HDFS 文件系统

1. 为实现高可用性，需要保证一个备用的namenode，以在active namenode失效的时候进行接管，而如果备用的也失效，才另行配置
2. HDFS就是一个文件系统，可以简单的看作普通操作系统的文件系统去操作

Hadoop 搭建注意事项

按照此博客进行搭建<http://www.powerxing.com/install-hadoop/>

看log以确定网站的地址

配置极简命令：

1. xml文件和env脚本的目录为./etc/hadoop/
2. 要在以上目录的hadoop-env.sh的末尾加上java home的目录JAVA\_HOME=\*\*\*
3. 配置一些xml文件：
   1. core-site.xml的配置

<configuration>

<property>

<name>hadoop.tmp.dir</name> #配置临时文件位置，防止开机重置

<value>file:/usr/local/hadoop/tmp</value>

<description>Abase for other temporary directories.</description>

</property>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

* 1. hdfs-core.xml的配置

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:/usr/local/hadoop/tmp/dfs/name</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>file:/usr/local/hadoop/tmp/dfs/data</value>

</property>

</configuration>

4)初始化文件系统

hdfs namenode –format

5)开启守护进程

start-dfs.sh

6)可以对hdfs文件系统开始操作了

如hdfs dfs –mkdir input 创建了一个目录

7）使用yarn管理

start-yarn.sh

mapred --daemon start historyserver

注意，如果有yarn的配置文件，就必须启动yarn，否则会报错

第六章，mapreduce的工作机制，

@存在调度，以及不同的调度方案@存在出错处理@其他一些机制

第七章，Mapreduce的类型与格式

# [第十章 Hive](hive学习.docx)