



Enabling Grids for E-science

EGEE – application support and identification

Vincent Breton, on behalf of NA4

Application identification and support manager

CNRS, France

Second EGEE review

December 6th 2005



Information Society



www.eu-egee.org

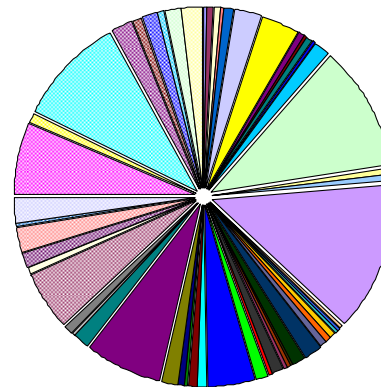
- **Status of applications: achievements since last review**
 - Deployment on the production infrastructure
 - Deployment on the PreProduction Service
 - Other achievements
- **Response to comments raised at last review**
- **Perspectives**
 - Short term: User forum March 1-3 2005
 - Mid term: EGEE-II
- **Conclusion**

What happened since last review

- **The number of users in VOs related to NA4 activity kept growing regularly**
 - from ~500 at PM9 to ~1000 at PM18
 - More than 20 applications are deployed on the production infrastructure
- **The usage of the grid by pilot applications has significantly evolved**
 - From data challenge to service challenge (HEP)
 - First biomedical data challenge (WISDOM)
- **Several existing applications have been migrated to the new middleware by the HEP, biomedical and generic teams**

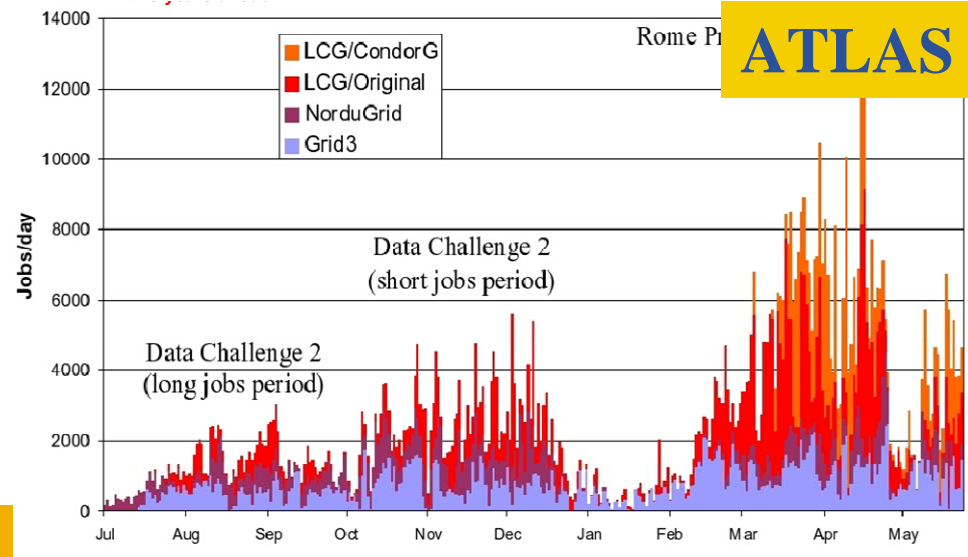
- **Fundamental activity in preparation of LHC start up**
 - Physics and detector studies
 - Computing systems commissioning
- **Examples:**
 - LHCb: >700 CPU/years in 2005 on the EGEE infrastructure
 - ATLAS: >10,000 jobs per day
 - Comprehensive analysis: see S.Campana et al., e-Science 2005, Melbourne, Australia
 - A lot of activity in all involved applications (including as usual a lot of activity within non-LHC experiments like BaBar, CDF and D0)
 - More details in DNA4.3.2

