



Amazon Managed Streaming for Apache Kafka

Build Kafka applications, not Kafka scaffolding

April 2021

Agenda

Why real-time analytics and data streaming?

Brief overview of Apache Kafka and use cases

Amazon Managed Streaming for Apache Kafka (MSK)

- Overview of current features
- Sizing your Amazon MSK clusters
- Migration and Best Practices

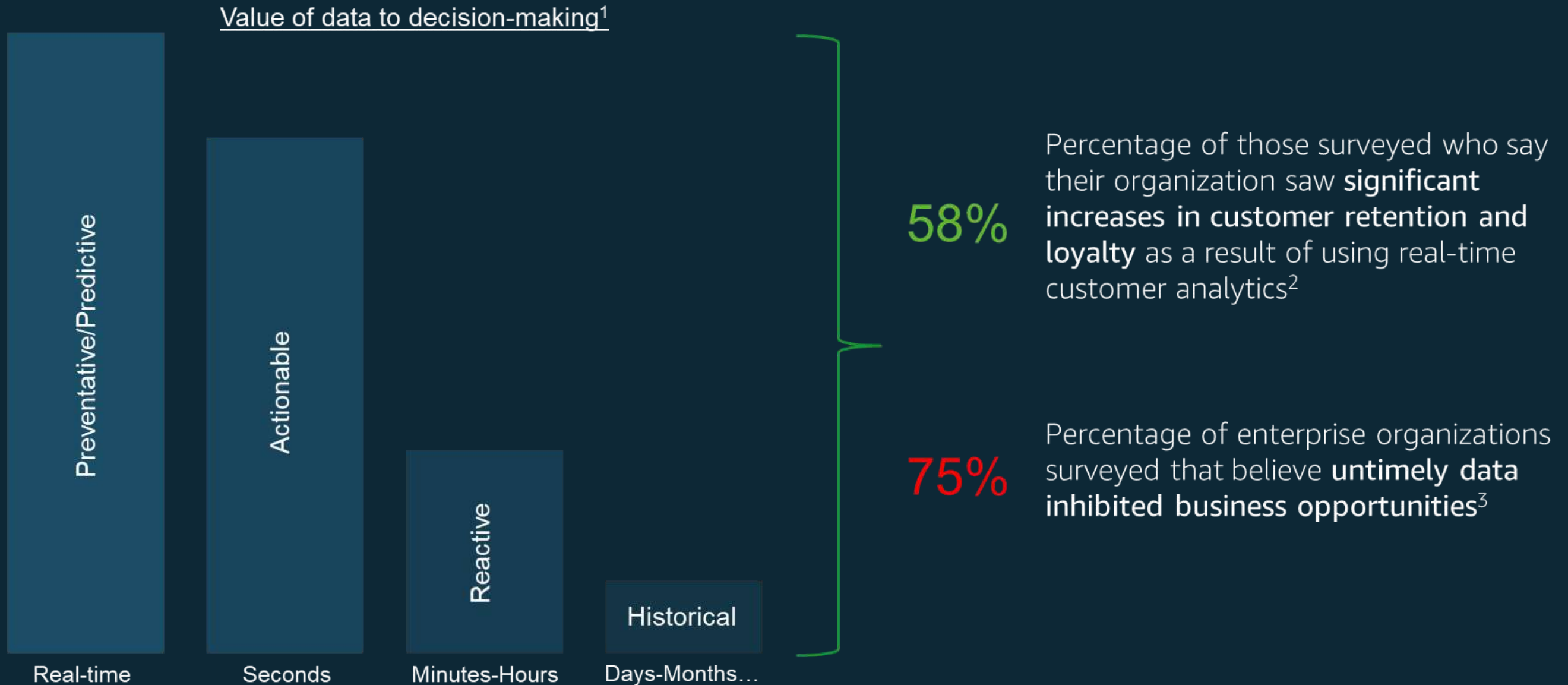
Customer successes

Getting started



The **value** of data diminishes over time

Data have a short shelf life of actionability¹. AWS lets you act on that data **as fast as the market dictates**.



