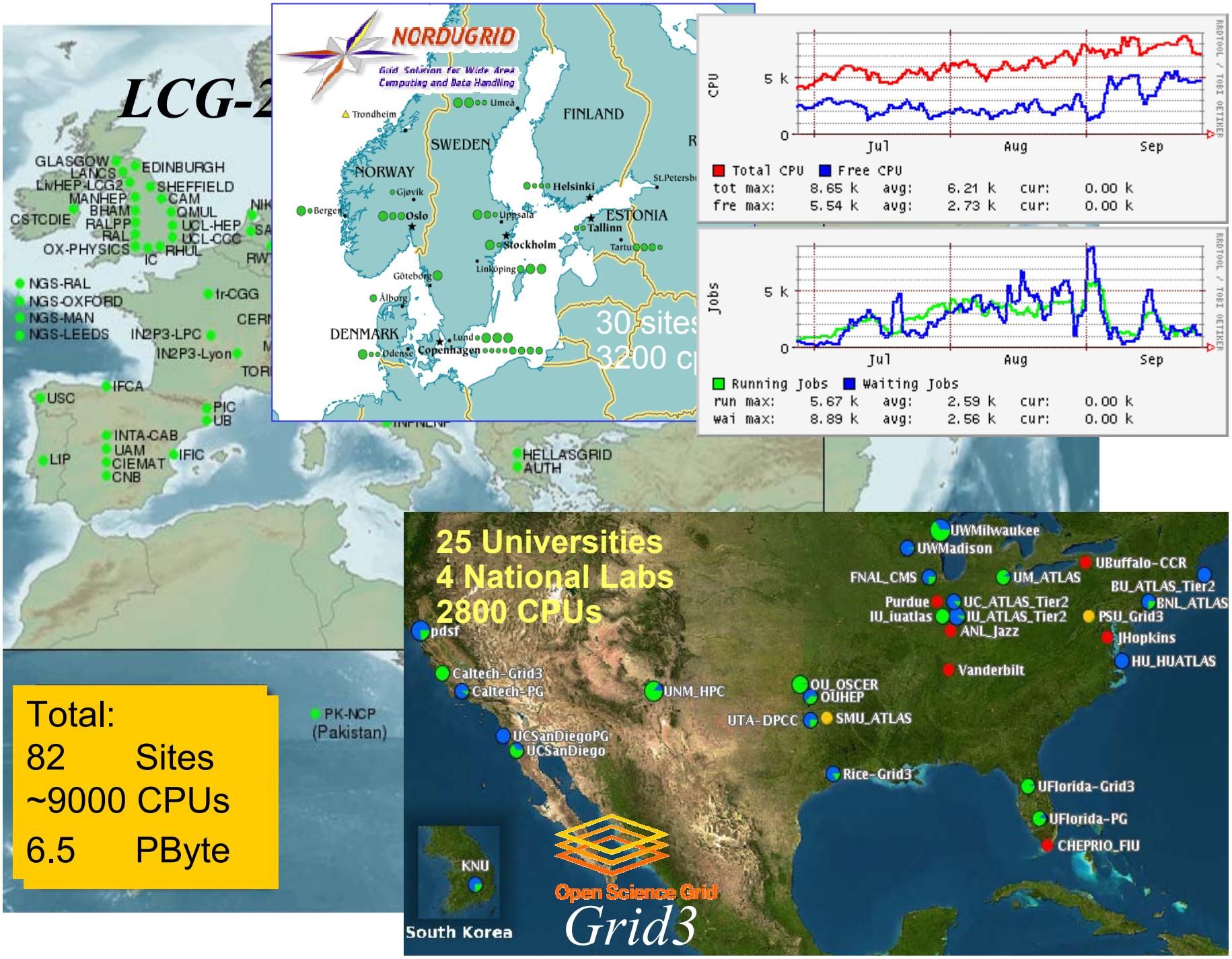
LCG Operations and Fabric Workshop
2-4 November 2004

