# 杭州电子科技大学高性能计算集群使用培训

2024-03-13

#### 中科曙光-坚持自主创新、服务中国的历史使命



中国计算机"玻璃房子"之耻 (1976年)

中国人的"争气机"-曙光1号 (1993年)



1976年: 美国总统福特确实批准了两台CDC公

司的超级计算机出口中国。

购买方: 中石油 "东方地球物理公司", BGP

条件:

第一: 美国出口中国的超级计算机性能必须是

大幅下降过的。

第二:美国的超算,只可以用于地质勘探,不

可以用于其他领域。

**第三**: 机房的钥匙由美国人掌握,每次使用,都必须向美国人报批,美国人同意以后才可以用。并且使用的时候要有美国人在边上监督。



李国杰 (身穿深色衬衣者) "曙光一号"研制组人员

#### 曙光5000A (2008年)



每秒230万亿次浮点运算 首次突破百万亿次



每秒6.4亿次浮点运算 **1995年曙光公司成立** 中国HPC走向市场化

#### 曙光星云 (2010年)



·中国首台实测性能超千万亿次 高性能计算机 每秒1270万亿次浮点运算

#### 曙光4000A (2004年)



每秒10万亿次浮点运算 继美、日之后的第三台

#### 曙光硅立方 (2016年)



下一代"E级"超级 每秒超亿万次浮点运算不多 全浸没式液冷计算机

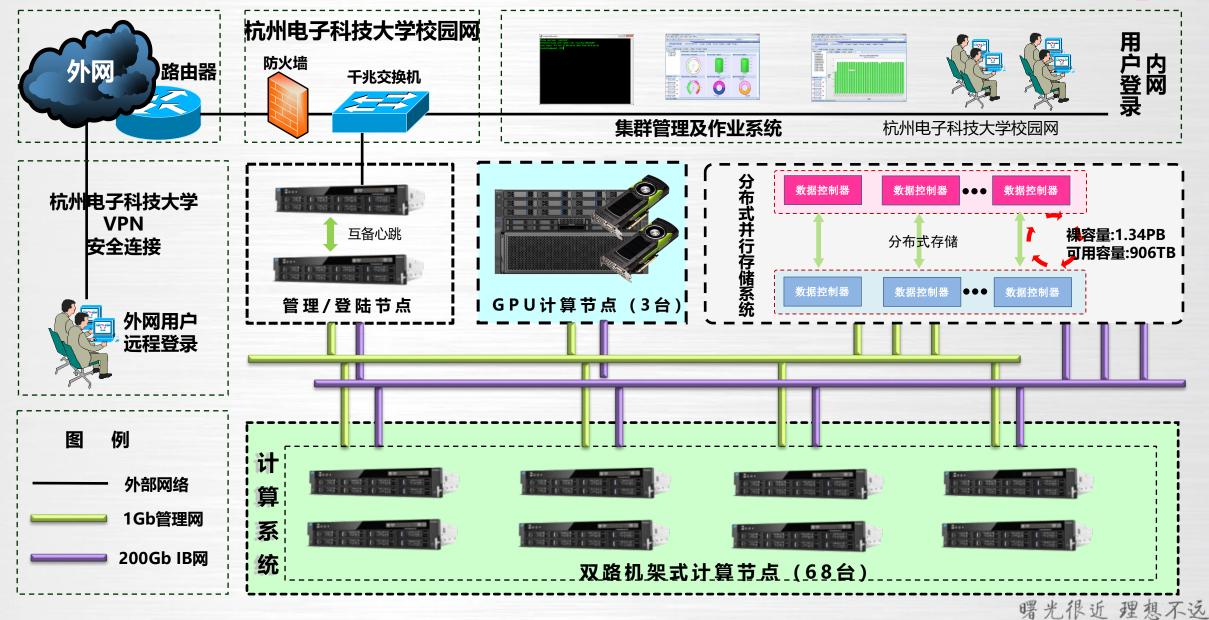




- 01 集群系统设置
- 02 集群用户登录
- 03 软件部署情况
- 94 集群使用

#### 系统拓扑连接图





### 集群系统设置



- 平台配置管理节点2台,分别配置2颗Intel(R) Xeon(R) Silver 4314 CPU @ 2.40GHz处理器,管理节点一(tc6000)是用户登入集群的门户;管理节点二(tc6001),部署宁畅SIMS 5.1集群管理系统,用户可通过IP地址实现Web访问;
- 平台配置CPU计算系统有60台,配置2颗Intel(R) Xeon(R) Platinum 8358P CPU @ 2.60GHz 32核处理器,256GB 内存; 节点号为node1~node60;
- 平台配置GPU计算系统有5台,节点号node191~node193配置2颗Intel(R) Xeon(R) Platinum 8358P CPU @ 2.60GHz 32核处理器,1TB内存;每台节点配置了8块NVIDIA NVLink A100 40GB GPU卡;节点号 node194~node195配置2颗Intel(R) Xeon(R) Platinum 6248R CPU @ 3GHz 24核处理器,256GB内存;每台节点配置了8块NVIDIA NVLink RTX 4090 24GB GPU卡。
- 平台配置一套**曙光Parastor**分布式存储系统,由**5**台ostor节点组成,**采用4+2:1容错机制**,可提供约**906TB** 的存储空间,挂载至每个节点/public目录,供整个集群使用。
- 所有节点使用2段网络进行连接,其中监控管理与IPMI管理网采用1Gb以太网互连,计算与存储网采用200Gb HDR Infiniband互连;
- 平台采用CentOS Linux release 7.8.2003版本操作系统,配置intel\GNU的C与fortran编译器、MPI并行环境、Python、java、数学库等软件。

# 集群系统设置-管理、计算、存储资源



计算子系统	节点名	节点描述	节点数	作业队列
管理节点1	TC6000	登录方式: Shell	1	-
管理节点2	TC6001	登录方式: Shell\WEB	1	-
CPU计算节点	node1-68	2颗Intel 8358P 2.60GHz 32核处理器, 256GB内存	68	normal
GPU计算节点	Node191-195	2颗Intel 8358P 2.60GHz 32核处理器, 1TB内存, 8块NVLink A100 40GB GPU卡	5	gpu
存储		/public共享目录,所有节点均共同挂载		

### 集群系统设置-软件部署路径约束



- 用户家目录/public/home/username,用户只对属于自己的家目录具有完全的读写权限,用户间家目录禁止互相访问,私有软件及个性化环境变量,可以部署在自己的家目录下。
- 软件安装目录/public/software/, 集群共享软件部署路径。

