
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

ARC314

Publishing real-time financial data feeds using Kafka

Rana Dutt

Principal Solutions Architect
AWS

Diego Soares

Senior Solutions Architect
AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Types of financial data streams

Who is using Kafka today

Kafka and Amazon Managed Streaming for Apache Kafka (Amazon MSK)

How to use Amazon MSK to publish a real-time data feed

Security model

Operations

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Types of financial data streams

- Asset prices
  - Stocks and ETFs
  - Options and commodities
  - Crypto
- Corporate actions and news
- Analytics



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Use cases

- Trading
- Surveillance and fraud detection
- Risk assessment
- Post-trade analytics and backtesting



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Latency considerations

We define the required end-to-end latency by transmission, networking, and computational delay



High-touch

Example:
retail trading



Built for the cloud

Examples:
crypto exchange
and auctions



Low latency

Examples:
hybrid architecture,
edges at venues,
single markets



Ultra-low latency

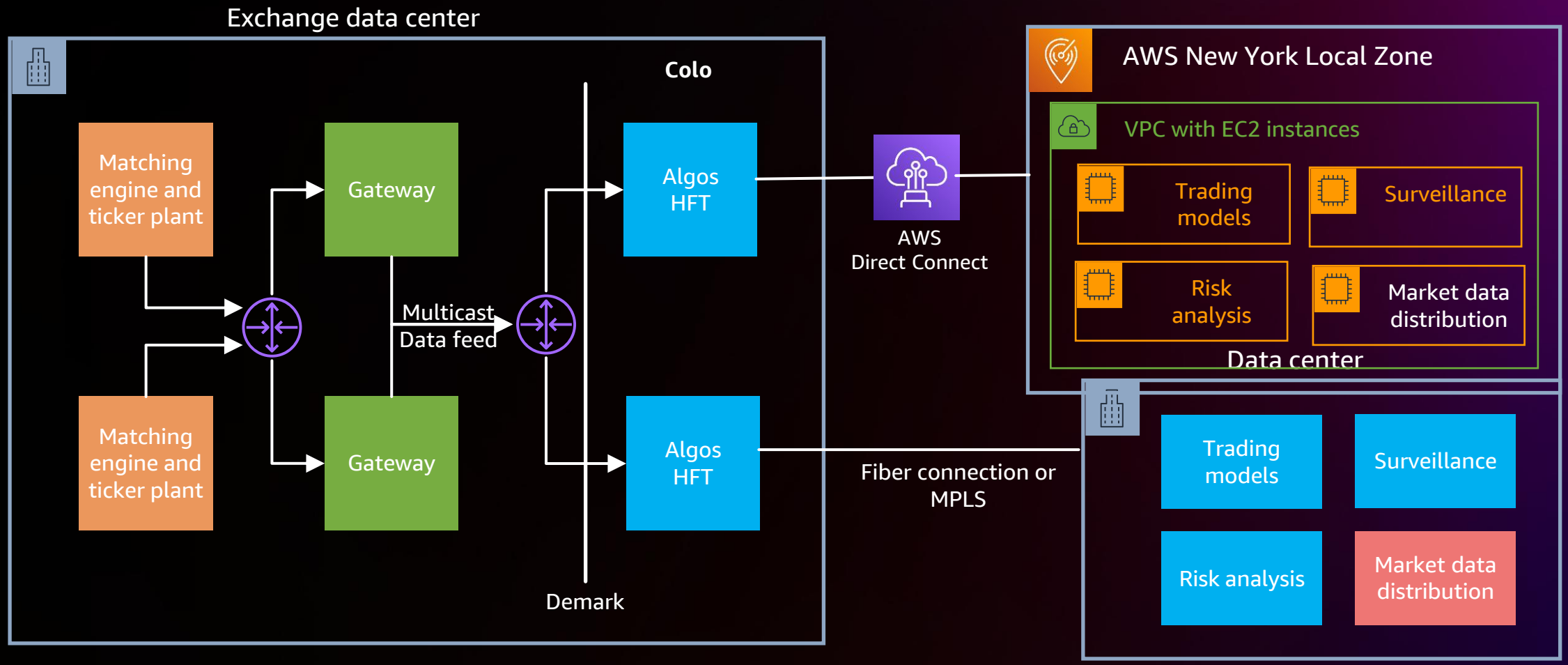
Examples:
hybrid architecture,
collocation replacement,
and rearchitected
applications

Widespread adoption today → Gaining traction

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Low-latency trading – AWS Local Zones

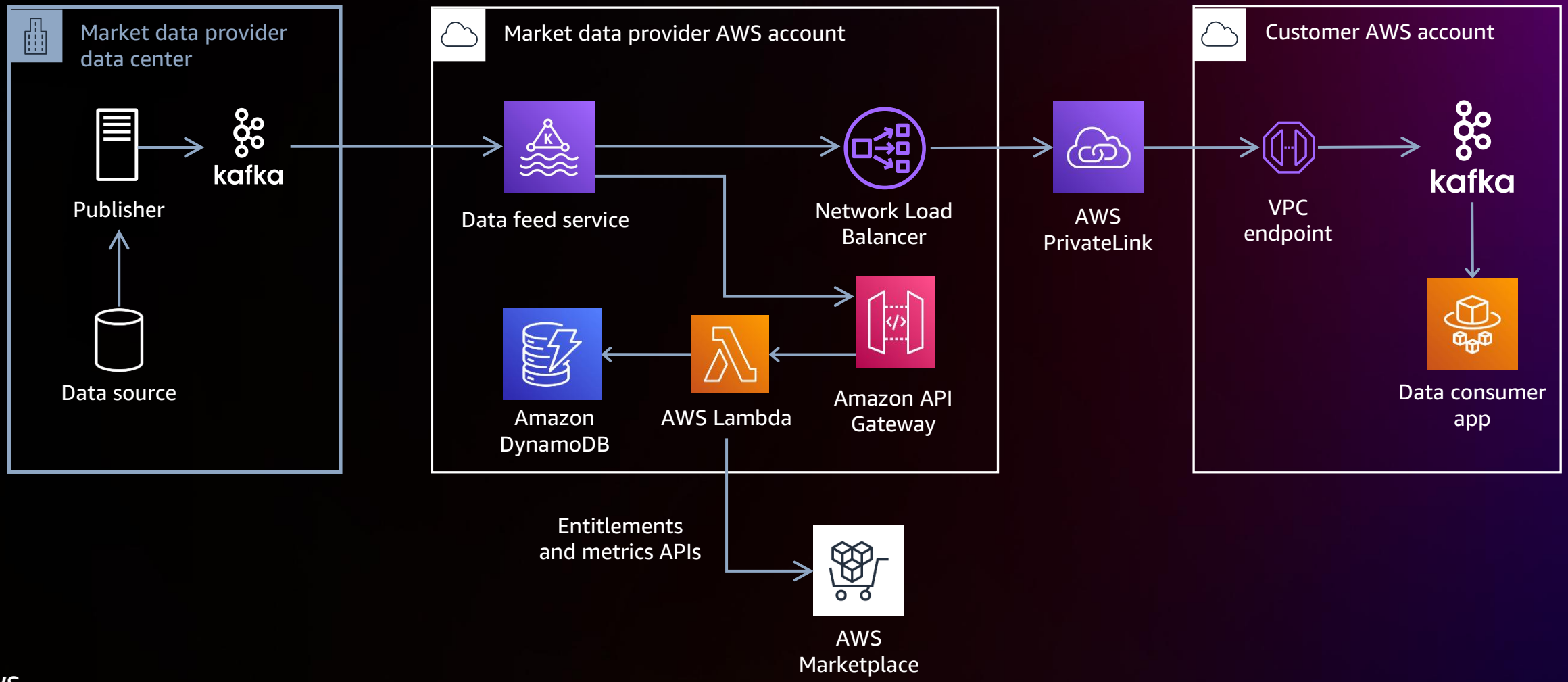
MIGRATE FROM DATA CENTER TO AWS NEW YORK LOCAL ZONE



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Publish real-time market data on AWS

AWS MARKETPLACE



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Real-time financial data providers on AWS



Nasdaq NCDS



REFINITIV real-time data

Bloomberg

Bloomberg B-PIPE



Cboe Global Cloud

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NCDS on AWS delivers real-time market data seamlessly

"Nasdaq Cloud Data Service is a significant advancement in the financial data space, as it uses the cloud to stream important, real-time market data tailored to our clients' specific needs. Adding the cloud to the data equation through our collaboration with AWS is a big win for investors."

