# 生产Greenplum集群扩容说明

# 当前生产环境说明

* 1. 硬件环境

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **服务器类型** | **数量** | **IP** | **CPU** | **内存** | **硬盘** |
| **硬件** | MASTER | 1 | 172.16.18.10 | 2U32C | 128G | 1.5T |
| SEGMENT | 2 | 172.16.18.11，172.16.18.12 | 2U32C | 128G | 5.8T |
| Standby | 1 | 172.16.18.13 | 与ETL服务器合并部署 | | |

|  |  |
| --- | --- |
| 数据库版本 | 4.3.8.1 |
| 操作系统版本 | redhat6.5 |

* 1. 当期部署模式（**Grouped Mirroring**）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **MASTER** |  |  |  |
|  |  |  |  | **Standby** |
|  |  |  |  |  |
|  |  |  |  |  |
| **SEGMENT1** |  | **SEGMENT2** |  |  |
| P1 |  | P5 |  |  |
| P2 |  | P6 |  |  |
| P3 |  | P7 |  |  |
| P4 |  | P8 |  |  |
| M5 |  | M1 |  |  |
| M6 |  | M2 |  |  |
| M7 |  | M3 |  |  |
| M8 |  | M4 |  |  |

# 扩容说明

根据当前我们的部署模式，一次最少扩容两台。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **MASTER** |  |  |  |  |  |
|  |  |  |  | **Standby** |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| **SEGMENT1** |  | **SEGMENT2** |  | **SEGMENT3** |  | **SEGMENT4** |
| P1 |  | P5 |  | P9 |  | P13 |
| P2 |  | P6 |  | P10 |  | P14 |
| P3 |  | P7 |  | P11 |  | P15 |
| P4 |  | P8 |  | P12 |  | P16 |
| M5 |  | M1 |  | M13 |  | M9 |
| M6 |  | M2 |  | M14 |  | M10 |
| M7 |  | M3 |  | M15 |  | M11 |
| M8 |  | M4 |  | M16 |  | M12 |

