

Developing Stream Processing Applications with AWS Kinesis

KINESIS FUNDAMENTALS

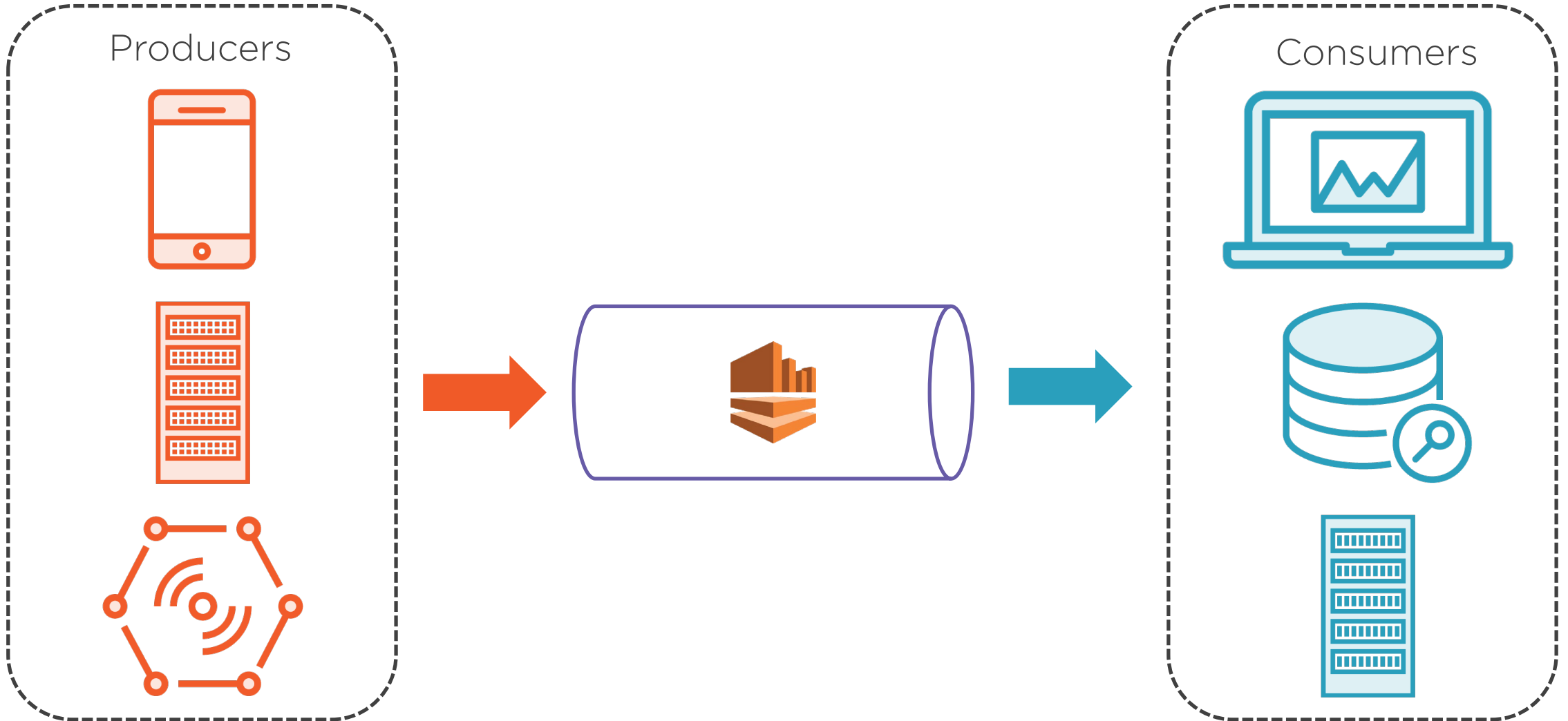


Ivan Mushketyk

@mushketyk brewing.codes



What Is AWS Kinesis



Kinesis Use-cases



Log data
collection and
processing



Messaging



Real-time
metrics



Activity tracking



Kinesis Features



Real-time
performance



High throughput



Elastic



AWS integrations



Low cost



Easy administration



Course Scenario



Process stream of tweets

Calculate analytics

Work with Kinesis using

- Java libraries
- Big Data frameworks
- Lambda
- SQL analytics

Course Plan

Advanced Kinesis
consumers



Kinesis Client Library



Kinesis Firehose



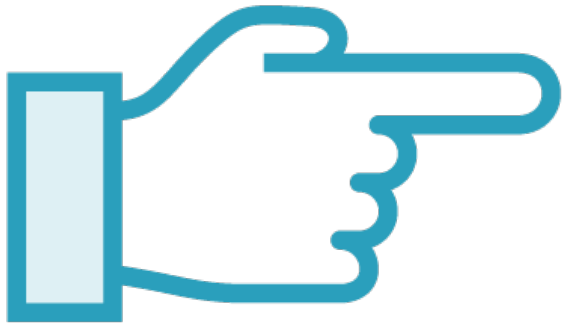
Kinesis Analytics



Kinesis Streams



Course Prerequisites



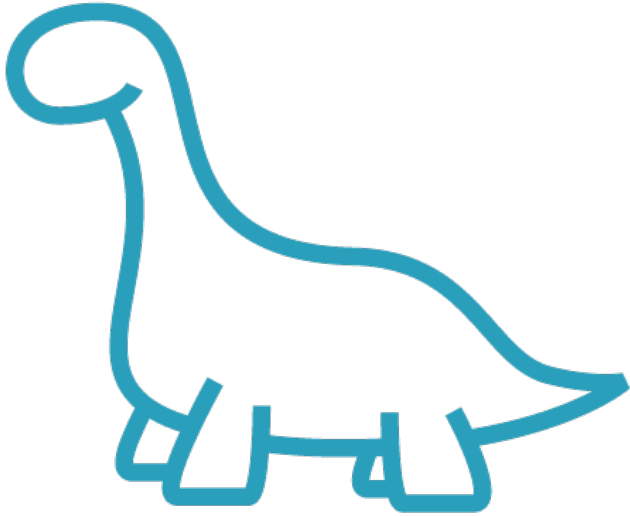
What you should know:

