Amazon Kinesis Data Streams

Fernando Medina Corey

CLOUD ARCHITECT

@fmc_sea   www.fernandomc.com

Overview

Core Kinesis Data Streams concepts

– Producers, shards, consumers, and records

Producers

– Kinesis Agent, Producer Library, and AWS SDKs

Consumers

– Kinesis Client Libraries, AWS Services, and AWS SDKs

Demo – Analyzing Kinesis Data Streams

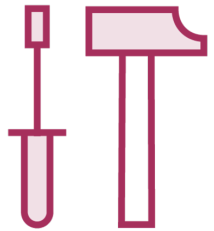# Kinesis Data Streams Basics

Producers
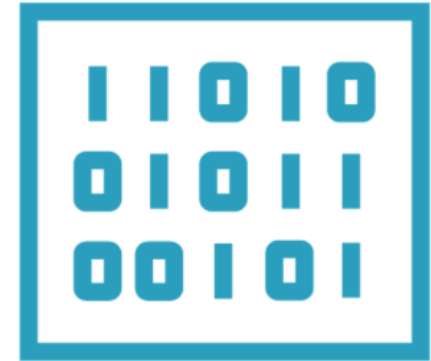
Kinesis Agent

Producer Library

AWS SDKs

Kinesis Data Streams Concepts

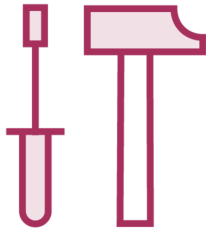**Data Stream**

**Kinesis Data
Stream Shards**
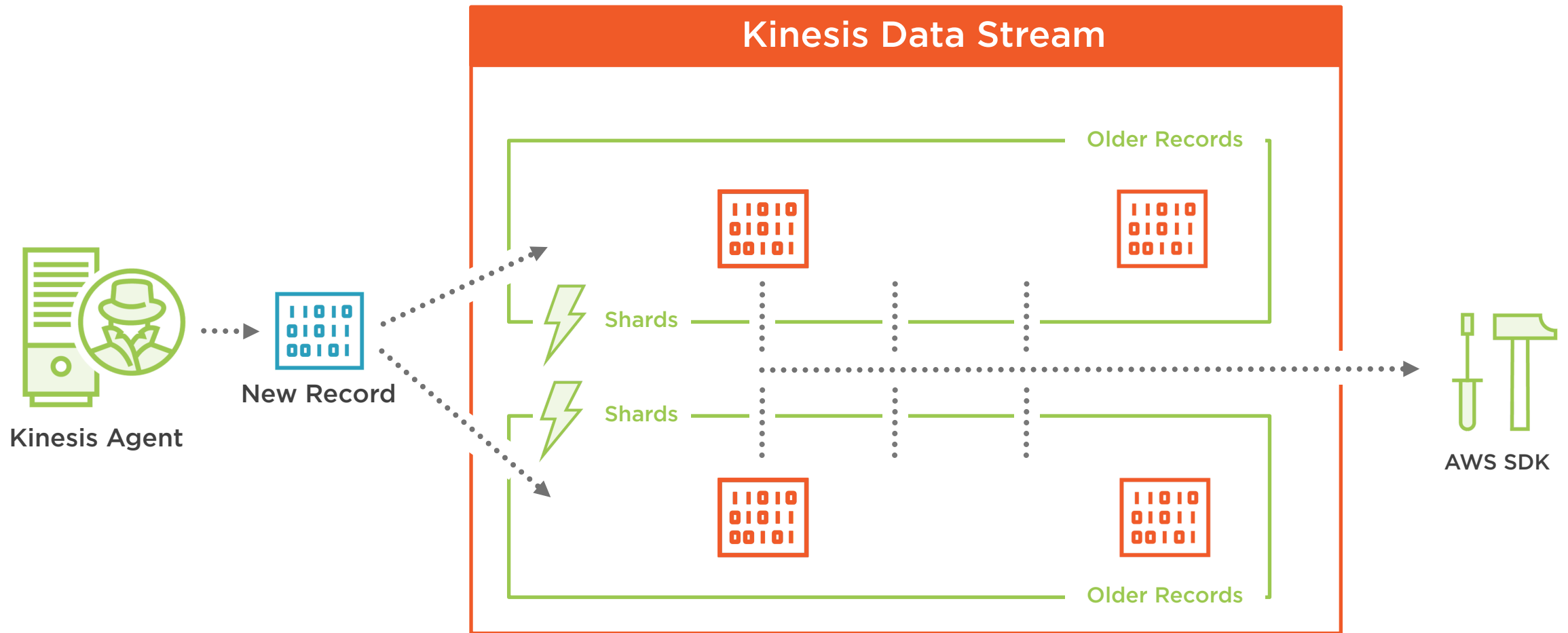
Records

Consumers

Client Libraries

AWS SDKs

AWS Services
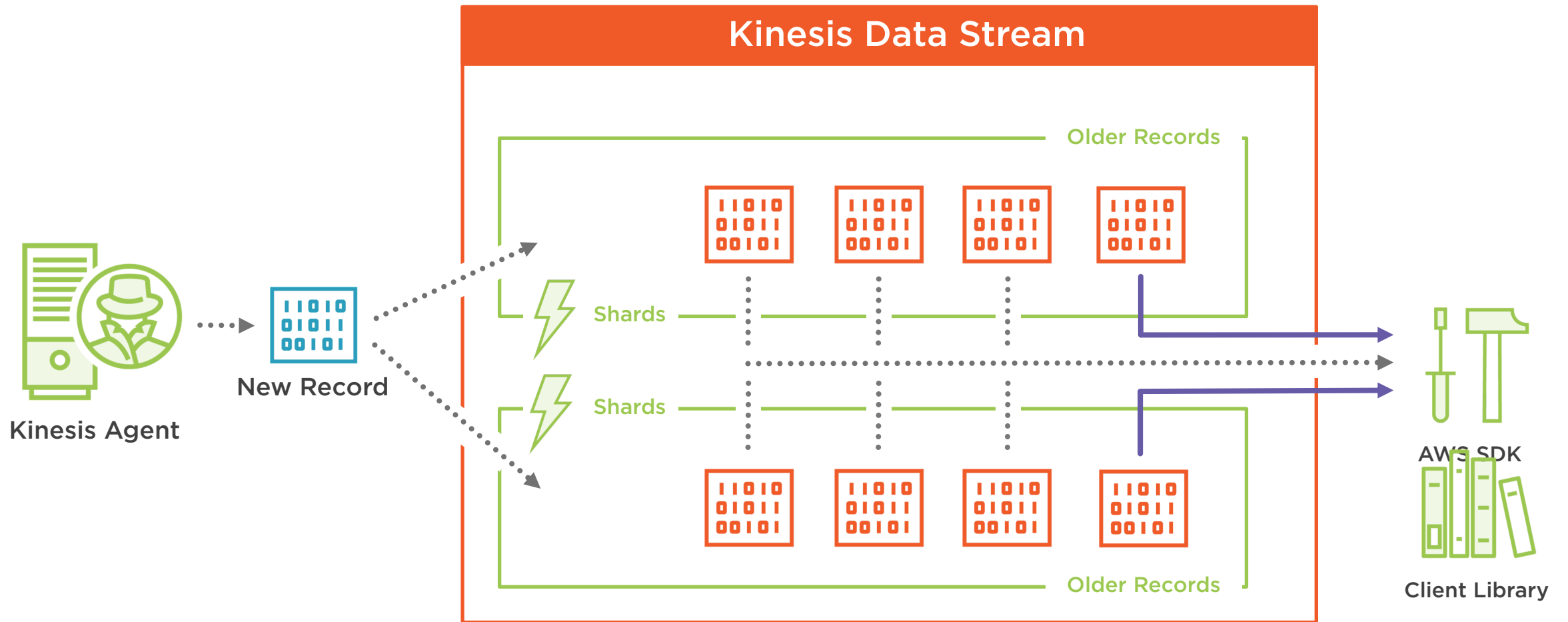
Kinesis Data Streams Shards and Records