Introduction

Ian Bird
CERN IT-GD


LCG-2
NORDUGRID
Grid Solution for Wide Area Computing and Data Handling
FINLAND
SWEDEN
NORWAY
ESTONIA
DENMARK
Trondheim
Umea
Helsinki
Gjovik
Bergen
Oslo
Uppsala
Tallinn
Stockholm
Tartu
Linköping
Göteborg
Ålborg
Lund
Odense
Copenhagen
30 sites
3200 cp
CPU
5 k
0
Jul
Aug
Sep
Total CPU
Free CPU
tot max: 8.65 k avg: 6.21 k cur: 0.00 k
fre max: 5.54 k avg: 2.73 k cur: 0.00 k
Jobs
Running Jobs
Waiting Jobs
run max: 5.67 k avg: 2.59 k cur: 0.00 k
wai max: 8.89 k avg: 2.56 k cur: 0.00 k
GLASGOW
EDINBURGH
LANCS
LivHEP-LCG2
SHEFFIELD
MANHEP
CAM
BHAM
QMUL
CSTCDIE
RALPP
UCL-HEP
RAL
UCL-CCC
OX-PHYSICS
RHUL
IC
NGS-RAL
NGS-OXFORD
NGS-MAN
NGS-LEEDS
Ir-CGG
IN2P3-LPC
IN2P3-Lyon
IFCA
USC
PIC
UB
INTA-CAB
UAM
IFIC
LIP
CIEMAT
CNB
HELLASGRID
AUTH
PK-NCP
(Pakistan)
Total:
82 Sites
~9000 CPUs
6.5 PByte
25 Universities
4 National Labs
2800 CPUs
UWMilwaukee
UWMadison
UBuffalo-CCR
FNAL_CMS
UM_ATLAS
BU_ATLAS_Tier2
Purdue
UC_ATLAS_Tier2
BNL_ATLAS
IU_iuatlas
IU_ATLAS_Tier2
PSU_Grid3
ANL_Jazz
JHopkins
pdsf
Vanderbilt
HU_HUATLAS
Caltech-Grid3
Caltech-PG
UNM_HPC
OU_OSCER
OUHEP
UTA-DPCC
SMU_ATLAS
UCSanDiegoPG
UCSanDiego
Rice-Grid3
UFlorida-Grid3
UFlorida-PG
CHEPRIO_FIU
KNU
South Korea
Open Science Grid
Grid3

Status – 1

- LCG covers many sites (>80) now – both large and small
  - Large sites – existing infrastructures – need to add-on grid interfaces, use existing tools, etc.
  - Small sites want a completely packaged, push-button, out-of-the-box installation (including batch system, etc)
  - Satisfying both simultaneously is hard – requires very flexible packaging, installation, and configuration tools and procedures
    - A lot of effort had to be invested in this area
- There are many problems – but in the end we are quite successful
  - Middleware is relatively stable and reliable
  - System is used in production
  - System is reasonably easy to install now – >80 sites
  - Now have a basis on which to incrementally build essential functionality
- This infrastructure forms the basis of the initial EGEE production service

Status – 2

- Level of complexity anticipated for LHC → O(100) sites
- We already have >80!

- We probably see most of the operations issues already now

- Data challenges over the past 10 months:
  - Probably the first time such a set of large scale grid productions has been done

Data challenges – summary

- Significant efforts invested on all sides – very fruitful collaborations
  - Unfortunately, DCs were first time the LCG-2 system had been used
  - Adaptations were essential – adapting experiment software to middleware and vice-versa – as limitations/capabilities were exposed
  - Many problems were recognised and addressed during the challenges
- Middleware is actually quite stable now
- But – job efficiency is not high – for many reasons
- Started to see some basic underlying issues:
  - Of implementation (lack of error handling, scalability, etc)
  - Of underlying models (workload management)
  - Perhaps also of fabric services – batch systems ?
- But – single largest issue is lack of stable operations

Problems during the data challenges

- Sites suffering from configuration and operational problems
  - inadequate resources on some sites (hardware, human..)
  - this is now the main source of failures
- Load balancing between different sites is problematic
  - jobs can be "attracted" to sites that have inadequate resources
  - modern batch systems are too complex and dynamic to summarize their behaviour in a few values in the IS
- Identification of problems is difficult
  - distributed environment, access to many logfiles needed…..
  - status of monitoring tools
- Handling thousands of jobs is time consuming and tedious
  - Support for bulk operation is not adequate
- Performance and scalability of services
  - storage (access and number of files)
  - job submission
  - information system
  - file catalogues
- Services suffered from hardware problems

Addressing some of the problems

- Services deployed redundantly (multiple instances)
  - addressing performance and availability problems
  - work started on fail over procedures
- Development efforts continue for improved data management tools
  - reliable data transfer services
  - catalogues
  - disk pool managers
    - dCache (FNAL/DESY)
    - disk pool manager (CERN) (simple solution for smaller sites)
- Installation procedures have been simplified and made more robust (incremental process)
  - less configuration problems

Addressing problems and next steps

- Outstanding middleware problems have been collected during the DCs
  - https://edms.cern.ch/file/495809/0.4/Broker-Requirements.pdf
  - And in the GAG document
  - 1st **systematic** confrontation of required functionalities with capabilities of the existing middleware
    - Some can be patched, worked around,
    - Most has to be direct input as essential requirements to future developments
- Port to Scientific Linux is finished
- Started to setup pre-production service for new EGEE middleware
- Moving to the new operations structure given by the EGEE project

