



Ken Chen 2015/04/28

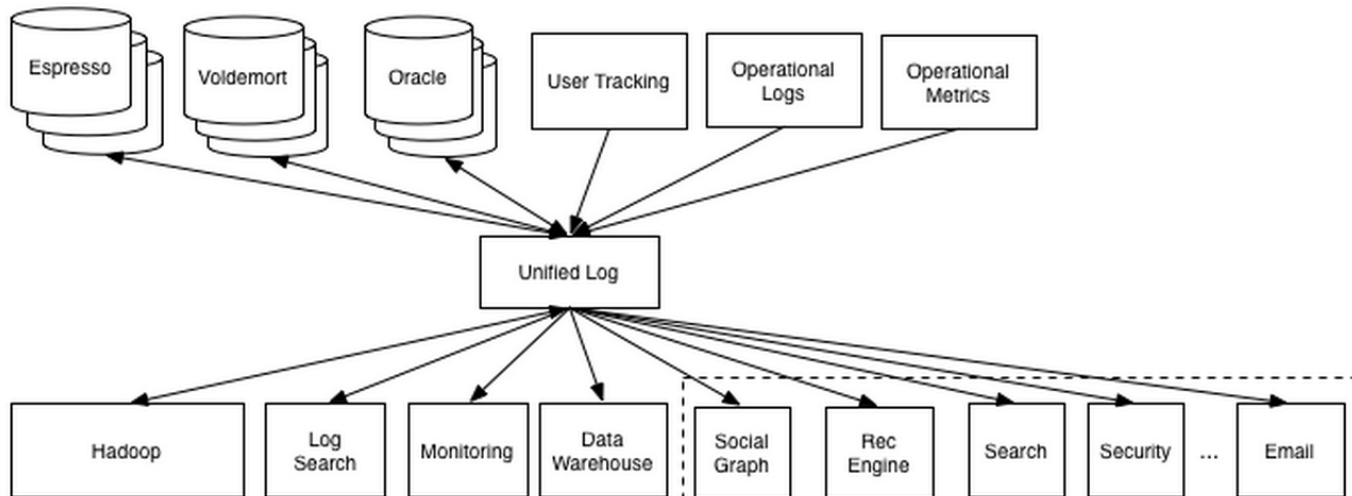
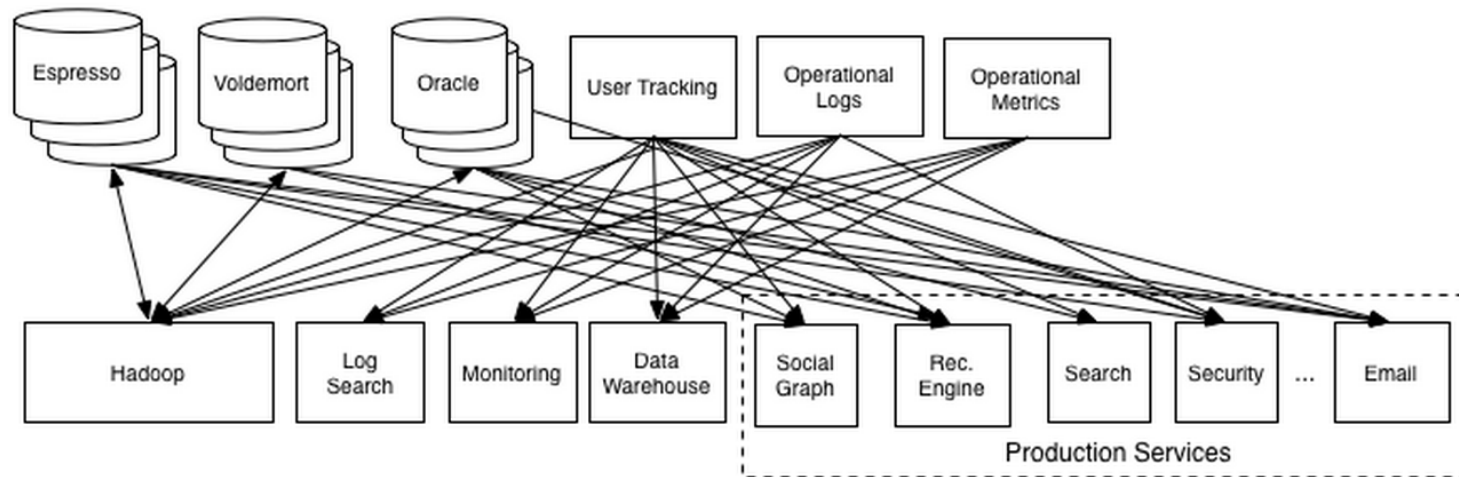
Agenda

- Kafka
 - What is Kafka
 - Key components, features designs
 - Deployment Checklist
 - Integration
 - Kafka in future

What is Kafka

- What is Kafka ?
 - Is a **distributed, partitioned, replicated** commit **log service**.
 - Provides **Push/Pull** messaging system semantics
 - High Performance
 - CAP (Consistency/Availability/Partition-tolerance)

