

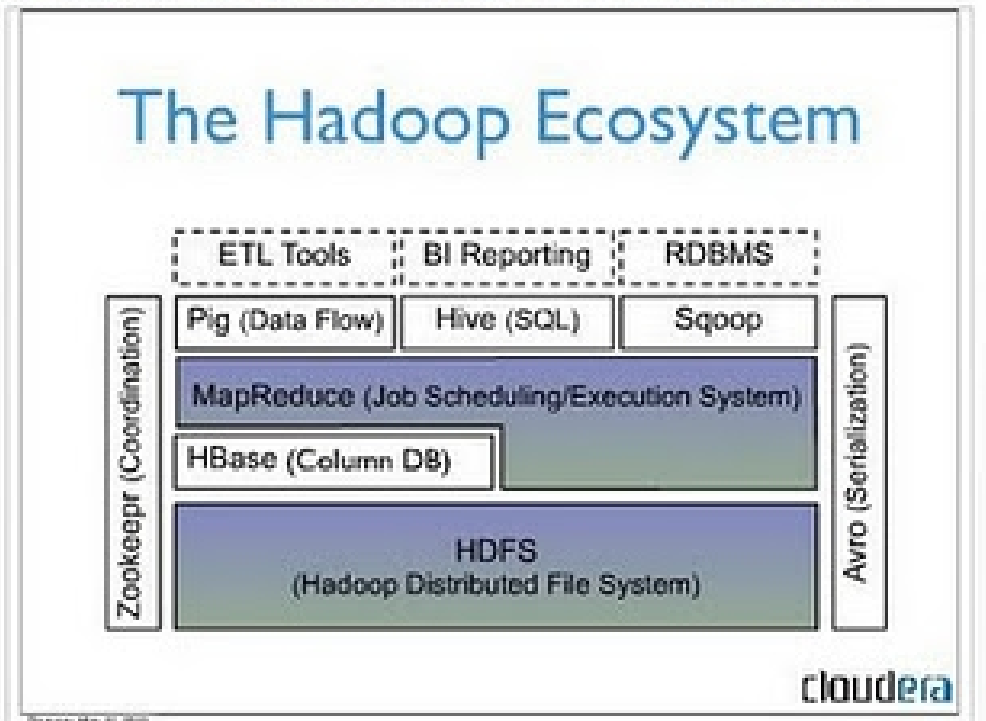


介绍

HBASE 定位

- HBASE 是存储
- 基于 HDFS
- 实时随机读写

Apache hadoop an introduction todd lipcom - gluecon 2010



HBASE 特性

- 线性扩展
- 行操作的强一致性
- 自动分表
- 支持 MapReduce
- Java,Thrift,REST-ful 接口

HBASE 基本性能参数

- 3 台 RegionServer. 每台 8G 内存，8 核
- 1 亿行

	Row/s	MB/s	Row/s Per node	BigTable Row/s per node
随机写	14789	14.789	4930	8850
随机写 (noLog)	22180	22.180	7393	8850
随机读	1996	1.996	665	1212
顺序读	10678	10.678	3559	4425

目录

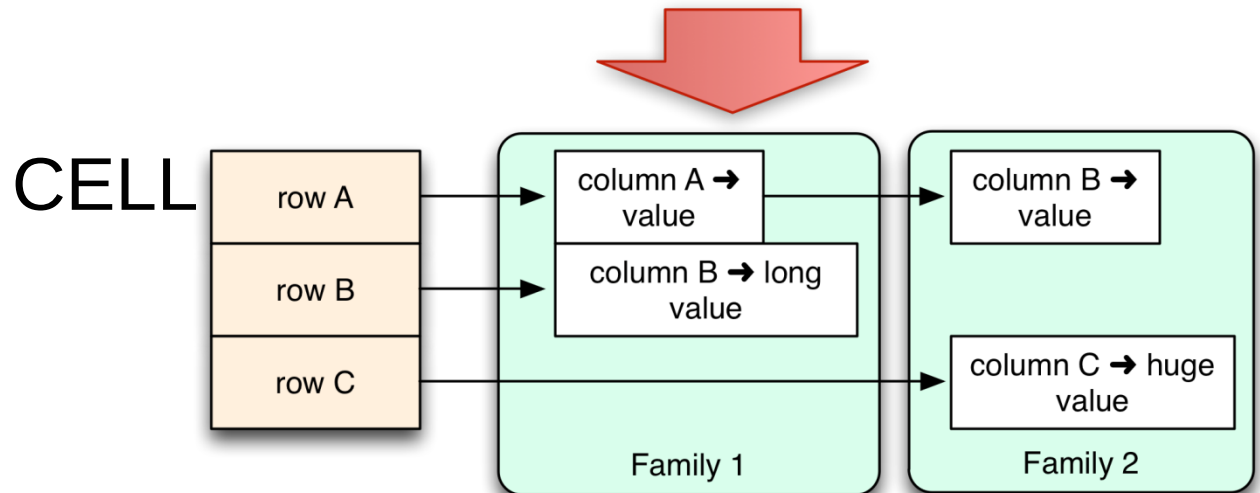
- HBASE 模型
- 架构设计
- 使用技巧
- 运维技巧
- 测试分析

HBASE 模型

Hbase 数据模型

- Table
- Region
- ColumnFamily
- Row
- Column
- Version
- Value

	column A (int)	column B (varchar)	column C (boolean)	column D (date)
row A				
row B				
row C			NULL?	
row D				



