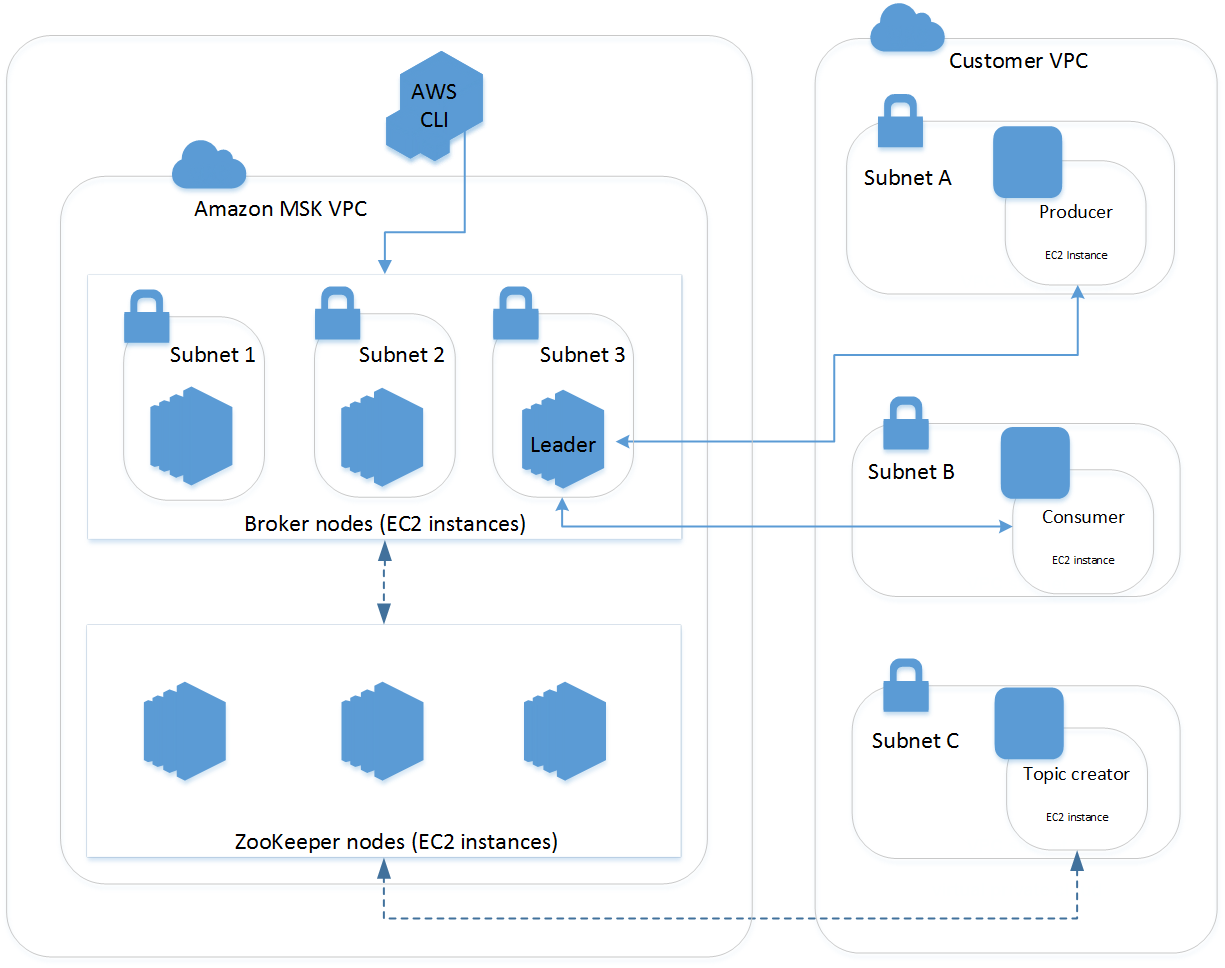
**Research Activity: Explore** [**https://aws.amazon.com/msk/**](https://aws.amazon.com/msk/) **to understand various case study & deployment**

Apache Kafka (Kafka) is an open-source platform that enables customers to capture streaming data like click stream events, transactions, IoT events, application and machine logs, and have applications that perform real-time analytics, run continuous transformations, and distribute this data to data lakes and databases in real time. We can use Kafka as a streaming data store to decouple applications producing streaming data (producers) from those consuming streaming data (consumers).

While Kafka is a [popular enterprise data streaming and messaging framework](https://cwiki.apache.org/confluence/display/KAFKA/Powered+By), it can be difficult to setup, scale, and manage in production. Amazon MSK takes care of these managing tasks and makes it easy to set up, configure, and run Kafka, along with [Apache ZooKeeper](https://zookeeper.apache.org), in an environment following best practices for high availability and security.

