The Green Computing Observatory

Michel Jouvin (LAL)

Cécile Germain-Renaud (LRI), Thibaut Jacob (LRI), Gilles Kassel (MIS), Julien Nauroy (LRI), Guillaume Philippon (LAL)

Outline

- Contexts
- Acquisition
- Status and roadmap
- Scientific issues
- Conclusions

GCO in a nutshell

- Research about sustainable computing is suffering the lack of representative experimental data
    - In particular about power consumption profiles
- The GCO project aims to provide scientific community with data about a large production grid computing center with an experimental cloud platform
    - GCO takes care of both data acquisition, data curation and a first data analysis
- GCO combines expertise in managing a production computing center, expertise in ontology for the semantics of data and expertise in machine learning for data interpretation
- GCO is a sub-project of the well established Grid Observatory
    - Will use the same HW and SW infrastructure to publish data

Who are we?

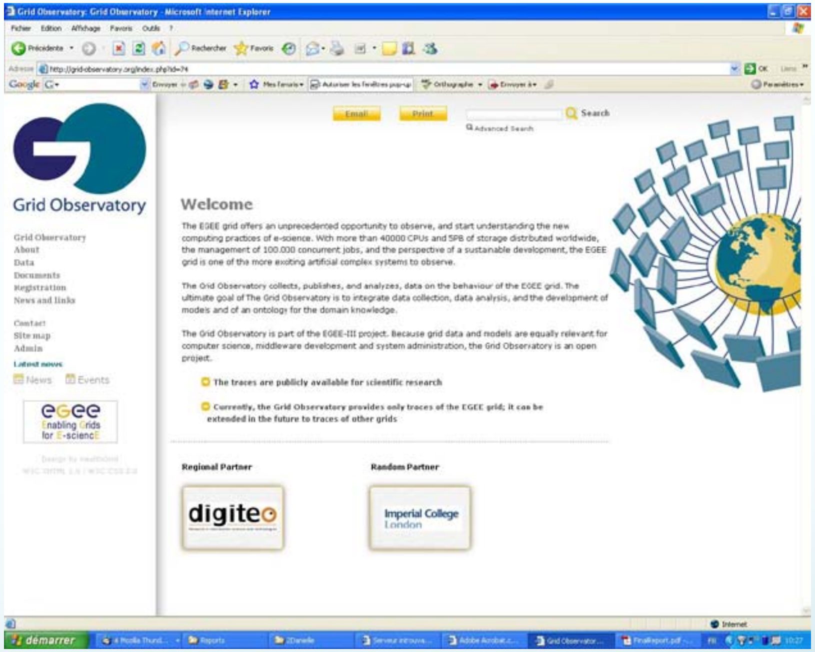
- A collaborative effort of
  - CNRS/UPS Laboratoire de Recherche en Informatique
  - CNRS/UPS Laboratoire de l'Accélérateur Linéaire (GRIF grid site)
  - U. Picardie MIS laboratory
- With the support of
  - France Grilles – French NGI member of EGI
  - EGI-Inspire (FP7 project supporting EGI)
  - INRIA – Saclay (ADT programme)
  - CNRS (PEPS programme)
  - University Paris Sud (MRM programme)

Motivation

- The metrics remain to be defined
  - “Energy efficient” means the delivery of the same or better service output with less energy input: how to define the service?
  - All costs should be considered : ideally should include building and recycling costs but probably too difficult to integrate
- Energy and power consumption are complex systems.
  - Sophisticated HW/SW mechanisms eg ACPI, dynamically over-clocking of active cores, and other optimisations based on on-line statistical monitoring.
  - Interaction with cooling provisioning (eg. fan speed), cooling efficiency (PUE)
  - Usefulness of powered IT
- Evaluation ideally requires behavioral models based on real data
  - Importance of *curated* data collection at various centers

The Grid Observatory (I): Digital Curation

- Behavioral data of the EGEE/EGI grid
  - Collection, preservation, indexing
  - Correlation with known operational events
  - Continuous and exhaustive datasets
- Portal allowing to download/query data
  - For scientific and engineering usage


Grid Observatory: Grid Observatory - Microsoft Internet Explorer
Fichier Edition Affichage Favoris Outils ?
Précédente Rechercher Favoris
Adresse http://grid-observatory.org/index.php?id=74
OK Liens
Google G- Envoyer Mes favoris Autoriser les fenêtres pop-up Orthographe Envoyer à Paramètres
Email
Print
Search
Advanced Search
Grid Observatory
Grid Observatory
About
Data
Documents
Registration
News and links
Contact
Site map
Admin
Latest news
News
Events
egee Enabling Grids for E-sciencE
Regional Partner
digiteo
Random Partner
Imperial College London
démarrer
Internet

