E-Infrastructure Development Trends in the Area of Grids, Clouds, HPC, Storage, Virtualization and IaaS

Peter Kacsuk, MTA-SZTAKI,
kacsuk@sztaki.hu

Peter Stefan, NIIFI,
stefan@niif.hu

Imre Szeberenyi, BME,
szebi@iit.bme.hu

Overview

- Background
- Infrastructure in a nutshell
- Scientific computing on desktop grids
- Scientific computing on clouds
- NIIFI Cloud
- Summary

Background

- Message: how to serve scientific computing with compute and data storage infrastructure in more efficient ways?
- Scientific computing is the quintessence of IT:
  - it is extensive, ranges from low level hw to higher level applications;
  - it is present in the history of IT since the very beginning (Neumann – solving differential equations);
  - its purpose is noble: find answers to scientific questions (cancer research, climate modeling, car industry…).
- Our purpose: to give suggestions on the answers and to consider about future trends in this field.

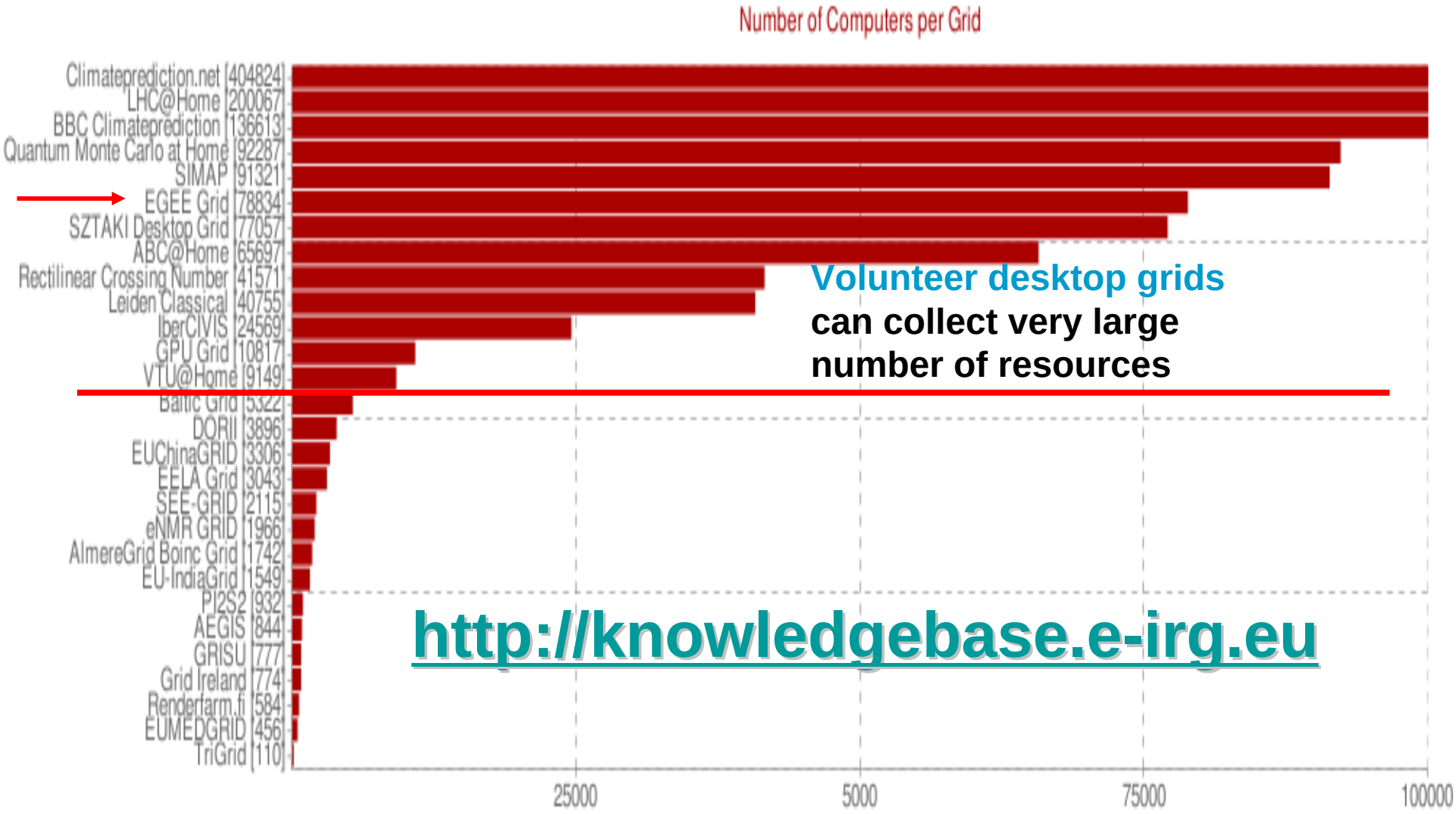
Infrastructure in a nutshell

- Single computers
- Parallel computers, vector computers
- Mainframes, metacomputers, geographically centered clusters,
- Data/resource centers
- Supercomputers
- Grids
- Clouds – Services

All are about serving the science!

Scientific Computing on Desktops

Why volunteer desktop grids are important?



Potential of volunteer computing

- Number of privately owned PCs:
  - currently 1.5 billion
  - grow to 2 billion by 2015
- **~1 million computers** are actively participating in volunteer computing → supplying ~10PetaFLOPS of computing power
- Fastest supercomputers supply few PetaFLOPS
- Largest grids (e.g. EGI) have several hundreds of thousands of hosts.
- In the area of so-called bag of tasks, parameter sweep applications volunteer computing is competitive.
- Near-term potential of volunteer computing goes well beyond Exa-scale.

