南大通用

GBase 8a MPP集群

使用入门

讨论稿

2020-07-11

www.gbase8.cn

南 大 通 用 数 据 技 术 股 份 有 限 公 司

General Data Technologies Co., Ltd.

版权所有©GBASE2004-2020

天津总公司：中国天津华苑产业区海泰发展六道6号海泰绿色产业基地J座5层

电　　 话：022-58815678

传　　　真：022-58815679

北京分公司：北京市朝阳区太阳宫中路12号太阳宫大厦10层1008室

电　　　话：010-88866866

传　　　真：010-88864556

http://www.gbase.cn

E-mail:info@gbase.cn



|  |  |  |
| --- | --- | --- |
| **日期** | **变更说明** | **作者** |
| 2020-07-11 | 新建 | 赵学庆 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

目录

[1 概述 4](#_Toc45359326)

[1.1 约定 4](#_Toc45359327)

[2 环境准备 4](#_Toc45359328)

[2.1 机器数量 4](#_Toc45359329)

[2.2 操作系统 5](#_Toc45359330)

[2.3 数据库版本 5](#_Toc45359331)

[3 安装 5](#_Toc45359332)

[3.1 上传 5](#_Toc45359333)

[3.2 解压 5](#_Toc45359334)

[3.3 配置 6](#_Toc45359335)

[3.4 安装 7](#_Toc45359336)

[3.5 设置分布策略 8](#_Toc45359337)

[3.6 初始化 9](#_Toc45359338)

[4 初始化业务 9](#_Toc45359339)

[4.1 创建数据库 9](#_Toc45359340)

[4.2 创建表 10](#_Toc45359341)

[4.3 插入数据 11](#_Toc45359342)

[4.4 查询 11](#_Toc45359343)

[5 加载 11](#_Toc45359344)

[5.1 构造数据 12](#_Toc45359345)

[5.2 加载入库 12](#_Toc45359346)

[5.3 其它一些协议的加载样例如下 13](#_Toc45359347)

[6 导出 13](#_Toc45359348)

# 概述

本文档是介绍南大通用GBase 8a MPP集群(后面简称8a集群)的安装、简单使用，加点运维等基本操作，目标人群是刚接触8a集群的开发设计人员。

此文档只介绍基本使用，部分功能的详细说明，请查看产品手册。

本文档只包含8a集群V8版本在redhat 7上的操作样例。

## 约定

|  |  |
| --- | --- |
| 名字 | 说明 |
| 管理节点  gcluster节点 | 协调节点，coordinator节点，用于对外提供服务，解析和下发执行计划的节点。  包含gcluster的目录和gclusterd进程。  管理节点只保存表结构，不保存真实数据。  小规模集群里，管理和数据节点可以部署在同一个机器上。 |
| 数据节点  gnode节点 | 负责保存数据，执行具体SQL的节点。  小规模集群里，管理和数据节点可以部署在同一个机器上。 |
| 协调节点  gcware节点 | 负责集群一致性的服务。 |
| Dba用户  dbaUser | 数据库安装的操作系统用户，一般都是gbase |

