Greenplum集群环境搭建

[(146条消息) greenplum安装部署-CentOS7.9下Greenplum6.19集群搭建详细过程-CSDN博客](https://blog.csdn.net/msh6453/article/details/124892605?ops_request_misc=%257B%2522request%255Fid%2522%253A%2522167998554416800227477064%2522%252C%2522scm%2522%253A%252220140713.130102334.pc%255Fblog.%2522%257D&request_id=167998554416800227477064&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2~blog~first_rank_ecpm_v1~rank_v31_ecpm-2-124892605-null-null.blog_rank_default&utm_term=greenplum-db-6.19.3-rhel-x86_64.rpm%E4%B8%8B%E8%BD%BD&spm=1018.2226.3001.4450)

https://blog.csdn.net/msh6453/article/details/124892605?ops\_request\_misc=%257B%2522request%255Fid%2522%253A%2522167998554416800227477064%2522%252C%2522scm%2522%253A%252220140713.130102334.pc%255Fblog.%2522%257D&request\_id=167998554416800227477064&biz\_id=0&utm\_medium=distribute.pc\_search\_result.none-task-blog-2~blog~first\_rank\_ecpm\_v1~rank\_v31\_ecpm-2-124892605-null-null.blog\_rank\_default&utm\_term=greenplum-db-6.19.3-rhel-x86\_64.rpm%E4%B8%8B%E8%BD%BD&spm=1018.2226.3001.4450

GP Greenplum

架构-

Share-nothing 节点之间不会共享数据

Master宕机，standyMaster active 需要心跳检测机制，配置虚拟ip分配

Client一般只能与Master节点进行交互，client将sql发送给Master，然后Master对sql进行分析，将其分配给segment进行操作，再将结果汇总返回给客户端

Greenplum集群部署：

采集数据为非实时采集，采集周期在半小时以上

采集的数据为结构化数据，没有图像或者视频类型数据

安装环境准备

集群内部联通需要 关闭防火墙Firewall及安全加强SELinux（所有节点）

[root@localhost /]# systemctl stop firewalld //停止firewalld防火墙

[root@localhost /]# systemctl disable firewalld //disable防火墙，使其开机不自启

[root@localhost /]# systemctl status firewalld //查看firewalld是否已经关闭,running：激活状态 dead：未激活状态

[root@localhost /]# vi /etc/sysconfig/selinux

SELINUX=disabled //修改成disabled

时钟设置 内部时间一致

修改内核配置参数（所有节点）

修改/etc/sysctl.conf文件，其中的配置文件是官方配置的

cat <<EOF > /etc/sysctl.conf

# kernel.shmall = \_PHYS\_PAGES / 2

kernel.shmall = 742216

# kernel.shmmax = kernel.shmall \* PAGE\_SIZE

kernel.shmmax = 3040116736

kernel.shmmni = 4096

vm.overcommit\_memory = 2

vm.overcommit\_ratio = 95

net.ipv4.ip\_local\_port\_range = 10000 65535

kernel.sem = 500 2048000 200 4096

kernel.sysrq = 1

kernel.core\_uses\_pid = 1

kernel.msgmnb = 65536

kernel.msgmax = 65536

kernel.msgmni = 2048

net.ipv4.tcp\_syncookies = 1

net.ipv4.conf.default.accept\_source\_route = 0

net.ipv4.tcp\_max\_syn\_backlog = 4096

net.ipv4.conf.all.arp\_filter = 1

net.core.netdev\_max\_backlog = 10000

net.core.rmem\_max = 2097152

net.core.wmem\_max = 2097152

vm.swappiness = 10

vm.zone\_reclaim\_mode = 0

vm.dirty\_expire\_centisecs = 500

vm.dirty\_writeback\_centisecs = 100

vm.dirty\_background\_ratio = 0 # See System Memory

vm.dirty\_ratio = 0

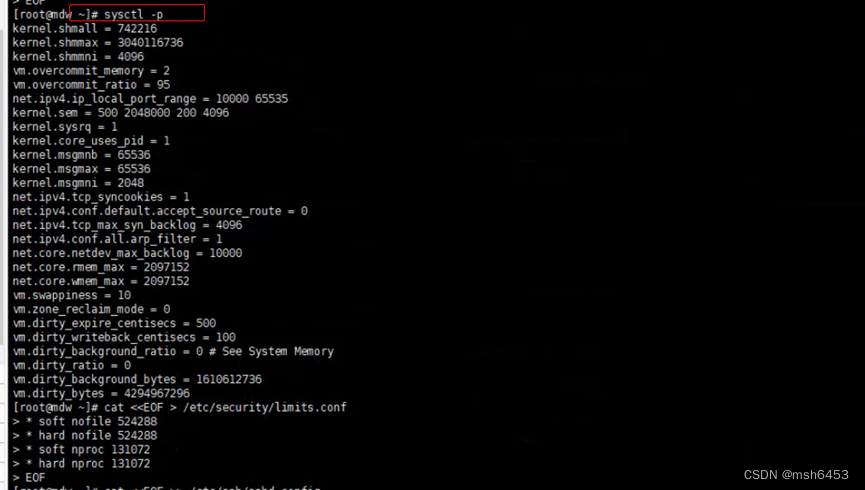
vm.dirty\_background\_bytes = 1610612736

vm.dirty\_bytes = 4294967296

EOF

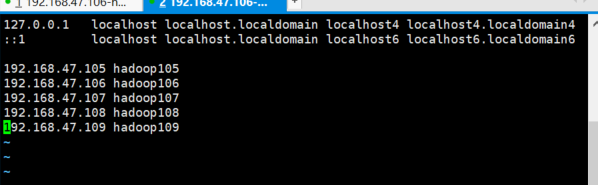
执行完上述操作后，执行sysctl -p 使配置生效

截图如下



配置/etc/hosts（所有节点）

Vi /etc/hosts



配置etc/sysconfig/network（所有节点）

每个节点的设置内容不同



注意其他节点也需要修改

配置资源限制参数etc/security/limits.conf(所有节点)