The most used pic

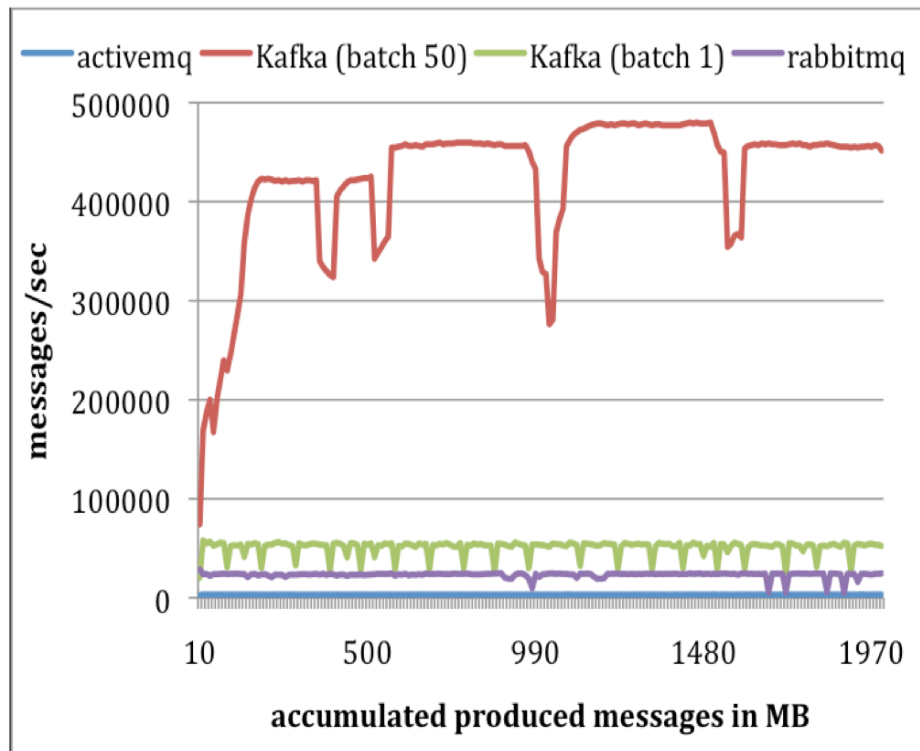


Kafka Perf

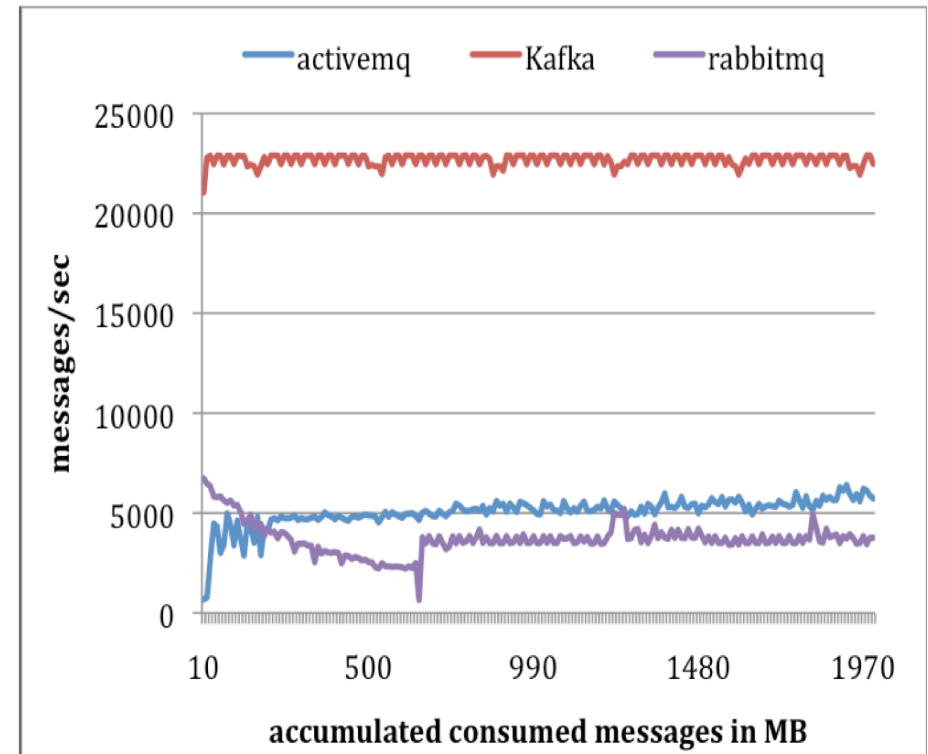
Apache Kafka @ LinkedIn

- ❑ 175 TB of in-flight log data per colo
- ❑ Low-latency: ~1.5ms
- ❑ Replicated to each datacenter
- ❑ Tens of thousands of data producers
- ❑ Thousands of consumers
- ❑ 7 million messages written/sec
- ❑ 35 million messages read/sec
- ❑ Hadoop integration

