



A Nordic Tier-1 for LHC

Mattias Wadenstein
Systems Integrator, NDGF
Grid Operations Workshop
Stockholm, June the 14th, 2007

- Background
 - Nordic Tier-1
- Organization / Governance
- Tier-1 Services:
 - Computing
 - Storage
 - ATLAS
 - ALICE
- Operation
 - current usage
 - tools used in daily operations

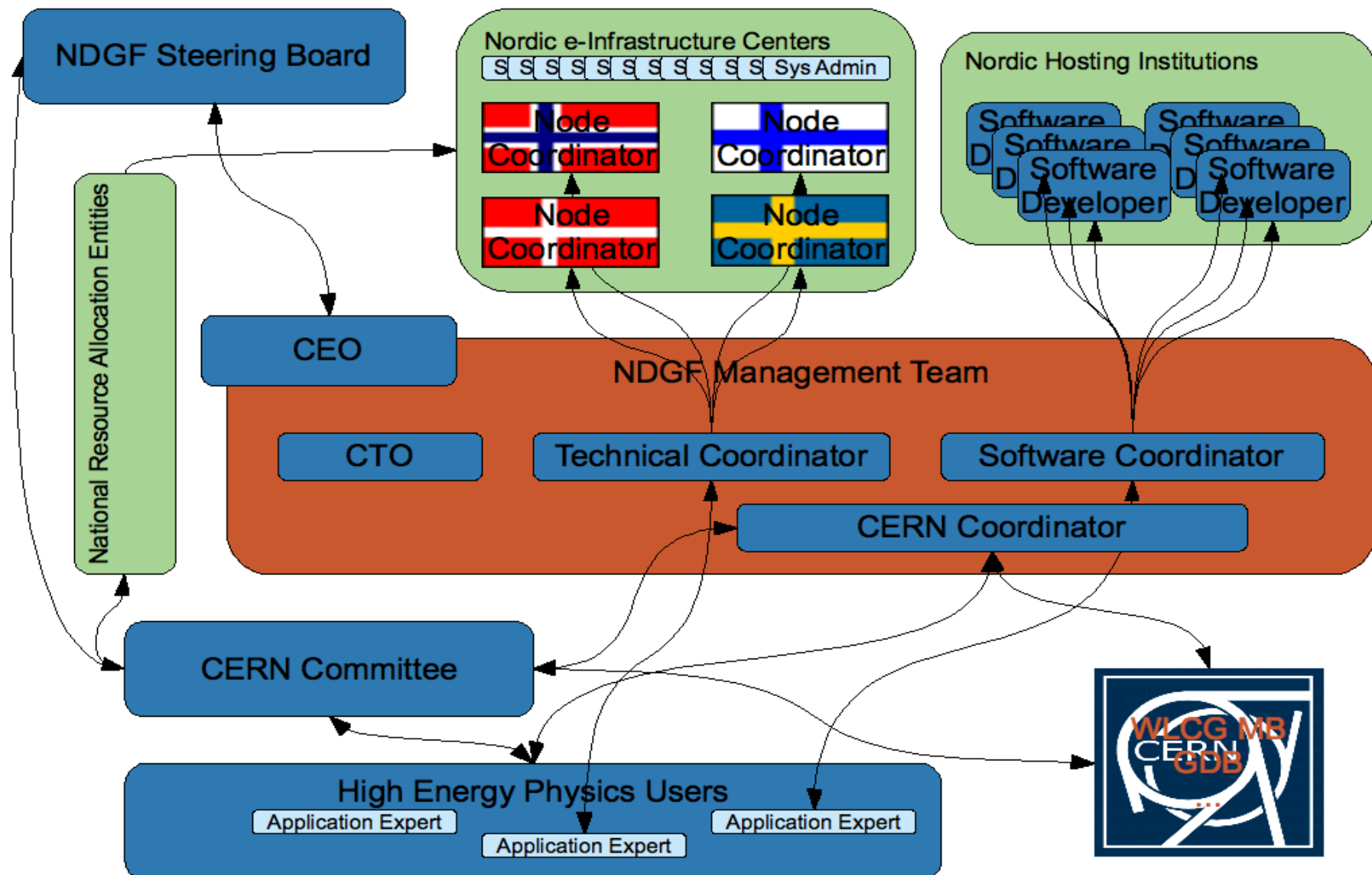


The Nordic Tier-1

- Nordic Countries constitute together 25Mio People
- No country is bigger than 10Mio People
- Nordic Tier-1 needs to utilize hardware at bigger Nordic compute sites
- Strong Nordic grid tradition: NorduGrid / ARC
 - Deployed at all Nordic compute centers
 - Used heavily also by non-HEP users (>75%)
- Need for a pan-Nordic organization for Tier-1 and possibly other huge inter/Nordic e-Science projects