- Core Java
- Basics of AWS
- AWS CLI

Not necessary to know:

- Kinesis
- Stream processing
- Big Data

History of Kinesis

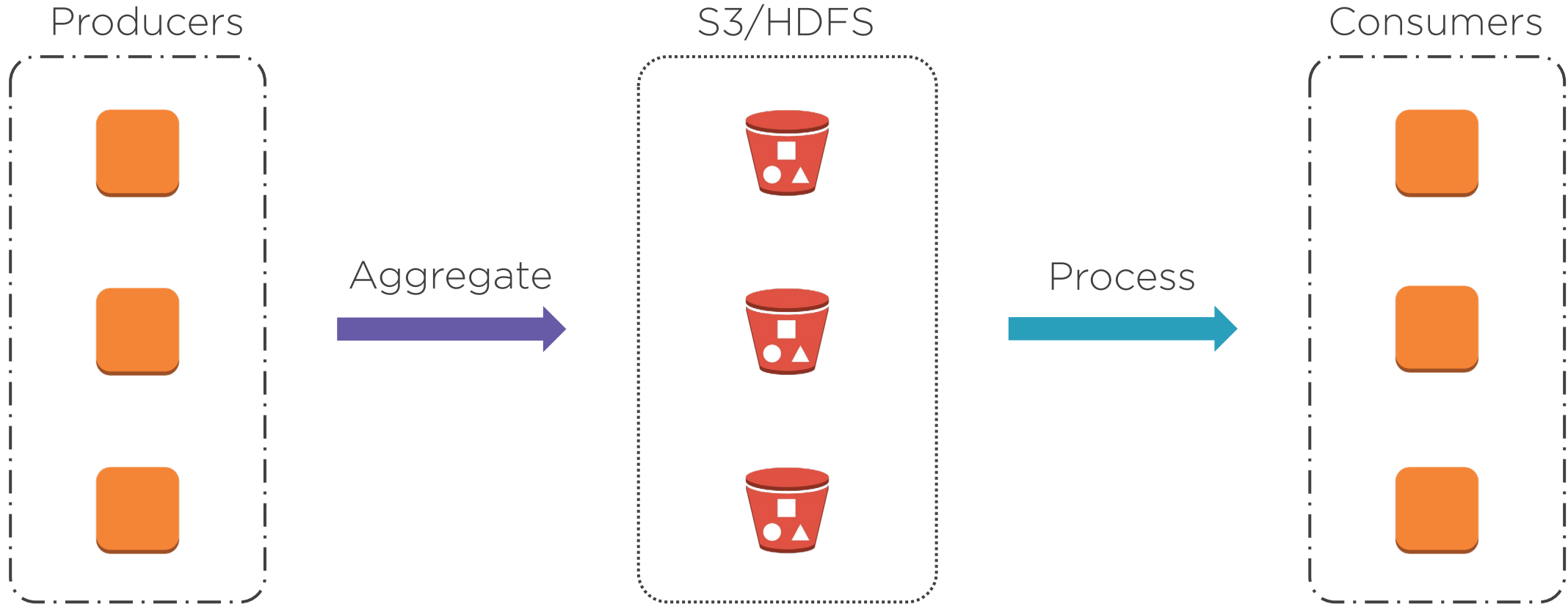


Where Kinesis came from

Why do we need new tool

How is it evolving

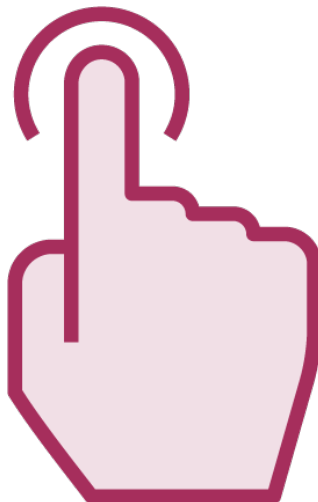
Batch Processing



Value of Recent Data



