EGEE and gLite

Andreas Unterkircher

openlab summer student lecture
CERN, Geneva 15 July 2008

Before we start…

- This talk mainly contains slides from other people (Markus Schulz, Maite Baroso, Diana Bosio, Oliver Keeble and many others)
- This talk is not about grid computing in general but focuses on the EGEE project and the gLite middleware
- This talk is not about LCG (will be treated in a separate lecture)

Content

- EGEE project and applications other than LCG
- gLite
- EGEE operations
- EGEE integration, testing and releases
- EGEEprojectstructure

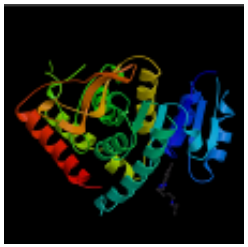
EGEE-III Goals

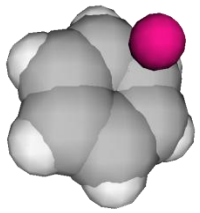
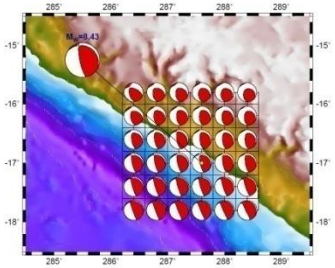
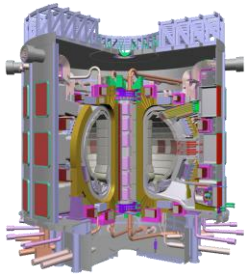
- **Maintain, enhance and simplify the use of the production quality computing infrastructure for an increasing range of researchers in diverse scientific fields**

- **Prepare the transition towards a sustainable infrastructure**

Reported Applications

- **1st year**
  - Growth in reported apps.
- **2nd year**
  - Transition: prototype to production



	6/2006	2/2007	1/2008
Astron. & Astrophysics	2	8	9
Comp. Chemistry	6	27	21
Earth Science	16	16	18
Fusion	2	3	4
High-Energy Physics	9	11	7
Life Sciences	23	39	37
Others	4	14	21
Total	**62**	**118**	**117**

Condensed Matter Physics
Comp. Fluid Dynamics
Computer Science/Tools
Civil Protection

Examples of Scientific Results

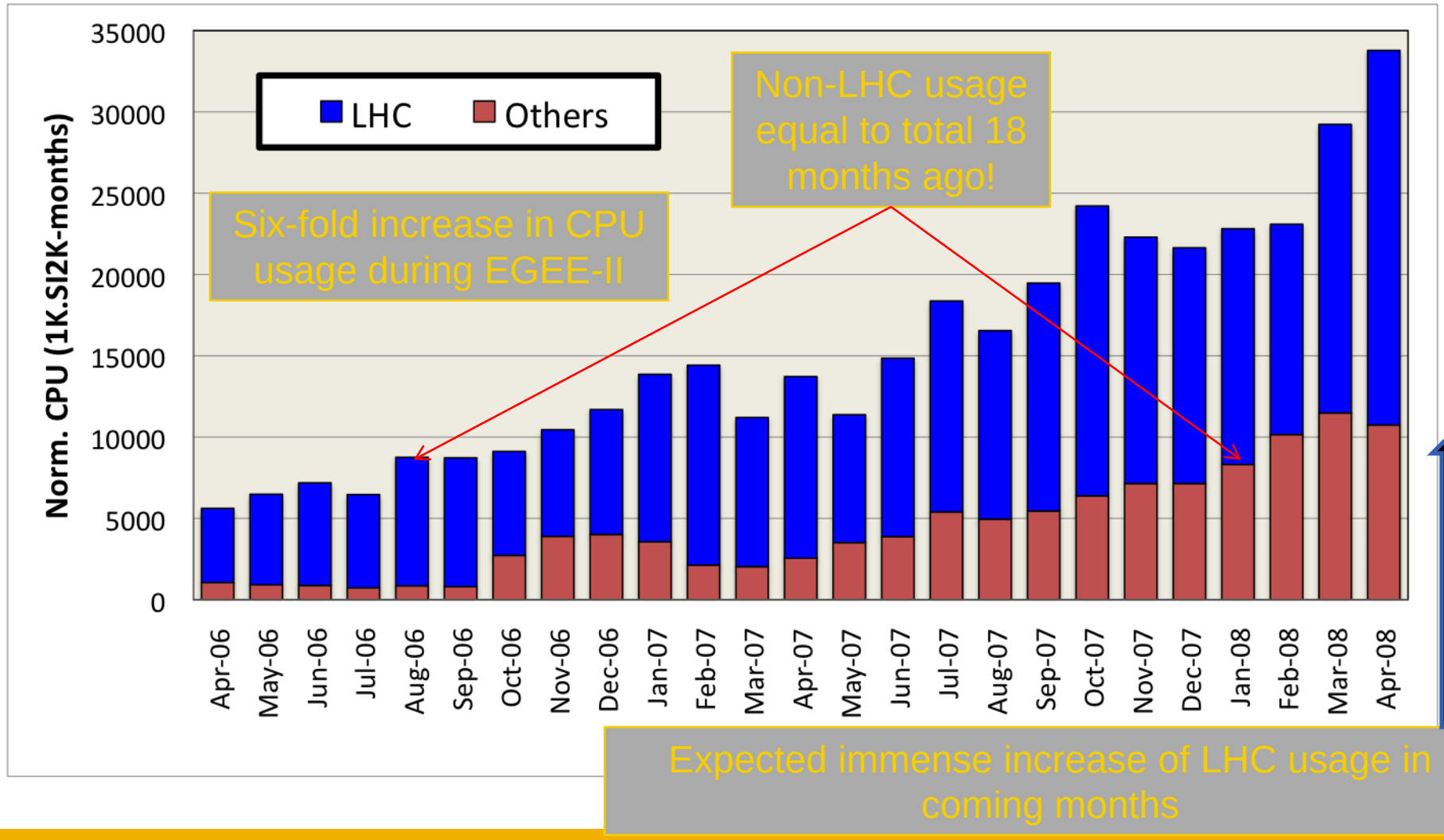
- **Earth Science**
  - Seismic noise calculation.
- **Fusion**
  - Ion kinetic transfer
  - Simulation of wall interactions
  - Stellarator optimization
- **Drug Discovery (WISDOM)**
  - Malaria: 6/30 compounds similar or better than PepstatinA
  - Avian flu: 20% of compounds better than Tamiflu
  - Ongoing tests with compounds from later calculations.

EGEE has been the driving force for achieving these scientific results by

- **Providing access to large amount of reliable computing resources**
- **Helping the establishment of new collaborations**

CPU Utilization

- **Recent level equal to ~32000 CPUs in continuous use.**

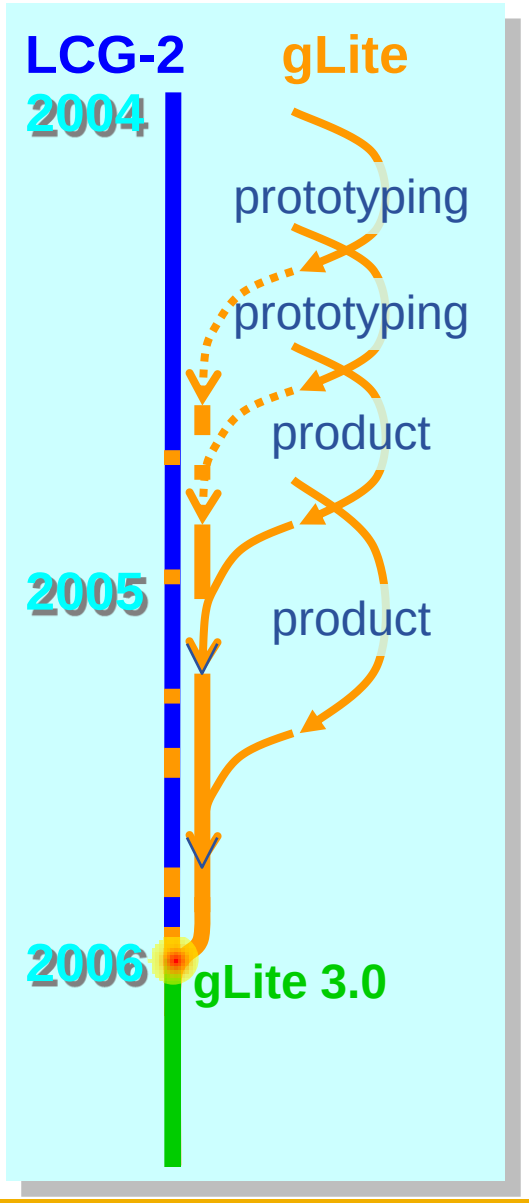


Content

- EGEE project and applications other than LCG
- **gLite**
- EGEE operations
- EGEE integration, testing and releases
- EGEEprojectstructure

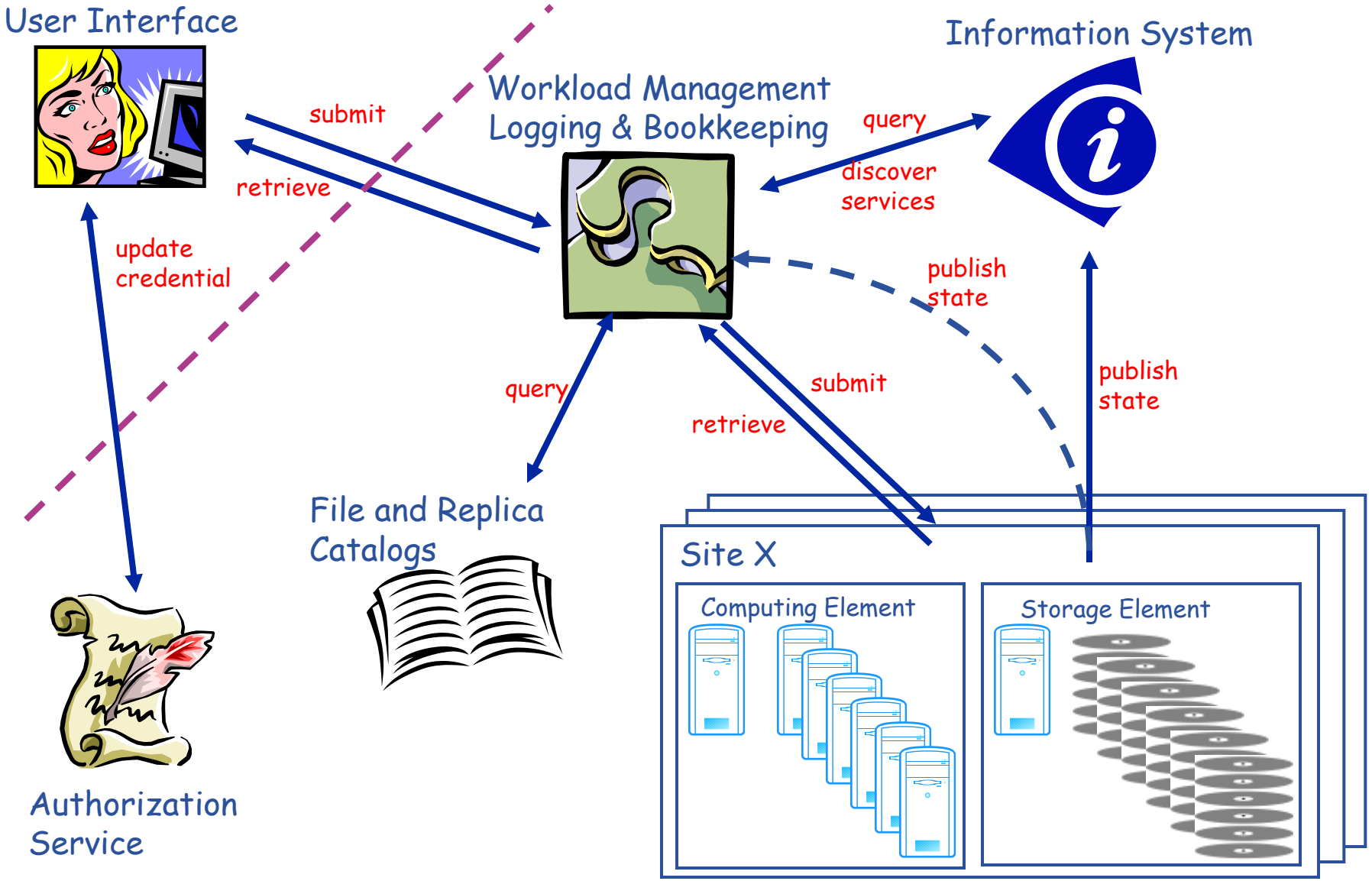
gLite Middleware Distribution

- **Combines components from different providers**
  - Condor and Globus 2 (via VDT)
  - LCG
  - EDG/EGEE
  - Others
- **After prototyping phases in 2004 and 2005 convergence with LCG-2 distribution reached in May 2006**
  - gLite 3.0
- **Focus on providing a deployable MW distribution for EGEE production service**

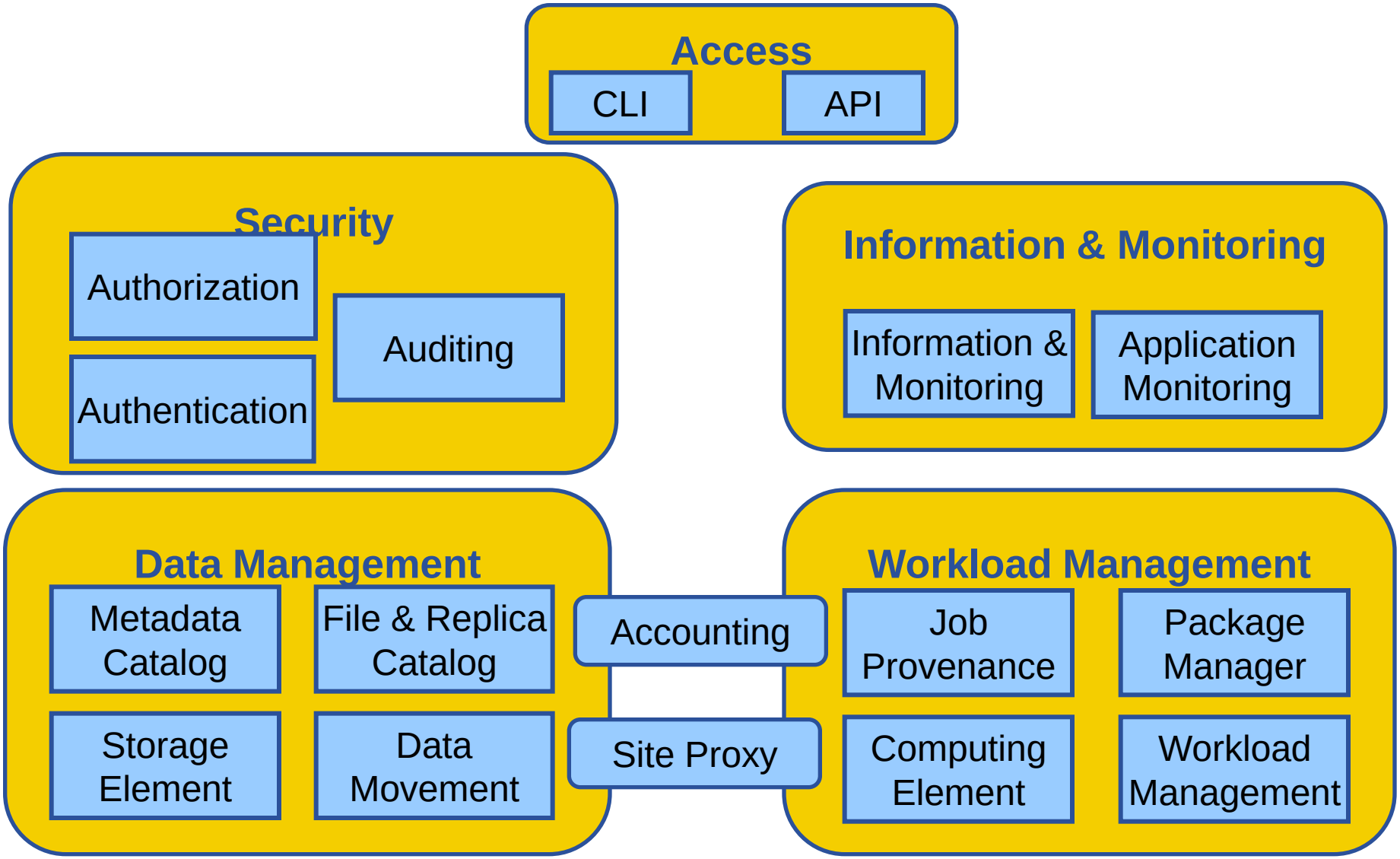


gLite
Lightweight Middleware for Grid Computing

The big picture

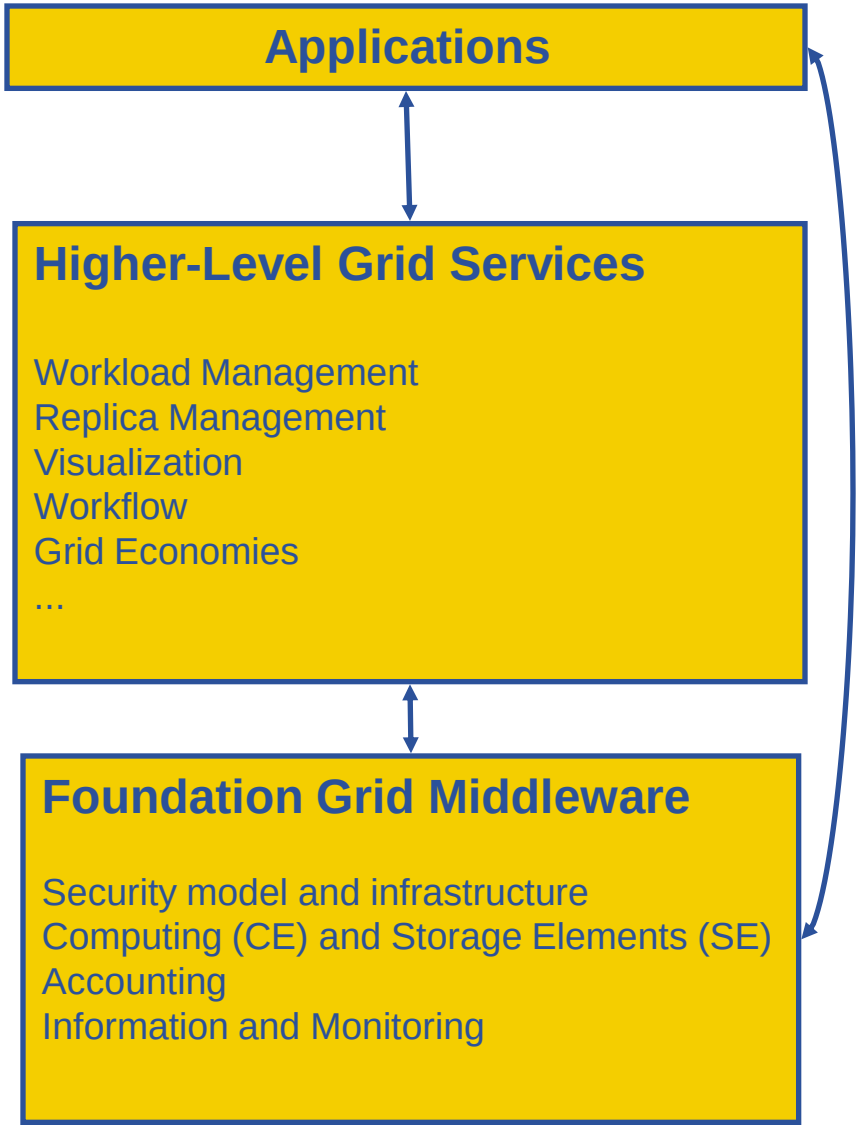


gLite Grid Middleware Services



Overview paper http://doc.cern.ch//archive/electronic/egee/tr/egee-tr-2006-001.pdf