操作如下

cat <<EOF > /etc/security/limits.conf

\* soft nofile 524288

\* hard nofile 524288

\* soft nproc 131072

\* hard nproc 131072

EOF



修改ssh的连接数(所有节点)

cat <<EOF >> /etc/ssh/sshd\_config

MaxStartups 200

MaxSessions 200

EOF



将所有服务器重启（所有节点)

执行reboot

添加组和用户 gpadmin用户及用户组 （所有节点）

如果已经存在，先删除

groupdel gpadmin

userdel gpadmin

创建新的用户和用户组

groupadd gpadmin

useradd -r -m -g gpadmin gpadmin

passwd gpadmin （设置密码） dlutpgadmin

集群相互之间ssh免密（所有节点机器 gpadmin 用户操作）

具体操作：（操作为master节点）

su gpadmin （进入gpadmin用户）

mkdir -p /home/gpadmin/conf（这个部分提前创建，后续会用到）

cd /home/gpadmin/

ssh-keygen -t rsa （生成密钥对）

ssh-copy-id Hadoop106

ssh-copy-id Hadoop107

ssh-copy-id Hadoop108

注：每个节点都需要这样操作。

ssh权限互通设置（Master节点）

在/home/gpadmin/conf的路径下生成 hostlist，seg\_hosts文件：

具体操作：

cd /home/gpadmin/conf

vi hostlist

#文件内容

hadoop106

hadoop107

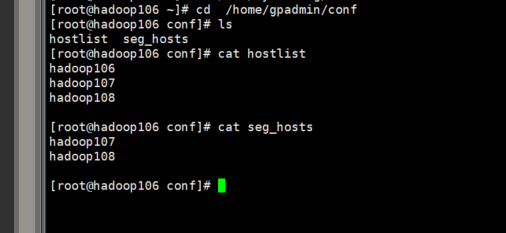
hadoop108

vi seg\_hosts

#文件内容

hadoop107

hadoop108



Greenplum集群安装（所有节点）

首先是将greenplum的安装包导入到服务器上（可以新建一个文件夹）；、

2、具体操作：

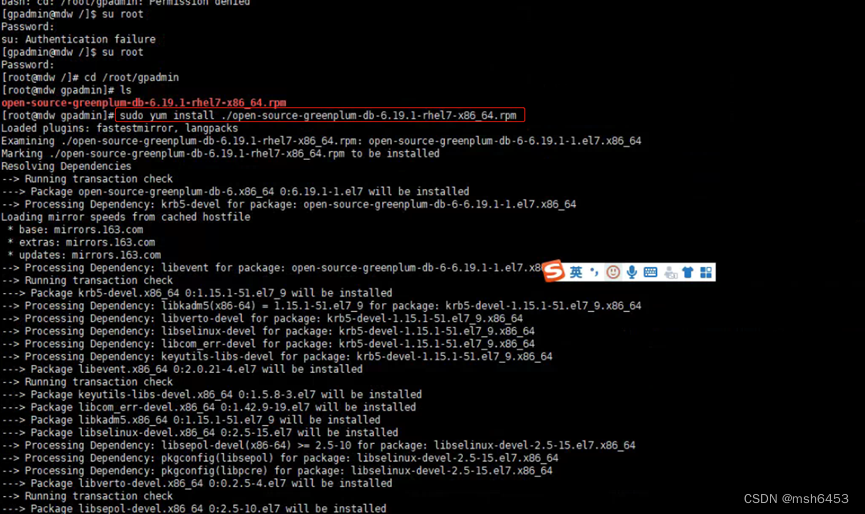
su root （进入root用户）

cd /root/gpadmin （这个安装包的所在位置）

ls （查看安装包，确认位置）

sudo yum install ./open-source-greenplum-db-6.19.1-rhel7-x86\_64.rpm （这个是官方的安装步骤，是不需要再去下载依赖的）

chown -R gpadmin:gpadmin /usr/local/greenplum\* （安装完之后，这个地方是将这个文件在gpadmin用户赋予权限）



中途需要输入一次y

注：所有节点都需要安装

使用 gpssh-exkeys 打通所有服务器 (Master节点)