# 环境准备

本章介绍简单【试用】8a集群所需要的环境。详细的环境需求，请参考产品手册的安装部分。

## 机器数量

1台即可，包括虚拟机。本文档只包含入门内容，如需进行性能测试或高可用测试，请主备最少3台服务器。

## 操作系统

本文档在redhat7版本操作样例，包括兼容的centos 7, cgsl5等。

内存建议最低4G，如果你准备做并发测试等，建议不低于16G。

磁盘空间请根据数据情况定，建议不低于10G，其中根目录可用空间不低于2G。

操作系统root密码。虽然支持sudo用户安装，如需要请参考产品安装手册。

本机的对外IP，本例是10.0.2.106 为一个ipv4地址。

关闭防火墙。

## 数据库版本

如您没有从其它渠道获得8a集群版本安装包，可以从官网下载免费版，([www.gbase8a.com](http://www.gbase8a.com))，比如：

GBase8a\_MPP\_Cluster-NoLicense-FREE-8.6.2\_build43-R7-redhat7.3-x86\_64.tar.bz2

其中RHEL7.3代表需要在redhat 7操作系统上运行。

x86\_64代表居于x86的64位CPU。

8.6.2是产品大版本号。

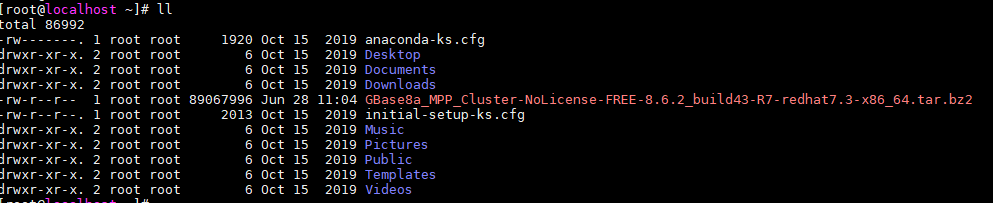
Build43-R7是小版本号。

# 安装

本章介绍8a集群的简单安装步骤，目标是一个可以建库，建表的使用环境。

## 上传

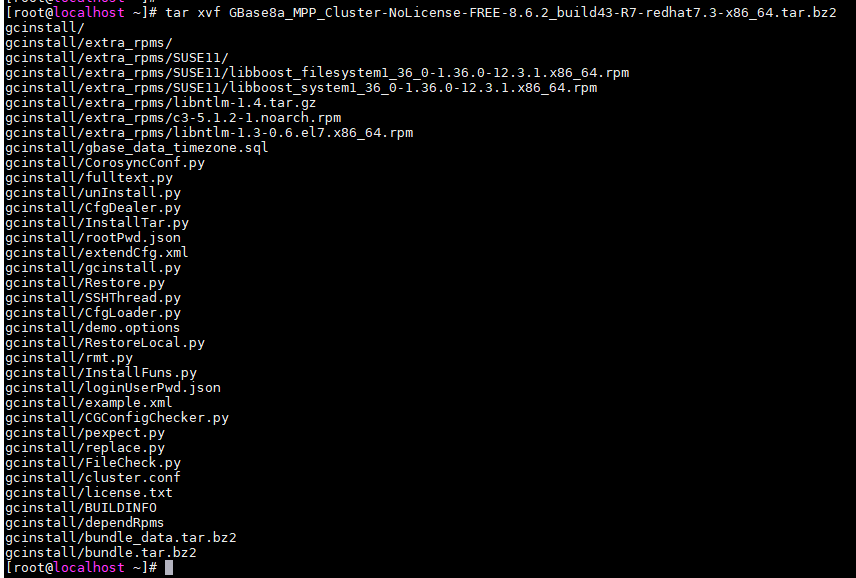
上传安装包到1台服务器的/root下面，如果有多台，【只需要】上传到任意一台即可。



## 解压

在当前位置解压缩：

tar xvf GBase8a\_MPP\_Cluster-NoLicense-FREE-8.6.2\_build43-R7-redhat7.3-x86\_64.tar.bz2



## 配置

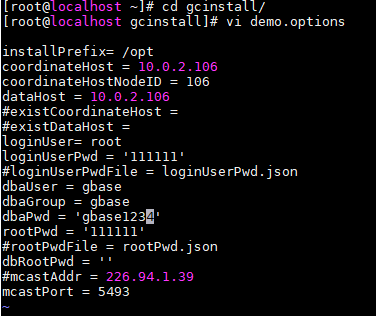
进入gcinstall 目录

cd gcinstall

编辑配置文件

vi demo.options

编辑后的配置文件样例如下(请忽略#开头的行)