Middleware structure



- **Applications have access both to Higher-level Grid Services and to Foundation Grid Middleware**
- **Higher-Level Grid Services are supposed to help the users building their computing infrastructure but should not be mandatory**
- **Foundation Grid Middleware will be deployed on the EGEE infrastructure**
  - Must be complete and robust
  - Should allow interoperation with other major grid infrastructures
  - Should not assume the use of Higher-Level Grid Services

Services and Clients

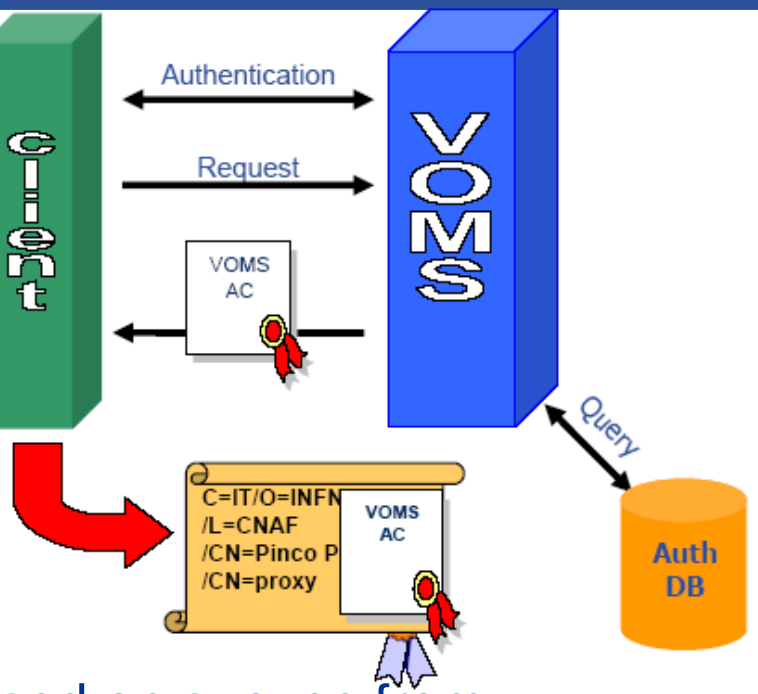
- Authentication and authorization: Myproxy, Voms
- Resource access: Computing Element, Worker Node, User Interface
- Workload Management system
- Logging and Bookkeeping
- Accounting
- Data Management: LCG File Catalog, Storage Element (DPM, dCache), GFAL, FTS, AMGA, Hydra
- Information Systems: BDII

Authentication

- **Authentication is based on X.509 PKI infrastructure**
  - Certificate Authorities (CA) issue (long lived) certificates identifying individuals (much like a passport)
    - Commonly used in web browsers to authenticate to sites
  - Trust between CAs and sites is established (offline)
  - In order to reduce vulnerability, on the Grid user identification is done by using (short lived) proxies of their certificates
- **Short-Lived Credential Services (SLCS)**
  - issue short lived certificates or proxies to its local users
    - e.g. from Kerberos or from Shibboleth credentials (new in EGEE II)
- **Proxies can**
  - Be delegated to a service such that it can act on the user's behalf
  - Be stored in an external proxy store (MyProxy)
  - Be renewed (in case they are about to expire)
  - Include additional attributes

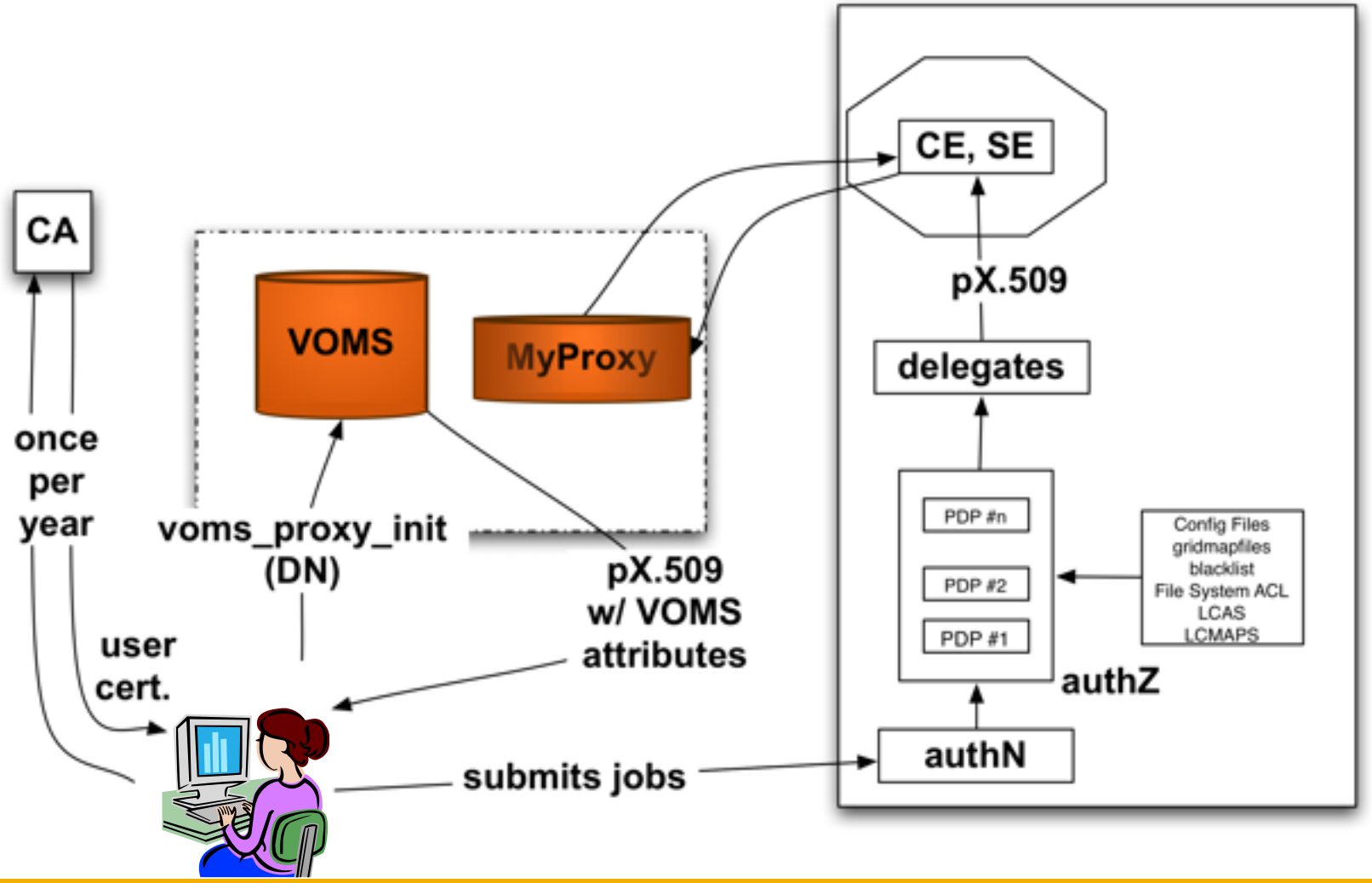
Authorization

- **VOMS service issues Attribute Certificates that are attached to certificate proxies**
  - Provide users with additional capabilities defined by the Virtual Organization
  - Base for the Authorization process
- **Authorization: via mapping to a local user on the resource or token**



  - glexec changes the local identity (based on suexec from Apache)
  - LCAS/LCMAPS use different plug-ins to determine if and how to map a grid user to a local user
    - mainly used for C-based applications
  - gLite Java Authorization Framework (XACML-compatible)
    - mainly used for Java-based applications

Security - overview



Services and Clients

- Authentication and authorization: Myproxy, Voms
- **Resource access: Computing Element, Worker Node, User Interface**
- Workload Management system
- Logging and Bookkeeping
- Accounting
- Data Management: LCG File Catalog, Storage Element (DPM, dCache), GFAL, FTS, AMGA, Hydra
- Information Systems: BDII

User Interface

- **Contains clients:**
  - Job management
  - Data management
  - Access to Information System
  - Authentication
- **Installation in user space (tarball) or rpm based**

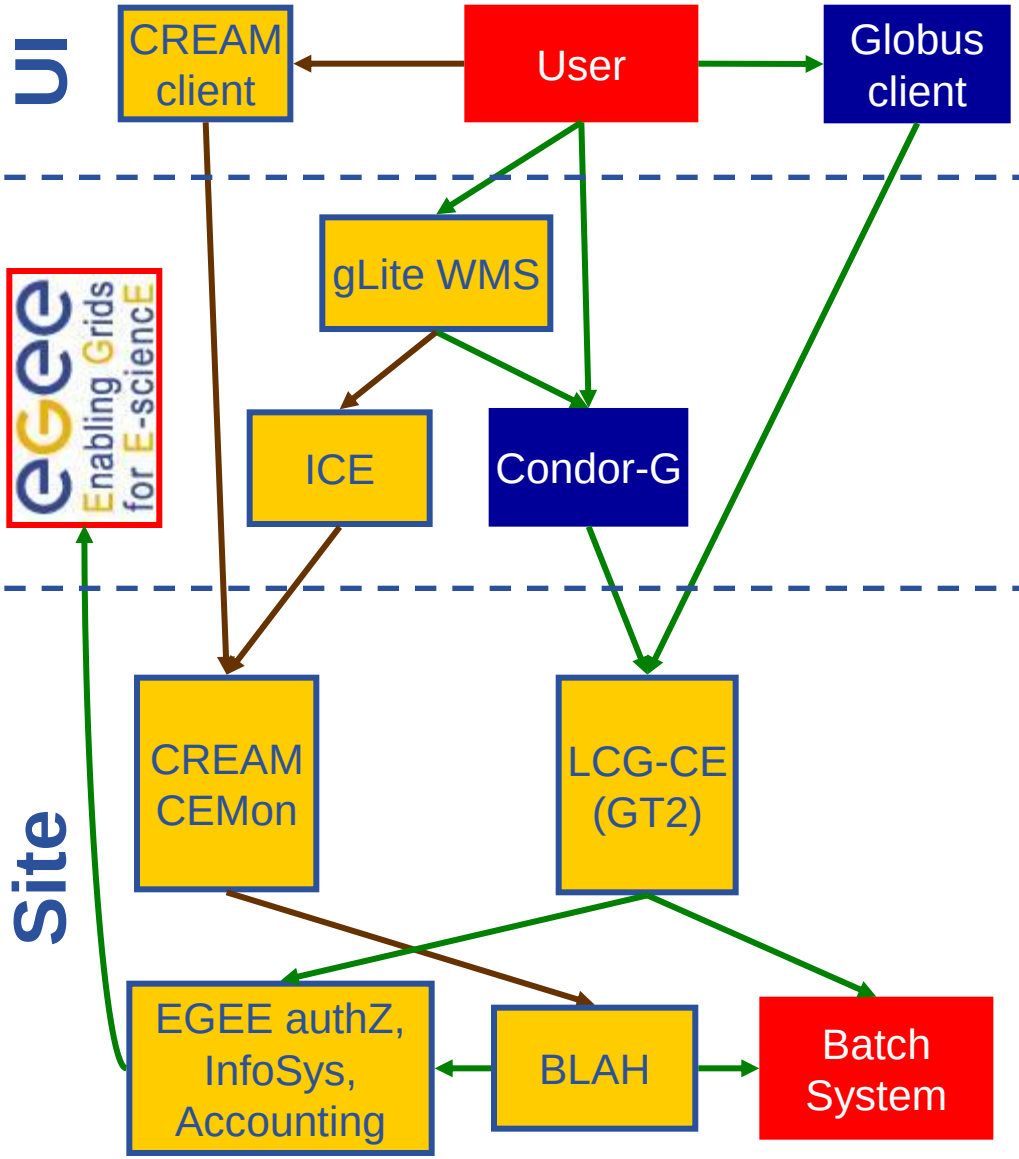
Worker Node

- That's where the jobs are being run
- Contains clients
  - Data management
- Has mechanism to install/manage VO specific software
- Currently available on SL4 i386 and x86_64
- Installs as tarball or rpm based

Resource Access in EGEE

- **LCG-CE (GT2 GRAM)**
  - Not ported to GT4. To be dismissed
- **CREAM *(WS-I)***
  - Prototype. OGF-BES (see demo at SC'06)
- **Possible developments:**
  - GT4 → BLAH submissions?

Choose your preferred path to the Batch System!



gLite component | non-gLite component | User / Resource

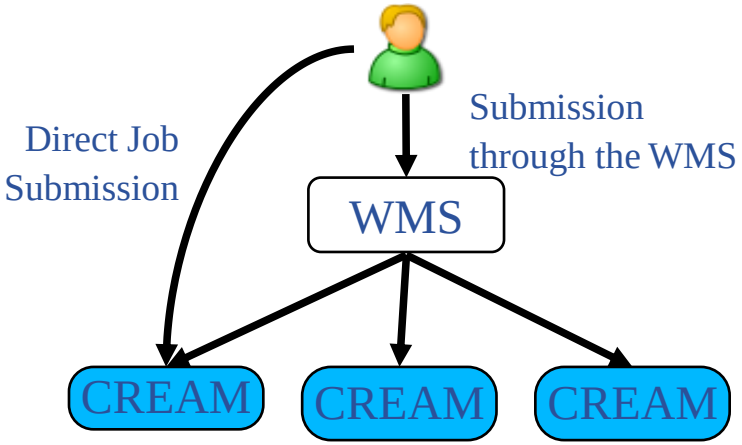
In production

In pre production

Possible development

CREAM usage scenario

- **CREAM service: Computing Resource Execution And Management service**
- **CREAM can be used:**
  - through the gLite WMS
  - by a generic client willing to interact directly with the CE
    - We provide and maintain an “official” CREAM CLI
      - *Very similar to the WMS CLI*
    - Users can build their own clients using a Web Service framework



CREAM functionality

- **Proxy delegation**
  - To delegate a proxy, which can be used by the job to do operations requiring security support (e.g. GridFTP file transfers)
  - Possibility to automatically delegate a proxy for each job submission
  - Possibility to delegate a proxy, and then using it for multiple job submissions
  - For submissions done via WMS, the proxy is delegated only when needed (i.e. only if the “same” proxy has not been delegated yet)
- **Job cancellation**
  - To cancel previously submitted jobs
- **Job list**
  - To get the identifiers of all your jobs submitted on a specific CREAM CE
- **Proxy renewal**
  - To renew proxies for previously submitted jobs
  - For jobs submitted to CREAM via the WMS, proxy renewal is done automatically, if it has been enabled
- **Job suspension and job resume**
- **Job purge**

CREAM functionality

- **Disable/enable new job submissions**
  - Can be used only by CREAM CE administrators
  - Useful for example for a scheduled shutdown of the CREAM CE

    > glite-ce-disable-submission grid005.pd.infn.it:8443

    > glite-ce-job-submit -a -r grid005.pd.infn.it:8443/cream-lsf-grid02 test.jdl

    MethodName=[jobRegister] ErrorCode=[0] Description=[The CREAM2 service cannot accept jobs anymore] Timestamp=[Tue 22 Jan 2008 16:28:47]

    > glite-ce-enable-submission grid005.pd.infn.it:8443
  - When submissions are disabled the other commands are still allowed
  - Submissions can be automatically disabled also when a certain condition (on the number of pending and/or idle and/or running jobs) specified in the CREAM conf file is met
    - E.g. a site administrator can decide to stop accepting new jobs when the site is already managing x jobs
- **Check if submissions are enabled**

  > glite-ce-allowed-submission grid005.pd.infn.it:8443

  Job Submission to this CREAM CE is disabled

CREAM job wrapper

- **The job submitted to the underlying batch system (via BLAH) is actually a job wrapper, very similar to the one considered in the submission to LCG-CEs**
  - Besides running the user job (the one specified as *Executable* in the JDL), it is responsible for transferring the sandboxes, for logging to LB, etc.
  - It also notifies CREAM about some job status changes
    - Running, Really-Running, Done
- **In the submission to the LCG CE the job wrapper is created on the WMS**
  - By the JobAdapter (Helper of WM)
- **In the CREAM CE the job wrapper is instead created on the CE by CREAM**

Interaction with LRMS

- **The interaction with the underlying local resource management system (LRMS) is fully managed by BLAH**
  - Implemented and maintained by INFN Milano group
- **BLAH used to submit, cancel, etc. jobs on the batch system**
- **BLAH also used, via the BLParser, to notify CREAM about job status changes**
  - Actually CREAM knows about (some) job status changes also from the job wrapper running on the Worker Node
- **Two BLParser implementation models:**
  - Old one: works parsing the batch system log files
  - New one: works referring to the batch system status/history commands
    - New model done also to facilitate the porting to new batch systems

Compliance to standards

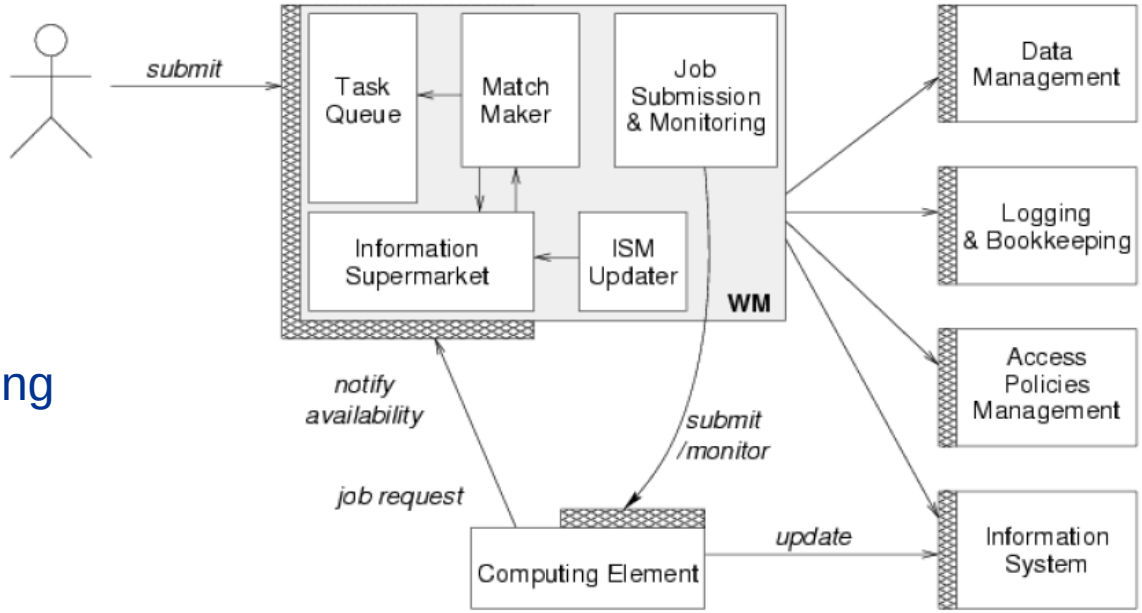
- **Besides the legacy interface, CREAM exposes also a BES-compliant interface**
- **BES (Basic Execution Service): recent OGF specification for a standard interface for Grid execution services**
  - Aim: favor interoperability between different Grids
- **BES defines basic operations for job submission and management**
  - BES itself does not mandate any specific security implementation
    - E.g. proxy delegation is not part of the BES specification
- **JSDL (Job Submission Description Language) used in BES to describe computational jobs**