CPU used: 6,389,638 h
Data Output: 77 TB



Site	Percentage	Site	Percentage
DIRAC.Barcelona.es	0.214%	DIRAC.Zurich.ch	0.003%
DIRAC.CERN.ch	0.571%	LCG.ACAD.bg	0.106%
DIRAC.CracowAgu.pl	0.001%	LCG.Barcelona.es	0.281%
DIRAC.LHCBOONLINE.ch	0.779%	LCG.Bologna.it	0.032%
DIRAC.PNPI.ru	0.000%	LCG.CESGA.es	0.528%
DIRAC.ScotGrid.uk	3.068%	LCG.CNAF-GRIDIT.it	0.012%
DIRAC.Zurich.ch	0.756%	LCG.CNB.es	0.385%
LCG.BHAM-HEP.uk	0.705%	LCG.CSCS.ch	0.282%
LCG.Bari.it	1.357%	LCG.CY01.cy	0.103%
LCG.CERN.ch	10.960%	LCG.Cambridge.uk	0.010%
LCG.CGG.fr	0.676%	LCG.Durham.uk	0.476%
LCG.CNAF.it	13.196%	LCG.FZK.de	1.708%
LCG.CPPM.fr	0.242%	LCG.Firenze.it	1.047%
LCG.CY01.cy	0.103%	LCG.GR-02.gr	0.226%
LCG.Cambridge.uk	0.010%	LCG.GR-04.gr	0.056%
LCG.Durham.uk	0.476%	LCG.HPC2N.se	0.001%
LCG.FZK.de	1.708%	LCG.IFCA.es	0.022%
LCG.Firenze.it	1.047%	LCG.IN2P3.fr	4.143%
LCG.GR-02.gr	0.226%	LCG.IPP.bg	0.033%
LCG.GR-04.gr	0.056%	LCG.Imperial.uk	0.891%
LCG.HPC2N.se	0.001%	LCG.JINR.ru	0.472%
LCG.IFCA.es	0.022%	LCG.Lancashire.uk	6.796%
LCG.IN2P3.fr	4.143%	LCG.Manchester.uk	0.285%
LCG.IPP.bg	0.033%	LCG.Montreal.ca	0.069%
LCG.Imperial.uk	0.891%	LCG.NSC.se	0.465%
LCG.JINR.ru	0.472%	LCG.Oxford.uk	1.214%
LCG.Lancashire.uk	6.796%	LCG.PNPI.ru	0.278%
LCG.Manchester.uk	0.285%	LCG.Pisa.it	0.121%
LCG.Montreal.ca	0.069%	LCG.RAL-HEP.uk	0.938%
LCG.NSC.se	0.465%	LCG.RHUL.uk	2.168%
LCG.Oxford.uk	1.214%	LCG.Sheffield.uk	0.094%
LCG.PNPI.ru	0.278%	LCG.Toronto.ca	0.343%
LCG.Pisa.it	0.121%	LCG.UCL-CCC.uk	1.455%
LCG.RAL-HEP.uk	0.938%		
LCG.RHUL.uk	2.168%		
LCG.Sheffield.uk	0.094%		
LCG.Toronto.ca	0.343%		
LCG.UCL-CCC.uk	1.455%		

LHCb

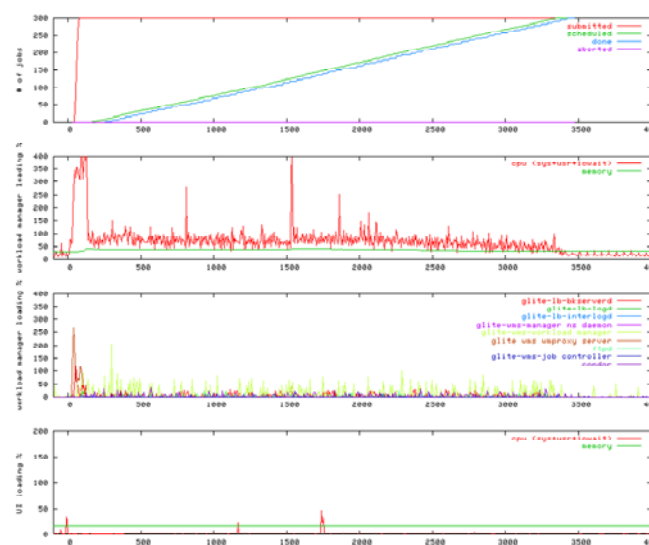


ATLAS

- **Task Forces**
 - EGEE + experiment people. Very close and positive collaboration. Emphasis on integration onto EGEE infrastructure

- **One example (direct contribution of NA4-HEP/ARDA within the ATLAS TF)**

- Detailed studies of advanced gLite feature: WMS bulk submission (ATLAS Task Force)
- Other middleware tested in the framework of other ARDA prototypes and TFs
- More under arda.cern.ch



