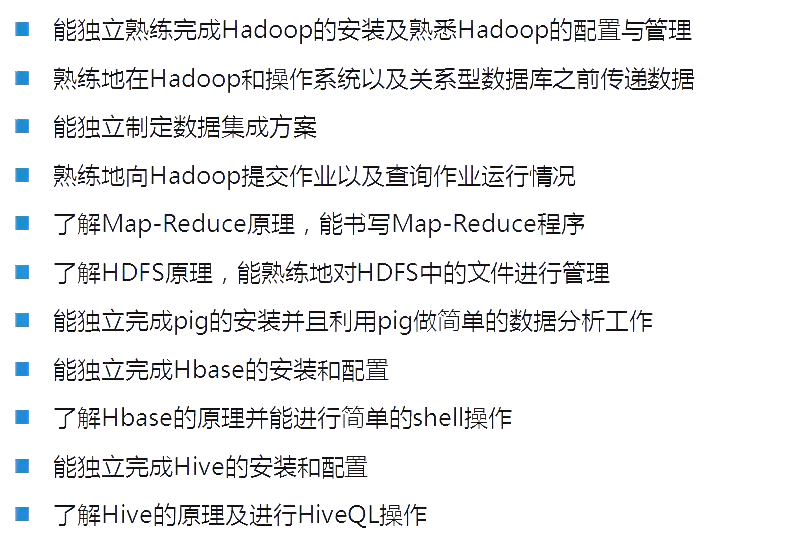
Hadoop笔记

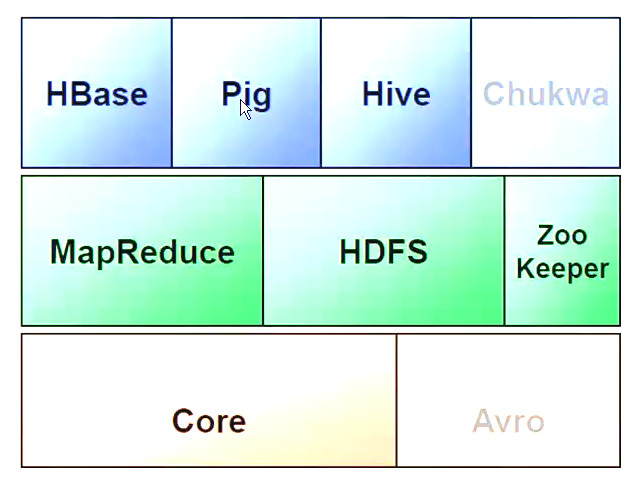
# Hadoop的技能



# Hadoop的特点

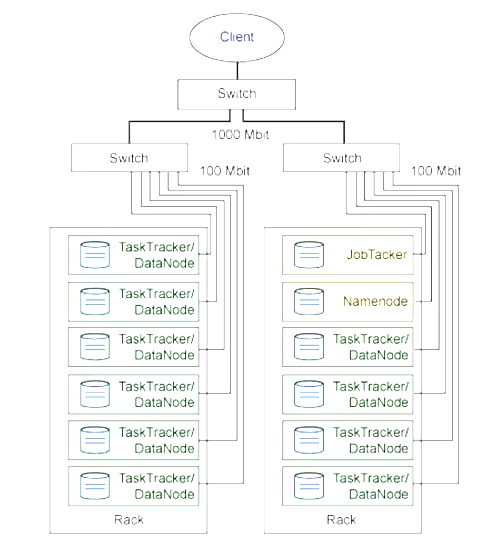
扩容能力、成本低、高效率、可靠性

# Hadoop家族



# Hadoop的架构

## Hadoop集群的物理分布

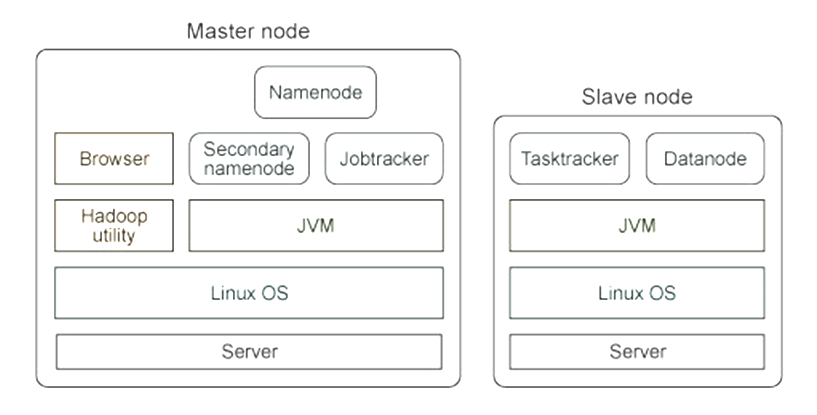


Master：NameNode、SecondaryNameNode、JobTracker

Slave：TaskTracker、DataNode

Master不是唯一的

## 单结点物理结构



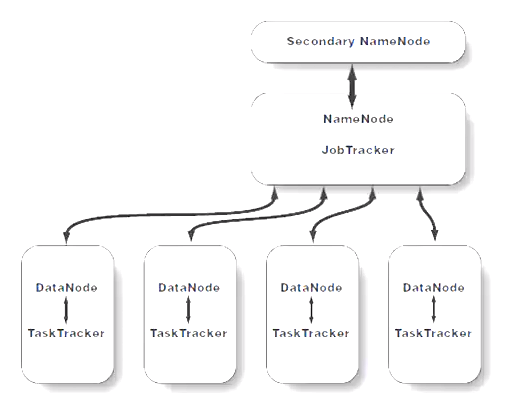
# MapReduce的架构

主从结构

主节点，一个：JobTracker

从节点，多个：TaskTracker

## JobTracker



用于处理作业的后台程序

接收客户提交的计算任务

把任务分配给TaskTracker执行

决定有哪些文件参与处理，然后切割task并分配节点

监控tasktracker，重启失败的task

每个集群只有一个JobTracker，位于master节点

## TaskTracker

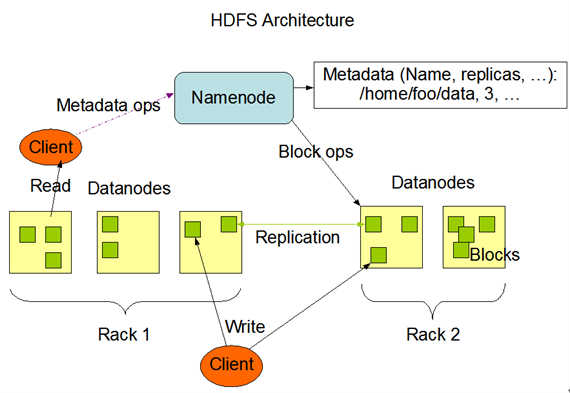
位于slave节点，与datanode结合

