

CIMENT et les grilles

Bruno Bzeznik

CIMENT - University of Grenoble

april 2012

CIMENT et les grilles

- 1 Introduction
- 2 CIGRI
 - La grille de CIMENT
 - OAR et le besteffort
 - L'intergiciel CiGri
 - Exemples
- 3 La grille de stockage de CIMENT
- 4 Ragrid
- 5 Grid5000

Outline

- 1 Introduction
- 2 CIGRI
 - La grille de CIMENT
 - OAR et le besteffort
 - L'intergiciel CiGri
 - Exemples
- 3 La grille de stockage de CIMENT
- 4 Ragrid
- 5 Grid5000

Introduction

Qu'est-ce qu'une grille de calcul ?

- Dans notre contexte (HPC), une grille de calcul permet d'accéder à des ressources de calcul dans une infrastructure mettant en oeuvre un grand nombre de machines **hétérogène** et/ou **délocalisées**.



Introduction

- Les grilles peuvent être très différentes (calcul, stockage, expérimentation, peer-to-peer, desktop-computing,...)
- Les grilles n'ont souvent pas d'administration centralisée
- Pas de vision unitaire des ressources de calcul, mais plutôt une vision globale, agrégée. En ce sens, ce ne sont pas des **clusters** (ou grappes)
- Les grilles sont gérées avec un **middleware** (ou intergiciel) qui permet aux utilisateurs de lancer des **jobs grille**.

Outline

- 1 Introduction
- 2 CIGRI
 - La grille de CIMENT
 - OAR et le besteffort
 - L'intergiciel CiGri
 - Exemples
- 3 La grille de stockage de CIMENT
- 4 Ragrid
- 5 Grid5000

- 1 Introduction
- 2 CIGRI
 - La grille de CIMENT
 - OAR et le besteffort
 - L'intergiciel CiGri
 - Exemples
- 3 La grille de stockage de CIMENT
- 4 Ragrid
- 5 Grid5000

CIGRI

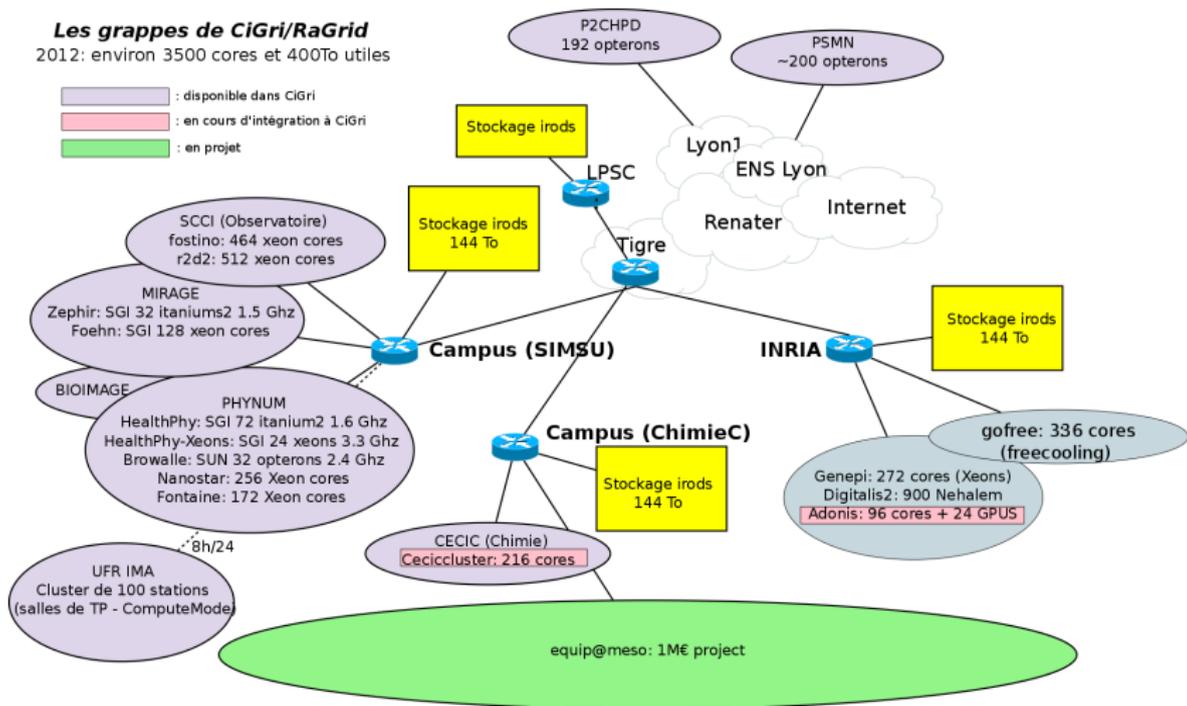
- CIGRI est la **grille locale** du mesocentre CIMENT
- CIGRI est aussi le nom de l'**intergiciel (middleware)** qui gère la grille locale de CIMENT

