



Contexte astrophysique

SDSS

BOSS

BOSS au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques
matérielles et logicielles

Avantages de
l'infrastructure

Difficultés rencontrées

Conclusions

Conclusion

Perspective

Retour d'expérience d'utilisation d'un Cloud en Astrophysique : le projet BOSS

Cécile Cavet,
Nicolás Busca, Michèle Detournay

Centre François Arago (FAcE), Laboratoire APC, Université Paris Diderot

27 Novembre 2012



Plan

Contexte astrophysique

SDSS
BOSS

BOSS au laboratoire APC

Code de calcul
Données en entrée
Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles
Avantages de l'infrastructure
Difficultés rencontrées

Conclusions

Conclusion
Perspective

- 1 Contexte astrophysique
- 2 BOSS au laboratoire APC
- 3 Utilisation du Cloud StratusLab
- 4 Conclusions

Relevés SDSS (Sloan Digital Sky Survey)

Contexte astrophysique

SDSS

BOSS

BOSS au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

Difficultés rencontrées

Conclusions

Conclusion

Perspective

- **SDSS I** : 1999 - 2005 ➔ 47 000 objets.
- **SDSS II** : 2005 - 2008 ➔ 1 M de galaxies et 120 000 quasars.
- **SDSS III** : 2009 - 2014 ➔ 4 expériences :
 - SEGUE-2 (halo de la Voie Lactée)
 - APOGEE (centre galactique de la Voie Lactée)
 - MARVELS (planètes extra-solaires)
 - **BOSS** (cosmologie)



Figure: Télescope SDSS. Au sol, à l'APO, 2,5 m de \varnothing . ©The Sloan Digital Sky Survey.

Expérience **BOSS** (Baryon Oscillation Spectroscopic Survey)¹

Contexte astrophysique

SDSS

BOSS

BOSS au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

Difficultés rencontrées

Conclusions

Conclusion

Perspective

- Cosmologie : étude de l'énergie sombre.
- Grand relevé photométrique et spectral : en 2014, 1,5 M de galaxies et 160 000 quasars ➔ 1/4 du ciel.
- Stockage des données :
 - À terme : 250 GB ➔ pas d'enjeu de stockage.
 - Base de données SQL.

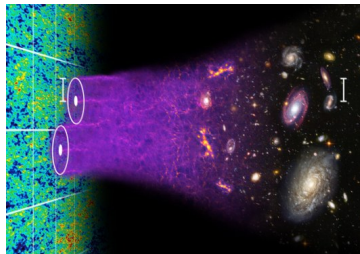


Figure: Illustration de l'effet des oscillations acoustiques de baryons sur le jeune univers (WMAP) identifiable aujourd'hui (relevé **BOSS**) © C. Blake & S. Moorfield.

¹<http://www.sdss3.org/surveys/boss.php>

Code de calcul porté sur

- Spectres de quasar → FFT → Spectre de puissance
- Language : JAVA
- Parallelisation : Parallel JAVA
- Algorithmme : FFT
 - Opération sur des tableaux 3D de données.
 - 25 GB de mémoire au minimum.
- En phase de développement (pas d'optimisation).

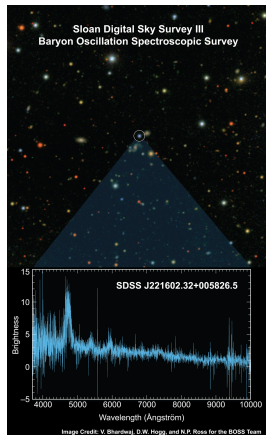


Figure: Spectre de quasar.

Données en entrée : spectres de quasars

Contexte astrophysique

SDSS

BOSS

BOSS au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

Difficultés rencontrées

Conclusions

Conclusion

Perspective

- Type de spectre : forêt Lyman- α \rightarrow $\sim 1/6$ du spectre total sélectionné en pré-processing.

■ Données :

- Cube d'Univers
- Discrétisation : 1024 pixels cubiques
- 1 pixel cubique = 1 spectre ou \emptyset .
- Volume : 60 000 fichiers FITS (700 MB) \rightarrow à terme 160 000 fichiers.

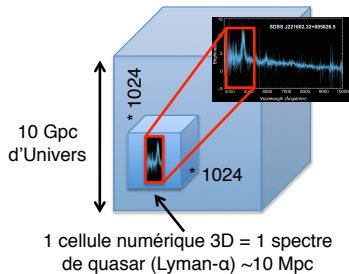
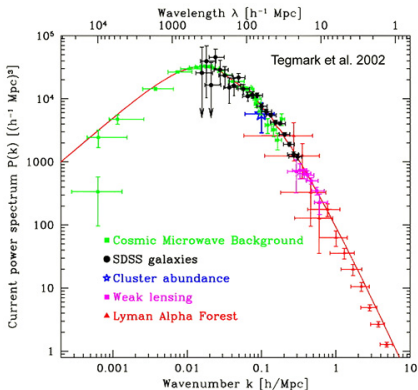


Figure: Cube d'Univers (forêt Lyman- α des spectres de quasar).

Données en sortie : spectre de puissance

- Données :
 - Détermination de nouveaux points et amélioration de l'incertitude sur les points existants.





