

腾讯云消息队列 CKafka

产品简介

产品文档



腾讯云

【版权声明】

©2013-2017 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

【商标声明】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或模式的承诺或保证。

文档目录

文档声明.....	2
产品简介.....	4
CKafka产品概述.....	4
CKafka技术原理.....	7
CKafka产品对比.....	10
服务等级协议（SLA）.....	12

产品简介

CKafka产品概述

产品概述

Cloud Kafka是基于腾讯自研CMQ消息队列引擎，提供高吞吐性能、高可扩展性的消息队列服务。Cloud Kafka完美兼容Apache kafka 0.9版本接口，在性能、扩展性以及业务安全保障、运维等方面具有超强优势，让您在享受低成本、超强功能的同时，免除繁琐运维工作。

应用场景

1. 网页追踪

Cloud Kafka通过实时处理网站活动（PV，搜索，用户其他活动等），并根据类型发布到topic中，这些信息流可以被用于实时监控或离线统计分析等。

由于每个用户的page view中会生成许多活动信息，因此网站活动跟踪需要很高的吞吐量，Cloud Kafka可以完美满足高吞吐、离线处理等要求。

2. 日志聚合

Cloud Kafka 提供了低延迟处理、易于支持多个数据源和分布式的数据处理（消费）的特性。相比于中心化的日志聚合系统，Cloud Kafka

可以在提供同样性能的条件下，实现更强的持久化保证以及更低的端到端延迟。

因此，Cloud Kafka 的特性决定它非常适合作为“日志收集中心”；多台主机/应用可以将操作日志“批量”“异步”的发送到 Cloud Kafka 集群中，而无需保存在本地或者 DB 中；Cloud Kafka

可以批量提交消息/压缩消息，这对生产者而言，几乎感觉不到性能的开支。此时

消费者可以使hadoop 等其他系统化的存储和分析系统对拉取日志进行统计分析。

3. 大数据场景

对于一些大数据相关的业务场景而言，需要对大量并发数据进行处理和汇总，此时对集群的处理性能和扩展性都有很高的要求。而Cloud Kafka在实现上，其数据分发机制，磁盘存储空间的分配、消息格式的处理、服务器选择以及数据压缩等方面，也决定其适合处理海量的实时消息，并能汇总分布式应用的数据，方便系统运维。

在具体的大数据场景中，Cloud

Kafka能够很好地支持离线数据、流式数据的处理，并能够方便地进行数据聚合、分析等操作。

优势

1.解耦

有效解耦生产者、消费者之间的关系。在确保同样的接口约束的前提下，允许独立扩展或修改生产者/消费者间的处理过程。

2.可扩展性

由于消息的处理过程被解耦，只需要水平扩展处理过程，即可有效增加消息的入队效率和处理效率，十分灵活。

3.削峰填谷

消息队列能够抵挡突增的访问压力，而不会因为突发的超负荷的请求而完全崩溃，有效提升系统健壮性。

4.可恢复性

当系统的部分组件出现故障时，整个系统不会因此收到影响，增加了系统的容错能力。及时某一处理消息的进程故障，队列中的消息依然可以在系统恢复后被处理。

5.顺序读写

Cloud Kafka能够保证一个Partition内消息的有序性，和大部分的消息队列一致，Cloud Kafka可以保证数据按照顺序进行处理，极大提升磁盘效率。

