

Projet Imageries du Vivant (IDV) Points de vue des infrastructures

Leila Abidi

leila.abidi@lipn.univ-paris13.fr



Plan de la présentation

- Positionnement
- Caractéristiques des tâches
- Vue fonctionnelle
- Vue technique
- Infrastructure du cloud
- Mise en place de la plateforme



Préambule

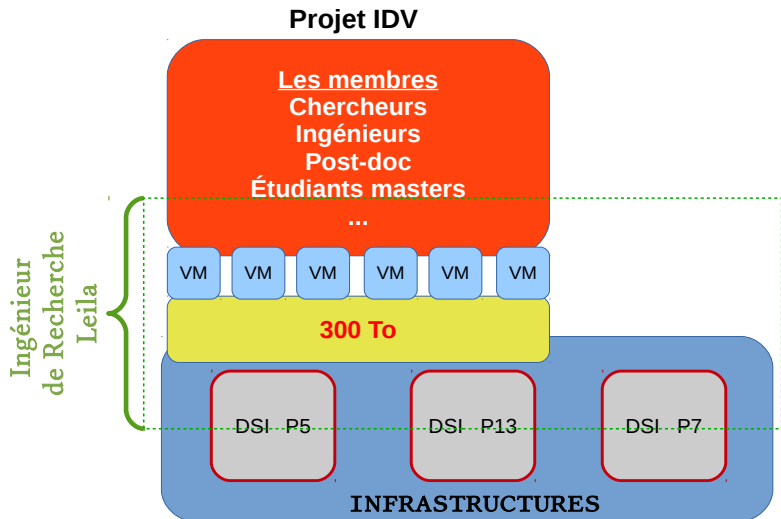
Vision d'un gestionnaire/administrateur d'infrastructures

- une **image** = **suite d'octets** dans un fichier
- les **logiciels** (sans distinction) sont des outils pour **traiter** / **analyser** les images
- une **infrastructure** = processeurs + stockage + réseau

Mon rôle

- Construire un **eco-système** logiciel
- Offrir un **service d'appui** aux recherches dans IDV pour exploiter au mieux les grandes infrastructures
- Permettre aux utilisateurs d'accéder à :
 - ▶ **une grande panoplie d'outils logiciels**
 - ▶ **une grande masse de données images**
- Assurer un **confort** aux utilisateurs :
 - ▶ perturber au minimum leurs façons de travail actuelles
 - ▶ s'adapter à leurs usages

Préambule (2) : Frontières entre responsabilités



Caractéristiques des tâches (première approximation)

- 1 Les applications avec des **dépendances** \Rightarrow les **clusters** (supercalculateurs)
- 2 Les applications avec **peu de dépendances** (BOT : Bag of Tasks) \Rightarrow les **grilles** (grids)
- 3 Les applications **sans dépendances** \Rightarrow les **Clouds**



Caractéristiques des tâches (raffinement de la notion)

(1) HPC : High Performance Computing

- Requiert une puissance de calcul (supercomputing)
- Courtes périodes de temps
- L'utilisateur est intéressé par savoir en combien de temps sa tâche va terminer.

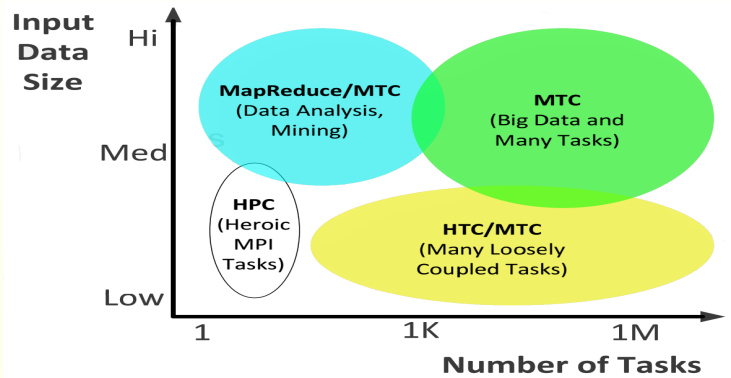
(2) HTC : High Throughput Computing

- Requiert une puissance de calcul (plusieurs ressources)
- Durées beaucoup plus longues (des mois plutôt que des heures)
- L'utilisateur est plus intéressé à savoir combien de tâches peuvent être complétées sur une longue période de temps.

