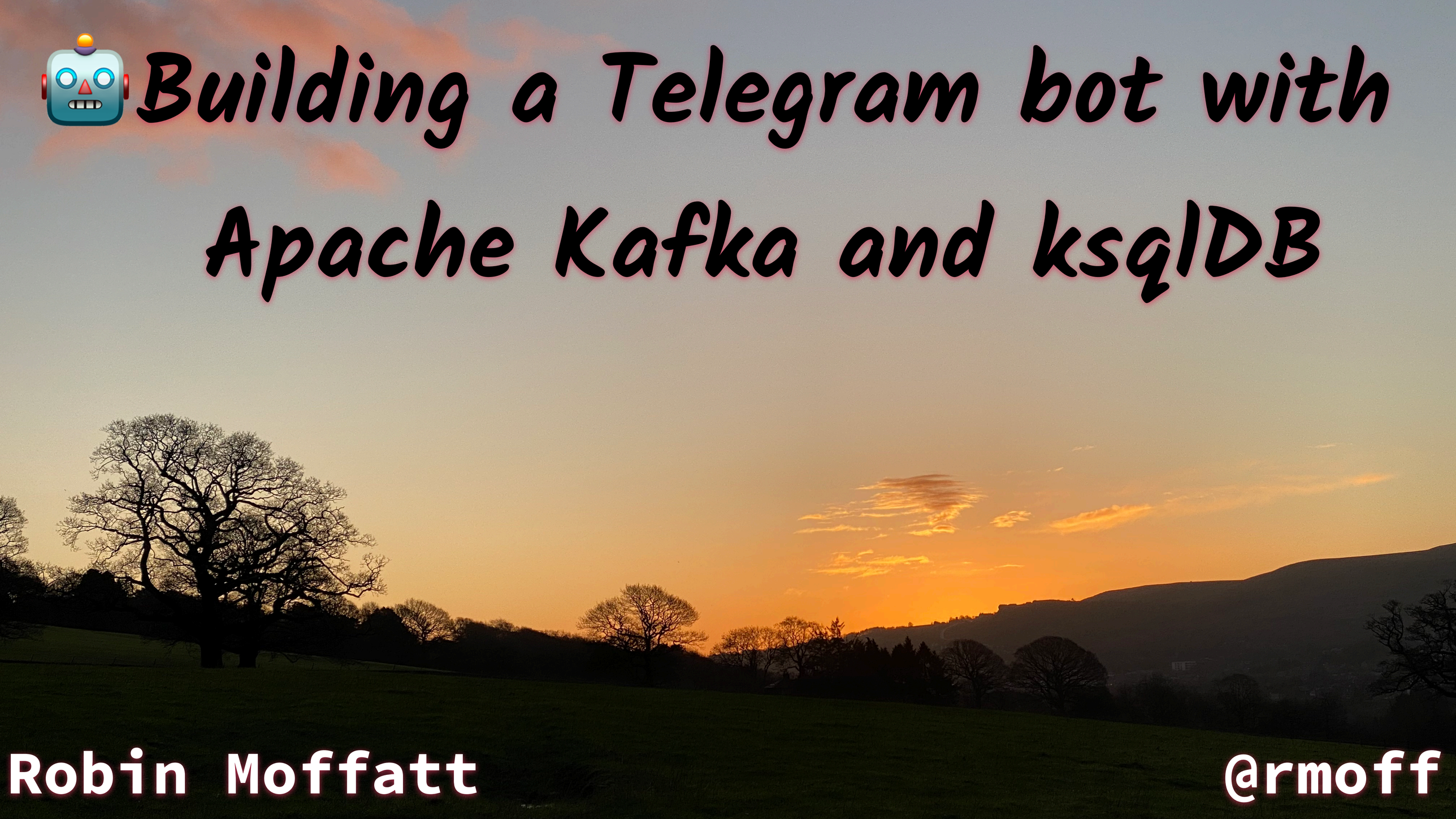
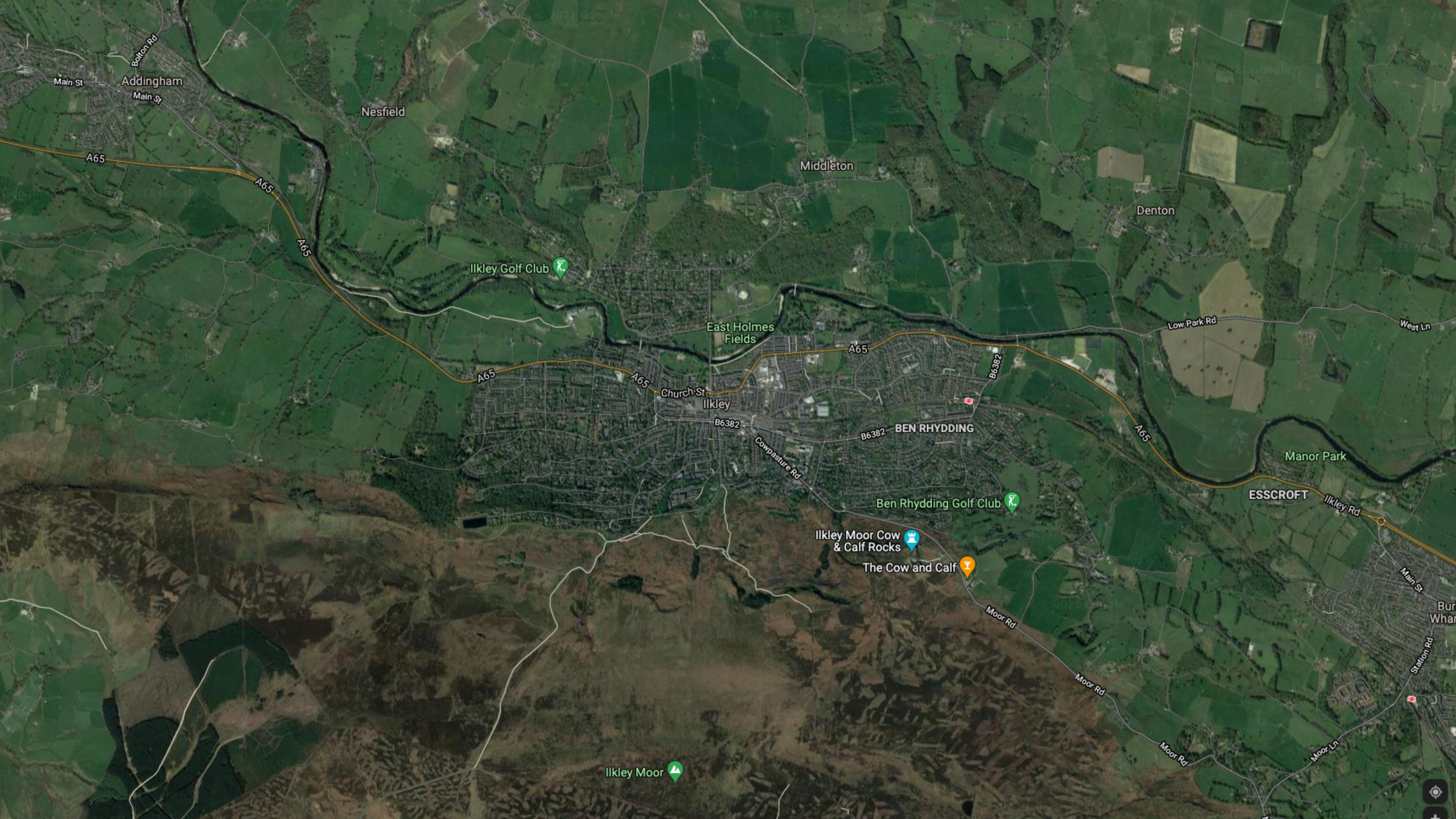

🤖Building a Telegram bot with Apache Kafka and ksqlDB
Robin Moffatt
@rmoff


Addingham
Main St
Bolton Rd
Nesfield
A65
Middleton
Denton
Ilkley Golf Club
East Holmes Fields
Low Park Rd
West Ln
Church St
Ilkley
B6382
BEN RHYDDING
Cowpasture Rd
Manor Park
ESSCROFT
Ilkley Rd
Ben Rhydding Golf Club
Ilkley Moor Cow & Calf Rocks
The Cow and Calf
Moor Rd
Main St
Station Rd
Moor Ln
Ilkley Moor


SWAIN HOUSE
RAVENSCLIFFE
HEATON
BOLTON OUTLANES
Woodhall Hills Golf
West Bradford Golf Club
Lister Park
ECCLESHILL
UNDERCLIFFE
Bradford Industrial Museum
Calverley
DAISY HILL
Bradford Royal Infirmary
MANNINGHAM
FAGLEY
Peel Park, Bradford
Undercliffe Cemetery, Bradford
Allerton
GIRLINGTON
WHITE ABBEY
BLACK ABBEY
ODEON Luxe Leeds Bradford
Myra Shay
LONGLANDS
WAPPING
BARKEREND
LOWER GRANGE
CROSSLEY HALL
Bradford
EASTBROOK
GOITSIDE
NEW LEEDS
HILL TOP
National Science and Media Museum
TYERSAL
LIDGET GREEN
SHEARBRIDGE
LAISTERDYKE
BROOMFIELDS
LITTLE HORTON
Clayton
Horton Park, Bradford
HOLME TOP
BOWLING
CANTERBURY
CUTLER HEIGHTS
GREAT HORTON
HOLME WOOD
Horton Bank
Bowling Park
Allerton Rd
Thornton Rd
Leeds Rd
Wakefield Rd
Manningham Ln
Canal Rd
Harrogate Rd
Bowling Back Ln
Great Horton Rd
Little Horton Ln
Clayton Rd
Bradford Rd
Cemetery Rd
Legrams Ln
A6177
A658
A6181
A641


N C P Welcome to Hall Ings Car Park
BRADFORD CAPITAL OF CYCLING

Where's my nearest carpark with available spaces?

How many spaces are available in this car park?

💡Tell me when a car park with spaces is available

How does occupancy vary over time?

NORTHERN DATA HUB

- Datasets
- Products
- District Dashboard (visual)
- Latest CQC Results (visual)
- Other open data sites
- Contact

Bradford car parks

City of Bradford Metropolitan District Council

Resources from the City of Bradford Metropolitan District Council (CBMDC) Parking service.

1 - Car park locations
a simple csv containing name and location including latitude / longitude

2 - Car park current status.
API that returns a csv dataset of the current status of 8 Bradford city centre car parks. The dataset returns capacity, empty places, status together with location details.
The dataset is updated every 3 minutes for a live view of spaces in these car parks.

3 - Car park historic status
API that returns a csv dataset building up the historic status of the 8 city centre car parks.
The dataset is updated every 30 minutes.