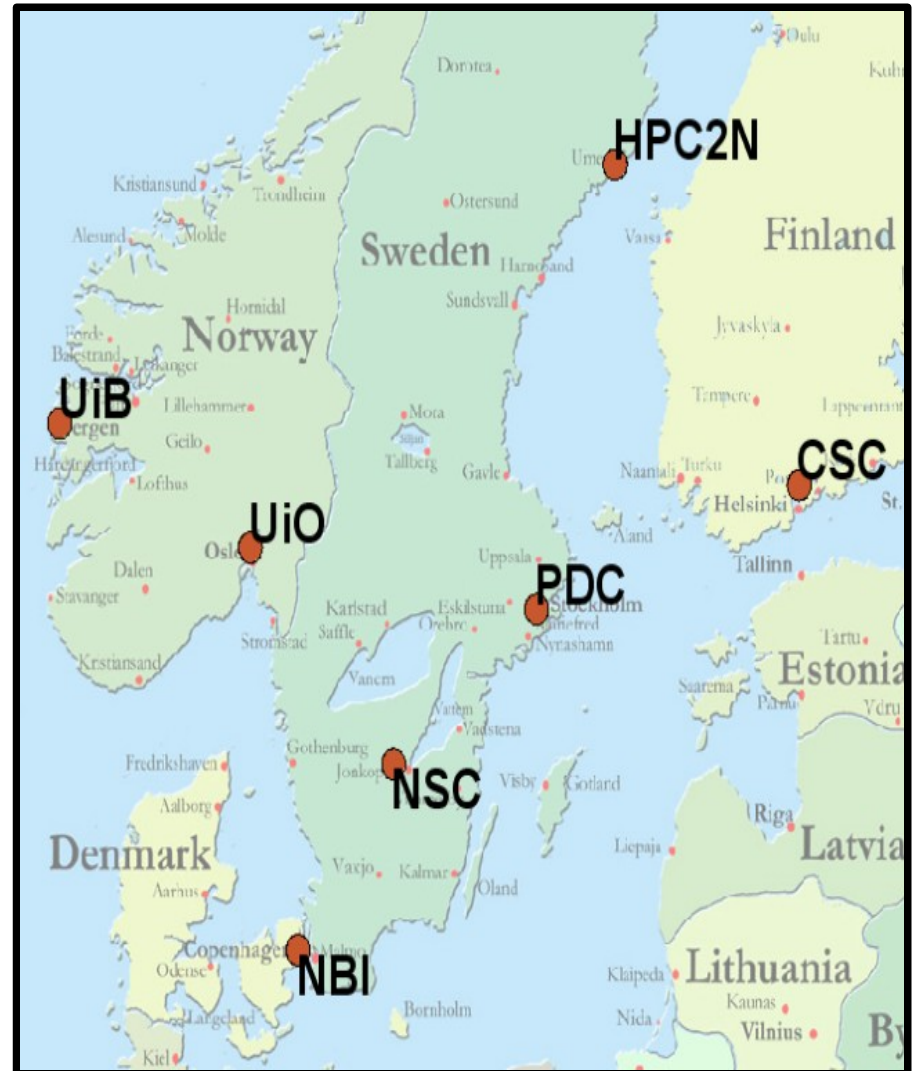
- Many (and sometimes stalled) discussions on a pan-Nordic grid organization resulted in the present NDGF
- Equally funded by the four Nordic Research Councils
- Hosted by the existing pan-Nordic RREN Organization NORDUnet
- Born June the 1st 2006
- Management team operational since September 2006
- Rapid staff ramp up

