Building the M-grid

Arto Teräs <arto.teras@csc.fi>

The 22nd NORDUnet Networking Conference

Svalbard, Norway, April 7th 2005

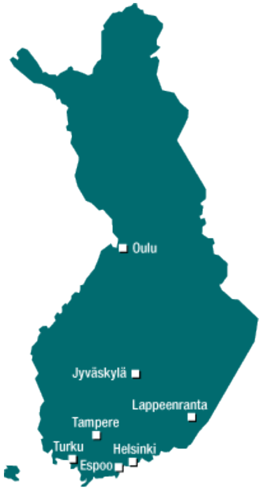
CSC

Contents

- The Finnish Material Sciences Grid (M-grid) overview
- Experiences in deploying the systems
- Rocks Cluster Distribution pros and cons
- Lessons learned in collaborative system administration
- Initial user experiences
- Security challenges
- Conclusion
- Questions

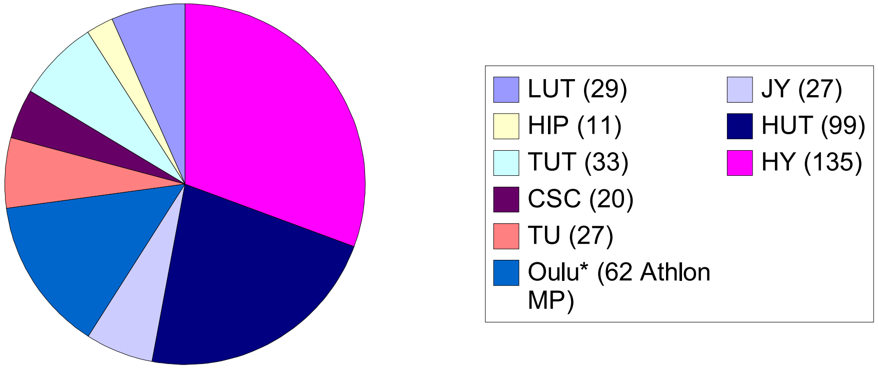
Finnish Material Sciences Grid (M-grid)

- **Joint project between seven Finnish universities, Helsinki Institute of Physics and CSC, the Finnish IT center for science**
- **Jointly funded by the Academy of Finland and the participating universities**
  - Funding application Nov 2003, deployment Oct 2004
- **First large initiative to put Grid middleware into production use in Finland**
- **Based on Linux clusters, targeted for serial and ”pleasantly parallel” applications**
- **Users mainly physicists and chemists**



