ThoughtWorks®

The Best of Apache Kafka Architecture

- publish-subscribe messaging service
- distributed commit/write-ahead log

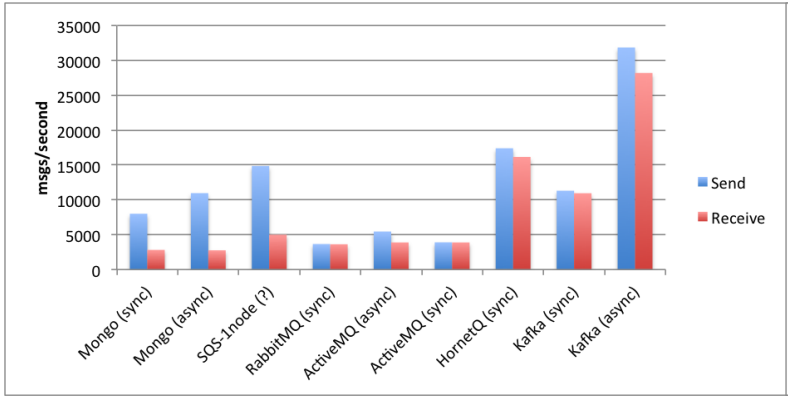
“producers produce, consumers consume, in large distributed reliable way -- real time”

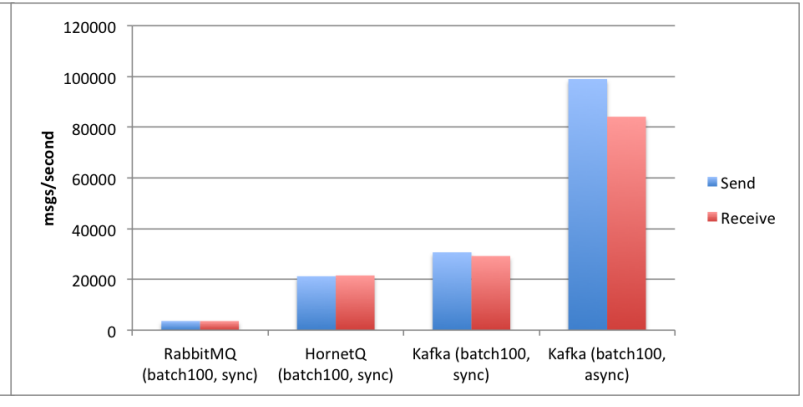
Why Kafka?

- DBs
- Logs
- Brokers
- HDFS

“For highly distributed messages, Kafka stands out.”