管理各自节点的task（JobTracker分配）

每个节点只有一个tasktracker，但一个tasktracker可以启动多个jvm，用于并行执行map或reduce任务

与jobtracker交互，执行JobTracker分配的计算任务

**# HDFS体系结构**



**HDFS的架构**

主从结构

**## NameNode**

HDFS的守护程序

本质是一个jetty服务器

对内存和IO进行集中管理

接收用户操作请求

提供名称查询服务，是一个jetty服务器。

保存的metadata信息（文件owership和permission、跟踪文件如何被分割成文件块，文件块又被哪些节点存储，即文件数据块列表）

NameNode是单点

**metadata**存储到**fsimage**



## DataNode:

保存块

HDFS数据块读写到本地文件系统

启动DN线程时向NN汇报block信息

向NN发送心跳保持联系

Block的副本放置策略：

第一个副本，放置在本地机架的节点上

第二个副本，第二个副本放在同一机架的另一个节点上

第三个副本，放置在不同机架的节点上

Block默认64MB，不够默认仍占用一个block的命名空间，但物理上并不会。

Block大小和副本数由client端上传文件时hdfs设置，其中副本数可以变更，block不可以上传后变更。

## 数据损坏

DN读取block的时候，会计算checksum

## SecondaryNameNode（SNN）

它不是NN的热备份

它可以作为冷备份：

