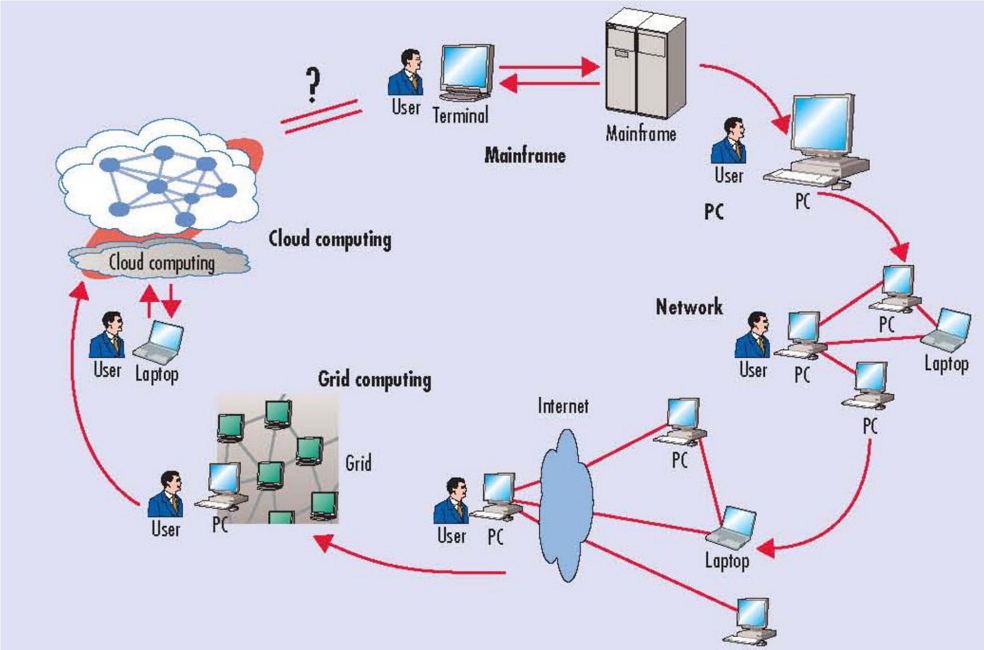


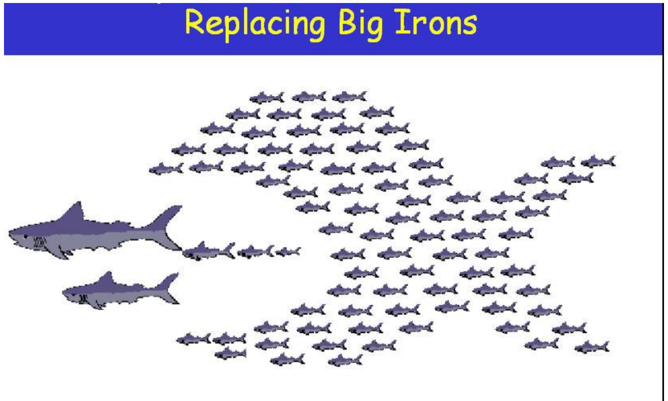
From: T.Dillon et al. 2010. Cloud Computing: Issues and Challenges. AINA'10

HPC in the Cloud

Dana Petcu, West University of Timisoara, Romania

Content

- Which is the biggest and more powerful?
- What I can do with the biggest and powerful?
- A use case at a small scale
- What's next?



Motto:

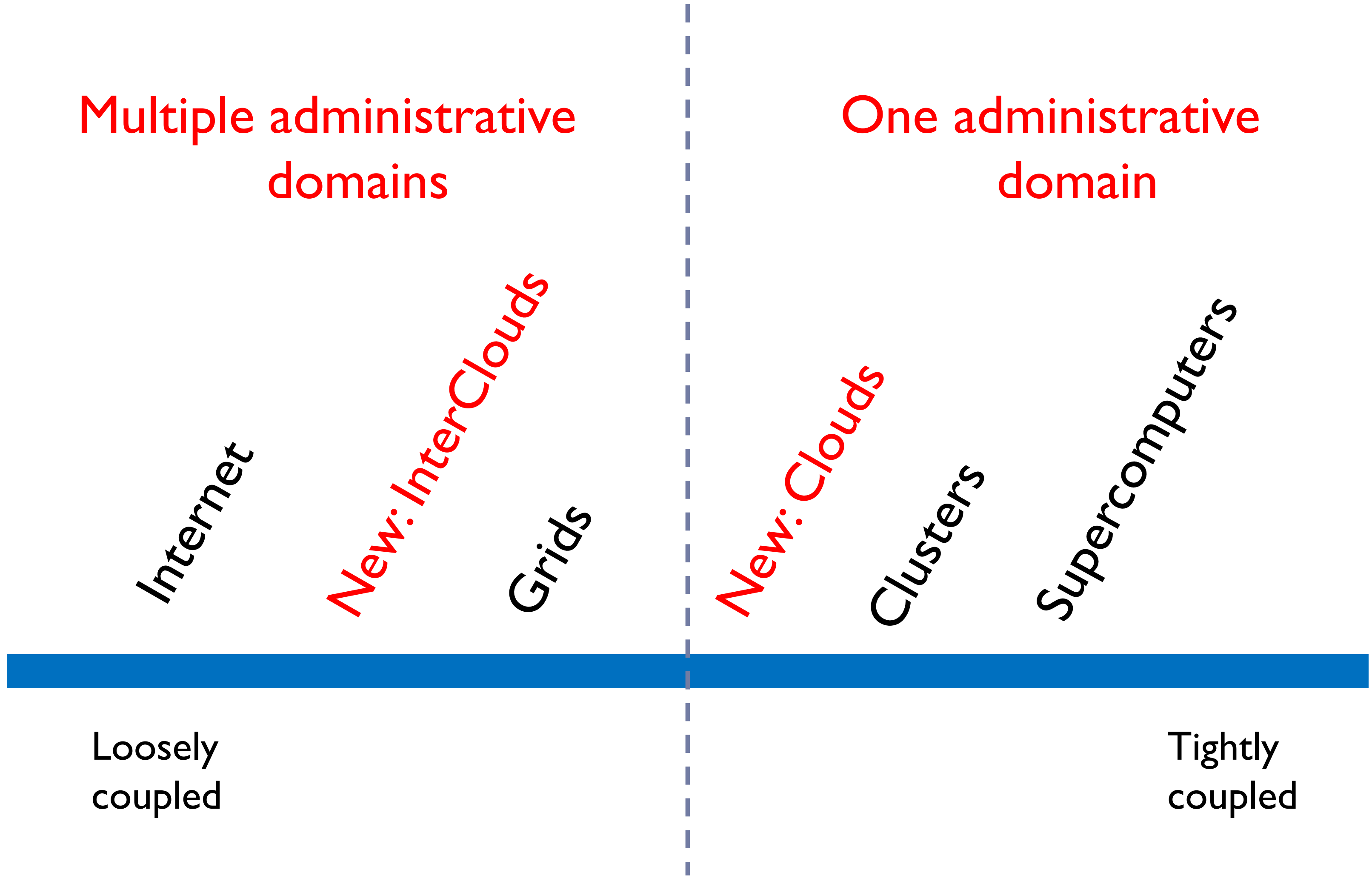
"The computer industry

is the only industry that is more fashion-driven

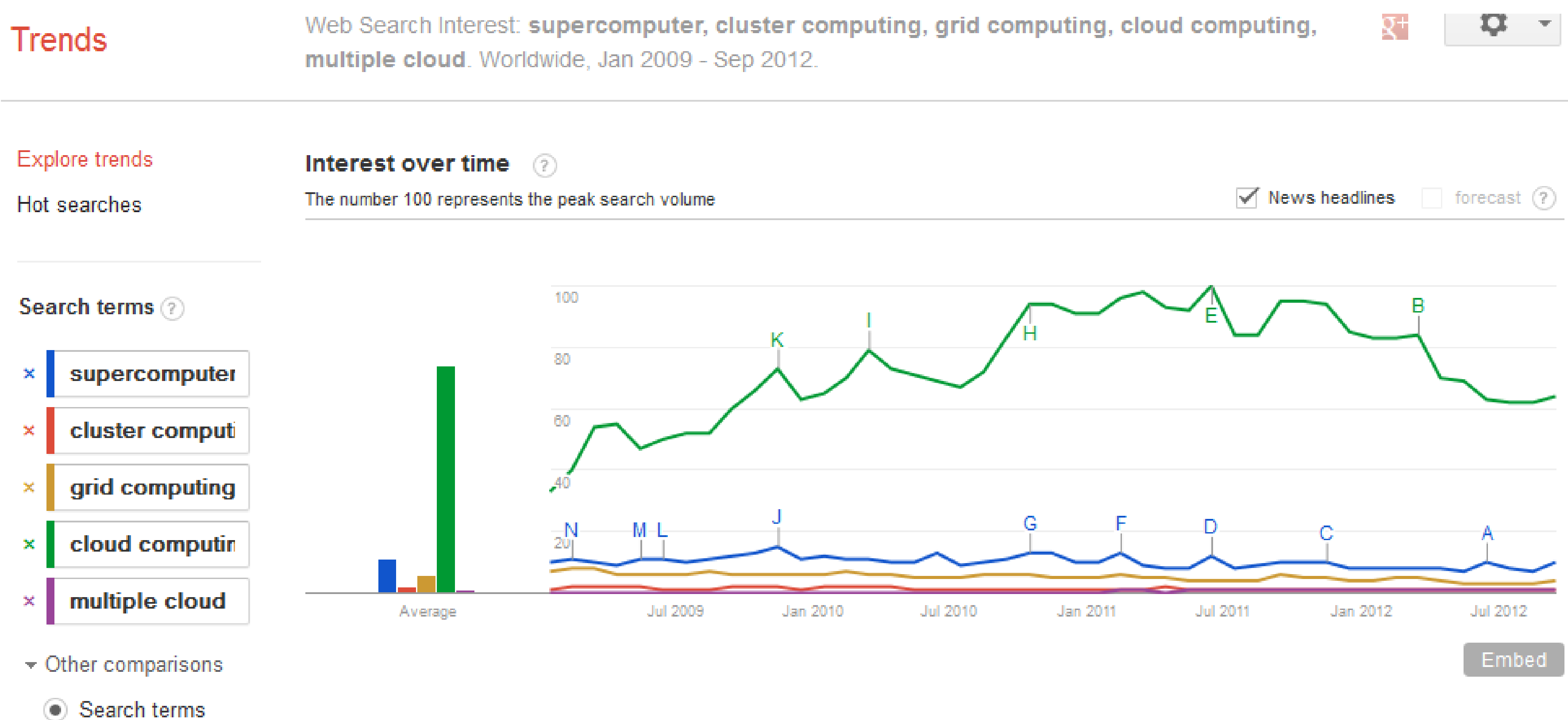
than women's fashion" [Oracle]"

