NorduGrid

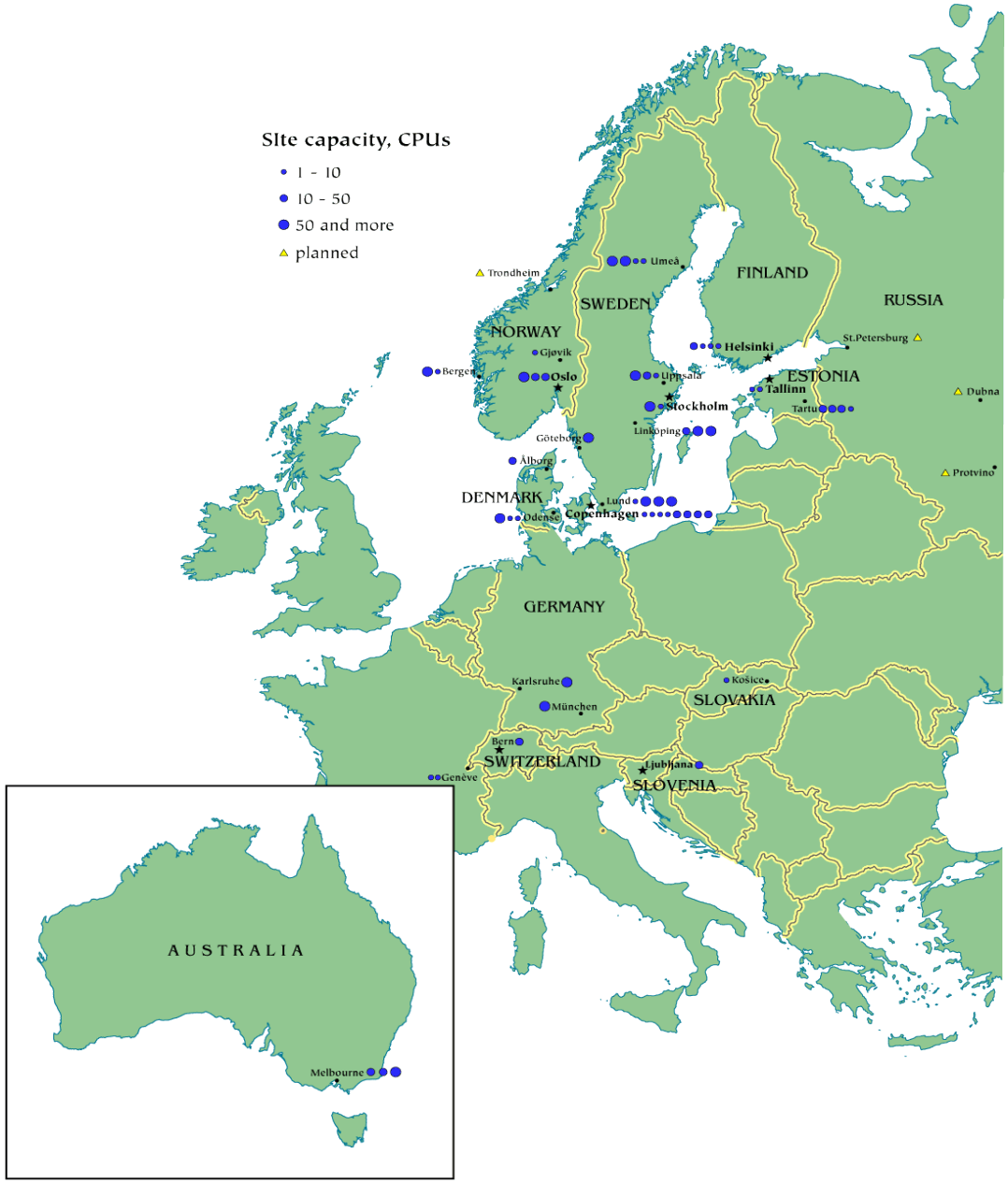
3 years of building Grid-like infrastructure in Nordic countries

Presented by
Aleksandr Konstantinov ***on behalf of NorduGrid collaboration***
Vilnius University/Lithuania and University of Oslo/Norway