Welcome

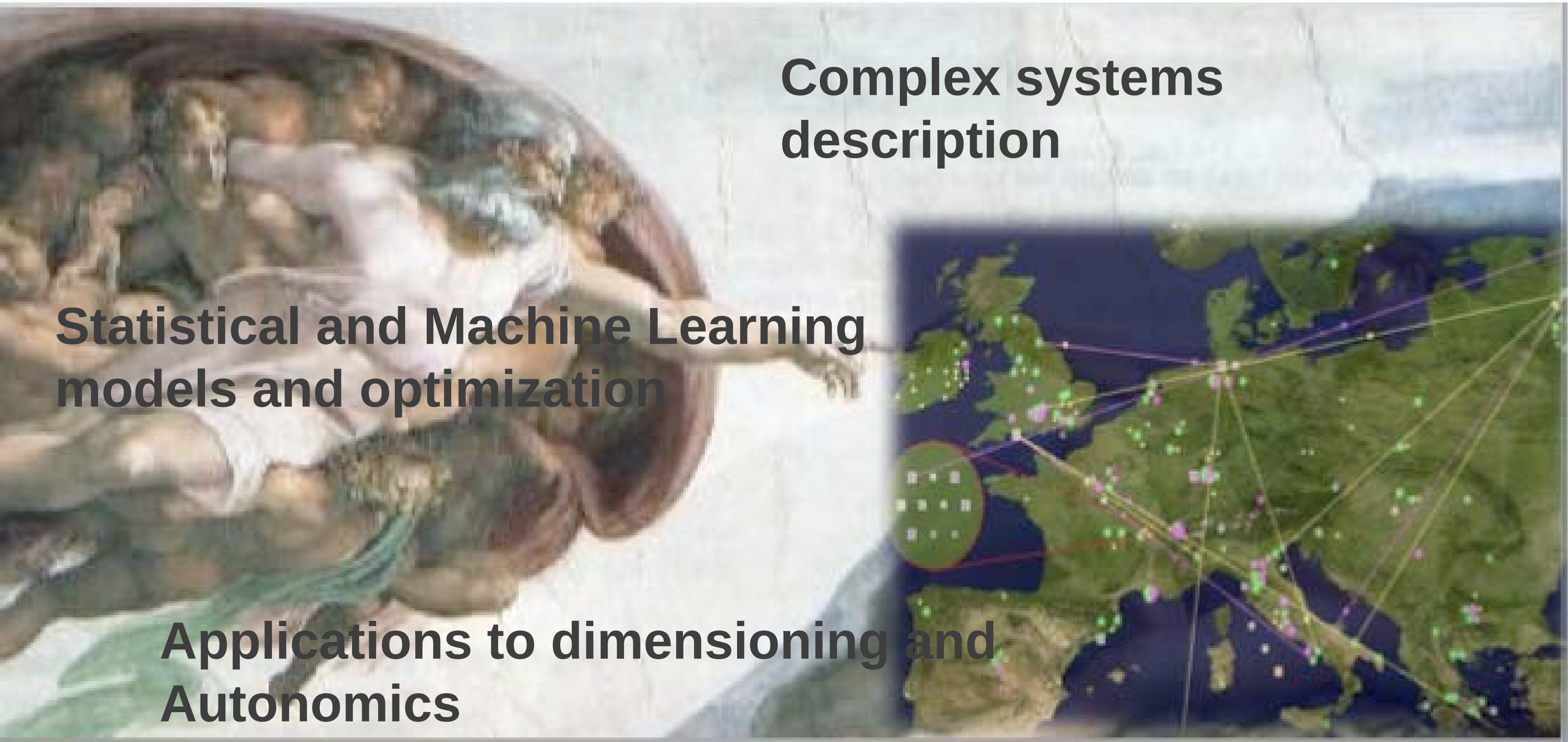
The EGEE grid offers an unprecedented opportunity to observe, and start understanding the new computing practices of e-science. With more than 40000 CPUs and 5PB of storage distributed worldwide, the management of 100.000 concurrent jobs, and the perspective of a sustainable development, the EGEE grid is one of the more exciting artificial complex systems to observe.

The Grid Observatory collects, publishes, and analyzes, data on the behaviour of the EGEE grid. The ultimate goal of The Grid Observatory is to integrate data collection, data analysis, and the development of models and of an ontology for the domain knowledge.

The Grid Observatory is part of the EGEE-III project. Because grid data and models are equally relevant for computer science, middleware development and system administration, the Grid Observatory is an open project.

- The traces are publicly available for scientific research
- Currently, the Grid Observatory provides only traces of the EGEE grid; it can be extended in the future to traces of other grids