Apache Kafka - Use Cases



Real-time web and log analytics



Transaction and Event Sourcing



Messaging



Decoupled Microservices



Streaming ETL



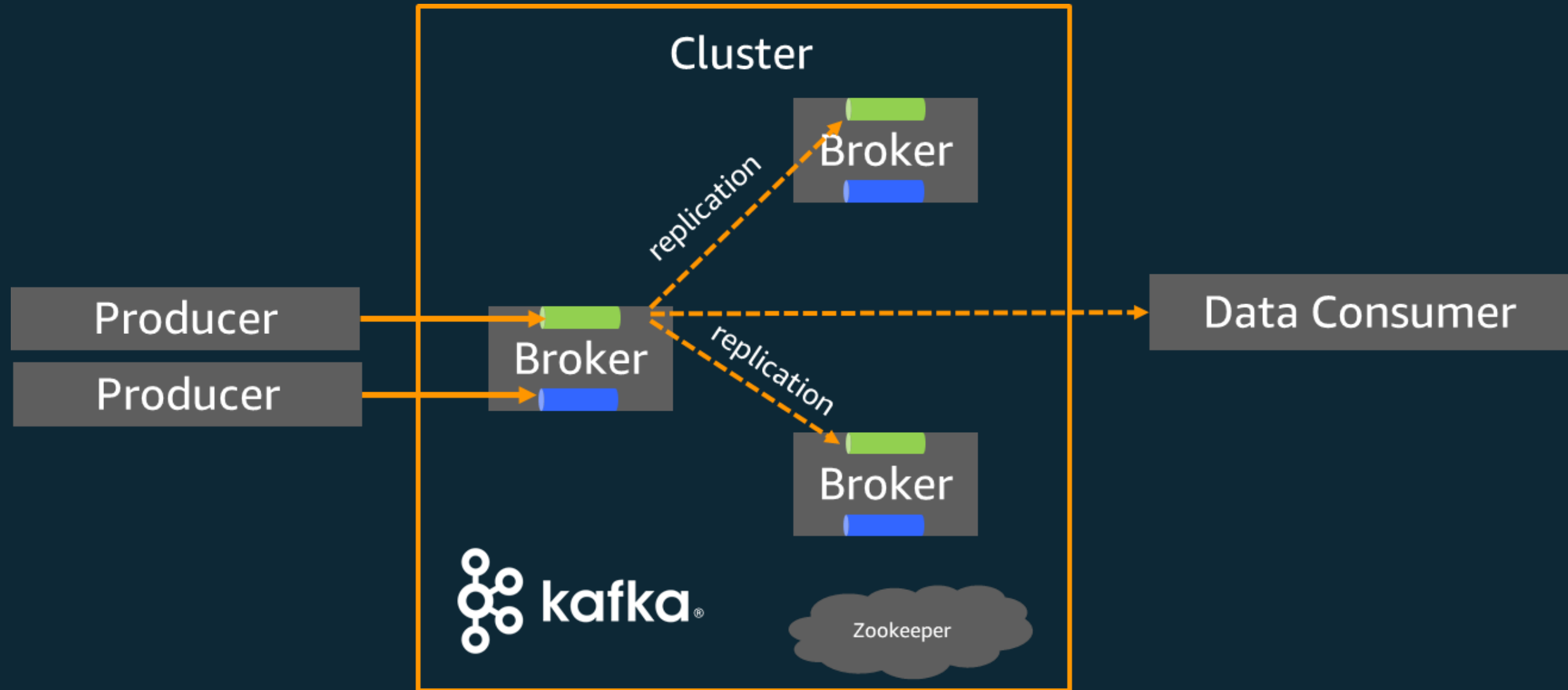
Metrics and Log Aggregation



