

HBase安装部署

www.huawei.com





目标

- 学完本课程后，您将能够：
 - 掌握**HBase**部署规划
 - 掌握**HBase**常用配置参数
 - 掌握**HBase**容灾配置参数与相关命令



目录

1. HBase部署规划
2. HBase关键参数配置
3. HBase监控介绍
4. HBase容灾配置与命令

HBase组网方案

概念	说明
管理节点	Management Node (MN) ，用于安装 FusionInsight Manager ，即 FusionInsight HD 集群的管理系统。 FusionInsight Manager 对部署在集群中的节点及服务进行集中管理。
控制节点	Control Node (CN) ，控制节点控制监控数据节点执行存储数据、接收数据、发送进程状态及完成控制节点的公共功能。 FusionInsight HD 的控制节点包括 HMaster 、 HiveServer 、 ResourceManager 、 NameNode 、 JournalNode 、 SlapdServer 等。
数据节点	Data Node (DN) ，执行管理节点发出的指示，上报任务状态、存储数据，以及执行数据节点的公共功能。 FusionInsight HD 的数据节点包括 DataNode 、 RegionServer 、 NodeManager 、 LoaderServer 等。

HBase组网方案

节点部署原则	适用场景
管理节点、控制节点和数据节点分开部署	(推荐) 节点数大于 200 节点的集群使用此场景，此方案至少需要 8 个节点
管理节点和控制节点合并部署 数据节点单独部署	(推荐) 节点数大于等于 6 ，小于等于 200 节点的集群使用此场景
管理节点、控制节点和数据节点合并部署	节点数小于 6 的集群使用此场景，此方案至少需要 3 个节点 说明： 不推荐使用此场景 如节点数量满足需求，建议将数据节点单独部署。 如节点数量不满足将数据节点单独部署的要求，必须使用此场景时，需要使用双平面组网方式。

HBase内存要求和部署原则

服务名称	角色名称	内存最小要求	依赖关系	角色业务部署原则
HBase	HM (HMaster)	2GB	依赖于HDFS ZooKeeper 和Yarn	分别部署在2个控制节点上，主备配置。
	RS (RegionServer)	4GB		部署在数据节点上，与HDFS的DataNode保持一致。
	TS (ThriftServer)	1GB		每个集群部署3个在控制节点上。若ThriftServer访问HBase延时不能满足用户需求的时候，可以部署多个在控制或者数据节点上。

HBase扩容

何时扩容RegionServer（前提）：

- 根据业务需求，理论分析当前**RegionServer**的读性能是否满足业务需求。假定网络状况足够好，最坏的情况下，对**HBase**的请求均为采用简单的**Get**接口，且每次**Get**请求所需数据均需从硬盘中读取，业务对集群的**IOPS**为**R**，**RegionServer**所在节点的单个硬盘的**IOPS**为**M**，单节点挂载业务硬盘为**N**块，则至少需要 $R / (M * N)$ 个**RegionServer**实例，如果当前实例个数无法满足此需要，建议进行**RegionServer**的扩容。
- 对**RegionServer**的运行状态进行分析，通过**FusionInsight Manager**获取**HBase**服务的“**RegionServer Queue Statistics**”报表，如果在最近**1-3**个月内发现“**Total Compaction Queue Size**”和“**Total Flush Queue Size**”在持续增长，且无明显的周期性，则说明**HBase**服务在写数据方面遇到性能瓶颈，需要进行扩容。

HBase扩容

- 扩容影响：
 - 扩容是为了更好满足当前业务需求，扩容期间不会影响当前集群业务。
- 扩容方式：
 - 如果集群内仍有部分安装了**DataNode**实例的节点未安装**RegionServer**，且在集群角色部署规划中允许这些节点安装**RegionServer**实例，可直接在这些节点上添加**RegionServer**实例完成扩容。
 - 节点扩容后在新增节点上添加**DataNode**实例与**RegionServer**实例。



目录

1. HBase部署规划
2. HBase关键参数配置
3. HBase监控介绍
4. HBase容灾配置与命令

HMaster关键参数配置

关键配置项	说明
GC_OPTS	运行 Master 进程的 JVM 参数，适当地调整改参数可以优化进程的性能。
hbase.master.logcleaner.ttl	设置 HLog 的保存期限，默认值为 3 天。如果需要尽快释放 Hlog 占用的磁盘空间，可以修改该值。
hbase.master.cleaner.interval	设置执行清除 HLog 动作的间隔时间。
hbase.master.handler.count	Master 处理事务的线程数，如果 Master 十分繁忙或者 Region Server 数很多时，可以适当增大该值。

HMaster关键参数配置

关键配置项	说明
hbase.replication	控制 HBase 集群是否使用 Replication 功能。
hbase.sessioncontrol.enable	开启或关闭 Master 连接数的控制策略。默认策略可以控制最大总连接数、每个用户最大连接数和单位时间内的最大连接数。
hbase.log.maxbackupindex	配置保留 Master 运行日志文件的最大数量。
hbase.log.maxfilesize	配置每个 Master 运行日志文件大小的最大值。

