北大集群普通用户操作手册

浪潮高性能实施工程师

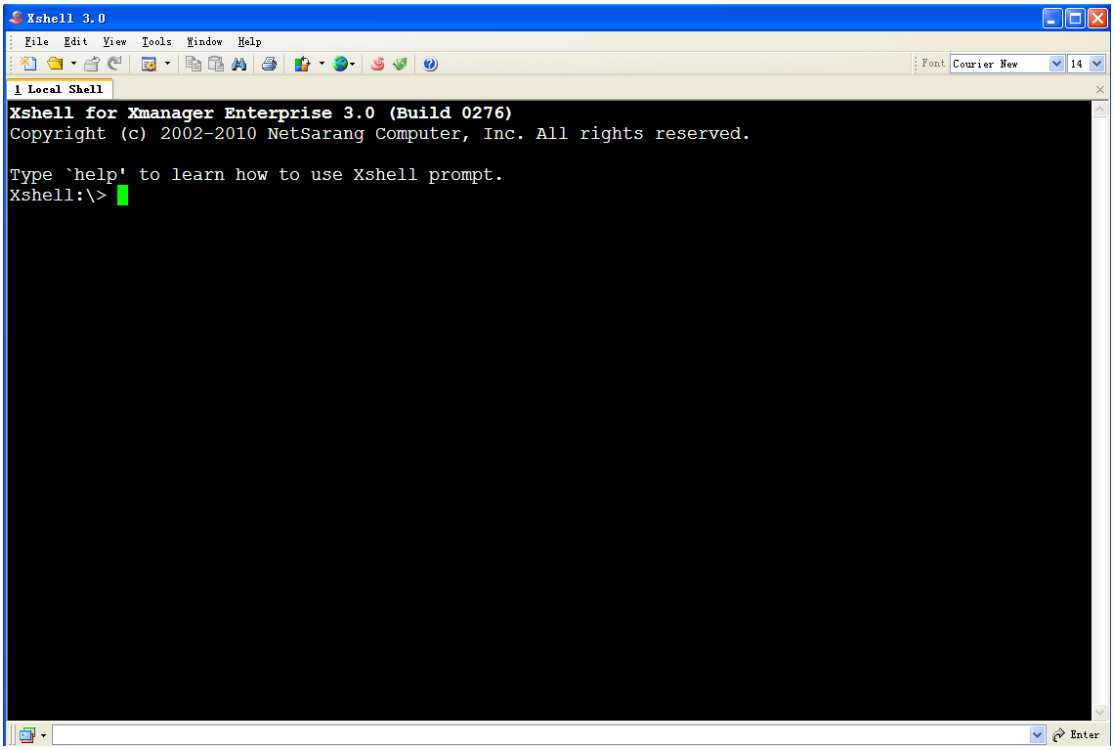
孙玉超：15953100795

1. 登陆作业提交节点

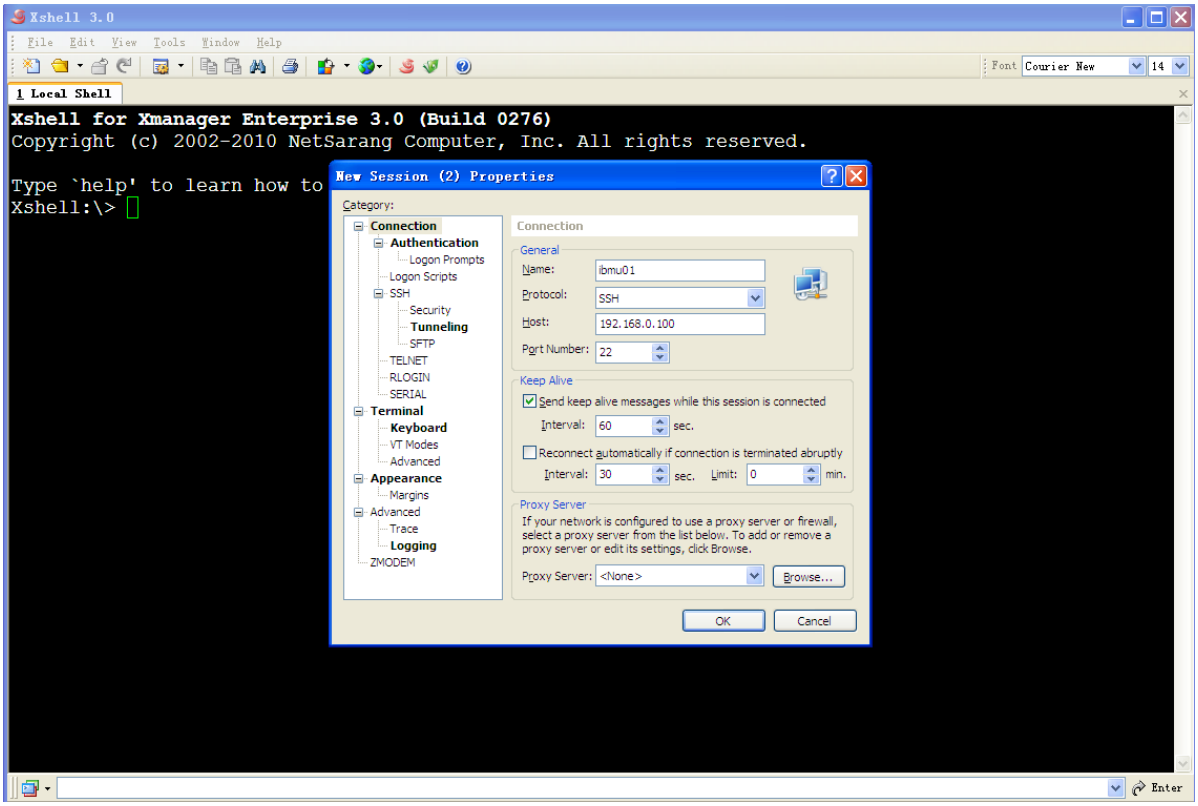
本套集群共设定了 2 个作业提交节点，sn 和 jn
Sn IP 为：162.105.13.252
Jn IP 为：162.105.13.250