Hardware and CPU Distribution

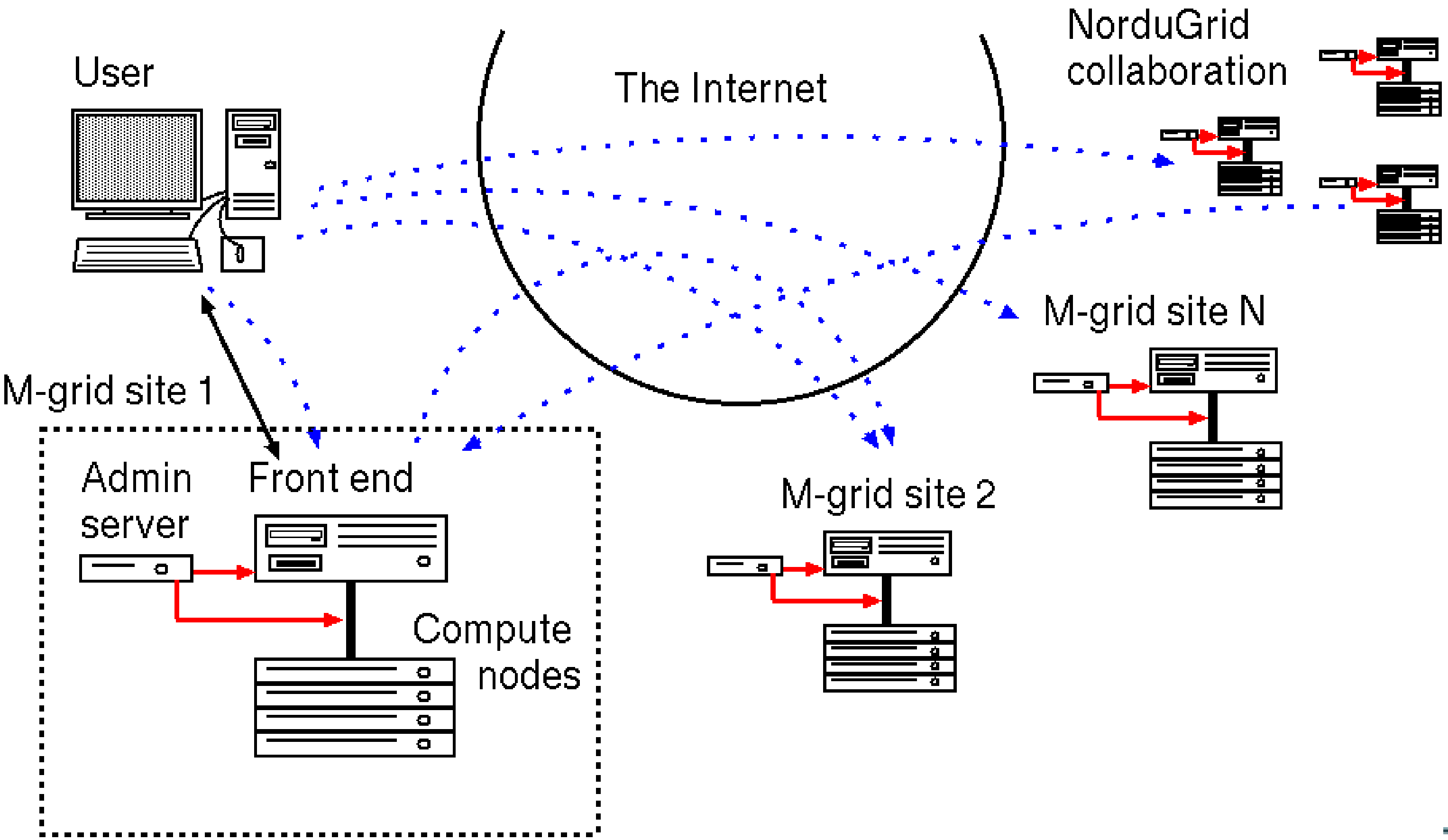
- **Dual AMD Opteron 1.8-2.2 GHz nodes with 2-8 GB memory, 80-320 GB local disk, 1-2 TB shared storage, 2xGbit Ethernet, remote administration hardware**
- **Number of CPUs: 410 (computing nodes only), 1.5 Tflops theoretical computing power**
- **9 sites, size of sites varies greatly**



Software Choices

- **NPACI Rocks Cluster Distribution**
  - Main developers in the San Diego Supercomputing Center, U.S.A.
  - Based on Red Hat Enterprise Linux 3.0 source packages, but customized for clusters
  - http://www.rocksclusters.org
- **NorduGrid ARC Grid middleware**
  - The most popular middleware in Nordic countries, one of the few suitable for production use
  - http://www.nordugrid.org

