腾讯云数据库 HBase

产品白皮书

[2016.06.26]

[V1.0]

腾讯云

【版权声明】

©2015-2016 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】

腾讯云

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。

目录

版权所有：腾讯云计算（北京）有限责任公司

第一章 前言

数据库是互联网服务的一个重要组成部分，存储了大量的资料和数据。随着互联网行业的高速发展，对数据库的需求也大量增加，数据容量也呈指数上升。在传统数据库应用中，一般都存在设备利用率低，资源池管理困难，故障切换和迁移对业务不够透明，无法按需部署，扩容建设周期长等问题。

随着云计算技术的不断成熟和腾讯云基础产品服务的不断发展，针对业界有潜力的数据库 HBase 在传统数据库应用的痛点和难点，腾讯云推出了云数据库 HBase。其包括了高性能、高可靠的服务，整合了自动化管理工具，最大程度减少开发人员在部署、监控、扩容和故障恢复等方面的投入，使开发者可以集中精力进行产品开发和运营。

第二章 产品定义

2.1 云数据库介绍

云数据库 HBase（Cloud HBase Service）是腾讯云基于全球广受欢迎的 HBase 打造的高性能，可伸缩，面向列的分布式存储系统，100%完全兼容 HBase 协议， 适用于写吞度量大，海量数据存储以及分布式计算的场景。为您提供稳定，丰富的集群管理，弹性可扩展的数据库服务

云数据库 HBase 主要具有以下一些特点：

1）支持节点数量按需配置，计算节点和存储节点可视化一键部署

2）集群管理与监控，可视化的数据指标趋势使监控更加立体化。

3）集群扩展性，任务调度集群下发运维操作指令各个节点以实现运维自动化，实现节点的快速扩缩容。

版权所有：腾讯云计算（北京）有限责任公司

4）服务高可用，存储层 HDFS 高可用 HA， 单点 出现故障，可以迅速切换至备节点，从而实现不间断对外提供服务。

2.2 云数据库优势

云数据库 HBase 作为一种服务提供给用户 ,使它相对于自建 Hbase 数据库更容易部署、管理和扩展，详见下表：

维度	云数据库 HBase	自建 HBase
自动化部署	支持节点数量按需配置，计算节点和存储节点可视化一键部署，您不需关注内部细节，省去了部署工作环节，为您节省 60% 的开发时间，您可以将更多的经历投入在业务之中	自行安装，部署多节点，自行调试运行，招聘专业的 DBA 搭建环境，时间和人力成本高
服务可用性	为了提供最佳的数据持久性和可用性，腾讯云 HBase 将会自动检测并替换您的数据集群中的任何故障节点。替换节点可立即使用，为您尽快地恢复数据查询，保证集群正常运行 ,您完全不需要做任何处理，服务会自动切换，不影响业务	HBase 自身环境复杂 ,遇到故障处理的难度大，恢复不及时影响业务数据
数据可靠	底层分布式文件系统 HDFS，可提供冗余存储多份数据（默认为 3 份数据）来保证	自行保障，依赖硬件的故障发生率，依赖技术人员的数据库管理水平

性	数据可靠性。您完全不用担心数据会丢失	
丰富的监控	集群状态，节点流量，读写请求量以及存储容量等关键数据指标可视化监控，让您完全掌握运行状态，同时提前规避风险	需自行开发数据库监控系统，运维人员需半夜处理故障
运维省心	用户无需要关心 HBase 运行过程中的故障处理，版本升级，节点的增删的操作，以上由云数据库运营团队全面负责。您只需关注业务数据的读出和写入。	自行操作增删节点，开发多系统保证节点稳定运行，手动备份数据，自行实现数据恢复

2.3 应用场景

- **构建海量数据存储系统**

  适用于 TB 级以上的数据存储，动态扩展节点应对持续增长的数据存储量。仅需在管理控制台中点击操作一下，就能在性能或容量需要改变时，轻松改变集群节点数或节点类型，方便您构建海量存储系统

