CBP
CENTRE BLAISE PASCAL

Aramis 2016

Stockage : des usages aux outils…
Mais pour quels enjeux ?

Emmanuel Quémener

Plan (& avertissements d'usage)

- Commençons par la fin : quels enjeux ?
  - Il y a bien bien longtemps, dans un système stellaire banal...
- Stockage & Centre Blaise Pascal : histoire & historique
  - Enquête « besoin de stockage » ENS-Lyon : le détonateur (presque)
- Retour d'expériences (comme demandé...)
  - Le stockage distribué : une évidence, mais à quelles conditions ?
  - ZFS comme socle : une alternative au Hardware & aux $€£
  - La visualisation distante : un mariage étonnamment efficace
  - L'accès simplifié aux ressources de calcul : le web comme portail
  - Traitements de données biologiques : les nouveaux défis
- Conclusion

En 1676, Römer ou il était la C élérité

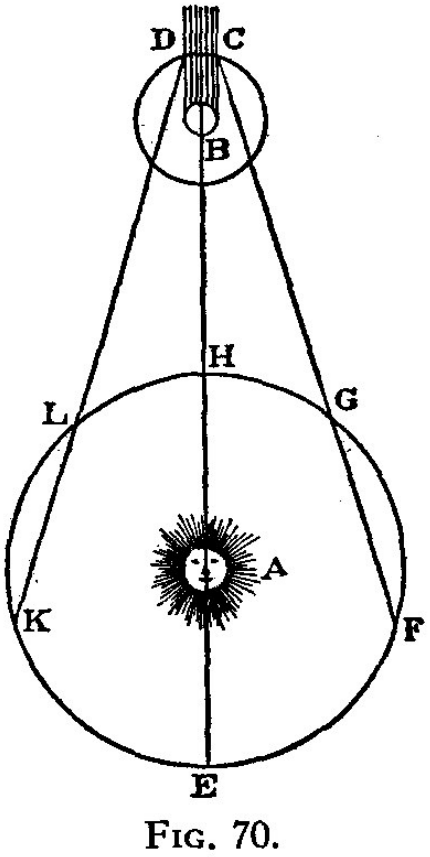
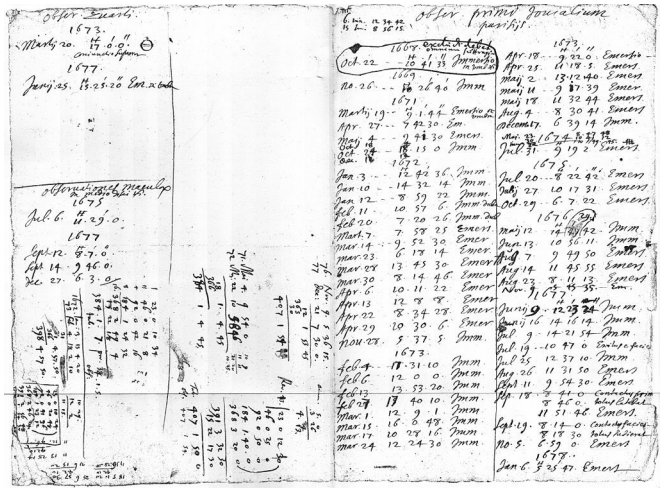


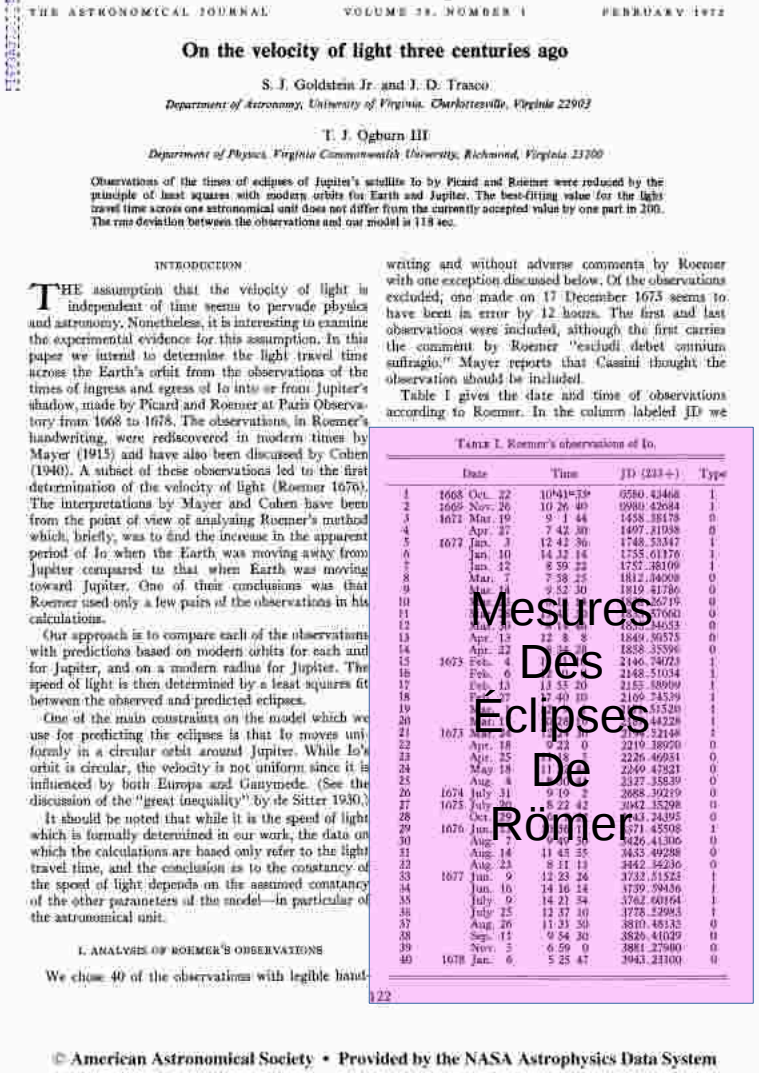
FIG. 70.



Vitesse de la lumière

- Finie ! (et c'est un pilier de la physique)
- Évaluée à 225000 km/s (25 % d'erreur)

En 1973, Goldstein & Trasco



THE ASTRONOMICAL JOURNAL VOLUME 78, NUMBER 1 FEBRUARY 1973

On the velocity of light three centuries ago

S. J. Goldstein Jr. and J. D. Trasco
Department of Astronomy, University of Virginia, Charlottesville, Virginia 22903

T. J. Ogburn III
Department of Physics, Virginia Commonwealth University, Richmond, Virginia 23220

Observations of the times of eclipses of Jupiter's satellite Io by Picard and Roemer were reduced by the principle of least squares with modern orbits for Earth and Jupiter. The best-fitting value for the light travel time across one astronomical unit does not differ from the currently accepted value by one part in 200. The rms deviation between the observations and our model is 118 sec.

INTRODUCTION

THE assumption that the velocity of light is independent of time seems to pervade physics and astronomy. Nonetheless, it is interesting to examine the experimental evidence for this assumption. In this paper we intend to determine the light travel time across the Earth's orbit from the observations of the times of ingress and egress of Io into or from Jupiter's shadow, made by Picard and Roemer at Paris Observatory from 1668 to 1678. The observations, in Roemer's handwriting, were rediscovered in modern times by Mayer (1915) and have also been discussed by Cohen (1940). A subset of these observations led to the first determination of the velocity of light (Roemer 1676). The interpretations by Mayer and Cohen have been from the point of view of analyzing Roemer's method which, briefly, was to find the increase in the apparent period of Io when the Earth was moving away from Jupiter compared to that when Earth was moving toward Jupiter. One of their conclusions was that Roemer used only a few pairs of the observations in his calculations.

Our approach is to compare each of the observations with predictions based on modern orbits for each and for Jupiter, and on a modern radius for Jupiter. The speed of light is then determined by a least squares fit between the observed and predicted eclipses.

One of the main constraints on the model which we use for predicting the eclipses is that Io moves uniformly in a circular orbit around Jupiter. While Io's orbit is circular, the velocity is not uniform since it is influenced by both Europa and Ganymede. (See the discussion of the "great inequality" by de Sitter 1930.)

It should be noted that while it is the speed of light which is formally determined in our work, the data on which the calculations are based only refer to the light travel time, and the conclusion as to the constancy of the speed of light depends on the assumed constancy of the other parameters of the model—in particular of the astronomical unit.

I. ANALYSIS OF ROEMER'S OBSERVATIONS

We chose 40 of the observations with legible handwriting and without adverse comments by Roemer with one exception discussed below. Of the observations excluded, one made on 17 December 1675 seems to have been in error by 12 hours. The first and last observations were included, although the first carries the comment by Roemer "excludi debet omnium suffragio." Mayer reports that Cassini thought the observation should be included.

Table I gives the date and time of observations according to Roemer. In the column labeled JD we

TABLE I. Roemer's observations of Io.

	Date	Time	JD (2333+)	Type
1	1668 Oct. 22	10 41 39	0580.43468	1
2	1668 Nov. 26	10 26 40	0980.42684	1
3	1672 Mar. 19	9 1 44	1458.38178	0
4	Apr. 27	7 42 30	1497.31938	0
5	1672 Jan. 3	12 42 36	1748.53347	1
6	Jan. 10	14 32 14	1755.61176	1
7	Jan. 12	8 59 22	1757.38109	1
8	Mar. 7	7 58 25	1812.34208	0
9	Mar. 14	9 52 30	1819.41786	0
10	[illegible]	[illegible]	[illegible]	0
11	[illegible]	[illegible]	[illegible]	0
12	[illegible]	[illegible]	[illegible]	0
13	Apr. 13	12 8 8	1849.50575	0
14	Apr. 22	[illegible]	1858.35596	0
15	1673 Feb. 4	9 22 0	2146.38973	1
16	Feb. 6	7 18 0	2148.31034	1
17	Feb. 13	13 53 20	2155.58909	1
18	Feb. 27	17 40 10	2169.74539	1
19	[illegible]	[illegible]	[illegible]	1
20	[illegible]	[illegible]	[illegible]	1
21	1673 Mar. 20	[illegible]	2190.52148	1
22	Apr. 18	9 22 0	2219.38950	0
23	Apr. 25	7 18 0	2226.46951	0
24	May 18	[illegible]	2249.47821	0
25	Aug. 4	[illegible]	2327.35839	0
26	1674 July 31	9 19 2	2688.39219	0
27	1675 July 26	8 22 42	3042.35298	0
28	Oct. [illegible]	[illegible]	3143.24395	0
29	1676 Jan. 17	[illegible]	3271.45508	1
30	Aug. 7	[illegible]	3426.41306	0
31	Aug. 14	11 45 55	3433.49288	0
32	Aug. 23	8 11 13	3442.34236	0
33	1677 Jan. 9	12 23 26	3732.51523	1
34	Jan. 16	14 16 14	3739.59436	1
35	July 9	14 21 34	3762.60164	1
36	July 25	12 37 10	3778.52993	1
37	Aug. 26	11 21 50	3810.48133	0
38	Sept. 11	9 54 30	3826.41029	0
39	Nov. 3	6 59 0	3881.27980	0
40	1678 Jan. 6	5 25 47	3943.23100	0

122

Mesures Des Éclipses De Römer

TABLE II. Heliocentric coordinates for Earth and Jupiter.