1、使用 gpssh-exkeys 将所有机器的通道打开，这样在每台机器之间跳转，就不需要输入密码。

具体操作：

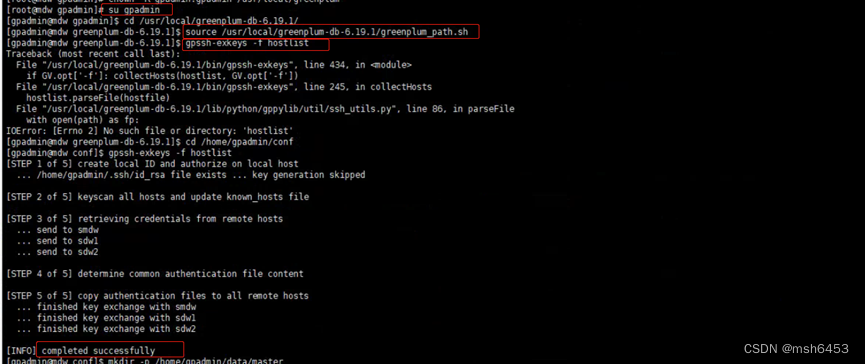
su gpadmin

cd /usr/local/greenplum-db-6.19.1/

source /usr/local/greenplum-db-6.19.1/greenplum\_path.sh

cd /home/gpadmin/conf

gpssh-exkeys -f hostlist



环境变量配置.bashrc和GPHOME（所有节点机器 gpadmin 用户操作）

1、配置环境变量.bashrc

具体操作：

su gpadmin

mkdir -p /home/gpadmin/data/master （这个部分可以提前创建）

cat <<EOF >> /home/gpadmin/.bashrc

source /usr/local/greenplum-db/greenplum\_path.sh

export PGPORT=5432

export PGUSER=gpadmin

export MASTER\_DATA\_DIRECTORY=/home/gpadmin/data/master/gpseg-1（注意：这个路径是我本人放置数据的文件夹；可以自行考虑，根据服务器内存调整位置）

export PGDATABASE=gp\_sydb

export LD\_PRELOAD=/lib64/libz.so.1 ps

EOF

source /home/gpadmin/.bashrc



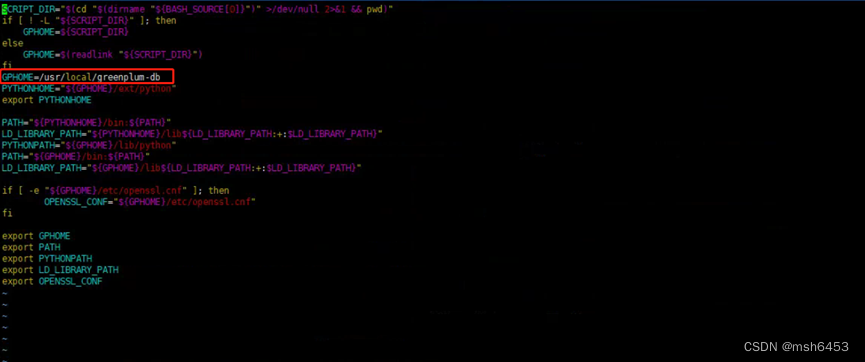
2、配置环境变量GPHOME，首先进到文件中直接修改

具体操作：

vi /usr/local/greenplum-db/greenplum\_path.sh

添加的路径：GPHOME=/usr/local/greenplum-db

截图如下：



创建数据文件夹（Master节点）

1、创建各个节点的数据文件夹，该步骤之后 hostlist 文件包含机器下都会创建 data目录，data目录下都会创建master、primary、mirror文件夹。

具体操作：

gpssh -f /home/gpadmin/conf/hostlist

mkdir data

cd data

mkdir master

mkdir primary

mkdir mirror

exit

截图如下：



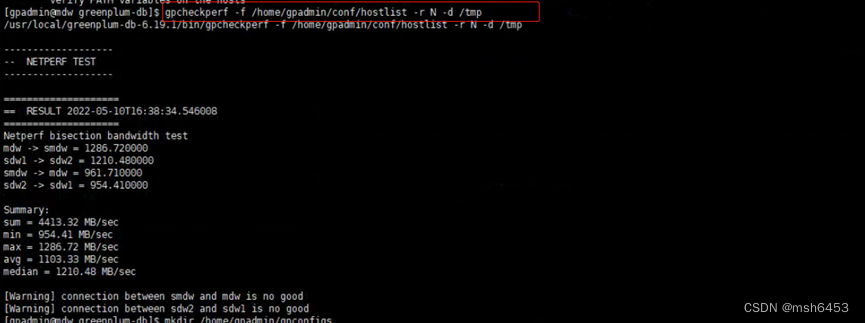
连通性检查（主节点 gpadmin 用户操作 ）

1、检查各个节点的连通性

具体操作：

gpcheckperf -f /home/gpadmin/conf/hostlist -r N -d /tmp

截图如下：



3.9 初始化配置文件编辑修改（主节点 gpadmin 用户操作 ）

1、这里修改初始化文件，首先拷贝一个文件gpinitsystem\_config，在修改。

具体操作：

mkdir /home/gpadmin/gpconfigs

cp /usr/local/greenplum/greenplum-db/docs/cli\_help/gpconfigs/gpinitsystem\_config /home/gpadmin/gpconfigs/gpinitsystem\_config （这个部分是一个命令，比较长）

vim /home/gpadmin/gpconfigs/gpinitsystem\_config

这里先新建在修改：

# ------------------------以下为配置内容------------------------

# 该项配置设置主节点数据存储位置，括号里边有几个代表每台主机创建几个postgresql数据库实例，即segment的实例数，上边示例是2个。

declare -a DATA\_DIRECTORY=(/home/gpadmin/data/primary /home/gpadmin/data/primary)

# 该项配置设置主节点机器名字

MASTER\_HOSTNAME=mdw

# 该项配置设置主节点数据存储位置

MASTER\_DIRECTORY=/home/gpadmin/data/master

# 该项配置设置是备节点数据存储位置,规则同DATA\_DIRECTORY，括号里边数量和DATA\_DIRECTORY保持一致。

# greenplum数据分主节点和备节点，主节点挂掉时候备节点数据会启用。

declare -a MIRROR\_DATA\_DIRECTORY=(/home/gpadmin/data/primary /home/gpadmin/data/primary)

