



科学与工程计算中心

Center for Computational Science and Engineering

“太乙”

超级计算机

使用手册

更新时间：2020 年 11月

## 目录

一、 集群基本情况.....	3
1.1 系统概况.....	3
1.2 硬件概况.....	3
二、 登陆集群.....	3
2.1 VPN账号.....	3
2.2 校外登录（校内不用vpn）.....	4
2.3 集群账号.....	4
2.4 登陆和ftp.....	5
2.5 登陆界面.....	5
三、 集群编译器及环境.....	6
3.1 查看模块.....	6
3.1 加载模块.....	6
3.1 卸载模块.....	7
四、 使用集群.....	7
4.1 集群目录.....	7
4.2 集群回收站.....	7
4.2 集群队列.....	8
4.3 查询存储使用量.....	8
4.4 作业范例.....	8
4.5 作业调度系统使用.....	9
4.6 计费与充值.....	11
4.7 查询机时及费用.....	11
4.8 问题反馈.....	12

## 一、 集群基本情况

### 1.1 系统概况

二期集群（太乙，Tai-Yi）包含815个双路刀片节点，2个大内存节点和4个GPU节点，GPFS 并行文件存储系统，Intel Omni-Path 高速(100 Gbps OPA)计算网络。太乙系统理论计算峰值超过2500万亿次，实测计算性能超过1600万亿次。在2018年11月的全球Top 500中排名第127。

### 1.2 硬件概况

系统峰值性能：2.5PFlops	实测持续运算性能：1.687PFlops
计算节点：815 台	配置：2 个 Xeon Gold 6148 CPU(2.4GHz/20c)，192 GB 内存)
大内存节点：2 台	配置：8 个 Xeon Platinum 8160 CPU (2.1GHz/24c)，6 TB 内存
GPU 节点：4 台	配置：2 个 Xeon Gold 6148 CPU(2.4GHz/20c)，384 GB 内存，2 张 NVIDIA V100
存储：5.5 PB	配置：GPFS 并行文件系统，实测读写带宽均超过 40 GB/s

## 二、 登陆集群

### 2.1 VPN账号

在学校外面登陆集群，首先需要VPN账号，登陆VPN账号后，再登陆集群。

2.1.1 申请VPN账号：发邮件至网络信息中心 [its@sustech.edu.cn](mailto:its@sustech.edu.cn);

2.1.2 图2.1描述VPN账号登陆过程：



(图：2.1)

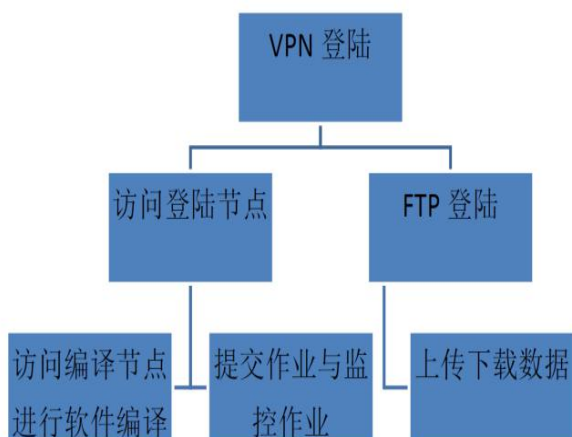
2.1.3      图：2.2红色框显示已连接和IP，表示vpn登陆成功



(图：2.2)

