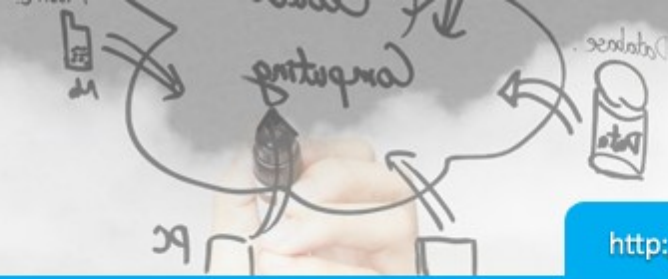




Déploiement d'un Cloud de calcul scientifique

Vincent Legoll et le groupe FG-Cloud
vincent.legoll@IDGRILLES.FR



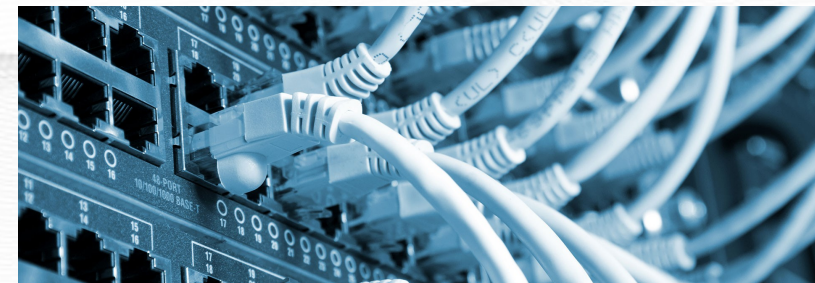
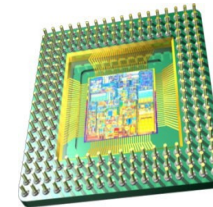
Éléments constitutifs d'un cloud

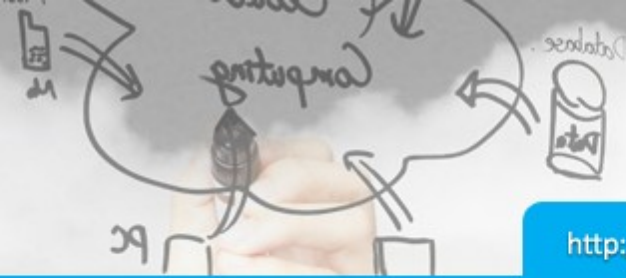
Spécificités liées au calcul scientifique

Cloud IaaS

- Partage des ressources hétérogènes d'une ferme de serveurs

- Cœurs de CPU
- Mémoire
- Stockage
- Réseau





Virtualisation

- **Avantages**

- Isolation : sécurité
- Mutualisation : rentabilité
- Approvisionnement : élasticité
- Abstraction du matériel : homogénéité
- Découplage host / guest : administration OS
- Utilisations de périphériques virtuels : performances

Calcul

- **Processeurs, Mémoire**

- Choix des hyperviseurs : KVM, Xen, LXC, QEmu, VMWare, etc.

Compromis : performance, isolation, fonctionnalités

- Sur-allocation des ressources :

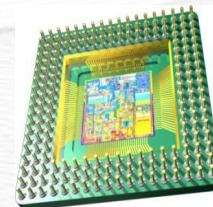
Par défaut, pour nova : 1600 % CPU et 150 % RAM - serveurs web peu chargés

Pour du calcul scientifique : dépasser 100 % fait perdre de la performance

nova.conf : `cpu_allocation_ratio = 1` `ram_allocation_ratio = 1`

- Modèle et caractéristiques des processeurs : natif ou émulé

nova.conf : `cpu_mode = host-passthrough`



Stockage

- Données, disques virtuels

- Catalogue d'images d'OS

- Permanent : périph. block formaté avec système de fichiers

Données utilisateur, taille moyenne

- Objet : non structuré, accès par APIs, métadonnées

Source de données pour le calcul, taille « Big Data »



Stockage

- Avantages, possibilités

- Sauvegardes
- Instantanés de VM
- Cache local des images
- Migration des VMs
 - à chaud possible si stockage partagé
 - à tiède sinon (VM en état pause)
- NFS, Ceph, GlusterFS, Lustre, autres?



