Apache Kafka Overview

Subtitle

Social handles

Team or presenters name

Date

Agenda

Apache Kafka use cases

Apache Kafka 101

Challenges operating Apache Kafka

aws

Apache Kafka

aws

Apache Kafka use cases

Real-time web and log analytics

Messaging

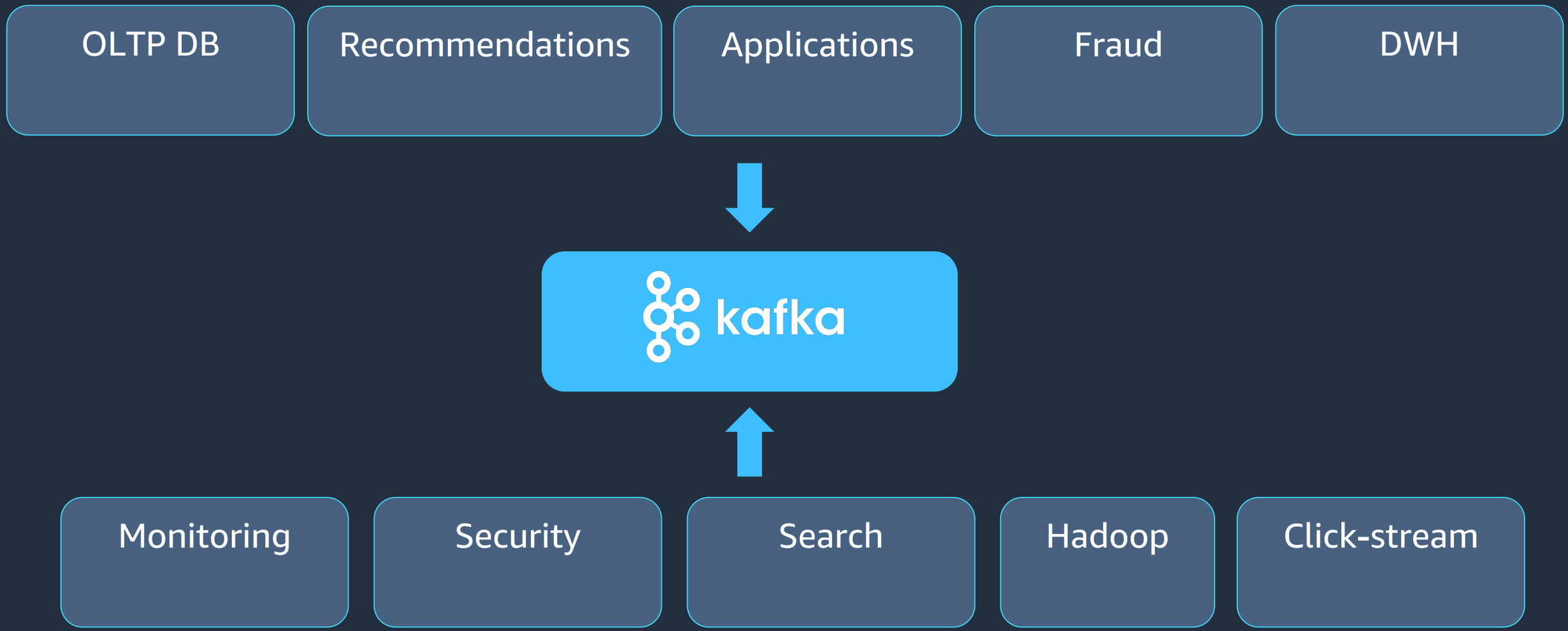
Transaction and event sourcing

Decoupled microservices

Streaming ETL

aws

Why Apache Kafka?



aws

Apache Kafka : anatomy

Apache Kafka is a “Distributed Streaming Platform”

Apache Kafka has 3 core capabilities:

1) Publish and subscribe to streams of records (message queue)

2) Store streams of records in a fault-tolerant way

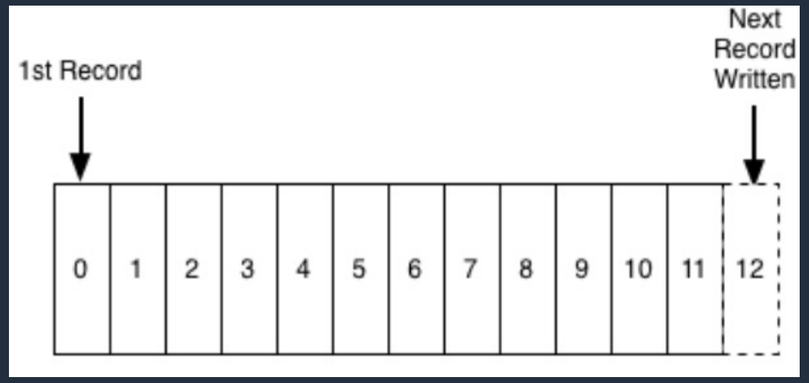
3) Process streams of records as they occur

