Service Challenge Meeting

Summary of Service Challenges

Based on slides and discussions at March 15 SC Meeting in Lyon

Agenda

- Status of Service Challenge 2
- Preparations for Service Challenge 3
- Involvement of Experiments in SC3
- Milestones and Workplan

Service Challenge Summary

- "Service Challenge 2" started 14th March
- Set up Infrastructure to **7** Sites
  - **BNL**, CNAF, FNAL, FZK, IN2P3, NIKHEF, RAL
- 100MB/s to each site
  - 500MB/s combined to all sites at same time
  - 500MB/s to a few sites individually
- Use RADIANT transfer software
  - Still with dummy data
  - 1GB file size
- Goal : by end March, sustained 500 MB/s at CERN

➢ **BNL status and plans in next slides**

BNL Plans

Week of 3/14: "Functionality" – Demonstrate the basic features:

- Verify infrastructure
- Check SRM interoperability
- Check load handling (effective distribution of requests over multiple servers)
- Verify controls (tools that manage the transfers and the cleanup)
- ???

Week of 3/21: "Performance" :

- Run SRM/SRM transfers for 12-24h at 80-100MB/sec

Week of 3/28: "Robustness":

- Run SRM/SRM transfers for the entire week at ~60% of available capacity (600Mbit/sec?)

BNL Status

Deployed dCache testbed for SC; the system is equipped with:

- One core node and four pool servers; SRM and gridftp doors.
- 4 x 300GB = 1.2TB space.
- Disk only as backend
  - the production dCache system is interfaced with HPSS, however during this SC phase we'd like to not involve tape access due to reduced total capacity and critical production schedule conflicts.

Tested successfully the following configurations:

- CERN SRM-CASTOR ---> BNL SRM-dCache
- CERN SRM-CASTOR ---> BNL GRIDFTP-dCache
- BNL SRM-dCache ---> CERN SRM-CASTOR
- BNL GRIDFTP-dCache ---> CERN SRM-CASTOR

Completed successfully a test using list-driven SRM 3rd party transfers between CERN SRM-CASTOR and BNL SRM-dCache. The duration of the test was two hours.

Plans and timescale - Proposals

- Monday 14th- Friday 18th
  - Tests site individually. 6 sites, so started early with NL !
    - Monday IN2P3
    - Tuesday FNAL
    - Wednesday FZK
    - Thursday INFN
    - Friday RAL
- Monday 21st – Friday 1st April
  - Test all sites together
- Monday 4th – Friday 8th April
  - Up to 500MB/s single site tests
  - Fermi, FZK, SARA
- Monday 11th - …
  - Tape tests ?
  - Fermi, INFN, RAL
- Overlaps with Easter and Storage Management w/s at CERN

This will be updated with BNL asap

Meetings

- SC2 Meetings
  - Phone Conference Monday Afternoon
  - Daily call with different sites
- Post SC2 Meetings
  - Tuesday 4pm (starting April)
- Storage Workshop
  - 5-7 April @ CERN
- Pre-SC3 Meeting
  - June 13th – 15th at CERN
- Taipei SC Meeting
  - 26th April
- Hepix @FZK
  - 9-13 May