Services and Clients

- Authentication and authorization: Myproxy, Voms
- Resource access: Computing Element, Worker Node, User Interface
- **Workload Management system**
- Logging and Bookkeeping
- Accounting
- Data Management: LCG File Catalog, Storage Element (DPM, dCache), GFAL, FTS, AMGA, Hydra
- Information Systems: BDII

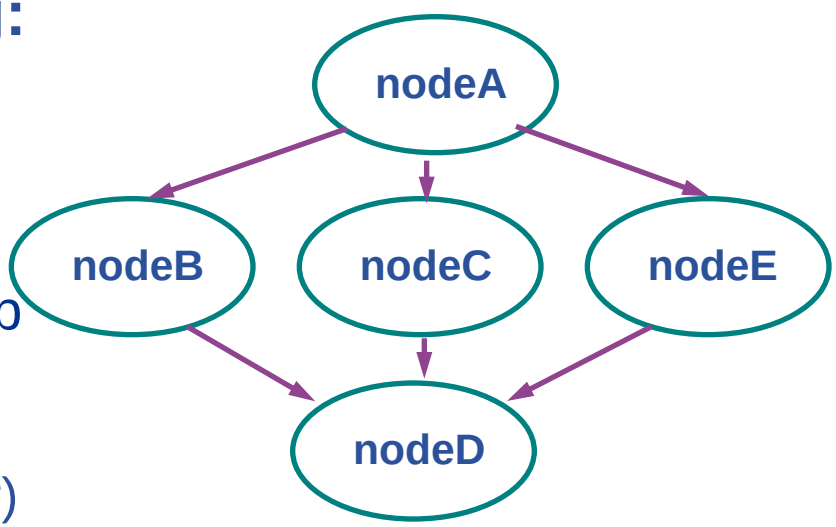
Workload Management System

- **WMS: Resource brokering, workflow management, I/O data management**
  - ➔ Web Service interface: WMProxy
  - – Task Queue: keep non matched jobs
  - – Information SuperMarket: optimized cache of information system
  - – Match Maker: assigns jobs to resources according to user requirements (possibly including data location)
  - – Job submission & monitoring
    - ➔ Condor-G
    - ➔ *ICE (to CREAM)*
  - – External interactions:
    - Information System
    - Data Catalogs
    - Logging&Bookkeeping
    - *Policy Management systems*



Workload Management System

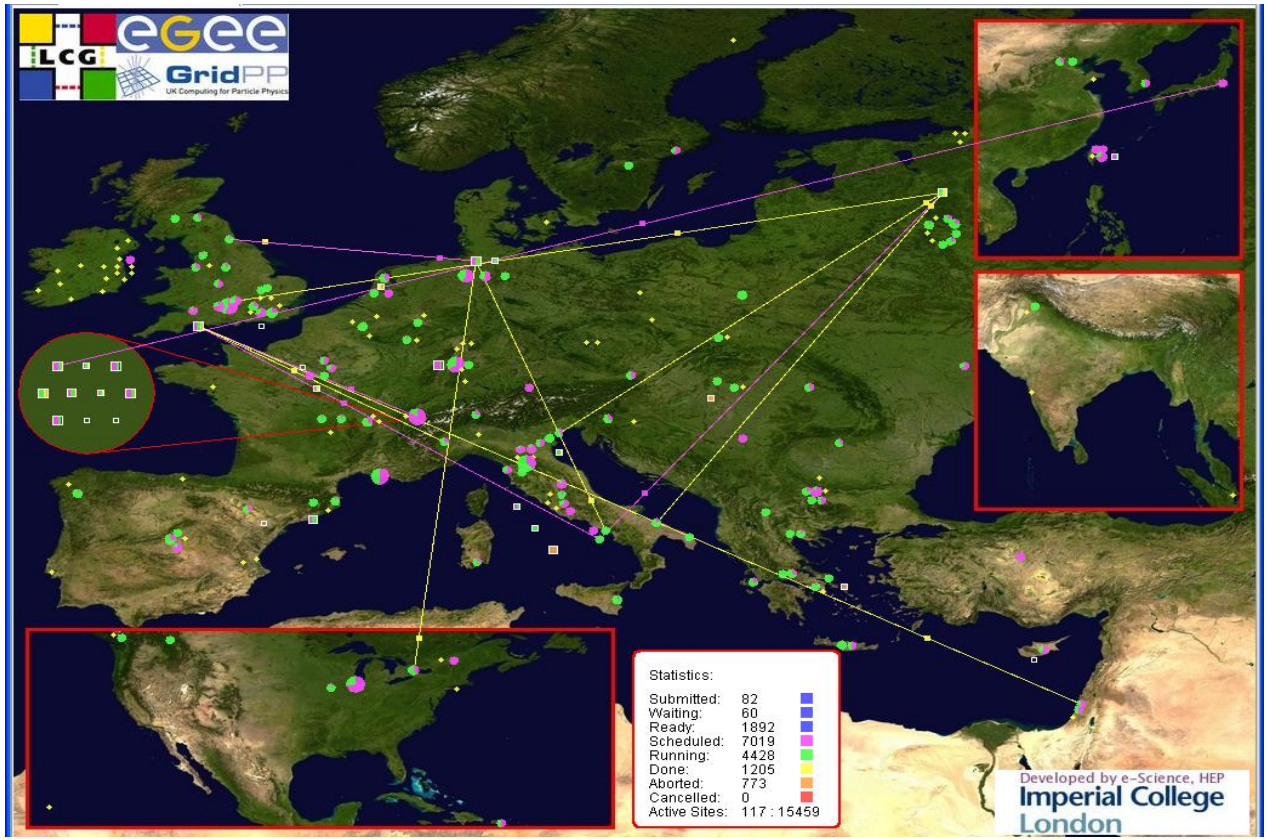
- **Not only resource brokering:**
- **Support for compound jobs**
  - Compound, Parametric, DAGs (Direct Acyclic Graphs)
  - One shot submission of a group of jobs (Bulk Submission)
    - Submission time reduction (single call to WMProxy server)



  - Shared input sandboxes
  - Single Job Id to manage the group (single job ID still available)
- **Support for 'scattered' input/output sandboxes**
- **Support for *deep* and *shallow resubmission***
  - Automatic resubmission in case of failure of the infrastructure
- **Automatic proxy renewal (including VOMS attributes)**

Logging & Bookkeeping

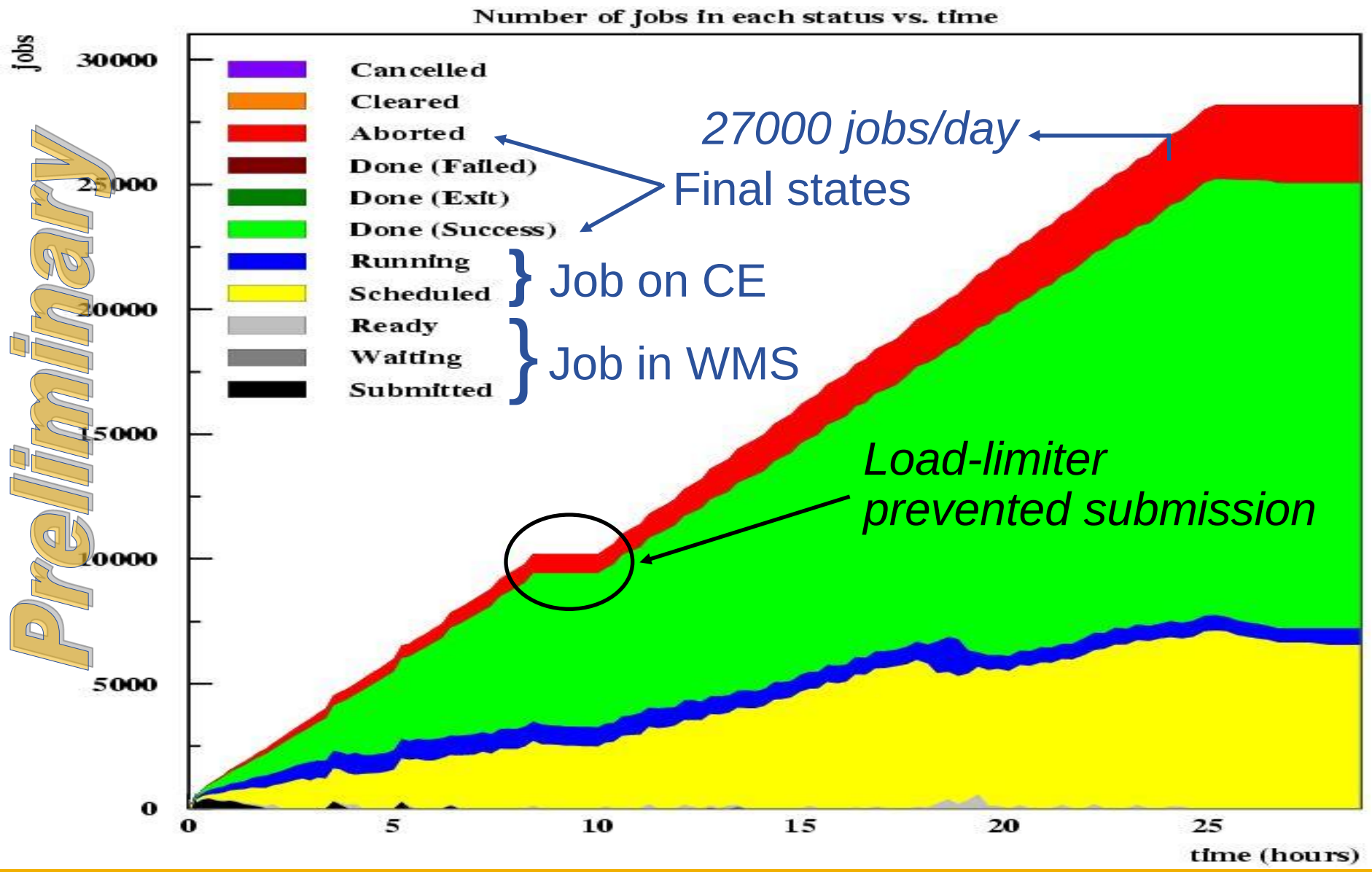
- **LB: Tracks jobs during their lifetime (in terms of events)**
  - Functional to WMS operations
    - but works also for jobs not submitted through the WMS
  - Web service Interface for querying



Job Provenance

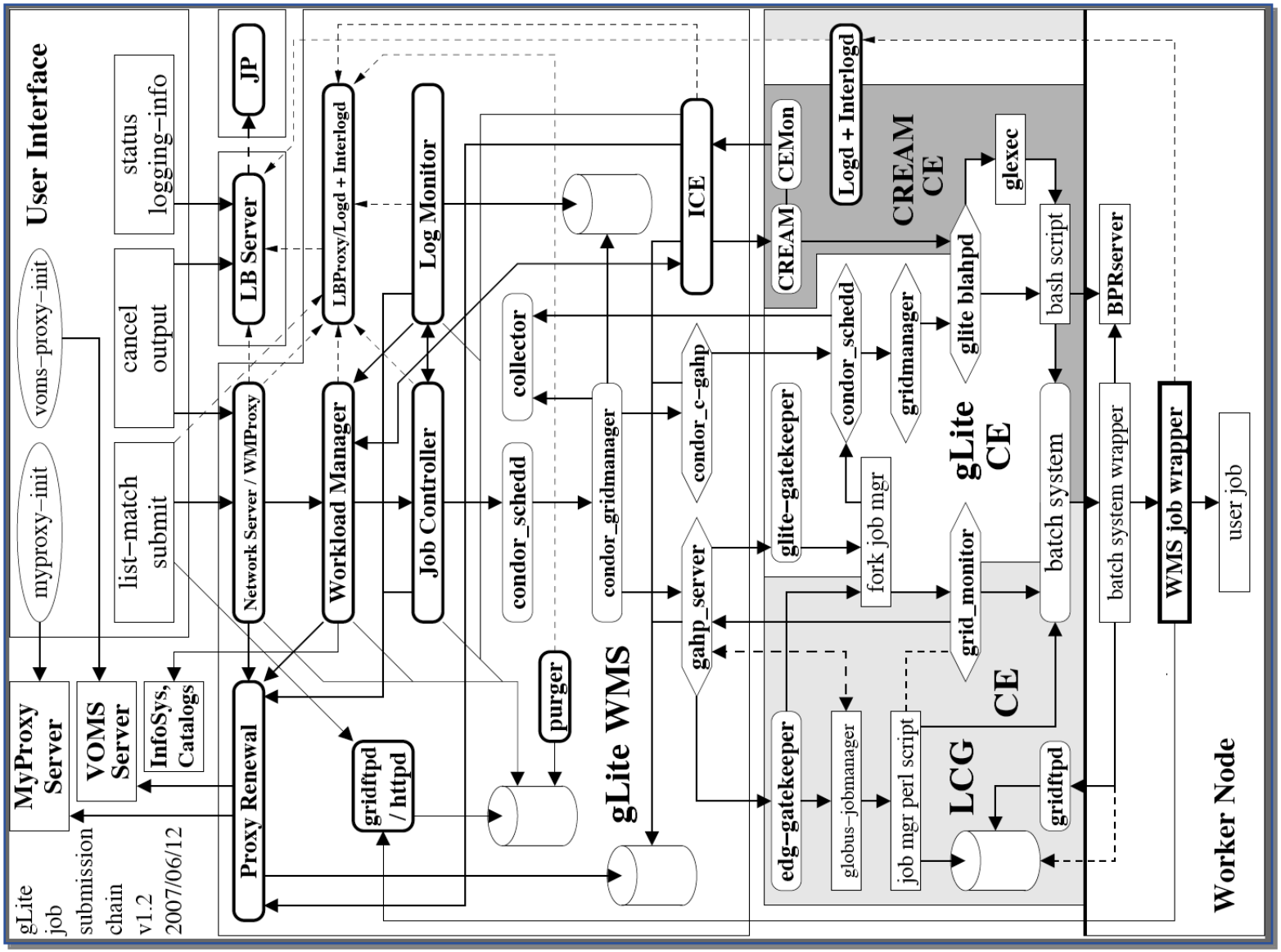
- **LB data is only stored for a limited amount of time**
- **Job Provenance allows to store LB data for a long time**
  - Can be enabled on a VO bases
  - Not yet in the release

WMS and LB performances



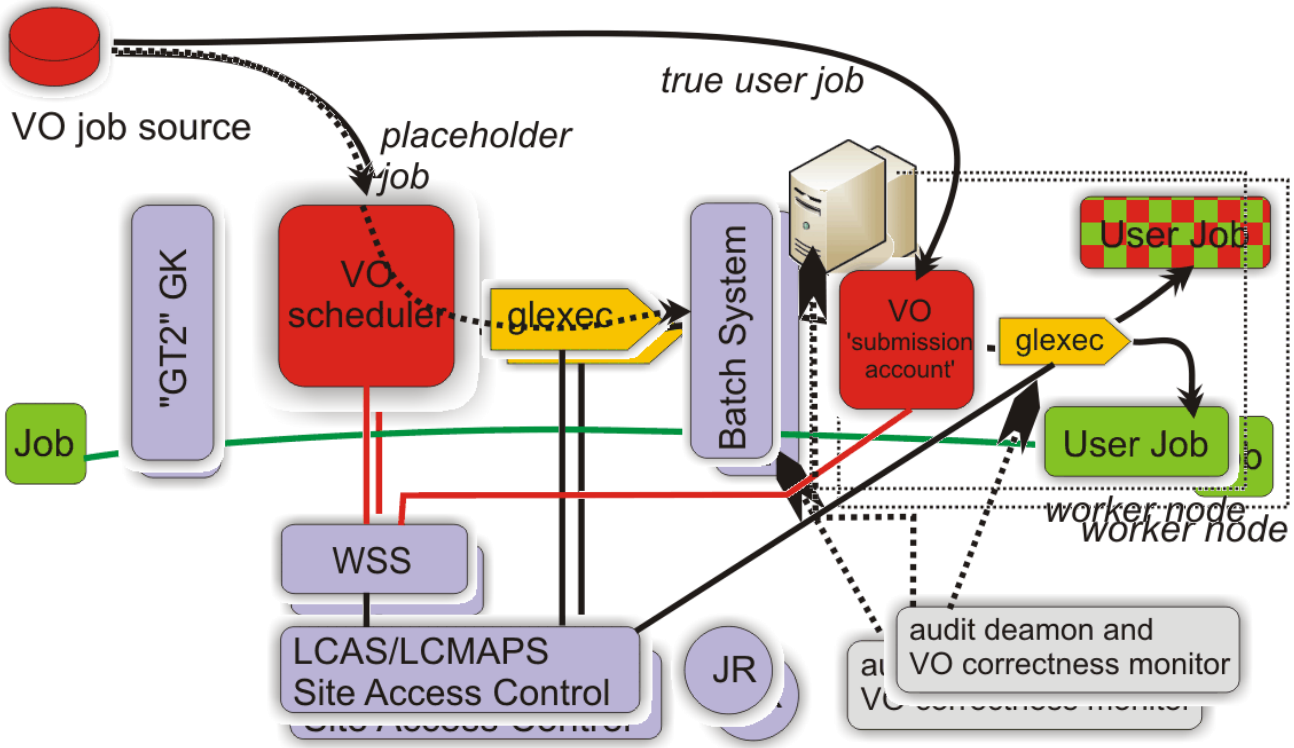
What it really does

- Thanks to Maarten we know:
  - Simplified view



Coming: support for pilot jobs

- **Several VOs submit *pilot jobs* with a single identity for all of the VO**
  - The pilot job gets the user job when it arrives on the WN and executes it
    - Just-in-time scheduling. VO policies implemented at the central queue



