**Apache Kafka:**

Apache Kafka is a distributed event streaming platform. With an ability to allow applications to manage large amounts of data,Kafka is also fault-tolerant and built to scale. Apache Kafka’s framework is based on Java and the Publish-Subscribe Messaging system. The framework allows data streaming at an unprecedented rate, that too, from multiple sources Kafka is famous in the data community for data streaming services because it can handle Big Data with large input volumes. And, with minimum downtime and low latency, Kafka services are easy to scale up and down.

**Some of Kafka’s most valuable features are as follows:**

**High Scalability:** The partitioned log model allows Kafka services to scale beyond a single server’s capability.

**Low Latency:** Kafka services separate data streams, allowing low latency and high throughput.

**Fault-Tolerant & Durable:** In Kafka, partitions are segregated then duplicated across servers. The segregation and duplication process makes Kafka services fault-tolerant by protecting them against ad-hoc server failures like master and database failures.

**High Extensibility:** Kafka is highly accessible through various other applications, allowing developers to add more features.

**problem statement:-** to create a data pipeline, suppose we have to pull data from the source, and through Kafka topic, we need to collect that data into target. Kafka has a pub-sub architecture, the producer is responsible to publish data into Kafka broker inside specific topics into specific partitions and the consumer is responsible for subscribing data from those partitions. so here use case is like here in terms of source we are pulling data from an excel sheet and dumping data into Kafka broker-topic and from there we are reading data and store in output excel.

