



Big Science, Grid and NDGF

Josva Kleist

Software Coordinator, NDGF

Aalborg University, November 6th, 2008



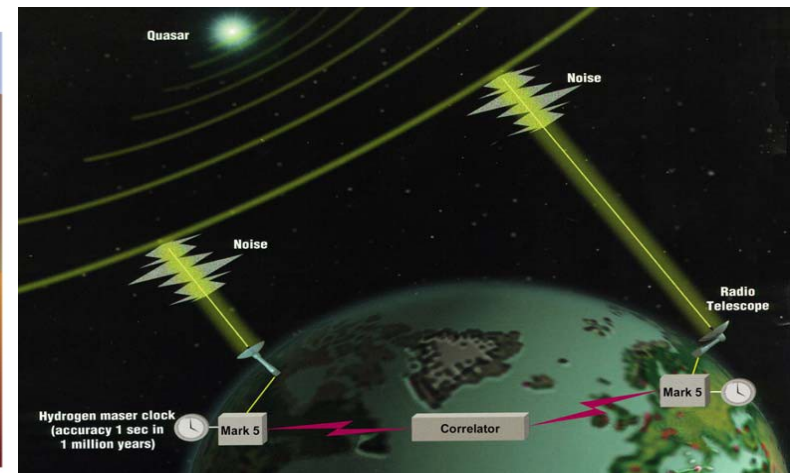
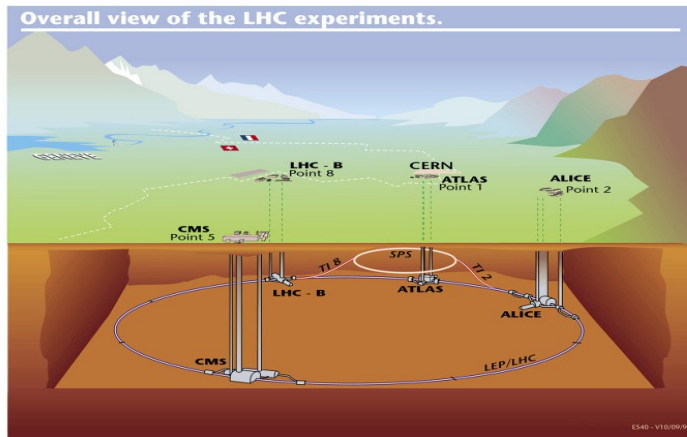
Watch the movie at:

www.ndgf.org

- Tycho Brahe made a sky survey late in the 16th century
- It listed planets with their orbit
- A huge effort with vast amounts of data
- Kepler did analysis on these data
- Found out about elliptic orbits
- Separation of:
 - Data Collection
 - Data Analysis

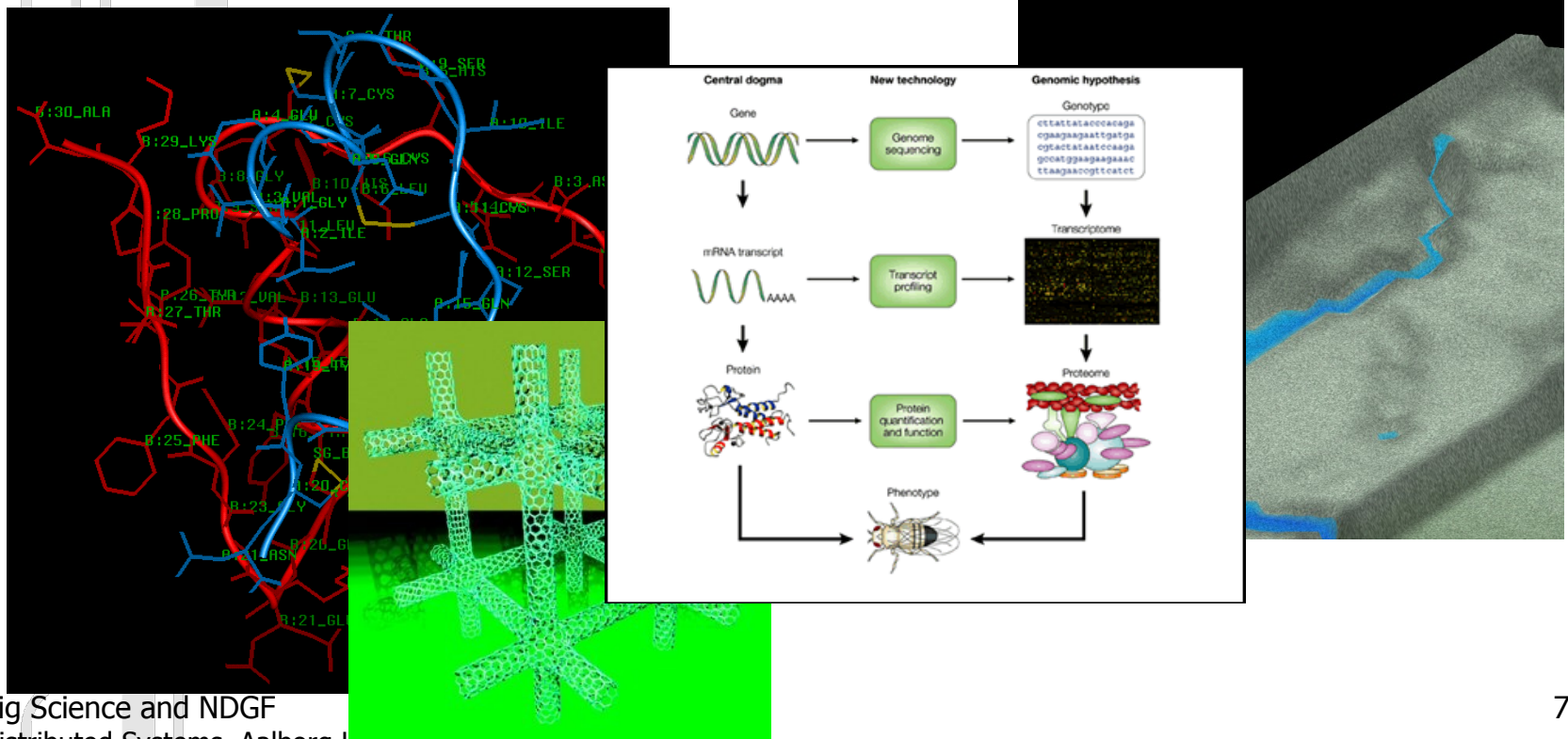


- Data Collection projects
 - Sloan Digital Sky Survey
 - Human Genome Project
 - The Icelandic screening of health records vs genes
 - e-VLBI
 - CERN: ALICE, ATLAS, CMS and LHCb
 - Others...



- Today's big science projects – eScience
- Grid computing – middleware for eScience
- eScience example: The Large Hadron Collider
- The Nordic DataGrid Facility
- NDGF: a Tier-1 for WLCG
 - Networking
 - Computations
 - Storage

”Science (increasingly) done through distributed global collaborations enabled by the Internet, using very large data collections, tera-scale computing resources and high performance visualisation.”

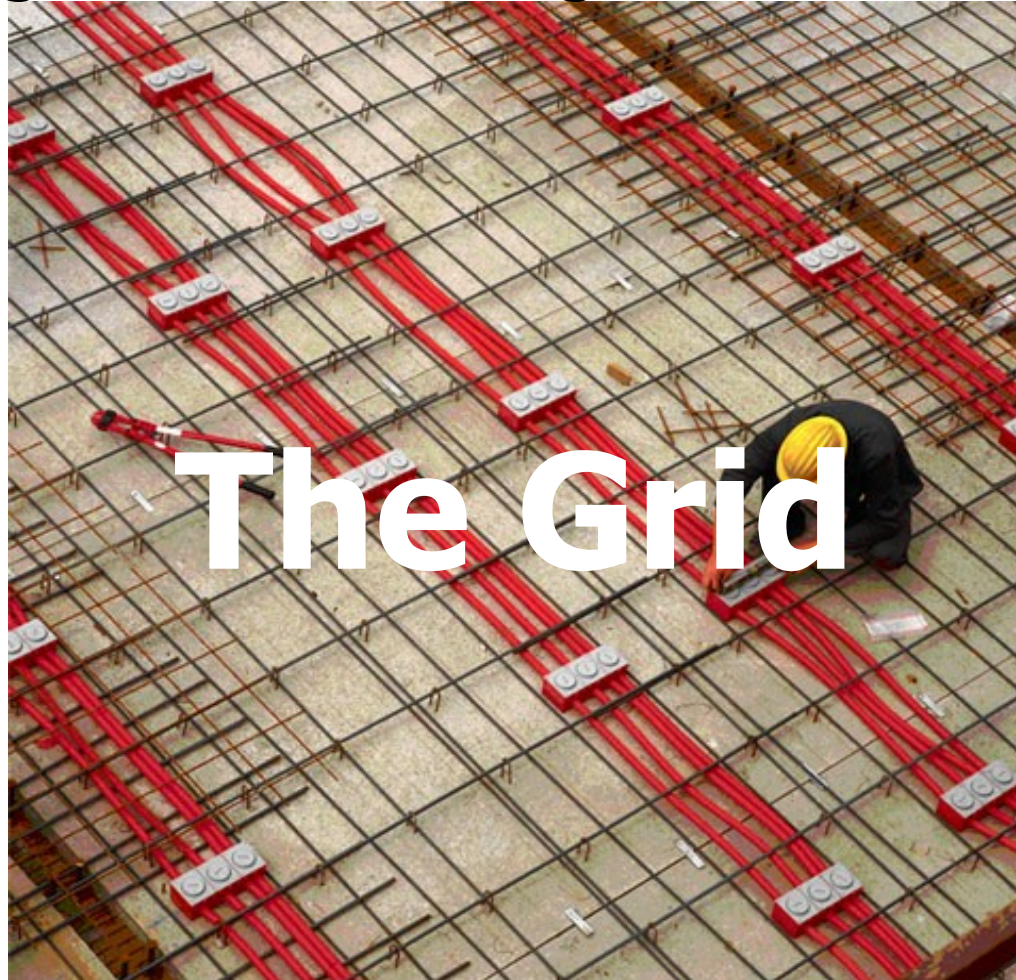




+



Data Storage and Processing



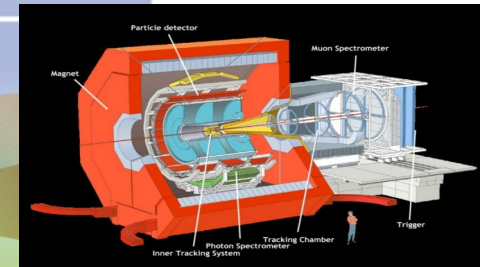
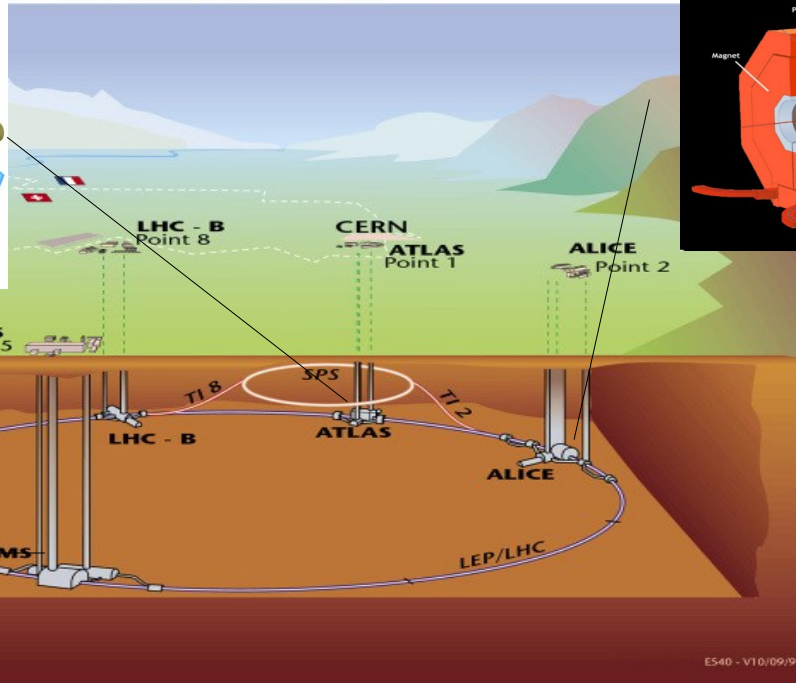
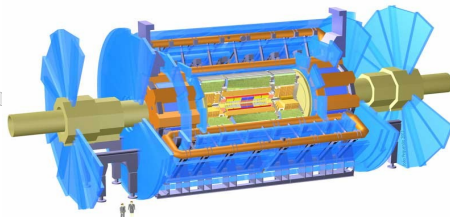
- An Infrastructure for:
 - Storing huge amounts of scientific data
 - Accessing distributed data
 - Processing huge amounts of data
 - High availability

- Connecting:
 - Users from multiple organizations
 - Resources from multiple resource providers
 - Multiple user groups

The Large Hadron Collider



Overall view of the LHC experiments.



