NORTHWESTERN
UNIVERSITY

Notes on Transferring 100 TB of Data Using Globus

William E. Mihalo; Anton Verlygo; Ryan K. Sisk
Northwestern University

NUIT
Northwestern University
Information Technology

NORTHWESTERN
UNIVERSITY

Transferring 100 TB of data

1. Background and description of the data
2. Procedure for transferring the data
3. Problems encountered
4. Recommendations

NUIT
Northwestern University
Information Technology

NORTHWESTERN UNIVERSITY

Transferring 100 TB of data Workstation and disk drives

1.2011 iMac (USB 2.0 connectors) with 8 GB of RAM.
2.3 TB Fantom or Lacie external drives attached via USB
3.Network connection was via 1 GigE ethernet
4.48 external disk drives were used for the transfer.

NUIT
Northwestern University Information Technology

NORTHWESTERN UNIVERSITY

Transferring 100 TB of data End point configuration

1. A total of 2 servers provide endpoint support.
2. One machine runs globus connect (server).
3. A second server provides login authentication into the quest cluster and also runs globus connect.
4. Each globus server has two Xeon X5650 CPUs with 24 GB of RAM.
5. Our globus servers are located at the Evanston Data Center.
6. Access to our storage volumes from the globus servers is via Infiniband/GPFS for our primary storage and a 10 Gb/s NFS connection for our supplemental storage.

NUIT
Northwestern University Information Technology

NORTHWESTERN UNIVERSITY

Transferring 100 TB of data Procedure

1. Transfers started in late October and finished in late February.
2. Disk drives were shipped to Northwestern's Chicago Campus.
3. Transfers were done on a lab computer.

Transfers started in the late afternoon and ran over night or over the course of a weekend.

NUIT
Northwestern University
Information Technology

NORTHWESTERN UNIVERSITY

Transferring 100 TB of data Procedure (cont'd)

1. Disk drives were attached to an iMac computer either singly or in pairs.
2. Typical transfer speed with a single disk drive attached was about 219 megabits/second
3. Transfer speed with two drives attached throttled down to 194 megabits/second

NUIT
Northwestern University Information Technology

NORTHWESTERN UNIVERSITY

Transferring 100 TB of data Description of Data

1. Average file size was about 747 MB.
2. 147,311 files were transferred
3. Transfers took 81,597 minutes or about 1,360 hours.

NUIT
Northwestern University Information Technology

NORTHWESTERN UNIVERSITY

Transferring 100 TB of data Description of the transfer

1. Data were transferred from Northwestern's Chicago Campus to our Evanston Campus.
2. Globus connect server nodes are located at the Evanston campus.
3. Between Chicago and Evanston campuses, we have dark fiber for the network communication

NUIT
Northwestern University Information Technology

Transferring 100 TB of data Problems

1.Transfers were done when the workstation was not busy. This meant transfers took place in the evening and over the weekend.
2.Transfers conflicted with backups that take place between Chicago and Evanston campuses and this had a substantial impact on our transfer speed. We had a high network load when we started the transfers.
3.USB 2 was a bottleneck.

NUIT
Northwestern University
Information Technology

NORTHWESTERN
UNIVERSITY

Transferring 100 TB of data Problems (cont'd)

1. Condition of the disk drive will affect the speed of the transfers.
2. Drives that were damaged during shipping had their transfer speeds cut in half.

NUIT
Northwestern University
Information Technology

Transferring 100 TB of data Recommendations

1.Do some test transfers before embarking on a massive set of transfers.
2.Run your test transfers at different times of the day.
3.Use smartmon tools to check the condition of a disk drive before starting a transfer. If you are doing a lot of transfers, you might want to defer transferring data from a failing drive until the end. If it is possible, you can also see about having the data re-shipped on a different disk drive (if it is still available).

NUIT
Northwestern University
Information Technology

NORTHWESTERN UNIVERSITY

Transferring 100 TB of data Recommendations (cont'd)

1.In our case, data stayed on the storage system of the sending institution for 30 days and then were automatically deleted. With serveral dozen disk drives arriving, it was difficult to check the condition of each drive until just before transferring data.

NUIT
Northwestern University
Information Technology

NORTHWESTERN UNIVERSITY

Transferring 100 TB of data Recommendations (cont'd)

1.Use smartmon tools to check the condition of a disk drive before starting a transfer. If you are doing a lot of transfers, you might want to defer transferring data from a failing drive until the end. If it is possible, you can also see about having the data re-shipped on a different disk drive (if it is still available).

NUIT
Northwestern University
Information Technology

NORTHWESTERN UNIVERSITY

Transferring 100 TB of data
Why did we use a modified sneaker net?

1. Many institutions use default port speeds of 100 MB/s.
2. We have run non-globus transfers from other sites where they could not ship a disk drive to us. We found these transfers took an extremely long time and could easily fail, or have corrupted data.
3. We have proselytized the use of globus technology for transfer purposes. Globus needs to make this technology more visible to potential end-users and institutions.

NUIT
Northwestern University
Information Technology

NORTHWESTERN UNIVERSITY

Transferring 100 TB of data Recommendation for globus

An online tool within globus connect that showed network statistics and the current transfer rate
would be extremely helpful.

NUIT
Northwestern University
Information Technology