

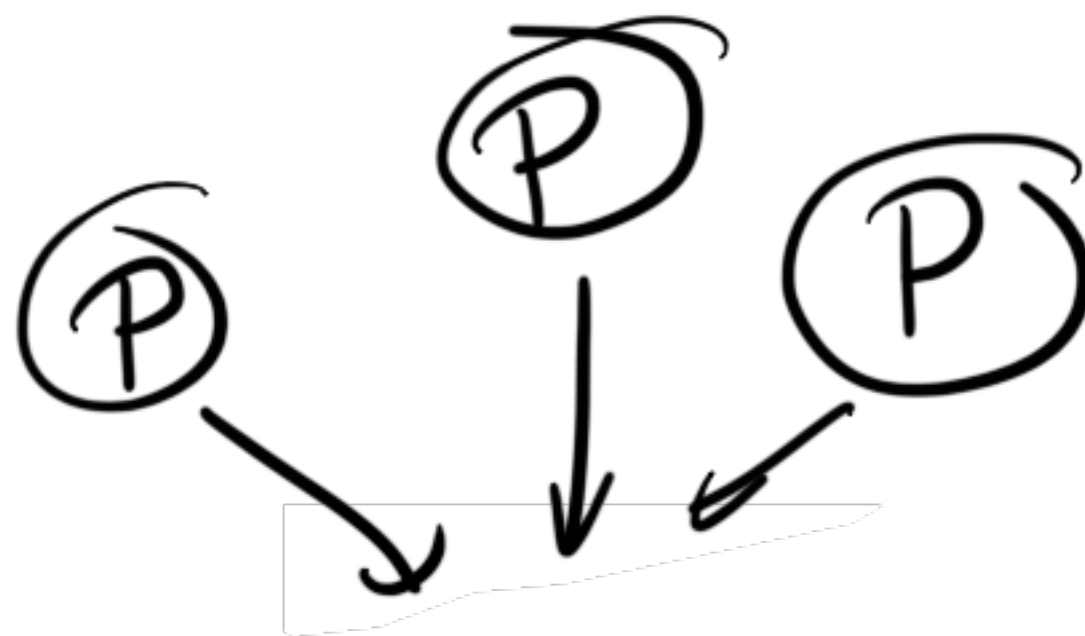
Kafka

A little introduction

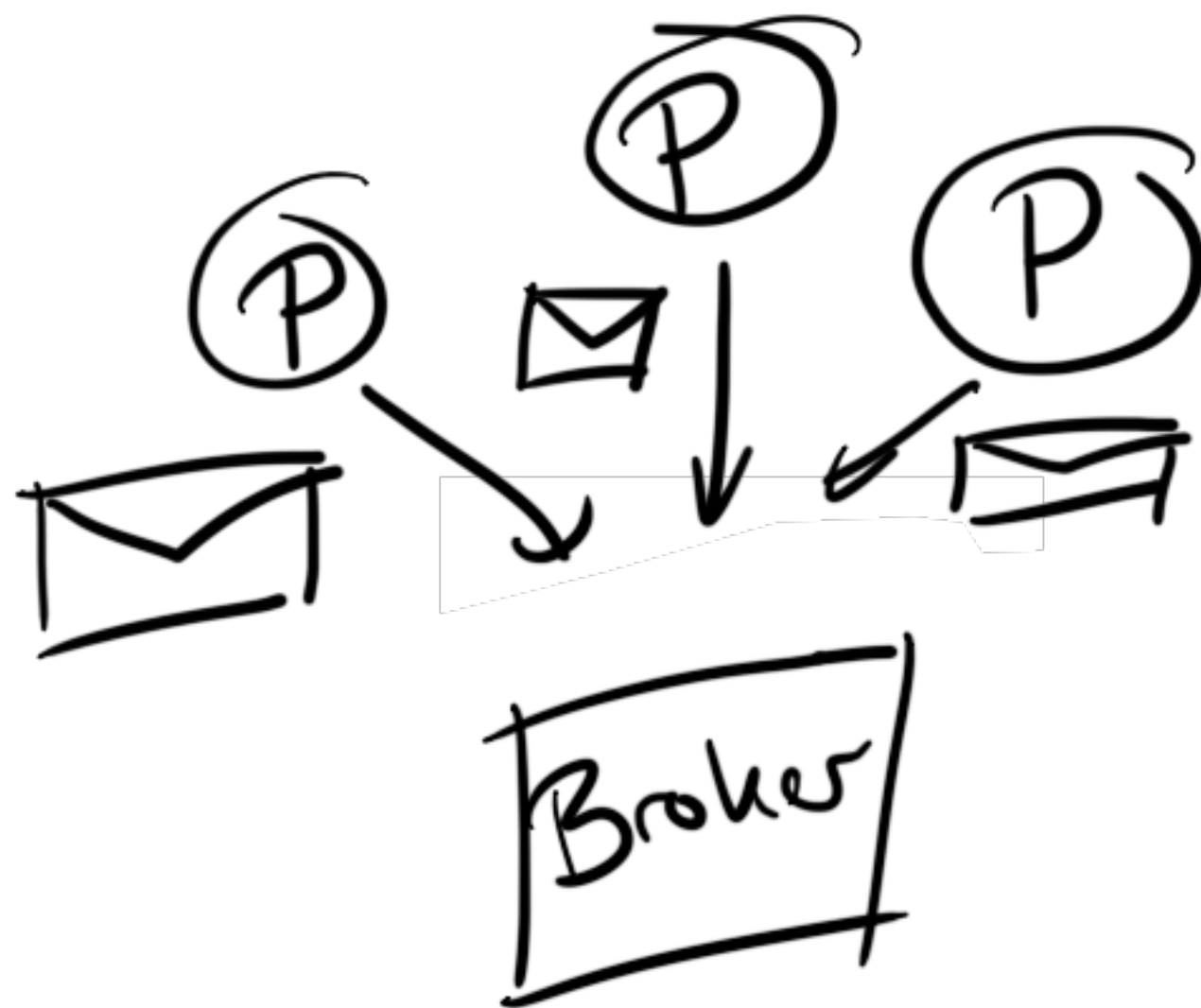


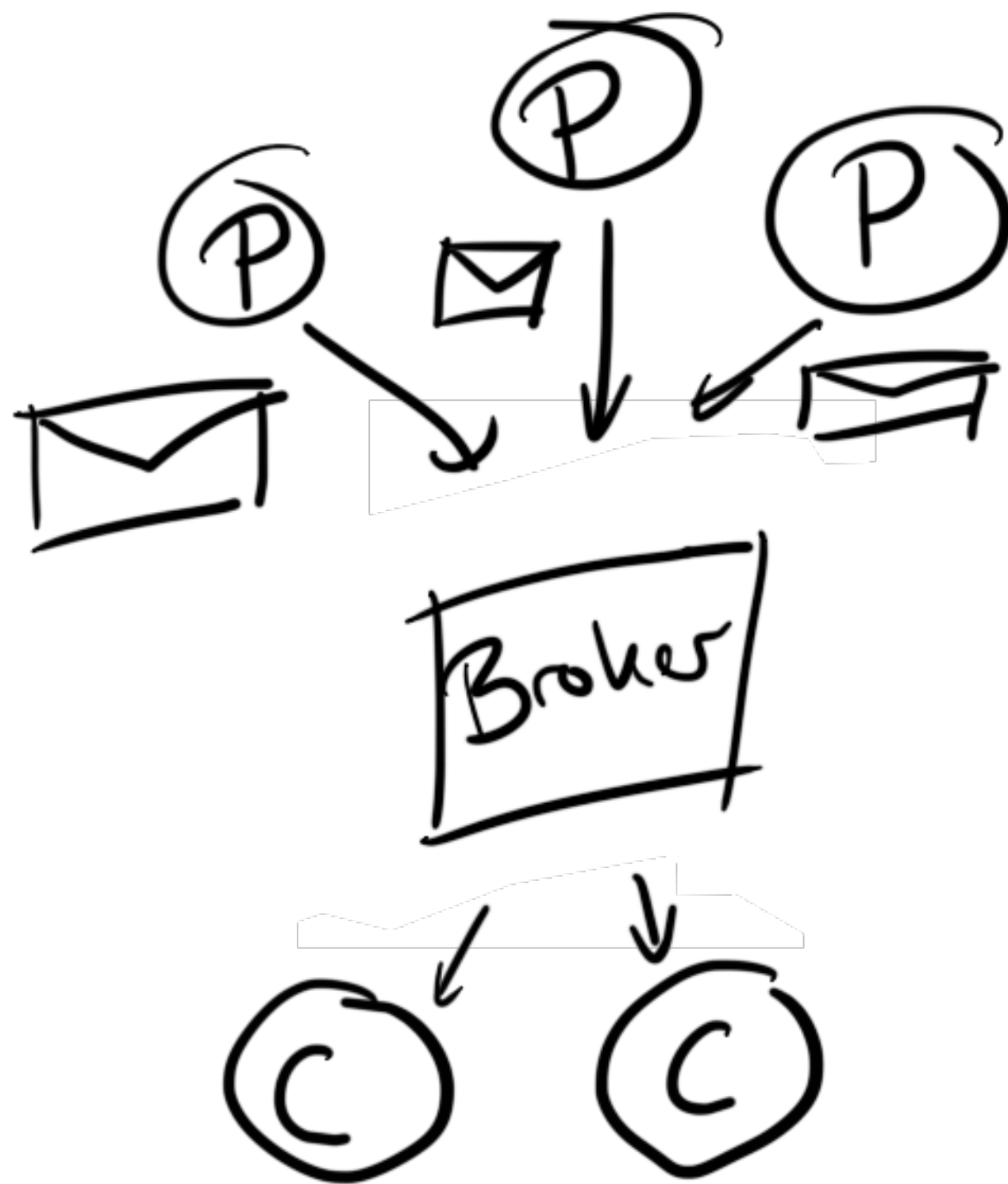
