

Genocluster

Haute Performance pour le traitement des données biologiques

24 octobre 2006

Anthony Assi

Anthony.Assi@irisa.fr



- **Présentation du personnel**
- **Présentation de l'architecture du Cluster**
- **Présentation du portail Web et des différents outils**
- **Cas pratique de l'utilisation des machines**

Les Ingenieurs Experts:

• **Anthony Assi** (Anthony.Assi@irisa.fr) :

Systeme et Développement

• **Laetitia Guillot** (Laetitia.Guillot@irisa.fr) :

Recherche et découverte de motifs

• **Olivier Filangi** (Olivier.Filangi@irisa.fr) :

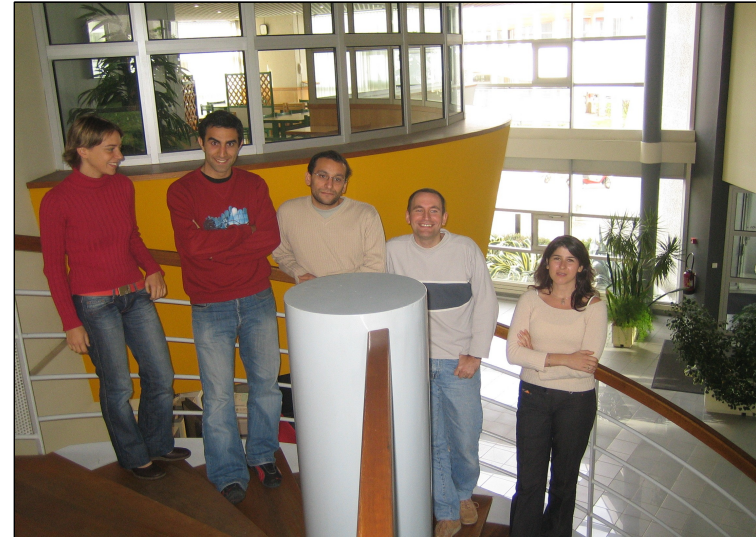
Développement Informatique

• **Gregory Ranchy** (Gregory.Ranchy@irisa.fr) :

Développement bio-informatique

• **Sophie Roucan** (Sophie.Roucan@irisa.fr) :

Logithèque, Web et Communication



Les Anciens Ingenieurs Experts:



Emmanuelle Morin
(Novembre 2005)



Anne-Sophie Valin
(Mai 2006)

Architecture matérielle

- Une ferme de calcul de 32 noeuds, chaque noeud est composé de 2 processeurs AMD Opteron 64 bits et dispose de 4Go de mémoire vive



- Deux serveurs Web redondants (Sun V40z), pour assurer la haute disponibilité



NOUVEAU

- Baie disque de 8TO pour les Banques de données

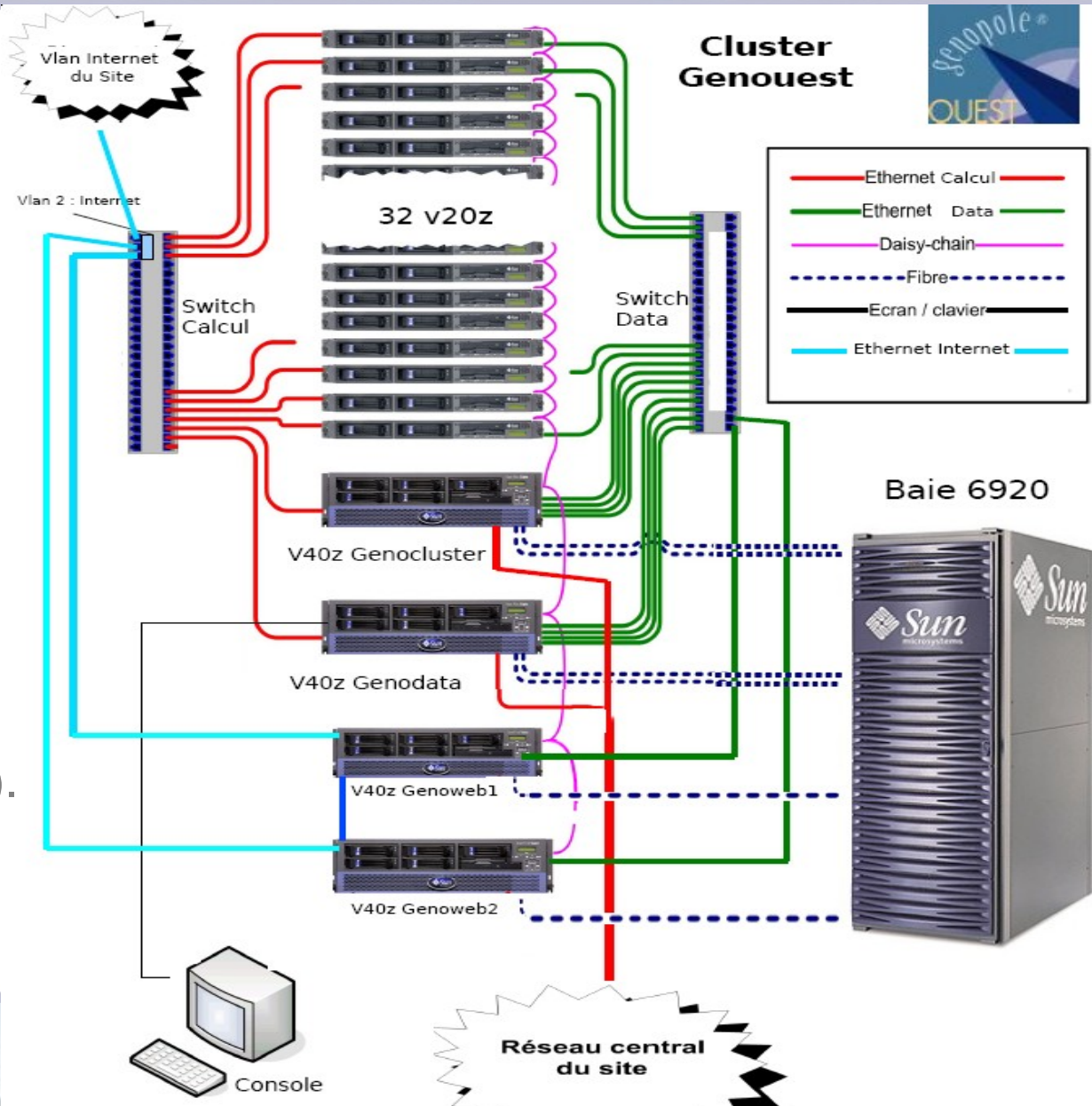
- Serveur de Stockage Sun StorageTek 6920 de 6 TO