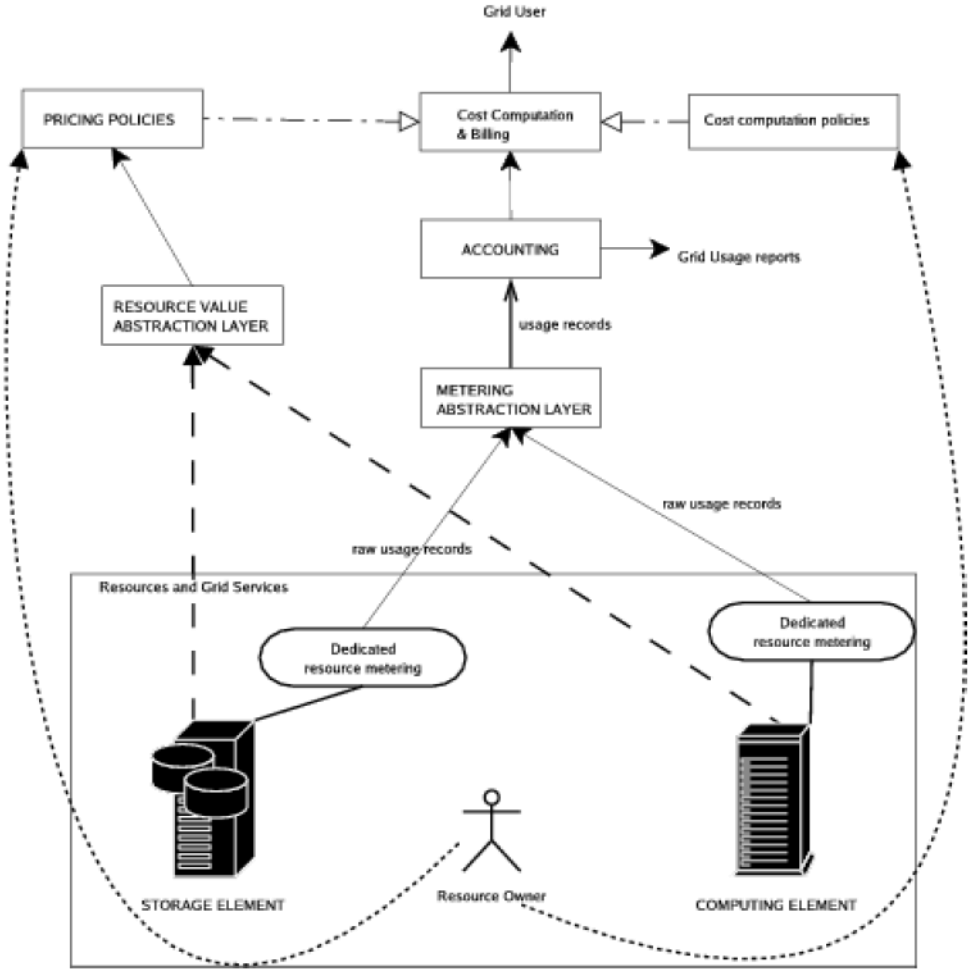
- **Use the same mechanism for changing the identity on the Computing Element also on the Worker Nodes (glexec)**
  - The site may know the identity of the real user

Services and Clients

- Authentication and authorization: Myproxy, Voms
- Resource access: Computing Element, Worker Node, User Interface
- Workload Management system
- Logging and Bookkeeping
- **Accounting**
- Data Management: LCG File Catalog, Storage Element (DPM, dCache), GFAL, FTS, AMGA, Hydra
- Information Systems: BDII

Job accounting

- **Resource usage by VO, group or single user**
  - *Resource metering*: sensors running on resources to determine usage
  - *Pricing policies*: associate a cost to resource usage
    - if enabled allowed market-based resource brokering
  - privacy: access to accounting data granted only to authorized people (user, provider, VO manager)



- **Information collected at the Grid Operations Centre (GOC)**
- **Basic functionality in APEL, full functionality in DGAS**

Accounting

GOC - VO Manager View

- Table shows CPU, WCT and Jobs of the Top 10 Anonymised Users
- This example shows that the largest WCT User has a job efficiency o 10%É clearly the VO Manager may wish to contact this person

EGEE View | VO MANAGER View | VO MEMBER View | SITE ADMIN View | USER View

January 2006 - December 2006

The following table shows the Usage of the Top 10 Users ov time and the Total Usage of the Other Users. A detailed view selecting an individual user.

Top 10 Us

#	User ID	Jobs #	Jobs %	CPU time Hrs	CPU time %	Wor Hrs
1	2daf992f6a1f68d0	16,629	26.2%	120	24.3%	8
2	671e5b277fa6d1ac	20	0.0%	98	19.8%	8
3	4cd78d6b029f7050	7,773	12.5%	80	16.2%	5
4	5d61bd2201bec5f9	7,816	12.6%	52	10.5%	4
5	2a153e141e98f06a	1,950	3.1%	36	7.3%	3
6	62d3866c3c8260d6	39	0.1%	27	5.5%	1
7	2756dfcb65975a47	95	0.2%	19	3.8%	1
8	748206ea352cce31	467	0.8%	12	2.4%	1
9	2d04c1be5d64c1b8	3	0.0%	11	2.2%	1
10	2d72edb286203a97	83	0.1%	9	1.8%	1
	Others (DN known)	25,073	40.5%	18	3.6%	1
	Others (DN unknown)	3,021	4.9%	12	2.4%	1
	Total	61,969		494		3

Click

Key: 0% <= eff < 75%, 75% <

EGEE-II INFSO-RI-031688

GOC - Site Admin View

- The Site Administrator can view usage of anonymous grid users who executed jobs at th

EGEE View | VO MANAGER View | VO

CESGA

Janua

The following table shows the Usage of the Top 10 Users orde time and the Total Usage of the Other Users. A detailed view c selecting an individual user.

Top 10 User

#	User ID	Jobs #	Jobs %	CPU time Hrs	CPU time %	Norm Hrs
1	007c482b7a509753	335	0.5%	6,069	30.9%	2,3
2	006f8bf719df068f	49,214	68.6%	2,769	14.1%	1,0
3	57684c0c3d621a53	1,598	2.2%	1,717	8.7%	6
4	43289fd45f650e5e	101	0.1%	1,616	8.2%	6
5	11e9316e4987c00c	541	0.8%	1,295	6.6%	4
6	6b641d611ad9af1e	101	0.1%	1,148	5.8%	4
7	2b87e56d01c14620	186	0.3%	787	4.0%	3
8	398ccd8b5d6c6e30	51	0.1%	678	3.5%	2
9	0e2564274a92c273	1,098	1.5%	608	3.1%	2
10	77f727593fa532af	625	0.9%	401	2.0%	1
	Others (DN known)	17,930	25.0%	2,553	13.0%	9
	Others (DN unknown)	0	0.0%	0	0.0%	
	Total	71,780		19,641		7,4

Click he

Key: 0% <= eff < 75%, 75% <=

EGEE-II INFSO-RI-031688

GOC - User View

- **Each Grid User can interrogate their own accounting data**
  - Tables showing what they did and when
  - Number of Jobs, CPU and WCT per Month (per VO)
  - Average Job Efficiency per VO
  - AccumulativeNjobs, CPU and WCT per VO
  - The sites which executed the jobs, and when they were done

EGEE View | VO MANAGER View | VO MEMBER View | SITE ADMIN View | USER View

dteam - VO | Exclude dteam jobs information

Refresh

USER Total number of jobs by VO and DATE.

January 2006 - December 2006.

The following table shows the distribution of the Total number of Your Jobs grouped by VO a DATE

Total number of jobs run by VO and DATE

VO	Jan 06	Feb 06	Mar 06	Apr 06	May 06	Jun 06	Jul 06	Aug 06	Sep 06	Oct 06	Nov 06	Dec 06	Total
dteam	1,447	1,554	972	1,450	1,535	1,642	1,392	1,280	926	0	0	0	12,198
otal	1,447	1,554	972	1,450	1,535	1,642	1,392	1,280	926	0	0	0	12,198
Percentage	11.86%	12.74%	7.97%	11.89%	12.58%	13.46%	11.41%	10.49%	7.59%	0.00%	0.00%	0.00%	

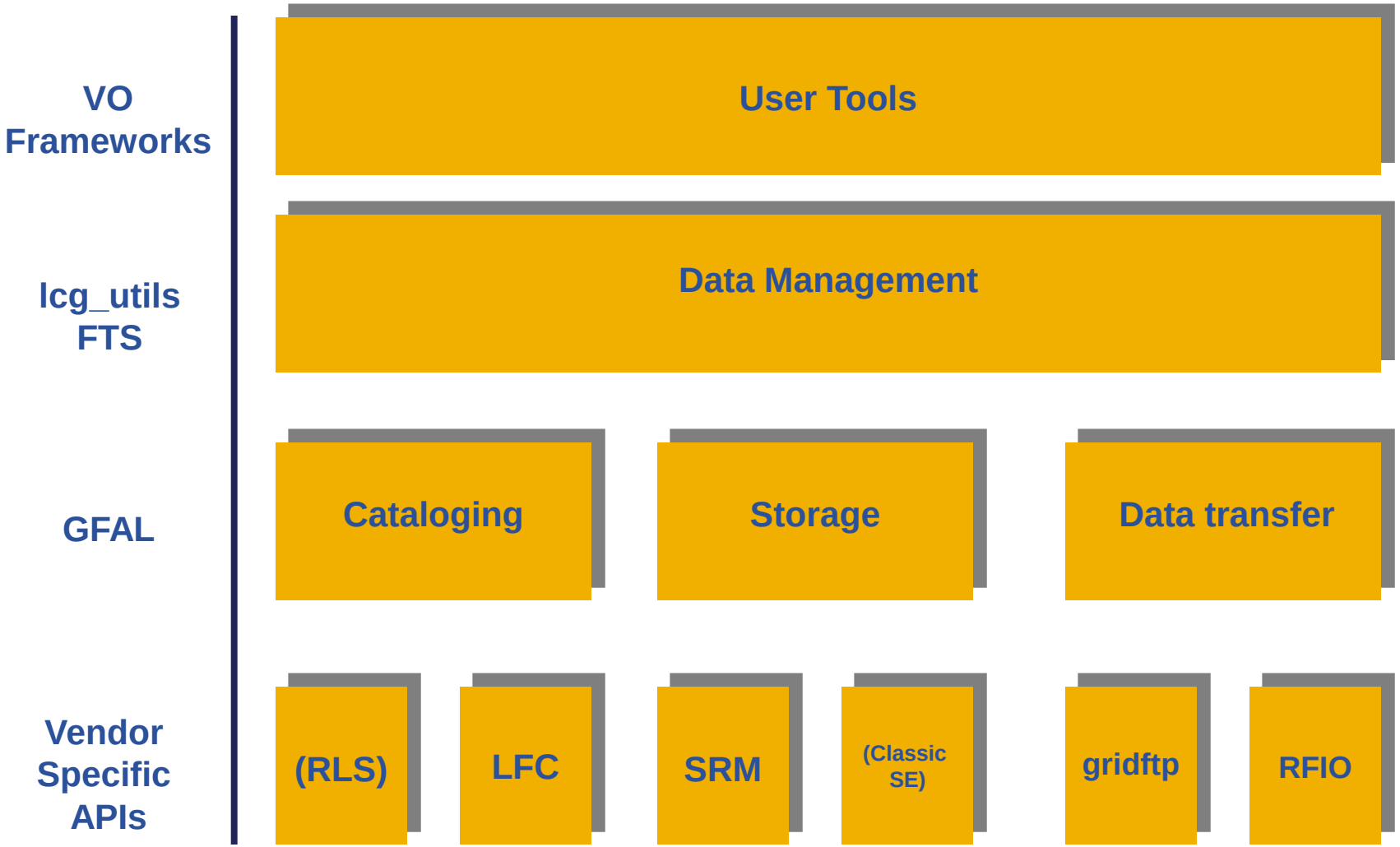
Click here for a csv dump of this table

EGEE-II INFSO-RI-031688 gLite - The EGEE Middleware Distribution 25

Services and Clients

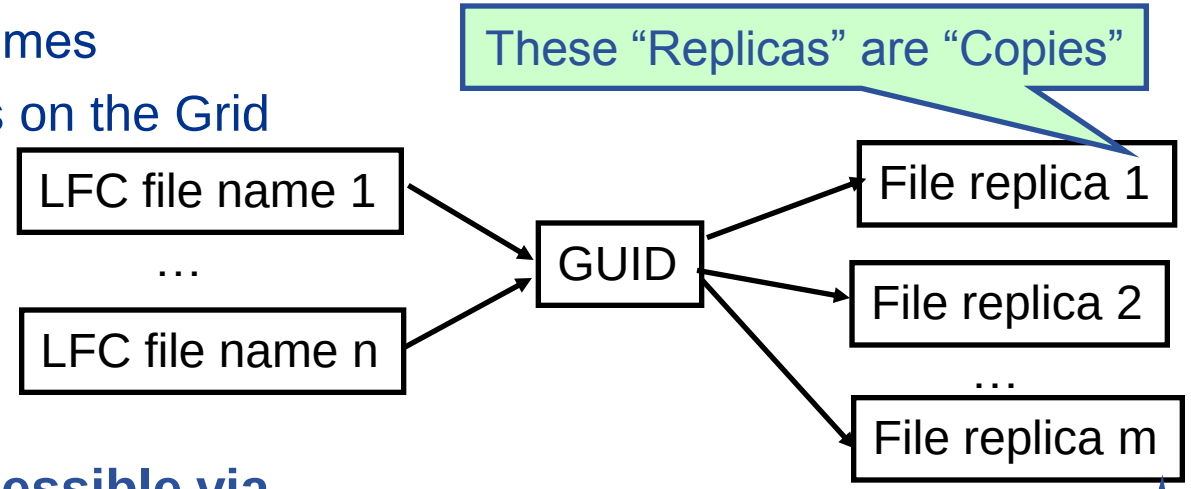
- Authentication and authorization: Myproxy, Voms
- Resource access: Computing Element, Worker Node, User Interface
- Workload Management system
- Logging and Bookkeeping
- Accounting
- **Data Management: LCG File Catalog, Storage Element (DPM, dCache), GFAL, FTS, AMGA, Hydra**
- Information Systems: BDII

EGEE Data Management



LCG "File" Catalog

- **The LFC stores mappings between**
  - Users' file names
  - File locations on the Grid



- **The LFC is accessible via**
  - CLI, C API, Python interface, Perl interface
    - Supports sessions and bulk operations
  - Data Location Interface (DLI)
    - Web Service used for match making:
      - *given a GUID, returns physical file location*
- **ORACLE backend for high performance applications**
  - Read-only replication support

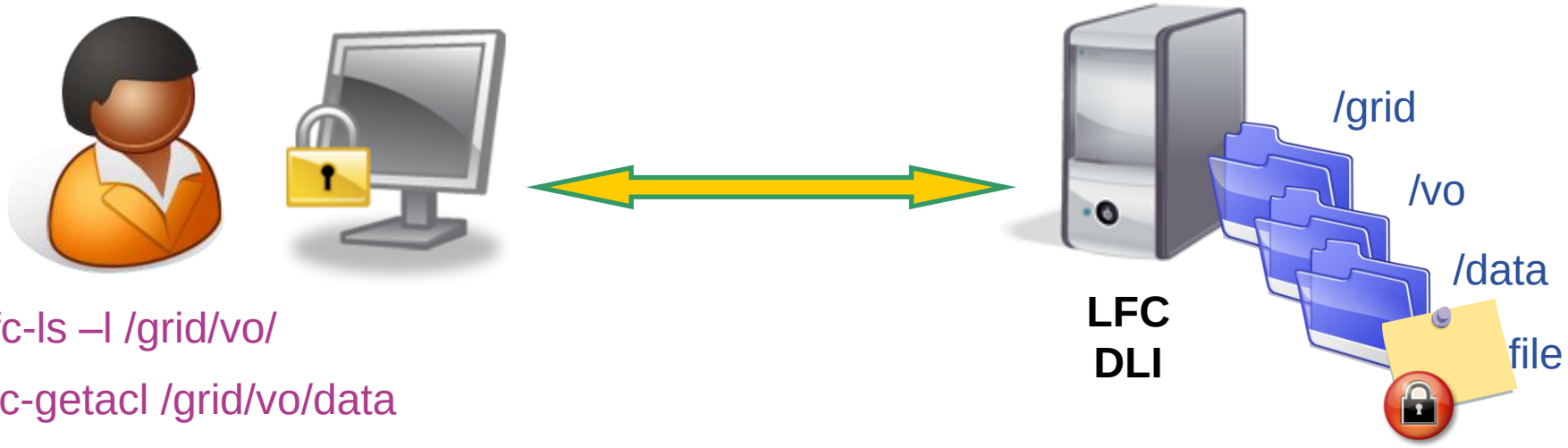
All files are "Write Once Read Many"

LFC features

Hierarchical Namespace
GSI security
Permissions and ownership
ACLs (based on VOMS)
Virtual ids
- Each user is mapped to (uid, gid)

VOMS support
- To each VOMS group/role corresponds a virtual gid

Bulk operations



lfc-ls –l /grid/vo/

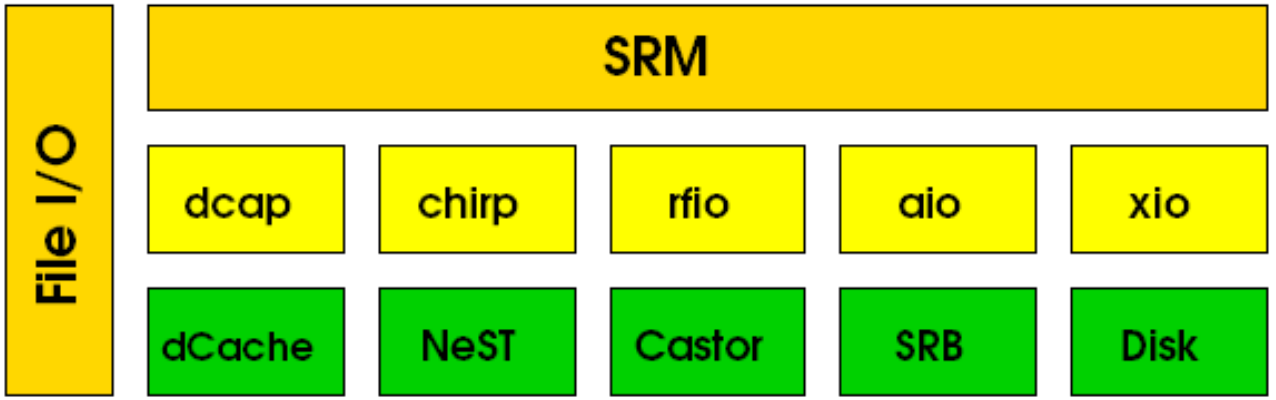
lfc-getacl /grid/vo/data

Services and Clients

- Authentication and authorization: Myproxy, Voms
- Resource access: Computing Element, Worker Node, User Interface
- Workload Management system
- Logging and Bookkeeping
- Accounting
- **Data Management: LCG File Catalog, Storage Element (DPM, dCache), GFAL, FTS, AMGA, Hydra**
- Information Systems: BDII

Storage Element

- **Storage Resource Manager (SRM)**
  - hides the storage system implementation (disk or active tape)
  - handles authorization
  - translates SURLs (Storage URL) to TURLs (Transfer URLs)
  - disk-based: DPM, dCache,+; tape-based: Castor, dCache
- ***File I/O*: *posix-like* access from local nodes or the grid**
  - ➔ GFAL (Grid File Access Layer)



What is a DPM ?

