EGI - High Throughput Computing cost and sustainability

The European Grid Infrastructure, High Throughput Computing cost aspects and sustainability perspectives or Lessons learned

Per Öster
CSC – IT Center for Science Ltd
(EGI Council Chair)

EGI-InSPIRE Project

Integrated Sustainable Pan-European Infrastructure for Researchers in Europe

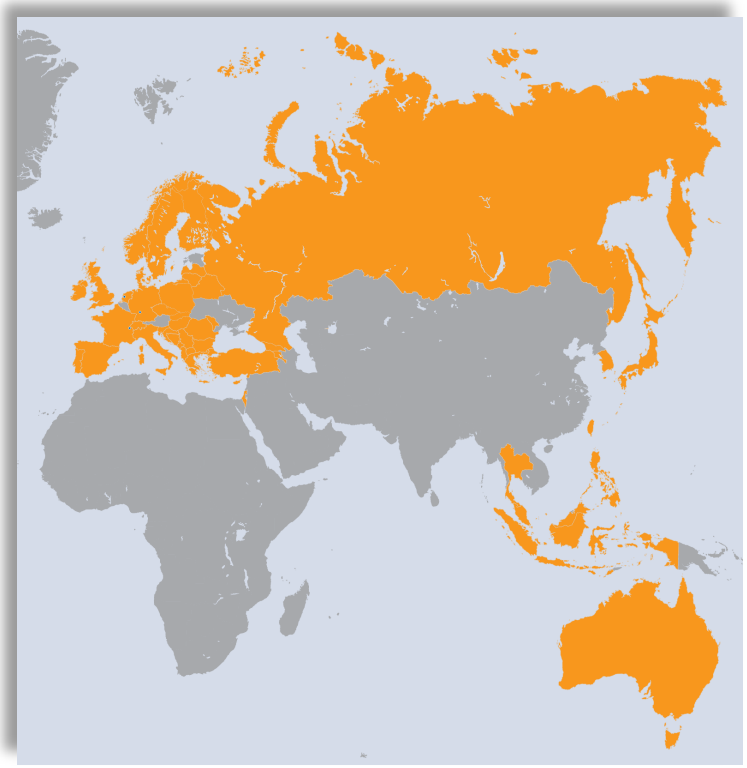
A 4 year project with €25M EC contribution

