

CDH 6.3.2 部署

作者: k8s

- 原文链接: https://ld246.com/article/1608271790792
- 来源网站: 链滴
- 许可协议:署名-相同方式共享 4.0 国际 (CC BY-SA 4.0)

Cloudera Manager安装

系统环境准备,安装基础环境

1. 选择四台已经安装 CentOS7 Linux系统的节点,分配资源。

安装CDH节点推荐内存为64G,大部分内存被Cloudera Management Service 占用,因为做了大量数据分析和整合。这里,划分四台节点如下:

IP地址	节点角色
192.168.0.50	Server,Agent
192.168.0.51	Agent
192.168.0.52	Agent
192.168.0.53	Agent
	IP地址 192.168.0.50 192.168.0.51 192.168.0.52 192.168.0.53

服务架构如下:



● Cloudera Manager Server是Cloudera Manager 的核心是, Server 管理控制台服务和托管应程序逻辑, 负责软件的安装、配置、服务的启动与关闭及管理集群。

- Agent安装在每台主机上,负责进程的启动和停止,解压配置,触发安装及监控主机。
- Management Service是由一组角色组成的服务,这些角色执行各种监视,警报和报告功能。
- DataBase用来存储配置及监控信息。
- ClouderaRepository是Cloudera Manager 分发软件的存储库。

● Clients分为AdminConsole (管理员web界面版) 和API (用于开发者创建Cloudera Manager 序) 两种形式。

2. 初始化节点

- 更改主机名分别为 cm-s1、cm-a1、cm-a2、cm-a3
- 配置 SSH 秘钥验证,达到任一节点无需密码验证登录其他站点 root 用户的效果

● 配置/etc/hosts 文件,使任一节点通过 cm-s1、cm-a1、cm-a2、cm-a3 即可域名解析到对应 点 IP

● 关闭SELinux和firewalld.service

3. 配置时间同步服务

在每个节点上安装chrony服务,并配置:

[root@cm-s1 ~]# for i in `seg 0 3`;do ssh root@192.168.0.5\$i yum -y install chrony;done [root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5\$i wget https://sto.17ker.top/sourc s/chrony.conf -O /etc/chrony.conf;done

[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5\$i systemctl start chronyd;done

4. 安装JDK

给每台节点安装jdk,这里我们安装的CDH版本为6.3.2,在官网描述中这里需要使用jdk8,且对小版 有要求,

Oracle JDK 8 versions that are tested and recommended

Oracle JDK Version	Notes
1.8u181	Recommended / Latest version tested
1.8u162	Recommended
1.8u141	Recommended
1.8u131	Recommended
1.8u121	Recommended
1.8u111	Recommended
1.8u102	Recommended
1.8u91	Recommended
1.8u74	Recommended
1.8u31	Minimum required

这里我们选择jdk8版本中的1.8u181版本安装。

[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5\$i rpm -ivh https://sto.17ker.top/s urces/jdk-8u181-linux-x64.rpm;done

在每台节点配置jdk的环境变量:

[root@cm-s1 ~]# tail -n5 /etc/profile #Java Environment variables export JAVA HOME=/usr/java/default export CLASSPATH=::\$CLASSPATH:\$JAVA HOME/lib export PATH=\$PATH:\$JAVA HOME/bin [root@cm-s1 ~]# source /etc/profile

5. 具体如下:

6. 安装MySQL数据库

在cm-s1节点安装mysql5.7

● 添加 mysql5.7 仓库

```
[root@cm-s1 ~]# rpm -ivh https://dev.mysql.com/get/mysql57-community-release-el7-11.no
rch.rpm
[root@cm_s1 ~]# yum makacacha fast
```

[root@cm-s1 ~]# yum makecache fast

● 安装MySQL

[root@cm-s1 ~]# yum -y install mysql-community-server

● 启动

[root@cm-s1 ~]# systemctl start mysqld

- 安全访问设置
 - 查看默认 root 密码:

[root@cm-s1 ~]# cat /var/log/mysqld.log | grep -i 'temporary password'

• 更改 root 密码,移除匿名用户:

[root@cm-s1 ~]# mysql_secure_installation

● 移除 root 用户远程登录限制:

[root@cm-s1 ~]# mysql -hlocalhost -pAz123456_ -e "UPDATE mysql.user SET host='192.168.0 %' WHERE user='root';FLUSH PRIVILEGES;"