Perf Comparison



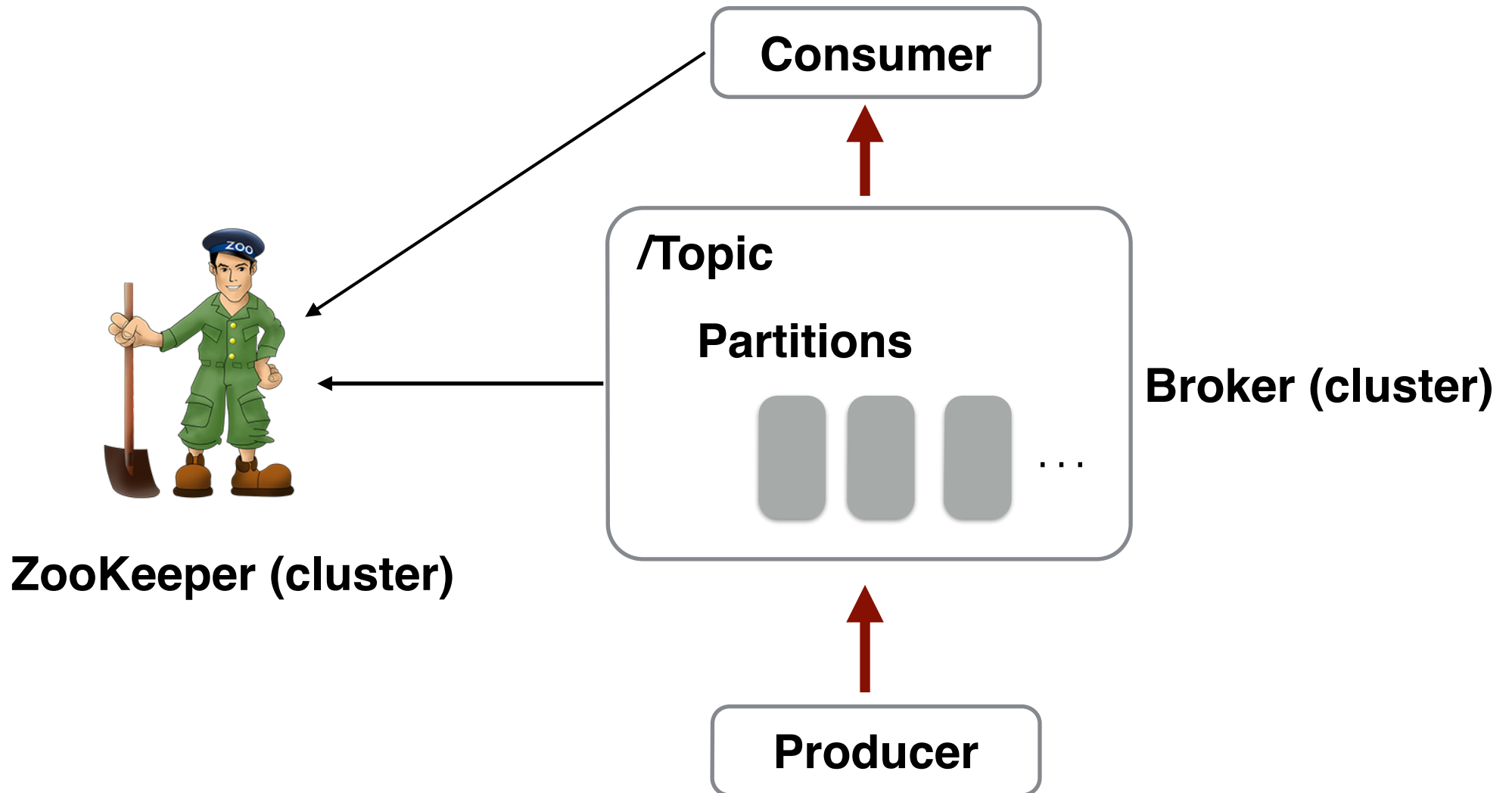
Producer Performance



Consumer Performance

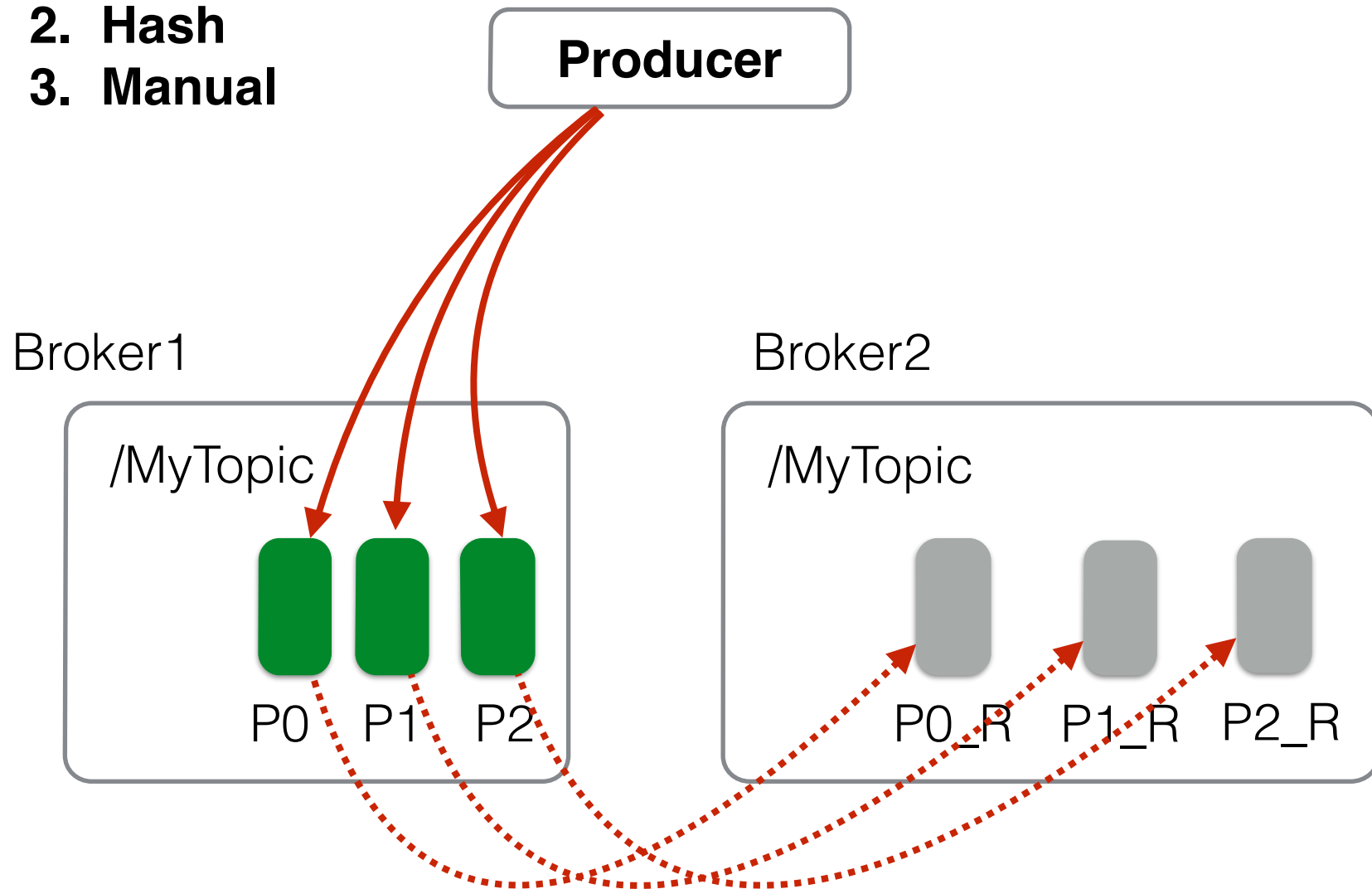
Key Components/Features

Key components



