L'Institut Français de Bioinformatique (IFB) : mettre en place une infrastructure informatique pour les sciences de la vie

J-F Gibrat
UMR 3601 – IFB-core,
CNRS Gif-sur-Yvette

Journée moyens mutualisés
d'accès au calcul intensif,
Paris, 11 janvier 2016

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

L'Institut Français de Bioinformatique

- Mission générale : fournir des ressources de base en bioinformatique à la communauté des sciences de la vie
- Infrastructure nationale de **service** en bioinformatique
  - **Données** : Fournir un accès à des collections de données spécialisées à haute valeur ajoutée issues de l'expertise du laboratoire d'accueil
  - **Outils** : Développer et mettre à disposition des outils et services en lignes pour analyser les données correspondant à l'expertise scientifique du laboratoire d'accueil
  - **Appui** aux projets scientifiques et hébergement sur une infrastructure informatique
  - **Infrastructure** : Mettre à disposition une infrastructure informatique dédiée à l'analyse des données des sciences du vivant (matériel, données, outils)
  - **Formations**

cf. http://france-bioinformatique.fr

ifb
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

Caractéristiques des analyses bioinformatiques…1

- *Les utilisateurs manquent d'une « culture » de base en informatique.*
- Beaucoup d'analyses sont distribuables (parallélisables par les données).
- Les analyses nécessitent souvent l'enchaînement de logiciels différents (pipelines, workflows)
- et l'utilisation de collections de données publiques qui sont mises à jour régulièrement.
- Grande variété de langages utilisés (perl, python…)
- Les logiciels utilisés ont souvent beaucoup de dépendances (version de bibliothèques)

Caractéristiques des analyses bioinformatiques…2

- Nécessité de suivre l'évolution très rapide des technologies de production des données.
- Foisonnement des logiciels d'analyse (98 logiciels pour aligner les lectures sur un génome).
- Typiquement, une PF bioinformatique met à disposition plusieurs centaines de logiciels différents.
- Utilisation de bases de données relationnelles (SGBD: postgreSQL, mySQL) et noSQL.
- Nécessité de mettre à disposition des interfaces conviviales pour les utilisateurs (portails web, logiciels avec interfaces graphiques).
- Résultat➔ très faible utilisation des centres de calcul intensif