# 新机扩容准备工作

* 1. 保证新增主机硬件与原机器保持一致，网络万兆，保持集群每台机器在统一网络内。
  2. 配置主机名

172.16.18.16@root> vi /etc/sysconfig/network

HOSTNAME=sdw3

172.16.18.16@root> vi /etc/hosts

172.16.18.10 mdw

172.16.18.11 sdw1

172.16.18.12 sdw2

172.16.18.13 ETLServer

#add

172.16.18.16 sdw3

172.16.18.17 sdw4

172.16.18.17@root> vi /etc/sysconfig/network

HOSTNAME=sdw4

172.16.18.17@root> vi /etc/hosts

172.16.18.10 mdw

172.16.18.11 sdw1

172.16.18.12 sdw2

172.16.18.13 ETLServer

#add

172.16.18.16 sdw3

172.16.18.17 sdw4

在集群其他机器/etc/hosts配置中增加

#add

172.16.18.16 sdw3

172.16.18.17 sdw4

* 1. 关闭防火墙

#服务器sdw3

sdw3@root>chkconfig iptables off

sdw3@root>chkconfig cups off

sdw3@root>chkconfig ip6tables off

#服务器sdw4

sdw4@root>chkconfig iptables off

sdw4@root>chkconfig cups off

sdw4@root>chkconfig ip6tables off

* 1. 修改操作系统参数

#sdw3/4 表示在 3跟4两台机器上执行

sdw3/4@root>vi /etc/sysctl.conf

# Kernel sysctl configuration file for Red Hat Linux

#

# For binary values, 0 is disabled, 1 is enabled. See sysctl(8) and

# sysctl.conf(5) for more details.

# Controls IP packet forwarding

net.ipv4.ip\_forward = 0

# Controls source route verification

net.ipv4.conf.default.rp\_filter = 1

# Do not accept source routing

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

# Controls the System Request debugging functionality of the kernel

kernel.sysrq = 0

# Controls whether core dumps will append the PID to the core filename.

# Useful for debugging multi-threaded applications.

kernel.core\_uses\_pid = 1

# Controls the use of TCP syncookies

net.ipv4.tcp\_syncookies = 1

# Disable netfilter on bridges.

net.bridge.bridge-nf-call-ip6tables = 0

net.bridge.bridge-nf-call-iptables = 0

net.bridge.bridge-nf-call-arptables = 0

# Controls the default maxmimum size of a mesage queue

kernel.msgmnb = 65536

# Controls the maximum size of a message, in bytes

kernel.msgmax = 65536

# Controls the maximum shared segment size, in bytes

kernel.shmmax = 68719476736

# Controls the maximum number of shared memory segments, in pages

kernel.shmall = 4294967296

#########for greenplum parameter###########################

kernel.shmmax = 500000000

kernel.shmmni = 4096

kernel.shmall = 4000000000

kernel.sem = 250 512000 100 2048

kernel.sysrq = 1

kernel.core\_uses\_pid = 1

kernel.msgmnb = 65536

kernel.msgmax = 65536

kernel.msgmni = 2048

net.ipv4.tcp\_syncookies = 1

net.ipv4.ip\_forward = 0

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

net.ipv4.tcp\_tw\_recycle = 1

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

net.ipv4.conf.default.rp\_filter = 1

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

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

net.ipv4.ip\_local\_port\_range = 1025 65535

net.core.netdev\_max\_backlog = 10000

vm.overcommit\_memory = 2

sdw3/4@root>sysctl -p

#修改磁盘调度策略

sdw3/4@root> vi /boot/grub/menu.lst

在kernel一行的最后加上elevator=deadline

#修改系统文件会话限制

sdw3/4@root>vi /etc/security/limits.conf

#add grennplum parameter

\* soft nofile 65536

\* hard nofile 65536

\* soft nproc 131072

\* hard nproc 131072

sdw3/4@root>ulimit -a

sdw3/4@root>vi /etc/security/limits.d/90-nproc.conf

\* soft nproc 1024

#将1024 修改为131072

sdw3/4@gpadmin>ulimit -a

#修改磁盘缓存设置

sdw3/4@root>vi /etc/rc.d/rc.local

blockdev --setra 16384 /dev/sdXX

#XX根据实际的数据目录挂载磁盘决定

#配置NTP服务

sdw3/4@root>service ntpd start

sdw3/4@root>chkconfig ntpd on

sdw3/4@root>vi /etc/ntp.conf

server mdw

sdw3/4@root>ntpdate -u mdw

#################重启服务器##############################

* 1. 创建用户及目录

挂载磁盘

sdw3/4@root>groupadd gpadmin

sdw3/4@root>useradd -g gpadmin -s /bin/bash -d /home/gpadmin gpadmin

sdw3/4@root>passwd gpadmin

sdw3/4@root>mkdir /greenplum

sdw3/4@root>mkdir -p /greenplum/gpdata

sdw3/4@root>chown gpadmin:gpadmin /greenplum/gpdata

#创建主节点数据目录

sdw3/4@gpadmin> mkdir -p /greenplum/gpdata/pdata

#创建备节点数据目录

sdw3/4@gpadmin> mkdir -p /greenplum/gpdata/mdata

* 1. 设置环境变量

sdw3/4@root>vi /etc/profile

#add greenplum parameter

source /home/gpadmin/greenplum-db/greenplum\_path.sh

sdw3/4@root> source /etc/profile

# 备份集群

需要在原集群中创建目录

gpadm@all:mkdir /greenplum/gpbak

gpadm@mdw:

gp\_dump sjck\_fb -n ods --gp-d=/greenplum/gpbak

gp\_dump sjck\_fb -n pdm --gp-d=/greenplum/gpbak

gp\_dump sjck\_fb -n int\_zx --gp-d=/greenplum/gpbak

gp\_dump sjck\_fb -n int\_fx --gp-d=/greenplum/gpbak

gp\_dump sjck\_fb -n int\_jl --gp-d=/greenplum/gpbak

gp\_dump sjck\_fb -n wb --gp-d=/greenplum/gpbak

# 新机集群环境准备工作

* 1. 新建新机hosts文件

mdw@gpadmin>cd /home/gpadmin/greenplum-db/gpconfigs/

mdw@gpadmin>vi new\_seg

sdw3

sdw4

* 1. 配置root用户的免密认证

mdw@root>cd /home/gpadmin/greenplum-db/gpconfigs/

mdw@root>gpssh-exkeys -e all\_hosts -x new\_seg

* 1. 配置gpadmin用户的免密认证

mdw@gpadmin>cd /home/gpadmin/greenplum-db/gpconfigs/

mdw@gpadmin>gpssh-exkeys -e all\_hosts -x new\_seg

* 1. 安装gp程序到新主机

mdw@gpadmin>cd /home/gpadmin/greenplum-db/gpconfigs/

mdw@gpadmin>gpseginstall -f new\_seg -u gpadmin

* 1. 创建扩容库

mdw@gpadmin>psql -d postgres

mdw@gpadmin#CREATE DATABASE gpexpand;

* 1. 检查

mdw@gpadmin>cd /home/gpadmin/greenplum-db/gpconfigs/

mdw@gpadmin>gpcheck -f new\_seg

20181102:16:07:15:050703 gpcheck:mdw:root-[INFO]:-dedupe hostnames

20181102:16:07:15:050703 gpcheck:mdw:root-[INFO]:-Detected platform: Generic Linux Cluster

20181102:16:07:15:050703 gpcheck:mdw:root-[INFO]:-generate data on servers

20181102:16:07:16:050703 gpcheck:mdw:root-[INFO]:-copy data files from servers

20181102:16:07:16:050703 gpcheck:mdw:root-[INFO]:-delete remote tmp files

20181102:16:07:16:050703 gpcheck:mdw:root-[INFO]:-Using gpcheck config file: /home/gpadmin/greenplum-db/./etc/gpcheck.cnf

