KAFKA

ASHISH GUPTA
@thriverashish

1. Event Streaming
2. Use of Event stream
3. Kafka as Event Streaming Platform
4. How does kafka Work
5. Concepts & Terminologies.

EVENT STREAMING

↳ Capturing data in Real-time from event-sources like databases, sensors, mobile devices, cloud services and software applications in the form of stream of events.

* Events Streams are stored durably for later retreival, manipulating, processing and reacting to the event streams in real time as well as retrospectively.

* Ensures continous flow and interpretation of data so that right information is at right place at the right time.

@thiveraskish

* What can I use event streaming for:

1. Process payments & transactions in real-time such as Stocks exchange, banks etc.

2. Track logistics info in real time.

3. React to customer interactions and orders such as in Retail, the hotel and travel industry.

4. Capture and Analyse sensor data from IOT devices or other equipment.

@thriverashish

* KAFKA is an event streaming platform:

↳ combines 3 key capabilities

1. To publish (write) and subscribe (read) stream of events.

2. To store streams of events durably and reliably for as long as you want.

3. To process streams of events as they occur or retrospectively.

** All these functionalities provided in a Distributed, highly scalable, elastic, fault-tolerant and secure manner.

@thriverashish

How does Kafka work ?

It is a distributed system consisting of servers and clients that communicate via high performance TCP Network Protocol.

SERVERS : Kafka is run as a cluster of one or more servers that can span multiple datacenters or cloud regions.
Some of these servers from the storage layer called the Brokers

Other servers run Kafka Connect to continously import and export data as event streams to integrate Kafka with your existing systems such as relational DBs as well as other Kafka clusters.

Clients : They allow you to write distributed application and microservices that lead, write and process streams of events in parallel, at scale and in fault-tolerant manner even in case of network problems.

@thriverashish

* Main Concepts and Terminology

Event : → Record or Message

↳ "something Happened"

Writing | Reading data in Kafka is in the form of Events

Event has a Key, Value, and timestamp and optional metadata.

PRODUCERS :

↳ clients that publish (write) events to Kafka.

CONSUMERS :

↳ clients that subscribe to (read or process) events.

Producers & Consumers are fully decoupled and Agnostic to each other.

@thriverashish

Topics

→ Events are organised and durably stored in topics.

