

APACHE KAFKA EVENT-STREAMING PLATFORM FOR .NET DEVELOPERS

OCTOBER, 2019



@GAMUSSA | #SPRINGONE | @CONFLUENTINC

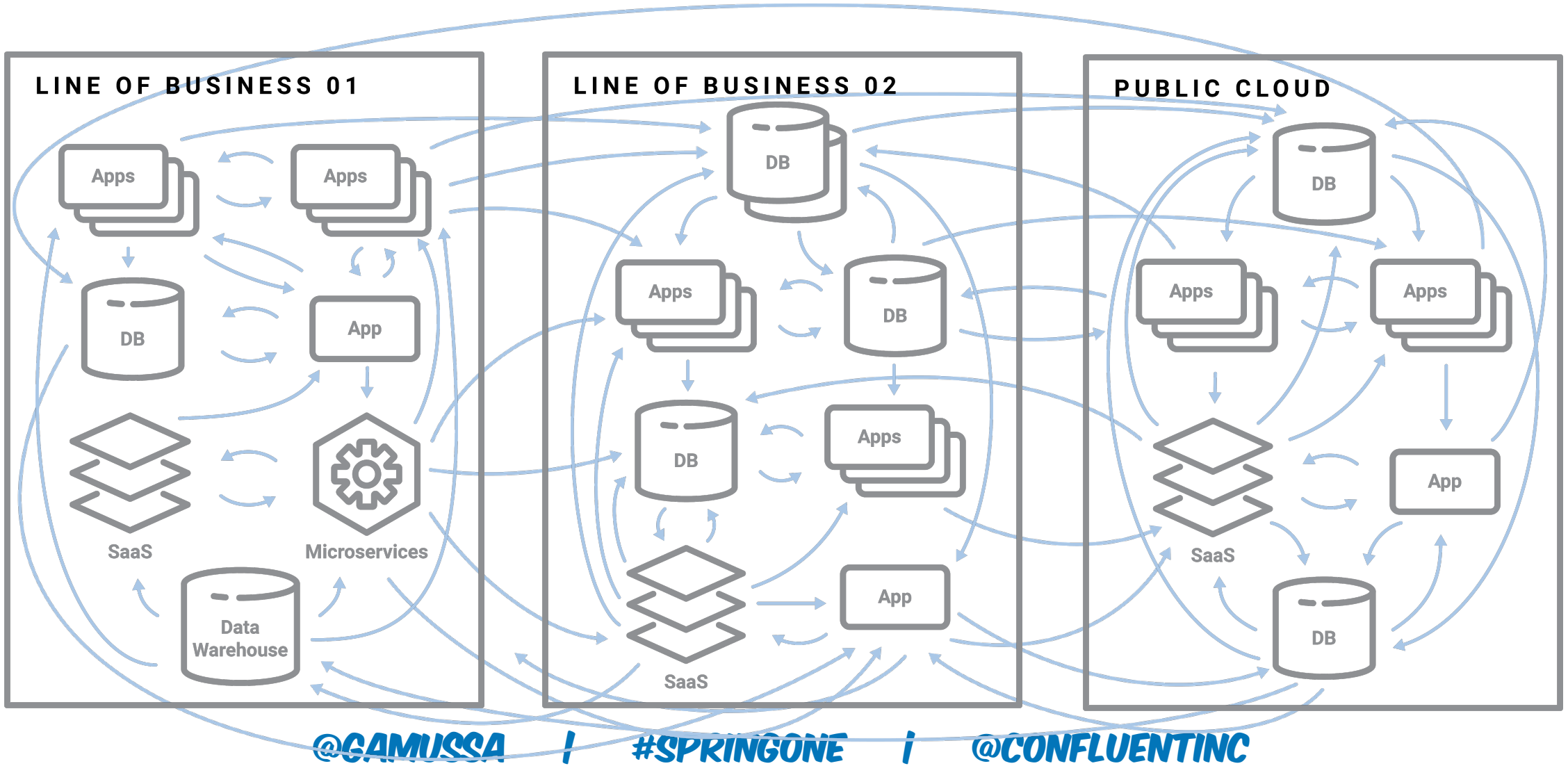
I BUILD HIGHLY SCALABLE

Hello World

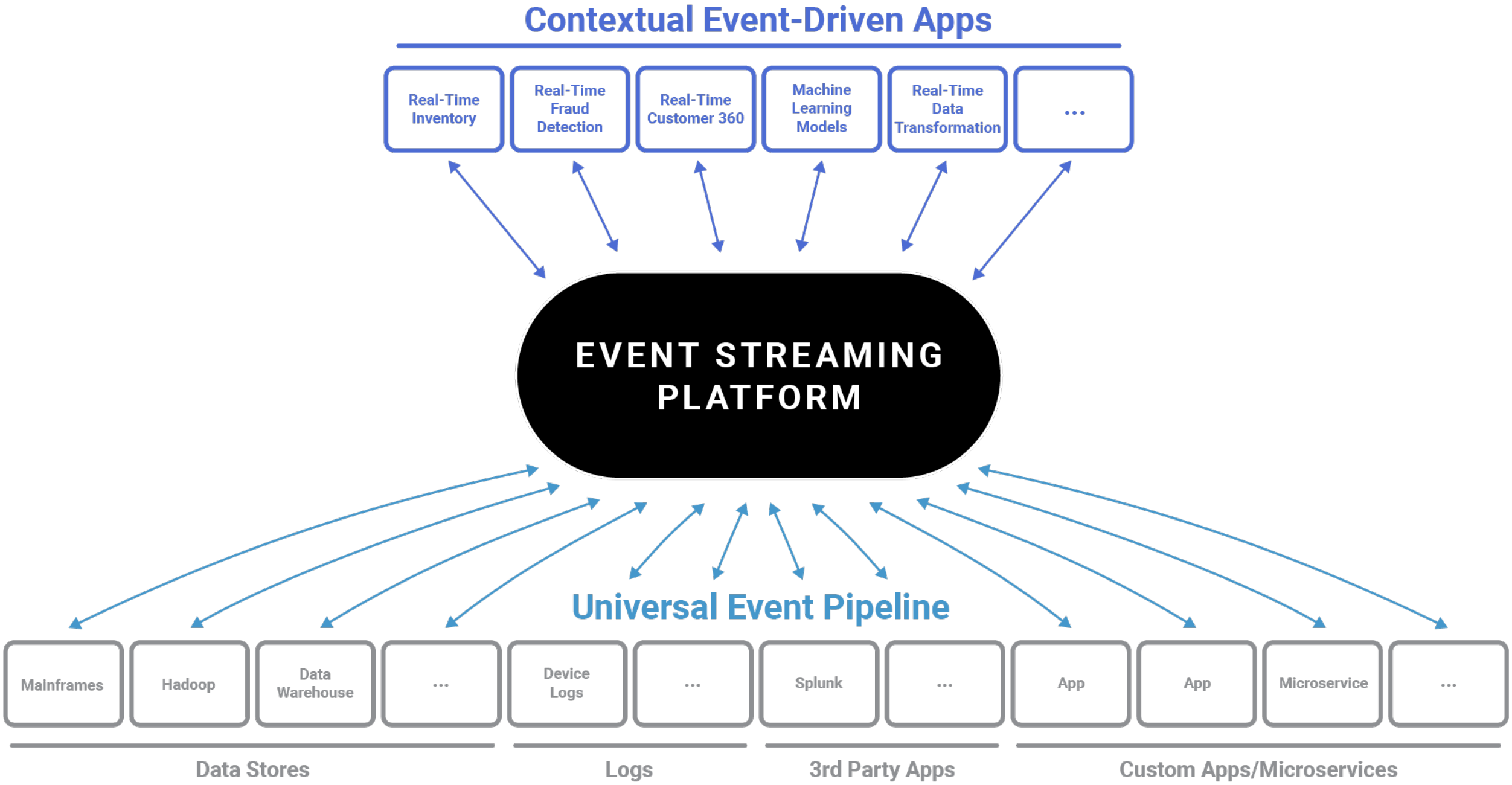
APPS

A COMPANY IS BUILD ON DATA FLOWS BUT ALL WE HAVE IS DATA STORES

PRE-STREAMING

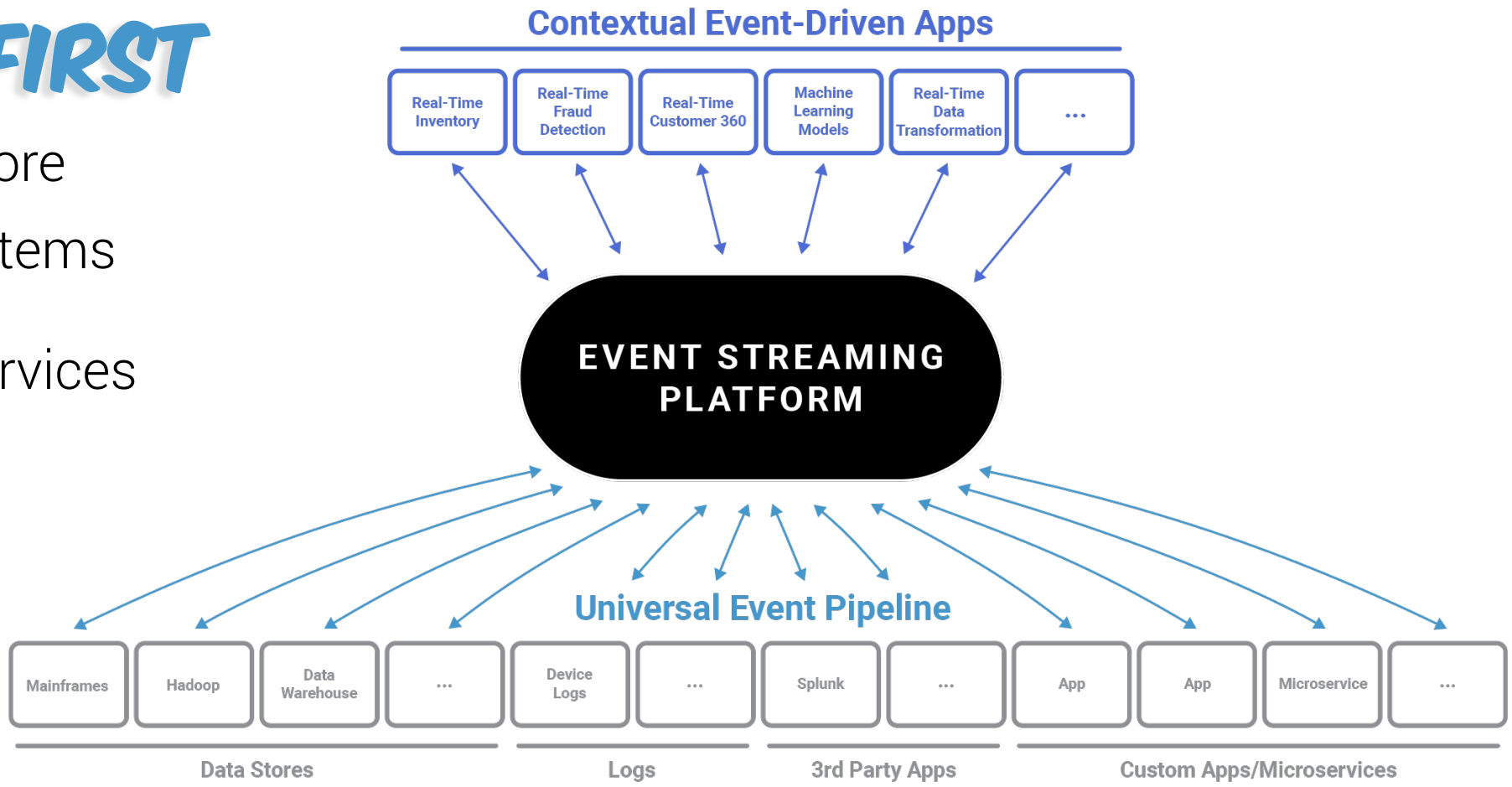



HIVE
hadoop
Map Reduce
STORM
kafka
APACHE
Spark
ZOO

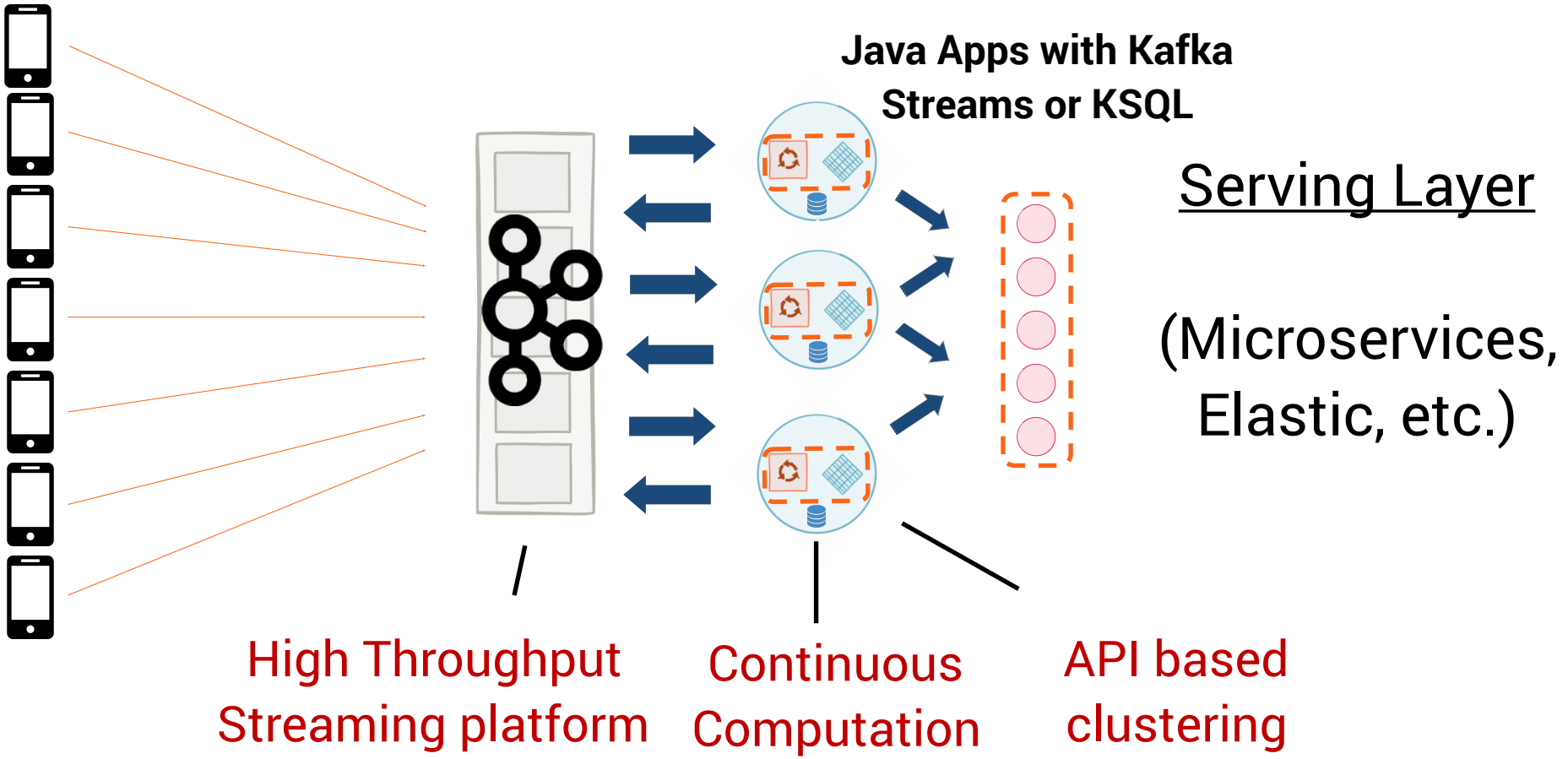

Contextual Event-Driven Apps
Real-Time Inventory
Real-Time Fraud Detection
Real-Time Customer 360
Machine Learning Models
Real-Time Data Transformation