	X_E	Y_E	Z_E	X_J	Y_J	Z_J
1	0.8217	0.5116	0.2223	1.9602	4.2957	1.7957
2	0.3506	0.8448	0.3670	−1.0325	4.6616	2.0263
3	−0.9958	−0.0503	−0.0219	−4.1044	3.1110	1.4362
4	−0.7574	−0.6103	−0.2652	−4.2897	2.9099	1.3544
5	−0.2945	0.8604	0.3737	−5.1673	1.4374	0.7439
6	−0.4102	0.8199	0.3561	−5.1835	1.3926	0.7251
7	−0.4381	0.8078	0.3509	−5.1875	1.3814	0.7204
8	−0.9843	0.1280	0.0555	−5.2962	1.0290	0.5718
9	−0.9959	0.0163	0.0070	−5.3080	0.9832	0.5524
10	−0.9895	−0.1233	−0.0536	−5.3221	0.9257	0.5281
11	−0.9746	−0.2060	−0.0895	−5.3302	0.8912	0.5135
12	−0.9678	−0.2332	−0.1013	−5.3329	0.8797	0.5086
13	−0.8821	−0.4410	−0.1916	−5.3528	0.7873	0.4694
14	−0.8020	−0.5956	−0.2587	−5.3641	0.7294	0.4449
15	−0.7661	−0.6506	−0.2827	−5.3678	0.7103	−0.3696
16	−0.7854	−0.5[illegible]	[illegible]	−5.[illegible]	−1.1767	−0.3746
17	−0.8550	−0.4[illegible]	[illegible]	−5.[illegible]	−1.2221	−0.3944
18	−0.9540	−0.2493	−0.1083	−5.2724	−1.3126	−0.4338
19	−0.[illegible]	−0.[illegible]	−0.0965	−5.[illegible]	−1.3239	−0.4388
20	−0.[illegible]	−0.[illegible]	−0.[illegible]	−5.[illegible]	−1.[illegible]	−0.4829
21	−0.[illegible]	−0.[illegible]	−0.[illegible]	−5.[illegible]	−1.[illegible]	−0.5025
22	−0.[illegible]	−0.[illegible]	−0.[illegible]	−5.[illegible]	−1.[illegible]	−0.5709
23	−0.7722	−0.5940	−0.2581	−5.1532	−1.6701	−0.5902
24	−0.4725	−0.8217	−0.3570	−5.0961	−1.8124	−0.6527
25	−0.[illegible]	−0.[illegible]	−0.[illegible]	−5.[illegible]	−1.[illegible]	−0.8592
26	−0.[illegible]	−0.[illegible]	−0.[illegible]	−5.[illegible]	−1.[illegible]	−0.6568
27	−0.[illegible]	−0.[illegible]	−0.[illegible]	−5.[illegible]	−1.[illegible]	−0.0477
28	0.7575	0.3868	0.2550	0.0323	−4.8182	−2.0689
29	−0.0484	−0.9311	−0.4044	1.7050	−4.4620	−1.9571
30	0.7732	−0.6003	−0.2607	2.0864	−4.3003	−1.8971
31	0.8441	−0.5114	−0.2221	2.1346	−4.2774	−1.8885
32	0.9158	−0.3900	−0.1694	2.1945	−4.2482	−1.8774
33	−0.1191	−0.9254	−0.4020	3.8890	−2.9144	−1.3465
34	0.0003	−0.9323	−0.4050	3.9223	−2.8738	−1.3299
35	0.3799	−0.8648	−0.3756	4.0271	−2.7394	−1.2748
36	0.6131	−0.7420	−0.3223	4.0968	−2.6443	−1.2357
37	0.9347	−0.3485	−0.1513	4.2293	−2.4490	−1.1551
38	0.9981	0.1066	0.0462	4.2916	−2.3495	−1.1139
39	0.6656	0.6718	0.2919	4.4864	−1.9961	−0.9670
40	−0.3468	0.8439	0.3666	4.6679	−1.5803	−0.7930

Notes: X is measured in the direction of the vernal equinox (1950); Y, in the celestial equator 90° eastward; Z, toward the celestial north pole.

Positions de La Terre et Jupiter Par l'observatoire naval

- Utilisation des mesures de Römer
- Exploitation des positions astronomiques de l'Observatoire Naval des USA
- Utilisation de méthodes statistiques
- Valeur de c : 300000 km/s à 0.5 %

Conclusion de l'introduction ?

- **Ole Römer était un excellent observateur :**
  - Et même pas de chronomètre à disposition !
- **« Ce n'est pas parce que la conclusion est erronée ou approximative que les mesures le sont. »**
  - Römer fait une erreur de 25 % mais l'erreur provient essentiellement d'autres paramètres à la précision insuffisante !
- **En Science, on peut faire du neuf avec du vieux :**
  - ***Astrophysical*** *Journal* n'est pas ***Science X*** (et heureusement)...
- **Pour retravailler des données archaïques, encore faut-il les avoir...**
  - Et c'est là que NOUS intervenons :-)

D'abord, c'est quoi le CBP ?

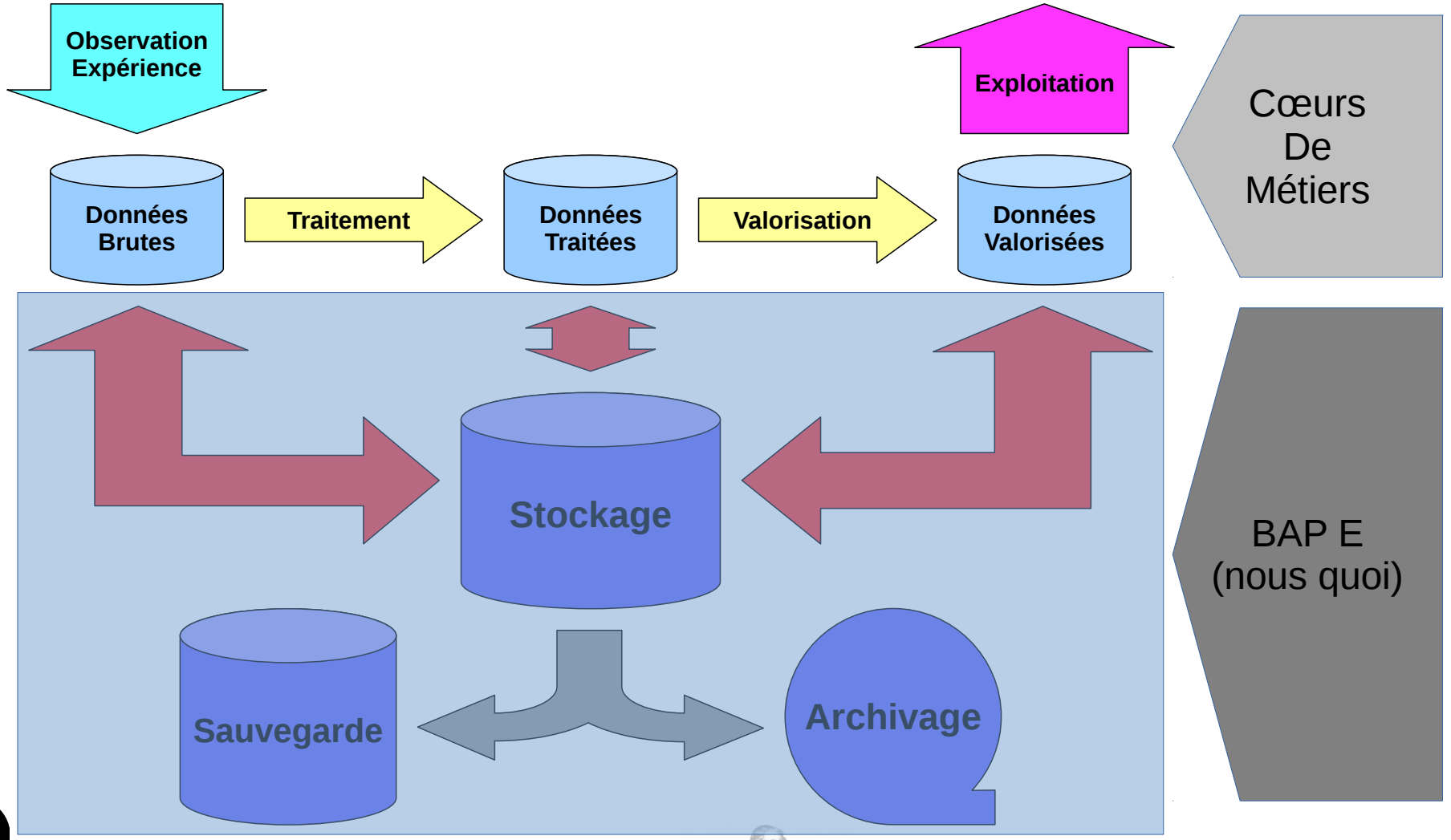
- Hôtel à projets, à conférences, à formations...
- Maison de la modélisation
- Plate-forme expérimentale / Centre d'essais :
  - Plateaux techniques multi-nœuds, multi-cœurs, multi-shaders, …
    - De 4 à 64 nœuds, de 2 à 20 cœurs, plus de 30 modèles de GPU
  - Plateaux techniques d'intégration : Debian, Ubuntu, …
  - Plateau technique « architectures exotiques »
  - « Paillasses numériques » :
    - Machines virtuelles pour les « Humanités Numériques »
    - Machines de visualisation : 3D avec lunettes, MorphoGraphX, …
    - Machines de traitement expérimentales : repeat*, Galaxy, ...