CMS Collaboration
36 Nations, 160 Institutions, 2008 Scientists and Engineers (November 2003)

TRIGGER & DATA ACQUISITION
Austria, Czech Republic, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

PRECISION MONITORING
Austria, Czech Republic, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

RETURN YOKES
Belgium, Canada, France, Germany, Greece, Russia, Sweden, Switzerland, USA, Brazil

SUPERCONDUCTING MAGNETS
Austria, Czech Republic, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

PHOTON CALORIMETER
Austria, Czech Republic, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

FAST CALORIMETER
Austria, Czech Republic, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

MUON CHAMBERS
Belgium, Canada, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

TRIGGER
Austria, Czech Republic, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

CRYSTAL BALL
Austria, Czech Republic, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

SPS
Belgium, Canada, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

LEP/LHC
Belgium, Canada, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

ALICE
Belgium, Canada, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

ATLAS
Belgium, Canada, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

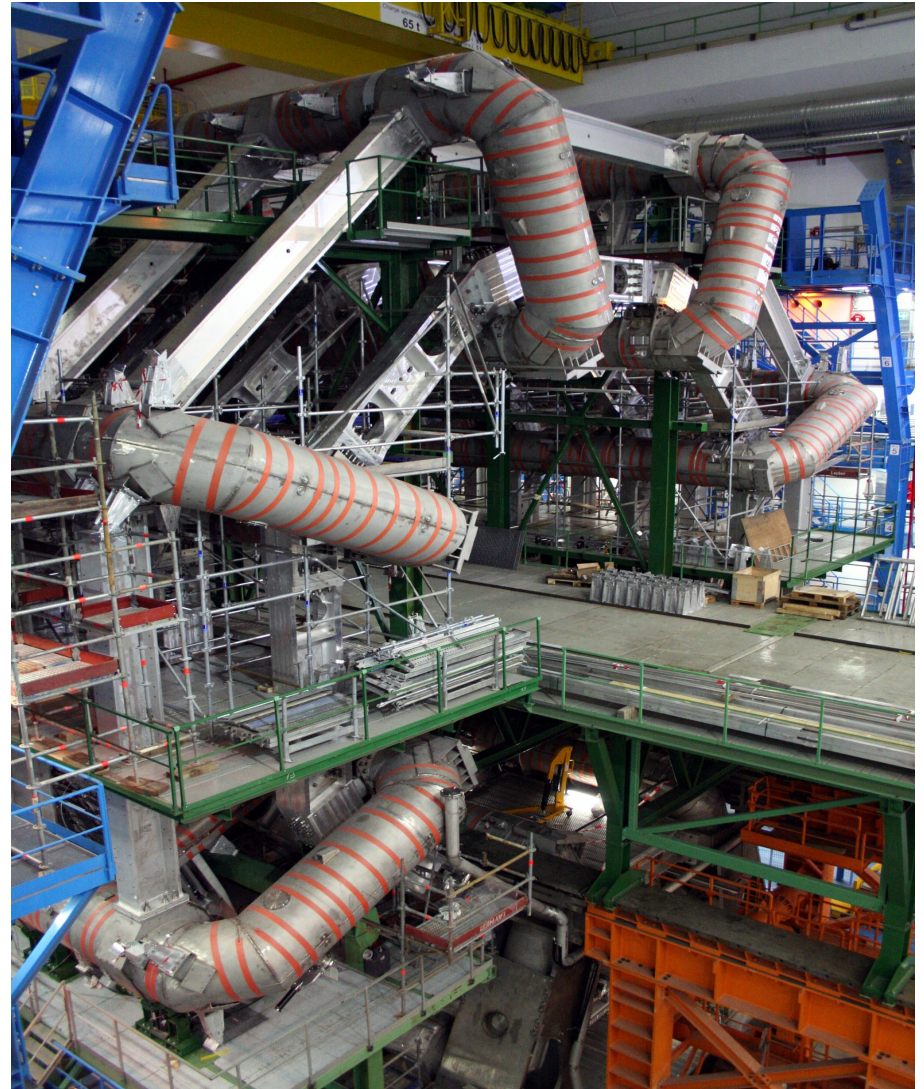
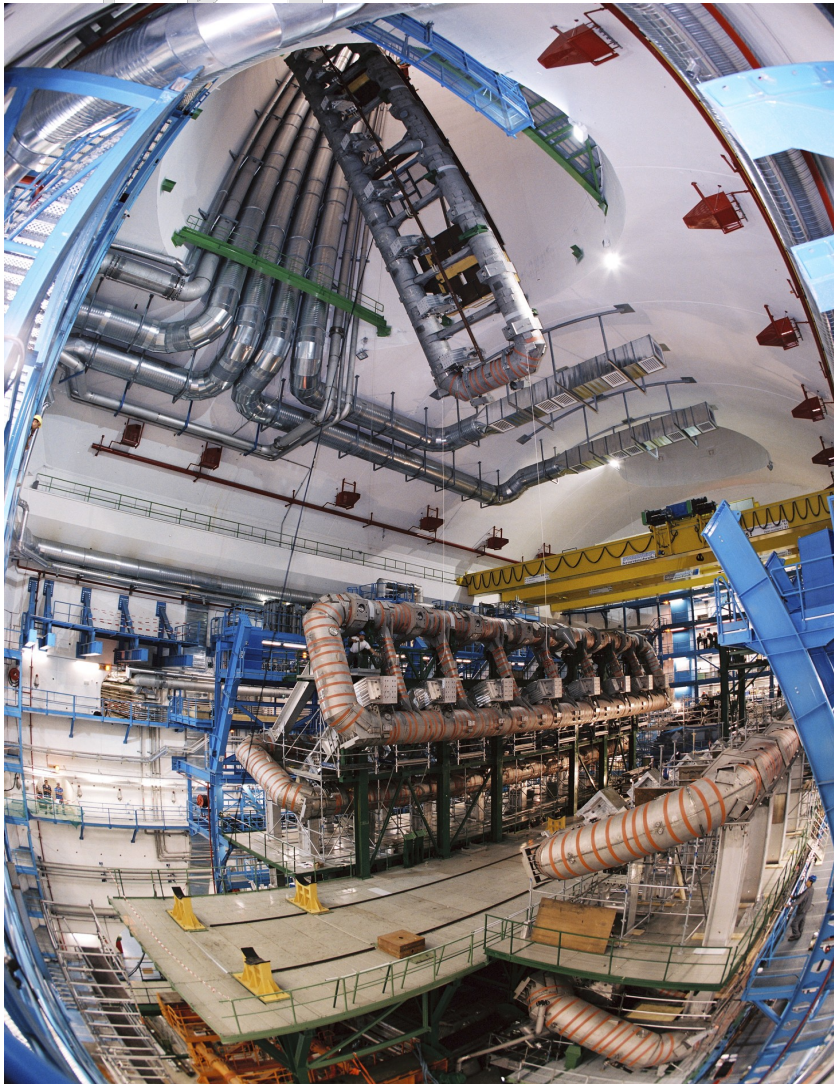
LHC - B
Belgium, Canada, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

CMS
Belgium, Canada, France, Germany, Hungary, Italy, Korea, Mexico, Poland, Spain, Switzerland, UK, USA

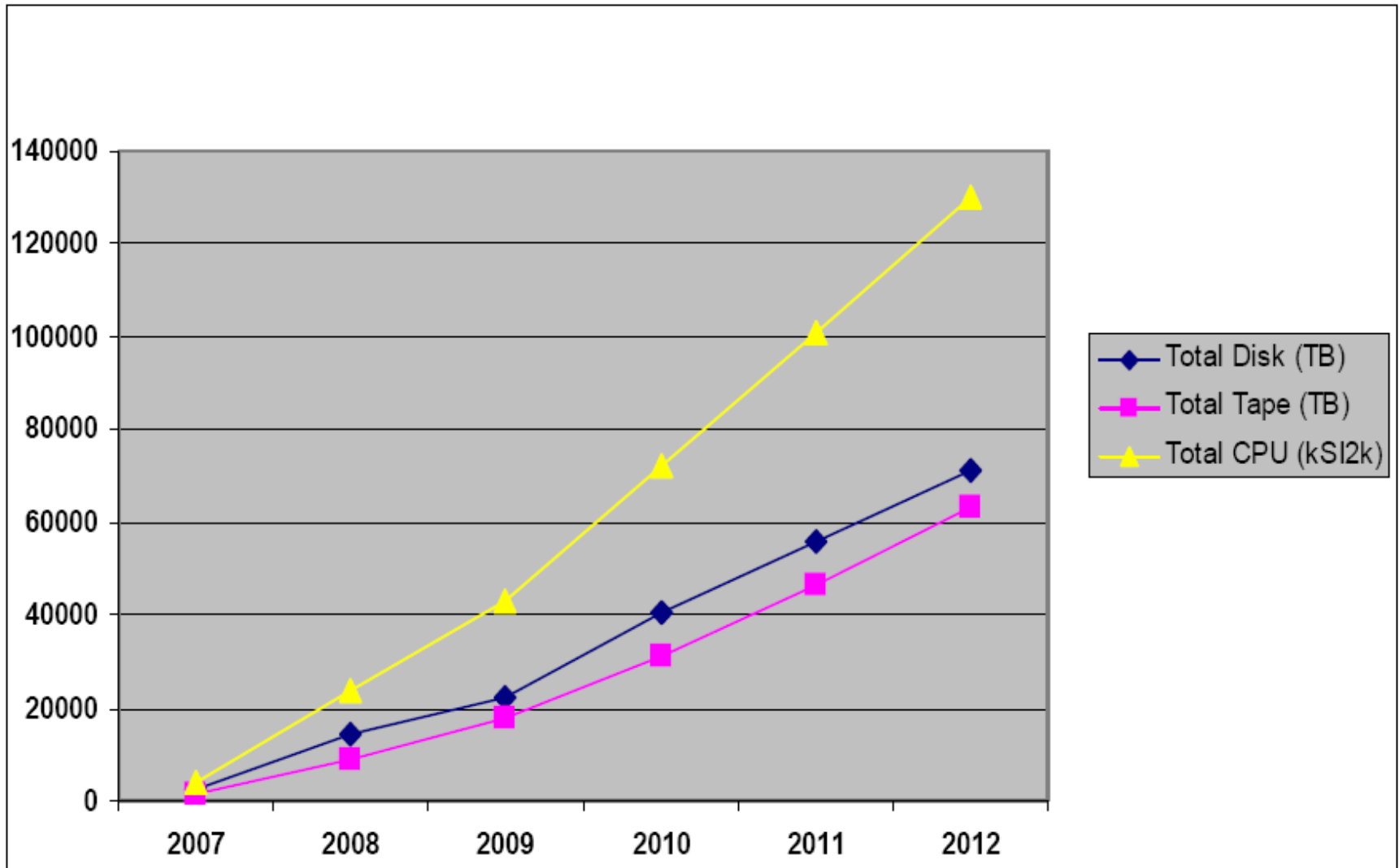
Summary:
Total weight: 12500 T
Overall diameter: 16.3 m
Overall length: 21.5 m
Magnetic field: 4 Tesla

E540 - V10/09/97

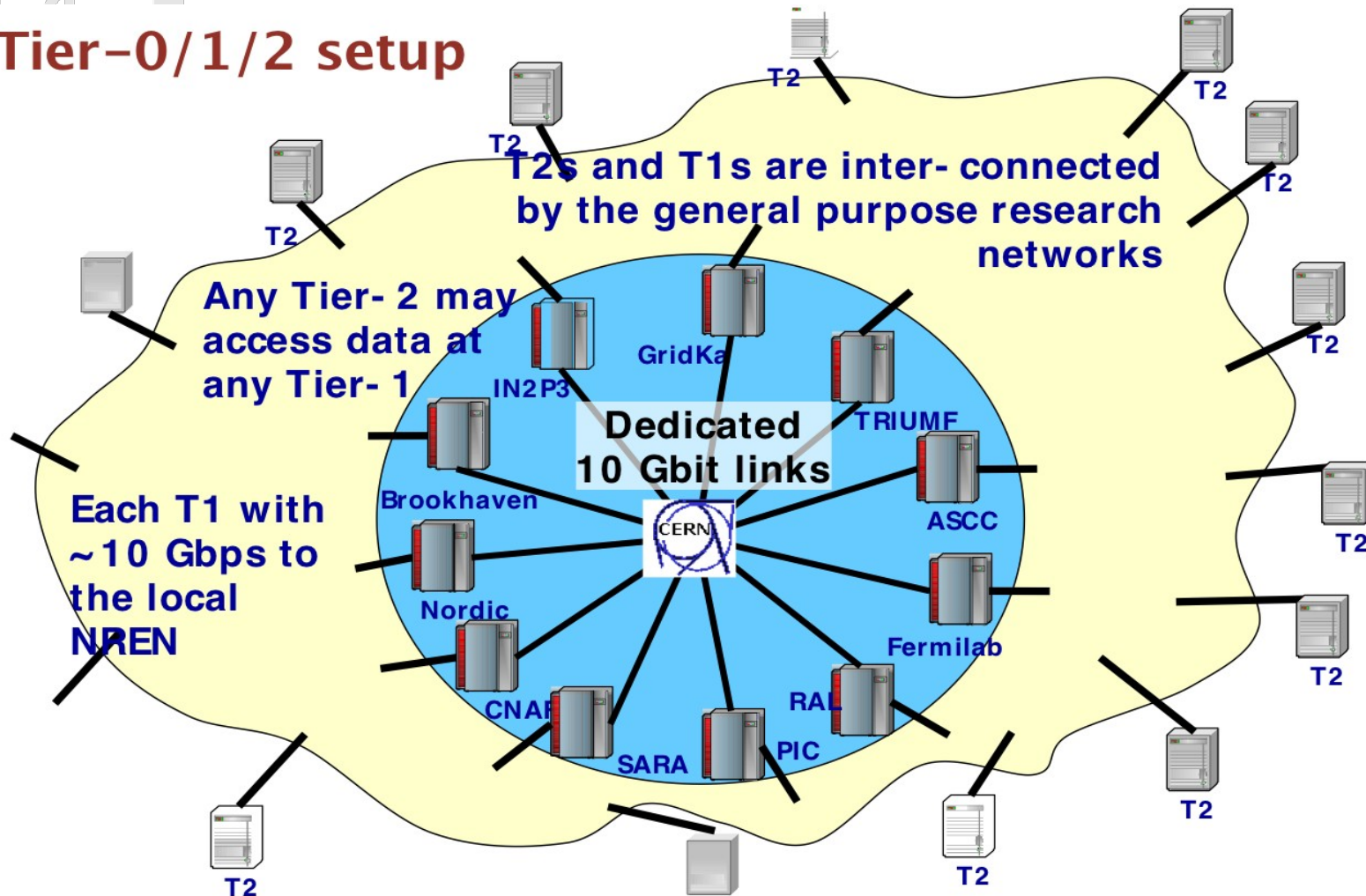




Big Science and NDGF
Distributed Systems, Aalborg University, November 6th 2008



Tier-0/1/2 setup



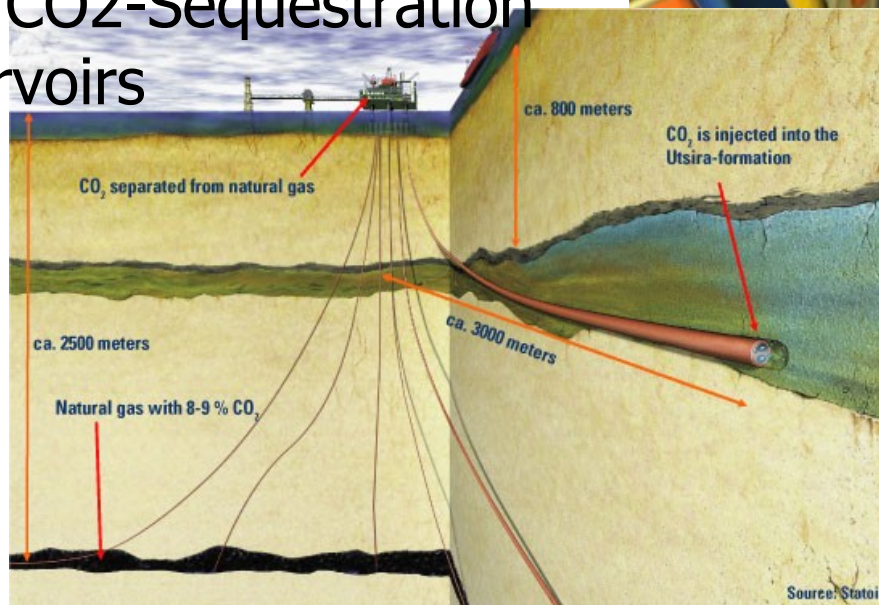
- Nordic Participation in *Big Science*:
 - WLCG – the Worldwide Large Hadron Collider Grid