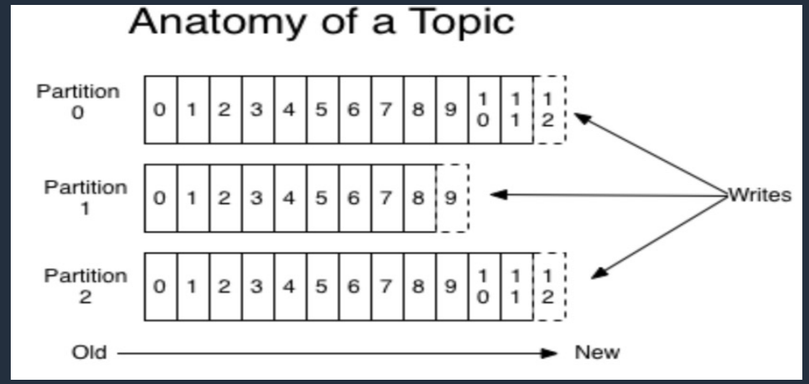
aws

Apache Kafka : anatomy

Key concepts

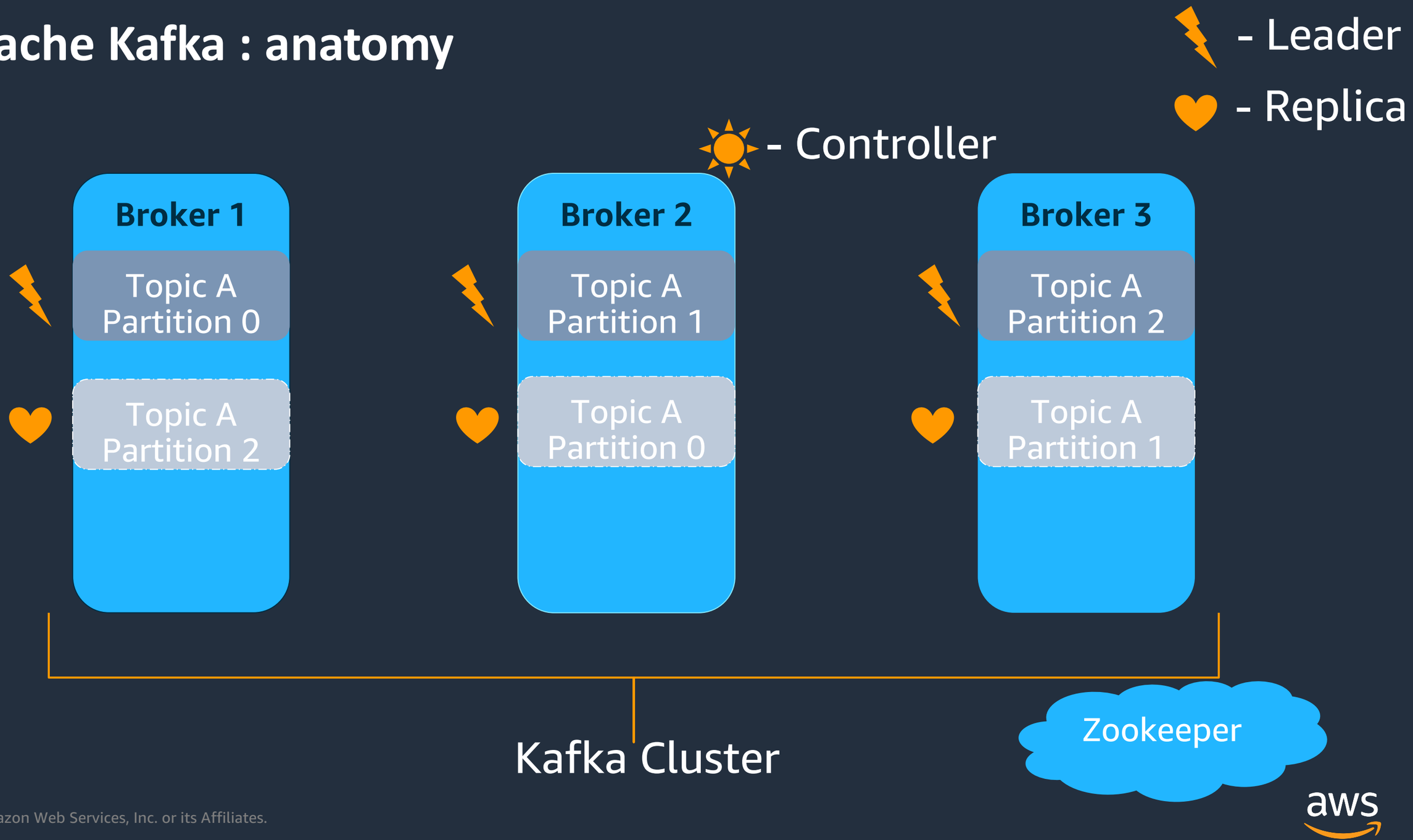
The log: An append-only, totally-ordered sequence of records ordered by time. It looks like this:





