HPC Cloud

Floris Sluiter
SARA computing & networking services

BiG Grid
the dutch e-science grid

About SARA, NCF and BiG Grid

- The foundation for **National Compute Facilities** is part of NWO, the Dutch Government Organization for Scientific Research
- The **BiG Grid** project is a collaboration between NCF, Nikhef and NBIC, and enables access to grid infrastructures for scientific research in the Netherlands.
- **SARA** is a national High Performance Computing and e-Science Support Center, in Amsterdam and the primary operational partner of BiG Grid

BiG Grid
the dutch e-science grid

SARA Project involvements

SARA
Scientific Infrastructure and support

High Performance Computing	Huygens, GPU cluster Lisa, Grid. Hadoop, HPC Cloud	
High Resolution Visualization	Tiled Panel Display Remote Visualization	
High Performance Networking	SURFnet 6 AMSix Netherlight	
Mass Storage	2*10 Petabyte Tape archive 4 Petabyte disk storage	

BiG Grid
the dutch e-science grid

Scientific Computing facilities SARA (Specs)

Huygens National Super

Power6, 3328 cores in 105 nodes
15.25 TB of memory,
Infiniband 160 Gbit/s
700 TB of disk space,
60 TFlop/s

NCF

LISA National Compute Cluster

Intel, 4480 cores in 512 nodes,
12 TB of memory,
Infiniband 20Gbit/s
50Tbyte disk space
20 TFlop/s

NCF

Grid Resources

Intel, 3000 Cores in 3000 nodes
9 TB memory
125 Mbit/s (1 Gbit/s burst) Ethernet
4.5 PB of disk space, 4 PB tape
60K specints

BiG Grid

GPU Cluster (part of LISA)

Tesla GPU 2000 cores in 8 nodes
32 Gbyte memory (total for GPU)
Infinband 20Gbit/s
2 Tbyte disk space
7 Tflop/s

NCF

HPC Cloud

?

BiG Grid

Innovative Infrastructure

Hadoop
CDMI
Webdav, iRODS
ClearSpeed

hadoop

iRODS

NCF

BiG Grid

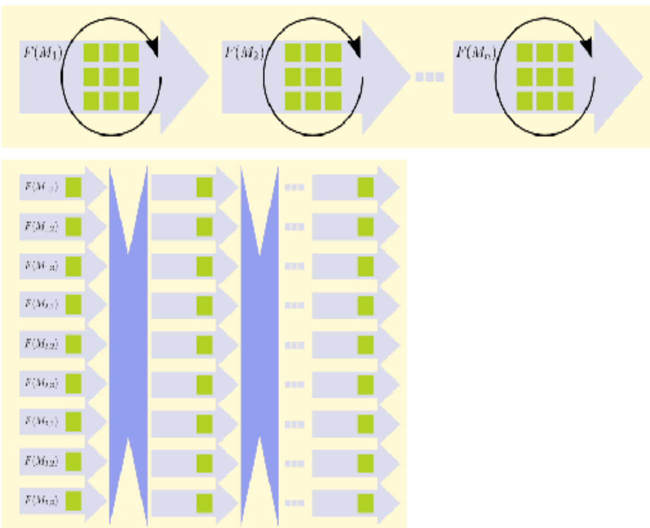
BiG Grid
the dutch e-science grid

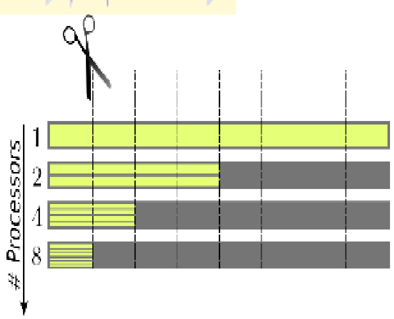
What is a HPC Cloud?

BiG Grid
the dutch e-science grid

High Performance Computing
Application Parallelization

Task & Data Parallelization

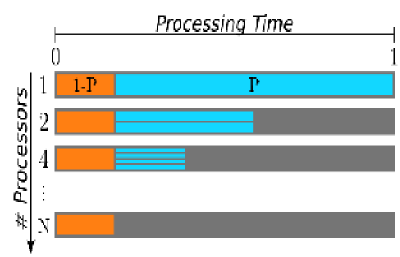


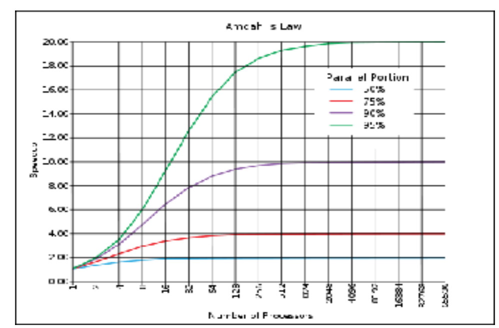


Amdahl's Law: The speedup of a program using multiple processors in parallel computing is limited by the sequential fraction of the program.

For example, if 95% of the program can be parallelized, the theoretical maximum speedup using parallel computing would be 20× as shown in the diagram, no matter how many processors are used.

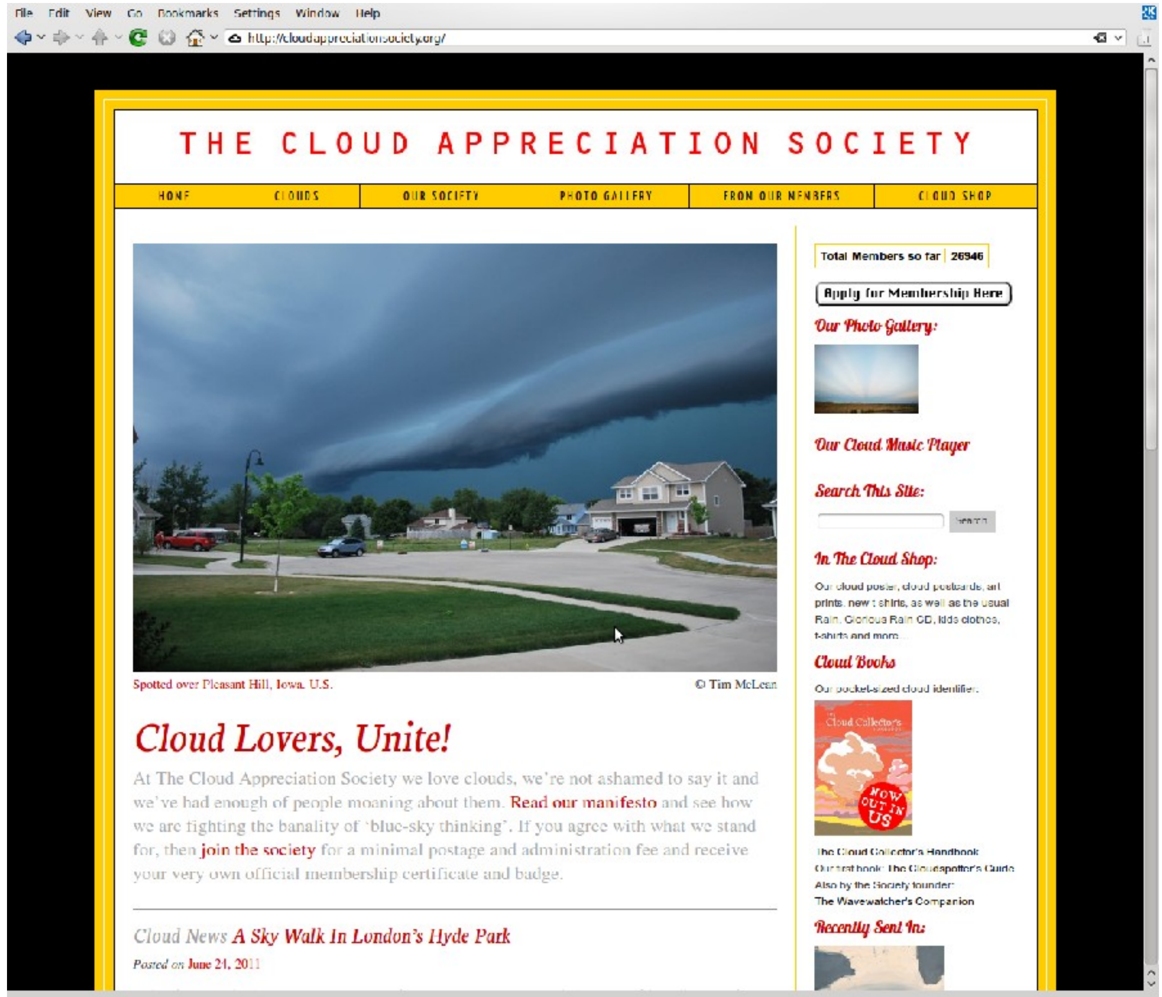
$$\frac{1}{(1-P)+\frac{P}{N}}$$





BiG Grid
the dutch e-science grid

What is a Cloud?



BiG Grid
the dutch e-science grid

What is Cloud Computing?

"Cloud computing is a model for enabling convenient, ondemand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction." (NIST-SP800-146)

Service Models:

- Cloud Software as a Service (SaaS)
- Cloud Platform as a Service (PaaS)
- Cloud Infrastructure as a Service (IaaS)

BiGGrid:

Allow users to

- freely instantiate a personal environment
- leap from laptop (small scale) to HPC (large scale)

BiG Grid
the dutch e-science grid

What is a Cloud?

[National Institute for Standards and Technology NIST]

[http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc]

- Resource Pooling,
  - Multiple concurrent users on a shared system
- Broad Network Access
  - Accesible from The Internet
- Measured Service
  - Pay per use
- Rapid Elasticity
  - Capabilities scaled up and down dynamically (pay-as-you-go)
- On-demand self-service
  - User is in full control

Is a Compute Centre a Cloud?

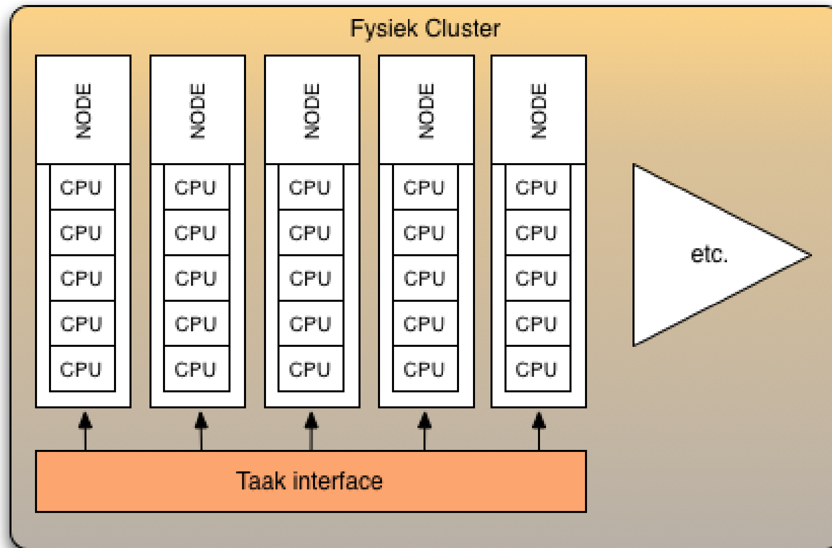
[National Institute for Standards and Technology NIST]
[http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc]

