

# 单细胞中心 高性能计算平台用户指南 CSCOmics-HPC User Guide

单细胞中心生物学平台  
逐码科技 SimpleHPC

# 目录

## Contents

### 一、集群概况

Cluster overview

### 二、服务器登录

Server login

### 三、软件环境管理

Software environment management

### 四、SLURM调度系统

SLURM scheduling system

# 一、集群概況

Part 1. Cluster overview

# 集群概况

## Cluster overview

- 目前集群由11台Dell服务器组成 (11 Dell servers in total)
  - 管理节点1台 (CentOS 7.6) (1 Management node)  
主要作用：运行集群用户认证、作业调度、存储挂载等服务，并向用户提供登录及提交作业的功能。(Function: login, submit SLURM jobs)
  - 计算节点8台 (CentOS 7.6) (8 Compute nodes)  
主要作用：运行用户提交的作业，进行计算任务。(Function: run jobs)
  - 存储节点2台 (CentOS 7.6) (2 storage nodes)  
主要作用：存储数据，向其他节点提供存储挂载的服务。(Function: mount directories, store amounts of data)

# 集群概况

## Cluster overview

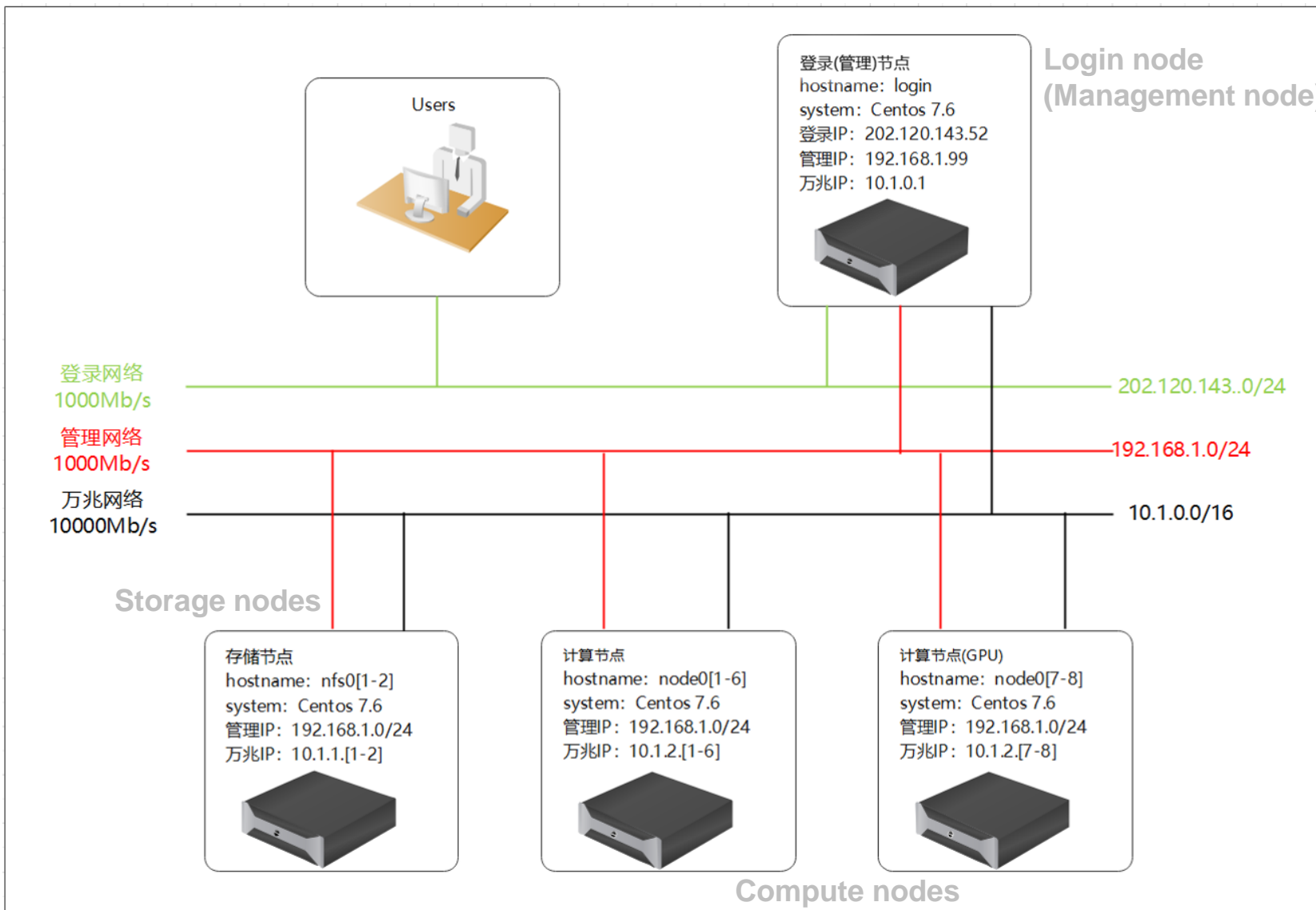
节点 Node	功能分类 Function	型号 Type	数量 Number	CPUs	Memory	GPU
管理节点 Management node	服务+登录节点	PowerEdge R740xd	1	6	64G	/
计算节点 Compute node	cpu节点	PowerEdge MX740c	6	80	512G	/
	gpu节点	PowerEdge R740	2	80	1T	NVIDIA Corporation Tesla V100 * 2 16GB
存储节点 Storage node	存储节点	使用beegfs向集群提供1.5P存储挂载空间				

