KAFKA-PART 2

MORE TERMS

https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Bootstrap Server

This provides the initial hosts that act as the starting point for a Kafka client to discover the full set of alive servers in the cluster.

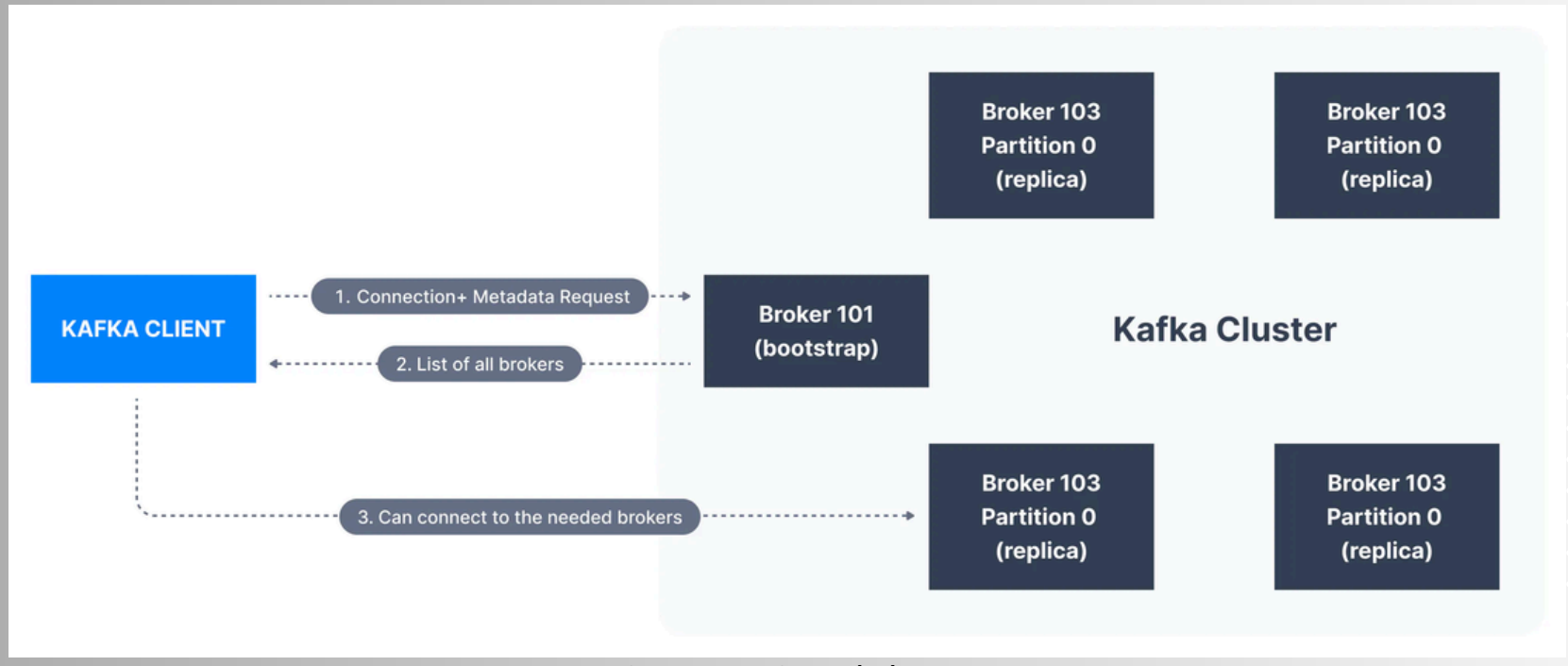
https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Bootstrap Server

It is list of host/port pairs to connect to Cluster



Source: Conduktor

https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Producer Serialization

Serialization is the process of converting data structures or objects into a format that can be transmitted over the network or stored in a file.

Its encoding the data in binary format before sending it to Kafka topic.