or

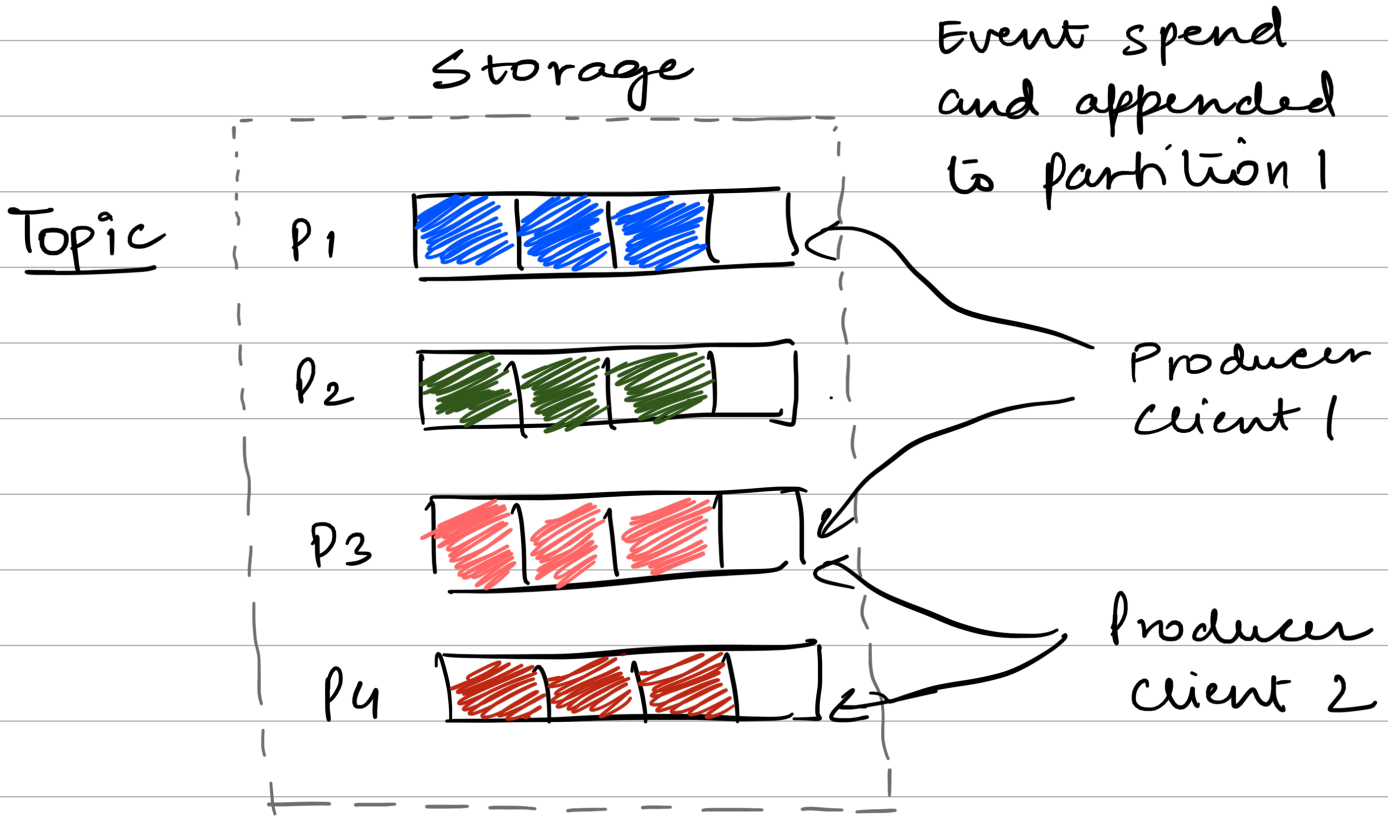
Topic is a ordered collection of events.

Topic in Kafka is always multi-producer and multi-subscriber

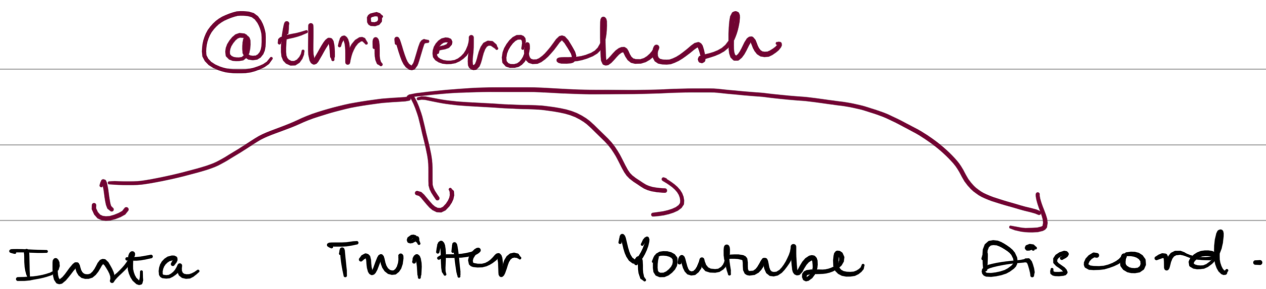
↳ it can have zero, one or many producers and consumers to write and read.

Topics are **Partitioned** meaning a topic is spread over number of buckets located on different Kafka **Brokers**.

Events with same event key (eg. a (customer or vehicle id) are written to same partition and Kafka guarantees that any consumer of a given topic-partition will always read partition event in same order they are written.



To make data fault-tolerant and highly available every topic can be replicated. even across geo-regions or data centers.



Ashish GUPTA