
confluent

@gamussa @NYJavaSig @confluentinc

APACHE kafka

spring by Pivotal™

confluent

# The agenda

- Quick intro to Streaming Platform
- Gentle intro to Stream processing
- Wire everything nicely with Spring

https://cnfl.io/streams-movie

# Who am I?

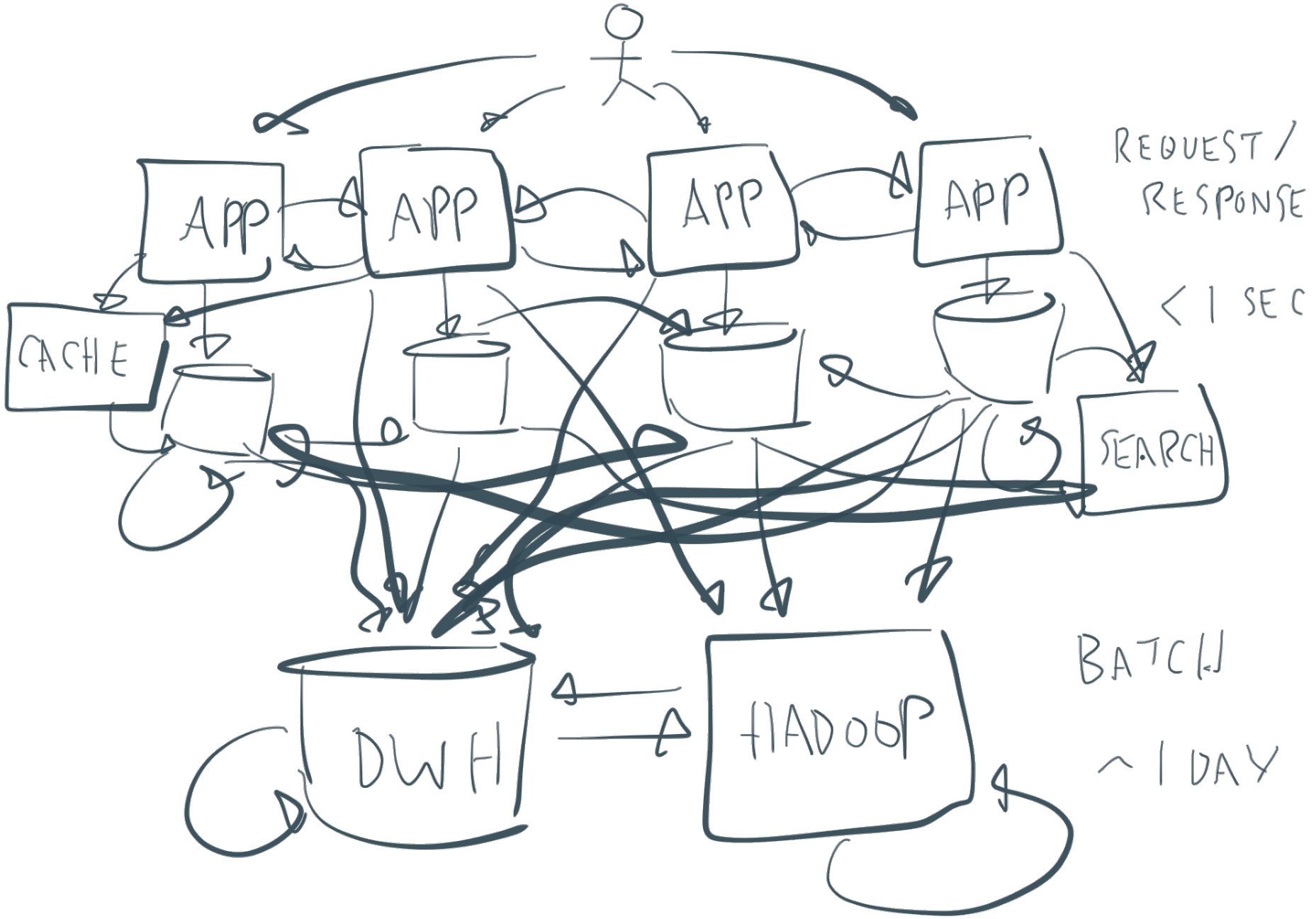
confluent

Solutions Architect
Developer Advocate

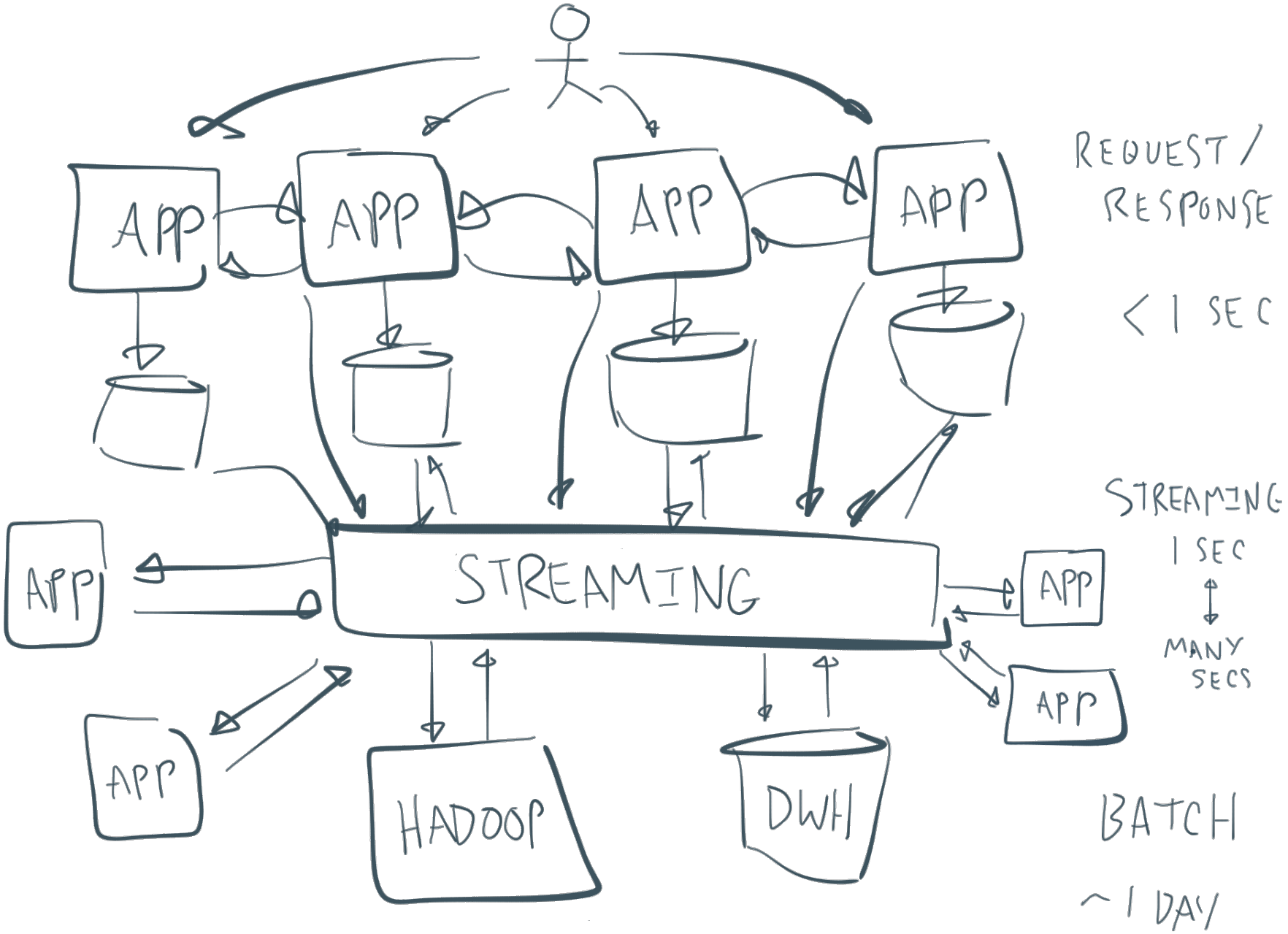
@gamussa in internetz


Hey you, yes, you,
go follow me in twitter ©

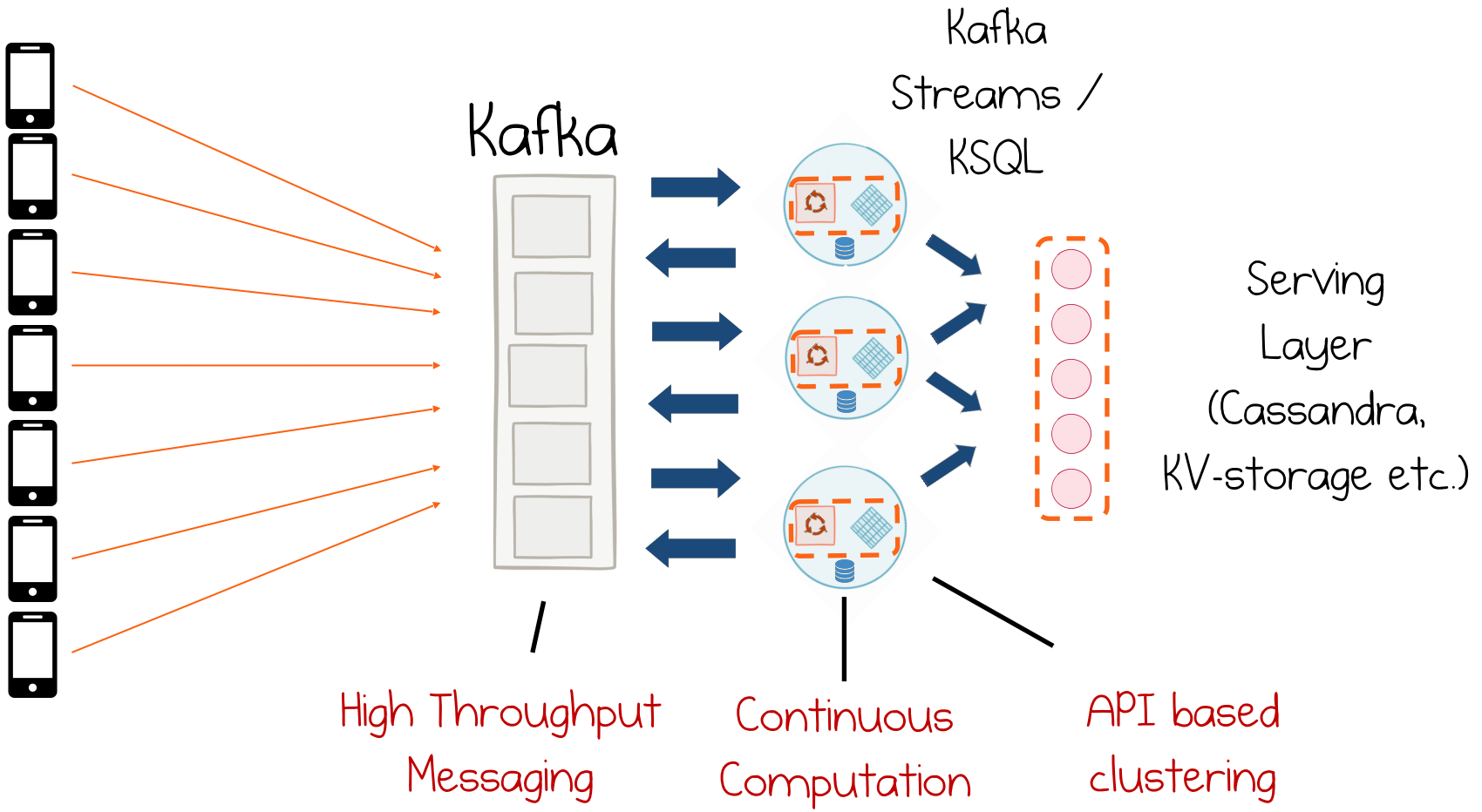
# Kafka & Confluent



APP
APP
APP
APP
REQUEST / RESPONSE
< 1 SEC
CACHE
SEARCH
DWH
HADOOP
BATCH
~ 1 DAY

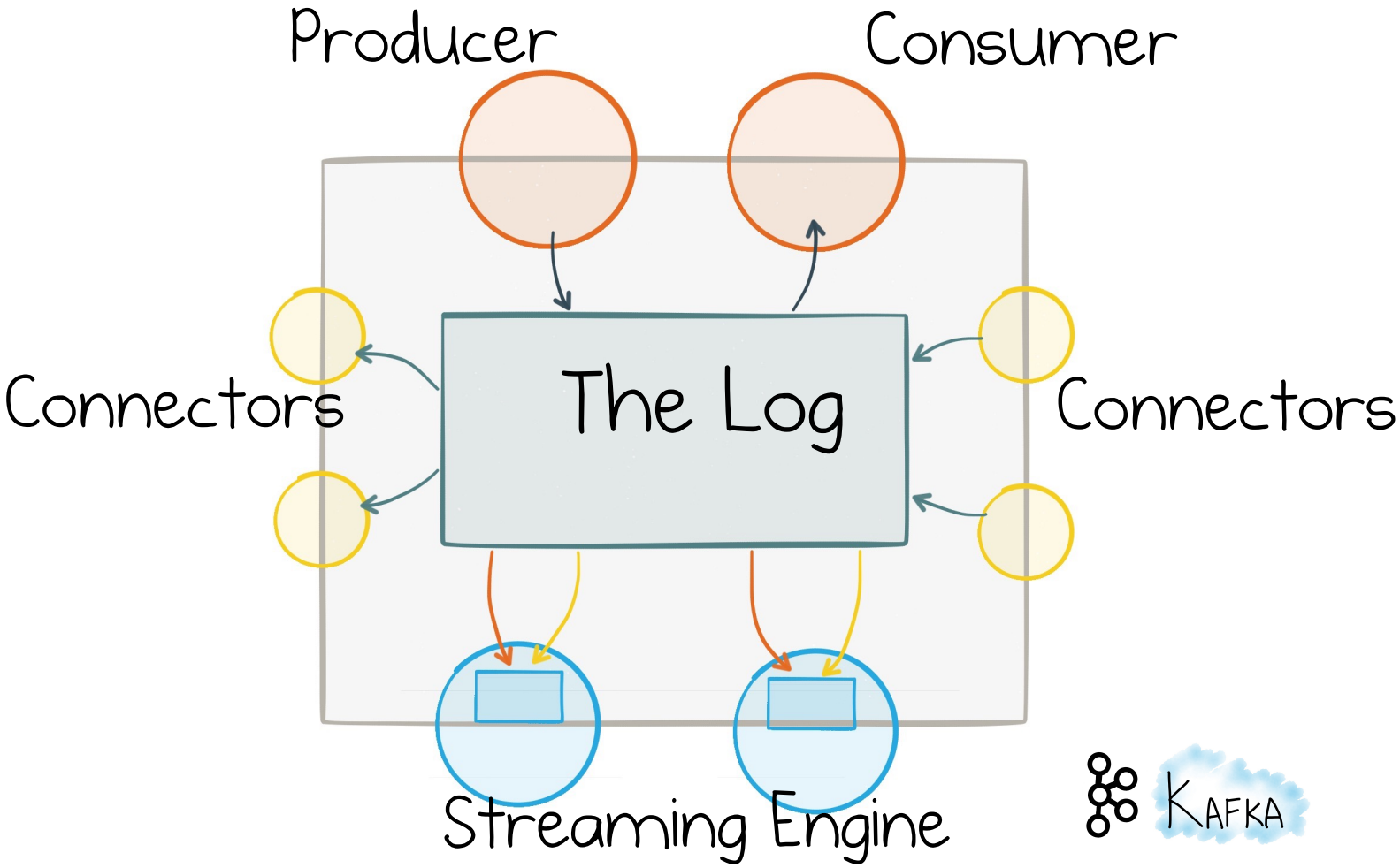

HIVE
PE PE SILVA
APACHE Spark
STORM
hadoop
Map Reduce
kafka


APP
APP
APP
APP
APP
STREAMING
APP
APP
APP
HADOOP
DWH
REQUEST / RESPONSE
< 1 SEC
STREAMING
1 SEC
MANY SECS
BATCH
~ 1 DAY

# Origins in Stream Processing

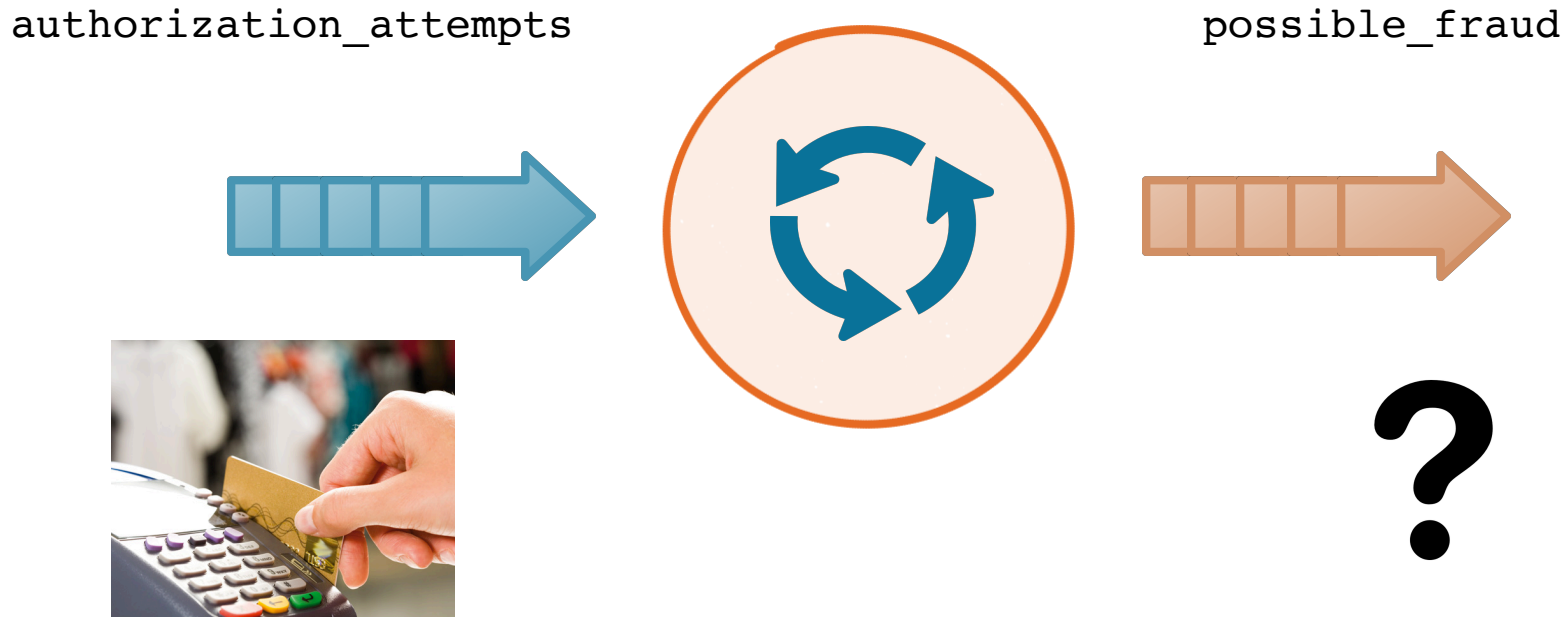


# Kafka is a Streaming Platform

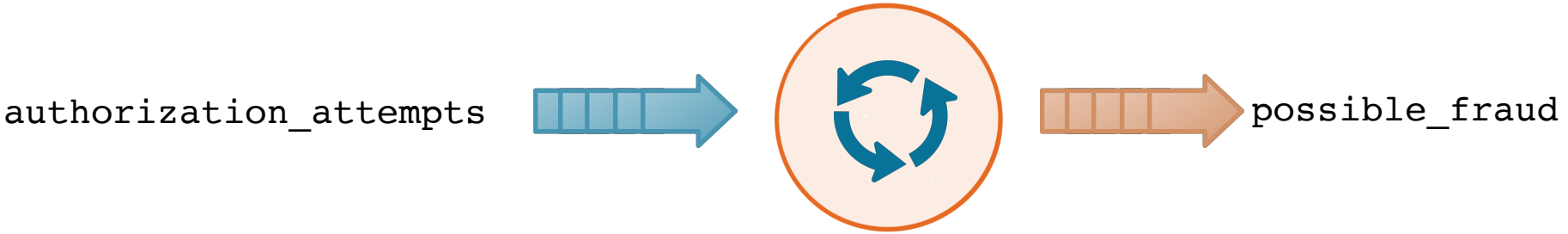


Streaming

# What exactly is Stream Processing?

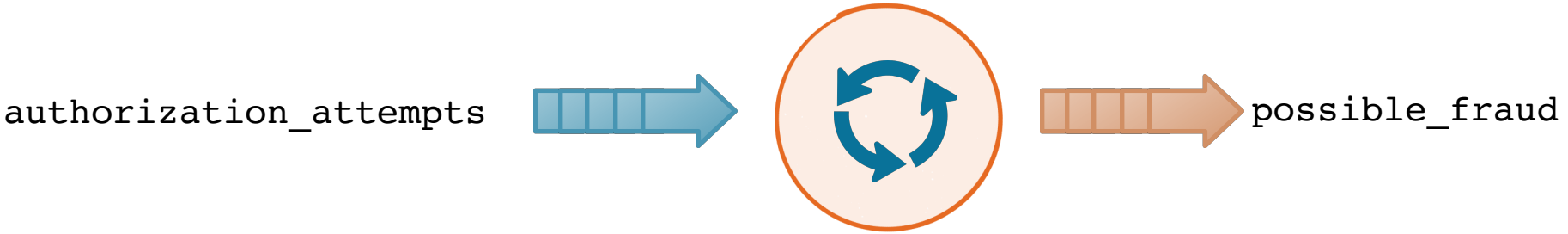


# What exactly is Stream Processing?



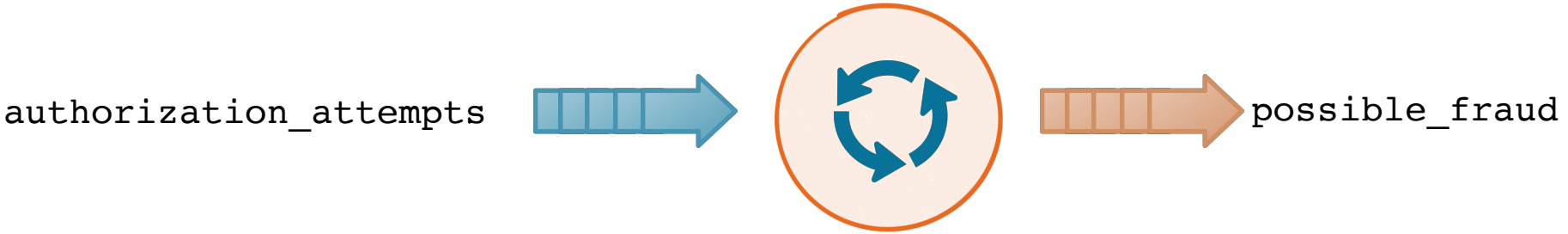

```
