

HBase理论文件

HBase优化

- 高可用
 - conf目录下创建backup-masters文件
 - 分发到其他集群机器上
- 预分区
- RowKey设计
 - 设计RowKey的目的
 - 设计RowKey的方法
- 内存优化
 - 一般会分配整个可用内存的70%给HBase的Java堆
 - 太大容易GC太久导致RegionServer长期不可用
- 基础优化
 - 允许在HDFS的文件中追加内容
 - 优化DataNode允许的最大文件打开数
 - 优化延迟高的数据操作的等待时间
 - 优化数据的写入效率
 - 设置RPC监听数量
 - 优化HStore文件大小
 - 优化HBase客户端缓存
 - 指定scan.next扫描HBase所获取的行数

HBase与Hive集成

- HBase与Hive的对比
 - Hive
 - HBase
- HBase与Hive的集成使用

HBase进阶

- RegionServer
- HBase写流程
- MemStore Flush
- HBase读流程
- StoreFile Compaction
 - 定义
 - 分类
 - 图示
- Region Split
 - 定义
 - 时机
 - 图示

HBase简介

- HBase定义
- 概述
- HBase的逻辑结构
 - 架构说明
 - 图示
- HBase的物理存储结构
- HBase数据模型
 - 数据模型
 - Name Space
 - Table
 - Row
 - Column
 - Time Stamp
 - Cell
 - Region Server
 - Master
 - Zookeeper
 - HDFS
- HBase基本架构
 - 架构说明
 - 图示