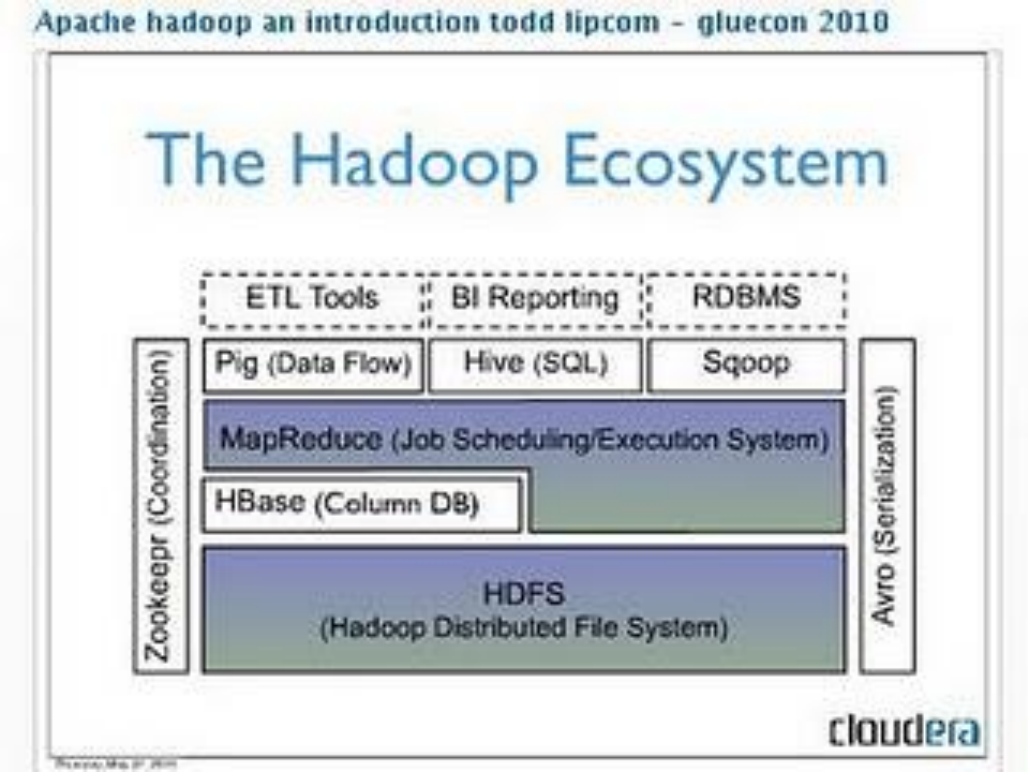
介绍

HBASE定位

- HBASE是存储
- 基于HDFS
- 实时随机读写



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

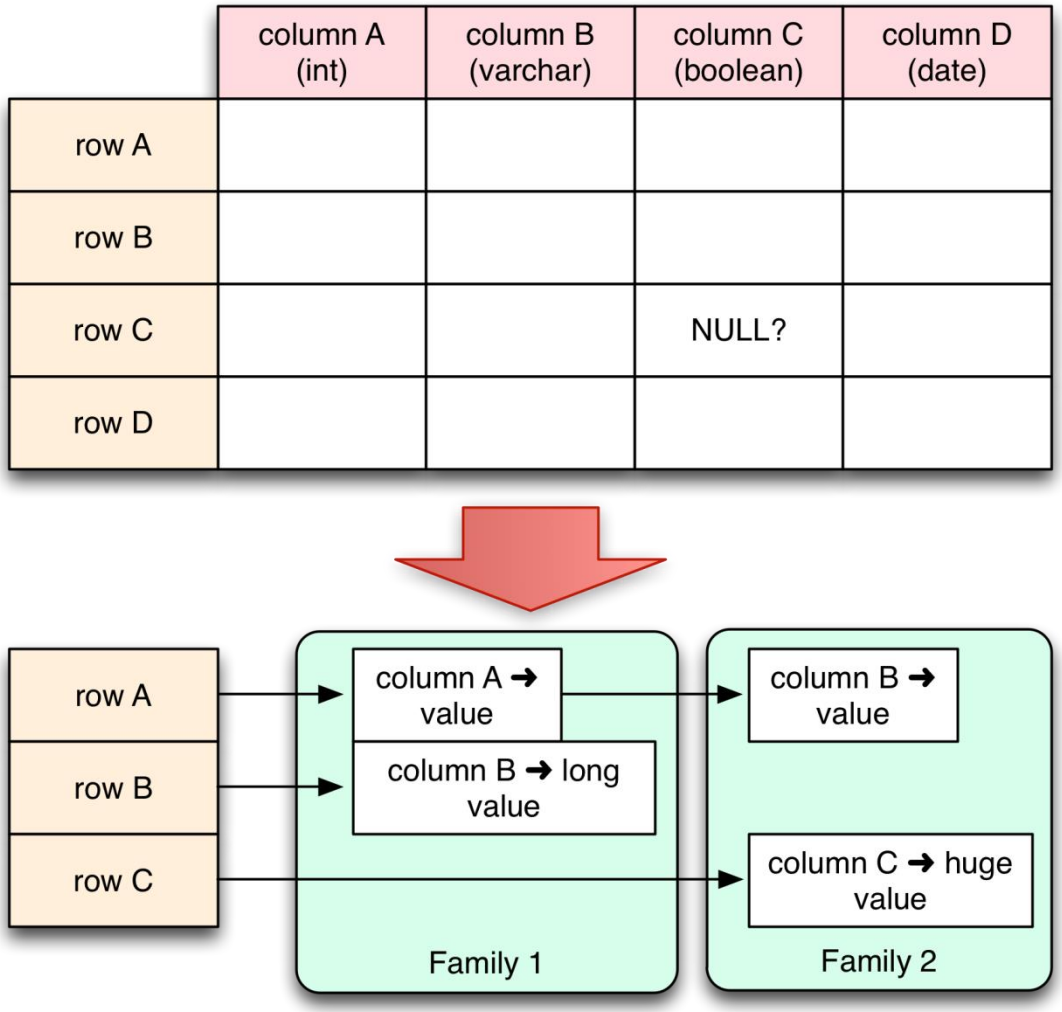
- HBASE模型
- 架构设计
- 使用技巧
- 运维技巧
- 测试分析

HBASE模型

Hbase数据模型

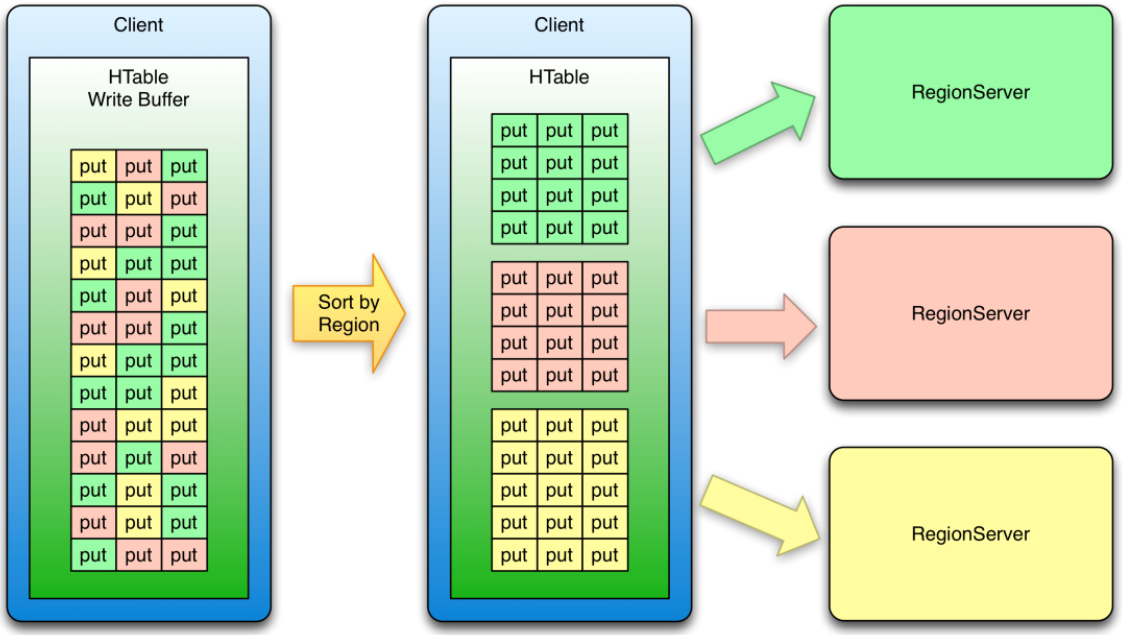
- Table
- Region
- ColumnFamily
- Row
- Column
- Version
- Value

CELL



HBASE操作

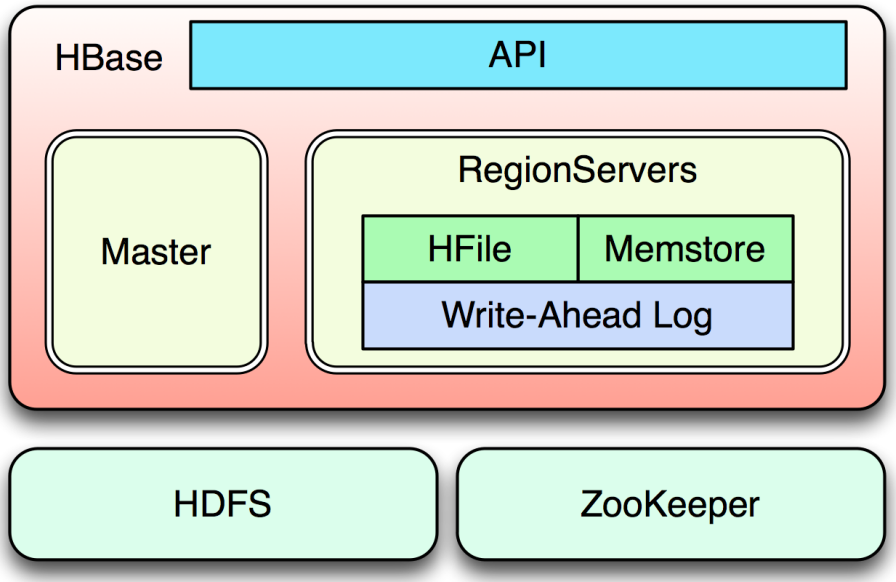
- Put
  - Delete
  - 原子操作
  - WAL
- Scan
  - Get
  - Filter
  - Cache/Batch
- 批量操作
- 行锁