Operations so far

- Existing mode is not sustainable
- Very few people involved
  - Monitoring run by (very few) people at CERN, RAL, Taipei
  - Tools not yet sufficient
- No good process in place
- Many problems reported
  - Most get bounced to CERN team (or left to them), but
  - Several other people quite active in posing and addressing questions
  - Need to harness and expand those efforts
- No control over "bad" sites
  - Remove from information system
    - … But which one?
- User support
  - Not at all clear where a user should report problems

LCG and EGEE Operations

- EGEE is funded to operate and support a research grid infrastructure in Europe

- The core infrastructure of the LCG and EGEE grids is now operated as a single service, growing out of LCG service
  - LCG includes US and Asia-Pacific, EGEE includes other sciences
  - Substantial part of infrastructure common to both
- LCG Deployment Manager is the EGEE Operations Manager
  - CERN team (Operations Management Centre) provides coordination, management, and 2nd level support
- Support activities are expanded with the provision of
  - Core Infrastructure Centres (CIC) (4)
  - Regional Operations Centres (ROC) (9)
  - ROCs are coordinated by Italy, outside of CERN (which has no ROC)

LCG → EGEE in Europe

Operations Management Centre

Core Infrastructure Centre

Regional Operations Centre

- Operational support:
  - The LCG GOC is the model for the EGEE CICs
    - CIC's replace the European GOC at RAL
    - Also run essential infrastructure services
    - Provide support for other (non-LHC) applications
    - Provide 2nd level support to ROCs
- User support:
  - Becomes hierarchical
  - Through the Regional Operations Centres (ROC)
    - Act as front-line support for user and operations issues
    - Provide local knowledge and adaptations
- Coordination:
  - At CERN (Operations Management Centre) and CIC for HEP

Summary

- Data challenges – demonstrated:
  - Many m/w functional and performance issues (documented)
  - Main problem is service stability
    - Site fabric management, configuration, change control
    - Etc
  - Grid3 report similar problems …
  - User support process needs improvement
- Now moving into continuous production + service & data challenges

Services

- Existing production service based on LCG-2 will remain in place
- In parallel run pre-production service
  - Demonstrate/test new middleware
  - Hopefully services can migrate to production and replace/supplement existing LCG-2 services
  - Many developments are independent of environment
    - Storage management, file transfer services
- Certification test-bed remains
  - But hopefully needs less integration effort – should now be done by EGEE/JRA1
  - Streamline deployment process
  - Better process to include external developments

How to move forward

- Build an agreed operations model for the next year
  - Should be able to evolve
  - Using EGEE SA1 infrastructure and resources
  - … and collaborations with US (Grid3/OSG), Asia (Taipei,…), and others
  - 5 working groups:
    - Operations support
    - User support
    - Operational security
    - Fabric management issues
    - SW needs and tools → requirements from operations
      - This is important for the long term
- Cannot cover everything – but we should outline the strategy and directions and indicate the major issues to be addressed
- Don't forget simple solutions: (e.g.)
  - Need fabric management training for many sites

Some issues

- Resource Centres:
  - Large sites – have operations staff and/or on-call support
  - Small sites – have no on-call and often little support at all
- Regional Operations Centres:
  - Probably do not provide after-hours or on-call support. If this were the case then the model of support could more include the ROCs. However, it is clear that most ROCs will not have this level of support.
- Core Infrastructure Centres:
  - Must have on-call support after-hours
    - To be rotated through the 4 or 5 active CICs

Thus, a basic question to answer is how much power or control can the CICs have in order to deal with problems when staff at RCs and ROCs are not available?

- Either CICs have rights to manage critical services on sites where there is no support, or
- Have the right to remove "broken" sites and services from the infrastructure.

- Likely that we have all combinations of these …

Issues

- Operational management
  - How much control can/should be assumed by an operations centre?
  - Small sites with little support – can GOCs restart services?
    - More intelligence in the services to recognise problems
- Strong organisation to take operational responsibility
  - Ensure that problems are addressed, traced, reported
- Need site management to take responsibility
- Ensure that Operational security group is in place with good communications
- Simplify service configurations – to avoid mistakes
- Weight of VOs
  - EGEE has many VOs (most still national in scope)
  - Deploying a VO is very heavyweight – must become much simpler

Format of the workshop

- Today and tomorrow morning
  - Plenary session to expose the issues and generate discussion
- Working groups (Wed pm + Thurs am)
  - Provide guidance on solutions/directions/strategy in 5 main areas
  - Each should produce a short document/slides
- Thurs pm
  - Summaries and conclusions