

Virtualisation

- Pascal PETIT
- Pascal.petit@univ-evry.fr
- Page prof.: <http://www.ibisc.fr/~petit>

Virtualisation: définitions

- Ensemble techniques logicielles et matérielles permettant de fournir un ensemble de ressources informatiques utilisable indépendamment de la plate forme matériel
- Domaines concernés :
 - Ressources CPU et mémoire
 - Stockage
- Surcouche permettant de s'abstraire des contraintes matérielles
- Exemple: les gestionnaires de volumes

Problème des datacenters d'hier

- Serveurs sous-utilisés (en moyenne 5 à 15%)
- Sécurité: 1 service = 1 serveur
 - Implique une multiplication des serveurs physiques
- Coût d'achat
- Coût d'exploitation
 - Énergie: alimentation (#80€/an/serveur), climatisation (> alimentation)
 - Gestion, place occupée,...
- Facilité de gestion
- Tolérance de panne

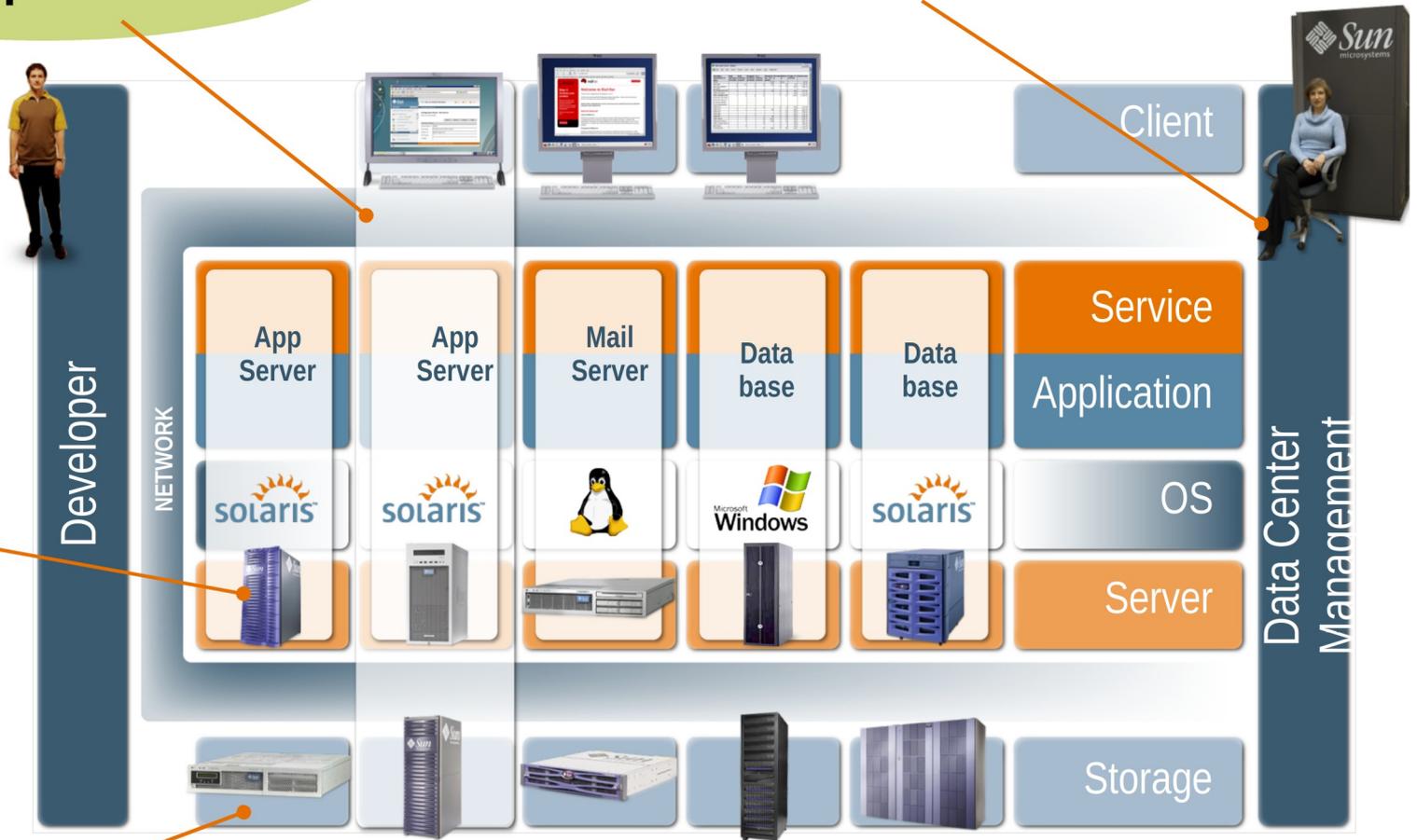
Le Data Center d'hier

1 server = 1 application

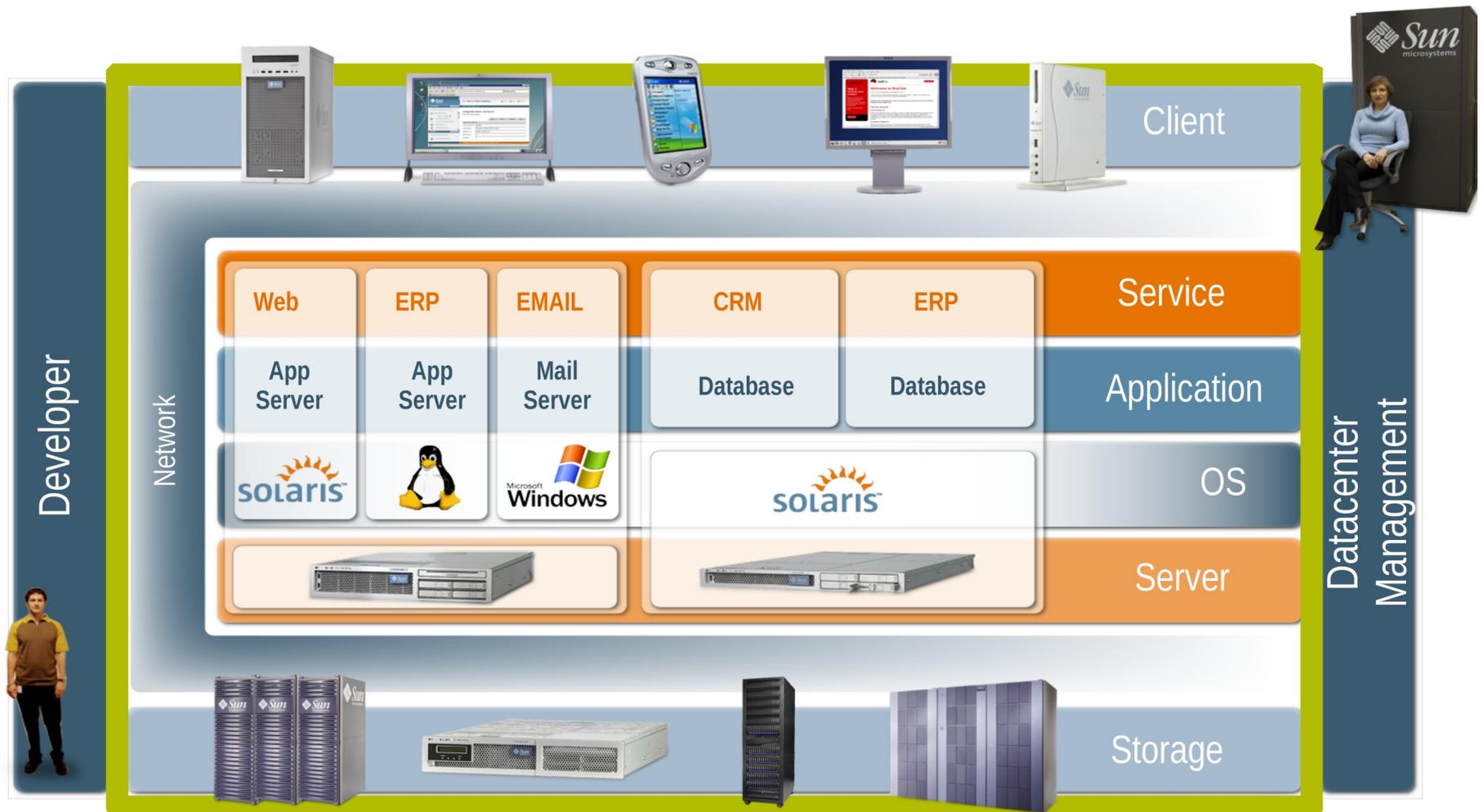
L'ensemble des serveurs est coûteux et difficile à gérer