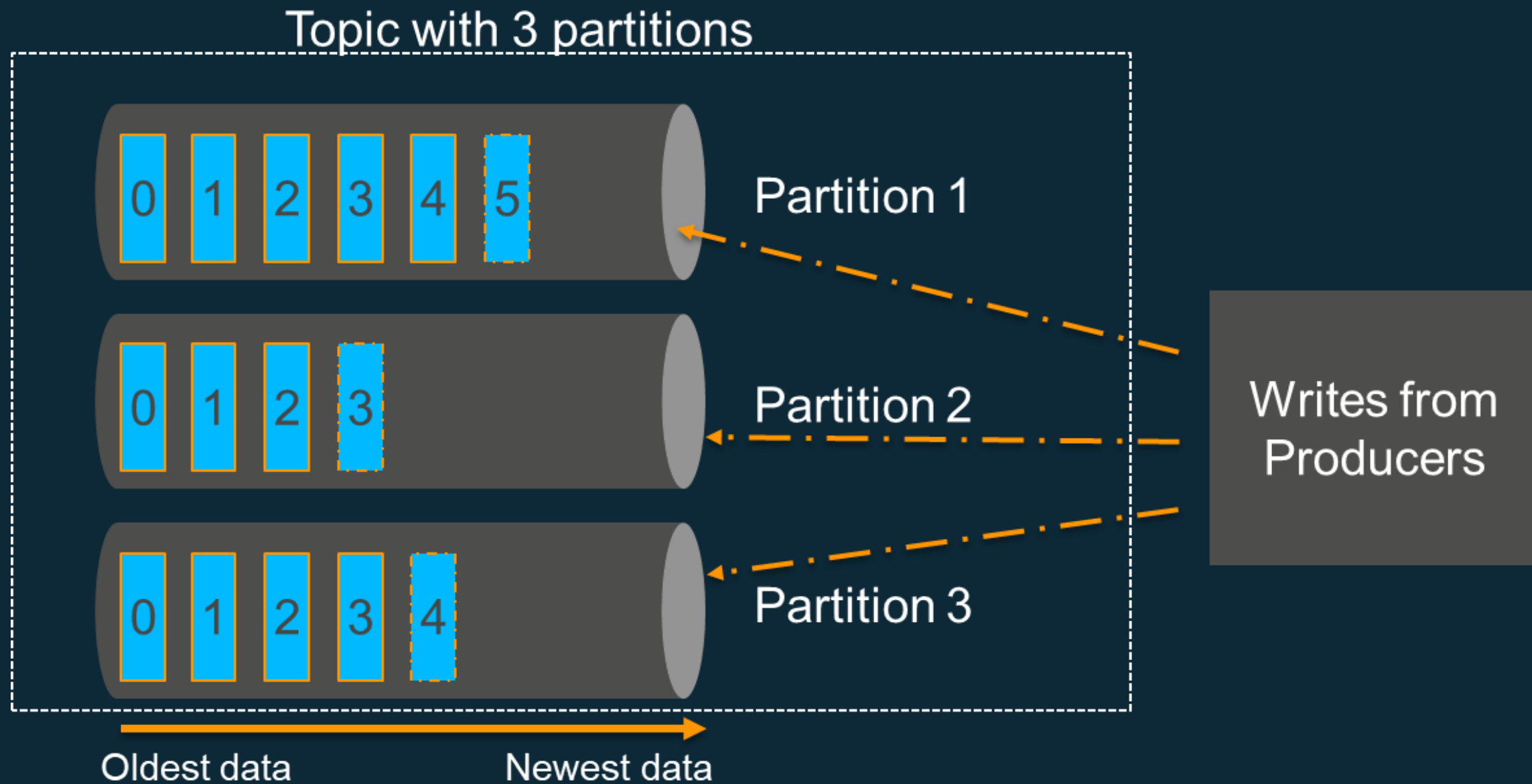
Apache Kafka



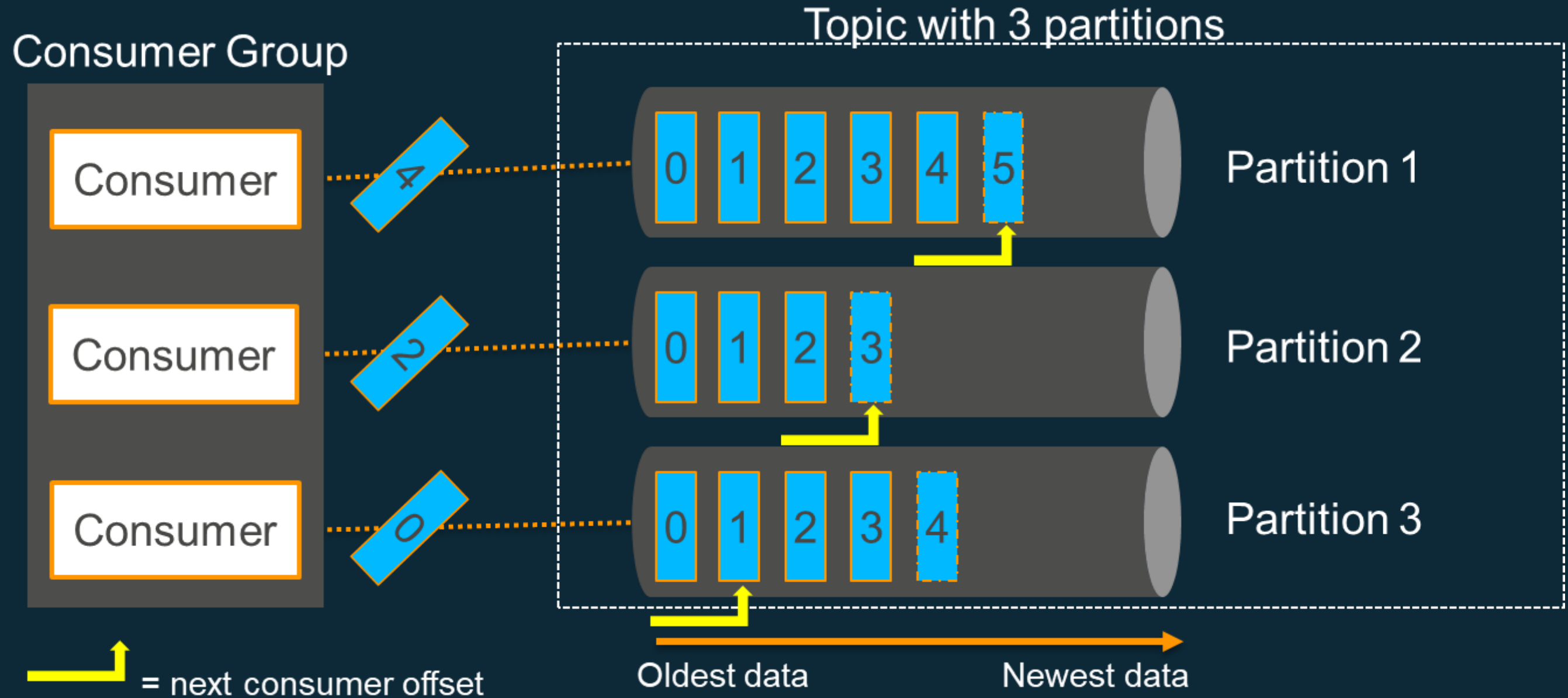
Apache Kafka anatomy 101



Apache Kafka anatomy - Writes to partitions



Apache Kafka anatomy - Reads from partitions



