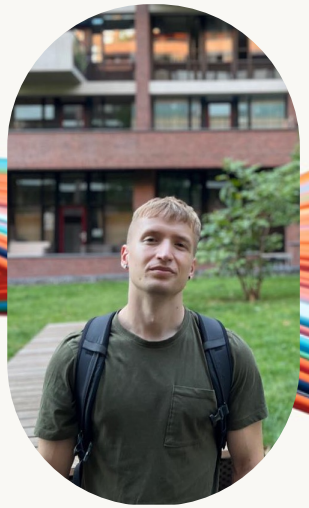
Kafka Internals

Architecture, APIs & Performance

Dominik Táskai, Senior Software Engineer
2025.07.23

INTRO

About Me



Dominik Táskai

SENIOR SOFTWARE ENGINEER, GENESYS

dtaskai

/in/dtaskai

- Working on globally distributed systems
- AWS, Data, Kubernetes
- Open-source software contributor
- Linkerd Ambassador
- KCD Organizer
- KubeCon Co-located Events Speaker

Agenda

01 **Topics, Partitions, Offsets**
Ordering, Data Distribution

02 **Producers**
Message Keys, Acknowledgements

03 **Consumers**
Serialization, Deserialization

04 **Brokers**
Broker Discovery, Metadata

05 **Replication**
Durability

06 **Security**
SSL, SASL, Kerberos

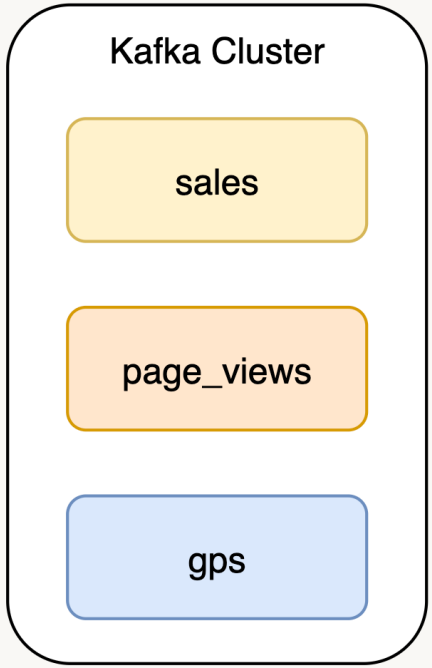
07 **Strimzi**
Kafka on Kubernetes

08 **Demo**

TOPICS, PARTITIONS, OFFSETS

Topics

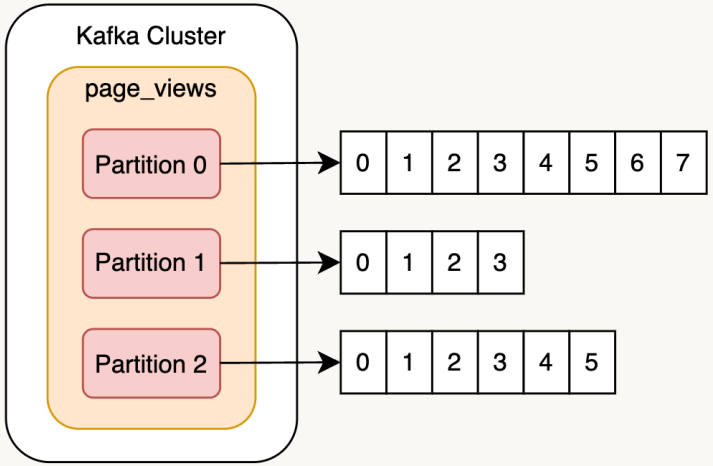
- Data streams
- Supports arbitrary message formats
- Categorized by business use cases
- Partitioned to scale
- Producers write, Consumers read
- Configured retention (NOT exactly a DB)