注:安装在/public/software/下在的软件以软件名加版本号来区分。

■ 环境变量目录/public/software/profile.d/集群中所有部署软件的环境变量文件均放置在该目录下,可按需调用。

注:每个软件对应1个环境变量配置文件,变量文件名以软件名加版本号来区分。 普通用户自已安装在自己家目录中的软件,环境变量可以写在自已家目录中的.bashrc文件中。

■ 软件源码目录/public/sourcecode, 存放安装程序源代码





- 01 集群系统设置
- 02 集群用户登录
- 03 软件部署情况
- 04 集群使用

### 集群系统登录方式



WEB浏览器登录地址: <a href="https://manage-sc.hdu.edu.cn/sso/login">https://manage-sc.hdu.edu.cn/sso/login</a>

作业提交、计算请使用普通账号, root账号仅用于服务器配置、软件安装使用。

登录工具: putty命令行登录工具 , (推荐Xmanager Enterprise)

VNC图形桌面登录: VNC Viewer;

Linux to Windows文件传输工具: winscp (推荐xftp)

Linux to Linux文件传输: scp

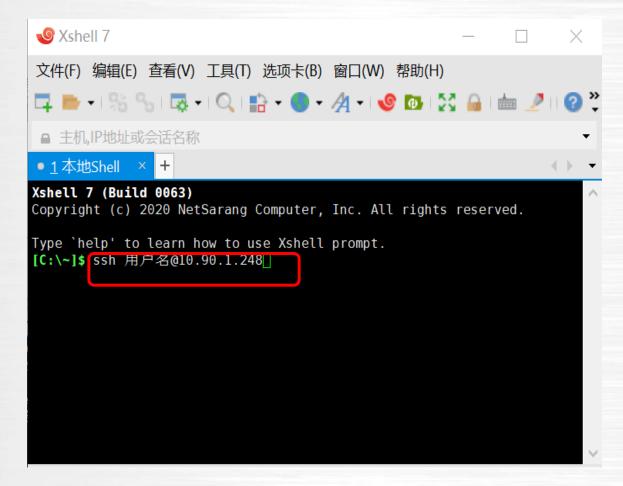
# 集群登录方式一: 命令行终端

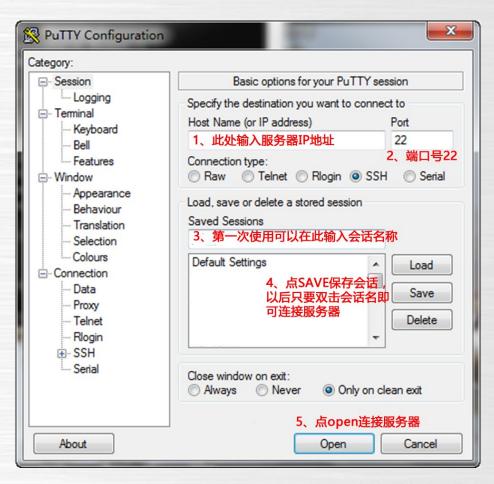


□SSH Secure Shell Client, PuTTY, SecureCRT、xshell等SSH客户端软件登录。

□Linux客户端可以直接在命令行终端中执行ssh命令进行登录:

\$>ssh username@登录节点IP地址





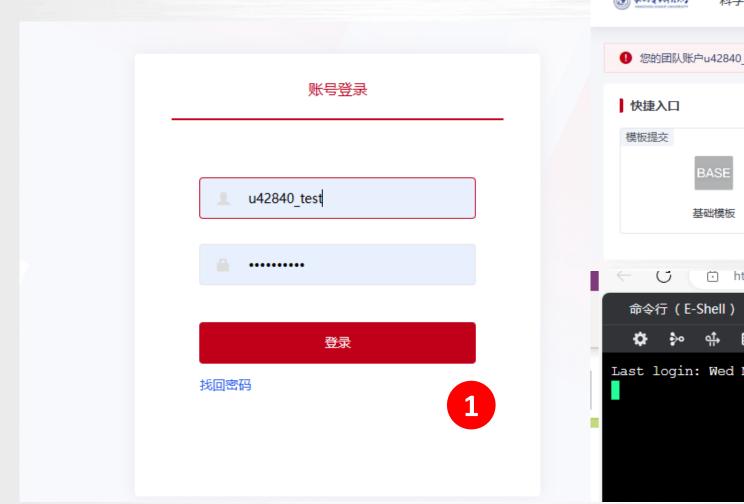
曙光很近 理想不远

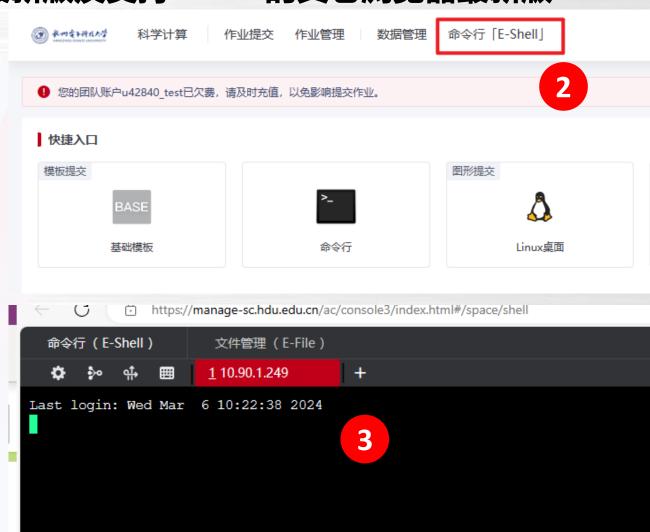
## 集群登录方式二: SIMS WEB登录



Web登录地址: https://manage-sc.hdu.edu.cn/sso/login

兼容浏览器: chrome\firefox\edge最新版及支持html5的其它浏览器最新版

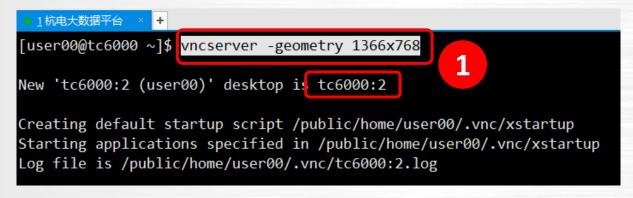




### 集群登录方式三: VNC桌面登录



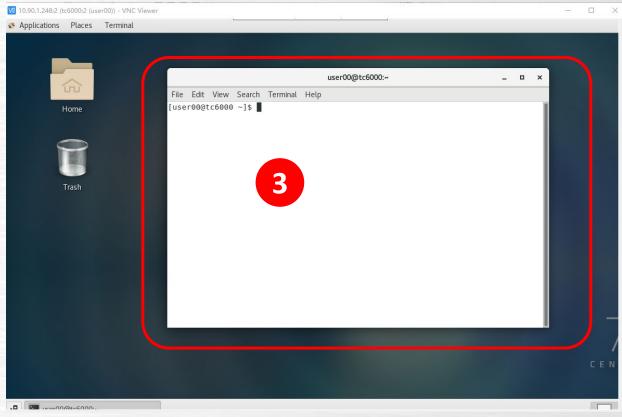
Shell登录集群,执行vncserver-geometry 1366x768,即开启一个分辨率为1366x768的VNC桌面(注意桌面ID)。



