The Supercomputing Facility for Bioinformatics & Computational Biology, IIT Delhi

User Manual

Dr. B. Jayaram (Professor of Chemistry and Coordinator),
Surojit Bose, Pankaj Sharma, Vidhu S Pandey, Praveen Agrawal

Supercomputing Facility for Bioinformatics & Computational Biology,
3rd Floor, Synergy Building,
IIT Delhi, Hauz Khas,
New Delhi -110 016
Ph: +91-11-26596786
www.scfbio-iitd.org

The resources and facilities provided by the Supercomputing Facility for Bioinformatics & Computational Biology, IIT Delhi (SCFBio) are geared to cater for specific scientific research work in life science area only. Understandably the resources at SCFBio are critical and limited. It is therefore expected that users play a constructive role ensuring a responsible usage conduct of the resources provided by SCFBio.

"Biology easily has 500 years of exciting problems to work on."

-- Donald E. Knuth

"It would appear we have reached the limits of what it is possible to achieve with computer technology, although one should be careful with such statements; they tend to sound pretty silly after 5 years."

-- John von Neumann

"Wherever you go, take your Supercomputer along with you."

-- SCFBio

TABLE OF CONTENTS

1 OVERVIEW 5

2 THE FACILITY 5

3 GETTING STARTED 5

3.1 LOGGING INTO THE REMOTE MACHINE 5
3.2 LOGGING OUT OF THE REMOTE MACHINE 5
3.3 LOGIN PROFILES AND SHELLS 5

4 MANIPULATING FILES AND DIRECTORIES 6

4.1 CHANGING DIRECTORY 6
4.2 COPY FILES 6
4.3 MAKING A LINK 6
4.4 MAKE A NEW DIRECTORY 6
4.5 MOVE / RENAME FILES 7
4.6 PRESENT WORKING DIRECTORY 7
4.7 REMOVE FILES 7
4.8 REMOVE DIRECTORY 7
4.9 LISTING FILES AND DIRECTORIES 7

5 FILE TRANSFER 8

5.1 ESTABLISHING REMOTE CONNECTION 8
5.2 FILE UPLOADING 8
5.3 FILE DOWNLOADING 8
5.4 MAKING A NEW DIRECTORY ON REMOTE HOST 9
5.5 CHANGING DIRECTORY IN LOCAL HOST 9
5.6 CHANGING DIRECTORY IN THE REMOTE HOST 9
5.7 CLOSING THE CONNECTION 9

6 COMPILING & RUNNING PROGRAMS 9

6.1 COMPILING PROGRAMS 9
6.2 RUNNING PROGRAMS 10

7 SUBMITTING JOBS 10

7.1 SUBMITTING SERIAL (SINGLE PROCESSOR) JOBS 10
7.2 SUBMITTING MPI JOBS 11
7.3 CHECK THE STATUS OF SUBMITTED JOBS 11
7.4 REMOVING SUBMITTED JOBS FROM QUEUE 11

8 OPTIMIZATION 12

9 AVAILABLE SOFTWARE 12

9.1 AMBER 8 12
9.2 GAMESS 12
9.3 GROMACS 12

10 FREQUENTLY ASKED QUESTIONS (FAQS) 12

10.1 WHEN I RUN A MPI JOB USING "MPRUN –NP X COMMAND", IT GIVES THE ERROR "MPRUN: MISSING ENVIRONMENT VARIABLE 'JOB_ID': INVALID REQUEST" 12
10.2 MY JOB TAKES A LONG TIME TO START OR AT SOME TIME IT DOESN'T START ALTOGETHER. 12
10.3 HOW DO I ACCESS SCFBIO FROM WINDOW MACHINES? 12
10.4 IN GROMACS SOME COMMANDS ARE SUFFIXED "_MPI" AND MANY OTHERS NOT. 13
10.5 HOW DO I CHECK THE STATUS OF A RUNNING PARALLEL JOB. 14

