

The Changing Face of ETL

Event-Driven Architectures
for Data Engineers

@rmoff









Photo by Rohit Tandon on Unsplash





It used to be so simple





More Sources



More Targets



More Data

Batches and Buckets





Applications
Respond

→ an order was placed!

Analytics

**Tell Us What
Happened**

→ how many orders
were placed







Events

An event is both:[”]

★ Notification

★ State transfer

A Customer Experience



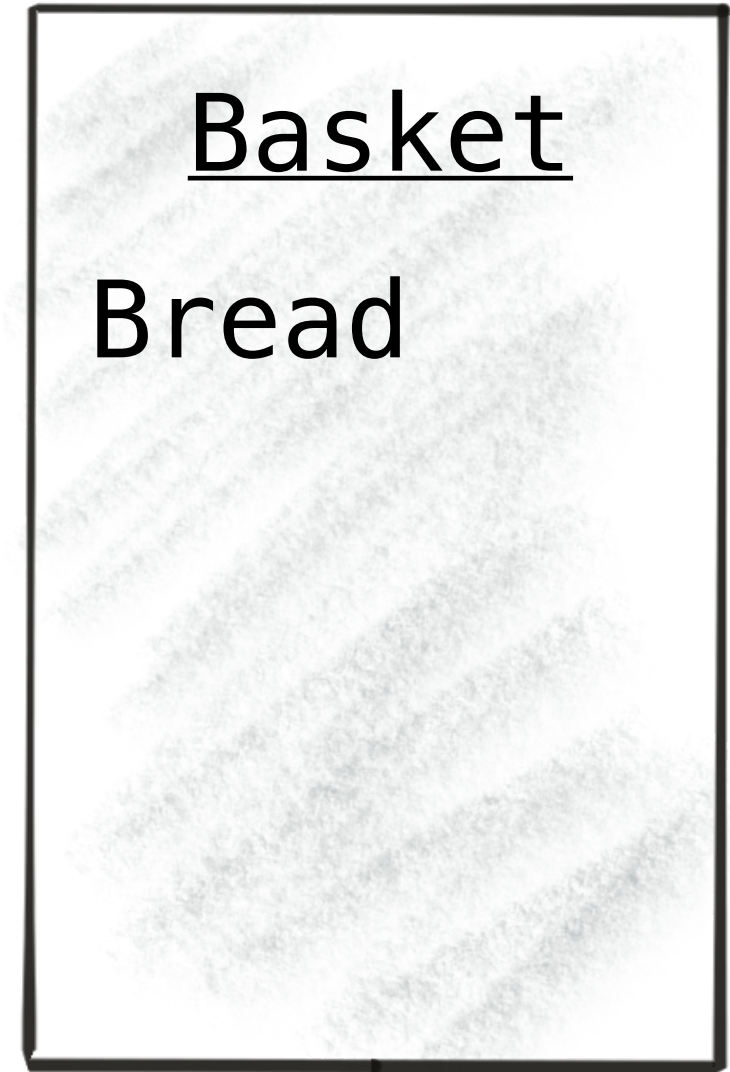
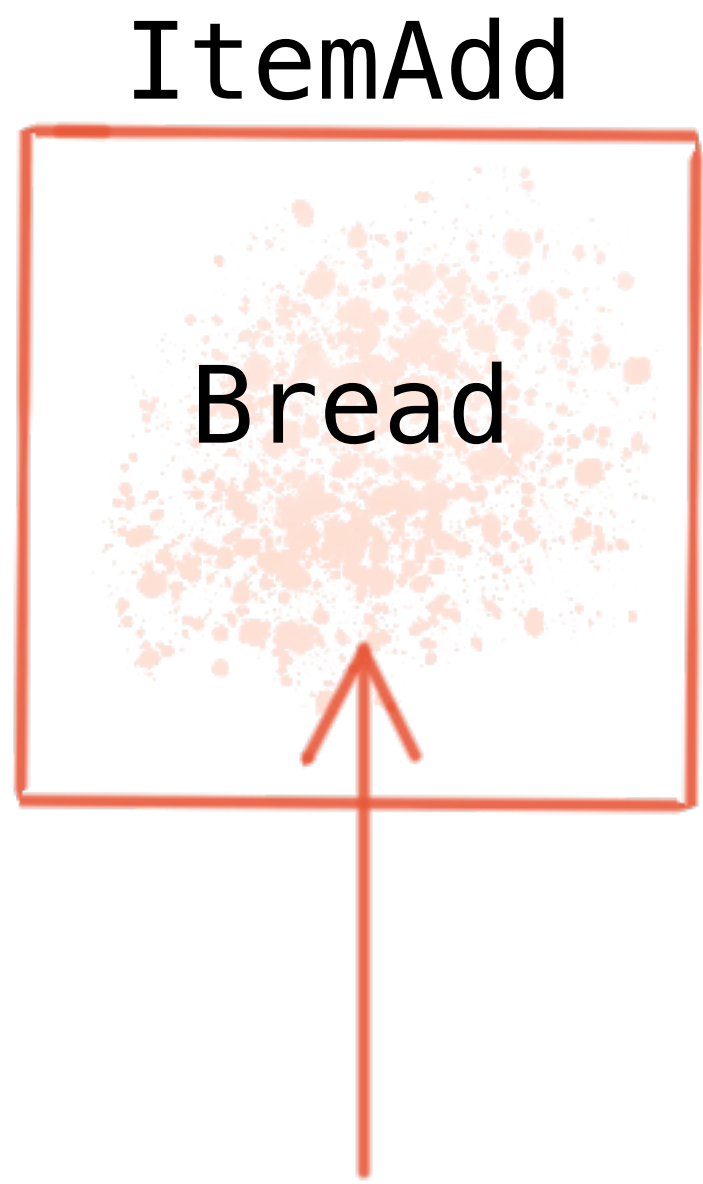
A Sensor Reading



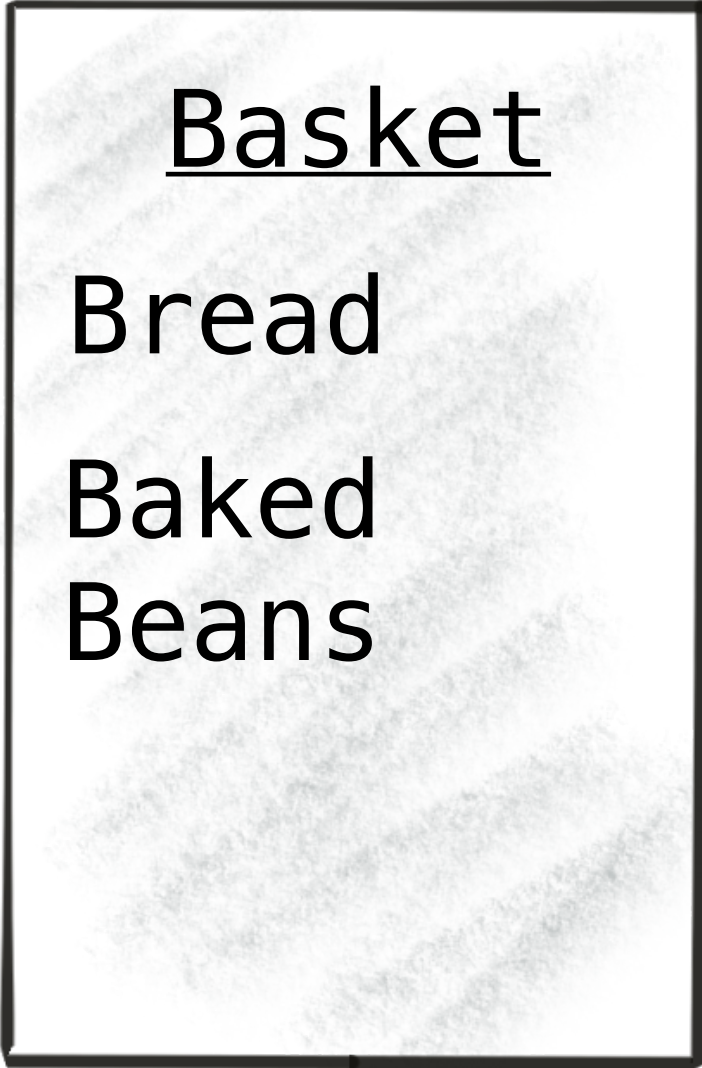
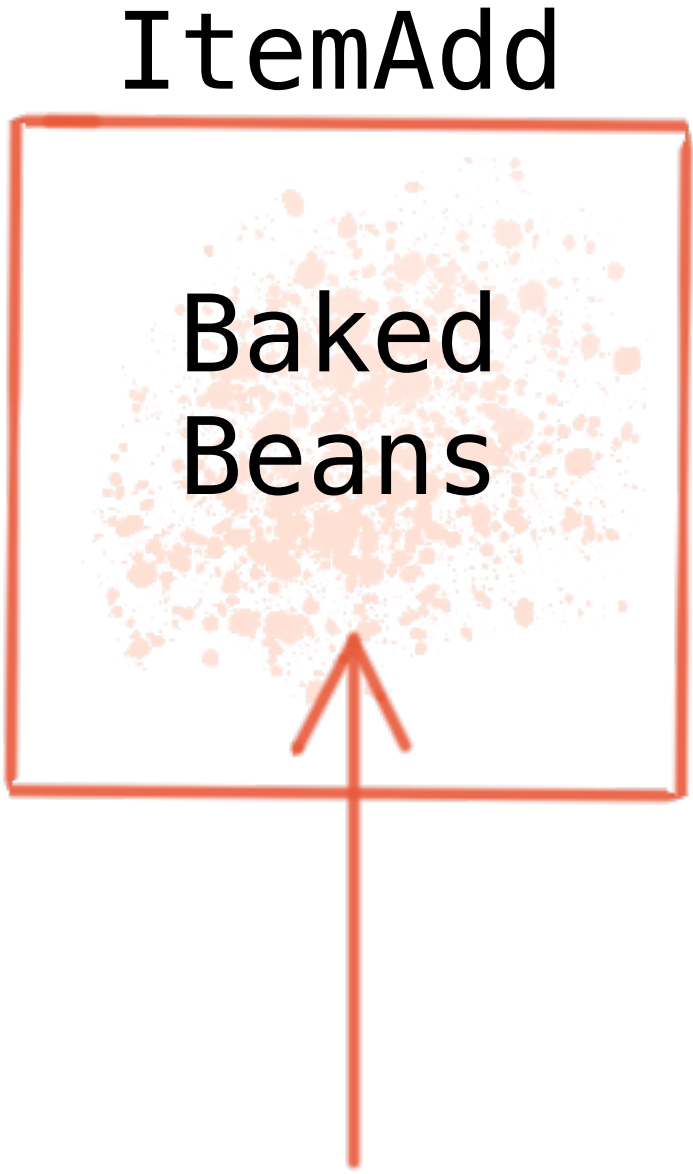
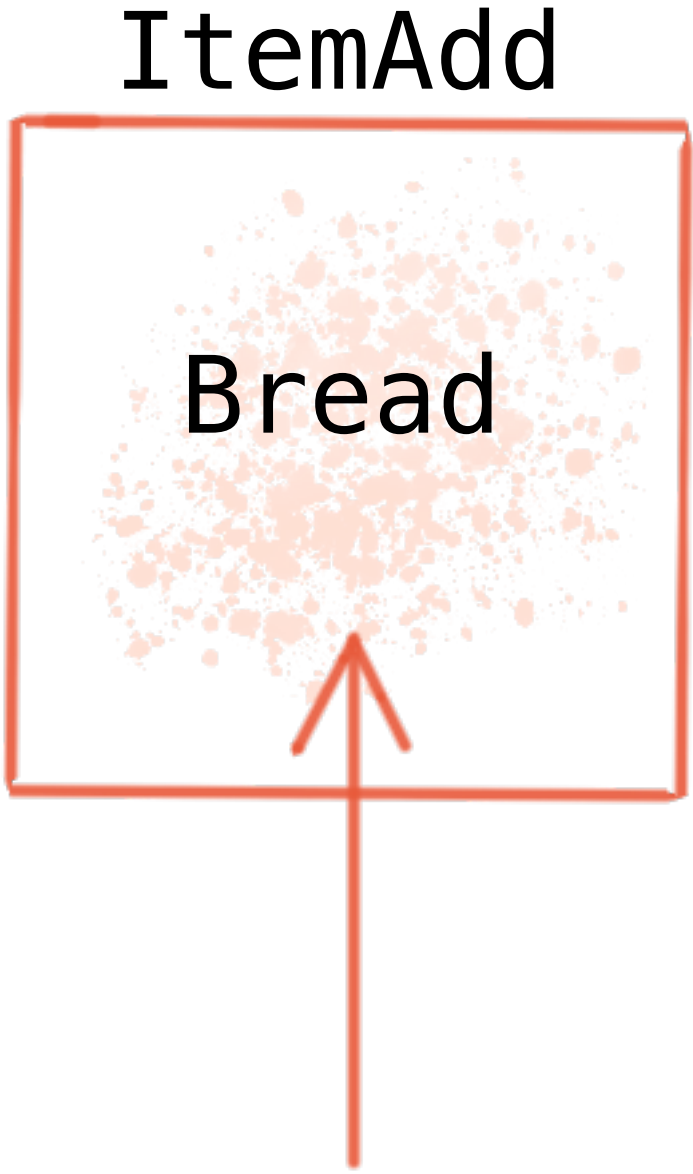
Events



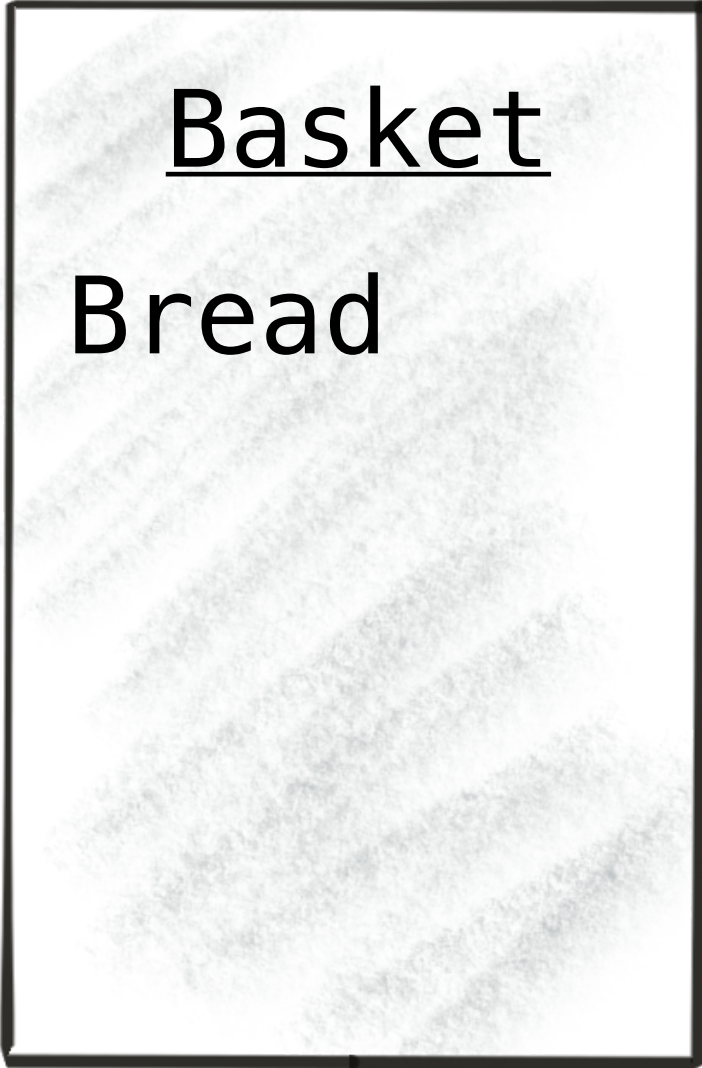
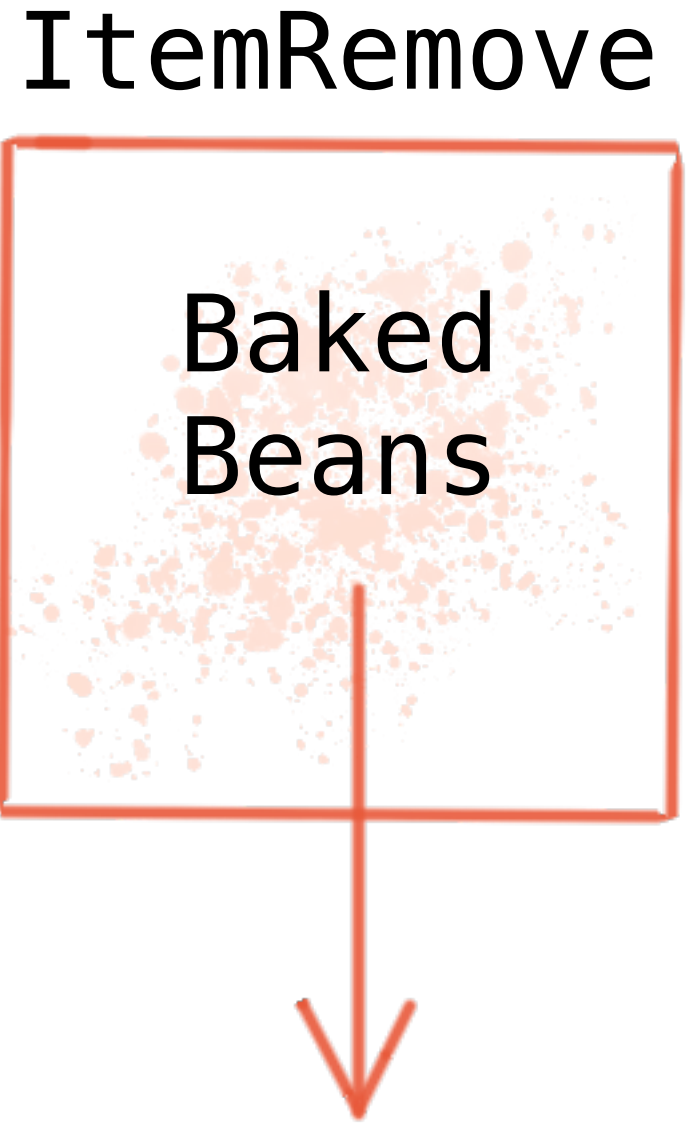
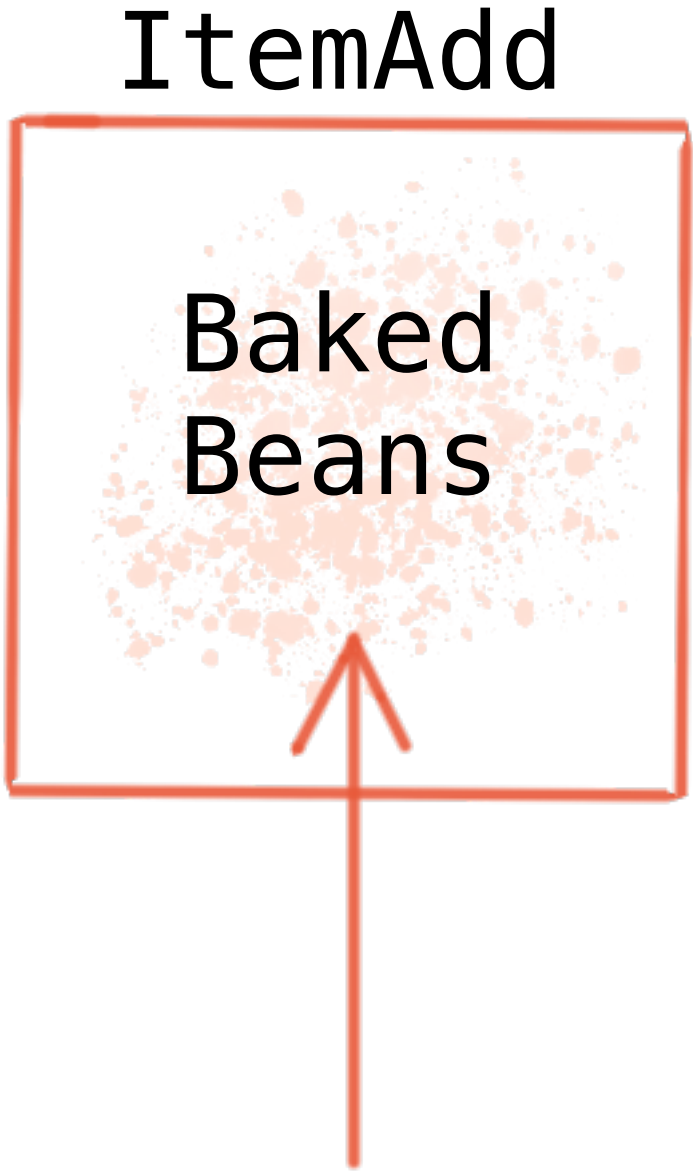
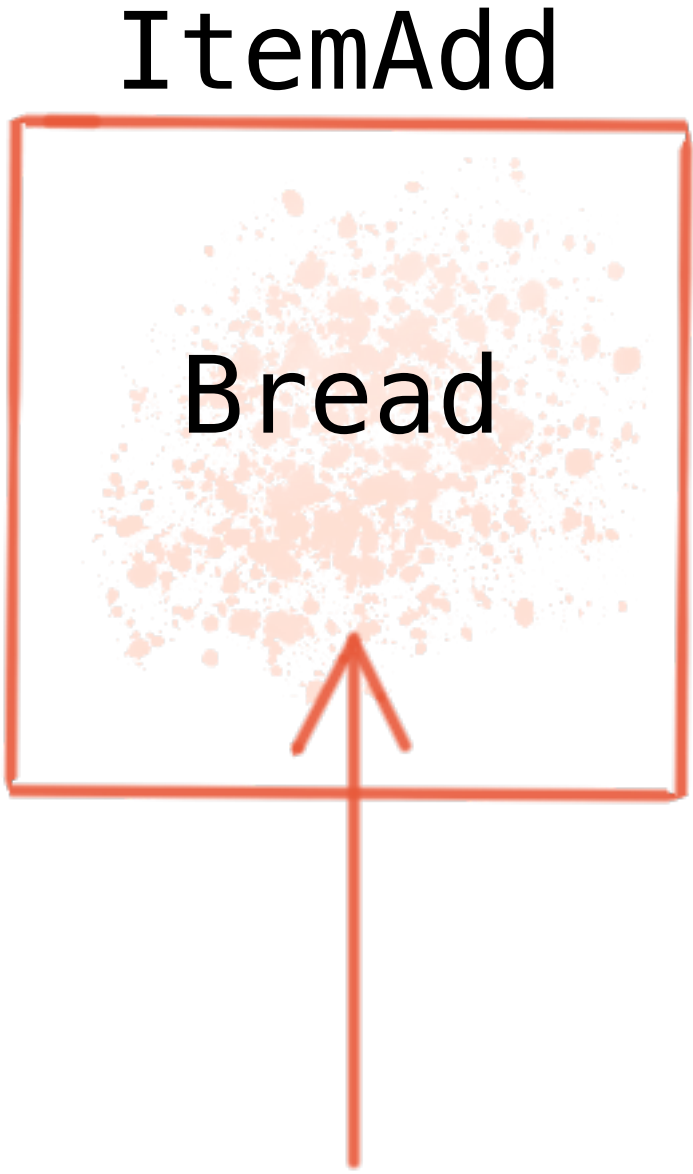
Events



