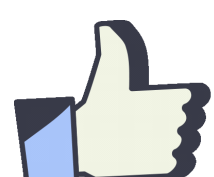
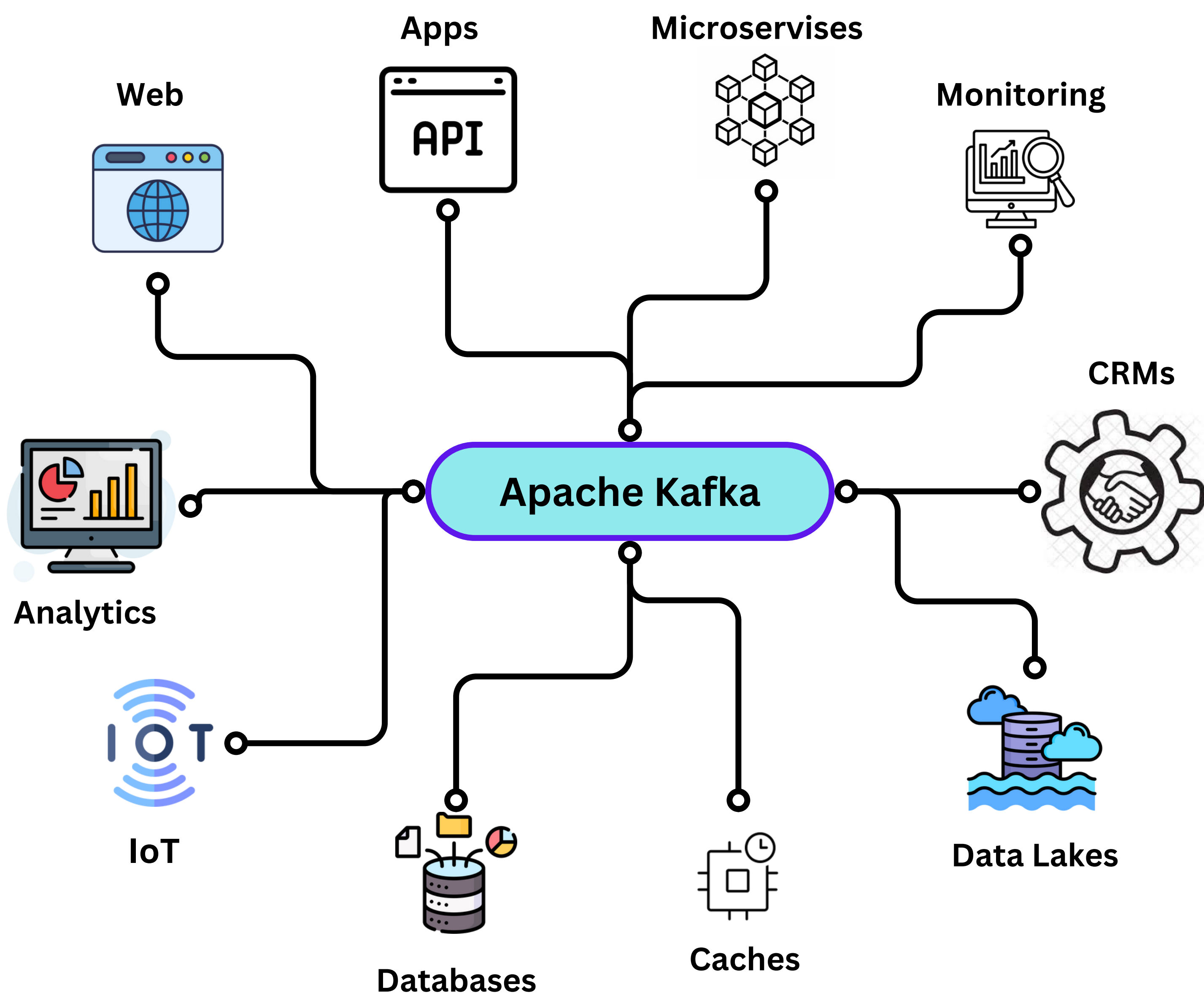