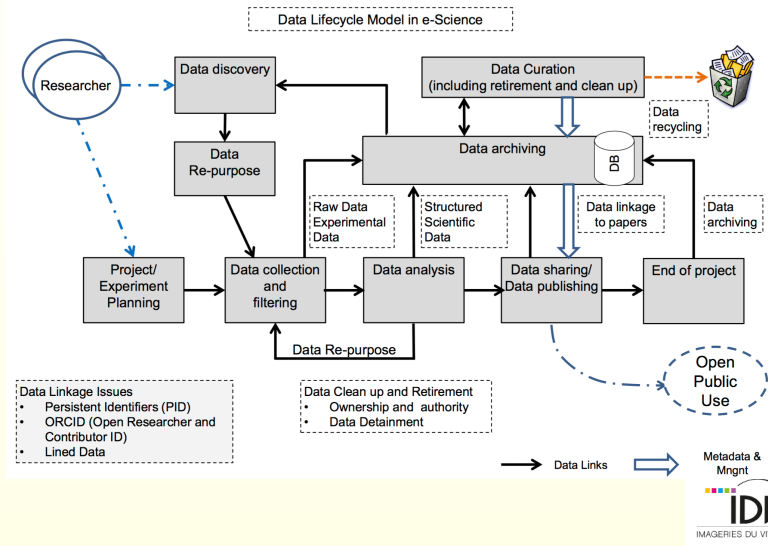
(3) MTC : Many Task Computing

- HPC comprenant plusieurs tâches distinctes
- Calculs couplés via des opérations d'écriture et de lecture sur un système de fichiers

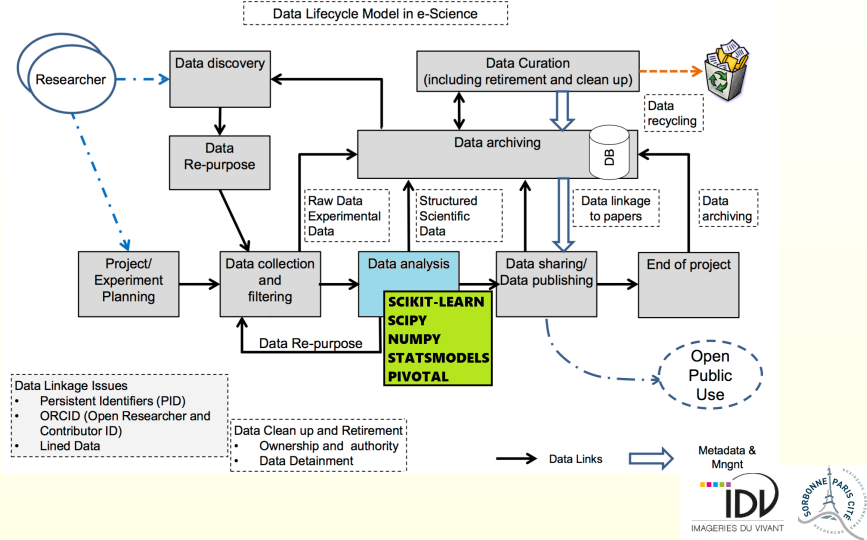
HPC ← MTC → HTC



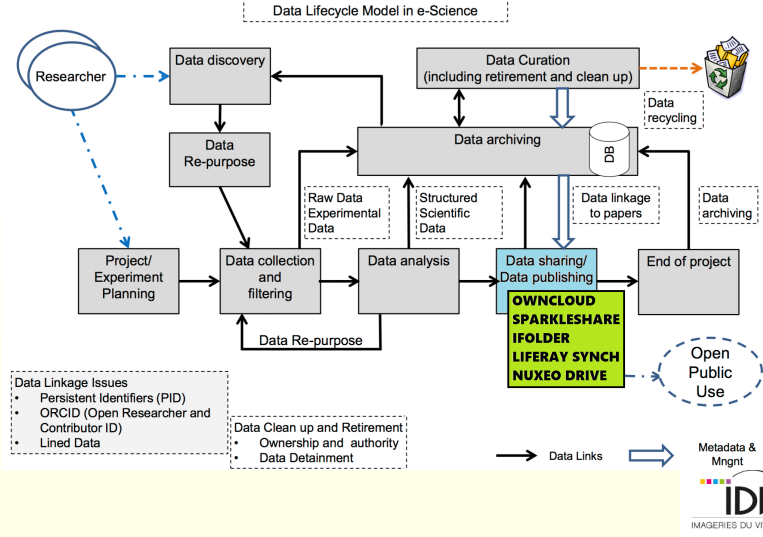
Data Life Cycle : vue fonctionnelle (NIST)