Enquête sur le stockage ENS-Lyon

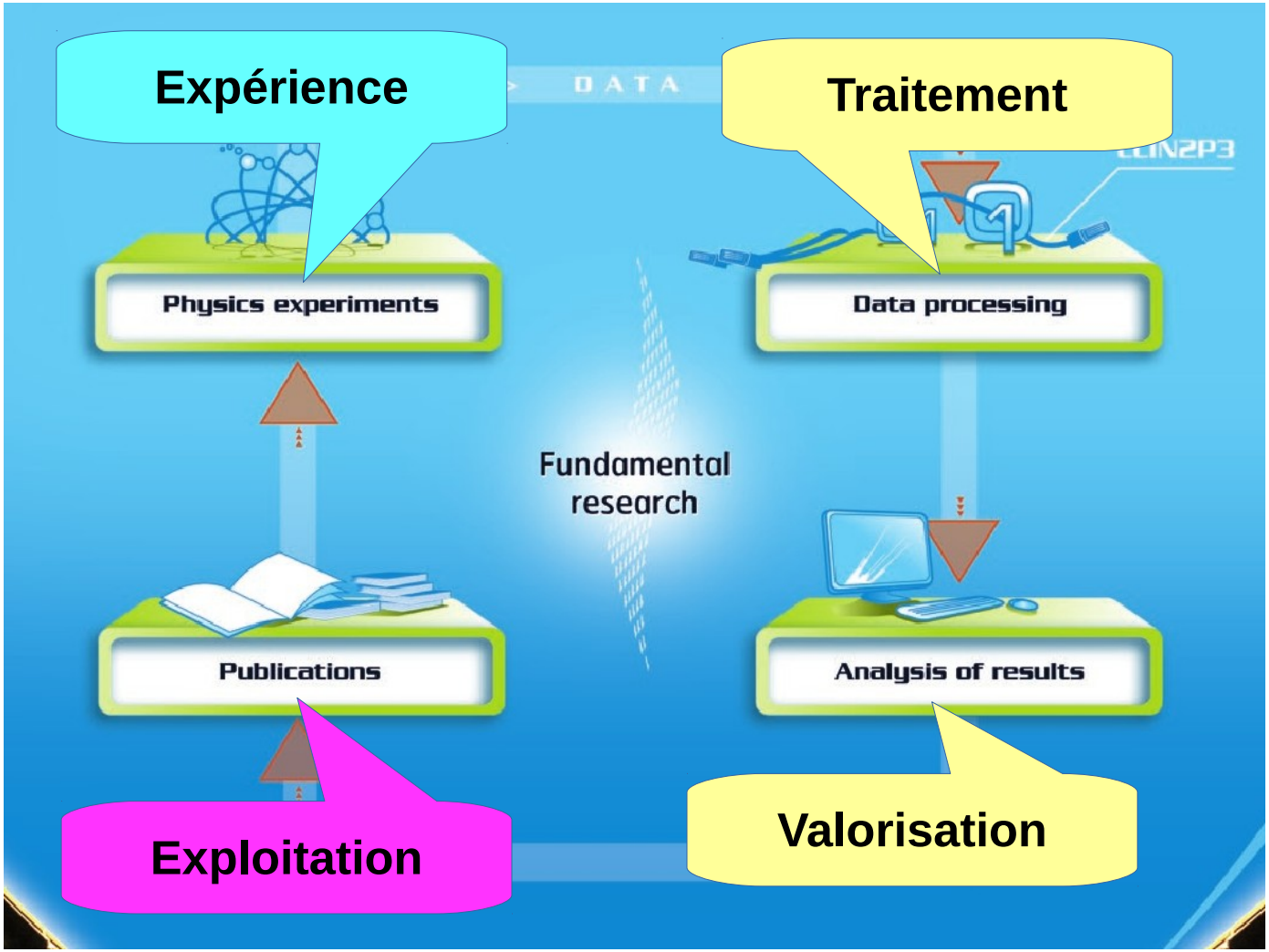
Direction de la recherche, le 20 décembre 2009

- **Mission d'étude confiée au CBP sur :**
  - Besoins en sauvegarde et de stockage des laboratoires de biologie ;
  - Besoins en sauvegarde et de stockage des autres laboratoires du site Monod ;
  - Les conséquences en terme de « froid » ;
  - Les conséquences en terme de locaux ;
  - Premières conclusions pour fin février 2010.

Modélisation du circuit de l'information



Une modélisation partagée ?
En septembre 2011 à l'IN2P3



L'approche projet CQQCOQP

- Méthode analytique standard
  - Systématique : questionnement à 7 questions
    - Qui fait Quoi ? Où ? Quand ? Comment ? Combien Pourquoi ?
  - Universelle : *five Ws* en anglais
    - *« Who, What, Where, When, Why ? »*
  - Source http://fr.wikipedia.org/wiki/QQOQCCP
- S'interroger sur l'objet (la « cible ») :
  - Un problème, une situation, un processus, une solution
- Grain choisi dans l'étude : le processus...

C**Q**QCOQ**P** : Pourquoi ? Quoi ?

- Pourquoi ?
  - A quoi « sert » la plate-forme expérimentale ?
- Quoi ?
  - Quelle est la nature des acquisitions ?
  - Entrée : Quelle est la nature « physique » de l'entrée ?
    - Exemple : Une caméra CCD de format 1024×768 à 25Hz
  - Sortie : Quelle est la nature « numérique » de la sortie ?
    - Exemple : Une image Tiff non compressée de 1024×768 sur 16bits

CQ**Q**COQP : Qui ?

- Quels sont les acteurs autour des processus ?
  - Par qui ? Quelle personne réalise l'expérience ?
  - Exemple : Un ITA réalise l'expérience suivant un protocole pré-établi
- Pour qui ? Quelle personne exploite les résultats ?
  - Exemple : Un doctorant exploite les données

CQQCO**Q**P : Quand ?

- Quels critères de temps en jeu ?
  - **Durée** : Quelle est la durée moyenne d'une expérience ?
    - Exemple : Une expérience dure en moyenne une demi-journée
  - **Récurrence** : Combien d'expériences sont-elles réalisées par semaine ?
    - Exemple : La plate-forme est utilisée en moyenne 5 fois par semaine
  - **Pérennité** : Quelle durée de vie attribuer aux données ?
    - Exemple : Les données brutes sont conservées pendant les 3 ans de doctorat.

CQQC**O**QP : Où ?

- Où se situent les équipements ?
  - **Source** : Les configurations d'une expérience sont-elles stockées localement ?
    - Exemple : les paramètres de chaque expérience sont sur un répertoire distant.
  - **Équipement** : Dans quel local se trouve la manipulation (contrainte d'accès) ?
    - Exemple : la manipulation, dans une salle blanche, ne permet pas l'entrée et la sortie de périphériques de stockage amovible. Le lien réseau est le seul utilisable.
  - **Destination** : A quel endroit seront stockés les résultats d'une expérience (localement, sur un serveur de stockage etc...) ?
    - Exemple : Les résultats ne sont stockés que de façon temporaire

CQQCOQP : Combien ?

- Ces questions concernent la « sortie » des données
  - **Volume** : Quel est le volume de données (en Mo) d'une expérience ?
    - Exemple : une manipulation génère 1500 images pour un volume de 2 Go.
  - **Débit** : Le réseau standard (100Mb/s) est-il suffisant pour les transferts ?
    - Exemple : non, le transfert dure en moyenne plus d'une heure
  - **Latence** : Existe-t-il des contraintes sur les outils pour une acquisition ?
    - Exemple : oui, l'utilisation d'un partage réseau est inadaptée pour des raisons d'accès direct aux données.

CQQCOQP : Comment ?

- Logiciels :
  - Quel type de logiciel est utilisé pour l'acquisition, le traitement ?
  - Exemple :
    - le logiciel est fourni avec l'équipement.
    - Ses bogues entraînent des dysfonctionnements sur certaines séquences.
    - Il est indispensable de réaliser les acquisitions en local puis de les télécharger sur le dossier distant.

De la consultation à l'analyse...

- Questionnaire en ligne hiver 2010T1
- Analyse & Synthèse des réponses 2010T1
- Rédaction des rapports 2010T2
- Publication : en ligne & JRES 2011
  - Directement : http://www.cbp.ens-lyon.fr/doku.php?id=developpement:projets:stockage4ens
  - Dans Google : « enquête stockage CBP », second lien
- Mais quelles spécifications fonctionnelles tirer ?
  - Il n'y a pas que les 150TB de la première année à fournir...

Spécifications fonctionnelles
Le « salon »

Le « front office », une affaire d'utilisateurs...

- gestion fine de l'accès aux données :
  - pour les accès en écriture à partir des postes de manipulation,
  - pour les accès en lecture à partir des postes de traitement,
  - pour les responsables face aux personnels temporaires
- accessibilité des données dépassant le cadre du laboratoire :
  - espace accessible de l'extérieur (simple & sécurisé)
- indexation indispensable des expériences et des traitements
- mise en place de plates-formes de traitements dédiées
- abstraction des volumes de stockage
- mise à disposition rapide

Spécifications fonctionnelles
La « cuisine »

Le « back-office », une affaire d'informaticiens :

- amélioration des conditions de transfert des données
- disponibilité accrue des dispositifs de stockage
- souscription « large » de contrat de maintenance
- procédures simplifiées (mise à disposition & extension)
- procédures simplifiées pour la restauration d'un volume
- « scalabilité » de la solution de stockage pour son extension
- modes « bloc » et « fichier » disponibles à discrétion

Pour les sauvegardes & archivages

- Pour la sauvegarde
  - une séparation physique du stockage primaire
  - une représentation la plus synchrone possible
- Pour l'archivage
  - Deux approches possibles :
    - un archivage basé sur le stockage originel ;
    - un archivage basé sur la sauvegarde.
  - Dans les deux cas, les archives sous la forme :
    - d'une série d'instantanés pris suivant une politique pré-établie
    - une copie complète sur un support tierce, archivée physiquement

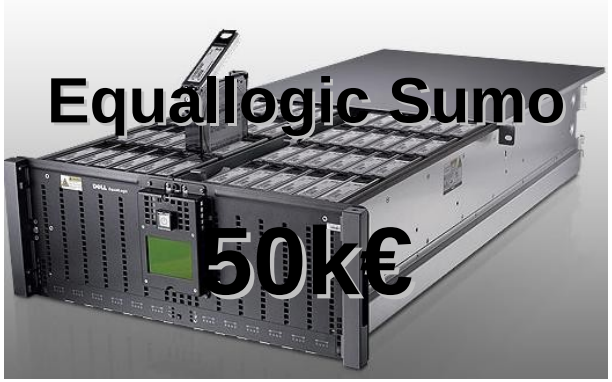
Des projections de l'enquête... A de nouvelles pratiques !

- Espace pour les laboratoires de biologie
  - Mise à disposition « lente » : presque 1 an pour un service incomplet...
    - Prolifération de disques nomades
    - Choix de protocoles « destructifs » (purge régulière des données brutes)
  - Accès réseau désastreusement lent (100 Mb/s)
  - Localité du traitement des données finalement efficace (en attendant)...
- Mais, nouvelles pratiques en croissance
  - Augmentation constante des données brutes
  - Séquençage : méthodes & traitements à cycle court (jetables)
  - Émergence d'un portail d'applications « standard » : Galaxy

Entre matériel, logiciel et Open Source : propositions de 2010

- Tout dans le logiciel (la boîte noire)
  - ISCSI : une « target », à chaque serveur de se raccrocher
  - NetApp/Isilon/PanaSAS/GPFS/…
- Tout dans le matériel :
  - Du RAID matériel : 3Ware, LSI, MegaRAID, ...
- Tout dans l'Open Source
  - **MD** Linux : RAID, **LVM** Linux : volumes physiques & logiques
  - ZFS, mais Solaris (ou OpenSolaris) sur x4500
    - Implémentation ZFS-Fuse : ben c'est du Fuse…
    - Implémentation ZFSonLinux : porté par LLNL & Brian Behlendorf
  - Des solutions distribuées : GlusterFS, CephFS, XtreemFS

Lorsque le budget invite à d'autres explorations...



