

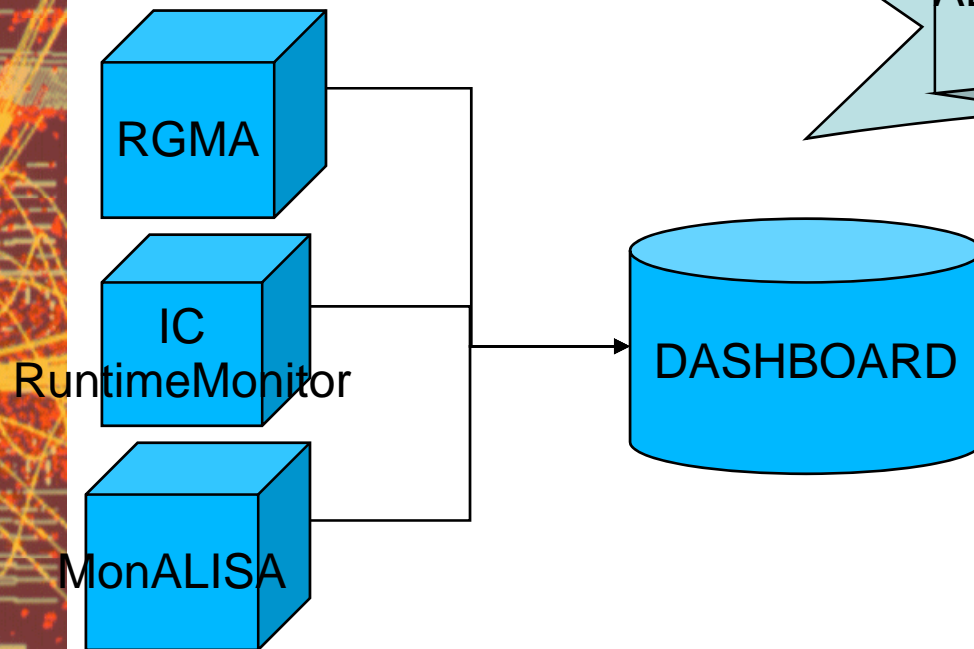
Grid Reliability

Pablo Saiz

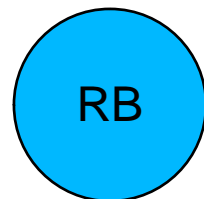
On behalf of the Dashboard team:
J. Andreeva, C. Cirstoiu, B. Gaidioz,
J. Herrala, E.J. Maguire, G. Maier,
R. Rocha, P. Saiz

- What is Grid reliability?
- How do we do it?
- What do we do?
 - Data Management
 - Workload Management
- To Do list
- Conclusions
- Useful links

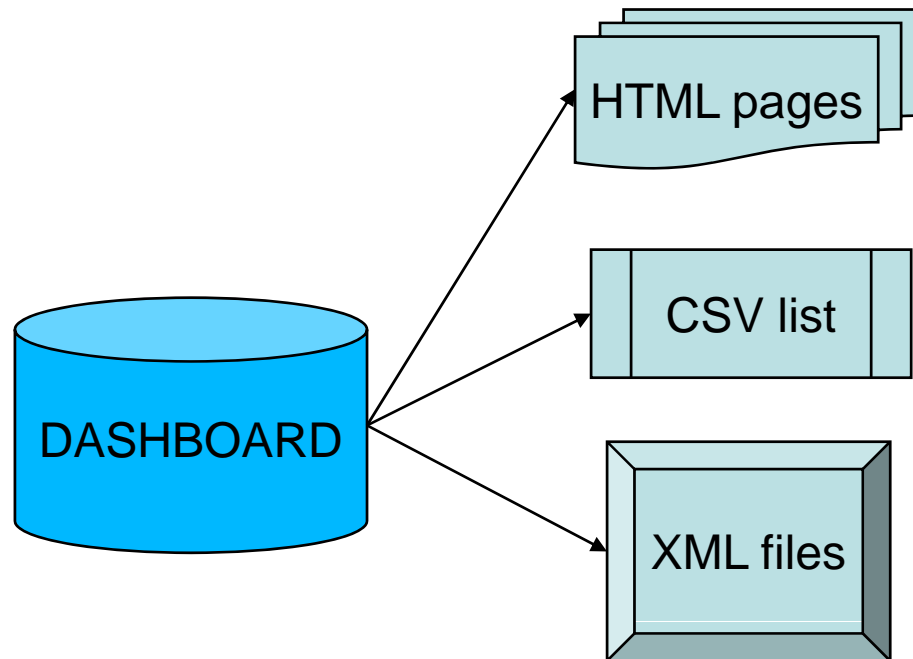
- Our goal:
 - Provide tools to detect, investigate and solve all the possible Grid errors.
- How:
 - Using the experiment's dashboards
 - Monitoring the user jobs
- Deliverables:
 - Efficiency tables
 - Site performances as seen by selected applications
 - Tools to monitor the sites “day-by-day” and augment the available information for more efficient debugging



Already deployed for
ALICE, ATLAS, LHCb and CMS
VleMed in its way...



