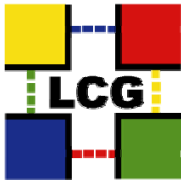
# WLCG GridKa+T2s Workshop

## Site Report

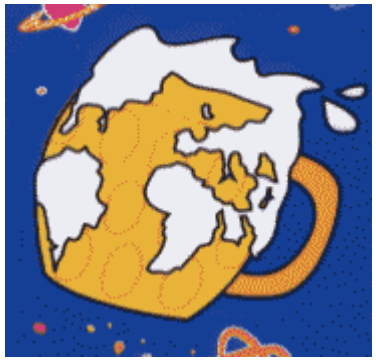
---

Presenter, Site, Country




Worldwide LHC Computing Grid

Distributed Production Environment for Physics data Processing



# Overview

- Grid services offered by this site
  - BDII, CE, SRM-enabled SE, …
  - LFC, VO-boxes (both local (e.g. ALICE) and remote (e.g. ATLAS)
  - 3D services (SQUID, local MySQL or other DB services)
  - O/S; middleware; hardware (CPU, disk, tape) status and outlook
  - Support and operations staff + expertise and outlook
  - Issues & Concerns
- Participation to date in SC4
  - Activities; results; issues
- Participation in remainder of 2006
  - and beyond??
- Some examples from June T2 survey follow…


| Problem Response Time and Availability targets Tier-2 Centres | | | |
|---|---|---|---|
| ***Service*** | ***Maximum delay in responding to operational problems*** | | ***availability*** |
| | ***Prime time*** | ***Other periods*** | |
| **End-user analysis facility** | **2 hours** | **72 hours** | **95%** |
| **Other services** | **12 hours** | **72 hours** | **95%** |



les.robertson@cern.ch

# T2 Size

- Big differences among T2s
- Level of resources planned at LHC startup
  - CPU : 400+ if 1 VO, 800+ if 4 Vos (very few exceptions)
  - Disk : 50 to 800 TB !!! (not proportional to number of Vos)
    - 3 T2s plan 2500+ CPUs
    - Some T2s probably devoted to MC
  - Network (external) : 1 Gb/s (1 10 Gb/s planned, 2 0.5 Gb/s)
  - No MSS planned : 2 exceptions (between 50 and 100 TB)
    - Less than disk space
- FTE : big discrepancies
  - From 1 to 13, majority between 4 and 6
  - Not related to T2 size (at first glance)
  - May be some confusion with the question : FTE vs. people

# Sites / T2

- Number of sites making the T2 : 1 to 8 !
  - Site : geographical
  - 1 site : ½ of (answering) T2s
  - 2 sites : 4
  - 3+ : 8
  - Number of sites seen by the MW : sometimes 1 for the whole T2, sometimes more than 1 / site…
    - Question not asked explicitly : assume generally 1 / site ?
- Largest T2s are federations
- National choices
  - Italy : all T2s are 1 site and support (mainly) 1 VO
  - Several countries have only one T2 made of several sites
  - UK has 4 federated T2s
  - Related to local institute configuration : lot of small labs vs. universities ?

# T2 OS / MW versions

- OS : SL(C)3 32-bit mainly
  - Majority (> 75%) using CERN SL (SLC)
  - RHEL3 x 2, CentOS planned at 1 T2
  - Interest in SL4 32- ou 64-bit but generally waiting for MW to be ready and/or CERN to do it first…
    - GRIF already has WNs running SL4 64-bit (LCG 2.7)
- MW : LCG 2.7 everywhere (almost)
  - 1 ½ T2 running gLite3, 1 LCG 2.6
  - NorduGrid using ARC
  - INFN using INFN-G (very close to LCG, same version)
  - gLite3 upgrade planned everywhere : ½ by end of June

# T2 Administration

- Questions focused on distributed/federated T2s
- Mainly “distributed administration” = each site independently
  - Often a technical coordinator able to act at each site
  - A few sites thinking about inter-site logins : ssh, gsissh, sudo…
  - Sometimes, vendor tools used
  - 1 federated T2 with totally independent sites : 1 meeting / year
- Deployment : site independence mainly
  - Sometimes agreement of minimum set of tools
  - GRIF exception : deployment managed by Quattor from a unique repository
    - More details during Quattor tutorial on Friday
  - Mainly YAIM (+KS), 4 Quattor, 1 Rocks
  - Not necessarily same batch scheduler or SE product