First gLite pre1.4 WMS performance measurements in Milano

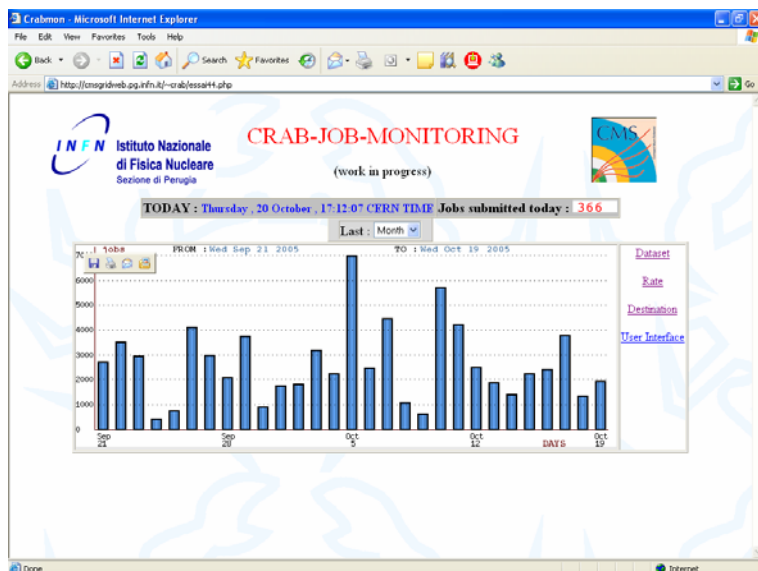
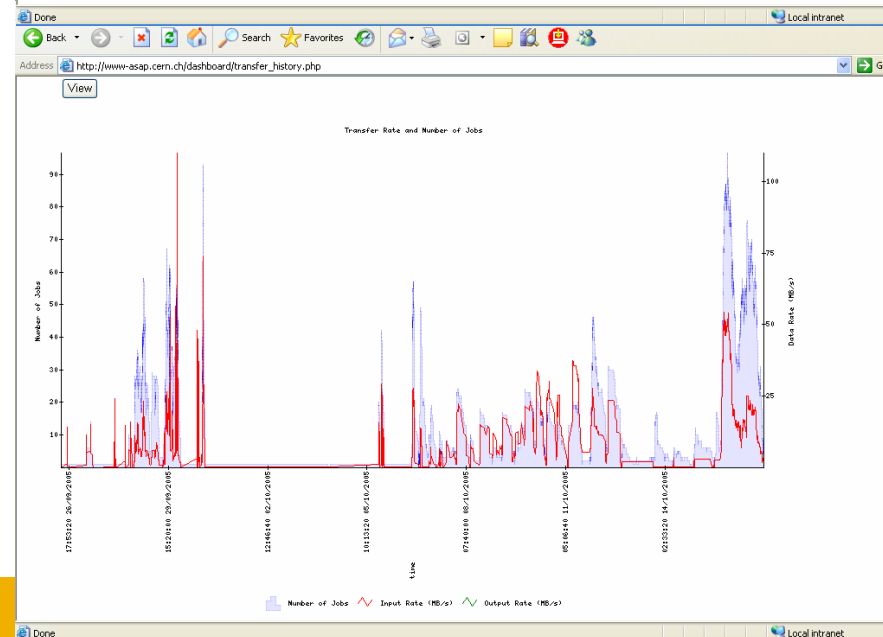
- gLite 1.4 WMS (4CPU, 3GB memory)
- 300 simple hello world jobs submitted by 3 parallel threads
- submission rate ~ 4.1 jobs/sec
- dispatching rate ~ 0.08 jobs/sec
- All jobs in bulk are submitted to the same CE: atlasce.inf.infn.it

• Thanks to Elisabetta Molinari for setting up the gLite WMS in Milan

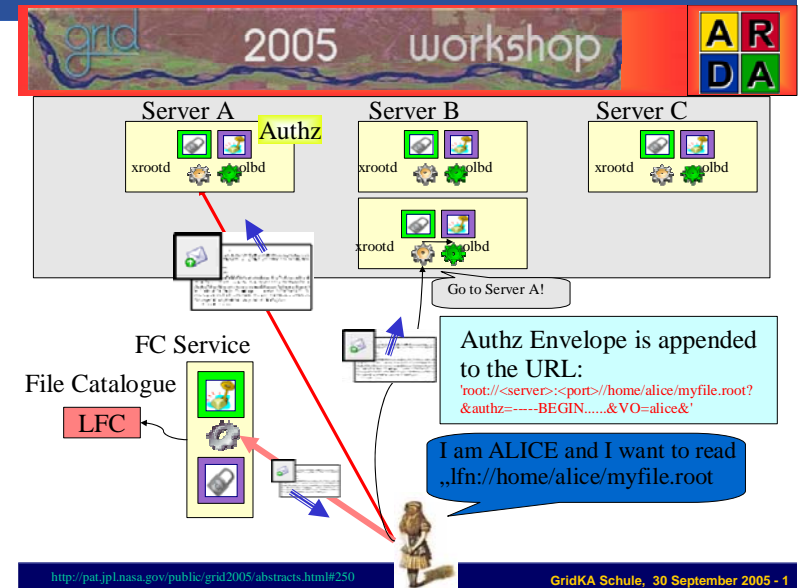
30 Sept. 2005

LCG/EGEE Taskforce Meeting

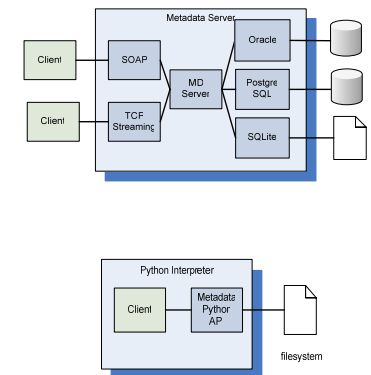
- From prototypes to coherent integration (CMS)
 - NA4 HEP (ARDA/ASAP) prototype → converging on the CMS CRAB system
 - SC3 activity, analysis jobs, productions jobs → CMS dashboard
 - Clear signs of wide user activity!!!
 - ASAP and CRAB
 - ~150k jobs/month