- Project cost €72M
- Total Effort ~€330M
- Effort: 9261PMs

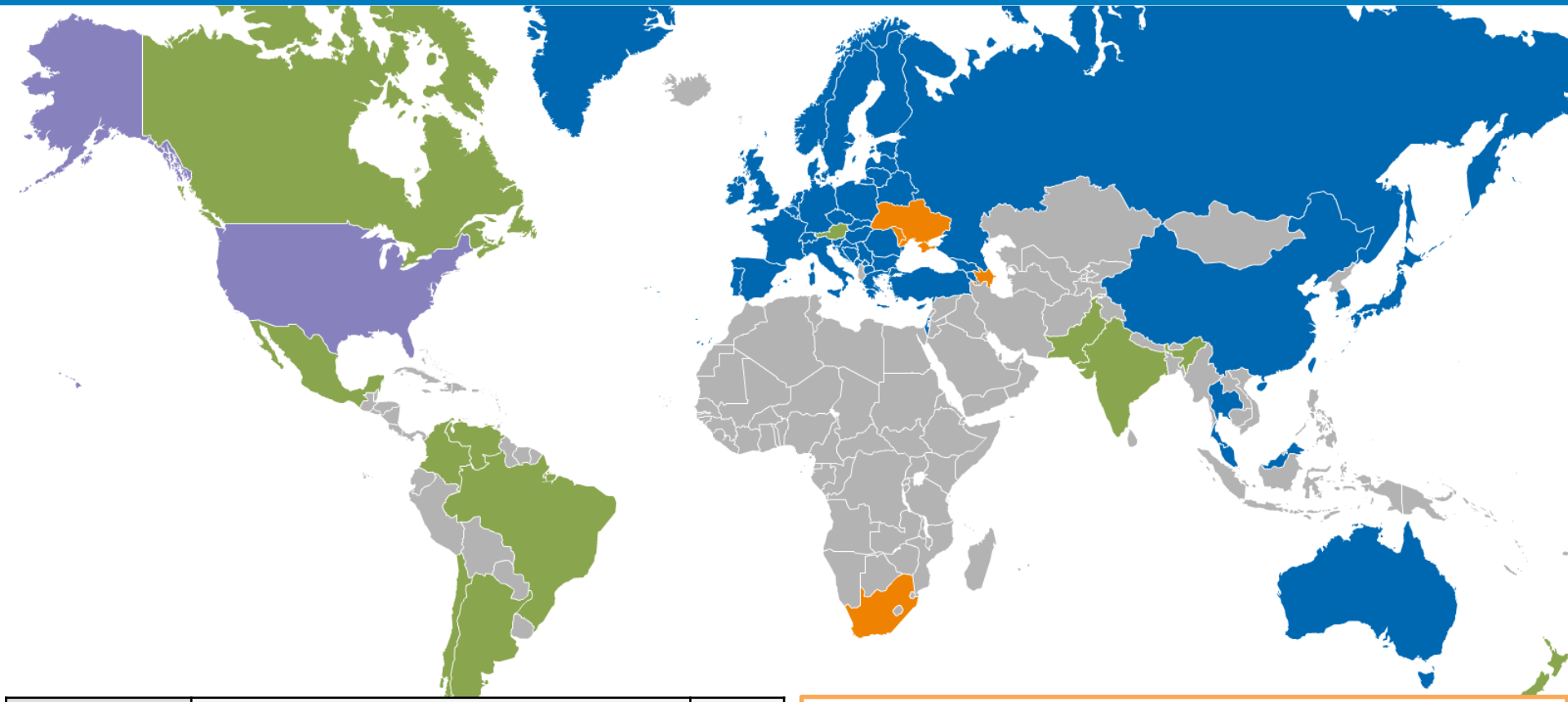
Project Partners (50)

EGI.eu, 38 NGIs, 2 EIROs

Asia Pacific (9 partners)