Lauren Dillard, EVP and Head of Global Information Services, Nasdaq

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Apache Kafka



Open-source distributed event streaming platform

Low latency and high throughput

Offers pub/sub and fans out to any number of consumers

Enables use cases such as real-time messaging, event-stream processing, and more

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is Amazon MSK?



Amazon Managed Streaming for Apache Kafka (Amazon MSK)

Securely stream data with a fully managed, highly available Apache Kafka service

- Automate provisioning, configuring, and tuning
- Fully compatible with open-source Apache Kafka
- Highly secure
- Lower cost

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Fully managed

On premises	Amazon EC2	Amazon MSK
App Dev/Optimization	App Dev/Optimization	App Dev/Optimization
Scaling	Scaling	Scaling*
High Availability	High Availability	High Availability
Kafka Install/ Patching	Kafka Install Patching	Kafka Install/Patching
Rolling Version Upgrades	Rolling Version Upgrades	Rolling Version Upgrades
Broker/ZK Maintenance	Broker/ZK Maintenance	Broker/ZK Maintenance
Within-cluster Data Xfer Cost	Within-cluster Data Xfer Cost	Within-cluster Data Xfer Cost
Encryption	Encryption	Encryption
OS Patching	OS Patching	OS Patching
OS Install	OS Install	OS Install
Hardware Maintenance	Hardware Maintenance	Hardware Maintenance
Hardware Lifecycle	Hardware Lifecycle	Hardware Lifecycle
Power/Network/ HVAC	Power/Network/HVAC	Power/Network/HVAC

Self-Managed Kafka

AWS Managed

More focus on creating streaming applications than managing infrastructure

* Cluster expansion and auto scaling storage

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon MSK Serverless



No servers to manage



Fully compatible



On-demand streaming capacity



Same security as Amazon MSK

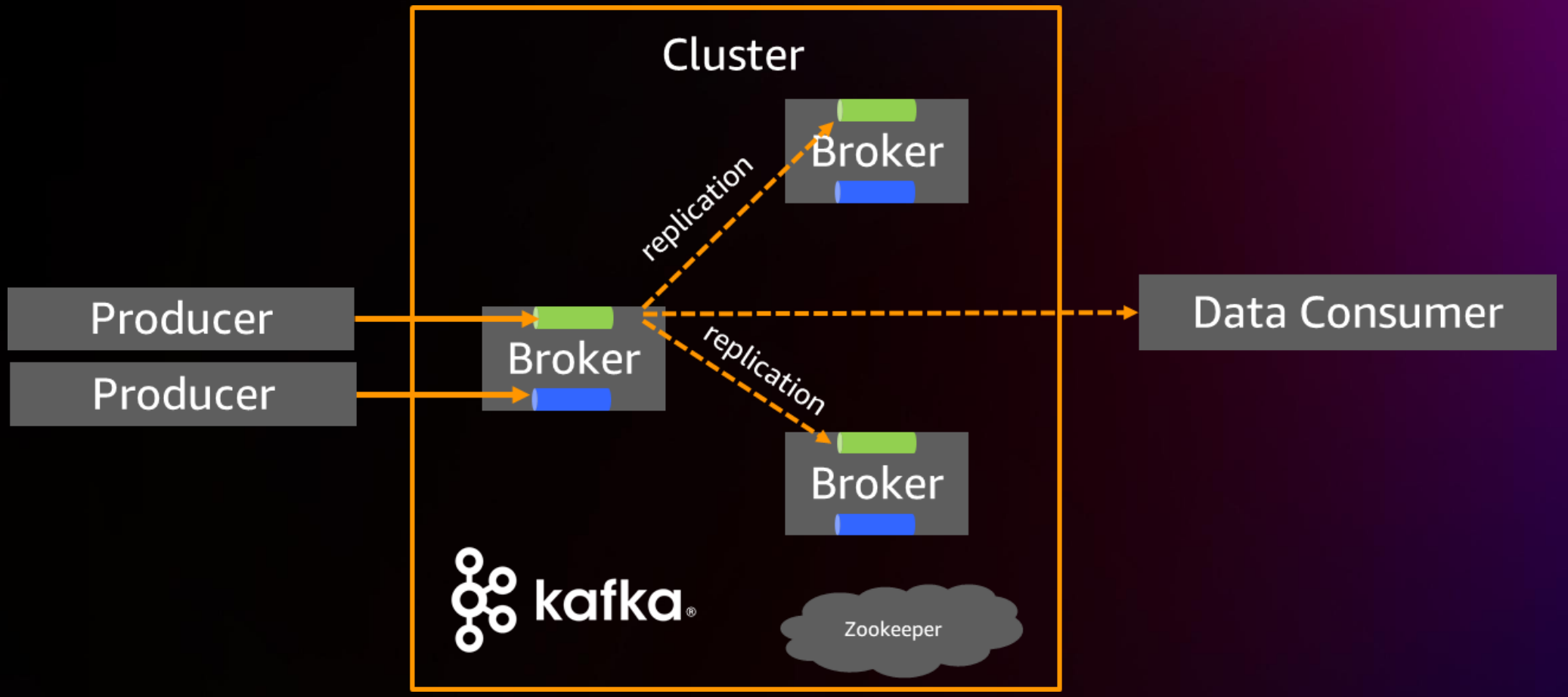


Pay-for-throughput pricing

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Apache Kafka review

CLUSTER ANATOMY



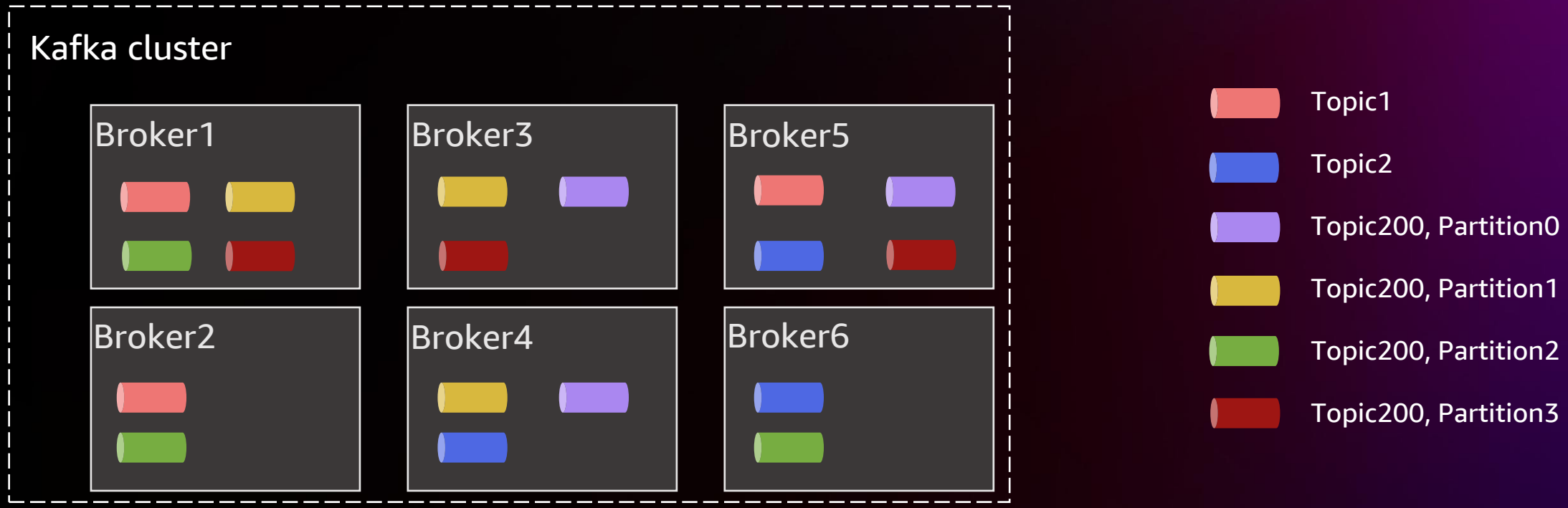
Recommendations:

