

## Contexte astrophysique

Mission Euclid

Galaxies

## Au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques  
matérielles et logicielles

Avantages de  
l'infrastructure

## Conclusions

Conclusion

Perspective

# Utilisation du Cloud StratusLab dans le cadre d'application astroparticule à l'APC

**Cécile Cavet & Michèle Detournay**

Centre François Arago (FACe), Laboratoire APC, Université Paris Diderot  
LabEx UnivEarthS

28 Mai 2013



# Plan

## Contexte astrophysique

Mission Euclid  
Galaxies

## Au laboratoire APC

Code de calcul  
Données en entrée  
Données en sortie  
Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles  
Avantages de l'infrastructure

## Conclusions

Conclusion  
Perspective

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

## Contexte astrophysique

### Mission Euclid

Galaxies

## Au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

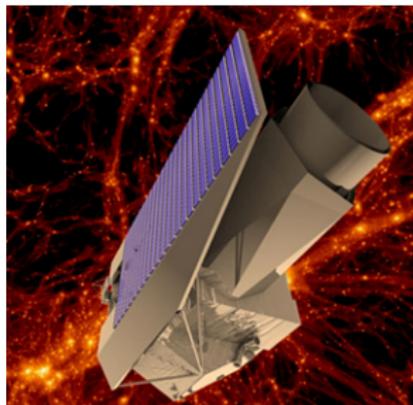
Avantages de l'infrastructure

## Conclusions

Conclusion

Perspective

- Origine de l'accélération de l'expansion de l'Univers.
- **Mission** de classe moyenne de l'ESA.
- Lancement en  $\sim 2020$  et  $\sim 10$  ans de relevé.
- Thématiques :
  - Cosmologie
  - Évolution des Galaxies
  - ...



**Figure:** Prototype du télescope spatial Euclid de 1,2 m de  $\phi$ .  
 © ESA.

# Étude des Galaxies

## Contexte astrophysique

Mission Euclid

Galaxies

## Au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

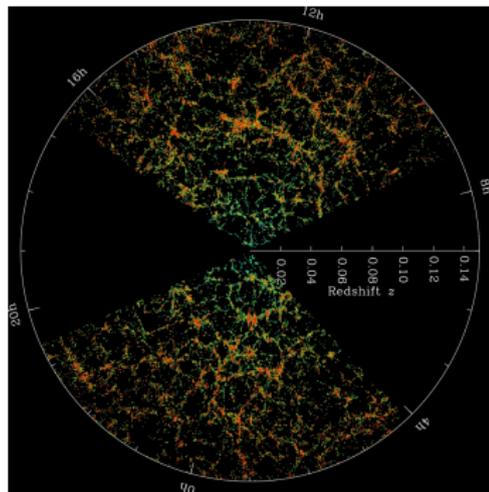
Avantages de l'infrastructure

## Conclusions

Conclusion

Perspective

- Grand relevé photométrique :
  - $\sim 1/4$  du ciel.
  - Nombre de Galaxies :  $N_G = 1,5 \times 10^9$ .
- Observables :
  - Morphologie  $\rightarrow$  mesure directe.
  - Distribution (*via* le décalage spectral  $z$ )  $\rightarrow$  nécessite un code de traitement de données.



**Figure:** Carte de la distribution des Galaxies de l'Univers proche ( $z_{max} = 0,15$ ) mesurée par le relevé SDSS. © SDSS.

# Code de calcul Le PHARE (PHotometric Analysis for Redshift Estimations)

## Contexte astrophysique

Mission Euclid  
Galaxies

## Au laboratoire APC

Code de calcul  
Données en entrée  
Données en sortie  
Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles  
Avantages de l'infrastructure

## Conclusions

Conclusion  
Perspective

- Détermination des décalages spectraux photométriques.
- Language : Fortran.
- Algorithmes :
  - Opérations classiques.
  - Minimisation de  $\chi^2$   
➔ Librairie MINUIT.
- Séquentiel.
- En phase de développement (pas d'optimisation).