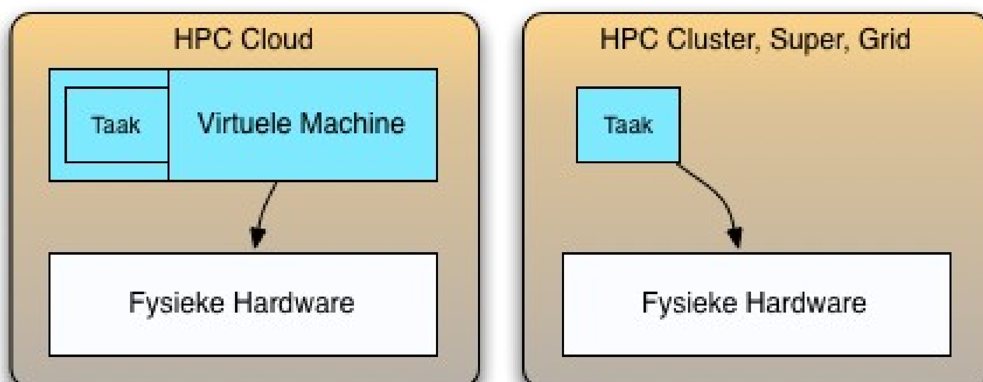
- Resource Pooling,
  - Multiple concurrent users on a shared system — **Yes**
- Broad Network Access
  - Accesible from The Internet — **Yes**
- Measured Service
  - Pay per use — **Yes**
- Rapid Elasticity
  - Capabilities scaled up and down dynamically (pay-as-you-go) Within pre-allocation — **Some**
- On-demand self-service
  - User is in full control
    No control over OS and adding resources not trivial — **NO**

BiG Grid
the dutch e-science grid

HPC Cloud vs HPC Cluster

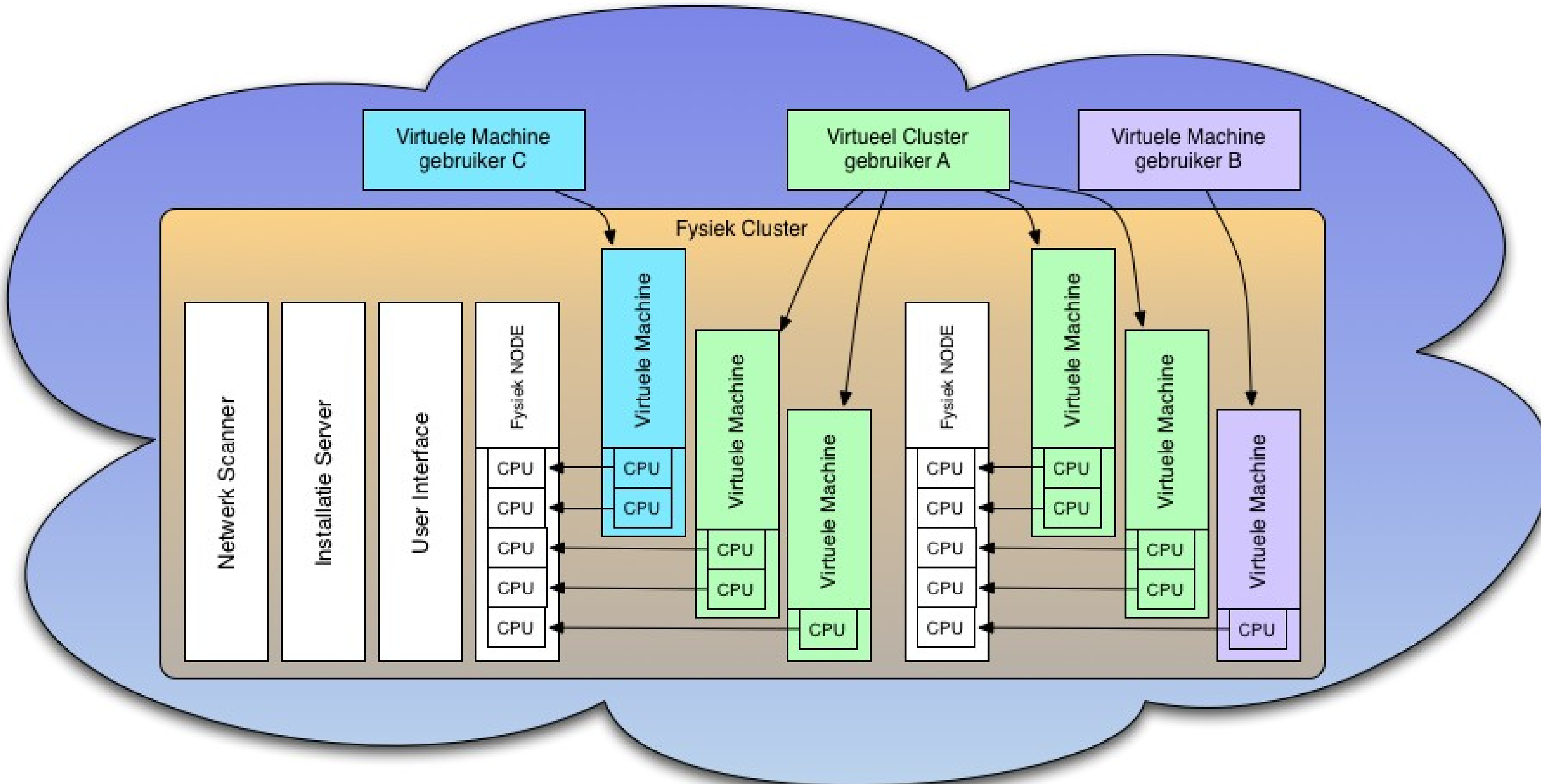
- A VM is a "job"





Virtual Machines in a HPC Cloud

- Multi-tenancy



BiG Grid
the dutch e-science grid

(HPC) Cloud Why?

World

- better utilization for infrastructure
- "Green IT" (power off under-utilization)
- easy management

BiGGrid

- free OS & software environment
- locked software can be used
- rapid availability
- HPC cloud for academic world

Massive interest and multiple early adopters prove the need for an academic HPC Cloud environment.

- beta-cloud running “production

BiG Grid
the dutch e-science grid

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

...At AMAZON?

- Cheap?
  - Quadruple Extra Large = 8cores and 64Gb ram: $2.00/h (or $5300/y + $0.68/h)
  - 1024 cores = $2.242.560/y (or $678k + $760k = $1.4M/y)

- Bandwidth = pay extra
- Storage = pay extra
- I/O guarantees?
- Support?
- Secure (no analysis/forensics)?
- **High Performance Computing??**

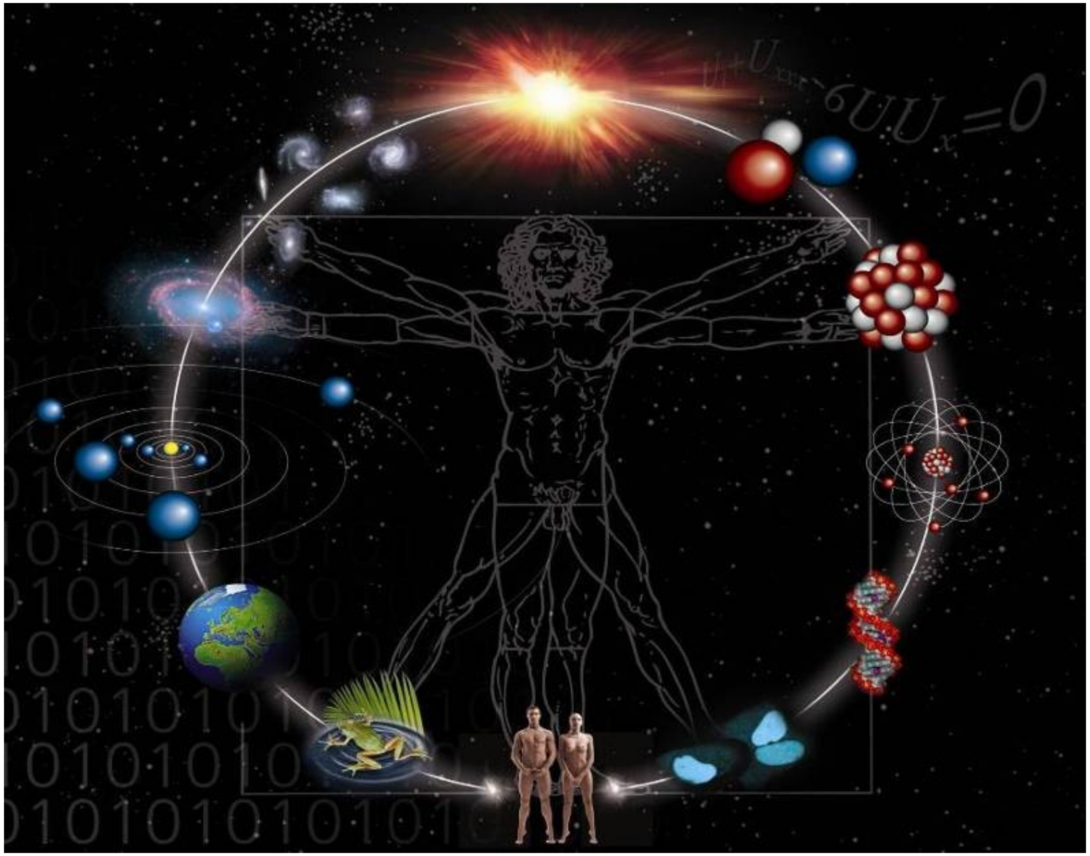
BiG Grid
the dutch e-science grid

What is needed to create a successful HPC Cloud?

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.

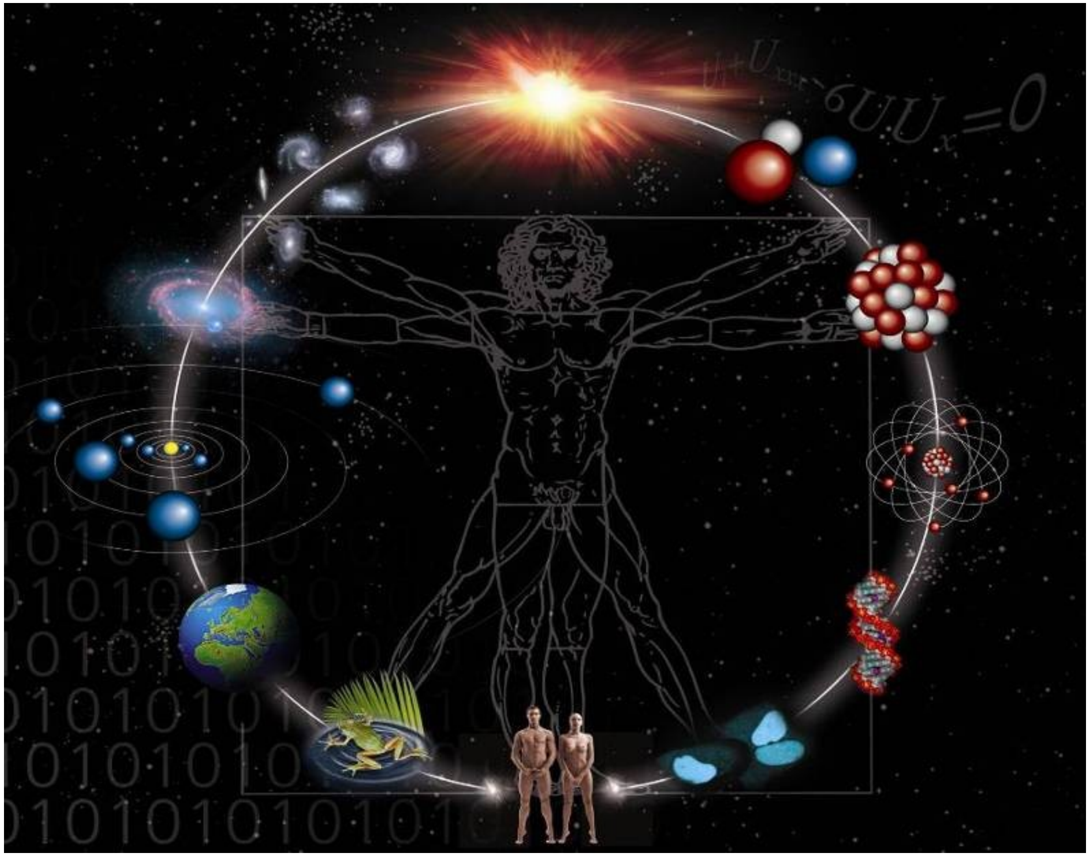


Slide courtesy of prof. F. Linde, Nikhef

(current) Users of HPC Cloud 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

