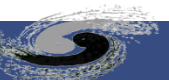
高能所计算平台培训

姜晓巍
代表 计算中心
中国科学院高能物理所



主要内容

- 计算平台简介
- 账号及登录
- 文件存储
- 作业系统
- 常见FAQ
- 典型使用示例

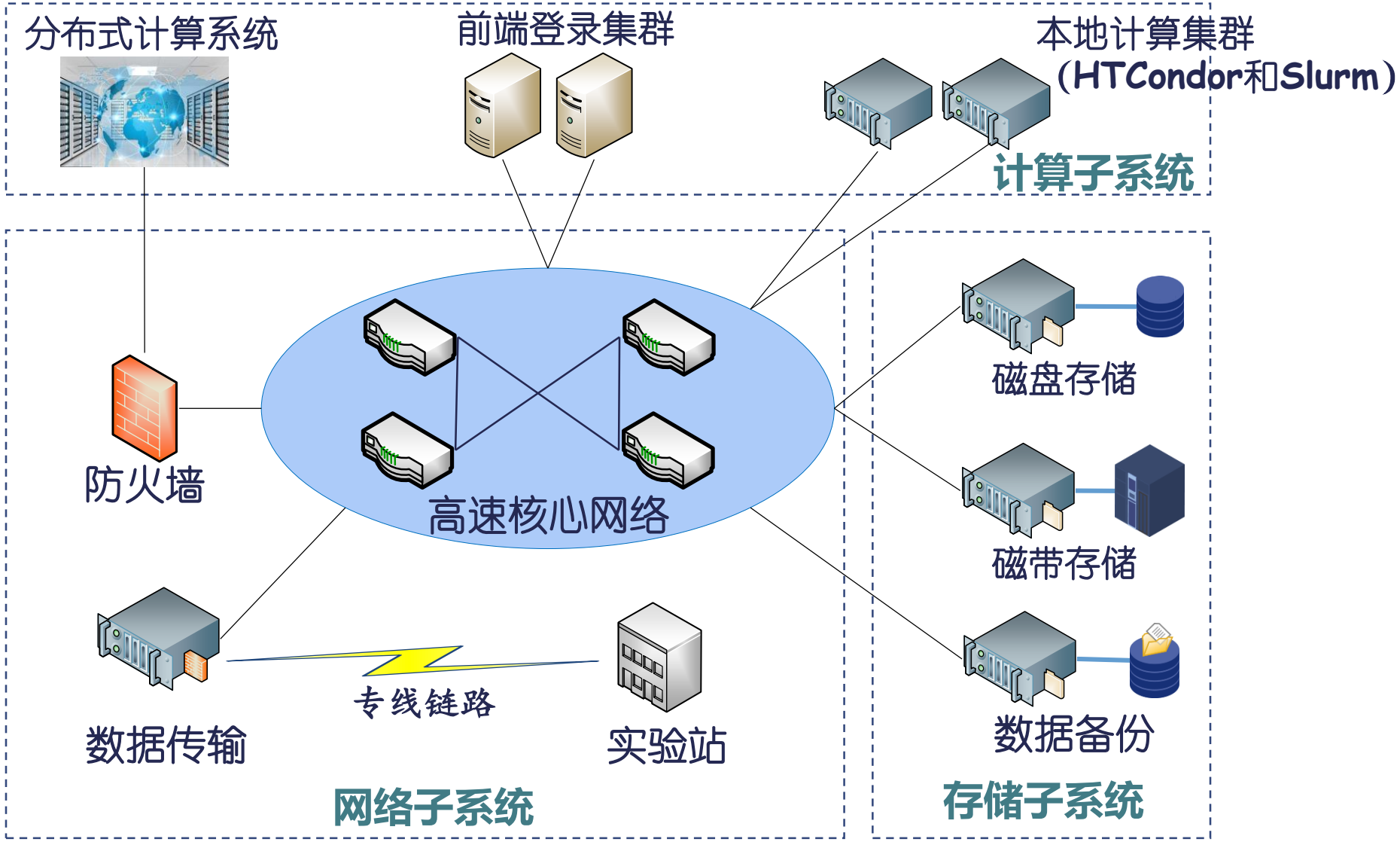
主要内容

- **计算平台简介**
- 账号及登录
- 文件存储
- 作业系统
- 常见FAQ
- 典型使用示例

计算平台概况

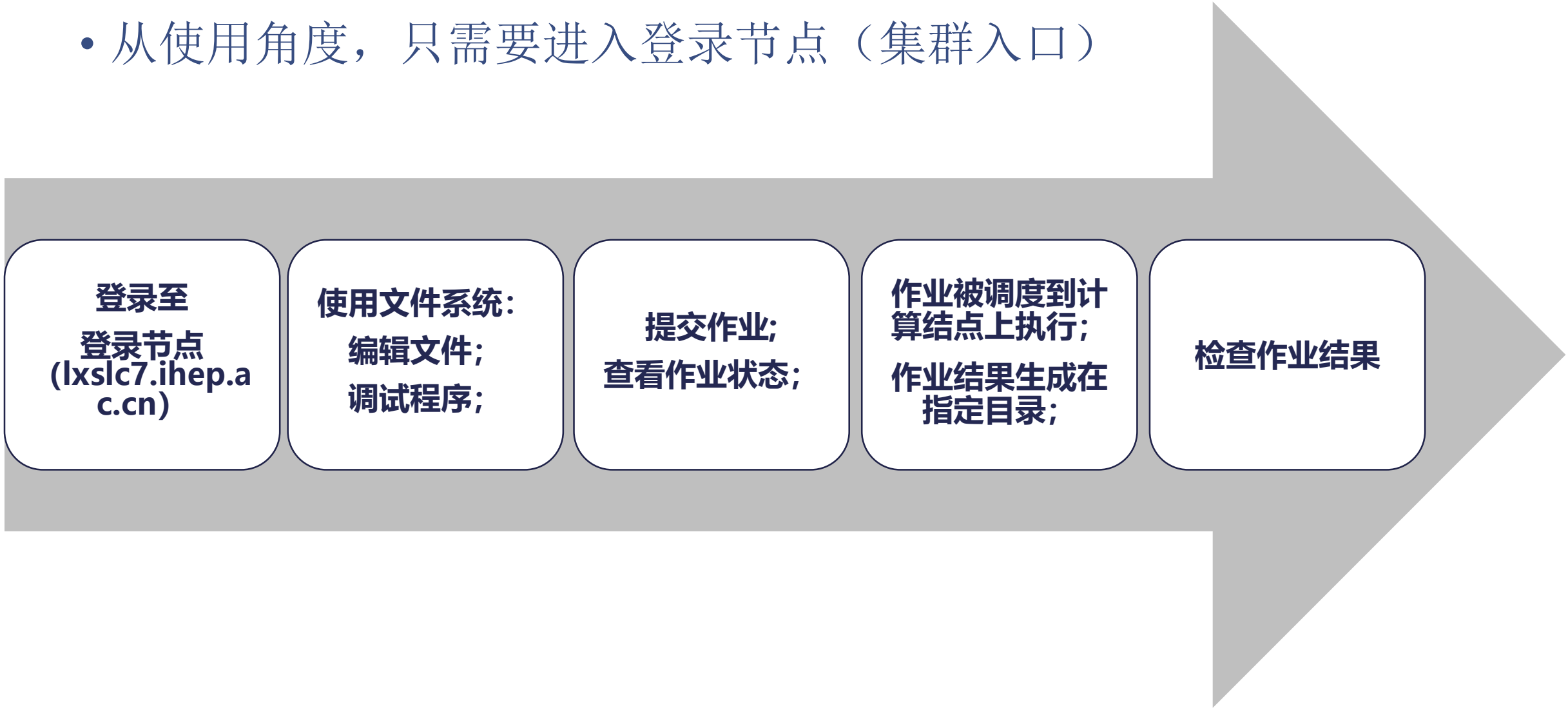
- 服务高能所10+个实验与应用
  - BES、JUNO、LHAASO、HXMT、LQCD、HEPS、CEPC、ATLAS、CMS、LHCb…
- 拥有约33000 CPU 核和190GPU卡
  - 支持高通量计算（HTC）和高性能计算（HPC）两种本地计算模式
  - 网格计算（WLCG二级站点）和分布式计算（Dirac和dHTC）
- 拥有约44PB 的磁盘空间和 21PB 的磁带存储系统
  - 磁盘存储（Lustre和EOS）
  - 磁带存储（Castor& EOS CTA）
- 网络
  - 数据区带宽为400Gbps，支持以太网/IB网络，支持IPV4/V6双栈网络
  - 出口带宽双栈共40Gbps，是LHCONE成员

计算平台架构


分布式计算系统
前端登录集群
本地计算集群
(HTCondor和Slurm)
计算子系统
防火墙
高速核心网络
磁盘存储
磁带存储
数据备份
专线链路
数据传输
实验站
网络子系统
存储子系统