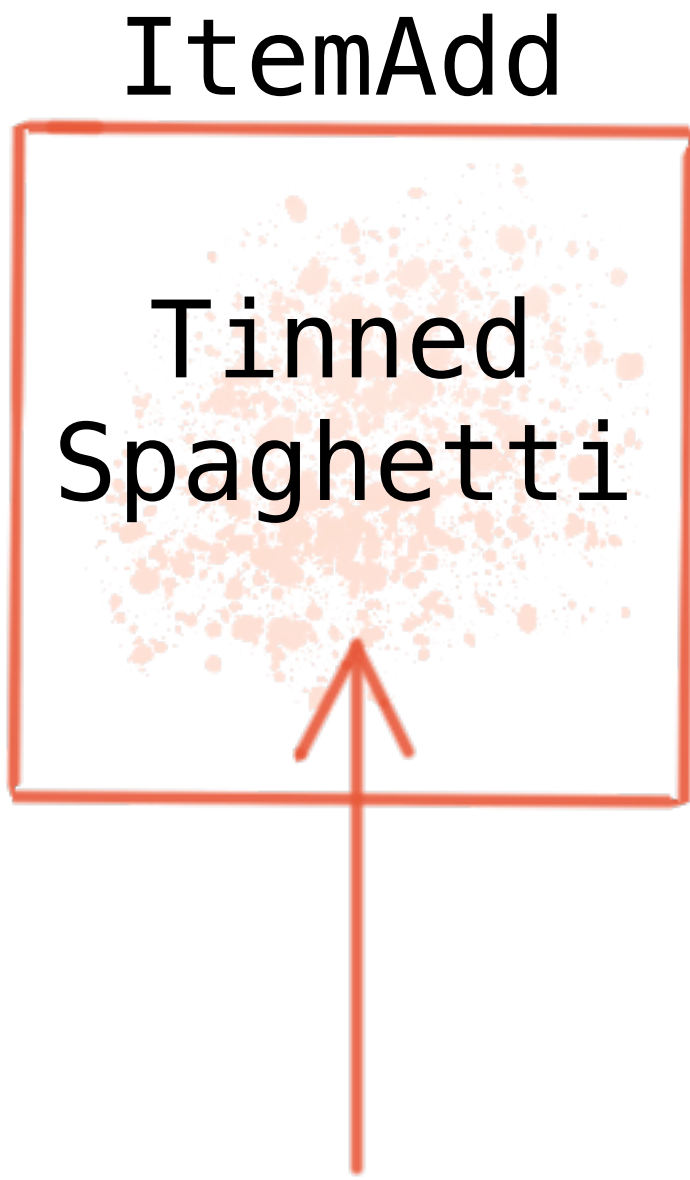
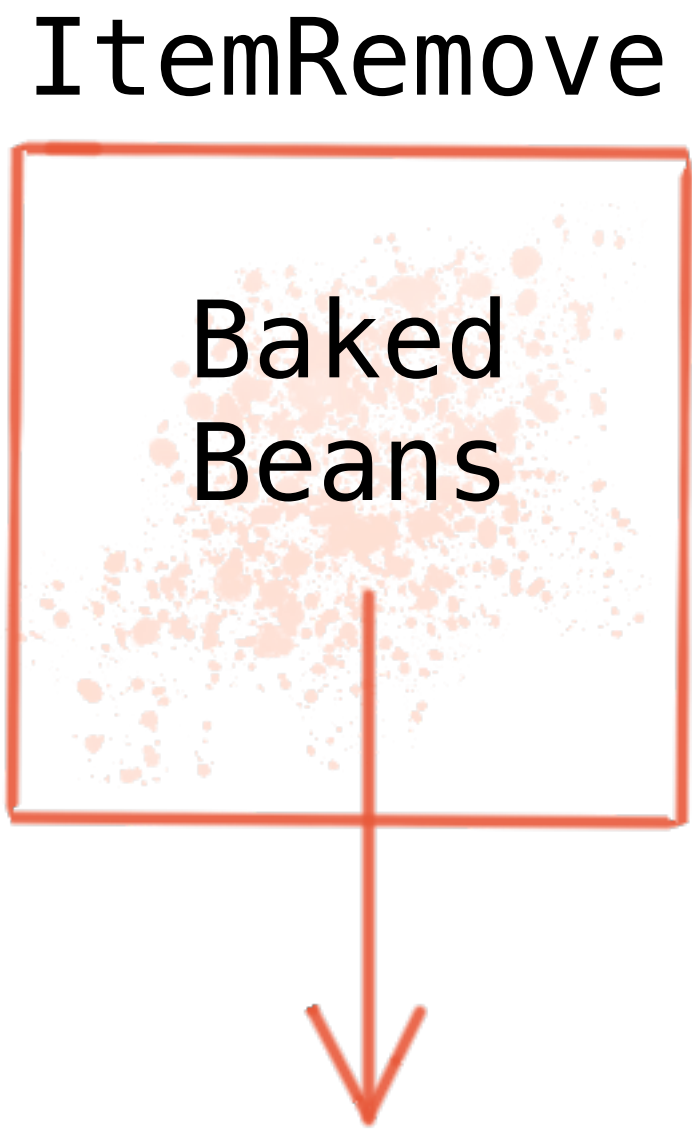
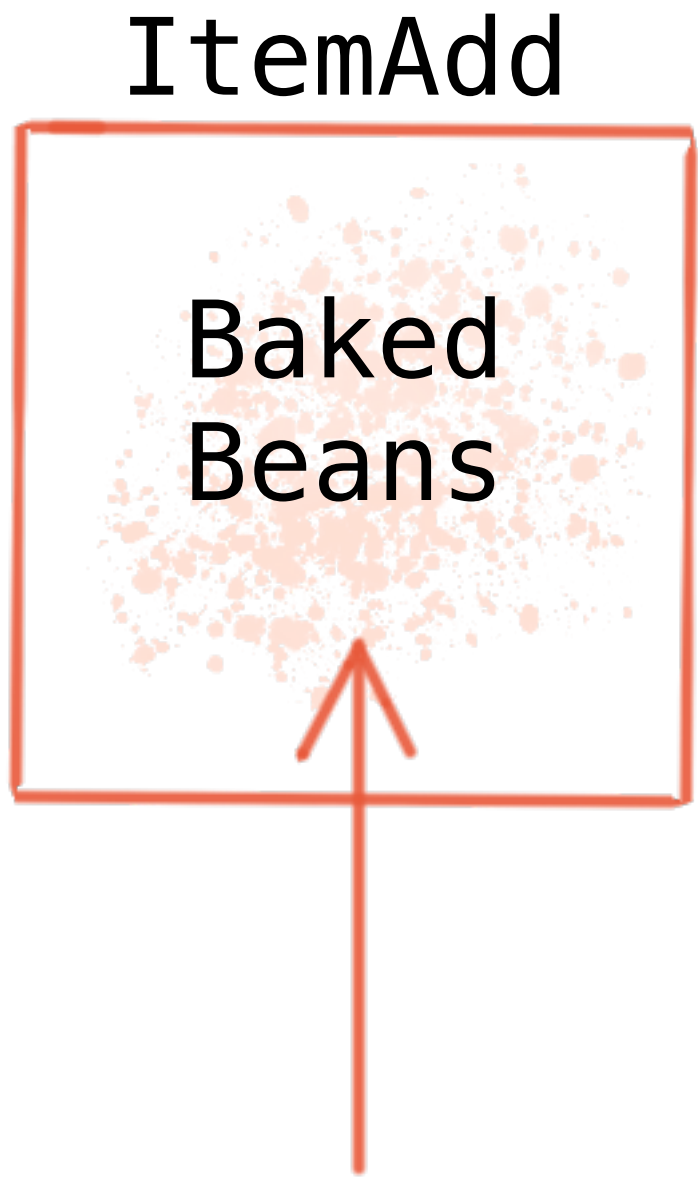
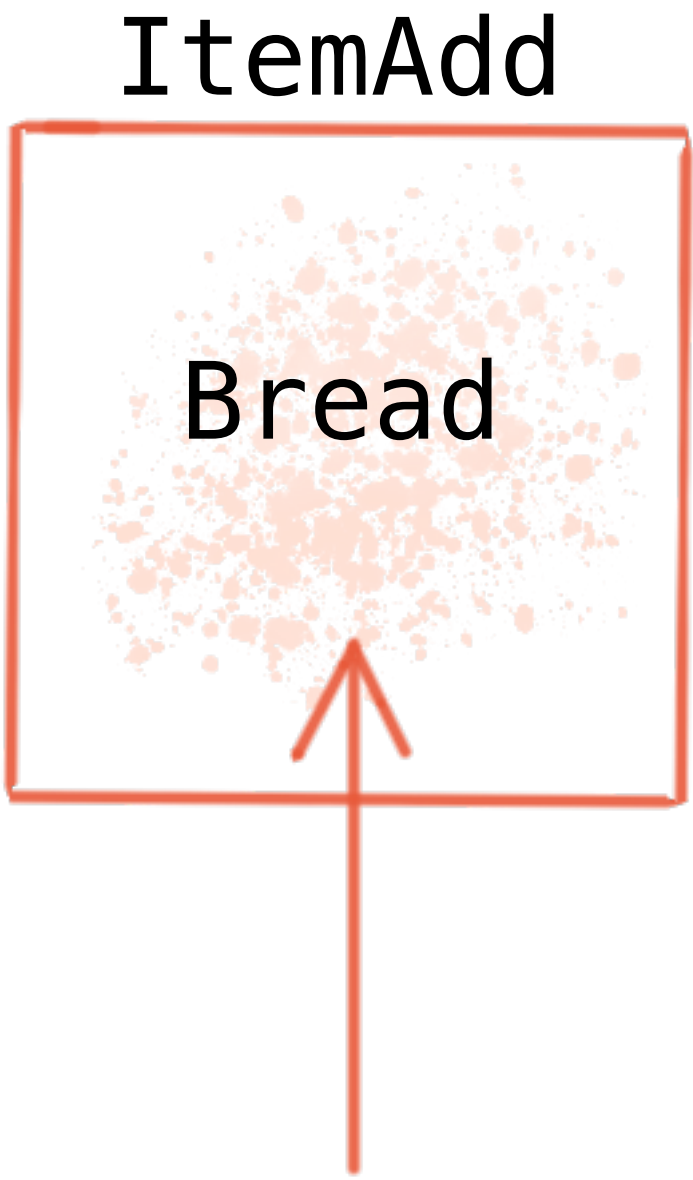
Events



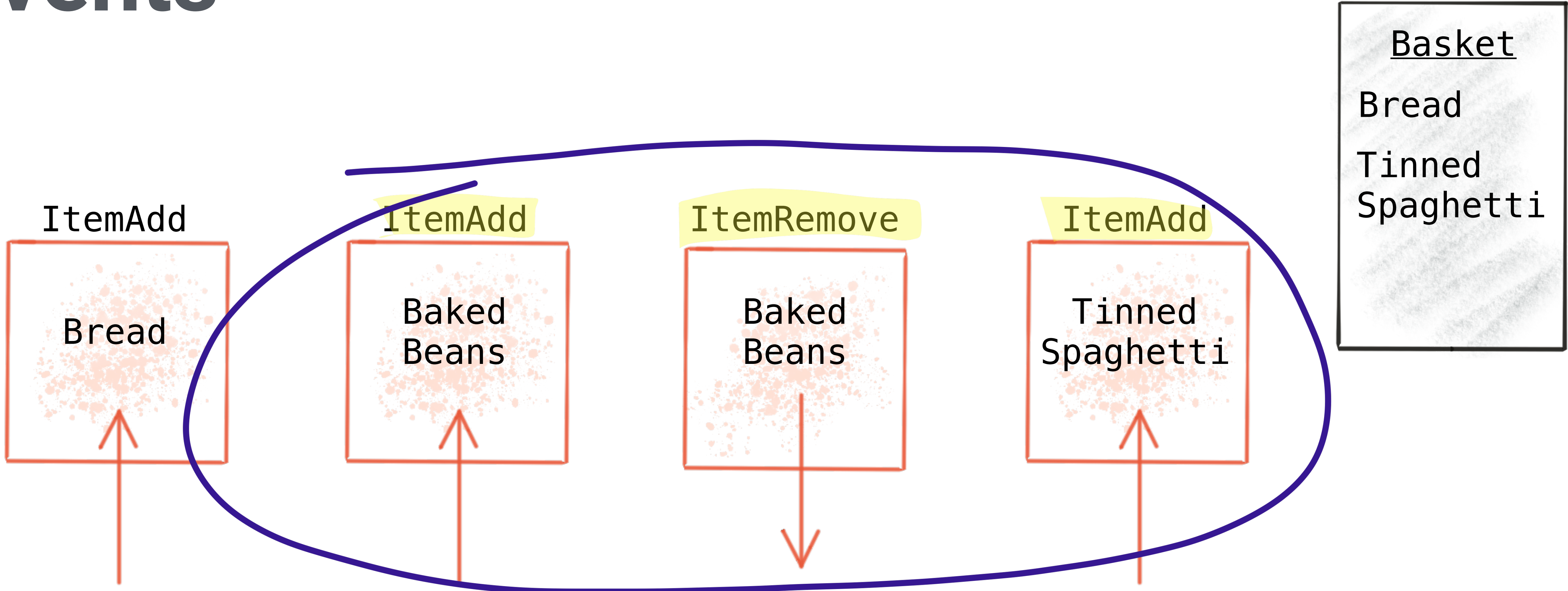
Events



Events

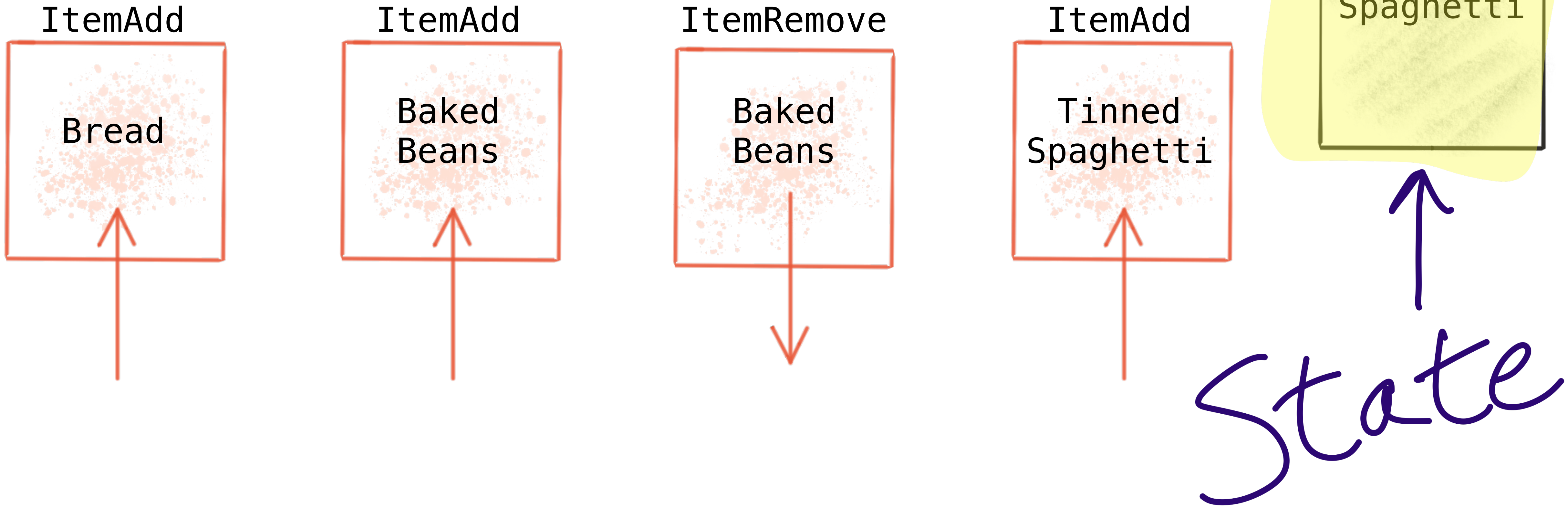


Events



Behaviour

Events

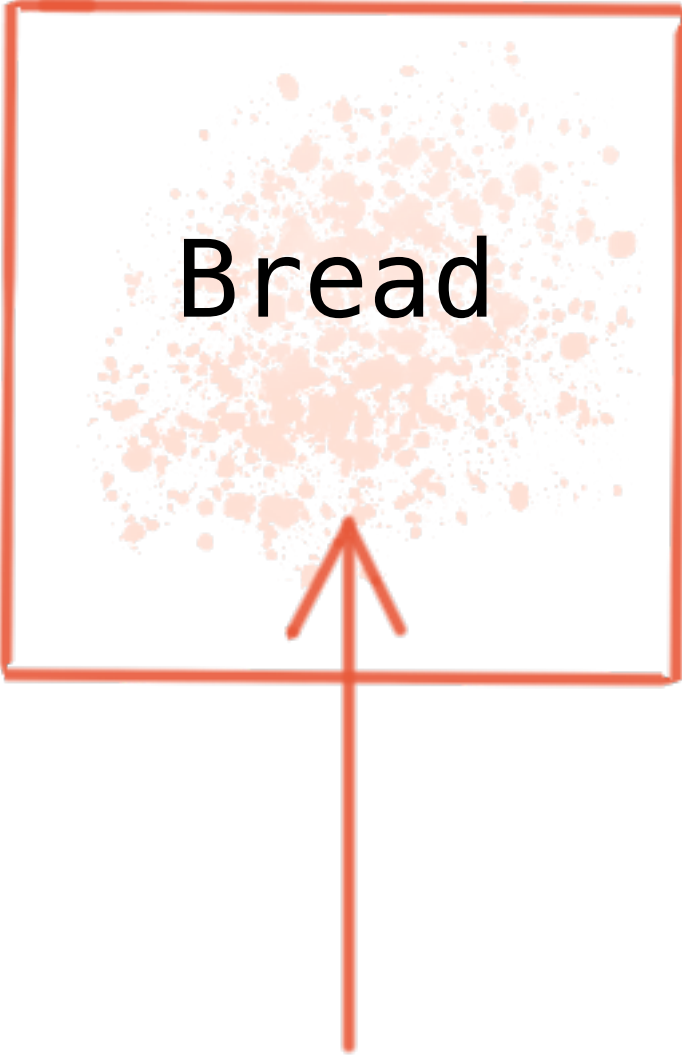


Events

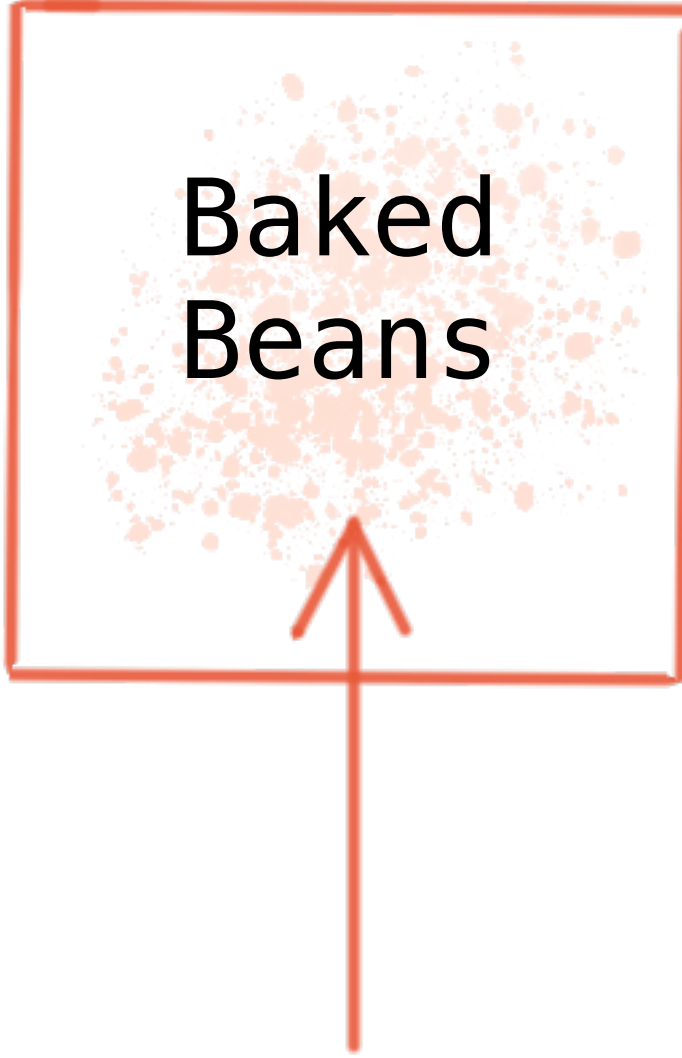
Event Stream



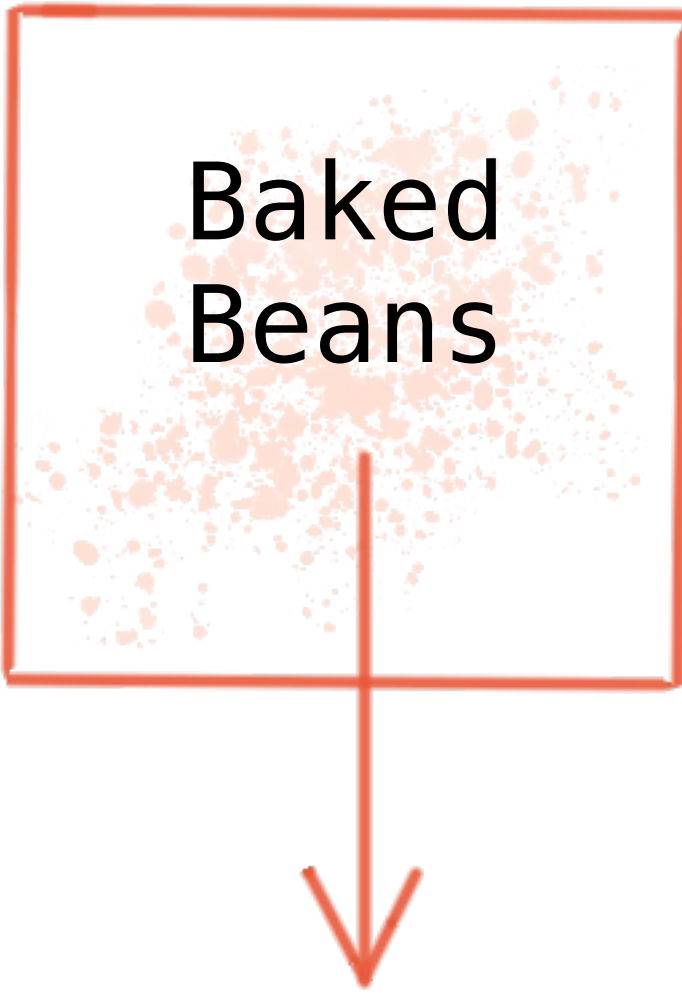
ItemAdd



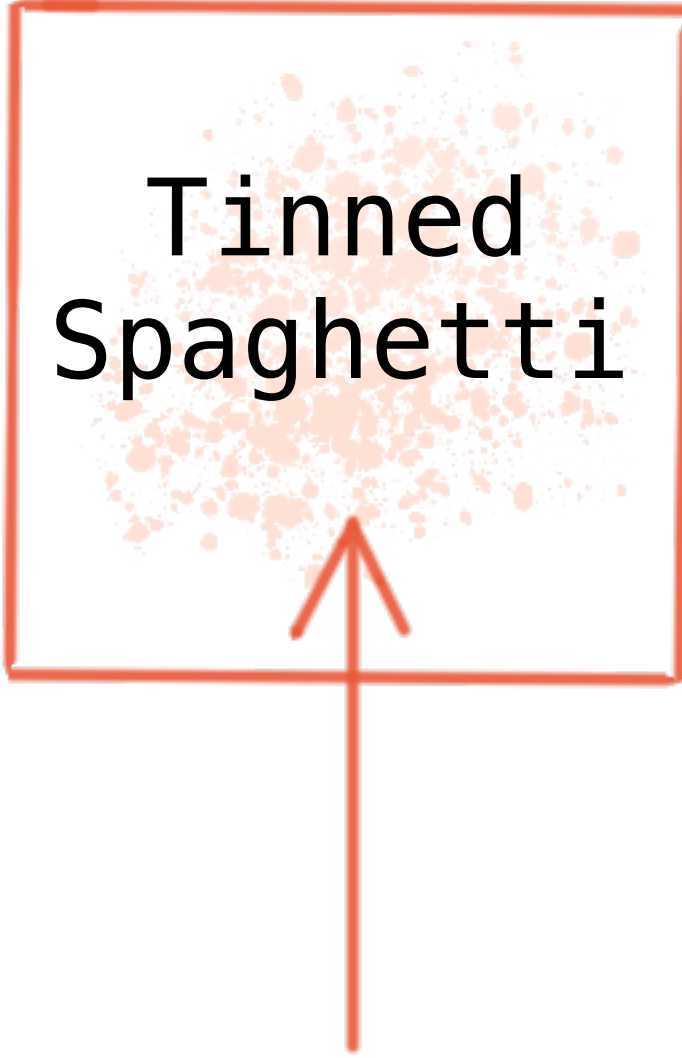
ItemAdd



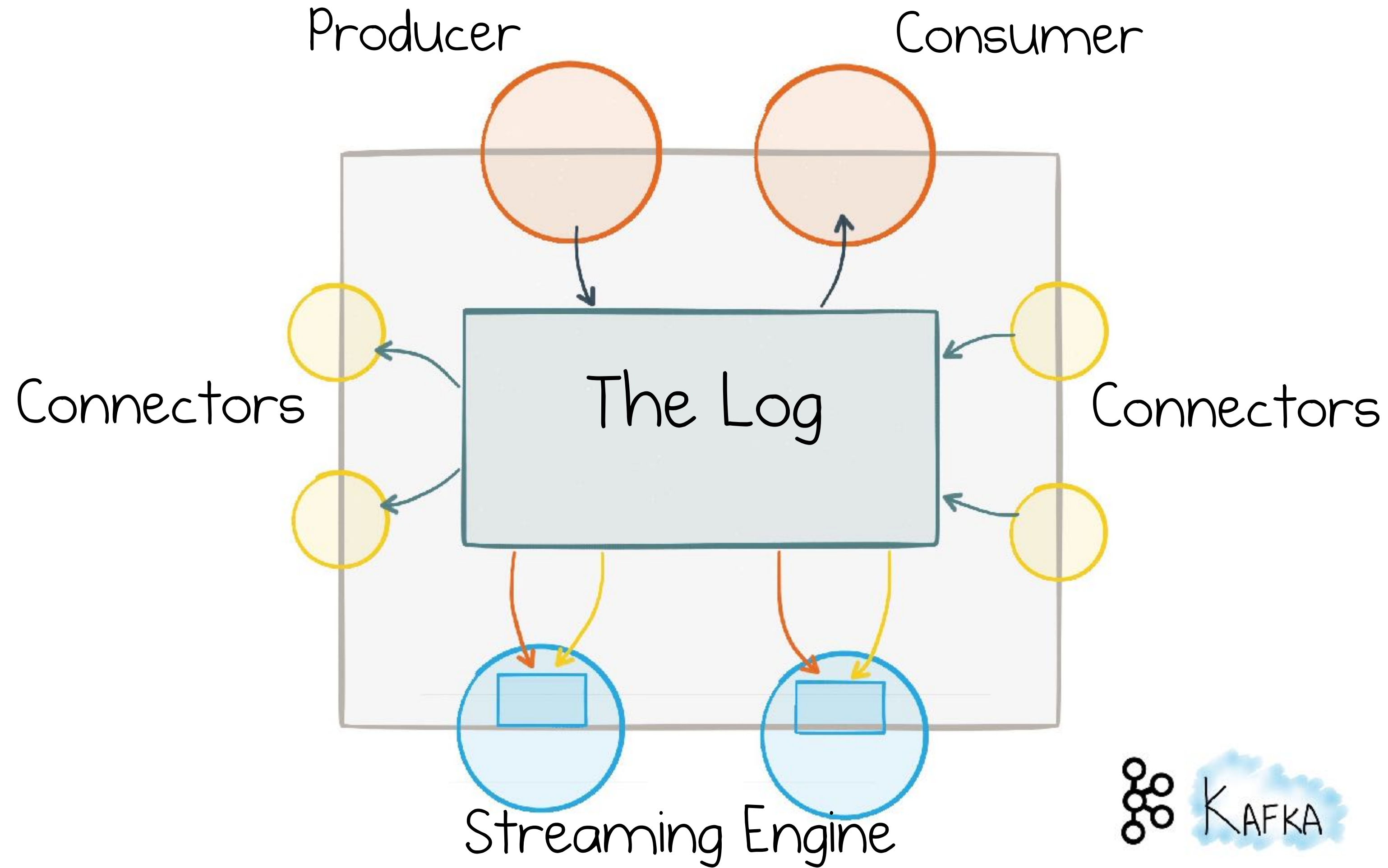
ItemRemove



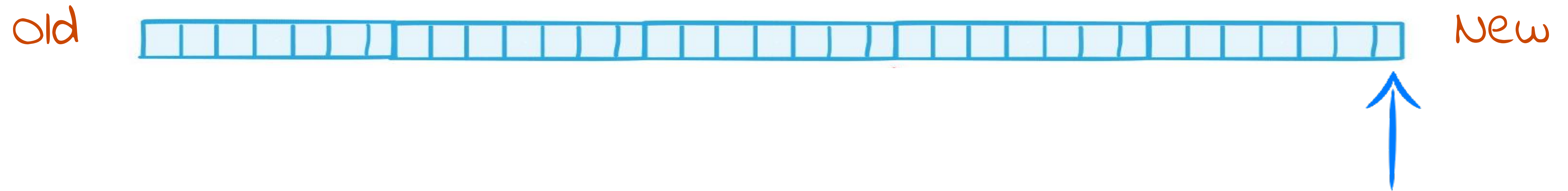
ItemAdd



What is an Event Streaming Platform?

