

Kafka

[零 参考资料 3](#_Toc58856172)

[一 概述 3](#_Toc58856173)

[1 体系结构 3](#_Toc58856174)

[1.1 概念 3](#_Toc58856175)

[2 时间轮 5](#_Toc58856176)

[3 ISR & HW & LEO 5](#_Toc58856177)

[二 Producer 7](#_Toc58856178)

[1 概述 7](#_Toc58856179)

[2 配置 7](#_Toc58856180)

[3 API 9](#_Toc58856181)

[三 Consumer 9](#_Toc58856182)

[1 概述 9](#_Toc58856183)

[2 配置 9](#_Toc58856184)

[3 API 11](#_Toc58856185)

[四 Topic&Partition 11](#_Toc58856186)

[1 概述 11](#_Toc58856187)

[2 Topic配置 11](#_Toc58856188)

[五 Log Storage 12](#_Toc58856189)

[1 概述 12](#_Toc58856190)

[2 日志V0 12](#_Toc58856191)

[3 日志V1 12](#_Toc58856192)

[4 日志V3 12](#_Toc58856193)

[六 Server 12](#_Toc58856194)

[1 概述 12](#_Toc58856195)

[2 Broker配置 12](#_Toc58856196)

[七 监控 17](#_Toc58856197)

[八 Connect 17](#_Toc58856198)

[1 概述 17](#_Toc58856199)

[2 配置 17](#_Toc58856200)

[九 应用 17](#_Toc58856201)

# 零 参考资料

https://kafka.apache.org/documentation/

https://kafka.apache.org/26/javadoc/index.html

《深入理解Kafka 核心设计与实践原理》朱忠华

PS：1 文档是2.6版本，书是2.0版本，参数配置以文档为准，具体差异不在本文体现

2 红色配置为关键配置

# 概述

## 体系结构

### 概念

* + - 1. Producer：生产者，即消息发送方，负责产生消息，投递到Kafka中
      2. Consumer：消费者，即消息接收方，连接到kafka上并接收消息，然后根据消息处理业务逻辑
      3. Broker：服务代理节点
         1. 可以简单的理解为一个单独的kafka服务节点或Kafka服务实例，
         2. 但如果一台服务器上只部署了一个kafka实例，才可以将broker看作一台Kafka服务器。
         3. 一个或多个Broker组成了kafka集群
      4. Topic：主题，逻辑上的概念，kafka消息按topic进行分类
         1. 生产者负责将消息发送到特定Topic
         2. 消费者通过订阅topic消费消息
         3. 一个Topic可以横跨多个Broker
      5. Partition: Topic的分区
         1. 一个Partition只属于单个Topic，因此也被称为Topic-Partition
         2. 同个Topic下，不同Partition包含的消息不同
         3. 在存储层面，Partition可以看作是一个可追加的Log文件
      6. Offset：偏移量，从0开始
         1. 消息被追加到Partition的log文件时，会分配到一个特定的Offset
         2. offset是消息在存储其的Partition中唯一标识。
         3. offset不跨分区，因此kafka通过offset保证消息在当前分区有序，但不保证消息在Topic中有序
      7. Replica：副本
         1. kafka基于Partition实现多Replica机制，提高容灾能力
         2. 同一个Partition的不同Replica中保存的消息最终相同，即同一时刻，同一个Partition的多Replica之间并非一定相同
         3. Replica之间是一主多从关系：

Leader 负责处理消息读写请求，follower只负责与Leader消息同步

当Leader出现故障时，从Follower中重新选举新的Leader对外服务

大部分时候follower的消息相对Leader有一定延后

* + - 1. Assigned Replica：Partition所有replica的统称
         1. ISR(In-Sync Replica)：由所有与Leader保持一定程度同步的replica组成，是AR的子集
         2. OSR(Out-of-Sync Replica)：由与Leader同步滞后过多的Replica组成，是AR的子集
         3. AR = ISR+OSR,正常情况下，AR=ISR，OSR为空
         4. Leader负责维护&跟踪所有follower的同步情况

当ISR中的follower滞后太多或失效时，leader会把其移出ISR到OSR中

当OSR中的follower同步消息追上leader时，将追上的follower从OSR移到ISR

* + - 1. HW(High Watermark)：高水位，标识了一个特定的消息offset，consumer只能获取到这个offset之前的消息
      2. LEO(Log End Offset)：标识当前log下一条待写入消息offset，相当于当前log最后条消息offset+1