Region Server关键参数配置

关键配置项	说明
GC_OPTS	运行 Region Server 进程的 JVM 参数。如果服务器的内存较大，可以适当增大运行内存，可有效改善性能。
hbase.sessioncontrol.enable	开启或关闭 Region Server 连接数的控制策略。默认策略可以控制最大总连接数、每个用户最大连接数和单位时间内的最大连接数。
hbase.regionserver.maxlogs	未 Flush 数据的 Hlog 的最大数量。
hbase.regionserver.handler.count	Region Server 处理事务的线程数，如果 Region Server 十分繁忙或者很多客户端同时连接集群时，可以适当增大该值。
hbase.regionserver.global.memstore.size	更新被锁定以及强制 flush 发生之前一个 RegionServer 上支持的所有 MemStore 的大小。

Region Server关键参数配置

关键配置项	说明
hfile.block.cache.size	设置读缓存空间占进程总内存大小的比例。
hbase.hregion.max.filesize	设置每个 Region 的总文件大小上限。
hbase.hregion.memstore.flush.size	每个 Region 的写缓存达到指定值的时候，就需要 flush 到文件系统中。
hbase.hregion.majorcompaction	设置自动运行 Major-Compaction 的间隔时间。 如果设置为0，将会禁用自动运行 Major-Compaction 功能。

Region Server关键参数配置

关键配置项	说明
hbase.hstore.compaction.min、 hbase.hstore.compaction.max	指定 Compaction 时的最小和最大文件数。
hbase.hstore.compaction.max.size	当 hfile 的大小超过指定值，在 Minor-Compaction 中该文件不会再进行合并。
hbase.regionserver.thread.compaction.throttle	指定 Compaction 时总数据量的阈值。

典型应用场景

- 用户业务场景需求有以下几点：
 - 数据导入采取**bulkload**批量导入方式；
 - 业务场景中没有实时插入场景；
 - 需要高并发的顺序读。
- 那么根据这个相对典型的应用场景，我们可以调整以下参数：
 - 由于并没有实时插入的场景，可以把 *hbase.regionserver.global.memstore.size* 适当调小；
 - 那由于同时还需要高并发的顺序读性能，可以把 *hfile.block.cache.size* 适当调大，并把 *hbase.regionserver.handler.count* 适当调大点；



目录

1. HBase部署规划
2. HBase关键参数配置
3. HBase监控介绍
4. HBase容灾配置与命令

HMaster信息页面

Master 189-120-24-250

Region Servers

Base Stats

Memory

Requests

Storefiles

Compactions

ServerName	Start time	Requests Per Second	Num. Regions
189-120-24-250,21302,1461644544809	Tue Apr 26 12:22:24 CST 2016	0	1
189-39-150-216,21302,1461644542786	Tue Apr 26 12:22:22 CST 2016	0	1
189-39-150-239,21302,1461644545571	Tue Apr 26 12:22:25 CST 2016	0	0
Total:3		0	2

Backup Masters

ServerName	Port	Start Time
189-39-150-239	21300	Tue Apr 26 12:22:13 CST 2016
Total:1		

用户表信息页面

Table test_table

Table Attributes

Attribute Name	Value	Description
Enabled	true	Is the table enabled
Compaction	NONE	Is the table compacting

