



who's Tweeting about
DataSciCon

DataSciCon

@gammussa

KSQL

is a

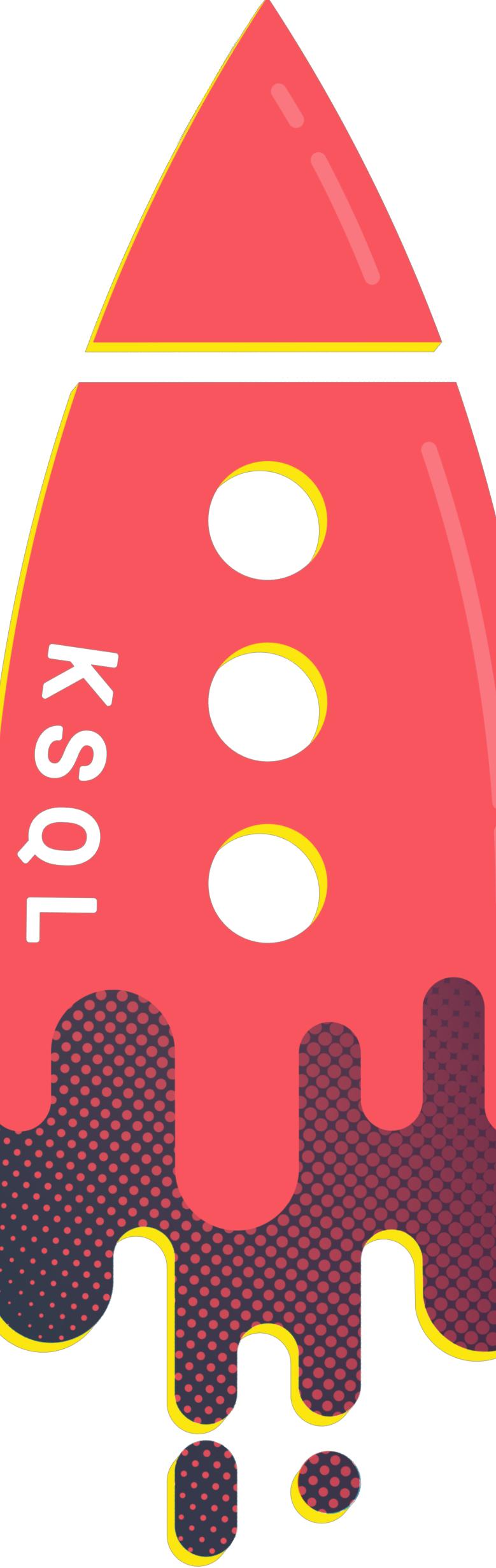
Declarative Stream Processing Language



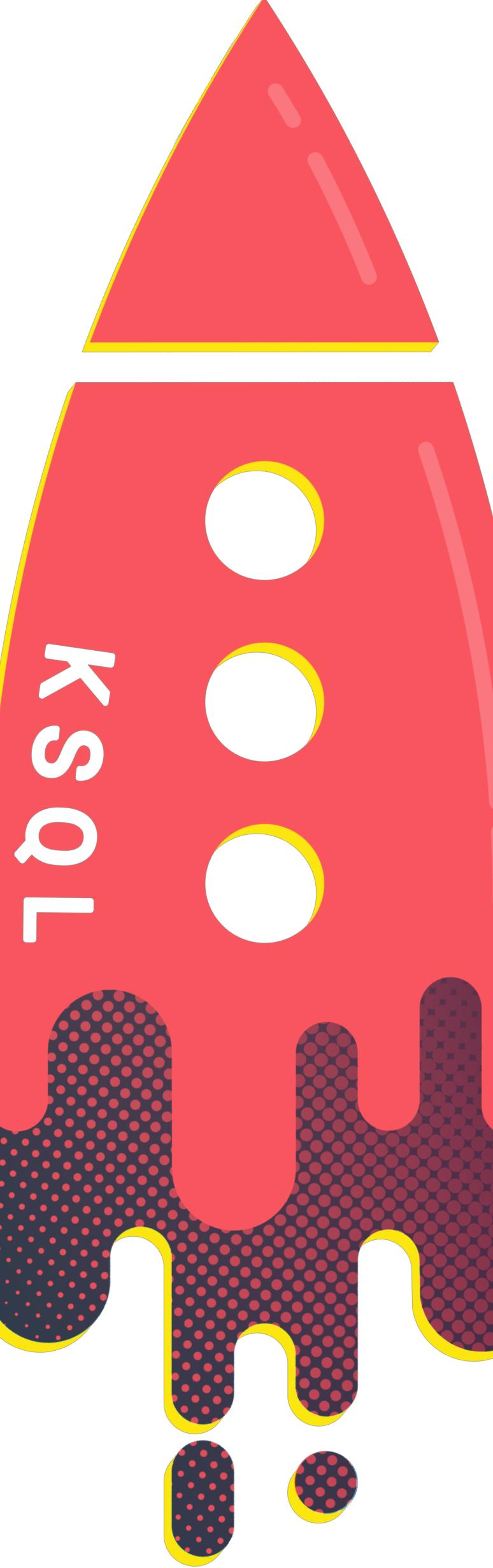
@gamussa

#DataSciCon

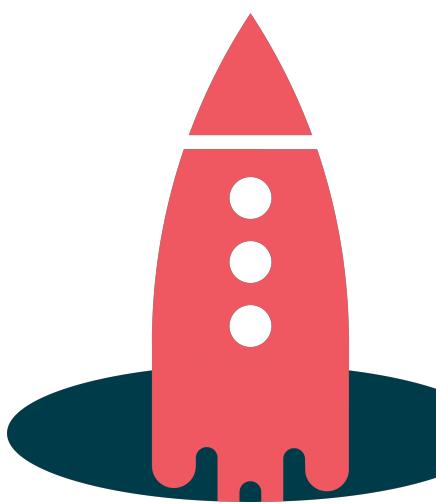
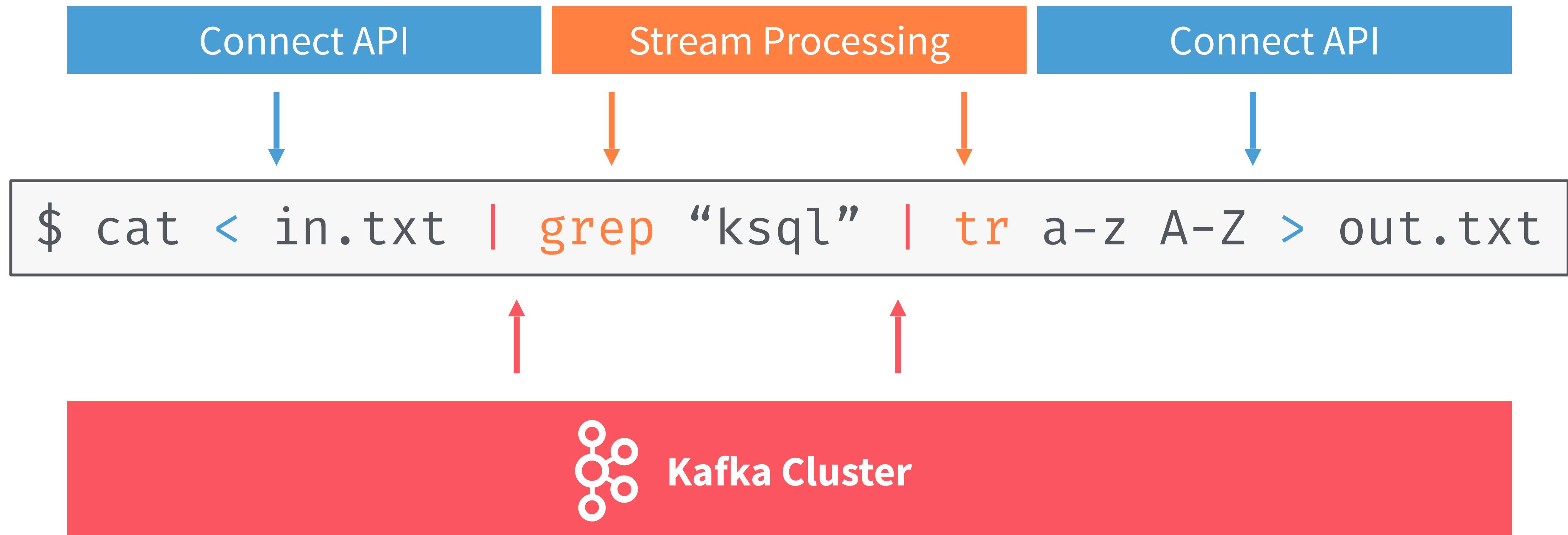
@confluentinc



