Kafka

Nakul Mishra

"Performance (or lack of it) happen in a specific context"

DTE SYSTEMS
D

FOCUS

on what matters

What are the typical root causes you most often experience

REBELLABS by ZEROTURNAROUND

Bias in Observations == bias in diagnostic

Measure over Guess

AND...

Reasoning

Wrong

"OMG! Log Flush time is 14ms!!"

Is this high?

How often do we flush logs?

Is it blocking?

Who cares?

Right

"OMG! We only process 50k requests per second! We need 10x this!"

Why? IO threads are busy.

Why? Waiting for flush.

Why are we flushing so often?

Log segments keep rotating.

Lets configure larger segments!

Kafka-distributed System

Kafka is super reliable

...If you know how to configure it that way

Mojaloop - Produce Durably

Imp. Producer Configs

- batch.size - size based
- linger.ms - time based
- compression.type
  - NOTE : better compression ratio -> higher throughput -> Higher latency
- max.inflight.requests.per.connection (ordering)
- acks (durability level while producing)

Sample config for Producing Durably & correct message Ordering

- Set replication factor: >= 3
- Set minimum ISRs: 2
- Acks:
  - all The producer gets an ack after all ISRs receive the data; the message is committed
- Retries:
  - How long of downtime I need to survive divide by retry.backoff.ms (MAX_INT)
- Memory for retries:
  - Have plenty of buffer.memory
  - max.block.ms = Long.MAX_VALUE
  - Or handle the BufferExhaustedException / TimeoutException yourself
- In flight requests could lead to message re-ordering.
  Max.inflight.requests.per.connection = 1

Think about Disaster Recovery +
Multi-datacenter

https://www.confluent.io/white-paper/disaster-recovery-for-multi-datacenter-apache-kafka-deployments/

Measure over Guess?

Monitoring

NOTE: Keep in mind that monitoring is NOT alerting!

What to monitor?

Monitor: Broker | Producer | Consumer | Zookeeper

Tons of metrics, all available via JMX

- Broker throughput
- Topic throughput
- Disk utilization
- Unclean leader elections
- Network pool usage
- Request pool usage
- Request latencies - 30 request types, 5 phases each
- Topic partition status counts: online, under replicated, offline
- Log flush rates
- ZK disconnects
- Garbage collection pauses
- Message delivery
- Consumer groups reading from topics
- ....

Reference

- Broker : https://docs.confluent.io/current/kafka/monitoring.html#broker-metrics
- Producer : https://docs.confluent.io/current/kafka/monitoring.html#producer-metrics
- Consumer : https://docs.confluent.io/current/kafka/monitoring.html#new-consumer-metrics
- Zookeerper : https://docs.confluent.io/current/kafka/monitoring.html#zk-metrics

Monitor: UnderReplicatedPartitions!!!

Monitor Consumer Lag

Understanding Anatomy helps

Understand how...

- Produce & Fetch request lifecycle

Measure over Guess: Host System

Monitor Host System

- CPU usage
- Memory usage
- Available disk space
- Disk IO
- Network IO
- Number of open file handles

Alert Host System

- 60% disk usage for disks storing the Kafka log (configured via log.dirs)
- 60% disk IO usage
- 60% network IO usage
- 60% file handle usage

Why 60% ?

60% is chosen to afford time and resources to add more nodes and move partitions

NOTE: The Kafka broker will only use JVM heap space for metadata and buffers. The broker relies on the OS' page cache for caching.

Kafka Performance tools?

- kafka-consumer-perf-test
- kafka-producer-perf-test

Thank you!

Questions?