基本使用流程

- 从使用角度，只需要进入登录节点（集群入口）

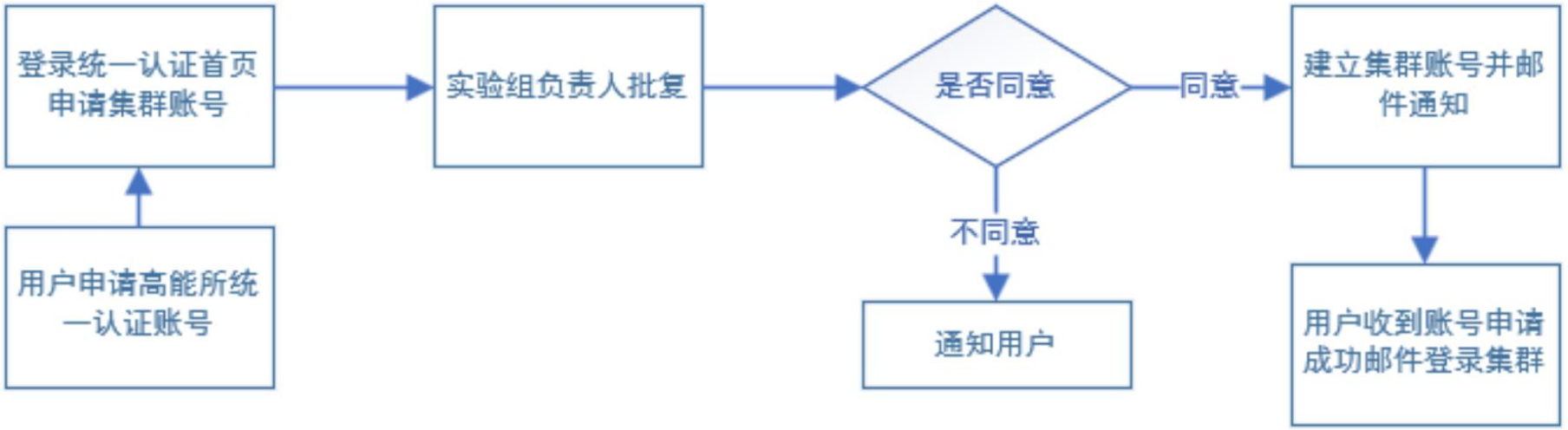


主要内容

- 计算平台简介
- **账号及登录**
- 文件存储
- 作业系统
- 常见FAQ
- 典型使用示例

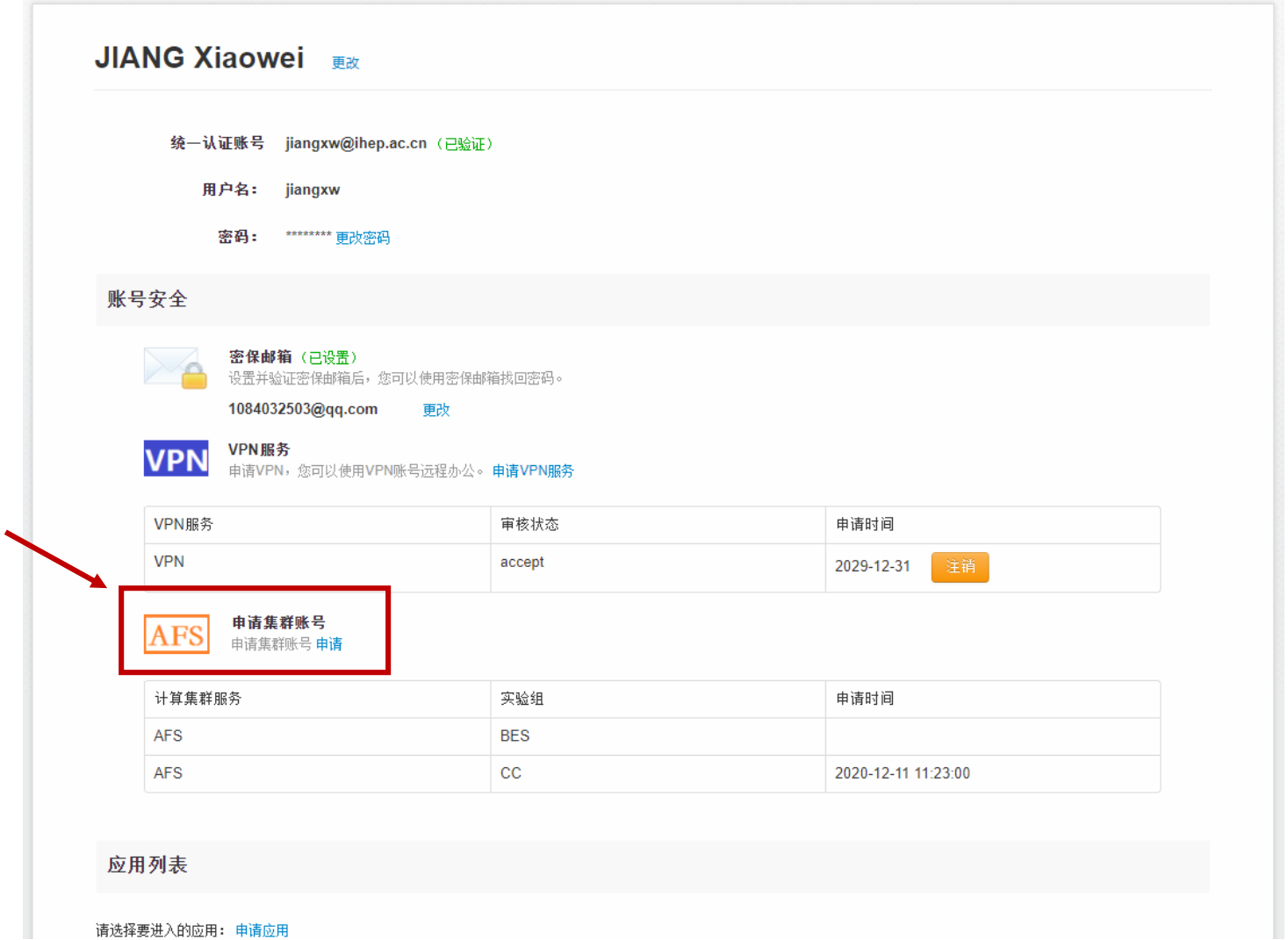
计算平台账号管理

- 计算平台每位计算用户均隶属于一个或几个实验，只有在得到各个实验计算负责人批准后，该用户才能拥有计算平台的个人账号（AFS账号），使用各个实验的计算及存储资源。



账号申请（1）

- 统一认证用户申请集群账号（https://login.ihep.ac.cn/）



账号申请（2）

账号登录后，默认的shell类型，建议选bash

账号隶属应用，用于判断账号的隶属关系

账号所属用户组，对应linux group，主要用于判断存储和计算资源使用权限

密码管理（仅限于近期使用）

- 密码遗忘&重置密码（此方式即将关闭，未来仅在统一认证界面中密码修改）
  - http://afsapply.ihep.ac.cn/ccapply/userfindpasswd.action
- 密码修改
  - 用户使用账号密码登录节点lxslc7.ihep.ac.cn，使用命令kpasswd命

```
-bash-4.1$ kpasswd
Password for username@IHEPKRB5:      #输入当前密码
Enter new password:      #输入新密码
Enter it again:      #输入新密码
```

```
$ kinit       # 初始化认证
$ aklog       # 获取tokens，默认有效期为25小时
$ tokens      # 查看tokens信息
