APACHE
kafka®
A distributed streaming platform

Streaming Platform?

pub/**sub** similar to message queues

Store streams of records in a **fault-tolerant** & **durable** way

Process streams of records as they **occur**

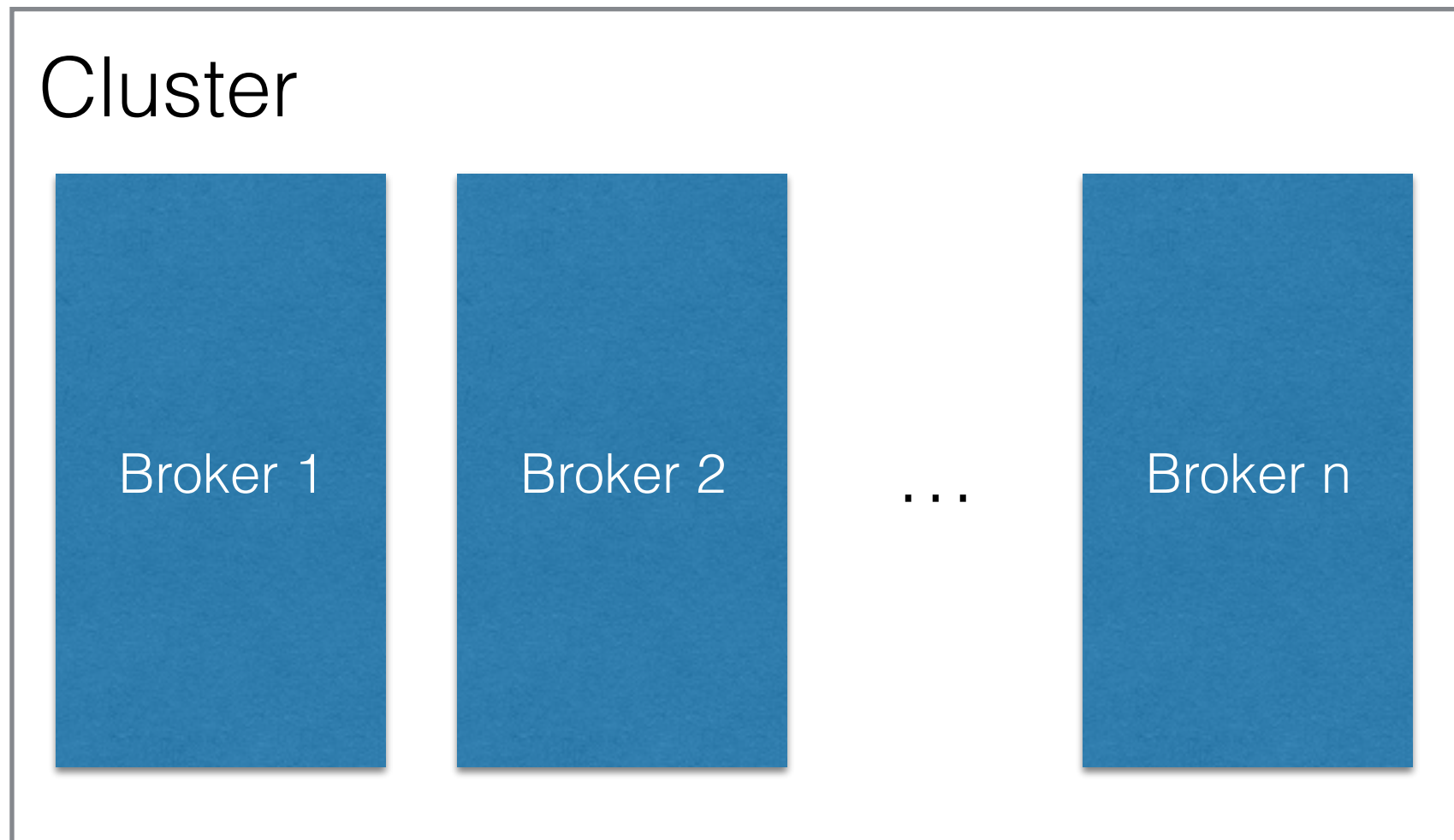# Typical Use Cases?