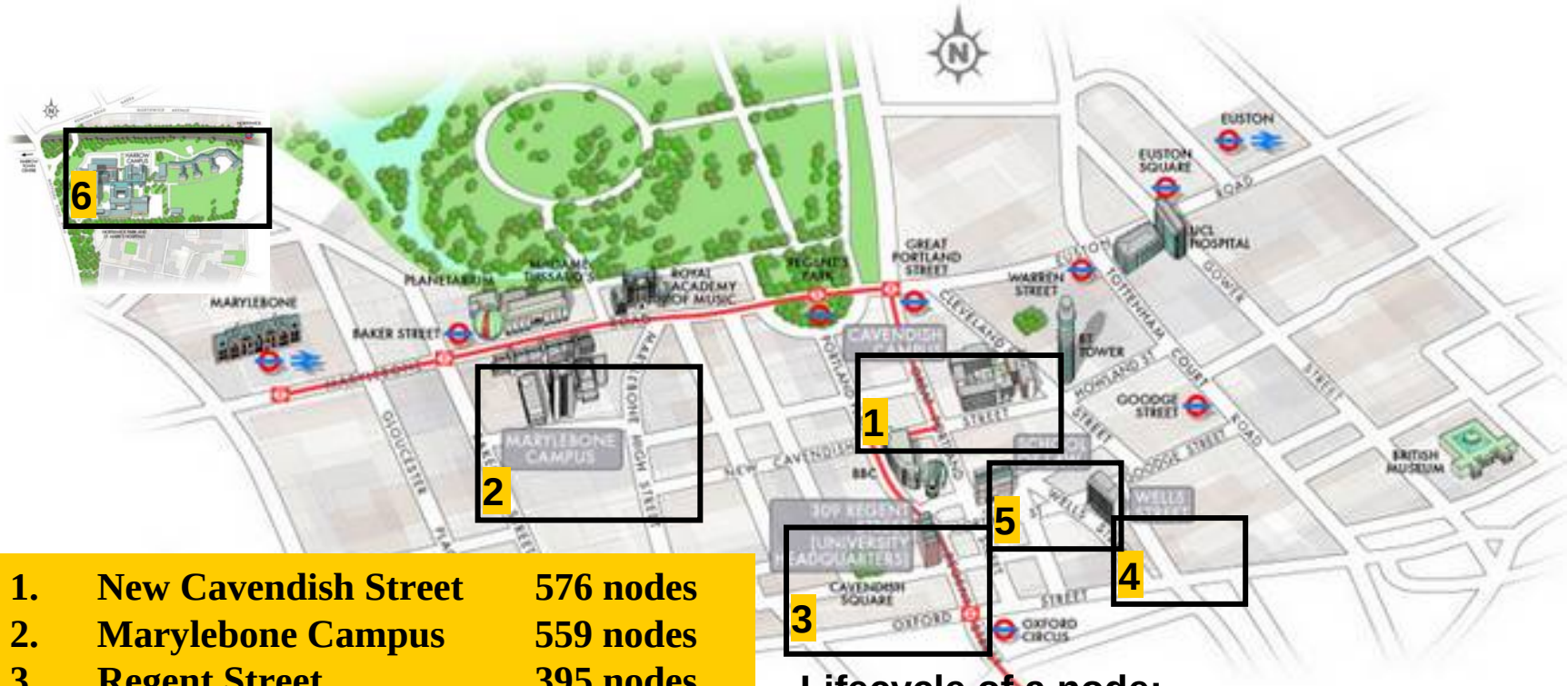
Types of Desktop Grids

- **Global (volunteer) Desktop Grid**
  - Aim is to collect resources for grand-challenge scientific problems.
- Examples:
  - SETI@home, Folding@home, Shakemovie@home, LHC@home
  - Community World Grid, IberCivis, SZTAKI Desktop Grid.
- **Local (institutional) Desktop Grid**
  - Aim is to enable the quick, easy and **inexpensive** creation of grid for any community (company, university, etc.) to solve their own applications.
- Example:
  - SZTAKI Desktop Grid (SZDG) local version (used within EDGeS, EDGI, DEGISCO).