What's in a Kinesis Data Stream

Kinesis Agent

New Record

Kinesis Data Stream
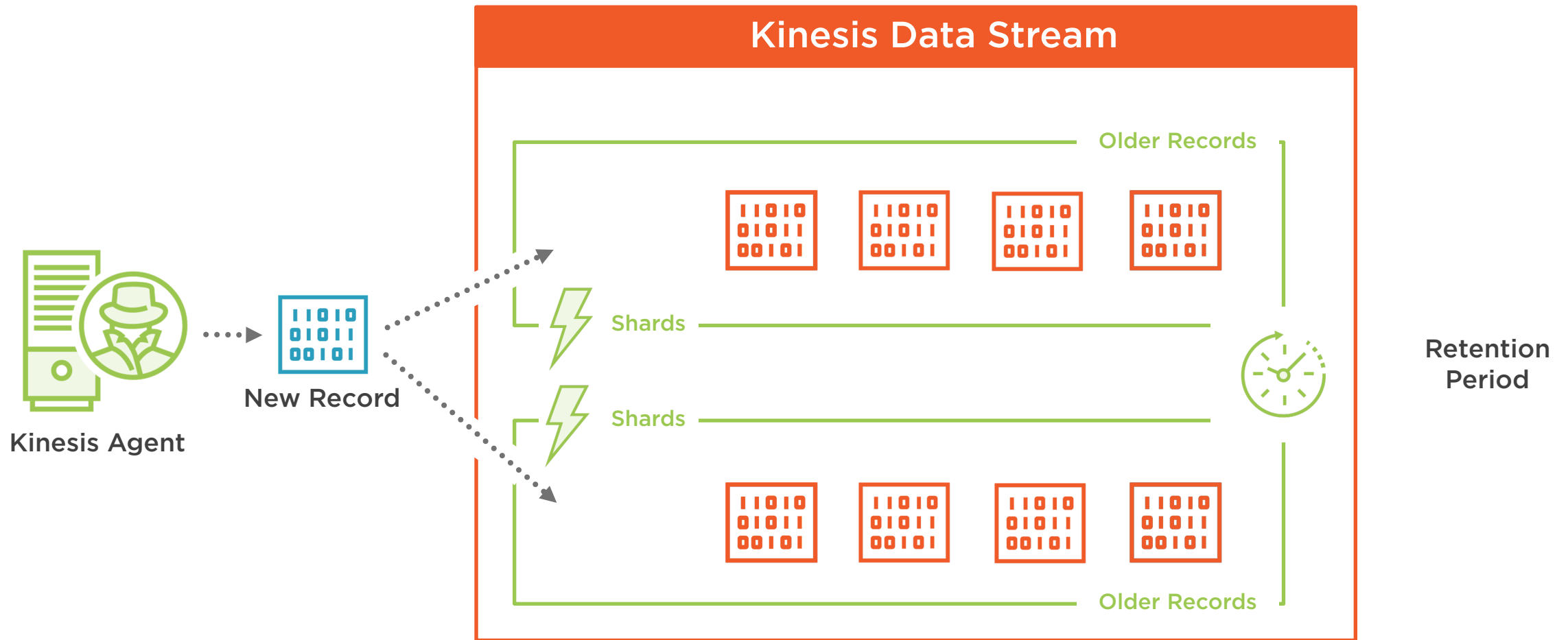
Older Records

Shards

Shards

Older Records

AWS SDK

What's in a Kinesis Data Stream

What's in a Kinesis Data Stream

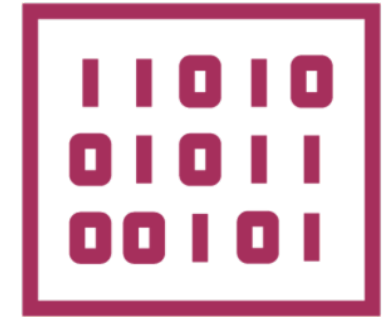# What's in a Data Record?

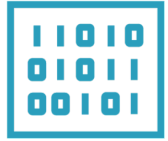**Sequence Number**   **Partition Key**   **Data Blob**

