

Conseil *Scientifique* des *Utilisateurs*



1. **Présentation générale de la plateforme**
 1. Equipe
 2. Missions
 3. Caractéristiques
2. Point sur les ressources fournies par la plateforme
 1. Matériel
 2. Outils
 3. Bases de données
 4. Formations
 5. Conception et développement
3. Points divers
 1. Certification ISO9001
 2. Galaxy
 3. Cluster
 4. Stockage
 5. Achats

Présentation générale de la PF

Equipe :

- ✓ **Jean-François GIBRAT** : Responsable scientifique. DR, **0.2 ETP**
- ✓ **Franck SAMSON** : Responsable technique. IR, **1 ETP**
- ✓ **Eric MONTAUDON** : Administrateur Système Linux. AI, **0.9 ETP**
- ✓ **Véronique MARTIN** : Ingénieur Outils et Banques de données. AI, **0.5 ETP**
- ✓ **Annie GENDRAULT** : Ingénieur Bases de données. IR, **0.5 ETP**
- ✓ **Sandra DEROZIER** : Ingénieur Conception et Développement. IE, **1 ETP**

Total : ~4,1 ETP

- ✓ **Point de vigilance 2014 : Fusion MIAj et MIG.**

Présentation générale de la PF

Missions :

3 types de missions :

- ✓ Infrastructure pour le calcul scientifique
 - Déploiement d'une **infrastructure de calcul scientifique**.
 - Mise à disposition des **collections génériques de données** issues des travaux de génomique.
 - Déploiement sur la plateforme de **logiciels d'analyse** de ces données.

- ✓ Savoir faire en bioinformatique
 - **Formation** des utilisateurs aux méthodologies de la bioinformatique et aux logiciels d'analyse.
 - **Assistance, conseil et expertise** aux utilisateurs
- ✓ Conception et développement d'applications
 - Développement d'outils de **visualisation**
 - Mise en place de **pipeline** pour automatiser certains traitements de données.

