Huma-Num

la TGIR des humanités numériques

Outils de virtualisation de serveurs

Joël Marchand

TGIR Huma-Num - CNRS UMS 3598

11 décembre 2014

1 Introduction

2 Généralités

3 VMware

4 KVM

5 Virtualisation légère

Huma-Num

- Huma-Num est une Très Grande Infrastructure de Recherche (TGIR) visant à faciliter le tournant numérique de la recherche en Sciences Humaines et Sociales (SHS).
- Elle développe également un dispositif technologique unique permettant le traitement, la conservation, l'accès et l'interopérabilité des données de la recherche en SHS.
- Ce dispositif est composé d'une grille de services dédiés, d'une plateforme d'accès unifié (ISIDORE) et d'une procédure d'archivage à long terme.
- La TGIR Huma-Num est portée par l'Unité Mixte de Services 3598 associant le CNRS, l'Université d'Aix-Marseille et le Campus Condorcet.

Préambules

- Je ne suis pas un expert système sur les techniques de virtualisation, d'isolation, d'allocation de ressources, etc
- Je suis loin d'avoir tout compris et digéré les pointeurs récoltés !
- Je ne suis pas un exploitant chevronné de larges plateformes
- Mon expérience : un petit peu de VMware Server et de Linux-Vserver de 2004 à 2010

Mon cahier des charges

- faire un état des lieux des outils de virtualisation serveurs
- pour une unité faisant de l'hébergement type Web
- pour une dizaine de serveurs physiques et un NetApp
- dans un environnement essentiellement Linux de type Redhat
- avec très peu d'ETP pour installer/administrer au long cours
- dans le but d'une exploitation paisible et simple à distance
- en phase avec ce qui se fait dans la communauté ESR
- en préservant au maximum les possibles pour l'avenir
- sans contrainte forte de disponibilité ou de performances ou d'optimisation de l'usage des ressources
- sans besoin au départ de délégation de gestion externe des VM
- avec envie de pouvoir faire de la virtualisation lourde et légère
- en ne s'engageant pas dans des dépenses financières récurrentes intenables à terme

Périmètre - 1/2

- J'ai limité mon étude à
  - écosystème VMware
  - écosystème KVM
  - technologies de conteneurs Linux : Linux Vserver, OpenVZ, LXC
- Donc je n'ai pas regardé
  - Xen et Citrix
  - Oracle
  - Hyper-V
  - jails BSD
  - zones Solaris

Périmètre - 2/2

- Je n'ai pas regardé non plus
  - sauvegarde de VM : snapshots, mirroring, Veeam
  - supervision de VM : par l'outil de gestion, par le host, Veeam
  - config et virtualisation de réseaux : agrégation, vswitchs, SDN
  - virtualisation pour postes clients : VirtualBox, VMware Workstation
  - virtualisation de postes clients (VDI) : VMware View
  - virtualisation et technologie de stockage centralisé
  - outils de niveau Cloud : Vcloud, OpenStack, OpenNebula
  - choses nouvelles : OpenNode OS, Docker, Nutanix
- Toutes les manipulations d'installation se sont faites sur un PC ordinaire
  - soit sous Ubuntu 14.04
  - soit depuis le hardware nu

Observations générales - 1/2

- Chacun est content des outils qu'il utilise
- Personne ne dit devoir en changer rapidement
- Tout le monde ou presque utilise un stockage centralisé, principalement en ISCSI ou NFS
- Beaucoup de gens utilisent des outils gratuits basés sur ESXi et KVM
- Plus l'infrastructure est grosse et critique, plus elle est outillée et sécurisée
- Il n'y a pas d'argument définitif pour choisir entre les technologies

Observations générales - 2/2

- Nombreuses variables
  - contexte : DSI / labos
  - besoins : systèmes, applicatifs, disponibilité, support
  - moyens financiers et RH
  - prise de risques
  - le bon outil est celui que l'on maitrise
- Pas mal de structures importantes (DSI, COMUE, UNR) travaillent pour automatiser, optimiser, déléguer, facturer : outils Cloud
- OpenStack semble s'imposer
  - Cloudwatt
  - Numergy
  - OVH
  - maquettes au sein de l'ESR