4 Resources ...

csv api geojson

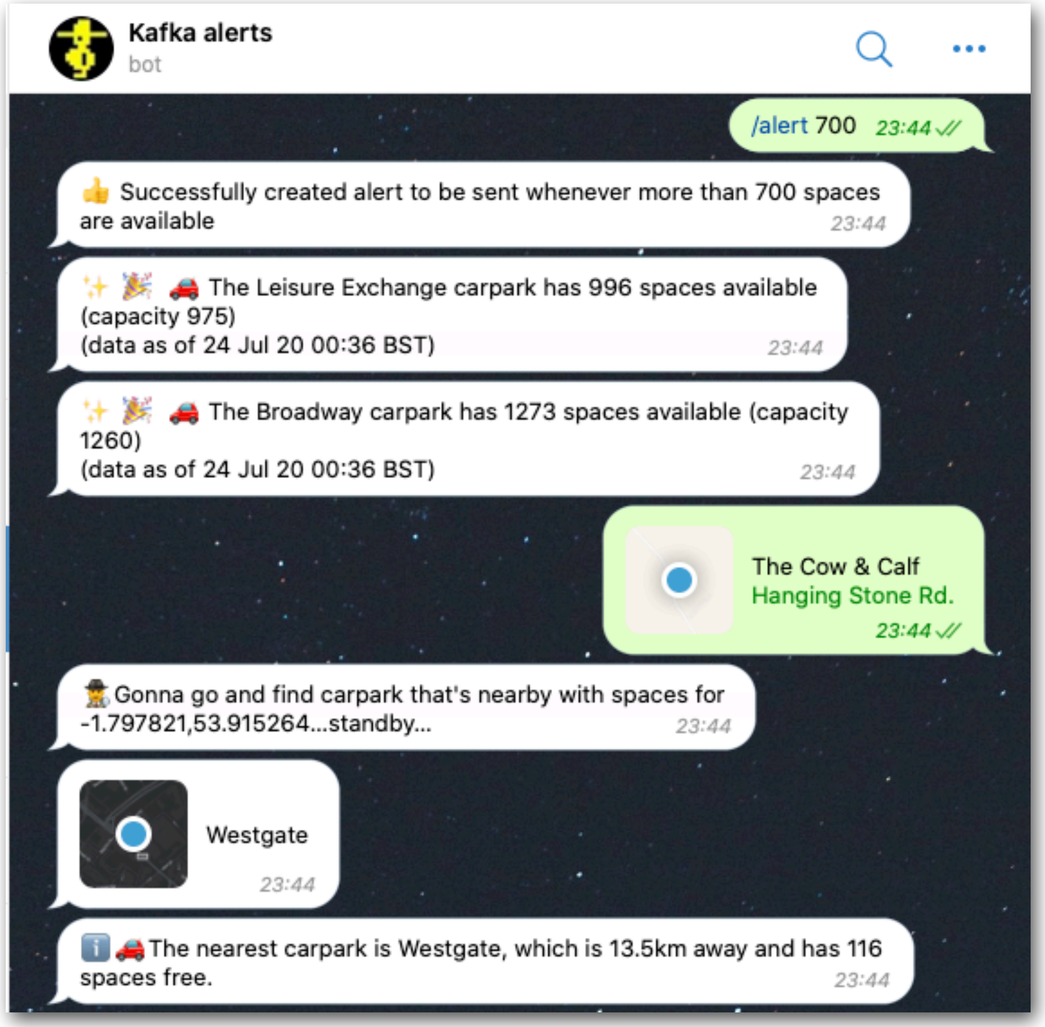
More Information

Map of car parks across Bradford district

html Live map of Bradford car parks

License UK Open Government Licence (OGL v3) **Frequency** daily

Telegram

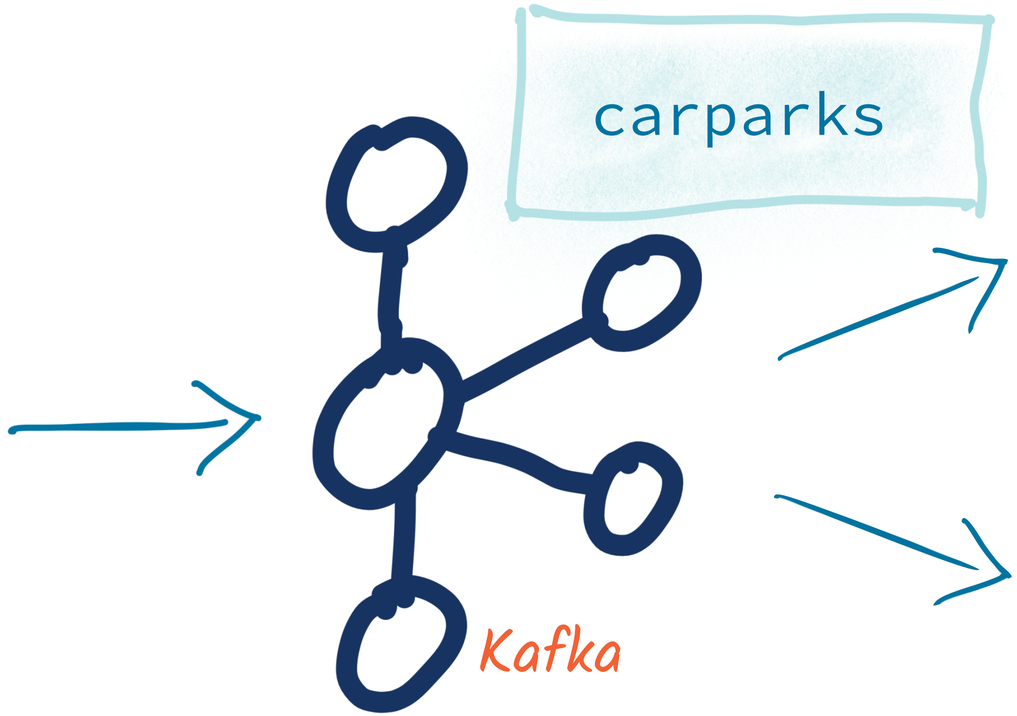

Kafka alerts
bot
/alert 700 23:44
Successfully created alert to be sent whenever more than 700 spaces are available 23:44
The Leisure Exchange carpark has 996 spaces available (capacity 975)
(data as of 24 Jul 20 00:36 BST) 23:44
The Broadway carpark has 1273 spaces available (capacity 1260)
(data as of 24 Jul 20 00:36 BST) 23:44
The Cow & Calf
Hanging Stone Rd.
23:44
Gonna go and find carpark that's nearby with spaces for -1.797821,53.915264...standby... 23:44
Westgate
23:44
The nearest carpark is Westgate, which is 13.5km away and has 116 spaces free. 23:44

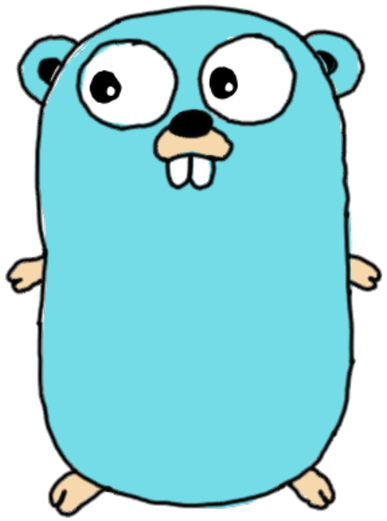
Don't just tell me...

show me!

Demo code: https://rmoff.dev/carparks

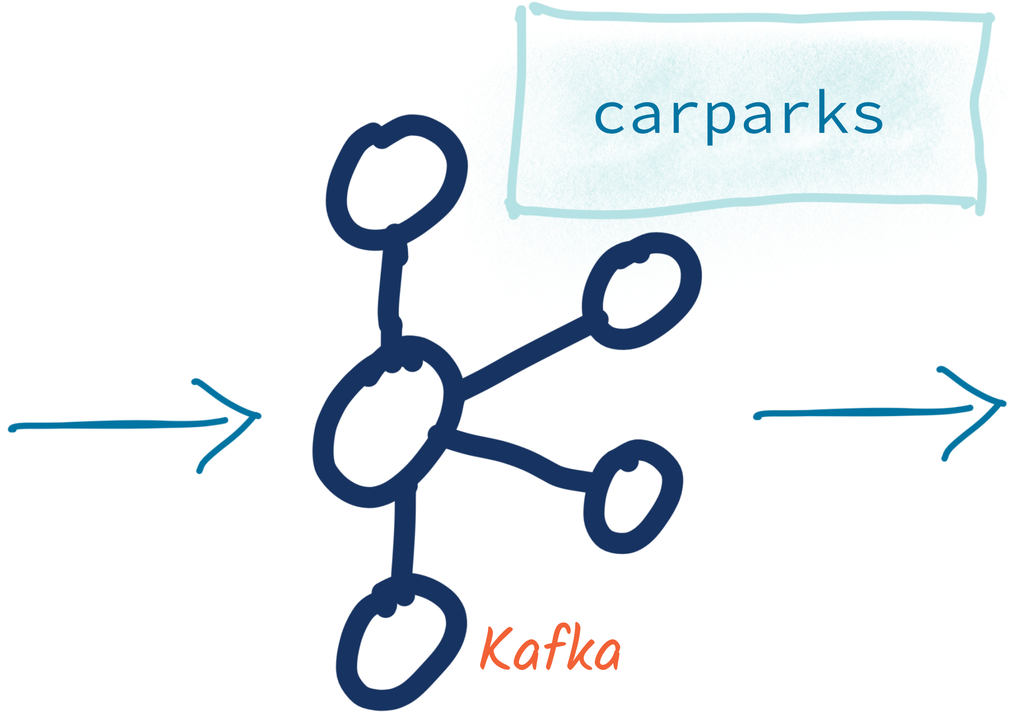

NORTHERN
DATA
HUB

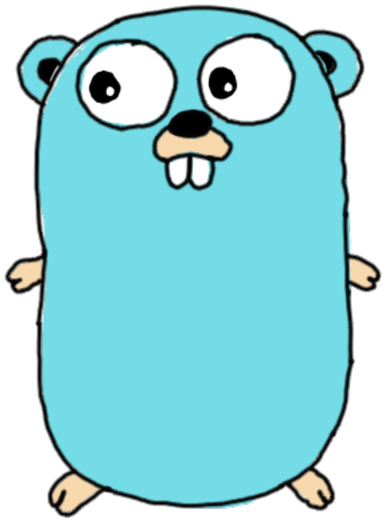

carparks
Kafka



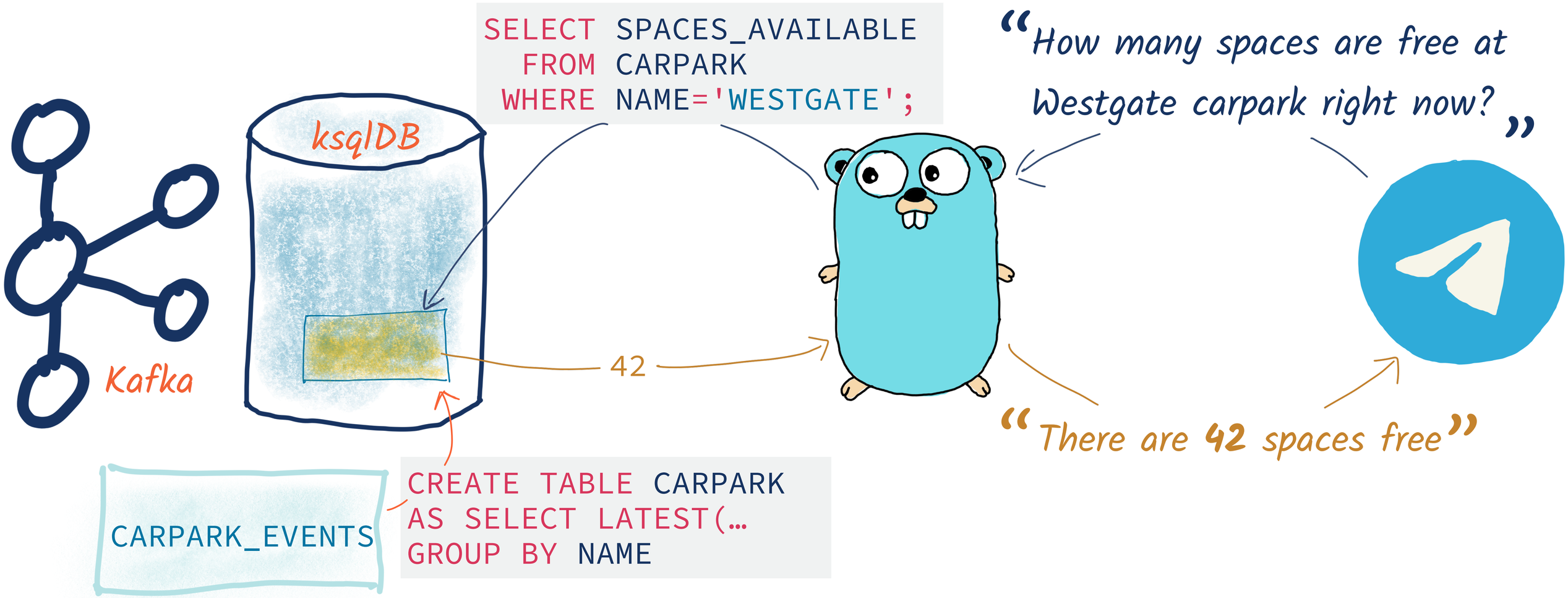
What are the key pieces of the design?