```

注意事项

- 用户申请账号填写的手机号码、电子邮件地址必需真实有效
- 为了个人账号安全，请使用强密码格式。计算平台规定用户密码长度不得少于10位，且必须包含字母、数字或特殊字符中的任意两种。不符合上述要求的密码将不被系统接受。
- 账号创建成功后，用户会收到通知邮件
- 在密码到期前的30天、7天和2天，用户将会分别收到三次邮件提醒
- 用户账号的信息如果发生改变，请及时与计算中心联系更新

登录集群

- 登录集群使用lxslc7.ihep.ac.cn
  - 系统依据当前节点负载情况，自动分配合适的登录节点
  - 避免使用具体登录节点域名，如（lxslc701.ihep.ac.cn）

容器使用（登录节点）

- 满足用户使用多种操作系统版本及环境的需求
- 镜像查看

```
$ hep_container images
```

- 支持用户组查看 – 容器中仅可访问用户组相关存储目录

```
$ hep_container groups
```

- 进入容器环境

```
$ hep_container shell SL5
```

- 直接使用容器执行操作

```
$ hep_container exec SL5 cat /etc/redhat-release
```

主要内容

- 计算平台简介
- 账号及登录
- **文件存储**
- 作业系统
- 常见FAQ
- 典型使用示例

文件存储

- 计算中心为各个实验组和个人用户提供多种级别的文件存储服务
- HOME目录
  - AFS文件存储：个人文件存放
- 软件存储
  - CVMFS文件存储：主要用于存放实验软件
- 数据存储
  - LUSTRE文件存储：主要用于存放海量实验数据；个人数据目录；临时目录
  - EOS文件存储：主要用于存放海量实验数据（LHAASO，JUNO,HXMT）

