



The Grid Observatory

Cécile Germain

Julien Nauroy

Laboratoire de Recherche en Informatique

Université Paris Sud, CNRS, INRIA

2013-09-18

The Grid Observatory Project

The G.O. collects, publishes and analyzes data on the behavior of the EGI grid since 2008

It provides scientists with large-scale, long-term, exhaustive, curated datasets

- BDII, EGI-wide, since 2008
- WMS data from GRIF-LAL
- Logging & Bookkeeping, GRIF-wide
- GridFTP logs from the GRIF-LAL SE, since 2010
- RTM backups (retired since 2012)

Data is anonymized & curated before publication

The Grid Observatory collaboration

- Operators and funding
 - CNRS/UPS Laboratoire de Recherche en Informatique
 - CNRS/UPS Laboratoire de l'Accélérateur Linéaire
 - Imperial College London
 - France Grilles – French NGI of EGI
 - EGI-Inspire
 - Ile de France council
(Software and Complex Systems programme)
 - INRIA – Saclay (ADT programme)
 - CNRS (PEPS programme)
 - University Paris Sud (MRM programme)
- Scientific Collaborations
 - NSF Center for Autonomic Computing
 - European Middleware Initiative
 - Institut des Systèmes Complexes
 - Cardiff University



The Grid Observatory Project

Traces are available through the web portal



Grid Observatory

Grid Observatory
About
Data
Presentation
Query
Documents
Registration
Resources and links
GCO Monitor
Events Calendar

Contact
Site map
Admin

Latest news

News Events



Design by HealthGrid
HTML 5 | CSS 3

[Search](#) [My Profile](#) [Logout](#)
[Users](#) [Queries](#) [Cache](#) [Stats](#) [Banners](#)

