Working Group 4: Fabric Management

- **Overall goals**: look at site experience in the area of fabric management; focus on fabric management/operations in the context of grid services; understand where common problems or weak areas are; try to converge on some best practices. How can we reduce the frequently heard lament that local problems are one of the main reasons for suboptimal grid reliability?

- Tentative list of topics:
  - **System installations** (LCFGng; Quattor; manual; script-based; what else?). Experience? Which requirements? (e.g.: why do you want/need Quattor? Or manual installation procedures? Or other things? Mix and match?) Which commitments?
  - **Batch/scheduling systems**; flavors? Work on testing / development / improvements in this area? Documentation/management tools? How do we define/enforce policies? What policies? (Hard limits? Fairshares?)
  - How do we **monitor the fabric** and what do we do with monitoring info; explore what sites are using / have developed; or what they would like to have, and miss. How do we certify/select fabric components? How do we provide status information to fabric users?
  - **Fabric set-ups** for the grid (shared file systems? MPI support? Parallel FS? Hyperthreading? Bad/good experiences to share? Use cases? How do we manage storage? Networking?)
  - **Upgrade procedures**, how are they triggered, and how do we cope with them? For example: security/performance patches to clusters; upgrades (or downgrades) to batch systems
  - How do/can we provide fabric-related **feedback to LCG/EGEE**?
  - How do we **disseminate information** about fabric/grid issues? (Fabric training for sites? Training on how batch systems are used – for app developers? Do we [want to] publish policies?)
  - **Conflicts** with other (non-grid / non-LCG) uses of the fabric?
  - Can we identify **points of contact** for (some of) these topics?
  - How do we **keep on discussing** these things? Hepix? Some sort of LCG/EGEE-specific forum?
  - ?

- Format
  - For each topic: informal presentations from site operations; discussion; actions, timeline

Initial comments on the list of topics

- Produce a document for site administrators with recipes and How-To
- Check related work being done in South-East Asia in the field of fabric set-ups
- Dissemination of information: should we explicitly tackle training for middleware developers?
- Keep in mind that conflicts with non-grid uses of the fabric may occur in both small and big sites
- Consider LAN issues, specifically related to the service challenges
- Can we clarify the distinction between fabric management (WP4) and software management (WP5)

WG4: Summary Report

- Attendance: ~25 people, 15 sites represented
- Initial list of topics proved to be too long to be discussed in its entirety
  - A lot of interest in fabric management issues
  - We need to find a way to continue these conversations (more on this later)
- We did cover:
  - System installations
  - Batch/scheduling systems
  - Fabric monitoring
  - Some considerations on fabric set-ups
  - Joint session WG4/WG5 on ETT-related problems and software installation methods (see report from WG5)
  - Communication channels
- Summary of actions at the end of this presentation

Working Group 4: Fabric Management

- **Overall goals**: look at site experience in the area of fabric management; focus on fabric management/operations in the context of grid services; understand where common problems or weak areas are; try to converge on some best practices. How can we reduce the frequently heard lament that local problems are one of the main reasons for suboptimal grid reliability?

- Tentative list of topics:
  - **System installations** (LCFGng; Quattor; manual; script-based; what else?). Experience? Which requirements? (e.g.: why do you want/need Quattor? Or manual installation procedures? Or other things? Mix and match?) Which commitments?
  - **Batch/scheduling systems**; flavors? Work on testing / development / improvements in this area? Documentation/management tools? How do we define/enforce policies? What policies? (Hard limits? Fairshares?)
  - How do we **monitor the fabric** and what do we do with monitoring info; explore what sites are using / have developed; or what they would like to have, and miss. How do we certify/select fabric components? How do we provide status information to fabric users?
  - **Fabric set-ups** for the grid (shared file systems? MPI support? Parallel FS? Hyperthreading? Bad/good experiences to share? Use cases? How do we manage storage? Networking?)
  - **Upgrade procedures**, how are they triggered, and how do we cope with them? For example: security/performance patches to clusters; upgrades (or downgrades) to batch systems
  - How do/can we provide fabric-related **feedback to LCG/EGEE**?
  - How do we **disseminate information** about fabric/grid issues? (Fabric training for sites? Training on how batch systems are used – for app developers? Do we [want to] publish policies?)
  - **Conflicts** with other (non-grid / non-LCG) uses of the fabric?
  - Can we identify **points of contact** for (some of) these topics?
  - How do we **keep on discussing** these things? Hepix? Some sort of LCG/EGEE-specific forum?
  - ?