AFS存储

- 用户卷
  - 用户成功登录后，默认进入用户home目录
  - 每个用户在个人afs home目录下拥有500MB空间
  - 路径：
    - /afs/ihep.ac.cn/users/a-z/username
- 如果tokens过期，将无法读写；需要更新tokens，获得权限
  - kinit; aklog
- 登录节点可读写，作业在计算节点运行对afs无写权限
  - 尽量不使用该目录作为作业提交目录

AFS存储常用操作

- 设置访问权限

```
$ fs setacl -dir /afs/ihep.ac.cn/users/h/huangql/mydir -acl huqb all
```

  - 只有afs命令设置的访问权限有效
  - Linux的访问权限设置（例如chmod 400)对AFS文件无效

- 查看目录Quota

```
$ fs listquota /afs/ihep.ac.cn/users/h/huangql/mydir
```

- 查看目录空间使用情况

```
$ fs quota /afs/ihep.ac.cn/users/h/huangql
```

- 用户tokens获取及延期

```
$ kinit username
Password:
$ aklog
```

LUSTRE存储（1）

- 提供BES，DYB，JUNO等多个实验海量数据存储服务
  - /scratchfs存放临时文件
  - /workfs2专用于用户交互使用，登录节点可读写，计算节点无写权限

	总空间（TB）	用途	用户限额	备份情况
/besfs5	1800	BESIII生产数据，BESIII group数据，用户数据	users目录每人50GB使用空间，group目录各组不同	原始数据有磁带备份，其它无备份
/bes3fs	1900	BESIII 生产数据		原始数据有磁带备份，其它无备份
/bes3fs	1900	BESIII 生产数据		原始数据有磁带备份，其它无备份
/besfs3	2600	BESIII 生产数据		原始数据有磁带备份，其它无备份
/besfs4	2500	BESIII 生产数据		原始数据有磁带备份，其它无备份

LUSTRE存储（2）

/publicfs	3000	ATLAS,CMS,LHCB,UCAS分池共享数据盘	每人5TB使用空间，30万文件数	各组资源完全隔离使用，无备份
/sharefs	1500	Alicpt,BES,HBKG,HEPS,MBH分池共享数据盘		各组资源完全隔离使用，无备份
/scratchfs	572	用户临时文件	每人500GB使用空间，20万文件数	无备份
/workfs2	22	用户个人文件（推荐保存重要文件或结果）	每人5GB使用空间，5万文件数	全盘备份
/cefs	2500	CEPC 实验数据，用户数据		无备份
/junofs	3100	JUNO 实验数据	每人500GB使用空间，30万文件数	无备份
/dybfs	2500	DYB 实验数据，用户数据	每人1TB使用空间，30万文件数	原始数据有磁带备份，其它无备份
/dybfs2	1400	DYB 实验数据，用户数据	每人1TB使用空间，30万文件数	原始数据有磁带备份，其它无备份
/gecamfs	1500	GECAM 实验数据		无备份
/hxmtfs	1300	HXMT 实验数据		无备份
/hpcfs	1800	Slurm集群GPU应用数据		无备份
/lhaasofs	610	LHAASO 用户数据	每人200GB使用空间，50万文件数	无备份

LUSTRE常用操作（1）

- 查看用户资源配额

```
$ lfs quota -u zhangsan -h /publicfsDisk
# 命令输出如下：（已用空间）（软空间配额）（硬空间配额）（已存文件数）（软文件数配额）
quotas for user zhangsan (uid XXXX): Filesystem kbytes quota limit grace files quota limit
grace /publicfs 3.3G 5G 5G - 232010 300000 300100
```

- Project空间配额（只适用于/lhaasofs/user与/besfs5）
  - 限制指定目录空间使用
  - 在特定目录下，对任意一个普通文件（非目录）执行命令：

```
$ lsattr -p nohup.out
1097 -----------------P nohup.out
```

  - 运行命令查看目录配额使用情况

```
# lfs quota -p 1097 -h /besfs5
Disk quotas for prj 1097 (pid 1097):
    Filesystem    used   quota   limit   grace   files   quota   limit   grace
       /besfs5  1.362T*     50G     50G       -  293249       0       0       -
```

LUSTRE常用操作（2）

- 设置目录的访问控制（ACL）
  - 设置访问控制

```
$ setfacl -m user:wanglu:rwx /besfs4/wanglutest
```

  - 查看acl权限

```
$ getfacl /besfs4/wanglutest
```

  - 删除acl权限

```
$ setfacl -x user:wanglu /besfs4/wanglutest
```

  - 恢复linux原来的权限设置

```
$ setfacl -b /besfs4/wanglutest
```

注意：

1. 只有目录的属主可以操作目录的ACL权限
2. 设置过ACL权限后，ls -l 目录会多一个”+”，此时，Linux原来的permission规则会失效
3. 可以对一个组添加ACL setfacl -m group:xxx /xxx/xxx

EOS存储

- 目前提供LHAASO、HXMT、JUNO等实验的海量数据存储服务。

实例名	挂载点	实例服务器地址	总空间	用途
LHAASO实验	/eos	eos01.ihep.ac.cn	14.18 PB	LHAASO本地实验数据
LHAASO稻城	/eos/daocheng	lhmteos01.ihep.ac.cn	2.54 PB	LHAASO稻城快速重建
HXMT实验	/mnt/hxmt	hxmteos01.ihep.ac.cn	806.22 TB	HXMT实验数据
JUNO实验	/eos/juno	junoeos01.ihep.cn	979.74TB	JUNO实验

