Running a Production Grid Site at the London e-Science Centre

David McBride
(Formerly) Site Administrator for UKI-LT2-IC-LeSC

Department of Computing,
Imperial College, London
Existing Infrastructure

- **Viking**: 260-node SMP Intel cluster, running RedHat 7.2 (as of Summer 2002).
- **Mars**: 204-node SMP AMD Opteron cluster, running RedHat Enterprise Linux 3 (RHEL3).
- **Sun GridEngine** batch system, supporting MPI.
- Bespoke configuration management infrastructure.
- Shared NFS-mounted home directories.
- Many pre-existing local users.
LCG Software Distribution

• Designed to satisfy the needs of their typical user.
  – A particle physicist not trained in the operation and management of a computer cluster.

• Provided as a set of binary RPM packages.
  – Initially, only for RedHat 7.3;
  – Later releases also supported Scientific Linux 3, a derivative of RedHat Enterprise Linux 3 (RHEL3).
  – Later provided monolithic tarball images.

• Automated bootstrap tools also provided, to assist with mass-installation.
LCG Software Distribution, cont.

- However, hard to integrate into an existing environment.
  - No source code available in readily-buildable form
    - SRPMS not provided by distributor
    - Large complicated CVS repository holds code; build process undocumented and hard to replicate.
    - Thus, must either use binary-compatible distribution or virtualisation.
  - Post-installation configuration handled by YAIM
    - Functions poorly documented, if at all.
    - Very little sanity-checking or error handling.
LCG Software Distribution, cont.

- File organisation unhelpful, lacking.
  - Several different top level directories – edg, globus, glite, etc..
  - 4 or 5 different tmp directories.
  - Logging information scattered in several places.
    - Should simply be using syslog!
- Supporting documentation only provided a superficial overview, didn't detail specifics.
Binary Compatibility

- Originally, published binaries were incompatible
  - RedHat 7.3 RPMs are binary incompatible with RedHat 7.2 system libraries, as used on Viking.
  - Reinstalling each Viking node was out of the question.
  - Experimented with using virtualisation – simple `chroot()`, User-Mode Linux (UML)
- Two things happened:
  - Mars, our new RHEL3 cluster was commissioned.
  - LCG-2_4_0 was released with support for RHEL3-compatible operating systems.
Batch System

- LCG ships with, and assumes the use of, the PBS batch system.
- Already had Sun's GridEngine batch system installed.
- Job Submission handled by a Globus 2 GRAM service
  - Uses “JobManager” plugins, implemented in Perl, to interact with batch system-specific interfaces.
  - Didn't provide a JobManager implementation for SGE.
- So we had to write our own.
  - (It helped that we'd already done some work in this area...)
Batch System, cont.

- There is also a separate Information Reporter component
  - Advertises, amongst other things, the current state of the queues provided by the batch system.
  - Again, no SGE implementation available.
- However, implementation was non-trivial!
  - Information system components expect each cluster to advertise a number of distinct 'queues' that jobs can be submitted to.
  - Each queue has different constraints and properties.
  - However, GridEngine doesn't use queues!
Job Processing Models

- **Queue-based model**
  - Batch system advertises individual queues.
  - Each queue specifies constraints on types of job that may be submitted.
  - Each queue served by a specific set of worker nodes.
- **Examples:** PBS

- **Attribute-based model**
  - Batch system has single job entry point.
  - Each job submitted is annotated with its requirements.
  - Each worker node can impose restrictions on what classes of job it will run.
- **Examples:** SGE, Condor
Solution: Virtual Queues

- Define a set of *virtual queues* using job constraints.
  - In effect, “bucket” job requirements into individual classes.
  - Dynamically calculate queue membership at runtime.
- Virtual queue definition completely independent of underlying batch-system configuration.
  - Batch system configuration changes immediately reflected in published statistics
  - Can modify the set of advertised queues with any modification of the batch system.
Virtual Queues: Diagram

Queue-based Model

Resource Broker

Virtual Queue Status

Information Reporter

Cluster Status

Job Manager

Batch System

Queue: 10min

max-wall_time=360

Virtual Queue Configuration

10min: max-wall_time=360
Virtual Queues, cont.