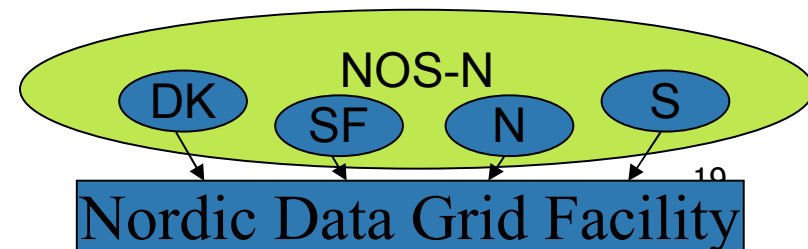
- Nordic Participation in *Big Science*:
 - WLCG – the Worldwide Large Hadron Collider Grid
 - BioInformatics



- Nordic Participation in *Big Science*:
 - WLCG – the Worldwide Large Hadron Collider Grid
 - BioInformatics
 - Screening of CO₂-Sequestration suitable reservoirs

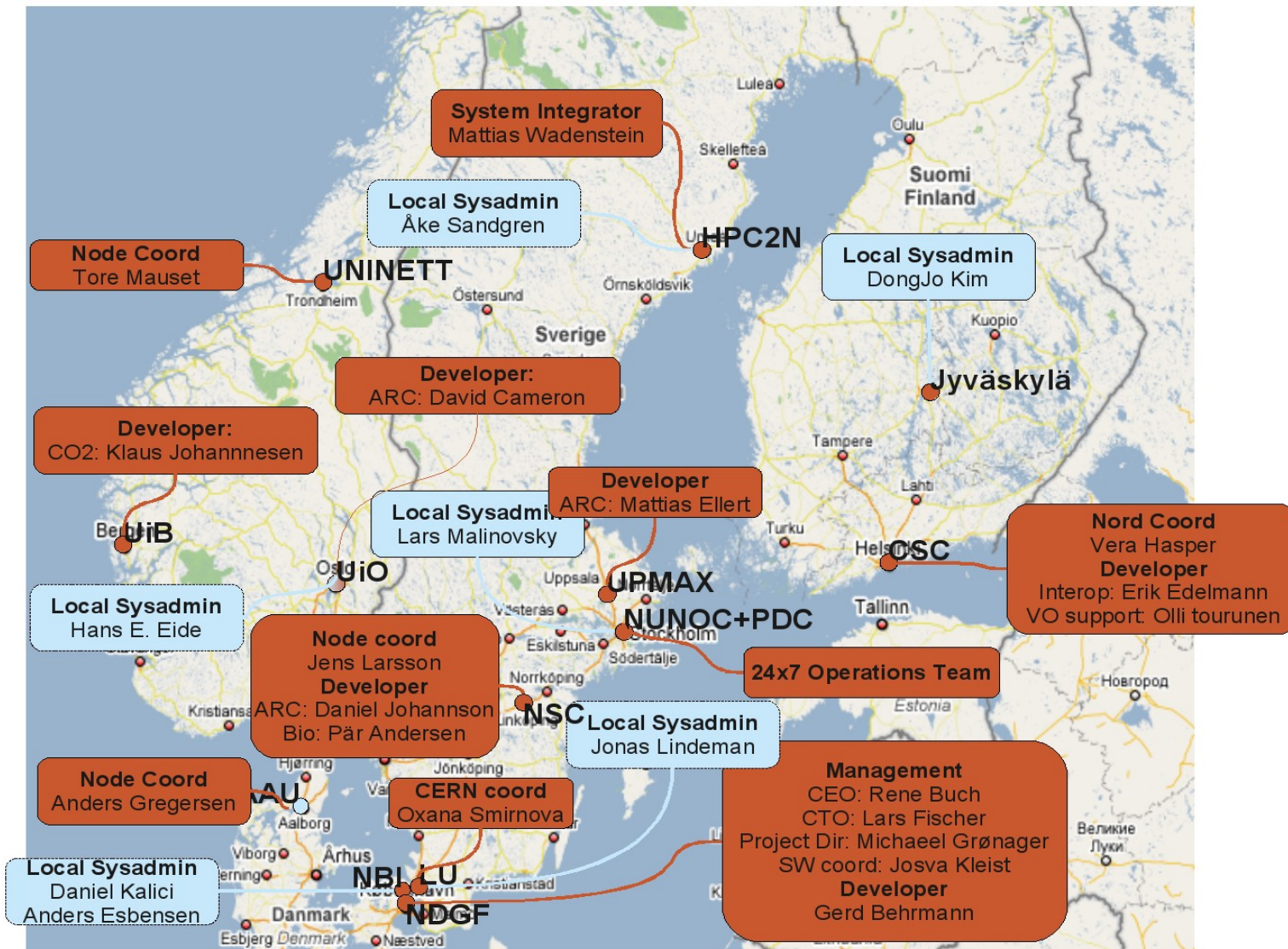


- A Co-operative Nordic Data and Computing Grid facility
 - Nordic production grid, leveraging national grid resources
 - Common policy framework for Nordic production grid
 - Joint Nordic planning and coordination
 - Operate Nordic storage facility for major projects
 - Co-ordinate & host major eScience projects (i.e., Nordic WLCG Tier-1)
 - Develop grid middleware and services
- NDGF 2006-2010
 - Funded (2 M€/year) by National Research Councils of the Nordic Countries

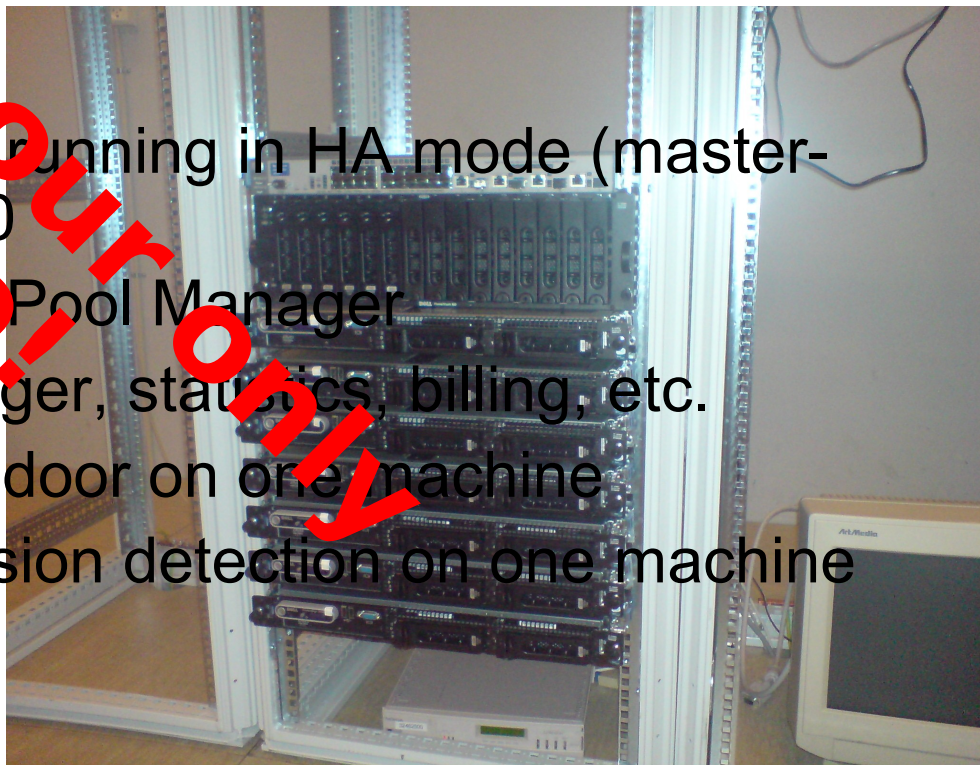


- "...to establish a Nordic data grid facility and to involve Nordic countries in European and global co-operation in data sharing in a variety of fields."
- To *coordinate* and *facilitate* the creation of a Nordic eInfrastructure sharing platform
- To enable Nordic researchers to participate in major international projects
- To optimize and standardize use of resources
- To optimize Nordic participation in international projects

A distributed organization



- Central Installation:
 - 7 Dell 1950 2xDual Core 2GHz Xeon, 4GB RAM, 2 x 73GB 15k SAS disks (mirrored) (one for spare)
 - 2 x Dell PowerVault MD-1000 direct attached storage enclosures with 7 x 143GB 15k SAS RAID-10 each
- Running:
 - 2 Postgress for PNFS running in HA mode (master-slave) DB on MD-1000
 - 1 PNFS Manager and Pool Manager
 - 1 SRM, location manager, statistics, billing, etc.
 - 1 GridFTP and xrootd door on one machine
 - 1 Monitoring and intrusion detection on one machine



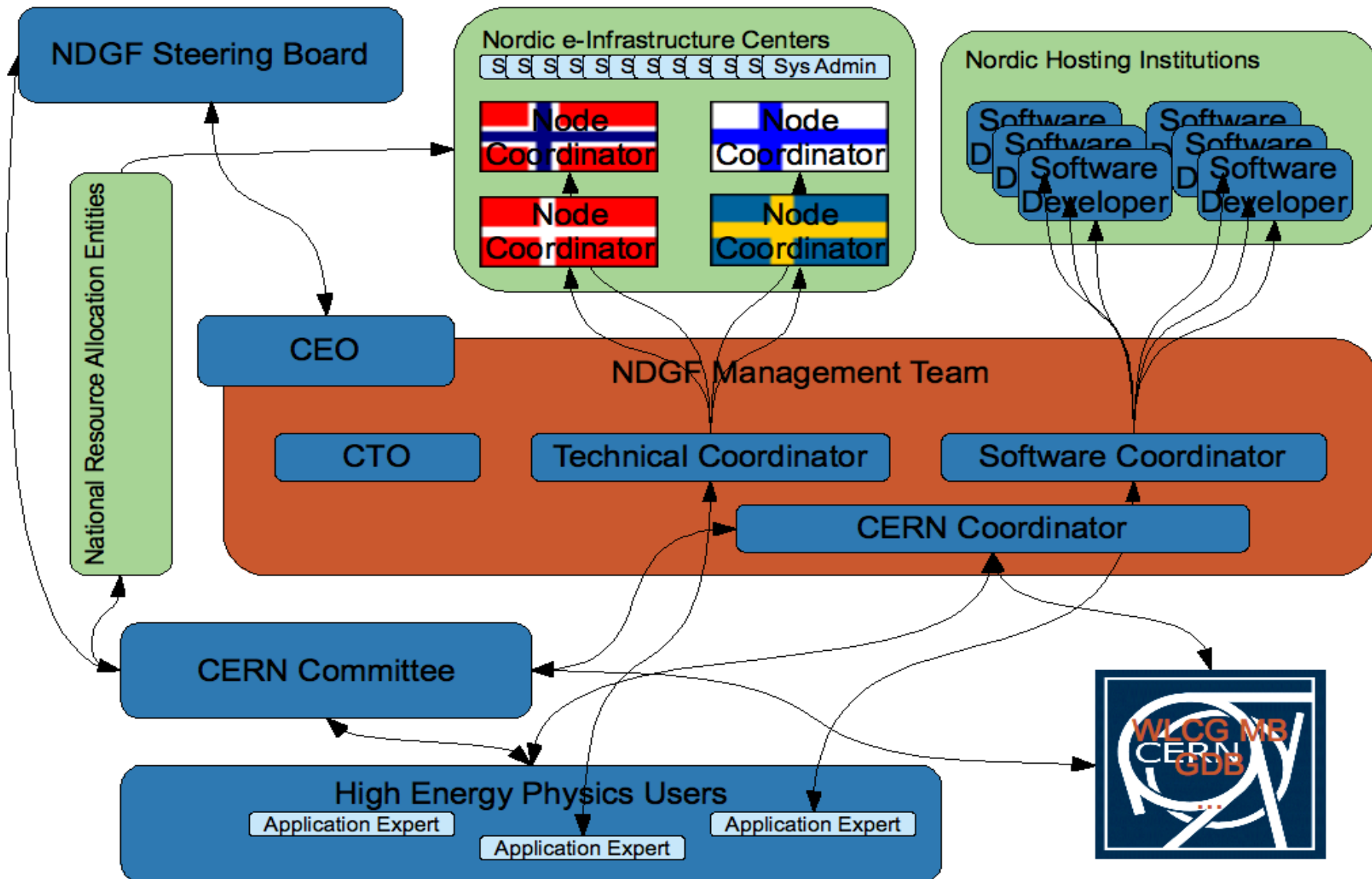
Tier-1 Services:

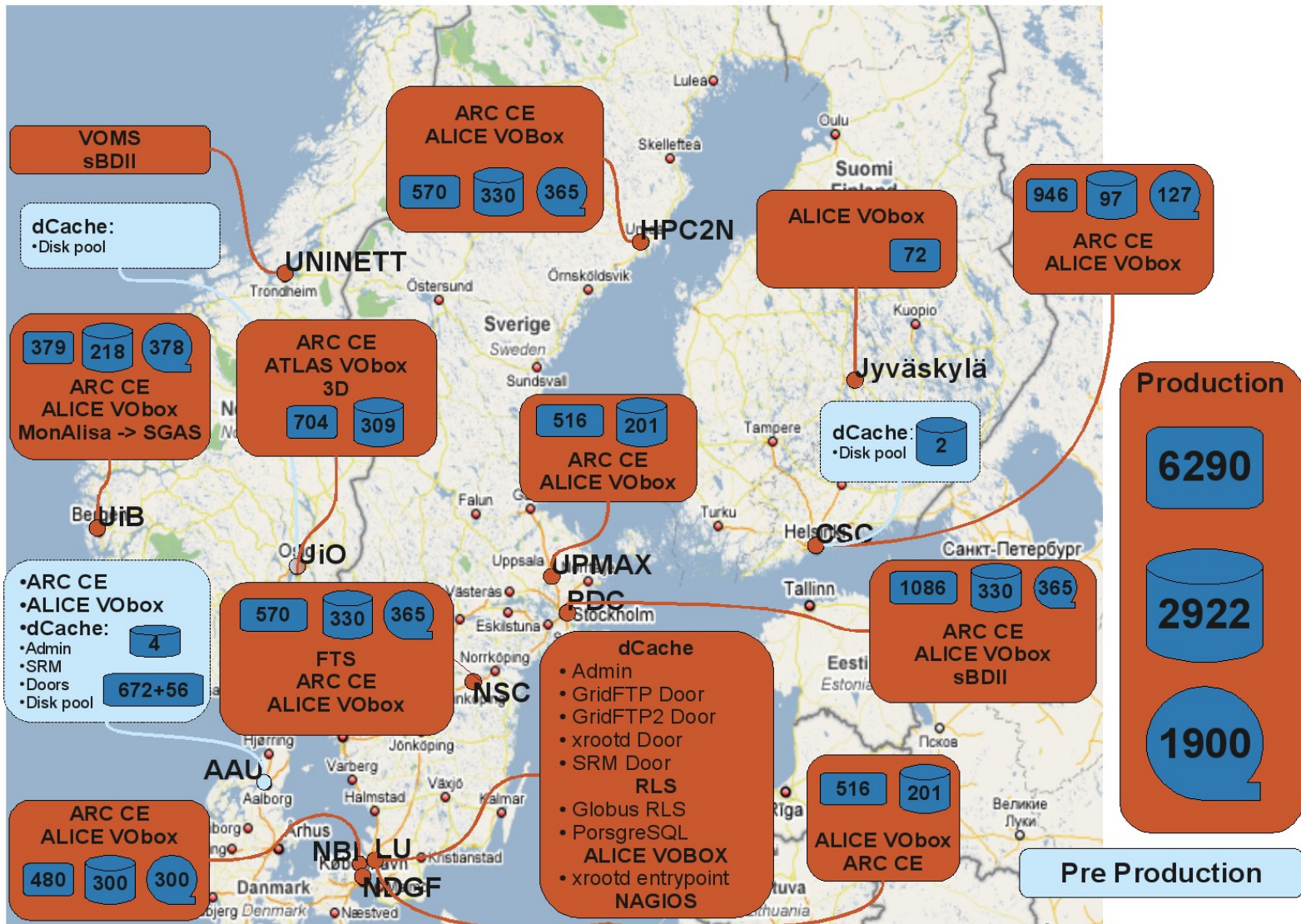
- Organization
- Network
- Computing
- Storage
- ATLAS
- ALICE
- Accounting
- Monitoring