Messaging

Website Activity Tracking

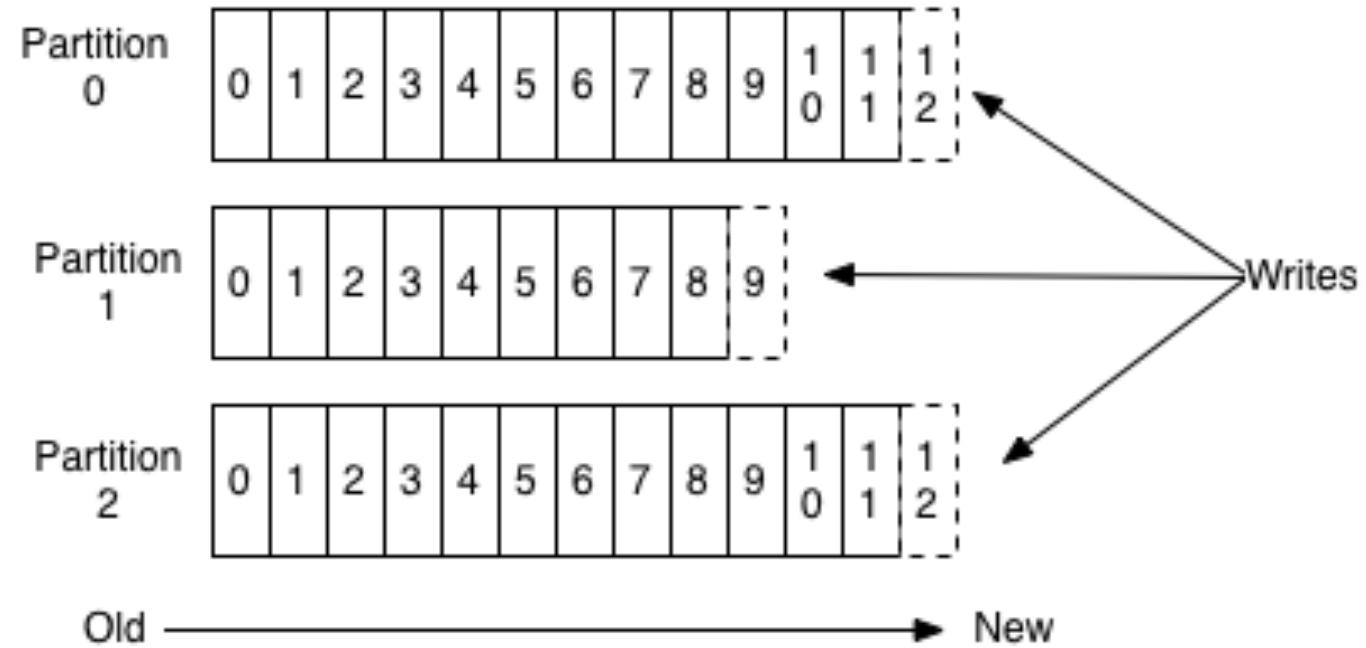# Log Aggregation

Stream Processing

Event Sourcing

Core Concepts

Cluster & Broker

Cluster

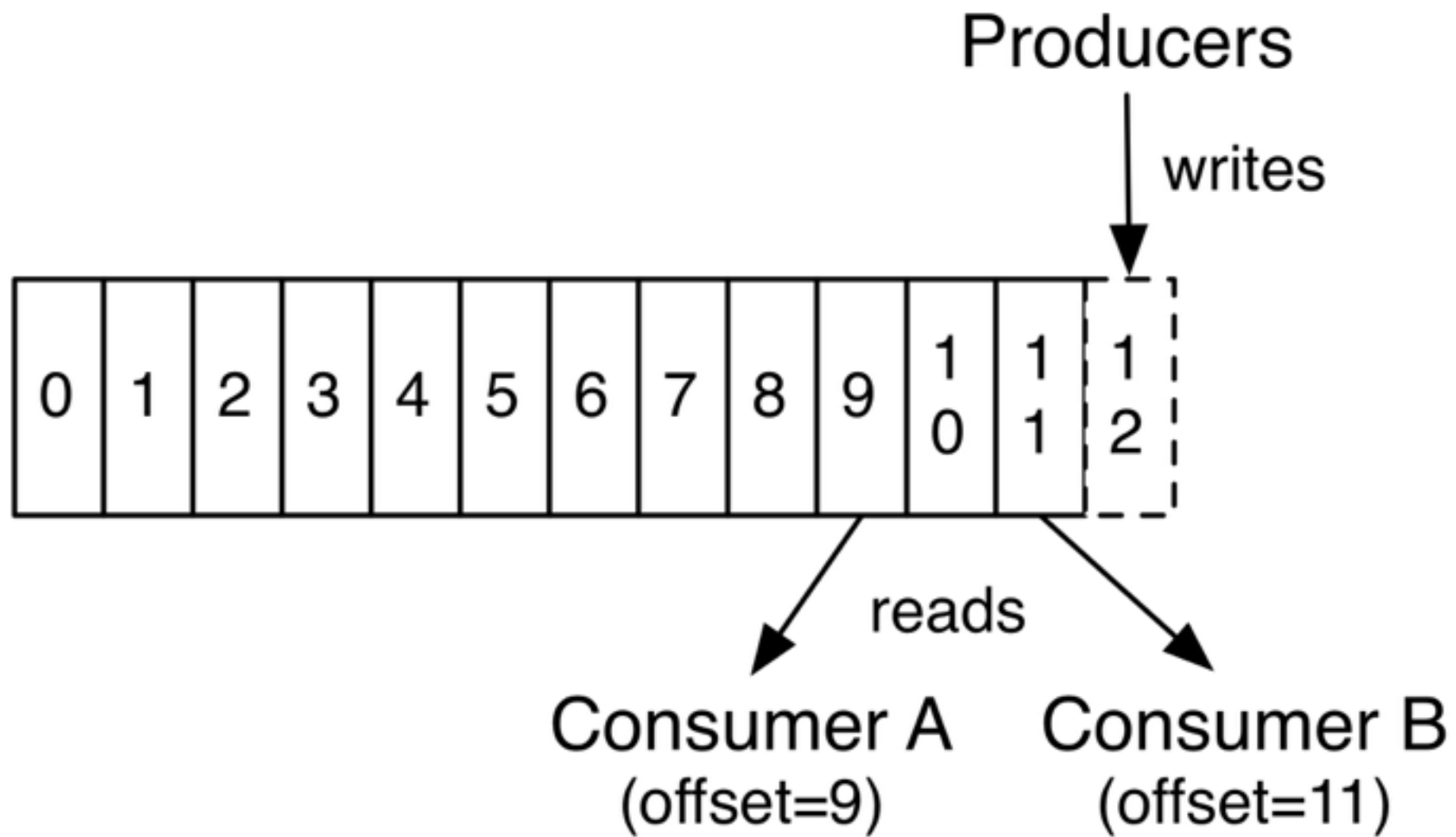Broker 1   Broker 2   …   Broker n

Topic

Anatomy of a Topic

Partition

Record
key;**value**;timestamp

Retention

- By time (e.g. max 2 days)

- By size (log may not grow beyond X GB)

- Infinite (totally fine)