CREATE STREAM possible_fraud AS
SELECT card_number, count(*)
FROM authorization_attempts
WINDOW TUMBLING (SIZE 5 MINUTE)
GROUP BY card_number
HAVING count(*) > 3;
```

# What exactly is Stream Processing?



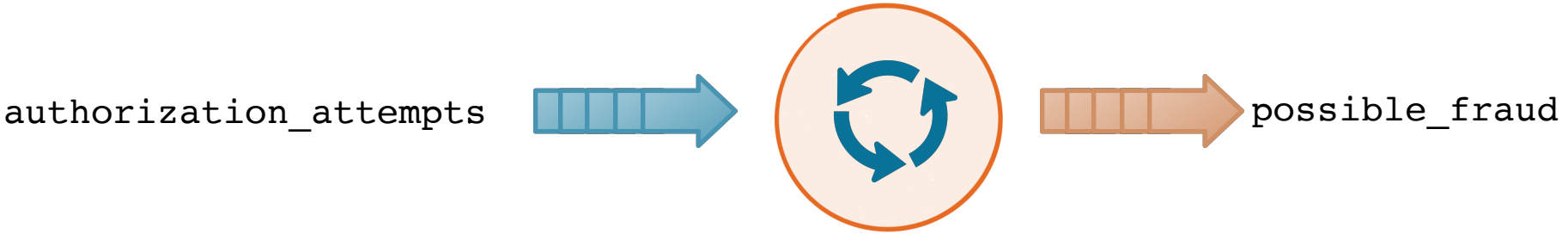

```
CREATE STREAM possible_fraud AS
SELECT card_number, count(*)
FROM authorization_attempts
WINDOW TUMBLING (SIZE 5 MINUTE)
GROUP BY card_number
HAVING count(*) > 3;
```

# What exactly is Stream Processing?



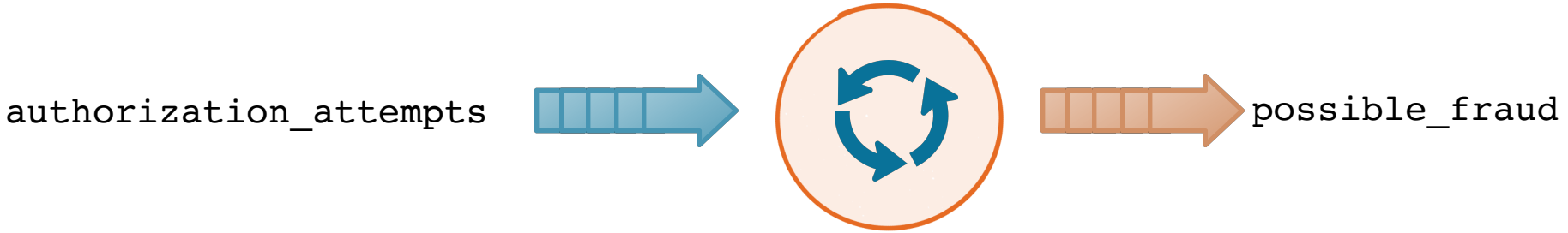

```
CREATE STREAM possible_fraud AS
SELECT card_number, count(*)
FROM authorization_attempts
WINDOW TUMBLING (SIZE 5 MINUTE)
GROUP BY card_number
HAVING count(*) > 3;
```

# What exactly is Stream Processing?



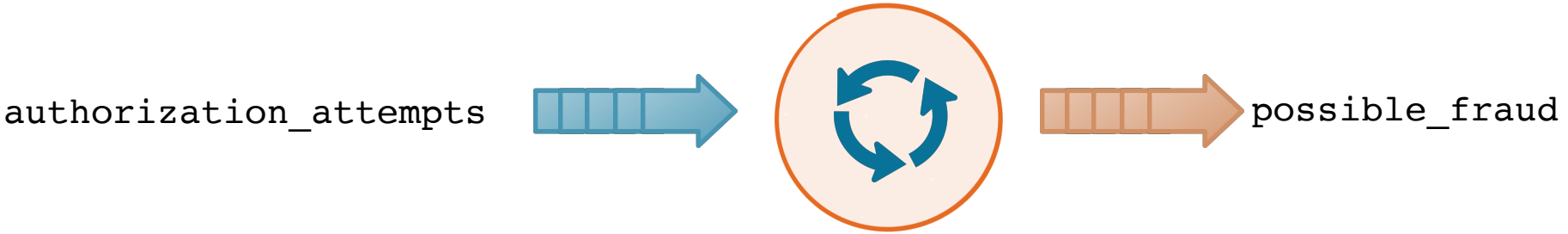

```
CREATE STREAM possible_fraud AS
SELECT card_number, count(*)
FROM authorization attempts
WINDOW TUMBLING (SIZE 5 MINUTE)
GROUP BY card_number
HAVING count(*) > 3;
```

# What exactly is Stream Processing?



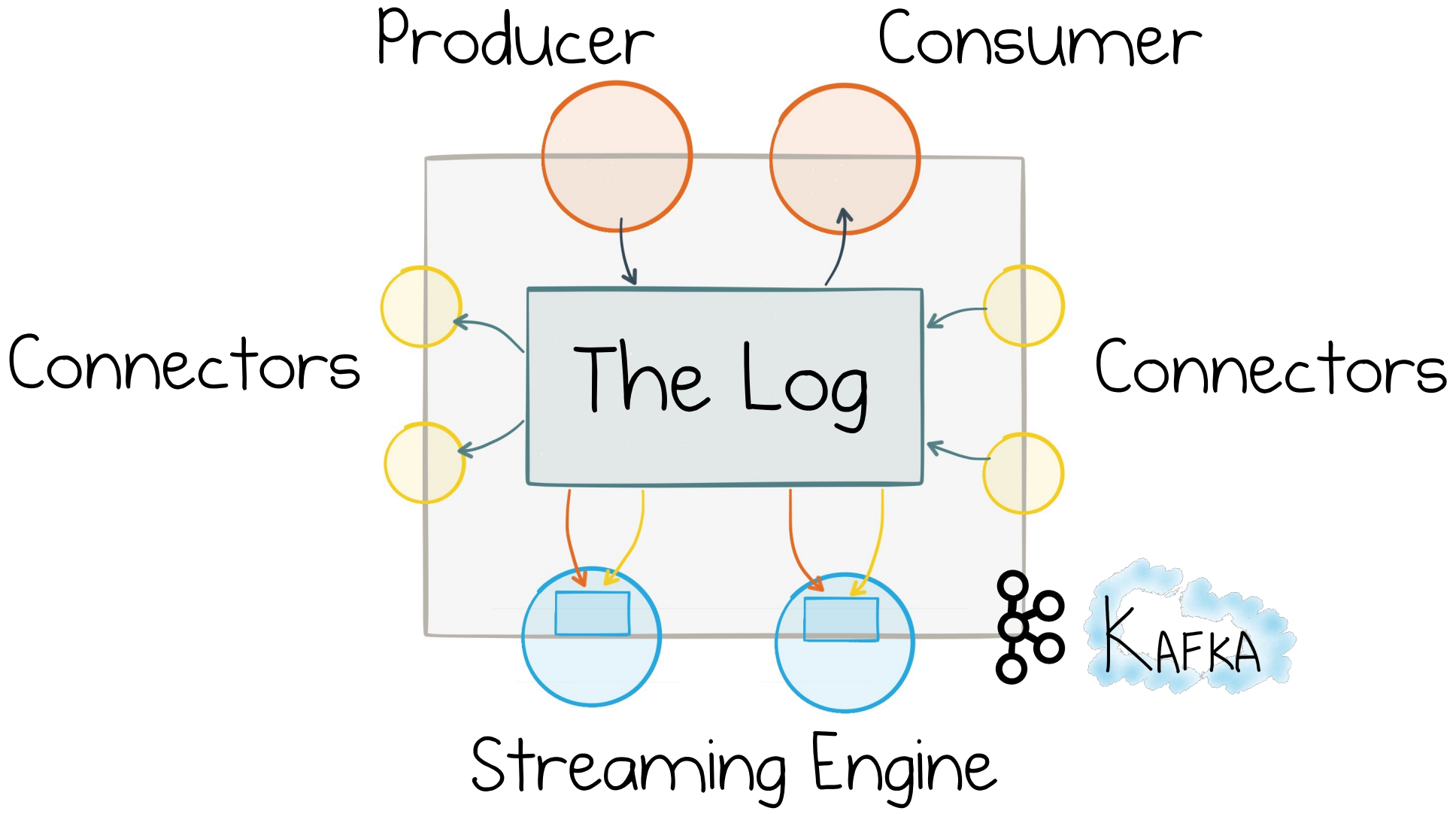

```
CREATE STREAM possible_fraud AS
SELECT card_number, count(*)
FROM authorization_attempts
WINDOW TUMBLING (SIZE 5 MINUTE)
GROUP BY card_number
HAVING count(*) > 3;
```