- **构建分布式计算平台**

  构建大数据管理与分析平台，可以轻松分析您的所有数据，无需担心大数据存储和计算瓶颈，您可以将更多精力投入数据分析和挖掘，如以下

  - 搭建企业 BI 以及报表系统
  - 分析多种产品的用户行为数据，挖掘用户潜在需求
  - 分析产品订单交易数据

- 分析移动产品或者站点广告流量和点击量
- 分析和汇总游戏内玩家和道具数据
- 评估车辆网领域的车辆地理位置、运营效率和最佳调度信息

第三章 产品功能

3.1 功能介绍

云数据库 HBase 本质上是开源数据库 HBase 的云化服务，不仅具有传统自建 HBase 的功能，而且也具有一些特有的功能，主要有以下几点：

1. 轻松管理数据库

   提供命令行和 Web 两种方式管理云数据库，并支持批量数据库的管理、权限设置。

2. 自动容灾功能

   热备架构，当主节点挂掉，备节点会自动接管服务；使用过程中不用担心服务终端

3. 专业的监控与告警

   多维度监控，自定义资源阈值告警，提供慢查询分析报告和服务器运行日志下载。

4. 集群扩展

   任务调度集群下发运维操作指令给各个节点以实现运维自动化，实现节点的快速扩缩容。

3.2 功能列表

特性	一级子特性	二级子特性	描述
实例管理	使用向导		Web 控制台提供数据库使用帮助文档
	新建实例	地域设置	可选择实例所在地域
		可用区设置	可根据地域选择可用区
		网络设置	可选择基础网络或者 VPC 网络
		配置类型设置	可选择高IO版和大容量版(具体地域售卖类型以页面展示为准)；
		自动部署	一键部署，立即生效
		节点数量设置	可选择节点数量
		内存设置	可根据性能或者负载能力选择相应规格
		硬盘设置	可根据性能或者数据量选择相应规格
		项目设置	可选择实例所属项目，便于分项目管理
	实例列表	显示实例信息	显示实例名称、运行状态、所属项目、可用区、网络类型、内网地址、集群类型、配置类型、内存、容量、创建时间、到期时间
		批量操作	续费、自动续费设置（开发中）
		单实例操作	初始化，数据访问
		单实例学费	续费（开发中）
实例访问	内网访问		提供内网 IP/PORT

版权所有：腾讯云计算（北京）有限责任公司

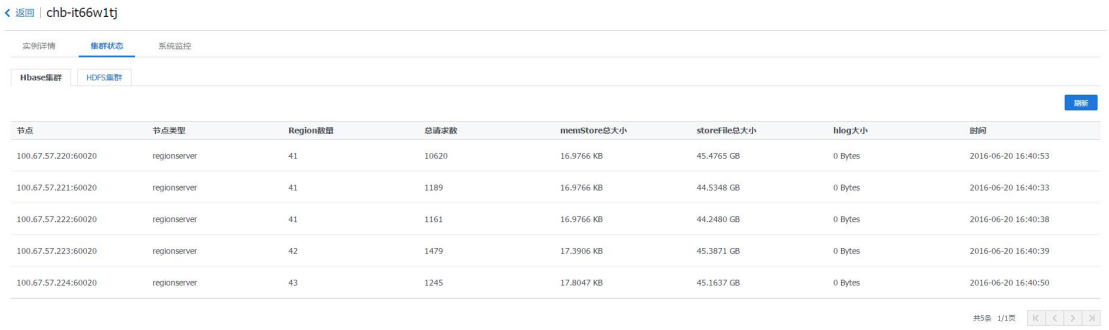
	外网访问		提供外网访问地址，需手动开通
实例监控	监控指标选择	集群监控	1. Region 数量 2. 总请求数 3. memstore 大小 4. storefile 总大小
		系统监控	1. Regionserver CPU 使用率 2. Regionserver 读请求次数 3. Regionserver 写请求次数 4. 主节点 CPU 使用率 5. 数据节点 CPU 使用率 6. 读流量 7. 写流量