- **Disk Pool Manager**
  - Manages storage on disk servers
  - SRM support
    - 1.1
    - 2.1 (for backward compatibility)
    - 2.2 (released in DPM version 1.6.3)
  - GSI security
  - ACLs
  - VOMS support
  - Secondary groups support (see LFC)

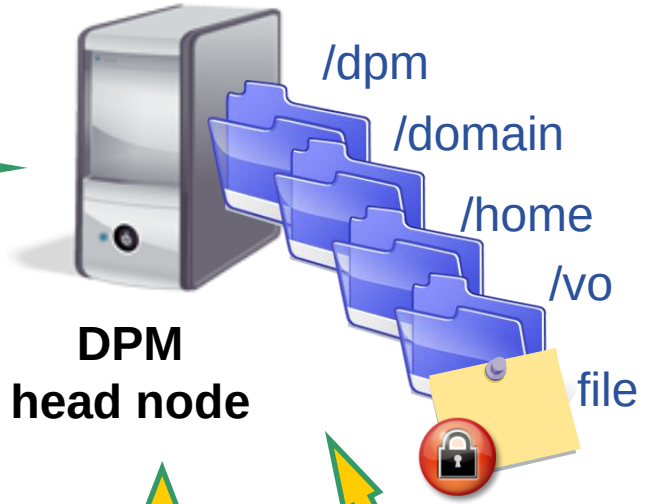
DPM strengths

- **Easy to use**
  - Hierarchical namespace
    - $ dpns-ls /dpm/cern.ch/home/vo/data
- **Easy to administrate**
  - Easy to install and configure
  - Low maintenance effort
  - Easy to add/drain/remove disk servers
- **Target: small to medium sites**
  - Single disks --> several disk servers

DPM: user's point of view

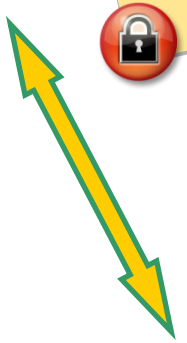
(uid, gid1, …)

CLI, C API, SRM-enabled client, etc.



DPM head node

data transfer



DPM disk servers

DPM Name Server
- Namespace
- Authorization
- Physical files location

Disk Servers
- Physical files

Direct data transfer from/to disk server (no bottleneck)

External transfers via gridFTP

GFAL & lcg_util

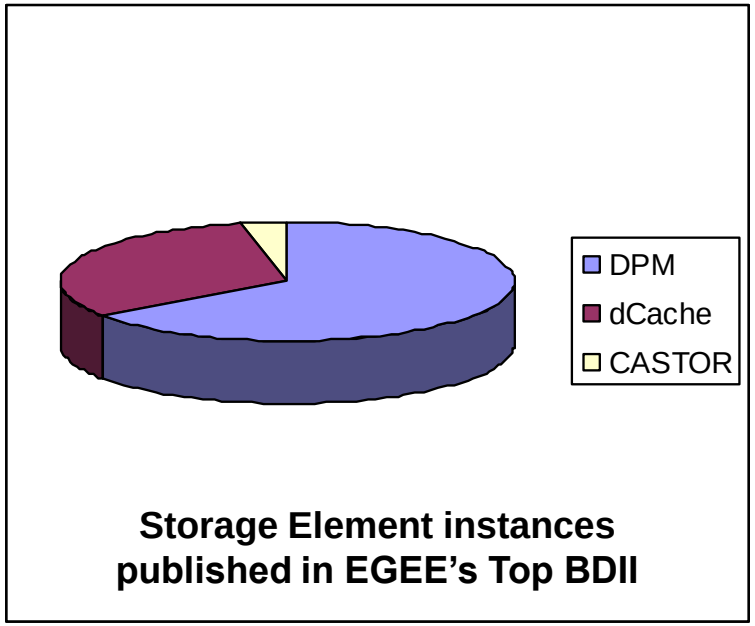
- **Data management access libs.**
  - Shield users from complexity
  - Interacts with information system, catalogue and SRM-SEs

- **GFAL**
  - Posix like C API for file access
  - SRMv2.2 support
  - User space tokens correspond to
    - A certain retention policy (custodial/replica)
    - A certain access latency (online/nearline)

- **lcg_util (command line + C API)**
  - Replication, catalogue interaction etc.

LFC & DPM deployment status

- **EGEE Catalog**
  - 110 LFCs in production
    - 37 central LFCs
    - 73 local LFCs
- **EGEE SRM Storage Elements**
  - CASTOR
  - dCache
  - DPM
    - 96 DPMs in production
    - Supporting 135 VOs



Storage Element instances published in EGEE's Top BDII

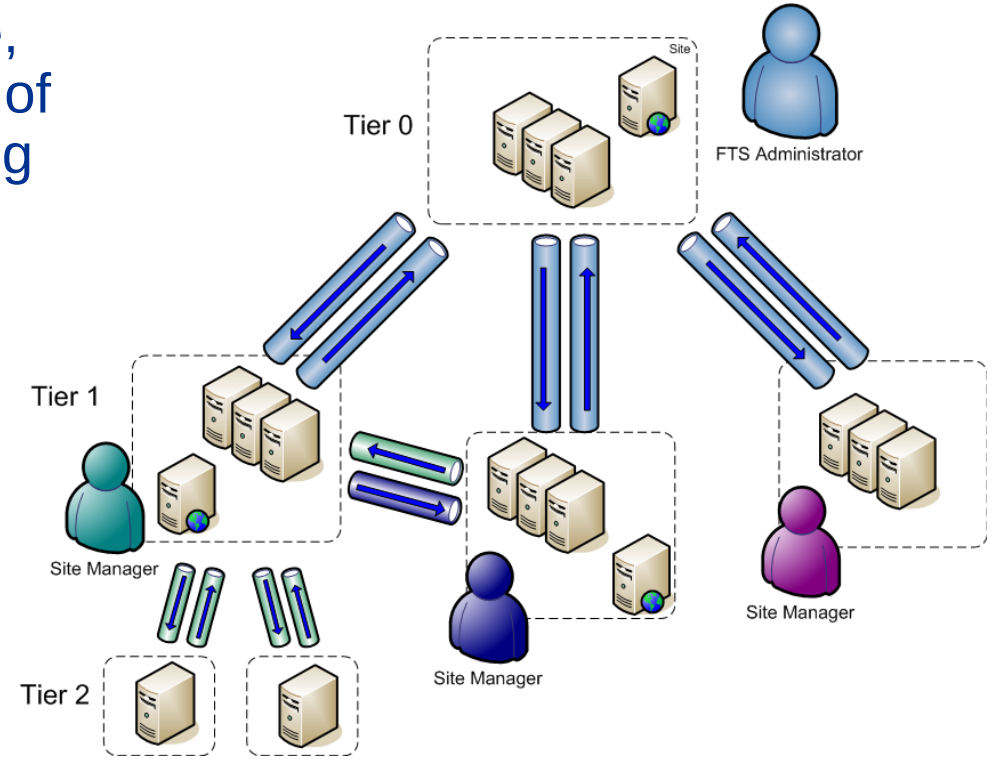
- **LFC and DPM**
  - Stable and reliable production quality services
  - Well established services
  - Require low support effort from administrators and developers

dCache

- Storage Management system developed at DESY and Fermilab
- Supports disk and tape
- Distributed within gLite

FTS overview

- **gLite File Transfer Service is a reliable data movement fabric service (batch for file transfers)**
  - FTS performs bulk file transfers between multiple sites
  - Transfers are made between any SRM-compliant storage elements (both SRM 1.1 and 2.2 supported)
  - It is a **multi-VO** service, used to balance usage of site resources according to the SLAs agreed between a site and the VOs it supports
  - VOMS aware



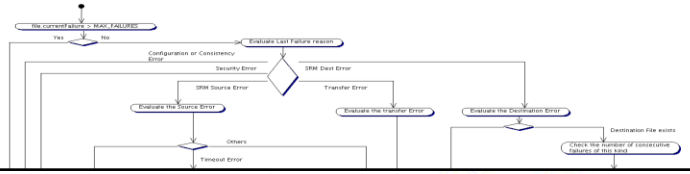
Grid File Transfer Service

FTS

- **Why is it needed ?**
  - For the user, the service it provides is the reliable point to point movement of Storage URLs (SURLs) and ensures you get your share of the sites' resources
  - For the site manager, it provides a reliable and manageable way of serving file movement requests from their VOs and an easy way to discover problems with the overall service delivered to the users
  - For the VO production manager, it provides ability to control requests coming from his users
    - Re-ordering, prioritization,…
  - The focus is on the "**service**" delivered to the user
    - It makes it easy to do these things well with minimal manpower

FTS: key points

- **Reliability**
  - It handles the retries in case of storage / network failure
    - VO customizable retry
  - Service designed for high-availability deploym
- **Security**
  - All data is transferred se SRM / gridFTP
  - Service audits all user /
- **Service and performan**
  - Service stability: it is des storage and network res
  - Service recovery: integra degradation



FTS Report

Disclaimer
This page contains a report generated from information stored in the FTS Database and is intended for reporting purposes only. Since the format will probably change in the future, it's therefore recommended not to use parsing robots on it.

Statistics concerning all the transfers performed yesterday
Between 2006-10-12 08:00:00 +02:00 and 2006-10-13 08:00:00 +02:00

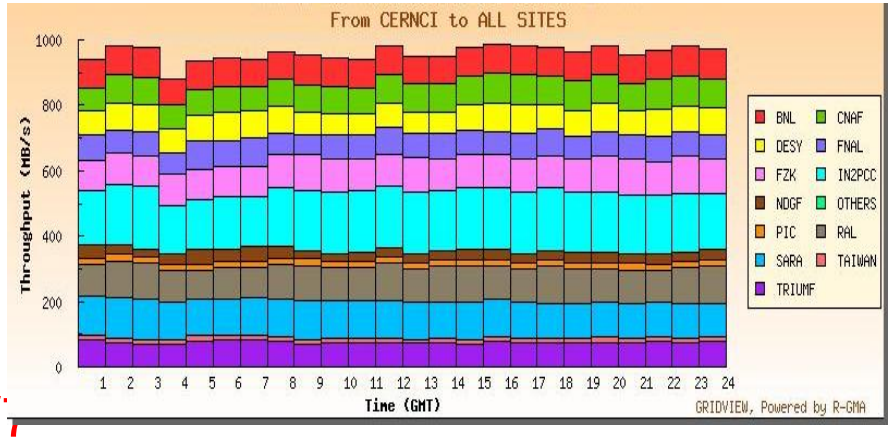
CERN-.* Filter Show VO details

Channel Name	VO Name	Total	% Failures	# Succ.	# Fail.	1st Failure Reason	% 1st Failure Reason	2nd Failure Reason	% 2nd Failure Reason	Avg. Size (GB)	Avg. Duration (sec)	Avg. Tx Rate (MB/sec)	Eff. Tx Bytes (GB)	Tx Bytes (GB)
CERN-PIC	[All]	12262	73.97	3192	9070	Dest SRM	56.22	Other	37.53	0.53	263.03	1.62	1700.41	1700.41
	atlas	8932	99.92	7	8925	Dest SRM	57.13	Other	38.08	0	220	0	0	0
	cms	208	0	208	0					2.7	767.55	3.64	561.26	561.26
	dteam	974	0.51	969	5	Other	80	Source SRM	20	0.95	356.31	2.88	923.83	923.83
	lhcb	2145	6.53	2005	140	Source SRM	99.29	Other	0.71	0.11	165.85	0.81	215.32	215.32
	ops	3	0	3	0					0	202.67	0	0	0
CERN-RAL	[All]	8699	59.26	3544	5155	Other	84.91	Source SRM	14.88	0.85	478.22	2.59	3026.81	3027.57
	alice	1155	82.6	201	954	Other	99.58	Dest SRM	0.31	1.86	1805.05	1.11	372.95	372.95
	atlas	4512	88.52	518	3994	Other	84.85	Source SRM	15.15	1.79	1428.94	1.49	926.26	926.57
	cms	227	3.08	220	7	Dest SRM	85.71	Source SRM	14.29	2.53	348.65	10.08	555.61	555.61
	dteam	1077	3.99	1034	43	Other	86.05	Source SRM	9.3	0.95	276.64	4.01	980.47	980.91
	lhcb	1725	9.1	1568	157	Source SRM	99.36	Other	0.64	0.12	146.03	1.16	191.52	191.52
	ops	3	0	3	0					0	27	0.01	0	0
CERN-SARA	[All]	8792	42.55	5051	3741	Dest SRM	83.77	Source SRM	12.22	1.34	108.02	15.4	6777.95	6784.92
	alice	3134	15.12	2660	474	Source SRM	57.17	Dest SRM	41.14	1.66	109.53	18.43	4426.44	4430.29
	atlas	2018	53.32	942	1076	Dest SRM	72.4	Source SRM	16.54	1.15	144.44	9.42	1085.07	1087.6
	dteam	3488	61.32	1349	2139	Dest SRM	98.74	Other	0.98	0.93	81.91	14.66	1260.74	1261.32
	lhcb	148	35.14	96	52	Dest SRM	92.31	Other	3.85	0.06	76.1	0.93	5.7	5.7
	ops	4	0	4	0					0	97.25	0.02	0	0
CERN-INFN	[All]	11492	42.31	6630	4862	Dest SRM	43.85	Other	37.7	1.13	395.77	3.21	7514.29	7614.84
CERN-CERN	[All]	1536	39.71	926	610	Source SRM	58.36	Dest SRM	15.9	0.07	287.71	0.38	67.89	69.08
CERN-ASCC	[All]	6851	23.54	5238	1613	Source SRM	50.84	Other	28.89	1.14	1098.6	1.08	5955.81	6080.58
CERN-GRIDKA	[All]	12755	21.38	10028	2727	Source SRM	64.36	Other	32.53	0.87	371.97	3.19	8762.02	8767.53
CERN-TRIUMF	[All]	2244	20.63	1781	463	Other	61.77	Source SRM	31.1	1.04	395.15	3.63	1847.25	1917.13
CERN-BNL	[All]	13975	19.42	11261	2714	Source SRM	69.97	Other	24.24	0.44	190.38	3.41	4951.59	4960.34
CERN-IN2P3	[All]	11697	13.76	10087	1610	Source SRM	48.57	Other	47.45	1.22	296.21	5.33	12329.63	12329.63
CERN-FNAL	[All]	917	4.58	875	42	Transfer	97.62	Other	2.38	0	379.88	0	0	0

Click on the Channel Name to show the VO details

Service scale

- **Designed to scale up to the transfer needs of very data intensive applications**
- **Currently deployed in production at CERN**
  - Running the production WLCG tier-0 data export
  - Target rate is ~1 Gbyte/sec 24/7
  - Over 9 petabytes transferred in last 6 months >10 million files
- **Also deployed at ~10 tier-1 sites running a mesh of transfers across WLCG**
  - Inter-tier1 and tier-1 to tier-2 transfers
  - Each tier-1 has transferred around 0.2 – 0.5 petabytes of data



AMGA

- **AMGA is a metadata catalogue**
- **Metadata is information about data stored in files**
  - Usually lives in relational databases
  - E.g. Medical image (data) and corresponding patient (metadata)
- **Why not accessing DBs directly on the Grid? Possible but**
  - Authentication (VOMS)
  - Logging, tracing
  - Connection pooling
  - Data replication

AMGA features

- **Implementation:**
  - SOAP and Text front-ends
  - Streamed Bulk Operations ----> performance
  - Supports single calls, sessions & connections
  - SSL security with gridcerts (X509)
    - and others, passwords, Kerberos
    - Own User & Group management + VOMS
  - PostgreSQL, Oracle, MySQL, SQLitebackends
  - APIs: C/C++, Python, Java, CLI
  - SOAP interface
- **Query parser supports good fraction of SQL:**
  - Access permissions per directory/entry via ACLs
- **AMGA integrates support for replication of metadata**
  - Asynchronous replication: Ideal forWAN
- **Performance required to be comparable to direct DB access by HEP applications**

Replication and Federation

- Replication: Transfer of data/changes master to one or more slaves
- Federation: Integration of data from many masters into a whole
- Replication allows scaling the reads, high availability
- Federation allows scaling the total size of the metadata
- Federation allows scaling the number of concurrent writes
- Replication within the metadata catalogue itself allows cross vendor database replication

Scale

- **LHCb (HEP VO use case)**
  - 120 Million entries successfully tested!
  - 150GB data
  - 100 000 entries/day insert rate expected
  - 10 entries/second read-rate
  - Uses ORACLE RAC backend
    - For most demanding use cases

Encrypted Data Storage

Motivation

Medical community as the principal user

- large amount of images
- privacy concerns vs. processing needs
- ease of use (image production and application)

Strong security requirements

- anonymity (patient data is separate)
- fine grained access control (only selected individuals)
- privacy (even storage administrator cannot read)

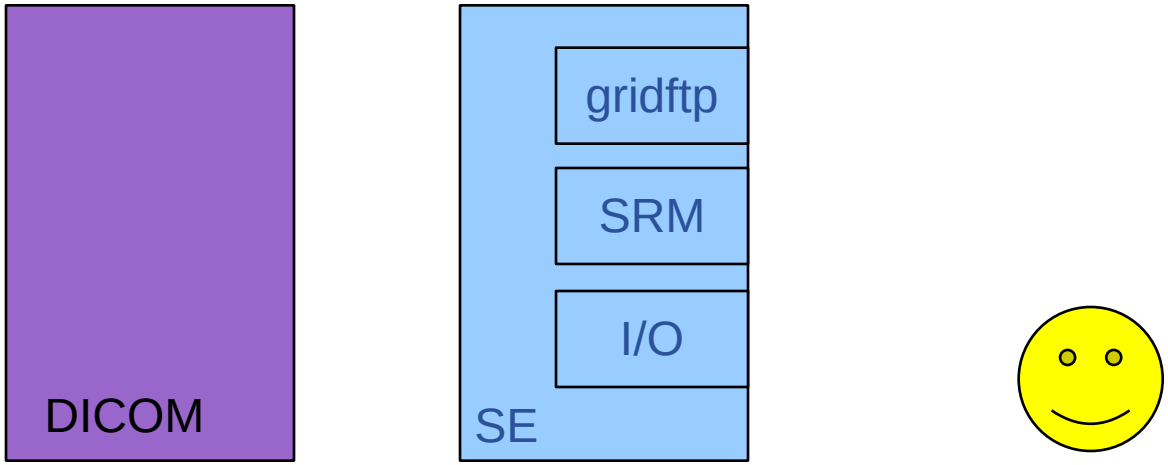
- Components partly still under development

HYDRA

- Encrypted Storage solution targeted towards Biomed/Medical Data Management.
- Encrypts files and stores them on normal Storage Elements.
- The encryption key is stored in the Hydra Keystore.
- Keys are split and distributed to at least three keystores.
- The Hydra client library makes use of Shamir's Secret Sharing Scheme to split encryption keys and store these pieces into multiple Hydra services.

