

BigData Series

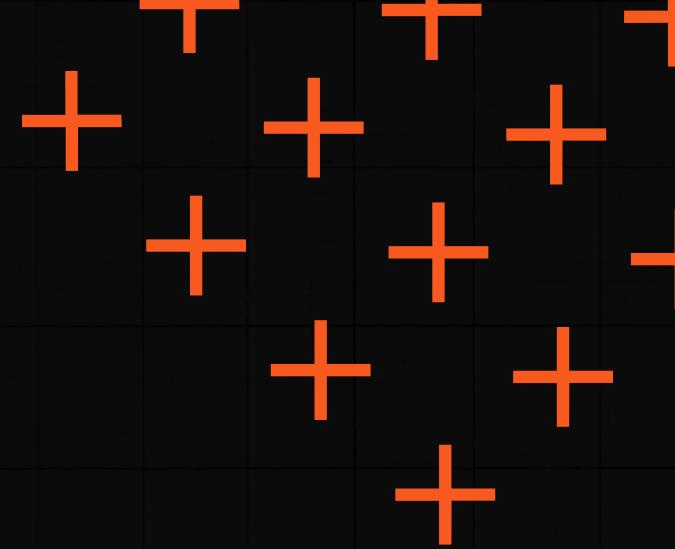
PART 18/xx

KAFKA API
& Cluster

@Abhijith



66



Recap :

consumers:



- pull records from partitions using tracked offsets.

consumer groups

- share partitions to scale parallel consumption.

Replication:

Replication:

- duplicates partitions across brokers for fault tolerance



Why Kafka APIs Matter?

- 1.** • define how applications interact with Kafka.
- 2.** • separate data ETL, processing, and integration.
- 3.** • write data
- 4.** • read data
- 5.** • process streams and integrate external systems





Producer API

- 1.** • publishes records to Kafka topics.
- 2.** • sends data to a topic. asynchronous by default.
- 3.** • chooses partition (key-based or round-robin)
- 4.** • supports batching and compression
- 5.** • Used when applications generate events.





Consumer API

- 1.** • reads records from topics. scales w consumer groups
- 2.** • pulls data from partitions
- 3.** • tracks progress using offsets
- 4.** • supports replay and recovery.
- 5.** • Used when applications react to events.





Streams API

- 1.** • enables stream processing inside applications.
- 2.** • transforms data in real time.
- 3.** • supports joins, aggregations, windows
- 4.** • state stored locally with changelogs
- 5.** • Used for real time analytics. No external cluster used.

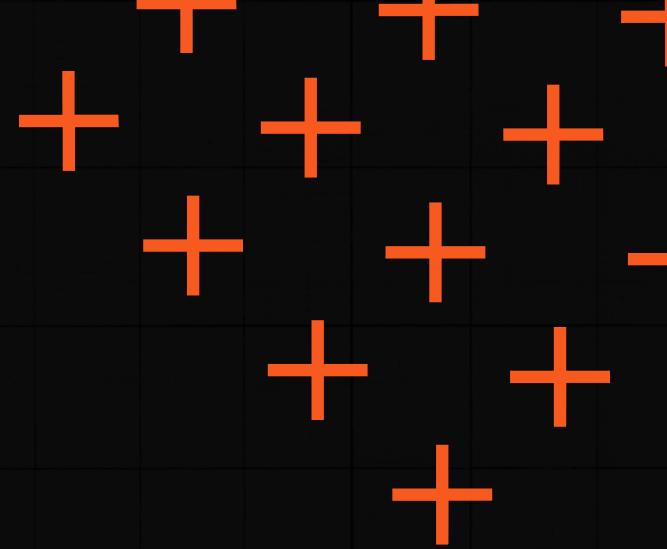




Connector API

- 1.** • integrates Kafka with external systems.
- 2.** • source connectors import data into Kafka.
- 3.** • sink connectors export data out. standardized.
- 4.** • minimal custom code, reusable integrations
- 5.** • Used for databases, storage, SaaS systems.



