

Evolution des techniques de virtualisation des systèmes informatiques

Avertissement

Marques et logos

IBM, le logo IBM, AIX, AS/400, Active Memory, BladeCenter, DB2, ESCON, i5/OS, OS/400, PowerPC, pSeries, WebSphere, xSeries, z/OS, zSeries, DS8000, iSeries, Micro-Partitioning, POWER, PowerVM, PowerVM (logo), POWER Hypervisor, POWER7, System Storage, System z, sont des marques déposées de la compagnie International Business Machines Corporation aux États-Unis et pour les autres pays.

Les noms :Power Architecture et Power.org sont des marques sous licence de Power.org.

UNIX est une marque déposée de l'Open Group aux États-Unis et pour les autres pays.

Linux est une marque déposée de Linus aux États-Unis et pour les autres pays.

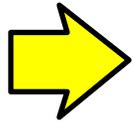
Microsoft, Windows sont des marques déposées de Microsoft Corporation aux États-Unis et pour les autres pays.

Intel, Itanium, Pentium et Xeon sont des marques déposées d'Intel Corporation ou de ses filiales aux États-Unis et pour autres pays.

Java et tous les logos sont des marques appartenant de Sun Microsystems, Inc.aux États-Unis et autres pays.

De nombreuses figures figurant dans ce document sont des extraits de documents publiés par IBM, Intel ou bien de pages de sites Internet. Leur réutilisation peut être soumise à l'autorisation des propriétaires des documents d'origine.

Définitions et types



Définitions et types de virtualisation

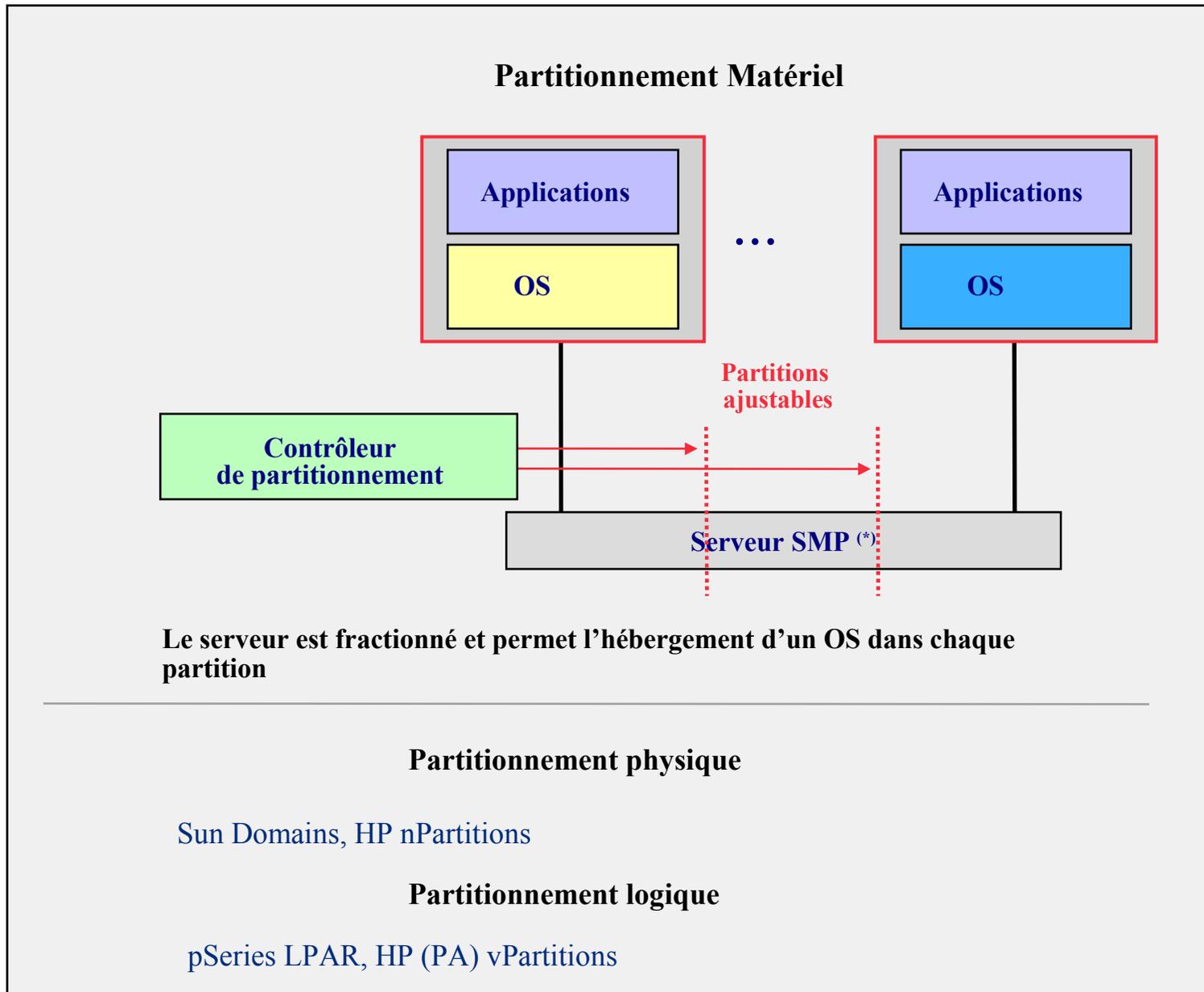
Les origines

Les besoins

Techniques mises en oeuvre

Situation actuelle

Partitionnement physique



(*) SMP : système multiprocesseur partageant une mémoire commune

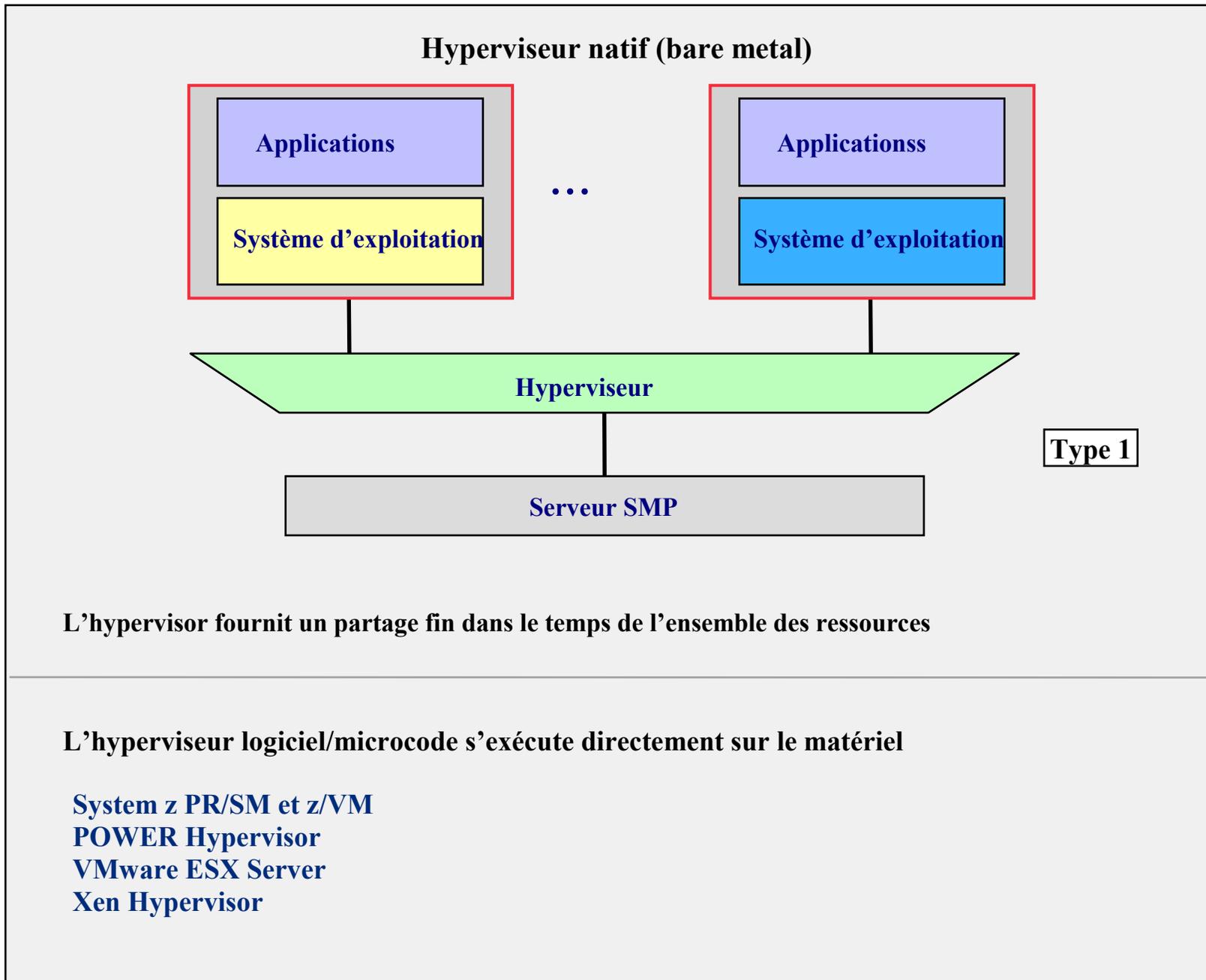
Systemes virtualisés : Type 1

Type 1

Popek and Goldberg (1974) :

- ❑ VMM (Virtual Machine Monitor) ↔ Hyperviseur
 - ❑ L'équivalence → fonctionne à l'identique en virtuel ou en réel
 - ❑ Le contrôle des ressources → le VMM assure le contrôle
 - ❑ L'efficacité → la majeure partie des instructions doivent ne pas être dépendante du VMM
 - ❑ Les instructions du processeur doivent être de 3 types :
 - ✓ Privilégiées (Mode hyperviseur)
 - ✓ Sensitives (Modification de configuration) sous ensemble des privilégiés
 - ✓ Non-Privilégiées (exécutées directement par l'OS virtualisé)
-
- Avantages :
 - Performance proche du système réel avec des ressources identiques
 - Inconvénient :
 - Systèmes virtuels compatibles avec l'architecture des processeurs réels