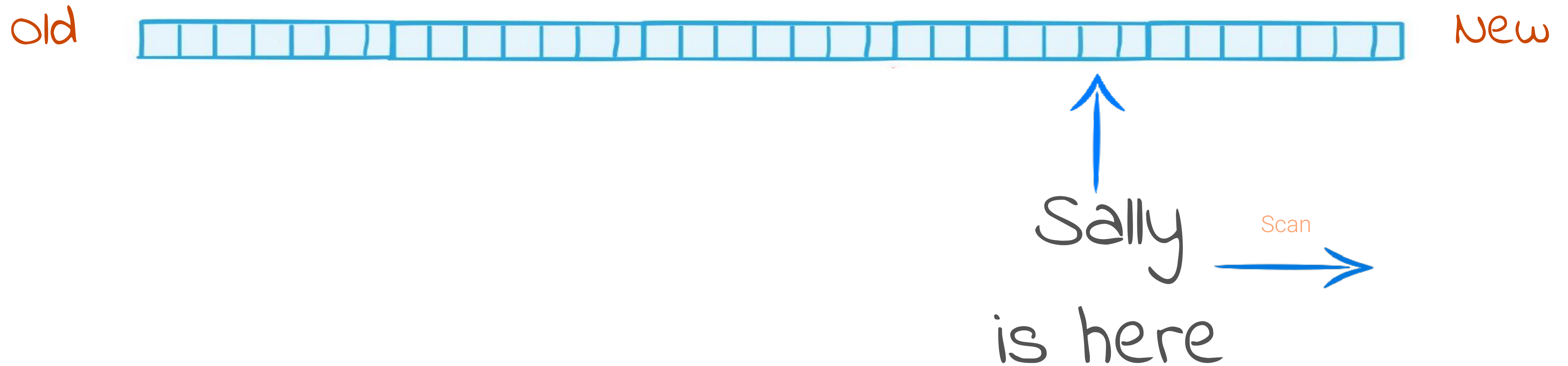


Immutable Event Log

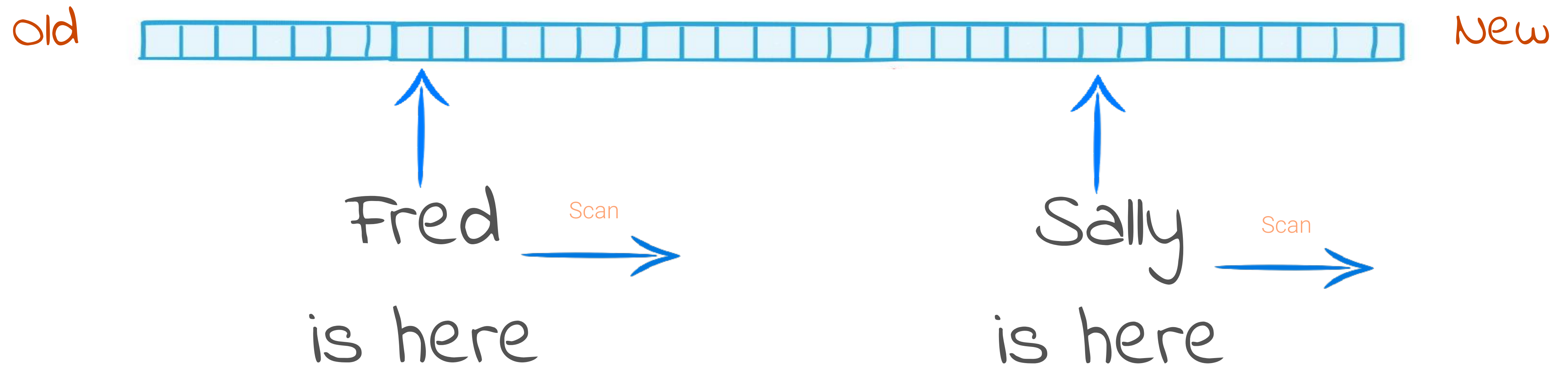


Messages are added at the end of the log

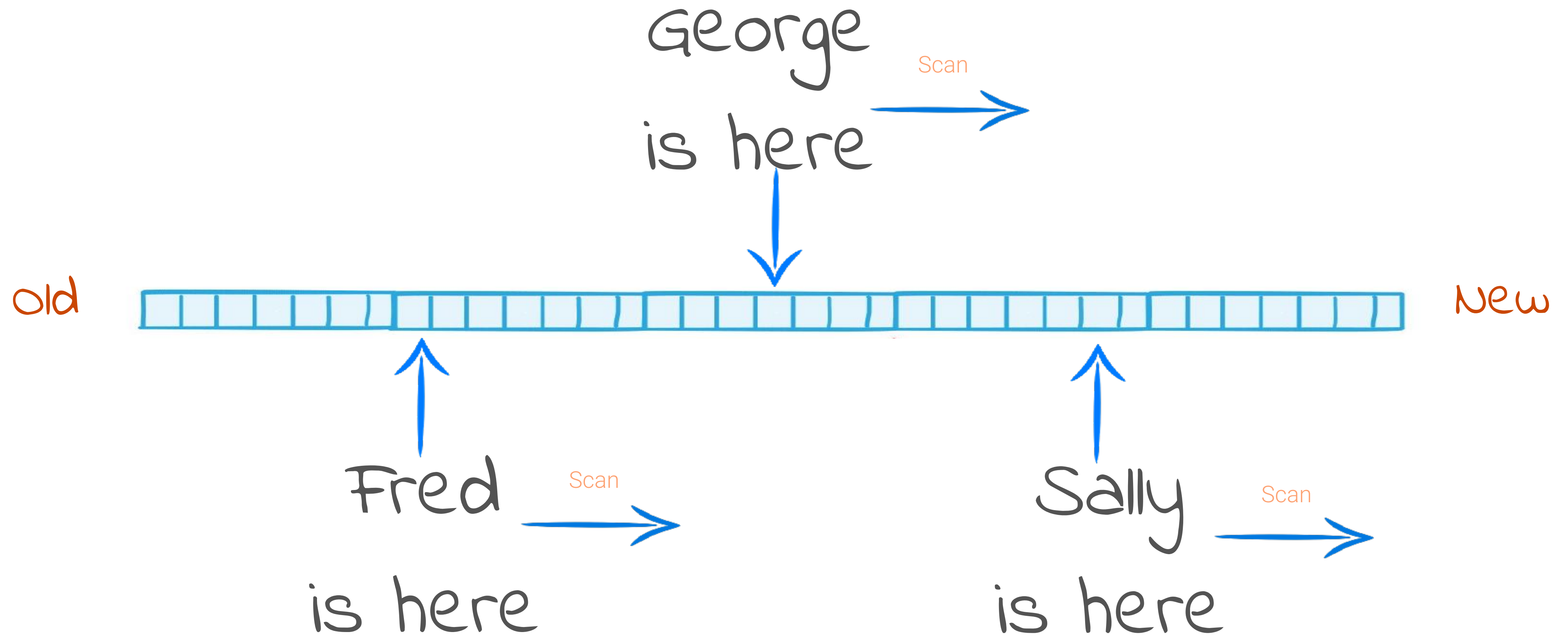
Consumers have a position all of their own



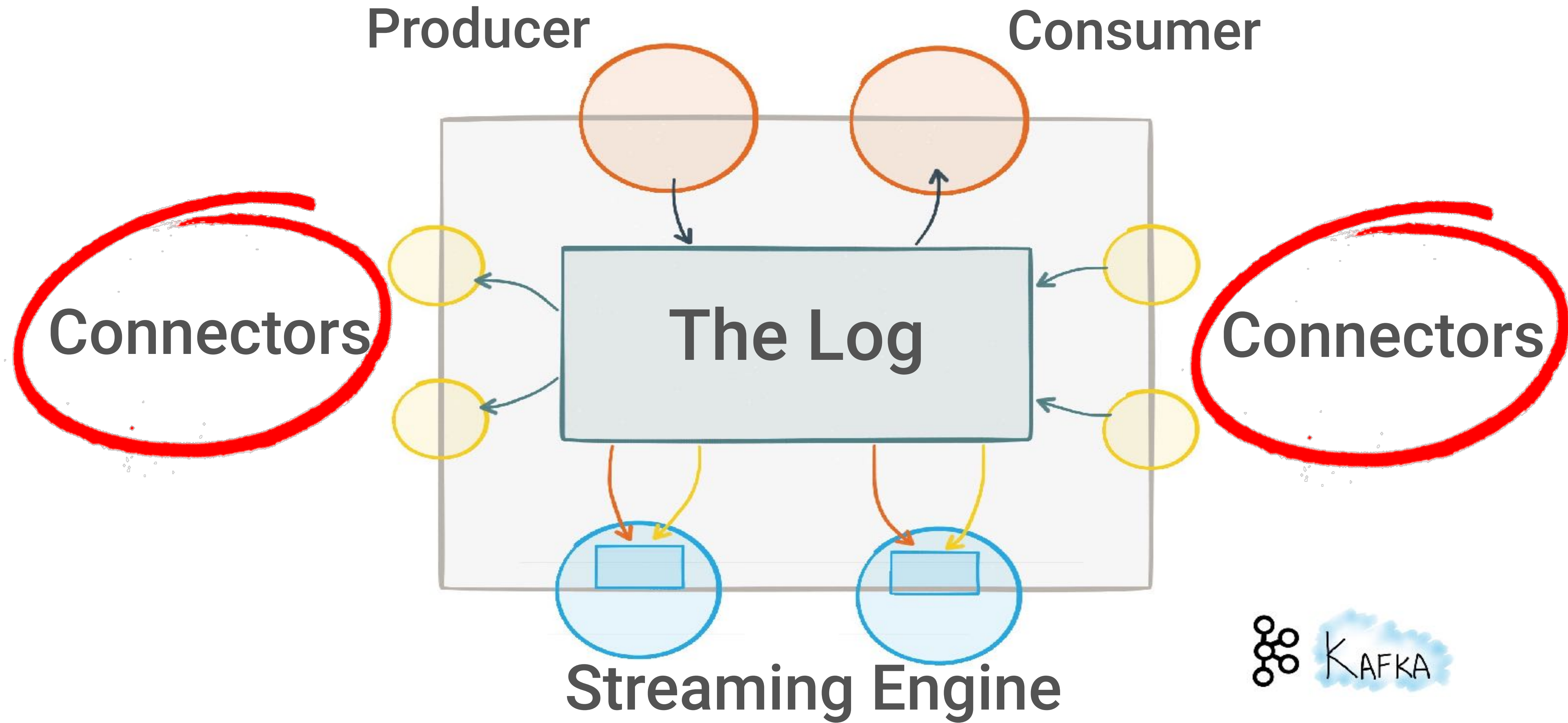
Consumers have a position all of their own



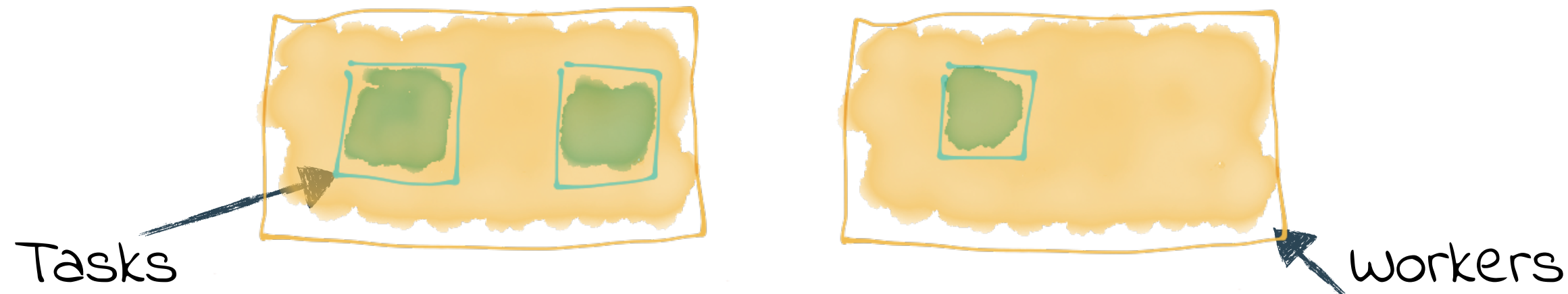
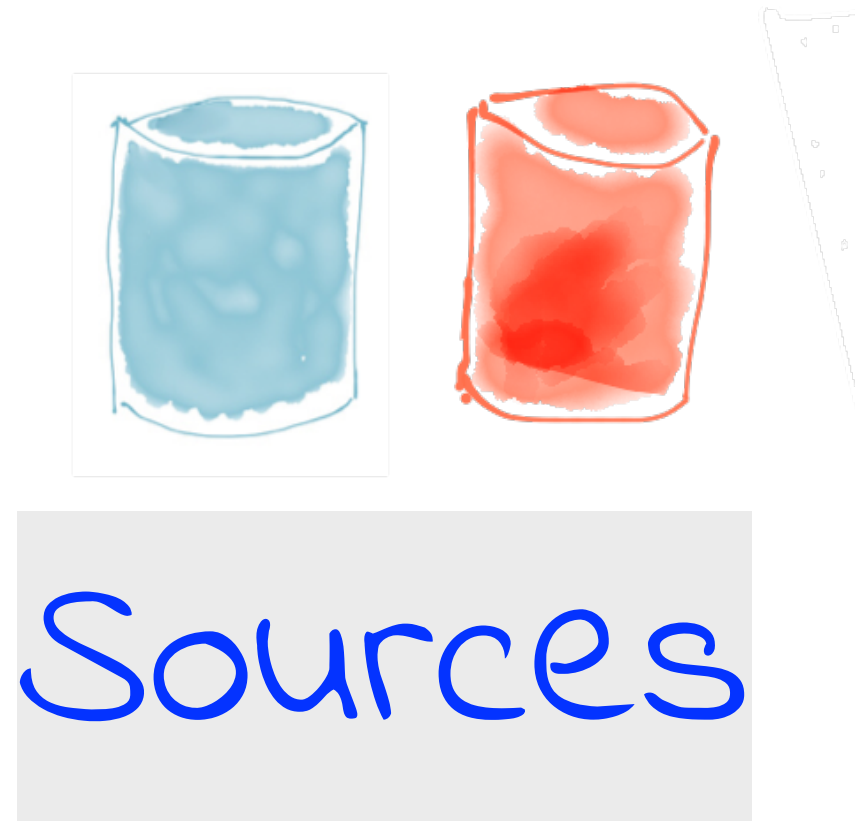
Consumers have a position all of their own



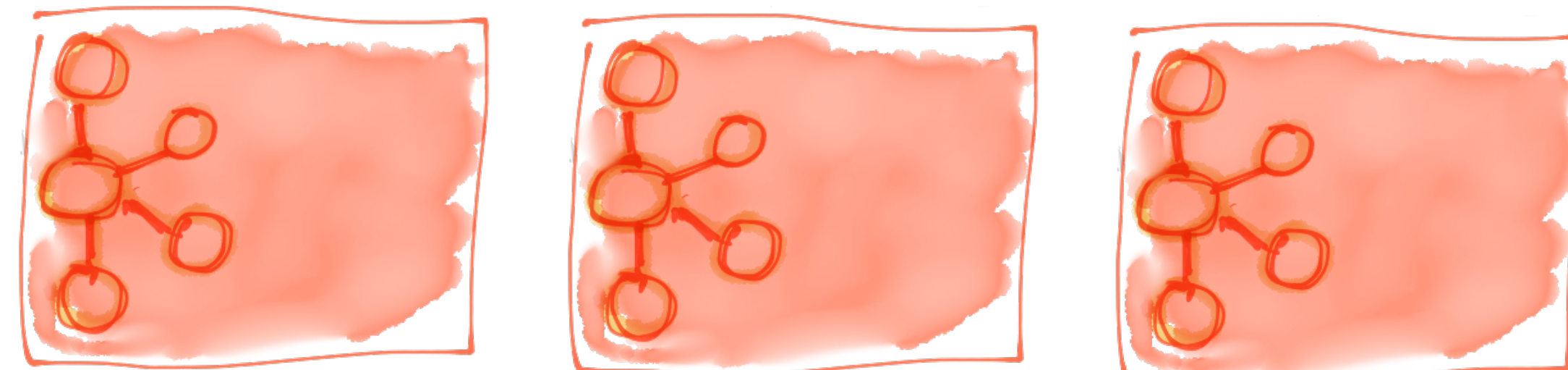
The Connect API



Streaming Integration with Kafka Connect

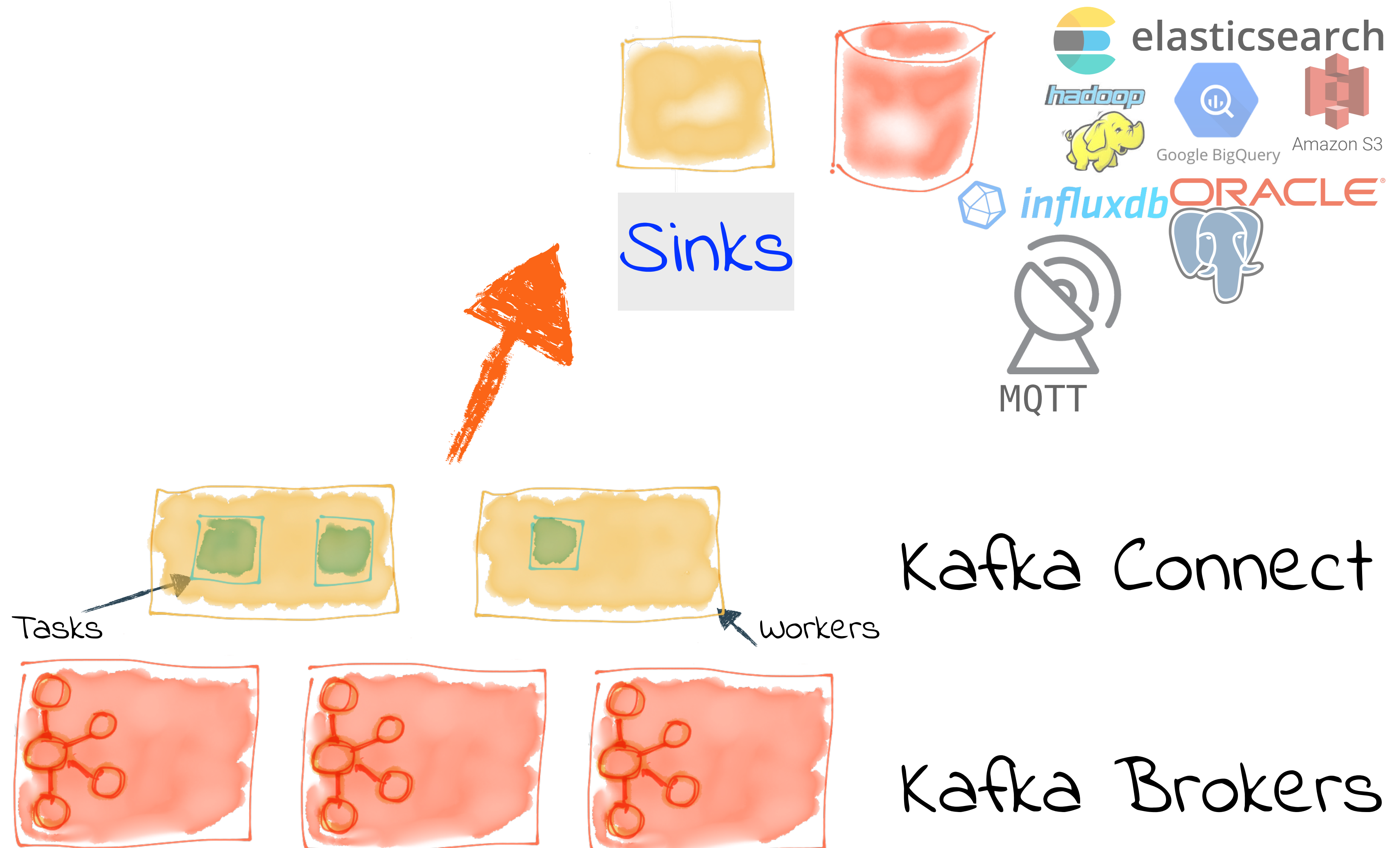


Kafka Connect

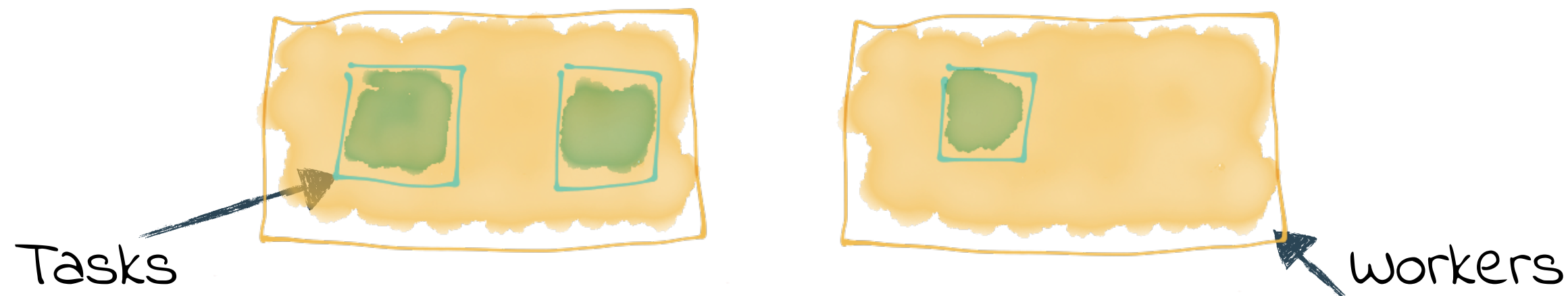
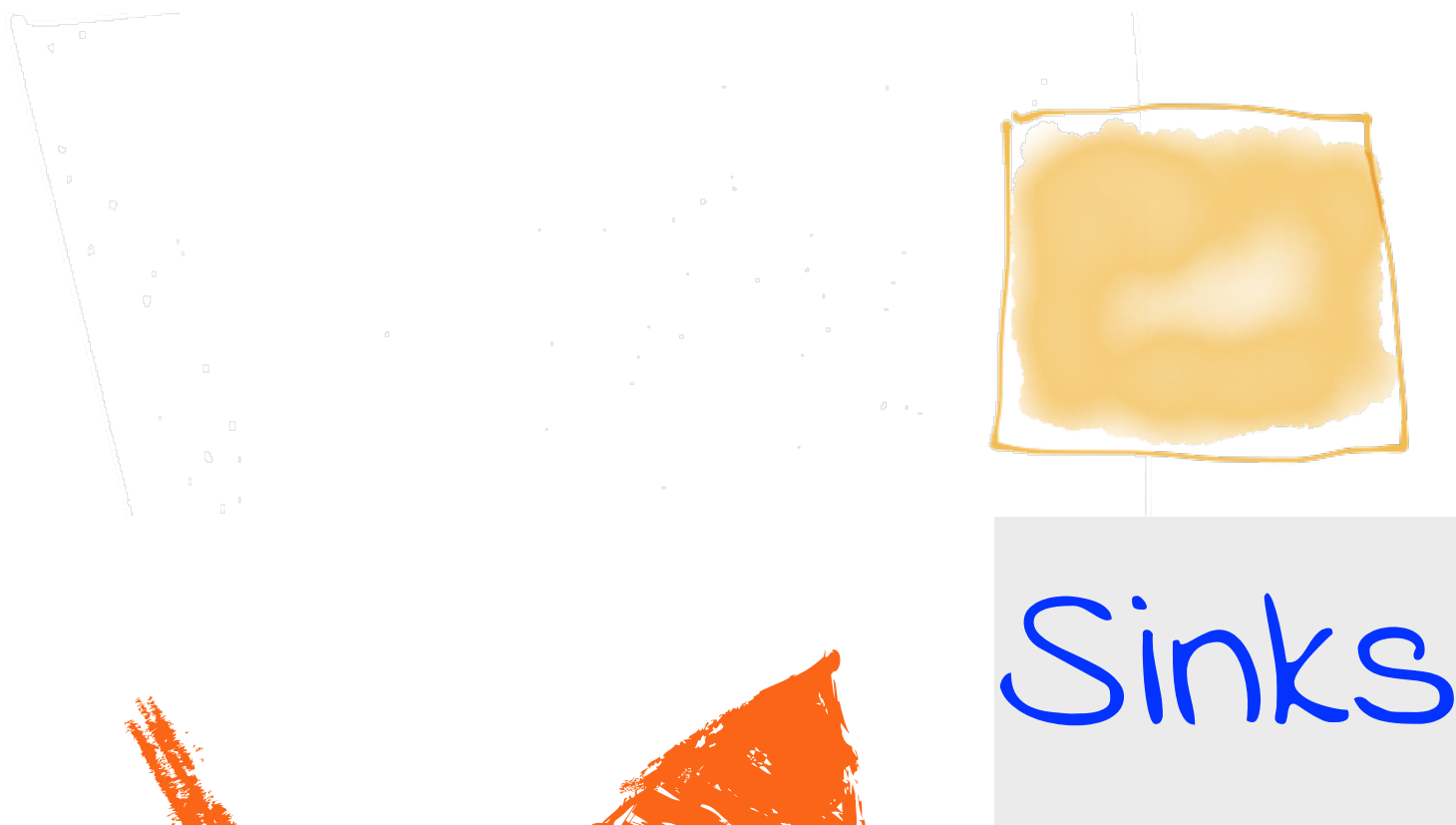
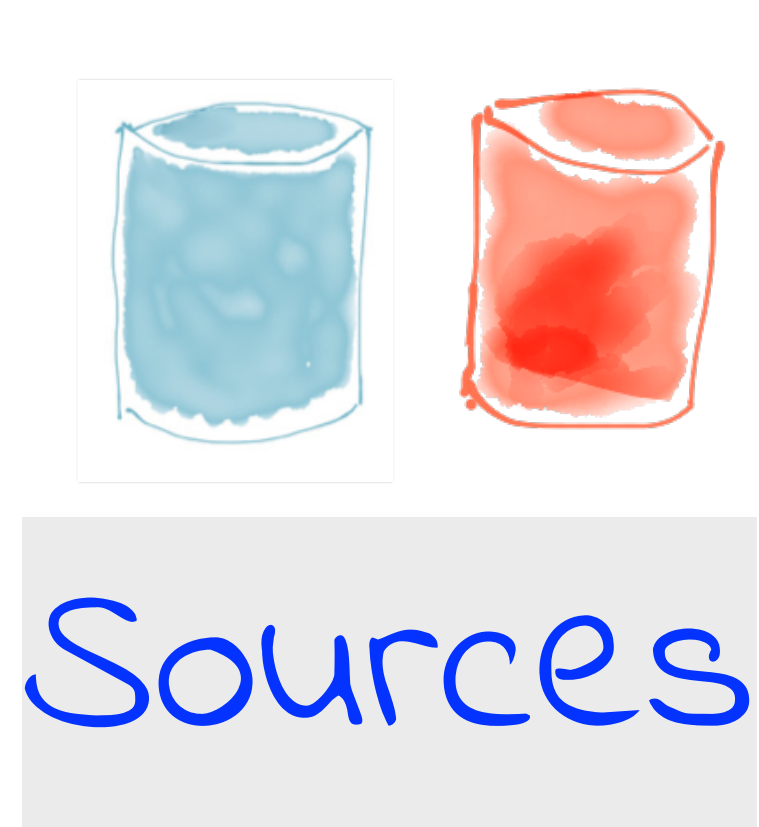