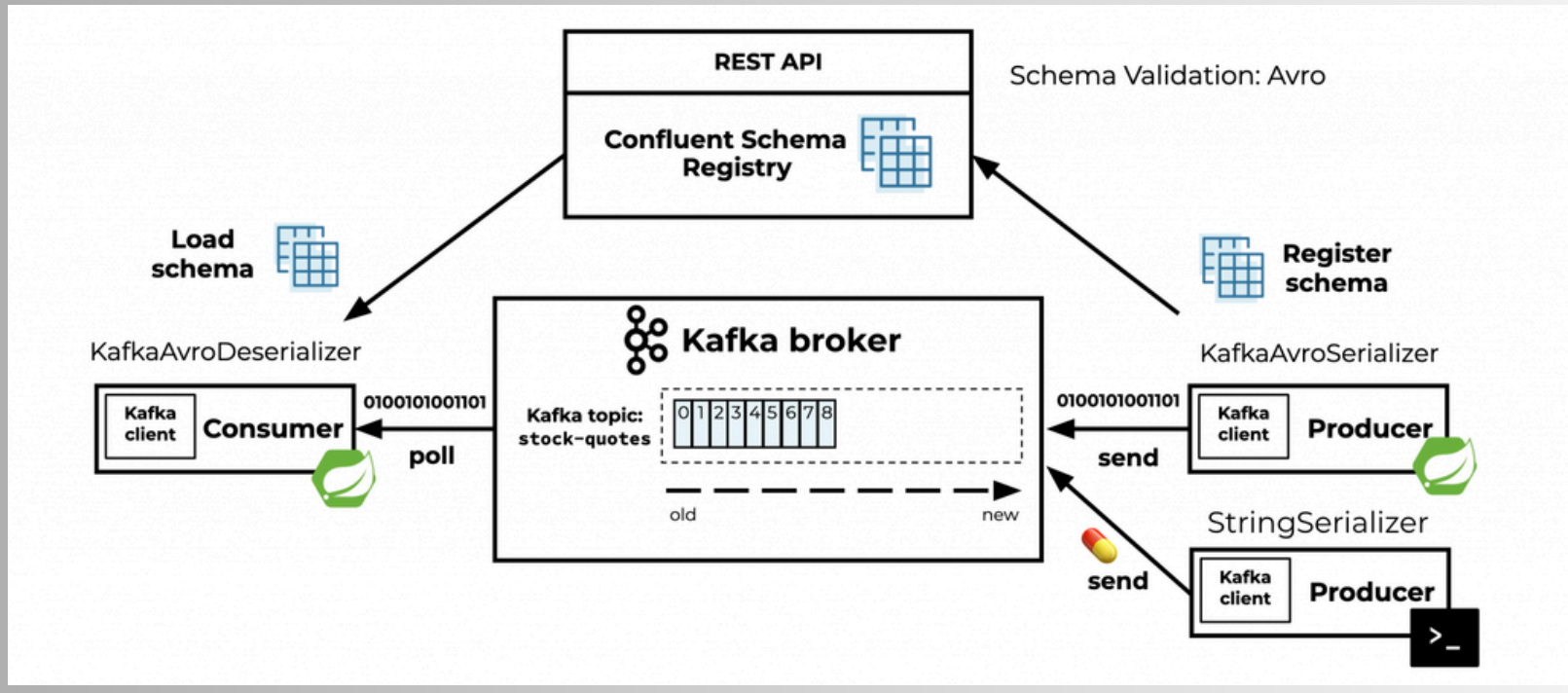
https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Producer Serialization

Kafka messages have key, value and needs to be converted to binary format before getting transmitted.