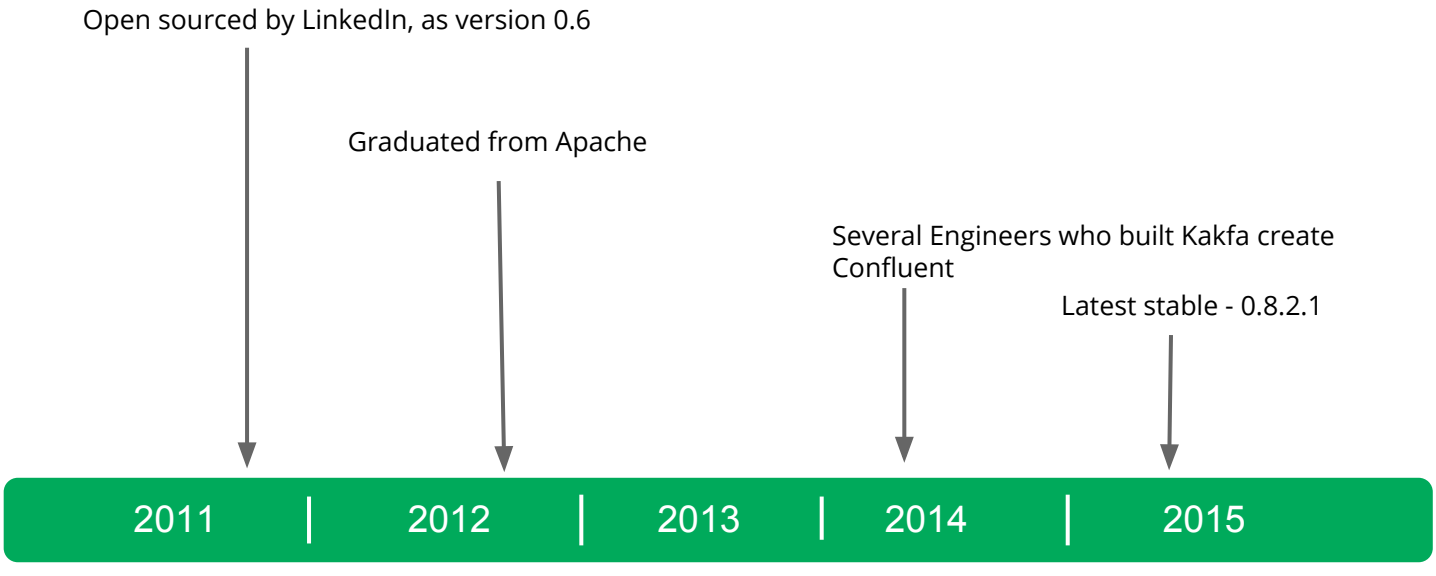
Kafka Vs _____





src: https://softwaremill.com/mqperf/

Timeline


Open sourced by LinkedIn, as version 0.6
Graduated from Apache
Several Engineers who built Kakfa create Confluent
Latest stable - 0.8.2.1
2011
2012
2013
2014
2015

