egee

GridICE:
a Monitoring Tool for Grid Systems

Recent Evolution, Use Cases and Interoperability.

Grid Monitoring Workshop

Monterey Bay, California, 25 June 2007

Antonio Pierro

INFN-BARI (Italy)

Antonio.pierro <at> ba.infn.it

www.eu-egee.org

Information Society
and Media

Enabling Grids for E-sciencE

1. **Recent evolution of GridICE**
   - New lightweight sensors
   - Handling the VOMS information
   - New Lemon release integration

2. **Interoperability of the monitoring tools**
   - Integration with local monitoring systems (LEMON)
   - Standard interface for publishing monitoring data
   - Data Exchange Standard
   - Grid Monitoring Probes Specification

3. **Grid monitoring from the different users' perspectives**
   - Normal users/ VO manager/Site Manager point of view
   - Details of the VOMS groups, roles and users.
   - Troubleshooting

4. **Reliability of the information**
   - First test period: Jan-Feb-Mar 2007
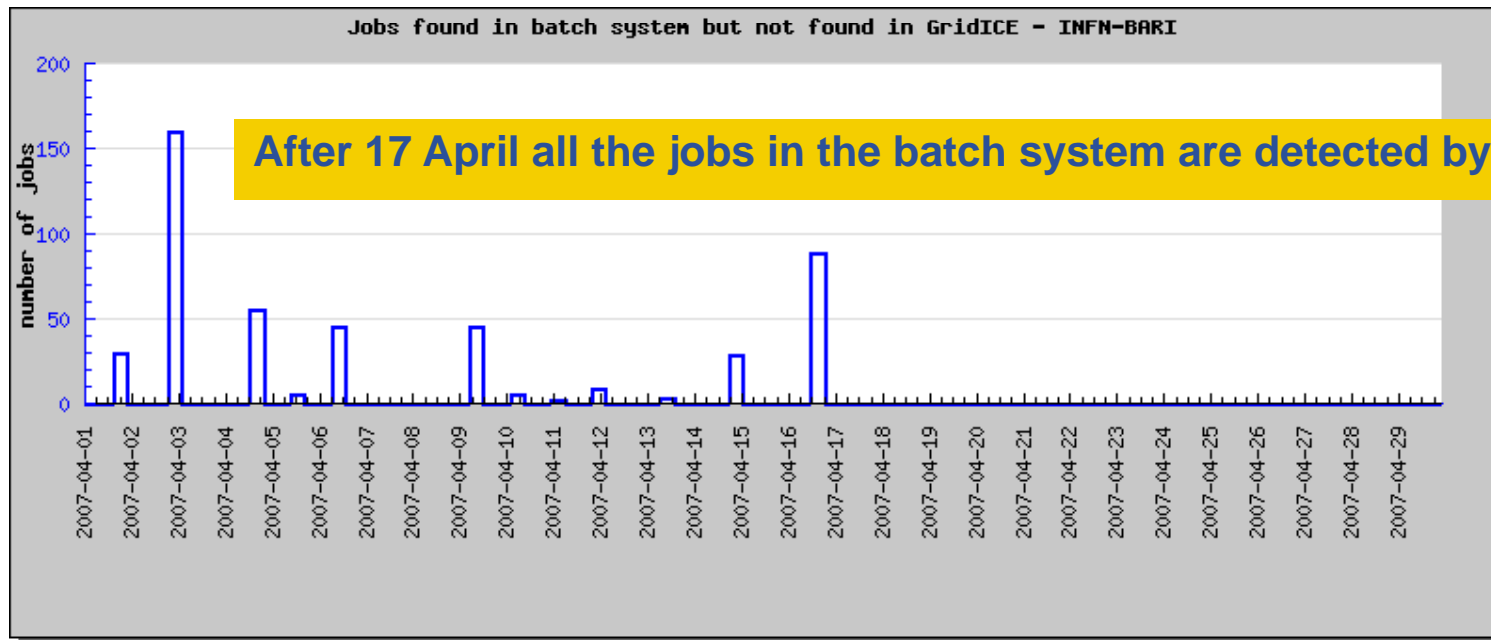   - Second test period (last sensors release): 1-14 June 2007

