Cloud OpenStack de l'IPHC : Retour d'expérience

Jérôme Pansanel
jerome.pansanel@iphc.cnrs.fr

Grenoble – 5 avril 2017

OpenStack à l'IPHC

Pourquoi un Cloud IaaS ?

Pour répondre à des demandes spécifiques

- Utiliser un OS ou des outils spécifiques
- Déployer des infrastructures de test (rapidement et simplement)
- Intégration avec des outils supportant le Cloud nativement
- Anticiper les besoins (conteneurs, …)
- Effectuer des tâches longues
- Savoir facilement intégrer les réponses aux besoins de calcul hors norme
- Gestion des logiciels propriétaires par équipe
- Pouvoir déborder sur les centres partenaires

Contraintes

- Pas d'investissement initial
- Pas de RH supplémentaire
- Être compatible avec nos partenaires historiques (LHC, IFB, EGI, …)
- Les performances doivent être là (calcul scientifique)
- Faible contrainte de temps
- Disponibilité annoncée > 95 %

Le choix d'OpenStack

Une solution industrielle

- Logiciel largement adopté par les communautés académiques et privées
- Documentation abondante
- Communauté très active
- Logiciel libre
- Développement actif (release tous les 6 mois)
- Des API implémentées dans plusieurs outils pour les utilisateurs
- Support de l'API EC2
- Possibilité de se connecter facilement à différents fournisseurs d'identité
- API accessible facilement avec n'importe quel langage
- Architecture modulaire et basée sur des technologies standards
- Extension avec un système de plugins (Python)
- Extension fournit par les fournisseurs (CISCO, Junyper, DELL EMC, ...)

Un support par les éditeurs des principales distributions Linux

- RedHat
- Ubuntu
- SUSE

Pour débuter : un *test bed*

Le *test bed* en 2013

- Réutilisation d'un châssis HP SL6500
- Un nœud pour les services Cloud
- Un nœud pour le service réseau
- Six hyperviseurs
- Scientific Linux 6
- OpenStack Grizzly, puis Havana

Limitations

- Pas de distinction entre les réseaux
- Pas de stockage permanent (Cinder)
- Connectivité limitée à 1 Gb/s

Infrastructure de production

Infrastructure 2014

- Financement de deux serveurs (*cloud controller* et *network*)
- Récupération d'un châssis M1000e de la grille
- Réseau dédié
- Scientific Linux 6
- OpenStack Icehouse
- 160 VMs (1 cœur, 2 Go de RAM et 20 Go de disque)
- Dédiée au calcul scientifique (Proxmox pour la virtualisation des services à l'IPHC)
- Déploiement centralisé avec Quattor

Infrastructure de production

Infrastructure 2017
- Hyperviseurs récents (jusqu'à 48 coeurs et 512 Go de RAM)
- ~ 500 coeurs, 3 To de RAM et 40 To de disques
- Connectivité à 10 Gb/s
- Neutron / OpenVSwitch

Logiciel
- Déploiement centralisé avec Quattor
- CentOS 7
- OpenStack Mitaka (bientôt Newton)
- Supervision Nagios (matérielle et fonctionnelle)
- Log centralisé

Et l'humain dans tout ça ?
- Trois personnes s'occupent de l'infra (~ 0,5 FTE)
- Coopération avec des partenaires (formations, échange techniques, projets, ...)

Infrastructure de production

Cloud Manager

- Nova (scheduler)
- Glance (image service)
- Keystone (identity service)
- Cinder (storage)
- Neutron (network)
- Heat (orchestration)
- Magnum (docker)

