SDFS 文档

14302010037 欧承祖

1. 使用说明
   1. 在与SimpleDistributedFileSystem.jar文件同级目录下放一个test.pdf文件，无参数运行这个jar包，就会在同级目录下看到”NameNode/”,”DataNode/”和”test\_copy.pdf”，其中”NameNode/”和”DataNode/”分别是两个服务器存放自己数据的根目录，”test\_copy.pdf”为将”test.pdf”写入到SDFS后再从SDFS中读取到的文件。
   2. 每次操作结束后请删除”NameNode/”和”DataNode/”两个目录下的所有文件，因为用来测试的Main函数中写死了创建的路径，再次创建会抛异常，详情见源代码的Main函数。此举只是为了方便测试，不涉及本质，系统并无异常。
2. 完成内容
   1. 实现第一阶段文档中所提出的基本功能，包括需求文档中要求的Client、NameNode和DataNode分别需要支持的接口
   2. 基本按照提供的接口实现，少数地方做了细微改动，改动主要有：
      1. 调整包结构，进一步将Client、NameNode、DataNode分离，方便后续修改为分布式系统时代码的兼容性
      2. 增加一个entity包，用来存放分布式系统中网络通信交换的数据结构
      3. 优化NameNode中FileNode和DirNode的实现，同时考虑到接口统一性和网络传输数据应尽可能少的要求，并没有直接修改FileNode和DirNode，而是在他们外面包了一层Entry类，方便NameNode管理
      4. 修改DirNode的存储格式：将记录名字长度的比特数从三个改为了四个。目的是方便直接用int操作
3. 遇到问题
   1. 尚未处理宕机恢复问题
   2. 尚未支持增加冗余block
4. 额外工作
   1. 优化NameNode处理请求的效率，实现方法为：在服务器启动时将全部节点加载进内存，处理请求时不需要额外多次磁盘读写，只需在特定时刻将内存中的改动持久化进磁盘中就行。这样比起每个请求都从磁盘上读取效率要高。
   2. 模块层次分明，耦合度低，设计时已做考虑，系统易于拓展成分布式的。