- Robot de sauvegarde Sun StorEdge L180



Architecture Reseaux:

- 1 subnet pour l'échange des données
- 1 subnet pour le calcul
- 1 réseau privé pour l'accès au SP de chaque machine.
- Plusieurs Sous-Réseaux Virtuels (VPN).



Serveurs Web

Resource	genoweb1	genoweb2
apache-genoweb	10 Active	1 StandBy
ip-genoweb	10 Active	1 StandBy
mysql	10 Active	1 StandBy
ip-genoweb-int	10 Active	1 StandBy
web	10 Active	1 StandBy
disk24950	10 Active	1 StandBy
web	10 Active	1 StandBy
disk24950	10 Active	1 StandBy

<-- genoweb2: Updating server state to: alive



- 2 Serveurs Web à configuration identique (Apache/MySQL)
- Exécution des applications Web sur les noeuds de calcul, et non sur le serveur Web.
- 2 Disques en Raid 1 (OS)
- 3 Disques en Raid 5 (Data)
- 2 Connexions FibreChannel à la baie de stockage
- 3 addr IP Publique, dont 1 virtuelle
- 3 addr IP Privée, dont 1 virtuelle

Pour plus d'informations:

webmaster@genouest.org



Welcome on GenQuest bioinformatics platform

Welcome on the OUEST-genopole® bioinformatics server

This platform aims at proposing **innovative software in Bioinformatics**. You will find here a very complete set of tools in bioinformatics, public databanks updated on a daily basis, and a range of links to seminars, training courses, platform news (content also available as an [RSS](#) newsfeed) ...

For a direct access to the tools, please use thumbnail "tools". You will access :

- Exclusive tools** such as [STAN](#) or [Wapam](#) (under heading "Pattern matching"), [GenoFrag](#) (under heading "Primers design") and [Domain Organizer](#) (under heading "Modulome").
- Specialized relational databases such as** [MIPDB](#) ...

The Search toolbox (up right) provides an easy access to documents and tools available on this site.

If you wish to give access to your software (possibly a prototype without interface) or on the contrary, if you are looking for some functionalities and do not find any practical tool, please use the [FAQ](#).

We propose you to subscribe to the bioinformatics platform's [mailing-list](#) : it will allow you to know news about software updates, events or trainings.

[Register](#) to the **bioinformatics platform fourth workshop - 24th October 2006**. (See the [program](#))

Platform events

OUEST-genopole® bioinformatics platform fourth workshop

Program and registration form for this workshop on Tuesday October 24th 2006 are available.

[Show more...](#)

Platform news

20 October 2006 : Mailing list creation

The bioinformatics platform proposes you to subscribe to its mailing list.