The biggest and the powerfull

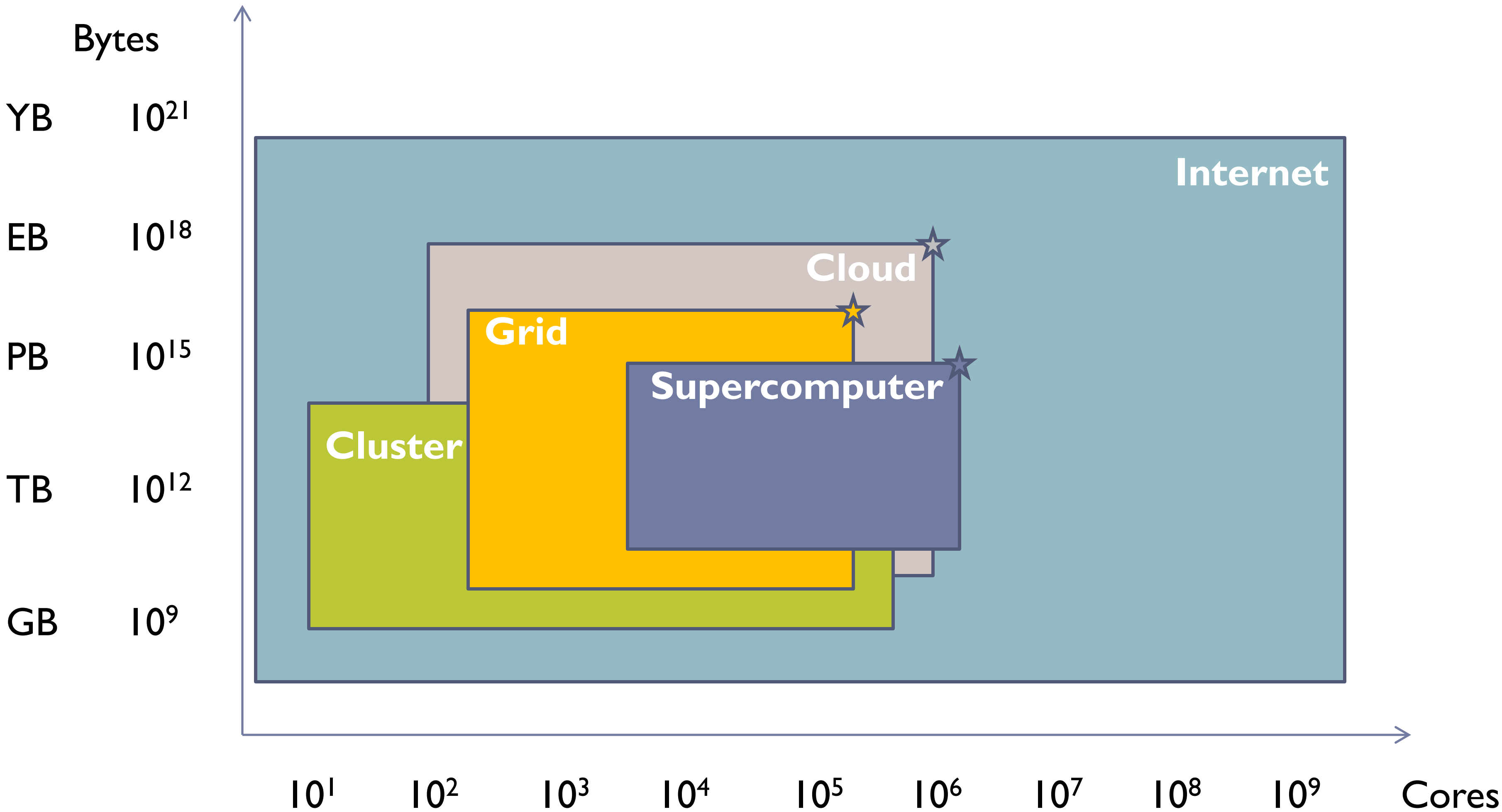
Updated Computing Continuum



Asked Google which is the most trendy



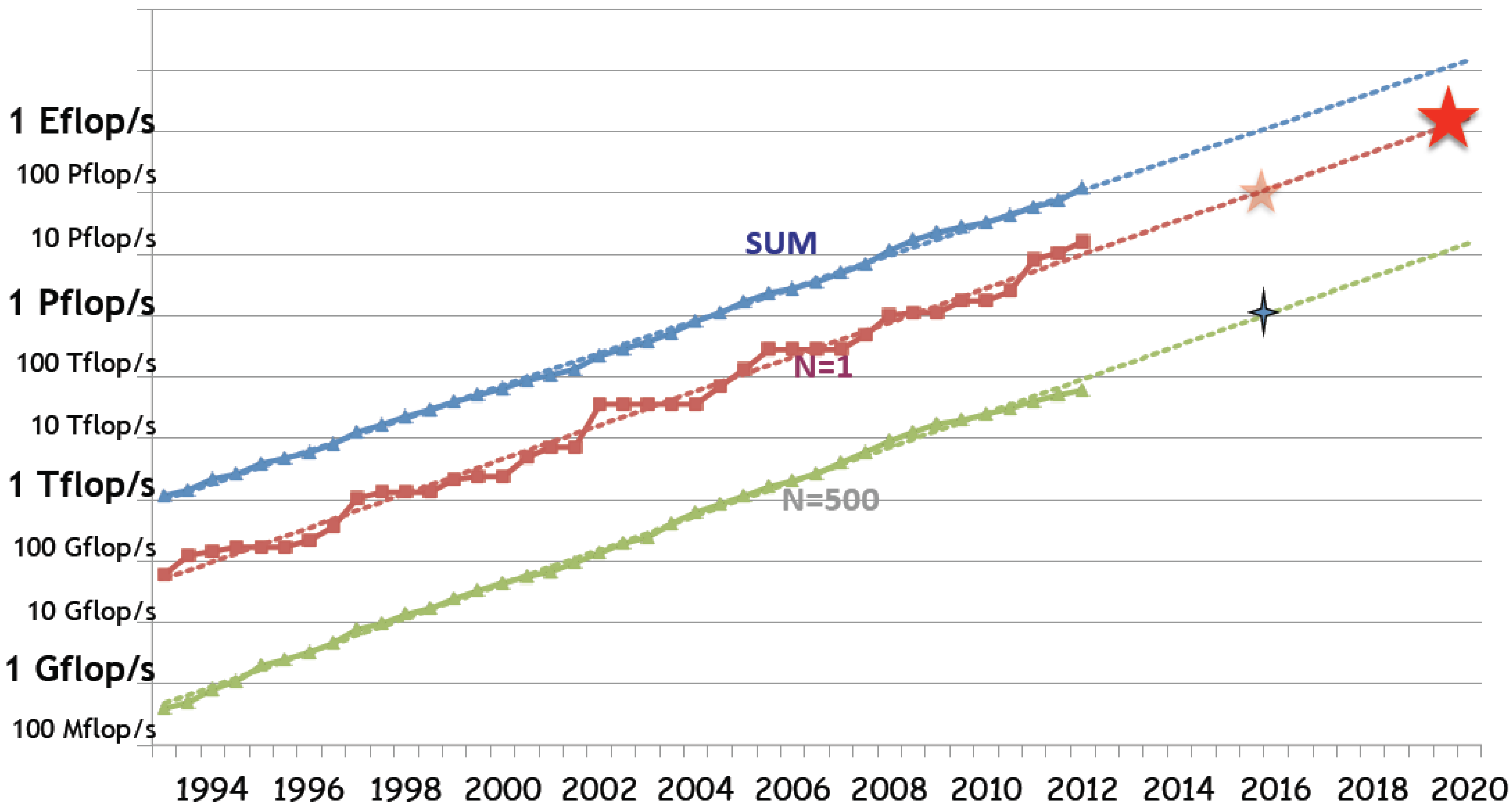
Characteristics: resources number



Top 500: the most powerful ones [June 2012]

Rank	Name	Site	Manufacturer	Country	Year	Segmen	Total Cor	Archi	Processor	Processor Tec	Process	OS F	Cores/	System Model	Intercon	Contine
1	Sequoia	DOE/NNSA/LLNL	IBM	United States	2011	Research	1572864	MPP	Power BQC 16C	PowerPC	1600	Linux	16	BlueGene/Q	Custom	Americas
2		RIKEN Advanced Institute for Com	Fujitsu	Japan	2011	Research	705024	Cluster	SPARC64 VIIIfx	Sparc	2000	Linux	8	K computer	Custom	Asia
3	Mira	DOE/SC/Argonne National Labora	IBM	United States	2012	Research	786432	MPP	Power BQC 16C	PowerPC	1600	Linux	16	BlueGene/Q	Custom	Americas
4	SuperMUC	Leibniz Rechenzentrum	IBM	Germany	2012	Academic	147456	Cluster	Xeon E5-2680 8	Intel SandyBridge	2700	Linux	8	iDataPlex DX360M4	Infiniband	Europe
5	Tianhe-1A	National Supercomputing Center i	NUDT	China	2010	Research	186368	MPP	Xeon X5670 6C	Intel Nehalem	2930	Linux	6	NUDT YH MPP	Proprietary	Asia
6	Jaguar	DOE/SC/Oak Ridge National Labo	Cray Inc.	United States	2009	Research	298592	Cluster	Opteron 6274 1	AMD x86_64	2200	Linux	16	Cray XK6	Cray Gemi	Americas
7	Fermi	CINECA	IBM	Italy	2012	Academic	163840	MPP	Power BQC 16C	PowerPC	1600	Linux	16	BlueGene/Q	Custom	Europe
8	JuQUEEN	Forschungszentrum Juelich (FZJ)	IBM	Germany	2012	Research	131072	MPP	Power BQC 16C	PowerPC	1600	Linux	16	BlueGene/Q	Custom	Europe
9	Curie thin	CEA/TGCC-GENCI	Bull SA	France	2012	Research	77184	Cluster	Xeon E5-2680 8	Intel SandyBridge	2700	Linux	8	Bullx B510	Infiniband	Europe
10	Nebulae	National Supercomputing Centre i	Dawning	China	2010	Research	120640	Cluster	Xeon X5650 6C	Intel Nehalem	2660	Linux	6	Dawning TC3600 Blade	Infiniband	Asia
11	Pleiades	NASA/Ames Research Center/NAS	SGI	United States	2011	Research	125980	MPP	Xeon E5450 4C	Intel Core	3000	Linux	4	SGI Altix ICE 8200EX/8	Infiniband	Americas
12	Helios	International Fusion Energy Resea	Bull SA	Japan	2011	Academic	70560	Cluster	Xeon E5-2680 8	Intel SandyBridge	2700	Linux	8	Bullx B510	Infiniband	Asia
13	Blue Joule	Science and Technology Facilities	IBM	United Kingdom	2012	Research	114688	MPP	Power BQC 16C	PowerPC	1600	Linux	16	BlueGene/Q	Custom	Europe
14	TSUBAME	GSIC Center, Tokyo Institute of Te	NEC/HP	Japan	2010	Academic	73278	Cluster	Xeon X5670 6C	Intel Nehalem	2930	Linux	6	Cluster Platform SL390	Infiniband	Asia
15	Cielo	DOE/NNSA/LANL/SNL	Cray Inc.	United States	2011	Research	142272	MPP	Opteron 6136 8	AMD x86_64	2400	Linux	8	Cray XE6	Custom	Americas
16	Hopper	DOE/SC/LBNL/NERSC	Cray Inc.	United States	2010	Research	153408	MPP	Opteron 6172 1	AMD x86_64	2100	Linux	12	Cray XE6	Custom	Americas
17	Tera-100	Commissariat a l'Energie Atomique	Bull SA	France	2010	Research	138368	Cluster	Xeon X7560 8C	Intel Nehalem	2260	Linux	8	bullx super-node S6010	Infiniband	Europe
18	Oakleaf-F	Information Technology Center, T	Fujitsu	Japan	2012	Academic	76800	Cluster	SPARC64 IXfx 1	Sparc	1848	Linux	16	PRIMEHPC FX10	Tofu interc	Asia
19	Roadrunn	DOE/NNSA/LANL	IBM	United States	2009	Research	122400	Cluster	PowerXCell 8i 9	Power	3200	Linux	9	BladeCenter QS22 Clu	Infiniband	Americas
495		IT Services Provider	Hewlett-Packard	United States	2012	Industry	13980	Cluster	Xeon E5620 4C	Intel Nehalem	2400	Linux	4	Cluster Platform 3000 E	Gigabit Et	Americas
496		IT Service Provider	Hewlett-Packard	United States	2010	Industry	10056	Cluster	Xeon X5670 6C	Intel Nehalem	2930	Linux	6	Cluster Platform 3000 E	Gigabit Et	Americas
497	Tsessebe	Centre for High Performance Com	Dell/Oracle	South Africa	2009	Academic	6336	Cluster	Xeon X5570 4C	Intel Nehalem	2930	Linux	4	Blade X6275/ PowerEd	Infiniband	Africa
498		Web Company (F)	Hewlett-Packard	United States	2012	Industry	11040	Cluster	Xeon X5650 6C	Intel Nehalem	2660	Linux	6	Cluster Platform SL160	Gigabit Et	Americas
499		Energy Company (A)	IBM	Italy	2012	Industry	4096	Cluster	Xeon E5-2650 8	Intel SandyBridge	2000	Linux	8	BladeCenter HS23 Clu	Infiniband	Europe
500		IT Service Provider	Hewlett-Packard	United States	2012	Industry	6064	Cluster	Xeon X5672 4C	Intel Nehalem	3200	Linux	4	Cluster Platform 3000 2	Infiniband	Americas

Projected performance [J. Dongarra, June'12]

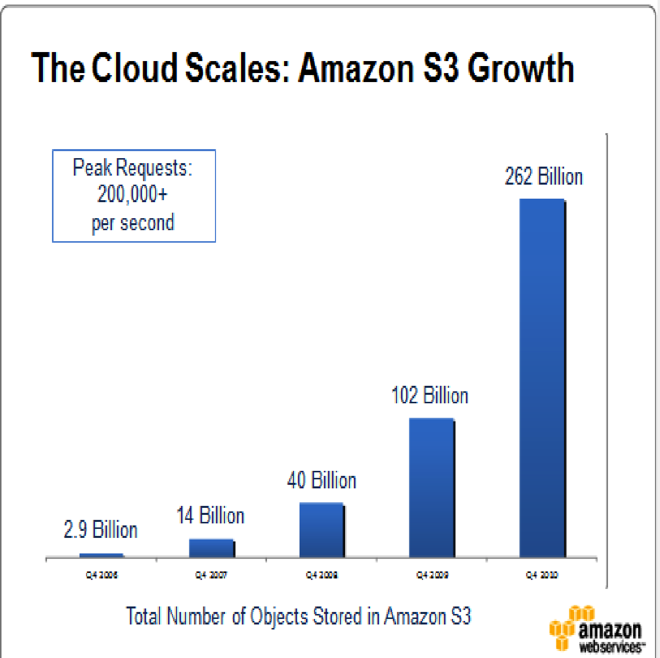


Top 500: the biggest supercomputers & clusters

Sequoia/DOE

RIKEN

Cloud: the biggest (?)