如下红色的参数是必须要修改的。

installPrefix 是安装的目录。

coordinateHost 是本次安装的机器IP地址

coordinateHostNodeID, 用于ipv6地址，改成和ipv6的最后部分相同即可。不能重复。ipv4可忽略。

loginUser 登陆操作系统的用户名，本例用root用户。

loginUserPwd 登陆操作系统用户的密码，本例是root用户的密码。

dbaUser 数据库使用的操作系统用户，默认是gbase。

dbaGroup 数据库使用的操作系统用户的所属用户组，默认gbase。

dbaPwd 数据库使用的操作系统用户的密码。

rootPwd 操作系统用户root的密码，本例中和前面的loginUserPwd相同。

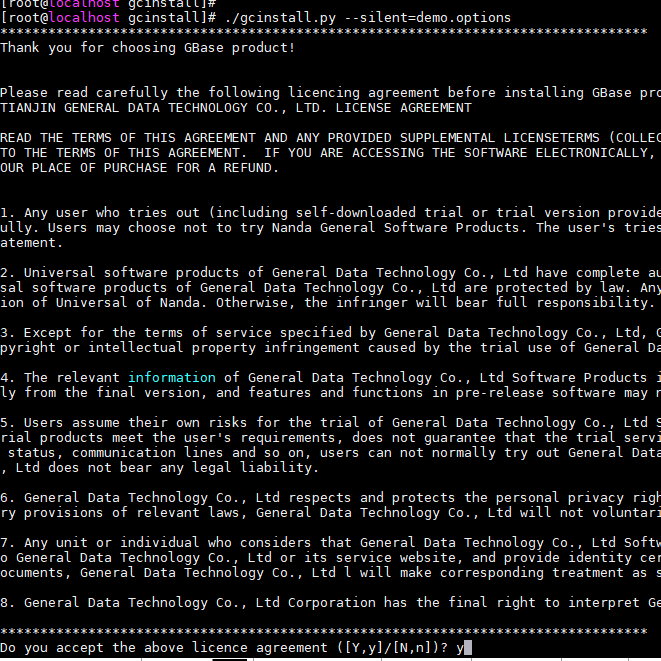
后面的配置忽略

## 安装

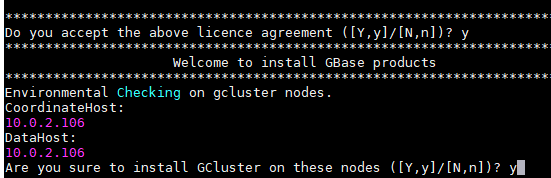
执行

./gcinstall.py --silent=demo.options

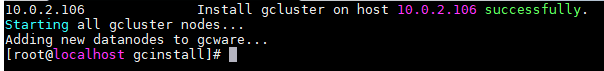
并在许可同意部分输入字母y后按回车



在系统检查成功后，在确认安装时输入字母y后按回车。



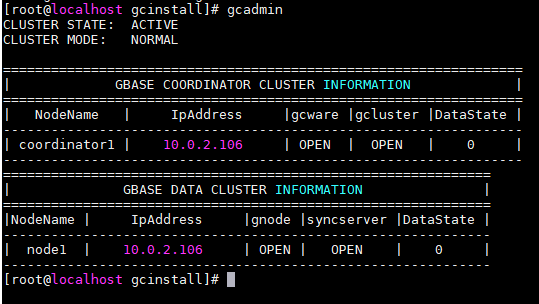
等待安装完成



确认安装结果

gcadmin

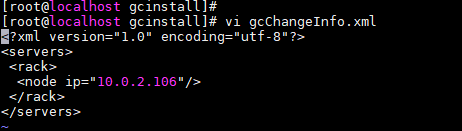
如果安装后立即运行如上命令报错，可以稍等10秒再尝试1次。



## 设置分布策略

编辑分布策略的xml文件。默认安装程序会在当前目录下生成一个gcChangeInfo.xml

vi gcChangeInfo.xml



里面的node ip 是我们新安装集群的IP, 本例中我们不修改这个。

切换到操作系统gbase用户

su gbase

，请注意如果你目录变了，请转到我们安装程序的目录。

新建数据库分布策略

gcadmin distribution gcChangeInfo.xml p 1 d 0

其中

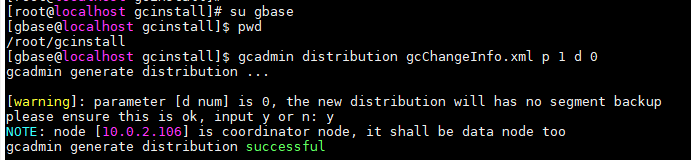
distribution 是创建新的分布策略的命令。

gcChangeInfo.xml收我们分布策略配置文件。

p 是每个节点的分片数量，我们设置为1。

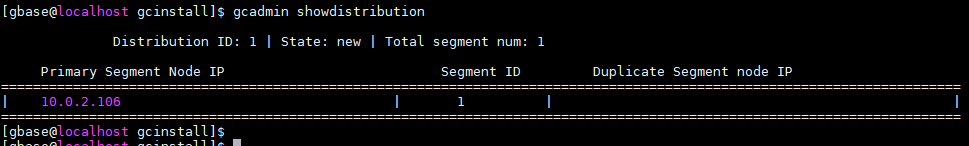
d 是副本数量，由于本例是单个节点，副本数为0. 如果你是多个节点，这里可以设置为1。