6.异步通信

在业务无需立即处理消息的场景下，Cloud Kafka提供了消息的异步处理机制，访问量高时仅将消息放入队列中，在访问量降低后再对消息进行处理，缓解系统压力。

名词解释

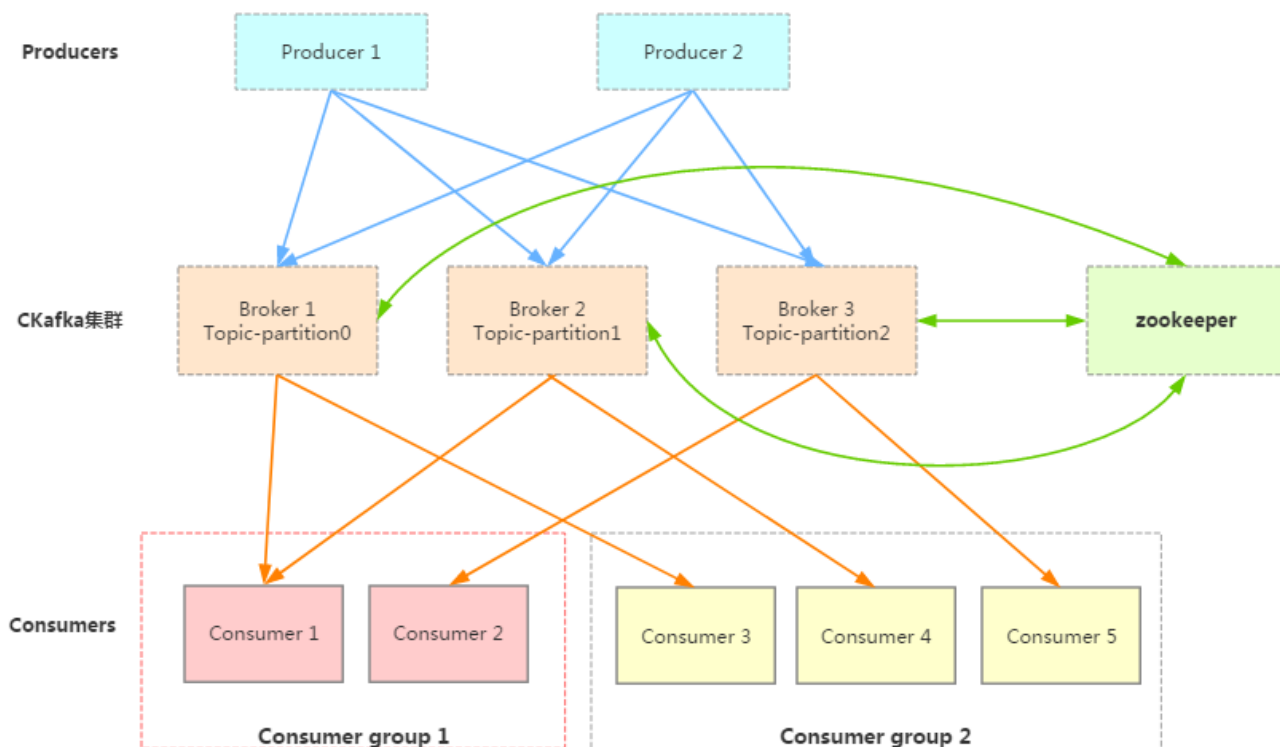
序号	名称	解释
1	Broker	Cloud Kafka集群中的服务器
2	Topic	消息类别，Cloud Kafka是面向消息的
3	Partition	一个物理上分区的概念，一个Topic

序号	名称	解释
		可以包含一个或者多个partition，Cloud Kafka以partition作为分配单位
4	Replica	partition 的副本，用于保障 partition 的高可用
5	Offset	消息在partition的唯一序号
6	Producer	生产者，负责发布消息
7	Consumer	消费者，从集群中消费消息
8	Consumer group	消费者分组，每个consumer必须属于一个consumer group。每条消息能被多个consumer group消费，但只能被该group中的一个consumer消费
9	Zookeeper	用于存储集群的meta数据、进行leader选举、故障容错等

CKafka技术原理

技术原理

Cloud Kafka的架构图如下所示：



一个典型的Cloud Kafka集群如上所示。其中的生产者Producer可能是网页活动产生的消息、或是服务日志等信息。生产者通过push模式将消息发布到Cloud

Kafka的Broker集群，消费者通过pull模式从broker中消费消息。消费者Consumer被划分为若干个Consumer Group，此外，集群通过Zookeeper管理集群配置，进行leader选举，故障容错等。

高吞吐的实现

Cloud Kafka中存在大量的网络数据持久化到磁盘和磁盘文件通过网络发送的过程。这一过程的性能直接影响Kafka的整体吞吐量，主要通过以下几点实现。

1. 高效使用磁盘
2. 磁盘中顺序读写数据，提高磁盘利用率
 - 写message
 - 消息写到page cache，由异步线程刷盘
 - 读message

消息直接从page cache转入socket发送出去

当从page cache没有找到相应数据时，此时会产生磁盘IO,从磁盘加载消息到page cache,然后直接从socket发出去

3. Broker的零拷贝(Zero Copy)机制

使用sendfile系统调用，将数据直接从页缓存发送到网络上

4. 减少网络开销

5. 数据压缩降低网络负载

6. 批处理机制：Producer批量向Broker写数据、Consumer批量从Broker拉数据

数据持久化

Cloud Kafka的数据持久化主要通过如下原理实现：

1. topic中partition存储分布

在Cloud Kafka文件存储中，同一topic有多个不同partition，每个partition在物理上对应一个文件夹，用户存储该partition中的消息和索引文件。例如，创建两个topic，topic1中存在5个partition，topic2中存在10个partition，则整个集群上会相应生成5+10=15个文件夹。