## 2.2 校外登录（校内不用vpn）

图2.3显示校外登陆集群流程

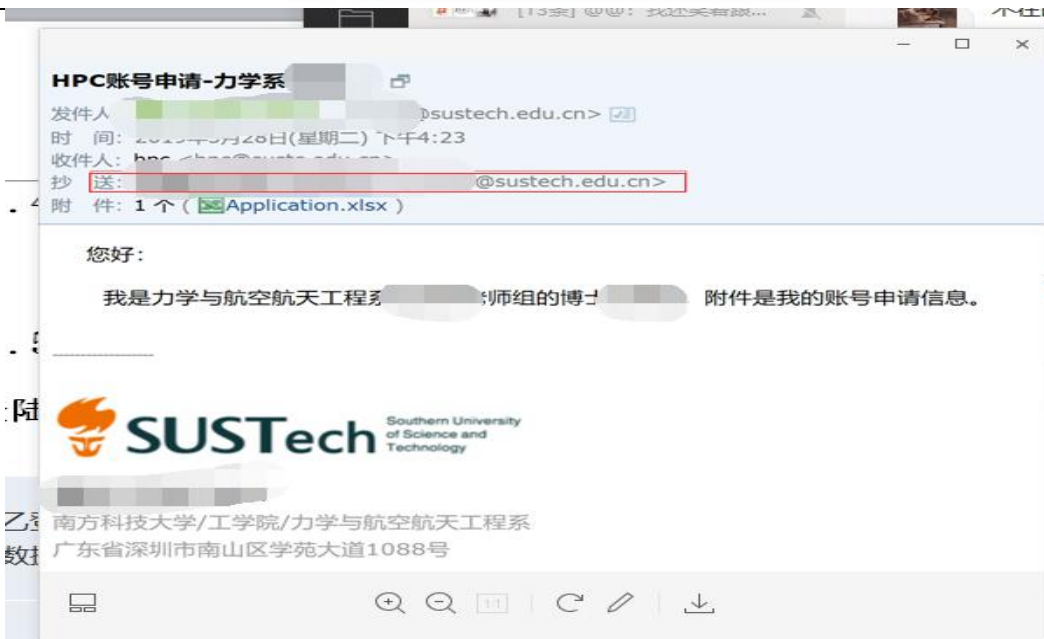


(图:2.3)

## 2.3 集群账号

### 2.3.1 集群账号申请

申请集群账号发邮件至 (hpc@sustech.edu.cn)；



(图：1.3-1)



(图：1.3.-2)

- 注：1、图1.3-1为申请账号样例；
- 2、申请表在<http://hpc.sustech.edu.cn/>首页下载；
- 3、申请邮件需抄送给PI（课题组长）；
- 4、账号开通后，有回复邮件，图1.3-2为样例；

## 2.4 登陆和ftp

太乙登陆和传输数据，采用的动态负载技术，使用IP：

172.18.6.175

## 2.5 登陆界面

```

NOTICE:
-----
CCSE solicits speakers for its users meeting to exchange HPC and related experience
link:https://hpc.sustech.edu.cn/newsdetail/notice05.html
-----

QUEUE'S RUNLIMIT NOTICE.
debug: 1h      Unlimited
ser:    360h    Unlimited
short:  360h    More Than 40c
medium: 720h    More Than 120c
large:  12h     More Than 1000c
smp:    360h    Unlimited
gpu:    168h    Unlimited

User--guide: http://hpc.sustech.edu.cn/process/user_guide.html
Only login ip: 172.18.6.175

Support Team Email: hpc@sustech.edu.cn
Support Team Phone: 0755-88015834, 15986647786, 15818796162

Attention:
1. Please do not run jobs that keep generating numerous small files for a long time before contacting the support team for a proper solution, because it will cause a high pressure to the storage system and slow down everyone.
2. Please do not run large memory jobs on the login node, because it will slow down everyone.
3. Please do not use MPICH in job, because it may cause a high pressure to the storage system and slow down everyone.

Warning: violation of the above rules may result in job termination and account blocking
  
```

- NOTICE是通知内容(QQ群:161195492和邮件:ccse@sustech.edu.cn也发通知);
- QUEUE'S RUNLIMIT NOTICE是队列时间限制和最小核数限制;
- Attention为上机注意事项: 1. 未经允许, 不能在集群上运行持续创建大量小文件的作业(文件数量级别在千万以上), 因为这样的作业会给共享存储带来巨大的IO压力从而影响正常的作业; 2. 在登陆节点上, 禁止运行消耗CPU或者内存资源的作业; 3. 禁止在作业中使用MPICH;
- 登陆界面的内容, 会根据实际情况进行调整。

## 三、 集群编译器及环境

由于用户可能需要使用不同的软件环境, 配置不同的环境变量, 因而在“太乙”上安装了“module 软件”来进行管理, 用户方便环境变量的设置从而提高移植软件效率。因为“太乙”的cpu为intel Xeon Gold, 建议使用Intel编译器。