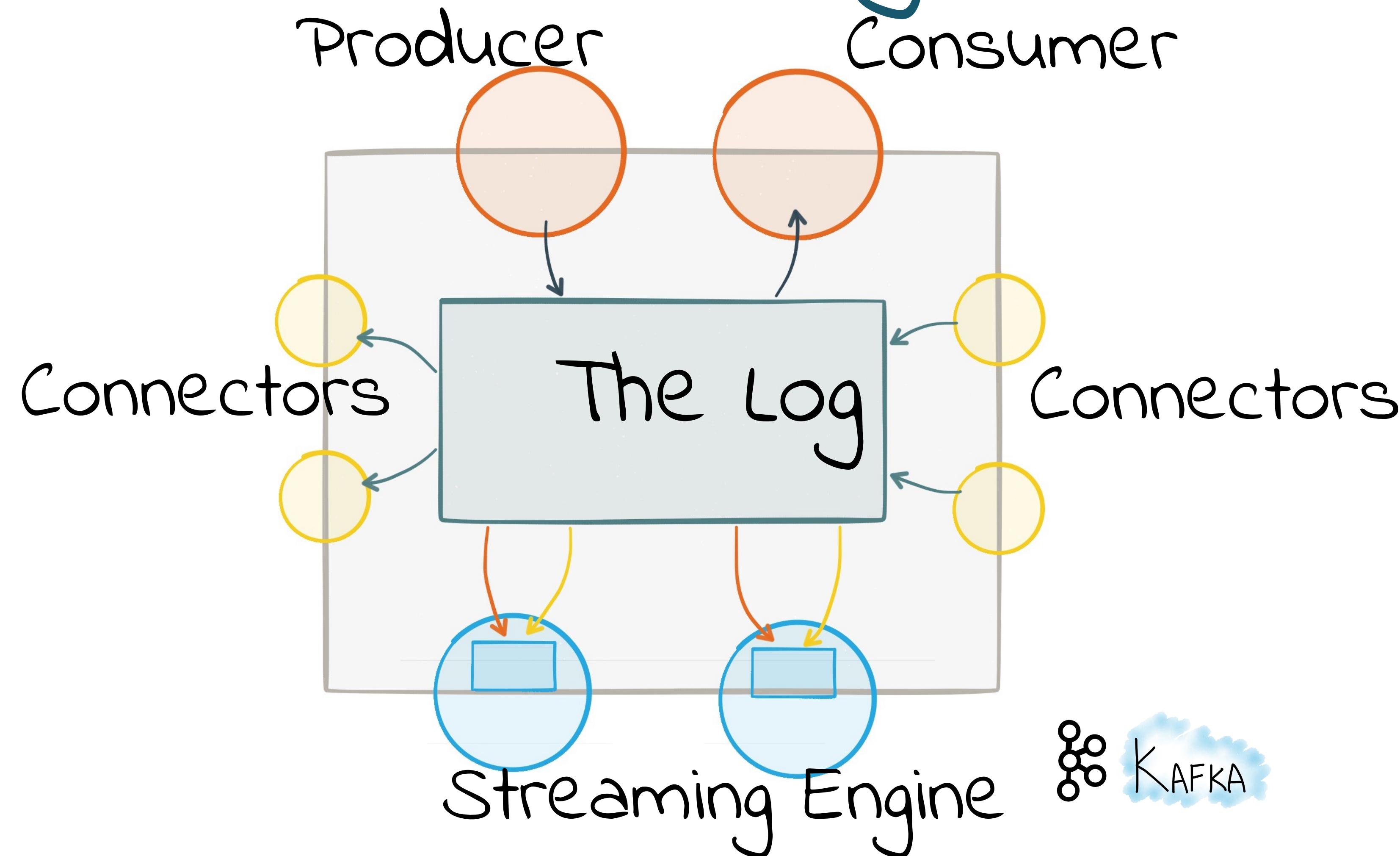
KSQL
is the
Streaming
SQL Engine
for
Apache Kafka