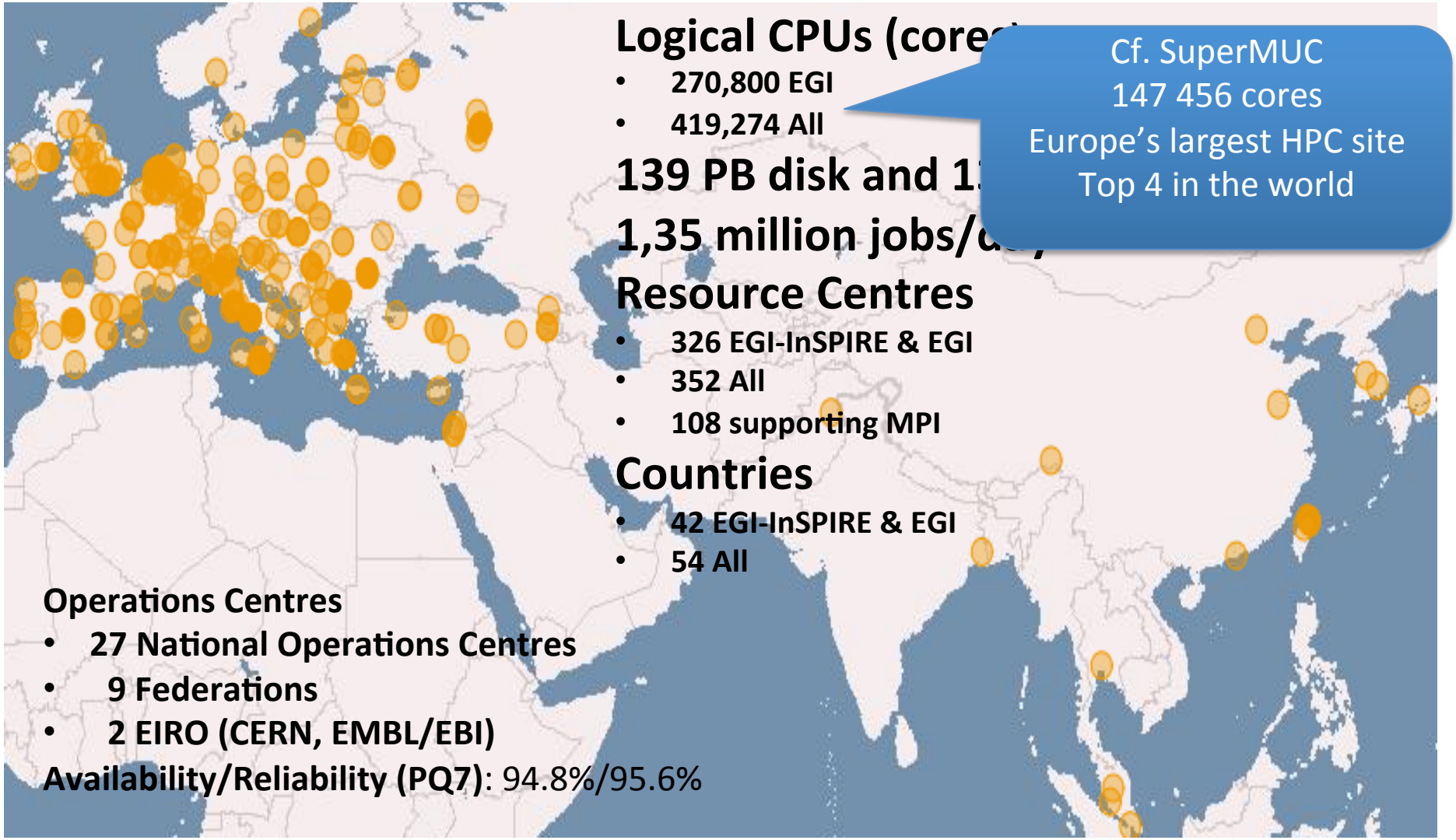
Resource Infrastructure Providers



Resource Centres	EGI-InSPIRE & EGI Council members	326
	Including integrated RPs	352
Countries	EGI-InSPIRE & EGI Council members	42
	Including integrated RPs	54

Integrated EGI-InSPIRE Partners and EGI Council Members
Internal/External Resource Providers (being integrated)
External Resource Providers (integrated)
Peer Resource Providers

European Grid Infrastructure (June 2012)



Logical CPUs (core
- **270,800 EGI**
- **419,274 All**

139 PB disk and 1

1,35 million jobs/d

Resource Centres
- **326 EGI-InSPIRE & EGI**
- **352 All**
- **108 supporting MPI**

Countries
- **42 EGI-InSPIRE & EGI**
- **54 All**

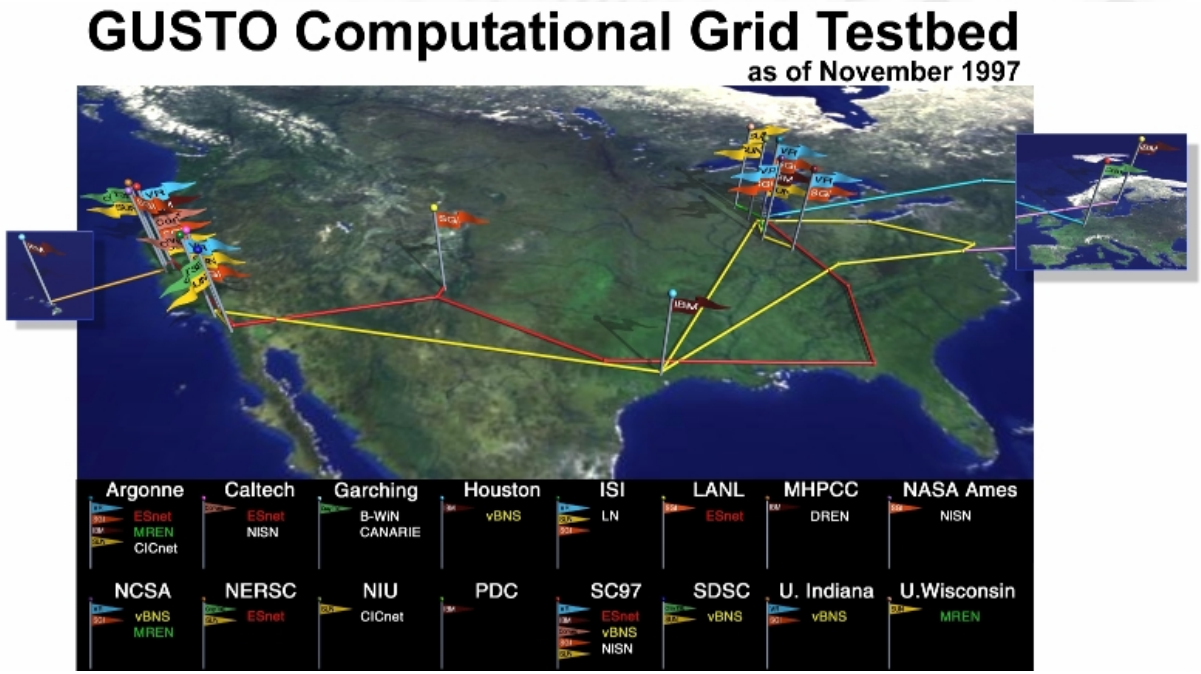
Cf. SuperMUC
147 456 cores
Europe's largest HPC site
Top 4 in the world