The Grid Observatory (II): analysis and modeling



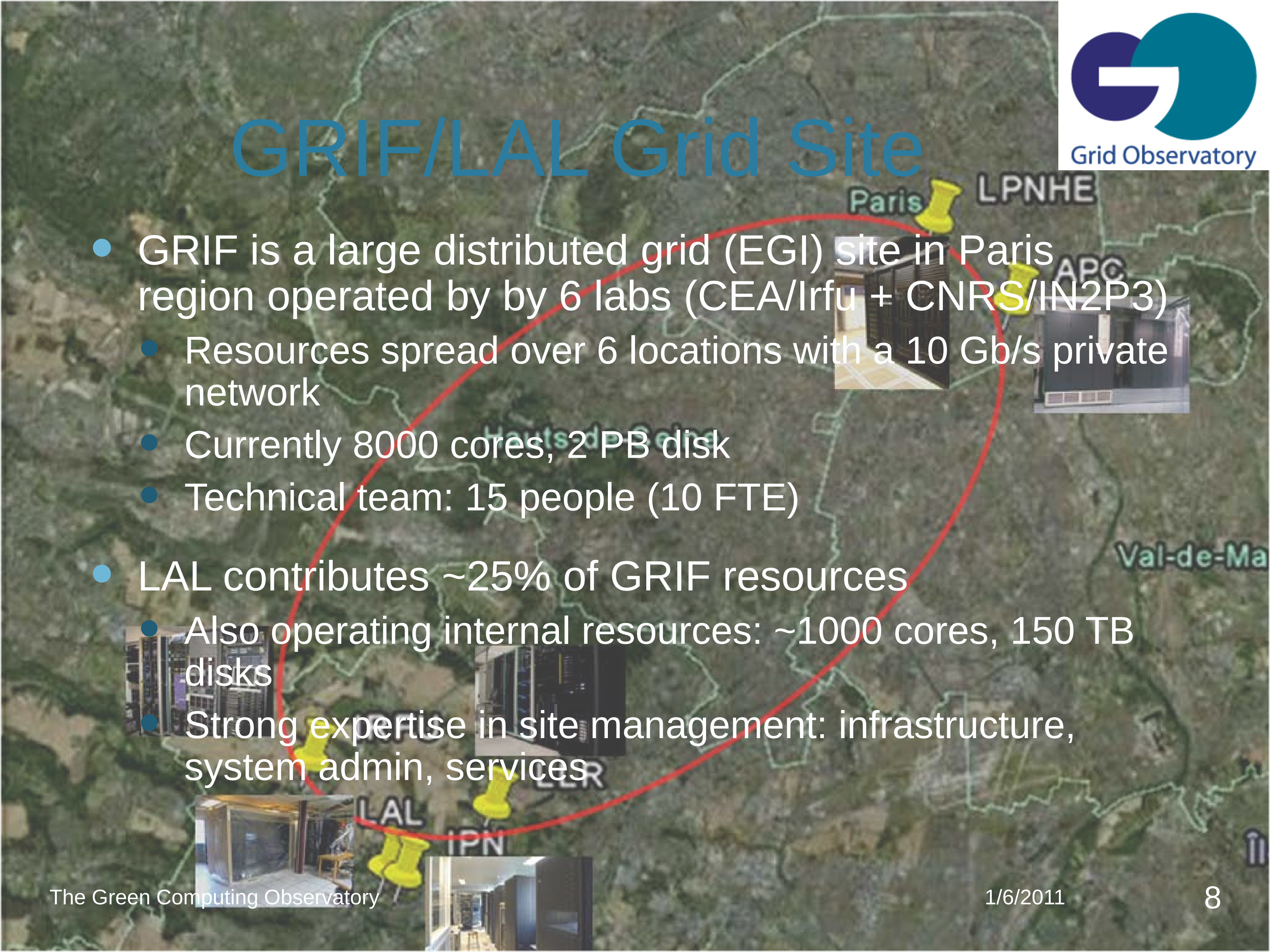
Complex systems description

Statistical and Machine Learning models and optimization

Applications to dimensioning and Autonomics

GRIF/LAL Grid Site

- GRIF is a large distributed grid (EGI) site in Paris region operated by by 6 labs (CEA/Irfu + CNRS/IN2P3)
  - Resources spread over 6 locations with a 10 Gb/s private network
  - Currently 8000 cores, 2 PB disk
  - Technical team: 15 people (10 FTE)
- LAL contributes ~25% of GRIF resources
  - Also operating internal resources: ~1000 cores, 150 TB disks
  - Strong expertise in site management: infrastructure, system admin, services



LAL Computing Room

- Mostly based on traditional racks + cooling
  - Cold-water based central cooling
  - 13 racks hosting 1U systems
  - 4 lower-density racks (network, storage)
- Recently introduced water-cooled racks
  - Cooling through back door (ATOS)

StratusLab

- Information
  - 1 June 2010—31 May 2012 (2 years)
  - 6 partners from 5 countries
  - Budget : 3.3 M€ (2.3 M€ EC)
- Goal
  - Create a comprehensive, open-source “private” cloud distribution
  - Focus on supporting grid services
- Contacts
  - Site web: http://stratuslab.eu/
  - Twitter: @StratusLab
  - Support: support@stratuslab.eu

CNRS (FR)

UCM (ES)

GRNET (GR)

SIXSQ (CH)

TID (ES)

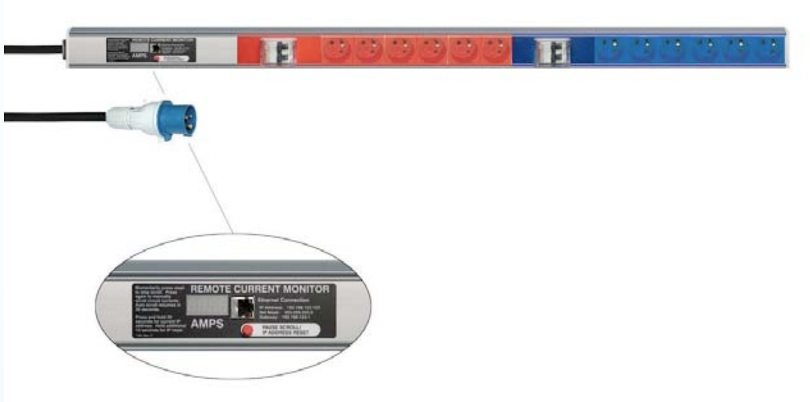
TCD (IE)

Acquisition

- Goal: monitoring the EGI GRIF/LAL site and the StratusLab testbed
  - Global energy usage based on room power distribution monitoring
    - Should include cooling power consumption
- 2 acquisition methods
  - PDU monitoring with outlet granularity
  - IPMI-based monitoring: fine grain information at motherboard level
- In-progress: correlating both to see if we can rely on IPMI