Pub-Sub Messaging System

Broker

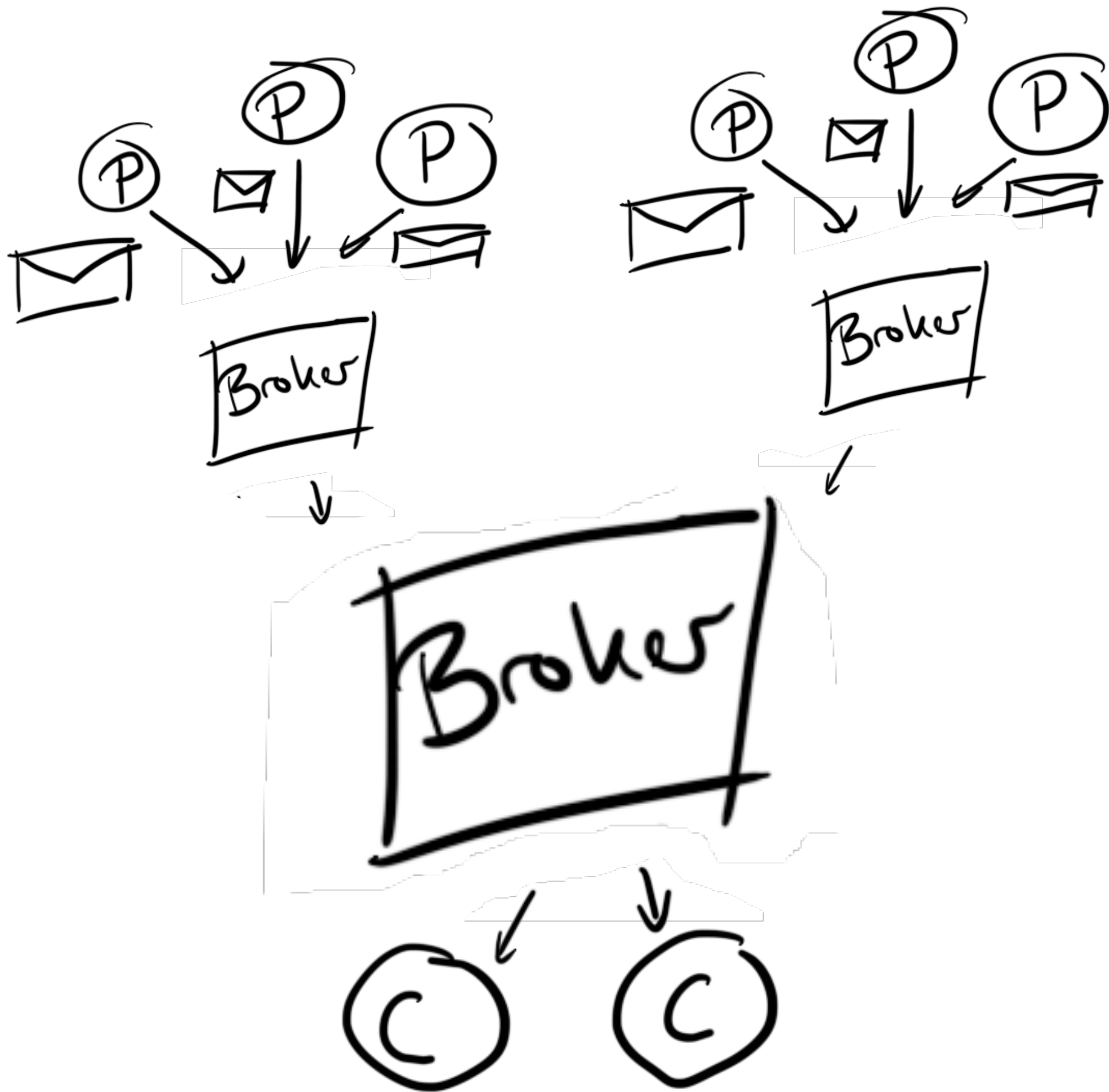


Broker



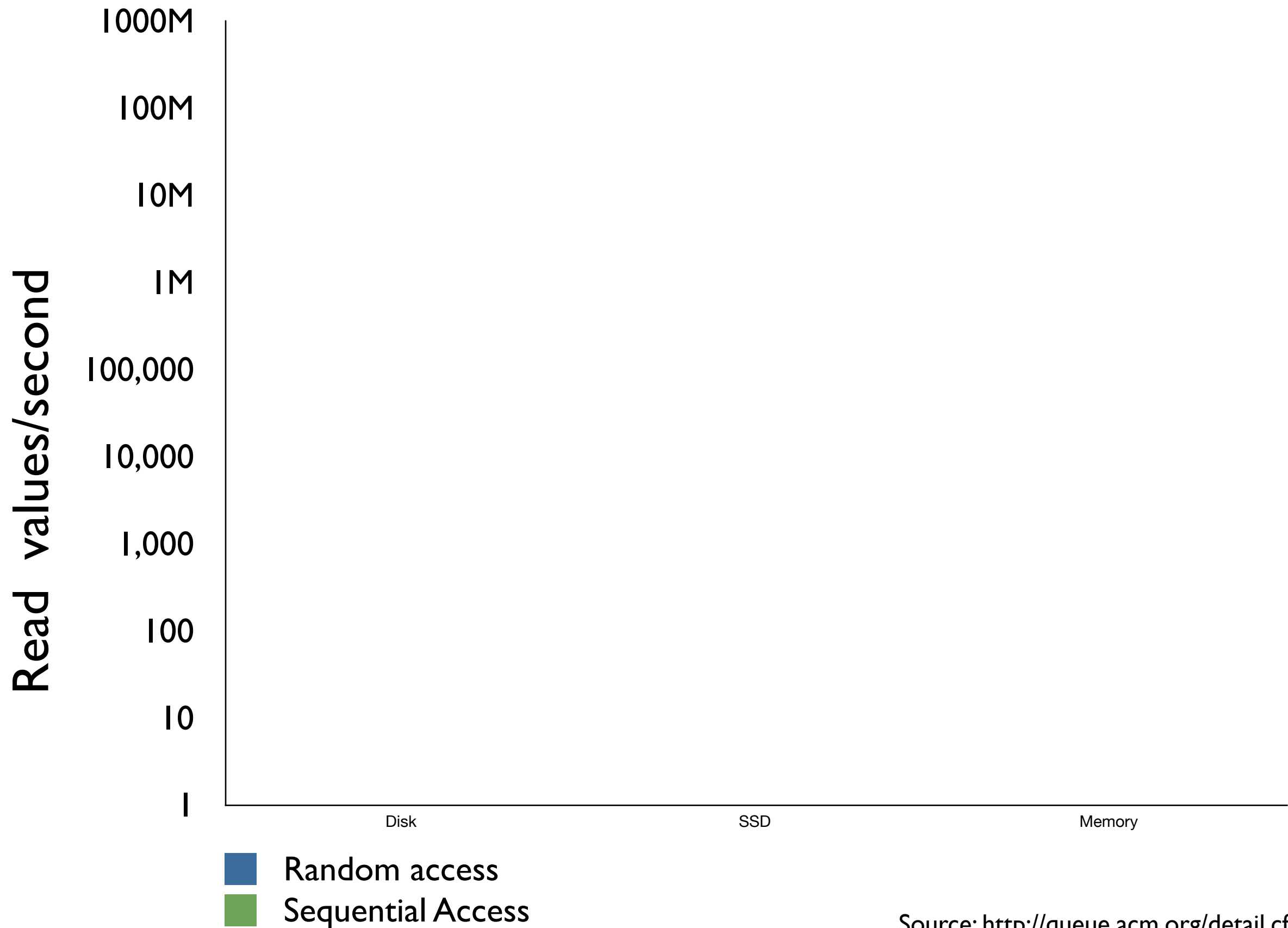


Distributed

