

Biosphère-12

Echanges Techniques

SOC



Ordre du jour - 05-06 février 2020

Echanges techniques

Gestion des données

- Interopérabilité, avancement IFB et Elixir (AG)
- Plan de gestion des données - DMP (OC)
- Ceph-recherche au CCIPL (YD)
- Projet DOMA-FR (JFG)

Intégration des outils et pipelines (appliances et conteneurs)

- Appliances basées sur des dépôts git (CB)
- Introduction à Terraform (SD)
- Goterra (OS)

- Création d'image de VM avec Packer (JL)
- Biocontainer (OS)

Fédération

- Migration slipstream et keycloak (CB, JL)

Gestion des utilisateurs

- quotas
- SSO et keycloak -> Biosphère-AAI

Administration de site

- configuration d'un nouveau site
- proxy de site
- suivi de l'usage effectif des VMs
- gabarits spéciaux
- GPU, ce qui marche et ne marche pas
- VMs "expirées?"
- affichage des slots disponibles en fonction des gabarits

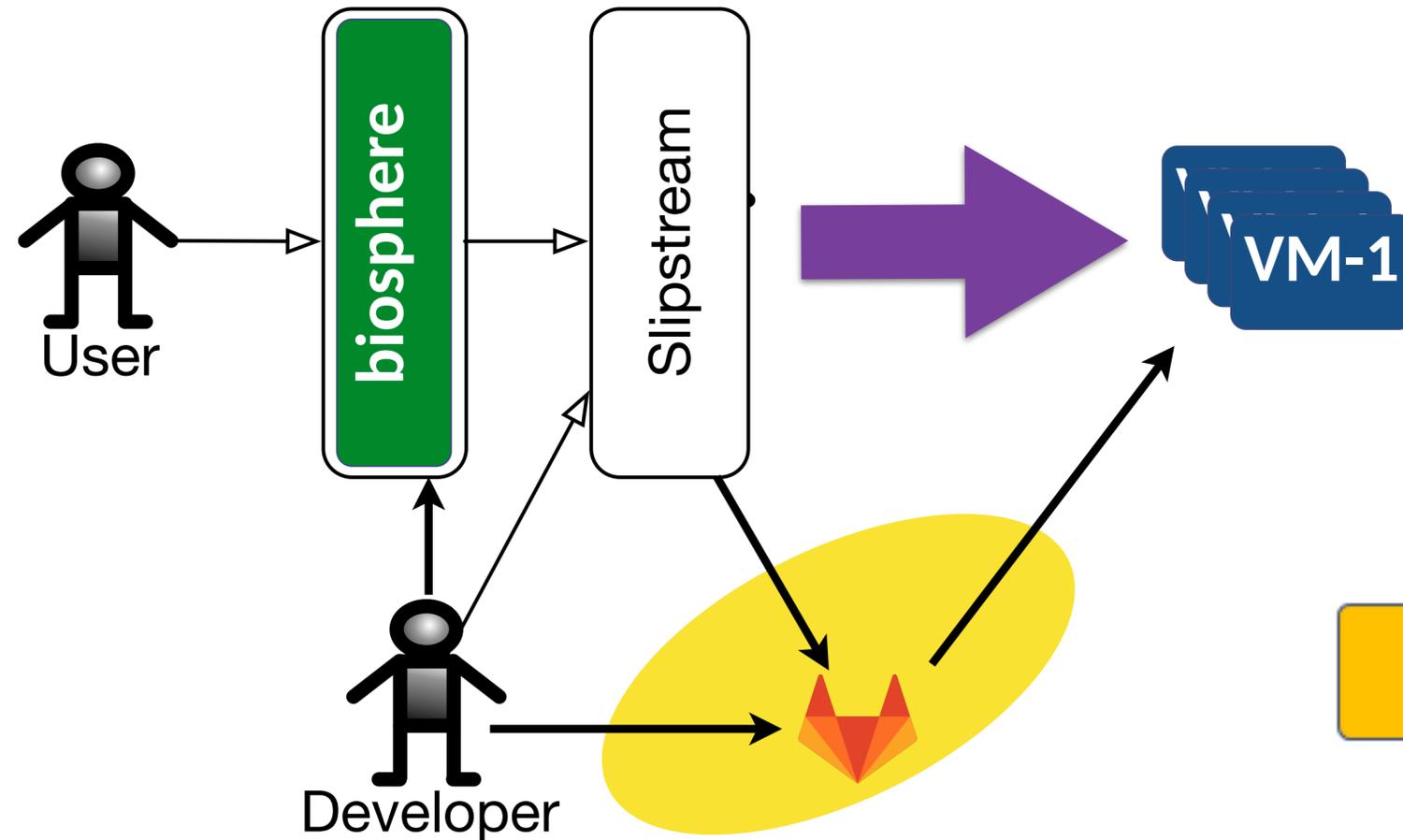
Stockage et gestion des données

Stockages disponibles pour les utilisateurs

- 4 types
- propres à un cloud, sans échange inter-sites
- nfs/autofs,manila
- visibles depuis la page DATA de Biosphère (BioMaJ)

Stockage	Droits	VM	Répertoire	Comm.