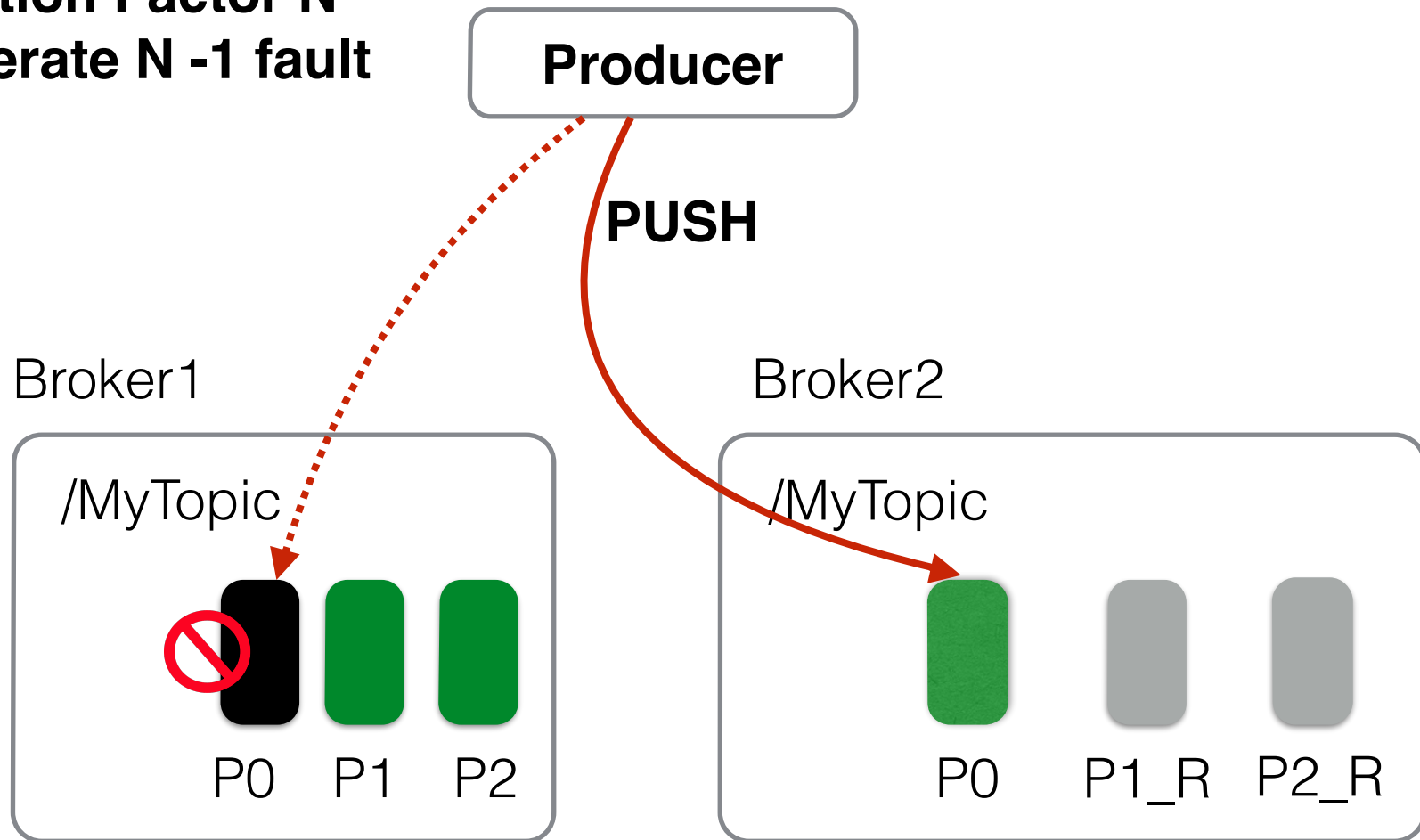
Producer

1. Random
2. Hash
3. Manual



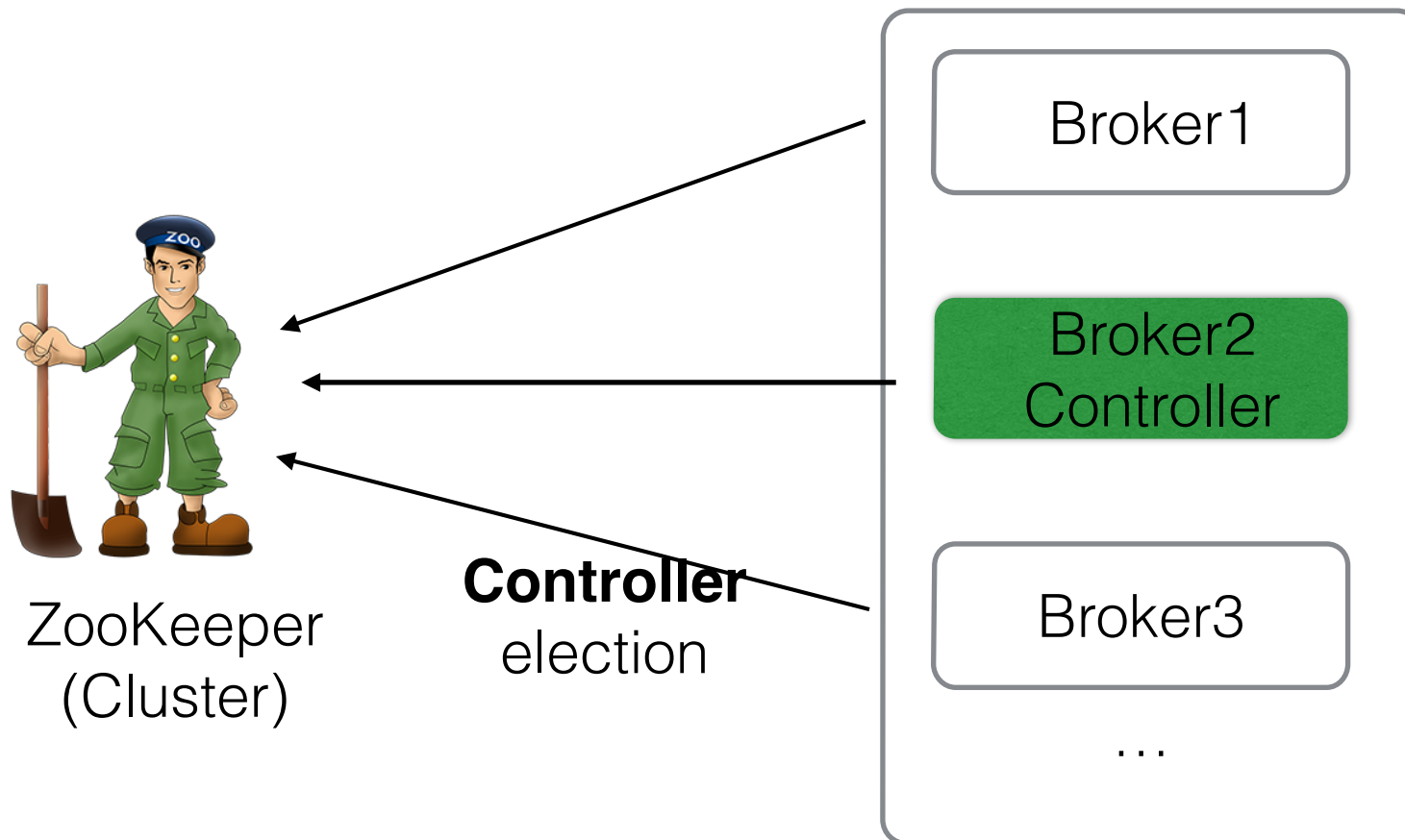
Producer

Replication Factor N
Can tolerate N - 1 fault