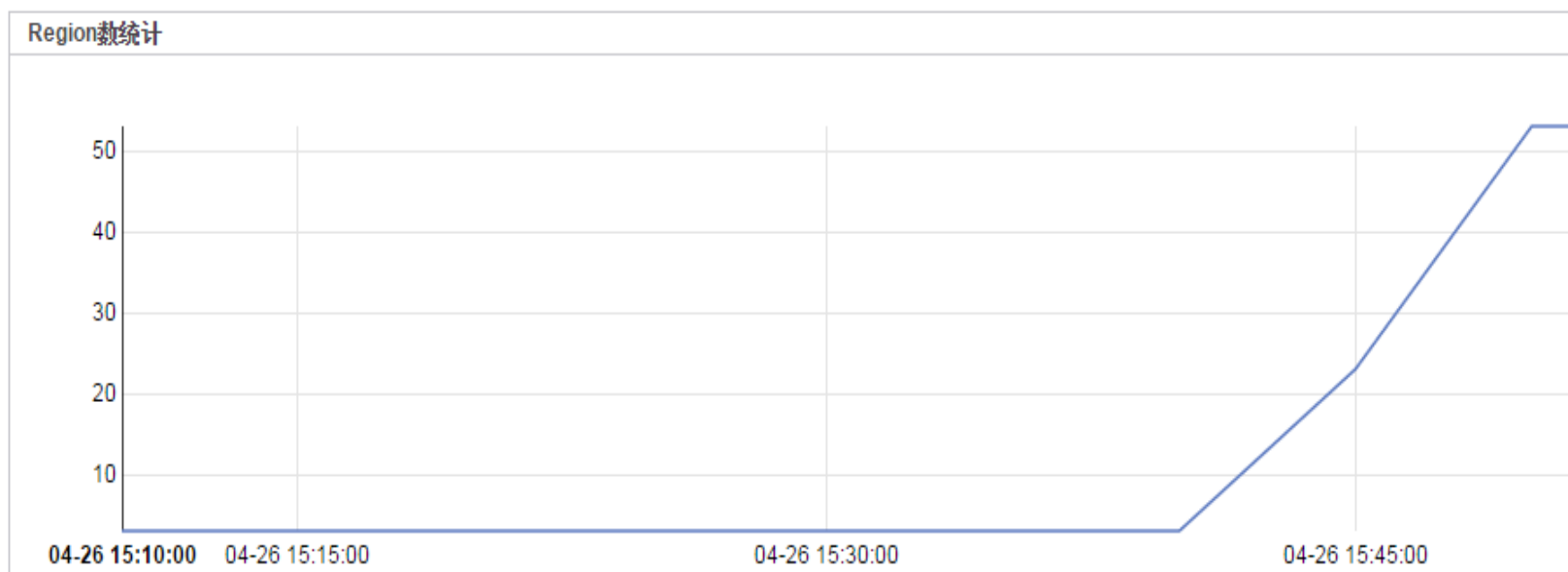
Table Regions

Name	Region Server	Start Key	End Key
test_table,,1461552119718.708e49cb24fff8d329150e37e399bec9.	189-39-150-239:21302		a
test_table,a,1461552119718.cd392f3f0b8fe1c7e07cb193972a5a58.	189-39-150-216:21302	a	b