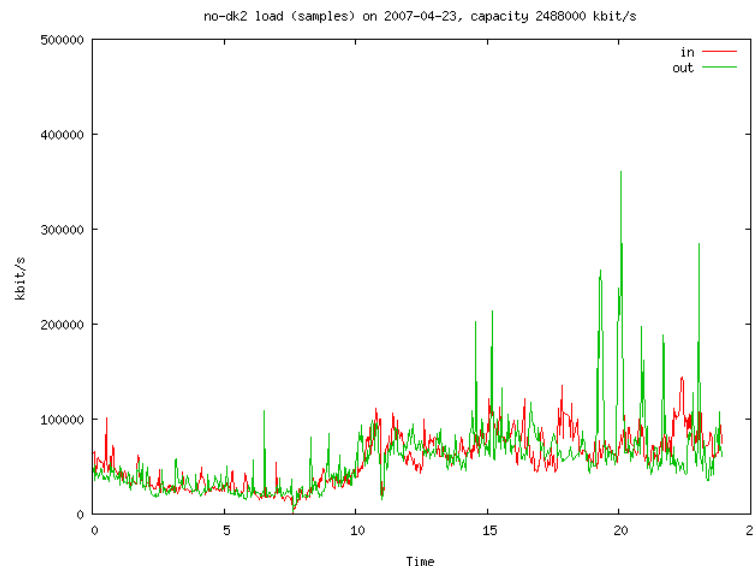
- The 7 biggest Nordic compute centers, dTier-1s, form the NDGF Tier-1
- Resources (Storage and Computing) are scattered
- Services can be centralized
- Advantages in redundancy
- Especially for 24x7 data taking



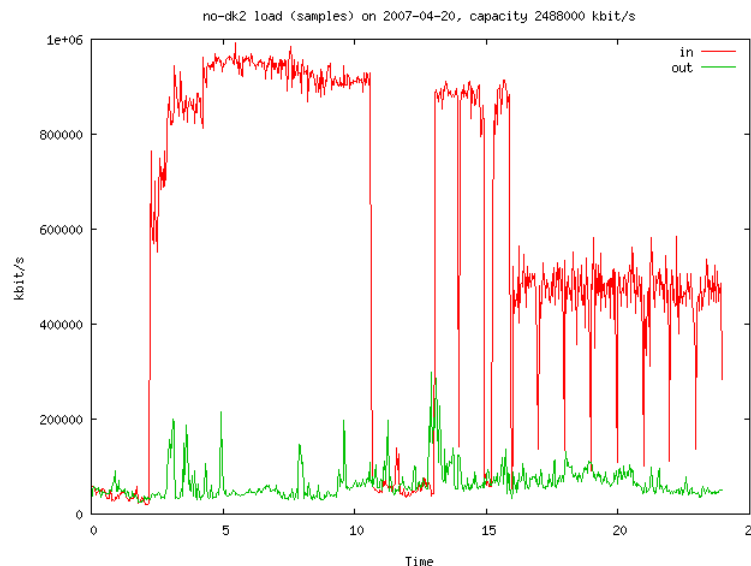




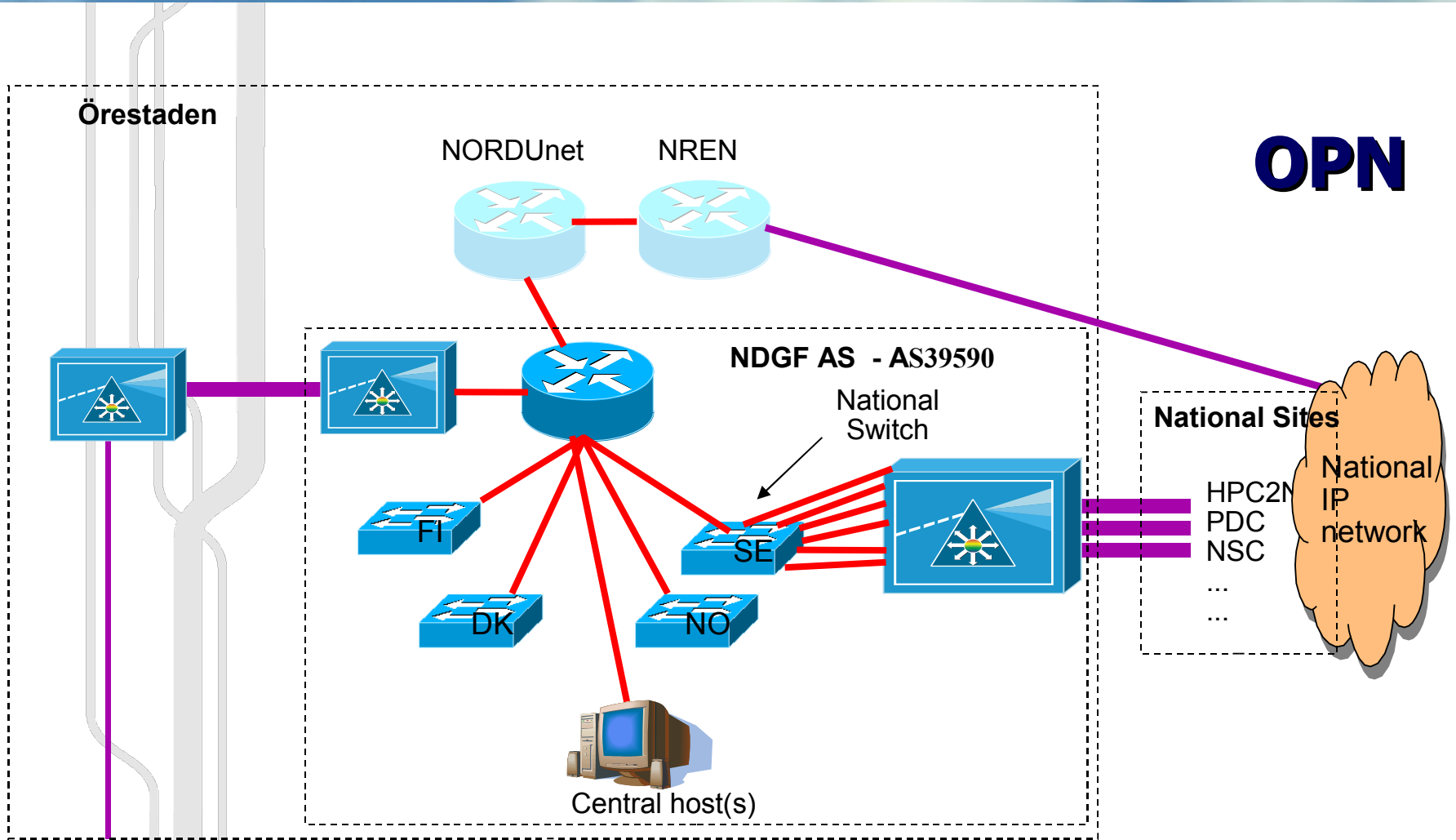
- Today NDGF is connected directly with GEANT 10Gbit fiber to CERN
- Inter-Nordic shared 10Gbit network from NORDUnet
- A Dedicated 10Gbit LAN covering all dTier-1 centers being build



- Today NDGF is connected directly with GEANT 10Gbit fiber to CERN
- Inter-Nordic shared 10Gbit network from NORDUnet
- A Dedicated 10Gbit LAN covering all dTier-1 centers being build



The Infrastructure: Networking



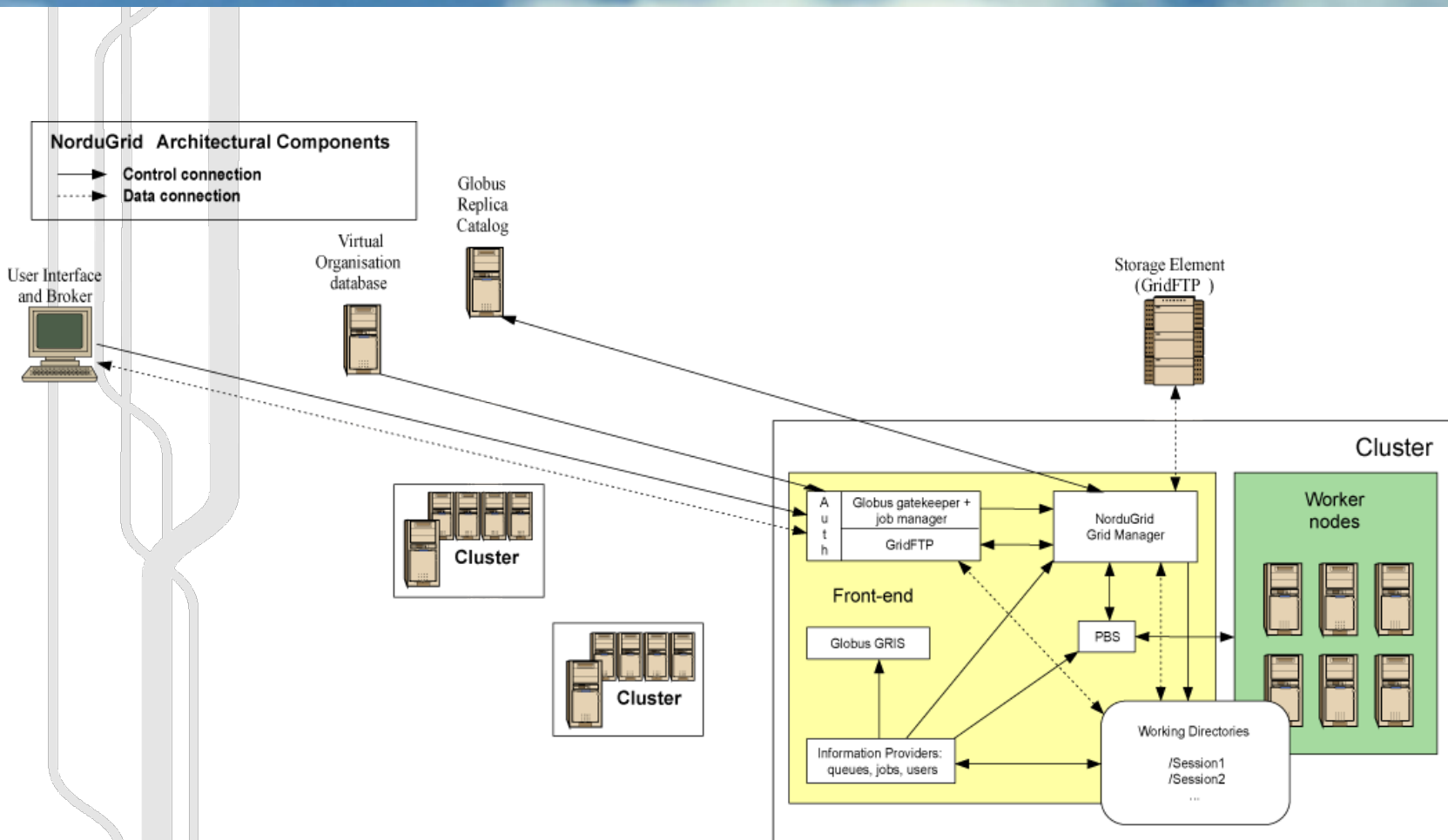
CERN
LHC

- NorduGrid / ARC middleware for Computing
- Used routinely since 2002 for e.g. ATLAS data challenges
- Deployed at all the dTier-1 sites and Tier-2 sites

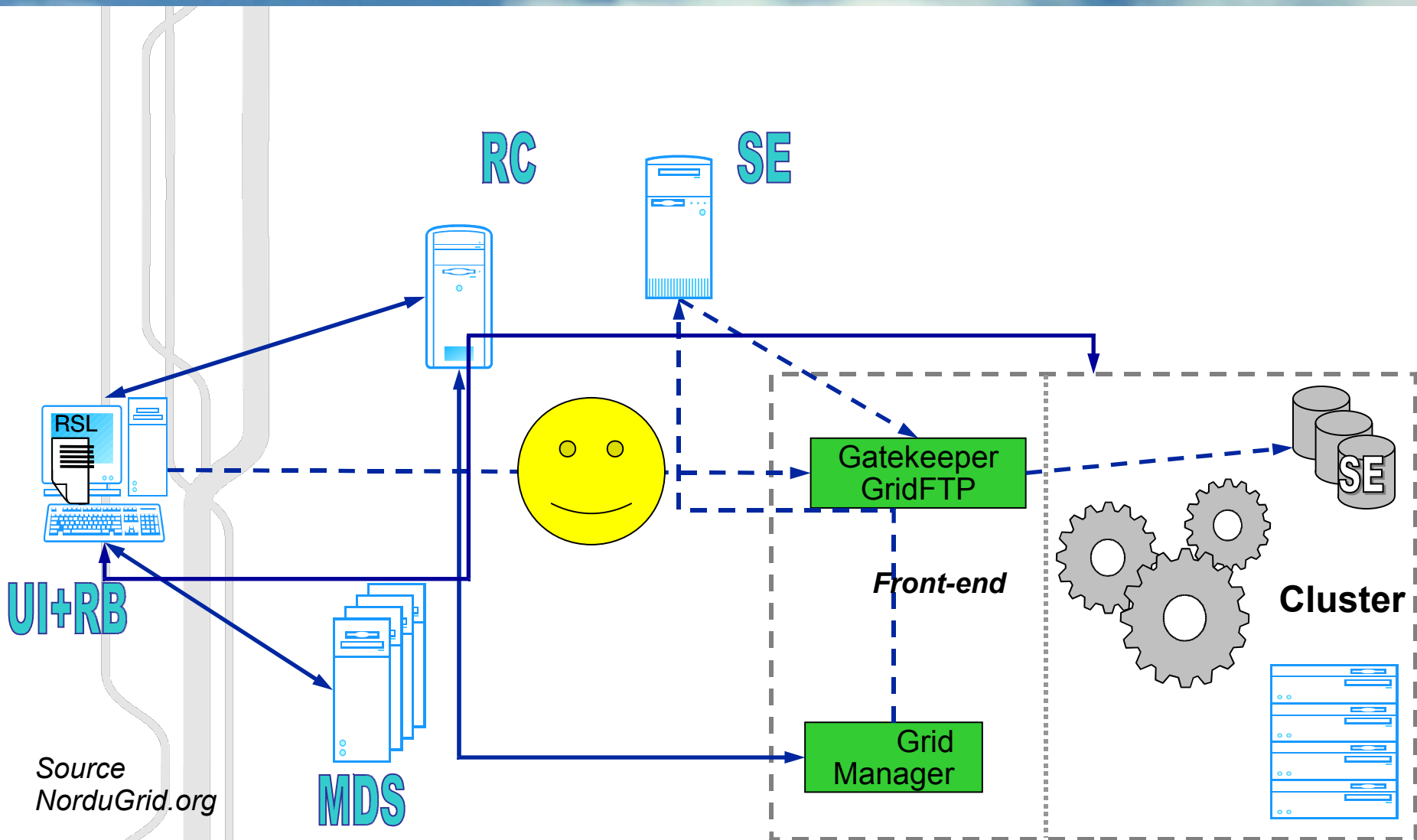
Grid Monitor - Microsoft Internet Explorer