- Avec un SAN (même iSCSI), il faut des frontales !
- Étude de ZFS comme socle d'éléments de SAN :
  - Natif (avec OpenSolaris) et ZFSonLinux
- Étude des systèmes de fichiers distribués :
  - AoE et iSCSI, GlusterFS, XtreemFS, CephFS

Maturation de ZFSonLinux

- Solution Open Source MDADM/LVM2
  - Mdadm pour créer, administrer les volumes RAID 0,1,5,6,10
  - LVM2 pour administrer les volumes logiques
    - Instantanés possibles, gestion des disques, du RAID (contraignant)
- Solution BTRFS : pas mature (à l'époque)
- Solution ZFSonLinux :
  - Couche d'abstration SPL : *Solaris Porting Layer*
  - Gestion des pools (groupes de disques) : zpool <options>
  - Gestion des volumes : zfs <options>

Match ZFS vs MDADM/LVM Petit exercice

- Assembler 5 disques en RAID :
  - 5 disques plus 1 spare
- Créer un volume en mode fichier
- Créer un volume disque de 100G

En MDADM/LVM : 8 commandes

1. mdadm --create /dev/md0 --level=5 --raid-devices=5 --spare-devices=1 /dev/sd[a-f]
2. mdadm --examine --scan >> /etc/mdadm/mdadm.conf
3. vgcreate raid5 /dev/md0
4. lvcreate -L100G -nmode-fichier raid5
5. mkfs.ext4 /dev/raid5/mode-fichier
6. mkdir /media/mode-fichier
7. mount -o noatime /dev/raid5/mode-fichier /media/mode-fichier
8. lvcreate -L100G -nmode-bloc raid5

En **ZFSonLinux : 3 commandes**

1. zpool create raid5 raidz /dev/sd[a-e] spare /dev/sdf
2. zfs create raid5 raid5/mode-fichier -o mountpoint=/media/mode-fichier
3. zfs create -V 100G raid5/mode-bloc

ZFSonLinux : ce qui kiffe grave !

- L'extensibilité : rajouter des pools à volonté !
  - Rajouter un pool avec des disques : zpool add MonPool /dev/MonNouveauDisque
- L'archivage :
  - zfs snapshot MonPool/MonVolume@MonSnapshot
- Le clonage : 2 étapes, l'instantané & le clonage
  - zfs snapshot MonPool/MonServeurISCSI@MonSnapshot
  - zfs clone MonPool/MonServeurISCSI@MonSnapshot MonPool/MonNouveauServeurISCSI
- La sauvegarde :
  - zfs send MonVolume@MonSnapshot | ssh MonServeur zfs receive MonBackup/MonServeur/MonVolume
- La compression : passer en lz4 par défaut
  - zfs set compression=lz4 MonVolume

ZFSonLinux : les trucs moins drôles

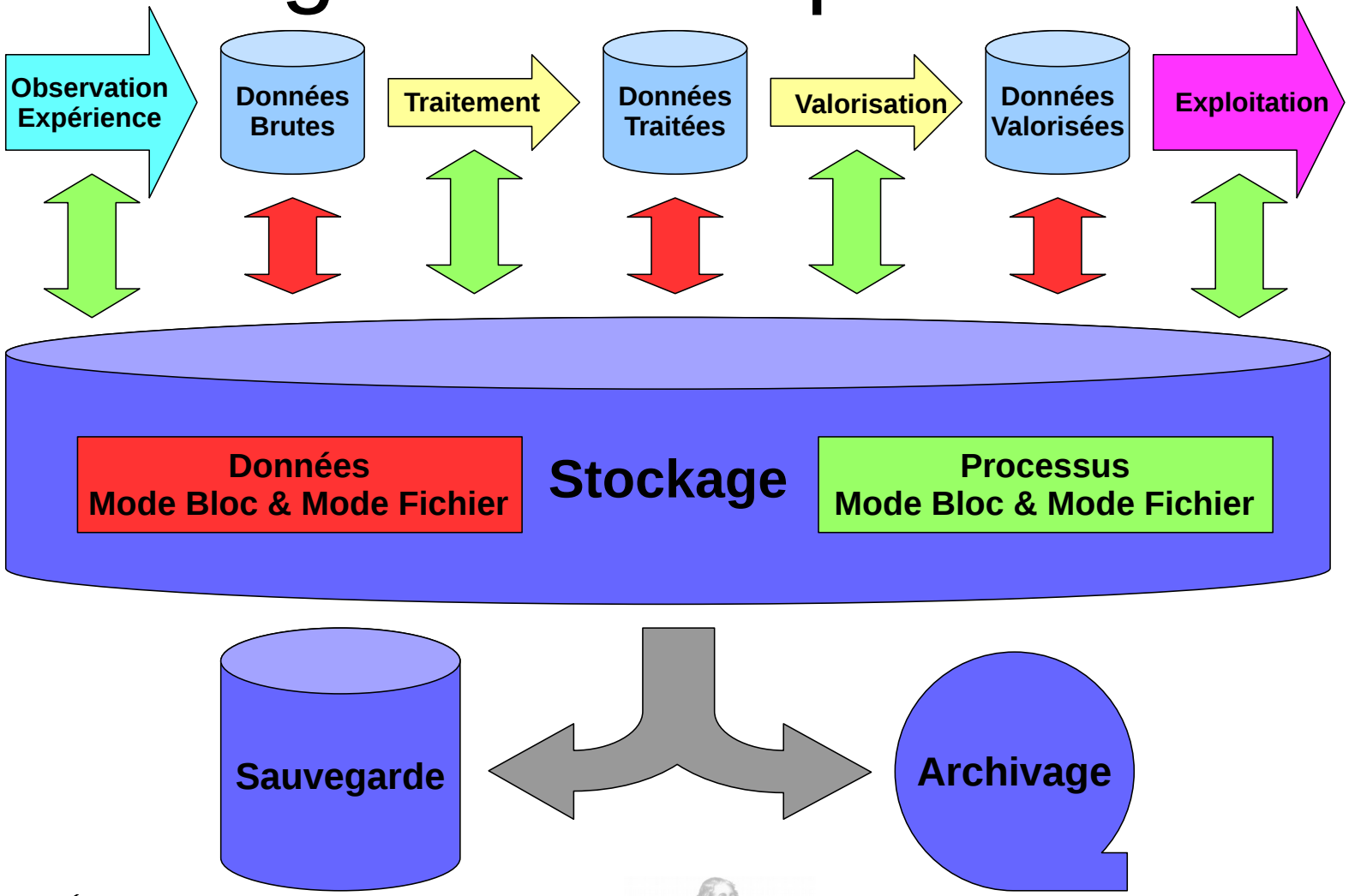
- Plus c'est rempli, moins c'est rapide
- Ne JAMAIS nommer ses disques /dev/sd? dans le pool
- Toujours régler le zfs_arc_max (50% RAM)
- Toujours créer un pool avec l'option ashift=12
- Des recommandations Solaris parfois overkill :
  - 1GHz & 1GB pour 1TB de disque brut
    - Mais en fait : 4 cœurs à 2.8GHz et 16GB de RAM pour 24TB brut
  - 60 disques de 4TB : 48 coeurs avec HT à 2.5GHZ et 256GB : cher !
    - Mais en reprenant les specs matériels Oracle : 16c 2GHz & 128GB RAM

ZFSonLinux : parfait !
En attendant BtrFS/CephFS

- Centre Blaise Pascal :
  - 306 disques ou partitions ZFS pour 381TB bruts
- Pôle Scientifique de Modélisation Numérique :
  - 600 disques pour 1.5PB bruts
    - Dont 700TB bruts sur 3 serveurs !
  - Bientôt 800 disques pour 2PB bruts
- Si certains sont intéressés… Formation à organiser !

Mais les usages « stockage » ont évolué !

De l'intégration des processus...



Des processus très « volatils » Aux « paillasses numériques »

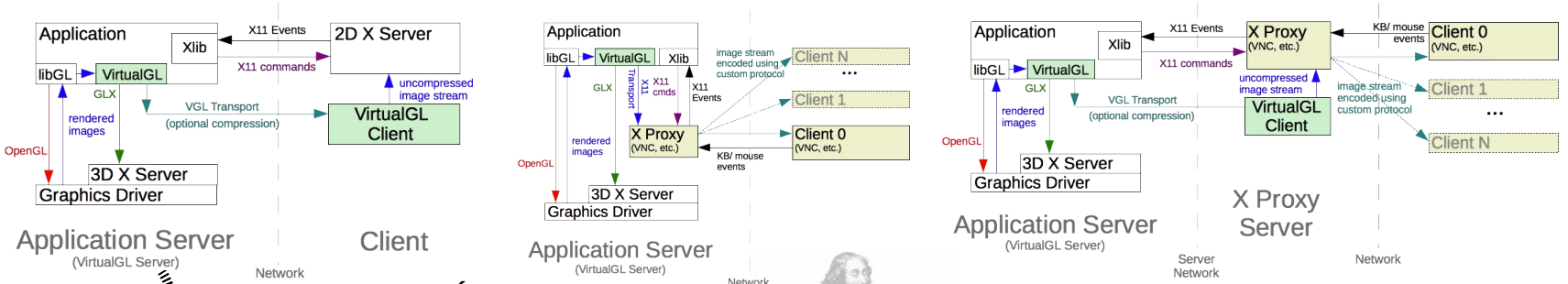
- La demande des « cœurs de métier » : du *AAS
  - Au départ du *Software*, mais virant vers les *Platform* ou *Infrastructure*
- Des socles de processus virtualisés ou déportés : IAAS
  - « mode fichier » pour *SIDUS* : *Read Only*
    - Machines physiques, KVM ou VirtualBox
    - NFSroot accessible par le réseau pour les stations de travail, les nœuds, les COMOD
    - Applications intégrées (SAAS) au système natif, accessibilité par x2go
  - « mode bloc » pour iSCSI et Virtual Disk : *Read/Write*
    - Machines physiques ou KVM : « Paillasses Numériques »
    - Flexibilité d'un accès complet, Sécurité d'une administration centralisée

Bureau à distance : un SAAS
Le couple : x2go/VirtualGL

- Pourquoi visualiser ?
  - Parce que la puissance d'analyse est derrière les yeux !
- Pourquoi accéder aux ressources à distance ?
  - Parce que les ressources ne sont pas accessibles 7j/7, 24h/24
  - Parce que c'est plus pratique pour suivre les évolutions...
- Quelles contraintes de la visualisation à distance :
  - Sur un canal RDP : multi-plateforme, assez efficace, restreint en 3D
  - Sur un canal TimeViewer : très efficace, mais double tunnel...
  - Sur un canal SSH : assez lourd, passage OpenGL, multi-plateforme difficile

VirtualGL : efficace mais pénible...

- *Shading* par la carte graphique
- Transport par canal SSH
- Utilisation :
  - Sous GNU/Linux, pas trop difficile :
    - Sur le poste client : vglconnect -s <MonLogin>@<MaVisu>
    - Sur le poste <MaVisu> : vglrun <MonAppli3D>
  - Sous les autres OS, bonne chance...