服务监控页面

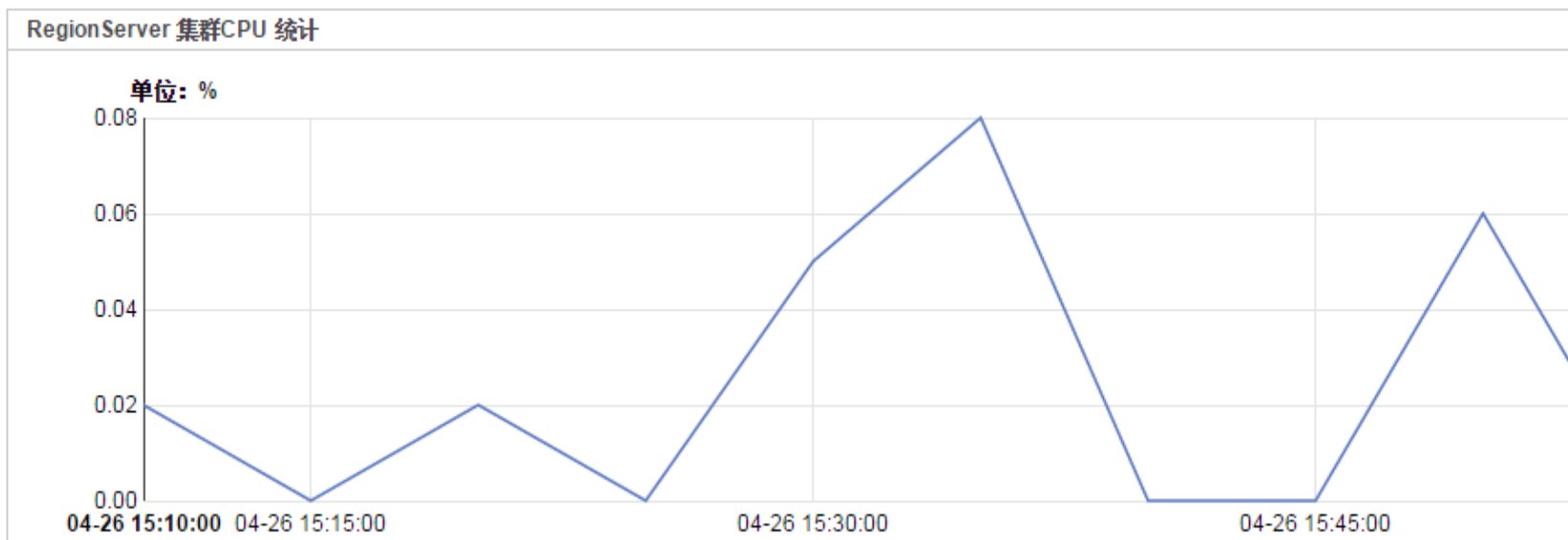
- **Region数统计**

- **HBase当前Region总数量**。可以评估集群**Region**数量的变化趋势，发现是否存在大量**Region**分裂等情况。



服务监控页面

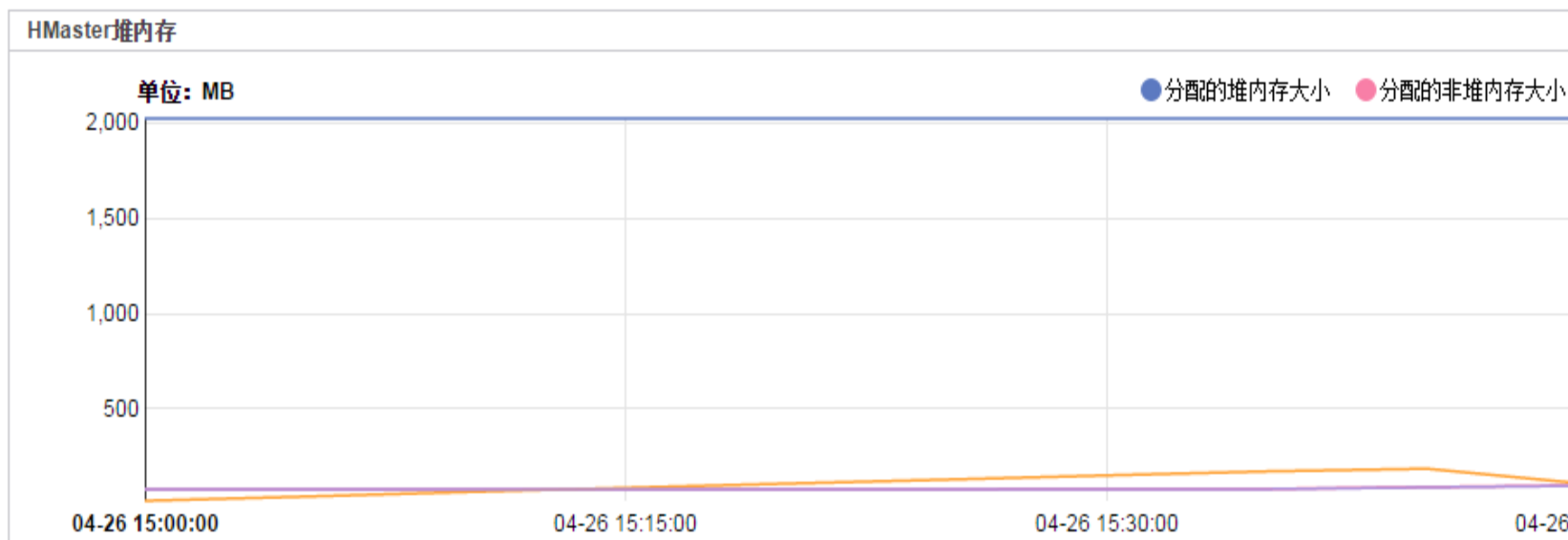
- **RegionServer 集群CPU统计**