Systeme virtualisés : Type 1



Systeme virtualisés : Type 2

Type 2

Émulation tournant comme application d'un système d'exploitation

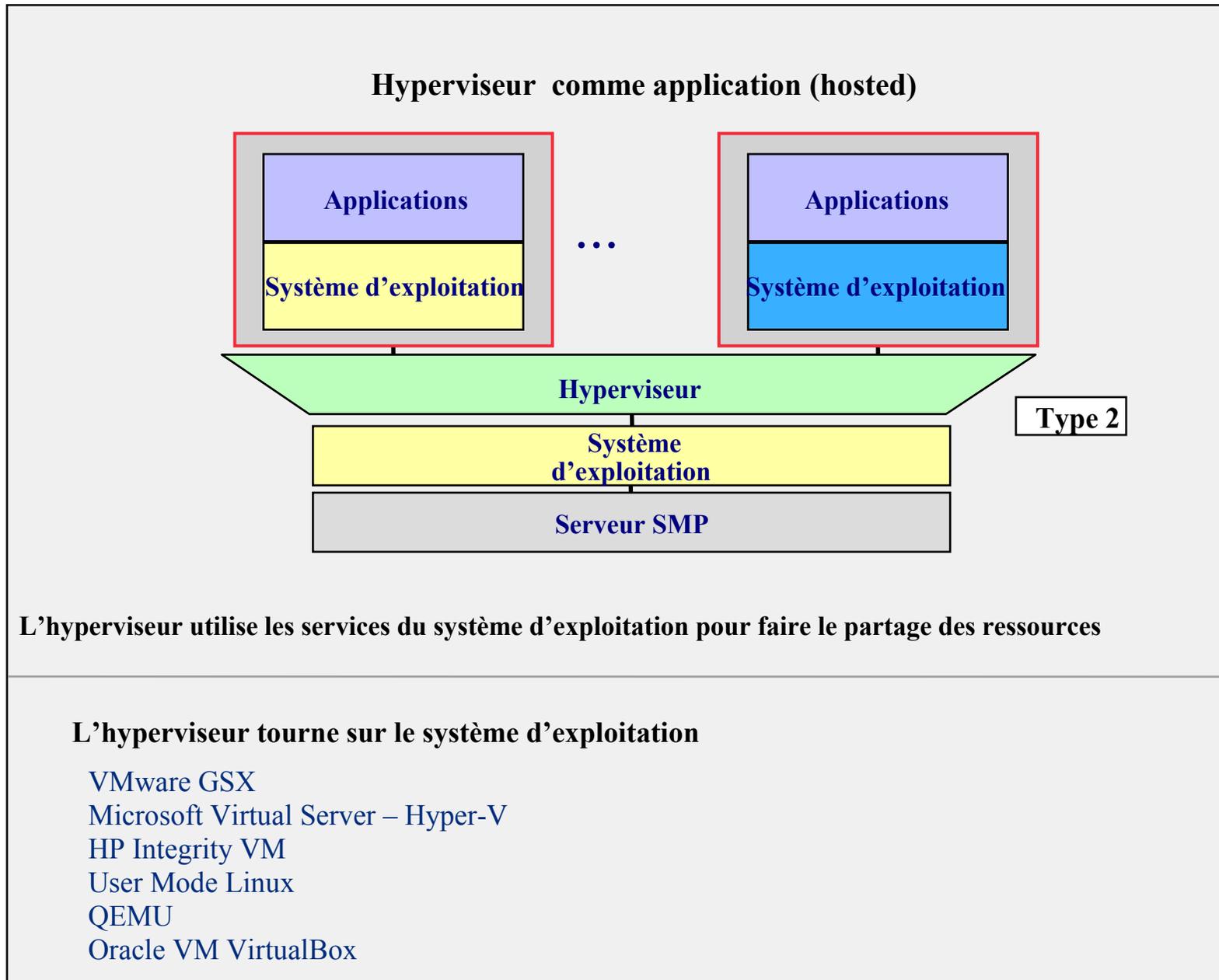
Avantages

- L'architecture matériel peut être différente de l'OS ou inexistante

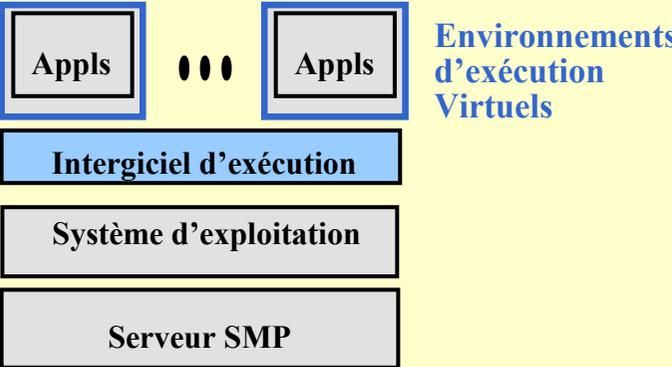
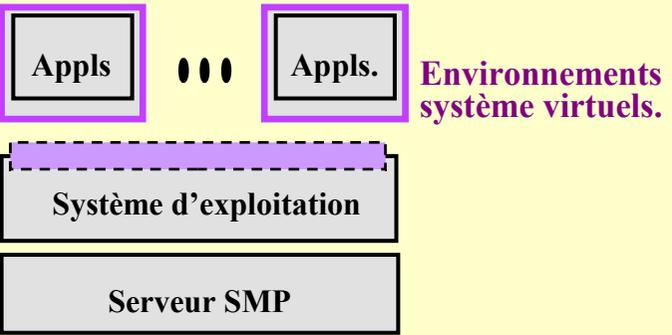
Inconvénient

- Performance moindre à très dégradée

Systeme virtualises : Type 2

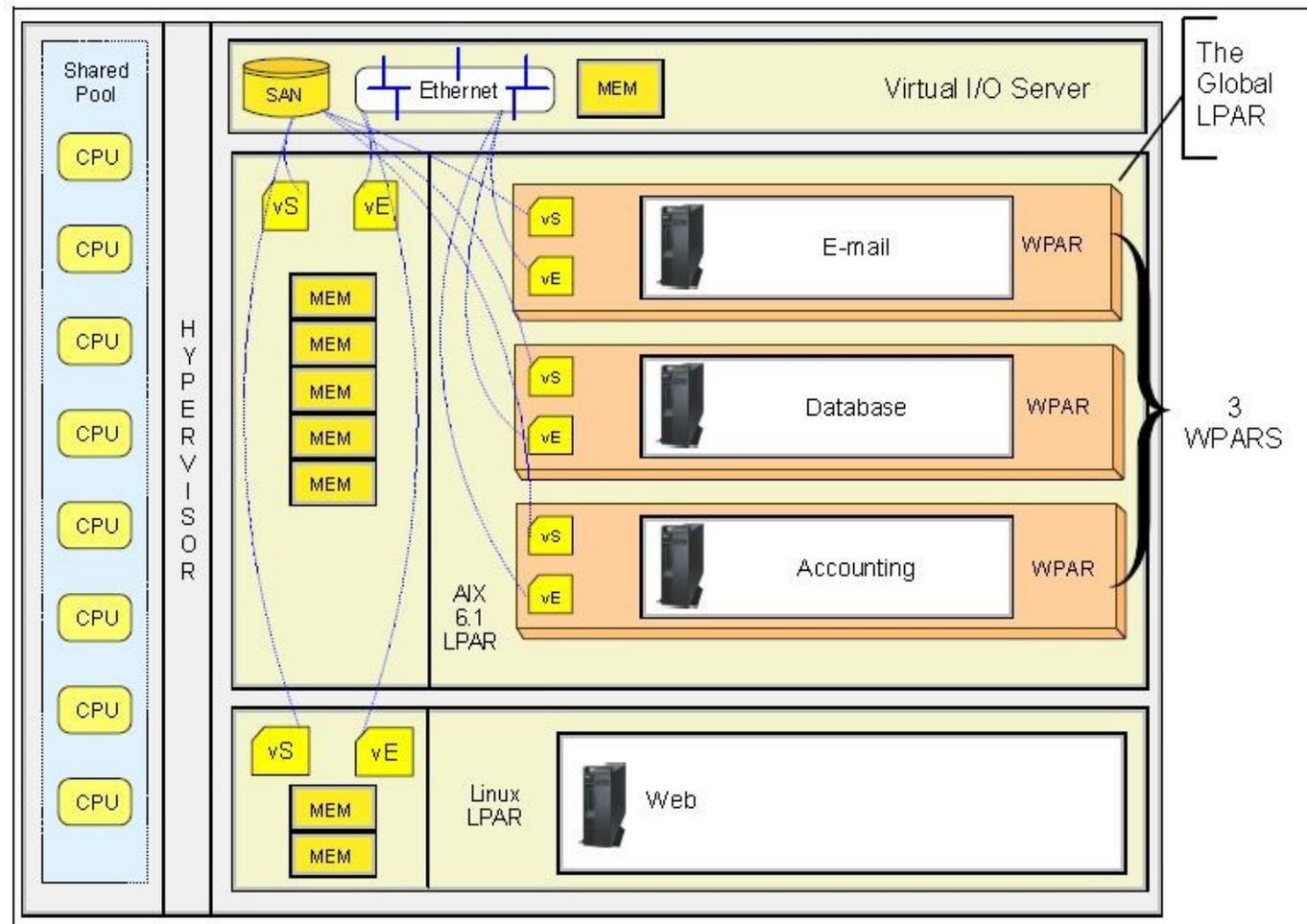


Virtualisation des traitements

	Méthode	Description	Exemples
Conteneurs d'applications	 <p>Environnements d'exécution Virtuels</p>	<ul style="list-style-type: none"> ▪ L'Intergiciel d'exécution (middleware Runtime) fournit des containers JVM, J2EE, ou CLR par application ▪ Les environnement d'exécution peuvent être indépendant du système 	<ul style="list-style-type: none"> ▪ WebSphere XD ▪ CORBA
	 <p>Environnements système virtuels.</p>	<ul style="list-style-type: none"> ▪ OS et middleware crée un environnement système virtuel par application ▪ Chaque container a son espace nommé, ses fichiers, attaches, ... 	<ul style="list-style-type: none"> ▪ z/OS Ad. Sp., i5/OS Sous-système. ▪ Solaris Containers ▪ AIX (WPAR) ▪ HP-UX Secure Resource Part. ▪ Windows & Linux solutions de Softricity, SWsoft, VERITAS, Trigence, ...

Ceci est plus de la gestion de traitement que de la virtualisation de systèmes

Virtualisation des traitements (WPAR)



AIX peut créer deux sortes de WPAR :

- ⇒ Des WPAR systèmes sorte d'instance indépendantes AIX
- ⇒ Des WPAR applicatives bien plus simples