Building Blocks

- **Hospitals:**
  - DICOM = Digital Image and COmmunication in Medicine
- **Grid: SE = SRM + gridftp + I/O**
  - and a client (application processing an image)



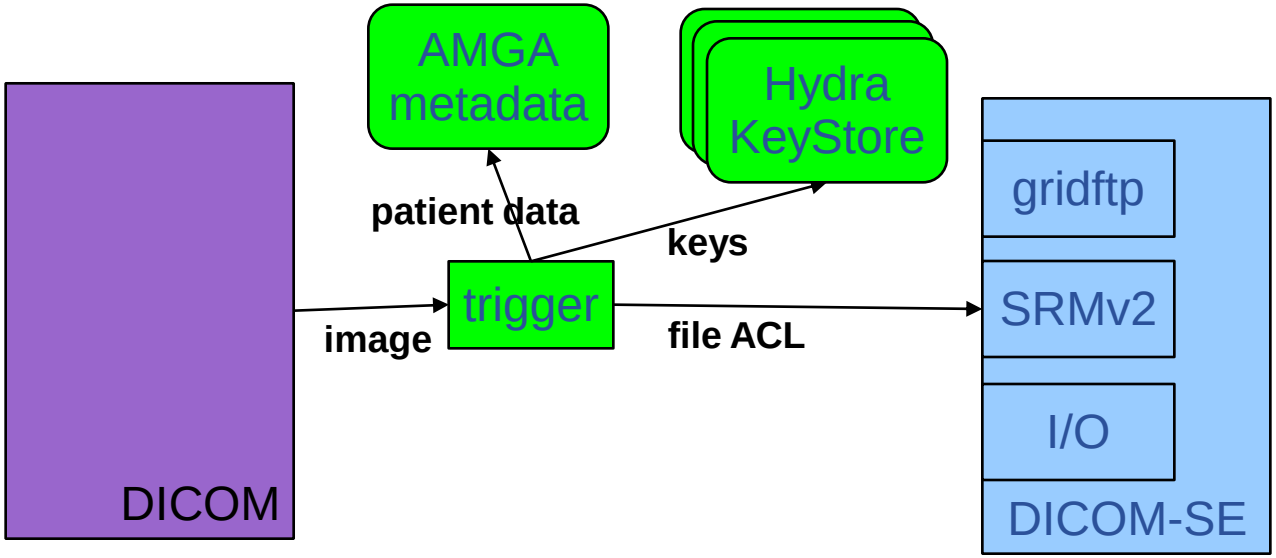
Goal: data access at any location

Exporting Images

"wrapping" DICOM :

- anonymity: patient data is separated and stored in AMGA
- access control: ACL information on individual files in SE (DPM)
- privacy: per-file keys
  - distributed among several Hydra key servers
  - fine grained access control

Image is retrieved from DICOM and processed to be "exported" to the grid.

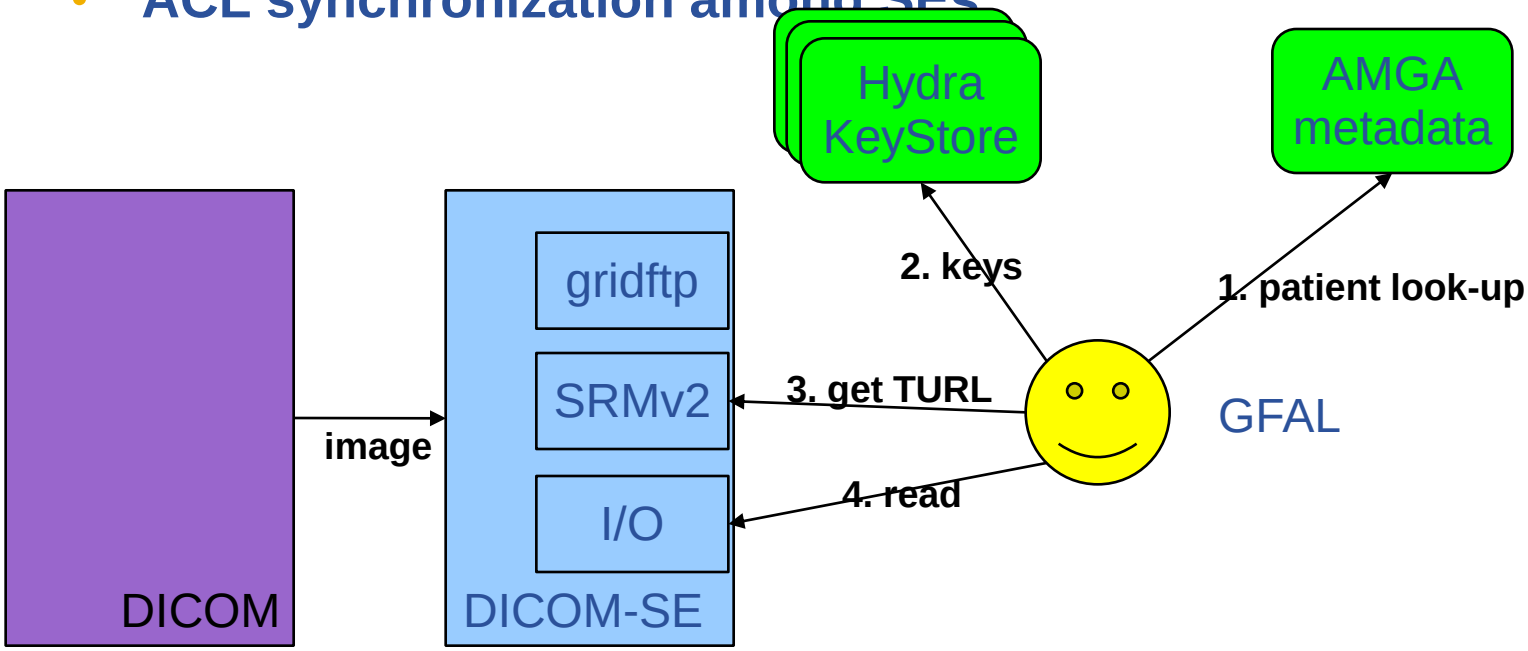


Accessing Images

- image ID is located by AMGA
- key is retrieved from the Hydra key servers
- file is accessed by SRM (access control in DPM)
- data is read and decrypted block-by-block in memory only (**GFAL** and hydra-cli)---> useful for all

Still to be solved:

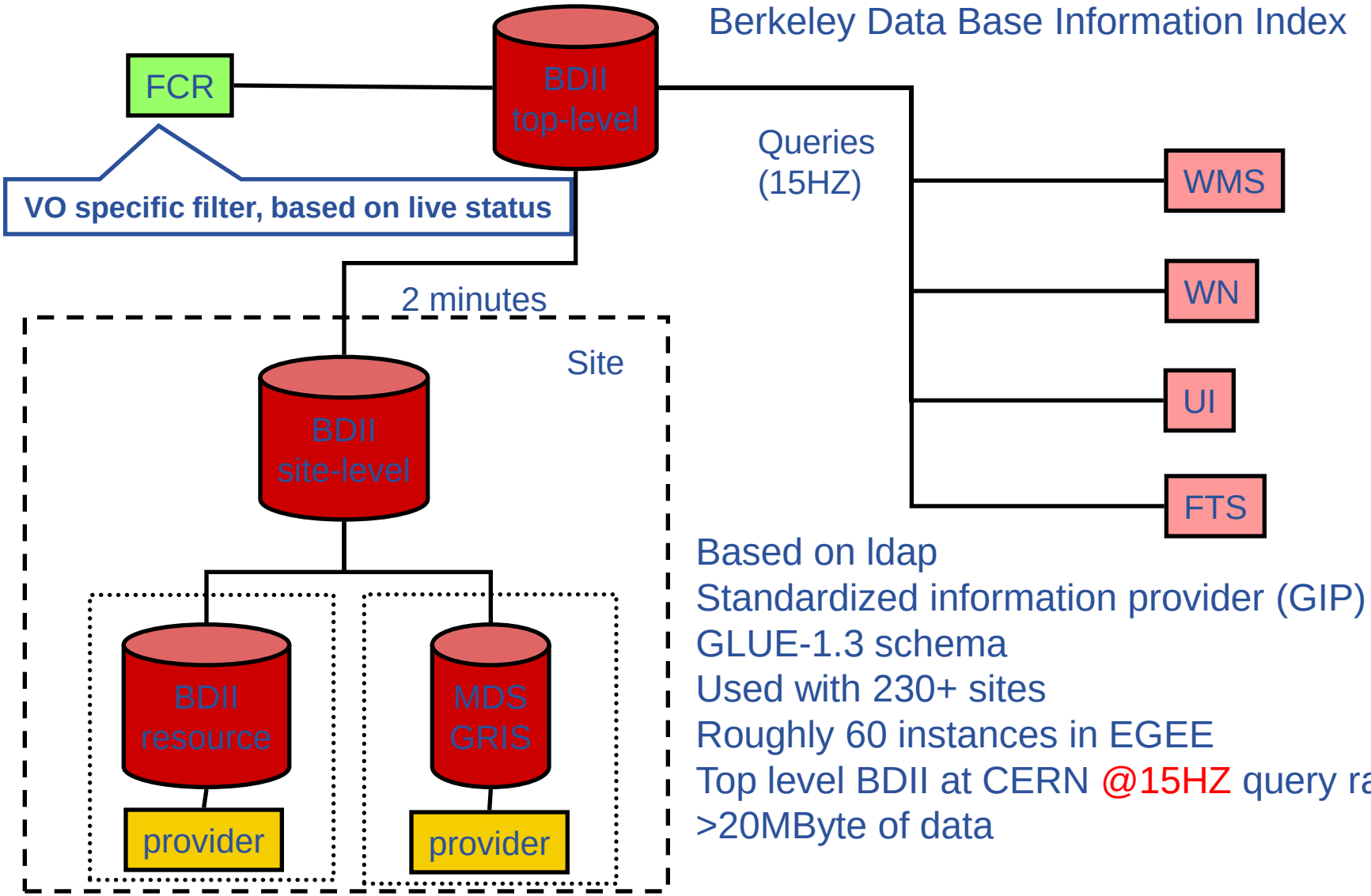
- ACL synchronization among SEs



Services and Clients

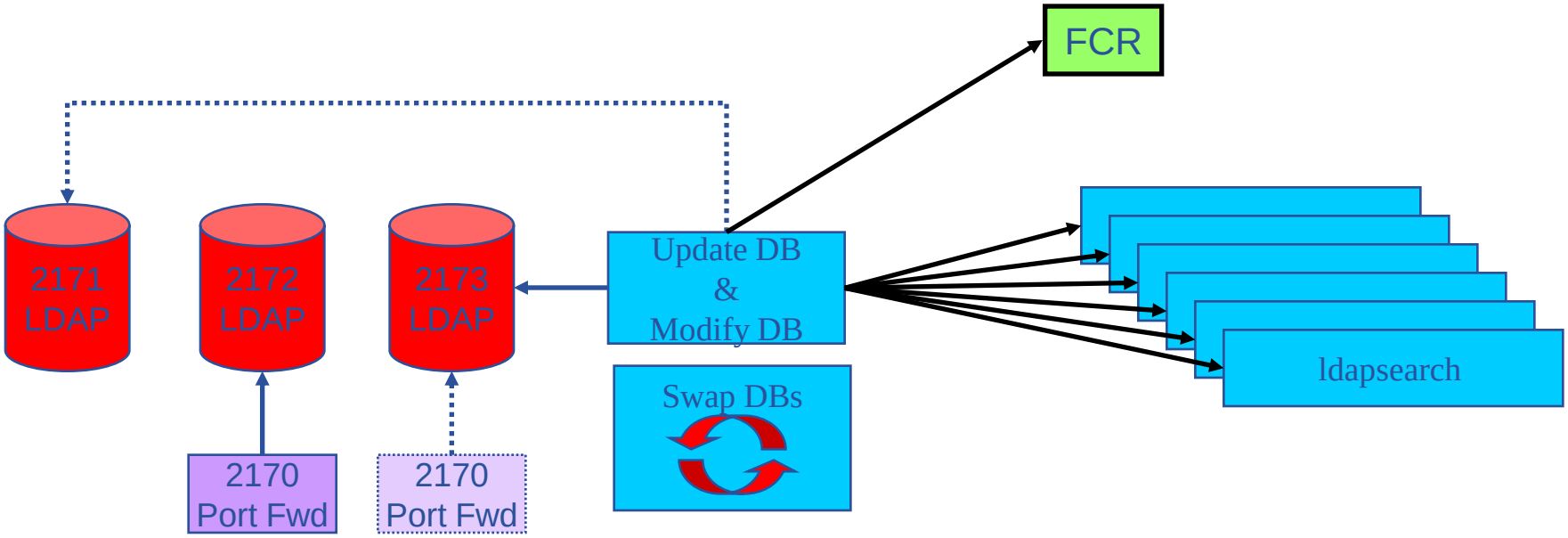
- Authentication and authorization: Myproxy, Voms
- Resource access: Computing Element, Worker Node, User Interface
- Workload Management system
- Logging and Bookkeeping
- Accounting
- Data Management: LCG File Catalog, Storage Element (DPM, dCache), GFAL, FTS, AMGA, Hydra
- **Information Systems: BDII**

The Information System

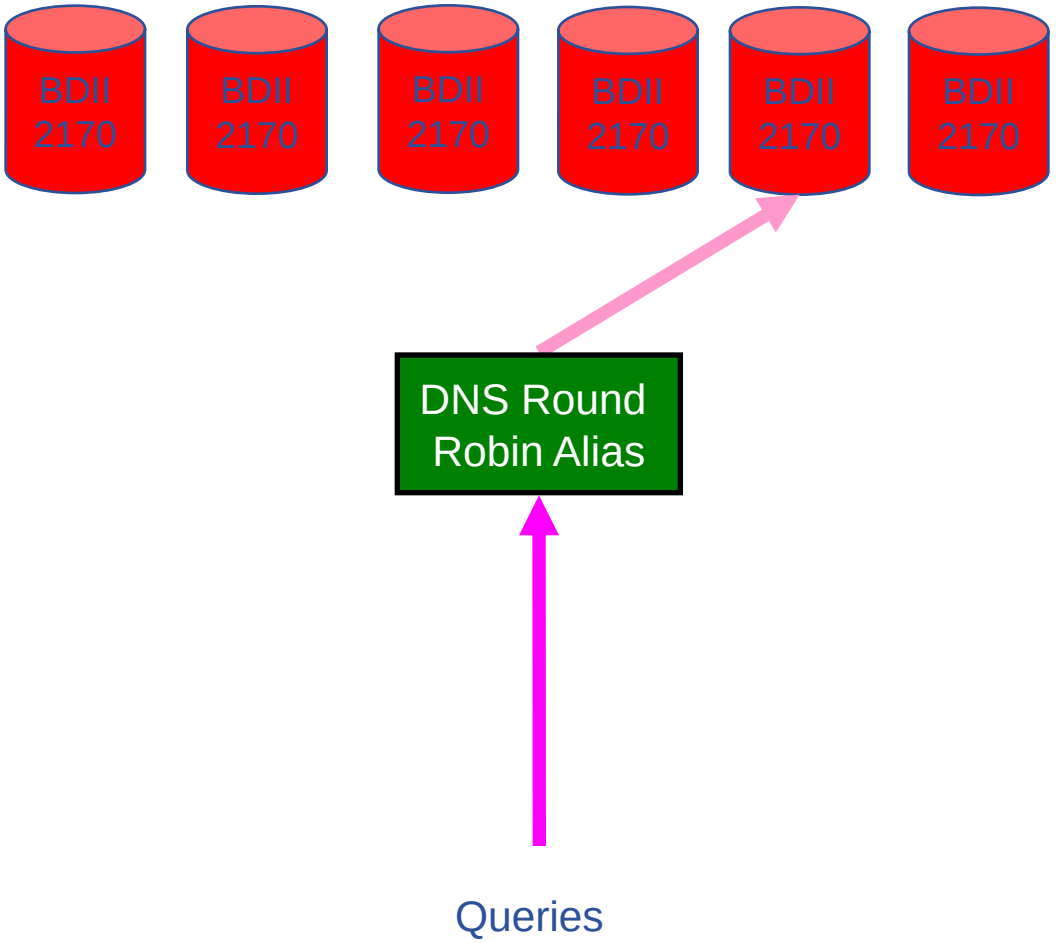


Based on ldap
Standardized information provider (GIP)
GLUE-1.3 schema
Used with 230+ sites
Roughly 60 instances in EGEE
Top level BDII at CERN @15HZ query rate
>20MByte of data

Inside A BDII

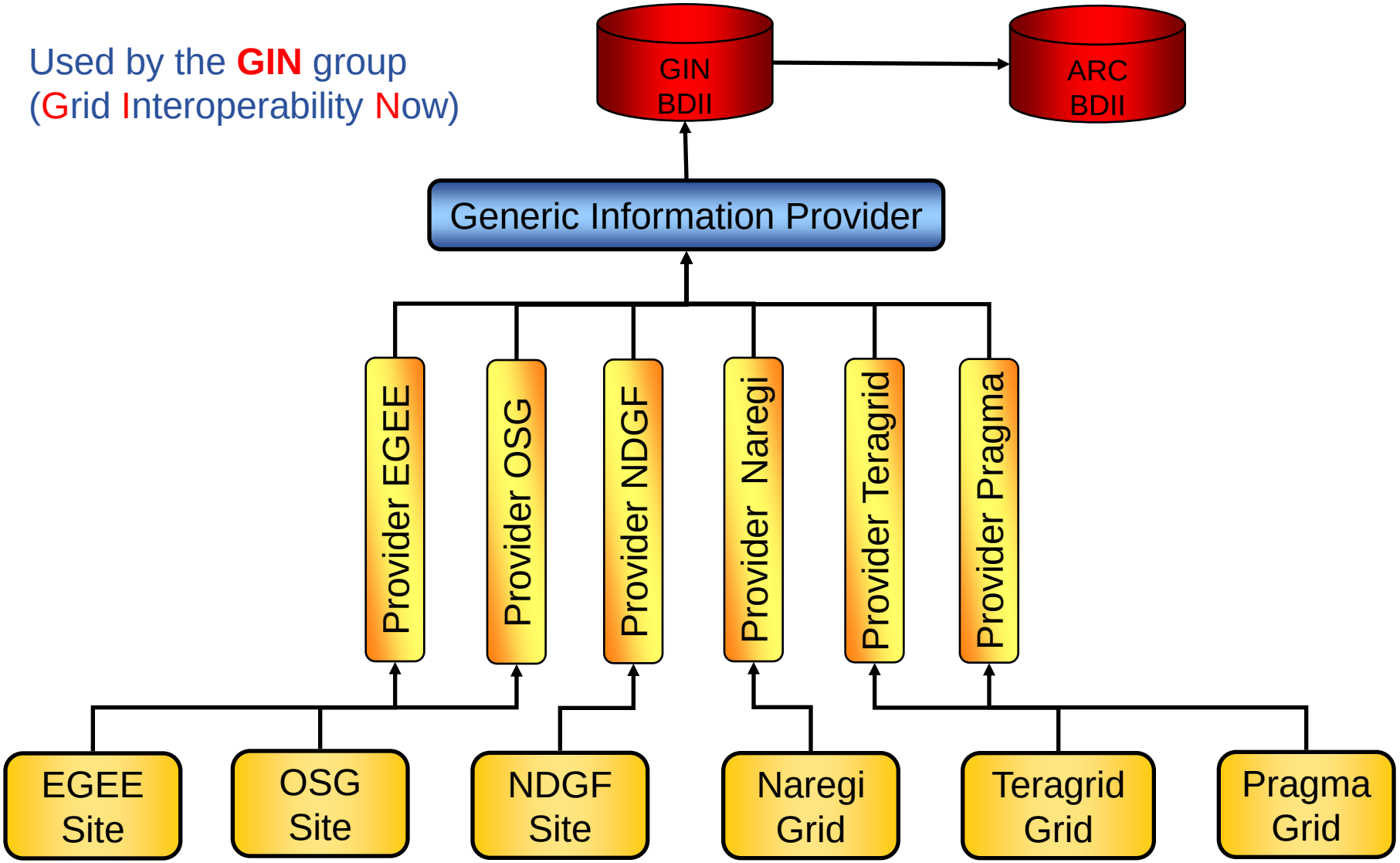


Load Balanced BDII



GIN BDII

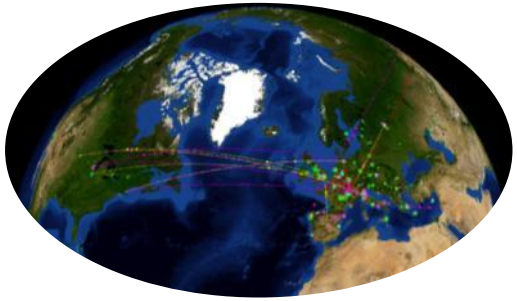
Used by the **GIN** group
(Grid Interoperability Now)



