

Stream Processing

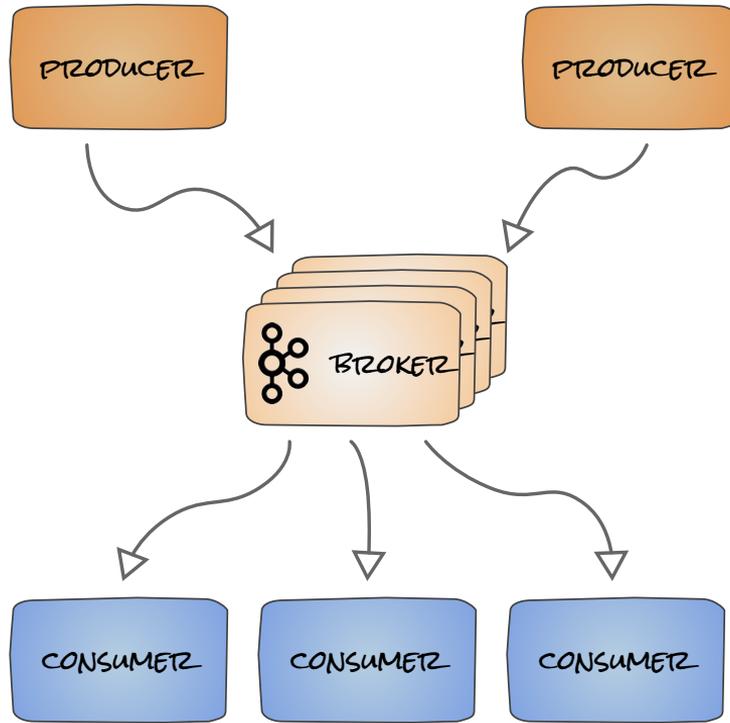
with

Kafka and KSQL

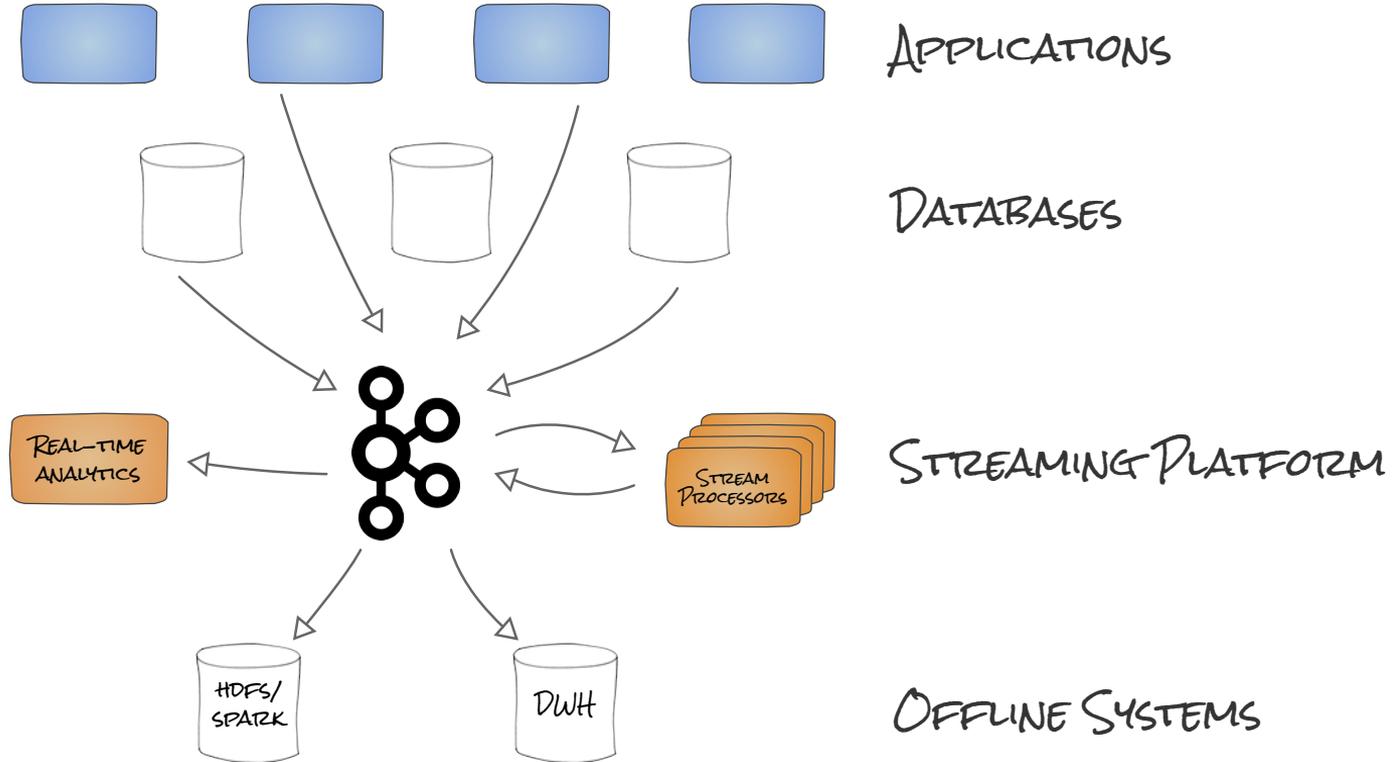
<https://github.com/confluentinc/kafka-workshop>