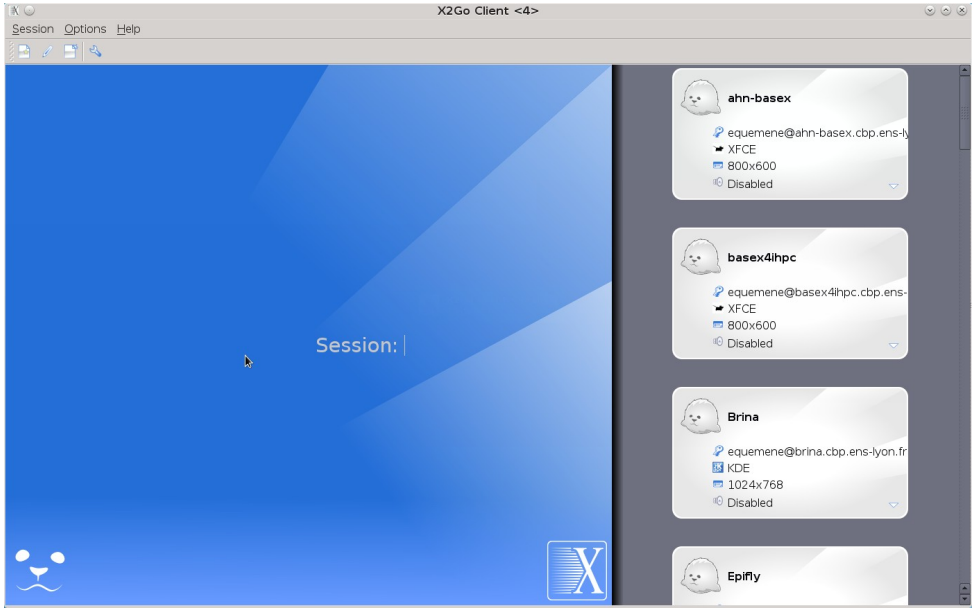
x2go : la transformation d'une session en vidéo

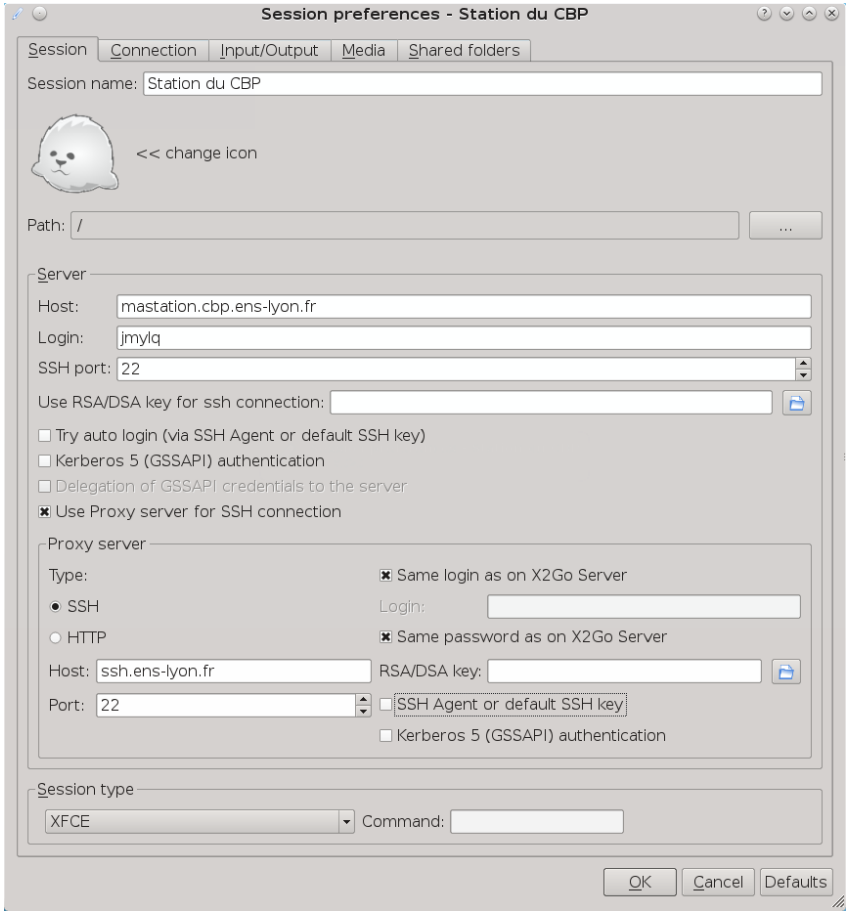
- 3 principes de base :
  - SSH, avec sa polyvalence, est le canal rêvé
  - Chaque image est transférée en jpg
  - Seule la modification de l'écran est transférée (comme MP4)
- Ses avantages :
  - Son multiplateforme (enfin presque)
  - Très faible bande passante utilisée
  - Passage de périphériques (son), espace de stockage, etc
  - Exploitation de MesaGL pour la 3D, mais insuffisante...

x2go & VirtualGL : enfin un mariage qui marche !

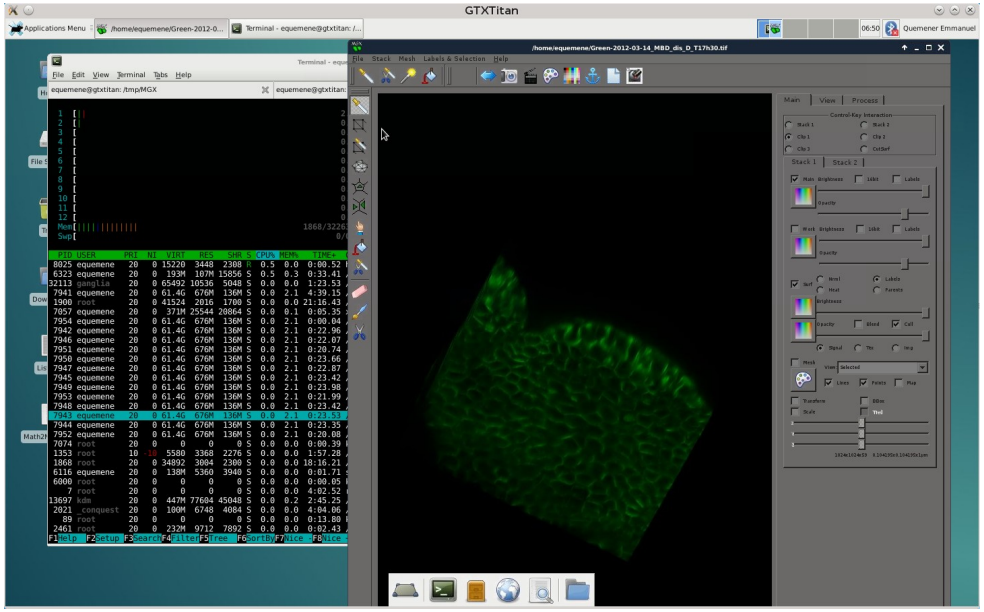
- Utiliser x2go pour :
  - Son multi-plateforme : Windows, MacOSX, GNU/Linux
  - Sa faible bande passante utilisée (300KB/s pour une vidéo HD)
  - Son passage de périphériques :
    - Même la carte son peut passer !
    - Mais pour MacOSX, inférieur à 8.5 pour passer les périphériques
- Utiliser VirtualGL pour :
  - Les grosses applications 3D & GPU (Cuda & OpenCL) : MorphoGraphX, Paraview, VMD, ...
  - Préfixer « juste » la commande par vglrun
- Comme ça :
  - Transfert des données primaires plus accessibles
  - Exploitation plus rationnelle des stations de travail

Connexion par x2go
Une configuration aisée...



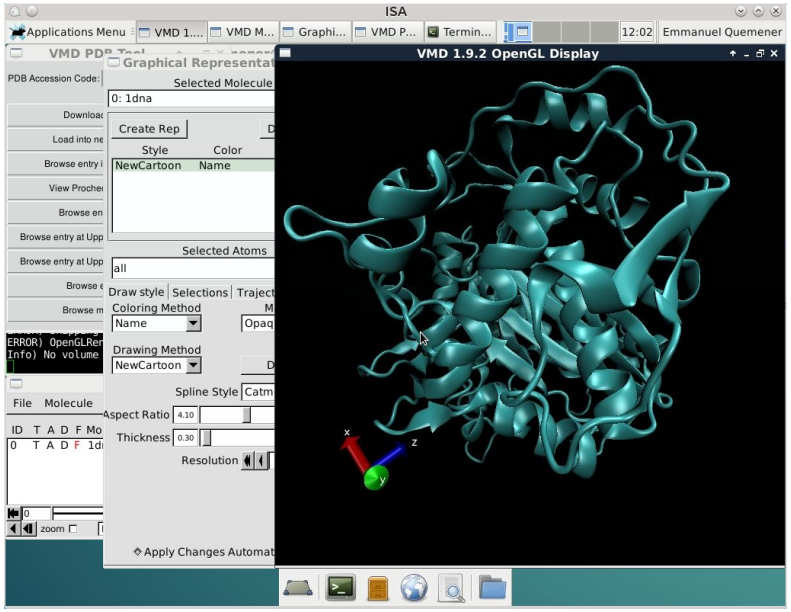


Exploitation x2go/VirtualGL
Pas seulement la visualisation...