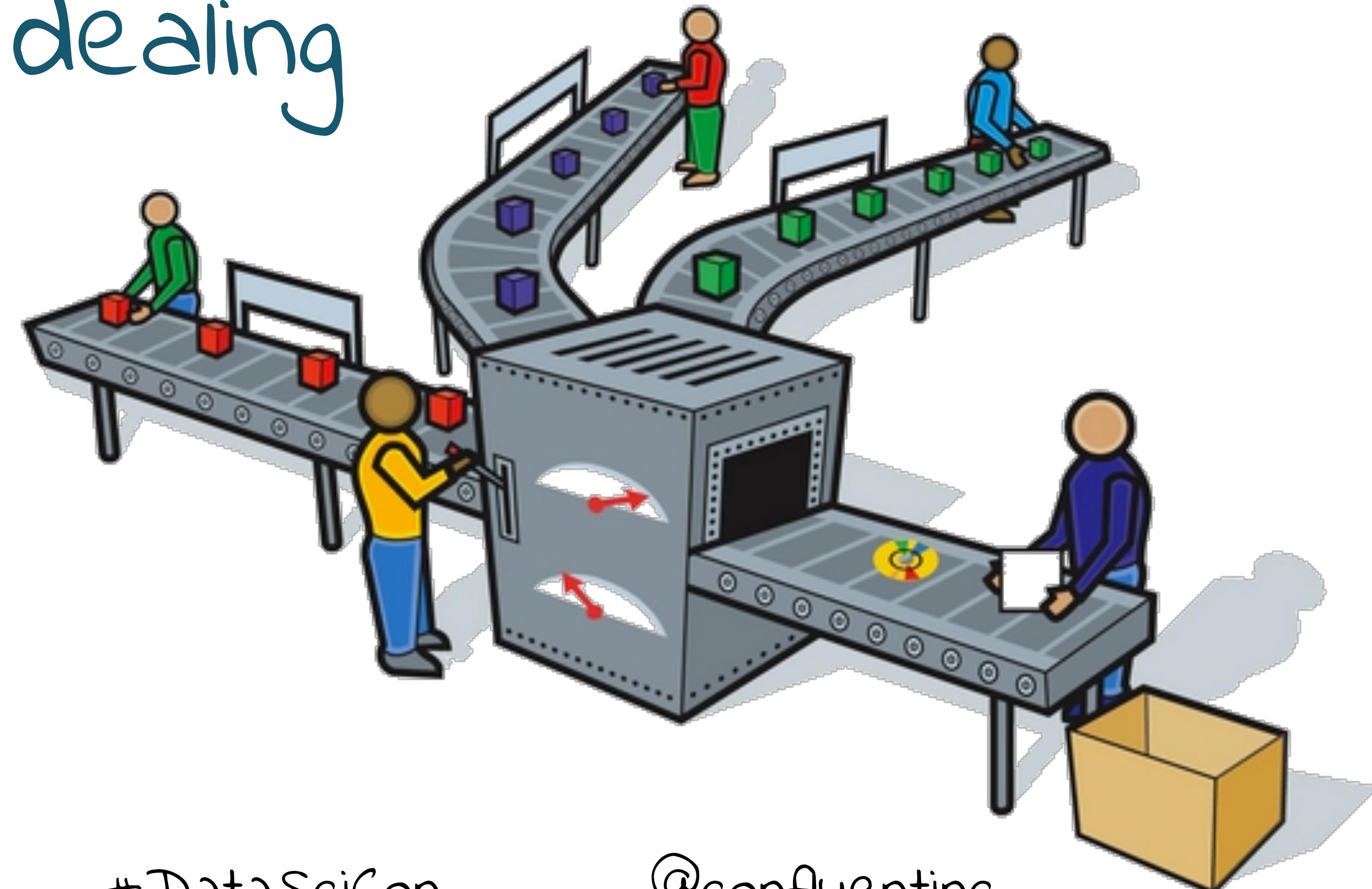
Stream Processing by Analogy



Kafka is a Streaming Platform

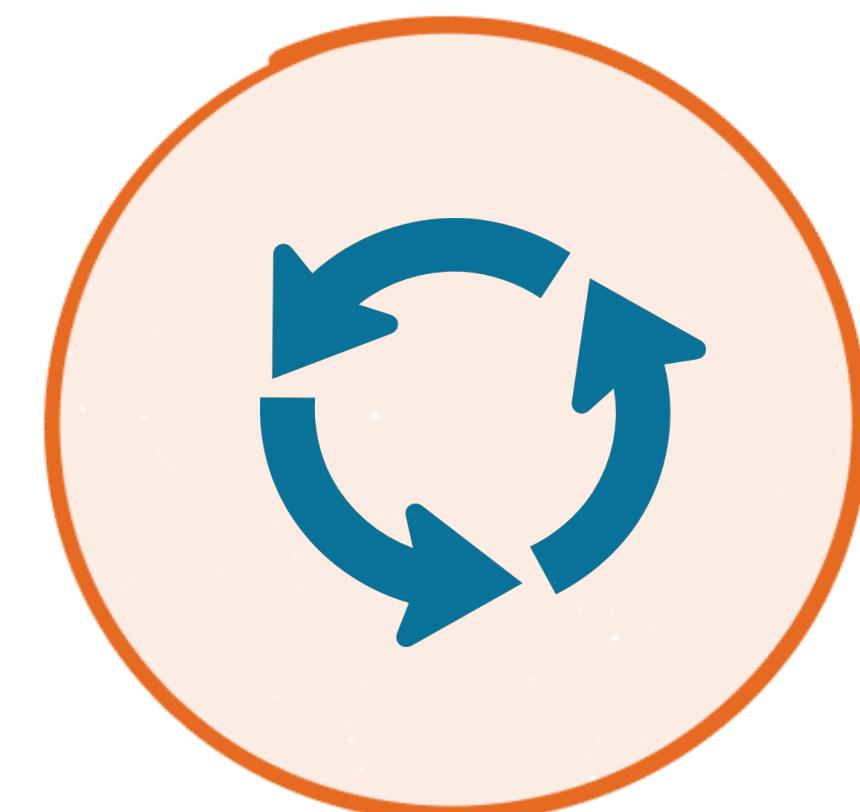
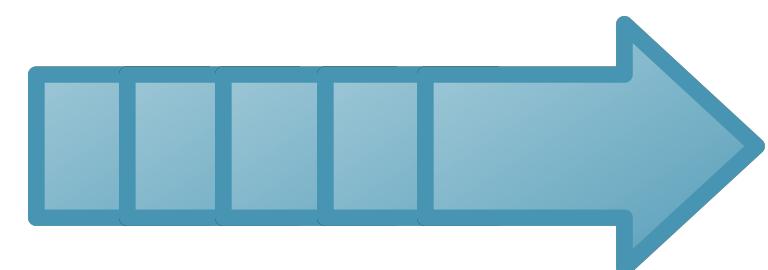


Streaming
is the toolset for dealing
with events
as they move!

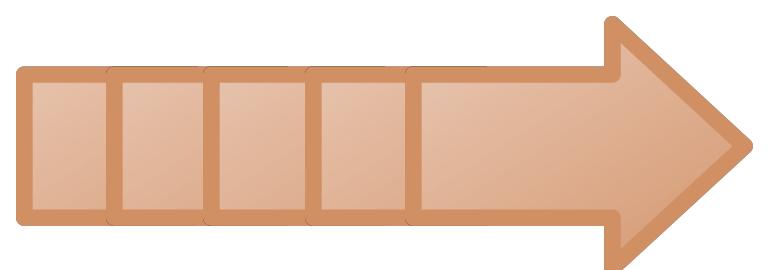


what exactly is Stream Processing?