Un peu d'histoire

Définitions et types de virtualisation



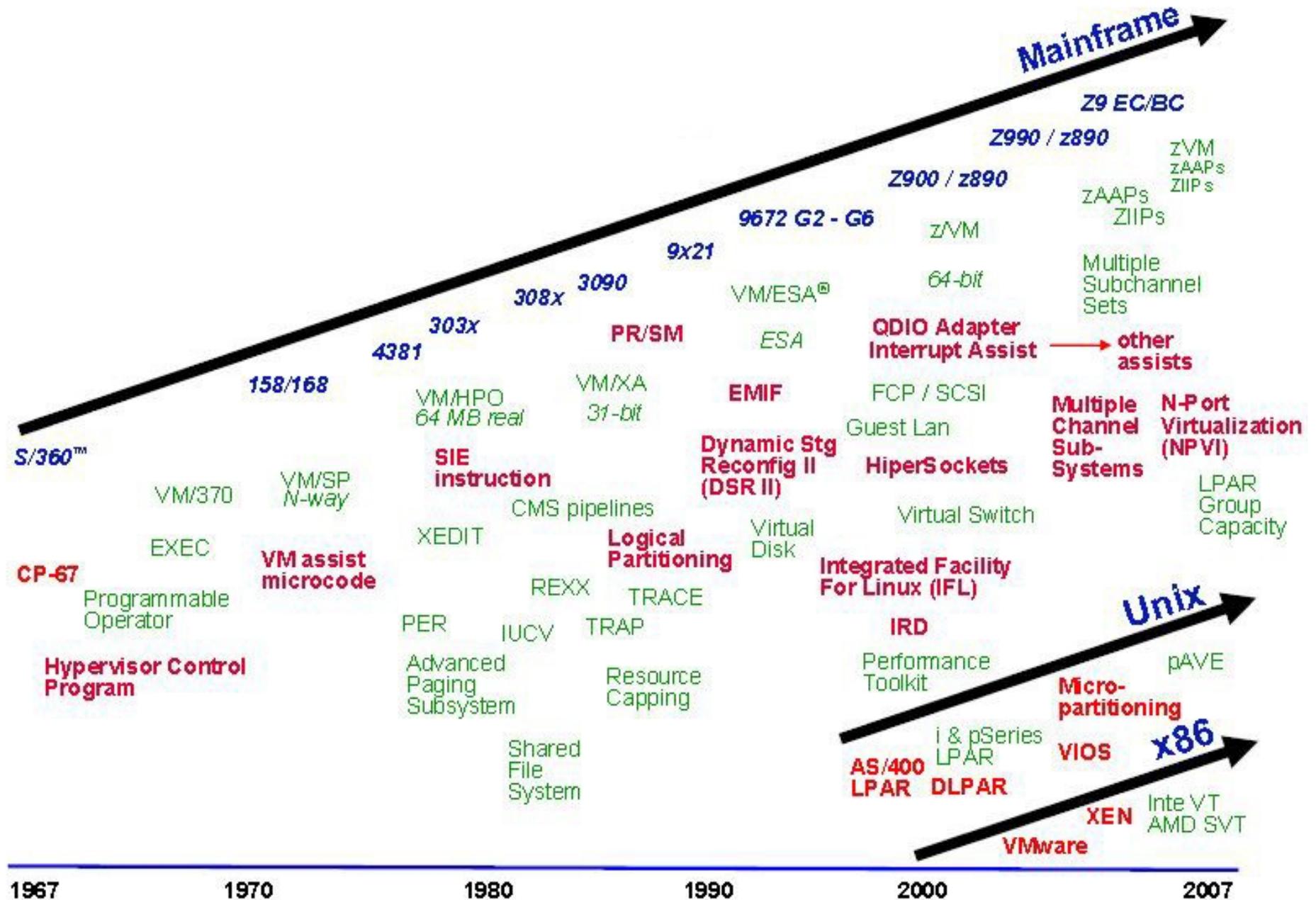
Les origines

Les besoins

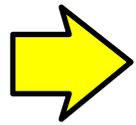
Techniques mises en oeuvre

Situation actuelle

Rétrospective



Les origines IBM



Les premières ébauches IBM : CTSS

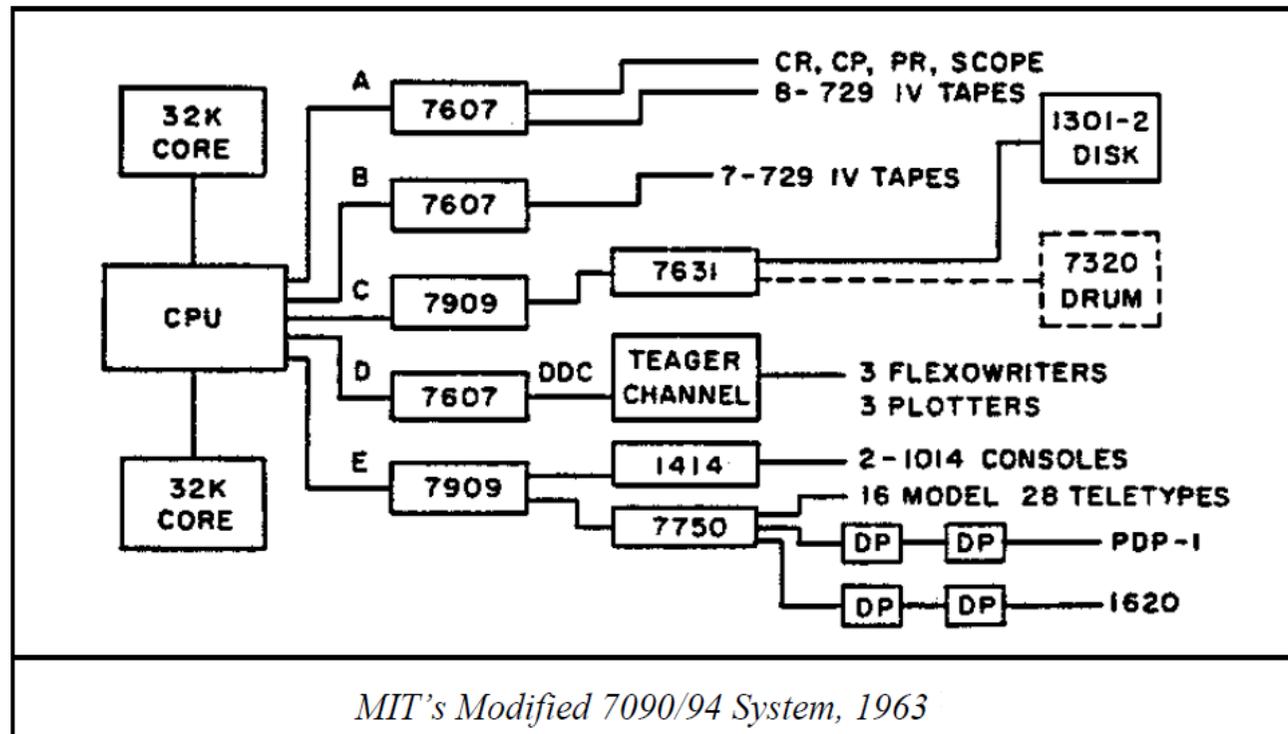
CP-40 et CP-67

VM/370

Les premières ébauches IBM

CTSS (Compatible Time-Sharing System)

- Développé par un petit groupe du M.I.T. dirigé par Fernando Corbato
- Premier papier sur un système “ temps partagé” en 1959
- Première version au MIT en novembre 1961, sur IBM 709, puis sur 7090 et 7094
- Apparaît rapidement la nécessité d’adapter le matériel (partage mémoire)



Superviseur

Gère les interruptions et les entrées/sorties

Mémoire

A-bank superviseur
B-bank utilisateurs
Réallocation
Protection

CMS

Fichiers définis par :

- Nom
- Type
- Emplacement

- Développement en cours de la famille des systèmes IBM 360
- Malgré des relations étroites entre l'équipe de développement du système IBM 360 et le MIT, les dirigeants d'IBM ne sont pas convaincus du besoin de systèmes en temps partagé ➡ système 360 annoncé sans traduction d'adresse de mémoire.
- Projet MAC financé par le DARPA (**D**efense **A**dvanced **R**esearch **P**rojects **A**gency) pour le développement d'un successeur de CTSS qui conduira à Multics, lequel tournera finalement sur un GE 645 en 1969 au MIT, soutenu par General Electric et les Laboratoires BELL.
- La synthèse par les Laboratoires Bell de CTSS et Multics conduira au système **Unix**.
- IBM réagit avec le système S/360 model 67 et le développement de TSS (the Time-Sharing System)
- En 1964 le MIT élabore avec le projet **CP-40** les conditions nécessaires pour virtualiser des systèmes S/360
 - Programme de contrôle du système : CP-40 Control Program
 - nombre de machines virtuelles fixe : 14 VM
 - mémoire virtuelle fixe : 256K
 - Partitionnement et contrôle de l'espace disque (minidisques et CCW translation)
 - Unités (unit record) en mode spoulé
 - Fonctions matériel communes disponibles
 - Nécessité de développer un système mono-utilisateur appelé CMS (**C**ambridge **M**onitor **S**ystem)

- **CMS**

- Choix du système de fichier original :
 - nommage par : filename, filetype, filemode
 - Enregistrements dans des blocs de taille fixe
 - lecture / écriture par numéro de bloc relatif du minidisque
 - Création de fichier par simple opération d'écriture
 - Dans beaucoup de commandes le filemode peut être omis : recherche des disques dans un ordre défini
- Design indulgent (arguments ou paramètres implicites, filemode recherché si pas précisé)
- Conversion de CP-40 pour tourner sur IBM S/360 modèle 67 (septembre 1966). Extension du nombre de machines et des ressources allouées.
- CP-40 et CMS en production en janvier 1967. Une pré-version d'OS/360 en machine virtuelle.
- Participation de l'université de Grenoble qui disposait de l'IBM s/360 modèle 67 sous la direction de Claude Hans
- Uniquement expérimental dans les premières années, on compte 18 systèmes installés avec CP-67 en février 1968
- Disponible pour les clients en Type III (pas d'engagement ni de support de la part d'IBM)
- CP-67 sera présent en final sur 44 systèmes IBM s/360 modèle 67 dont ¼ internes IBM.

IBM/370

- Annonce du système IBM s/370 sans translation d'adresse en juin 1970
- En mai 1971, IBM annonce l'abandon de TSS, et du développement d'une version s/370 de CP/CMS.
- Une réorganisation interne permet au développement de VM de disposer gratuitement de 30 inspecteurs logiciel pendant 2 années.
- Adaptation des machines virtuelles pour accueillir des systèmes s/370
 - ⇒ Dépendance des autres laboratoires pour développer et tester leurs systèmes, y compris pour des matériels inexistantes pour l'instant.
 - ⇒ L'état major d'IBM se rend à l'évidence. Les laboratoires d'IBM ont besoin d'un système de virtualisation de machines.
- Le 02 aout 1972 IBM annonce :
 - ✓ la mémoire virtuelle sur toute la gamme des systèmes s/370
 - ✓ VM/370 et 3 autres systèmes d'exploitation : DOS/VS, OS/VS1 et OS/VS2
- VM/370 release 1 permet de tourner VM/370 en machine virtuelle
- VM/370 release 2 en avril 1974 exploite les fonctions du microcode « Virtual Machine Assist » ainsi que le support des terminaux IBM 3270
- Support de VS1 handshaking (PAGEX)

Evolution années 70 - 80

Technique

- Gestion des défauts de page (shadow tables)
- Optimisation du code de réflexion des défauts de programme
- Gestion du timer
- Connivence DOS et MVS (PAGEX)
- Extension des assistances matériel : CP assist et VM Assist
- Révision du « scheduler » de CP
- Optimisation de CMS
- Exécution de la machine virtuelle par une instruction (SIE : Start Interpretive Execution)

Contribution

Forte participation des universités, des sociétés clientes, et des clubs d'utilisateurs au développement de fonctions, d'outils d'analyse et de gestion dans VM

- Plus de 400 systèmes VM en 1976, 1000 en 1978, et 10000 en 1985
- Développement interne IBM comme système personnel dans les années 80
 - ⇒ Messagerie (PROFS, OfficeVision)
 - ⇒ Recherche documentaire
 - ⇒ Connexion et échange de fichiers à distance
- Généralisation de la présence de VM avec les systèmes DOS/VSE