For some RB:
`edg-get-logging-info -v 2`
`edg-get-status -v 2`



We can display the same data in different formats

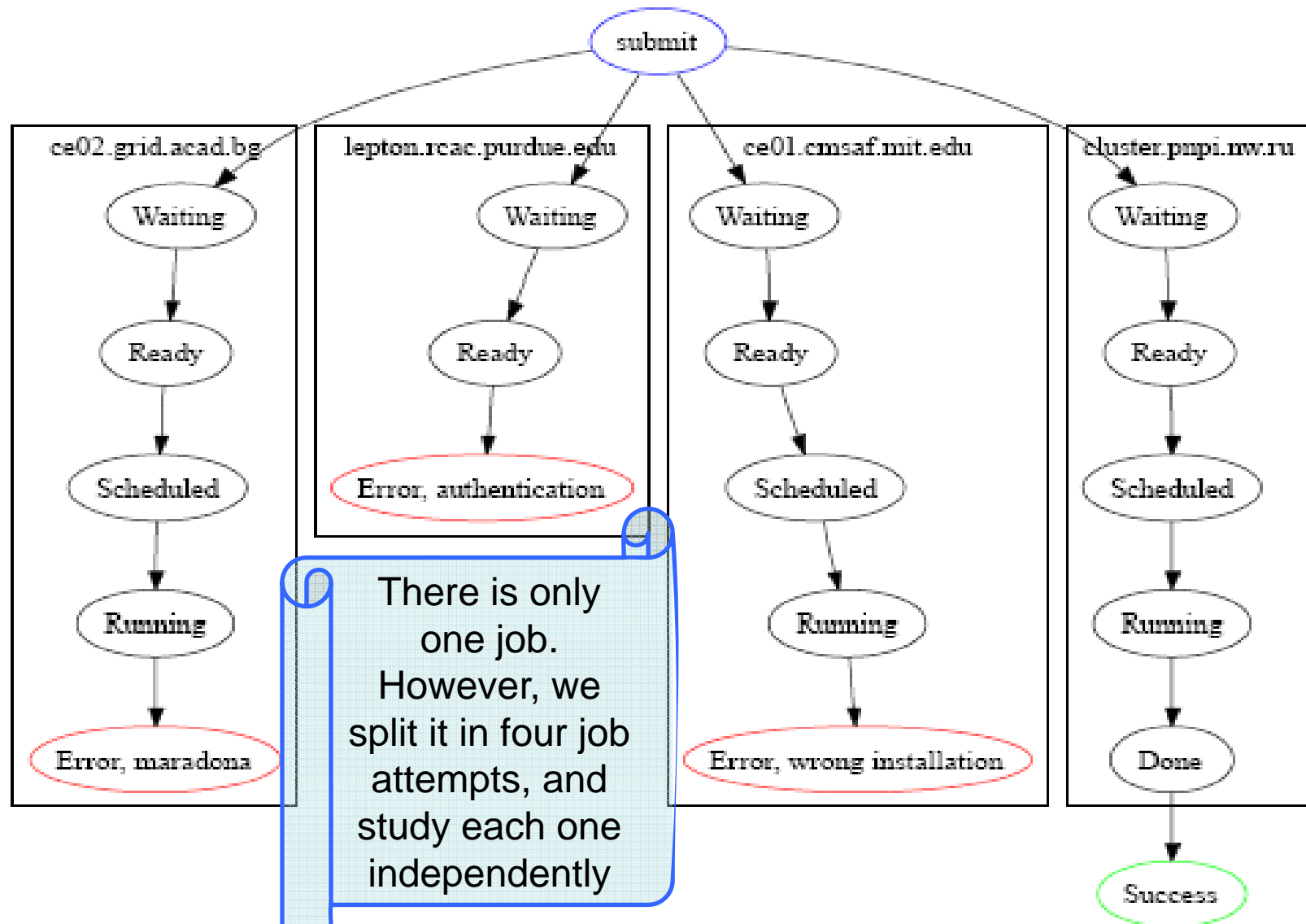
For more details, see Julia Andreeva's talk, Wednesday at 17:10:
[Grid Monitoring from the VO/
User perspective. Dashboard for the LHC experiments.](#)

- Workload management
 - Deployed for ALICE, ATLAS, CMS and LHCb
 - Monitor jobs through RGMA and Imperial College Runtime Monitor
- Data management:
 - FTS for ALICE
 - Deployed in September 2006
 - Heavily used during the service challenges
 - DDM for ATLAS