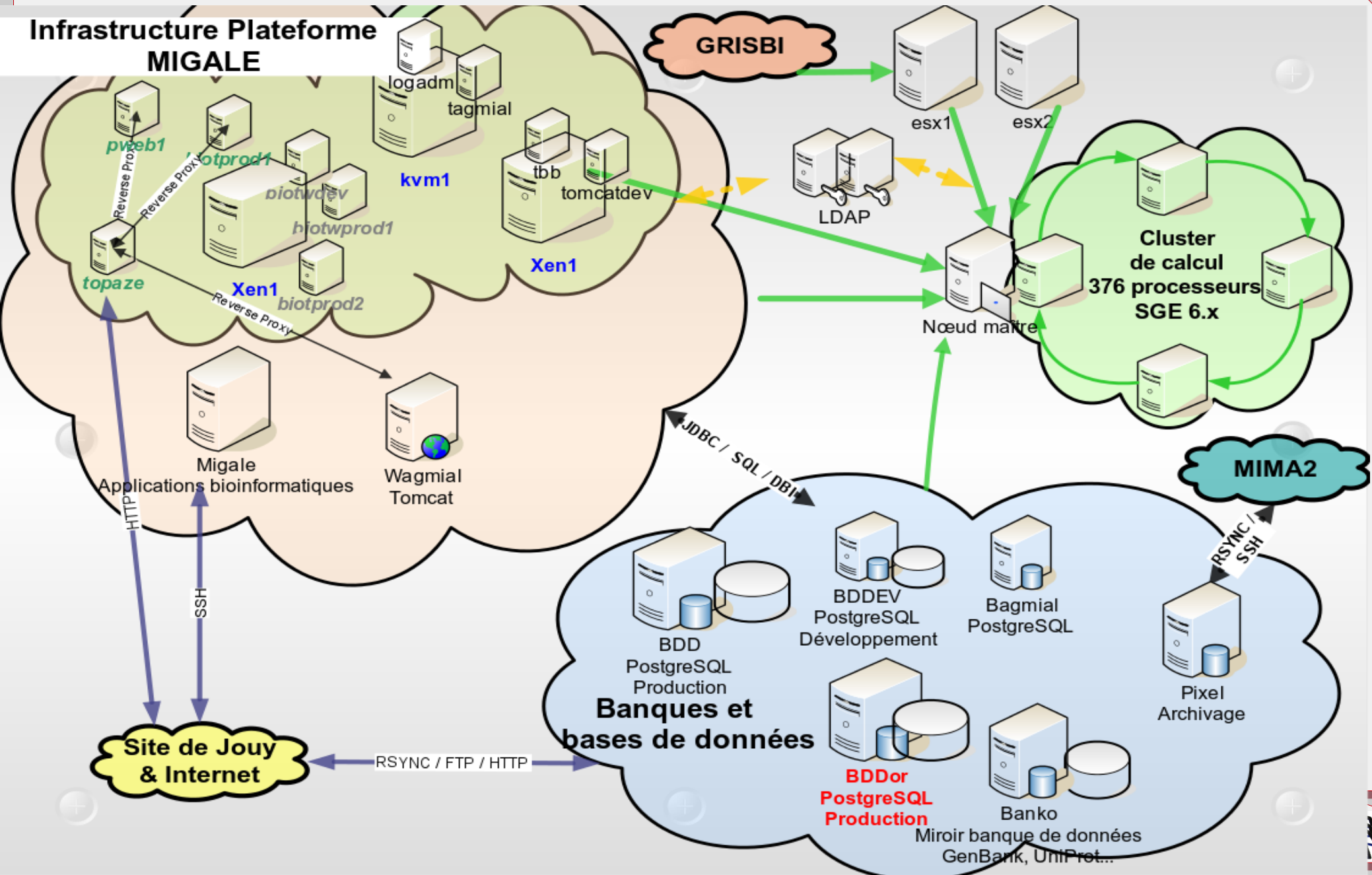
Présentation générale de la PF

Caractéristiques :

- Plus de 380 utilisateurs identifiés
 - 131 créations de compte
 - ~80 utilisateurs supplémentaires du cluster
 - 95 utilisateurs Galaxy
- RIO Opérationnelle depuis 2006, labelisée IBISA en 2010
- Membre de ReNaBi, du pôle régional ApliBio et de l'IFB
- Certifiée ISO9001 en Avril 2011
- Plateforme Stratégique CNOC depuis 2011, renouvelé en 2013.
- Appartient au CATI MIAGO, Projet de développement du portail IGO.

Ressources fournies par la PF

Infrastructure Plateforme MIGALE



Conseil *Scientifique* des *Utilisateurs*



1. Présentation générale de la plateforme
 1. Equipe
 2. Missions
 3. Caractéristiques
2. Point sur les ressources fournies par la plateforme
 1. Matériel
 2. Outils
 3. Bases de données
 4. Formations
 5. Conception et développement
3. Points divers
 1. Certification ISO9001
 2. Galaxy
 3. Cluster
 4. Stockage
 5. Achats

Travaux et incidents majeurs

- **Travaux :**
 - Ajout de fibres optiques dans notre salle
 - Modification de l'infrastructure reseau
- **Incidents**
 - *Probleme Climatisation*
 - *Probleme de memoire sur Banko*
 - *Reboot de migale => freeze (4x)*
 - *Reboot de noeuds => freeze*

Les Outils

Outils :

BRCLims logiciel de gestion de collections de ressources biologiques

RefComp Referentiel de compétences du département MIA

YeastIP Phylogenie et taxonomie pour la levure

basysbio Hébergement du projet basysbio

AnovArray Analyse de données d'expression (microarray et macroarray)

Class2g Ségrégation des gènes en deux groupes selon leur expression

DOMIRE Identification de domaines protéiques à partir de récurrences

FROST Reconnaissance de repliements

ISLAND Simulation de progrès d'un projet de cartographie physique de génomes

KAKSI Assignation de la structure secondaire des protéines

LP2Asium Analyse syntaxique de composants indépendants

Les Outils

Outils :