Search for Traces

Search filters

Site*
Service*
Start date*
End date*

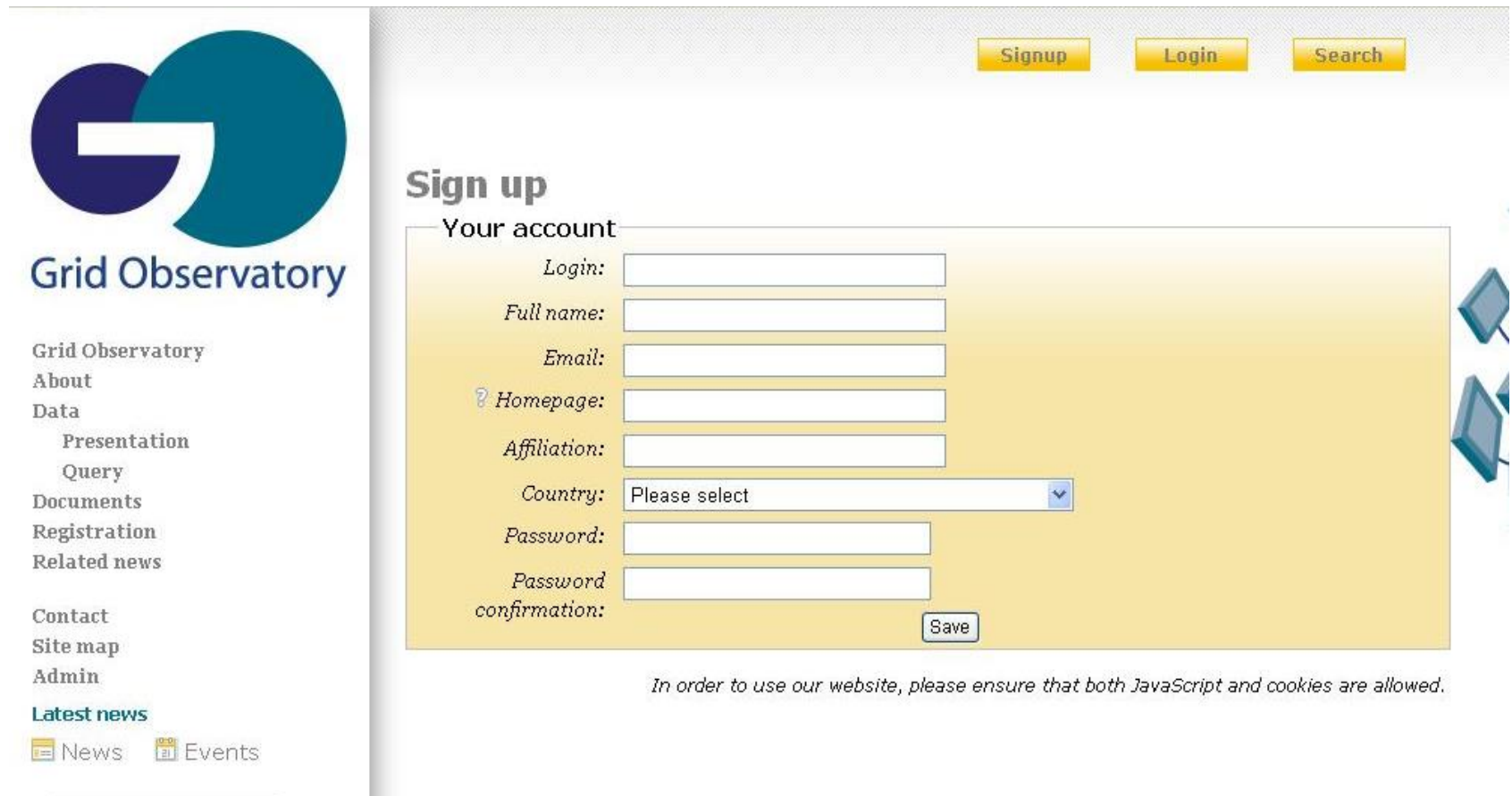
Query results


Site	Service	Start Date	End Date	Size	
EGI	Information System	2013-08-26	2013-09-01	237 MB	Download trace (555 MB compressed)
EGI	Information System	2013-08-19	2013-08-25	870 MB	Download trace (638 MB compressed)
EGI	Information System	2013-08-12	2013-08-18	256 MB	Download trace (564 MB compressed)
EGI	Information System	2013-08-05	2013-08-11	2.18 GB	Download trace (794 MB compressed)
EGI	Information System	2013-07-29	2013-08-04	741 MB	Download trace (634 MB compressed)
EGI	Information System	2013-07-22	2013-07-28	1.4 GB	Download trace (710 MB compressed)
EGI	Information System	2013-07-15	2013-07-21	1.23 GB	Download trace (676 MB compressed)
EGI	Information System	2013-07-08	2013-07-14	1.22 GB	Download trace (675 MB compressed)



The Grid Observatory Project



Basic registration is required




Grid Observatory

Grid Observatory
About
Data
 Presentation
 Query
Documents
Registration
Related news

Contact
Site map
Admin

Latest news
 News  Events

Sign up

Your account

Login:

Full name:

Email:

? Homepage:

Affiliation:

Country: Please select

Password:

Password confirmation:

In order to use our website, please ensure that both JavaScript and cookies are allowed.

Summary

- Recent developments
- Use cases and results
- Preparing for the future: 2013 - 2014