Communications

- Réseaux virtuels

- Matériel : greffons pour switches pilotés
- Routeurs virtuels (OpenVSwitch, OpenFlow)
- Interne : un sous-réseau distinct par groupe d'utilisateurs
- Externe / public : # IPs disponibles / nécessaires -> NAT



Communications

- Possibilités
 - Encapsulation : tunnels GRE, coûts associés
 - Étiquetage : VLAN, pas de coûts supplémentaire, supporté par le matériel
 - JumboFrames 9k, attention aux problèmes DHCP, MTU
`dnsmasq.conf : dhcp-option-force=26,1454`
- Linux sur vos commutateurs (cumulusnetworks.com)



Sécurité

- Définition des conditions d'accès au services
- Responsabilités
 - Images OS : vérifiées, certifiées
 - VMs : utilisateurs
 - Infrastructure : Vous
 - Traçabilité (centralisation & rétention logs, BD) : Vous
- Alertes vulnérabilités => maintenance (vous et utilisateurs)



Sécurité

- Responsable sécurité de France Grilles : Sophie Ferry
- Piratage
 - Détection VM exploitées, alerter les responsables
 - Blocage d'accès / extinction VMs
 - Analyse
- Procédure de réponse :
<https://forge.in2p3.fr/projects/francegrilles-ops/wiki/Security>



Sécurité

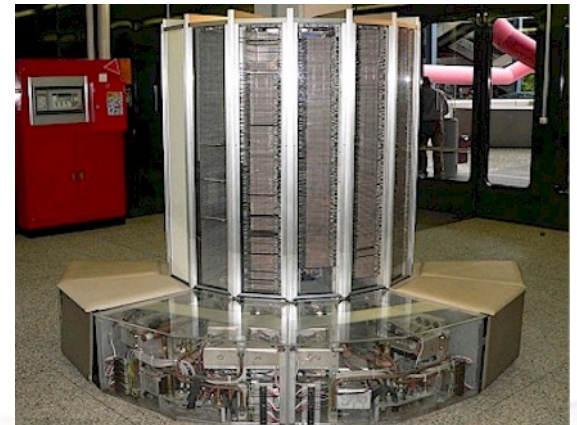
- **Outils**
- Infrastructure :
 - Surveillance (Nagios, Ganglia, pakiti)
 - Analyseur de trafic : ZNeTS pour loguer les connections
- Cloud :
 - Groupes de sécurité, RBAC
 - Règles de pare-feu



HPC / HTC et cloud ?

- **Problématique**

- Scalabilité des calculs
- Besoins en bande passante
- Sensibilité à la latence
- Ratio : IO / CPU



HPC / HTC et cloud

- **Possibilités**
 - Ordonnanceur de VM : quelles VMs sur quel HV
 - PCI-e passthrough : carte dédiée à une VM
 - GPGPU, SIMD, OpenCL, Cuda (Xen et VmWare : OK)
 - Infiniband, SR-IOV, Eth10G
 - Bare-metal : fonctionnement sans hyperviseur,
pas de virtualisation,
coût d'installation de machine