#### 在VNC Viewer客户端中输入IP:ID, 登录VNC桌面



#### 在VNC桌面桌面中可以使用终端或图形化执行任务。



#### VNC工具使用



#### 启用VNC步骤:

- 1、以自己的用户名命令行登录系统
- 2、输入vncserver (注意,第一次运行会提示设置密码) 运行完后,注意生成的ID号,然后即可以用VNC viewer客户端工具连接。

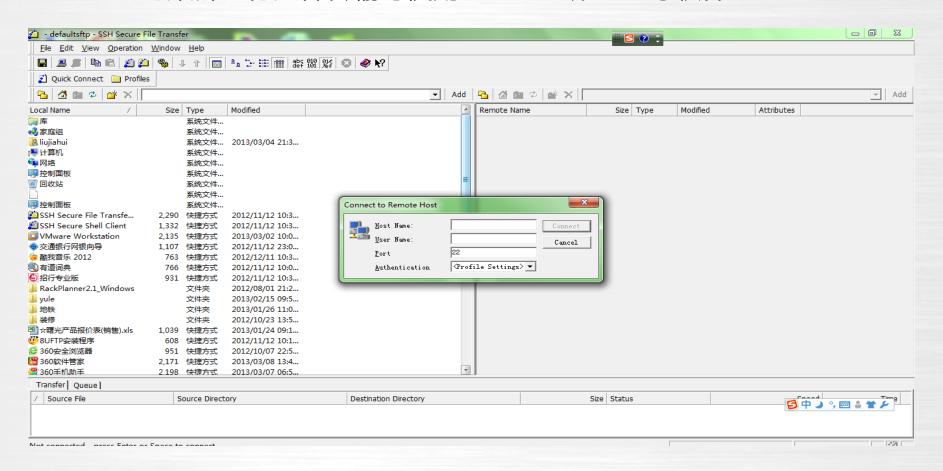
#### 高级操作:

- 1、vncserver -kill:ID (杀掉某个ID的VNC桌面)
- 2、示例使用: vncserver -geometry 1920x1080 24:5
- 指定分辨率1920x1080,颜色位深度为24,连接端口ID为5;
- 3、改图形化登录密码: vncpasswd
- 注:为了安全,建议图形VNC登录密码不要与命令行登录密码设置为同一个。

### 上传下载数据: WinSCP简介



■ Windows与集群进行文件传输可使用WinSCP或SFTP等软件



■ Linux操作系统间复制文件,直接使用命令 scp filename test@ip:/home/test





- 01 集群系统设置
- 02 集群用户登录
- 03 软件部署情况
- 04 集群使用

# 集群系统-基础软件部署



软件	安装目录
编译器	Intel2021、GCC-7.3.1、Cmake-3.20.1; /public/software/compiler/
数学库	Fftw-3.3.9、Gsl-2.6、Hdf5-1.12.0、Jasper-1.900.1、Lapack-3.9.1、Netcdf-4.7.4、Petsc-3.15.0; /public/software/mathlib/
MPI	Intelmpi-2021.3.0、Openmpi-4.0.3; /public/software/mpi/intelmpi/2021.3.0
anaconda3	/public/software/apps/anaconda3/2021.0
ansys	/public/software/apps/ansys_inc/v211
OpenFOAM6	/public/software/apps/OpenFOAM-6/

# 编译器



软件名称		软件信息
	软件版本	4.8.5
	安装路径	/usr
GNU编译器	调用方式	C: gcc C++: g++ F77: g77 F90: gfortran
	软件版本	2021
	安装路径	/public/software/compiler/intel-compiler/2021.3.0/
Intel编译器	环境变量配置文件	/public/software/profile.d/compiler_intel-compiler-2021.3.0.sh
	调用方式	C: icc C++: icpc F77: ifort F90: ifort





软件名称	软件信息									
	说明	OpenMPI,与Intel编译器关联,支持以太网和InfiniBand。								
	软件版本	openmpi-4.0.3								
On an MDI	安装路径	/public/software/mpi/openmpi/intel/4.0.3								
OpenMPI (与Intel编	环境变量配置文件	/public/software/profile.d/mpi_openmpi-intel-4.0.3.sh								
译器关联)	调用方式	C: mpicc C++: mpicxx F77: mpif77 F90: mpif90 mpirun -np <n> -machinefile <machinefile></machinefile></n>								

# 数学库



软件名称	软件信息								
	说明	单精度版FFTW3数学库							
FFTW3-	软件版本	fftw-3.3.9							
float	安装路径	/public/software/mathlib/fftw/intelmpi/3.3.9_single/							
	调用方式	<编译器> -I/public/software/profile.d/mathlib_fftw-intelmpi- 3.3.9_single.sh							
	说明	双精度版FFTW3数学库							
FFTW3- double	软件版本	fftw-3.3.9							
double	安装路径	/public/software/mathlib/fftw/intelmpi/3.3.9_double/							
	调用方式	<编译器> -I/public/software/profile.d/mathlib_fftw-intelmpi- 3.3.9_double.sh							

# 数学库



软件名称		软件信息			
	说明	Intel MKL, 包含BLAS、LAPACK、FFT、ScaLAPACK、BLACS等			
	软件版本	Intel Compiler 2021自带版本			
	安装路径	/public/software/compiler/intel-compiler/2021.3.0/MKL			
Intel MKL	环境变量配置文件	/public/software/profile.d/compiler_intel-compiler- 2021.3.0.sh			
	调用方式	ifort/icc –L\${MKLROOT}/lib/em64t -lmkl_intel_lp64 - lmkl_sequential -lmkl_core			