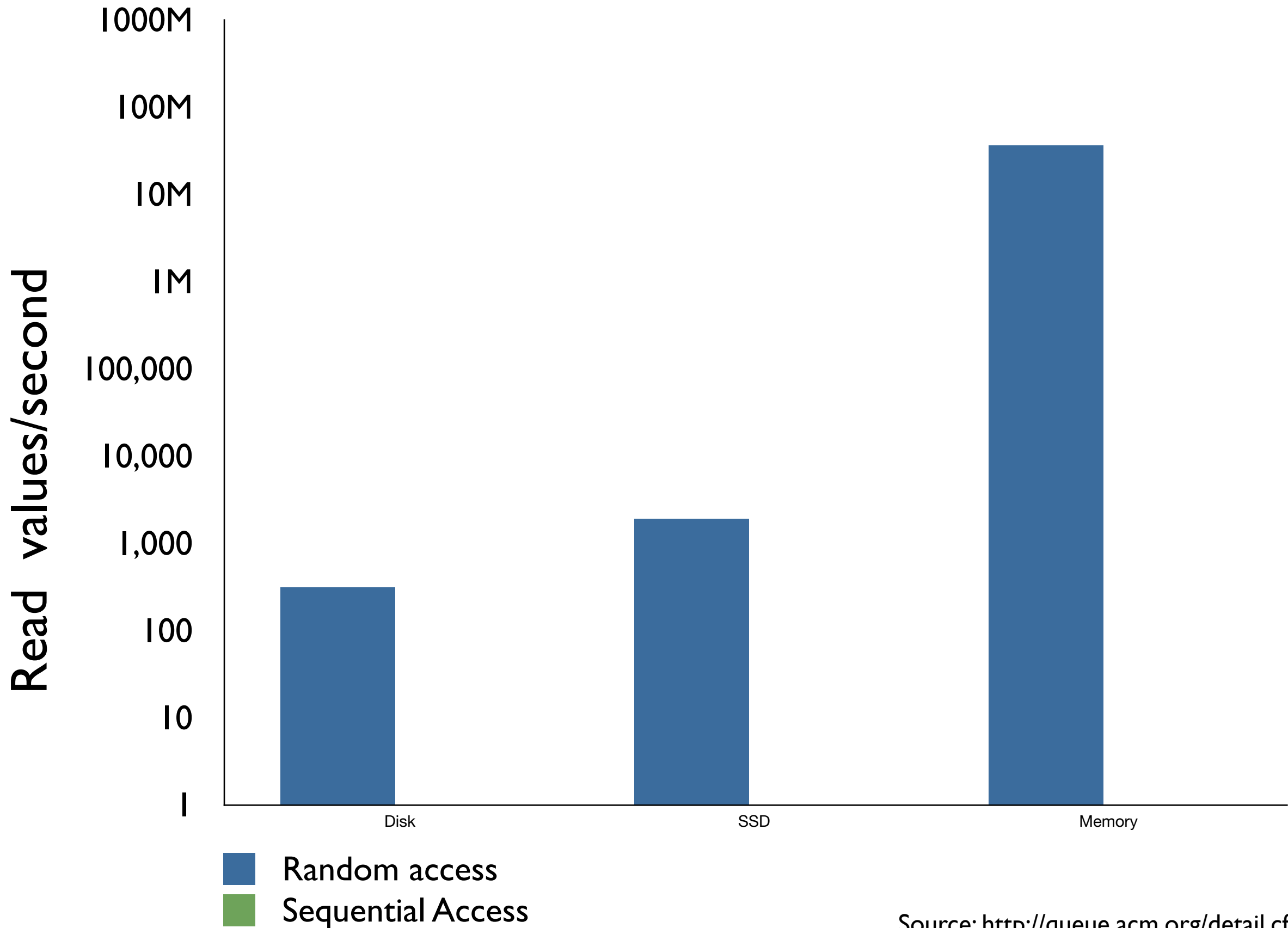


Performance

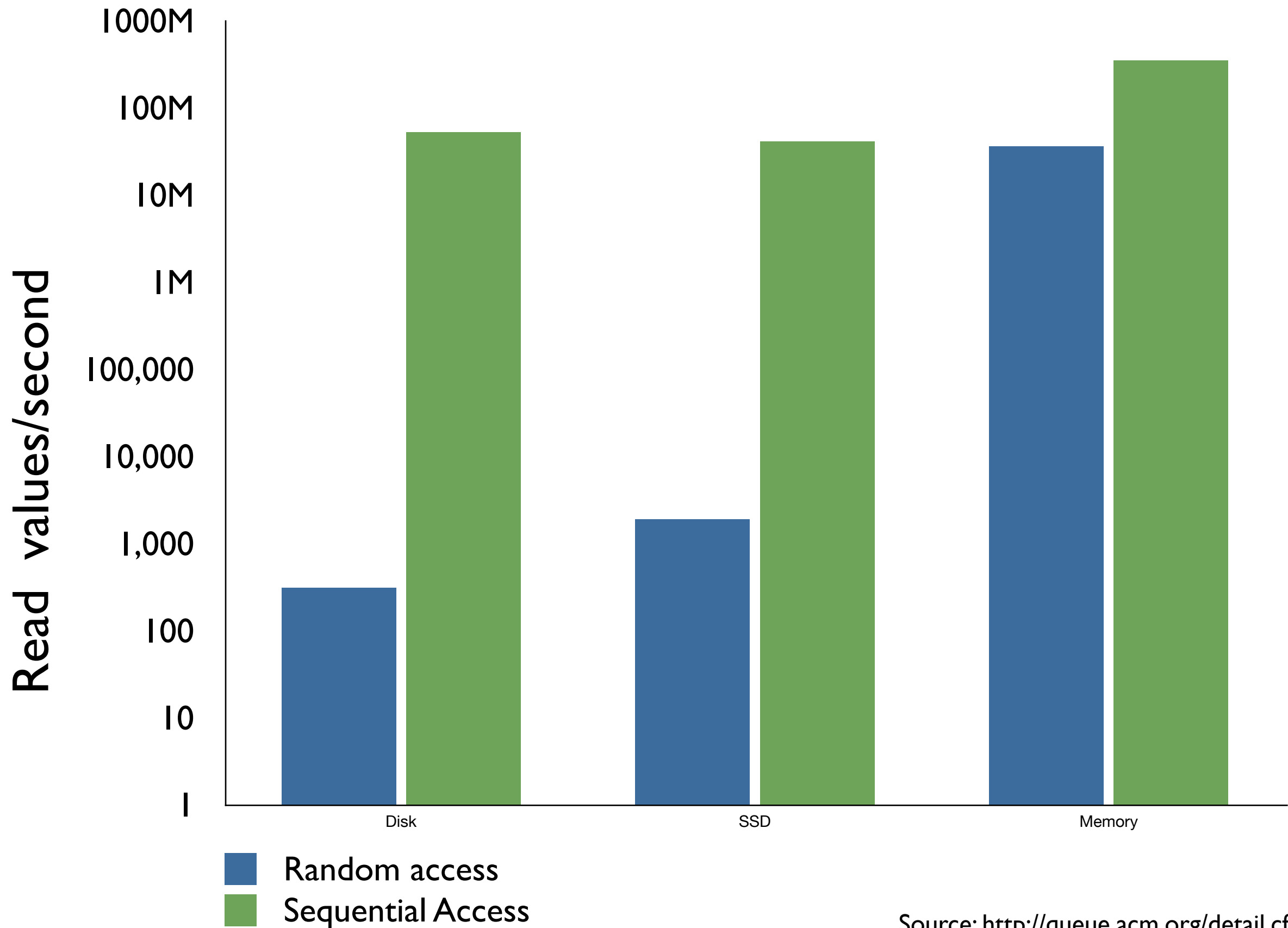
Disk/Memory Performance



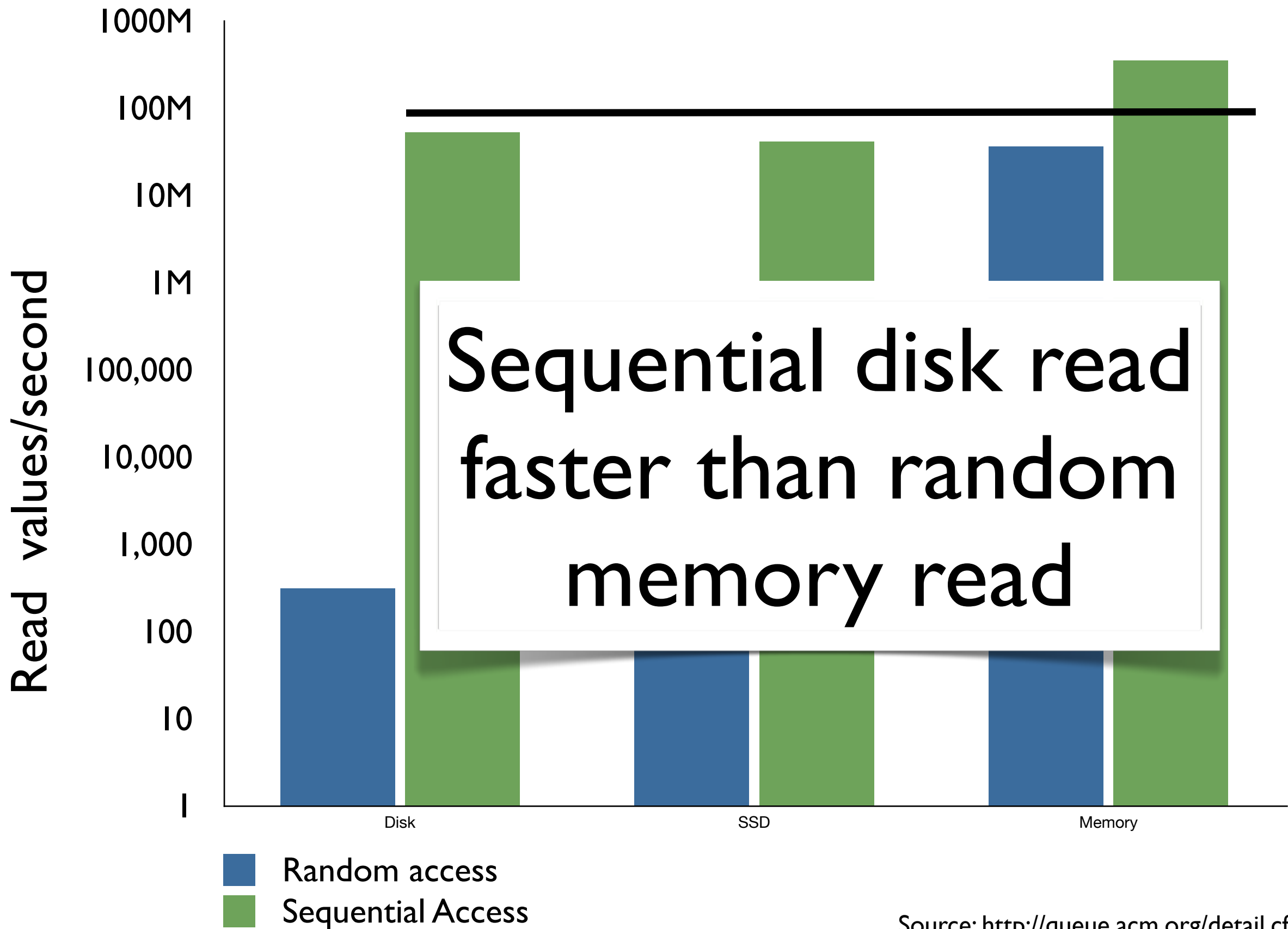