## 时间轮

## ISR & HW & LEO

* + - 1. 假设某个分区的ISR集合中有3个副本，即1个leader和两个follow
         1. 此时LEO和HW都为3
         2. 生产者发出消息3和消息4，会被先存入leader



* + - 1. 消息3&4写入Leader后，Follower会请求拉取消息3&4以进行同步



* + - 1. 同步过程中，不同Follower同步效率不同，下图情况：
         1. Leader的LEO=5，Follower1 LEO=5，Follower2 LEO=4， HW=4
         2. Consumer可以获取0-3消息



* + - 1. 当所有Replica都同步到消息4，则所有HW和LEO都为5，Consumer可以获得到消息4了



# Producer

## 概述



## 配置

|  |  |
| --- | --- |
| 配置 | 说明 |
| key.serializer |  |
| value.serializer |  |
| acks |  |
| bootstrap.servers |  |
| buffer.memory |  |
| compression.type |  |
| retries |  |
| ssl.key.password |  |
| ssl.keystore.location |  |
| ssl.keystore.password |  |
| ssl.truststore.location |  |
| ssl.truststore.password |  |
| batch.size |  |
| client.dns.lookup |  |
| client.id |  |
| connections.max.idle.ms |  |
| delivery.timeout.ms |  |
| linger.ms |  |
| max.block.ms |  |
| max.request.size |  |
| partitioner.class |  |
| receive.buffer.bytes |  |
| request.timeout.ms |  |
| sasl.client.callback.handler.class |  |
| sasl.jaas.config |  |
| sasl.kerberos.service.name |  |
| sasl.login.callback.handler.class |  |
| sasl.login.class |  |
| sasl.mechanism |  |
| security.protocol |  |
| send.buffer.bytes |  |
| ssl.enabled.protocols |  |
| ssl.keystore.type |  |
| ssl.protocol |  |
| ssl.provider |  |
| ssl.truststore.type |  |
| enable.idempotence |  |
| interceptor.classes |  |
| max.in.flight.requests.per.connection |  |
| metadata.max.age.ms |  |
| metadata.max.idle.ms |  |
| metric.reporters |  |
| metrics.num.samples |  |
| metrics.recording.level |  |
| metrics.sample.window.ms |  |
| reconnect.backoff.max.ms |  |
| reconnect.backoff.ms |  |
| retry.backoff.ms |  |
| sasl.kerberos.kinit.cmd |  |
| sasl.kerberos.min.time.before.relogin |  |
| sasl.kerberos.ticket.renew.jitter |  |
| sasl.kerberos.ticket.renew.window.factor |  |
| sasl.login.refresh.buffer.seconds |  |
| sasl.login.refresh.min.period.seconds |  |
| sasl.login.refresh.window.factor |  |
| sasl.login.refresh.window.jitter |  |
| security.providers |  |
| ssl.cipher.suites |  |
| ssl.endpoint.identification.algorithm |  |
| ssl.engine.factory.class |  |
| ssl.keymanager.algorithm |  |
| ssl.secure.random.implementation |  |
| ssl.trustmanager.algorithm |  |
| transaction.timeout.ms |  |
| transactional.id |  |

## API

# Consumer

## 概述

## 配置

|  |  |
| --- | --- |
| 配置项 | 说明 |
| key.deserializer |  |
| value.deserializer |  |
| bootstrap.servers |  |
| fetch.min.bytes |  |
| group.id |  |
| heartbeat.interval.ms |  |
| max.partition.fetch.bytes |  |
| session.timeout.ms |  |
| ssl.key.password |  |
| ssl.keystore.location |  |
| ssl.keystore.password |  |
| ssl.truststore.location |  |
| ssl.truststore.password |  |
| allow.auto.create.topics |  |
| auto.offset.reset |  |
| client.dns.lookup |  |
| connections.max.idle.ms |  |
| default.api.timeout.ms |  |
| enable.auto.commit |  |
| exclude.internal.topics |  |
| fetch.max.bytes |  |
| group.instance.id |  |
| isolation.level |  |
| max.poll.interval.ms |  |
| max.poll.records |  |
| partition.assignment.strategy |  |
| receive.buffer.bytes |  |
| request.timeout.ms |  |
| sasl.client.callback.handler.class |  |
| sasl.jaas.config |  |
| sasl.kerberos.service.name |  |
| sasl.login.callback.handler.class |  |
| sasl.login.class |  |
| sasl.mechanism |  |
| security.protocol |  |
| send.buffer.bytes |  |
| ssl.enabled.protocols |  |
| ssl.keystore.type |  |
| ssl.protocol |  |
| ssl.provider |  |
| ssl.truststore.type |  |
| auto.commit.interval.ms |  |
| check.crcs |  |
| client.id |  |
| client.rack |  |
| fetch.max.wait.ms |  |
| interceptor.classes |  |
| metadata.max.age.ms |  |
| metric.reporters |  |
| metrics.num.samples |  |
| metrics.recording.level |  |
| metrics.sample.window.ms |  |
| reconnect.backoff.max.ms |  |
| reconnect.backoff.ms |  |
| retry.backoff.ms |  |
| sasl.kerberos.kinit.cmd |  |
| sasl.kerberos.min.time.before.relogin |  |
| sasl.kerberos.ticket.renew.jitter |  |
| sasl.kerberos.ticket.renew.window.factor |  |
| sasl.login.refresh.buffer.seconds |  |
| sasl.login.refresh.min.period.seconds |  |
| sasl.login.refresh.window.factor |  |
| sasl.login.refresh.window.jitter |  |
| security.providers |  |
| ssl.cipher.suites |  |
| ssl.endpoint.identification.algorithm |  |
| ssl.engine.factory.class |  |
| ssl.keymanager.algorithm |  |
| ssl.secure.random.implementation |  |
| ssl.trustmanager.algorithm |  |

