

Streaming Data with Apache Kafka



Janani Ravi

Co-founder, Loonycorn

www.loonycorn.com

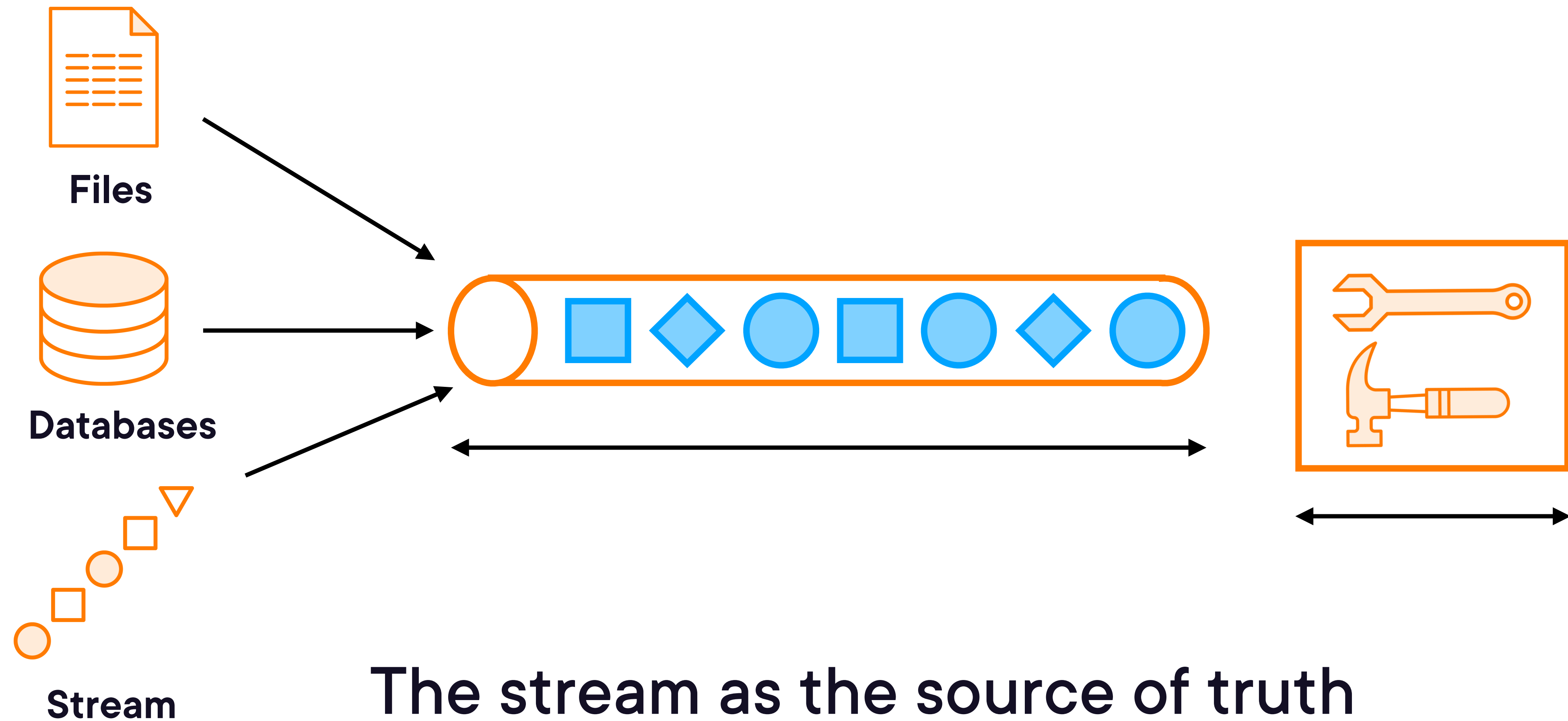




Introducing Apache Kafka



Stream-first Architecture

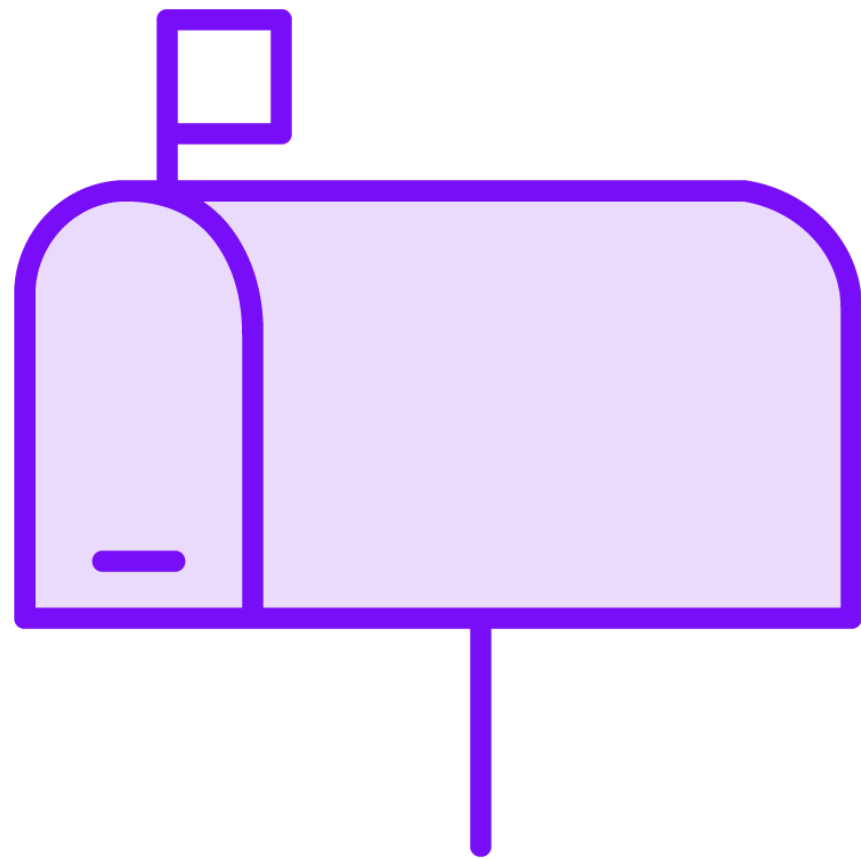


Apache Kafka

A distributed, highly scalable, elastic, fault-tolerant, and secure event streaming platform