Migration d'une VM

- D'après les quelques tests effectués, on peut faire passer une VM d'un outil à un autre : VirtualBox, VMware, les outils KVM
- Ceci relativise l'impact dans le temps de choisir telle ou telle solution de gestion de VM
- Migration = recopie du disque virtuel + redéfinition dans l'outil de gestion + parfois usage d'outils de conversion du disque virtuel
  - vmkfstools (ex vmware-vdiskmanager)
  - VBoxManage
  - qemu-img
- Plus difficile apparemment au niveau Windows
  - nombreuses pistes essayées
  - seules solutions effectives
    - VM Converter (lent)
    - passage de sysprep avant de copier la VM

VMware - points positifs

- c'est le premier acteur historique
- c'est la référence : usagers, documentations, formations, compétences
- c'est le choix du DSI
- fonctionnalité de snapshots très au point et très confortable
- délégation possible de gestion externe des VM
- outils de couplage et d'intégration pour les baies de stockage (NetApp, ...)
- fonctionnalités aussi évoluées que l'on veut, suivant ce qu'on paie : vmotion, storage vmotion, DRS, HA, FT, ...
- avance fonctionnelle sur l'écosystème KVM
- cf. offre IAAS de la DSI du CNRS basée sur vCloud

VMware - points négatifs

- solution propriétaire fermée
- complexité de l'écosystème
- coût des briques payantes
- besoin d'une VM Windows
  - soit pour vSphere Client
  - soit pour avoir le bon Flash pour vSphere Web Client (!)
- besoin d'installer les VMware Tools dans chaque VM
- logique de ne pas "mettre ses doigts" dans le host (sauf pour quelques commandes CLI)
- pas de possibilité d'y installer des outils supplémentaires (supervision, métrologie)

VMware - observations

- limite de 1 To de RAM par VM
- usage d'un stockage centralisé
  - en FC (hier ?)
  - en ISCSI (ex : Dell EqualLogic)
  - en NFS (ex : NetApp)
- stockage NFS
  - tendance montante ?
  - plus simple à administrer, pas de réseau SAN
  - allocation dynamique du stockage plus simple
  - robustesse et performances sur un NetApp
- certains sites importants, anciens utilisateurs, envisagent de se séparer de VMware et de migrer vers KVM et/ou des conteneurs

VMware - prix 1/2

- produits gratuits
  - hyperviseur nu ESXi
  - client vSphere
  - vSphere Command-Line Interface
  - VM Converter
- marché national via le Groupe Logiciels
  - SCC pour la fourniture de licences
  - Cheops pour les prestations d'assistance technique
- conditions
  - dixit SCC : 40% de remise (tarif académique) sur l'ensemble des produits, puis une remise de 17,89% pour la licence acquise et 20,83% pour le support associé en Basic
  - dans le tableau Excel récupéré par le Groupe Logiciels : 3% sur prix académique

VMware - prix 2/2

- choix possibles
  - 3 niveaux de vSphere
  - 3 niveaux de vSphere with Operations Management
  - 2 niveaux de vCenter Server
  - 2 niveaux de support
  - packs VMware vSphere Essentials kits pour 3 serveurs bi-sockets
- exemple
  - en support basic
  - vCenter Server Standard
  - sur 5 ans
  - pour 8 serveurs bi-sockets
  - de 18 k€ pour vSphere Standard à 65 k€ pour vSphere with Operations Management Entreprise

VMware - avis personnel

- écart trop important entre ce qui est gratuit et ce qui est payant
- mais quid de gérer n»1 hyperviseurs indépendants à la main ?
- rapport entre prix du matériel et prix du logiciel excessif
- crainte de la situation d'acteur dominant ayant changé plusieurs fois de politique commerciale