Les besoins

Définitions et types de virtualisation

Les premiers systèmes virtuels

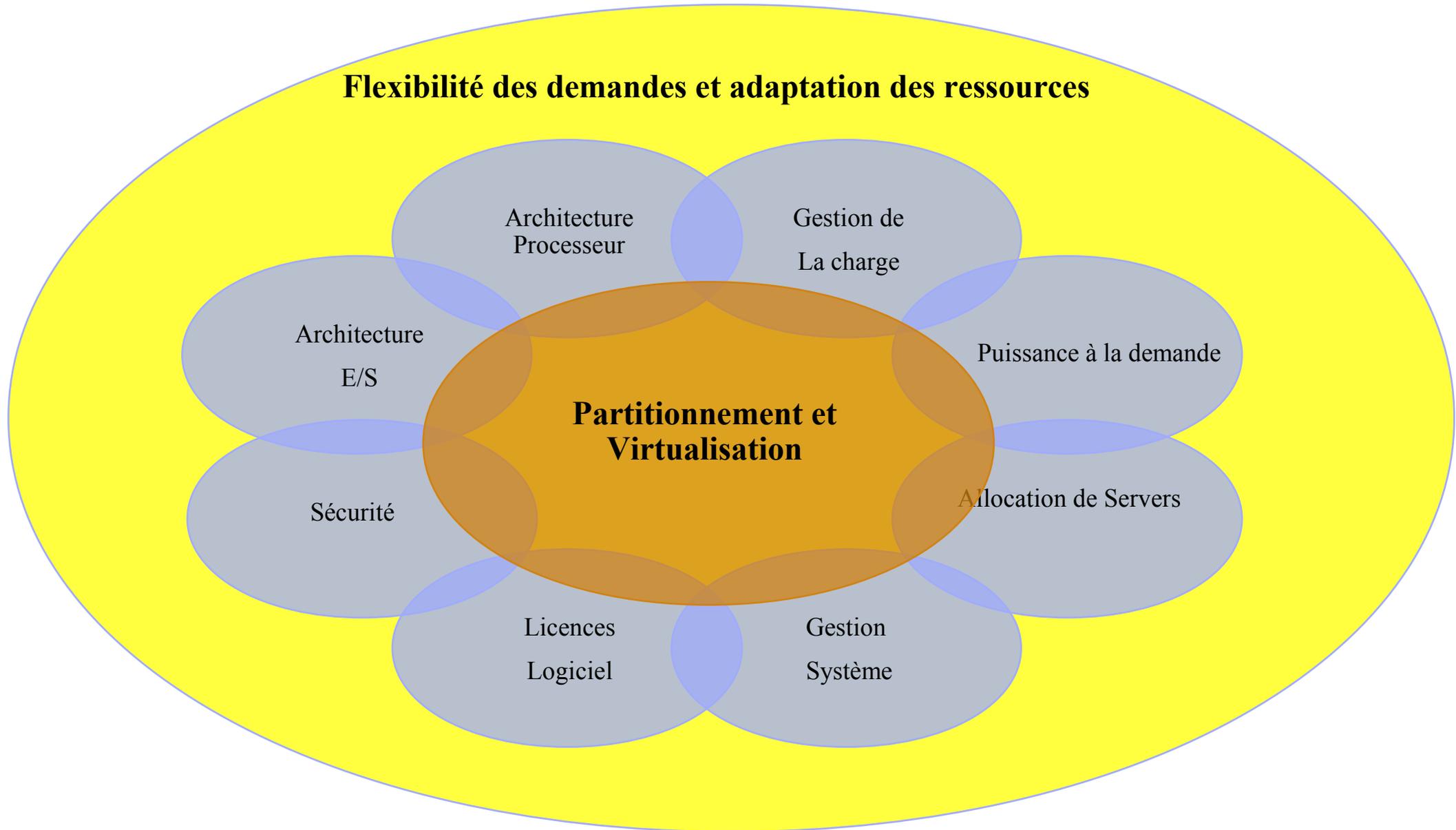


Les besoins

Techniques mises en oeuvre

Situation actuelle

Optimisation des ressources et de déploiement



Bénéfices

- Réduction des coûts matériel et environnementaux
 - ✓ Optimisation et mutualisation des ressources
 - ✓ Gain d'espace et de consommation électrique
- Réduction des coûts de gestion
 - ✓ Moins de serveurs physiques à gérer → Tâches de gestion communes
- Flexibilité et réactivité selon à la demande
 - ✓ Ajustement des ressources pour satisfaire aux nouvelles demandes
 - ✓ L'ajout ou le retrait de serveurs en quelques minutes
- Isolation des applications et des serveurs au niveau matériel
- Cohabitation de systèmes de production et de test
- Migration vers de nouvelles releases/versions ou de système d'exploitation
- Support d'anciennes versions
- Clonage de serveurs facilité
- Serveurs de secours polyvalent
- Mise au point de nouveaux systèmes grâce aux outils de l'hyperviseur
 - ✓ Débogage interactif impossible sur le serveur réel
- Éducation des professionnels dans des conditions réelles

Contraintes et choix

Contraintes :

- Disponibilité dépendant de la solidité du matériel et du logiciel de l'hyperviseur
- Centralisation des ressources, nécessité de sites de secours
- Modification de certains systèmes pour optimiser leur exécution en machine virtuelle
 - Exemple : Linux jiffies
- Coût logiciel supplémentaires dans certains cas
- Configuration matériel et logiciel spécifiques pour bénéficier de toutes les fonctions

Virtualisation ?

- Serveurs déjà très chargés ou I/O intensifs
- Applicatifs sachant bien gérer des requêtes nombreuses :
 - Bases de données
 - Transactionnels performants
- Applications critiques (temps réel)
- Systèmes conçus pour des serveurs individuels

Techniques mises en oeuvre

Définitions et types de virtualisation

Les premiers systèmes virtuels

Les besoins



Techniques mises en oeuvre

Situation actuelle

Dispatching des machines virtuelles :

- Nécessité de réduire la part de l'hyperviseur dans la virtualisation :
 - ✓ Descriptif de la machine virtuelle connue du processeur
 - Intel : Virtual Machine Control Structure (VMCS)
 - IBM système z : SIE State Descriptor Block (SIEBK)
 - ✓ Extensions suivant l'architecture
 - IBM : SIE sous SIE
 - Intel : VMCS shadowing

Mémoire :

- ✓ Un seul niveau de virtualisation → affectation de mémoire réelle

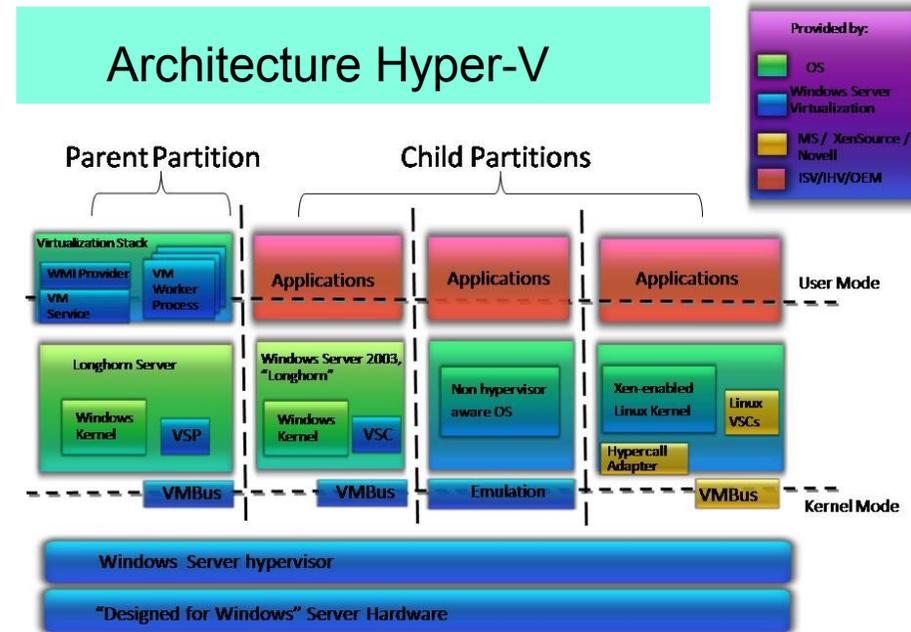
Entrées/Sorties :

- ✓ Exécution des demandes d'entrées/sorties sans intervention de l'hyperviseur
- ✓ Préférez les unités virtuelles simulées quand cela est possible
- ✓ Partage direct des unités sans dégradation ou matériel supplémentaire
- ✓ Utilisation de la gestion par l'hyperviseur quand c'est possible
- ✓ Affectation de mémoire réelle pour des accès direct aux données (réseaux)

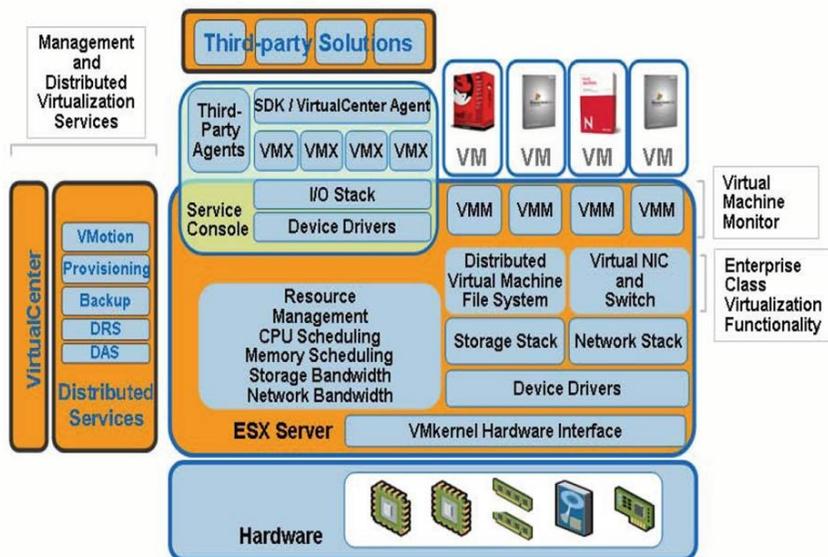
Les systèmes basés sur processeurs Intel

Utilisés pour virtualiser de 2 à 30 serveurs

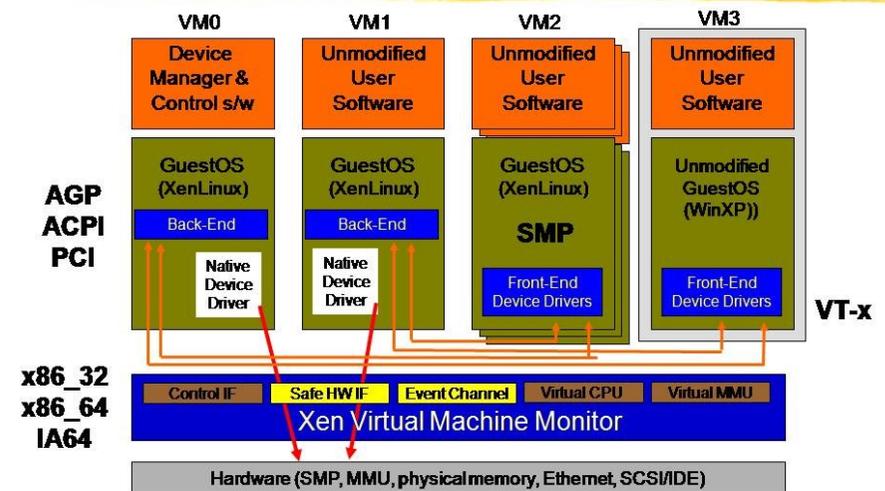
- Hyper-V (Microsoft)
 - ✓ Hyperviseur de type 2
 - ✓ Maximum 64pCPU, 4vCPU
 - ✓ Mémoire 1To
 - ✓ Windows, Linux
- VMware (EMC)
 - ✓ Hyperviseur de type 1
 - ✓ Maximum 32 vCPUs et 96Go de mémoire
 - ✓ Windows, Linux, Solaris
- Xen (CITRIX)
 - ✓ Hyperviseur de type 1
 - ✓ Maximum 8vCPUs et 256Go de mémoire
 - ✓ Windows, Linux, Solaris