Charge moyenne par serveur: 5 à 15%

Stockage parcellaire non mutualisé



Le DATA Center aujourd'hui



Source schéma: Sun microsystems

Apport de la virtualisation

- Coût
 - Acquisition
 - Exploitation
 - Mais : attention au coût des licences des OS, ...
- Facilité de gestion
 - Sauvegarde des machines virtuelles
 - Déplacement de machines virtuelles
- Sécurité
 - Isolation des machines virtuelles

Historique de la virtualisation

- 1965 IBM M44/44X **paging system**
- 1965 IBM System/360-67 **virtual memory hardware**
- 1967 IBM CP-40 (January) and CP-67 (April) **time-sharing**
- 1972 IBM VM/370 **run VM under VM**
- 1997 Connectix **First version of Virtual PC**
- 1998 VMWare U.S. Patent 6,397,242
- 1999 VMware Virtual Platform for the Intel IA-32 architecture
- 2000 IBM z/VM
- 2001 Connectix **Virtual PC for Windows**
- 2003 Microsoft acquired Connectix
- 2003 EMC acquired VMware
- 2003 VERITAS acquired Eascent
- ◆ 2005 Intel VT
- ◆ 2006 AMD VT
- ◆ 2005 XEN
- ◆ 2006 VMWare Server
- ◆ 2006 Virtual PC 2006
- ◆ 2006 HP IVM Version 2.0
- ◆ 2006 Virtual Iron 3.1
- ◆ 2007 InnoTek VirtualBox
- ◆ 2007 KVM in Linux Kernel
- ◆ 2007 XEN in Linux Kernel
- ◆ 2008 microsoft Hyper V

I - Les différentes techniques

- ▶ Redondance
- ▶ Virtualisation du stockage
- ▶ Réimplémentation de bibliothèques
- ▶ Isolateur
- ▶ Noyau en espace utilisateur
- ▶ Superviseur en mode natif
- ▶ Superviseur en mode émulé
- ▶ Hyperviseur (para-virtualisation)

Redondance et tolérance de panne :

exemple des routeurs

- Fait en live au tableau
- HSRP/VRRP/CARP
- Mac et IP virtuelles
- Protocole Hello
- Problème des états du routeurs HS (NAT, FW)

Virtualisation du stockage

- Gestionnaire de volumes logiques
 - Ajoute une couche qui permet de ne plus voir les périphériques physique
 - Volume logique:
 - Composer de plusieurs volumes physiques provenant de sources (disques) variés
 - Extensibles, support du RAID, ...

Virtualisation du stockage

- NAS/SAN

- NAS:

- Connecté au réseau
 - Propose un accès via des protocoles de partages réseau (CIFS, NFS, AFS, ...)
 - Accès en mode fichiers