版权所有：腾讯云计算（北京）有限责任公司

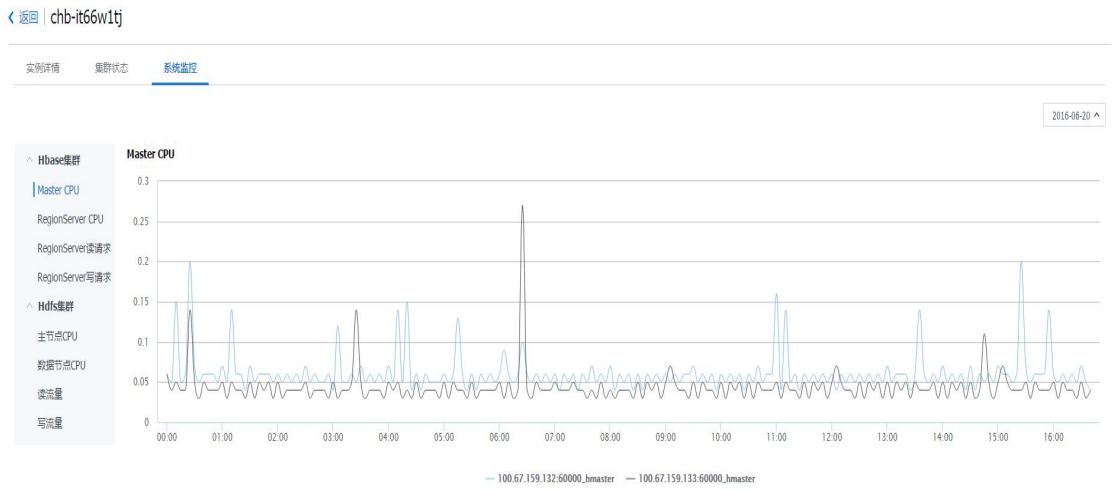
3.3 产品示意图

1. 实例列表

2. 集群状态

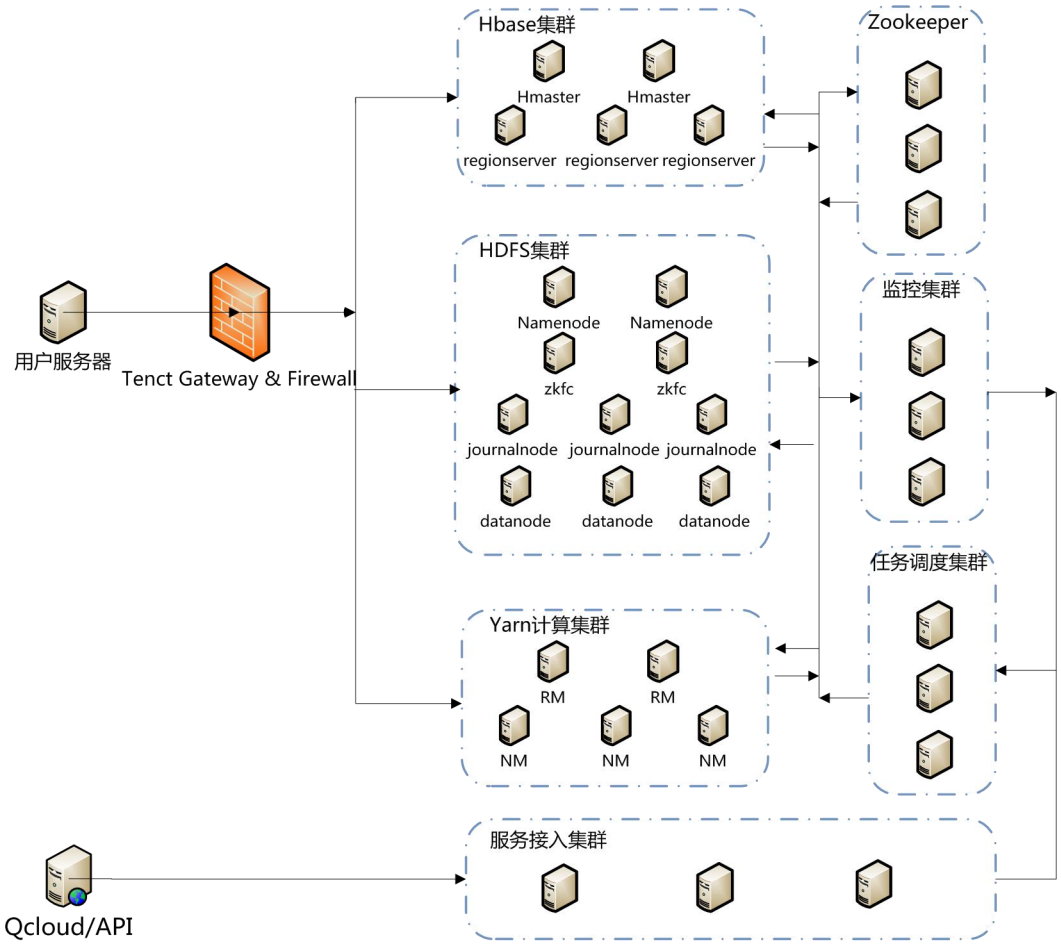


3. 系统监控



第四章 系统架构及技术原理

4.1 系统架构



云数据库 Hbase 系统包括如下几大模块：

4.1.1 接入模块：Tencent Gateway & firewall

用于云数据库 Hbase 的整体接入，主要屏蔽 IP/PORT 的变化，使用户无感知，对业务逻辑透明，并针对未授权的访问进行隔离和管控

4.1.2 Hbase 集群

Hbase 集群 Hmaster 节点采用一主一备来提供高可用，通过 zookeeper 分布式协调器来监控和切换主备，整个系统不存在单点问题以达到整个系统的高可用，regionserver 节点采用 2N+1 的方式部署 ,通过 zookeeper 来监控节点的服务状态，当一个 regionserver 下线的时候其负责的 region 自动托管到其他服务器，同时 regionserver 保证了对 HDFS 的 namenode 的 HA，保证在 HDFS 的主 namenode 发生故障的时候，regionserver 在秒级对 namenode 切换感知到达高可用。

4.1.3 HDFS 集群