Disk/Memory Performance



Disk/Memory Performance



Disk/Memory Performance



Persistent

Topic

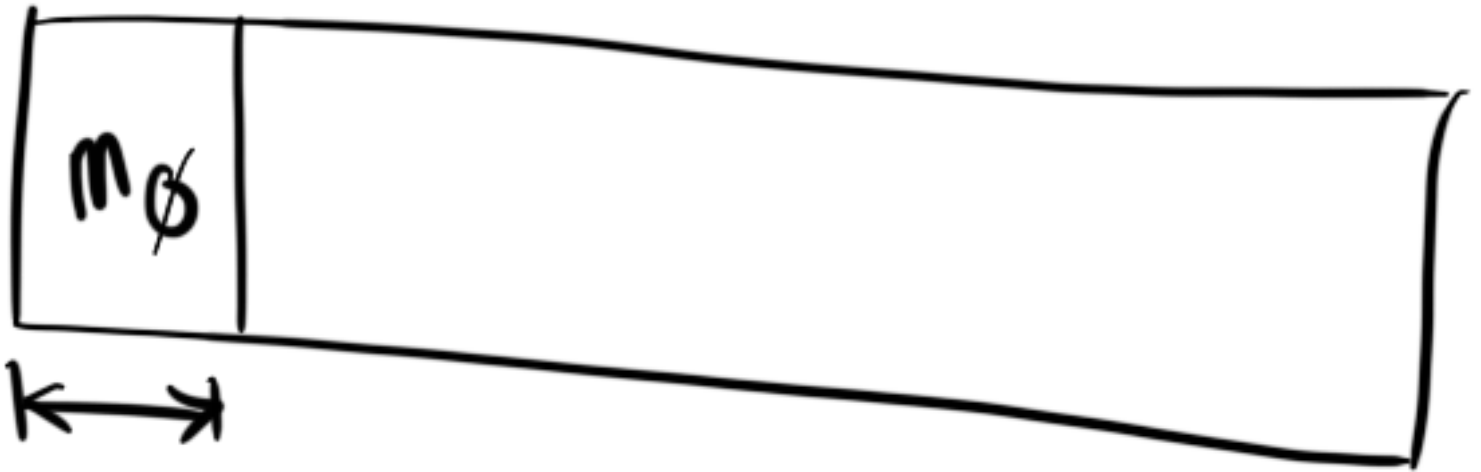
 → Topic

