



Virtualisation sous Linux

L'age de raison

Daniel Veillard
veillard@redhat.com

Solution Linux 2009

Une jungle d'acronymes

Xen

QEmu

OpenVZ

UML

VServer

KVM

LXC

VMWare

VirtualIron

VirtualBox

Iguest

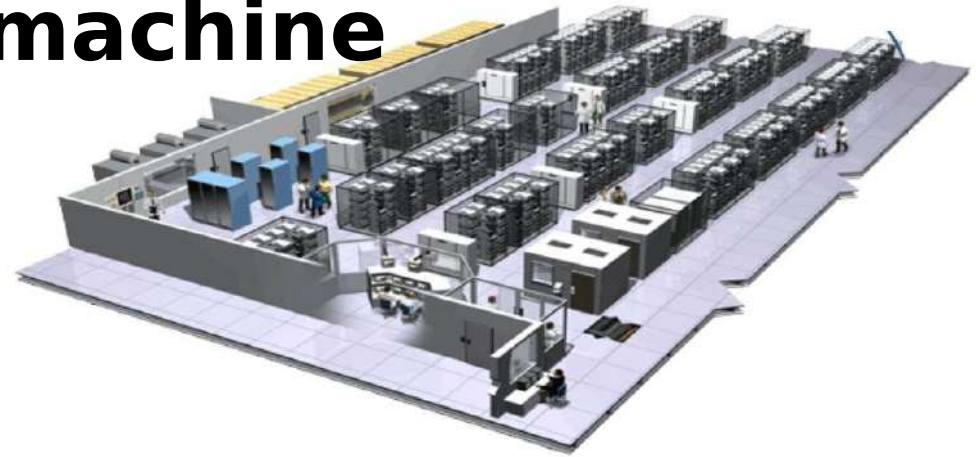
Définition et but

- Définition:
 - Découpler les ressources informatiques de leur implémentation
- But: abstraire le matériel pour s'en affranchir
- Machine Virtuelle
 - Processeurs virtuels
 - Mémoire
 - Périphériques réels ou émulés
- Stockage Virtualisé
 - Disque distants
 - Disques locaux émulés
- Transition du multitâche au multi-OS

Benefices de la virtualisation

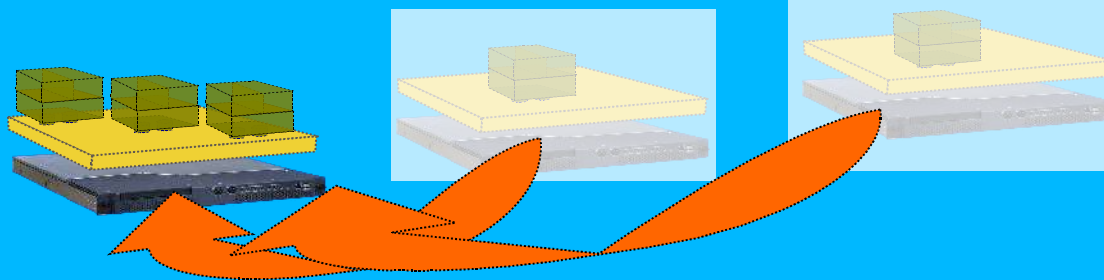
- Les bénéfices sont de deux ordres :
- Réduction des couts
 - Meilleure gestion de la capacité
 - Simplification de la maintenance
 - Mise a l'échelle
- Réduction des risques
 - Isolation et sécurité des services
 - Indépendance matérielle
 - Redondance aisée

Exemple: une salle machine

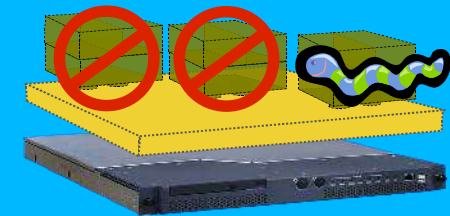


- Contraintes physiques
 - Taille, énergie, climatisation
- Surdimensionnée
 - Utilisation moyenne 10 - 30%
- Un serveur par application/client
 - Simplification de la gestion, facturation
 - Sécurité
- Problèmes
 - Coûts
 - Administration