Kafka

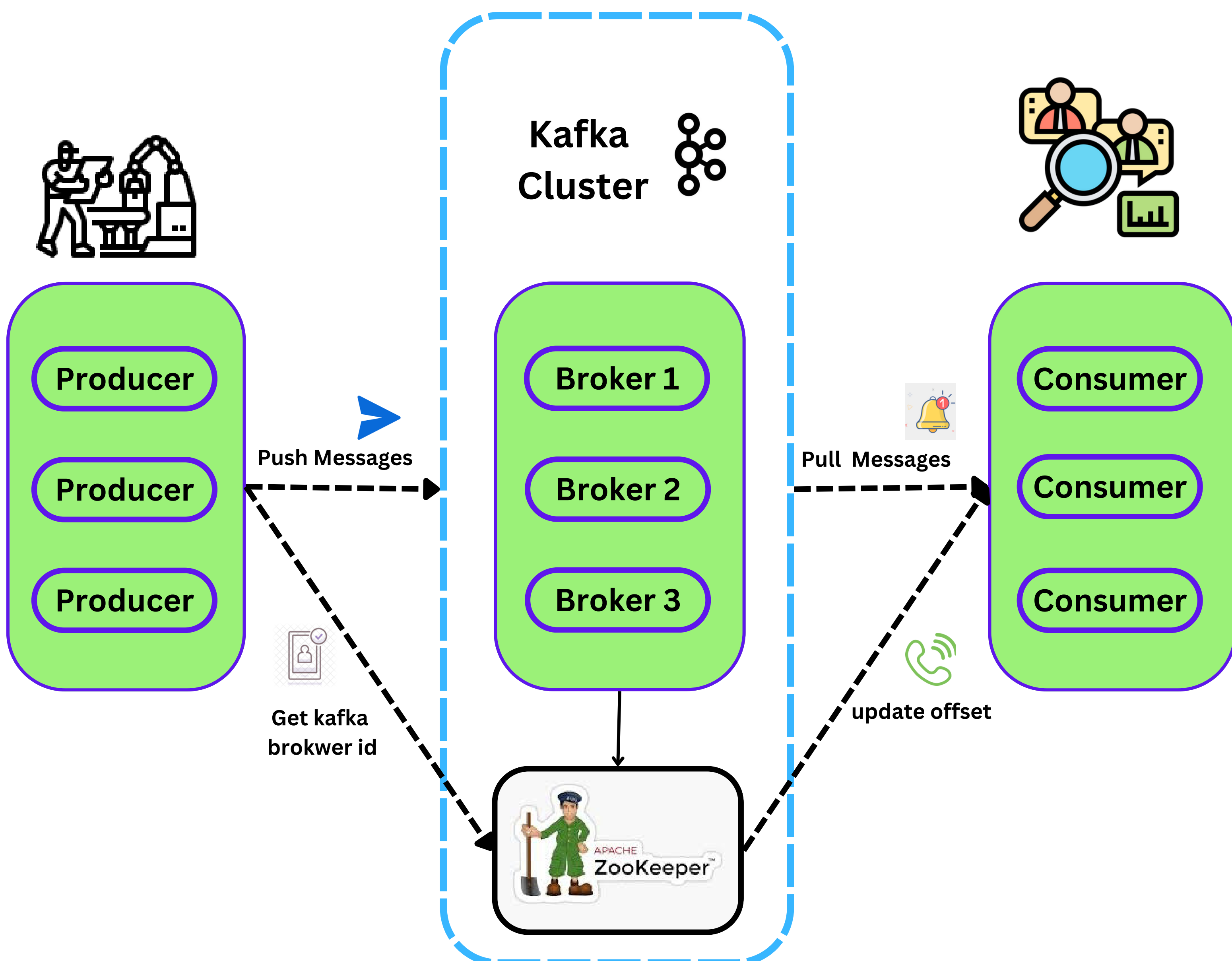


Follow

Share

Kafka Ecosystem

- Apache Kafka is an open-source distributed streaming platform. Kafka is widely used for building realtime data pipelines, event sourcing, log aggregation, and stream processing applications.



