

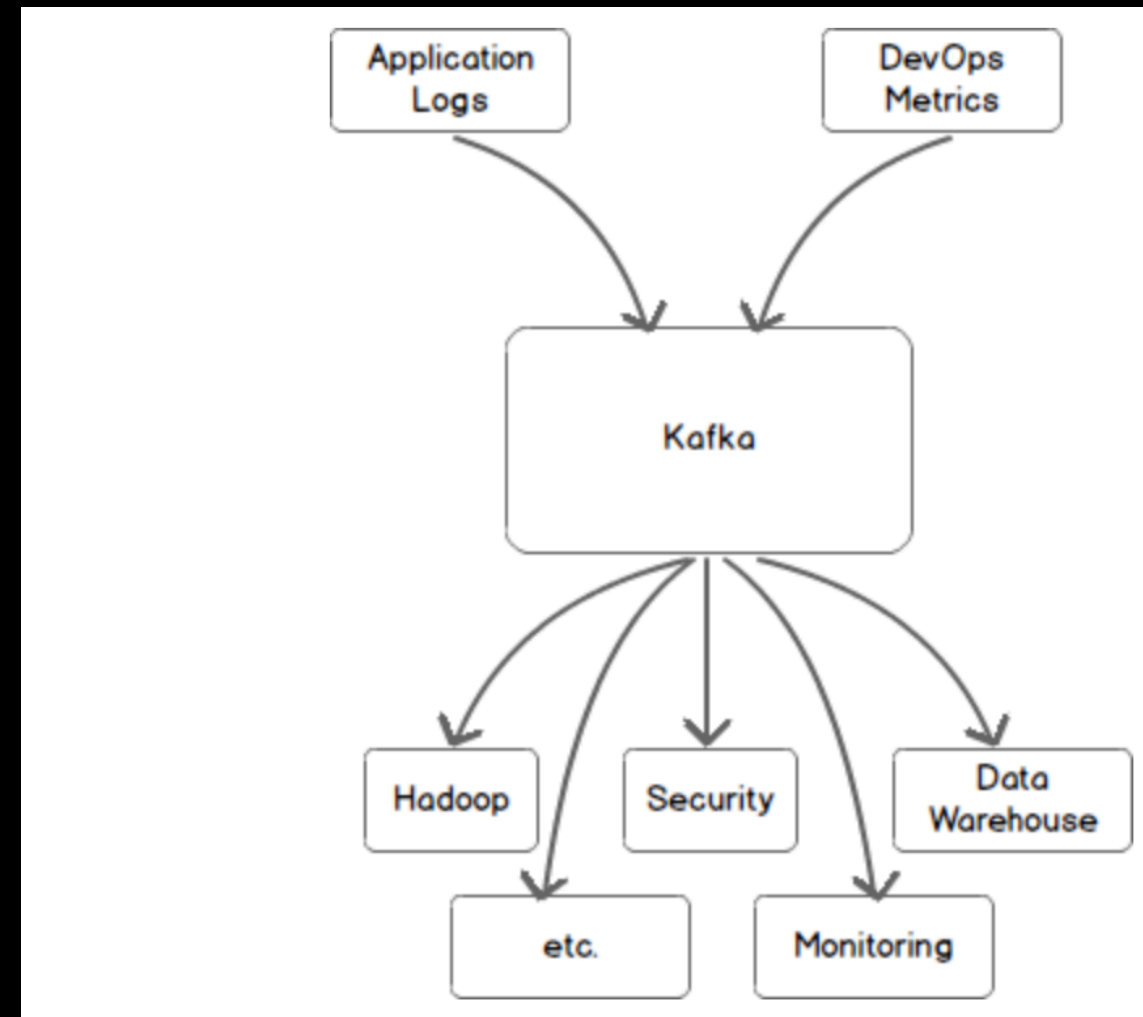
MESSAGING

WITH

KAFKA

WHAT IS KAFKA

- **Circular Distributed** Messaging System
- Purpose to move data between systems
- Allow huge number of ad-hoc & permanent consumers
- High available and resilient to node failure, Automatic recovery
- Communication & Integrations between components



USE CASES

- Messaging
- Website Activity Tracking
- Metrics
- Log Aggregation
- Sensor Data IOT