/var/log/topic/6000000.hatna



→ Topic

100
bytes



message



Length

Magic Value

Checksum

Payload

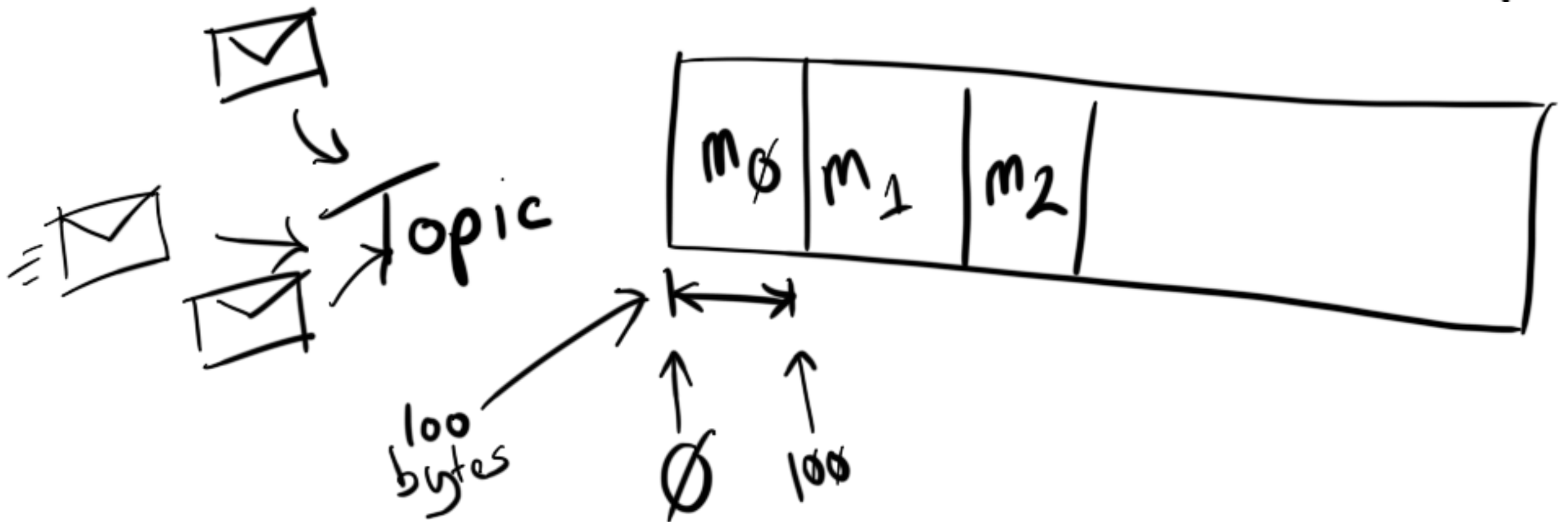
4 bytes

1 byte

