



kokchun giang

working with real-time data with
apache kafka - a distributed event streaming platform

traditional databases represents **physical & logical entities**



id	first_name	last_name	email

customer table

id	name	price	stock_qty

products table

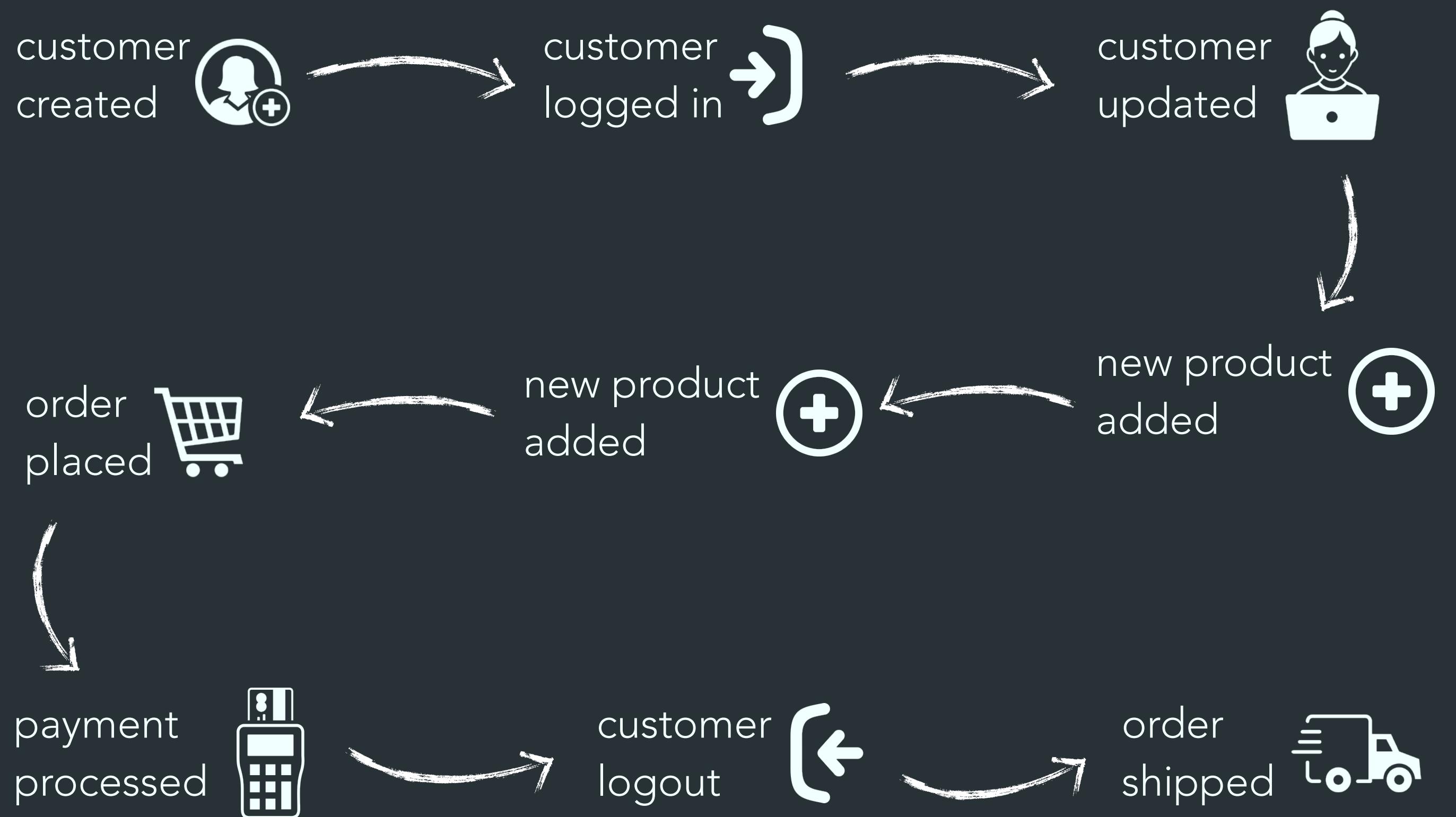
order_id	cust_id	order_date	total_amt