- R720xd
- 16 coeurs
- 64 Go de RAM
- 4 To stockage iSCSI
- 2,2 To stockage VMs

Network Node

- Neutron (network)
- R720
- 8 coeurs
- 32 Go de RAM
- 2 x 10 Gb

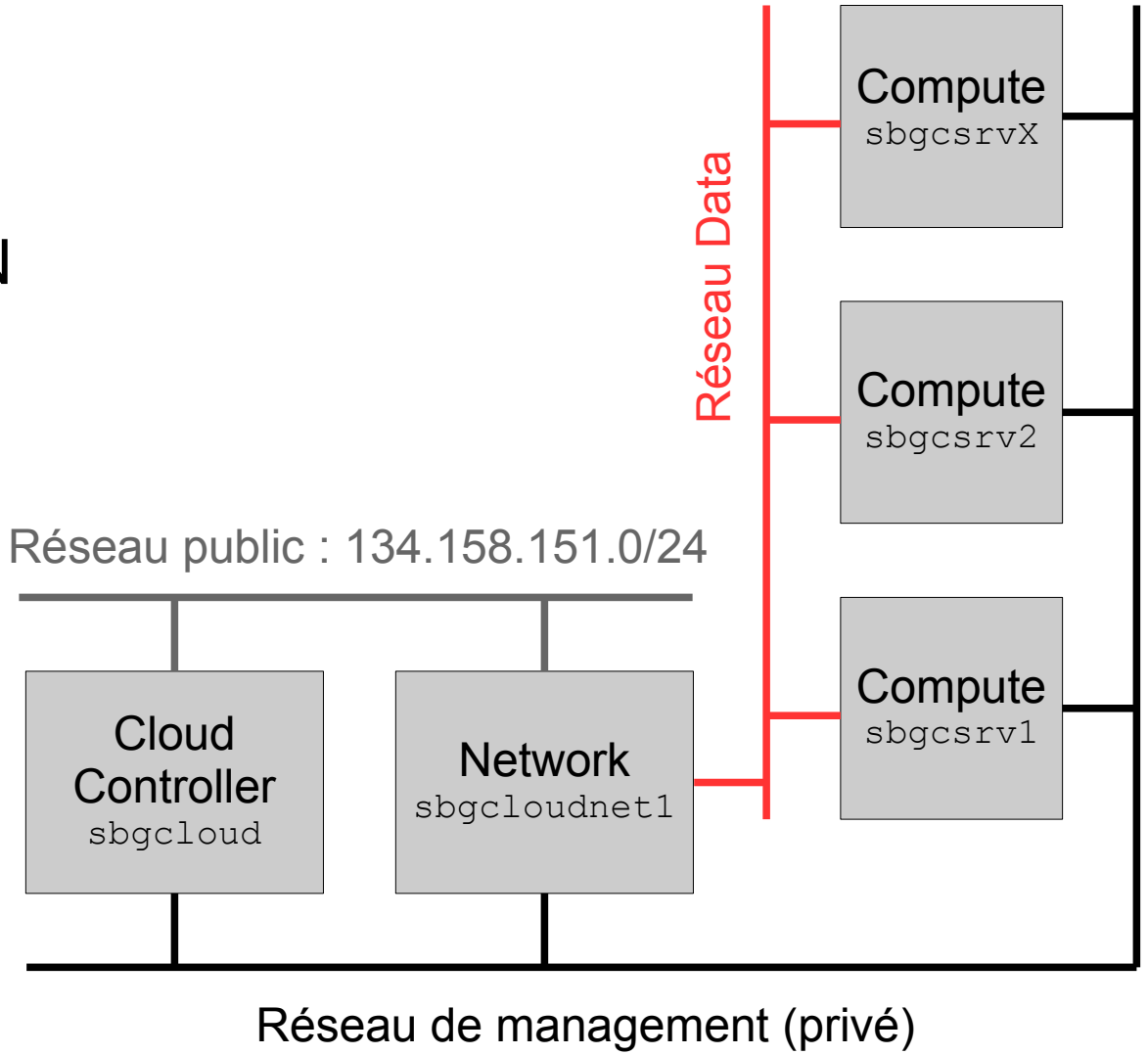
Compute Nodes

- Nova (compute)
- Neutron (network)

Infrastructure de production

Neutron

- Plugin ML2
- OpenVSwitch
- Isolation VLAN



OpenStack en pratique

Dans la pratique

De manière générale

- Installation simple des composants
- Mais configuration du réseau peut être complexe
- La documentation abondante peut aussi être un piège
- La communauté est très active
- Suivre les releases – les mises à jour se passent de mieux en mieux :-)
- Ne restez pas seul, le projet bouge vite !
- Les outils d'administration manquent de *polish*, il arrive encore de devoir modifier directement la base de données

Gestion de l'infrastructure

- Les services sont très stables
- Faire attention lors des déploiements des nouveaux services : charge supplémentaire sur MySQL et RabbitMQ
- Problème de passage à l'échelle avec iscsi / lvm : s'orienter vers un autre backend pour Cinder (CEPH, ...)
- Scheduling / extinction des VMs : pré-emption, VM à durée de vie limitée

En pratique

Exploitation

- Migration entre les nœuds nécessitent des CPUs aux fonctionnalités équivalentes
- Espace disque partagé pour la migration à chaud
- Ne pas sous-estimer la place pour le stockage des VMs
- Certaines erreurs ne sont pas correctement logguées (il faut activer le mode debug et verbose)
- Les logs !

