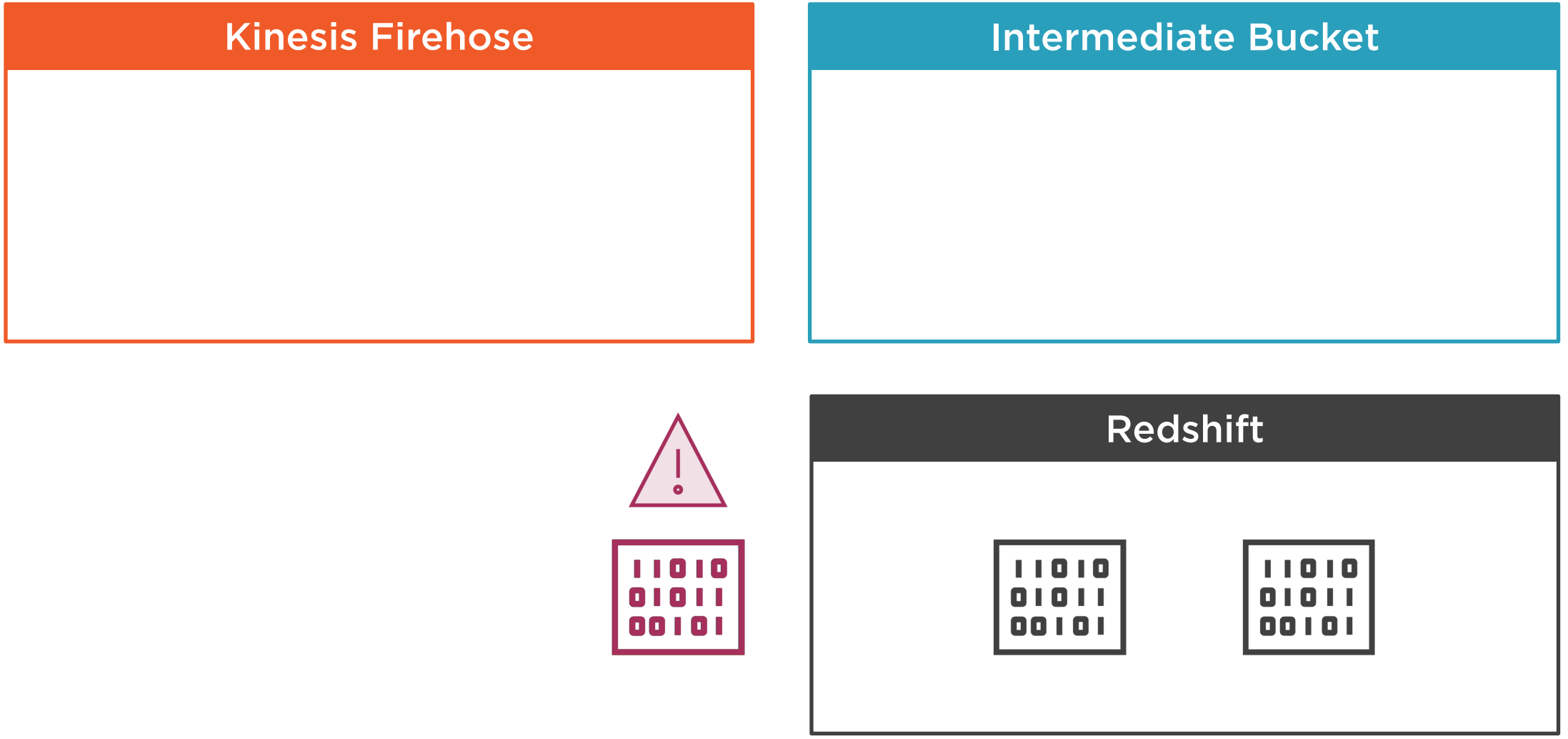




Kinesis Firehose to Amazon Redshift



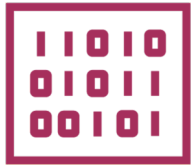
Kinesis Firehose to Amazon Redshift



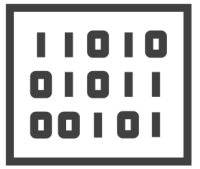
Kinesis Firehose to Amazon Redshift

Kinesis Firehose

Intermediate Bucket

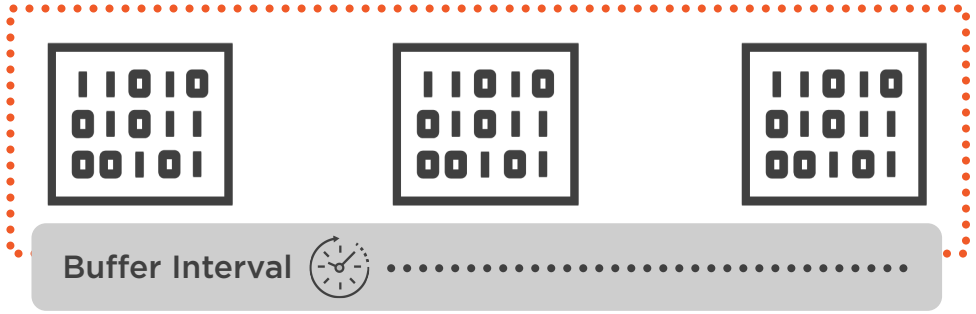


Redshift



Kinesis Firehose to Amazon Redshift

Kinesis Firehose



Intermediate Bucket

Lambda

Redshift

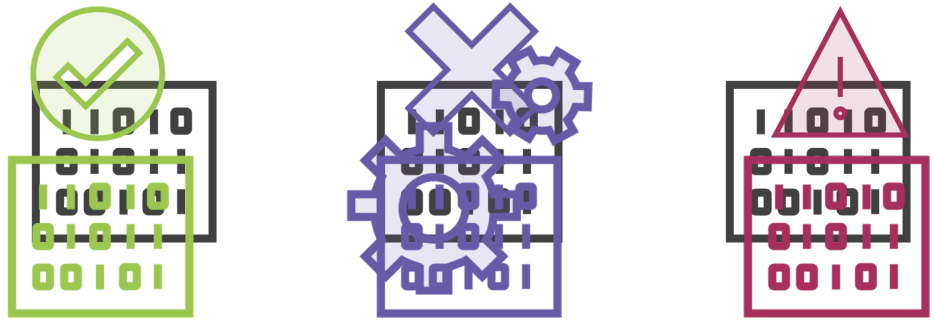
Kinesis Firehose to Amazon Redshift