HDFS 集群搭建的软件版本为社区版 hadoop-2.6.4,为了保证整个集群的高可用性，采用双 namenode、2N+1 个 journalnode、双 zfkc、2N+1 个数据节点来部署整个集群，其中 zkfc 负责对 namenode 的 Failover，journalnode 负责 namnode 的 editorlog 同步，同时为了数据安全，建议在购买集群的时候数据备份最小为 3 备份，存储节点最小为 5 个节点。

4.1.4 YARN 计算集群

YARN 计算集群是选择性的功能，主要功能是在 YARN 框架下运行 hadoop 的 MR 任务，可用于数据分析、离线数据计算等，该集群采用双 resourcemanager 方式部署，通过分布式协调器 zookeeper 实现自动 HA 来保证高可用

版权所有：腾讯云计算（北京）有限责任公司

4.1.5 监控集群

监控集群负责对 hbase 集群、hdfs 集群、yarn 集群的健康情况监控以及监控数据的采集和处理，实现了节点故障秒级发现和以分钟为单位的监控数据采集，并最终以图表的形式给 web 前端提供监控数据服务。

4.1.6 任务调度集群

任务调度集群负责处理服务集群生成的业务逻辑任务并把结果反馈给服务集群，同时还接受运维系统下发的运维操作并形成运维指令下发给 hbase 集群、hdfs 集群、yarn 集群的各个节点以实现运维自动化，实现节点的快速扩缩容。