authorization_attempts

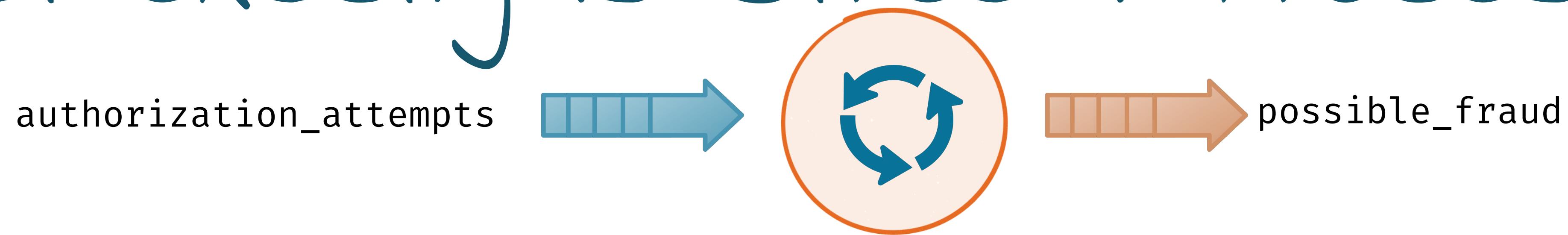


possible_fraud



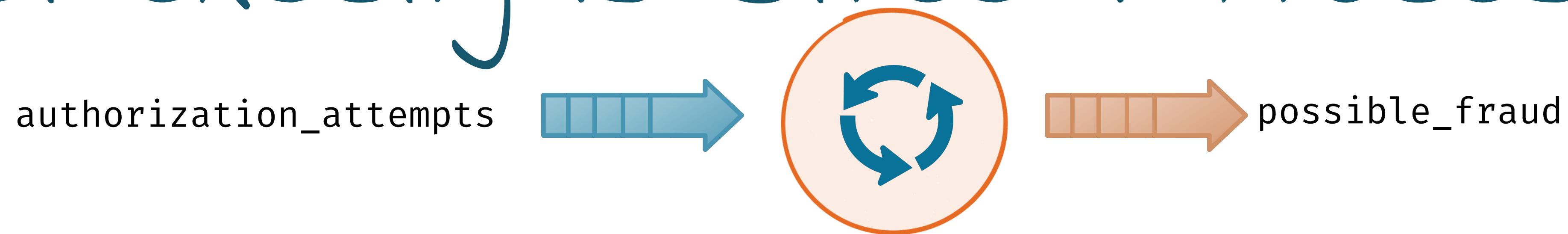
?

what exactly is Stream Processing?



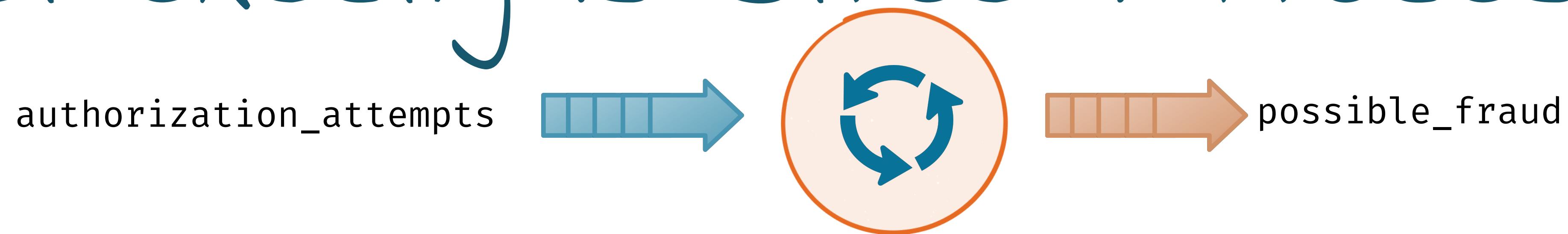
```
CREATE STREAM possible_fraud AS
SELECT card_number, count(*)
FROM authorization_attempts
WINDOW TUMBLING (SIZE 5 MINUTE)
GROUP BY card_number
HAVING count(*) > 3;
```

what exactly is Stream Processing?



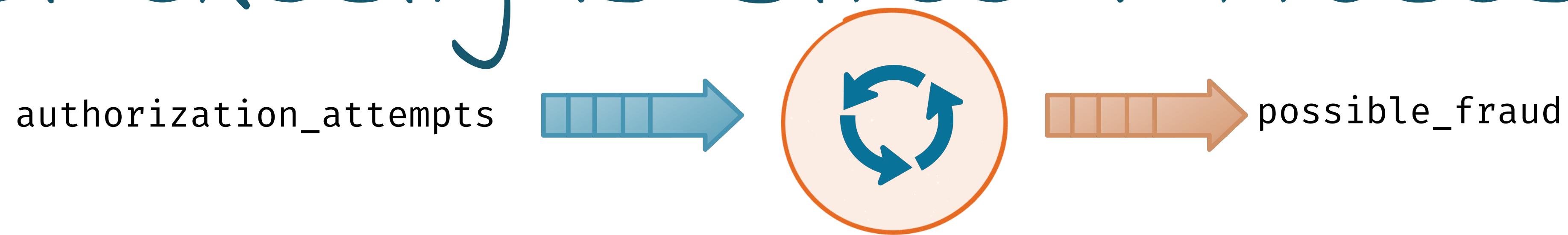
```
CREATE STREAM possible_fraud AS  
SELECT card_number, count(*)  
FROM authorization_attempts  
WINDOW TUMBLING (SIZE 5 MINUTE)  
GROUP BY card_number  
HAVING count(*) > 3;
```

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?

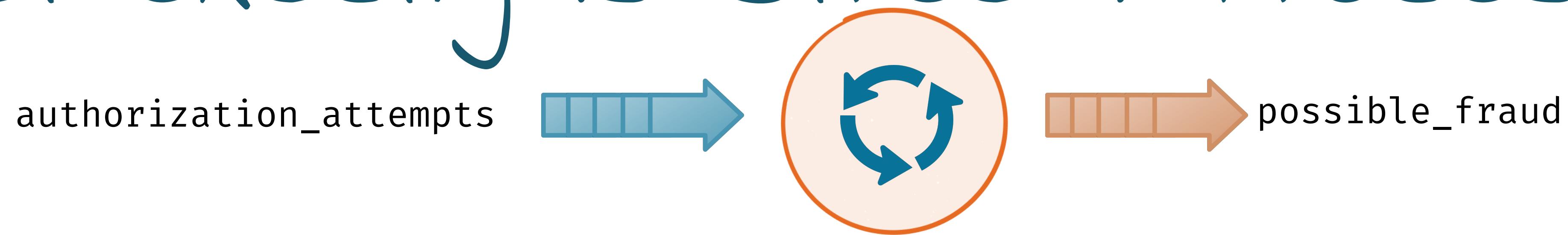


```
CREATE STREAM possible_fraud AS  
SELECT card_number, count(*)  
FROM authorization_attempts
```

WINDOW TUMBLING (SIZE 5 MINUTE)

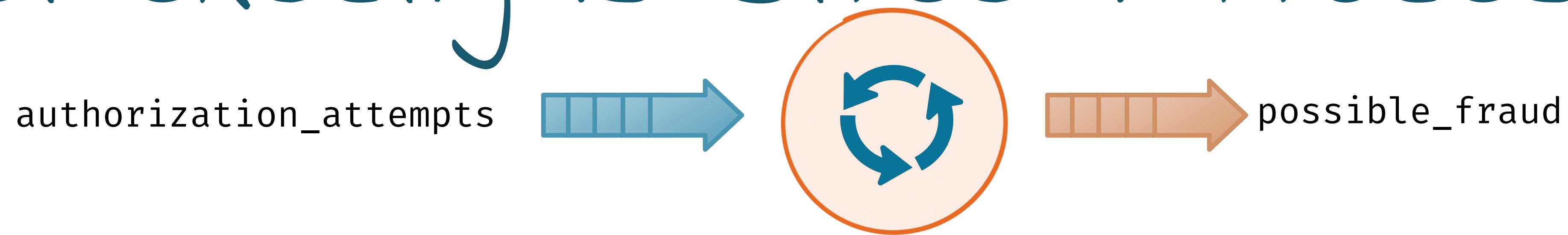
```
GROUP BY card_number  
HAVING count(*) > 3;
```