KVM - généralités 1/2

- Natif depuis de nombreuses années dans le noyau Linux
- Plusieurs dispositifs de gestion
  - libvirt/virsh/virt-manager
  - Proxmox
  - ovirt/RHEV
  - OpenNode OS
  - OpenStack
- Quelle raison reste-t-il d'utiliser directement les commandes qemu ?
- Nombreux usages dans l'ESR, surtout dans la Recherche
- Beaucoup de présentations déjà faites

KVM - généralités 2/2

- Plusieurs formats de disque virtuel
  - raw : plus de performance, sécurité, pas de snapshots
  - qcow2 : performance si cache en écriture sur le host, snapshots
- Drivers virtio
  - les préférer autant que possible, pour les performances
  - natifs sur les Linux récents
  - disponibles pour Windows
- Modification de la taille des images : libguestfs
- Déduplication des pages mémoire entre les VM : KSM
- Memory balloning
- Hyperviseur de référence pour OpenStack

Libvirt/virsh/virt-manager - 1/2

- Projet libre, fortement soutenu par Redhat
- Large usage dans la communauté ESR
- Gestion par CLI et GUI X-Window
- Multi-hyperviseurs, y compris ESX, Xen, VMware, ...
- Migration possible de VM d'un host à un autre
- Possibilité de faire ce qu'on veut sur le host
- Pas de notion de template, mais possibilité de clonage
- Pas de notion de cluster, de H-A, de configuration centralisée
- virt-manager permet de gérer aussi des conteneurs LXC

Libvirt/virsh/virt-manager - 2/2

- Points positifs
  - totalement libre
  - maturité du projet
  - simplicité
  - interface CLI
  - configurations explicites en XML
  - outil le plus ouvert et qui obère le moins l'avenir
- Points négatifs
  - passage à l'échelle ?
  - pas de gestion des rôles et donc de délégation possible

Proxmox - 1/2

- Projet développé par 4 frères autrichiens au sein d'une société avec des fonds publics
- Outil libre, mais plusieurs niveaux de support payant (à partir de qqs dizaines d'euros par an et par serveur)
- Large usage dans la communauté ESR
- Librairie différente de libvirt
- Gestion par interface Web bien faite et simple à appréhender
- Stockages LVM, ISCSI, NFS, Ceph
- Migration à chaud et haute-disponibilité possibles
- Installation "bare-metal" conseillée : on ne touche plus au host ensuite
- Configuration centralisée, logique de cluster jusque 16 noeuds
- Gestion équivalente de VM KVM et de conteneurs OpenVZ
- Notion de templates fournis de base, par Turnkey, ...

Proxmox - 2/2

- Points positifs
  - très bon retour des utilisateurs
  - maturité du projet
  - réputation d'OpenVZ
  - gestion des rôles : délégation de gestion
  - gestion intégrée des sauvegardes
  - cluster
  - interface Web simple et fonctionnelle
  - interface CLI disponible pour KVM et OpenVZ
  - hyperviseur opensource basé sur Debian
  - API REST
- Points négatifs
  - pas encore de support de LXC
  - crainte que demain le produit ne soit plus gratuit
  - quid demain pour reprise par OpenStack ?

oVirt/RHEV - 1/2

- Projet développé par Redhat
- Idem que Fedora/RHEL : version libre/version payante supportée
- Basé aussi sur libvirt
- Vise à concurrencer directement VMware vSphere
- Stockages locaux, ISCSI, NFS, GlusterFS, Posix
- Fonctionnalités évoluées
  - migration à chaud, haute-disponibilité,
  - équilibrage de charge, gestion des alimentations électriques
  - gestion des rôles avec bornage des ressources consommées, ...
- Marché national via le Groupe Logiciels
  - SCC pour la fourniture de licences : remise de 18% ?
  - Think pour les prestations d'assistance technique : pas de réponse !
- Prix 2782 € HT pour 5 ans et un serveur bi-sockets

oVirt/RHEV - 2/2

- Points négatifs
  - hyperviseur fermé (comme avec ESXi) basé sur RHEL
  - pas de support de virtualisation légère
  - pas réussi à créer un stockage centralisé et donc pas de VM
  - pas d'exemple d'usage en production dans l'ESR
  - prix de la version commerciale élevé
- Points positifs probables
  - gestion centralisée par interface Web et REST
  - outil d'un niveau équivalent à vSphere, mais sur KVM
  - richesse fonctionnelle la plus grande des 3 solutions
  - outil pertinent à grande échelle