@tberglund
@rmoff

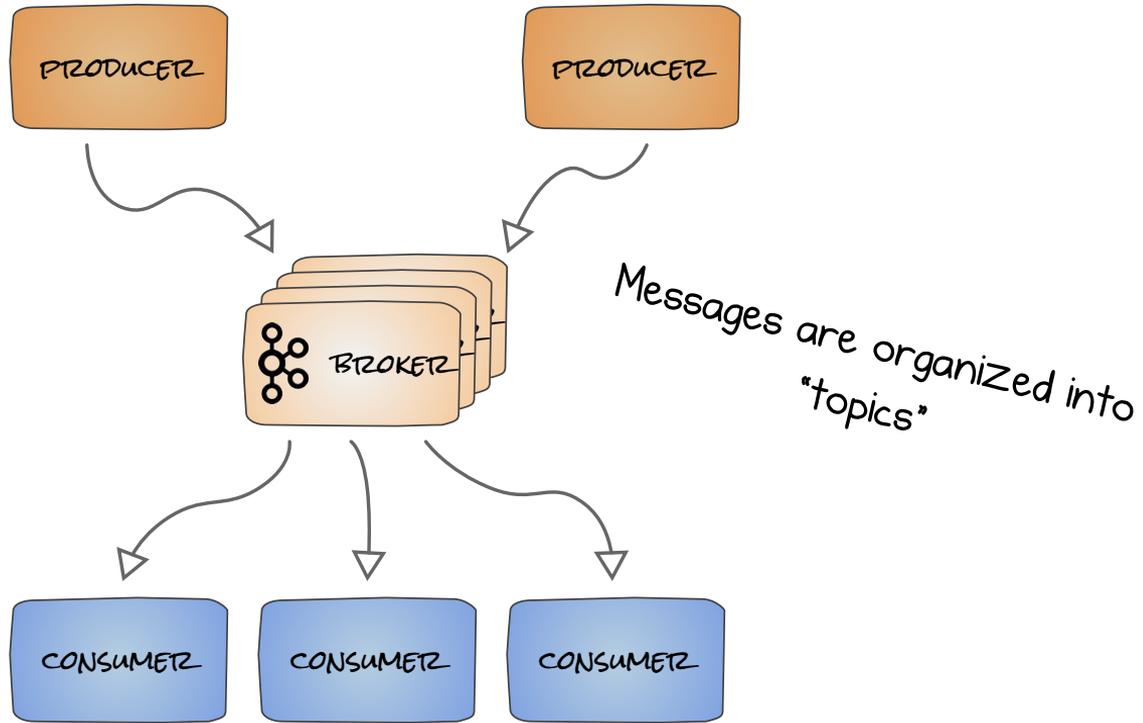
Basic Architecture Model



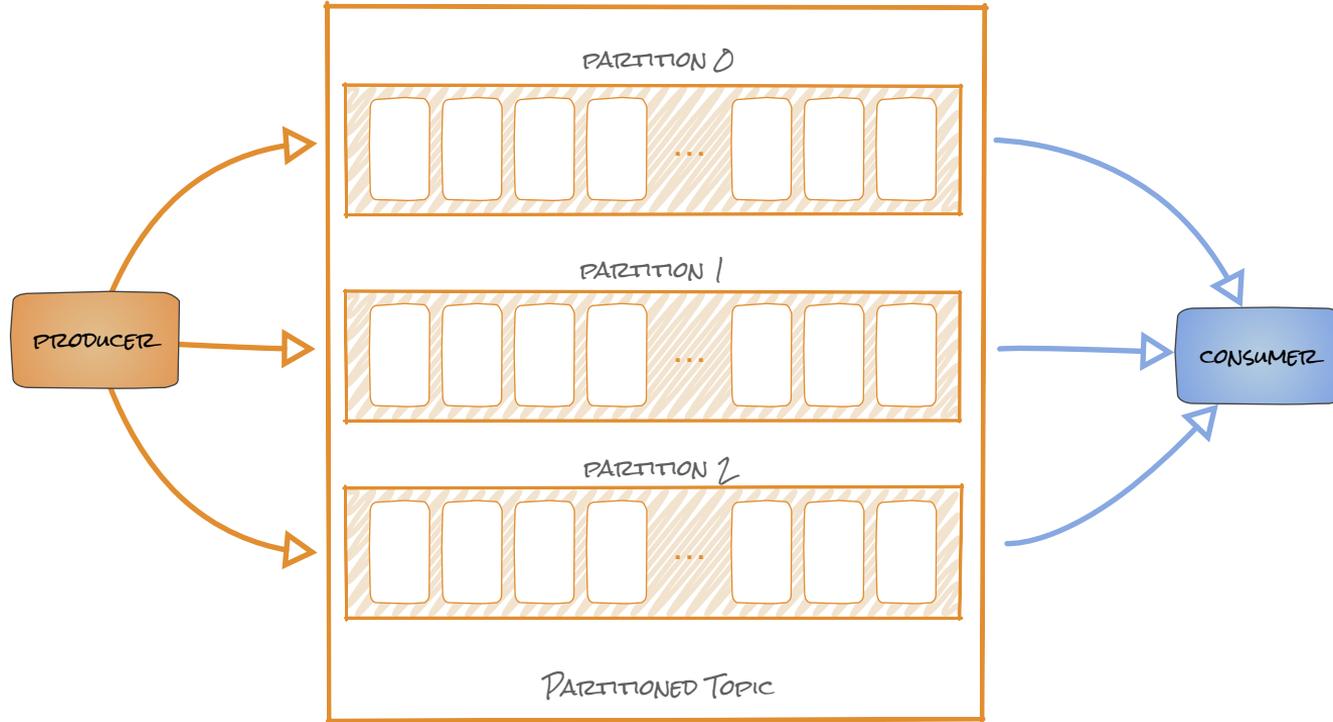
Streaming Platform



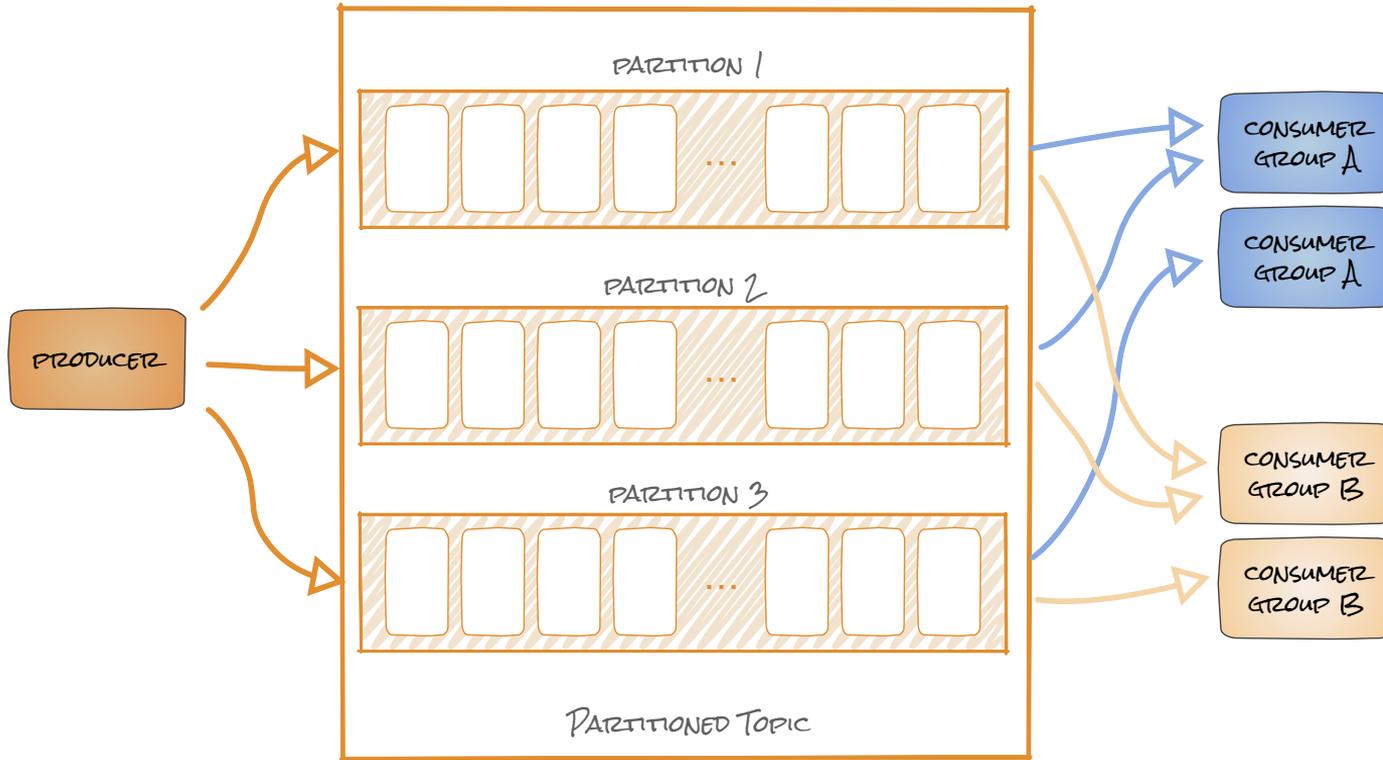