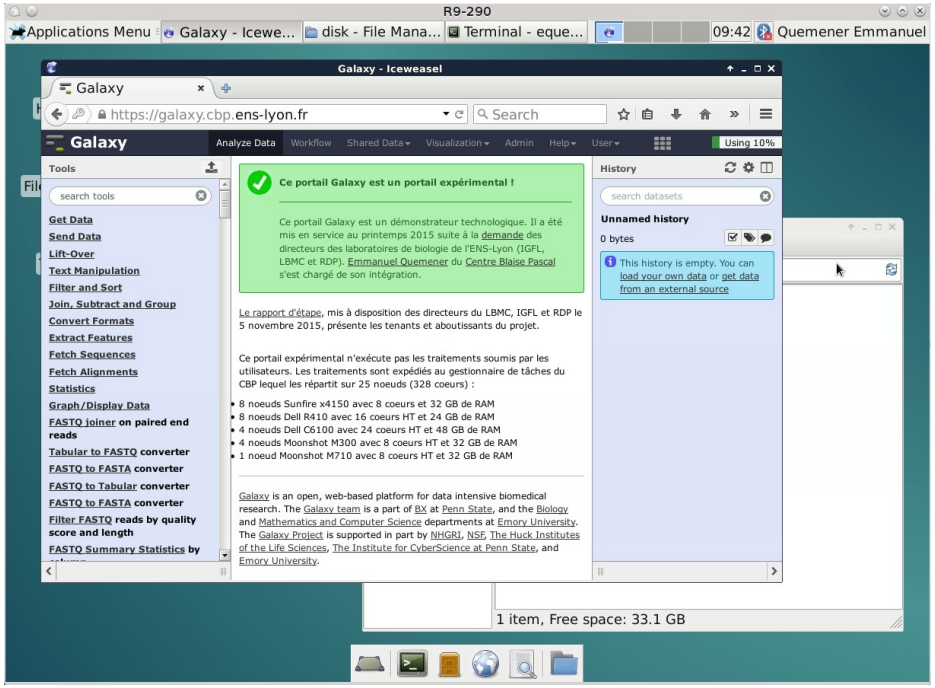


Traitement avec MorphoGraphX

Visualisation avec VMD

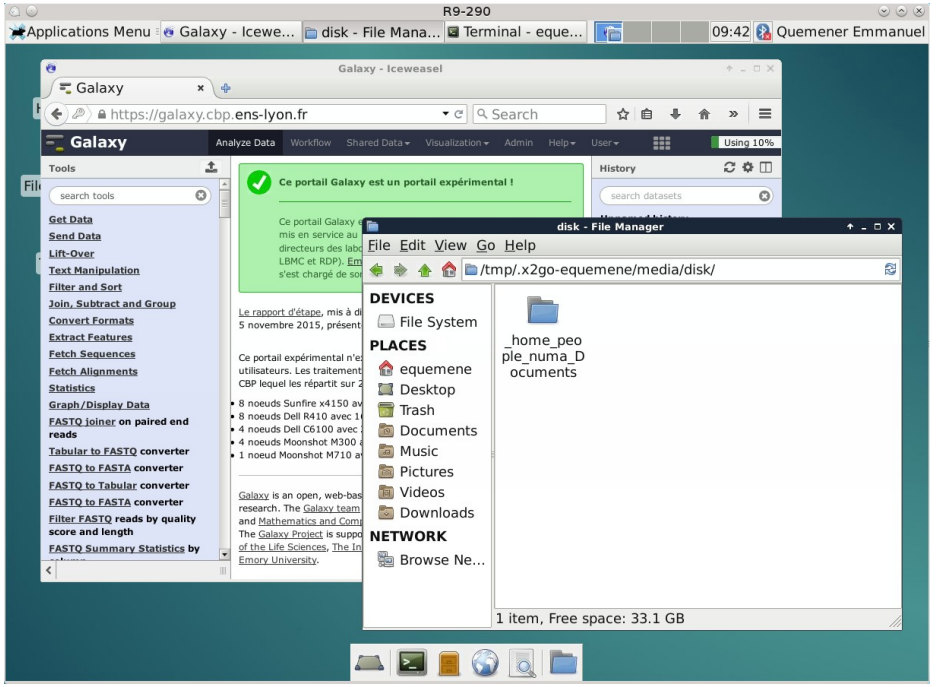


Connexion avec x2go
Le bureau ultime à distance ?



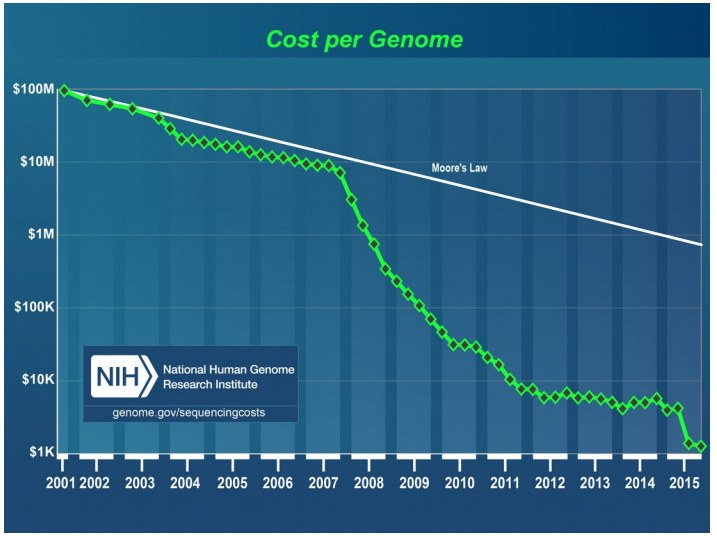
Accès
Portail Galaxy

Récupération
des Données



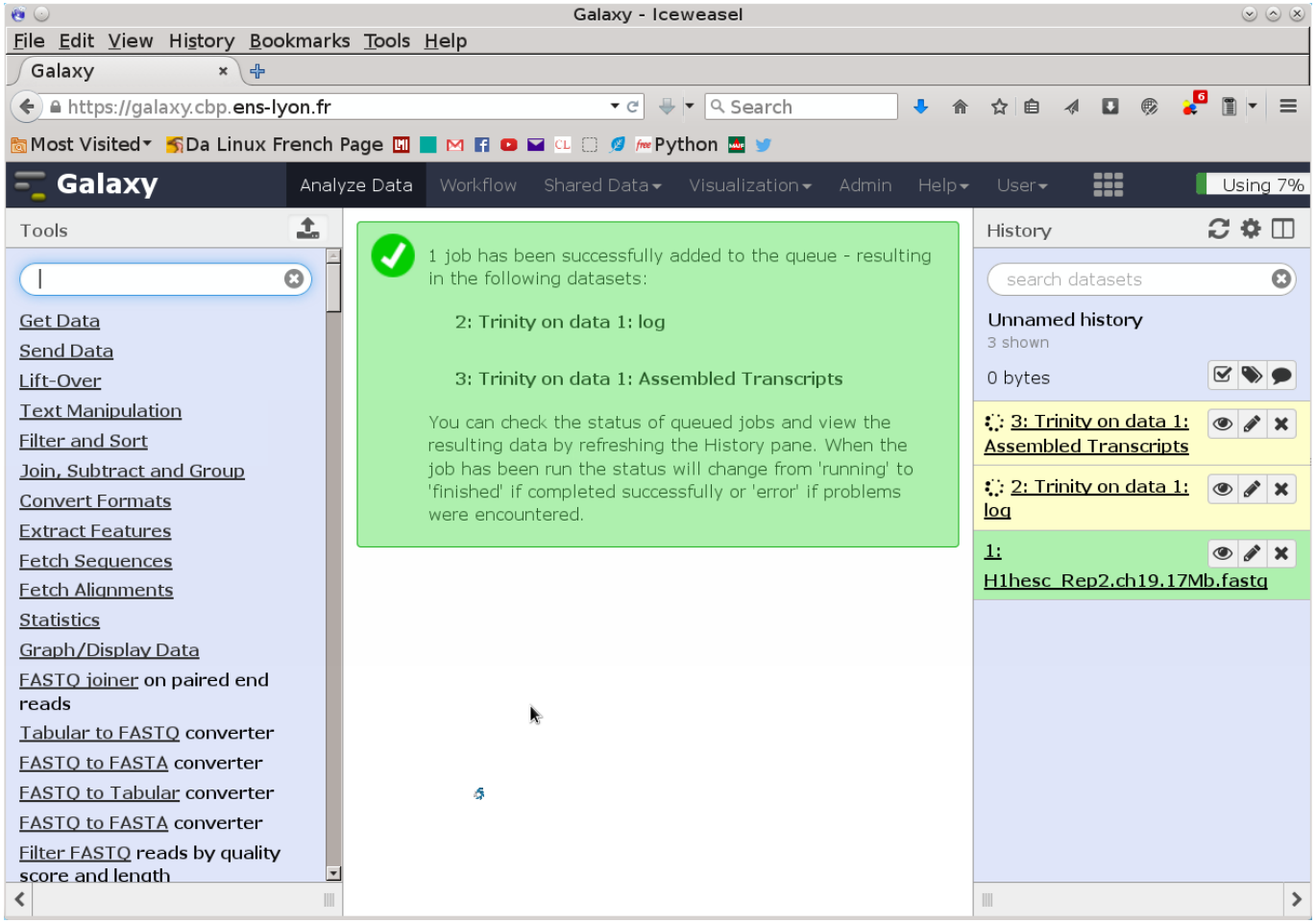
Pourquoi un portail Galaxy ?
Un PAAS pour la biologie (et +)

- Pourquoi ?
  - Coût séquençage /100000 en 15 ans
    - x10k pour les GPU, x30 pour les CPU
  - Ligne de commande : insurmontable
    - « Univers » clickodrome majoritaire.

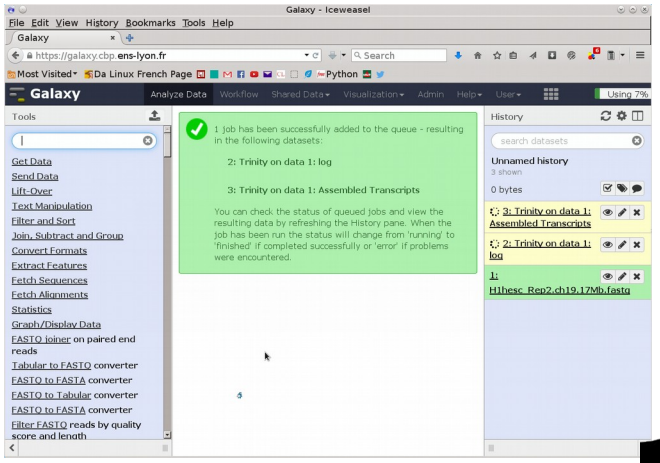


- Comment ?
  - Exploiter l'interface la plus « naturelle » : le navigateur Web
  - Réaliser une interface entre portail & applications existantes
    - Émergence d'un standard : Galaxy Project de Sandia

A quoi ça ressemble un « Portail Galaxy » ?



Galaxy@CBP : un Iceberg
Sur et sous la ligne de flottaison



- 4 serveurs physiques
- Une frontale virtuelle pour Galaxy
- Une frontale virtuelle pour les nœuds
- Une trentaine de nœuds
- Une dizaine d'équipements réseaux
- Une authentification DSI

Plate-forme Galaxy : le Graal ?

- « Galaxy, je l'ai installé sur mon portable, ça marche ! »
- « Pas de souci, ça s'installe tout seul ! »

Galaxy « personnel »

Galaxy « universel » ?

Pour un portail Galaxy universel
Quels outils ?