Données référence	ro	Toutes	/ifb/data/public	BioMaJ, appdata, <i>teachdata</i>
Ephémère	rw	Interne à la VM	/ifb/data/local	Variable par cloud
Partage usager	rw	Toutes VMs de l'utilisateur	/ifb/data/mydisk	Manila
Partage groupe	rw	Toutes VMs des membres du groupe	/ifb/data/ <nom_groupe>	Manila

Développement d'apps



Recettes de déploiement

- Dépôts git, integration continue
- Conteneurs docker, référencés dans des catalogues
- Workflows, avec docker
- Jupyter Notebooks
- Scripts R...

- Biosphère-RAINBio
 - ★ description de l'app
- Slipstream
 - ★ provision des VMs
 - ★ installation et configuration des outils et de l'environnement
- Dépôt git (option)
 - ★ description install et config

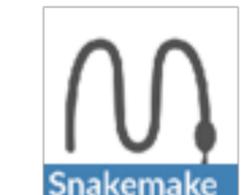
Set of tools

Pipeline, workflow, recipe

Docker (opt.)

Operating System

CONDA



nextflow



CentOS, Debian, Ubuntu

Dev Apps avec git

Architecture

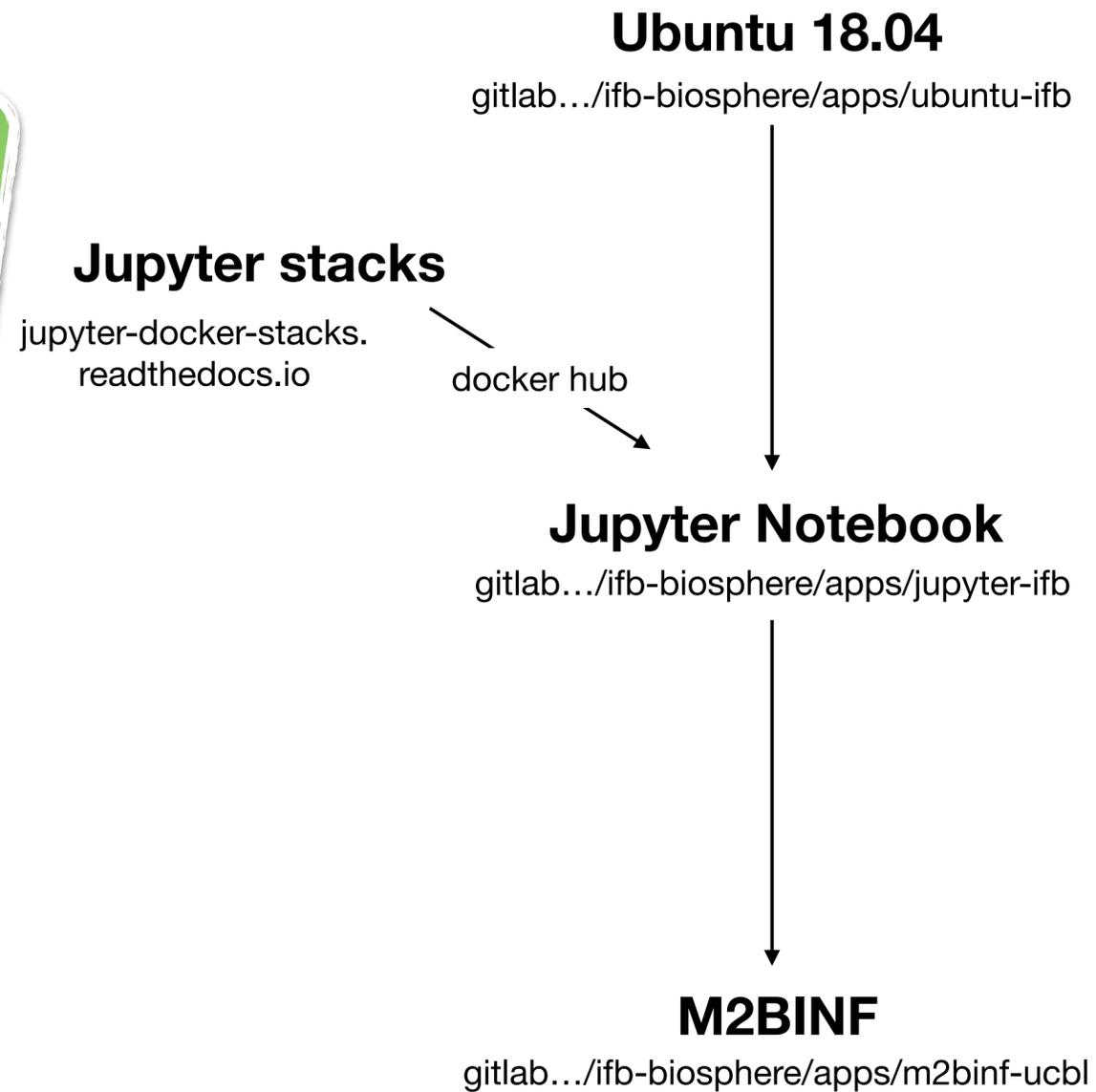
- Description
 - ★ `README.md`
- Installation
 - ★ `install.sh`
- Déploiement
 - ★ `deploy.sh`
- Intégration continue
 - ★ `Dockerfile`
 - ★ `.git-ci.yaml`
- Utilitaires : configuration et templates
 - ★ `ansible playbook`
 - ★ `miniconda-install.sh`
 - ★ `profile-ifb.sh`
 - ★ `nginx...`