Enabling Grids for E-sciencE

- **Attributes measured by the Job Monitoring sensor**

Field	Description
NAME	Job name
JOB_ID	Local LRMS job id
GRID_ID	Grid job unique id
USER	Local mapped username
VO	User VO name
QUEUE	Queue
QTIME	Job creation time
START	Job start time
END	Job end time
STATUS	Job status
CPUTIME	Job CPU usage time
WALLTIME	Job walltime
MEMORY	Job memory usage
VMEMORY	Job virtual memory usage
EXEC_HOST	Execution host (WN)
EXIT_STATUS	Batch system exit status
SUBJECT	User DN
VOMS_ROLES	User VOMS roles

To reduce its intrusiveness in terms of **resources consumption**:

- Two daemons running and a probe executed periodically

- They listen to a set of log files and collect the relevant information

- Few LRMS commands to retrieve jobs status

- The status of all jobs is stored in a cache (stateful behaviour)

- Substantial effort to integrate the new version (v.2.13.x)
- Revision of the whole set of GridICE-specific sensors to comply with the new version of LEMON: the most of the sensor were substituted with the Lemon ones (to reduce the load on machines)
- Significant rewriting of the fmon2glue
- Testing at INFN-BARI since February 2007

Jobs found in batch system but not found in GridICE - INFN-BARI

After 17 April all the jobs in the batch system are detected by GridICE

Enabling Grids for E-sciencE

- Grid monitoring **integrated with local monitoring**
- The last server version is very simple to install
  - The client installation may be turned on in the **standard middleware LCG installation** (no additional operation are needed)
- The LEMON monitoring system and alarm management are integrated in the new version of the GridICE server
- The local sensor currently used for farm monitoring can be interfaced with GridICE to collect all the available data
- The back-end is realized with LEMON
  - Local farm monitoring that are using LEMON can be integrated with GridICE
  - Possible integration of data collected from GridICE with Lemon RRD framework (LRF) - (very similar with Ganglia, for those familiar with the tool) [5]

- We have chosen an approach already in use in current Grid systems: the **GLUE Schema** [3].

- This schema is the **result of** a joint **collaboration between large European and American Grid projects**.

- **It includes** a conceptual schema modeled in the form of **UML class diagrams and** mappings to specific technologies such as **LDAP** (Lightweight Directory Access Protocol) [4], XML, and relational data models.

- **The most part of the metrics defined in the "Standard GLUE Schema"** (used by gLite as Information System) **are** currently **measured by GridICE**.

- Moreover, a rich set of metrics related to a **computer system** has been defined

Enabling Grids for E-sciencE

- Gridice team is **working to follow the standard suggested by LCG Monitornig working group**.
    - https://twiki.cern.ch/twiki/bin/view/LCG/GridMonitoringDataExchangeStandard
- Conventions
    - **Passing** list in **URL** (graph, xml file, web page)
- URL **Data types**
    - scalar values- (examples: width=570&height=340, days=3)
    - string - quite intuitive (examples: farmName,VOName)
    - number - optionally two sub-types: int, float
    - boolean – they are represented as string "true" or "false"
- Example
    - http://gridice3.cnaf.infn.it:50080/gridice/chart/statsResUsage-rocRep-API.php?farm_name=INFN-BARI&vo_name=ALL&shift_time=5%20Days&date_stop=June%208,%202007&width=570&height=340

Enabling Grids for E-sciencE

- each view of the web-based interface **offers the same data in XML format**
  - You can retrieve data in XML format file that can be used to **exchange information** among **different monitoring tool analysis**
  - Attributes in XML file are well commented and self-explaining.

```
– <SiteInfo Created="June 13 2007 18:35:02" Expire="2">
   – <CEList>
      – <CE>
           <CEUniqueID>gridba2.ba.infn.it:2119/jobmanager-lcgpbs-cert</CEUniqueID>
           <Site>INFN-BARI</Site>
           <Status>1</Status>
           <FreeSlots>9</FreeSlots>
           <RunningJobs>0</RunningJobs>
           <WaitingJobs>4444</WaitingJobs>
           <TotalSlots>80</TotalSlots>
```

- You can retrieve data from database through a store procedure function

- using W3C standards to offer easy access to monitoring data

- Concerning the Grid Monitoring Probes Specification, Gridice is already working to follow the standard suggested by LCG Monitornig working group.
  - https://twiki.cern.ch/twiki/bin/view/LCG/GridMonitoringProbeSpecification

- We're going to change our sensors in order to have two output format:
  - Output format 1: used by our tool for **our servers**
  - Output format 2: it will be used by other monitoring complying with specification suggested by LCG Monitornig working group.