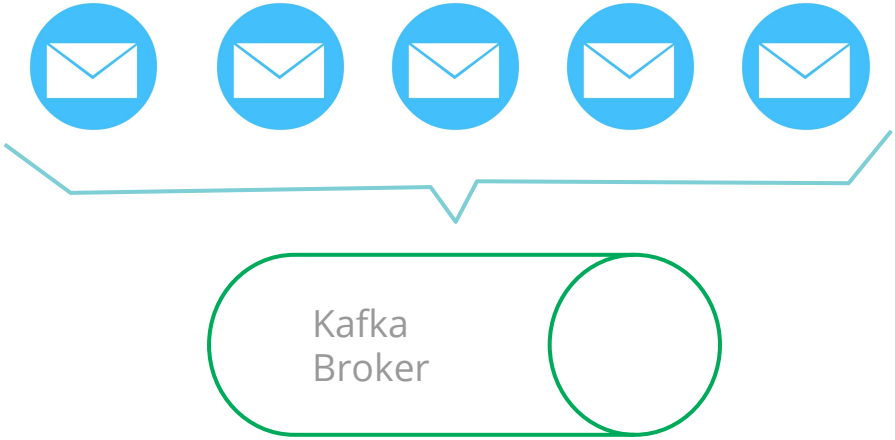
A Kafka Message

kafka.message.Message

Change requested:KAFKA-2511

Producers - push

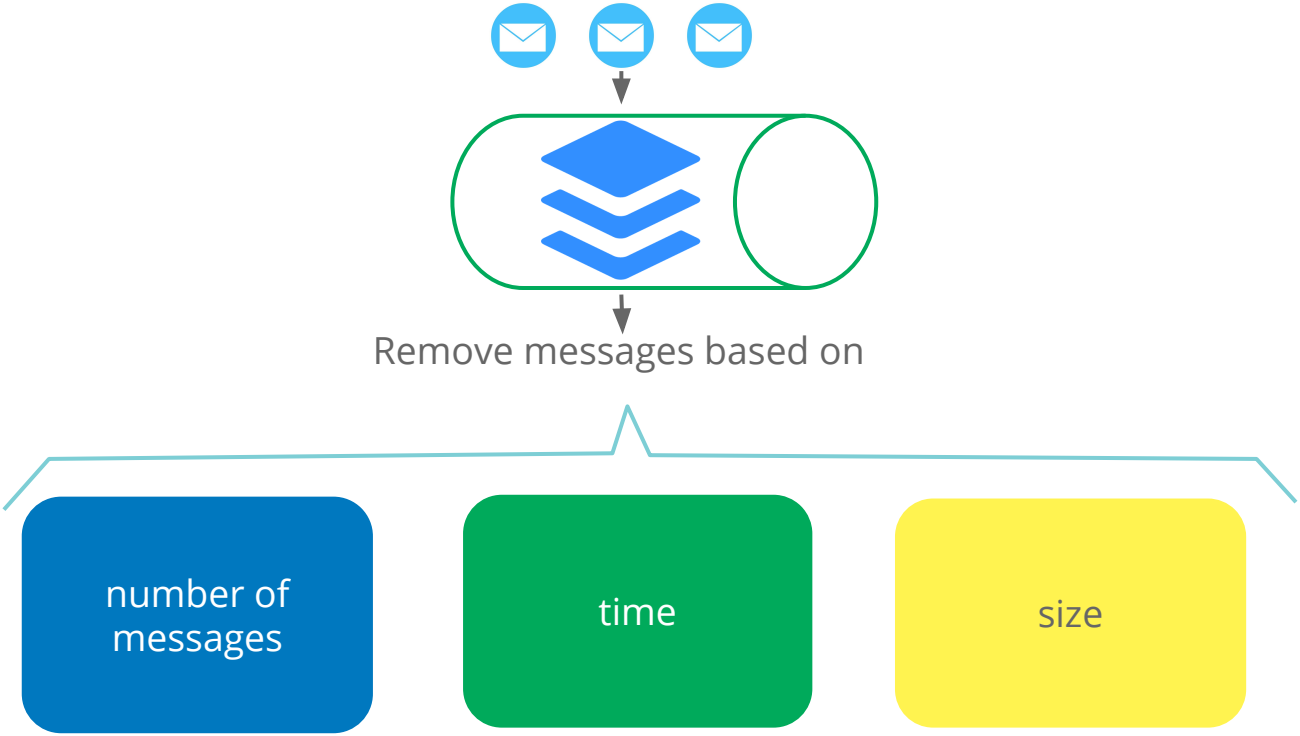
Request => RequiredAcks Timeout [TopicName [Partition MessageSetSize MessageSet]]



Response => [TopicName [Partition ErrorCode Offset]]

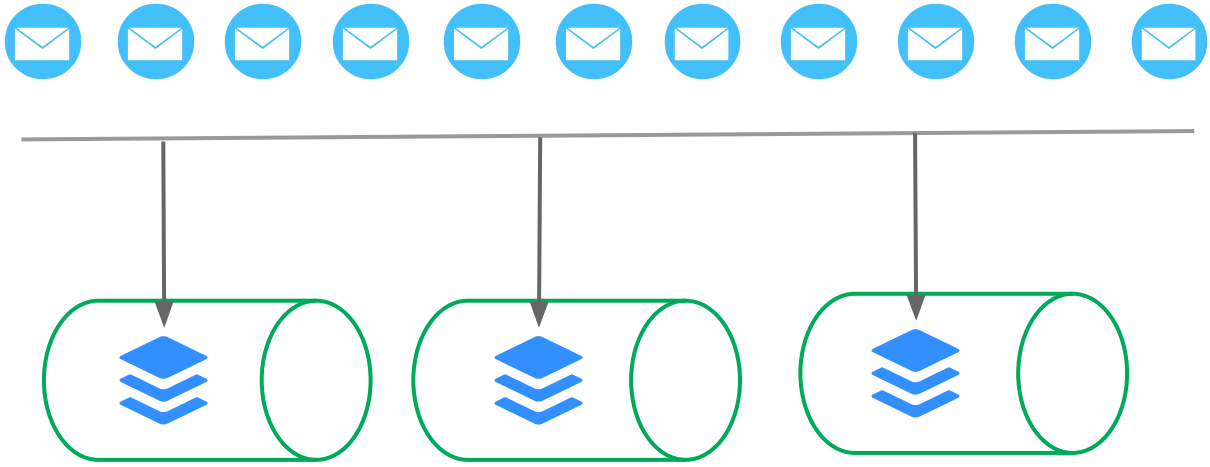
org.apache.kafka.clients.producer.KafkaProducer