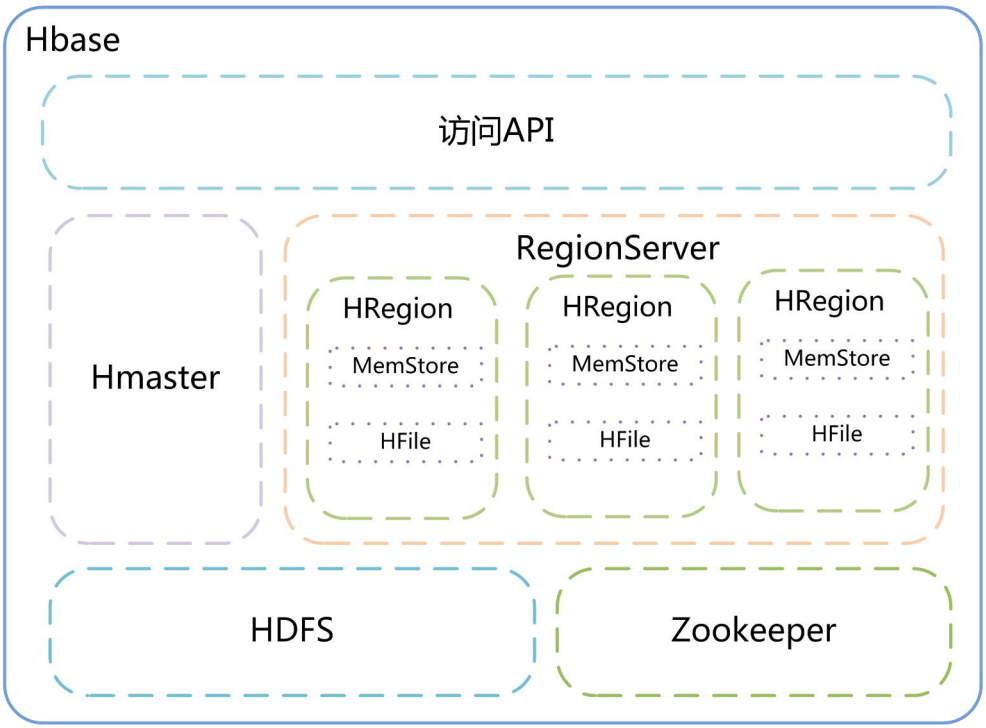
4.1.7 zookeeper 集群

分布式协调器，实现 hmaster、namenode 以及 resourcemanager 的自动 HA，保证集群的高可用，zookeeper 集群最小节点数为 2N+1 其中 N 大于等于 1

版权所有：腾讯云计算（北京）有限责任公司

4.2 高可用技术原理

4.2.1 Hbase 高可用

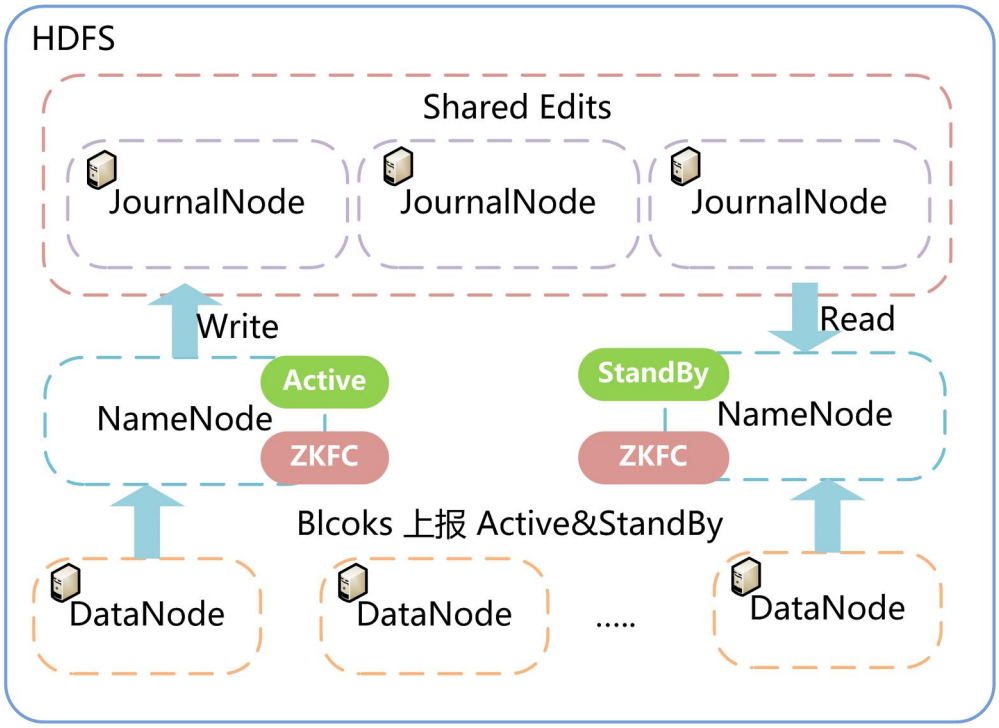


如上图，HBase 在完全分布式环境下，由 Master 进程负责管理 RegionServers 集群的负载均衡以及资源分配，ZooKeeper 负责集群元数据的维护并且监控集群的状态以防止单点故障，每个 RegionServer 会负责具体数据块的读写，HBase 所有的数据存储在 HDSF 系统上，对 Hmaster 高可用，在 Hmaster（可能有多个节点）启动后会去分布式协调器的根路径（依赖于 hbase 的配置文件默认为/hbase）下创建 master 节点如果创建成功则当前节点为 active 节点 ,其余节点为 standby 状态，当 active 状态的节点死掉的时候会因为和 zookeeper 链接超时导致注册的节点失效而其他节点会继续抢注该节点而成为新的 active 节点。

