成都超算中心用户手册 V1.2

一、账号注册

1. 账号申请

账号注册入口：使用网页浏览器访问：https://hpc.nscc-cd.cn，点击图1-1中的“立即注册”。

图1-1

邮箱地址和手机号码是作为认证和找回的主要途径。如未填写手机号码/邮箱地址，将无法申请计算资源。

图1-2

2. 计算资源申请

1.2.1 申请计算资源

完成图1–2的注册后，第一次登陆先进计算平台必须 “申请资源”，如图1–3。

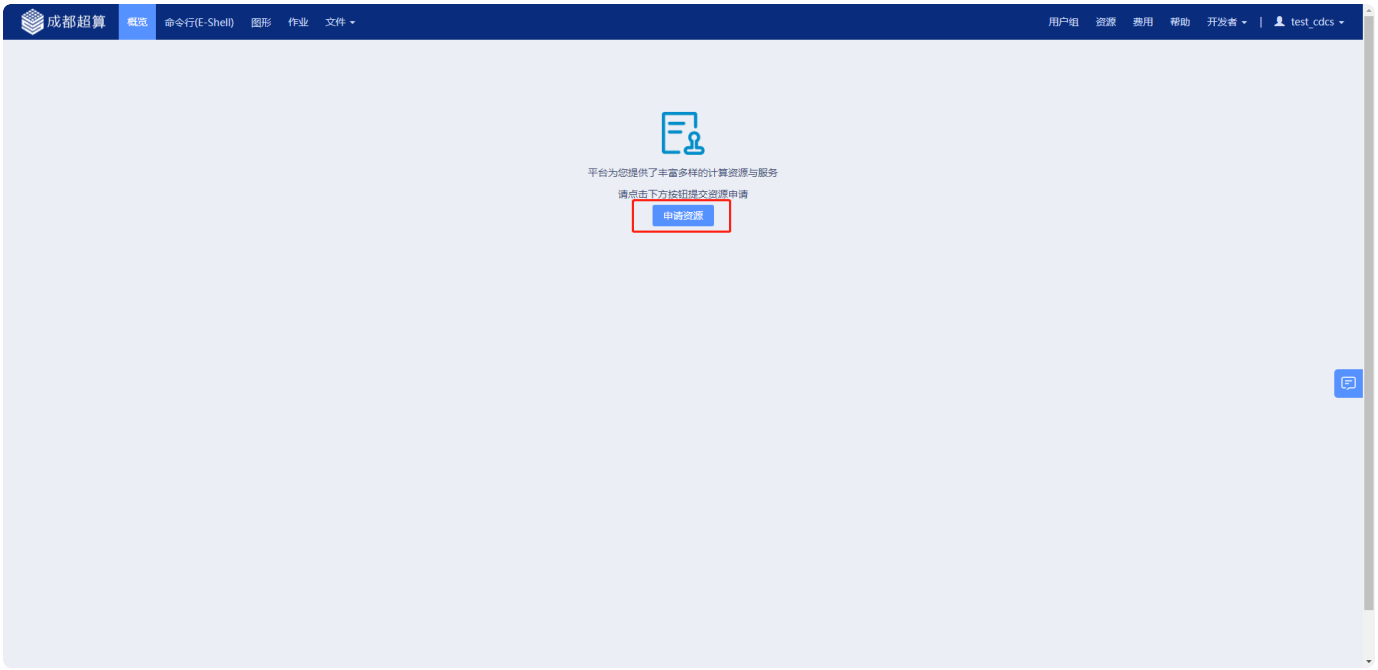

图1–3

图1–4中选择 "科学与工程计算资源" ，然后找到“试用申请”，点击 "申请试用" 按钮。申请提交后，后台管理员会进行账号审批。

注：异构节点包含了CPU资源和DCU加速卡资源。

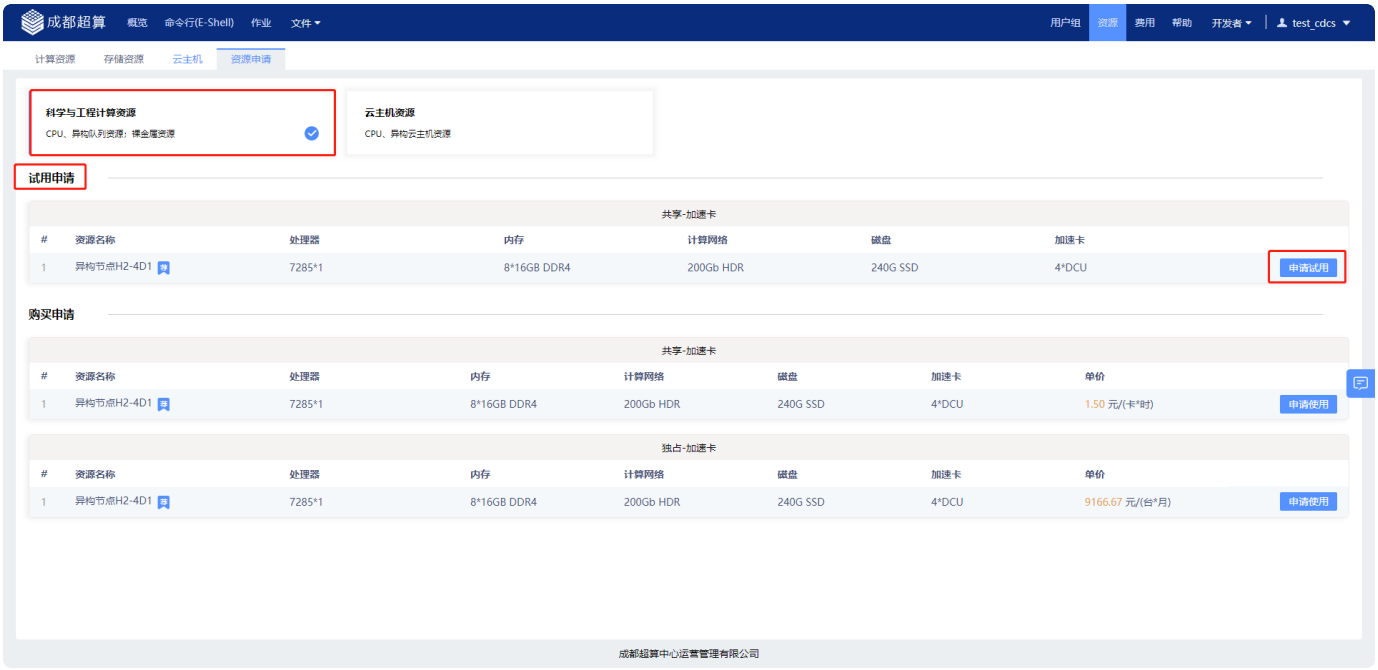

图1–4

如图1–5，请务必在备注中填写院系、研究方向和常用软件。

