



HPC Cloud Interactive User support

Floris Sluiter
Project leader
SARA computing & networking services

BiG Grid

the dutch e-science grid

SARA Project involvements



European Grid Infrastructure
Towards a sustainable grid infrastructure



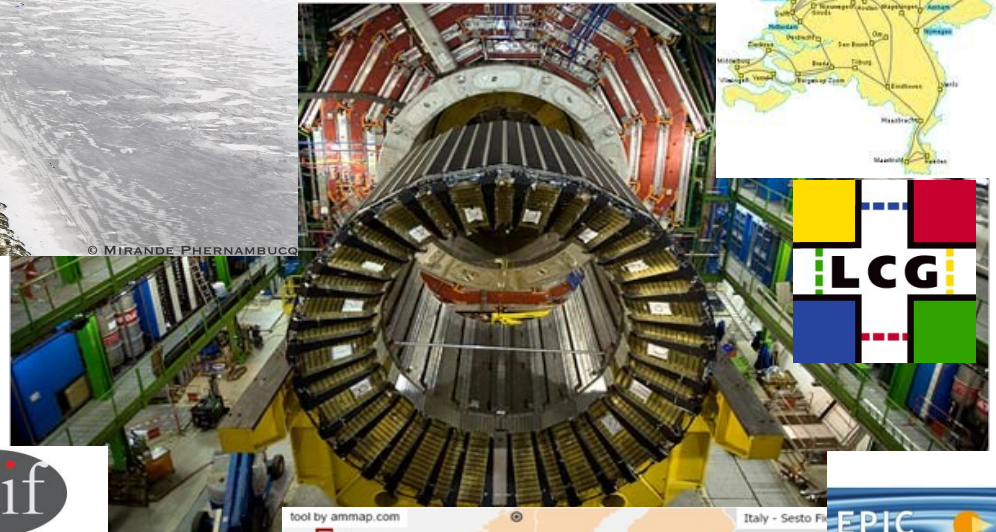
LIGHTHOUSE



BiG Grid
the dutch e-science grid



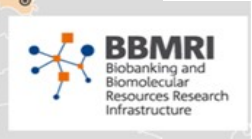
© MIRANDE PERNAMBUCO



Global Lambda Integrated Facility



High Performance Digital Media Network



Biobanking and Biomolecular Resources Research Infrastructure



Pan-European Research Infrastructure on High Performance Computing



HPC Cloud Philosophy

HPC Cloud Computing:

***Self Service Dynamically Scalable Computing
Facilities***

***Cloud computing is not about new technology, it is
about new uses of technology***



BiG Grid

the dutch e-science grid





Our starting point for BiG Grid HPC Cloud

- Easy & standard(familiar) access protocol
 - name&password (or x509 certificates)
 - Support ad hoc collaborations
 - Support Cloud standards (OCCI, OVF, CDMI, WebdDAV)
- Zero client software install
 - Standard browser with java applets & javascript enabled
 - Additional tools optional: VNC viewer, ssh/putty etc
- User has free choice
 - Operating System & applications
 - Root rights in VM and on private network
 - Configuration of private cluster
 - Anything goes: Multi core, multi node, long running (services, databases)
- It doesn't have to be optimal, great is good enough
 - Virtualization overhead acceptable, only thousands of users not millions , only terabytes not petabytes