Kinesis Firehose

Buffer Interval

Intermediate Bucket

Lambda



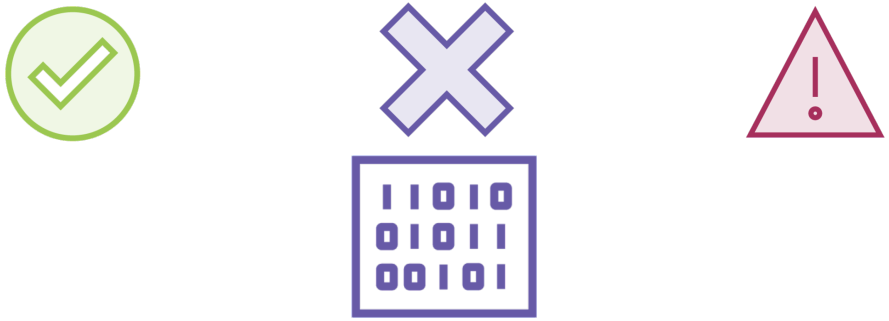
Redshift

Kinesis Firehose to Amazon Redshift

Kinesis Firehose

Intermediate Bucket

Lambda



Redshift

Kinesis Firehose to Amazon Redshift

Kinesis Firehose

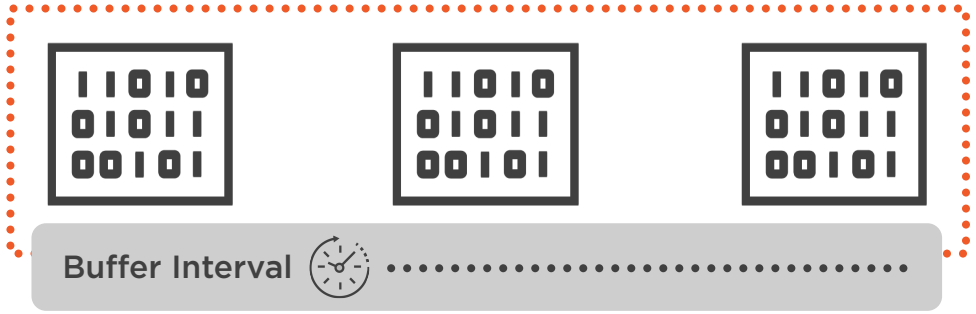
Intermediate Bucket

Lambda

Redshift

Kinesis Firehose to Amazon Redshift

Kinesis Firehose



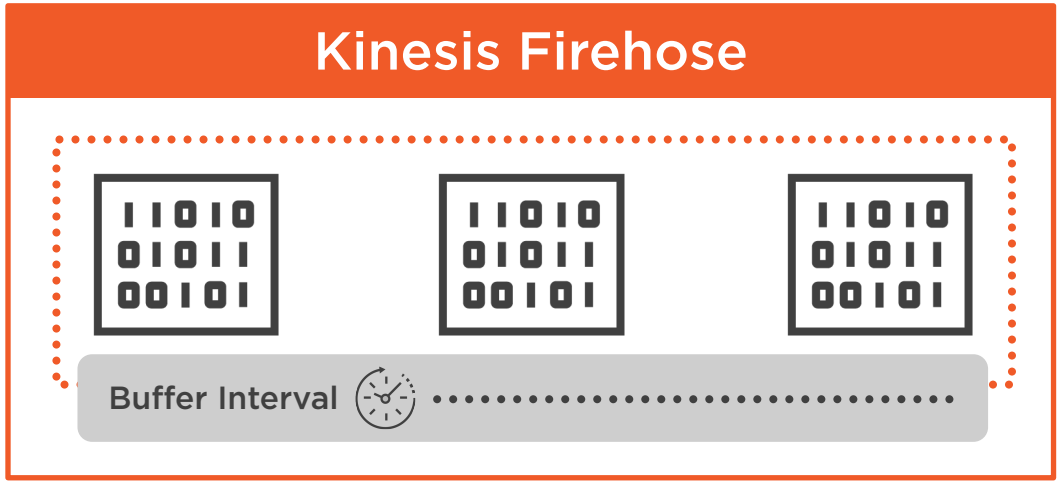
Intermediate Bucket