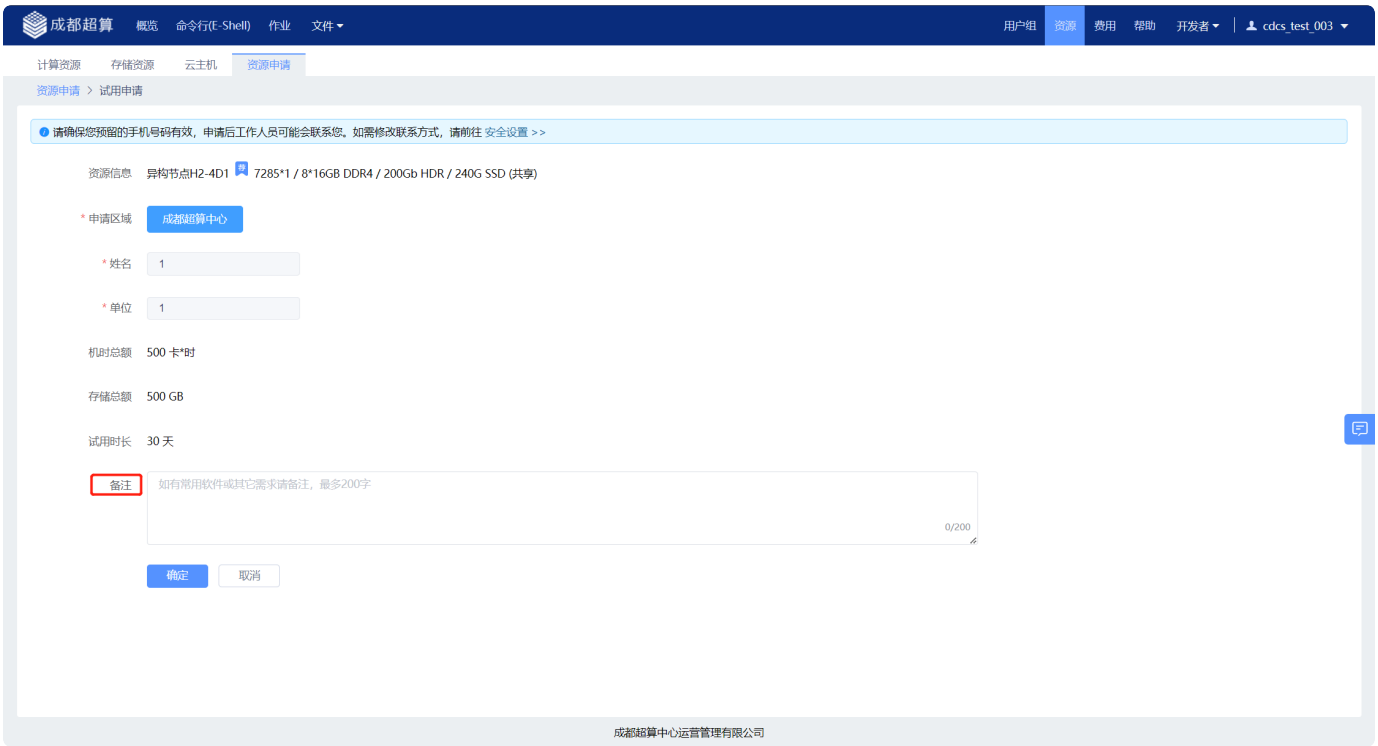

图1–5

1.2.2 主界面展示

资源申请审批通过后，再次登陆成都超算中心–先进计算平台，显示界面如图1–6。

至此，用户已可以使用成都超算中心的计算资源。

注：由于版本更新，现图1-6中第四点为”智能计算服务“，第五点为”科学与工业计算服务“。

图1-6

二、登录使用

2.1 命令行（E-Shell）

E-Shell是一个web版的Linux终端，如图2-1所示：

点击图2-1中右侧的“齿轮”按钮，可以修改字体大小、设置主题、切换登录节点。

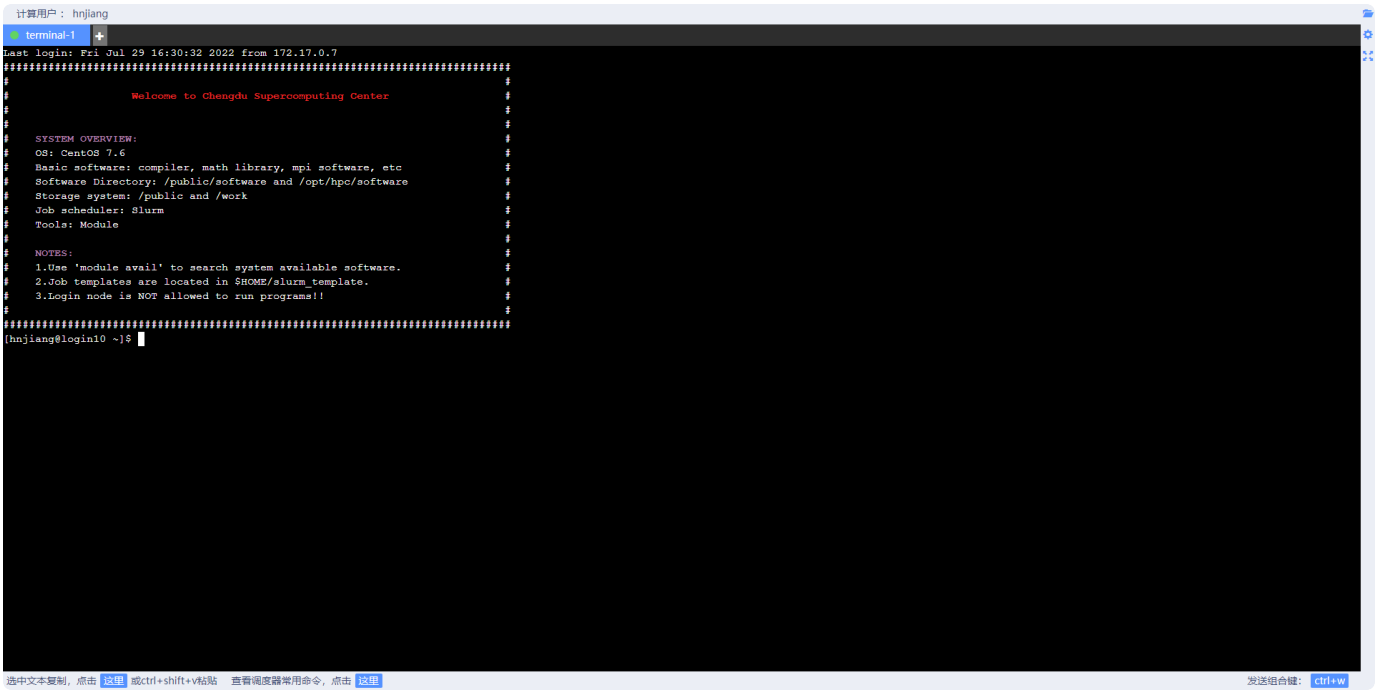

图2–1

2.2 文件（E–File）

E–File提供文件管理功能，如图2–2所示：

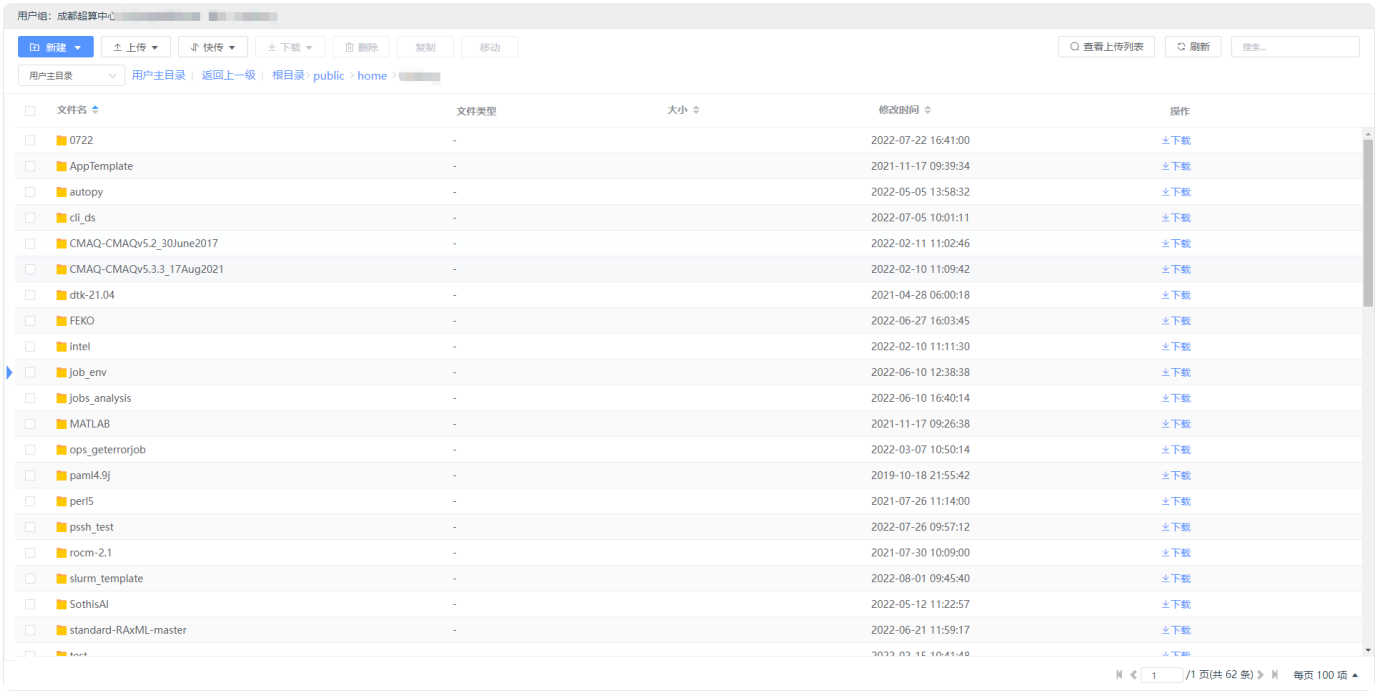

图2–2

选中文件后可执行下载、删除、重命名操作，重命名一次只能操作一个文件。选中多个文件下载将打包成一个。