2. partition中文件存储方式

partition物理上由多个segment组成，每个segment大小相等，顺序读写，快速删除过期segment，提高磁盘利用率。

水平扩展Scale Out

一个Topic可分多个Partition, 分布在一个或多个Broker上

一个消费者可订阅其中一个或者多个Partition

Producer负责将消息均衡分配到对应的Partition

Partition内消息有序的

Consumer Group设计

Cloud Kafka不删除已消费的消息

任何consumer必须属于一个group

同一Consumer Group中的多个Consumer不同时消费同一个partition
不同Group同时消费同一条消息，多元化（队列模式、发布订阅模式）

多副本

为何需要多副本设计？

增强系统可用性、可靠性

Replica均匀分布到整个集群

Replica的算法如下：

- 1、将所有Broker（假设共 n 个Broker）和待分配的Partition排序
- 2、将第 i 个Partition分配到第 $(i \bmod n)$ 个Broker上
- 3、将第 i 个Partition的第 j 个Replica分配到第 $((i + j) \bmod n)$ 个Broker上

Leader Election选举机制

Cloud Kafka在Zookeeper中动态维护了一个ISR（in-sync replicas），
ISR里的所有Replica都跟上了leader
只有ISR里的成员才有被选为Leader的可能。

ISR中 $f+1$ 个Replica，一个Partition能在保证不丢失已commit的消息的前提下
容忍 f 个Replica的失败。

共有 $2f+1$ 个Replica（包含Leader和Follower），commit之前必须保证有 $f+1$ 个
Replica复制完消息，为了保证正确选出新的Leader，fail的Replica不能超过 f 个。

CKafka产品对比

Cloud Kafka与其他消息服务产品的性能对比详情如下：

特性	CKafka	Apache Kafka	RabbitMQ	RocketMQ	CMQ
优点	吞吐量非常大 扩展性非常灵活 运维成本极低	吞吐量大	可靠性高	可靠性高	可靠性非常高 金融等强一致性场景
缺点	极端情况可能丢失消息	可能丢失消息 扩展性不够灵活 依赖组件多，运维量大 安全防护功能有限，隔离和兼容性差	性能较差 扩展不灵活	HA切换需要手动支持，不能自动化	为保证强一致性，吞吐量一般
开发语言	C++	Scala	Erlang	Java	C++
可扩展性	非常灵活、易于扩展，发送消息只需指明vip地址，broker的变化对于收发消息都透明	不够灵活，发送消息需指明broker地址，接收消息需zookeeper协调调度	不够灵活，发送消息需要指明broker地址	较灵活，发送方、接收方和name server连接	灵活、平滑、水平扩展，逻辑上单个Queue可跨多个集群提供服务
吞吐量	非常大	较大	一般	一般	一般
常规性能	百万级QPS	百万级QPS	十万级QPS	十万级QPS	十万级QPS
2C 4GB压测	读写22万 QPS	读写20万 QPS	读写 10万 QPS	读写 10万 QPS	读写12万 QPS
同步算法	ISR (Replica)	ISR (Replica)	GM	同步双写	Raft
可用性	可用性很高，主从自动切换，腾讯云消息服务承诺可用性99.95%	可用性高，主从自动切换，但由于异步刷盘和复制，切换后可能会丢失消息	主备自动切换，用mirror queue支持m/s，master提供服务，slave仅备份	不支持主从自动切换，master不可用时slave只读不写	可用性很高，broker中存在2节点即可提供高可用服务
消费方式	拉取方式	拉取方式	拉取和推送方式	拉取方式	拉取和推送方式
消息可靠性	较高 可通过三副本方式提升可靠性，集群容灾性能好，故障情况极少发生	低 broker只有异步刷盘机制并主备只有异步复制，可能会导致丢失部分消息	高 发送消息时，指定消息为持久化就会写入到磁盘	高 broker同步双写，主备都写成功才返回成功	极高 保证消息不丢失同步刷盘，数据持久性99.999999%

