Spark 编译与部署(上) --基础环境搭建

第1页共31页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan

1	运行环境说明	3
	1.1 硬软件环境	3
	1.2 集群网络环境	3
	1.3 安装使用工具	4
	1.3.1 Linux文件传输工具	4
	1.3.2 Linux命令行执行工具	4
2	搭建样板机环境	5
	2.1 安装操作系统	6
	2.2 设置系统环境	15
	2.2.1 设置机器名	15
	2.2.2 设置IP地址	15
	2.2.3 设置Host映射文件	17
	2.2.4 关闭防火墙	18
	2.2.5 关闭SElinux	19
	2.3 配置运行环境	19
	2.3.1 更新OpenSSL	19
	2.3.2 修改SSH配置文件	20
	2.3.3 增加hadoop组和用户	21
	2.3.4 JDK安装及配置	22
	2.3.5 Scala 安装及配置	24
3	配置集群环境	25
	3.1 复制样板机	25
	3.2 设置机器名和IP地址	26
	3.3 配置SSH无密码登录	26
	3.4 设置机器启动模式(可选)	29
4	问题解决	30
	4.1 安装CENTOS64 位虚拟机 THIS HOST SUPPORTS INTEL VT-X, BUT INTEL VT-X IS DISABLED	30
	4.2 *** IS NOT IN THE SUDOERS FILE解决方法	30

Spark 编译与部署(上)--基础环境搭建

1 运行环境说明

1.1 硬软件环境

- 主机操作系统: Windows 64 位, 双核 4 线程, 主频 2.2G, 10G 内存
- 虚拟软件: VMware® Workstation 9.0.0 build-812388
- 虚拟机操作系统: CentOS6.5 64 位, 单核, 1G 内存
- 虚拟机运行环境:
 - ▶ JDK: 1.7.0_55 64 位
 - ➢ Hadoop: 2.2.0(需要编译为 64 位)
 - > Scala : 2.10.4
 - ➢ Spark: 1.1.0(需要编译)

1.2 集群网络环境

集群包含三个节点,节点之间可以免密码 SSH 访问,节点 IP 地址和主机名分布如下:

序号	IP 地址	机器名	类型	核数/内存	用户名	目录
1	192.168.0.61	hadoop1	NN/DN/RM Master/Worker	1核/3G	hadoop	/app 程序所在路径
2	192.168.0.62	hadoop2	DN/NM/Worker	1核/2G	hadoop	/app/scala /app/hadoop
3	192.168.0.63	hadoop3	DN/NM/Worker	1核/2G	hadoop	/app/complied

- 所有节点均是 CentOS6.5 64bit 系统,防火墙/SElinux 均禁用,所有节点上均创建了一个 hadoop 用户,用户主目录是/home/hadoop,上传文件存放在/home/hadoop/upload 文件夹中。
- 2. 所有节点上均创建了一个目录/app 用于存放安装程序,并且拥有者是 hadoop 用户,对其必须有 rwx 权限(一般做法是 root 用户在根目录下创建/app 目录,并使用 chown 命令修改该目录拥有者为 hadoop), 否则 hadoop 用户使用 SSH 往其他机器分发文件会出现权限不足的提示

1.3 安装使用工具

1.3.1 Linux 文件传输工具

向 Linux 系统传输文件推荐使用 SSH Secure File Transfer,该工具顶部为工具的菜单和快捷方式,中间部分左面为本地文件目录,右边为远程文件目录,可以通过拖拽等方式实现文件的下载与上传,底部为操作情况监控区,如下图所示:



1.3.2 Linux 命令行执行工具

 SSH Secure Shell SSH Secure 工具的 SSH Secure Shell 提供了远程命令执行,如下图 所示:



● SecureCRT SecureCRT 是常用远程执行 Linux 命令行工具,如下图所示:

hadoop1 - SecureCRT	
文件(F) 编辑(E) 查看(V) 选项(O) 传输(T) 脚本(S) 工具(L) 帮助(H)	
🖏 🖏 🕞 🖏 🔊 🖻 🐁 🗚 🛛 🥃 🍠 🖙 🌋 📍 🞯 🖙 💂	
hadoop1 hadoop2 hadoop3	
<pre>[hadoop@hadoop1 ~]\$ ifconfig eth0 Link encap:Ethernet HWaddr 00:0C:29:03:0A:AC inet addr:192.168.0.61 Bcast:192.168.0.255 Mask:255.255.255.0 inet6 addr: fe80::20c:29ff:fe03:aac/64 Scope:Link UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1 RX packets:4968 errors:0 dropped:0 overruns:0 frame:0 TX packets:795 errors:0 dropped:0 overruns:0 carrier:0 collisions:0 txqueuelen:1000 RX bytes:6442720 (6.1 MiB) TX bytes:82691 (80.7 KiB)</pre>	
<pre>lo Link encap:Local Loopback inet addr:127.0.0.1 Mask:255.0.0.0 inet6 addr: ::1/128 Scope:Host UP LOOPBACK RUNNING MTU:16436 Metric:1 RX packets:80 errors:0 dropped:0 overruns:0 frame:0 TX packets:80 errors:0 dropped:0 overruns:0 carrier:0 collisions:0 txqueuelen:0 RX bytes:10394 (10.1 KiB) TX bytes:10394 (10.1 KiB) [hadoop@hadoop1 ~]\$</pre>	H
就绪 ssh2: AES-256-CTR 17, 21 21行, 87列 Linux	大写数字。

2 搭建样板机环境

本次安装集群分为三个节点,本节搭建样板机环境搭建,搭建分为安装操作系统、设置系统环境和配置运行环境三个步骤。

2.1 安装操作系统

第一步 插入 CentOS 6.5 的安装介质,使用介质启动电脑出现如下界面



- Install or upgrade an existing system 安装或升级现有的系统
- install system with basic video driver 安装过程中采用基本的显卡驱动
- Rescue installed system 进入系统修复模式
- Boot from local drive 退出安装从硬盘启动
- Memory test 内存检测

