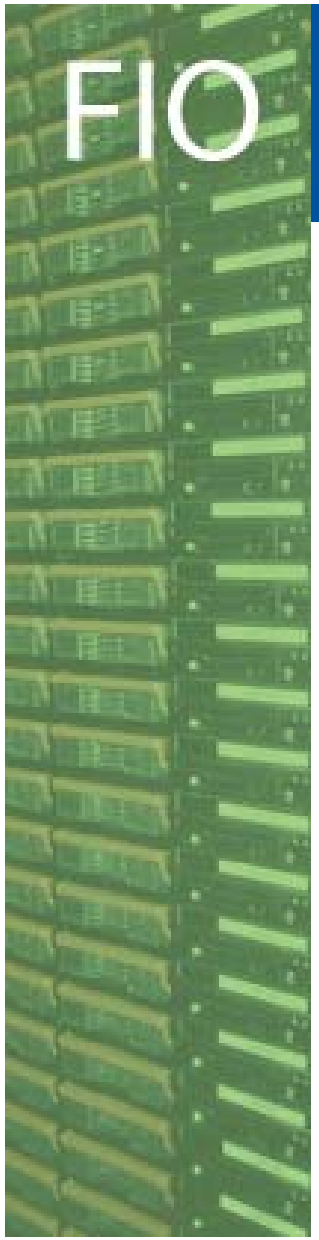


Fabric Infrastructure and Operations

CERN IT Department

Operations procedures

CERN Site Report

Grid operations workshop
Stockholm 13 June 2007

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

FIO

Many questions to answer in 15'!

CERN IT Department

- Tools used in daily operations
- Features missing
- Examples of the most frequent scheduled interventions at CERN
- Examples of the most frequent unscheduled interventions at CERN
- Points to improve in communication with ROC, other sites, VOs, rest of the world...
- Plan for deployment of updates/new versions in continuous operation
- Communication with users
- Correlation of cross-site issues
- Percentage of real site problems are detected and reported by the COD before we know about them
- Usefulness of operations bodies/meetings

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Overview

- CERN fabric by numbers
  - CE (DNS loadbalanced clusters)
    - 17 LCG CE
    - 5 gLite CE
  - WNs: ~2800 WNs, 7000 CPUs
  - SE
    - SRM – 20 nodes (v11, v22)
    - Storage system: CASTOR version 2
      - ~1.5PB disk cache (LHC VOs only)
      - 3PB tape used, ~140 tape drives (T10K, 3592B, 9940B)
- Service not covered in this presentation
  - FTS
  - gLite WMS
  - VOMS
  - Database services

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

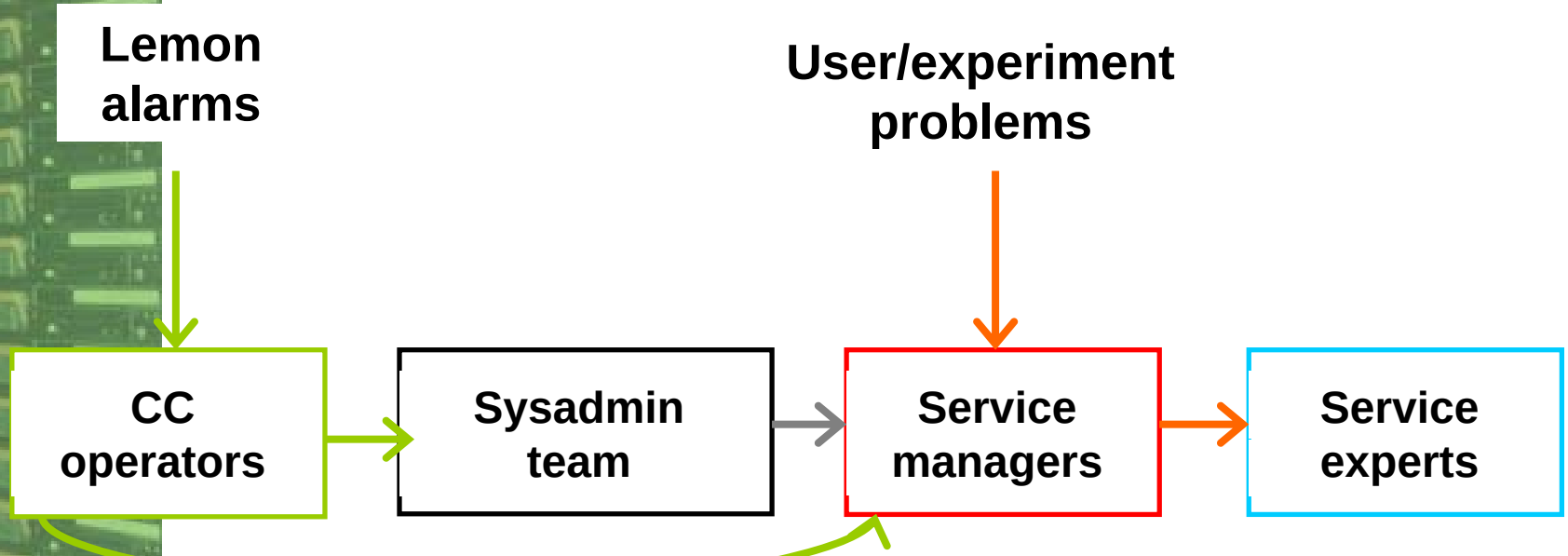
Tools used in daily operations

- ELFms toolsuite (http://cern.ch/elfms)
  - Monitoring, exceptions & alarms
  - System installation and configuratio
  - Hardware and state management
  - Service display