Recent developments 2011 - 2013

Building of the GCO

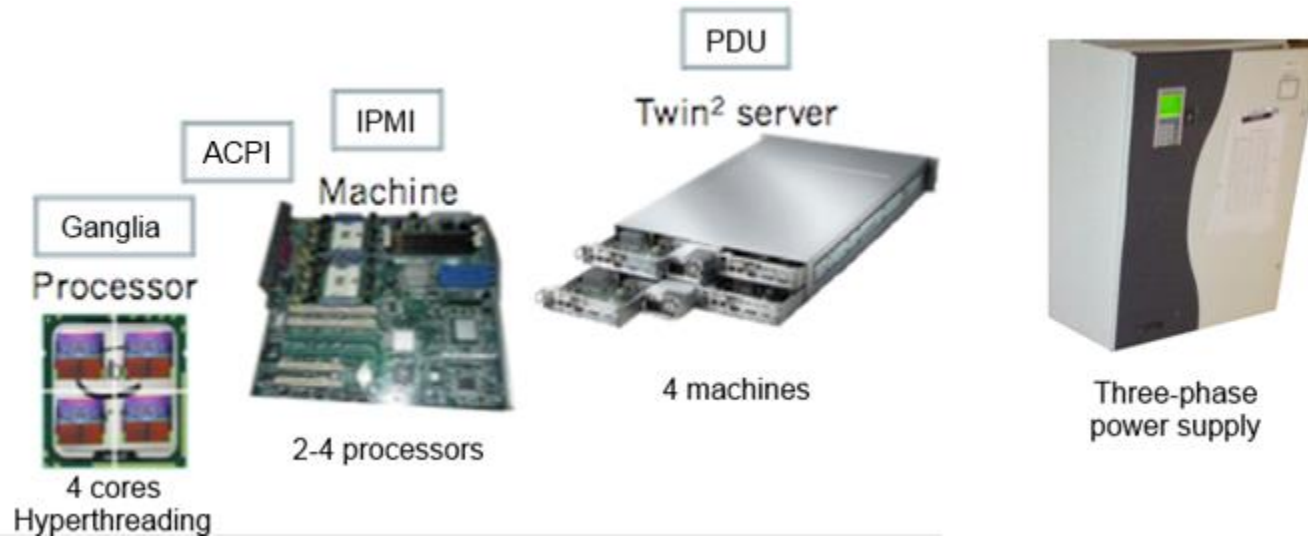
The Green Computing Observatory Project

- Electric consumption, heat production
- Monitoring of ~250 computing machines at LAL
- Built from the ground up VS G.O. on top of gLite monitoring
- Integration of the LAL's Stratuslab Cloud instance