第二步 介质检测选择"Skip",直接跳过



第三步 出现引导界面,点击"next"



第四步 选择安装过程语言,选中"English (English)"

第7页共31页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan

What language would you like to use during the installation process?		
Bulgarian (Български)		^
Catalan (Català)		
Chinese(Simplified) (中文(简体))		
Chinese(Traditional) (中文(正體))		=
Croatian (Hrvatski)		
Czech (Čeština)		
Danish (Dansk)		
Dutch (Nederlands)		
English (English)		
Estonian (eesti keel)		
Finnish (suomi)		
French (Français)		
German (Deutsch)		
Greek (Ελληνικά)		
Gujarati (ગુજરાતી)		
Hebrew (עברית)		
Hindi (हिन्दी)		~
	Back	Next

第五步 键盘布局选择 "U.S.English"

Select the appropriate keyboard for the system.		•
Portuguese		^
Romanian		
Russian		
Serbian		
Serbian (latin)		
Slovak (qwerty)		
Slovenian		
Spanish		
Swedish		
Swiss French		
Swiss French (latin1)		
Swiss German		
Swiss German (latin1)		
Turkish		
U.S. English		=
U.S. International		
Ukrainian		
United Kingdom		~
	Back	Next 🔷

第六步 选择 "Basic Storage Devies"点击"Next"

第 8 页 共 31 页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan

lasic St	orage Devic	25				
nstalls or u his is prob	upgrades to typic ably it.	al types of storage device	s. If you're not sure	which option is right for	r you,	
speciali	zed Storage	Devices				
nstalls or u ou to add	upgrades to ente FCoE / iSCSI / zf	rprise devices such as Sto CP disks and to filter out o	rage Area Networks (levices the installer sl	SANs). This option will a hould ignore.	allow	
					Back	
					-	1.1

第七步 询问是否覆写所有数据,选择"Yes,discard any data"

The	Storage Device Warning
ne	storage device below may contain data.
	VMware, VMware Virtual S 20480.0 MB pci-0000:00:10.0-scsi-0:0:0:0
/e co	ould not detect partitions or filesystems on this device.
his o r vii ot b emo	could be because the device is blank , unpartitioned , r tual . If not, there may be data on the device that can e recovered if you use it in this installation. We can ve the device from this installation to protect the data.
re y	ou sure this device does not contain valuable data?
Z Ap	oply my choice to all devices with undetected partitions or filesystems
	Yes, discard any data No, keep any data

第八步 Hostname 填写格式 "英文名.姓"

第9页共31页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan

Please name this computer. The hostname identifies the computer on a network.		
Hostname: hadoop1		
Configure Network		
	a ck	Next 🔶

第九步 时区可以在地图上点击,选择"Shanghai"并取消 System clock uses UTC 选择

Please select the nearest city in your time zone:	
Selected city: Shanghai, Asia (east China - Beijing, Guangdong, Shanghai, etc.)	
Asia/Shanghai 🗘	
System clock uses UTC	
	Back Next

第十步 设置 root 的密码

第 10 页 共 31 页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan

The roo the syst user.	t account is used for administering em. Enter a password for the root		
Root Password:	•••••		
Confirm:	•••••		
		▶	
		Back Nex	٢t

第十一步 硬盘分区,一定要按照图示点选

Vhich type	e of installation would you like?
• •	Use All Space Removes all partitions on the selected device(s). This includes partitions created by other operating systems.
	Tip: This option will remove data from the selected device(s). Make sure you have backups.
0	Replace Existing Linux System(s) Removes only Linux partitions (created from a previous Linux installation). This does not remove other partitions you may have on your storage device(s) (such as VFAT or FAT32).
	Tip: This option will remove data from the selected device(s). Make sure you have backups.
0 🚺	Shrink Current System Shrinks existing partitions to create free space for the default layout.
0	Use Free Space Retains your current data and partitions and uses only the unpartitioned space on the selected device (s), assuming you have enough free space available.
° ?	Create Custom Layout Manually create your own custom layout on the selected device(s) using our partitioning tool.
Encryp	system
Review	and modify partitioning layout
	Rack
	Rea Back

第十二步 询问是否改写入到硬盘,选择"Write changes to disk"

第 11 页 共 31 页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan



第十三步 选择系统安装模式为"Desktop"

The default installation of CentOS is a minimum install. You can optionally select a different set of software now.		I
 Desktop Minimal Desktop Minimal Basic Server Database Server Web Server Virtual Host Software Development Workstation Please select any additional repositories that you want to use for software installation.		=
CentOS		
You can further customize the software selection now, or after install via the software management application. Customize later Customize now		
	e Back	Next

第十四步 桌面环境就设置完成了,点击安装

第 12 页 共 31 页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan



第十五步 安装完成,重启

Congratulations, your CentOS installation is complete. Please reboot to use the installed system. Note that updates may be available to ensure the proper functioning of your system and installation of these updates is recommended after the reboot.
Back Reboot

第十六步 重启之后,的 License Information

第 13 页 共 31 页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan

A Home × A CentOS 64-bit ×	
🚾 🗝 Eile Edit View VM Iabs Help 🔐 🕶 📇 🏷 🕤 🖓 🗉 🖃 💭 🔚	
To release input, press Ctrl+Alt.	
CentoS-6 EULA	
Kdump CentOS-6 comes with no guarantees or warranties of any sorts,	
either written or implied.	
The Distribution is released as GPLv2. Individual packages in the distribution come with their own licences. A copy of the GPLv2 license is included with the distribution media.	
Var Larree to the License Arreement	
Ine lide pet agree	
O No, i do not agree	
	Back Forward

第十七步 创建用户和设置密码(这里不进行设置用户和密码)