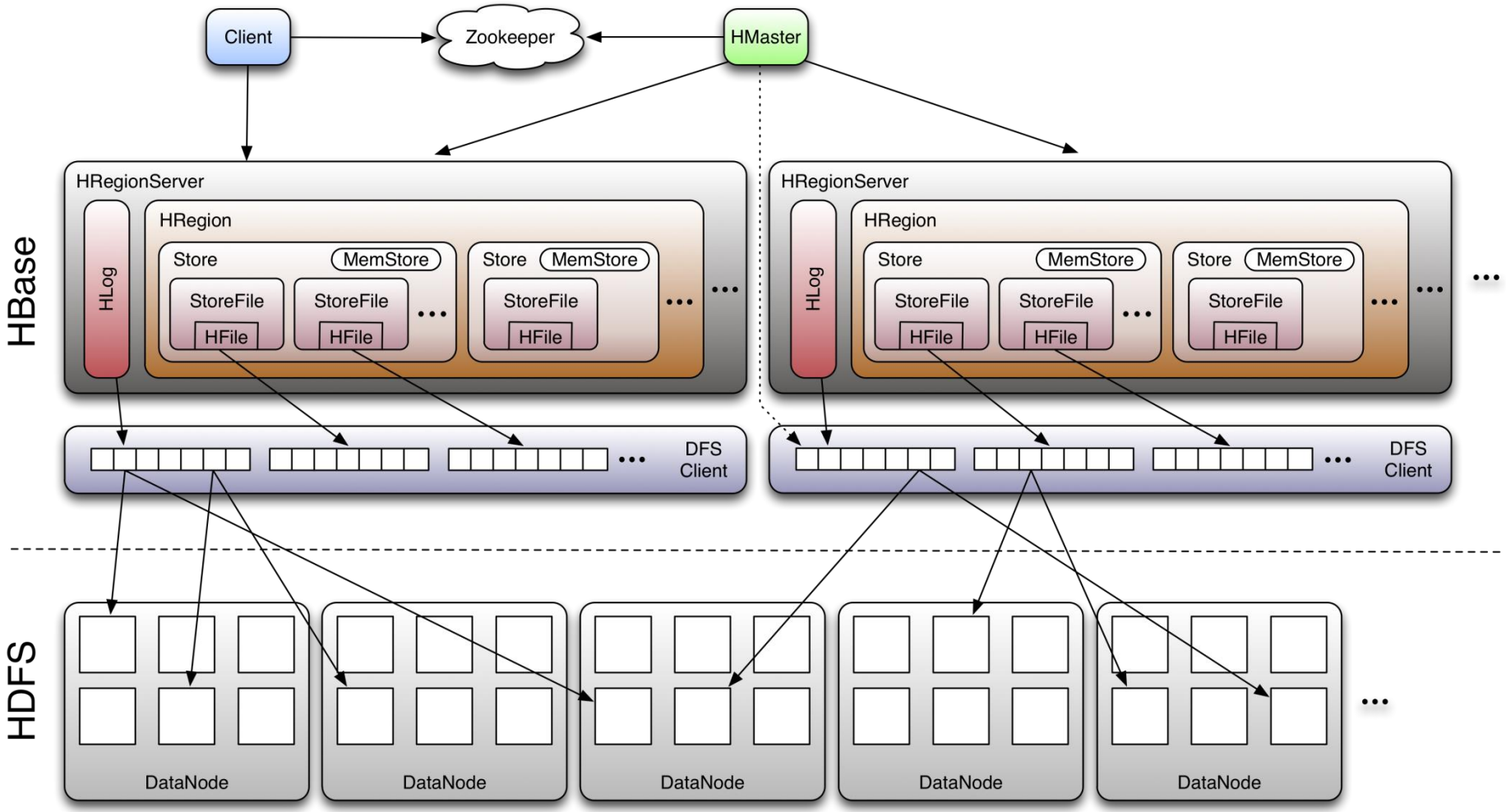
HBASE架构设计

总体结构

- Master
  - Region之上的操作
  - Put/Get不经过Master
- RegionServer
  - Region之下的操作
- HDFS
  - HFile
  - HLog
- ZooKeeper
  - 状态信息