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;
```

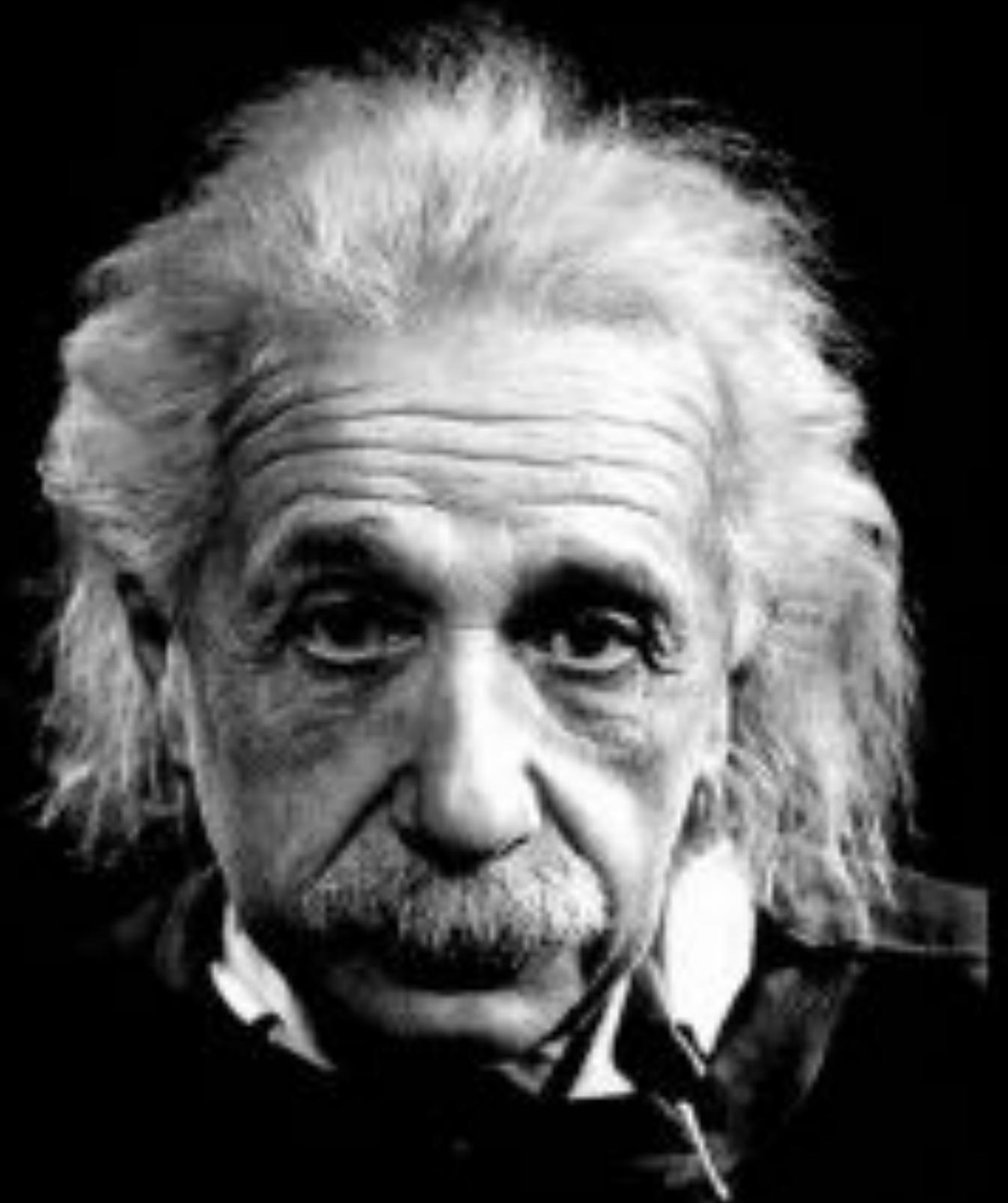
what exactly is Stream Processing?