Processes: ■ Grid ■ Local

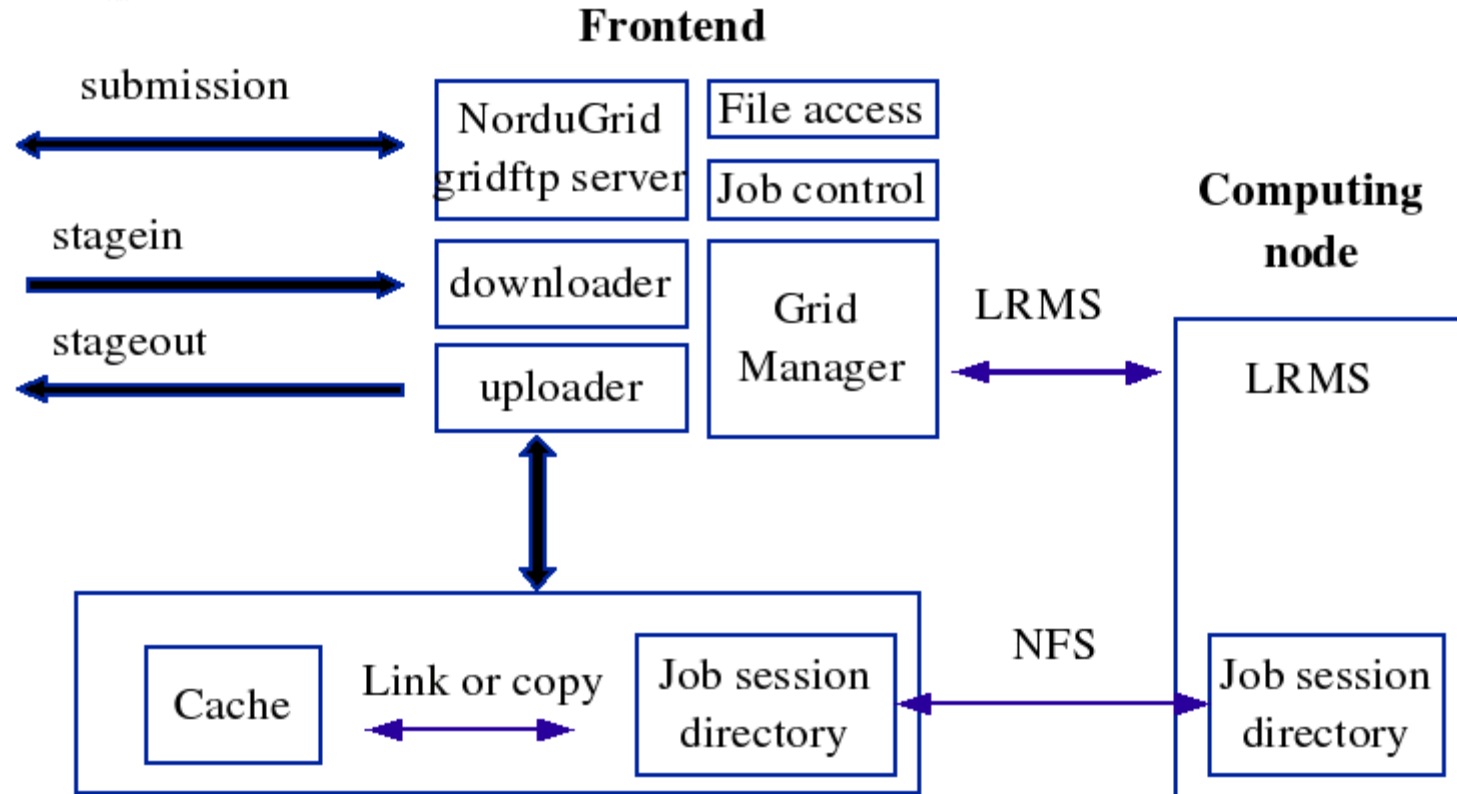
Country	Site	CPUs	Load (processes: Grid+local)	Queueing
Australia	Atlas (UniMelb)	26	0+2	0+0
	Charm (UniMelb)	36	0+0 (queue down)	0+0
	Alfred (UniMelb)	90	0+6	2+1
Denmark	DistLab (DIKU)	10	0+0	0+0
	Aalborg Grid Gateway	46	38+0	0+0
	Niflheim (DCSC/DTU)	902	0+898	0+17
	Horseshoe (DCSC/SDU)	1192	0+873	0+3
	HEPAX1	1	0+0	0+0
	Morpheus	18	15+0	23+0
	Theory (DCSC/IKU)	112	0+42	0+1
	VCR (VideoRecorder)	1	1+0 (queue down)	0+0
Estonia	UT IMCB Anakonda clus>	15	3+0	0+0
	UT CS Antarctica Clus>	20	6+0	0+0
	CMS on CERN Linux	1	0+0	0+0
	CMS Production server	5	0+0	0+0
	UT DOUG Cluster	2	0+0	0+0
	CMS test cluster	1	0+0	0+0
	EENet cluster	6	0+0	0+0
	UT Physics Cluster	3	3+0	0+0
Finland	CSC Kirppu	1	1+0	6+0
	Mill (Physicum)	60	0+15	0+0
	Alpha (HIP)	1	0+0	0+0
	Testbed0 (HIP)	1	0+0	4+1
Germany	FZK cluster	996	83+349	0+0
	LRZ cluster	234	0+230	0+243
Norway	Oslo Temp Cluster	11	0+0	25+0
	Parallab IBM Cluster	58	0+57	0+75
	Bergen Grid Cluster	2	2+0	7+0
	Oslo Grid Cluster	41	9+15	51+0
UiO Grid	100	0+98	0+1	
Slovenia	SIGNET	40	6+31	6+0
Sweden	Bluesmoke (SweGrid,NS>	99	95+0	187+0
	Kosufy farm	60	36+0	0+0
	ISV	4	4+0	14+0
	Hagrid (SweGrid, Uppm>	100	50+0	68+0
	Ingrid (SweGrid,HPC2N)	101	69+0	124+0
	Monolith (NSC)	398	0+342	0+121
	Quark Cluster	7	0+0	0+0
	Beppe (SweGrid PDC KT>	96	92+0	49+0
	Sigrid (SweGrid, Luna>	99	49+50	19+25
Toto7/Whenim64 (Lunar>	192	0+161	0+11	
Switzerland	Bern ATLAS Cluster	8	8+0	12+0
TOTAL	42 sites	5196	570 + 3169	597 + 499



October 2002



Source
NordGrid.org



arcsub

to submit a task

arcstat

to obtain the status of jobs and clusters

arccat

to display the stdout or stderr of a running job

arcget

to retrieve the result from a finished job

arckill

to cancel a job request

arcclean

to delete a job from a remote cluster

arc renew

to renew user's proxy

arcsync

to synchronize the local job info with the MDS

arccopy

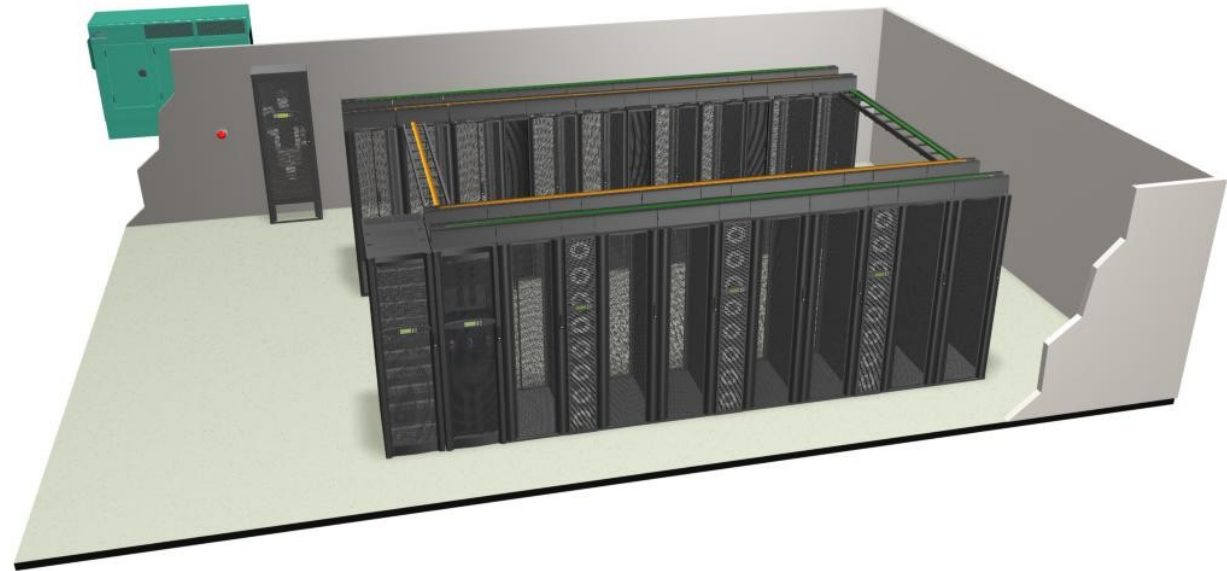
to transfer files to, from and between clusters

arc remove

to remove files

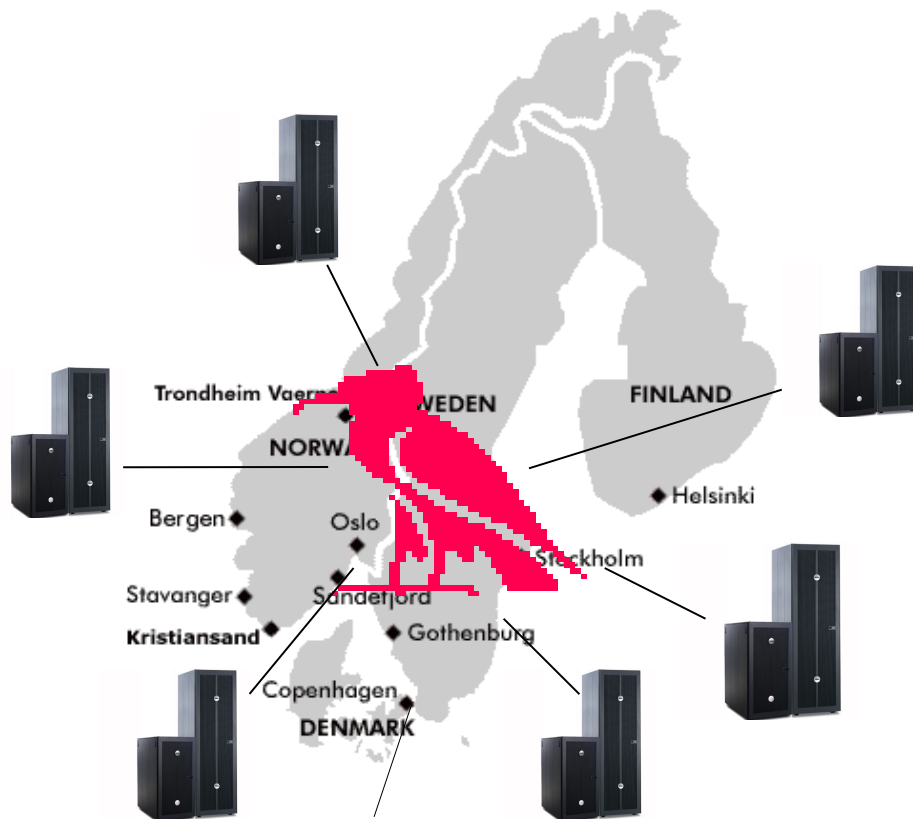
- The user must be authorized to use the cluster and the queue
- The cluster's and queue's characteristics must match the requirements specified in the xRSL string (max CPU time, required free disk space, installed software etc)
- If the job requires a file that is registered in a Replica Catalog, the brokering gives priority to clusters where a copy of the file is already present
- From all queues that fulfills the criteria one is chosen randomly, with a weight proportional to the number of free CPUs available for the user in each queue
- If there are no available CPUs in any of the queues, the job is submitted to the queue with the lowest number of queued job per processor