Dépôts actuels

- **Biosphere Apps (7)**
 - ★ <https://gitlab.in2p3.fr/ifb-biosphere/apps>
- Genopole 2019
 - ★ <https://github.com/IFB-ElixirFr/biosphere-genopole-genome-variation>
- Microcloud
 - ★ <https://github.com/IFB-ElixirFr/biosphere-microcloud>
- Integron Finder
 - ★ <https://github.com/C3BI-pasteur-fr/IFB-playbook>

```
1) git clone ...  
   /ifb/apprepo/app_name  
2) install.sh  
3) deploy.sh
```

App M2BINF à base de Jupyter



Name	Last commit	Last update
<code>.gitlab-ci.yml</code>	Add CI	4 months ago
<code>Dockerfile</code>	Add ipympl (pip)	4 months ago
<code>README.md</code>	Add App data (version, OS)	4 months ago
<code>deploy.sh</code>	Fix docker binds for shared IFB volumes	4 months ago
<code>install.sh</code>	Version initiale avec build local	4 months ago

App M2BINF UCBL

Description

Environnement cloud pour le support des cours du Master 2 de bioinformatique de Lyon1 (<https://www.bioinfo-lyon.fr>).

Développée par l'équipe pédagogique du Master 2 de bioinformatique de Lyon1 et l'équipe cloud de l'IFB-core

Tools

- Jupyter notebook [ref](#)
- ETE (Environment for Tree Exploration) [ref](#)

Gestion des usagers et des ressources

Utilisateur

- **"Se connecter"**
 - ★ Identifiant des organismes (fédération eduGAIN + service Nuvla-IFB)
 - ★ Compte « local » : username/password, interne (django)
 - ★ Compte de service dans nuvla et les clouds (keystone)
- Compte actif si membre d'un groupe valide
 - ★ Permanent (affiliation) ou à durée limitée (formation, projet)

Groupe

- Création et adhésion en ligne
- Actif après validation par les admins
- Permanent (affiliation) ou à durée limitée (formation, projet)

Quota

- Quota utilisateur par défaut : 36 000 vCPU.h
- Allocation supplémentaire
 - ★ CPU/utilisateur : projet moyen, développeur, formateur
 - ★ CPU/groupe : projet lourd
 - ★ BigMem : suivant les sites



Utilisez vos identifiants de votre organisme (CNRS, INRA, Inserm, Universités...)