对于 regionserver 节点死机后 zookeeper 节点会感知到立刻通知 master 进行 RS 死机处理把宕机节点的 region 重新指派给新的 regionserver 进行托管，数据安全

性是由 HDFS 来保证真的，同时用户也可以根据自己的需要来设定数据的安全程度，建议 HDFS 数据备份数为 3，在写 WAL 日志的时候默认为强制同步写，用户也可以设置为异步写，风险是 regionserver 宕机的时候存在丢数据的可能，zookeeper 集群作为一个分布式协调器支持平行扩展，部署节点为 2N+1 个节点，建议部署 5 个以上的节点。

4.2.2 HDFS 集群高可用



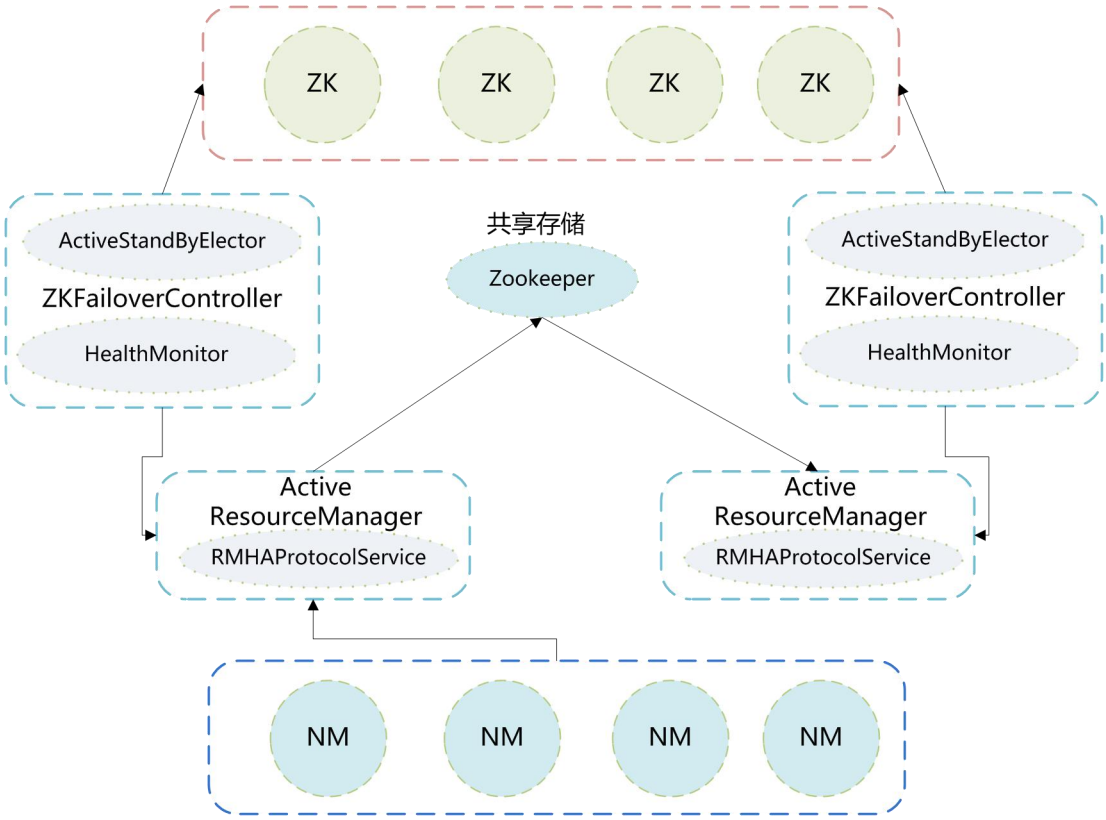
如上图为存储层 HDFS 高可用 HA，HA 即为 High Availability,用于解决 NameNode 单点故障问题,该特性通过热备的方式为主 NameNode 提供一个备用者,一旦主 NameNode 出现故障,可以迅速切换至备 NameNode, 从而实现不间断对外提供服务,在该方案中 HA 的 namenode 节点通常由两个 NameNode 组成,一个处于 active 状态, 另一个处于 standby 状态。Active NameNode 对外提供服务,比如处理来自客户端的 RPC 请 求,而 Standby NameNode 则不对外提供服务,仅同步 active namenode 的状态,以