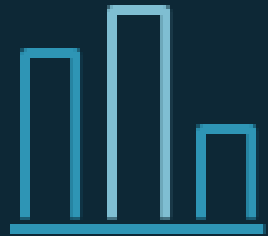
Challenges operating Apache Kafka



Difficult to setup



Tricky to scale



Hard to achieve high availability



Integration required development

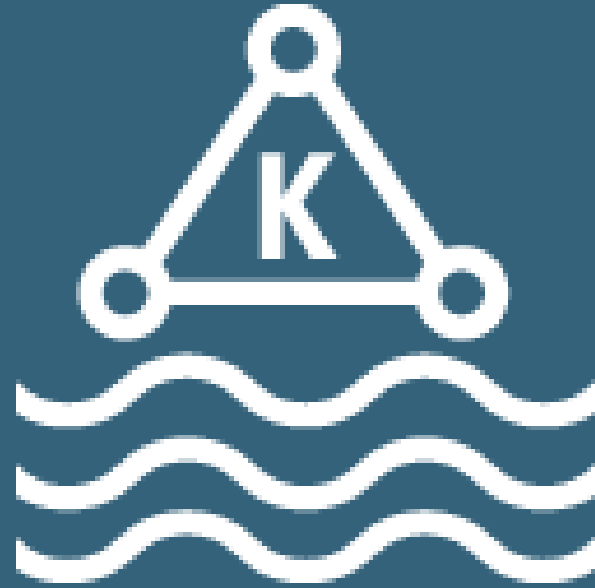


Error prone and complex to manage



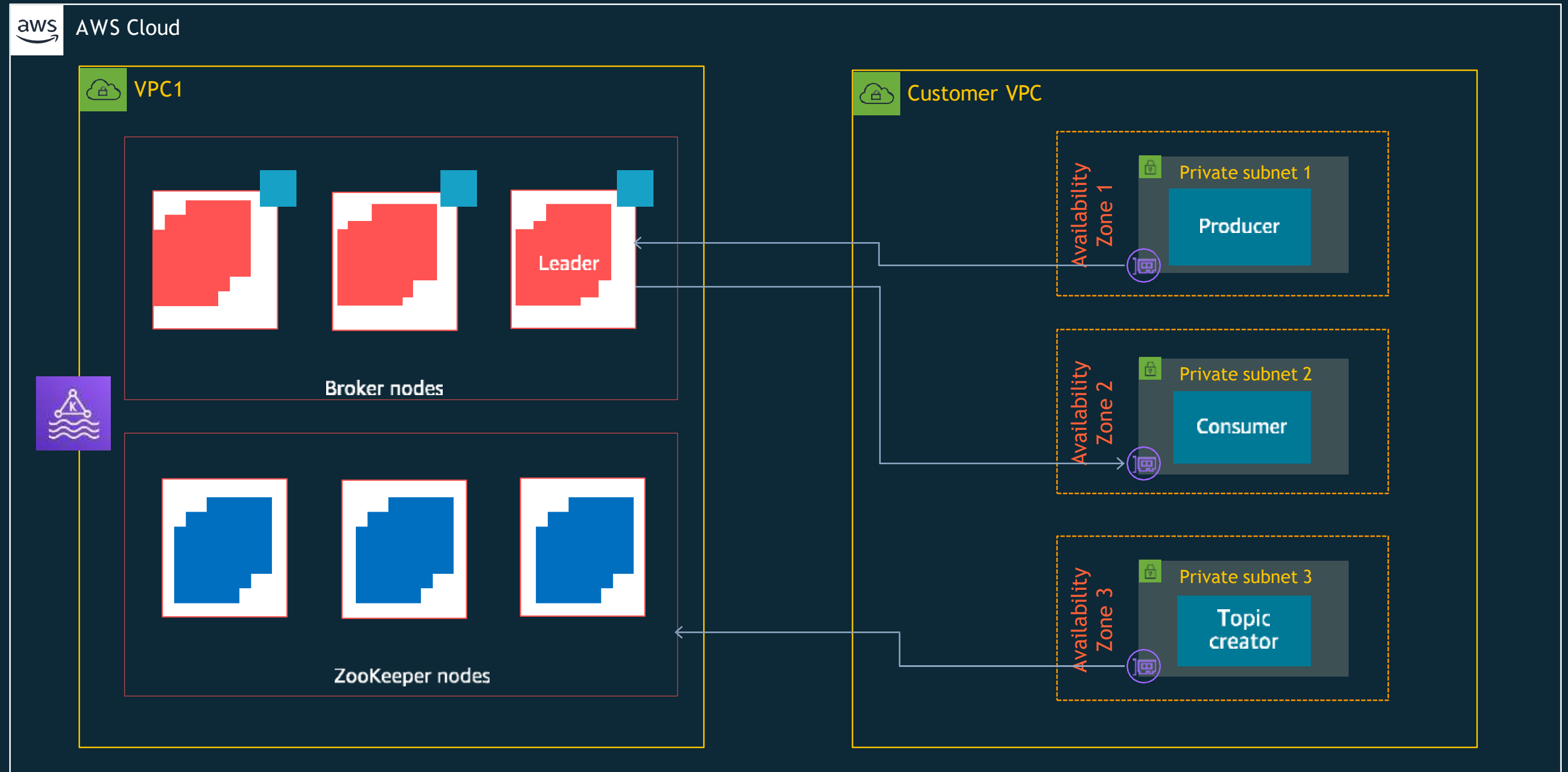
Expensive to maintain





Amazon MSK: A fully managed,
highly available, and secure service
for Apache Kafka