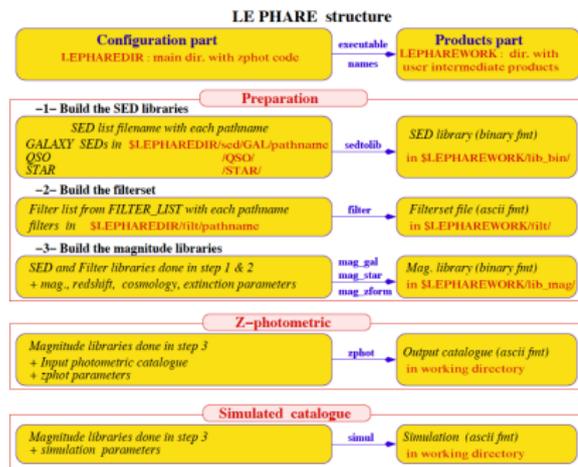


Figure: Structure du code de traitement de données Le PHARE. © Arnouts & Ilbert.

# Données en entrée

- Données < 1 GB.
- Bibliothèque de données ancillaires.
- Librairies et catalogues sauvegardés lors d'étapes de calcul intermédiaires.

## Contexte astrophysique

Mission Euclid

Galaxies

## Au laboratoire APC

Code de calcul

**Données en entrée**

Données en sortie

Pipeline

## Utilisation du Cloud StratusLab

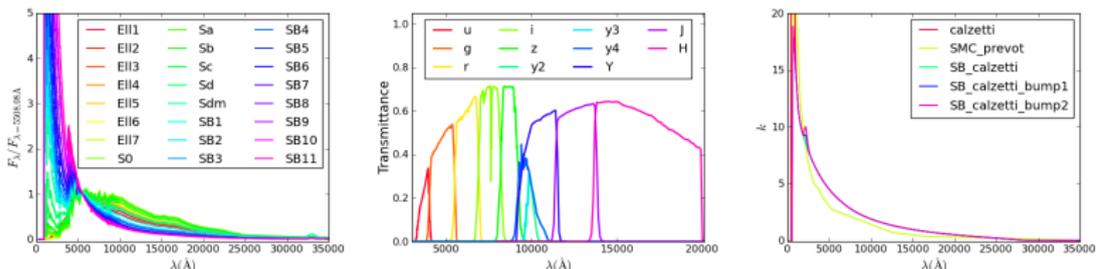
Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

## Conclusions

Conclusion

Perspective



**Figure:** Données ancillaires : spectres références de Galaxies, transmission des filtres de télescopes (LSST et Euclid) et extinction galactique.

# Données en sortie

- Catalogue de Galaxies et autres informations.
- Étude en cours :  $N_G \sim 5 \times 10^4$  Galaxies  $\rightarrow$  4 GB d'espace disque.
- À terme :  $N_G = 1,5 \times 10^9$  Galaxies  $\rightarrow$  au moins  $\sim 100$  TB à stocker.

$\rightarrow$  Le volume des données produites est la contrainte principale du portage de l'application.

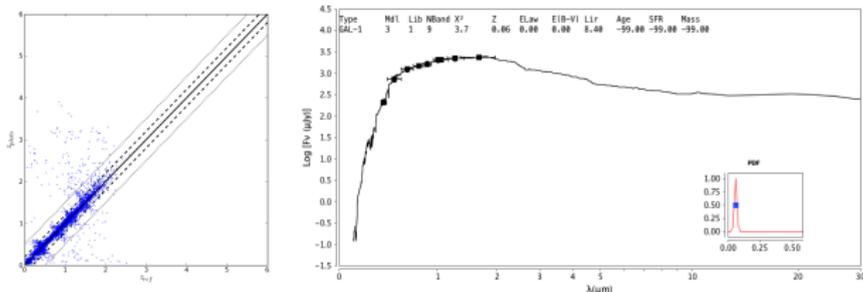


Figure: Données en sortie : décalages spectraux photométriques et spectre.