# 该项配置设置默认数据库名字，和环境变量数据库名字保持一致，不然会失败。

DATABASE\_NAME=gp\_sydb



2、在/home/gpadmin/gpconfigs新增一个配置文件hostfile\_gpinitsystem

具体操作：

cd /home/gpadmin/gpconfigs

vi hostfile\_gpinitsystem

#新增的内容

sdw1

sdw2

截图如下：

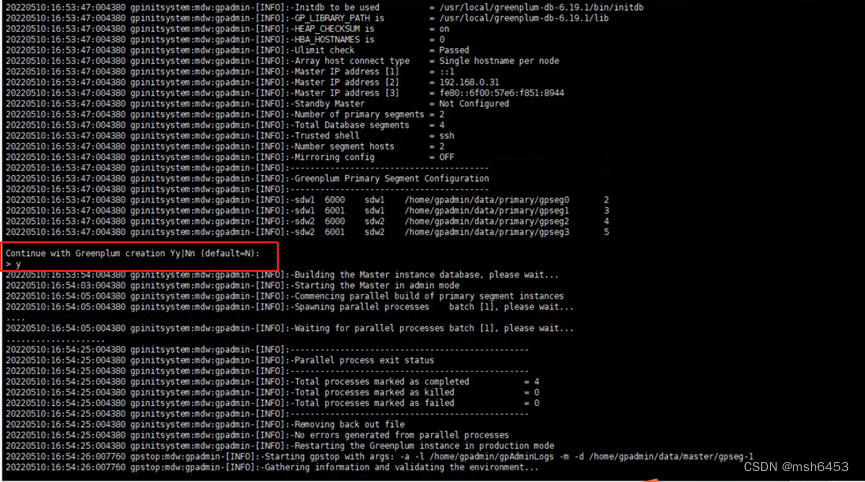


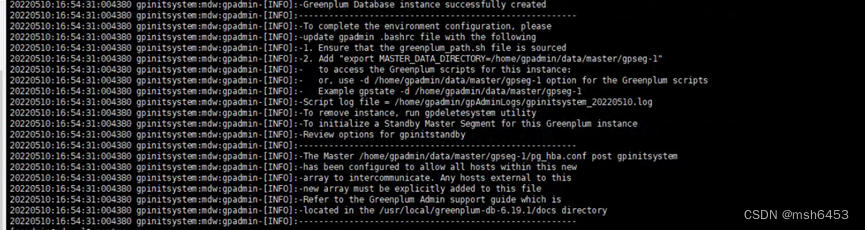
3.10 初始化数据库（主节点 gpadmin 用户操作 ）

1、初始化数据库

具体操作：

gpinitsystem -c gpconfigs/gpinitsystem\_config -h gpconfigs/hostfile\_gpinitsystem





上述应该就是安装好了。

3.11设置远程登录（主节点 gpadmin 用户操作 ）

1、如需要Navicat能够链接，需要配置如下：

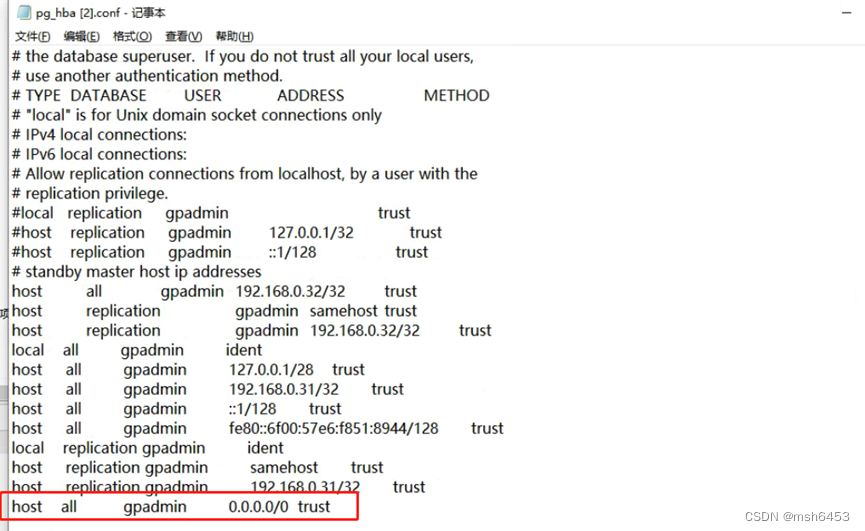
具体操作：

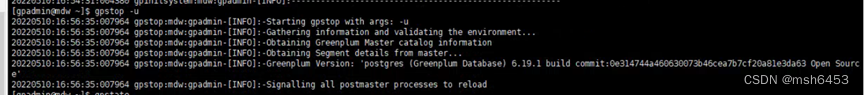
echo "host all gpadmin 0.0.0.0/0 trust" >> /home/gpadmin/data/master/gpseg-1/pg\_hba.conf