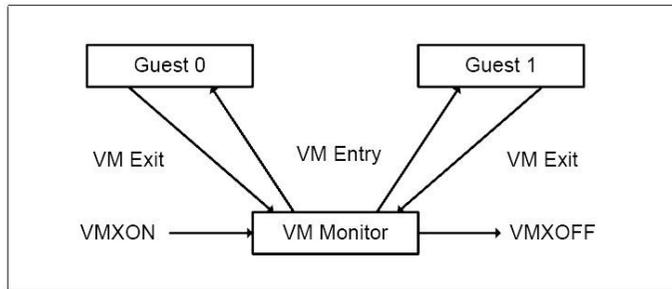
Architecture VMware



Architecture Xen

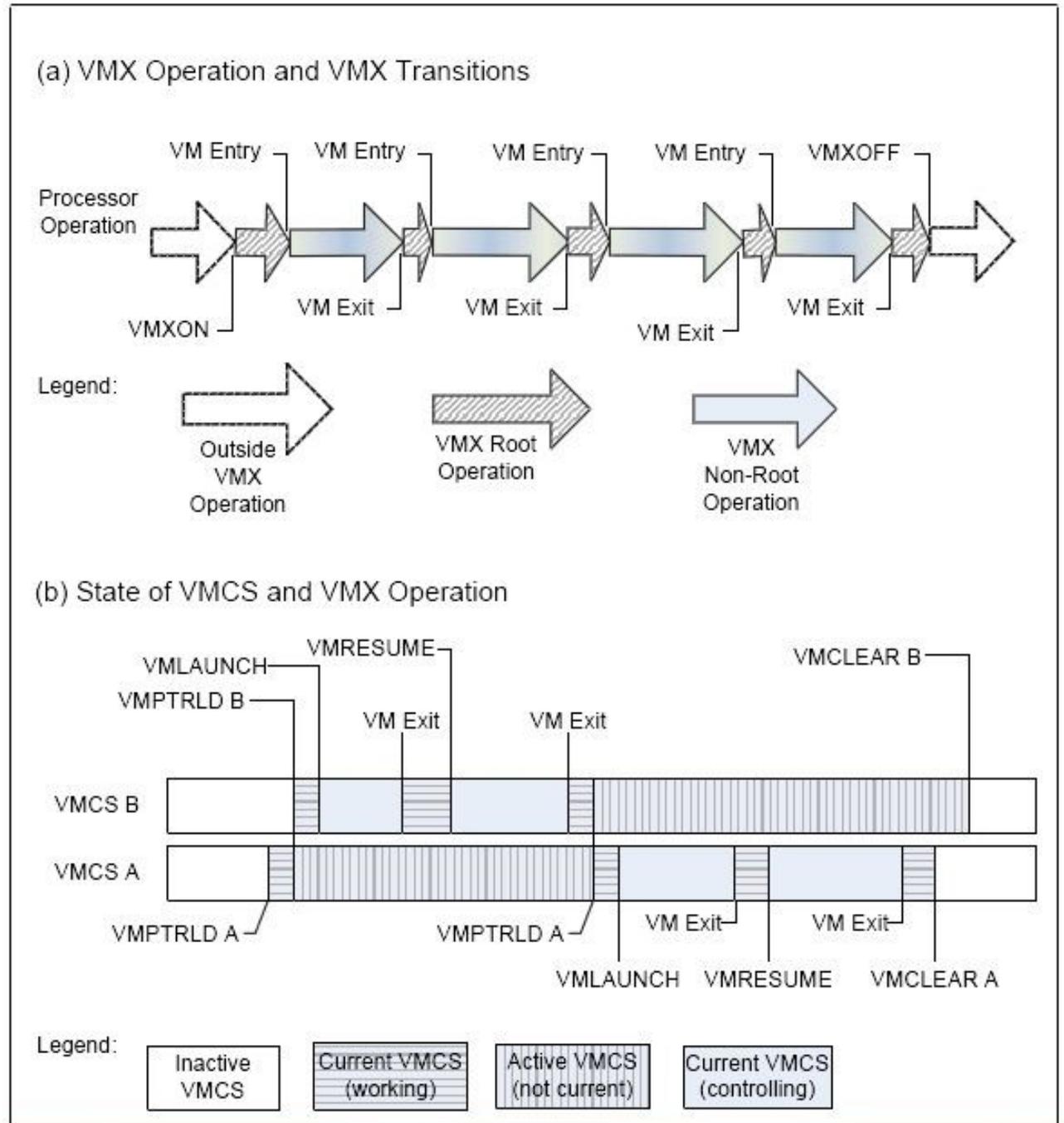


Exécution de la machine virtuelle (Intel)

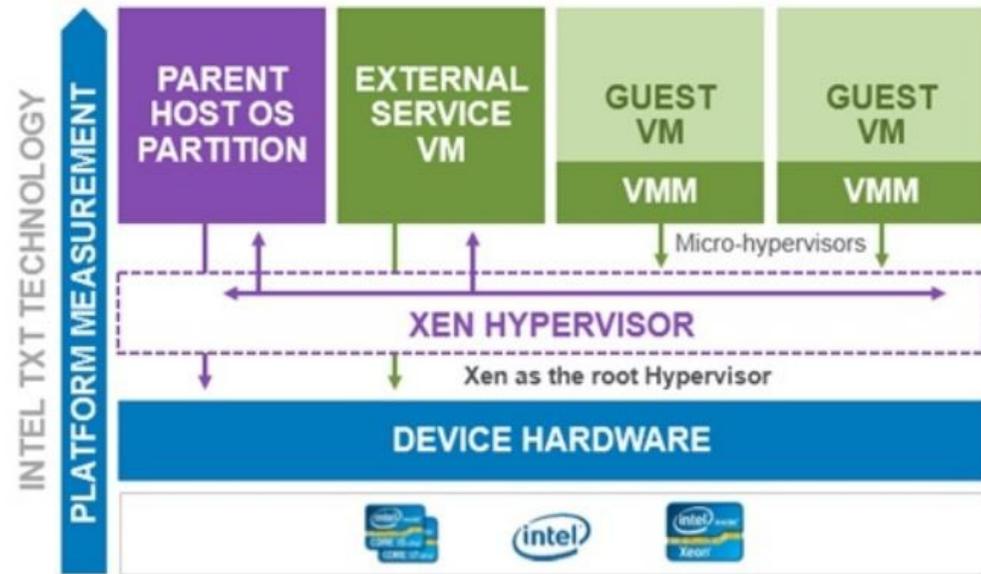
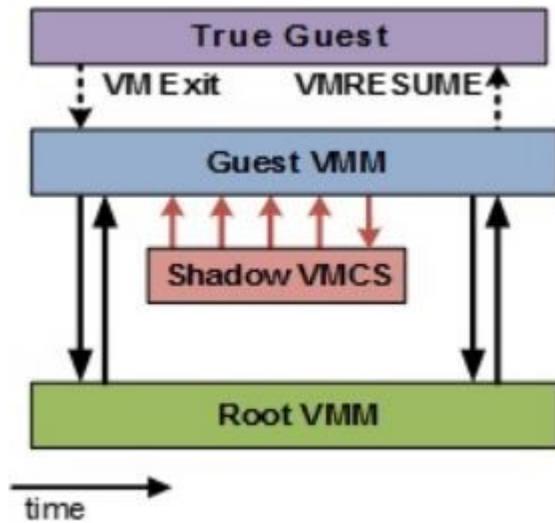


Démarrage du mode virtuel (VMX) par l'instruction VMXON

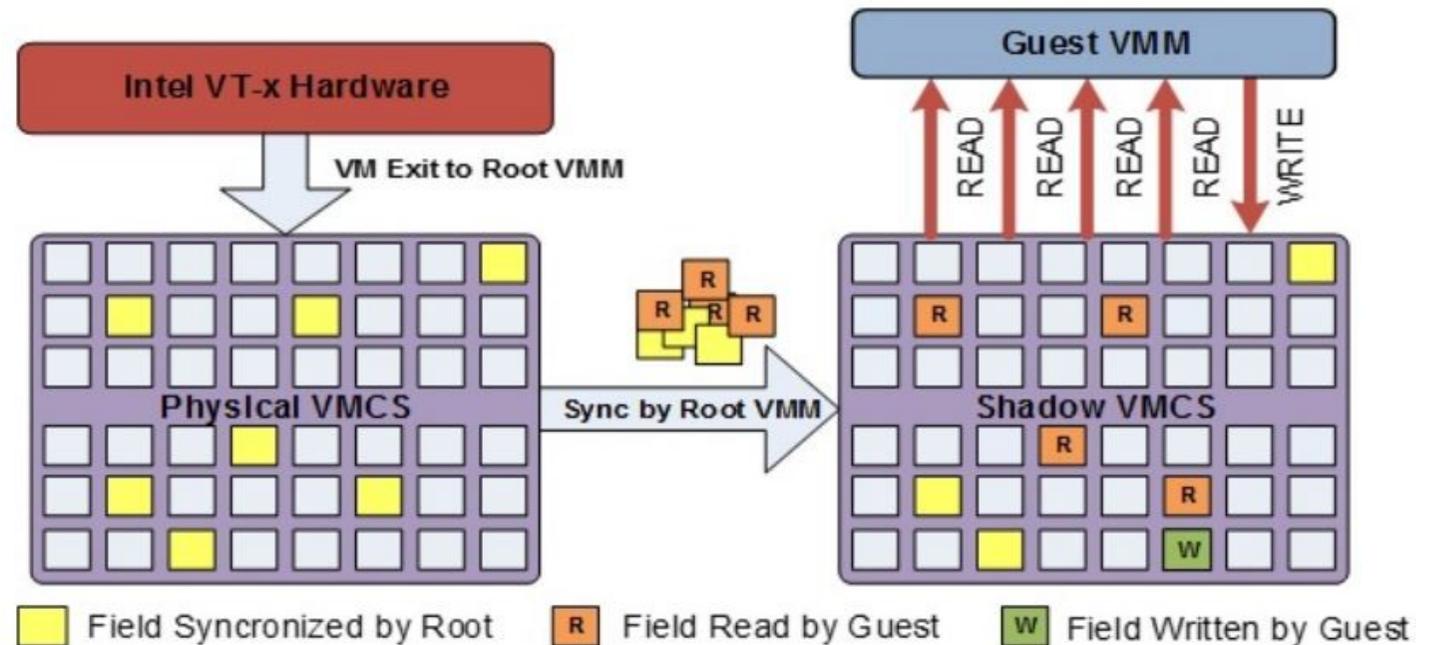
Sortie du mode virtuel par l'instruction VMXOFF



Exécution de la machine virtuelle (Intel)

