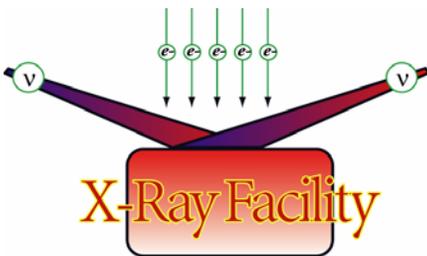


Soma's Computer Notes

X-Ray data transfer using *rsync*



*Procedure for transferring data from an
old to a new computer using rsync*

Table of contents

Introduction 3
UNIX command: *rsync* 3
Conclusion..... 6

© 2000-2006 Thayumanasamy Somasundaram
414, Kasha Laboratory
Institute of Molecular Biophysics, Florida State University,
Tallahassee, FL 32306-4380
E-mail: soma@sb.fsu.edu • URL: <http://www.sb.fsu.edu/~soma>
Phone 850.644.6448 • Fax 850.644.7244
March 8, 2006
Logos, Figures & Photos © of the respective Instrument Manufacturers

X-Ray Data Transfer Using *rsync*

Procedure for transferring data from an old to a new computer using rsync

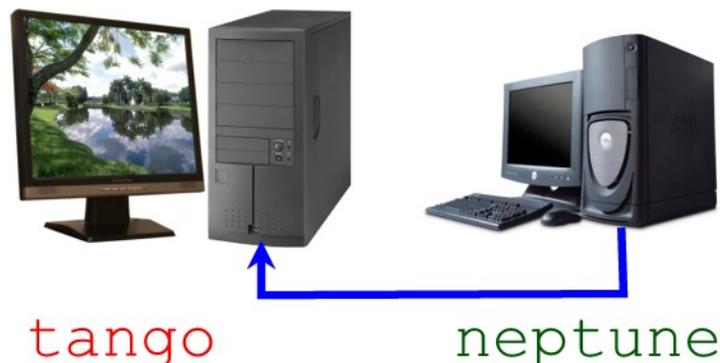
Version: March 08, 2006

Introduction

This note is intended to help the [X-Ray Facility](#) (XRF) users transfer their data from an old computer to a new computer using a Linux/UNIX utility called `rsync`. A copy of this Note will be posted in [XRF Resources page](#) shortly after receiving suggestions and corrections from the users. This note was first written in March 7, 2006 and updated in March 8, 2006. The update includes several improvements over the original version (dropping `ssh` and compression options while using `rsync` since the both computers mount each other using Network File System, *aka* NFS).

UNIX command: *rsync*

`rsync` is an archiving UNIX/Linux utility that synchronizes directories and files in location with another location (both local and remote). It has several options (secure copy, file compression, directory tree, and file comparison) allowing a user to copy files from a source to a destination either one-time transfer or incrementally over the network.



X-Ray Facility (XRF) will soon be retiring the single processor RedHat Linux based machines `neptune.sb.fsu.edu` and `raccoon.sb.fsu.edu` replacing them with Ubuntu Linux based dual AMD 2.0 GHz Athalon processor machines `tango.sb.fsu.edu` and `gauss.sb.fsu.edu`. Users are urged to archive their data in the older machines using tape and/or DVD, or transfer their data (Note: data here refers to the processed data only and NOT the original diffraction images and the latter should NOT be transferred due to size limitations) to the new machines.

Currently, `tango.sb.fsu.edu` is on-line and users should first establish that they can log-in to this machine using their `flame.sb.fsu.edu` username and password (or `mailer.sb.fsu.edu` username and password). Once logged-in the users should create a new directory under `/data/users/` called `username` (i.e., your username) and a sub-directory (say, `My-Data`) and other sub-sub-directories depending upon your project. Next step will be to transfer the “required” data from the older machines to the new machine.

The following procedure is one of the ways of accomplishing the transfer using `rsync` protocol from the old machine to new machine while preserving the ownership, symlinks, date of creation, and other characteristics of your old files. In this procedure, we will use `rsync` with `exclude-from` option to selectively transfer the “required” directories or files. For more details of how `rsync` works and other options the user is referred to [rsync web page](#) or manpage. For this example we are assuming a typical scenario where the user has several directories in the old machine (referred to as **Source**) but wants to transfer only some of them to the new machine (referred to as **Destination**).