# Pipeline Photo-z avec Pipelet

## Contexte astrophysique

Mission Euclid  
Galaxies

## Au laboratoire APC

Code de calcul  
Données en entrée  
Données en sortie

## Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles  
Avantages de l'infrastructure

## Conclusions

Conclusion  
Perspective

## ■ Chaîne de traitement de données :

- Etude paramétrique.
- Exécution des différentes étapes du code **Le PHARE** + Visualisation.

- Language : Python.
- Base de données SQL.
- Interface Web.

➔ **Le produit de la base de données SQL doit être conservé.**

Figure: Interface Web du pipeline Photo-z utilisant l'outil Pipelet.

## Contexte astrophysique

Mission Euclid

Galaxies

## Au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques  
matérielles et logicielles

Avantages de  
l'infrastructure

## Conclusions

Conclusion

Perspective



# stratuslab

- Image disque de base
- Image disque spécialisée
  - Disque persistant
  - Machine virtuelle



# StratusLab : Image disque de base

## Contexte astrophysique

Mission Euclid  
Galaxies

## Au laboratoire APC

Code de calcul  
Données en entrée  
Données en sortie  
Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

## Conclusions

Conclusion  
Perspective

- Référencée sur le [MarketPlace](#) ➔  
Identifiant : `CM4KOK22Im9zs5PwW6Vu37SkjSI`
- Hyperviseur (Virtualiseur) : VirtualBox.
- 10 GB d'espace disque.
- Scientific Linux 6.3.
- Contextualisée pour [StratusLab](#) et avec Cloud-Init (utilisation multi-Cloud).
- Paquets Linux : `acpid`, `open-ssh`...
- Sauvée en format `qcow2`.
- Stockée sur le site Web de transfert de fichiers de l'APC.



# StratusLab : Image disque spécialisée

## Contexte astrophysique

Mission Euclid  
Galaxies

## Au laboratoire APC

Code de calcul  
Données en entrée  
Données en sortie  
Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

## Conclusions

Conclusion  
Perspective

- Référencée sur le [MarketPlace](#) ➔  
Identifiant : NYUUiZmum8Tj77kMhkxbPwxl5tA
- Créée avec le client [StratusLab](#).
- Mêmes caractéristiques que l'image disque de base.
- Paquets Linux scientifiques : compilateur, Python et ses modules (Ipython : visualisation; CherryPy : interface Web...).
- Création d'un utilisateur générique.
- Logiciels : le code [Le PHARE](#), l'outil [Pipelet](#) (Gitorious), le pipeline [Photo-z](#) (en développement).
- Stockée sur le serveur pdisk de [StratusLab](#).



# StratusLab : Disque persistant

## Contexte astrophysique

Mission Euclid

Galaxies

## Au laboratoire APC

Code de calcul

Données en entrée

Données en sortie

Pipeline

## Utilisation du Cloud StratusLab

### Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

## Conclusions

Conclusion

Perspective

- Privé.
- Créé avec le client [StratusLab](#).
- Au moins **4 GB** d'espace disque.
- Formaté en Ext 3.
- Stocké sur le serveur pdisk de [StratusLab](#).
- Permet de sauver :
  - Le produit de la base de données SQL.
  - Les données issues de l'exécution de la chaîne de traitement de données.



# StratusLab : Machine virtuelle

## Contexte astrophysique

Mission Euclid  
Galaxies

## Au laboratoire APC

Code de calcul  
Données en entrée  
Données en sortie  
Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

Avantages de l'infrastructure

## Conclusions

Conclusion  
Perspective

- Pas de contrainte forte sur les ressources de calcul :  
1 cœur de calcul, 4 GB de mémoire, 2 GB de swap.
- Cycle de fonctionnement :
  - 1 Démarrage avec l'image disque spécialisée.
  - 2 Montage du disque persistant.
  - 3 Ouverture d'un port (par ex. 8080) pour l'utilisation de l'interface Web.
  - 4 Utilisation (connexion au compte utilisateur générique avec une clé ssh).
- Accessibilité : dépend du type d'utilisateur.
  - Avancé (Root + User) ➔ 4 étapes du cycle.
  - Lambda (User) ➔ seulement l'étape 4.