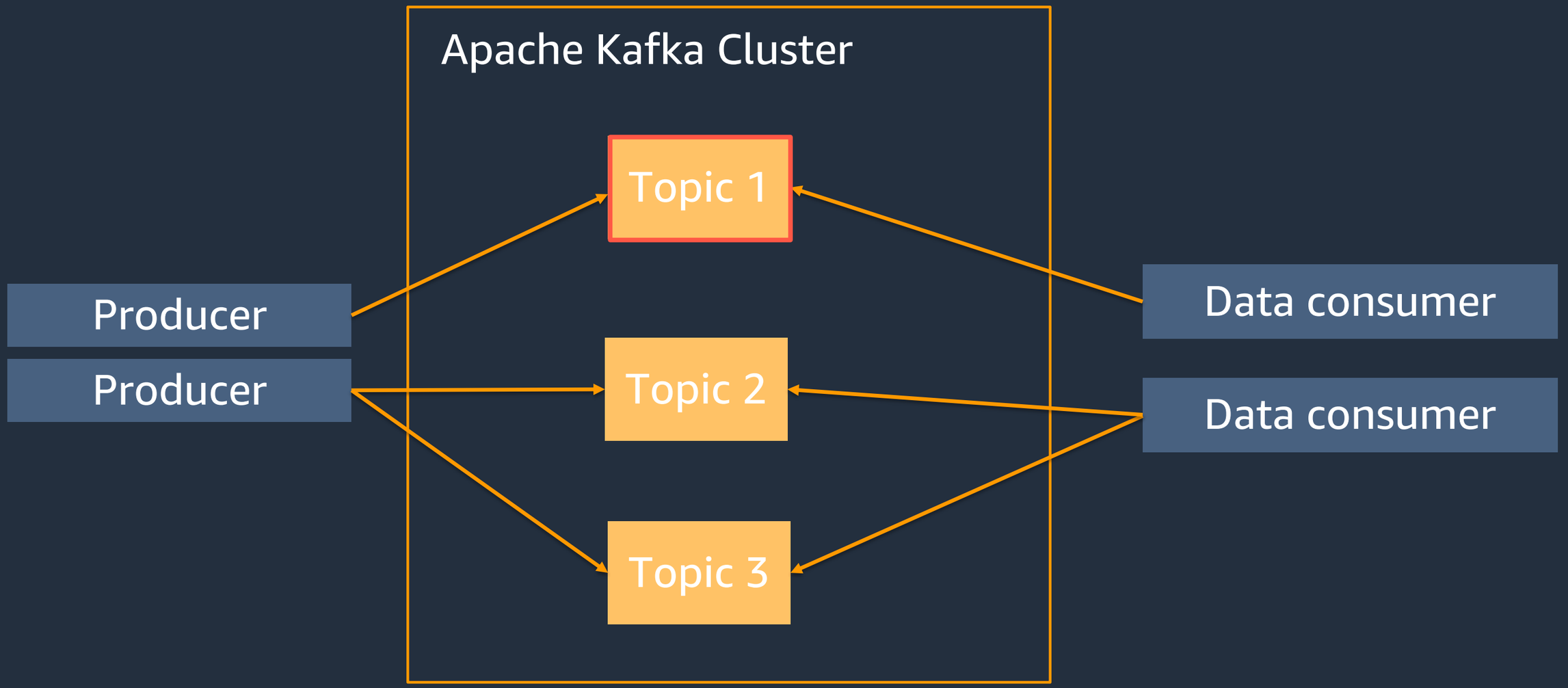
* Kafka is run as a cluster, one or more servers, multiple datacenters. (AZ's)
* Each record consists of a key, a value, and a timestamp.
* The Kafka cluster stores streams of *records* in categories called *topics*.