Construction d'un portail Galaxy

- Objectif : intégrer un portail Galaxy à un mésocentre
  - Ça veut dire :
    - Interfacer la passerelle à des ressources distribuées de type « cluster »
    - Distribuer les requêtes des utilisateurs sur les nœuds des clusters
    - Examiner le « comportement » en fonction de la charge
- Expériences : 4 déploiements successifs
  - Premier déploiement en local dans un dossier dédié
  - Second déploiement dans le dossier de l'utilisateur « galaxy »
  - Troisième déploiement dans un dossier partagé avec les nœuds
  - Quatrième déploiement (migration) vers une machine « confortable »

Autour du portail Galaxy : les autres services nécessaires

- Services locaux :
  - Portail Galaxy : serveur d'application en Python sur port particulier
  - Redirecteur (Proxy) Web : NGINX pour accès Web facilité
  - Serveur FTP : chargement de gros volumes de données
  - Serveur NFS : partage dossier Galaxy avec nœuds
- Services externes :
  - Authentification avec LDAP par ENS-Lyon & filtrage par login par CBP
  - Serveur de Batch avec GridEngine
  - Serveur SIDUS des nœuds

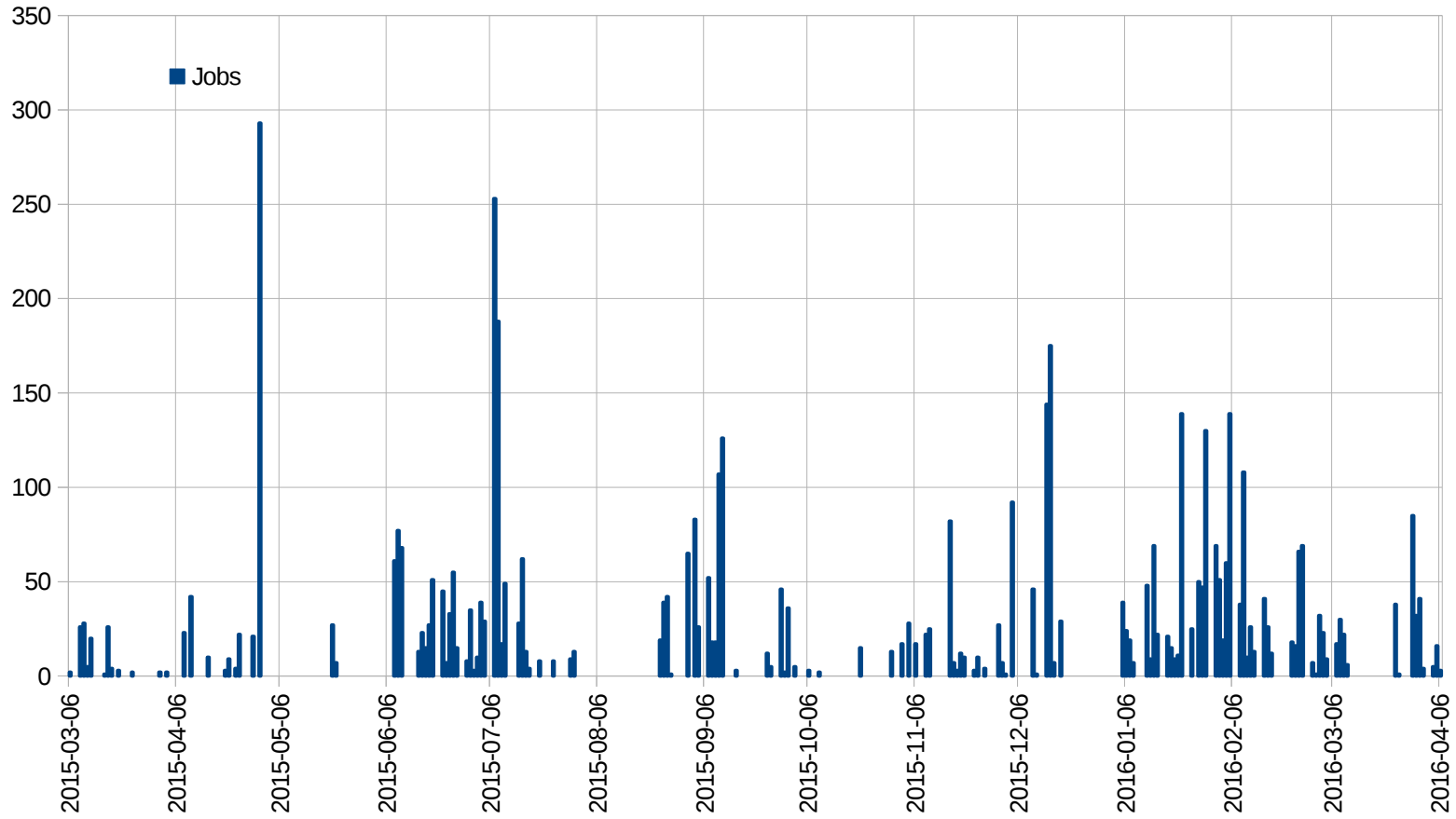
Galaxy pour Recherche & Enseignement
Mariage de la carpe & du lapin

- Recherche
  - Activité « creuse » : quelques utilisateurs simultanés
  - Volumes entrée/sortie difficiles à anticiper
  - Calculs de durée difficile à anticiper
- Enseignement
  - Activité « dense » : plusieurs dizaines d'utilisateurs simultanés
  - Volumes entrée/sortie raisonnable et prévisible
  - Calculs de durée raisonnable et prévisible : quelques minutes

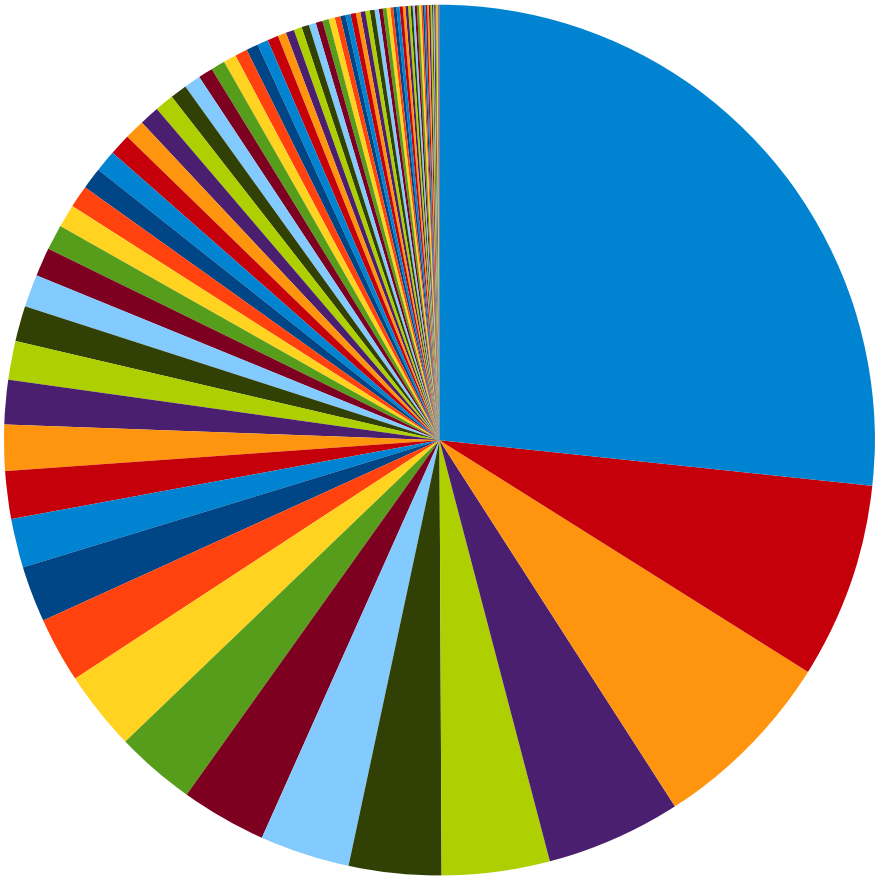
Soucis lors des expériences

- Exploration :
  - Choix du gestionnaire de batch : nécessité respect DRMAA
    - OAR : actuel CBP, pressenti PSMN : API non opérante
    - Slurm : très commun utilisé, API DRMAA inopérante (malgré la doc)
    - GridEngine : jamais utilisé au CBP, aide à la configuration PSMN
  - Dossier partagé « galaxy » contre le portail :
    - Mauvais choix du $HOME de l'utilisateur Galaxy : passage dans dossier
      - Déni de service sur la frontale des clusters CBP
    - Mauvais choix d'un volume disque « standard » :
      - Réactivité des disques insuffisante à la sollicitation
    - Mauvais choix d'un réseau Gigabit Ethernet :
      - Passage sur plate-forme de production à l'IB

Galaxy : après 1 an et 1 mois
Jobs : 5192 de « production »