Kafka Cluster

- Since Kafka is a distributed system, it acts as a cluster.
- A Kafka cluster consists of a set of brokers.
- A cluster has a minimum of 3 brokers.



Kafka Broker

- The broker is the Kafka server. It's just a meaningful name given to the kafka server and this name makes sense as well because all that kafka does is act as a message broker between producer and consumer.
- The producer and consumer don't interact directly. they use kafka server as an agent or a broker to exchange messages.



Producer

- **Producer is an application that sends messages. it does not send messages directly to the recipient. it send messages only to the kafka server.**



Consumer

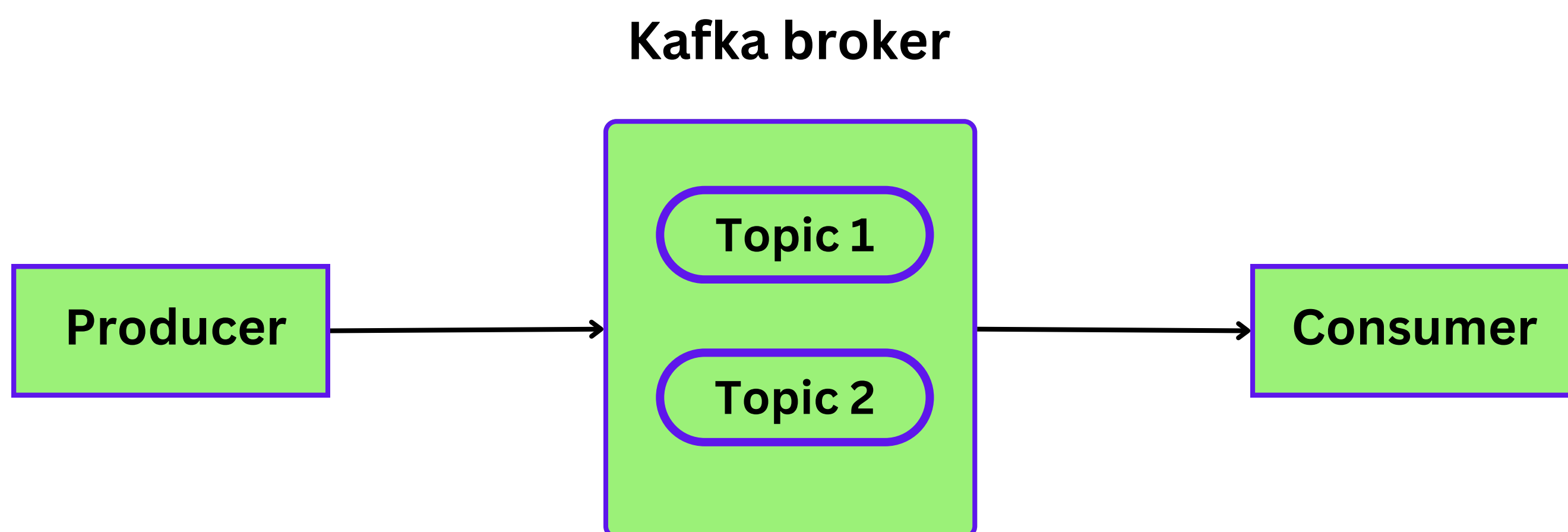
- Consumer is an application that reads messages from the kafka server.
- If producers are sending data, they must be sending it to someone, right? The consumers are the recipients. but remember that the producers don't send data to a recipient address. they just send it to kafka server. and anyone who is interested in that data from a kafka server is a **consumer**, and they can ask for data sent by any producer provided they have permissions to read it.