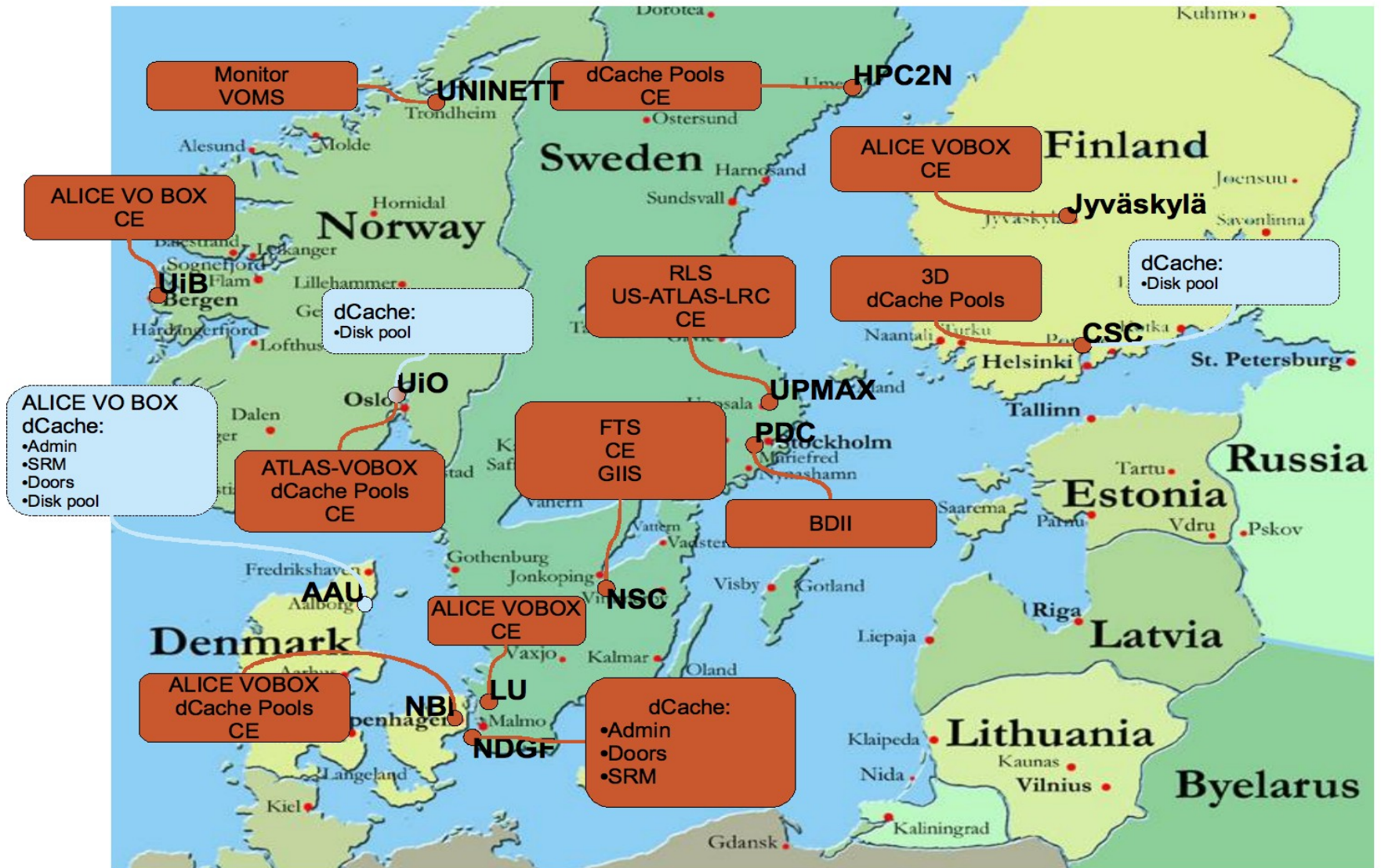
Organization – CERN related



NDGF Tier-1 Resource Centers

- 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 is planned

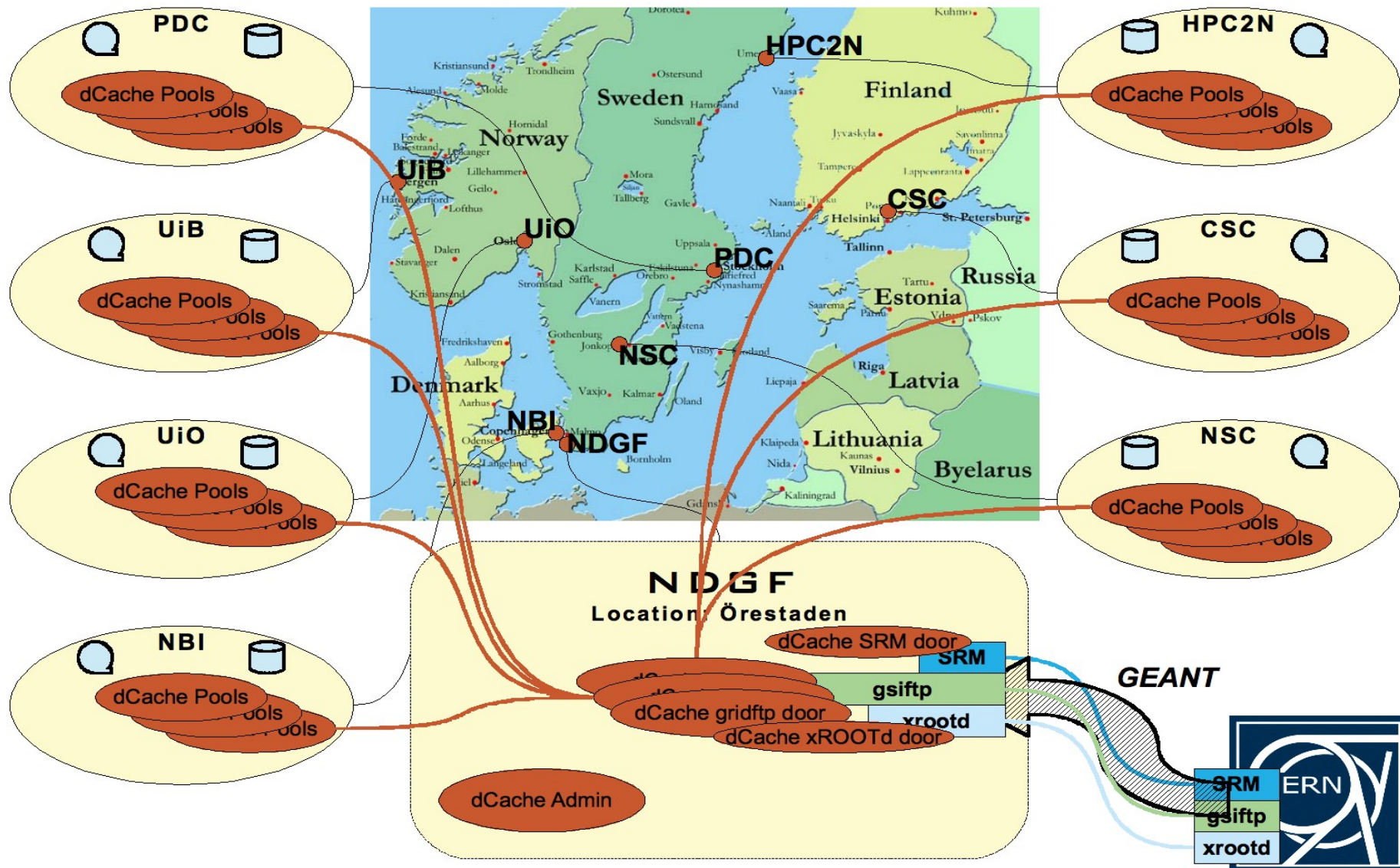
- NorduGrid / ARC middleware for Computing
- Used routinely since 2002 for e.g. ATLAS data challenges
- Deployed at all the Tier-1 sites
- Interoperability tackled pr. VO

