Kafka Fundamentals
In this first learning block, we are going through Kafka 101.

1. We'll have a very necessary all-theory section, during which we'll learn all the fundamentals of Kafka
2. We'll then go ahead and set up Kafka on our computers and in the cloud
3. We'll spend a lot of time learning how to use the Kafka CLI
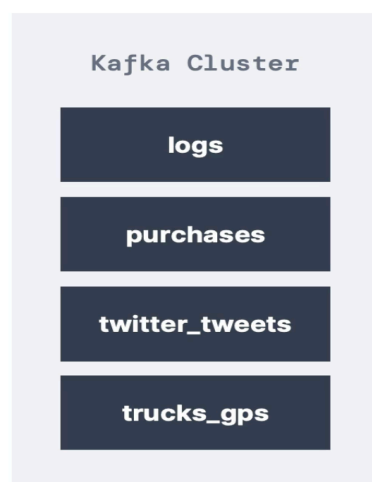4. We'll write some Java Code to create our first Producers and Consumers

By the end of this learning block, you'll be "proficient" with Kafka.

Hope you're ready, let's get started!
==
==

Kafka Topics:
- Kafka topics are a particular stream of data within your kafka cluster
- It's like a table in a database(without all the constraints, meaning you can send anything you want to kafka topic, there is no data verification)
- You can have as many topics as you want
- A topic is identified by its name in the kafka cluster
- Support any kind of message format, i.e- json,avro, text, binary…
- Sequence of messages in a topic is called a data stream(that is why kafka is called as a data streaming platform)
- You cannot **query** topics, instead use kafka **Producers** to send data and kafka **consumers** to read the data
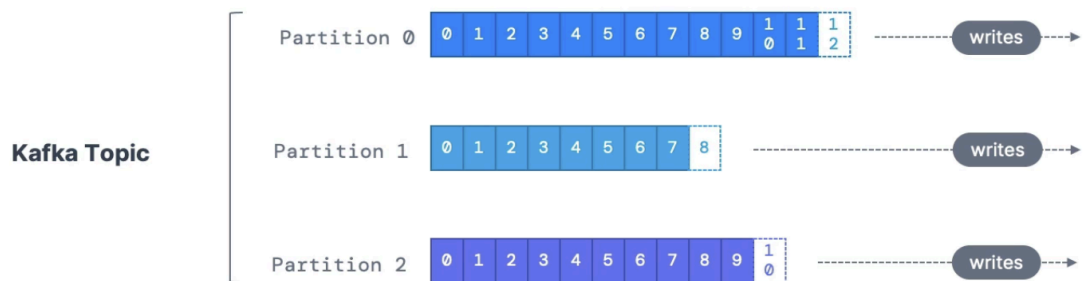

These are all different topics within the same kafka cluster.

==
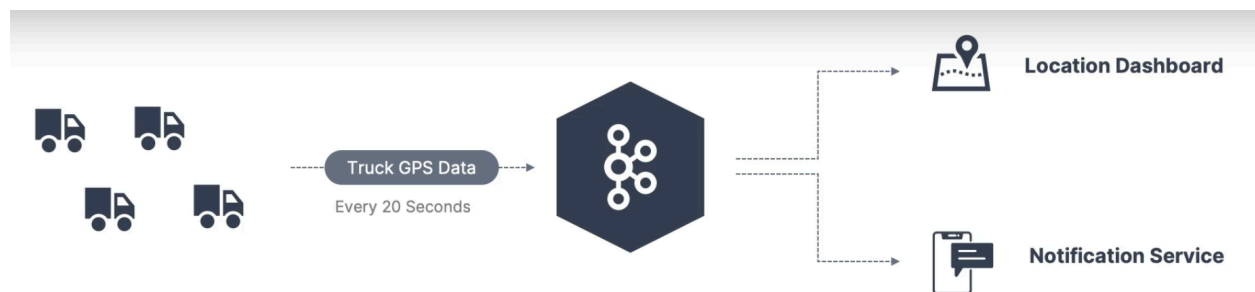==

Partitions and Offsets:
- Topics are split in **partitions**(for example - 100 partitions)
  - The messages sent to kafka topic are going to end up in these partitions and messages within each partition are going to be ordered
  - Each message within the partition, gets an incremental id, called **offset**

- Kafka topics are **immutable:** once data is written to a partition, it cannot be changed.

Topics Example: truck_gps

- Say you have a fleet of trucks, each truck reports its position to apache kafka
- Each truck will send a message to kafka every 20 seconds, each message will contain the truckId and truck position (latitude/longitude)
- You can have a topic truck_gps that contains the position of all the trucks.
- We choose to create the topic with 10 partitions(arbitrary number for now)

Important Notes - Topics, Partitions and Offsets:
- Once the data is written to a partition, it cannot be changed(immutability)
- Data is kept only for a limited time (default is one week - configurable)
- Offset only has meaning for a specific partition.
  - Offset 3 in partition 0 will b hold different data than offset 3 in partition 1 or 2.
- Offsets are not **reused** even if previous messages have been deleted.
- Order is guaranteed only within a partition (not across partitions)
- Data is assigned randomly to a partition, unless a key is provided(More on this later)
- You can have as many partitions per topic as you want