• Works well in practice, with some minor issues:
  - Can violate assumptions made by other tools.
  • GSTAT monitor assumes that jobs belong to at most one queue – calculates total number of jobs by summing queue totals.
    - Can produce wildly inflated numbers if virtual queues are nested or overlap.
    - Users don't always specify job requirements properly
  • Many dteam monitoring jobs exceeded 10min runtime, yet were still submitted to our 10min queue.
Integration issues

- Essential that we do not unnecessarily modify existing cluster environment.
  - Already running in production, used by departments throughout the College.
- So used tarball images, and stored live distribution within a globally-accessible NFS volume.
  - Use 'job preamble' extension to JobManager to source the necessary extra configuration files for Grid users.
  - Existing users see no changes whatsoever.
  - Also allows several LCG installations to co-exist.
Architecture differences

- Binaries provided are for 32-bit systems; however Mars is a 64-bit cluster running 32-bit/64-bit bi-arch operating system.
  - Providing extra 32-bit versions of Perl and the like is straightforward;
  - However, several LCG experiments had significant difficulties adapting their code and control scripts.
I discovered a hole in our internet security.

What?!!

Good grief, man! How could you put a hole in our internet?

Actually, that's not my job. But I'll inform our network management group.

Passing the buck!!! You're a buck passer!!!

Actually, that's not my job. But I'll inform our network management group.

Passing the buck!!! You're a buck passer!!!

Forget it! There's no hole! It got better!

Forget it! There's no hole! It got better!

That's more like it.

I fixed the internet.
Security: Incidents

• Did have one security incident that affected us:
  – A dteam member in Germany had given a copy of his credentials to a colleague, who was then using our site (and others) as a testbed for his software.
  – Spotted unusual job manually, dealt with accordingly.

• Another security incident affected a large number of sites
  – Originally detected by Kostas Georgiou, systems administrator at our sister site in Imperial HEP
  – Dutifully notified other sites, and flagged suspect traffic from other sites.
Security: Incidents, cont.

- “fig.chaw.com” incident:
  - Investigations later determined that passwords and credentials had been exposed for users belonging to more than 30 sites.
  - Futher attacker activity had been detected at 13 of these sites.
  - Root-level compromise of several machines, including shell servers located at CERN.
- First compromise had occurred more than 40 days prior to first detection.
Some of the software shipped with LCG simply hasn't been designed and built with security in mind.

Example: R-GMA

- Database-backed distributed data store.
- Originally deployed without any access controls whatsoever.
- Later added authentication requirement, though no authorization capabilities – anyone with valid credentials may do whatever they please.
Security: Vulnerabilities Handling

- Has not been handled well.
  - Several serious security flaws have not been fixed, despite being known for years.
  - No central Vulnerabilities group; GridPP started their own.
- Didn't have sufficient manpower to deal with issues effectively
- Didn't have necessary leverage to force developers to resolve problems.
- Unwilling to disclose information regarding unfixed vulnerabilities, even when no fix is forthcoming.
Results
Results, cont.

- **Documentation:**
  - Detailed installation notes:
    - [https://www.gridpp.ac.uk/wiki/IC-LeSC](https://www.gridpp.ac.uk/wiki/IC-LeSC)
  - Includes patches to APEL accounting system, currently being adopted upstream.

- **Code:**
  - GridEngine JobManager implementation:
  - GridEngine Information Reporter:
Final Remarks

• Software Engineering:
  – Build process should be easy to replicate.
  – Opaque 'black box' design makes deployment, debugging and re-engineering difficult.
  – Many advantages of open-source model are lost.

• Security:
  – Little compartmentalisation within Grid infrastructure
  – Monoculture: Break Once, Break Anywhere.
  – Continuous monitoring essential; more tools needed.
  – Responsible Disclosure policy should be adopted.
Questions?