Operations Centres
- **27 National Operations Centres**
- **9 Federations**
- **2 EIRO (CERN, EMBL/EBI)**

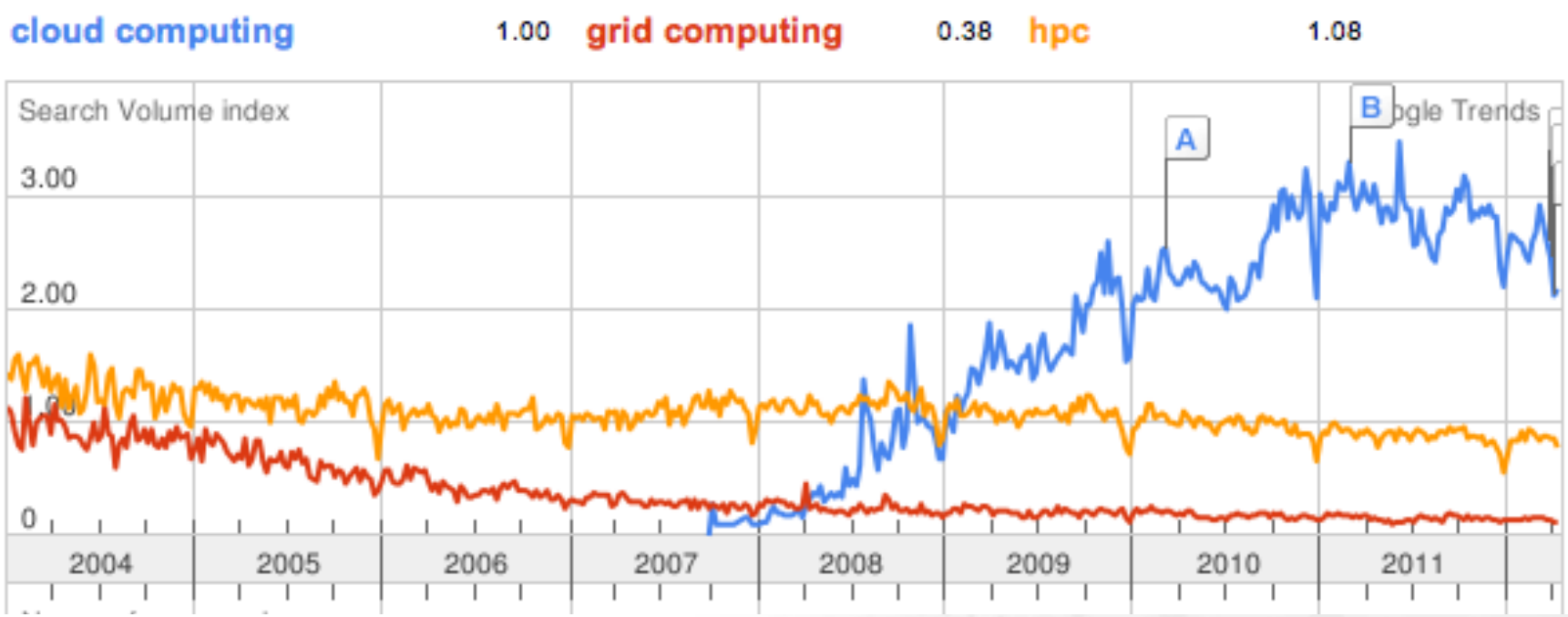
Availability/Reliability (PQ7): 94.8%/95.6%