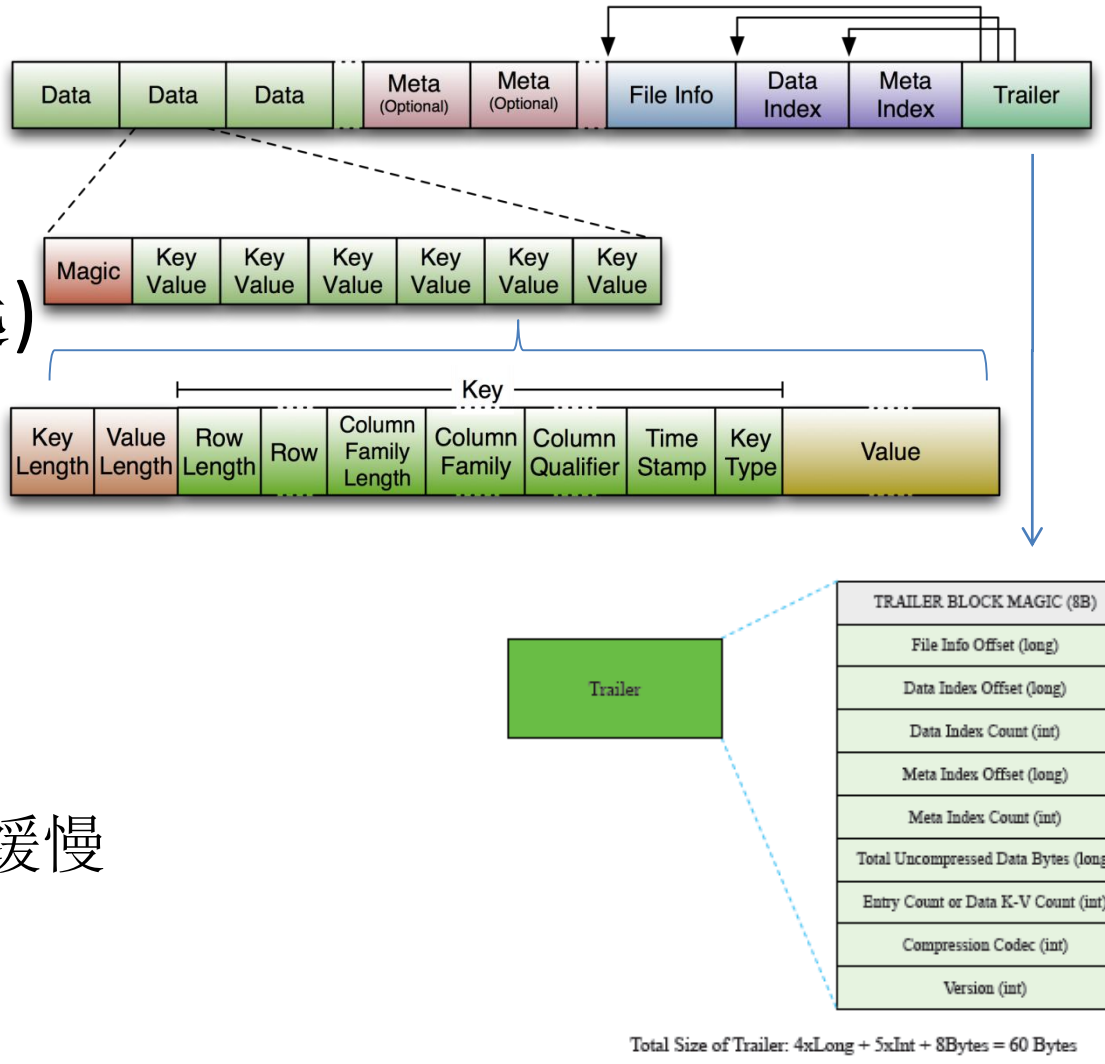


RegionServer结构



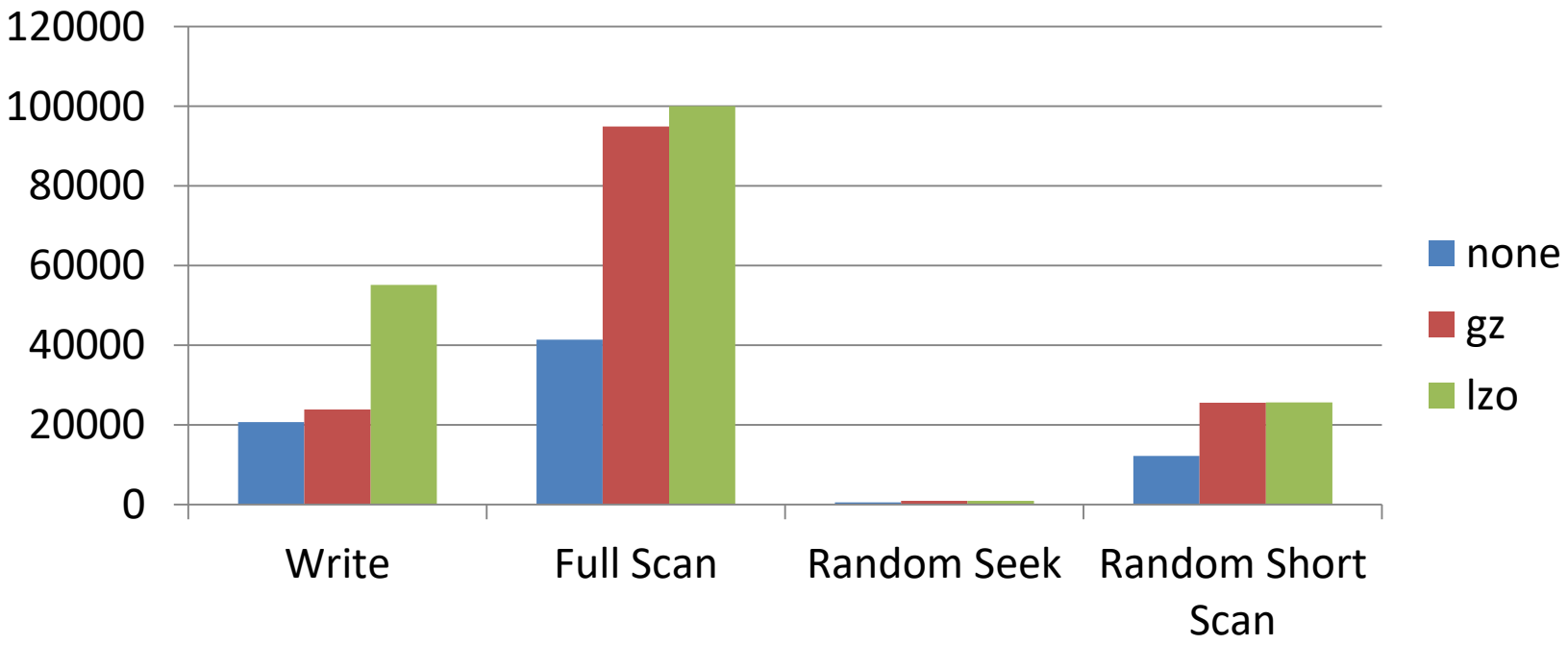
HFile结构

- 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



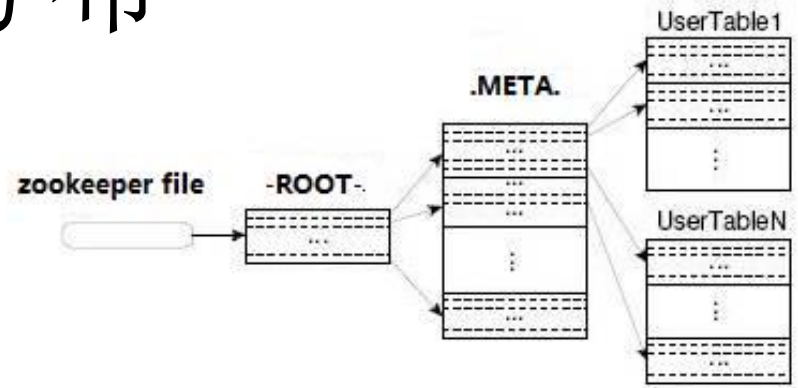
Hfile性能测试

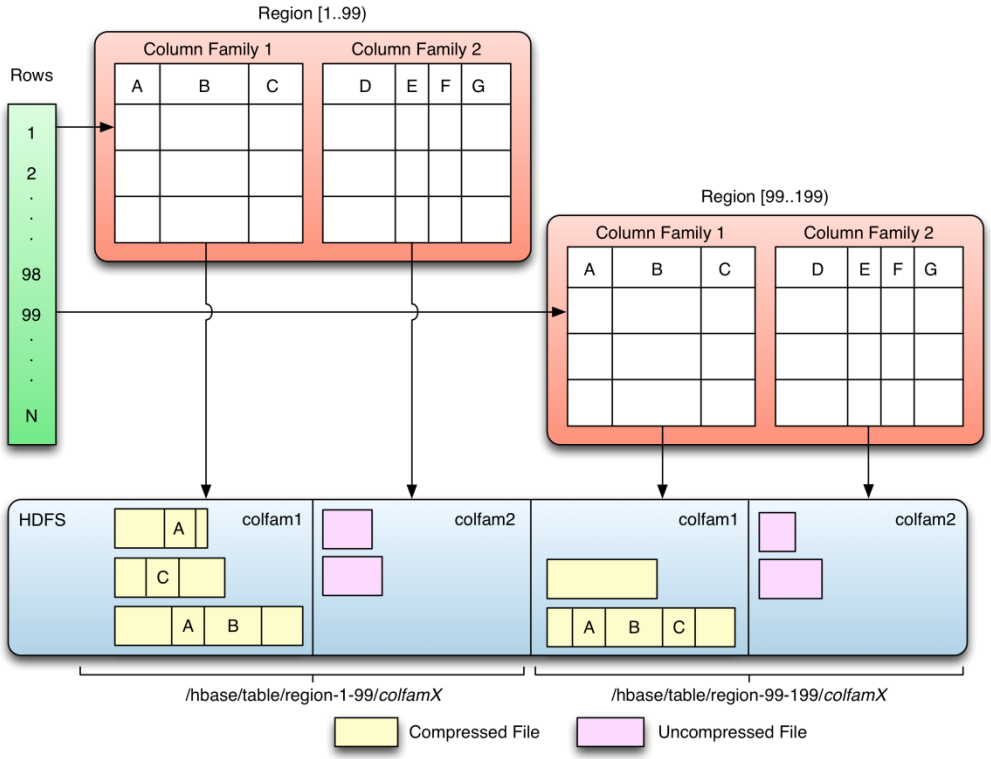