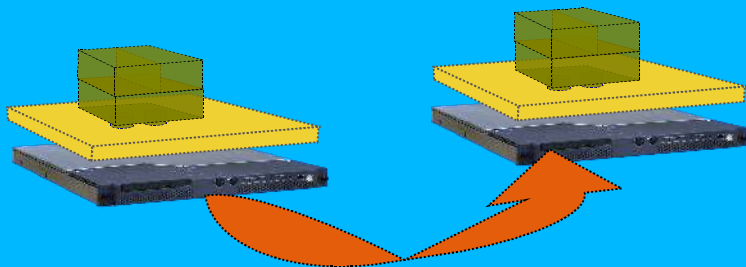
La virtualisation en action



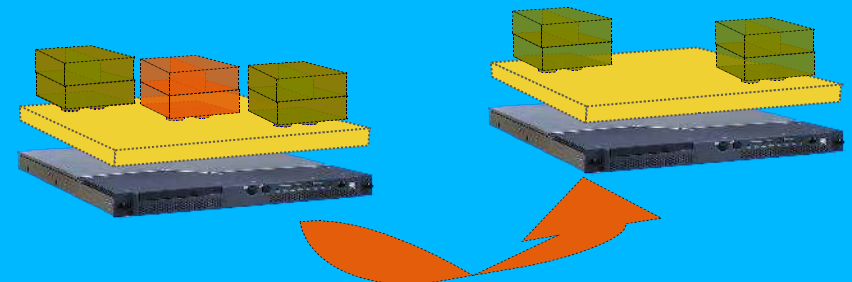
Consolidation de serveurs sous-utilisés



Politique de
Sécurisation



Pas d'interruption de service,
migration



Répartition dynamique de
charge

Des virtualisations

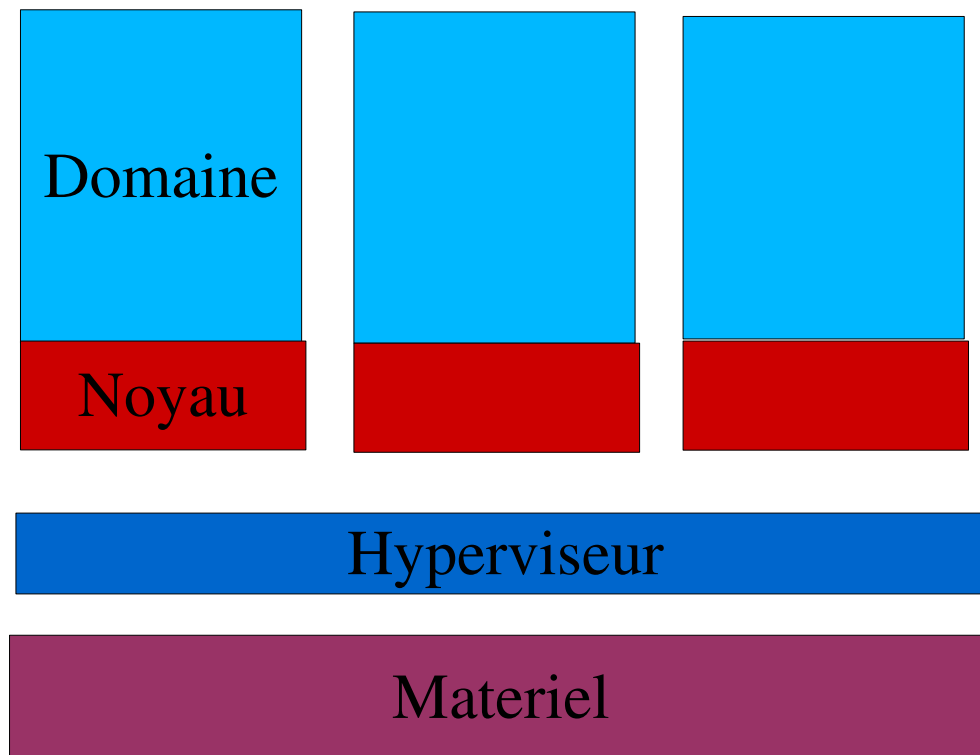
- Émulateurs: ex. Qemu, Hercules
 - simulation complète de la machine
- Virtualisation complète: ex. KVM, Xen, VirtualBox
 - Réutilise le processeur natif quand possible
 - OS client non modifié, requiers support matériel
- Para-virtualisation: ex. Xen-PV
 - OS client recompilé pour la machine virtuelle
- Conteneurs: ex. LXC, OpenVZ
 - Un seul OS, avec des primitives de cloisonnement
- Autres:
 - Ré-écriture au vol (VMWare)
 - Matériel virtualisé (mainframes)

