EGEE and HPC: Pride and Prejudice?

Peter Kunszt,
Swiss National Supercomputing Centre CSCS

Istambul, EGEE08

www.eu-egee.org

CSCS
Swiss National Supercomputing Centre

Information Society and Media

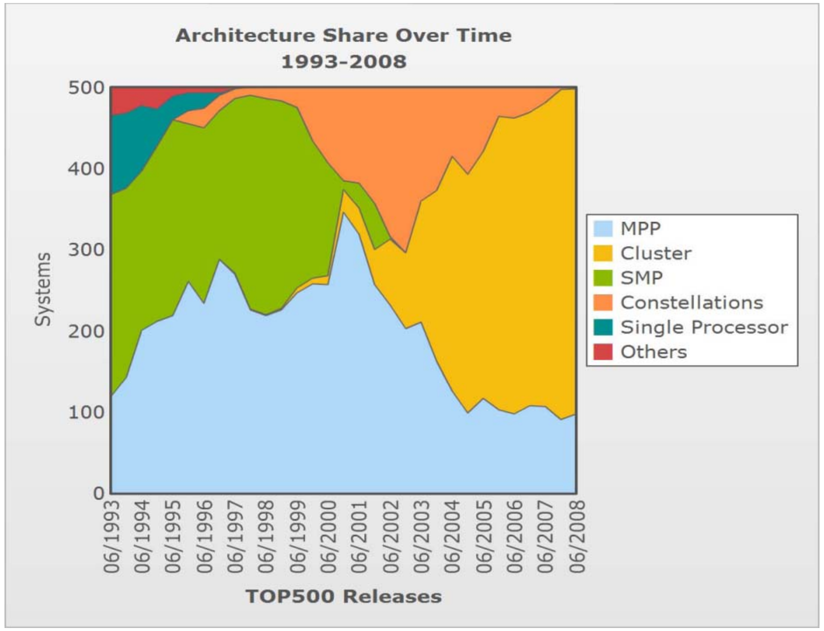
Content

- EGEE and Supercomputers
- Basic differences
- Prejudices debunked

CSCS
Swiss National Supercomputing Centre

Supercomputing Centers in EGEE

- **There are many classical supercomputing centers as EGEE partners**
  - Netherlands: SARA
  - Finland: CSC
  - UK: EPCC, STFC
  - Sweden: KTH PDC
  - Poland: PSNC, WCS
  - Spain: CESGA
  - Germany: LRZ, Karlsruhe
  - Switzerland: CSCS
  - And more... Sorry if you‘re not listed here



- **There may even be some clusters used also by EGEE in the top500. Clusters are dominating the picture.**

CSCS
Swiss National Supercomputing Centre

Basic differences

EGEE

- Virtual Organizations
- Certificates
- Self-organized access
- Hierarchical support
- Mostly data-centric
- ~75‘000 CPUs in production (available as shown in gstat, 21.9.2008)

Supercomputing Centers

- Projects
- User accounts
- Peer reviewed access
- Central support
- Mostly computation
- ~750‘000 CPUs in production just in top10 2.4MCPU in top500

EGEE would be among top10
if it was a single system, but not top5

More differences

EGEE

- Many single-CPU jobs
- Cannot run many-CPU supercomputing jobs
- Heterogeneous
- Off-the-shelf market-proven technology

Supercomputers

- Fewer many-CPU jobs
- Can also run EGEE single CPU jobs
- Homogeneous
- Top of the notch latest hottest technology

CSCS
Swiss National Supercomputing Centre

Do comparisons make sense?

EGEE

- Many single-CPU jobs
- Cannot run many-CPU supercomputing jobs
- Heterogeneous
- Off-the-shelf market-proven technology

Supercomputers

- Fewer many-CPU jobs
- Can also run EGEE single CPU jobs
- Homogeneous
- Top of the notch latest hottest technology

Both enable World Class Science
Both are needed by World Class Scientists
Both can be proud.
To argue with one AGAINST the other makes no sense

CSCS
Swiss National Supercomputing Centre

Part of the same Ecosystem

- **Final goal: computing ‚instruments' to enable research.**

- **Supercomputers have a very specific use case for science**
  - Many domains can only progress using large-scale simulations that cannot be executed on losely coupled systems
- **Very large clusters have another specific use case**
  - Parameter studies, statistical studies, data mining problems don't need tightly coupled systems
- **Grids provide the support for collaborations**
  - Resource and data sharing
  - Make use of complementary resources
  - Interface standardization

Capacity and Capability

Capacity resources

- Used simultaneously by many people or by many jobs
- Standardized versions of common operating systems
- Well-understood and mainstream
- ‘Work-horses’ in computing providing a lot of ***capacity***.

Capability resources

- Used by very few people simultaneously, or even just by one person at any given time.
- Have special properties, like very large memory, a lot of interconnected CPUs, etc
- Have the **capability** to run exceptionally large or difficult problems.