Stream and Record Limitations

Record data blob size limit - 1 MB

Number of streams per account – No limit

Soft limit of shards per account - 200 or 500 depending on region

Data retention length - 24 hours to 7 days

Shard Limitations

Ingest 1 MB per second

Ingest 1000 records per second

Read 2 MB per second

5 read transactions per second

10 MB or 10k records per read transaction
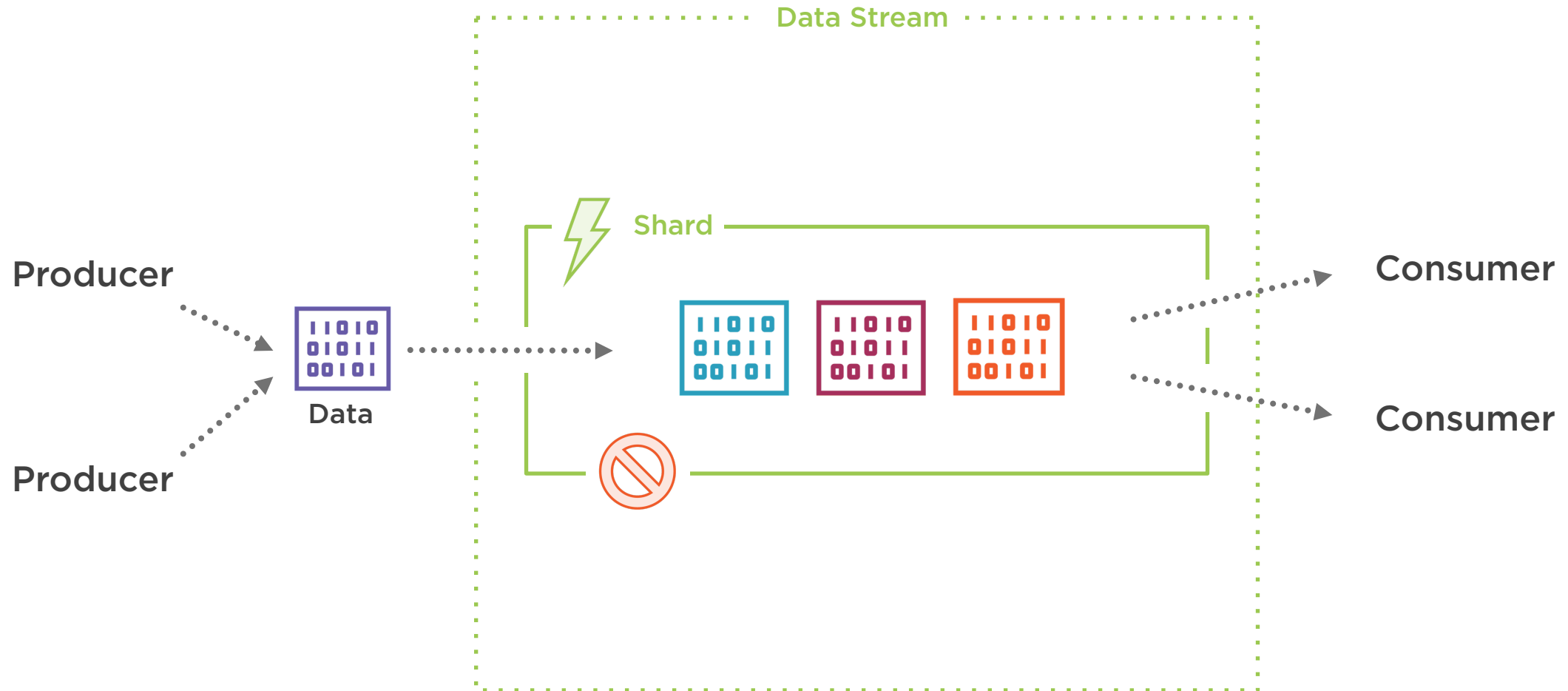
Resharding : Merging or Splitting a Shard
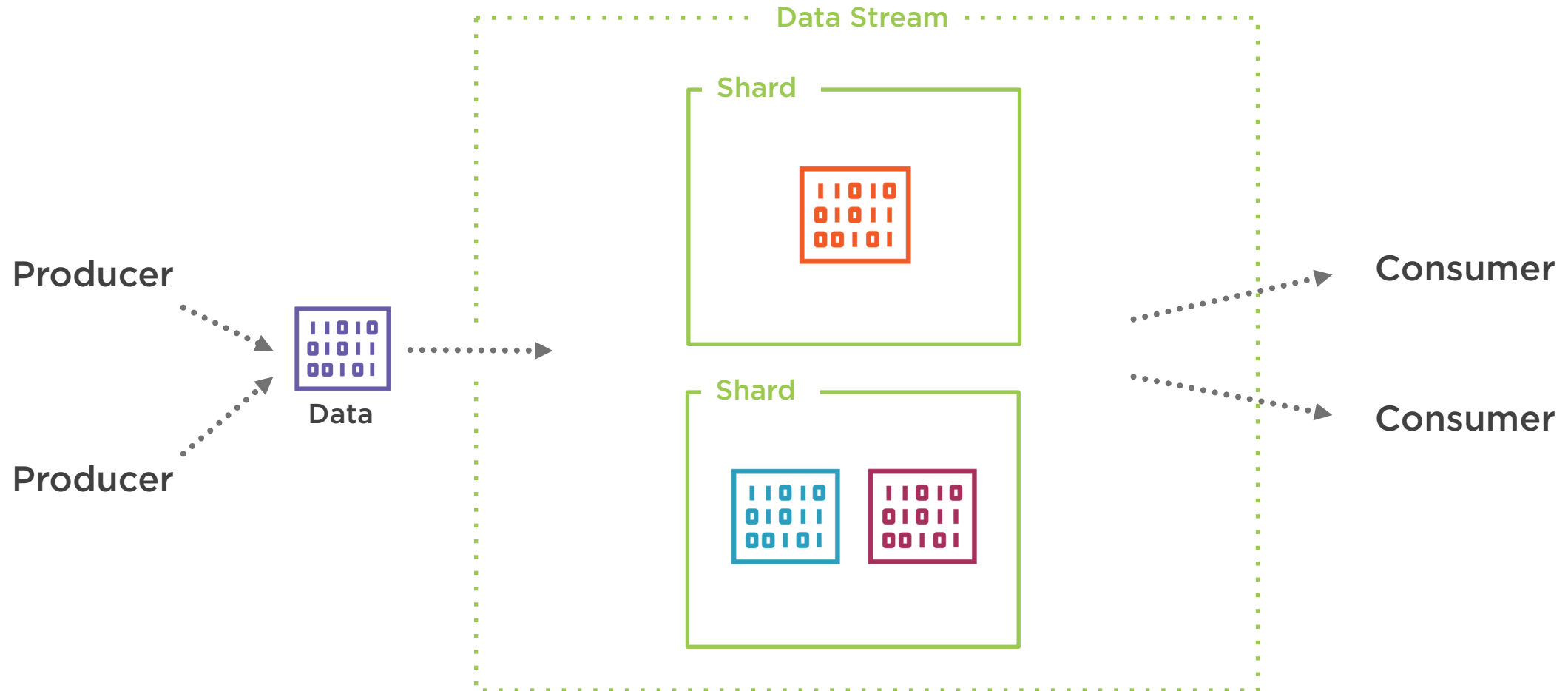
Shard Split