10.6 I GET THE ERROR MESSAGE TIMEOUT WHILE LAUNCHING A MPI JOB AND IT ASKS ME TO USE '-T' FLAG WITH MPRUN. 14
10.7 I AM A BEGINNER AND I WANT TO LEARN MORE ABOUT UNIX IN GENERAL. CAN YOU PLEASE GIVE SOME REFERENCES? 14
10.8 WHENEVER I DELETE MULTIPLE FILES USING WILD CARD SAY BY RM *, I NEED TO ANSWER 'Y' FOR EACH FILE. 14
10.9 I DON'T LIKE THE DEFAULT EXECUTABLE NAME I.E. A.OUT. HOW CAN I CHANGE IT? 14
10.10 IS THERE ANY AVAILABLE DEBUGGER? 14
10.11 I AM UNABLE TO TELNET/FTP/SSH/SFTP TO ANY OTHER MACHINE FROM THE SUPERCOMPUTER. 14
10.12 WHEN I OPEN A FILE USING VI EDITOR, I FIND ^M AT THE END OF EVERY LINE. 15
10.13 HOW DO I CHECK MY QUOTA STATUS? 15
10.14 I AM UNABLE TO CONNECT TO THE PROVIDED IP ADDRESS AND THE IP ADDRESS ISN'T RESPONDING TO PING REQUEST. 15
10.15 HOW CAN I LEARN MORE ABOUT A COMMAND? 15

1 Overview

This manual is provided for the users of SCFBio as the primary reference. The manual covers the supercomputer configuration, available software and hardware, access methods, and user support. Suggestions for additions or corrections to this manual may please be directed to biogrid@scfbio-iitd.org

2 The Facility

The facility currently hosts a 70 processor (2GB shared memory dual CPU Sun 280R, 900 MHz 64 bit Ultra Sparc III Cu) cluster over a dual network of 1 Gbps & 100 Mbps switched Ethernets. Sun Grid Engine and HPC Cluster Runtime Environment (CRE) software together provide a single window for parallel job creation, load-balancing capabilities across the nodes, queue functionality and share entitlement.

3 Getting Started

3.1 Logging into the remote machine

Interactive access to the remote machine is obtained through the **ssh** command

```
$>ssh user@aa.bb.cc.dd
```

3.2 Logging out of the remote machine

To logoff the remote machine, enter **exit** and a return

```
$>exit
```

3.3 Login profiles and shells

At login the user's default shell is **bash**. We have setup your PATH and a few important aliases in /etc/profile. Though if you want to customize something to your taste, please make corresponding entries in ~/bash_profile

4 Manipulating Files and Directories

4.1 Changing directory

cd without arguments puts the user in the users home directory. With a directory name as an argument, the command moves the user to that directory

```
$>cd directorypath
```

4.2 Copy files

cp makes copies of files in two ways.

```
$>cp file1 file2
```

makes a new copy of **file1** and names it **file2**.

```
$>cp [list of files] directory
```

puts copies of all the files listed into the directory named. Contrast this to the **mv** command which moves or renames a file.

4.3 Making a link

ln creates a link between files.
Example:
The following links the existing file example.c to ex.c.

```
$>ln example.c ex.c
```

The following creates symbolic links.

```
$>ln -s /usr/include incl
```

See the online **man** pages for many other ways to use **ln**.

4.4 Make a new directory

mkdir makes a new subdirectory in the current directory.

```
$>mkdir directoryname
```

makes a subdirectory called **directoryname**.

4.5 Move / rename files

mv moves or changes the name of a file.

$>mv *file1 file2*

changes the name of **file1** to **file2.** If the second argument is a directory, the file is moved to that directory. One can also specify that the file have a new name in the directory 'direc':

$>mv *file1 direc/file2*

would move **file1** to directory **direc** and give it the name **file2** in that directory.

4.6 Present working directory

pwd returns the name of the current working directory. It simply tells you the current directory.

4.7 Remove files

rm removes each file in a list from a directory. By default option **-i** to **rm** inquires whether each file should be removed or not. Option **-r** causes rm to delete a directory along with any files or directories in it.

$>rm *filename*

4.8 Remove directory

rmdir removes an empty directory from the current directory.

$>rmdir *directoryname*
removes the subdirectory named **directoryname** (if it is empty of files). To remove a directory and all files in that directory, either remove the files first and then remove the directory or use the **rm –r** option described above.