# 集群概况

Cluster overview

Network Topology

网络架构



# 相关存储路径

## Cluster directories

- 集群统一存储 (Master directory)

/cluster

- 软件存储 (Software directory)

/cluster/apps

- 用户家目录 (Home directory)

/cluster/home

# 二、服务器登录

Part 2. Server login



# 登录方式

## Login (in campus)

- Windows (Xshell)

使用终端软件，SSH登录，输入IP、端口号（port）和密码。

- Mac (iterm2)

方法同上。端口号前面加参数“-p”。

\*医学院师生在校外需要先连接VPN才能访问服务器。

```
[C:\~]$ ssh gurobi@202.120.143.52 1103

Connecting to 202.120.143.52:1103...
Connection established.
To escape to local shell, press 'Ctrl+Alt+J'.

Last login: Tue May 23 17:58:33 2023



#####
#
#                               Welcome to CSCComics-HPC!
#
#-----#
# This is the login node. To improve user experience, please do not transfer #
# large files and run computationally intensive processes on this node.     #
# Please submit your jobs to SLURM, and transfer large files through SFTP(ip: #
# 10.110.0.43, port: 1103).
#-----#
#                               Features
#
#
# 1. Cloud platform
# Website: http://202.120.143.52:9000
# Apps: Jupyter, RStudio, matlab
# Functions: Compute resources (SLURM) and storage resources statistics
#
# 2. Module
# Name: R/4.2.2, anaconda3/2022.10, etc.
# Location: /cluster/apps
#
# Usage example:
# Step1. module load anaconda3/2022.10
# Step2. source activate scanpy
# or just 'module load software_name'
#
# 3. Reference genomes
# Location: /cluster/apps/Refs
#-----#
# If you have any problems, please contact csccomicshpc@shsmu.edu.cn. Thanks. #
#-----#
* Version = CentOS Linux release 7.6.1810 (Core)
* Mem/free = 62.3341/53.1381 GB
```

# 登录方式

Login (for users from other institutions)