大家根据情况随便登陆此 2 节点都行，最好管理员根据实际使用情况，指定用户登陆哪个节点，以使负载平均开。

请使用 ssh 工具登陆，以下示例为 xmanger 工具。

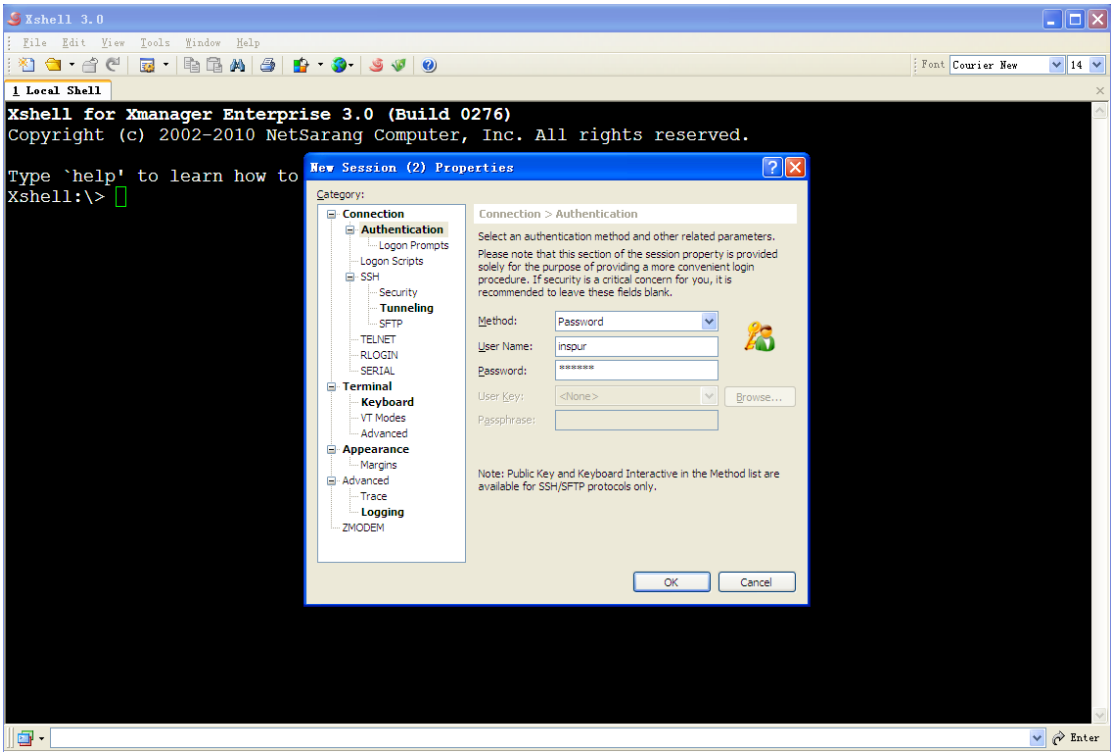


添加主机
点击 file 里的 new 选项

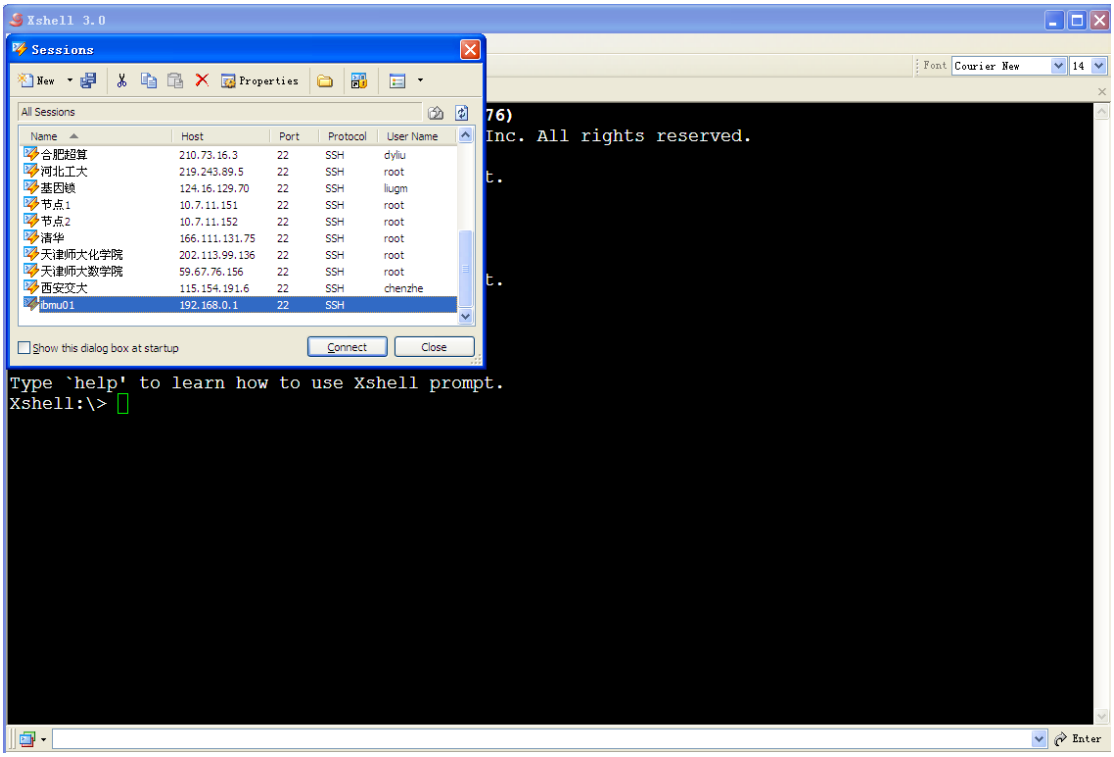


在此界面下，Connection 选项里，Name 里随便填写一个名字用来识别你所添加的机器即可，Host 选项填写远程主机的 IP 地址

然后点击 Authentication 选项，如下图



此选项里，user Name 填写登陆用户名，password 填写登陆密码，填完后点击 OK，添加主机完毕



直接点解 connect 即可连上远程主机的 shell 里

以后连接主机，直接点击 open 选项里所保存的主机即可直接登陆，无需再输入用户名和密码。

2，加载环境变量

登陆进来后若需要使用 module 命令加载编译器，mpi 等的环境变量
请在~/.bashrc 里加入以下 2 行，以改变 module 的模版路径，放到共享目录上
module unuse /usr/share/Modules/modulefiles
module use /home/用户名/modules

Module 的环境变量模版，请根据自己需要在/home/用户名/modules 里自行设定
[inspur@polaris ~]$ module avali（请使用此命令查看已经设置好的环境变量）
如 :