- **The Grid involves a huge number of worldwide distributed resources**

- **Monitoring of those distributed resources is a vital determinant for the whole system**

- **Different actors require different views of monitoring information:**

  - **Virtual Organization managers** require the ability of observing and analyzing the **performance of** the "actual " **system they are using** (this can dynamically change over time)

  - Both **site administrators** and **grid operation center** managers require performance analysis and fault detection of the resources for which they are responsible

  - **Grid Service developers** require the ability of analyzing the **behaviour of their applications** (e.g.,how does a resource broker dispatch jobs over a set of available resources)

- The users are identified with the digital certificate installed in its browser
  - *a valid CA certificate*
  - *server based on https protocol*

- The new sensor are able to retrieve the VOMS information
  - *VOMS information: groups and roles of users submitting the jobs*

- Users not registred in GridICE DB (because they don't submit jobs)
  - *The related role (e.g., site manager, vo manager) can be retrieved by the GOC database.*

Finding information about personal jobs by means of personal browser certificate

Finding information about exit status of personal jobs
by means of personal browser certificate

VO: ALL   Site: INFN-BARI   exit status jobs   startTime: 03 Jun 2007 00:00
                                                       endTime: 05 Jun 2007 00:00

User:

55%

Exit status Occurrences
- 0   162
- 271   129
- -200   4
- Total: 295

1%

44%
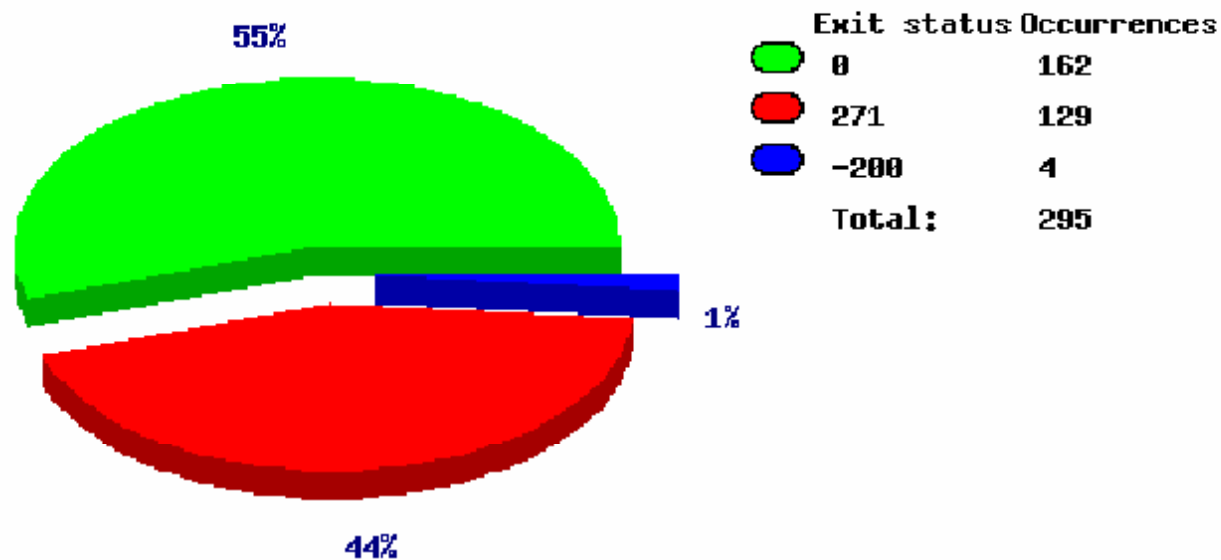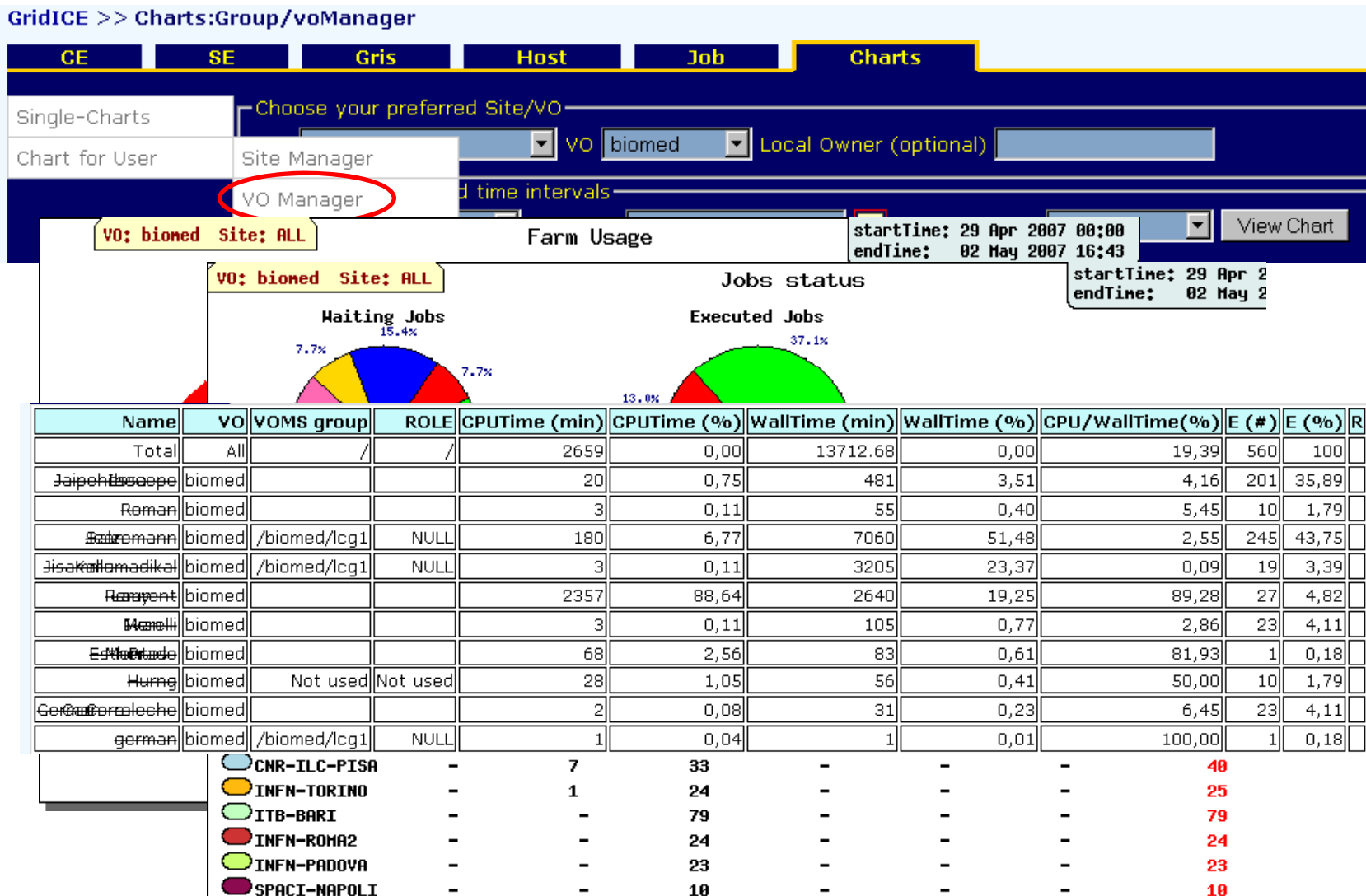
GridICE >> Charts:Group/voManager

| CE | SE | Gris | Host | Job | Charts |

Single-Charts

Chart for User

Site Manager

VO Manager

Choose your preferred Site/VO

VO biomed Local Owner (optional)

d time intervals

VO: ALL Site: ITB-BARI   VO Usage   startTime: 29 Apr 2007 00:00 endTime: 02 May 2007 18:33