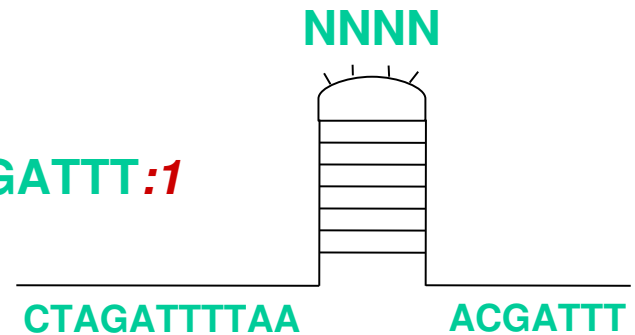
[Read more](#)

- Le cluster, héberge différents outils accessibles via le site Web
- Recherche et découverte de motifs: STAN, WAPAM, PRATT
- Wisconsin Package – EMBOSS
- Recherche de similarités : BLAST, mpiBlast
- Découpage de séquences biologiques : GenoFrag, DomainOrganizer
-

- Outil de recherche motifs complexes dans les génomes

Exemple de motif complexe :

CTAGATTTTAA:2 - X:[7] - x(4) - ~X:5 - ACGATTT:1



- Indexation des génomes sous forme d'arbre des suffixes
- STAN 200x plus rapide sur genoweb que sur Idefix (Ancien serveur)
- 1 noeud de Genocluster plus puissant que idfix avec ses 16 procs.

Pour plus d'informations sur STAN :

<http://www.irisa.fr/symbiose/STAN>

- Version parallélisée du BLAST de NCBI.
- Gratuite et open-source
- mpiBLAST segmente les indexes des banques de données utilisés par BLAST et les distribue sur les noeuds du cluster
- mpiBLAST permet une exécution des requêtes BLAST sur différents noeuds simultanément
- mpiBLAST plus rapide que BLAST




Plate-forme
Bio-informatique
GENOUEST





Accueil FAQ Plan Contacts English Imprimer

La plateforme Outils Banques Séminaires Formations Aide


Sont décrites dans cette page toutes les banques de séquences disponibles sur notre serveur ainsi que les outils qui y sont liés. Sont également présentées les bases de données spécialisées réalisées par nos ingénieurs.

Vous êtes ici : [Accueil](#) > [Banques](#) > [Banques publiques](#) > [Banques nucléiques](#)

Banques

-  **Projet BioMaj**
- Informations sur BioMaj
-  **Banques publiques**
- Banques nucléiques
- Banques protéiques
- Banques de motifs protéiques
- Banques d'enzymes
-  **Génomes complets**
- Génomes eucaryotes
- Génomes bactériens
- Génomes archaïes
-  **TRANSFAC®**
- Demande d'accès TRANSFAC®
- Accès

Banques nucléiques

 **Banques nucléiques**

Name	Version
GenBank	153.0 (04/2006)
EMBL	86.0 (03/2006)
Imgt	200606-6 (01/2006)
Imgt-HLA	2.12 (01/2006)
GoldenH	17 (NCBI build 35) (11/2004)
GoldenM	5 (NCBI build 33) (11/2004)
Flybase*	7.02 (10/2002)
Unigene	182 (04/2005)

Last update of this informations : April 25 2006 11:49.

Bases de données spécialisées

Ces bases sont en accès restreint.
Pour les consulter, nom d'utilisateur : **public** et mot de passe : **public**.

- Xénope**
- Semences en phase de germination**
- Huitre**
- MIPDB**

Copyright © 2005 QUEST-genopole®. Mentions Légales Contact : webmaster@genouest.org

Les Banques de données

- Projet **Biomaj** (Olivier Filangi)
- Rapatriements de différentes banques de données
- Mise à jour régulières et automatisées
- Formatage des banques sous différents formats
- Automatisation des workflows de mise à jour
- ...

Démarche à faire pour avoir accès:

- Remplir la demande d'ouverture de compte

<http://genouest.org/acces.php>

- Connexion a distance au serveur:

```
$ssh -X username@genocluster.univ-rennes1.fr
```