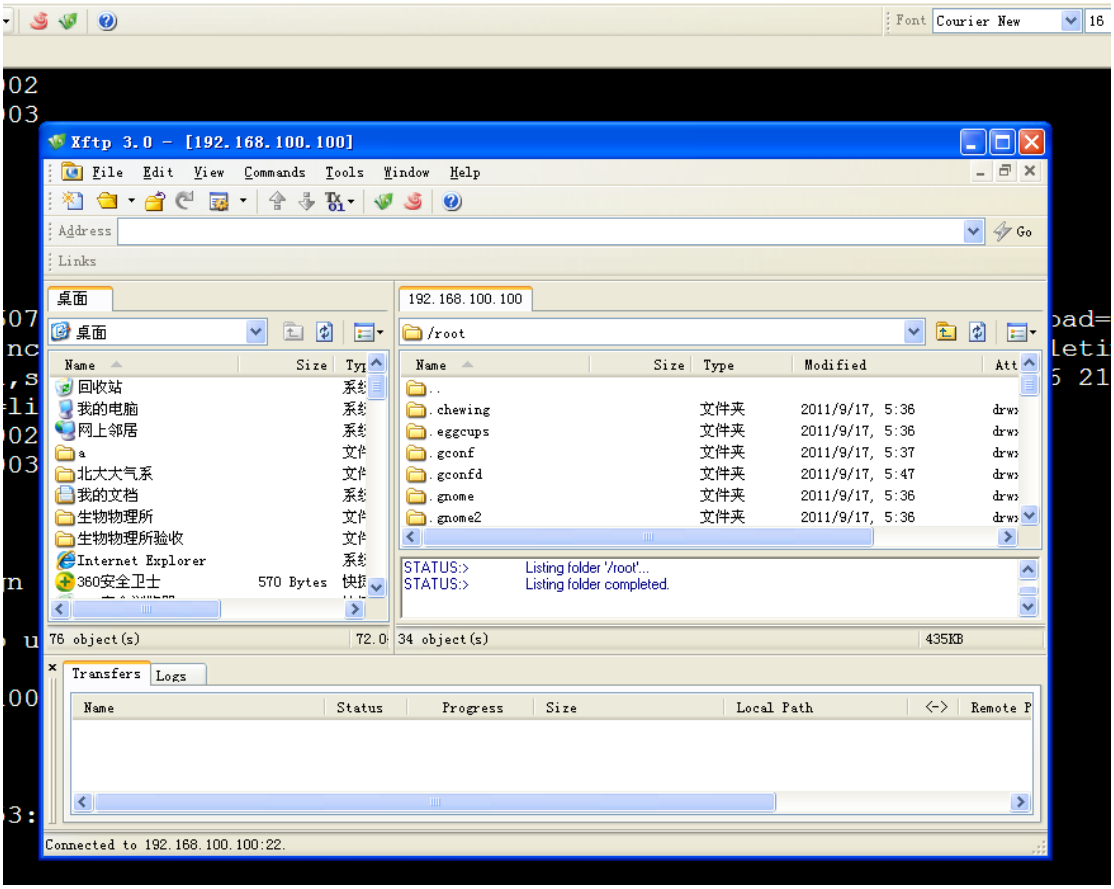
```
[inspur@polaris ~]$ module avali

------------------------------------------------------------ /lustre/inspur/modules -------------------
impi-intel    openmpi-intel openmpi-pgi
```