Smart PDU

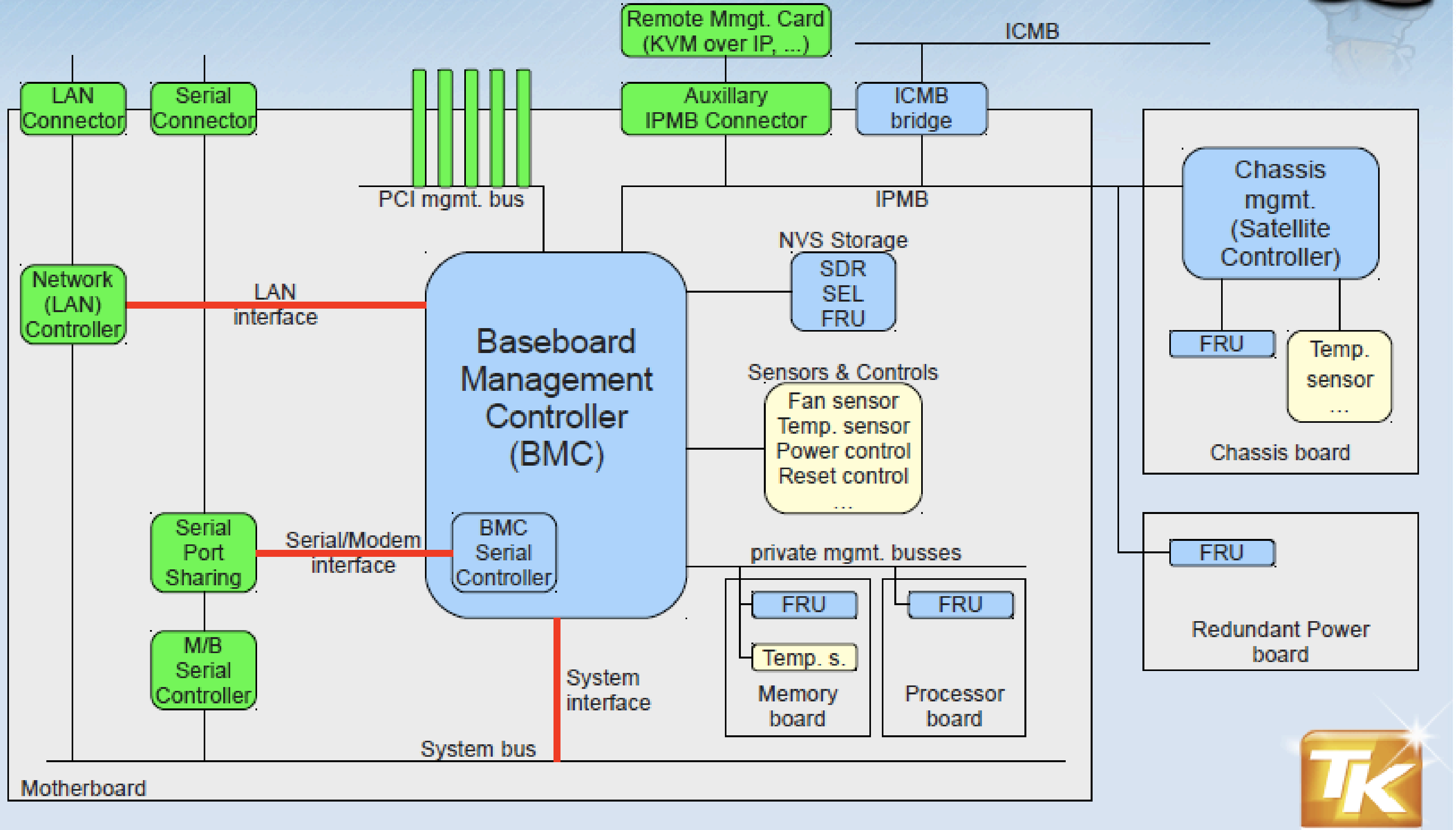
- PGEP PULTI
  - 16 outlets
  - Each PDU outlet managed separately
  - Query protocol : SNMP
  - Embedded Web server
- 1 rack (32U over 36) equiped
  - 1U system
  - Grid worker nodes
- Issue: last systems are Twin^2
  - 4 systems in 2U
  - 2 redundant power supplies



IPMI

- IPMI = Intelligent Platform Management Interface,
- Based on a specialized processor card (BMC)
  - 1998: IPMI v1.0, 2001: IPMI v1.5, originally by Intel, HP, NEC, Dell
  - 2004: IPMI v2.0 (matured version of IMPI)
  - De facto standard implemented by all motherboard vendors
- Allows fine grain monitoring of individual system parts…
  - Temperatures, fans, voltages, etc.
- And many other things: http://www.intel.com/design/servers/ipmi
  - Recovery Control (power on/off/reset a server)
  - Logging (System Event Log)
  - Inventory (FRU information)