Source: Confluent

Kafka supports built-in & custom serialization

https://www.linkedin.com/in/vijay-mahawar

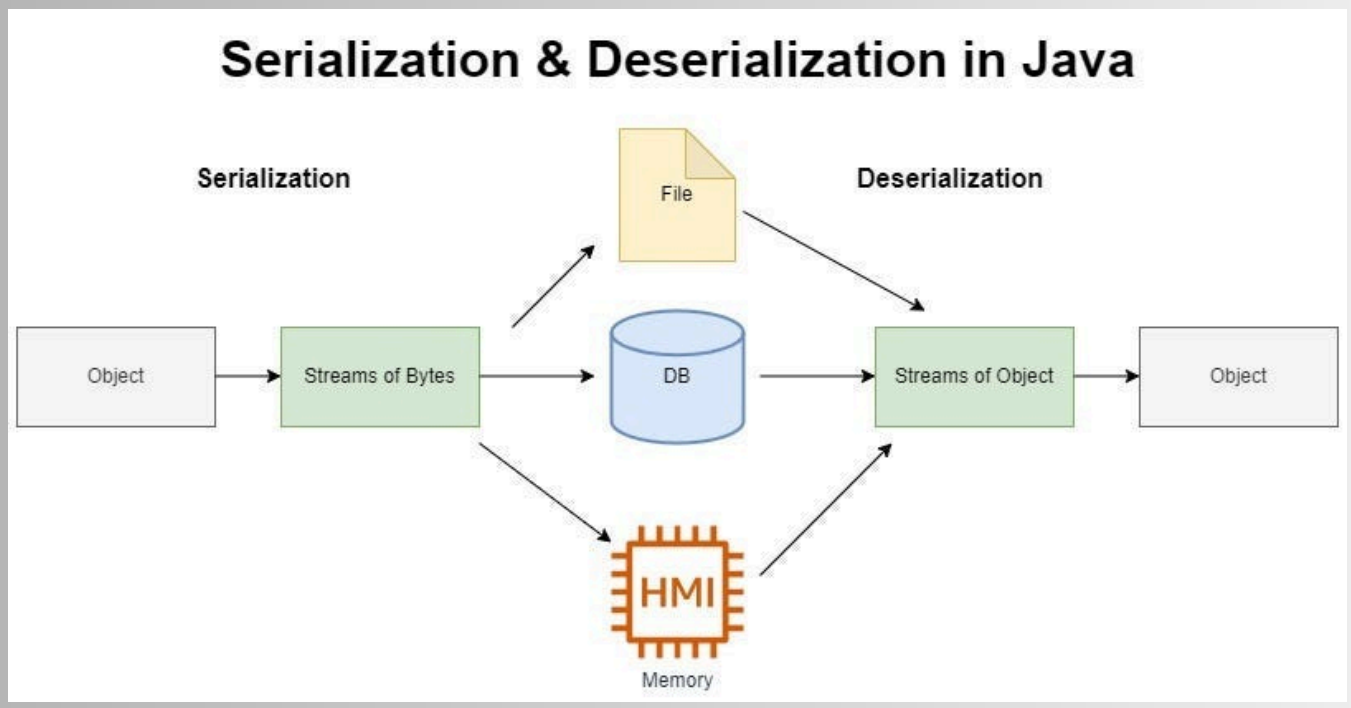
KAFKA-PART 2

MORE TERMS

Consumer Deserialisation

It is the process of converting data structures or objects back to its original format.

Its decoding the data from binary format to original format once pulled from Kafka topic.



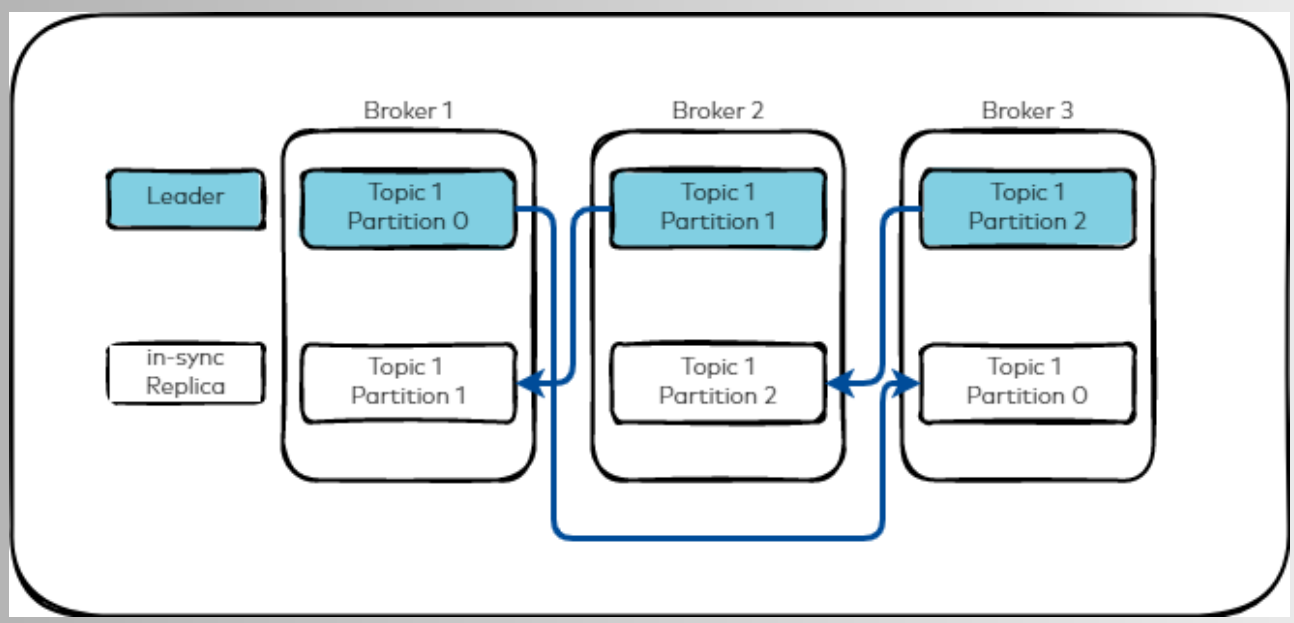
https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Topic Replication