Fédération de Clouds de l'IFB

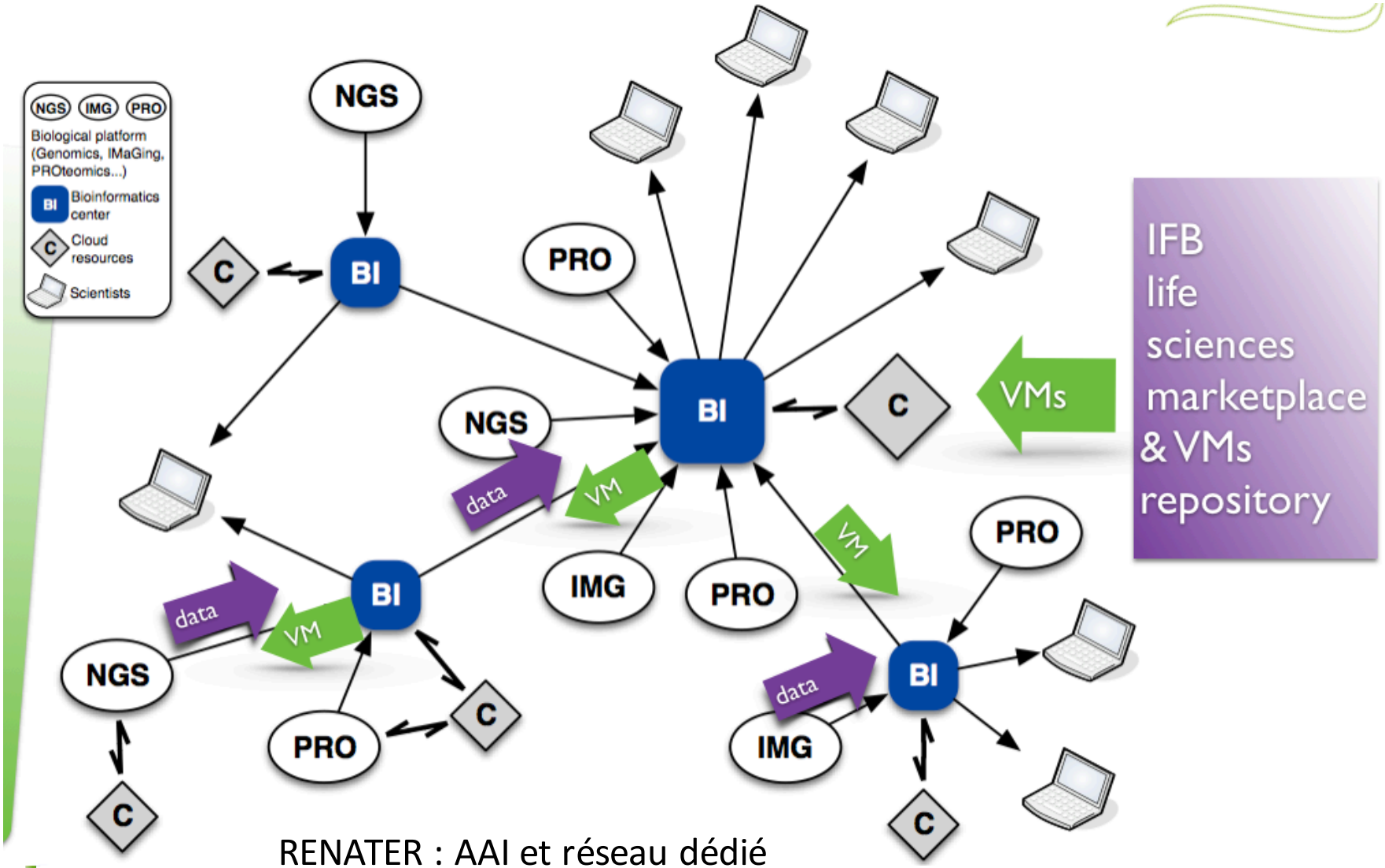
Propriétés du Cloud

- Accès standard par le réseau
- Accès en self-service (à la demande)
- « Élasticité » : les ressources informatiques (stockage, calcul, mémoire, bande passante réseau) sont évolutives
- Modèle économique basé sur une mesure fine de l'utilisation des ressources

Déploiement d'une fédération de Clouds académiques

- Infrastructures régionales :
  - 15 000 cœurs et 6 Po de stockage; > 5 000 utilisateurs
- Infrastructure nationale hébergée à l'IDRIS :
  - Pilote 200 cœurs, 50 To de stockage
  - Début 2017 : 5 000 cœurs, 1Po de stockage
  - Début 2018 : 10 000 cœurs, 2 Po de stockage
  - Mutualisation du stockage sur bande avec IDRIS (2 Po)