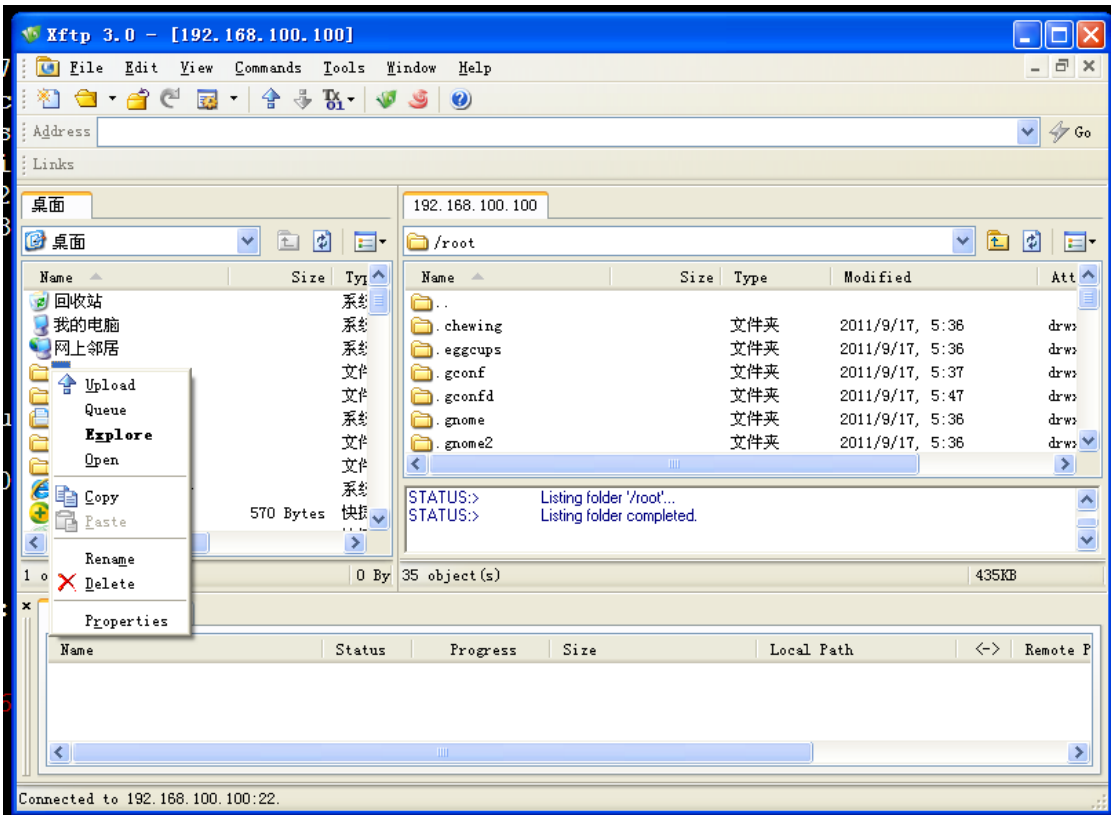
如要加载 impi-intel 的环境变量，请运行
[inspur@polaris ~]$ module load impi-intel
测试环境变量是否生效（以 intel mpicc 为例）
[inspur@polaris ~]$ which mpicc
/lustre/intel/impi/4.0.0.028/intel64/bin/mpicc

3 上传文件到集群

点击绿色的 new file transfer 按钮打开 xftp 工具



右键单击需要上传的文件或者文件夹，选择 upload 即可把文件上传到用户家目录下。



注意，若是比较大的文件，请上传到自己目录的 work 目录下，work 目录软连接到/lustre 目录下。以后计算或者保存计算结果，把相关文件都保存到 work 目录下，本身 home 目录下空间太小，

4. 进行计算

作业运算注意：

1，所有作业的运算都要通过作业调度系统来提交，不准绕过作业调度系统直接 mpirun 运行。

2. 若作业需要先调试测试，请用 pnodes 命令查看目前节点的状态，然后 ssh 登陆 free 节点进行程序调试，禁止在 2 个作业登陆节点上进行任何程序的计算。

```
[root@m1 ~]# pnodes
free nodes: 17
============================
executing job nodes: 0
============================
down nodes: 1
============================
offline nodes: 0
============================
============================
 Down:     node12
```

Free: node1 node2 node3 node4 node5 node6 node7 node8 node9 node10 node11 node13 node14 node15 node16 node17 node18

Full:

Offline:

作业调试修改完毕后，修改 pbs 脚本，用作业调度系统来提交作业即可。
请修改合适 PBS 脚本，示例如下。

#PBS -N test （作业的名字）
#PBS –e /home/inspur/error ###错误输出路径
#PBS –o /home/inspur/work/stand ####标准结果输出路径
#PBS -l nodes=2:ppn=12 （定义启动 2 个节点，每个节点启动 12 个核）
#PBS -l walltime=120:00:00（作业运行最大时间 12 个小时，估算自己的作业最大运行时间，这里指定一下。）
#PBS -q batch （作业提交到 batch 队列）
#PBS –V （定义环境变量范围）
#PBS -S /bin/bash （使用 bash)
############指定环境变量###########
source /opt/intel/Compiler/11.1/072/bin/iccvars.sh intel64(定义 intel MPI 环境变量）
source /opt/intel/Compiler/11.1/072/bin/ifortvars.sh intel64
source /opt/intel/impi/4.0.0.028/bin64/mpivars.sh
source /opt/intel/Compiler/11.1/072/mkl/tools/environment/mklvarsem64t.sh
cd $PBS_O_WORKDIR （进入到你的工作目录，当前作业提交的目录）
EXEC=/home/inspur/test/test.exe
##############脚本部分############
NP=`cat $PBS_NODEFILE | wc -l` ##########计算 cpu 使用核数
NN=`cat $PBS_NODEFILE | sort | uniq | tee /tmp/nodes.$$ | wc -l`##########计算使用节点数
cat $PBS_NODEFILE > /tmp/nodefile.$$
mpirun -genv I_MPI_DEVICE rdma -machinefile /tmp/nodefile.$$ -n $NP $EXEC
rm –rf /tmp/nodefile.$$