Technique de virtualisation

- Problème
 - Le jeu d'instruction du i386 n'est pas virtualisable
- Émulation complète (QEmu)
 - Contrôle total, mais lent
- Réécriture du code au vol (VMWare)
 - Plus rapide mais très complexe
- Paravirtualisation
 - Recompilation pour la plateforme virtuelle, idéal
- Virtualisation matérielle
 - Nouveaux CPUs
 - Secondé par des driver paravirtuel

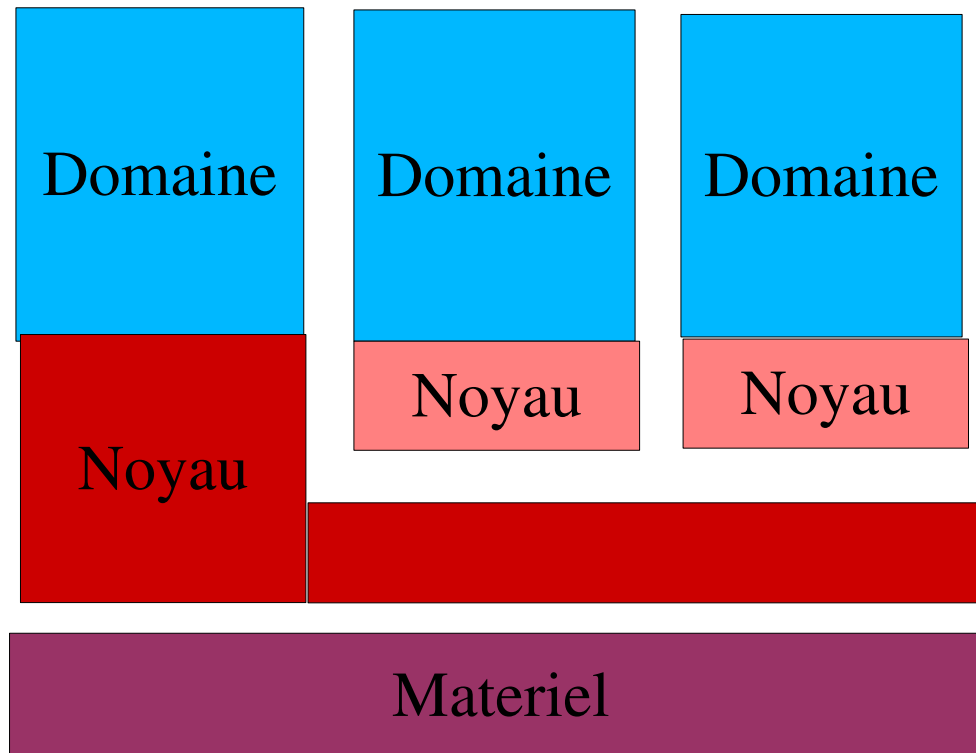