- default.replication.factor=3
- min.insync.replicas=2

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Apache Kafka review

TOPIC ANATOMY

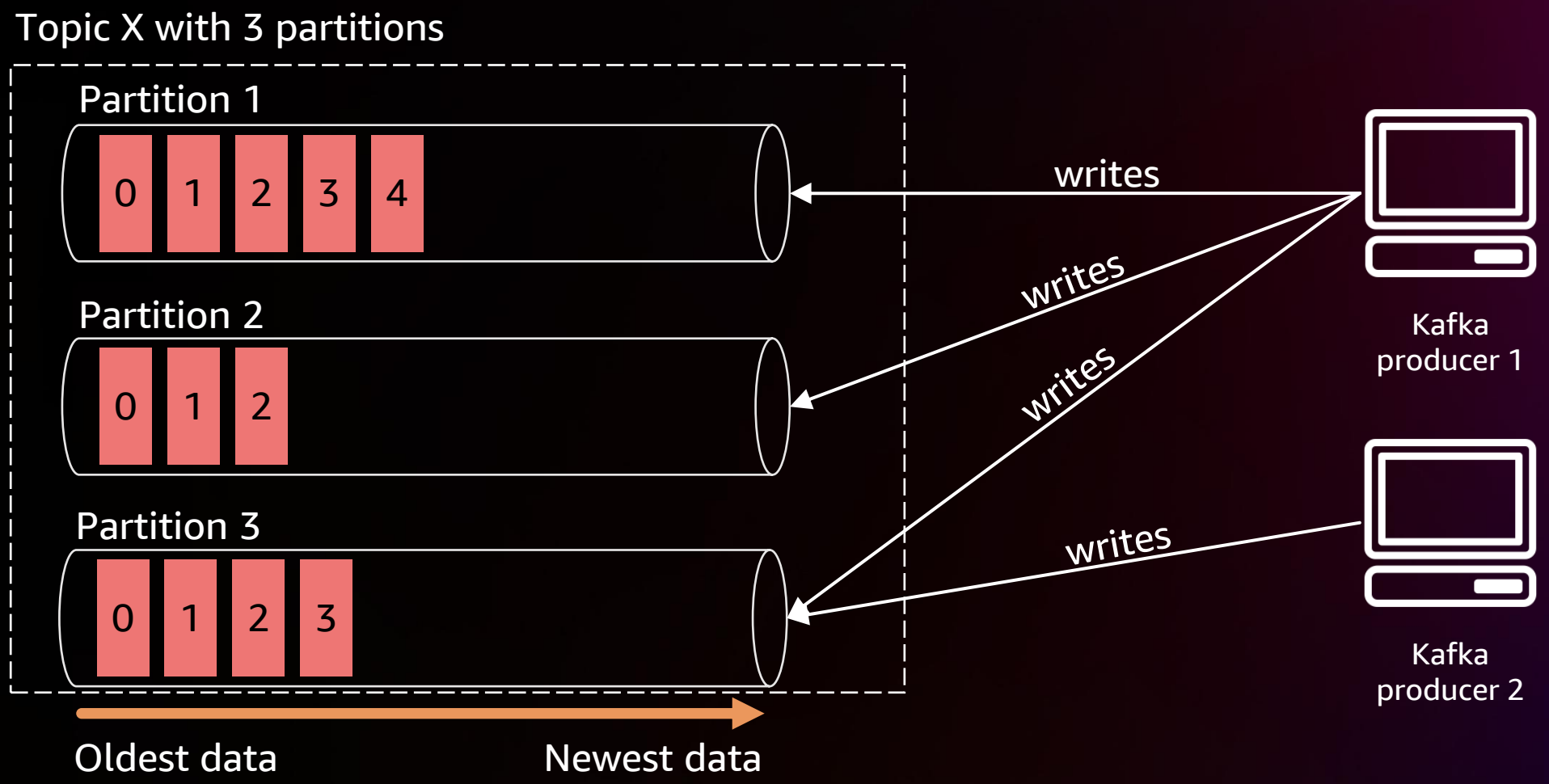



```
AdminHost@ kafka$ bin/kafka-topics.sh --bootstrap-server $BROKERS --create --replication-factor
3 --partitions 4 --topic topic200
AdminHost@ kafka$ bin/kafka-topics.sh --bootstrap-server $BROKERS --describe --topic topic200
Topic: topic200  PartitionCount: 4ReplicationFactor: 3     Configs: min.insync.replicas=2
        Topic: topic200  Partition: 0    Leader: 5       Replicas: 3,5,4  Isr: 3,5,4
        Topic: topic200  Partition: 1    Leader: 1       Replicas: 3,1,4  Isr: 3,1,4
        Topic: topic200  Partition: 2    Leader: 2       Replicas: 1,2,6  Isr: 1,2,6
        Topic: topic200  Partition: 3    Leader: 3       Replicas: 5,3,1  Isr: 5,3,1
```

