Apache Flume

- Apache Flume is a tool/service/data ingestion mechanism for collecting aggregating and transporting large amounts of streaming data such as log data, events (etc...) from various web serves to a centralized data store.
- It is a highly reliable, distributed, and configurable tool that is principally designed to transfer streaming data from various sources to HDFS.
- Apache Flume is a distributed, reliable and available system for efficiently collecting, aggregating and moving large amounts of log data from many different sources to a centralized data store.

The use of Apache Flume is not only restricted to log data aggregation. Since data sources are customizable therefore Flume can be used to transport massive quantities of event data including but not limited to network traffic data, social-media-generated data, email messages and pretty much any data source possible.

Apache Kafka

- Kafka has been developed by Apache Software Foundation. It is an open-source message broker. Using Kafka, we can handle feeds with high-throughput and low-latency.
- It is an open-source stream-processing software platform written in Java and Scala.
- Apache Kafka aims to provide a high throughput, unified, low-latency platform for handling the real-time data feeds.
- Kafka generally used TCP based protocol which optimized for efficiency. It is very fast and performs 2 million writes per second.
- Apache Kafka generally used for real-time analytics, ingestion data into the Hadoop and to spark, error recovery, website activity tracking.
- Guarantees zero percent data loss.