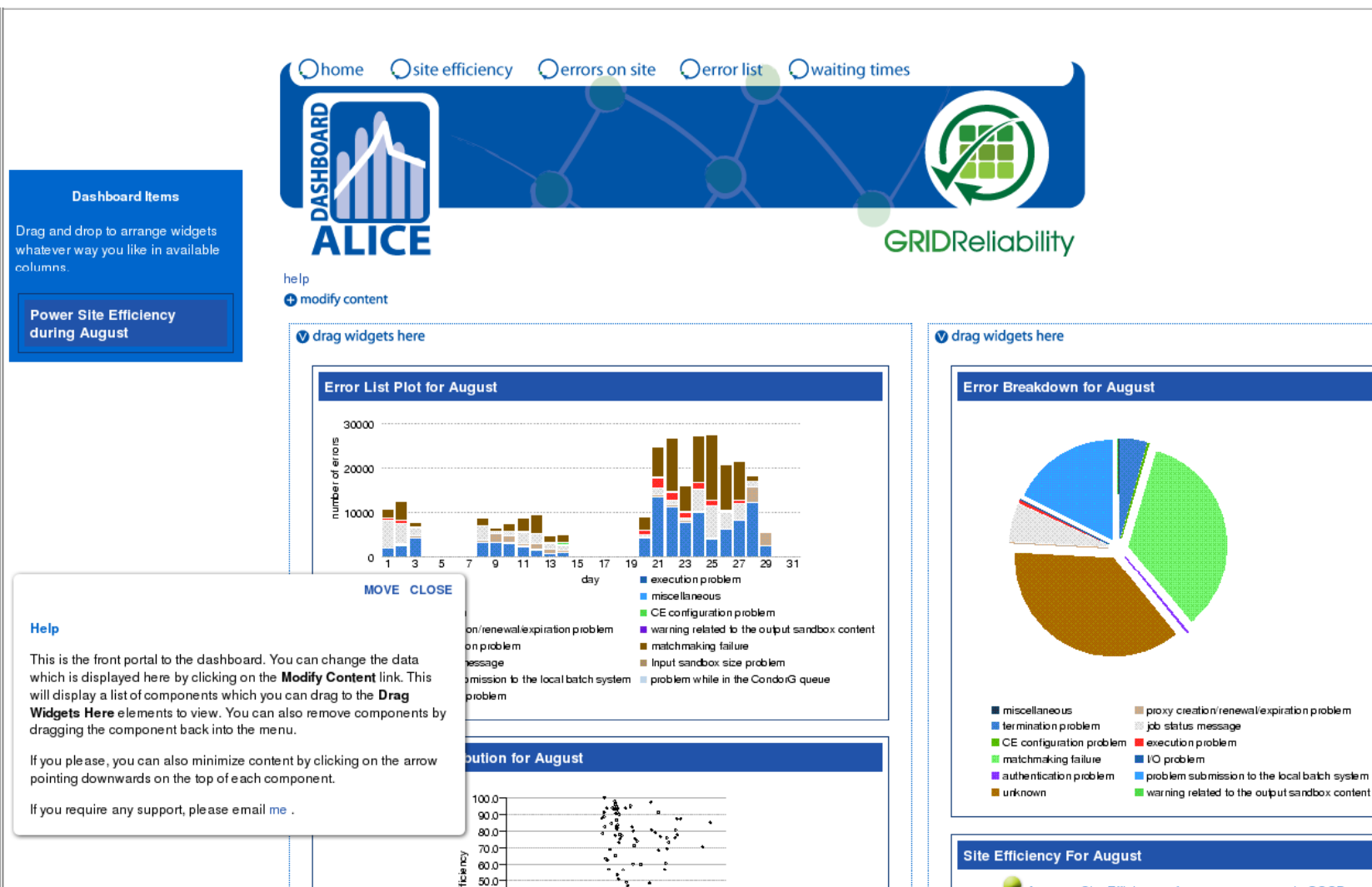
For more details, see Ricardo Rocha's talk, Thursday at 17:10:
[Monitoring the Atlas Distributed Data Management System](#)

- Looking at the final state:
 - Simple:
 - Reliability of the whole system
- Looking at all the status changes:
 - More information
 - Possibility to catch errors solved by the middleware
 - Reliability of different sites

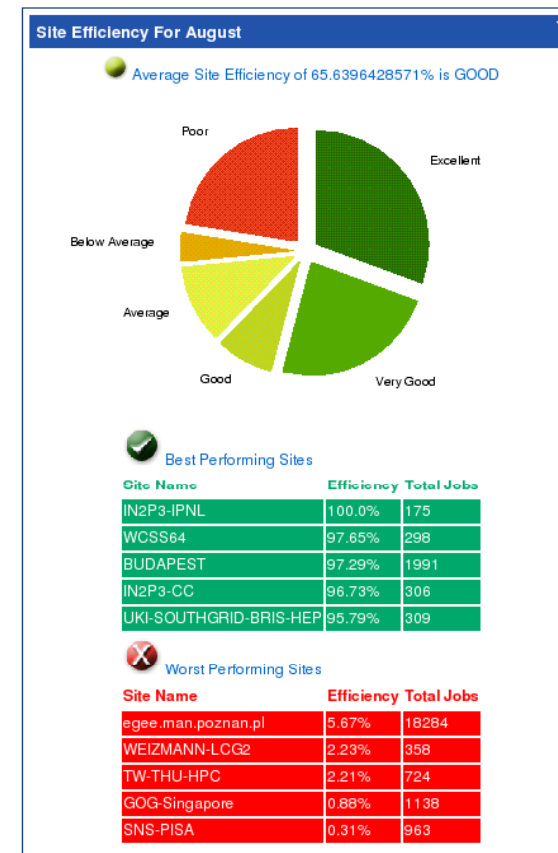
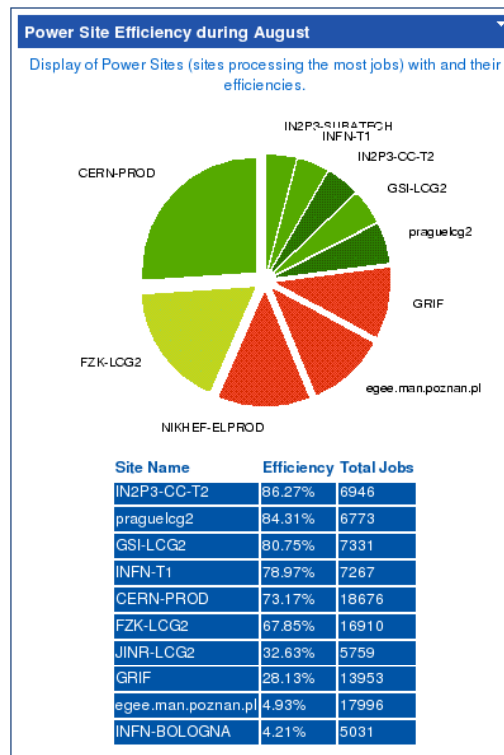
Job vs. Job Attempt