Event Driven Alerts

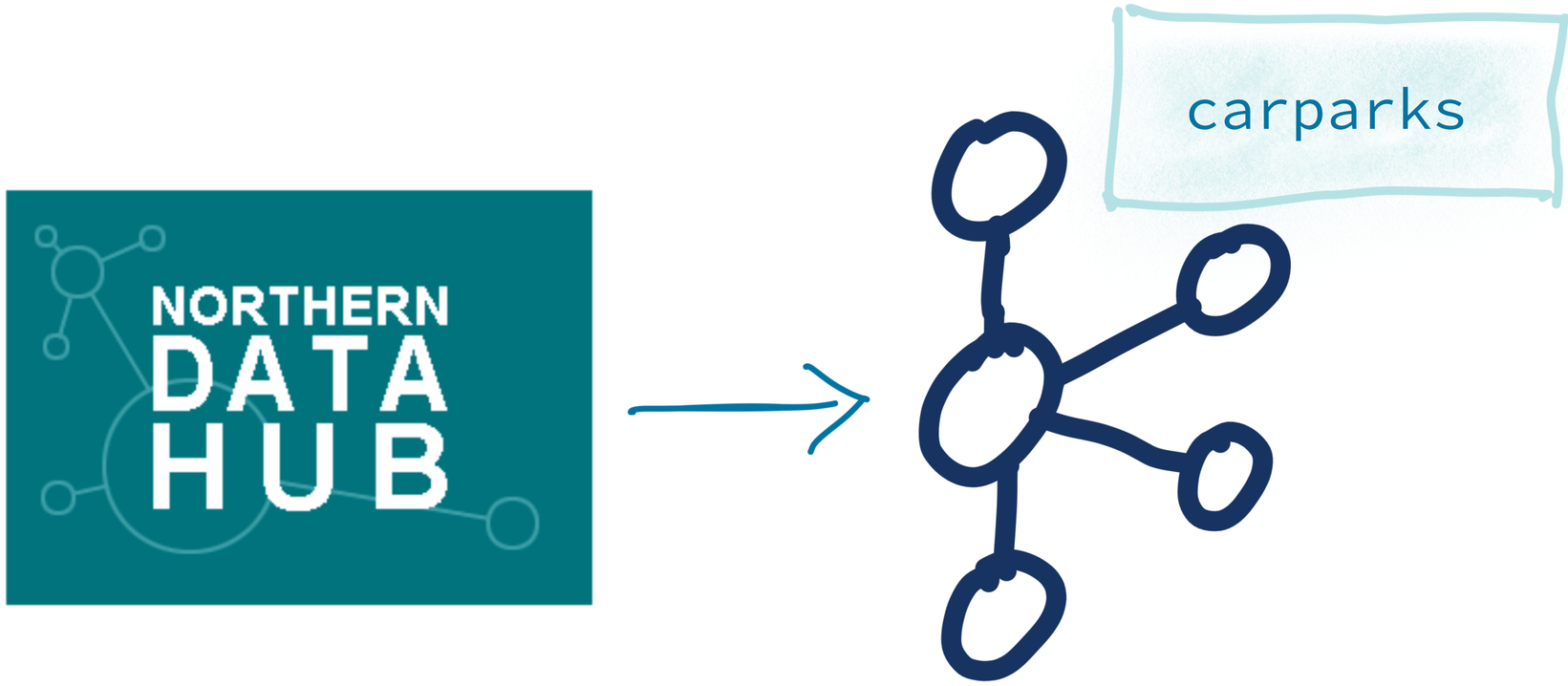




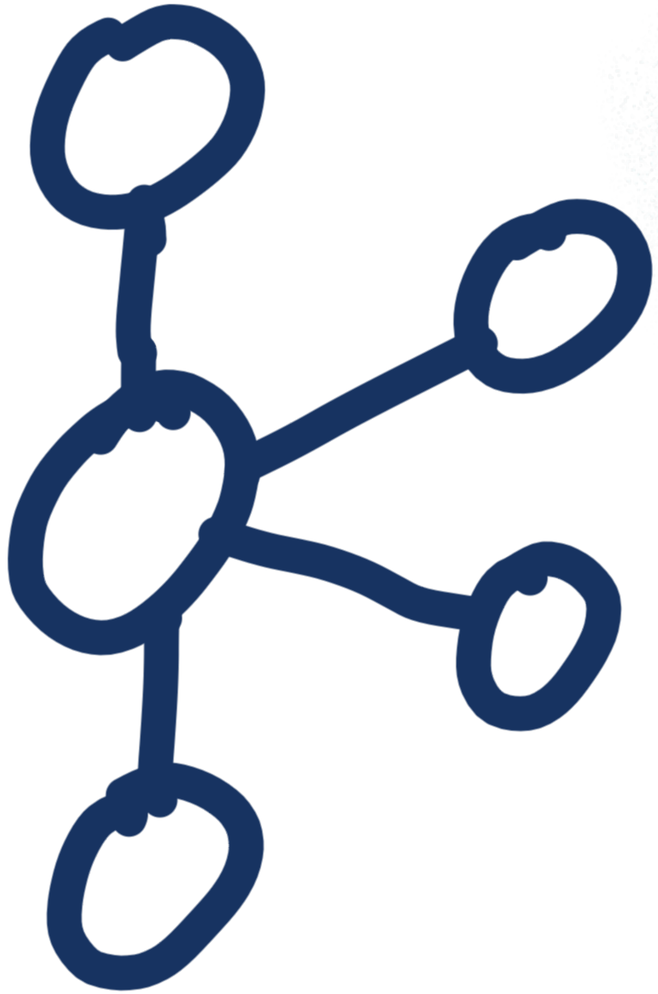
KN Lookups (materialised views)



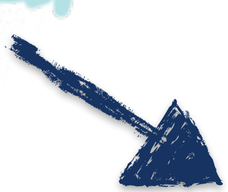
A schema...

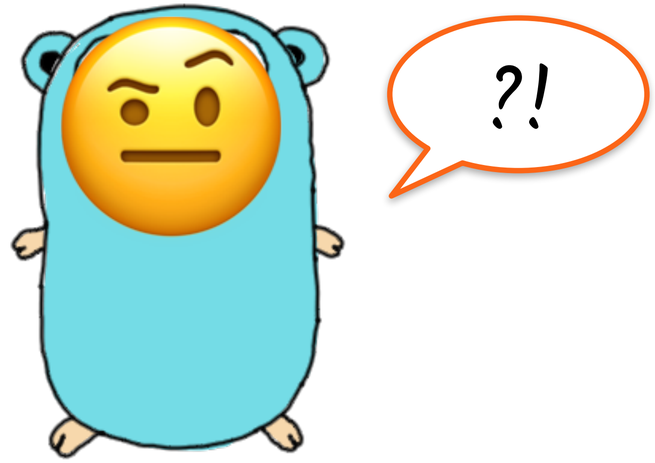


A schema...

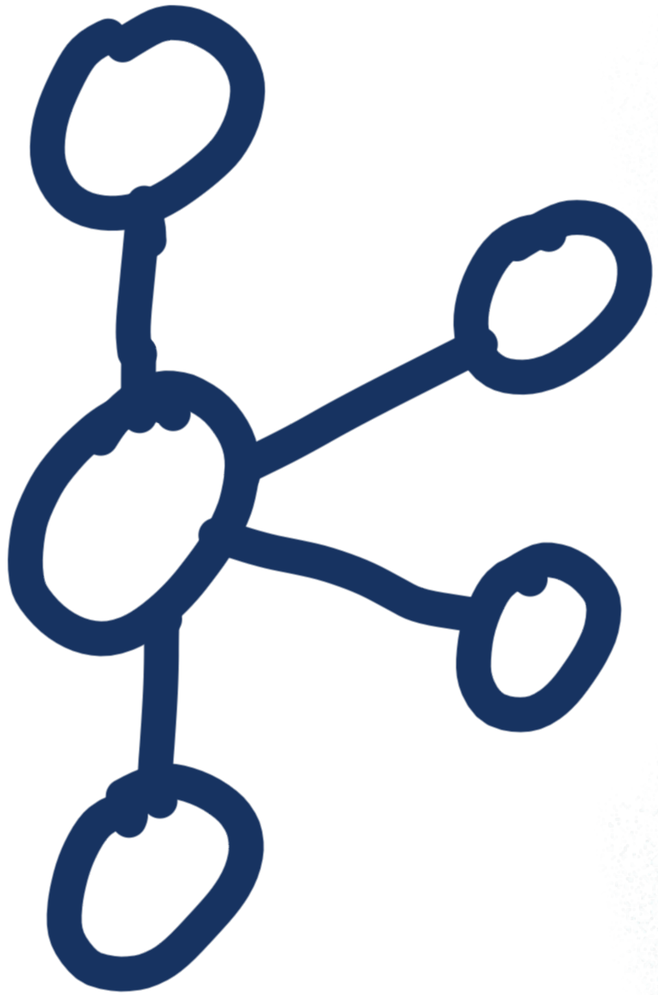


```
2020-10-14,12:28,Broadway,1132,921
2020-10-14,12:28,Kirkgate Centre,611,474
2020-10-14,12:28,Sharpe Street,98,63
```

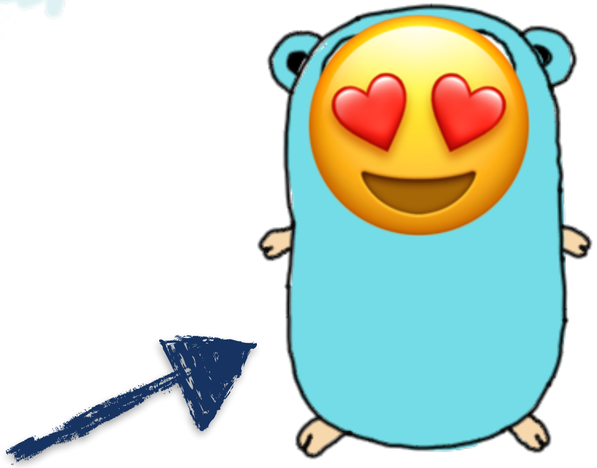




My kingdom for a schema!

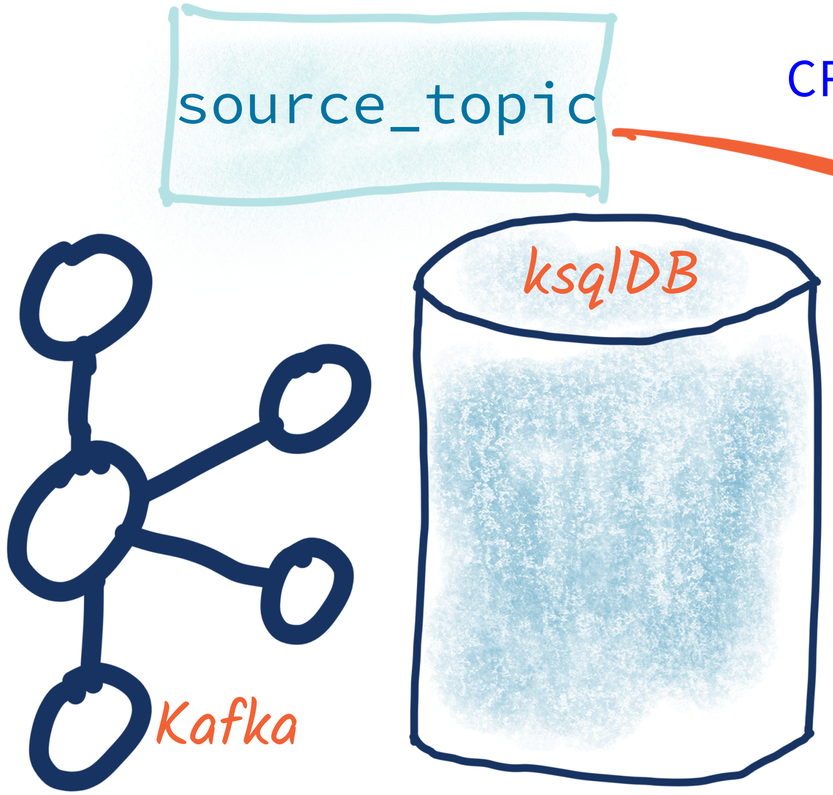


```
2020-10-14,12:28,Broadway,1132,921
2020-10-14,12:28,Kirkgate Centre,611,474
2020-10-14,12:28,Sharpe Street,98,63
```



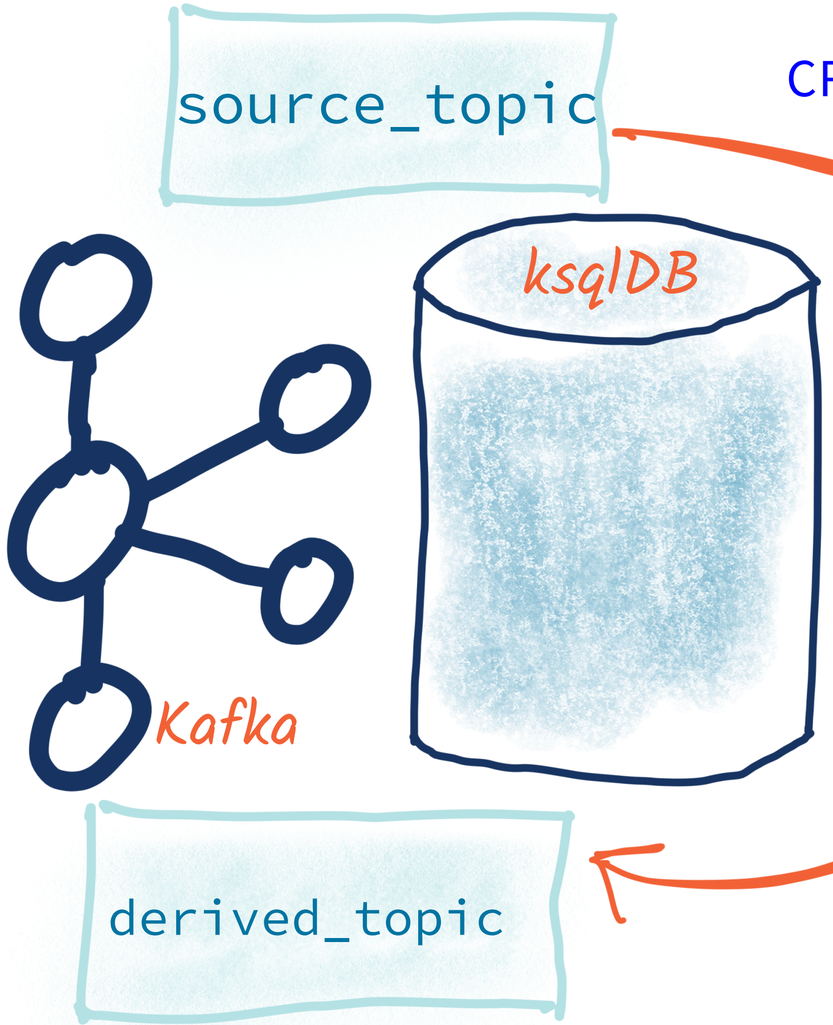
```
{
    "ts": "2020-10-14T12:28 UTC+1",
    "name": "Broadway",
    "capacity": 1132,
    "empty": 921
}
…
```