Topic



kafka.common.Topic

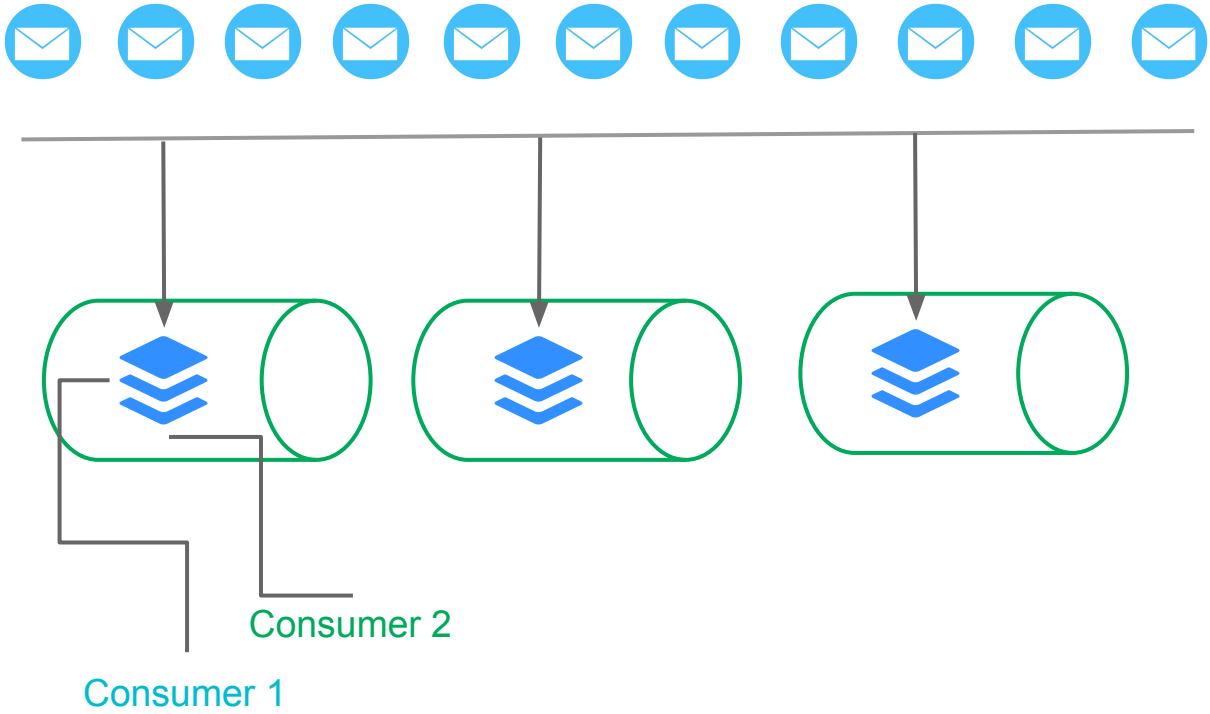
Partitions



Serves: Horizontal scaling, Parallel consumer reads

kafka.cluster.Partition

Consumers - pull



kafka.consumer.ConsumerConnector,
kafka.consumer.SimpleConsumer

Consumer offsets

committing and fetching consumer offsets

img src: http://www.reynanprinting.com/photos/undefined/impresion-offset1.jpg

kafka:// - protocol

“Binary protocol over TCP”

- Metadata
- Send
- Fetch
- Offsets
- Offset commit
- Offset fetch

Mechanical Sympathy

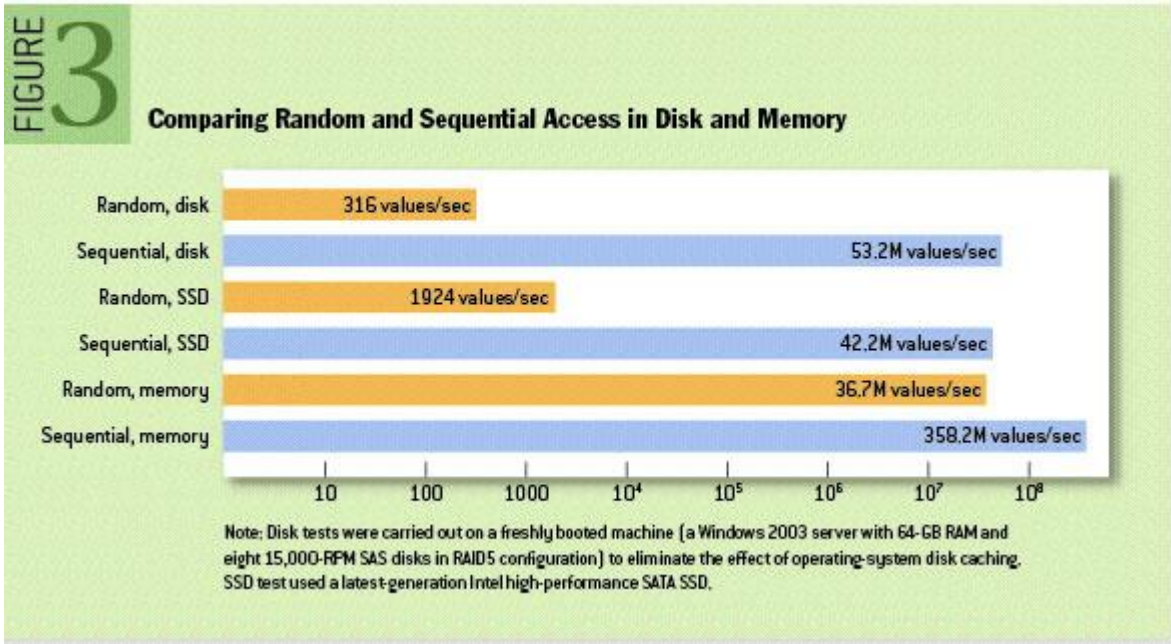
"The most amazing achievement of the computer software industry is its continuing cancellation of the steady and staggering gains made by the computer hardware industry." - Henry Peteroski