20181102:16:07:16:050703 gpcheck:mdw:root-[ERROR]:-GPCHECK\_ERROR host(sdw4): on device (/dev/sda4) blockdev readahead value '256' does not match expected value '16384'

20181102:16:07:16:050703 gpcheck:mdw:root-[ERROR]:-GPCHECK\_ERROR host(sdw4): on device (/dev/sda) blockdev readahead value '256' does not match expected value '16384'

20181102:16:07:16:050703 gpcheck:mdw:root-[ERROR]:-GPCHECK\_ERROR host(sdw4): on device (/dev/sda1) blockdev readahead value '256' does not match expected value '16384'

20181102:16:07:16:050703 gpcheck:mdw:root-[ERROR]:-GPCHECK\_ERROR host(sdw4): on device (/dev/sda2) blockdev readahead value '256' does not match expected value '16384'

20181102:16:07:16:050703 gpcheck:mdw:root-[ERROR]:-GPCHECK\_ERROR host(sdw4): on device (/dev/sda3) blockdev readahead value '256' does not match expected value '16384'

20181102:16:07:16:050703 gpcheck:mdw:root-[ERROR]:-GPCHECK\_ERROR host(sdw3): on device (/dev/sda3) blockdev readahead value '256' does not match expected value '16384'

20181102:16:07:16:050703 gpcheck:mdw:root-[ERROR]:-GPCHECK\_ERROR host(sdw3): on device (/dev/sda1) blockdev readahead value '256' does not match expected value '16384'

20181102:16:07:16:050703 gpcheck:mdw:root-[ERROR]:-GPCHECK\_ERROR host(sdw3): on device (/dev/sda2) blockdev readahead value '256' does not match expected value '16384'

20181102:16:07:16:050703 gpcheck:mdw:root-[ERROR]:-GPCHECK\_ERROR host(sdw3): on device (/dev/sda) blockdev readahead value '256' does not match expected value '16384'

20181102:16:07:16:050703 gpcheck:mdw:root-[INFO]:-gpcheck completing...

mdw@gpadmin>gpcheckos -f new\_seg

mdw@gpadmin>gpcheckperf -f new\_seg -r N -d /greenplum > checknetwork.out

# 生成扩容配置文件

* 1. 检查集群状态

mdw@gpadmin>gpstate

mdw@gpadmin>gpstate -m

#保证集群状态都正常,不能有主备实例切换或者一个机器宕机的情况，一定要保持集群状态是完美正常运行！！

* 1. 生成配置文件

mdw@gpadmin>cd /home/gpadmin/greenplum-db/gpconfigs/

mdw@gpadmin>gpexpand -f new\_seg -D gpexpand

Would you like to initiate a new System Expansion Yy|Nn (default=N):

>y

What type of mirroring strategy would you like?

spread|grouped (default=grouped):

>grouped

How many new primary segments per host do you want to add? (default=0):

>0

# 执行扩容集群

#查看上一步骤中生成的配置文件

mdw@gpadmin>cd /home/gpadmin/greenplum-db/gpconfigs/

mdw@gpadmin>gpexpand -f new\_seg -i gpexpand\_inputfile\_XXX -D gpexpand -S -V -v -n 16 -B 1 -t /tmp

#gpexpand\_inputfile\_XXX 为生成的配置文件名字

# 修改数据重分布策略

* 1. 当前各层数据分布

mart 31 GB

ods 101 GB

pdm 144 GB

sdata 2285 GB

summ 287 GB

* 1. 分布策略

用gpadmin用户登陆数据库gpexpand数据库做如下操作

SDATA层表不做数据成分布

Delete from gpexpand.status\_detail where fq\_name like ‘sdata.%’;

Update gpexpand.status\_detail set rand =10;

Update gpexpand.status\_detail set rand =1 where fq\_name like ‘mart.%’;

Update gpexpand.status\_detail set rand =2 where fq\_name like ‘ods.%’;

Update gpexpand.status\_detail set rand =3 where fq\_name like ‘pdm.%’;

Update gpexpand.status\_detail set rand =4 where fq\_name like ‘summ.%’;

#删除准备用分区重构的表

Delete from gpexpand.status\_detail where fq\_name like ‘pdm.p03\_cotr\_repay\_seq’;

* 1. 执行数据重分布

mdw@gpadmin>gpexpand -d 02:30:00 -n 2 -D gpexpand

#小时:分钟:秒

* 1. 查看重分布进度

select \* from gpexpand.expansion\_progress

# 清理扩展库

#数据重分布完成后,检查select \* from gpexpand.expansion\_progress进度

mdw@gpadmin>gpexpand -c -D gpexpand

#会清理gpexpand下的两张表

#会将gpexpand.status\_detail 备份至

🡪/greenplum/gpdata/master/gpseg-1/gpexpand.status\_detail