gpstop -u （重新加载数据库配置）

截图如下：

结果截图





检测postgres服务启动了

ps -ef | grep postgres

Greenplum数据库验证

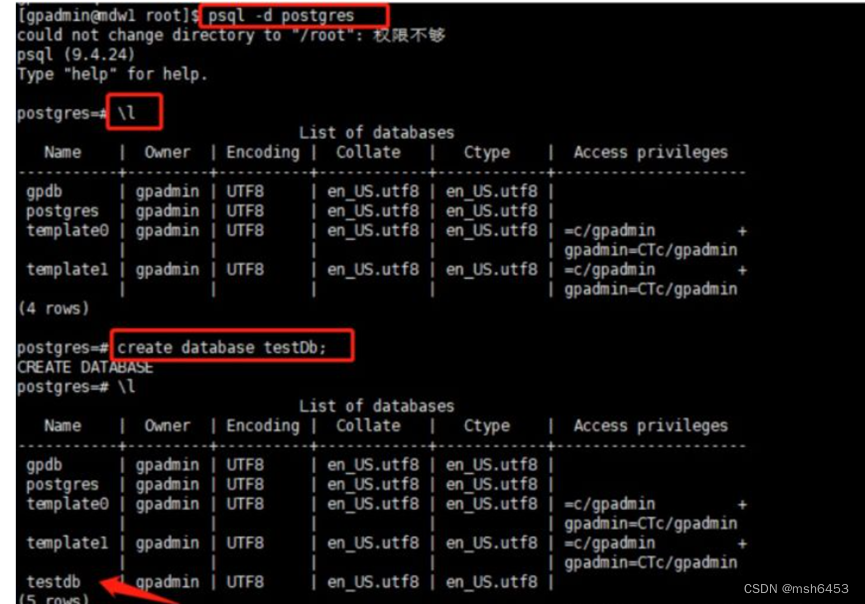
4.1登录验证（主节点 gpadmin 用户操作 ）

1、登录验证

具体操作：

psql -d postgres

截图如下：



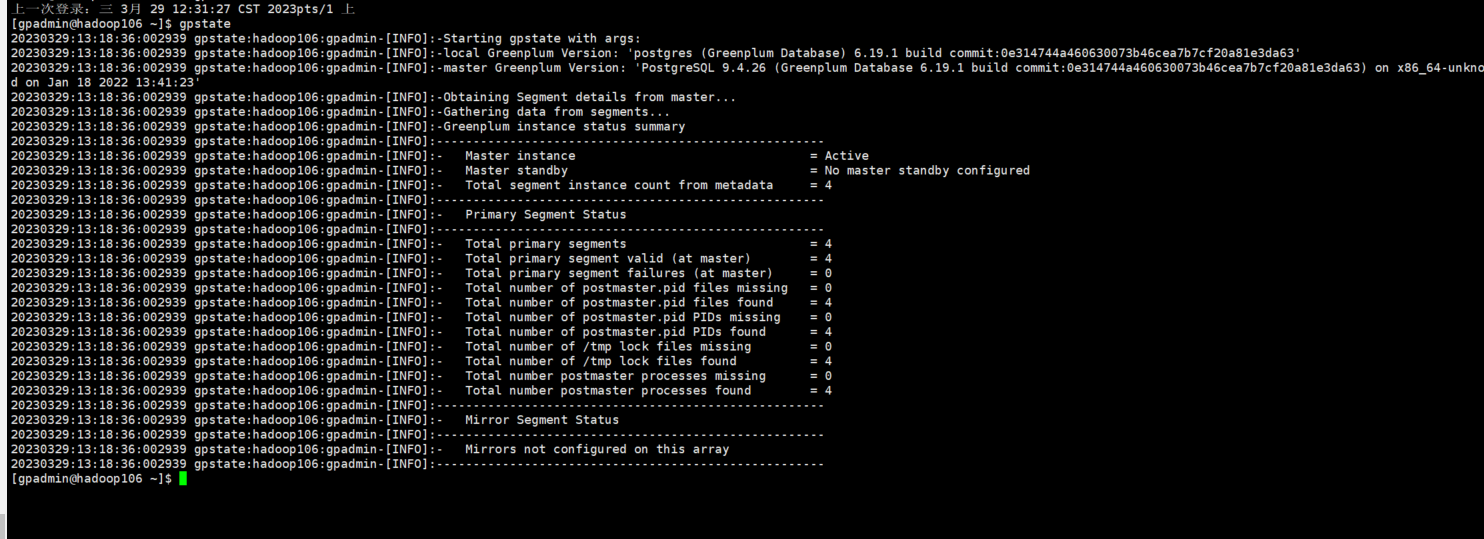
navicat登录验证

查看集群状态验证，在主节点上gpadmin用户操作命令

具体操作：

gpstate

截图如下



其他greenplum命令（主节点 gpadmin 用户操作 ）

1、命令如下

关闭gpstop 或者gpstop -a

启动gpstart或者gpstart-a （-a的话不用输入yes）

状态查看gpstate

help命令查看所有命令

psql -d postgres 进入psql命令行执行help

\h 查看SQL命令

? 查看psql命令

服务器关闭后 启动greenplum集群方式

主节点切换到gpadmin用户

使用 gpstart启动greenplum集群

启动后登录数据库 psql -d postgres

-----------pg--------------

su - postgres

psql -U postgres

\l 查看有哪些库

\c wn\_test 使用wn\_test库

\d 查看有哪些表

\d+ wn\_test\_stu 查看表详细信息

\d temp\_test\_wn

\di+ 查看索引

\timing 计时

\q 退出

说明：

1.greenplum中节点分布设置

Vim /home/gpadmin/gpconfigs/gpinitsystem\_config

2.数据文件存放位置

/home/gpadmin/data

Segment配置目录

/home/gpadmin/conf

3.不增加集群节点，每台机器扩容segment实例个数

参考网站

[Greenplum扩容详解 - 一寸HUI - 博客园 (cnblogs.com)](https://www.cnblogs.com/zsql/p/14602563.html)

<https://www.cnblogs.com/zsql/p/14602563.html>

操作记录

现在的状态是有3台主机，一个是master节点，还有两个segment的机器，每个segment的机器上都有一个primary和mirror的segment，现在计划在现有的集群上进行segment的扩容，在每台机器上的segment的数量翻倍