4.9 Listing files and directories

ls lists the files in the current directory or the directory named as an argument. There are many options:

ls -a [directory]
lists all files, including files whose names start with a period.

ls -c [directory]
lists files by date of creation.

ls -l [directory]
lists files in long form: links, owner, size, date and time of last change.

ls -p [directory]
subdirectories are indicated by /.

ls -r [directory]
reverses the listing order.

l**s -s [directory]**
gives the sizes of files in blocks.

ls -C [directory]
lists files in columns using full screen width.

ls -R [directory]
recursively lists files in the current directory and all subdirectories.

5 File Transfer

5.1 Establishing remote connection

To establish a connection to a remote system use the **sftp** command. After the connection is established provide the valid password.

```
$>sftp user@aa.bb.cc.dd
```

5.2 File uploading

Move a file from the local host to remote host

```
$>put filename
```

To put multiple files using wild cards

```
$>mput pattern*
```

5.3 File downloading

Move a file from remote host to local host

```
$>get filename
```

To get multiple files using wild cards

```
$>mget pattern*
```

5.4 Making a new directory on remote host

$>mkdir *directoryname*

5.5 Changing directory in local host

$>lcd *directorypath*

5.6 Changing directory in the remote host

$>cd *directorypath*

5.7 Closing the connection

$>bye

6 Compiling & Running Programs

The following compilers are available
C (cc)
C++ (CC)
FORTRAN 90 (f90)
FORTRAN 77 (f77)
JAVA

6.1 Compiling programs

$>cc *source.c*

compiles C program source.c and produces executable **a.out**

$>CC *source.cpp*

compiles C++ program source.cpp and produces executable **a.out**

$>f77 *source.f*

compiles FORTRAN program source.f and produces executable **a.out**

$>f90 *source.f*

compiles FORTRAN program source.f and produces executable **a.out**

$>javac *source.java*

compiles Java program source.java and produces byte code source.class

6.2 Running programs

After compiling a program the default executable is created as **a.out** file which can be run using the following syntax

$>./a.out

Java programs after compilation produces intermediate byte code (.class files) which need to be interpreted using the following command

$>java *source.class*

Note: You are discouraged from running a job in interactive mode for an extended period of time by using **./a.out OR java source.class** . It's OK to do so for testing purpose but not for more than 1-2 minutes. Such job slow down the entire system and are liable to be killed.

7 Submitting Jobs

7.1 Submitting serial *(single processor)* jobs

For firing a job through SGE, you need to use a submit script say submit.sh
Example of submit.sh

```
#!/bin/bash
#$ -S /bin/bash
#$ -cwd
source /etc/profile
source ~/.bash_profile

./a.out [arg1] [arg2] [arg3]
```

source ~/.**bash_profile ,** only if you have any local settings in .bash_profile

$>**qsub *submit.sh***

The **qsub** command is issued by the user to submit a file into the batch execution queue.

7.2 Submitting MPI Jobs

For firing a job through SGE, you need to use a submit script say submit.sh
Example of submit.sh

```
#!/bin/bash
#$ -S /bin/bash
#$ -cwd
#$ -pe cre np
source /etc/profile
source ~/.bash_profile

mprun mpi.exe [args1] [args2] [args3] ……
```

np to be replaced by number of processors. Recommended **4/8**
source ~/.**bash_profile ,** only if you have any local settings in .bash_profile

$>**qsub *submit.sh***

The **qsub** command is issued by the user to submit a file into the batch execution queue.