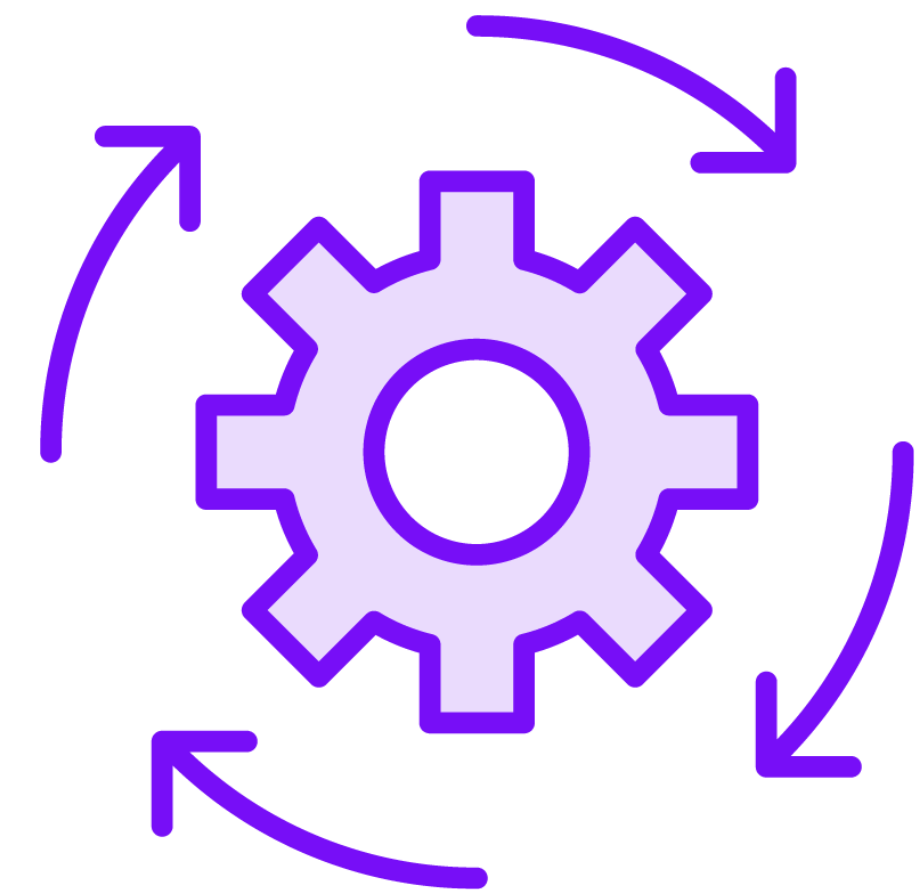
Spark Streaming



Publish and Subscribe
Write and read a stream
of events



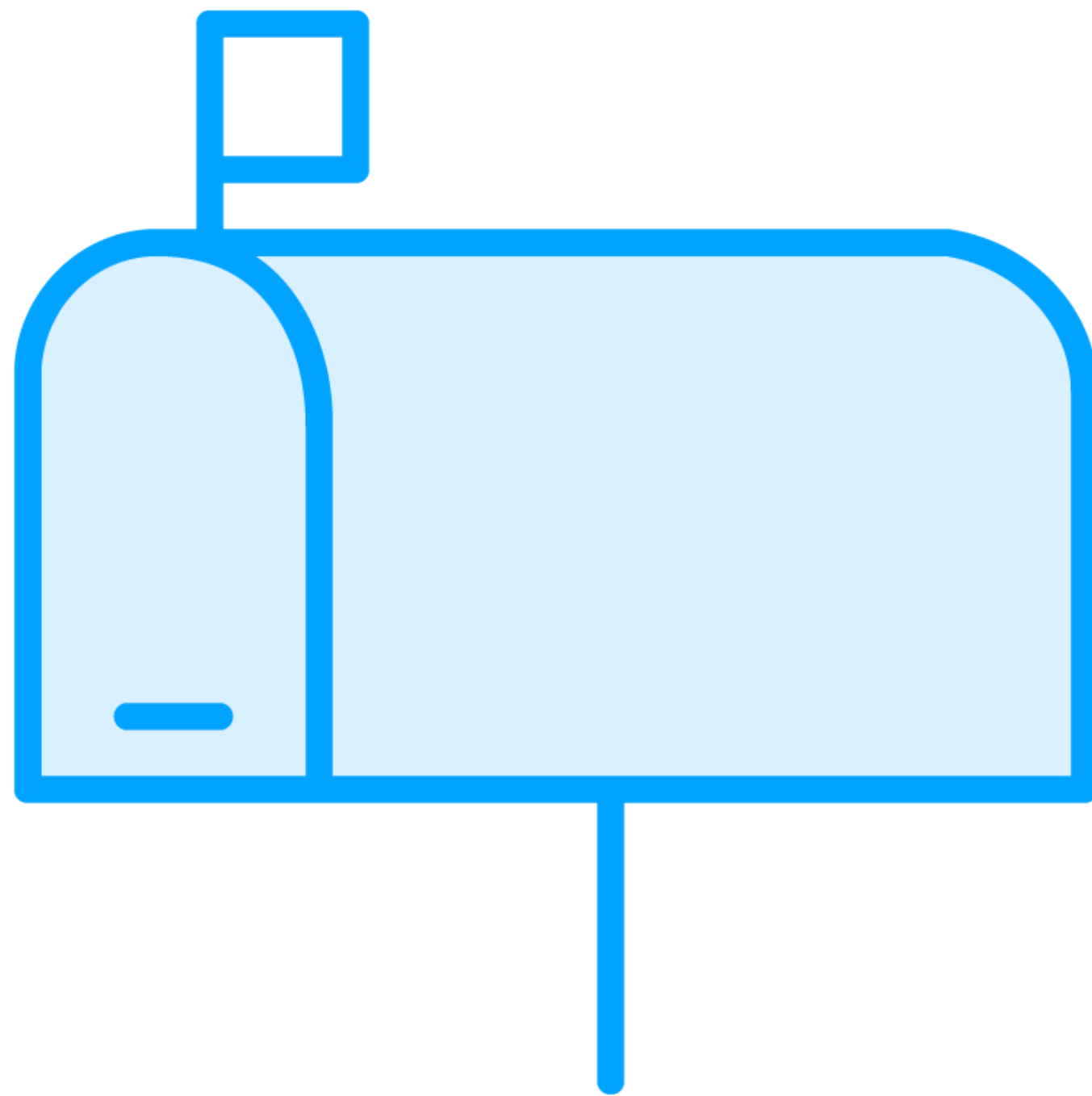
Store
Store events durably
and reliably



Process
Process as they occur
or retrospectively