How connectivity works



What Amazon MSK does for you



Makes Apache Kafka more accessible to your organization



Drives best practices through design, defaults, and automation



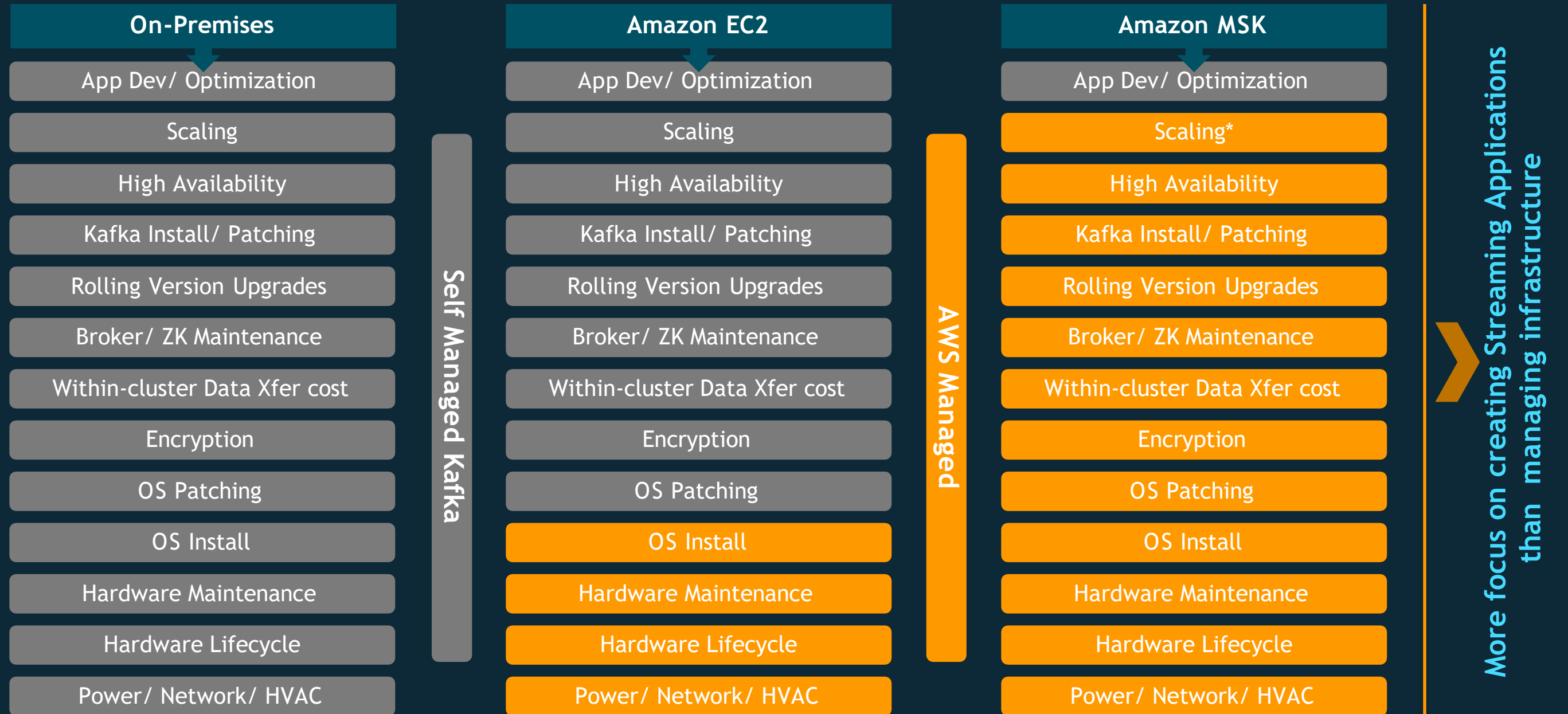
Allows developers to focus more on application development and less on infrastructure management



Amazon MSK is committed to improving open-source Apache Kafka



Fully Managed



* Cluster expansion, auto scaling storage and vertical scaling



Scalability

Scale out clusters



Scale out existing clusters by adding additional brokers without affecting cluster availability

Auto scale storage



Scale up broker attached EBS storage without affecting broker availability

Vertical Scaling ^{New}



Change the size or family of your brokers without reassigning Apache Kafka partitions



How we handle encryption

Encryption at rest



AWS KMS

Amazon MSK uses Amazon EBS server-side encryption and AWS KMS keys to encrypt storage volumes

If customers do not use self managed keys, the Amazon MSK service uses a default CMK

Encryption in transit



TLS

TLS Encryption in transit enabled by default

Certificate-based mutual TLS authentication using AWS Certificate Manager Private Certificate Authority (AWS PCA)