# StratusLab : avantages de l'infrastructure I

## Contexte astrophysique

Mission Euclid  
Galaxies

## Au laboratoire APC

Code de calcul  
Données en entrée  
Données en sortie  
Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

## Avantages de l'infrastructure

## Conclusions

Conclusion  
Perspective

## Image disque spécialisée :

- Automatisation de l'installation de logiciels.
- Utilisation collective de l'image disque :
  - Lancement par plusieurs utilisateurs de machines virtuelles avec la même image disque.
  - Utilisation par plusieurs utilisateurs de la même machine virtuelle.
  - Lancement d'une machine virtuelle avec la même image disque sur un autre Cloud.



# StratusLab : avantages de l'infrastructure II

## Contexte astrophysique

Mission Euclid  
Galaxies

## Au laboratoire APC

Code de calcul  
Données en entrée  
Données en sortie  
Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

## Avantages de l'infrastructure

## Conclusions

Conclusion  
Perspective

## Disque persistant :

- Taille à la demande.
- Montable et démontable.
- Pas de perte de données en cas de défaillance de la machine virtuelle.
- Pas de perte de temps en cas de modifications des ressources de calcul de la machine virtuelle.



# StratusLab : avantages de l'infrastructure III

## Contexte astrophysique

Mission Euclid  
Galaxies

## Au laboratoire APC

Code de calcul  
Données en entrée  
Données en sortie  
Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles

## Avantages de l'infrastructure

## Conclusions

Conclusion  
Perspective

## Machine virtuelle :

- Instantiation et connexion SSH immédiate.
- Durée de vie infinie (ou presque).
- Utilisateur **Root** ➔ très pratique pour les utilisateurs avancés.
- Gestion des connexions ➔ pas de restriction sur les ports.
- Portage et exécution du code ➔ instantané.
- Temps d'exécution du calcul ➔ aucune dégradation de performance.
- Toutes les erreurs d'administration système sont permises... ➔ apprentissage du système.



# Conclusion

## Contexte astrophysique

Mission Euclid  
Galaxies

## Au laboratoire APC

Code de calcul  
Données en entrée  
Données en sortie  
Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles  
Avantages de l'infrastructure

## Conclusions

Conclusion  
Perspective

Le Cloud **StratusLab** est adapté à cette application astroparticule.

- Ressources de calcul adéquates.
- Environnement de travail similaire à l'environnement de travail usuel.
- Aucune perte de temps lors du portage du code.

**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 et particulièrement quand l'obtention de résultats doit être rapide.**



# En perspective

## Contexte astrophysique

Mission Euclid  
Galaxies

## Au laboratoire APC

Code de calcul  
Données en entrée  
Données en sortie  
Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles  
Avantages de l'infrastructure

## Conclusions

Conclusion  
Perspective

- Utilisation d'une machine virtuelle multi-coeur :
  - Ré-écriture du code **Le PHARE** en Python et implémentation dans le pipeline **Photo-z**.
  - Parallélisation du code **Le PHARE** avec OpenMP.
- Stockage des grands volumes de données en sortie :
  - Augmentation de  $N_G$   $\rightarrow$  plusieurs **TB** à stocker.
  - Utilisation future :
    - d'un disque persistant de  $\sim 1$  **TB**.
    - d'un système de données distribuées de type IRODS.

# Merci pour votre attention.

## Contexte astrophysique

Mission Euclid  
 Galaxies

## Au laboratoire APC

Code de calcul  
 Données en entrée  
 Données en sortie  
 Pipeline

## Utilisation du Cloud StratusLab

Caractéristiques matérielles et logicielles  
 Avantages de l'infrastructure

## Conclusions

Conclusion  
 Perspective