OpenStack

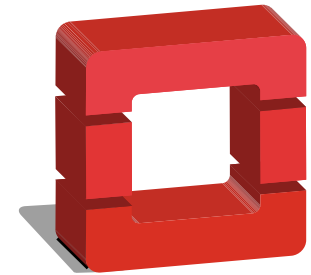
Déploiement



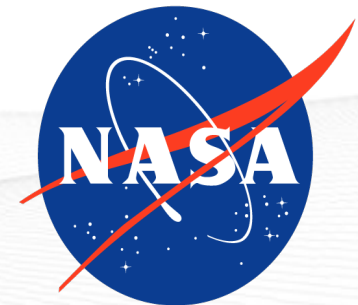
OpenStack

• Historique

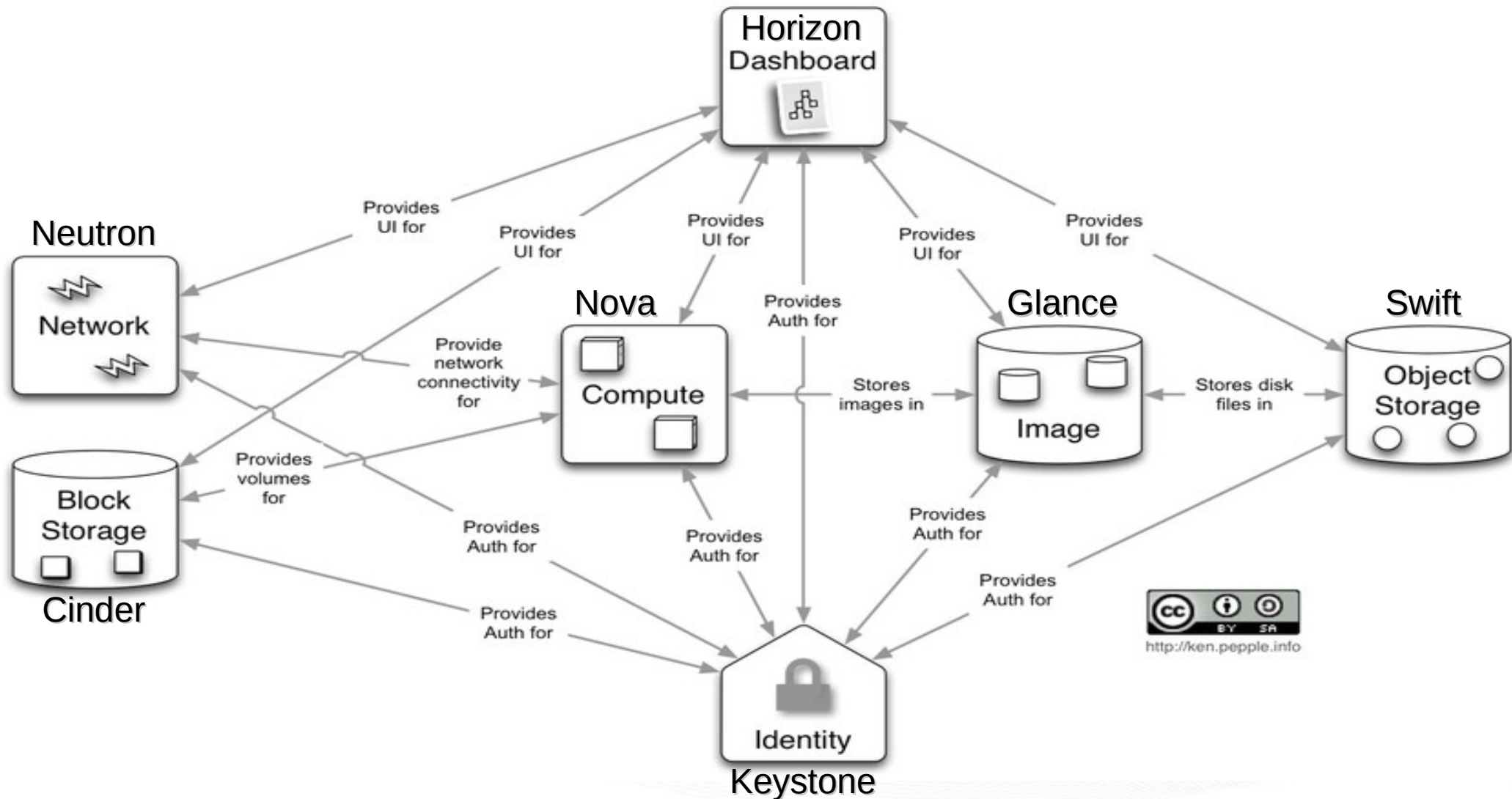
- Nasa & Rackspace
- Fondation non-commerciale
- Support de nombreuses sociétés :
AMD, Canonical, Cisco, HP, Intel, IBM, Oracle, RedHat, etc.
- Multiple composants, principalement en python



openstack®
CLOUD SOFTWARE

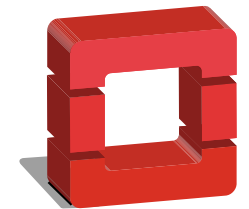


Architecture d'OpenStack



OpenStack

- Rythme : nouvelle version tout les 6 mois
- Supportée pendant 1 an 1/2



openstack
CLOUD SOFTWARE

<u>Nom de code</u>	<u>Date de sortie</u>	<u>Support</u>
Havana	Octobre 2013	Plus supportée
Icehouse	Avril 2014	Supportée (*)
Juno	Octobre 2014	Supportée (*)
Kilo	Avril 2015	Courante, stable (*)
Liberty	Octobre 2015	En développement
Musashi ???	Avril 2016	Futur