Deep AWS service integrations



Amazon Glue Schema Registry for schema evolution



Amazon Kinesis Data Analytics for Flink for complex stream processing



Amazon Data Migration Service for change data capture use cases



AWS Lambda with MSK as an event source



AWS Certificate Manager for Private CAs used for client TLS authentication



Amazon IAM for authentication of cluster APIs



Amazon KMS for storage volume encryption



AWS CloudFormation for describing/provisioning Amazon MSK clusters using code



Amazon CloudWatch for metrics



Monitoring MSK



CloudWatch Metrics

You can set three levels of monitoring with in CloudWatch for MSK, Default (at no cost to you), PER_BROKER and PER_TOPIC_PER_BROKER level.



Open Monitoring with Prometheus

You can enable open monitoring with Prometheus and expand your monitoring capability to third party compatible tools such as Datadog, Lenses, New Relic and Sumo Logic

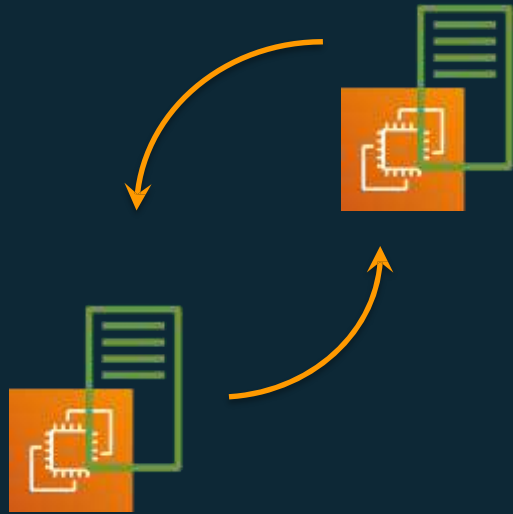


Broker Logs to CW logs, S3 and AES

Continuously stream Apache Kafka broker logs to Amazon CloudWatch Logs, Amazon S3, or Amazon Elasticsearch Service via Amazon Kinesis Data Firehose



Rolling Version Upgrades



Upgrade your Kafka versions
without any downtime for
your clusters*

* Deployment follows Apache Kafka best practices for replication



AWS Glue Schema Registry



Centrally discover, control, and evolve your data schemas



- *Improve data quality for your data streaming applications*
- *Enforce schemas and schema evolution to prevent downstream application failures*
- *Easily integrates with Amazon Managed Streaming for Apache Kafka (MSK), Amazon Kinesis Data Streams, and Amazon Kinesis Data Analytics for Apache Flink for convenient setup*
- *Use provided open source libraries to compress data and save on storage and data transfer costs*



Amazon MSK key features



Fully compatible with Apache Kafka toolsets, frameworks and utilities as long as they are using standard Kafka libraries and APIs



SASL/ SCRAM authentication with user names and passwords secured by AWS Secrets Manager



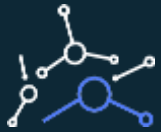
Cruise Control for partition management



Consumer Lag metrics by default for new clusters



Amazon MSK key features cont.



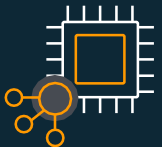
Define custom cluster configurations



Choice of 2 AZ or 3 AZ deployments



HIPAA, PCI, ISO and SOC (1,2,3) compliant



AWS CloudFormation and Terraform support



Zookeeper is automatically provisioned and included with each cluster at no additional cost



Lowest TCO in the Industry



Up to 60% lower cost compared to self managed Kafka

On-demand, hourly pricing is prorated to the second

Broker and storage pricing

- Broker pricing starts with kafka.m5.large at \$0.21 per hour
- Storage pricing is \$0.10 per GB-month

Data transfer from replication within the cluster and ZooKeeper nodes are included at no additional cost

Now, you can create Amazon MSK clusters with T3 brokers for less than \$2.50/day!!!