	none	gz	lzo
Write	20718	23885	55147
Full Scan	41436	94937	100000
Random Seek	600	989	956
Random Short Scan	12241	25568	25655



存储分布

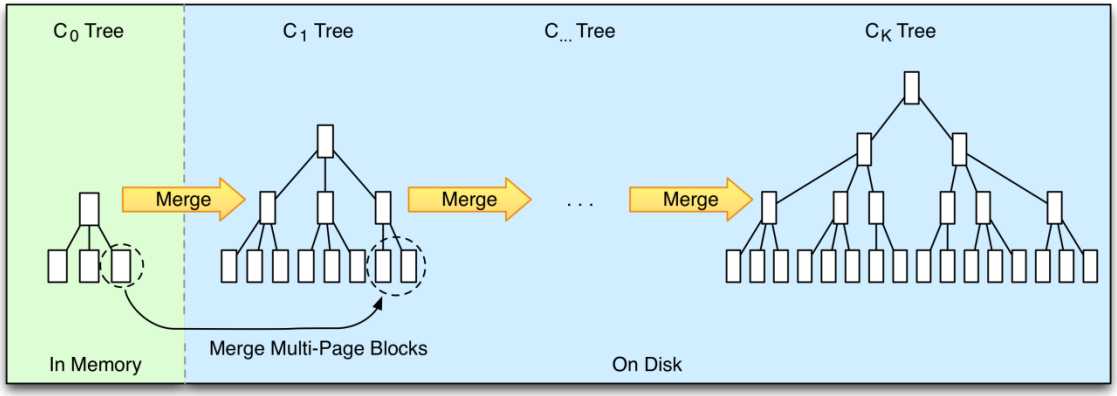
- 寻找RegionServer
  - ZooKeeper
  - -ROOT-(单Region)
  - .META.
  - 用户表