上传文件可以使用“上传”，右下角会显示进度框，如图2-3。

快传尚处于测试阶段。

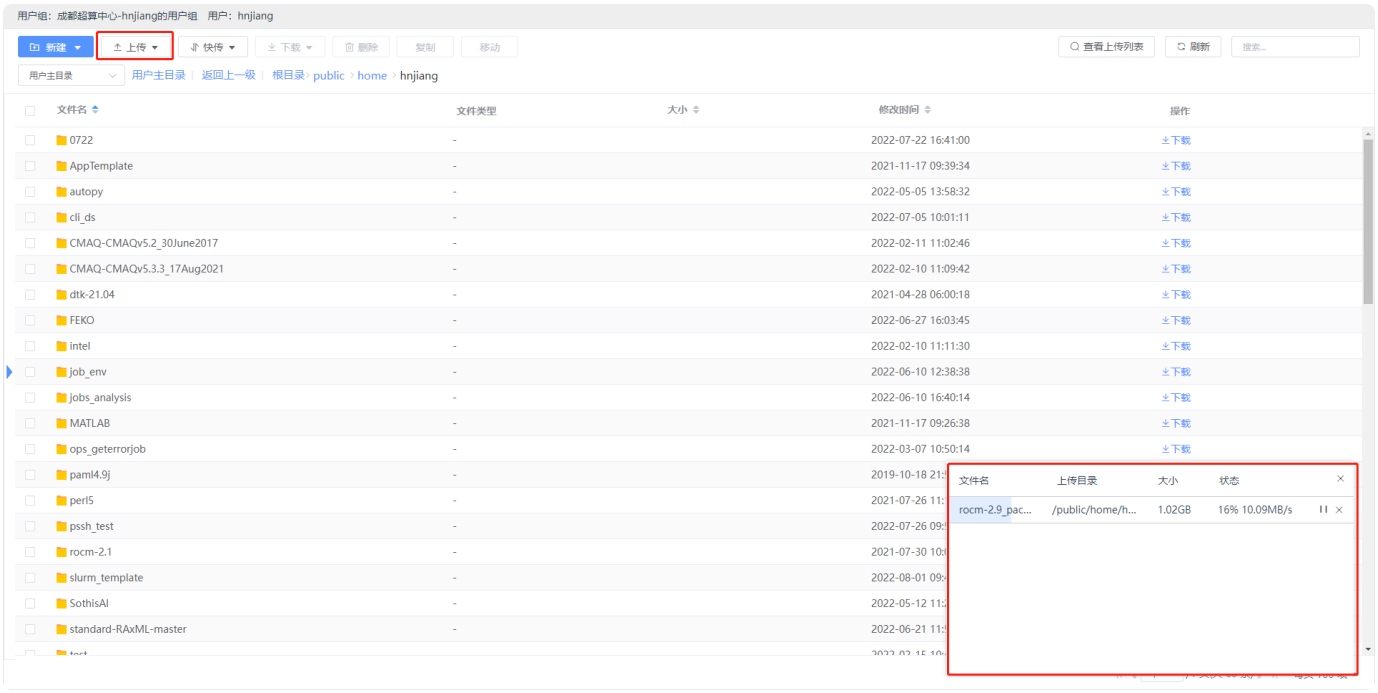

图2-3

2.3 作业

“作业”显示当前正在运行或排队的作业，如图2-4：

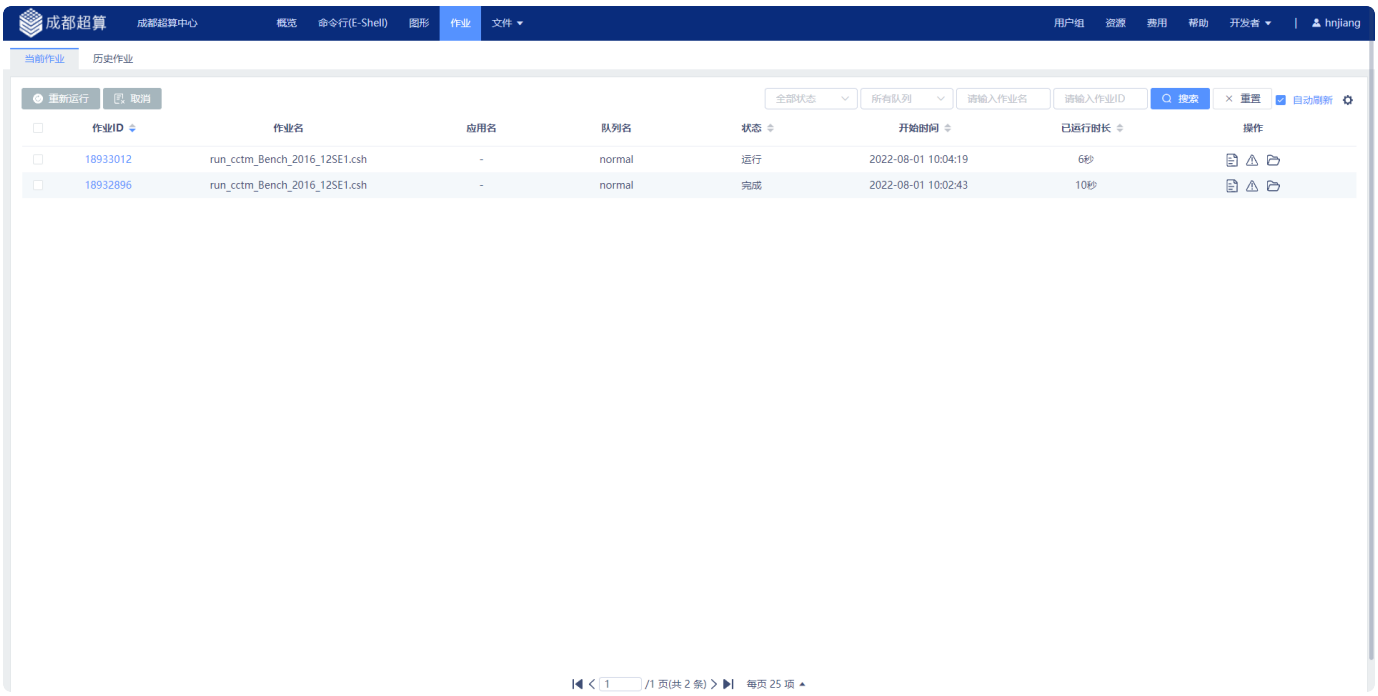

图2-4

点击一条当前正在运行的作业或历史作业可查看作业详情，如图2–5：

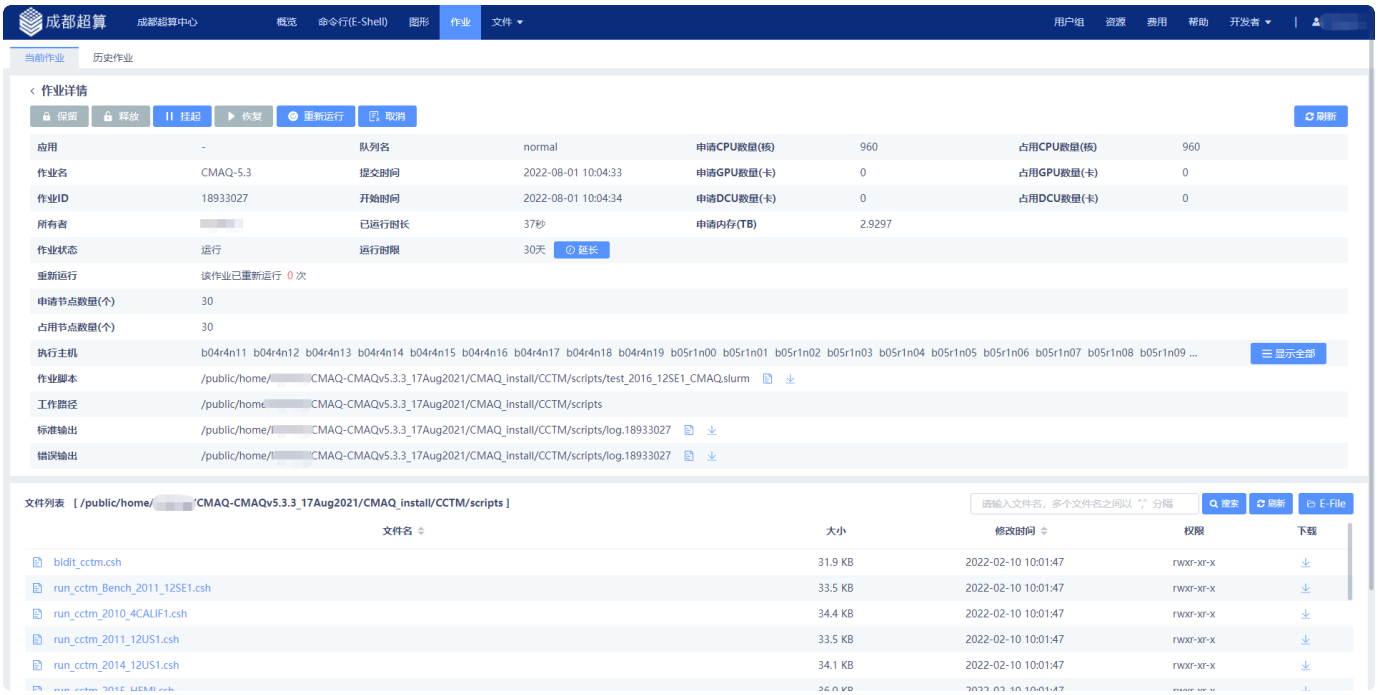

图2–5

2.4 图形

2.4.1 打开VNC界面

点击“图形”后，显示如图2–6。

点击”VNC“后，显示如图2–7，点击”打开“按钮，即可进入。

注：浏览器可能会拦截弹出窗口，请自行根据浏览器设置不拦截。

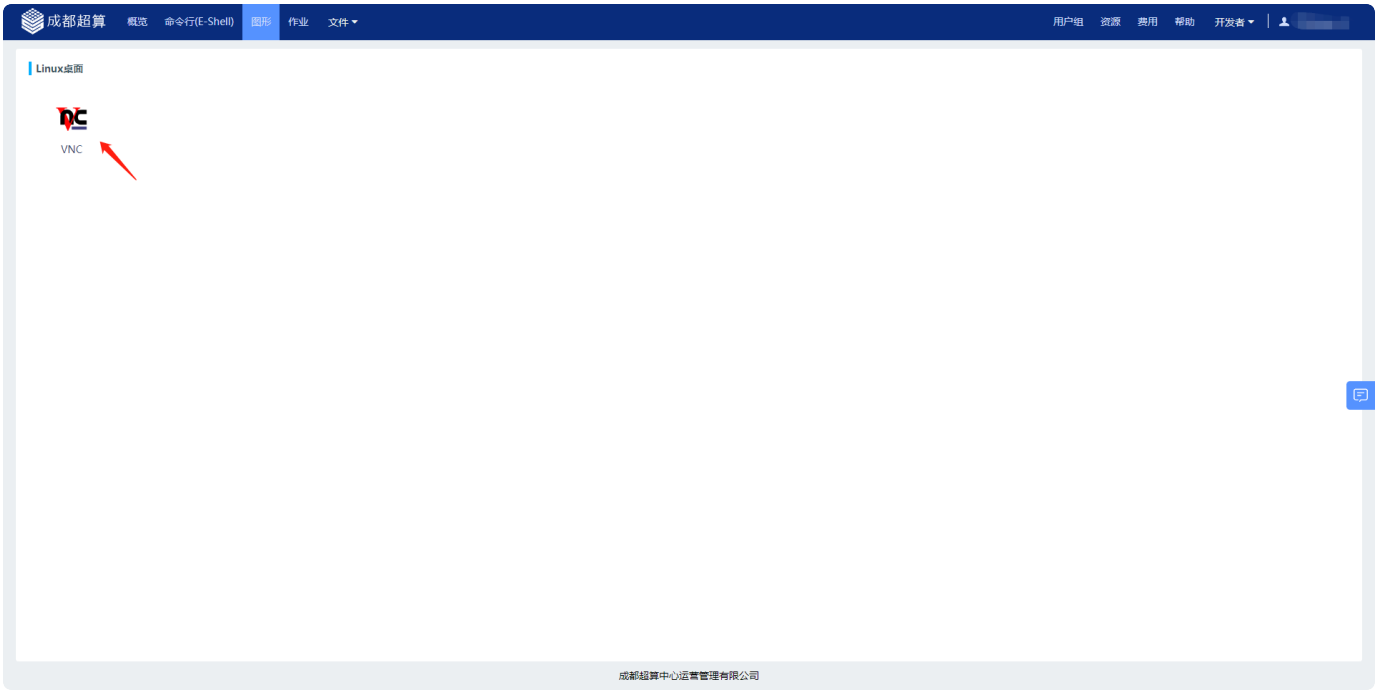
成都超算
概览
命令行(E-Shell)
图形
作业
文件
用户组
资源
费用
帮助
开发者
Linux桌面
