



## Introduction

Sondage

## Comparaison

Changement de

paradigme

Performance

## Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur

# Portage d'applications sur le Cloud IaaS

**Cécile Cavet**

`cecile.cavet@apc.univ-paris7.fr`

Centre François Arago (FACe),

Laboratoire AstroParticule et Cosmologie (APC), LabEx UnivEarthS

*APC, Univ. Paris Diderot, CNRS/IN2P3, CEA/Irfu, Obs. de Paris,*

*Sorbonne Paris Cité, France*

3 Juillet 2014



# Sondage

## Introduction

### Sondage

## Comparaison

Changement de paradigme

Performance

## Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur

Quelle infrastructure de calcul utilisez-vous ?

- PC uniquement.
- Cluster (MPI, Hadoop).
- Centre HPC/HTC.
- Grille.
- Cloud (un peu...) / serveur virtualisé.

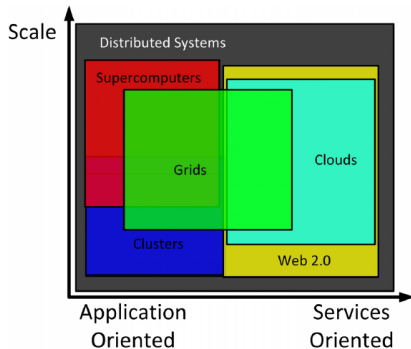


Figure: Infrastructures de calcul distribué.



# Changement de paradigme

## Introduction

Sondage

## Comparaison

Changement de  
paradigme

Performance

## Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur

## Portage d'application :

sur le cloud  $\neq$  sur un cluster/supercalculateur.



VS





# Changement de paradigme

## Introduction

Sondage

## Comparaison

Changement de  
paradigme

Performance

## Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur

## Cluster :

- OS et ressources fixes.
- Simple utilisateur.
- Gestionnaire de job.
- Compilateur très performant (Intel...).
- Bibliothèques déjà installées.
- Sauvegarde automatique des données.

➔ Portage = modification du code pour l'adapter à l'environnement informatique.



# Changement de paradigme

## Introduction

Sondage

## Comparaison

Changement de  
paradigme

Performance

## Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur

## Cloud :

- OS et ressources à la demande.
- Utilisateur **root** et normal (**cloud-user...**).
- Gestionnaire de MV.
- Compilateur gratuit (les licences ne sont pas encore très adaptatives au concept de MV).
- Librairies à installer.
- Gestion par l'utilisateur de la sauvegarde des données.

➔ Portage = modification de l'environnement informatique pour l'adapter au code.



# Et la performance ?

## Introduction

Sondage

## Comparaison

Changement de  
paradigme

Performance

## Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur

Cluster virtuel du cloud **VS** cluster physique :

- Bande-passante de la mémoire : **STREAM**.
- Accès aux I/O : **IOzone**.
- Communications MPI : Intel MPI Benchmark (**IMB**).
- Transfert de fichiers : Integrated Rule-Oriented Data System (**IRODS**).
- High Performance Computing (HPC) :
  - NASA Parallel Benchmark (**NPB**).
  - High Performance LINPACK (**HPL**).

➔ La virtualisation ne dégrade pas les performances, les ressources virtualisées transcrivent les ressources physiques.



# Méthodologie du portage d'application

## Introduction

Sondage

## Comparaison

Changement de

paradigme

Performance

## Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur

- Créer / utiliser une image disque de base : choix de l'OS, de la taille du disque.
- Customisation de cette image :
  - Installation de paquets Linux.
  - Installation de paquets scientifiques : compilateur, librairie, outils de développement, détachement de job, visualisation...
  - Installation de l'application : dépôt ou serveur Web.
- Partage de l'image disque sur un/plusieurs catalogue(s).
- Création d'un disque persistant pour sauver de manière permanente les données, une structure...