TOPICS, PARTITIONS, OFFSETS

Partitions

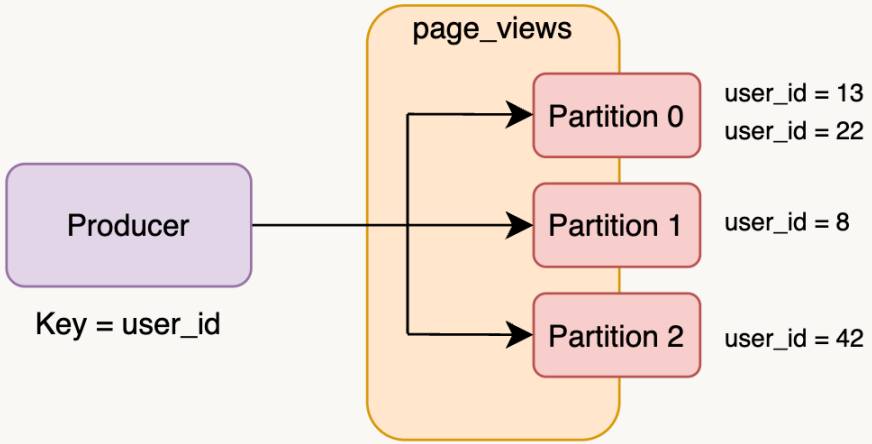
- Topics can be split into **partitions**
- Enables horizontal scaling
- Messages identified by **offset**
- Guaranteed order within partition
- Immutable
- Data assigned randomly (unless **key** is specified)



PRODUCERS

Producers, Message Keys

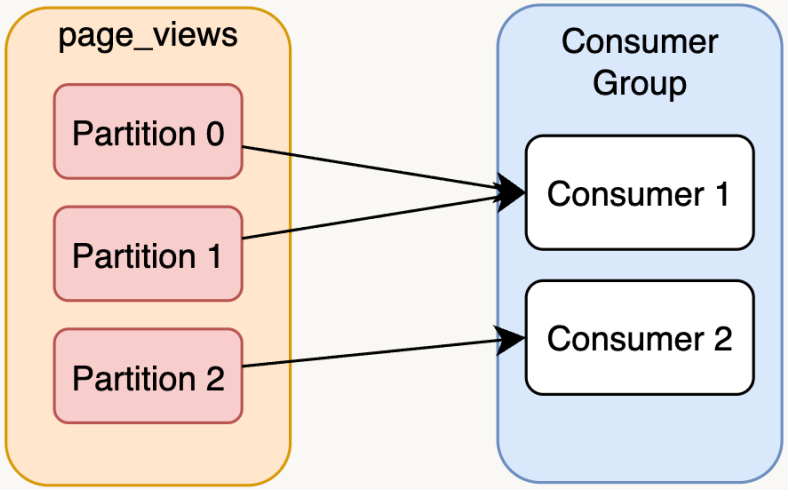
- Producers write to topics
- Optional message key
- Acknowledgement configuration
- Retry mechanisms
- Select the key based on ordering needs



CONSUMERS

Consumers

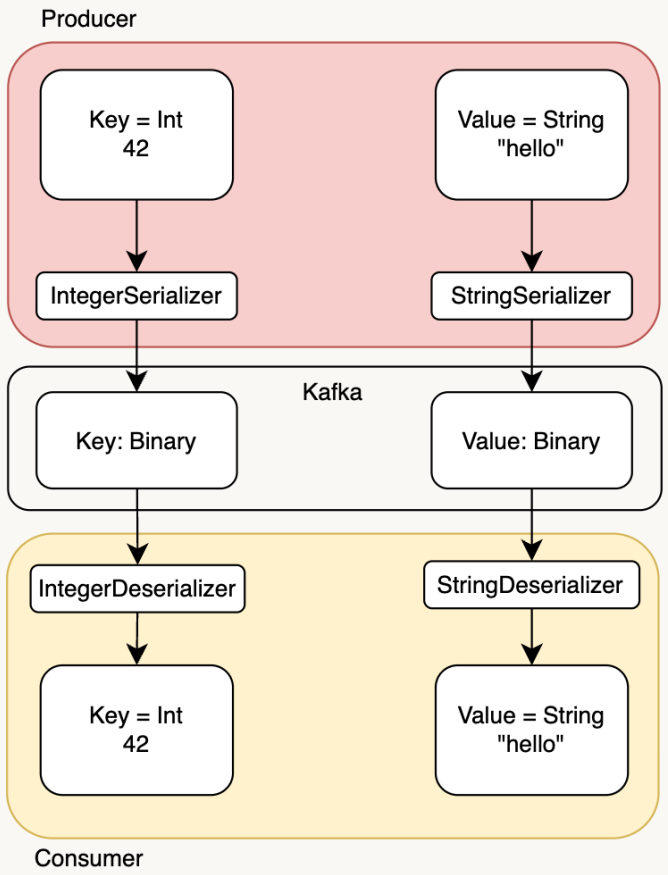
- Consumers read from topics
- Internally load balanced
- One partition per consumer
- Offsets are committed after processing
- Processing should be idempotent