University of Westminster

MTA SZTAKI

Local DGs in practice – University of Westminster DG system based on SZDG

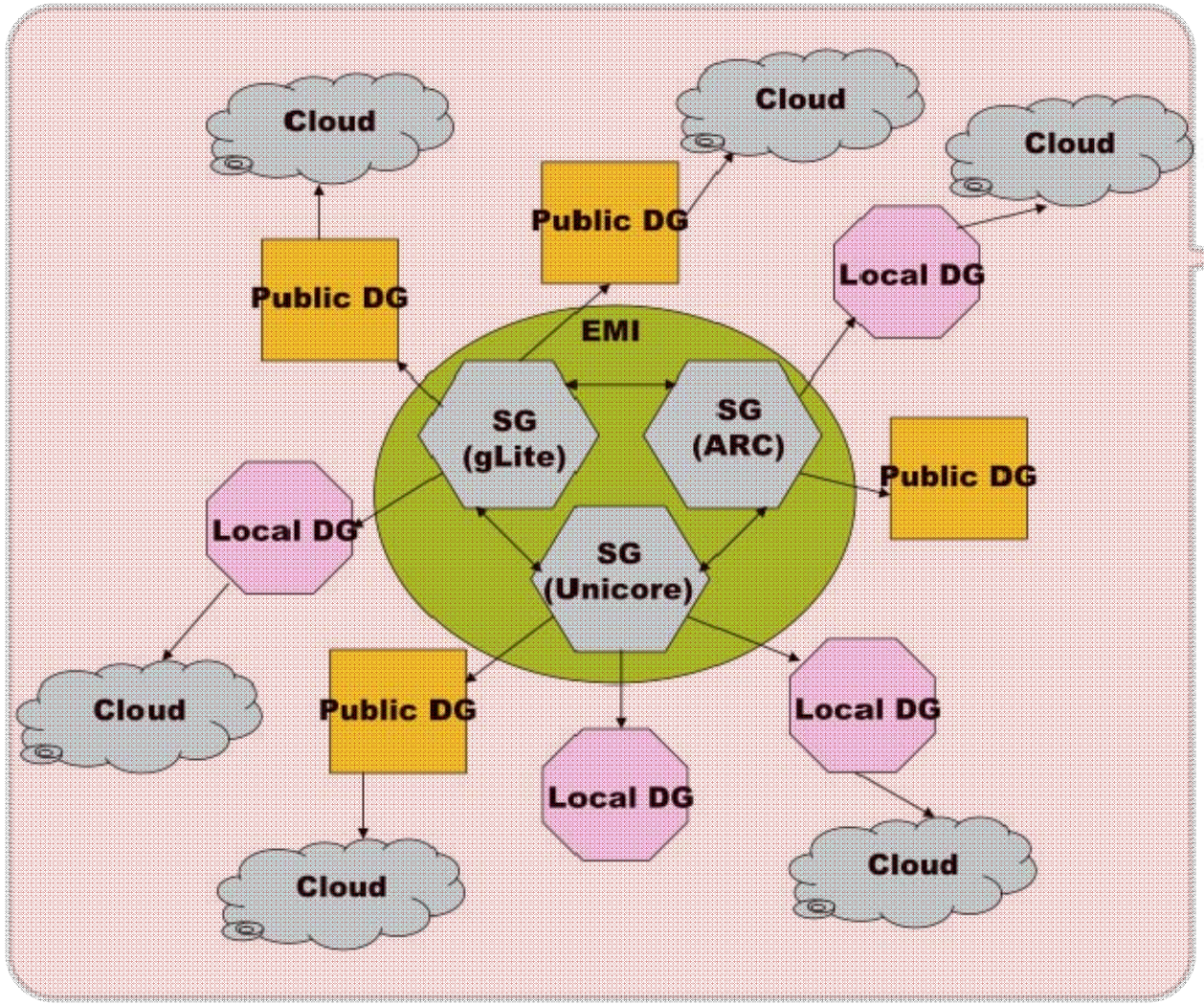


1.	New Cavendish Street	576 nodes
2.	Marylebone Campus	559 nodes
3.	Regent Street	395 nodes
4.	Wells Street	31 nodes
5.	Little Tichfield Street	66 nodes
