

HDFS理论文件

大数据概论

- 概念
- 特点 (4V)
- 大数据部门组织结构
- 小Tips: 版本号看法

Hadoop优化

- MapReduce运行过慢的原因
- 参数调优
- MapReduce优化方法
 - 数据输入阶段
 - Map阶段
 - Reduce阶段
 - I/O阶段
 - 整体
- 数据倾斜问题
 - 概述
 - 优化方法
- HDFS小文件
 - 小文件问题
 - 解决方案

Hadoop数据压缩

- 压缩概述
- 压缩格式
- 各压缩格式对比

Hadoop框架的大数据生态

- 发展起源
 - GFS ---> HDFS
 - Map Reduce ---> MR
 - BigTable ---> HBase
- Hadoop三大发行版
- Hadoop的优势
- Hadoop组成
- 大数据技术生态系统
- 图示

HDFS运作原理

- HDFS (Hadoop Distributed File System) 概述
- HDFS读写流程
 - HDFS写数据的流程
 - HDFS读数据的流程
 - 网络拓扑距离
- NameNode和SecondaryNameNode
- DataNode
- HDFS高可用配置
 - HA概述
 - HDFS-HA工作机制