HBASE 操作

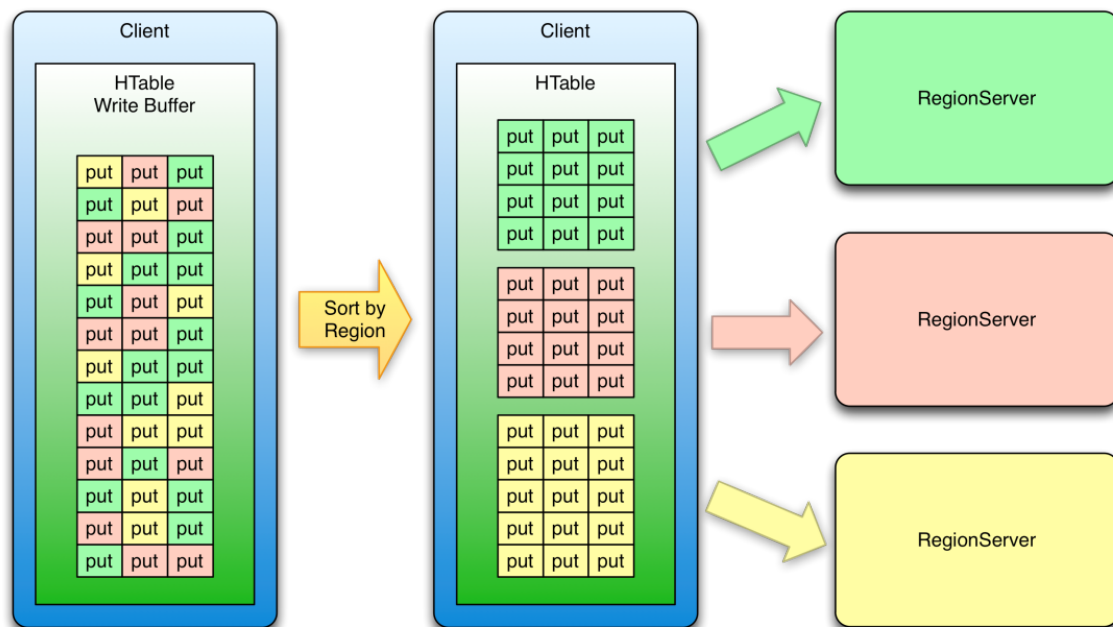
- Put

- Delete
- 原子操作
- WAL

- Scan

- Get
- Filter
- Cache/Batch

- 批量操作



Hbase 架构设计

总体结构

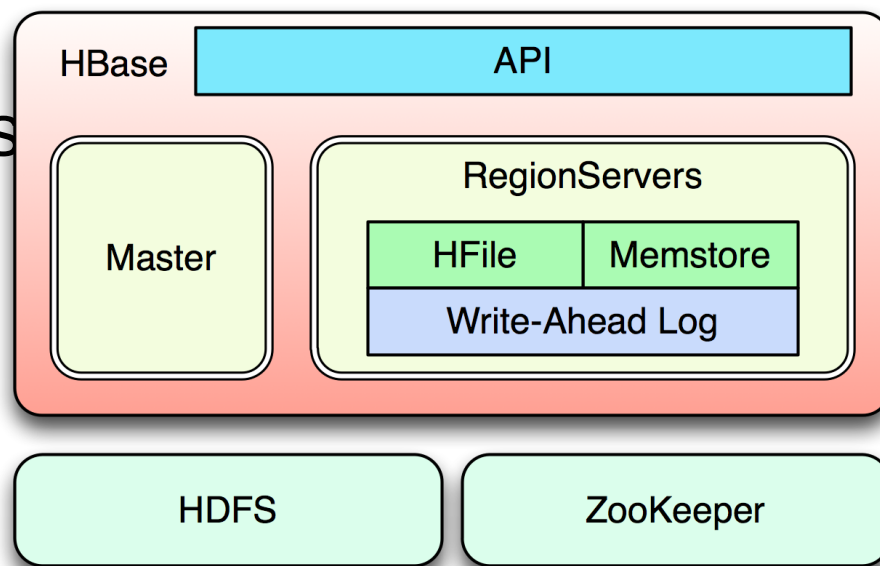
- Master
 - Region 之上的操作
 - Put/Get 不经过 Mas