```
CREATE STREAM possible_fraud AS  
SELECT card_number, count(*)  
FROM authorization_attempts  
WINDOW TUMBLING (SIZE 5 MINUTE)
```

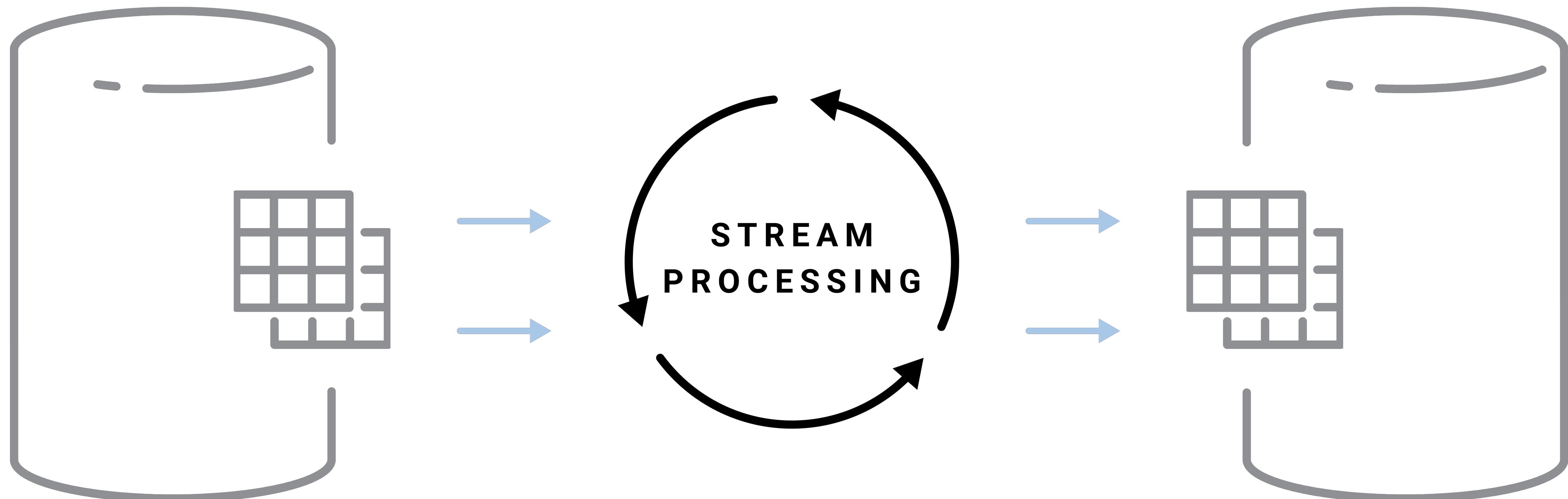
```
GROUP BY card_number  
HAVING count(*) > 3;
```

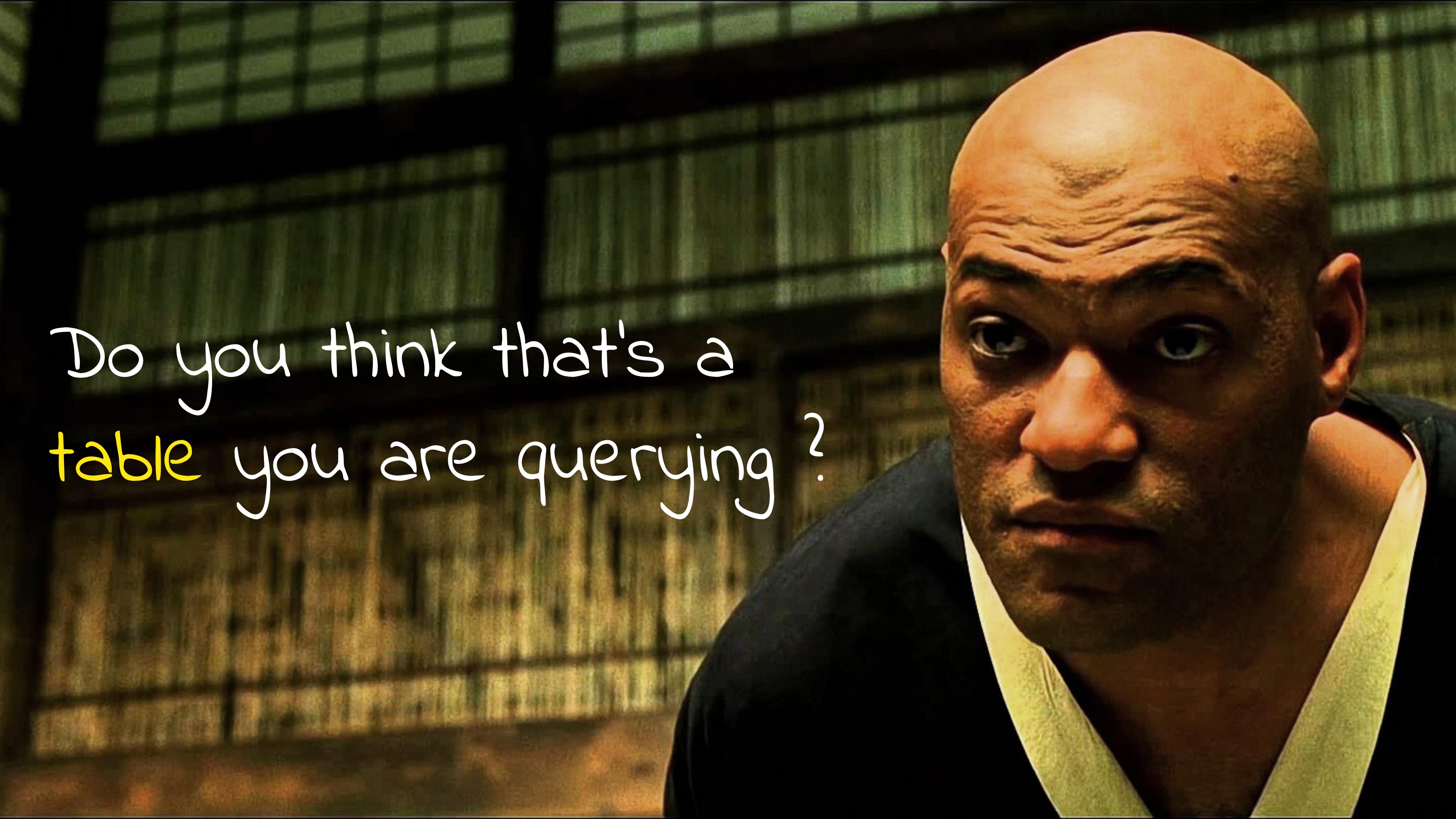
SPECIAL RELATIVITY



OF STREAMS AND TABLES

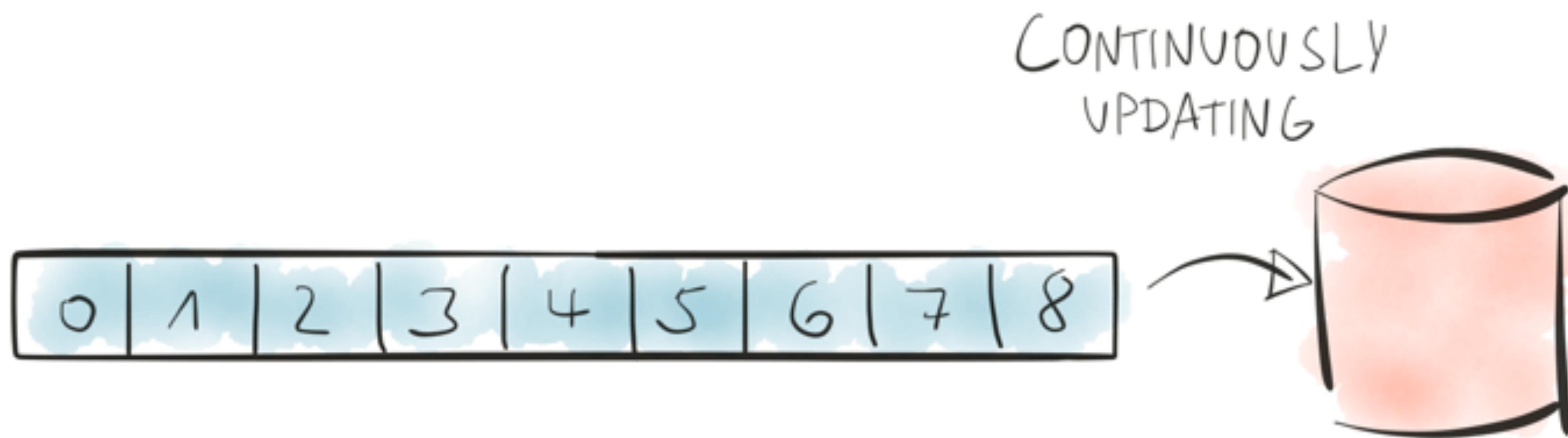
Table-Stream Duality

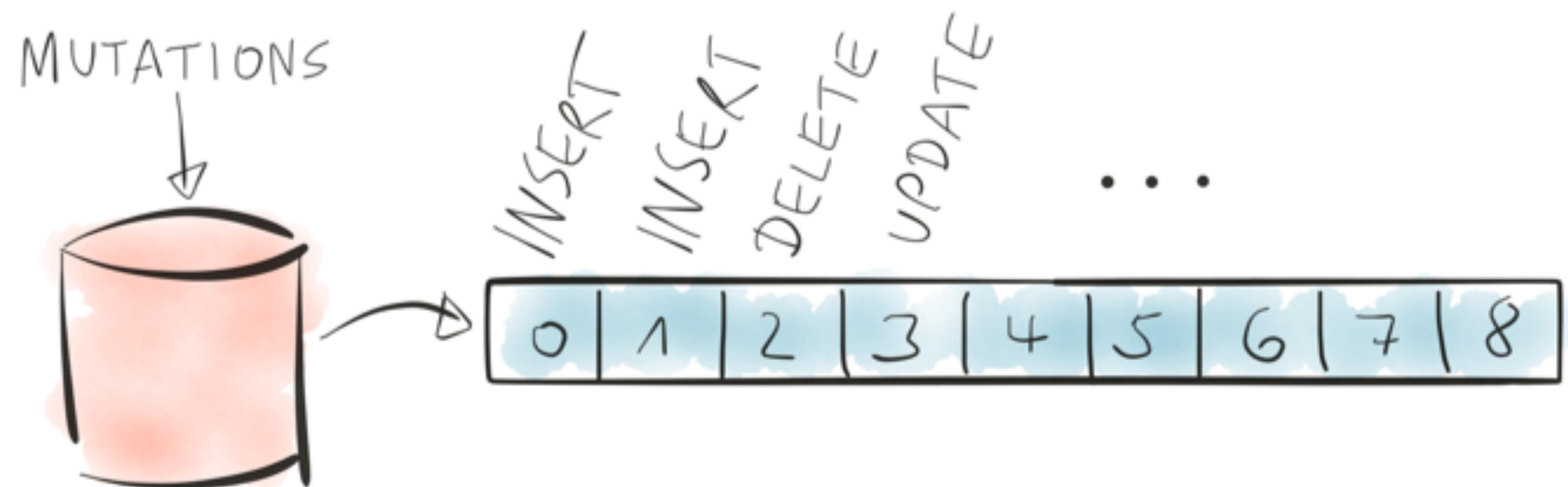


A close-up photograph of a man with a bald head and a serious, slightly shocked expression. He has deep-set eyes and a thin mustache. He is wearing a dark suit jacket over a white shirt. The background is blurred, showing what appears to be a row of bookshelves or a wall of books.

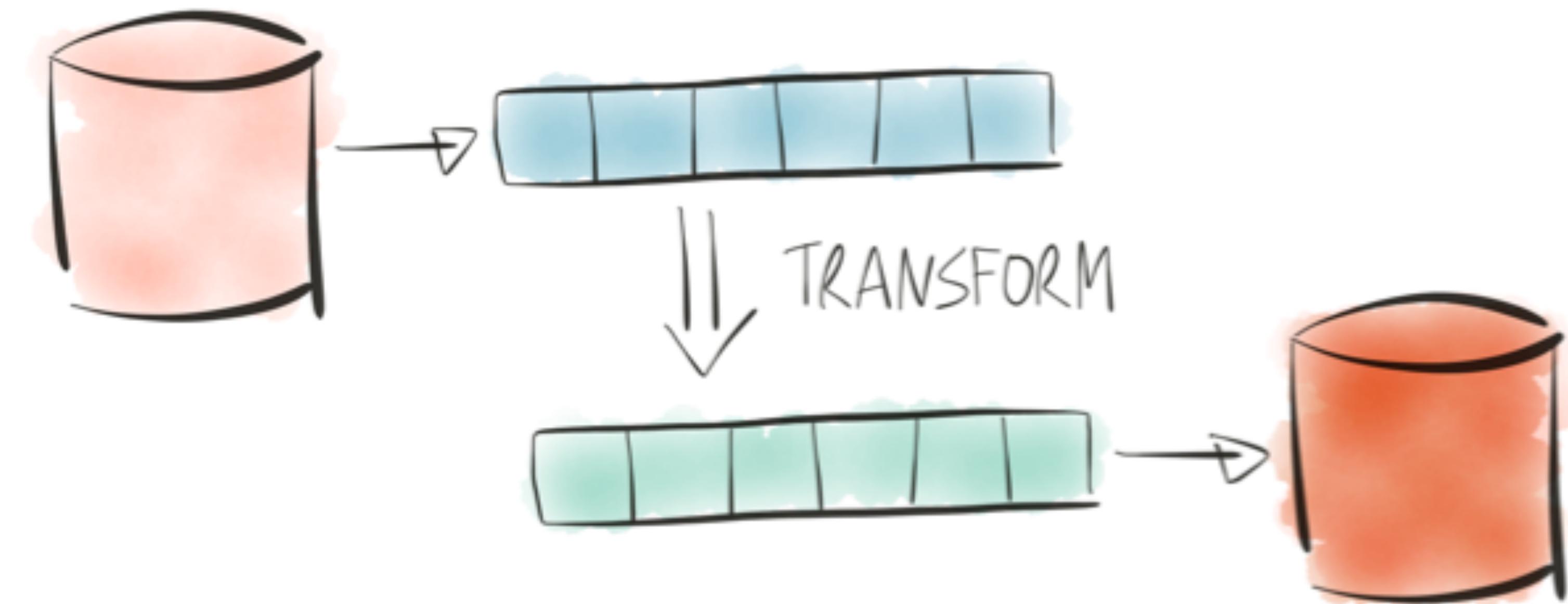