🗖 🛆 Homa 🗙 🕞	Contro Cd. hit	_ 0	52
W == File Edit View			
License			
Information • Create User Date and Time Kdump	You must create a 'username' for regular (non-administrative) use of your system. To create a system 'username', please provide the information requested below. Username: Full Name: Password: Confirm Password: If you need to use network authentication, such as Kerberos or NIS, please click the Use Network Login button. Use Network Login		
	If you need more control when creating the user (specifying home directory, and/or UID), please click the Advanced button.		
	Advanced		
		Back	orward

第十八步 "Date and Time" 选中 "Synchronize data and time over the network" Finsh 之后系统将重启

₩ → File Edit View License Information Create User > Date and Time Kdump	Centos sci-bit × Image: Secion sci-bit × Image: Secion sci-bit × Image: Secion sciebit × Image: Secion sci-bit × Image: Secion sciebit × Image: Secion sciebit × Image: Secion sciebit × Image: Sec							
	U synchronize date and time of your system:							
	Date Time							
	\langle July \rangle \langle 2014 \rangle Hour : 15 \bigcirc							
	Sun Mon Tue Wed Thu Fri Sat 29 30 1 2 3 4 5							
	6 7 8 9 10 11 12 Second : 20 ↓							
	13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 1 2 3 4 5 6 7 8 9							
	Ва	ack F	orward					

2.2 设置系统环境

该部分对服务器的配置需要在服务器本地进行配置,配置完毕后需要重启服务器确认配置是 否生效,特别是远程访问服务器需要设置固定 IP 地址。

2.2.1设置机器名

以 root 用户登录,使用#vi/etc/sysconfig/network 打开配置文件,根据实际情况设置该服务器的机器名,新机器名在重启后生效



2.2.2设置 IP 地址

1. 点击 System --> Preferences --> Network Connections, 如下图所示:



2. 修改或重建网络连接,设置该连接为手工方式,设置如下网络信息:

IP 地址:	192.168.0.61

子网掩码: 255.255.255.0

网关: 192.168.0.1

DNS: 221.12.1.227 (需要根据所在地设置 DNS 服务器)

【注意】

1、网关、DNS 等根据所在网络实际情况进行设置,并设置连接方式为"Available to all users",否则通过远程连接时会在服务器重启后无法连接服务器;

2、如果是运行在 VM Ware 虚拟机,网络使用桥接模式,设置能够连接到互联网中,以方便后面 Hadoop 和 Spark 编译等试验。

	Network C	opportions	🖻 Editing Wired connection 1 🛛 🗙
	Network C		Connection name: Wired connection 1
Computer	Name	Last Used 🔶 Add	✓ Connect automatically
	⊽ Wired	Edit	☑ Available to all users
root's Home	Wired connection 1	now Delete	Wired 802.1x Security IPv4 Settings IPv6 Settings
		=	Method: Manual
			Addresses
Trash			Address Netmask Gateway Add
			192.168.0.61 255.255.255.0 192.168.0.1 Delete
		Close	
Terminal			DNS servers: 221.12.1.227
			Search domains: 221.12.33.227
			DHCP client ID:
			Require IPv4 addressing for this connection to complete

第 16 页 共 31 页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan

 在命令行中,使用 ifconfig 命令查看设置 IP 地址信息,如果修改 IP 不生效,需要重启机器 再进行设置(如果该机器在设置后需要通过远程访问,建议重启机器,确认机器 IP 是否生效:

Σ					root@hadoop1:~	_ 0	×
File	Edit	View	Search	Terminal	l Help		
[root	@hado	oop1 ~]# ifco	nfig			(\land)
eth0		Link inet UP BR RX pa TX pa colli RX by	encap:E addr:192 addr: 0ADCAST ckets:59 ckets:60 sions:0 tes:1400	thernet 2.168.0.0 fe80::200 RUNNING 98 errors 19 errors txqueue 940 (136	HWaddr 00:0C:29:03:0A:AC 61 Bcast:192.168.0.255 Mask:255.255.255.0 c:29ff:fe03:aac/64 Scope:Link MULTICAST MTU:1500 Metric:1 s:0 dropped:0 overruns:0 frame:0 s:0 dropped:0 overruns:0 carrier:0 len:1000 .7 KiB) TX bytes:46587 (45.4 KiB)		
ιο		Link inet UP LO RX pa TX pa colli RX by	encap:Lo addr:12 OPBACK F ckets:50 ckets:50 sions:0 tes:3952	Dcal Loop 7.0.0.1 ::1/128 S RUNNING 6 errors 6 errors txqueue 2 (3.8 K	pback Mask:255.0.0.0 Scope:Host MTU:16436 Metric:1 :0 dropped:0 overruns:0 frame:0 :0 dropped:0 overruns:0 carrier:0 len:0 iB) TX bytes:3952 (3.8 KiB)		III

2.2.3 设置 Host 映射文件

1. 使用 root 身份编辑/etc/hosts 映射文件,设置 IP 地址与机器名的映射,设置信息如下:

#vi /etc/hosts

- 192.168.0.61 hadoop1
- 192.168.0.62 hadoop2
- 192.168.0.63 hadoop3

5	root@hadoop1:~	_ 0	×
File Edit View	Search Terminal Help		
127.0.0.1 loca ::1 loca 192.168.0.61 had 192.168.0.62 had 192.168.0.63 had	lhost localhost.localdomain localhost4 localhost4.locald lhost localhost.localdomain localhost6 localhost6.locald oop1 oop2 oop3	domain4 domain6	<

2. 使用如下命令对网络设置进行重启

#/etc/init.d/network restart

或者 #service network restart

第 17 页 共 31 页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan



3. 验证设置是否成功

E root@hadoop1:~	-	×
File Edit View Search Terminal Help		
[root@hadoop1 ~]# ping hadoop1		^
PING hadoop1 (192.168.0.61) 56(84) bytes of data.		
64 bytes from hadoop1 (192.168.0.61): icmp_seq=1 ttl=64 time=0.027 ms		
64 bytes from hadoop1 (192.168.0.61): icmp_seq=2 ttl=64 time=0.030 ms		
64 bytes from hadoop1 (192.168.0.61): icmp_seq=3 ttl=64 time=0.028 ms		
^C		
hadoop1 ping statistics		
3 packets transmitted, 3 received, 0% packet loss, time 2812ms		
rtt min/avg/max/mdev = 0.027/0.028/0.030/0.004 ms		