Stock market



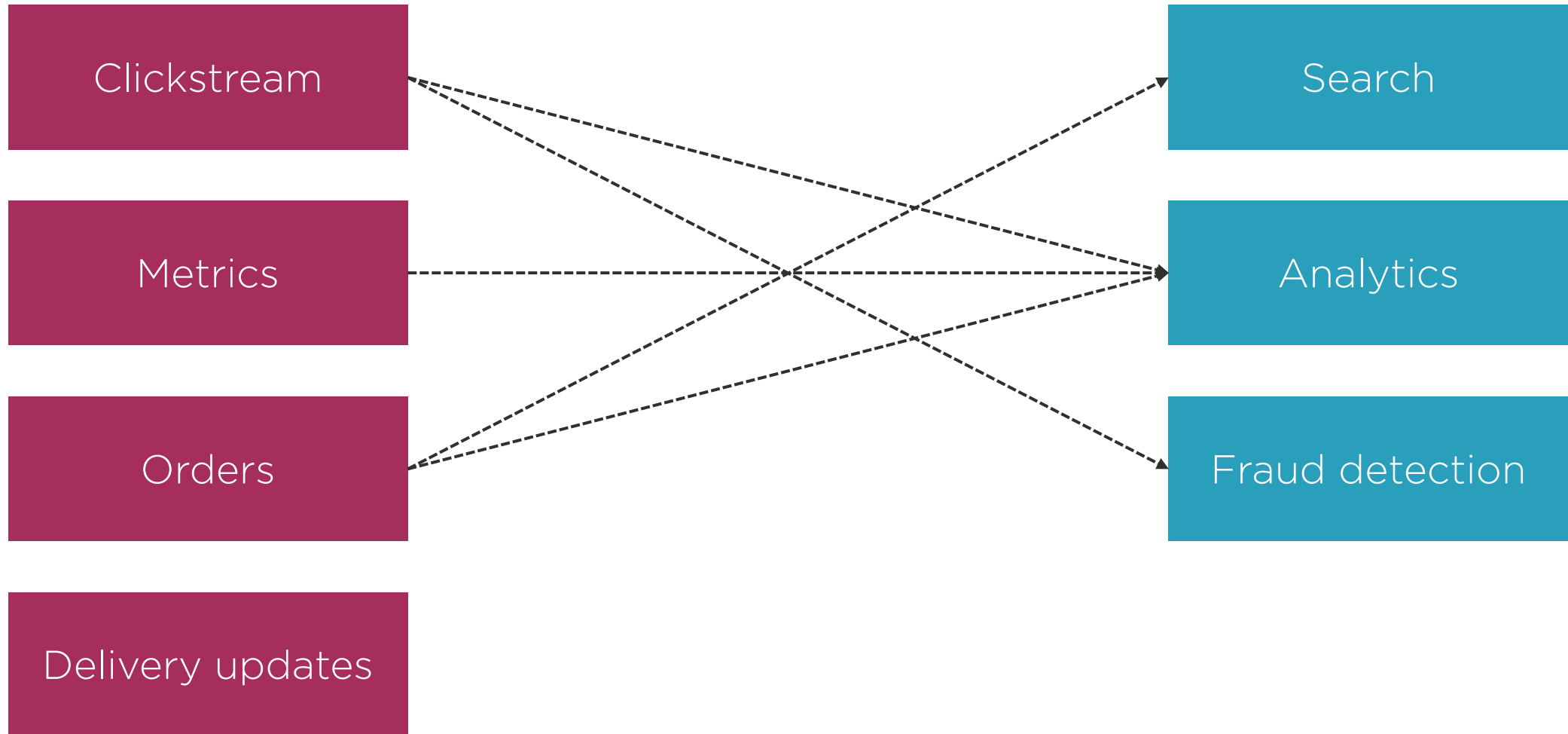
Clickstream analysis



Fraud detection



Old Architecture



Not a Good System Design



Need to implement each point-to-point connection

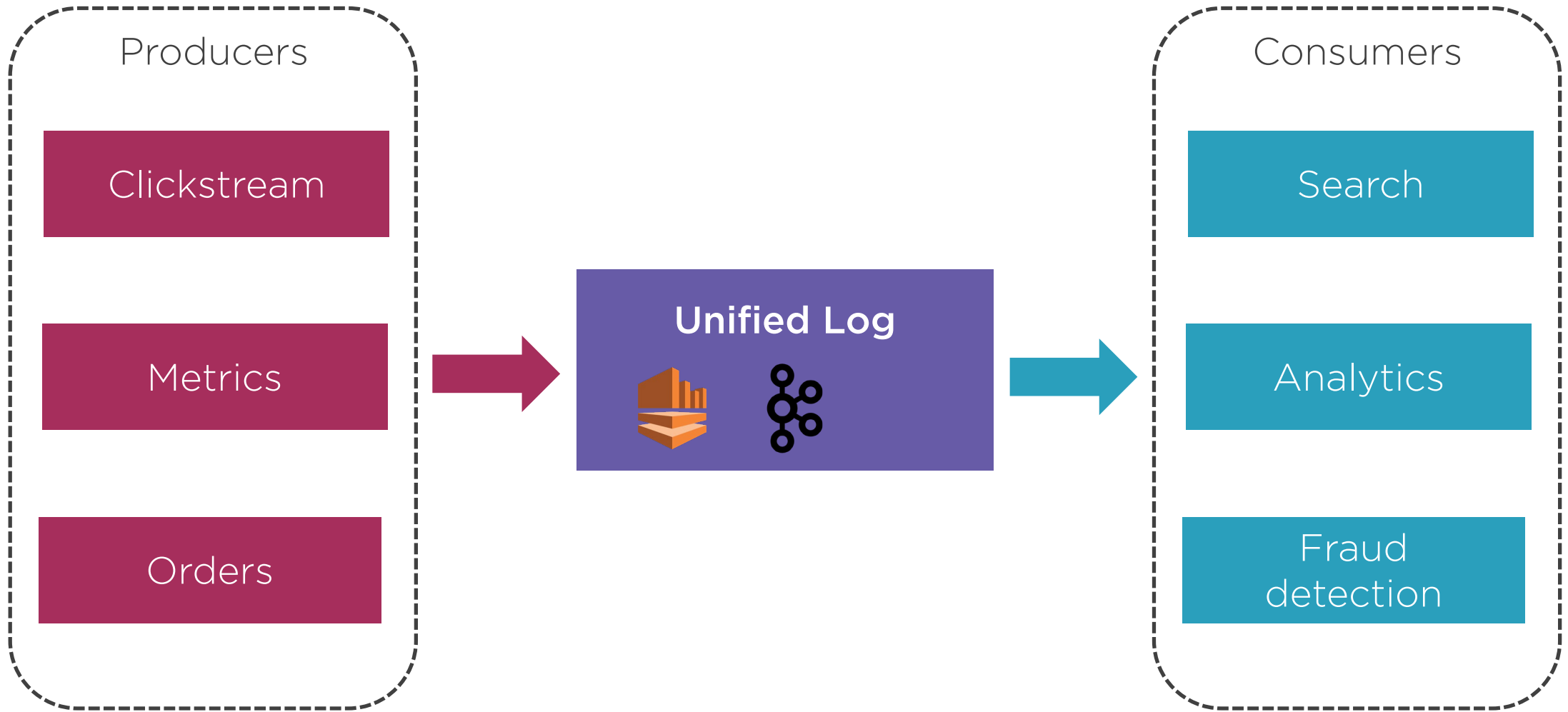
Each one is complex to implement

Should be maintained

Up to N^2 connections to maintain

A mess!