***For submitting a Gamess job on job.inp*: submit.sh**

```
#!/bin/bash
#$ -S /bin/bash
#$ -cwd
#$ -pe gamess np
source /etc/profile
source ~/.bash_profile

/opt/soft/gamess_parallel/gamess job 01 np >& submit.log
```

$>qsub -l g=1 submit.sh

7.3 Check the status of submitted jobs

$>qstat

7.4 Removing submitted jobs from queue

$>qdel job_id

job_id is obtained using **qstat**
output and errors (if any) can be checked in submit.sh.o<job_id> & submit.sh.e<job_id>

8 Optimization

Compile using **–fas**t option
Example

```
CC –fast test.cpp
mpCC –fast –lmpi test.cpp
```

9 Available Software

9.1 Amber 8

Installed Path: **`/opt/amber8/exe/`**
http://amber.scripps.edu/

9.2 Gamess

Installed Path: **`/opt/soft/gamess_parallel/gamess`**
http://www.msg.ameslab.gov/GAMESS/GAMESS.html

9.3 Gromacs

Installed Path :
`/opt/soft/gromacs/sparc-sun-solaris2.8/ultrasparc3/bin/`
`http://www.gromacs.org/`

10 Frequently Asked Questions (FAQs)

10.1 When I run a MPI job using "mprun –np x command", it gives the error "mprun: missing environment variable 'JOB_ID': Invalid request"

We have prohibited such direct running of processes as it by passes the queuing facility provided by SGE and overloads the machine. That effectively slows down every other job running on that node. Please follow the correct procedure given in the Submitting MPI Jobs.

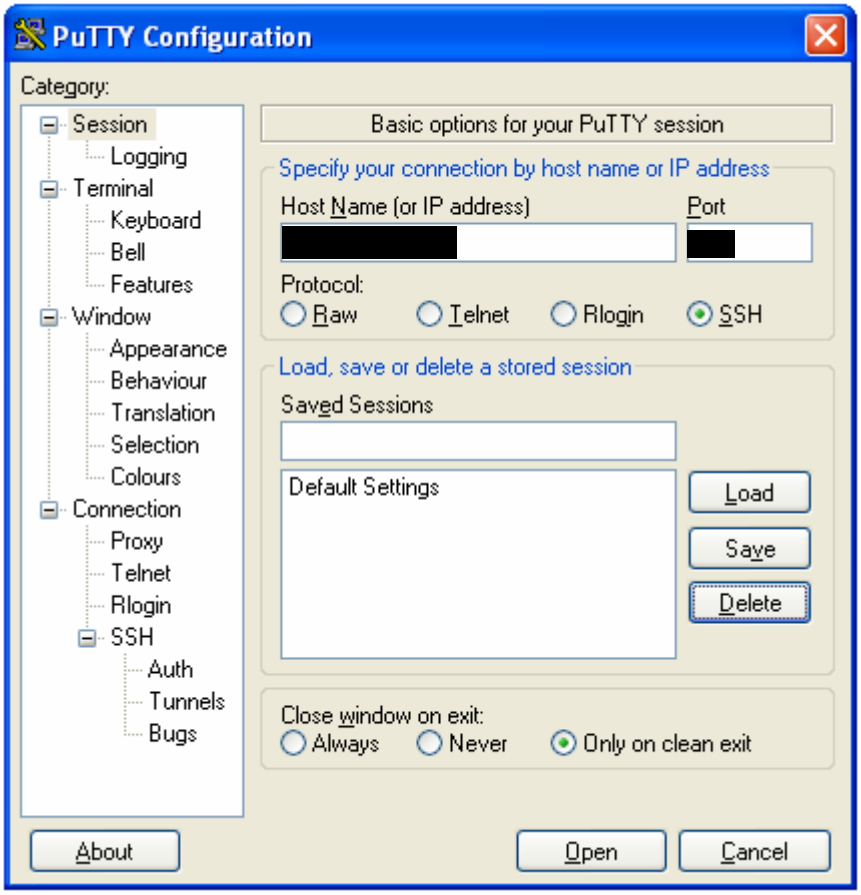
10.2 My job takes a long time to start or at some time it doesn't start altogether.

The waiting period of a job is directly dependent upon the number of jobs already waiting in the queue and the number of processors required by your job. A job is started only when the required number of processors is free. If your job requires large number of processors, it may very much happen that those number of processors never be free. With our testing & benchmarking we have found that the optimum number of processors to be 4/8. For 4, your queue waiting time should be smaller.