GlobusWORLD 2005
February 7-11, Boston, Massachusetts, USA

NorduGrid

- NorduGrid is a research collaboration established by universities in Denmark, Estonia, Finland, Norway and Sweden
  - Focuses on providing production-capable Grid-like middleware for academic researchers
  - Currently supports one of the largest Grid production systems
    - 10 countries, 40+ sites, ~4000 CPUs, ~30 TB storage



ARC

- ARC (Advanced Resource Connector) is the Grid middleware developed by the NorduGrid
  - Collection of tools and services
  - Based on Globus Toolkit™ 2 libraries and services
    - Can be built using GT3 pre-WS code too
  - Using available utilities whenever possible
    - Some services which could/can not provide functionality outlined in NorduGrid architecture were replaced
    - Some other services not provided in Globus Toolkit™ 2 were developed

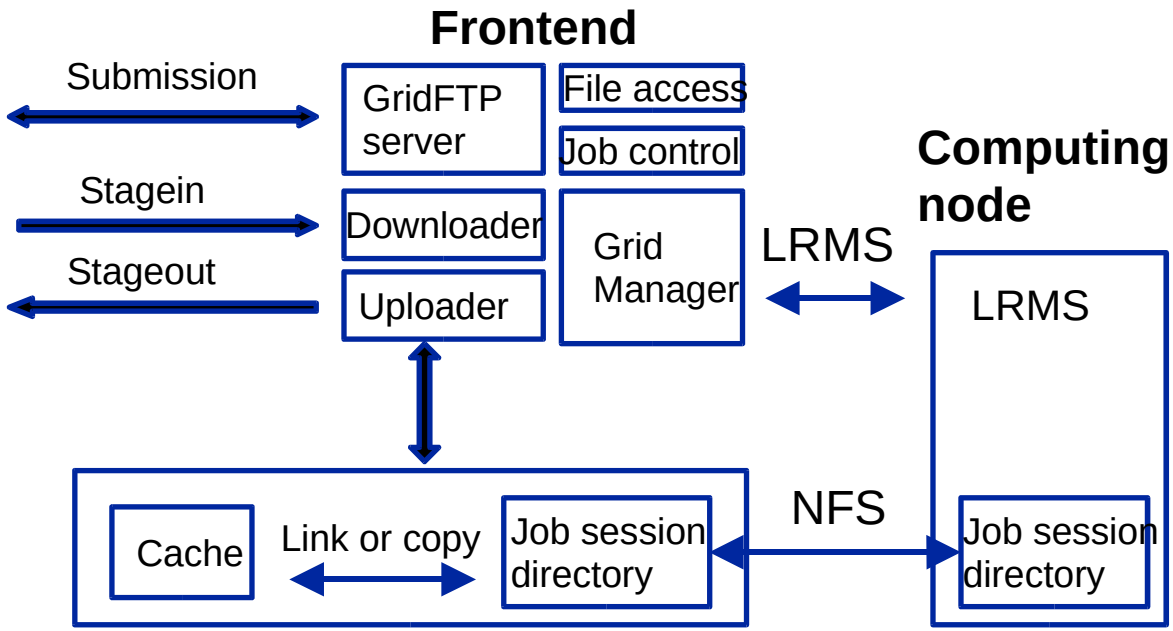
- Initial development principles:
  - Simple
  - Stable
  - Non-invasive

ARC components

- Job management
  - Lightweight **U**ser **I**nterface with built-in Personal Resource Broker
    - Basic and complete support for single job management
    - Basic functionality for data management
  - Resource frontend – **G**rid **M**anager – accessible through GridFTP interface
- Information
  - Information System based on Globus MDS 2 with a modified schema
  - WS based logging service – Logger
- Data management
  - Supported Indexing Services include legacy Globus' **R**eplica **C**atalog (RC) and **R**eplica **L**ocation **S**ervice (RLS)
    - All operations involving Indexing Services are done automatically
    - Looking for better solutions
  - GridFTP server implementation with pluggable backends
  - "**S**mart" Storage **E**lement (SSE) – Web Service based data service with direct control of Indexing Services
  - Every piece of code has support for full set of protocols
- System monitoring
  - Web interface to current state of the system – Grid Monitor
  - System's history and statistics – NGLogger