# Les outils qui facilitent le travail

## Introduction

Sondage

## Comparaison

Changement de

paradigme

Performance

## Portage

### d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur

- En général : connaissances d'administration système.
- Sur la MV :
  - **root** bien sûre...
  - CloudInit :
    - Contextualisation multi-cloud : StratusLab, OpenStack, OpenNebula...
    - Utilisateur générique sur la MV : **cloud-user**.
  - **ping** : si la MV perd sa connexion réseau il ne reste plus qu'à la détruire (**StratusLab**) ou la re-booter (**OpenStack**) si le problème est temporaire.
  - **kill/delete** : permet de tout tester (même le pire...)...et de tout détruire.





# Méthodologie de l'utilisation d'une MV pour faire du calcul

## Introduction

Sondage

## Comparaison

Changement de paradigme

Performance

## Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur

### ■ Choisir suffisamment de ressources :

- CPU, Mémoire, Swap.
- Espace disque.
- Stockage temporaire et/ou permanent.

➔ en général en plusieurs itérations (sauf si on connaît très bien son code).

### ■ Utiliser le même « workflow » que sur l'infrastructure habituelle :

- Si pipeline : MV avec gestionnaire de pipeline.
- Si cluster : cluster virtuel avec gestionnaire de soumission de job.

➔ Customisation plus contraignante mais gain très intéressant : **assure le pic de productivité car extensibilité à la demande.**



# Les outils qui facilitent le travail

## ■ Sur la MV :

- **nohup, screen** : détachement de job.
- **df, free, fdisk...** : surveillance de l'utilisation des ressources.
- **snapshot** : sauvegarder régulièrement l'environnement de travail (modifications du code...) et les données produites.

### Introduction

Sondage

### Comparaison

Changement de paradigme

Performance

### Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

### Projet

Rôle de l'ingénieur



# Le problème du multi-cloud

## Introduction

Sondage

## Comparaison

Changement de

paradigme

Performance

## Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur

- L'utilisateur doit connaître les fonctionnalités de X clients de cloud.
  - ➔ Outils multi-cloud comme **SlipStream** (voir samedi).
- Chaque image disque doit être référencée sur le catalogue correspondant à l'infrastructure.
  - MarketPlace pour **StratusLab** ou Catalogue d'image géré par Glance pour **OpenStack**.
  - Si instantané d'une image disque (customisation)
    - ➔ ne peut pas être exportée.
- ➔ Un catalogue unique serait idéal (à venir ?).



# Le problème du multi-cloud

## Introduction

Sondage

## Comparaison

Changement de paradigme

Performance

## Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur

- Le disque persistant est lié à une infrastructure de cloud.
  - Le rapatriement de grand volume de données est problématique.
  - Si le cloud est down, pas d'accès aux données.
- ➔ Image ISO pour le transfert d'une infrastructure vers une autre, le rapatriement des données.
- ➔ Système de fichiers distribués : IRODS, AFS en essai au CC-IN2P3 (système de fichiers partagés entre MV/serveur stockage).



# Dans le cadre d'un projet, le rôle de l'ingénieur

## Introduction

Sondage

## Comparaison

Changement de paradigme

Performance

## Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur

- Création et partage de l'image disque : **utilisateurs avancés**.
  - Utilisateurs instancient eux-mêmes la MV.
  - 1 utilisateur / MV .
- Création et partage de la MV : **utilisateurs lambda**.
  - 1 admin + X utilisateurs .
  - Utilisateurs se connectent seulement à la MV.
  - X utilisateurs / MV.
  - Conseil : mettre le `\home` des utilisateurs sur un disque persistant.



# Philosophie : l'économie de moyen

## Introduction

Sondage

## Comparaison

Changement de

paradigme

Performance

## Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur

- Réutiliser les images disques disponibles sur les catalogues.
- Automatiser le plus possible les différentes étapes car la MV a une durée de vie limitée.
- Préférer les **utilisateurs avancés** dans un projet (surtout si on veut partir en vacances...).



# Et maintenant, les TP développement...

## Introduction

Sondage

## Comparaison

Changement de

paradigme

Performance

## Portage d'application

Méthodologie du portage

Méthodologie du calcul

Interopérabilité du cloud

## Projet

Rôle de l'ingénieur



Figure: @Software Carpentry.