- RegionServer

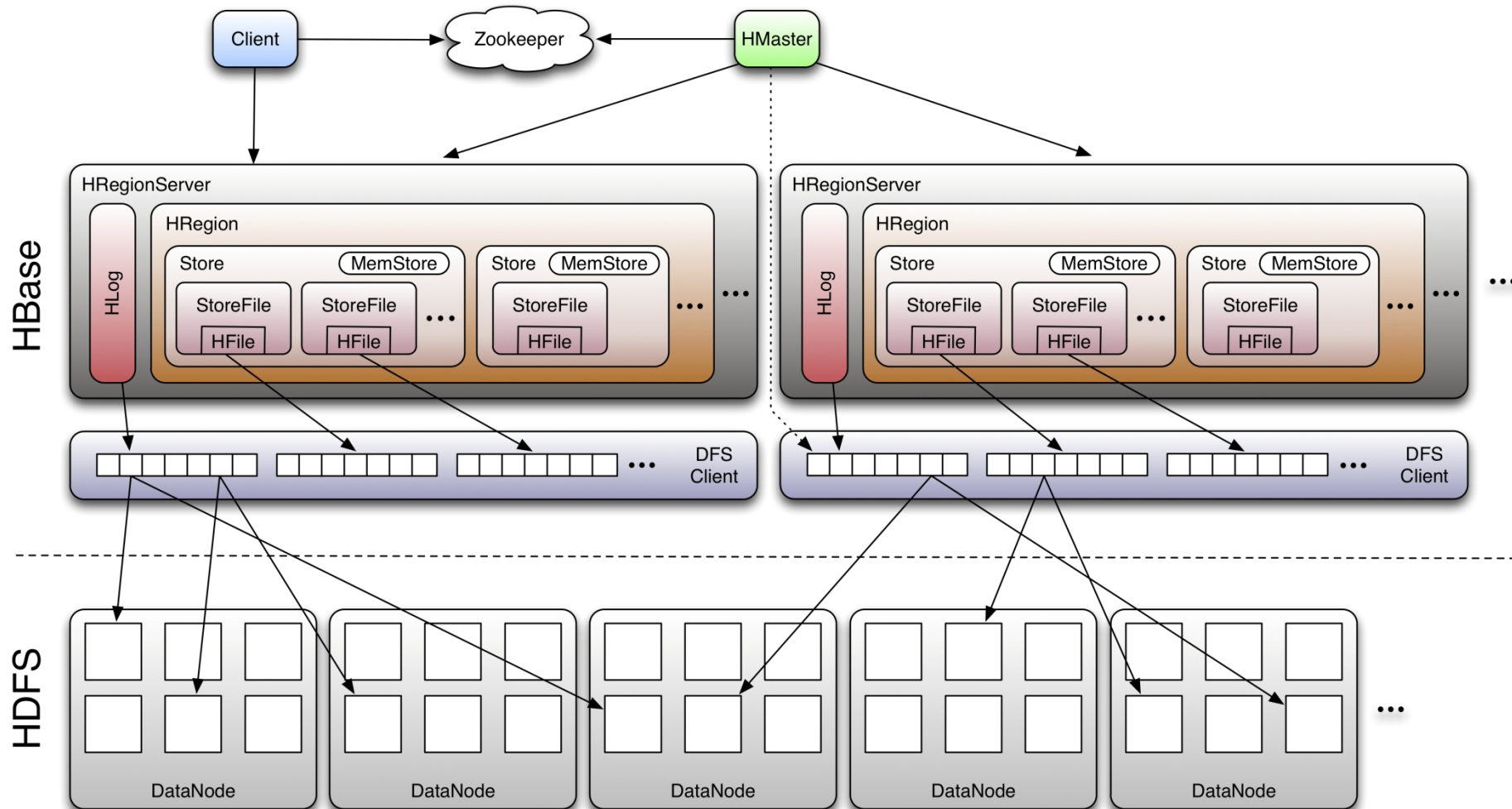
- Region 之下的操作

- HDFS

- HFile
 - HLog



RegionServer 结构



HFile 结构

- DataBlock

- 存储 Key-Value

- MetaBlock(可选)

- 存储 BloomFilter

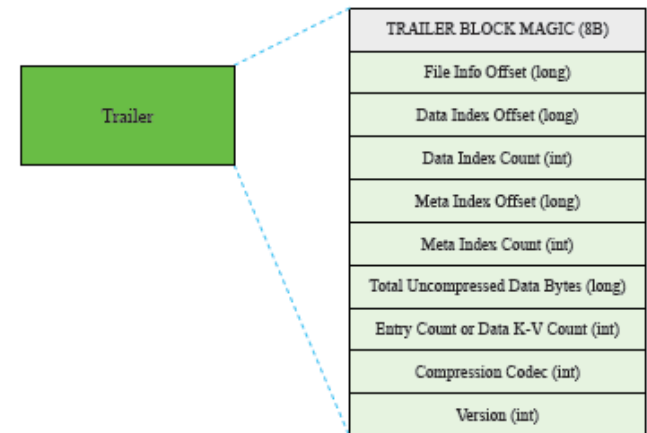
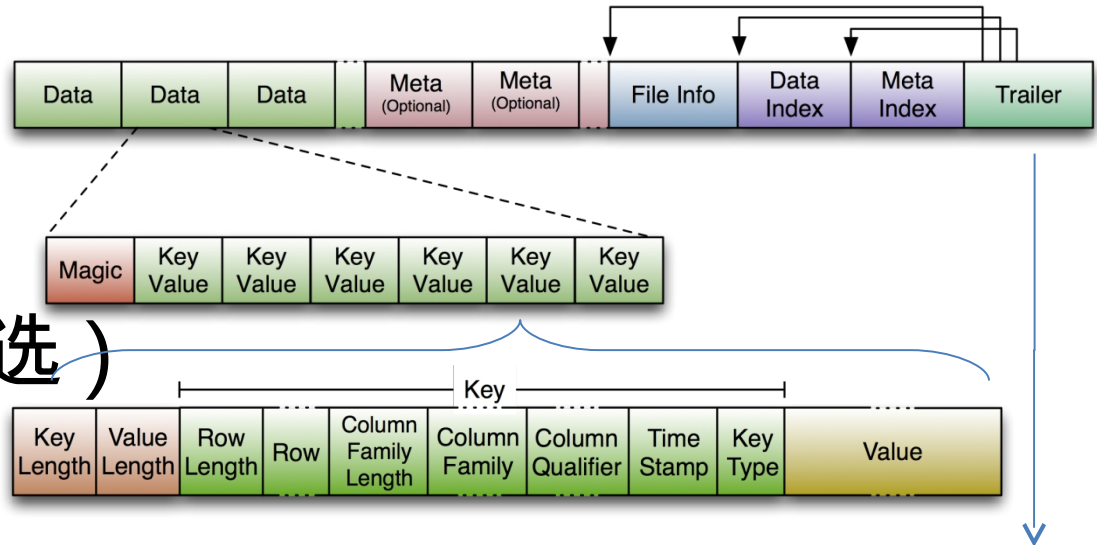
- DataBlockIndex