Shard Merge

Visualizing Kinesis Data Streams

Producer

Producer

Data

Data Stream

Shard

Consumer

Consumer

Visualizing Kinesis Data Streams

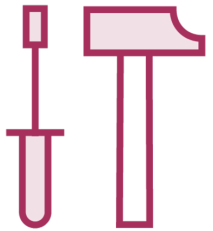# Kinesis Data Streams Producers

Producers – Kinesis Agent

Kinesis Agent
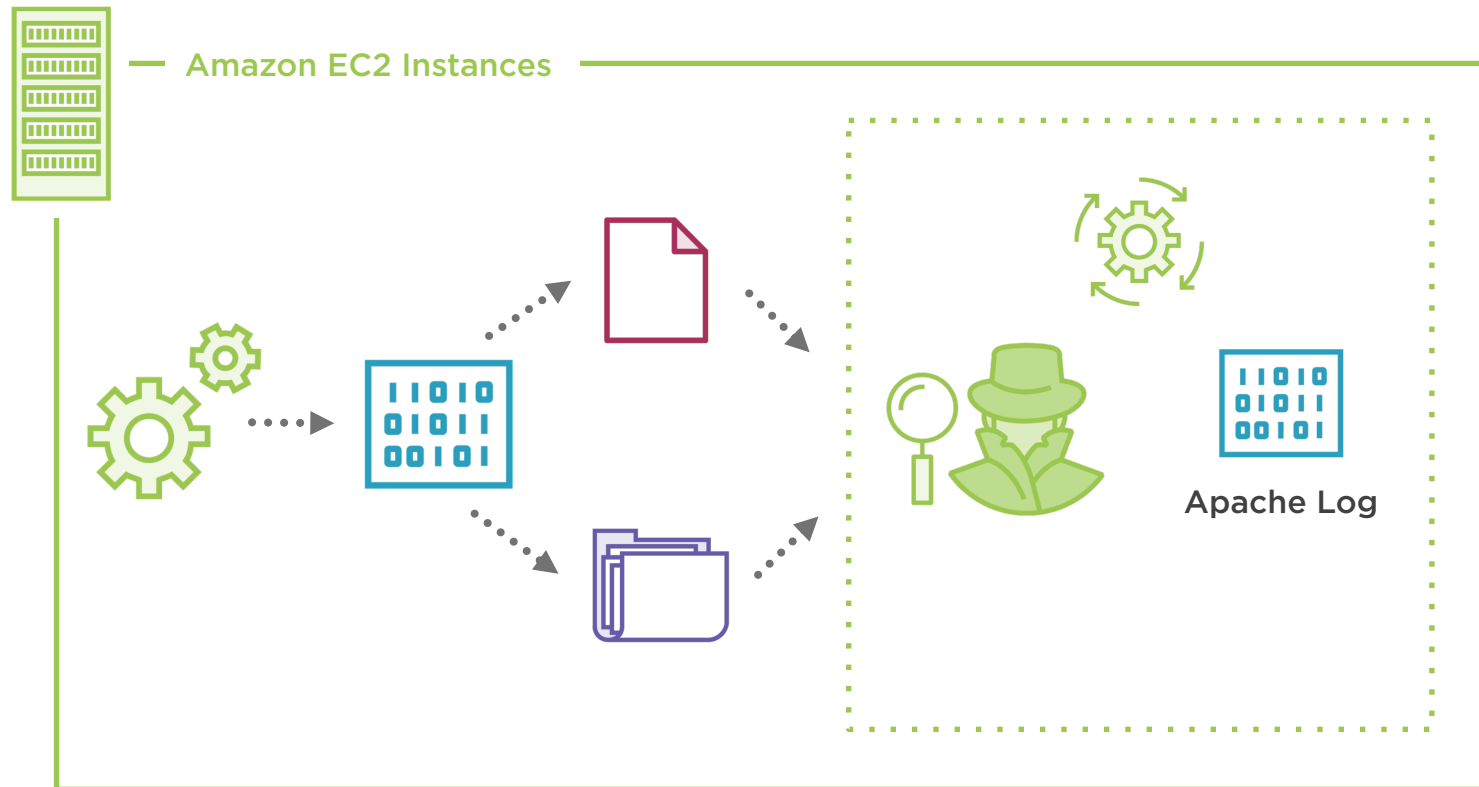
Producer Library

AWS SDKs

What Is the Kinesis Agent?

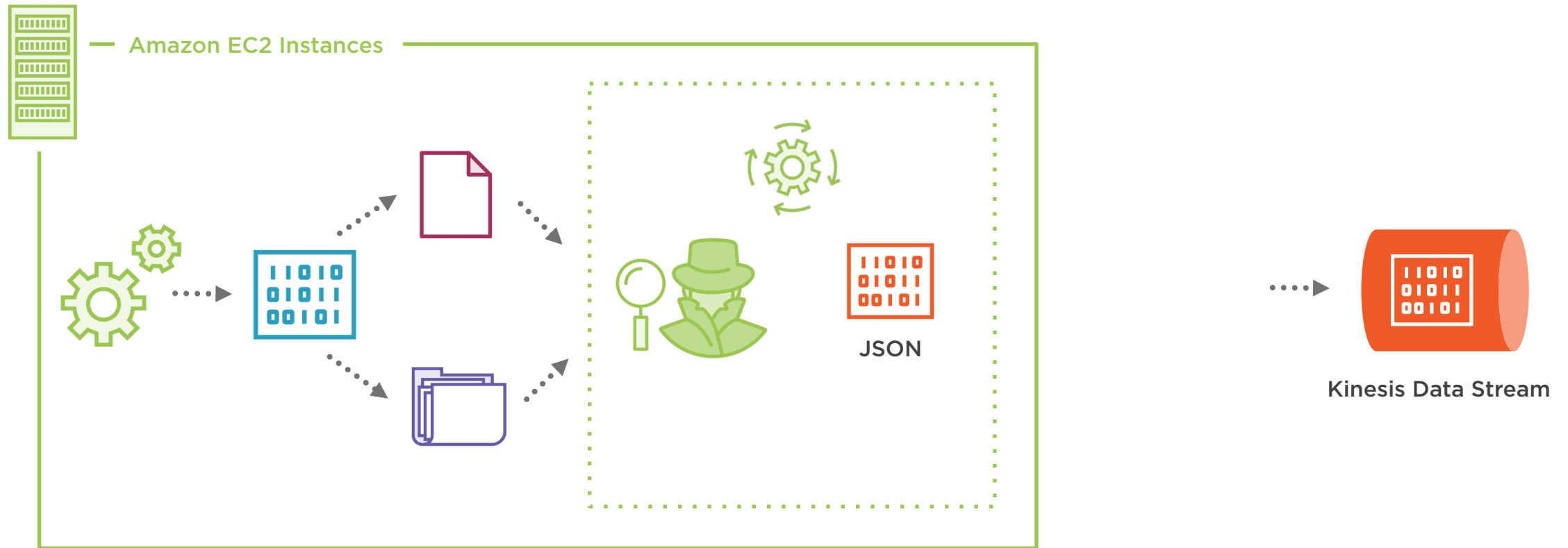| Java application | Streams logs from one or many machines | Lots of useful features and integrations |

Visualizing the Kinesis Agent in Action

Amazon EC2 Instances

Apache Log

Visualizing the Kinesis Agent in Action

Amazon EC2 Instances

JSON

Kinesis Data Stream

Kinesis Agent Log Formats

Multi-line to Single Line	**CSV/PSV/DSV to JSON**	**Apache Log or Syslog formats to JSON**

Why Use the Kinesis Agent?

Performance benefits
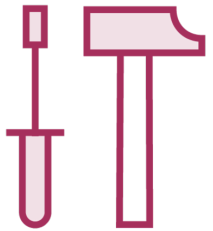
Lots of built-in functionality

CloudWatch integration

Producers – Producer Library

Kinesis Agent

Producer Library

AWS SDKs

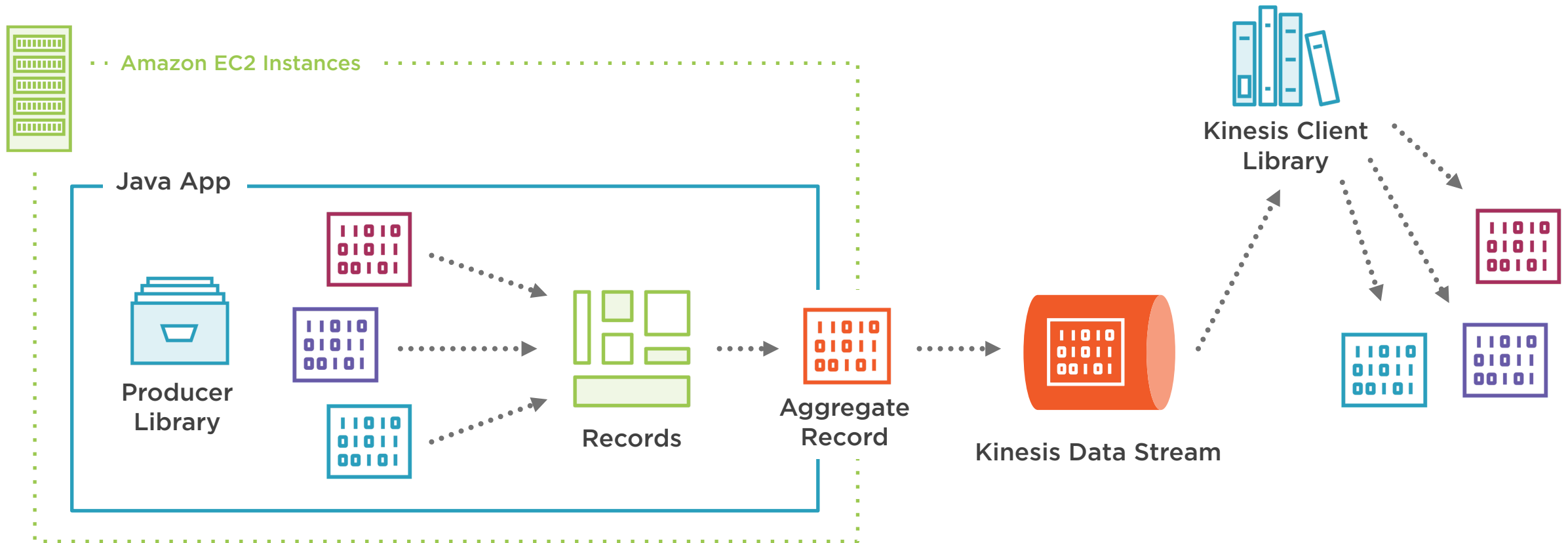# What Is the Producer Library?

