Apache Kafka & Streaming Jobs

Apache Kafka: A distributed event streaming platform for handling real-time data feeds. It allows data to be read and written in real-time from various systems.

Streaming Jobs: Kafka processes continuous data streams, making it ideal for real-time applications like stock trades, user activity tracking, or IoT data. Preferred when data needs to be processed in real-time or near-real-time, such as in monitoring systems or fraud detection.

Kafka Components

Brokers: Kafka servers that store data and serve client requests. A cluster can have multiple brokers for scalability.

Zookeeper: Manages and coordinates Kafka brokers. It keeps track of the cluster's health and configuration.

Kafka Server: A machine running the Kafka broker, responsible for storing data and serving requests.

Kafka Clients: Applications that produce (write) or consume (read) messages to/from Kafka.

KRaft: New consensus algorithm that removes the need for Zookeeper, making Kafka more efficient.

Kafka Cluster: A set of brokers working together to distribute data storage and processing load.