Unified Log



History of Kinesis

2008

LinkedIn
Internal data pipeline

2013

Amazon
Kinesis



2015

Amazon
Kinesis Firehose



2011

LinkedIn
Apache Kafka



2014

Google
Cloud Dataflow

2016

Amazon
Kinesis Analytics



Kinesis Building Blocks



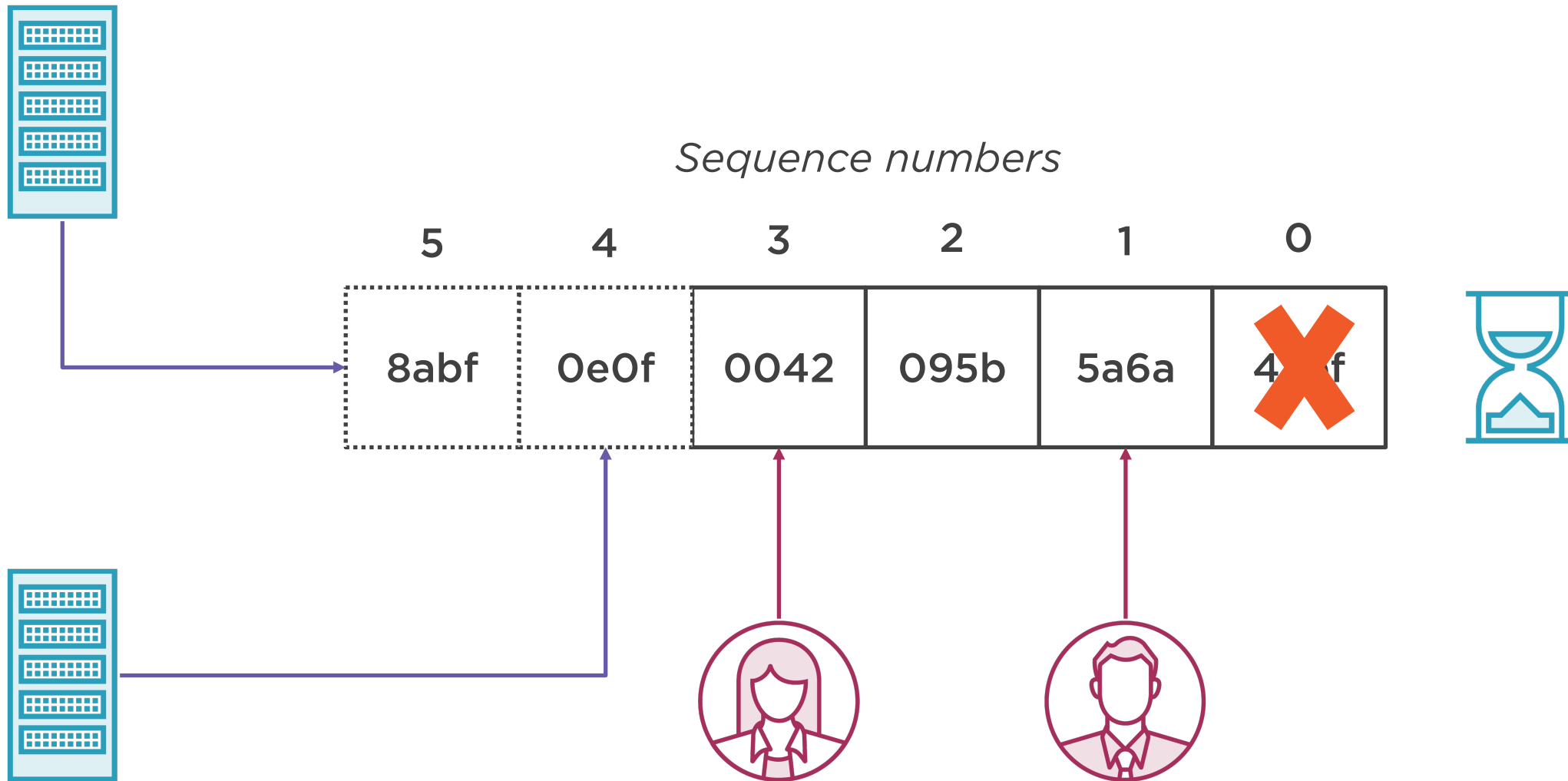
Streams in Kinesis

How to write data to Kinesis

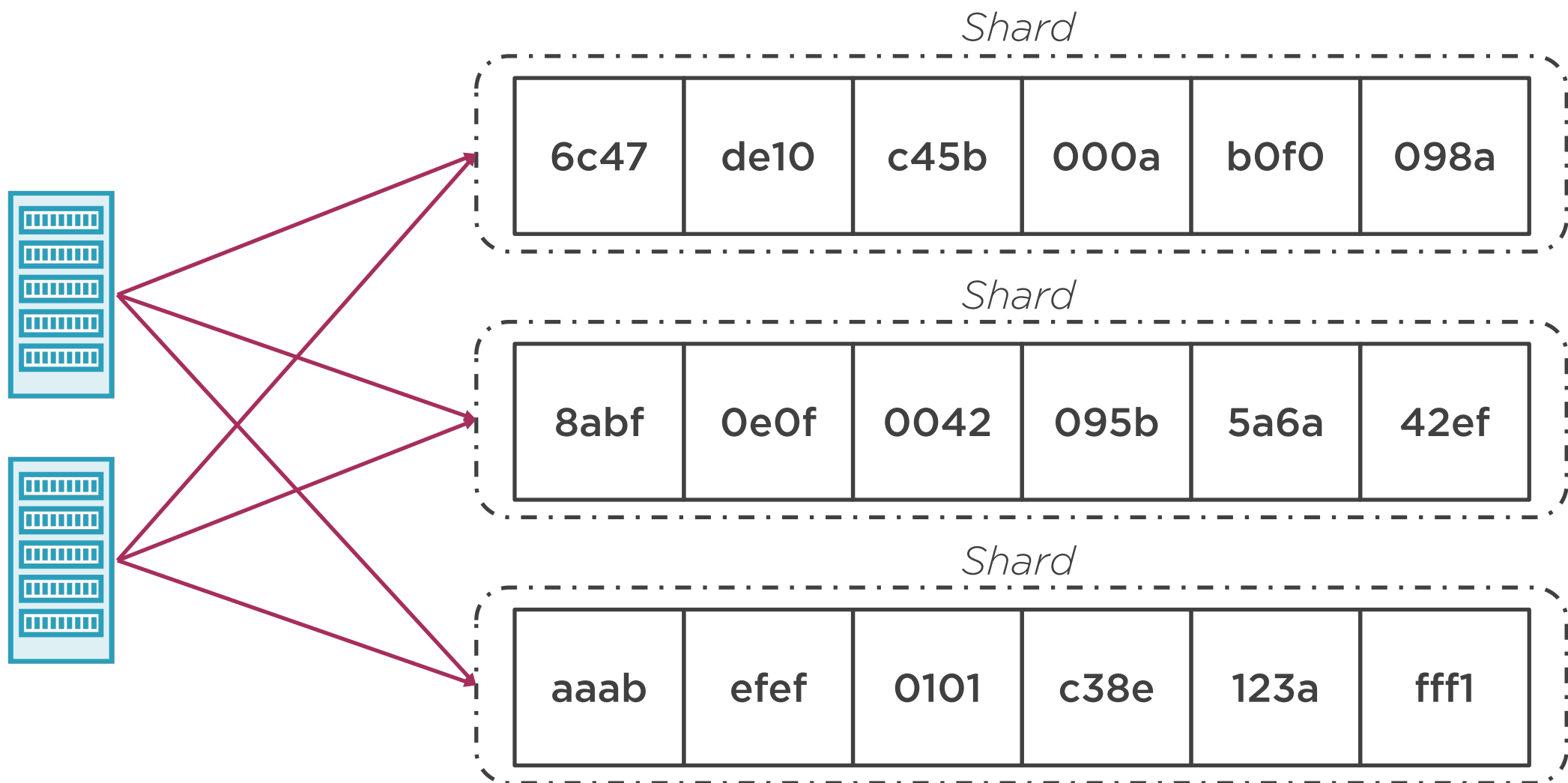
How data is stored in Kinesis

How data is read from Kinesis

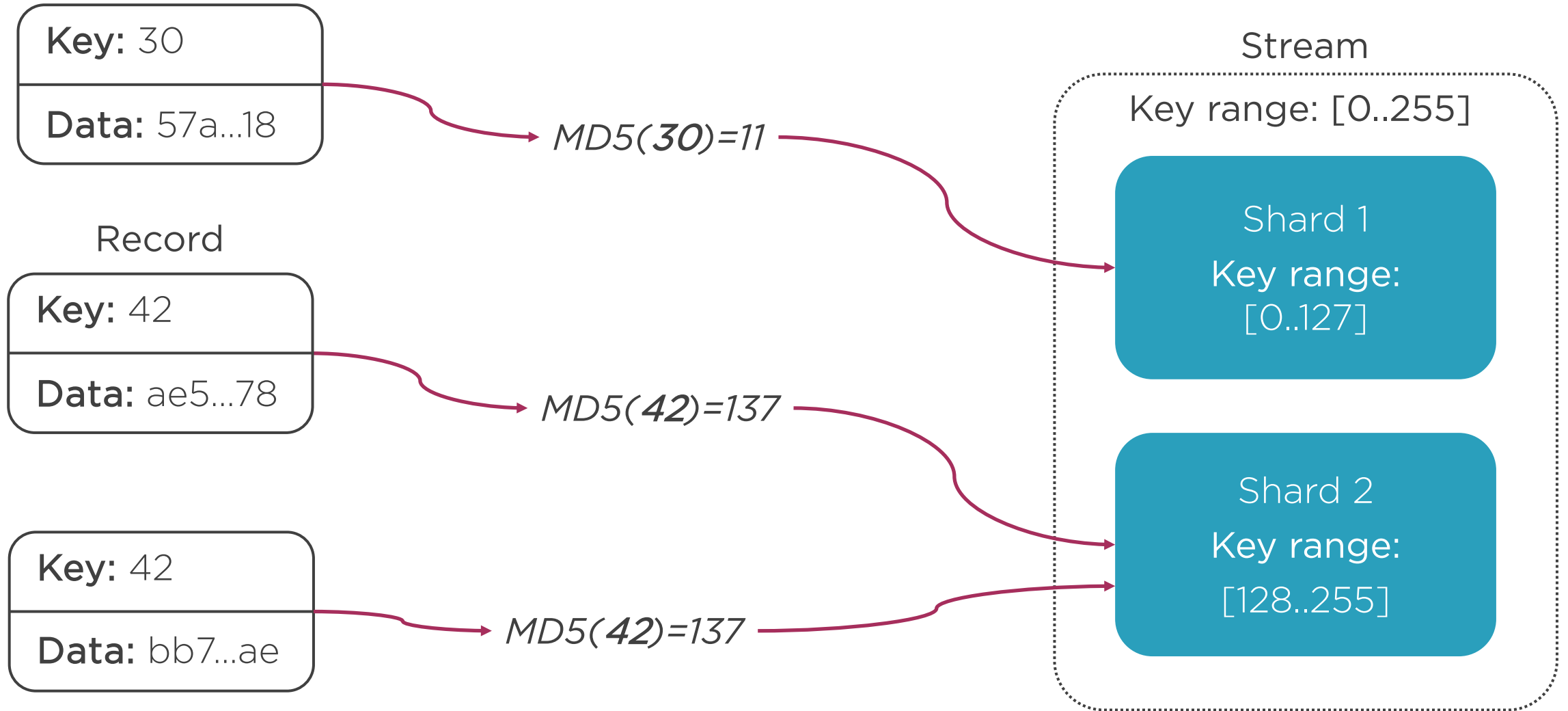