Kafka Brokers

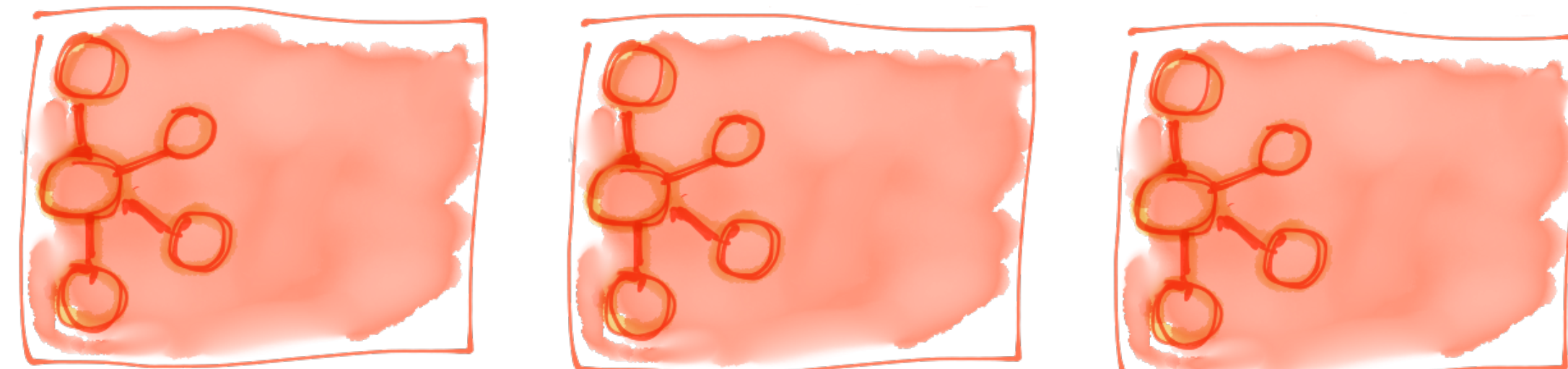
Streaming Integration with Kafka Connect



Streaming Integration with Kafka Connect

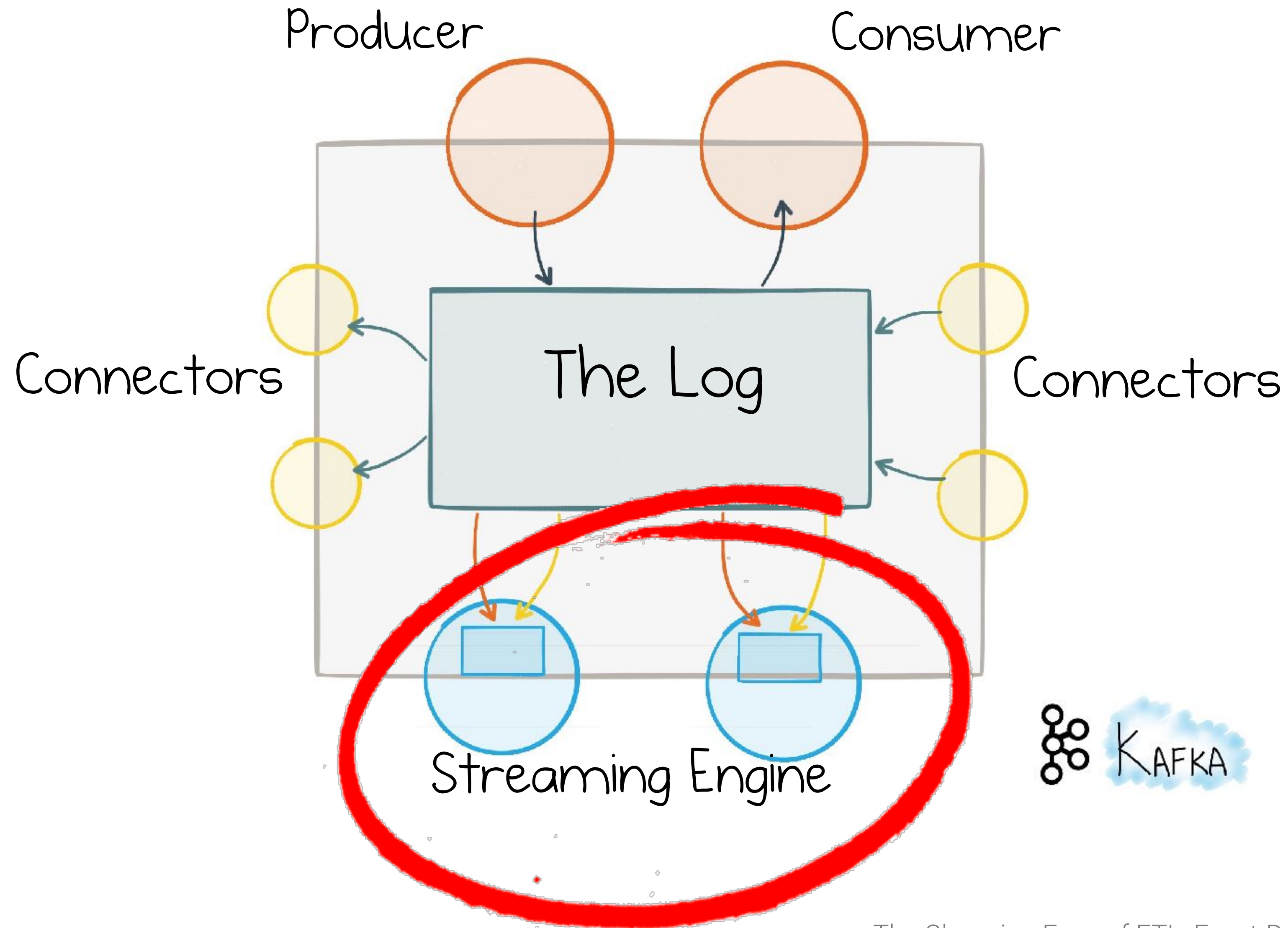


Kafka Connect



Kafka Brokers

Stream Processing in Kafka



Kafka Streams API



```
final StreamsBuilder builder = new StreamsBuilder()  
    .stream("orders", Consumed.with(stringSerde, ordersSerde))  
    .filter( (key, order) -> order.getStatus().equals("COMPLETE") )  
    .to("complete_orders", Produced.with(stringSerde, ordersSerde));
```



Stream Processing with KSQL



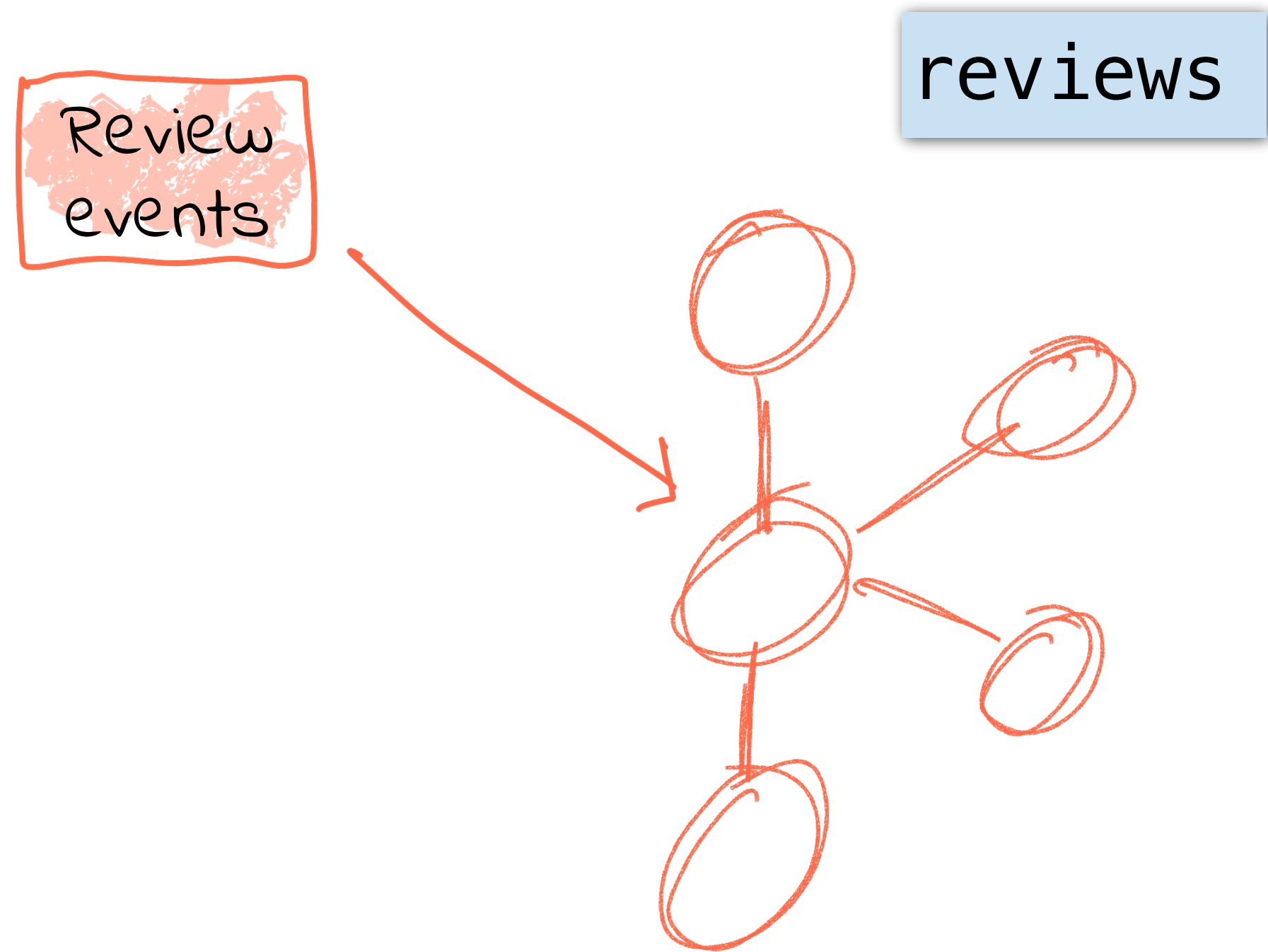
```
CREATE STREAM completedOrders AS  
SELECT *  
FROM orders  
WHERE status='COMPLETE' ;
```



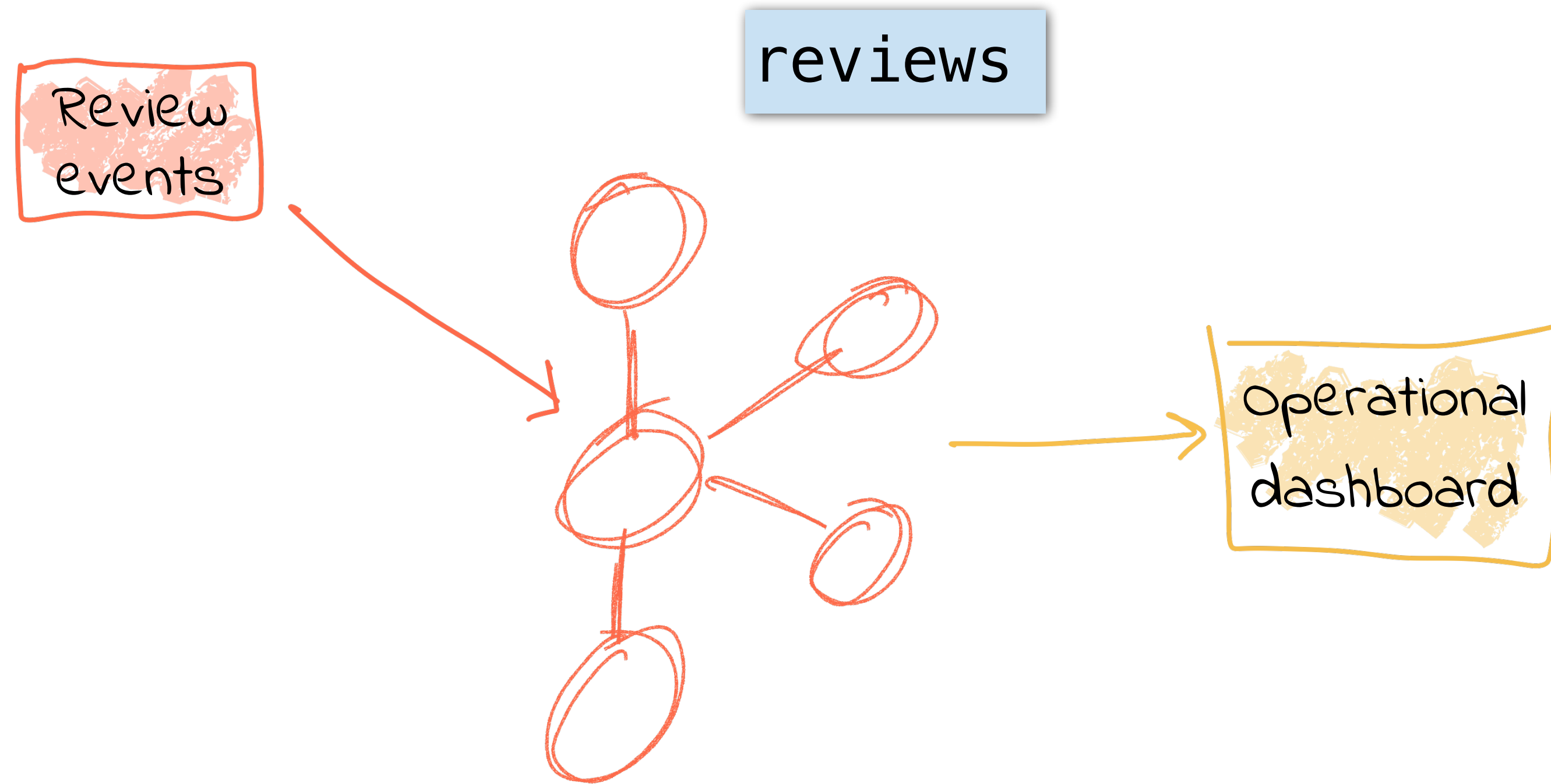


**This is
Something
New**

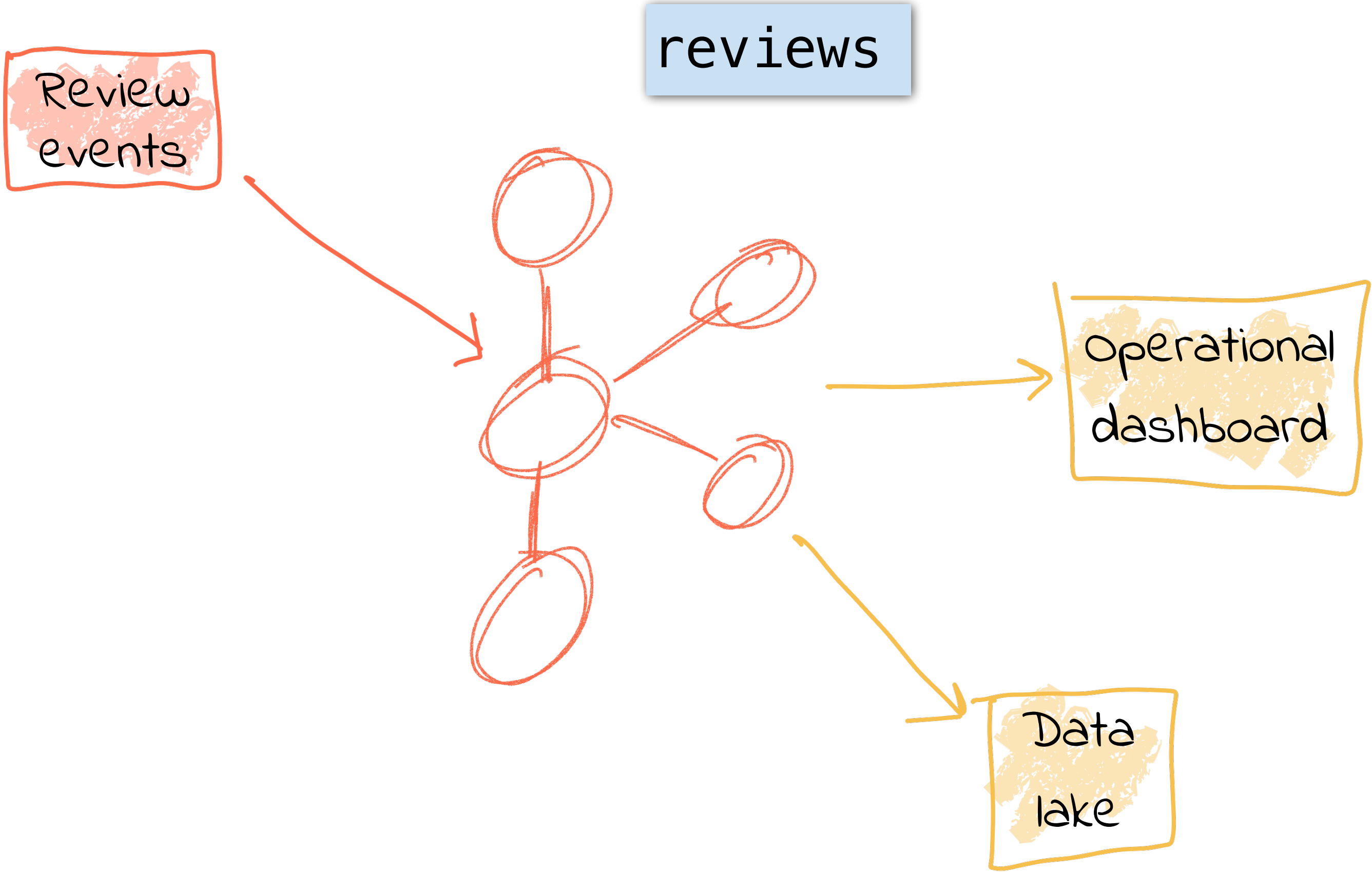
Events in Action



Events in Action



Events in Action



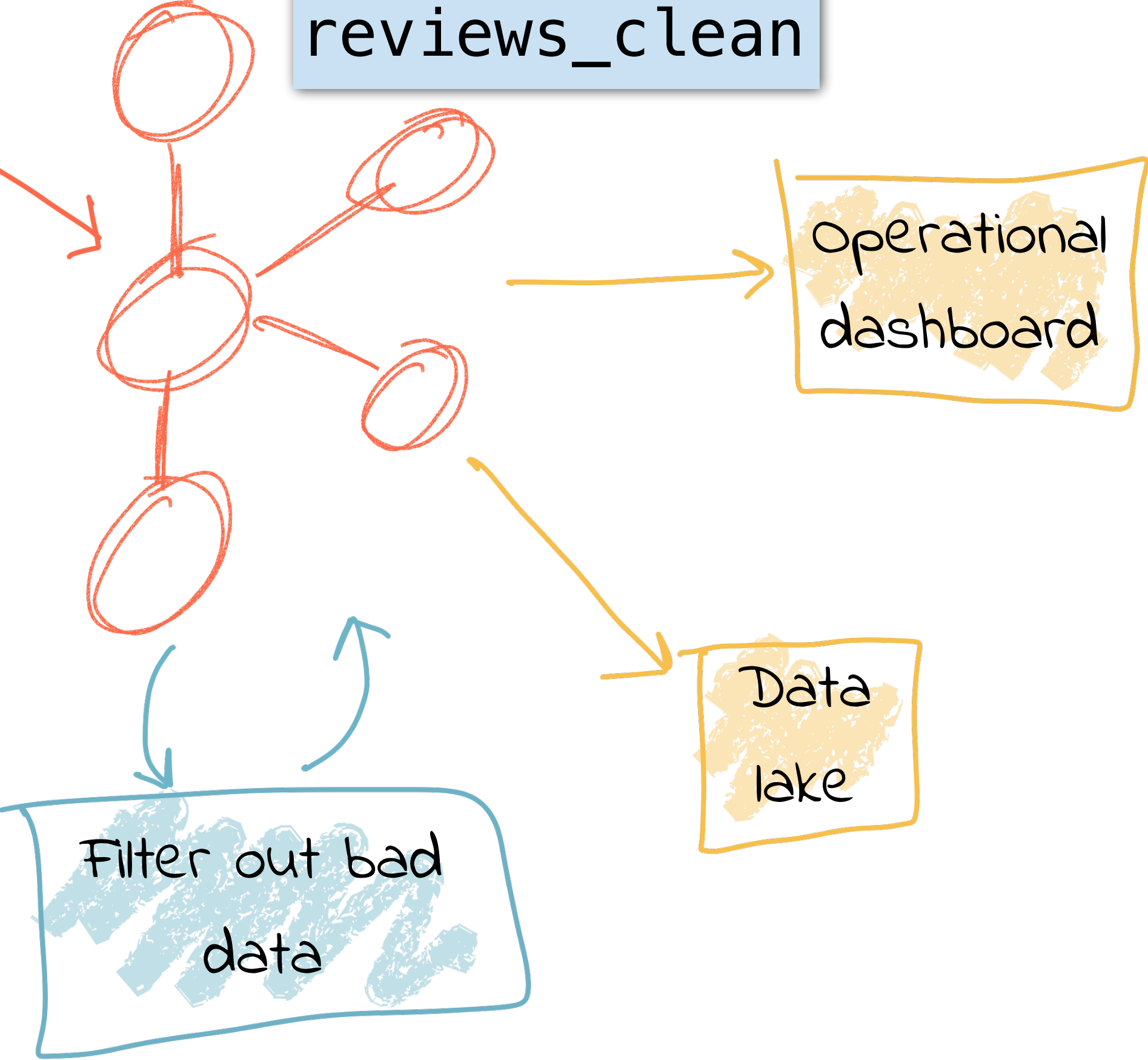
Events in Action

```
CREATE STREAM reviews_clean AS  
SELECT * FROM reviews  
WHERE id IS NOT NULL;
```