2) IPMI basics



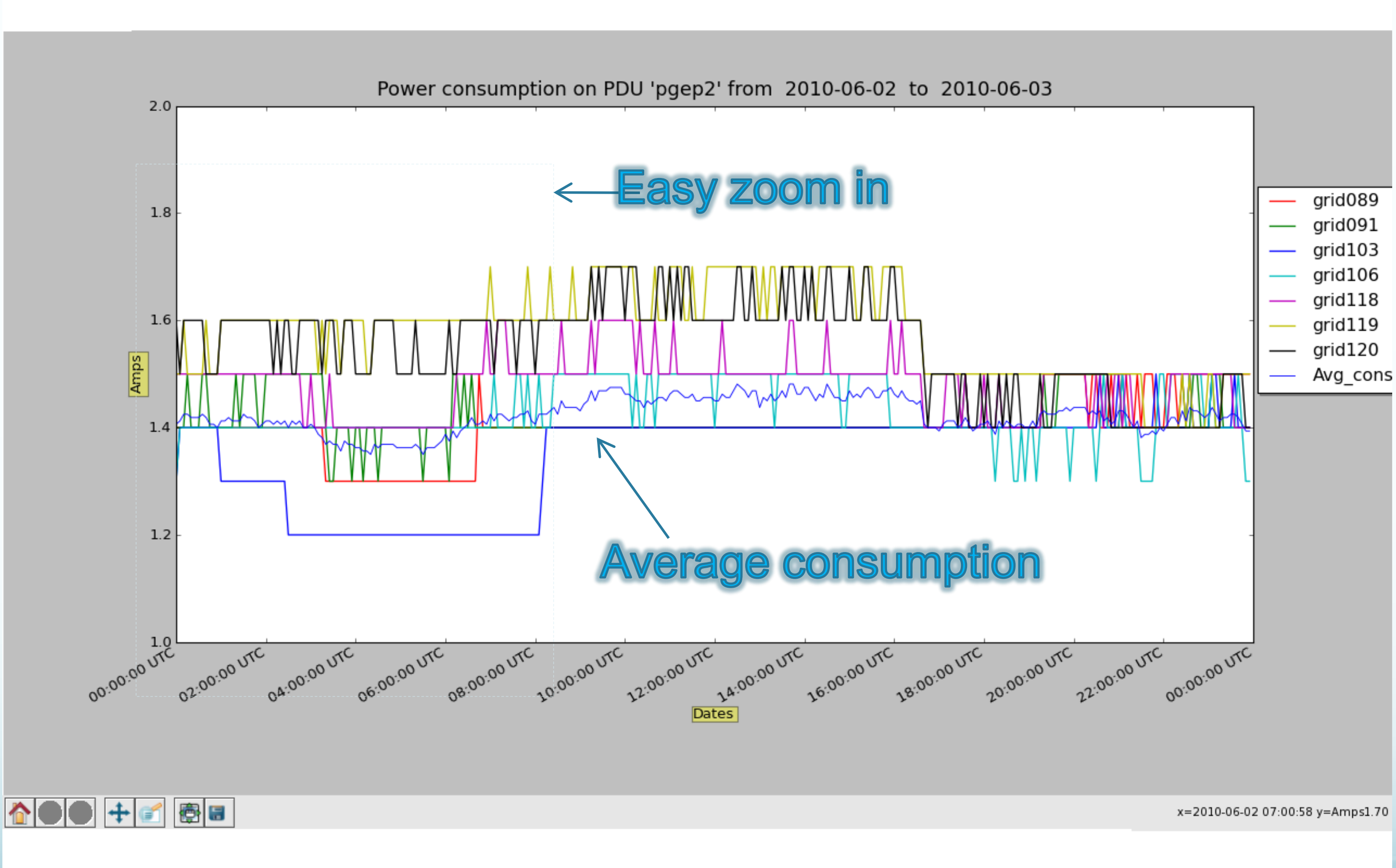
Source: http://www.netways.de/uploads/media/Werner_Fischer_-The-Power-Of-IPMI.pdf

slide 6/26

PowerMon Prototype

- A set of tools to collect and visualize the data about individual machine power consumption and load
- Written in Python, using SNMP for power data acquisition
  - Easy to extend for supporting new PDU HW
  - IPMI-based data acquisition to be added soon
  - Machine load retrieved from RRD tools DB generated by Ganglia, Nagios or other load monitoring tools
  - Consolidated data stored in a SQL db with a fixed sampling interval (currently 5 mn)
- Visualization for exploring correlations between load and power data