ARC – Grid Manager

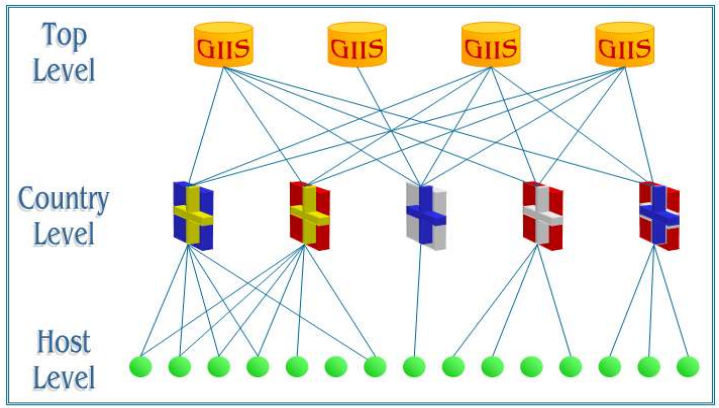
- Solution similar to Globus' gatekeeper reimplemented in order to have needed functionality
- GridFTP interface for job control
  - Each job is presented as virtual subdirectory
  - FTP commands are mapped to job management operations
- Handles pre- and post-staging of files with integrated support for data indexing services (RC, RLS).
- Shared cache of pre-staged files with automatic registration in Indexing Services
- Application-specific **R**untime **E**nvironments
- Limitations due to architecture – data staging only at beginning and end of job

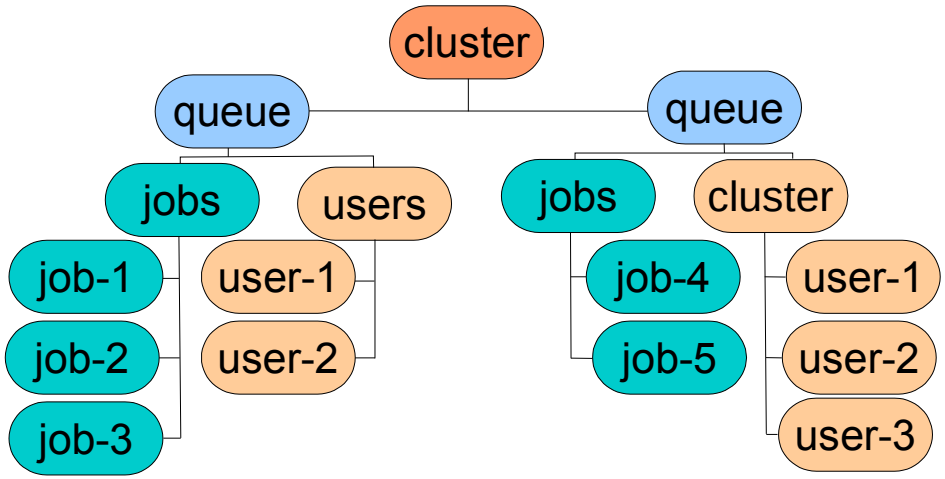


ARC –Information System

- Uses Globus' MDS 2.2
  - Soft-state registration allows creation of dynamic structure
  - Multi-rooted tree
  - GIIS caching is not used by the clients due to buggy implementation
  - Several patches and bug fixes are applied
- Mostly cluster-oriented
  - A new schema and information providers were developed, to serve clusters
- All queries are anonymous
  - Authenticated queries are very inflexible
- Not very scalable
- Looking for new solution now