## API

# Topic&Partition

## 概述

## Topic配置

|  |  |
| --- | --- |
| 配置项 | 说明 |
| cleanup.policy |  |
| compression.type |  |
| delete.retention.ms |  |
| file.delete.delay.ms |  |
| flush.messages |  |
| flush.ms |  |
| follower.replication.throttled.replicas |  |
| index.interval.bytes |  |
| leader.replication.throttled.replicas |  |
| max.compaction.lag.ms |  |
| max.message.bytes |  |
| message.format.version |  |
| message.timestamp.difference.max.ms |  |
| message.timestamp.type |  |
| min.cleanable.dirty.ratio |  |
| min.compaction.lag.ms |  |
| min.insync.replicas |  |
| preallocate |  |
| retention.bytes |  |
| retention.ms |  |
| segment.bytes |  |
| segment.index.bytes |  |
| segment.jitter.ms |  |
| segment.ms |  |
| unclean.leader.election.enable |  |
| message.downconversion.enable |  |

# Log Storage

## 概述

## 日志V0

## 日志V1

## 日志V3

# Server

## 概述

## Broker配置

|  |  |
| --- | --- |
| 配置项 | 说明 |
| zookeeper.connect | 指定zk地址hostname:port  多个用英文逗号隔开  也可以指定zk路径  hostname1:port1,hostname2:port2,hostname3:port3/chroot/path |
| advertised.host.name | 不推荐使用  仅当advertised.listeners或listeners配置为空时有效  建议用advertised.listeners替代  将指定外网地址推送到zk  如果没有配置则取host.name值  如果也没有配置host.name则取java.net.InetAddress.getCanonicalHostName()值  在IaaS等外网地址需要特指的环境中使用 |
| advertised.listeners | 将指定外网地址推送到zk  如果没有配置，则推送listeners配置  不能配置为0.0.0.0 |
| advertised.port | 不推荐使用  仅当advertised.listeners或listeners配置为空时有效  建议用advertised.listeners替代  将指定端口推送到zk  如果没有配置，则推送broker绑定端口 |
| auto.create.topics.enable | 是否允许自动创建topic  默认true  如果为true，则producer发送消息到一个未存在的topic，broker会自动创建一个topic |
| auto.leader.rebalance.enable | 是否允许leader重新负载  默认为true 当为true时，后台线程会根据leader.imbalance.check.interval.seconds定期检查当前leader，如果当前leader负载超过了leader.imbalance.per.broker.percentage百分比，则自动重新负载分配leader |
| background.threads | 配置后台处理辅助任务的线程数  默认10，取值为大于0的正整数 |
| broker.id | 设置broker唯一标识  默认值-1， 取值int型  如果不设置，则自动生成一个  为保证同个zk下broker.id唯一，自动生成的id从reserved.broker.max.id+1开始 |
| compression.type | 指定topic消息默认压缩方式  默认值 producer  取值gzip, snappy, lz4, zstd，producer，uncompressed  如果设置为uncompressed，则为不压缩  如果设置为producer，则根据producer配置决定压缩方式 |
| control.plane.listener.name |  |
| delete.topic.enable | 是否允许非admin权限删除topic  默认为true  如果设置为false，不影响admin tool删除topic |
| host.name |  |
| leader.imbalance.check.interval.seconds | 定期检查leader负载情况  默认300s |
| leader.imbalance.per.broker.percentage | 每个leader最高负载率  默认10 |
| listeners |  |
| log.dir | 指定log目录  当log.dirs未设置时，生效 |
| log.dirs | 指定多个log目录  如果未设置则取log.dir配置 |
| log.flush.interval.messages | log partation积累的消息flush到disk的阈值  默认9223372036854775807  取值为大于0的整数 |
| log.flush.interval.ms | 定时将消息从memory flush to disk  默认不设置  如果不设置，则取log.flush.scheduler.interval.ms值 |
| log.flush.offset.checkpoint.interval.ms |  |
| log.flush.scheduler.interval.ms |  |
| log.flush.start.offset.checkpoint.interval.ms |  |
| log.retention.bytes |  |
| log.retention.hours |  |
| log.retention.minutes |  |
| log.retention.ms |  |
| log.roll.hours |  |
| log.roll.jitter.hours |  |
| log.roll.jitter.ms |  |
| log.roll.ms |  |
| log.segment.bytes |  |
| log.segment.delete.delay.ms |  |
| message.max.bytes |  |
| min.insync.replicas |  |
