From: Maarten Litmaath
Sent: Monday, May 08, 2006 00:15
To: Jamie Shiers
Subject: brief SC4 summary per site

ASGC

Still running CASTOR-1. Experimented with various upgrades to get improved performance per node: kernel from 2.4 to 2.6 for TCP,
ext3 to XFS file system, newer drivers for RAID controllers, TCP buffer and window sizes, latest CASTOR gridftpd, etc.
This did not converge in time, and in the end various components had to be downgraded again, after which stable running at 120 MB/s to disk was achieved for a few days, but followed by instabilities that are not all understood yet.
The tests and investigations are useful preparations for the CASTOR-2 upgrade foreseen to take place in a few months.

BNL

Changed their dCache setup to allow for stable running at nominal rates using 3rd party transfers (i.e. deploying a sufficient number of powerful gridftp door nodes), not relying on srmCopy. SC4 exercise Exposed a few dCache issues/bugs, e.g. a single stuck transfer can block all others, and recursive PNFS listings can cause a DoS.
In the end, with occasional admin interventions, met both disk-disk and disk-tape nominal rates for many days.

CNAF

Had a mixed CASTOR-1/-2 setup during April, suffering a lot from problems fixed in later CASTOR-2 versions. Upgraded to the latest release at the end of April and were finally able to run at 200 MB/s disk-disk for many hours in a row, but followed by new instabilities.

DESY

An excellent reference site. During April their rate was limited to a constant 70 MB/s disk-disk by their network. Then they upgraded their uplink from 1 to 10 Gbps, which boosted their rate to a constant 170 MB/s.

FNAL

At this time the only site that ran with srmCopy, exposing a few bugs in the FTS that subsequently got fixed. Reached the highest rates of all sites, peaking at 450 MB/s, but needing a large number (70-160) of concurrent transfers and a large number (20) of streams per file.
They are using some thirty-odd ordinary machines to receive the data.
Easily reached stable nominal rates for disk-disk and disk-tape, but also suffered from a recursive PNFS listing.

FZK/GridKa

For all of April were limited to about 150 MB/s disk-disk for reasons that are not completely understood, though a significant problem in the disk striping configuration was only fixed early May, along with one or two other changes.
The rates then finally managed to exceed the nominal 200 MB/s disk-disk, but there were new instabilities, e.g. a dip every 6 hours for reasons not

understood. Switched to tape last week, using a single drive for now, not quite stable yet.

IN2P3

The very first site that was set up and ready even before SC4 started.
They came very close to their nominal 200 MB/s for many days in a row, but only came to steadily exceed that rate during the last two weeks, after having switched back from tape to disk, with some configuration changes. They now have met both disk-disk and disk-tape target rates.

NDGF

Started halfway through SC4, but immediately wrote all data to tape, even during the disk-disk phase, and very quickly reached their target rate of 50 MB/s, usually doing 60.

PIC

Still running CASTOR-1. Limited by their 1-Gbps shared uplink to some 70 MB/s disk-disk, in the end also reached during the disk-tape phase.
Had quite a few instabilities, but demonstrated about the maximum that could be achieved with their current setup. Now focusing on the upgrade to CASTOR-2.
Network uplink to be upgraded in September.

RAL

Still running dCache instead of CASTOR-2. Using their production setup, shared with other users. Network situation was not completely clear, lightpath capacity got doubled in the midst of SC4, but did not have a very noticeable effect. Their disk-disk target rate was 150 MB/s, which was just met most of the time, but never exceeded. Tape target rate was 50 MB/s, which is met most of the time, but with high error rates.

SARA

Suffered some problems with their dedicated 10-Gbps link to CERN.
Exceeded their nominal 200 MB/s disk-disk only for 2 days in a row.
Discovered some bottleneck in their SAN configuration during the tape phase, limiting them to not much more than 30 MB/s. Switched their channel off to investigate short- and long-term solutions.

TRIUMF

Very stable reference site. Easily reached their 50 MB/s target rate both disk-disk and disk-tape.