Apache Kafka



Many-to-many asynchronous messaging

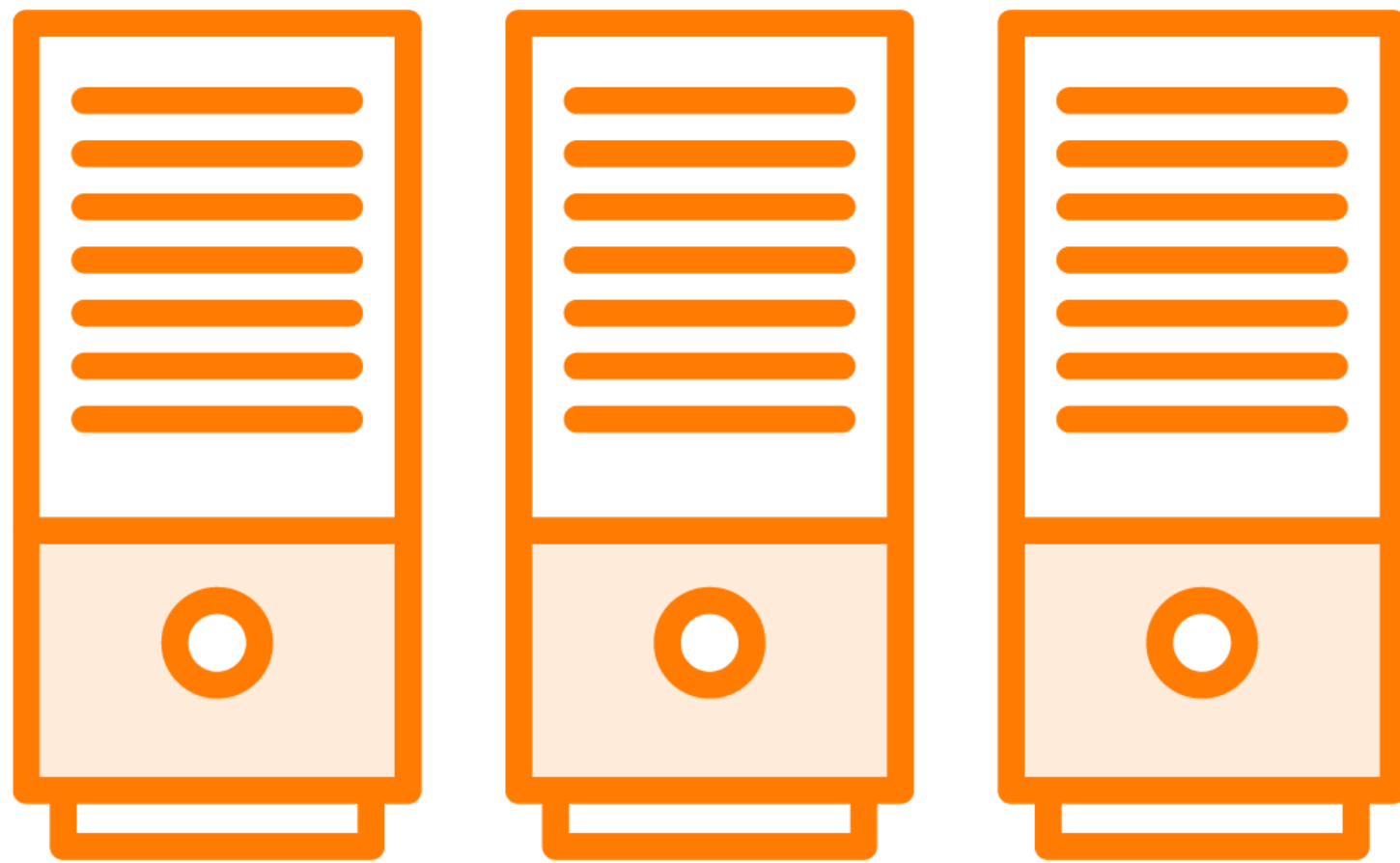
Decouples senders and receivers

Reliable, scalable, secure

Fault-tolerant, durable record storage



Servers and Brokers



Kafka is a distributed system