- 连接vpn后，登录堡垒机，在主机运维里ssh连接登录节点

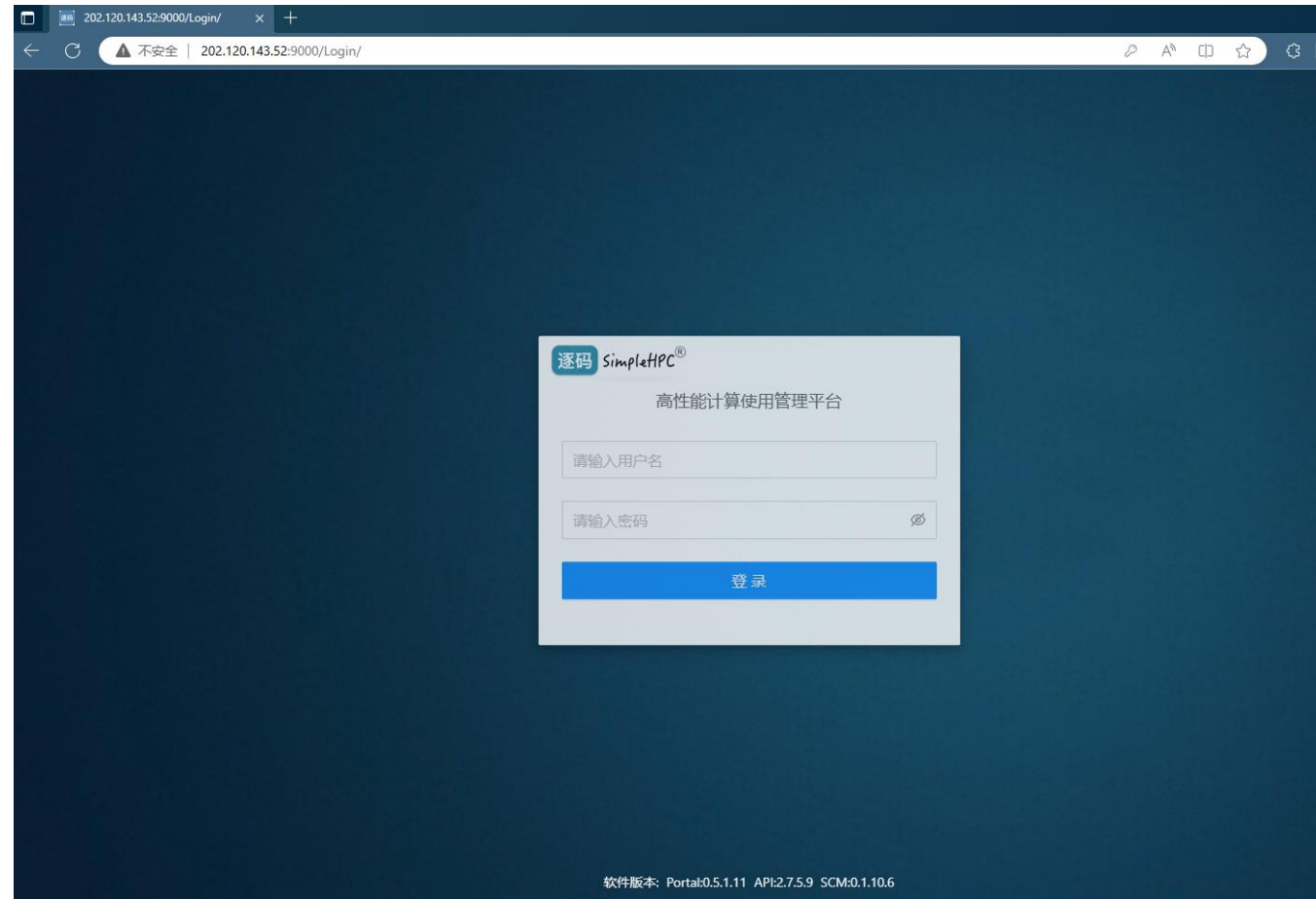
First, connect to vpn, then log in to the bastion host, choose “host” and click computer icon (will start a local client) or earth icon (will open a webpage).

名称	地址	类型	访问方式
SFTP传输文件专用节点	10.110.0.43	Linux	 [SFTP] [EMPTY]
单细胞中心服务器登录节点	202.120.143.52	Linux	 [SSH] [EMPTY]

# 登录方式

Login (to cloud platform, for all users)

- 也可以连接云平台 (<http://202.120.143.52:9000>)



# 三、软件环境管理

Part 3. Software environment management

# Module版本控制

## Environment Modules

- module 是 linux 下的一个用来管理不同环境的软件，每个用户所需要的环境不同，并且需要在软件的不同版本之间切换，module 就可以方便的管理当前用户所处的环境。
- module is a user interface to the Modules package. The Modules package provides for the dynamic modification of the user's environment via modulefiles.