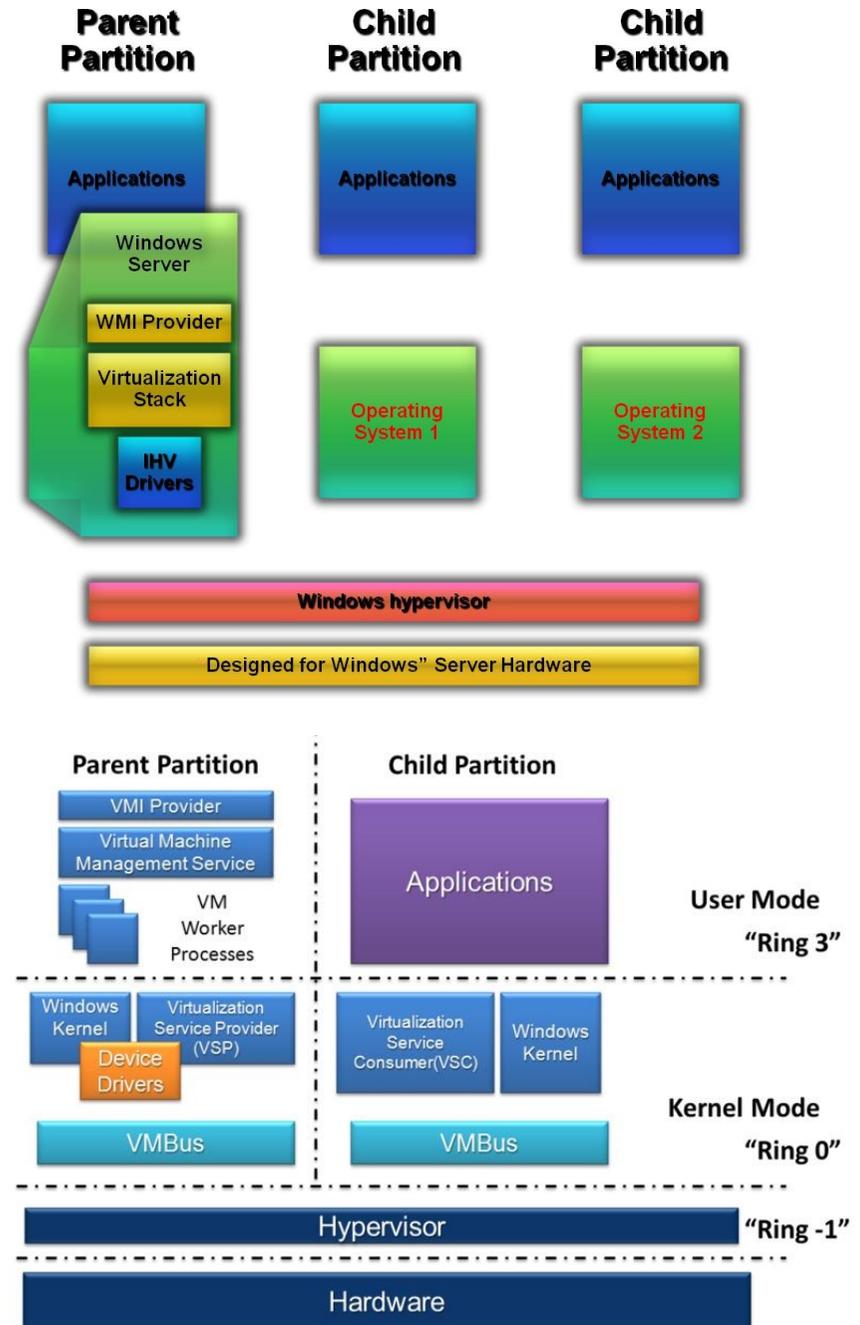
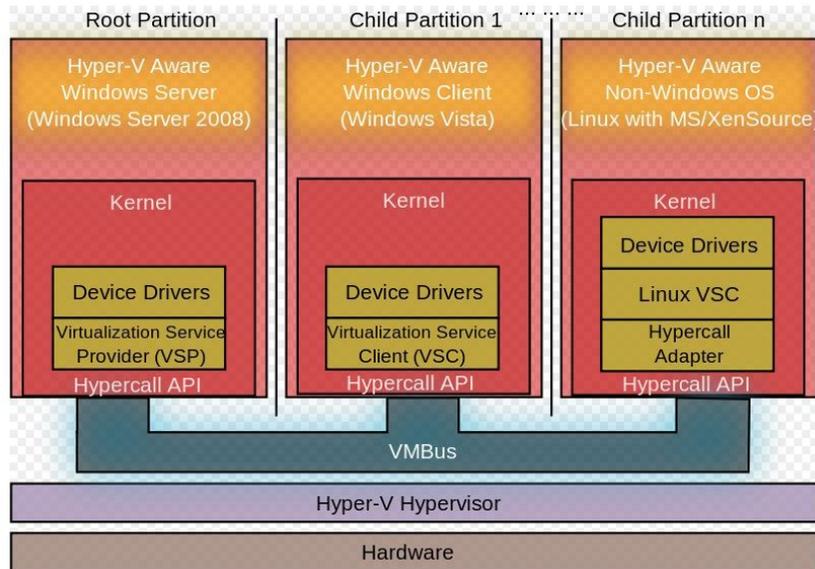


L'utilisation de bitmap réduit le temps de synchronisation des VMCS

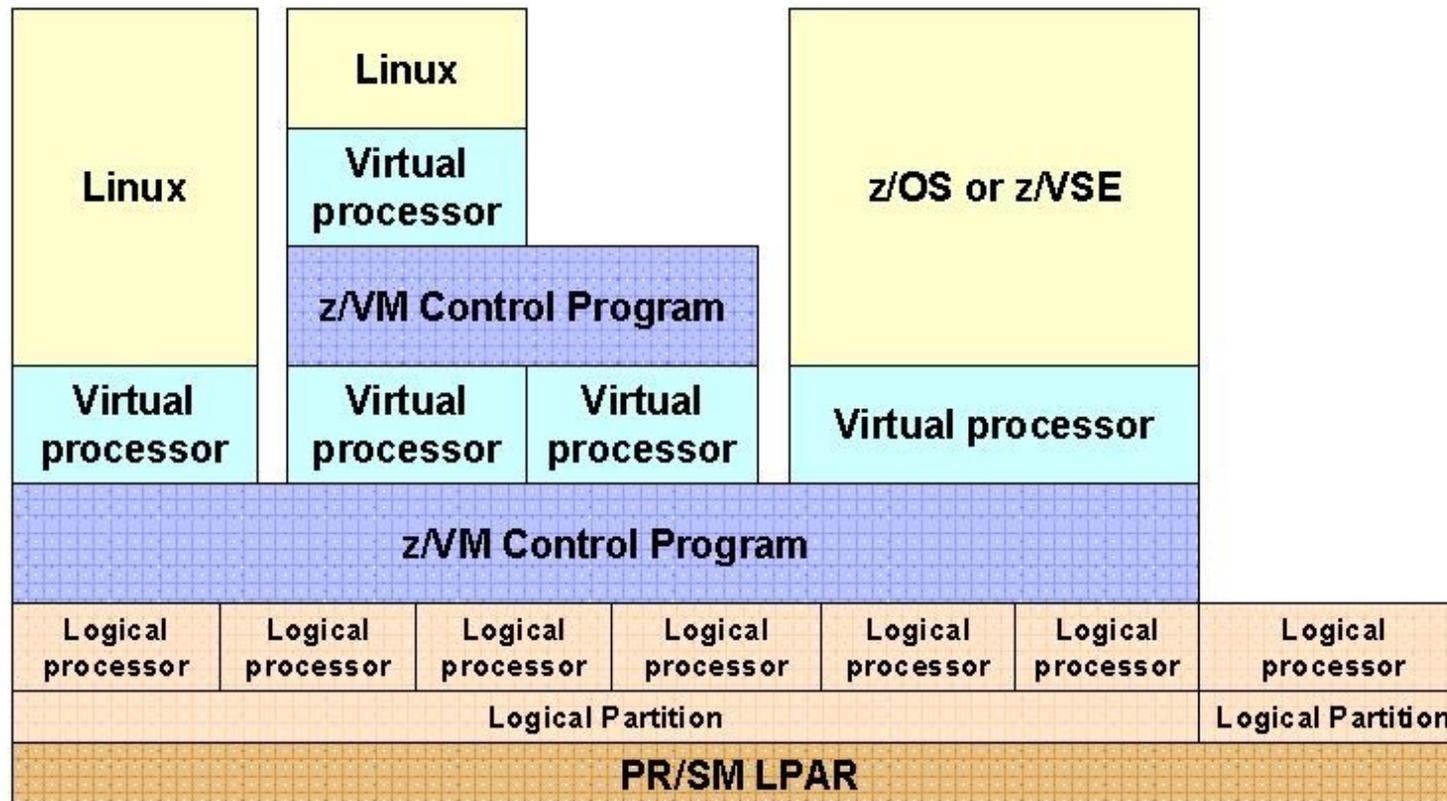


Exécution de la machine virtuelle (Microsoft)

- L'accès aux périphériques matériel est assuré par la partition Parent



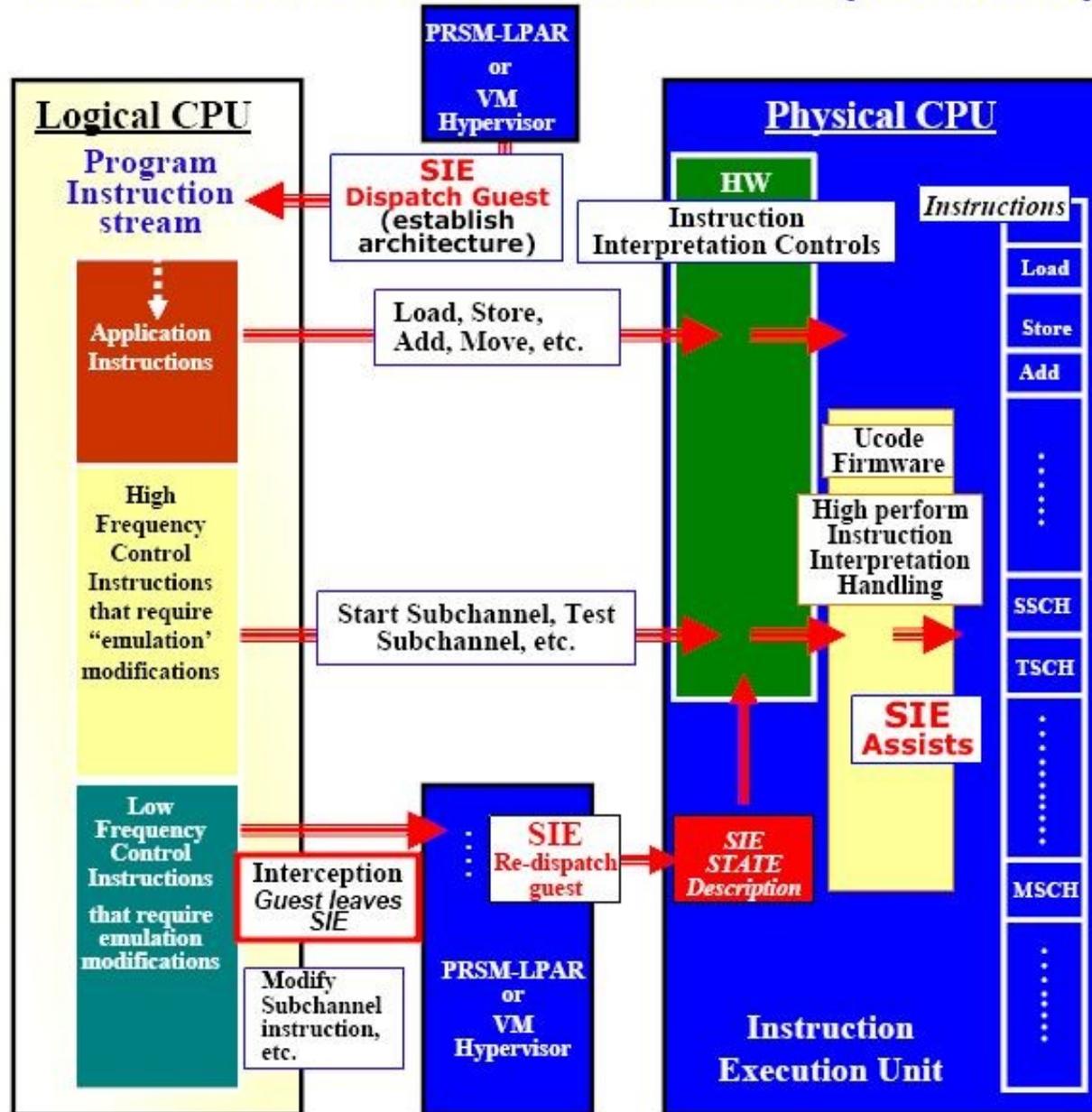
Architecture IBM système z



- PR/SM virtualise les ressources matériel et gère l'exécution des partitions logiques
- PR/SM est le seul mode disponible sur les systèmes IBM z actuels