Source:

This is the old machine (`neptune.sb.fsu.edu`). I log-in and move to the location where my processed files are stored. There are four sub-directories under `Old-Data` but only contents of some directories (say `PDB`) need to be transferred but not others (say, `Data`, `Denzo` and `Phaser`) from this old computer to the new computer called `tango.sb.fsu.edu`.

Here I am listing the directories in the old machine to find out what needs to be transferred and what can be excluded. As I mentioned above, we decide to transfer all the contents of `PDB` (i.e. `PDB` and all its sub-directories) but none of the other.

```
soma@neptune[3:02pm]/d6/Old-Data>/bin/ls -lt
total 20
drwxr-xr-x    2 soma    users      4096 May 12  2005 PDB
drwxr-xr-x    2 soma    users      4096 May  5  2005 Phaser
drwxr-xr-x    3 soma    users      8192 May  5  2005 Denzo
drwxr-xr-x    4 soma    users      9096 May  3  2005 Data
```

Destination:

This is the new machine (`tango.sb.fsu.edu`). I log-in first and create a directory under `/data/users` called `soma` (this is my username). Then I move into that directory. Then I create a file called `ex.txt` in which I type names of all the directories of old computer (one directory per line) that I do NOT want to be transferred to the new machine and save it. Then I create a new directory called `My-Data`. Next I issue the `rsync` command with `exclude-from=` and other options. The transfer should begin, proceed and complete. Since there are several ways to transfer the data between two computers in a local area network, as opposed to wide area network (NFS mounted or not, forgo or keep compression and secure or regular transfer), I tried out all the combinations and determined the time taken to transfer the same amount of data and the results are given in the table below:

Compression	SSH	NFS	Time (bytes/sec)
Yes	No	No	14720.13
Yes	Yes	No	20072.91
No	No	No	79406.36
No	Yes	No	97052.22
Yes	No	Yes	220972.00
No	No	Yes	873640.00

Since we are transferring data between the XRF computers we will choose the fastest way. This means that there is no need for secure shell (since the directories are NFS mounted), no compression needed since we are not concerned about band-width (internal transfer) but want to reduce the cpu load.

A Typical Session (transcript from the new computer): Using `exclude-from` option

What a computer prints-out is shown as **courier-bold** and what the user types-in is shown in **courier-regular**.

```
soma@tango:/data/users/soma$ pwd
/data/users/soma
soma@tango:/data/users/soma$ mkdir My-Data
soma@tango:/data/users/soma$ ls -lt
drwxr-xr-x  2 soma Domain Users  39 2006-03-06 11:55 My-Data
soma@tango:/data/users/soma$ emacs ex.txt
Contents of ex.txt
  Data
  Denzo
  Phaser
soma@tango:/data/users/soma$ rsync -av --exclude-from=ex.txt
/neptune/d6/Old-Data/ ./My-Data/
building file list ... done
./
PDB/
PDB/1JQZ.pdb
PDB/1jqz-a.pdb
PDB/1jqz-b.pdb
PDB/1jqz-h20.pdb
PDB/Readme.txt
PDB/moleman.log
PDB/rasmol-distance-query.txt

sent 436634 bytes  received 186 bytes  873640.00 bytes/sec
total size is 436014  speedup is 1.00

soma@tango:/data/users/soma$ cd My-Data
soma@tango:/data/users/soma/My-Data$ ls -lt

drwxr-xr-x  2 soma Domain Users  141 2005-05-12 10:17 PDB
```

The command and the explanation:

```
rsync -av --exclude-from=ex.txt /neptune/d6/Old-Data/ ./My-Data/
```

and the explanation for the command

`rsync`: the command itself
`-a`: archive mode (back-up)
`-v`: verbose mode
`--exclude-from=`: directories and files that need to be excluded; wild card permitted
`ex.txt`: the list of directories that are to be excluded, one line per directory
`/neptune/d6/Old-Data`: source directory for copying (NFS mounted, old machine)
`./My-Data/`: destination directory for storing (local computer, current directory, new machine)

Conclusion

The users will be given one-month to transfer their data from neptune to tango. Please send your suggestions and comments to [Soma](#).