### anaconda3



软件名称	软件信息							
	python	3.8						
anaconda3	软件环境	# conda environments: base * /public/software/apps/anaconda3/2021.05 pytorch-cpu /public/software/apps/anaconda3/2021.05/envs/pytorch-cpu tf-cpu /public/software/apps/anaconda3/2021.05/envs/tf- cpu tf-gpu /public/software/apps/anaconda3/2021.05/envs/tf- gpu						
	安装路径	/public/software/apps/anaconda3						
	调用软件方法	conda activate tf-gpu						





- 01 集群系统设置
- 02 集群用户登录
- 03 软件部署情况
- 04 集群使用

### 集群环境变量文件路径



安装软件的环境变量设置脚本路径(/public/software/profile.d)

```
[root@tc6001 profile.d]# pwd
/public/software/profile.d
[root@tc6001 profile.d]# ls
                                       mathlib fftw-intelmpi-3.3.9 double.sh
apps anaconda3-2021.05.sh
apps ansys194.sh
                                       mathlib_fftw-intelmpi-3.3.9_single.sh
apps_ansys211.sh
                                       mathlib_gsl-gnu-2.6.sh
apps openfoam6.sh
                                       mathlib gsl-intel-2.6.sh
benchmark hpl-intelmpi-2.2.sh
                                       mathlib hdf5-gnu-1.12.0.sh
benchmark htop-gnu-2.1.0.sh
                                       mathlib hdf5-intel-1.12.0.sh
benchmark iozone-gnu-3.430.sh
                                       mathlib netcdf-gnu-4.7.4.sh
compiler cmake-compiler-3.20.1.sh
                                       mathlib netcdf-intel-4.7.4.sh
                                       mpi intelmpi-2021.3.0.sh
compiler gcc-7.3.1.sh
compiler intel-compiler-2017u8-env.sh
                                       mpi openmpi-4.1.2-intel2021-env.sh
compiler_intel-compiler-2021.3.0.sh
                                       mpi openmpi-gnu-3.1.3.sh
cuda-10.1-env.sh
                                       mpi openmpi-gnu-4.0.3.sh
                                       mpi openmpi-intel-4.0.3.sh
cuda-11.4.3-env.sh
[root@tc6001 profile.d]#
```

环境变量文件中记录了软件安装的路径等信息,通过加载不同软件的环境变量,操作系统才可以识别并使用相关软件。

#### 用户自定义调用默认环境变量



• 各用户在.bashrc中选择需要的进行设置

### 临时调用所需的软件环境变量



#### 使用参考:

source /public/software/profile.d/mpi\_openmpi-intel-4.0.3.sh ##openmpi 环境变量激活

### 运行计算程序



■运行一个串行程序a.out 需要注意该程序是否为单进程多线程程序,即openmp并行程序,如高斯软件

[test@tc6000 test]\$ ./a.out

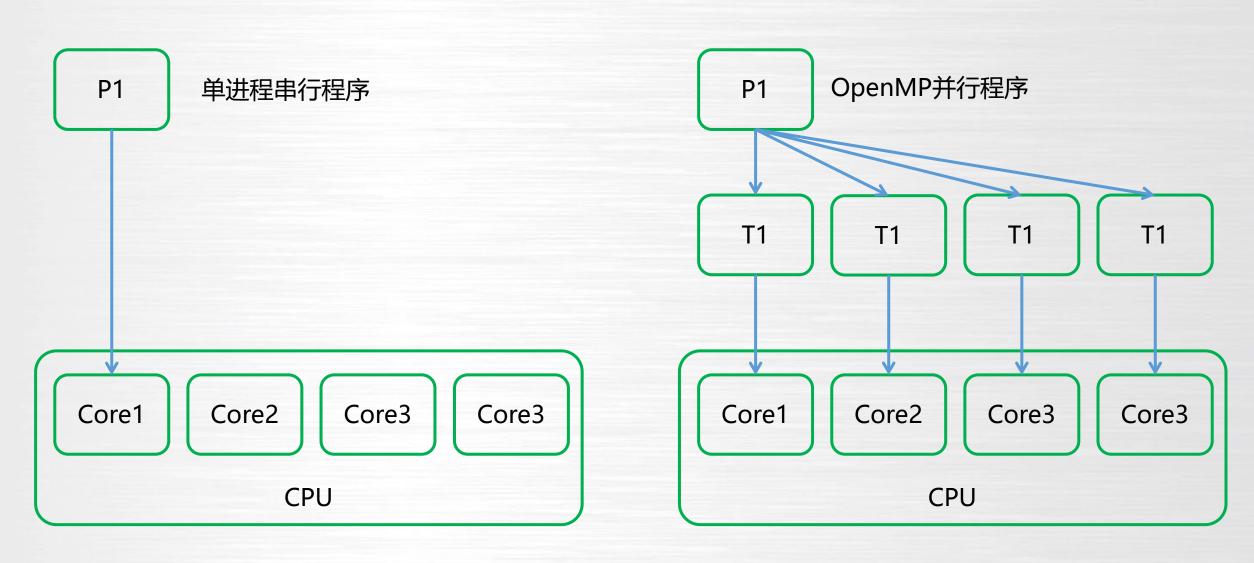
■运行一个标准的mpi并行程序a.out

[test@tc6000 test]\$ mpirun -np 180 -machinefile ./nodelist ./a.out

■运行一个商业版并行程序fluent

[test@tc6000 test]\$ fluent 2d -t240 -p -cnf=./nodelist -ssh -mpi=intel -g -i Input >out.log