Data is written down not just to one broker, but many.



Source: Confluent

Automatic failover to these replicas when a server in the cluster fails, so messages remain available.

https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Replica Factor

The replication factor includes the total number of replicas including the leader.

Image in previous shows 2 replicas for each partition.

Topics with a replication factor of one are topics that are not replicated.

https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Leader Replica

Under non-failure conditions, each partition in Kafka has a single leader and zero or more followers.

When a leader fails, a new leader must be chosen from the followers.

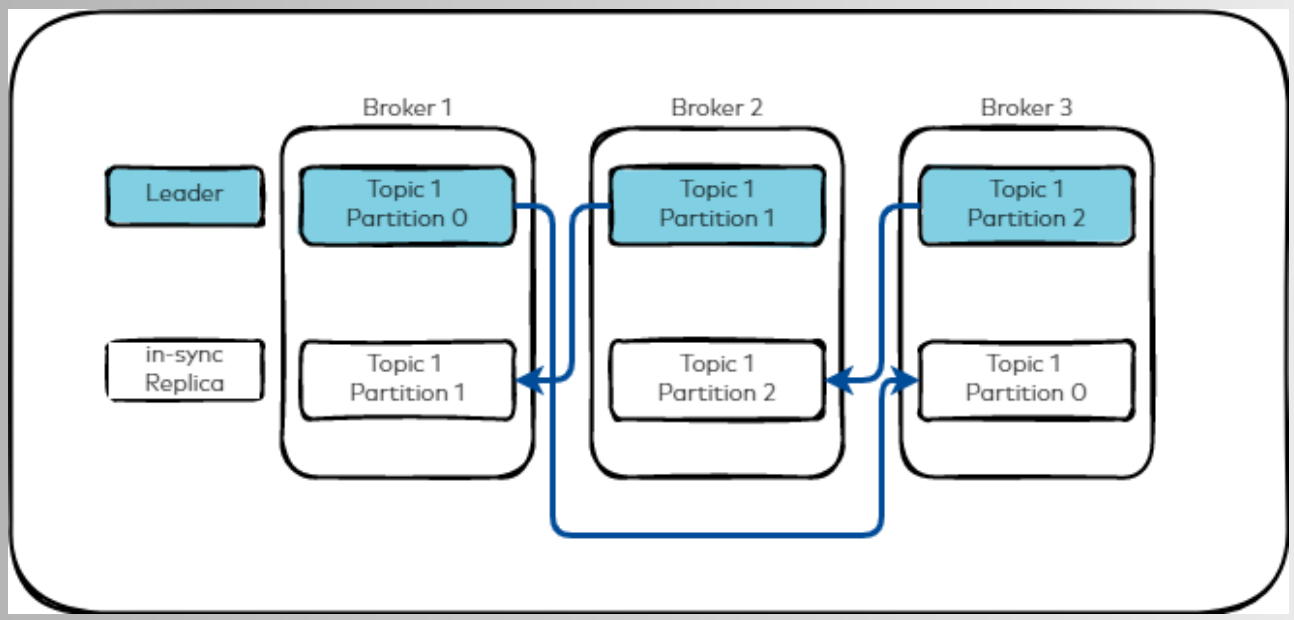
https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