Kinesis Stream



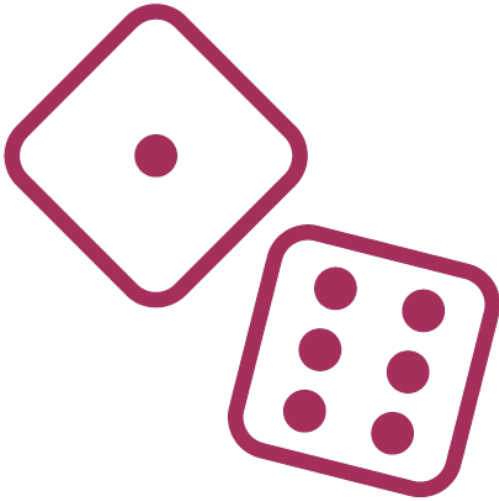
Scaling Kinesis Stream



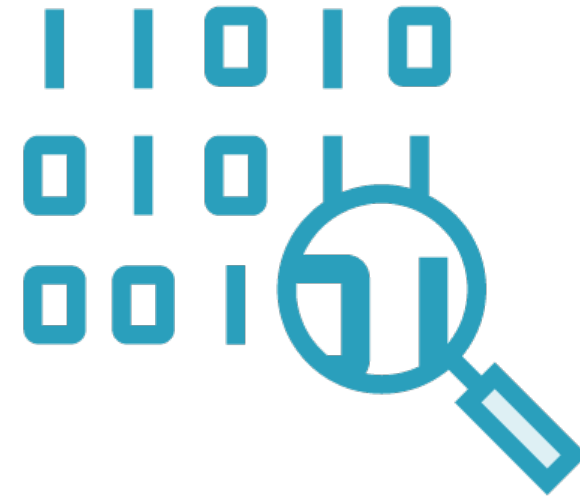
Writing Data to Kinesis



How to Select a Key

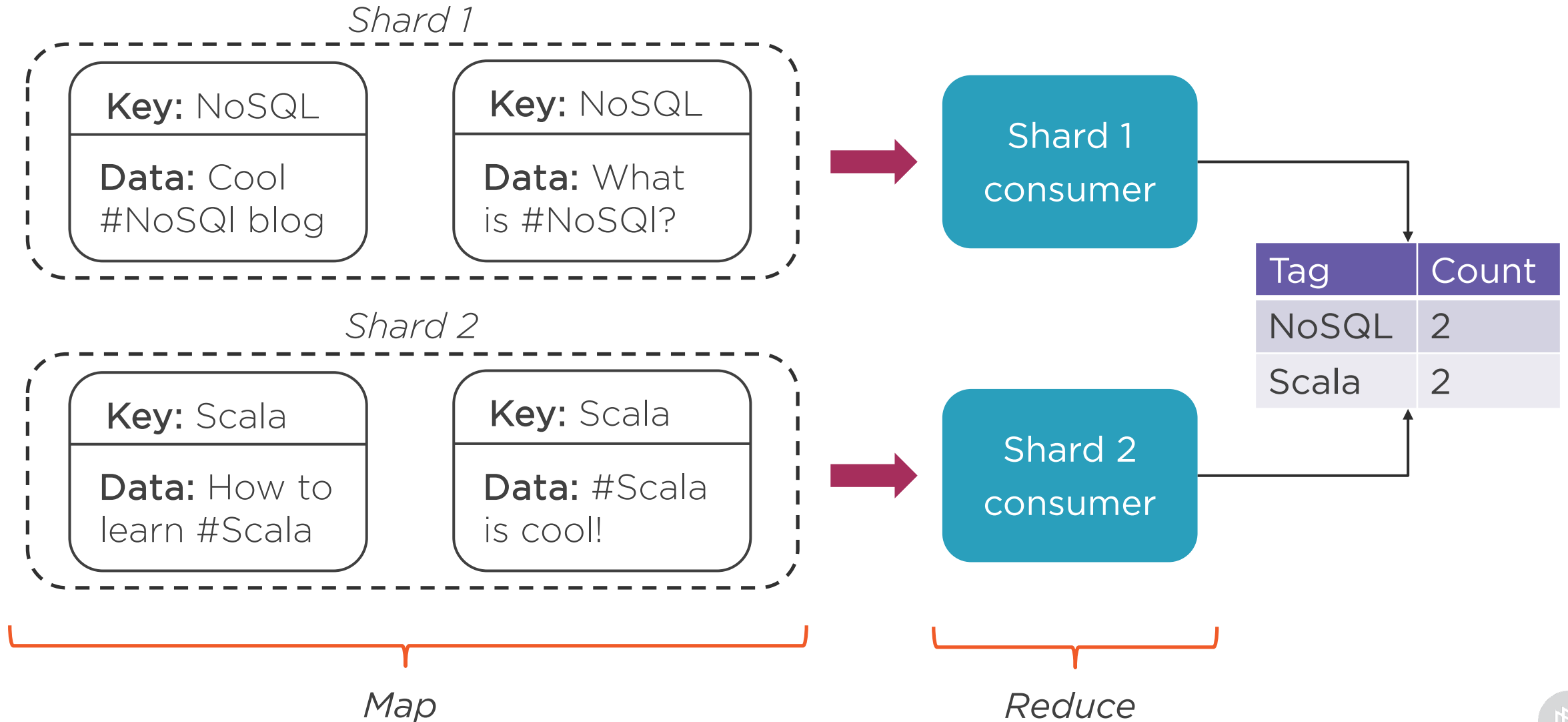