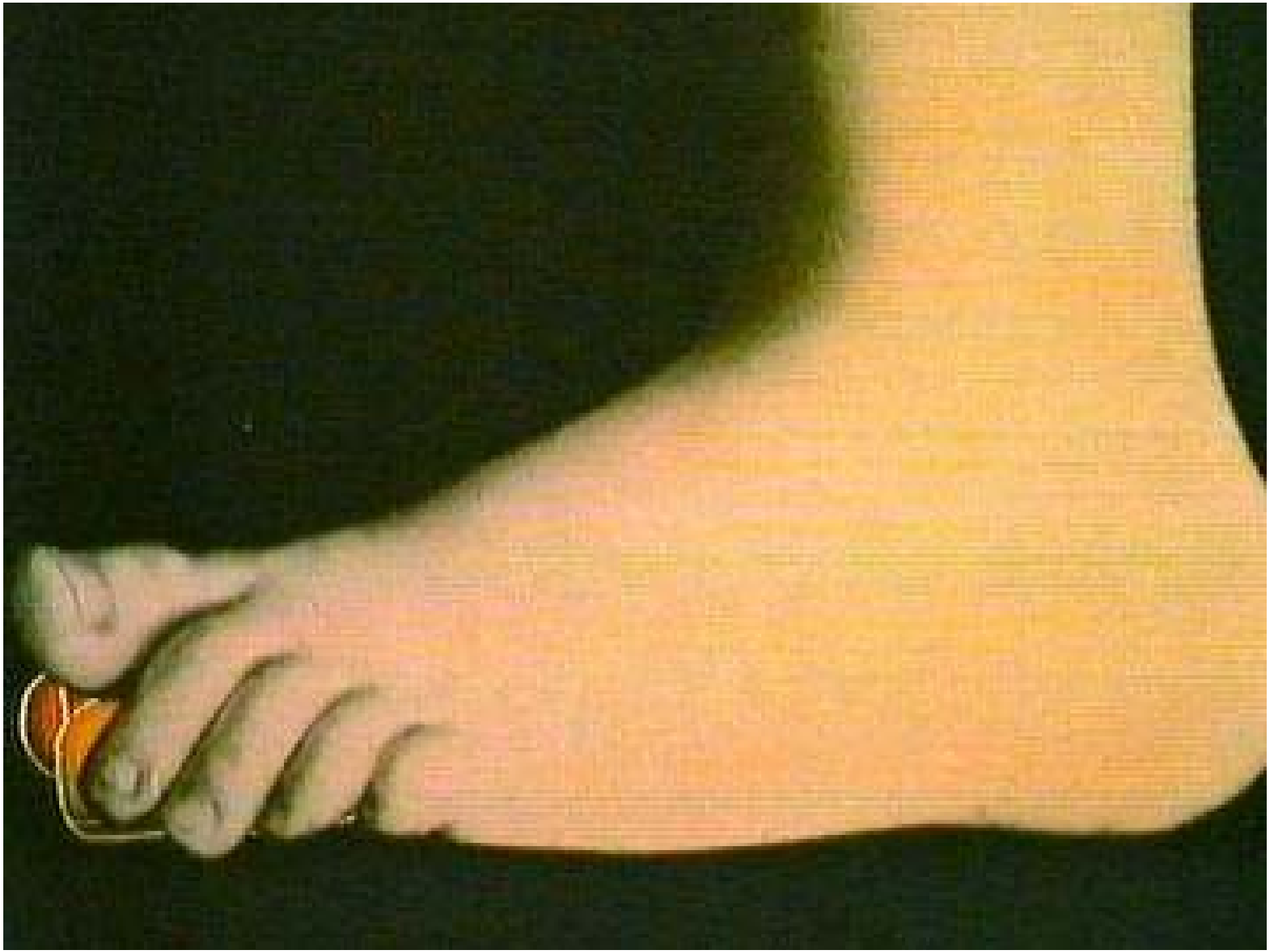
There are also others:

- **Weekly SC meetings at CERN;**
- **T2 discussions with UK**
  - **next June in Glasgow?**
- **T2 INFN workshop in Bari**
  - **26-27 May**
- **etc. etc.**

Longer Term Planning

- Mid April – Mid June – SC2
  - Taipei
  - **BNL – now!**
  - PIC
  - TRIUMF
- SC3 deployment

- SRM Tests
  - FZK, SARA, IN2P3 BNL (end March?)
- CERN
  - Tape Write at given rate (300MB/s)
  - Split over 6 sites for remote write



SC3 – Milestone Decomposition

- File transfer goals:
  - Build up disk – disk transfer speeds to 150MB/s
    - SC2 was 100MB/s – agreed by site
  - Include tape – transfer speeds of 60MB/s
- Tier1 goals:
  - Bring in additional Tier1 sites wrt SC2
    - PIC and Nordic most likely added later: SC4? US-ALICE T1? Others?
- Tier2 goals:
  - Start to bring Tier2 sites into challenge
    - Agree services T2s offer / require
    - On-going plan (more later) to address this via GridPP, INFN etc.
- Experiment goals:
  - Address main offline use cases ***except*** those related to analysis
    - i.e. real data flow out of T0-T1-T2; simulation in from T2-T1
- Service goals:
  - Include CPU (to generate files) and storage
  - Start to add additional components
    - Catalogs, VOs, experiment-specific solutions etc, 3D involvement, …
    - Choice of software components, validation, fallback, …

LCG Service Challenges: Tier2 Issues

A Simple T2 Model

N.B. this may vary from region to region

- Each T2 is configured to upload MC data ***to*** and download data ***via*** a given T1
- In case the T1 is logical unavailable, **wait and retry**
  - MC production might eventually stall
- For data download, **retrieve** via **alternate** route / T1
  - Which may well be at lower speed, but hopefully rare
- Data residing at a T1 other than 'preferred' T1 is transparently delivered through appropriate network route
  - T1s need at least as good interconnectivity as to T0 (?)