A Little Grid History

- Grid computing
  - Since more than a decade
  - Since more than two decades as meta-computing
  - Origin in HPC needs
  - Early tools for parametric studies -> HTC
- e.g. Condor
  - popular framework for HTC
  - Industry & academy



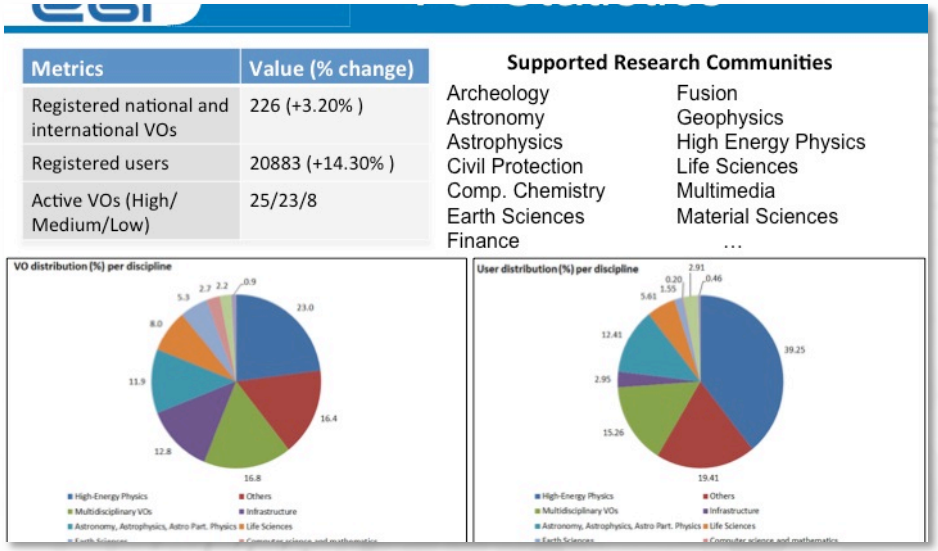
Lesson 1: What goes up…



Caveat: No validation of presented curves

Is Cloud Computing a Game Changer for HTC?

- 226 VOs in EGI
  - Very different size, scope, and funding
  - Mix of resource providers (also within VOs)
    - Univ labs, computing centres,…
  - Move from capex to opex not obvious



Metrics	Value (% change)
Registered national and international VOs	226 (+3.20%)
Registered users	20883 (+14.30%)
Active VOs (High/ Medium/Low)	25/23/8

Supported Research Communities

Archeology
Astronomy
Astrophysics
Civil Protection
Comp. Chemistry
Earth Sciences
Finance
Fusion
Geophysics
High Energy Physics
Life Sciences
Multimedia
Material Sciences
…

Maybe for specific users…

Then (Feb 2012)…

Bioinformatics Startup Scale Genomics Touts Flexibility as Key Differentiator of Cloud Platform

February 03, 2012

Like 1 Tweet 10 +1 1 Share 0

Type size: Email Printer-friendly version RSS Feed

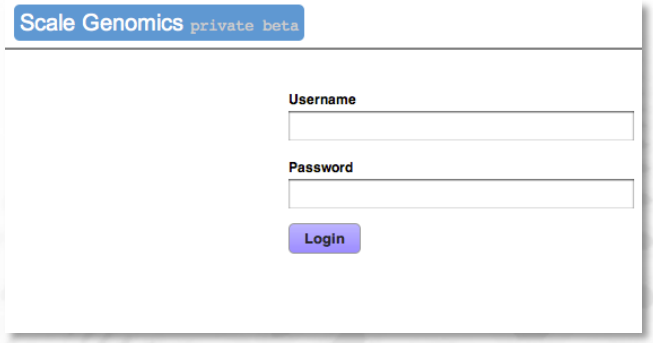
By Uduak Grace Thomas

Bioinformatics startup Scale Genomics is prepping to roll out later this month a new cloud-based storage and analysis system for next-generation sequencing data.