Image source: http://www.theguide2surrey.com

Persistence

“Everything is faster till the disk IO.”

Disk faster than RAM



src: http://queue.acm.org/detail.cfm?id=1563874

Linear Read & Writes

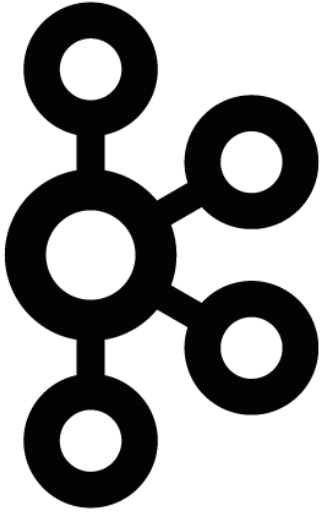
On high level there are only two operations:

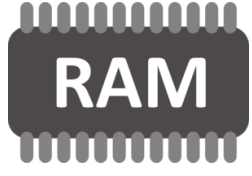
Append to end of log

fetch messages from a partition beginning from a particular message id

sequential file I/O

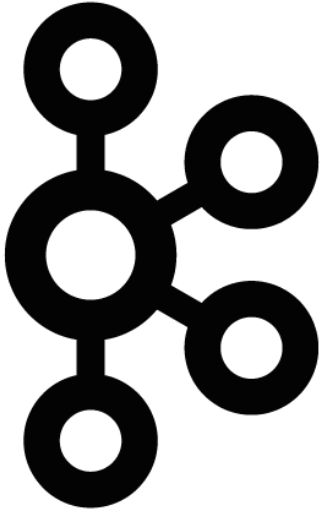
“Let us play pictionary”