Library for Java apps

Helps collect and aggregate logs to Kinesis

Useful when working with high-rate producers
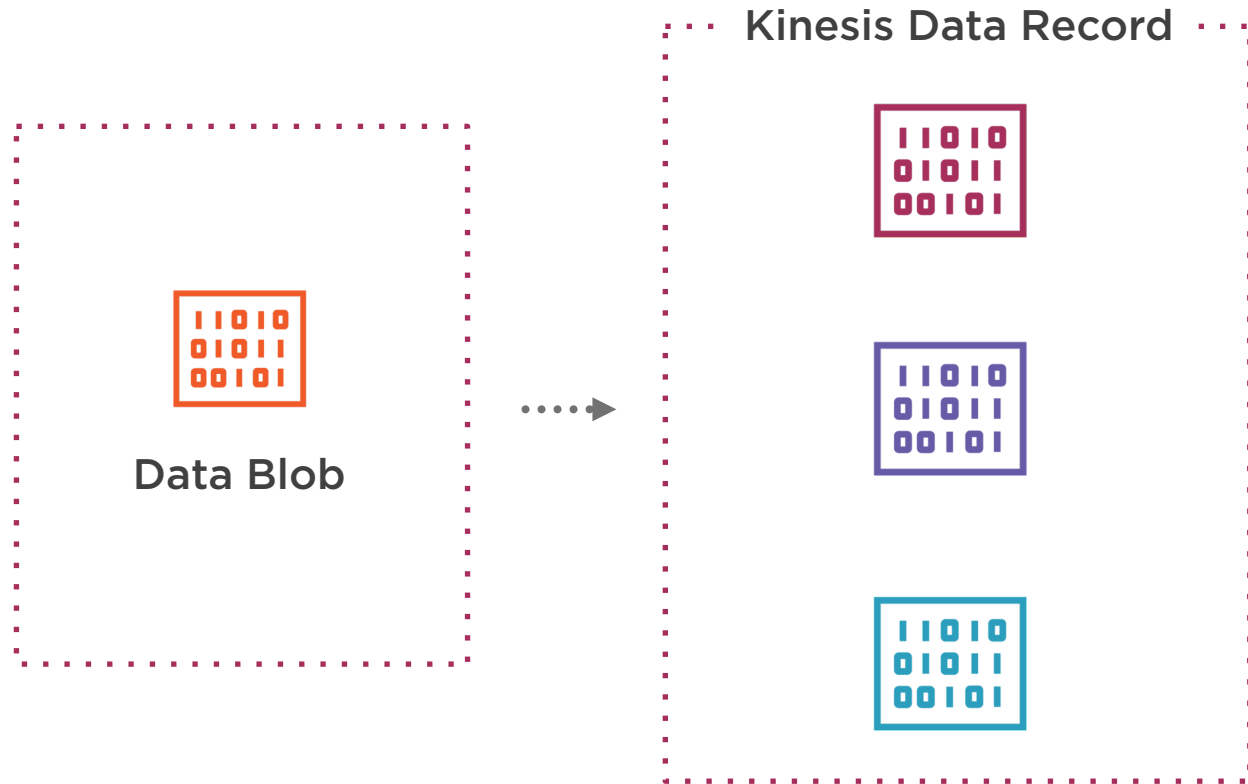
Visualizing the Producer Library

Amazon EC2 Instances

Java App

Producer Library

Records

Aggregate Record

Kinesis Data Stream

Kinesis Client Library

Kinesis Producer Library – Batching

Aggregation

Collection

Batching - Aggregation

Data Blob

Kinesis Data Record
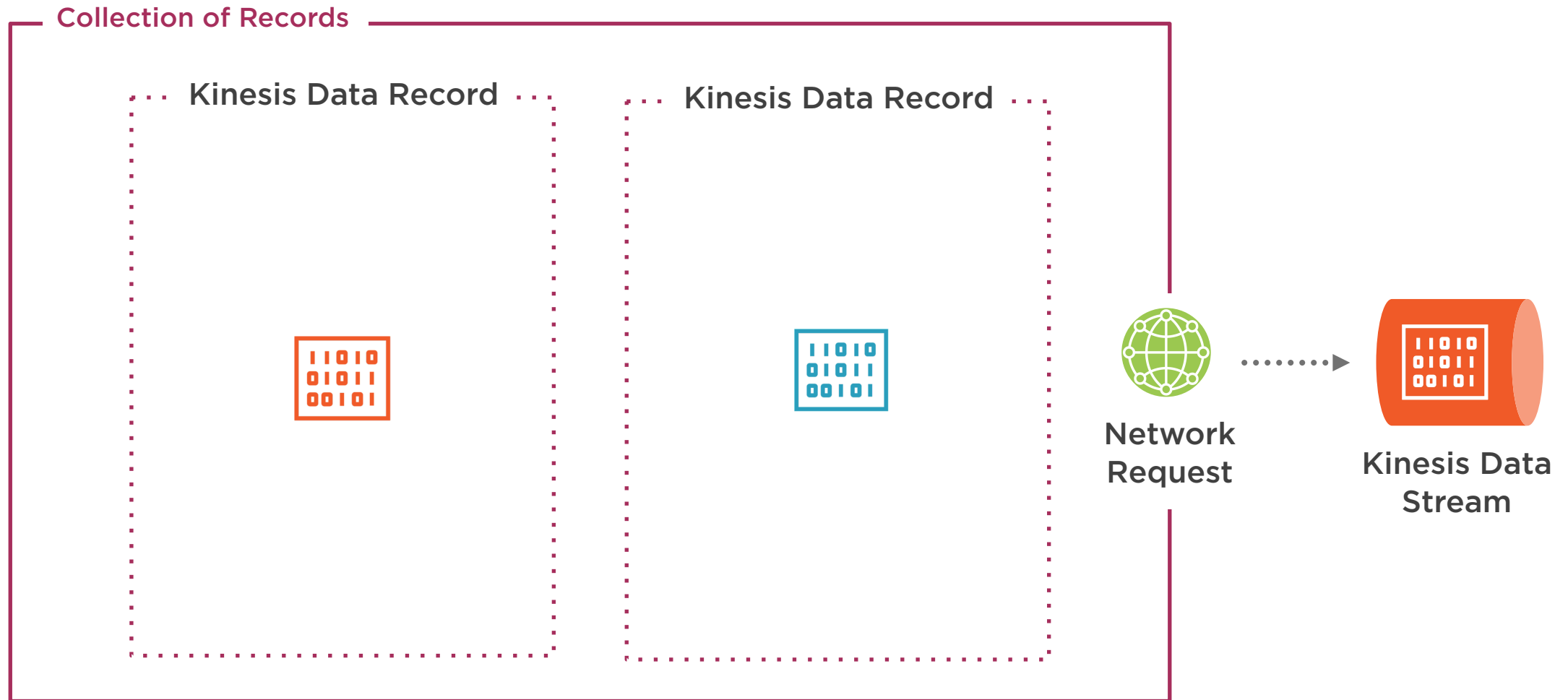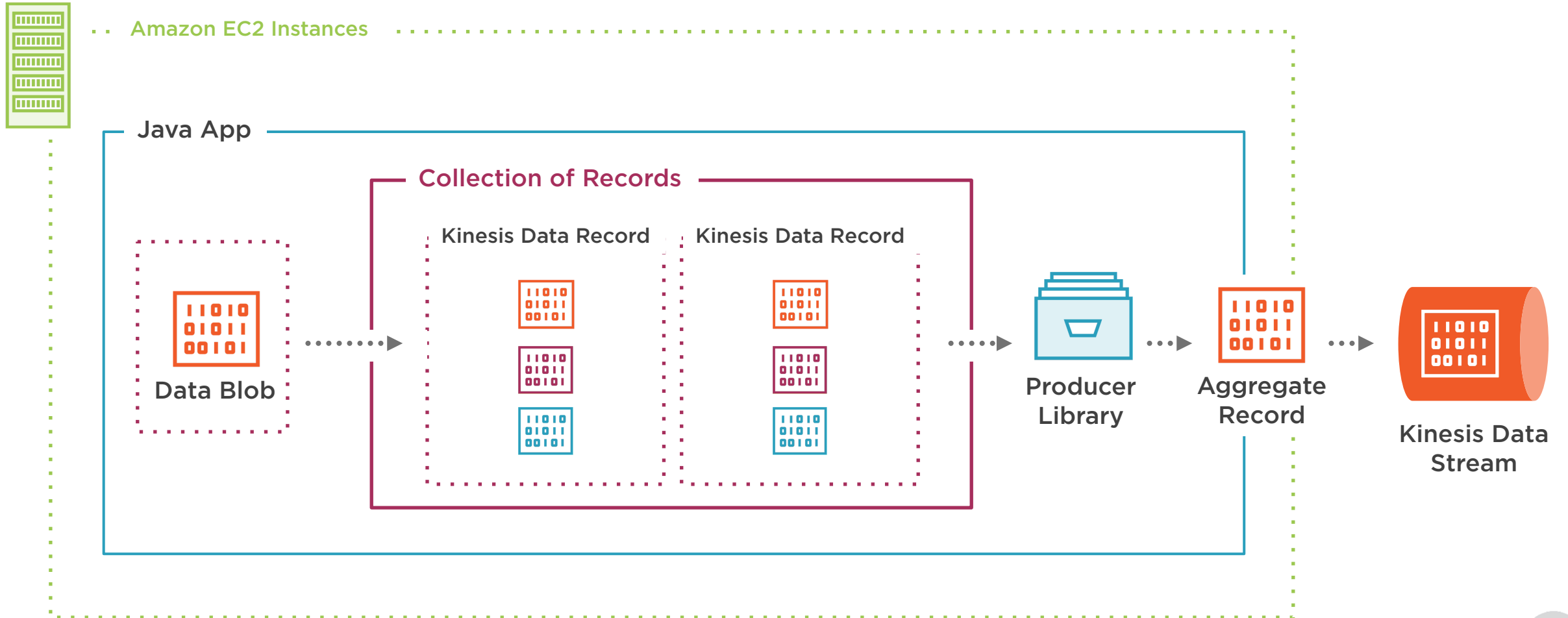
Batching - Collection