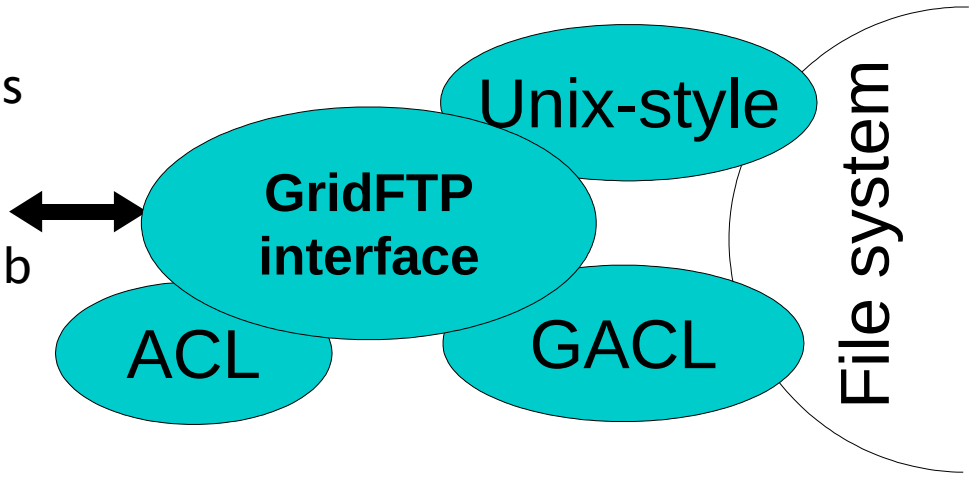
NorduGrid Hierarchy





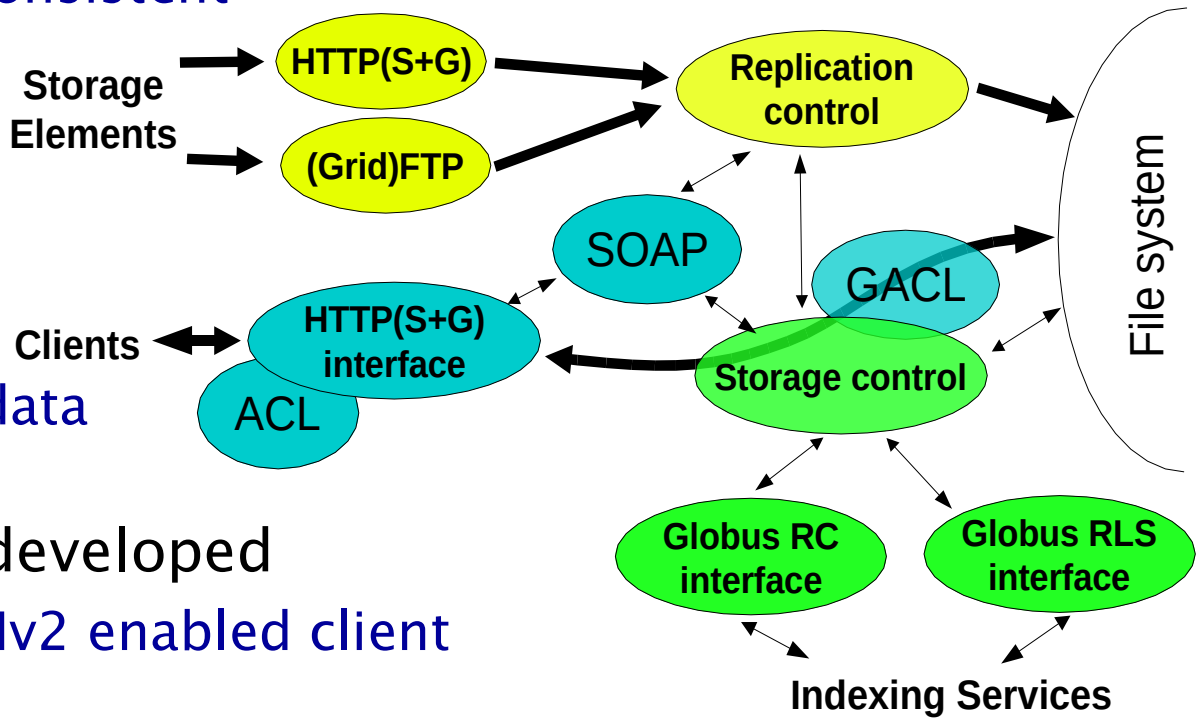
ARC – GridFTP server

- Own implementation of GridFTP server
  - Protocol is implemented using globus_ftp_control library from Globus Toolkit™
  - User-dependent virtual tree of directories with flexible access rules evaluated against
    - Grid identity of the user
    - **V**irtual **O**rganization **M**embership **S**ervice credentials
    - Any external module
  - Local file/object access is through pluggable modules
    - ***fileplugin*** – ordinary file access with static access rules
      - Based on Grid-UNIX identity mapping
      - Based on Grid identity only
    - ***gaclplugin*** – access to each object is controlled through GACL object maintained by object's owner
    - ***jobplugin*** – control/access user's job