Estimated 900 PB

huanliu.wordpress.com/2012/03/13/amazon-data-center-size/

← Host server CPU utilization in Amazon EC2 cloud   Amazon DynamoDB use cases →

Amazon data center size

MARCH 13, 2012   94 COMMENTS

*(Edit 3/16/2012: I am surprised that this post is picked up by a lot of media outlets. Given the strong interest, I want to emphasize what is measured and what is derived. The # of server racks in EC2 is what I am directly observing. By assuming 64 physical servers in a rack, I can derive the rough server count. But remember this is an *assumption*. Check the comments below that some think that AWS uses 1U server, others think that AWS is less dense. Obviously, using a different assumption, the estimated server number would be different. For example, if a credible source tells you that AWS uses 36 1U servers in each rack, the number of servers would be 255,600. An additional note: please visit my disclaimer page. This is a personal blog, only represents my personal opinion, not my employer's.)*

Similar to the EC2 CPU utilization rate, another piece of secret information Amazon will never share with you is the size of their data center. But it is really informative if we can get a glimpse, because Amazon is clearly a leader in this space, and their growth rate would be a great indicator of how well the cloud industry is doing.

Although Amazon would never tell you, I have figured out a way to probe for its size. There have been early guesstimates on how big Amazon cloud is, and there are even tricks to figure out how many virtual machines are started in EC2, but this is the first time anyone can estimate the real size of Amazon EC2.

The methodology is fully documented below for those inquisitive minds. If you are one of them, read it through and feel free to point out if there are any flaws in the methodology. But for those of you who just want to know the numbers: Amazon has a pretty impressive infrastructure. The following table shows the number of server racks and physical servers each of Amazon's data centers has, as of Mar. 12, 2012. The column on server racks is what I directly probed (see the methodology below), and the column on number of servers is derived by assuming there are 64 blade servers in each rack.

data center\size	# of server racks	# of blade servers
US East (Virginia)	5,030	321,920
US West (Oregon)	41	2,624
US West (N. California)	630	40,320
EU West (Ireland)	814	52,096
AP Northeast (Japan)	314	20,096
AP Southeast (Singapore)	246	15,744
SA East (Sao Paulo)	25	1,600
Total	**7,100**	**454,400**

The first key observation is that Amazon now has close to half a million servers, which is quite impressive. The other observation is that the US east data center, being the first data center, is much bigger. What it means is that it is hard to compete with Amazon on scale in the US, but in other regions, the entry barrier is lower. For example, Sao

Cloud:
the biggest (?)

- June 2012:

www.extremetech.com/extreme/131962-google-compute-engine-for-2-millionday-your-company-can-run-the-third-faste

Google Compute Engine: For $2 million/day, your company can run the third fastest supercomputer in the world

By Sebastian Anthony on June 28, 2012 at 3:18 pm | 3 Comments

Share This Article

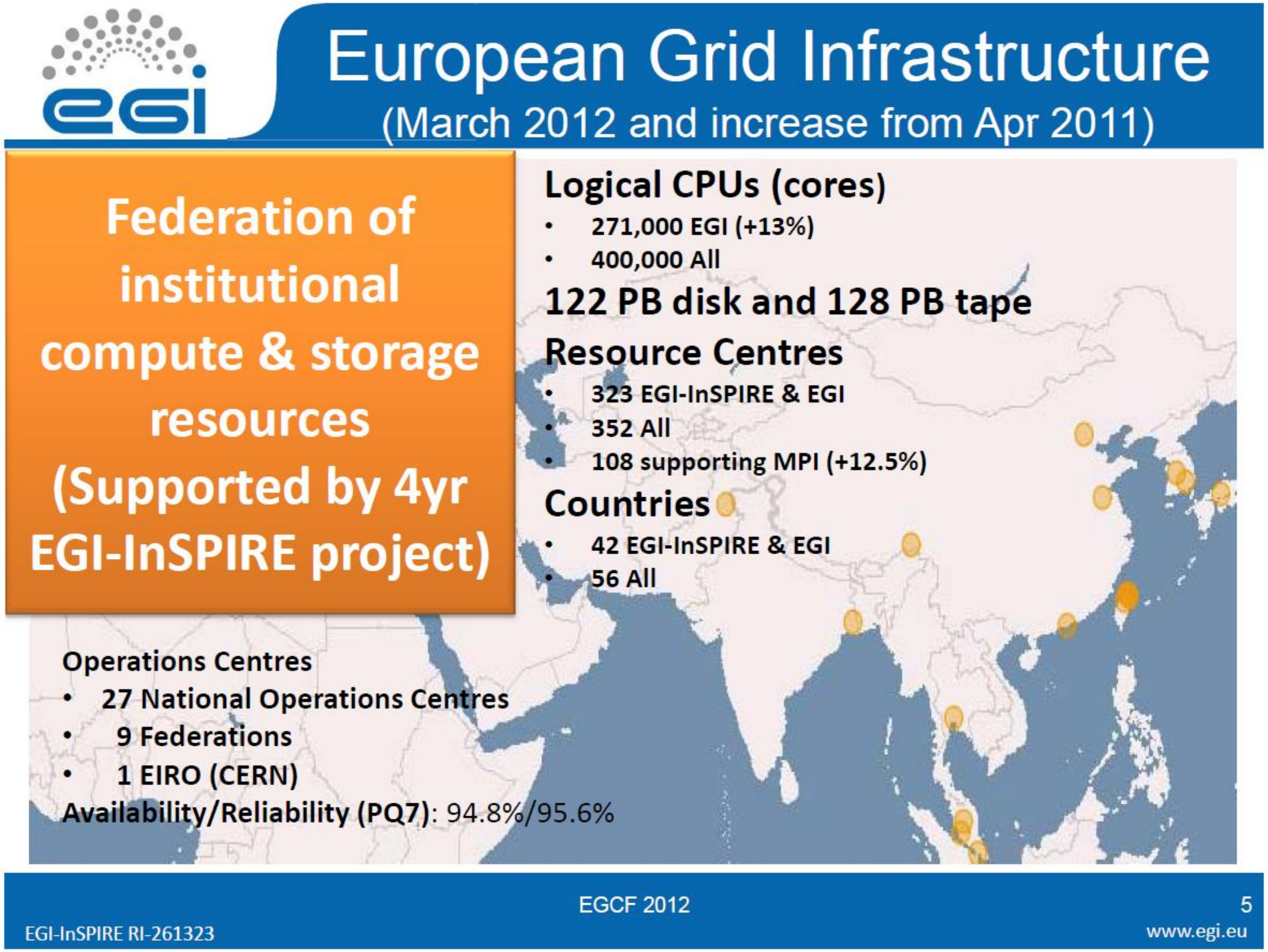
128 Like | 53 Tweet | 2,613 | 20 +1 | reddit

At the Google I/O conference in San Francisco, Google has announced the immediate availability of Compute Engine, an infrastructure-as-a-service (IAAS) product that directly competes with Amazon EC2 and Microsoft Azure. Citing more than a decade of running and optimizing its own data centers and network infrastructure, Google is claiming that the Compute Engine is more scalable, more stable, and cheaper than the competition.

For this story, we'll focus on scalability and cost (I'm sure that Compute Engine is stable, but Google just hasn't given us any figures to work with). Google says that Compute Engine has access to 770,000 cores — a figure that will surely grow over time. In one demo at Google I/O, a genomics app (it analyzed the human genome) was shown to use 600,000 cores. These cores are made available as Linux virtual machines (VMs), with 1, 2, 4, or 8 cores each. Each core apparently has access to 3.75GB of RAM each — and, of course, each VM is connected together using Google's advanced networking technologies and topologies.