2.2.4 关闭防火墙

在 hadoop 安装过程中需要关闭防火墙和 SElinux, 否则会出现异常

Σ						root@had	oop1:~		_ 0	×
File	Edit	View	Search	Te	erminal	Help				
[root	@hado	op1 ha	adoop]#	se	ervice	iptables	status	5		^
Table	e: filt	ter								
Chair	INPU	Г (ро]	licy AC	CEF	PT)					
num	targe	t	prot o	pt	source	•		destination		
1	ACCEP	Г	all -	-	0.0.0.	0/0		0.0.0.0/0	state RELATED	,
ESTAE	BLISHE	D								
2	ACCEP	Г	icmp -	-	0.0.0.	0/0		0.0.0.0/0		
3	ACCEP	Г	all -	-	0.0.0.	0/0		0.0.0.0/0		
4	ACCEP	Г	tcp -	-	0.0.0.	0/0		0.0.0.0/0	state NEW tcp	
dpt:2	22									
5	REJEC	Г	all -	-	0.0.0.	0/0		0.0.0.0/0	reject-with i	C
mp-ho	ost-pro	ohibit	ted							
Chair	n FORW/	ARD (p	policy	ACC	CEPT)					
num	targe	t	prot o	pt	source	;		destination		Ξ
1	REJEC	Г	all -	-	0.0.0.	0/0		0.0.0.0/0	reject-with i	C
mp-ho	ost-pro	ohibit	ted							
Chair	1 OUTPI	UT (po	olicy A	CCE	EPT)					
num	targe	t	prot o	pt	source	•		destination		
	Å	第 18 〕	页共31	页	出自	石山园,博	客地址:	http://www.cnblogs.com	n/shishanyuan	

1. service iptables status 查看防火墙状态,如下所示表示 iptables 已经开启

2. 以 root 用户使用如下命令关闭 iptables

#chkconfig iptables off

[root@hadoop1 hadoop]# chkconfig iptables off
[root@hadoop1 hadoop]#

2.2.5 关闭 SElinux

1. 使用 getenforce 命令查看是否关闭

Σ	root@hadoop1:~	-	×
File	Edit View Search Terminal Help		
[root [root Enfor	@hadoop1 hadoop]# @hadoop1 hadoop]# getenforce cing		^
2	root@hadoop1:~	_	×
File	Edit View Search Terminal Help		

2. 修改/etc/selinux/config 文件

将 SELINUX=enforcing 改为 SELINUX=disabled,执行该命令后重启机器生效

#vi /etc/selinux/config

E root@hadoop1:~	-	×
File Edit View Search Terminal Help		
<pre># This file controls the state of SELinux on the system. # SELINUX= can take one of these three values: # enforcing - SELinux security policy is enforced. # permissive - SELinux prints warnings instead of enforcing. # disabled - No SELinux policy is loaded. #SELINUX=enforcing SELINUX=disable # SELINUXTYPE= can take one of these two values: # targeted - Targeted processes are protected, # mls - Multi Level Security protection. SELINUXTYPE=targeted</pre>		11 1

2.3 配置运行环境

2.3.1 更新 OpenSSL

CentOS 系统自带的 OpenSSL 存在 bug , 如果不更新 OpenSSL 在 Ambari 部署过程会出现无 法通过 SSH 连接节点,使用如下命令进行更新:

#yum update openssl

E root@hadoop1:~		_ 0	x
File Edit View Search Terminal Help			
[root@hadoop1 hadoop]# yum update openssl			^
Loaded plugins: fastestmirror, refresh-packagekit, securit	у		
<pre>* base: centos.ustc.edu.cn * extras: centos.ustc.edu.cn * updates: centos.ustc.edu.cn base base/primary_db extras extras </pre>	3.7 kB 4.4 MB 3.4 kB 19 kB	00:00 00:04 00:00 00:00	Ξ
updates	3.4 kB	00:00	
updates/primary_db Setting up Update Process Resolving Dependencies > Running transaction check > Package openssl.x86_64 0:1.0.1e-15.el6 will be update > Package openssl.x86_64 0:1.0.1e-16.el6_5.7 will be an > Finished Dependency Resolution	3.1 MB d update	00:02	
<pre>Package: centos-release-6-5.el6.centos.11.1.x86_64 (@anact 149.x86_64/6.5) From : /etc/pki/rpm-gpg/RPM-GPG-KEY-CentOS-6 Is this ok [y/N]: y Running rpm_check_debug Running Transaction Test Transaction Test Succeeded Running Transaction Updating : openssl-1.0.1e-16.el6_5.7.x86_64 Cleanup : openssl-1.0.1e-15.el6.x86_64 Verifying : openssl-1.0.1e-15.el6.x86_64</pre>	onda-CentOS	5-20131127 1/2 2/2 1/2 2/2	
Updated: openssl.x86_64 0:1.0.1e-16.el6_5.7			Ш

2.3.2 修改 SSH 配置文件

1. 以 root 用户使用如下命令打开 sshd_config 配置文件 *#vi /etc/ssh/ssd_config*

开放三个配置,如下图所示:

RSAAuthentication yes PubkeyAuthentication yes AuthorizedKeysFile .ssh/authorized_keys #MaxSessions 10
RSAAuthentication yes
PubkeyAuthentication yes
AuthorizedKeysFile .ssh/authorized_keys
#AuthorizedKeysCommand none
#AuthorizedKeysCommandRunAs nobody
For this to work you will also need host keys in /etc/ssh/ssh_known_hosts
#RhostsRSAAuthentication no
similar for protocol version 2
#HostbasedAuthentication no
Change to yes if you don't trust ~/.ssh/known_hosts for
RhostsRSAAuthentication and HostbasedAuthentication
#IgnoreUserKnownHosts no
To disable tunneled clear text passwords, change to no here!
#PasswordAuthentication yes
#PermitEmptyPasswords no
PasswordAuthentication yes

