Tools for Data Transfer

Dr. David Swanson
Director, Holland Computing Center
University of Nebraska
Outline

• Data Transfer Context
• Data Transfer Tools
• Example Cases
• New Research
Data Transfer Tools (Swanson)
Broadening the Reach Workshop, Kansas City, MO
02/17/14 – 02/19/14
All running jobs
1 Weeks from 2014-02-16 06:52 to 2014-02-18 06:52

Maximum: 6,270, Minimum: 0.00, Average: 5,064, Current: 5,900
Data Transfer Tools (Swanson)
Broadening the Reach Workshop, Kansas City, MO
02/17/14 – 02/19/14
Storage Options

- /home ZFS, quotas, backed up
- /work Lustre, no quota, purged
- /shared ZFS, not backed up
- HDFS (CMS only)
- Library custodial storage
- Offsite Tape increased security, per year
Data Transfer Tools

• GridFTP ... FTW!
  – requires consideration admin intervention
• Also use: http, condor file transfer, SCP, winSCP
• Aspera (not using due to nature of network)
• Have ordered transfer boxes for Globus Online endpoints – much more accessible for users
Data Transfer Options

• **Batch Transfers**
  - < 100 MB: Condor File Transfer [UWisc]
  - < 500 MB: http caching (squid)
  - < 2 GB: GridFTP [Globus]

• **Streaming Transfers**
  - “small”, generic: Parrot/Chirp [CCTools, ND]
  - “big”, HEP: xrootd [SLAC]
Data Transfer Tools (Swanson)
Broadening the Reach Workshop, Kansas City, MO
02/17/14 – 02/19/14
Any Data, Any Time, Anywhere

• NSF-sponsored project aiming to increase data access for HEP

• While HEP has a long history in bulk data transfers between sites, want to enable direct remote IO to users’ applications

• With remote IO, users can access the experiments’ files in seconds, not hours
AAA technology - Xrootd

• Similar in spirit to a content delivery network, although unique in that the access patterns are inherently cache-unfriendly and the volume of the data (~30PB)
• allows CMS to run jobs at one site while streaming data from a second site
• currently plays an integral role in the global CMS computing evolution plans
AAA technology - Xrootd
LARK

- NSF grant in the CC-NIE program, Lark aims to improve network usage for HTCondor distributed high-throughput computing
- integrates and tests advanced network technologies (IPv6, perfSONAR, network management via SDN), with the goal of bringing them into production at UNL and University of Wisconsin Madison
LARK

• For example, provide each HTCondor batch job with an IP address so the network layer can manage at a per-job, not per-host, level.
Credit Due

• Not possible without help/support from I2, GPN, ITS (Mark Askren, CEO) networking (Kent Christensen)

• OSG, UW-Madison, ND, Globus

• NSF

• Dr. Brian Bockelman, Dr. Adam Caprez, Dr-2-b Derek Weitzel, Garhan Attebury

• NU Administration, NRI, Holland Foundation