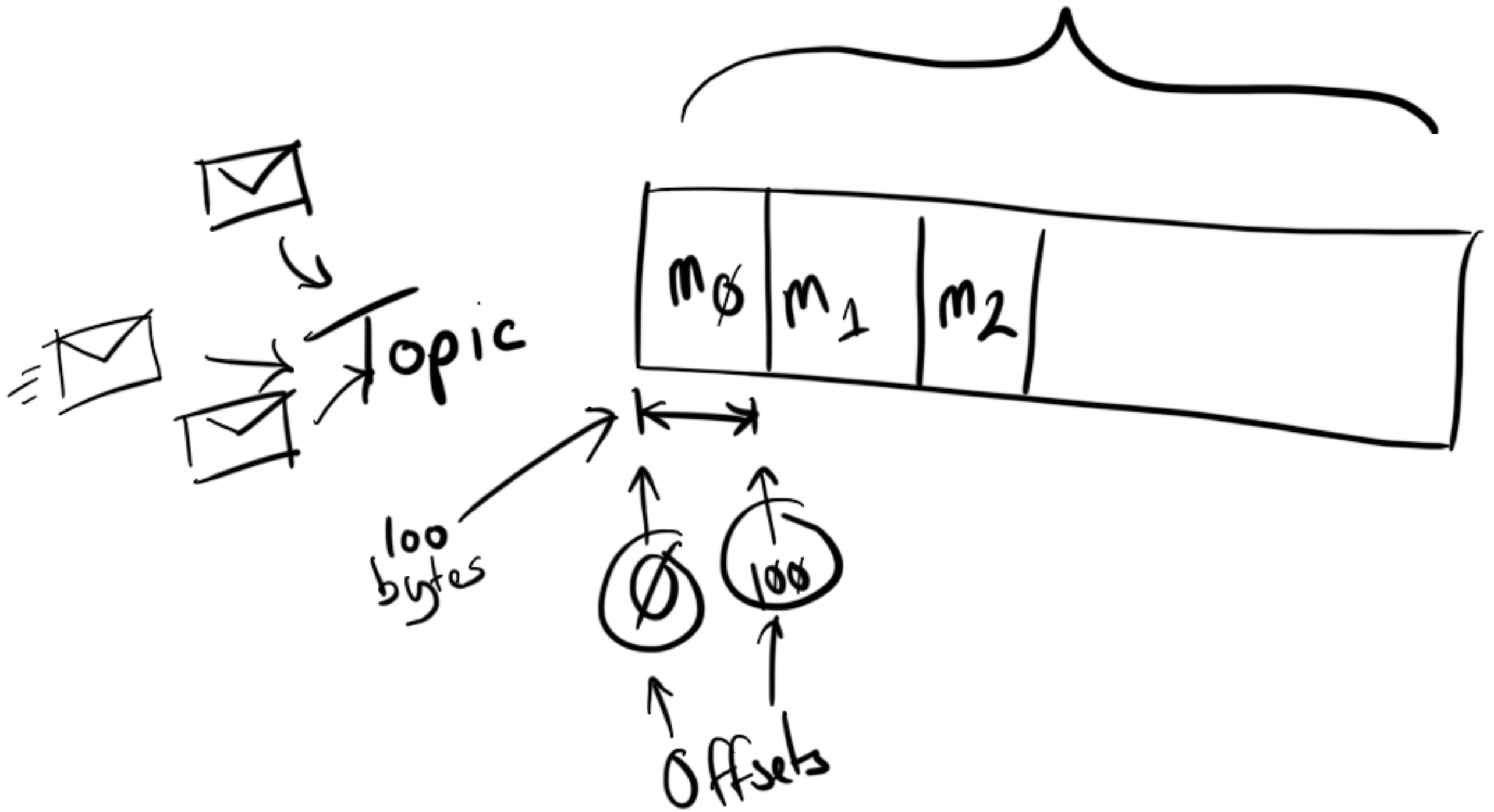
4 bytes

n bytes

/var/log/topic/6000000.kafka



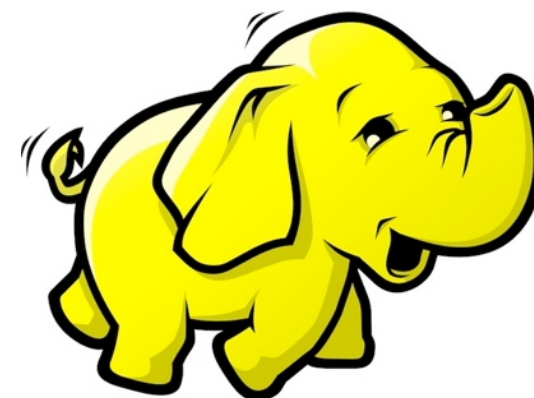
/var/log/topic/0000000.kafka





232 →

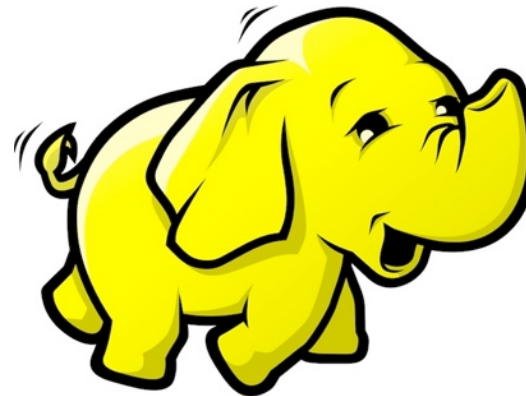
Consumer



Token

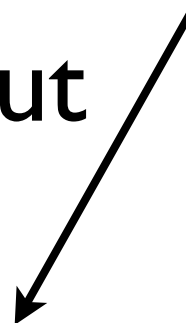
Offset: 0
Broker: kafka.local
Topic: Testing