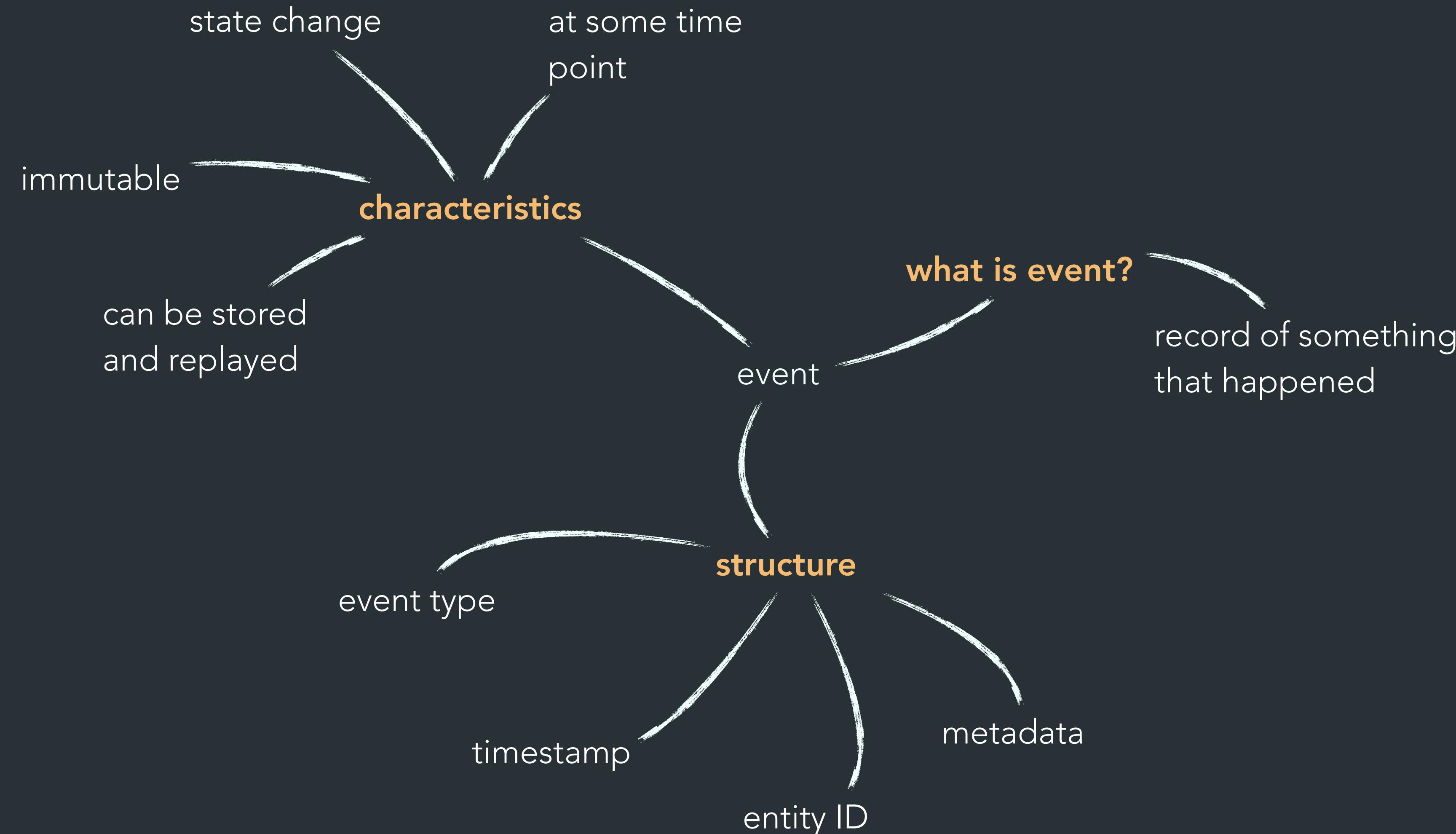
orders table

payment_id	order_id	payment_method	amount

payment table

real-time **events** online shopping example





example of an **event** in json structure

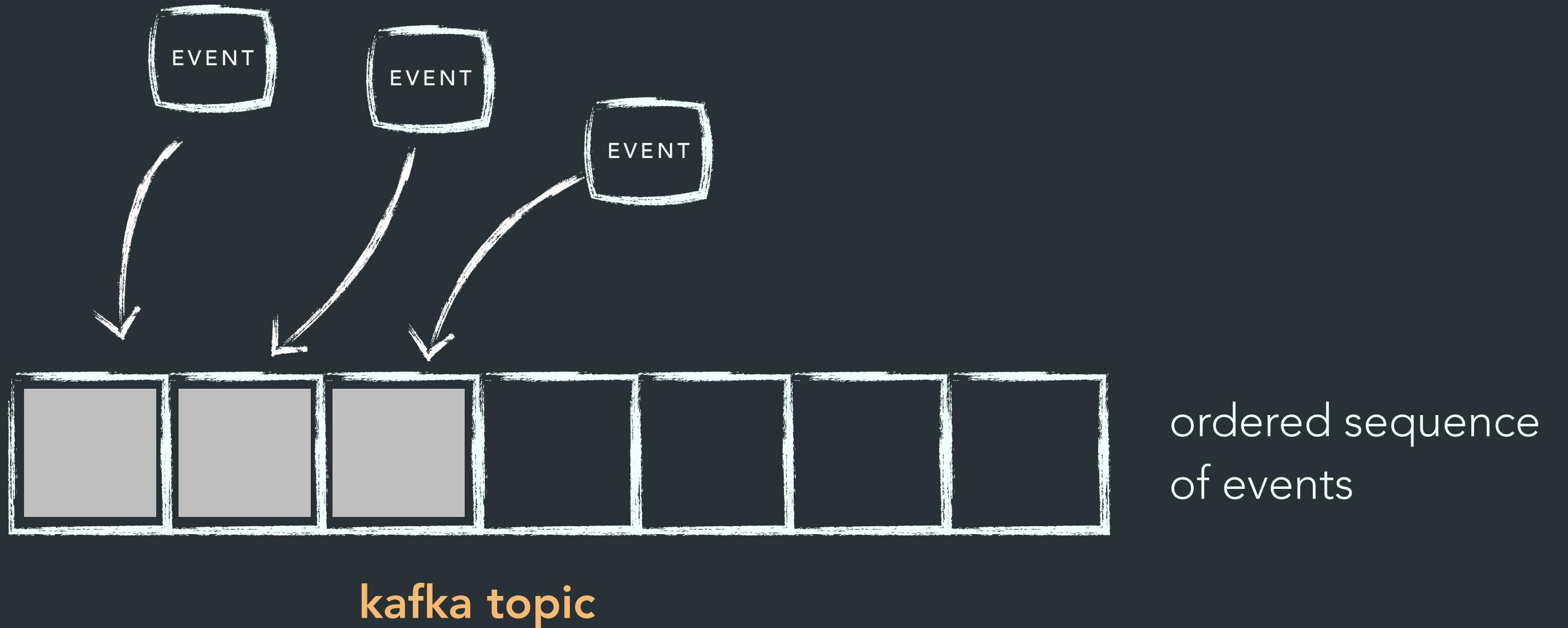
```
1  {
2      "event_type": "order_placed",
3      "timestamp": "2025-01-26T14:20:00Z",
4      "customer_id": 1001,
5      "order_id": 301,
6      "items": [
7          {
8              "product_id": 201,
9              "quantity": 1,
10             "price": 1200.0
11         },
12         {
13             "product_id": 202,
14             "quantity": 1,
15             "price": 800.0
16         }
17     ],
18     "total_amount": 2090.0
19 }
20 |
```

what happened?