Basic Architecture Model



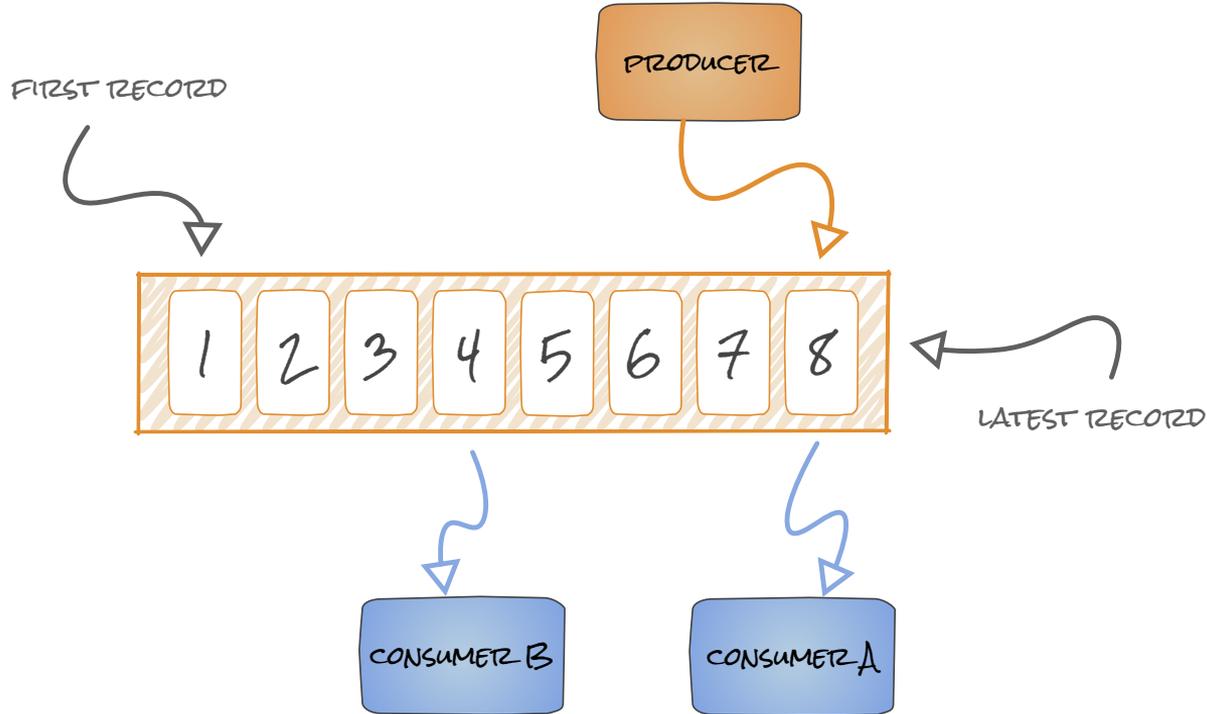
Partitioned Logs



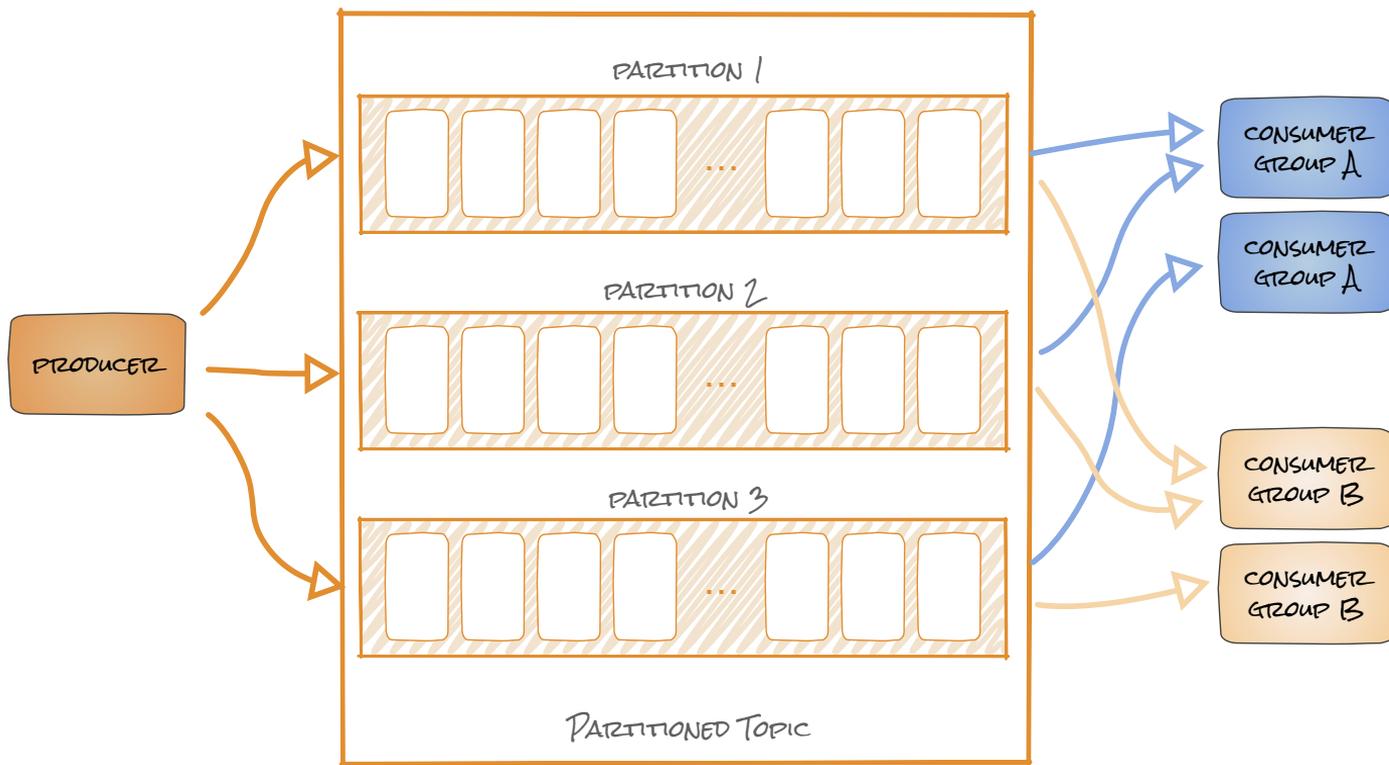
Scalable Consumption



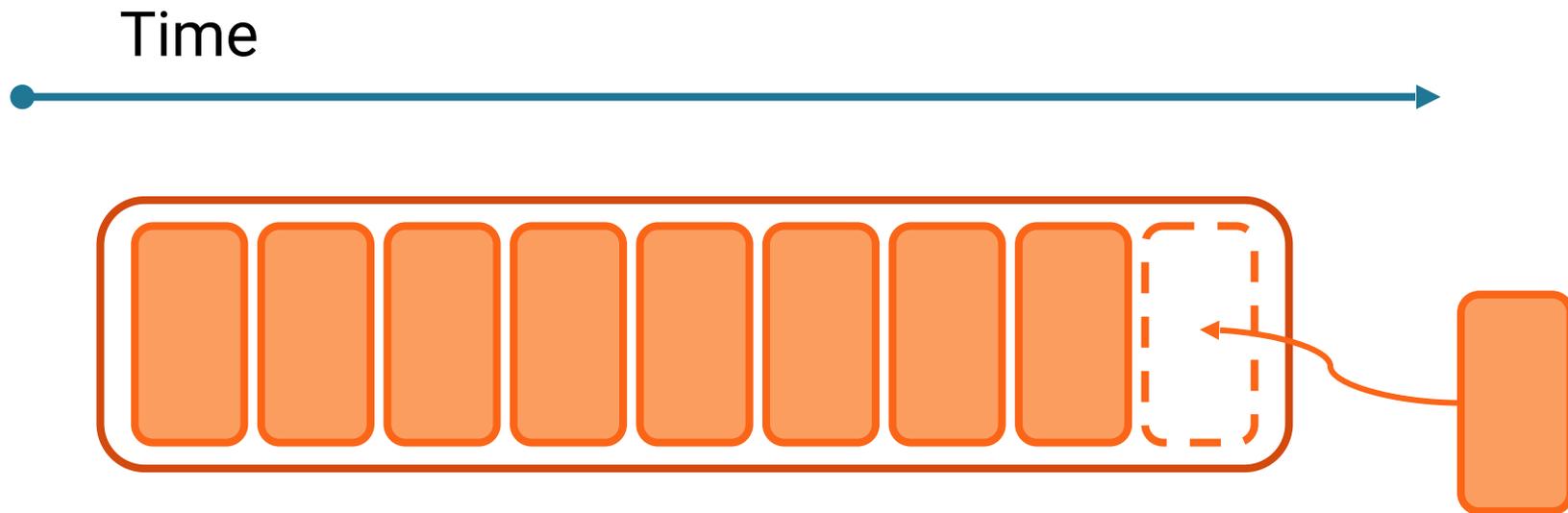