VO: ALL Site: ITB-BARI   Jobs status   startTime: 29 Apr 200 endTime: 02 May 200

Waiting Jobs 50.0% Executed Jobs 52.0%

Name	VO	VOMS group	ROLE	CPUTime (min)	CPUTime (%)	WallTime (min)	WallTime (%)	CPU/WallTime(%)	E (#)	E (%)	R	Q
Total	All	/	/	18149.75	0,11	59436.63	0,37	30,54	628	100	46	1248
Mieba	cms	/cms	NULL	19	0,10	26417	44,45	0,07	133	21,18	16	95
Filippo	cms	/cms/Commissioning/Tracker/cmsTAC	production	9308	51,28	12518	21,06	74,36	109	17,36	20	566
PatMetaberze	alice	Not used	Not used	5566	30,67	10458	17,60	53,22	26	4,14	9	22
Andreaba	cms	/cms	NULL	3063	16,88	4247	7,15	72,12	68	10,83		99
elisabeth	bio	/biomed/lcg1	NULL	1	0,01	1183	1,99	0,08	3	0,48	1	
Nadrok	dteam	/dteam/cern	NULL	28	0,15	1105	1,86	2,53	66	10,51		2
Quadri	atlas	/atlas/lcg1	production	2	0,01	783	1,32	0,26	3	0,48		142
DieLrkel	atlas	/atlas/lcg1	NULL	6	0,03	472	0,79	1,27	44	7,01		77
DaRothotte	atlas	/atlas/lcg1	production	2	0,01	389	0,65	0,51	3	0,48		229
Nipetryk	atlas	/ops	NULL	2	0,01	105	0,18	1,90	5	0,80		4
TWmaus	atlas	Not used	Not used	1	0,01	45	0,08	2,22	7	1,11		6
JaipehEsoaepe	biomed	Not used	Not used	1	0,01	35	0,06	2,86	2	0,32		1
stefanforte	cms	/cms	NULL	86	0,47	942	1,58	9,13	105	16,72		
Drulager	cms	/cms/dcms	NULL	59	0,33	306	0,51	19,28	22	3,50		

		-	-	1	-	-	-	-	1	
cms		-	-	1	-	-	-	-	1	
Total:		**0**	**2**	**196**	**0**	**0**	**0**	**198**		

– **Site Admins** can easily use GridICE for both **detect fault situations** related to the own resources (es. is there any grid services down?) and control how the own resources are used and appear to the Grid (es. how many jobs are running or waiting in my site?).

General Site view quickly gives you hints on **how is geographically composed your Grid**, it also tells you what are going on your Grid in terms of **Job and CPU load percentage**. To detect problems on your site you can select it and navigate Gris, Host, Job and Charts tabs contextualized for your grid site resources.[2]

GridICE >> Site::ALL

General

Site ▼		Region	GK#	Q#	RunJob	WaitJob	JobLoad	Power	WN#	CPU#	CPULoad	Available	Total	%	MH#
CERN-PROD	🇨🇭	CERN	2	32	5078	4351	-	-	-	-	-	647.3 GB	1.7 TB	64%	-
CNR-ILC-PISA	●	Italy	1	6	1	2	5%	3K	2	4	28%	753.9 GB	763.2 GB	1%	4
CREAM-PADOVA	○	Italy	1	1	50	8991	-	-	-	-	-	-	-		
ENEA-INFO	●	Italy	2	6	0	0	-	-	-	-	-	43.2 GB	54.7 GB	21%	1
ESA-ESRIN	●	Italy	1	4	0	0	-	-	-	-	-	870.3 GB	870.3 GB	0%	7

IN2P3-CC
INAF-TRIESTE

Through GridICE the **site manager** has a view of the average load of the WNs at the site. One of the load parameters is CPULoad. **If CPULoad is very high,** e.g. the value is stable around 100%, **it is good practice to look at the queue status and check job distribution**.

GridICE >> Site::ALL >> Site::ENEA-INFO

	CE	SE		Gris	Host	Job	Charts	
Hostname		🌐	Site ▼	Domain	Middleware	Type	LastCheck	Conn	Since	Entries	Scheduling