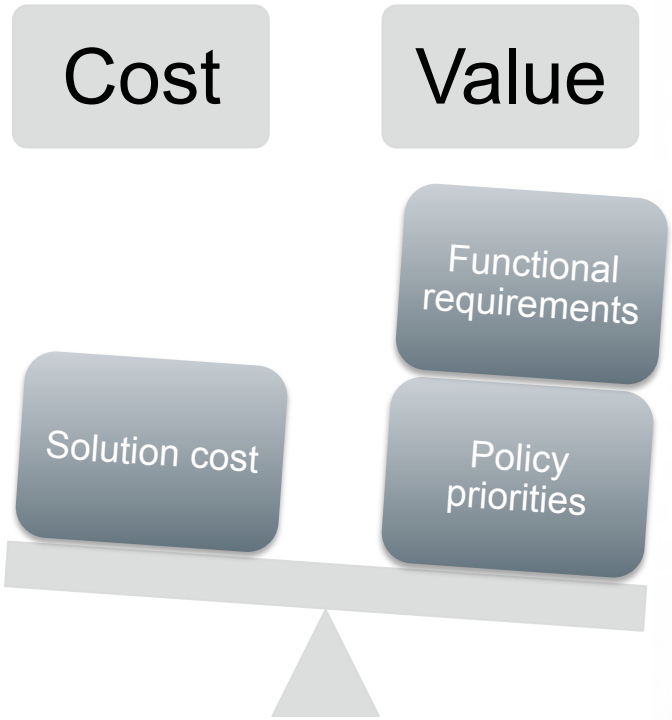
The platform is targeted at customers who have some bioinformatics expertise but who do not have access to the compute resources needed to run their analysis, Dmitri Petrov, co-founder of the company and professor of biology at Stanford, told *BioInform*.

When it launches its private cloud platform later this month, Scale Genomics will compete indirectly with another Stanford bioinformatics spin-out, DNAnexus, in the sense that both companies offer a cloud-based infrastructure for analyzing next-gen sequence data.

Now (Jul 2012)…



Public non-profit vs. Commercial



Lesson 2: There are no free lunches

$4,829-per-hour supercomputer built on Amazon cloud to fuel cancer research

By Jon Brodkin | Published 3 days ago

- 51 132 cores
- 9.4¢ per core hour

- Rough comparison: 800USD/h (capex) + 100% (opex) = 1600 USD/h to buy and operate a similar cluster =>
- ¿14M USD/yr for a 51 000 core cluster?

Lesson 2: There are no free lunches (cont.)

- EGI-InSPIRE total cost ~330M€, an estimate based on
  - 4 years
  - 170 000 cores (now 270 800)
  - Effort, power and renewal of HW included
- 0.055 € per core hour (opex + x% capex)
- Cf. Amazon cloud cost
  - 0.094 $ per core hour (0.075 €)

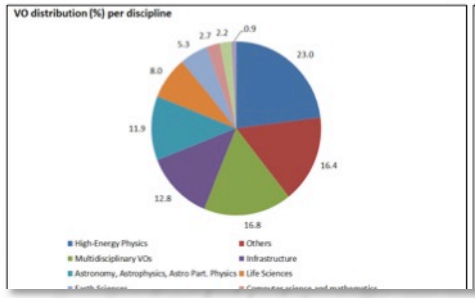
Is Cloud Computing a Game Changer for HTC?

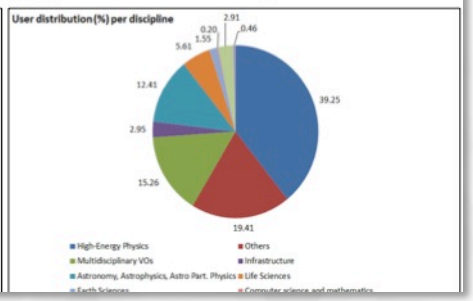
- 226 VOs in EGI
  - Very different size, scope, and funding
  - But, for each VO collaboration is key

Metrics	Value (% change)
Registered national and international VOs	226 (+3.20%)
Registered users	20883 (+14.30%)
Active VOs (High/ Medium/Low)	25/23/8

Supported Research Communities

Archeology
Astronomy
Astrophysics
Civil Protection
Comp. Chemistry
Earth Sciences
Finance
Fusion
Geophysics
High Energy Physics
Life Sciences
Multimedia
Material Sciences
…