Kafka Topic

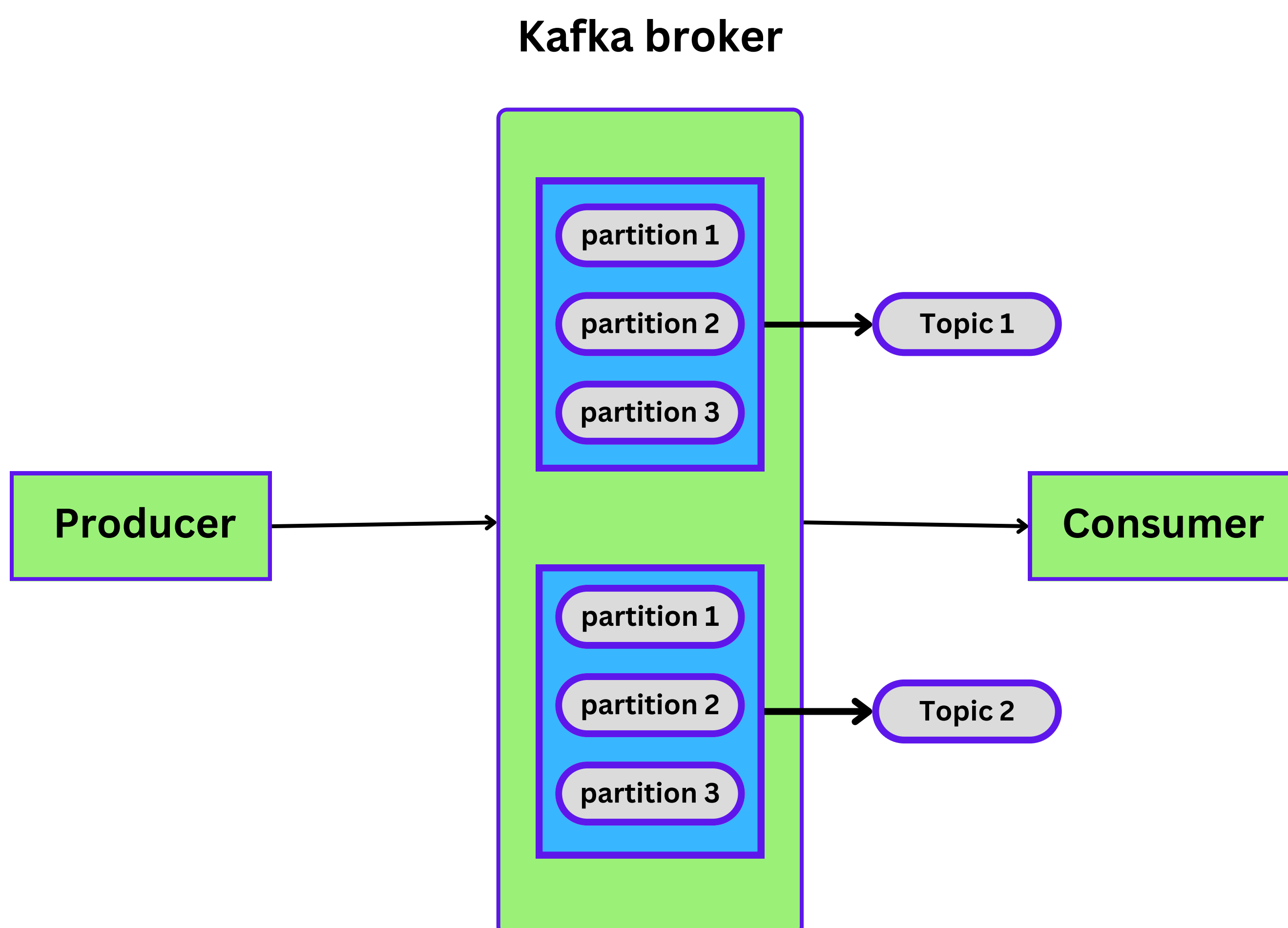
We learned that producer sends data to the kafka broker. then a consumer can ask for data from the kafka broker. but the question is, which data? we need to have some identification mechanism to request data from a broker. there comes the notion of the **topic**.

- Topic is like a table in database or folder in a file system.
- Topic is identified by a name.
- You can have any number of topics.