IGO	Portail des ressources développées au sein de MIG
MuGen	Outil interactif d'exploration dans plusieurs génomes annotés
OSS-HMM	Prédiction de la structure secondaire des protéines selon 3 états : H, b et C
PCM	Méthode de corrélation par paire
R'HOM	Recherche de régions homogènes dans les séquences ADN
R'MES	Recherche de mots exceptionnels dans les séquences ADN
SHOW	Recherche de régions homogènes dans les séquences ADN
•	
SIMPA	Prédiction de la structure secondaire des protéines selon 3 états : H, b et C

Les bases de données

Bases de données :

MICADO	Gestion de génomes microbiens et de données d'analyse fonctionnelle de <i>Bacillus subtilis</i>
MOSAIC	Comparaison de génomes bactériens proches
PAREO	Intégration des connaissances sur les voies métaboliques issues de la base japonaise Kegg
PARIS	Analyses protéomiques
PFAM	Alignements multiples de séquences et domaines protéiques
PROSE	Version relationnelle de la banque de données UNIPROT
SPiD	Interactions protéine-protéine de <i>Bacillus subtilis</i>

Formations

Conception et développement

Conception et Développement :

- Projet développement :
 - Projet de portail pour l'unité, première version disponible depuis Janvier 2013, une nouvelle version est en cours de développement.
 - Interfaçage d'outils dans les environnements Mobylye et Galaxy
 - Projet de création d'interfaces de visualisation => évolution de Mugen, projet en cours mais ralenti par le succès de galaxy
 - Maintien des sites drupals
- Mini projet :
 - Mise en place d'un site gbrowse pour des données de la truite

Conseil *Scientifique* des *Utilisateurs*



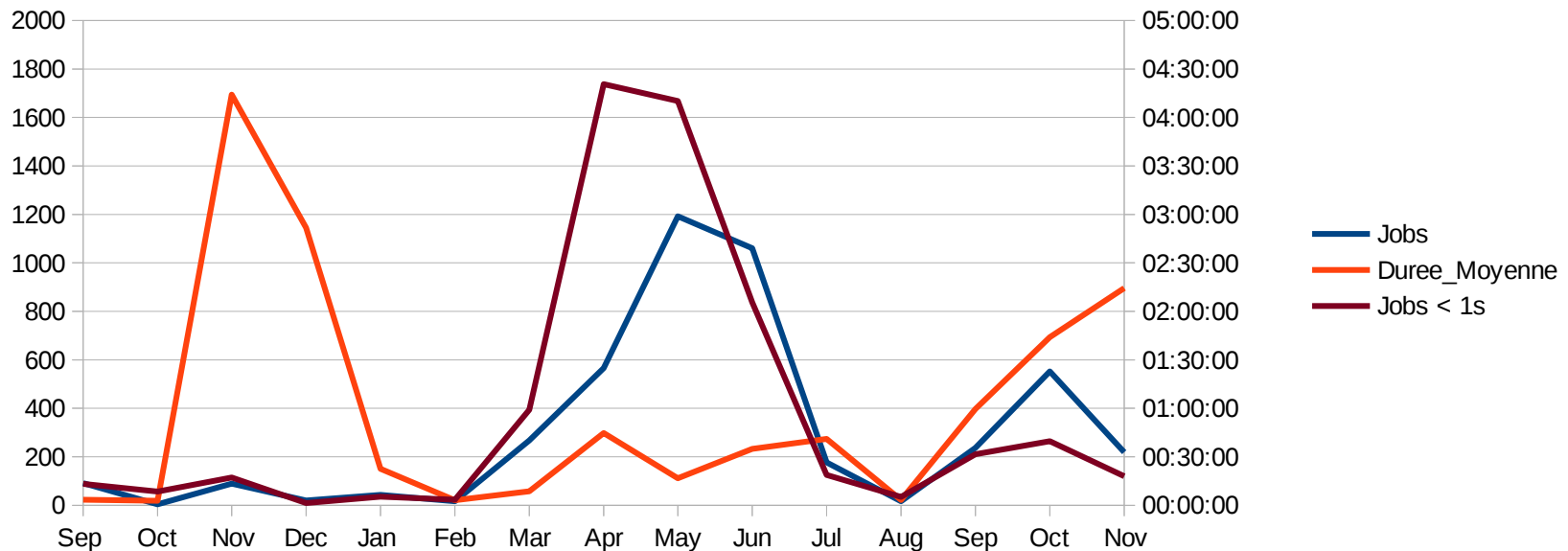
1. Présentation générale de la plateforme
 1. Equipe
 2. Missions
 3. Caractéristiques
2. Point sur les ressources fournies par la plateforme
 1. Matériel
 2. Outils
 3. Bases de données
 4. Formations
 5. Conception et développement
3. **Points divers**
 1. Certification ISO9001
 2. Galaxy
 3. Cluster
 4. Stockage
 5. Achats

