学习总结

1812975 曹沛鑫 2020.6.25

1. HDFS命令行基本操作

·查看帮助 hdfs dfs -help

·查看当前目录信息 hdfs dfs -ls [-h][-r] <path> #-h以更友好的方式列出，主要针对文件大小显示成相应单位K、M、G等

·查看文件内容 hdfs dfs -cat <hdfsfile>

·上传文件到hdfs hdfs dfs -put [-f] <local src> <hdfs dst> #-f如果hdfs已经存在要上传的文件，则覆盖

·剪切文件 hdfs dfs -moveFromLocal a.txt /aa.txt

·下载hdfs文件到本地 hdfs dfs -get <hdfs path> <localpath>

·合并下载 hdfs dfs -getmerge <hdfs path> <localdst>

·创建目录 hdfs dfs -mkdir <path>

·创建多级目录 hdfs dfs -mkdir -p <path>

·新建空文件 hdfs dfs -touchz <filename>

·删除文件或目录 hdfs dfs -rm [-r][-f] [-skipTrash] <hdfs path> #-r递归删除目录下所有文件，-f直接删除，不予提示，-skipTrash彻底删除文件，不放入回收站

·拷贝 hdfs dfs -cp [-r] <hdfs path> <hdfs path1>

·移动 hdfs dfs -mv <hdfs path> <hdfs path1>

·统计目录下对象数 hdfs dfs -count <hdfs path>

·统计目录下的对象大小 hdfs dfs -du[-s][-h] <hdfs path>

·显示hdfs的容量、数据块和数据节点信息 hdfs dfsadmin -report

二、Hadoop分布式文件系统

1. HDFS是为以流式数据访问模式存储超大文件而设计的文件系统，在商用硬件上运行。
2. HDFS的块默认大小64MB。好处是文件可以大于网络中任意一个磁盘的容量，可以利用集群上的任意一个磁盘；可以简化存储子系统；适合于为提供容错和实用性而做的复制操作。执行命令：hadoop fsck / -files -blocks能列出文件系统中各个文件的块。

三、Hadoop集群搭建（伪分布式）

1.Linux基本环境配置

·修改主机名：vim /etc/sysconfig/network

·修改ip：vim /etc/sysconfig/network-scripts/ifcfg-eno16777736

·修改主机名和IP的映射关系：vim /et/hosts

·关闭防火墙：service iptables status

·修改/etc/sudoers文件权限：chmod u+w /etc/sudoers

·打开/etc/sudoers文件：vim /etc/sudoers；给用户添加执行的权限：

# hadoop 用户名

hadoop ALL=(root) NOPASSWD:ALL

2.Java环境安装，同昨天一样

四、总结

1.伪分布没做完，明天继续

2.Pdf看的云里雾里的，实践也云里雾里的