[Se connecter](#)

Nous utilisons la fédération d'identité européenne eduGAIN. Si votre organisme académique n'est pas dans cette fédération, vous pouvez utiliser un compte local avec votre adresse professionnelle.

Quotas

	Catégorie	Validité	Nb vCPU.h	Max vCPU	Max IP pub.
Utilisateur	Standard	Affiliation	36010 / an	<i>15% quota / j</i>	5
	Formateur	1 an	100000 (*)	<i>10% quota / j</i>	-
	DevApps	1 an	300000 (*)	<i>5% quota / j</i>	10
	Ext. moy.	1 an	400000 (*)	<i>5% quota / j</i>	10
Groupe	Affiliation	Existence	<i>variable</i>	<i>2x vCPU-eq-an</i>	-
	Formation	jours-mois	<i>variable</i>	<i>1,2x nb vCPUs demandés</i>	<i>2 x part.</i>
	Projet	mois-ans	<i>variable</i>	<i>2x vCPU-eq-an</i>	-
	PF IFB cloud	<i>Contribution</i>	<i>Nb vCPU-PF.an</i>	<i>2x vCPU-PF</i>	<i>ext. : idem user local : illim. pour ses mbs</i>

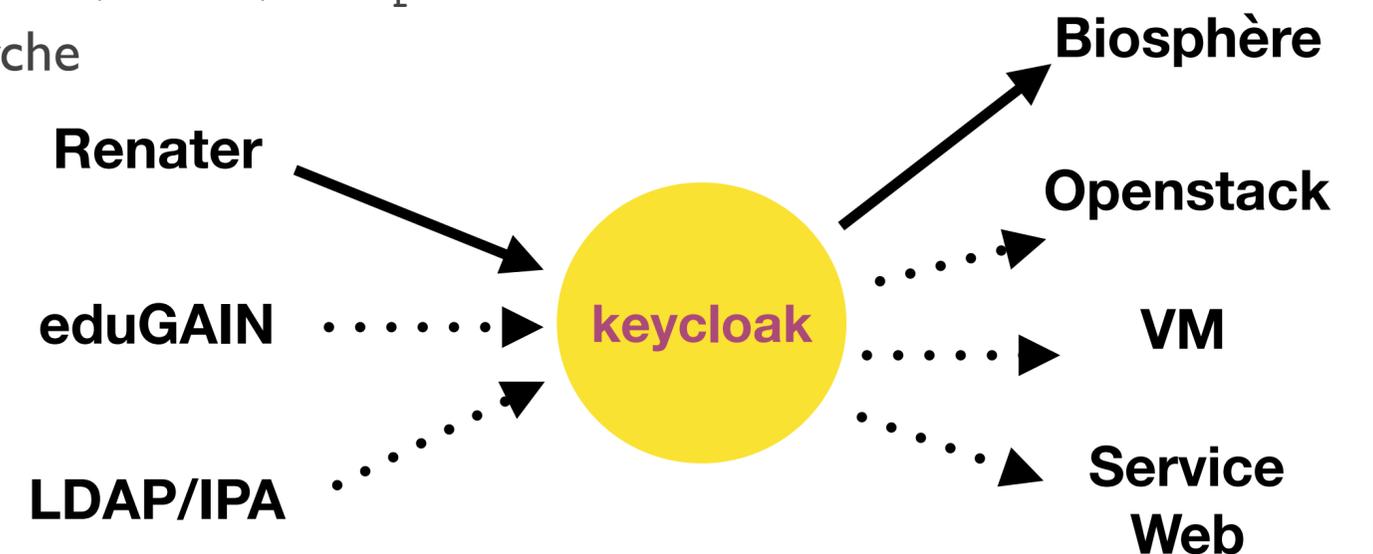
Migration slipstream et keycloak

Souscription NUVLA-IFB (6/2/2020)

- slipstream
 - ★ <https://nuv.la>
- keycloak
 - ★ <https://fed-id.nuv.la/auth/admin/Biosphere>
 - ★ SAMLbridge : eduGAIN

Services IFB-Biosphère

- slipstream
 - ★ <https://slipstream.france-bioinformatique.fr>
- keycloak
 - ★ <https://keycloak.france-bioinformatique.fr/auth/admin/Biosphere>
 - ★ SAMLbridge : Fédération Renater Education-Recherche