# What exactly is Stream Processing?

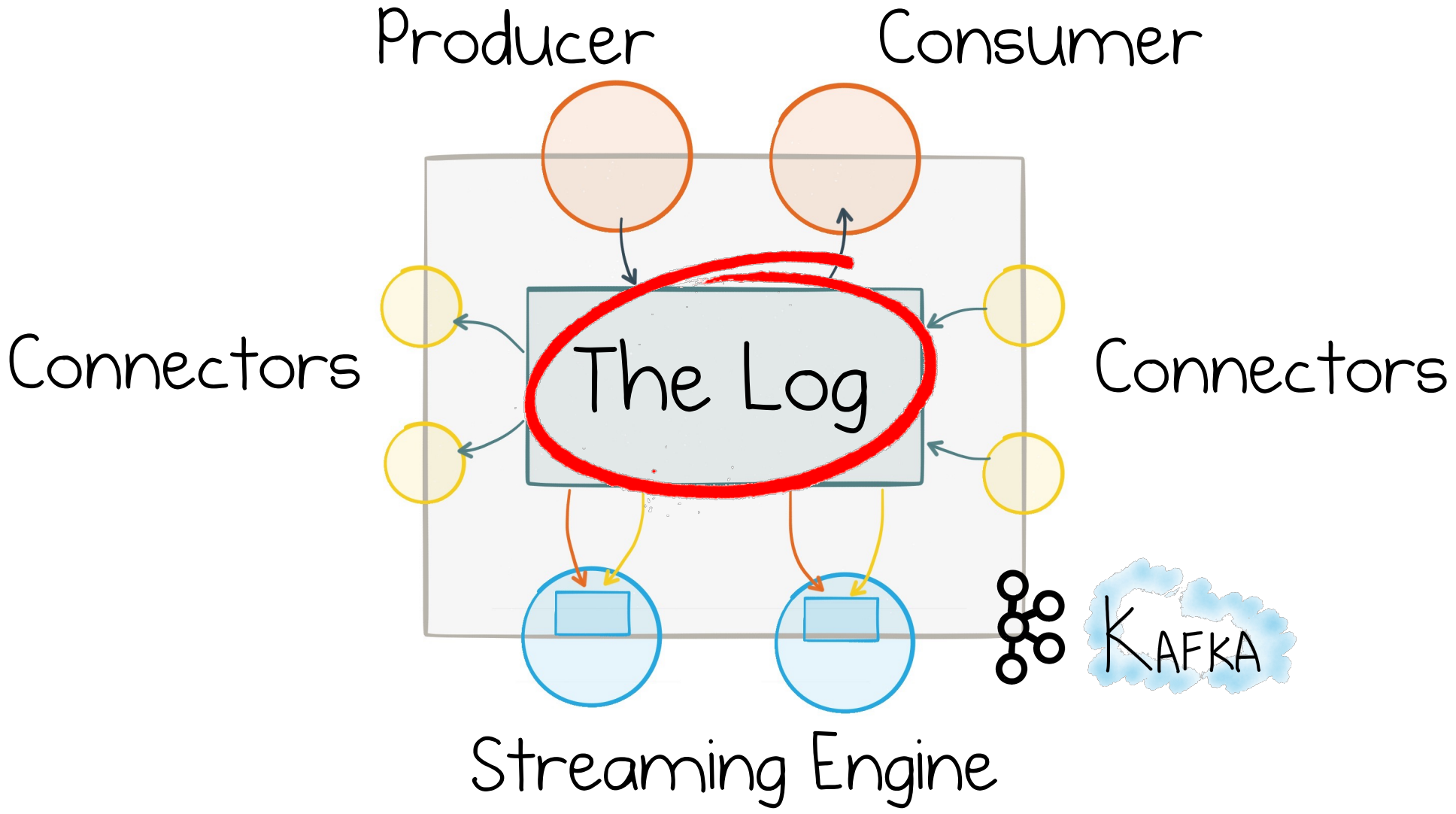



```
CREATE STREAM possible_fraud AS
SELECT card_number, count(*)
FROM authorization_attempts
WINDOW TUMBLING (SIZE 5 MINUTE)
GROUP BY card_number
HAVING count(*) > 3;
```

# What is a Streaming Platform?



# Kafka's Distributed Log



# The log - durable messaging system

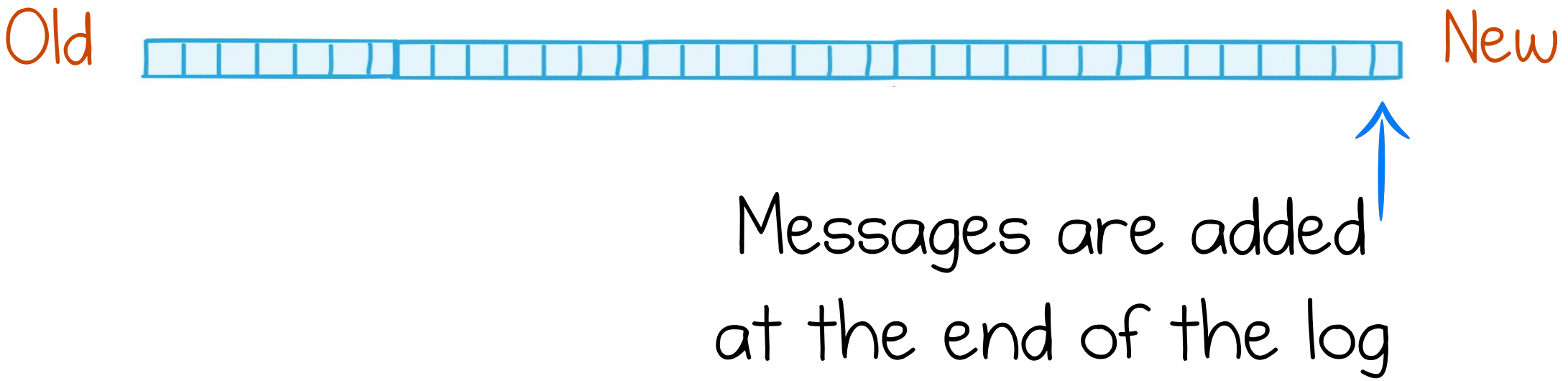
Similar to a traditional messaging system (ActiveMQ, Rabbit) but with:

(a) Far better scalability

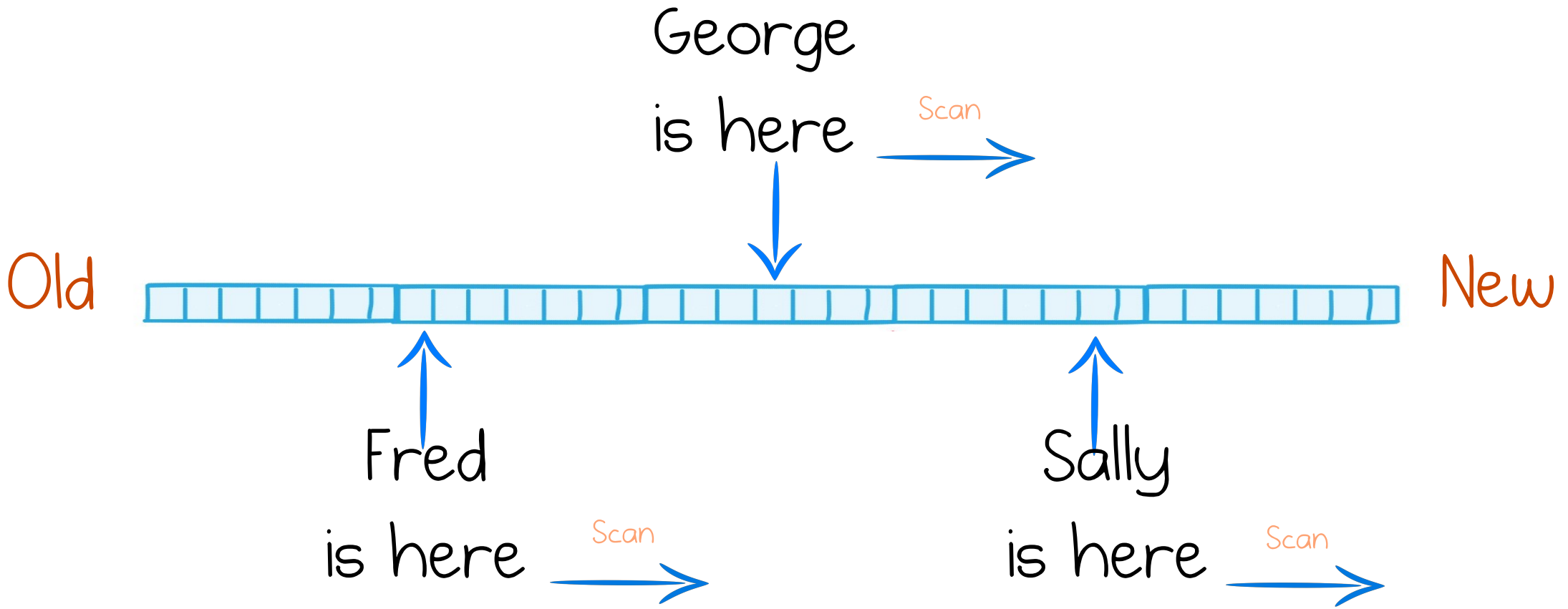
(b) Built in fault tolerance / HA

(c) Storage

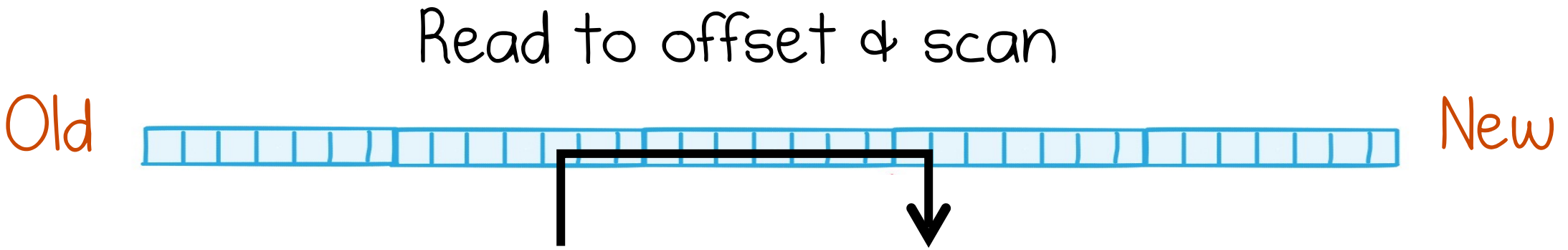
# The log is a simple idea



# Consumers have a position all of their own

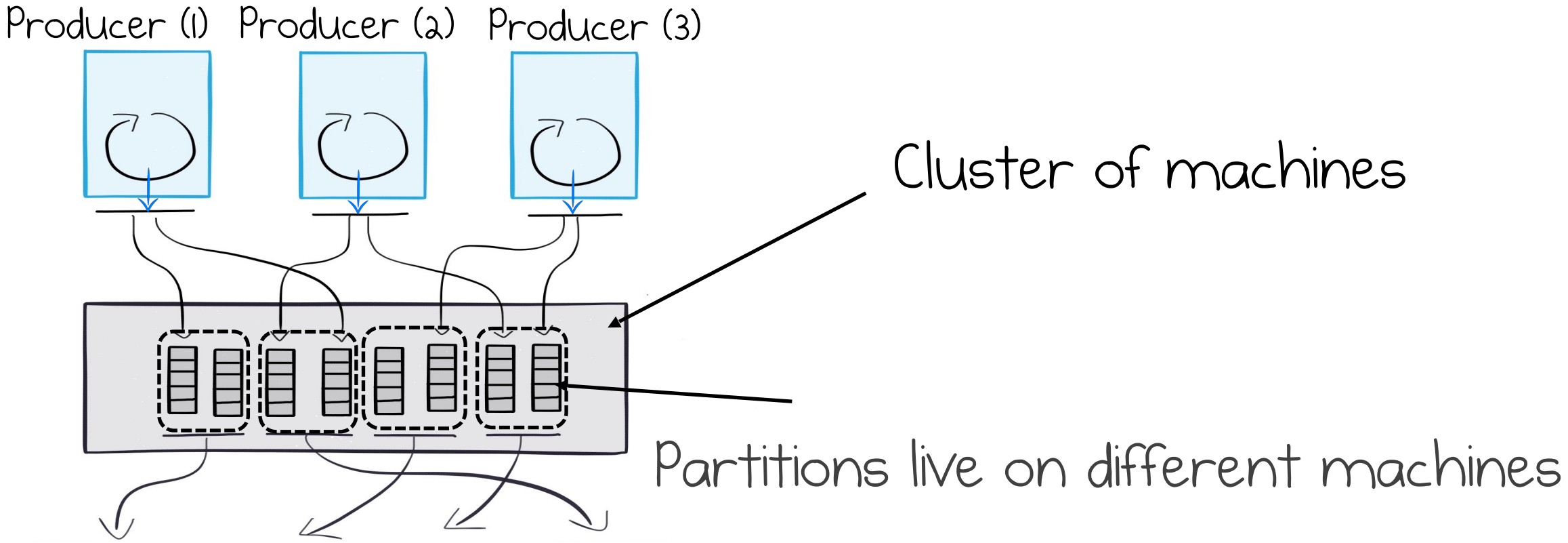


# Only Sequential Access

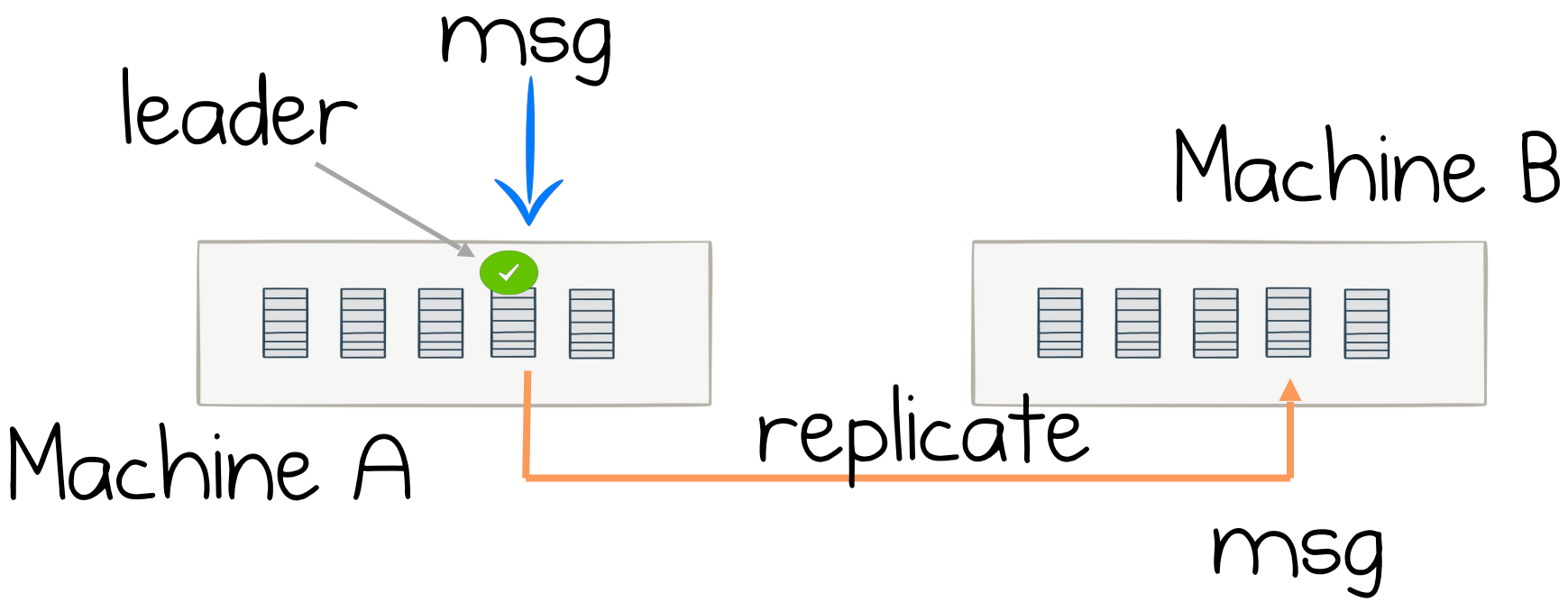


# Shard data to get scalability

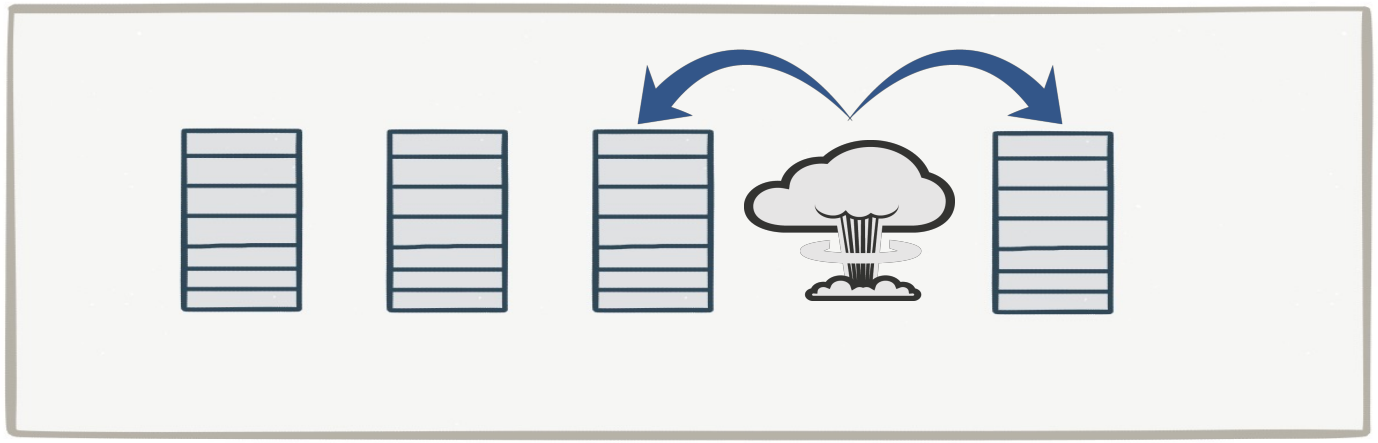
Messages are sent to different partitions



# Replicate to get fault tolerance

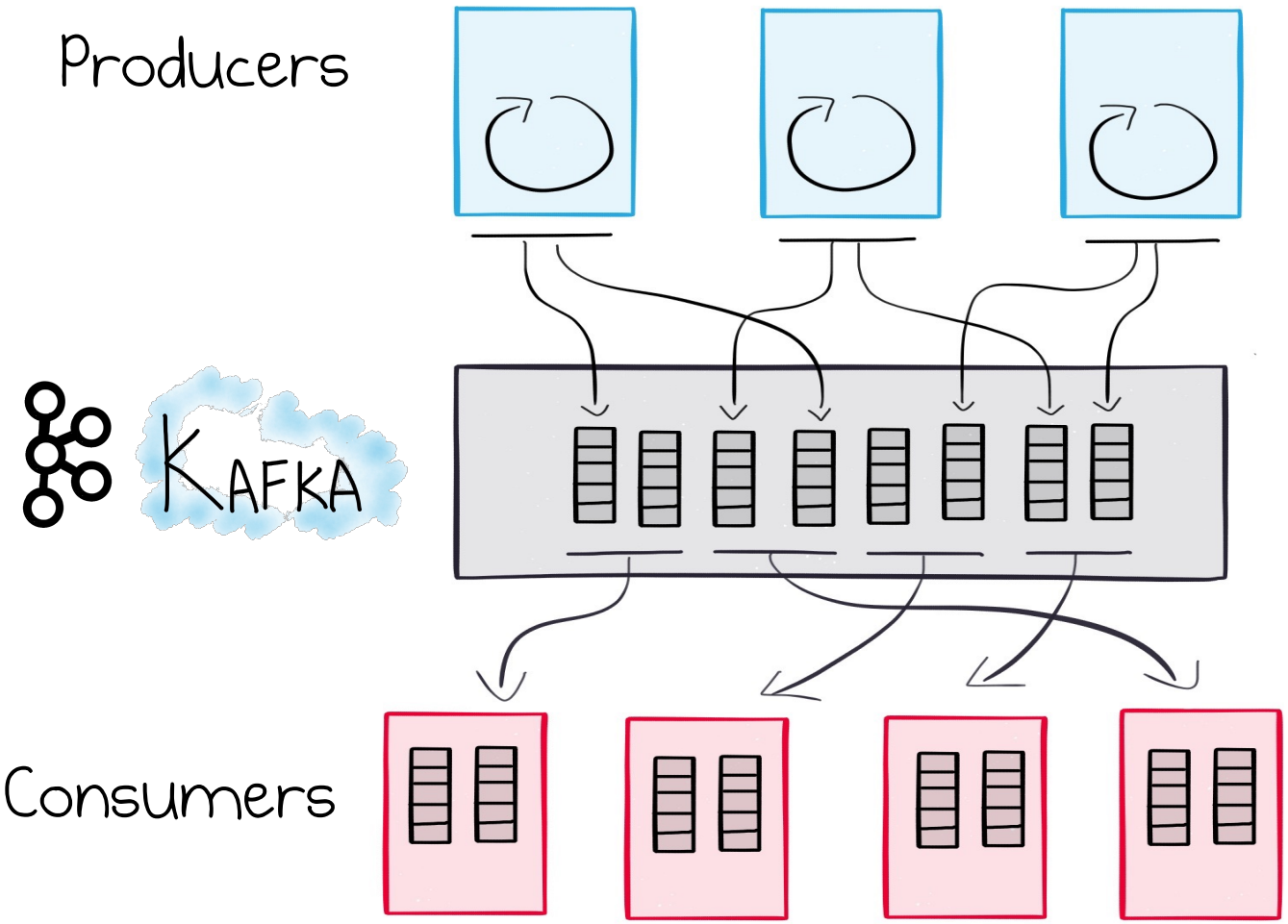


# Replication provides resiliency



A 'replica' takes over on machine failure

# Linearly Scalable Architecture

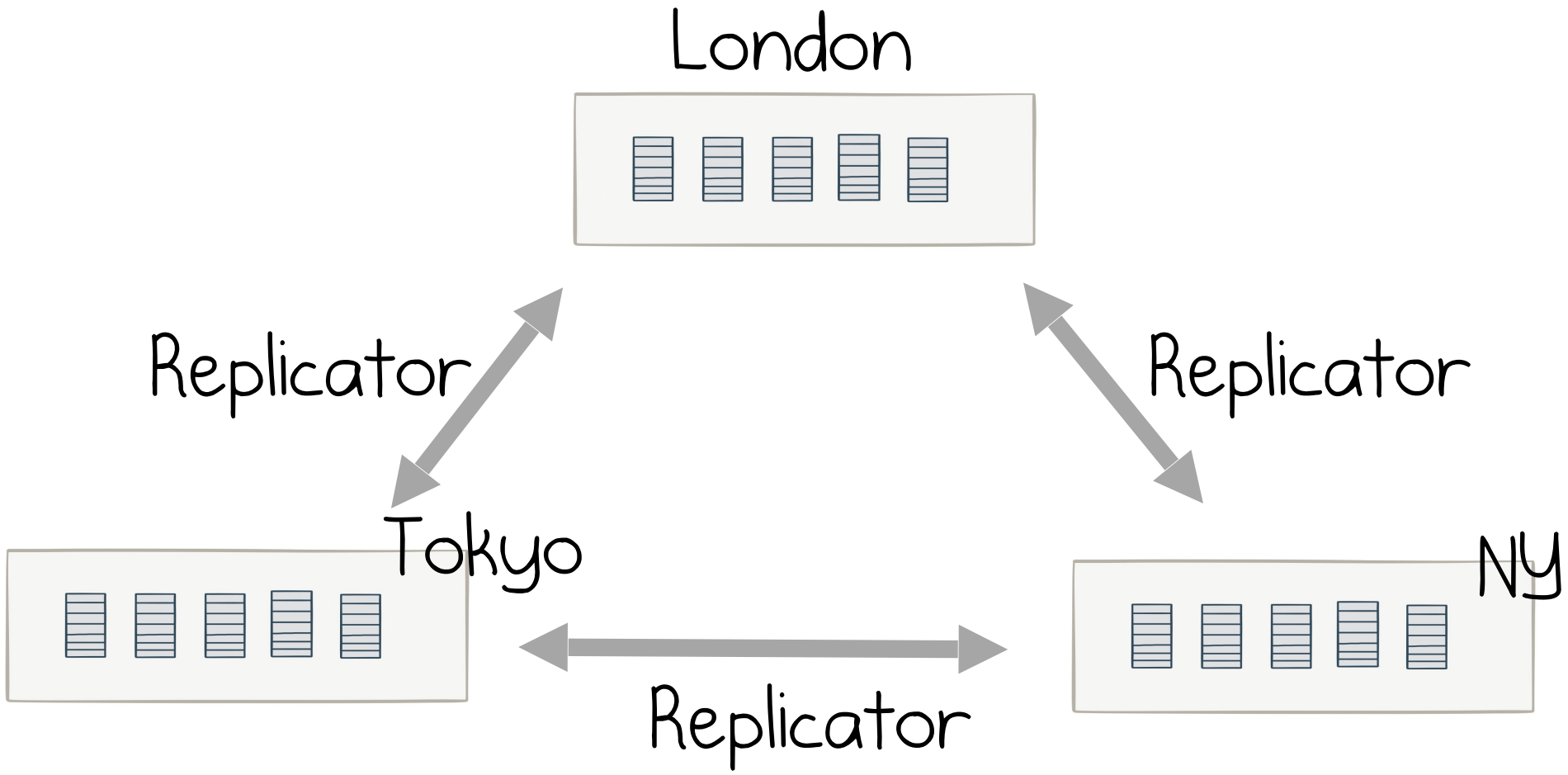



Single topic:

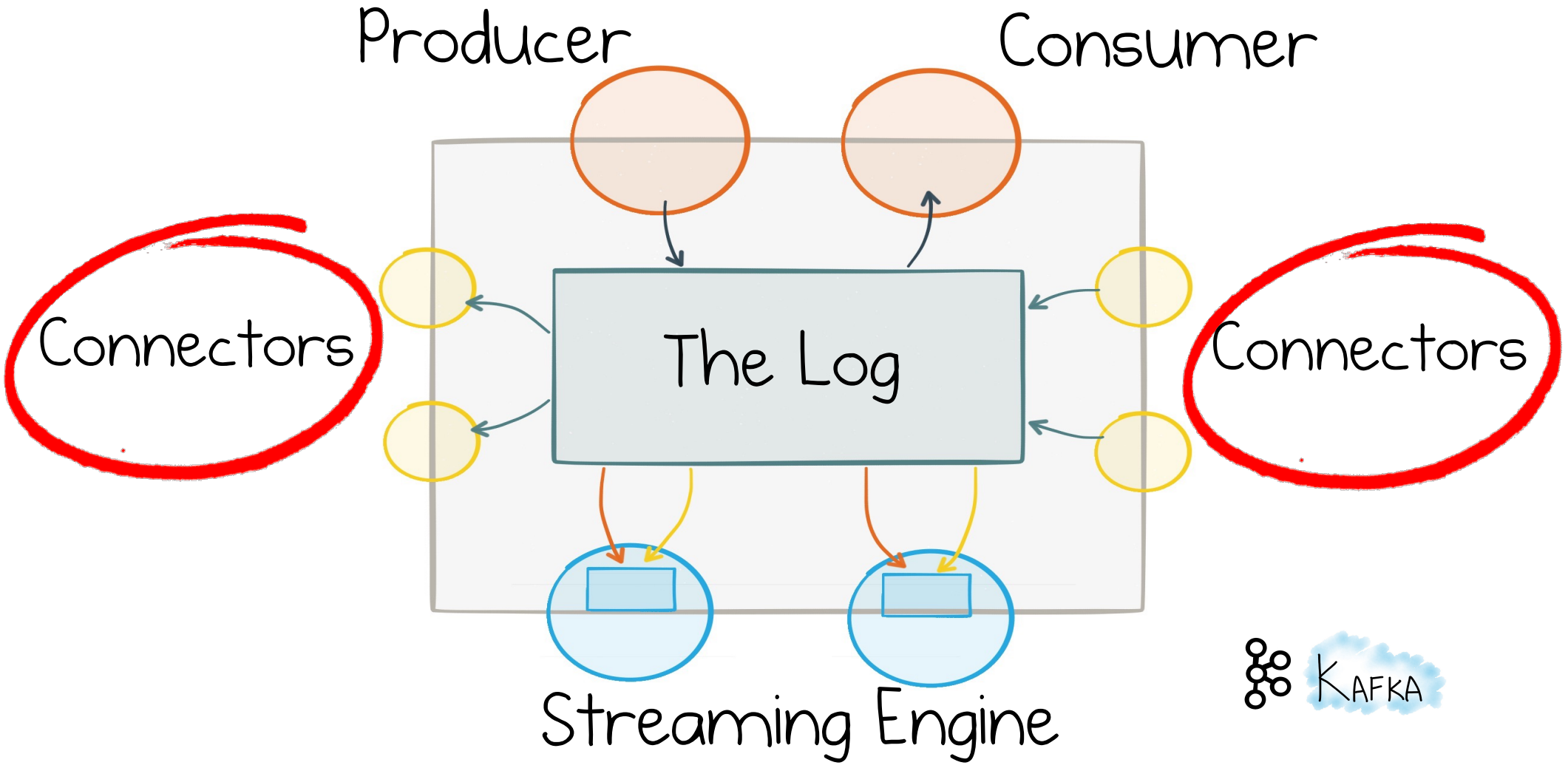
- Many producers machines
- Many consumer machines
- Many Broker machines