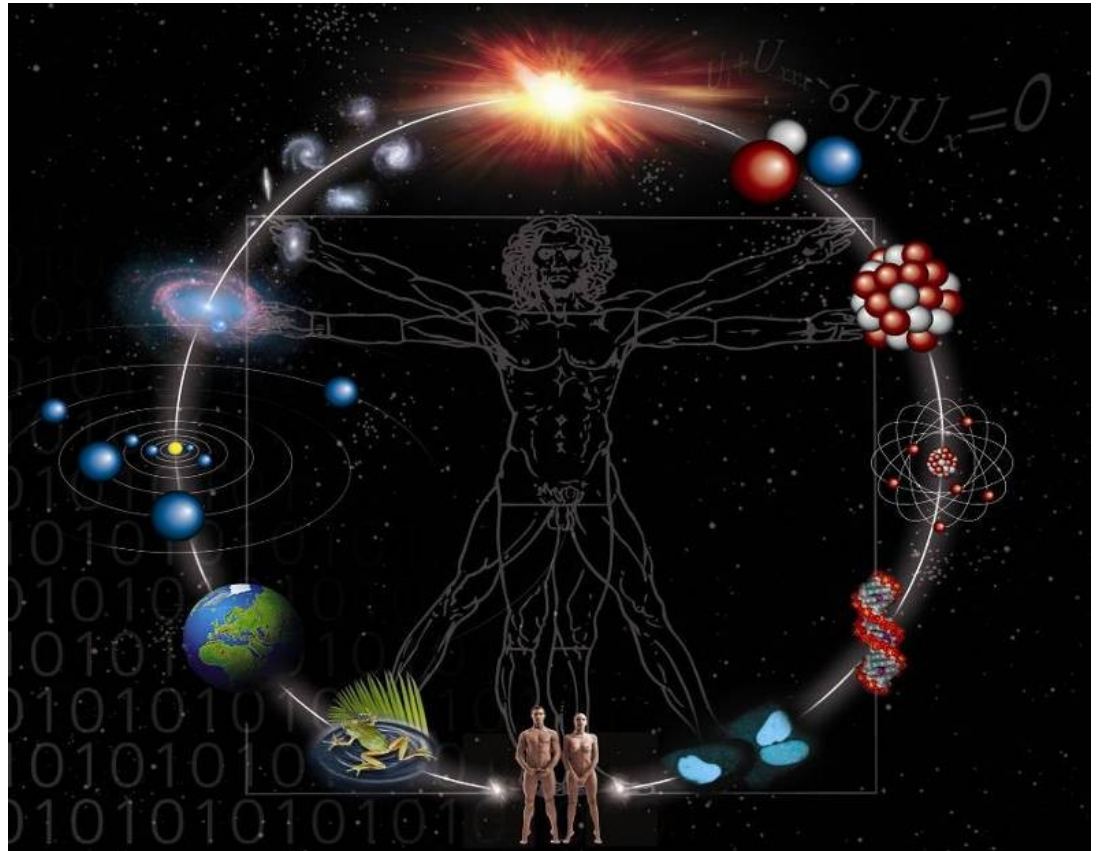
BiG Grid

the dutch e-science grid

Users of Scientific Computing

- High Energy Physics
- Atomic and molecular physics (DNA);
- Life sciences (cell biology);
- Human interaction (all human sciences from linguistics to even phobia studies)

- from the big bang;
- to astronomy;
- science of the solar system;
- earth (climate and geophysics);
- into life and biodiversity.



Slide courtesy of prof. F. Linde, Nikhef

BiG Grid

the dutch e-science grid





Users in pilot and beta phase

- From the start at least 50% in use
- Currently between 70-80%

- 50 user groups
 - 30 % from lifesciences (bio-informatics)
 - Psychology
 - Geography
 - Linguistics
 - Econometrists
- Currently 19 requests on waitinglist (!)
- Festive Launch at 4 th October in Amsterdam (www.sara.nl → Agenda)

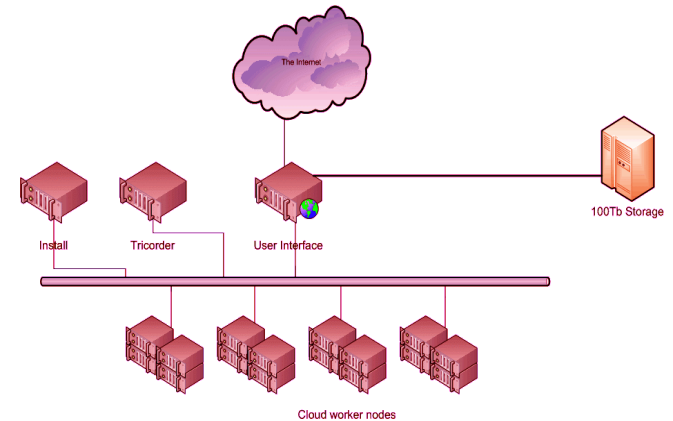


BiG Grid

the dutch e-science grid

The product: Virtual Private HPC Cluster

- We offer:
 - Fully configurable **HPC** Cluster (a cluster from scratch)
 - Fast CPU
 - Large Memory (256GB/32 cores)
 - High Bandwidth (10Gbit/s)
 - Large and fast storage (400Tbyte)
- Users will be **root** inside their own cluster
- **Free** choice of OS, etc
- And/Or **use** existing VMs:
Examples, Templates, Clones of Laptop, Downloaded VMs, etc
- **Public** IP possible (subject to security scan)