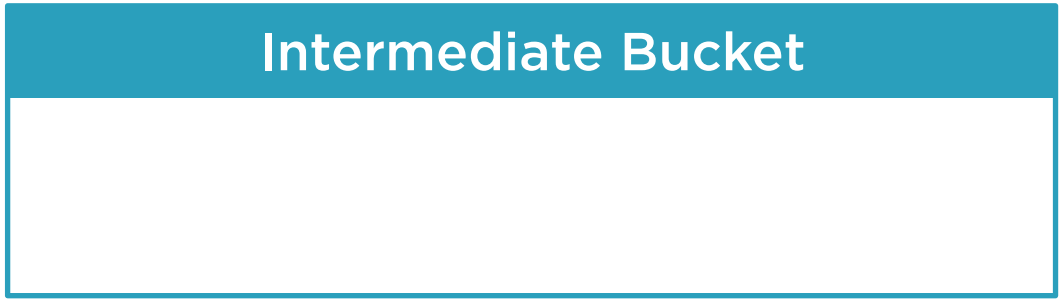
Backup Bucket

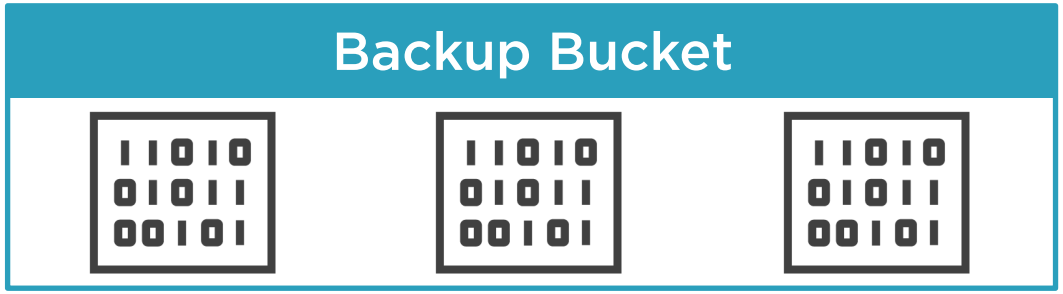
Lambda

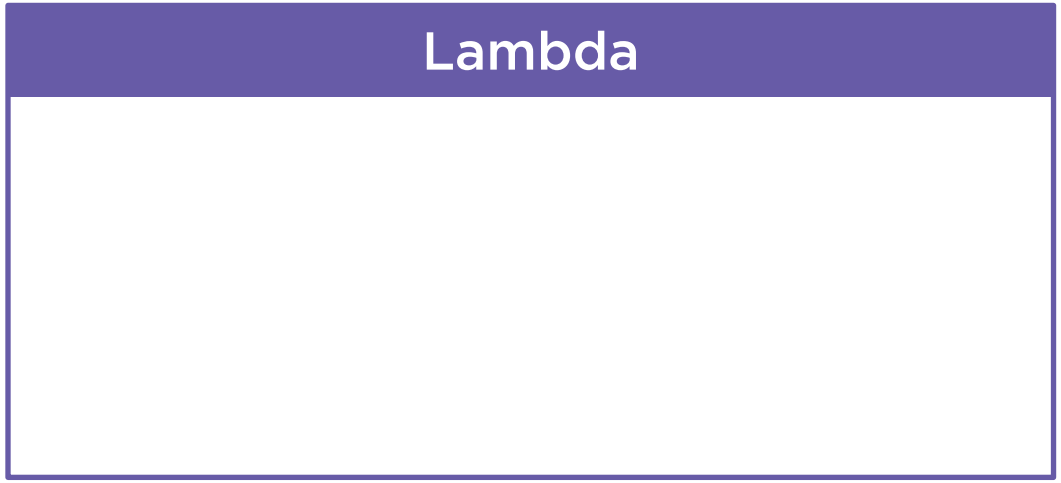
Redshift

Kinesis Firehose to Amazon Redshift







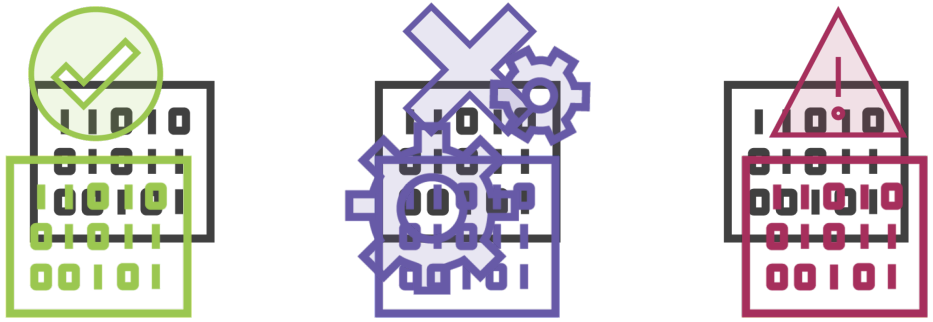


Kinesis Firehose to Amazon Redshift

Kinesis Firehose

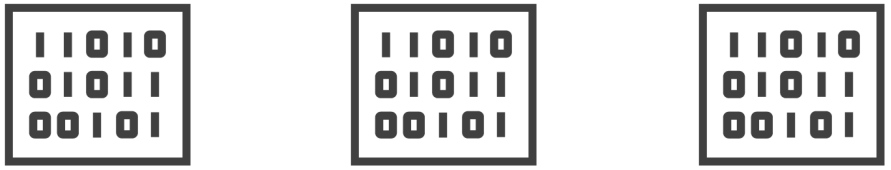
Buffer Interval

Lambda



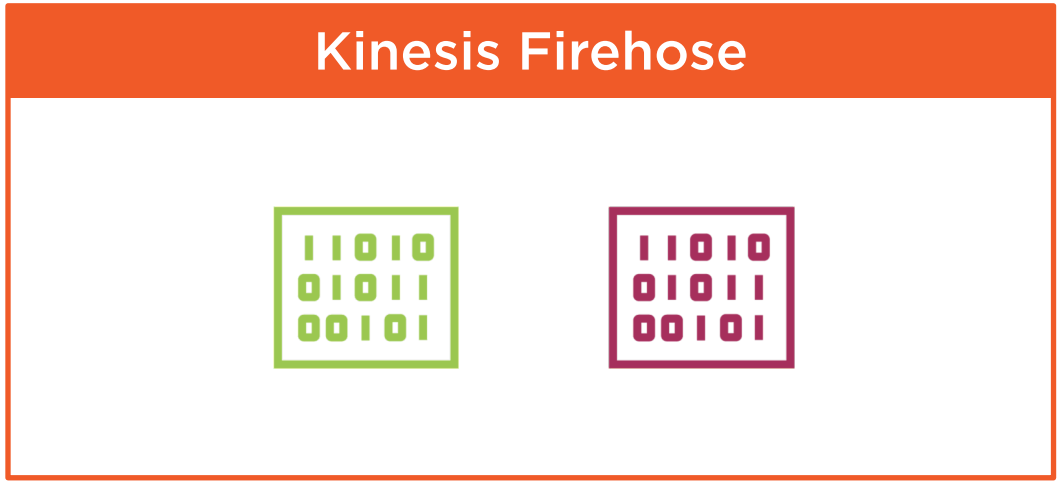
Intermediate Bucket

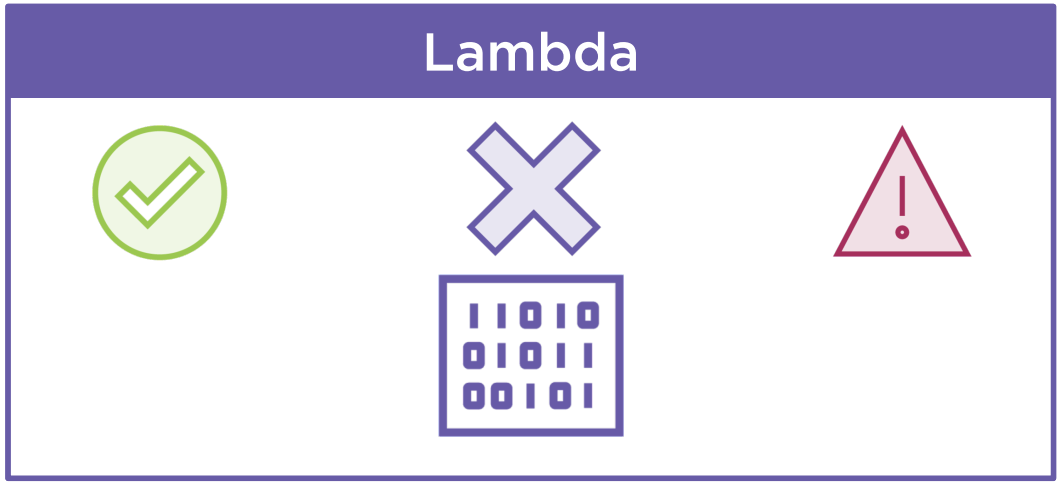
Backup Bucket

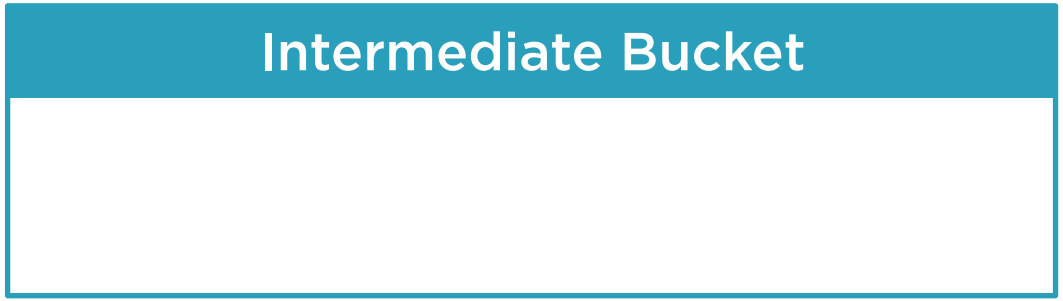