CONSUMERS

Serialization, Deserialization

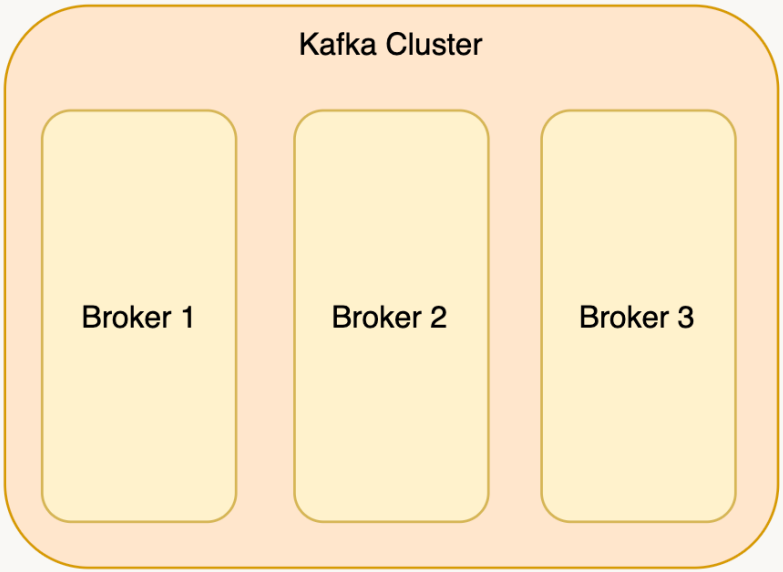
- Indicates how to transform objects
- Must be compatible pairs
- Type must not change!
- Supports lots of types (JSON, Avro, Protobuf)
- Custom serializer/deserializer