Platform and tools:

- Redmine collaboration portal
- Custom GUI (Open Source)
- Open Nebula + custom add-ons
- CDMI storage interface

BiG Grid

the dutch e-science grid

HPC Cloud, what is it good for?

- Interactive applications
- High Memory, Large data
- Same data, many different applications
(Cloud reduces porting efforts!)
- Dynamic, fast changing and complicated applications
- Clusters with Multi Operating Systems

- Collaboration
- Flexible and Versatile

- System architecture is expandable and scalable



BiG Grid

the dutch e-science grid

User collaboration Portal

- Redmine (www.redmine.org)

#	Project	Tracker	Subject
82	SARA Cloud Infrastructure	Support	provide documentation on "Clone my laptop" (Feedback)
106	SARA Cloud infrastructure	Feature	Proper User Agreement (Feedback)

Home

This is the main community support site for the HPC Cloud Infrastructure at SARA; it contains several sub-projects to separate the relevant information streams.

- **Documentation**
A project containing all documentation
- **Support**
The support channel for all question, feature requests and bugs.
- **All Projects**
An overview of all other projects, including the community projects and public vm templates
- **The management console**
Brings you to the main management console to manage your virtual machines, virtual networks etc.

You can always contact the support team via cloud-support@sara.nl

Latest news

Evaluating OpenNebula 2.0: Wiki update
Installation process description added to the wiki.
Added by Jhon Masschelein 6 months ago

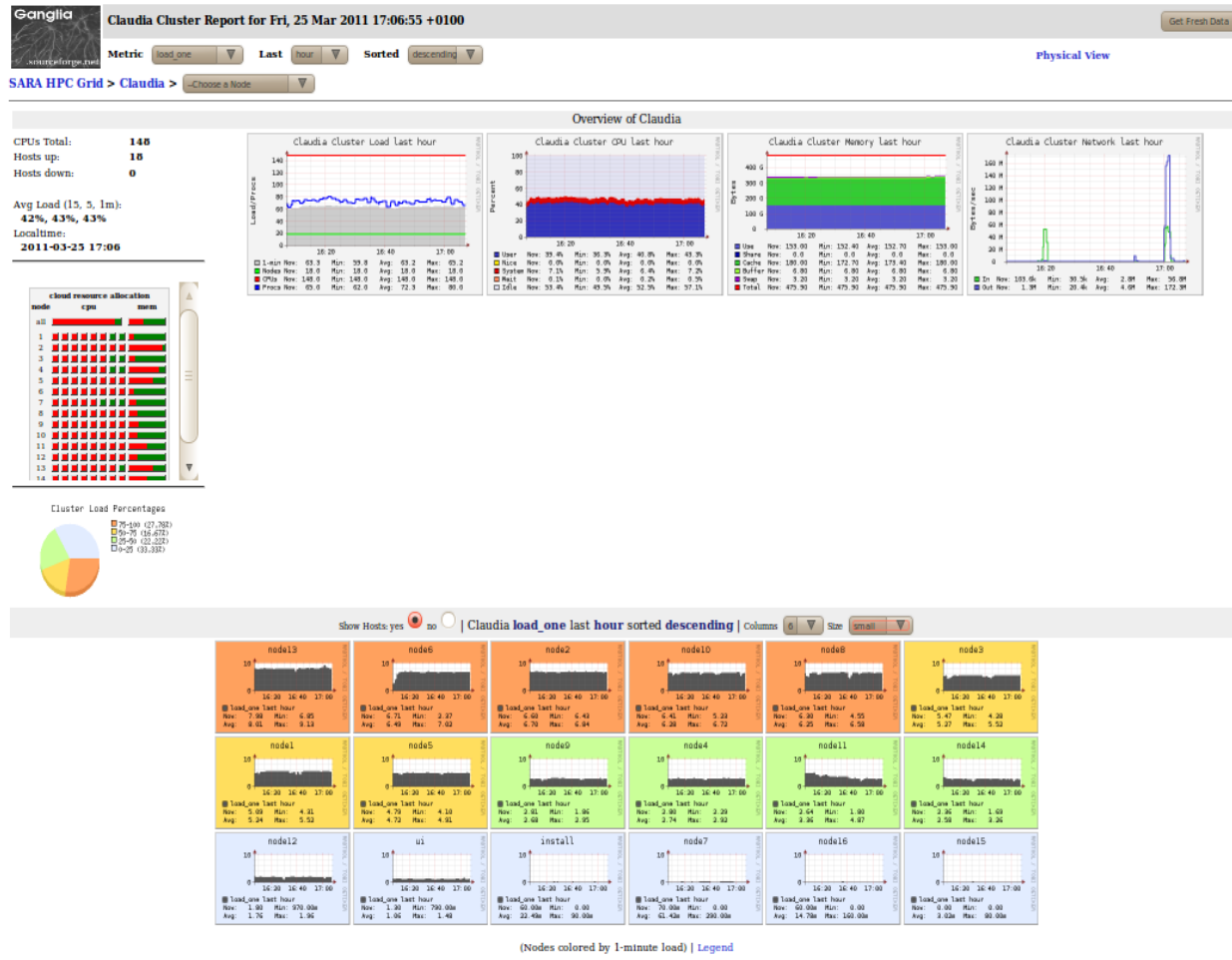
[View all news](#)