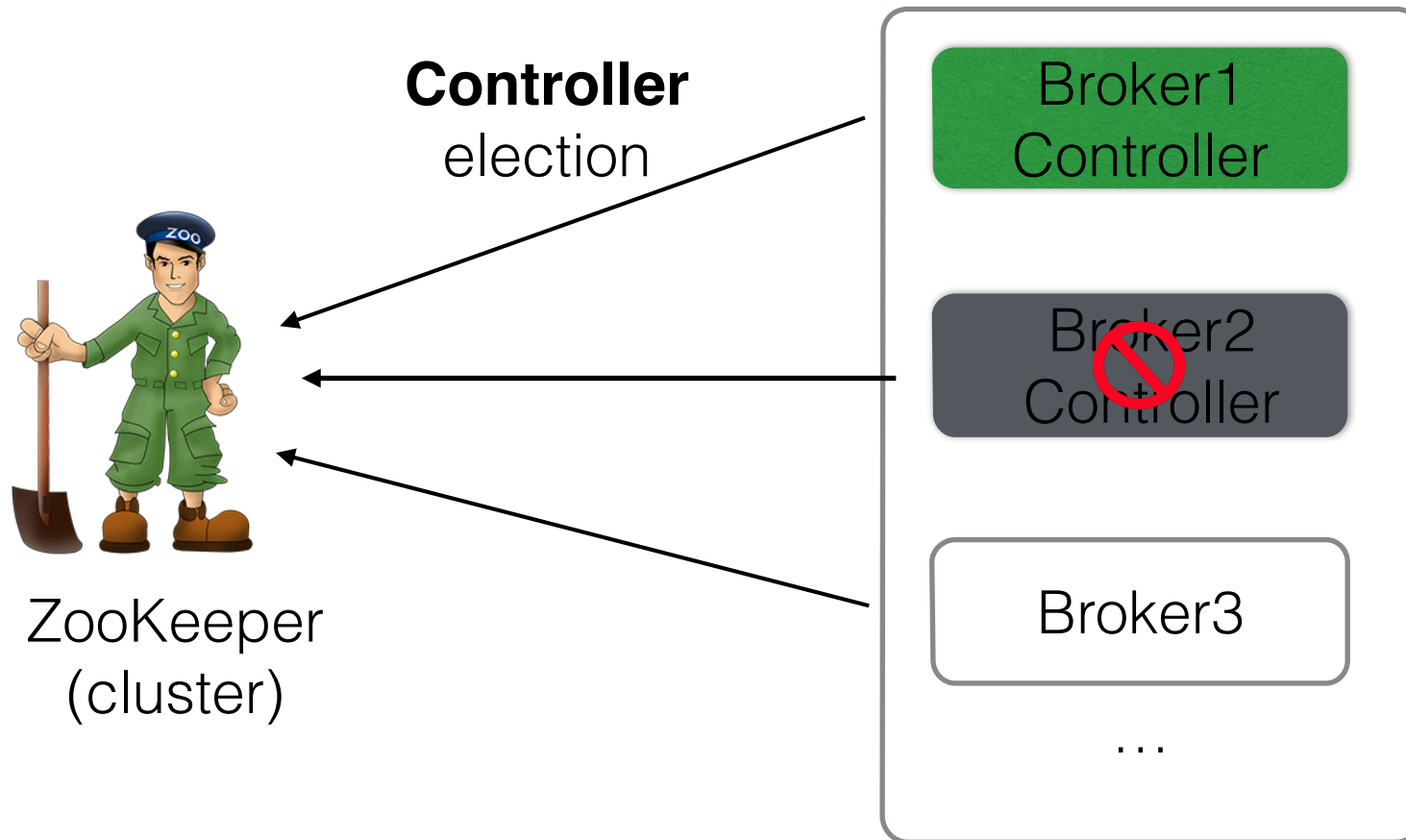


Broker

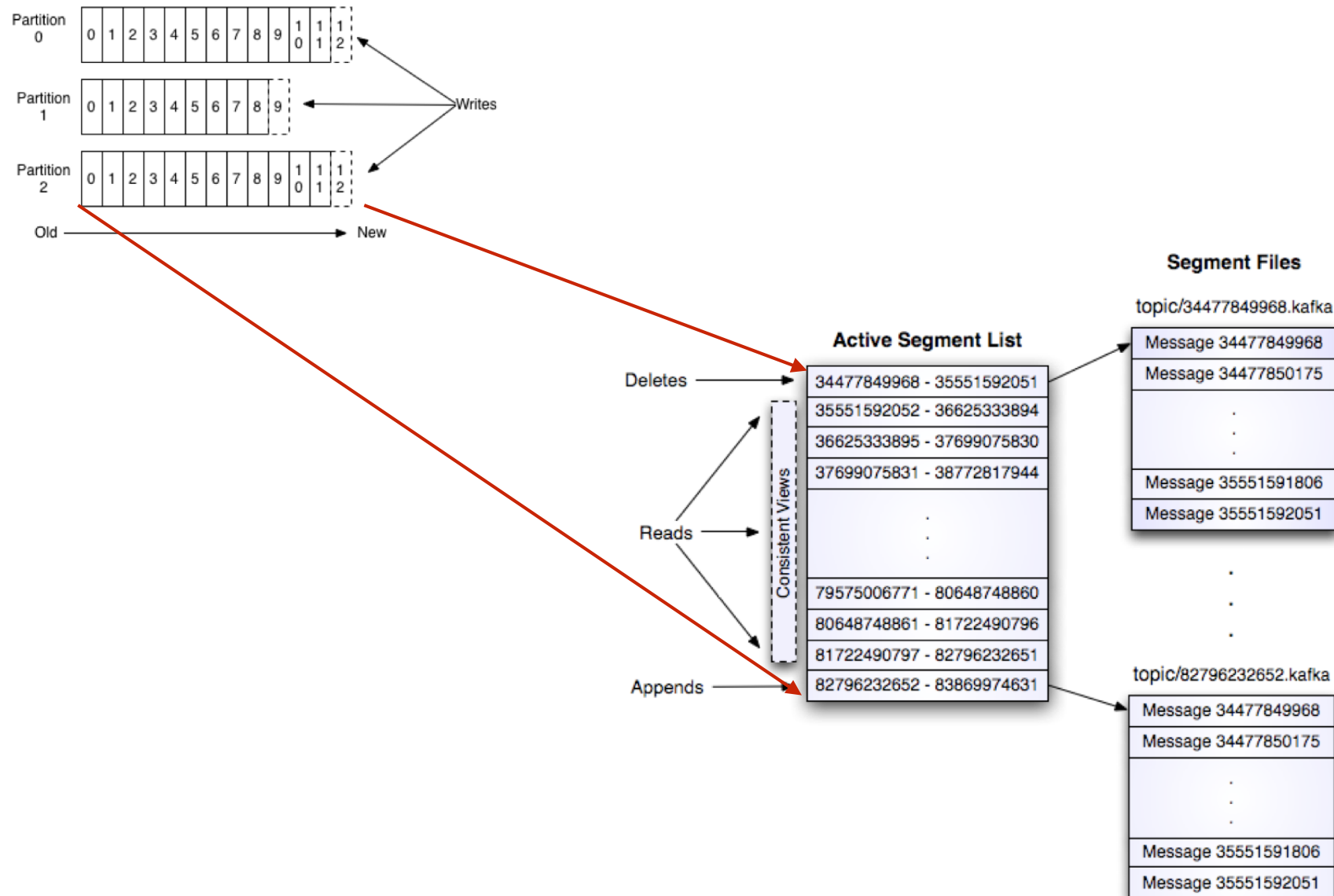
1. **Handle Read/Write requests**
2. **Controller handles partition states/replicas/re-assignment**