EOS存储常用操作（1）

- 仅在登录节点交互使用(FUSE方式)
- 访问方式(xrootd方式)

```
$ xrdfs root://eos01.ihep.ac.cn  ls /eos/user/file.txt
```

```
$ eos root://eos01.ihep.ac.cn  ls /eos/user/file.txt
```

- EOS_MGM_URL环境变量：实例服务器地址
  - 在北京集群查看：

```
$ echo $EOS_MGM_URL
root://eos01.ihep.ac.cn
```

  - 如果不存在，可自行设置：

```
$ export EOS_MGM_URL=root://eos01.ihep.ac.cn
```

EOS存储常用操作（1）

- 访问方式(xrootd方式)
  - 作业程序（C++）

```
TFile *filein = TFile::Open("root://eos01.ihep.ac.cn//eos_absolute_path_filein_name.root")
 或
TFile *fileout = TFile::Open("root://eos01.ihep.ac.cn//eos_absolute_path_fileout_name.root")
```

  - 非root格式文件参考：
    - http://afsapply.ihep.ac.cn/cchelp/zh/local-cluster/storage/EOS/

注意：打开的文件使用后，应使用TFile:Close()及时关闭

EOS存储常用操作（2）

- 查看资源配额情况

```
$ eos quota /eos/user/z/zhangsan
```

- 输出

```
$ eos quota /eos/user/z/zhangsan
By user ...
# ____________________________________________________________________________________________________
# ==> Quota Node: /eos/user/z/zhangsan/
# ____________________________________________________________________________________________________
user       used bytes logi bytes used files aval bytes aval logib aval files filled[%]  vol-status
ino-status
zhangsan  800 GB  800 GB  10 k-   1.00 TB    200GB    25M -  9.08       ok         ok
         (已用空间)           (已使用文件数)  (空间配额) (文件数配额)
By group ...
# ____________________________________________________________________________________________________
# ==> Quota Node: /eos/user/z/zhangsan/
# ____________________________________________________________________________________________________
# ....................................................................................................
group      used bytes logi bytes used files aval bytes aval logib aval files filled[%]  vol-status
ino-status
u07        800 GB  800 GB  10 k-   0 B        0 B        0 -        100.00     ignored    ignored
```

EOS存储常用操作（3）

- 查看回收站文件

```
$ eos recycle ls
```

- 清空回收站中的文件

```
$ eos recycle purge
```

- 恢复回收站中的某个文件

```
$ eos recycle restore 0000000008b0f7bf
```

注意：目前/eos回收站中的文件只保留3天时间。

CVMFS存储（只读）

- 软件卷

路径	用途
/cvmfs/bes.ihep.ac.cn	提供bes所需的软件库
/cvmfs/bes3.ihep.ac.cn	提供bes3所需的软件库
/cvmfs/cepc.ihep.ac.cn	提供cepc实验所需的数据分析软件
/cvmfs/exo.ihep.ac.cn	提供exo实验所需的数据分析软件
/cvmfs/dcomputing.ihep.ac.cn	提供分布式计算实验所需的数据分析软件
/cvmfs/gluex.ihep.ac.cn	提供gluex实验所需的数据数据分析软件
/cvmfs/hxmt.ihep.ac.cn	提供HXMT实验所需的数据数据分析软件
/cvmfs/heps_ap.ihep.ac.cn	提供HEPS实验所需的数据数据分析软件
/cvmfs/juno.ihep.ac.cn	提供juno实验所需的数据分析软件
/cvmfs/lhaaso.ihep.ac.cn	提供lhaaso实验所需的数据分析软件
/cvmfs/lqcd.ihep.ac.cn	提供lqcd实验所需的数据分析软件
/cvmfs/mlgpu.ihep.ac.cn	提供gpu机器学习相关软件
/cvmfs/raq.ihep.ac.cn	提供raq实验所需的数据分析软件

- 公共卷

路径	用途
/cvmfs/common.ihep.ac.cn	存储系统组公共脚本和软件

CVMFS存储（只读）

- 如果需要安装应用软件，请联系计算中心

备份

- 如需恢复数据：发送具体的目录名、文件名及恢复的日期到 helpdesk@ihep.ac.cn

应用	目录	备份策略
BES	/home/bes	每天一次备份，可恢复一个月之内的数据
BSRF	/home/bsrf	每天一次备份，可恢复两周之内的数据
LHC	/home/lhc	每天一次备份，可恢复两周之内的数据
CC	/home/cc	每天一次备份，可恢复两周之内的数据
ATLAS	/publicfs/atlas/codesbackup	每天一次备份，可恢复两周之内的数据
ATLAS	/afs/ihep.ac.cn/soft/atlas	每周一次备份，可恢复一个月之内的数据
CMS	/afs/ihep.ac.cn/soft/CMS	每周一次备份，可恢复一个月之内的数据
公共目录	/afs/ihep.ac.cn/users	每天一次备份，可恢复两周之内的数据
公共目录	/workfs2	每天一次备份，可恢复一个月之内的数据

存储目录组织(1)

- 针对使用文件系统的几点建议（1）
  - 单一目录下不要有过多(几万以上)数据或脚本等文件，应按照一定规律创建子目录，将文件放在子目录下，**单目录文件数量控制在3000以内**。
  - **作业中避免使用ls *或rm ***之类的操作；如果只需查看文件名信息，可以使用/bin/ls代替ls命令，可以加快速度；如果需要查看/eos目录，则使用“eos ls 目录绝对路径”，速度会更快。
  - 任务脚本写成一个模板，将要分析的文件名、数据目录和其他程序参数**作为脚本参数在用 hep_sub 提交时传递给脚本**，例如 hep_sub my_job.sh -argu "aaa bbb"，而不用生成许多类似的脚本，如my_job_aaa.sh、my_job_bbb.sh。

存储目录组织(2)

