NTNU – Trondheim
Norwegian University of
Science and Technology

AGENDA

Norstore

- **-**What is Notur and Norstore
- -Interview
- -File transfer applications
- -File transfer time
- -Best pratices

Demonstration of iRODS

Demonstration of R

What is Notur and Norstore

Notur:

-The Notur project provides the national infrastructure for computational science in Norway.

Norstore:

-Norstore is to establish and maintain a national infrastructure for the curation of digital scientific data.

See www.notur.no and http://docs.notur.no/

There are 2 storage sites in Norway to day:

-Norstore-trd (NTNU)
(http://docs.notur.no/ntnu/norstore-ntnu/norstore-trd)

-Norstore-osl (UiO)
(https://wiki.uio.no/usit/suf/vd/hpc/index.php/NorStore)

It is 600TB both in Trondheim and Oslo.
Each site use 300TB for replicate the other site.

-UNINETT Sigma is the coordinator and project responsible of the Notur/Norstore project and entered into agreements with partners

Apply for a project on Norstore.

Send a application to Notur Resource Allocation Committee (RFK)

https://www.notur.no/application/login.php

INTERVIEW

The information that will be gathered during the interview includes (but is not limited to) the following

A. Data characterization:

1.Describe the characteristics of the information, in particular:

2.If available, specify the location of documentation that describes the characteristics of the information.

3.If the information will be modified while it resides on the NorStore resources, describe

B.Services, interfaces and performances:

1.Which applications are used to transfer and manipulate data?

2.Specify the required interfaces, protocols and tools (e.g., gridftp, sftp, scp, http(s), ldap, ...).

3.Specify the need for other Linux, Windows, Web and/or other interfaces.

4.Specify if the data *must* be 'mounted' on a remote system, such that it is readily available for use on that remote system and appears as if it is locally available (without using explicit transfer methods to access the data).

5.Indicate the amount of information (in TeraBytes) that is operated on every day on average.

6.If you consider your data to be I/O intensive, please provide details, e.g., in terms of required performances. If possible, specify concrete performance requirements, e.g.,

-response times, availability

-bandwidths and latencies

7.Specify any non-default requirements for availability of data and uptime requirements for services.

-By default, all data and services are available 24/7.

However, errors that occur may only be repaired during regular working hours (roughly 8am – 4pm).

8.Specify the preferred mechanisms to provide/restrict access to data (e.g., password or certificates)

9.Specify the services that are necessary and sufficient to avoid loss of data. This includes mirrors and back-up policies (e.g., number of copies, frequency of back-up). The following defaults apply:

-For static (non-changing) data sets, one copy is made.

-For data that itself is already a copy, no copies / back-ups are made.

Storing structures

To day the norstore is organized in projects as_

Norstore->projects->project1

->project2 etc

It is up to user to organize and maintains its project space.

Project data should be organized and structured by content

(This system will be changed!)

Norstore trd replicate all its data to Norstore_osl.

Therefor, please use the "_noreplica" options on files and folders to EXCLUDE any unnecessary files or folder from being replicated to norstore-osl

Norstore->projects->project1->_nobackup_tmp_data

->images

->analyzed_data

File Transfer application

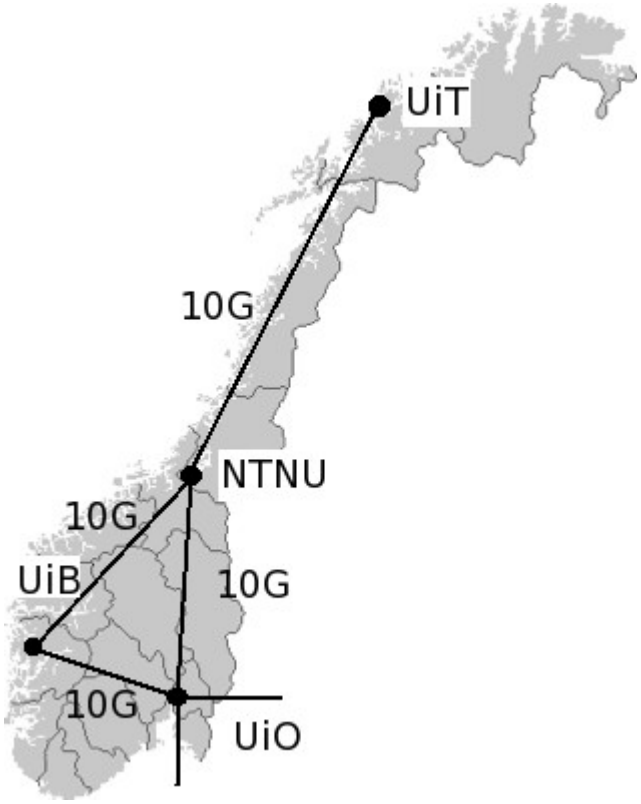
Norstore trd support several file transfer application as

-HPN-scp (Secure copy.)

-iRODS

-GridFTP

Transmission speed (10Gb/s)



File transfer time.

Table 1. Transfer time between HPC centers and Norstore-trd for 1TB data.

Data collection	Hexagon scp	Hexagon iRODS	Stallo scp	Stallo iRODS	Njord scp	Njord iRODS	Osl scp	Osl iRODS	Osl gridftp
Small (1TB)	113hours (4.7days)	151hours (13days)	170h 7.1d	355h 14.8d	64h 2.7d	313h 13d	198h 8d	315h 13d	244h 10d
Big (1TB)	4h	1,2h	3.5h	3h	3.7h	48 min	5.3h	29 min (4.6Gb/s)	54min

Small data files: Large amount of small files of kBytes.
Big data files: Few files of GBytes.

Best practices

The NorStore best practices (general rules of thumb) are given below:

NorStore distinguishes between project, user and scratch spaces.

1.All new users associated with a project must register (NorStore access and “Application for user account”) before access can be established. Users who are already associated with a Notur or NorStore project do not normally need to fill in the registration form.

2.Project space should not include personal files (mails, photos, documents etc). This should be placed in the user spaces

3.Project data should as a rule be compressed and/or stored in widely accepted or standardized formats

see wiki for a list of suggested standard formats

http://en.wikipedia.org/wiki/Scientific_data_format#Scientific_data_formats_.28data_exchange.29

4.Project data should be organized and structured by content (rather than by user)

5.Temporary project data, resulting from analysis, computations or similar should be reduced to a minimum and not left on disk unattended for longer periods

6.Project data which has long-term value or is considered critical (eg. raw observational data) should be collected in a “data package” (eg. tar-ball) and accompanied by documentation of the data

7.Data documentation should include, as a minimum, a description of the data; such as origin, date of creation, type of data (observational, simulation), data format, owner or PI and a persistent reference to the science project (or a summary of the science project, goals etc). In order to secure the data is meaningful for the future a package containing the software that is used to process and/or analyze it, should accompany the data package.