现在segment的目录为：

/home/gpadmin/data/primary

/home/gpadmin/data/mirror

/home/gpadmin/data/primary

/home/gpadmin/data2/primary

/home/gpadmin/data2/mirror

上面的目录的所属组和用户均为gpadmin:gpamdin，这里创建目录可以使用gpssh创建也可以一个一个的创建

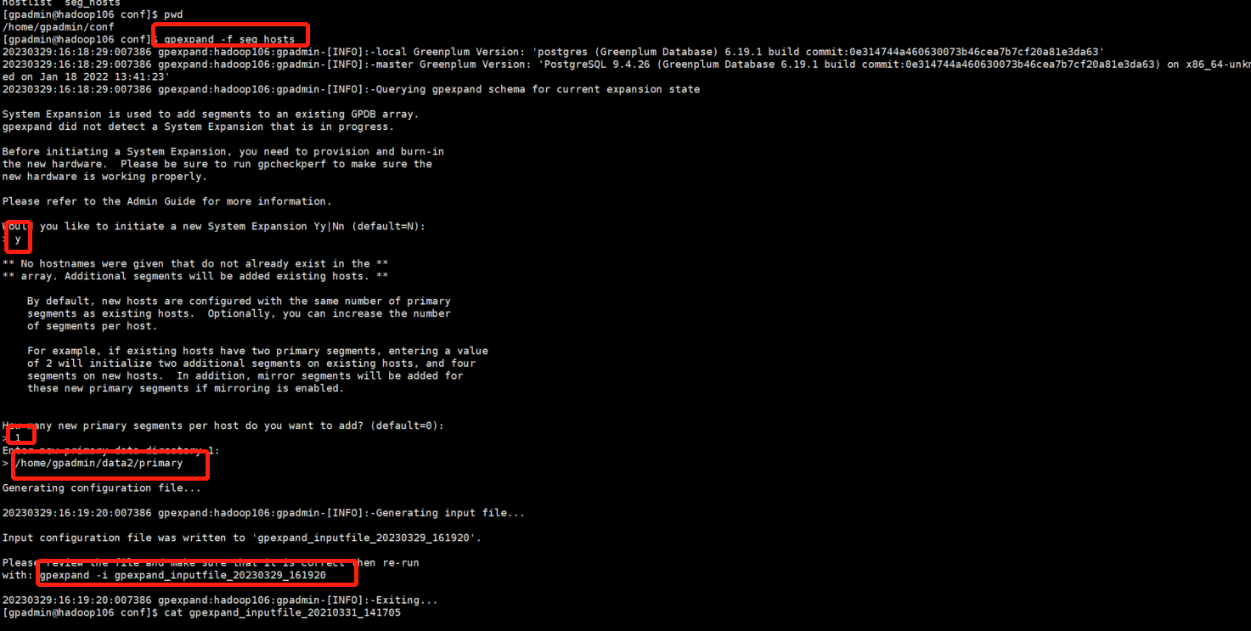
创建初始化文件

查看目前segment的主机

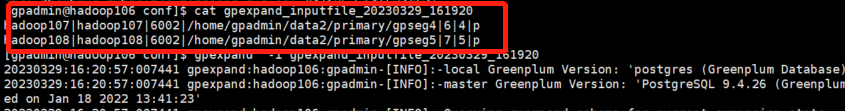
cd /home/gpadmin/conf

cat seg\_hosts

执行：gpexpand -f seg\_hosts

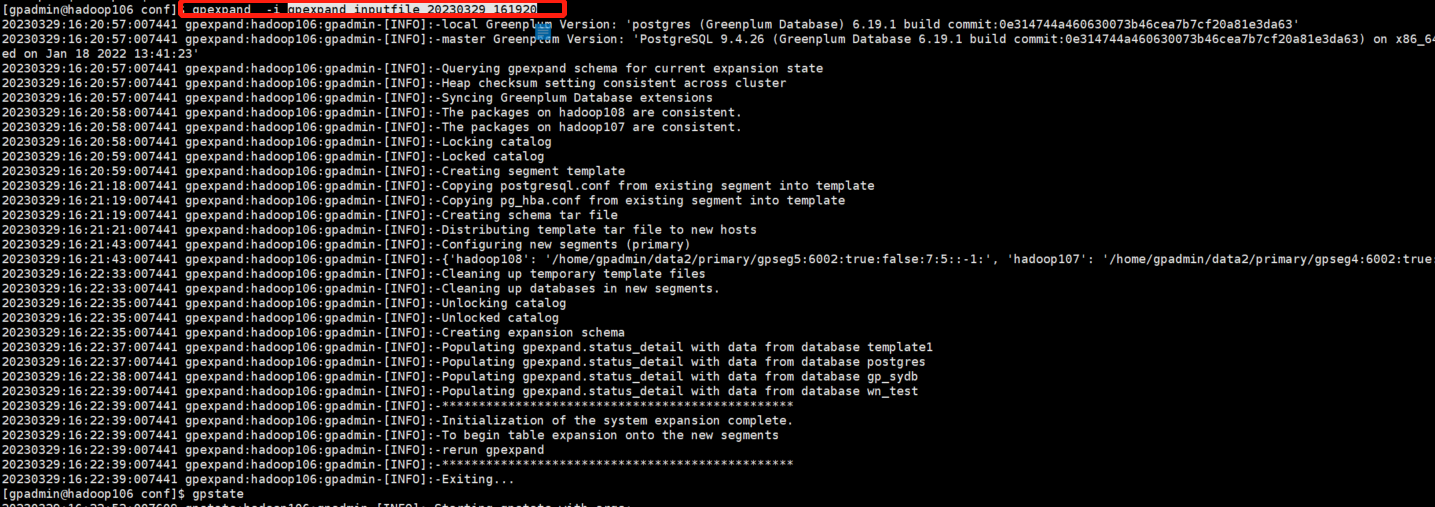


查看初始化文件：



初始化Segment并且创建扩容schema

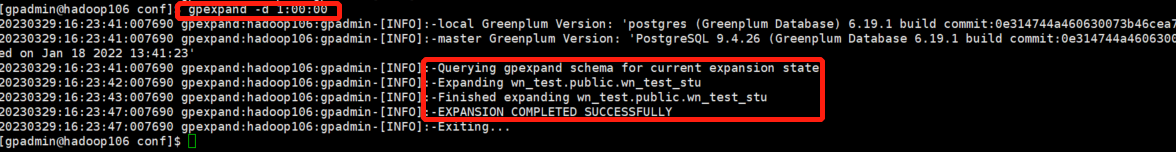
执行命令：gpexpand -i gpexpand\_inputfile\_20230329\_161920



使用gpstate验证下：（segment为6了，成功）

重分布数据

执行命令：gpexpand -d 1:00:00 #不动命令回去看gpexpand命令说明，这里没有业务表，所以很快就重分布完成了，如果数据量很大，可以增加线程



4. Greenplum 添加mirror步骤

概述