Kafka Partitions

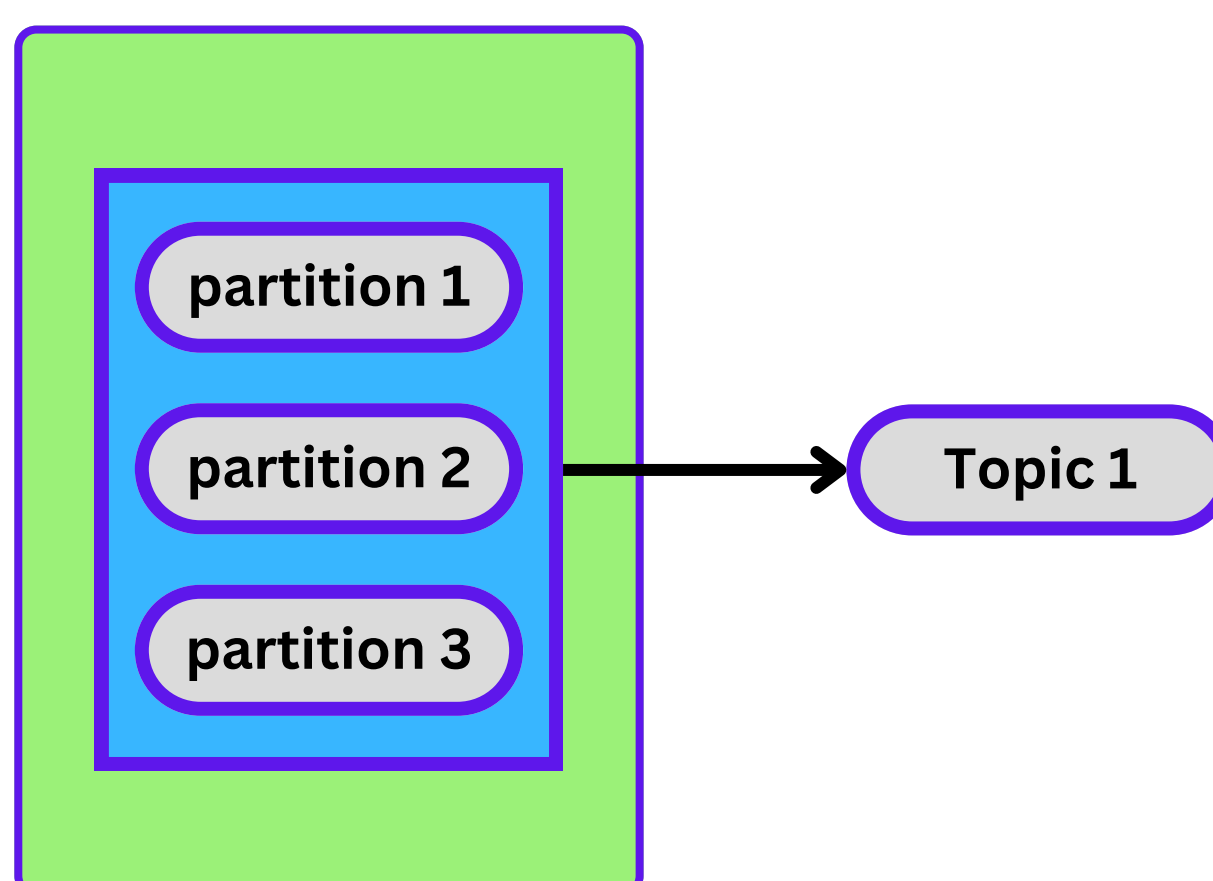
- Kafka topics are divided into a number of partitions, which contain records in an **unchangeable** sequence.
- Kafka brokers will store messages for a topic. but the capacity of data can be enormous and it may not be possible to store in single computer. therefore it will partitioned into multiple parts and distributed among multiple computers, since kafka is a distributed system.