**No Bottleneck!!**

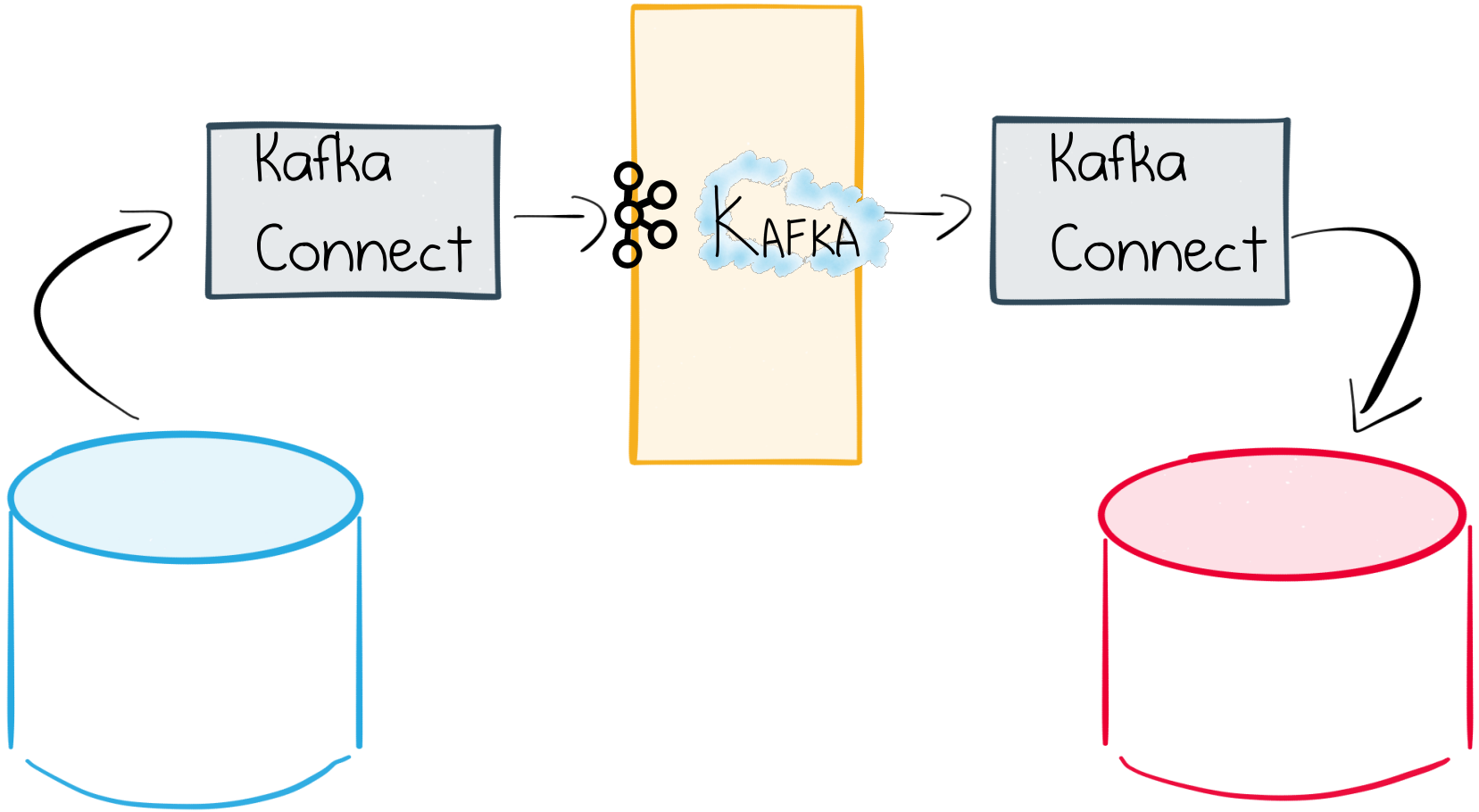
# Worldwide, localized views



# The Connect API



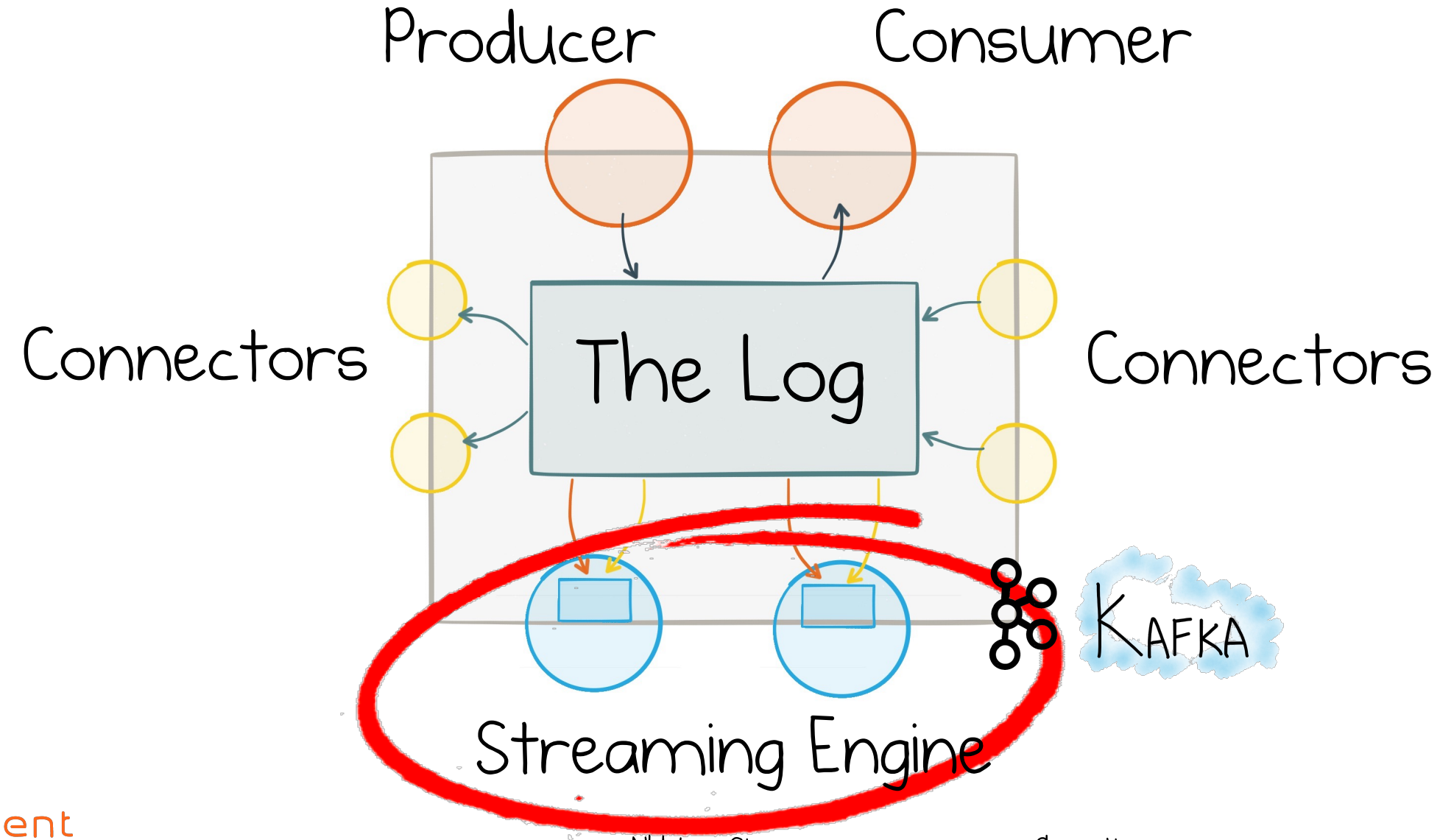
# Ingest / Egest into any data source

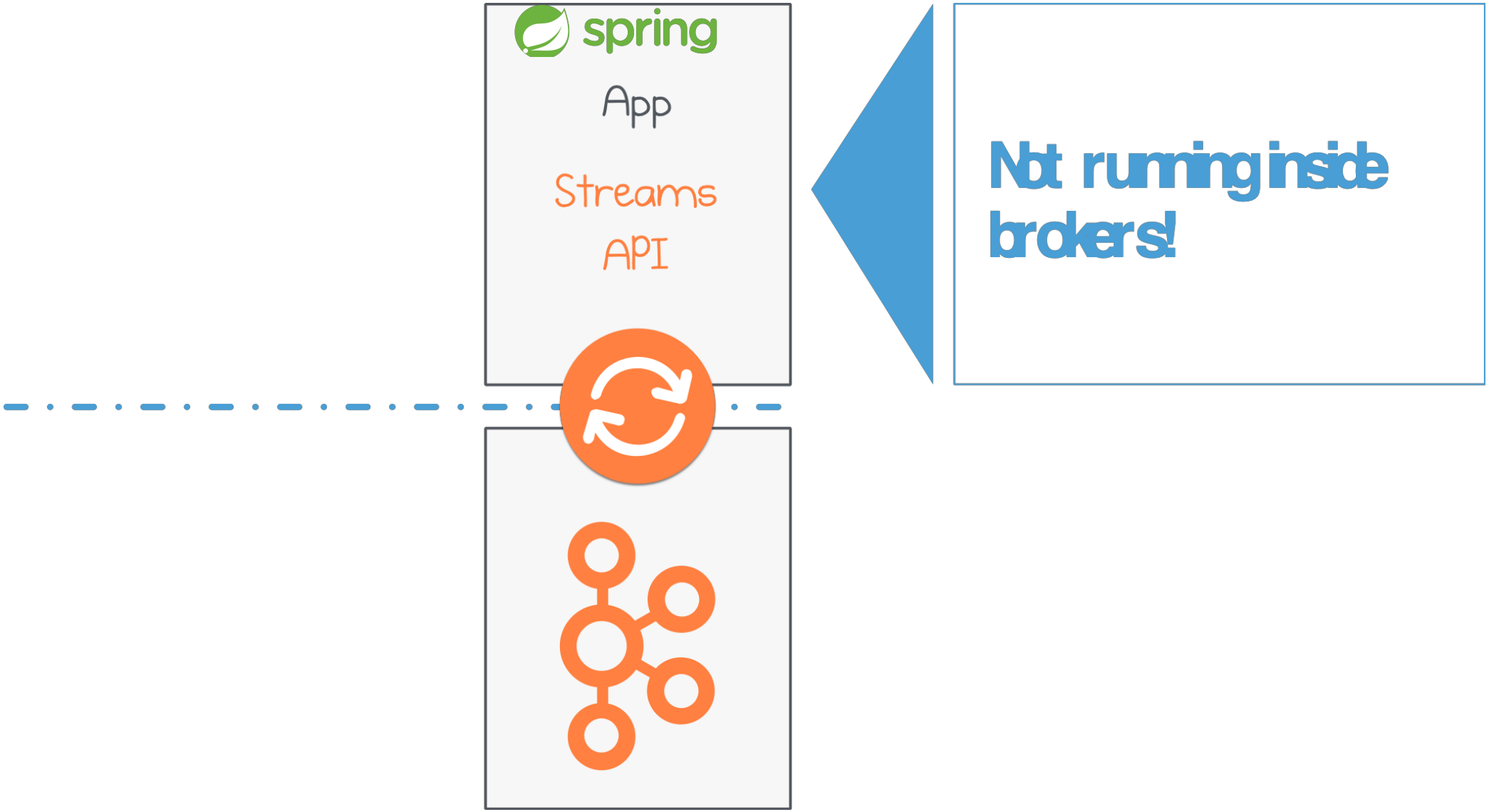


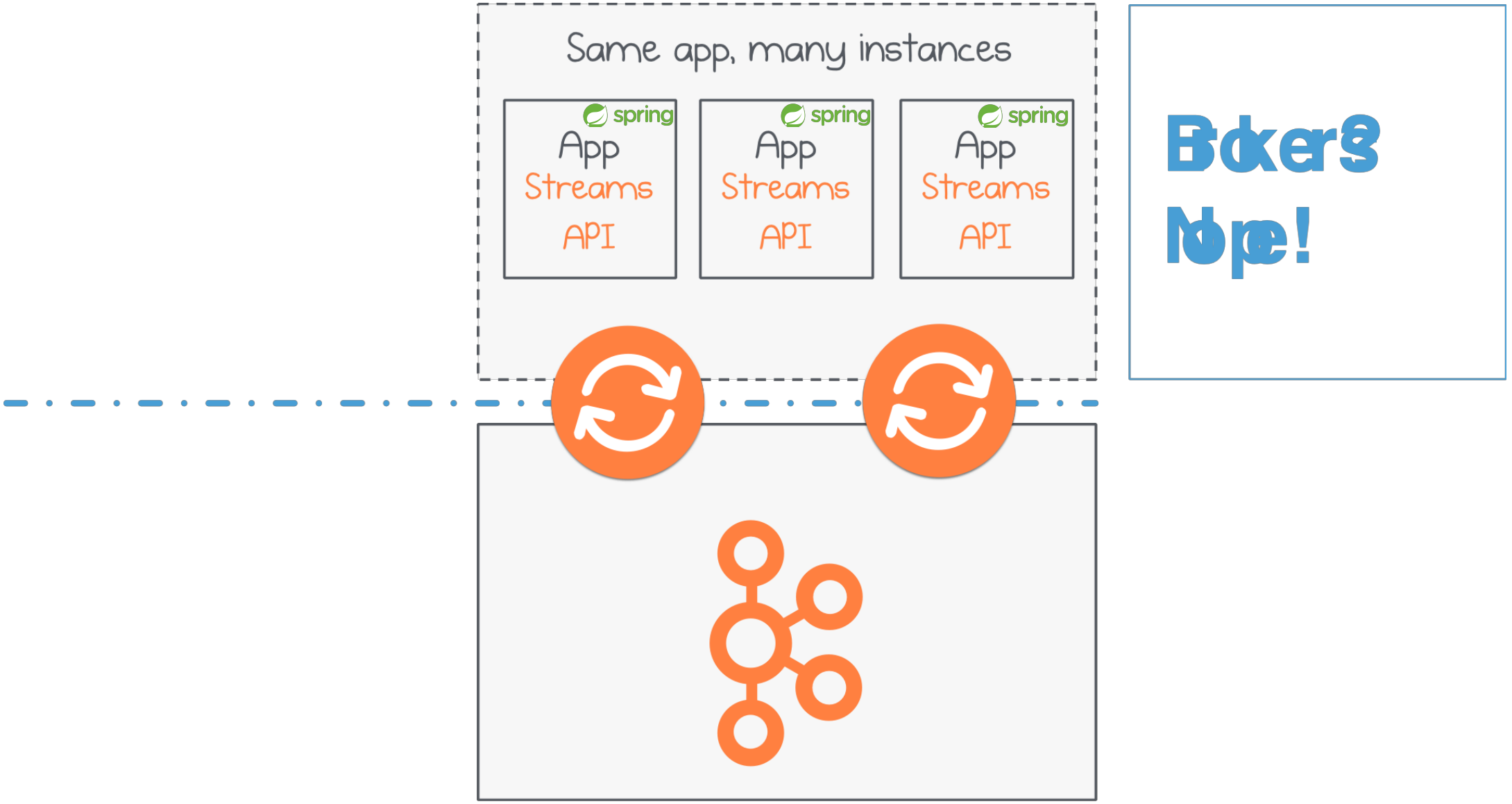
# Ingest/Egest data from/to data sources