*

×

~

2. 配置后重启服务

#service sshd restart

[root@hadoop1 hadoop]# [root@hadoop1 hadoop]# service sshd restart Stopping sshd: Starting sshd: [root@hadoop1 hadoop]#

2.3.3 增加 hadoop 组和用户

使用如下命令增加 hadoop 组和 hadoop 用户 (密码), 创建 hadoop 组件存放目录

[OK]

#groupadd -g 1000 hadoop #useradd -u 2000 -g hadoop hadoop #mkdir -p /app/hadoop #chown -R hadoop:hadoop /app/hadoop #passwd hadoop

File Edit View Search Terminal Help
[root@hadoop1 ~]# useradd -u 2000 -g hadoop hadoop
[root@hadoop1 ~]# mkdir -p /app/hadoop
[root@hadoop1 ~]# ls /app
hadoop
[root@hadoop1 ~]# passwd hadoop
Changing password for user hadoop.
New password:
BAD PASSWORD: it is too simplistic/systematic
BAD PASSWORD: is too simple
Retype new password:
passwd: all authentication tokens updated successfully.

创建 hadoop 用户上传文件目录,设置该目录组和文件夹为 hadoop

#mkdir /home/hadoop/upload #chown -R hadoop:hadoop /home/hadoop/upload

2	root@hadoop1:~	-	×
File Edit View Search Termina	l Help		
<pre>[root@hadoop1 ~]# mkdir /home/ [root@hadoop1 ~]# chown -R had [root@hadoop1 ~]# ll /home/had total 4 drwxr-xr-x 2 hadoop hadoop 409 [root@hadoop1 ~]#</pre>	hadoop/upload oop:hadoop /home/hadoop/upload oop 6 Jan 14 09:39 <mark>upload</mark>		^

2.3.4 JDK 安装及配置

1. 下载 JDK1.7 64bit 安装包

打开 JDK1.7 64bit 安装包下载链接为:

<u>http://www.oracle.com/technetwork/java/javase/downloads/jdk7-downloads-1880</u> 260.html

打开界面之后,先选中 Accept License Agreement,然后下载 jdk-7u55-linux-x64.tar.gz,如下图所示:

🏉 Java SE Development Kit 7 - Dov	wnloads Oracle Technology Network Oracle -	Windows Inte	rnet Explorer	
O ⊂ O http://www.oracle	.com/technetwork/java/javase/downloads/jdk7	-downloads-1	880260.html 🛛 👻 😽 🗙 🖸 <i>360</i>	/渡索 👂 🔻
☆ 收藏夹 □ Java SE Developr	ment Kit 7 - Downloads Or		🟠 🕶 🗟 👻 🚍 🖶	▼ 页面(P)▼ 安全(S)▼ 工具(O)▼ 🕢▼
	Java SE Development Kit 7u55 You must accept the Oracle Binary Co	de License Agr software. line License Ag	reement for Java SE to download this greement	Get it now for FREE! Subscribe Today
	Product / File Description	File Size	Download	Introducing Java 8
	Linux x86	115.67 MB	tidk-7u55-linux-i586.rpm	
	Linux x86	133 MB	idk-7u55-linux-i586.tar.gz	🖉 📄 🚊 Java
	Linux x64	116.97 MB	idk-7u55-linux-x64.rpm	
	Linux x64	131.82 MB	idk-7u55-linux-x64.tar.oz	
	Mac OS X x64	179.56 MB	jdk-7u55-macosx-x64.dmg	
	Solaris x86 (SVR4 package)	138.86 MB	jdk-7u55-solaris-i586.tar.Z	
	Solaris x86	95.14 MB	jdk-7u55-solaris-i586.tar.gz	
	Solaris x64 (SVR4 package)	24.55 MB	idk-7u55-solaris-x64.tar.Z	
	Solaris x64	16.25 MB	idk-7u55-solaris-x64.tar.gz	Watch Now
	Solaris SPARC (SVR4 package)	138.23 MB	tidk-7u55-solaris-sparc.tar.Z	and watch now
	Solaris SPARC	98.18 MB	idk-7u55-solaris-sparc.tar.gz	
	Solaris SPARC 64-bit (SVR4 package)	24 MB		
	Solaris SPARC 64-bit	18.34 MB	idk-7u55-solaris-sparcy9.tar.oz	
	Windows x86	123.67 MB	idk-7u55-windows-i586.exe	
	Windows x64	125.49 MB	idk-7u55-windows-x64.exe	
	Java SE Development Kit 7u55 D Java SE Development Kit 7u55 Demos a	emos and S nd Samples Do BSD License	amples Downloads	
	· · · · · · · · · · · · · · · · · · ·	-	A Internet 保护模式:	: 启用

2. 赋予 hadoop 用户/usr/lib/java 目录可读写权限,使用命令如下:

\$sudo chmod -R 777 /usr/lib/java

[hadoop@hadoop1 ~]\$ [hadoop@hadoop1 ~]\$ sudo chmod -R 777 /usr/lib/java [sudo] password for hadoop: [hadoop@hadoop1 ~]\$ ■

第 22 页 共 31 页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan

该步骤有可能遇到问题 2.2, 可参考解决办法处理

3. 把下载的安装包,使用 1.1.3.1 介绍的 ssh 工具上传到/usr/lib/java 目录下,使用如下命令 进行解压

\$tar -zxvf jdk-7u55-linux-x64.tar.gz

```
[hadoop@hadoop1 java]$
[hadoop@hadoop1 java]$ ls
jdk-7u55-linux-x64.tar.gz
[hadoop@hadoop1 java]$ tar -zxvf jdk-7u55-linux-x64.tar.gz
```