Broker

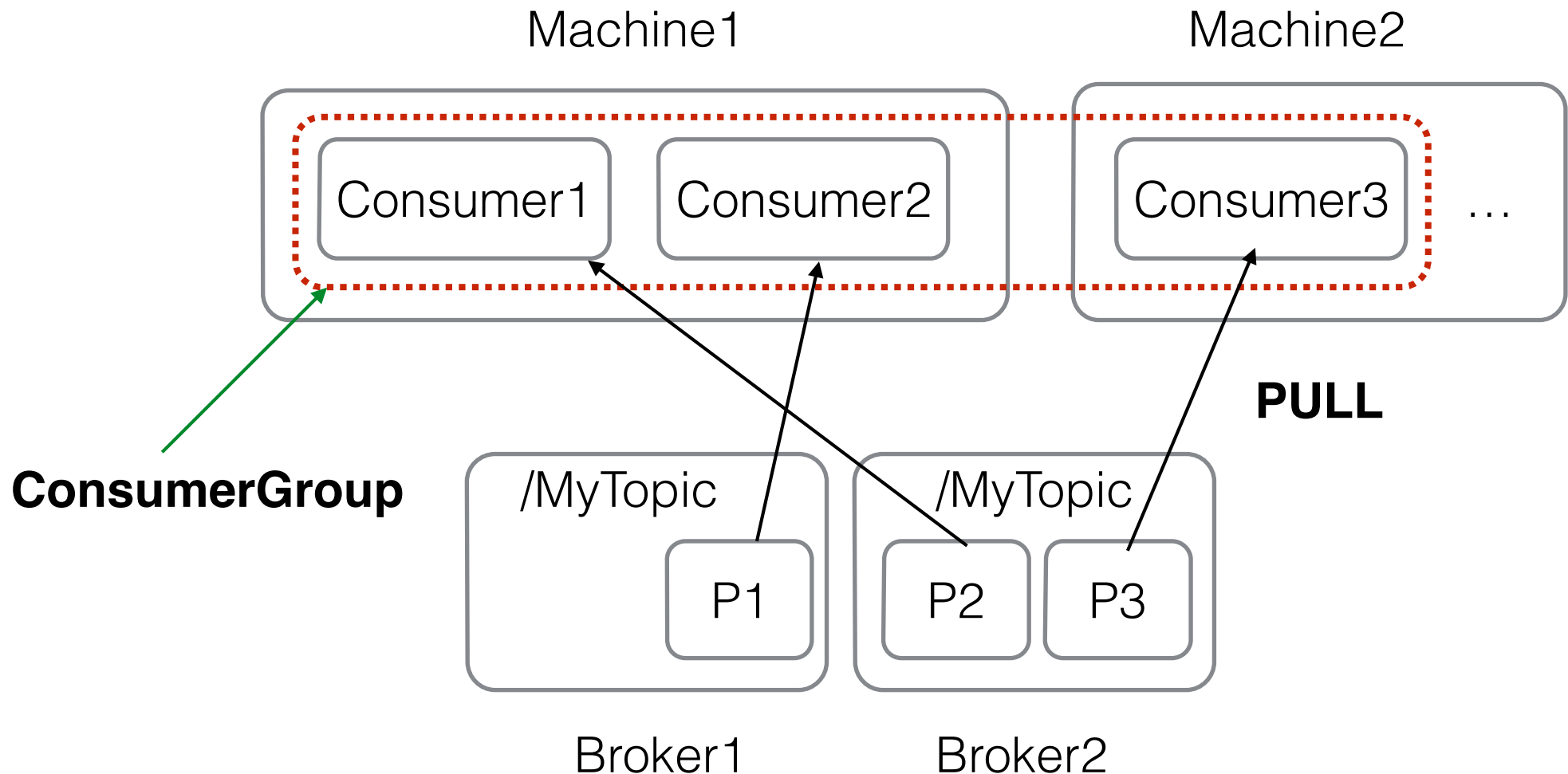


Partition



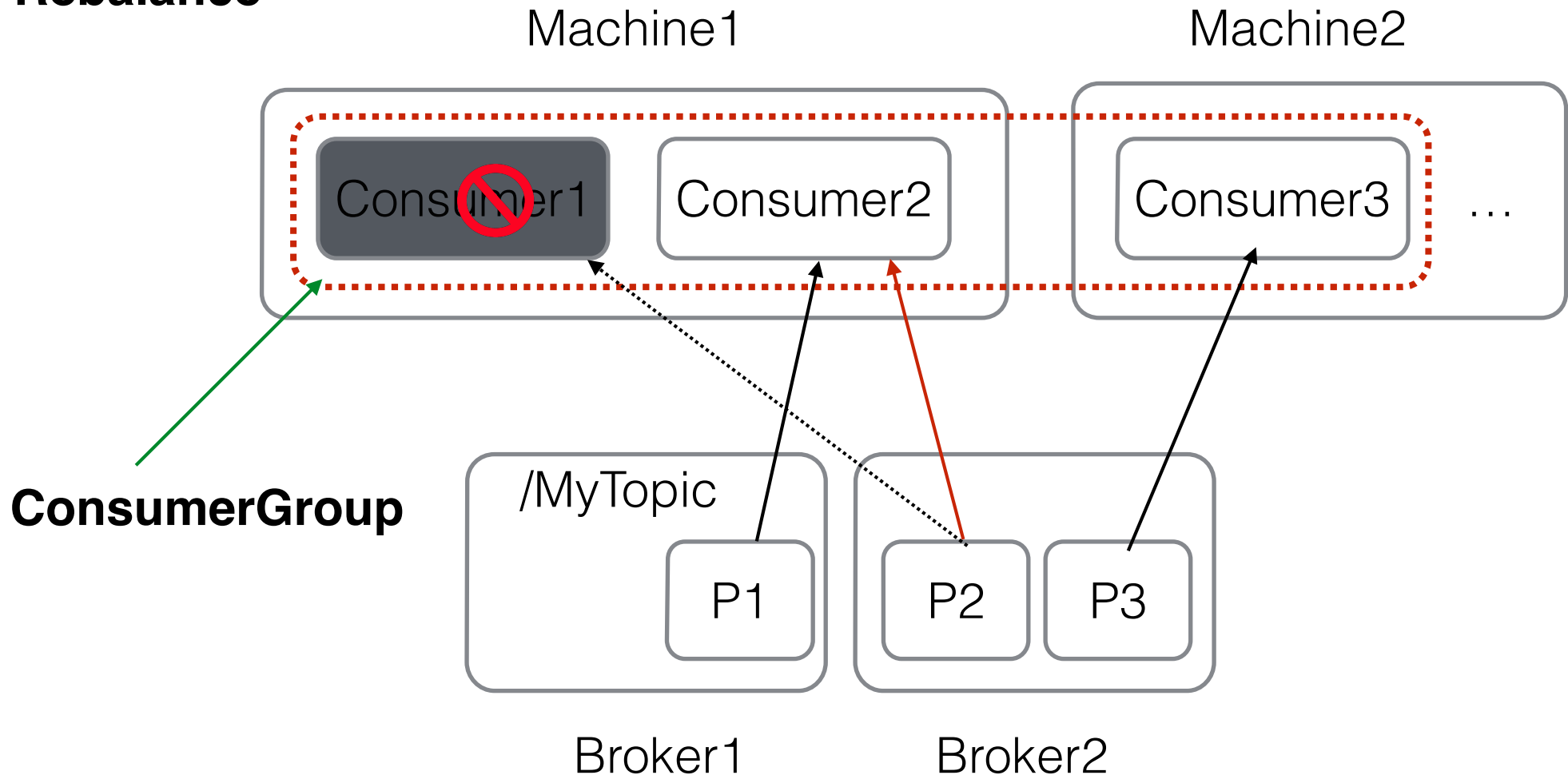
Consumer (Group)