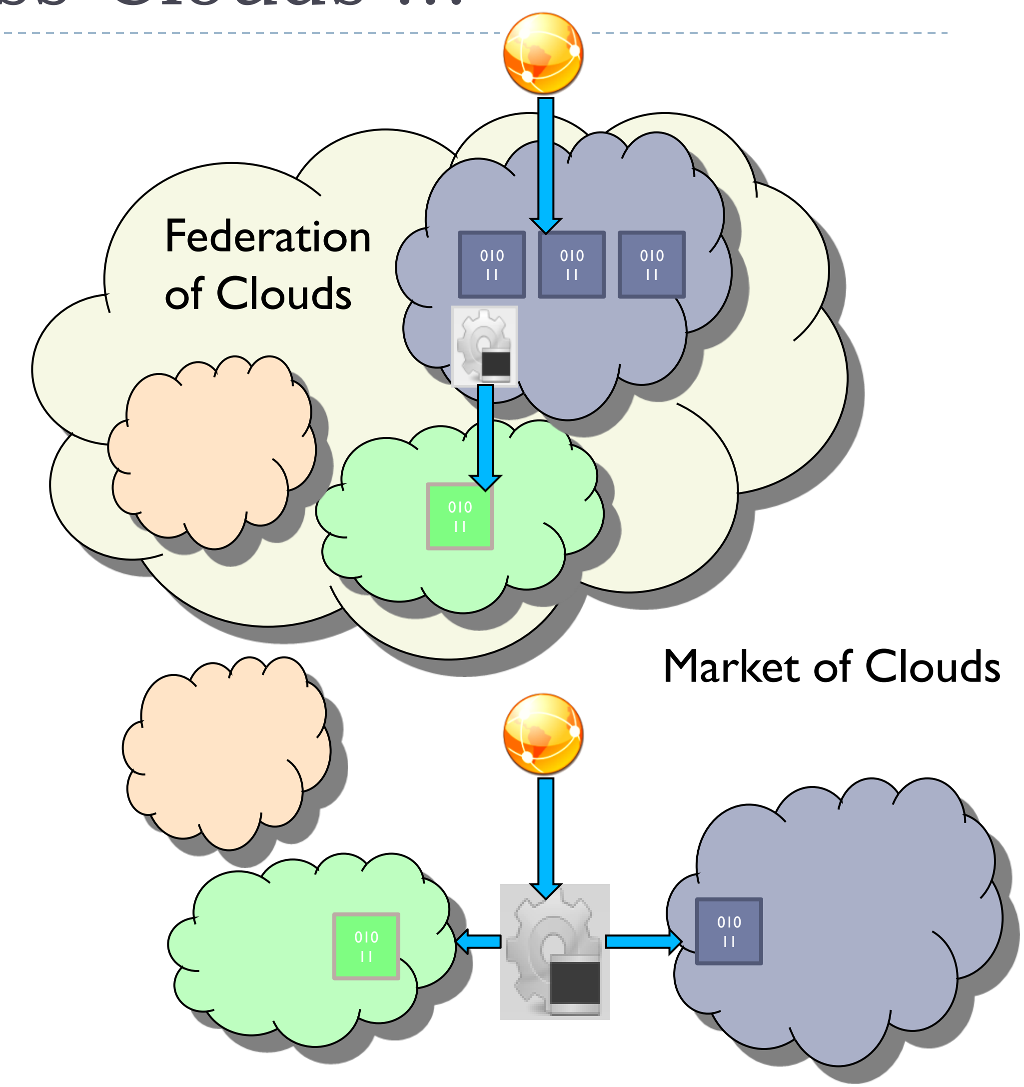
777,000 cores, assuming the entire Compute Engine cluster consists of 8-core CPUs, equates to 96,250 computers. This is a huge number — probably equal to the total number of servers operated by Intel, or data centers such as The Planet or Rackspace, but

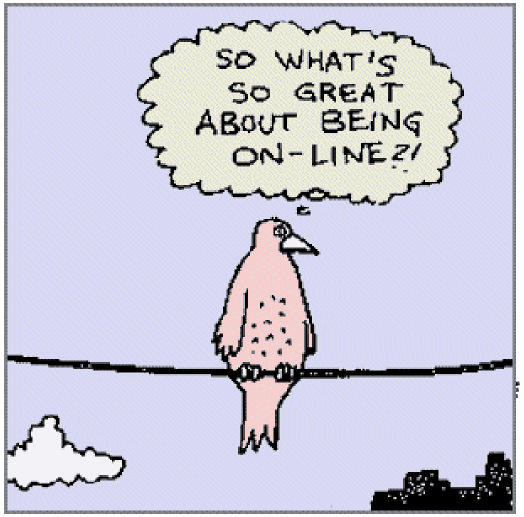
Grids: the biggest



InterCloud, multiple Clouds, Sky computing, Cross-Clouds …

- interconnected
  "cloud of clouds"
- extension of the Internet "network of networks"





Following the giants: ‘Big’ and Famous Applications

on e-Infrastructures

“Classical” HPC applications

- Type of applications:
  - Weather forecast and climate research
  - Molecular modeling (e.g. crystals, biology, chemistry)
  - Quantum physics and physical simulations (e.g. nuclear fusion)
- Open HPC applications:
  - Bio-informatics:
    - mpiBLAST, MPI-HMMER
  - Molecular Dynamics:
    - GROMCAS, NAMD, Desmond, OpenAtom
  - Environment/Weather
    - POP, WRF, MM5

Appls/supercomputers & big clusters [Top500]

Application Area	Count	System Share (%)	Rmax (GFlops)	Rpeak (GFlops)	Cores
Not Specified	209	41.8	60037590.69	82863853.8	6440642
Research	105	21	40532017.25	53789204.6	4213217
Finance	25	5	1801282.97	3512856.38	335444
Web Services	21	4.2	1755179.7	3249561	295844
Energy	17	3.4	3221250.39	4311936.38	276356
Geophysics	14	2.8	1225886	2918282.4	100624
Services	14	2.8	988820.5	1753013.4	164756
Defense	13	2.6	2588070.4	3138660.08	319536
Weather and Climate Research	13	2.6	3934162	5152868.06	351460
Logistic Services	8	1.6	531532.93	1013975.9	92722
IT Services	8	1.6	566670.5	1098033.52	106572
Entertainment	7	1.4	497856	692428.4	61824
Aerospace	7	1.4	1903523	2528001.47	202508
Environment	6	1.2	754030	1250227.72	59776
Benchmarking	6	1.2	911196	1127694.4	66176
Information Service	5	1	402436.66	722117.44	63452
Information Processing Service	5	1	345266	569035.52	85856
Automotive	2	0.4	177240	200833.92	17136
Telecommunication	2	0.4	150995.72	277047.36	26796
Internet Provider	2	0.4	162555	306390.66	31776
Transportation	2	0.4	126084	237144.32	22288
Semiconductor	2	0.4	180472	253384.72	18360
Electronics	2	0.4	124488	139937.28	13152
Software	1	0.2	172691	209715	16384
Medicine	1	0.2	63830	94208	10240
Cloud Services	1	0.2	89940	155079	4968
Life Science	1	0.2	97071	159948.8	18176
Retail	1	0.2	75649	145705	11904

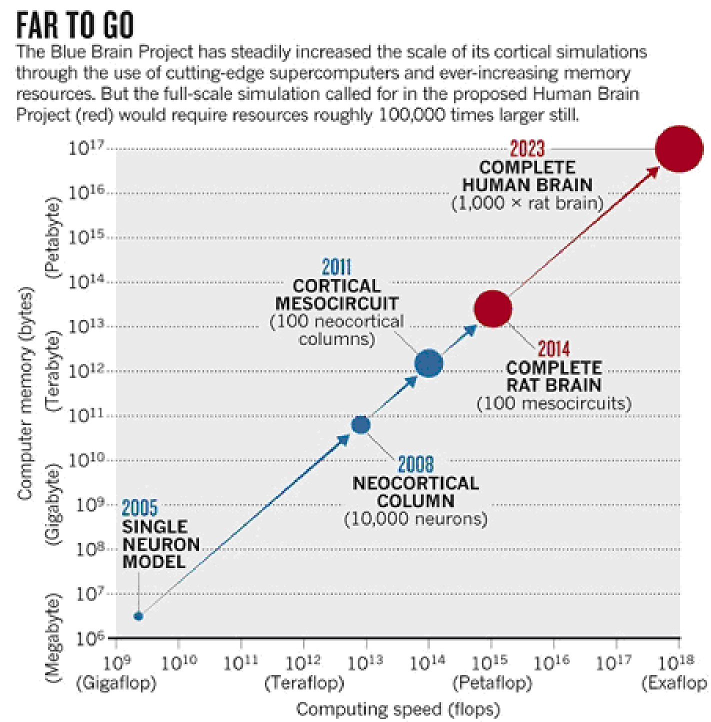
There are appls which can reach exascale?