Publication started in January 2012

Building of the GCO

Heterogeneous data sources (7), hardware & software



Automation requires careful selection of data

- e.g. association PDU-machine, 3-phase PS

Failures arise

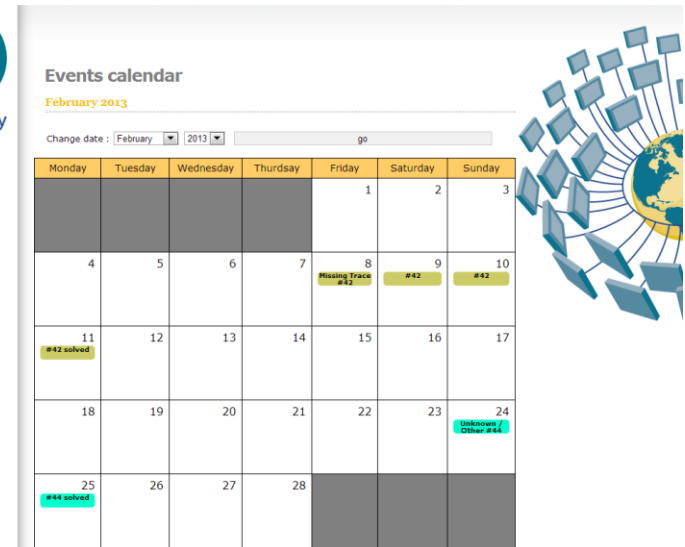
- Network, hardware, software

Building of tools for the GCO

Created analysis tools with advanced visualization



GCO Monitor



Event Reporter



Acquisition of data from a cloud



Started acquisition of an IaaS cloud

- Instance at LAL: 18 machines with 16-32 cores & 32-64GB RAM

Hosts monitored through common G.C.O. tools

- Ganglia, IPMI & ACPI monitoring

VMs lifecycle acquired from logs & database

Publication expected October 2013

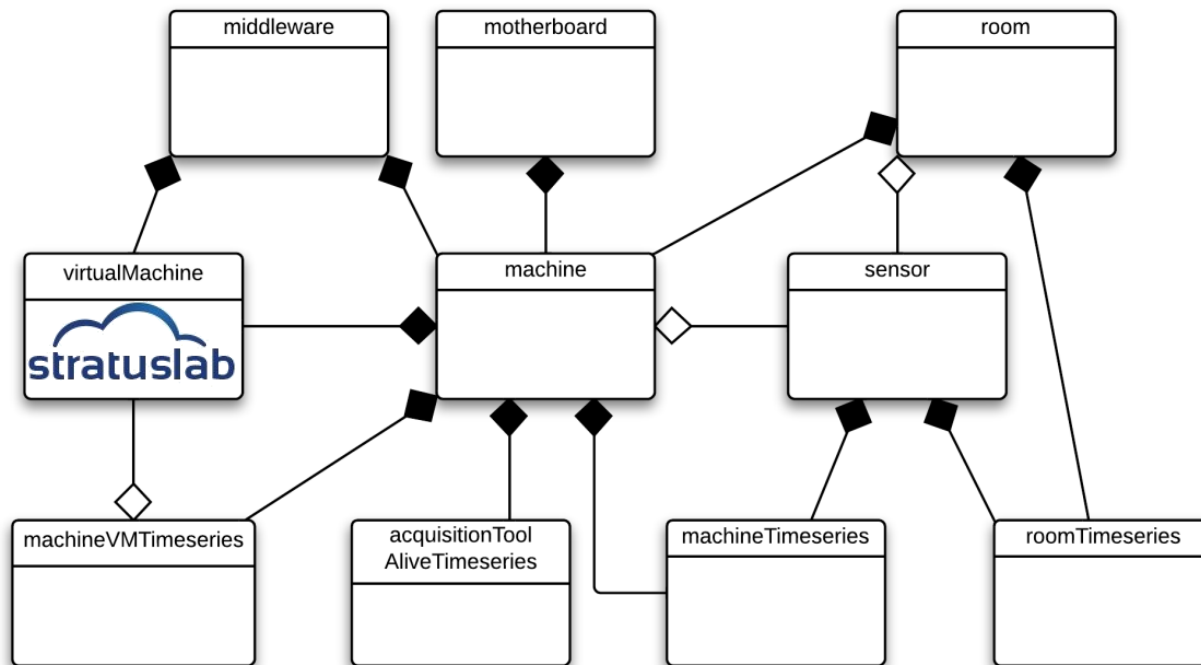
- Need to ensure the data is consistent

Integration within a common data model

Data model built on top of an ontology

- Collaboration with G. Kassel & F. Fürst, UPJV

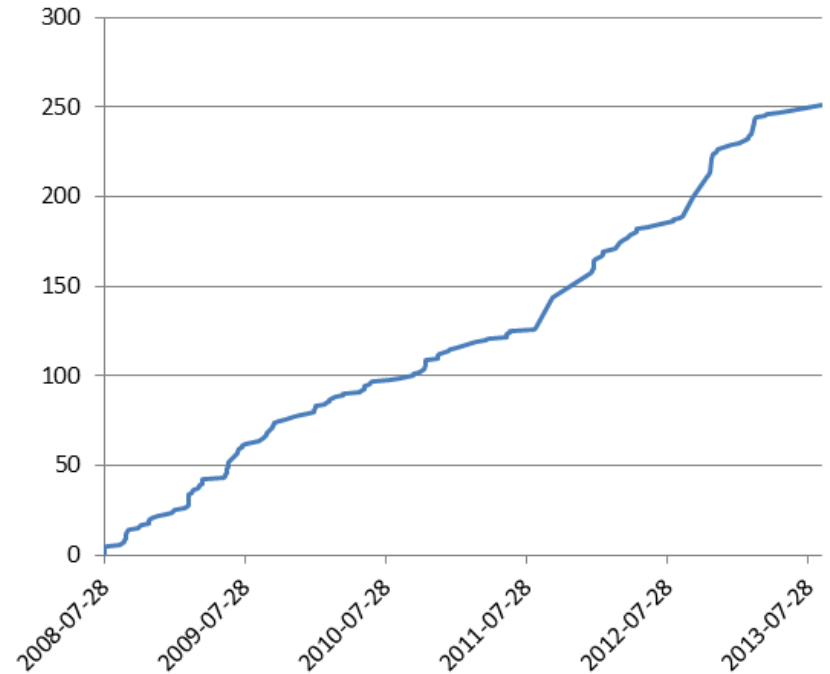
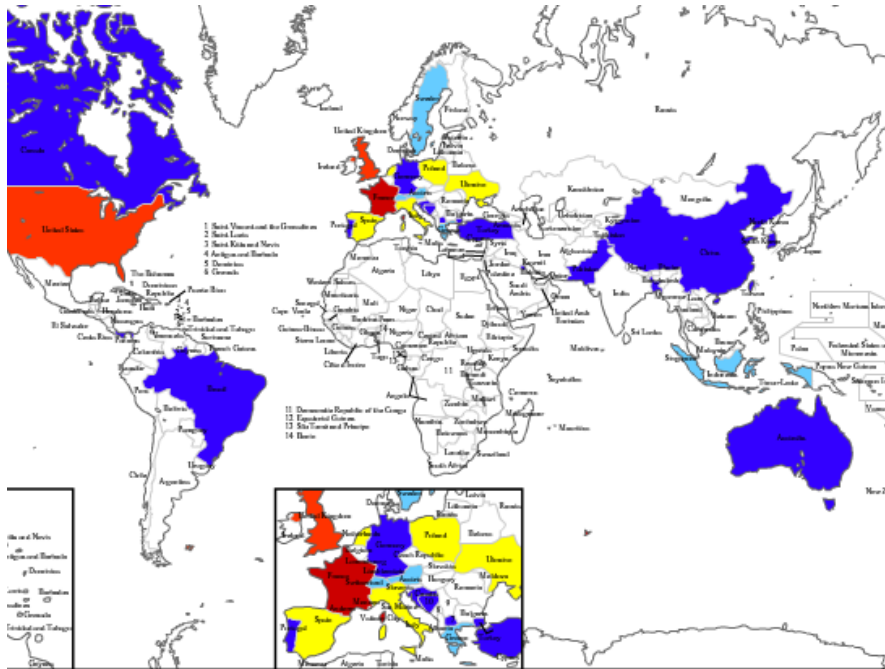
Publication in XML format