解压后目录如下图所示:

```
[jan@hadoop ~]$
[jan@hadoop ~]$ cd /usr/lib/java
[jan@hadoop java]$ ll
total 134988
drwxr-xr-x. 8 uucp 143 4096 Mar 18 11:04 jdk1.7.0_55
-rwxrw-rw-. 1 jan jan 138220064 Apr 29 22:24 jdk-7u55-linux-x64.tar.gz
[jan@hadoop java]$
```

4. 使用 root 用户配置/etc/profile 文件,并生效该配置

export JAVA_HOME=/usr/lib/java/jdk1.7.0_55 export PATH=\$JAVA_HOME/bin:\$PATH export CLASSPATH=.:\$JAVA_HOME/lib/dt.jar:\$JAVA_HOME/lib/tools.jar

/etc/profile # System wide environment and startup programs, for login setup # Functions and aliases go in /etc/bashrc # It's NOT a good idea to change this file unless you know what you # are doing. It's much better to create a custom.sh shell script in # /etc/profile.d/ to make custom changes to your environment, as this # will prevent the need for merging in future updates. export JAVA_HOME=/usr/lib/java/jdk1.7.0_55 export PATH=\$JAVA_HOME/bin:\$PATH export CLASSPATH=.:\$JAVA_HOME/lib/dt.jar:\$JAVA_HOME/lib/tools.jar pathmunge () { case ":\${PATH}:" in *:"\$1":") if ["\$2" = "after"] ; then PATH=\$PATH:\$1 else PATH=\$PATH:\$1 else PATH=\$I:\$PATH fi esac

5. 重新登录并验证

\$logout

\$java -version

```
Last login: Tue Sep 23 08:37:39 2014 from 10.88.147.231

[hadoop@hadoop1 ~]$

[hadoop@hadoop1 ~]$ java -version

java version "1.7.0_55"

Java(TM) SE Runtime Environment (build 1.7.0_55-b13)

Java Hotspot(TM) 64-Bit Server VM (build 24.55-b03, mixed mode)

[hadoop@hadoop1 ~]$

[hadoop@hadoop1 ~]$
```

٨

2.3.5 Scala 安装及配置

1. 下载 Scala 安装包

Scala2.10.4 安装包下载链接为:<u>http://www.scala-lang.org/download/2.10.4.html</u>,因为在 Scala2.11.4 下IDEA有些异常,故在这里建议安装Scala2.10.4 版本

cala 2.10.4 The Scala 🗙 📃 👘	10 80 10 100 and	
C 🗋 www.scala-lang.org	/download/2.10.4.html	
Other resources You can find the installer downlo documentation and source code	oad links for other operating systems, as v e archives for Scala 2.10.4 below.	vell as
Archive	System	Size
scala-2.10.4.tgz	Mac OS X, Unix, Cygwin	28.55M
scala-2.10.4.msi	Windows (msi installer)	60.00M
scala-2.10.4.zip	Windows	28.60M
scala-2.10.4.deb	Debian	24.83M
scala-2.10.4 rpm	RPM package	24.83M

2. 上传 Scala 安装文件

把下载的 scala 安装包使用 SSH Secure File Transfer 工具(如 1.3.1 介绍)上传到 /home/hadoop/upload 目录下,如下图所示:

<u>File</u> Edit <u>V</u> iew <u>O</u> peration <u>W</u> i	indow <u>H</u> elp				
🖬 🎩 🍠 🛍 🛍 🖄 📁 🕫	🎭 J û 💼 🖷	1 <u>□</u> <u>-</u>	10 9/e 🖸 🥔 🦑		
🛛 🖉 Quick Connect 🗀 Profiles					
🛛 🔁 🖄 🛍 🌣 🔤 🛣 📉 🚺 5.на	adoop安装环境\Spark	र 💌 🛛 Add 🛛 🔁 🔤	🗿 🖻 🌣 📑 🗙 🛛	/home/hadoop/u	upload 💌 Add
Local Name	Size Type	Modified 🔶 Rem	ote Name	🛆 Size	Type Modified
sbt-0.13.7.tgz 1,0)58,904 360压缩	2015/01/1	cala-2.10.4.tgz	29,937,5	360压缩 2015/01/2
sbt-launch-0.13.5.jar 1,2	208.380 LAR × 14	2014/11/2(
📑 scala-2.10.4.tgz 29,9	37,534 360压缩	2015/01/1:			
📑 scala-2.11.4.tgz 26,5	09,669 360压缩	2015/01/1: =			
d shark-0.9.1.zip 4	58,339 360压缩	2015/02/0			
📑 shark-0.9.1-bin-hadoop 142,2	264,7 360压缩	2015/02/06			
shark-0.9.2.zip 4	60,525 360压缩	2015/02/0!			
shark-0.9.3.zip 4	61,356 360压缩	2015/02/0!			
📑 shark-assembly-0.9.2-h 142,5	585,3 360压缩	2015/02/0!			
spark-001-bin-badoon 1761	125.0 260 🖽 🕼	2015/02/01			
•					P
Transfer Queue					
🛆 Source File Source Dire	ectory Dest	ination Directory	Size Status		Speed Time
爺 scala-2.10.4.tgz F:\9.tools\	5.Hadoop安 /hor	me/hadoop/upload	29,937,534 Compl	ete 11492.	.3 kB/s 00:00:02

3. 解压缩

到上传目录下,用如下命令解压缩:

\$cd /home/hadoop/upload

\$tar -zxf scala-2.10.4.tgz

```
[hadoop@hadoop1 ~]$ cd /home/hadoop/upload
[hadoop@hadoop1 upload]$ tar -zxf scala-2.10.4.tgz
[hadoop@hadoop1 upload]$ ls
scala-2.10.4 scala-2.10.4.tgz
[hadoop@hadoop1 upload]$ ■
```

迁移到/app 目录下:

\$sudo mv scala-2.10.4 /app/

4. 使用 root 用户配置/etc/profile 文件,并生效该配置

export SCALA_HOME=/app/scala-2.10.4

export PATH=\$PATH:\${SCALA_HOME}/bin

export JAVA_HOME=/usr/lib/java/jdk1.7.0_55 export HADOOP_HOME=/app/hadoop/hadoop-2.4.1 export PATH=\$PATH:\$JAVA_HOME/bin:\$HADOOP_HOME/bin:\$HADOOP_HOME/sbin export CLASSPATH=.:\$JAVA_HOME/lib/dt.jar:\$JAVA_HOME/lib/tools.jar Export SCALA_HOME=/app/scala-2.10.4 export PATH=\$PATH:\${SCALA_HOME}/bin

5. 重新登录并验证

\$exit

\$scala -version

```
[hadoop@hadoop1 ~]$ scala -version
scala code runner version 2.10.4 -- Copyright 2002-2013, LAMP/EPFL
[hadoop@hadoop1 ~]$ scala
Welcome to Scala version 2.10.4 (Java HotSpot(TM) 64-Bit Server VM, Java 1.7.0_55).
Type in expressions to have them evaluated.
Type :help for more information.
scala> ■
```

3 配置集群环境

复制样板机生成其他两个节点,按照规划设置及其命名和 IP 地址,最后设置 SSH 无密码登录。

3.1 复制样板机

复制样板机两份,分别为 hadoop2 和 hadoop3 节点

第 25 页 共 31 页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan



3.2 设置机器名和 IP 地址

以 root 用户登录,使用 vi /etc/sysconfig/network 打开配置文件,根据 1.2 规划修改机器名,修改机器名后需要重新启动机器,新机器名在重启后生效



	S Networ	k Connections	×	Editing Wired connection 1 ×
			~	Connection name: Wired connection 1
Computer	Name	Last Used	Add	Connect automatically
_	⊽ Wired		Edit	✓ Available to all users
	Wired connection 1	4 minutes ago		Wind 202 by Socurity IPV4 Settings IDv6 Sottings
root's Home	when connection 2	12 minutes ago	Delete	When boz.ix security in the security in the security in the security
			Ξ	Method: Manual
				Addresses
Trash				Address Netmask Gateway Add
		[\checkmark	192.168.0.62 255.255.255.0 192.168.0.1 Delete
			Close	
<u>}</u>				
Terminal				DNS servers: 221.12.1.227
				Search domains: 221.12.33.227
				DHCP client ID:
				Require IPv4 addressing for this connection to complete
				Deutee
				Routes

3.3 配置 SSH 无密码登录

1. 使用 hadoop 用户登录在三个节点中使用如下命令生成私钥和公钥;

\$ssh-keygen -t rsa

第 26 页 共 31 页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan



 进入 /home/hadoop/.ssh 目录在三个节点中分别把公钥命名为 authorized_keys_hadoop1、authorized_keys_hadoop2和authorized_keys_hadoop3, 使用命令如下:

\$cd /home/hadoop/.ssh

\$cp id_rsa.pub authorized_keys_hadoop1

[hadoop@hadoop1 .ssh]\$ [hadoop@hadoop1 .ssh]\$ cp id_rsa.pub authorized_keys_hadoop1 [hadoop@hadoop1 .ssh]\$ ls authorized_keys_hadoop1 [hadoop@hadoop1 .ssh]\$ id_rsa id_rsa.pub

 把两个从节点(hadoop2、hadoop3)的公钥使用 scp 命令传送到 hadoop1 节点的 /home/hadoop/.ssh 文件夹中;

\$scp authorized_keys_hadoop2 hadoop@hadoop1:/home/hadoop/.ssh
\$scp authorized_keys_hadoop3 hadoop@hadoop1:/home/hadoop/.ssh

[hadoop@hadoop2 .ssh]\$
[hadoop@hadoop2 .ssh]\$ scp authorized_keys_hadoop2 hadoop@hadoop1:/home/hadoop/.ssh]
hadoop@hadoop1's password:
authorized_keys_hadoop2 100% 396 0.4KB/s 00:00
[hadoop@hadoop2 .ssh]\$
[hadoop@hadoop1 .ssh]\$ ls
authorized_keys_hadoop1 authorized_keys_hadoop2 id_rsa

authorized_keys_hadoop1 authorized_keys_hadoop2 id_rsa [hadoop@hadoop1 .ssh]\$ ls authorized_keys_hadoop1 authorized_keys_hadoop2 authorized_keys_hadoop3 id_rsa

4. 把三个节点的公钥信息保存到 authorized_key 文件中

使用\$cat authorized_keys_hadoop1 >> authorized_keys 命令

<u>.</u>

把该文件分发到其他两个从节点上 5.

使用 \$scp authorized_keys hadoop@hadoop2:/home/hadoop/.ssh 把密码文件分发出去

	•	
[hadoop@hadoop1 .ssh]\$		*
[[hadoop@hadoop1 .ssh]\$ scp_autnor1zed_keys_hadoop	@hadoop2:/home/hadoop/.ssn	
The authenticity of host hadoop2 (10.88.147.222)	can't be established.	
RSA key fingerprint is 02:3a:d5:e7:2c:1/:60:ce:tb	:40:6f:ce:a4:15:93:12.	
Are you sure you want to continue connecting (yes	/no)? yes	
Warning: Permanently added Thadoop2,10.88.147.222	(RSA) to the list of known hosts.	
hadoop@hadoop2's password:		
author1zed_keys	100% 1188 1.2KB/S 00:00	
[nadoop@nadoop1 .ssn]\$	ale de se De lle se lle de se lle se l	
[nadoop@nadoop1 .ssn]} scp authorized_keys nadoop	@nadoop3:/nome/nadoop/.ssn	
The authenticity of nost nadoops (10.88.147.223)	can't be established.	
RSA Key fingerprint is 02:3a:us:e/:20:1/:00:ce:10	:40:67:Ce:a4:15:93:12.	
Warming, Dormanontly added 'badoon? 10,98,147,222	(REA) to the list of known bests	
hadoon@hadoon?'s password.	(RSA) to the fist of known hosts.	
authorized keys	100% 1188 1 2KB/c 00.00	
[hadoon@hadoon1_ssh]	100% 1100 1.2KB/3 00.00	
[hadoopenadoop1 .33h]3		
[hadaan@hadaan]ab]f	·	_
[[nadoop@nadoop2 .ssn]\$		-
[[nadoop@nadoop2].ssn]\$ is	lungua hasta	
Tauthorized_keys_authorized_keys_nadoopz_id_rsa	known_nosts	
[nadoop@nadoop2 .ssn]\$		

[hadoop@hadoop2 .ssh]\$

在三台机器中使用如下设置 authorized_keys 读写权限 6.