- Key 到 Block Offset

- Read

- 占用内存，加载缓慢

- Write



Total Size of Trailer: 4xLong + 5xInt + 8Bytes = 60 Bytes

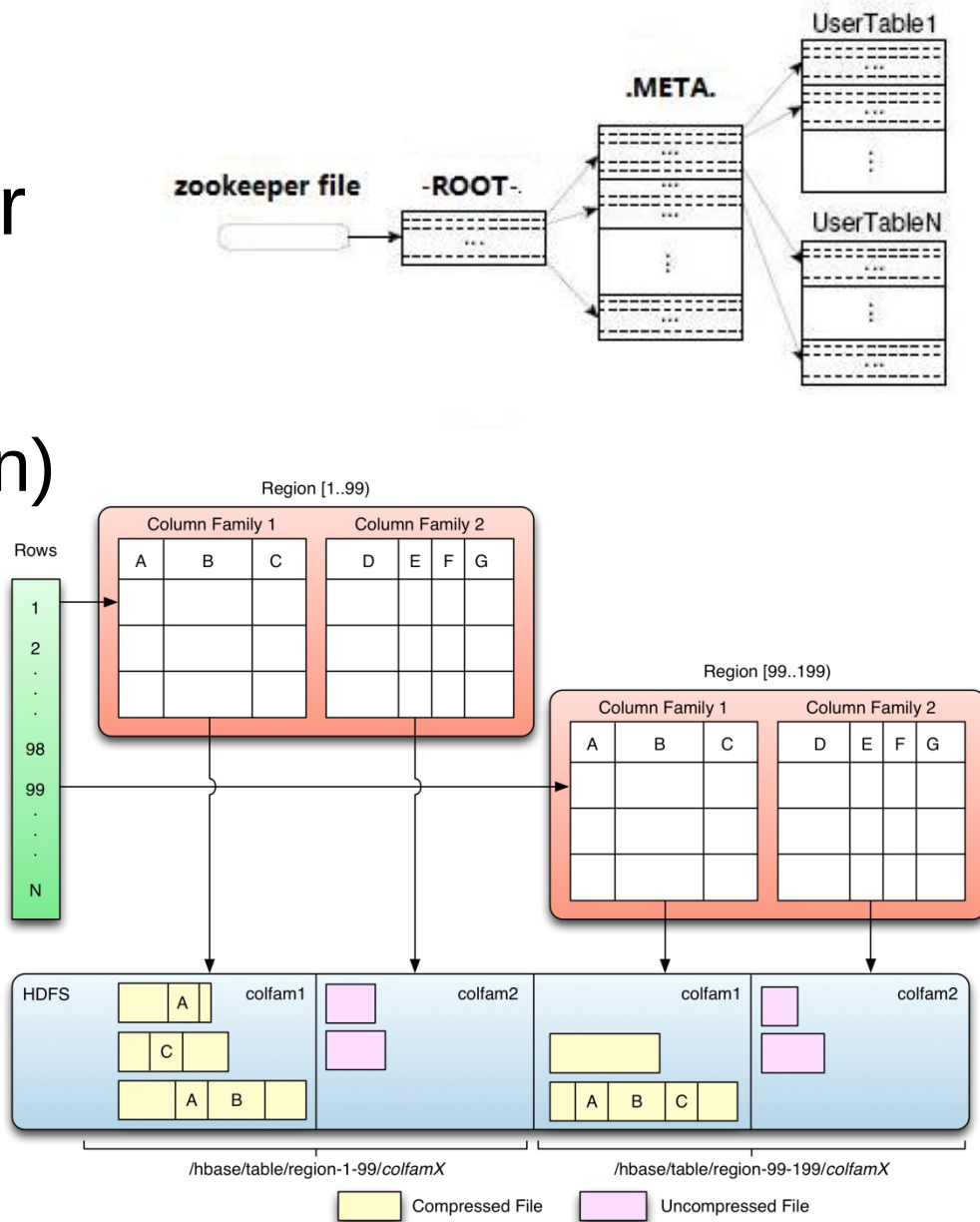
Hfile 性能测试

	none	gz	lzo
Write	20718	23885	55147
Full Scan	41436	94937	100000
Random Seek	600	989	956
Random Short Scan	12241	25568	25655



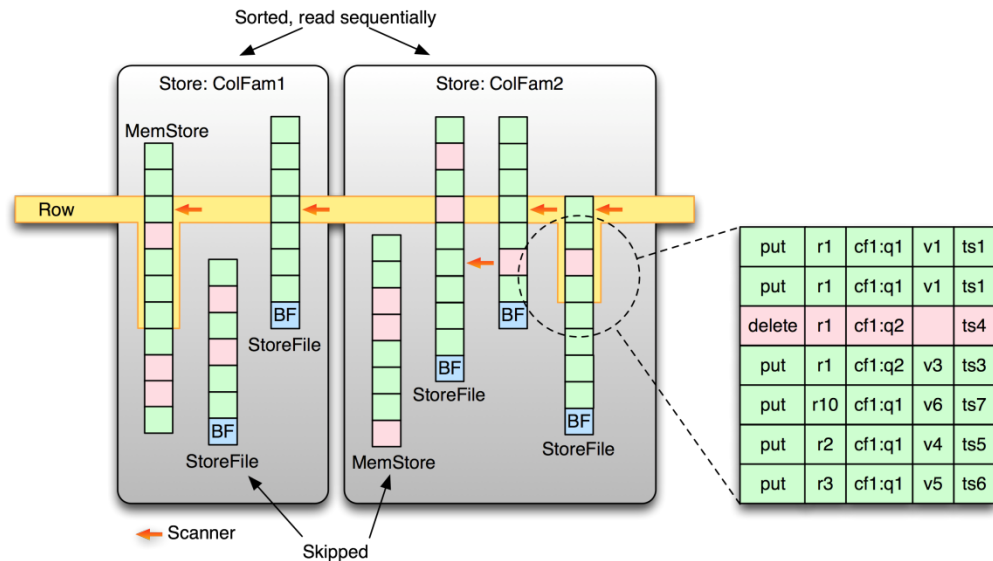
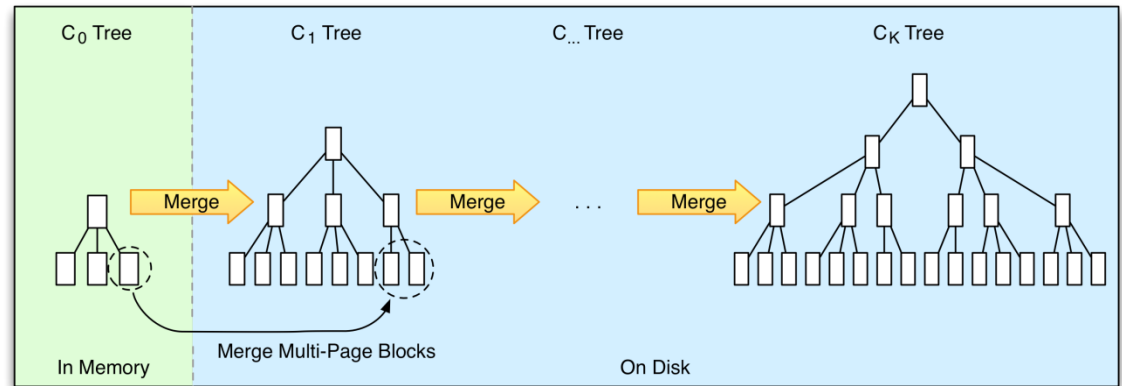
存储分布