Grid Monitor - Microsoft Internet Explorer

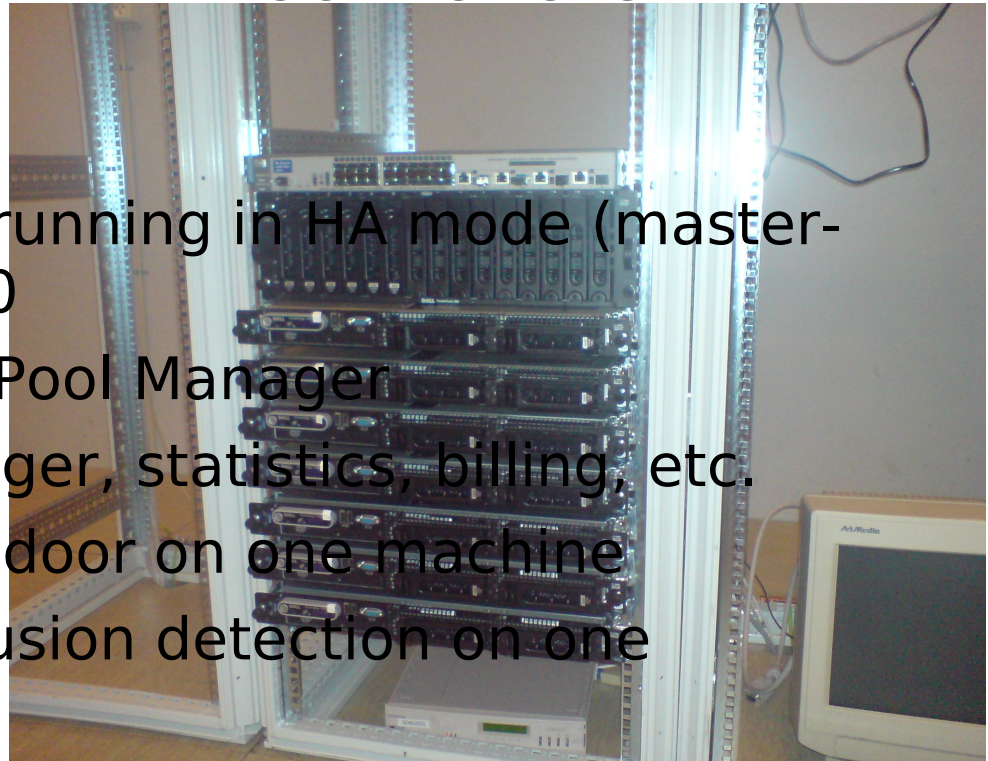
Processes: ■ Grid ■ Local

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)	30	0+6	2+1
	Diablo (DIII)	10	0+0	0+0
Denmark	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/KU)	112	0+42	0+1
Finland	UT/CPH (idea Recorder)	1	1+0 (queue down)	0+0
	UT/MSB Anaxonda clus>	15	3+0	0+0
	UT/CSA Antiochia clus>	20	0+0	0+0
	CMS on CERN Linux	1	0+0	0+0
	CMS Production server	5	0+0	0+0
	CTDOUG Cluster	2	0+0	0+0
	GLUE test cluster	1	0+0	0+0
	EENet cluster	6	0+0	0+0
Germany	UT Physics Cluster	3	3+0	0+0
	CSC Kippu	1	1+0	6+0
	Mill (Physics im)	60	0+15	0+0
	Alpha (HIP)	1	0+0	0+0
Norway	Testbed0 (HIP)	1	0+0	4+1
	FZK cluster	996	83+349	0+0
	LRZ cluster	23	0+230	0+243
	Oslo Temp. Nuste	11	0+0	25+0
Slovenia	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
Sweden	SIGNET	40	6+31	6+0
	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
Switzerland	Sigrid (SweGrid, Luna>	99	49+50	19+25
	Toto7/Whenim64 (Lunar>	192	0+161	0+11
	Bern ATLAS Cluster	8	8+0	12+0
TOTAL		42 sites	5196 570 + 3169	597 + 499

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



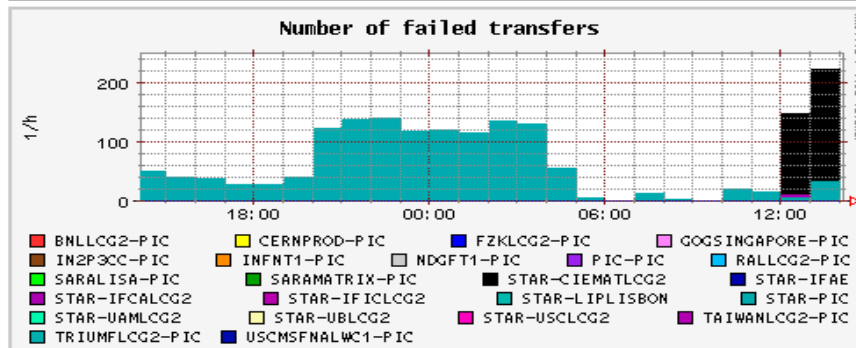
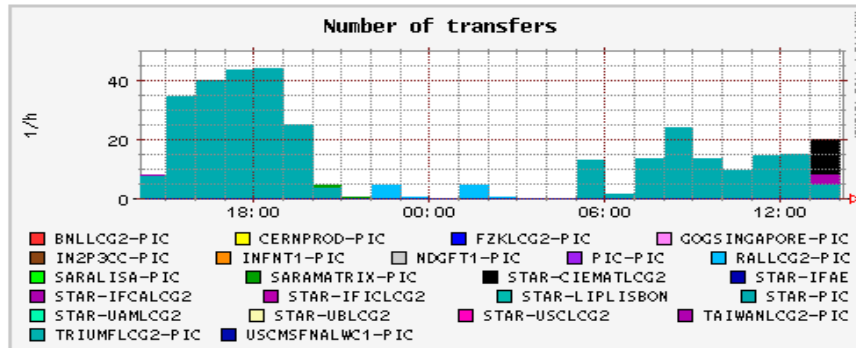
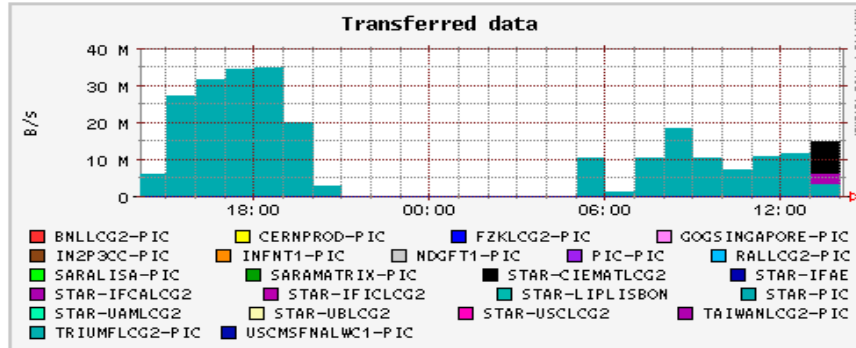
- Central Installation:
 - 7 Dell 1950 2xDual Core 2GHz Xeon, 4GB RAM, 2 x 73GB 15k SAS disks (mirrored) (one for spare)
 - 2 x Dell PowerVault MD-1000 direct attached storage enclosures with 7 x 143GB 15k SAS RAID-10 each
- Running:
 - 2 Postgress for PNFS running in HA mode (master-slave) DB on MD-1000
 - 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