Use cases & results

A growing community

- ~250 users from 40 countries



Results from the users' community

Ph.D. theses:

- D. Bandara – simulator (2012)
- W. Mulder – machine learning (2011)
- X. Zhang – data clustering (2010)
- J. Perez – machine learning for scheduling (2010)

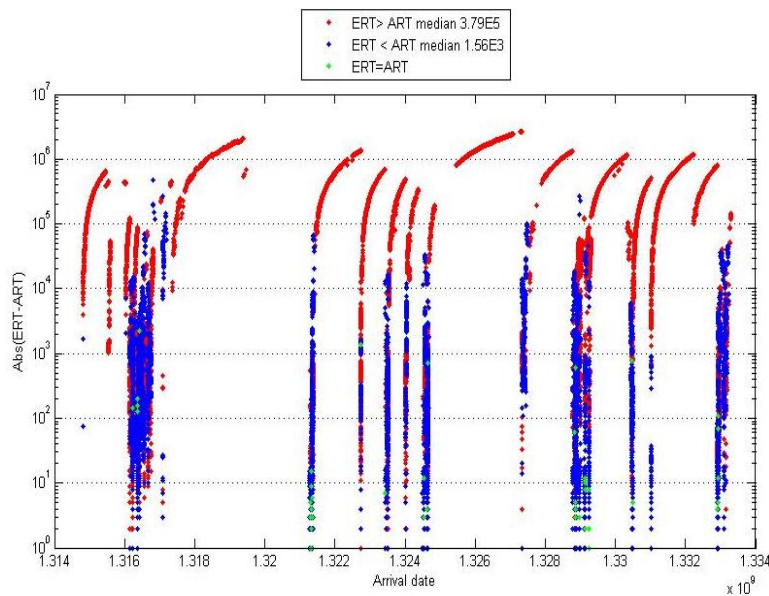
Numerous publications in machine learning
& autonomic computing

- Some available on grid-observatory.org/index.php?id=72
- Most recent one: X. Zhang et al. Data Stream Clustering with Affinity Propagation. To appear in Transactions on Knowledge and Data Engineering