aws

Apache Kafka review

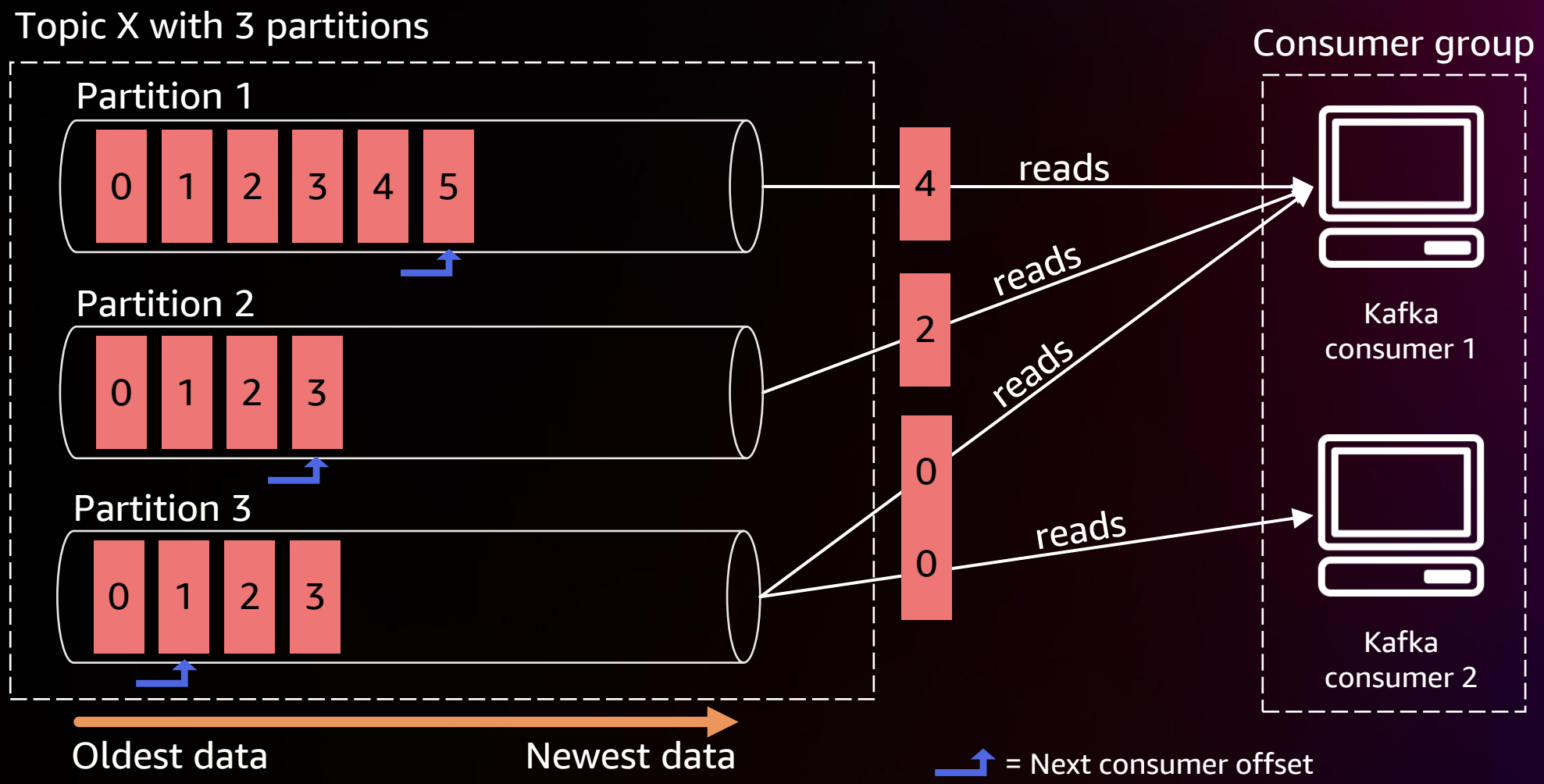
WRITES TO PARTITIONS



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Apache Kafka review

READS FROM PARTITIONS



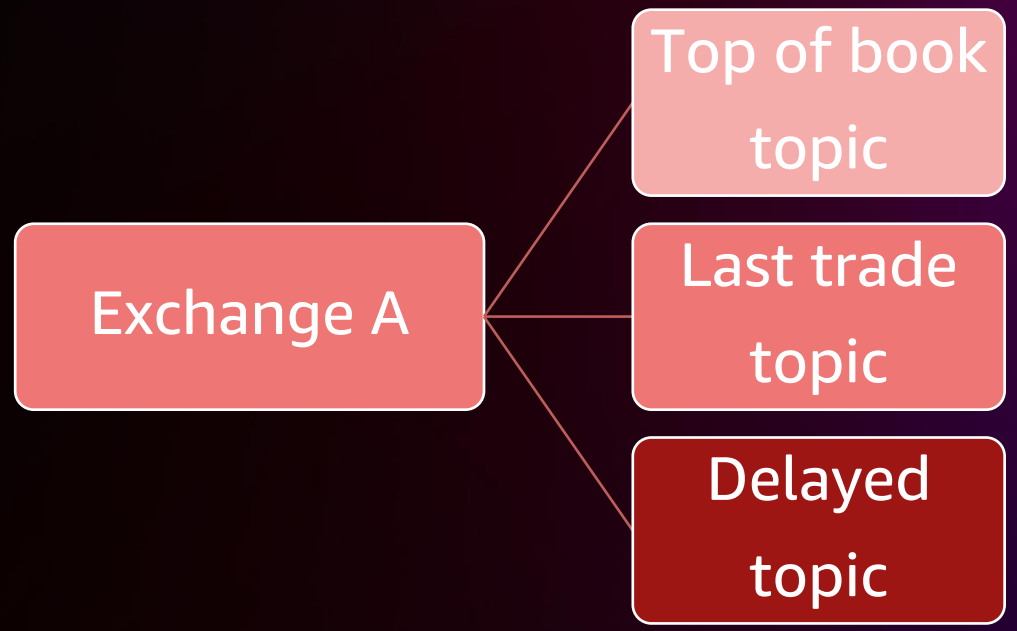
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Real-world example

SECURITIES EXAMPLE

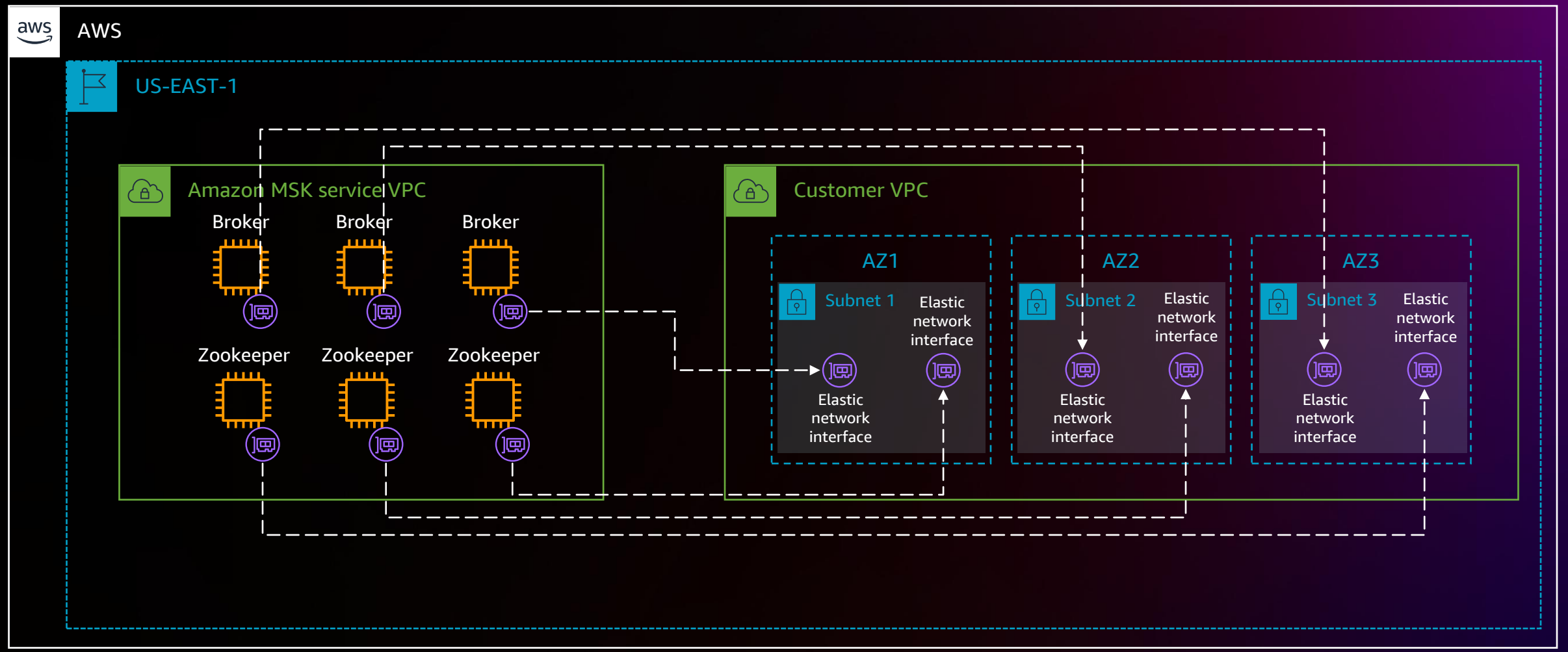
Scenario

- An exchange with 3 feeds
  - Top of book
  - Last trade
  - Delayed 15 minutes
- Asset class is securities