Log compaction

```
1    123 => bill@microsoft.com

2                .

3                .

4                .

5    123 => bill@gatesfoundation.org

6                .

7                .

8                .

9    123 => bill@gmail.com
```
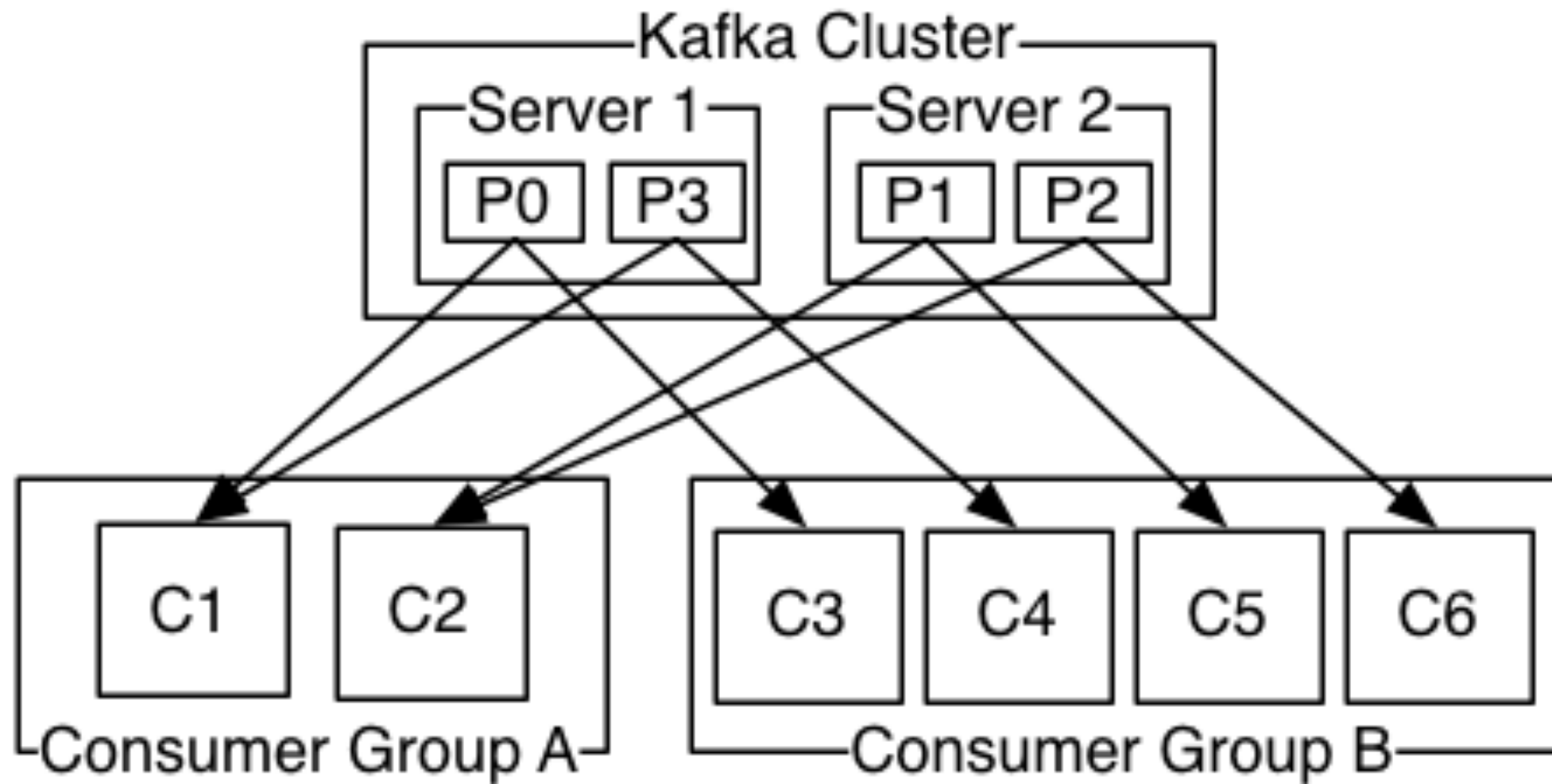