- Operational strategy
  - Automate where possible
    - Lemon actuators
    - Automated state changes upon declared exceptions
  - If not, document the procedure, if possible for non service-expert use (sysadmin, operators)
- Operation and support workflows
  - Remedy
  - GGUS tickets are fed into local Remedy

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Grid services support flow



- 24 x 7 coverage
- 1st level alarm handling
- Driven by procedures

- 2nd level alarms handling
- System installation, configuration
- Manage Hardware repairs
- 24 x 7 coverage through Piquet service

- Service responsible
- Applies s/w upgrades, changes, provides procedures
- Entry point for support lines
- SMod, GMod – person on rota

- Different experts
- May end up in bug to developers

Problem reports come to the Service Managers via many different flows, using many different tools, directly and indirectly.
This still needs some tuning

Piquet Working Group looking into the possibility to establish piquet services for critical services

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Features missing

- Support for scheduled upgrades (ongoing)
  - Needed for partitioning of clusters into “prod” / “new” / “test” areas
  - Moving from 1-2 linux upgrade/week to ‘scheduled upgrades’ where we only apply software updates between LHC runs
  - Aim to validate the complete software stack (OS + middleware + applications). Need experiment help!
  - Planned for ~end summer
- Lemon GUI service views (ongoing)
  - Allows to tailor Lemon GUI to show service specific metrics (e.g. castor request rate)
  - Planned for July
- Alarm SAM failures for CERN-PROD

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Examples of the most frequent scheduled interventions at CERN

- Upgrades to storage system
  - Intrusive oracle intervention on the castor name server database (e.g. moving to new hardware, quarterly patches)
    - Affects all VOs
    - Batch normally paused during the intervention
  - New castor stager version, intrusive oracle interventions
    - Usually only one VO at a time
  - Tape library interventions
    - Degradation (long wait time for tape recalls) for all VOs
- Router/switch upgrades or changes
- Changes to the CE cluster
  - Draining of nodes for h/w intervention
  - Only degradation, no service interruption

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Examples of the most frequent unscheduled interventions at CERN

- Power cuts / cooling failures
- CASTOR stager meltdowns
  - Usually only one VO
  - If left for longer period (e.g. over a Weekend) it can result in global degradation of the SRM service affecting all VOs
- CE overload

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Points to improve in communication with ROC, other sites, VOs, rest of the world...

- Weekly site reports
  - Need to streamline better local reporting to fit the grid
    - Content
      - Consistency
      - Avoid duplication where possible
      - Flag relevance (grid/non-grid)
    - Synchronization in time
    - Avoid duplication
  - Site availability
    - Test categories can be confusing (CE – SE correlations)
- VO communication: when to use broadcast?
  - Currently using a mixture of mailing lists, hn, broadcasts
  - Trying to move to use broadcast more consistently but today this is not sufficient
- Weekly GGUS ticket review with ROC is very useful!

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Plan for deployment of updates/new versions in continuous operation

- We plan to only apply software updates between LHC runs
- This means that we will prepare updates during LHC runs
- Aim to validate the complete software stack (OS + middleware + applications). Need experiment help!
- To be done in 2007: make a test cluster available to validate upcoming releases
- Hope to move to this mode in ~September

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Communication with users

- Channels for service announcements agreed with each experiment VO managers individually
  - XXX-support@cern.ch, hn-YYY@cern.ch
  - List of named people
  - Broadcasts
  - + combinations of above
- Support: power-users know the shortcuts but we are trying to insist on use of support flows

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Correlation of cross-site issues

- Particularly relevant for data transfers problems
  - Who owns the problem from start?
  - Who determines what site/component is at fault?
  - Successful problem determination usually depends on a coordinated investigations between the involved sites
  - Access to log files and other relevant information
  - Login access to remote hosts for ‘bare’ transfer tests

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Percentage of real site problems are detected and reported by the COD before we know about them

FIO

CERN IT Department

- Have not gathered any statistics but most probably <50%
- What is a ‘real site problem’?
  - Degradation
    - Failure of one CE node out of a loadbalanced cluster of 10
  - Partial outage
    - SE for one VO is unavailable due to CASTOR problems
  - Full outage
- Even if the problem is known, the COD reports are in general useful for us
  - Probing can detect configuration problems that we can’t see ourselves
    - Recent case with groups.conf file on the CEs

Usefulness of operations bodies/meetings

- Weekly grid operations meeting is of little / no use for CERN-PROD service mgrs
  - Prefer expanding our weekly meeting with ROC mgrs (ticket review) to include
    - Review site availability
    - Weekly service reports
    - Coordination of scheduled interventions
    - Review GGUS tickets
- Operations bodies
  - User support
    - Good filtering but sometimes slow escalation
  - Operation
    - COD reports are in general useful

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

FIO

The Baltic see COD is delicious, try it while you're in Stockholm

CERN IT Department

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it