- Format
  - For each topic: informal presentations from site operations; discussion; actions, timeline

Fabric Management WG – CERN, 2-4/11/2004 – Davide Salomoni

System Installations (1)

- Unanimous desire to have a community-supported way of dealing with system installations
  - Rather than reinventing/discovering solutions independently
- Very strong interest in using Quattor
  - Most sites currently use LCFG, with a couple of sites mix and match systems (LCFG + scripts – not the ones presented on Tuesday by Laurence/Louis)
  - 4 sites (including CERN) have already installed Quattor servers
    - Key point: provide at the very least installation, configuration of Worker Nodes
    - Installation reports being written (NIKHEF, RAL, CNAF)
  - Pros
    - Modularity, possibility of switching on/off individual components
    - Load balancing, scalability capabilities
  - Cons
    - Perfectible installation guide, perceived complexity in setting the system up
    - Several components have already been written at CERN, but there is no clear publicity, nor testing outside of CERN (example: afs component)
    - LCG components are also available (Cal Loomis), but, again, with limited documentation

System Installations (2)

- Positioning: generic installation procedures (formerly: "manual installations") are not competing with Quattor
  - Rather, we expect to build on top of the generic installation scripts
- Resources
  - Within November, we'll have a formal "Quattor WG" under the auspices of the GDB
  - SA1 people can officially work on Quattor (i.e., account for time spent on Quattor in the EGEE timesheets). There is no "free lunch" here either; we expect the formation of a virtuous circle with this WG to help with configuration and support of Quattor components.
- Actions
  - We need not wait for the formation of the Quattor WG to start doing something
  - Provide details on Quattor scalability (comparable to an analogous work made in EDG times for LCFG) and resilience, with an How-To – German Cancio
  - Better document which components are available at CERN, and encourage their usage outside of CERN – German Cancio
  - Organize a Quattor training for interested parties – CERN (around December)

Batch/Scheduling Systems

- Several sites use vanilla OpenPBS, as shipped with LCG2
  - For these sites, a common suggestion was to upgrade to Torque/Maui
  - And use qstat caching
  - There is much interest in understanding how to configure fairshares, and awareness that fairshare configurations can be complex (plus, maui docs are generally not an easy read)
  - There is no clear answer on which tools one could use to simplify fairshare configurations (regardless of the scheduling system, be it e.g. maui or LSF)
  - While most of the discussion has focused on maui/torque, several batch systems are in use; the majority is running OpenPBS/Torque (w/ or w/o maui), but there's also LSF, Condor, BQS, Sun Grid Engine
- Some sites still evaluating/comparing various batch/scheduling systems
  - But need to reproduce stress conditions similar e.g. to DC
  - Interest in exploring reduncancy at the pbs server level – nobody currently doing that
- Actions
  - Fairshare session scheduled for Hepix, May 2005. We want to do something before then.
  - Write a document on how to configure maui for simple use cases, targeting mostly smaller sites. There are several ongoing drafts that already partially deal with this – we need to integrate them. RAL, CERN-GD (Sophie Lemaitre), NIKHEF.
  - There are useful patches to both Torque and Maui, developed by some of the sites: integrate them into a set of RPMs, and also try to get those patches accepted by the official Torque/Maui distributions. RAL, NIKHEF (+ CERN-GD?)