Content

- EGEE project and applications other than LCG
- gLite
- **EGEE operations**
- EGEE integration, testing and releases
- EGEEprojectstructure

The EGEE Infrastructure



Test-beds & Services

- Production Service
- Pre-production service
- Certification test-beds (SA3)
- Training infrastructure (NA4)

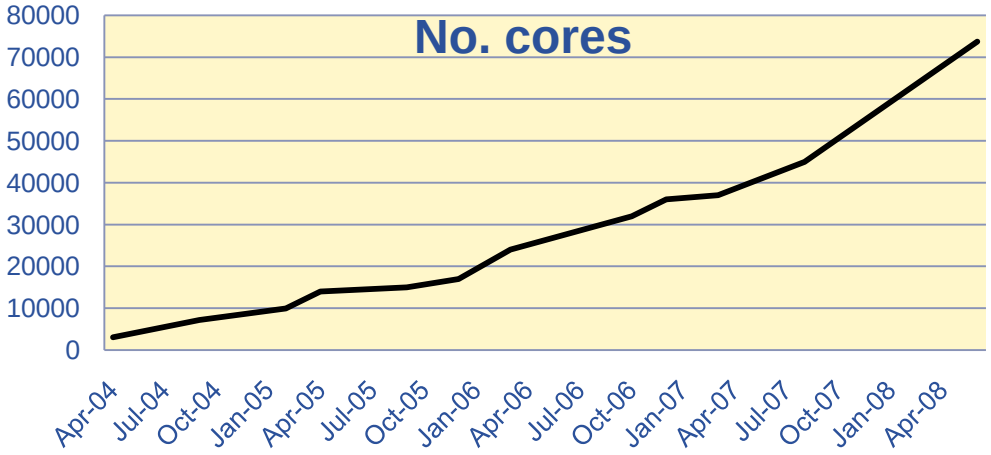
Support Structures & Processes

- Operations Coordination Centre
- Regional Operations Centres
- Global Grid User Support
- EGEE Network Operations Centre (SA2)
- Operational Security Coordination Team
- Training activities (NA3)

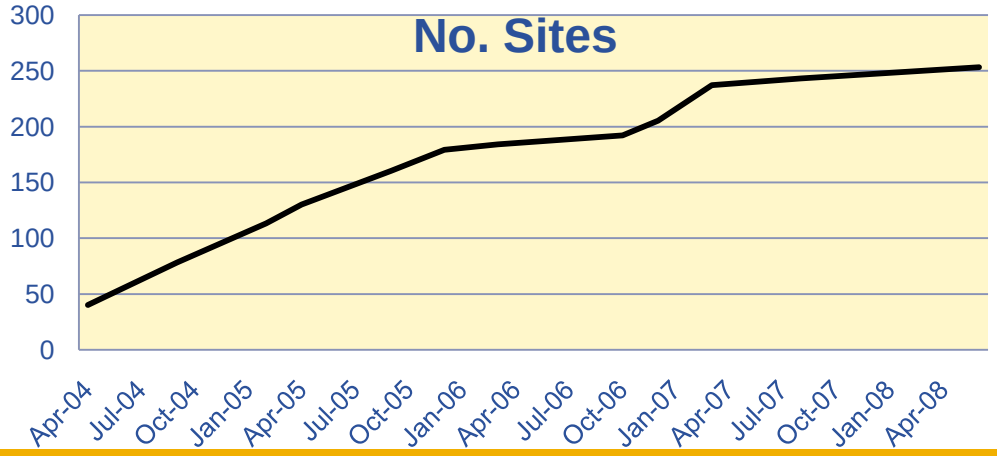
Security & Policy Groups

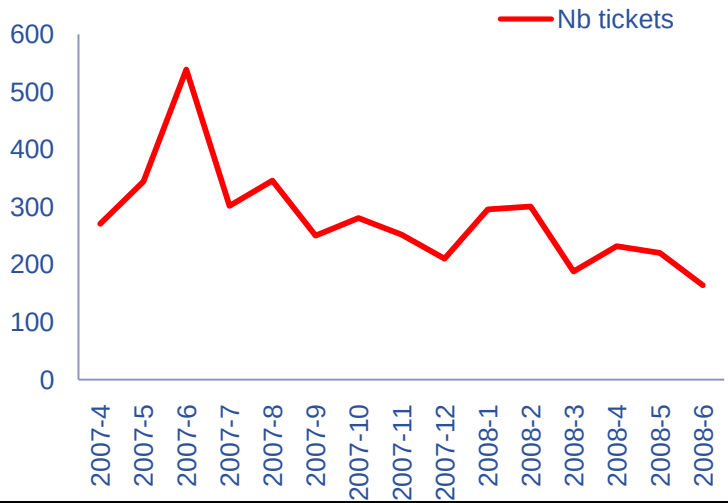
- Joint Security Policy Group
- EuGridPMA (& IGTF)
- Grid Security Vulnerability Group
- Operations Advisory Group (+NA4)

Cores, Sites, ROCs



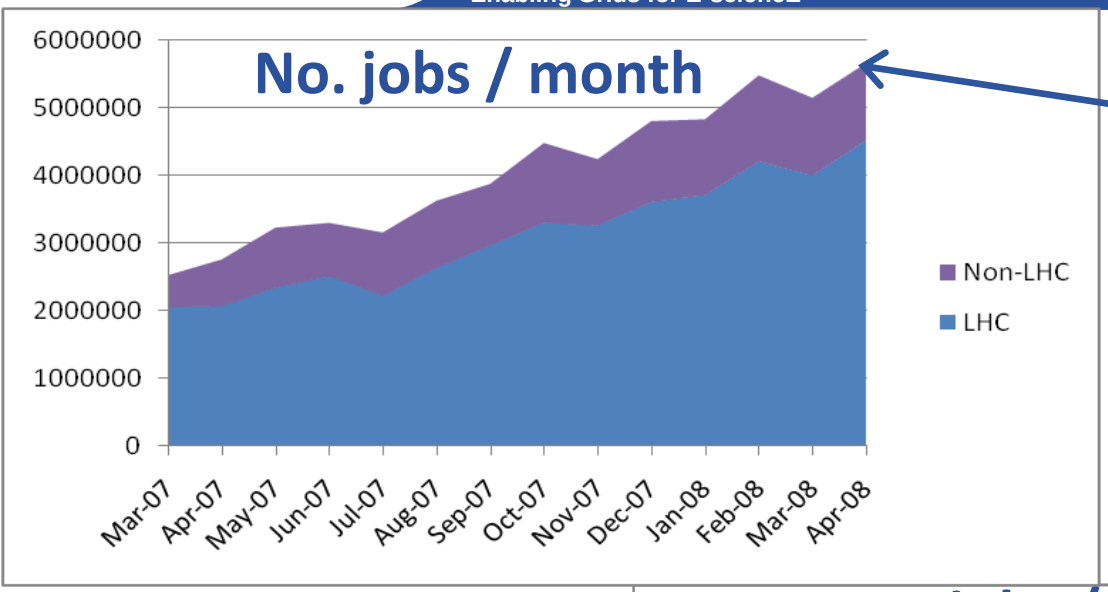
- 73709 cores
- 255 sites (145 partner sites)
- 48 countries (33 partner countries)





ROC	Partner - DoW	Partner - actual	Total	% non partner
CERN	1800	4856	6676	27%
France	1252	16203	16203	0%
De/CH	1852	8075	12536	36%
Italy	2280	6548	6571	0.4%
UK/I	2010	6618	12040	45%
CE	1163	2959	4711	37%
NE	1860	3207	4110	22%
SEE	1289	3606	3608	0.1%
SWE	898	1699	2280	25%
Russia	445	1378	1601	14%
A-P	801	1912	3373	43%
Total	**15650**	**57061**	**73709**	**23%**

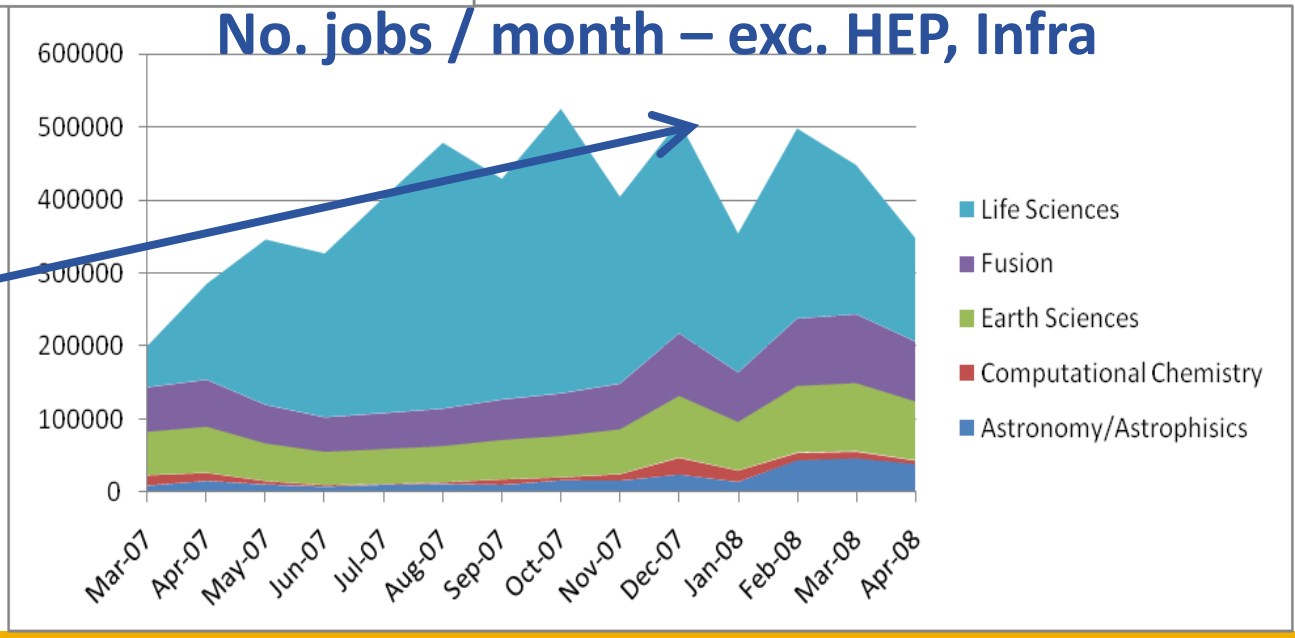
Workload



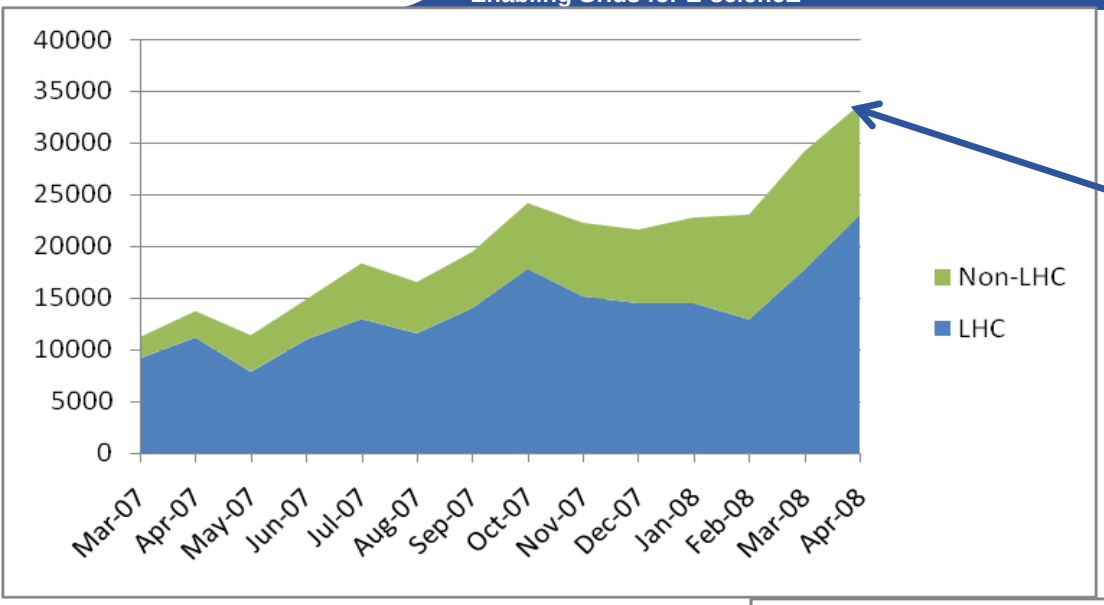
188.000 jobs/day
(98000 jobs/day 1y ago)

54 million jobs in the 2nd year
150K per day sustained average

17.000 jobs/day
(13000 jobs/day 1y ago)



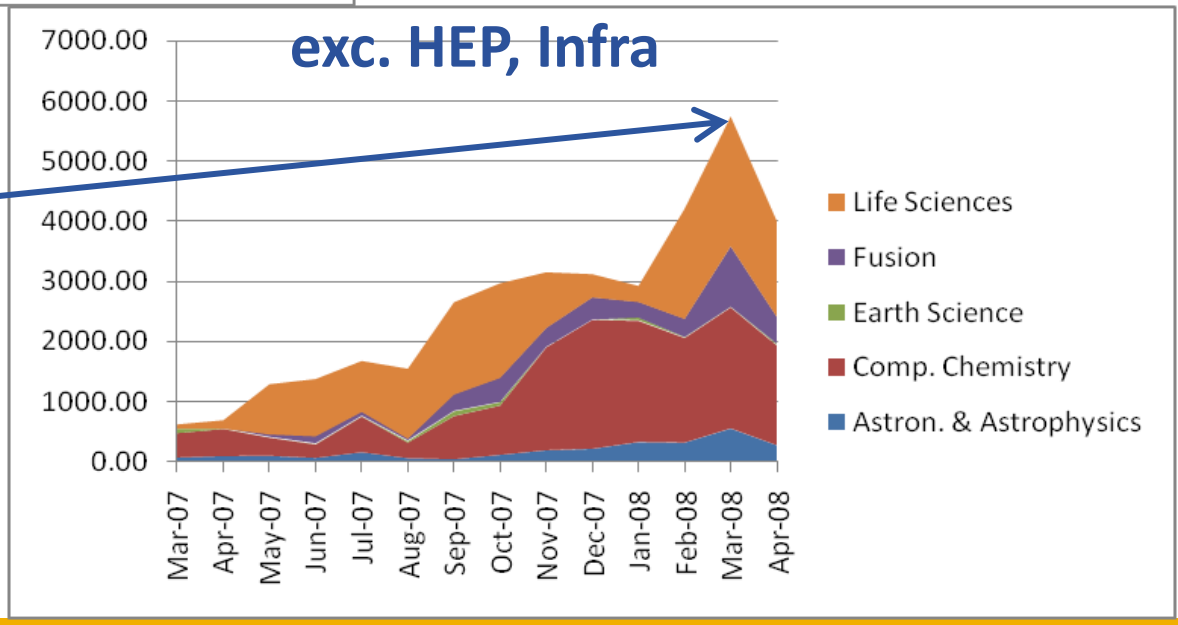
CPU time delivered (CPU months)



33.700 CPU-month
(14.000 CPU-month)

exc. HEP, Infra

Peak of 5700 CPU-month
(3600 CPU-month)



Operations: tools

- **GGUS: the main ticketing system**
- **GOCDB: the site database**
- **SAM: the site monitoring**
- **CIC portal: the “operations website”**
  - EGEE broadcast
  - COD dashboard
  - SAMAP: (one of the many) SAM user interfaces
- **Gstat: simple overview and statistics**
- **Gridview: a monitoring interface**
- **Gridmap: a view of the grid**

For all these tools you need a certificate by one of the grid CA to access the full functionality

The tools: GOCDB

- **Grid Operations Centres Database**
- **It is the DB for all sites in EGEE**
  - Used to declare the names of
    - Sites
    - Service nodes
    - Service managers
  - Used to declare (un)scheduled downtimes
- **http://goc.gridops.org**

If you are not listed there, you are not in the EGEE grid (production and pre-production)!

Gridview screeshot: data transfer

Gridview: Visualization and Monitoring Tool for LCG - Mozilla Firefox

http://gridview.cern.ch/GRIDVIEW/index.php

Monitoring and Visualization Tool for LCG

Data Transfer | Job Status | FTS | Service Availability

<< ABOUT ABOUT >>

What do you want ?