Logs and Pub/Sub



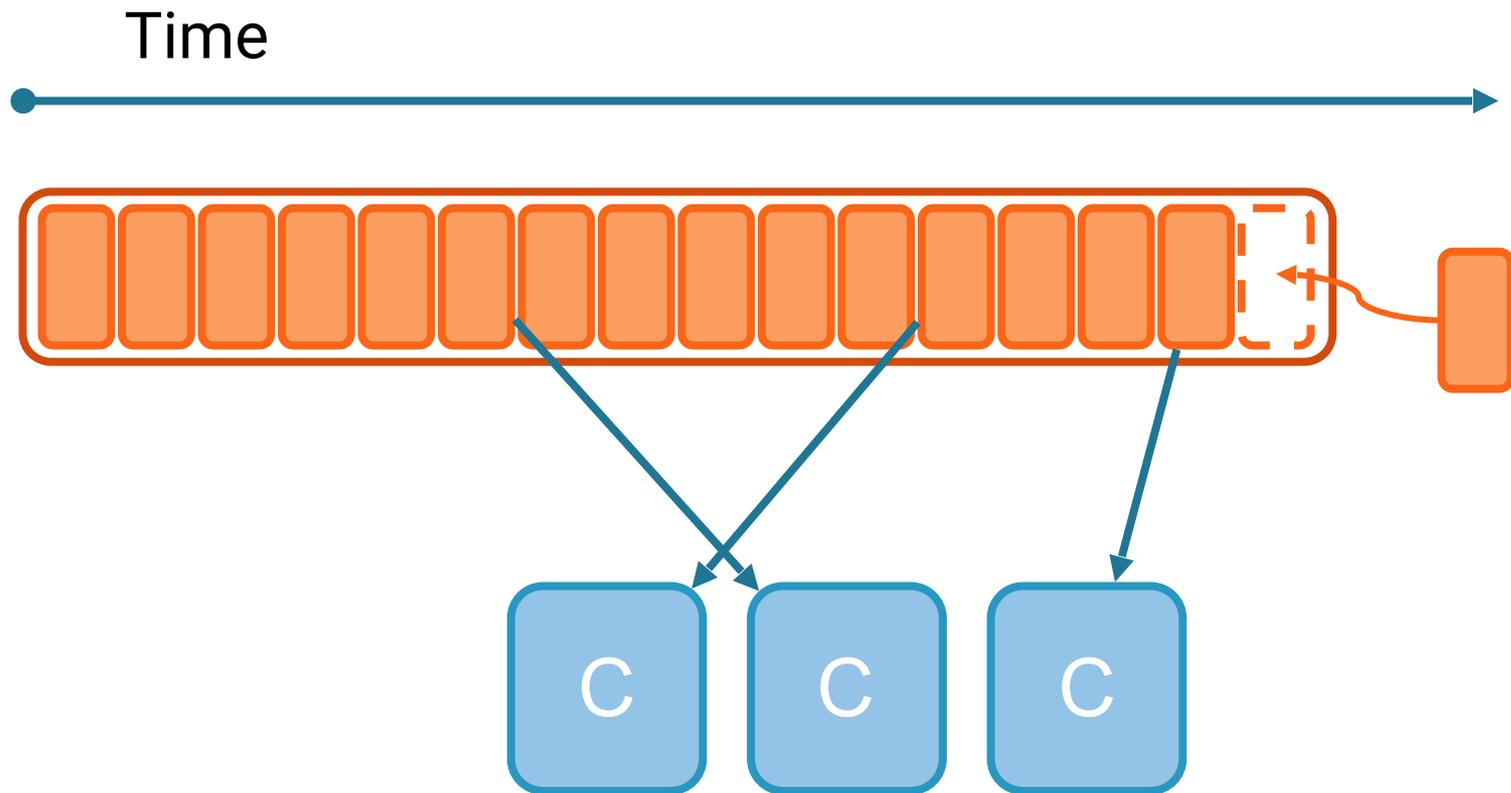
Scalable Consumption



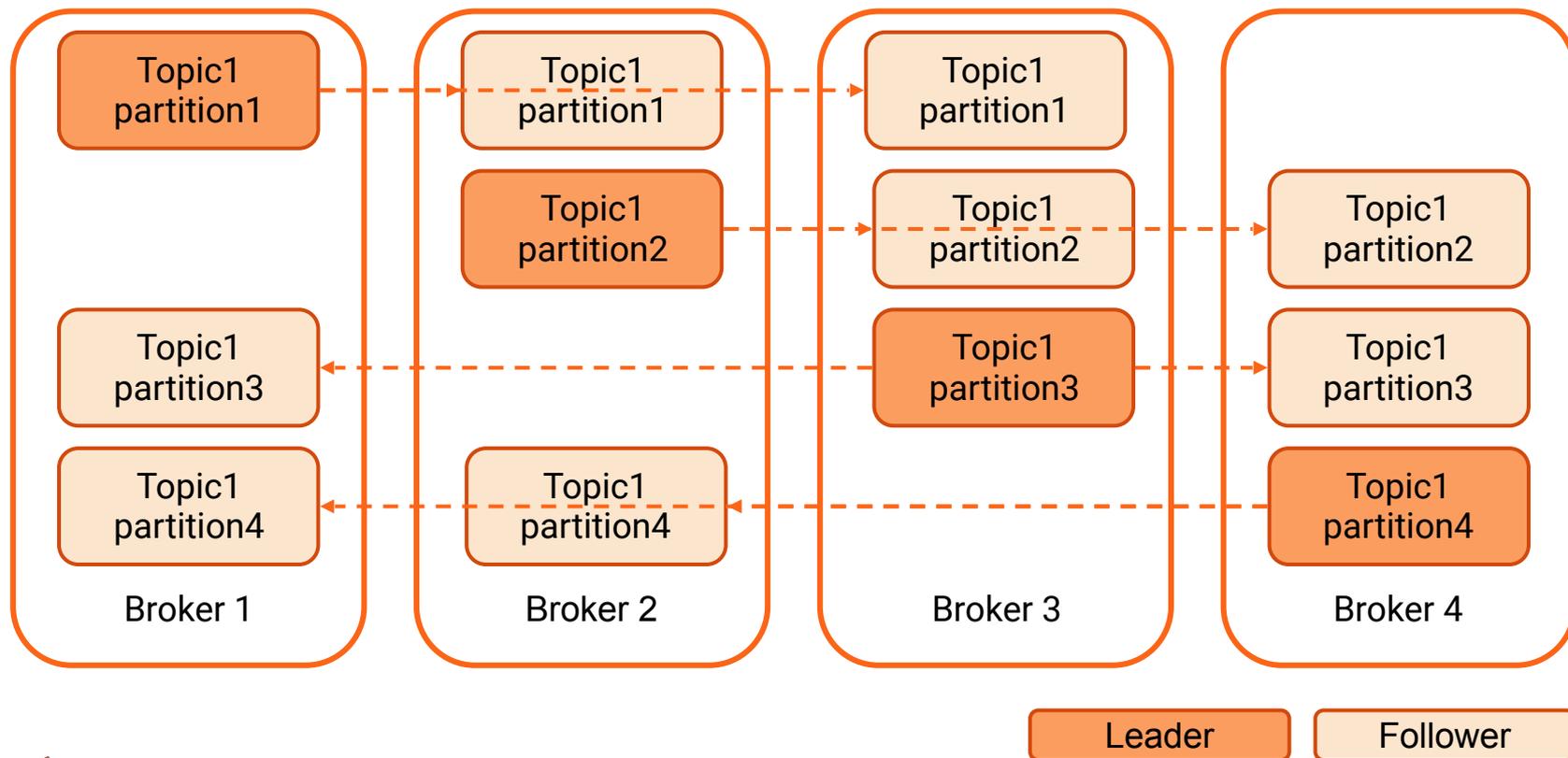
Producing to Kafka



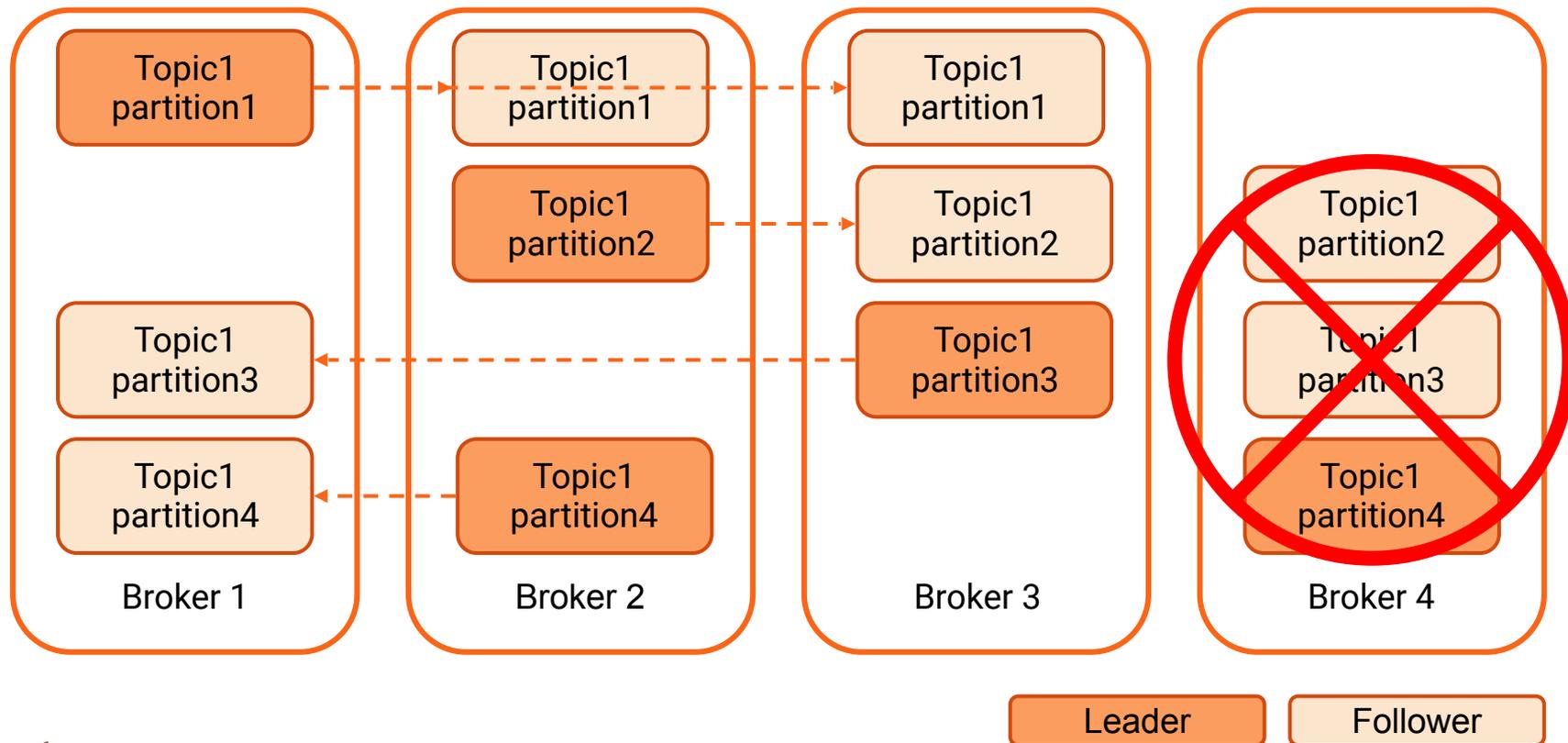