便能够在 它失败时快速进行切换。

为了能够实时同步 Active 和 Standby 两个 NameNode 的元数据信息(实际上 editlog), 需提供一个共享存储系统,可以是 NFS、QJM(Quorum Journal Manager)或者 Zookeeper, Active Namenode 将数据写入共享存储系统,而 Standby 监听该系统,一旦发现有新数据写 入,则读取这些数据,并加载到自己内存中,以保证自己内存状态与 Active NameNode 保持 基本一致,如此这般,在紧急情况下 standby 便可快速切为 active namenode , 主备 NameNode 之间通过一 组 JournalNode 同步元数据信息,一条数据只要成功写入多数 JournalNode 即认为写入成功。 通常配置奇数个(2N+1)个 JournalNode,这样,只要 N+1 个写入成功就认为数据写入成功, 此时最多容忍 N-1 个 JournalNode 挂掉,比如 3 个 JournalNode 时,最多允许 1 个 JournalNode 挂掉,5 个 JournalNode 时,最多允许 2 个 JournalNode 挂掉。

版权所有 : 腾讯云计算 (北京) 有限责任公司

4.2.3 YARN 集群高可用

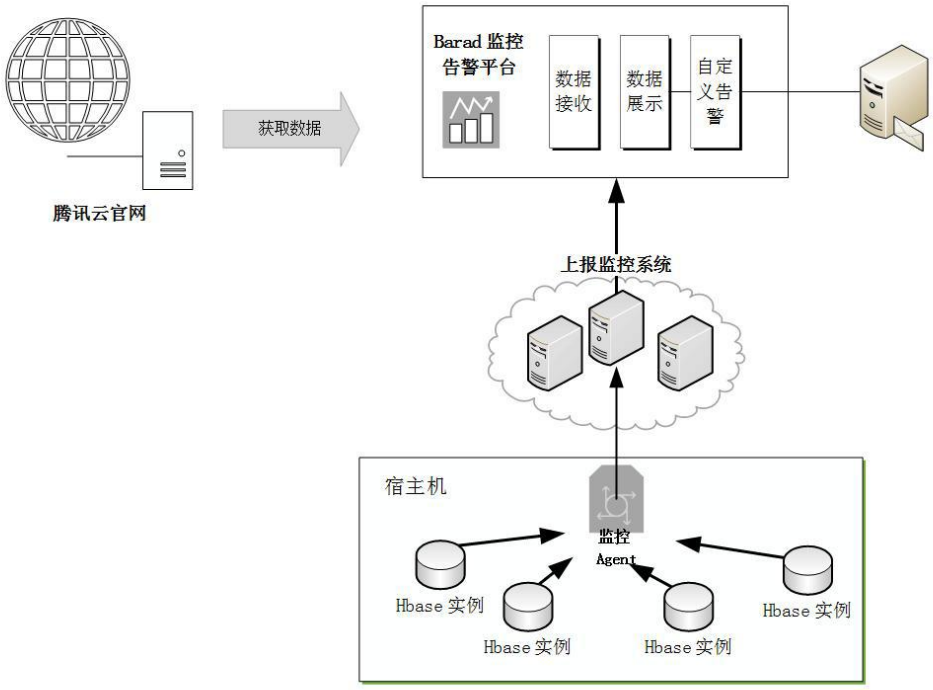


如上图，YARN 集群高可用主要是针对 ResourceManager 做容灾处理，ResourceManager 负责整个系统的资源管理和调度，内部维护了各个应用程序的 ApplictionMaster 信息，NodeManager 信息，资源使用信息等。考虑到这些信息绝大多数可以动态重构，因此解决 YARN 单点故障要比 HDFS 单点容易很多。与 HDFS 类似，YARN 的单点故障仍采用主备切换的方式完成，不同的是，备节点不会同步主节点的信息，而是在切换之后，才从共享存储系统读取所需信息。之所以这样，是因为 YARN ResourceManager 内部保存的信息非常少，大部分可以重构，且这些信息是动态变化的，很快会变旧，它借助 Zookeeper 完成主备节点信息共享。它仅在 Zookeeper 上保存 Application ID 和 ApplicationAttempt ID，以便故障恢复后重新创建这些 Application，其他信息则动态重构或者丢弃，比如 NodeManager 信息（包括可用资