- SAN

- Les baies apparaissent comme un disque local
 - Accès en mode bloc

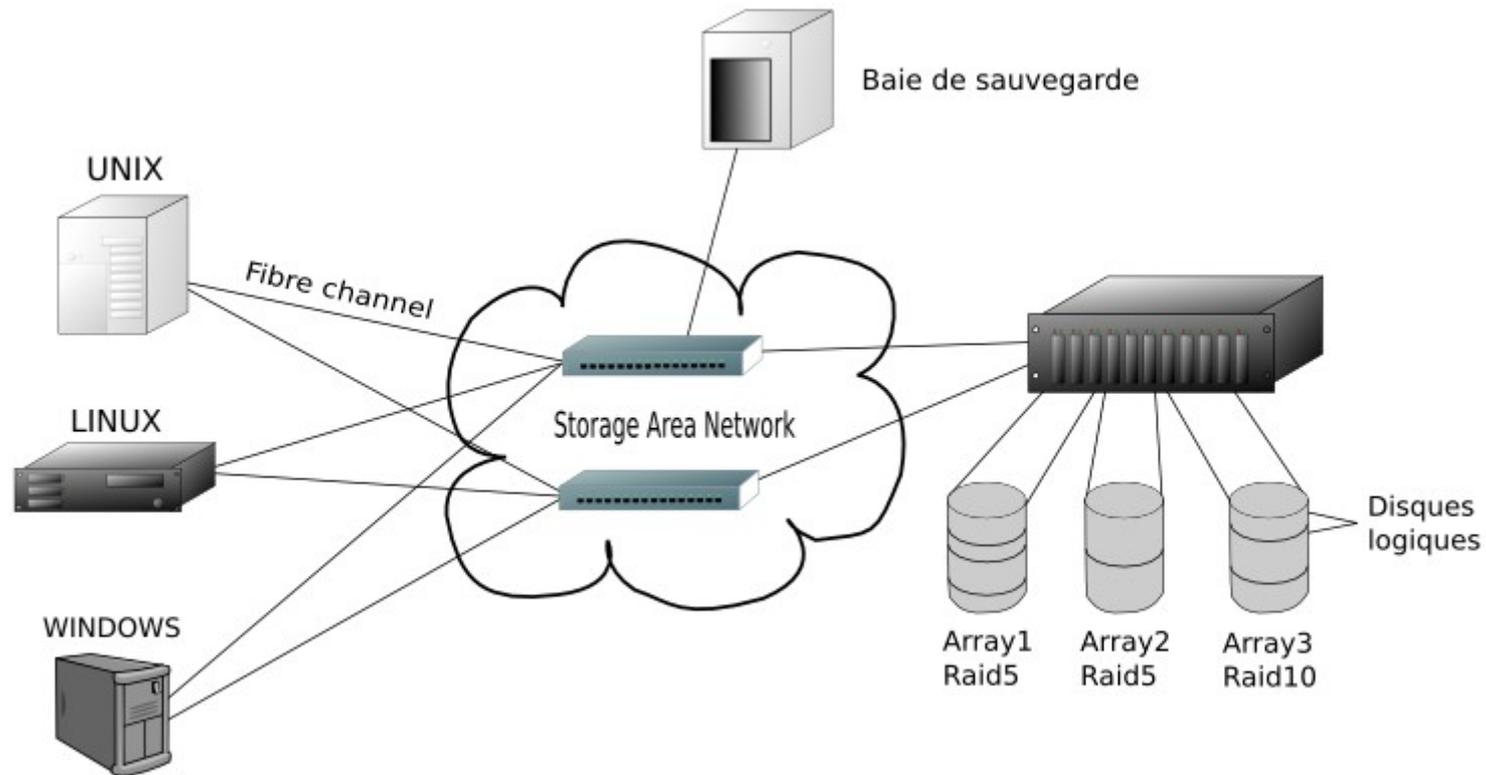
Virtualisation du stockage: SAN

- SAN:
 - Gestion globale et mutualisée de l'espace disque pour l'ensemble des serveurs: souplesse
 - Ajout facile d'espace à un serveur donné
 - Ajout facile d'espace sur le SAN (ajout de disques, de baies, ...)
 - Sauvegarde, tolérance de panne facilitées (raid, redondance de la liaison avec le SAN)
 - Réplication vers une autre baie (éventuellement distante), historisation, snapshot, ...

Virtualisation du stockage: SAN

- Protocoles des SAN : Lien serveurs-SAN
 - Fiber channel : rapide mais cher
 - Iscsi: scsi sur TCP/IP, supporté par les OS tant en client qu'en serveur
 - RFC 3720 - Internet Small Computer Systems Interface (iSCSI), Avril 2004.
 - RFC 3721 - Internet Small Computer Systems Interface (iSCSI) Naming and Discovery, Avril 2004.
 - RFC 3722 - String Profile for Internet Small Computer Systems Interface (iSCSI) Names, Avril 2004.
 - AoE: ATA sur ethernet (peu utilisé)

Virtualisation du stockage



Virtualisation du stockage: choix de la taille des disques

- Paramètres :
 - Espace disque
 - Débit du disques
 - Nombre d'I/O par seconde
- Dans la vraie vie :
 - SATA:
 - De l'espace pour pas cher
 - Vitesse de rotation modeste
 - Temps d'accès élevé (8ms)
 - Nb d'I/O par seconded modeste