...
EVENT STREAMING PLATFORM
Universal Event Pipeline
Mainframes
Hadoop
Data Warehouse
...
Device Logs
...
Splunk
...
App
App
Microservice
...
Data Stores
Logs
3rd Party Apps
Custom Apps/Microservices

NEW WORLD
STREAMING FIRST

- DB/DWH + Many more distributed data systems
- Monolith -> Microservices
- Batch -> Real-time

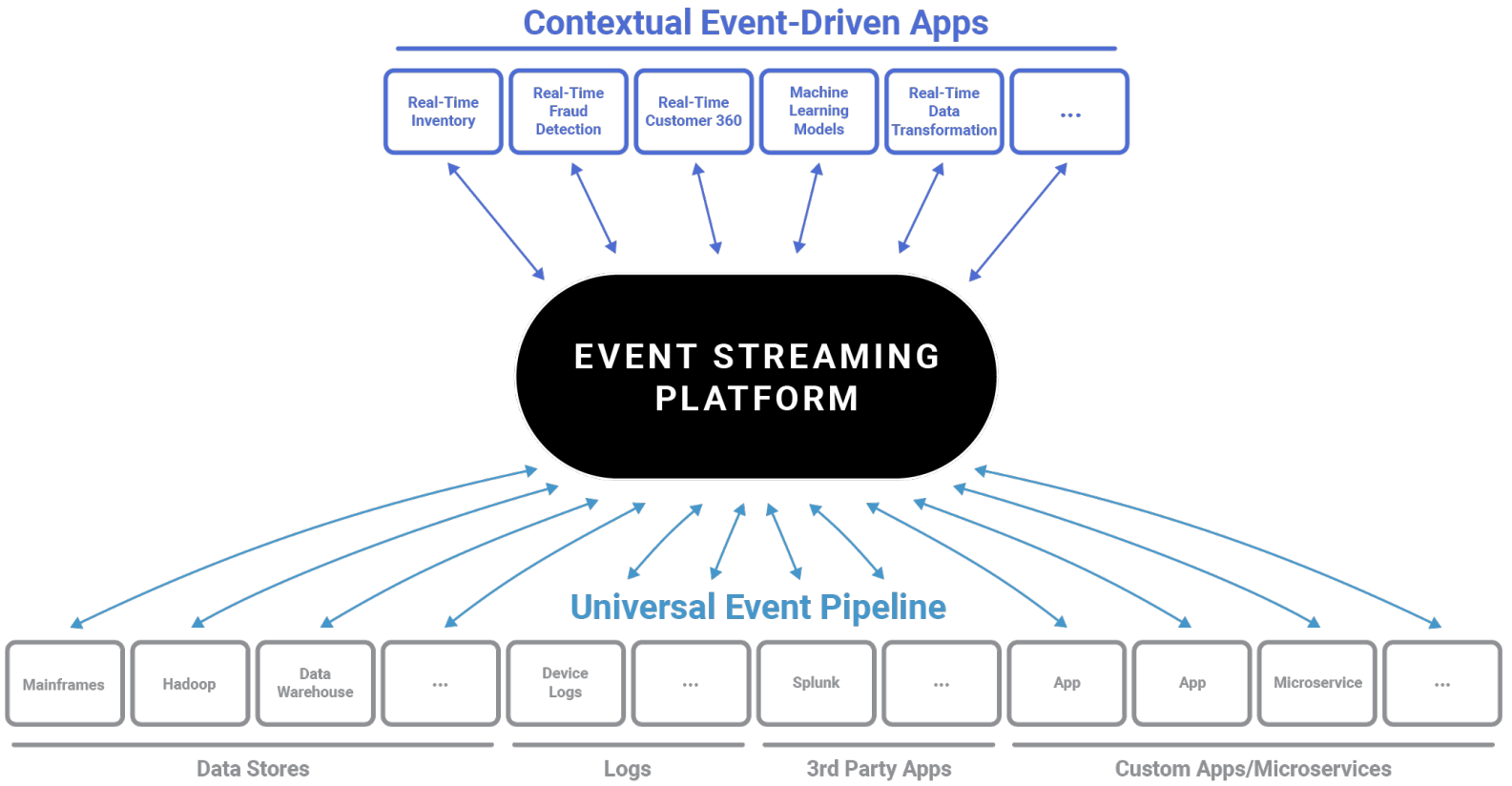


ORIGINS IN STREAM PROCESSING



STREAMING PLATFORM

- STORAGE
- PUB / SUB
- PROCESSING

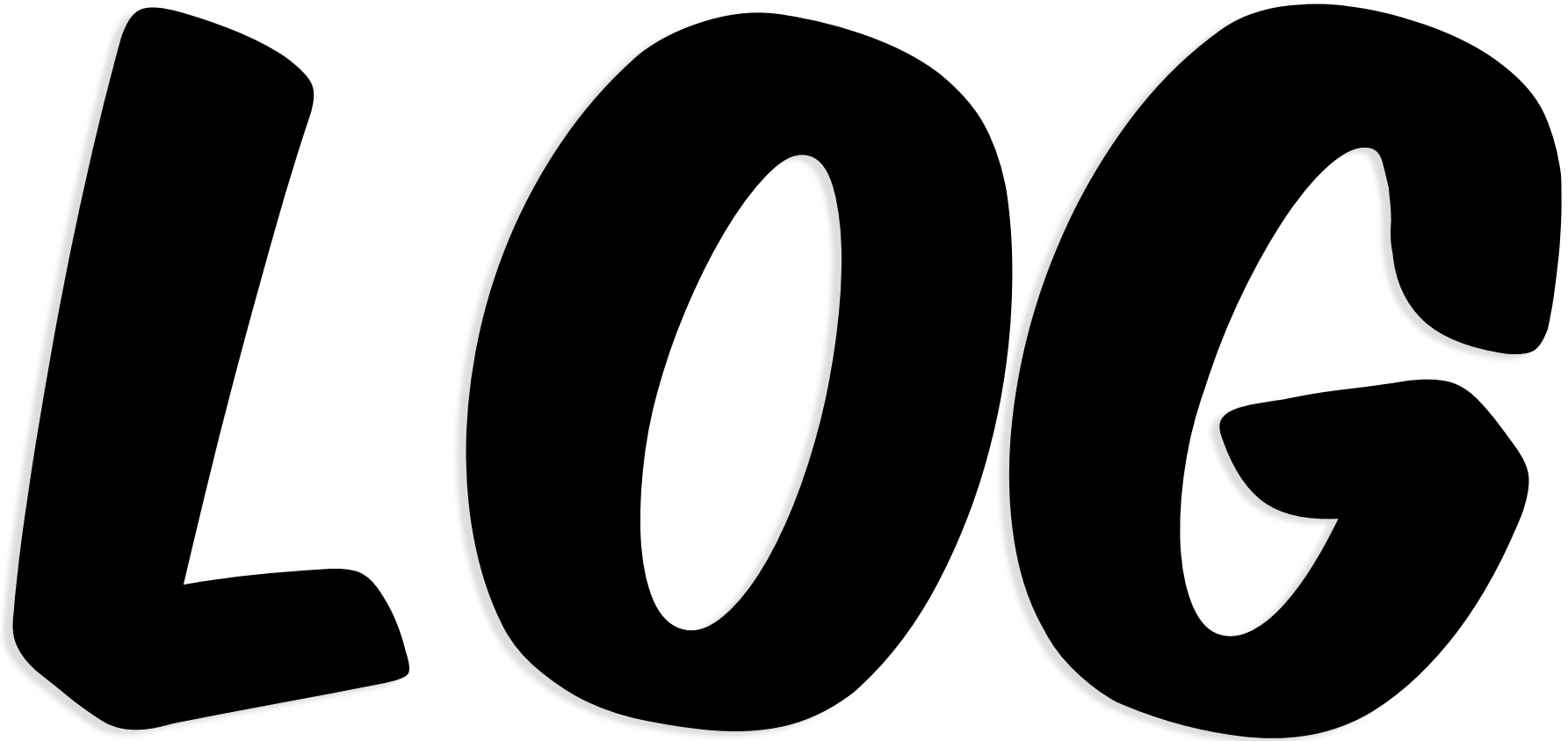


STORAGE

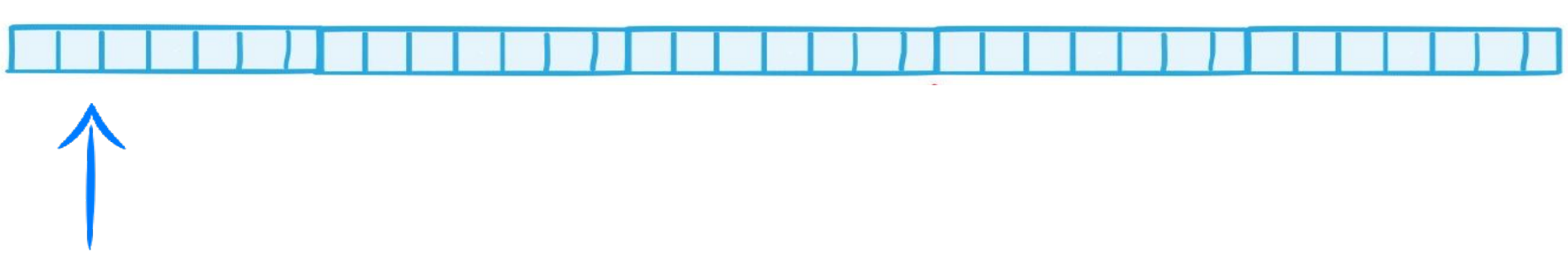
CORE ABSTRACTION

- DB - table
- Hadoop - file
- Kafka - ?