# 注:单进程串行程序和单进程多线程程序运行特点 Sugish



#### 注: 单进程串行程序运行特点



```
top - 11:42:30 up 147 days, 2:14, 1 user, load average: 28.81, 28.73, 28.76
Tasks: 516 total, 29 running, 487 sleeping,
                                               0 stopped,
                                                            0 zombie
%Cpu(s): 99.6 us, 0.4 sy, 0.0 ni, 0.0 id, 0.0 wa, 0.0 hi, 0.0 si, 0.0 st
KiB Mem : 39468774+total, 36435603+free, 20222688 used, 10108992 buff/cache
KiB Swap: 16777212 total, 16400576 free, 376636 used. 36916816+avail Mem
   PID USER
                 PR NI
                           VIRT
                                   RES
                                          SHR S %CPU %MEM
                                                               TIME+ COMMAND
 97802 jingq LJ
                20
                         620604 380932
                                         7136 R 100.0 0.1 40:59.02 SUBSPACE PB DIA
 99185 jingg LJ 20
                         459884 271024
                                         7116 R 100.0 0.1 33:45.90 SUBSPACE PB DIA
 99994 jingq LJ 20
                         453884 234624
                      0
                                         7072 R 100.0
                                                       0.1 30:07.19 SUBSPACE PB DIA
101082 jingq LJ 20
                         479264 252884
                                         8632 R 100.0
                                                      0.1 22:16.81 SUBSPACE PB DIA
101310 jingq LJ 20
                      0
                         453744 235036
                                         8880 R 100.0 0.1 20:35.79 SUBSPACE PB DIA
101415 jingq LJ 20
                         497548 273504
                                         8808 R 100.0 0.1 20:31.47 SUBSPACE PB DIA
                      0
102706 jingq LJ 20
                         382800 169512
                                         8800 R 106
                                                                           8
                                                                                                 15
                                                                                                              100.0%
                                                                                                                        22
102934 jingq LJ 20
                                         8556 R 106
                         467872 244968
                                                                                                                        23
                                                                 100.0%
                                                                           9
                                                                                                 16
                                                                                                              100.0%
                                                                                                                                     100.0%
 97252 jinga LJ 20
                         495388 234168
                                         7192 R 99
                                                                 100.0%
                                                                                                17
                                                                                                             100.0%]
                                                                           10
                                                                                                                        24
                                                                                                                                     100.0%
 98362 jingg LJ 20
                         581924 353736
                                         7152 R 99
                      0
                                                                                                 18
                                                                 100.0%
                                                                           11
                                                                                                              100.0%
                                                                                                                        25 T
                                                                                                                                     100.0%
                                         7008 R 99 5
 99307 jingq LJ 20
                        554260 319788
                                                                 100.0%
                                                                           12
                                                                                                 19
                                                                                                                       26
                                                                                        100.0%
                                                                                                              100.0%
                                                                                                                                     100.0%
100239 jingg LJ 20
                         494168 267944
                                                 99 6
                                         6868 R
                                                                           13
                                                                                                 20
                                                                                                                       27
                                                                 100.0%
                                                                                        100.0%
                                                                                                              100.0%
                                                                                                                                     100.0%
100938 jingq LJ 20
                      0
                        346816 155136
                                         8556 R 99 7
                                                                                                 21
                                                                           14
                                                                                        100.0%
                                                                                                            | | 100.0%
                                                                                                                        28
                                                                                                                                     100.0%
101573 jingg LJ 20
                         441012 223792
                                         8796 R 99 Mem[
                                                                                                 Tasks: 127, 56 thr; 30 running
                                                                                    23.2G/376G
                                                                                                Load average: 29.23 28.86 28.81
                                         8868 R 99 Swp[
101978 jingq LJ 20
                         422520 207260
                                                                                    368M/16.0G
                                                                                                 Uptime: 147 days(!), 02:15:20
102136 jingg LJ 20
                        369516 163264
                                         8804 R 99
102262 jingq LJ 20
                      0
                         422748 205772
                                         8852 R 99
                                                                                           CPU%
                                                                          VIRT
102408 jingg LJ 20
                         451208 227244
                                         8792 R
                                                    02136 jingq LJ
                                                                                165M
                                                                          367M
                                                                                         R 101. D.0 15:02.71
                                                 990938 jingq_LJ
102531 jingg LJ 20
                         474460 246368
                                         8868 R
                                                                 20
                                                                                     8556 R
                                                                         376M
                                                                               173M
                                                                                                 0.0 25:03.23 WSUBSPACE PR DIAG POPO X V
                                                                  20
                                                                       0 665M 427M
                                                   98136 jingq LJ
                                                                                    7152 R 99.9 0.1 40:20.07 ../SUBSPACE PB DIAG POP9 X V
                                                   L00239 jingq LJ
                                                                  20
                                                                         482M
                                                                               261M
                                                                                    6868
                                                                                           99.9 0.1 28:25.15 ../SUBSPACE PB DIAG POP9 X V
                                                                       0 361M
                                                                                    8784 R 99.9 0.0 6:19.74 ../SUBSPACE_PB_DIAG_POP9_X_V
                                                         jingg LJ
                                                                  20
                                                                               158M
                                                   L02828
                                                   99185 jingg LJ
                                                                  20
                                                                       0 570M
                                                                               346M
                                                                                    7116
                                                                                          R 99.9 0.1 34:37.69 ../SUBSPACE PB DIAG POP9 X V
                                                                       0 383M
                                                   L03325
                                                         jingg LJ
                                                                  20
                                                                               180M
                                                                                    9704 R 99.9 0.0 1:41.34 ../SUBSPACE PB DIAG POP9 X V
                                                                       0 443M
                                                                               229M
                                                                                    7072 R 99.9 0.1 30:58.97 ../SUBSPACE PB DIAG POP9 X V
                                                   99994 jingq LJ
                                                                  20
                                                                  20
                                                                       0 370M
                                                                               181M
                                                                                    8544
                                                   L00712 jingq LJ
                                                                                          R 99.9 0.0 25:29.09 ../SUBSPACE PB DIAG POP9 X V
                                                         jingq LJ
                                                                       0 417M
                                                                               207M
                                                                                    8544
                                                   01188
                                                                                          R 99.9 0.1 22:50.26 ../SUBSPACE PB DIAG POP9 X V
                                                   L02706 jingq LJ
                                                                 20
                                                                       0 379M
                                                                               153M
                                                                                    8800
                                                                                          R 99.9 0.0 6:55.53 ../SUBSPACE PB DIAG POP9 X V
                                                                      0 526M 304M 7196 R 99.9 0.1 45:23.57 ../SUBSPACE PB DIAG POP9 X V
                                                   97252 jingg LJ
                                                   Help F2Setup F3SearchF4FilterF5Tree F6SortByF7Nice -F8Nice +F9Kill F10Quit
```