Virtualisation légère - généralités

- Partage du même noyau Linux + isolation des ressources
- Existe depuis plus de 10 ans
- Intégration assez récente des notions de namespaces et de cgroups dans le noyau Linux : donne la capacité de limiter la consommation de chaque VM
- Pas de disque virtuel, mais des arborescences de fichiers

Virtualisation légère - les outils - 1/2

- Linux Vserver
  - implémentation la plus ancienne
  - robuste et simple
  - nécessite un noyau 2.6 ad-hoc, plus fourni par les distributions
  - pas de gestion des ressources
  - pas d'outil de gestion centralisé
- OpenVZ
  - développé par la société Parallels, base du produit Virtuozzo
  - nécessite un noyau 2.6 ad-hoc, plus fourni par les distributions
  - surtout utilisé via Proxmox
  - gestion des ressources
  - bonne réputation
  - notion de checkpoints et de templates

Virtualisation légère - les outils - 2/2

- LXC
  - intégré dans le noyau Linux
  - gestion des ressources
  - le plus récent des trois
  - projet porté par Ubuntu
  - 2 implémentations
    - celle d'Ubuntu : lxc-tools
    - celle visible dans virt-manager
  - au départ, utilisé par Docker, mais ce n'est plus le cas
  - projet Flockport : repository d'applications préconfigurées
  - maturité réputée croissante
  - probablement le projet qui va émerger

Virtualisation légère - les plus/les moins

- Points positifs
  - simplicité de déploiement pour un applicatif et son environnement logiciel
  - moindre consommation des ressources sur le host
  - démarrage plus rapide des VM
  - remplissage des hosts en mode "best effort"
  - gestion en CLI
- Points négatifs
  - impose que la VM tourne sur le même noyau Linux du host
  - pas de migration à chaud ou de H-A
  - pas de délégation de gestion (demain avec LXC ?)
  - aucun support commercial connu
  - bornage de la consommation des ressources moindre ?
  - sécurité moindre ?

Personnes consultées - 1/2

- Les listes
  - virtualisation@services.cnrs.fr
  - asr@services.cnrs.fr
  - mathrice@math.cnrs.fr
- VMware
  - zouhir.hafidi@dt.insu.cnrs.fr
  - david.tsang-hin-sun@univ-tlse3.fr
  - jean-marc.vinet@ifsttar.fr
  - maxime.baudin@ac-rennes.fr
  - albert.shih@obspm.fr
  - olivier.thibault@lmpt.univ-tours.fr

Personnes consultées - 2/2

- KVM
  - francis.monestier@u-pem.fr
  - vishaal.golam@u-pem.fr
  - sebastien.geiger@iphc.cnrs.fr
  - jacquelin.charbonnel@univ-angers.fr
  - jean-luc.bellon@cpht.polytechnique.fr
  - stephane.vaillant@obspm.fr
  - francis.hulin-hubard@lsv.ens-cachan.fr
  - herve.ballans@ias.u-psud.fr
  - sylvain.maurin@isc.cnrs.fr
- Virtualisation légère
  - albert.shih@obspm.fr
  - emmanuel.halbwachs@obspm.fr
  - julien.brule@obspm.fr
- Grand grand merci à tous ces collègues !

Huma-Num

Compléments

- J'ai compilé l'ensemble de mes notes techniques et les pointeurs transmis par la communauté en `http://sharedocs.huma-num.fr/wl/?id=sB&filename=Notes-virtualisation`
- Je suis preneur de vos corrections et compléments sur ces notes via `mailto:joel.marchand@huma-num.fr`
- Merci pour votre attention et tout retour

Huma-Num