ARC – "Smart" Storage Element

- Uses HTTPS/HTTPG + SOAP
  - Firewall friendly
- Integrated flexible access control
  - Per stored object
  - Evaluated against Grid identity of a client
- Direct interface to data indexing services
  - Indices are kept more consistent
- Data transfer tasks
  - Integrated support for data replication
- SRM v2 interface being developed
  - Still waiting for any SRMv2 enabled client



ARC – User Interface

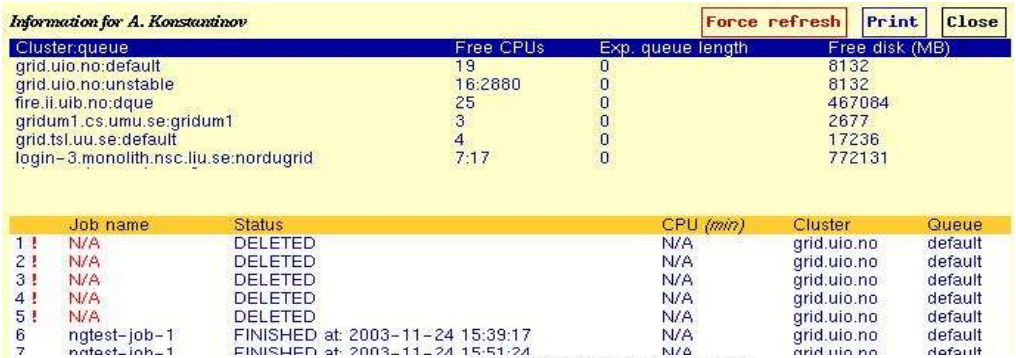
- Collection of command line utilities:
  - ***ngsub*** – find suitable resource and start job
  - ngstat – check status of job
  - ngcat – monitor job by looking at its stdout/stderr
  - ngget – get results of finished job
  - ngkill – stop job
  - ngclean – delete job from computing resource
  - ngsync – find user's jobs
  - ngrenew – update remote credentials
  - ngls – list files on storage element or in job's directory
  - ngcopy – data transfer
  - ngrequest – third-party transfers or data tasks
  - ngremove – delete remote files

ARC – User Interface (cont.)

- Contains a Personal Resource Broker – job submission sequence:
  - Collects information about Computing Resources (clusters) through MDS network
  - Collects information about requested input data (size and availability at Computing Resources if available)
  - Collects Information about resources at Computing Resources
  - Selects suitable Computing Resource and submits job request
    - 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 RSL string (max CPU time, required free disk space, installed software etc.)
    - If the job requires a file that is registered in a data indexing service, 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 resources or shortest queue
  - Uploads locally available input data

ARC – Grid Monitor

- Web based interface captures current state of the system
  - Implemented in PHP
  - Very rich interface
  - From statistics to detailed view
    - Summary per cluster
    - Jobs per cluster
    - Jobs per user
    - etc.



Information for A. Konstantinov Force refresh Print Close

Cluster:queue	Free CPUs	Exp. queue length	Free disk (MB)
grid.uio.no:default	19	0	8132
grid.uio.no:unstable	16:2880	0	8132
fire.ii.uib.no:dque	25	0	467084
gridum1.cs.umu.se:gridum1	3	0	2677
grid.tsl.uu.se:default	4	0	17236
login-3.monolith.nsc.liu.se:nordugrid	7:17	0	772131