VNC
成都超算中心运营管理有限公司

图2-6


Linux桌面
VNC
已创建的图形界面
会话名称: admin07i:26
创建时间: 2022-08-03 14:43:43
已运行时长:
打开
日志
删除

图2–7

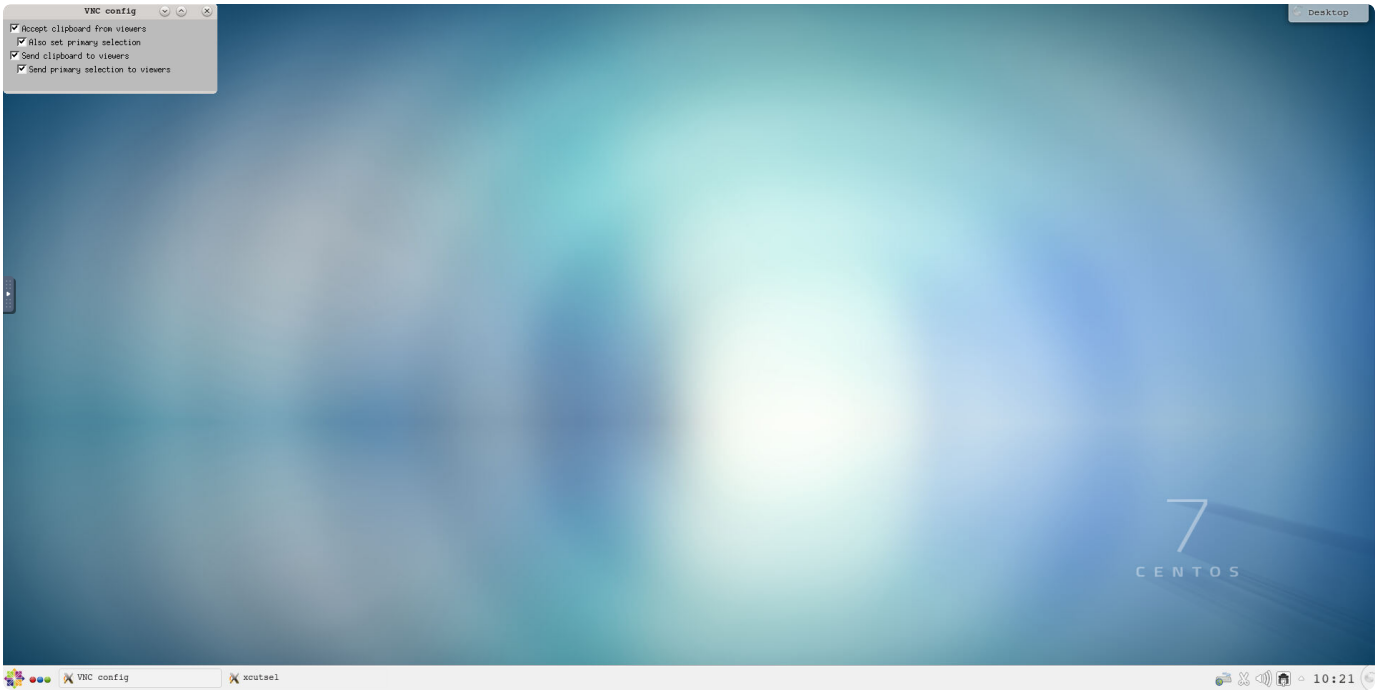

图2-8

打开后显示界面如图2-8，则说明打开成功。

2.4.2 在计算节点开启应用并输出显示至VNC可视化窗口的两种方式，任选其一即可

2.4.2.1 slurm脚本

根据图2-7中的”会话名称“修改代码框中的”DISPLAY“端口号。

```bash
#!/bin/bash
#SBATCH -J vnc-test
#SBATCH -N 1
#SBATCH -n 32
#SBATCH -t 0

ssh $HOSTNAME "export DISPLAY=admin07i:26;module load apps/vesta/gtk2;VESTA-gui"
```