Do you think that's a
table you are querying ?

Streams to Tables



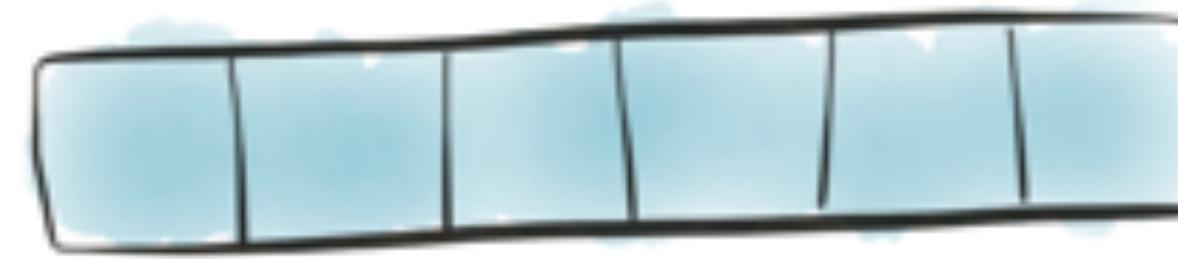


Stream/Table Duality

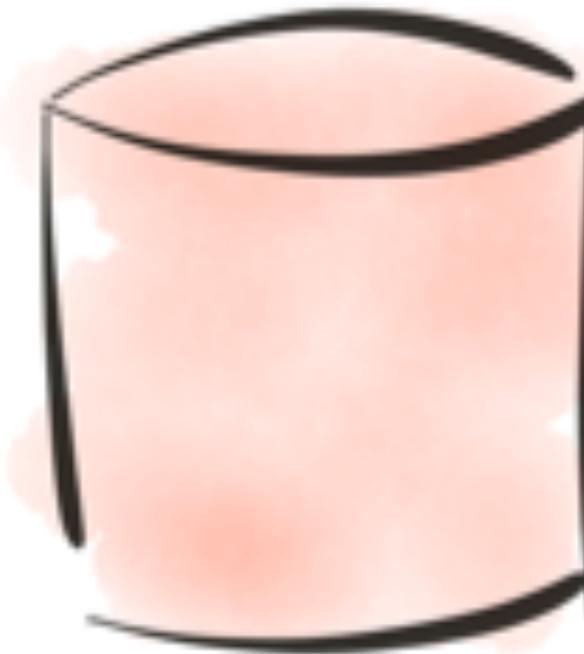


Stream/Table Duality

FACTS



JOIN



DIMENSIONS



ENRICHED
STREAM

TABLE

| | |
|------|---|
| Gwen | 1 |
|------|---|

| | |
|----------|---|
| Gwen | 1 |
| Matthias | 1 |

| | |
|----------|---|
| Gwen | 2 |
| Matthias | 1 |

| | |
|----------|---|
| Gwen | 2 |
| Matthias | 1 |
| Viktor | 1 |

STREAM

(“Gwen”, 1)

TABLE

| | |
|------|---|
| Gwen | 1 |
|------|---|

| | |
|----------|---|
| Gwen | 1 |
| Matthias | 1 |

| | |
|----------|---|
| Gwen | 2 |
| Matthias | 1 |

| | |
|----------|---|
| Gwen | 2 |
| Matthias | 1 |
| Viktor | 1 |

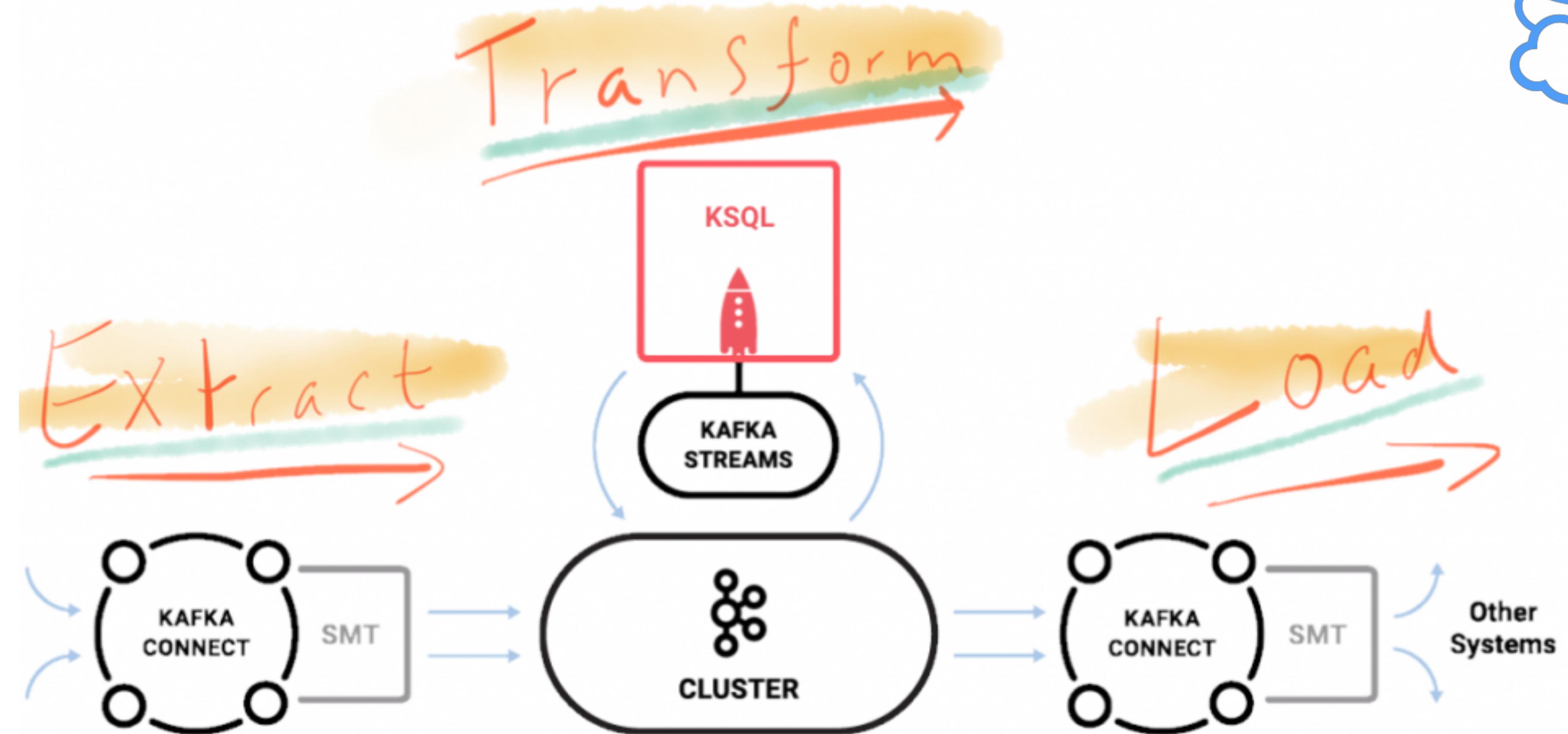