Apache Kafka : anatomy



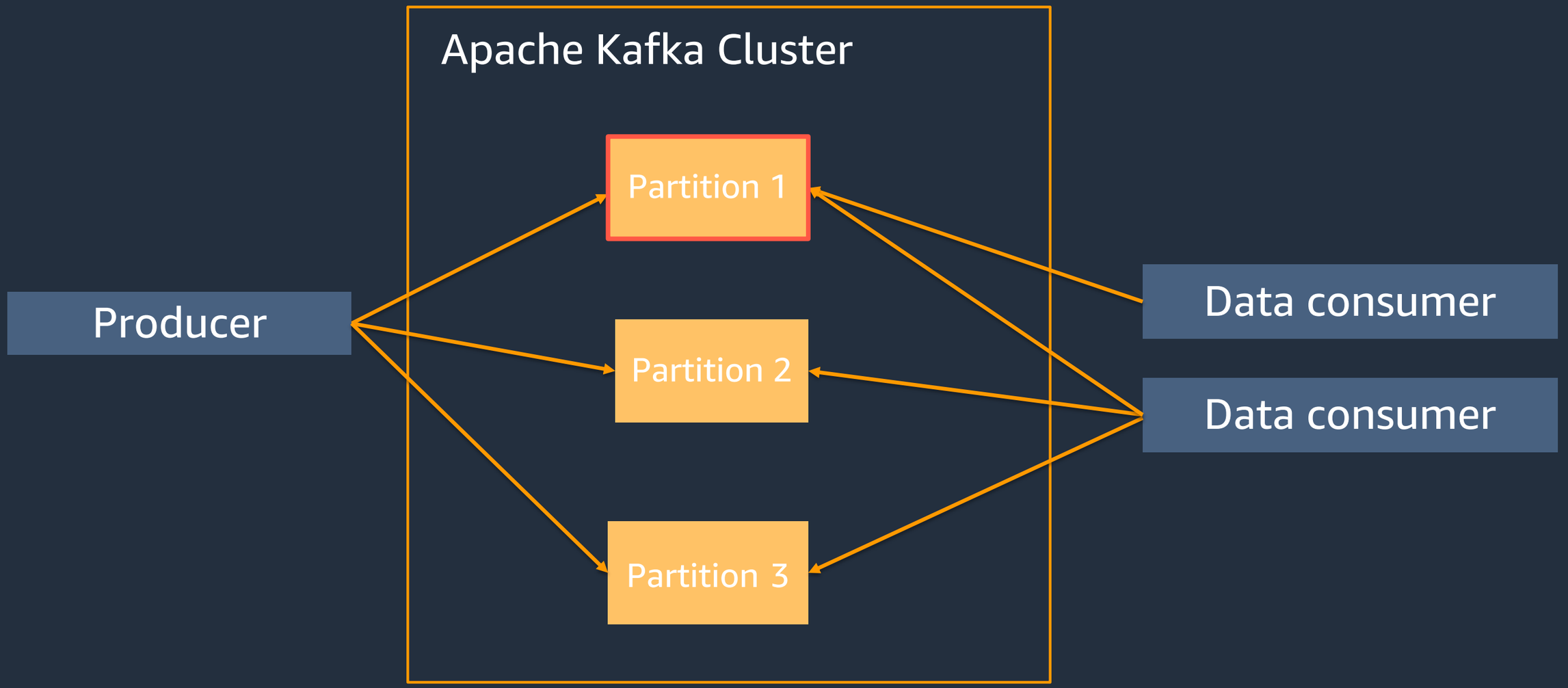
aws

Apache Kafka anatomy 101: Topics



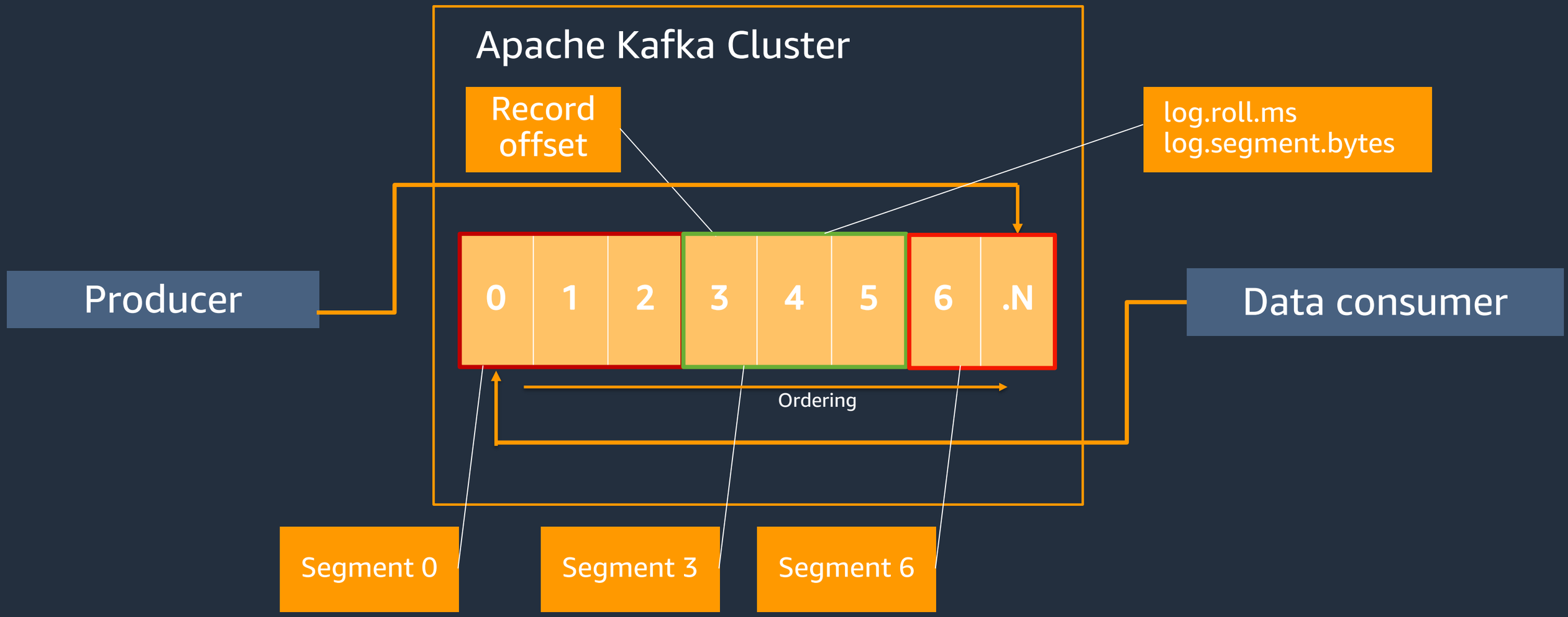
aws

Apache Kafka anatomy 101: Partitions