GO files as a test platform for SAGA

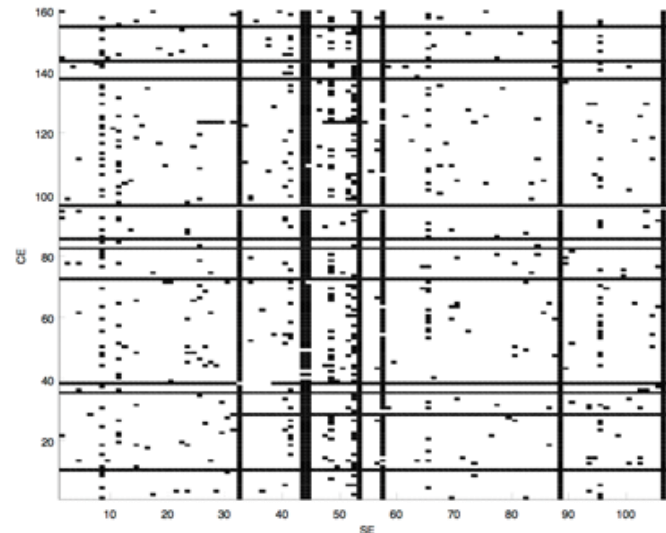
Results from the G.O.

Queues' waiting time



[The Grid Observatory.
In CCGRID 2011]

Smart probing

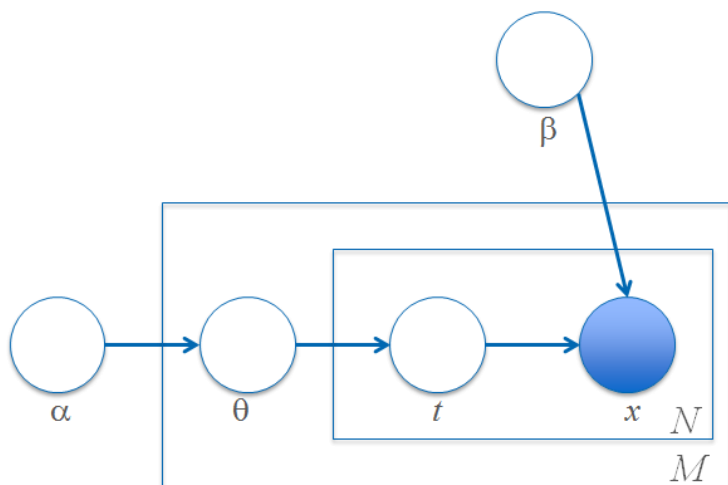


[Distributed Monitoring with
Collaborative Prediction.
IEEE/ACM CCGrid 2012]



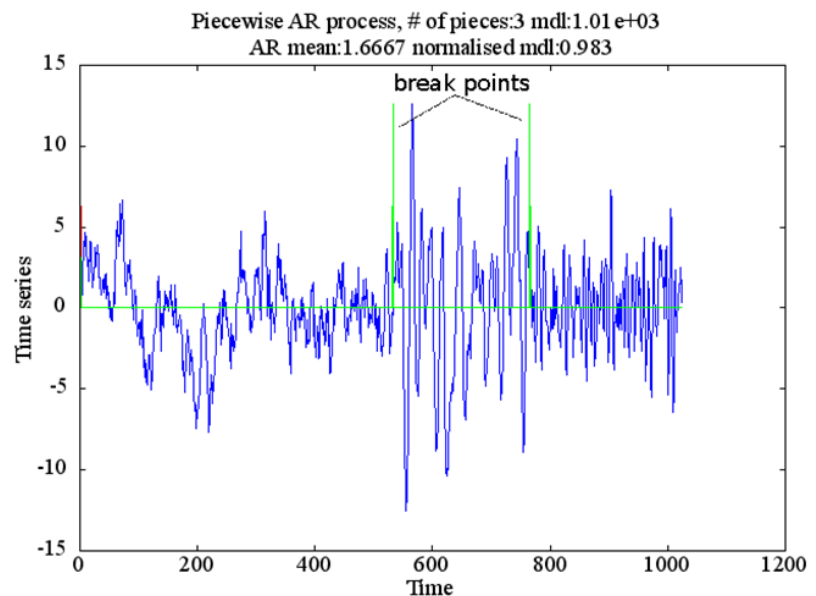
Results from the G.O.