ISO 9001

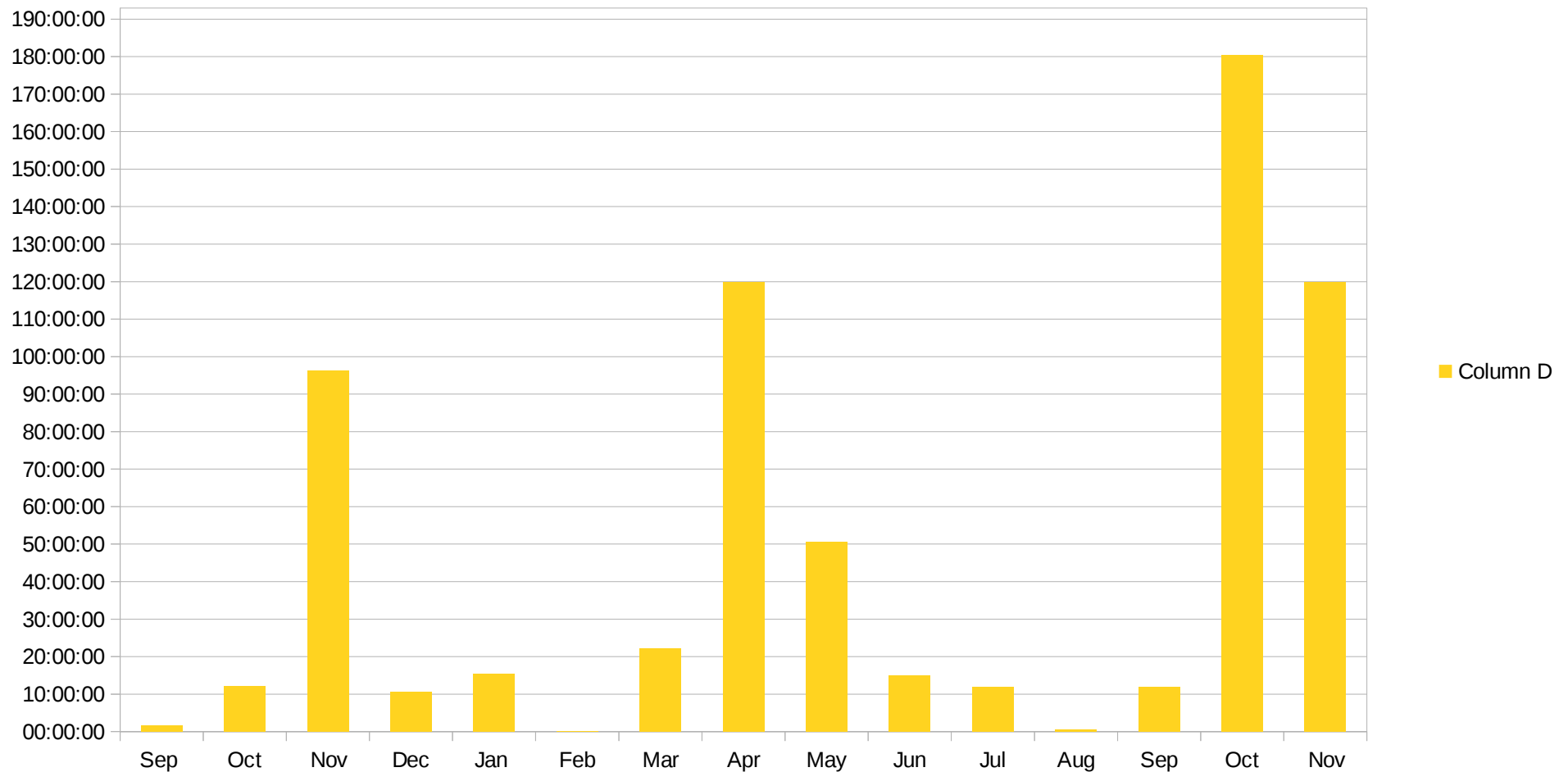
- Certifié en 2011
 - Audit de suivi en 2012
 - Audit de suivi en 2013
 - Audit de renouvellement en 2014
- Quelques évolutions mineures
 - 3ieme mission developpement,
 - Répartition du processus Assistance, Conseil, et Expertise.