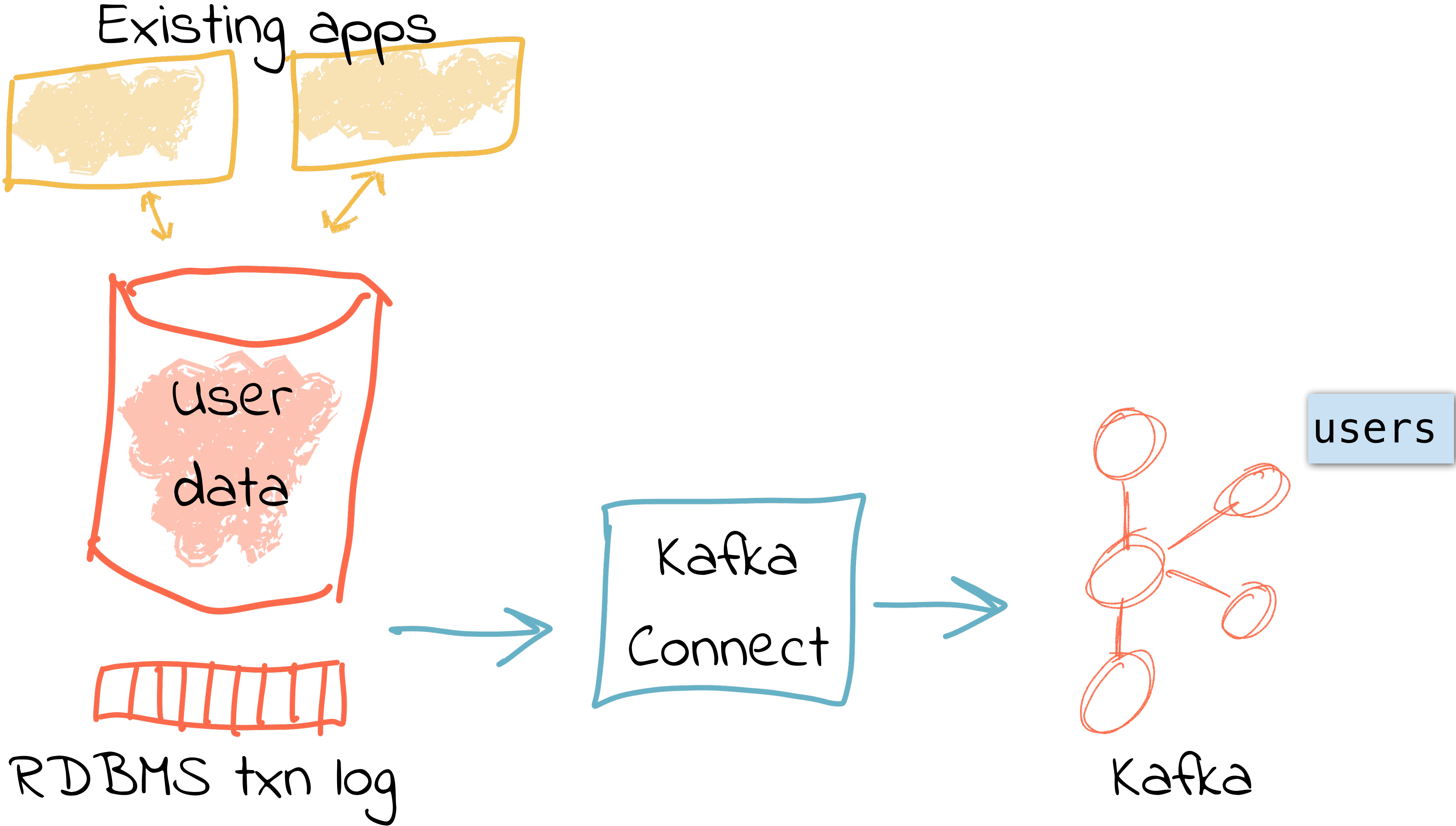
Review events

reviews

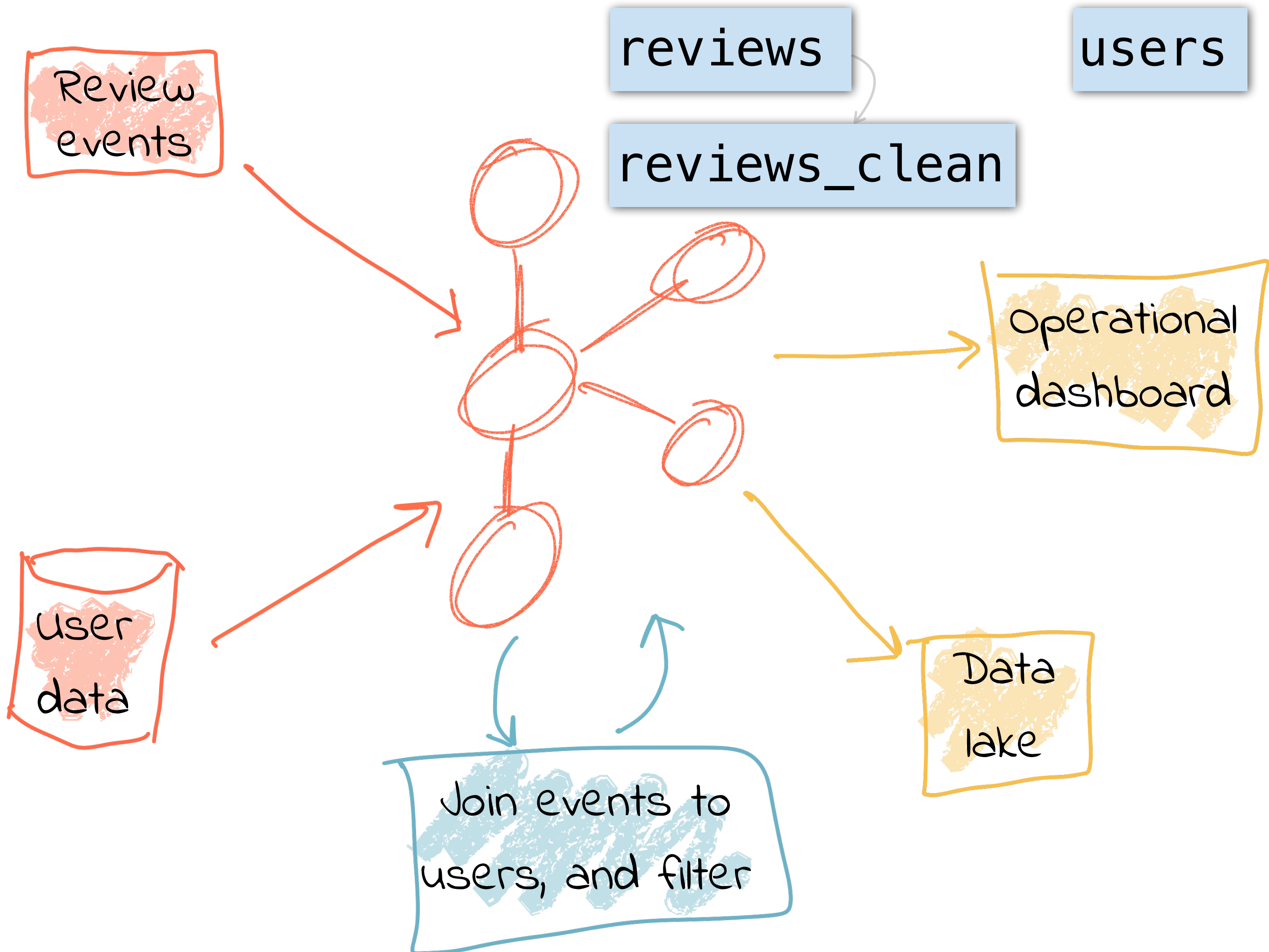
reviews_clean



Events in Action

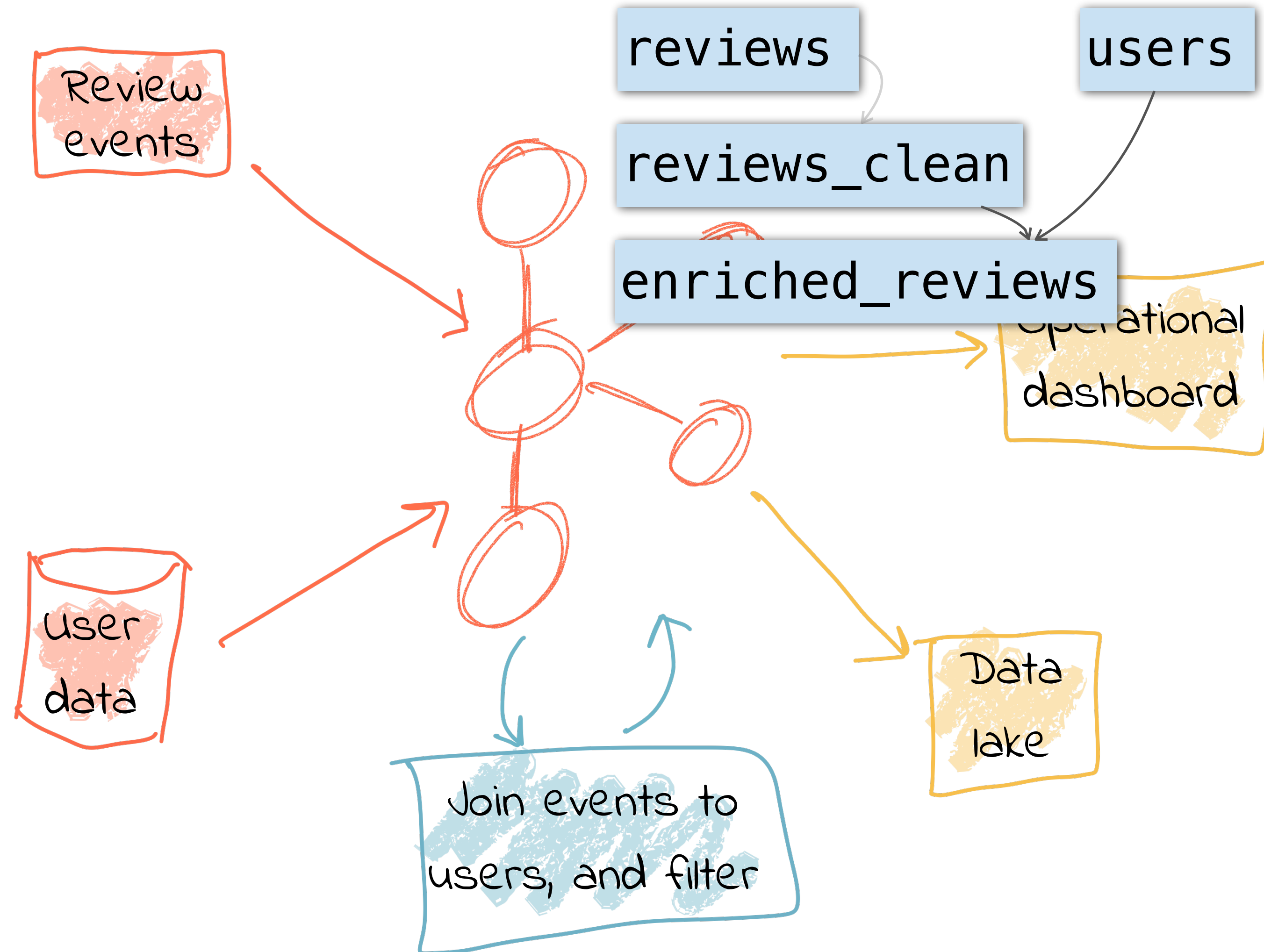


Events in Action

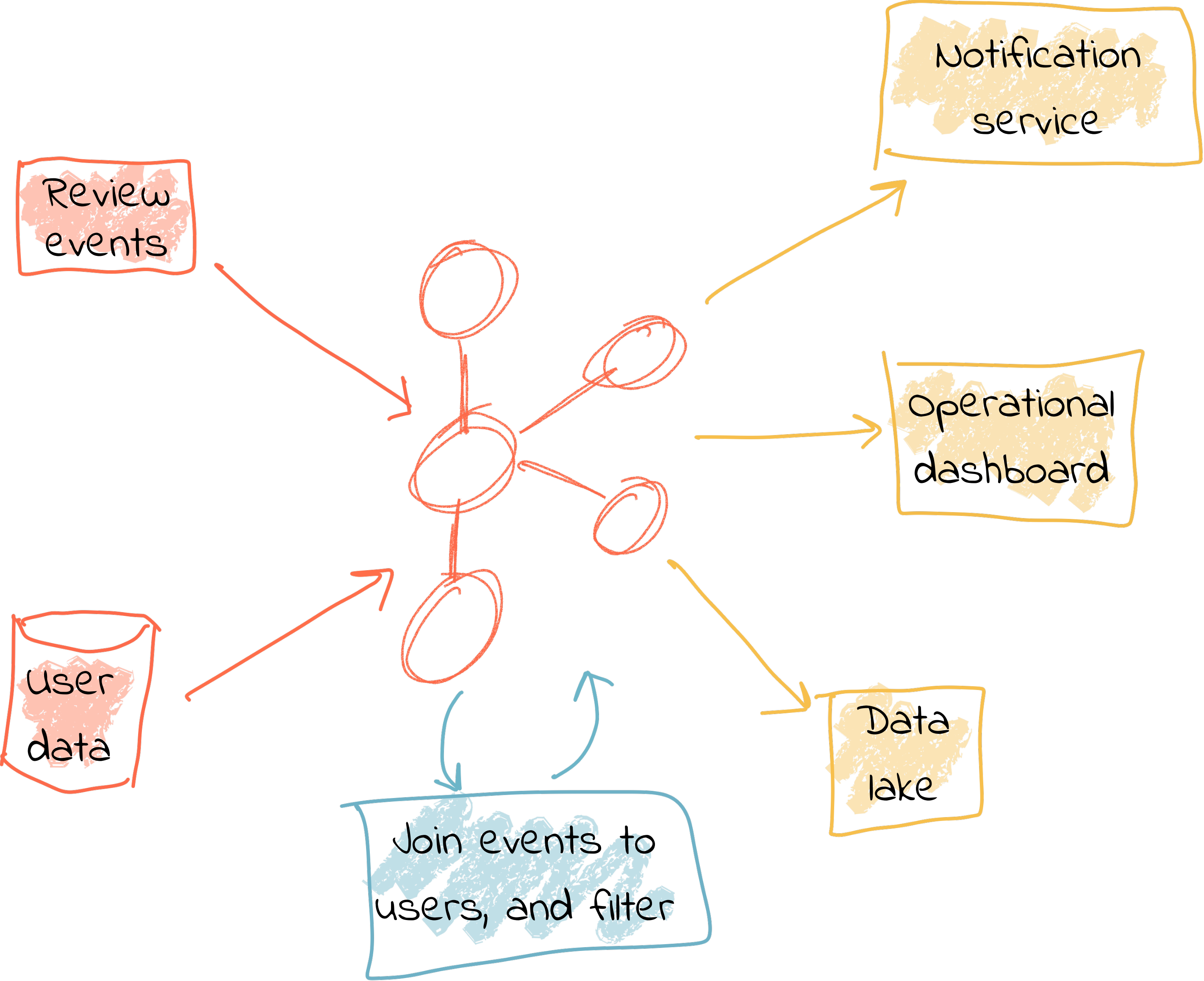


Events in Action

```
CREATE STREAM enriched_reviews AS  
SELECT * FROM reviews_clean r  
INNER JOIN users u  
ON r.userid=u.userid;
```

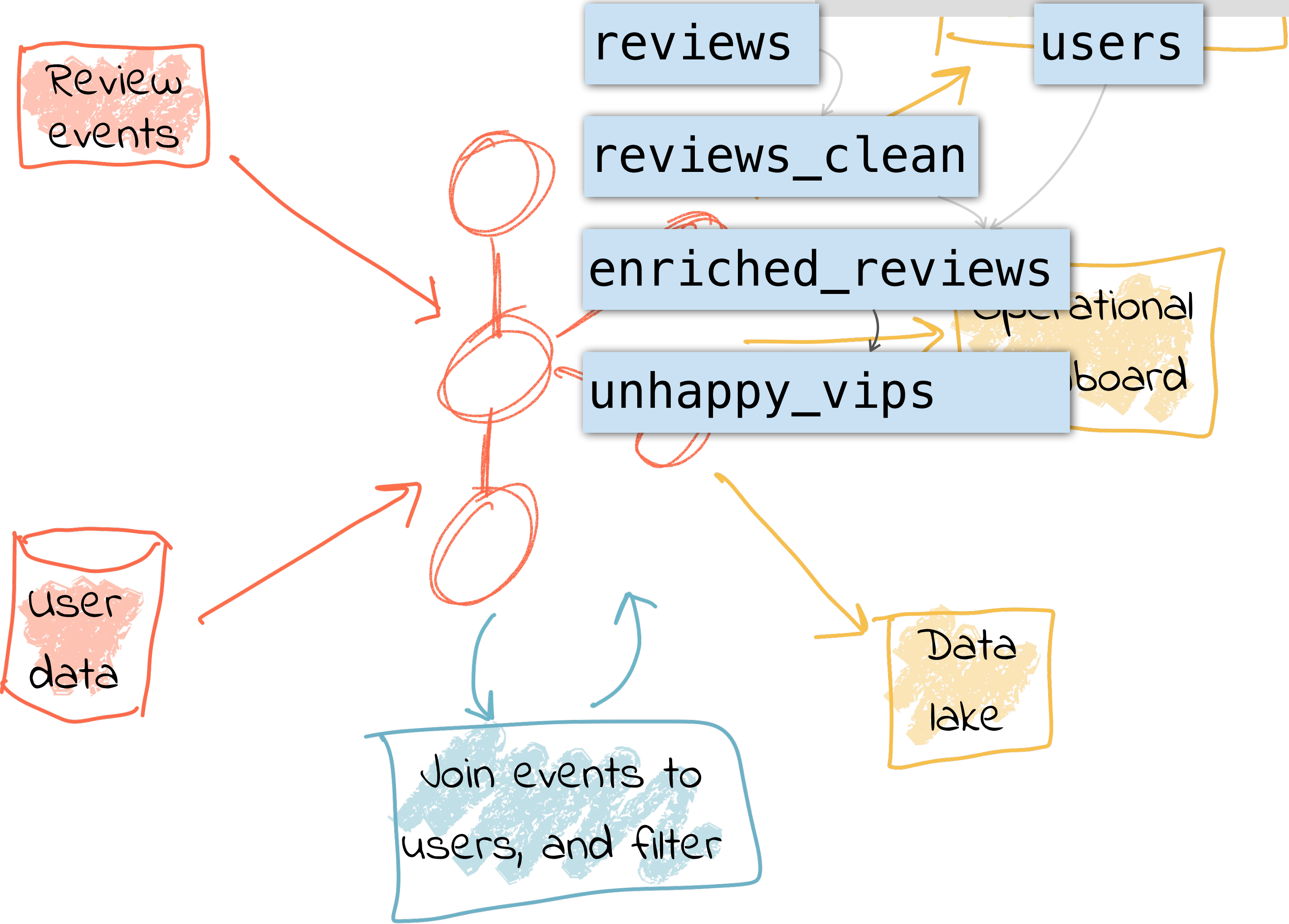


Events in Action



Events in Action

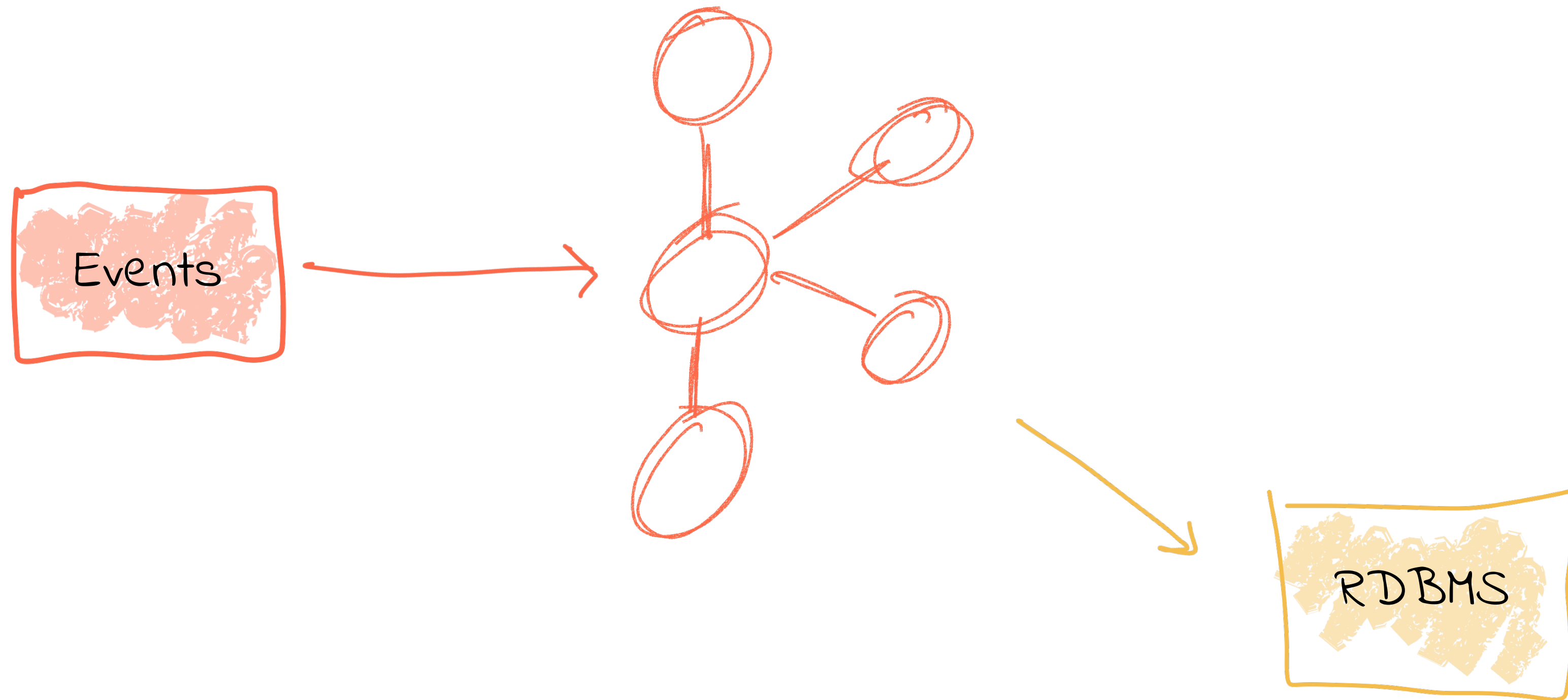
```
CREATE STREAM unhappy_vips AS  
SELECT * FROM enriched_reviews  
WHERE rating < 3  
AND status = 'Platinum';
```



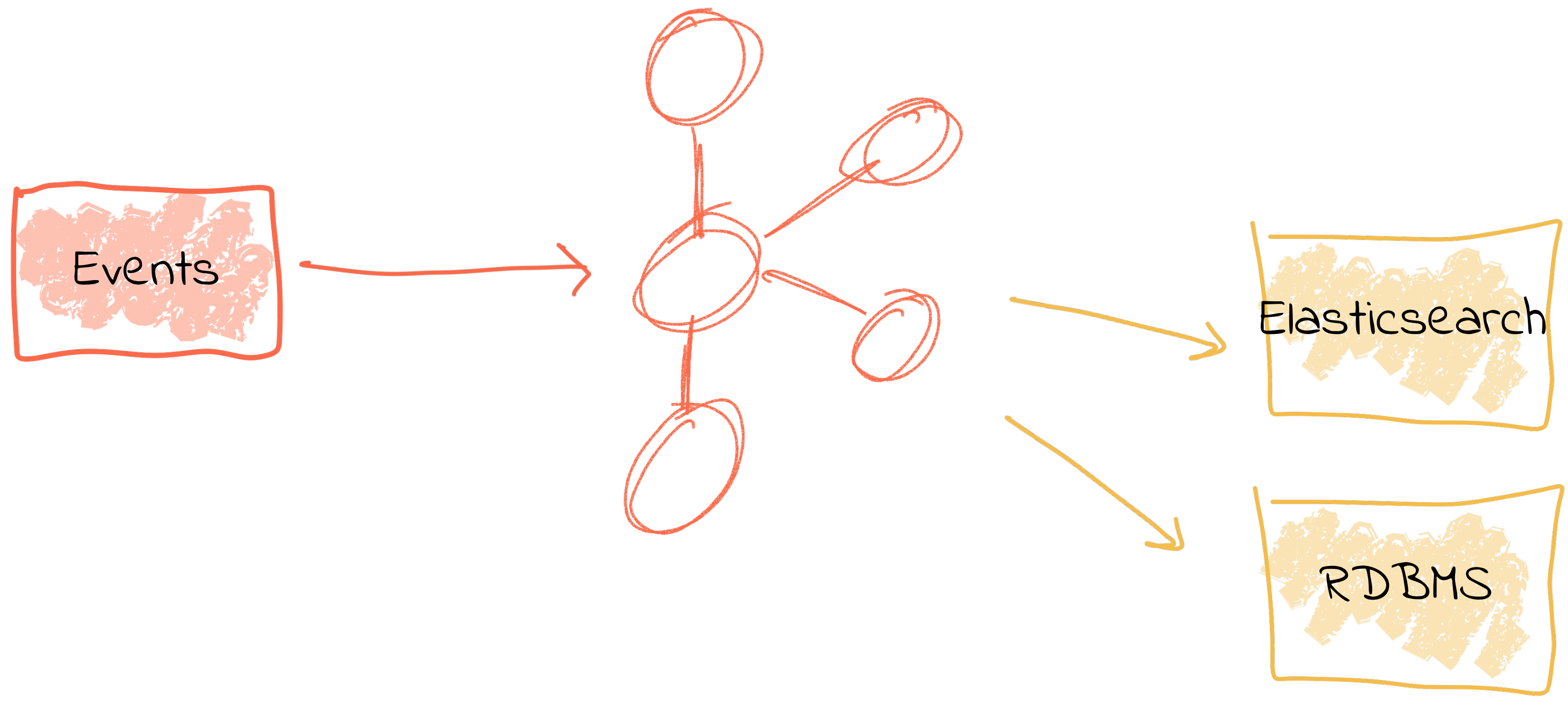
The Power of an Event-Driven Architecture