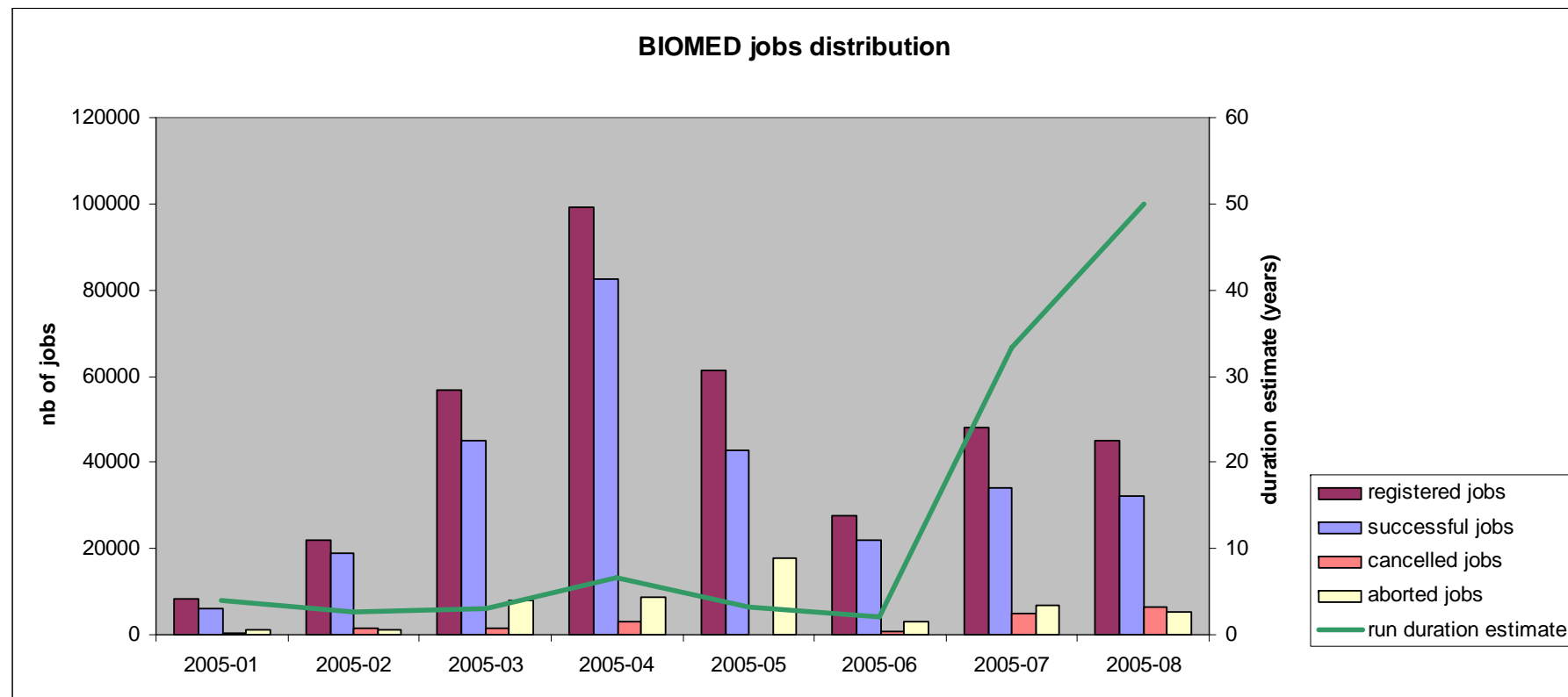
- Other examples:
 - Efficient data access integration: File catalogue ACL and fast data access via xrootd (ALICE prototype and Task Force). Presentation at SuperComputing 2005
 - Improved user access: GANGA (ATLAS and LHCb activities). Good feedback and demo in Pisa
 - Contribution to services used also outside HEP and contributed to gLite (AMGA). Used in the demo of Biomed and “Generic” in Pisa EGEE Conference