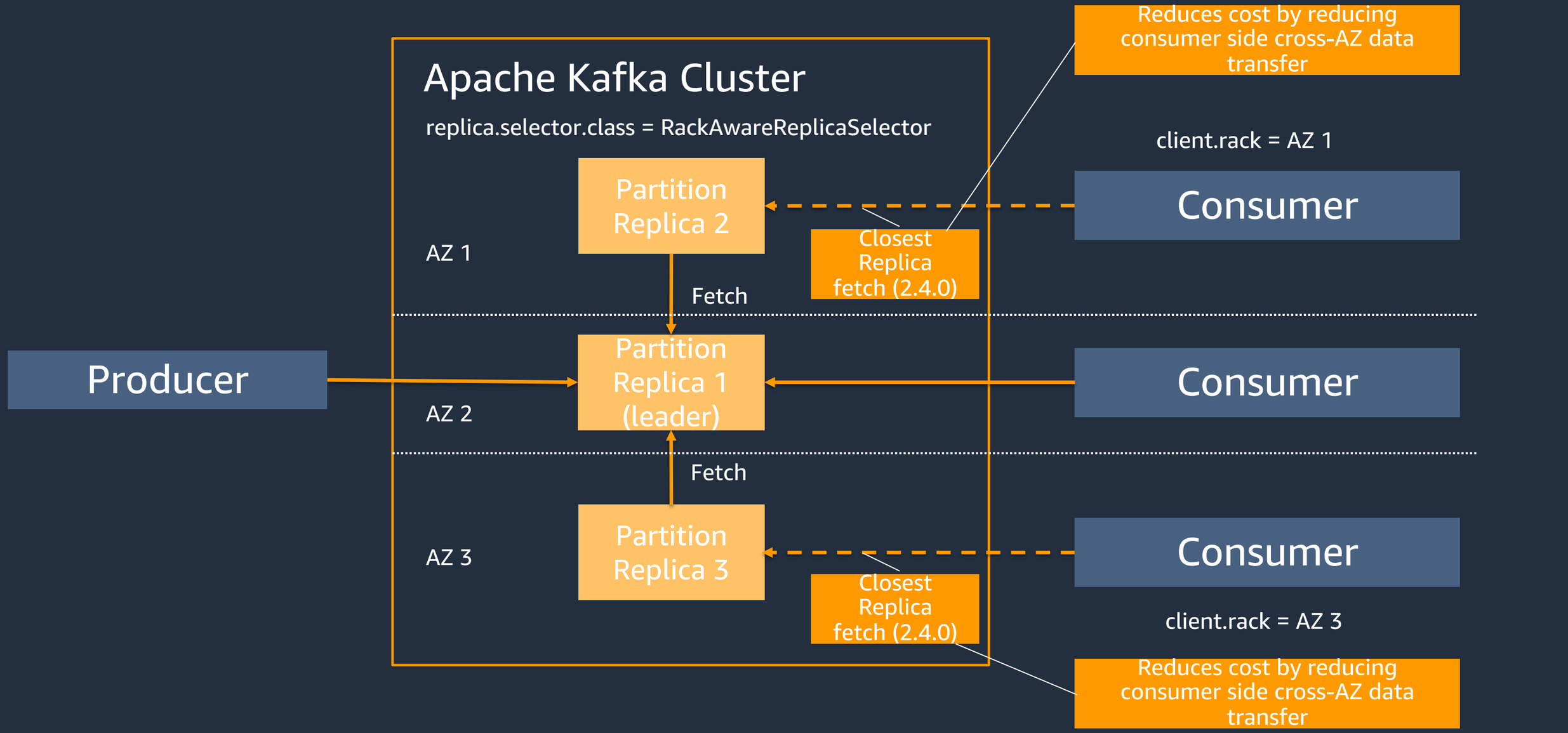
aws

Apache Kafka anatomy 101: Offsets

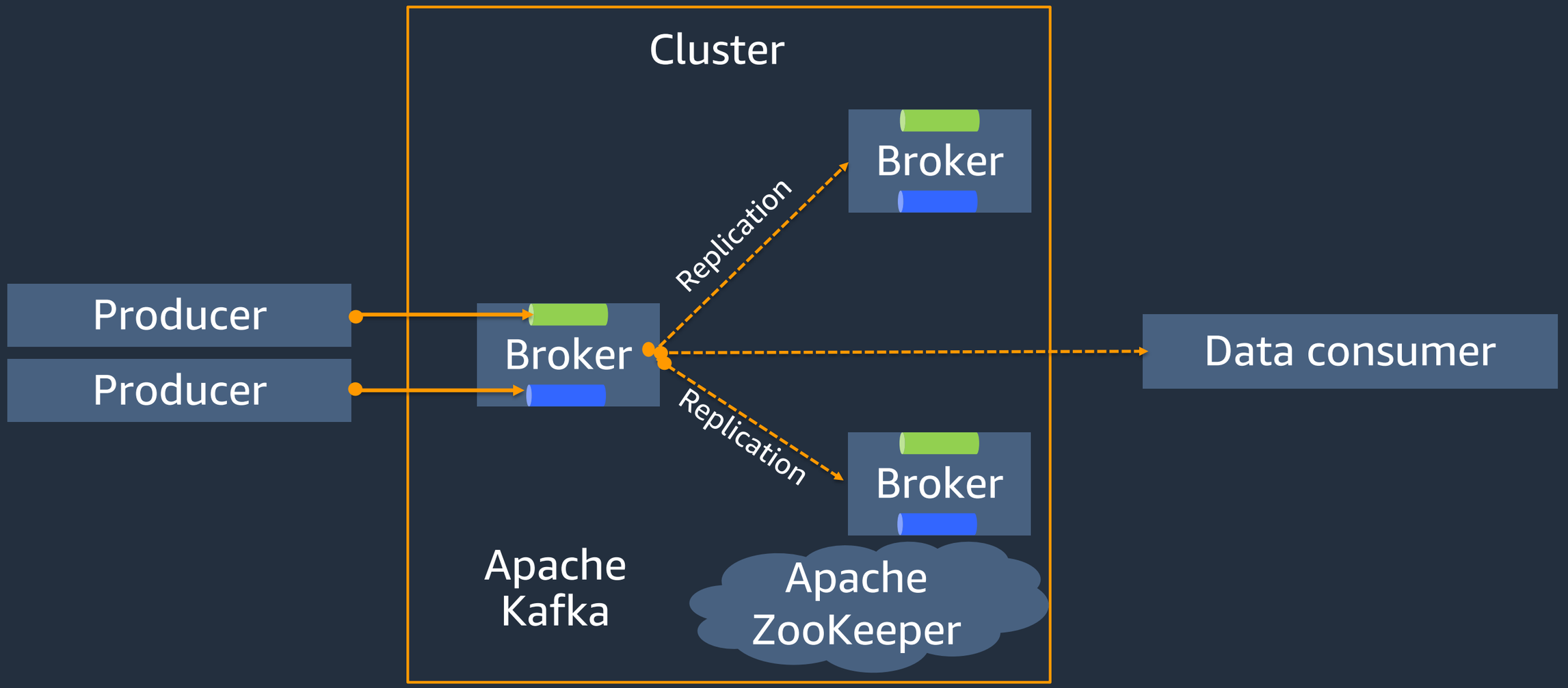


aws

Apache Kafka anatomy 101: Replicas

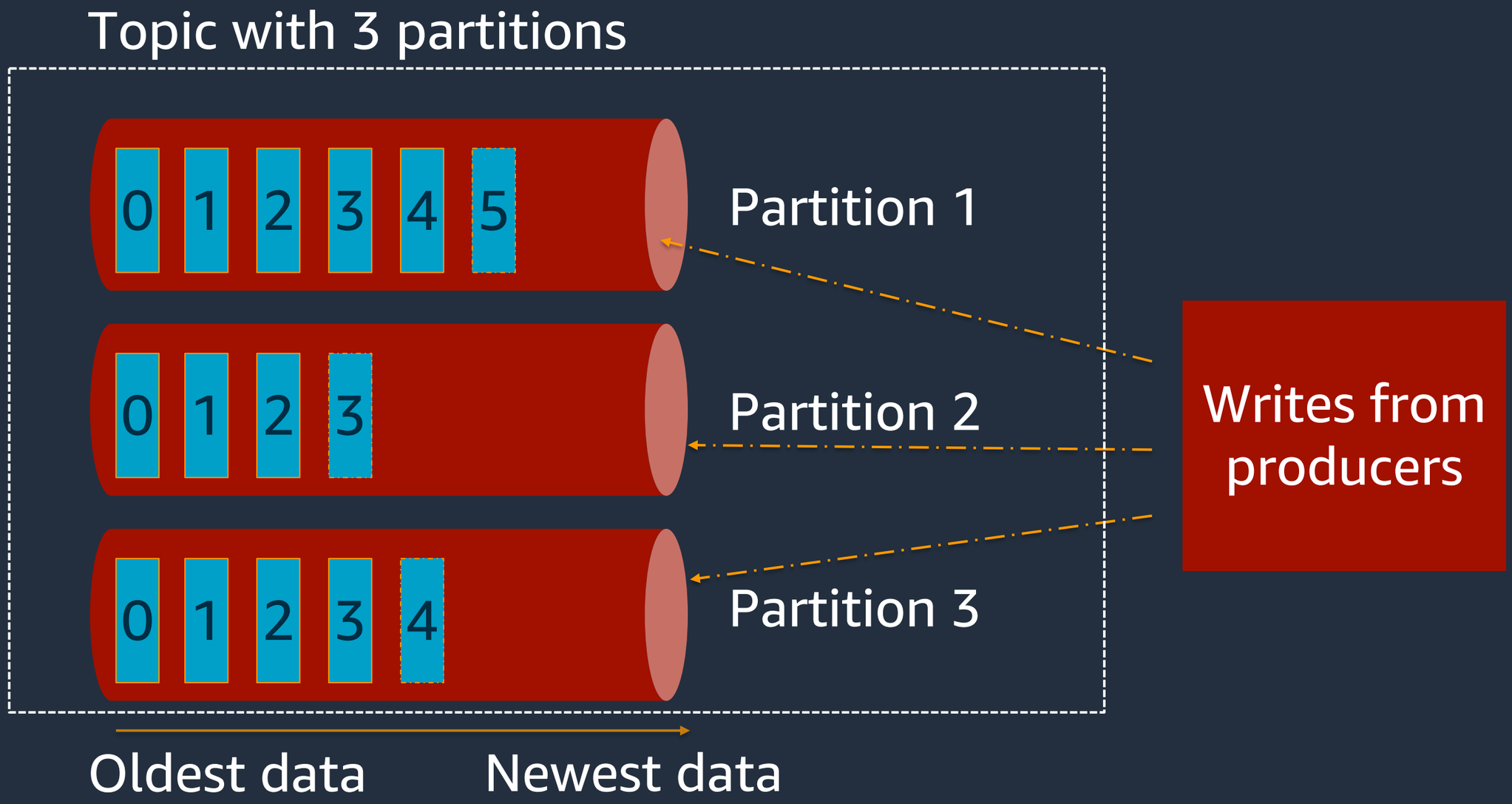