将本地保存的fsimage导入

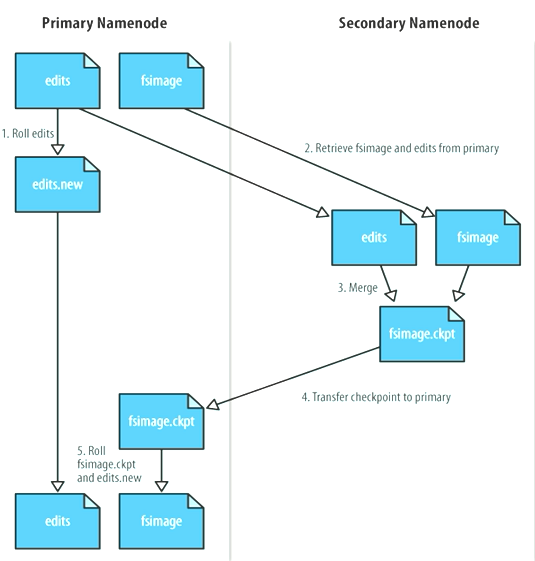
修改cluster所有DN的NameNode地址

修改所有client端NameNode地址

或者修改SNN IP为原NN IP

它的工作是帮助NN合并edits，减少NN启动时间

## SecondaryNameNode



简单描述一下具体流程：

步骤一：SSN在一个checkpoint时间点和NameNode进行通信，请求NameNode停止使用edits文件记录相关操作而是暂时将新的Write操作写到新的文件edit.new来。

步骤二：SSN通过HTTP GET的方式从NameNode中将fsimage和edits文件下载回来本地目录中。

步骤三：SSN中合并edits和fsimage。SSN将从NameNode中下载回来的fsimage加载到内存中，然后逐条 执行edits文件中的各个操作项，使得加载到内存中的fsimage中包含edits中的操作，这个过程就是所谓的合并了。

步骤四：在SSN中合并完fsimage和edits文件后，需要将新的fsimage回传到NameNode上，这个是通过HTTP POST方式进行的。

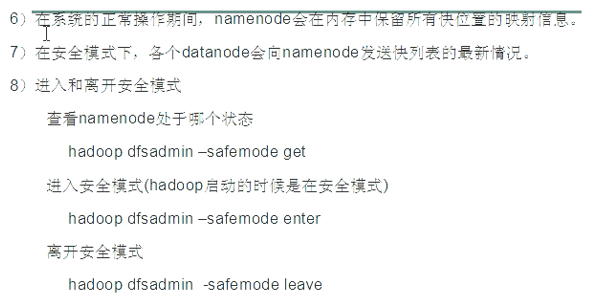
步骤五：NameNode将从SSN接收到的新的fsimage替换掉旧的fsimage。同时将edits.new文件转换为通常的edits文件，这样edits文件的大小就得到减少了。SSN整个合并以及和NameNode的交互过程到这里已经结束。