- Amazon S3
- Elasticsearch
- HDFS
- JDBC
- Couchbase
- Cassandra
- Oracle
- SAP
- Vertica
- Blockchain
- DynamoDB
- FTP
- Github
- BigQuery
- Google Pub Sub
- RethinkDB
- Salesforce
- Solr
- Splunk
- JMX
- Kenesis
- MongoDB
- MQTT
- NATS
- Postgres
- Rabbit
- Redis
- Twitter
- Bintray

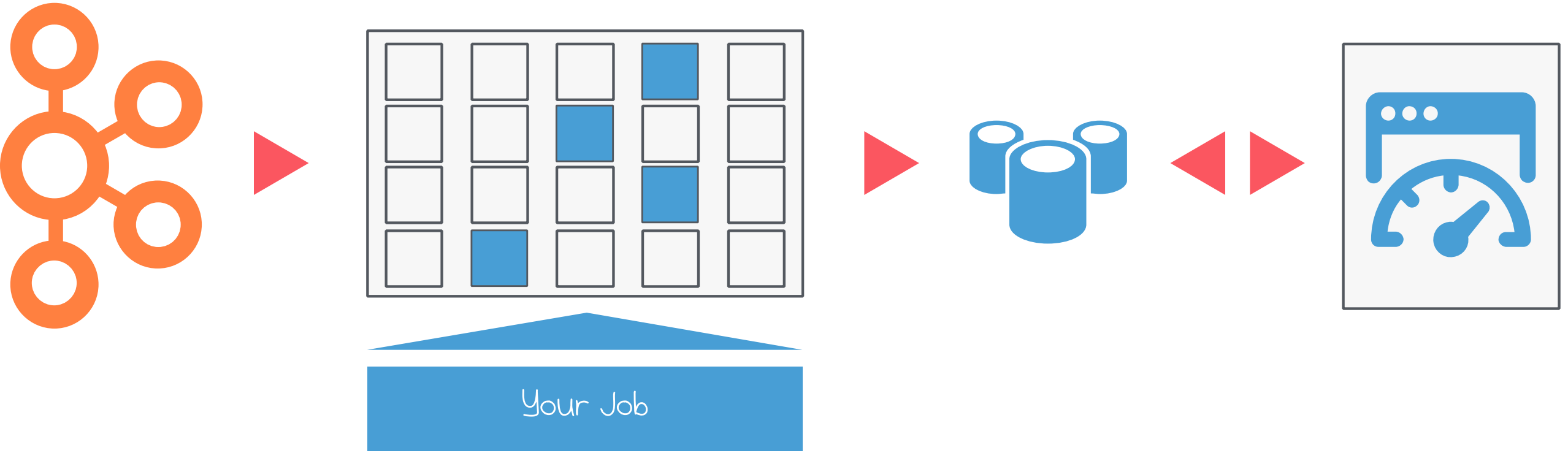
# Kafka Streams and KSQL



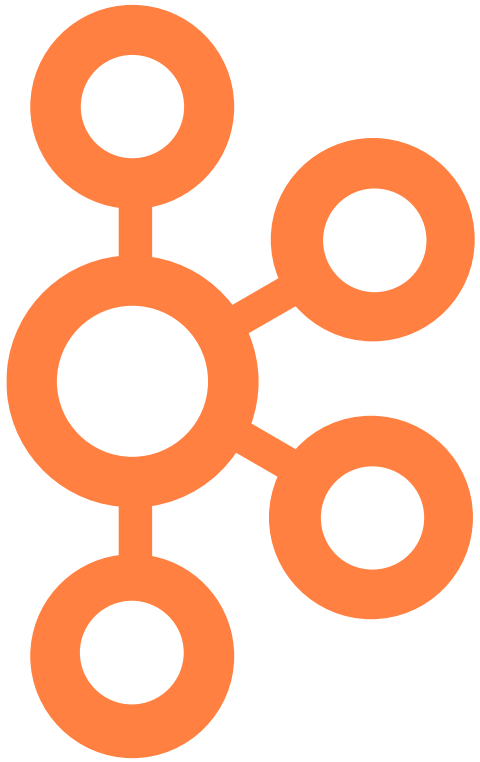

spring
App
Streams
API
Not running inside brokers!

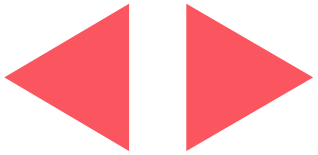

Same app, many instances
spring
App
Streams
API
spring
App
Streams
API
spring
App
Streams
API
Brokers
Nope!