In-Sync Replica(ISR)

The ISR is simply all the replicas of a partition that are “in-sync” with the leader in the last 10 seconds or time period as per *replica.lag.time.max.ms*



Source: Confluent

https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Producer Acknowledgement

Acknowledgment is also known as confirmation.

acks=0 (possible data loss): Producers sends and does not wait for acknowledgement

acks=1 (limited data loss): Producer sends and awaits acknowledgement only from Leader Replica. Does not guarantee replication

acks=all (no data loss): Producer sends and awaits acknowledgement from all replica

https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Consumer group

Consumer groups allow Kafka consumers to work together and process events from a topic in parallel.

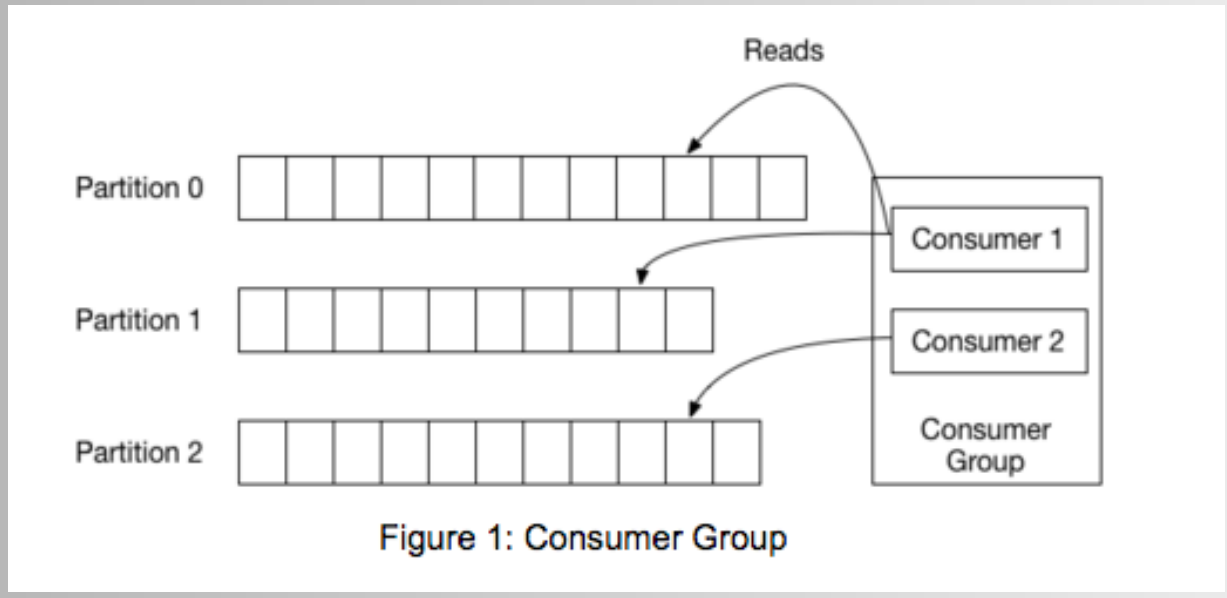


Figure 1: Consumer Group

Consumers are assigned a subset of partitions from a topic or set of topics and can parallelize the processing of those events.

https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Consumer Offset

When a consumer reads data from a Kafka topic, it processes messages sequentially.

To keep track of what has already been read, Kafka records a marker called the consumer offset.

It acts as a pointer to the last record that Kafka has already sent to the consumer during the most recent poll.