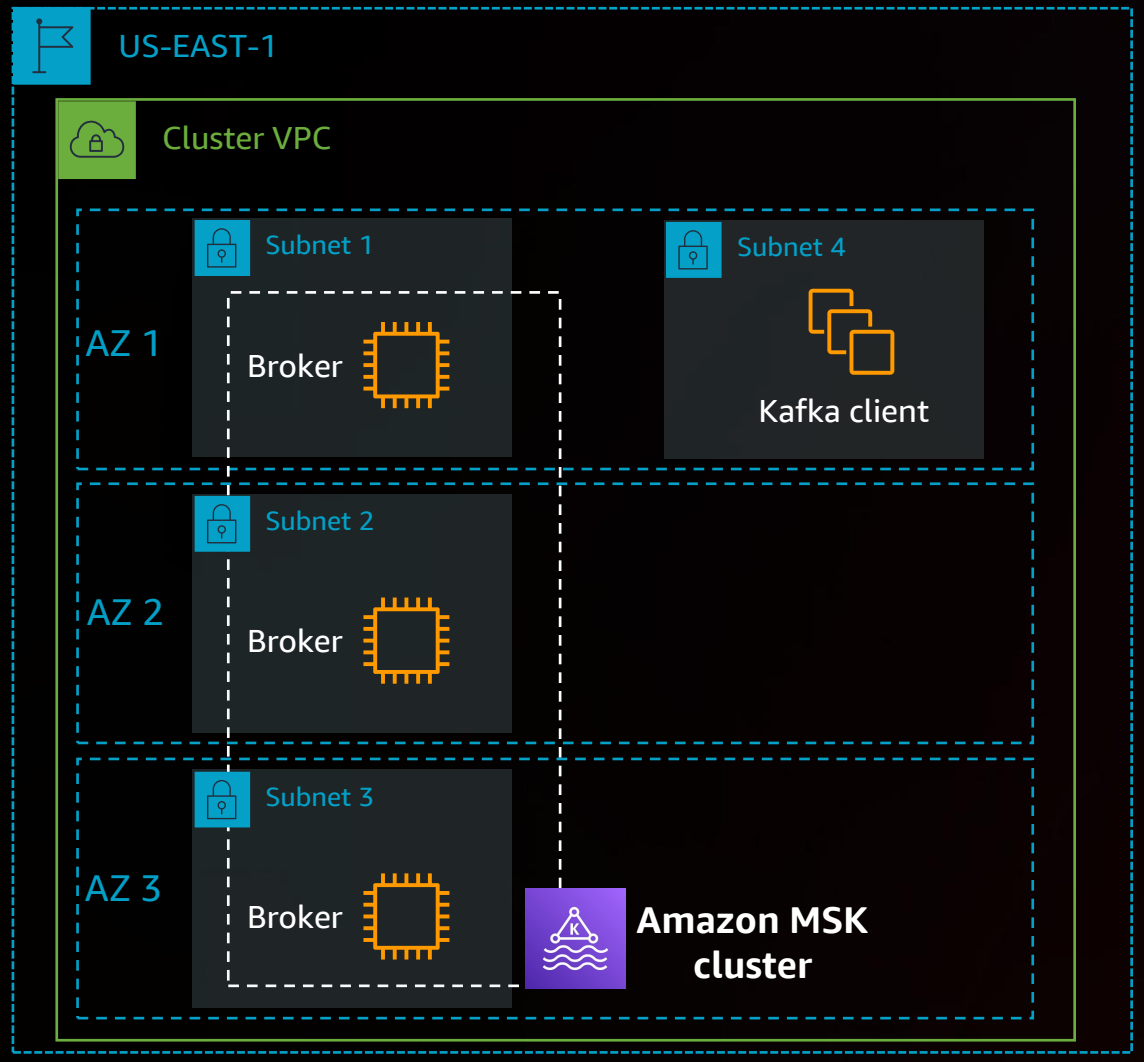
aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon MSK cluster architecture



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon MSK cluster architecture

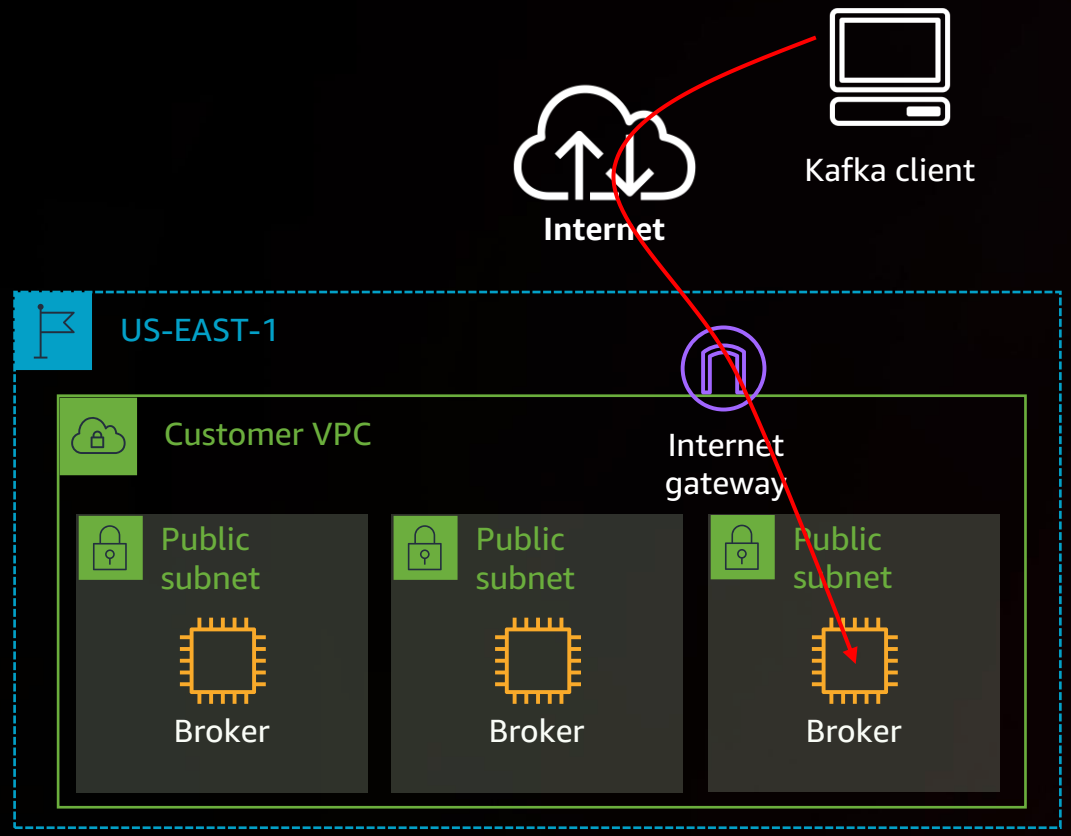


- Brokers reside in the service VPC
- Connectivity to cluster happens via network interfaces in the cluster VPC
- Connectivity control enforced via security groups
- Clients connect to
  x.y.z.kafka.<cluster region>.amazonaws.com
- Amazon MSK supports native Kafka tools

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon MSK cluster connectivity – Public

DIRECT PUBLIC ACCESS



Private endpoint:
b-1.<cluster name>.y.z.kafka.<region>.amazonaws.com:9094
Public endpoint:
b-1-public.<cluster name>.y.z.kafka.<region>.amazonaws.com:9194

Available since November 2021

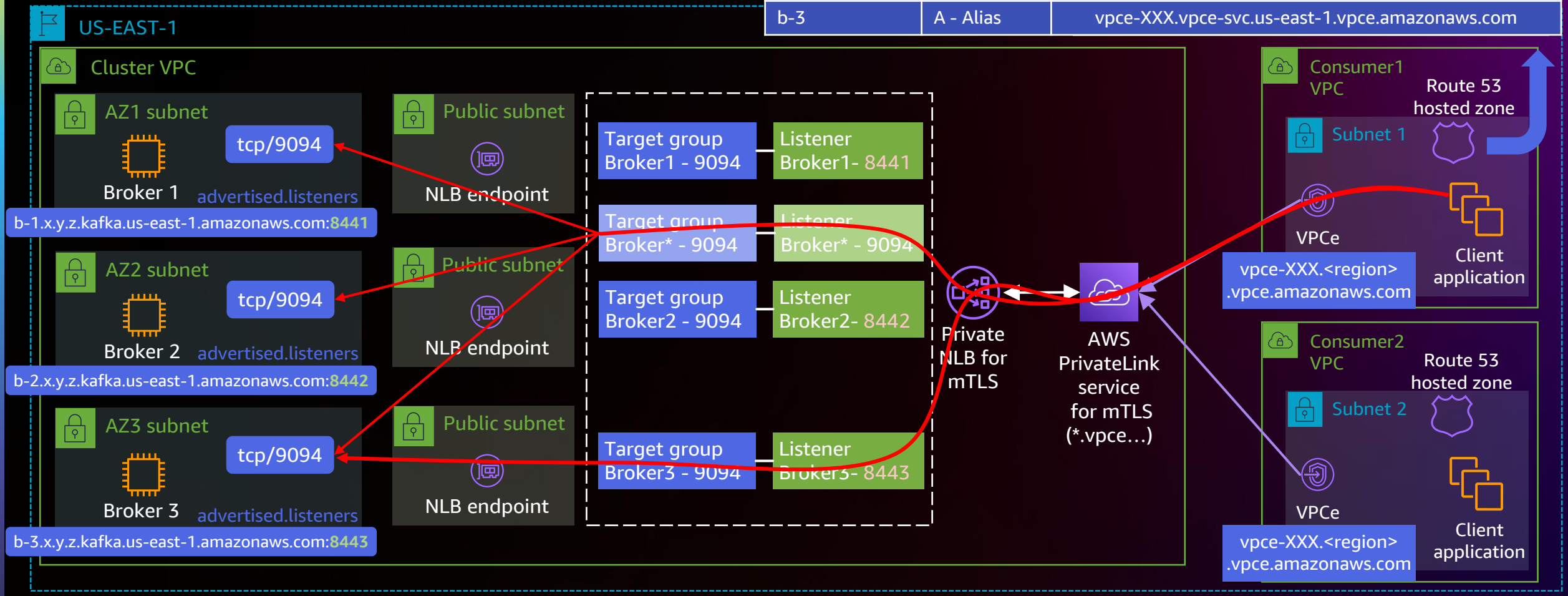
- Benefits
  - No need for additional services
  - Lower cost
  - Highly scalable
- Requirements
  - Requires public subnets
  - Disable unauthenticated connections
  - Disable PLAINTEXT(non-tls) traffic
  - Encryption must be turned on
  - Kafka version >= 2.6.0
  - Enable Kafka ACLs
    allow.everyone.if.no.acl.found=false