 - **RegionServer**当前**CPU**使用率总和。可以评估集群是否处于较高的负载压力情况、或者用于定位性能问题。



HMaster进程监控页面

- HMaster堆内存

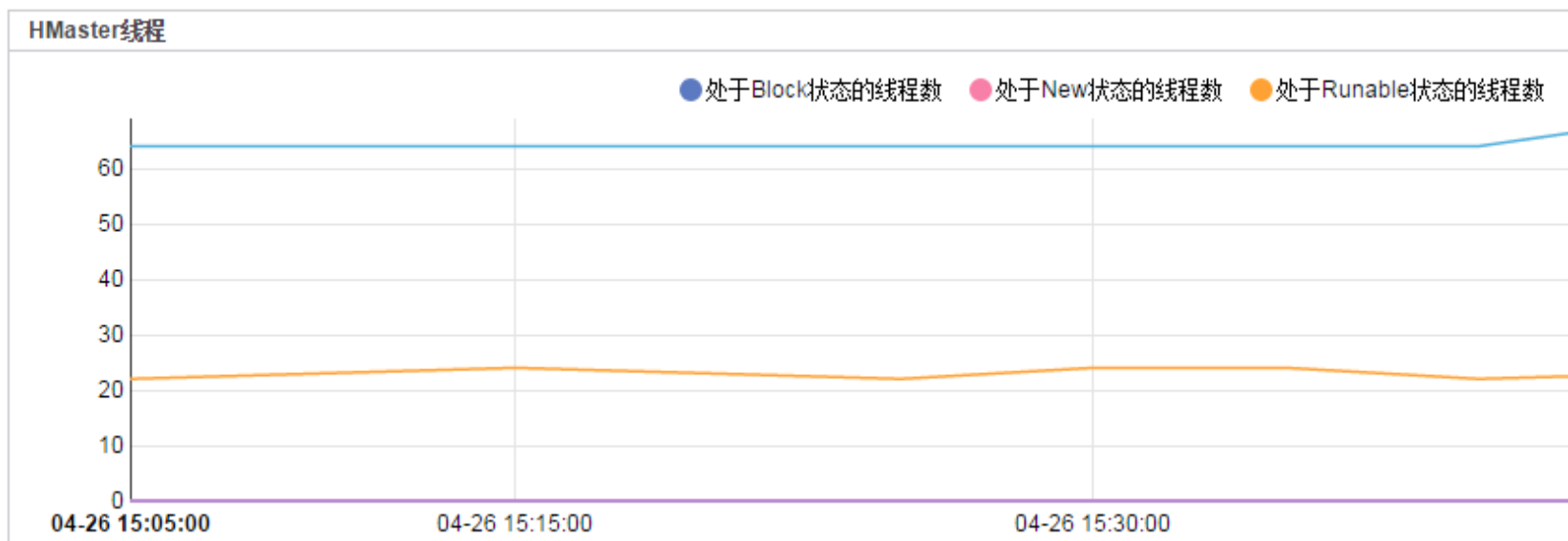
- **Master**进程内存使用情况。如果已用内存占总内存比例一直较高，说明内存资源比较紧张，可以考虑优化。