Exécution de la machine virtuelle (IBM système z)

Basic Direct HW virtualization – transparent to applications/OS



System z with SIE (Start Interpretive Instruction Execution)

* System z runs ALWAYS in PR/SM-LPAR mode under SIE

* LPAR is the "only" game in town, meaning performance items and other functionalities is developed accordingly

zVM invokes SIE to run VM's (SIE under SIE)

* Efficient for performance - and new version of OS and Hypervisor

Positioning of System z & Intel/AMD

System z:

the requirement for the underlying HW support is NOT an issue for System z, - since the basic System z Architecture & HW design has implemented this for "decades".

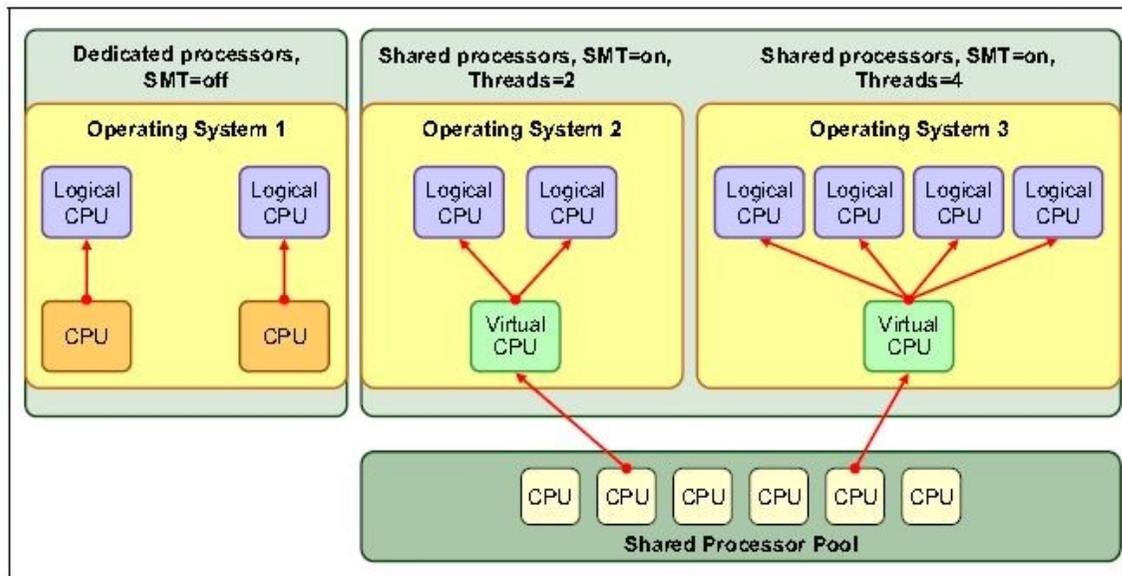
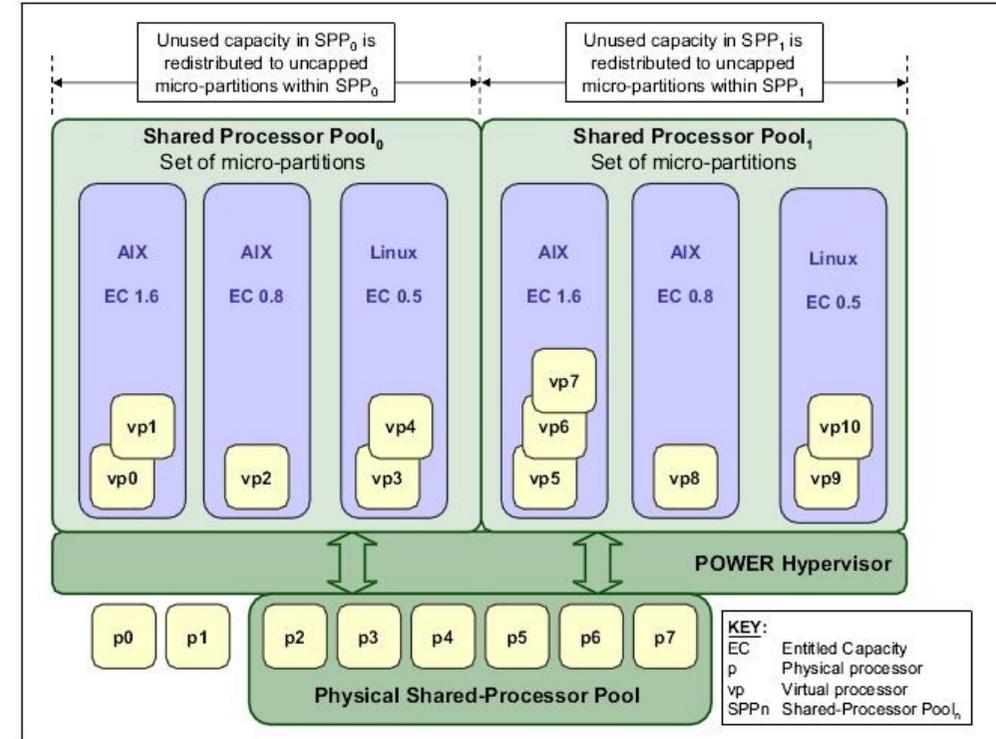
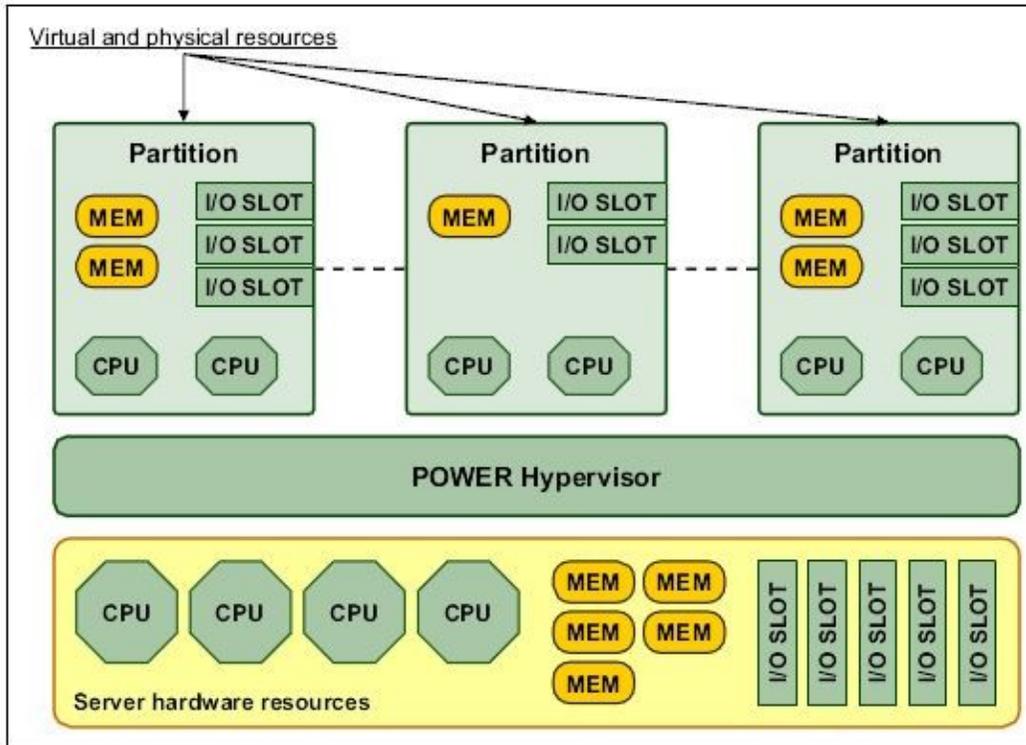
Intel and AMD

is developing some virtualization HW support based on a similar structure like the SIE architecture as it was externally documented 25 years ago.

The amount of HW and SIE assist functionalities are seemingly rather limited at this point.

No "SIE" under "SIE"