- 寻找 RegionServer
 - ZooKeeper
 - ROOT-(单 Region)
 - .META.
 - 用户表



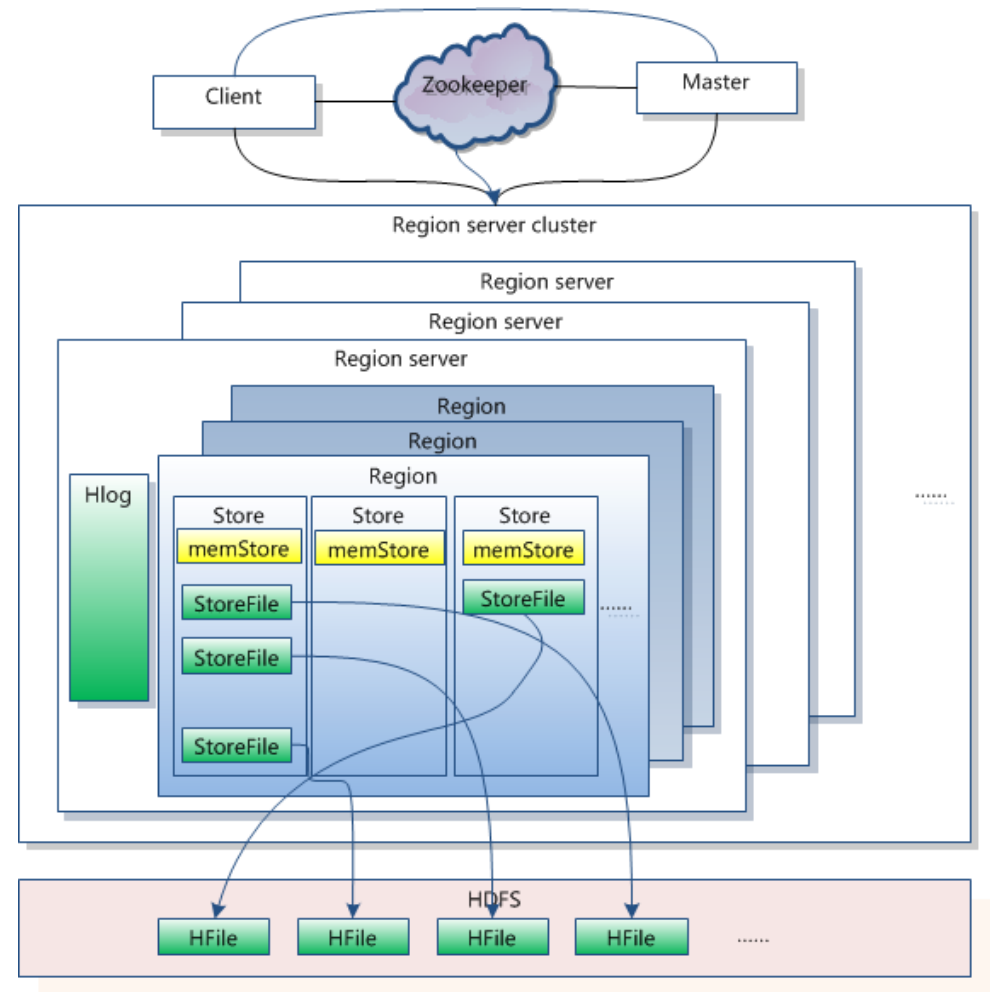
Put/Get 操作

- PUT
- DELETE
- GET
- SCAN



Region 操作

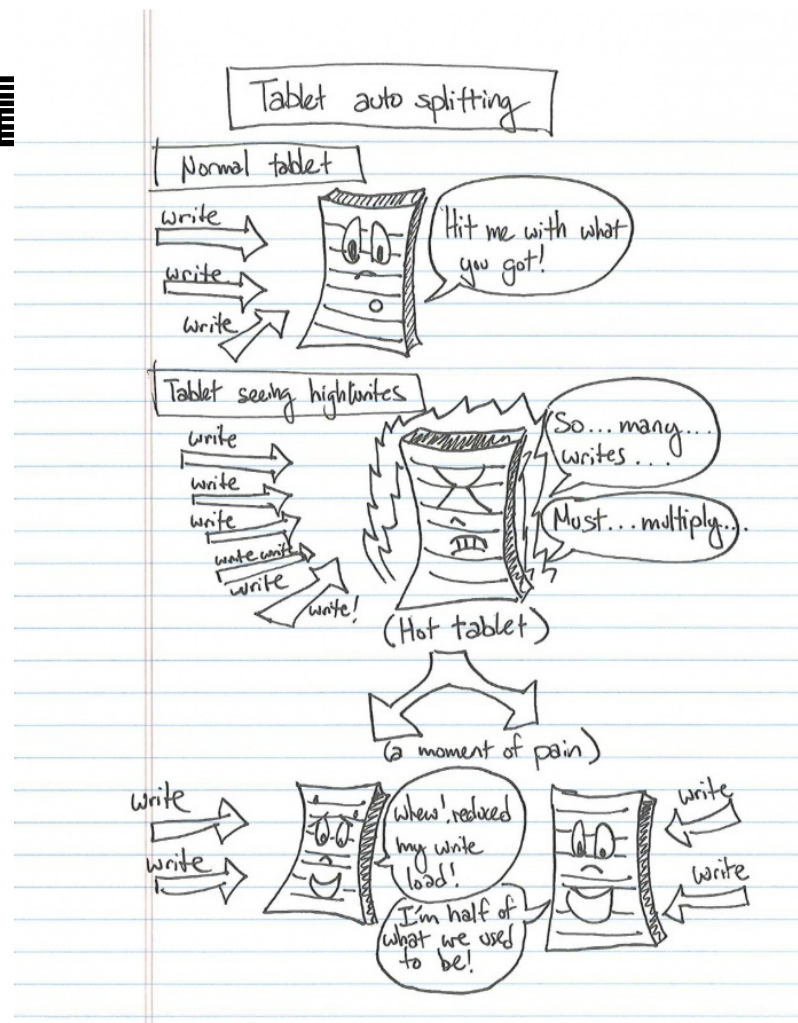
- Flush MemStore
- Compact
- Major Compact
- Split



使用技巧

Schema 设计

- Column Family 的数量
 - 最好为 1
- Key 的设计
 - 避免单调递增
 - 最小化
- 最小化 Column



Schema 设计

- Column Family 的数:

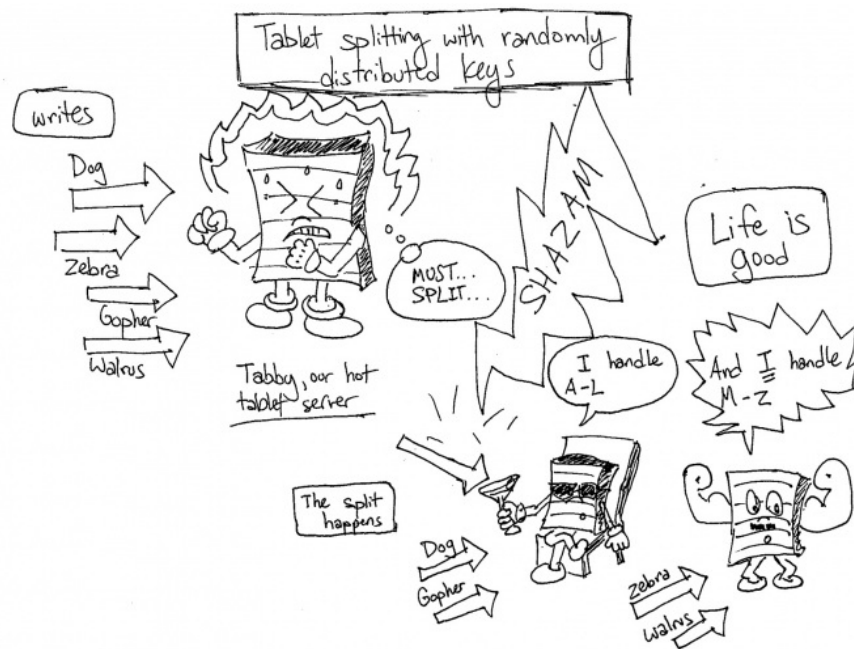
- 最好为 1

- Key 的设计

- 避免单调递增

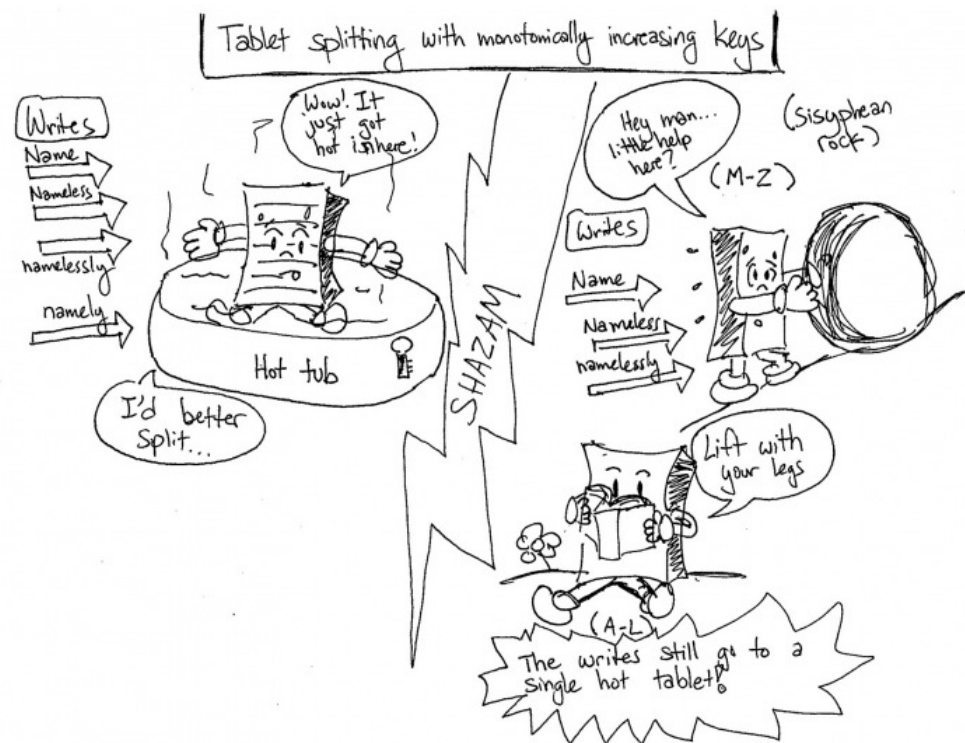
- 最小化

- 最小化 Column



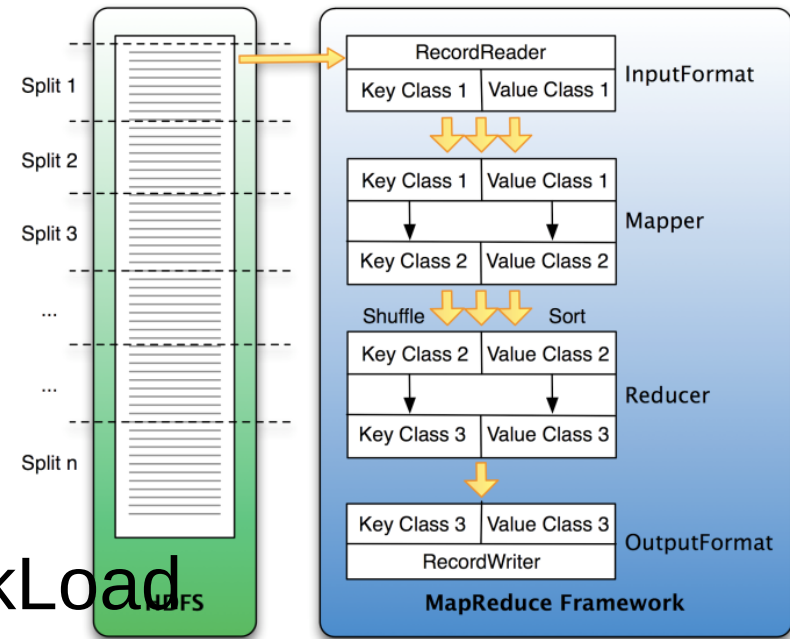
Schema 设计

- Column Family 的数
 - 最好为 1
- Key 的设计
 - 避免单调递增
 - 最小化
- 最小化 Column



MapReduce 结合

- Mapper
 - Region 数 = Mapper 数
- Reducer
 - Region 数 = Reducer 数
 - Reducer 写 Hfile, 再 BulkLoad
- Hive/Pig



建立索引

- 单列索引

- 组合索引

- Join?

- Key \Leftrightarrow Kind:ID

Index	Key
Column:Value	Key...