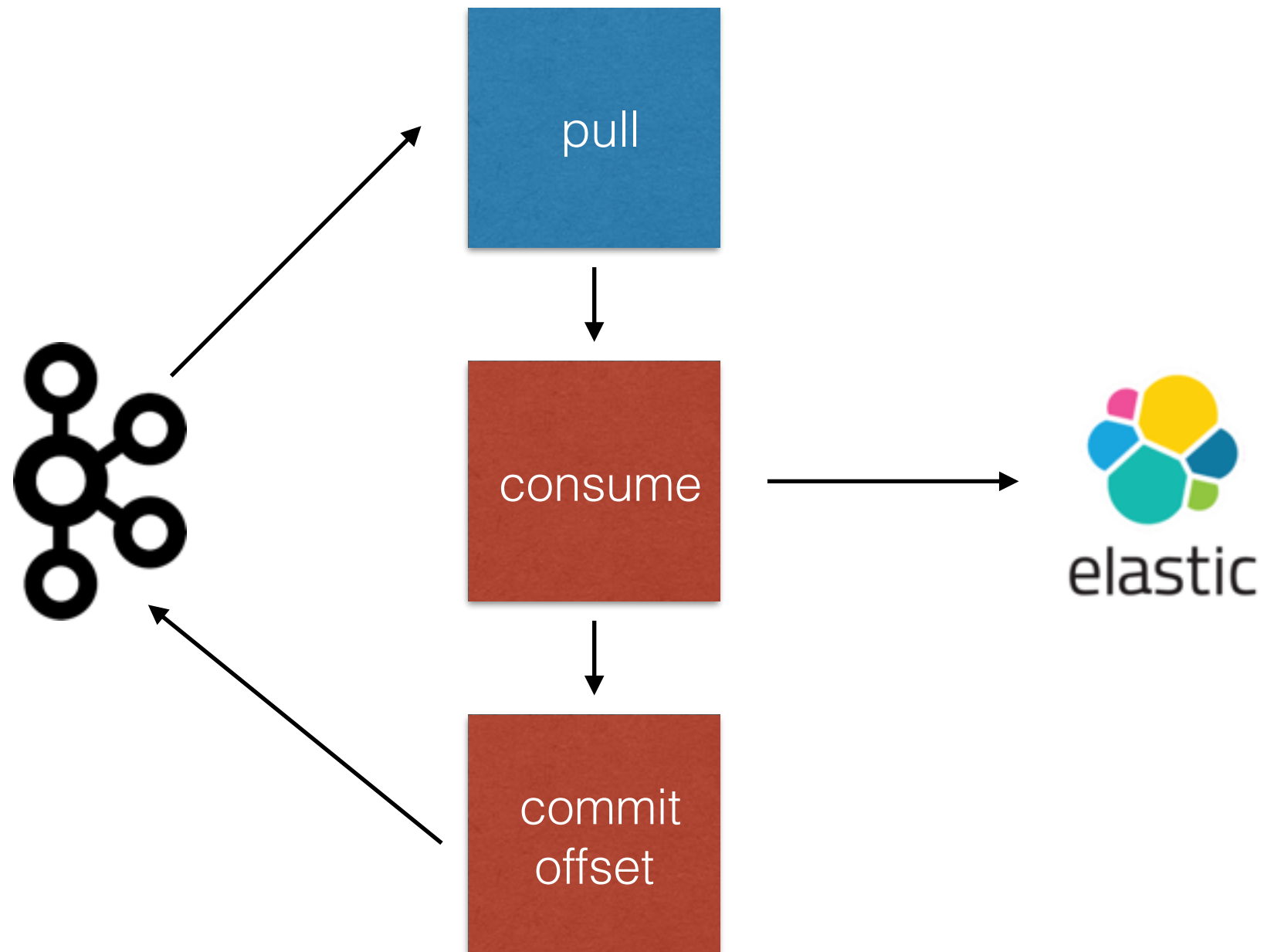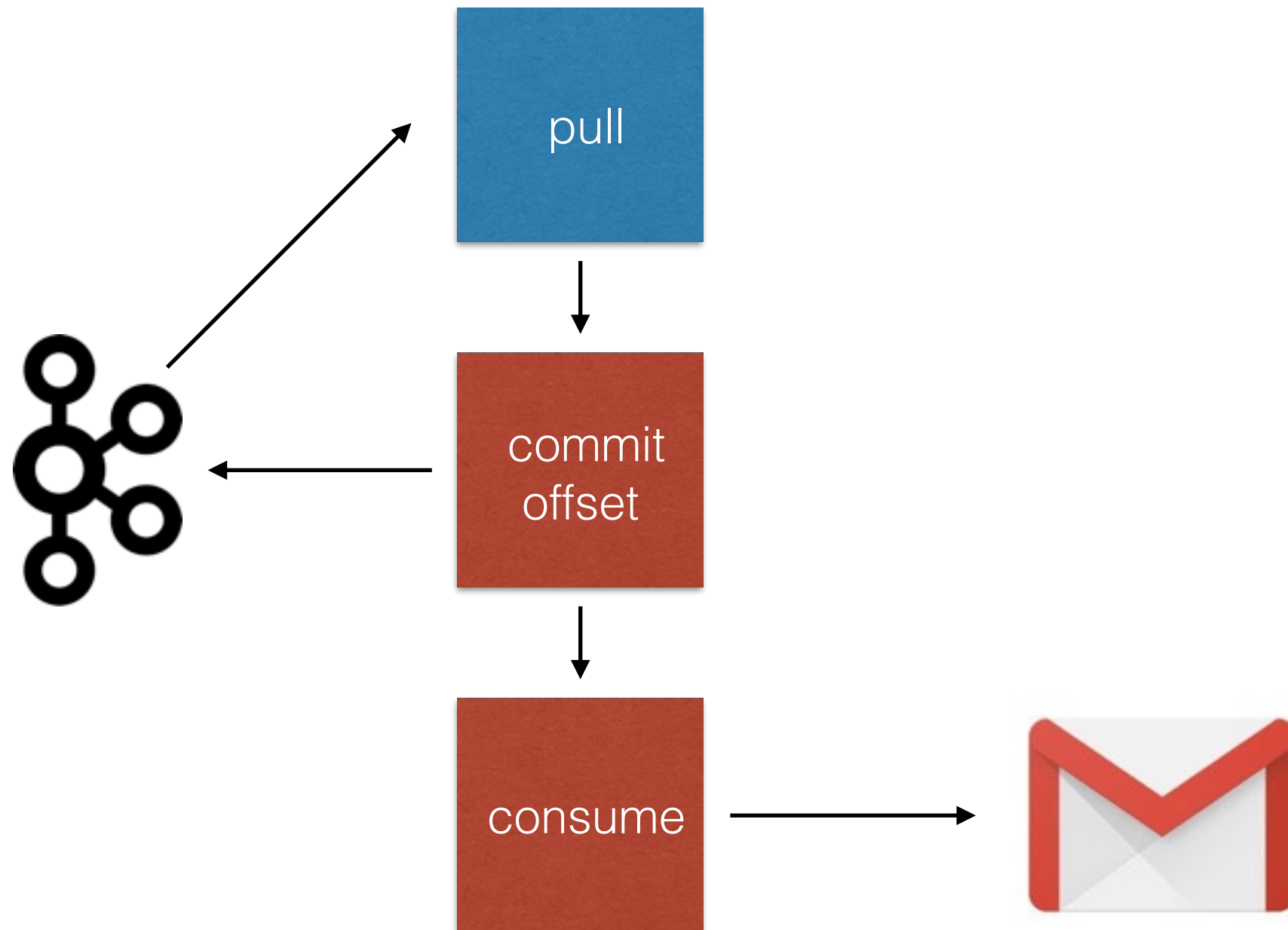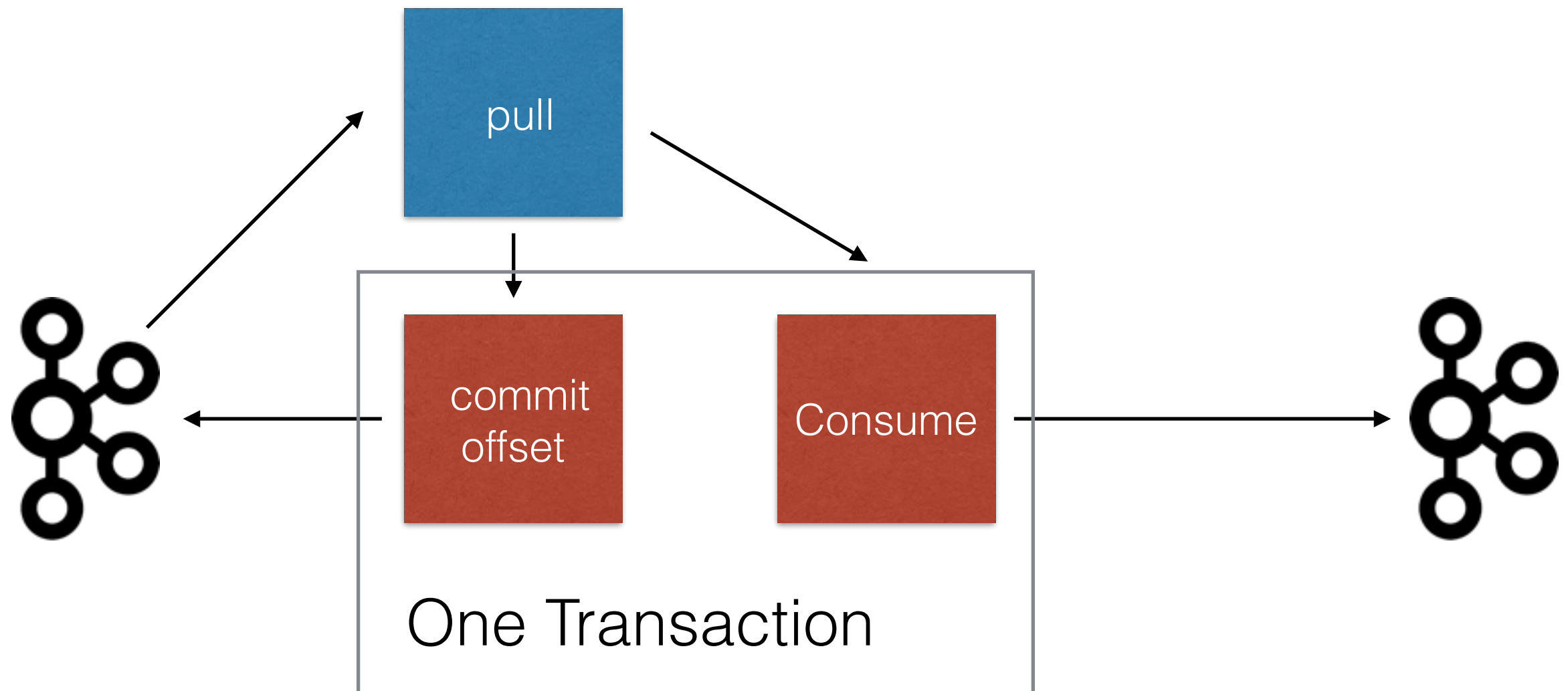
Quotas

- Network bandwidth

- Request rate

Producer

Consumer

At least once

At most once

Exactly once

Let's code!
https://github.com/Sebruck/kafka-talk