新安装的greenplum集群只有primary节点，没有mirror。高可用性没得到保证。所以就需要为集群添加mirror节点

注意：数据量过大时添加mirror，会有较大的磁盘压力，因为后台会一直同步数据，这一点线上系统需要注意

添加前情况

初始化的时候只有primary

gpstate -b

查看系统配置表，发现只有primary节点

postgres= # SELECT \* from gp\_segment\_configuration ;

添加步骤

在所有需要添加mirror的主机，创建存放mirror的数据目录

可以使用gpssh命令

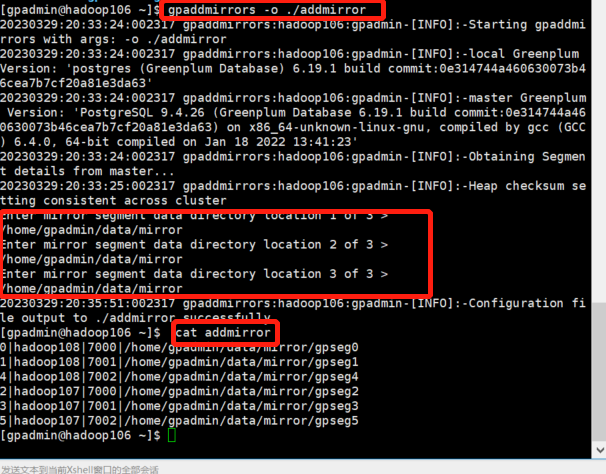
gpssh -f seg\_hosts -e 'mkdir -p /data/mirror'

也可手动添加

生成addmirror文件

执行gpaddmirrors 命令，键入存放mirror节点的日志，生成配置文件

[gpadmin@gw\_mdw1 ~]$ gpaddmirrors -o ./addmirror



执行添加命令

[gpadmin@gw\_mdw1 ~]$ gpaddmirrors -i addmirror

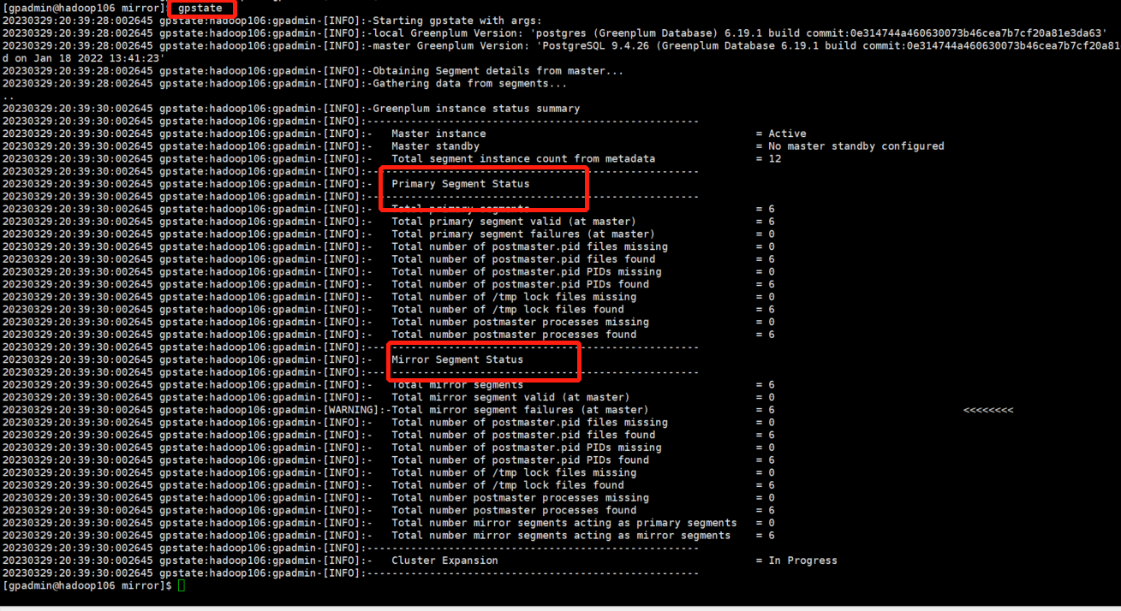
命令没有报错，查看mirror节点的情况

使用gpstate -m查看，发现所有的mirror正在同步数据，因为我的集群新建，很快就同步完成了，此时再执行gpstate -m就可以看到Data Status的状态是Synchronized（已同步的）