Virtualisation avec hyperviseur

- Hyperviseur isole les noyaux du matériel, ex. Xen



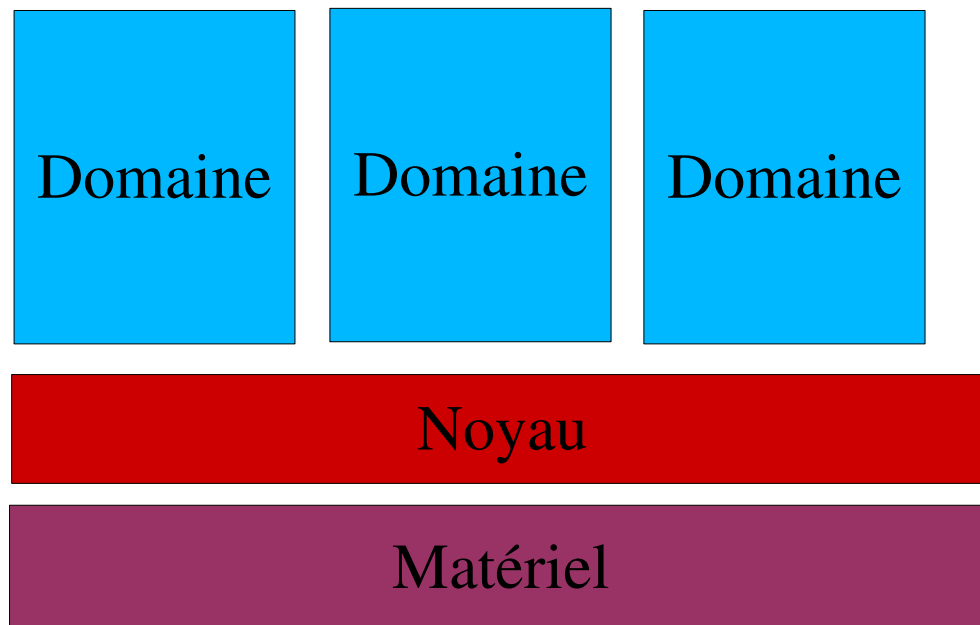
Virtualisation sans hyperviseur dédié

- Le noyau principal est l'hyperviseur, ex. KVM, VirtualBox



Virtualisation avec un seul noyau

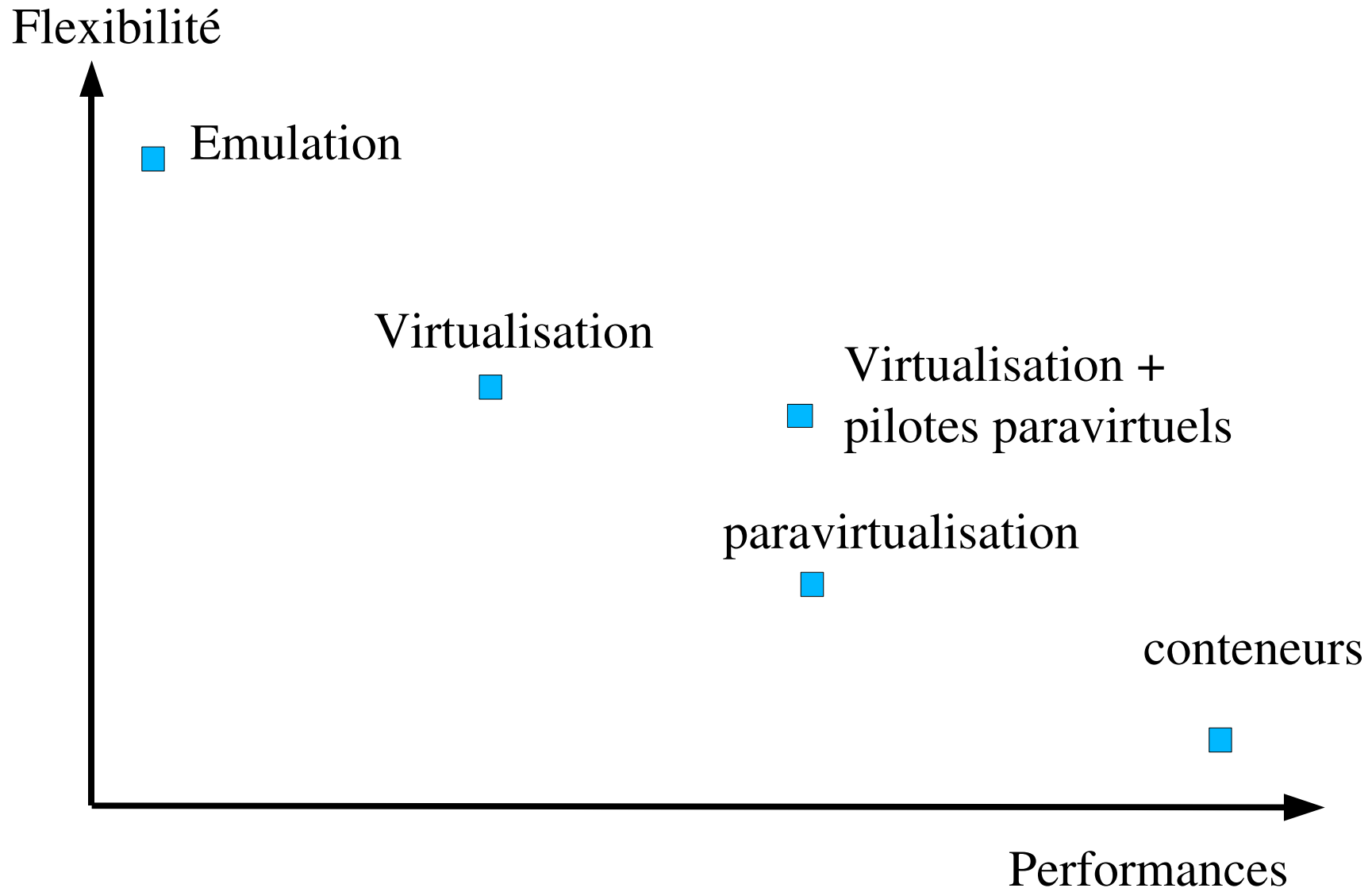
- Le noyau sert tous les domaines e.g. OpenVZ, LXC
- Partition des périphériques
- Allocations mémoire/CPU unifiées