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon MSK cluster connectivity – Private

CUSTOMER-DEPLOYED PRIVATELINK

Amazon Route 53 hosted zone domain name: x.y.z.kafka.us-east-1.amazonaws.com		
b-1	A - Alias	vpce-XXX.vpce-svc.us-east-1.vpce.amazonaws.com
b-2	A - Alias	vpce-XXX.vpce-svc.us-east-1.vpce.amazonaws.com
b-3	A - Alias	vpce-XXX.vpce-svc.us-east-1.vpce.amazonaws.com

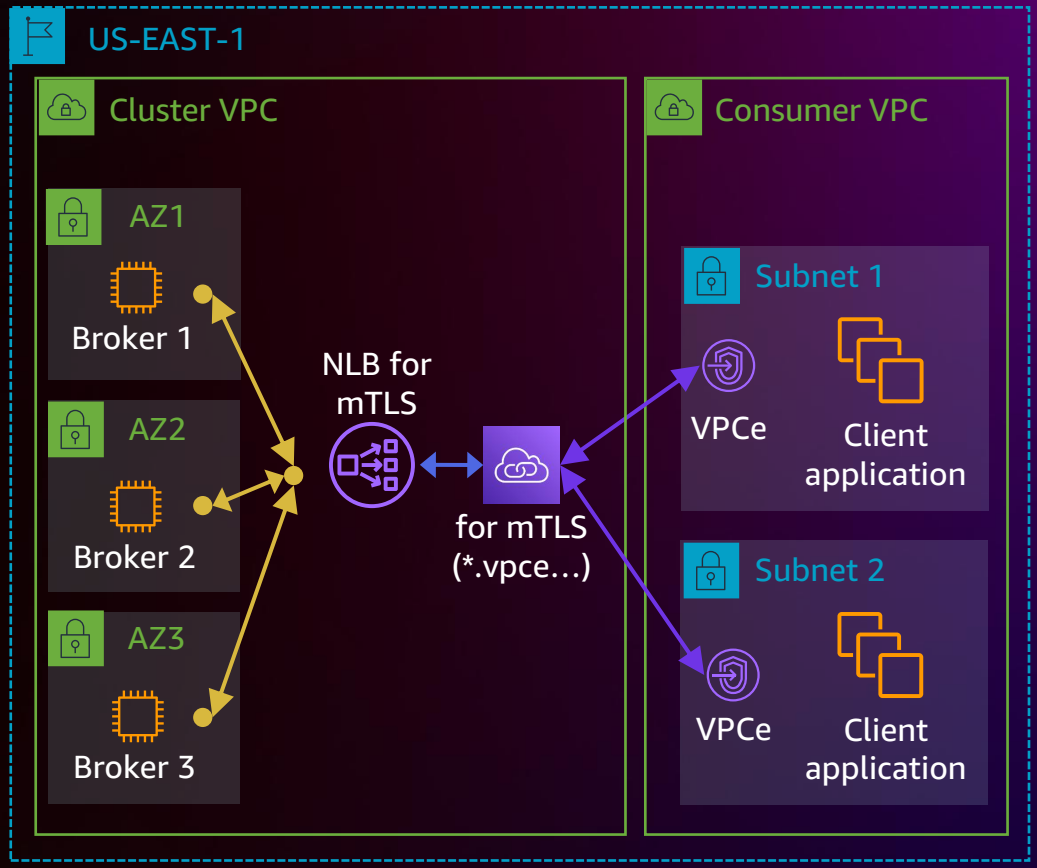


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon MSK cluster connectivity – Private

CUSTOMER-DEPLOYED PRIVATELINK

- Flexible architecture
- Additional steps required to scale the cluster (add Network Load Balancer listeners)
- Support all authentication methods
- Support non-TLS communication
- Cluster account pays for Network Load Balancer cost
- Requires Kafka configuration changes
  advertised.listeners
- Consumer account pays for PrivateLink cost

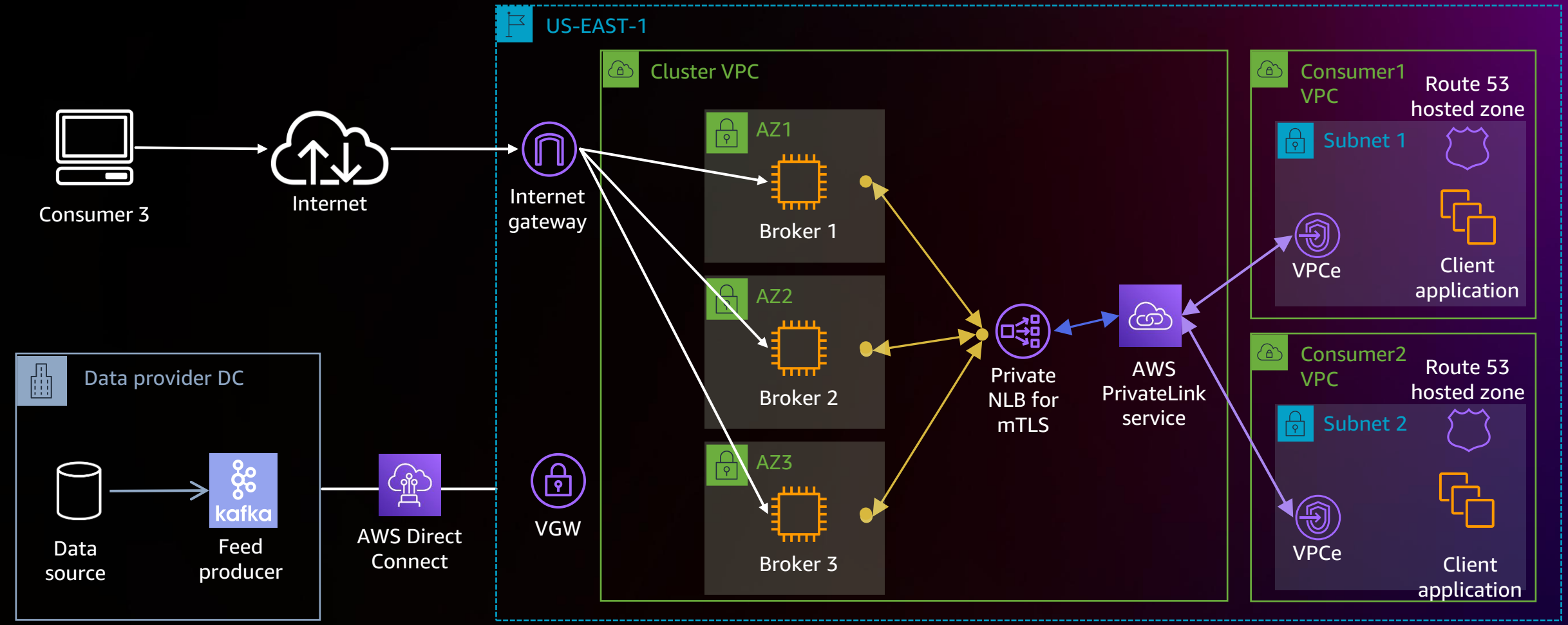


aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Putting all the pieces together