BROKERS

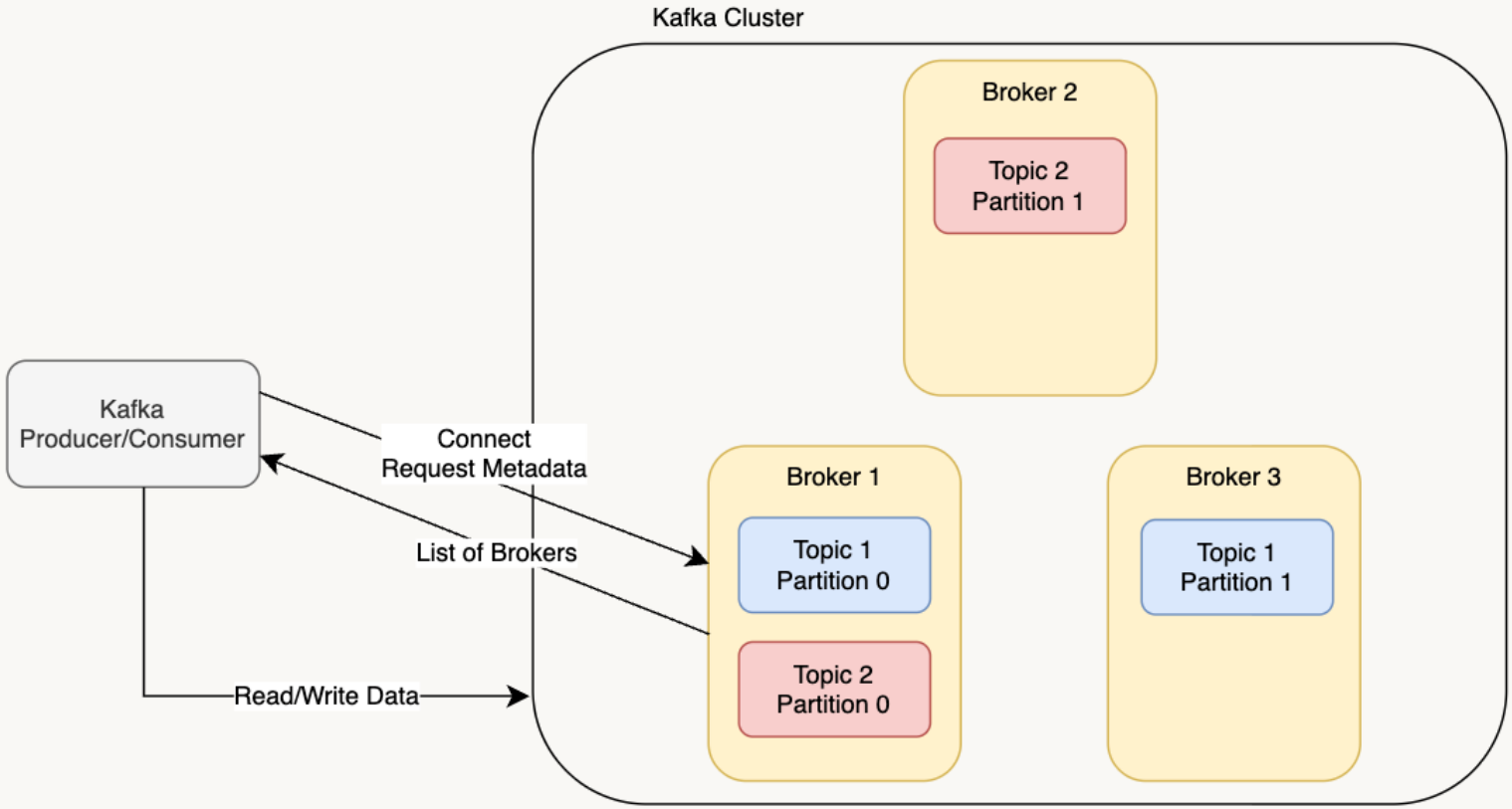
Brokers, Broker Discovery, Metadata

- Cluster consists of multiple server (brokers)
- Identified by ID
- Contains certain **partitions**
- Each serves metadata about the entire cluster
- No need to know metadata it beforehand



BROKERS

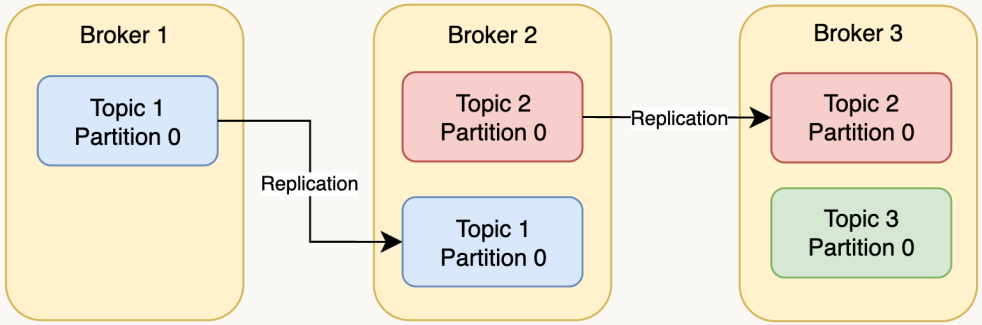
Brokers, Broker Discovery, Metadata



REPLICATION

Replication, Durability

- Configured with replication factor
- Recommended: > 1
- Can’t exceed broker count
- Storage considerations
- Network bandwidth
- Consistency settings (acks=all)



SECURITY

SSL, ACLs, Kerberos

- Data is ”human readable” by default
- **SSL** encrypts data in transit
- SSL can be used for authentication
  - Required for ACL
- **ACLs**: Topic, Consumer Group, Cluster permissions
- **Kerberos**: network authentication protocol
  - Complicated, cryptic errors