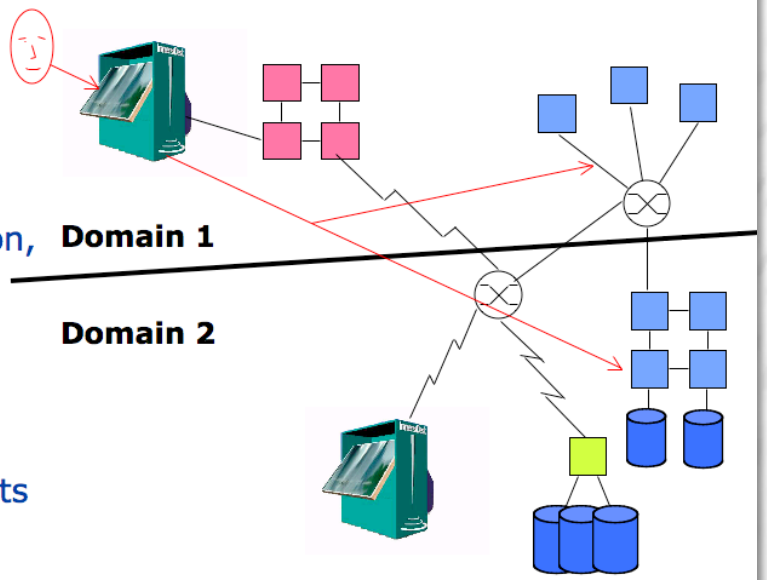


Lesson 3: Collaboration requires horizontal infrastructure

the globus project
www.globus.org

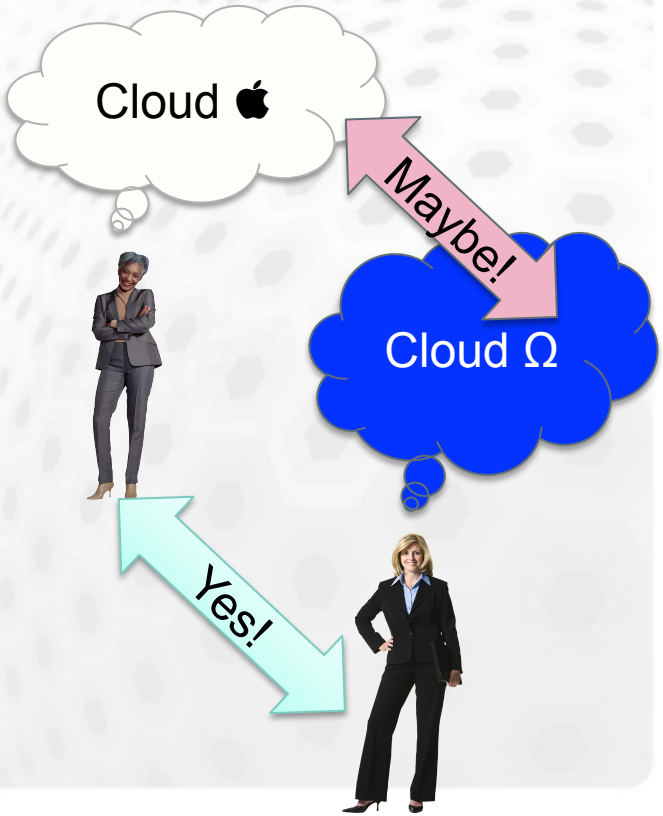
Issues

- Authenticate once
- Specify simulation (code, resources, etc.)
- Locate resources
- Negotiate authorization, acceptable use, etc.
- Acquire resources
- Initiate computation
- Steer computation
- Access remote datasets
- Collaborate on results
- Account for usage