- Operational with two sites:
 - Denmark, Univ. of Copenhagen: 70 TB
 - but currently offline due to SAN issues
 - Norway, Univ. of Oslo: 70TB
- Running whatever OS the local admins want
 - Currently:
 - 64-bit Ubuntu 6.06
 - 64-bit RHEL4
 - 32-bit RHEL4-compatible

- Located in Linköping:
 - 1 Server for FTS
 - 1 Server for Oracle database
- Channels from T1s to NDGF for T1-T1 transfers
- Also being used internally for migrating data, in particular from legacy gsiftp sources
- Web interface to statistics and error messages
 - screenshots in the next two pages

[Last day](#) [Last week](#) [Last 3 months](#) [Last year](#)



- Web interface also deployed at PIC
 - screenshot from this installation as the NDGF fts was boring right now

[Failure summary last 24 hours](#)

[Failure summary last hour](#)

[Failure summary last 10 minutes](#)

Summary of FTS transfer failures last hour

Failures per site

Site	#errors
PIC	749
TAIWAN-LCG2	468
FZK-LCG2	415
RAL-LCG2	401
INFN-T1	388
CERN-PROD	365
IN2P3-CC	266
IFIC-LCG2	176
IFAE	107
NIKHEF-ELPROD	44
SARA-MATRIX	13
CIEMAT-LCG2	9
USCMS-FNAL-WC1	6
GRID002.FT.UAM.ES	6
INFN-MILANO	4
INFN-FRASCATI	3

Grand total for all sites

Colour legend: Problem seems to be at NDGF Problem seems to be at remote site Undecided

#errors Error type

3420 Total number of errors

455 [FINAL:TRANSFER: Transfer failed. ERROR the server sent an error response: 425 425 Cannot open port: java.lang.Exception: Pool manager error: Best pool <dc004 2> too high : 2.](#)

381 [FINAL:TRANSFER: Transfer failed. ERROR the server sent an error response: 425 425 Cannot open port: java.lang.Exception: Pool manager error: Best pool <dc005 1> too high : 2.](#)

273 [FINAL:TRANSFER: Transfer failed. ERROR an end-of-file was reached](#)

64 [FINAL:ABORTED: Operation was aborted \(the gridFTP transfer timed out\).](#)

58 [FINAL:TRANSFER: Transfer failed. ERROR the server sent an error response: 425 425 Cannot open port: java.lang.Exception: Pool manager error: Best pool <dc005 1> too high : 2.](#)

54 [FINAL:SRM DEST: Failed on SRM put: SRM getRequestStatus timed out on put, also failing to do 'advisoryDelete' on target.](#)

45 [FINAL:TRANSFER: Transfer failed. ERROR the server sent an error response: 425 425 Cannot open port: java.lang.Exception: Pool manager error: Best pool <dc020 1> too high : 2.](#)

42 [FINAL:SRM SOURCE: Failed on SRM get: SRM getRequestStatus timed out on get](#)

32 [FINAL:TRANSFER: Transfer failed. ERROR the server sent an error response: 500 500 java.lang.reflect.InvocationTargetException: <stor>](#)

26 [FINAL:TRANSFER: Transfer failed. ERROR the server sent an error response: 425 425 Cannot open port: java.lang.Exception: Pool manager error: Best pool <dc020 2> too high : 2.](#)