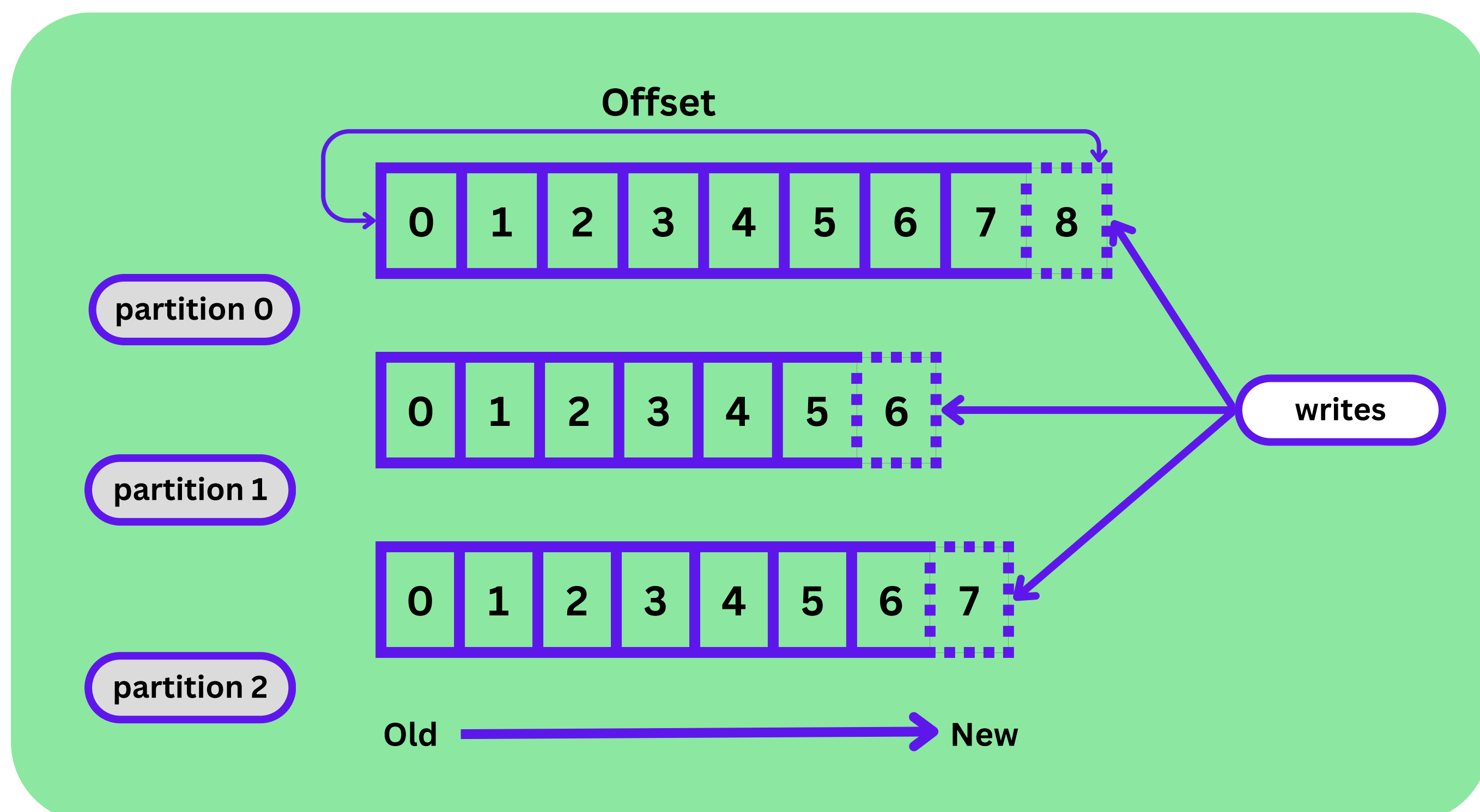
Offsets

- Offset is a sequence of ids given to messages as they arrive at a partition. once the offset is assigned it will never be changed. the first message gets an offset zero. the next message receives an offset one and so on.