Virtualisation du stockage: choix de la taille des disques

- SAS :

- Coût au Go plus élevé que SATA
- Vitesse de rotation élevée (15Krpm)
- Temps d'accès faible (4 ms)
- Nb d'I/O par seconde plus élevé que SATA
 - File d'attente de taille correcte
 - Temps d'accès faible
- Lien SAS bicanal

Virtualisation du stockage: choix de la taille des disques

- SSD:

- Coût au Go très élevé
- Temps d'accès très très faible => nombre d'I/O par secondes très élevé (8000 à 3000000 iops (2019))
- Durée de vie ?

- Choix de la taille des disques

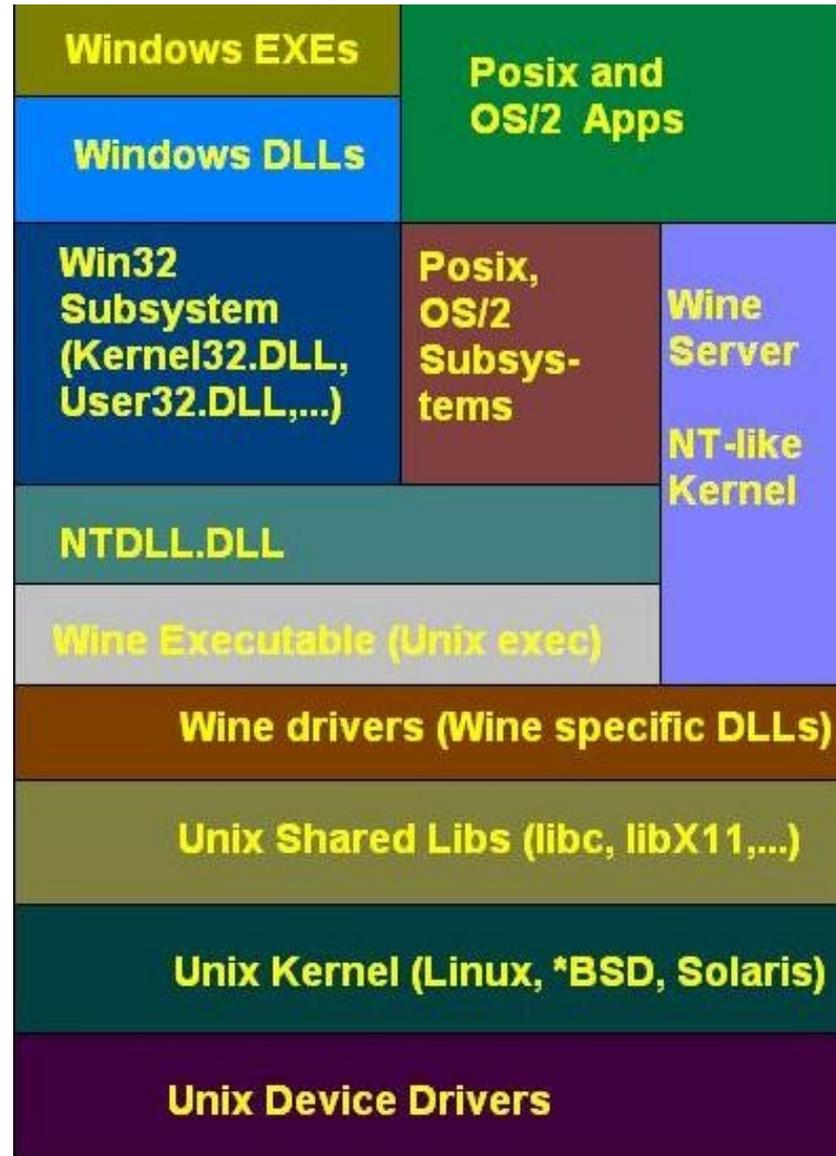
- Déterminer le facteur bloquant (I/O/s, capacité,)
- Prendre des disques de taille modeste pour augmenter le nombre d'O/I/s de l'ensemble
- Raid 0, raid5,6, 10, 01, ... permettent de combiner des disques pour « additionner » les débits et/ou les nombre d'I/O par seconde des disques