HMaster进程监控页面

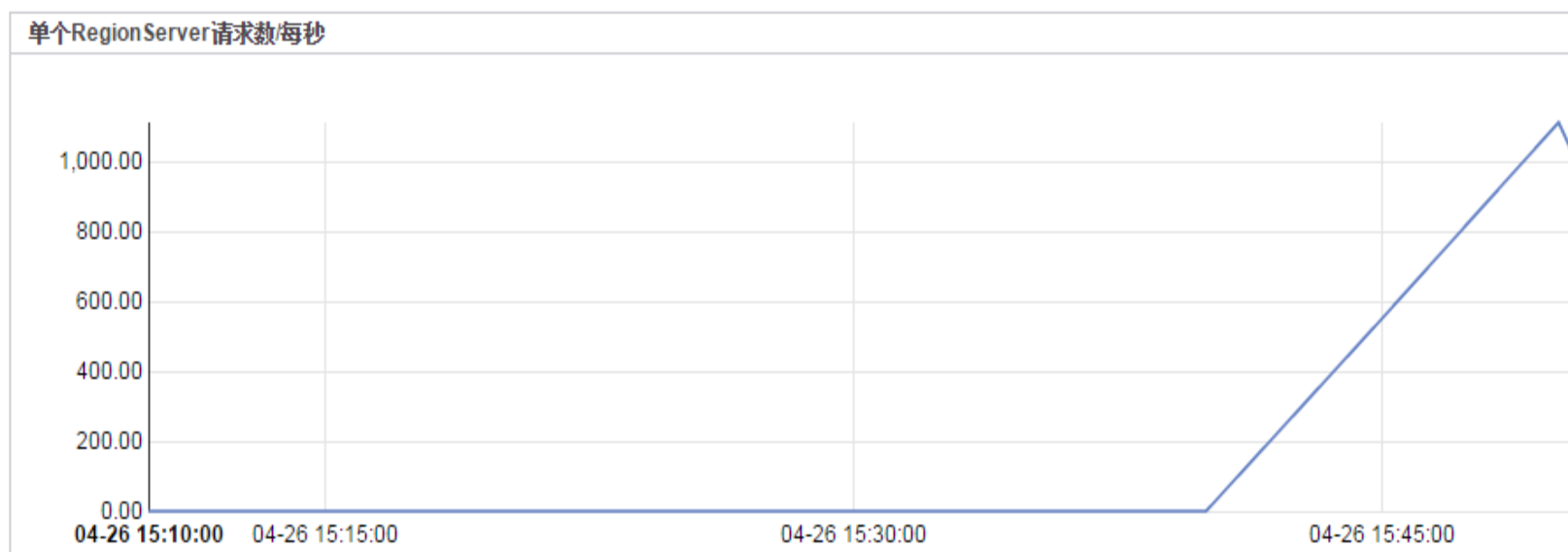
- HMaster线程

- **Master**进程中线程运行情况。如所有线程一直都处于运行状态，说明**Master**业务较为繁忙，可以考虑优化。



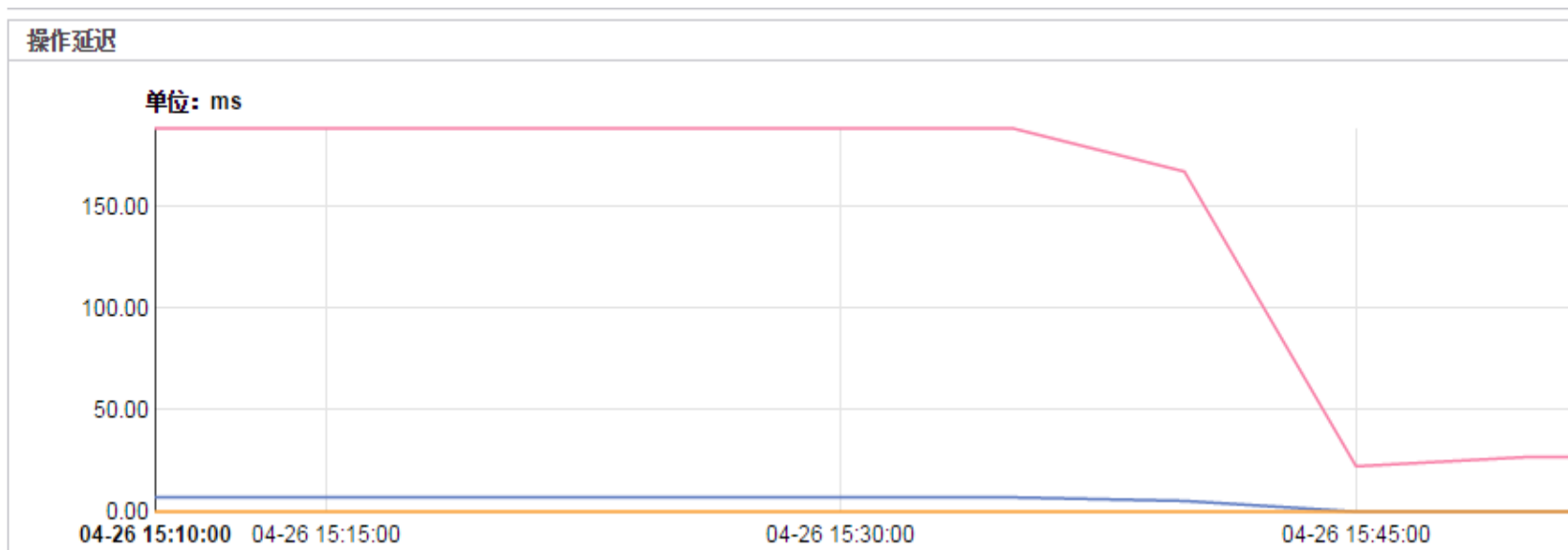
RegionServer进程监控页面

- 单个RegionServer请求数/每秒
 - RegionServer当前一秒内处理的数据量，这是一个可以评估该Region Server进程的负载情况。



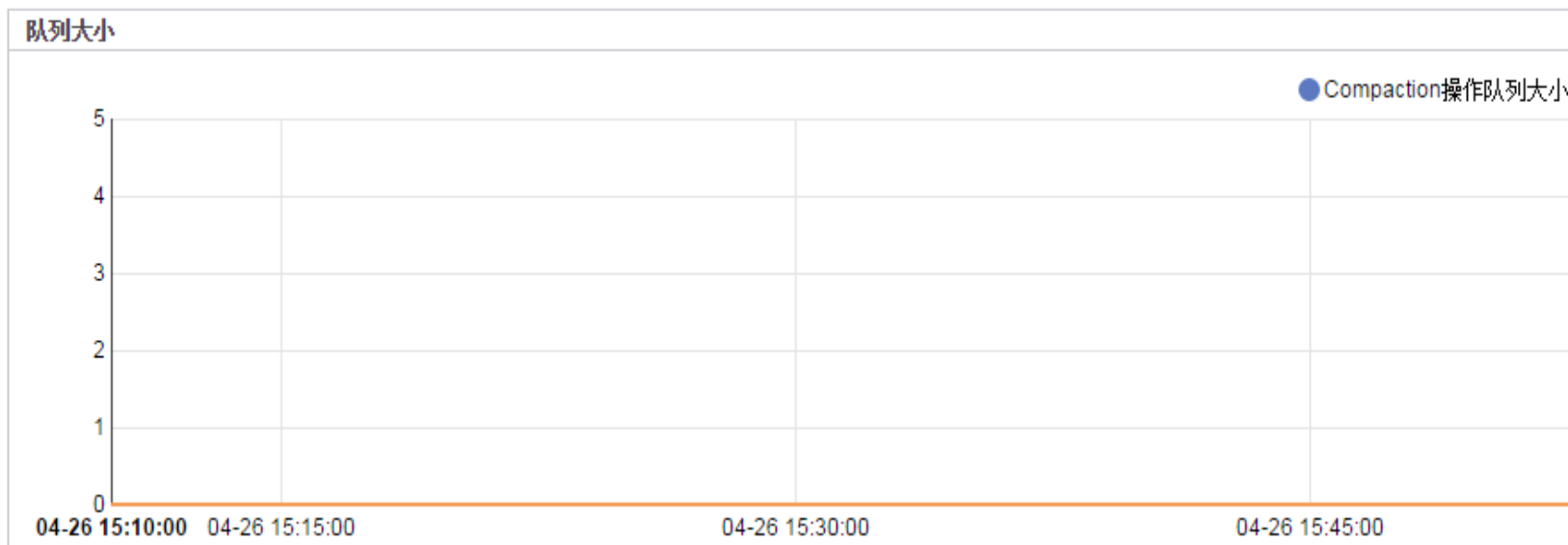
RegionServer进程监控页面

- 操作延迟
 - **RegionServer**对请求的处理时延。可以根据当前值或趋势判断**Region Server**的性能状态。



RegionServer进程监控页面

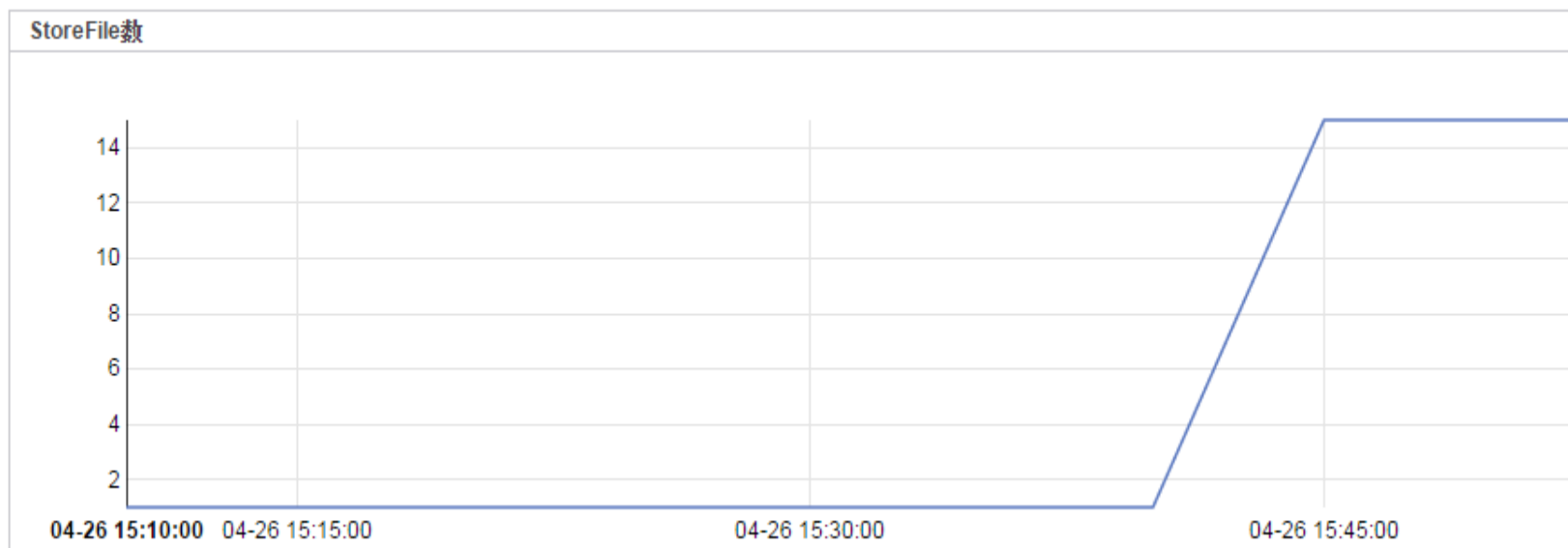
- 队列大小
 - **RegionServer**中**Compaction**和**Flush**队列的大小。如果任务队列一直较大，可能需要进行优化。



RegionServer进程监控页面

- StoreFile数