	Job name	Status	CPU *(min)*	Cluster	Queue
1 !	N/A	DELETED	N/A	grid.uio.no	default
2 !	N/A	DELETED	N/A	grid.uio.no	default
3 !	N/A	DELETED	N/A	grid.uio.no	default
4 !	N/A	DELETED	N/A	grid.uio.no	default
5 !	N/A	DELETED	N/A	grid.uio.no	default
6	ngtest-job-1	FINISHED at: 2003-11-24 15:39:17	N/A	grid.uio.no	default

NorduGrid Storage Elements

2003-11-27 CET 00:42:16 Force refresh Print Close

#	Alias	Space (MB)	Name	Base URL	Type
1	**Oslo SE dc1**	0	dc1.uio.no	gsiftp://dc1.uio.no:2811/dc1/	gridftp-ba
2	**Oslo SE dc4**	616169	dc4.uio.no	gsiftp://dc4.uio.no:2811/dc1/	gridftp-ba
3	**Oslo SE dc2**	845	dc2.uio.no	gsiftp://dc2.uio.no:2811/dc1/	gridftp-ba
4	**Oslo SE dc3**	959730	dc3.uio.no	gsiftp://dc3.uio.no:2811/dc1/	gridftp-ba
5	**Oslo Test SE dc>**	959730	dc3.uio.no	gsiftp://dc3.uio.no:2811/test/	gridftp-ba
6	**Uppsala Atlas M>**	55	atlas:grid.tsl.uu.se	gsiftp://grid.tsl.uu.se:2811/atlas	gridftp-ba
7	**Uppsala General>**	17236	files:grid.tsl.uu.se	gsiftp://grid.tsl.uu.se:2811/files	gridftp-ba