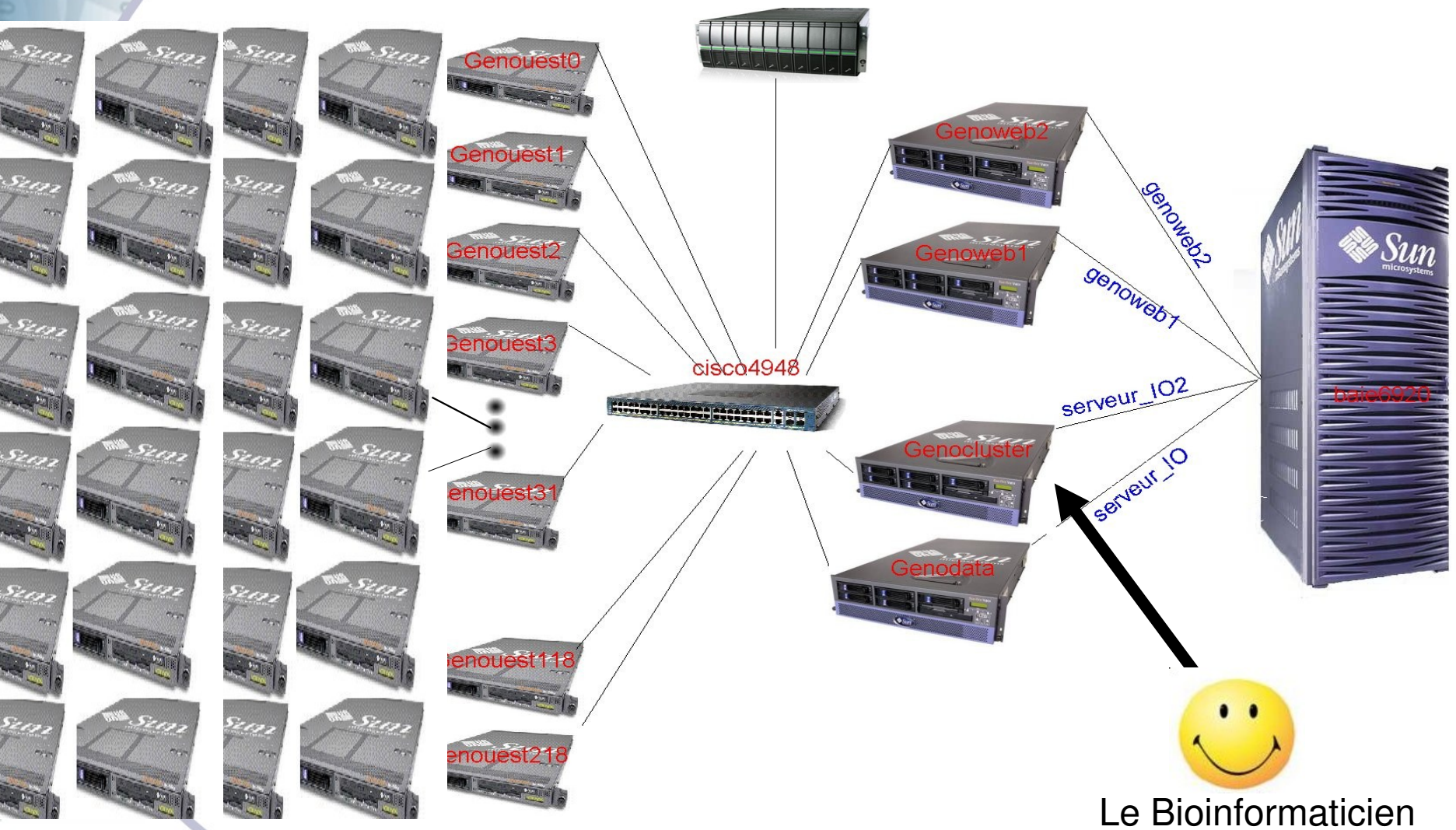


Ne pas executer des programmes sur le noeud Genocluster,
mais plutot passer par SGE (Système de gestion de charge
et de distribution des taches) !

Demande d'aide :

admin-genocluster@irisa.fr

Point d'entrée:



Le Bioinformaticien

Interactif :

- [aassi@genocluster ~]\$ **qrsh**

Last login: Tue May 9 08:07:11 2006 from genocluster

[aassi@genouest22 ~]\$

- Initialisation des environnements:

[aassi@genouest22 ~]\$ **source /local/env/envjava**

- Lancer l'application:

[aassi@genouest22 ~]\$ **java MonAppli**

Interactif vs Batch:

Batch:

```
[aassi@genocluster home]$qsub calcule_pi -o fichier.resultat -e fichier.erreur
```

```
[aassi@genocluster home]$qstat
```

job-ID	prior	name	user	state	submit/start at	queue	slots	ja-task-ID
--------	-------	------	------	-------	-----------------	-------	-------	------------

117	0.00000	calculer_p	aassi	qw	05/09/2006 10:29:41		1	
-----	---------	------------	-------	----	---------------------	--	---	--

```
[aassi@genocluster home]$qstat
```

117	0.55500	calculer_p	aassi	r	05/09/2006 10:29:51	batch.q@genouest1-data		1
-----	---------	------------	-------	---	---------------------	------------------------	--	---

```
[aassi@genocluster home]$qstat
```

```
[aassi@genocluster home]$cat fichier.resultat
```

```
3.14159265358979323846264338327950288419716939937510582097494459230781640628628  
9986280...
```

Questions ?



Presentation disponible à l'adresse:

http://genoweb.univ-rennes1.fr/Serveur-GPO/presentations/20061024_genocluster.pdf

Les Ingénieurs vous accueillent !