PowerMon Visualisation

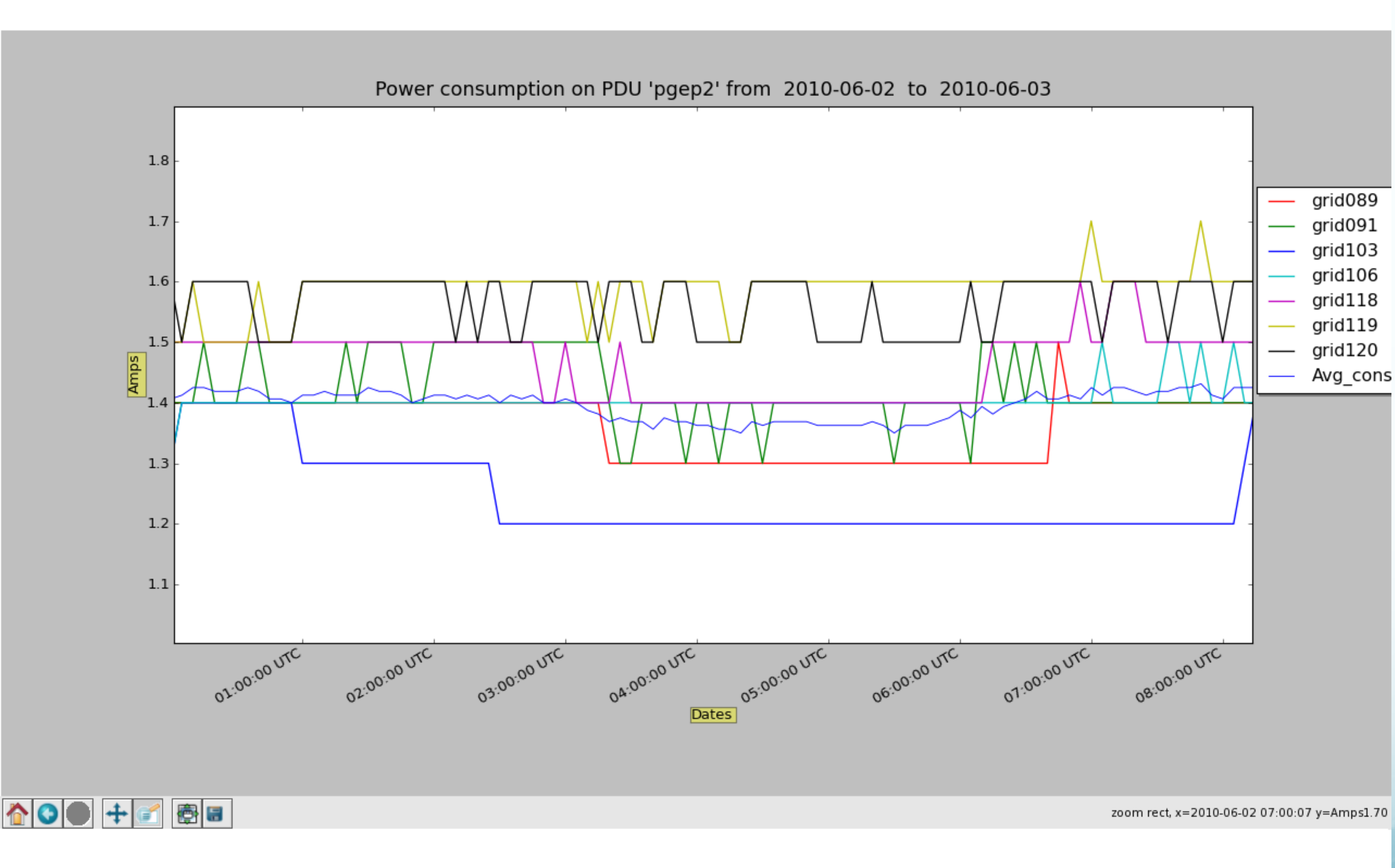


x=2010-06-02 07:00:58 y=Amps1.70

Easy zoom in

Average consumption

PowerMon Visualisation



Zoommed results

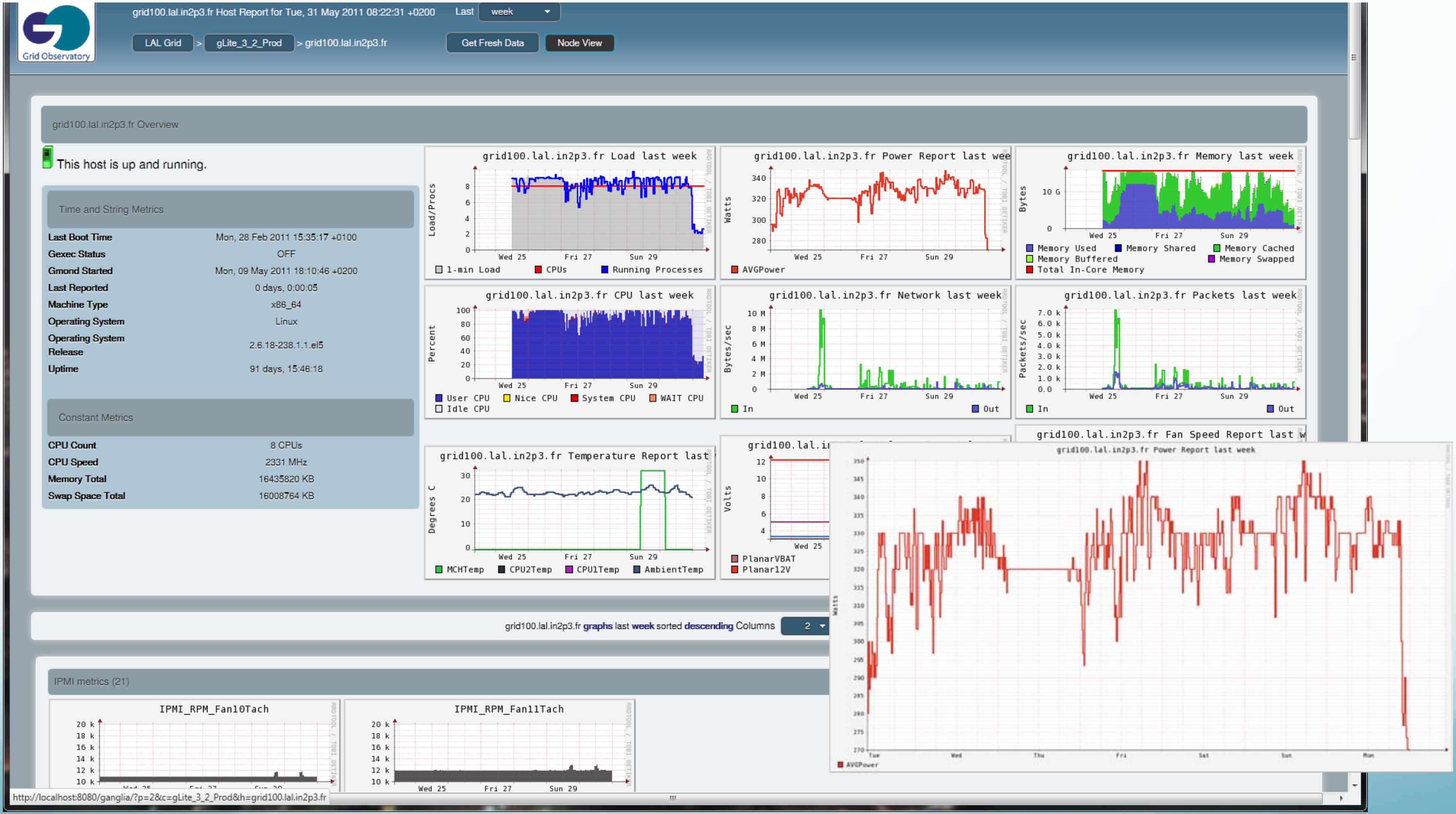
Status and Roadmap…

- Currently monitoring 1 rack through PDU and 8 through IPMI
  - 200 IBM 3550 (1600 cores) and in 5 Dell C6100 (400 cores)
  - Focus on assessing IPMI reliability
  - Collecting 400MB/day with a sampling interval of 5 mn
  - Data available: power consumption/machine, CPU load
- Short term plans (funding by CNRS PEPS)
  - PDU-based acquisition for Dell C6100 systems (Twin^2)
  - Collect information about global power consumption, ambiant temperature, fan speeds
    - Cooling inefficiency leads to increased fan speed which leads to +20% in power consumption
  - Integration of IPMI-based acquisition into PowerMon