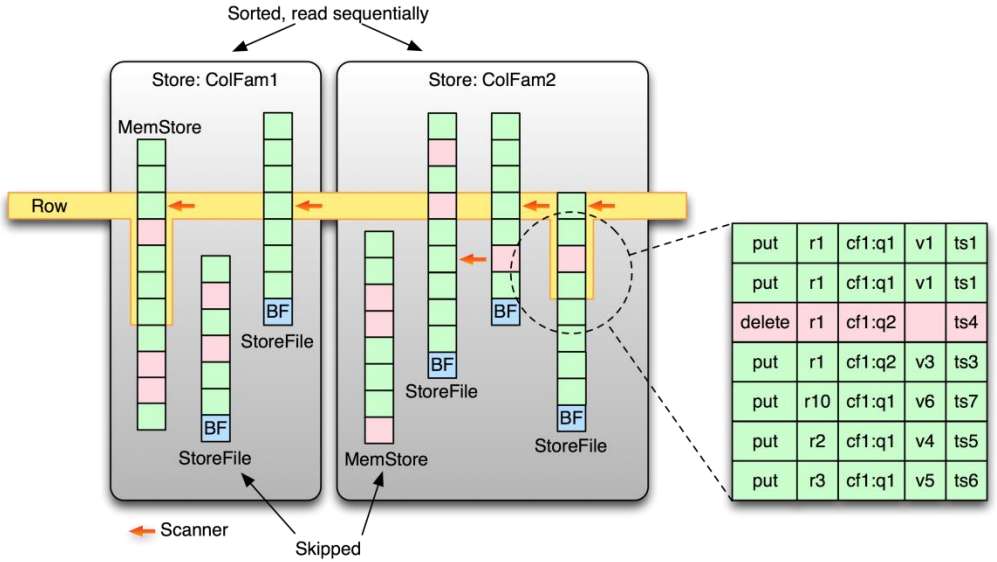




Put/Get操作

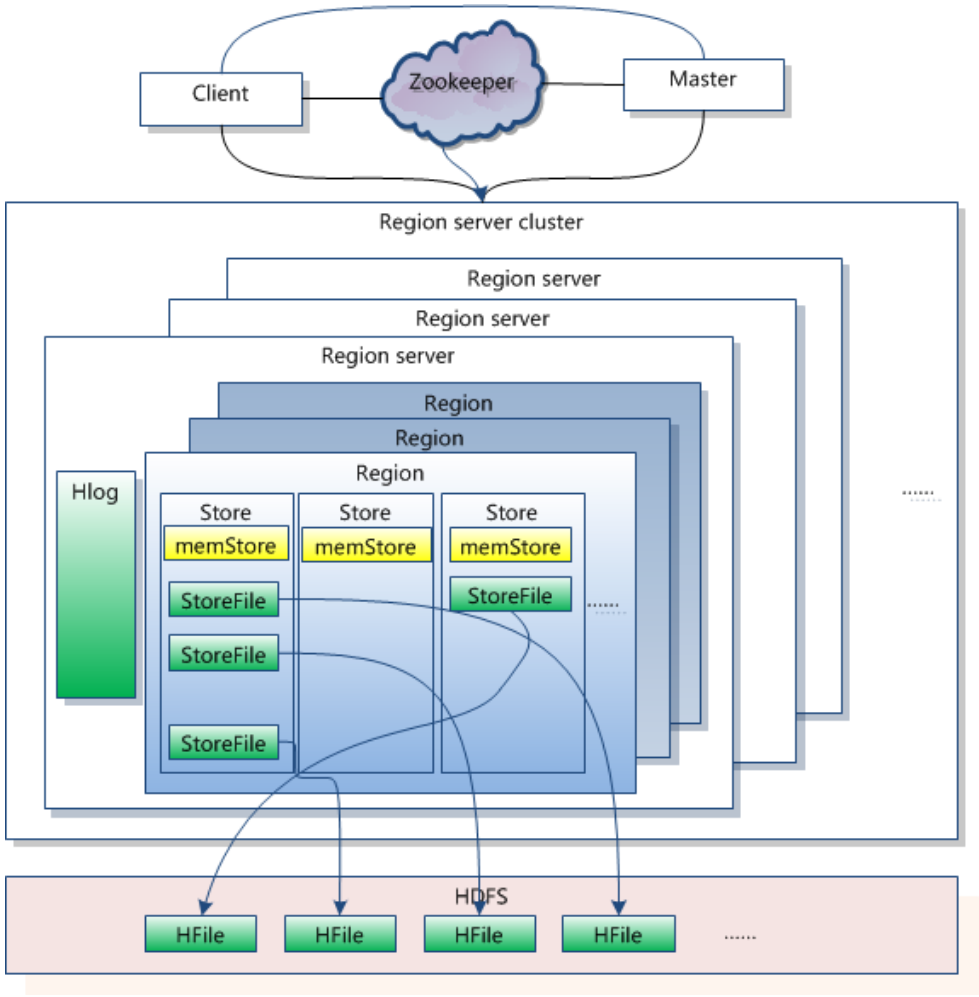
- PUT
- DELETE
- GET
- SCAN





Region操作

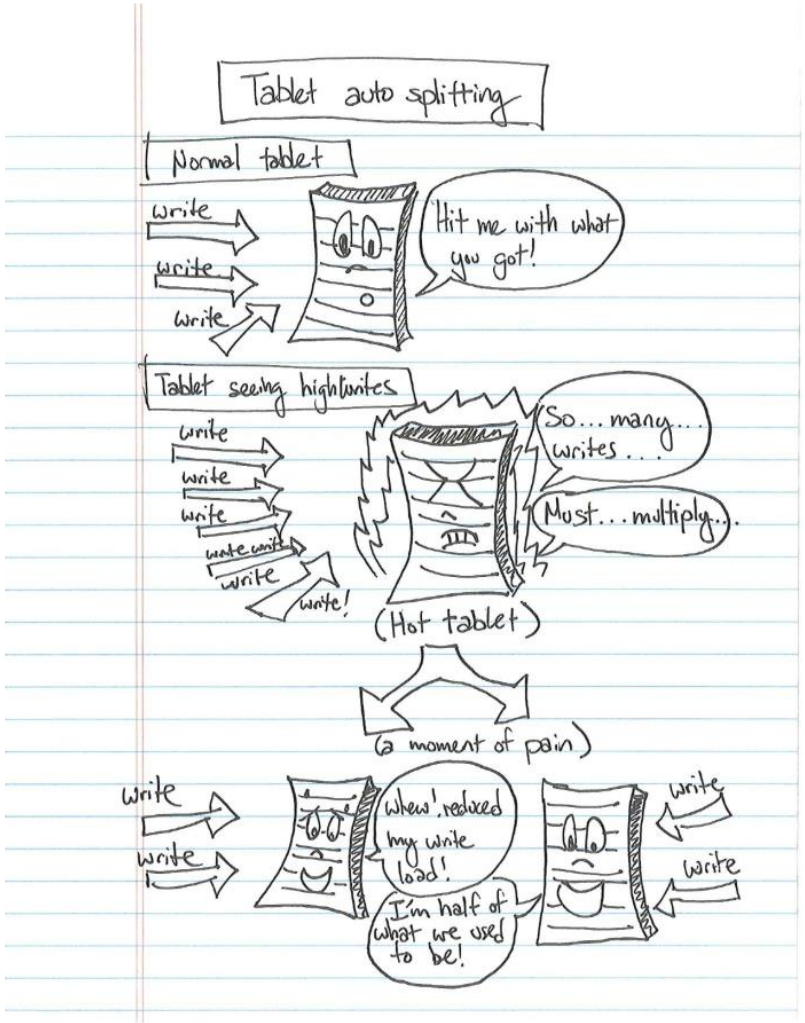
- Flush MemStore
- Compact
- Major Compact
- Split