Self Service GUI

Developed at SARA
Open Source, available at
www.opennebula.org

BiG Grid
the dutch e-science grid

Monitoring workload





Advantages of HPC Cloud

- Only small overhead from virtualization (5%)
- easy/no porting of applications
- Applications with different requirements can co-exist on the same physical host
- Long running services (for example databases)
- Tailored Computing
- Service Cost shifts from manpower to infrastructure
- Usage cost in HPC stays Pay per Use
- Time to solution shortens for many users



BiG Grid

the dutch e.science grid

Observations

- Usage: Scientific programmer prepares environment, Scientist uses
- Several “heterogenic clusters” Microsoft Instances combined with Linux
- Modest parallelism (maximum 64)
- User wishlist: Possibility to share a collection of custom made virtual machines with other users
- Added value: support by your trusted HPC centre.
- HPC Cloud on HPC hardware is necessary addition to a complete HPC eco-system
- Interactive support works (some users do read tickets and documentation)

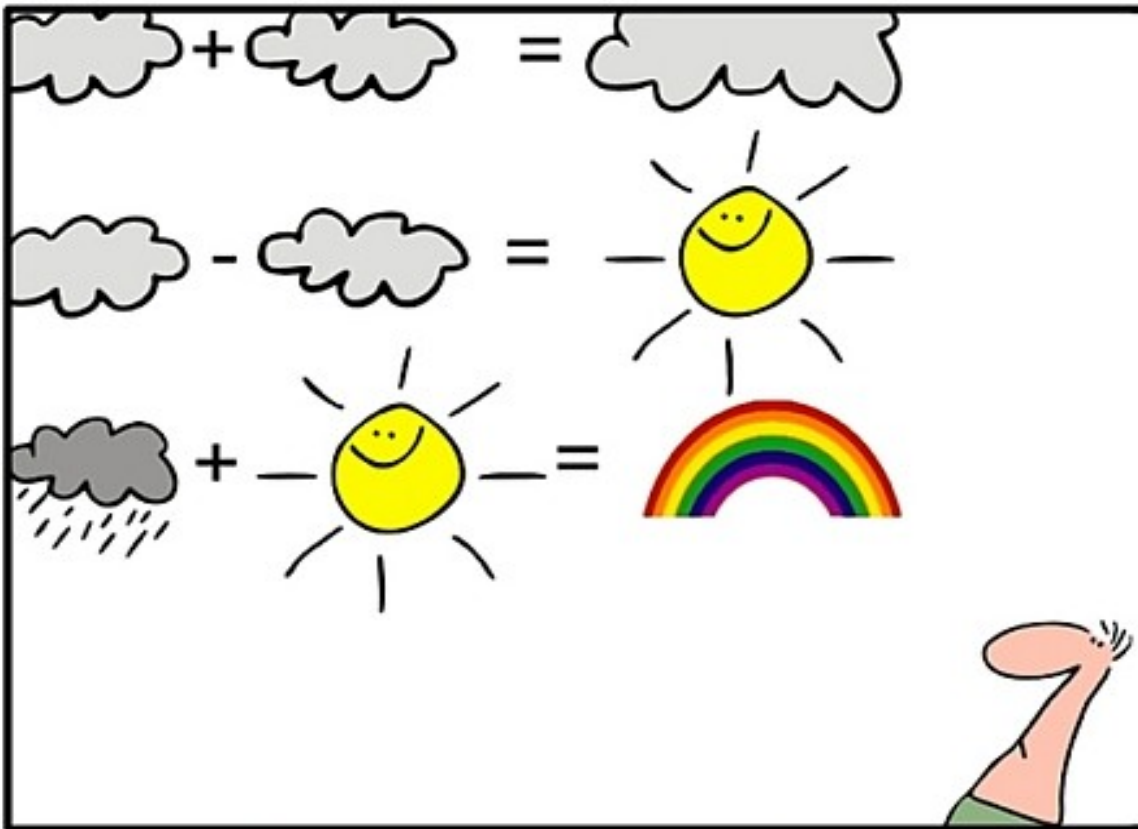


BiG Grid

the dutch e-science grid

Thank you!

Questions?



geek and poke



www.cloud.sara.nl

***SIMPLY EXPLAINED - PART 17:
CLOUD COMPUTING***

BiG Grid

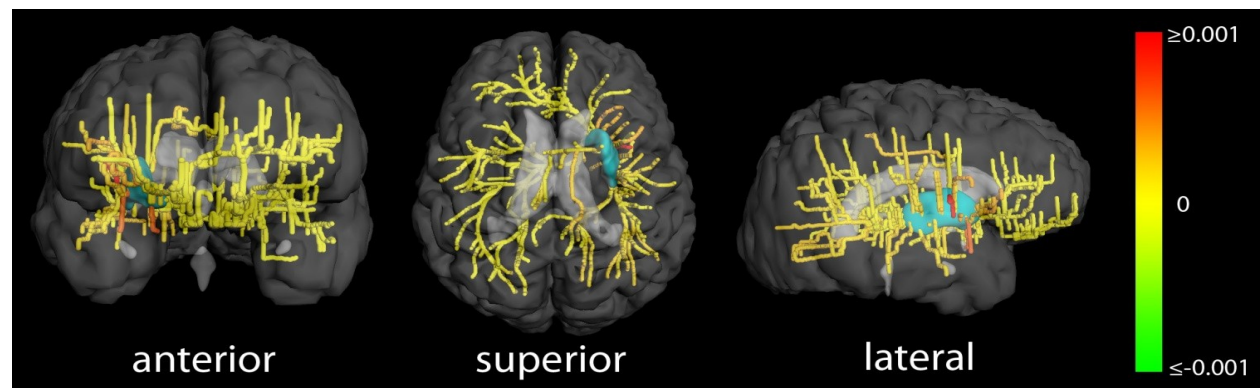
the dutch e-science grid

photo: <http://cloudappreciationsociety.org/>



Example Project 1

- Medical data MRI Image processing pipeline
 - Cluster with custom imaging software
 - Dynamic scaling up depending on the load
 - Added 1 VM with web service for user access, data upload and download