## HDFS权限

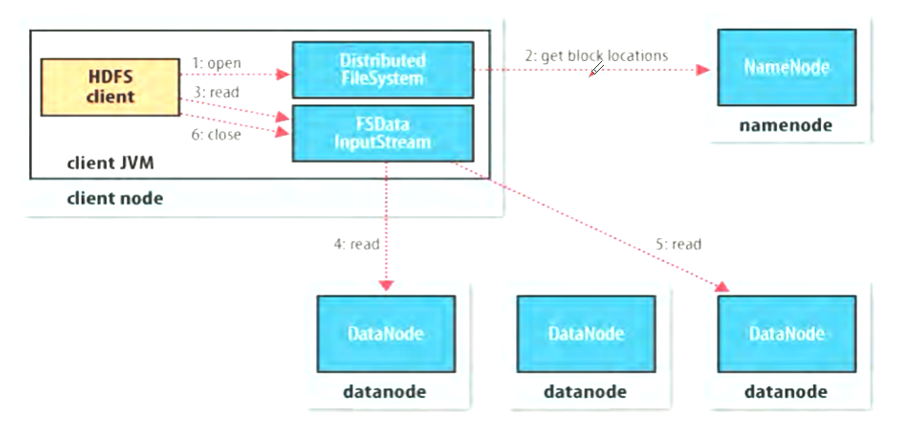
rwx权限

## 安全模式

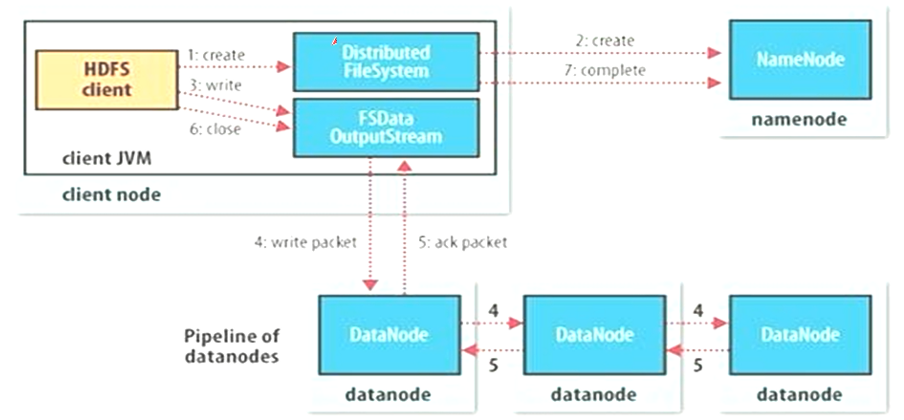




## HDFS文件读取



## HDFS文件写入



## HDFS开发常用命令

列出hdfs下的文件

hadoop fs –ls 目录

如列出HDFS目录下某个文档中的文件

hadoop fs –ls 目录（HDFS目录）

递归查看

hadoop fs –lsr 目录

创建文件夹

hadoop fs –mkdir /d1

上传文件

hadoop fs –put <linux-source> <hdfs-destination>

hadoop fs –put /root/a.txt /d1

注意：若/d1文件夹不存在，则/d1默认为文件名

同理hadoop fs –put /root/a.txt /d1/newa.txt 则被修改为文件名

将HDFS中的文件复制到本地系统中

hadoop fs –get <hdfs -souce> <linux-destination>

hadoop fs -get /d1/a.txt /root

删除HDFS下的文档

hadoop fs -rm /d1/a.txt

hadoop fs -rmr /d1 递归删除

查看HDFS下的某个文件

hadoop fs –text /d1/a.txt

管理与更新

hadoop dfsadmin -report

安全模式

退出 hadoop dfsadmin -safemode leave

进入 hadoop dfsadmin -safemode enter

拷贝

hadoop fs –cp <src><des>

拷贝多个文件时<des>需要是文件夹

移动

hadoop fs –mv <src><des>

其他命令

hadoop fs自行查看

hadoop fs –help

注意hdfs应使用绝对路径 如果是相对路径 默认是/user/<currentUser>

HDFS适用于一次写入多次查询，不支持并发，小文件不适合

# 安装和配置

## hadoop环境

虚拟机VMware Workstation10

ubuntu 14.04

jdk-6u24-linux-.bin

hadoop-1.1.2.tar.gz

## 伪分布式安装

注意：以下操作均使用root用户登录，不是su root

**关闭防火墙** （主机和虚拟机ufw disable）

关闭虚拟机防火墙

相关命令ufw disable|enable ufw status

**修改ip （桥接模式 静态IP）**

**编辑文件/etc/network/interfaces:  
sudo vi /etc/network/interfaces  
  
auto eth0  
iface eth0 inet static  
address 192.168.3.90  
gateway 192.168.3.1  
netmask 255.255.255.0**