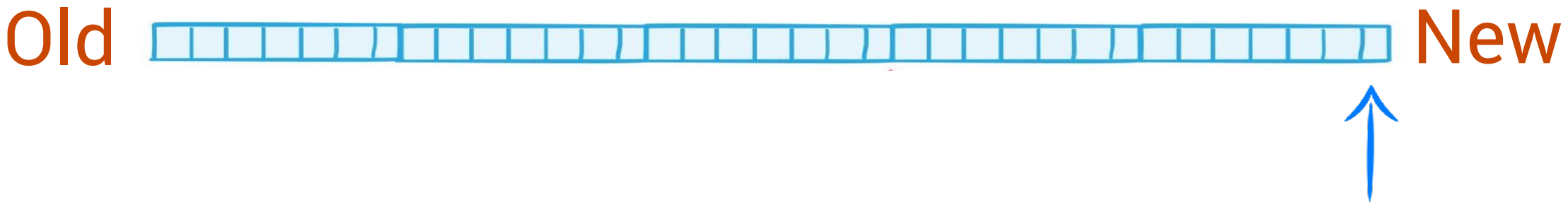

LOG

THE LOG IS A SIMPLE IDEA

Old  New

Messages are added at the end of the log

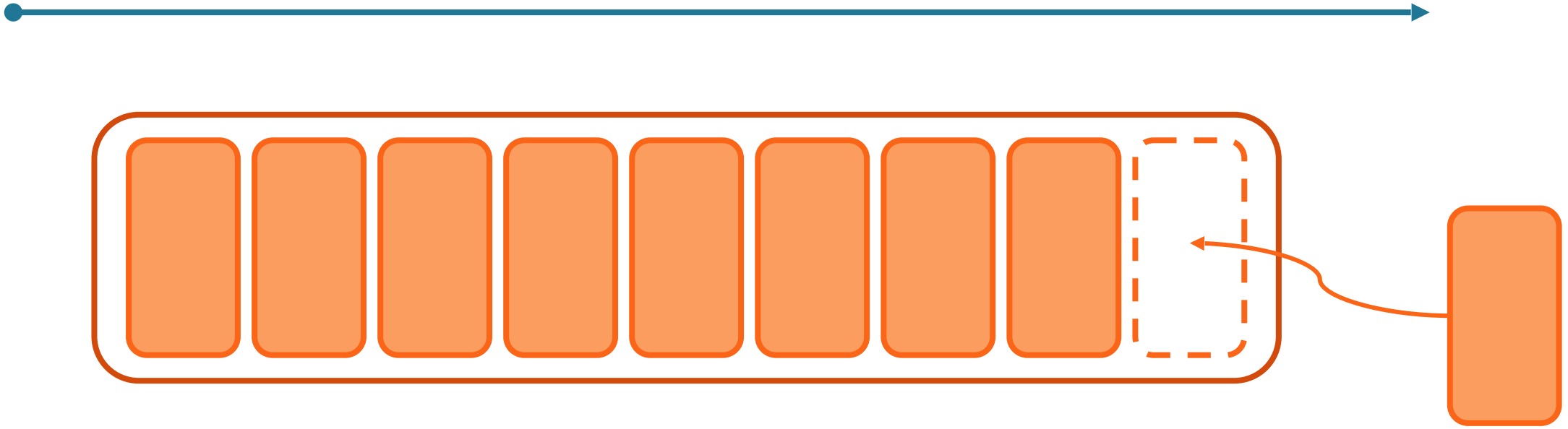
THE LOG IS A SIMPLE IDEA

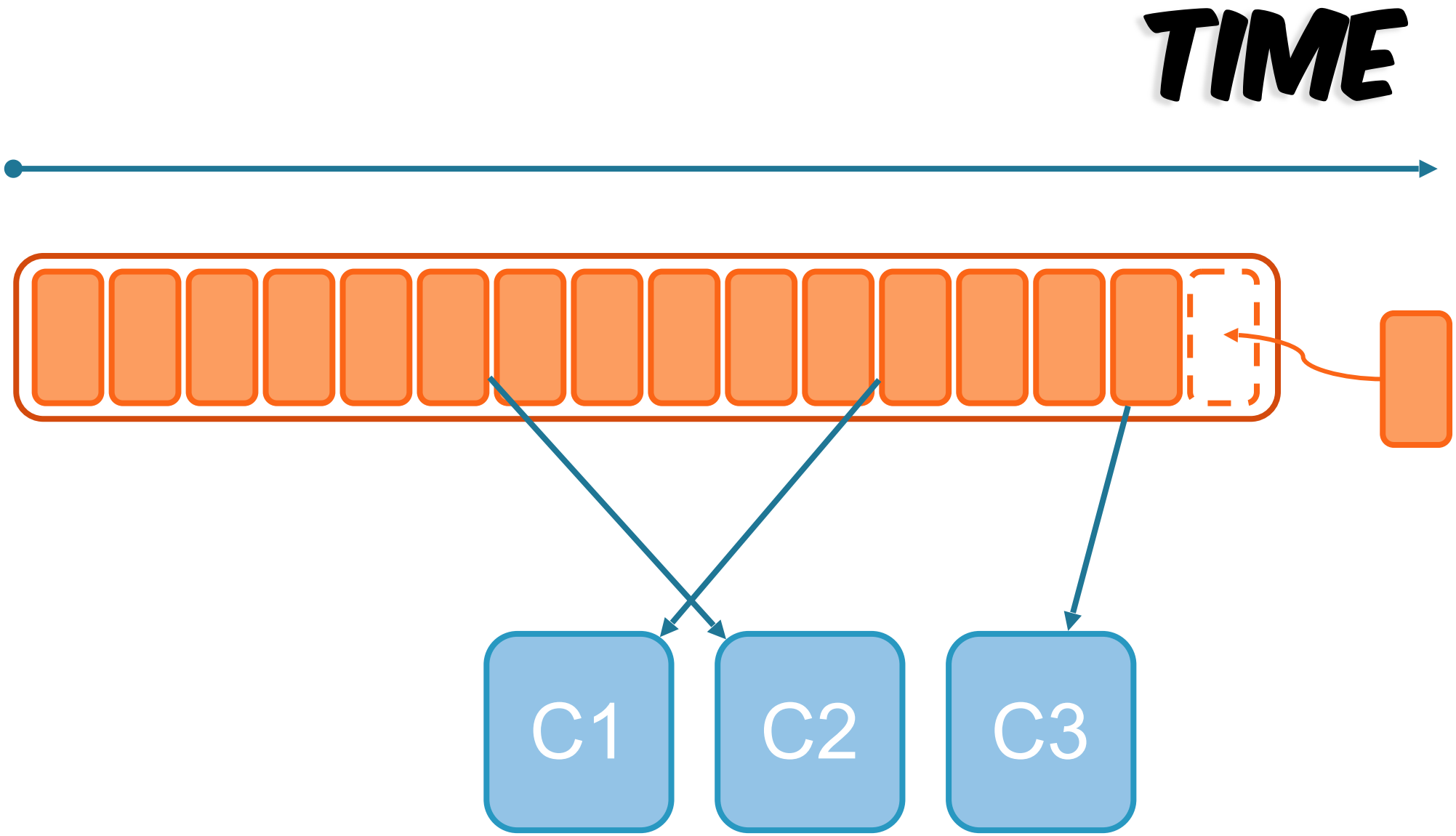


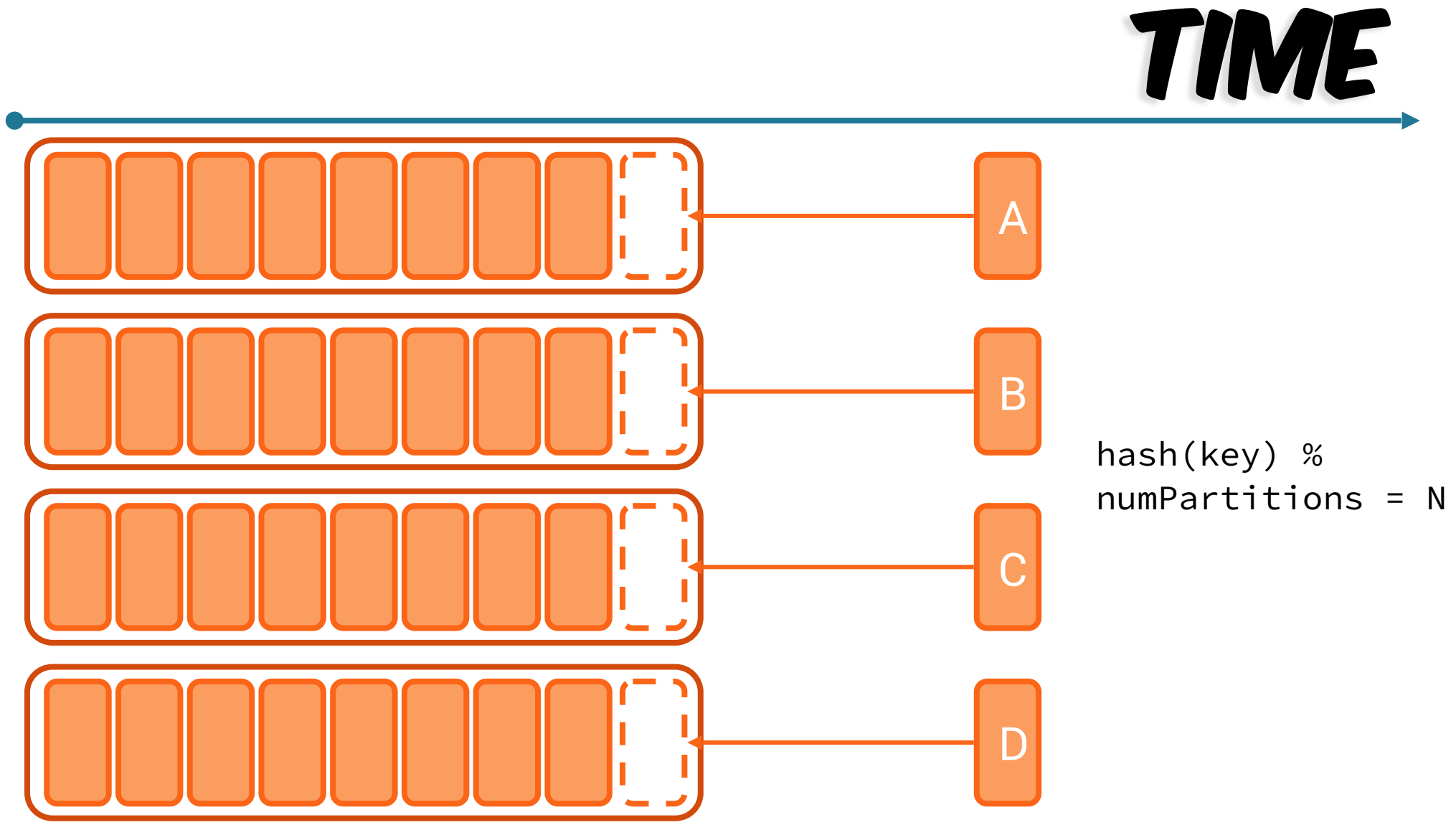
Messages are added at the end of the log

PUB / SUB

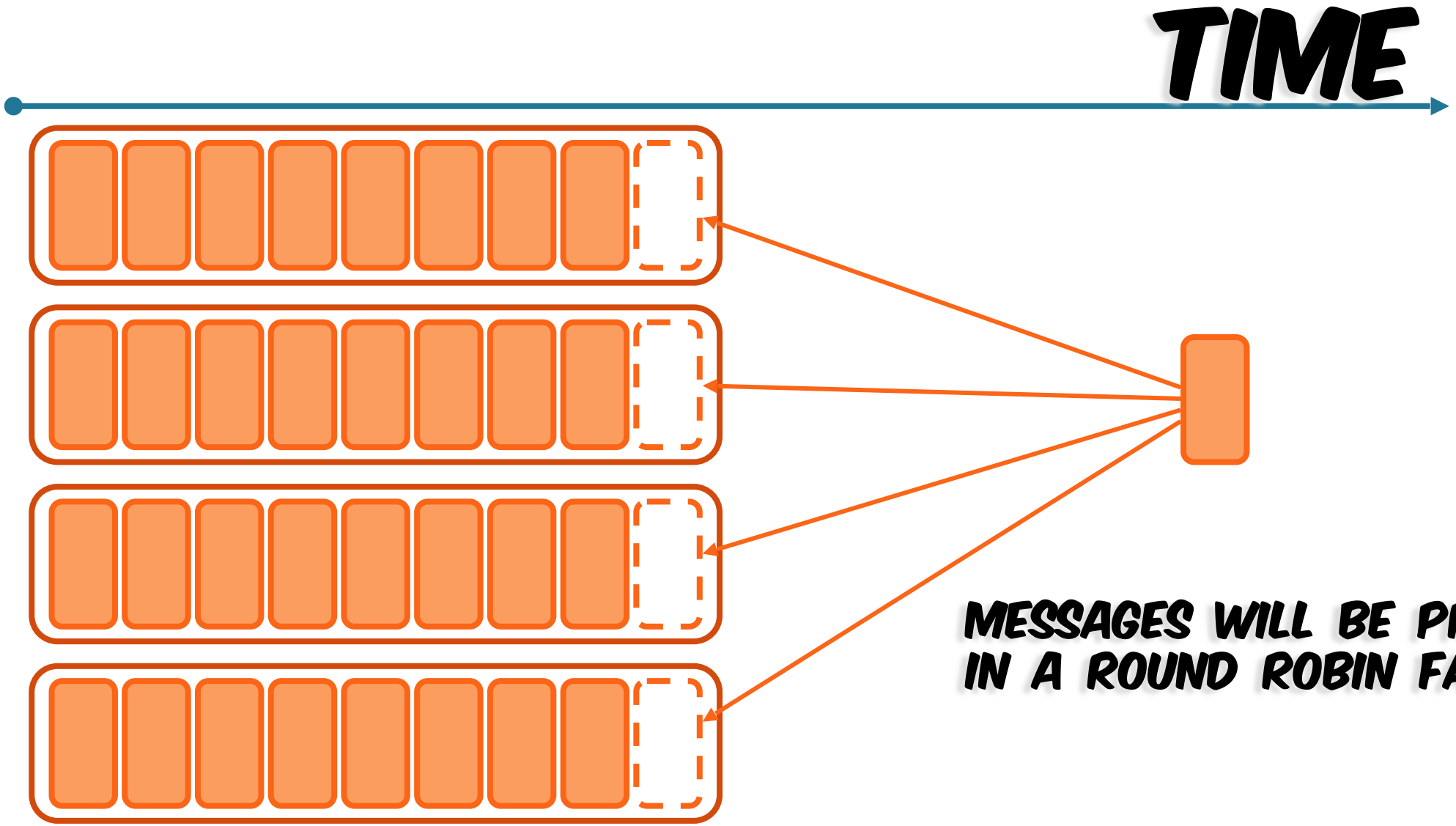
TIME




TIME
C1
C2
C3

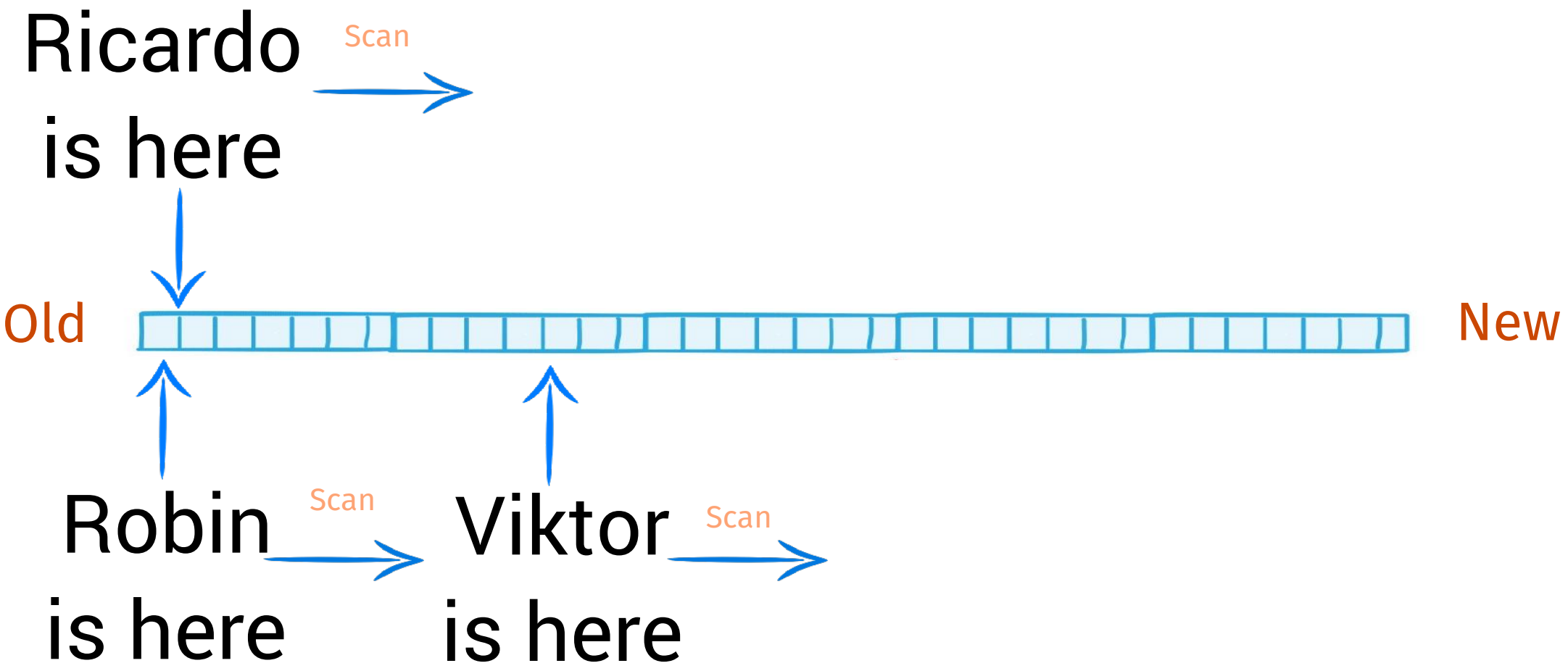



```
hash(key) %
numPartitions = N
```