Batching - Collection and Aggregation

Why Use the Producer Library?

Helps handle errors with retry logic

Batching and performance with high-rate producers

CloudWatch integration

Integrates with the Kinesis Client Libraries

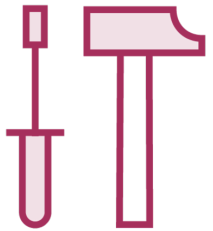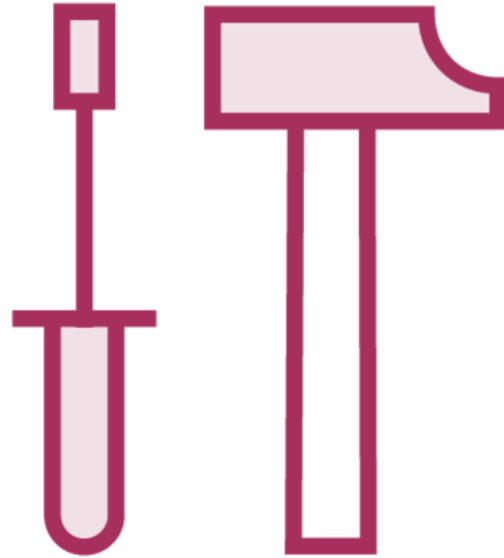# Producers – AWS SDKs

Kinesis Agent

Producer Library

AWS SDKs

Why Use the AWS SDKs?

Supports multiple programming languages

Embed inside mobile apps

Low rate producers

Options for the AWS SDKs

PutRecord

PutRecords

Visualizing the AWS SDKs

Amazon EC2 Instances

Python App

Boto 3

Mobile App

iOS SDK

IoT Device

IoT Device SDK

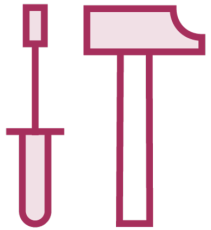Kinesis Data Stream
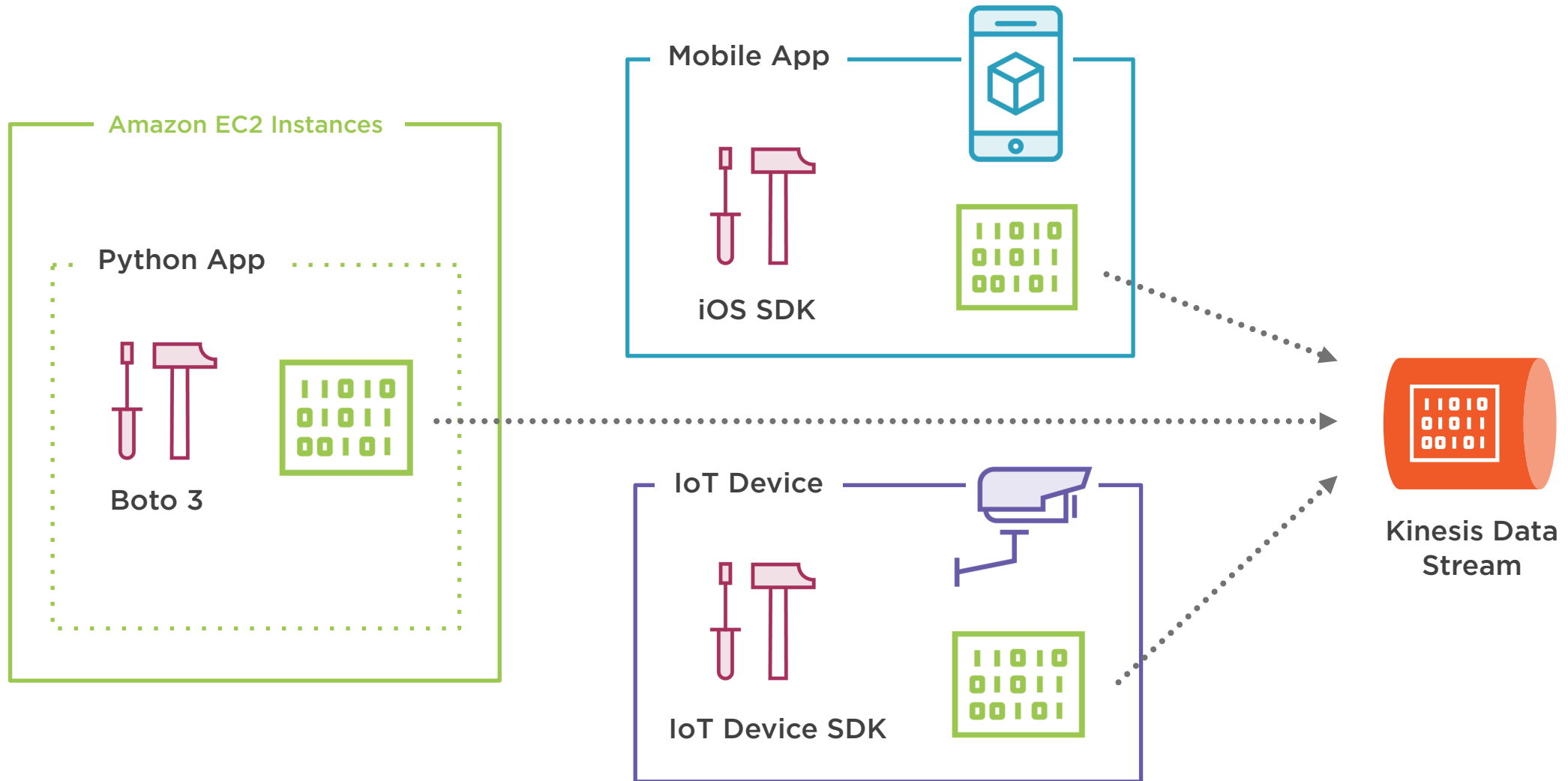
Kinesis Data Streams Consumers