Kafka broker

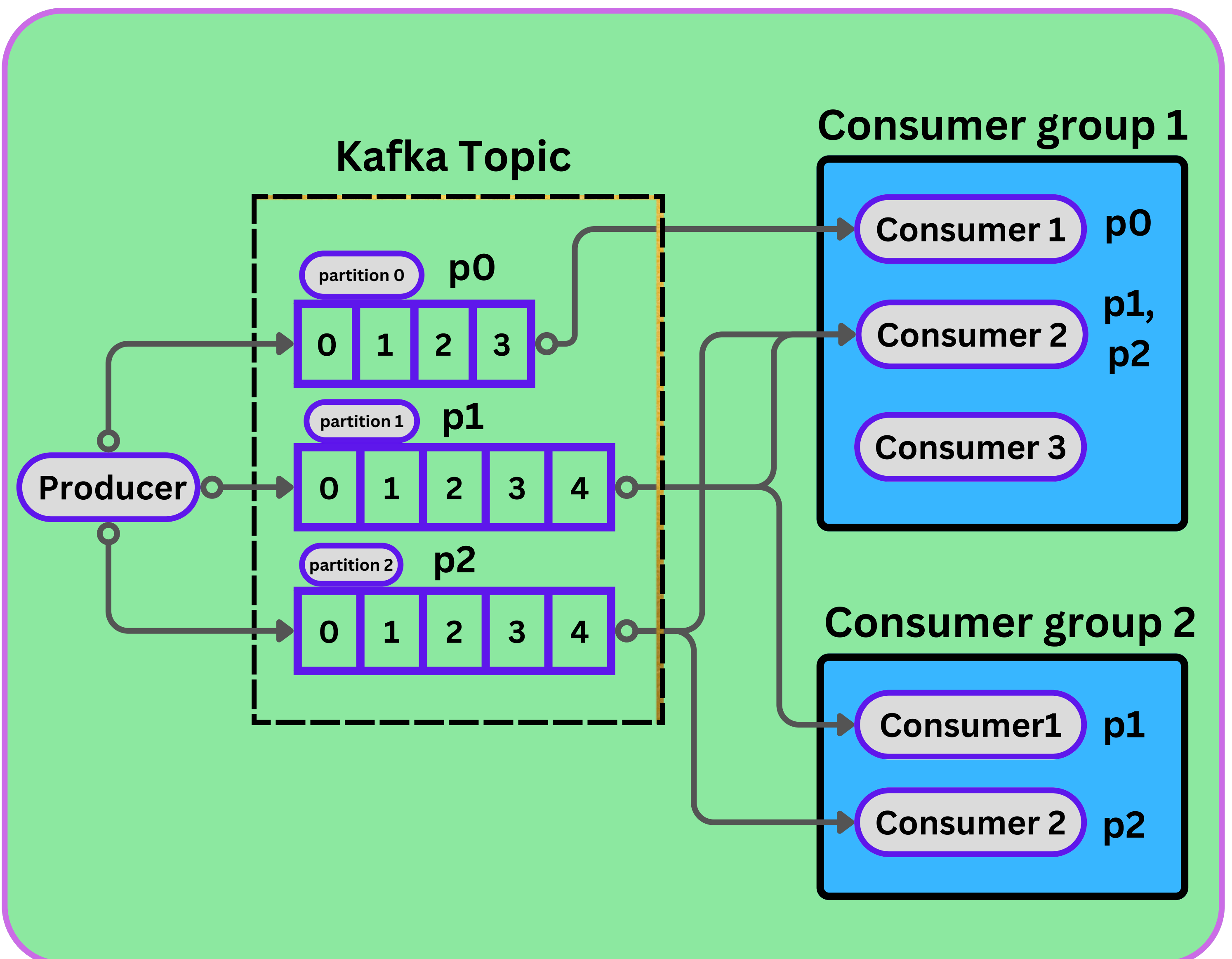


Kafka Topic



Consumer Groups

- A consumer group contains one or more consumers working together to process the messages.
- Diagram that shows a kafka architecture. A producer is feeding into a kafka topic over three partitions and the messages are then being subscribed to by consumers.



Thank you!



vishal-bramhankar



techwithvishalraj



Vishall0317

