

# Environnement informatique d'imagerie microscopique en biologie structurale à l'IGBMC

Jonathan Michalon, 15 décembre 2015

Rencontre « Cycle de vie des images en microscopie : les enjeux  
techniques »

# Contexte

- IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire), 4 départements, près de 750 personnes ;
- département de biologie structurale intégrative ;
- problématiques informatiques de stockage et de calcul :
  - principal microscope : 2 To de données brutes par jour ;
  - autres systèmes ;
- Instruct (européen), FRISBI (French Infrastructure for Integrated Structural Biology).

# Stockage

# Besoins

- Stockage capacitif, capable de supporter des débits importants sur les jeux de données ;
- accès depuis différents types de machines (calcul, visualisation, poste de travail, . . . ) ;
- disponibilité suffisante (si des données doivent être envoyées immédiatement, il faut pouvoir) ;
- maîtrise totale du système pour pouvoir optimiser et dédier spécifiquement l'ensemble à des données d'imagerie biologique.

# Matériel utilisé

Nous avons opté pour du stockage sur matériel générique :

- 3 serveurs Dell R720xd avec 64 Go de RAM ;
- 5 baies de disques Dell MD1200 ;
- 288 To en disques capacitifs de 4 To à 7200 tpm (72 disques).

# Mise en place

Pour maîtriser la chaîne « acquisitions - stockage - traitements » et travailler au plus proche des données, nous utilisons du logiciel libre sur des serveurs GNU/Linux standards.

- RAID6 par ensemble de 12 disques ;
- GlusterFS :
  - système de fichiers distribué ;
  - agrégation des briques de stockage ;
  - réplication simple ;
  - montage natif ou NFS, API d'accès implémentée dans Samba/CIFS et qemu.

Nous obtenons ainsi un stockage central, accessible par une variété de méthodes, redondé (sans interruption de service en mode natif) et rapide (750 Mo/s en transferts parallélisés).

# Calcul

# Besoins

Il y a différents types de besoins, principalement :

- Des tâches extrêmement parallélisables (classification 2D de particules par exemple) ;
- des tâches gourmandes en mémoire (reconstruction 3D) ;
- de la visualisation 3D demandant des ressources graphiques (affichage et GPGPU).



# Mise en place

Pour répondre à cela les chercheurs ont maintenant accès à plusieurs systèmes :

- Deux serveurs R920 comprenant chacun 60 cœurs de calcul et 1 To de RAM ;
- le méso-centre de l'Université de Strasbourg :
  - contribution directe de 256 cœurs de calcul ;
  - mutualisation efficace ;
- des serveurs comportant des cartes graphiques, avec affichage à distance.

# Réseau

# Besoins

- Supporter les transferts rapides depuis les machines d'acquisition ;
- accéder au stockage depuis les machines de calcul locales ;
- travailler sur le méso-centre rapidement ;
- mettre en place des machines virtuelles au plus proche des données.

# Mise en place

- 10 gbps généralisé :
  - bâtiment récent donc câblage le supportant ;
  - switch et cartes réseau en conséquence ;
- montage du stockage permanent sur les machines de traitement locales (autofs + client natif GlusterFS) ;
- VLAN d'interconnexion vers le méso-centre ;
- transferts parallélisés accessibles au moyen de scripts et d'interfaces web minimalistes.

# Global

Pour obtenir cette structure le workflow sera...

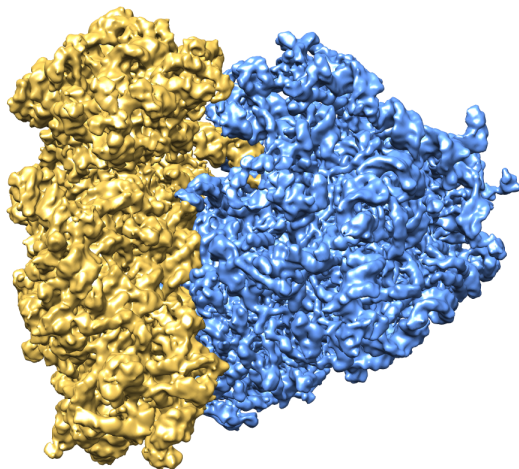


FIGURE 1: Structure 3D d'un ribosome

## Étapes pour obtenir l'image de ribosome précédente :

- Acquisition (microscope Titan Kryos) :
  - plus de 1 To de micrographes sur la passerelle-tampon ;
  - transfert parallélisé via un formulaire web vers le stockage centralisé ;
- Picking (sélection des particules) :
  - eman2 avec visualisation à distance, directement sur le stockage ;
- Classification 2D (groupement des particules en classes) :
  - eman2 sur le méso-centre, transferts aller-retour en 10 gbps ;
- Reconstruction 3D (affinage itératif de la structure) :
  - relion sur R920.

# Conclusion

- Infrastructure informatique forte ;
- répondant aux diverses problématiques inhérentes aux travaux des chercheurs ;
- faire évoluer l'infrastructure en augmentant les capacités (le stockage est le prochain travail) ;
- se recentrer sur les traitements en étant assuré d'une base de travail solide et évolutive.



Merci pour votre attention.<sup>1</sup>

---

1. Support de présentation réalisé en Markdown, utilisant la classe Beamer de  $\text{\LaTeX}$  via Pandoc.