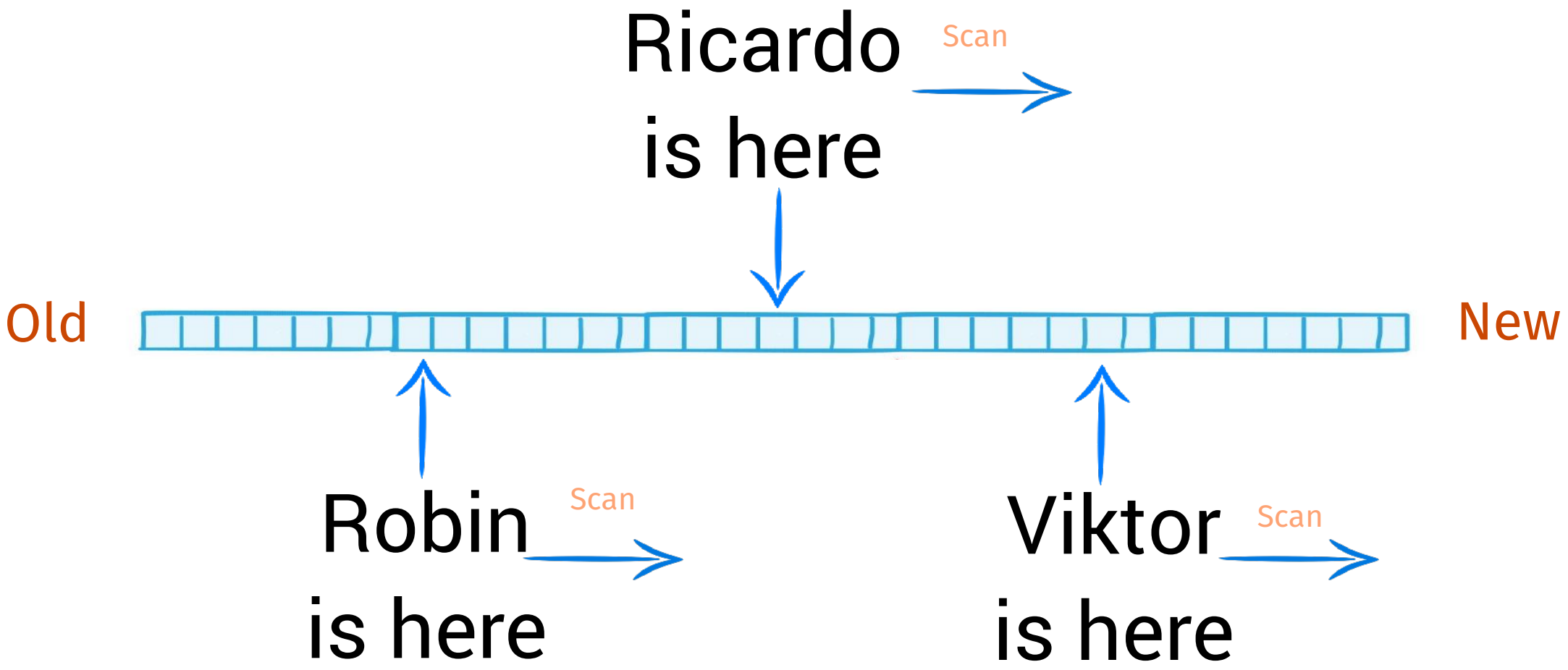


MESSAGES WILL BE PRODUCED IN A ROUND ROBIN FASHION

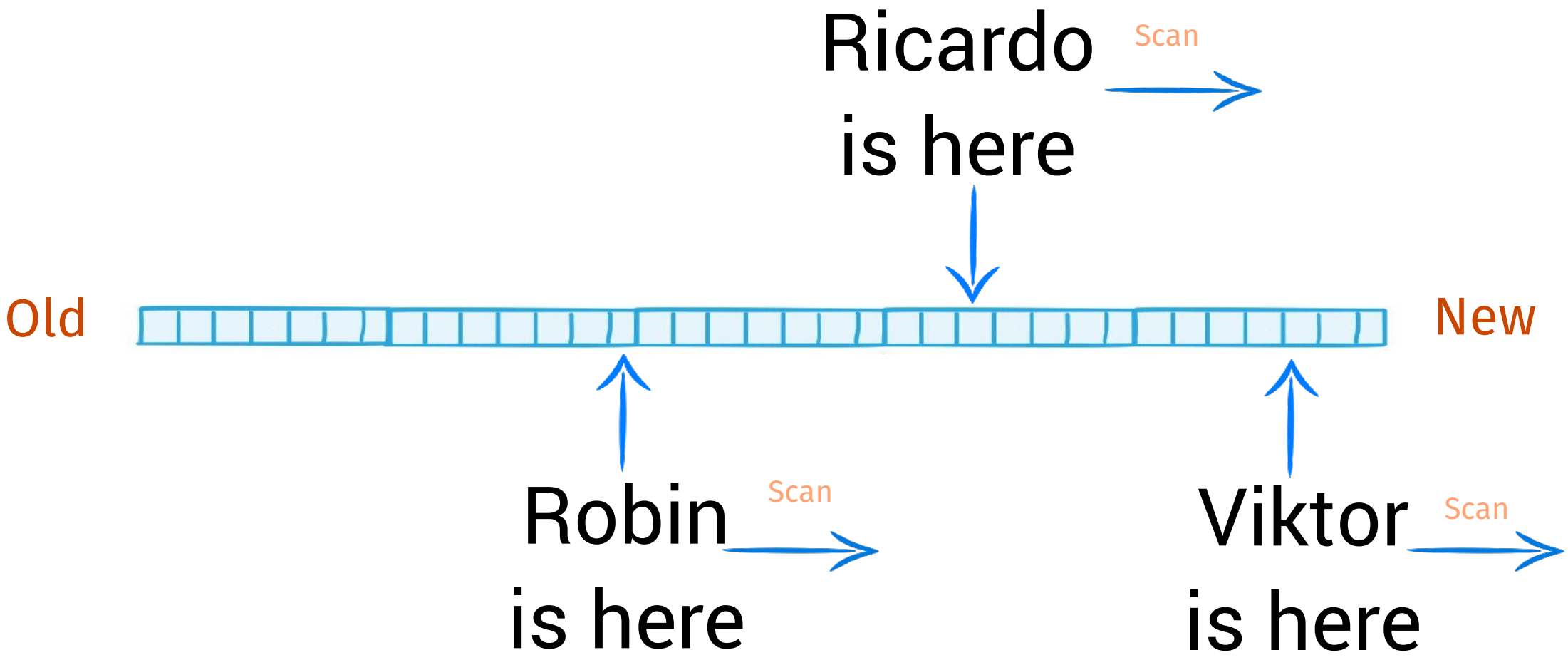
CONSUMERS HAVE A POSITION ALL OF THEIR OWN



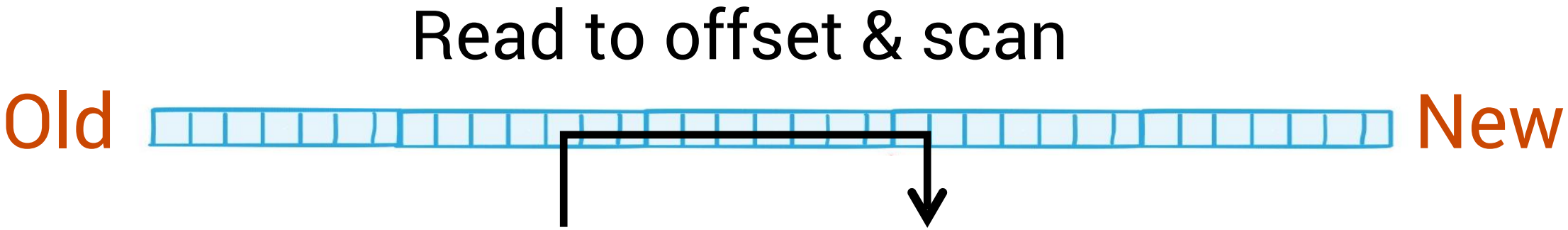
CONSUMERS HAVE A POSITION ALL OF THEIR OWN



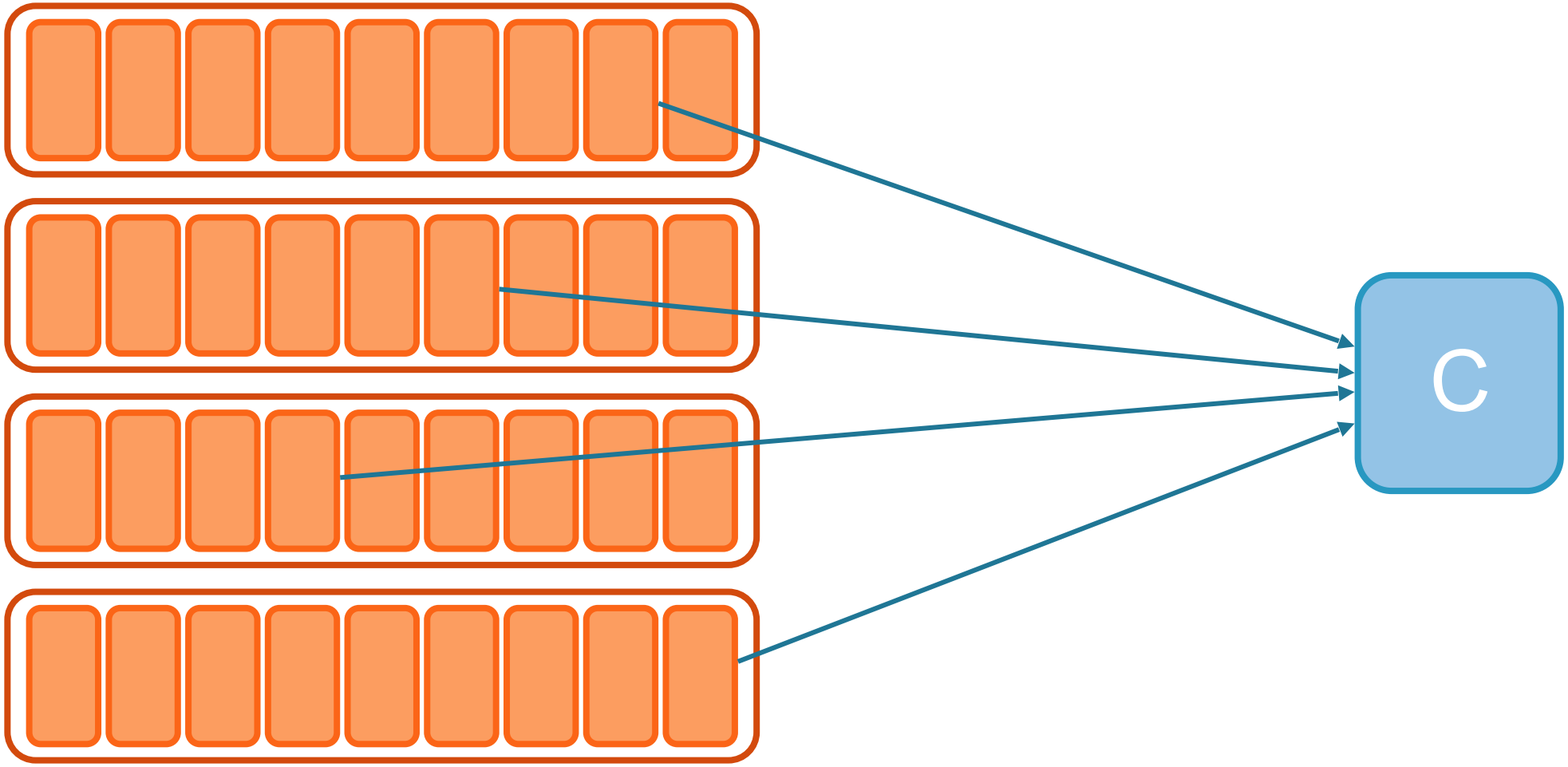
CONSUMERS HAVE A POSITION ALL OF THEIR OWN