7. 安装第三方依赖包

在每台节点上安装ClouderManager需要的第三方依赖包,每台节点执行如下命令:

yum install -y chkconfig bind-utils psmisc cyrus-sasl-plain cyrus-sasl-gssapi portmap /lib/lsb/ nit-functions httpd mod_ssl openssl-devel python python-psycopg2 MySQL-python libxslt zli sqlite fuse fuse-libs redhat-lsb

至此,安装Cloudera Manager的基础环境准备完成。

Cloudera Manager 组件安装

1. 添加yum源

\$ sudo wget https://archive.cloudera.com/cm6/6.3.1/redhat7/yum/cloudera-manager.repo -P
/etc/yum.repos.d/
\$ sudo rpm --import https://archive.cloudera.com/cm6/6.3.1/redhat7/yum/RPM-GPG-KEY-cl
udera

\$ sudo yum makecache fast

2. 在cm-s1节点安装 CM Server

[root@cm-s1 ~]# yum install -y cloudera-manager-daemons cloudera-manager-server cloude a-manager-agent

3. 在cm-a1、cm-a2、cm-a3节点安装daemons、agent RPM包

[root@cm-a1 ~]# yum install -y cloudera-manager-daemons cloudera-manager-agent 当完成安装agent后,会在当前节点生成"/opt/cloudera/cm-agent"目录。

4. 配置CM Agent的server host

在所有的节点上,修改Agent的配置文件 /etc/cloudera-scm-agent/config.ini,修改 server_host=m-s1,指定所有Agent节点的Server节点为cm-s1。

[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5\$i sed -i 's/server_host=localhost/s rver_host=cm-s1/' /etc/cloudera-scm-agent/config.ini;done [root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5\$i grep server_host /etc/cloudera-s m-agent/config.ini;done

5. 配置CM Server 数据库

在cm-s1节点上,修改 /etc/cloudera-scm-server/db.properties文件,配置内容如下:

com.cloudera.cmf.db.type=mysql com.cloudera.cmf.db.host=cm-s1 com.cloudera.cmf.db.name=cmf com.cloudera.cmf.db.user=cmf com.cloudera.cmf.db.password=Az123456_ com.cloudera.cmf.db.setupType=EXTERNAL

上传 mysql-connector-java-5.1.48.jar文件到cm-s1节点的/usr/share/java目录下,如果没有目录 要创建,并且软链接到 mysql-connector-java.jar

[root@cm-s1 ~]# mkdir -p /usr/share/java [root@cm-s1 ~]# wget https://sto.17ker.top/sources/mysql-connector-java-5.1.48.jar -P /usr/ hare/java [root@cm-s1 ~]# cd /usr/share/java/ [root@cm-s1 java]# ln -s mysgl-connector-java-5.1.48.jar mysgl-connector-java.jar

为MySQL服务器创建cmf用户

[root@cm-s1 ~]# mysql -uroot -hcm-s1 -pAz123456_ -e "grant all on *.* to 'cmf'@'%' identifi d by 'Az123456_' with grant option;"

在cm-s1节点上执行如下命令初始化数据库:

[root@cm-s1 ~]# cd /opt/cloudera/cm/schema/ # -u用户, -p密码,数据库类型,数据库,用户,密码 [root@cm-s1 schema]# ./scm_prepare_database.sh -ucmf -pAz123456_ mysql cmf cmf Az123 56_

All done, your SCM database is configured correctly!

6. 准备 CDH Parcels 本地源

在cm-s1节点上创建 /opt/cloudera/parcel-repo目录,并下载如下文件:

[root@cm-s1 ~]# mkdir -p /opt/cloudera/parcel-repo [root@cm-s1 ~]# wget -P /opt/cloudera/parcel-repo https://archive.cloudera.com/cdh6/6.3.2 parcels/CDH-6.3.2-1.cdh6.3.2.p0.1605554-el7.parcel [root@cm-s1 ~]# wget -O /opt/cloudera/parcel-repo/CDH-6.3.2-1.cdh6.3.2.p0.1605554-el7.p rcel.sha https://archive.cloudera.com/cdh6/6.3.2/parcels/CDH-6.3.2-1.cdh6.3.2.p0.1605554-el7 parcel.sha1

[root@cm-s1 ~]# wget -P /opt/cloudera/parcel-repo https://archive.cloudera.com/cdh6/6.3.2 parcels/manifest.json

7. 启动 CM Server、Agent

在Server节点启动ClouderaManager Server, 分别在Agent节点启动ClouderaManager Agent:

[root@cm-s1 ~]# systemctl start cloudera-scm-server.service [root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5\$i systemctl start cloudera-scm-ag nt.service;done