Consumers – Client Libraries

Client Libraries

AWS SDKs

AWS Services

What Are the Kinesis Client Libraries?

Tool for processing data from Kinesis Streams

Variety of programming languages supported

KCLs help with common use cases

Kinesis Client Libraries – Supported Languages

Ruby

.Net

Node.js

Java

Python

Visualizing the Kinesis Client Library

Amazon EC2 Instances

Java App

Collection of Records

Kinesis Data Record

Kinesis Data Record

Data Blob

Producer Library

Aggregate Record

Kinesis Data Stream
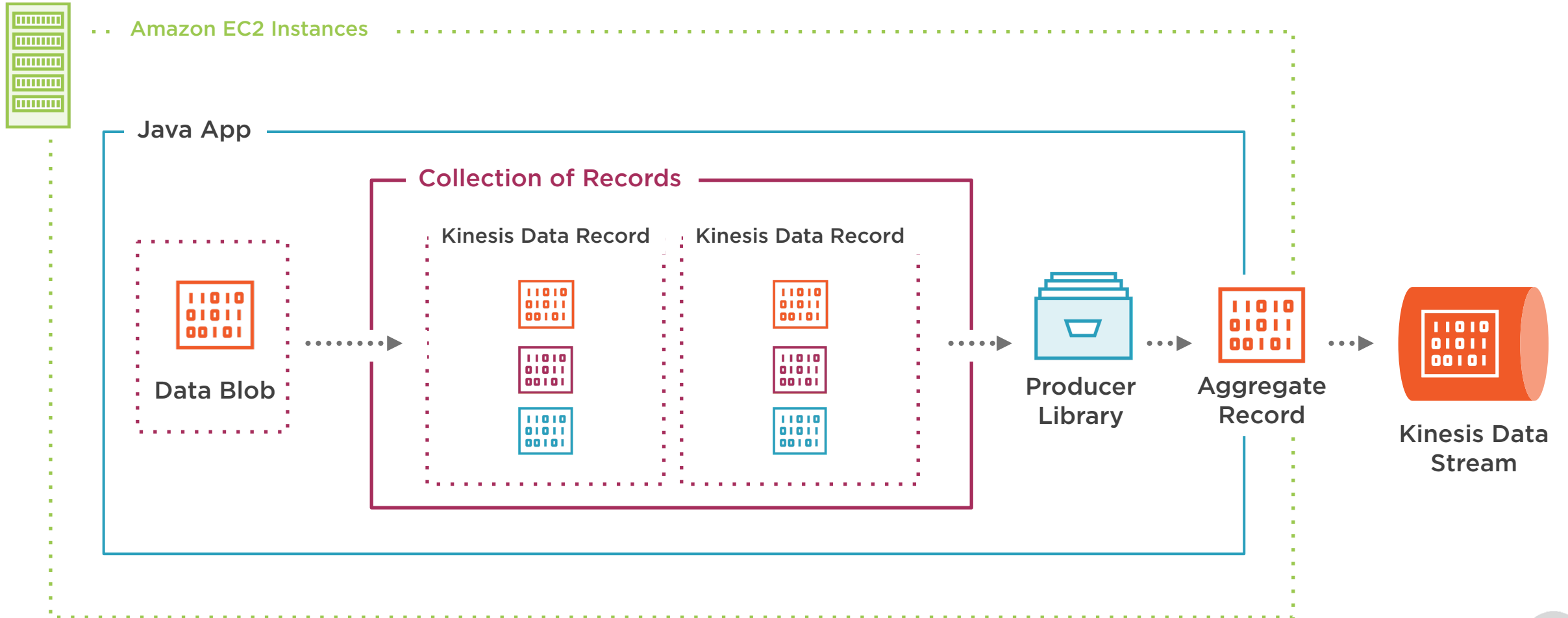
Visualizing the Kinesis Client Library
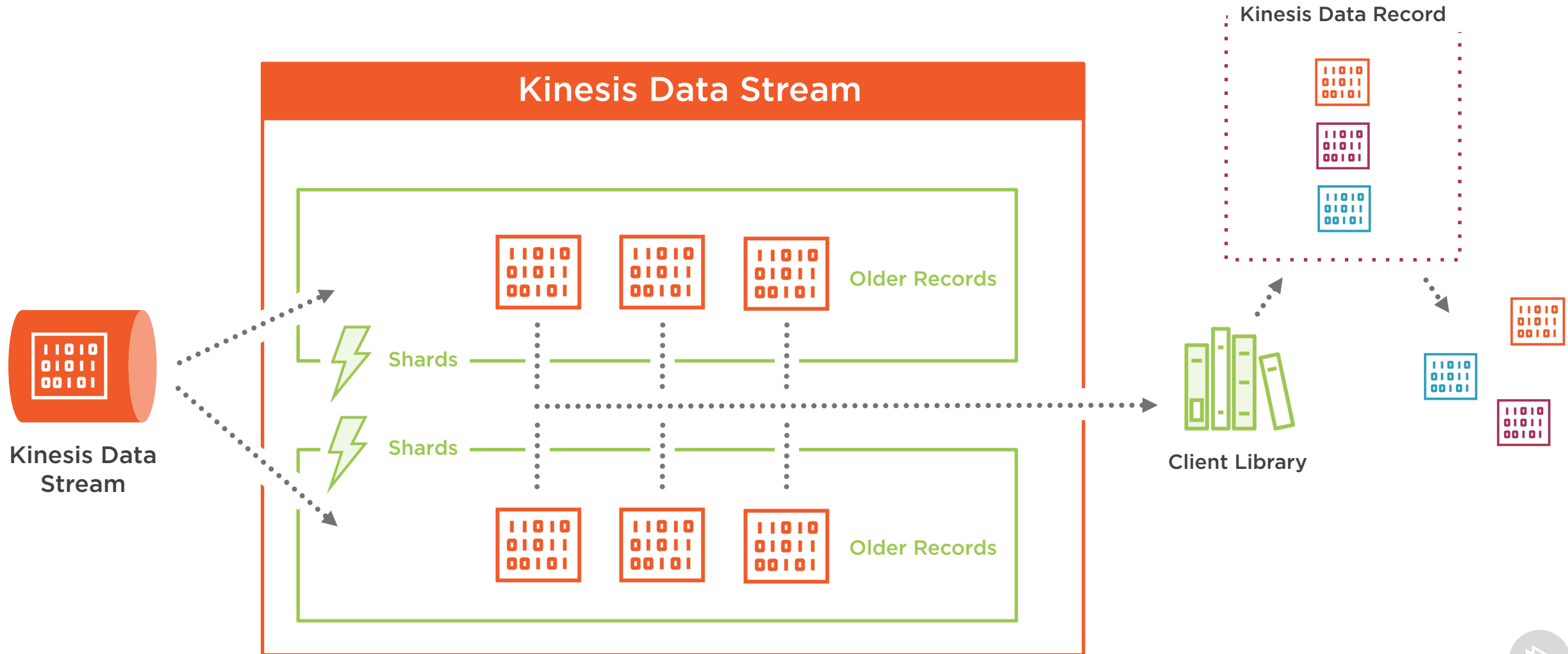
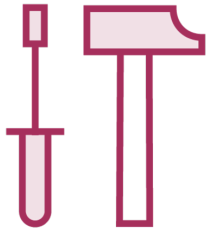# Throughput Options