| num.io.threads |  |
| num.network.threads |  |
| num.recovery.threads.per.data.dir |  |
| num.replica.alter.log.dirs.threads |  |
| num.replica.fetchers |  |
| offset.metadata.max.bytes |  |
| offsets.commit.required.acks |  |
| offsets.commit.timeout.ms |  |
| offsets.load.buffer.size |  |
| offsets.retention.check.interval.ms |  |
| offsets.retention.minutes |  |
| offsets.topic.compression.codec |  |
| offsets.topic.num.partitions |  |
| offsets.topic.replication.factor |  |
| offsets.topic.segment.bytes |  |
| port |  |
| queued.max.requests |  |
| quota.consumer.default |  |
| quota.producer.default |  |
| replica.fetch.min.bytes |  |
| replica.fetch.wait.max.ms |  |
| replica.high.watermark.checkpoint.interval.ms |  |
| replica.lag.time.max.ms |  |
| replica.socket.receive.buffer.bytes |  |
| replica.socket.timeout.ms |  |
| request.timeout.ms |  |
| socket.receive.buffer.bytes |  |
| socket.request.max.bytes |  |
| socket.send.buffer.bytes |  |
| transaction.max.timeout.ms |  |
| transaction.state.log.load.buffer.size |  |
| transaction.state.log.min.isr |  |
| transaction.state.log.num.partitions |  |
| transaction.state.log.replication.factor |  |
| transaction.state.log.segment.bytes |  |
| transactional.id.expiration.ms |  |
| unclean.leader.election.enable |  |
| zookeeper.connection.timeout.ms |  |
| zookeeper.max.in.flight.requests |  |
| zookeeper.session.timeout.ms |  |
| zookeeper.set.acl |  |
| broker.id.generation.enable |  |
| broker.rack |  |
| connections.max.idle.ms |  |
| connections.max.reauth.ms |  |
| controlled.shutdown.enable |  |
| controlled.shutdown.max.retries |  |
| controlled.shutdown.retry.backoff.ms |  |
| controller.socket.timeout.ms |  |
| default.replication.factor |  |
| delegation.token.expiry.time.ms |  |
| delegation.token.master.key |  |
| delegation.token.max.lifetime.ms |  |
| delete.records.purgatory.purge.interval.requests |  |
| fetch.max.bytes |  |
| fetch.purgatory.purge.interval.requests |  |
| group.initial.rebalance.delay.ms |  |
| group.max.session.timeout.ms |  |
| group.max.size |  |
| group.min.session.timeout.ms |  |
| inter.broker.listener.name |  |
| inter.broker.protocol.version |  |
| log.cleaner.backoff.ms |  |
| log.cleaner.dedupe.buffer.size |  |
| log.cleaner.delete.retention.ms |  |
| log.cleaner.enable |  |
| log.cleaner.io.buffer.load.factor |  |
| log.cleaner.io.buffer.size |  |
| log.cleaner.io.max.bytes.per.second |  |
| log.cleaner.max.compaction.lag.ms |  |
| log.cleaner.min.cleanable.ratio |  |
| log.cleaner.min.compaction.lag.ms |  |
| log.cleaner.threads |  |
| log.cleanup.policy |  |
| log.index.interval.bytes |  |
| log.index.size.max.bytes |  |
| log.message.format.version |  |
| log.message.timestamp.difference.max.ms |  |
| log.message.timestamp.type |  |
| log.preallocate |  |
| log.retention.check.interval.ms |  |
| max.connections |  |
| max.connections.per.ip |  |
| max.connections.per.ip.overrides |  |
| max.incremental.fetch.session.cache.slots |  |
| num.partitions |  |
| password.encoder.old.secret |  |
| password.encoder.secret |  |
| principal.builder.class |  |
| producer.purgatory.purge.interval.requests |  |