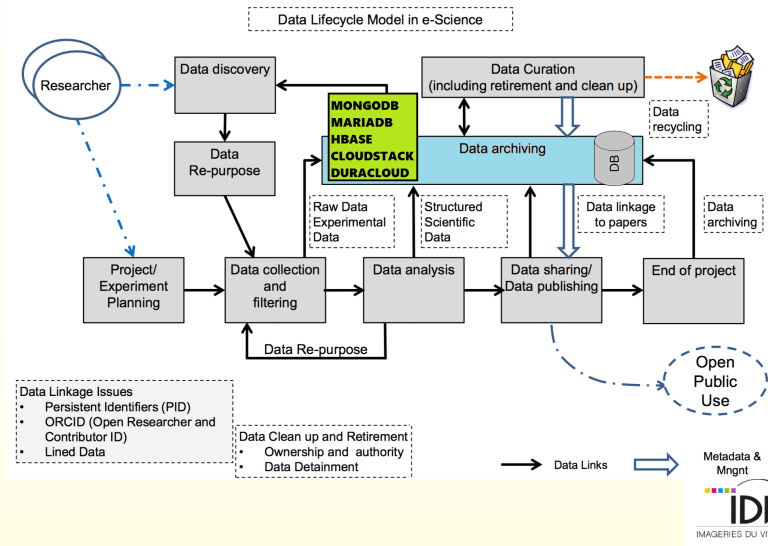
Data Life Cycle : vue technique



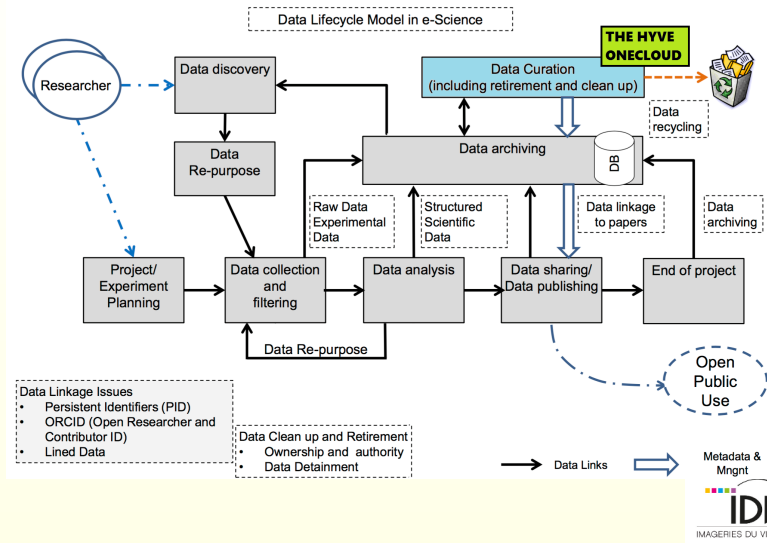
Data Life Cycle : vue technique



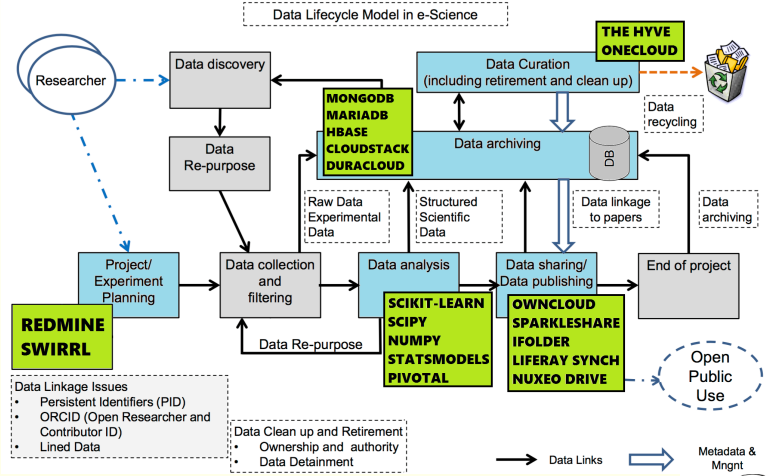
Data Life Cycle : vue technique



Data Life Cycle : vue technique



Data Life Cycle : vue technique



Le cloud : infrastructure orientée Services

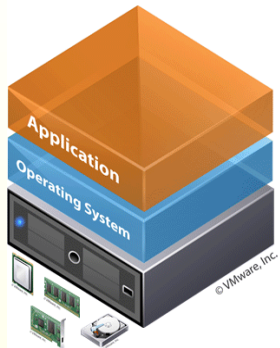
- Tout est **service** (processeurs, stockage et de réseau)
- On parle de "Infrastructure as a Service" (**IaaS**)

Acquis et avantages

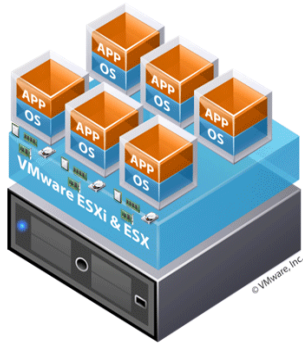
- Accéder à des **niveaux de services** beaucoup plus élevés :
 - ▶ Fiabilité et performance, non atteints avec une installation standard sur site
 - ▶ Se concentrer sur les aspects les plus importants de la gestion des activités
 - ▶ Pas de perte de temps due au maintien et bon fonctionnement de l'infrastructure. **AUTOMATISATION**
 - ▶ Le coût d'entretien du service Cloud comparé à l'exécution et l'entretien d'une infrastructure interne à un laboratoire est moindre. **MUTUALISATION**
- l'installation et l'exploitation (infrastructures et logiciels techniques) sont à la portée de tous

Virtualisation

Sans virtualisation



Avec virtualisation



- **Cohabitation** de plusieurs systèmes sur le même support physique
- Utilisation **optimale** des ressources systèmes