单列索引

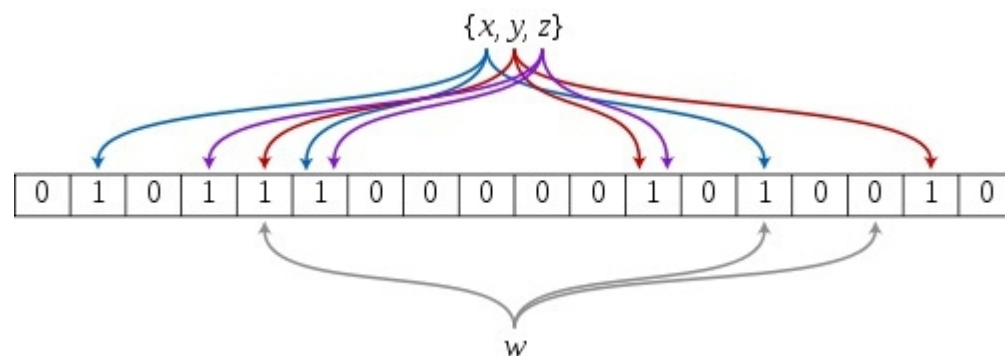


Index	Key
Column:Value/Column:Value	Key...

组合索引

开发调优

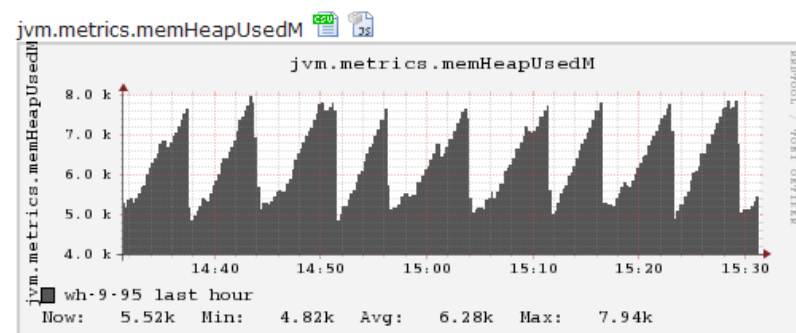
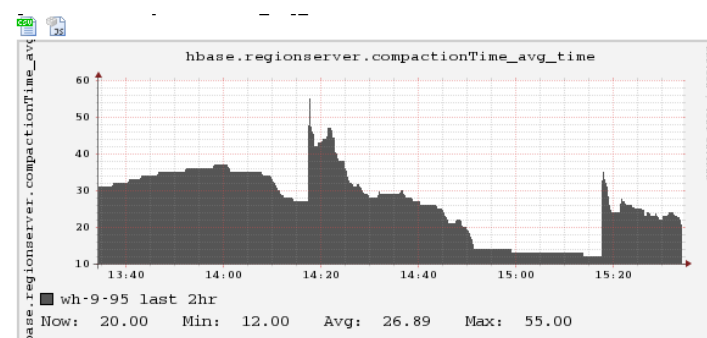
- Table 属性
 - BlockSize
 - BloomFilter
 - BlockCache
 - InMemory
- 尽可能使用 Bulk Load
- Put 使用客户端 Cache
- Scan 使用 Cache/Batch



运维技巧

HBASE 部署

- Hadoop 版本
 - Hadoop 0.20.x
 - Append 补丁
- ZooKeeper
- Metric
- 内存
 - RegionServer 12GB
 - MemStore $\leq 40\%$
 - UFile DataIndex



Region 管理

- 预创建 Region
- Region 的大小
 - `hbase.hregion.max.filesize=256MB,1GB,4GB`
 - 手动 Split, 交错负载
- Region 合并
 - `hbase.hstore.compactionThreshold=3`
 - `hbase.hstore.blockingStoreFiles=7`(阻塞 , 超时)
 - `hbase.hstore.compaction.max=10`

运维调优

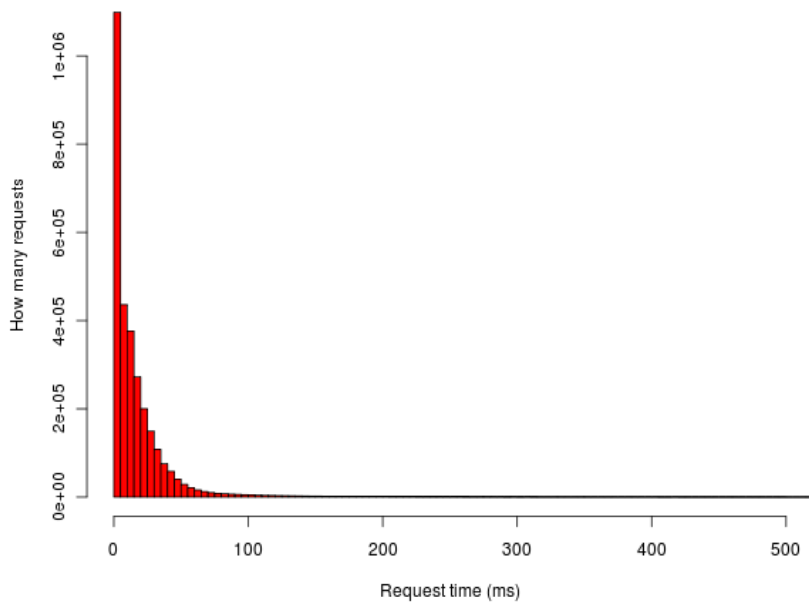
- Java GC
 - JVM GC 调整 (ParNewGC+CMS)
 - Full GC-10s/GB
 - MemStore 本地分配 (2MB, 减少碎片)
- LZ0 压缩
 - 压缩单位为 Block
 - 提高性能
- 并发数调整

