Hadoop核心项目：

1. HDFS：Hadoop Distributed File System 分布式文件系统
2. MapReduce:并行计算框架

HDFS的架构：

主从结构：

主节点：只有一个，namenode

从节点：有很多个，datanodes

Namenode负责：

1. 接收用户操作请求
2. 维护文件系统的目录结构
3. 管理文件与block之间的关系，block与datanode之间的关系

Datanode负责：

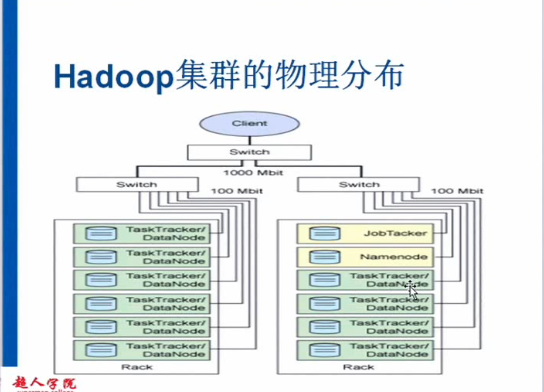
1. 存储文件
2. 文件被分为block存储在磁盘上
3. 为保证数据安全，文件会有多个副本

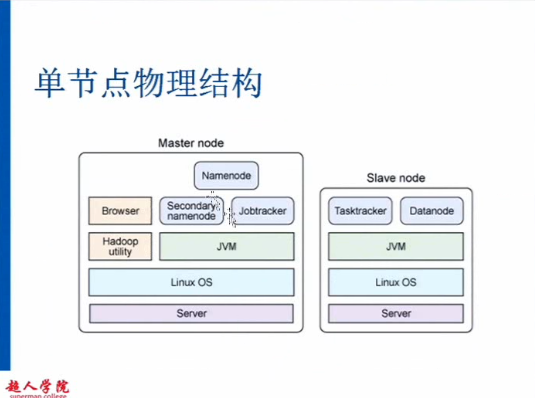
MapReduce架构：

1. 主从结构
   1. 主节点，只有一个：JobTracker（项目经理）
   2. 从节点：有很多TaskTrackers(开发人员)
2. JobTracker 负责
   1. 接收客户提交的计算请求
   2. 把计算任务分配给TaskTrackers执行
   3. 监控TaskTracker的执行情况
3. TaskTrackers负责
   1. 执行JobTracker分配的计算任务

Hadoop特点：

1. 扩容能力（Scalable）：PB级数据
2. 成本低（Economical）：普通机器组成服务群
3. 高效率(Efficient)： 分发数据，在各个节点并行处理他们，使得整体处理速度很高
4. 可靠性(Reliable)：自动维护任务多个副本，并且在任务失败后能自动重新部署





主从节点的安装机器是没有区别的，主节点还是从节点根据机器上的进程是否为JobTracker和NameNode

Hadoop部署方式：

1. 本地模式：生产中不用
2. 伪分布模式：学习使用
3. 集群模式：开发生产用

伪分布模式安装步骤：

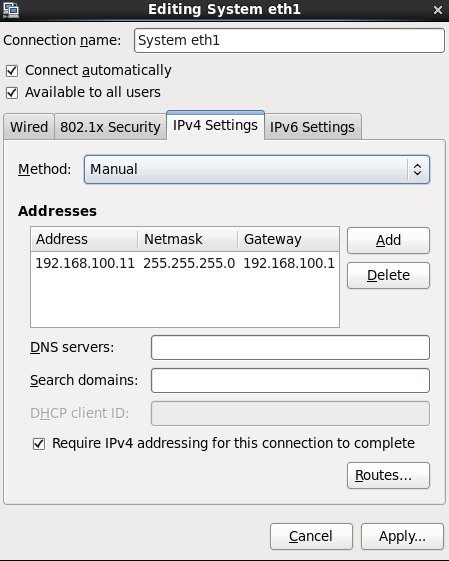
1. 关闭防火墙
2. 修改IP
3. 修改hostName
4. 设置ssh自动登录
5. 安装jdk
6. 安装hadoop

虚拟机的上网方式改为桥接上网。

学习时可设置为host-only方式，多个虚拟机只连接宿主机，和外网没有接触。

虚拟机的ip修改：

1. 进入虚拟机Centos🡪桌面最顶端右击联网图标（两个电脑）🡪编辑连接🡪选择System Eth0类似的ip接口🡪点编辑🡪选择Ipv4 Setting 🡪method设置为Manual,address处add添加上设置的ip地址，掩码和网关，网关需要是宿主机Vm的虚拟ip地址VMnet1的地址🡪点击Apply



1. 重启网络服务：#service network restart
2. Ifconfig 检查配置是否生效
3. 从CentOS上#ping 192.168.100.1

从宿主机cmd窗口：ping 192.168.100.11

都通了证明Ip配置好了

远程登录：

1. 检查CentOS上的ssh是否开启

ps –e | grep ssh

显示出以sshd结尾的一行就证明sshd服务已经启动，

否则用#service sshd start开启服务，#service sshd stop是关闭服务的。

1. 在宿主机打开PieTTY 输入CentOS的ip地址，点打开

输入root用户和密码，就会登录成功

1. 安装winscp软件，是在linux和windows系统之间传输文件的工具
2. 关闭防火墙

#service iptables stop 都报OK就是关闭成功

#service iptables status 检查下防火墙关闭状态

这种关闭只是暂时的，可以通过修改配置文件，在开机启动时不启动防火墙。

#vim /etc/selinux/config (默认你会使用vim编辑器)

将SELINUX=enforcing (开机启动) 修改为SELINUX=disabled (开机不启动)。

也可以通过命令关闭防火墙自动启动

#chkconfig --list| grep iptables

查看iptables的开启状态 0-6共7个

#chkconfig iptables off 关闭防火墙

1. 修改主机名：方便连接，不用输入ip了

#hostname 查看主机名

#hostname hadoop1 设置主机名为hadoop1

这样的设置只对当前操作生效，重启后会恢复为原来的主机名，需要修改配置文件才能一劳永逸

#vim /etc/sysconfig/network 文件下

HOSTNAME=hadoop1 保存退出，这样重启后就仍然是hadoop1了，不重启该文件修改是不生效的，可以直接使用hostname 在本次修改主机名

1. SSH 远程免登陆设置
   1. 生成RSA公钥

#ssh-keygen –t rsa

一直回车

* 1. 查看ssh公钥存在的位置

公钥在当前用户(root)主目录下的.ssh目录下

为id\_rsa.pub

#cp ~/.ssh/id\_rsa.pub ~/.ssh/authorized\_keys

就可以实现本机的免密码登录

复制到远程主机，在接下来用scp命令添加公钥

* 1. 绑定主机名和IP地址

#vim /etc/hosts 添加一行

192.168.100.11 hadoop1 就可以了

#ssh hadoop1 就可以免密码登录了，第一次会确认公钥，输入yes即可