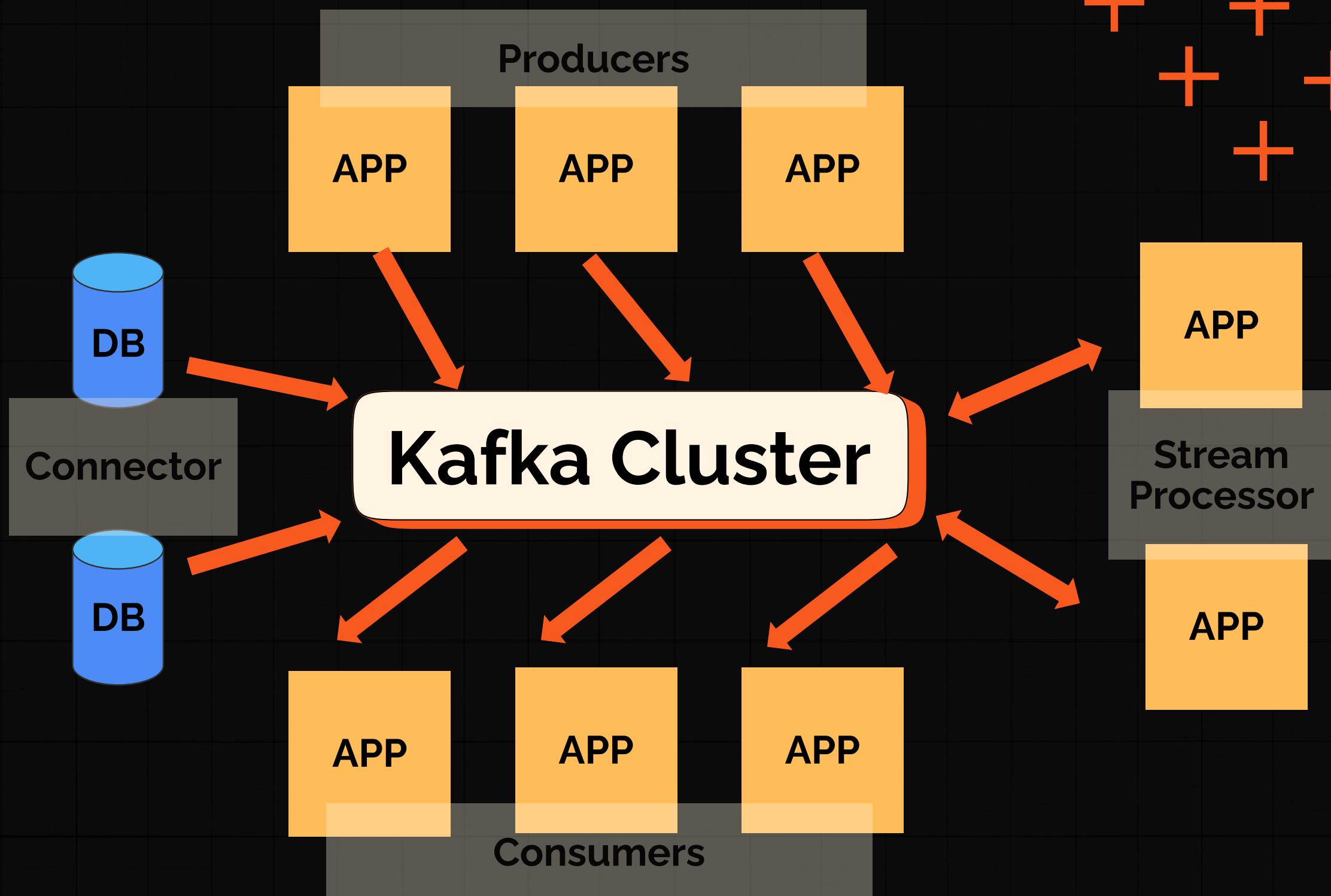


Kafka Cluster

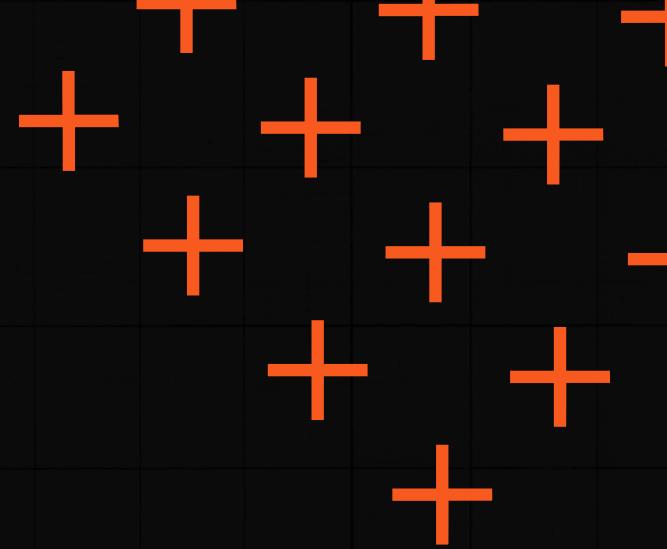
- 1.** • a group of brokers working together.
- 2.** • brokers store partitions.
- 3.** • partitions replicated for fault tolerance
- 4.** • leaders handle reads and writes
- 5.** • The cluster ensures scalability and durability.



66



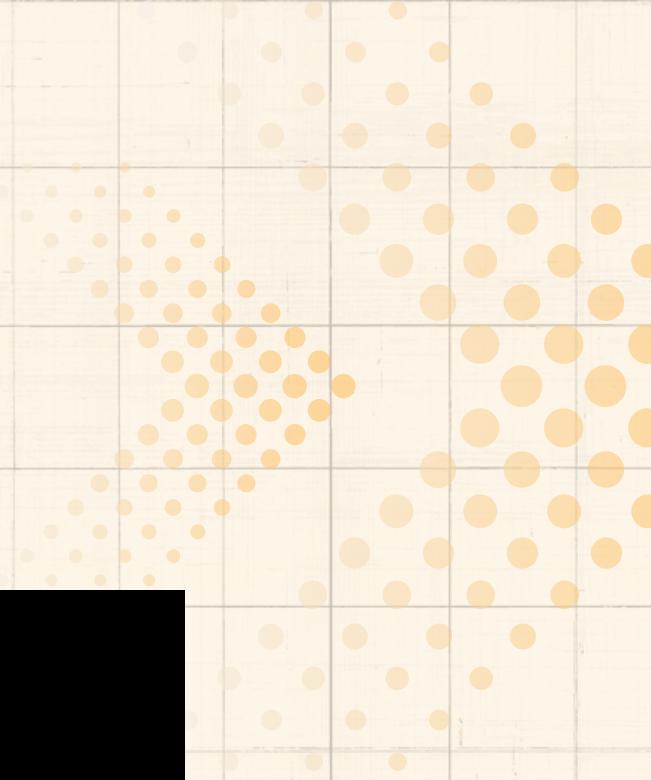
66



API + Cluster

- 1.**
 - APIs operate on top of the cluster.
- 2.**
 - Cluster + APIs = end-to-end data platform.
- 3.**
 - write, read, process and move data using API's.

66



Up Next:

Kafka in practice

**Follow for more and
share this post with
others who might find
it helpful!**

#HappyLearning #BigDataSeries #ML

@Abhijith