Fédération de Clouds pour la bioinformatique



RENATER : AAI et réseau dédié

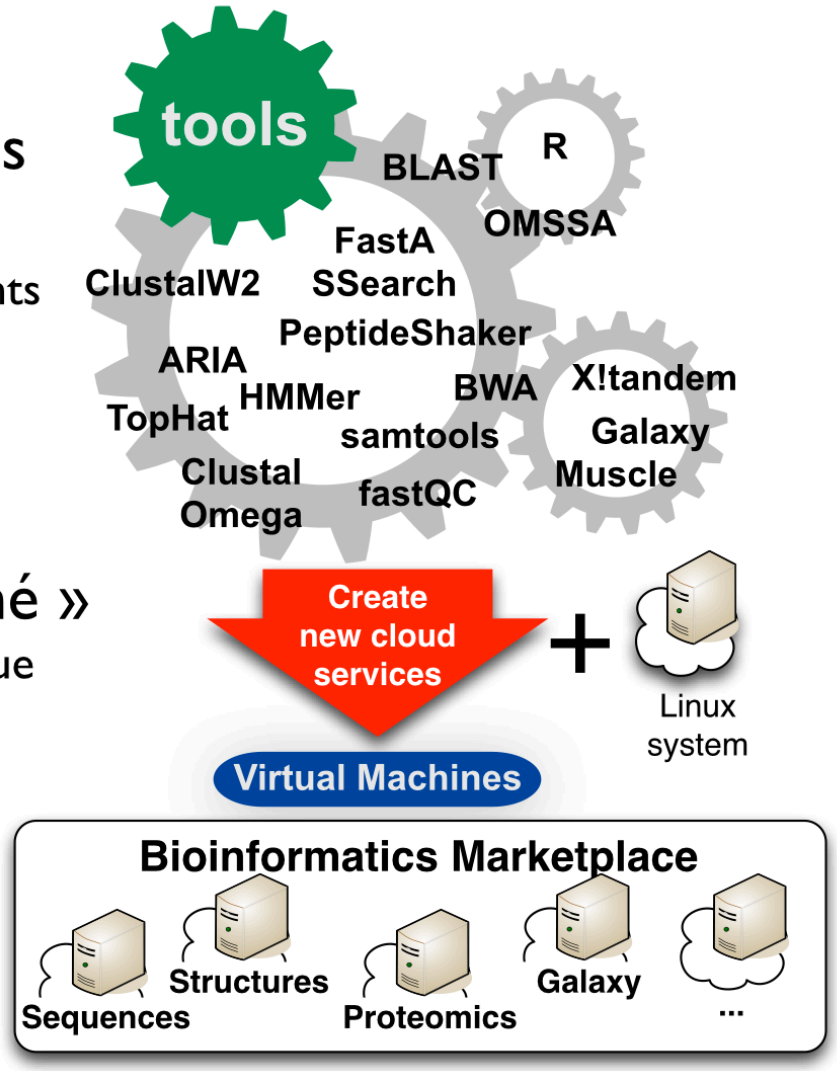
Accès aux différents Clouds via un « broker » (slipstream)

Les « appliances » bioinformatiques

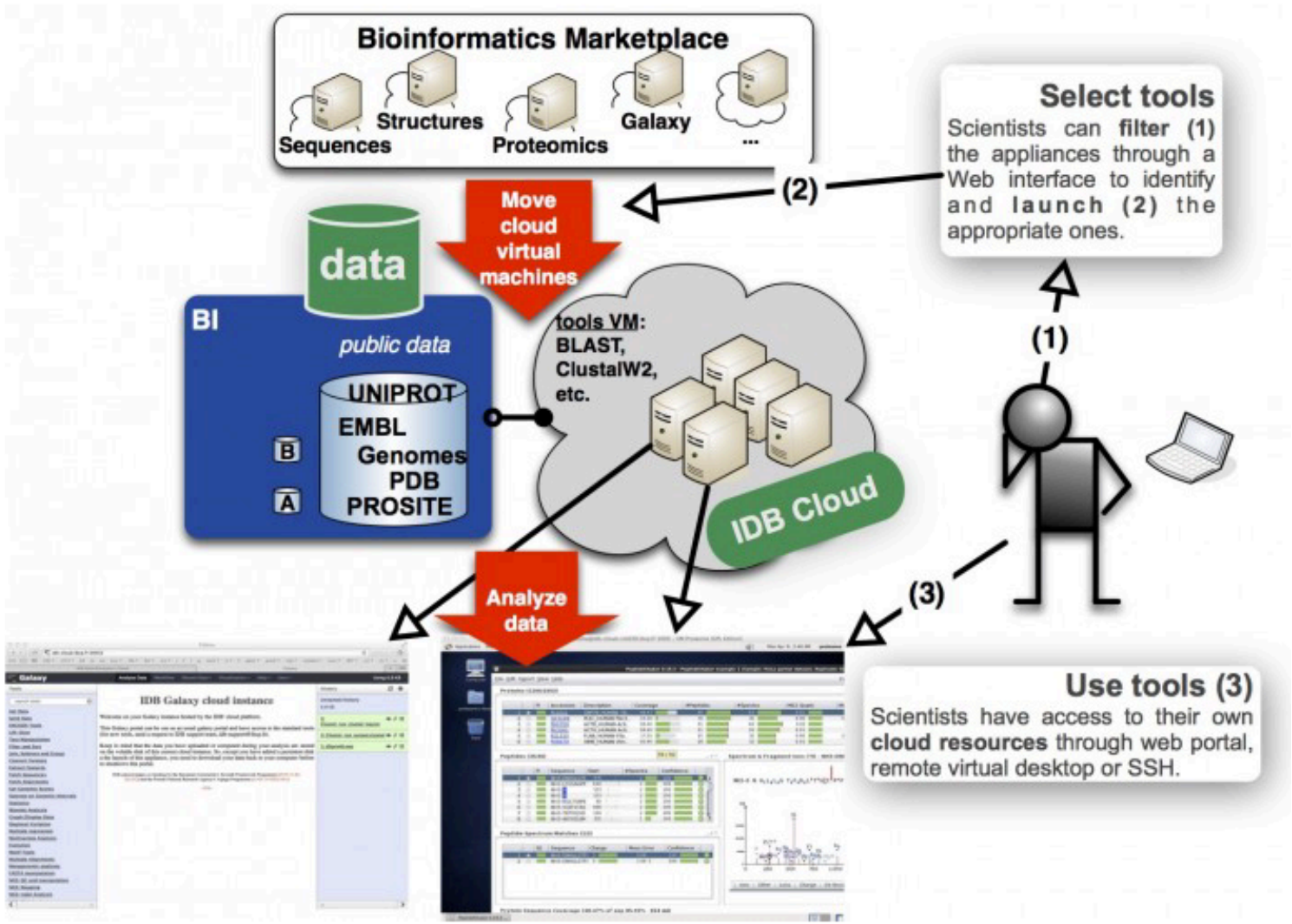
- Un Cloud fournit des machines virtuelles configurables, par ex. le Cloud commercial d'Amazon
- Valeur ajoutée de l'IFB : fournir des « solutions » bioinformatiques clés en main.
- Création d'« appliances » dédiées :
  - à un domaine particulier (protéomique, métabolomique, etc.)
  - à un type d'analyse particulier (analyse de variants, expression différentielle de gènes, ChIP-seq, etc.)
- Appliance (image) / MV <–> Classe / objet
- Création d'un catalogue d'appliances national (market place)