Galaxy

- 95 Utilisateurs sur l'interface.
- Principaux outils mis en ligne :
 - Suite R'mes, Rmes Format, SurfG+,Prodigal,Sed,FASTA Ssats,Blastall,HMMScan,RiboPicker,SortmeRNA,Suite NCBI Tools,Sickle,Prinseq,Quast,FASTQc,Suite Velvet,...



Galaxy Durée Max



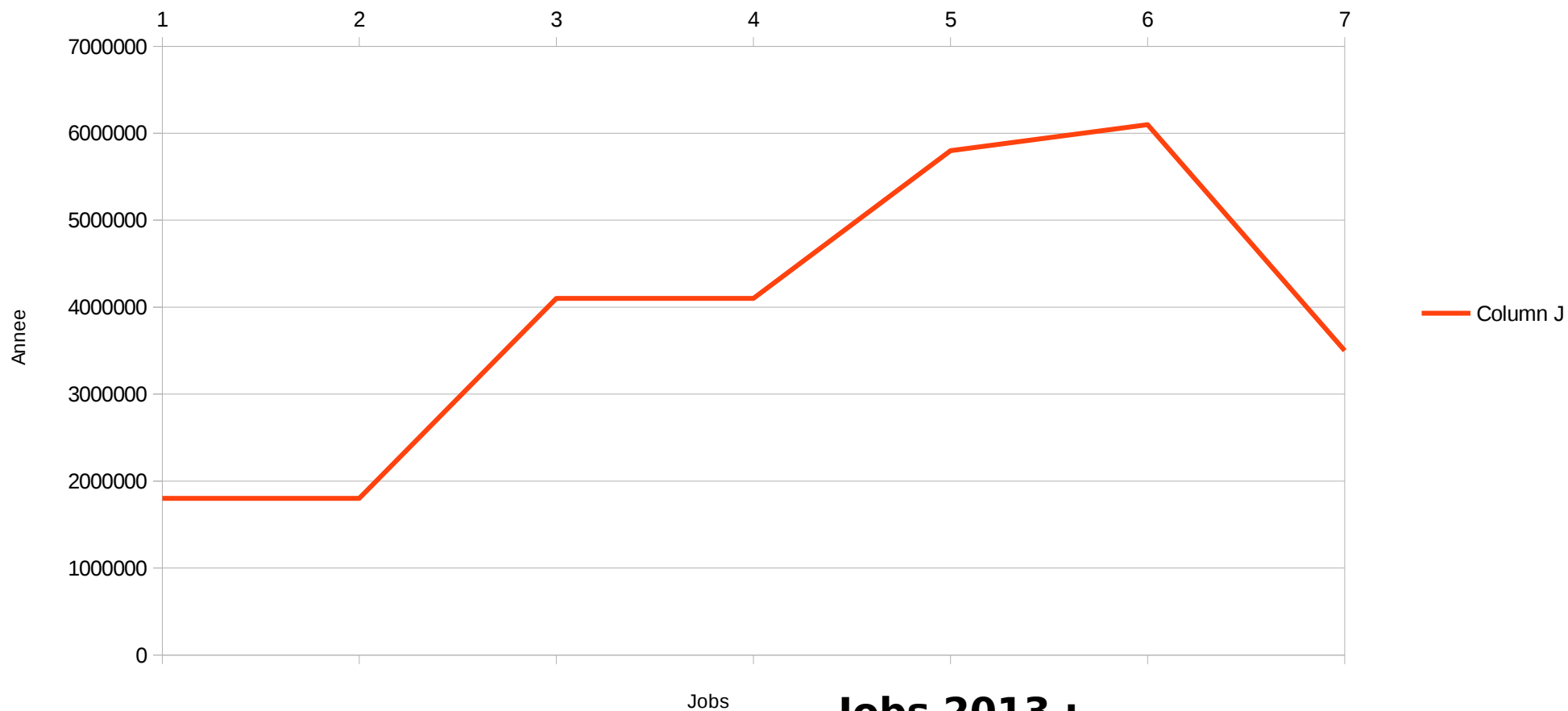
Politique d'utilisation du Cluster

Groupe	Noeuds de calcul	Processeur	Mémoire
@2sixcore-2.4 (120 coeurs)	n50-n59	Intel Six Core 2.4 Ghz	48 Go
@dualcore-2.33 (40 coeurs)	n60-n69	Intel Dual Core 2.33 Ghz	8 Go
@quadcore-2.33 (80 coeurs)	n70-n79	Intel Quad Core 2.33 Ghz	16 Go
@4xquadcore-2.8 (128 coeurs)	n80-n87	AMD Quad Core 2,8 Ghz	96 Go
@2xquadcore-2.27 (128 coeurs)	n88-n103	Intel Quad Core 2.27 Ghz	32 Go

nous sommes à **496** coeurs!