- 'Site of the day':
 - daily report on number of successful/failed job attempts
- Site performance
 - Evolution of a site over a period of time
- Error list
 - Most Common list of error messages, with pointers to documentation
 - Evolution of the error over time
- Waiting time
 - Time that users have to wait from the moment they submit the job until they get the results back
- Aggregated reports:
 - Monthly reports
 - Multi vo reports



- Ranking of the efficiency of the sites



- For each VO:

home site efficiency errors on site error list waiting times

DASHBOARD
ALICE

GRIDReliability

help
Site Efficiency details the reliability of sites and their jobs. You may elect to receive data on efficiency over a day or month and are also able to compare site efficiencies amongst ALICE, ATLAS, CMS and LHCb.

[Daily Reports]
[Monthly Reports]
[VO Comparison]



Click on any Site, and you will have a breakdown of the jobs according to the CEs

If you click on the CEs, you will have a breakdown of the jobs of that CE according to their workflow

Warning! This table does not represent the number of jobs, but the number of job attempts. For example, if a user submits one job, the job lands twice on site A where it fails, and then it lands on site B where it is successful, that job would produce three entries in the following table: two failures for site A, and one success for site B.

Displaying the sites with more than 100 jobs

Displaying the values of the date: 26-Aug-2007

If you want a similar report for any other day, click [here](#) to go back and make a query

We also provide a list of worker nodes where jobs failed. To see the list, click [here](#)

help
Reports for 2007

Aug >1 >2 >3 >4 >5 >6 >7 >8 >9 >10 >11 >12 >13 >14 >15 >16 >17 >18 >19 >20 >21 >22 >23 >24 >25 >26 >27
Jul >1 >2 >3 >4 >5 >6 >7 >8 >9 >10 >11 >12 >13 >14 >15 >16 >17 >18 >19 >20 >21 >22 >23 >24 >25 >26 >27 >28 >29 >30 >31
Jun >1 >2 >3 >4 >5 >6 >7 >8 >9 >10 >11 >12 >13 >14 >15 >16 >17 >18 >19 >20 >21 >22 >23 >24 >25 >26 >27 >28 >29 >30
May >1 >2 >3 >4 >5 >6 >7 >8 >9 >10 >11 >12 >13 >14 >15 >16 >17 >18 >19 >20 >21 >22 >23 >24 >25 >26 >27 >28 >29 >30 >31
Apr >1 >2 >3 >4 >5 >6 >7 >8 >9 >10 >11 >12 >13 >14 >15 >16 >17 >18 >19 >20 >21 >22 >23 >24 >25 >26 >27 >28 >29 >30
Mar >1 >2 >3 >4 >5 >6 >7 >8 >9 >10 >11 >12 >13 >14 >15 >16 >17 >18 >19 >20 >21 >22 >23 >24 >25 >26 >27 >28 >29 >30 >31
Feb >11 >12 >13 >14 >15 >16 >17 >18 >19 >20 >21 >22 >23 >24 >25 >26 >27 >28

Please select a date to view the site efficiency report for that day.

Day 1st Month Jan Year 2007 Submit

SiteName (click on any site)	Successful jobs	Failed jobs	Efficiency
SNS-PISA	0	101	0.00%
NIKHEF-ELPROD	1	2906	0.03%
unknown	83	253	24.70%
praguecg2	129	23	84.87%
GRIF	516	80	86.58%
SARA-MATRIX	201	21	90.54%
FZK-LCG2	3265	217	93.77%
IN2P3-SUBATECH	1075	3	99.72%
CERN-PROD	6283	17	99.73%
RAL-LCG2	325	0	100.00%

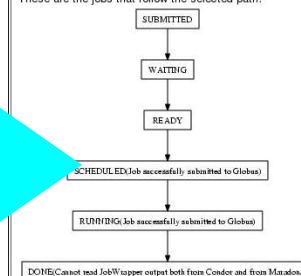
- Getting the job ids (and history) of jobs that failed.

try also provide a list of failed jobs (more jobs failed, to see the list, click here)

SiteName (click on any site)	Successful jobs	Failed jobs	Efficiency
RAL-LCG2	0	1507	0.00%
lcgc01.gridpp.rl.ac.uk:2119/jobmanager-lcgpbs-aliceL2000	0	1451	0.00%

JobIds	# jobs	Successful?	Error message
See all the jobids...	1175	Ignored	Job MaxRetryCount hit
See all the jobids...	3	Ignored	Job MaxRetryCount hit
See all the jobids...	1	Ignored	Job MaxRetryCount hit
See all the jobids...	1	Ignored	Job MaxRetryCount hit
See all the jobids...	1198	No	Cannot read JobWrapper output both from Condor and from Maradona
See all the jobids...	90	No	Job successfully submitted to Globus
See all the jobids...	73	No	unknown