### 3.1 查看模块

module avail: 查看可用的模块列表, 如图:3-1

```

[root@login01 Li-bulk]# module avail
----- /share/base/modulefiles/compilers -----
gcc/8.2.0(default)  intel/2015.6      intel/2018.3      intel/2019.1
gcc/1.11.1          intel/2017.8      intel/2018.4      intel/2019.4bk
----- /share/base/modulefiles/parallel -----
dapl/2.1.7          itac/2018.3.022    mpi/intel/2019.1  mpi/openmpi/3.1.2_gcc  mpi/platform/9.1.4.r3
dapl/2.1.9          knem/1.1.3         mpi/intel/2019.4bk mpi/openmpi/3.1.2_intel mpi/4.3.1/impi
dapl/2.1.9-fca      mpi/intel/2017.8   mpi/mvapich2/2.3b  mpi/platform/2.3.0
ga/5-4b/mpi         mpi/intel/2018.3   mpi/mvapich2/2.3_gcc mpi/platform/8.1.1
ipsxe/2018.3.222    mpi/intel/2018.4   mpi/openmpi/2.1.2_gcc mpi/platform/9.1.2
----- /share/base/modulefiles/bio -----
amber/16
----- /share/base/modulefiles/chem -----
vasp/5.4.1          vasp/5.4.4         vasp/5.4.4-vtst
----- /share/base/modulefiles/cae -----
ansys/192          CFX/192            fluent/192        matlab/2017b        OpenFoam/3.0
----- /share/base/modulefiles/graphics -----
gnuplot/5.2.6       jasper/1.900.1/gnu jpeg/9a            ncl/6.5.0-gnu485    png/1.6.17
  
```

(图: 3-1)

### 3.1 加载模块

module load [modulesfile]: 加载使用的 modulefiles, 如图:3-2



```
[ccse-xiezy@login01 ~]$ module load intel/2018.4  
[ccse-xiezy@login01 ~]$ which icc  
/share/intel/2018u4/compilers_and_libraries_2018.5.274/linux/bin/intel64/icc  
[ccse-xiezy@login01 ~]$
```

(图：3-2 )

### 3.1 卸载模块

`module unload [modulesfile]`: 移除使用 `module` 加载的软件环境, 如图:3-3

```
[ccse-xiezy@login01 ~]$ module unload intel/2018.4  
[ccse-xiezy@login01 ~]$ which icc  
~/intel/compilers_and_libraries_2019.4.243/linux/bin/intel64/icc  
[ccse-xiezy@login01 ~]$
```

(图：3-3 )

## 四、 使用集群

### 4.1 集群目录

三个目录可以使用(work,data,scratch), 限额是针对课题组而言, 数据量小(例如: 不到10G), 建议直接使用/work目录, 以下内容可以忽略。如果数据量大, 请仔细耐心阅读以下内容:

- /work目录限额为1TB, 存放核心数据, 例如: 源代码, 安装部署软件;
- /data目录限额为8TB(可用容量4TB), 存放运算数据, 支持存储容量扩容, 扩容费用见收费政策;
- /scratch目录不限额, 用于中间数据量大的作业。例如: Gaussian运算的过程中, 产生10T的大文件, 而/work和/data空间限额没有10T, 作业运算不了。这时可以使用/scratch目录;
- 为了保证/scratch有充足的存储空间, 设置自动清理10天前目录的数据, /scratch每天会自动产生一个日期目录, 用户可以将数据写到这个目录, 但是数据最多保存10天, 例如/scratch/2020-01-01这个目录在2020年1月1日自动创建, 2020年1月11日会自动迁移目录/scratch/2020-01-01至回收站, 并且2020年1月11日前, 用户能收到数据删除的邮件提醒通知。如果有数据需要保存, 请在十天内及时把数据迁移到/data或者/work。如作业持续运行时间超过10天, 请不要使用/scratch目录, 使用/data或/work目录;
- /data目录采用双副本的存储方式(存储系统为了保证数据安全采用的一种存储方式), 例如: 1T的数据, 其实占有2T的空间; 扩容收费还是按照1T来收, 详细情况请阅收费政策;
- 基于存储空间容量的限制和对数据安全的考虑, 请用户及时把重要数据或敏感数据保存到自己的计算机中, 并及时清理自己的存储空间。

### 4.2 集群回收站

`rm`命令删除数据, 如果发生删除错误, 或者scratch目录数据被自动清理, 数据能否找回? 为了应对这个问题, 系统设置了回收站, 回收站的位置:

/scratch/recyclebin/. 当天和前天目录, 分别回收当天和前天删除的数据, 回收站里面的数据保存2天, 2天后, 数据被彻底被清理。

因为回收站是放在目录为/scratch,该目录没设限额，所以不占用用户的存储空间。如果不想使用回收站，想直接删除数据，那么使用绝对路径命令：/bin/rm,这种方式删除数据，是不能找回。

4.2 集群队列

“太乙”集群目前共开放以下七个队列：

队列	队列时长限制	应用场景	单作业核心数限制
large	12小时	大规模	1000核以上（含1000）
medium	720小时	中等规模	120核以上（含120）
short	360小时	小规模	40核以上（含40）
ser	360小时	串行	无限制
smp	360小时	大内存	无限制
debug	1小时	调试	无限制
gpu	24小时	GPU	无限制

注：以上数据可能会根据实际情况，进行调整更改，请以2.6节：QUEUE’ S RUNLIMIT NOTICE为准。

4.3 查询存储使用量

使用mmlsquota,若是个人，使用参数-u。若是课题组，使用参数-g;

```
logout
[ccse-xiezy@login01 work]$ mmlsquota -u ccse-xiezy --block-size auto
Block Limits
Filesystem type blocks quota limit in_doubt grace | File Limits
files quota limit in_doubt grace Remarks
data USR no limits

Filesystem Fileset type blocks quota limit in_doubt grace | File Limits
files quota limit in_doubt grace Rema
rks scratch root USR no limits

Filesystem type blocks quota limit in_doubt grace | File Limits
files quota limit in_doubt grace Remarks
work USR no limits
[ccse-xiezy@login01 work]$ mmlsquota -g ccse-xiezy --block-size auto
Disk quotas for group ccse-xiezy (gid 1284):
Block Limits
Filesystem type blocks quota limit in_doubt grace | File Limits
files quota limit in_doubt grace Remarks
data GRP 0 1.953T 2.051T 0 none | 1 0 0 0 none

Filesystem Fileset type blocks quota limit in_doubt grace | File Limits
files quota limit in_doubt grace Rema
rks scratch root GRP no limits

Filesystem type blocks quota limit in_doubt grace | File Limits
files quota limit in_doubt grace Remarks
work GRP 47.5G 500G 600G 0 none | 83140 0 0 0 none
[ccse-xiezy@login01 work]$
```

(查询范例)

4.4 作业范例

范例1-"太乙"-vasp



注意:mpi建议采用2018.4版本, 范例可参考: /share/user\_guide/exam/mytest.lsf

```
#!/bin/sh
#BSUB -J N_F                                ##job name
#BSUB -q short                               ##queue name
#BSUB -n 80                                  ##number of total cores
#BSUB -R "span[ptile=40]"                   ##40 cores per node
#BSUB -W 12:00                               ##walltime in hh:mm
#BSUB -R "select[hname!='r13n18']"          ##exclusive r13n18
#BSUB -e err.Log                             ##error log
#BSUB -o H.Log                              ##output log
module load intel/2018.4 mpi/intel/2018.4 vasp/5.4.4
mpirun vasp_std &>log
```

## 范例2-"太乙"-自编mpi代码

```
#!/bin/bash
#BSUB -J test
#BSUB -q short
#BSUB -n 320
#BSUB -e %J.err
#BSUB -o %J.out
#BSUB -R "span[ptile=40]"
#BSUB -R "select[hname!='r03n43']"
#BSUB -R "select[hname!='r03n55']"
#BSUB -R "select[hname!='r03n64']"

module load fftw/2.1.5
module load intel/2018.4
module load mpi/intel/2018.4

cd $LS_SUBCWD
echo "processes will start at:"
date

mpirun -machinefile $LSB_DJOB_HOSTFILE -np 320 ./main > $LSB_JOBID.log 2>&1

echo "processes end at:"
date
```

- 编译 mpicc 必须与 mpirun 一致,这里是 2018.4
- 不推荐使用自己编译 mpi 软件和 fftw 库
- main 必须是可执行文件(chown a+x main)
- 更多范例, 请参考科学与工程计算中心官网: <http://hpc.sustech.edu.cn/userguide.html>
- 经过测试, 建议使用系统自带的 2018.4 版本 intel mpi