can be used to
track what this
customer has done

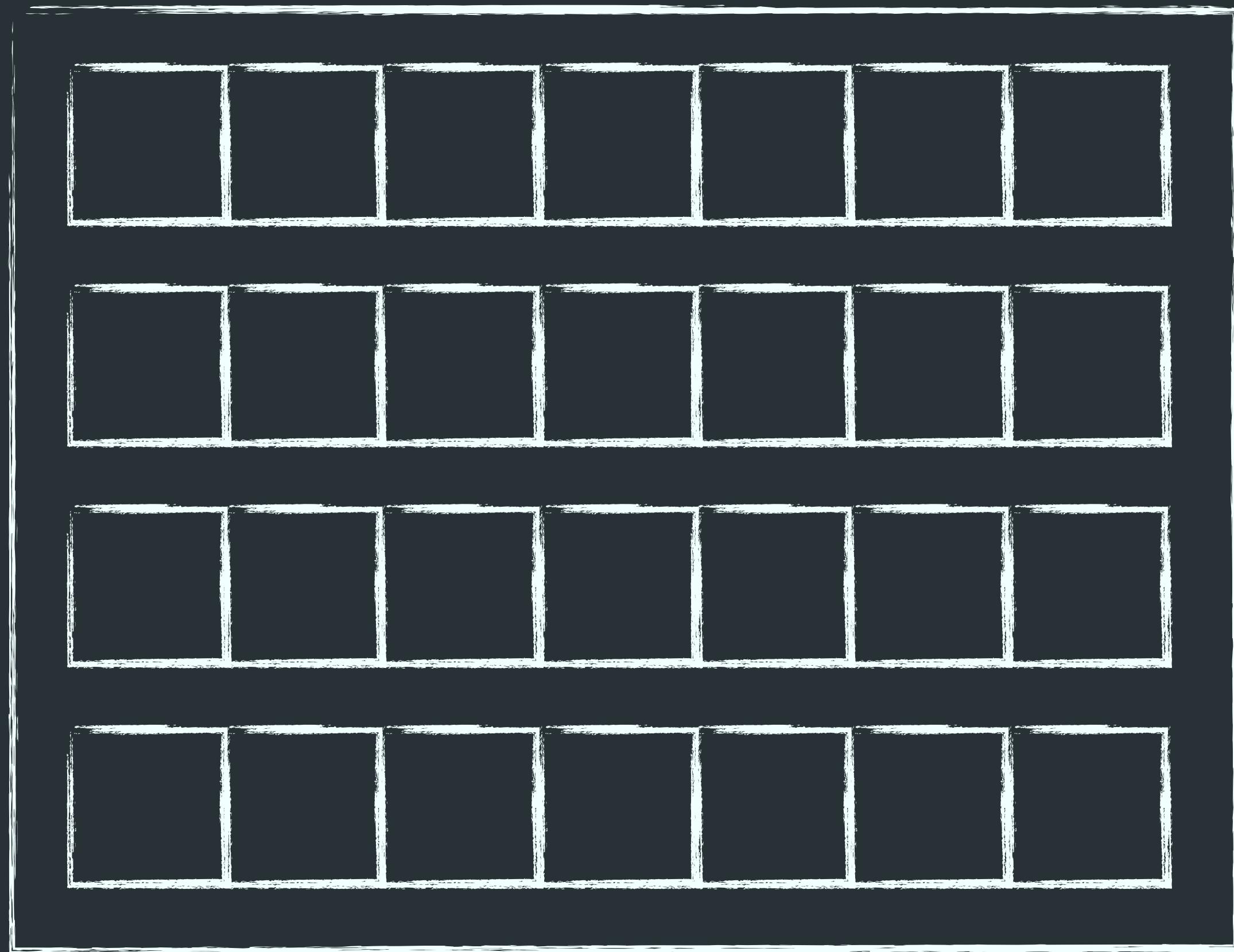
so how to store
these events?

these events are stored in **kafka topics**



analogy to topic is a folder in your
filesystem and events are files

a **kafka broker** is a server that stores and manages multiple **kafka topics**



no hard limit to how many events can exist in a topic

a kafka broker could be run on the cloud or locally with or without a docker container

kafka broker

a **kafka cluster** contains multiple **brokers** so
the platform is very scalable

