Service Challenge Meeting

"Review of Service Challenge 1"

James Casey, IT-GD, CERN

RAL, 26 January 2005

Overview

- Reminder of targets for the Service Challenge

- What we did

- What can we learn for SC2?

Milestone I & II Proposal

From NIKHEF/SARA Service Challenge Meeting

Service Challenge Schedule

From FZK Dec Service Challenge Meeting:

SARA – Dec 04

- Used a SARA specific solution
  - Gridftp running on 32 nodes of SGI supercomputer (teras)
  - 3 x 1Gb network links direct to teras.sara.nl
  - 3 gridftp servers, one for each link
  - Did load balancing from CERN side
    - 3 oplapro machines transmitted down each 1Gb link
  - Used radiant-load-generator script to generate data transfers
- Much efforts was put in from SARA personnel (~1-2 FTEs) before and during the challenge period
- Tests ran from 6-20th December
  - Much time spent debugging components

Problems seen during SC1

- Network Instability
    - Router electrical problem at CERN
    - Interruptions due to network upgrades on CERN test LAN
- Hardware Instability
    - Crashes seen on teras 32-node partition used for challenges
    - Disk failure on CERN transfer node
- Software Instability
    - Failed transfers from gridftp. Long timeouts resulted in significant reduction in throughput
    - Problems in gridftp with corrupted files
- Often hard to isolate a problem to the right subsystem

SARA SC1 Summary

- Sustained run of 3 days at end
  - 6 hosts at CERN side. single stream transfers, 12 files at a time
  - Average throughput was 54MB/s
  - Error rate on transfers was 2.7%
- Could transfer down each individual network links at ~40MB/s
  - This did not translate into the expected 120MB/s speed
  - Load on teras and oplapro machines was never high (~6-7 for a 32 node teras, < 2 for 2-node oplapro) Load on oplapro machines
- See Service Challenge wiki for logbook kept during Challenge

Gridftp problems

- 64 bit compatibility problems
  - logs negative numbers for file size > 2 GB
  - logs erroneous buffer sizes to the logfile if the server is 64-bits

- No checking of file length on transfer
  - No error message doing a third party transfer with corrupted files

- Issues followed up with globus gridftp team
  - First two will be fixed in next version.
  - The issue of how to signal problems during transfers is logged as an enhancement request
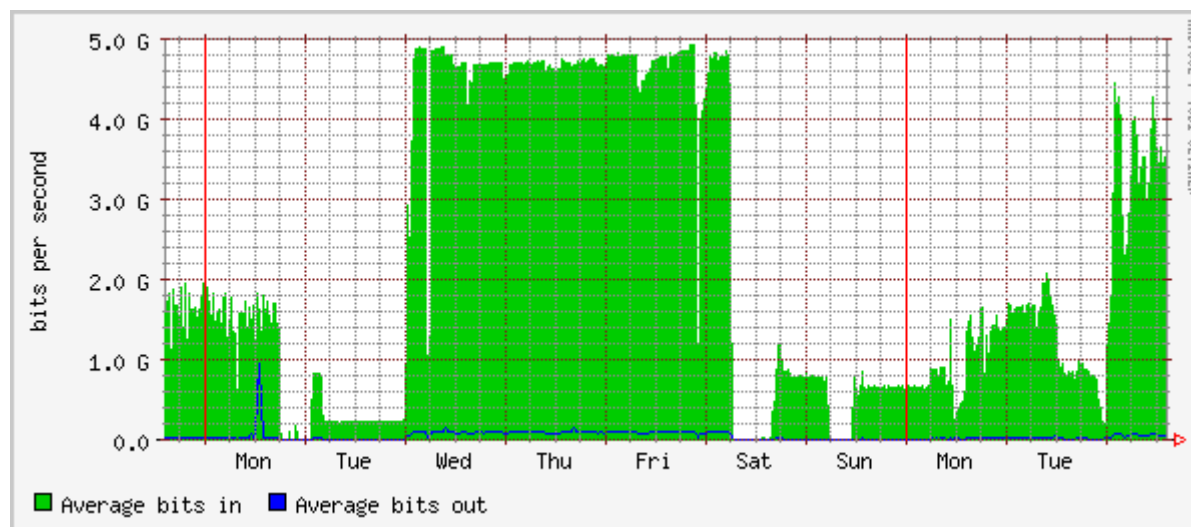
FermiLab & FZK – Dec 04/Jan 05

- FermiLab declined to take part in Dec 04 sustained challenge
  - They had already demonstrated 500MB/s for 3 days in November

- FZK started this week
  - Bruno Hoeft will give more details in his site report

What can we learn ?

- SC1 did not succeed
  - We did not meet the milestone of 500MB/s for 2 weeks
- We need to do these challenges to see what actually goes wrong
  - A lot of things do, and did, go wrong
- We need better test plans for validating the infrastrcture before the challenges (network throughtput, disk speeds, etc…)
  - Ron Trompert (SARA) has made a first version of this
- We need to proactively fix low-level components
  - Gridftp, etc…
- SC2 and SC3 will be a lot of work !