The Infrastructure: Storage

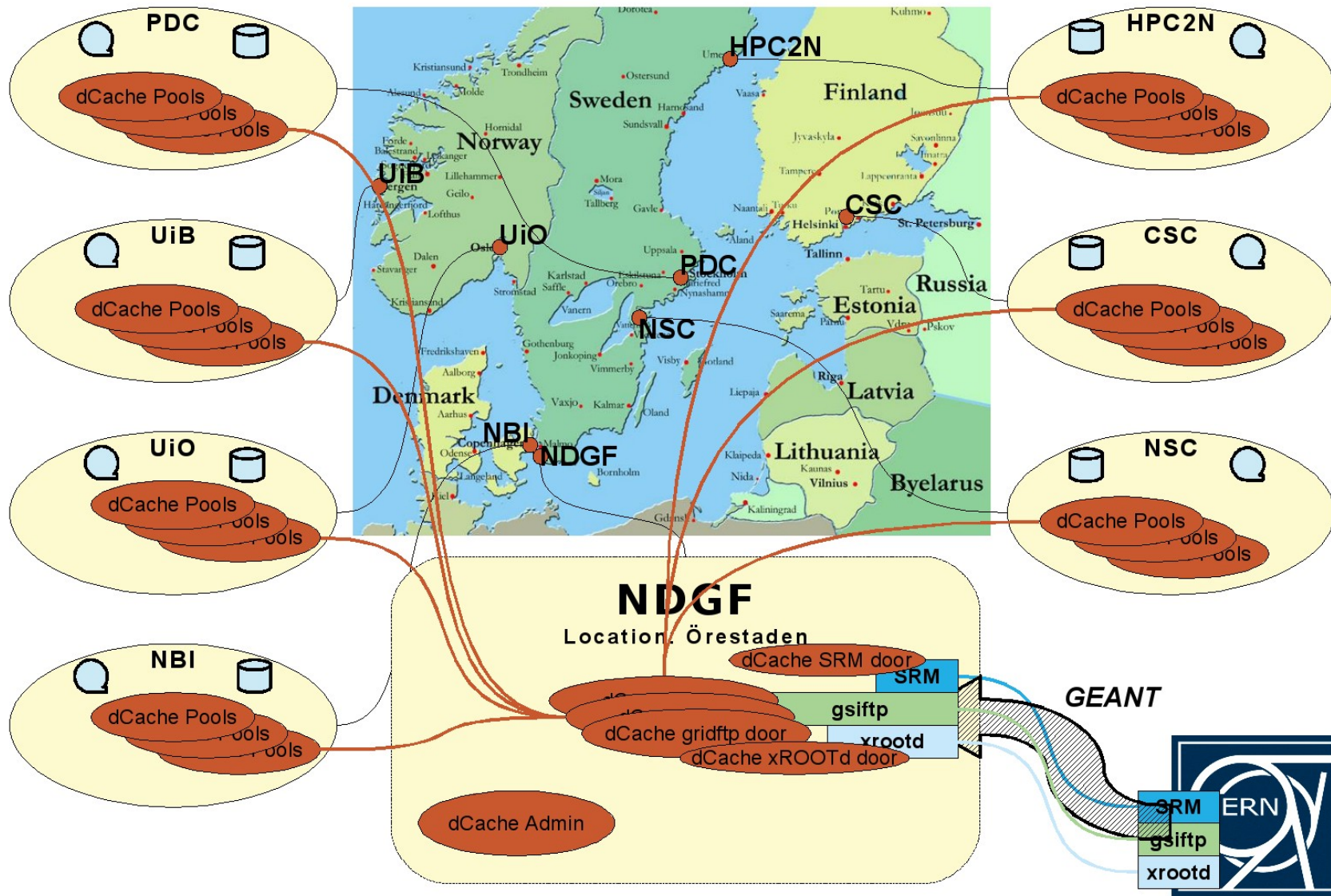


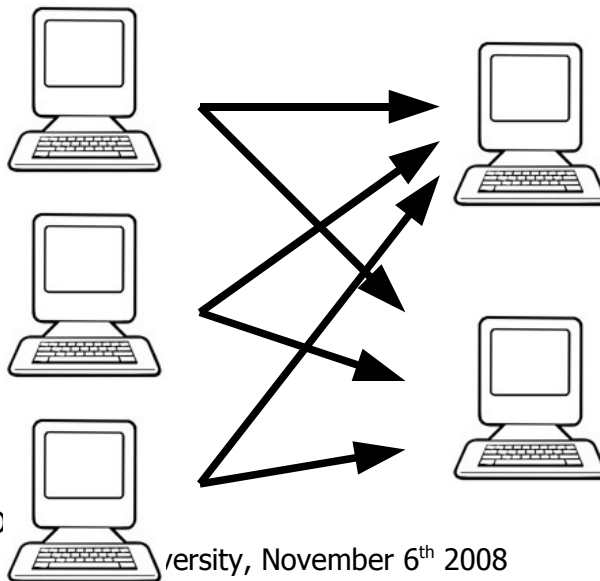
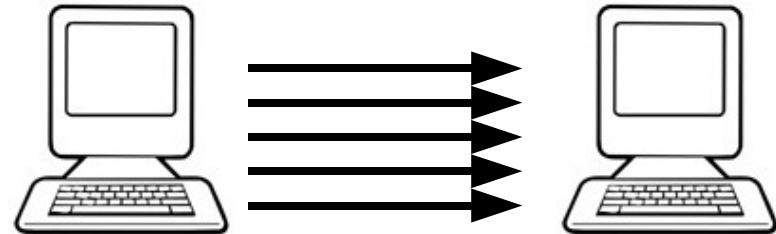
The Infrastructure: Storage

- dCache Installation
- Admin and Door nodes at GEANT endpoint
- Pools at sites
- Very close collaboration with DESY and FermiLab ensure dCache is suited also for distributed use