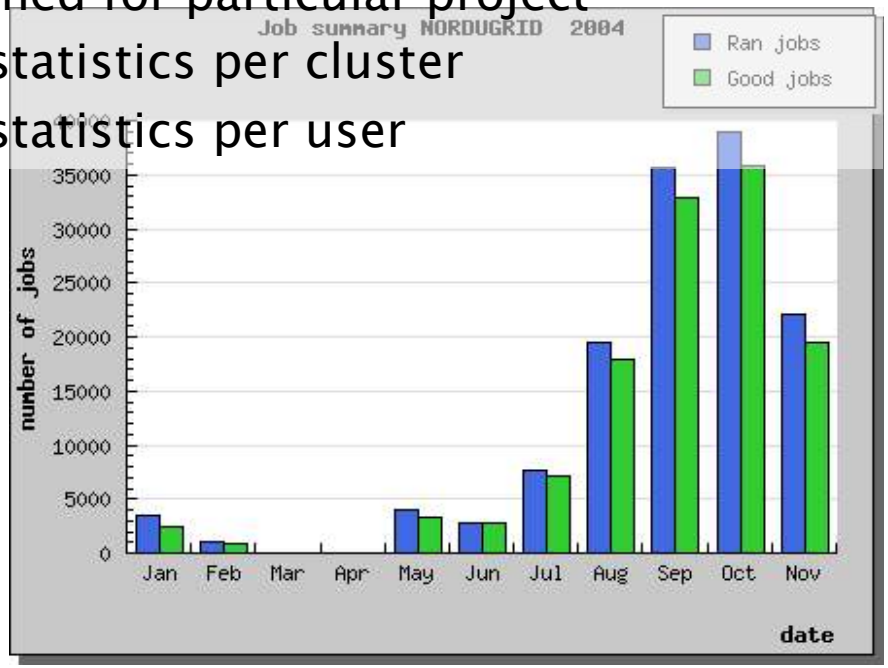
2004-09-24 CEST 00:47:24

Processes: Grid Local

Country	Site	CPUs	Load (processes: Grid+local)	Queueing
Australia	Atlas (UniMelb)	26	0+2	0+0
	Charm (UniMelb)	19	0+0	0+0
	Alfred (UniMelb)	88	0+26	8+1
Denmark	DistLab (DIKU)	10	0+0	0+0
	Aalborg Grid Gateway	46	46+0	49+0
	Horseshoe (DCSC/SDU)	1088	9+646	0+2
	HEPAX1	1	0+0	0+0
	Morpheus	18	17+0	16+0
	Theory (DCSC/KU)	112	0+52	0+0
	VCR (VideoRecorder)	1	1+0 (queue down)	0+0
Estonia	UT IMCB Anakonda clus>	15	6+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	2+0	0+0
	UT Physics Cluster	17	3+0	0+0
Finland	CSC Kirppu	1	0+0	0+0
	Mill (Physicum)	64	0+7	0+0
	Alpha (HIP)	1	0+0	0+0
	Testbed0 (HIP)	1	0+0	4+1
Germany	FZK cluster	886	176+153	0+0
	LRZ cluster	234	0+222	34+366
Norway	Oslo Temp Cluster	13	0+0	5+0
	Parallab IBM Cluster	58	6+51	2+69
	Oslo Grid Cluster	43	16+1	37+0
	UiO Grid	93	48+43	0+41
Slovakia	UPJS GRID	1	0+0(no queue info)	0+0
Slovenia	SiGNET	51	23+28	77+0
	Bluesmoke (Swegrid,NS>	98	94+0	545+0
	osufy farm	60	58+0	90+0
	V	4	4+0	2+0
	agrid (SweGrid, Uppm>	0		372+0
	grid (SweGrid,HPC2N)	101	101+0	528+0
	onolith (NSC)	400	0+328 (queue down)	0+132
	uark Cluster	7	3+0	3+0
	ppe (SweGrid PDC KT>	96	92+0	279+0
	grid (SweGrid, Luna>	99	67+32	42+43
	oto7/Whenim64 (Lunar>	191	0+157	0+632
	ern ATLAS Cluster	8	8+0	12+0
	47 sites	3986	884 + 1748	2105 + 1287

ARC – Logger and NGLogger

- Logger
  - Central MySQL database with WS interface
  - Information is pushed from Grid Managers
- NGLogger
  - Web interface to Logger's database
  - Complementary to Grid Monitor
  - Shows statistics computed from history
    - Can be tuned for particular project
    - Provides statistics per cluster
    - Provides statistics per user



Number of jobs allowed to run	4166	
Good jobs	3860	32.7 %
Failed in files download (pre-processing)	120	2.3 %
Job submission to LRMS failed	86	2.1 %
Job exit code is 127	37	0.9 %
Job exit code is 6	20	0.5 %
Job exit code is 1	10	0.2 %
Job exit code is 271	8	0.2 %
Failed running plugin at state PREPARING	6	0.1 %
Job exit code is 255	6	0.1 %
	5	0.1 %
Job exit code is -4	3	0.1 %
Job exit code is 2	3	0.1 %
Job exit code is -1	1	0.0 %
Job exit code is 155	1	0.0 %
User requested to cancel the job (% of the total number of submitted jobs)	136	3.2 %
Total number of jobs	4302	

ARC – Ongoing efforts

- arcLib – effort to create well defined set of C++ classes to interface ARC-enabled resources
  - Currently being developed
  - Already minimal functionality available
  - Will provide all UI functionality in a flexible and portable way
  - ***Pluggable brokering algorithms***
- jarcLib – set of Java classes
  - Because ARC is mostly written in C/C++ Java developers had problems with adding ARC support into their products
  - Implements only basic functionality
  - Uses Java CoG – causes problems due to incompatibilty with mainstream Globus Toolkit™
- Distributed data and replicas indexing service.

ARC – Middleware status

- Current stable release: 0.4.4 (2004-10-18)
  - GPL license
  - Binaries available for 12+ Linux distributions
  - Builds on top of Globus Toolkit 2 and 3 – NorduGrid-patched Globus Toolkit is preferred for stability
  - EDG VOMS integrated (voms-1.1.39-5ng)
  - Being maintained
- Current development series: 0.5.x
  - Contains the "Smart" Storage Element and other newly introduced features
  - Quite stable already – some sites are using it
- Anybody is free to use; best-effort support is guaranteed
  - Support: nordugrid-support@nordugrid.org
  - Download at http://ftp.nordugrid.org and http://cvs.nordugrid.org
  - Bug reports: http://bugzilla.nordugrid.org
- Any constructive contribution is welcome
  - Join nordugrid-discuss@nordugrid.org (**very** busy list!)