- VO-wise graphs
- Site-wise graphs
- Average Throughput
- Aggregate Data X-fer

VO: All VOs, Alice, Atlas

- Use Site Abbreviations
- Use Full Site Names

Source Site: All Sites

Destination Site: All Sites

- Current Summary
- Hourly Report
- Daily Report
- Weekly Report
- Monthly Report

From 00 Hours On 16 Jun 2008

To 11 Hours On 16 Jun 2008

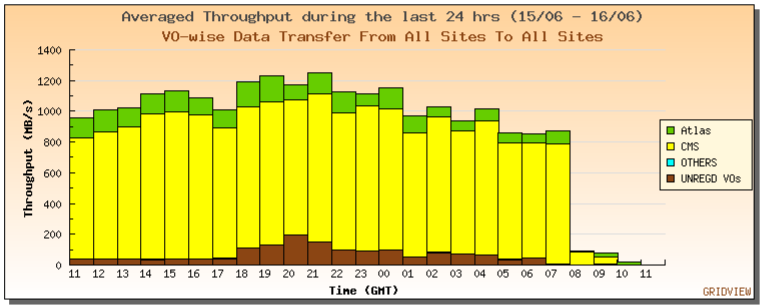
Display Graphs

Sites Abbreviations

Latest Gridftp Records

Current Status

(VO-wise Data Transfer From All Sites To All Sites)

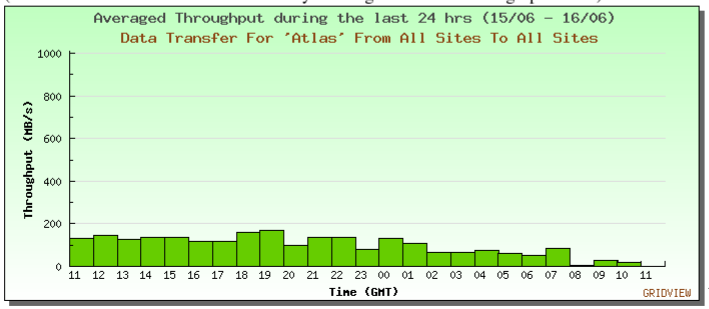


(**OTHERS:** VOs giving throughput less than 1% of max, click here for names)

Graphs for Individual VOs:-

Atlas | CMS | OTHERS | UNREGD VOs |

(Site-wise details for a VO can be seen by clicking over the relevant graph below)

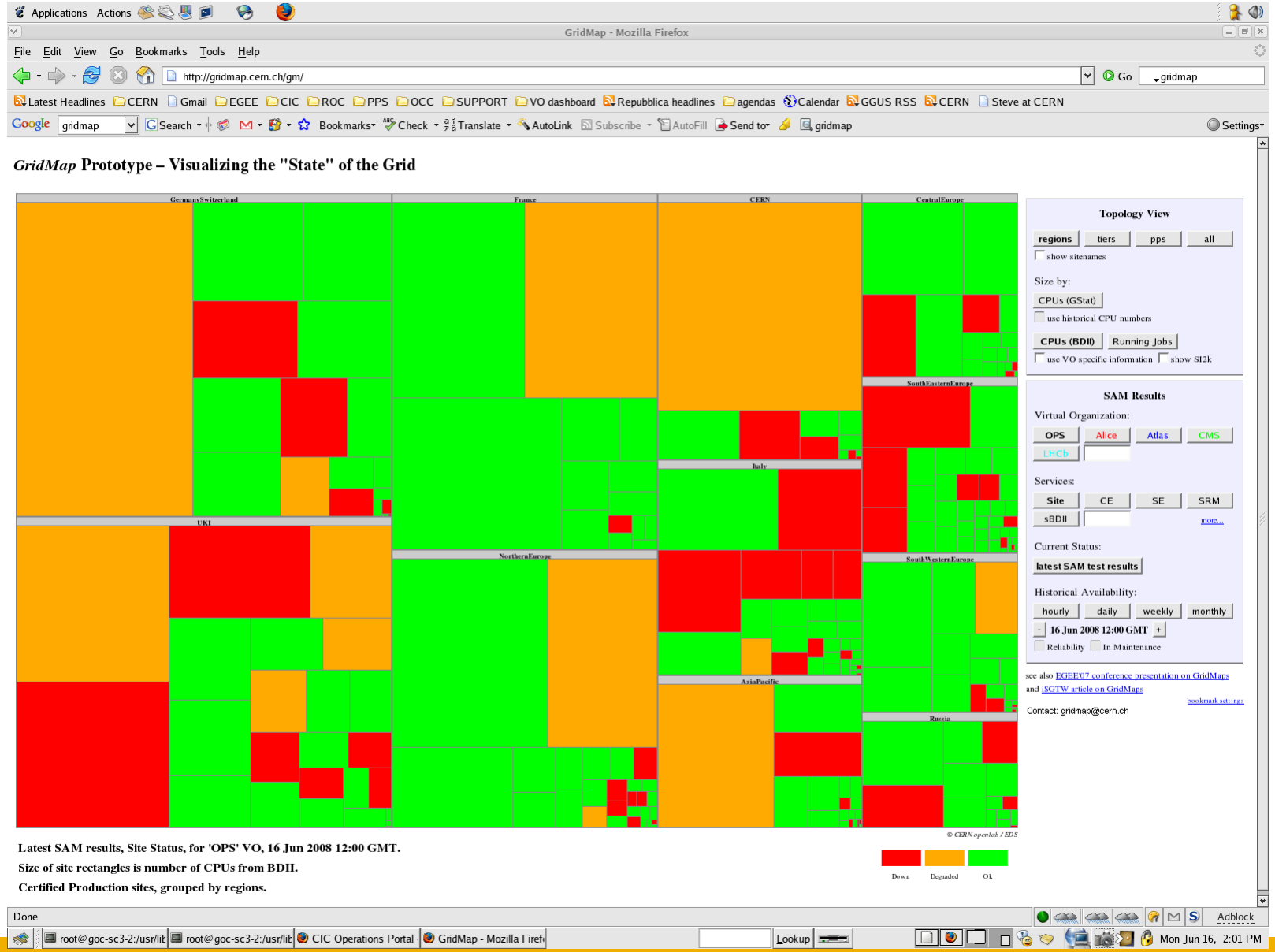


Averaged Throughput during the last 24 hrs (15/06 - 16/06)

Done

Mon Jun 16, 2:00 PM

Gridmap: screenshot



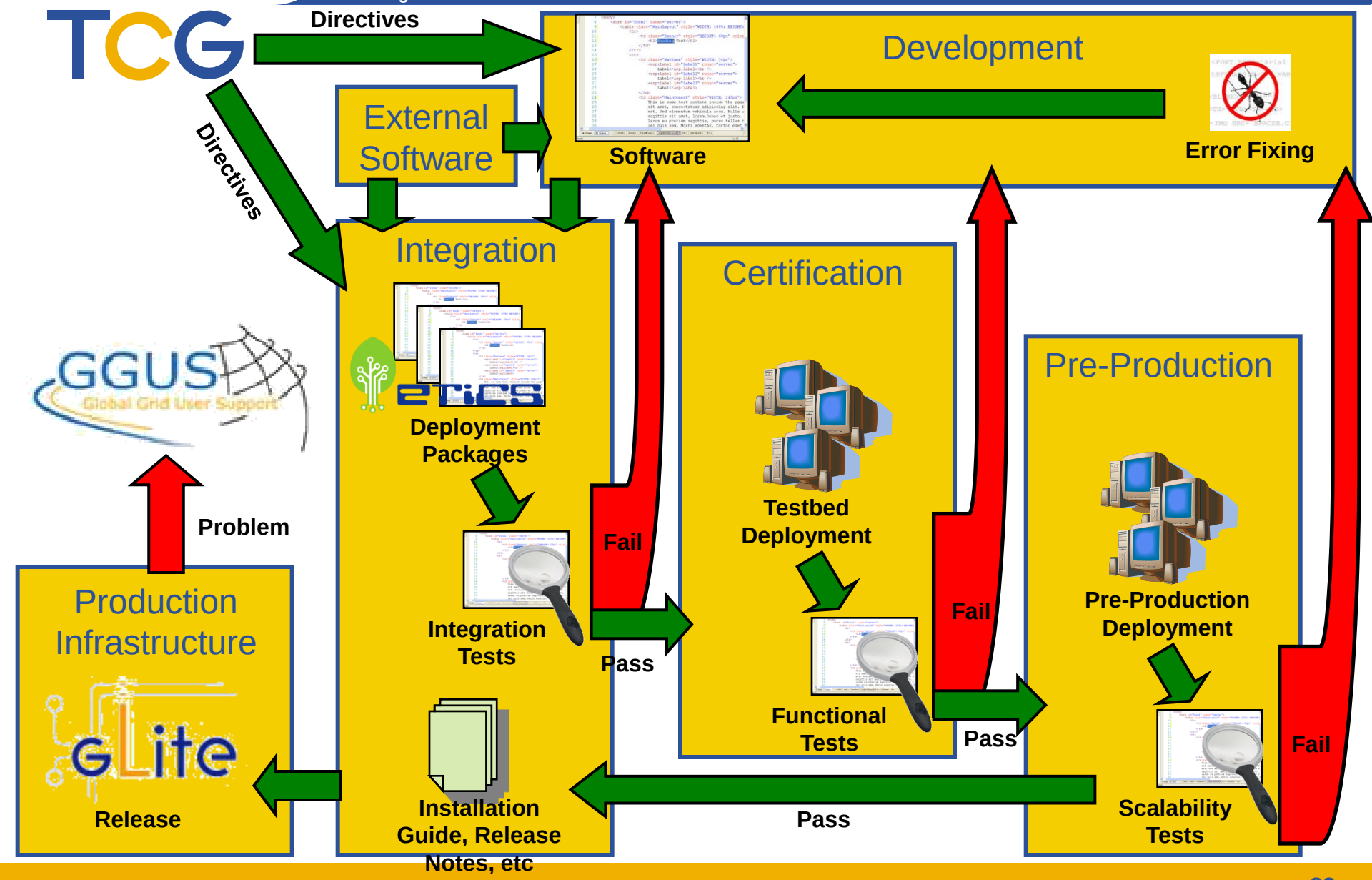
Content

- EGEE project and applications other than LCG
- gLite
- EGEE operations
- **EGEE integration, testing and releases**
- EGEEprojectstructure

Integration, testing and releases

- Integration and Packaging
- Testing and Certification
  - ***Functional and Stress Testing***
  - ***Security, Vulnerability Testing***
  - ***Operate Certification and Testing Test Beds***
  - ***Project Testing Coordination***
- Debugging, Analysis, Support
- Interoperation
- Support for porting
- Participate in standardization efforts

gLite Process

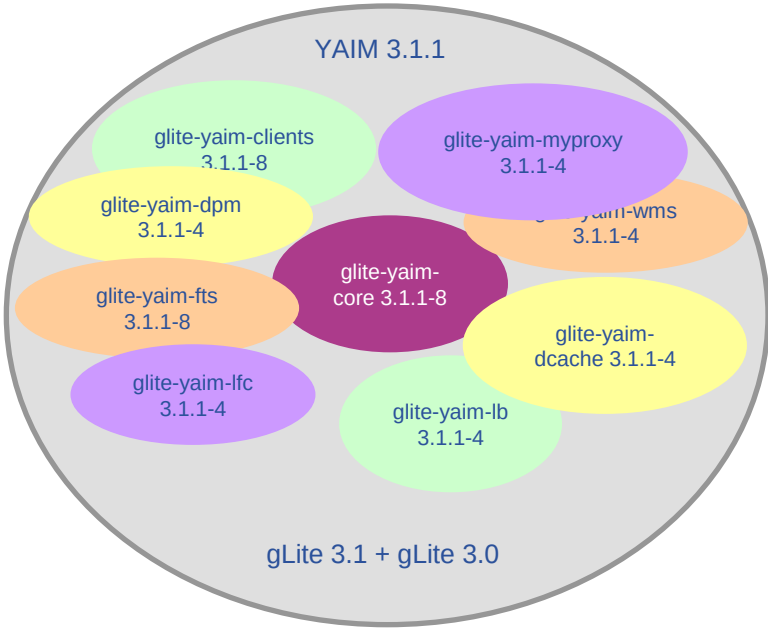


Usage

- **Process is in active use since July 2006**
  - Produced 26 sets of updates to the system in the first year
  - Second year:
    - Produced 23 sets of updates to gLite-3.0
    - Produced 17 sets of updates to gLite-3.1
  - Processed a total of **565** Patches
    - 361 for gLite-3.0, 204 for gLite-3.1
    - First year: **269** Patches
      - *Addressing 835 Change Requests*
  - During EGEE-II 3099 change requests have been opened
    - Increased usage and new use cases have uncovered more issues
    - 14% related to enhancements
    - 86% related to defects
    - Closed bugs: 1464 EGEE-II and 1002 EGEE-I

Configuration Management

- **YAIM: Simplicity**
  - Key-Value pairs + bash
- **Popular with site administrators**
  - Result of a survey
  - Easy to integrate with local tools
  - Easy to modify



- **Moved all components to YAIM**
  - Initially monolithic architecture
  - Every configuration change required an update to all components

Build Systems

- **Started with 3 systems**
  - LCG, gLite, ETICS
  - Complicate dependency management, release management
- **Moved to 1**
- **ETICS**
  - Used for the gLite-3.1 branch
  - Migration process to ETICS started in early August 06
    - Finished for almost all components September 2007
    - Last component moved February 2008
  - Overall experience has been positive
    - Functionality and performance has improved significantly over time
    - Multiplatform build support was very helpful

Test Cases

- **Central repository for tests**
  - Contains more than 250 test cases
  - During the second year we almost doubled the number of tests
  - Most progress has been achieved for the following components:
    - Clients (many options, quite good coverage)
    - Data management tests: SRM, DPM, LFC, FTS
    - Stress tests: WMS/LB, CE
- **Test development is mainly done by partners**
  - Formal follow-up on test development
    - *Progress is monitored and documented every 2 weeks*
- **Many tests (about 30%) come from outside sources**
  - Volunteers, other projects,…

- Status table at TCD:
  - http://cagraidsvr06.cs.tcd.ie/autobuild

Builds using ETICS version: 1.3.6-1

Worker Node Build Status													
ARCH	OS TYPE	VERSION	DISTRO	torque	VDT	deps	GridIre	Basic	RGMA	VOMS	DM	gfal	WN-dev
ia32	CentOS	4	yum	3/3	0/1	30/30	2/2	12/12	41/41	13/13	17/17	21/23	107/109
	CentOS	5	yum	3/3	4/4	30/30	2/2	12/12	41/41	14/14	17/17	20/20	106/109
	Debian	4	debs	3/3	1/1	29/30	1/1	12/12	41/41	14/14	16/17	16/20	95/107
	Solaris	10	pkg/tarball	3/3	1/1	23/23	2/2	12/12	33/41	0/11	7/17	7/20	N/A
	SuSE	10	apt	3/3	4/4	30/30	1/1	12/12	41/41	13/13	17/17	18/20	N/A
x86_64	CentOS	4	yum	3/3	1/1	26/26	2/2	9/9	41/41	15/15	18/18	21/21	90/108
	CentOS	5	yum	3/3	4/4	24/30	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	SuSE	10	apt	3/3	4/4	1/30	N/A	N/A	N/A	N/A	N/A	N/A	N/A
powerpc	Mac OS X	10.4	fink/tarball	3/3	1/1	1/30	1/1	12/12	33/41	0/11	0/17	11/29	57/109
	AIX	5	rpm/tarball	3/3	1/1	22/30	1/1	10/11	0/6	0/4	7/17	7/20	N/A
	Yellow Dog	6	yum	3/3	0/3	0/27	N/A	N/A	N/A	N/A	N/A	N/A	N/A

PSNC Build Results

Worker Node Build Status												
ARCH	OS TYPE	VERSION	DISTRO	torque	VDT	deps	Basic	RGMA	VOMS	DM	gfal	WN-dev
x86_64	Debian	4	debs	3/3	1/1	22/22	12/12	41/41	14/14	16/17	18/21	75/107

Obselete OS Build Results

Worker Node Build Status												
ARCH	OS TYPE	VERSION	DISTRO	torque	VDT	deps	GridIre	Basic	RGMA	VOMS	DM	WN-dev
ia32	SuSE	9	apt	3/3	3/3	28/28	2/2	12/12	41/41	14/14	17/17	107/107
x86_64	SLES	9	apt	3/3	1/1	24/24	1/1	9/9	37/41	10/15	12/18	74/109
powerpc	Mac OS X	10.3	fink/tarball	3/3	1/1	23/23	1/1	12/12	32/41	0/11	0/17	60/109

Legend	Colour	.	.	.	.
	Meaning	To be Started	Started	DONE	Not Applicable

Batch System Support

- **SA3 supports now:**
- **Torque/PBS -> reference platform**
  - LCG-CE, CREAM-CE
- **SGE**
  - LCG-CE, gLite-CE
- **Condor**
  - LCG-CE
- **LSF**
  - No direct support by a defined partner
  - LCG-CE, CREAM

Content

- EGEE project and applications other than LCG
- gLite
- EGEE operations
- EGEE testing, integration and release
- **EGEEproject structure**

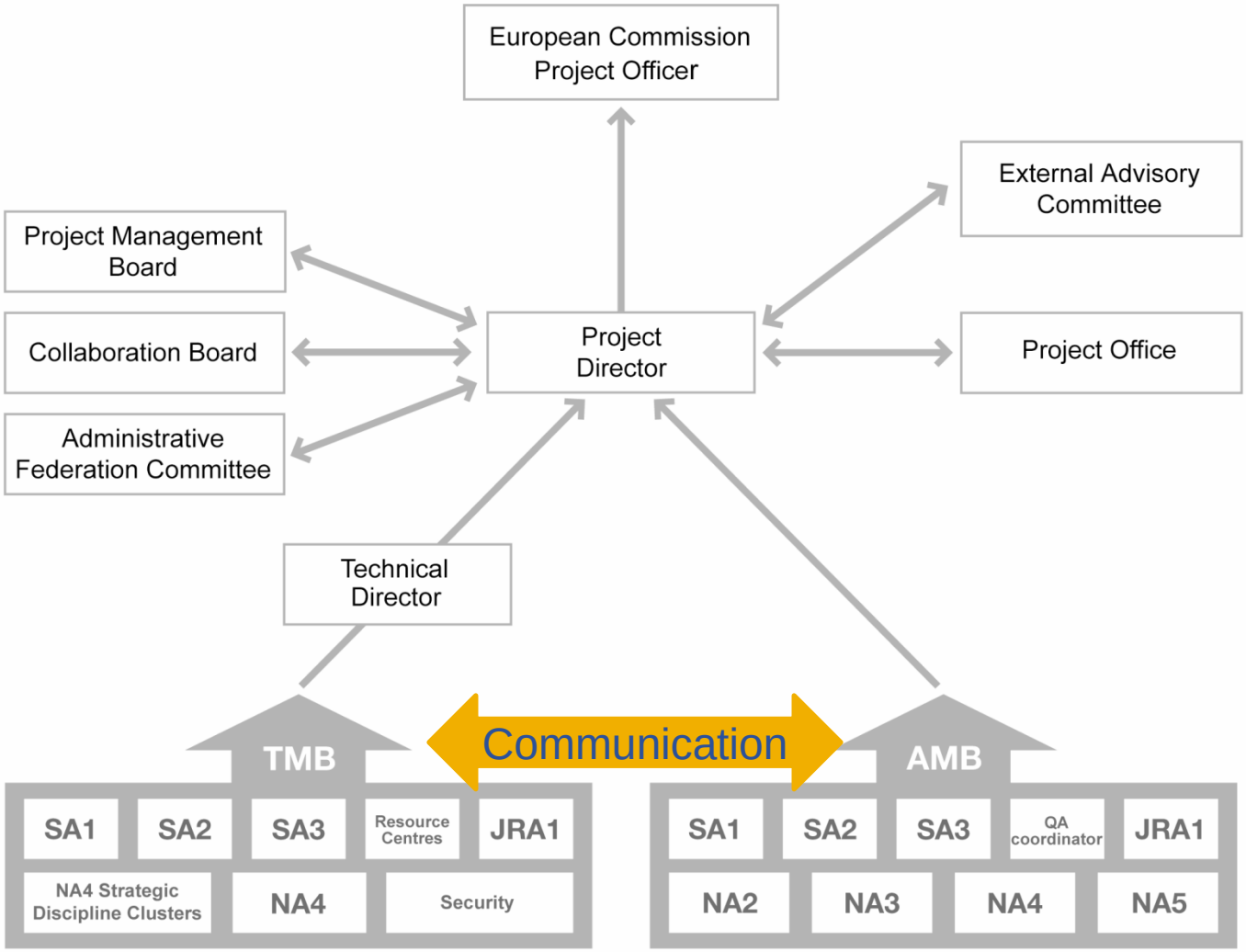
EGEE II in numbers (April 08)

- Total Budget: 52 792 100 Euros (EC funding: 36 971 365 Euros)
- Total partner institutions: 139
- Number of countries represented by project partners: 32
- Number of contributing staff: > 1000
- Number registered VOs using the EGEE infrastructure: > 130

EGEE-III Activities

Networking activities	*Specific Service Activities*
NA1: Management Bob Jones, CERN	SA1: Operations Maite Baroso Lopez, CERN
NA2: Dissemination Hiring in progress, CERN	SA2: Networking Support Xavier Jeannin, CNRS
NA3: Training Robin McConnell, UEDIN	SA3: Integration, testing & cert. Oliver Keeble, CERN
NA4: Applications Cal Loomis, CNRS	*Joint Research Activities*
NA5: International Coop. & Policy Panos Louridas, GRNET	JRA1: Middleware engineering Francesco Giacomini, INFN

EGEE-III management structure



The European Grid Initiative

- Need to prepare permanent, common Grid infrastructure
- Ensure the long-term sustainability of the European e-Infrastructure independent of short project funding cycles
- Coordinate the integration and interaction between National Grid Infrastructures (NGIs)
- Operate the production Grid infrastructure on a European level for a wide range of scientific disciplines