Facteurs déterminants

- Capacités d'émulation
- Performances
- Sécurité, isolation
- Occupation mémoire, disque
- Fonctionnalités:
 - Tuning
 - drivers
 - Snapshots
 - Migrations a froid/a chaud
- Outils d'administration
- Coûts

Compromis flexibilité/performances



QEmu

- Emulateur, CPU et Matériel
- Nombreuse plateformes émulées
- Auteur principal Fabrice Bellard
- Réutilisé par d'autres outils (Xen, KVM)

Xen

- Projet de Recherche, Cambridge UK
- Paravirtualisation de Windows et Linux
- Licence GPL
- Présentation a OLS 2004
- Performances intéressantes
- Virtualisation complète (Qemu + support processeur)
- Intégration partielle dans Linux 2.6.23
- Rachat par Citrix
- Coopération avec Microsoft
- Distance par rapport a Linux

Xen: nombreuses fonctionnalités

- Cycle de vie classique de la machine virtuelle
- Sauvegarde/restauration de l'image mémoire
- Migration a froid et a chaud
- Redimensionnement mémoire et CPUs a chaud
- Pilotes de périphériques complets via Linux
- Bon support i386/x86_64 , ia64
- Robuste, beaucoup de tests
- Isolation correcte mais pas absolue
- Bonne performances en paravirtualisation
- Support avancé de la virtualisation matérielle

KVM: l'autre solution

- Intégré au sources de linux depuis 2.6.20
- Virtualisation complète
- Nécessite support matériel (Intel VT or AMD-V)
- Utilise QEMU pour la partie émulation
- Qumranet racheté par Red Hat en 2008
- Développé avec de nombreux partenaires
- Fonctionnalités augmentent:
 - Support processeurs S390, ia64, ppc64
 - Migration
 - Drivers paravirtuels linux/Windows
 - Port FreeBSD
- A peu près a parité avec Xen désormais

UML User Mode Linux

- Inclus dans les noyau 2.6.x
- linux est un programme
- Faible performances
- Très simple a mettre en oeuvre

VServer

- Containers pour Linux depuis très longtemps (patch)
- Partage du système
- Interception des appels systèmes
- Pas d'émulation, très rapide
- Ordonnanceur unique
- Un seul noyau
- Modèle réseau très simple