6.	Harrow Campus	254 nodes
Total:		1881 nodes

Lifecycle of a node:

1. PCs basically used by students/staff
2. If unused, switch to Desktop Grid mode
3. No more work from DG server -> shutdown (green solution)

The EDGI Infrastructure



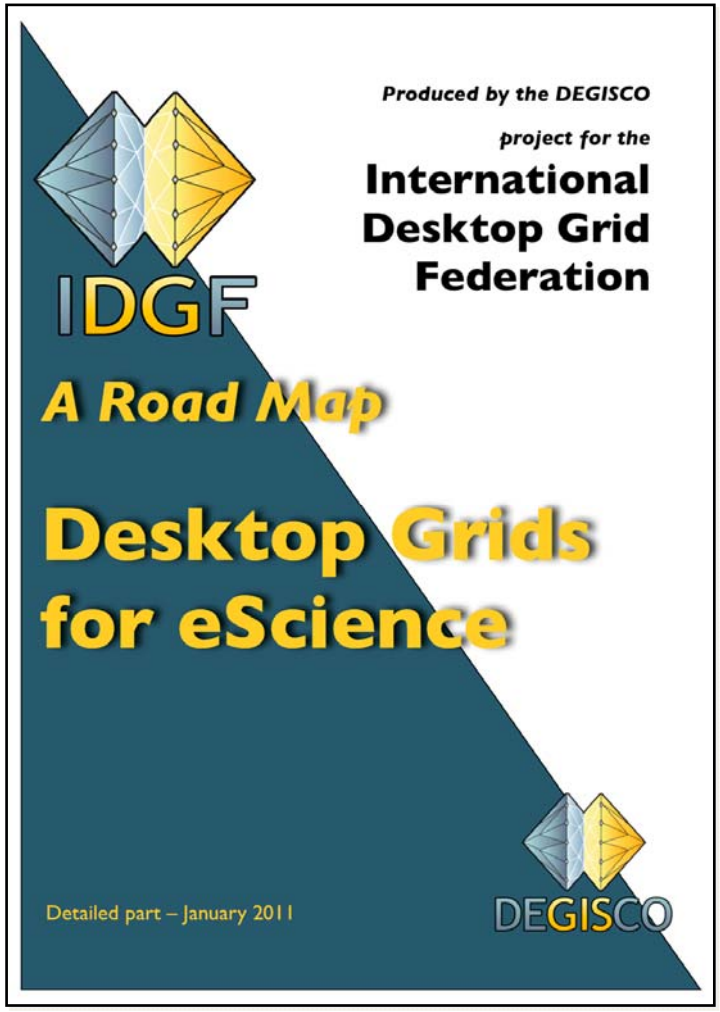
Extends EMI/EGI (gLite, ARC, Unicore) grids with DGs