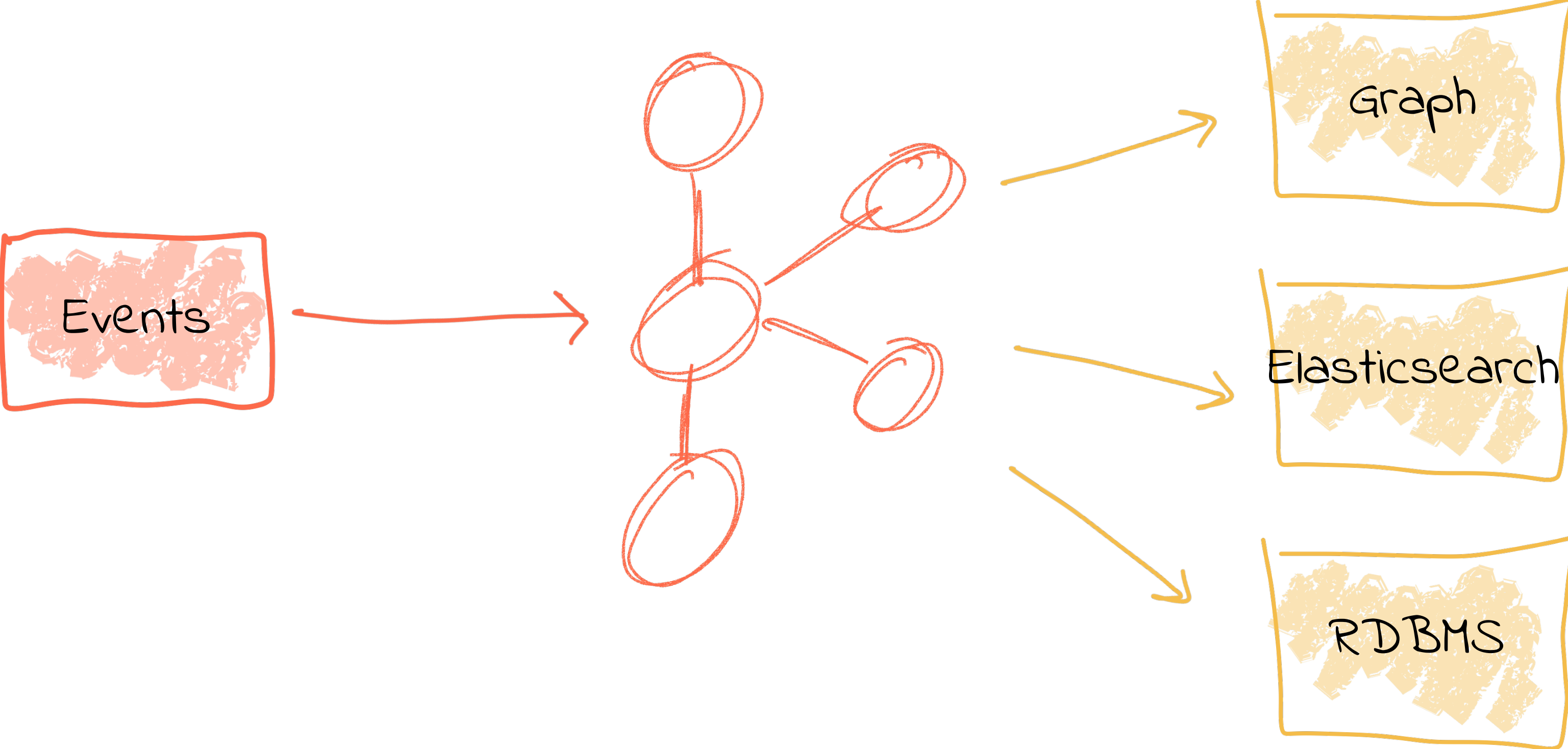
Not Everything is a Nail



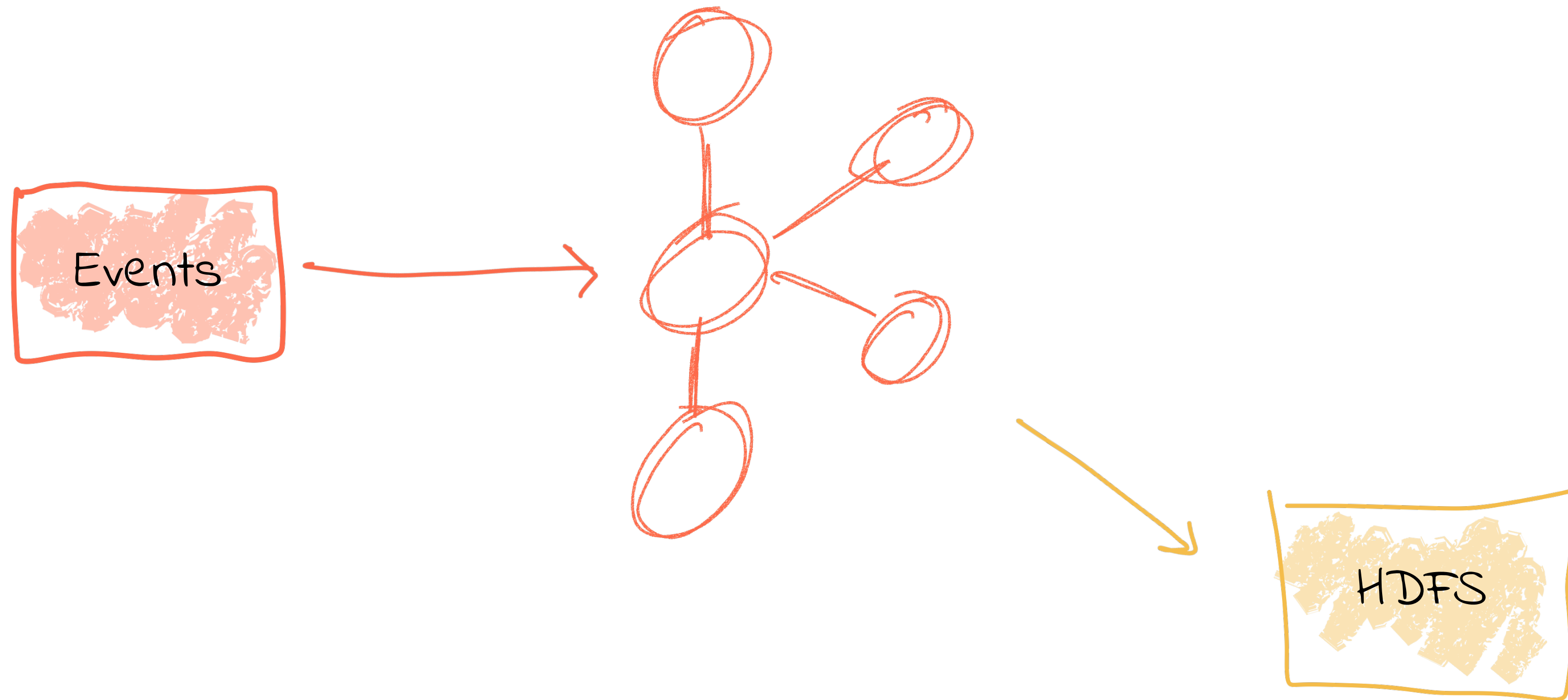
Not Everything is a Nail



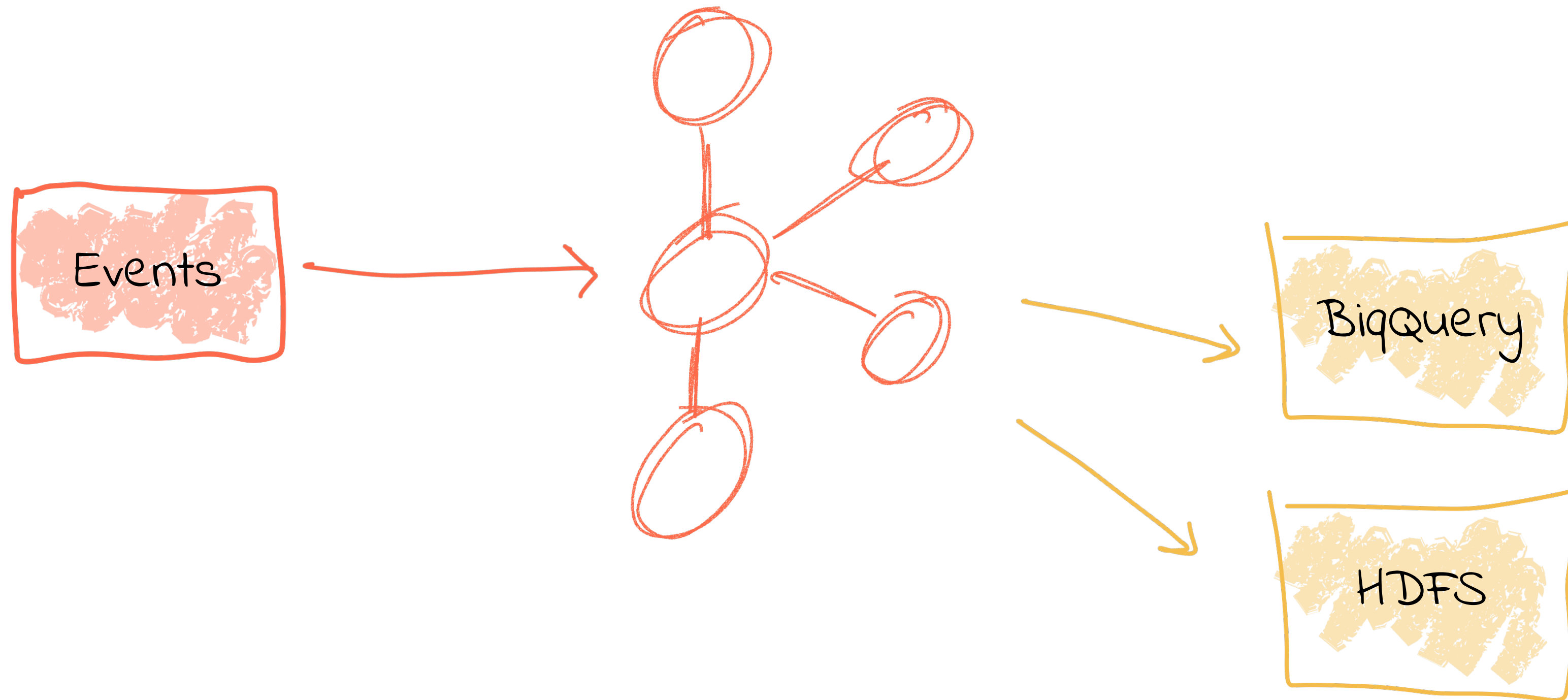
Not Everything is a Nail



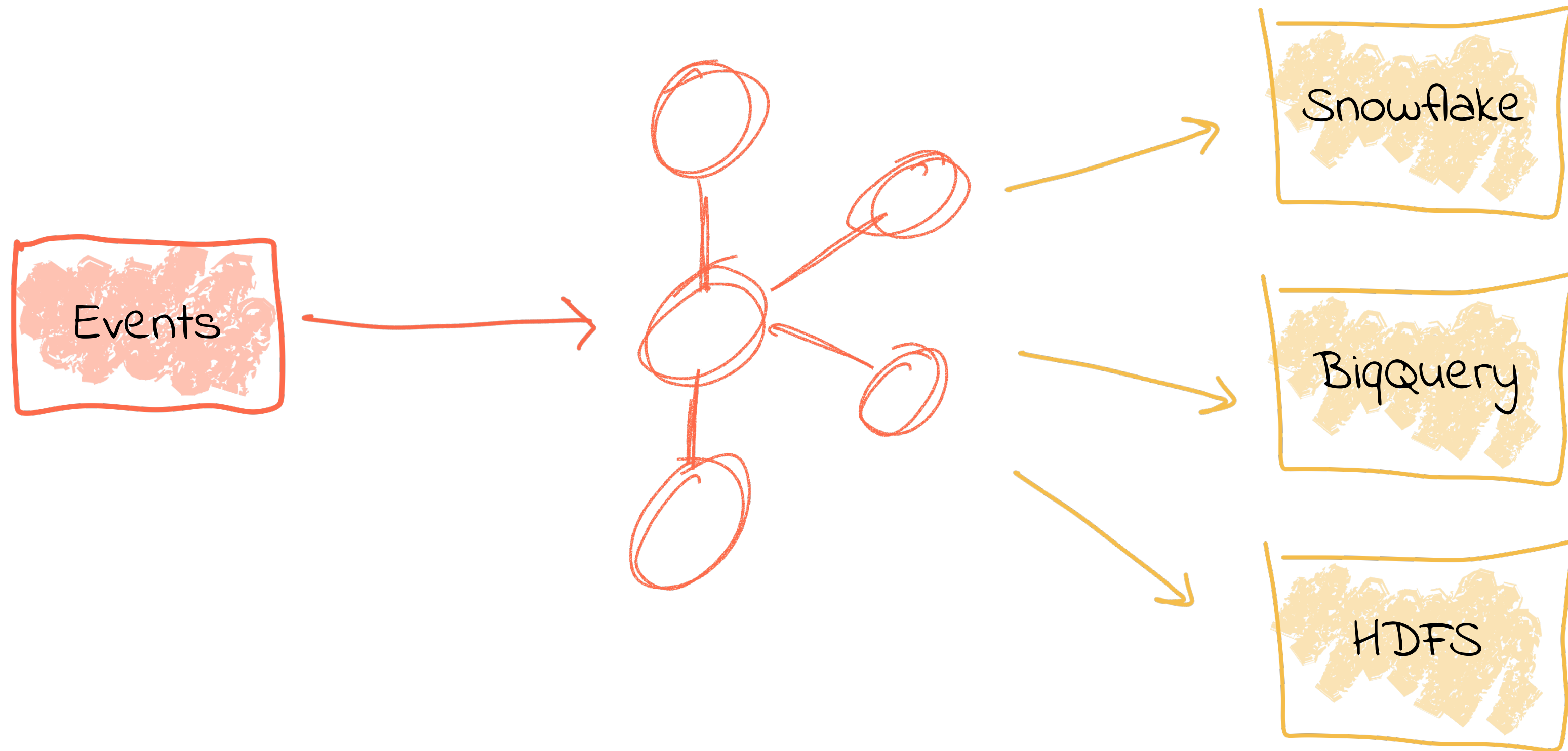
Side-by-Side Tech Evaluation



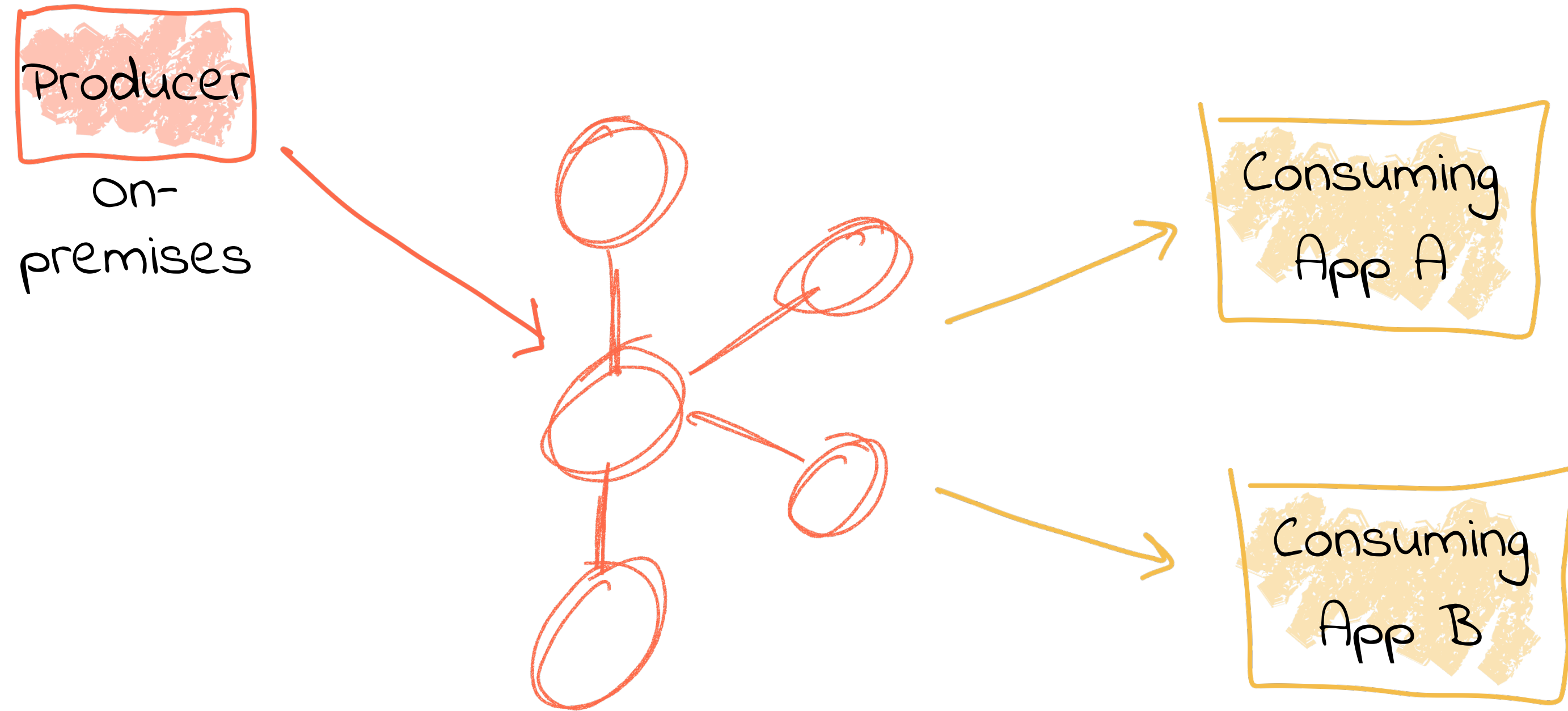
Side-by-Side Tech Evaluation



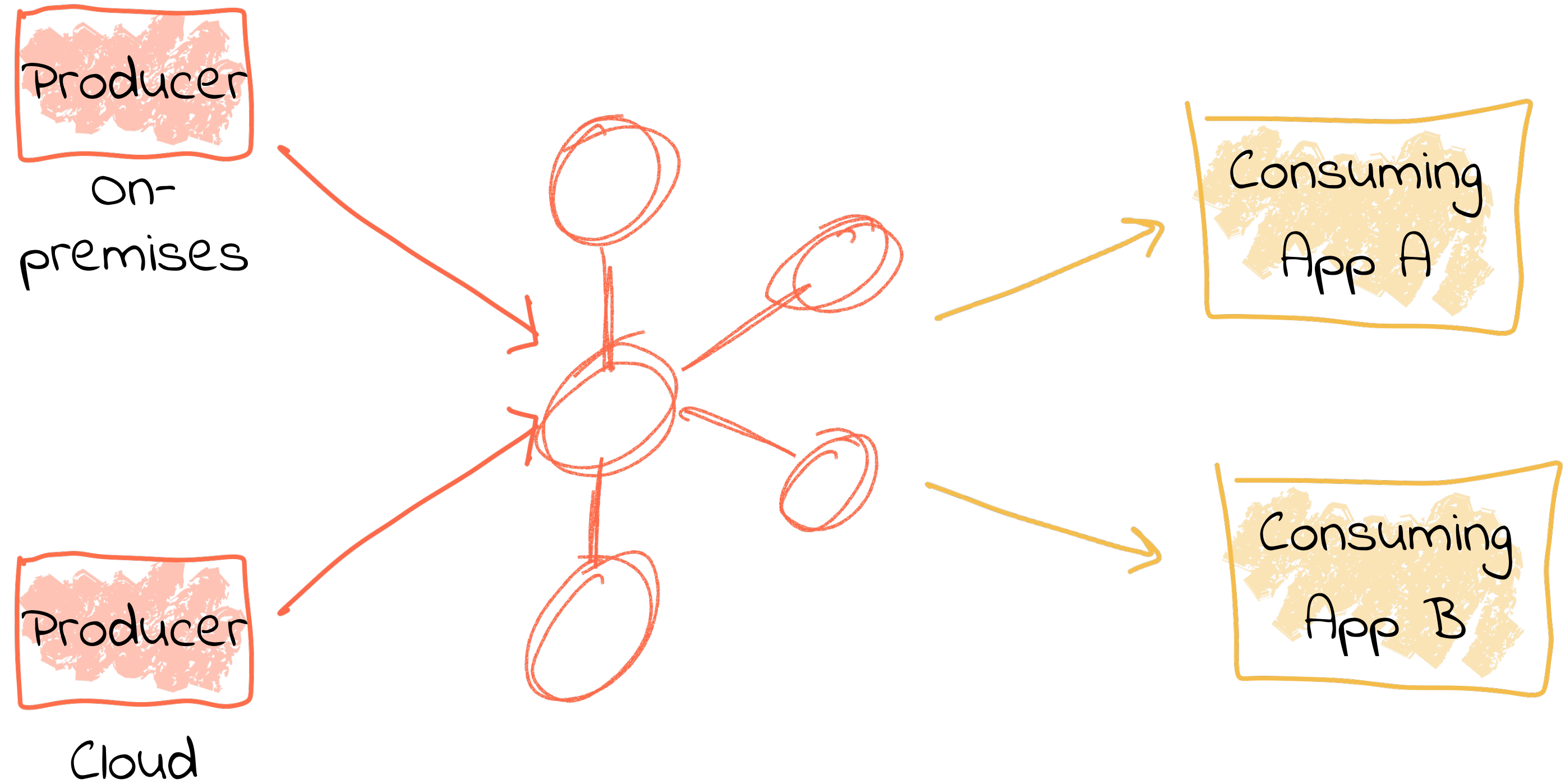
Side-by-Side Tech Evaluation



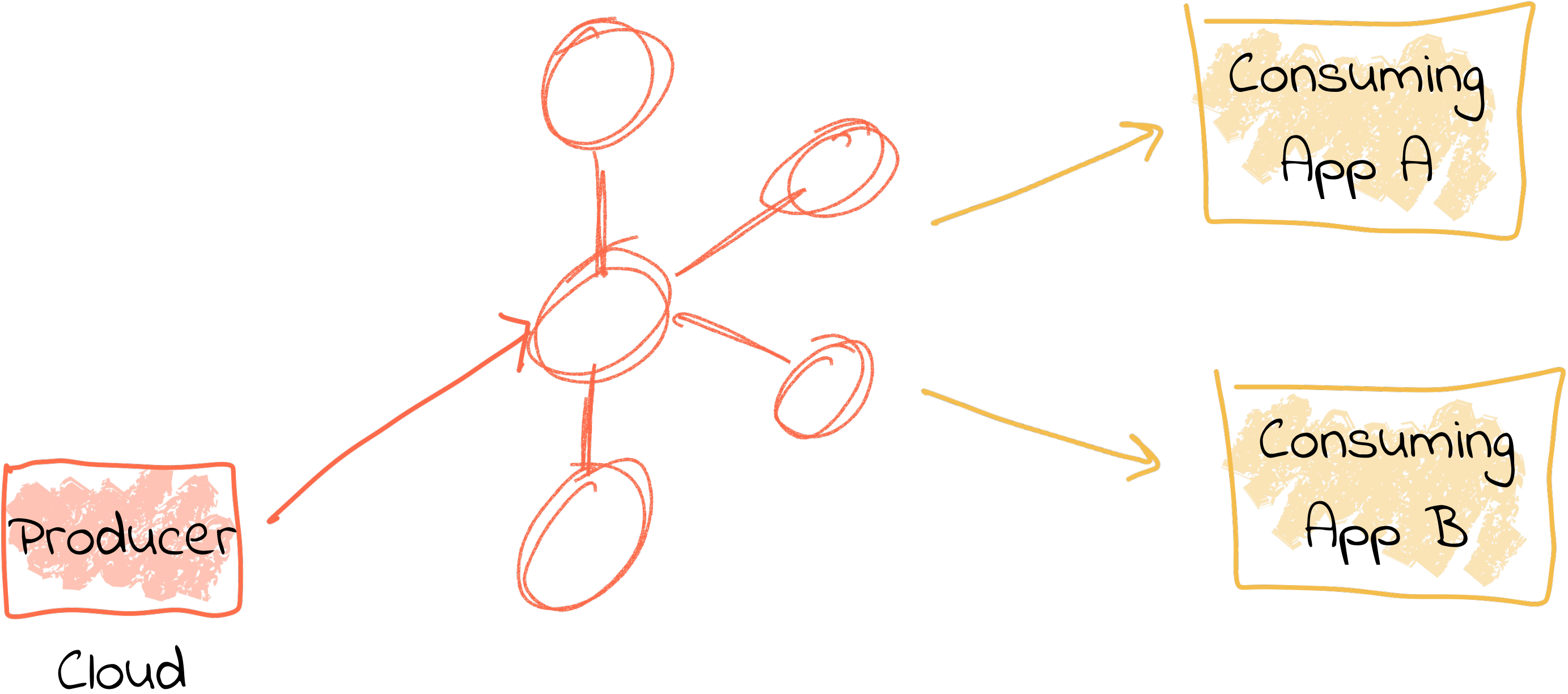
Evolve Data Sources



Evolve Data Sources



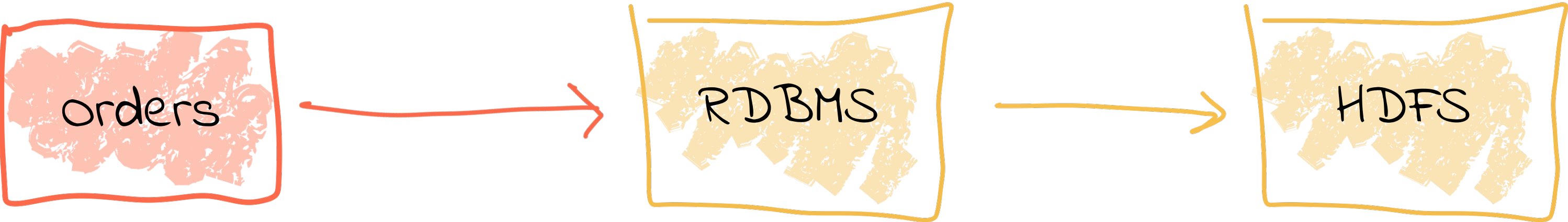
Evolve Data Sources



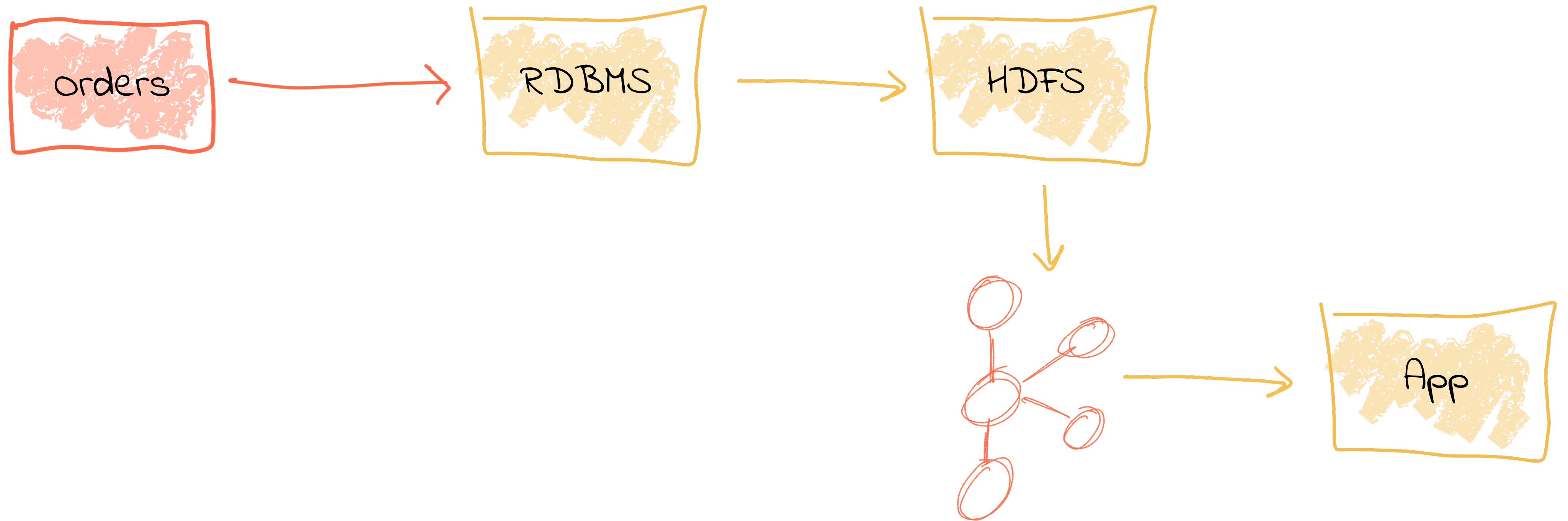
Tight Coupling != Flexible



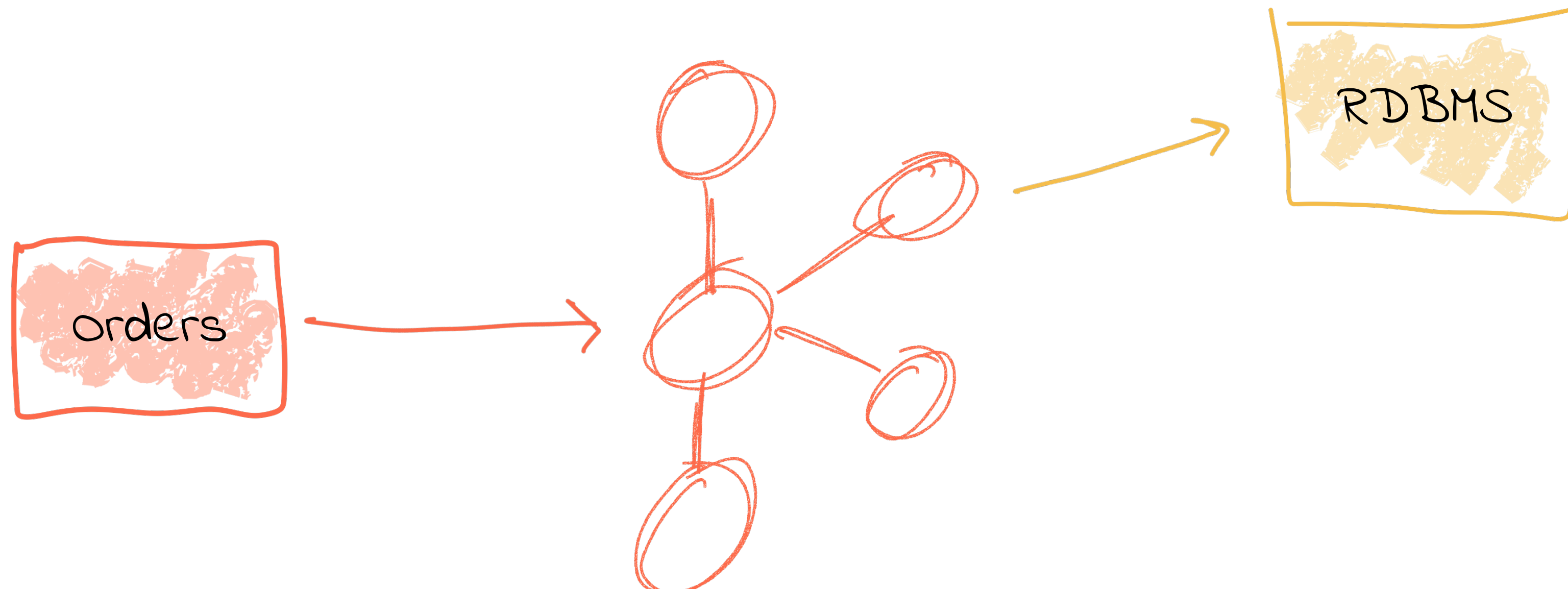
Tight Coupling != Flexible