通过E-Shell登录至终端，将代码框中的内容随意复制进一个文件，并提交该作业，如图2-9：

```
[hnjiang@login10 lj]$ sbatch vesta.slurm
Submitted batch job 19115934
[hnjiang@login10 lj]$ cat vesta.slurm
#!/bin/bash
#SBATCH -J vnc-test
#SBATCH -N 1
#SBATCH -n 32
#SBATCH -t 0

ssh $HOSTNAME "export DISPLAY=admin07i:26;module load apps/vesta/gtk2;VESTA-gui"
```

图2–9

通过sbatch提交作业后，即可通过图2–8 VNC窗口显示APP图形化界面，如图2–10：

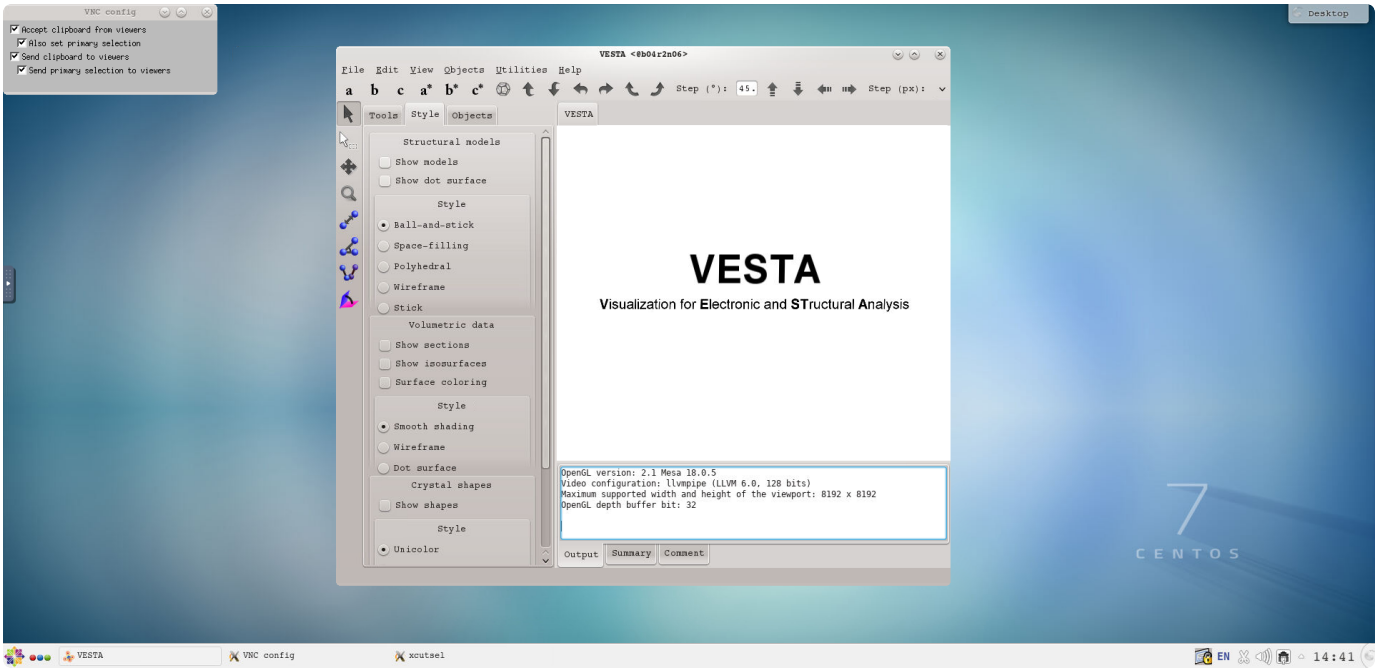

图2–10

之后需要使用scancel命令删除作业，如图2–11：

```
[hnjiang@login10 lj]$ squeue
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
          19115934    normal vnc-test  hnjiang  R       0:39      1 b04r2n06
[hnjiang@login10 lj]$
[hnjiang@login10 lj]$
[hnjiang@login10 lj]$ scancel 19115934
```

图2–11

2.4.2.2 E–SHELL内申请计算节点，通过ssh至计算节点打开应用

在命令行界面，使用salloc命令申请计算节点，如图2–12。

```bash
salloc -n 32 -N 1 -p normal
```

```
[hnjiang@login10 lj]$ salloc -n 32 -N 1 -p normal
salloc: Pending job allocation 19116042
salloc: job 19116042 queued and waiting for resources
salloc: job 19116042 has been allocated resources
salloc: Granted job allocation 19116042
salloc: Waiting for resource configuration
salloc: Nodes a01r4n18 are ready for job
[hnjiang@login10 lj]$ squeue
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
          19116042    normal     bash  hnjiang  R       0:13      1 a01r4n18
```

图2-12

根据申请到的节点名，通过命令行界面SSH至计算节点。

```bash
ssh a01r4n18
#设置DISPLAY端口
export DISPLAY=admin07i:26
#启用可视化程序
module load apps/vesta/gtk2
VESTA-gui
```

```
[hnjiang@login10 lj]$ squeue
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
          19116042    normal     bash  hnjiang  R       0:15      1 a01r4n18
[hnjiang@login10 lj]$ ssh a01r4n18
Warning: Permanently added 'a01r4n18' (ED25519) to the list of known hosts.
[hnjiang@a01r4n18 ~]$ export DISPLAY=admin07i:26
[hnjiang@a01r4n18 ~]$ module load apps/vesta/gtk2
[hnjiang@a01r4n18 ~]$ VESTA-gui
```