Random

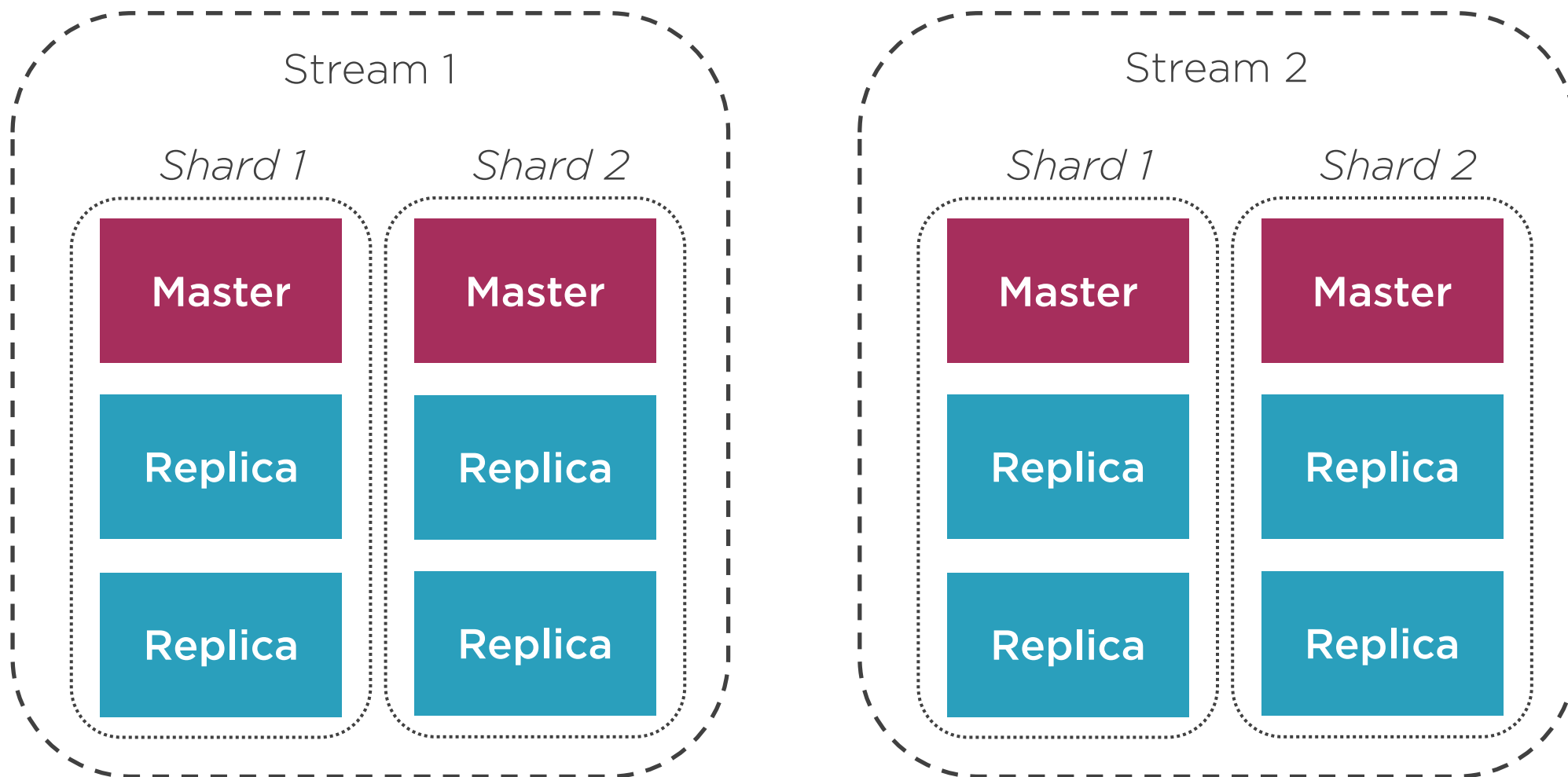


Deduct from data

Using Meaningful Keys



Kinesis Streams



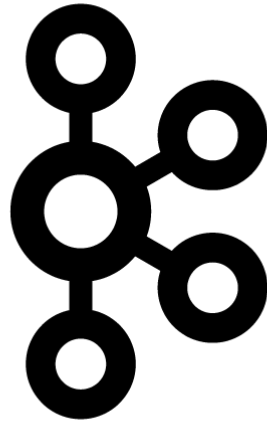
Number of shard in Kinesis stream is dynamic and can be changed at any time.