- **Isolation**

Mise en place de la plateforme Cumulus et du projet IDV dans cette plateforme

- Installation d'**OpenNebula** → **DSI**
- Configuration d'un profil Administrateur de la plateforme **Cumulus** → **DSI**
- Accès à un profil Administrateur du projet IDV pour configurer les **Machines Virtuelles** (VM) → **moi**
- **Déploiement** des VM → **DSI/moi**
- Configuration de l'URL afin de rendre la plateforme joignable de l'extérieur → **DSI/moi**
- Mise en place d'une interface d'**authentification** permettant aux utilisateurs à se connecter → **DSI/moi**
- Chaque utilisateur est potentiellement **super-utilisateur** dans sa VM

Infrastructure

Configuration

6 serveurs lames Dell M630 :

- 700 Go RAM
- 2 x Xeon 2.6 GhZ



IMAGERIES DU VIVANT

Copies écran de l'interface

☰ OpenNebula

- Dashboard
- System
- Virtual Resources
 - Virtual Machines**
 - Templates
 - Images
 - Files & Kernels
- Infrastructure
- Marketplace
- OneFlow
- Settings

Support

Not connected

Sign in

oneadmin OpenNebula

Virtual Machines

Refresh + Search [] [Play] [Pause] [Stop] [Refresh] [Grid] [Print] [User] [Delete]

<input type="checkbox"/>	ID ▾	Owner	Group	Name	Status	Host	IPs	
<input type="checkbox"/>	3	oneadmin	oneadmin	IDV-1	RUNNING	antimoine.dig.univ-paris5.fr	172.17.34.60	🗨