# Module版本控制

## Environment Modules

- 加载要使用的软件 (load a module)  
module load gcc/12.2.0
- 卸载已加载的软件 (unload a module)  
module unload gcc/12.2.0

```
[root@login ~]# module load gcc/12.2.0
[root@login ~]# which gcc
/cluster/apps/gcc/12.2.0/bin/gcc
[root@login ~]# module unload gcc/12.2.0
[root@login ~]# which gcc
/usr/bin/gcc
```

# Module版本控制

## Environment Modules

- 查看当前已经加载了哪些软件 (list loaded modules)  
module li

```
[root@login ~]# module load anaconda3/2022.10
[root@login ~]# module load gcc/12.2.0
[root@login ~]# module li
Currently Loaded Modulefiles:
  1) anaconda3/2022.10   2) gcc/12.2.0
```



# Anaconda

- Anaconda 是一个python的数据科学平台，anaconda 的特点是拥有大量的科学包和一些依赖项，相对于python不用自行下载相关的包，anaconda支持创建虚拟环境，可以让不同用户在自己的环境下安装和管理包。

Anaconda is a free and open-source distribution of the programming languages Python and R. The distribution comes with the Python interpreter and various packages related to machine learning and data science.

# Anaconda

- 加载anaconda (load anaconda)

module load anaconda3/2022.10

- 查看anaconda下的环境 (list environments)

conda env list

一般公有环境位于: /cluster/apps/anaconda3/2022.10/envs/

私有环境位于: /cluster/home/USERNAME/.conda/envs/

# Anaconda

- 创建私有环境 (create env)  
conda create -n ENV\_NAME

\*建议先srun到计算节点再操作conda环境。

srun --pty /bin/bash

```
[pm@head1 ~]$ conda create -n test
Collecting package metadata (current_repodata.json): done
Solving environment: done

==> WARNING: A newer version of conda exists. <==
  current version: 4.12.0
  latest version: 23.1.0

Please update conda by running

  $ conda update -n base -c defaults conda

## Package Plan ##

  environment location: /cluster/home/pm/.conda/envs/test

Proceed ([y]/n)? y

Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
#   $ conda activate test
#
# To deactivate an active environment, use
#
#   $ conda deactivate
```

# Anaconda

- 激活私有环境 (activate env)  
source activate ENV\_NAME
- 在环境中安装相关python包 (install package)  
conda install PACKAGE\_NAME
- 查看当前环境已安装的包 (list installed packages)  
conda list

# Anaconda

- 退出conda环境 (exit env)  
source deactivate
- 删除指定的环境 (delete env)  
conda remove -n ENV\_NAME --all

# 四、SLURM调度系统

Part 4. SLURM scheduling system

# SLURM

- SLURM是当前成熟、功能强大的作业调度系统，被广泛的应用于大规模的高性能计算平台。

SLURM is an open source, fault-tolerant, and highly scalable cluster management and job scheduling system for large and small Linux clusters.

# SLURM

- 查看集群节点状态 (check status of nodes)

```
[root@login ~]# sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
debug      up 30-00:00:0 1    idle node01
normal*    up 7-00:00:00 5    mix  node[02-06]
gpu        up 7-00:00:00 2    mix  node[07-08]
Jupyter    up 7-00:00:00 5    mix  node[02-06]
App        up 7-00:00:00 5    mix  node[02-06]
```

normal为队列名，\*号代表该队列为默认队列

状态：

idle 空闲

mix 有作业在运行但还有空闲资源

alloc 有作业运行并且无空闲资源



# SLURM

- 查看节点负载情况 (check load of nodes)