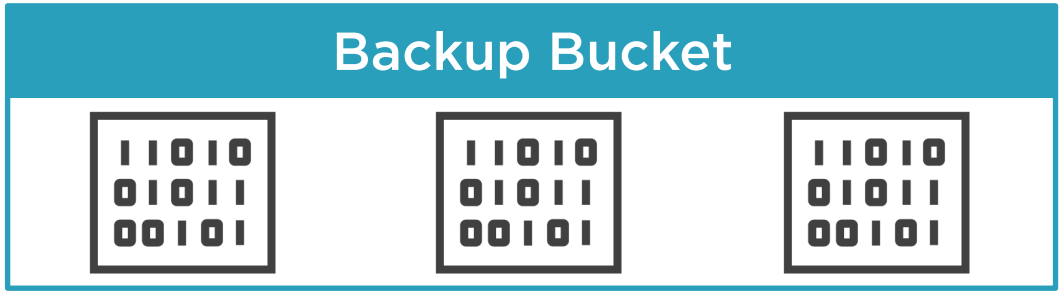
Redshift

Kinesis Firehose to Amazon Redshift

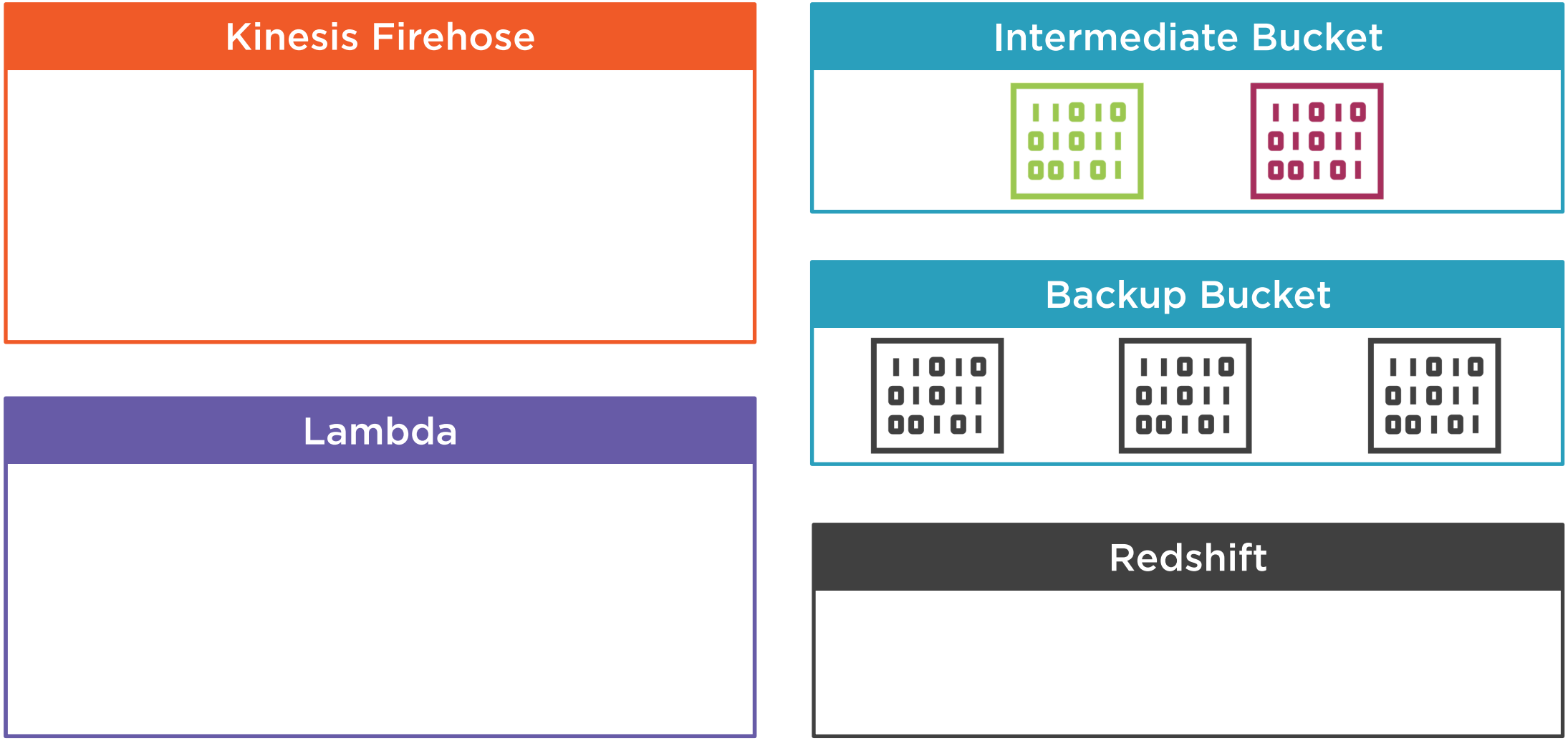






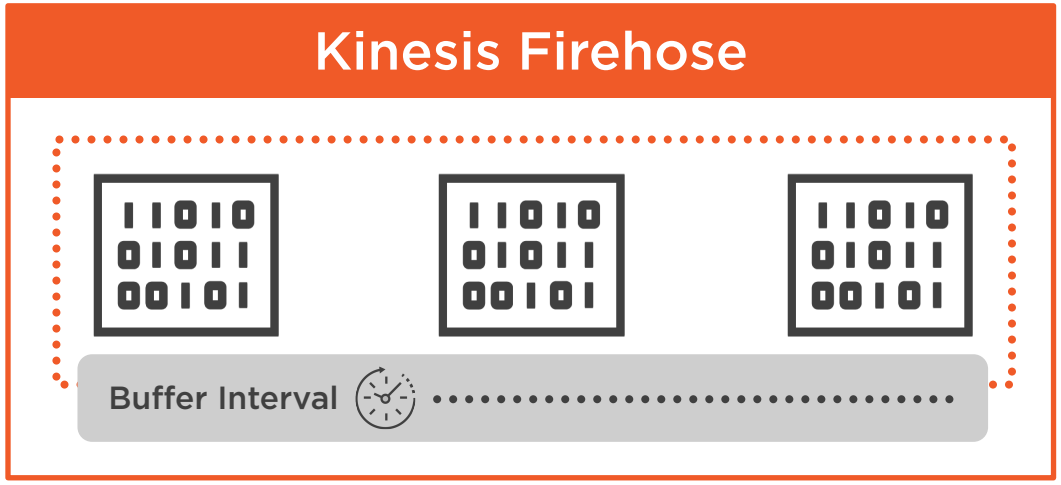


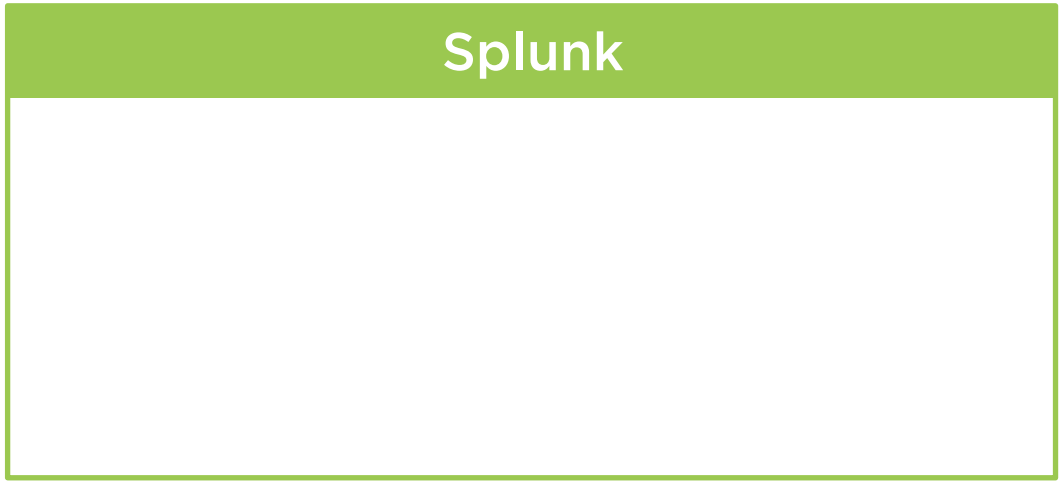
Kinesis Firehose to Amazon Redshift

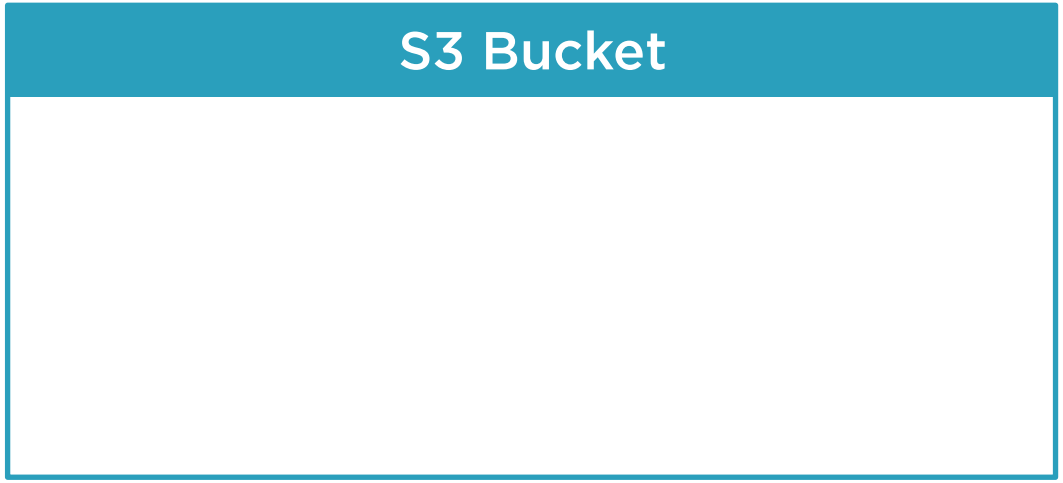


Kinesis Firehose to Splunk

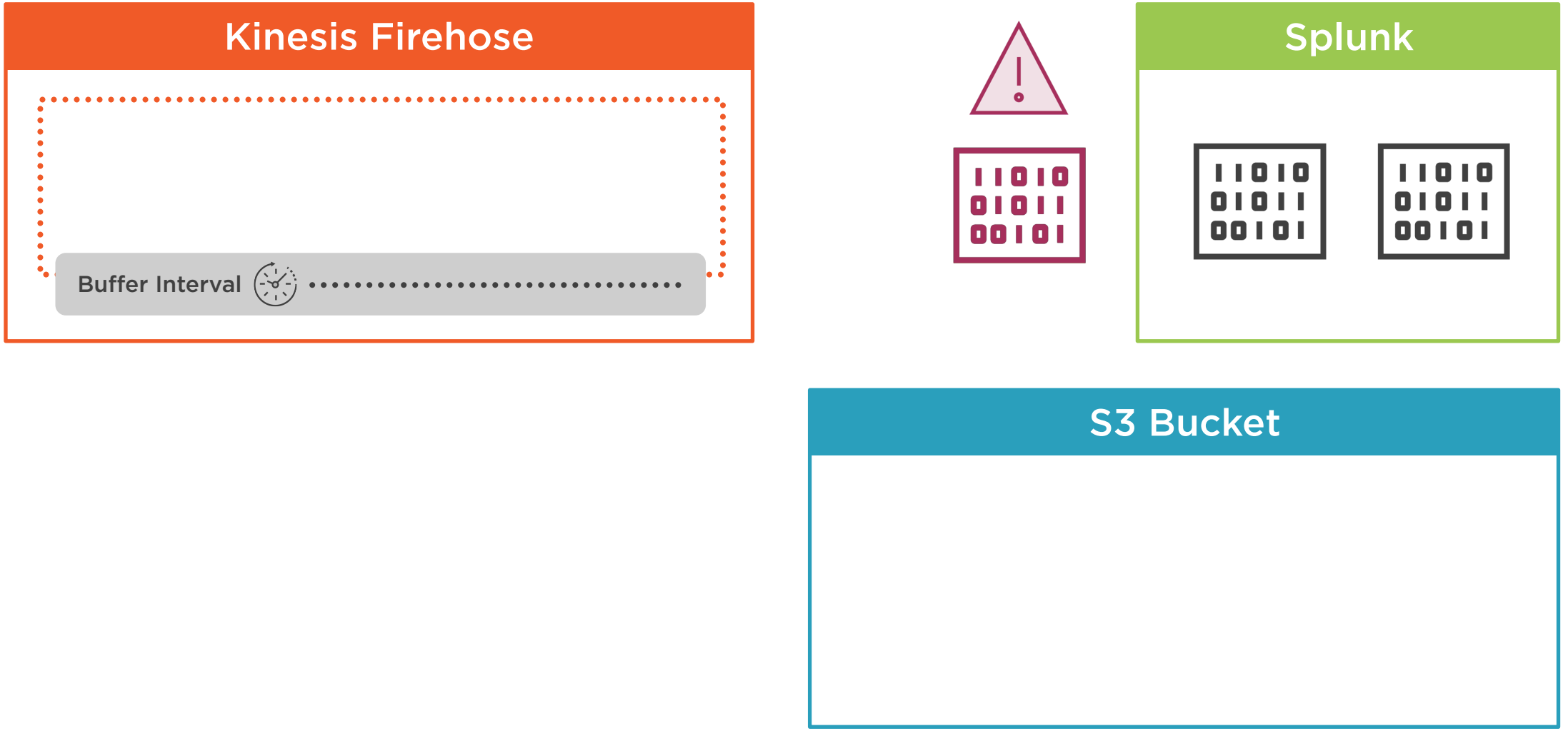
Kinesis Firehose to Splunk







Kinesis Firehose to Splunk



Kinesis Firehose to Splunk