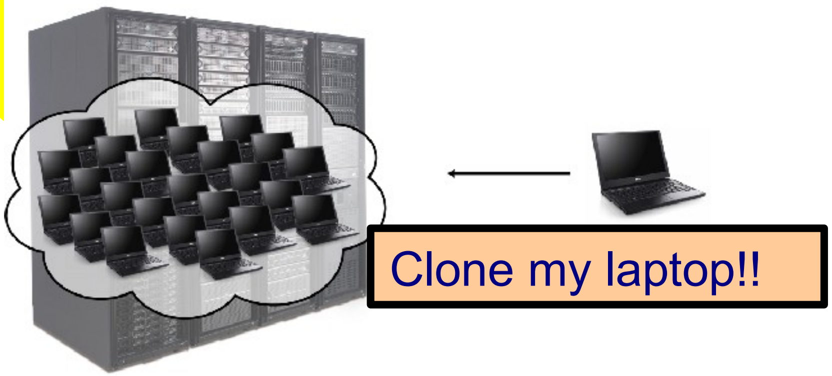
BiG Grid
the dutch e-science grid

HPC (Cloud) Application types

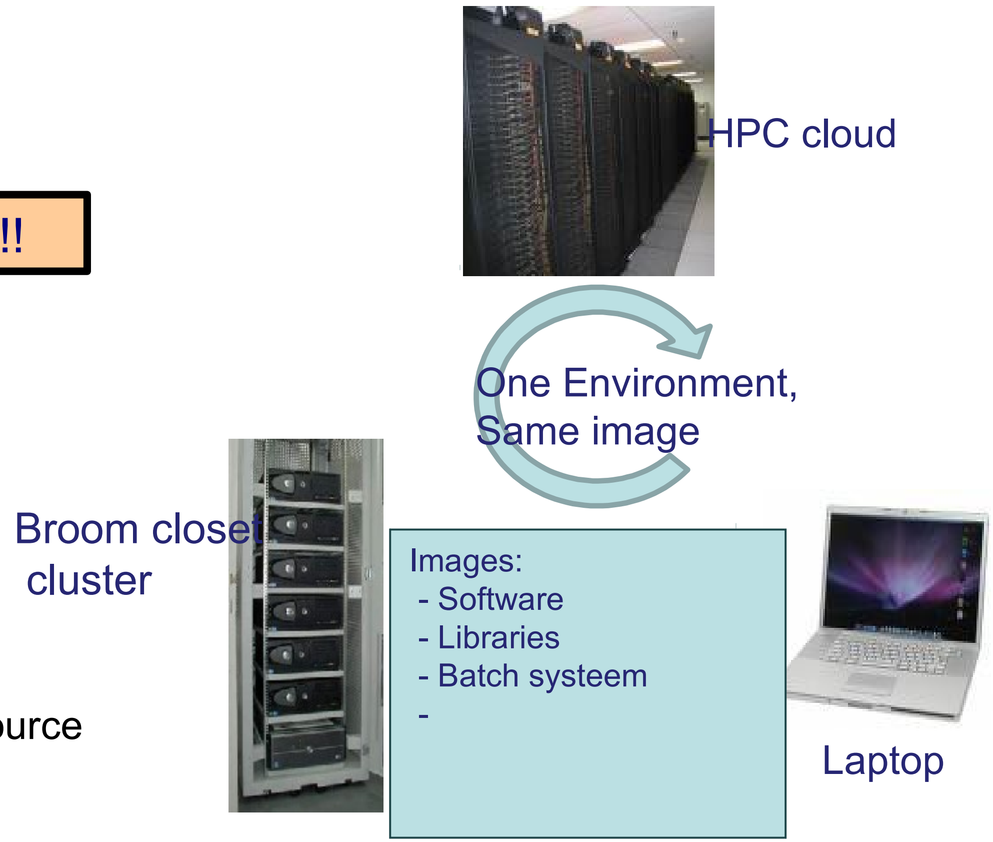
Type	Examples	Requirements
Compute Intensive	Monte Carlo simulations and parameter optimizations, etc	CPU Cycles
Data intensive	Signal/Image processing in Astronomy, Remote Sensing, Medical Imaging, DNA matching, Pattern matching, etc	I/O to data (SAN File Servers)
Communication intensive	Particle Physics, MPI, etc	Fast interconnect network
Memory intensive	DNA assembly, etc	Large (Shared) RAM
Continuous services	Databases, webservers, webservices	Dynamically scalable

BiG Grid
the dutch e-science grid

HPC Cloud: Concepts

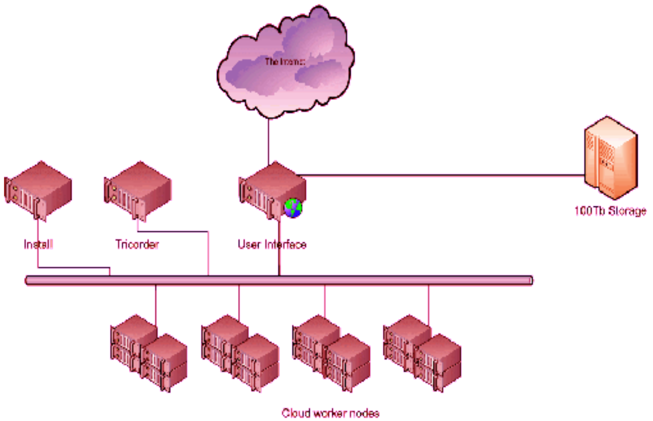


- HPC Hardware
- No overcommitting (reserved resources)
- Secured environment and network
- User is able to fully control their resource (VM start, stop, OS, applications, resource allocation)
- Develop together with users



The product: Virtual Private HPC Cluster

- We (plan to) offer:
  - Fully configurable **HPC** Cluster (a cluster from scratch)
  - Fast CPU
  - Large Memory (64GB/8 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)

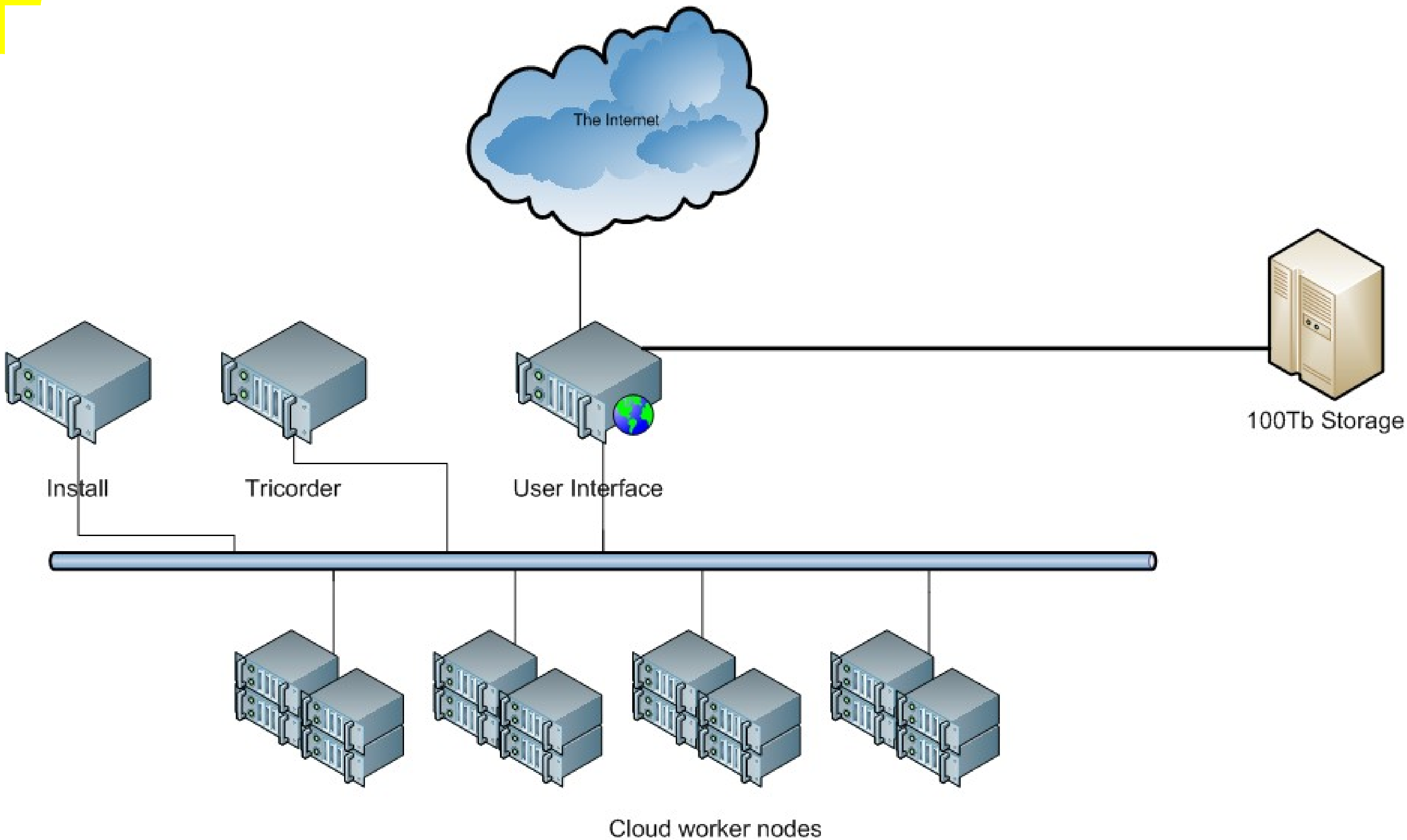


Platform and tools:

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

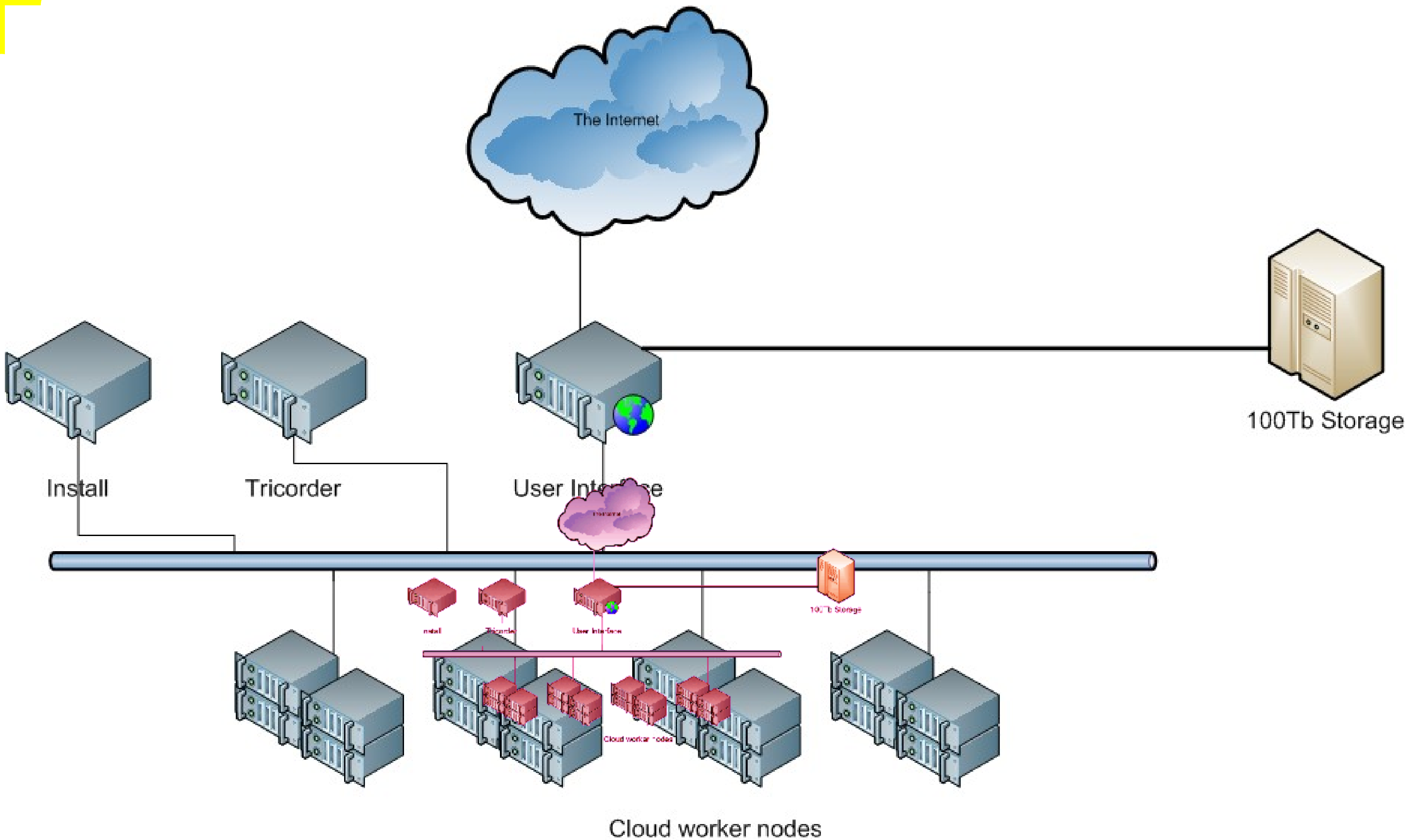
BiG Grid
the dutch e-science grid

Physical architecture (testbed)