使用技巧

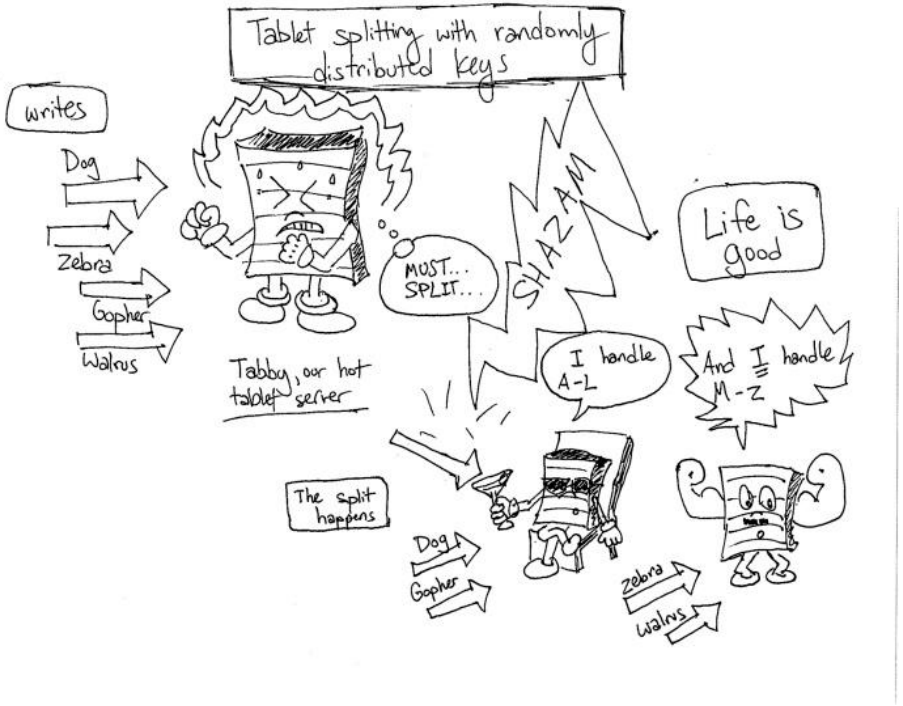
Schema设计

- 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



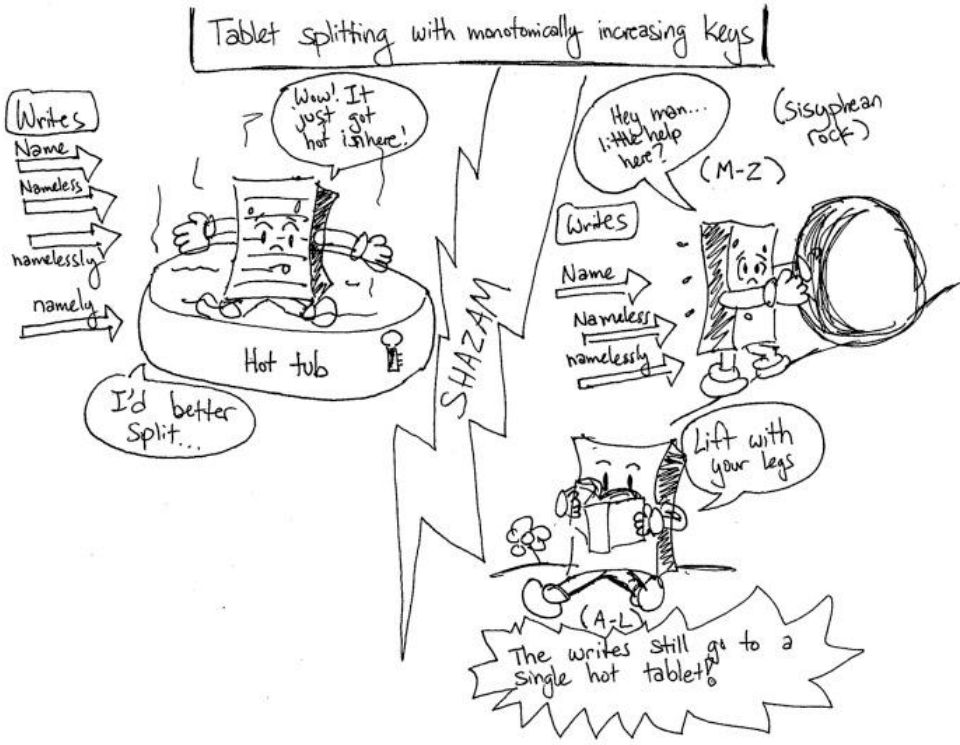
Schema设计

- 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



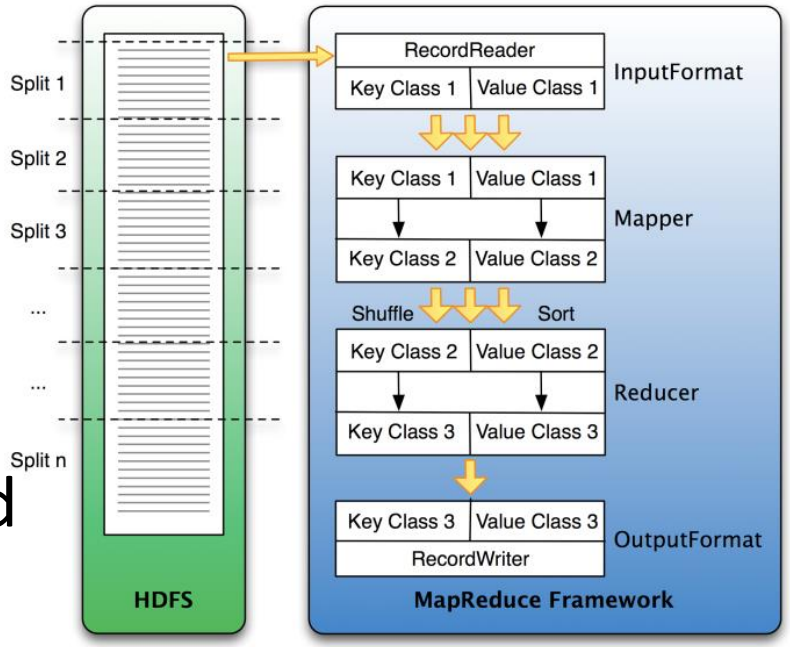
Schema设计

- Column Family的数量
  - 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 <=> Kind:ID

Index	Key
Column:Value	Key…

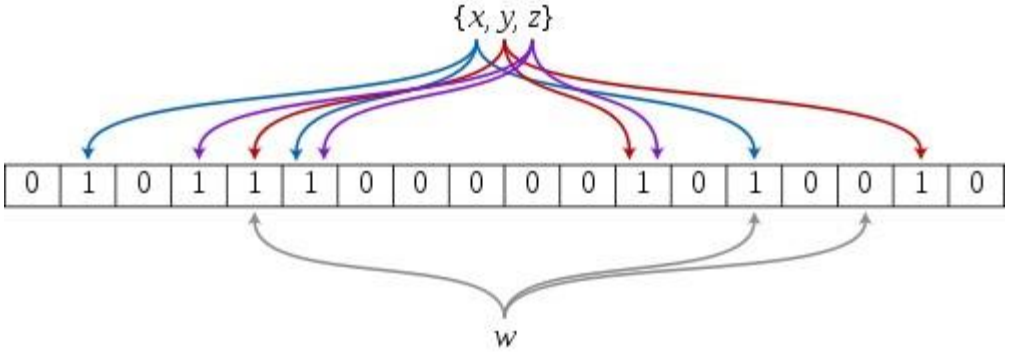
单列索引

Index	Key
Column:Value/Column:Value	Key…

组合索引

开发调优

- Table属性
  - BlockSize
  - BloomFilter
  - BlockCache
  - InMemory

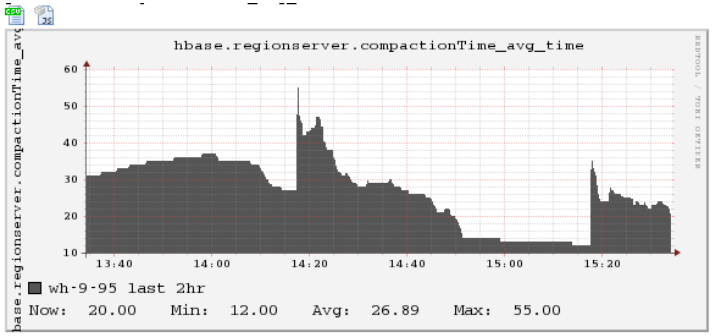


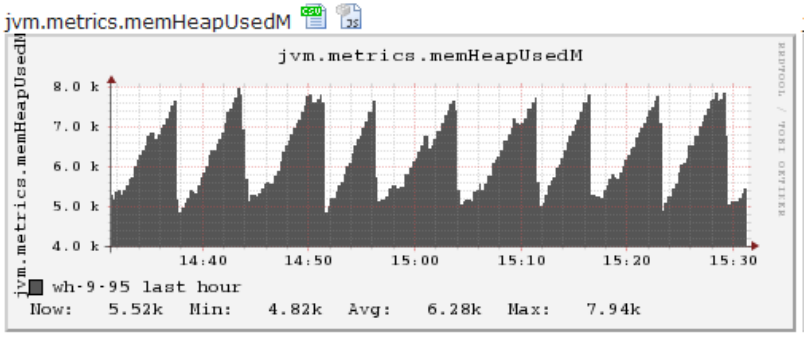
- 尽可能使用Bulk Load
