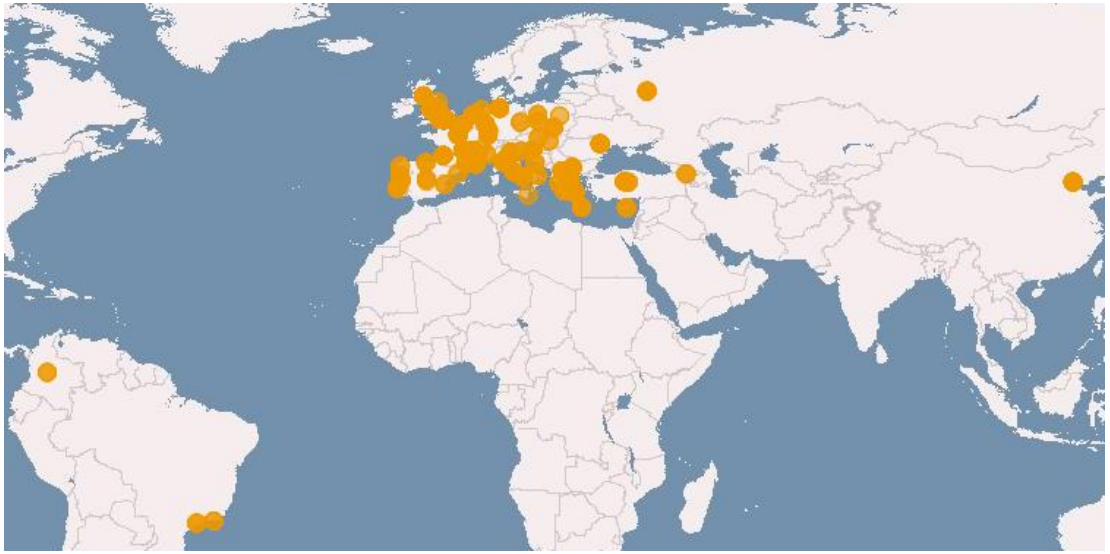
Gestion de données : Besoins de la VO Biomed

Sorina Pop

Laboratoire Creatis
Université de Lyon, CREATIS; CNRS UMR5220; Inserm U1044; INSA-Lyon; Université Lyon 1, France

VO Biomed en bref

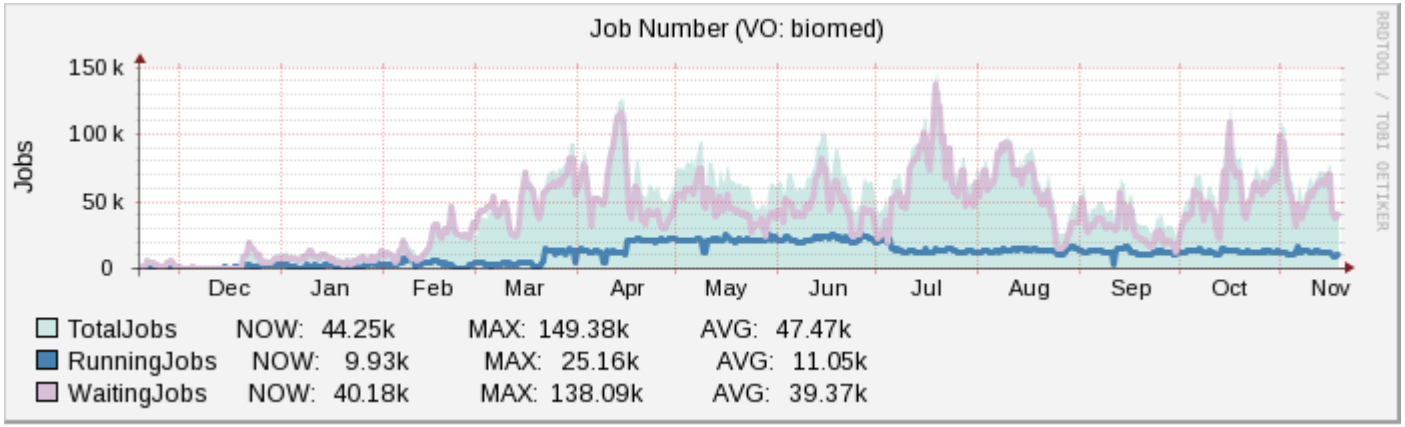
- Plus de 100 sites
  - *~190 CEs*
  - *~100 SEs*
  - *36 WMS*
  - *1 LFC*
  - *1 serveur VOMS*