21 [FINAL:TRANSFER: Transfer failed. ERROR the server sent an error response: 425 425 Can't open data connection. timed out\(\) failed.](#)

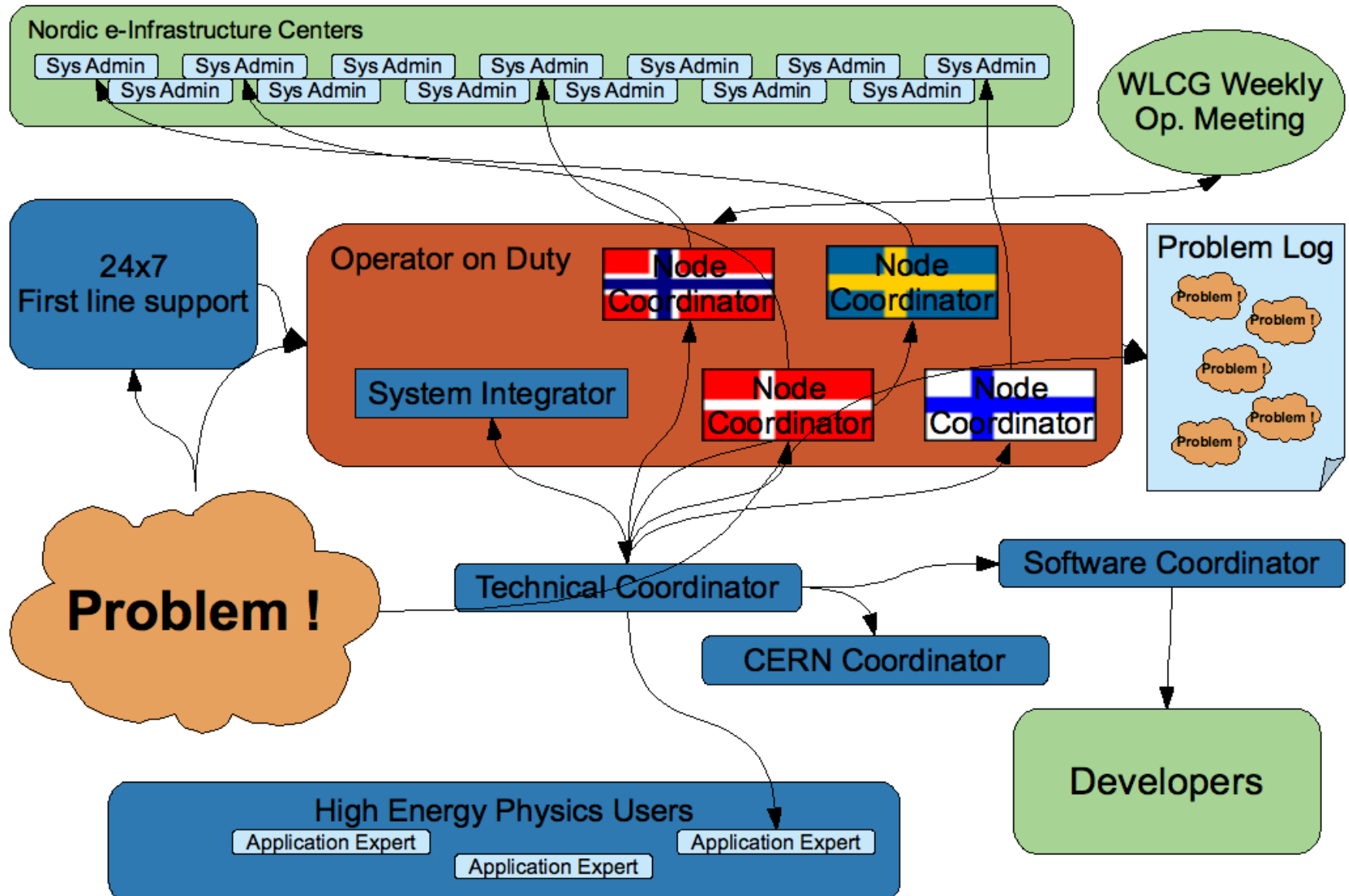
6 [FINAL:TRANSFER: Transfer failed. ERROR the server sent an error response: 500 500 java.lang.reflect.InvocationTargetException: <retr>](#)

0 [FINAL:SRM SOURCE: Failed on SRM get: Failed on SRM get on \(TRANSFER URL\): id failed: null - Error in RemoteFileAccessImpl: Failed with exception: \(TRANSFER URL\) access failed : 500.](#)