Hostname			Site ▼	Domain	Middleware	Type	LastCheck	Conn	Since	Entries	Scheduling
egseaix.frascati.enea.it	🇮🇹		ENEA-INFO	frascati.enea.it	GLITE-3_0_0	SE	0h19m27s	✓	2007-06-02 02:46	12	0
egse.frascati.enea.it	🇮🇹		ENEA-INFO	frascati.enea.it	GLITE-3_0_0	SE	0h19m26s	✓	2007-05-23 14:15	12	0
egseaix.frascati.enea.it	🇮🇹		ENEA-INFO	frascati.enea.it	GLITE-3_0_0	EX	0h17m17s	✓	2007-06-02 02:39	14	0
egse.frascati.enea.it	🇮🇹		ENEA-INFO	frascati.enea.it	GLITE-3_0_0	EX	0h17m18s	✗	2007-05-29 12:36	1	0
egce.frascati.enea.it	🇮🇹		ENEA-INFO	frascati.enea.it	GLITE-3_0_0	CE	0h4m46s	✓	2007-06-01 18:01	29	0
egceaix.frascati.enea.it	🇮🇹		ENEA-INFO	frascati.enea.it	GLITE-3_0_0	CE	0h4m46s	✓	2007-06-02 16:29	29	0

Observing the Gris tab a site admin can have **information on local Grid Information System status**. An intuitive icon will advice you what is the last result for a set of ldap queries to your gris types, take also a look to **the number of ldap "Entries" (DNs) for each gris**. In a normal situation CE, SE and SB GRIS's publish entries in the order of 10, while the EX GRIS publish entries in the order of 100.

The possibility to detect same grid service down on your site can be found in the Host tabs...
Intuitive progress bars will also tell you **statistic information on CPU/RAM usage**
to check the actual load for your machines.
To spot same details on the type of grid process stopped, simply select the hostname...

GridICE >> Site::ALL >> Site::INFN-T1

CE	SE	Gris	Host	Job	Charts	XM

View 1 View 2 View 3

Hostname	Site	Role ▼	Procs	Load15Min	CPU Usage	RAM Free	RAM Usage	Virtual Free	Virtual Usage	Last Check
ce01-lcg.cr.cnaf.infn.it	INFN-T1	CE	STOP	0.14	1%	0.6 GB	68%	8 GB	0%	0h2m19s
ce02-lcg.cr.cnaf.infn.it	INFN-T1	CE	STOP	0.19	9%	228 MB	94%	8 GB	0%	0h2m19s
ce03-lcg.cr.cnaf.infn.it	INFN-T1	CE	STOP	0	1%	132 MB	97%	8 GB	0%	0h2m19s
ce04-lcg.cr.cnaf.infn.it	INFN-T1	CE	✓	0.53	8%	227 MB	89%	8 GB	0%	0h2m19s

You will see what kind of grid middleware related process is stopped.
In this case glite-dgas-ceServerd-had process is not running .

Processes	Role	Proc Name	Status	Inst#	First	Last	CPU1Max	CPUAll	Mem1Max	MemAvg	Time1Max	TimeAll
	ce-access-node	edg-gatekeeper	S	1	8-18:02	8-18:02	0	0	0	0	0-00:00	0-00:00
	ce-access-node	fmon-agent	S	1	8-18:02	8-18:02	0	0	0	0	0-00:01	0-00:01
	ce-access-node	glite-dgas-ceServerd-had	STOP	0	0-00:00	0-00:00	0	0	0	0	0-00:00	0-00:00

Grid site admin can have a lot of benefit using this tool in concert with **SFTs**;
giving an example, if in **your site SFTs jobs fail in job list match**
 you can investigate on GridICE Gris
related tab view about possible reasons (es. bind error, empty ldap query responce),
or simply verify that **grid services related to the information system**
**are running properly** looking in the Host related tab.

Enabling Grids for E-sciencE

Period: Jan-Feb-Mar 2007

1. **We retrieve info from batch-system PBS/Torque and LSF**

2. **The info retrieved are stored in a MySQL DB**

3. **The batch-system logs have been compared to info retrieved by GridICE sensors**

4. **The comparison has been performed job per job**

5. **In the bigger sites, we have also performed some test on integral data (INFN-T1). The Reliability is about 90%.**

a) **The main reasons of fault are:**

    a) Daemons crashes

    b) Transfer buffer limit exceeded (500KB)

Enabling Grids for E-sciencE

Period: Jan-Feb-Mar 2007

FARM-NAME	Period	BatchSystem	num di jobs not found in GRICE	total	diff %
INFN-BARI	20 gen 07 -> 03 mar 07	PBS	871	19623	4.44
INFN-ITB	20 gen 07 -> 03 mar 07	PBS	4	1086	0.37
INFN-PADOVA	20 gen 07 -> 03 mar 07	LSF	1142	11169	10.22
INFN-PISA	20 gen 07 -> 03 mar 07	PBS	315	15142	2.08
INFN-LNL-2	20 gen 07 -> 03 mar 07	LSF	923	33641	2.74

egee
Enabling Grids for E-sciencE

For the bigger sites,

we have also performed some test on integral data (INFN-T1).