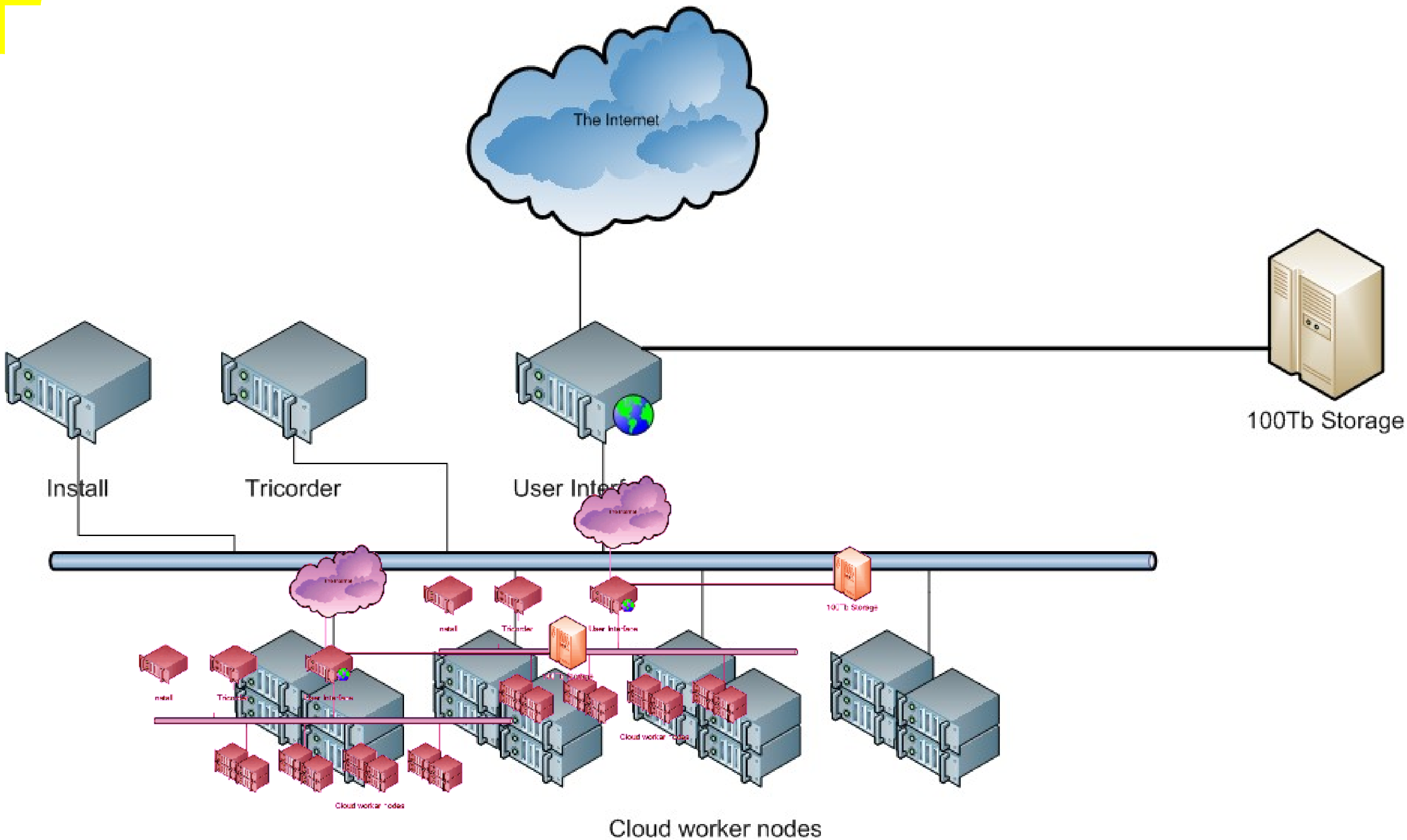


BiG Grid
the dutch e-science grid

Virtual architecture

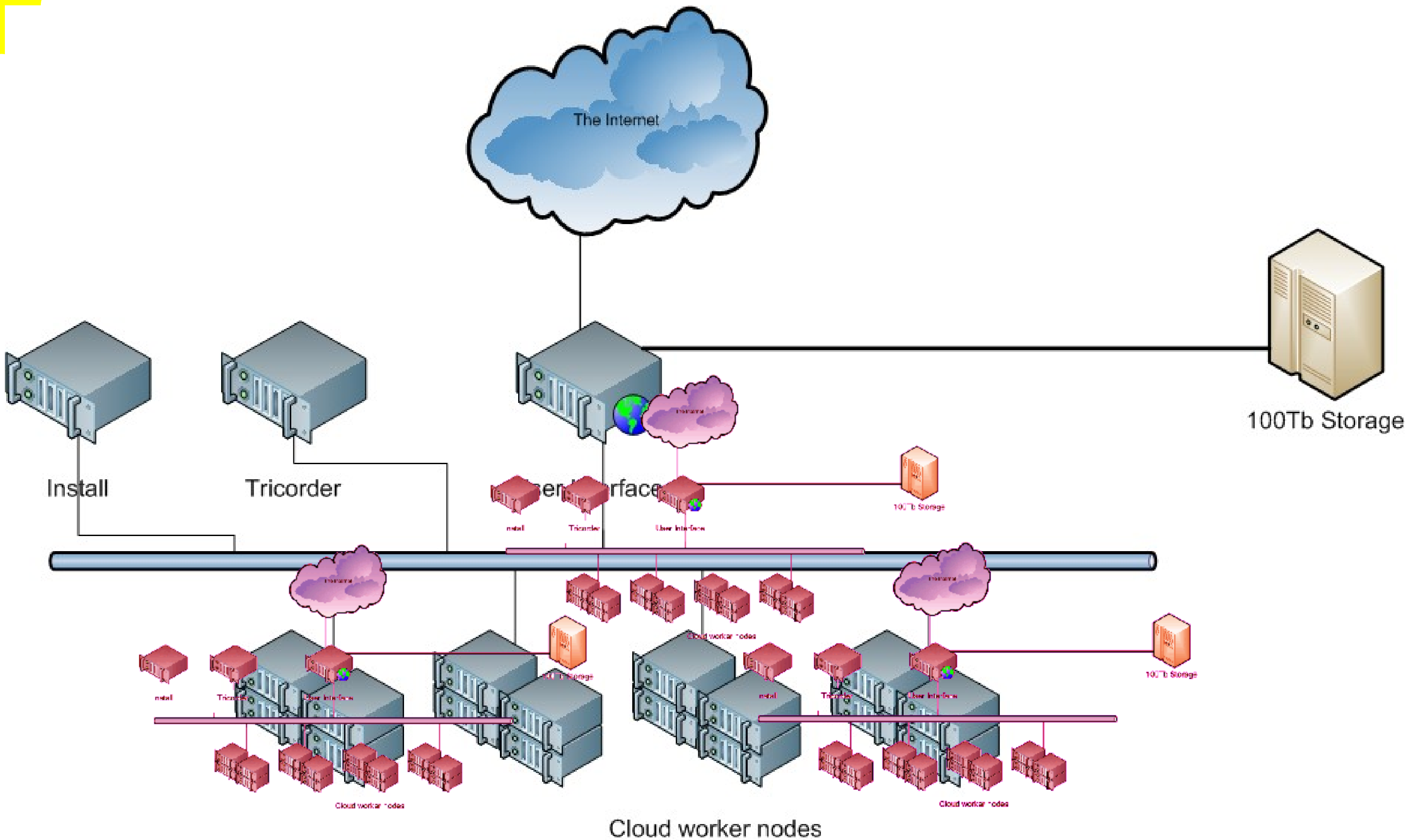


Virtual architecture cont…



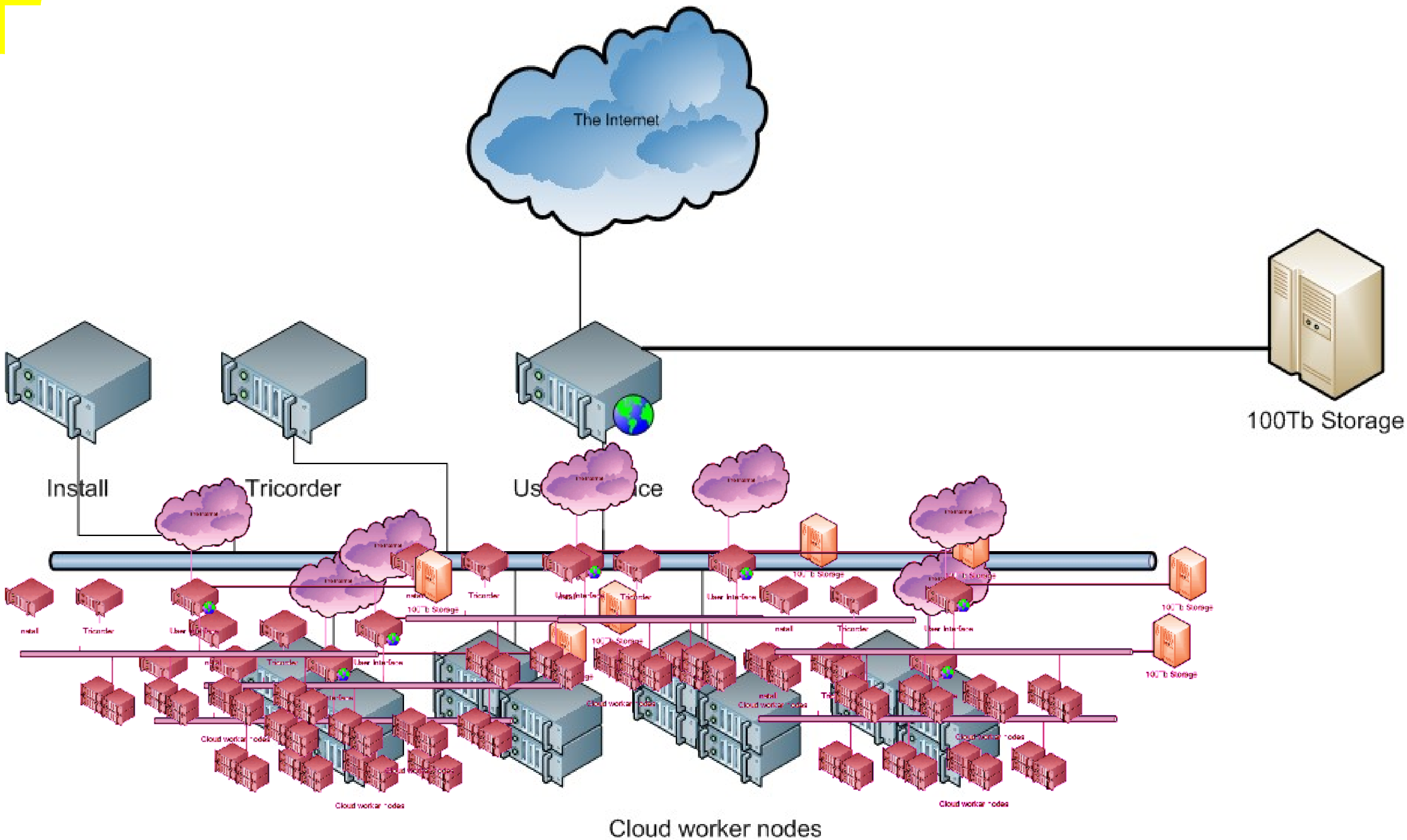
BiG Grid
the dutch e-science grid

Virtual architecture cont...