CIGRI : la grille de CIMENT

Les grappes de CiGrid/RaGrid

2012: environ 3500 cores et 400To utiles

- : disponible dans CiGrid
- : en cours d'intégration à CiGrid
- : en projet



CIGRI : la grille de CIMENT

Current CIMENT hardware resources

Name	Brand	Number of cpu cores	Total memory	Max memory/node	Total storage (net)	Computing network	Total Gflop/s	Buy date
Fostino	IBM	464	464 GB	8 GB	27.5 TB	Gigabit ethernet	5196.8	2008-09-01
R2d2	IBM	512	1088 GB	32 GB	19.24 TB	Infiniband DDR	5120	2008-09-01
Genepi	Bull	272	272 GB	8 GB	5.44 TB	Infiniband DDR	2720	2008-10-10
Nanostar	SGI	256	512 GB	16 GB	7 TB	Infiniband DDR	2560	2009-01-01
Edel	Bull	576	1728 GB	24 GB	0 TB	Infiniband DDR	5230.08	2009-01-01
Adonis	Bull	96	288 GB	24 GB	0 TB	Infiniband DDR	3621.68	2010-01-01
Foehn	SGI	128	480 GB	48 GB	7 TB	Infiniband DDR	1367.04	2010-03-01
Global_storage	Dell		216 GB	24 GB	400 TB		0	2010-09-01
Fontaine	Dell	144	288 GB	24 GB	12 TB	Infiniband QDR	1307.52	2010-11-01
Gofree	Dell	336	2016 GB	72 GB	30 TB	Infiniband QDR	3177.6	2011-01-01
Ceciccluster	Dell	216	432 GB	24 GB	12.5 TB	Infiniband QDR	1961.28	2011-12-01

Total CIMENT resources

- Total number of cores: **3140**
- Total memory: **7.928 TB**
- Total disk space: **520.68 TB**
- Total op/s: **32.262 TFlop/s**

- 1 Introduction
- 2 CIGRI
 - La grille de CIMENT
 - **OAR et le besteffort**
 - L'intergiciel CiGri
 - Exemples
- 3 La grille de stockage de CIMENT
- 4 Ragrid
- 5 Grid5000

OAR

- OAR est le gestionnaire de ressources installé sur les machines de CIMENT
- OAR est développé au LIG (est issu de CIMENT et Grid5000 à l'origine)
- OAR offre un mode de fonctionnement intéressant : **best-effort**

Best-effort jobs

- Best effort jobs allow us to exploit idle resources of production clusters
- When a node is free, a best-effort job may be placed on it.
- If a "normal" job needs the node, the best-effort job is killed
- as a consequence, a best-effort job must be "short enough" (lt 30 minutes) or there's too much chance for it to be killed
- Checkpointing may help

- 1 Introduction
- 2 CIGRI
 - La grille de CIMENT
 - OAR et le besteffort
 - **L'intergiciel CiGri**
 - Exemples
- 3 La grille de stockage de CIMENT
- 4 Ragrid
- 5 Grid5000

CIGRI : l'intergiciel

- Une grille **légère**
- A la base, pour les jobs multiparamétriques uniquement
- Administration simple, bien adaptée à la taille d'un meso-centre
- Gère les jobs best-effort : re-soumission automatique
- Interface web pour les utilisateurs
- Prochainement (v3) : API pour applications tierces

Cigri efficiency (1/2)



The load of the clusters is not constant and peaks are often not at the same time...