Configuration d'un site

1. Créer un tenant 'biosphere'

- utilisateur 'biosphere.admin' avec les droits de créer des comptes pour ce tenant

2. Ouvrir les ports réseaux

- instances : 22 et 443, depuis tout internet
- contrôleur : API publiques, depuis le cloud broker (nuvla/slipstream)
 - ★ keystone/identity (5000)
 - ★ nova/compute (8774)

3. Demander l'enregistrement d'un nouveau cloud dans la fédération.

- <https://biosphere.france-bioinformatique.fr/cloud/add>

4. Référencer les images de base dans le cloud broker

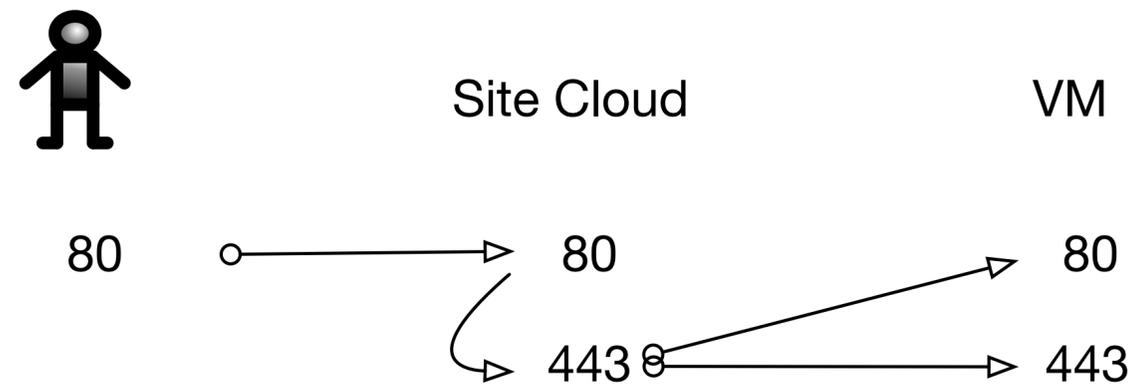
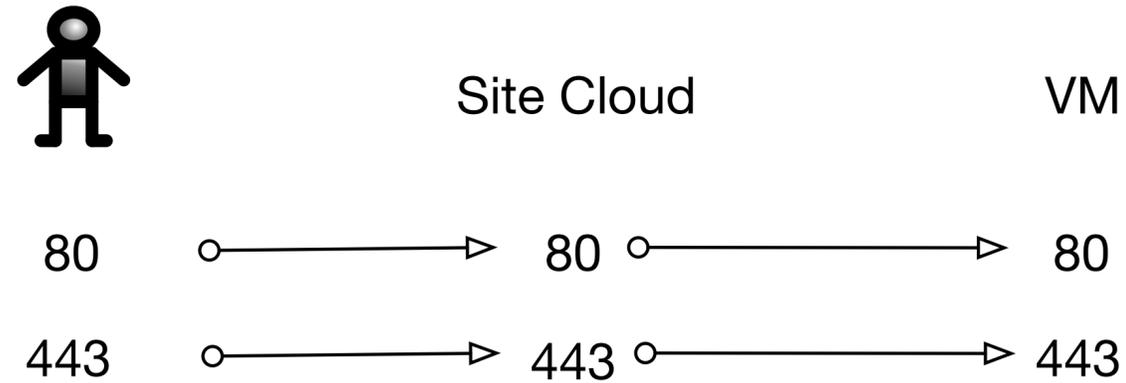
- ubuntu-16.04, ubuntu-18.04, debian-9, centos-7

Options

- volumes permanents : Manila+NFS/cephFS
- données de références: biomaj, apps, formations
- proxy de site : web/ssh