TERMINOLOGY

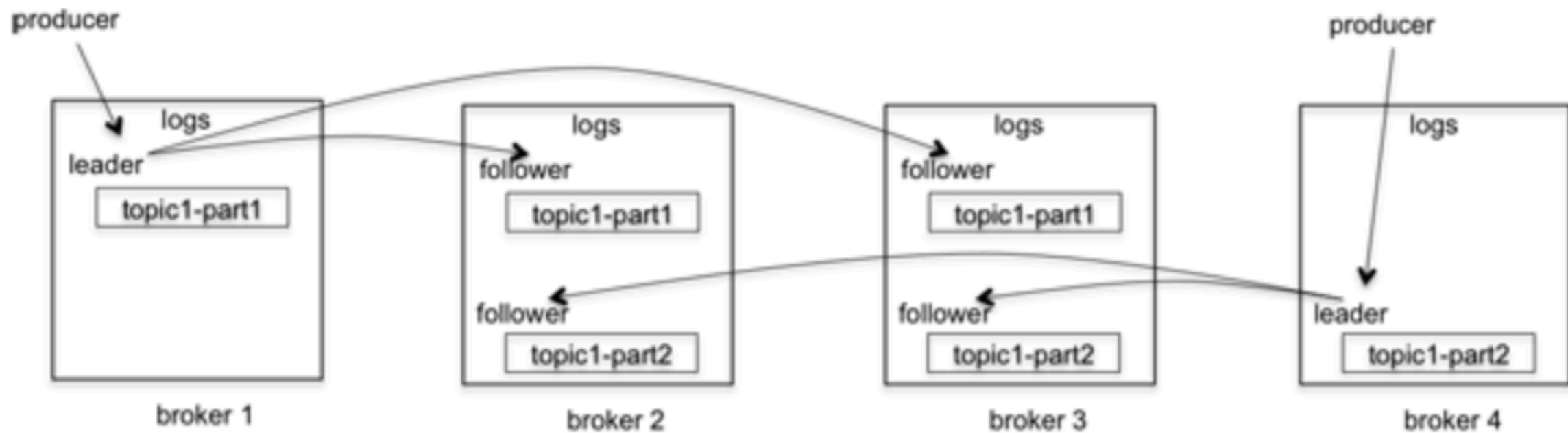
- Topics
- Producers
- Consumers
- Key
- Message
- Brokers
- Partitions
- Offset
- Consumer Groups

WHY KAFKA?

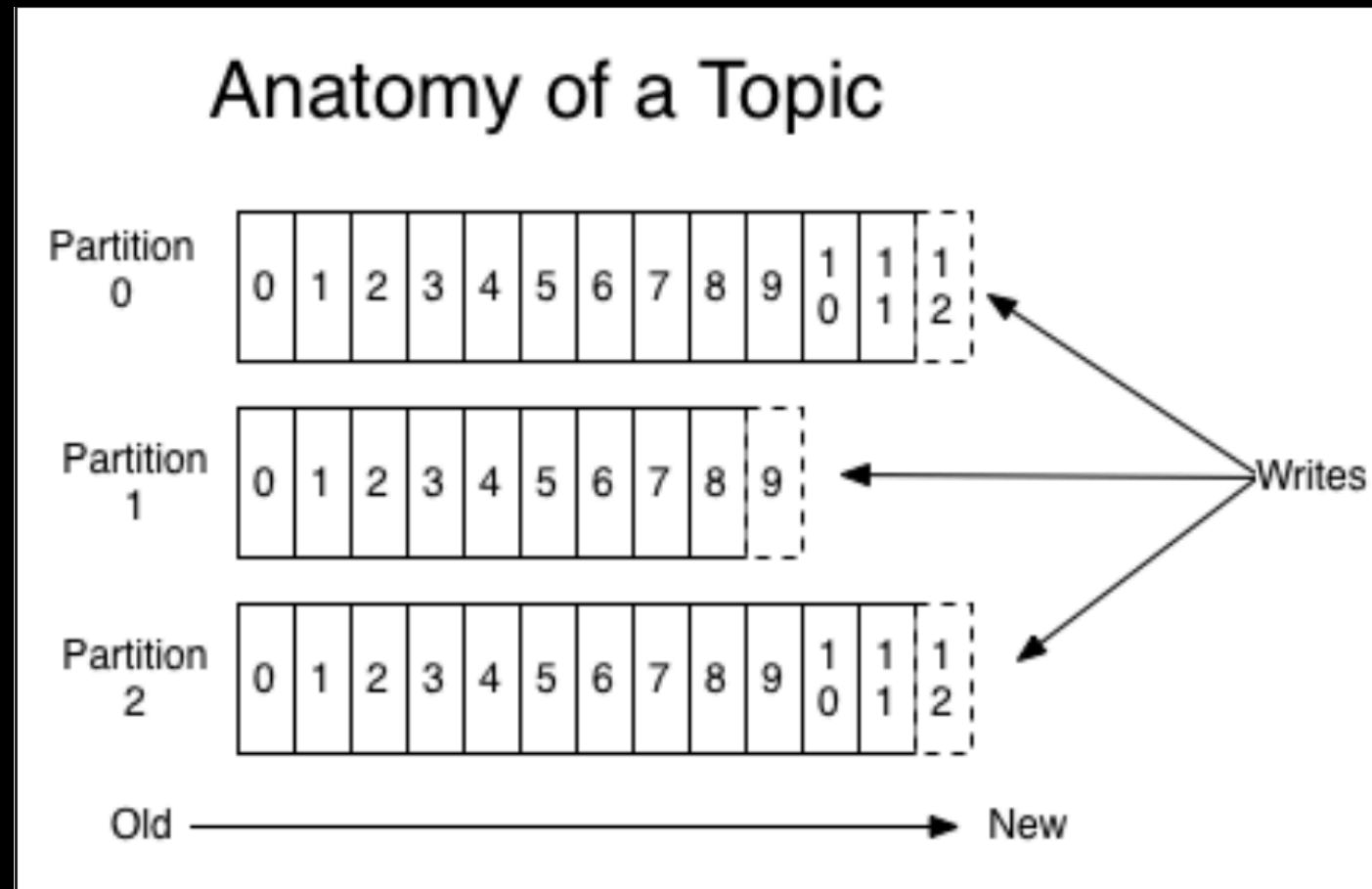
- Limitation of RAM, Disk, CPU core
- Central repository
- Maintain code base & test every time
- Adding new sources to our applications.