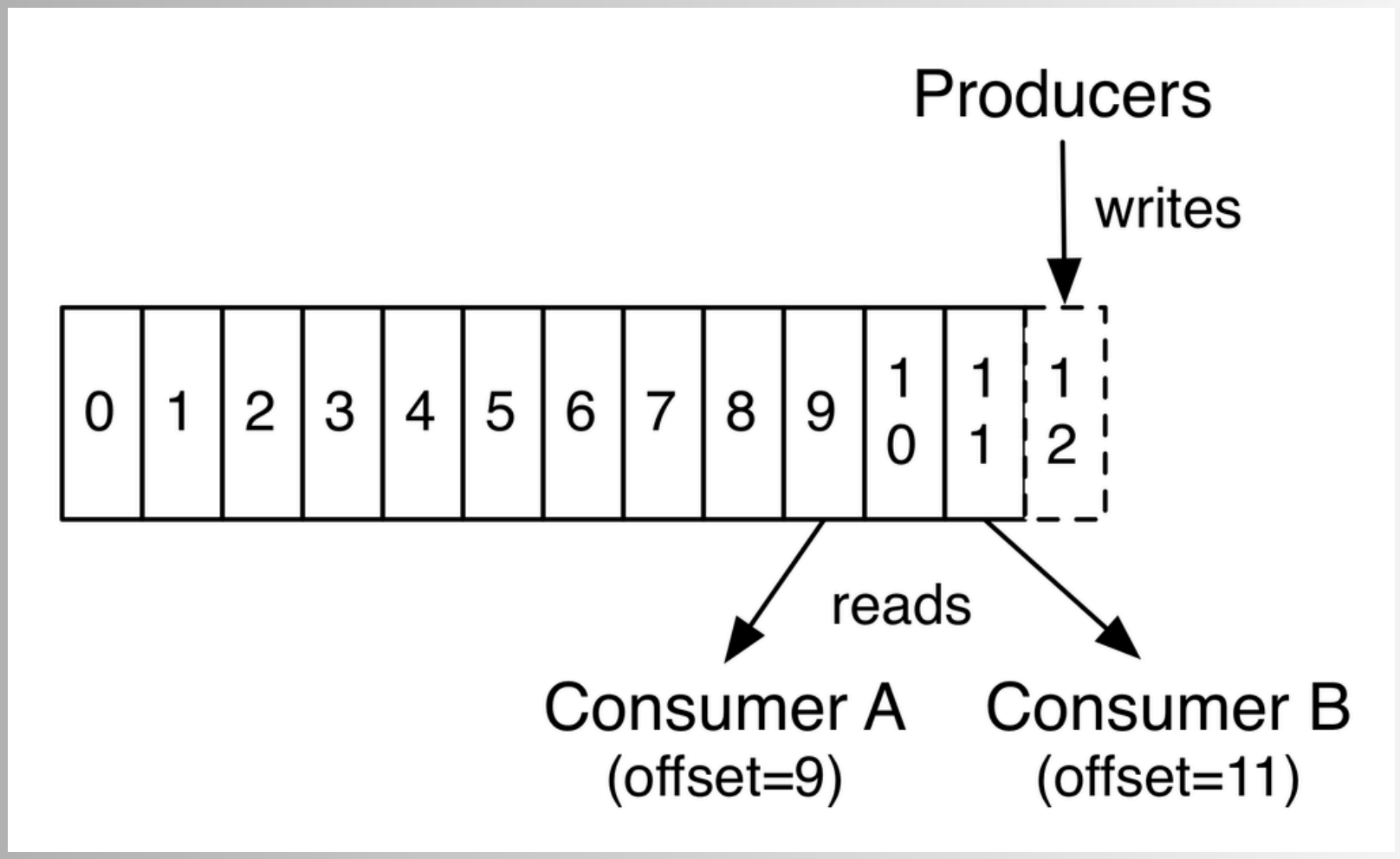
https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Resetting Offset

The pointer/mark to the message in Kafka can be controlled manually.



https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Resetting Offset

You can also use this tool to reset the consumer offset in scenarios where a consumer is stalled or significantly lagging.

<u>Use:</u> ***-reset-offsets with –shift-by –to-earliest***

https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Producer Idempotent

This ensures that messages published on Kafka topics are not duplicated from the producer side.

This ensures exactly-once semantic.

https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Ktable

KTable in Kafka Streams is an abstraction of a changelog stream that saves state in Kafka Streams.

KTable represents a primary-keyed table where each record in the changelog stream corresponds to an update on the table.