Extend Desktop Grids with Clouds for QoS

Supports both compute and data intensive applications


WEB2GRID
Nemzeti Technológia Program
DocMark4WEB2 on Facebook - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://apps.facebook.com/docmark/?ref=bookmarks&count=0
EGI-InSPIRE
DocMark4WEB2 on Facebook
facebook
Search
myphotos machine
Logout
?
Select
Queue
Finished job
Photos
The following photos were uploaded to be watermarked:
Hungarian project to integrate the advantages of Web2 and DGs
The selected photos have to be paid, then they are forwarded to execute watermark embedding. The running time of procedure depends on the size of selected photo and parameters of watermarking algorithm (between 10 minutes and 100 hours). You can display the result of watermarked photos on the Finished job screen.
Tick the checkbox if you would like to protect this photo by embedding invisible watermark and creating electronic signature!
All unchecked photos can be submitted later!
Protect it for 1.00 EGC!
Tick the checkbox if you would like to protect this photo by embedding invisible watermark and creating electronic signature
All unchecked photos can be submitted later!
Protect it for 1.00 EGC!

Desktop grids for eScience - A Road map



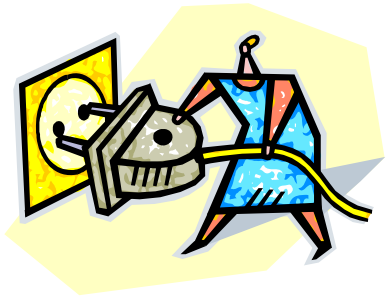
- Helps organisations in deciding on setting up a Desktop Grid.
- Linking it into eScience. infrastructures: EGI, Clouds, Grids.
- Advice on all levels from technical to legal and political.

Source: http://europa.eu/volunteering

Scientific Computing on Clouds

Scientific Computing on Clouds

- Cloud is invented by the business area.
  - Main goal is: selling everything as services
- Clouds – still developing
  - SaaS (Google Docs, app. portals);
  - PaaS (web services, grid, databases, HA clusters);
  - IaaS (virtualized resources).



- Furthermore the three above are not layers!
- In SC we tend to take the IaaS definition.
- European initiative: Stratuslab.

