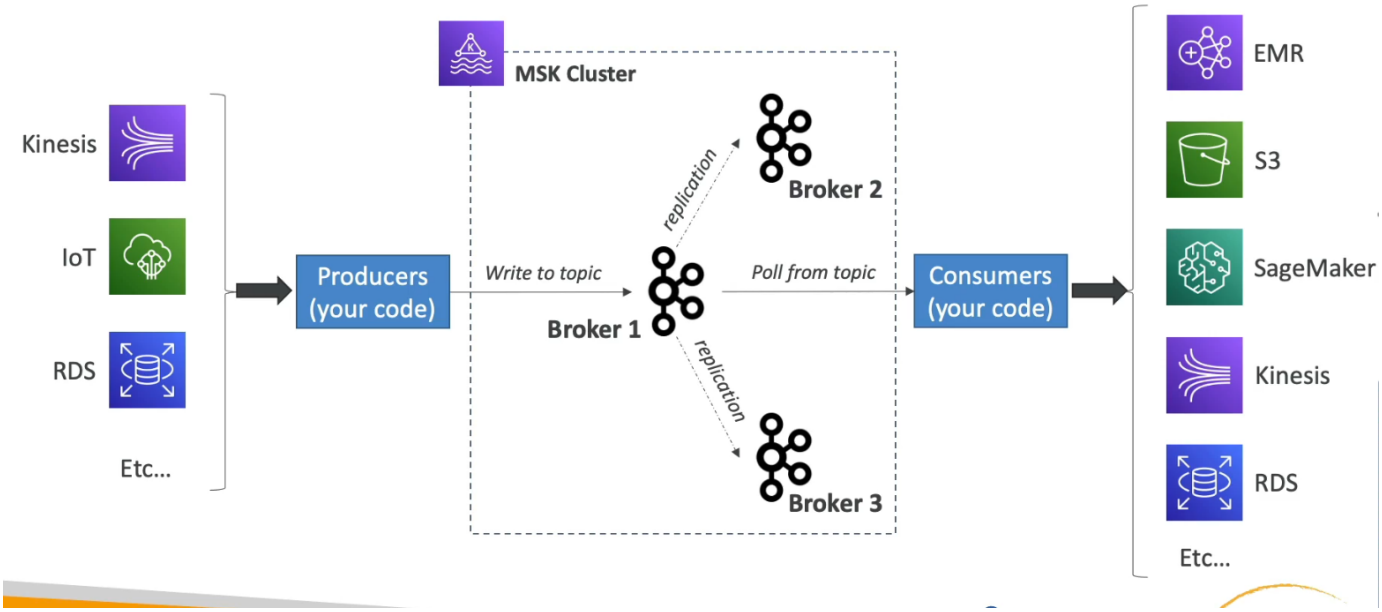
amazon msk

Amazon MSK (Managed Streaming for Apache Kafka) is a fully managed service provided by Amazon Web Services (AWS) that enables you to build and run applications using Apache Kafka without the need to manage the underlying infrastructure. It simplifies the deployment, configuration, and management of Apache Kafka clusters, allowing you to focus on building your applications

Amazon Managed Streaming for Apache Kafka (Amazon MSK)

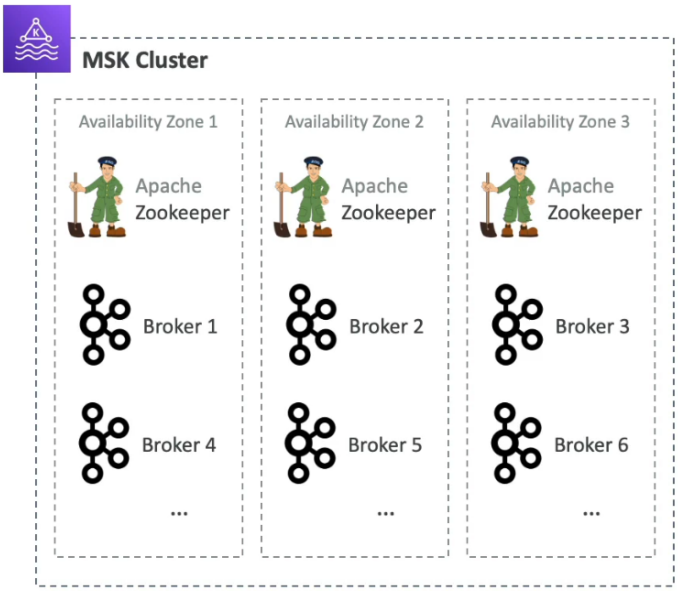
- Alternative to Kinesis (Kafka vs Kinesis next lecture)
- Fully managed Apache Kafka on AWS
  - Allow you to create, update, delete clusters
  - MSK creates & manages Kafka brokers nodes & Zookeeper nodes for you
  - Deploy the MSK cluster in your VPC, multi-AZ (up to 3 for HA)
  - Automatic recovery from common Apache Kafka failures
  - Data is stored on EBS volumes
- You can build producers and consumers of data
- Can create custom configurations for your clusters
  - Default message size of 1MB
  - Possibilities of sending large messages (ex: 10MB) into Kafka after custom configuration