Apache Kafka anatomy 101



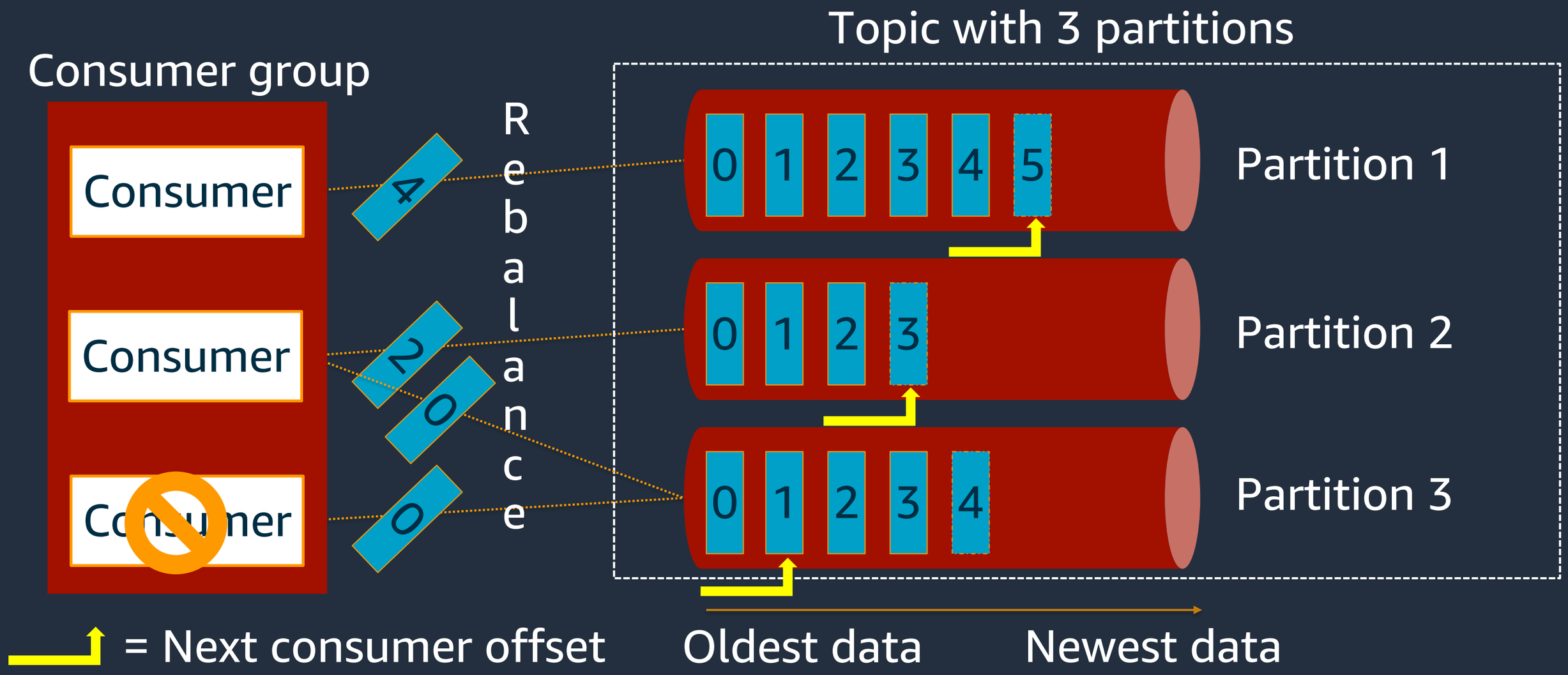
aws

Apache Kafka anatomy: Writes to partitions



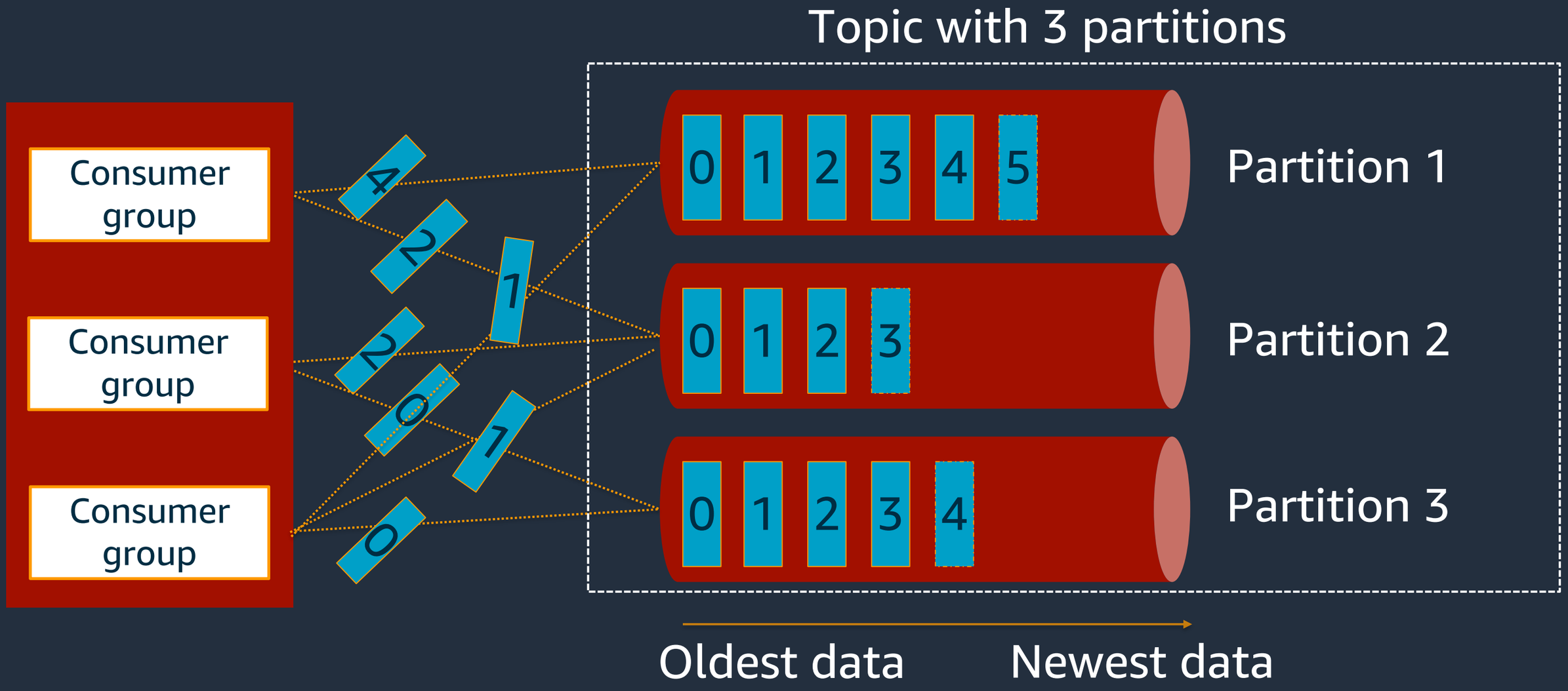
aws

Apache Kafka anatomy: Reads from partitions