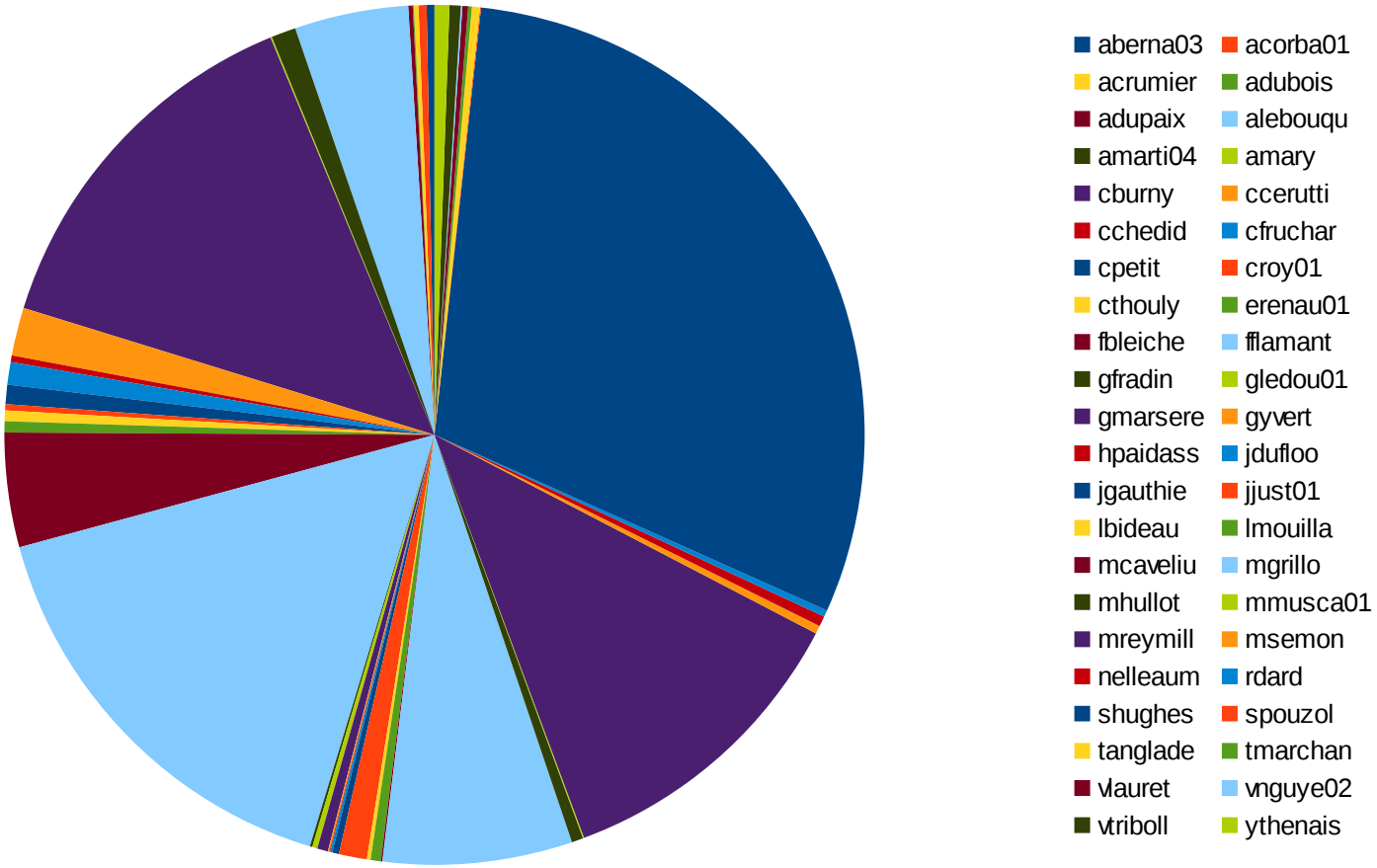


Galaxy : après 1 an et 1 mois
Applications : 106 utilisées



cshl_fastx_reverse_complement
peakcalling_macs
tabular_to_fastq
cshl_fastq_to_fasta
sam_bw_filter
fastq_quality_trimmer
Show_beginning1
cshl_fastq_quality_converter
Paste1
trinityrnaseq
cshl_fastx_trimmer
trimmer
wc_gnu
samtools_flagstat
samtools_mpileup
tophat2
mira_4_0_de_novo
Add_a_column1
ucsc_table_direct1
cshl_fastq_quality_filter
fasta_compute_length
cat1
fastq_qual_convert
tab2fasta
fastq_stats
fastq_groomer
Grep1
fastqc
Fetch_Taxonomic_Ranks
sam_merge2
barchart_gnuplot
cshl_fastx_collapser
wiggle2simple1
runMapping
bedtools_bamtobed
CONVERTER_sam_to_bam
sort1
CONVERTER_bed_to_gff_0
ebi_sra_main
bedtools_intersectbed
Remove_beginning1
bedtools_sortbed
addValue
cshl_fastx_nucleotides_distribution
samtool_filter2
CONVERTER_interval_to_bed_0
fasta2tab
fasta_filter_by_length
comp1
ncbi_makeblastdb
Filter1
ncbi_blastn_wrapper
bowtie2
bam_to_sam
bedtools_multiintersectbed
join1
Paired_fastQ_trimmer
seq_filter_by_mapping
Count1
EXPORT_HISTORY
CONVERTER_Bam_Bai_0
samtools_filter
bedtools_overlapbed
deseq2_hts
Summary_Statistics1
cshl_fastx_clipper
mira_4_0_mapping
revcompl
runAssembly
fastq_to_tabular
Cut1
edgeR
cshl_fastq_quality_boxplot
Convert_characters1
cshl_fastx_quality_statistics
cshl_fastx_barcode_splitter
cshl_fasta_formatter
SET_METADATA
fastq_to_fasta_python
Grouping1
cutadapt
fastq_filter
htseq_count
upload1

Galaxy : après 1 an et 1 mois
Utilisateurs : 44 différents mais 5 gros



Galaxy : et la suite ?

- Fonctionnellement :
  - Ouverture à plus de personnels, voire hors site
  - Délégation auprès de bioinformaticiens pour les greffons Sandia
  - Création de greffons pour d'autres communautés
- Techniquement :
  - Consolidation de la frontale : CPU/RAM/Espace disque/Réseau
  - Changement du gestionnaire de tâches : GridEngine vers Slurm
  - Consolidation du réseau entre frontale & nœuds : InfiniBand
  - Extension du nombre de nœuds de traitement

Applications « raoul » de biologie Ou la scalabilité des traitements...

- Faire tourner la suite Repeat(Masker|Modeler)
  - Processus de (3|6) « produits » très hétérogènes
  - A petite échelle (portable ou station de travail) : OK
  - A plus grande échelle, passage au PSMN : KO
    - Plantage du serveur NFS (avec un Loïs « zorgieusement désappointé »)
  - Exploration du « comportement » au CBP
    - Lancement sur station de travail, disque SSD
    - Lancement sur nœud de cluster, espace GlusterFS
    - Lancement sur serveur rapide, espace Ramdisk

Traitements RepeatModeler 2013
Le gros « stress » des GlusterFS

2013	Bases	Sequences	Files Temp	User	Elapsed	Input/Ouput
Pelodiscus	2202483752	19904	**309531**	620394	394995	2073771400
Tetraodon	358618246	27	**403873**	1040280	453677	1684136336
Killifish	1230898532	6012	**376817**	773775	531665	2531555536
2015						
Amphiprion	870234352	1081012	**1193669**	1244091	1036729	530305664
Roussette	4311024850	3502619	**743440**	1662646	1906300	271617672

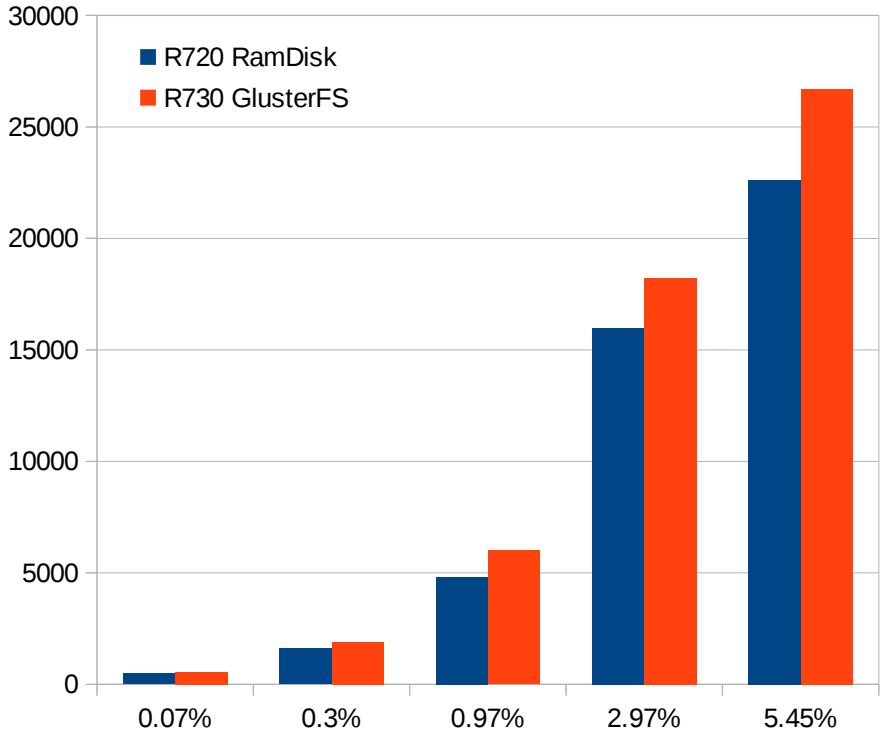
En regardant plus finement avec « /usr/bin/time »

2013	User	System	Elapsed	Input/Ouput	Data Rate IO/s
Pelodiscus	620394	2209	394995	2073771400	**5250**
Tetraodon	1040280	1565	453677	1684136336	**3712**
Killifish	773775	2316	531665	2531555536	**4762**
2015					
Amphiprion	1244091	193623	1036729	530305664	512
Roussette	1662646	372336	1906300	271617672	142

Traitement RepeatModeler 2015
Le Match : RamDisk vs GlusterFS

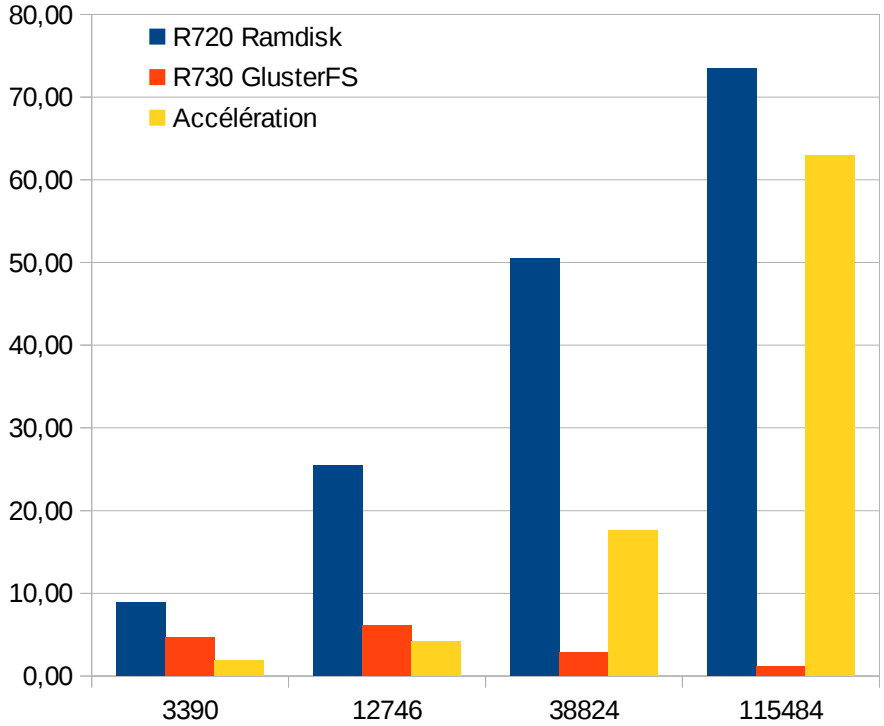
Progression « Input Database Coverage »

Less is Better !



Progression « Family Refinement »

Best is Better !



Quel enseignement de repeat* ? A chaque traitement son infra...

- La panacée n'existe pas :
  - en stockage, en traitement, en « modes » (bloc & fichiers)
- Deux nécessités :
  - Exploration la plus fine possible sur ensemble « pertinent »
  - Diversité des installations pour explorer les différentes approches
  - Application du « big data » sur le traitement des data elles-mêmes
- Et quelles implémentations ?

Quels enjeux des données ? Mais pas seulement...

- Uniquement pour les données ?
  - Pour toutes les sciences : un principle universel, la causalité
    - « Les mêmes causes produisent les mêmes effets. »
    - Un corollaire à la causalité : la reproductibilité
  - Reproductibilité : enjeu « stratégique » pour la Maison Blanche
    - Controverses scientifiques : pas seulement les résultats
      - Pas seulement les données brutes : la méthodologie, les outils, etc.
  - Observations, expériences, traitements, valorisation : protocole
- Pérennité
  - Éviter le syndrome des « bandes Apollo »
- Reproductibilité :
  - Pouvoir « rejouer » les traitements voire publier la « machine » complète.

White House takes notice of reproducibility in science, and wants your opinion

with 31 comments

The White House's Office of Science and Technology Policy (OSTP) is taking a look at innovation and scientific research, and issues of reproducibility have made it onto its radar.

Here's the description of the project from the *Federal Register*:

> The Office of Science and Technology Policy and the National Economic Council request public comments to provide input into an upcoming update of the *Strategy for American Innovation*, which helps to guide the Administration's efforts to promote lasting economic growth and competitiveness through policies that support transformative American innovation in products, processes, and services and spur new fundamental discoveries that in the long run lead to growing economic prosperity and rising living standards. These efforts include policies to promote critical components of the American innovation ecosystem, including scientific research and development (R&D), technical workforce, entrepreneurship, technology commercialization, advanced manufacturing, and others. The strategy also provides an important framework to channel these Federal investments in innovation capacity towards innovative activity for specific national priorities. The public input provided through this notice will inform the deliberations of the National Economic Council and the Office of Science and Technology Policy, which are together responsible for publishing an updated *Strategy for American Innovation*.

Conclusion

- « Inventaire à la prévert » des activités ?
  - Retour d'expérience de 6 années d'activités « tous azimuts »
- CBP @ ENS-Lyon : approche pragmatique
  - Toujours avec la perspective de sa généralisation (au PSMN)
  - Toujours ascendante et progressive
  - Toujours orientée vers la simplicité et la liberté
- Les « données » indissociables des « processus »
  - Conservation des « processus » comme les « données »