Sizing guidance in minutes

<https://bit.ly/2m3hlwz>

Here is a link to it in the Amazon MSK Best Practices section in the user documentation

<https://docs.aws.amazon.com/msk/latest/developerguide/bestpractices.html>

https://www.youtube.com/watch?v=4C_FT2le9E4



Up to 40% lower cost
compared to self managed
Kafka

The calculator gets you in the ballpark for most workloads

Costs are normalized to \$/GB ingested to make pricing comparisons easy

*Your latency requirements will impact how much you should provision,
adjust accordingly*

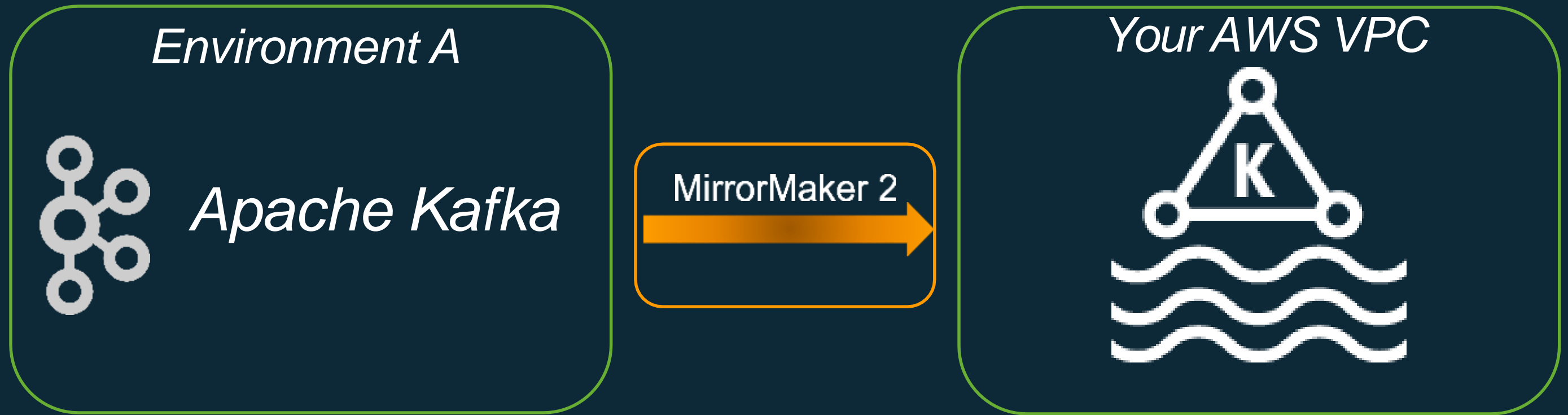
We still recommend that you test

*You can find a link to the calculator in the best practices section of the
Amazon MSK user documentation*



Easy lift and shift migration

Amazon MSK runs and manages Apache Kafka that makes it easy for you to migrate and run your existing Apache Kafka applications on AWS **without changes to the application code**



<https://amazonmsk-labs.workshop.aws/en/migration.html>



Best Practices

Monitor disk usage

Monitor the `KafkaDataLogsDiskUsed` metric in CloudWatch and set an alarm at 85% utilization

Adjust data retention parameters

Fine tune retention time period or retention log size at both the cluster and topic levels

Confine Zookeeper usage to Amazon MSK

Do not add non-MSK brokers or app config related metadata to MSK provisioned Zookeeper

Enable In-Transit encryption

For secure exchange of data between your clients and MSK

Partitions

Leverage a partition key strategy that balances load and supports higher throughput



Kafka Migration Program

Packaged best practices, expertise, hands-on technical assets, and migration incentives to accelerate your MSK migration



Free 2 day deep dive workshop

Get familiar with MSK, understand how MSK will fit in to your architecture and get hands-on migration guidance



Engage with our professional services experts

Get a jumpstart on your migration with a 'Launch' (paid) engagement with our experts from professional services



Ready to use technical assets

Get hands-on access to migration labs that detail best practices around configuring and migrating using MM2



Migration Incentives

If needed, get POC credits to get the ball rolling on your migration

* Work with your account teams to check if you [qualify](#)



Getting Started

Learn MSK using our new education classes

<https://aws.amazon.com/msk/resources/>

Get hands on using our additional Labs

<https://amazonmsk-labs.workshop.aws/en/>

Get started in minutes with our ready to use templates

<https://aws.amazon.com/solutions/implementations/aws-streaming-data-solution-for-amazon-msk/>

Watch how compatible toolsets work with MSK

https://www.youtube.com/watch?v=4C_FT2le9E4&t=6s

Need a deep dive?

Contact us at msk-bd@amazon.com