Don't create more consumers than partitions in a CG



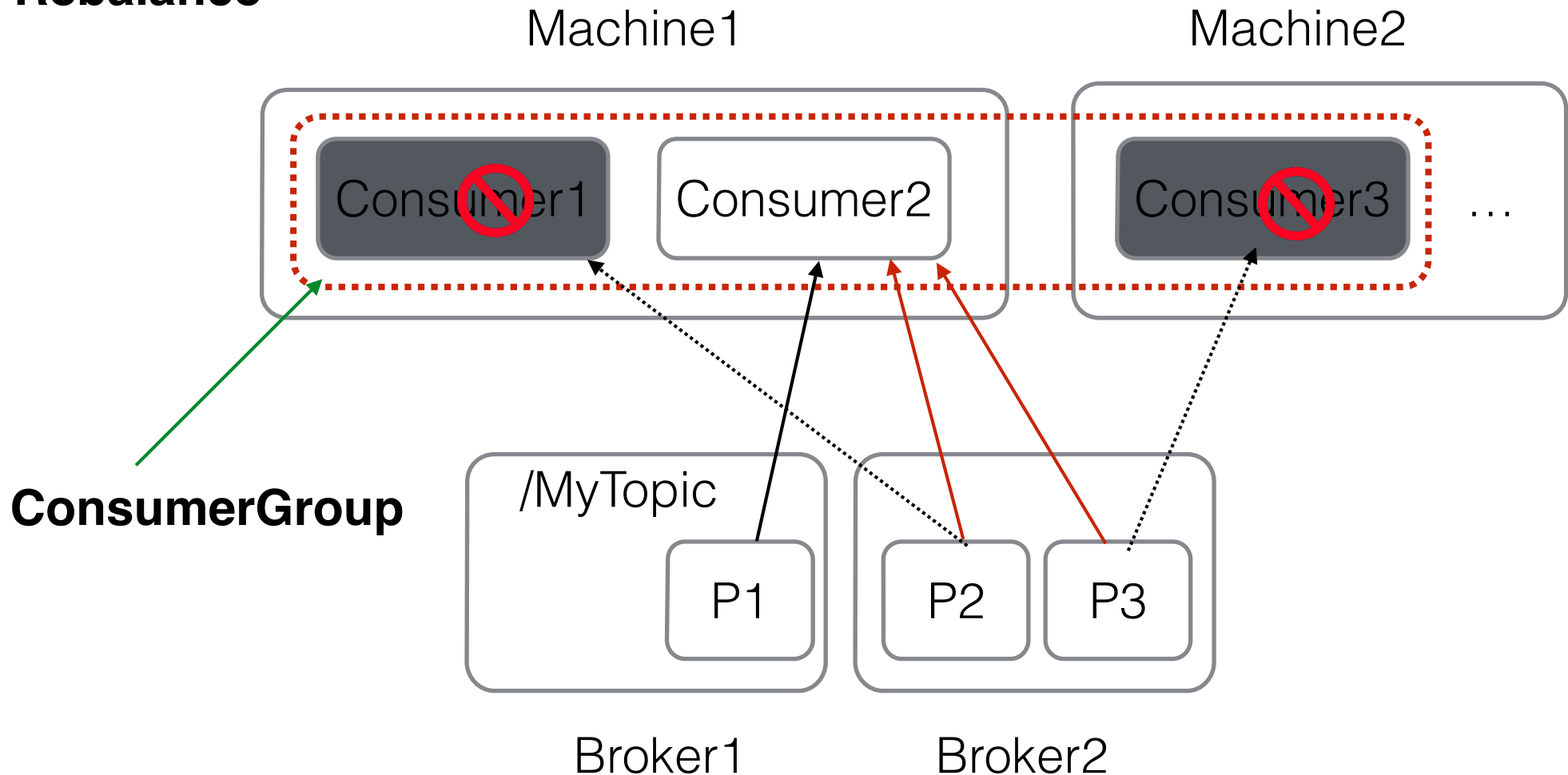
Consumer (Group)