BiG Grid

the dutch e-science grid

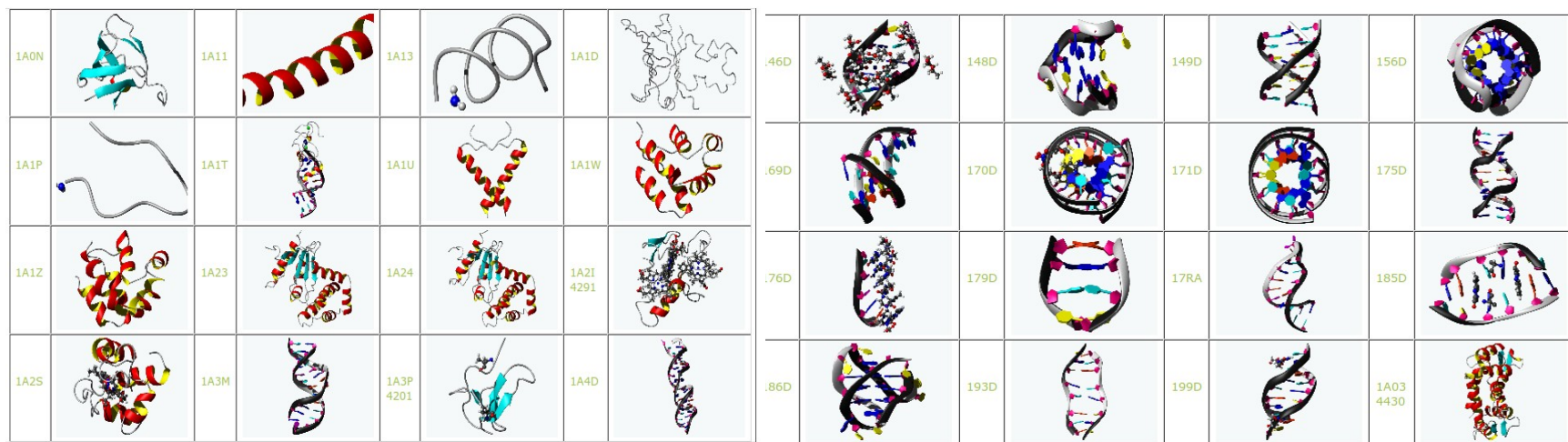
Pictures from H. Vroegman, Erasmus MG

Example project 2

NMR spectroscopy: Virtual Cing by J. Doreleijers

With NMR spectroscopy the 3D structure of biomolecules such as proteins and DNA are solved in solution. It thus provides a structural view of the chemical reactions that underly most diseases.

NMR structure determination needs a solid validation of the experimental data in relation to the resulting 3D coordinates because the process in many labs has not and often -can- not be automated fully. A virtual machine called VirtualCing (VC for short) interfaces to the best 24 NMR validation programs, together with CING's internal unique checks. VC was developed because installing the external programs on a traditional grid would take too long in development and would be cumbersome to maintain. We were able to validate all the 8,000+ structures currently available in the worldwide database Protein Data Bank (wwPDB) in just a week. The same strategy is applied to recalculate, improve and validate several thousand protein structures in a new project named NMR_REDO.



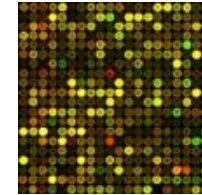
BiG Grid

the dutch e-science grid

User Experience

(slides from Han Rauwerda, transcriptomics UVA)

Microarray analysis: *Calculation of F-values in a 36 * 135 k transcriptomics study using of 5000 permutations on 16 cores. Over 10 week period **30.000** core-hours*
Data analysis using R (statistical analysis) with specialized plugin



Ageing study - conditional correlation

dr. Martijs Jonker (MAD/IBU), prof. van Steeg (RIVM), prof. dr. v.d. Horst en prof.dr. Hoeymakers (EMC)

- 6 timepoints, 4 tissues, 3 replicates and 35 k measurements + pathological data
- Question: find per-gene correlation with pathological data (staining)
- Spearman Correlation conditional on chronological age (not normal)
- p-values through 10k permutations (**4000 core hours** / tissue)

Co-expression network analysis

- 6k * 6k correlation matrix (conditional on chronological age)
- calculation of this matrix parallellized. (**5.000 core hours** / tissue)

Development during testing period (real life!)

Conclusions

- Many ideas were tried (clusters with 32 - 64 cores)
- worked out of the box (including the standard cluster logic)
- no indication of large overhead
- Cloud cluster: like a real cluster
- Virtually no hick-ups of the system, no waiting times
- User: it is a very convenient system

BiG Grid

the dutch e-science grid