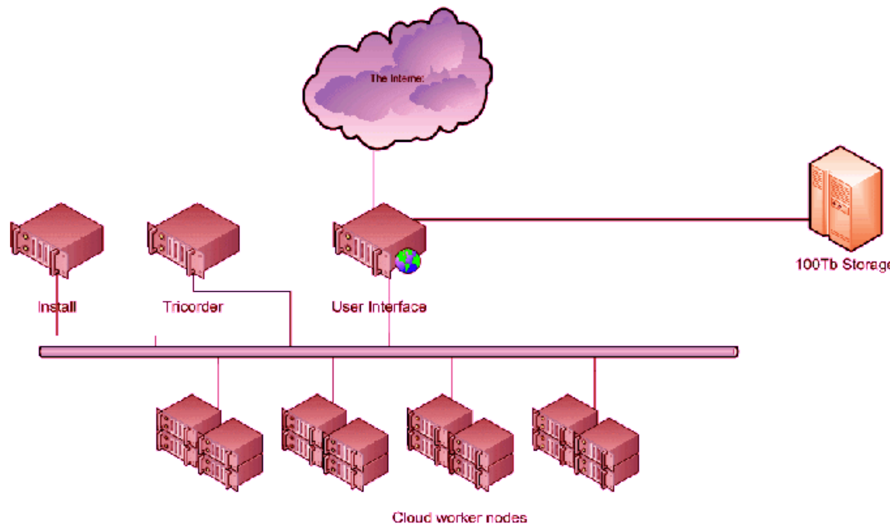
BiG Grid
the dutch e-science grid

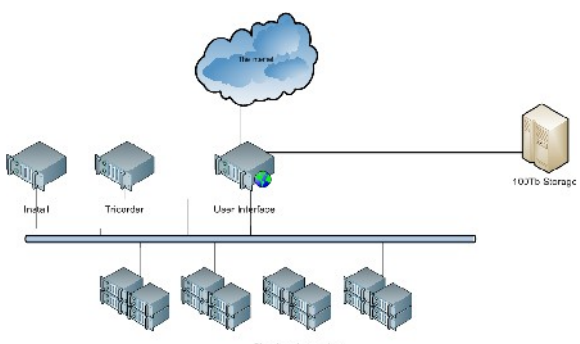
Virtual architecture cont...



BiG Grid
the dutch e-science grid

Virtual architecture
User view





BiG Grid
the dutch e-science grid

HPC Cloud trust (1/2)

Security is of major importance
- cloud user confidence
- infrastructure provider confidence

Protect
- the outside from the cloud users
- the cloud users from the outside
- the cloud users from each other

Not possible to protect the cloud user from himself
- user has full access/control/responsibility
  ex. virus research must be possible

HPC Cloud trust (2/2)

- Firewall
  - fine-grained access rules ("closed port" policy)
  - non-standard ports open on request only and between limited network ranges
- Scanning of new virtual templates
  - catches initial problems, but once the VM is live...
- Port scanning
  - catches well-known problems
- State-full Package Inspection
  - random sample based

BiG Grid
the dutch e-science grid

Project Development Goals

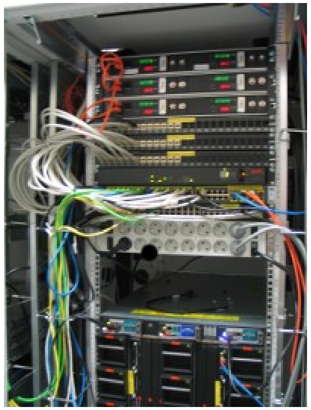
- Physical Architecture
  - HPC Cloud needs High I/O capabilities
  - Performance tuning: optimize hard- & software
  - Scheduling
- Usability
  - Interfaces
  - Templates
  - Documentation & Education
  - Involve users in pre-production (!)
- Security
  - Protect user against self, fellow users, the world and vice versa!
  - Enable user to share private data and templates
  - Self Service Interface
    - User specifies "normal network traffic", ACLs & Firewall rules
  - Monitoring, Monitoring, Monitoring!
    - No control over contents of VM
    - monitor its ports, network and communication patterns

ROADMAP

1) SARA Innovation project in 2009,
2) Pre-production for BiGGrid in 2010
3) In 2011 (summer) Production Infrastructure
4) Development continues 2011/2012

BiG Grid
the dutch e-science grid

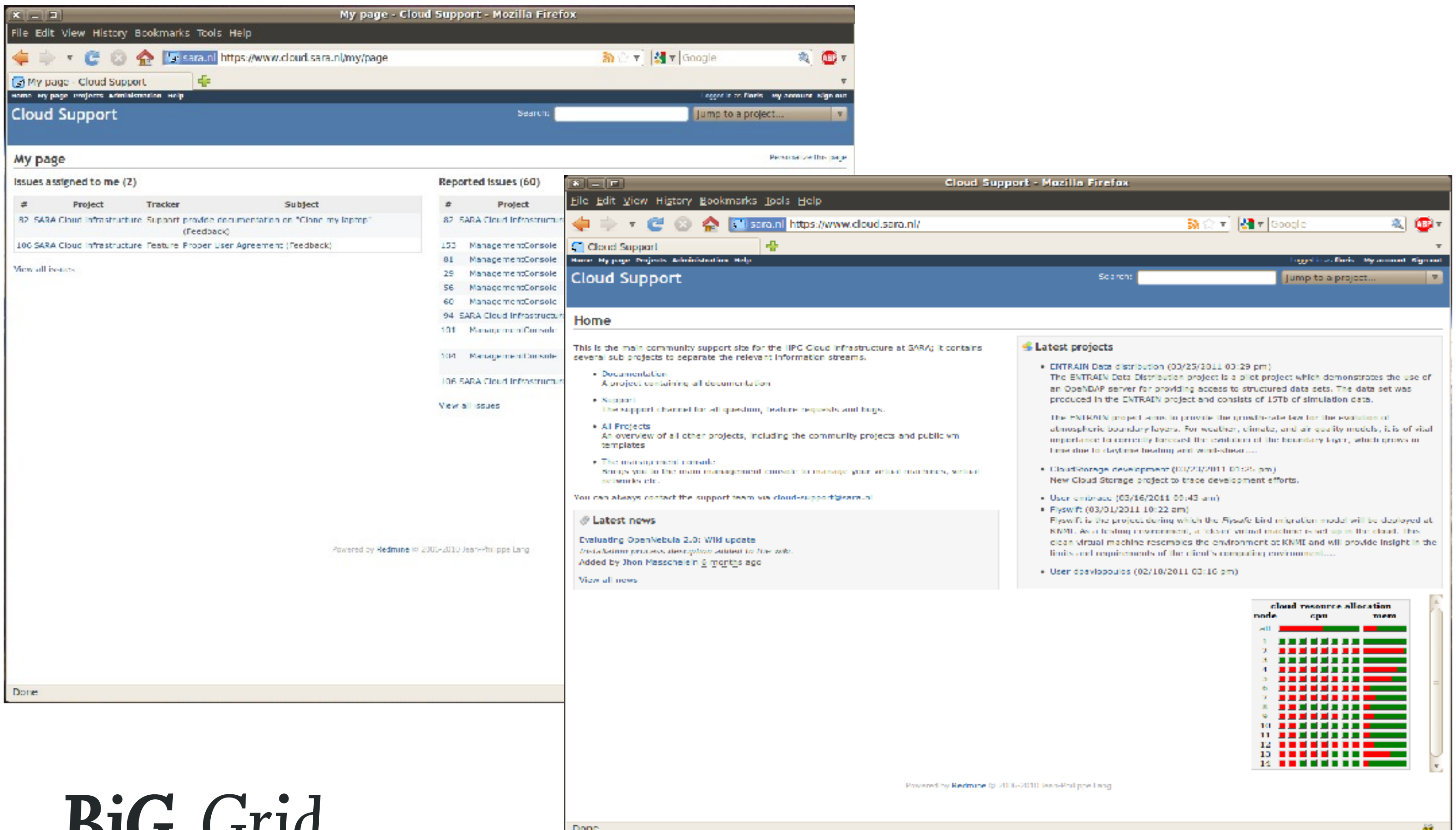
A Team and a bit of Hard Labour



BiG Grid
the dutch e-science grid

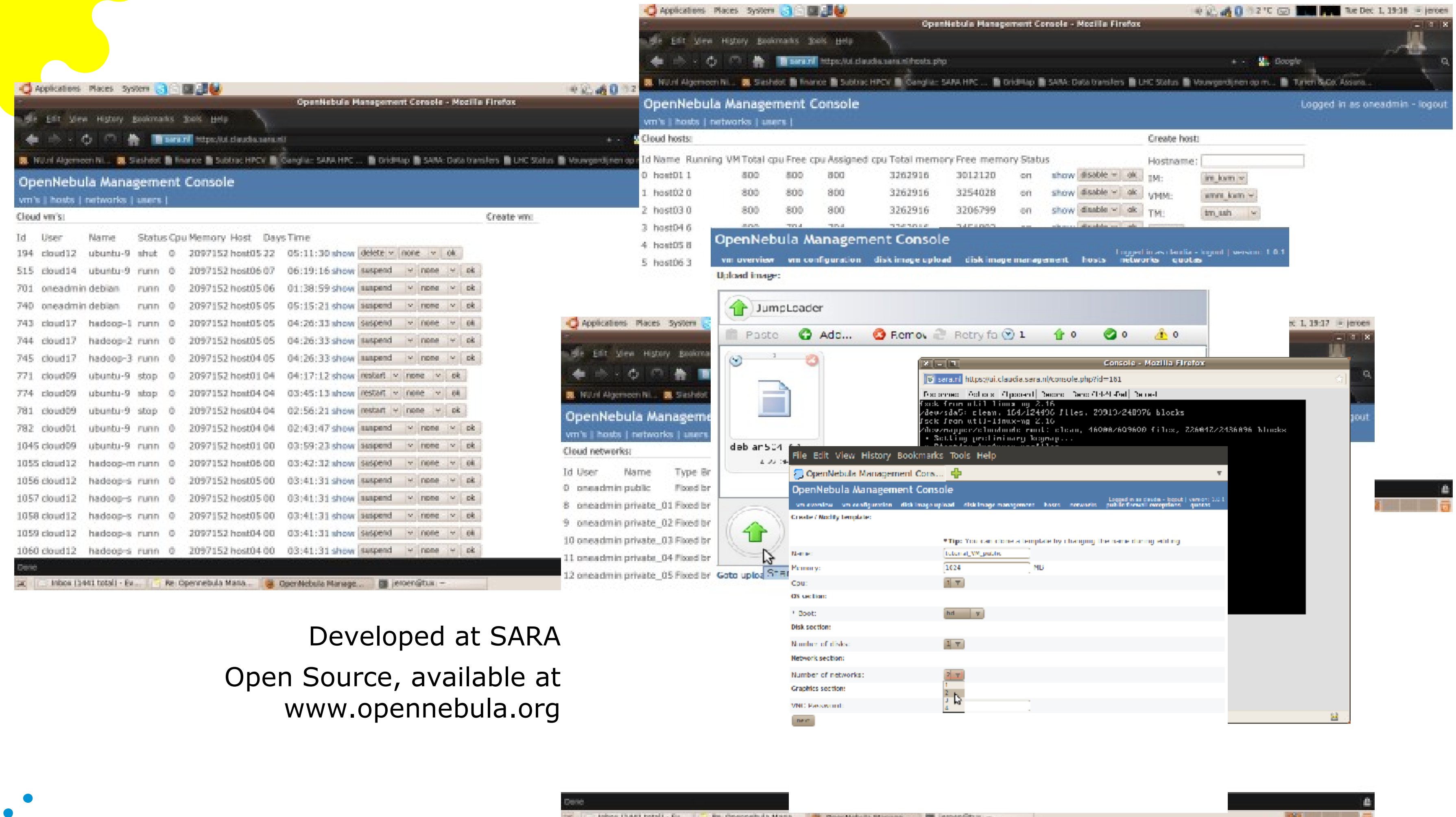
User collaboration Portal

- Redmine (www.redmine.org)