源，健康状态等信息），需由 NodeManager 重新通过 RPC 汇报，而资源使用信息（每个节点资源使用情况，每个任务资源获取情况等）则全部重置，也就是说，故障恢复或者 ResourceManager 主备切换后，整个集群跟重启过一样，只不过是之前正在运行的应用程序不需要重新提交，但已经分配的资源信息则全部丢失。这意味着，NodeManager 重新跟新的 ResourceManager 连接后，ResourceManager 发送的第一个指令是让 NodeManager 重启，而 NodeManager 会杀死所有正在运行的 Container，外，需要补充说明的是，YARN ResourceManager 只负责 ApplicationMaster 的状态维护和容错，ApplicationMaster 内部管理和调度的任务，比如 MapTask 和 ReduceTask 则需要由 AppicationMaster 自己容错 这不属于 YARN 这个系统管理的范畴。比如 MapReduce 的 ApplicationMaster—MRAppMaster 会在 HDFS 上记录 Task 运行日志，这样，当它运行失败重新被调度到另外一个节点运行时，会重新从 HDFS 上读取日志，恢复已经运行完成的 Task，而只需为那些未运行完成的任务申请资源和二次调度（注意，之前正在运行的 Task 会被杀死，重新执行）

版权所有：腾讯云计算（北京）有限责任公司

4.3 监控与告警技术原理



云数据库 Hbase 监控提供全方位的监控数据和自定义告警功能，监控指标包括负载监控，访问统计，网络流量等重要指标。

监控数据通过部署在每台母机上的 Agent 进行定时采集，然后上报给数据中转节点，通过中转节点进行数据检查，汇总，然后批量上报给云监控系统 Barad，Barad 提供数据展示、数据查询 API 以及自定义告警等功能。

第五章 实例规格与性能说明

第六章 服务等级协议(SLA)

服务内容

腾讯云数据库 Hbase 是腾讯云基于全球最有潜力的开源 NoSQL 数据库 Hbase 专业打造的高性能分布式数据存储服务，100%完全兼容 Hbase 协议，适用于面向非关系型数据库的场景。

同时腾讯云数据库 Hbase 提供了高性能、高可靠、易用、便捷的 Hbase 集群服务，每一个实例都是数据至少 2 个备份 master 双主的形式搭建的，保证了用户数据高可用和系统的高稳定。

6.2 数据持久性

服务周期内(即用户购买的 Hbase 的服务期内)承诺每月用户申请实例的数据存储的持久性为 99.9996%。即用户每月每 1000000 个实例的存储的文件，每月只有 4 个实例有数据丢失的可能性

6.3 数据可销毁性

用户主动删除数据或用户服务期满后需要销毁数据，删除数据后或设备弃置、转售前腾讯云将采取磁盘低级格式化操作彻底删除用户所有数据，并无法复原，硬盘到期报废时将进行消磁。

6.4 数据知情权

A.数据存储的数据中心位置（可以通过提交工单进行咨询确认）。

B.数据备份数量以及备份数据存储的数据中心位置（可以通过提交工单进行咨询确认）。

C.帮助用户选择网络条件合适的数据中心存储数据，冷备则是根据资源利用情况动态分配，

版权所有：腾讯云计算（北京）有限责任公司

用户默认无需选择数据中心和冷备中心位置，如果需要选择，可以通过提交工单进行咨询确认。

D.数据中心要遵守的当地的法律和中华人民共和国相关法律（可提交工单进行咨询确认）。

E.用户所有数据不会提供给任意第三方，除政府监管部门监管审计需要。用户的行为日志会用于数据库运行状态的数据分析，但不会对外呈现用户个人信息数据。

6.5 数据私密性

腾讯云通过配置防火墙策略，采用白名单过滤机制进行网络隔离，通过各个 Hbase 集群相互独立的访问各自的 HDFS 存储来保证同一资源池用户数据互不可见。

6.6 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，可以提供数据库相关信息，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

6.7 服务可用性

A.腾讯云数据库 Hbase 承诺 99.95%的业务可用性。即单个数据库实例每个服务周期所有可用时间/服务周期不低于 99.95%。其中业务不可用的统计单元为用户单数据库实例。

B.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。

6.8.故障恢复能力

腾讯云提供专业团队 7x24 小时帮助维护。