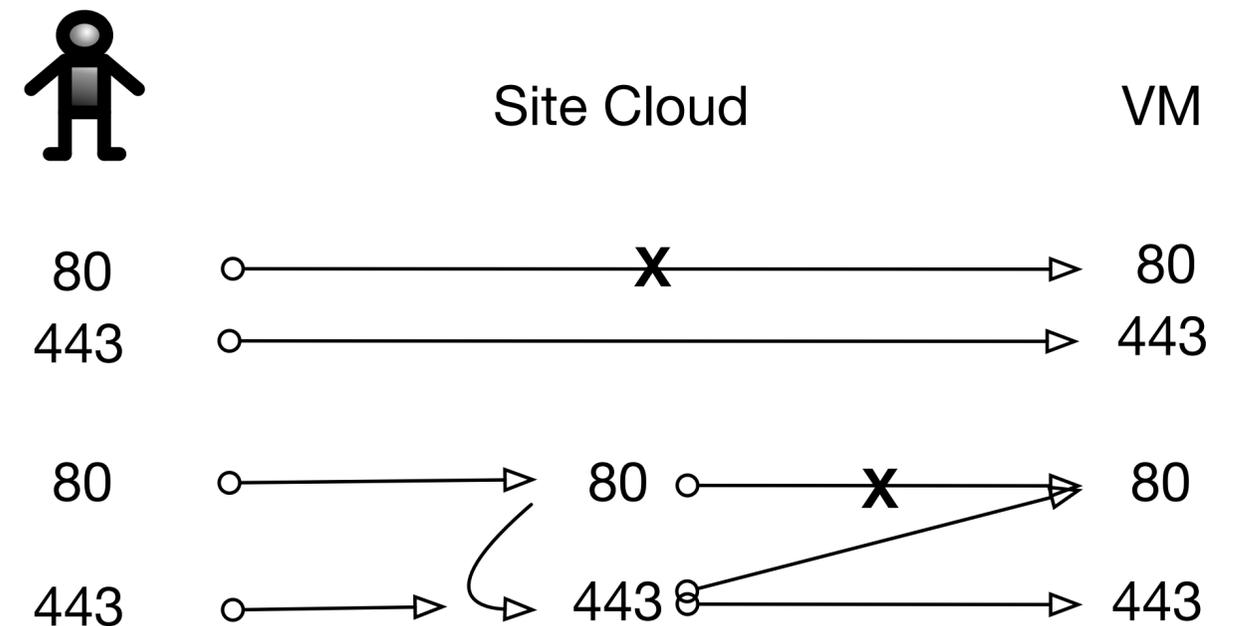
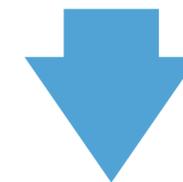
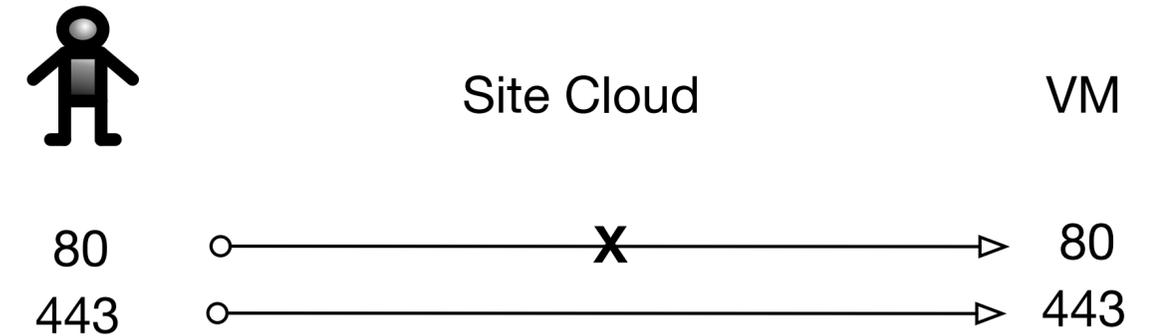
Proxy de site