注意：请把计算结果路径指向 work 目录
Pbs 作业脚本里几个变量值
例如，定义#pbs –l nodes=2.ppn=4,那么 PBS_NODEFILE 内容就是
cu06
cu06
Cu06
cu06
cu05
cu05

```
cu05
cu05
NP=`cat $PBS_NODEFILE | wc -l`
echo $NP
8
NN=`cat $PBS_NODEFILE | sort | uniq | tee /tmpnodes.$$ | wc -l`
echo $NN
2
```

5. 修改普通用户密码

管理员新建用户后初始密码为 111111，请用户妥善更改自己用户的密码，更改密码命令为
[test@polaris ~]$yppasswd #####使用 nis 提供用户管理。

6. PBS 命令

PBS 提供 3 条命令用于作业管理。

qsub 命令

—用于提交作业脚本

命令格式：

```
qsub  [-a date_time] [-c interval] [-C directive_prefix]
      [-e path] [-I] [-j join] [-k keep] [-l resource_list]  [-m
mail_options]
      [-M user_list][-N name] [-o path] [-p priority] [-q destination] [-r
c]
      [-S path_list] [-u user_list][-v variable_list] [-V]
      [-W additional_attributes] [-z]
      [script]
```

参数说明：因为所采用的选项一般放在 pbs 脚本中提交，所以具体见 PBS 脚本选项。

例：# qsub aaa.pbs 提交某作业，系统将产生一个作业号

qstat 命令

用于查询作业状态信息

命令格式：qatat [-f][-a][-i] [-n][-s] [-R] [-Q][-q][-B][-u]

参数说明：

-f jobid 列出指定作业的信息

-a 列出系统所有作业

-i 列出不在运行的作业

-n 列出分配给此作业的结点

-s 列出队列管理员与 scheduler 所提供的建议

-R 列出磁盘预留信息

-Q 操作符是 destination id，指明请求的是队列状态

-q 列出队列状态，并以 alternative 形式显示

-au userid 列出指定用户的所有作业

-B 列出 PBS Server 信息

-r 列出所有正在运行的作业

-Qf queue 列出指定队列的信息

-u 若操作符为作业号，则列出其状态。

若操作符为 destination id，则列出运行在其上的属于 user_list 中用户的作业状态。

例：# qstat -f 211 查询作业号为 211 的作业的具体信息。

qdel 命令

用于删除已提交的作业

命令格式：qdel [-W 间隔时间] 作业号

命令行参数：

例：# qdel -W 15 211 15 秒后删除作业号为 211 的作业

7. 统计节点状态信息

用户在登录节点上可以使用如下命令查看计算节点的状态和信息。

pnodes 命令

```
[root@m1 ~]# sh pnodes
free nodes: 17
==========================
executing job nodes: 0
==========================
down nodes: 1
==========================
offline nodes: 0
==========================
==========================
 Down:     node12
 Free:     node1 node2 node3 node4 node5 node6 node7 node8 node9 node10 node11 node13
node14 node15 node16 node17 node18
 Full:
 Offline:
```

Pbsnodes 命令

查看节点的概况信息。

```
[root@polaris ~]# pbsnodes -l all
c01b01                  free
c01b02                  free
c01b03                  free
.....................................
c18b09                  job-exclusive
c18b10                  job-exclusive

[root@polaris ~]# pbsnodes  -l down
c07b05                  down

[root@polaris ~]# pbsnodes -l free
c01b01                  free
c01b02                  free
c01b03                  free
```

关于个人用户 home 目录存放数据的问题。
集群整个 home 目录是共享的存储一部分空间，大小为 2.6T 左右。
所以空间有限，大数据不要直接放到家目录下，每一个家目录下都由管理员建立了一个 work 目录，此 work 目录指向/lustre 目录，为并行文件系统，空间大约为 46T 左右，在计算的过程中，也需要把计算结果也指向 work 目录。
对于占用空间不大，但很重要的数据，可以直接放到家目录下，work 目录不如家目录安全。