BiG Grid
the dutch e-science grid

Self Service GUI

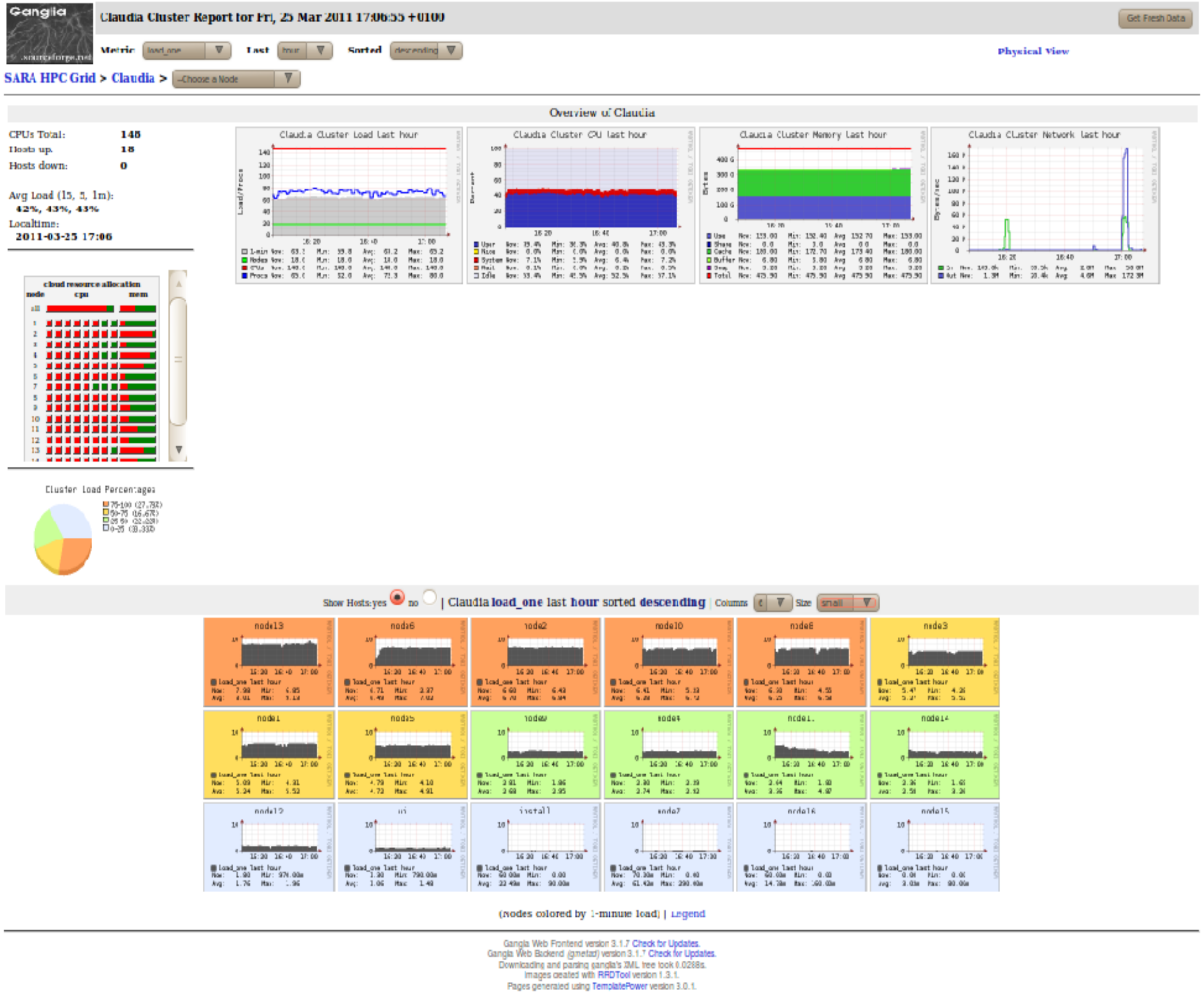


Developed at SARA

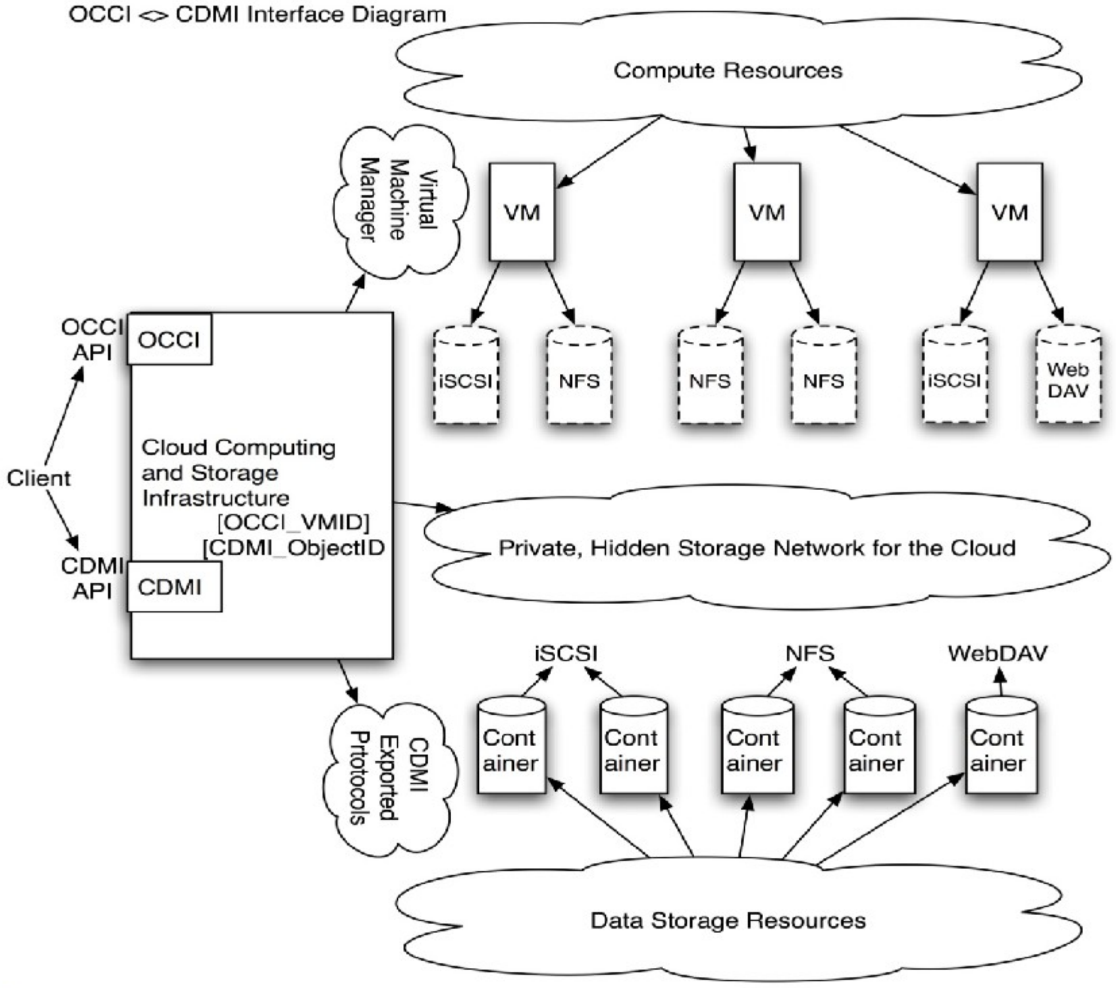
Open Source, available at www.opennebula.org

BiG Grid
the dutch e-science grid

Monitoring workload



Standards: OCCI + CDMI + OVF + CNMI = CMI



BiG Grid
the dutch e-science grid

Development plans/effort @ SARA

- Storage
  - CDMI server application
- Network
  - Dynamic provisioning
  - QoS
  - ACL/Firewall rules
  - Dynamic DNS
  - "CNMI"
  - Network benchmarking
- Compute
  - OCCI server with AAA?
- GUI
  - New & improved on OCCI/CDMI
- Security
  - Flow analysis
  - Dynamic ACL/Firewall

Real world network virtualization tests with qemu/KVM

- 20 gbit/s DDR infiniband (IPoIB) is compared with 1 Gbps Ethernet and 10 Gbps Ethernet
- Virtual network bridged to physical (needed for user separation)
- "real-world" tests performed on non optimized system
- Results
  - 1GE: 0,92 Gbps (1 Gbs)
  - IpoIB: 2,44 Gbps(20Gbs)
  - 10GE: 2,40 Gbps (10Gbs)
- Bottleneck: virtio driver
- Likely Solution: SRIOV

- Full report on www.cloud.sara.nl

User participation
30 involved in testing

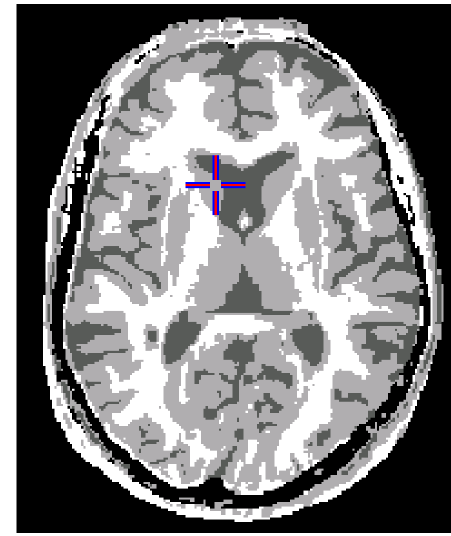
nr.	Title	Core Hours	Storage	Objective	Group/instiute
1	Cloud computing for sequence assembly	14 samples * 2 vms * 2-4 cores * 2 days = 5000	10-100GB / VM	Run a set of prepared vm's for different and specific sequence assembly tasks	Bacterial Genomics, CMBI Nijmegen
2	Cloud computing for a multi-method perspective study of construction of (cyber)space and place	2000 (+)	75-100GB	Analyse 20 million Flickr Geocoded data points	Uva, GPIO institute
3	Urban Flood Simulation				al Science
4	A user friendly cloud-based inverse modelling environment				o-ecology,
5	Real life HPC cloud computing experiences for MicroArray analyses				artment, rmatics Unit,
6	Customized pipelines for the processing of MRI brain data				ng Group, mus MC
7	Cloud computing for historical map collections: access and georeferencing				raphy, UvA
8	Parallellization of MT3DMS for modeling contaminant transport at large scale				
9	An imputation pipeline on Grid Gain				formatics f groningen
10	Regional Atmospheric Soaring Prediction				o-ecology,
11	Extraction of Social Signals from video	160	630GB	Video Feature extraction	Laboratory, TU Delft
12	sequencing data from mouse tumors	?	150-300GB	Run analysis pipeline to create mouse model for genome analysis	Chris Klijn, NKI