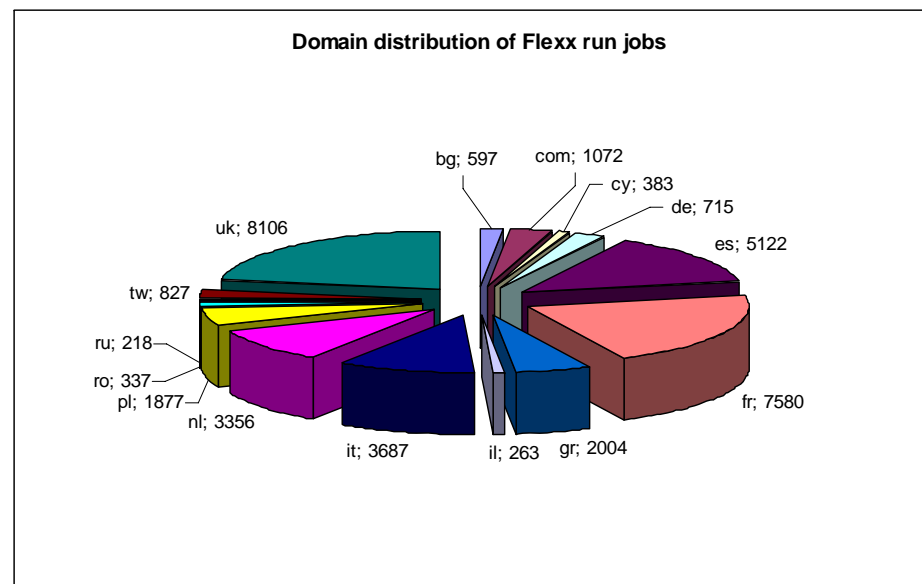
- **Prototype**
 - Validate our ideas and expose a concrete example to interested parties
- **Multiple back ends**
 - Currently: Oracle, PostgreSQL, SQLite, MySQL
- **Dual front ends**
 - TCP Streaming
 - Chosen for performance
 - SOAP
 - Formal requirement of EGEE
 - Compare SOAP with TCP Streaming
- **Also implemented as standalone Python library**
 - Data stored on the file system



- ~ 70 users, 9 countries
- > 12 Applications (medical image processing, bioinformatics)
- ~3000 CPUs, ~12 TB disk space
- ~100 CPU years, ~ 500K jobs last 6 months



- **Significant biological parameters**
 - two different molecular docking applications (Autodock and FlexX)
 - about one million virtual ligands selected
 - target proteins from the parasite responsible for malaria
- **Significant numbers**
 - Total of about 46 million ligands docked in 6 weeks
 - 1TB of data produced
 - Up 1000 computers in 15 countries used simultaneously corresponding to about 80 CPU years
 - Average crunching factor ~600