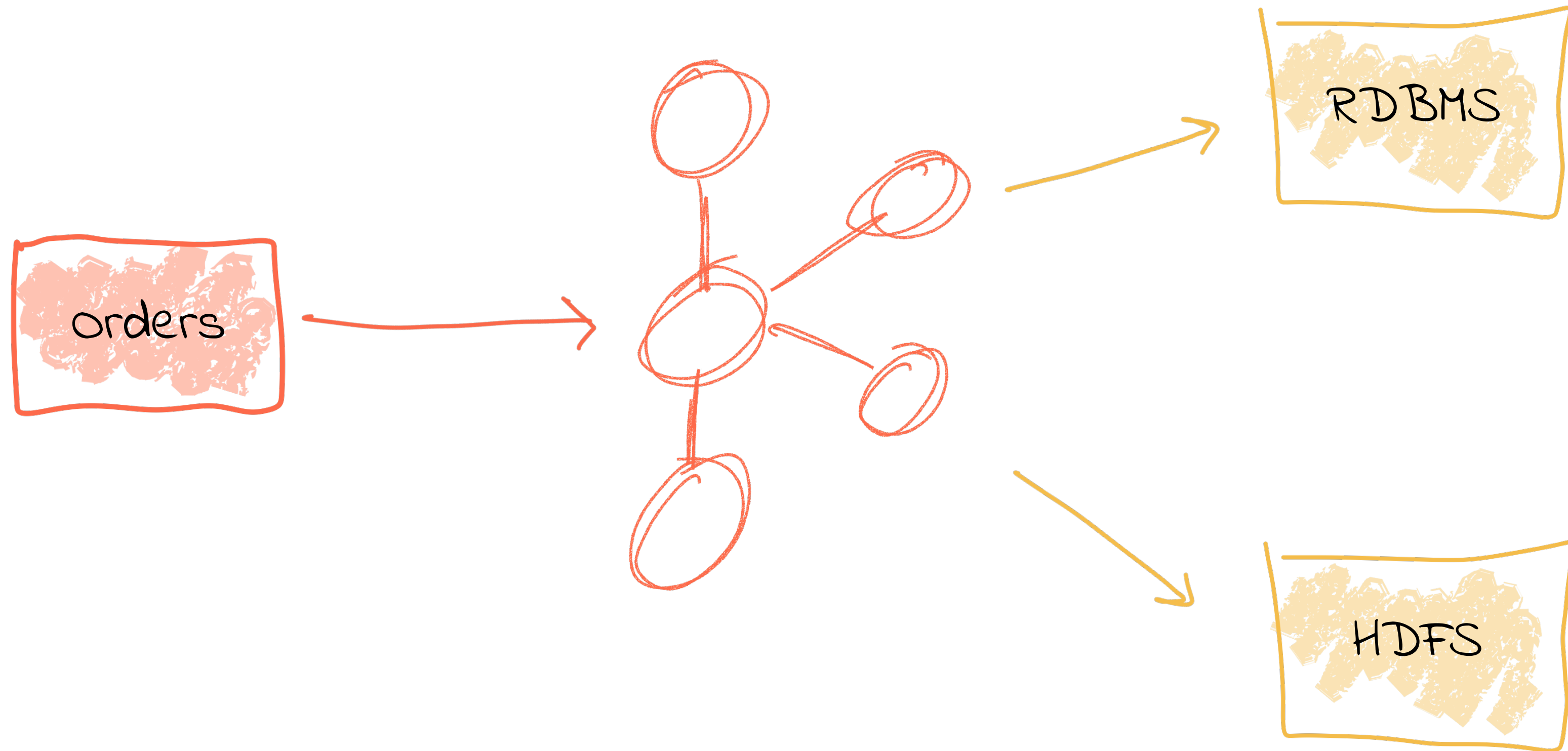
Tight Coupling != Flexible



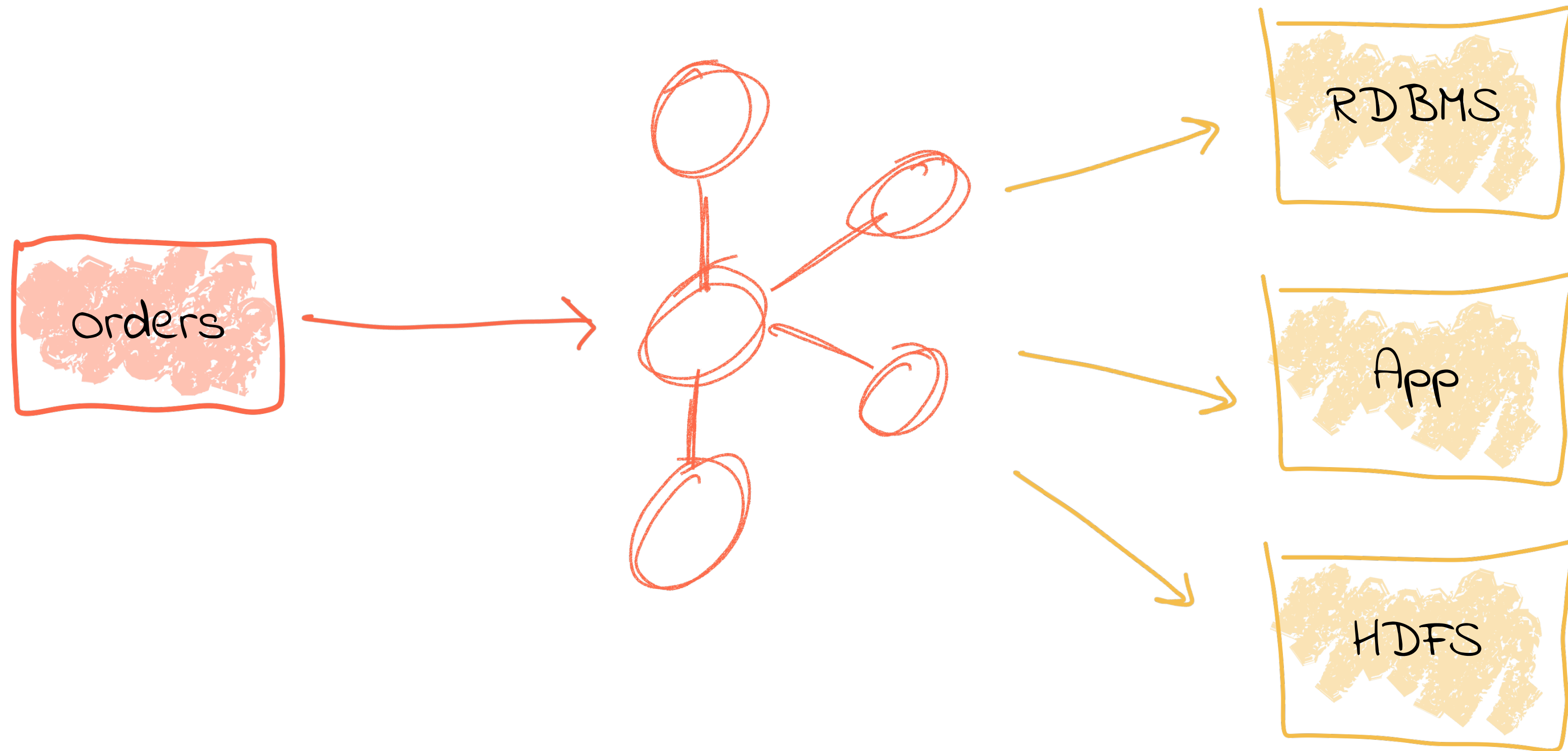
Loose Coupling == Freedom to Evolve



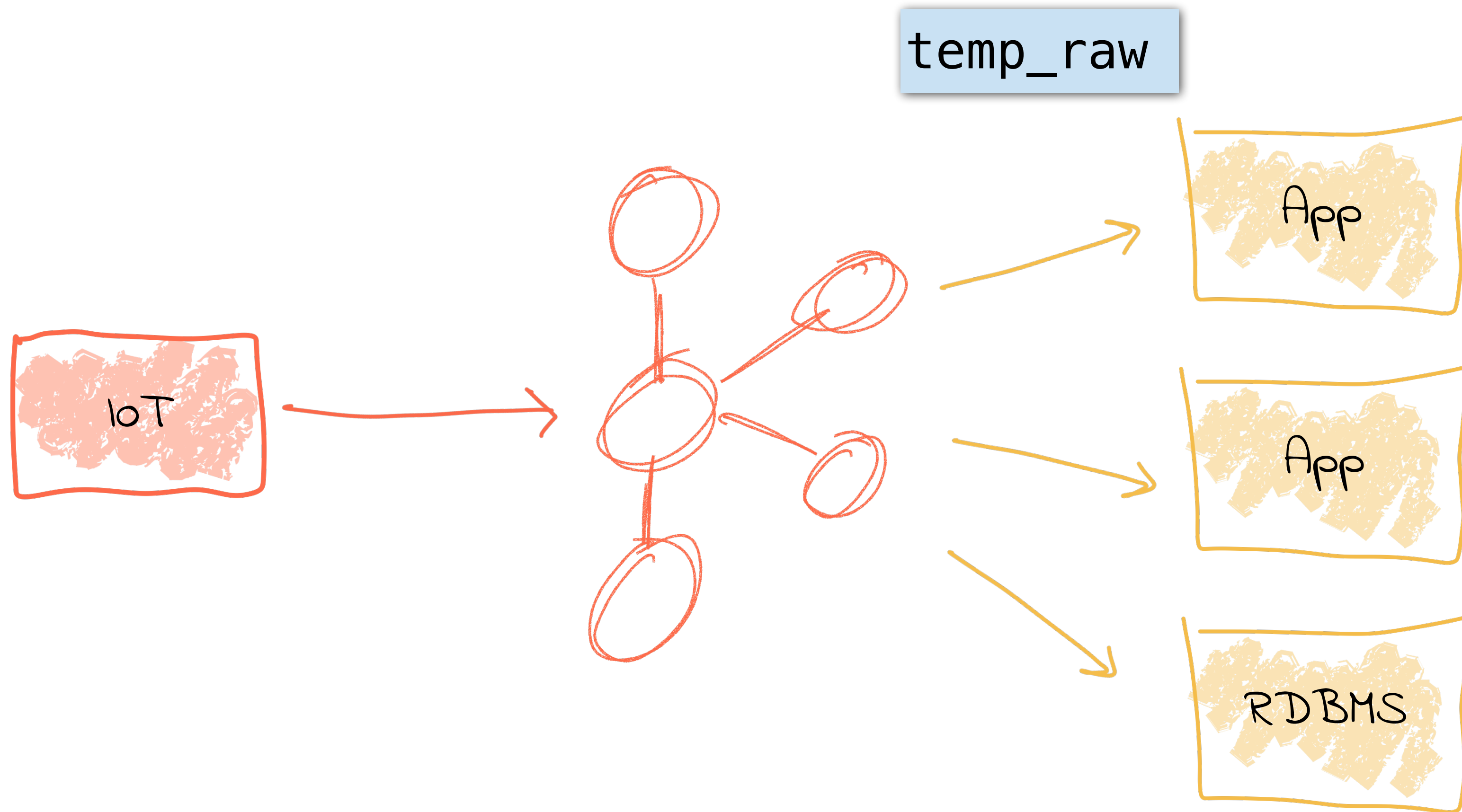
Loose Coupling == Freedom to Evolve



Loose Coupling == Freedom to Evolve

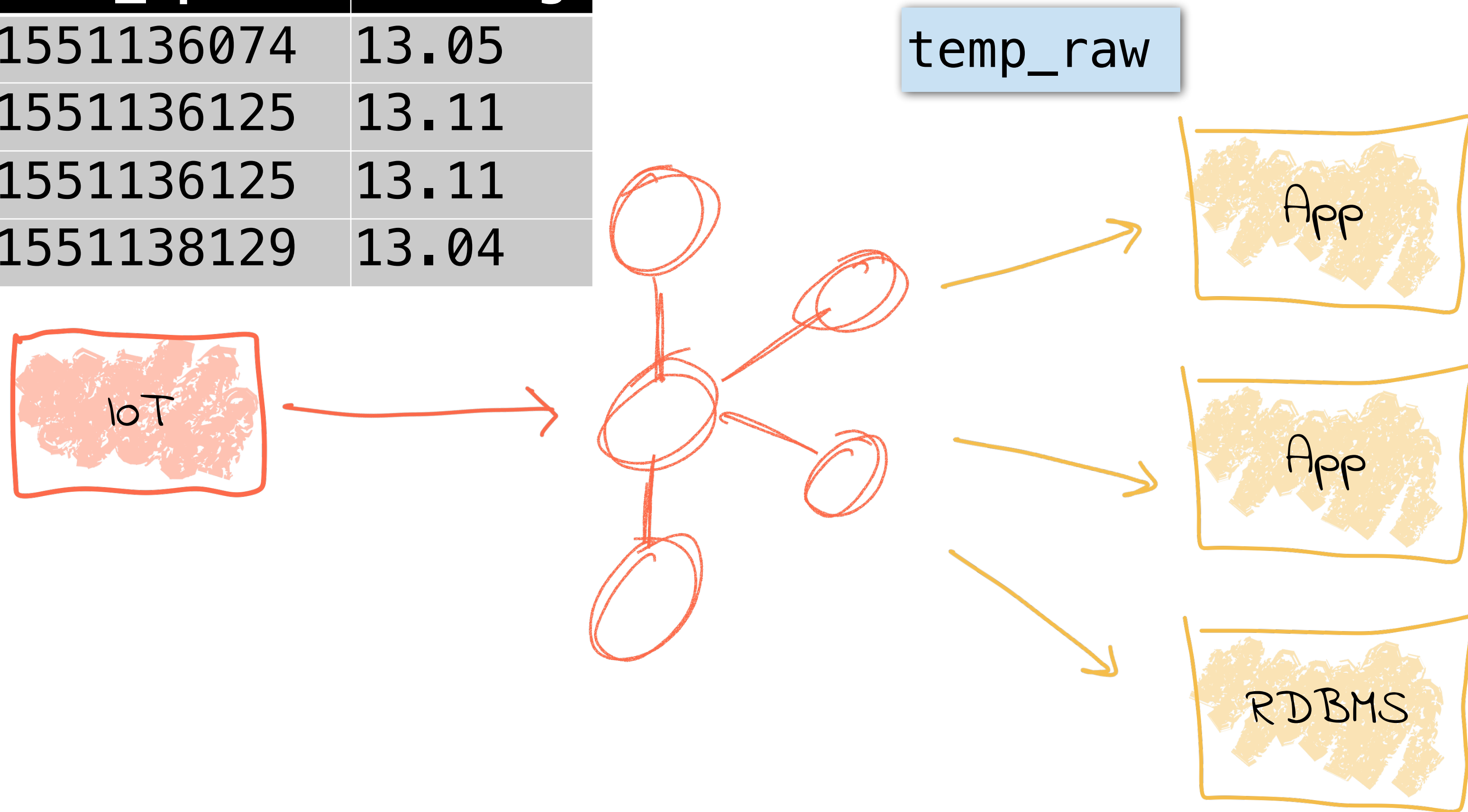


Transform Once, Use Many: Data Cleansing



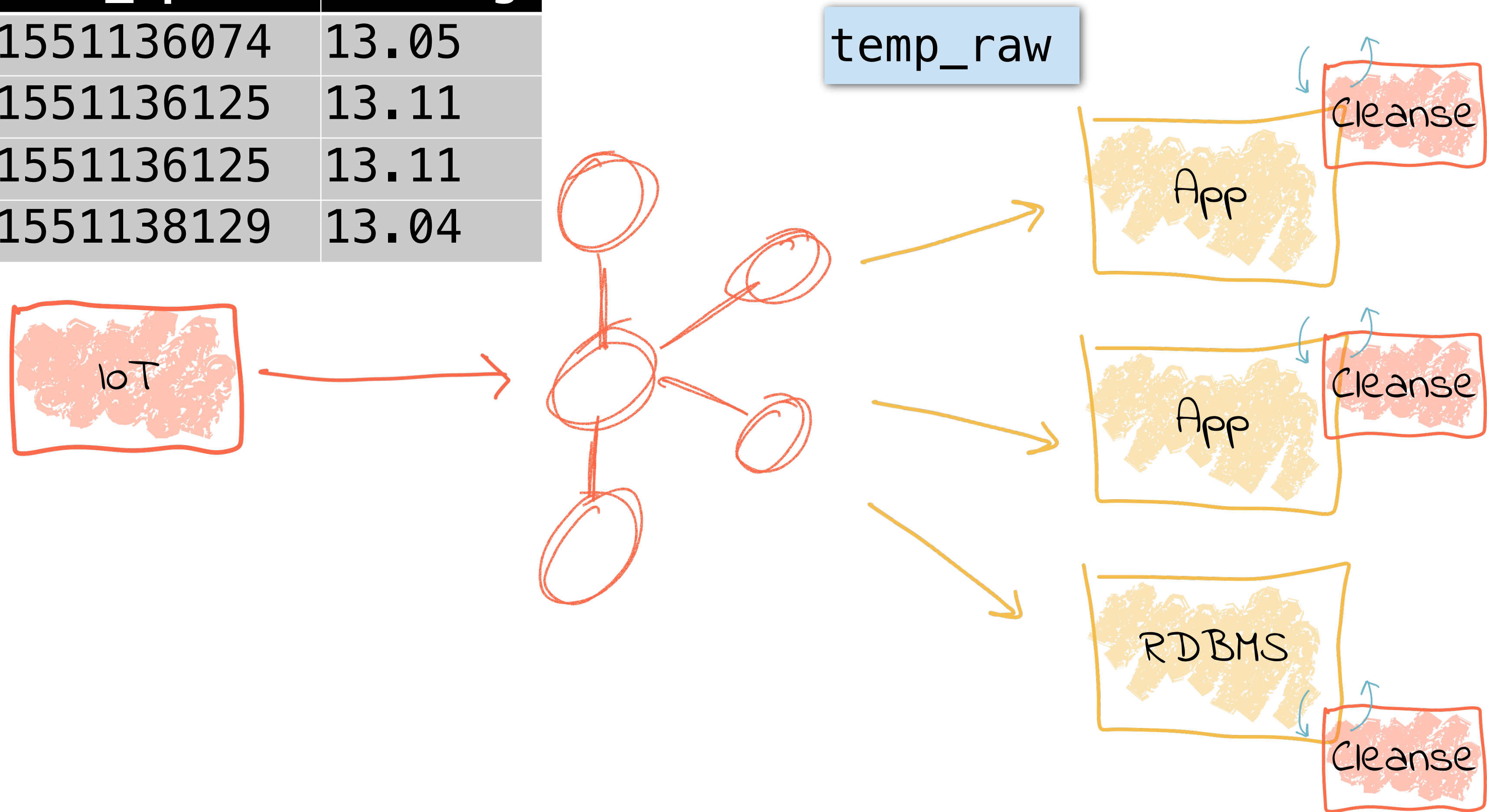
Transform Once, Use Many: Data Cleansing

sensor_id	time_epoch	reading
42	1551136074	13.05
42	1551136125	13.11
	1551136125	13.11
42	1551138129	13.04



Transform Once, Use Many: Data Cleansing

sensor_id	time_epoch	reading
42	1551136074	13.05
42	1551136125	13.11
	1551136125	13.11
42	1551138129	13.04