# Before

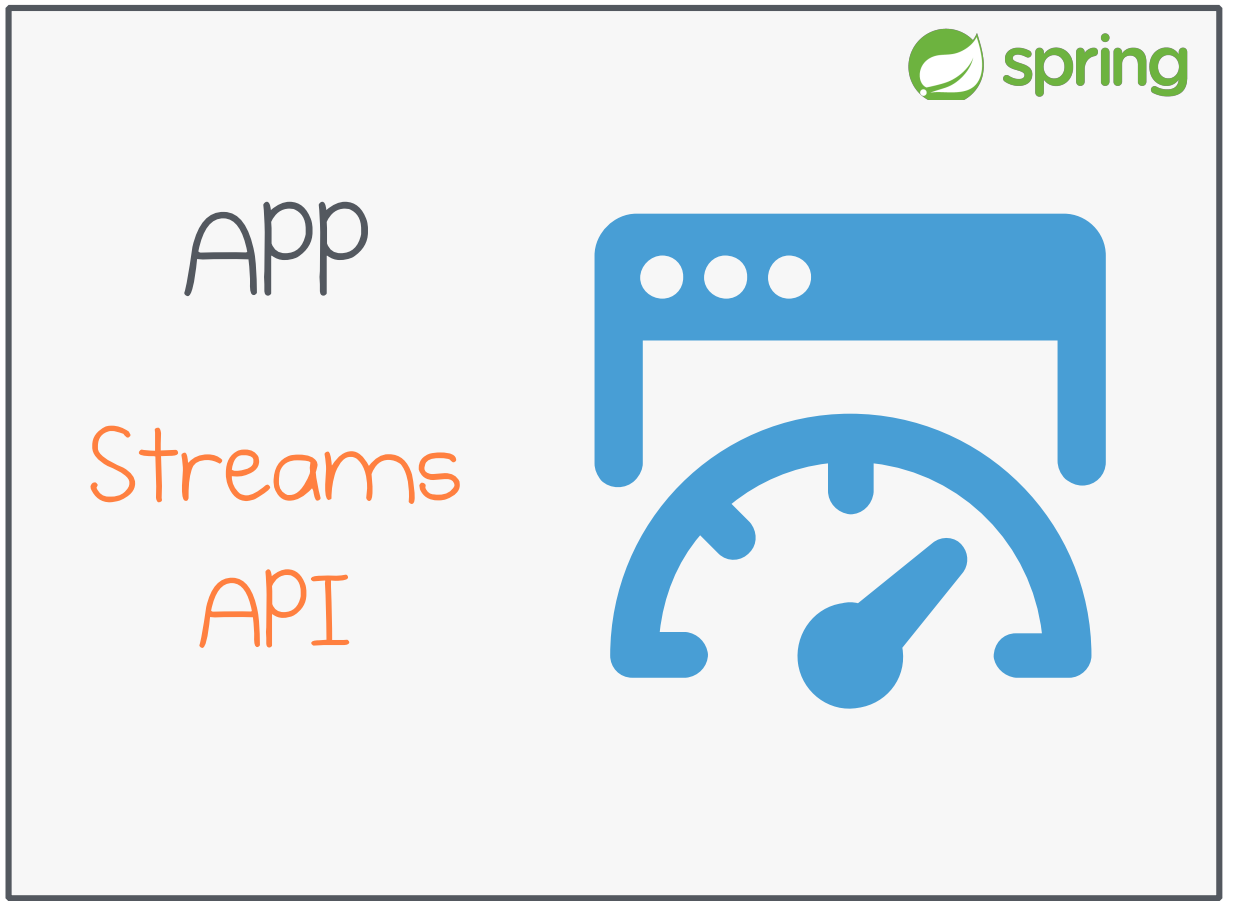


# After



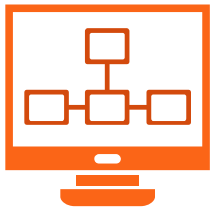


Dashboard



# Things Kafka Streams Does

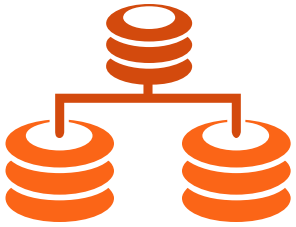
Runs everywhere



Clustering done for you

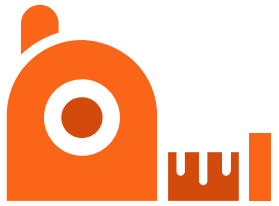
Exactly-once processing

Event-time processing



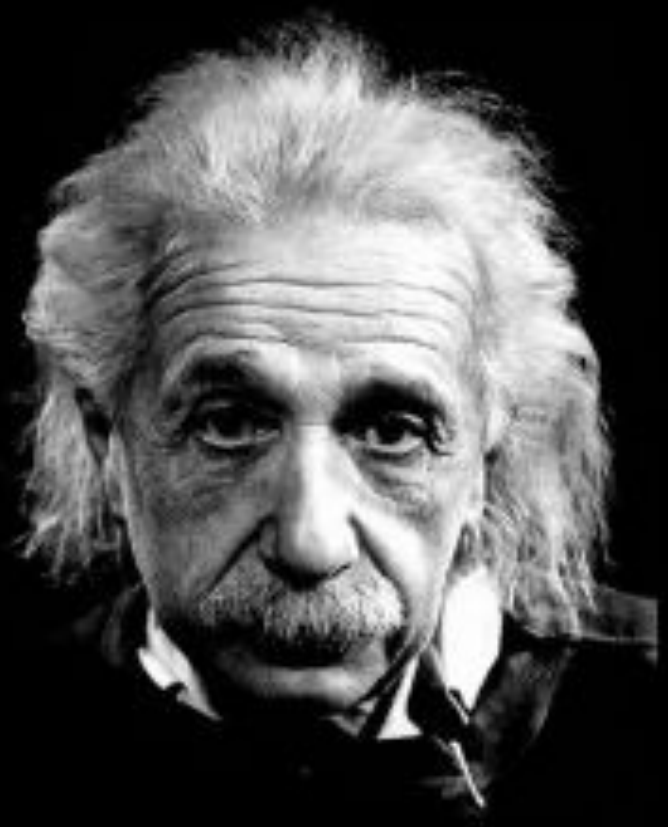
Integrated database

Joins, windowing, aggregation



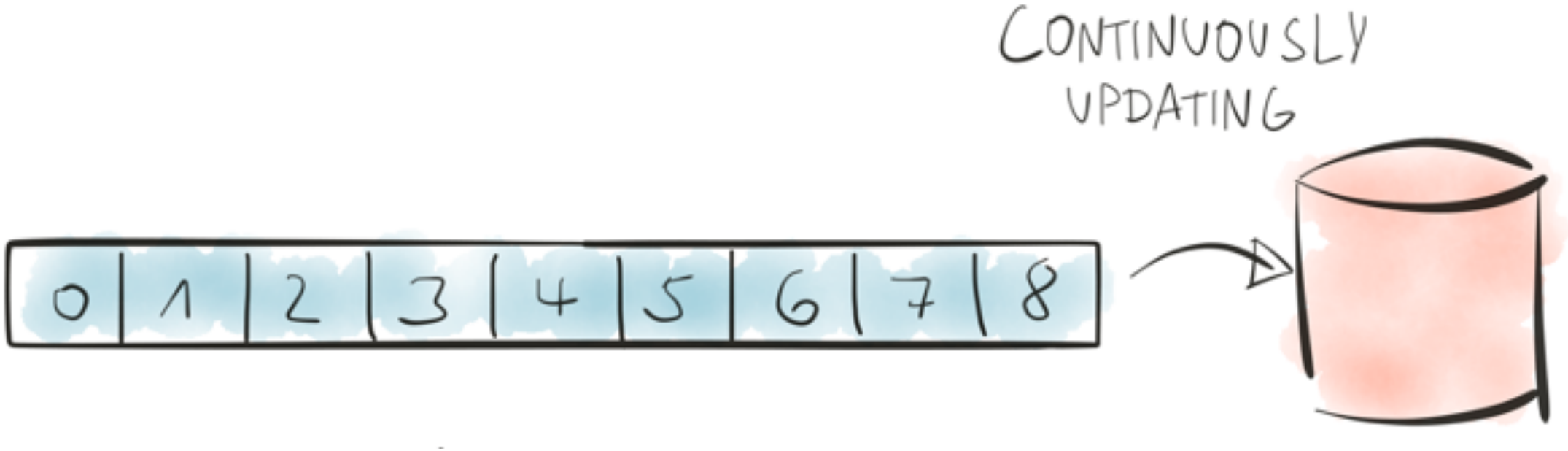
S/M/L/XL/XXL/XXXL sizes

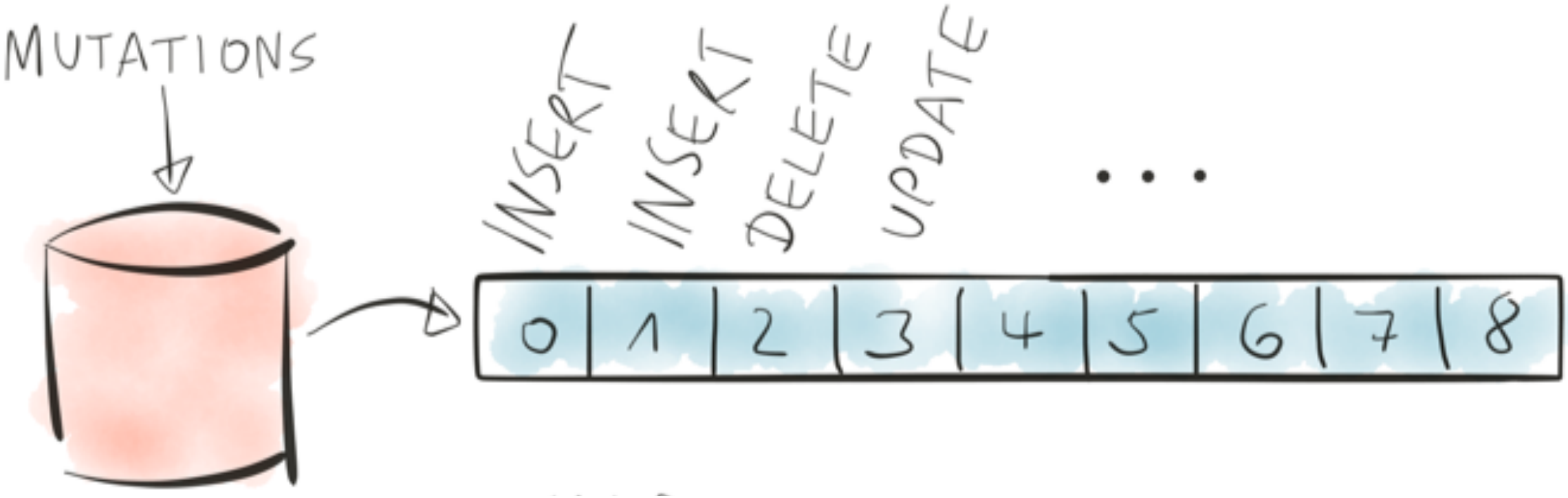
# SPECIAL RELATIVITY



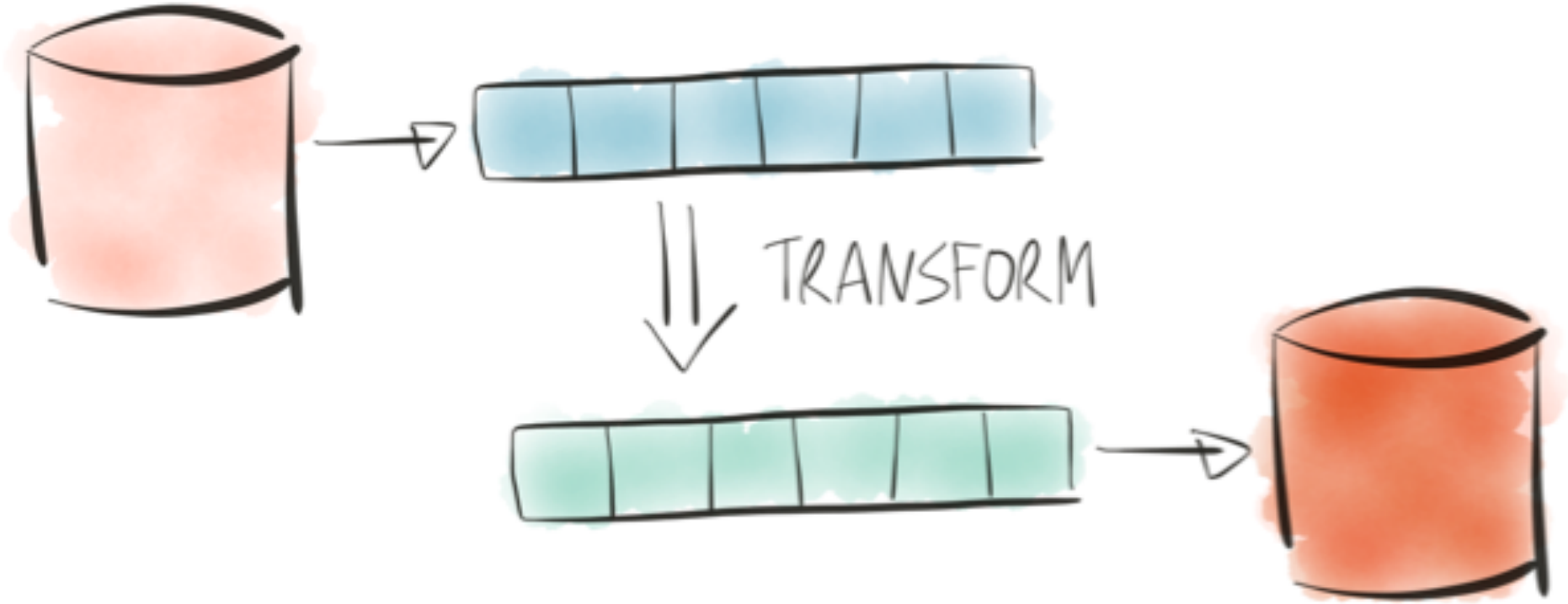
# OF STREAMS AND TABLES

# Streams to Tables

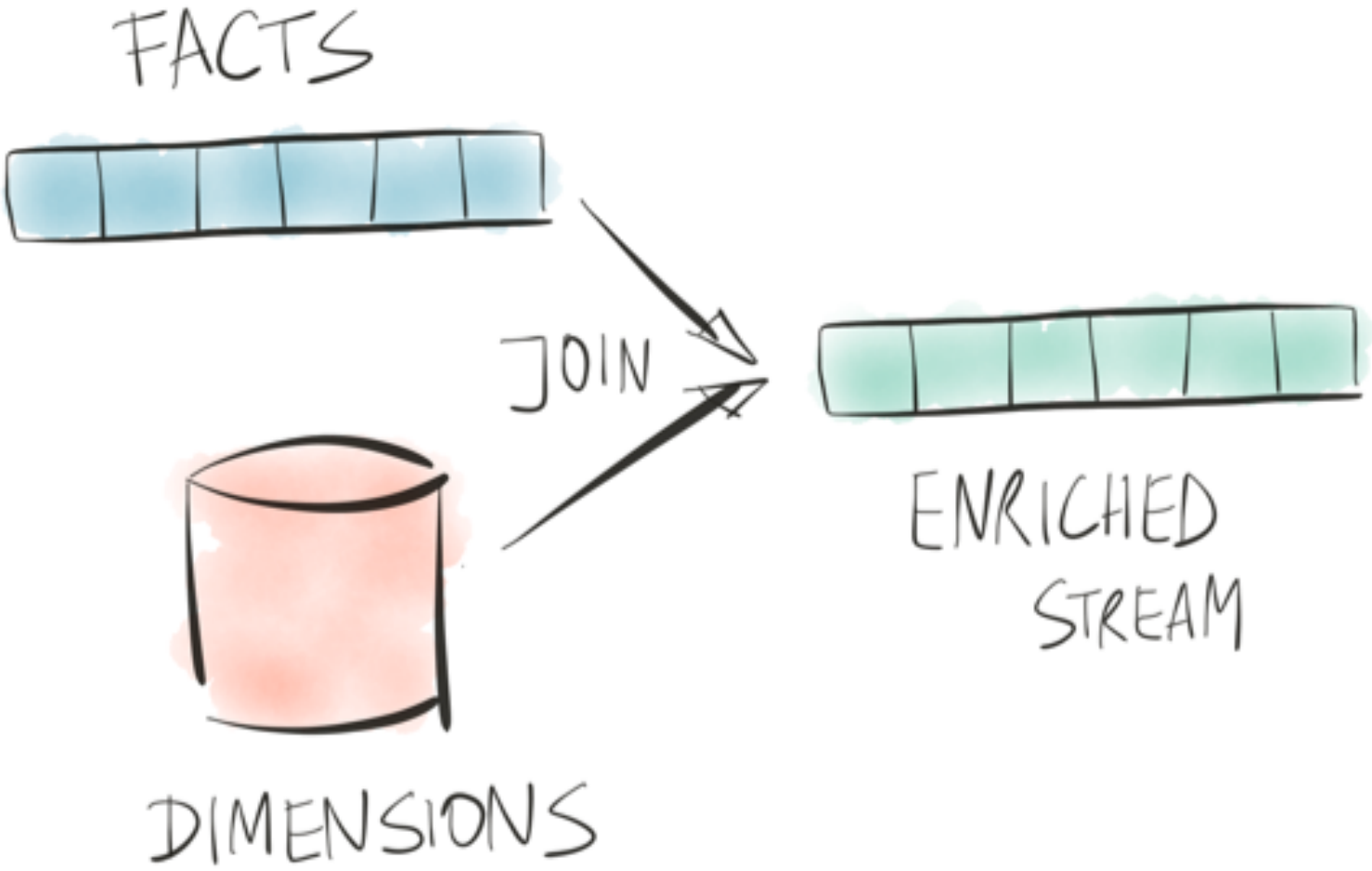



MUTATIONS
INSERT
INSERT
DELETE
UPDATE
. . .
0
1
2
3
4
5
6
7
8

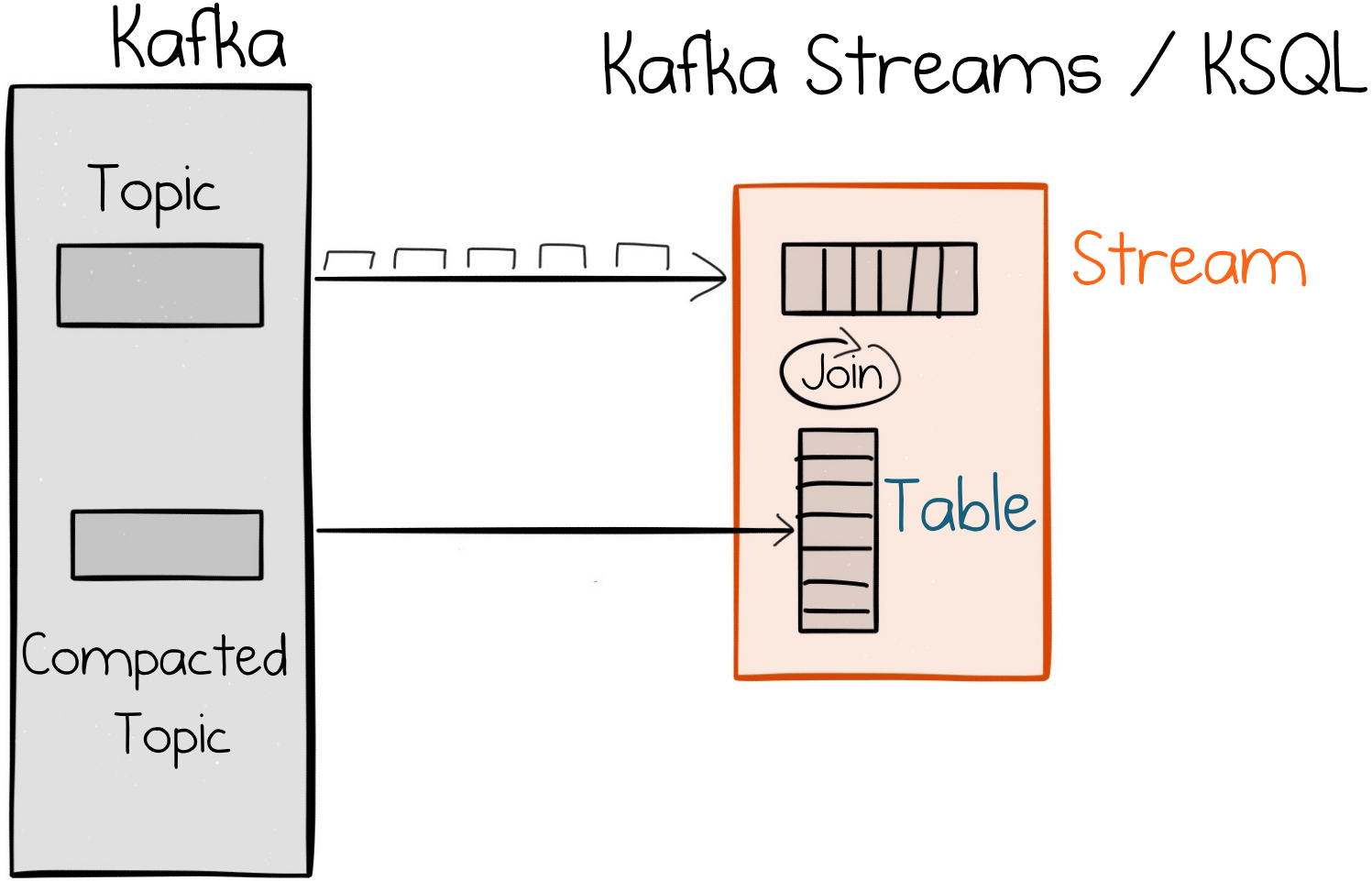
# Stream/Table Duality



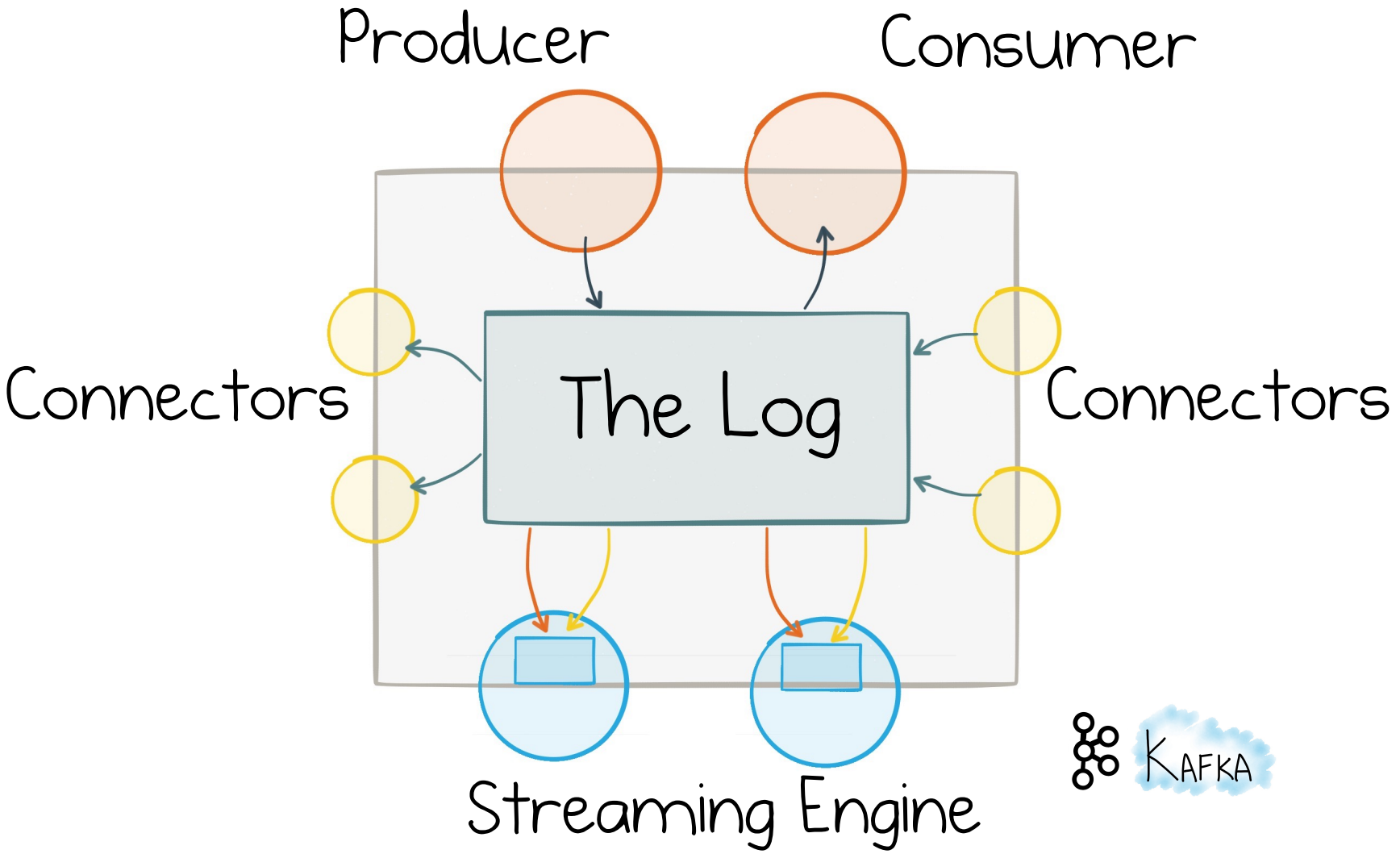
# Stream/Table Duality



# Join Streams and Tables



# Kafka is a complete Streaming Platform



THE LEADING DISTRIBUTION OF APACHE KAFKA

# Downloads https://www.confluent.io/download/

Choose between Confluent Open Source or Confluent Enterprise edition. Both are built on the world's most popular streaming platform, Apache Kafka®.