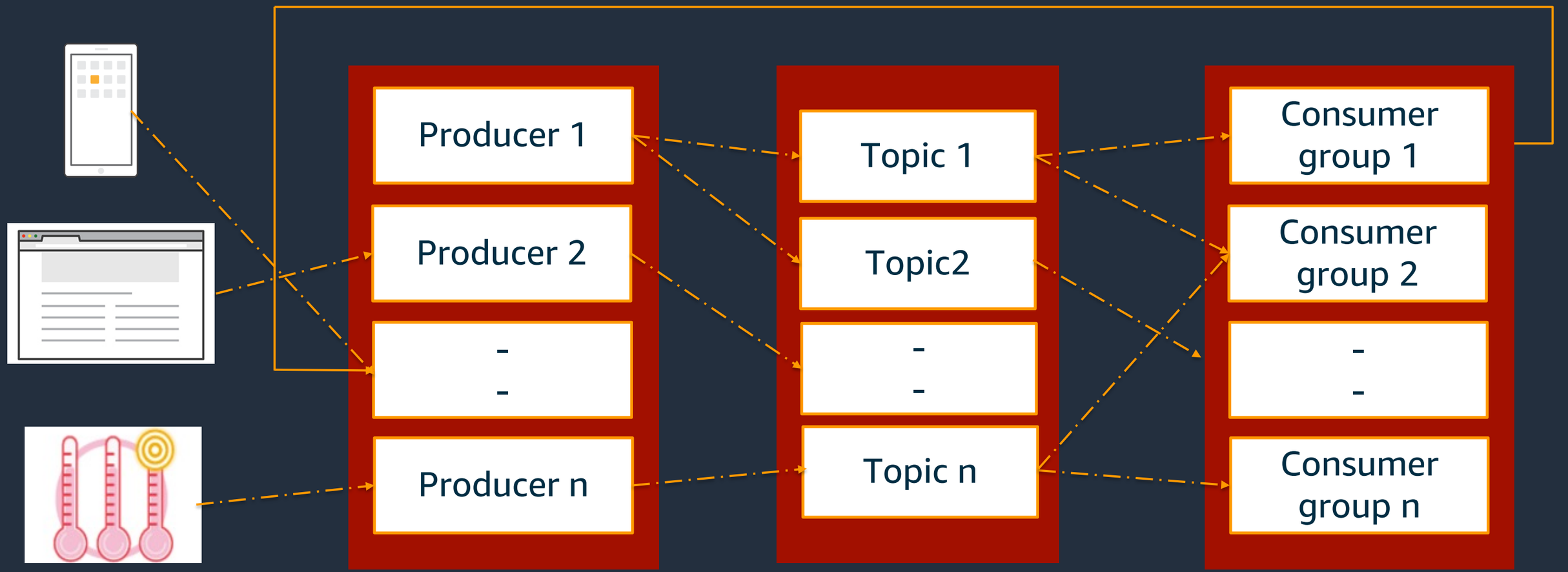
aws

Apache Kafka in pub/sub mode: Reads from partitions



aws

Apache Kafka in Event-driven Architecture



aws

Challenges operating Apache Kafka

Difficult to setup

Tricky to scale

Hard to achieve high availability

Integration required development

Error prone and complex to manage

Expensive to maintain

aws

Q & A

Thank You!

aws