- Environ 300 utilisateurs

VO Biomed : calcul

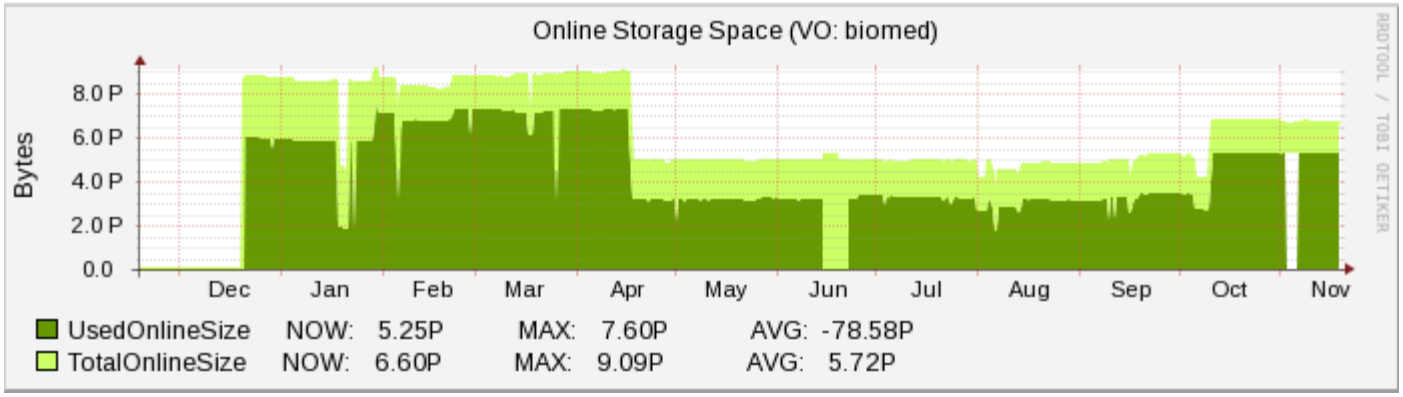
- Jobs de calcul (source GStat2)
  - *En moyenne, 40k jobs en attente versus 11k jobs en exécution*



- Dirac est utilisé par une partie des utilisateurs
  - *Instance nationale mise à disposition par France-Grilles*

VO biomed : stockage

- Utilisation des ressources de stockage
  - *5.25 PB utilisés sur les 6.6 PB disponibles (source GStat2)*



Gestion des données

- SE plein
- Fichiers anciens
  - *« oubliés » par les utilisateurs*
- Zombies (« dark data »)
  - *Réplicas physiques qui n'existent pas dans le LFC*
- Ghosts (« fantômes »)
  - *Données qui existent dans le LFC, mais n'ont pas de réplica physique*

Biomed technical shifts

- Participation volontaire de 8 laboratoires
  - *http://lsgc.org/en/Biomed:home*
- Suivre les ressources de la VO (100+ sites)
  - *LFC, VOMS, CE (188 CEs, 36 WMS), stockage (100 SEs)*
- Interface entre les utilisateurs et les sites
- Mise en place et exécution de procédures
  - *Ex: SE plein ou retiré de la production*

Procédures actuelles : SE plein

- Identification of full SEs
  - *To identify full Storage Elements, refer to the status of biomed on-line storage space report, and select only SEs with **less than 10% AND less than 500 GB of free space left**.*
  - *If biomed uses more than 20% of the total space, then apply the clean-up biomed files procedure*

- **Clean-up biomed files (in case biomed is a heavy user)**
  - *Get the list of LFNs and DNs having files stored on this SE using LFCBrowseSE. Note that this may take a while (few hours)*
  - *Submit a GGUS **Team** ticket, assign it to **VOSupport**,*
  - *Put a link or attach the files list in the ticket.*
  - *Notify the concerned users by email, and require them to move or delete their data*
  - *Send reminders to users every week until at least 90% (or at least 500 Go for big SEs) of the SE space is free. Note that you may need to contact the site admins in order to get the accurate value of the storage space for biomed*
  - *The ticket is solved when at least 90% (or at least 500 Go for big SEs) of the SE space for biomed is free.*

Procédures actuelles (suite)

- SEs retirés de la production
  - *Récupérer la liste des fichiers sur le SE*
  - *Notifier les utilisateurs*
  - *Effacer si besoin les entrées du LFC*
- Gestion de ghosts et zombies
  - *Listing du LFC avec LFCBrowseSE*
  - *Listing des SEs à l'aide des admins des sites*
  - *Mise en correspondance des résultats*
  - *Demandes d'effacement des ghosts et zombies*
- Procédures lourdes et peu efficaces…

VAPOR

- Vo Administration and operations PORtal
  - *Mini-projet EGI-InSPIRE démarré en avril 2013 pour 12 mois*
  - *Porteur : Franck Michel, I3S*
- Buts
  - *Mutualiser l'effort opérationnel des VOs de petite ou moyenne taille*
  - *Fournir les outils appropriées*
- Axes principales
  - *Indicateurs de l'état des ressources, rapports*
  - *Gestion des utilisateurs*
  - ***Gestion des données***

Pistes de réflexion

- GFAL 2.0
  - *Bibliothèque pour l'accès et la gestion des données*
  - *Permet le listing récursif des données sur les SEs*
    - Avec des infos type « posix »
- CVMFS : CernVM File System
  - *Meilleure disponibilité et gestion des fichiers ?*
  - *Contact Catalin Condurache*
    - Déploiement du stratum 0 et 1 au RAL
- DIRAC

Merci pour votre attention !
Questions ?