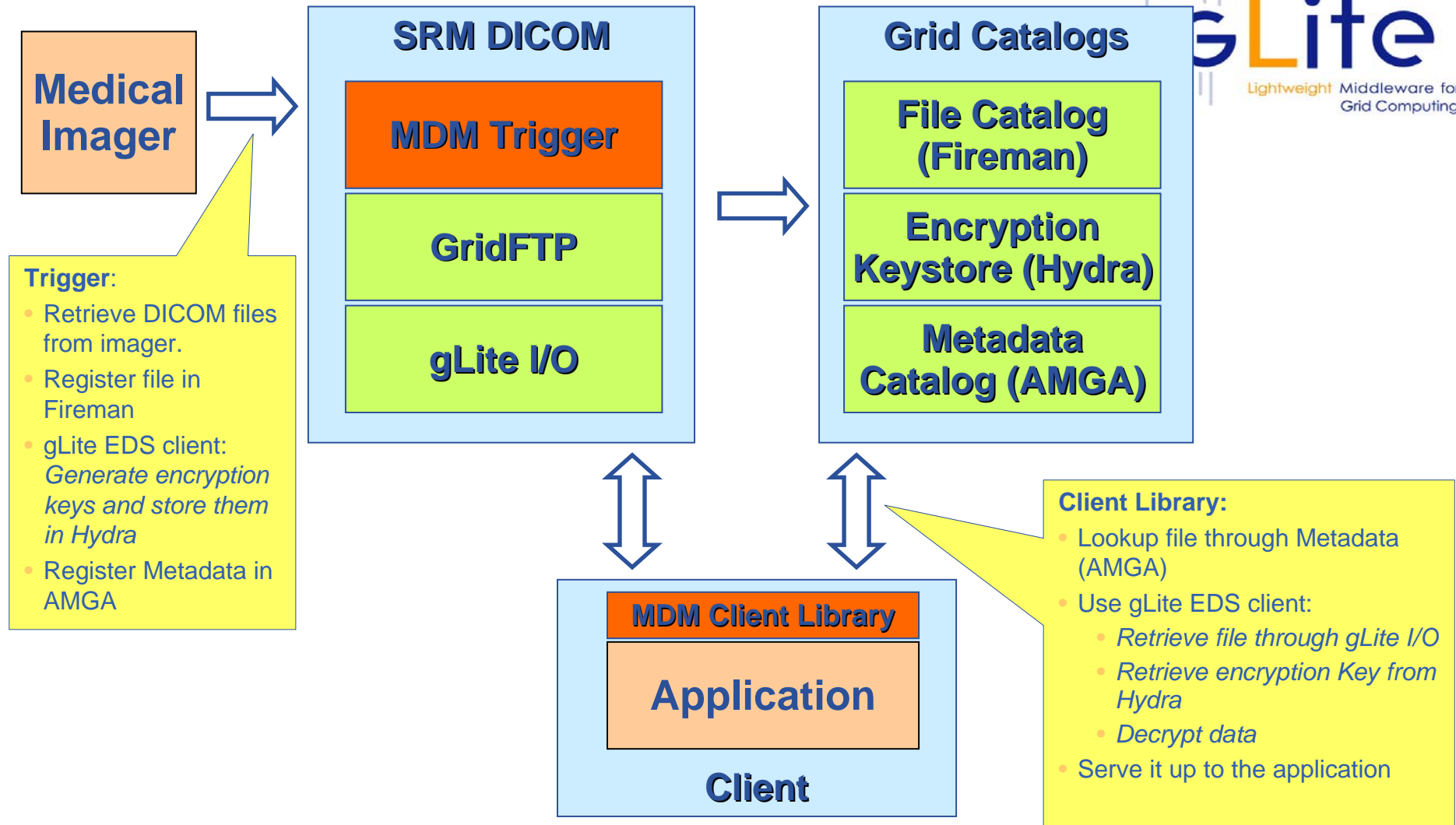


WISDOM open day

December 16th, 2005, Bonn (Germany)

Discuss Data Challenge results
Prepare next steps towards a malaria Grid (EGEE-II, Embrace, Bioinfogrid)
Information: <http://wisdom.eu-egge.fr>

- **PPS components tested**
 - Weekly testing of WMS
 - FiReMan
 - ACLs for file access
 - R-GMA
- **Other gLite testing**
 - Site installation (v1.0, 1.1, 1.2 and 1.3)
 - Advanced data management on prototype testbed
 - FiReMan, File Transfer Service
 - Metadata (AMGA)
 - File encryption
 - dCache SRM (SRM-DICOM interface)
 - Security features really needed for biomedical applications: expected on PPS soon!

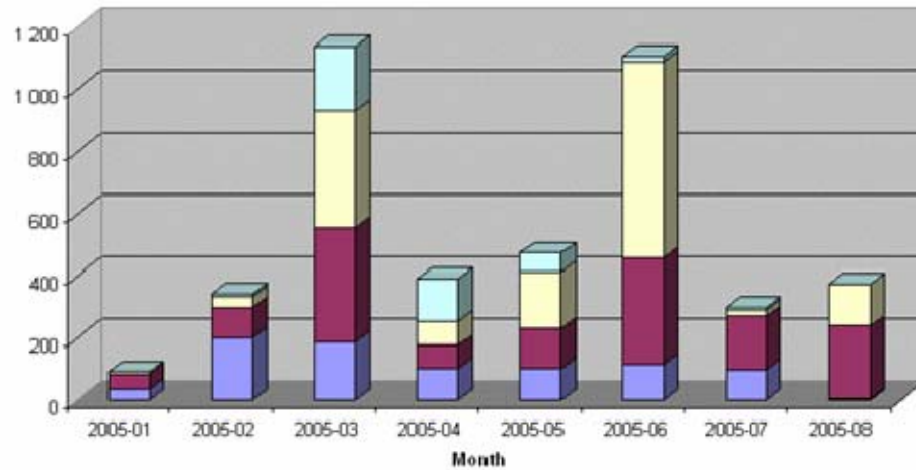


- **Applications accepted by Project Executive Board**
 - Earth Science Research (Earth Observation, Hydrology, Climate)
 - Geophysics (Industry)
 - Computational Chemistry
 - Astrophysics (MAGIC and Planck collaborations)
 - Finance (EGRID)

- **New Applications recommended to PEB by EGEE Generic Applications Advisory Panel (EGAAP)**
 - Fusion (ITER)
 - Archaeology
 - EC projects (EELA, EUMEDGRID, EUCHINAGRID, BIOINFOGRID)