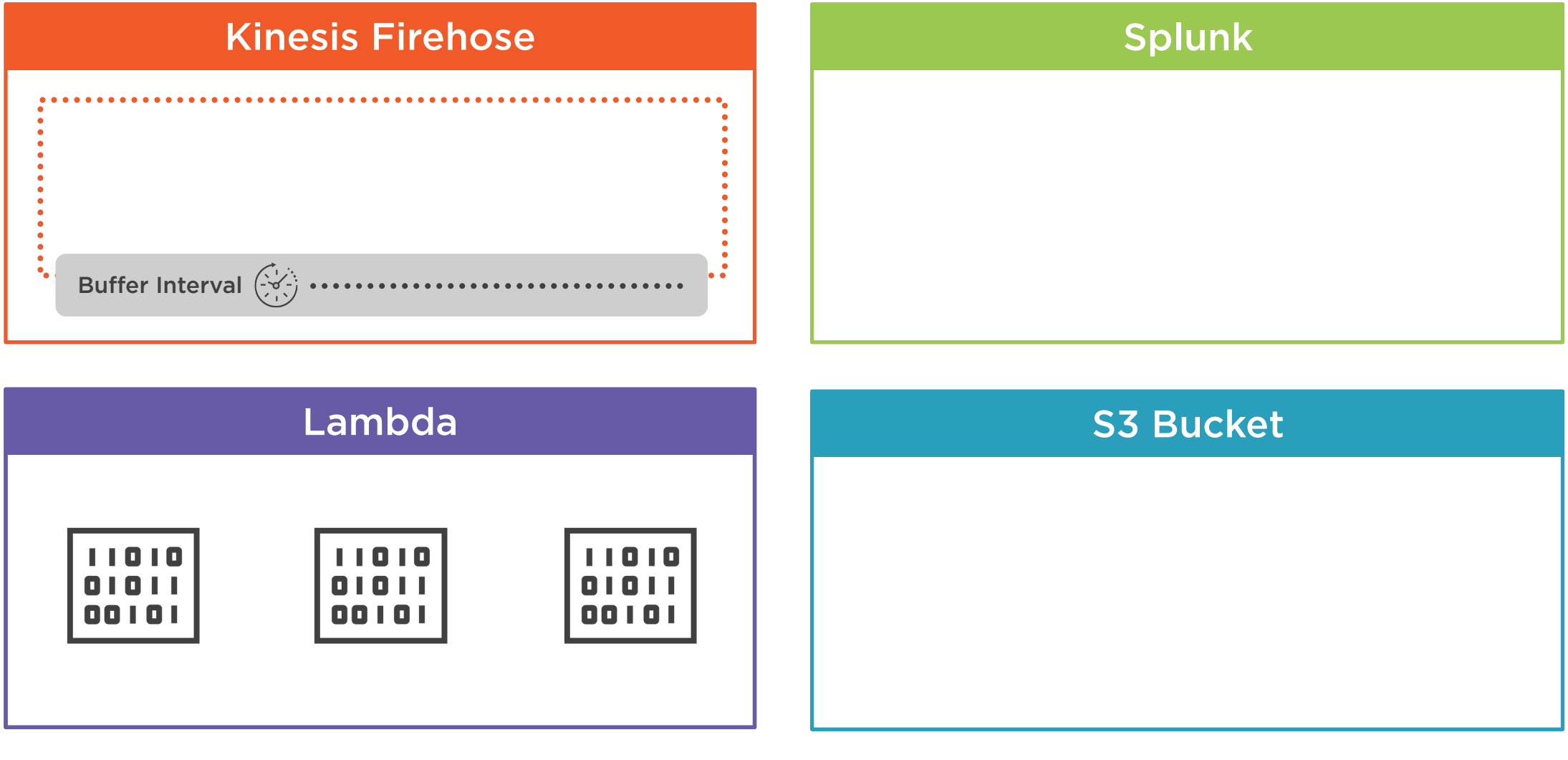
Kinesis Firehose

Splunk

Lambda

S3 Bucket

Kinesis Firehose to Splunk

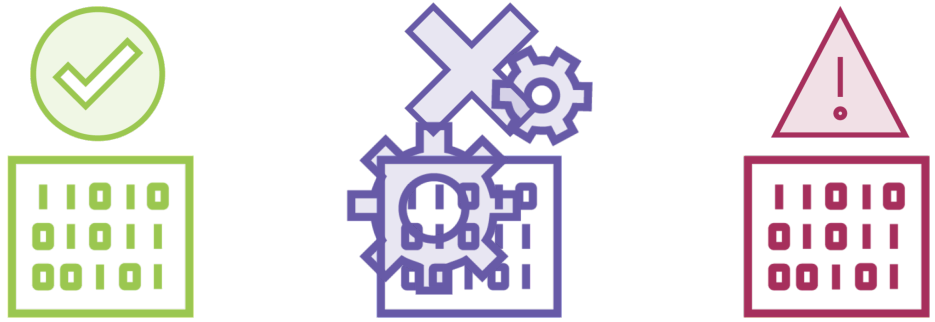


Kinesis Firehose to Splunk

Kinesis Firehose

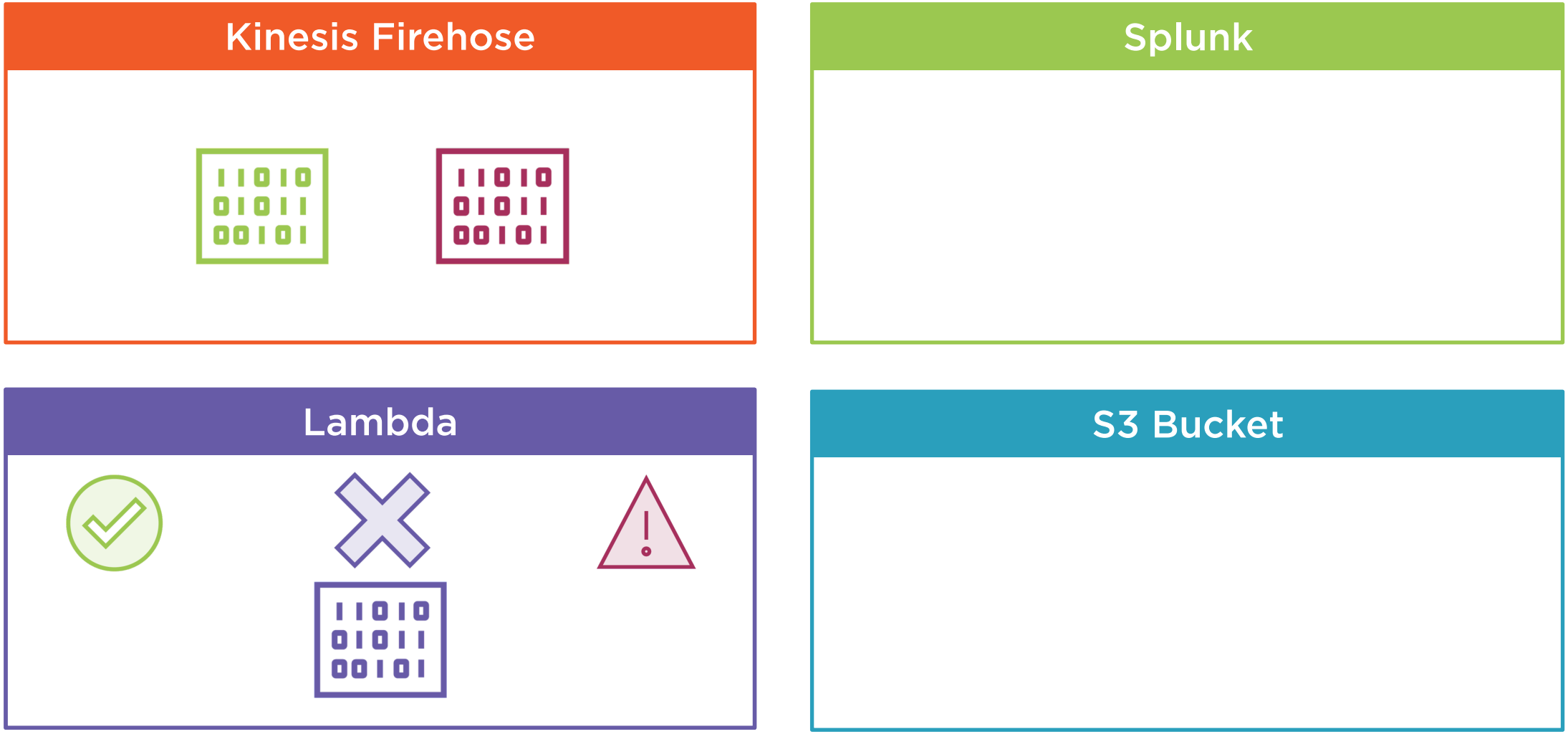
Splunk

Lambda



S3 Bucket

Kinesis Firehose to Splunk



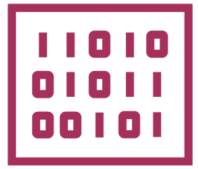
Kinesis Firehose to Splunk

Kinesis Firehose

Splunk

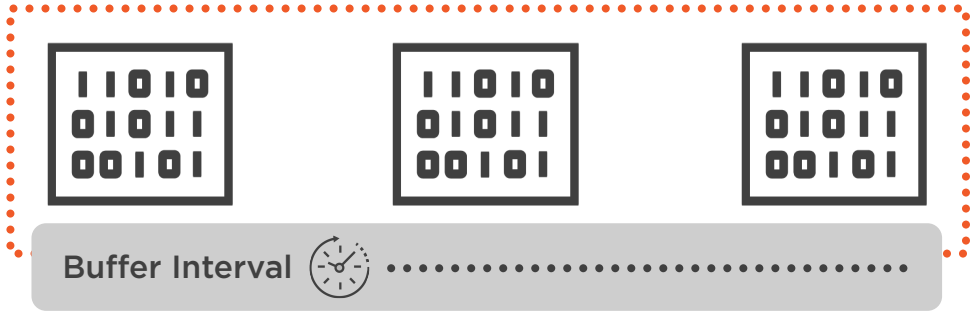
Lambda

S3 Bucket



Kinesis Firehose to Splunk

Kinesis Firehose