- **RegionServer**当前数据文件数。可用于定位性能问题，如数据文件数越来越多，有可能导致性能降低，可以考虑对**RegionServer**扩容。





目录

1. HBase部署规划
2. HBase关键参数配置
3. HBase监控介绍
4. HBase容灾配置与命令

HBase容灾

- 容灾操作场景

HBase集群容灾作为提高**HBase**集群系统高可用性的一个关键特性，为**HBase**提供了实时的异地数据容灾功能，可以把本**HBase**集群中的数据备份到另一个集群。支持**HBase**表普通写数据与**Bulkload**批量写数据场景下的容灾。

它对外提供了基础的运维工具，包含灾备关系维护，重建，数据校验，数据同步进展查看等功能。

HBase容灾

- 容灾使用约束

- 尽管容灾提供了实时的数据复制功能，但实际的数据同步进展，由多方面的因素决定的，例如，当前主集群业务的繁忙程度，备集群进程的健康状态等。因此，在正常情形下，主备数据并非绝对实时同步，备集群不应该接管业务。极端情形下是否可以接管业务，可由系统维护人员以及决策人员根据当前的数据同步指标来决定。
- 容灾功能当前仅支持一主一备。
- 通常情况下，不允许对备集群的灾备表进行表级别的操作，例如修改表属性、删除表等，一旦误操作备集群后会造成主集群数据同步失败、备集群对应表的数据丢失。
- 主集群的**HBase**表已启用容灾功能同步数据，用户每次修改表的结构时，需要手动修改备集群的灾备表结构，保持与主集群表结构一致。