The Reliability is about 95% when there isn't buffer limit problem.

Jobs running on INFN-T1

Buffer limit problem 500KB

— GridICE
— LSF
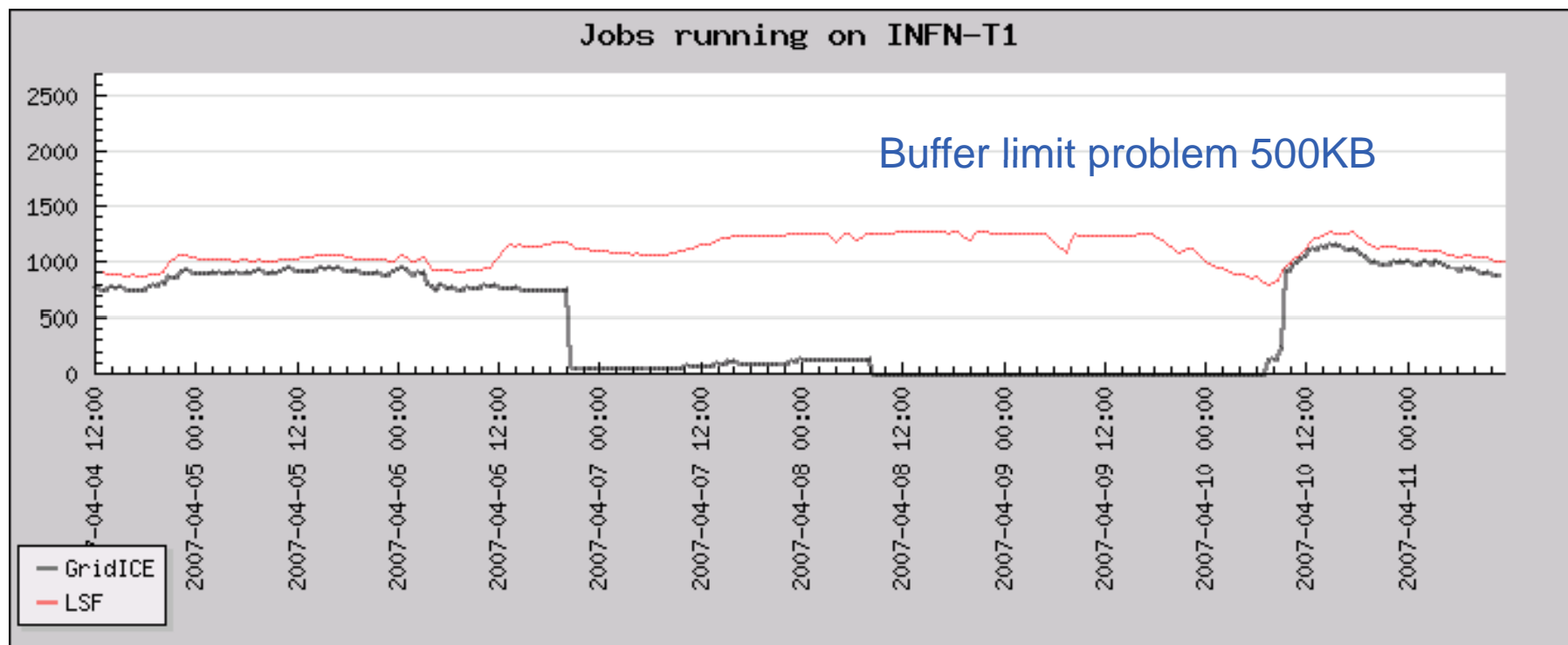
Period: 1-15 June

FARM-NAME	periodo esaminato	BatchSystem	num di jobs not found in GRICE	total	diff %	improvement
INFN-PADOVA	1 June 07 ->7 June 07	LSF	243	4789	5	5.22
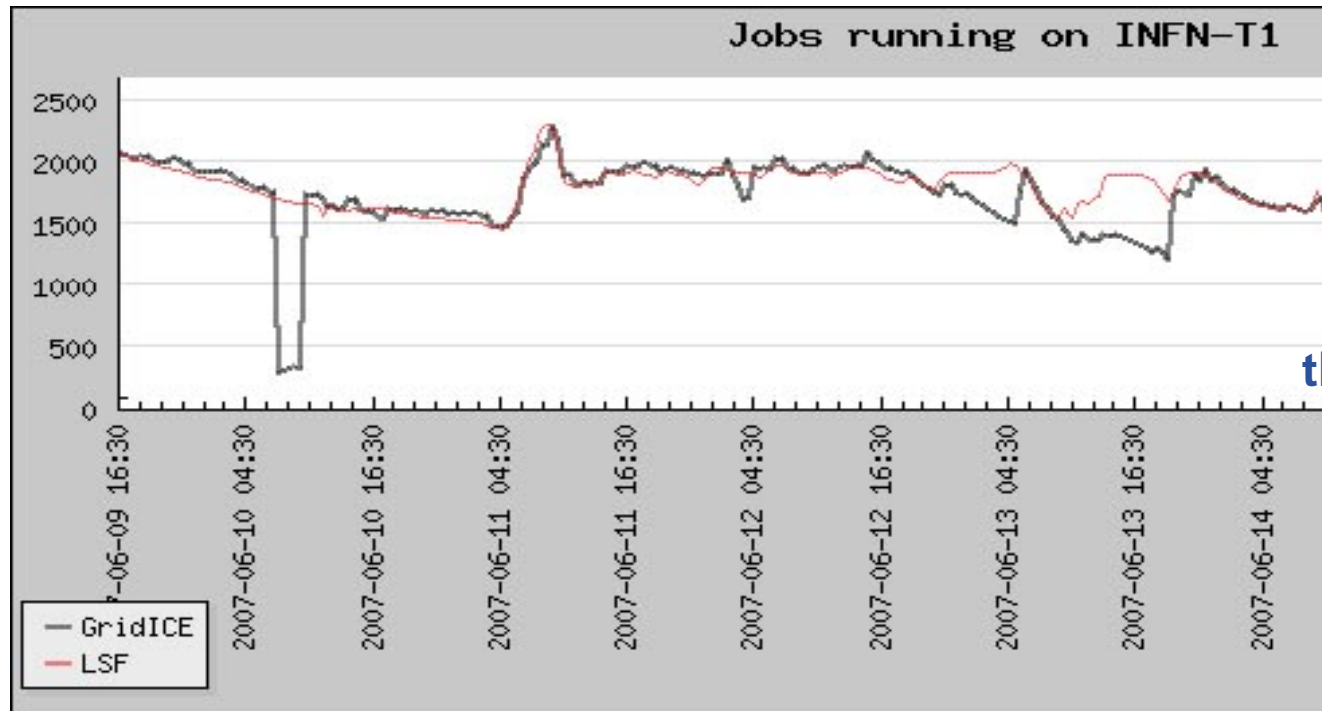
The Reliability is about 99% when there isn't buffer limit problem.

Enabling Grids for E-sciencE

- **To improve reliability and performance**
- **To increase interoperability with other monitoring tools**
  - Data Exchange Standard
  - Grid Monitoring Probes Specification
- **To implement a notification system**

…

- **We are open to collect and work on any new requirements your monitoring needs**

Enabling Grids for E-sciencE

- **[1] Recent Evolutions of GridICE: a Monitoring Tool for Grid Systems**

- **[2]G.Misurelli https://grid.ct.infn.it/twiki/bin/view/GILDA/SiteMonitoringTools**

- **[3] S. Andreozzi, M. Sgaravatto, and C. Vistoli. Sharing a Conceptual Model of Grid Resources and Services. In Proceedings of the Conference on Computing in High Energy and Nuclear Physics (CHEP 2003), La Jolla, CA, USA, Mar 2003.**

- **[4] S. Andreozzi. GLUE Schema Implementation for the LDAP Model, Technical Report, INFN, May 2003. http://www.cnaf.infn.it/ » sergio/publications/Glue4LDAP.pdf.**

- **[5] S. Andreozzi, N. De Bortoli b , S. Fantinel c ,A. Ghiselli a , G. Rubini a , G. Tortone b ,C. Vistoli GridICE: a Monitoring Service for Grid Systems**