**网卡设置生效sudo /etc/init.d/networking restart**

**修改hostname以及绑定**

1、ssh登录需要安装openssh

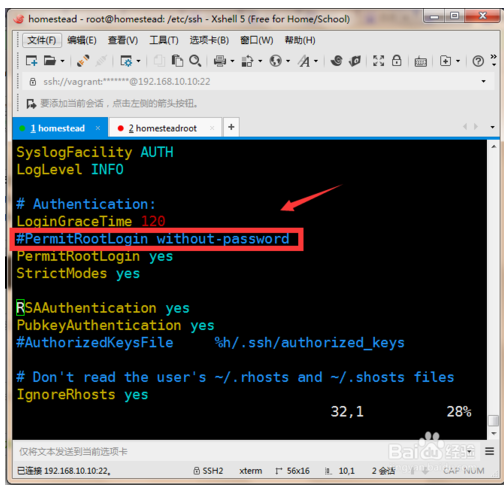
2、xshell普通用户直连 root用户拒绝

========解决方法==============

a、确认ssh服务已安装，通过普通用户连接成功；

b、确认ssh配置是否对root进行特殊设置，修改/etc/ssh/sshd\_config文件中

PermitRootLogin without-password将 without-password改为yes；



c、重启ssh服务/etc/init.d/ssh restart；

======================

3、解决修改主机名重启失效

========解决方法=============

a、centos里的/etc/sysconfig/network对应ubuntu里的/etc/network/interfaces文件，修改NETWORKING=yes HOSTNAME=hadoop 无效。（适用centos）