Producing to Kafka



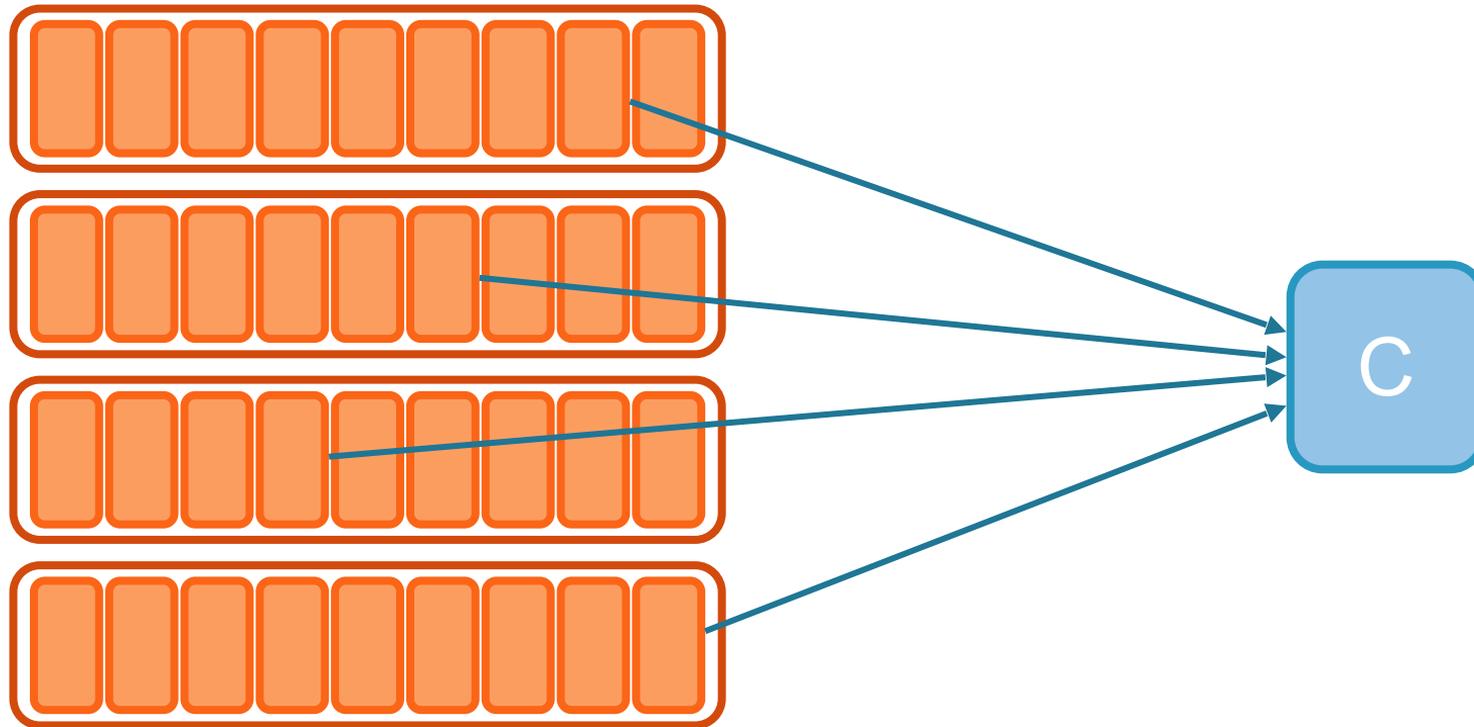
Partition Leadership and Replication



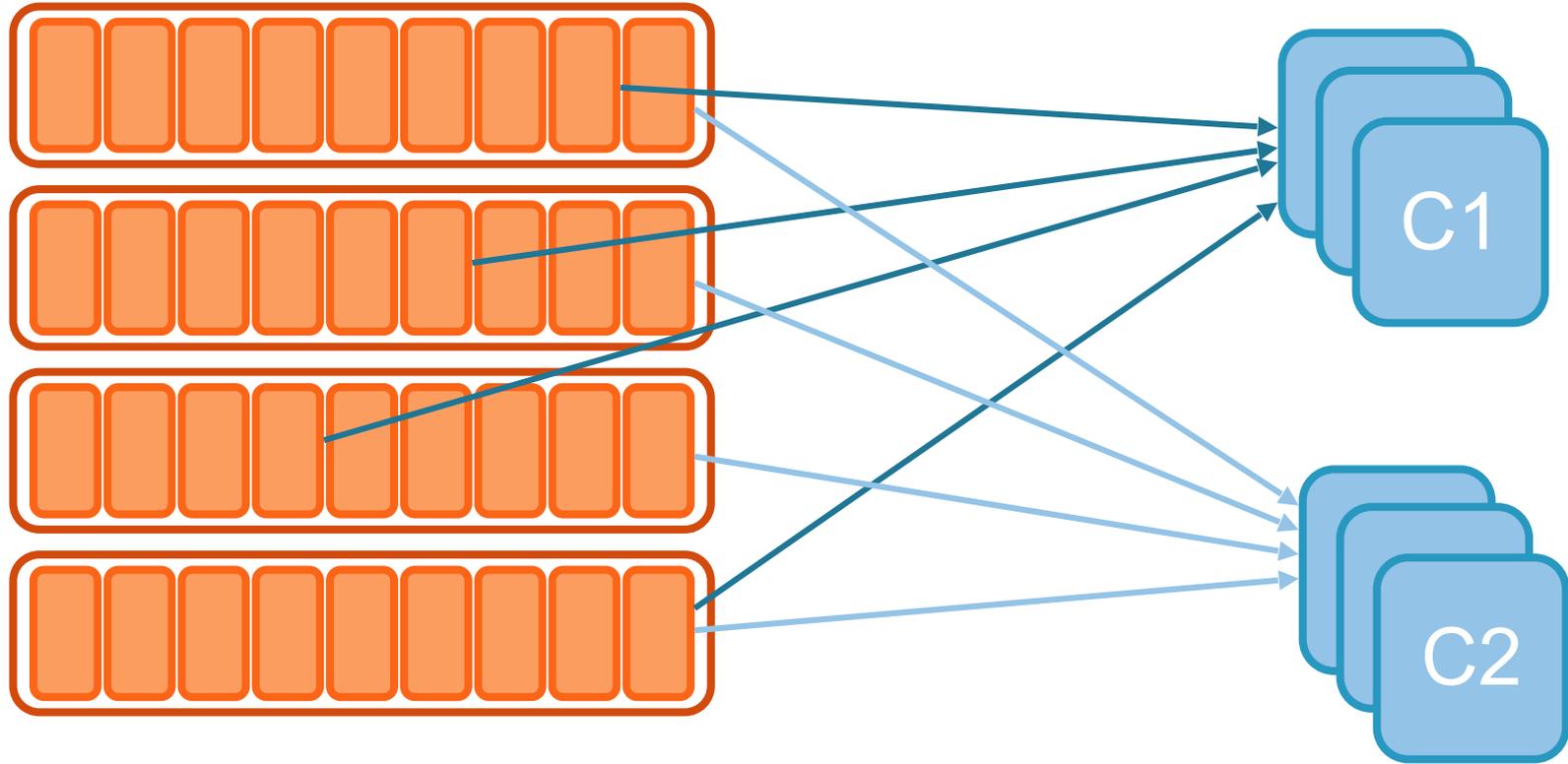
Partition Leadership and Replication - node failure