10.3 How do I access SCFBio from Window machines?

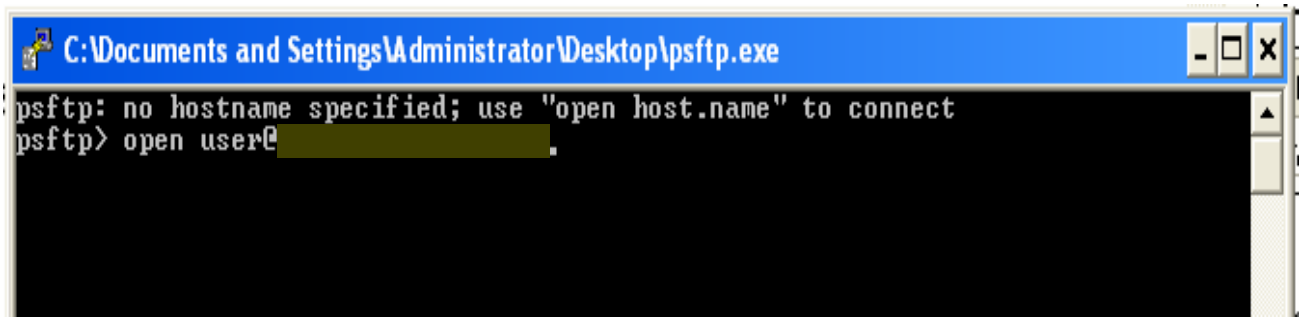
There are many free ssh or sftp applications available for window machines.
e.g. you can download **putty**(ssh) & psftp(sftp) from
http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html

10.3.1 putty



Please use the values from the figure (Host Name & Port.)

10.3.2 psftp



Please use the above command to work, replacing user by your provided userID.

10.4 In GROMACS some commands are suffixed "_mpi" and many others not.

Gromacs doesn't have all the programs implemented in parallel. There are only 4, which run parallel namely disco_mpi, ffscan_mpi, g_disre_mpi & mdrun_mpi . To run any of these 4

programs, please use Submitting MPI Jobs, while for others
Submitting serial (single processor) jobs

10.5 How do I check the status of a running parallel job.

```
$> mpps –Apf
```

If you find any process in the status of EXIT/CORING/SIGNALLING, kindly check your submit script. If any be in the status of SPAWNING for a prolonged period of time, please mail us detailing time & date.

10.6 I get the error message timeout while launching a MPI job and it asks me to use '-t' flag with mprun.

Kindly report this error to us with time & date and any other output messages you may have.

10.7 I am a beginner and I want to learn more about UNIX in general. Can you please give some references?

There are many good tutorials available on internet for beginners.
e.g. http://www.ee.surrey.ac.uk/Teaching/Unix/
You can also refer to books such as "UNIX Concepts & Applications by Sumitabha Das"

10.8 Whenever I delete multiple files using wild card say by rm *, I need to answer 'y' for each file.

Please use '\rm *' for the purpose. Be careful as it would delete all the files in your current working directory.

10.9 I don't like the default executable name i.e. a.out. How can I change it?

Please use –o flag to compiler with the executable name.
e.g. CC my_program.cpp –o my_program.exe
This will compile the program and executable will be formed with the name my_program.exe

10.10 Is there any available debugger?

```
$>dbx
```

10.11 I am unable to telnet/ftp/ssh/sftp to any other machine from the Supercomputer.

We have denied the permission to make a connection from the supercomputer to any other machine. You are only allowed to make connection to supercomputer from any other machines but not the other way.

10.12 When I open a file using vi editor, I find ^M at the end of every line.

Whenever a file is opened in Windows, it appends ^M at the end of every line. It is just a matter of format difference between windows & UNIX. To remove ^M, please use

```
$> dos2unix originalfile convertedfile
```

10.13 How do I check my quota status?

```
$> quota –v
```

10.14 I am unable to connect to the provided IP address and the IP address isn't responding to ping request.

For the security reasons, the IP address doesn't reply to ping request. If you are unable to connect, the primary reason would be that the network is down on your end. To check please ping www.mail.yahoo.com . If yahoo.com responds and you are unable to connect even after waiting for more than 1-2 minutes then please drop us a mail detailing time & date.

10.15 How can I learn more about a command?

```
$>man command_name
```