These are the jobs that follow the selected path:



Displaying the values of the date: 18-Mar-07

JobId:	# resubmission	Job successful?	Timestamp	Worker Node*
https://lcg02.gridpp.rl.ac.uk:9000/-fC1Xm1D5Ok3hCIWzsexg	2	NO	2007-03-18 09:34:10	
https://lcg02.gridpp.rl.ac.uk:9000/-DIX425mOhY28qthqVfEg	2	NO	2007-03-18 20:49:37	
https://lcg02.gridpp.rl.ac.uk:9000/-HvSHBCIDGH5qE8ESi6r2Q	2	NO	2007-03-18 20:38:12	
https://lcg02.gridpp.rl.ac.uk:9000/-Hvmmpa_az1GrYVj8h0wEg	2	NO	2007-03-18 09:17:22	
https://lcg02.gridpp.rl.ac.uk:9000/-K4c8vHNgPO89rlinC1Hbw	2	NO	2007-03-18 07:43:27	
https://lcg02.gridpp.rl.ac.uk:9000/-OcqslUY8HATYIU-oMEp4g	2	NO	2007-03-18 04:34:06	
https://lcg02.gridpp.rl.ac.uk:9000/-WEBN3JcYhzLu7Q9e1MCG	2	NO	2007-03-18 22:02:47	
https://lcg02.gridpp.rl.ac.uk:9000/-ZJlh3jzHbbXJqNfa0o-pw	2	NO	2007-03-18 18:49:04	
https://lcg02.gridpp.rl.ac.uk:9000/-auxa-B3LzaGhuNx-CZLOg	2	NO	2007-03-18 07:37:18	
https://lcg02.gridpp.rl.ac.uk:9000/-byTuCyW2lp4e-6YC6RfTg	2	NO	2007-03-18 09:22:17	
https://lcg02.gridpp.rl.ac.uk:9000/-kHsmmqZ1p46Mc-eMCRqmRg	2	NO	2007-03-18 01:49:56	
https://lcg02.gridpp.rl.ac.uk:9000/-ownFN0wLsZaYACUwORKg	2	NO	2007-03-18 10:13:09	

- Reliability of a site over a period of time

Monthly Reports

Choose the site and the month from which you want the report:

CERN-PROD January 2007 Create report

AUVERGRID
 BG04-ACAD
 BUDAPEST
 CERN-PROD
 CGG-LCG2
 CNR-ILC-PISA
 CY-01-KIMON
 CYFRONET-IA64
 CYFRONET-LCG2
 EFDA-JET
 FMPhi-UNIBA
 FZK-LCG2
 GOG-Singapore
 GR-01-AUTH
 GR-03-HEPNTUA
 GR-04-FORTH-ICS
 GR-05-DEMOKRITOS
 GR-06-IASA
 GRIF
 GSI-LCG2

CERN-PROD DETAILS

Click on any cell, and you will have a breakdown of the jobs according to the CEs

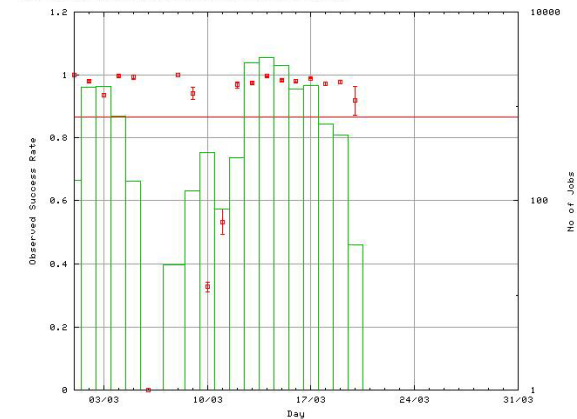
If you click on the CEs, you will have a breakdown of the jobs of that CE according to their workflow

Warning! This table does not represent the number of jobs, but the number of job attempts. For example, if a user submits one job, the job lands twice on site A where it fails, and then it lands on site B where it is successful, that would produce three entries in the following table: two failures for site A, and one success for site B. If you want a similar report for any other day, click [here](#)

We also provide a list of worker nodes where jobs failed. To see the list, click [here](#)

Date (click on any row)	Successful jobs	Failed jobs	Efficiency
20-MAR-2007	23	0	79.31%
19-MAR-2007	204	206	57.96%
19-MAR-2007	530	93	86.62%
17-MAR-2007	517	149	77.63%
16-MAR-2007	544	223	70.93%
15-MAR-2007	328	431	43.21%
14-MAR-2007	1098	785	58.31%
13-MAR-2007	591	331	64.10%
12-MAR-2007	235	0	100.00%
11-MAR-2007	80	59	57.56%
10-MAR-2007	293	164	38.71%
09-MAR-2007	132	2	99.58%
08-MAR-2007	14	0	100.00%
06-MAR-2007	1	1	0.00%
05-MAR-2007	133	0	100.00%
04-MAR-2007	99	0	100.00%
03-MAR-2007	1226	35	97.22%
02-MAR-2007	1144	13	98.86%
01-MAR-2007	0	0	100.00%

The plot below presents the efficiency of the site. The green boxes are the number of successful jobs (please note that we put jobs and not job attempts). It is in logarithmic scale up to 10,000 (see the right side of the plot). The red dots are the efficiency for each day. The red bar is the average efficiency