Cigri efficiency (2/2)



CiGri uses the idle cpus

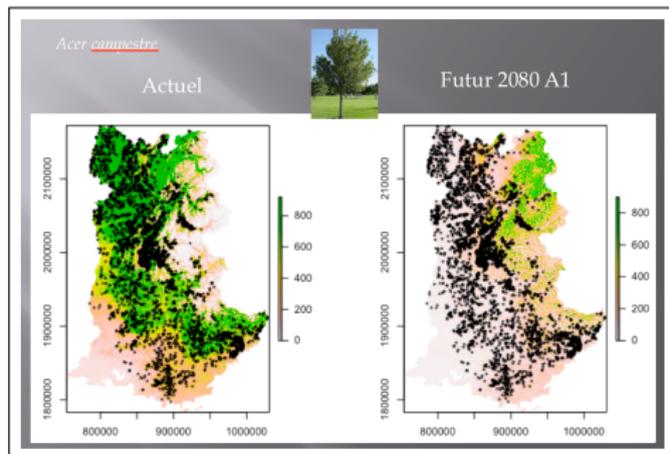
Alors CIGRI, c'est pour qui ?

- Vous avez de nombreux jobs séquentiels (des milliers, des millions,...), assez courts, avec éventuellement des paramètres qui varient. **ALORS CIGRI EST FAIT POUR VOUS !**
- Vous avez de nombreux jobs paramétriques, faiblement parallèles (dont le niveau de parallélisme ne met pas en jeu plus de 2 noeuds). **Alors Cigri peut vous aider**
- Vous avez de nombreux jobs paramétriques séquentiels longs. **Si vous pouvez gérer des points de reprise (checkpoint applicatif) alors Cigri peut vous aider**
- Vous avez "quelques" très gros jobs parallèles. **Cigri n'est pas fait pour ça.**

- 1 Introduction
- 2 CIGRI
 - La grille de CIMENT
 - OAR et le besteffort
 - L'intergiciel CiGri
 - **Examples**
- 3 La grille de stockage de CIMENT
- 4 Ragrid
- 5 Grid5000

Cigri exemples

Biologie : effet des changements climatiques sur la biodiversité européenne. Utilisation de R.
Campagnes de 3000 jobs (pour 3000 espèces dont l'absence ou la présence est étudiée). Un job correspond à l'application d'un modèle pour une espèce et de nombreuses campagnes sont réalisées en fonction des variables du milieu.



Wilfried Thuiller, Tamara Muenkemueller

Cigri exemples

Astro : Analyse interactive des données en provenance de l'équipement CONSERT de la sonde ROSETTA envoyée sur une comète. Développement d'un framework autour de l'API CIGRI.



(source : esa.int)

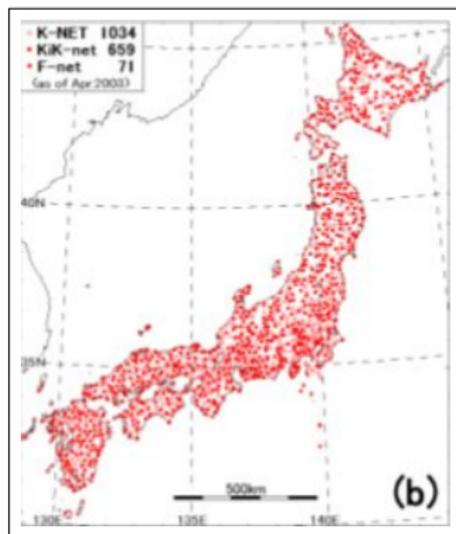
Yves Rogez, Philippe Bollard

CiGri examples

Geologie : Analyse des données sismiques du Japon sur 4 ans. ANR projet **Namazu**.

Most of the data of Japan of the recent years, including the Tohoku seism (Hi-Net, Tiltmeter and F-Net). Processing of sesimograms, computation of correlations, analysis of the results with the **Whisper project, Project FP7 ERC Advanced grant 227507**

The main reserve is the transfert of the data



Xavier Briand, Florent Brenguier, Michel Campillo, Philippe Roux, Nicolai Shapiro (IPGP and Isterre for France)

En savoir plus sur Cigri

Fiche plume "cigri"

Outline

- 1 Introduction
- 2 CIGRI
 - La grille de CIMENT
 - OAR et le besteffort
 - L'intergiciel CiGri
 - Exemples
- 3 La grille de stockage de CIMENT**
- 4 Ragrid
- 5 Grid5000