- 针对使用文件系统的几点建议（2）
  - 可以直接将生成的数据文件存放在 eos 上，建议使用xrootd方式读写文件
  - 避免使用类似“hadd *.root”，*.root文件数量过多的前提下，该操作对文件系统压力会非常大。建议提前将需要合并的文件生成到一个列表里，直接遍历特定文件；而对于存放在EOS上的root文件，支持xrootd方式访问，。
  - 文件尺寸不大的个人程序文件(几MB) 如 my_program，在作业脚本中用eos cp命令将程序文件复制到运行节点的/tmp/目录下，在复制之前可以先判断是否存在，不存在时再复制，然后运行/tmp/my_program my_para。

主要内容

- 计算平台简介
- 账号及登录
- 文件存储
- **作业系统**
- 常见FAQ
- 典型使用示例

作业系统

- HPC与HTC集群
- HTCondor
  - HTCondor集群支持高通量计算，采用HTCondor作为负载管理系统，绝大多数作业为单核或单节点作业
- Slurm
  - Slurm集群支持高性能作业（High Performance Computing），采用Slurm作为负载管理系统，绝大多数作业为多核并行、GPU作业

HTCondor作业

- 高能所计算集群针对HTC作业提供一个作业前端工具HepJob，封装了HTC集群的优化，如非特殊需求，建议只使用HepJob命令
- 环境设置作业准备
  - 加载HepJob环境

```
# bash用户
$ export PATH=/afs/ihep.ac.cn/soft/common/sysgroup/hep_job/bin:$PATH
# tcsh用户
$ setenv PATH /afs/ihep.ac.cn/soft/common/sysgroup/hep_job/bin:$PATH
```

  - 作业脚本需有可执行权限

```
# 查看作业脚本是否有可执行权限
$ /bin/ls -l job.sh
-rw-r--r-- 1 jiangxw u07 85 Aug 29 18:23 job.sh

# 赋予作业脚本可执行权限
$ /bin/chmod +x job.sh
```

HTCondor作业常用命令（1）

- 作业提交

```
$ hep_sub job.sh
```

- 作业查询

```
$ hep_q -u <username>
```

- 作业删除

```
$ hep_rm 3745232 3745233.0
```

- 挂起作业释放

```
$ hep_release 3745233.0
```

- 修改作业需求

```
$ hep_edit 3745233.0 -m 8000
```

HTCondor作业常用命令（2）

- 按组查询作业时长限制

```
$ hep_clus -g juno --walltime
```

实验	短作业(short)时长限制(小时)	普通作业时长限制(小时)	mid作业时长限制(小时)及资源使用量限制(百分比)
BES	<0.5	<40	<100:10%
JUNO	<0.5	<20	<100:10%
DYW	<0.5	<10	<100:10%
CEPC	<0.5	<10	<100:10%
ATLAS	<0.5	<10	<100:10%
CMS	<0.5	<10	<100:10%
HXMT	<0.5	<14	<100:10%
GECAM	<0.5	<24	<100:10%
LHCb	<0.5	<100	
LHAASO	<0.5	<15	<100:10%

注意，未设置mid作业的实验，默认提交mid作业时，资源使用量限制为1.

在作业中获取作业信息

- 获取作业ID

```
#!/bin/bash
JobId=$_CONDOR_IHEP_JOB_ID
```

- 获取运行节点

```
#!/bin/bash
ExecWorkNode=$_CONDOR_IHEP_REMOTE_HOST
```

- 获取作业提交时间

```
#!/bin/bash
SubmissionTime=$_CONDOR_IHEP_SUBMISSION_TIME
```

提交大内存作业

- 默认情况下，作业分配资源不会考虑内存大小（随机分配）
- 如果作业有大内存特殊需要，使用-mem参数指定内存，命令如下：

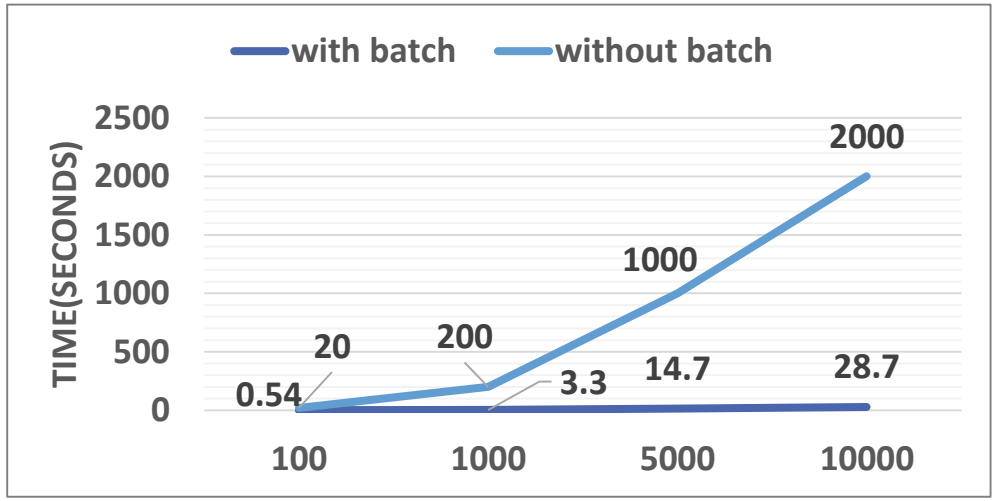
```
$ hep_sub -mem 3000 job.sh
```

- 其中，-mem参数值单位为MB，实例中3000表示3GB内存
- 注意，大内存节点相对较少，尽量控制单作业的内存使用

批量作业提交(1)

- 批量作业提交，即一次性提交多个作业，可支持一次性提交大量的作业。参数-n指定一次性批量提交的作业数量。
- 示例1：
  - 已有作业脚本

```
real_job_20191204_0.sh
real_job_20191204_10.sh
real_job_20191204_1.sh
real_job_20191204_2.sh
real_job_20191204_3.sh
real_job_20191204_4.sh
real_job_20191204_5.sh
real_job_20191204_6.sh
real_job_20191204_7.sh
real_job_20191204_8.sh
real_job_20191204_9.sh
```



