Questions and Answers

Questions and Answers

Q: What is live streaming?

A: Data streaming is the process of continuously transferring and processing large volumes of data in real-time.

It involves sending data from a source to one or more destinations, such as a data warehouse or an application, as a continuous flow of small, incremental updates rather than as a batch of data.

Q: What is Kafka?

A: Apache Kafka is an open-source distributed streaming platform that is used for building real-time data pipelines and streaming applications

Q: What is Kafka Topic?

A: Topics are created and managed by the Kafka cluster, and messages can be produced to and consumed from one or more topics by various applications or systems.

Topics can have multiple consumers, and each consumer can read messages from one or more partitions of a topic. Additionally, topics can be replicated across multiple brokers to provide fault tolerance and high availability.

Q: What is Kafka producer and consumer?

A: The producers are the source of data. It can be an IoT sensor, some files that appear somewhere in the cloud, or data being scraped from a web site(i.e., yFinance provide live and historical stock data).

The consumers are applications that process the data from producers.

Q: What are other live streaming applications?

A: There are various data streaming technologies available, including Apache Kafka, Apache Spark Streaming, and Apache Flink. These tools provide features such as fault tolerance, scalability, and real-time data processing, making them popular choices for building data-intensive applications.

Q: What kind of hardware is needed for live streaming?

A: Depending on the size of the application, we can use a single PC for very small/practice application to clusters of distributed data centers for very large applications such as in banks, stock market, social media, retail, etc.

Q. What are the technical challenges to live streaming?

A: There are many challenges that exist in live data streaming. Latency stands out the most. Delivering data in real-time or as fast as possible. Other challenges are fault tolerance, data integrity, security, and other challenges shared by large computer and software applications.

Q. What is event streaming?

A: Similar to data streaming, event streaming involves capturing and processing events in real-time. Events are defined as significant changes in the state such as clicking on a button or staring/pausing on a screen longer than a set period. IoT devices can issue events for example.

Q. Can use of live data streaming be dangerous?

A: Anything that has to do with data and data capturing and distributing can be dangerous. Data breaches and invasion of privacy of the public are the paramount dangers. Dissemination of sensitive data to wrong destinations can be disastrous at all levels from individuals to corporations to governments.

Q: Can Kafka run in the cloud?

A: Amazon, Google, Microsoft, and other offer cloud servers to kafka. Use of cloud servers offers benefits such as flexibility, scalability, fault tolerance, security and many other IT challenges that businesses need not be concerned about.