Number of jobs

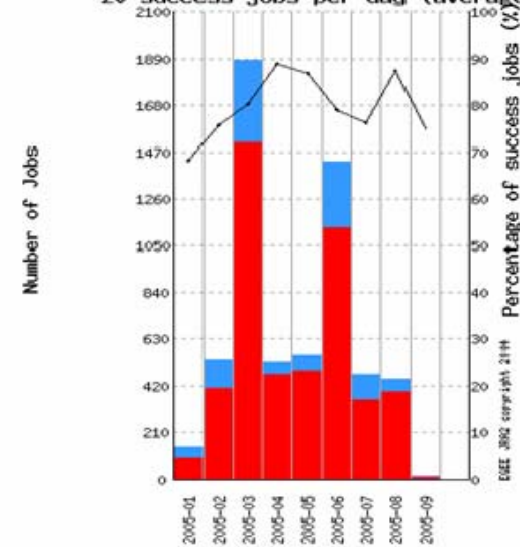
ESR



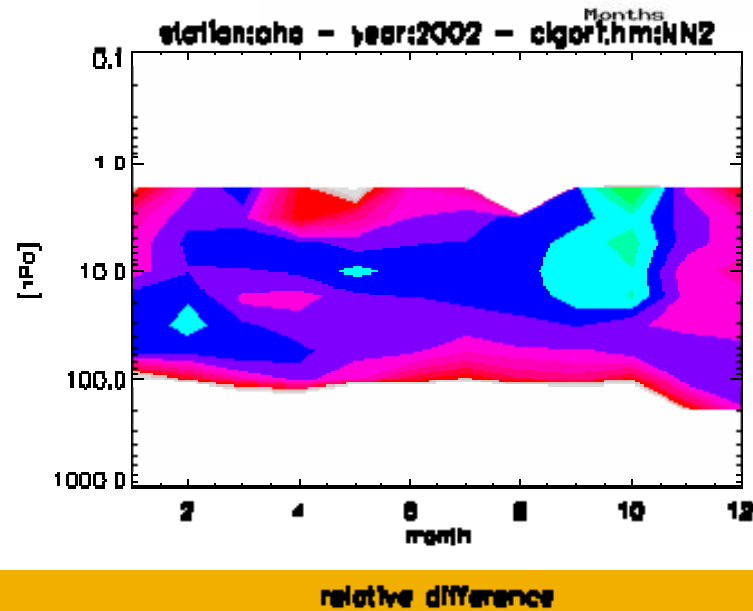
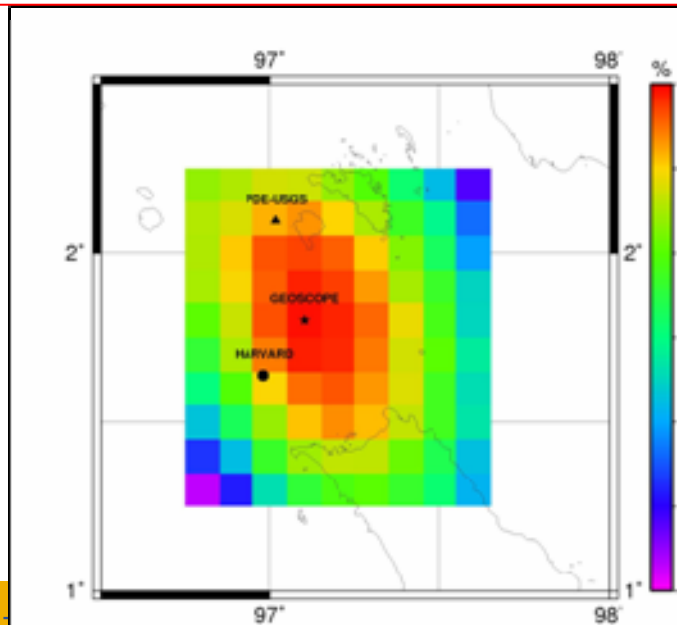
V0 Stats (ESR)

Success/(Registered-Cancelled) Jobs = 81 %

20 success jobs per day (average)



Earthquakes epicenter determination



Ozone maps

gLite Services

LCG-2 Resource Broker (RB)	grid013.ct.infn.it
LCG-2 Resource Broker (RB)	skurut2.cesnet.cz
LCG-2 Resource Broker (RB)	rb.be.itu.edu.tr
gLite Resource Broker (RB)	glite-rb.ct.infn.it
gLite Resource Broker (RB)	glite-rb2.ct.infn.it
Information Index (BDII)	grid004.ct.infn.it
Backup Information Index (BDII)	grid013.ct.infn.it
GILDA VO server	grid-vo.cnaf.infn.it:10389
GILDA gLite VOMS server	cert-voms-01.cnaf.infn.it:15008
GILDA gLite DGAS Price Authority	grid-demo1.ct.infn.it:56568
GILDA gLite DGAS HLR	grid-demo1.ct.infn.it:56567
GridICE Monitoring System	alifarm7.ct.infn.it:50080
LCG-2 Replica Location Service (RLS)	grid008.ct.infn.it
LCG-2 File Catalog (LFC)	lfc-gilda.ct.infn.it
Backup LCG-2 File Catalog (LFC)	lfc-gilda.cern.ch
gLite FiReMan Catalog	grid017.ct.infn.it
gLite File Transfer/Placement Service	fts.ct.infn.it:8443
gLite R-GMA server	rgmasrv.ct.infn.it:8443
MyProxy Server	grid001.ct.infn.it