在确认时输入字母y然后回车



通过如下命令检查集群分布策略是否生效

gcadmin showdistribution



其中右半部分的Duplicate Segment node IP是副本的IP,我们无副本，所以是空的。

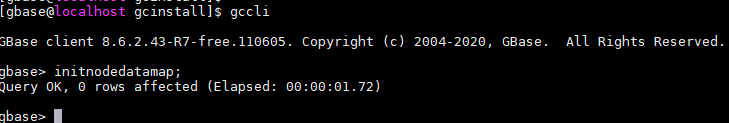
## 初始化

使用命令行客户端登陆集群

gccli

执行如下命令

initnodedatamap;



至此，数据库已经安装完毕，可以正式提供服务了。

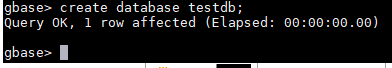
# 初始化业务

本部分是介绍数据库的最基本操作，创建数据库，创建表，插入数据并查询。

## 创建数据库

执行如下命令，创建指定名字的数据库。

create database testdb;



切换到testdb数据库

use testdb;

C:\Users\hasee\AppData\Local\Temp\1593325147(1).png

## 创建表

创建表，我们简单设置一个人员名字表，包含如下几个字段：

id 人员编号， int

name 姓名，varchar

joinDate 加入日期，date

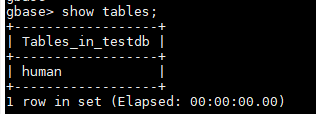
salary 薪水，Decimal

create table human(id int,name varchar(100),joindate date,salary decimal(10,2));

C:\Users\hasee\AppData\Local\Temp\1593325200(1).png

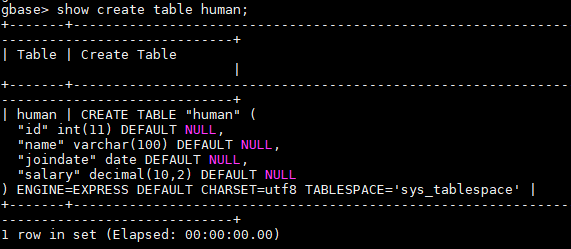
查看当前库有哪些表

show tables;



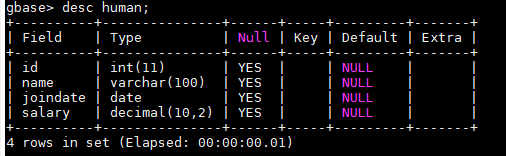
查看表的建表语句

show create table human;



查看表结构

desc human



确认我们创建的表是正确的。

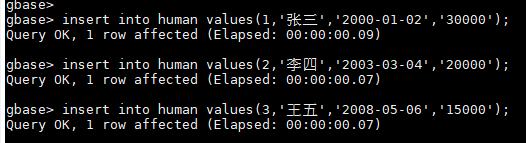
## 插入数据

通过insert方式，插入3行数据。

insert into human values(1,'张三','2000-01-02','30000');

insert into human values(2,'李四','2003-03-04','20000');

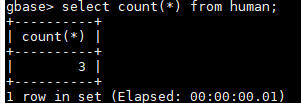
insert into human values(3,'王五','2008-05-06','15000');



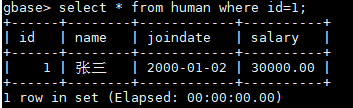
## 查询

通过select语句，查询我们保存到数据库的数据。

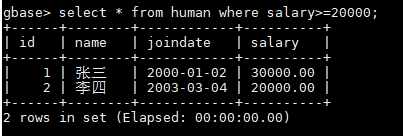
select count(\*) from human;



select \* from human where id=1;



select \* from human where salary>=20000;



# 加载

由于8a集群的单条insert性能很差，每秒最高也就几十到几千条，在大数据下(建议千行以上)，都是通过加载的方式，批量入库。

本部分仅用于8a集群加载功能的演示，如关心性能，请自行构造业务表以及业务数据，进行测试。

## 构造数据

我们人工构造了几行数据，文件为/home/gbase/human.txt

[gbase@localhost ~]$ cat human.txt

1,First,2012-12-23,10000.12

2,Second,2013-11-12,8888.88

3,中国人,2019-08-09,9999.99

[gbase@localhost ~]$ pwd

/home/gbase

## 加载入库

本次加载只用于演示，采用基于ssh的sftp入库。如测试性能，请使用ftp,hdaoop等协议。

load data infile 'sftp://gbase:gbase1234@10.0.2.106/home/gbase/human.txt' into table testdb.human fields terminated by ',';