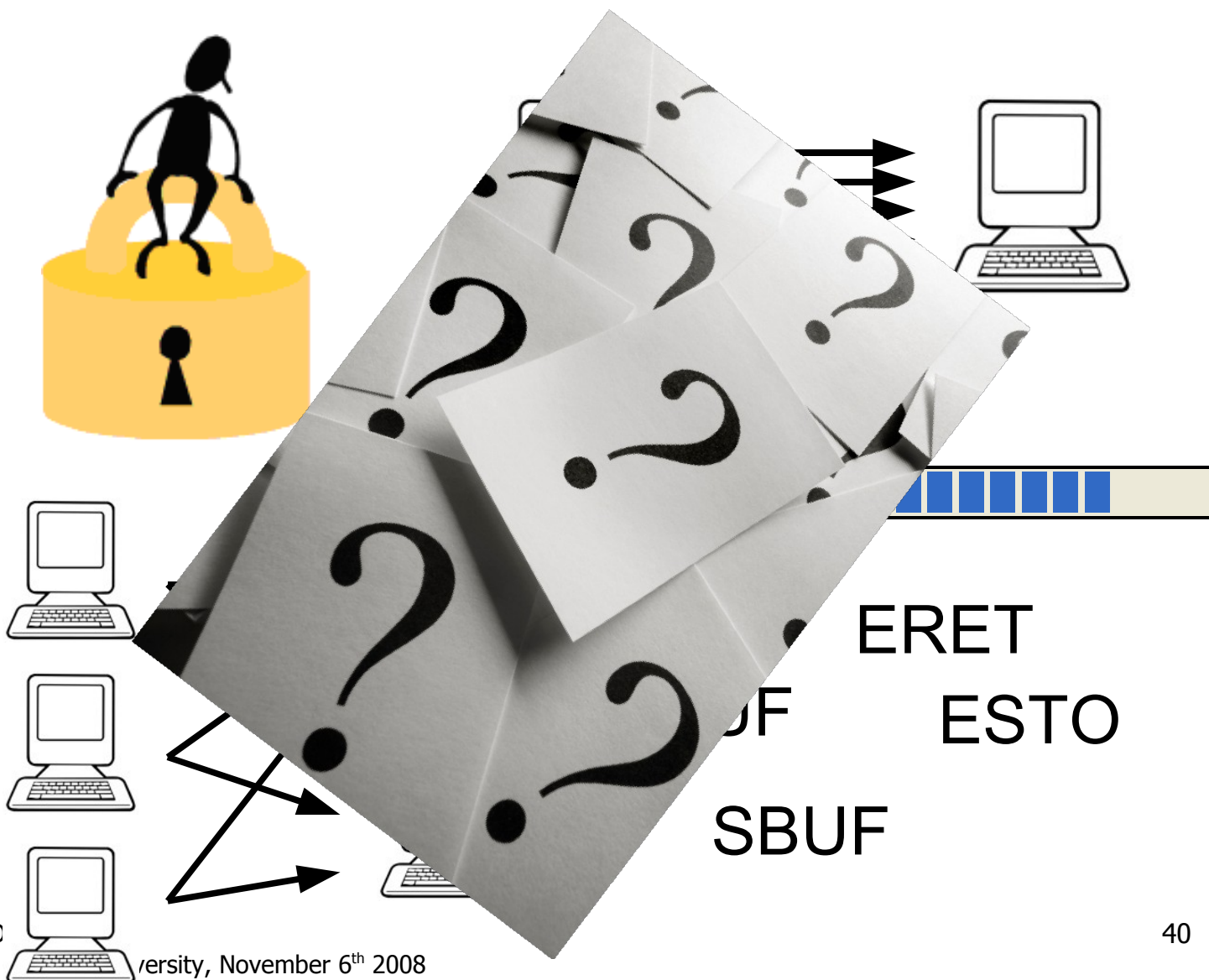


The Infrastructure: Storage

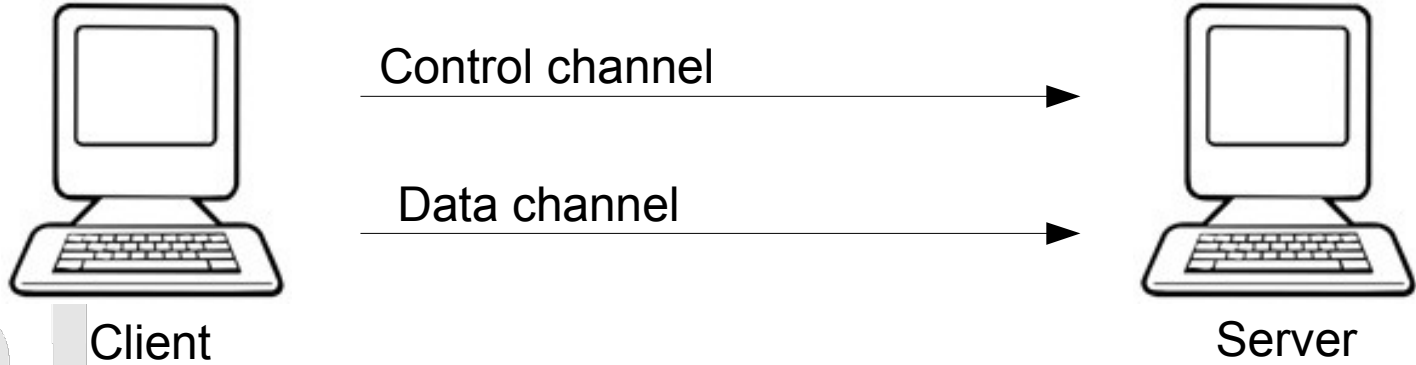




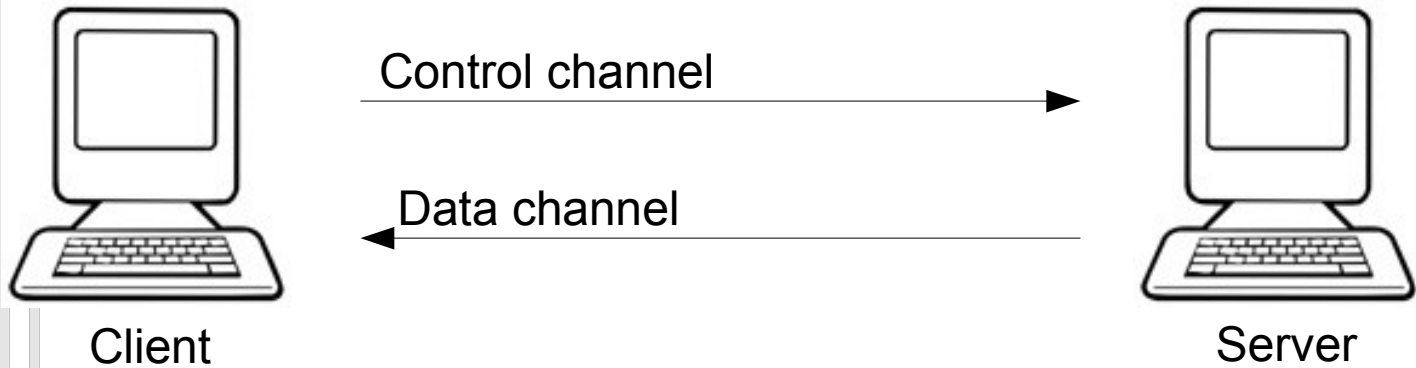
ERET
ABUF ESTO
SBUF



Passive servers



Active servers





Client

/disk/my_file

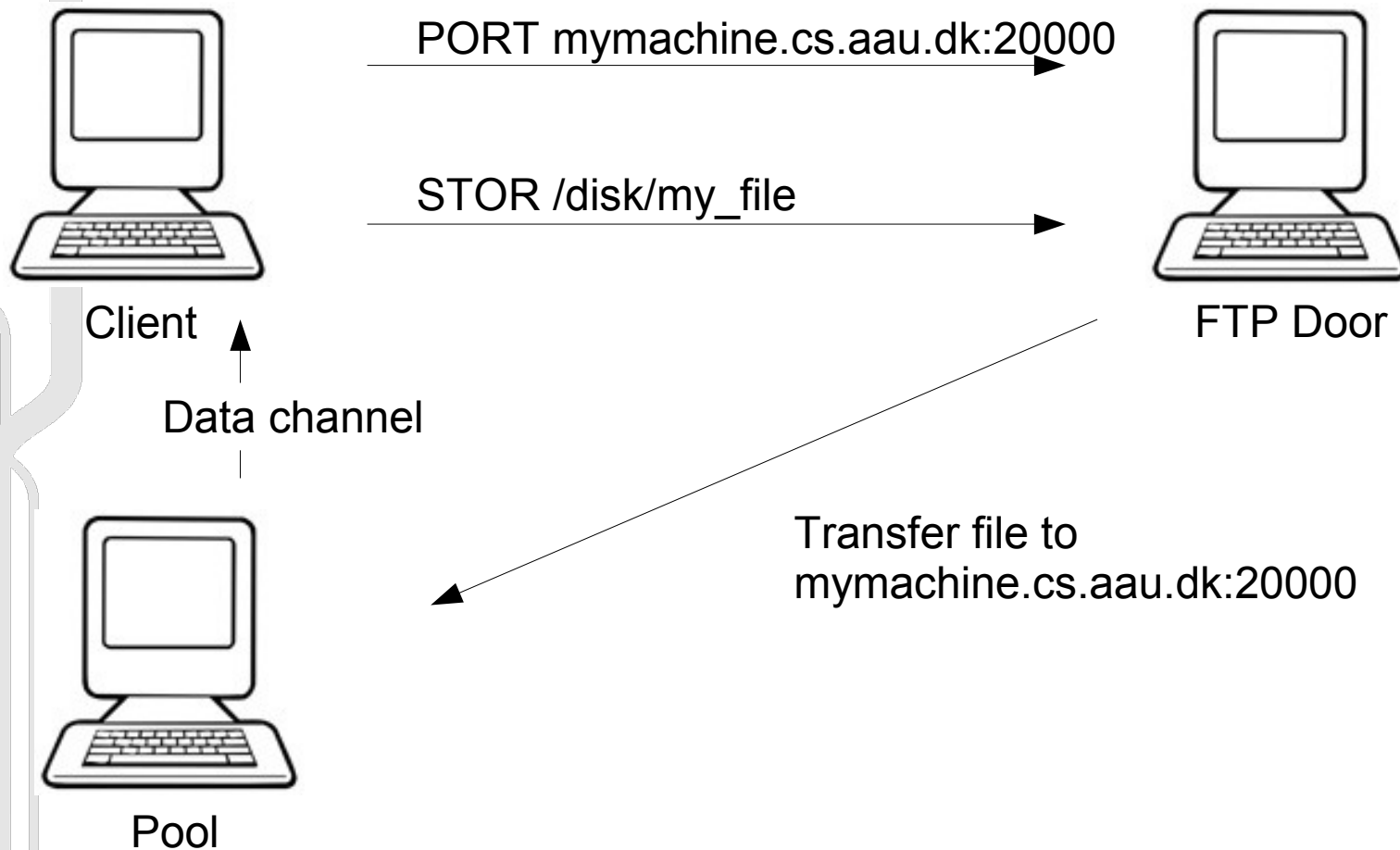


Pool

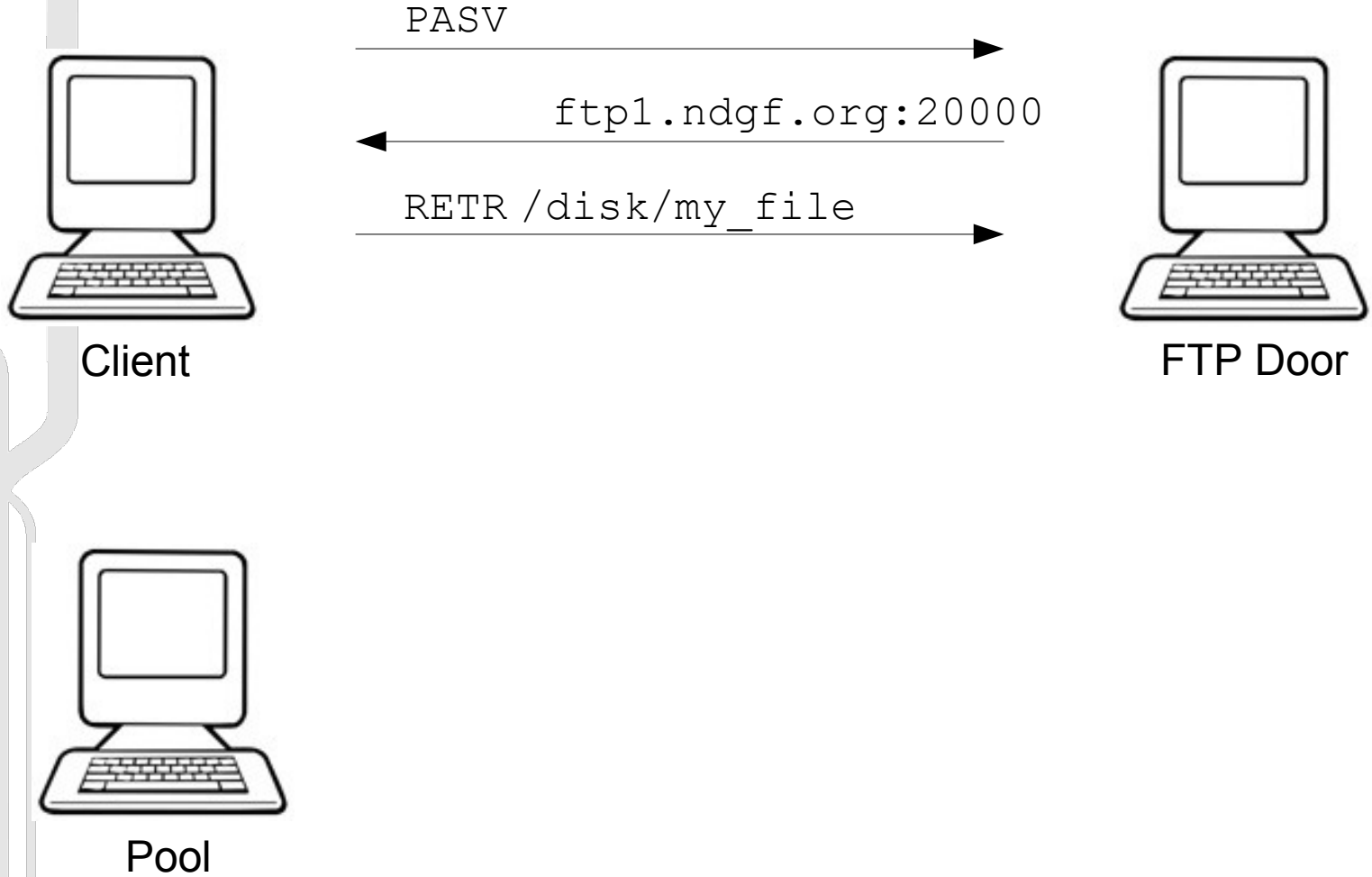


FTP Door

Active transfers in dCache



Passive transfers in dCache





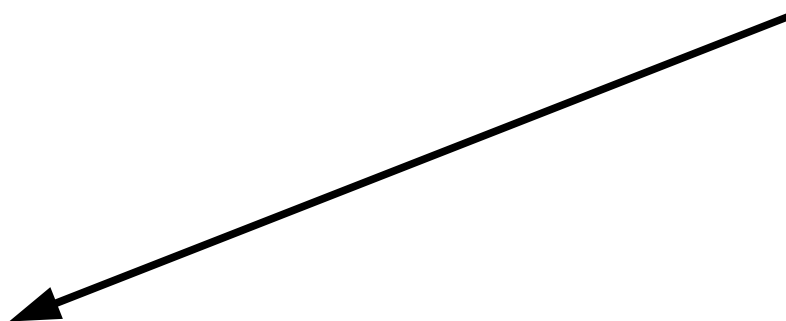
Client



Pool



FTP Door



Passive transfers in dCache



Client



Pool



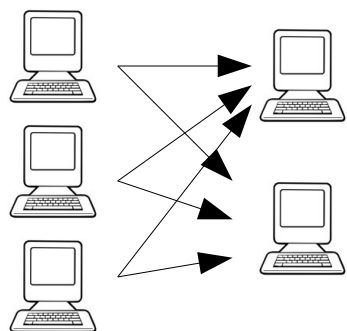
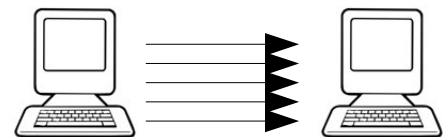
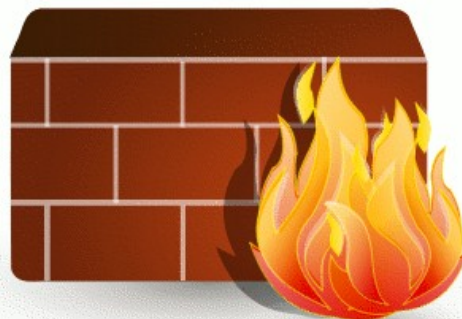
FTP Door



Use active transfers!

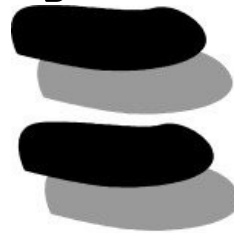


Active transfers won't work

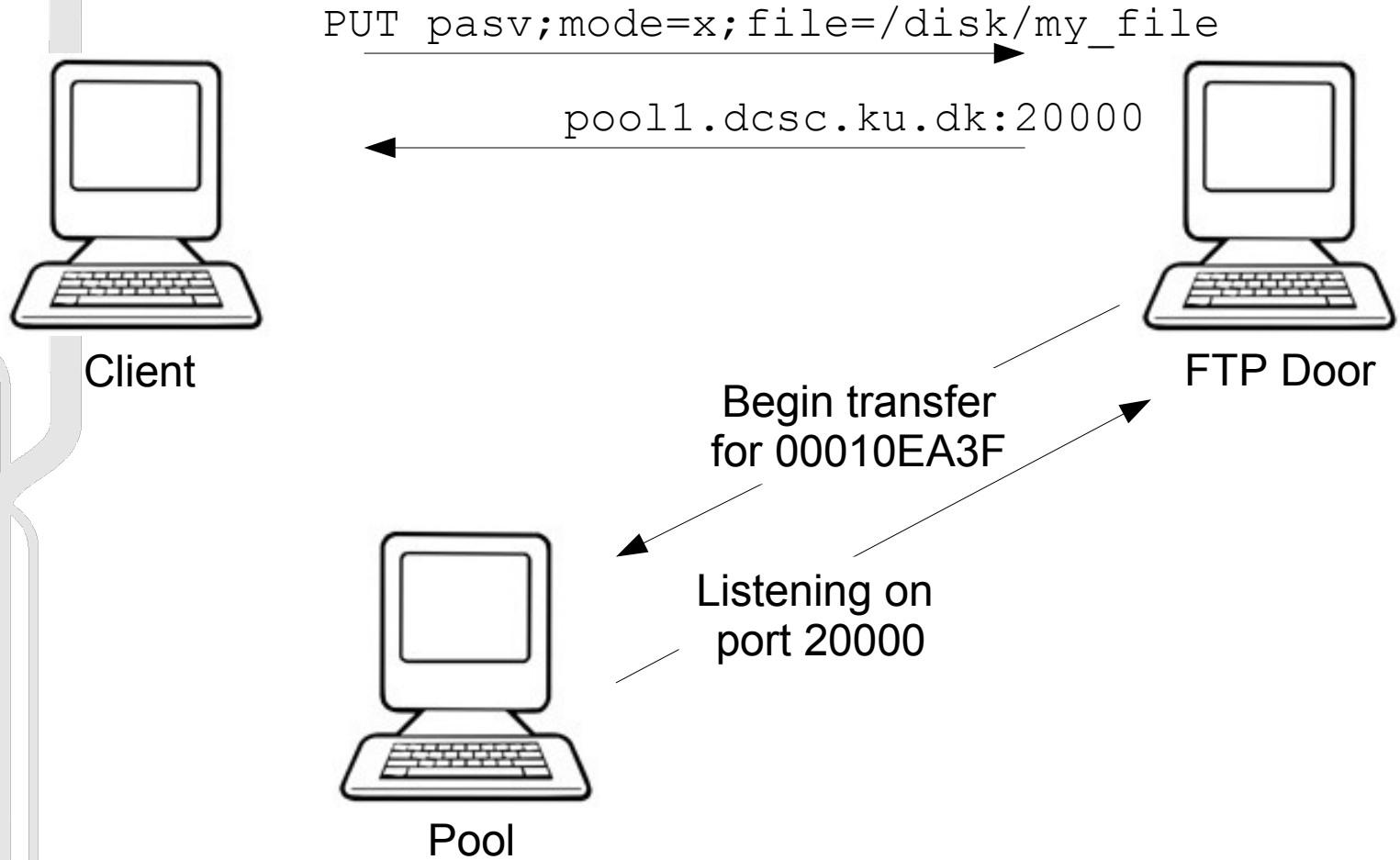


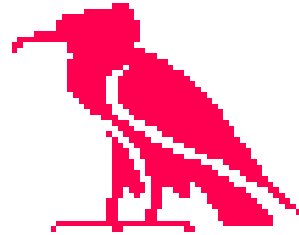
ERET
ABUF ESTO
SBUF


```
PASV  
227 PORT=(a.b.c.d)  
MODE E  
200 OK  
STOR /disk/my_file  
150 Opening data channel
```



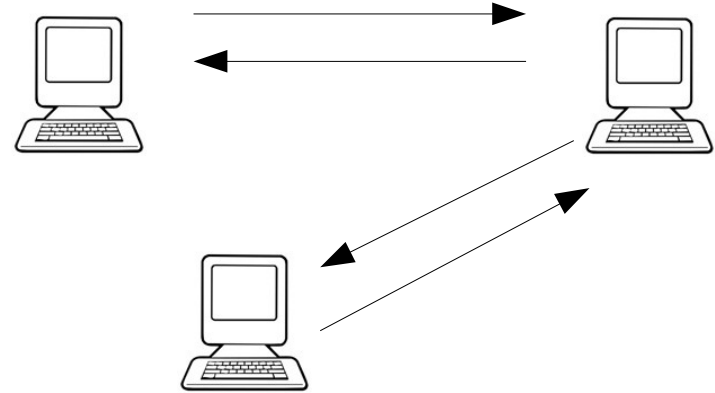
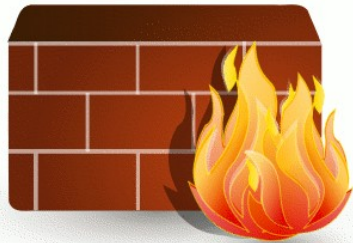
```
PUT pasv;mode=x;file=/disk/my_file  
127 PORT=(a.b.c.d)  
150 Opening data channel
```

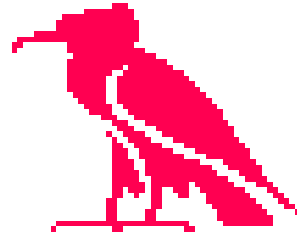




KnowARC







EGEE Accounting Portal

http://www3.egee.cesga.es/gridsite/accounting/CESGA/tier1_view.html

Bonjour SMHI - Ocea... HIROMB System Mana... HEP sites Aktuelt - F...andsvæsenet The 3G Ipod...d 3G iPhone EGEE Accounting Portal NORDUnet DB

Marc Livanage - Blog - ... Welcome to ISGC 2008 NORDUnet2008 Confer... Meeting-2008-01-25 N... EGEE Accounting Portal riska site:www.aka.fi - ...

EGEE ACCOUNTING PORTAL

GLOBAL View
VO MANAGER View
VO MEMBER View
SITE ADMIN View
USER View
REPORTS
LINKS

Hierarchical Tree

- ▼ Tier1
 - CA-TRIUMF
 - CH-CERN
 - DE-KIT
 - ES-PIC
 - FR-CCIN2P3
 - IT-INFN-CNAF
 - NDGF
 - NL-T1
 - TW-ASGC
 - UK-T1-RAL
 - US-FNAL-CMS
 - US-T1-BNL
- Tier2
- Countries
- ▼ EGEE
 - Production
 - PPS
 - OSG
 - UNREGISTERED
 - VO_Discipline
 - VO_Metrics

The following table shows the distribution of Normalised CPU time grouped by TIER1 and VO (only information about LHC VOs is showed in detail. The rest of VOs will be grouped in a new category).

Normalised CPU time [units 1K.SI2K.Hours] by TIER1 and VO							
TIER1	alice	atlas	cms	lhcb	Other VOs	Total	%
CA-TRIUMF	0	4,463,305	0	0	24,809	4,488,114	6.73%
CH-CERN	2,425,120	1,962,054	2,623,014	904,796	395,677	8,310,661	12.46%
DE-KIT	2,529,727	1,510,853	2,154,989	1,004,171	71,809	7,271,549	10.90%
ES-PIC	0	904,885	850,552	510,898	15,223	2,281,558	3.42%
FR-CCIN2P3	2,408,281	2,052,169	1,928,082	696,294	4,218,657	11,303,483	16.95%
IT-INFN-CNAF	1,162,184	1,698,409	1,560,147	833,114	2,245,540	7,499,394	11.25%
NDGF	1,230,314	2,618,708	0	0	7	3,849,029	5.77%
NL-T1	483,529	932,599	225	908,622	1,172,615	3,497,590	5.25%
TW-ASGC	0	498,288	412,239	0	514,375	1,424,902	2.14%
UK-T1-RAL	304,466	1,292,661	601,991	571,041	444,599	3,214,758	4.82%
US-FNAL-CMS	0	2,752	6,488,710	0	0	6,491,462	9.74%
US-T1-BNL	0	7,049,143	11	0	0	7,049,154	10.57%
Total	10,543,621	24,985,826	16,619,960	5,428,936	9,103,311	66,681,654	
Percentage	15.81%	37.47%	24.92%	8.14%	13.65%		

Click here for a csv dump of this table

The information in the previous table is also shown in the following graph.