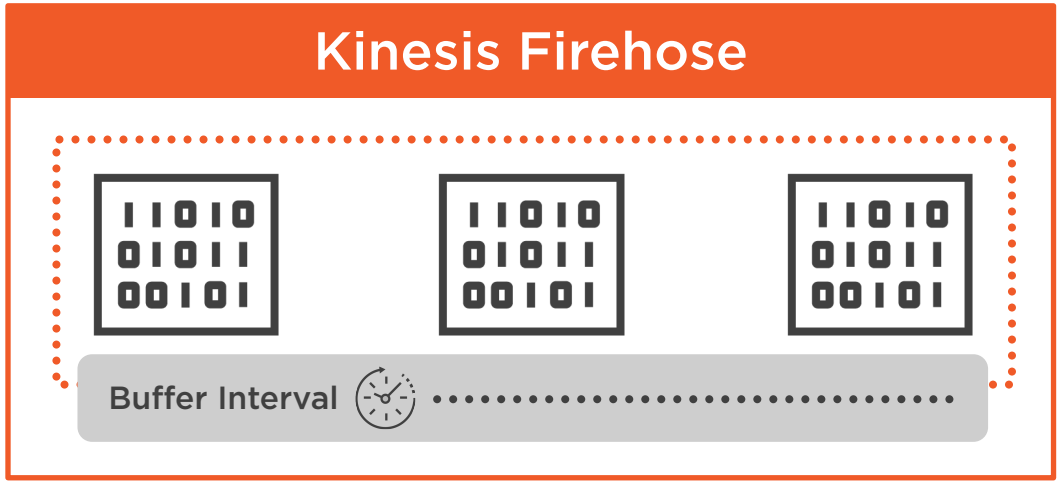


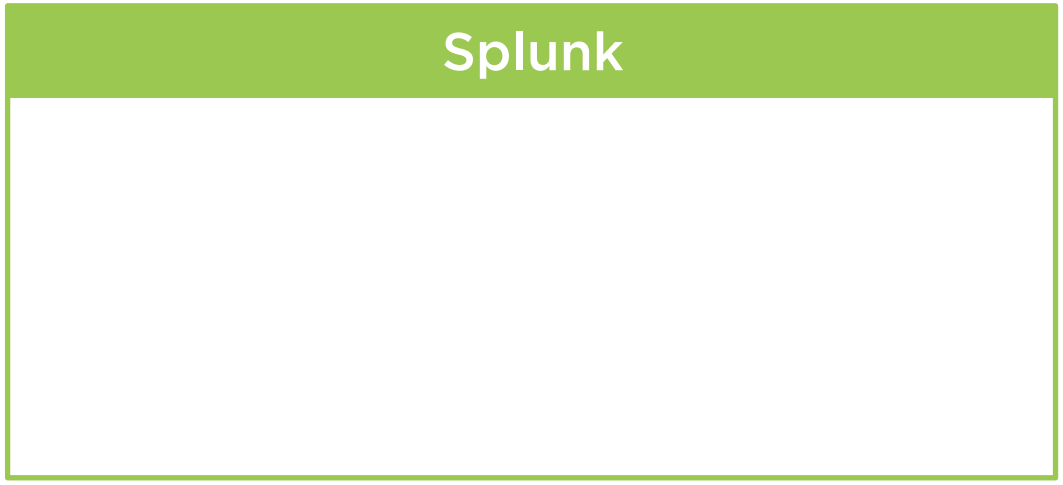
Splunk

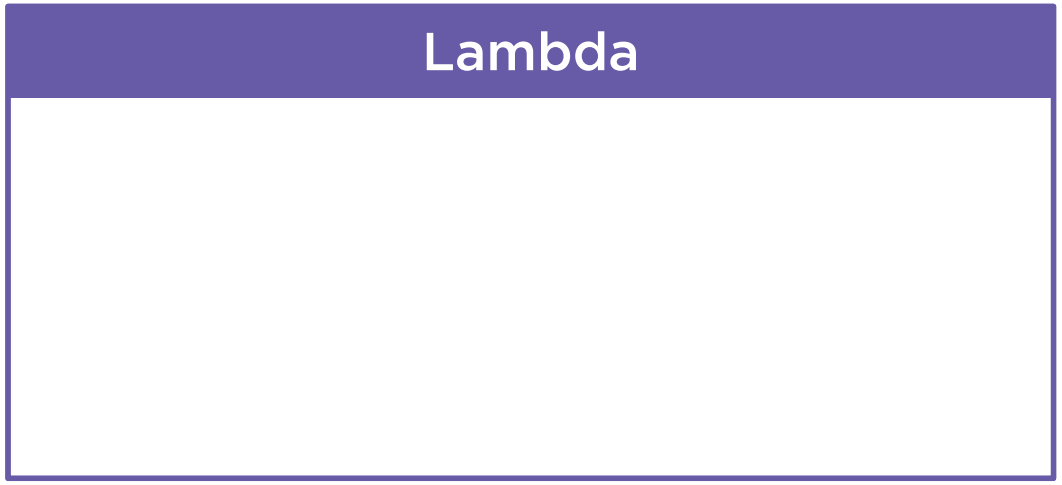
Lambda

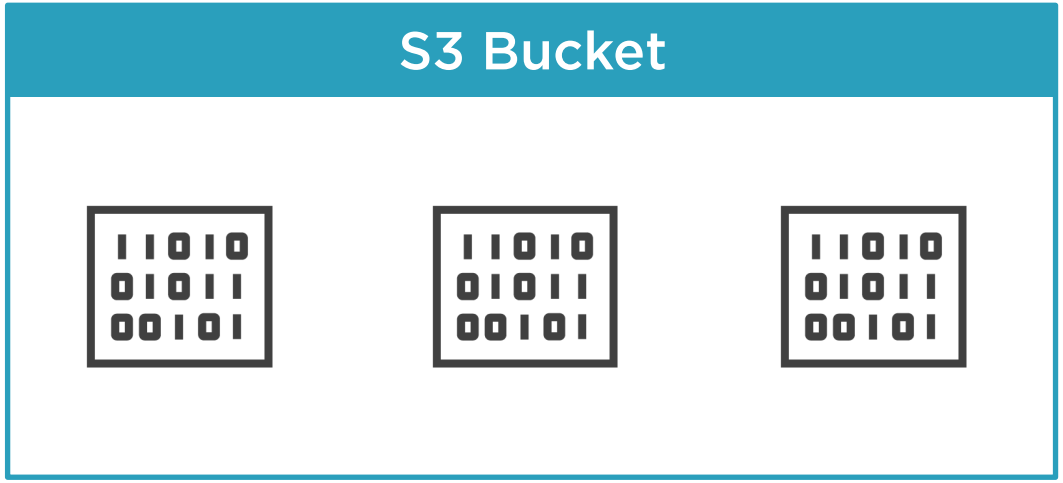
S3 Bucket

Kinesis Firehose to Splunk









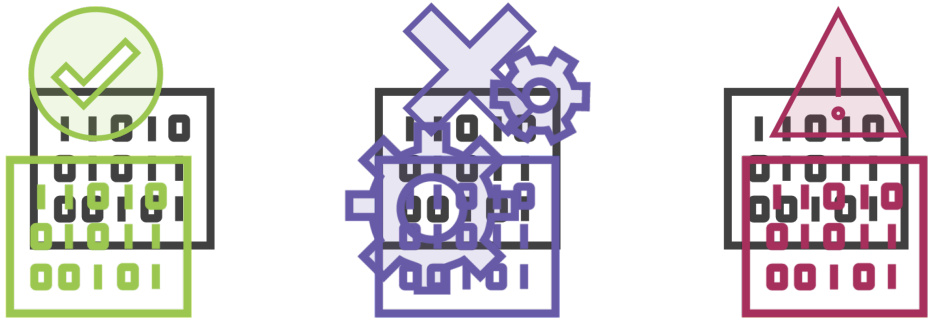
Kinesis Firehose to Splunk

Kinesis Firehose

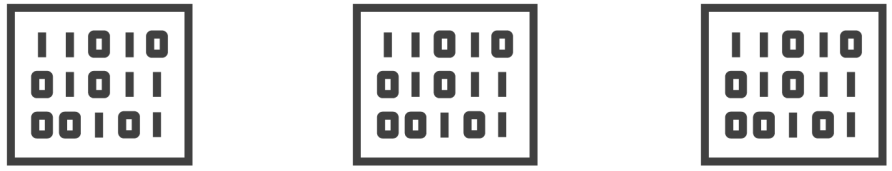
Buffer Interval

Splunk

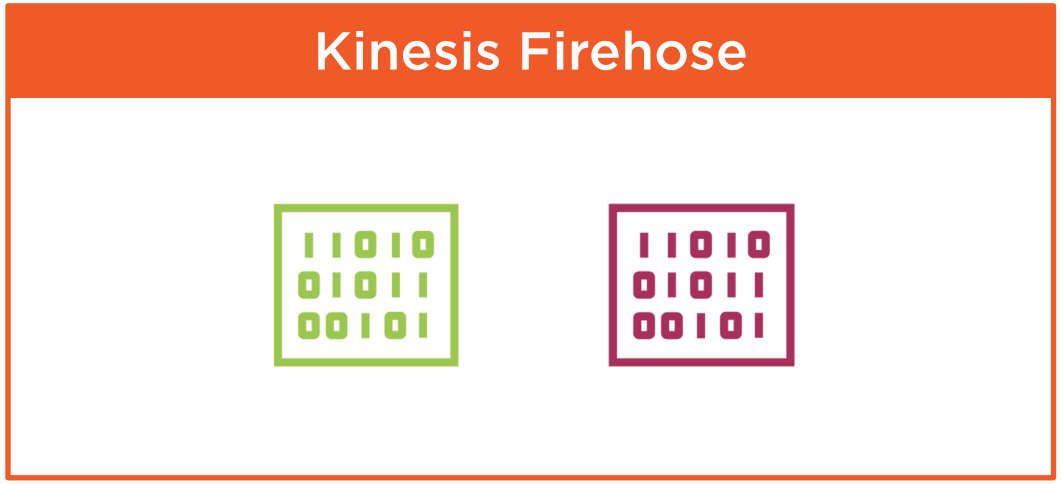
Lambda



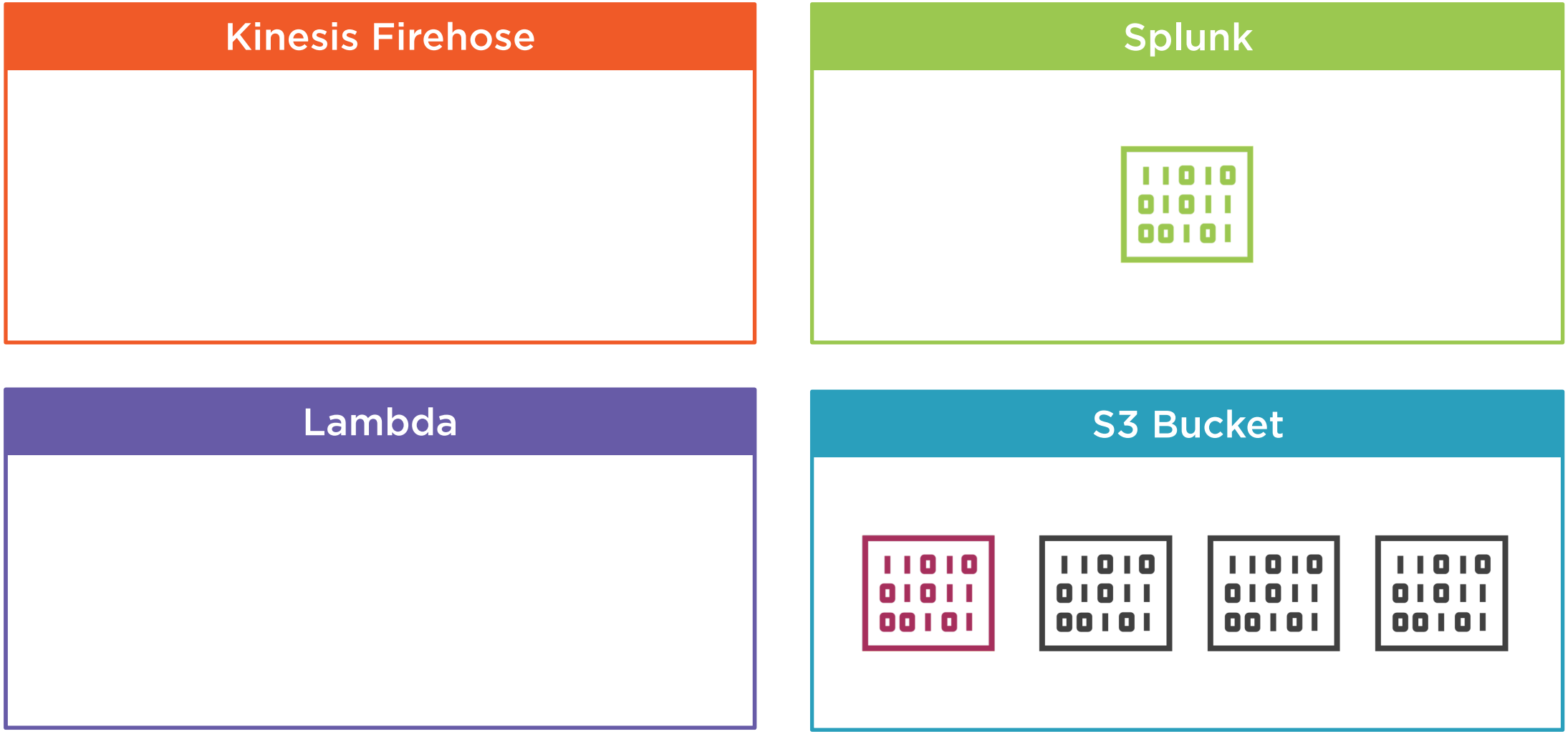
S3 Bucket