正确的方法：主机名存放在/etc/hostname（适用ubuntu）

b、hostname绑定

vim /etc/hosts

===============================

**设置ssh自动登录 （虚拟机ubuntu安装openssh）**

<1> 执行命令ssh-keygen –t rsa产生密钥，位于目录~/.ssh/

<2> cp id\_rsa.pub authorized\_keys

<3> 验证 ssh localhost

**安装jdk**

**安装hadoop**

<1> 配置HADOOP\_HOME环境变量

<2> 修改hadoop的4个配置文件 $HADOOP\_HOME/conf下

**hadoop-env.sh**

配置

```export JAVA\_HOME=/usr/local/jdk1.6.0\_24```

**core-site.xml**

配置

```

<configuration>

<property>

<name>fs.default.name</name>

<value>hdfs://hadoop:9000</value>

<description>改为自己的主机名</description>

</property>

<property>

<name>hadoop.tmp.dir</name>

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

</property>

</configuration>

```

**hdfs-site.xml**

配置

```

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

</configuration>

```

**mapred-site.xml**

配置

```

<configuration>

<property>

<name>mapred.job.tracker</name>

<value>hadoop:9001</value>

<description>改为自己的主机名</description>

</property>

</configuration>

```

<3> Hadoop格式化

命令hadoop namenode -format

<4> 启动与验证

start-all.sh jps（5个守护进程）

NameNode界面: 浏览器查看hadoop:50070

Map/Reduce界面: 浏览器查看hadoop:50030

Windows/mac下查看 修改hosts

其他问题：多次格式化hadoop错误？

解决方法：删除/usr/local/hadoop/tmp文件夹，重新格式化

解决$HADOOP\_HOME is deprecated

令$HADOOP\_HOME\_WARN\_SUPPRESS=0

## **Vmware的三种网络方式**

Vmnet0 桥接

Vmnet1 主机模式

Vmnet8 NAT模式

Vmware虚拟机三种网络模式详解

*地址：http://blog.csdn.net/noob\_f/article/details/51099040*

桥接模式Vmnet0

什么是桥接模式？桥接模式就是将主机网卡与虚拟机虚拟的网卡利用虚拟网桥进行通信。在桥接的作用下，类似于把物理主机虚拟为一个交换机，所有桥接设置的虚拟机连接到这个交换机的一个接口上，物理主机也同样插在这个交换机当中，所以所有桥接下的网卡与网卡都是交换模式的，相互可以访问而不干扰。在桥接模式下，虚拟机ip地址需要与主机在同一个网段，如果需要联网，则网关与DNS需要与主机网卡一致。

适用： 桥接模式配置简单，但如果你的网络环境是ip资源很缺少或对ip管理比较严格的话，那桥接模式就不太适用了。

NAT地址转换模式（Vmnet8）

如果你的网络ip资源紧缺，但是你又希望你的虚拟机能够联网，这时候NAT模式是最好的选择。NAT模式借助虚拟NAT设备和虚拟DHCP服务器，使得虚拟机可以联网。

主机模式（VMnet1）

Host-Only模式其实就是NAT模式去除了虚拟NAT设备，然后使用VMware Network Adapter VMnet1虚拟网卡连接VMnet1虚拟交换机来与虚拟机通信的，Host-Only模式将虚拟机与外网隔开，使得虚拟机成为一个独立的系统，只与主机相互通讯。

###

**Vmware虚拟机-Ubuntu14.04**

**IP地址：192.168.3.117**

**账户saber密码1992**

**账户root密码root**

4、其他

ubuntu重启网卡：

**先关闭接口:sudo ifconfig eth0 down**

**然后打开:sudo ifconfig eth0 up**

## MapReduce

1、MapReduce工作机制



2、调度机制

缺省为先入先出作业调度

支持公平调度器

支持容量调度器

3、任务执行优化机制

推测式执行：即如果jobtracker发现有拖后腿的任务，会再启动一个相同的备份任务，然后哪个先执行完就会kill掉另一个。因此在监控网页上经常能看到正常执行完的作业有被kill掉的任务。

推测式执行缺省打开，但如果是代码问题，并不能解决问题，而且会使集群变慢，通过在mapred-site.xml配置文件中设置mapred.map.tasks.speculative.execution和mapred.reduce.tasks.speculative.execution可为map或reduce任务开启或关闭推测式执行。

重用JVM，可以省去启动新的JVM消耗的时间和资源，在mapred-site.xml中设置mapred.job.reuse.jvm.num.tasks设置单个JVM上运行的最大任务数（1，>1和-1），-1表示没有限制，系统安全。

忽略模式，任务在读取数据失败2次后，会把数据位置告诉jobtracker，后者重新启动该任务并且在遇到所记录的坏数据时自动跳过。

4、错误处理机制：硬件故障

·硬件故障是指jobtracker故障或tasktracker故障

·jobtracker是单点，若发生故障目前hadoop还无法处理，所以选择牢固的硬件为jobtracker

·jobtracker通过心跳（周期为1分钟）了解tasktracker是否发生故障或负载过于严重

·jobtracker从任务节点中移除发生故障的tasktracker

·如果发生故障的节点在执行map任务并且尚未完成，jobtracker会要求其他节点重新执行此map任务

·如果故障节点在执行reduce任务并且尚未完成，jobtracker会要求其他节点继续执行尚未完成的reduce任务

错误处理机制：任务失败

·由于代码缺陷或进程崩溃引起任务失败

·JVM自动退出，向tasktracker父进程发送错误信息，错误信息被写入日志

·tasktracker监听程序若发现进程退出，或进程很久没有更新信息送回，将任务标记为失败

·标记失败任务后，任务计数器减一以便接受新任务，并通过心跳信号告诉jobtracker任务失败的信息。

·jobtracker获悉任务失败后，将把该任务重新放入调度队列，重新分配出去再执行

·如果一个任务失败超过4次（可以设置），将不会再被执行，同时作业也宣布失败

5、