Basic T2 Services

- T2s will need to provide services for data up- and down-load
- ➢ **Assume that this uses the same components as between T0 and T1s**
- Assume that this also includes an SRM interface to local disk pool manager
  - This / these should / could also be provided through LCG
- Networking requirements need to be further studied but current estimates suggest 1Gbit connections will satisfy needs of a given experiment
  - Can be dramatically affected by analysis model, heavy ion data processing etc

Which Candidate T2 Sites?

- Would be useful to have:
  - Good local support from relevant experiment(s)
  - Some experience with disk pool mgr and file transfer s/w
  - 'Sufficient' local CPU and storage resources
  - Manpower available to participate in SC3+
    - And also define relevant objectives?
  - 1Gbit/s network connection to T1?
    - Probably a luxury at this stage…
- First T2 site(s) will no doubt be a learning process
- Hope to (semi-)automate this so that adding new sites can be achieved with low overhead

Possible T2 Procedure(?)

- **Work through existing structures where possible**
  - **INFN in Italy, GridPP in the UK,** HEPiX, etc.
  - USATLAS, USCMS etc.
  - Probably need also some 'regional' events
  - Complemented with workshops at CERN and elsewhere
- **Work has already started with GridPP + INFN**
  - Other candidates T2s (next)
  - Choosing T2s for SC3 together with experiments
- **Strong interest also expressed from US sites**
  - Work through BNL and FNAL / US-ATLAS and US-CMS **+ ALICE**
  - US T2 sites a solved issue?

Tier 2s

- New version of 'Joining the SC as a Tier-1'
  - Test with DESY/FZK – April
- 'How to be a Tier-2' document
  - DESY (CMS + ATLAS)
  - Lancaster (ATLAS)
  - London (CMS)
  - Torino (ALICE)
  - Scotgrid (LHCb)
- Each Tier-2 is associated with a Tier-1 who is responsible for getting them set up
- Services at Tier-2 are **managed storage** and **reliable file transfer**

File Transfer Software SC3

Gavin McCance – JRA1 Data Management Cluster

Service Challenge Meeting

March 15, 2005, Lyon, France

Reliable File Transfer Requirements

- LCG created a set of requirements based on the Robust Data Transfer Service Challenge
- LCG and gLite teams translated this into a detailed architecture and design document for the software and the service
  - A prototype (radiant) was created to test out the architecture and used in SC1 and SC2
  - gLite FTS ("File Transfer Service") is an instantiation of the same architecture and design, and is the candidate for use in SC3
    - Current version of FTS and SC2 radiant software are interoperable

Did we do it?

- The goal was to demonstrate ~ 1 week unattended continuous running of the software for the SC meeting
  - On setup 1: demonstrated 1 week ~continuous running
  - But, 2 interventions were necessary
  - Setup 2: higher bandwidth, newer version of software
  - After false start has been stable for a few days
    - 1 intervention necessary (same cause as on setup1)
- Issues uncovered and bugs found (and fixed!)
- We now have a test setup we can use to keep stressing the software

Test setups

- Plan is to maintain these test setups 24/7
  - They've already been vital in finding issues that we could not find on more limited test setups
- Test 1 (low bandwidth) should be used for testing the new version of the software
  - Make sure no bugs have been introduced
  - Try to maintain stable service level, but upgrade to test new versions
- Test 2 (higher bandwidth) should be used to demonstrate stability and performance
  - Maintain service level
  - Upgrade less frequently, only when new versions have demonstrated stability on smaller test setup
  - Ramp up this test setup as we head to SC3
    - More machines
    - More sites
    - Define operating procedures

LCG Service Challenges: Experiment Plans

Experiment involvement

- SC1 and SC2 do not involve the experiments
- At highest level, SC3 tests all offline Use Cases ***except*** for analysis
- Future Service Challenges will include also analysis

➢ **This adds a number of significant issues over and above robust file transfer!**