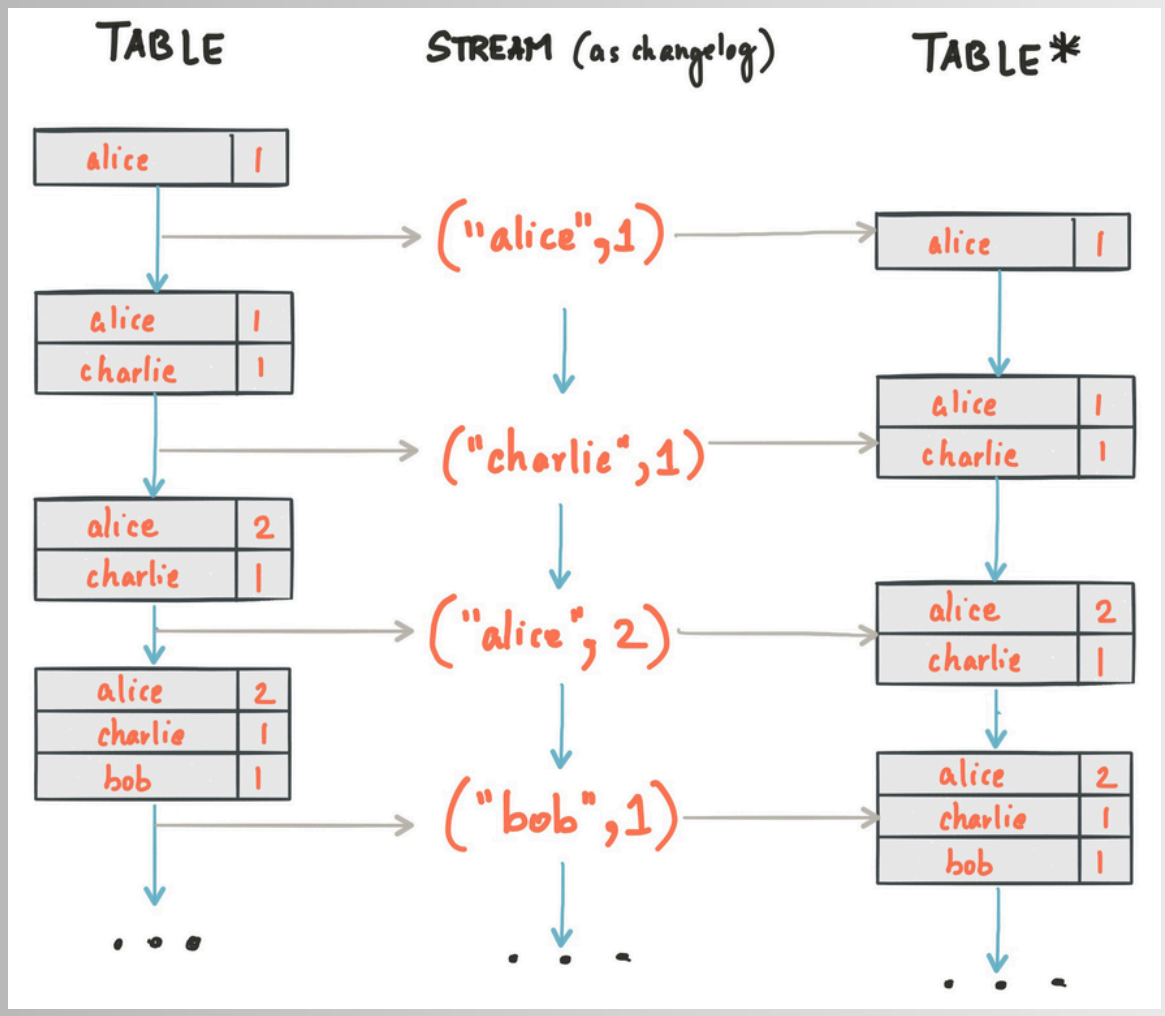
https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Ktable

The record key serves as the primary key for the table. This is similar to database tables



Source: Confluent

https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Zero copy

Moving the data from one buffer to another without involving the operating system process is called zero copy.

https://www.linkedin.com/in/vijay-mahawar

KAFKA-PART 2

MORE TERMS

Thanks

Have A Nice Day

https://www.linkedin.com/in/vijay-mahawar