Grid — The Whole Picture

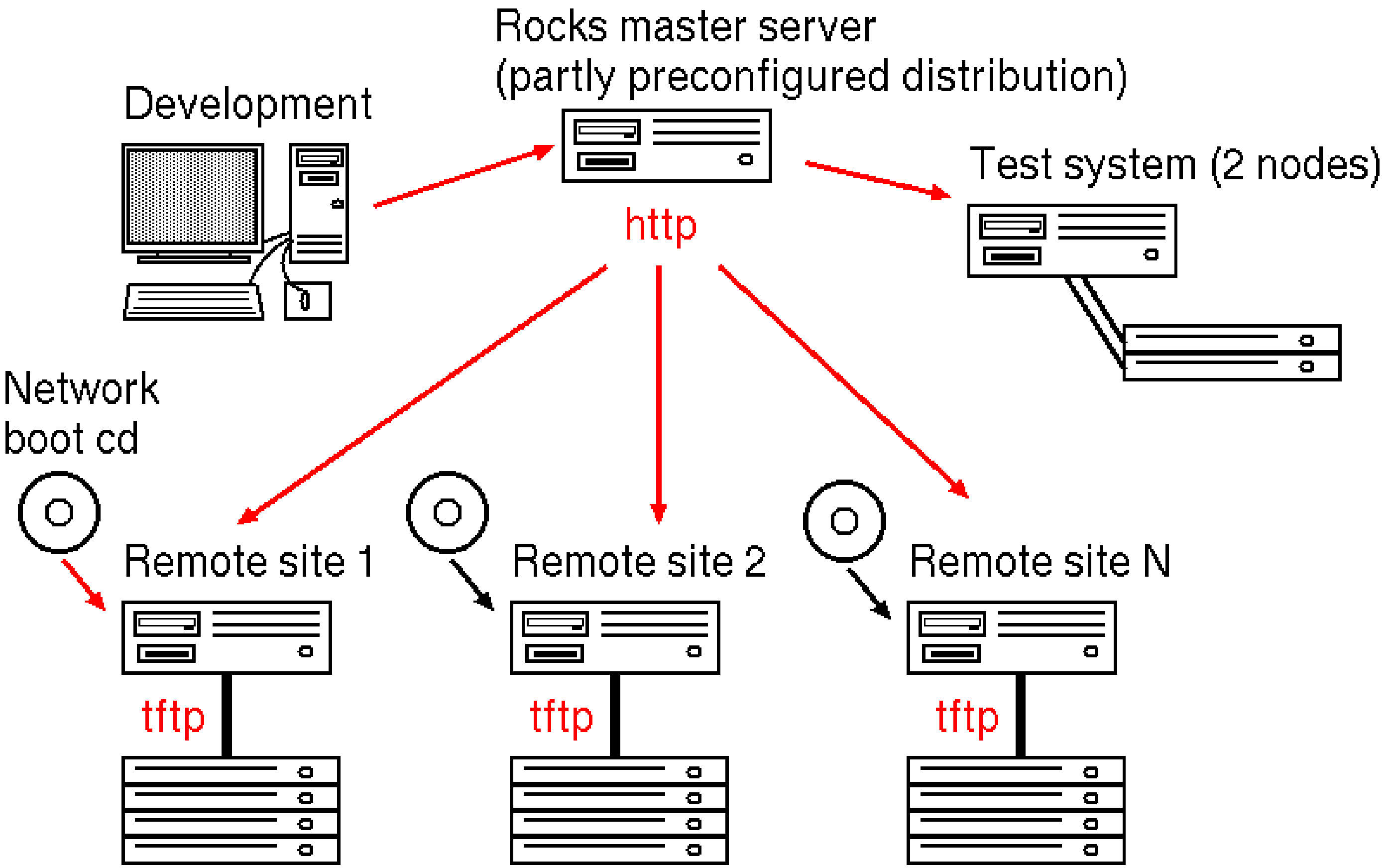


System Administration in M-grid

- **Tasks divided between CSC and site administrators**
- **CSC administrators**
  - Maintain (remotely) the OS, batch queue system, Grid middleware and certain libraries for all sites except Oulu
  - Separate small test cluster for testing new software releases
- **Site administrators**
  - Local applications and libraries, system monitoring, user support
- **Regular meetings of administrators & support network**

CSC

Installation Plan



Deployment Experiences

- **CSC prepared the distribution and a boot cd, local administrators responsible for installing their own cluster**
- **Preparing the distribution took more time than expected**
- **Actual deployment went quite smoothly**
  - Most sites spent less than a day installing the OS and nodes, larger sites took two days
  - One site had strange problems taking more time
- **A few settings which we didn't have preconfigured properly were fixed manually afterwards**

Rocks Pros and Cons

Good:

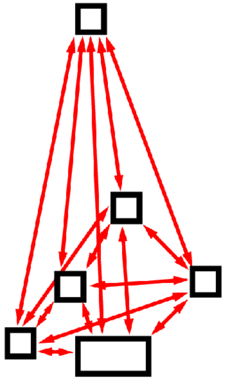
- **Easy to get started, designed for clusters**
- **Nice monitoring tools, many things work out of the box**
- **Most major vendors have their hardware certified for RHEL => Rocks usually works too**

Something to improve:

- **Security updates not provided by the Rocks team (patching using RHEL source rpms ok)**
- **Diagnosis and debugging difficult when customizing the distribution**

Goals of Shared System Administration

- **Centrally administered foundation while maintaining local control**
  - A new paradigm — traditionally in Finland HPC resources have been centralized at CSC
- **Easier for universities than setting up their own system from scratch**
  - However, needs a significant amount of work both from CSC and the local sysadmins
- **Take advantage of the local sysadmin expertise in software used by the local researchers**
  - Faster and better user support



36 pairs for collaboration!

Positive Experiences

- **Local sysadmins have found CSC support valuable**
  - Having also local control (root) is important psychologically
- **Participants have used their expertise to pick up suitable tasks, fruitful discussion on the mailing list**
- **Collaboration has strengthened relationships between groups also in their research**
- **Systems are closer to the user**
  - Easier to talk to the own group sysadmin, less support requests to CSC
- **Local sysadmins are often also users => direct usability feedback to CSC**

Negative Experiences

- **Sun Grid Engine v. 5.3 batch system configurability**
  - Version 6.0 is better designed for clusters
- **Wiki-based FAQ hasn't really taken off**
- **User documentation became scattered**
  - Mainly due to lack of human resources (people assigned to other projects before finishing the docs)
  - Compiling the documentation needs central coordination
- **Some users found support poor**
  - Clearly divided between sites: on some sites users are very happy

Initial User Experiences

- **Users got started relatively quickly: the average total load of the M-grid is above 50%**
- **Performance has been quite satisfactory**
- **Problems centered around Fortran compiler and MPI runs**
  - MPI works, but killed jobs can leave unfreed resources behind
  - Several Fortran compilers available (PGI, Pathscale, Intel, G95++, ...): difficult to find one which would be satisfactory for all users

Grid Use

- **Policy: Users may submit jobs both locally and through Grid interface**
  - Grid jobs have higher priority than local jobs in 20% of each system, and may fill all available free nodes
- **Reality: Middleware installation got delayed so no real experience on Grid use yet**
  - Problems with the 64 bit environment and Sun Grid Engine support took time (solved now)
  - Time will show how users adopt the Grid environment; our collaborative network will hopefully be helpful

CSC

Security Challenges

- **Grid crosses organizational boundaries**

  => Collaboration and mutual trust needed!

- **Some new risks and all the old ones with wider impact area**
  - Compromised user account most probable method of intrusion
- **Definitions of responsibilities necessary to build trust**
  - Risk analysis
  - Acceptable use policy
  - Incident response

Security Challenges (cont.)

- **Getting all the relevant parties involved**
  - Computing centers, university IT departments, local admins, CERTs and also users
  - International collaboration
- **Distributed systems with hundreds of users are always vulnerable**
  - Focus on detecting break-ins quickly
  - Clear procedures how to act when a system is compromised

Conclusion

- **Sharing system administration tasks can work**
  - Partners need to know each other — face to face meetings are very useful in avoiding flame wars
- **User support in a distributed system potentially very good but needs special attention**
- **Grid projects strengthen ties between groups even before actual Grid use**
- **Rocks is a good choice for a cluster toolkit (among others)**
- **International collaboration on security and policies needed**

More Information

- **M-grid home page: http://www.csc.fi/proj/mgrid/**
- **Rocks home page: http://www.rocksclusters.org**
- **NorduGrid home page: http://www.nordugrid.org**
- **Contact people:**
  - Arto Teräs <arto.teras@csc.fi>
  - Kai Nordlund <kai.nordlund@helsinki.fi>
  - Olli-Pekka Lehto <oplehto@csc.fi> (Rocks)
  - Urpo Kaila <urpo.kaila@csc.fi> (security)
- **Thank you! Questions?**