# T2 CE + LRMS

- Most common configuration = 1 CE / site
  - No CE spanning sites (some expression of interest : 2)
  - Sometimes several per site, e.g. 1 CE / VO
  - Generally not seen as problem : let MW / experiment SW deal with the situation
- LRMS : Torque/PBS w/ or without MAUI
  - Several SGE, 1 Condor : better integration into MW asked
  - No question on fairshare usage
    - GRIF experience : critical for efficient sharing of resources between VOs
  - No question on simulation/analysis co-existence
  - GRIF would like to look at multi-cluster technologies to allow transparent cross-submission preserving fairshare
    - Not easy to deal with data location
    - Probably efficient only with 10 Gb/s connections between T2 sites

# T2 SE

- Only ½ answered questions about SE : difficult to interpret
  - Answers : 2/3 using DPM, 1/3 dCache, 1 Classic (?), ARC
  - No consistency inside a federated T2. Some plan T2 to choose in the future
- 1 SE / site everywhere (almost)
  - 1 T2 with 1 SE / VO
  - 1 site with 2 SEs
  - No plan for a unified SE across a federated T2

# T1 Relationship

- Not all T2s have a preferred T1 yet
  - CMS has too many T2s in Europe compared to number of T1
  - No T1 doing management at T2
  - Some federated T2s (2) have a different reference T1 at each site
    - Matrix consistency between experiments ?
  - Main T1s (from answers) : CNAF, CC IN2P3, FZK, RAL
    - Some being reference T1 for very far T2 (e.g. CC IN2P3 for Tokyo)

# T2 Helpdesk and Support

- Question a little bit imprecise : wide range of answers
- Helpdesk : majority has nothing special set up
  - Rely on GGUS or national helpdesk generally
  - Sometimes not really formal
- Support : from 0.5 to 3 people
  - Mainly 9x5, 1 24x7
  - Some T2s : participation to national helpdesk

# T2 Participation to SC

- SC3 : 1/3 participated
- SC4 : majority will participate
  - 1 No, 1 may be, 1 did not answer

# Requirements for MW…

- Wide range of wishes, requests…
- MW quality : several asked for better tested releases
  - Reliable, dependable, documented upgrade
  - Simpler docs and how-to
- Improved MW support for distributed T2s
  - Consolidated view : resource usage, fairshare, job status
  - Guidelines / Best practices for distributed T2 set-up
  - GOC DB : should support notion of site in resource description , **should allow downtime on a resource without suspending SFT for the whole MW site (T2)**
    - Big concern for federated T2s seen as 1 MW site (e.g. GRIF)
- Enhancements
  - Improved support for SGE and Condor in MW
  - Xroot support integrated into MW

# ... Requirements for MW

- Miscellanous
  - DPM srmCopy
  - Central logging (instead of 10s of files)
  - Drop of VO box (not from me !)
  - Yum instead of apt (1)
  - Quattor templates for gLite3 (for me !)

# Conclusions…

- Picture is complex… but we already knew that
  - Resource size, FTE, number of VOs supported…
- Many T2s are ready for production and hope to participate to SC4
  - Critical for experiments (except LHCb)
  - Cannot postpone T2 participation to SC5…
  - Most of them have no experience with data transfers
- Federated T2s are not an exception
  - No major MW obstacle but they are mainly separated MW sites
  - Sometimes only a political/administrative coordination
  - A (declared) large T2 can hide several small sites : actual impact remains to be seen
  - If successful, could allow setup of new T2s in the future by federation of small sites

# … Conclusions

- Some interest to build “distributed T2s”
  - 1 MW site with resources geographically distributed
    - E.g. GRIF (Paris region)
  - Critical : tighter technical coordination and good inter-site connexion (1+ Gb/s)
  - Need better consolidated reports from MW
    - Monitoring tool for the whole T2 (e.g. Lemon)
    - Consolidated accounting
    - Consolidated site view (job status…)
  - Site BDII redundancy is critical
    - Recommend BDII sub-hierarchy per site (done by GRIF) ?
  - Not clear if 1 CE/SE per site is optimal or if there are other viable options
    - Avoid defeating co-location of jobs with data done by experiment frameworks
    - Would provide benefits if able to share the load between sites in case one is overloaded (multi-cluster features only in commercial products : LSF and Moab)