

Machines virtuelles a tetaneutral.net avec du
libre de Ganeti a Proxmox VE + Ceph en
passant par Openstack et libvirt

Laurent GUERBY

- Ingénieur de formation
- Développeur GCC (compilateur GNU, langage Ada)
- Ancien vice-président de l'APRIL
- Libriste invétéré
- Mais aussi open data, open hardware, enercoop

Association tetaneutral.net

- Création 3 janvier 2011 à Toulouse <https://tetaneutral.net>
- Membre fondateur de la Fédération FDN <https://ffdn.org>
- Défense de la neutralité du réseau
- Éducation populaire sur le réseau et les infrastructures
- 420 adhérent-es cotisant-es, plusieurs centaines gratuit
- Modèle prix participatif pour les services
- Pas de subventions, auto financé

Services tetaneutral.net

- Accès internet par réseau radio “de fenêtre à fenêtre”
- Sur Toulouse et quelques communes
- Accès internet fibre (encore balbutiant, ~ 70)
- Hébergement de machines physiques format libre (~100)
- VPN wireguard (~20)
- Et hébergement de machines virtuelles (~210)

La virtualisation

- Pour gérer des machines virtuelles (“VM”) il faut un logiciel de type hyperviseur
- L’hyperviseur peut être un système d’exploitation indépendant, ou s’ajouter à un système existant
- En libre : libvirt, Ganeti, OpenStack, proxmox VE, oVirt, open nebula, xen, ...
- Monde propriétaire : VMware (considéré comme le leader du marché), HyperV
- Tetaneutral.net fait toujours le choix du logiciel libre

Première partie : ganeti

- Hyperviseur développé par Google pour ses besoins propres (environnement de dev), GPLv2, 2007
- Intègre Xen ou KVM, LVM, DRDB
- 4 PC Tour avec réseau gigabit et stockage DRDB
- Données copiées sur deux machines
- A Mix'Art Myrys
- Premier service de VM associatif à prix libre

Deuxième partie : openstack

- Logiciel libre ambitieux, lancé par NASA et rackspace en 2010
- Licence Apache 2.0
- Installation en 2014 sur de nouvelles machines, version Havana
- Stockage ceph (redondance triple : disque et machine)
- Migration des VMs de ganeti à openstack
- Aide de Mehdi et Loïc qui travaillaient sur openstack à titre professionnel
- Développement sur mesure pour la partie réseau d'openstack

Réseau BGP openstack

- Idée : ne pas partager le réseau “niveau 2” entre les VMs
- Chaque VM a une interface réseau virtuelle “tap”
- Ces “tap” ne sont pas mis dans un “bridge” (sécurité)
- Un “hook” openstack rajoute la route explicite vers l’IP de la VM sur le tap
- Le serveur BGP “bird” distribue les routes des VMs au cluster
- Et au routeur tetaneutral.net

Openstack sur la durée

- Logiciel complexe
- Debian vs Ubuntu
- Bugs demandant des interventions manuelles
- Difficulté de certaines mises à jour
- Adhérents compétents en openstack moins disponibles
- Openstack est basé sur libvirt pour la gestion locale des VMs
- Décision d'abandonner la couche openstack

Troisième partie : libvirt + ceph

- Désinstallation de la couche openstack
- En gardant les “XML” qui décrivent les VMs
- Développement de scripts autour de libvirt
- Gestion a minima mais suffisante
- Le stockage inchangé sur ceph donc pas de migration sur ce point là
- Reflexion sur le matériel

Interlude quincaillerie

- 14 machines PC Tour avec disques SATA interne
- 60+ disques, remplacements périodiques
- Temps de bénévole conséquent démontage/remontage
- Une seule alimentation 230V
- Décision de changer le matériel
- Achat de 7 boîtiers avec 16 tiroirs hot “swap” et double alimentation 230V, Ryzen, 128GB RAM ECC
- Migration des VMs et disques sur le nouveau matériel

Quatrième partie : Proxmox VE

- Proxmox VE est un logiciel libre AGPL
- Basé sur debian, PVE 7.x = debian 11, PVE8.x = debian 12
- Kernel Ubuntu avec ZFS
- QEMU/KVM dernière version “upstream”
- Codé en perl, listes de diffusion, contributions de la communauté
- Support commercial disponible avec un tarif d’entrée raisonnable (~ 100 EUR/an).
- Support GTR 4h et gestion SSH pour le critique

Proxmox VE - 2

- Interface web plutôt bien faite
- Plus familier pour nos bénévoles que libvirt avec des scripts maison
- Intégration de ZFS et de ceph, aussi en web
- API, “CLI”, édition des fichiers “synchrones”
- Gestion des migrations de version bien documentée et rodée (outillage et wiki)
- Décision de migration de libvirt+ceph à proxmox+ceph

Jeu de taquin

- 7 hyperviseurs, 210+ VM, 180 TB de stockage brut, 80 TB utilisé
- Sortie d'une machine du cluster libvirt+ceph et réinstallation en PVE 7.4 avec 6x4TB SSD en ZFS
- Remplissage de la machine PVE avec les plus grosses VMs
- Jusqu'à pouvoir vider et installer en PVE deux machines de plus
- 3 machines en PVE donc création du stockage ceph
- Migration des VMs restantes sur les 4 machines de l'ancien cluster vers les 3 machines PVE

Fin de la migration

- Enfin intégration des 4 machines dans PVE, cluster de 7 hyperviseurs avec toutes les VMs
- Impact sur nos adhérent-e-s : quelques lenteurs, un seul redémarrage de leur machine virtuelle
- Le 11 août 2023 les 216 VMs sont migrées
- Le 18 août 2023 mise à jour de PVE 7.4 en PVE 8.0 sans impact adhérent
- Isolation réseau en utilisant un VLAN par machine virtuelle

Petite peur

- Lors des premiers tests PVE sur nos 3 machines
- Après 6 à 10 jours “idle” sans VM, “freeze” d’une machine mais pas les deux autres
- De longues recherches, mises à jour de BIOS, kernel, etc...
- Le problème persiste (test pénible vu le délai)
- Décision de migrer quand même
- Ouf ... si pas “idle” pas de “freeze”

Bilan

- 3 mois sans problème avec PVE+ceph
- Utilisation de cloudimage pour créer les VMs
- Avec l'API REST PVE, librairie python "proxmoxer"
- Réseau purement systemd-networkd, suppression de netplan.io
- (qui génère un fichier systemd in fine plus petit que la configuration netplan.io)
- Tetaneutral.net peut continuer à créer des machines virtuelles pour ses adhérent-e-s.