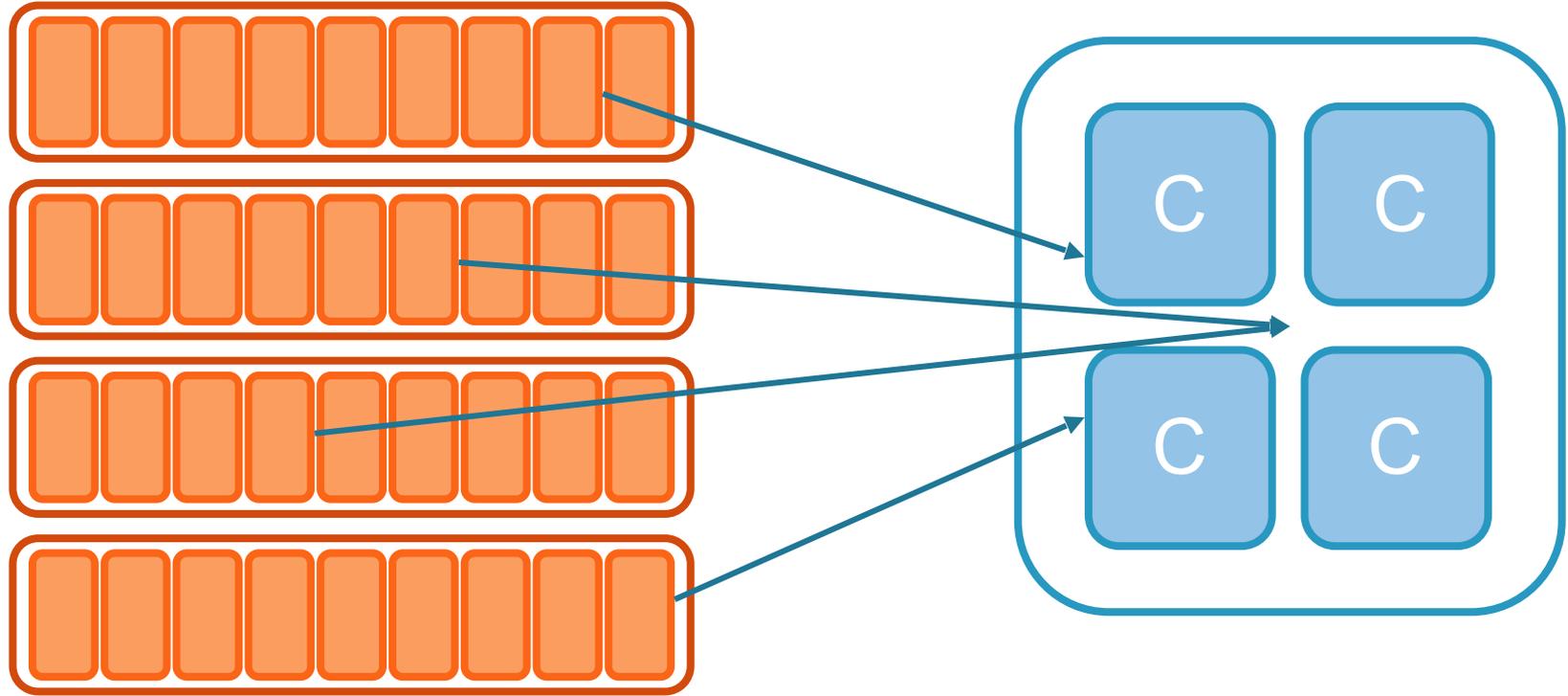
Consuming From Kafka - Single Consumer



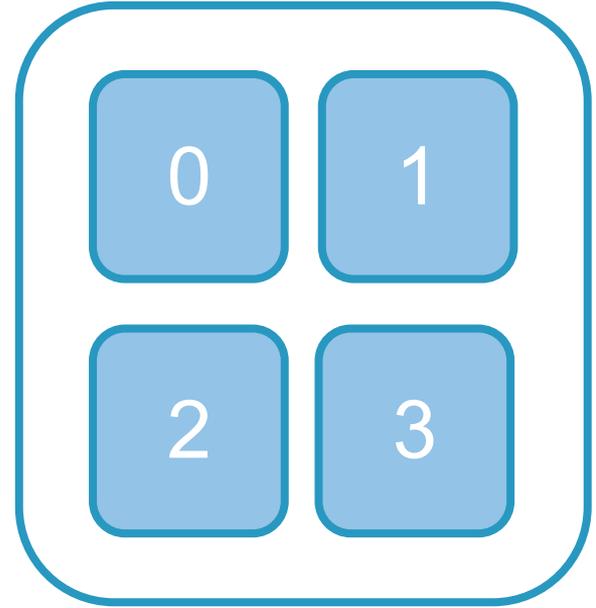
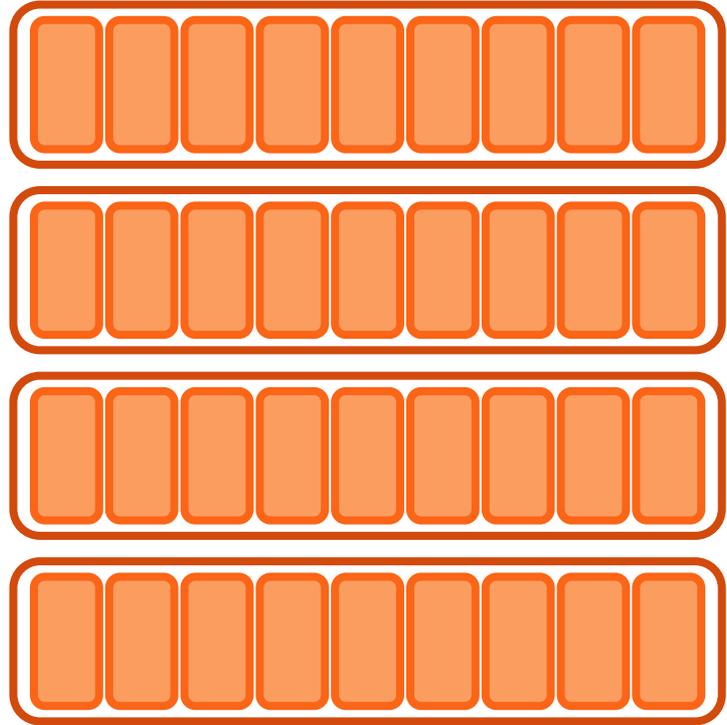
Consuming From Kafka - Grouped Consumers



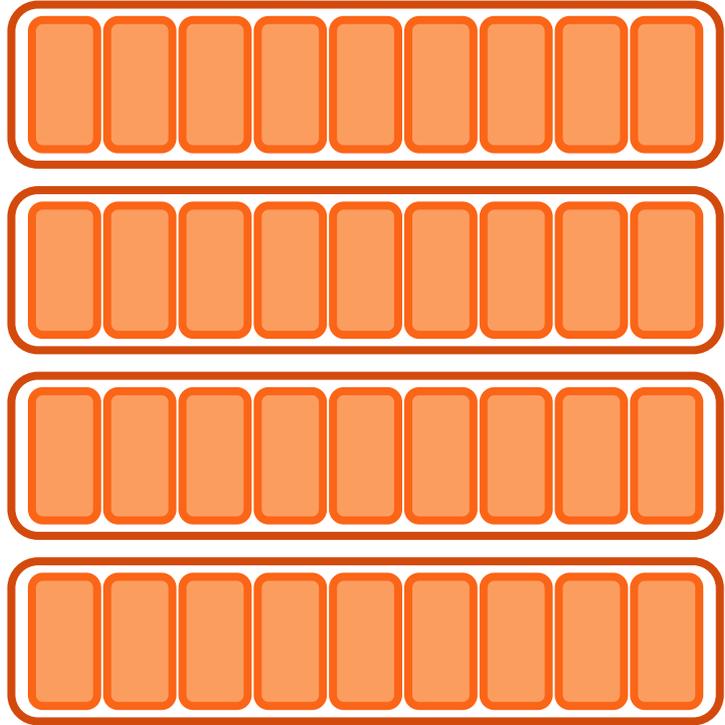
Consuming From Kafka - Grouped Consumers

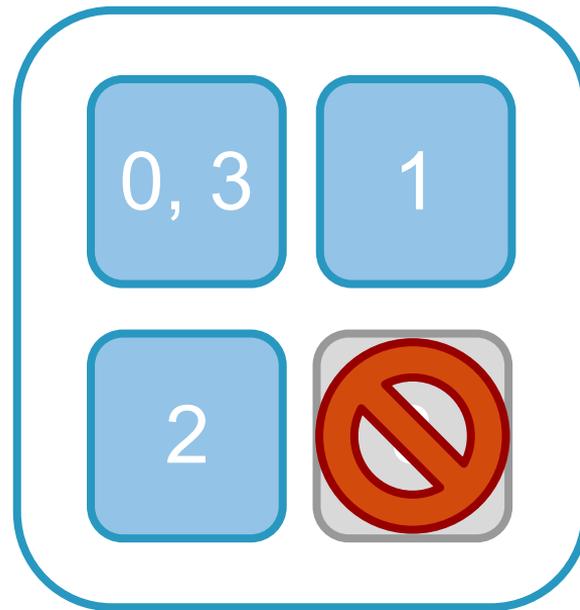
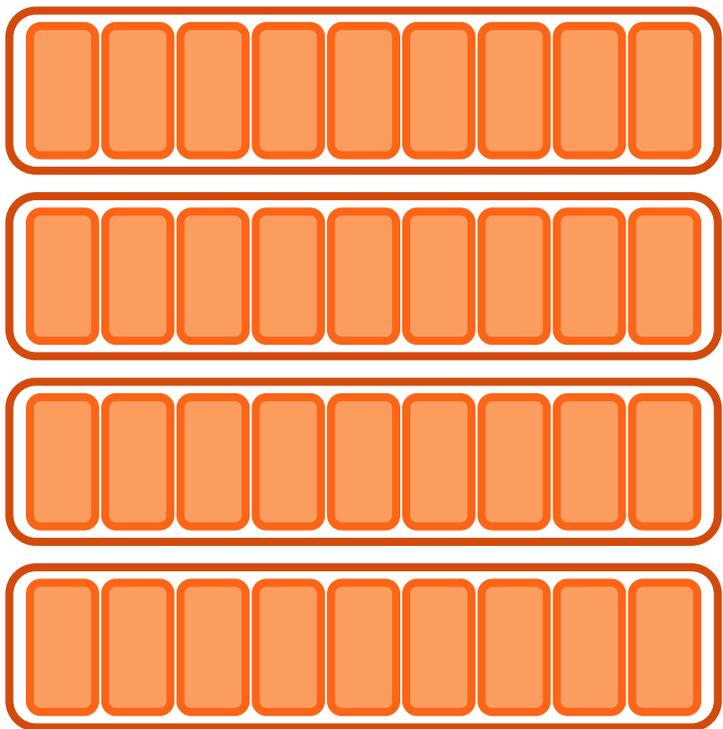