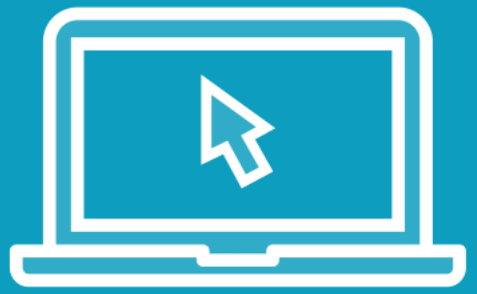
Kinesis Firehose to Splunk



Kinesis Firehose to Splunk



Demo



WARNING

- Not in the AWS Free Tier

Creating a Firehose Delivery Stream

- Create Kinesis Data Stream Source
- Script to send data into data stream
- Create Firehose Delivery Stream
- Create Lambda function
  - Transform data from JSON to CSV
- Output data in S3

Kinesis Firehose vs. Kinesis Data Streams

Comparing Kinesis Options

Kinesis Firehose	Kinesis Data Streams
1-minute delay for data is ok	You want data in near real time
More managed service	More control of capacity yourself
Only supports specific destinations	Supports multiple data consumers
Use Lambda for data transformation	Consumers processing data streams
Retries deliveries automatically	24 hours to 7 days of data retention

Summary

- Basic Concepts
- Firehose Producers and Destinations
- Transforming data
- Demo: Working with Firehose
- Firehose vs. Data Streams

Up next:

Data Transfer with Snowball and Direct Connect