## 4.5 作业调度系统使用

"太乙"采用的作业调度平台是LSF 10.1。用户需要熟悉一些基本的Linux 命令行操作，特别是文件目录操作，并能熟练使用一种编辑器（vi 或emacs 等）。本文档介绍LSF常用命令，更加详细的内容请参考LSF文档：

[http://hpc.sustech.edu.cn/ref/lsf\\_users\\_guide\\_v10.1.pdf](http://hpc.sustech.edu.cn/ref/lsf_users_guide_v10.1.pdf)

#### 4.5.1 "太乙"-提交作业的命令

`bsub<test.lsf`

注意：需要有"<"符号

提交 LSF 任务，成功后会给出此任务的 JOBID。

`bjobs`查看自己的所有运行任务情况；输入`bjobs`后，会列出当前用户正在运行的所有作业，最左边一列数字是每个作业的 JOBID，一些其他命令使用的时候需要先调用`bjobs`查看JOBID。

```

bjobs -l 查看所有运行任务的详细情况
bjobs -l JOBID 查看 JOBID 这个任务的详细情况
bpeek -f JOBID 跟踪查看某任务屏幕输出
bkill JOBID 终止某任务运行
bkill JOBID1 JOBID2 JOBID3 终止多个任务运行
bqueues 查看所有任务队列的状态
bstop JOBID 临时挂起某个计算作业，为其它计算腾出资源
bresume JOBID 恢复由 bstop 挂起的作业
lshosts 查看节点的信息
bhosts 查看节点的作业使用信息
lsload 查看节点的即时负载信息
    
```

作业结束后，会在计算目录下生成`output.JOBID`，此文件为程序输出内容。

#### 4.5.2 "太乙"-任务监控

使用`bjobs`可以监控到以下参数：

RUN-作业已经分派到节点机上，正在运行；

DONE-作业已经正常结束；

EXIT-作业已经结束，但可能异常退出或者被终止；

UNKWN-作业状态不确定；

作业提交完后的5分钟，建议查看作业是否正常(A和B查看节点负载,C查看输出文件)：

A 在login01上: `jobToLoad 150614`，能查到这个作业所有节点负载：

```
login01 ~]$ jobToLoad 150614
```

HOST NAME	status	rl5s	rlm	rl5m	ut	pg	ls	it	tmp	swp	mem
r07n47	ok	40.1	41.2	40.6	100%	0.0	0	46944	213G	3.3G	157G
r05n03	ok	40.1	40.7	40.6	100%	0.0	0	46944	212G	3.4G	157G
r07n27	ok	40.1	40.2	40.6	100%	0.0	0	28048	212G	3G	157G
r05n07	ok	40.2	40.2	40.6	100%	0.0	0	46944	213G	3.2G	157G
r07n23	ok	40.2	40.5	40.5	100%	0.0	0	46944	212G	2.9G	157G
r07n21	ok	40.3	41.1	40.7	100%	0.0	0	46944	212G	3.1G	157G

B 查看当前用户所有作业负载: `jobToLoad -a`

```
@login01 ~]$ jobToLoad -a
```

HOST_NAME	status	r15s	rlm	r15m	ut	pg	ls	it	tmp	swp	mem
r09n04	ok	40.0	40.7	40.6	100%	0.0	0	46944	212G	2.8G	157G
r09n44	ok	40.0	40.2	40.5	100%	0.0	0	46912	212G	3.4G	157G
r08n31	ok	40.1	41.2	40.7	100%	0.0	0	46944	212G	2.7G	157G
r11n11	ok	40.1	40.7	40.6	100%	0.0	0	41344	212G	3.3G	157G
r08n37	ok	40.2	40.7	40.6	100%	0.0	0	46944	212G	2.3G	157G
r11n54	ok	40.2	40.5	40.6	100%	0.0	0	46912	212G	3G	157G
r09n50	ok	40.3	40.9	40.9	100%	0.0	0	46912	212G	2.6G	157G
r08n58	ok	40.3	40.5	40.6	100%	0.0	0	15656	212G	3.5G	157G
r11n17	ok	40.4	40.7	40.5	100%	0.0	0	46912	212G	2.9G	157G
r09n01	ok	40.4	40.3	40.6	100%	0.0	0	46944	212G	3.1G	157G
r08n56	ok	41.0	40.7	40.8	100%	0.0	0	46944	211G	3.5G	157G
r08n60	ok	42.3	40.2	40.6	100%	0.0	0	46944	212G	3G	157G
HOST_NAME	status	r15s	rlm	r15m	ut	pg	ls	it	tmp	swp	mem
r09n12	ok	40.0	40.0	40.2	100%	0.0	0	46912	212G	2.9G	157G
r09n55	ok	40.0	40.3	40.3	100%	0.0	0	46912	212G	3G	157G
r10n53	ok	40.0	40.1	40.3	100%	0.0	0	46912	212G	3G	157G
r11n06	ok	40.0	40.4	40.3	100%	0.0	0	46080	212G	3.3G	157G
r09n17	ok	40.3	40.2	40.2	100%	0.0	0	46944	213G	3.2G	157G
r07n12	ok	40.5	40.5	40.2	100%	0.0	0	46944	212G	3.1G	157G