Trace mining



[Characterizing E-Science File Access Behavior via Latent Dirichlet Allocation, IEEE/ACM UCC 2011]

Regime identification



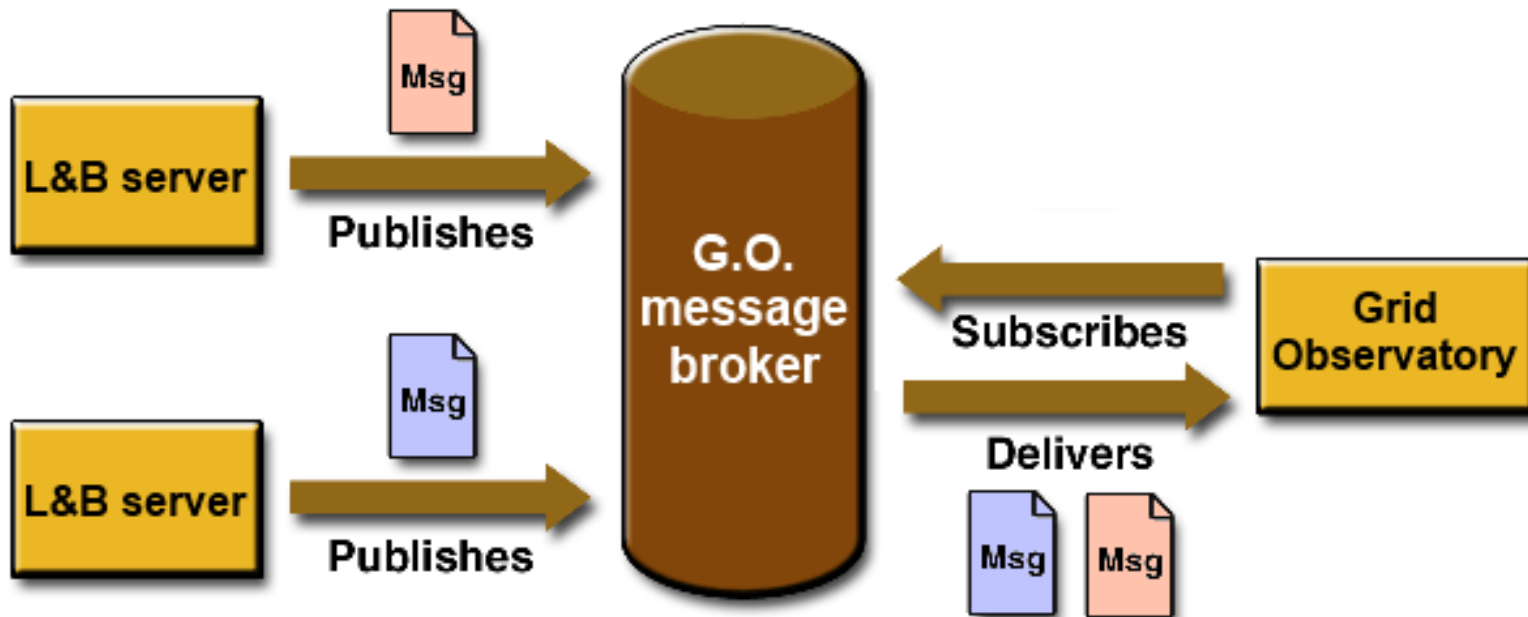
[Towards non stationary Grid Models, Journal of Grid Computing 9(4) Dec. 2011]

Preparing for the future 2013 - 2014

Integration with EGI Monitoring

Monitoring of EGI services through messaging

- G.C.O. data already conveyed through ActiveMQ
- L&B is the primary goal, experiment under way
- Other services from EGI moving to messaging



Towards a Grid Observatory v2.0

Integration of sources & new access facilities

- Goals: Linked Open Data & RDF

Integration of current data sources

- L&B, WMS, BDII,...
- G.C.O. data model will be the starting point



Scale up acquisition of existing data sources

- Starting with the L&B

Thank you
Any questions ?