Réimplémentation de bibliothèques

- ▶ Réécriture des bibliothèques d'un OS pour faire fonctionner des programmes sous un autre OS
- ▶ Exemple: Wine (utiliser des applications windows sous unix)

Architecture de Wine

- Suit l'architecture de NT
 - Implèmente toutes les “core” DLLs (ntdll, user32, kernel32)
- Le “Wine server” fournit l'infrastructure NT :
 - Transmission de messages
 - Synchronisation
 - Handles des objects



Isolateur

- ▶ Cloisonnement au sein d'un OS
- ▶ Une seule instance du noyau de l'OS commune
- ▶ Exemples (de technologies différentes) :
 - ▶ Chroot: changement de racine (faible)
 - ▶ BSD Jail: isolation en espace utilisateur
 - ▶ Conteneur linux, doker, Solaris Zones: partitionnement au niveau du noyau de l'OS
- ▶ Très performant mais le cloisonnement est-il suffisamment solide ?

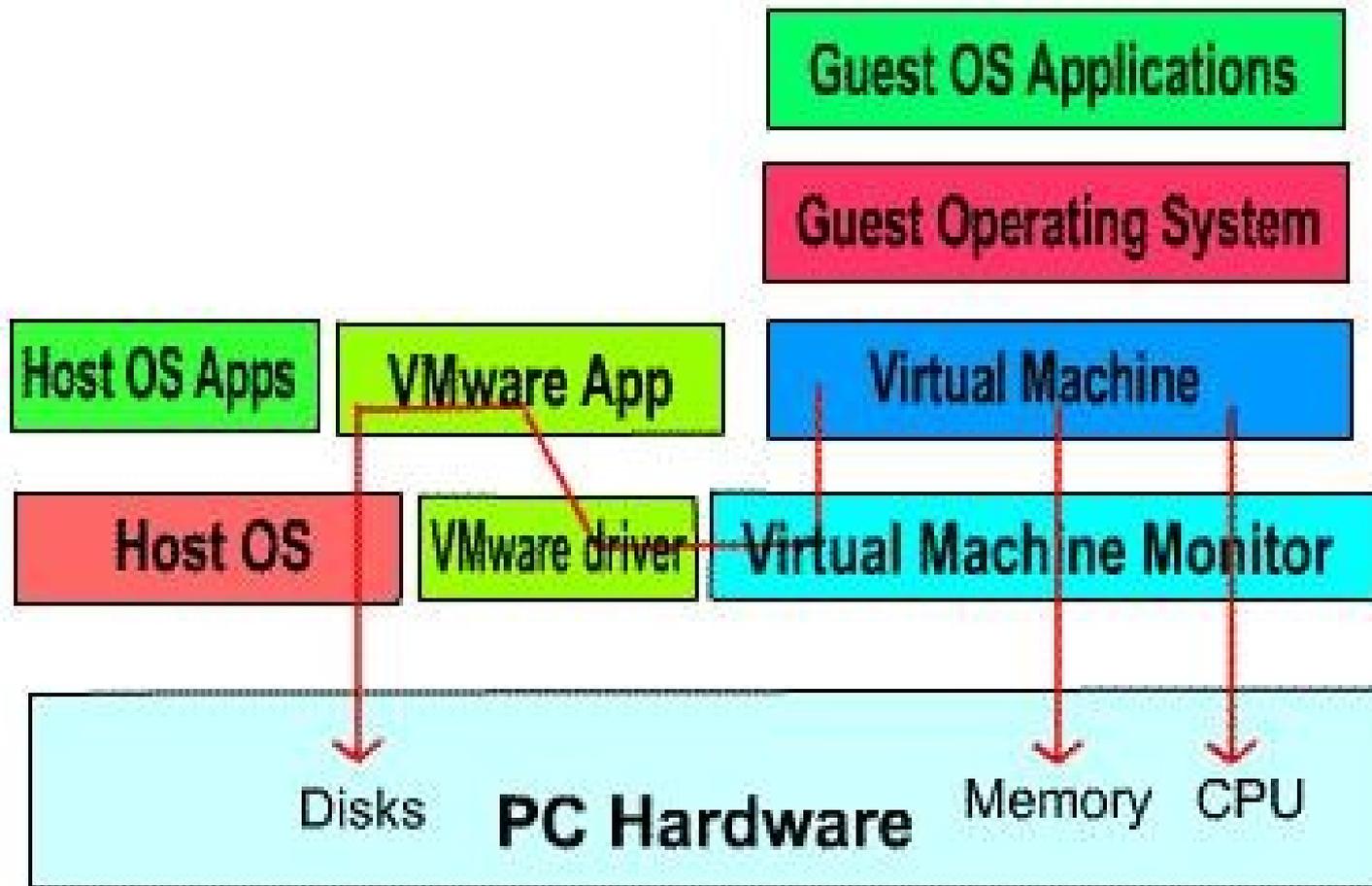
Superviseur en mode natif

- Appelé aussi “**machine virtuelle**, en mode natif”. Logiciel complexe qui permet l'exécution de plusieurs OS invités de même architecture processeur que la machine hôte en offrant un accès générique (émulé) aux ressources physiques.
- Exécution des instructions en mode natif (sauf exceptions difficiles à gérer)
- Bonnes performances