Fabric Monitoring

- Very diverse picture, all the usual suspects are present: Nagios, Ganglia, GridICE, LEMON, a variety of self-made scripts
  - A complaint: there are too many monitoring systems. A suggestion: this needs to be discussed in detail within SA1.
    - But most of the sites also say that they plan to invest time and resources in finding more comprehensive monitoring system (so, potentially, we'll have even more monitoring systems)
  - Most of these systems focus on hardware-related metrics (CPU Temperature, fan rotation speed, etc)
  - Very worrying: only 2 sites actually have systems that proactively try to deal with "black holes"
    - On the other hand, several sites have experienced black holes, and rely on just manual intervention following automatic notification mechanisms (best case), or user notifications (worst case). This can lead to varying degrees of unreliability.
- Actions
  - Create a subgroup to document which tests should be made at the fabric level to verify that WNs, batch system are in good order
    - Try out Piotr's script locally
    - Validate and integrate existing scripts
    - Weizmann, RAL, CNAF, IN2P3, NIKHEF, CERN (German Cancio)

Some Farm Set-up Issues (1)

- Hyper-Threading
  - Most sites have it disabled
  - One site (CNAF) enabled it after an explicit request from an experiment. But this does not really scale or work if:
    - You have multi-purpose center, or fairshare allocations (rather than statically defined sub-clusters)
    - You do not have properly configured nodes (for example, enough memory to avoid resource contention, enough jobs running on nodes to avoid suboptimal cpu usage, suitable OS)
    - You have conflicting requests from experiments
  - A difficulty is that experiments do not seem interested in benchmarking their applications (is this feasible at all)
  - Disk servers supposedly benefit from HT
  - Actions
    - Gather existing HT data, and perform some tests – CNAF, IN2P3, RAL

Some Farm Set-up Issues (2)

- What do you specify as technical requirements in procurement tenders?
  - Thermal properties (horror stories heard)
- Serial-ATA vs. SCSI disks: is the price difference justified? This is an example of a topic being discussed in several places (CHEP, Hepix), and we need to channel the discussions into this forum/WG
- Operational fabric security: a site would like to have an up-to-date list of IP networks to which allow outbound connectivity (to reduce the impact of a DoS attack). A topic for WG1? (operational security)
- Survey of common use cases for farm set-ups: review document being written in SE Asia (to be out approx. in 1 week) – All

Communication Channels

- We would like to have a single point of contact for grid-sysadmin-related topics. But the format for this is not agreed yet
  - A web page/portal? Who is going to maintain it?
    - Yet another suggestion to SA1?
  - LCG-ROLLOUT is too generic
    - Do we need/want another mailing list?
- There are topics we would have liked to discuss, and did not have the time
  - Should getting together in a similar WS be repeated [more frequently]? Should we wait 6 months for the Hepix meeting in Karlsruhe?
    - We should report about what we discuss and do in the context of this WG's activities at Hepix
  - An item for discussion in the GDB? How is WG2 (operational support) going to address this?

WG4: Summary of Actions

- Provide details on Quattor scalability (comparable to an analogous work made in EDG times for LCFG) and resilience, with an How-To – German Cancio
- Better document which components are available at CERN, and encourage their usage outside of CERN – German Cancio
- Organize a Quattor training for interested parties – CERN (around December)
- Write a document on how to configure maui for simple use cases, targeting mostly smaller sites. There are several ongoing drafts that already partially deal with this – we need to integrate them. RAL, CERN-GD (Sophie Lemaitre), NIKHEF.
- There are useful patches to both Torque and Maui, developed by some of the sites: integrate them into a set of RPMs, and also try to get those patches accepted by the official Torque/Maui distributions. RAL, NIKHEF (+ CERN-GD?)
- Create a subgroup to document which tests should be made at the fabric level to verify that WNs, batch system are in good order
  - Try out Piotr's script locally
  - Validate and integrate existing scripts
  - Weizmann, RAL, CNAF, IN2P3, NIKHEF, CERN (German Cancio)
- Gather existing HT data, and perform some tests – CNAF, IN2P3, RAL
- Survey of common use cases for farm set-ups: review document being written in SE Asia (to be out approx. in 1 week) – All
- Communication channels? (open point)