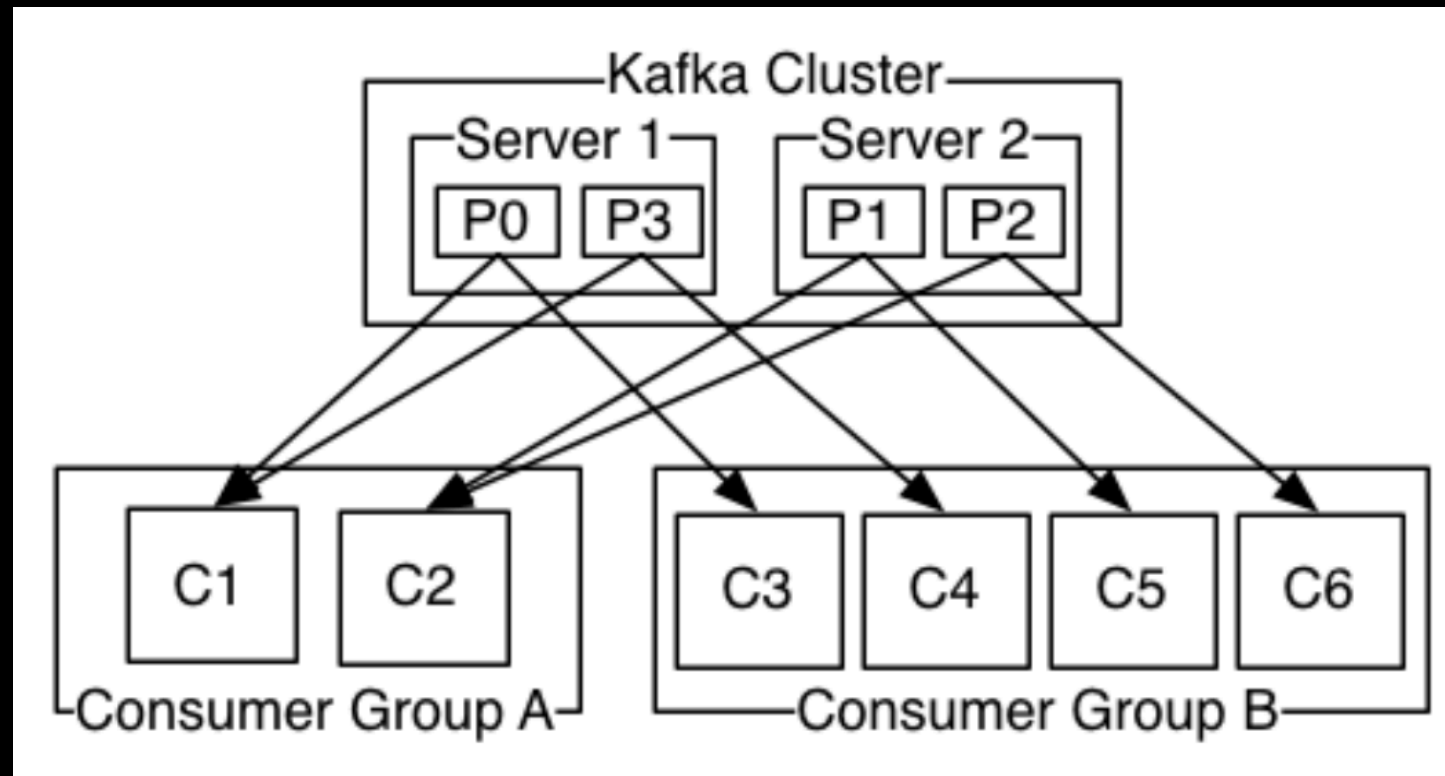
CORE CONCEPTS



CORE CONCEPTS



CONSUMER GROUPS



- Load balanced across all consumers in a groups

LOG COMPACTION

- How to recover yourself when application crashes?
- Full snapshot of the final value for every key
- Commit Log

ROLE OF ZOOKEEPER

- Actually key value store
- Heartbeats requests and save the state or health
- If some brokers fail Zookeeper comes to elect new.
- Following replications
- Default storage engine for consumer offset
- Rebalance the replicas adding new one