Sécurité

- Surveiller l'utlisation du réseau
- En cas d'utilisation du NAT par les VMs → logguer !
- La traçabilité n'est pas triviale dans OpenStack
- Chiffrage des disques pour les utilisateurs

En pratique

Intégration dans les fédérations de Cloud

- Création de projets, utilisateurs et / ou domaines
- Gestion des images et de l'authentification
- Supervision
- Multiplication des cas d'utilisation == amélioration de la robustesse de la plateforme
- EGI
  - Nécessite l'installation d'outils spécifiques (paquets RPMs ou DEB)
  - Ouverture de ports spécifiques à certaines API (EC2, OCCI)
  - Authentification spécifique (VOMS)
  - Accounting et publication des informations

Développements

Historique de l'assignation des IPs

- Traçabilité des IPs
- Historique par VM
- *Trigger* MySQL
- Licence Apache 2.0
- https://github.com/FranceGrilles/openstack-triggers

Outils d'exploitation

- Suivre l'utilisation des ressources par les utilisateurs
- Licence Apache 2.0
- https://github.com/Pansanel/openstack-user-tools

device id	user name	associating date	disassociating date
3e2767b0-c0f7-43e6-a1aa-c92e0e016190	/C=IT/O=INFN/OU=Personal Certificate/L=Bari/CN=Vincenzo Spinoso	2015-11-30 14:11:24	2015-11-30 14:19:03
bca349ba-a3cb-4311-8722-797600630090	/O=GRID-FR/C=FR/O=ISCPIF/CN=Seyyedmaziyar Shariatpanahi	2016-02-02 11:58:45	2016-02-11 10:34:01
431e52c2-9141-48fd-b7e3-c1d59646ff0a	fg_formation_user20	2016-04-28 16:01:47	2016-04-29 09:44:50

Développements

Isolation des VMs dans Nova

- Par défaut, les utilisateurs peuvent accéder aux VMs dans le même tenant (projet)
- Depuis la version 2.1 de l'API Nova, ce comportement ne peut plus être limité
- Vincent Gatignol a développé un ensemble de patches pour retrouver le comportement précédent
- Fonctionnel pour le code distribué via RDO
- Disponible sur Github : https://github.com/FranceGrilles/cloud-security
- Préparation de paquets RPM en cours

Pour les autres modules ?

- Importance de bien comprendre le fonctionnement de la sécurité par rôle
- Droit définis dans le fichier `policy.json`
- L'API v3 de Keystone (domain) va permettre de faciliter la gestion des droits (en la complexifiant ?)

https://blueprints.launchpad.net/nova/+spec/user-id-based-policy-enforcement
https://pdfs.semanticscholar.org/af46/9a71e17ad74a74496d59b8c412633587eabf.pdf

Cas d’utilisation

Bioinformatique

Biologie des systèmes

- Étude de la transmission d'informations entre animaux
- Analyse de réseaux sociaux
- Basé sur des scripts R et des bibliothèques externes (*RSiena*)
- Nécessite des VMs avec une grande quantité de mémoire (> 200 Go)
  → http://dx.doi.org/10.3389/fpsyg.2016.00539

Institut Français de Bioinformatique (IFB)

- Noeud français du projet Elixir
- Opère un Cloud communautaire pour la bioinformatique
- IFB-Core, un Cloud PaaS / SaaS hébergé à l'IDRIS adossé à des sites secondaires (satellites)
- Deux de ces satellites sont également membre de FG-Cloud (IPHC et Université de Lille 1)
- Noeud régional strasbourgeois BISTRO

LHC et imagerie

CMS
- Expérience du LHC
- Utilisation opportuniste des infrastructures Cloud
- Accès via EGI (OCCI) et direct (EC2)

Phénotypage à haut-débit
- Projet collectant de nombreux données phénotypiques
- Stockage basé sur iRODS (~ 1PB)
- Analyse de données utilisant les infrastructures de grille et de Cloud
- Fonctionnement concluant du workflow Cloud

Virtual Imaging Platform (VIP)
- Portail Web pour la simulation médical et l'analyse d'image
- Utilisation des outils DIRAC et de la VO Biomed
- Utilisation des ressources Cloud de l'IPHC et du CC-IN2P3
  → http://dx.doi.org/10.1109/TMI.2012.2220154

Creatis

Partenaires

Associé à plusieurs projets

- Mésocentre de l'Université de Strasbourg
- Cloud France Grilles :
  https://conf-ng.jres.org/2015/document_revision_2500.html
- Cloud IFB :
  http://www.france-bioinformatique.fr/fr/cloud
- Cloud fédéré EGI :
  https://www.egi.eu/infrastructure/cloud/
- Expérience WLCG (CMS)

Travaux en cours & perspectives

En cours
- Déploiement d'Ironic
- Test d'une solution de stockage objet (CEPH)
- Documenter Magnum
- Migration Newton
- Security Challenge
- Déploiement d'outils pour EGI (distribution d'images, sécurité)

Perspectives
- Security Challenge
- Augmenter la volumétrie du stockage objet
- Intégrer des outils pour le scheduling des VMs
- Neutron DVR

Questions ?