图2-13

VNC窗口，如图2-14：

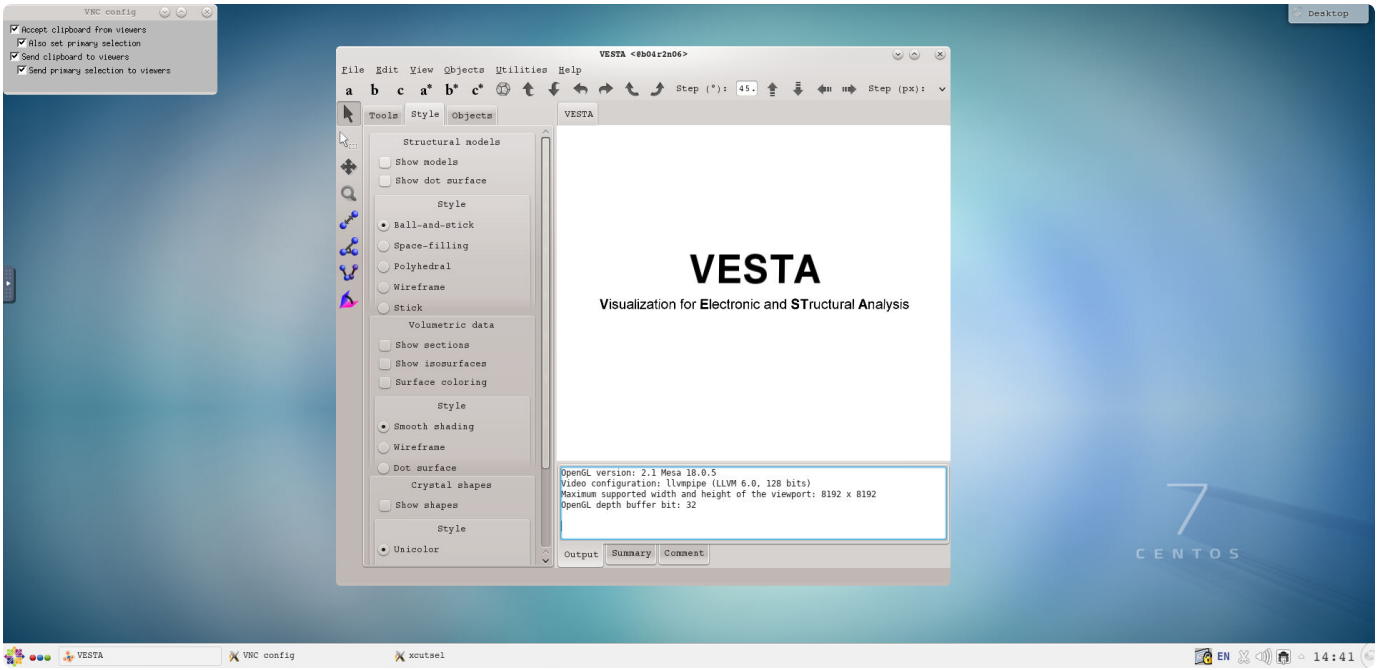

图2–14

注意：不能直接在VNC窗口内执行程序或打开终端，若管理员发现节点占用过高会立即删除。VNC窗口仅作为可视化输出窗口。

2.5 账号资源变更（资源）

回到主界面点击”资源“按钮，可进行扩容申请，如图2–15，包括计算资源和存储资源。

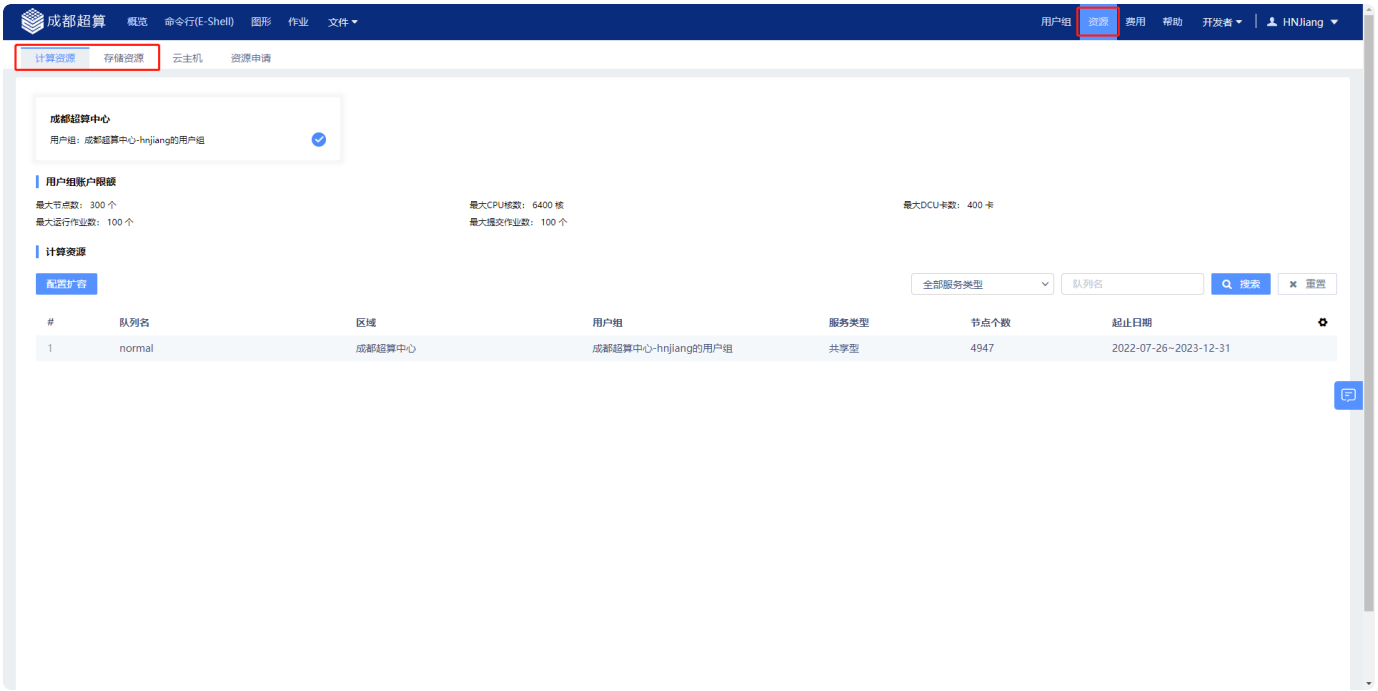

图2–15

2.5.1 扩容申请

点击图2–15中的”配置扩容“后，即可对计算资源进行申请，如图2–16。

注：申请计算资源扩容，需要联系您的运营经理进行后台审批。

注：图2–16中的”用户组名额“指当前账号可容纳的计算账户数。

注：无GPU卡。

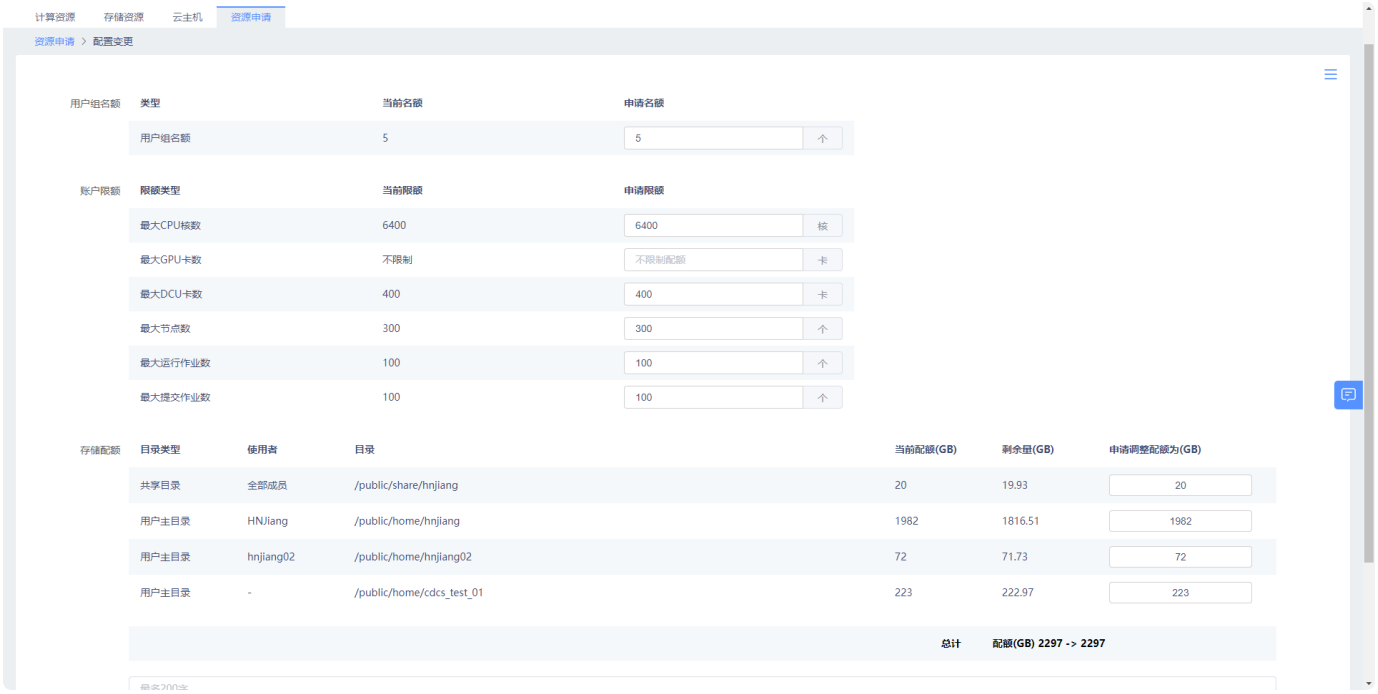

图2–16

2.5.2 账户组内存储资源分配

如果该组账号下有多个计算账号，可对组下的各个计算账号进行存储资源的分配。

如图2–17，该账号共有三个计算账号，可查看到各个账号的使用情况等信息。

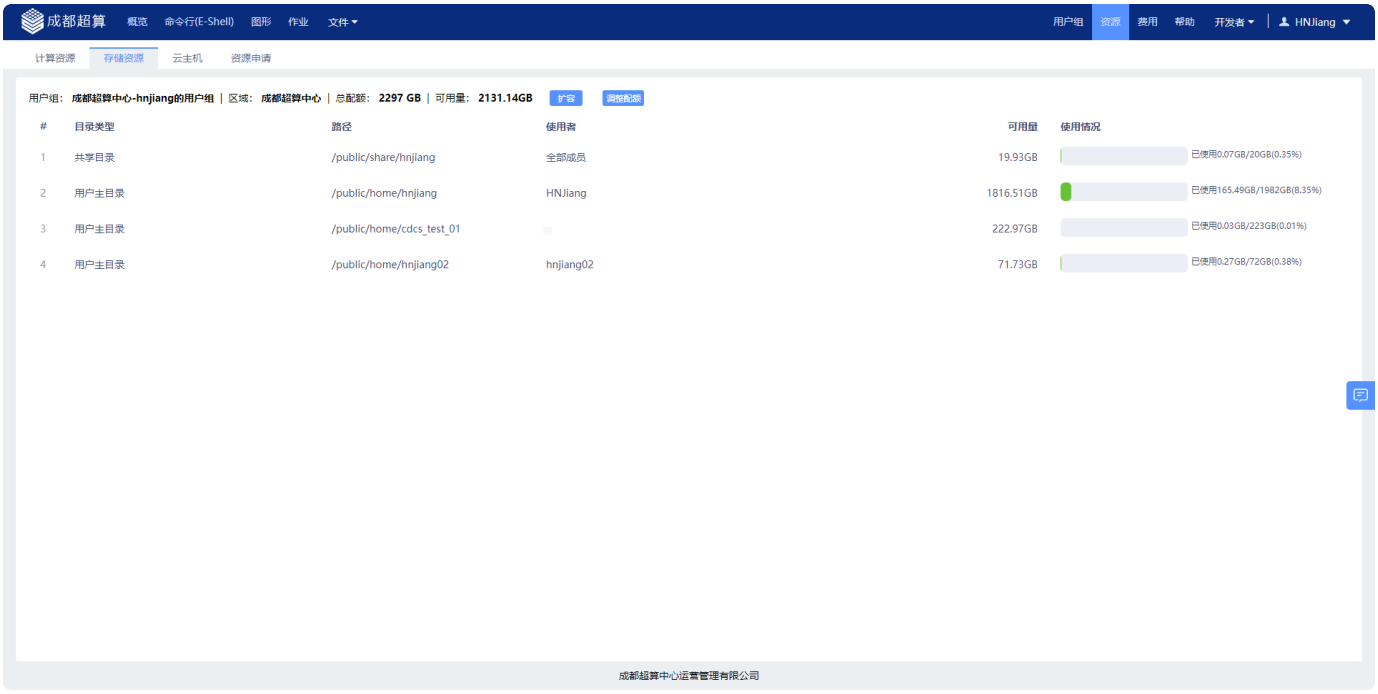

图2–17

点击图2–17中的”调整配额“既可对组下的各个计算账号进行存储资源的调整。

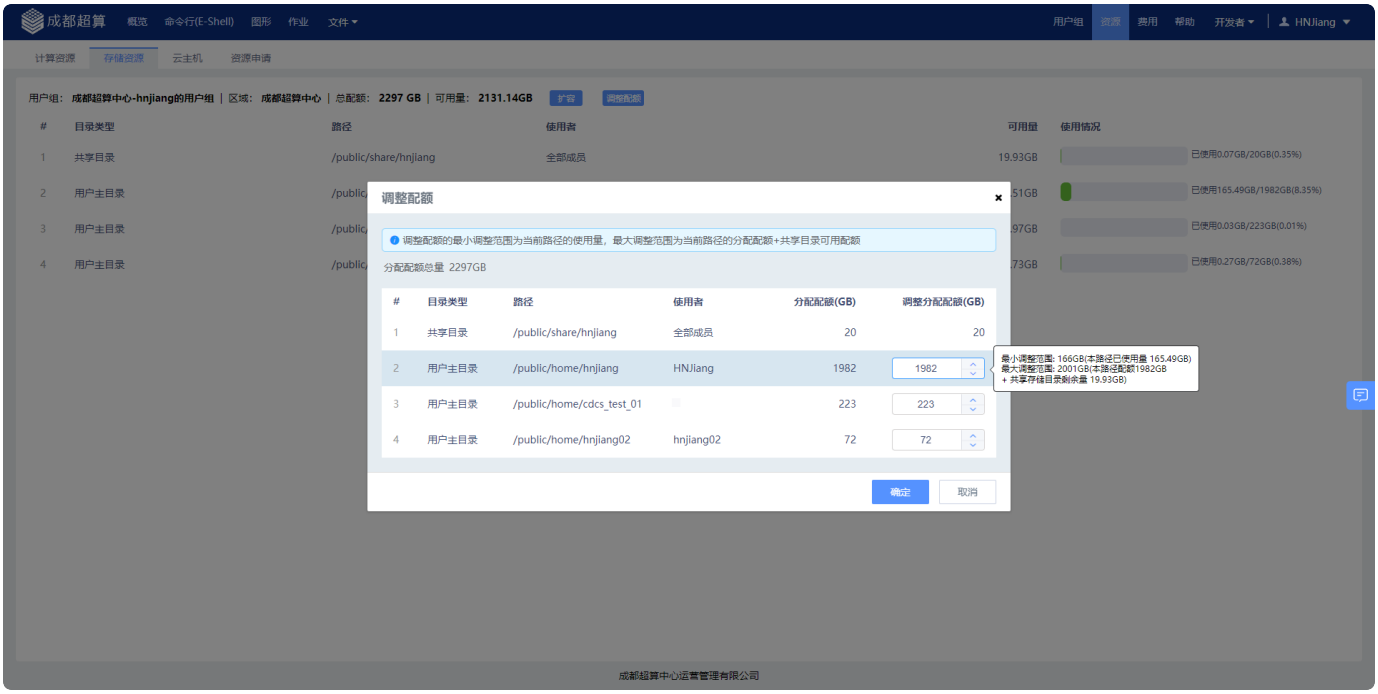

图2–18

2.6 费用

如图2–19“费用”界面，其中包括了“总览、账单、消费明细、收支明细、导出记录“。

可通过“消费明细”查看或导出多种数据，例如共享机时费等，也可通过“账单”查看或导出当月或历史使用情况。

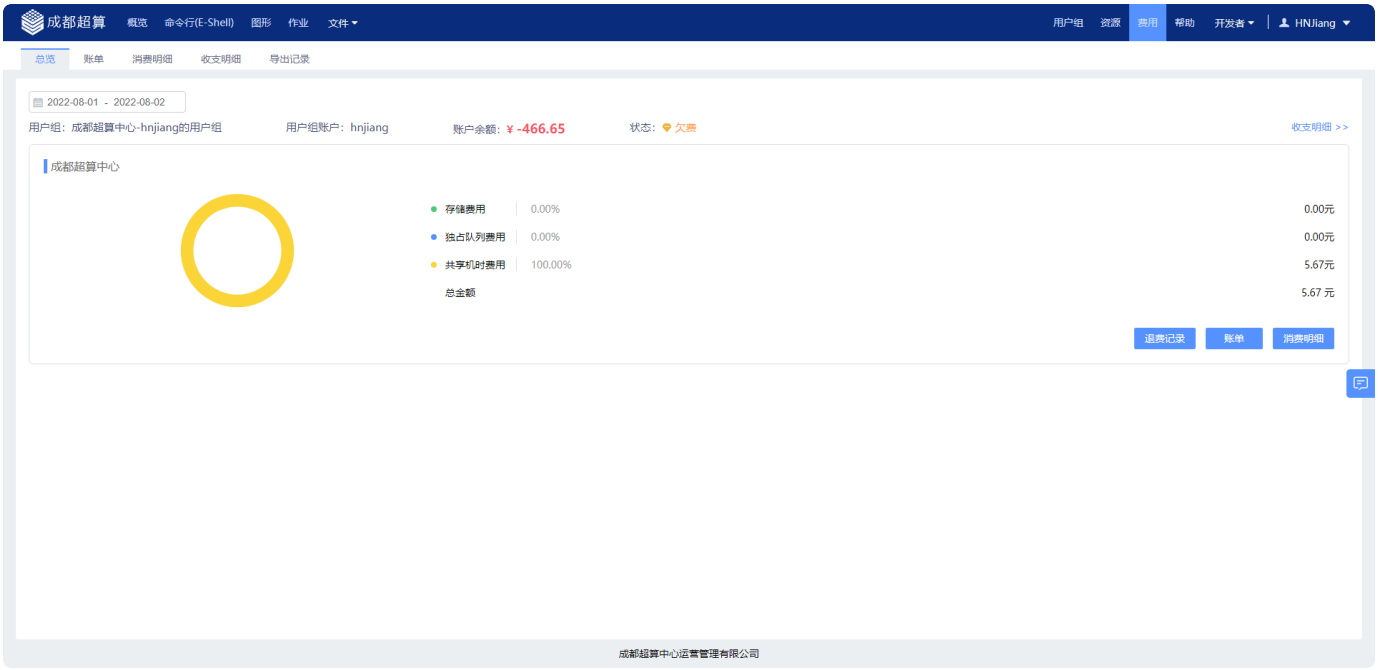

图2–19

三、作业相关

3.1 环境的加载

成都超算环境中，正确加载计算环境是计算软件正常运行的必要条件。

module avail 命令可用于查看系统可用软件清单。如图3–1，列表名称规则："应用类型/软件名/版本号/编译器"。

例如：apps/abinit/8.10.3/hpcx–2.4.1–intel2017 ，代表hpcx 并行环境和intel2017编译的应用软件abinit（版本 8.10.3）。

图3–1

介绍几种modulefile工具的常用指令：

```Bash
module av ## 查看系统中可用的软件
module load ## 加载环境，例如 module load apps/miniconda/3
module li ## 查看当前已加载的环境
module unload ## 卸载不需要的环境，例如 module unload apps/miniconda/3
module purge ## 卸载当前所有已加载的环境
module show ## 显示环境的配置文件，例如 module show apps/miniconda/3
```

一款并行程序通常需要串行编译器加并行编译器来编译，然后通过类似 mpirun 的并行执行命令来运行。如果运行过程中缺少动态库就会报错。因此，在设置依赖关系，加载软件模块时，必须预先加载依赖的其它模块（如编译器模块等），然后再加载模块本身，以免报错。

建议按照编译器、并行环境、所需库（如有需要）、应用软件的顺序加载相应环境。

3.2 作业调度系统Slurm

3.2.1 sinfo 分区查询

在先进计算平台中，使用 sinfo 命令查询队列信息。根据命令输出，可以看到集群作业调度系统的队列情况，例如：

```Bash
sinfo -p normal -t idle ##查看normal队列的空闲节点
```

3.2.2 srun 提交交互式作业

```Bash
srun -p normal -n 2 hostname ## 在normal队列上指定任务数运行 hostname