Création des « IFB apps »

- Apps bioinformatiques = machines virtuelles usuelles
  - quelques Go, faciles à convertir dans différents formats de virtualisation
- Pré-configurées avec outils bioinformatiques
- Référencées sur « place du marché »
  - Catalogue de MV dédiées à la bioinformatique
- MV « docker »
  - Installation de programmes dans containers



Utilisation du Cloud

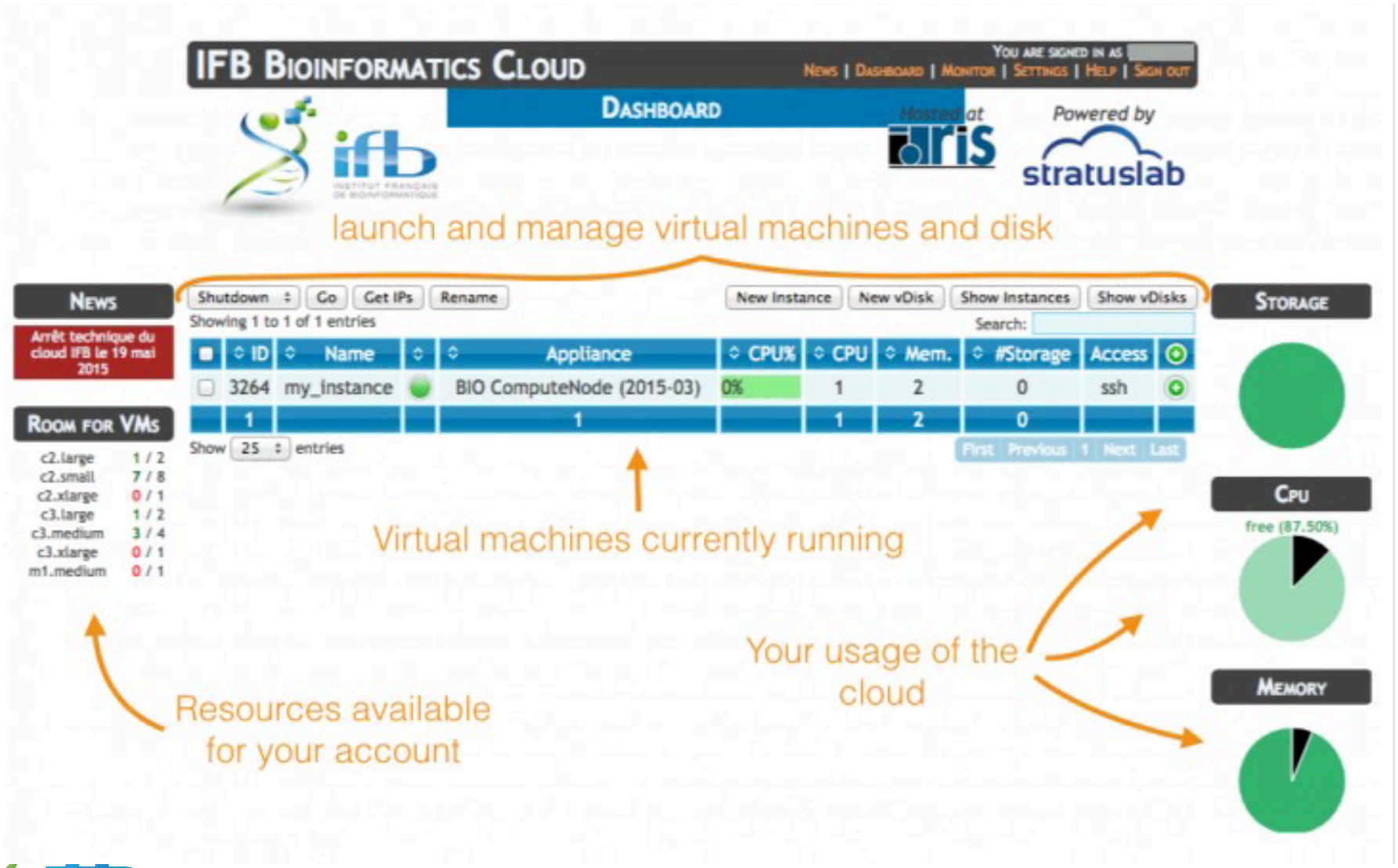


INSTITUT FRANÇAIS DE BIOINFORMATIQUE

Le Cloud pour les utilisateurs

- Faciliter l'utilisation des appliances pour les utilisateurs en utilisant des environnements
  - Portails Web (serveur Galaxy)
  - Bureau virtuel à distance
- Faciliter l'utilisation du Cloud pour les utilisateurs
  - « Tableau de bord » permettant de réserver les MV et de créer des disques virtuels persistants
  - Création automatique d'un cluster de VM
  - Export de données grâce à un serveur NFS
- Avantages :
  - Utilisation à la demande de ressources informatiques
  - Bibliothèques d'appliances répondant à divers types d'analyse
  - Assure la reproductibilité des analyses
  - Permet de faire facilement des formations

Tableau de bord du Cloud



INSTITUT FRANÇAIS DE BIOINFORMATIQUE

Le Cloud pour les utilisateurs

- Faciliter l'utilisation des appliances pour les utilisateurs en utilisant des environnements conviviaux :
  - Portails Web (serveur Galaxy)
  - Bureau virtuel à distance
- Faciliter l'utilisation du Cloud pour les utilisateurs
  - « Tableau de bord » permettant de réserver les MV et de créer des disques virtuels persistants
  - Création automatique d'un cluster de MV
  - Export de données grâce à un serveur NFS
- Avantages :
  - Utilisation à la demande de ressources informatiques
  - Catalogue d'appliances répondant à divers types d'analyse
  - Assure la reproductibilité des analyses
  - Permet de faire facilement des formations

Le Cloud pour les développeurs

- Faciliter la création d'appliances pour les développeurs :
  - Définitions de bonnes pratiques, formation des développeurs
  - Utilisation de containeurs Docker et d'un dépôt (BioShaDock)
  - Automatisation de la création d'appliances avec des logiciels de gestion de configurations (puppet, quattor, Ansible,…)
- Avantages
  - Installation *unique* d'un pipeline ou workflow dans une appliance.
  - Mécanisme « d'héritage » : création d'appliances à partir d'appliances déjà existantes.
  - Découplage des aspects administration système des aspects développements (problématiques DevOps)