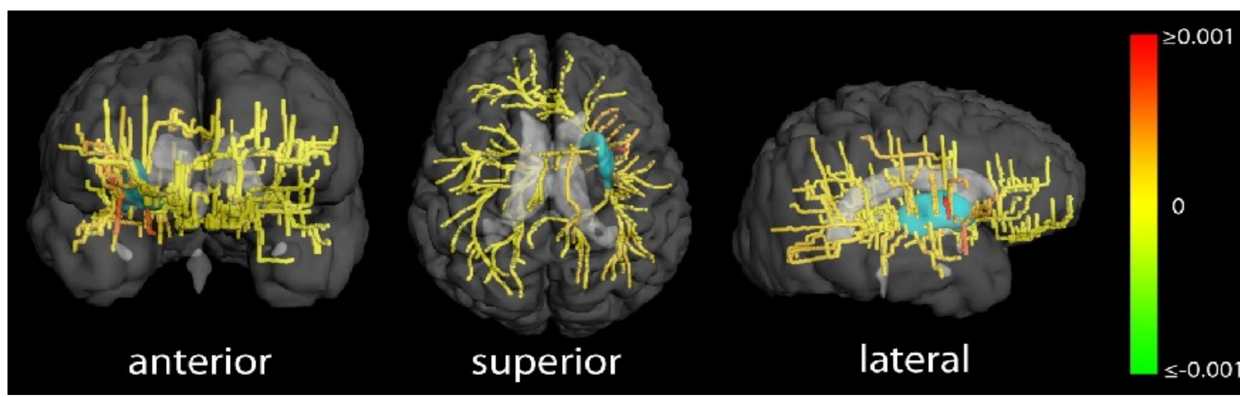
Field	# projects
Bioinformatics	8
Ecology	3
Geography	3
Computer science	5
Linguistics	4
Other	7

BiG Grid
the dutch e-science grid

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





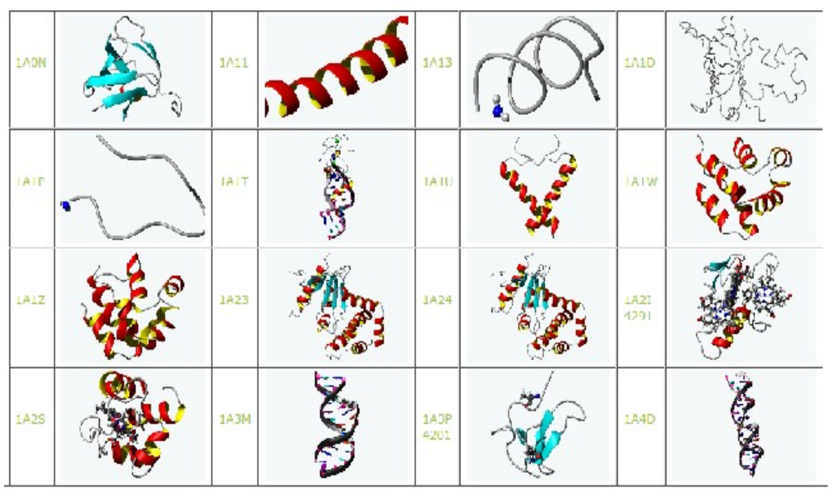
Pictures from H. Vrooman, Erasmus MC

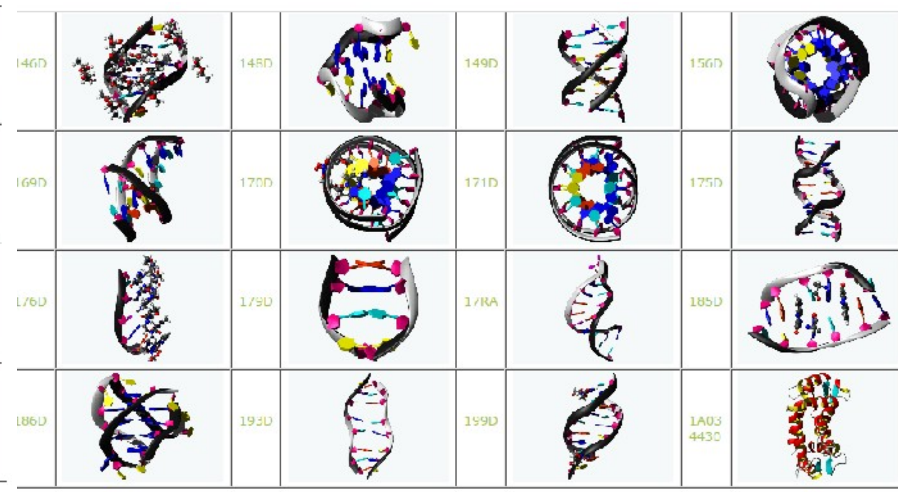
BiG Grid
the dutch e-science grid

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.

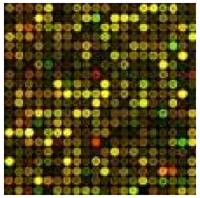




BiG Grid
the dutch e-science grid

User Experience

(slide**s** 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

Usage statistics in beta phase

Users liked it:

- ~90.000 core-hours used in first 10 weeks (~175.000 available)
- currently 500k core-hours
- 50% occupation *during beta testing*
- *Currently 80-90%*
- Science is being done!
- Some pioneers paved the way for the rest ("Google" launch approach)

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

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

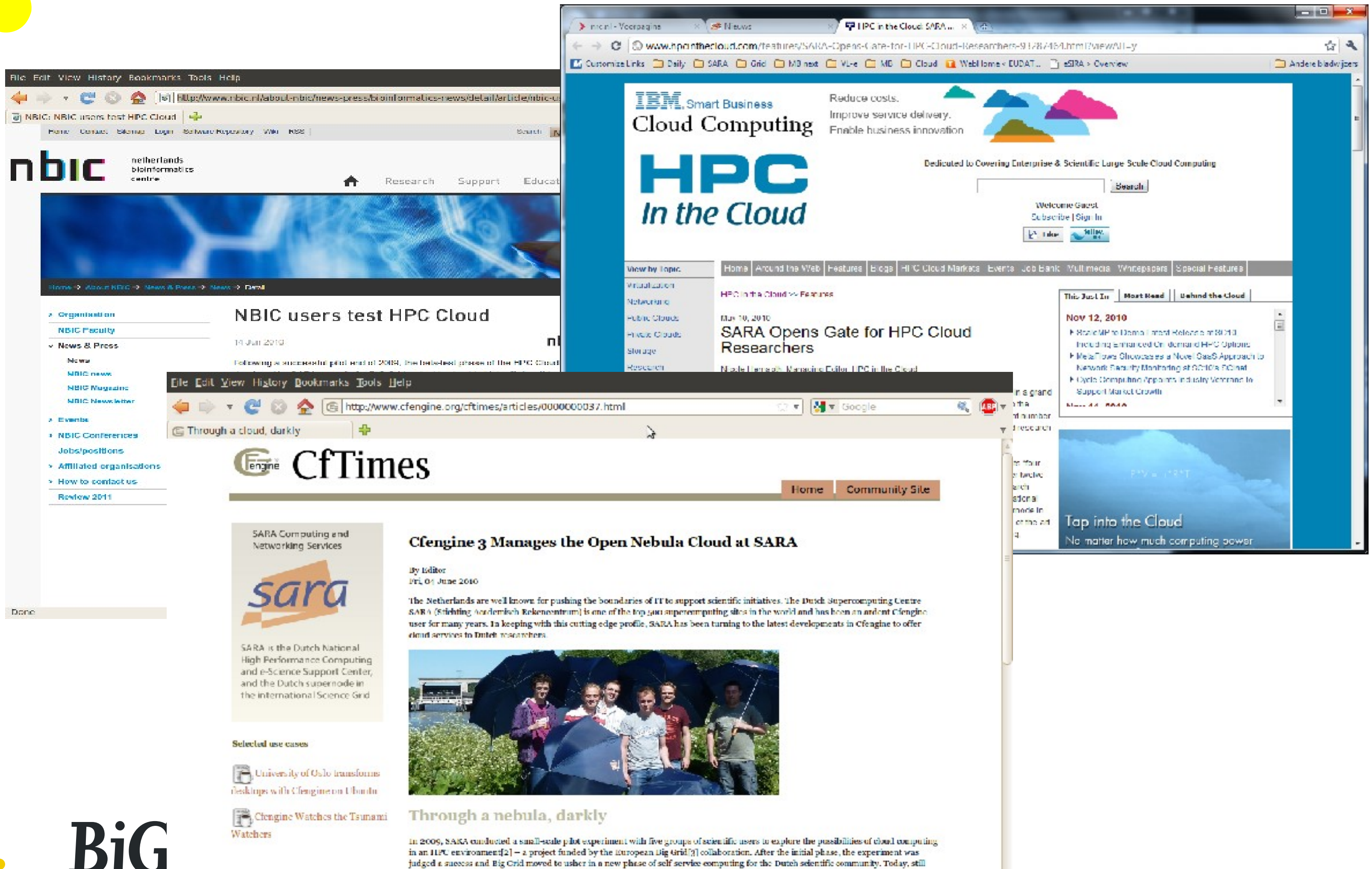
HPC Cloud, what is it good for?

- 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 HPC cloud in international media



BiG
the dutch e-science grid

SNEAK PREVIEW
(What is an ideal system for an HPC Cloud)

Calligo

"I make clouds"

19 Nodes:

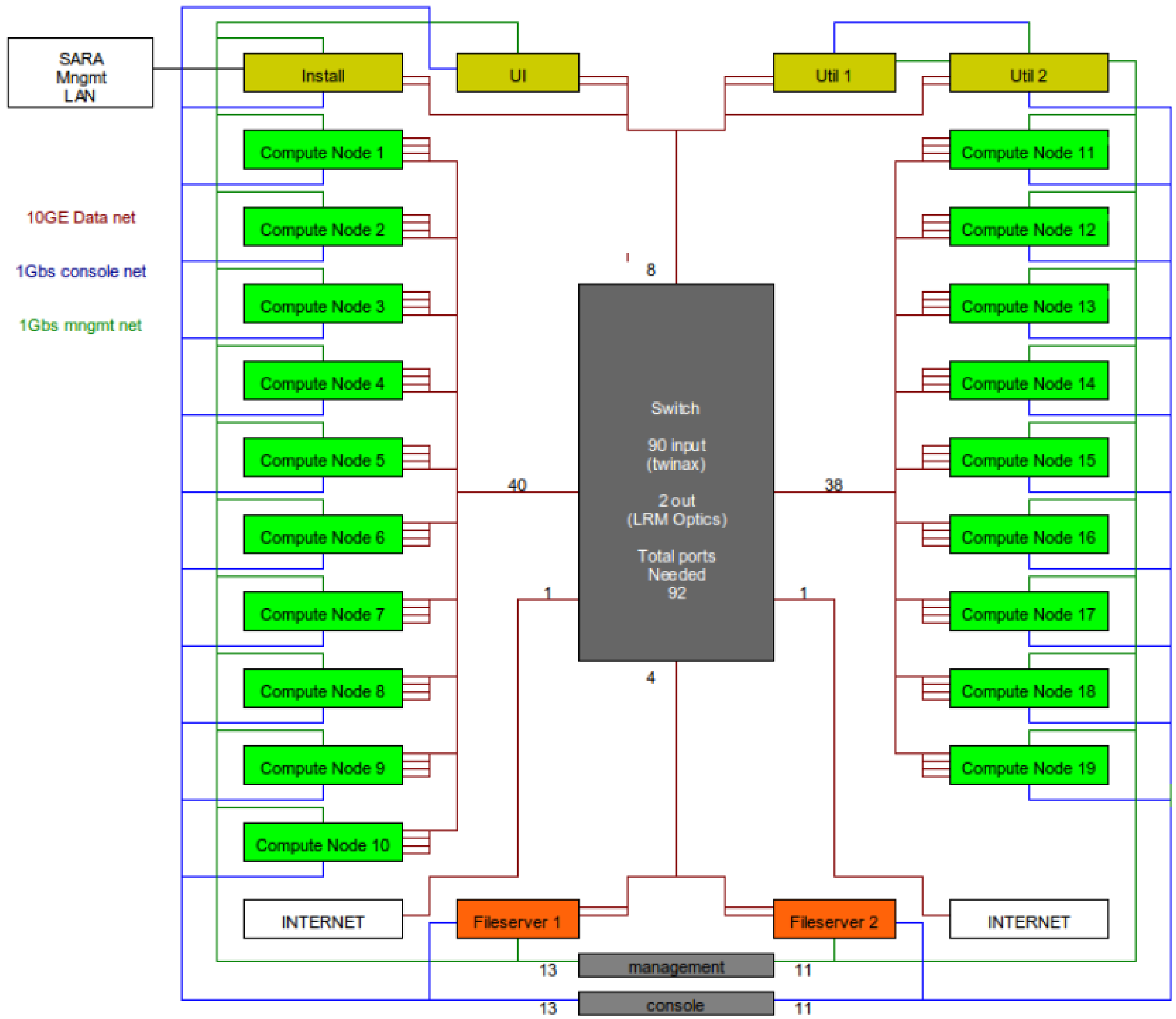
- CPU Intel 2.13 GHz 32 cores (Xeon-E7 "Westmere-EX")
- RAM 256 Gbyte
- "Local disk" 10 Tbyte
- Ethernet 4*10GE