- E.g. ExaScience Lab, Leuven
  - Space weather predictions
- DOE – Grand challenge workshop 2011

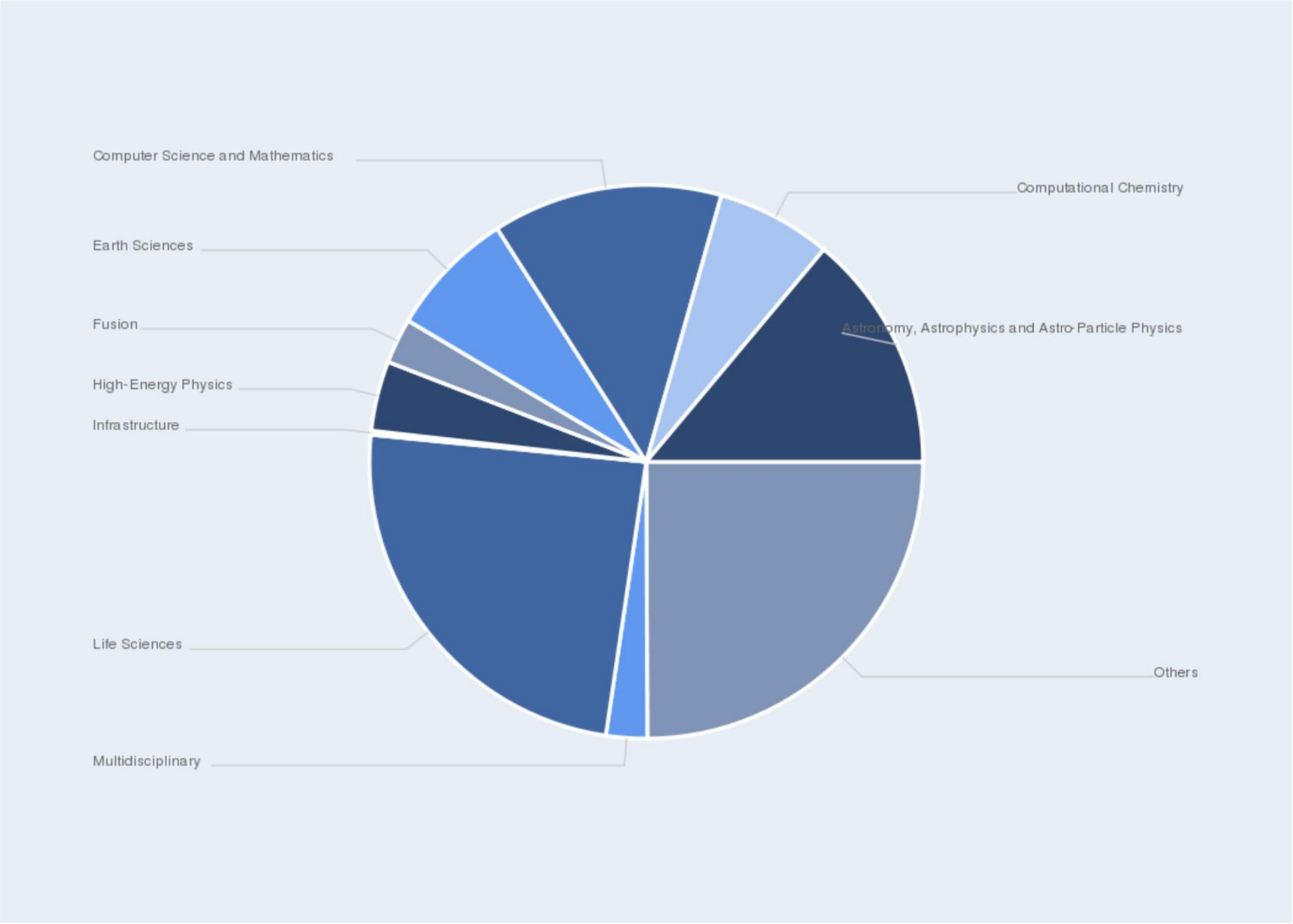
http://science.energy.gov/ascr/news-and-resources/workshops-and-conferences/grand-challenges/

  - Blue Brain project

Driving Applications Areas	
Circa 1990	Circa 2018-2025
Weather & Ocean Modeling	Climate Modeling
Chemistry & Materials	Chemistry & Materials
Plasma Modeling	Fusion Energy Sciences
Computational Biology & Bioinformatics	Computational Biology & Bioinformatics
High Energy/Quantum Physics	High Energy/Quantum Physics
Combustion Systems	Combustion Systems
Computational Electromagnetics	National Security Applications
Computational Fluid Dynamics (various)	Computational Fluid Dynamics (various)
Semiconductor Modeling & Design	
Superconductor Modeling	
Pharmaceutical Design	
Speech & Natural Language	
Vision & Cognition	
	Nuclear Physics
	Nuclear Energy Systems

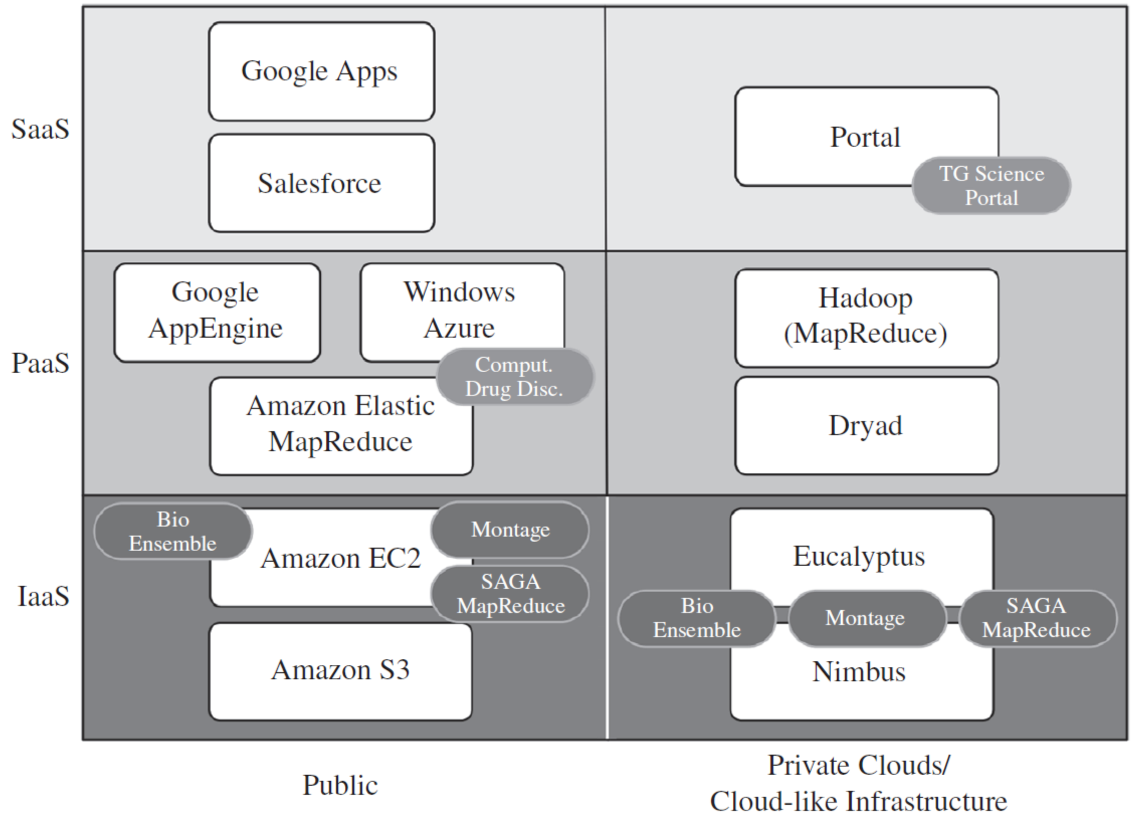


Applications on Grids [EGI statistics]

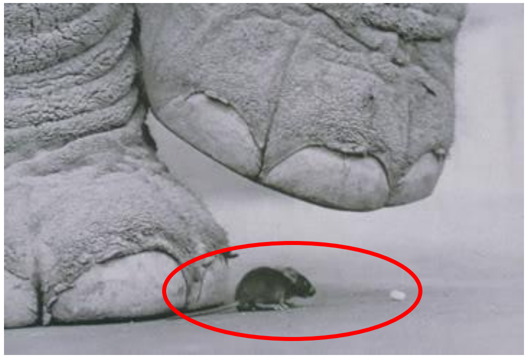


Scientific applications on Clouds

- A typical example:



From: UNDERSTANDING SCIENTIFIC APPLICATIONS FOR CLOUD ENVIRONMENTS
S. JHA, D.S. KATZ, A. LUCKOW, A. MERZKY, K. STAMOU,
Cloud Computing: Principles and Paradigms,
Edited by R. Buyya, J. Broberg and A. Goscinski
2011 John Wiley & Sons, Inc.



Appl supported on own e-Infras:
- Crystal growing simulations
- Airfoil design
- Data mining in medical databases
- Expert systems for numerics
- Membrane computing simulations
- Earth observation services
- …

Tools for supporting appls:
- EpODE, NESS
- PVMMaple, Maple2Grid
- Parallelless
- GiSHEO, ESIP
- mOSAIC
- …

To port or not to port my application?