Politique d'utilisation du Cluster

Jobs depuis 2007

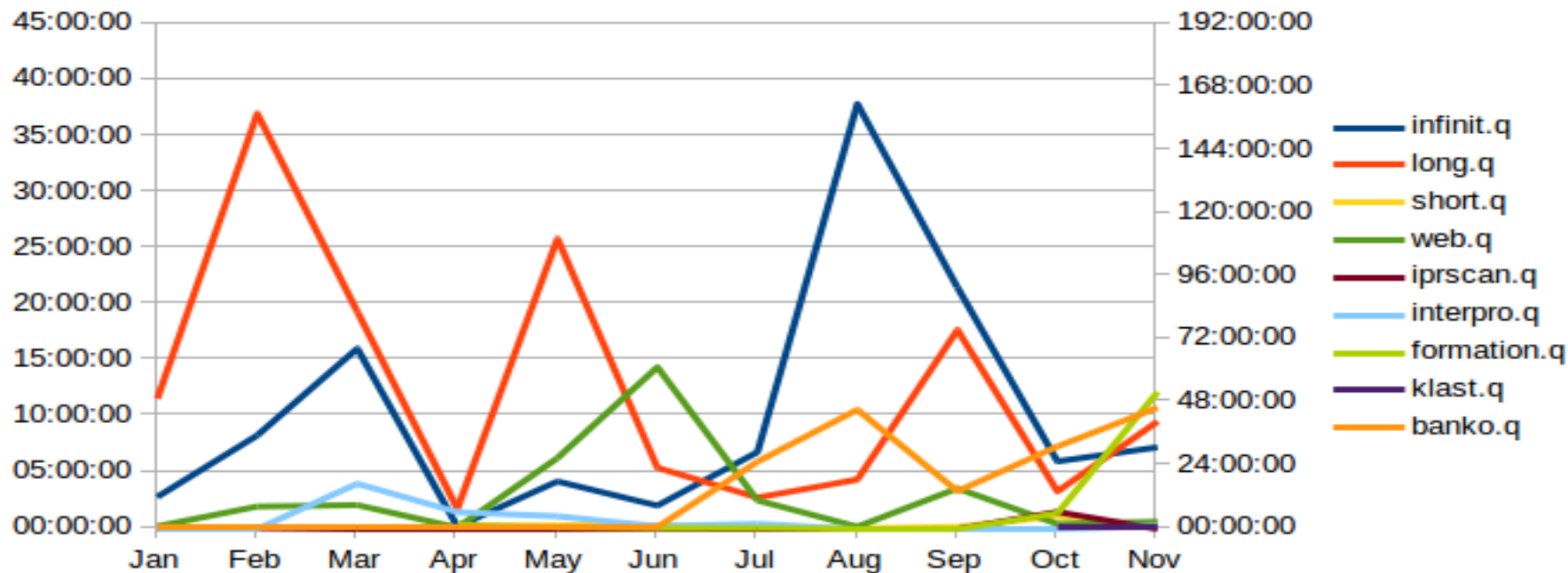


Jobs 2013 :
Total= 3623768 dont 1381217 < 1s
moyenne= 302000

Politique d'utilisation du Cluster

- **Point fils d'attentes**
 - **short.q** (210): jobs courts dont la durée estimée d'exécution ne dépassera pas 4H00 de temps de calcul, c'est la file par défaut.
 - **long.q** (355): jobs dont la durée d'exécution ne va pas dépasser 5 jours.
 - **infini.q** (110): jobs dont la durée d'exécution est supérieure à 5 jours.
 - **web.q** (28): applis web (user apache uniquement).
 - **iprscan.q** (16/3%): suite InterProScan.
 - Pour les calculs nécessitant des ressources mémoires supérieures à 8 Go :
commande `qsub -l big_mem -q short.q blast.sh`.
- **Quota :**
 - Max slots/all users : infini 110, long 355, short 210
 - Max slots/user : infini 52, long 151, short 100