- 因为这些作业脚本名字里都是这类格式 real_job_20191204_*.sh，且关键字符是从0开始递增的数字，提交这些作业只需要运行：

```
$ hep_sub real_job_20191204_"%{ProcId}".sh -n 11
```

批量作业提交(2)

- 示例2：
  - 1.已有脚本

```
real_job_20191201.sh
real_job_20191202.sh
real_job_20191203.sh
real_job_20191204.sh
real_job_20191205.sh
real_job_20191206.sh
real_job_20191207.sh
...
real_job_20191230.sh
real_job_20191231.sh
```

  - 2.额外准备脚本

```
#!/bin/bash

# get procid from command line
procid=$1

# map 0,1,2,...,30 to 1,2,3,...,31
sub_name_number=`expr $procid + 1`

# format 1,2,3,...,31 to 01,02,03,...,31
sub_name=`printf "%02d\n" $sub_name_number`

# run the real job script by the formatted file name
bash real_job_201912"${sub_name}".sh
```

  - 3.运行命令提交作业

```
$ hep_sub real_job_parent.sh -argu "%{ProcId}" -n 31
```

批量作业提交(3)

- 示例3：1.已有脚本

```
abcd.sh efgh.sh ijkl.sh mn.sh opq.sh rst.sh uvw.sh xyz.sh
```

- 2.额外生成脚本 00_htc_parent_job.sh

```
#!/bin/bash

procid=$1

job_scripts=(abcd.sh efgh.sh ijkl.sh mn.sh opq.sh rst.sh uvw.sh xyz.sh)

"/scratchfs/cc/jiangxw/tmp/""${job_scripts[$procid]}"
```

- 用于生成00_htc_parent_job.sh的脚本可参考：
  - http://code.ihep.ac.cn/shijy/ComputingCluster/-/blob/master/batch_submission/create_parent_script.sh
- 3.运行命令提交作业

```
$ hep_sub 00_htc_parent_job.sh -argu "%{ProcId}" -n 8
```

SLURM

- 申请权限
  - 提交作业前，用户需建立好AFS账号，账号申请页面是：申请页面
  - 账号申请成功后，上述组别的成员请分别向各组别计算负责人发送邮件申请slurm集群使用授权。未经授权，报错如下：

```
sbatch: error: Batch job submission failed: Invalid account or account/partition
combination specified
```

  - 经计算负责人与集群管理员授权后，用户方可提交作业至集群中运行。

SLURM作业常用命令

- 作业提交

```
$ sbatch slurm_sample_script_1.sh
```

- 作业查询

```
$ squeue 或 sacct -u <user_name>
```

- 作业删除

```
$ scancel <job_id>
```

- 查看集群状态

```
$ sinfo
```

SLURM作业准备

- 样例：/cvmfs/slurm.ihep.ac.cn/slurm_sample_script

GPU作业

```
#! /bin/bash
########## Part 1 #######
#SBATCH --partition=gpu
#SBATCH --qos=normal
#SBATCH --account=lqcd
#SBATCH --job-name=gres_test
#SBATCH --output=job-%j.out
#SBATCH --ntasks=2
#SBATCH --mem-per-cpu=2048
#SBATCH --gres=gpu:v100:2
########## Part 2 #######
echo "hello world!"
```

队列信息

资源信息

GPU作业

CPU作业

```
#! /bin/bash
########## Part 1 #######
#SBATCH --partition=mbh
#SBATCH --qos=regular
#SBATCH --account=mbh
#SBATCH --job-name=mbh_test
#SBATCH --output=job-%j.out
#SBATCH --ntasks=20
#SBATCH --mem-per-cpu=2048
########## Part 2 #######
echo "hello world!"
```

SLURM-GPU作业

- 应用组
  - lqcd,gpupwa,junogpu,mlgpu,higgs,bldesign
- 各组的资源分区（partition）、作业队列（QOS）、计算节点

partition（节点分区）	qos（队列）	group	资源限制	节点资源
lgpu	long	lqcd	QOS long - 作业运行时间不超过**30天** - 每组作业数量（运行+排队）不超过**64个** - 每个作业每CPU核最大可使用**40GB**内存	- 1个节点，每个节点384GB 内存 - 共8张GPU卡，36个CPU核
gpu	normal, debug	lqcd, gpupwa, junogpu, mlgpu,higgs	QOS normal - 每个作业运行时间不超过**48小时** - 每组作业数量（运行+排队）不超过**512个**，每组可使用的GPU卡数量不超过**128**张 - 每个用户作业数量（运行＋排队）不超过**96个**，每用户可使用的GPU卡数量不超过**64**张 - 每个作业每CPU核最大可使用**40GB**内存 QOS debug - 作业运行时间不超过**15分钟** - 每组作业数量（运行＋排队）不超过**256个**，每组可使用的GPU卡不超过**64**张 - 每个用户作业数量（运行＋排队）不超过**24个**，每个用户可使用的GPU卡不超过**16**张 - 每个作业每CPU核最大可使用**40GB**内存 - QOS debug 优先级**高于** QOS normal优先级	- 23个节点，每个节点384GB 内存 - 共182张GPU卡, 840个CPU 核

SLURM-CPU作业

- 应用组
  - mbh,bio,cac,nano,heps,cepcmpi,alicpt,bldesign,raq
- 各组的资源分区（partition）、作业队列（QOS）、计算节点

partition(节点分区)	QOS（作业队列）	account / group（组别）	worker nodes（计算节点）
mbh,mbh16	regular	mbh	16个节点，共256个CPU核
cac	regular	cac	8个节点，共384个CPU核
nano	regular	nano	7个节点，共336个CPU核
bioq	regular	bio	16个节点，共256个CPU核
biofastq	regular	bio	12个节点，共288个CPU核
heps	regular,advanced	heps	34个节点，共1224个CPU核
hepsdebug	hepsdebug	heps	1个节点，共36个CPU核
cepcmpi	regular	cepcmpi	36个节点，共1696个CPU核
ali	regular	alicpt	16个节点，共576个CPU核
bldesign	blregular	bldesign	3个节点，共108个CPU核
raq	regular	raq	12个节点，共672个CPU核