- Put使用客户端Cache
- Scan使用Cache/Batch

运维技巧

HBASE 部署

- Hadoop版本
  - Hadoop 0.20.x
  - Append补丁
- ZooKeeper
- Metric
- 内存
  - RegionServer 12GB
    - MemStore <=40%
    - HFile DataIndex
    - BlockCache <=20%
  - Master 4GB
  - ZooKeeper 1GB





Region管理

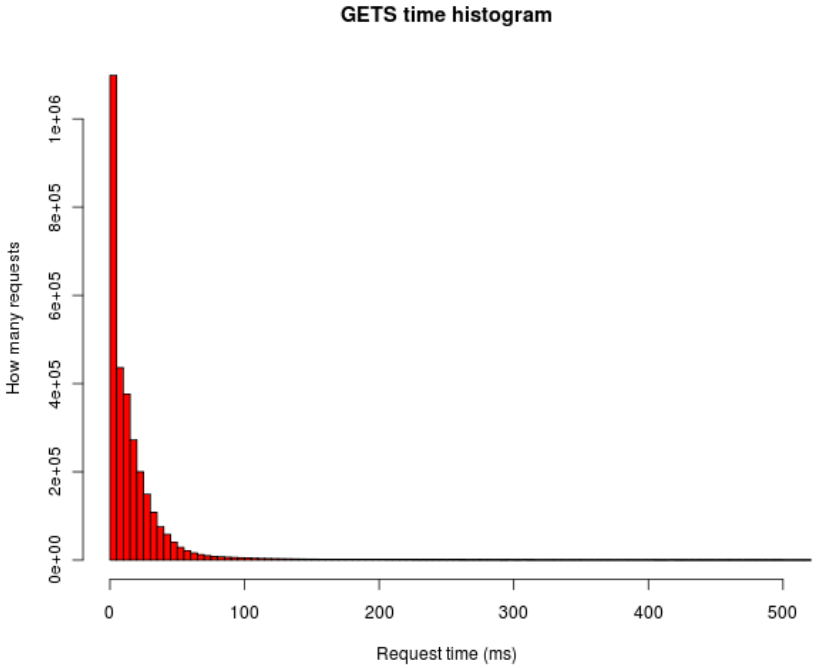
- 预创建Region
- Region的大小
  - hbase.hregion.max.filesize=256MB,1GB,4GB
  - 手动Split,交错负载
- Region合并
  - hbase.hstore.compactionThreshold=3
  - hbase.hstore.blockingStoreFiles=7(阻塞,超时)
  - hbase.hstore.compaction.max=10
  - hbase.hregion.majorcompaction=86400,0
- MemStore Flush
  - hbase.regionserver.global.memstore.upperLimit
  - hbase.regionserver.global.memstore.lowerLimit