Demo



@gamussa

#DataSciCon

@confluentinc

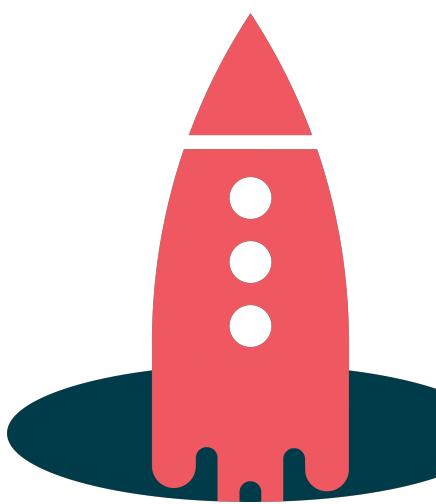


confluent

@gamusss

#DataSciCon

@confluentinc



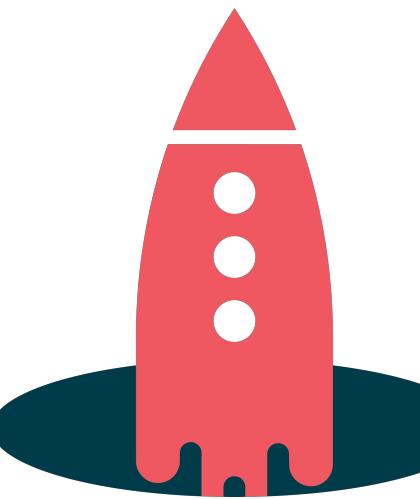
where is KSQL not such a great fit?

Ad-hoc queries

- Limited span of time usually retained in Kafka
- No indexes

BI reports (Tableau etc.)

- No indexes
- No JDBC (most BI tools are not good with continuous results!)



Resources and Next Steps



<https://github.com/confluentinc/ksql>



<http://confluent.io/ksql>



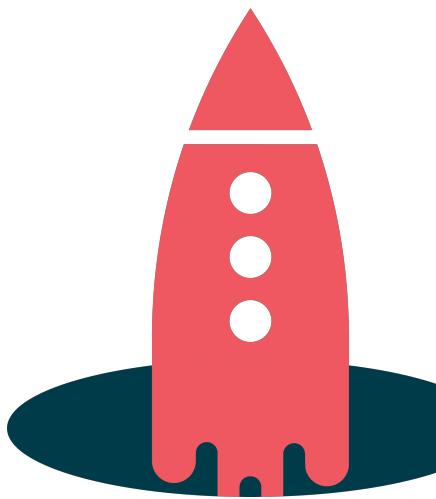
<https://slackpass.io/confluentcommunity> #ksql



@gamussa

#DataSciCon

@confluentinc



THANKS!



@gamussa
viktor@confluent.io

we are hiring!
<https://www.confluent.io/careers/>