lsload

```
[root@login ~]# lsload
HOSTNAMES      STATE    CPUS CPU_LOAD  S:C:T    CPUS(A/I/O/T)  MEMORY  GRES      PARTITION  AVAIL_FEATURES
node01         idle     80    0.01      2:20:2   0/80/0/80      515237  (null)   debug      (null)
node02         mix      80    351.61    2:20:2   40/40/0/80     515216  (null)   normal*    (null)
node03         mix      80    4.96      2:20:2   52/28/0/80     515216  (null)   normal*    (null)
node04         mix      80    1.00      2:20:2   40/40/0/80     515216  (null)   normal*    (null)
node05         mix      80    11.18     2:20:2   40/40/0/80     515237  (null)   normal*    (null)
node06         mix      80    4.90      2:20:2   34/46/0/80     515216  (null)   normal*    (null)
node07         mix      80    0.02      2:20:2   12/68/0/80     1031448 gpu:v100  gpu        (null)
node08         mix      80    1.14      2:20:2   16/64/0/80     1031448 gpu:v100  gpu        (null)
```

CPUS(A/I/O/T)

A: alloc 已分配

I: idle 空闲

O: other 其他

T: total 全部

# SLURM

- 查看作业信息 (check status of jobs)

sq

```
[root@login ~]# srun sleep 30 &
[1] 203499
[root@login ~]# sq | grep root
1494      sleep          N/A      R      root      normal    root      normal    4294900329 1      0:05      2      node06
[root@login ~]#
```

可看到作业相关信息：

作业id、作业名、用户、队列、使用cpu数、使用节点名

作业状态：

R 运行中

PD 等待中

CG 完成中

# SLURM

- 交互式作业 (submit an interactive job)

`srun --pty /bin/bash`

```
[root@login ~]# srun --pty /bin/bash
[root@node06 ~]# hostname
node06
```

注：通过`srun --help`查看命令帮助。

# SLURM

- 脚本式作业 (edit a script)

编辑slurm脚本 vim run.sh

```
#!/bin/bash
#SBATCH -J jobname
#SBATCH -p partition
#SBATCH -N 1          # 1 node
#SBATCH -n 1          # 1 task
#SBATCH -c 1          # 1 cpu
#SBATCH -o jobname.%j.out
#SBATCH -e jobname.%j.err
```

```
module load anaconda3/2022.10
source activate test
python /path/train.py
```

# SLURM

- 脚本式作业 (submit a background job)

递交slurm脚本

```
sbatch run.sh
```

sbatch的参数也可以像srun那样写在命令行中，但我们一般会写在脚本中。

可通过sbatch --help查看具体的帮助信息。

# SLURM

- 取消作业 (cancel jobs)  
scancel JOB\_ID

# 附加材料

Additional materials

# 用户应用程序

## Software

目前在/cluster/apps下已安装如下应用程序:

```
[root@login ~]# ls /cluster/apps/  
anaconda3  bwa          gmp          mpc          picard       SCcaller    tabix  
annovar    cmake        hdf5         mpfr         pipelines    snpEff      TenXGenomics  
backup     demuxlet     hisat2       nf_core      plink        sratoolkit  Utility  
bcftools   fastp        htlib        old          Python       STAR        varscan  
bcl2fastq  FastQC       java         openssl      R            stor        vcftools  
bedtools   gatk         libpng       parallel     Refs         stor4       vep  
bowtie2    gcc          modulefiles  perl         samtools    subread    VS_Code
```



