

The Grid Observatory: Status and Issues

EGI Technical Forum - September 14, 2010



Grid Observatory

1.1

Cécile Germain-Renaud, for the GO team
Laboratoire de Recherche en Informatique
and Laboratoire de l'Accélérateur Linéaire
Université Paris-Sud and CNRS



A collection of data for Computer Science and Engineering

The Grid Observatory
Status and Issues

Cécile
Germain-Renaud, for
the GO team

2003 NSF Blue-Ribbon Advisory Panel on Cyberinfrastructure Report : Revolutionizing Science and Engineering through Cyberinfrastructure

- Grids of computational centers
- Comprehensive libraries of digital objects
- **Well-curated collections of scientific data**
- Online instruments and vast sensor arrays
- Convenient software toolkits



Grid Observatory

1.2



Grid digital assets curation

- Collecting verifiable digital assets
- Providing digital asset search and retrieval
- Certification of the trustworthiness and integrity
- Semantic and ontological continuity and comparability

Building the domain knowledge

- Answering operational issues
- Descriptive/generative/predictive models
- Design and validation of model-free policies



Support and collaborations



The Grid Observatory Portal

The Grid Observatory
Status and Issues

Cécile
Germain-Renaud, for
the GO team



Grid Observatory Search for Traces - Windows Internet Explorer

http://query.grid-observatory.org/

Grid Observatory Search for Traces

Grid Observatory

Grid Observatory
About
Data
Presentation
Query
Documents
Registration
Related news

Contact
Site map
Admin

Latest news

News Events

egée
Enabling Grids
for e-science

Design by HealthGrid
W3C XHTML 1.0 | W3C CSS 2.0

Search for Traces

Filters [Help](#)

Site: GRIF-LAL
Service: Information System
From: 2008-06-23 To: 2008-09-09
[Search](#)

Documentation about the traces [Help](#)

Query results [Help](#)

Affiliation	Service	Period	Download?
GRIF-LAL	Information System	2008-07-28 2008-08-03	Download (288 MB)
GRIF-LAL	Information System	2008-08-11 2008-08-17	Download (198 MB)
GRIF-LAL	Information System	2008-08-18 2008-08-24	Download (295 MB)
GRIF-LAL	Information System	2008-08-25 2008-08-31	Download (306 MB)
GRIF-LAL	Information System	2008-09-01 2008-09-07	Download (408 MB)
GRIF-LAL	Information System	2008-09-08 2008-09-14	Download (299 MB)

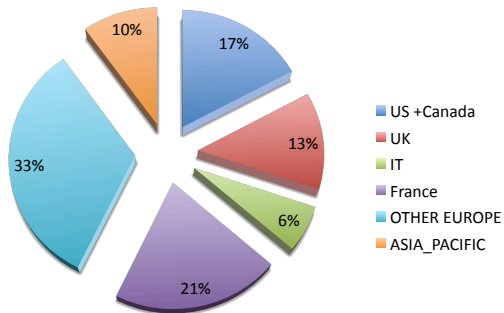
Displaying **all 6** traces

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

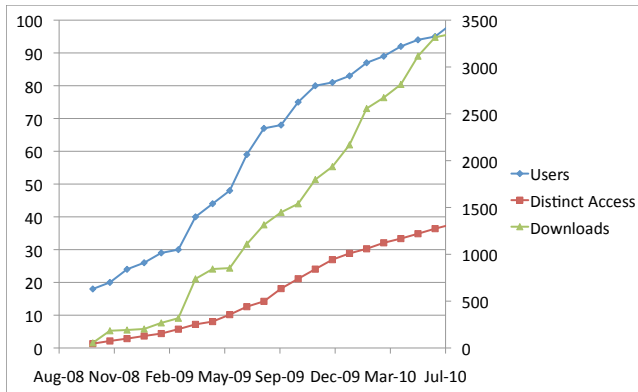
démarrer | Gestionnaire des tâches | Connexion réseau sa... | Logon | Mulberry (Connected) | Grid Observatory sta... | Grid Observatory Sta... | FR | 22:56

1.5





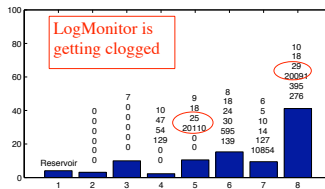
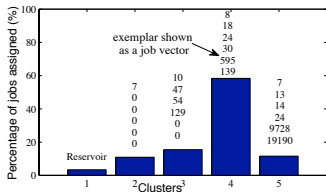
Activity



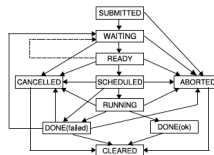
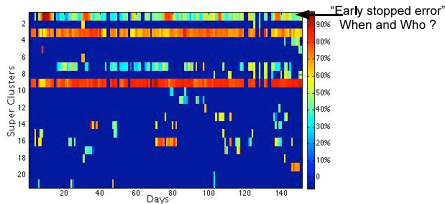
~ 20 peer-reviewed papers **that we know of**, latest M. Parashar et al. *Energy-Efficient Application-Aware Online Provisioning for Virtualized Clouds and Data Centers*
Proceedings of the 1st IEEE International Green Computing Conference.