[gpadmin@gw\_mdw1 ~]$ gpstate -m

查看节点状态

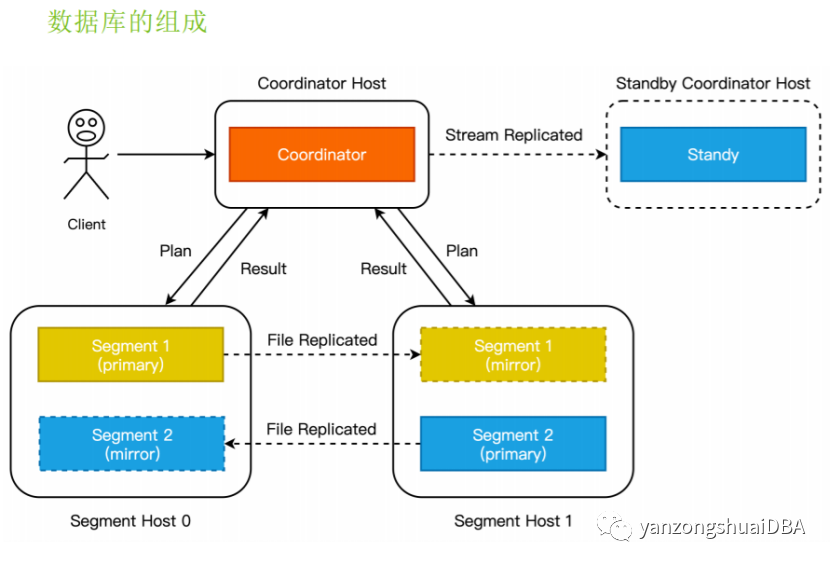
gpstate



发现所有的mirror已经启动

postgres= # SELECT \* from gp\_segment\_configuration ;

5. GreenPlum数据分布机制



数据默认使用hash分布。

Greenplum 6中 哈希分布 随机分布 新增 复制分布

哈希分布：

要使用这一策略，需要在创建表使用 “DISTRIBUTED BY（column，[...]）” 子句。

散列算法使分布键将每一行分配给特定segment。相同值的键将始终散列到同一个segment。选择唯一的分布键（例如Primary Key）将确保较均匀的数据分布。哈希分布是表的默认分布策略。

随机分布：

要使用这一策略，需要在创建表使用 “DISTRIBUTED RANDOMLY” 子句。

随机分布会将数据行按到来顺序依次循环发送到各个segment上。与哈希分布策略不同，具有相同值的数据行不一定位于同一个segment上。虽然随机分布确保了数据的平均分布，但只要有可能，应该尽量选择哈希分布策略，哈希分布的性能更加优良。

复制分布：

这种分布策略是GPDB 6的新增特性。

Greenplum数据分布和分区策略

要使用这一策略，需要在创建表使用 “DISTRIBUTED REPLICATED” 子句。

Greenplum数据库将每行数据分配到每个segment上。这种分布策略下，表数据将均匀分布，因为每个segment都存储着同样的数据行。适合于小表。

根据预定义的分布键(distributed by key)计算用户数据的哈希值，然后把哈希值映射到某个segment上。 分布键可以包含多个字段。分布键选择是否恰当是Greenplum能否发挥性能的主要因素，适合的分布键将数据均匀分布到各个 segment 上，避免数据倾斜。

重要结构体和函数

Greenplum 计算分布键哈希值的代码在src/backend/cdb/cdbhash.c文件中。结构体CdbHash是处理分布键哈希的主要数据结构。 计算分布键哈希值的逻辑为：

使用CdbHash \*makeCdbHash(int numsegs, int natts, Oid \*hashfuncs)创建一个 CdbHash 结构体

然后对每个 tuple 执行下面操作，计算该 tuple 对应的哈希值，并确定该tuple应该分布到哪个segment上：

○cdbhashinit()：执行初始化操作，仅仅是初始化hash初始值

○cdbhash()：这个函数会调用hashDatum()针对不同类型做不同的预处理，最后将处理后的列值添加到哈希计算中

○cdbhashreduce() ：映射哈希值到某个 segment



6.greenplum删除segment节点操作

[(23条消息) Deepgreen/Greenplum 删除节点步骤\_weixin\_30329623的博客-CSDN博客](https://blog.csdn.net/weixin_30329623/article/details/101092337?utm_medium=distribute.pc_relevant.none-task-blog-2~default~baidujs_utm_term~default-0-101092337-blog-52692280.235%5ev27%5epc_relevant_3mothn_strategy_recovery&spm=1001.2101.3001.4242.1&utm_relevant_index=3)

[https://blog.csdn.net/weixin\_30329623/article/details/101092337?utm\_medium=distribute.pc\_relevant.none-task-blog-2~default~baidujs\_utm\_term~default-0-101092337-blog-52692280.235^v27^pc\_relevant\_3mothn\_strategy\_recovery&spm=1001.2101.3001.4242.1&utm\_relevant\_index=3](https://blog.csdn.net/weixin_30329623/article/details/101092337?utm_medium=distribute.pc_relevant.none-task-blog-2~default~baidujs_utm_term~default-0-101092337-blog-52692280.235%5ev27%5epc_relevant_3mothn_strategy_recovery&spm=1001.2101.3001.4242.1&utm_relevant_index=3)

未做

硬件性能检测-安装完gp尽量先执行gpcheckperf来检查gp

监控工具-gpcc