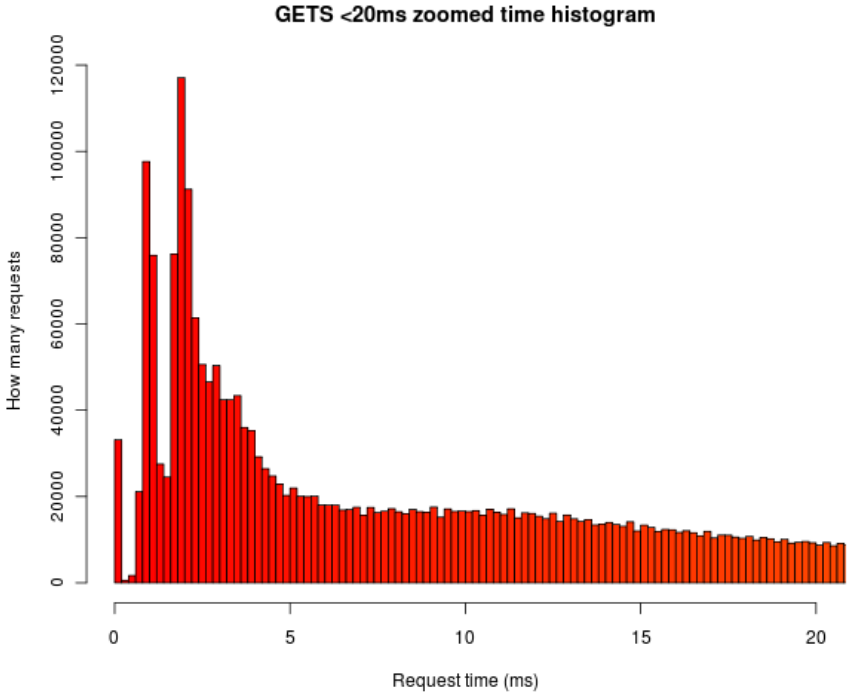
运维调优

- Java GC
  - JVM GC调整(ParNewGC+CMS)
  - Full GC-10s/GB
  - MemStore本地分配(2MB,减少碎片)
- LZO压缩
  - 压缩单位为Block
  - 提高性能
- 并发数调整
  - **hbase.regionserver.handler.count**
- Cache设置
  - **hfile.block.cache.size**

测试分析

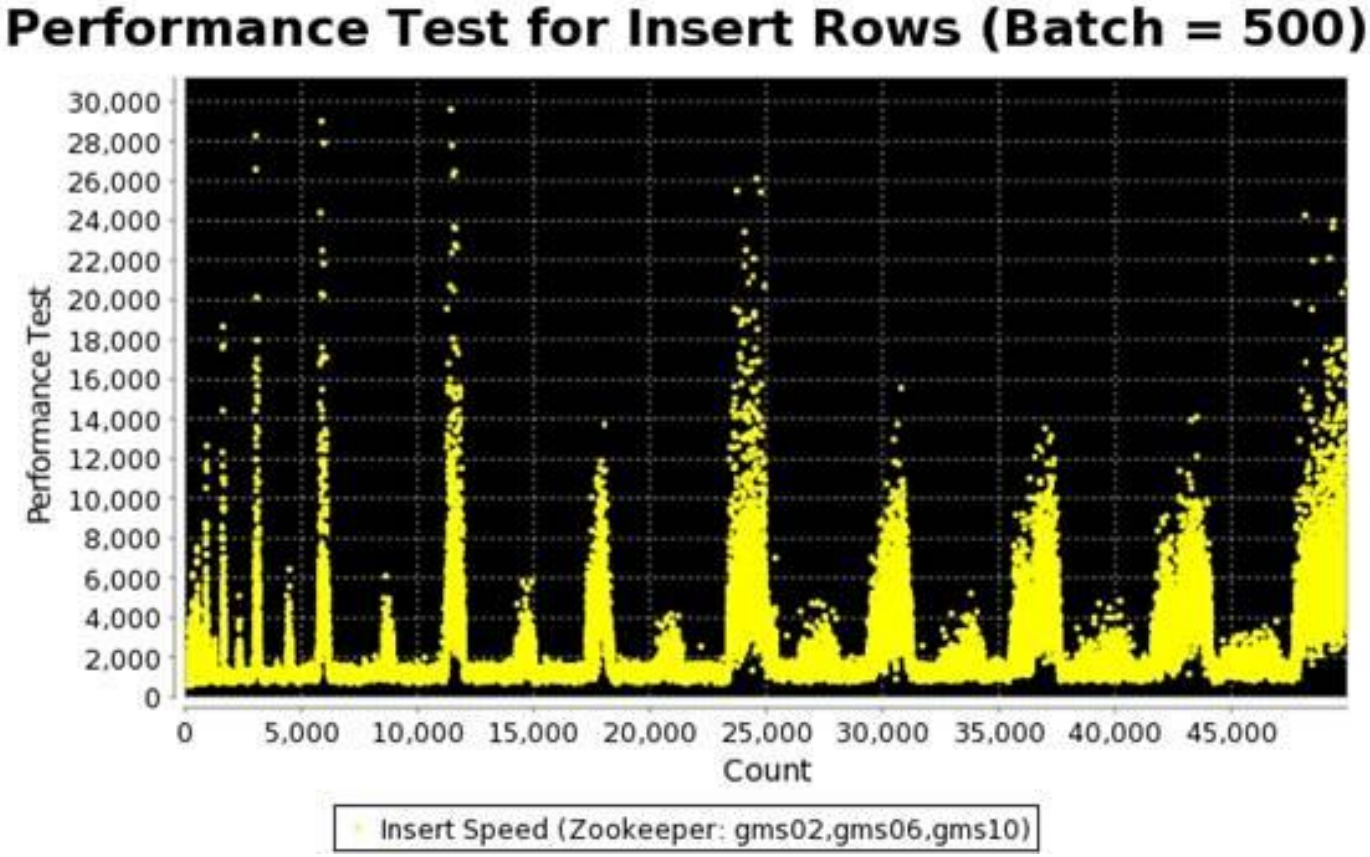
随机Get测试





- Get波动不是很大

Put测试



- 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