CSCS
Swiss National Supercomputing Centre

Prejudices

- **Capacity is cheaper, you get more TFlops/$**
  - Supercomputers are more expensive
  - Supercomputers have a short lifetime and are fragile
  - Clusters have much more competitive and aggressive pricing

- **No expensive porting of applications**
  - Supercomputers use latest hardware, apps need to be rewritten
  - Commercial codes are available for clusters but only ‚too late' for supercomputers, only when they are outside of their lifecycle

- **Most SC applications can be rewritten to run on capacity**
  - With clever new algorithms capacity clusters can run supercomputing applications
  - Indeed the supercomputing applications of 4 years ago are running now on clusters

CSCS
Swiss National Supercomputing Centre

Debunk 1: Capacity is Cheaper

- **Clusters run separate identical entities, loosely coupled**
  - Energy consumption scales linearly, always several steps behind supercomputers
  - Today, counting 3 years of operations, clusters are more expensive than supercomputers due to increased energy cost
  - EGEE has many dozens of FTEs to operate its infrastructure, a supercomputer of the same power needs much less people
- **Supercomputers are not more fragile**
  - They run large workloads, upon system degradation those workloads cannot be run anymore. Clusters simply degrade.
  - MTBF is better with clusters since they use hardware previously developed in supercomputers. Without supercomputers in production, clusters would be just as fragile.
- **Supercomputers have competitive pricing**
  - Supercomputing firms are not making money with top-end machines
  - Many government contracts esp. in U.S.
  - Access to top people inside companies, not just the local salesman

CSCS
Swiss National Supercomputing Centre

Debunk 2: Application Porting

- **Clusters have also application porting issues**
  - Also regular technology updates
  - Simply profiting from work done on supercomputers, so if those would not be there, the same effort would be needed
  - Updates in operating systems and libraries still there

- **Commercial codes not always off-the-shelf**
  - Many need very specific hardware and expect you to buy a dedicated cluster
  - Most will however gladly work with you to port their codes to new machines
  - Again, clusters profit from work done with supercomputers

Debunk 3: Smart Algorithms

- **Algorithms always need enhancing**
  - Smart new algorithms that run the same app just as fast on a cluster as on the supercomputer will always run even faster on the supercomputer
  - The best code of today running on a computer of 1980 would be MUCH faster than the code of 1980 running on the best computer of today
- **Having new hardware you have new possibilities**
  - Many more opportunities for enhanced algorithms
- **Everyone profits from better algorithms**
  - Riding the tech wave
  - Feedback into new hardware design

Prejudices – the other way

- **Grids are just a hype**
  - What remains when the funding is all used up?
- **Grids are complicated to use**
  - Users need too much time to learn how it‘s done
- **Grids have weak support**
  - Too many people involved, users need personal contact
- **Grids are maintained by amateurs**
  - Most clusters are very small installations maintained by a grad student, low quality of service

Debunk 4: Grids

- **Grids have been around for 10 years now**
  - Address a basic need of science: Collaborations and sharing, this will not go away
  - Term has simply been overloaded – be careful with its use!
- **There is a very large well-organized user community**
  - There have been dozens of schools and workshops over the years
  - Many people have been trained and user interfaces have been improved
- **Standards are driven by Grids**
  - GGF-OGF has achieved a lot, also with industry
- **Professionalized support and monitoring**
  - EGEE has demonstrated how to do it, still a lot to learn
  - First instance of such distributed heterogeneous infrastructure
- **Memorandum of Understandings**
  - Between sites and VOs – see LCG

CSCS
Swiss National Supercomputing Centre

We‘re all on the Same Side

- In many (smaller) countries EGEE and supercomputers is no contradiction and these are maintained by the same entitiy
- Both computing infrastructures are needed and have different roles in the computing ecosystem
- The science has to be put first!

Examples: learn from each other

- **What Supercomputing centers do well (and Grids dont)**
  - Peer reviewed resource allocation. The resources are fairly allocated in a competitive fashion.
  - User support. The users get to see their supporters and work closely together with them in joint projects.
  - Technology previews. Planning ahead for the next phase.

- **What Grids do well (and Supercomputing centers dont)**
  - Standardization. Uniform look-and-feel for users.Designing interfaces that will last through the next few technology upgrades.
  - Organizing scientific (sub)domains into Virtual Organizations.
  - Interdisciplinary collaborations. Scientists learning to apply each other‘s methods, entering into new projects.

CSCS
Swiss National Supercomputing Centre

The Message to You

- **Supercomputers, Clusters and Grids are integral part of the computing ecosystem. Don‘t compare them!**
- **EGEE and Supercomputing centers both have strengths and weaknesses. Learn from each other!**
- **Scientists don‘t care about our prejudices**
  - Strategically, always put the application before the infrastructure
  - Technology evolution management needs all the tech layers to be around
- **Policy makers need to be made aware of this**
  - HPC and Grids should not need to compete for funding
  - Unified policies need still to be worked out – see EGI_DS and PRACE