SLURM-CPU作业

- 队列资源使用限制

QOS	作业最大运行时间	可提交的最大作业数量	优先级
regular	60天	每个用户4000个，每个组8000个	低
advanced	60天	-，-	高
hepsdebug	30分钟	每个用户10个，-	中
blregular	30天	每个用户200个，每个组1000个	低

主要内容

- 计算平台简介
- 账号及登陆
- 文件存储
- 作业系统
- **常见FAQ**
- 典型使用示例

常见FAQ

- 有问题时如何寻求帮助
  - 电话服务支持（工作时间）：88236855
  - 发送邮件咨询：helpdesk@ihep.ac.cn
    ihep_computing_service@ihep.ac.cn
  - 网页咨询：http://helpdesk.ihep.ac.cn （推荐）

常见FAQ（集群账号）

- 登录lxslc7.ihep.ac.cn时，无法编辑文件，出现“.Xauthority does not exist 或 ”unauthorized“的报错
- 解决：
  - 更新tokens，再尝试删除

```
-bash-4.2$ kinit huqb
Password for huqb@IHEPKRB5:
-bash-4.2$ aklog -d
Authenticating to cell ihep.ac.cn (server afsdb1.ihep.ac.cn).
Trying to authenticate to user's realm IHEPKRB5.
Getting tickets: afs/ihep.ac.cn@IHEPKRB5
Using Kerberos V5 ticket natively
About to resolve name huqb to id in cell ihep.ac.cn.
Id 10517
Set username to AFS ID 10517
Setting tokens. AFS ID 10517 @ ihep.ac.cn
-bash-4.2$ rm -f ~/.Xauthority
-bash-4.2$ exit
```

常见FAQ（集群账号）

- 我的密码忘记了，如何重置？
  - 访问http://afsapply.ihep.ac.cn/ccapply/userfindpasswd.action
- 我的密码正确，但无法正常登录
  - 访问helpdesk.ihep.ac.cn，发送ticket询问
- 账号过期，如何处理？
  - 发送过期账号信息和延期申请至实验组负责人，抄送ihep_computing_service@ihep.ac.cn，
  - 实验组负责人同意延期后，账号管理员会对账号进行延期操作。

常见FAQ（文件系统）

- **我的目录突然无法正常写入了？**
  - 用户个人目录或是用户组的公共目录都被设置了最大可用份额。当使用空间超过最大可用份额时，相关人员会收到邮件提醒，需要尽快清理目录下文件。
- **我的文件被不小心删除了，还能恢复吗？**
  - 参见备份（第30页）
  - /eos/user有回收站功能，参见（第几页）
- **怎么查看我已经使用的存储空间份额？**
  - AFS存储：`fs quota /afs/ihep.ac.cn/users/z/zhangsan`
  - Lustre：`lfs quota -u zhangsan -h /publicfsDisk`
  - EOS存储：`eos quota /eos/user/z/zhangsan`

常见FAQ（文件系统）

- **运行ls卡住，不能正常显示目录下的文件？**
  - 可能是目录下，文件数过多。例如：有用户同一目录下，存放80万个文件
  - 尽量少使用全部匹配这类规则，如“ls *” 或 “rm *”
  - 推荐使用/bin/ls代替ls，只查看文件名的情况，速度快很多
  - 如果需要查看/eos目录，则使用“eos ls 目录绝对路径”，速度会更快

常见FAQ（作业）

- **我的作业排队很久，还是无法运行**
  - 对于近期内运行过大量作业的用户，调度系统会实时计算并调低其优先级，以保证用户间的公平性
  - 高能所计算集群长期维持于满负荷状态，在个别时期（如作业峰值期、存在高优先级公共服务作业等）资源极度紧张，难免出现长时间排队情况，只能耐心等待
  - 特殊作业（如长作业、大内存作业等）可用资源有限，可能导致较长排队时间
  - 排除前面原因后，可联系管理员寻求帮助

常见FAQ（作业）

- **我需要Scientific Linux 5（或SL6、SL7）的系统环境调试我的程序，但是登录结点只有CentOS7系统，该如何操作？**
  - 因安全问题，我们不再提供SL5等登录节点，但提供SL5/SL6/SL7容器供用户调试软件
  - 如果作业需要使用容器，提交时执行命令：hep_sub job.sh -os SL6
- **查询作业时，显示状态为hold(HTCondor)，Failed(Slurm) 是什么原因？**
  - 最常见原因是向afs、workfs2等目录下写出作业数据或日志，而这些目录在计

```
对HTCondor作业，若确认不是前述原因，可使用下面命令查询作业号为JobID的hold原因：
$ hep_q -i $JobID -hold
或
$ hep_q -u $user -hold
若对前述命令给出的原因说明存在疑问，请保留错误作业并联系管理员寻求帮助。
```

常见FAQ（作业）

- **当发现作业出现问题，希望得到管理员帮助时，需要提供哪些信息？**
  - 请提供尽可能详尽的作业信息，包括但不限于作业号（JobID）、大概的作业提交时间、错误现象及提示、作业日志、作业运行路径和脚本等内容，并尽量保留作业现场不删除。管理员获取的信息越多，问题越容易查找。
  - 提交ticket至helpdesk.ihep.ac.cn

建议及反馈

- 电话：88236855 （工作时间）
- 邮箱：
  - helpdesk@ihep.ac.cn
  - ihep_computing_service@ihep.ac.cn
- 网站：helpdesk.ihep.ac.cn （推荐）

参考来源

- 高能所计算环境使用手册：
  http://afsapply.ihep.ac.cn/cchelp/zh/
- 如有网格计算、分布式计算、虚拟云平台等方面的需求，也可参考该手册

谢谢

Q&A