CRITICAL CONFIG

- Batch size -> high throughput, higher latency
- Linger.ms -> time based batch, higher latency
- Compression.type
- Acks (default -1)

WHAT IS IMPORTANT

- Monitoring
- Scale Kafka & Zookeeper together

MONITORING TOOLS

- Sematext
- Kafka Monitor
- Grafana - Prometheus

DISADVANTAGE

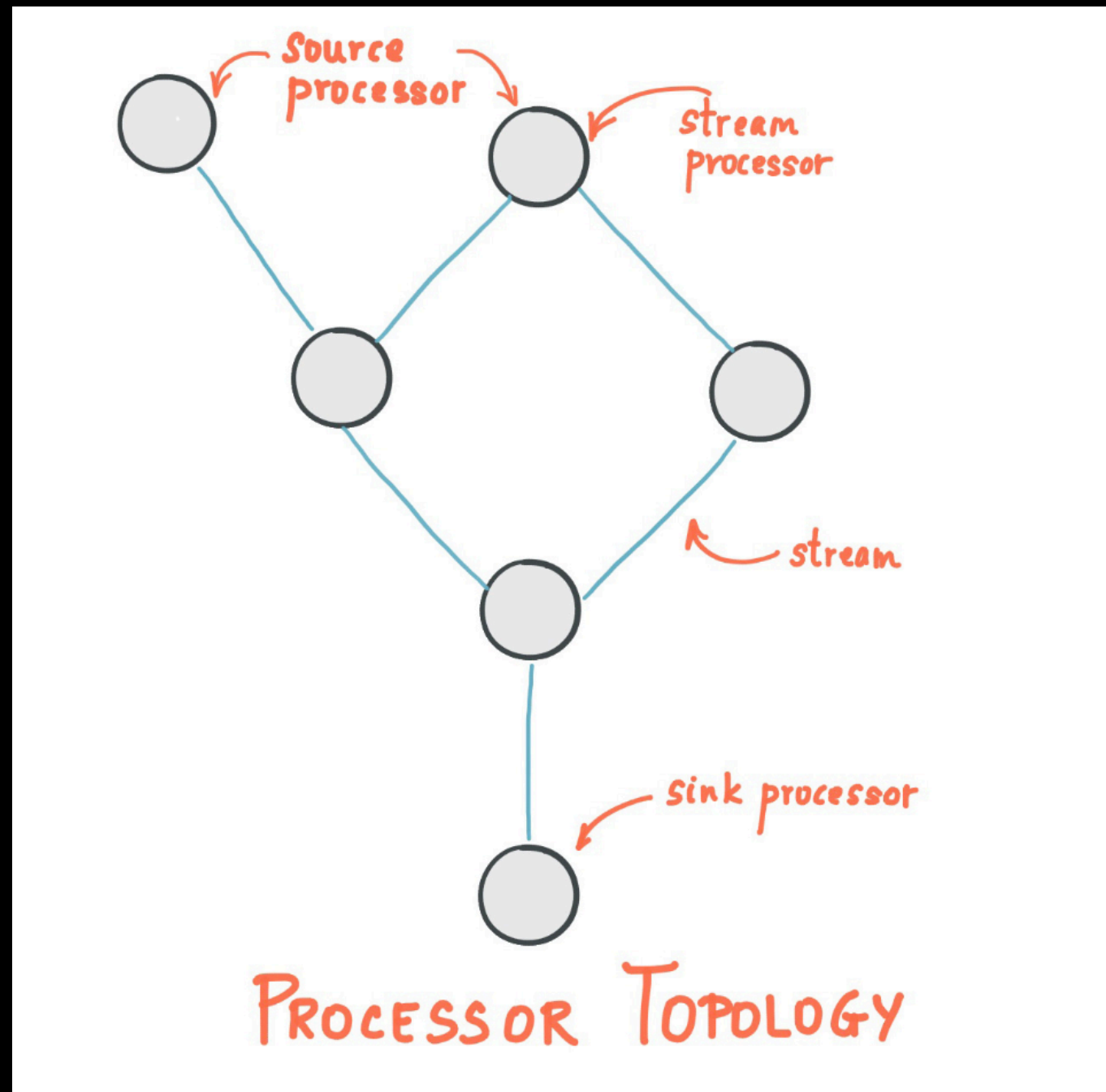
- Zookeeper Dependency
- Not support wildcard topic selection
- Reduces performance when message size increases
 - Compressing & Decompressing
- Insufficient monitoring tools