#### 注: 单进程多线程程序运行特点



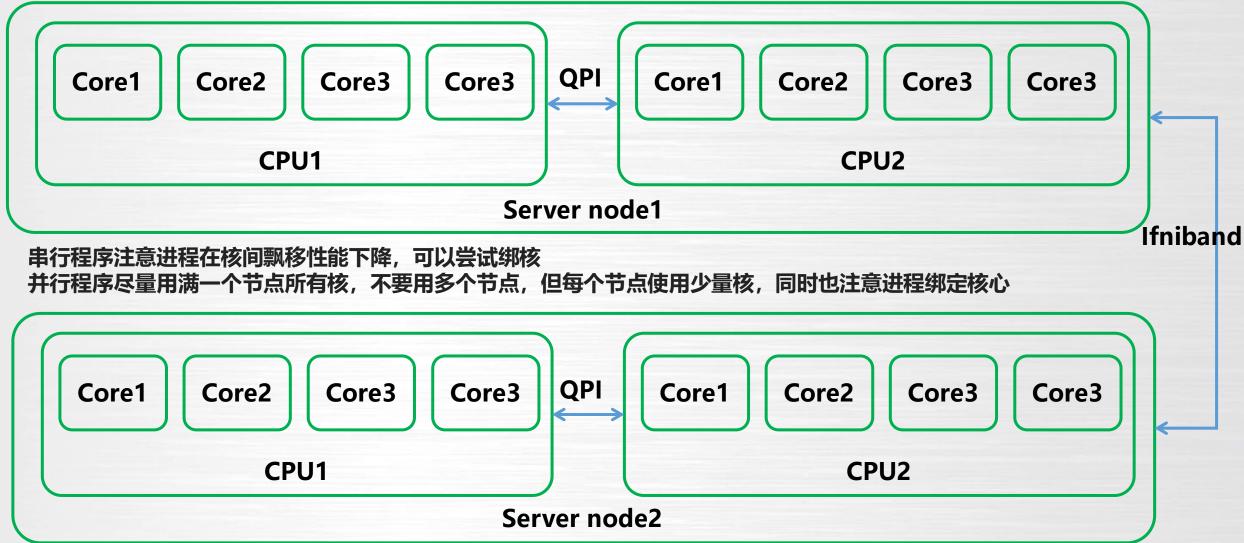
```
top - 11:41:03 up 147 days, 2:05, 1 user, load average: 29.67, 29.15, 28.94
                                             0 stopped, 0 zombie
Tasks: 423 total, 2 running, 421 sleeping,
%Cpu(s): 99.9 us, 0.1 sy, 0.0 ni, 0.0 id, 0.0 wa, 0.0 hi, 0.0 si, 0.0 st
KiB Mem : 13175065+total, 11128040+free, 11066500 used, 9403752 buff/cache
KiB Swap: 15624188 total, 15624188 free,
                                              0 used. 11757093+avail Mem
 PID USER
                                       SHR S %CPU %MEM
               PR NI
                         VIRT
                                 RES
                                                            TIME+ COMMAND
18515 ZCY HG
                                      8916 R 2796 1.9 42140:09 l502.exe
                    0 59.367g 2.442g
               20
1768 root
               20
                    0 851172 51736
                                      1712 S
                                               1.3 0.0
                                                          1878:35 oMgcd
2267 root
                                     11884 S
                                               0.3 0.0 321:28.91 pbs mom
               20
                    0 113328
                               49928
5298 root
                                                          0:00.04 top
               20
                    0 157980
                                2520
                                      1544 R
                                               0.3 0.0
               20
                    0 192892
                                5908
                                       2504 S
   1 root
   2 root
               20
                    0
                                         0 S
                            0
                                   0
                                                                                           15
                                                                                                                 22
               20
                    0
                            0
                                         0 S
                                                 2
                                                             100.0%]
                                                                       9
                                                                                           16
                                                                                                                 23
   3 root
                                                 3
                                                             100.0%
                                                                       10
                                                                                           17
                                                                                                                 24
                0 -20
                                         0 S
   5 root
                            0
                                                                      11
                                                             100.0%
                                                                                           18
                                                                                                                 25
   8 root
                    0
                            0
                                   0
                                         0 S
               rt
                                                 5
                                                                       12
                                                             100.0%
                                                                                   100.0%
                                                                                           19
                                                                                                                 26
                    0
                                   0
                                         0 S
               20
                            0
   9 root
                                                                                           20
                                                             100.0%
                                                                       13
                                                                                   100.0%
                                                                                                                 27
                                                 6
  10 root
               20
                    0
                            0
                                   0
                                         0 S
                                                                                                                 28
                                                             100.0%
                                                                       14
                                                                                   100.0%]
                                                                                           21
  11 root
               rt
                            0
                                         0 S
                                                 Mem
                                                                                           Tasks: 44, 494 thr; 32 running
                                                                               12.9G/126G
                                                                                           Load average: 29.39 29.15 28.95
                                                 Swp
                                                                                 0K/14.9G
                                                                                           Uptime: 147 days(!), 02:06:05
```

PID U	JSER	PRI	NI	VIRT	RES	SHR	S	CPU%	MEM%	TIME+	Command
18515 2	CY_HG	20	0	59.4G	2504M	8916	R	2800	1.9	702h	/public/software/g16/g16/l502.exe
2063 2	CY_HG	20	0	59.4G	2504M	8916	R	100.	1.9	10:55.70	/public/software/g16/g16/l502.exe
2089 2	CY_HG	20	0	59.4G	2504M	8916	R	100.	1.9	10:55.71	/public/software/g16/g16/l502.exe
2064 2	CY_HG	20	0	59.4G	2504M	8916	R	100.	1.9	10:54.05	/public/software/g16/g16/l502.exe
2088 2	CY_HG	20	0	59.4G	2504M	8916	R	100.	1.9	10:55.07	/public/software/g16/g16/l502.exe
2065 2	CY_HG	20	0	59.4G	2504M	8916	R	100.	1.9	10:55.59	/public/software/g16/g16/l502.exe
2074 2	CY_HG	20	0	59.4G	2504M	8916	R	100.	1.9	10:55.40	/public/software/g16/g16/l502.exe
2075 2	CY_HG	20	0	59.4G	2504M	8916	R	100.	1.9	10:55.43	/public/software/g16/g16/l502.exe
2080 2	CY_HG	20	0	59.4G	2504M	8916	R	100.	1.9	10:55.52	/public/software/g16/g16/l502.exe
2090 2	CY_HG	20	0	59.4G	2504M	8916	R	100.	1.9	10:55.62	/public/software/g16/g16/l502.exe
2087 2	CY_HG	20	0	59.4G	2504M	8916	R	100.	1.9	10:55.49	/public/software/g16/g16/l502.exe
2066 2		20			2504M			100.		10:55.57	
F1Help	F2Setup	F3Sea	rch	F4Filt	er <mark>F5</mark> Tr	ee F	Sc	rtBv	7Nice	e -F8Nice	+FOKill FIOQuit

