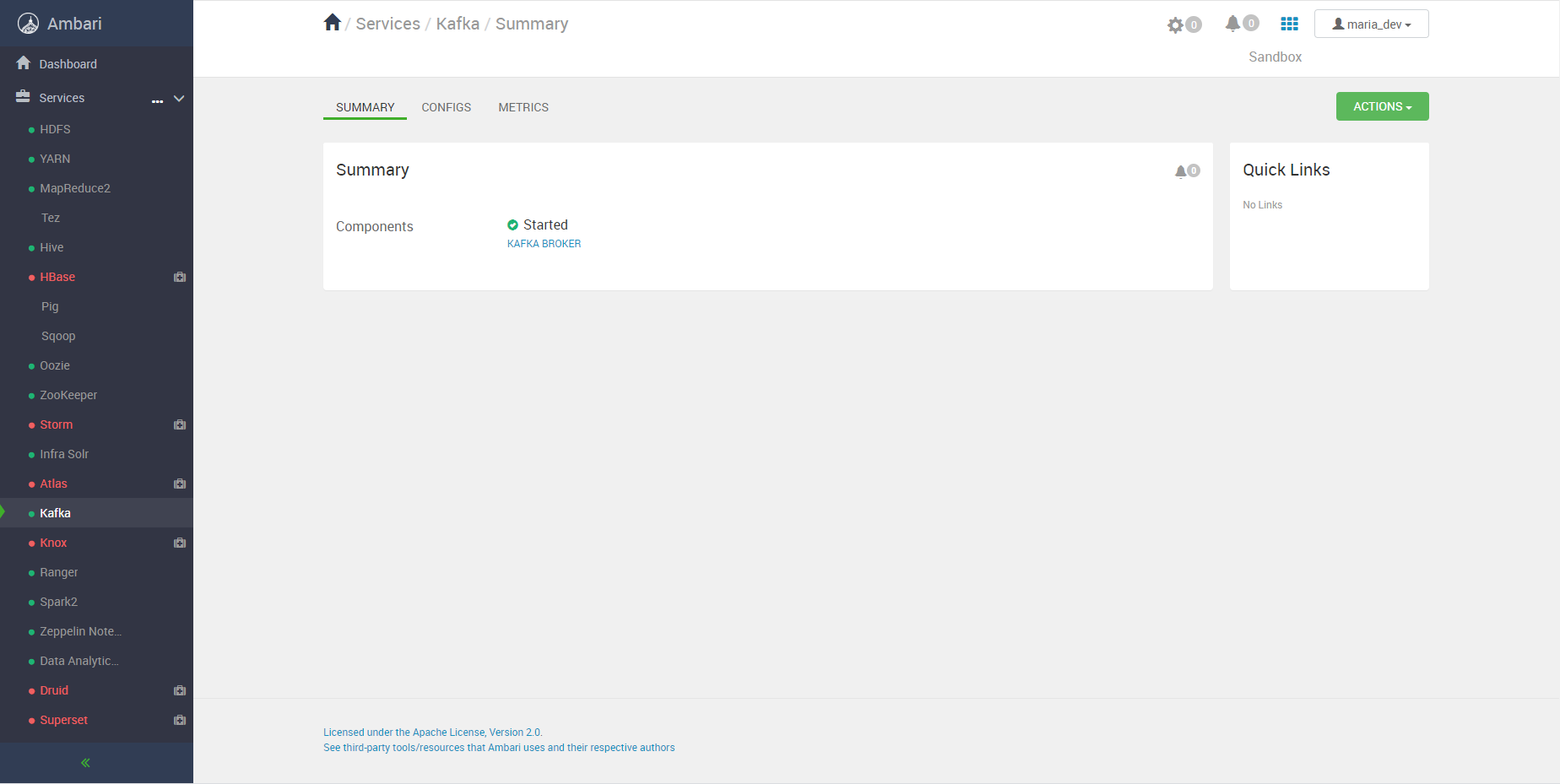
**Introduction to Big Data- Task 2**

**Kafka:**

It’s a **distributed stream platform**. A streaming platform has three key capabilities:

1. **Publish and subscribe to streams of records**, like a message queue or enterprise messaging system.
2. **Store streams of records in a fault-tolerant durable way**.
3. **Process streams of records** as they occur.

****

In the above image, Kafka is running.

**Kafka has four core APIs:**

1. **Producer API**

The Producer API allows an application to **publish a stream of records to one or more Kafka topics**.

1. **Consumer API**

The Consumer API **allows an application to subscribe to one or more topics and process the stream of records produced to them**.

1. **Streams API**

The Streams API **allows an application to act as a stream processor, consuming an input stream from one or more topics and producing an output stream to one or more output topics, effectively transforming the input streams to output streams.**

1. **Connector API**

The Connector API **allows building and running reusable producers or consumers that connect Kafka topics to existing applications or data systems**.

**Topic in Kafka:**

A topic is a **category name to which records are published**. Topics in **Kafka are always multi-subscriber; that is, a topic can have zero, one, or many consumers that subscribe to the data written to it**.



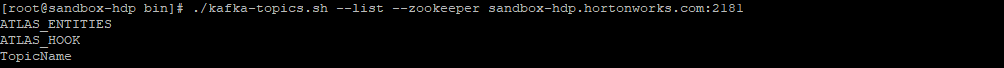
In the above image we have following details:

Partitions: 1

Replication factor: 1

Topic name: TopicName

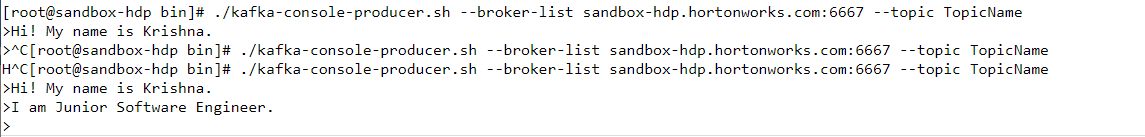
**Topic Listing in Kafka:**

****

In the above image, we can list the Topic names currently exists.

**Producer in Kafka:**

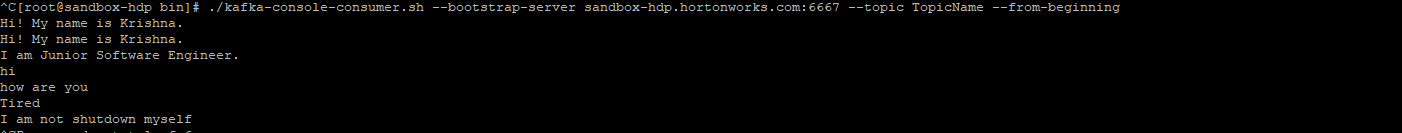
Producers **publish data to the topics of their choice**. The producer is responsible for choosing which record to assign to which partition within the topic. This can be done in a round-robin fashion simply to balance load or it can be done according to some semantic partition function.



In the above image, we are sending data using Producer to consumer.

**Consumer in Kafka:**

Each **record published to a topic is delivered to one consumer instance within each subscribing consumer group**. Consumer instances can be in separate processes or on separate machines.



In the above image, we are getting data from Producer to consumer. We are getting whole data include previous session data.