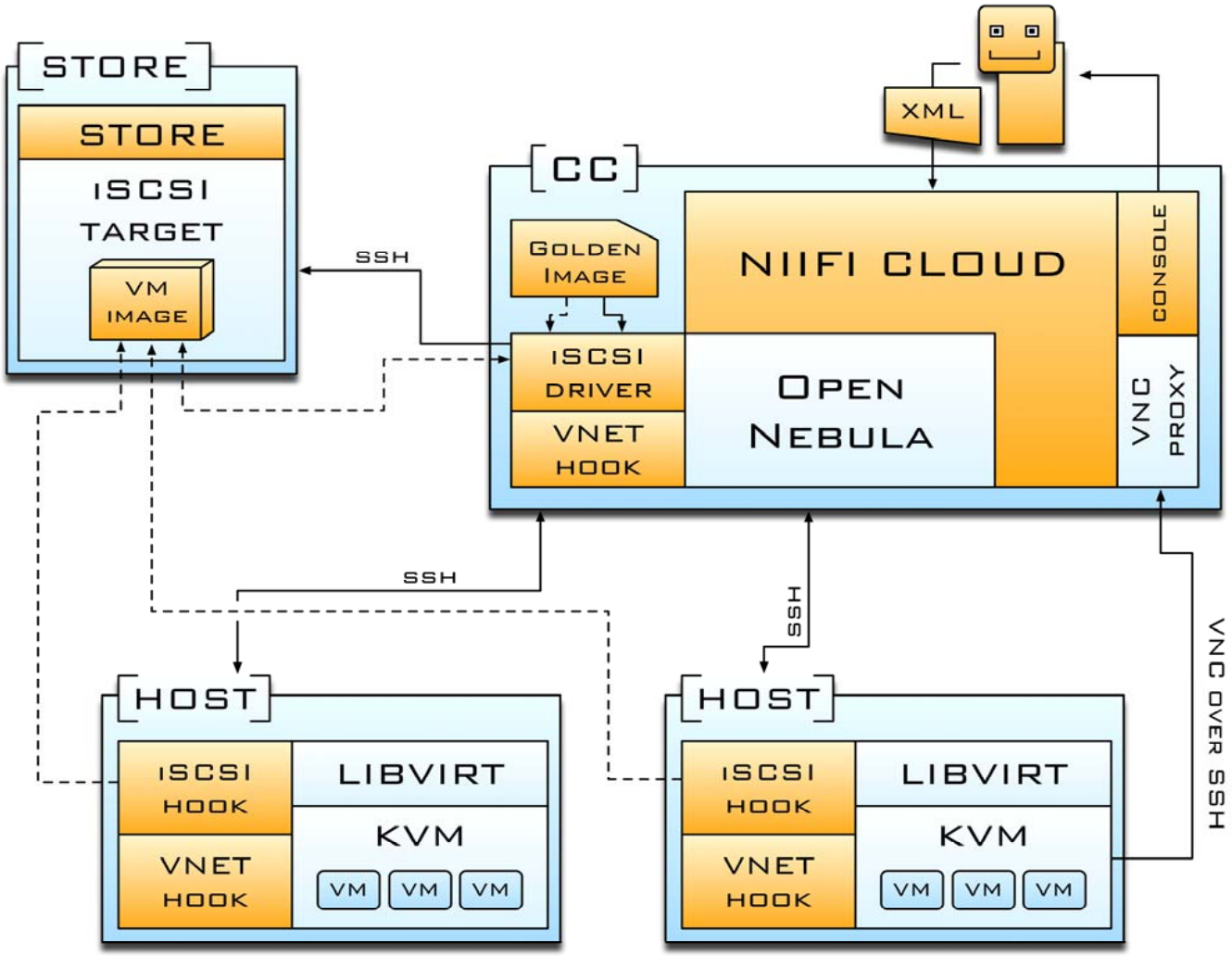
IaaS Benefits in Scientific Computing

- IaaS = decoupling from hw + managing OS images.
- Debate over the benefits here: pure hw services vs. virtualized resources.
- There are some obvious:
  - Better utilization of hw resources, more even spread of load as well as functionalities;
  - Lower price, lower sustainability costs, lower TCO;
  - Gradual scalability;
  - User ownership and involvement via better sandboxing;
  - Better customization in the sandbox.
- NIIF Cloud.

NIIF Cloud

- Designed to be a private cloud that can be transformed to public.
- Mostly for NIIFI services.
- Builds upon OpenNebula + KVM + libvirt. Significant extensions.
- OS images are stored on storage infrastructure and are served over iSCSI protocol.
- Geographically redundant layout:
  - Servers are located in provincial regional centers;
  - Interconnected by high quality data network.

NIIF Cloud Architecture



NIIF Cloud Features

- Virtual machines and a set of virtual machines interconnected by virtual networks.
- L2 connectivity and L2 separation of networks.
- Live migration of virtual machines between sites (machines, networks, images).
- Guaranteed CPU performance for virtual machines.
- Web based console.
- Grid and cluster software stack over the virtual clusters (ARC+SGE).

Future Cloud Trends

- The purpose of IaaS and scientific computing is different.
- Yet, they can coexist on the infrastructure.
- From the IaaS viewpoint: grid is a powerful tool connecting geographically distant sites and allow a specific service on it.
- From the grid viewpoint: IaaS cloud pushes into the grid software stack below the operating system level, and facilitates easy job migration, better resource utilization and better computing environment customization (support for multiple operating systems).

Summary

- Infrastructure solutions for serving scientific computing has been presented:
  - Desktop PC integration via SZTAKI DG;
  - Virtualization via NIIF Cloud;
- Future of grids and clouds.

Thanks & Questions