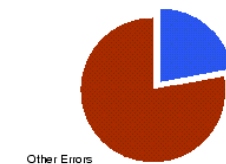
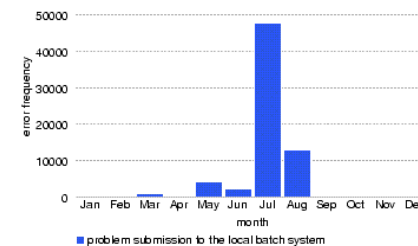
- Select the list of most common error
- Progression of error over time
- Pointers to the gocwiki (wherever possible)
- Restrict to a site and/or month

Sites which have had problems with the [problem submission to the local batch system](#) error type

Click on the site to view it's details

Display Name		Frequency of Error On Site	
egce.man.poznan.pl (Poznan, Poland)		66910	
Site Name	Site WWW	Site Email	Site Location
egce.man.poznan.pl	http://www.egce.man.poznan.pl	mailto:egce@man.poznan.pl	Poznan, Poland

Error Progression on Site over the Past Year and the proportion this Error covers over all errors on this site



NIKHEF-ELPROD (Amsterdam, Netherlands)	39655
LKI-SOUTHGRID-BHAM-HEP (Birmingham, UK)	36195
Amsterdam, Netherlands	26038

This table presents all the last error messages. It is ordered according to the number of times each error occurred.

If you click on the counter, you will see all workflows that finished with that error message

[Click here to view Pie Chart](#)

Error message	Counter	More info	Site Details
unknown	942848		Site Details
matchmaking failure	349349		Site Details
problem submission to the local batch system	304754		Site Details
job status message	144728		Site Details
termination problem	54978		Site Details
execution problem	13547		Site Details
miscellaneous	5047		Site Details
authentication problem	1938	See more info about this error...	Site Details
CE configuration problem	566	See more info about this error...	Site Details
proxy creation/renewal/expiration problem	491		Site Details
I/O problem	388		Site Details
warning related to the output sandbox content	36		Site Details

If you are interested in a particular site or date, please select it with the form below

Cannot read JobWrapper output...

Full message

Cannot read jobwrapper output, both from console and from hardware

Diagnosis

The user job exit status failed to be delivered to the RB, when two independent methods should have been tried:

1. The user job exit status is written into an extra "Hardware" file that is copied to the RB with globus-url-copy.
2. The job wrapper script writes the user job exit status to stdout, which is supposed to be sent back to the RB by Globus.

When both methods fail, it usually means that the job did not run to completion!

That means it either did not start at all:

- batch system submission problem (e.g. batch system in crazy state)
- WB disk full
- home directory absent or unwritable
- home directories not shared between CE and WB, while using standard job manager
- home directories on CE and WB have different paths (symlinks may not work)
- time not synchronized between CE and WB
- mismatch between forward and reverse DNS for CE name/IP-address
- WB cannot globus-url-copy from/to CE
- WB cannot scp to/from CE
- ...



- Total time (from submission to completion) for a type of jobs

Please enter the following information to view a plot with the time that it took for jobs to be executed, from the moment of the submission, until the job was finished.

NOTE: MINIMUM DURATION TIME MUST BE SMALLER THAN MAXIMUM DURATION TIME

Select a Site (Optional)

Enter the minimum duration Select Unit Of Time

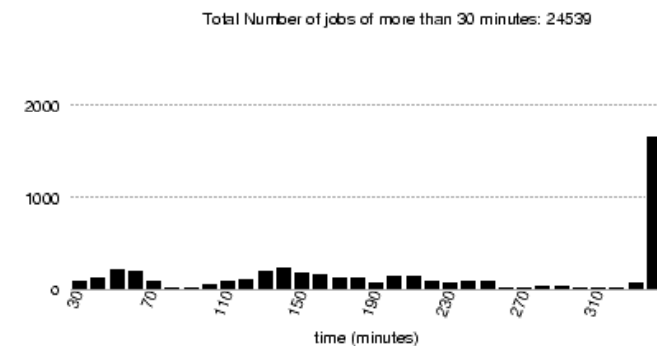
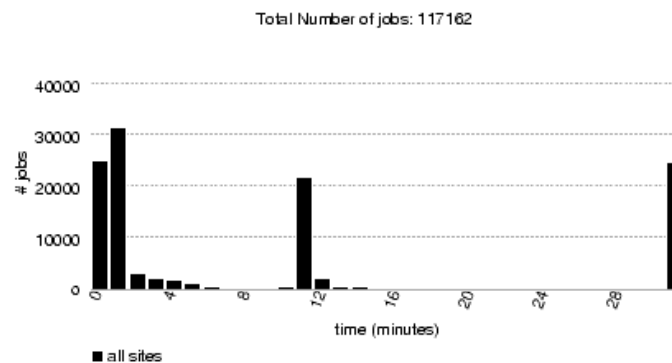
Enter the maximum duration Select Unit Of Time

Enter the number of columns to include in plot

Select Date Span (Optional) to

Please click on what you wish to retrieve
Request Plot ☐ Request Data ☐

This presents the waiting time for jobs Jobs that run for at least 3SECONDS and no more than 10SECONDS
There is a lot of data to process, so please be patient while the plot is created...