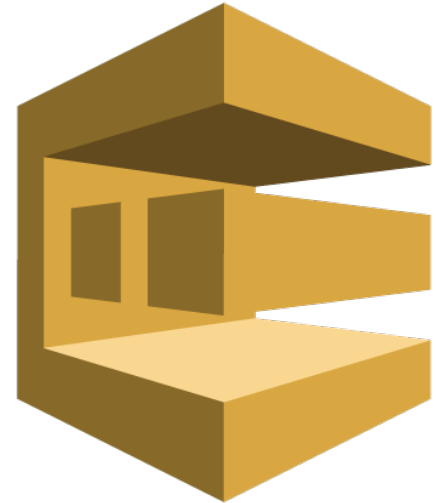
Technology Comparison



Kinesis



Kafka



SQS



Kinesis vs. Kafka

Kinesis

Messaging system

Stream/Shard

Proprietary

No operational load

Stores data for up to 7 days

Kinesis Client Library

More integrated with AWS ecosystem

Kafka

Messaging system

Topic/Partition

Open Source

Low to heavy operational load

Can store data indefinitely

Various Kafka clients

Has additional features



Kinesis vs. SQS

Kinesis

Messaging system

Built by AWS

No operational load

State tracking by client

Ordered

Message can be read many times

Unified log/Stream processing

SQS

Messaging system

Built by AWS

No operational load

State tracking by SQS

Order not guaranteed

Processed message is removed

Balancing tasks among workers



Demo

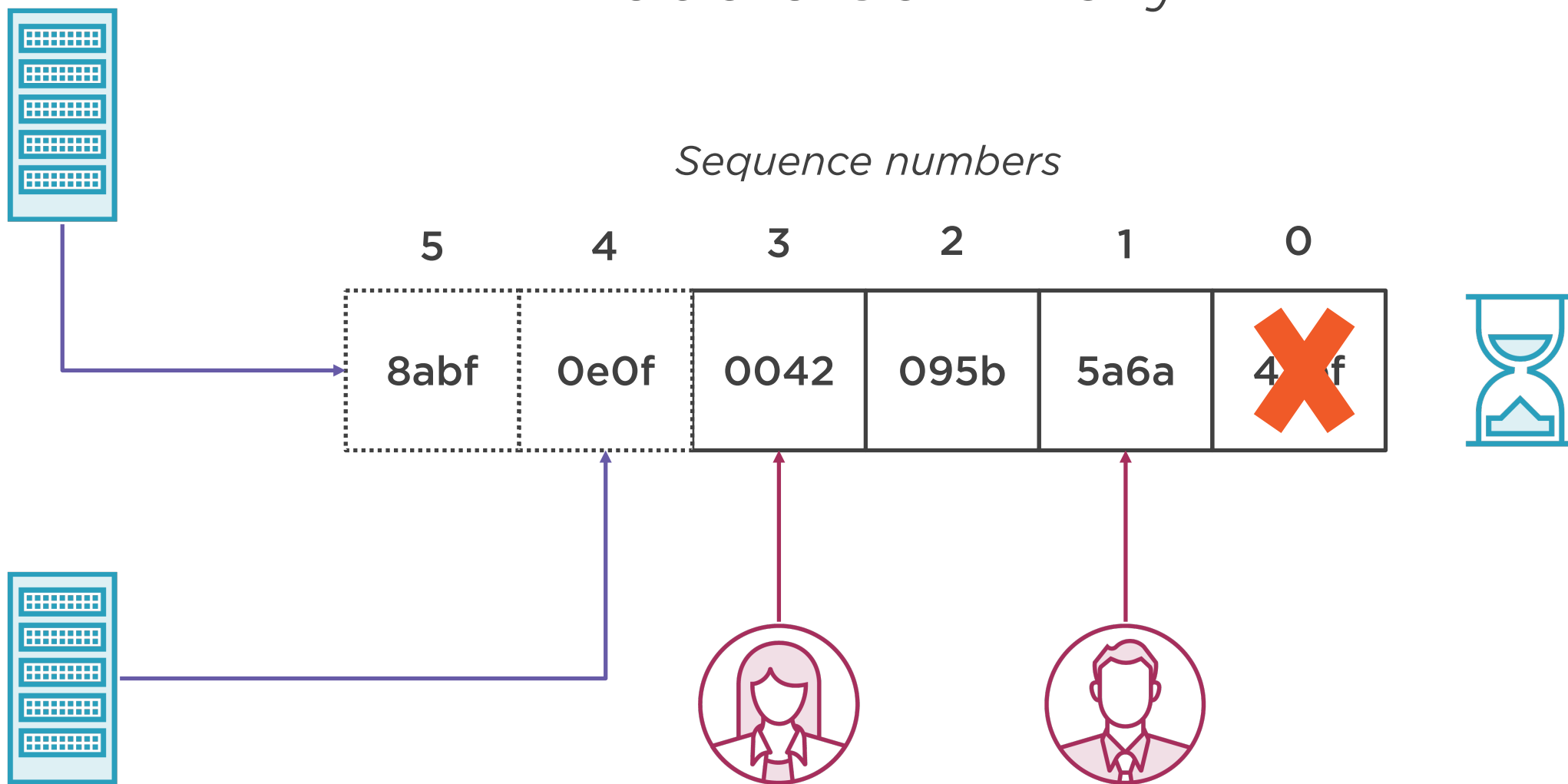


Create AWS account

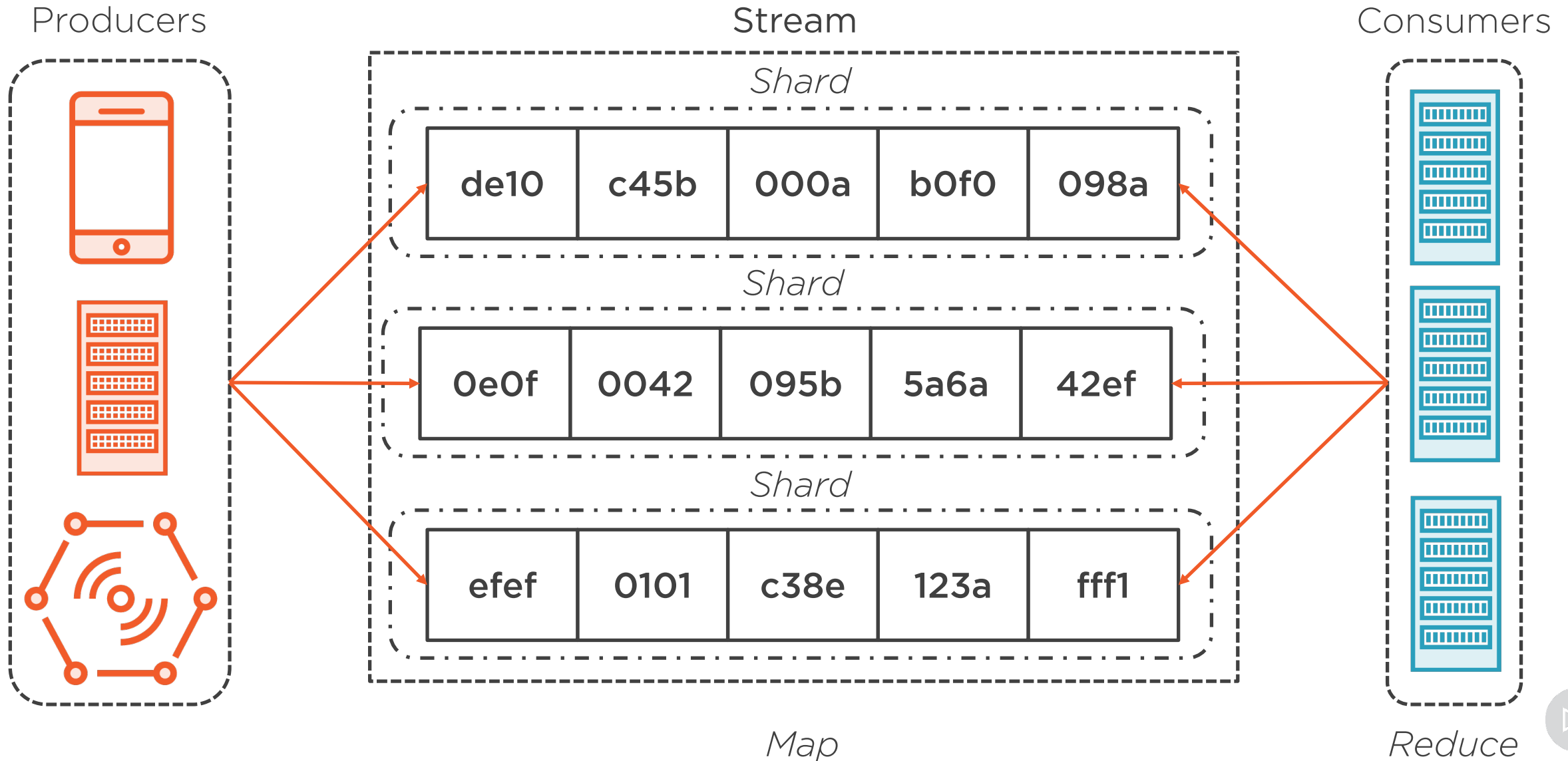
Important for upcoming examples



Module Summary



Module Summary



Kinesis Streams