```

3.2.3 sbatch（推荐使用）提交批处理作业

用户使用sbatch命令提交作业脚本，其基本格式为sbatch jobfile，jobfile为作业脚本文件。

在批处理作业脚本中，脚本第一行以"#!"字符开头，并指定脚本文件的解释程序，如 sh，bash。

sbatch常用选项如下表：

选项	含义	示例

–J	作业名称	–J test
–n	作业申请的总cpu核心数	–n 240
–N	作业申请的节点数	–N 10
–p	指定作业提交的队列	–p normal
––ntasks–per–node=	指定每个节点运行的进程数	––ntasks–per–node=32
––cpus–per–task=	指定任务需要的处理器数目	––cpus–per–task=1
–o	指定作业标准输出文件的名称，不能使用shell环境变量	–o %J.out，%J表示作业号
–e	指定作业标准错误输出文件的名称，不能使用shell环境变量	–e %J.err
–w	指定分配特定的计算节点	–w a01r01n01
–x	指定不分配特定的计算节点	–x a01r01n01
––exclusive	指定作业独占计算节点	––exclusive
––mem=	指定作业在每个节点的内存限制	––mem=2G
––mem–per–cpu=	限定每个进程占用的内存数。	––mem–per–cpu=512M
–d	作业依赖关系设置	–d after:1234表示本作业须待作业1234开始以后再执行
––gres=	指定每个节点使用通用资源名称及数量	––gres=dcu:4 表示本作业使用DCU加速卡

3.2.3.1 sbatch 使用示例

串行作业示例

单节点单核心执行“sleep”命令，用户可根据自己的实际情况进行更改。

sleep.slurm需复制至家目录下，然后通过sbatch命令进行提交。

sleep.slurm Bash

```
#!/bin/bash
#SBATCH -J test
#SBATCH -N 1
#SBATCH -n 1
#SBATCH -p normal

date
hostname
sleep 100
date
```

Bash

```
sbatch sleep.slurm
```

mpi并行作业示例

案例代码如下：

montecarlo.cpp Bash

```cpp
#include <mpi.h>
#include <iostream>

int main(int argc, char** argv)
{
    int rank, size, n, i, total;
    double x, y, pi;
    int count = 0;

    MPI_Init(&argc, &argv);
    MPI_Comm_rank(MPI_COMM_WORLD, &rank);
    MPI_Comm_size(MPI_COMM_WORLD, &size);

    if (rank == 0)
    {
        total = 1000000;
        if (argc >= 2){
            total = atoi(argv[1]);
        }
    }

    n = total / size;
    MPI_Bcast(&n, 1, MPI_LONG_LONG_INT, 0, MPI_COMM_WORLD);

    for (i = 0; i < n; i++)
    {
        x = rand() / double(RAND_MAX);
        y = rand() / double(RAND_MAX);
        if (x*x + y*y <= 1)
        {