# 用户应用程序

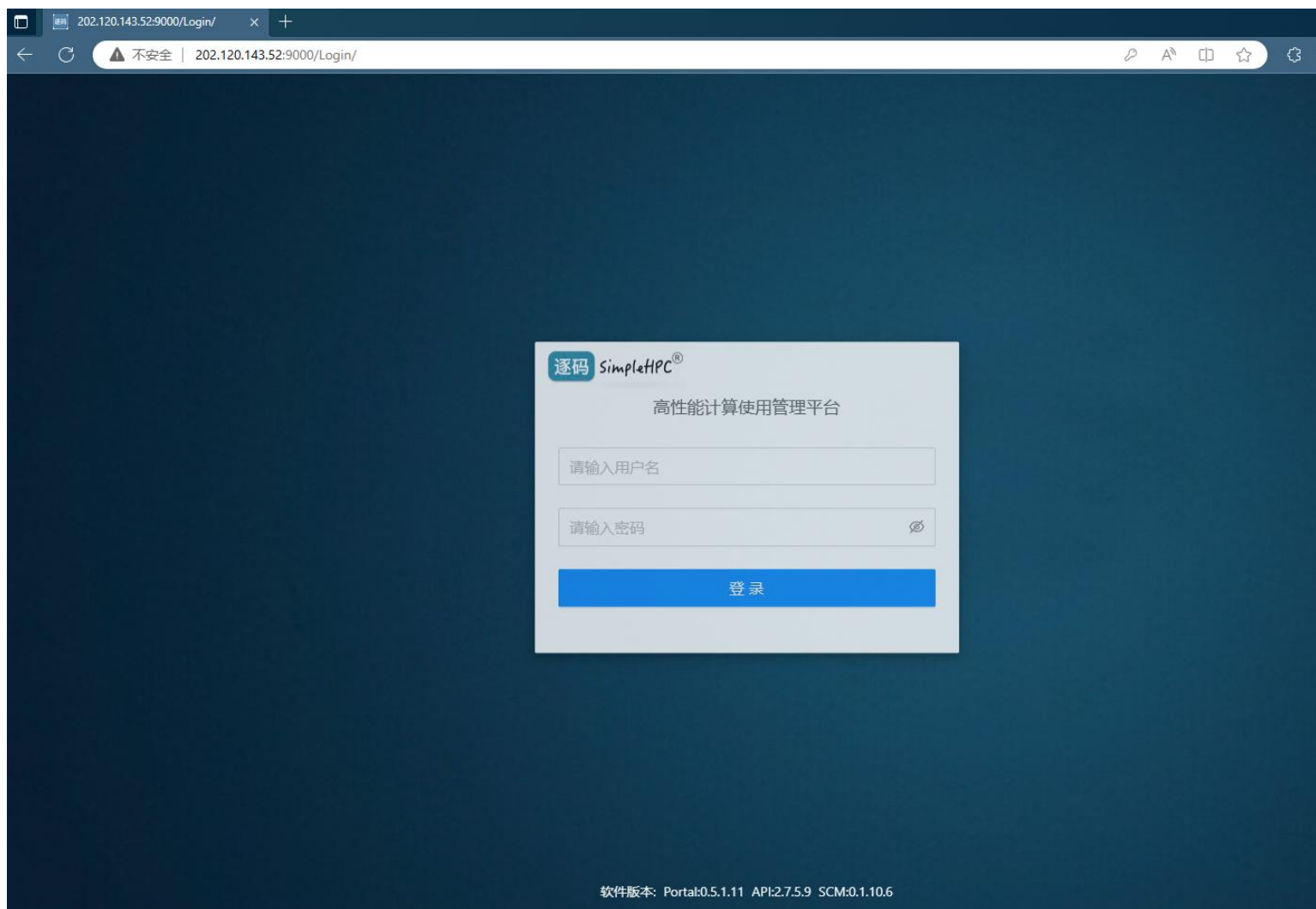
## Software

### 公共conda环境

- base /cluster/apps/anaconda3/2022.10
- HTSeq /cluster/apps/anaconda3/2022.10/envs/HTSeq
- Multicore-TSNE /cluster/apps/anaconda3/2022.10/envs/Multicore-TSNE
- R-4.2.2 /cluster/apps/anaconda3/2022.10/envs/R-4.2.2
- gcc /cluster/apps/anaconda3/2022.10/envs/gcc
- kallisto /cluster/apps/anaconda3/2022.10/envs/kallisto
- mono /cluster/apps/anaconda3/2022.10/envs/mono
- pyscenic /cluster/apps/anaconda3/2022.10/envs/pyscenic
- python-3.11.0 /cluster/apps/anaconda3/2022.10/envs/python-3.11.0
- pytorch-1.12 /cluster/apps/anaconda3/2022.10/envs/pytorch-1.12
- salmon /cluster/apps/anaconda3/2022.10/envs/salmon
- scanpy /cluster/apps/anaconda3/2022.10/envs/scanpy
- tools-v1 /cluster/apps/anaconda3/2022.10/envs/tools-v1
- trim-galore /cluster/apps/anaconda3/2022.10/envs/trim-galore