Applying a schema to streams of data




```
CREATE STREAM mySource (date        VARCHAR ,
                        time        VARCHAR ,
                        name        VARCHAR ,
                        capacity    INT )
  WITH (KAFKA_TOPIC='source_topic',
        VALUE_FORMAT='DELIMITED');
```

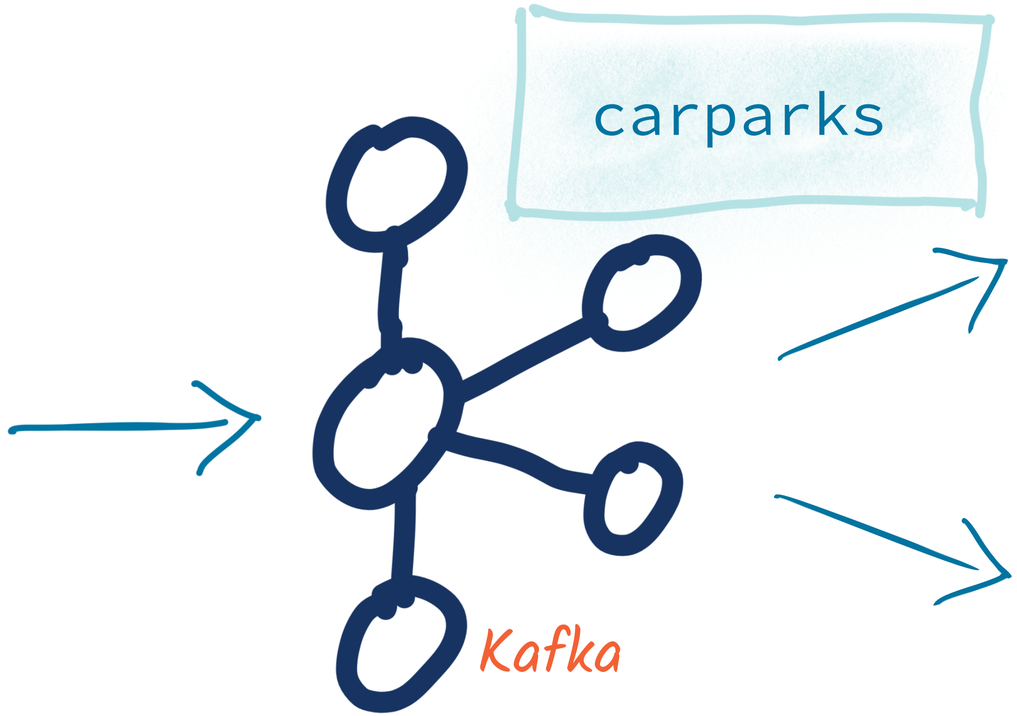
Applying a schema to streams of data

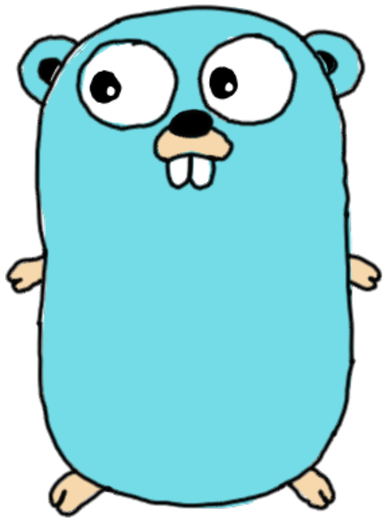



```
CREATE STREAM mySource (date            VARCHAR ,
                        time            VARCHAR ,
                        name            VARCHAR ,
                        capacity        INT )
    WITH (KAFKA_TOPIC='source_topic',
          VALUE_FORMAT='DELIMITED');
```

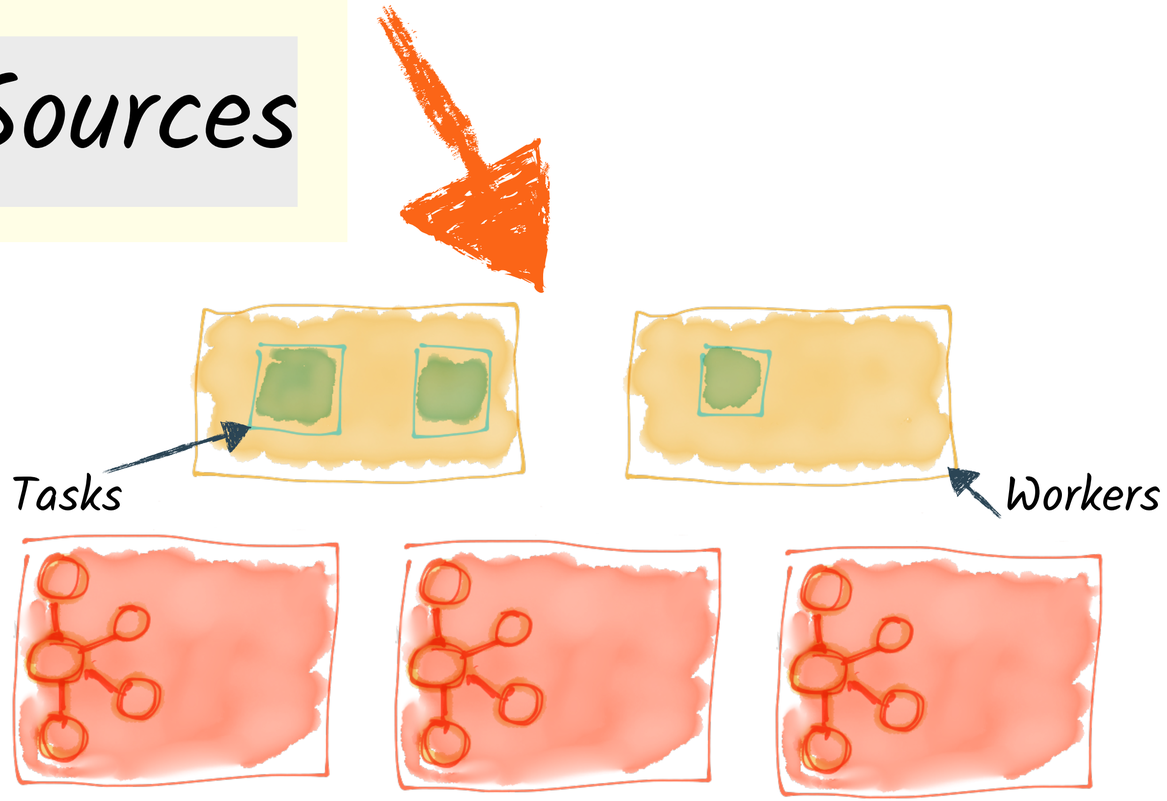
```
CREATE STREAM myTargetStream
WITH (VALUE_FORMAT='PROTOBUF',
      KAFKA_TOPIC='derived_topic') AS
SELECT  *
   FROM mySource;
```

Integration





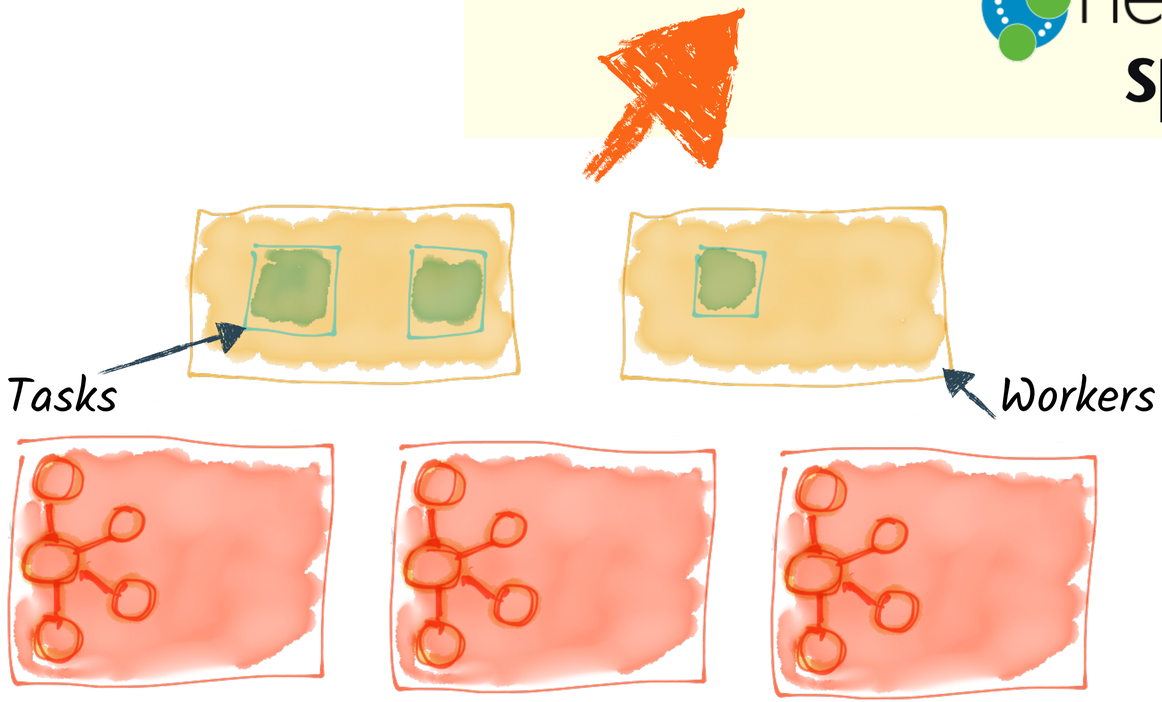
Streaming Integration with Kafka Connect



