华东师范大学超算中心云计算平台

用户快速入门

1. 系统结构和配置

师大之云（CLOUD@ECNU）高性能云计算平台，采用曙光 5000A 高性能计算机最新技术，由以下几部分组成：

——64 个刀片计算节点，每片刀片配置 2 颗 EXON E5450 3.0GHz 四核 CPU，16GB 内存；

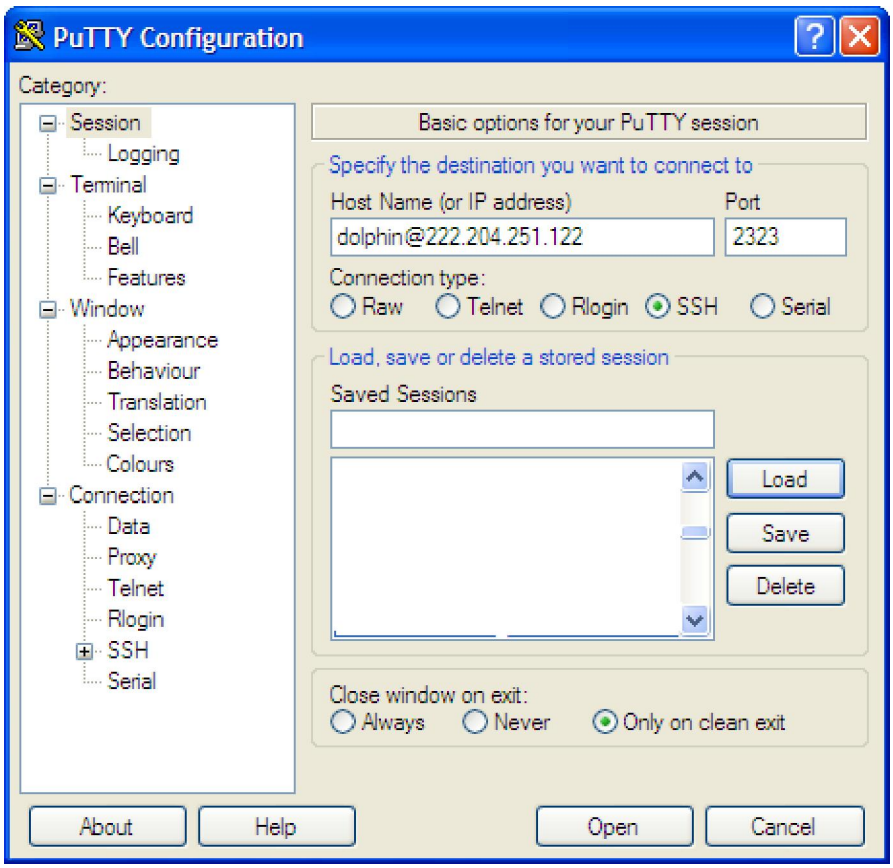
——1 台 Dawning I650r-H 双 CPU 节点作为头结点和 I/O 节点；

——10TB 高速存储；

——20Gbps Infiniband 全线速互联所有节点；

2. 访问师大之云

CLOUD@ECNU 上的账号可以通过发邮件给 jzheng@cc.ecnu.edu.cn，联系郑老师进行申请。有了自己的账号和密码后，可以通过 SSH 登陆到集群的接入节点，方法如下图示：



通常情况下，除了登陆节点和 debug 队列上的节点，不允许用户登录到集群的其它节点。除非作业调度系统通过你正在执行的作业把这些节点分配给你。可以用如下命令得到这些作业调度分配给你的节点列表：

qstat –rn

然后通过

ssh

访问你想要访问的节点。

头结点上的每个交互进程都只有 30 分钟的 CPU 时间限制，除非“ ulimit -a ”有不同的限制。

默认的登陆 SHELL 是/bin/bash，另一个可以切换的 SHELL 是 CSH，可以通过执行 csh 实现切换。

从远程系统上传输文件可以通过 scp [-pr]， 两个可选的选项-p 和-r 分别是“preserve permission”和“recursive copy”（对于目录 copy 来说）的意思，更详细的参考 man scp。

3. 文件系统

每个用户登录后都有一个高速文件系统可供使用，即$HOME 家目录。这个目录是用户登录到 CLOUD@ECNU 后的默认目录。该目录在整个师大之云集群上共享，每个用户有 200GB~1TB 的存储空间，具体情况根据申请时的磁盘 Quota 情况来定。如果需要修改密码，请使用 yppasswd 对自己的用户密码进行修改。

4. 编译器和基础库

除了 GNU gcc/g++/g77/gfortran 编译器外，“师大之云”高性能计算平台还安装了 INTEL F90/C++编译器和 CMKL 数学库。

	GNU	Intel
f77 programs	g77 prog.f	ifort prog.f

f90 programs	gfortran prog.f90	ifort prog.f90
C programs	gcc prog.c	icc prog.c
C++ programs	g++ prog.cpp	icpc prog.cpp

ifort/icc 几个有用的编译选项：

—— -O3 -xW -ip (最高级别的优化，man ifort，man icc 查看)

—— -Vaxlib （可移植的 Fortran 库，像 getargs, etime, ranf 等等）

要使用 CMKL 的 LAPACK/BLAS 数学库函数（文档位于/data/soft/libs/cmkl/9.1/doc/下），可以如下：

-L/data/soft/libs/cmkl/9.1/lib/em64t/ -lmkl_lapack -lmkl_em64t -lguide -lpthread

MPI 程序应该使用 MPI 的封装程序 mpiifort/mpiicc/mpiicpc 来编译和链接，用 mpirun 来运行。

mpiicc -show #显示将要运行的命令

mpiifort -O3 -xW -o prog prog.f90

mpirun -np 8 prog

需要注意的是，上述的 mpirun 将把 8 个进程都发起在登陆节点上，而通过作业调度系统提交则可以把其分发到调度系统分配的节点上去运行。关于作业调度系统的使用，请参考另外两篇文档《华师大高性能集群 MPI 使用情况说明》和《华师大作业调度策略设定及使用情况说明》。

5. 提交作业

曙光的 Gridview 作业调度系统为 CLOUD@ECNU 集群提供了功能强大的作业调度系统。要向 CLOUD@ECNU 集群提交作业，只需创建如下简单的批处理文件，其中包含有 PBS 提示符以及 SHELL 命令，然后通过 qsub 递交即可。

```
#!/bin/bash
#PBS -l walltime=2:00:00
#PBS -l nodes=2:ppn=8
#PBS -j oe
#PBS -q debug

#
#define variables
#
n_proc=$(cat $PBS_NODEFILE | wc -l)

#
#running jobs
#
cd   $PBS_O_WORKDIR    # Change to where the executable "prog" is

#
# Setup the MPI topology
#

time -p /data/soft/compiler/mpi/impi/3.2.2.006/bin64/mpirun --rsh=ssh -env I_MPI_DEVICE
rdma:OpenIB-cma -np ${n_proc} ./prog
exit 0
```

注意：

1）精确的估计程序需要的墙钟时间非常重要，因为调度器的调度策略里短作业比那些需求更多墙钟时间的长作业等待的时间要短。

2）通常情况下，建议大家把一个大作业通过系统或者程序提供的 Checkpoint 功能拆分成若干个更小的独立作业，以避免硬件故障导致的大作业等待时间过长的问题。

6. 监控作业

作业的监控可以通过如下命令进行：

——xpbsmon （图形监控终端）

——qstat [-a] [-f]

——tracejob

——qdel

——showq

——showstart

——showbf

——showres