(*) Correctifs de bogues et de sécurité

OpenStack

- **Système d'exploitation**

- RHEL, Ubuntu, OpenSuse, Scientific Linux, CentOS, etc...
- Dépôts de paquets OpenStack
 - RedHat : RDO (www.rdoproject.org) + EPEL (fedoraproject.org/wiki/EPEL)
 - Ubuntu (ppa:cloud-installer/stable)
- La version d'OS n'est pas toujours "au choix"
 - Centos 7 minimum pour Juno
 - Ubuntu :
 - 12.04 Havana
 - 14.04 Icehouse
 - Sinon : « `add-apt-repository cloud-archive:juno` »



OpenStack

• Installation, déploiement

- La doc officielle n'est pas parfaite, fragmentée :
 - Voir les autres versions
 - Cherchez sur ask.openstack.org
 - bugs.launchpad.net
 - Blogs, Wikis, etc...
- Prenez des notes exhaustives
- Utilisez toutes les options de log verbeux



OpenStack

- Installation, déploiement

- Fragilité de la config :

Utilisez quattor si possible

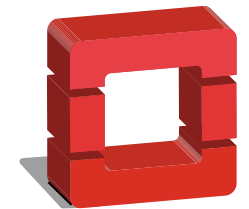
compute / network (validé / testé) / controller (pas encore validé)

- La partie réseau est la plus complexe

- Si vous testez dans des VMs, c'est encore pire

- Désactivez le firewall

- Testez à chaque étape

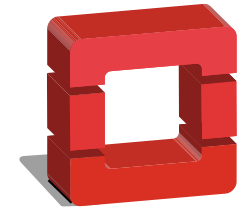


openstack[®]
CLOUD SOFTWARE

OpenStack

- Installation, déploiement

- NTP est important
- BD UTF-8

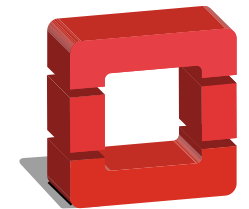


openstack
CLOUD SOFTWARE

OpenStack

- Administration, maintenance

- Migrations :
 - Matériel nécessaire pour tester (spare)
 - Interruption de service
 - MaJ qui casse encapsulation GRE (rencontré au LPSC)
- Redémarrage après modification(s)
- Ordonnancement des services



openstack
CLOUD SOFTWARE

OpenStack

• Problèmes

- Fichiers de log dispersés
- Configuration dupliquée entre les nœuds
- Attention aux typos
- Inconsistance des tables BD : à Recréer
- Définitions des points d'accès API
- CirrOS & clavier non-US



OpenStack

• Debug réseau

- Tcpcmdump br-int
- ip netns qrouter-XXXXXX exec ip a
- ifconfig
- OpenVSwitch
 - ovs-vsctl show
 - Port miroir
- Recréer certaines parties du réseau :
 - Suivre les dépendances (c'est pas automatique)
 - attention aux tables BD



OpenStack

- Problématique utilisateurs

- Démarrage manuel des VMs
- Confidentialité des données (par ex. médicales) nécessitant un agrément
- Cycle de vie, nettoyage : VMs, images, UUIDs, réseaux, etc...

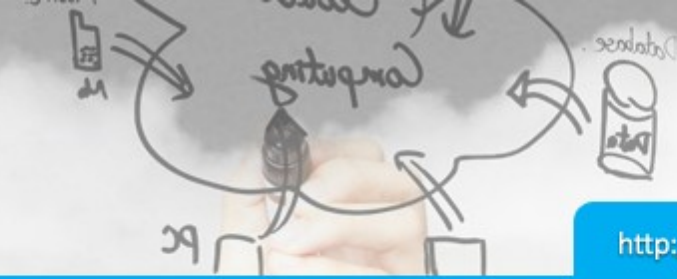


OpenStack

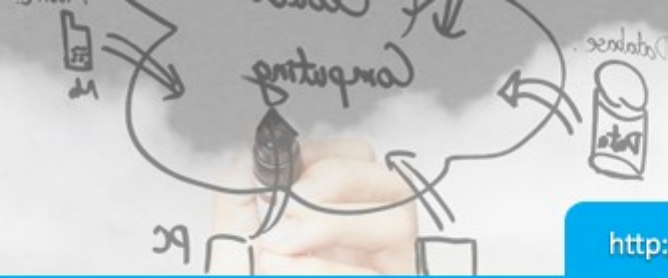
- Problématique des utilisateurs

- Accès console distant : VNC, Spice
- VM Windows & licenses
- Interface / orchestration (dashboard, heat, Slipstream, FogBow)
- Graphes d'utilisation par : utilisateur / VM / tenant
- SSH pour voir l'état d'avancement, de succès





Questions ?



Merci