\$chmod 400 authorized_keys

[hadoop@hadoop1 .ssh]\$ [hadoop@hadoop1 .ssh]\$ chmod 400 authorized_keys [hadoop@hadoop1 .ssh]\$ ■

测试 ssh 免密码登录是否生效 7.

[hadoop@hadoop1 .ssh]\$
[hadoop@hadoop1 .ssh]\$ ssh hadoop1
Last login: Fri May 23 22:32:55 2014 from 10.88.147.181
[hadoop@hadoop1 ~]\$ exit
logout
Connection to hadoop1 closed.
[hadoop@hadoop2 ~]\$ exit
logout
Connection to hadoop2 closed.
[hadoop@hadoop1 .ssh]\$ ssh hadoop3
Last login: Fri May 23 22:32:09 2014 from hadoop3.hwdomain
[hadoop@hadoop3 ~]\$ exit
logout
Connection to hadoop3 closed.
[hadoop@hadoop3 ~]\$ exit
logout
Connection to hadoop3 closed.
[hadoop@hadoop1 .ssh]\$ ssh hadoop3
Last login: Fri May 23 22:32:09 2014 from hadoop3.hwdomain
[hadoop@hadoop1 .ssh]\$ exit
logout

3.4 设置机器启动模式(可选)

设置好集群环境后,可以让集群运行在命令行模式下,减少集群所耗费的资源。以 root 用户使用#vi /etc/inittab,将 id:5:initdefault: 改为 id:3:initdefault:

hadoop1 hadoop2 hadoop3 23 # ٠ Default runlevel. The runlevels used are: 0 - halt (Do NOT set initdefault to this) # # # - Single user mode 1 2 - Multiuser, without NFS (The same as 3, if you do not have networking) 3 - Full multiuser mode # # # 4 – unused # 5 _ X11 - reboot (Do NOT set initdefault to this) # id:3:initdefault: INSERT

Linux 系统任何时候都运行在一个指定的运行级上,并且不同的运行级的程序和服务都不同,所要完成的工作和所要达到的目的都不同。CentOS 设置了如下表所示的运行级,并且系统可以在这些运行级别之间进行切换,以完成不同的工作。运行级说明

- 0 所有进程将被终止,机器将有序的停止,关机时系统处于这个运行级别
- 1 单用户模式。用于系统维护,只有少数进程运行,同时所有服务也不启动
- 2多用户模式。和运行级别 3 一样,只是网络文件系统(NFS)服务没被启动
- 3多用户模式。允许多用户登录系统,是系统默认的启动级别
- 4 留给用户自定义的运行级别
- 5多用户模式,并且在系统启动后运行 X-Window,给出一个图形化的登录窗口
- 6所有进程被终止,系统重新启动

4 问题解决

4.1 安装 CentOS64 位虚拟机 This host supports Intel VT-x, but Intel VT-x is disabled

在进行 Hadoop2.X 64bit 编译安装中由于使用到 64 位虚拟机,安装过程中出现下图错误:



按F1 键进入BIOS 设置实用程序 使用箭头键 security 面板下找 virtualization 按 Enter 键 进去 Intel Virtualization Technology 改成 Enabled 按 F10 键保存并退出 选择 Yes 按 Enter 键 完全关机(关闭电源)等待几秒钟重新启动计算机此 Intel 虚拟化技术开启成功

4.2 *** is not in the sudoers file 解决方法

当使用 hadoop 用户需要对文件夹进行赋权,使用 chmod 命令出现 "hadoop is not in the sudoers file. This incident will be reported"错误,如下所示:

🔚 10.88.147.221 - SecureCRT	
文件(F) 编辑(E) 查看(V) 选项(O) 传输(T) 脚本(S) 工具(L) 帮助(H)	
🖏 況 💭 🖏 🍋 🛍 🖓 🦻 😹 🎒 🌁 🌋 📍 🞯 🔤 🖕	
10.88.147.221	×
[hadoop@hadoop1 ~]\$ [hadoop@hadoop1 ~]\$ sudo chmod -R 777 /usr/lib/java [sudo] password for hadoop: hadoop is not in the sudoers file. This incident will be reported. [hadoop@hadoop1 ~]\$ ■	•

1. 使用 su 命令进入 root 用户

第 30 页 共 31 页 出自石山园,博客地址: http://www.cnblogs.com/shishanyuan



2. 添加文件的写权限,操作命令为: chmod u+w /etc/sudoers

10.88.147.221 - SecureCRT	
文件(F) 编辑(E) 查看(V) 选项(O) 传输(T) 脚本(S) 工具(L) 帮助(H)	
编33 🖓 🖓 🖓 🕒 🗈 8 🖌 😼 5 🔿 1 🚰 💥 📍 1 🞯 1 🔤 🖕	
10.88.147.221	×
[root@hadoop1 hadoop]# [root@hadoop1 hadoop]# chmod u+w /etc/sudoers [root@hadoop1 hadoop]# vi /etc/sudoers	*

3. 编辑/etc/sudoers 文件 ,使用 命令"vi /etc/sudoers"进入编辑模式 ,找到 :"root ALL=(ALL) ALL"在起下面添加"hadoop ALL=(ALL) ALL" ,然后保存退出。