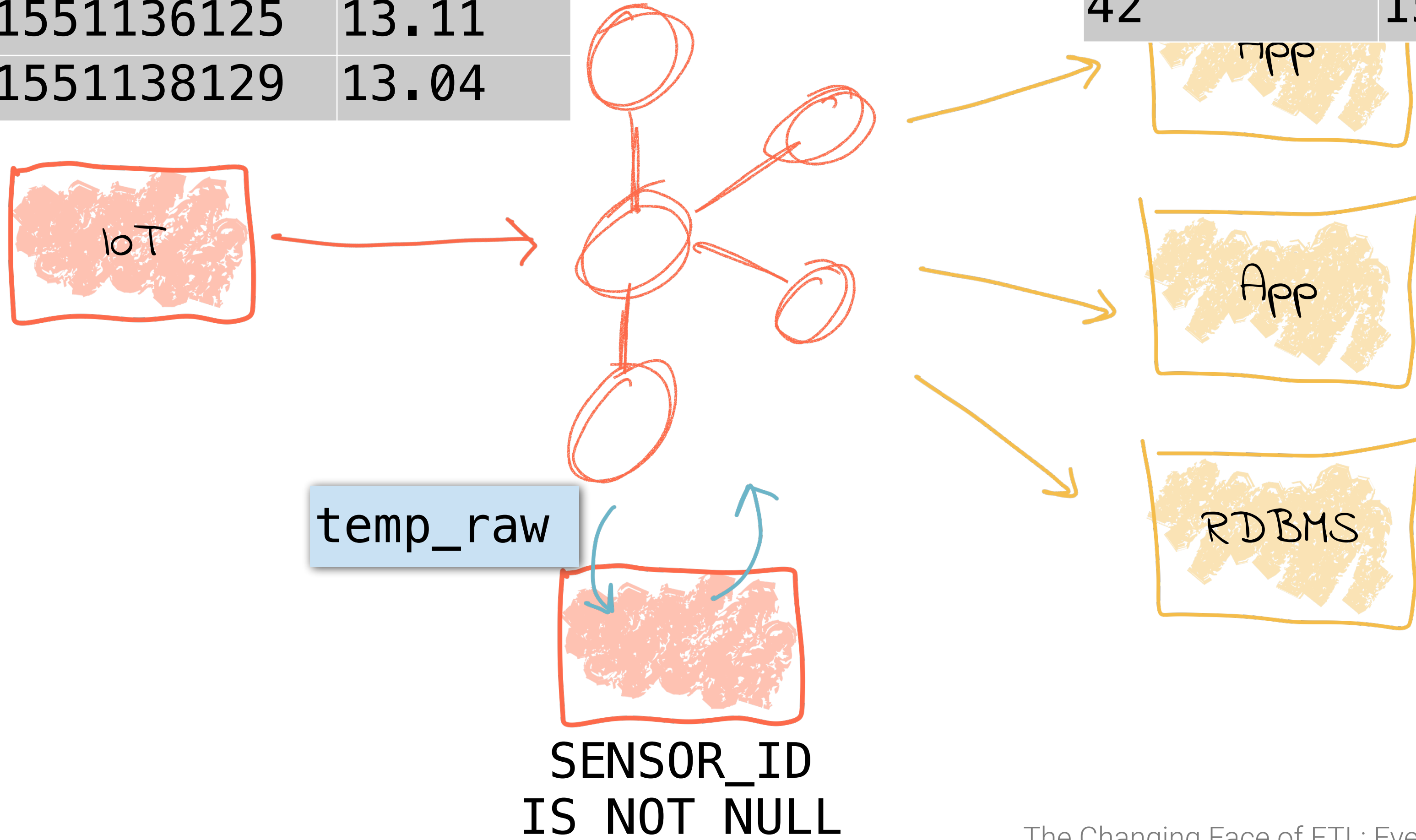


Transform Once, Use Many: Data Cleansing

sensor_id	time_epoch	reading
42	1551136074	13.05
42	1551136125	13.11
	1551136125	13.11
42	1551138129	13.04

temp_clean

sensor_id	time_epoch	reading
42	1551136074	13.05
42	1551136125	13.11
42	1551138129	13.04



Say NO to brittle pipelines



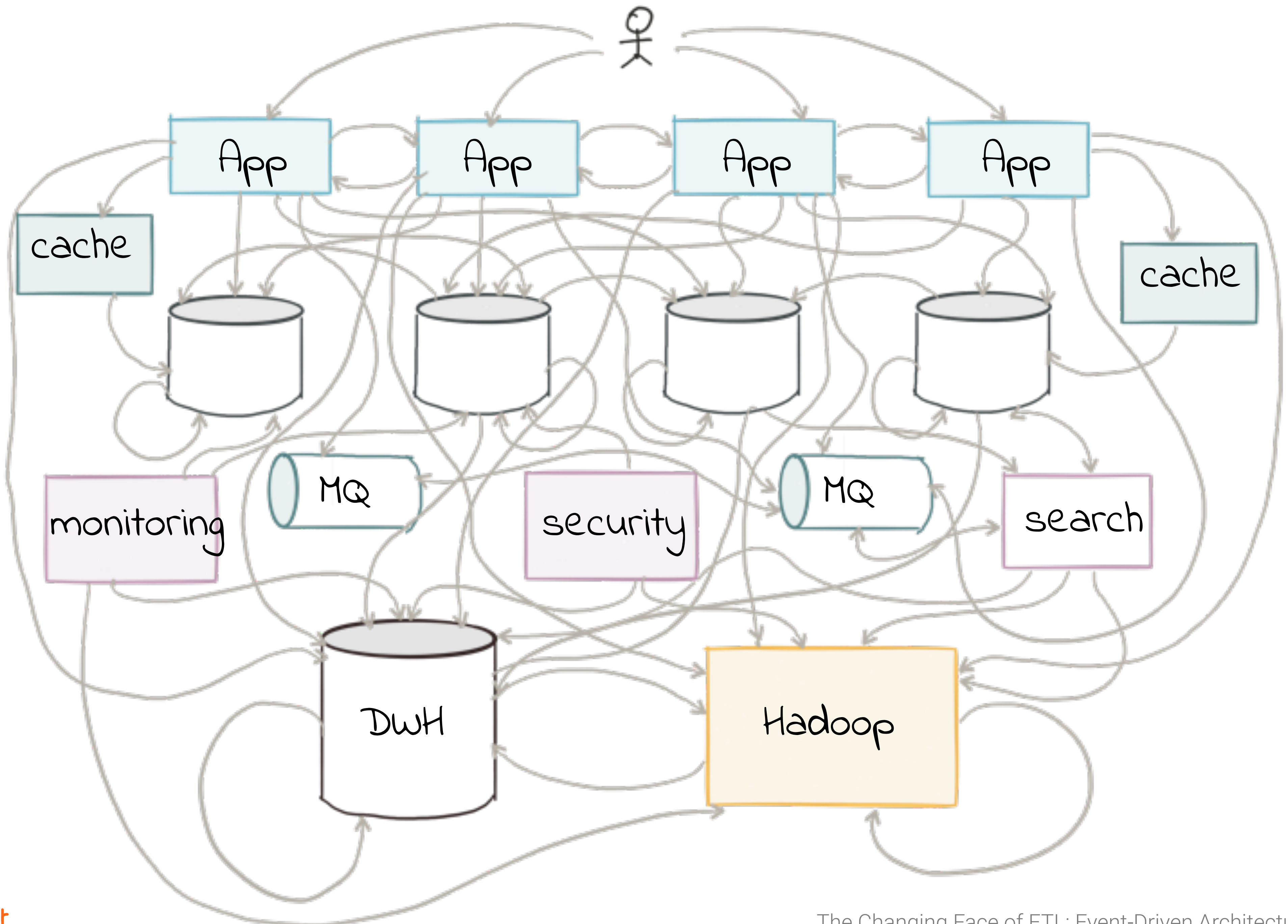
Latency requirements

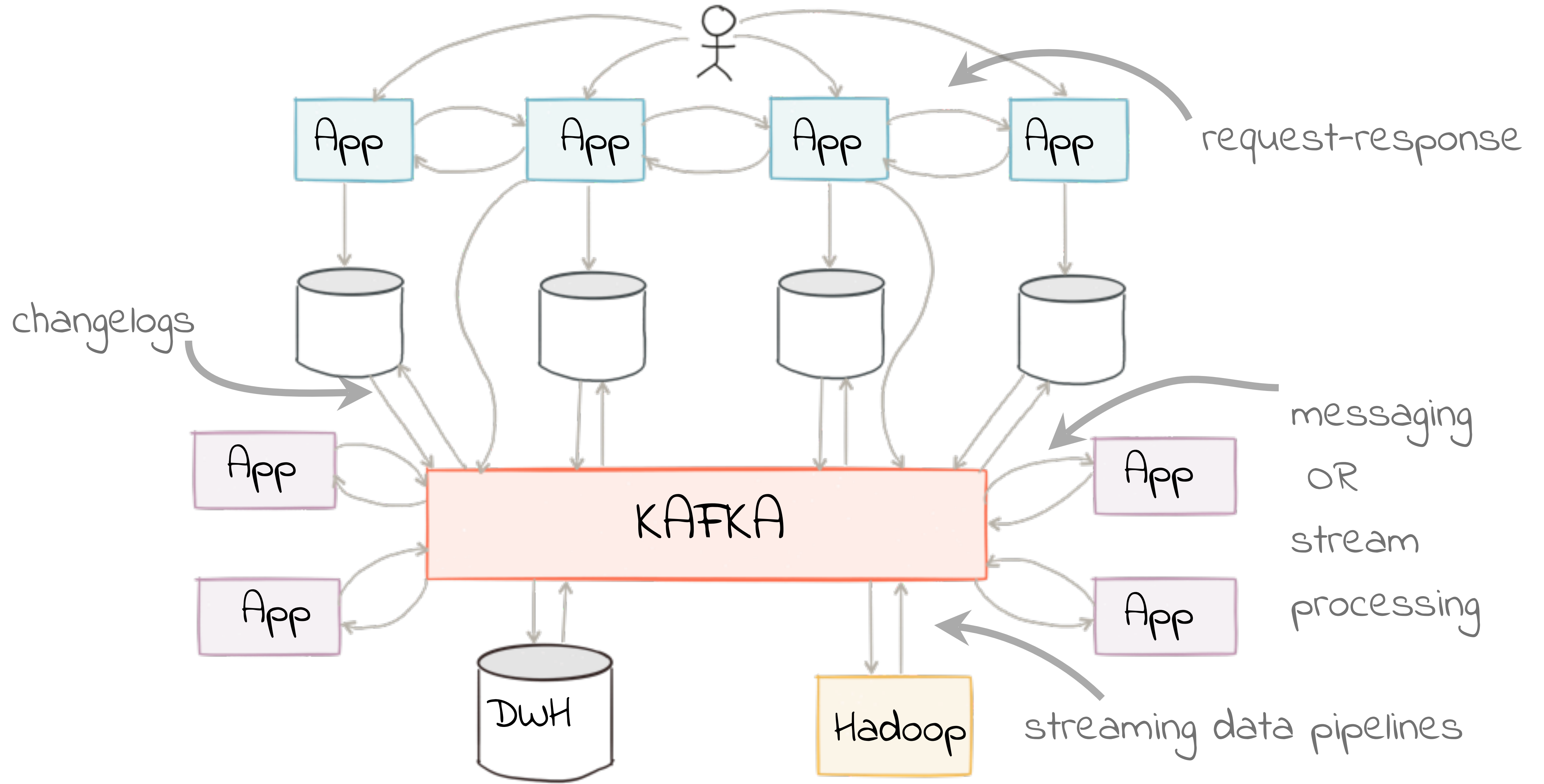
Users of the data

Scale

Data fidelity







EVENTS MODEL THE REAL WORLD



EVENT STREAMING PLATFORM

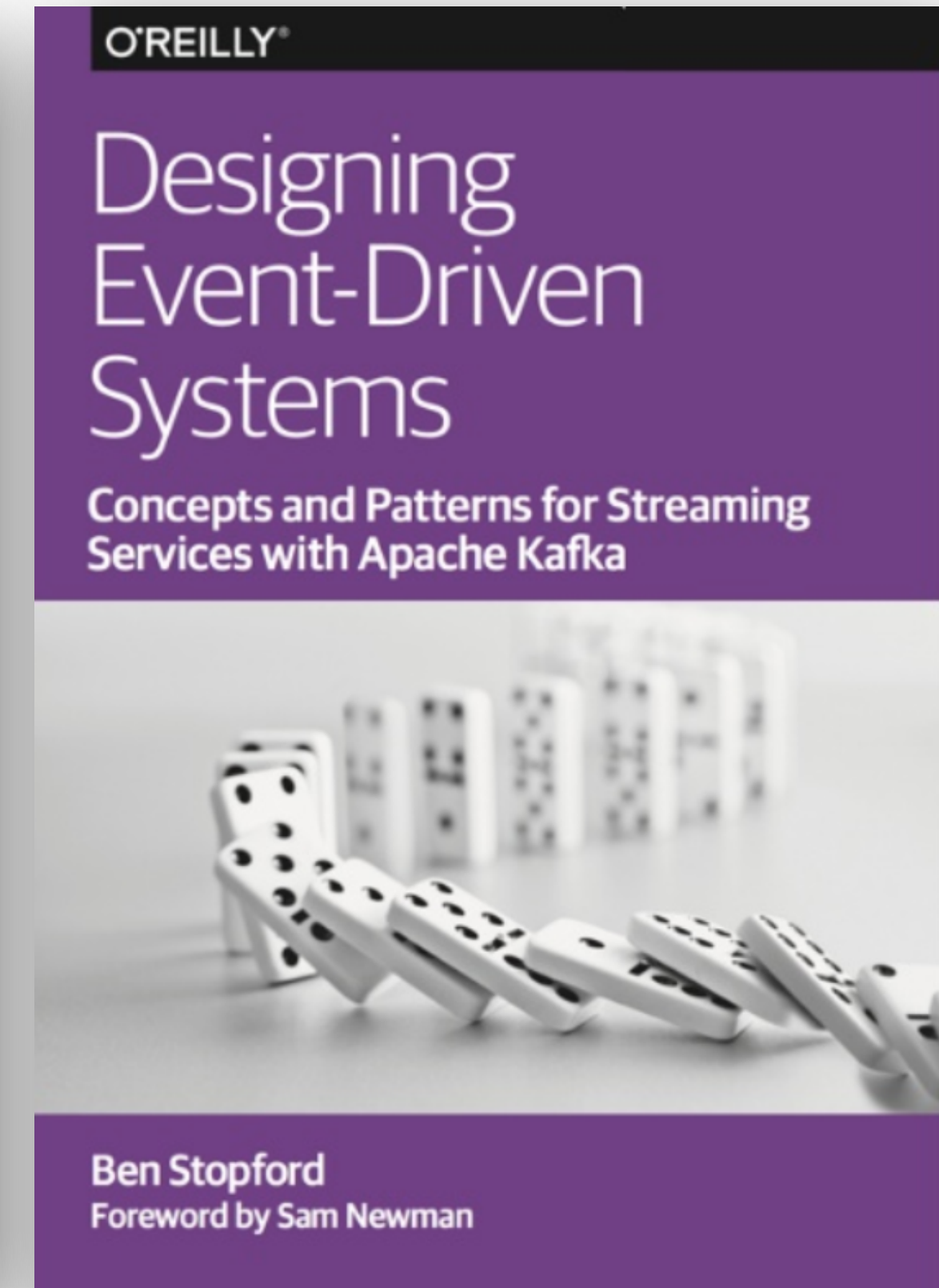
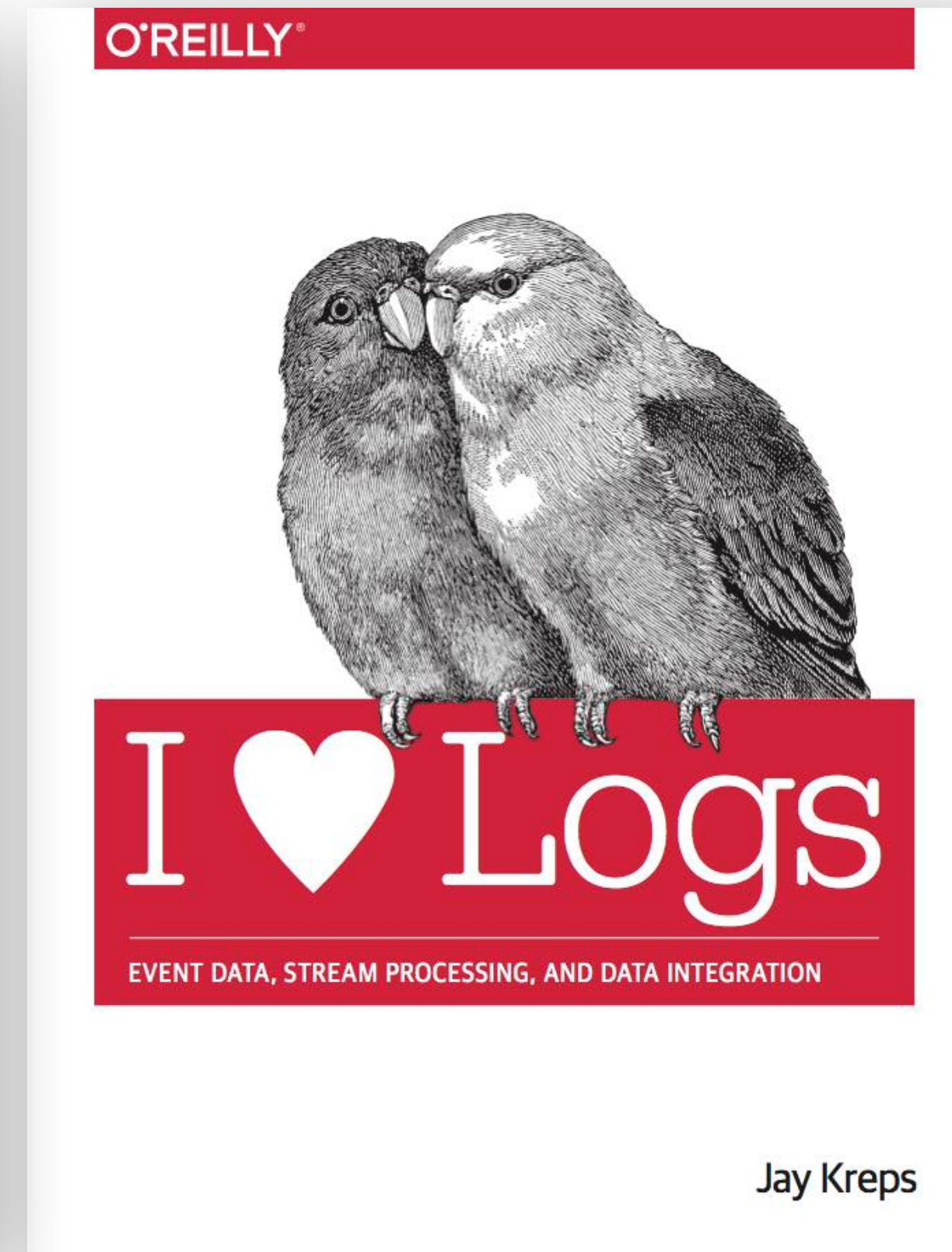
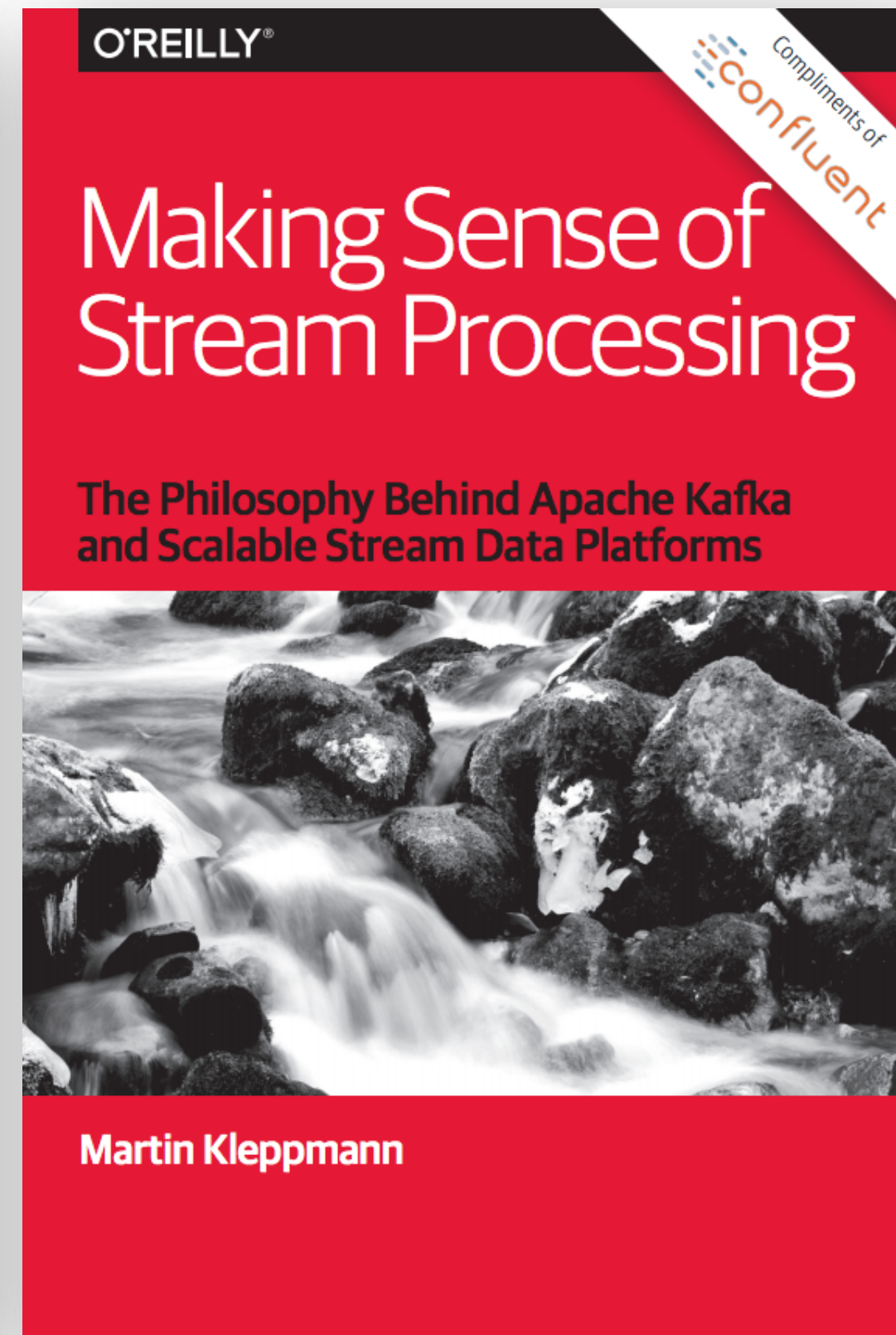
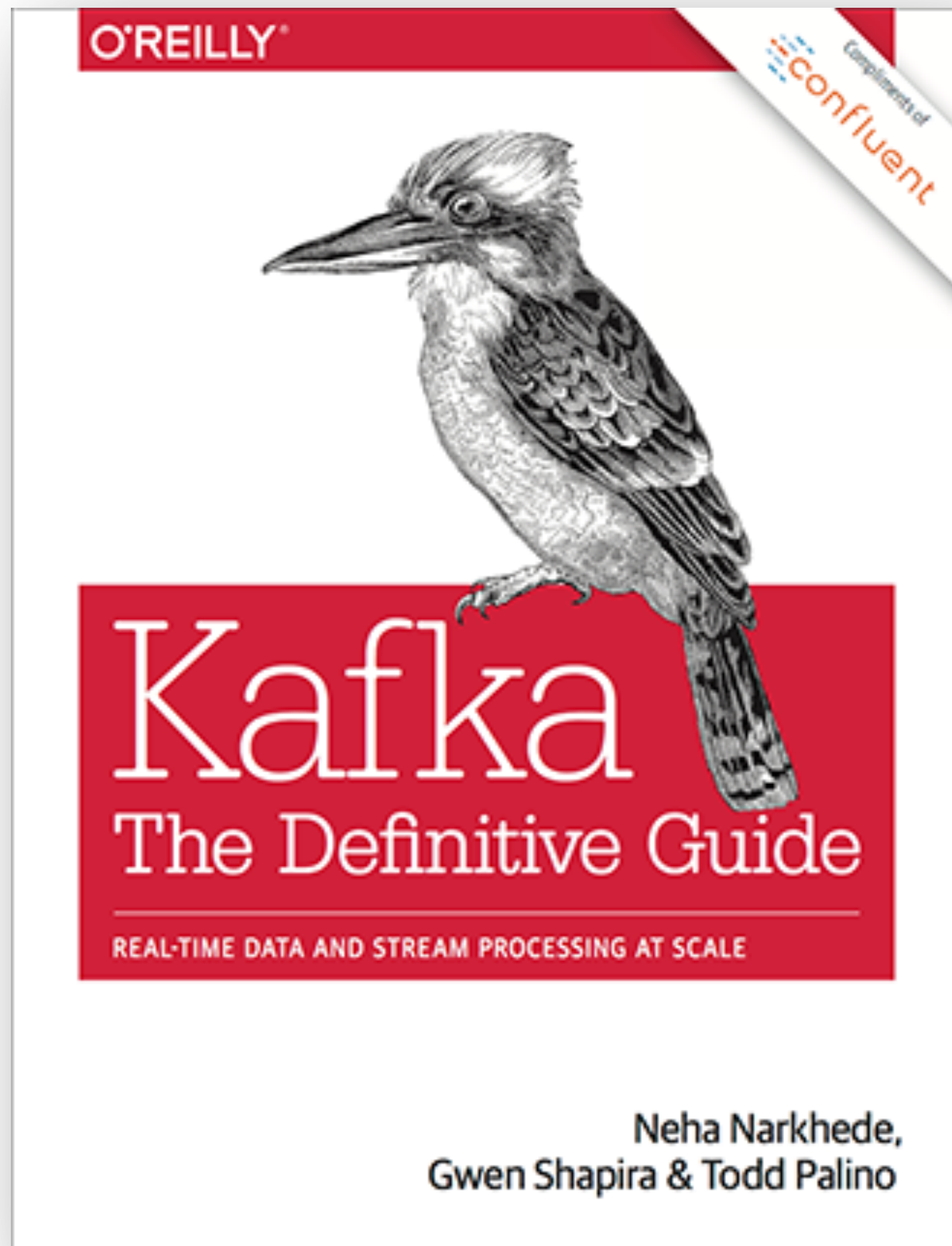
**NATIVE STREAM
PROCESSING**

DATA PERSISTENCE

**DATA WHEN
YOU NEED IT**

**FLEXIBILITY
+ SCALABILITY**

<http://cnfl.io/book-bundle>



confluent.io/download

<http://cnfl.io/book-bundle>

<http://cnfl.io/slack>

@rmoff

**EOF**