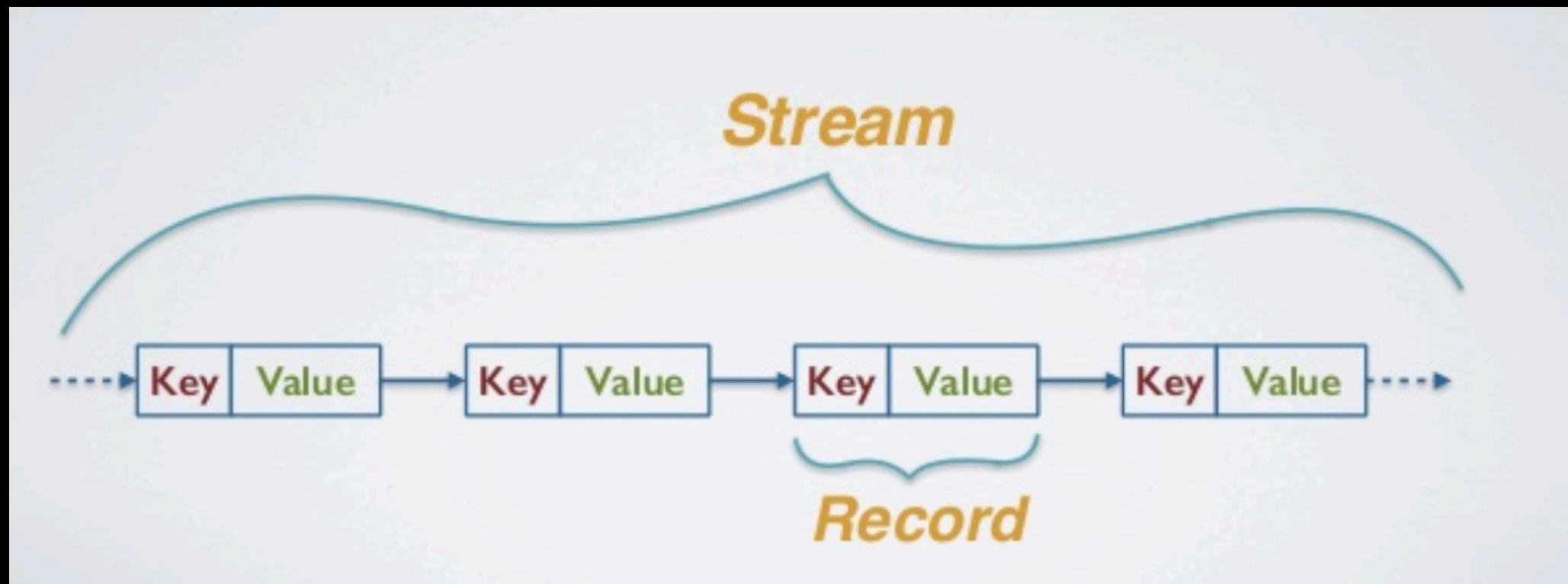
KAFKA STREAMS

KAFKA STREAMS

- Library for data processing and transformation
- There is no batch operation
- Scalable

CORE CONCEPTS

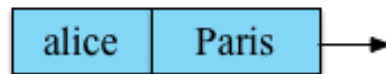




KSTREAMS & KTABLE

Table

Stream



older data —————> newer data

STATELESS & STATEFUL

- Stateful operations need another result or external data
 - Count, sum, max
- Stateless operations don't need any external data
 - multiplication number with a const or itself

INTERNAL TOPICS

- Control on the Kafka Streams
- Reporting
 - When key is changed (Transformation)
- Change Log
 - When run a aggregation such as count, sum, avg

COMPANIES

- LinkedIn
- Airbnb
- Netflix
- Spotify
- Uber
- Skyscanner
- Trivago
- Datadog