HBase容灾

- **HMaster配置项**

配置项	缺省值	描述
hbase.master.logcleaner.ttl	600000	指定 HLog 的保存期限。如果配置值为“ 604800000 ”（单位：毫秒），表示 HLog 的保存期限为 7 天。
hbase.master.cleaner.interval	60000	用于定义 HLog 的删除周期，即超过设置的时间的 HLog 会被自动删除。建议尽可能配置大的值来保留更多的 Hlog 。

HBase容灾

- RegionServer配置项

配置项	缺省值	描述
replication.source.size.capacity	67108864	当主集群同步数据到备集群中时，主集群会从 HLog 中读取数据，此时会根据本参数配置的大小读取并发送
replication.source.nb.capacity	25000	当主集群同步数据到备集群中时，主集群会从 HLog 中读取数据，此时会根据本参数配置的个数读取并发送。与“ replication.source.size.capacity ”一起配置使用
replication.source.maxretriesmultiplier	10	发送 Log 数据失败后，重新尝试的次数限制
replication.source.sleepforretries	1000	当发送 Log 数据失败后的休眠时间（毫秒）
hbase.regionserver.replication.handler.count	3	备用集群用于接收数据的 RPC 处理程序的线程数

HBase容灾

- 容灾相关命令

操作	命令	描述
建立灾备关系	add_peer '备集群ID','备集群地址信息' 示例: add_peer '1','zk1,zk2,zk3:24002:/hbase' add_peer '1','zk1,zk2,zk3:24002:/hbase1' 24002表示集群中ZooKeeper的端口号。 add_peer '1','zk1,zk2,zk3:24002:/hbase1',{HDFS_CON FS => true}	建立主集群与备集群的关系， 让其互相对应。如果启用 Bulkload 批量写数据容灾，则 命令为 add_peer '备集群ID',' 备集群地址信息 '',{HDFS_CONFS => true}
移除灾备关系	remove_peer '备集群ID' 示例: remove_peer '1'	在主集群中移除备集群的信息

HBase容灾

- 容灾相关命令

操作	命令	描述
启用用户表实时同步	enable_table_replication '表名' 示例: enable_table_replication 't1'	在主集群中, 设置已存在的表同步到备集群
禁用用户表实时同步	disable_table_replication '表名' 示例: disable_table_replication 't1'	在主集群中, 设置已存在的表不同步到备集群
查询灾备关系	list_peers	在主集群中查询已经设置的备集群的信息, 主要为 Zookeeper 信息

HBase容灾

- 容灾相关命令

操作	命令	描述
主备集群数据校验	<p>在安装好HBase客户端的Linux环境中执行，bin/hbase</p> <p>org.apache.hadoop.hbase.mapreduce.replication.VerifyReplication -starttime=开始时间 -endtime=结束时间 列族名称 备集群ID 表名</p>	<p>检查指定的表在主备集群间的数据是否一致</p> <p>命令行中参数说明如下：</p> <ul style="list-style-type: none">• 开始时间：如果未设置，则取默认的开始时间为0• 结束时间：如果未设置，则取默认的结束时间为当前操作提交的时间• 表名：如果未输入表名，则默认校验所有的启用了实时同步的用户表

HBase容灾

- 容灾相关命令

操作	命令	描述
切换数据写入状态	set_clusterState_active set_clusterState_standby	设置集群 HBase 表是否可写入数据。 Active 可以读写， Standby 只能读，不能写。

HBase容灾

- 容灾相关命令

操作	命令	描述
新增或更新已经在对端集群保存的主集群中HDFS配置	在HBase Shell中执行， <code>set_replication_hdfs_confs 'PeerId', {'key1' => 'value1', 'key2' => 'value2'}</code>	<p>启用包含Bulkload数据的容灾，在主集群修改HDFS参数时，新的参数值默认不会从主集群自动同步到备集群，需要手动执行命令同步。受影响的参数如下：</p> <ul style="list-style-type: none">fs.defaultFSdfs.client.failover.proxy.provider.haclusterdfs.client.failover.connection.retries.on.timeoutsdfs.client.failover.connection.retries <p>例如，fs.defaultFS 修改为 hdfs://hacluster_sale，同步HDFS配置到id为1的备集群时执行： <code>set_replication_hdfs_confs '1', {'fs.defaultFS' => 'hdfs://hacluster_sale'}</code></p>



总结

本章首先介绍了**HBase**部署和规划，以及部署完成后常用的参数配置，然后讲解了如何监控**Hbase**,最后展示了**Hbase**的容灾配置。

思考题

1. 没有配置互信的集群能否配置**HBase**容灾?
2. 在什么情况下需要考虑对**Hbase**扩容?

培训与认证部

习题

多选题

1. 部署**Hbase**需要依赖哪些组件？

- A. HDFS
- B. Hive
- C. ZooKeeper
- D. Yarn

Thank you

www.huawei.com