Showing 1 to 1 of 1 entries

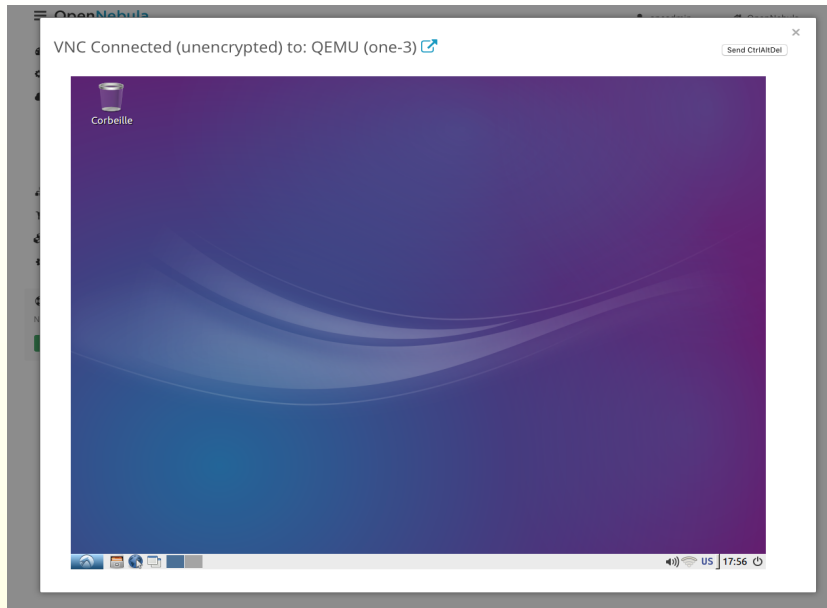
Previous **1** Next 10 ▾

1 TOTAL 1 ACTIVE 0 OFF 0 PENDING 0 FAILED

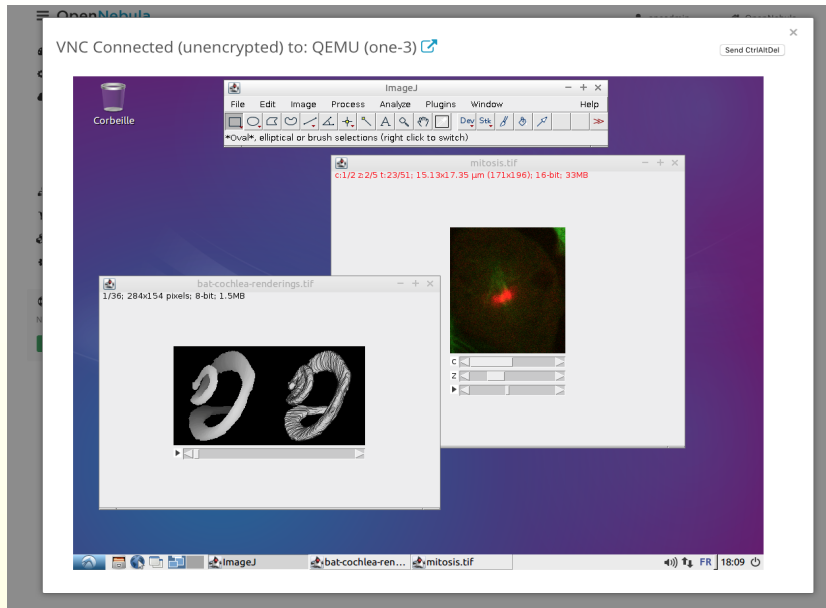
OpenNebula 4.14.0 by OpenNebula Systems.



Copies écran de l'interface



Copies écran de l'interface



VM pour Spark

Un algorithme de segmentation d'image qui passe à l'échelle

Mean-Shift-LSH



Image originale
481x321



Image à 43 clusters
200 blocs
20 min (4 cœurs)
4 min (20 cœurs)



Image à 40 clusters
500 blocs
13 min (4 cœurs)

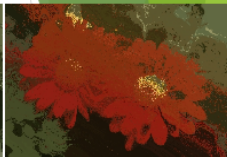


Image à 34 clusters
1000 blocs
10 min (4 cœurs)

**La même qualité d'image est
obtenue en 8h de calcul sous R
soit 24 fois plus.**

Beck Gaël, Hanene Azzag, Tarn Duong, Mustapha Lebbah (LIPN)

1

Perspectives

à court terme

- **Transfert des données** depuis / vers un site extérieur à P5 / Cumulus
 - SSH vers les VMs (IP publique) ?
 - OwnCloud, Globus online ?
- **Stockage** dans le cloud :
 - connecter la VM a un espace de stockage 100, 200, 300 To ?
 - quel protocole pour le montage des disques ?
- Traiter le cas des **logiciels propriétaires** les plus utilisés tel que **Matlab**
- **Traitement parallèle** d'images
 - réseau de VMs avec des outils de traitement massifs de données
 - enjeu : fouiller plus, fouiller mieux → découvrir plus

à long terme

- Implémenter la vision **Data Life Cycle** précédemment introduite

Projet Imageries du Vivant (IDV) Points de vue des infrastructures

Leila Abidi

leila.abidi@lipn.univ-paris13.fr