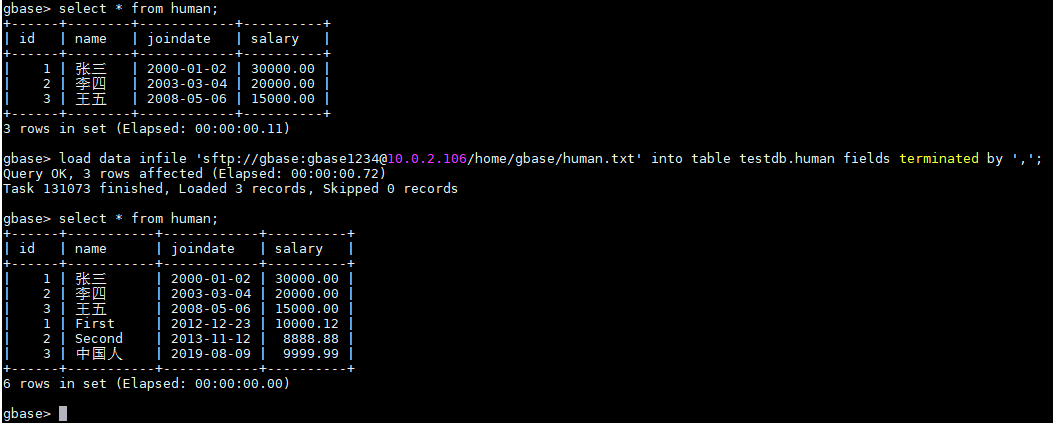
其中数据源采用了sftp格式

sftp://gbase:gbase1234@10.0.2.106/home/gbase/human.txt

入库目标是 testdb.human

fields表示字段的一些定义，此处只设定了字段之间的分隔符为单引号。

terminated by ','



## 其它一些协议的加载样例如下

ftp://gbaseload:gbaseload@192.168.1.1/2.csv

ftp://gbase:gbase@127.0.0.1//home/gbase/data/a.tbl

sftp://gbase:gbase@127.0.0.1/data/a.tbl

hdp://hadoop@192.168.10.1:50070/data/test.tbl

请参考 <http://www.gbase8.cn/?p=121>

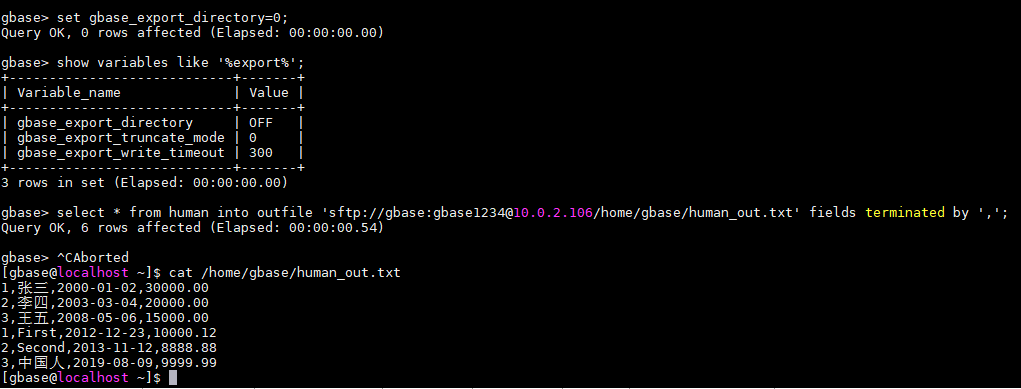
# 导出

将数据库的数据，导出为平稳本格式的数据文件。

如下例子是用逗号分割的样例

select \* from human into outfile 'sftp://gbase:gbase1234@10.0.2.106/home/gbase/human\_out.txt' fields terminated by ',';

参数与加载一样的。



由于默认导出时，会创建一个和导出文件相同的目录，比如导出为human\_out.txt, 实际上是 XXX/human\_out.txt/human\_out.txt

如上通过

set gbase\_export\_directory=0;

关闭导出时生成目录的参数。

参考 <http://www.gbase8.cn/?p=119>