La grille de stockage de CIMENT

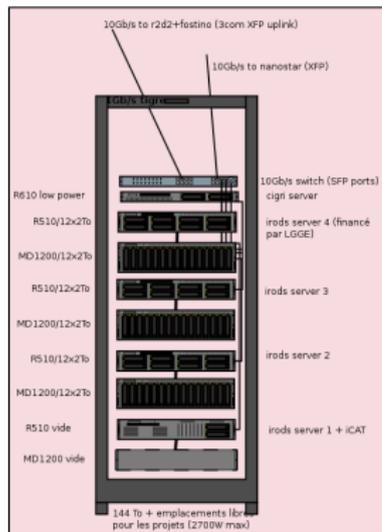
- Projet commun LPSC et CIMENT en 2010
- 288 TB CIMENT + 384 TB LPSC (brut)
- Données distribuées sur 4 sites, accessibles depuis tous les clusters de CIMENT à 10Gb/s ou 1Gb/s
- Gestion sous IRODS (<http://irods.org>)
- Extensible "à l'infini" : les labos peuvent contribuer à l'achat de noeuds ou de disques de stockage
- Offre des fonctionnalités depuis les clusters jusqu'aux desktops (par exemple : webdav)

La grille de stockage de CIMENT : contributions

- Initial : 288 TB brut
- Contrib LGGE 2011 : 48 TB brut
- Contrib LGGE 2012 : 48 TB brut + 18 TB brut
- Contrib ISTER 2012 : 48 TB brut
- **Total avril 2012 : 450 TB brut** (216 disques de 2TB et 6 disques de 3TB)

CIGRI storage hardware

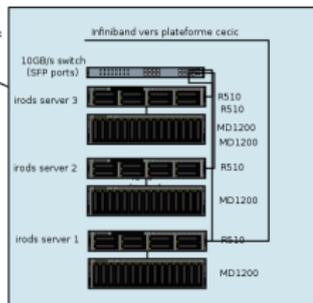
Stockage CIGRI Irods 04/2012: 450 TB (brut)



Consommations estimées (source Dell):
 R510+12 DD SAS: 395 W
 MD1200 + 12 DD SAS: 289 W

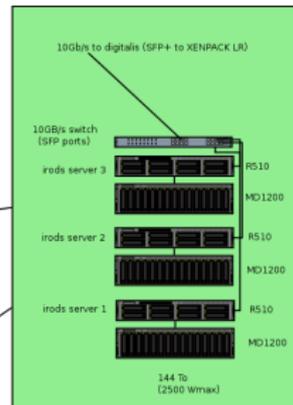
1Gb/s tigre

10Gb/s UJF



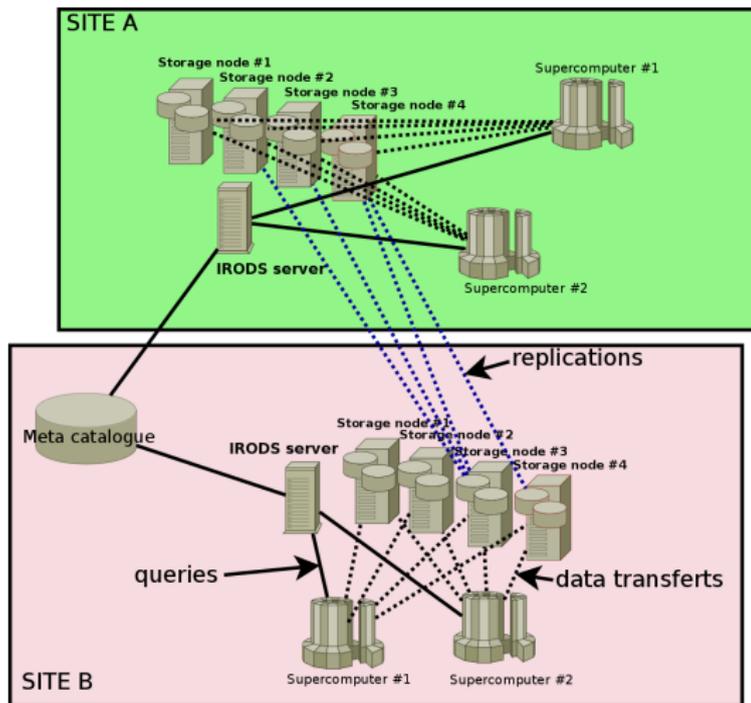
CECIC (Campus)

1Gb/s tigre



INRIA (Montbonnot)

IRODS Concept



IRODS

- Irods est un intergiciel de grille de stockage
- Ce n'est pas un système de fichiers !
- "On fait des put et des get"
- De nombreux moyens d'accès, dont une interface web
- Gestion de meta-données

IRODS Exemple

All the resources are seen as a unique user space

```
user@cluster1:~$ iput test_file
#
user@cluster2:~$ ils -l
/cigri/home/user:
user  0 smh-simsu-s2-cigri2  0 2010-12-06.16:43  test_file
user@cluster2:~$ iget test_file
```