C 查看输出文件:

```
#~~~~~
#lattice_vector -0.0000000000000000 5.4630289765190732 5.4630289765190732
#lattice_vector 5.4630289765190732 -0.0000000000000000 5.4630289765190732
#lattice_vector 5.4630289765190732 5.4630289765190732 -0.0000000000000000
# constrain_relaxation full x y z ...
#
#atom_frac 0.1250000000000000 0.6250000000000000 0.6250000000000000 Ag
#atom_frac 0.3773919634347005 0.3773919634347005 0.8678251096958962 S
# constrain_relaxation full x y z ...
#~~~~~
lattice_vector 18.00000000 0.00000000 0.00000000
lattice_vector 0.00000000 18.00000000 0.00000000
lattice_vector 0.00000000 0.00000000 18.00000000

atom_frac 0.91666667 0.08333333 0.91666667 Fe
initial_moment -4
atom_frac 0.41666667 0.41666667 0.58333333 Fe
initial_moment -4
atom_frac 0.08333333 0.91666667 0.08333333 Fe
initial_moment -4
atom_frac 0.91666667 0.25000000 0.75000000 Fe
initial_moment -4
atom_frac 0.91666667 0.25000000 0.08333333 Fe
```

### 4.5.3 "太乙"-任务实时输出

bpeek 输出所有内容; bpeek -f JOBID 通过重定向的方式将标准输出和标准错误重定向到指定的文件; 如果任务有异常退出, 可以查看计算目录文件下的 output. JOBID来查看具体的原因。

## 4.6 计费与充值

"太乙"目前实施充值收费政策。相关内容链接:

<http://hpc.sustech.edu.cn/newsdetail/recharge.html>

## 4.7 查询机时及费用

科学与工程计算中心自研开发了《南科大高性能集群用户管理系统》, 可使用该系统进行查询: <http://172.18.6.195:18090/>

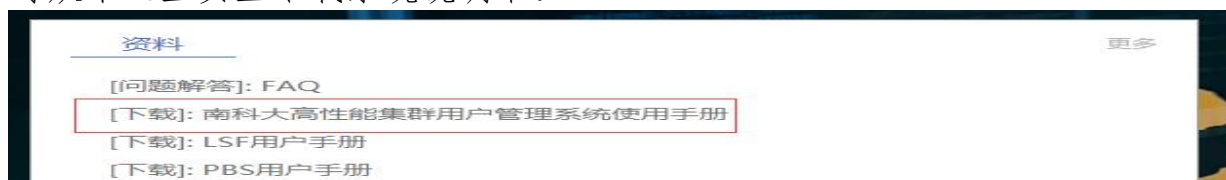
该系统使用说明书:

[http://hpc.sustech.edu.cn/ref/HPMS\\_UserGuide.pdf](http://hpc.sustech.edu.cn/ref/HPMS_UserGuide.pdf)

可从中心主页上进行登陆系统：



可从中心主页上下载系统说明书：



## 4.8 问题反馈

用户对集群使用的相关问题，[可以发邮件至hpc@sustech.edu.cn](mailto:hpc@sustech.edu.cn)，或者在QQ群里进行提问（QQ群号码：161195492）。另外经过对问题的总结和提炼，整理出来FAQ，建议用户先从FAQ里面寻找答案。

FAQ链接：<http://hpc.sustech.edu.cn/userguide.html>

主页链接：