OpenVZ

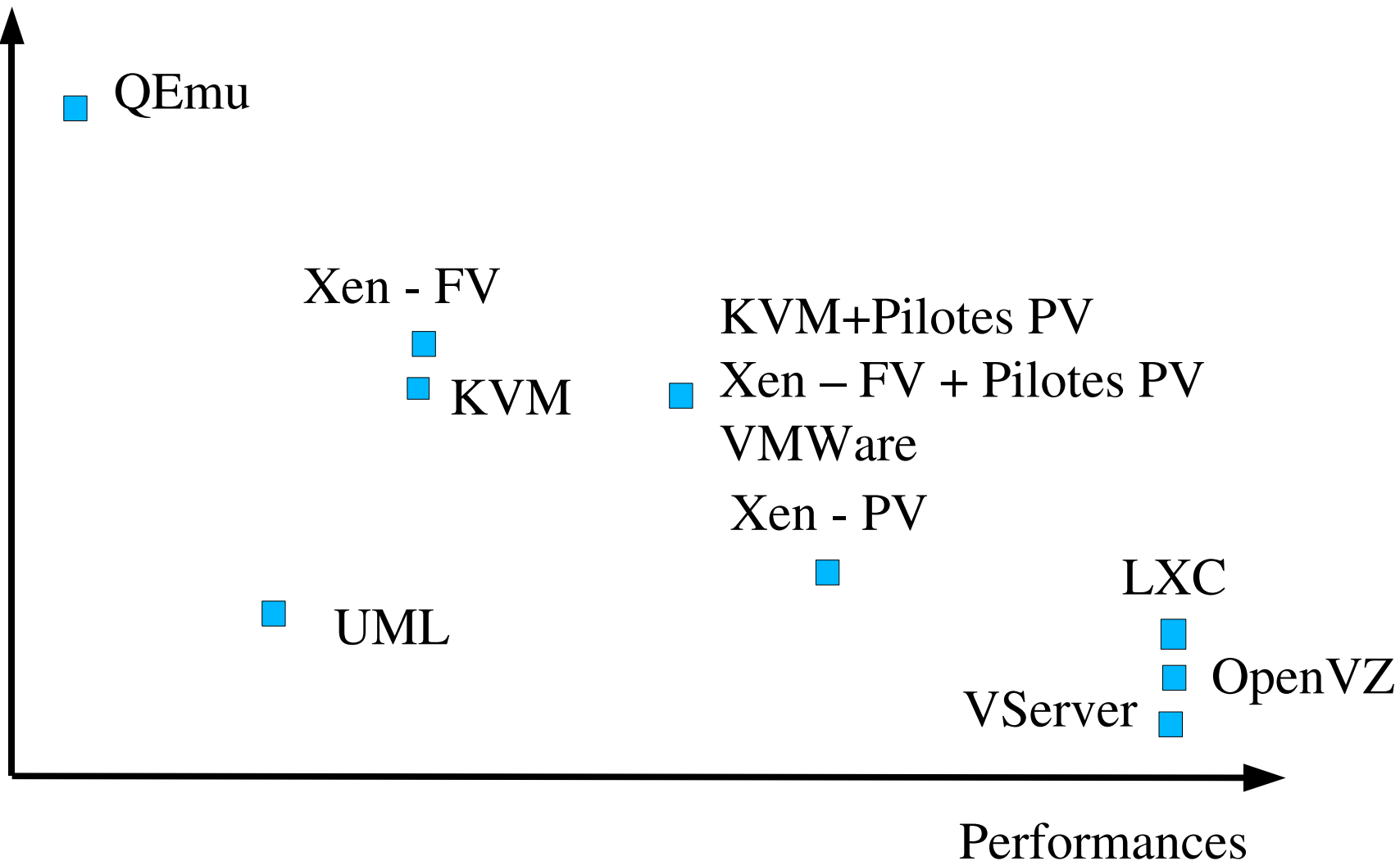
- Patch pour noyaux récents, modèle a container
- Implémentation plus complexe que Vserver
- Ordonnanceur a deux niveau
- Virtualisation du réseau
- Tuning très avancé, I/O, réseau et CPU
- Checkpointing, migration a chaud
- Base de Virtuozzo solution propriétaire

LXC: Linux Containers

- Mécanismes de groupe dans noyaux Linux récent
- Groupes de processus
- Contrôle de l'ordonnancement
- Contrôle de l'accès aux ressources
- Chroot amélioré