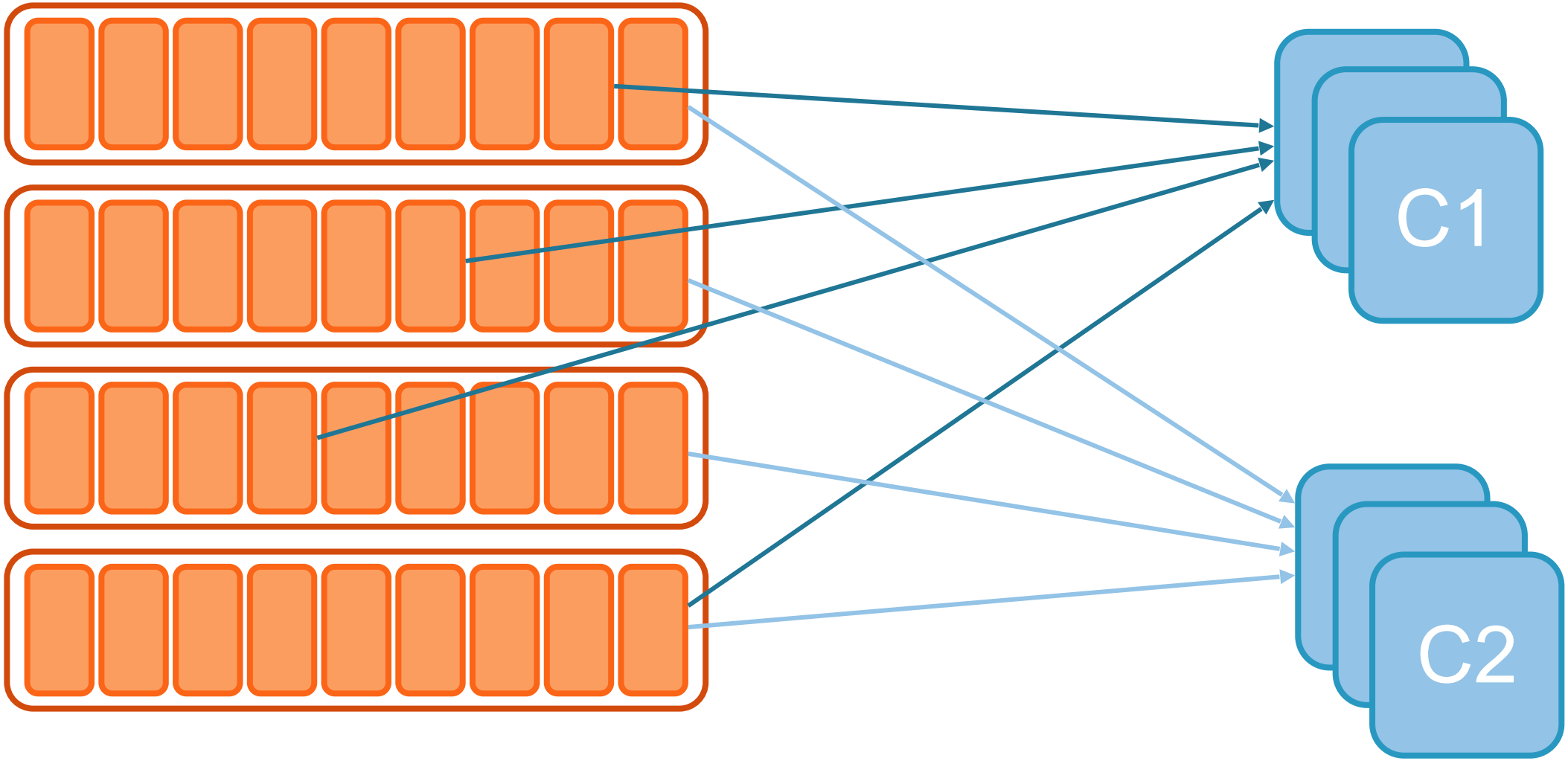


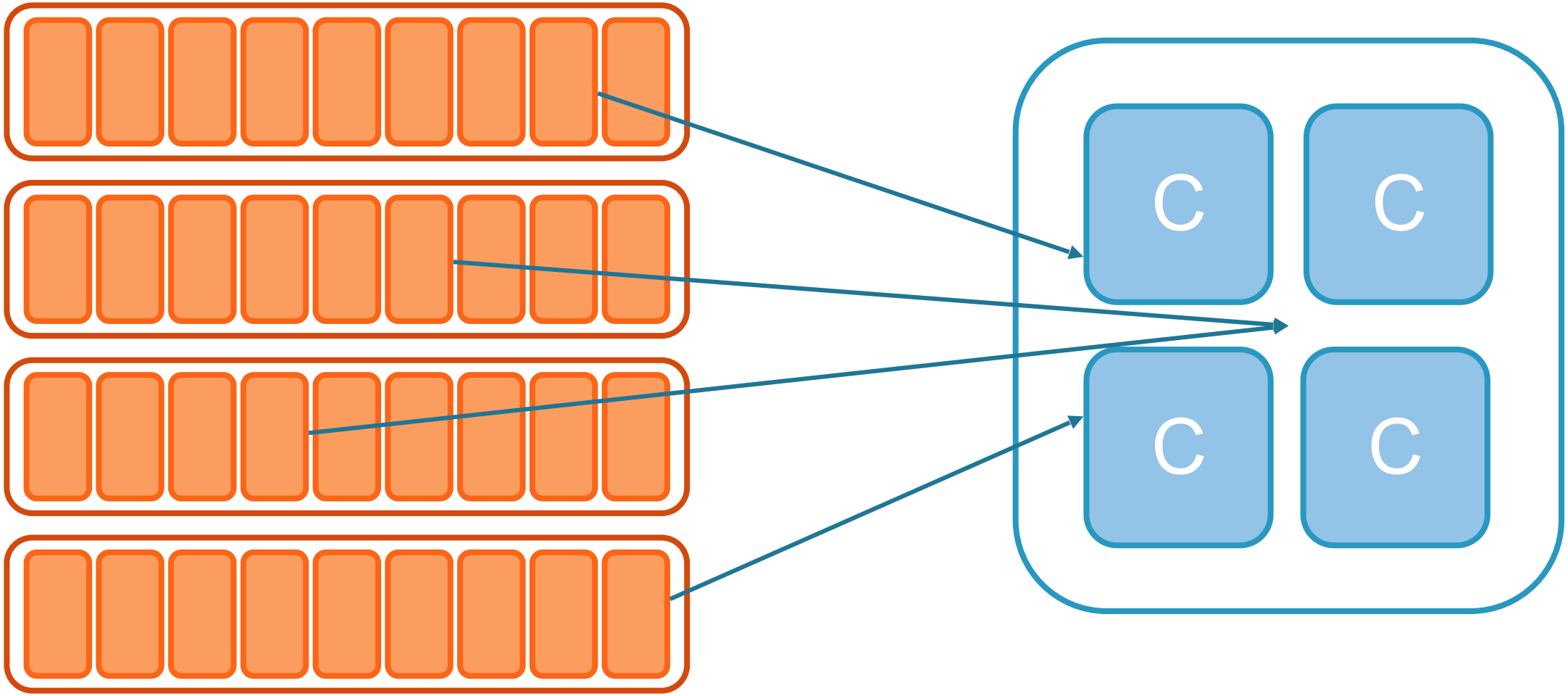
ONLY SEQUENTIAL ACCESS

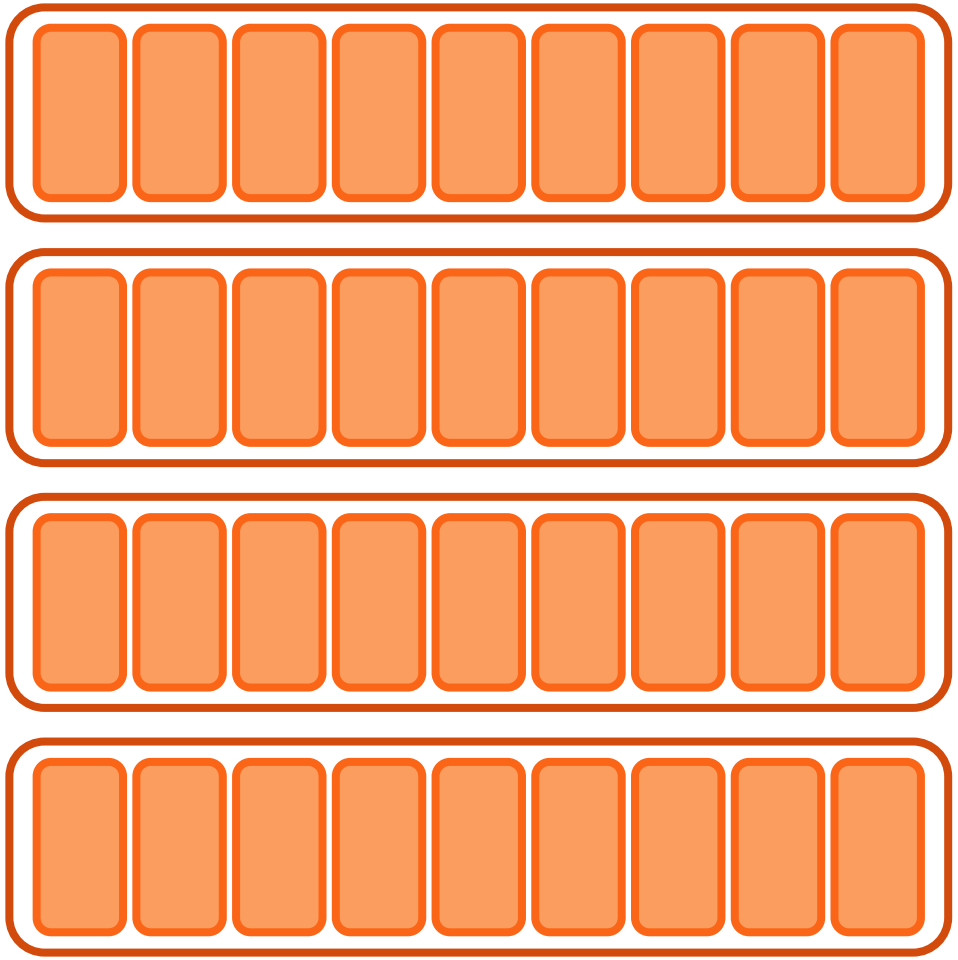


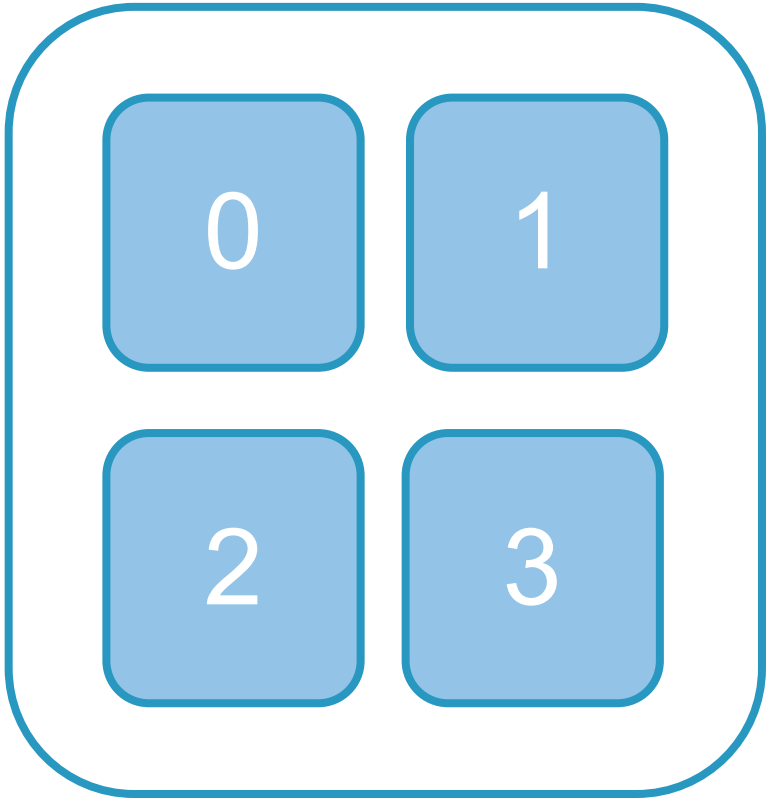

CONSUMERS
CONSUMER GROUP COORDINATOR
CONSUMER GROUP

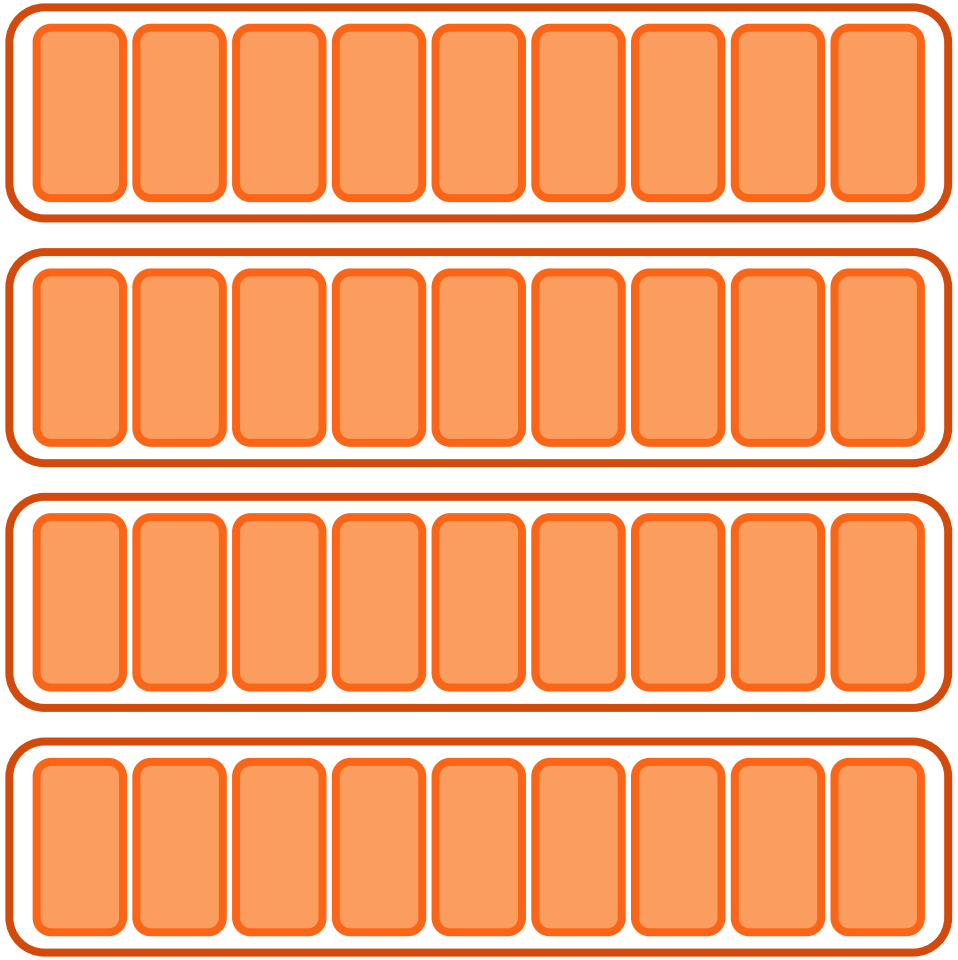

C