StratusLab : caractéristiques matérielles et logicielles requises I

Contexte astrophysique

SDSS

BOSS

BOSS au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

Difficultés rencontrées

Conclusions

Conclusion

Perspective

Image disque référencée sur le [MarketPlace](#) :

- Identifiant : HDKRNwTo_jOy305dkvswqAoSDG1
- Hyperviseur (Virtualiseur) : VirtualBox.
- 10 GB d'espace disque.
- Scientific Linux 6.3
- Contextualisée pour StratusLab (stratuslab-one-context...).
- Paquets Linux scientifiques : JAVA et Screen (gestionnaire de fenêtre).
- Sauvée en format qcow2.
- Stockée sur le site Web de transfert de fichiers de l'APC.



StratusLab : caractéristiques matérielles et logicielles requises II

FFT 3D ➔ 25 GB de mémoire au minimum

Machine virtuelle :

- 28 GB de mémoire
- 18 cœurs de calcul ➔ parallélisation sur le maximum de cœurs de calcul possible.
- 4 GB de swap

Contexte astrophysique

SDSS

BOSS

BOSS au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

Difficultés rencontrées

Conclusions

Conclusion

Perspective



StratusLab : avantages de l'infrastructure I

Contexte astrophysique

SDSS

BOSS

BOSS au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

Difficultés rencontrées

Conclusions

Conclusion

Perspective

Environnement :

- Création du compte pour utiliser **StratusLab** ➔ immédiat !
- Instantiation et connexion SSH à une machine virtuelle ➔ en quelques minutes !
- Durée de vie de la machine virtuelle : infinie ! (ou presque)



StratusLab : avantages de l'infrastructure

II

Contexte astrophysique

SDSS
BOSS

BOSS au laboratoire APC

Code de calcul
Données en entrée
Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

Difficultés rencontrées

Conclusions

Conclusion
Perspective

Machine virtuelle :

- 28 GB de mémoire.
- Utilisateur **Root** ➔ très pratique pour administrer sa machine virtuelle (installer des paquets Linux, vérifier la configuration du système...)
- Portage et exécution du code ➔ instantané !
- Temps d'exécution du calcul : similaire à une machine physique dédiée à cette tâche et équivalente en CPU et en mémoire ➔ aucune dégradation de performance.
- Toutes les erreurs d'administration système sont permises... (la machine virtuelle subissant un incident est tuée).



StratusLab : difficultés rencontrées

Contexte astrophysique

SDSS

BOSS

BOSS au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

Difficultés rencontrées

Conclusions

Conclusion

Perspective

■ Disponibilité des ressources :

- une requête d'une machine virtuelle de **28 GB** de mémoire : **24 cœurs** ➔ reste en **pending**
18 cœurs ➔ passe en **running**.
- Jusqu'à combien de cœurs de calcul l'utilisateur a-t-il accès ?

■ Description des ressources :

- Quelles sont les caractéristiques des machines physiques supportant le Cloud : réseau, CPU, mémoire, stockage, etc.
- Limitation des ressources virtualisées dépend des machines physiques.

Solution ➔ Monitoring + Caractéristiques matérielles



Conclusion I

Contexte astrophysique

SDSS

BOSS

BOSS au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

Difficultés rencontrées

Conclusions

Conclusion

Perspective

Le Cloud **StratusLab** est bien adapté à cette application astrophysique.

- Répond bien à une requête spécifique en terme de ressources de calcul.
- Environnement de travail sur la machine virtuelle similaire à l'environnement de travail usuel.
- Aucune perte de temps lors du portage du code.
- Aucun problème technique du Cloud StratusLab pendant la période d'utilisation.
- Utilisateur heureux... :-)



Conclusion II

Contexte astrophysique

SDSS
BOSS

BOSS au laboratoire APC

Code de calcul
Données en entrée
Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles
Avantages de l'infrastructure
Difficultés rencontrées

Conclusions

Conclusion
Perspective

Mais une optimisation du code de calcul peut aussi permettre l'accès à d'autres infrastructures de calcul distribué.

- Changement de langage de programmation et/ou interfaçage
- Parallélisation
- Pré-processing...

Dans le contexte de la recherche en astrophysique, l'infrastructure de Cloud est bien adaptée au prototypage et à la phase de production du code de calcul quand l'obtention de résultats doit être rapide.

En perspective

Contexte astrophysique

SDSS

BOSS

BOSS au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

Difficultés rencontrées

Conclusions

Conclusion

Perspective

- Reprise prochaine de ce projet sur **StratusLab**.
- Amélioration du spectre de puissance (pour la partie forêt Lyman- α) peut-être grâce à **StratusLab** ? (à suivre...)
- Utilisation de données issues de simulations numériques cosmologiques \rightarrow création d'un disque de stockage de **1 TB** sur le serveur de **disque persistant** (pdisk).

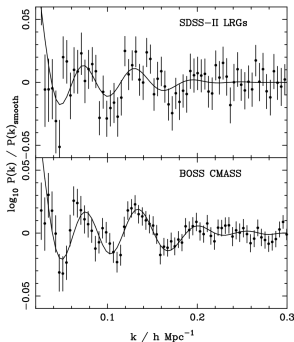


Figure: Logarithme du spectre de puissance pour SDSS II et SDSS III.

Merci pour votre attention.

Contexte astrophysique

SDSS

BOSS

BOSS au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Utilisation du Cloud StratusLab

Caractéristiques
matérielles et logicielles

Avantages de
l'infrastructure

Difficultés rencontrées

Conclusions

Conclusion

Perspective