ifb-genouest-genostack

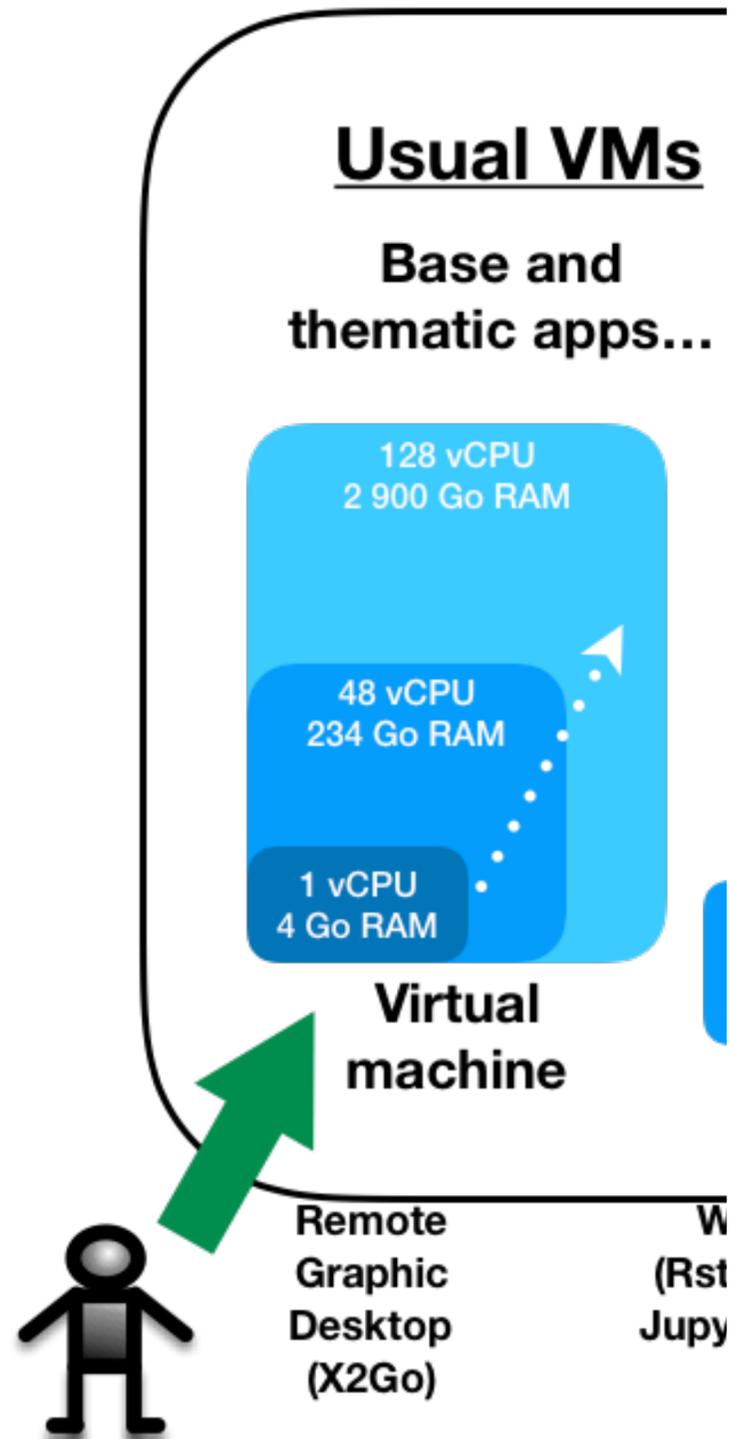


ifb-bird-stack

SANS PROXY



Gabarits spéciaux



Gabarits des VM

ifb.m4.small	1 vCPU, 4Go RAM, 70Go local disk
ifb.m4.large	2 vCPU, 8Go RAM, 120Go local disk
ifb.x1e.large (BigMem)	2 vCPU, 48Go RAM, 20Go local disk
ifb.m4.xlarge	4 vCPU, 16Go RAM, 220Go local disk
ifb.x1e.xlarge (BigMem)	4 vCPU, 96Go RAM, 20Go local disk
ifb.m4.2xlarge	8 vCPU, 32Go RAM, 470Go local disk
ifb.x1e.2xlarge (BigMem)	8 vCPU, 192Go RAM, 20Go local disk
ifb.m4.4xlarge	16 vCPU, 64Go RAM, 920Go local disk
ifb.x1e.4xlarge (BigMem)	16 vCPU, 384Go RAM, 20Go local disk
ifb.m4.6xlarge	24 vCPU, 128Go RAM, 1.3To local disk

- **BigMem (Go/c)** => +200 Go ?
- GPU, SSD
- *manycore (>50c)* ?

Administration de site

- *configuration d'un nouveau site*
- *proxy de site*
- *gabarits spéciaux*
- suivi de l'usage effectif des VMs
- GPU, ce qui marche et ne marche pas
- VMs "expirées?"
- affichage des slots disponibles en fonction des gabarits

Perspectives

Usage natif (API)

- opentack client, web horizon, API
- Programmatisation : terraform, ansible
- Comment concilier avec l'existant Biosphère ?

Stockage

- DMP
- Mode fichier (Manila)
- SSD

Traitement

- accélération 3D GPU
- GP-GPU

Services permanents