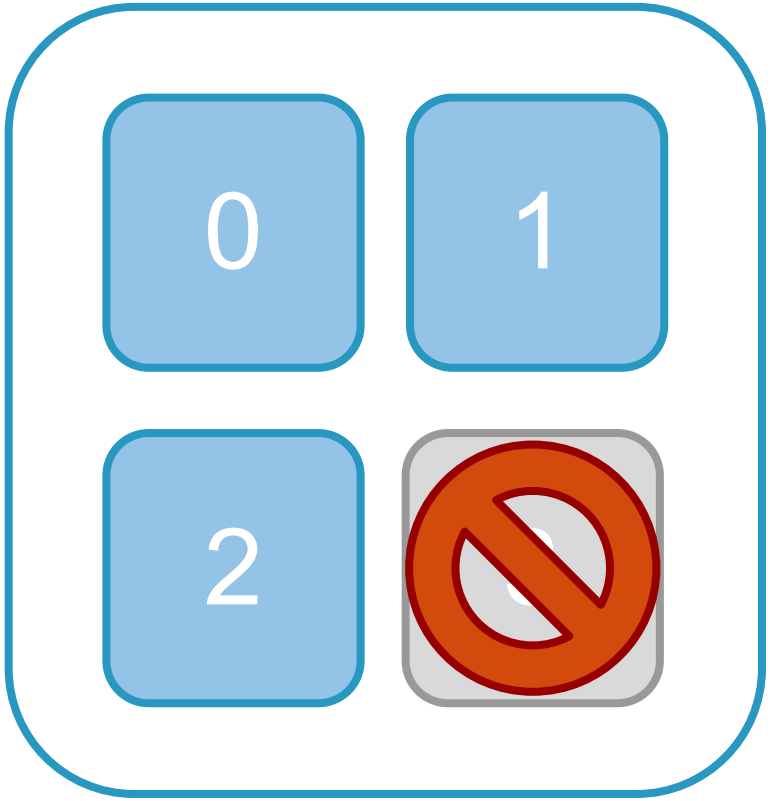

C1
C2

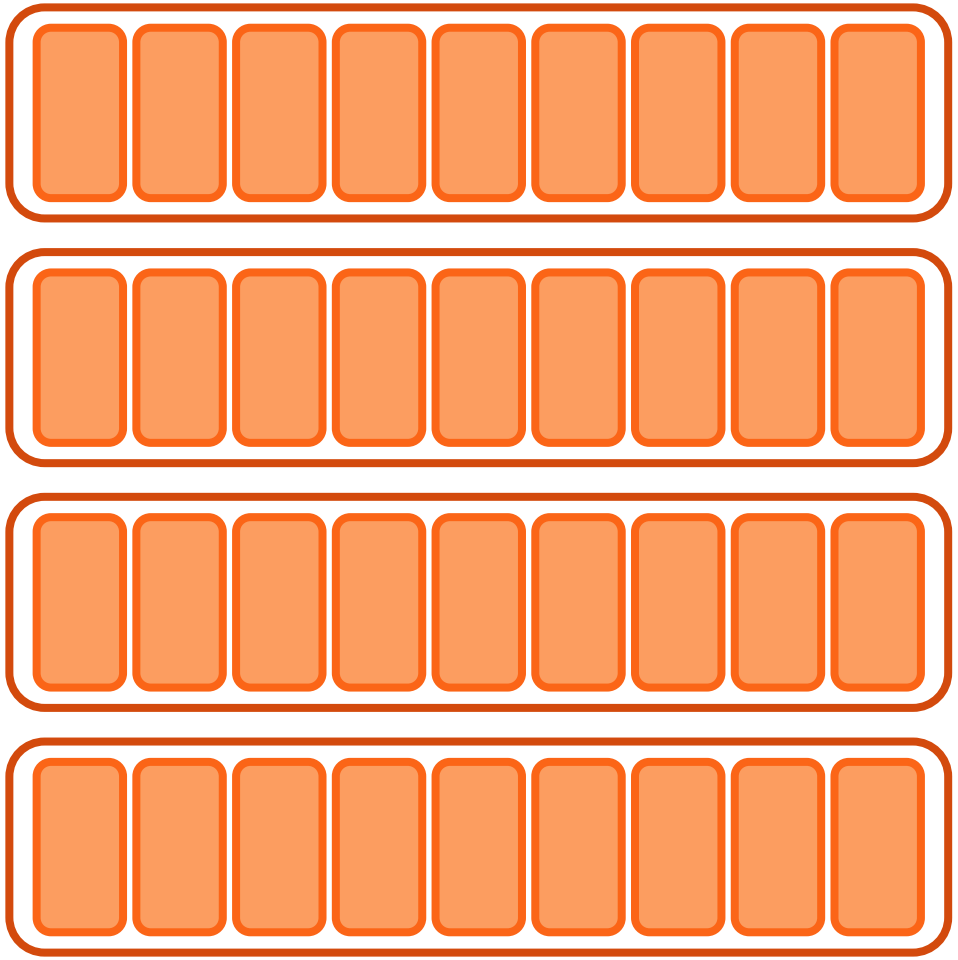

C
C
C
C

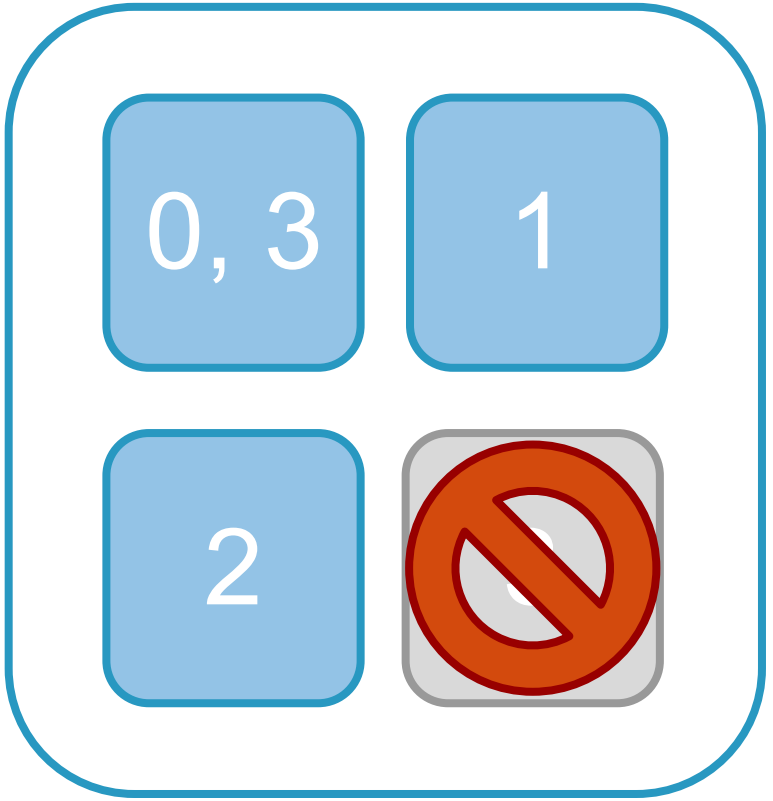



0
1
2
3



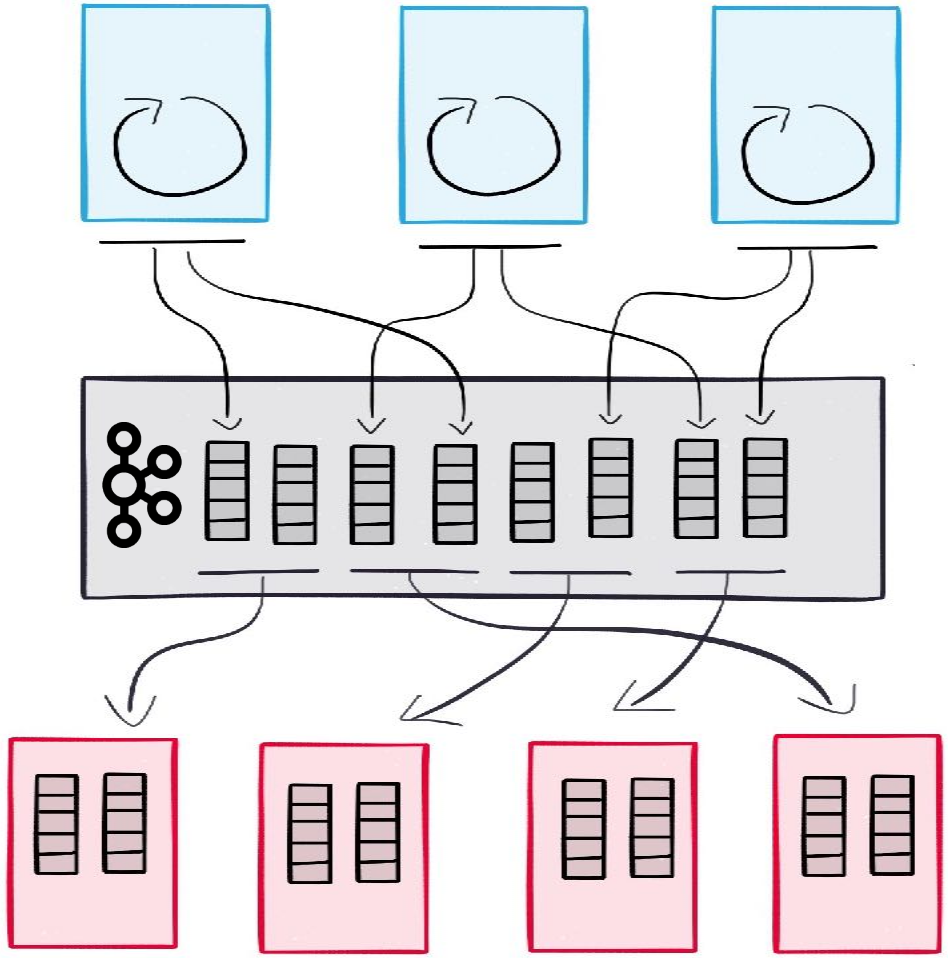

0
1
2




0, 3
1
2

LINEARLY SCALABLE ARCHITECTURE

Producers



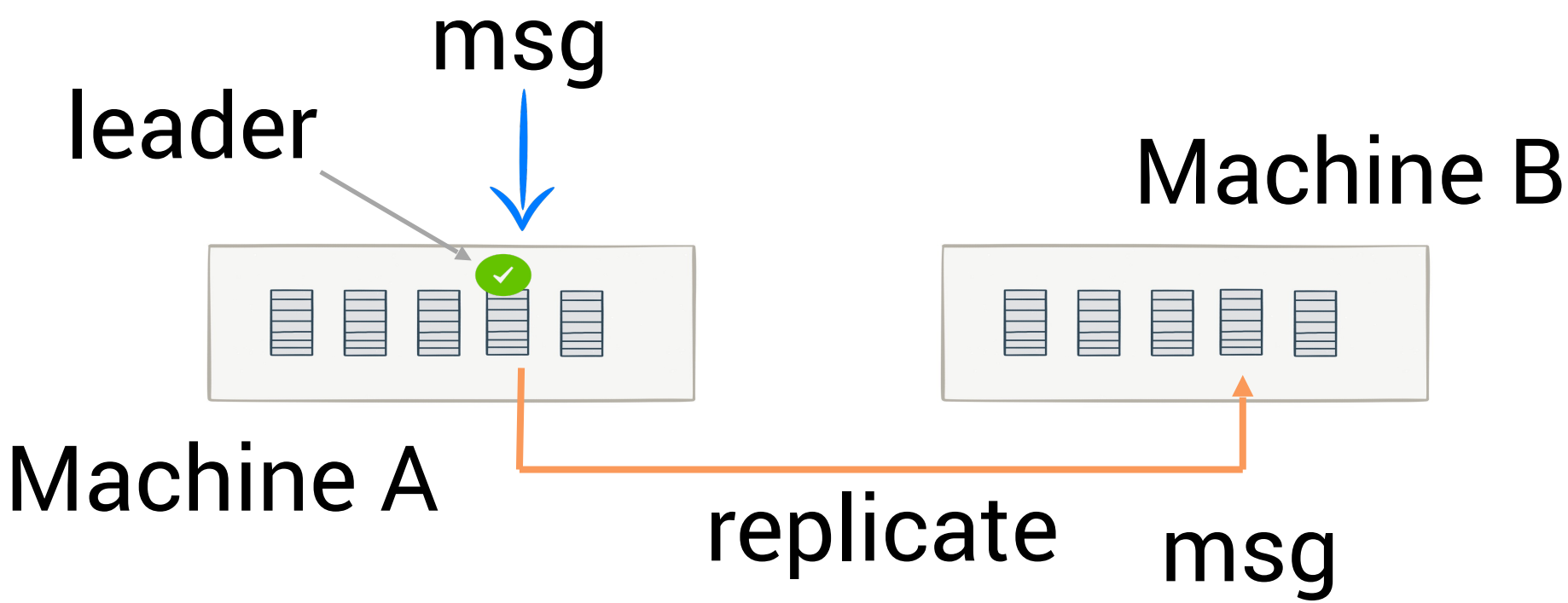
Consumers

Single topic:

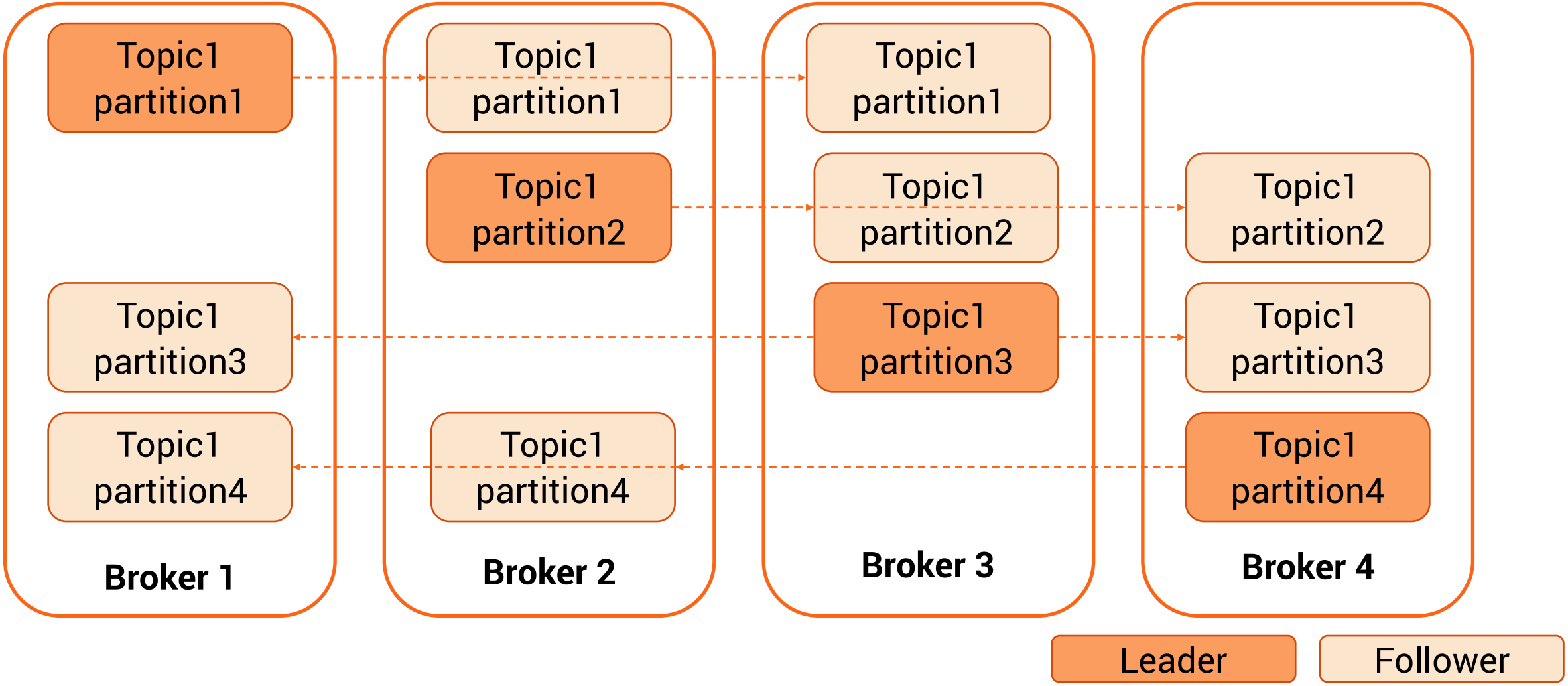
- Many producers machines
- Many consumer machines
- Many Broker machines

No Bottleneck!!

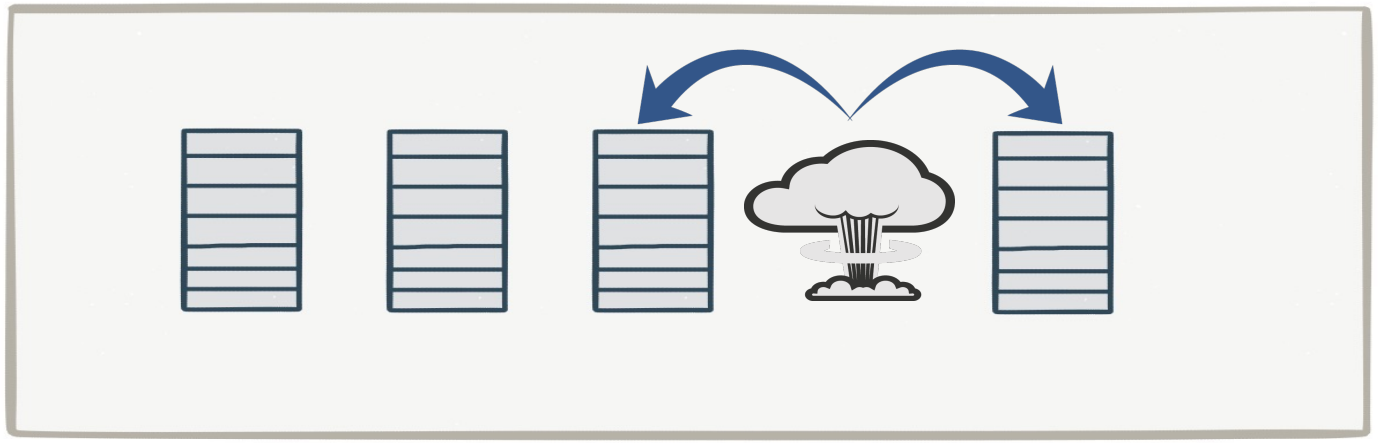
REPLICATE TO GET FAULT



PARTITION LEADERSHIP AND REPLICATION

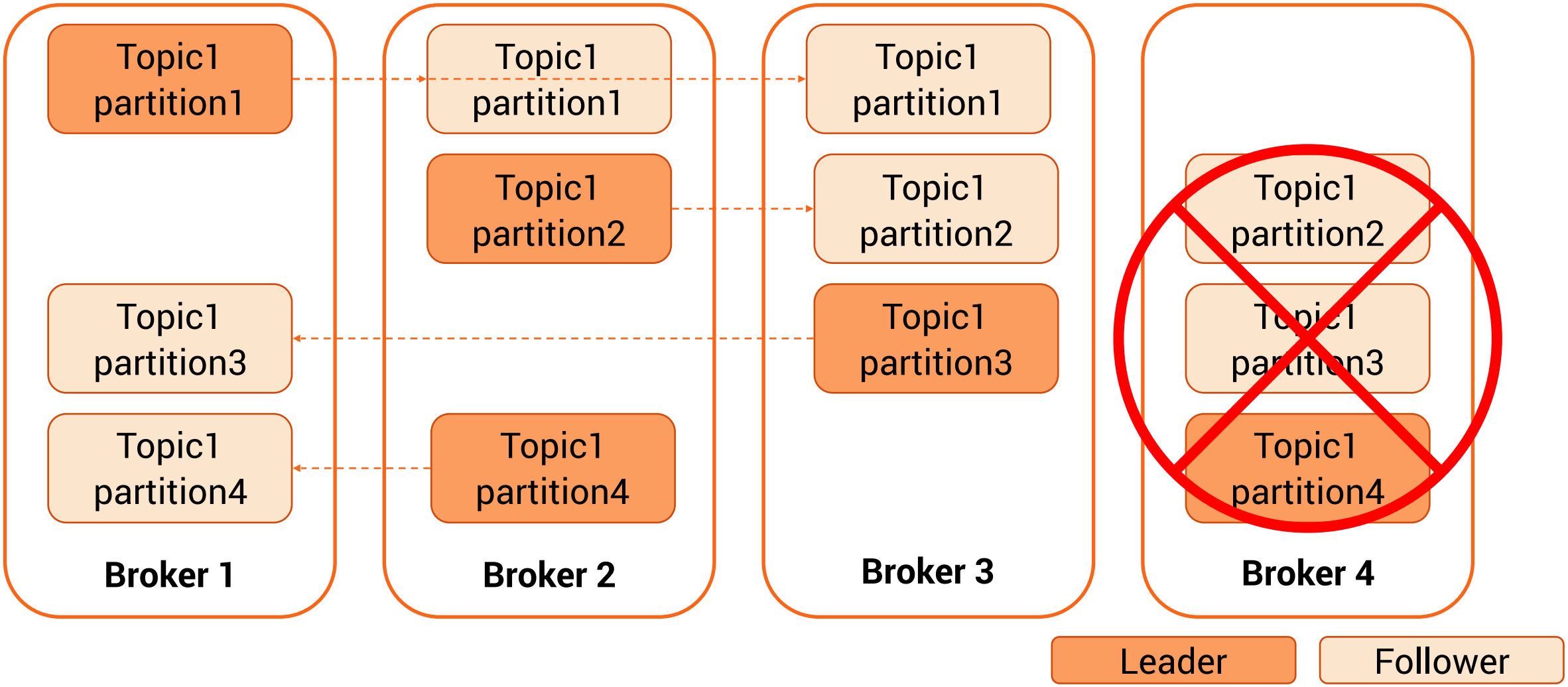


REPLICATION PROVIDES RESILIENCY



A replica takes over on machine failure

PARTITION LEADERSHIP AND REPLICATION - NODE FAILURE



THE LOG IS A TYPE OF DURABLE MESSAGING SYSTEM

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

(a) Far better scalability

(b) Built in fault tolerance / HA

(c) Storage

STOP!
DEMO TIME!

PROCESSING

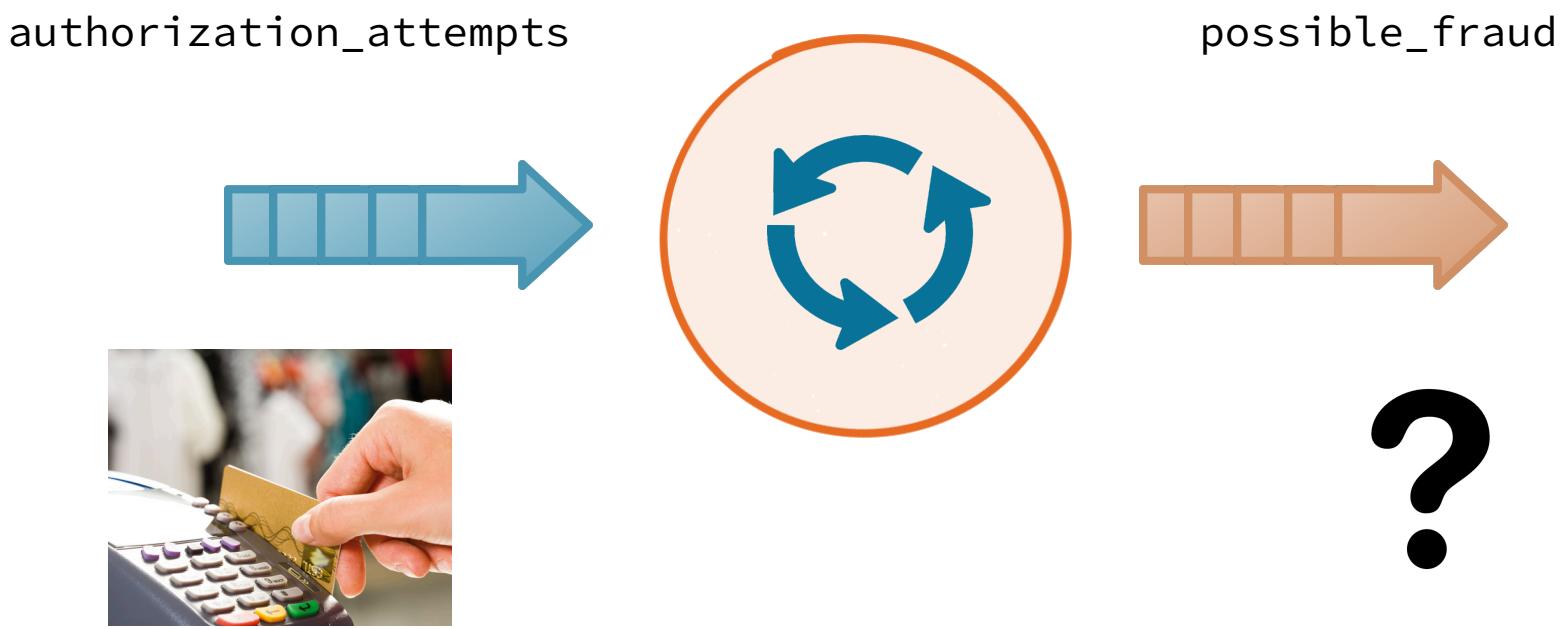
Streaming

is the **toolset** for dealing with **events**

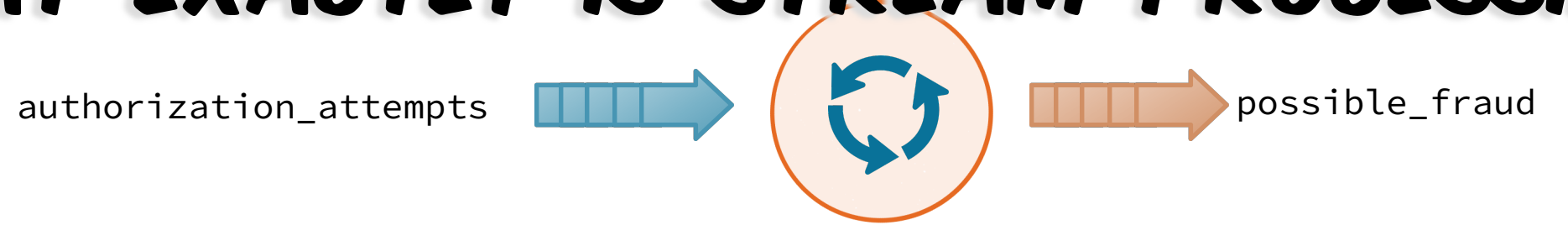
as **they move!**



WHAT EXACTLY IS STREAM PROCESSING?