- Must be discussed and planned now

ALICE vs. LCG Service Challenges

- We could:
  - Sample (bunches of) "RAW" events stored at T0 from our Catalogue
  - Reconstruct at T0
  - Ship from T0 to T1's
  - Reconstruct at T1 **with calibration data**
  - **Store/Catalogue the output**
- As soon as T2's start to join SC3:
  - Keep going with reconstruction steps

    +
  - Simulate events at T2's
  - Ship from T2 to T1's
  - Reconstruct at T1's and store/catalogue the output

ALICE vs. LCG Service Challenges

- What would we need for SC3?
  - AliRoot deployed on LCG/SC3 sites - ALICE
  - Our AliEn server with: - ALICE
    - task queue for SC3 jobs
    - catalogue to sample existing MC events and mimic raw data generation from DAQ
  - UI(s) for submission to LCG/SC3 - LCG
  - WMS + CE/SE Services on SC3 - LCG
  - Appropriate amount of storage resources - LCG
  - Appropriate JDL files for the different tasks - ALICE
  - Access to the ALICE AliEn Data Catalogue from LCG

ALICE vs. LCG Service Challenges

- Last step:
  - Try the analysis of reconstructed data
- That is SC4:
  - We have some more time to think about it

ATLAS & SC3

- July: SC3 phase 1
  - Infrastructure performance demonstration
  - Little direct ATLAS involvement
  - ATLAS observes performance of components and services to guide future adoption decisions

- **Input from ATLAS database group**

ATLAS & SC3

- September: SC3 phase 2
  - Experiment testing of computing model
  - Running
    - 'real' software
    - production and data management systems
    - but working with throw-away data.
    - ATLAS production involvement
    - Release 11 scaling debugging
    - Debug scaling for distributed conditions data access, calibration/alignment, DDM, event data distribution and discovery
    - T0 exercise testing

ATLAS & SC3

- Mid-October: SC3 phase 3
  - Service phase: becomes a production facility
  - Production, adding more T2s
  - Operations at SC3 scale producing and distributing useful data
  - New DDM system deployed and operating
  - Conduct distributed calibration/align scaling test
  - Conduct T0 exercise
  - Progressively integrate new tier centers into DDM system.
  - After T0 exercise move to steady state operations for T0 processing and data handling workflows

CMS & SC3

- Would like to use SC3 ‘service’ asap
- Are already shipping data around & processing it as foreseen in SC3 service phase
- More details in e.g. FNAL SC presentation

Service Challenge III

Service Challenge III - stated Goal by LCG:

- **Achieve 50% of the nominal data rate**
- Evaluating tape capacity to achieve challenge while still supporting the experiment
- **Add significant degrees of additional complexity**
- file catalogs
  - OSG currently has only experiment supported catalogs, CMS has a prototype data management system at the same time, so exactly what this means within the experiment needs to be understood
- VO management software
  - Anxious to reconcile VO management infrastucture in time for the challenge. Within OSG we have started using some of the more advanced functionality of VOMS that we would clearly like to maintain.
- Uses the CMS's offline frameworks to generate the data and drive the data movement.
- **Requirements are still under negotiation**

LHCb

- Would like to see robust infrastructure before getting involved
- Timescale: October 2005
- Expecting LCG to provide somewhat more than other experiments
  - i.e. higher level functionality that file transfer service

Experiment plans - Summary

- SC3 phases
  - Setup and config - July + August
  - Experiment software with throwaway data - September
  - Service phase
    - ATLAS – Mid October
    - ALICE – July would be best…
    - LHCb – post-October
    - CMS – July (or sooner)
- Tier-0 exercise
- Distribution to Tier-1
- …

Milestones and Workplan

- Presentation to PEB next Tuesday
- Draft SC TDR chapter on Web
  - http://agenda.cern.ch/askArchive.php?base=agenda&categ=a051047&id=a051047s1t3%2Fmoreinfo%2FService-Challenge-TDR.doc
  - Will be updated following discussions this week
- Need to react asap to experiments' SC3 plans
- More discussions with T1s and T2s on implications of experiment involvement in SC3.

Summary

- We continue to learn a lot from the on-going SCs
- In practice, even 'straight-forward' things involve significant effort / negotation etc.
  - e.g. network configuration at CERN for SC2/SC3
- There is a clear high-level roadmap for SC3 and even SC4
- But a frightening amount of work buried in the detail…
- ➢ **Should not underestimate additional complexity of SC3/4 over SC1/2!**
- Thanks for all the great participation and help!