Compromis flexibilité/performances

Flexibilité

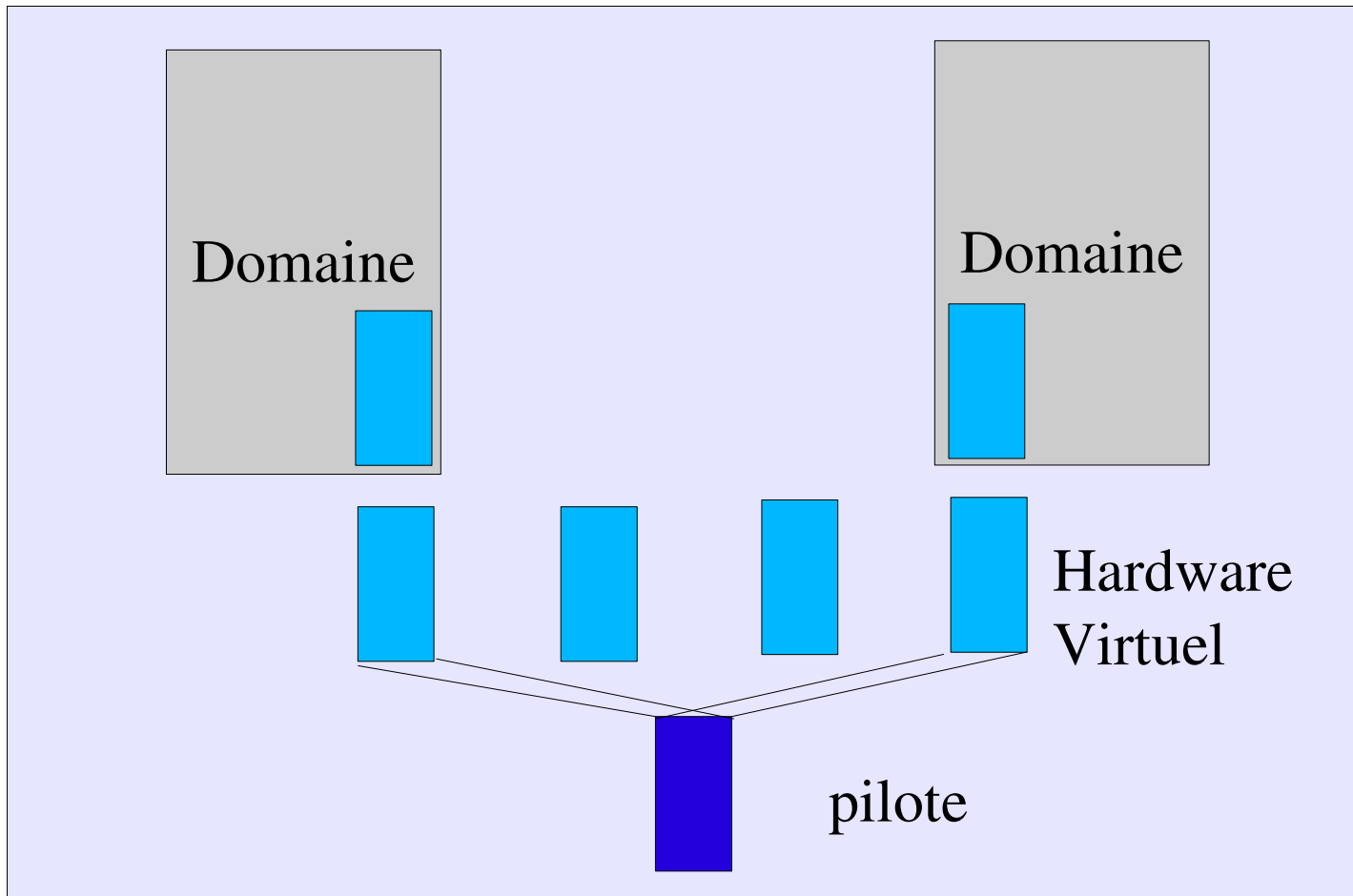


Evolution: integration, uniformisation

- Au niveau du matériel
 - Support pour la virtualisation, partitionnement des CPU
 - Partitionnement, QoS pour les entrées/sorties
- Au niveau du noyau linux
 - Primitives de virtualisation pv-ops
 - Intégration de Xen ... ?
 - Amélioration de KVM
- Au niveau applicatif
 - API d'isolation libvirt
 - Problème de l'affichage, accès GPU
 - Outils systèmes
- 'Appliances' Système + Application prêtes a l'emploi