- Daily multi-VO report for a selected number of sites (T1)
 - See at a glance if everything is working
- Monthly automatic reports with:
 - Efficiency tables
 - Summary of job attempts per site

- Clicking on any cell expands the information
- Available since January

SiteName	ALICE	ATLAS	CMS	LHCb	TOTAL
BNL-LCG2	no jobs	9 vs. 598 (1.48%)	no jobs	no jobs	9 vs 598 (1.48%)
CERN-PROD	56390 vs. 1473 (97.45%)	2163 vs. 881 (71.06%)	36470 vs. 12282 (74.81%)	16645 vs. 4258 (79.63%)	111668 vs 18894 (85.53%)
FZK-LCG2	29601 vs. 5483 (84.37%)	2092 vs. 987 (67.94%)	3065 vs. 4041 (43.13%)	13950 vs. 10550 (56.94%)	48708 vs 21061 (69.81%)
IN2P3-CC	10640 vs. 277 (97.46%)	756 vs. 602 (55.67%)	4174 vs. 2131 (66.20%)	4013 vs. 787 (83.60%)	19583 vs 3797 (83.76%)
INFN-T1	29104 vs. 960 (96.81%)	1077 vs. 1651 (39.48%)	11049 vs. 1122 (90.78%)	5257 vs. 1058 (83.25%)	46487 vs 4791 (90.66%)
NIKHEF-ELPROD	493 vs. 2 (99.60%)	687 vs. 216 (76.08%)	15 vs. 1 (93.75%)	5089 vs. 176 (96.66%)	6284 vs 395 (94.09%)
RAL-LCG2	3793 vs. 4140 (47.81%)	1454 vs. 804 (64.39%)	5606 vs. 2805 (66.65%)	7344 vs. 15806 (31.72%)	18197 vs 23555 (43.58%)
Taiwan-LCG2	no jobs	524 vs. 430 (54.93%)	3240 vs. 412 (88.72%)	no jobs	3764 vs 842 (81.72%)
USCMS-FNAL-WC1	no jobs	no jobs	6262 vs. 872 (87.78%)	no jobs	6262 vs 872 (87.78%)
pic	no jobs	1193 vs. 302 (79.80%)	2482 vs. 1459 (62.98%)	3748 vs. 7199 (34.24%)	7423 vs 8960 (45.31%)
TOTAL	130021 vs 12335 (91.34%)	9955 vs 6471 (60.61%)	72363 vs 25125 (74.23%)	56046 vs 39834 (58.45%)	268385 vs 83765 (76.21%)

- Reports automatically created at the end of each month

Summary table
Evaluation of site performance Summary table
Evaluation of cost performance
Evaluation of final performance
Evaluation of safety performance
Evaluation of QA performance
Evaluation of EHS

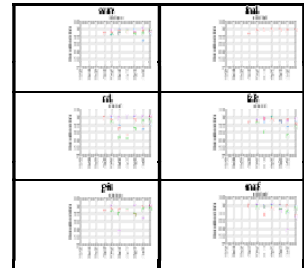
Summary table

The tables show the efficiency mean values over the reference period, grouped according to substrate type, with the relevant applications. The charts used to make it for the following specific application.

	com	mal	rl	rah	pk	swat	utfrd	avg	maps
sub	0.5	0.5	0.5	0.4	0.4	0.4	0.6	0.5	0.5
rule	0.0	***	0.7	0.7	0.6	0.2	0.0	***	0.0
count	0.0	***	0.6	0.5	***	0.4	0.0	***	0.0
score	0.5	***	0.0	0.0	0.6	0.4	0.0	0.0	0.0

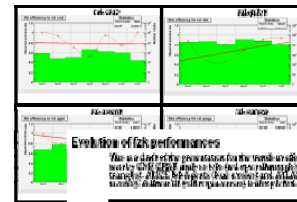
Evolution of site performances (summaries)

The Δ is a direct measure of the population for the transfer of drugs. However, distance is another factor that affects the rate of spread of drugs. The Δ is a direct measure of the transfer of drugs. The Δ is a direct measure of the transfer of drugs. The Δ is a direct measure of the transfer of drugs.



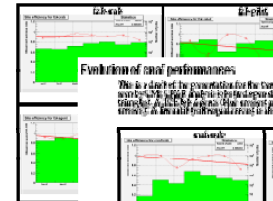
Evolution of rail performances:

This is a detailed presentation for the teacher's edition showing an example of the CPM and the other two standards. The first two standards are shown. The third standard is the same as the first two standards and is not shown. The fourth standard is the same as the first two standards and is not shown.



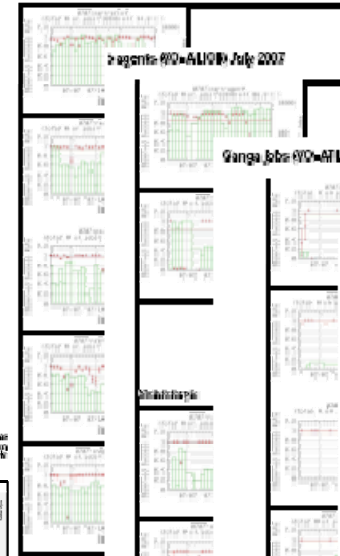
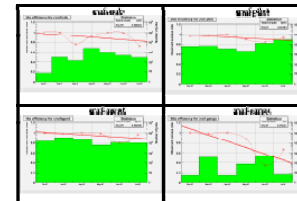
Evolution of fzk performances

This is a draft of the presentation for the 2008 annual meeting of the Society for the Study of the History of the Geological Sciences (SSGHS) and is for informational purposes only. It is not intended for distribution outside the SSGHS community. It is subject to change and is not to be used for any other purpose without the permission of the SSGHS.