… Status and Roadmap

- Visualisation: integration of power consumption into standard monitoring tools like Ganglia
  - Mostly a matter of producing RRD files
  - A prototype produces RRD files directly, could also be derived from PowerMon SQL DBs
- Data export to a common agreed format
  - Probably XML-based
  - Aim should be comparison between sites
  - Target date : January 2012
- Open questions: do we need motherboad and CPU temperatures

Ganglia-based Visualisation



grid100.lal.in2p3.fr Host Report for Tue, 31 May 2011 08:22:31 +0200 Last week

LAL Grid > gLite_3_2_Prod > grid100.lal.in2p3.fr

Get Fresh Data | Node View

grid100.lal.in2p3.fr Overview

This host is up and running.

Time and String Metrics

Last Boot Time	Mon, 28 Feb 2011 15:35:17 +0100
Gexec Status	OFF
Gmond Started	Mon, 09 May 2011 18:10:46 +0200
Last Reported	0 days, 0:00:05
Machine Type	x86_64
Operating System	Linux
Operating System Release	2.6.18-238.1.1.el5
Uptime	91 days, 15:46:18

Constant Metrics

CPU Count	8 CPUs
CPU Speed	2331 MHz
Memory Total	16435820 KB
Swap Space Total	16008764 KB

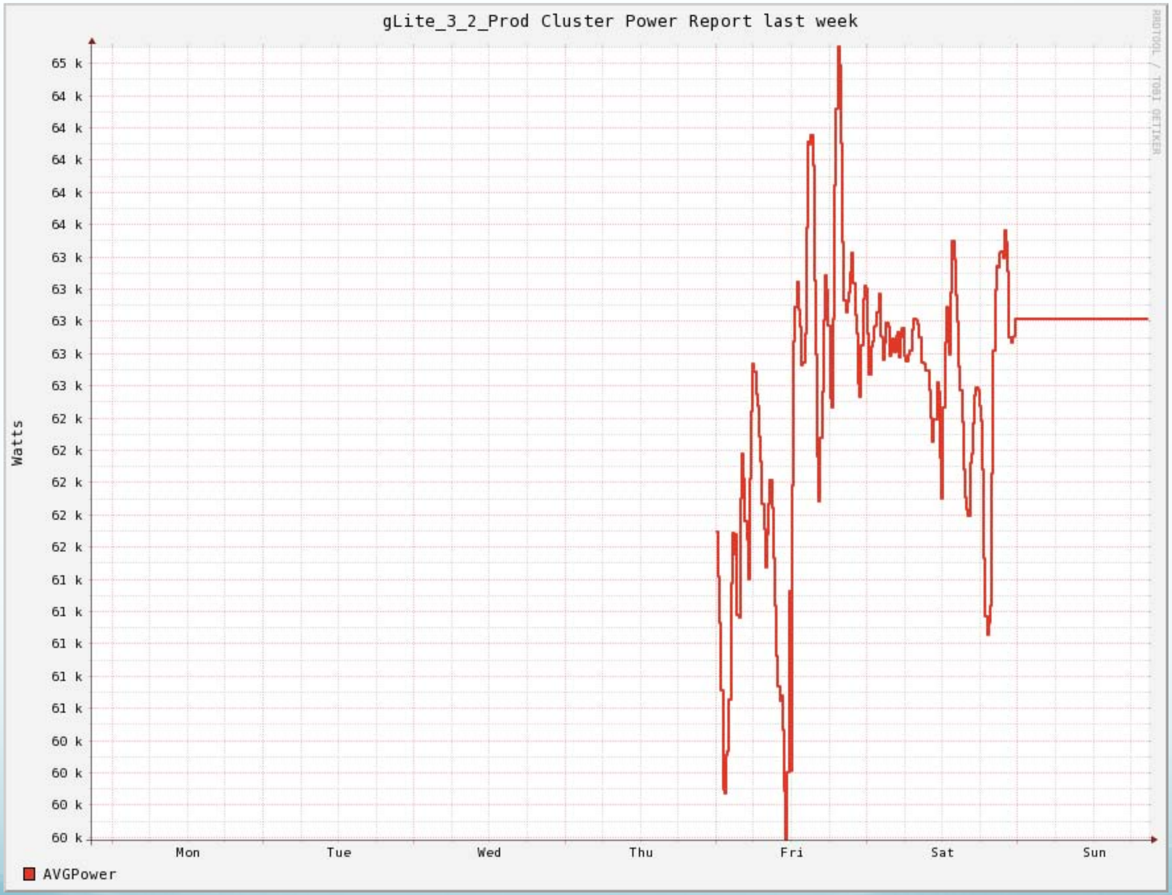
grid100.lal.in2p3.fr graphs last week sorted descending Columns 2