| queued.max.request.bytes |  |
| replica.fetch.backoff.ms |  |
| replica.fetch.max.bytes |  |
| replica.fetch.response.max.bytes |  |
| replica.selector.class |  |
| reserved.broker.max.id |  |
| sasl.client.callback.handler.class |  |
| sasl.enabled.mechanisms |  |
| sasl.jaas.config |  |
| sasl.kerberos.kinit.cmd |  |
| sasl.kerberos.min.time.before.relogin |  |
| sasl.kerberos.principal.to.local.rules |  |
| sasl.kerberos.service.name |  |
| sasl.kerberos.ticket.renew.jitter |  |
| sasl.kerberos.ticket.renew.window.factor |  |
| sasl.login.callback.handler.class |  |
| sasl.login.class |  |
| sasl.login.refresh.buffer.seconds |  |
| sasl.login.refresh.min.period.seconds |  |
| sasl.login.refresh.window.factor |  |
| sasl.login.refresh.window.jitter |  |
| sasl.mechanism.inter.broker.protocol |  |
| sasl.server.callback.handler.class |  |
| security.inter.broker.protocol |  |
| ssl.cipher.suites |  |
| ssl.client.auth |  |
| ssl.enabled.protocols |  |
| ssl.key.password |  |
| ssl.keymanager.algorithm |  |
| ssl.keystore.location |  |
| ssl.keystore.password |  |
| ssl.keystore.type |  |
| ssl.protocol |  |
| ssl.provider |  |
| ssl.trustmanager.algorithm |  |
| ssl.truststore.location |  |
| ssl.truststore.password |  |
| ssl.truststore.type |  |
| zookeeper.clientCnxnSocket |  |
| zookeeper.ssl.client.enable |  |
| zookeeper.ssl.keystore.location |  |
| zookeeper.ssl.keystore.password |  |
| zookeeper.ssl.keystore.type |  |
| zookeeper.ssl.truststore.location |  |
| zookeeper.ssl.truststore.password |  |
| zookeeper.ssl.truststore.type |  |
| alter.config.policy.class.name |  |
| alter.log.dirs.replication.quota.window.num |  |
| alter.log.dirs.replication.quota.window.size.seconds |  |
| authorizer.class.name |  |
| client.quota.callback.class |  |
| connection.failed.authentication.delay.ms |  |
| create.topic.policy.class.name |  |
| delegation.token.expiry.check.interval.ms |  |
| kafka.metrics.polling.interval.secs |  |
| kafka.metrics.reporters |  |
| listener.security.protocol.map |  |
| log.message.downconversion.enable |  |
| metric.reporters |  |
| metrics.num.samples |  |
| metrics.recording.level |  |
| metrics.sample.window.ms |  |
| password.encoder.cipher.algorithm |  |
| password.encoder.iterations |  |
| password.encoder.key.length |  |
| password.encoder.keyfactory.algorithm |  |
| quota.window.num |  |
| quota.window.size.seconds |  |
| replication.quota.window.num |  |
| replication.quota.window.size.seconds |  |
| security.providers |  |
| ssl.endpoint.identification.algorithm |  |
| ssl.engine.factory.class |  |
| ssl.principal.mapping.rules |  |
| ssl.secure.random.implementation |  |
| transaction.abort.timed.out.transaction.cleanup.interval.ms |  |
| transaction.remove.expired.transaction.cleanup.interval.ms |  |
| zookeeper.ssl.cipher.suites |  |
| zookeeper.ssl.crl.enable |  |
| zookeeper.ssl.enabled.protocols |  |
| zookeeper.ssl.endpoint.identification.algorithm |  |
| zookeeper.ssl.ocsp.enable |  |
| zookeeper.ssl.protocol |  |
| zookeeper.sync.time.ms |  |

# 监控

# Connect

## 概述

## 配置

|  |  |
| --- | --- |
| 配置项 | 说明 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# 应用