Evaluation of anal performance

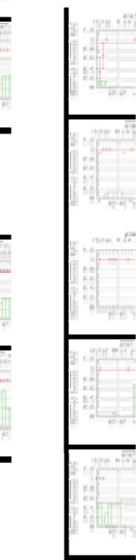
This is a direct representation for the variables being observed (things) as well as the independent variables (things that are being manipulated). As such, the system (the system and all its surroundings) is the system, as the whole system is being observed.



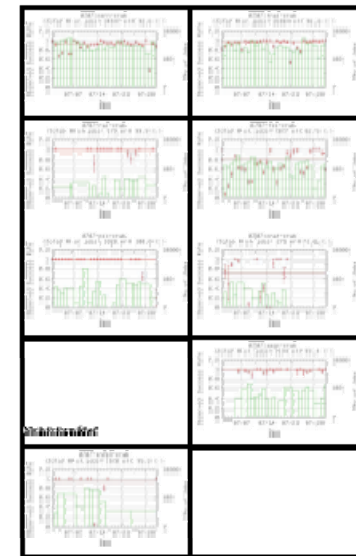
Abstract



Wanga Joba (YO-ATLAS) July 2007



ORCID iD: <https://orcid.org/0000-0001-9000-0001>



- Daily reports on successful/failed transfers
- Last 24 hours report updated every hour

FTS EFFICIENCY

Click on any Site, and you will have a breakdown according to the errors transferring files to that site

This table presents the transfers that have been done from CERN to the ALICE T1

Transfers done on: Thu 12 Oct 2006

Site (click on any site)	Successful transfers	Failed transfers	Efficiency
ALICE::LCG::SARA	0	2479	0.00 %
Error message			Counter
The file has size _size_ and should have _size_			1634
The FTS transfer _transferid_ failed (Failed on SRM put: Cannot Contact SRM Service. Error in srm_ping: SOAP-ENV:Client - CGSI-gSOAP: Error reading token data: Success)			510
The FTS transfer _transferid_ failed (Failed on SRM get: Failed SRM get on http://castorgridsc.cern.ch:8443/srm/managerv1			319
The FTS transfer _transferid_ failed (Failed on SRM put: Failed SRM put on http://srm.grid.sara.nl:8443/srm/managerv1			13
The FTS transfer _transferid_ failed (Failed on SRM put: Cannot Contact SRM Service. Error in srm_ping: SOAP-ENV:Client - CGSI-gSOAP: Could not open connection			2
The FTS transfer _transferid_ failed (Failed on SRM put: Cannot Contact SRM Service. Error in srm_ping: SOAP-ENV:Client - CGSI-gSOAP: Error reading token data: Connection reset by peer)			1
ALICE::LCG::RAL	120	166	41.96 %
Error message			Counter
The FTS transfer _transferid_ failed (Transfer failed. ERROR the server sent an error response: 451 451 Local resource failure: malloc: Cannot allocate memory.)			117
The FTS transfer _transferid_ failed (Transfer failed. ERROR the server sent an error response: 425 425 Cannot open port: java.lang.Exception: Pool request timed out : csfnfs62_1)			48
The FTS transfer _transferid_ failed (Failed on SRM put: Failed SRM put on http://dcache-tape.gridpp.rl.ac.uk:8443/srm/managerv1			1
ALICE::LCG::FZK	1774	392	81.90 %
Error message			Counter
The FTS transfer _transferid_ failed (Failed on SRM get: Failed SRM get on http://castorgridsc.cern.ch:8443/srm/managerv1			335
contacting the Broker/Transfer			53
The file has size _size_ and should have _size_			2
syntax error at line 1, column 0, byte 0 at /grid/fzk.de/mounts/nfs/software/alice/lcg2/alien2/lib/perl5/site_perl/5.8.7/i686-linux/XML/Parser.pm line 187 500 Can't connect to aliendb1.cern.ch:8095 (connect: timeout)			1
The FTS transfer _transferid_ failed (Failed on SRM put: Failed SRM put on http://gridka-dcache.fzk.de:8443/srm/managerv1			1
ALICE::LCG::CNAF	2030	366	84.72 %
ALICE::LCG::CCIN2P3	2171	343	86.36 %

- Study job workflow from other sources:
 - Pass the information from DIRAC (LHCb)
- X509 authentication in the web interface
- Group similar job attempts into patterns
- Data management reports for ATLAS
- Support its usage by sites and VOs
 - Always open to suggestions ☺
 - Adjust the tools according to the requests

- We provide several tools to investigate the grid reliability
- Data Management:
 - FTS reliability for ALICE
- Workload Management:
 - ‘Site of the day’, ‘Error messages’, ‘Site performance’
 - Aggregated views
 - Multi VO and automatic d monthly reports
- Already in use by ALICE, ATLAS, CMS and LHCb
 - Deployment for Vmed on its way

- <http://dashboard.cern.ch>

- FTS:

<http://dboard-gr.cern.ch/dashboard/data/fts/index.html>

- WMS:

<http://dashb-alice.cern.ch/jr.html>

<http://dashb-atlas-job.cern.ch/jr.html>

<http://dashb-lhcb.cern.ch/jr.html>