Rebalance



Consumer (Group)

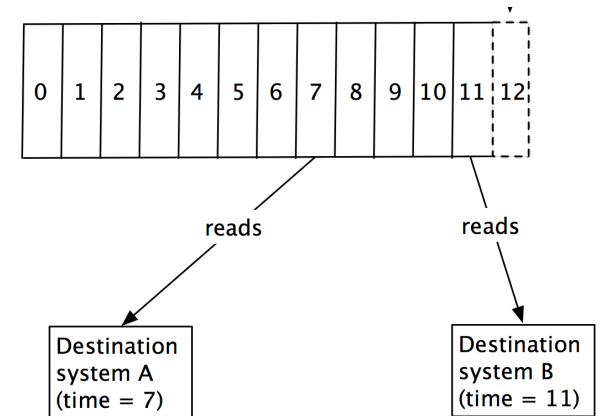
Rebalance



Consumer Offset Mgmt

Consumer Offsets Mgmt matters

1. Maintain them in ZooKeeper
 - ▶ Naturally supported by Kafka but not recommended
2. Maintain them in Kafka
 - ▶ Newly supported. Improvement is still undergoing.
3. **Maintain them by ourselves (External database)**



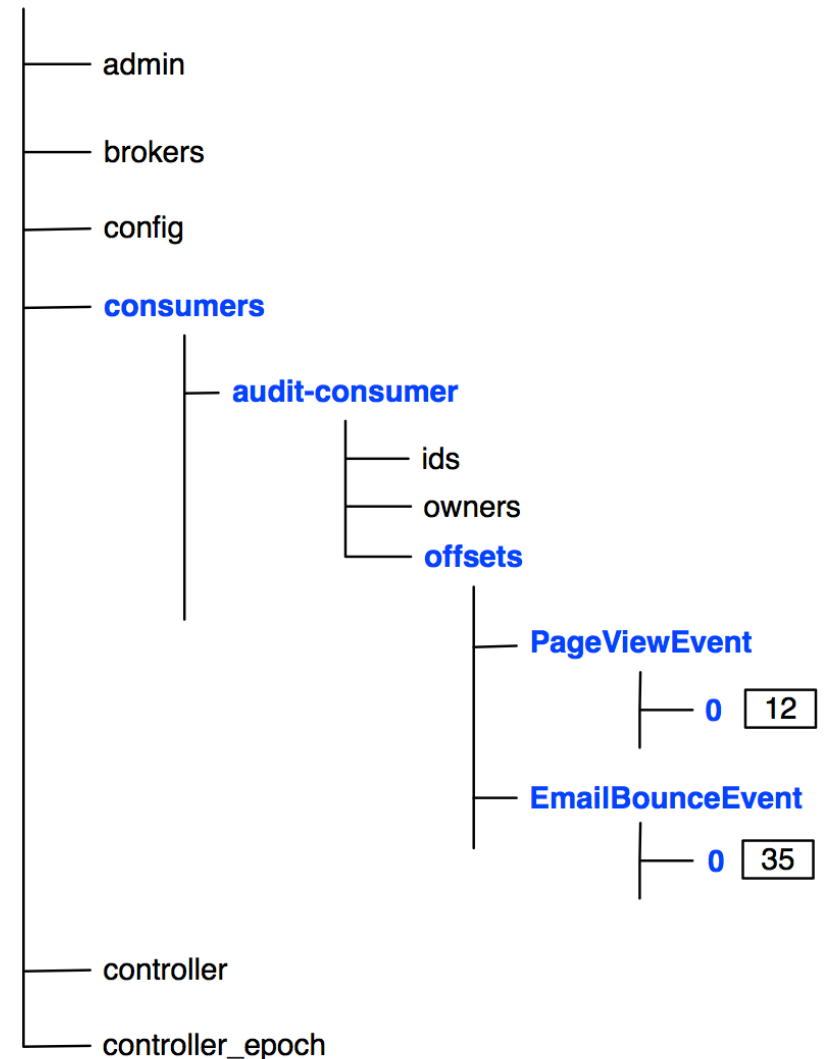
ZooKeeper

- **Heartbeat/Leader election etc**

- Ephemeral znode
- Notification/Callback when something changes

- **Global Registry**

- Broker Registry
- Topic Registry
- Consumer Registry
- Partition State Registry
- Controller Registry
- Consumer Offsets
- Partition Owner Registry
- ...



Recap

- **Producer**
- **Consumer**
- **Topic/Partition**
- **Broker**
- **ZooKeeper**

Deployment Checklist

ZooKeeper

1. Dedicate decent Servers
2. 5 Nodes Ensembles
3. SSDs
4. Java heap size

Brokers

1. RAID 10 vs RAID 5
2. XFS
3. Page cache
4. File handler number
5. CPU/Memory
6. Network bandwidth
7. How many brokers in a cluster

Topics/Partitions

1. How many partitions per broker and per cluster
2. Topic name length and number (Zookeeper node size is 1MB max)
3. Topic data retention policy
4. Segment size

Cluster expansion/shrink

1. Pretty hard for now. Future improvement
2. Consider using Mirror Maker