特性	CKafka	Apache Kafka	RabbitMQ	RocketMQ	CMQ
		息			
数据校验	CRC	CRC	无	CRC	checksum
消息回溯	支持	支持	不支持	不支持	支持
安全防护	支持	不支持	不支持	不支持	支持
监控告警	支持	不支持	不支持	不支持	支持
服务支持	支持	不支持	不支持	不支持	支持

注：“2C 4GB 压测”表示在 2核 4GB 内存服务器上压测的结果。

服务等级协议 (SLA)

1 腾讯云CKafka消息服务

CKafka (Cloud Kafka) 是一个分布式的、高吞吐量、高可扩展性的消息系统，兼容开源 Kafka API (0.9 0.10版本)。Ckafka

基于发布/订阅模式，通过消息解耦，使生产者和消费者异步交互，无需彼此等待。Ckafka 具有数据压缩、同时支持离线和实时数据处理等优点，适用于日志压缩收集、监控数据聚合等场景。

2 服务保障指标

腾讯云为您所购买的云服务制定服务等级指标，并向您承诺提供数据管理和业务质量方面最大程度的保障。同时，腾讯云有权根据变化适时对部分指标作出调整。若无特殊约定，本条款中的“月”均指30个自然日，按自然月计算。

2.1 CKafka消息服务

2.1.1 数据存储的持久性

每月您申请的 CKafka 的数据存储持久性为

99.999999%

。

2.1.2 数据可销毁性

在您要求删除数据或设备在弃置或转售前，腾讯云将采取磁盘低级格式化操作彻底删除您所有数据，并无法复原，硬盘到期报废时将进行消磁。

2.1.3 数据知情权

目前用户 CKafka 服务部署在六大数据中心，分别是：上海数据中心、广州数据中心、北京数据中心、成都数据中心、上海金融数据中心、深圳金融数据中心。

帮助用户选择网络条件最好的数据中心存储数据，用户购买云服务器时刻选择归属的地域（广州、上海、北京、成都、上海金融、深圳金融）。

用户使用已知的数据中心均遵守的当地的法律和中华人民共和国相关法律。

用户所有数据腾讯云不会提供给任意第三方，除政府监管部门监管审计需要。

2.1.4 数据可审查性

腾讯云在依据现有法律法规体系下，出于配合政府监管部门的监管或安全取证调查等原因的需要，在符合流程和手续完备的情况下，可以提供云服务器相关信息，包括关键组件的运行日志、运维人员的操作记录、用户操作记录等信息。

2.1.5 业务可用性

CKafka消息服务承诺

99.95%

的业务可用性，即用户每月可用时间应为

$30 \times 24 \times 60 \times 99.95\% = 43178.4$ 分钟

，即存在

$43200 - 43178.4 = 21.6$ 分钟

的不可用时间，其中业务不可用的统计单元为用户单业务实例。

业务故障的恢复正常时间

5分钟

以下，不计入业务不可用性计算中，不可用时间指业务发生故障开始到恢复正常使用的时间，包括维护时间。

超过5分钟的，纳入不可用时间。

2.1.6 故障恢复能力

腾讯云 CKafka 具备故障迁移能力，可在物理服务器故障发生时，自动将业务迁移至新的母机，无需用户参与，保证客户服务的连续性。同时，腾讯云提供专业团队

7x24小时

帮助维护。

3 服务计量准确性

腾讯云服务的费用在用户的选购页面和订单页面均有明确展示，用户可自行选择具体服务类型并按列明的价格进行购买。具体价格以腾讯云官网公布的价格为准，腾讯云按照用户选购的云服务规格和使用时长进行收费。

4 补偿

4.1 适用范围

适用于由腾讯云故障原因导致用户的服务不可正常使用或完全不可访问，以及腾讯云故障导致的网站（开发者服务网站）无法访问时，用户要求腾讯云针对事故/故障而进行的补偿。

4.2 补偿标准原则

故障时间 = 故障解决时间 - 故障开始时间。按分钟计算故障时间，故障时间小于1分钟的按

1分钟

计算。

例如，故障时间为

1分01秒

, 按

2分钟

算。

CKafka消息服务故障百倍赔偿：

后付费：赔偿方式是补偿现金券，

现金券金额 = 故障实例每天费用 ÷ 24 ÷ 60 × 故障时间（以分钟为单位）× 100

。赔偿现金券的上限，不超过 CKafka 服务的总费用。