IPMI metrics (21)

http://localhost:8080/ganglia/?p=2&c=gLite_3_2_Prod&h=grid100.lal.in2p3.fr

Ganglia-based Visualisation

- But also consolidation at cluster level



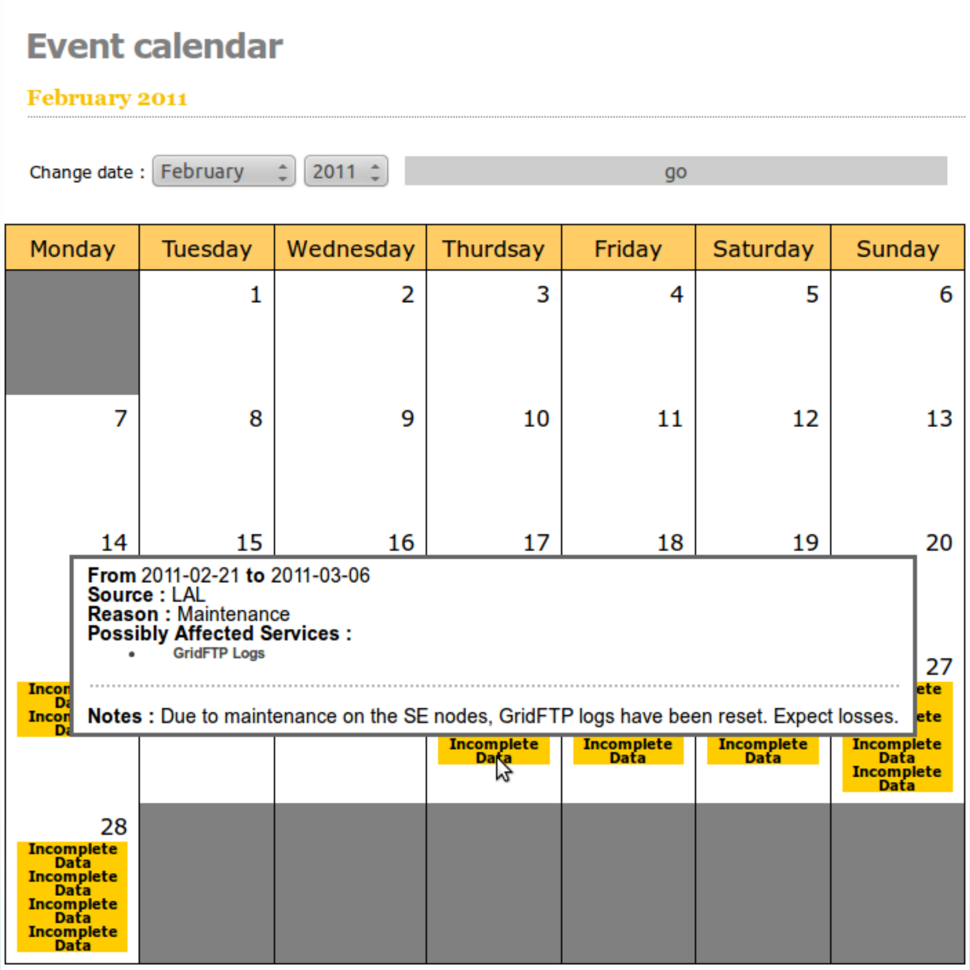
Data Curation…

- *Digital curation is the selection, preservation, maintenance, collection and archiving of digital assets* [Wikipedia]

- An important feature is to eliminate obvious outliers
  - Difficult, mostly a manual process
  - Importance of annotations (metadata)

- First implementation is based on an annotated calendar of known operational events
  - GRIF events are published by GRIF in a Google Calendar for its internal use: important for its accuracy
  - Google calendar is imported in a SQL DB and allows event annotation

… Data Curation



Event calendar

February 2011

Change date : February 2011 go

Monday	Tuesday	Wednesday	Thurdsay	Friday	Saturday	Sunday
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
Incon Da Incon Da			Incomplete Data	Incomplete Data	Incomplete Data	27 ete ete Incomplete Data Incomplete Data
28 Incomplete Data Incomplete Data Incomplete Data Incomplete Data						

From 2011-02-21 **to** 2011-03-06
Source : LAL
Reason : Maintenance
Possibly Affected Services :

- GridFTP Logs

Notes : Due to maintenance on the SE nodes, GridFTP logs have been reset. Expect losses.

Metrics, Measures and Models

- First step: behavioral descriptive models i.e. parsimonious representations from the large dimension space available from the detailed monitoring
  - Stationarity should not be assumed -> detection of ruptures
  - On-line, dynamic clustering with GStrAP
- Next: identify optima in the resulting complex landscape
- Requires the developement of a framework for automated analysis, in particular data correlations/clustering
  - 200+ systems!

Ontologies

- A requirement for data analysis and correlation
- Characterization of processes, services and collections do exist to model computational usages.
- These concepts are integrated in the ontological resources of the OntoSpec method defined by MIS.
  - They are linked to an ontology of Quantities and Units of Measure

Conclusions

- The GCO is build upon the Grid Observatory experience in grid behavioral data collection and publishing
  - Participates to the trend to Open Data
  - GCO is a task in Cloud benchmarking Activity Proposal for ICTLabs 2012
- GCO started a prototype for data collection at GRIF/LAL production grid site
  - Collection tool available and easy to extend to new HW
  - IPMI will be used for data collection extension to the whole site
    - Required for a fine enough granularity with Twin^2 systems
- We are willing to collaborate with “green computing” community and are open to community requirements

Useful Links

- Grid Observatory: http://www.grid-observatory.org/
- GRIF: http://grif.fr
- StratusLab: http://stratuslab.eu
- IPMI:
  - http://www.netways.de/osdc/y2010/programm/v/the_power_of_ipmi/
- OntoSpec : construction of ontologies
  - http://www.laria.u-picardie.fr/IC/site/?lang=en