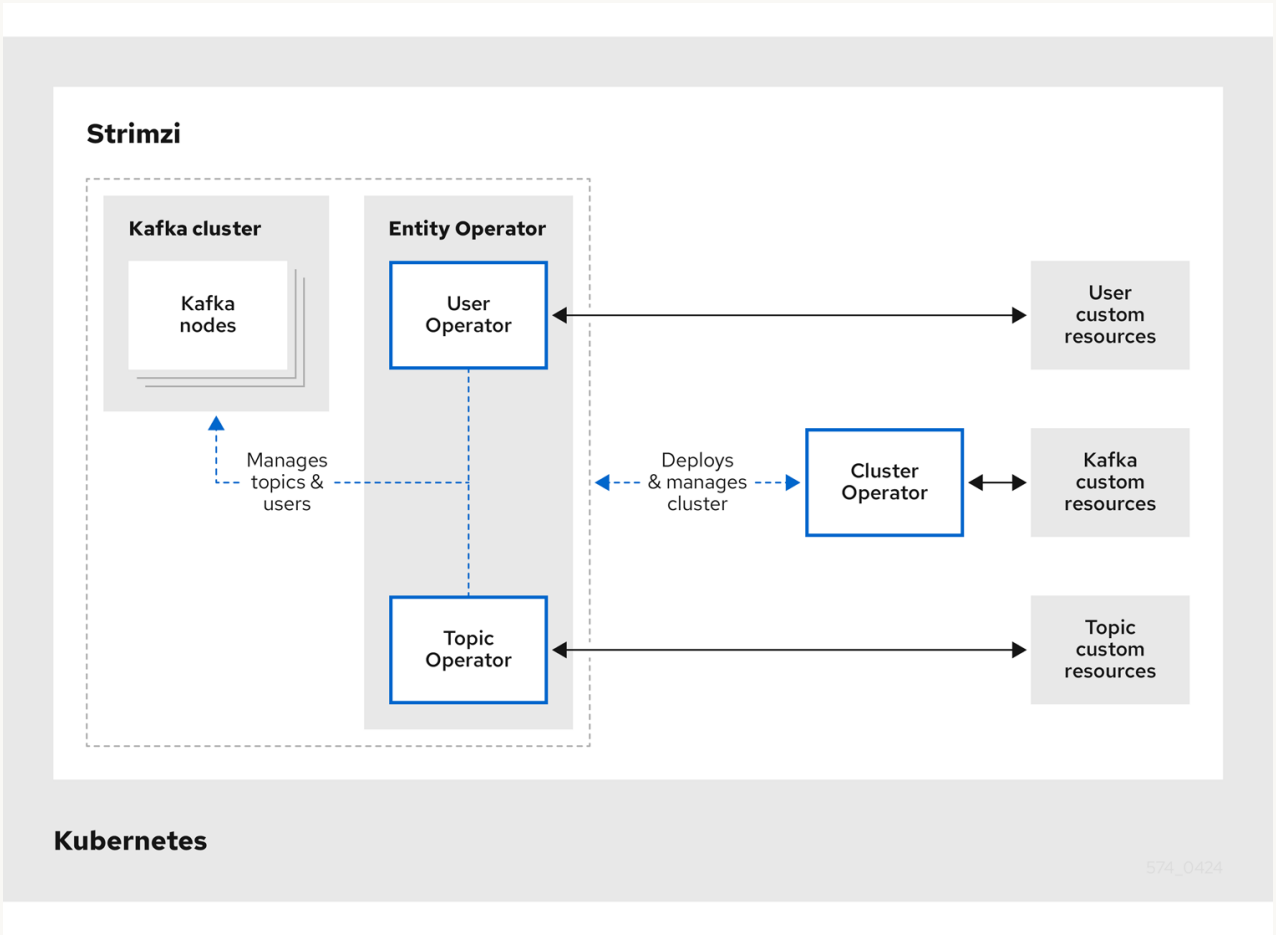
STRIMZI

Strimzi

- Kubernetes Operator Pattern
- Declarative Kafka resource management
- Cluster, Topic and User Operator
- Consistent deployments
- Automated scaling
- Automated TLS and RBAC

STRIMZI

Strimzi



Demo

https://github.com/dtaskai/kafka-hwsw

We are hiring!

- Associate Mobile Software Engineer
- Senior Software Engineer
- Software Development Engineer (Front-end)

Thank you