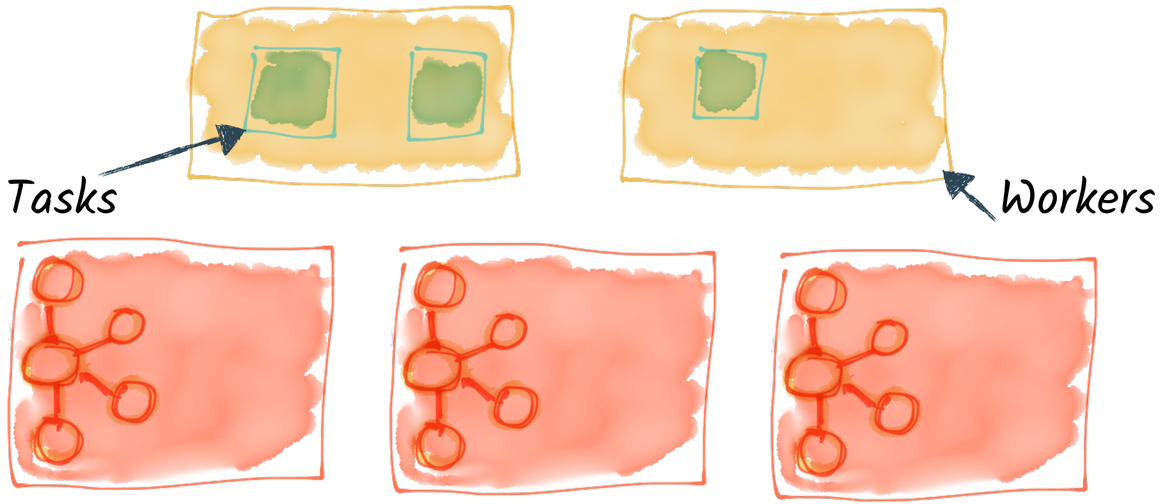
Kafka Connect

Kafka Brokers

Streaming Integration with Kafka Connect



Streaming Integration with Kafka Connect

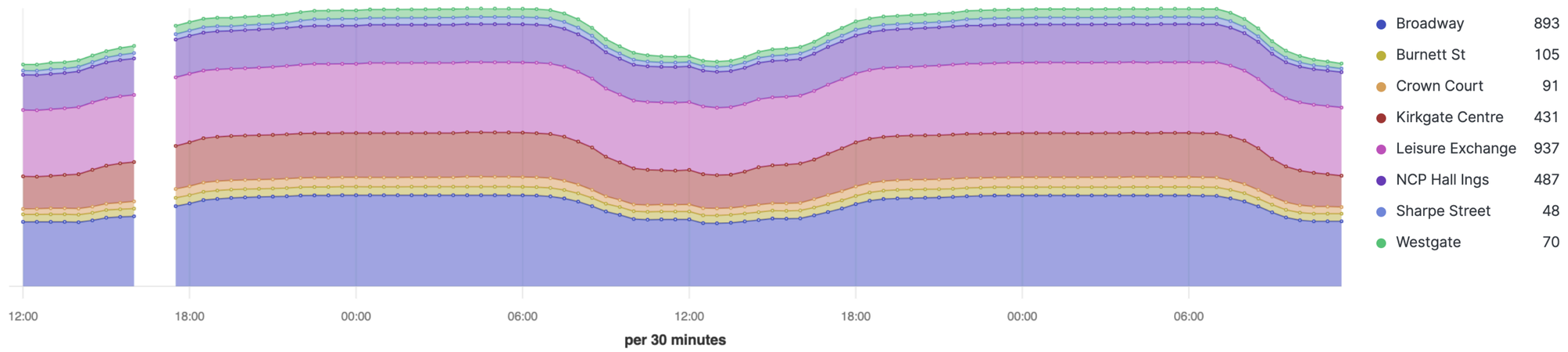


Kafka Connect

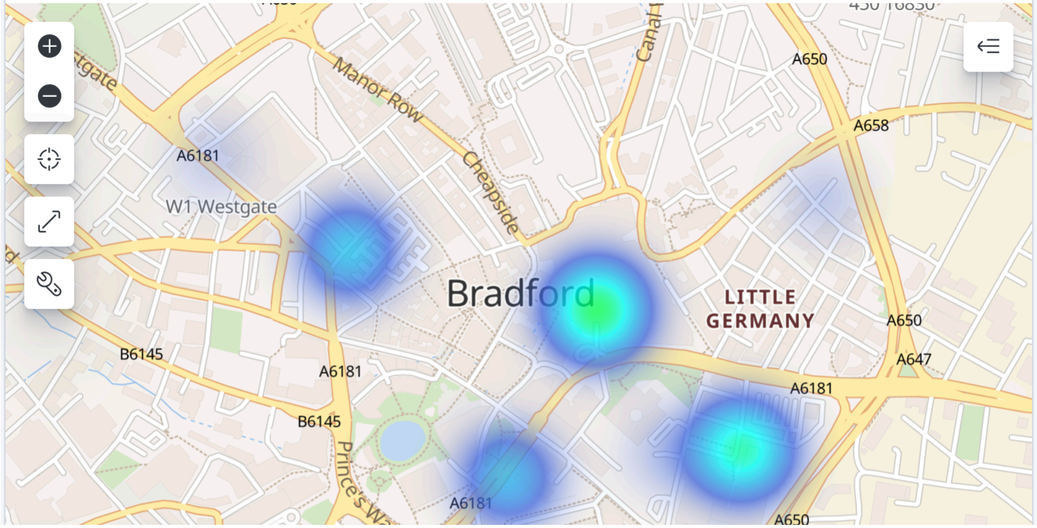
Kafka Brokers

Streaming Analytics

Median empty spaces over time



Carpark map



Current empty spaces

Min of EMPTY_PLACES **893** Broadway	Min of EMPTY_PLACES **106** Burnett St	Min of EMPTY_PLACES **92** Crown Court
Min of EMPTY_PLACES **439** Kirkgate Centre	Min of EMPTY_PLACES **939** Leisure Exchange	Min of EMPTY_PLACES **488** NCP Hall Ings

Why build it this way?

Events

Streams of Events

We want to react to them as they happen

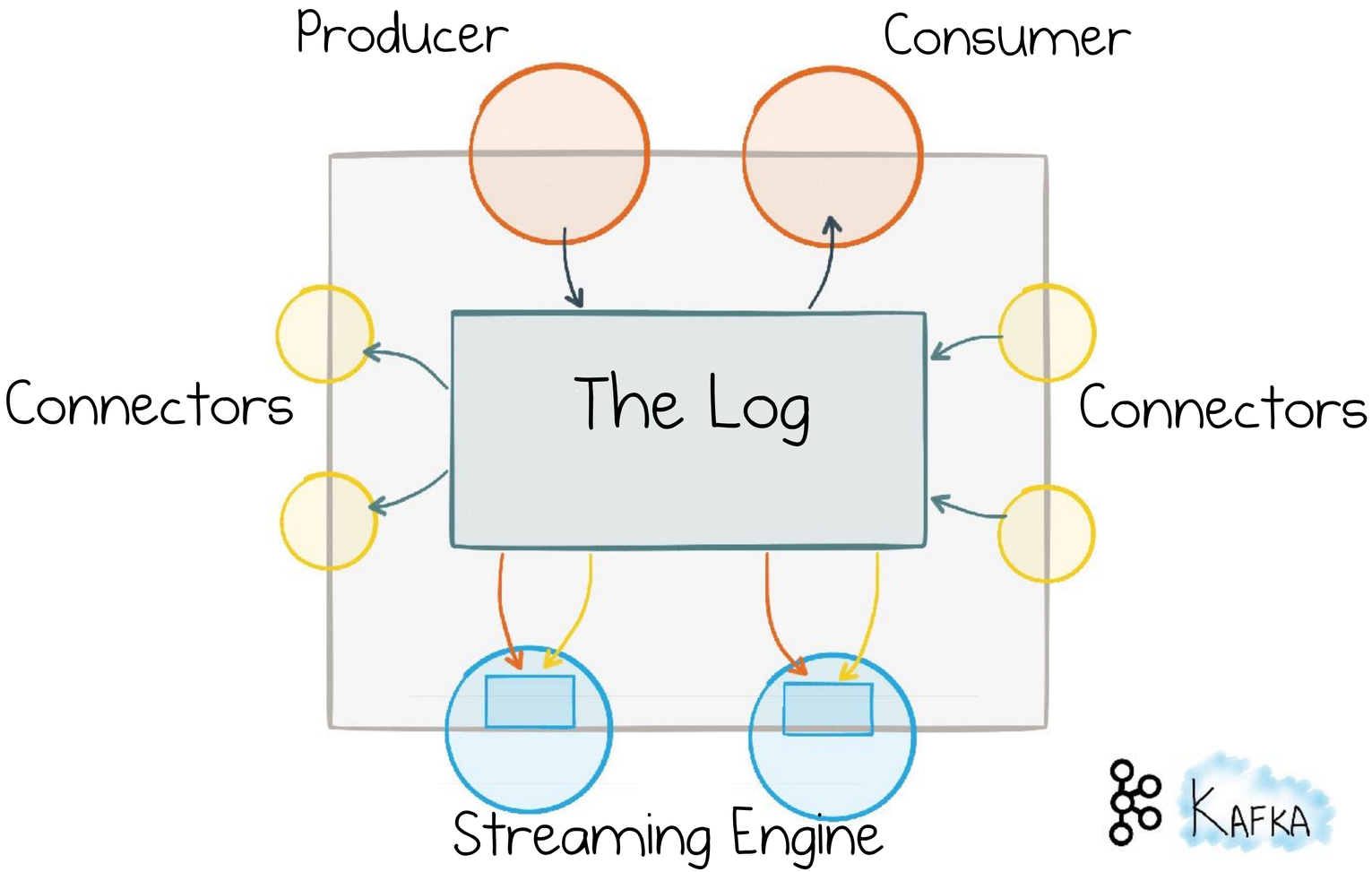
We want to build

state from a stream

of events

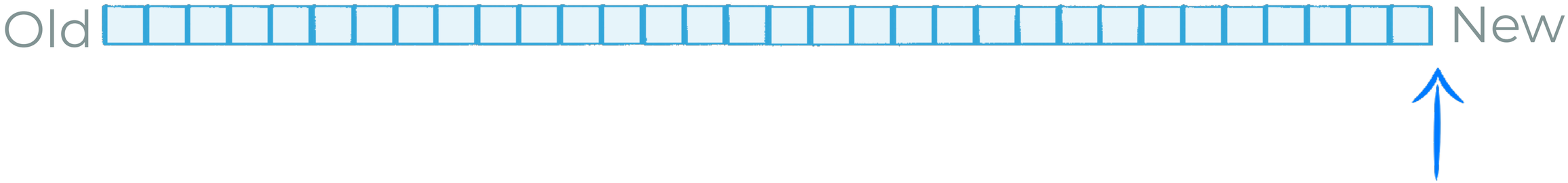
We want to provide the latest data in our analytics

Apache Kafka - an Event Streaming Platform



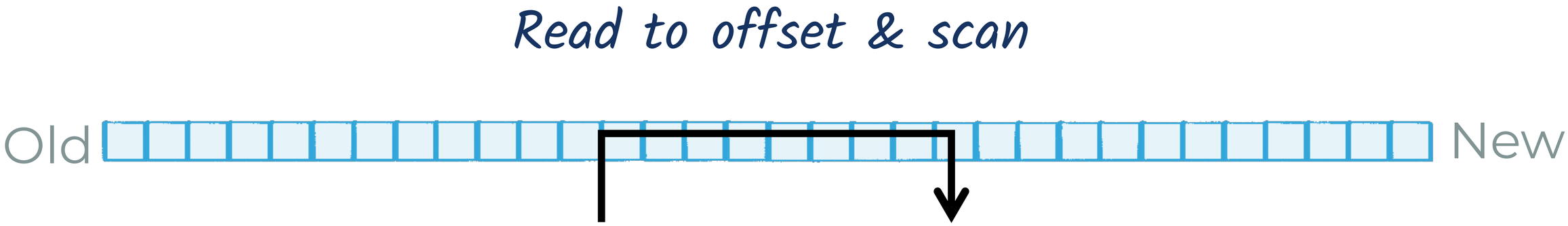
Why Kafka?