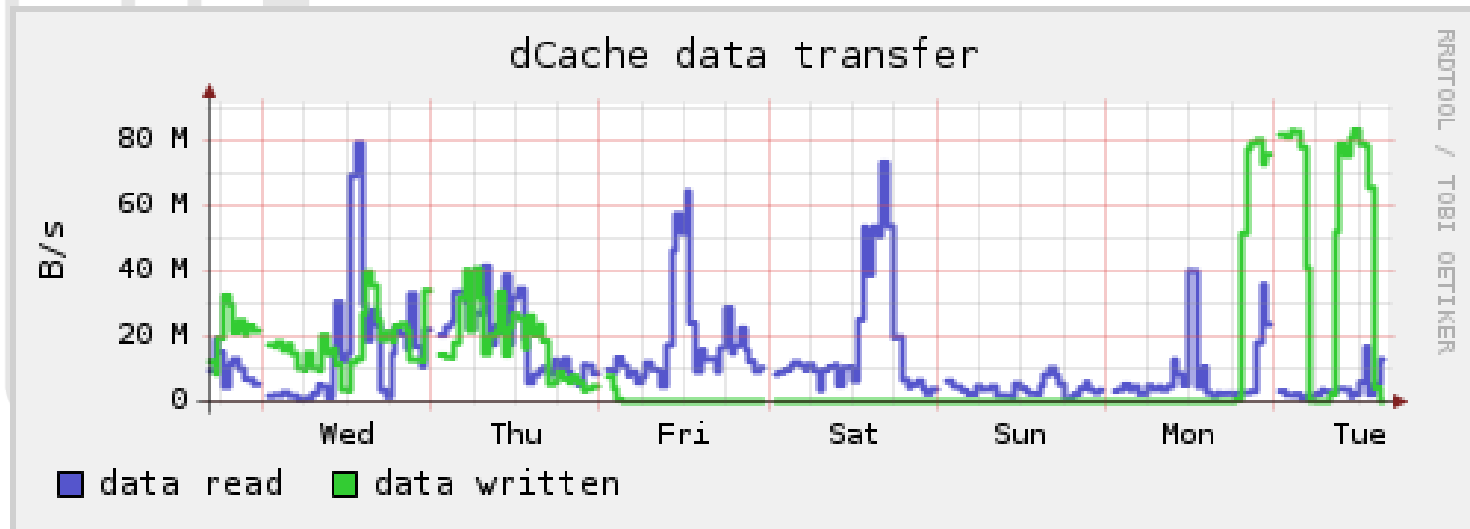
- Initial setup located in Helsinki:
 - One dual core dual Xeon box with 4GB of memory
 - no RAC, just one server
 - High availability SAN storage
 - a bit more than one TB of space allocated for data, more to come after March

- ATLAS VOBox (ARC flavor) services fully implemented
 - ARC uses Globus RLS
 - US-ATLAS-LRC mysql records almost identical to Globus RLS
 - Online sync'ed US-ATLAS-LRC and RLS in production since November
 - Enables outside ATLAS subscription to data stored on old SEs
- In operation since Nov 2006

- ALICE VOBox boxes:
 - Installed at:
 - Univ. of Copenhagen, Univ. of Bergen, Univ. of Aalborg and National Supercomp. Center in Sweden.
- Integration with ARC and the distributed NDGF Storage ongoing



- Current usage (beginning of June, 2007)
 - ATLAS data being read and written from compute jobs on the grid
 - ATLAS tier0-tier1 transfer tests



- CPU usage is depending on available participant clusters
 - Last weekend Atlas had >1200 concurrent jobs running on NorduGrid, a large part of those on NDGF-affiliated resources
 - The day-to-day computational jobs does not require much participation from NDGF operations – as long as the core services are running (dcache, rls, giis, etc)

- Interfacing with EGEE for operations
 - Read and respond to GGUS tickets
 - SAM tests running on NDGF
 - Cooperating with NEROC

- NorduGrid monitor for computing
- Storage – distributed dCache
 - ganglia graphing of involved nodes and some dCache sources
 - email and jabber communication with local admins and NDGF staff
 - CIC portal and GOCDB for downtime announcements etc
- Dozens of webpages for service status

- RT or mailinglist for user problems
- GGUS for COD tickets
 - And the occasional EGEE user reading data from srm.ndgf.org
- Nagios being set up
 - will share probes on the sysadmin wiki
- 24/7 first line being set up
 - with experts on call during weekends

- Rapidly ramping up diskspace within the next few months
- Large computational resources planned for end of 2007
- In 2010 we will have more than 4MSI2k, and 4PB of storage