The MSK clusters always run within an [Amazon VPC](https://aws.amazon.com/vpc/) managed by the MSK service. The MSK resources are made available to the own VPC, subnet, and security group through [elastic network interfaces (ENIs)](https://docs.aws.amazon.com/vpc/latest/userguide/VPC_ElasticNetworkInterfaces.html) which will appear in the account, as described in the following architectural diagram:

[](https://d2908q01vomqb2.cloudfront.net/da4b9237bacccdf19c0760cab7aec4a8359010b0/2019/05/13/msk-architecture-visio.png)

The diagram demonstrates the interaction between the following components:

* **Broker nodes** — When creating an Amazon MSK cluster, you specify how many broker nodes you want Amazon MSK to create in each Availability Zone. In the example cluster shown in this diagram, there's one broker per Availability Zone. Each Availability Zone has its own virtual private cloud (VPC) subnet.
* **ZooKeeper nodes** — Amazon MSK also creates the Apache ZooKeeper nodes for you. Apache ZooKeeper is an open-source server that enables highly reliable distributed coordination.
* **Producers, consumers, and topic creators** — Amazon MSK lets you use Apache Kafka data-plane operations to create topics and to produce and consume data.
* **AWS CLI** — You can use the AWS Command Line Interface (AWS CLI) or the APIs in the SDK to perform control-plane operations. For example, you can use the AWS CLI or the SDK to create or delete an Amazon MSK cluster, list all the clusters in an account, or view the properties of a cluster.

Amazon MSK detects and automatically recovers from the most common failure scenarios for clusters so that your producer and consumer applications can continue their write and read operations with minimal impact. When Amazon MSK detects a broker failure, it mitigates the failure or replaces the unhealthy or unreachable broker with a new one. In addition, where possible, it reuses the storage from the older broker to reduce the data that Apache Kafka needs to replicate. Your availability impact is limited to the time required for Amazon MSK to complete the detection and recovery. After a recovery, your producer and consumer apps can continue to communicate with the same broker IP addresses that they used before the failure.

Customers can create a cluster in minutes, use [AWS Identity and Access Management (IAM)](https://aws.amazon.com/iam/) to control cluster actions, authorize clients using TLS private certificate authorities fully managed by [AWS Certificate Manager (ACM)](https://aws.amazon.com/certificate-manager/), encrypt data in-transit using TLS, and encrypt data at rest using [AWS Key Management Service (KMS)](https://aws.amazon.com/kms/) encryption keys.

Amazon MSK continuously monitors server health and automatically replaces servers when they fail, automates server patching, and operates highly available ZooKeeper nodes as a part of the service at no additional cost. Key Kafka performance metrics are published in the console and in [Amazon CloudWatch](https://aws.amazon.com/cloudwatch/). Amazon MSK is fully compatible with Kafka versions 1.1.1 and 2.1.0, so that we can continue to run the applications, use Kafka’s admin tools, and and use Kafka compatible tools and frameworks without having to change the code.

Amazon MSK added many new features such as:

* Encryption in-transit via TLS between clients and brokers, and between brokers
* Mutual TLS authentication using ACM private certificate authorities
* Support for Kafka version 2.1.0
* 99.9% availability SLA
* HIPAA eligible
* Cluster-wide storage scale up
* Integration with AWS CloudTrail for MSK API logging
* Cluster tagging and tag-based IAM policy application
* Defining custom, cluster-wide configurations for topics and brokers

Some use cases of Amazon MSK are as follows:

**Nutmeg**

Nutmeg is Europe’s largest digital wealth manager, helping consumers to grow their wealth and reach their financial goals by using cost effective technology to boost returns. Nutmeg uses Apache Kafka to underpin their journey towards an event driven architecture.

**Poshmark**

Poshmark is a leading social commerce platform for the next generation of retailers and shoppers. Through technology, our mission is to build the world’s most connected shopping experience, while empowering people to build thriving retail businesses.

**Vonage**

Vonage is a global business cloud communications leader providing integrated communications solutions that improve how business gets done by enhancing a company’s customer and employee experience. Vonage uses Apache Kafka for real-time communication between various micro-services

**Secureworks®**

Secureworks® (NASDAQ: SCWX) is a technology-driven cybersecurity leader that protects organizations in the digitally connected world.

**Compass**

Compass is a real estate technology company with a powerful end-to-end platform that supports the entire buying and selling workflow. Compass uses Apache Kafka to provide its agents with quick access to fresh and accurate data from hundreds of real estate data sources nationwide.

**ZipRecruiter**

ZipRecruiter is a leading online employment marketplace using AI-driven matching technology to actively connect millions of businesses and job seekers

**Delhivery**

Enabling Digital Commerce – Delhivery aims to become India’s largest and most profitable fulfilment company for digital commerce. Delhivery brings reach, speed, and the power of their customizable technology toolkit to their customer’s online logistics operations

**Vortexa**

Vortexa provides real-time analytics on changes in seaborne oil supply, helping energy market participants make better-informed trading and shipping decisions and see opportunities before others.