Consuming From Kafka - Grouped Consumers



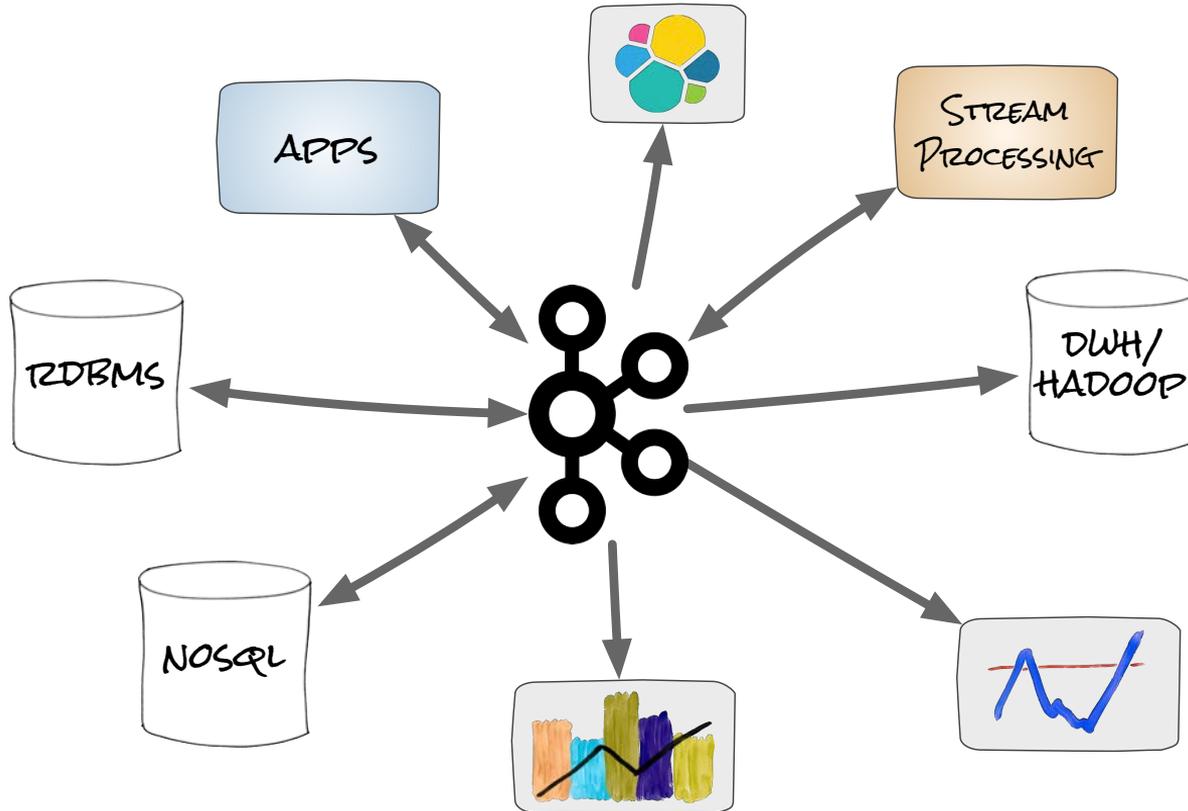
Consuming From Kafka - Grouped Consumers