Total System

- 608 cores
- RAM 4,75TB
- 96 ports 10GE, 1-hop, non-blocking interconnect
- 400TB shared storage (ISCSI,NFS,CIFS,CDMI...)
- 11.5K specints / 5TFlops

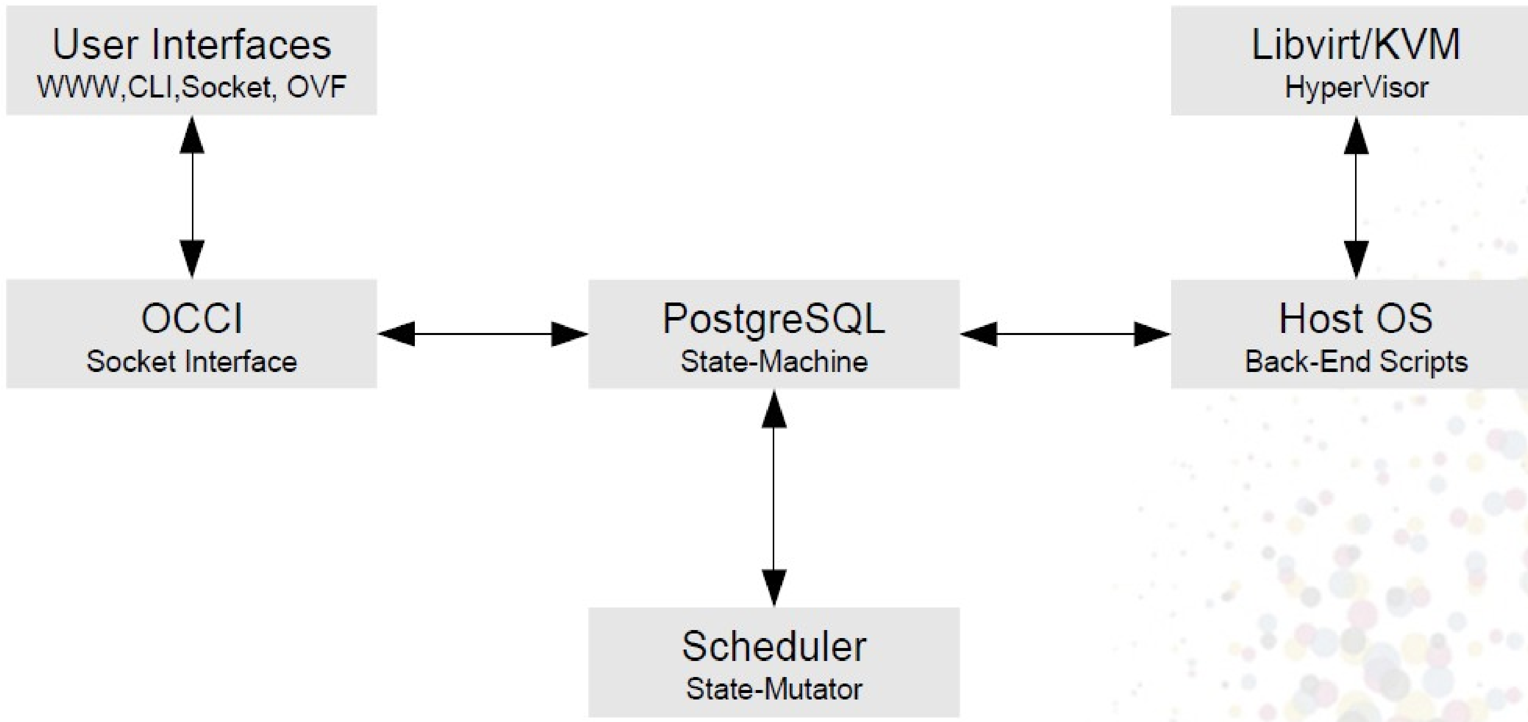
BiG Grid
the dutch e-science grid

Calligo, system architecture



BiG Grid
the dutch e-science grid

Calligo software architecture

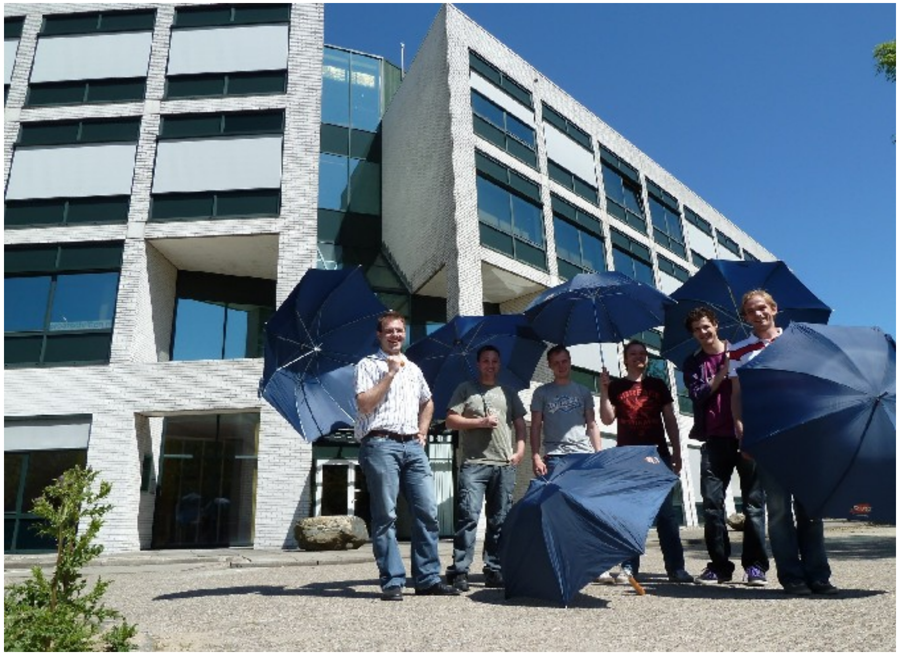


HPC Systems @ SARA

System	Node	Total
Huygens National Super	CPU Power6, 4.7Ghz, 32/64 cores RAM 128/256 GB "local disk" 8Tbyte Infiniband 8*20 Gbit/s	3456 cores in 105 nodes 15.25 TB of memory, 700 TB of disk space, 60 TFlop/s
LISA National Compute Cluster	CPU Intel 2.26Ghz 8/12 cores RAM 24 Gbyte Local disk 65Gbyte/200Gbyte Infiniband 20Gbit/s	4480 cores in 512 nodes, 12 TB of memory, Infiniband 20Gbit/s 50Tbyte disk space 20 TFlop/s
Grid	CPU Intel 2.2 Ghz 8 cores RAM 24 Gbyte Local disk 300Gbyte Ethernet 1 Gbit/s	3000 Cores in 3000 nodes 9 TB memory 125 Mbit/s (1 Gbit/s burst) Ethernet 4.5 PB of disk space, 4 PB tape 60K specints
Cloud	CPU Intel 2.13 GHz 32 cores (Xeon-E7 "Westmere-EX") RAM 256 Gbyte "Local disk" 10 Tbyte Ethernet 4*10GE	608 cores RAM 4,75TB 400TB shared storage (ISCSI,NFS,CIFS,CDMI...) 11.5K specints / 5TFlops 96 ports 10GE, 1-hop, non-blocking interconnect

BiG Grid
the dutch e-science grid

Acknowledgements



Our Sponsor: NL-BiGGrid

Our Brave & Entrepeneurial **Beta Users**

And the HPC Cloud team:

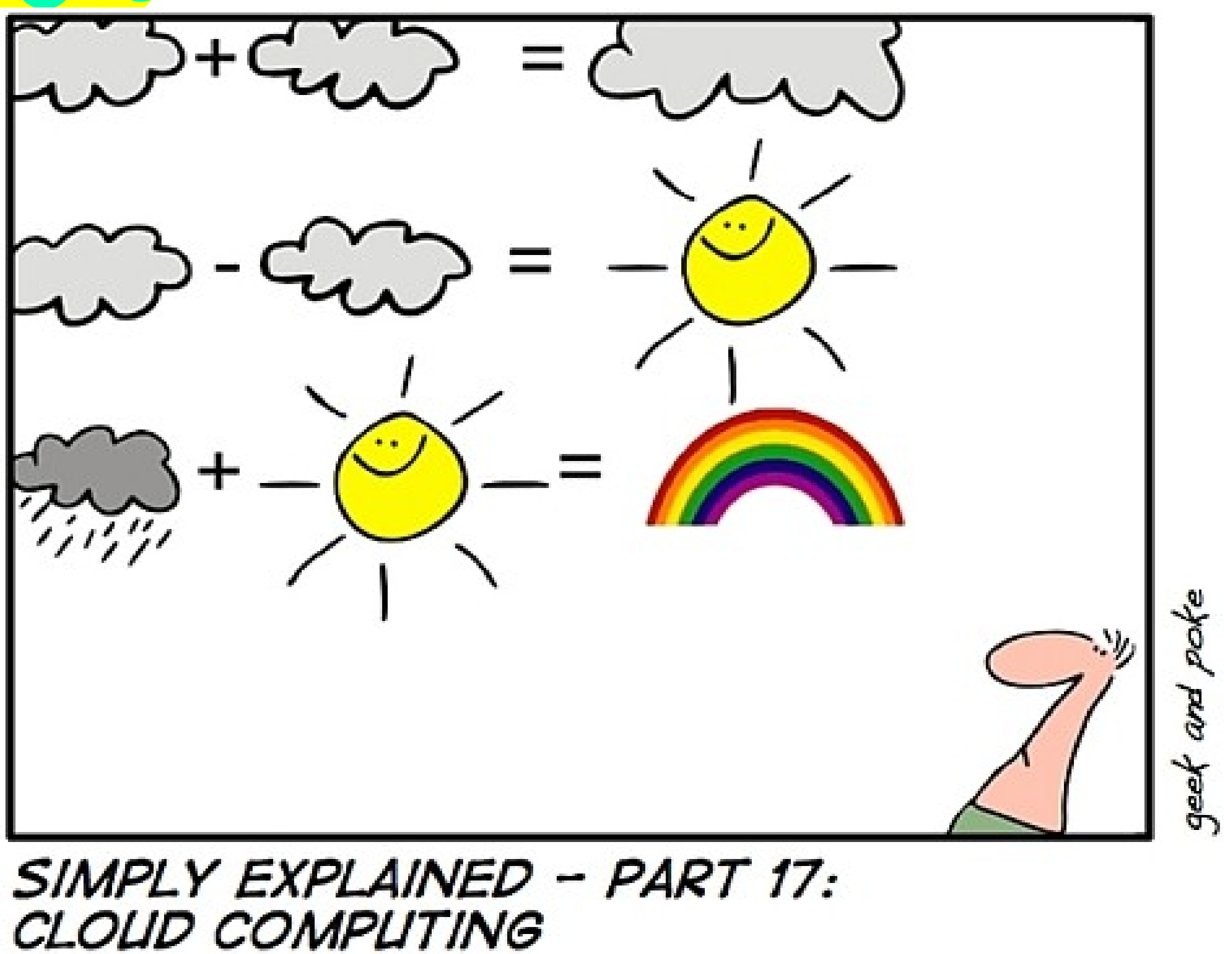
Jhon Masschelein, Tom Visser, Dennis Blommesteijn, Neil Mooney, Jeroen Nijhof, Alain van Hoof, Floris Sluiter et. al.

http://www.cloud.sara.nl

BiG Grid
the dutch e-science grid

Demo

Thank you!



Questions?

www.cloud.sara.nl

photo: http://cloudappreciationsociety.org/

BiG Grid
the dutch e-science grid