Superviseur en mode natif

- ▶ Exemples:
 - ▶ Vmware Player, workstation, fusion
 - ▶ Virtuabox
 - ▶ kvm

Vmware workstation



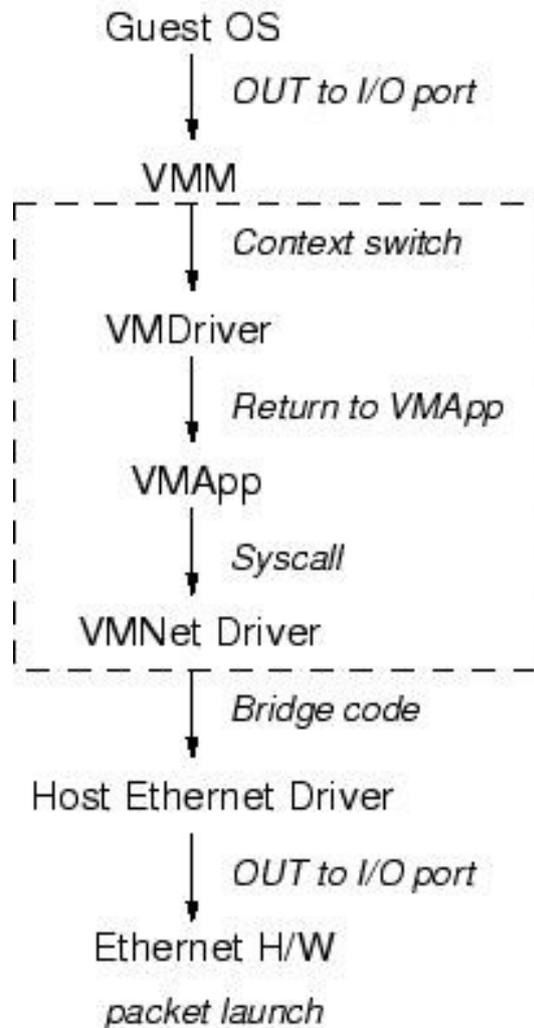
VMware Workstation Architecture

Vmware workstation

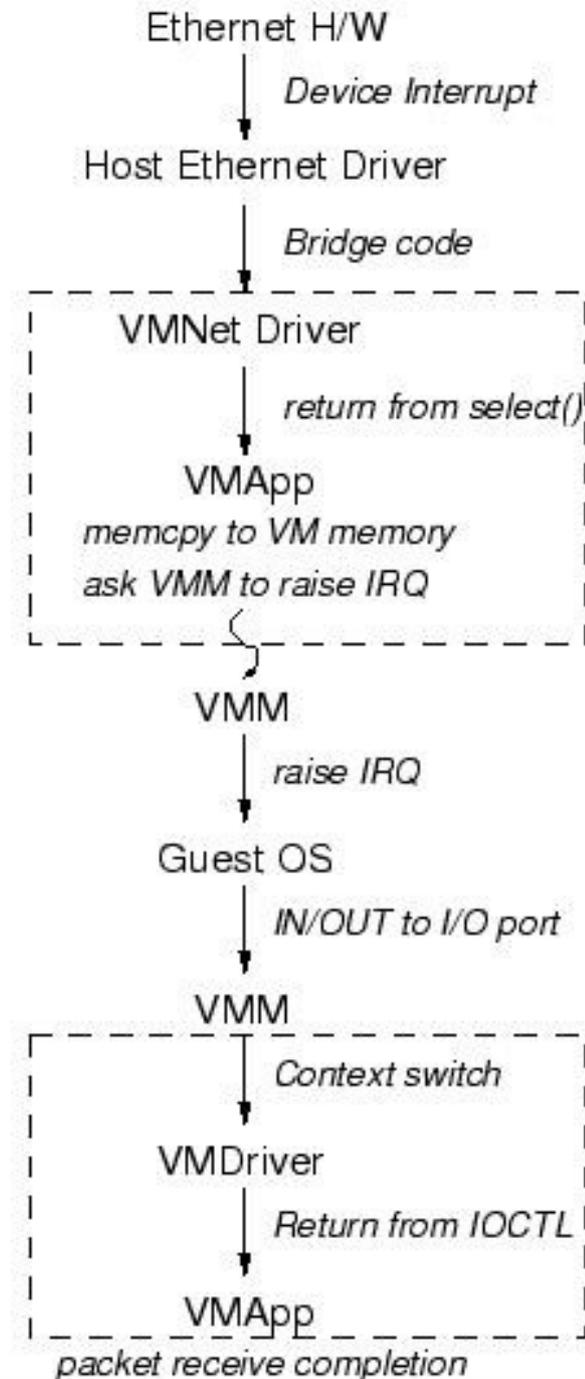
- ▶ VMM does not have access to I/O
- ▶ I/O in “host world”
 - ▶ Low level I/O instructions (issued by guest OS) are merged to high-level I/O system calls
 - ▶ VM Application executes I/O SysCalls
- ▶ VM Driver works as the communication link between VMM and VM Application
- ▶ World switch needs to “save” and “restore” machine state
- ▶ Additional techniques to increase efficiency

Vmware workstation

Network Packet Send



Network Packet Receive



Superviseur en mode émulé

- ▶ Appelé aussi “**machine virtuelle**, en mode émulé” ou encore “**émulateur**”

A la différence de la machine virtuelle en mode natif, le processeur est aussi émulé rendant ainsi possible l'exécution d'OS pour des plateformes différentes de celle de l'hôte