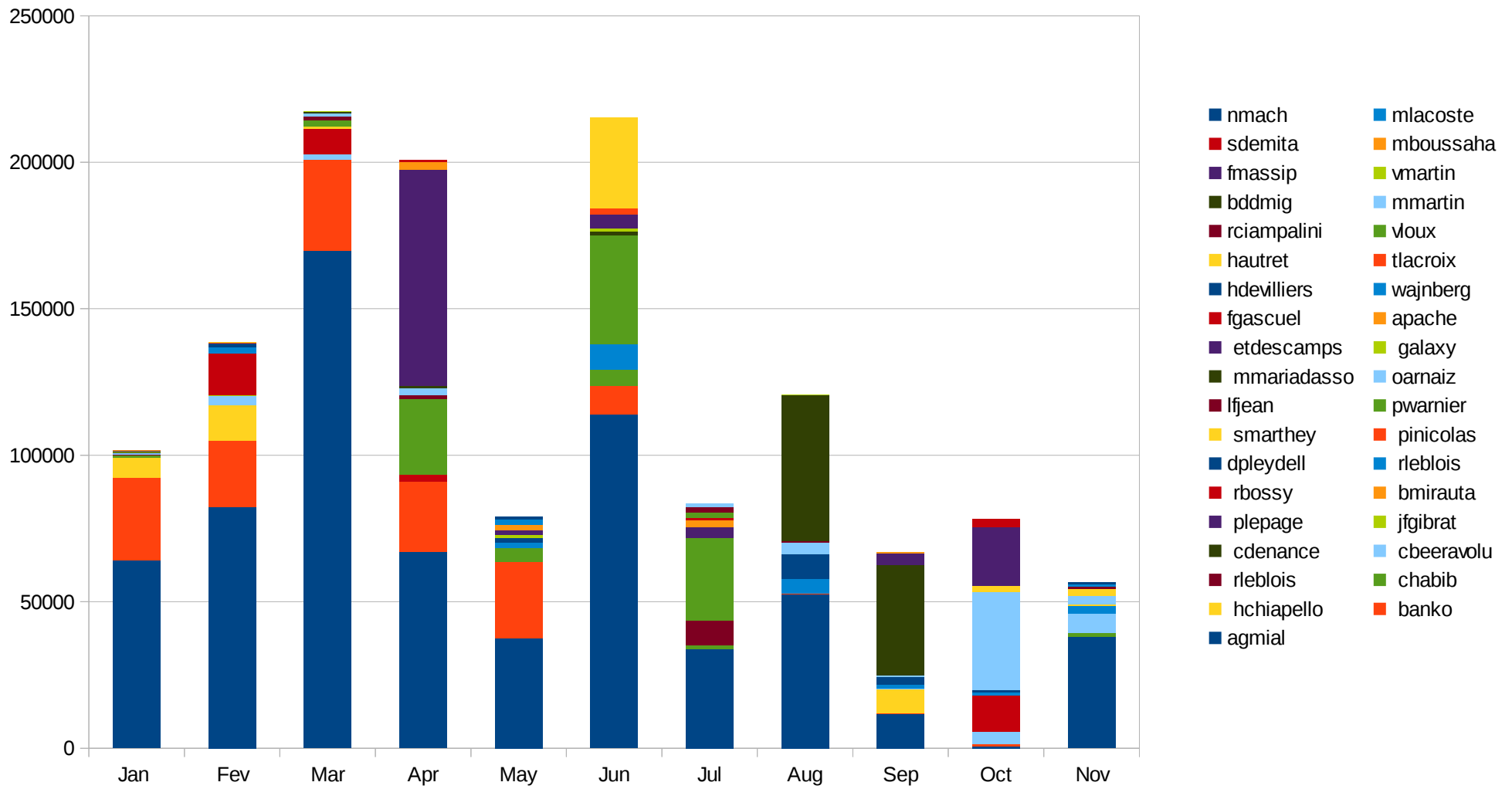
Politique d'utilisation du Cluster



Point fils d'attentes

- **short.q** 848653 jobs
- **long.q** 195992 jobs
- **infinet.q** 6187 jobs
- **web.q** 7268 jobs

Clusteur TOP10



Stockage

- Volumetrie totale : 98To
- Volumetrie banque de données : 20To
- Volumetrie sauvegardée : 45 To
- Mise en place des espaces save/work
- Mise en place d'un espace galaxy

Point achats

Quelques noeuds (3) ont été achetés et sont utilisés actuellement pour la migration de notre système de Centos 5 à Centos 6.

Modification du serveur agmial, et d'un serveur postgresql.

Modification d'un serveur bibliome. (2 R620)

Ajout d'un serveur FranceGenomique. (1 R720)

Sorties :

1 serveur de virtualisation sous xen remplacé par un serveur sous kvm avec une extension mémoire.

Modification de notre stockage, avec notre stockage unifié type "NetApp"

Achat effectué premier semestre 2013, une 30aine de To utiles.

Extension de la baie avec 5 disques achetés par MIAj pour leurs données.

Extension fin d'année 2013 avec l'ajout d'un tiroir de disque supplémentaire avec des disques. +20To :

70To brut/ 55To Utile

Points divers

Achat d'une licence KEGG Institutionnelle 5000 \$:

En 2012 :

- Participation du département MIG : 975€
- Participation de MGP : ~2000€
- MIG finance la partie restante

En 2013 :

Participation de MGP : ~ 2000€

MIG finance le reste.

Nombre d'utilisateurs effectifs de la ressource 2.

Points divers

DataCenter .

Points divers

Achat d'une licence Klast