            count++;
        }

    }

    int total_count = 0;
    MPI_Reduce(&count, &total_count, 1, MPI_LONG_LONG_INT, MPI_SUM, 0 , MPI_COMM_WORLD);
    pi = (4.0 * total_count) / total;

    if (rank == 0)
    {
        std::cout << "PI ≈ " << pi << std::endl;
    }

```

```
    MPI_Finalize();
    return 0;

}
```

run.slurm Bash

```
#!/bin/bash
#SBATCH --job-name=openmpitest
#SBATCH -p normal
#SBATCH -N 2
#SBATCH -n 60
#SBATCH --ntasks-per-node=30
#SBATCH --output=%j.out
#SBATCH --error=%j.err

ulimit -s unlimited
ulimit -l unlimited

module purge
module load compiler/gcc/9.3.0
module load mpi/openmpi/4.0.5/gcc-9.3.0

mpic++ montecarlo.cpp -o montecarlo -fopenmp

mpirun -np 60 ./montecarlo
```

更多使用方法可通过sbatch --help帮助手册查询。

3.2.4 salloc 节点资源获取

该命令支持用户在提交作业前，先获取所需计算资源，例如在第2.4.2.2节中，使用该命令申请了计算节点。

更多使用方法可通过salloc --help帮助手册查询。

3.2.5 squeue 作业信息查询

用户使用 squeue 命令，可以查看当前作业信息，如图3-2

```
[cdcs_test_01@login01 ~]$ squeue
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
           8253895    normal     bash cdcs_tes  R       0:57      1 d12r4n04
```

图3-2

图3-2中，参数含义如下表：

JOBID	作业 ID
PARTITION	分区名称
NAME	作业名
USER	用户名
ST	作业状态
TIME	运行时长
NODES	作业占用节点数
NODESLIST (REASON)	节点列表（原因）

其中作业**状态（ST）**有如下几种状态：

R (Runing)	正在运行
PD (PenDing)	作业排队中
CG (COMPLETING)	作业正在完成中
CA (CANCELLED)	作业被人取消
CD (COMPLETED)	作业运行完成
F (FAILED)	作业运行失败
NF (NODE_FAIL)	节点问题导致作业运行失败
PR	作业被抢占
S	作业被挂起
TO	作业超时被杀

其中**NODELIST（REASON）**有如下几种常见原因：

AssociationJobLimit	作业达到其最大允许的作业数限制
AssociationResourceLimit	作业达到其最大允许的资源限制
AssociationTimeLimit	作业达到时间限制

QOSJobLimit	作业的QOS达到其最大的作业数限制
QOSResourceLimit	作业的QOS达到其最大资源限制
QOSTimeLimit	作业的QOS达到其最大时间限制
JobHeldUser	作业被用户自己挂起
Resource	作业等待期所需资源可用