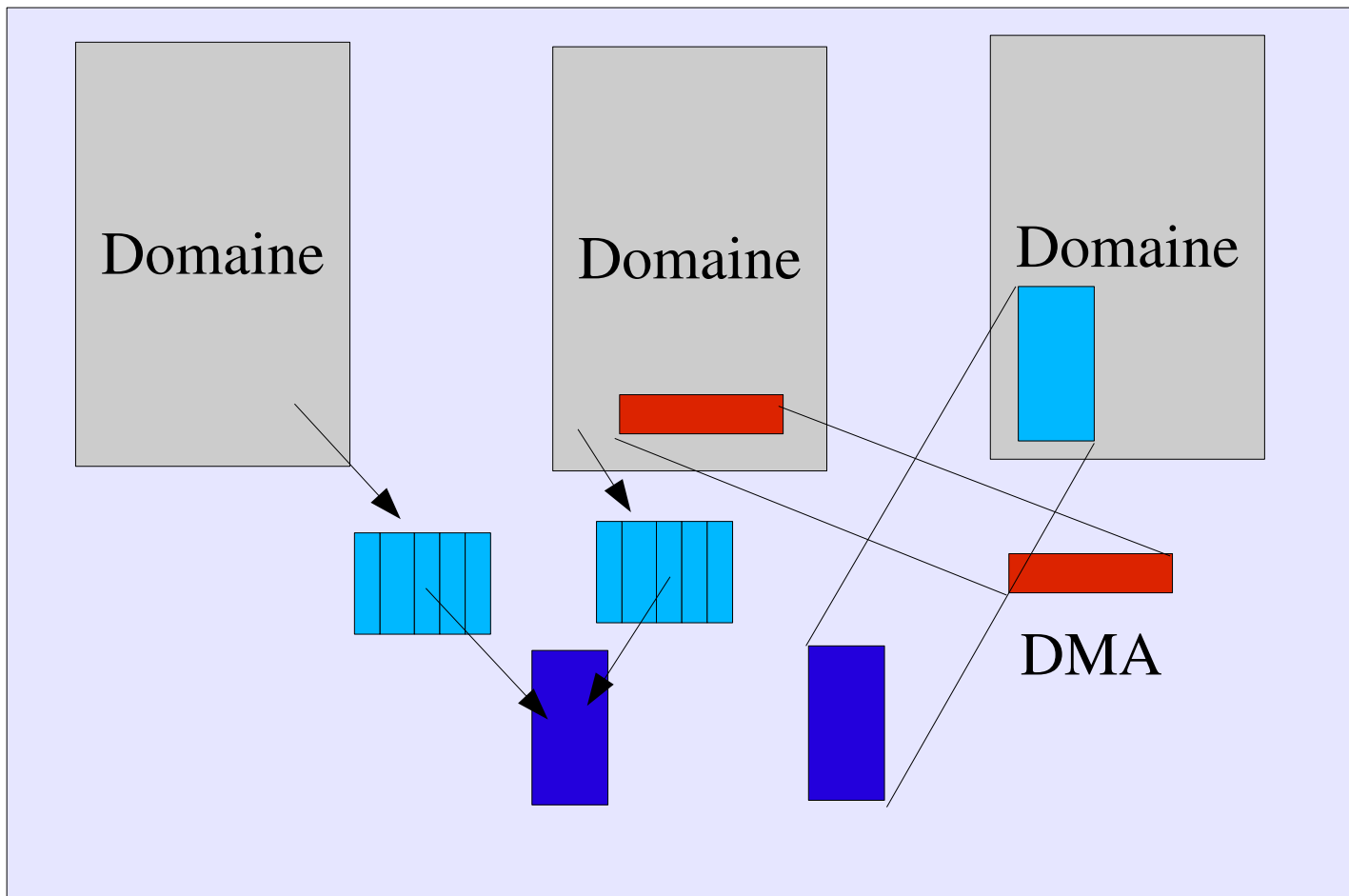
Virtualisation du materiel

- SR-IOV , virtualisation PCI, périphériques virtuels



Confinement et QoS

- Contrôle des accès via le matériel VT-D
- Multiple queues pour un périphérique



PV-Ops

- Linus 'mettez-vous d'accord !'
- API dans le Noyau pour la virtualisation
 - Virtualisation de l'espace mémoire, CPU et I/O
 - Réutilisable par Xen, VMWare, KVM ...
 - Plus de patch noyaux
 - Interface stable
- pv-ops clients dans le noyau standard
- Pv-ops serveur manquants
- Noyau unifié pour Domaine 0 et les noyaux clients
 - => Ca bloque un peu !

Libvirt



- API stable pour la virtualisation
 - LGPL <http://libvirt.org/>
 - API Generique (Xen, Qemu, KVM, OpenVZ, ...)
 - Relativement portable Solaris, Windows, OS-X...
- Operations de base
 - Créer/détruire/suspendre/sauver/migrer
 - Description XML des machines virtuelles
 - Statistiques, support NUMA
 - Gestion des périphériques et pools
 - Intégration de SELinux
- Ecrit en C
- Support des acces distants sécurisés
- Bindings python, perl, Java, Ocaml, Ruby

Libvirt et compagnie

- Virsh: shell de contrôle de la virtualisation
- Virt-manager: interface graphique de management
 - Cycle de vie complet
 - Création, migration
 - Console locale et distantes
- Libvirt-CIM
 - implémentation CIM par IBM (LGPL)
 - Pegasus et SFCB
 - Xen, KVM et LXC
- Cobbler:
 - Environnement d'installations automatiques
 - Koan agent de réinstallation

Conclusions

- De nombreuses solutions, couvrant tous les types
- Des besoins très divers
- Intégration s'accélère
- Le matériel se transforme, mais reste standard
- Changement des habitudes de travail

Plus d'obstacle techniques au déploiement



Questions?

veillard@redhat.com | www.redhat.com