IRODS replication

A file may be replicated (irepl command) for faster local access or for backup

```
user@edel-9:~$ ils -l test_dir
/cigri/home/user/test_dir:
  user  0 smh-simsu-s2-cigri2  0 2010-12-06.16:43  test
  user  1 mbt-inria-s1-cigri2  0 2010-12-06.17:19  test
```

IRODS Web interface

The screenshot shows a web browser window displaying the IRODS web interface. The address bar shows the URL: `https://ciment-grid.ujf-grenoble.fr/irods/browse.php#ruri=bzizou.cigr@ci...`. The interface includes a navigation pane on the left with a tree view of folders, a main content area with a file list, and a modal window for file details.

File List:

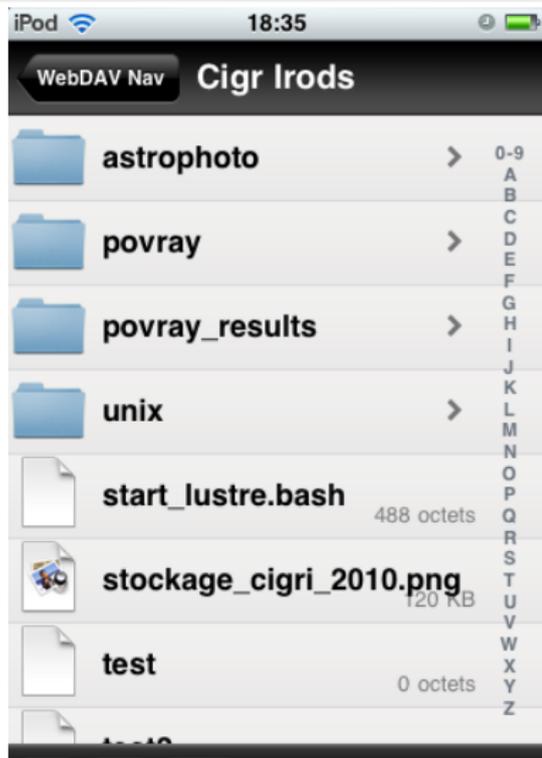
Name	Resource	Size	Date Modified
landscape_7323.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7326.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7329.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7332.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7325.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7328.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7313.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7316.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7318.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7319.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7322.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7317.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7315.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7284.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7293.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7297.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7309.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm
landscape_7312.png	smh-simsu-s2-cigr1	132.28 KB	December 7, 2010, 6:37 pm

File Details Modal (landscape_7319.png):

- Force Download**
- Size:** 132.28 KB (135453 Bytes)
- RODS URI:** bzizou.cigr@ci... /povray_results/landscape_7319.png
- Resource:** smh-simsu-s2-cigr1
- Type:** generic
- Date Modified:** Tue Dec 07 2010 18:37:32 GMT+0100 (CET)
- Date Created:** Tue Dec 07 2010 18:37:32 GMT+0100 (CET)

Buttons: overview | metadata | Copies | More

IRODS DAV access



IRODS et CIGRI

- Irods est évidemment le complément idéal de Cigri
- Récupération des résultats "au fil de l'eau"
- L'accès aux données devient transparent, quel que soit le cluster qui les produit ou qui y accède
- Permet aussi le déploiement des binaires via le "prologue"

Outline

- 1 Introduction
- 2 CIGRI
 - La grille de CIMENT
 - OAR et le besteffort
 - L'intergiciel CiGri
 - Exemples
- 3 La grille de stockage de CIMENT
- 4 **Ragrid**
- 5 Grid5000

Grilles en Rhone-Alpes

Ragrid est un groupe de travail régional, au sein de CIRA, pour collaborer au niveau des usages et des développements sur les grilles de notre région.

- Cigri (Grenoble)
- Diet (ENS Lyon)
- Tidra (IN2P3)

Outline

- 1 Introduction
- 2 CIGRI
 - La grille de CIMENT
 - OAR et le besteffort
 - L'intergiciel CiGri
 - Exemples
- 3 La grille de stockage de CIMENT
- 4 Ragrid
- 5 **Grid5000**

CIMENT et Grid5000

CIMENT a de nombreuses interactions avec Grid5000

- Le site grid5000 de Grenoble peut recevoir des jobs cigri
- Adonis : plateforme GPU accessible aux utilisateurs de CIMENT
- Développement de OAR
- Développement de Cigri v3
- Echange d'expertises