# 云平台

## Cloud platform



# 云平台

## Cloud platform

逐码

- 应用列表
- 调度系统
- 计费管理
- 文档管理
- 故障台(测试)

SimpleHPC 高性能计算平台

### 应用列表

请输入应用名

[应用列表](#)

 <p>Jupyter</p> <p>4 cpu</p> <p>4</p> <p>启动   关闭</p>	 <p>Rstudio</p> <p>4 cpu</p> <p>0</p> <p>启动   关闭</p>	 <p>matlab</p> <p>4 cpu</p> <p>0</p> <p>启动   关闭</p>
---	---	--

# 云平台-RStudio

## Cloud platform (RStudio)

RStudio Server — Mozilla Firefox

172.51.99.158:9000/app/rstudio/e1fc6dbc-12dd-4c3f-b38f-7de66d5f5ac8/

File Edit Code View Plots Session Build Debug Profile Tools Help pm Project: (None)

Console Terminal Background Jobs

```
R 4.2.1 ~/  
R version 4.2.1 (2022-06-23) -- "Funny-Looking Kid"  
Copyright (C) 2022 The R Foundation for Statistical Computing  
Platform: x86_64-conda-linux-gnu (64-bit)  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
  
Natural language support but running in an English locale  
  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
  
> system("pwd")  
/cluster/home/nm  
> hist(rnc 截图(Alt + A) = 'blue',border='yellow',main='',xlab='')  
Fontconfig error: Cannot load default config file: No such file: (nul  
l)  
> hist(rnorm(200),col='green',border='black',main='',xlab='')  
>
```

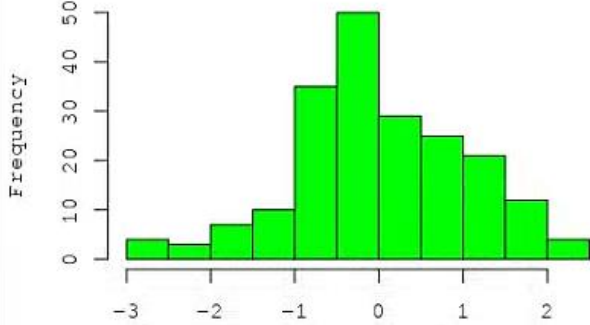
Environment History Connections Tutorial

138 MiB

Environment is empty

Files Plots Packages Help Viewer Presentation

Zoom Export



Bin Range	Frequency
-3.0 to -2.5	5
-2.5 to -2.0	3
-2.0 to -1.5	8
-1.5 to -1.0	12
-1.0 to -0.5	35
-0.5 to 0.0	50
0.0 to 0.5	28
0.5 to 1.0	25
1.0 to 1.5	20
1.5 to 2.0	12
2.0 to 2.5	5

# 云平台-Jupyter

## Cloud platform (Jupyter)

JupyterLab — Mozilla Firefox

202.120.143.52:9000/app/xuwanxing/lab

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last Modified
aliyunpan	8 months ago
analysis	a year ago
apache_related	2 years ago
apache-log4j-2.14.1-src	2 years ago
changepassword-0.9	2 years ago
downloads	a month ago
Downloads	8 months ago
edqeturbo-client	7 months ago
edirect	a year ago
ega_related	8 months ago
filtered_gene_bc_matrices	7 years ago
GenomeAnalysisTK-3.8-0-ge9d...	6 years ago
gmp-6.2.1	2 years ago
mono_test	7 days ago
nettle-3.7.3	2 years ago
Packages	a year ago
perl5	a year ago
Practice	2 years ago
Python-3.9.7	2 years ago
R	2 years ago

Launcher

Notebook

Python 3 (ipykernel)

Console

Python 3 (ipykernel)

Other

Terminal

Text File

Markdown File

Python File

Show Contextual Help

# 云平台-调度系统

## Cloud platform (SLURM)