#### CPU程序优化运行的一些思路



通信性能: C2C>QPI>Infiniband



串行程序可以执行taskset –c 0 ./a.out将程序绑定在CPU核Core0上运行,并行程序绑定具体参照各类mpi手册如openmpi可以使用:mpirun -np <N > -machinefile <machinefile > --bind-to-core --bycore <executable > 理想不适

### 集群作业管理调度系统



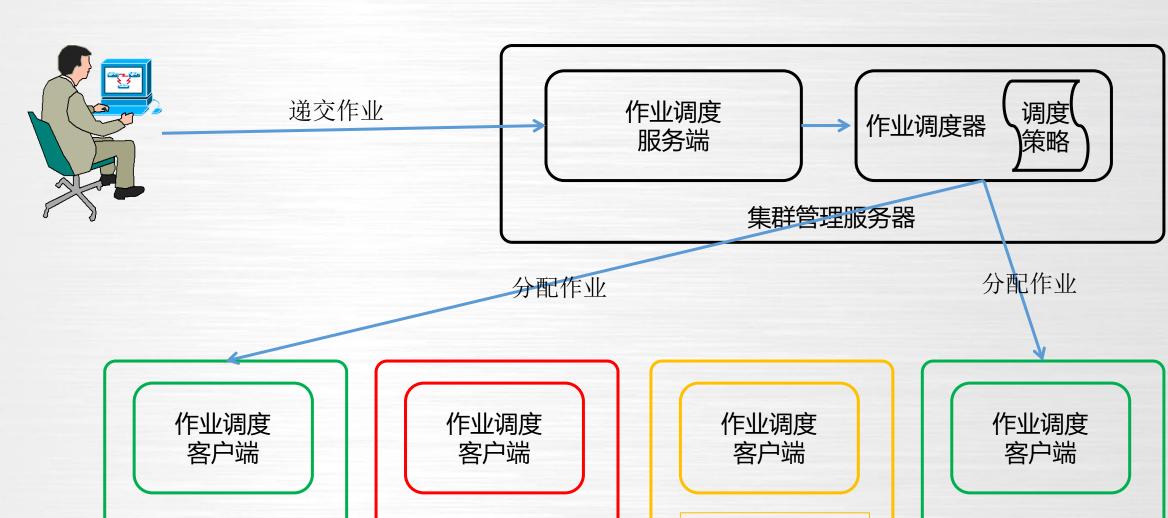
HPC计算集群作为公共的平台,当有多个用途或是多个使用者时,必须引入作业管理调度系统,以便对集群进行资源配额,和任务的自动化管理,集群使用者应避免出现以下三种问题:

- 1、在管理节点上长时间运行计算程序;
- 2、避免绕开调度系统在计算节点运行程序,这样会因为没有统一分配调度资源,导致某些计算节点执行超负载的计算任务,而部分节点却没有计算任务而资源闲置。
- 3、OpenMP多线程程序,请合理申请资源,避免单节点超负载,程序运行时,释放多少线程,在作业运行脚本中就申请多少个核心的资源。

#### 作业管理系统的结构

集群计算服务器





集群计算服务器

集群计算服务器

任务繁忙节点

集群计算服务器

曙光根近理想不远

#### slurm作业提交步骤



1. 准备:编写描述改作业的脚本,包括作业名,需要的资源等。

2. 提交: 使用sbatch命令将该作业提交给slurm服务器

3. 排队:服务器将该任务排入适当的队列

4. 调度:服务器检查各工作节点的状态是否符合该作业的要求,并进行调度。

5. <u>执行</u>: 当条件满足时,作业被发给相应的执行服务器执行。程序运行时执行服务器会收集 程序的标准输出和标准 错误流,等程序结束时,将这些信息返回给用户。

6. <u>查询和调整</u>: 当作业在运行时,用户可以使用squeue进行状态查询。用户发现作业提交错误 时,可以使用scancel 删除正在运行的作业。

7. <u>查看结果</u>:使用文本编辑软件vi或者系统命令cat, less等查看输出及错误信息显示。

#### slurm作业脚本

Sugon

运行程序

- 口本质是一个SHELL脚本
- 口注释以"#"开头
- 口slurm运行参数,以"# SBATCH"开头,此处的"#"不是注释作用,具有实际意义

mpirun -np 180 -machinefile ./ma -rdma ./xhpl intel64 static >\$log.out

口可以直接调用SHELL命令和系统命令

```
#!/bin/sh
#SBATCH -J GENERAL 0331 080311
#SBATCH -N 5
                                      申请资源
#SBATCH -n 36
#SBATCH -t 2400:0:0
#SBATCH -p gpu v100
cd $SLURM SUBMIT DIR
                           进入作业脚本所在的目录,以这个目录为计算目录
                            加载运行所需的环境变量
module load intelmpi/2017.4.239
```

#### 作业脚本编写



```
#!/bin/bash
#SBATCH -J jobname
                           #作业名为 jobname
                           #输出文件重定向到 test.out
#SBATCH -o test.out
                           #作业提交的队列为 gpu
#SBATCH -p gpu
#SBATCH -N 10
                           #作业申请10个节点
#SBATCH -n 320
                           #总计申请320个核心
#SBATCH --ntasks-per-node=32
                           #每节点申请32个核心
                           #任务运行的最长时间为1小时
#SBATCH -t 1:00:00
                           # 单个节点使用 1 块 GPU 卡
#SBATCH --gres=gpu:1
```

### 作业脚本编写其它语句 (可选)



```
#SBATCH --qos=debug

#SBATCH --cpus-per-task=4

#SBATCH -w node6

#SBATCH -x node6
```

#作业使用的 QoS 为 debug

#单任务使用的 CPU 核心数为 4

#指定运行作业的节点是node6,若不填写系统自动分配节点

#计算时排除node6节点

### slurm作业提交基本命令



□ 在作业调度系统中,用户使用 sbatch命令提交用户程序。用户运行程序的命令及slurm环境变量设置组成slurm作业脚本,作业脚本使用如下格式提交到slurm系统运行:

sbatch <slurm作业脚本>

sbatch -J wrf --nodes=N -p batch <slurm作业脚本>

# PBS的环境变量



变量名	说 明
登陆SHELL继承来的变量	包括\$HOME, \$LANG, \$LOGNAME, \$PATH, \$MAIL, \$SHELL和\$TZ。
\$SLURM_SUBMIT_HOST	sbatch提交的节点名称
\$SLURM_SUBMIT_DIR	sbatch提交的作业的绝对路径
\$SLURM_JOBID	作业被slurm系统指定的作业号
\$SLURM_NPROCS	申请的总核心

