**Difference between Apache Kafka and Flume**

Kafka and Flume both are used for real time event processing system. They both are developed by Apache. Kafka is a publish-subscribe model messaging system. It can be used to communicate between publisher and subscriber using topic. One of the best features of **Kafka** is, it is highly available and resilient to node failures and supports automatic recovery.

On the other hand, **flume** is mainly designed for Hadoop and it is a part of Hadoop ecosystem. It is used to collect data from different sources and transfer data to the centralized data store. Flume was mainly designed in order to collect streaming data (log data) from various web servers to HDFS.

| **Sr. No.** | **Key** | **Apache Kafka** | **Flume** |
| --- | --- | --- | --- |
| 1 | Basic | Apache Kafka is a distributed data store optimized for ingesting and processing streaming data in real-time | 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. |
| 2 | scalable | It is easy to scale | It is not scalable as Kafka |
| 3 | Push /Pull | Kafka is basically working as a pull model | Flume is basically working as a push model |
| 4 | Recovery | It is highly available and resilient to node failures and supports automatic recovery | In case of flume-agent failure, you will lose events in the channel |
| 5. | Flexibility | Kafka is a general purpose  publish-subscribe model messaging system | It is specially designed for Hadoop |