**Confluent Open Source**
Download v4.0

**Confluent Enterprise**
Download v4.0

**Free & Open Source**

Confluent Open Source is a tested, complete package around Apache Kafka that includes all the features needed to build a streaming platform.

- Apache Kafka
- Build in any language
- Manage your schema
- Connect popular systems

**30-Day Free Trial**

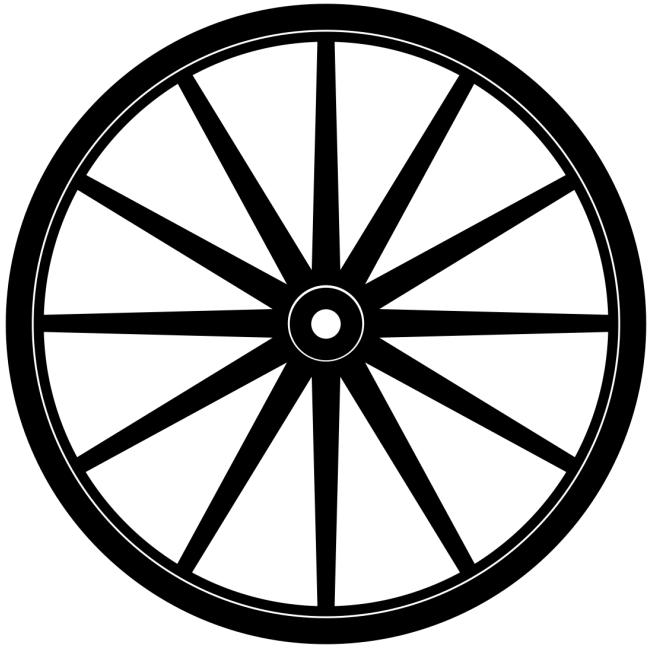
Confluent Enterprise includes Confluent Open Source, and adds features that improve scale, performance and reliability in production.

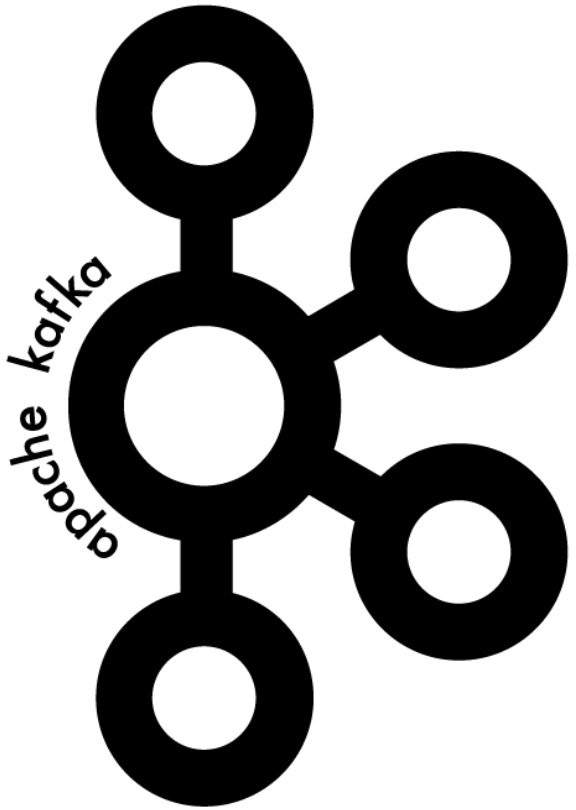
- Monitor your deployment
- Achieve high availability
- Automate resource allocation
- Have 24/7 support

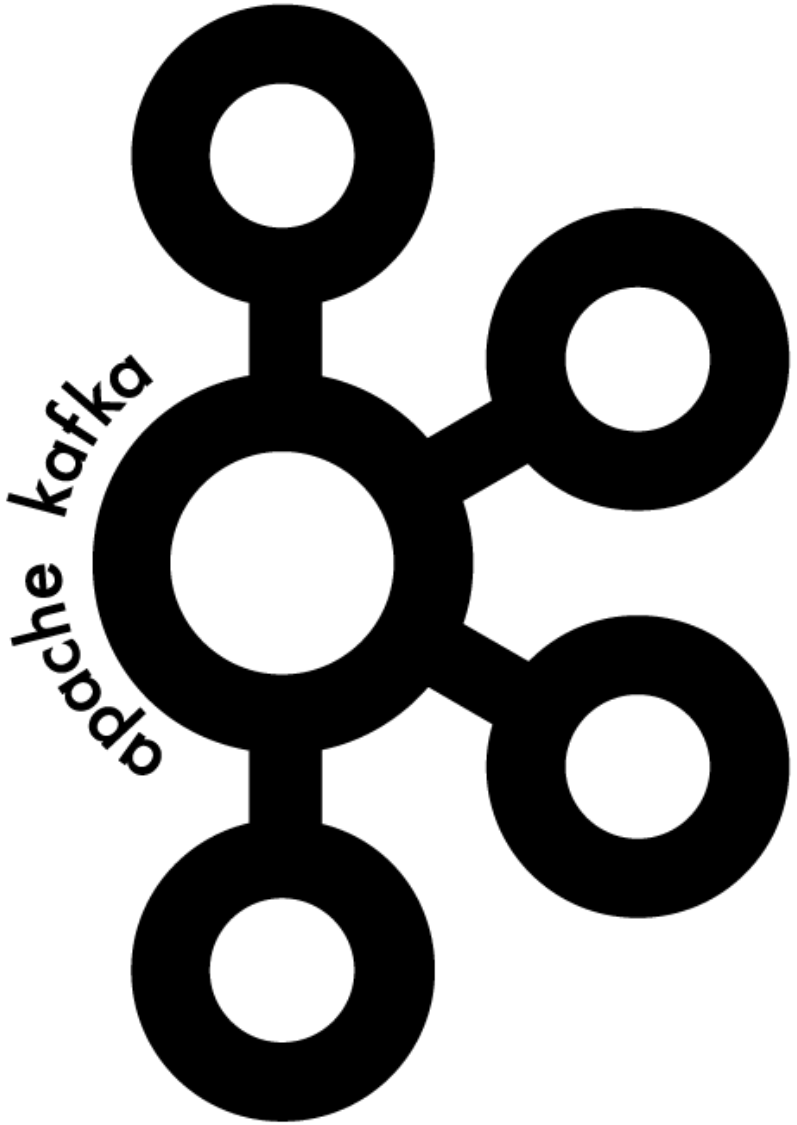
# Will it fly?

# Let's try!

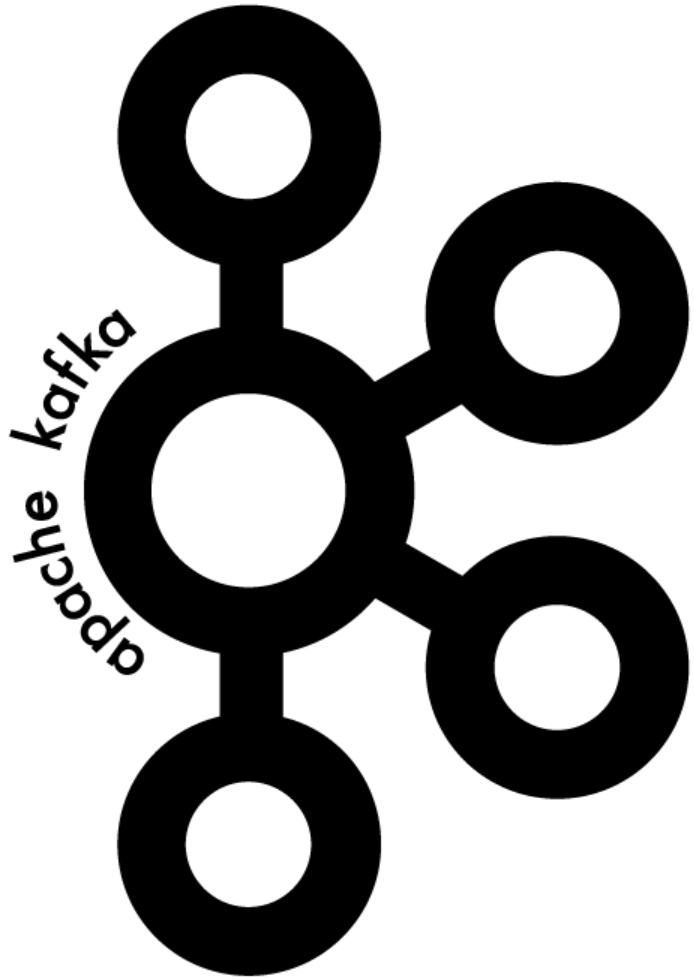
# One more thing...

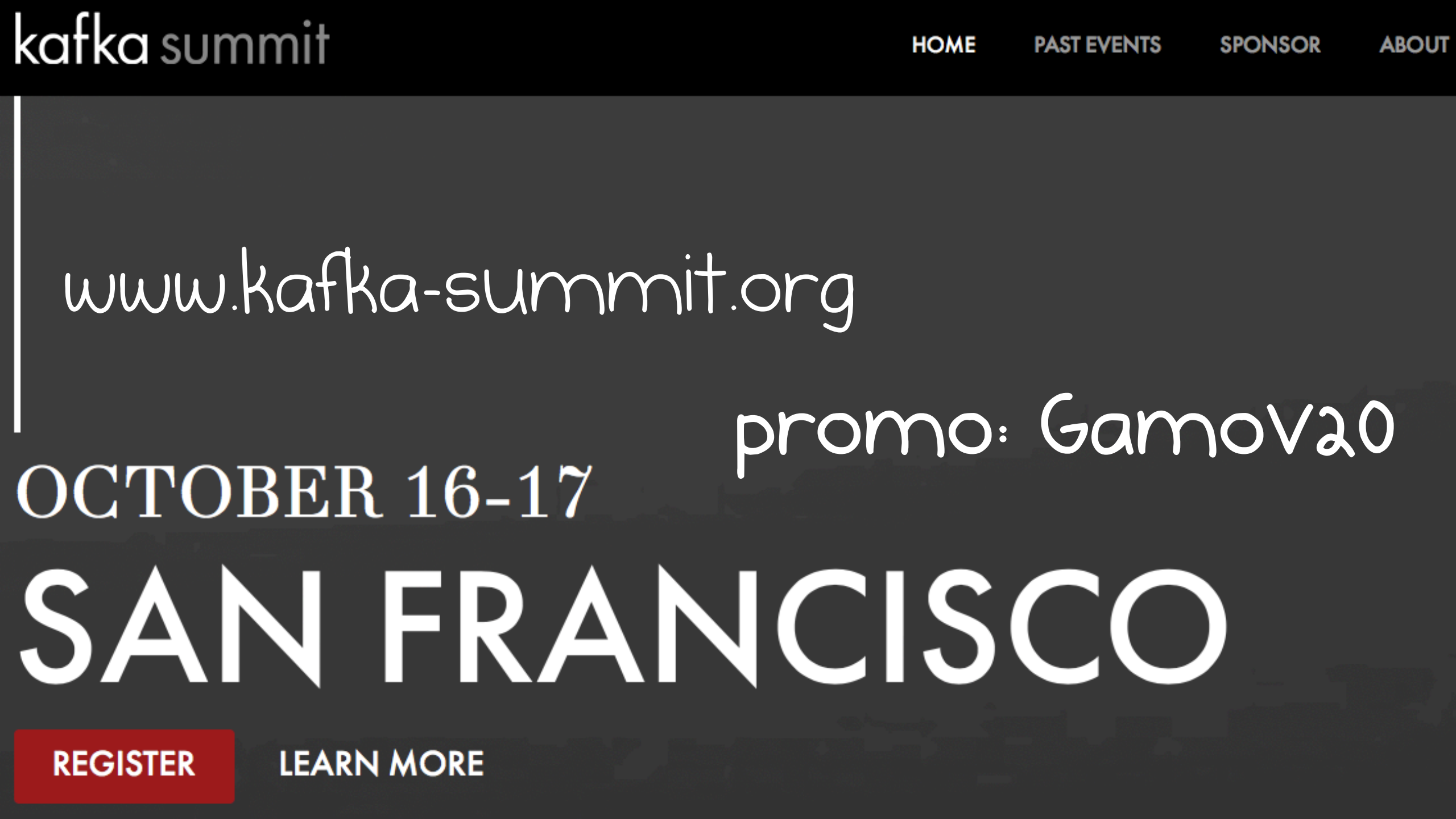



apache kafka


apache kafka

# A Major New Paradigm




kafka summit
HOME
PAST EVENTS
SPONSOR
ABOUT
www.kafka-summit.org
promo: Gamov20
OCTOBER 16-17
SAN FRANCISCO
REGISTER
LEARN MORE

# THANKS!

@gamussa
viktor@confluent.io

We are hiring!
https://www.confluent.io/careers/