END-TO-END ARCHITECTURE

CDK code



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Kafka client

GETTING STARTED WITH PYTHON

Python code

Kafka consumer

```python
from kafka import KafkaConsumer
import json,os

#Import Environment Variables
tlsbrokers = os.environ.get('TLSBROKERS')

#Create the Kafka Consumer
consumer = KafkaConsumer(
      'topic1', #topic to consume
      group_id='consumer_python', #local consumer name
      bootstrap_servers=tlsbrokers, #Brokers List
      # For mTLS auth:
      security_protocol='SSL',
      ssl_check_hostname=True,
      #Client Cert Signed by PCA
      ssl_certfile='kafkacert/client_cert.pem',
      #Client Key used on cert above
      ssl_keyfile='kafkacert/private_key.pem',
      #Public Keychain with Public Amazon PKI
      ssl_cafile='kafkacert/truststore.pem' )

# Loop to consume messages and Print details.
for msg in consumer:
      #Print message metadata
      print ("%s:%d:%d" % (msg.topic, msg.partition,
msg.offset))
      #check if massage is in JSON format if not print
string
      try:
          print(json.loads(msg.value))
      except:
          print(msg.value.decode('utf-8'))
```

Kafka producer

```python
from kafka import KafkaProducer
import json, os

#Import Environment Variables
tlsbrokers = os.environ.get('TLSBROKERS')

#Create the Kafka Producer
producer = KafkaProducer(
      bootstrap_servers=tlsbrokers, #Brokers List
      # For mTLS auth:
      security_protocol='SSL',
      ssl_check_hostname=True,
      #Client Cert Signed by PCA
      ssl_certfile='kafkacert/client_cert.pem',
      #Client Key used on cert above
      ssl_keyfile='kafkacert/private_key.pem',
      #Public Keychain with Public Amazon PKI
      ssl_cafile='kafkacert/truststore.pem',
      #Serialization Method
      value_serializer=lambda v: json.dumps(v).encode('utf-
8'),
      #Number of ACKs to wait on 0=None, 1=Leader, All= All
Brokers
      acks=(1))

#Message to send
msg = {"hello":"world"}

#Send message to Kafka Brokers
producer.send('topic1', value=msg)
producer.flush()
```

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon MSK security

PROTECTING YOUR CLUSTER



Network security



Encryption



Authentication



Authorization

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Authentication and authorization





- Kafka authentication methods
  - Unauthenticated
  - SASL/GSSAPI (Kerberos)
  - SASL/PLAIN
  - SASL/SCRAM
  - SASL/OAUTHBEARER
- Kafka authorization uses resource ACL

- Amazon MSK authentication methods
  - Unauthenticated
  - SASL/IAM
  - SASL/SCRAM
  - Mutual TLS
- Amazon MSK authorization uses resource ACL

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Security – Authentication

MUTUAL TLS AUTHENTICATION

TLS guide



Amazon
Public CA



AWS Private
Certificate Authority



Kafka
cluster



Kafka
client

1. Create a private certificate authority
   - Root CA or subordinate CA
2. Enable mTLS authentication on your cluster
3. Client certificates issued by AWS Private CA
4. Client needs to trust Amazon Public CA
5. Kafka ACLs can now be configured using the certificate distinguished name

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Security – Authorization

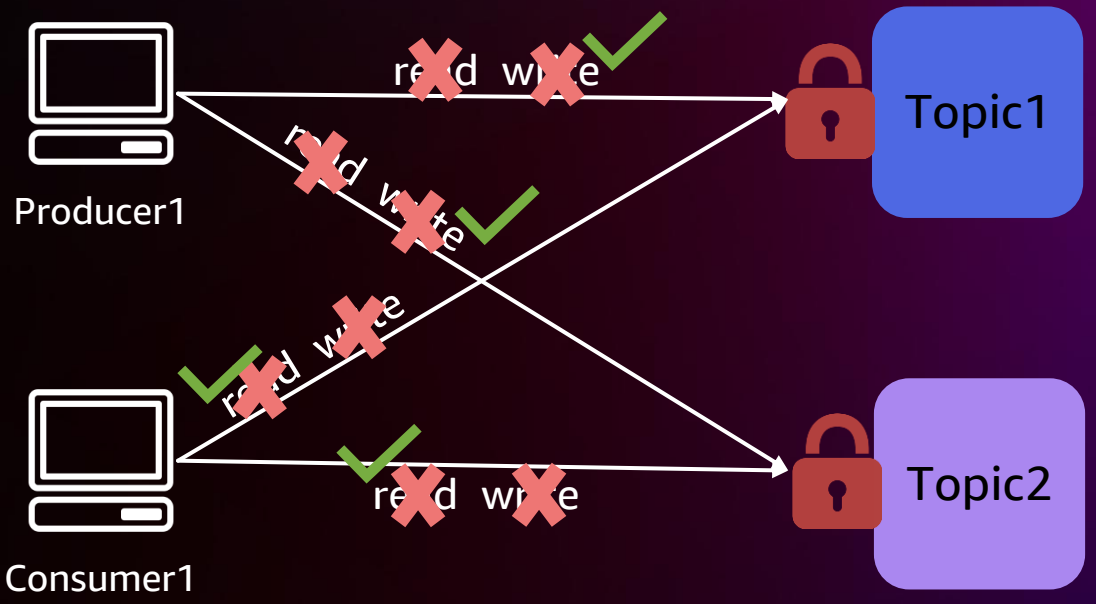
AUTHORIZATION WITH ACLS

By default Amazon MSK public access denies access to topics

allow.everyone.if.no.acl.found = false



Cluster admin




```
bin/kafka-acls.sh --authorizer-properties zookeeper.connect=$ZKNodes --add --allow-principal "User:CN=producer1" --operation Write --topic topic1
```

```
bin/kafka-acls.sh --authorizer-properties zookeeper.connect=$ZKNodes --add --allow-principal "User:CN=producer1" --operation Write --topic topic2
```

```
bin/kafka-acls.sh --authorizer-properties zookeeper.connect=$ZKNodes --add --allow-principal "User:CN=consumer1" --operation Read --group=* --topic topic1
```

```
bin/kafka-acls.sh --authorizer-properties zookeeper.connect=$ZKNodes --add --allow-principal "User:CN=consumer1" --operation Read --group=* --topic topic2
```

Principal	Permission	Operation	Resource	Host
User:CN=producer1	Allow	WRITE	TOPIC: topic1	-

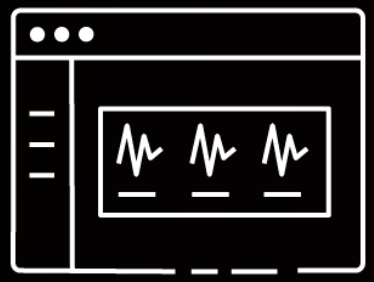
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon MSK monitoring

DEEP VISIBILITY

Monitoring best practices



Metrics

Amazon CloudWatch

Open-source Prometheus

4 monitoring levels

Amazon CloudWatch metrics for this cluster

Enhanced metrics are available at an additional cost. **Learn more**

- Basic monitoring
  Includes basic cluster-level and broker-level monitoring. Available for free.
- Enhanced broker-level monitoring
  Also includes basic monitoring. Available at an additional cost.
- Enhanced topic-level monitoring
  Also includes basic and enhanced broker-level monitoring. Available at an additional cost.
- Enhanced partition-level monitoring
  Also includes basic, enhanced broker-level monitoring and topic-level monitoring. Available at an additional cost.