hbase.regionserver.handler.count

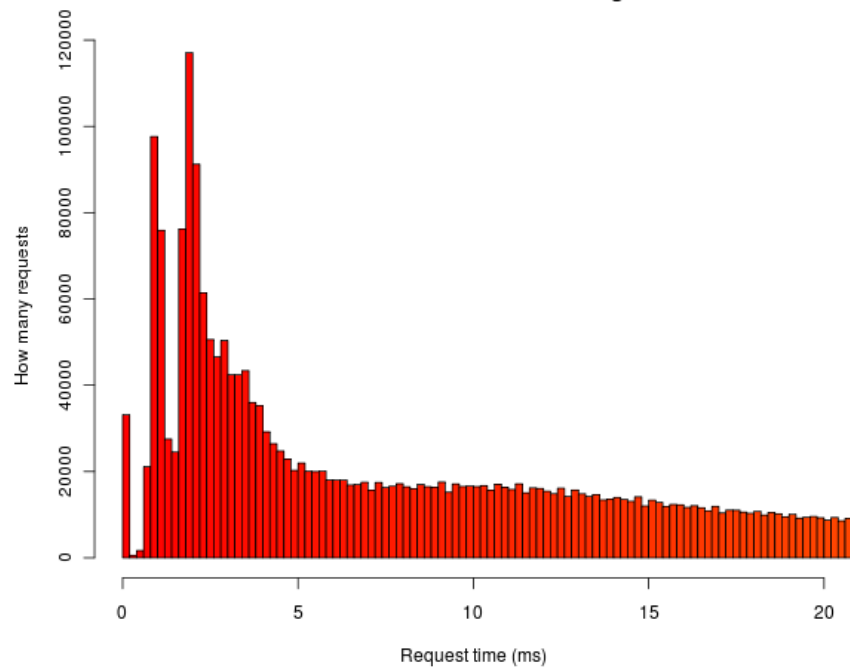
测试分析

随机 Get 测试

GETS time histogram



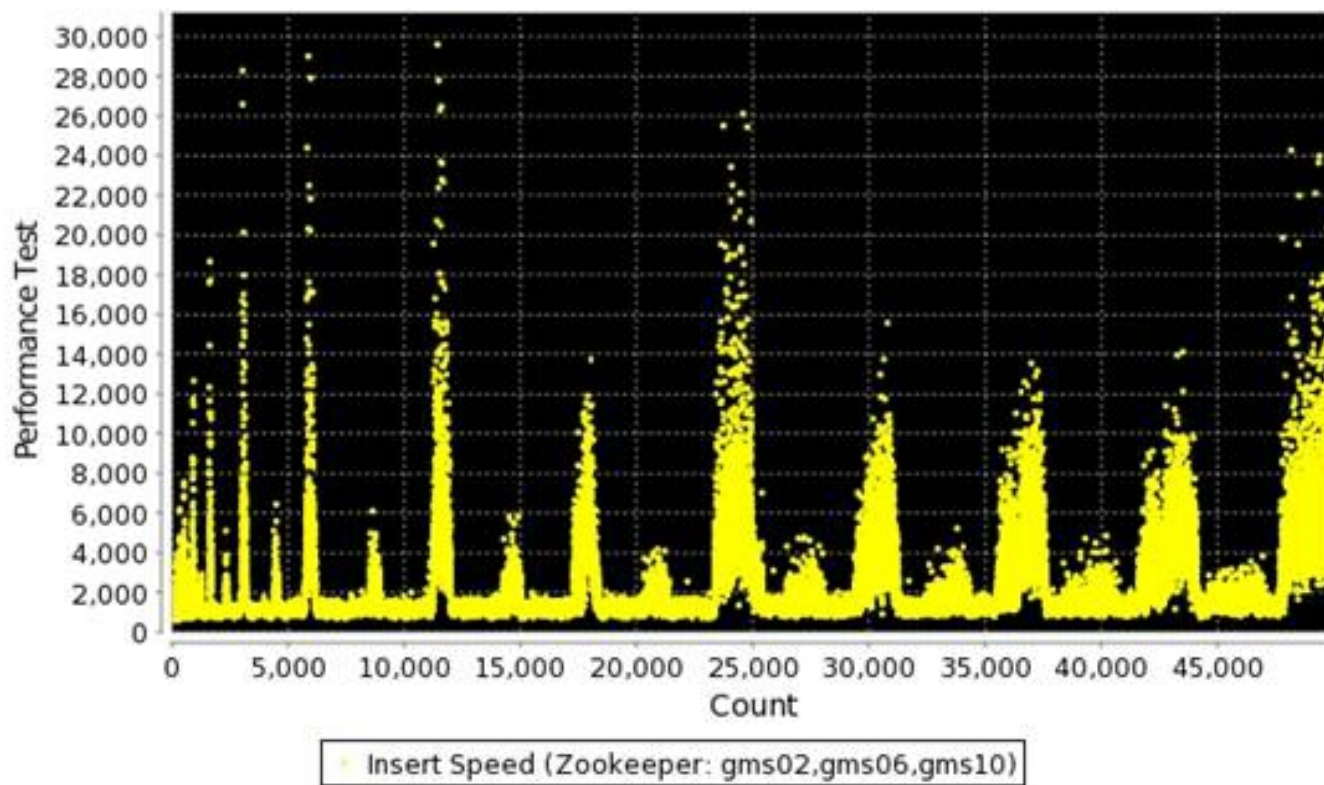
GETS <20ms zoomed time histogram



- Get 波动不是很大

Put 测试

Performance Test for Insert Rows (Batch = 500)

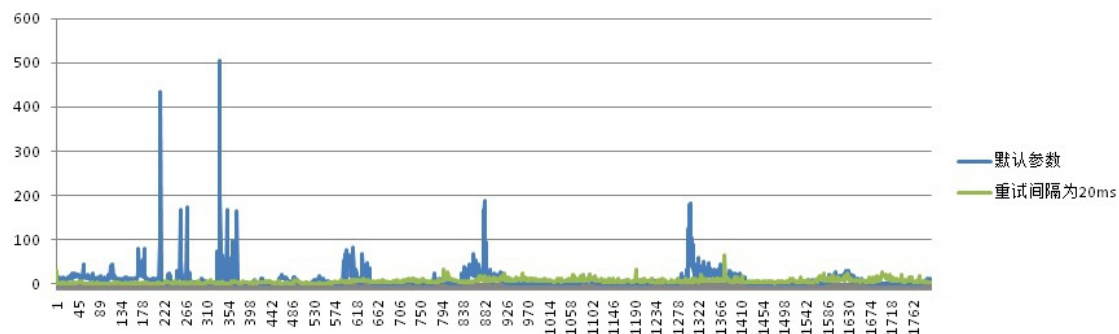


- Put 有波动
- Region 操作导致阻塞

Put 测试

- Client 重试波动
- HLog 拖慢速度
- Split 波动
- Compact 波动

不同client参数下写速度的对比



不同Server参数下写速度的对比