G-StrAP: A Grid Dashboard

Online Monitoring



Off-line Analysis



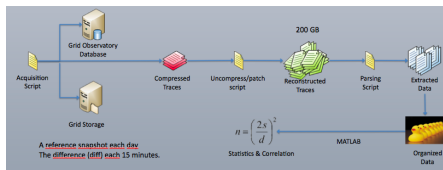
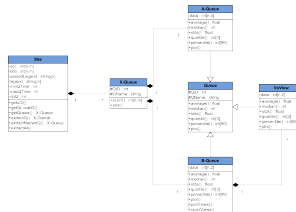
The Grid Observatory Status and Issues



Grid Observatory

1. Integrating heterogeneous data

- ERT - expected response time-published by the BDII
- ART- Actual Response Time-from local schedulers (eg PBS)



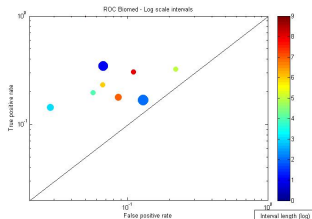
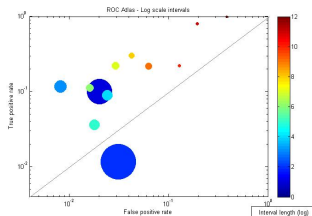
1.9

Example of use: is the ERT a good predictor for ART?

A request from EGEE JRA-3

2. What is **your** definition of good predictor?

- À la BQP (Batch Queue Predictor): how often does the prediction lie within a reasonable distance of the actual?
- ERT is a classifier, the classes are intervals of the value range Intervals
- ROC: True Positive Rate vs False Positive Rate

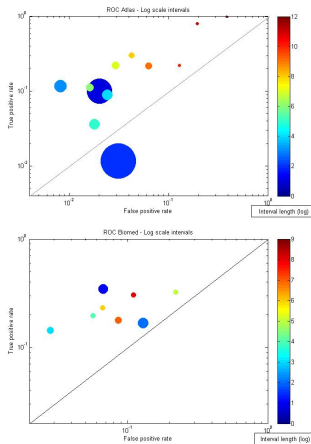


Example of use: is the ERT a good predictor for ART?

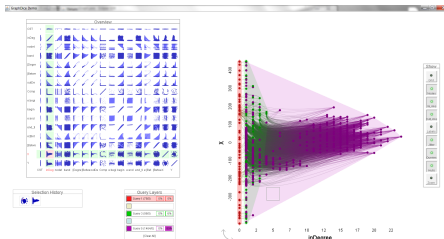
A request from EGEE JRA-3

3. Systematic Analysis

- On representative sites
- Not yet possible: ART data available only for one site (LAL)
- while comprehensive ERT are available



- Gridftp logs from the LAL site
- Collected and formatted
- Publication coming soon...
- Top priority for exhaustive collection

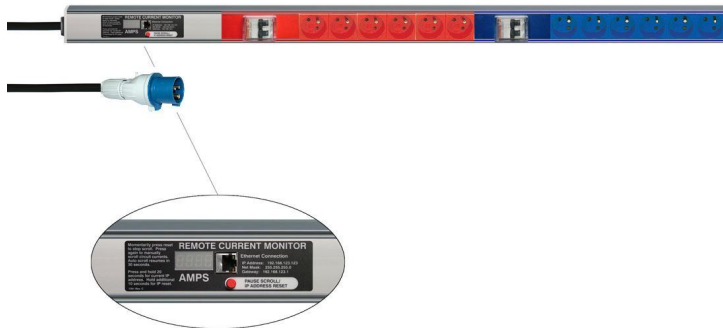


Users/Files/Clients worker nodes graph display with AVIZ GraphDice

Monitoring of energy consumption at the LAL computing center

- Provide fine grain data in the form of time series
- Operational infrastructure planning and optimization
- Prototype based on 32 machines, 300 (\sim 1500 cores) planned

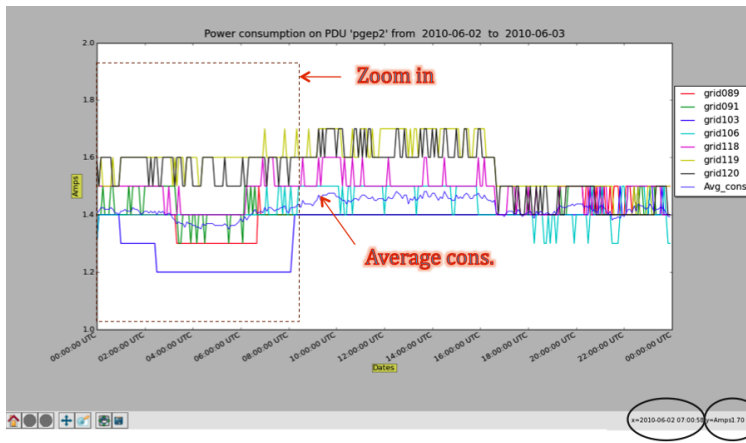




Power Distribution Unit (PDU) with SNMP-based remote monitoring

- Python scripts collect the data about individual machine power consumption and load, running every 5 minutes
- All these data are stored in a unique SQL database
- Visualization tools allowing correlations between these series

First results





- **Users Community** building
- More **user-friendly formats**, automatic consistency checking, more semantics
- More **participating sites**. Technology and process?
- **Actors Community** building