- **Demonstrated scientific benefits of the EGEE grid**
- **Capture of requirements and assessment of middleware**
- **Improved process to select new application areas**
- **Migration to gLite**

- **Demonstrated added value in terms of deployment scale**
 - Heavy production of simulated data for LHC Computing Grid
 - Essential for the commissioning of the experiments
- **Demonstrated added value in terms of response time**
 - Computation of seism epicenter
 - Search for new drugs: large scale in silico docking (WISDOM)
- **Demonstrated added value in terms of data distribution**
 - Distribution of LHC data on the grid and sizeable users activity (~150k jobs/month in the CMS case)
- **Coming:**
 - medical data management

- **Status at PM18**
 - A database of requirements has been setup by the Project Technical Forum
 - This database includes already >400 requirements
 - This database is populated by the scientific communities deploying applications on EGEE and by FP6 projects (eg Diligent, Seegrid,...)

- **capture of requirements from FP5 projects**
 - Several FP5 projects were granted access to the database (Grace Mammogrid)
 - But it generally turns out to be difficult to collect explicitly requirements
 - No formalization of requirements
 - Projects finished
 - Requirements are mostly collected through the scientific communities (Crossgrid, Datagrid)

- **assessment of middleware w.r.t requirements**
 - On-going work with JRA1 within the framework of PTF and TCG

- **Situation at PM18**
 - No new call for applications since EGEE review
 - New application communities (finance, archeology, fusion) asked for access to infrastructure
 - EC funded projects are also asking for collaboration with EGEE
- **Improved process under design**
 - Creation of lightweight recognized VOs vs supported VOs
 - Only supported VOs would be asked to write a Memorandum of Understanding
 - Decentralized integration vs centralized EGAAP like approach
 - Initial deployment at regional level
- **Improved Memorandum of Understanding between EGEE and the scientific communities**

Actions	Comp Chem	Magic	Planck	Drug Discovery	Egeode	ESR
MoU filled by applications	OK	OK	OK	OK	Waiting feedback	OK
SA1 questionnaire	OK	OK	OK	OK	OK	OK
SA1	OK	OK	OK	OK	-	OK
SA2	OK	OK	OK	OK	-	OK
NA3	OK	OK	OK	OK	-	OK

- **Status at PM18**

- gLite offers unprecedented functionalities for user communities
- The migration of several existing applications to gLite is achieved (see demos)
 - ARDA prototypes for HEP
 - Several biomed applications (Pharmacokinetics, CDSS, ...)
- gLite available for early deployment on GILDA
- gLite brought to 100s during EGEE tutorials with GILDA

- **Issues**

- The availability of gLite software on the pre-production and production services

- **Dates: March 1-3 2006**
- **Location: CERN, Switzerland**
- **Target attendance: 150 participants**
- **Goals**
 - Get a consistent understanding across the EGEE related projects of expectation, present status and possible evolution
 - Promote cross-application fertilisation
 - Prepare EGEE-II
- **Participation open to external projects and EGEE members**
- **Format: 3-day workshop**
 - Presentations by thematic areas selected by invitation and through a call for contributions
 - EGEE presentations (integration of new applications, access to resources, status of middleware,...)
 - With a lot of time for discussion
- **URL:**
 - <http://egee-intranet.web.cern.ch/egee-intranet/User-Forum/index.html>

- Ongoing support to 2 pilot areas (High Energy Physics, Biomedicine) with established user communities and large scale applications
- On-going support to GILDA, the virtual grid laboratory
- Increased support to 4 scientific disciplines (Astrophysics, Computational Chemistry, Earth Sciences, Fusion), 3 of them coming out of EGEE generic applications
- Increased effort to foster a cohesive application community at NA4 management level
- Increased complexity from the wider regional distribution and the associated projects

Task	Goal	Resource allocation in PM	Fraction
TNA4.1	Support teams to help with consulting and porting of applications	1005	51%
TNA4.2	Evolution of pilot applications with evolving middleware services	215	11%
TNA4.3	Validation of the efficient production and continuous availability of data for the scientific communities	282	14%
TNA4.4	Grid laboratory to attract and expose new applications to grid computing	192	10%
TNA4.5	Management of the large and diverse set of application and institutes involved in NA4	269	14%

- **Achievements**

- The number of users in VOs related to NA4 activity has doubled
- The usage of the grid by pilot applications has significantly evolved
- The migration of several existing applications to gLite is achieved

- **Issues**

- The availability of gLite software on the pre-production and production services;
- The amount of effort required by the VOs in order to obtain useful results from the use of the grid
- The availability, accessibility and quality of user documentation

- **Perspectives: User Forum, EGEE-II**