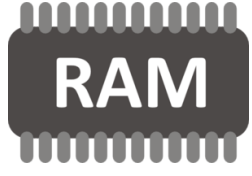




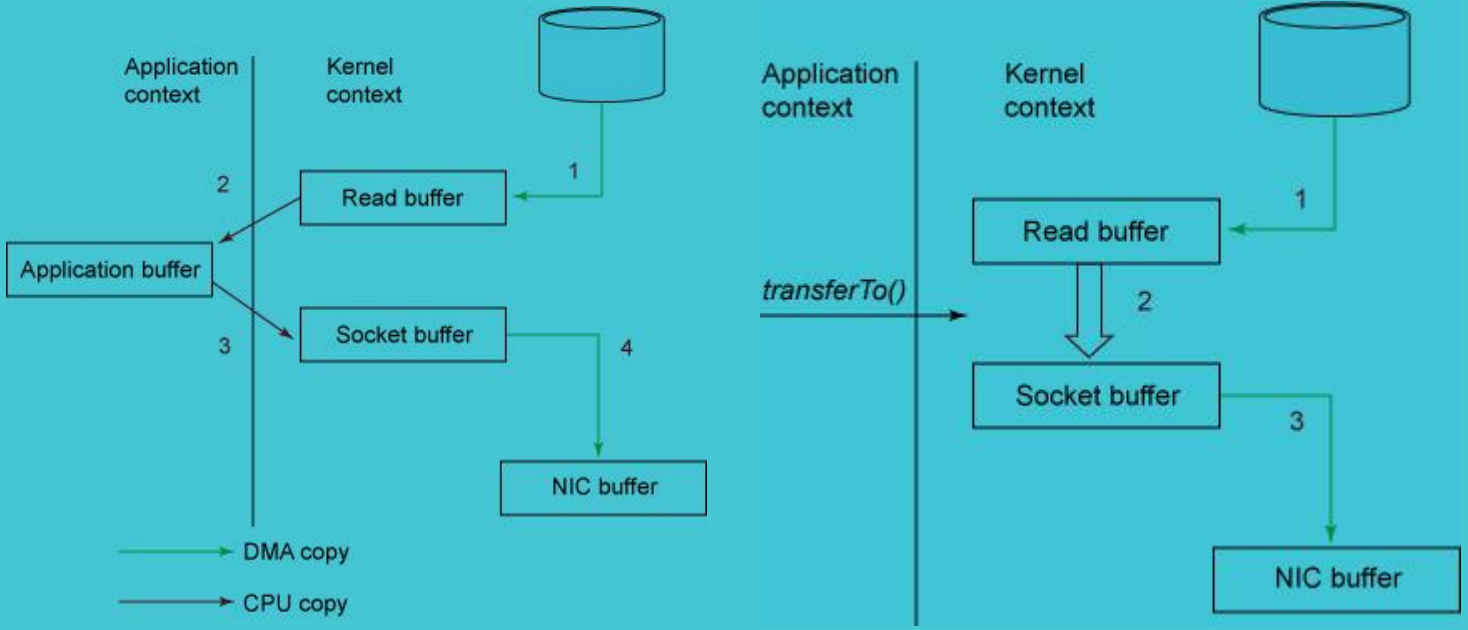
Linux Page Cache

"Kafka ate my RAM"





ZeroCopy



src: http://www.ibm.com/developerworks/library/j-zerocopy/

Batching

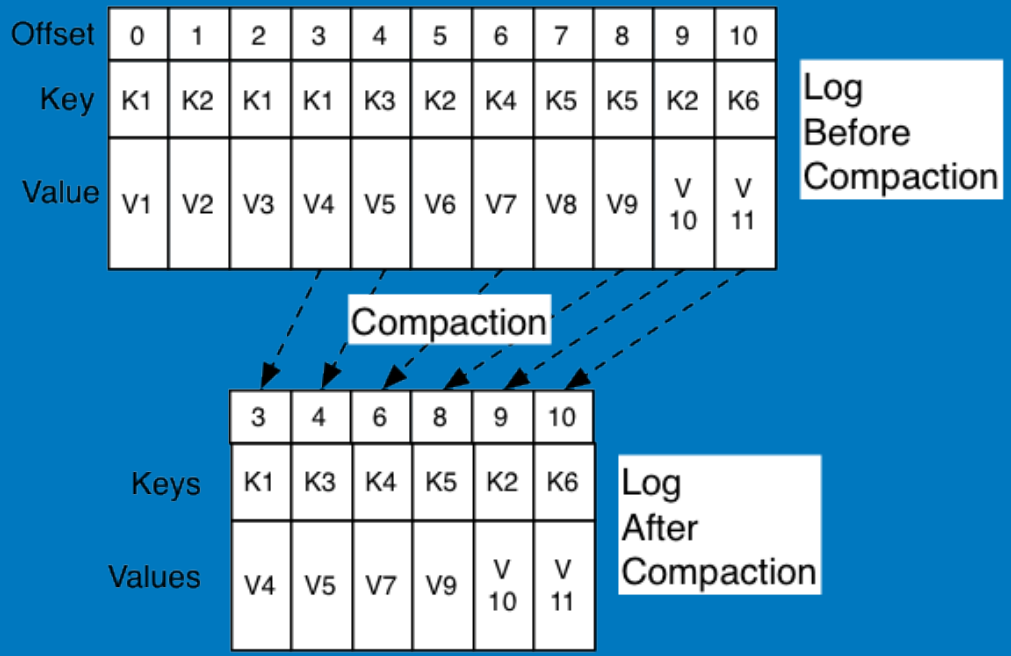
small latency to improve throughput

img src: https://prashanthpanduranga.files.wordpress.com/2015/05/tirupati.jpg

Compression

bandwidth is more expensive per-byte to scale than disk I/O, CPU, or network bandwidth capacity within a facility