A use case. UVT team experience

What we can do with these? [UVT equipments]

Blue Gene/P

4096 cores

Cluster

400 cores

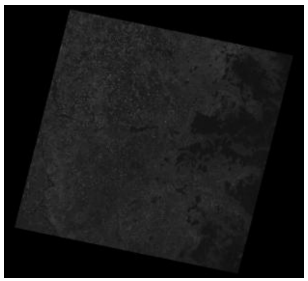
Earth Observation problems

- Both computational and data intensive
- Real time processing confronts several difficulties in one single computer and even impossibility
- Need of a computational environment handling
  - hundreds of distributed databases,
  - heterogeneous computing resources,
  - and simultaneous use

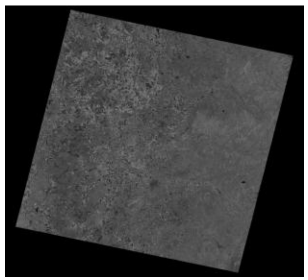
From the small to the big

- Simple algorithms: merge

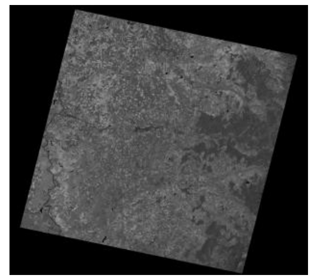
Band 3



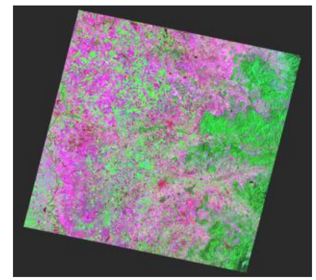
Band 4



Band 5

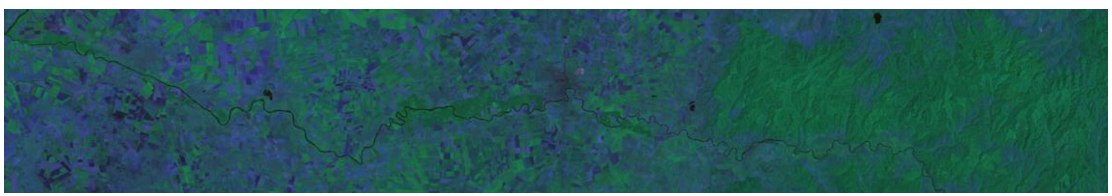


Pseudo-image

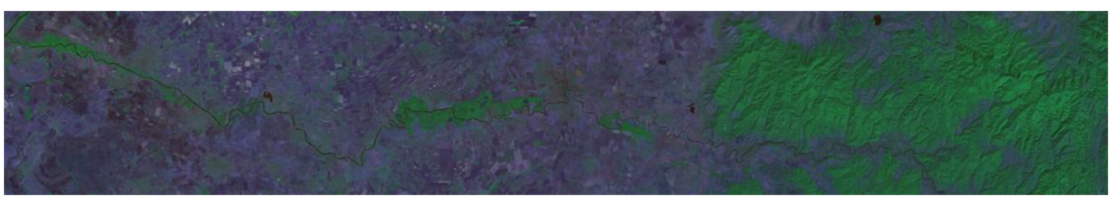


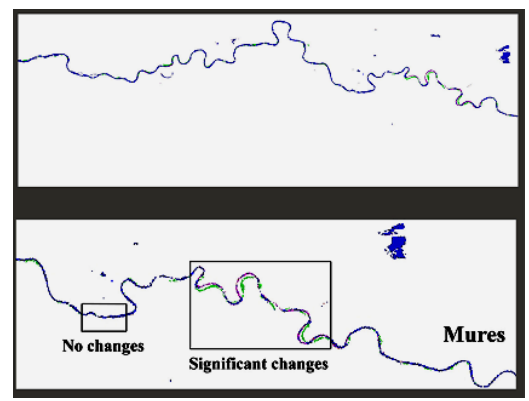
- Computational intensive algorithms

1992



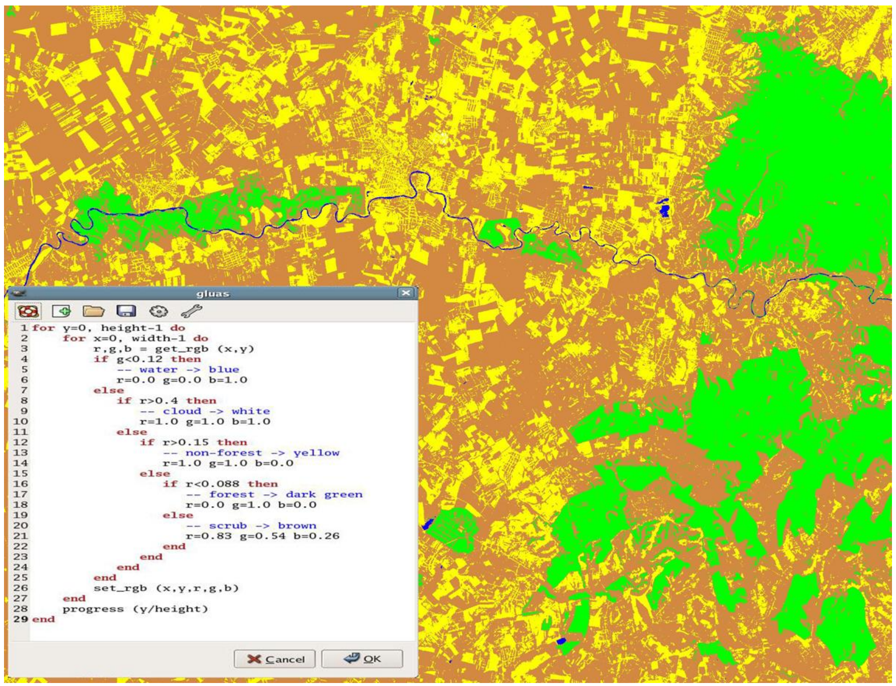
2000





Why Clusters [medioGRID project]

- Store the big data
- Process the data where they are



No. of processors	1	2	4
Time (s)	457	256	168
Speedup	-	1.78	2.72
Efficiency	-	89%	68%

E.g. D.Petcu, V. Iordan, Service based on GIMP for Processing Remote Sensing Images, SYNASC 2006

Why Supercomputer [HP-SEE project]

Algorithm 4 The general structure of the parallel Fuzzy c-Means

```
Read image slice X(p) = X_{i∈S_p}
Initialize the local membership values u_ij(p), i ∈ S_p, j = \overline{1,c}
iter = 0
repeat
   Compute C_j(p) = Σ_{i∈S_p} u_ij^m(p) X_i(p), j = \overline{1,c}
   Compute C'_j(p) = Σ_{i∈S_p} u_ij^m(p), j = \overline{1,c}
   Call MPI_Allreduce to compute C_j = C_j(1) + ... + C_j(P) for all j = \overline{1,c}
   Call MPI_Allreduce to compute C'_j = C'_j(1) + ... + C'_j(P) for all j = \overline{1,c}
   Compute V_j = C_j / C'_j, j = \overline{1,c}
   Update the local membership values u_ij^new, i ∈ S_p, j = \overline{1,c}
   Compute Err(p) = max_{i∈S_p, j=\overline{1,c}} |u_ij^new(p) − u_ij(p)|
   Call MPI_Allreduce to compute Err = max{Err(1), ..., Err(P)}
   iter = iter + 1
   u_ij = u_ij^new, i ∈ S_p, j = \overline{1,c}
until Err < ε or iter > iterMax
Compute the cluster validation measure(s)
if p==1 then
   Construct the classified image
end if
```