Runs on a **cluster of servers**

Some of these servers form the storage layer
and are called **brokers**



Publishers, Topics, and Consumers

Publisher



Topic



Consumer



Messages



Publishers, Topics, and Consumers

Producers



Topics



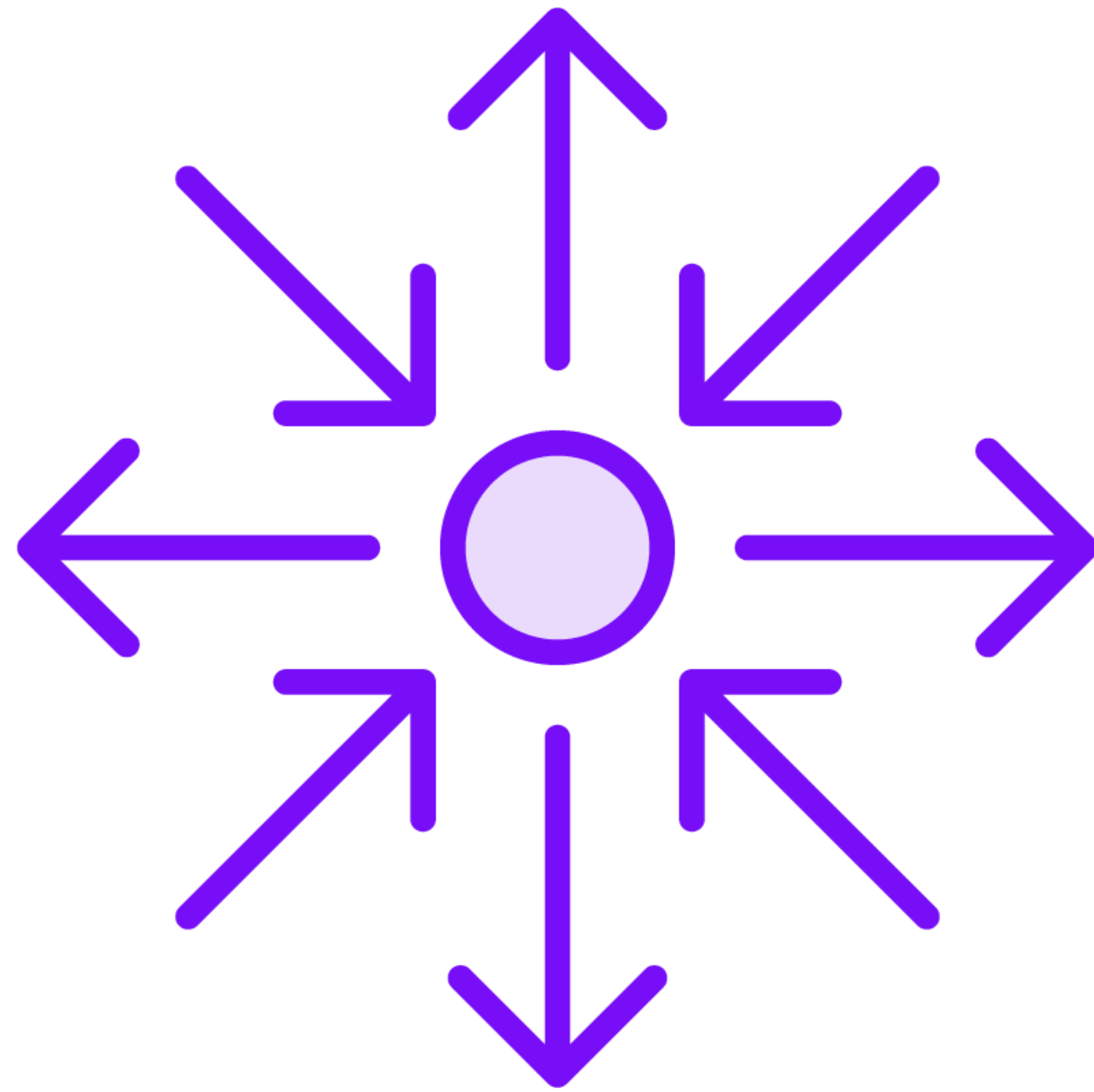
Consumers



Records



High Availability and Fault Tolerance



High availability using ZooKeeper for managing cluster leaders and followers

Fault tolerance using partitioned and replicated topics