Server启动比较慢,需要等待一会才能访问对应的web界面,可以查看启动日志,Sever启动的日志 认在路径*/var/log/cloudera-scm-server/cloudera-scm-server.log下。Agent启动的日志默认在路径 var/log/cloudera-scm-agent/cloudera-scm-agent.log*下。可以通过以上日志来检查启动中是否 错。

Sever首次启动会自动创建表以及数据,不要立即关闭或重启,否则需要删除所有表及数据重新安装。

至此, Cloudera Manager的安装完成。

CDH安装

1. 登录 Cloudera Manager WebUI界面

Cloudera Manager的WebUI界面登录地址为 http://cm-s1:7180, 默认的用户名和密码都是 admin 。输入完成后,点击登录:



2. 选择安装 CDH 的节点 登录之后,一直选择"继续"即可





选择Cloudera Express,点击"继续",弹出页面后继续点击"继续":

欢迎使用 Cloudera Mana	ger					
		Select Edition 升级别 Cloudera Exterprise 将提供可以和助型在关注	建任务环境下管理和监控 Hadoop 群集)	1.重要功能。		
	Accept License		Cloudera Express	Cloudera Enterprise Cloudera Enterprise 试用版	Cloudera Enterprise	
	Select Edition	件可证	88	40 天 在试用能之后,这个品质能说作为 Glouders Express 运行,如此都来和面描评合保 用不业影响。	年度17日 上世界可述 13日本年点文介	
		节点限制	无用制	EBN	无限制	
		СОН	~	~	~	
		Cloudera Manager 核心功能	~	~	~	
		Cloudera Manager 真级功能		~	~	
		Cloudera Navigator		~	~	
		Cloudera Navigator Key Trustee	1		*	
		Cloudera 变持			~	
		道参照 Clouders Express 和 Clouders Enterprise 中	કારણાં અને છે.		2	



P##95 Paral Cutter Basic Distor Basic <

弹出如下页面,该页面是为CDH指定主机,可以使用"模式"通配来选择主机,也可以选择"当前管的主机",这里"当前管理的主机"中有节点内容的原因正是由于之前我们在这三台节点配置过agen。如果未来集群增加机器,可以在新主机中搜索添加,后期会自动将agent安装到选中的新节点。 这里我们选择"当前管理的主机"中的所有节点,点击"继续"

Cloudera Manager							▲ 応 元持 ・ ▲ admin ・
群集安装							
	Custer Reset General Parcels Install Parcels Install Parcels Install Parcels	Specify Hosts In Couters Supress,) If the Couter	ma raw install CDH 6.8 or higher on up to 100 転(4) 1 市場市場面目上で一部活用にした構成で が登録: 1922,168.0.52 1922,168.0.53 1922,168.0.53 1922,168.0.53	Norm, 85,81 - Jahawi Jahawi Jahawi Adelawi	CDH 服本 1 元 元 元 元	7981 : 4 4 3 3	
						返回 地线	

3. 集群安装

选择 CDH 版本为 CDH 6.3.2, 点击"继续",如下图:

群集安装	
P Curier States P P P P </th <th></th>	

点击"继续"之后,Cloudera Manager主节点会将下载好的CDH分发到各个Agent节点,这时可以 三个Agent节点上看到路径/opt/cloudera/parcel-cache目录中有Cloudera Manager主节点传过来 CDH安装包,同时在完成传输之后,每个Agent节点还会将CDH安装包解压到/opt/cloudera/parcel 路径下,此时,界面显示如下:

群集安装	
• 520 Install Parcels 0200 0200 Parcel 26 T 1019 92/05 0480/98 2.01.2 • Cluster Bayers • CON1.5 2.0.1.005.3.2.0.1005.5.2 1010.100.100.100.100.000.000.000.000.00	
Providy Heats Providy Heats Provide Heats Provide Heats Provide Heats Provide Heats Provide Heats Provide Heats	
December 200 - Anno 1900 -	

点击"继续",显示正在检查主机健康状况,稍等片刻,会显示检查结果:

Cloudera Manager				▲ 22 8.0 -	A scheme +
群集安装					
	Trill Custor Energy Custor Energy Custor Energy Custor Energy Custor Energy Marcel Energy Marcel Energy	Inspect Cluster Image: Cluster	essaturant, Claudias recontinents Not Bay an performed 网络环境检查结果全部正常 主机环境检查结果有损量信息。		

点击检查结果,查看检查主机出现的问题:

0	没有存在已知错误的内核版本在运行。
0	Cloudera 建议将 /proc/sys/vm/swappiness 设置为最大值 10。当前设置为 30。使用 sysctl 命令在运行时更改该设置并编辑 /etc/sysctl.conf, 以在重启后保存该设置。您可以 继续进行安装,但 Cloudera Manager 可能会报告您的主机由于交换而运行状况不良。以下主机将受到影响: ◆ 查看详细信息 cm.a[1-3]
0	已启用透明大页面压缩,可能会导致重大性能问题。请运行"echo never > /sys/kernel/mm/transparent_hugepage/defrag"和"echo never > /sys/kernel/mm/transparent_hugepage/enabled"以禁用此设置,然后将同一命令添加到 /etc/rc.local 等初始化脚本中,以便在系统重启时予以设置。以下主机将受到影响: ◆ 查看详细信息 cm.a[1-3]
•	已满足 CDH 5 Hue Python 版本依赖关系。

这里建议swappiness=10的时候表示最大限度使用物理内存。由于透明超大页面已知会导致意外的 点重新启动,可以在每台节点中执行如下命令,在检查主机正确性后面点击"重新运行"即可解决。 令如下:

[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5\$i sed -i '\\$a\vm.swappiness\ =\ 10 /etc/sysctl.conf;done

[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5\$i sysctl -p;done # 在每个节点执行如下命令,并写入到/etc/rc.local中,注意软链接源文件/etc/rc.d/rc.local文件的 执行性

echo never > /sys/kernel/mm/transparent_hugepage/defrag

echo never > /sys/kernel/mm/transparent_hugepage/enabled

安装	
Custer Basics Custer	
Counter Bases Counter Bases Counter Bases Counter Bases Inspect Network Performance	
S Inspect Network Performance	
Specify Houts Specify Houts HID ● 上式设计 几秒第一种很巧呢 20.45% 组织检查数据第一代 展开设计 其5-	
C Inservance Inspect Hosts	
No instants were directed, review the inspector results to see what directs same performed.	

最后,点击"继续",进入"集群设置"

4. 集群设置

在进入的集群设置界面中,点击"自定义服务",选择"HDFS"、"YARN"、"Zookeeper"进 安装,注意:这里提示还会安装Cloudera Management Service,用于后期的监控、报告、事件和 报。

	服务类型	说明
	HBase	Apache HBase 提供对大型数据集的随机、实时的读/写访问权限(需要 HDFS 和 ZooKeeper)。
	HDFS	Apache Hadoop 分布式文件系统 (HDFS) 是 Hadoop 应用程序使用的主要存储系统,HDFS 创建多个数据块副本并将它们分布在整个群集的计算主机上,以应用可算 且极其快速的计算功能。
	🗆 😵 Hive	Hive 是一种数据仓库系统,提供名为 HiveQL I的 SQL 类语言。
	C the Hue	Hue 是与包括 Apache Hadoop 的 Cloudera Distribution 配合使用的图形用户界面(需要 HDFS、MapReduce 和 Hive)。
	🗆 🦞 Impala	Impala 为存储在 HDFS 和 HBase 中的数据提供了一个实时 SQL 重询接口。Impala 需要 Hive 服务,并与 Hue 共享 Hive Metastore。
	🗇 🐠 tsilon	EMC failon is a distributed filesystem.
	🗆 🏂 Kafka	Apache Kafka is publish-subscribe messaging rethought as a distributed commit log.
	C 🔅 Key-Value Store Indexer	键/值 Store Indexer 侦听 HBase 中所含衷内的数度变化,并使用 Solr 为其创建索引。
	🗆 🍕 Kudu	Kudu is a true column store for the Hadoop ecosystem.
	🖸 👩 Oozie	Oozie 是群集中管理数据处理作业的工作流协调服务。
	🗆 👩 Solr	Solr 是一个分布式服务,用于编制存储在 HDFS 中的数据的索引并搜索这些数据。
	🗆 💠 Spark	Apache Spark is an open source cluster computing system. This service runs Spark as an application on VARN.
	YARN (MR2 Included)	Apache Hadoop MapReduce 2.0 (MRv2) 或 YARN 是支持 MapReduce 应用程序的数据计算框架(需要 HDFS)。
	ZooKeeper	Apache ZooKeeper 是用于维护和同步配置数据的集中服务。

以上点击"继续",进入"角色分配"页面,显示的是Cloudera Manager自动划分的集群分配,当这里也可以自己分配节点。如下图所示:

Select Services	自定义角色分配				
	非可在点些自定义新群集的角色分配。但 建立改变分配错误。	如果分配不正确(例如、分配到某个主机上的角)	(大多)公影納服务性能。諸母后有特殊能式。	第已为特定角色很先造厚特定主机。否则 Cloudera 干	
- HEAMENE	进可以放主机首都角色分配。 医工机机				
O REAL	O HDFS				
	Diversel	SecondaryNumeNode + 1 #58	Dalancer + 1 IEB	C Hespf's	
				100 PR. 2. To	
● 非接近	C NFS Gateway B IR ± 41	Constanto de + 3 Mill			
🔿 womenn	Clouders Management Service				
	Gervice Monitor + 1 Bill	Activity Monitor	Host Monitor + 1 IEI	Event Server + 7 Britt	
0.00		BH-715	CHAI	100 S-1	
0.14	Akert Publisher × 1 #12				
	E YANN (MKZ Included)				
	ResourceManager + 1 858	II JobHatory Server + 1 Bill	U NodeManager × 3 新聞		
	1 ZooKeeper				
	A Server + 1 Mill				
	(Sec. 1)				

继续点击"继续",进入"审核更改页面",在这里可以手动修改一些配置文件,取默认配置。 继续点击"继续",进入执行集群命令,启动各个组件页面,当所有组件启动完成之后,如下图示:

Cloudera Manager						m /2 ≥35 × 4×0
群集设置						
	Select Services	首次送行 命令 ※② ○日洗成 首 12月11,24005下年 ③ 2.2m Finished First Run of the following services successfully: 2ookeeper, HDF3, YAB ◇ 日用瓜 11分類 (月17) ● Them AT Bage ○ Show Only Failed Shops ○ Show Only Running Shops	N (MR2 Included), Clouders Management	Service		
	() ₩48.82	Gun a set of services for the feat time 日式切りは、4、今日頃、		12月17,240.057年	2.3m	
		✓ ⑤ 気法法行き参援 已成功物成 4 个参照。		12月 17, 2:42:05 下平	2.2m	
		Ensuring that the expected software releases are installed on hosts.		12月 17.2.4005 下平	5.00e	
	0 58	> G 并行运动于 4 步程		12月 17, 2,40:10 下平	15.84	
		> 6 #1153357 2 #48		12月 17.2.4026 下平	14.9x	
		> 6 并行派行之参保		12月 17,24041 下午	76.06#	
		> G BUT VARN (MR2 included)	Of VARW (MRD Included) Of	12月 17,243 57 7年	5.89a	
		Verfying successful startup of services 比較比較高額 4 今世間。		12月17,242027年	34.054	
		 Werkying successful startup of service 2colikerper 	Interpret of	12月 17, 242:03 下平	137ms	
		 Werkying successful startup of service HDFS. 	HERE OF	12月 17, 2,42:03 平平	14410	
		 Verifying successful startup of service VABN (MB2 included) 	VARW (MR2 Included) Of	12月17.242.02下午	14.854	
		 Verifying successful startup of service Couliers Management Service 	Clouders Management Genote Of	12月 17, 2,42:03 下平	145ma	
					15 001	

点击"继续",完成集群设置。

Cloudera Manager				n /2 3/1-	A admin -
群集设置					
311/20 00 AB	briect Services all X-R R 118 write RX write RX	Cē ♥ MACRO. REHeba+aff.			
			道殿 完成		

5. CDH主页面

loudera Manager 818 - ±4. + 98 -	BH 西市· 安理-			11	m /2 RB+ Astron+
EQ 🗱 HAGHMAR E 🔂 -	所有继续命令				間 Switch to Table Vew O 活化 -
o mycluster	图表			30 11 10	t 2 cat a cat 12 cat 1.0. Ja 104 🖍
CDH 0.3.2 (Pantel)	III CPU	BRES O	ETIR-RHIE IO	H0/510	
0 = 4 TH	1005	Turnel	1	a state	
6 6 HD/2 -	I m		5.7563		
O III YAAN (MAIZ L	1	E SAMA	\$ 1.000s	E ADDA	
O d Zookeeper A1 -	02.45	0.0	02.48	2.4	
	-mpchatel #7 215 Hitt 216 CPU (2018 1.3%	-0402101. 44404 -04021018.0. #	-0-RM101 . 17465 -0-RM1011 . 5165	- 6 DataVision 2.864 - 8 Sanahorina 1844	
Cloudera Management Service					
C El Carden Ma					