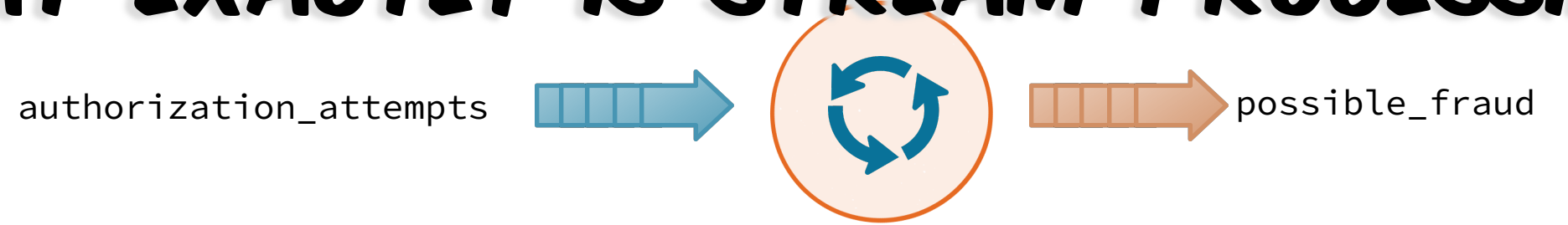


WHAT EXACTLY IS STREAM PROCESSING?



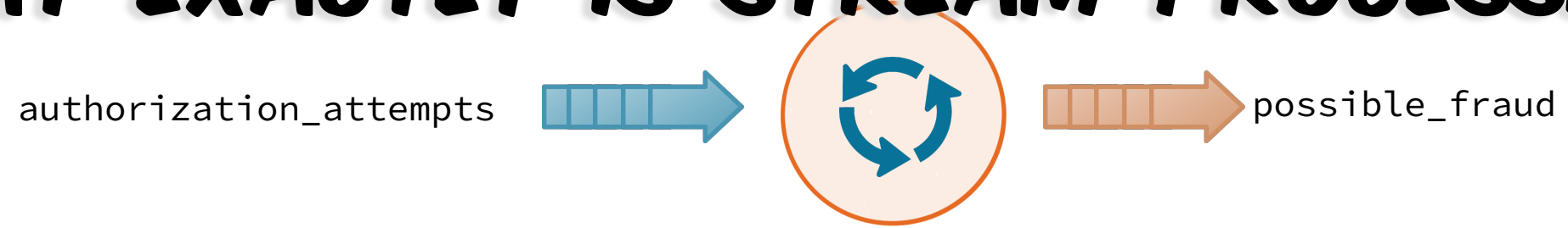

```
CREATE STREAM possible_fraud AS
SELECT card_number, count(*)
FROM authorization_attempts
WINDOW TUMBLING (SIZE 5 MINUTE)
GROUP BY card_number
HAVING count(*) > 3;
```

WHAT EXACTLY IS STREAM PROCESSING?



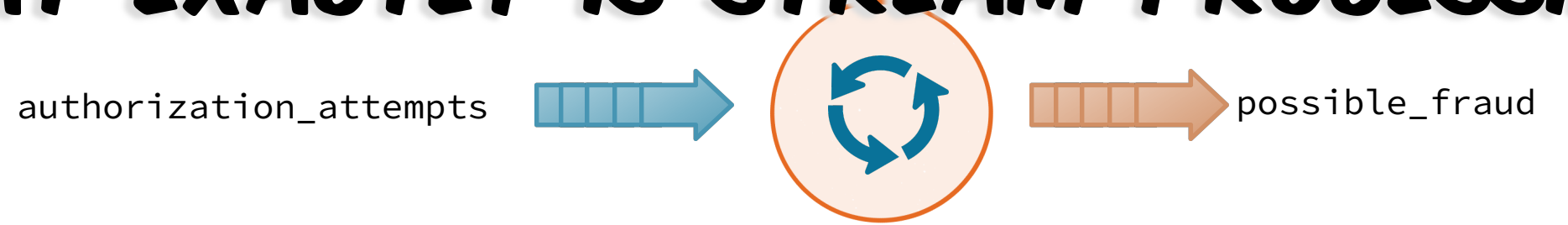

```
CREATE STREAM possible_fraud AS
SELECT card_number, count(*)
FROM authorization_attempts
WINDOW TUMBLING (SIZE 5 MINUTE)
GROUP BY card_number
HAVING count(*) > 3;
```

WHAT EXACTLY IS STREAM PROCESSING?



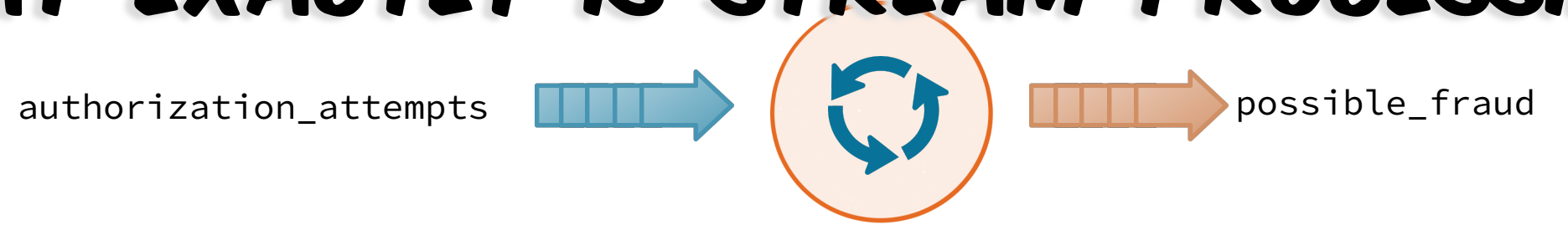

```
CREATE STREAM possible_fraud AS
SELECT card_number, count(*)
FROM authorization_attempts
WINDOW TUMBLING (SIZE 5 MINUTE)
GROUP BY card_number
HAVING count(*) > 3;
```

WHAT EXACTLY IS STREAM PROCESSING?



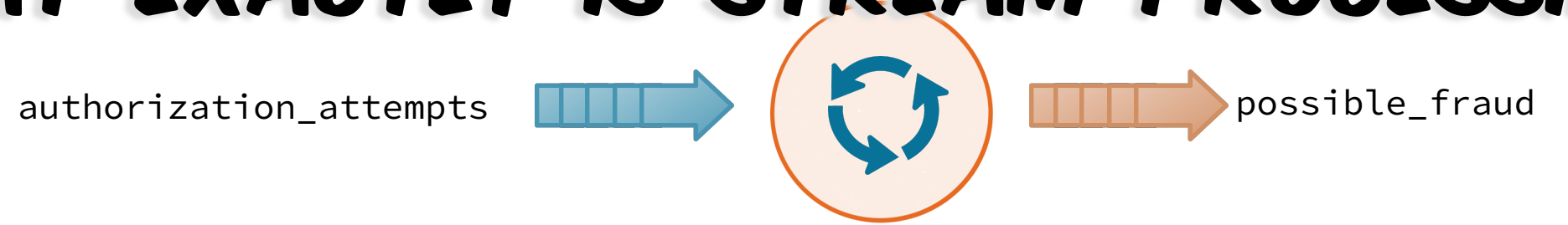

```
CREATE STREAM possible_fraud AS
SELECT card_number, count(*)
FROM authorization_attempts
WINDOW TUMBLING (SIZE 5 MINUTE)
GROUP BY card_number
HAVING count(*) > 3;
```

WHAT EXACTLY IS STREAM PROCESSING?



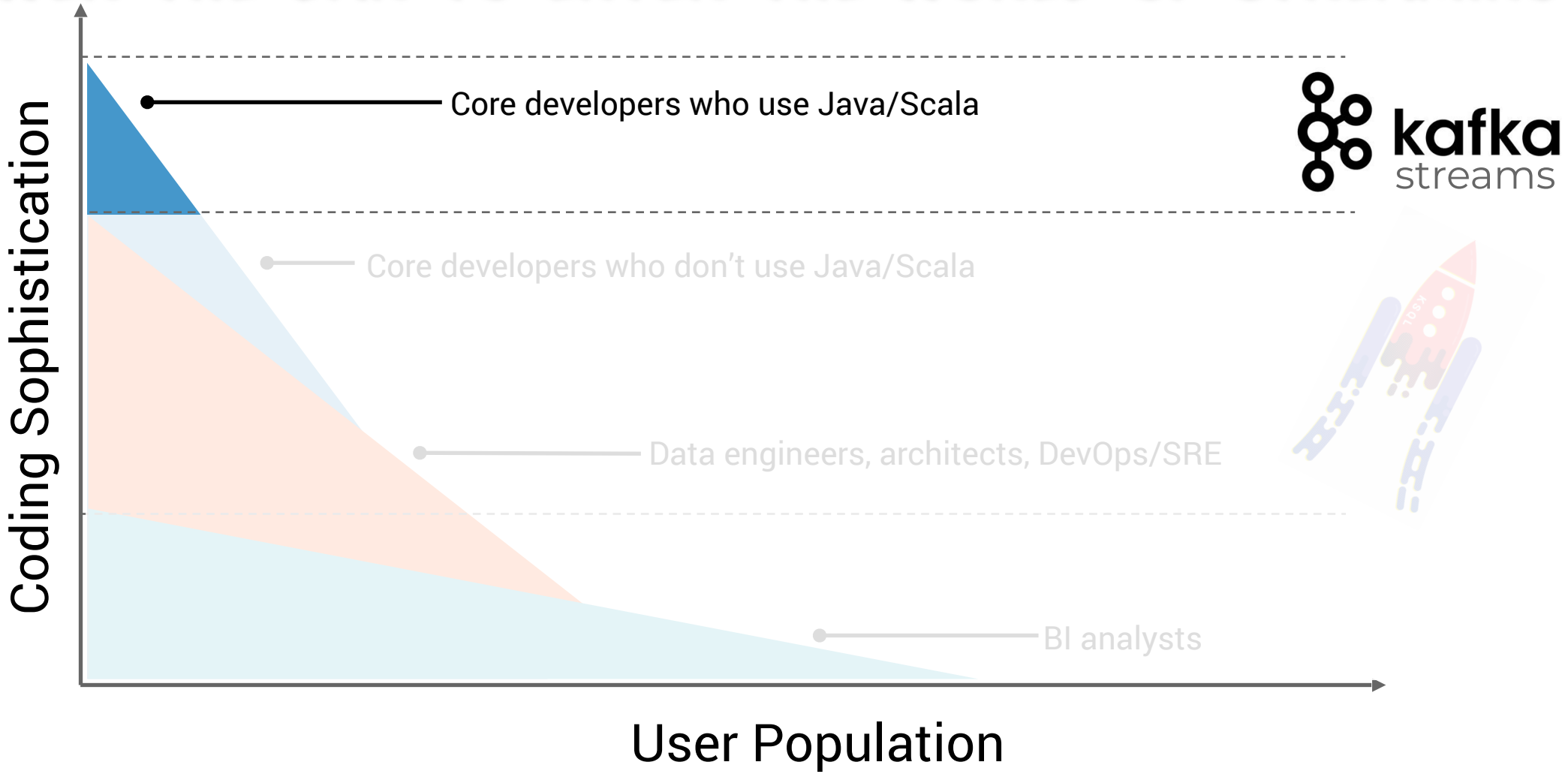

```
CREATE STREAM possible_fraud AS
SELECT card_number, count(*)
FROM authorization_attempts
WINDOW TUMBLING (SIZE 5 MINUTE)
GROUP BY card_number
HAVING count(*) > 3;
```

WHAT EXACTLY IS STREAM PROCESSING?

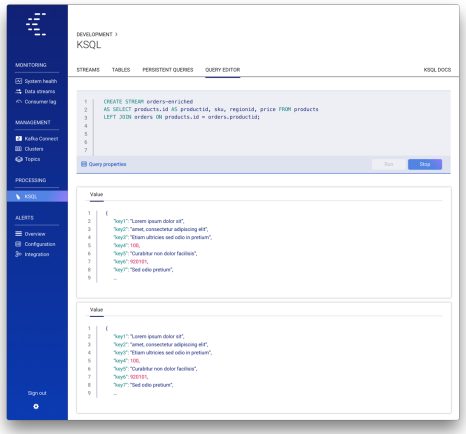



```
CREATE STREAM possible_fraud AS
SELECT card_number, count(*)
FROM authorization_attempts
WINDOW TUMBLING (SIZE 5 MINUTE)
GROUP BY card_number
HAVING count(*) > 3;
```

LOWER THE BAR TO ENTER THE WORLD OF STREAMING



KSQL #FTW



1 UI



2 CLI



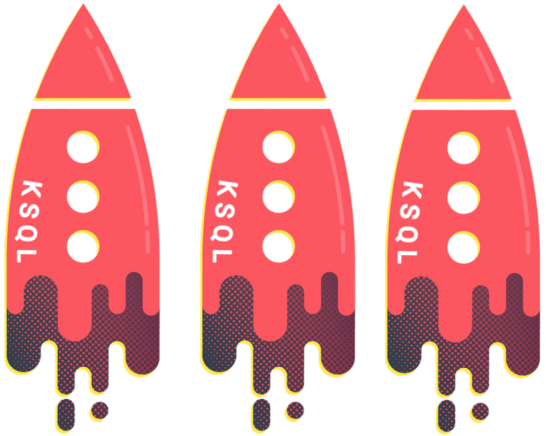
3 REST

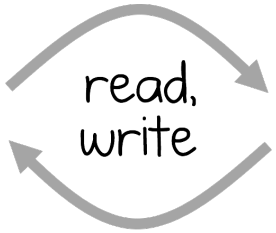


4 Headless

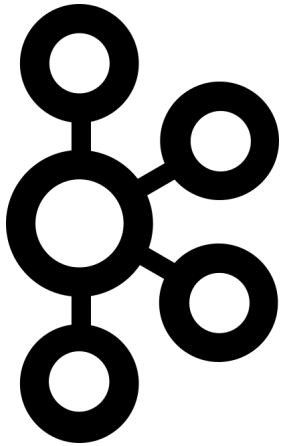
INTERACTION WITH KAFKA

KSQL
(processing)





Kafka
(data)





Application
(processing) Jva/KStreams, .NET



Does not run on Kafka brokers

Does not run on Kafka brokers



Find your local Meetup Group

https://cnfl.io/kafka-meetups

Grab Stream Processing books



https://cnfl.io/book-bundle



Join us in Slack

http://cnfl.io/slack



THANKS!

@gamussa
viktor@confluent.io