Distributed, Immutable, Event Log

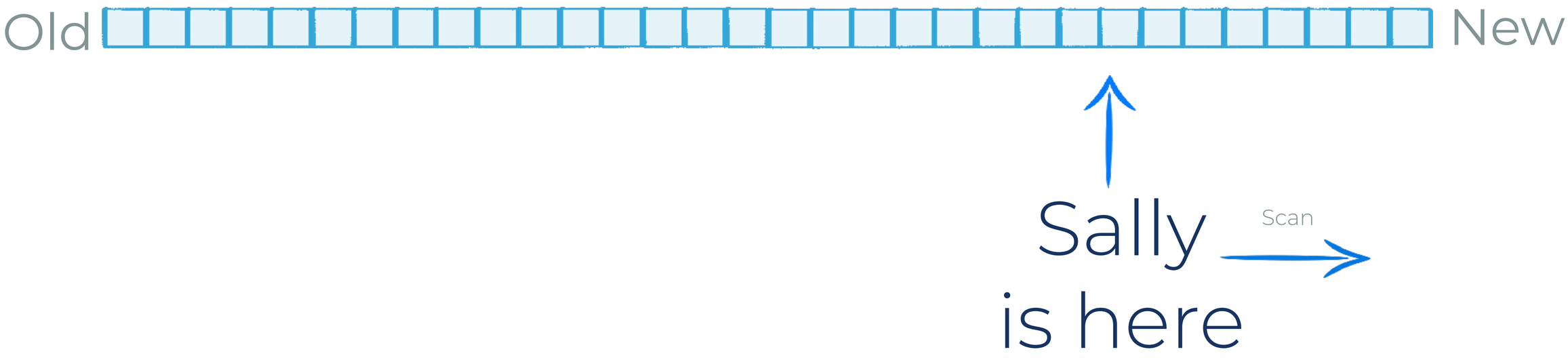


Events are added at the end of the log

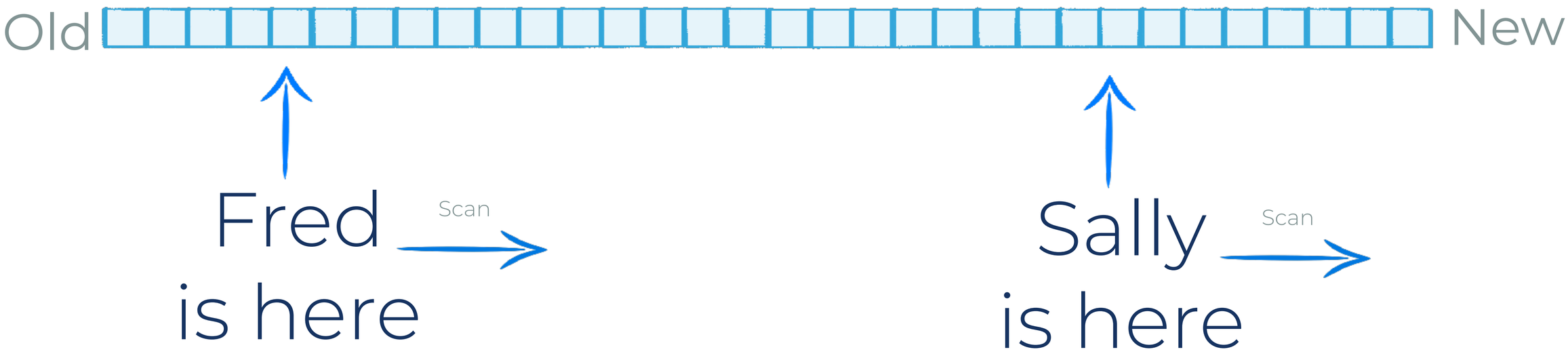
Consumers can seek to any point



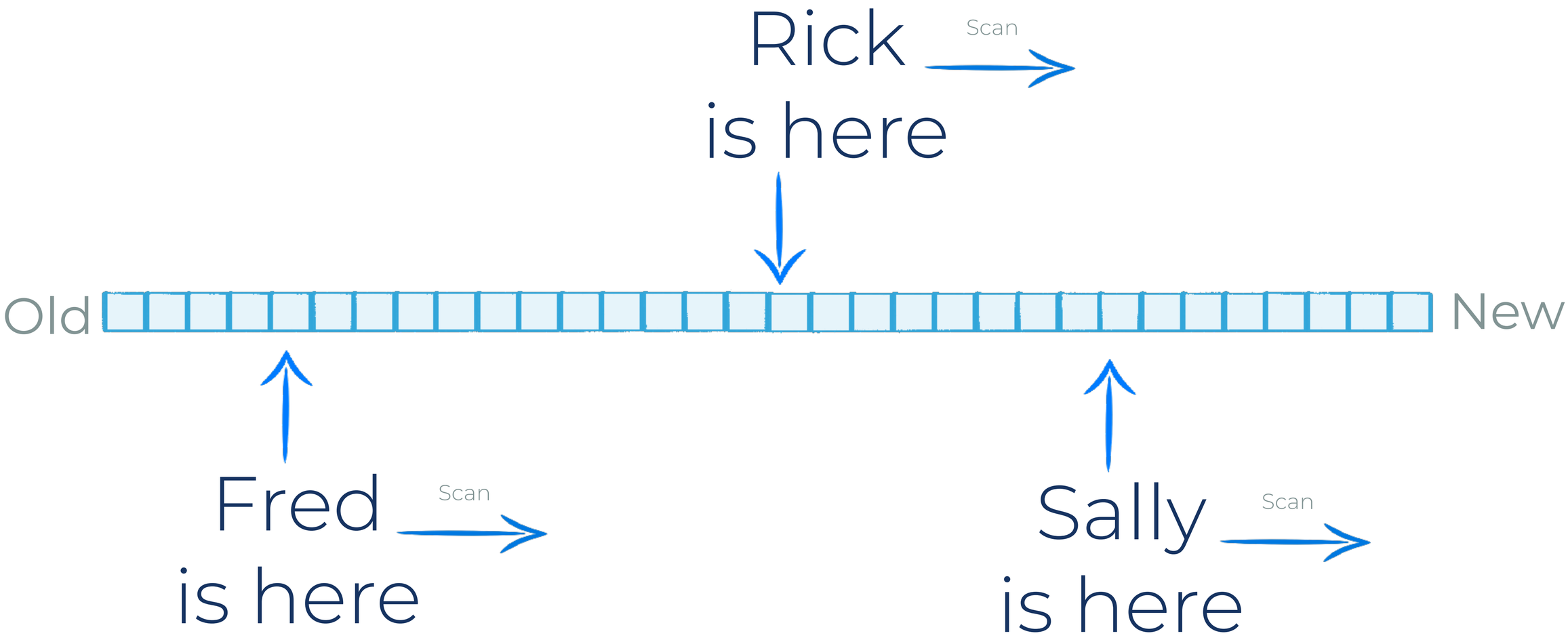
Data is not deleted once read



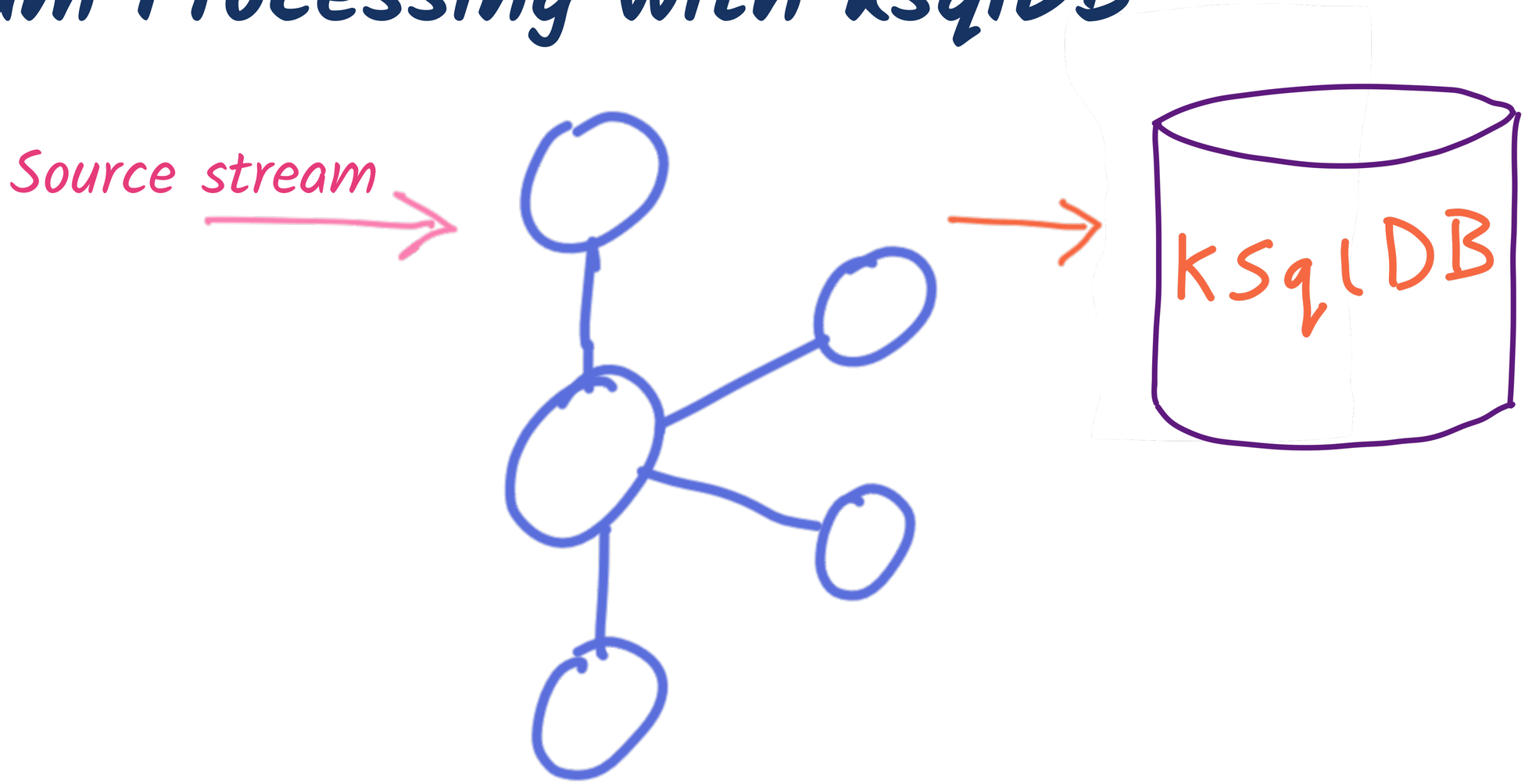
Consumers are independent of each other



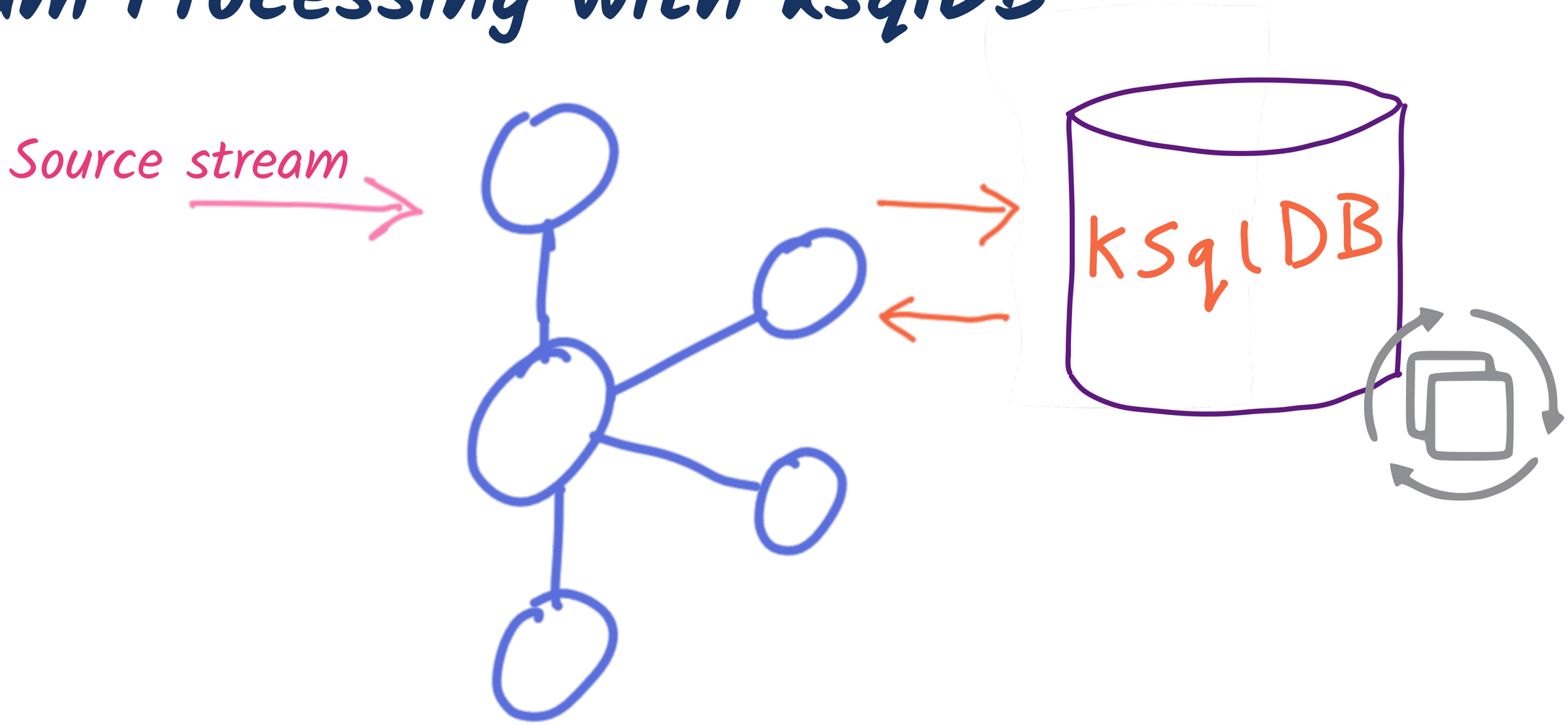
Consumers can be added later



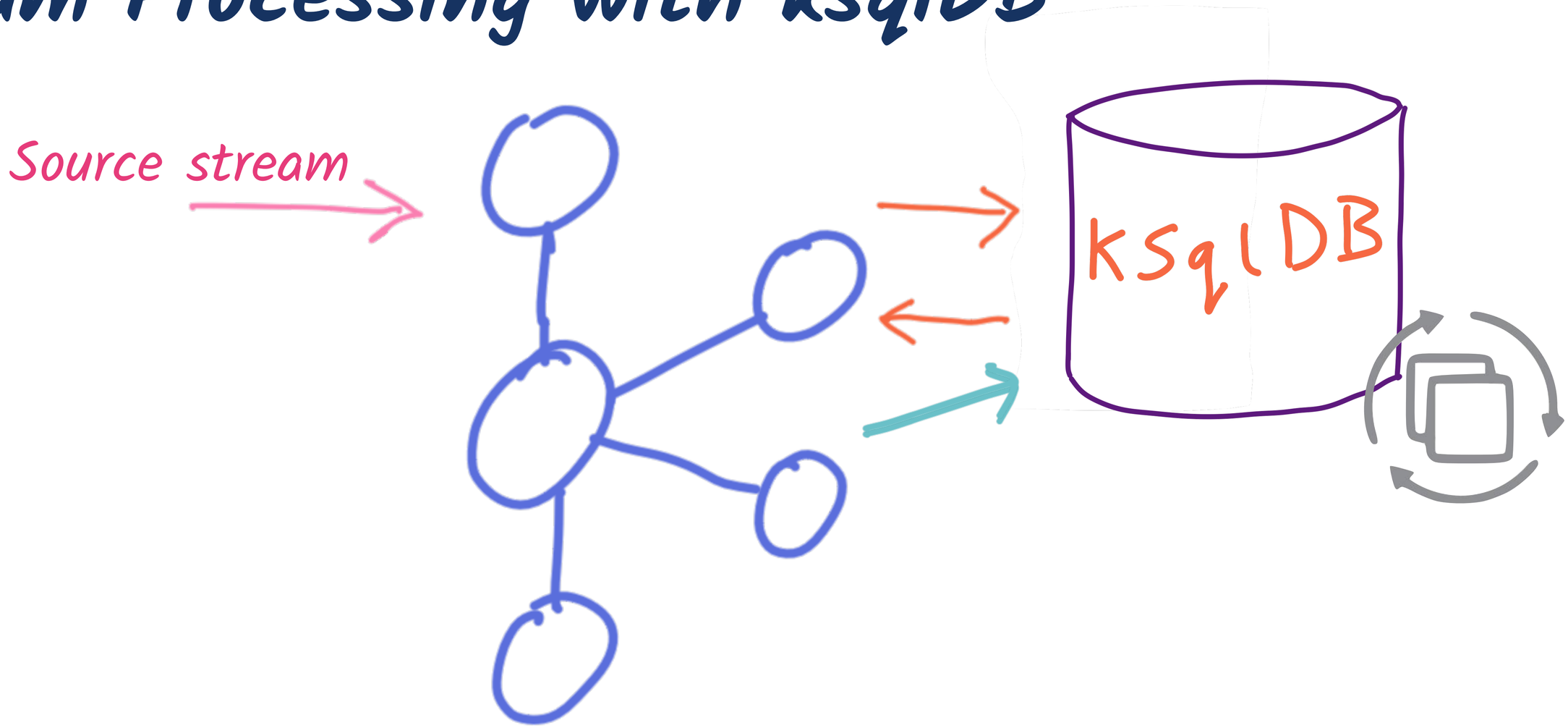
Stream Processing with ksqlDB



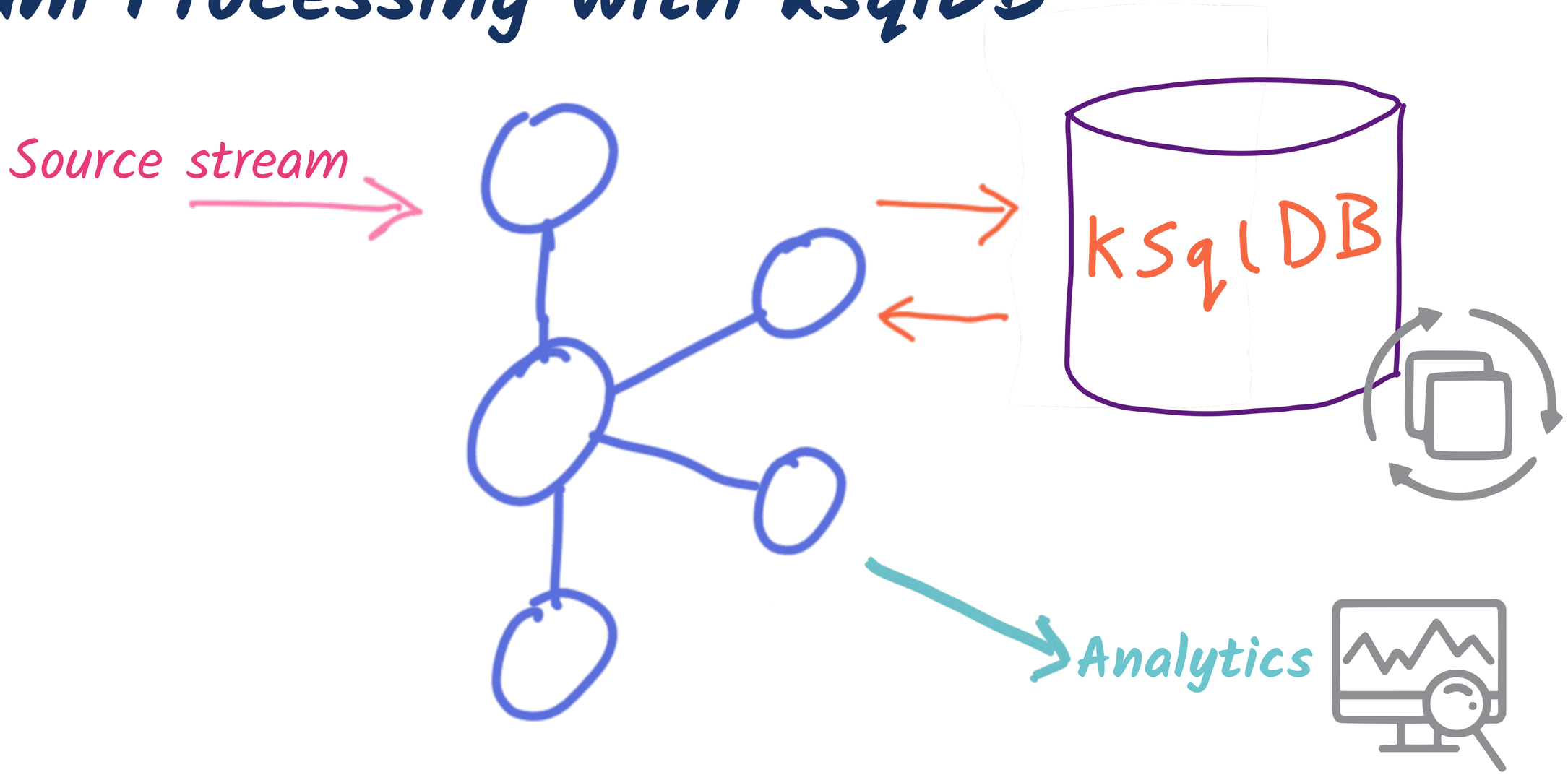
Stream Processing with ksqlDB



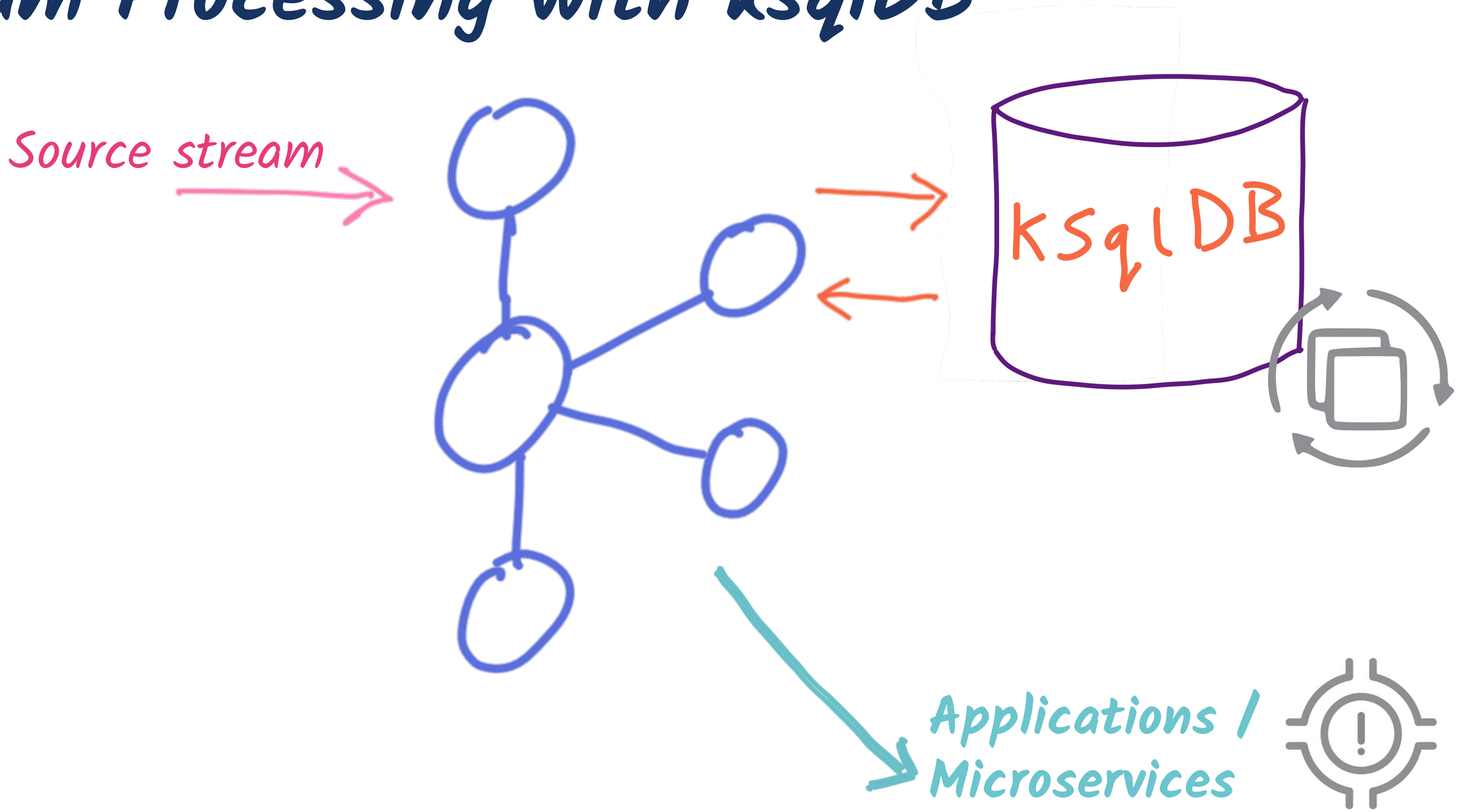
Stream Processing with ksqlDB



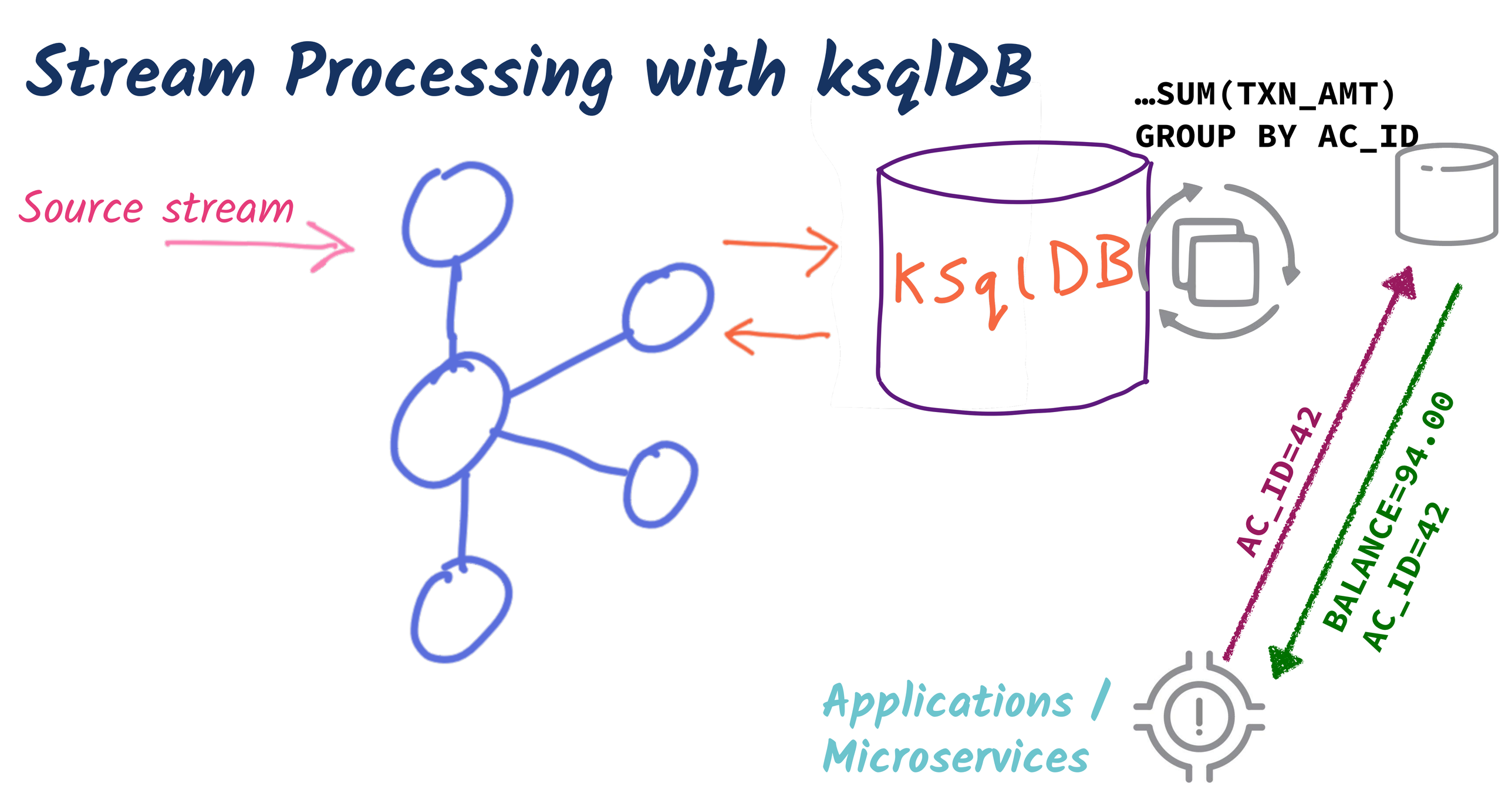
Stream Processing with ksqlDB



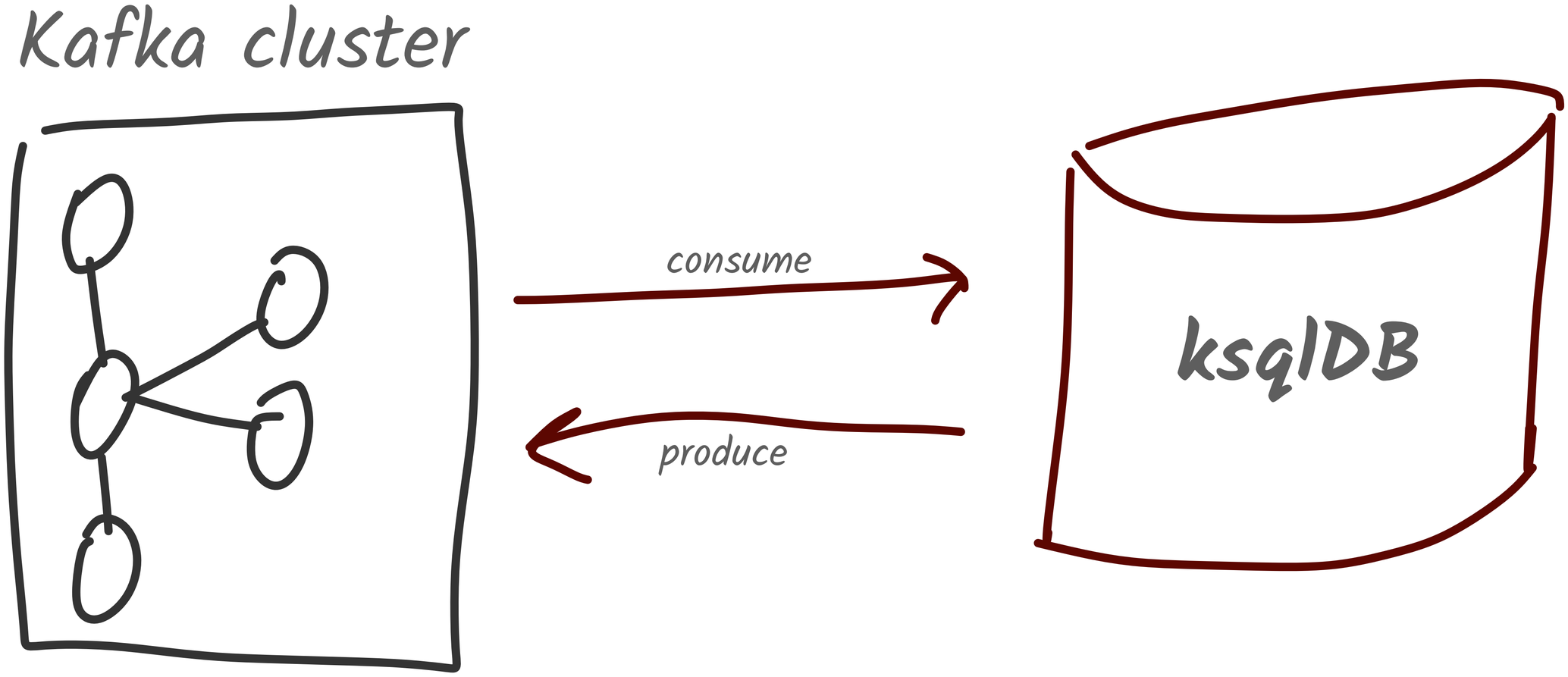
Stream Processing with ksqlDB

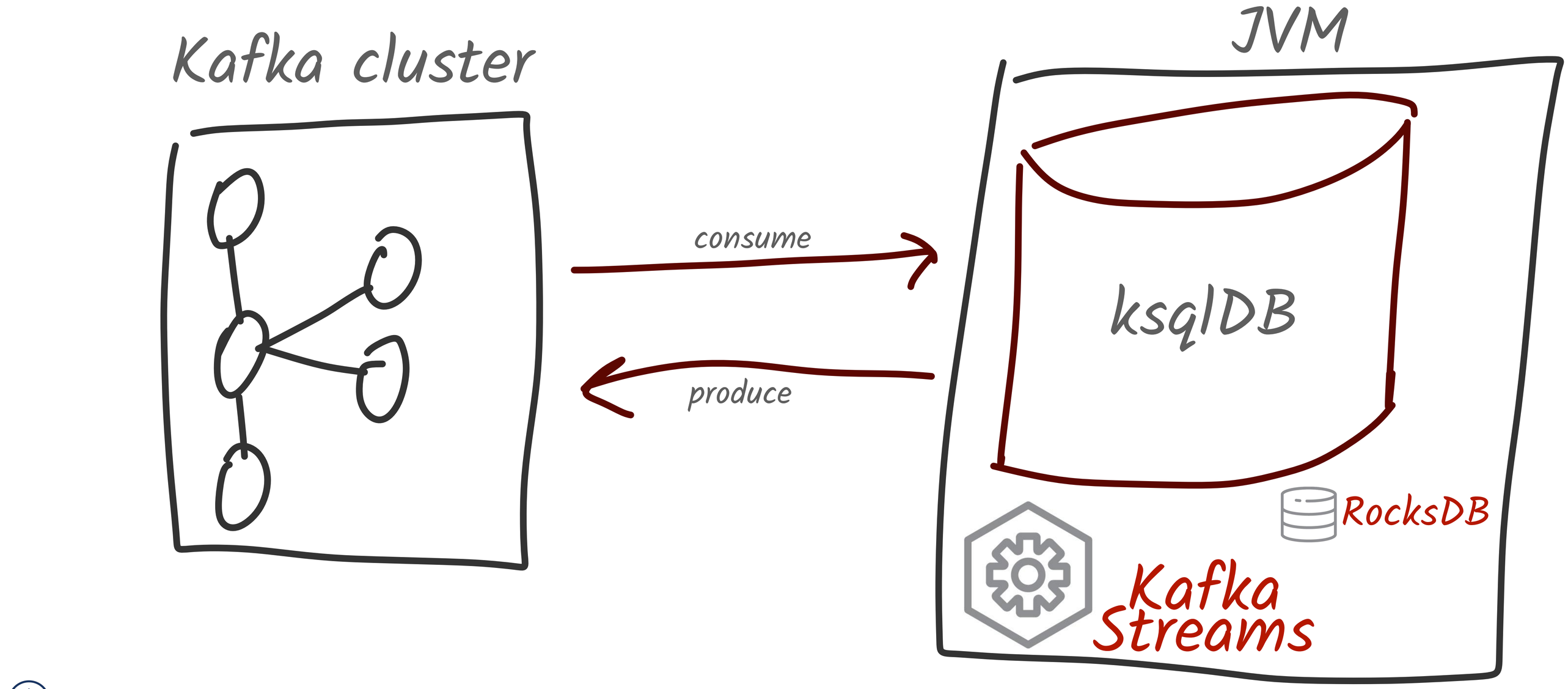


Stream Processing with ksqlDB



Under the covers of ksqlDB


Kafka cluster
consume
produce
ksqlDB


Kafka cluster
consume
produce
JVM
ksqlDB
RocksDB
Kafka Streams


WHERE CAN YOU RUN KSQLDB?
ANYWHERE!

confluent cloud

& ksqlDB