- ▶ Pb: performances modestes

- ▶ Exemples:

- ▶ QEMU, Plex86, Bochs

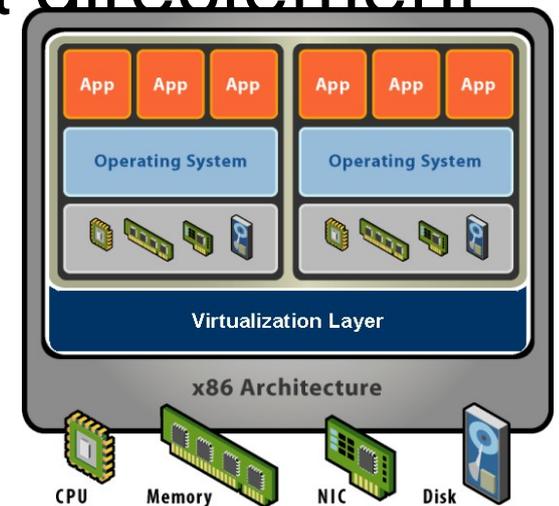
- ▶ Kego-fusion (émulateur console SEGA)

Hyperviseur (para-virtualisation)

- C'est un noyau dédié à la gestion de machines virtuelles. C'est une optimisation de la technique “superviseur”: au lieu d'avoir un OS et un logiciel de virtualisation, on a directement un noyau dédié à la virtualisation.

- Exemples:

- VMware ESX server
- XEN



Hardware Level

Virtualisation matérielle

- ▶ Support de la virtualisation directement dans le processeur. Pour les x86, il s'agit surtout de rendre possible l'exécution des instructions privilégiées du mode protégé par les guests directement afin d'optimiser les performances ou de ne pas avoir à modifier l'OS invité.
- ▶ Exemples:
 - ▶ Mainframes VM/CMS
 - ▶ Intel VT , AMD-V