TIER1 Normalised CPU time by TIER1 and VO
LHC VOs. January 2007 - December 2007

Developed by CESGA

(C) CESGA. 2006

EGEE Accounting Portal

http://www3.egee.cesga.es/gridsite/accounting/CESGA/tier1_view.html
Google

Bonjour SMHI - Ocea... HIROMB System Mana... HEP sites Aktuell - F...andsvæsenet The 3G Ipod...d 3G iPhone EGEE Accounting Portal NORDUnet DB

Marc Livanage - Blog - ... Welcome to ISGC 2008 NORDUnet2008 Confer... Meeting-2008-01-25 N... EGEE Accounting Portal riska site:www.aka.fi - ...

EGEE ACCOUNTING PORTAL

GLOBAL View

VO MANAGER View

VO MEMBER View

SITE ADMIN View

USER View

REPORTS

LINKS

Hierarchical Tree

- ▼ Tier1
 - CA-TRIUMF
 - CH-CERN
 - DE-KIT
 - ES-PIC
 - FR-CCIN2P3
 - IT-INFN-CNAF
 - NDGF
 - NL-T1
 - TW-ASGC
 - UK-T1-RAL
 - US-FNAL-CMS
 - US-T1-BNL
- Tier2
- Countries
- ▼ EGEE
 - Production
 - PPS
 - OSG
 - UNREGISTERED
 - VO_Discipline
 - VO_Metrics

The following table shows the distribution of Normalised CPU time grouped by TIER1 and VO (only information about LHC VOs is showed in detail. The rest of VOs will be grouped in a new category).

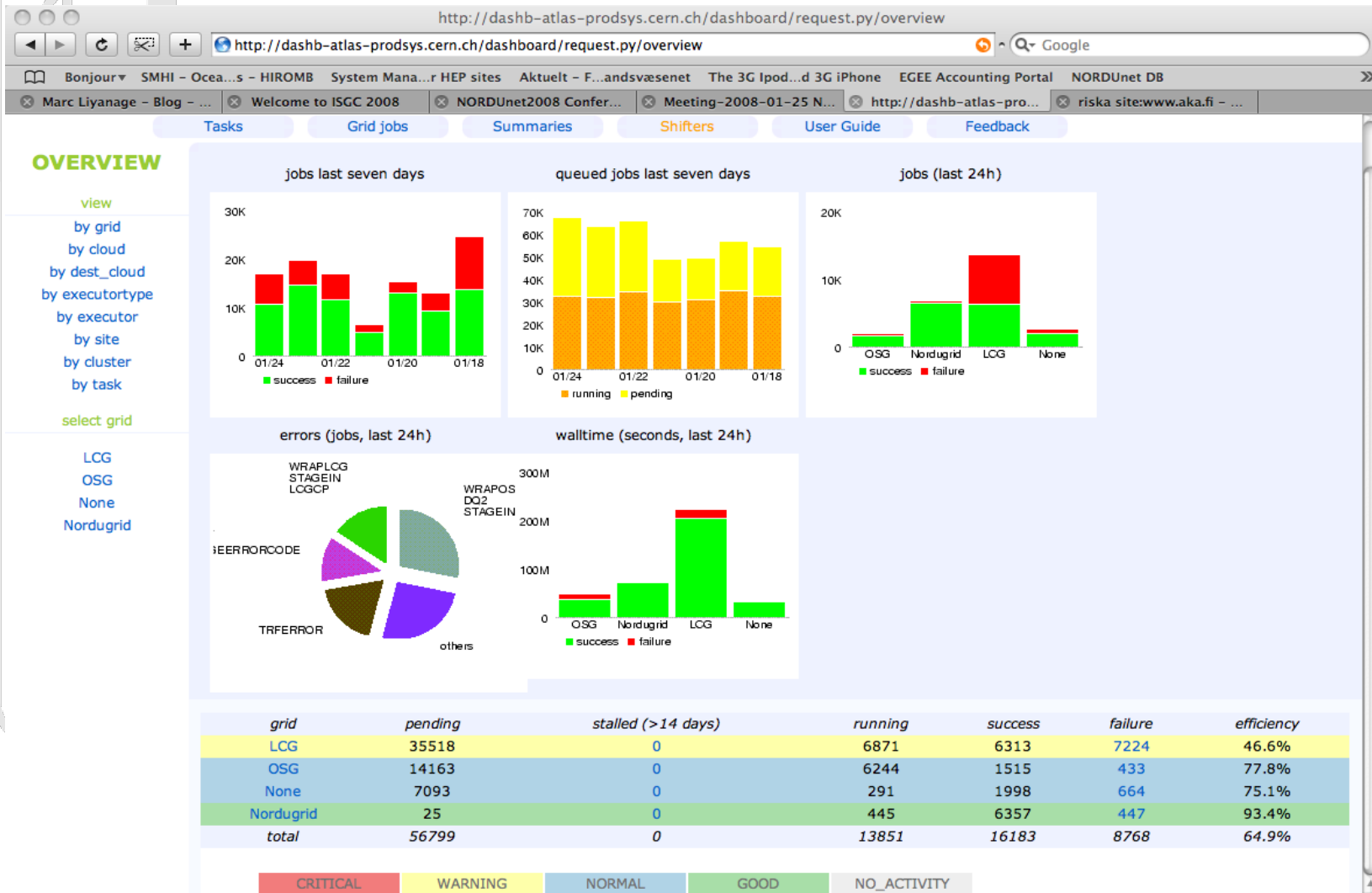
Normalised CPU time [units 1K.SI2K.Hours] by TIER1 and VO							
TIER1	alice	atlas	cms	lhcb	Other VOs	Total	%
CA-TRIUMF	0	4,463,305	0	0	24,809	4,488,114	6.73%
CH-CERN	2,425,120	1,962,054	2,623,014	904,796	395,677	8,310,661	12.46%
DE-KIT	2,529,727	1,510,853	2,154,989	1,004,171	71,809	7,271,549	10.90%
ES-PIC	0	904,885	850,552	510,898	15,223	2,281,558	3.42%
FR-CCIN2P3	2,408,281	2,052,169	1,928,082	696,294	4,218,657	11,303,483	16.95%
IT-INFN-CNAF							11.25%
NDGF	NDGF			1,230,314		2,618,708	5.77%
NL-T1	483,529	932,599	225	908,622	1,172,615	3,497,590	5.25%
TW-ASGC	0	498,288	412,239	0	514,375	1,424,902	2.14%
UK-T1-RAL	304,466	1,292,661	601,991	571,041	444,599	3,214,758	4.82%
US-FNAL-CMS	0	2,752	6,488,710	0	0	6,491,462	9.74%
US-T1-BNL	0	7,049,143	11	0	0	7,049,154	10.57%
Total	10,543,621	24,985,826	16,619,960	5,428,936	9,103,311	66,681,654	
Percentage	15.81%	37.47%	24.92%	8.14%	13.65%		

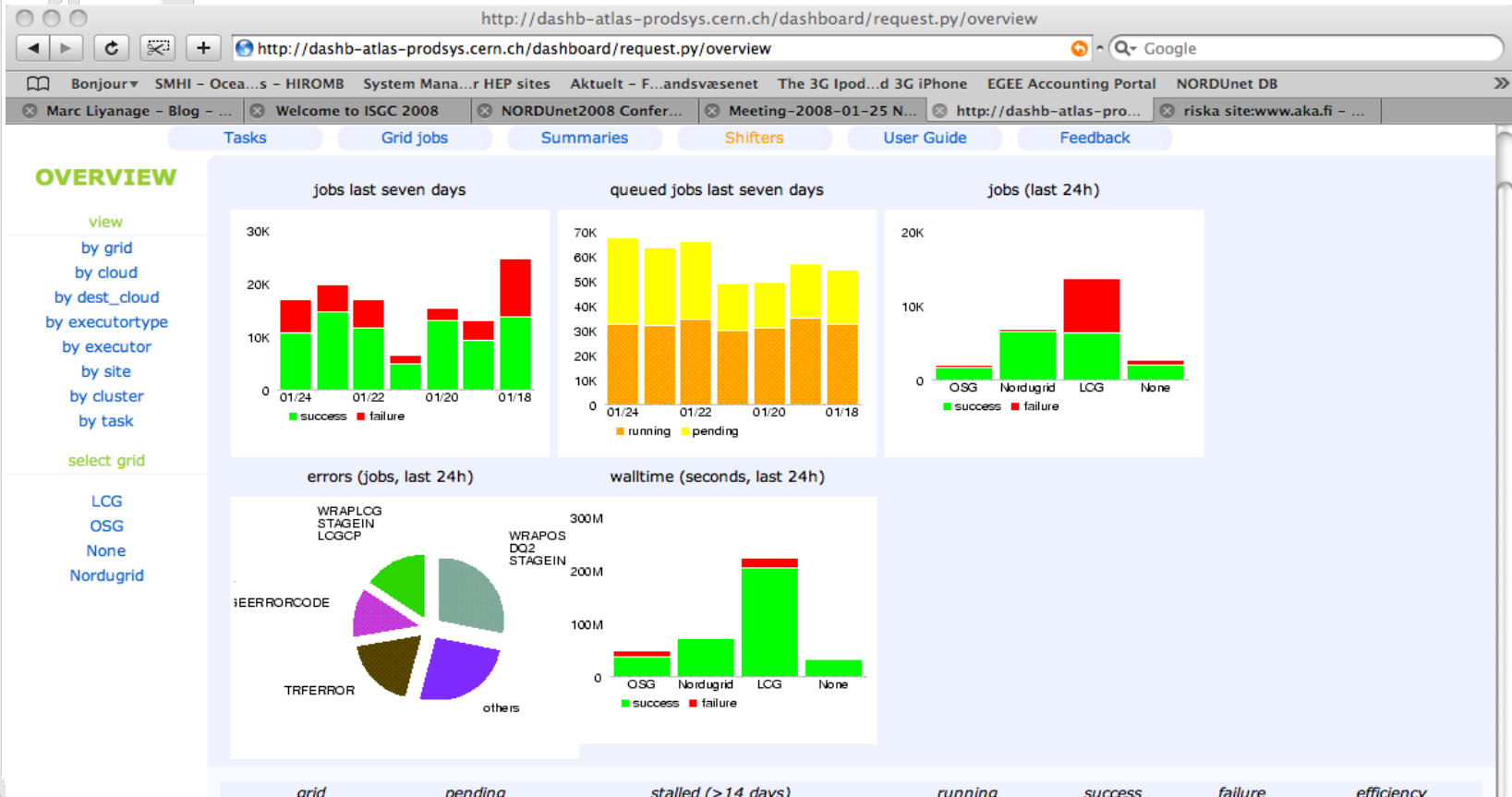
Click here for a csv dump of this table

The information in the previous table is also shown in the following graph.

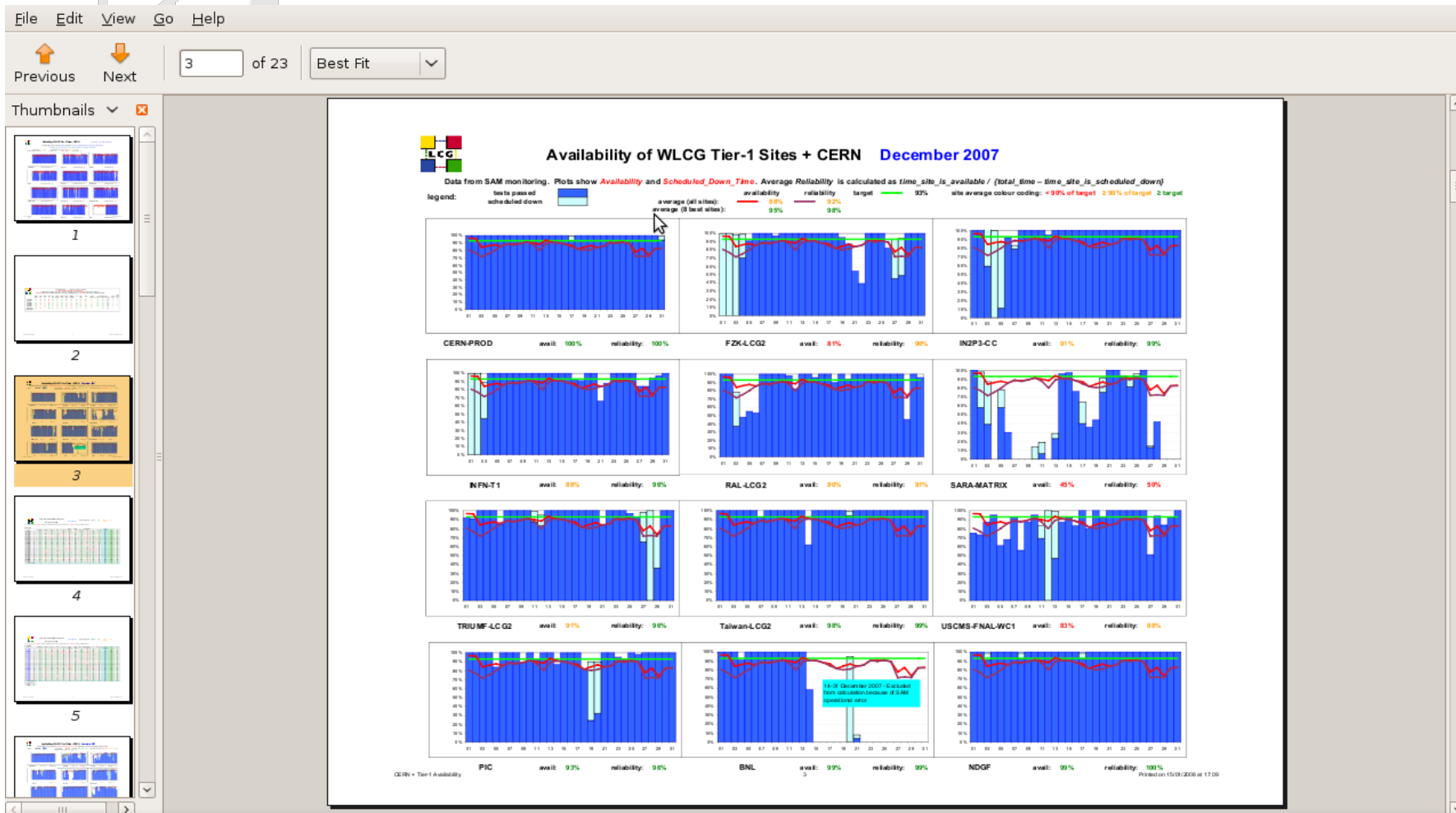
TIER1 Normalised CPU time by TIER1 and VO
LHC VOs. January 2007 - December 2007

Developed by CESGA





grid	pending	stalled (>14 days)	running	success	failure	efficiency
LCG	35518	0	6871	6313	7224	46.6%
OSG	14163	0	6244	1515	433	77.8%
None	7093	0	291	1998	664	75.1%
Nordugrid	25	0	445	6357	447	93.4%
total	56799	0	13851	16183	8768	64.9%







Questions?