User Tutorial (1.1.3) Introduction 6

Cf. Cloud Computing



Lesson 4 to 42

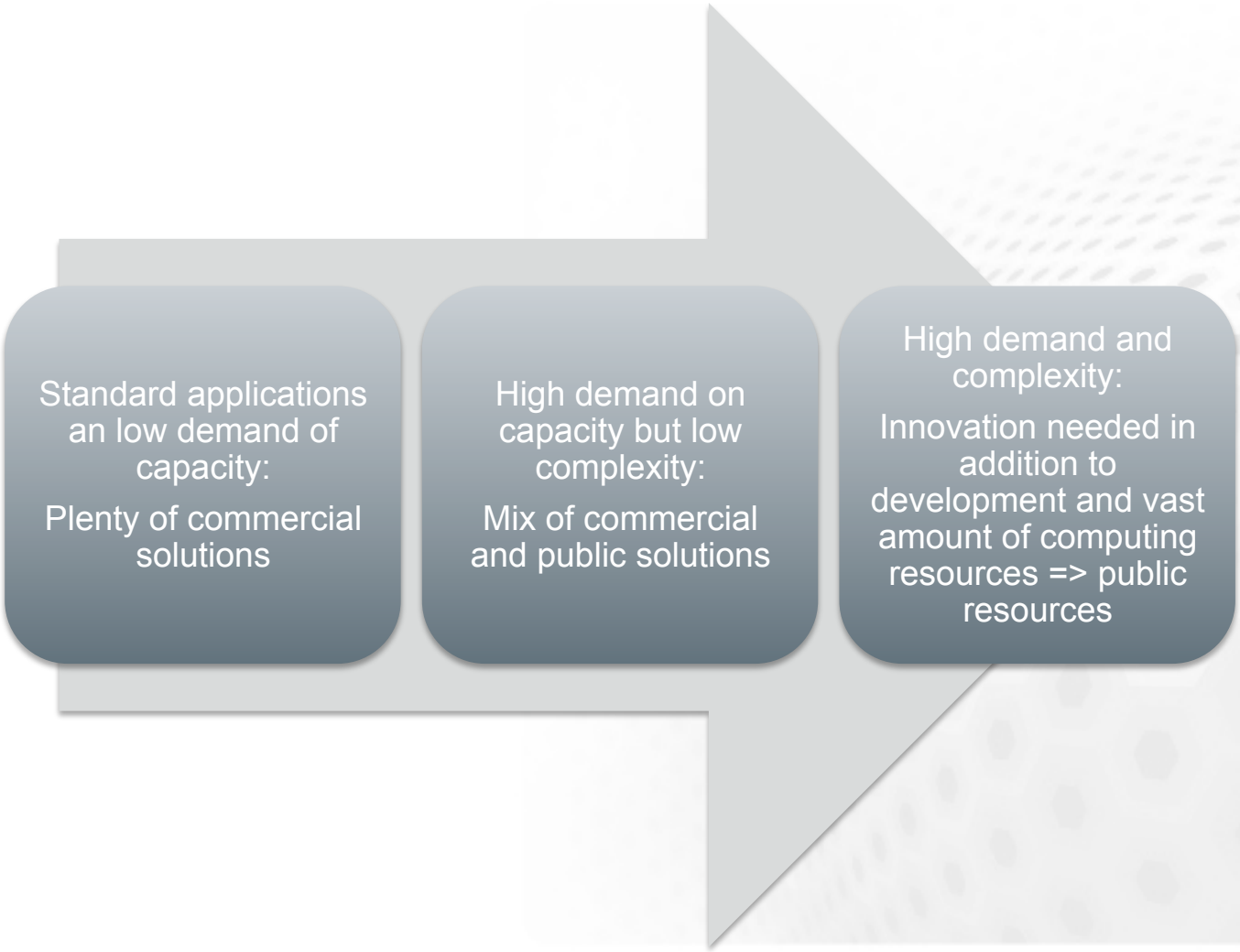
To be continued…

But, maybe a conclusion of the three first lessons:

Is Cloud Computing a Game Changer for HTC?

- My prediction:
  - Governments will not stop to invest in computing resources, operations and support, i.e. move CAPEX to OPEX
  - But, we will see a mix…

Commercial vs. Public non-Profit computing solutions applicability for research



BTW, I didn't say anything about sustainability…

- EGI can be considered a research infrastructure
- Sustainability of a research infrastructure is built from a number of factors
  - Value for researchers
  - Political priorities (societal value)
  - Funding streams
    - Governments
    - EC
    - Sales
  - A lot of work to realize the above

Acknowledgment

- Steven Newhouse and the EGI.eu staff, EGI.eu Amsterdam (EGI slides)