#### slurm脚本举例



```
# 这是一个串行作业脚本的例子
#SBATCH -J test
#SBATCH -N 1
#SBATCH -n 1
cd $SLURM SUBMIT DIR
./a.out
# 这是一个并行作业脚本的例子
#SBATCH -J wrf
#SBATCH -N 2
#SBATCH -n 20
#SBATCH -p test
cd $SLURM SUBMIT DIR
source /public/software/profile.d/mpi-openmpi-intel.sh
srun hostname -s | sort -n >slurm.hosts
mpirun -np $SLURM NPROCS -machinefile slurm.hosts ./a.out
```

#### 提交作业并查看



作业提交后,会生成一个作业号,如:

[dawning@tc6000 ~]\$ sbatch test.pbs

Submitted batch job 89

#### 查看集群作业运行状态:

[test@tc6000 hpl-6230]\$ squeue

JOBID PARTITION NAME USER ST TIME NODES NODELIST(REASON)

89 v100 GENERAL test R 0:03 5 node[17-21]

#### 作业状态说明:

E: 退出

Q: 排队

H: 挂起

R: 运行

C: 结束

### 命令行查询作业状态(续)



查询作业命令 squeue [参数], 其中参数可为:

-p: 列出系统队列信息

-a: 列出队列中的所有作业及其分配的节点

-t R: 列出正在运行的作业

-j jobid:列出指定作业的信息

### 命令行查看删除作业



```
[test@tc6000 hpl-6230]$ squeue

JOBID PARTITION NAME USER ST TIME NODES NODELIST(REASON)

89 v100 GENERAL_ test R 10:09 5 node[17-21
```

[test@tc6000 hpl-6230]\$ scancel 89

注: 用户只能删除自己的作业,管理员可以删除所有用户作业

### 集群信息查看命令

- 存储可用资源
  - df -h

```
[root@tc6000 ~]# df -h
Filesystem
                       Size Used Avail Use% Mounted on
devtmpfs
                        126G
                                0 126G
                                         0% /dev
tmpfs
                       126G
                              11G 116G
                                         9% /dev/shm
tmpfs
                       126G 1.4G 125G
                                         2% /run
tmpfs
                       126G
                                0 126G
                                         0% /sys/fs/cgroup
/dev/mapper/centos-root 376G
                              53G 324G
                                         14% /
/dev/sda2
                        1.9G 173M 1.7G
                                        10% /boot
/dev/sda1
                        200M
                              12M 189M
                                         6% /boot/efi
cloudstor public
                        906T
                              30T 877T
                                        4% /public
```

- 集群计算资源
  - pestat



[root@tc6000 ~]# pestat  HOSTNAME	
State Use/All NUMBER (MB) (MB) JobId User  nodel idle 0 64 0.14 257436 236906  node2 idle 0 64 20.78* 257436 229133  node3 idle 0 64 0.11 257436 237930  node4 idle 0 64 0.18 257436 239663  node5 idle 0 64 0.17 257436 239816  node6 idle 0 64 0.10 257436 239725  node7 idle 0 64 0.16 257436 239794	
node1       idle       0       64       0.14       257436       236906         node2       idle       0       64       20.78*       257436       229133         node3       idle       0       64       0.11       257436       237930         node4       idle       0       64       0.18       257436       239663         node5       idle       0       64       0.17       257436       239816         node6       idle       0       64       0.10       257436       239725         node7       idle       0       64       0.16       257436       239794	
node2       idle       0       64       20.78*       257436       229133         node3       idle       0       64       0.11       257436       237930         node4       idle       0       64       0.18       257436       239663         node5       idle       0       64       0.17       257436       239816         node6       idle       0       64       0.10       257436       239725         node7       idle       0       64       0.16       257436       239794	
node3       idle       0       64       0.11       257436       237930         node4       idle       0       64       0.18       257436       239663         node5       idle       0       64       0.17       257436       239816         node6       idle       0       64       0.10       257436       239725         node7       idle       0       64       0.16       257436       239794	
node4 idle 0 64 0.18 257436 239663 node5 idle 0 64 0.17 257436 239816 node6 idle 0 64 0.10 257436 239725 node7 idle 0 64 0.16 257436 239794	
node5 idle 0 64 0.17 257436 239816 node6 idle 0 64 0.10 257436 239725 node7 idle 0 64 0.16 257436 239794	
node6 idle 0 64 0.10 257436 239725 node7 idle 0 64 0.16 257436 239794	
node7 idle 0 64 0.16 257436 239794	
node8 idle 0 64 0.11 257436 239809	
node9 idle 0 64 0.17 257436 239660	
node10 idle 0 64 0.22 257436 239792	
nodell idle 0 64 0.10 257436 238033	
node12 idle 0 64 0.14 257436 239330	
node13 idle 0 64 0.16 257436 244586	
node14 idle 0 64 0.13 257436 239533	
node15 idle 0 64 0.06 257436 239693	
node16 idle 0 64 0.15 257436 239627	
node17 idle 0 64 0.16 257436 239673	
node18 <u>drain*</u> <b>0</b> 64 <u>293.90*</u> 257436 176690	
node19 idle 0 64 0.15 257436 239988	
node20 idle 0 64 0.09 257436 239562	
node21 idle 0 64 0.19 257436 239964	
node22 idle 0 64 0.19 257436 239992	
node23 idle 0 64 0.12 257436 239920	
node24 idle 0 64 0.21 257436 239858	
node25 idle 0 64 0.15 257436 239856	
node26 idle 0 64 0.15 257436 239342	
node27 idle 0 64 0.14 257436 239832	
node28 idle 0 64 0.18 257436 239832	
node29 idle 0 64 0.15 257436 239342	
node30 idle 0 64 0.18 257436 239276	
node31 idle 0 64 0.22 257436 239347	
node32 idle 0 64 0.20 257436 232870	
node33 idle 0 64 0.16 257436 239336	

# WEB查看作业



← C https://manage-sc.hdu.edu.cn/ac/console3/	index.html#/job-list/curre	nt			
② 株+1年+2月16人で 科学计算 作业提交 作业管理	数据管理 命令行「E	Shell			
当前作业 历史作业					
全部集群         展示组成员作业					作业ID ∨ 搜索.
□ 作业ID 作业名	应用	队列	集群	开始时间 ◆	运行时长 ♦
				暂无数据	



# 谢谢!

数字经济基础设施的领导者 数据中国百城百行的发起者 中科院产业化联盟的推动者 信息技术应用创新的践行者

曙光很近 理想不远