kafka.message.CompressionCodec



Log compaction

kafka.log.LogCleaner, LogCleanerManager

img src: http://kafka.apache.org/083/documentation.html

Message Delivery

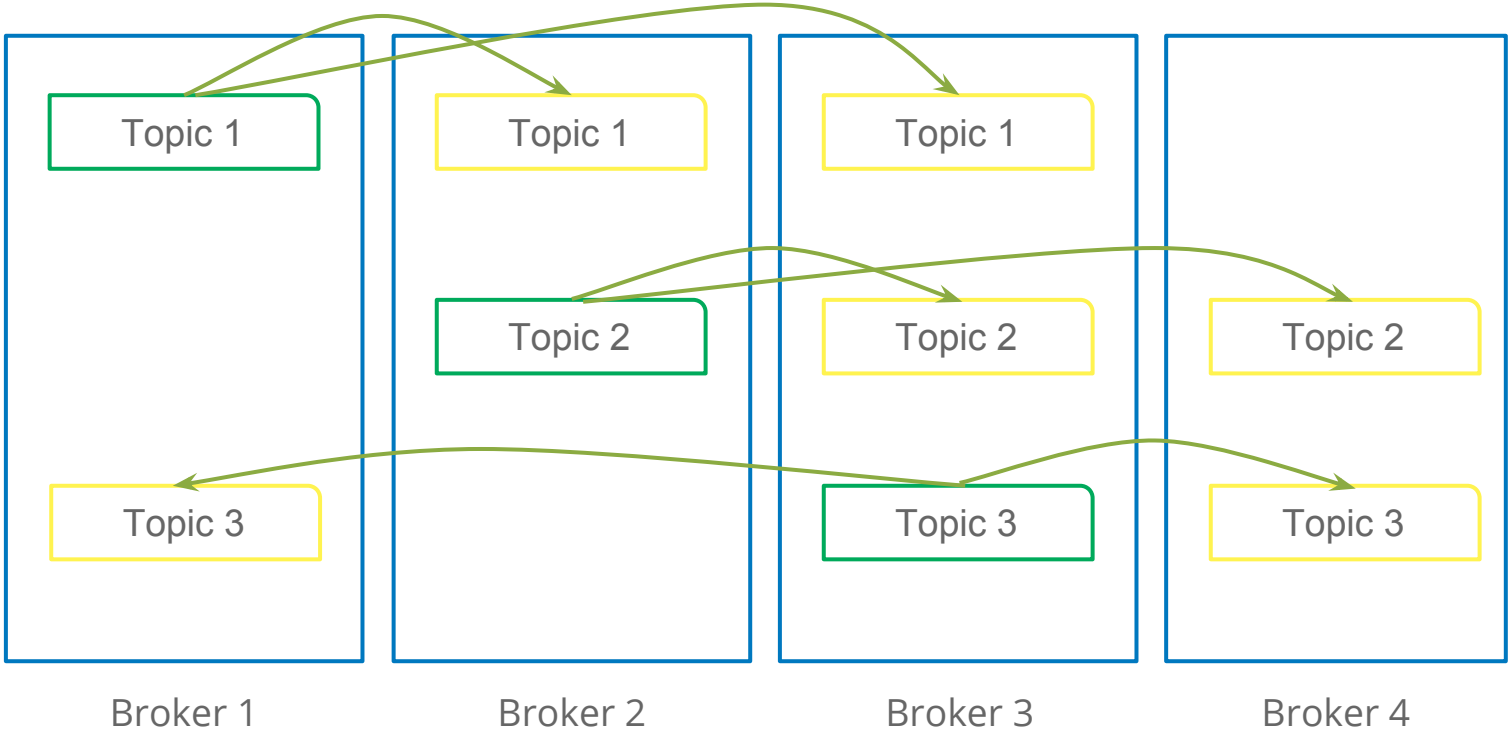
Replication

un-replicated = replication factor of one

Quorum based

- Better latency
- To tolerate **“f”** failures, need **“2f+1”** replicas

Primary-backup replication



ZooKeeper

cluster coordinator