Amazon CloudWatch

Amazon S3

Amazon Kinesis

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Marketplace

A curated digital catalog helping customers **find, subscribe to, deploy,** and **govern** third-party software, data, and services

Over **12,000** transactable listings

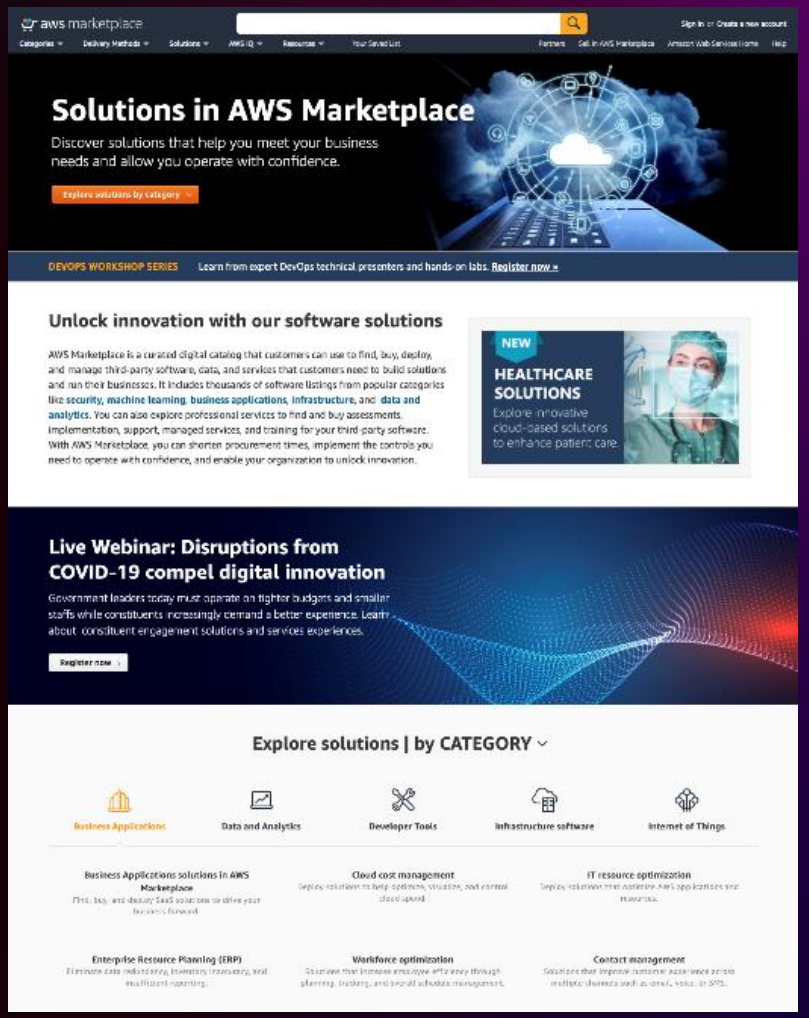
Deployed in **25** Regions

2,000+ ISVs, **250+** data providers, **1,000+** consulting partners

65+ categories (monitoring, security, databases, etc.)

2M+ active subscriptions

325,000+ active subscribers

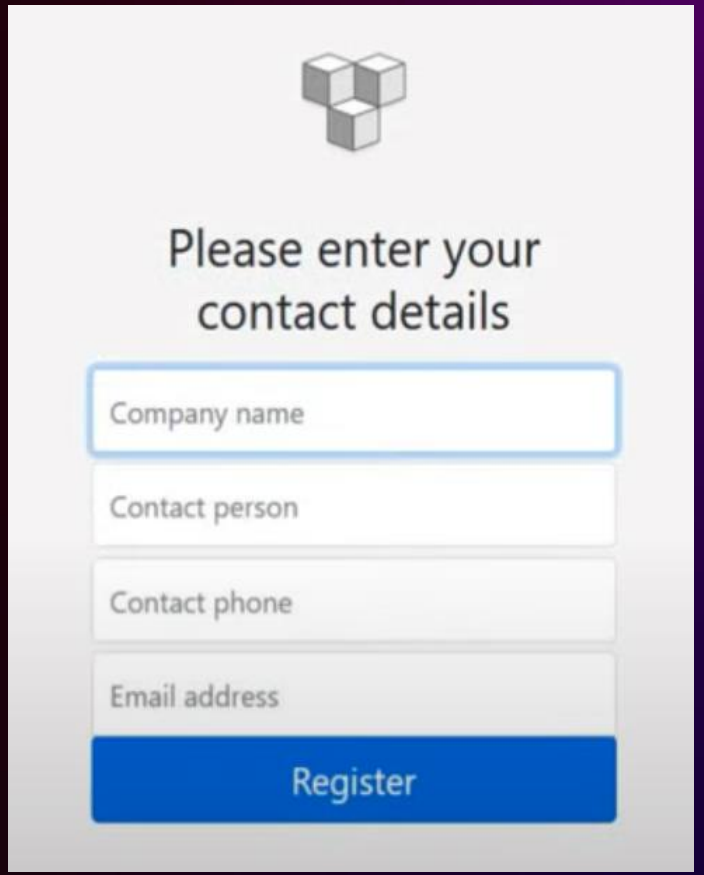


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Quick Start for AWS Marketplace

- Deploys AWS Marketplace SaaS integration
- Accepts new client registrations
- Handles customer entitlements and usage metering
- Clients are auto-redirected to registration app URL

https://go.aws/3Up2Tu1

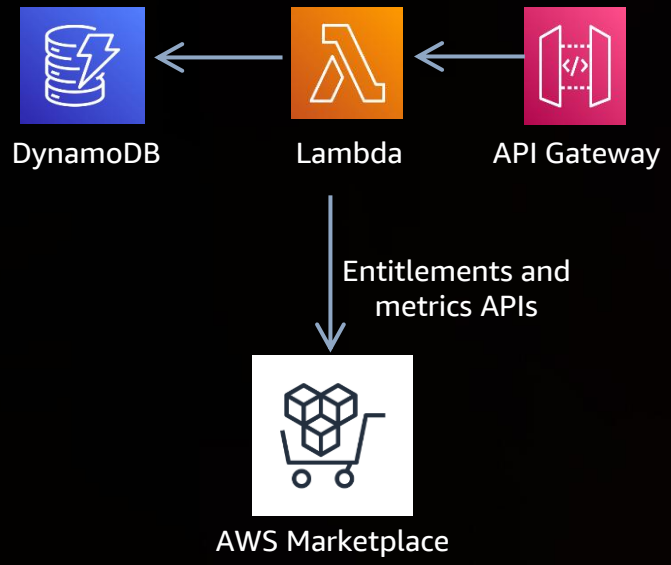


aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Steps for client registration

1

Deploy registration app using SaaS Quick Start for AWS Marketplace



2

Generate client certificate on AWS Private CA
Enable client on Amazon MSK ACL



3

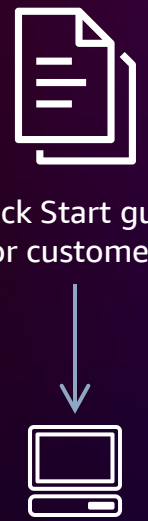
Store cert in AWS Secrets Manager
Store client info in DynamoDB





4

Client downloads Quick Start guide



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Summary

- Learned about financial data streams
- Learned about who is using Kafka today in the cloud
- Learned about Kafka and Amazon MSK
- Learned how to use Amazon MSK to publish a real-time data feed
- Learned best practices to deploy a real-time financial data feed using Kafka

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Additional resources



Amazon MSK best practices documentation

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

Best practices for right-sizing your Apache Kafka clusters blog

https://aws.amazon.com/blogs/big-data/best-practices-for-right-sizing-your-apache-kafka-clusters-to-optimize-performance-and-cost/

Github repository on AWS Samples (with CloudFormation templates, deployment reference architectures, Python clients, Cloudwatch dashboards, Grafana dashboards, and more)
https://github.com/aws-samples/msk-powered-financial-data-feed

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Rana Dutt
/ranadutt

Diego Soares
/diegogsoares



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.