- Scalable algorithms

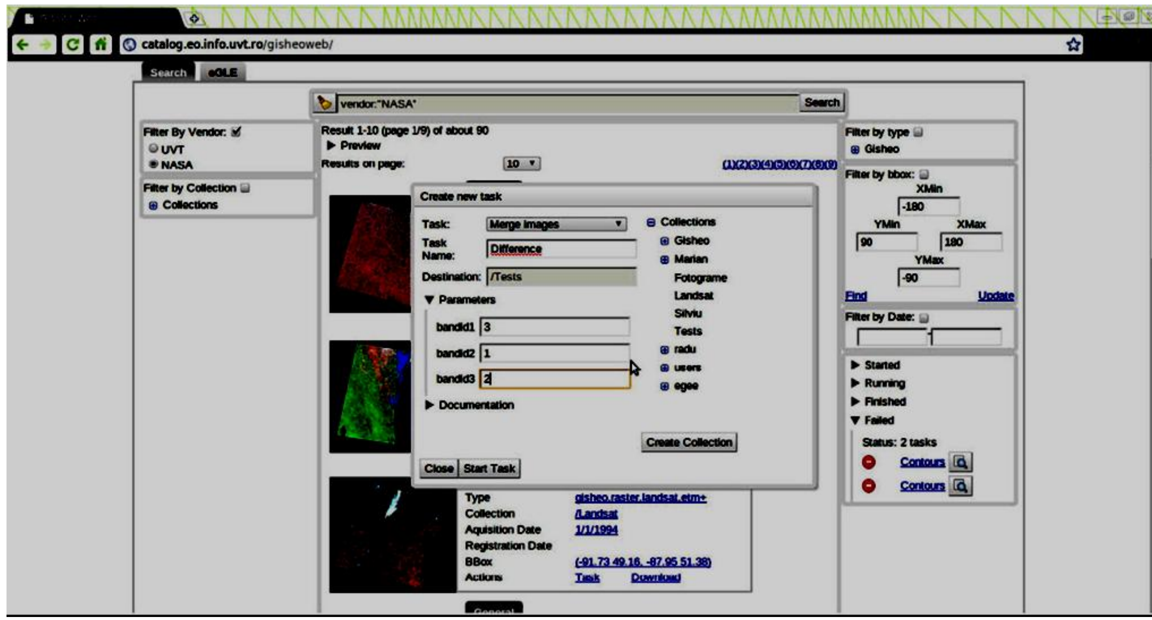
Table 8. Results on BlueGene/P for the parallel version of SFCM (100 iterations, 5 clusters, neighborhood size equal to 5). Test image: AVIRIS image (224 spectral bands, 1087 × 614 pixels)

No. Proc	K_w	K_h	P/16	Time(16)/ Time(P)	Total Time(s)	Time Send(s)	Time Reduce(s)	Time Send(%)	Time Reduce(%)
1024	32	32	64	40.38	4.94	0.10	1.64	2.09	33.27
1024	2	512	64	17.02	11.71	7.30	0.06	62.35	0.48
512	16	32	32	27.55	7.24	0.97	0.05	13.34	0.76
512	1	512	32	10.23	19.50	10.75	0.05	55.14	0.26
256	16	16	16	15.84	12.58	0.10	0.05	0.83	0.43
256	256	1	16	8.59	23.21	7.83	0.05	33.75	0.23
128	8	16	8	7.68	25.96	1.38	0.05	5.31	0.19
128	1	128	8	6.63	30.09	3.97	0.05	13.21	0.16
64	8	8	4	3.90	51.08	1.95	0.05	3.82	0.10
64	1	64	4	3.81	52.34	3.04	0.05	5.81	0.10
32	4	8	2	2.02	98.80	0.08	0.05	0.08	0.05
32	1	32	2	2.01	99.11	0.55	0.04	0.56	0.04
16	4	4	1	1.01	197.65	0.08	0.05	0.04	0.03
16	16	1	1	1.00	199.39	0.69	0.04	0.35	0.02

D. Petcu et al,
Fuzzy Clustering of Large Satellite Images using High Performance Computing,
In Procs of SPIE Volume 8183, no. 818302 (2011), SPIE Remote Sensing Conference: High-Performance Computing in Remote Sensing, 19-22 September 2011, Prague,
Doi:10.1117/12.898281

Why Grids [GiSHEO project]

- Remote services that can be combined
- Process the distributed data where they are

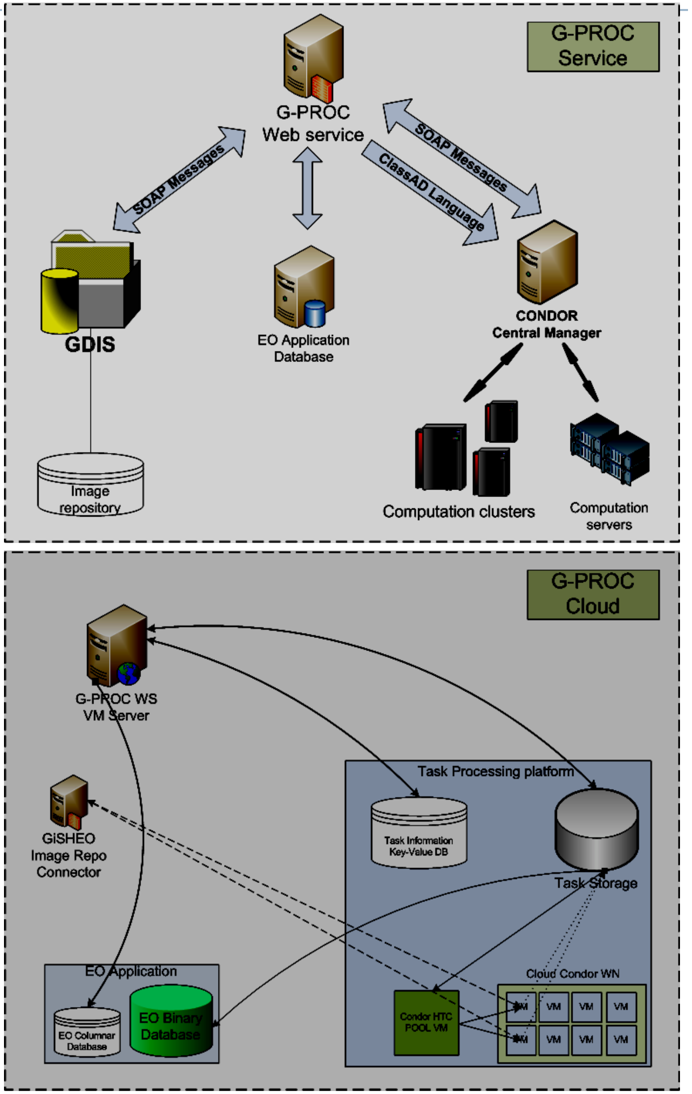


http://gisheo.info.uvt.ro

D. Petcu et al, Experiences in building a Grid-based platform to serve Earth observation training activities, Computer Standards Vol. 34 (6), 2012, 493-508, 10.1016/j.csi.2011.10.010.

Why Clouds [mOSAIC project]

- Store old data
- Share the data
- Reprocess according new algs



- Roberto Cossu, Claudio Di Giulio, Fabrice Brito, Dana Petcu, Cloud Computing for Earth Observation to appear in the book *Data Intensive Storage Services for Cloud Environments, 2012*

HPC services based on mOSAIC PaaS [HOST project]

- On-going work
  - First prototype in July 2013
- Reason:
  - offer services to consume the available resources

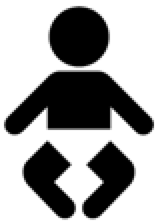
Scientific computing: Clouds vs. HPC

HPC [Batch processing]

- Advantages:
  - Fast communications
  - Full capacity usage
  - Reliability
  - Predictable performance
- Disadvantages:
  - Accounting procedures
  - Queues
  - Expensive maintenance
  - Large installations available in few countries

Clouds [Services]

- Advantages:
  - Fast availability
  - High level of accessibility
  - Programmable e-infrastructure
- Disadvantages:
  - Virtualization overheads
  - Costs charged to the users
  - Large installation usage still on request
  - Data transfer is prohibit
  - Non-predictable performance



What's next?

From the provider point of view

- Elastic Cluster
  - G. Mateescu, W. Gentzsch, C.J. Ribbens, "Hybrid Computing—Where HPC meets grid and Cloud Computing", FGCS 27, 2011
  - Unified model of managed HPC and Cloud resources

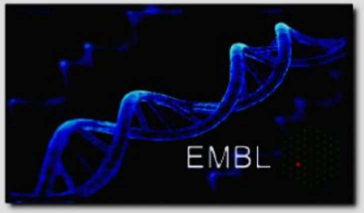
1. dynamic infrastructure management services (of which virtual infrastructure management services are a special case);
2. cluster-level services such as workload management;
3. intelligent modules that bridge the gap between cluster-level services and dynamic infrastructure management services.

  - Goal: execute scientific applications such that it satisfies the timing requirements of the applications
    - Timing constraints: deadlines, advance reservations, and best-effort

From an application point of view

Initial flagships use cases

HELIX NEBULA THESCIENCECLOUD

ATLAS High Energy Physics Cloud Use	Genomic Assembly in the Cloud	SuperSites Exploitation Platform
		
To support the computing capacity needs for the ATLAS experiment	A new service to simplify large scale genome analysis; for a deeper insight into evolution and biodiversity	To create an Earth Observation platform, focusing on earthquake and volcano research

- **Scientific challenges with societal impact**
- **Sponsored by user organisations**
- ***Stretch* what is possible with the cloud today**

http://www.helix-nebula.eu/

To do at application side

- Elastic scientific applications
  - E.g. simulators of membrane computing
  - Start with few machines and expand as needed
- Make elastic the components of the applications
  - Follow the example of the loosely coupled components of web applications

Take-away

- HPC in the Cloud needs to be improved in term of services
- Need to exploit elasticity