Apache Kafka at a high level



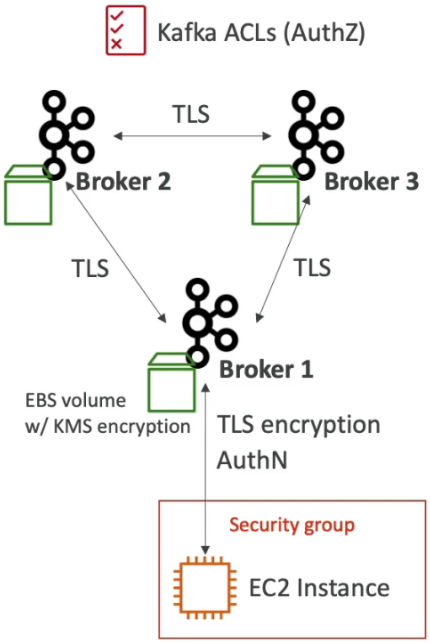
MSK – Configurations

- Choose the number of AZ (3 – recommended, or 2)
- Choose the VPC & Subnets
- The broker instance type (ex: kafka.m5.large)
- The number of brokers per AZ (can add brokers later)
- Size of your EBS volumes (1GB – 16TB)



MSK – Security

- Encryption:
  - Optional in-flight using TLS between the brokers
  - Optional in-flight with TLS between the clients and brokers
  - At rest for your EBS volumes using KMS
- Network Security:
  - Authorize specific security groups for your Apache Kafka clients
- Authentication & Authorization (important):
  - Define who can read/write to which topics
  - Mutual TLS (AuthN) + Kafka ACLs (AuthZ)
  - SASL/SCRAM (AuthN) + Kafka ACLs (AuthZ)
  - IAM Access Control (AuthN + AuthZ)



MSK – Monitoring

- CloudWatch Metrics
  - Basic monitoring (cluster and broker metrics)
  - Enhanced monitoring (++enhanced broker metrics)
  - Topic-level monitoring (++enhanced topic-level metrics)
- Prometheus (Open-Source Monitoring)
  - Opens a port on the broker to export cluster, broker and topic-level metrics
  - Setup the JMX Exporter (metrics) or Node Exporter (CPU and disk metrics)
- Broker Log Delivery
  - Delivery to CloudWatch Logs
  - Delivery to Amazon S3
  - Delivery to Kinesis Data Streams

MSK Connect

- Managed Kafka Connect workers on AWS
- Auto-scaling capabilities for workers
- You can deploy any Kafka Connect connectors to MSK Connect as a plugin
  - Amazon S3, Amazon Redshift, Amazon OpenSearch, Debezium, etc…
- Example pricing: Pay $0.11 per worker per hour

MSK Serverless

- Run Apache Kafka on MSK without managing the capacity
- MSK automatically provisions resources and scales compute & storage
- You just define your topics and your partitions and you're good to go!
- Security: IAM Access Control for all clusters

- Example Pricing:
  - $0.75 per cluster per hour = $558 monthly per cluster
  - $0.0015 per partition per hour = $1.08 monthly per partition
  - $0.10 per GB of storage each month
  - $0.10 per GB in
  - $0.05 per GB out

Kinesis Data Streams vs Amazon MSK

Kinesis Data Streams

- 1 MB message size limit
- Data Streams with Shards
- Shard Splitting & Merging
- TLS In-flight encryption
- KMS At-rest encryption
- Security:
  - IAM policies for AuthN/AuthZ

Amazon MSK

- 1MB default, configure for higher (ex: 10MB)
- Kafka Topics with Partitions
- Can only add partitions to a topic
- PLAINTEXT or TLS In-flight Encryption
- KMS At-rest encryption
- Security:
  - Mutual TLS (AuthN) + Kafka ACLs (AuthZ)
  - SASL/SCRAM (AuthN) + Kafka ACLs (AuthZ)
  - IAM Access Control (AuthN + AuthZ)