Fully Managed Kafka as a Service

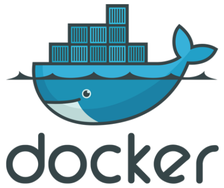
Running ksqlDB - self-managed

ksqlDB Server
(JVM process)

DEB, RPM, ZIP, TAR downloads
http://confluent.io/download

Docker images
confluentinc/ksqldb-server


Physical




…and many more…

Why Kafka?

Stream Store

Process Integrate

Stream

Store

Process

Integrate

Stream

Store

Process

Integrate

Stream

Store

Process Integrate

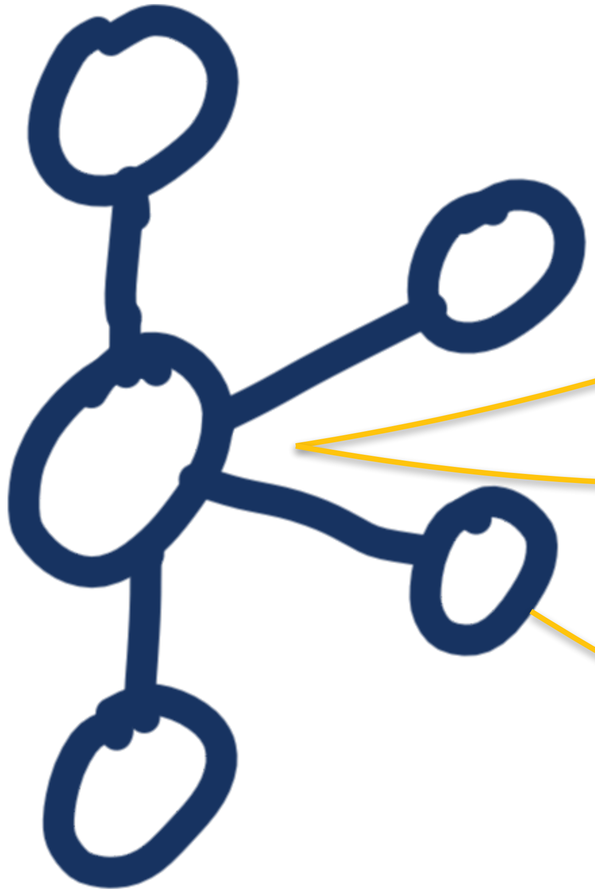
Stream

Store

Process

Integrate

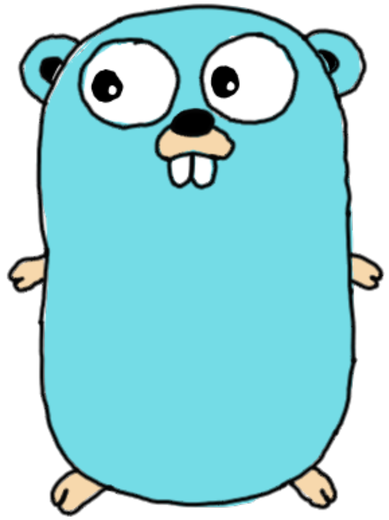
Flexible, event-driven applications



Event-driven alerts

Key/Value lookups

Streaming ETL



Want to learn more?

CONFLUENT

CTAs, not CATs
(sorry, not sorry)

Try it out for yourself

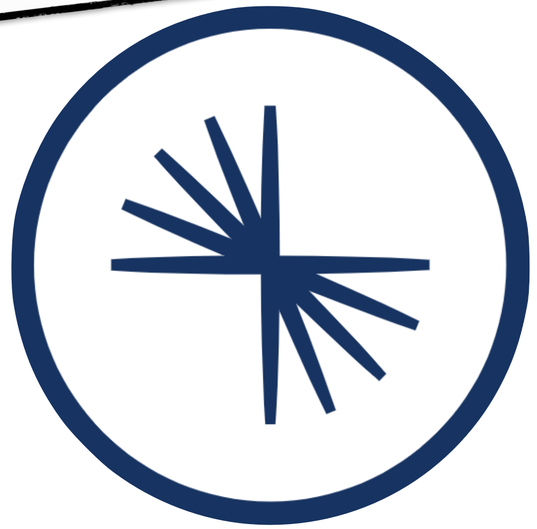
https://rmoff.dev/carparks

Deep-dive articles

- What is Apache Kafka?
- Event Streaming vs. Related Trends
- KRaft: Kafka without ZooKeeper
- Transactions & Guarantees in Kafka
- Processing & Storage Fundamentals
- Kafka Performance
- Cloud-native Kafka
- Streaming Database Systems
- Testing Apache Kafka
- Explore Kafka's Internals

Over 10 hours of free courses

- Apache Kafka 101
- Kafka Connect 101
- Kafka Streams 101
- ksqlDB 101
- Inside ksqlDB
- Spring Framework and Kafka
- Building Data Pipelines with Kafka
- Event Sourcing with Kafka
- Data Mesh 101



CONFLUENT

Developer

Plus: Hands-on Quick Starts and Client Language Guides + Event Streaming Patterns + More

developer.confluent.io

#EOF

@rmoff rmoff.dev/talks youtube.com/rmoff

Further reading / watching

https://rmoff.dev/kafka-talks

Real-life examples

Here's a nice example using real data to solve a real problem - is my train late now? What are the routes most likely to be delayed?

🚂On Track with Apache Kafka: Building a Streaming Platform solution with Rail Data

Moving from 🚆to 🚗 let's take another real data feed and build some realtime location-based notifications 🚉

🤖Building a Telegram bot with Go, Apache Kafka, and ksqlDB

Integration and data pipelines

Integration between Kafka and other systems? Kafka Connect has you covered ⚡

⏱️Kafka Connect in 60 seconds

🦸From Zero to Hero with Kafka Connect