Input



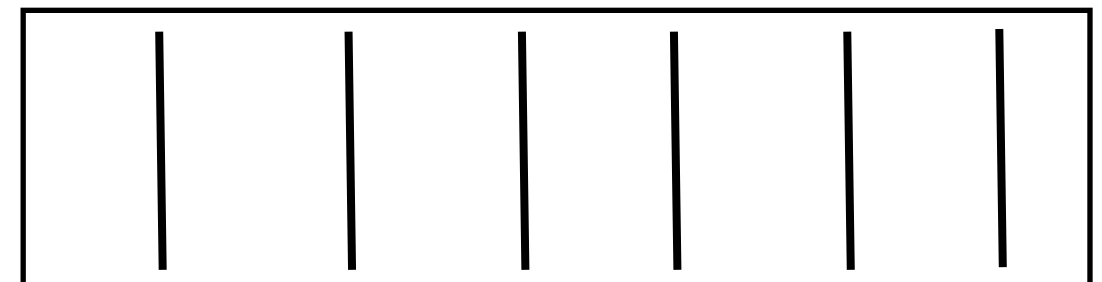
MR Job

Output



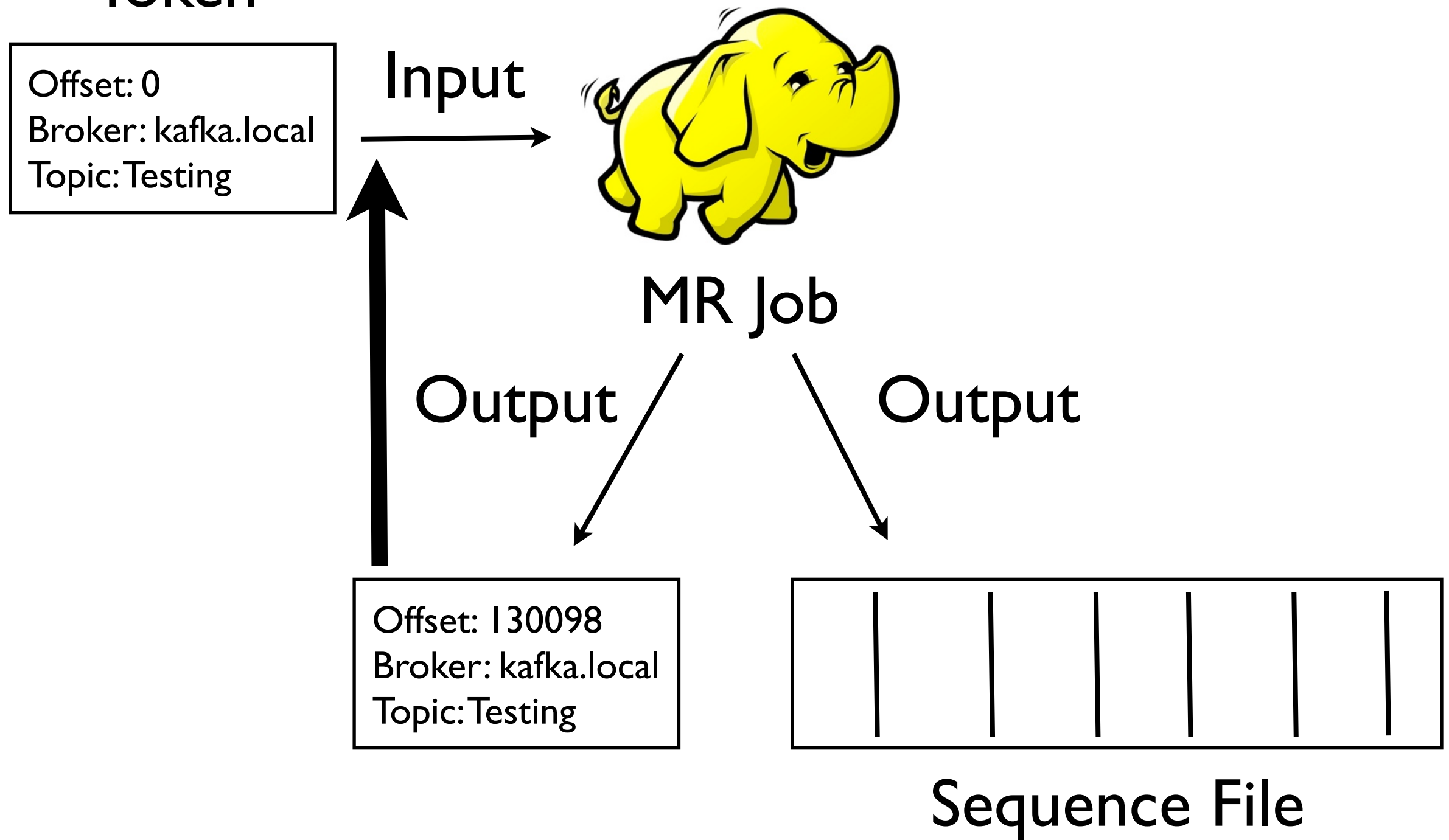
Offset: 130098
Broker: kafka.local
Topic: Testing

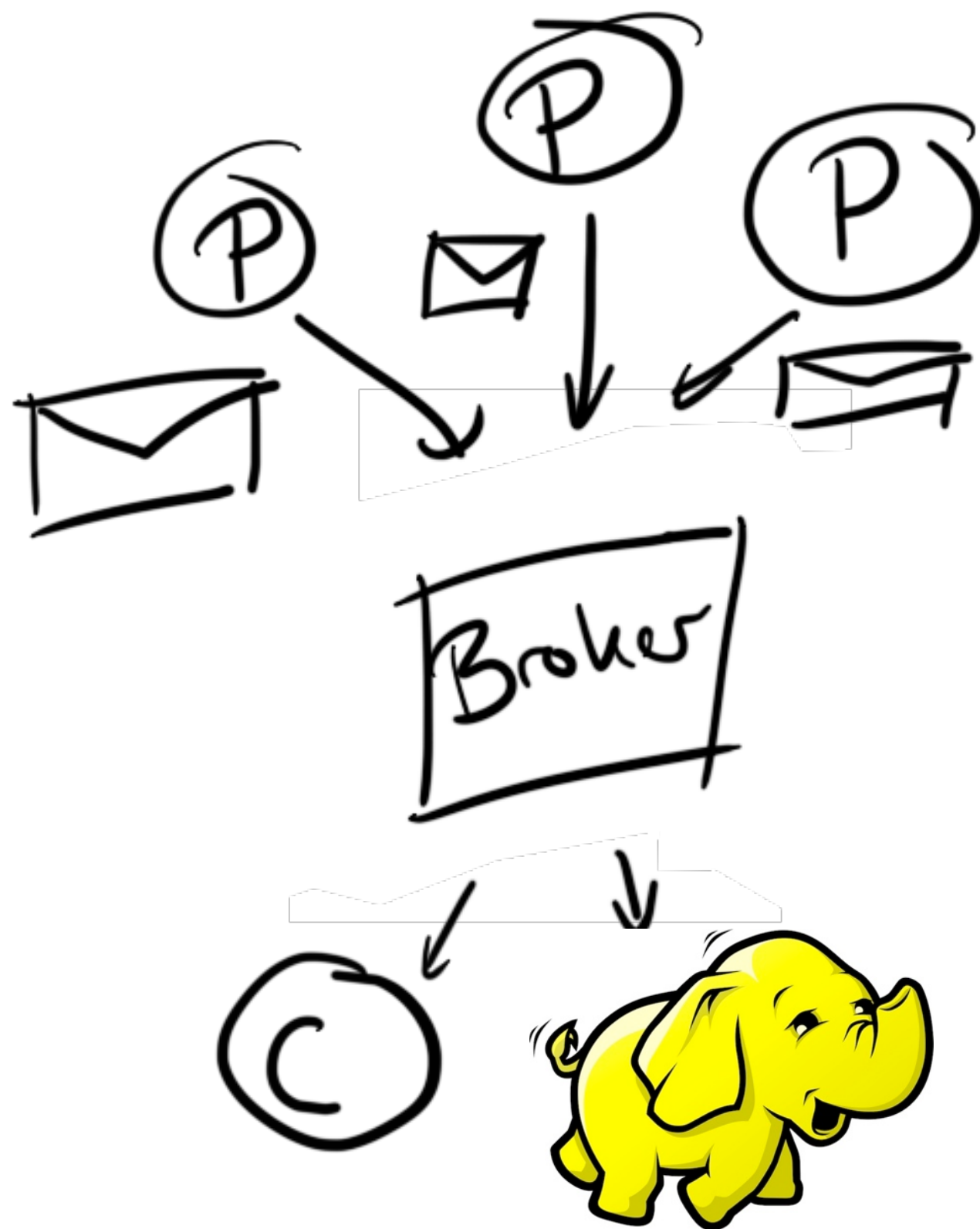
Output



Sequence File

Token





Useful Things

- <http://incubator.apache.org/kafka/>
- <https://github.com/pingles/clj-kafka>