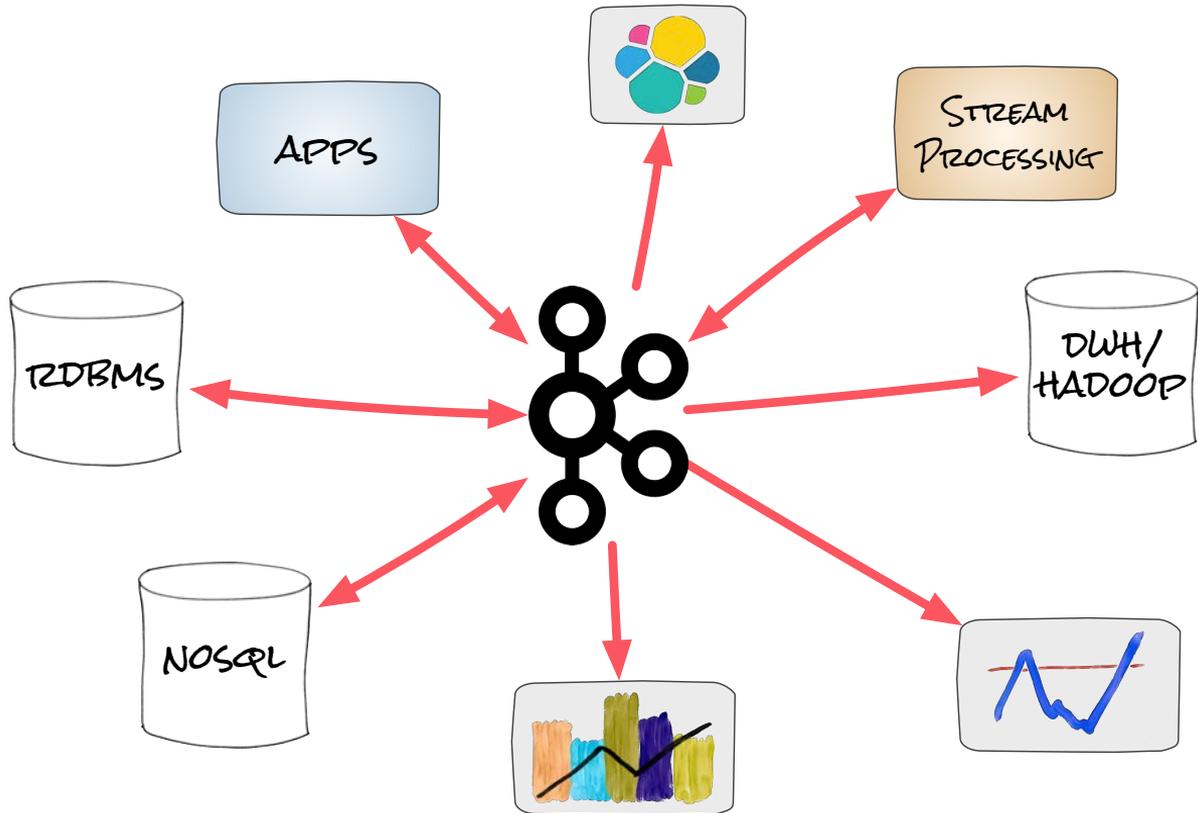




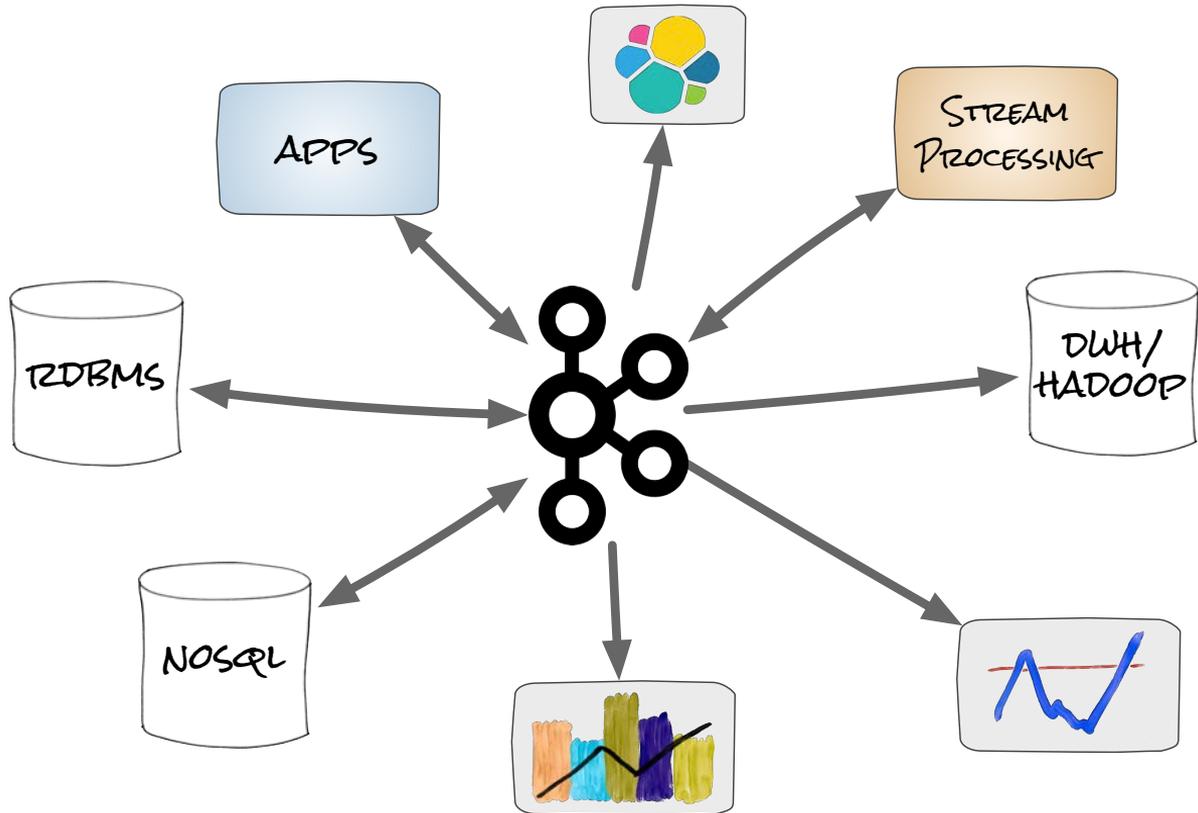
A Simpler Platform



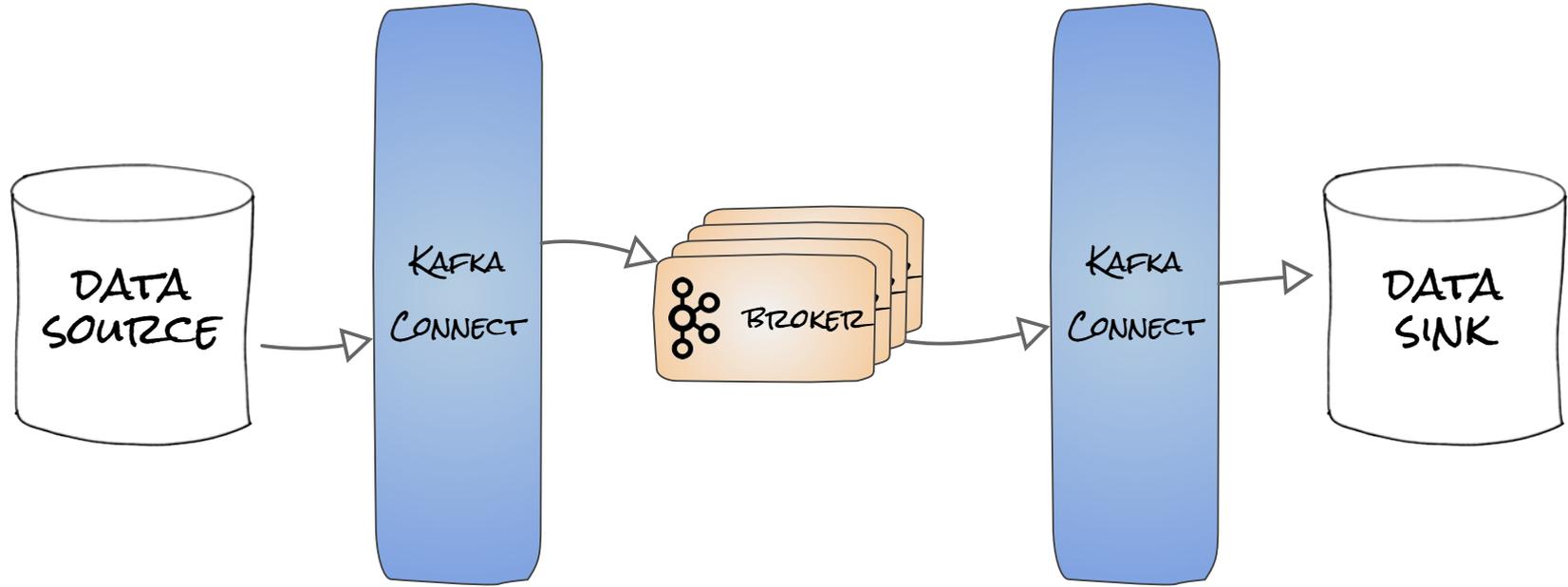
A Simpler Platform



A Simpler Platform



Kafka Connect

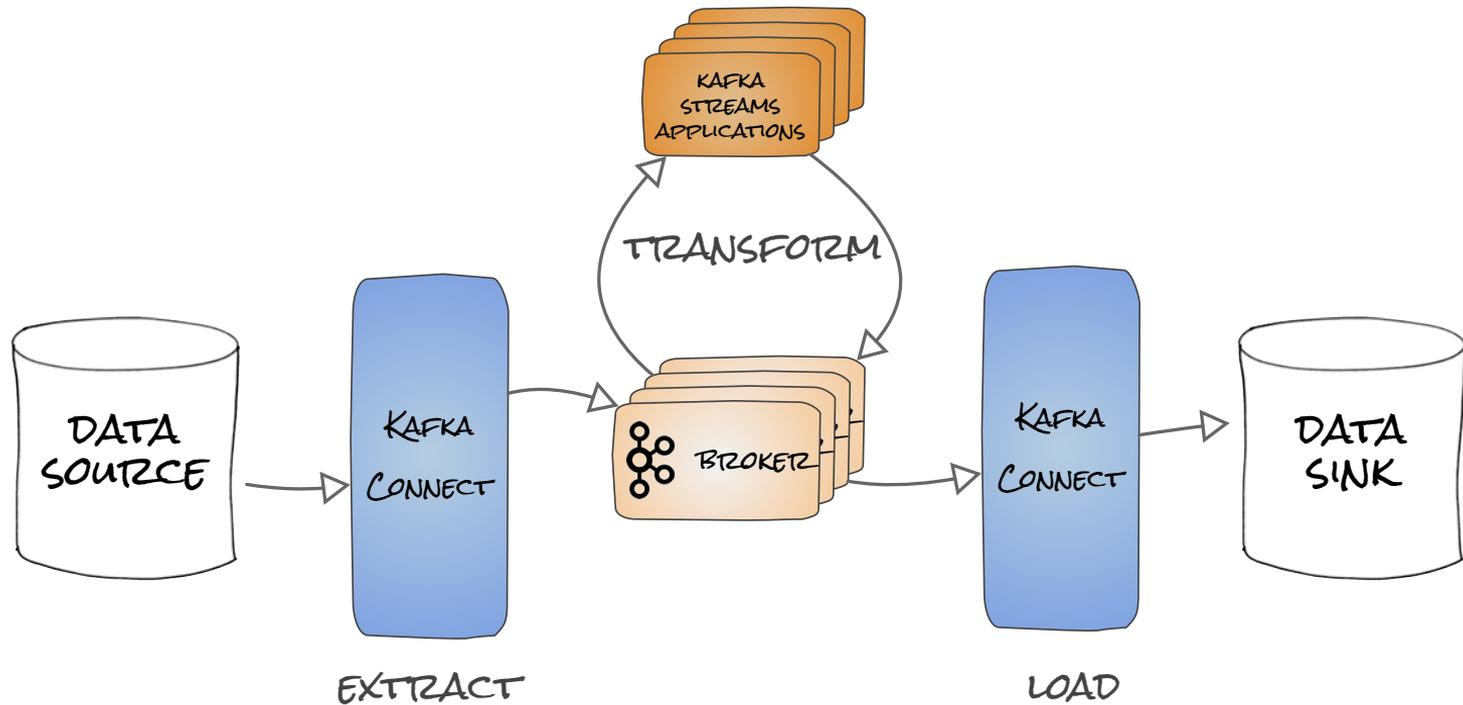


Kafka Connect

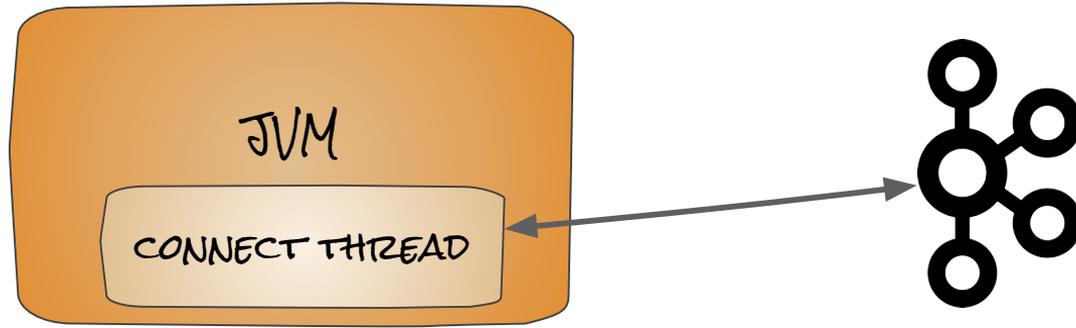
- a data integration framework
- scalable and fault-tolerant
- exactly-once delivery in many cases
- integrates Kafka with other data systems
- library of existing “connectors” for common data sources and sinks
- necessary component of modern streaming ETL systems



Streaming ETL



Standalone Mode



Distributed Mode