Shared throughput

Dedicated throughput
(Enhanced Fan-out)

Consumers – AWS SDKs
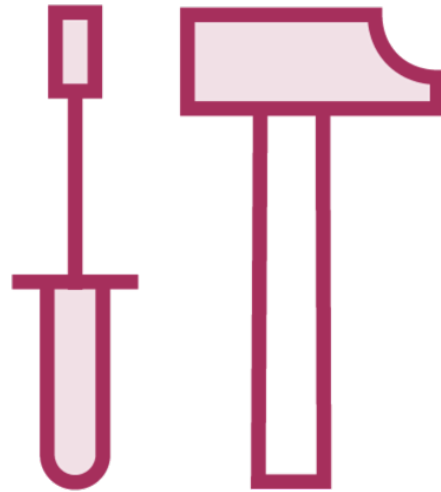
Client Libraries

AWS SDKs

AWS Services

Why Use the AWS SDKs?

Control over specific API actions

Can integrate with any SDK programming language
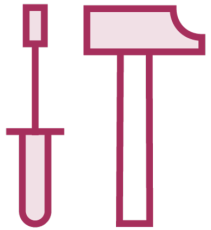
Run apps in EC2, AWS Lambda, containers or elsewhere

Consumers – AWS Services

Client Libraries

AWS SDKs

AWS Services

Consumers – AWS Services

AWS Lambda

Kinesis Data Analytics

Kinesis Delivery Firehose
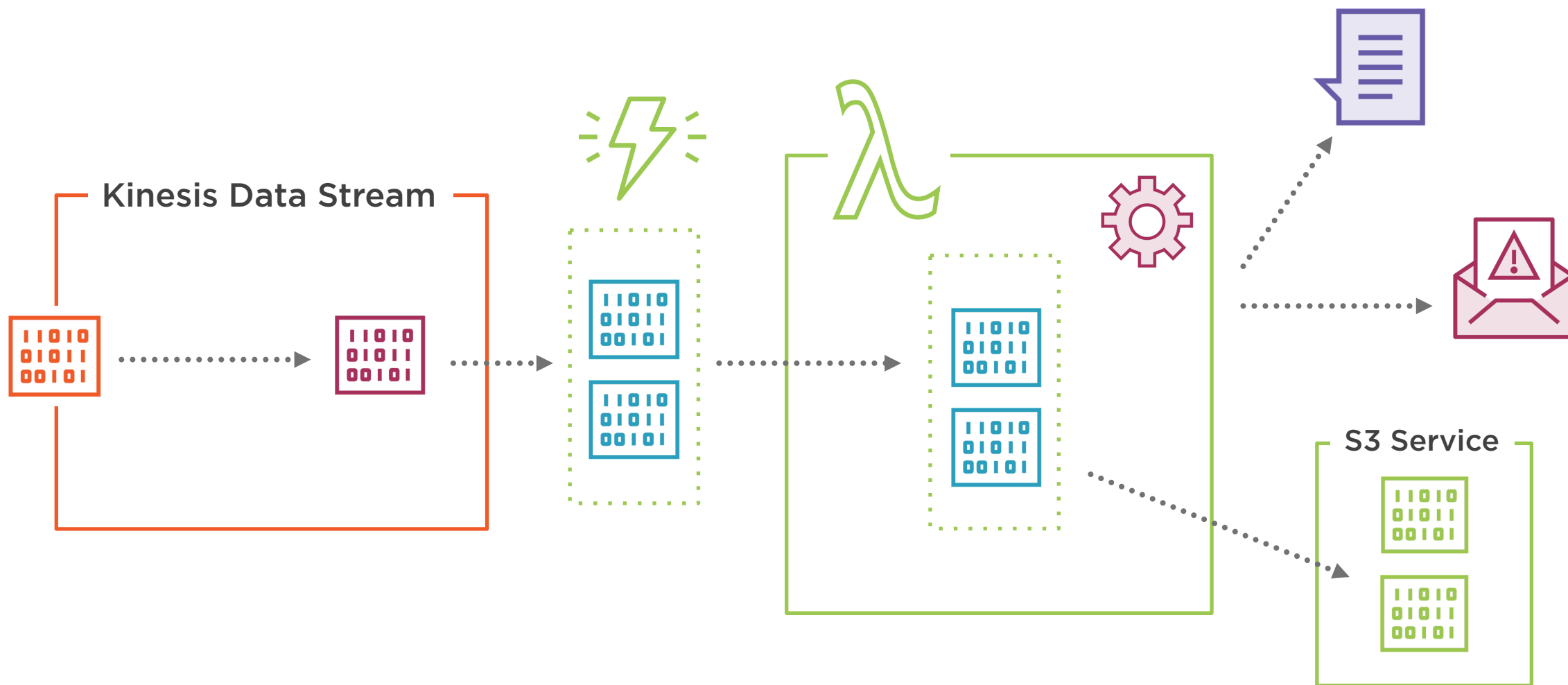
AWS Lambda Consumer

"Serverless" compute

Runs in response to new records

Records are batched and send to Lambda

AWS Lambda Consumer

Kinesis Data Stream

S3 Service

Kinesis Data Analytics

Stream data to Kinesis Data Analytics

Run SQL queries

Send output to other streams

Kinesis Data Firehose

Data flows from Kinesis Data Stream to Firehose

Firehose can process and transform data

Delivery to other destinations (S3, ElasticSearch, others)

Demo

WARNING
- Not in AWS free tier

Storyline
- Sales data from web API backend
- Searching for popular products
- Live analytics of streaming data

Create Kinesis Data Stream
- AWS Console
- Send data in with Boto3

Create Kinesis Data Analytics Application
- AWS Console
- Writing SQL queries

Summary

Kinesis Data Streams concepts

- Producers
- Shards and records
- Consumers
- Shared and dedicated throughput

Hands-on with:

- Kinesis Data Streams
- Kinesis Data Analytics

Up next:

Amazon Kinesis Data Firehose