Architecture POWERVM



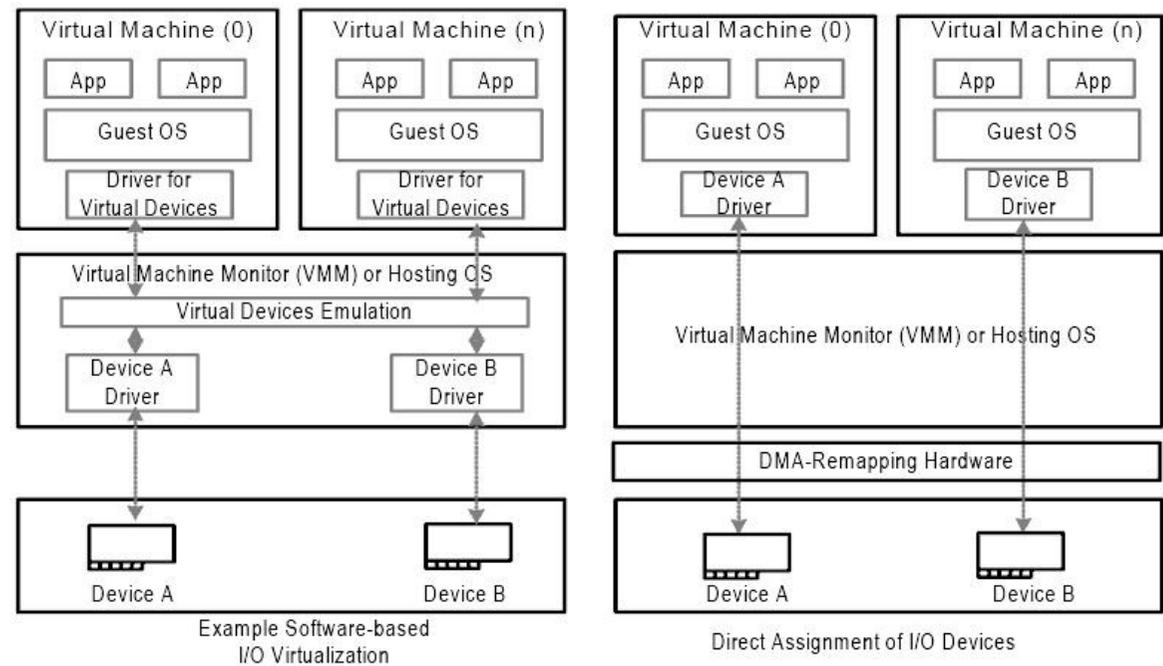
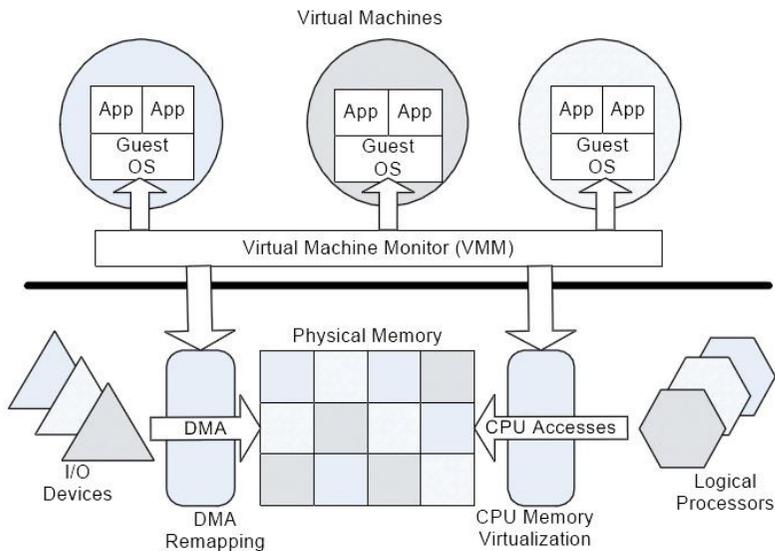
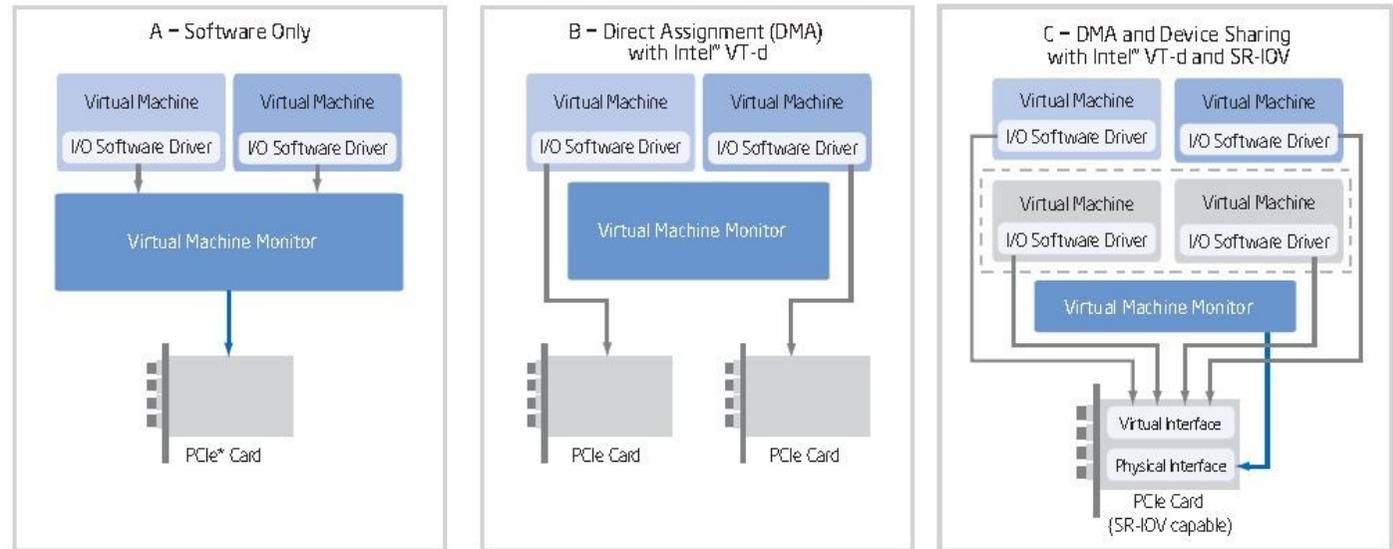
- Les processeurs partagés peuvent être groupés en pool
- Chaque processeur virtuel gère de 0 à 4 tâches

Exécution directe des Entrées/sorties

IBM : I/O Passthru, DIAG98

Intel : DMA

I/O Virtualization Solutions



Techniques de virtualisation

Les processeurs

La mémoire

Les unités de stockage

L'interface de commande de l'ordinateur

Les terminaux

Les interfaces réseau

Autres unités et fonctionnalités

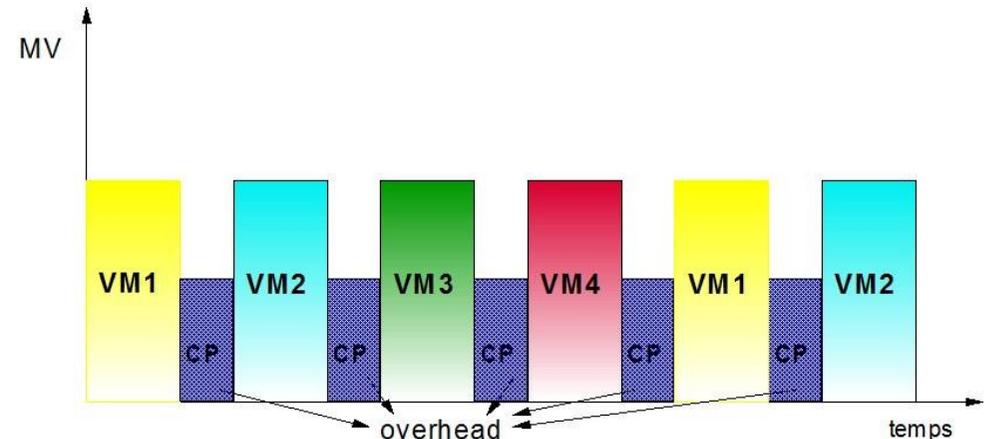
Les processeurs

- Processeur(s) dédié(s) ou partagés
- Distribution du processeur par tranche de temps « time slice ou quantum » pour chaque machine selon sa priorité ou son poids (relatif ou absolu).

- Contrôle repris par l'hyperviseur en cas de :

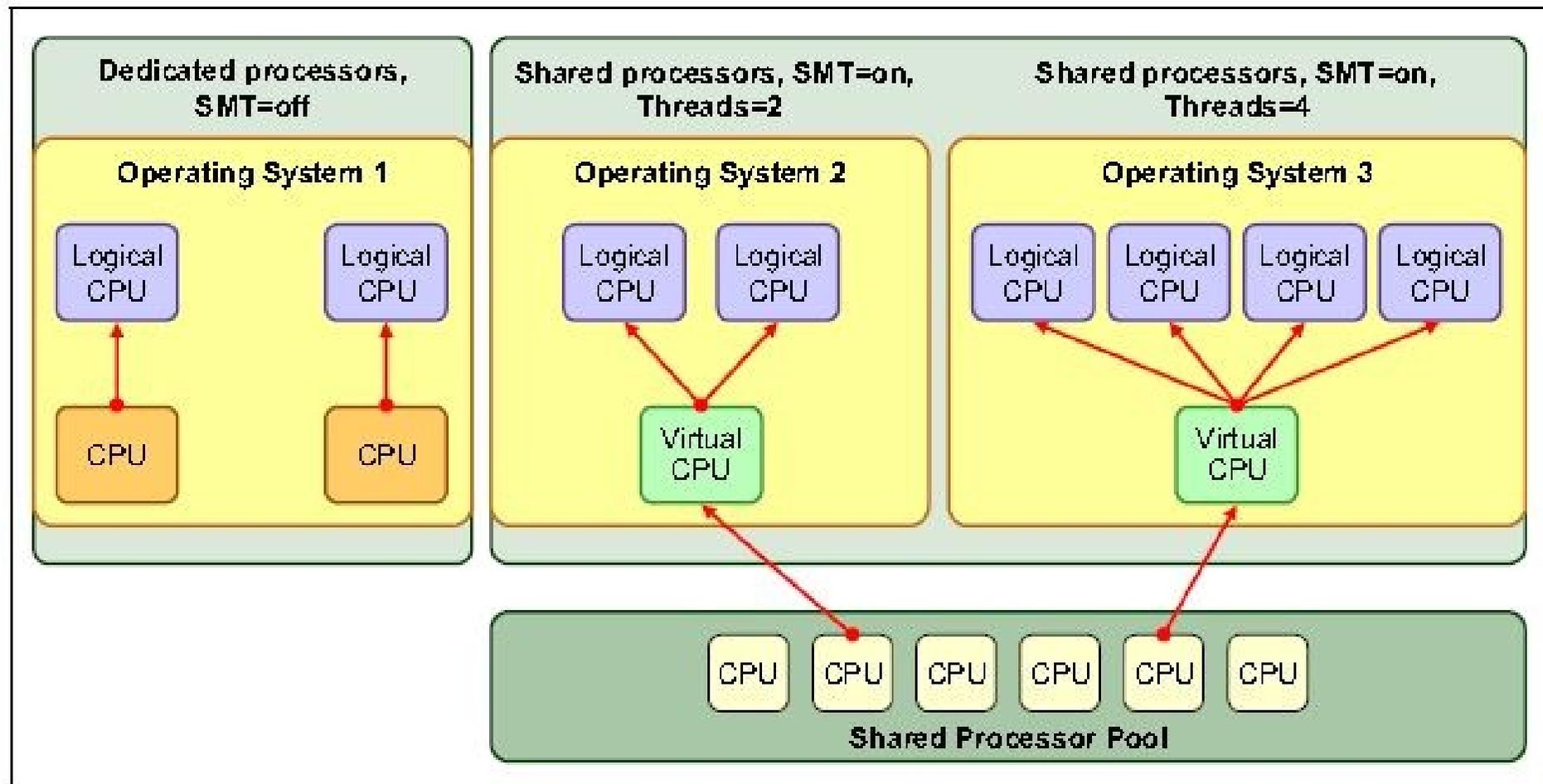
- ✓ Fin de la tranche de temps
- ✓ État d'attente système
- ✓ Interruption non masquée
- ✓ Défaut de page, simulation d'unité d'entrée/sortie,
- ✓ simulation d'instruction pas prise en charge
- ✓ Appel hyperviseur (Diagnose (z/architecture), VMCALL (Intel), Hcall (PowerVM))
- ✓ Prémption par une machine plus prioritaire
- ✓ Abandon volontaire de la machine virtuelle (active wait) ou attente sur une ressource détenue par un autre processeur virtuel

- Scheduling lié au profil de la machine virtuelle (interactive, batch, long batch)
- Priorité relative ou absolue de la machine virtuelle
- Utilisation plafonnée (capping)



Processeurs logiques (POWER7)

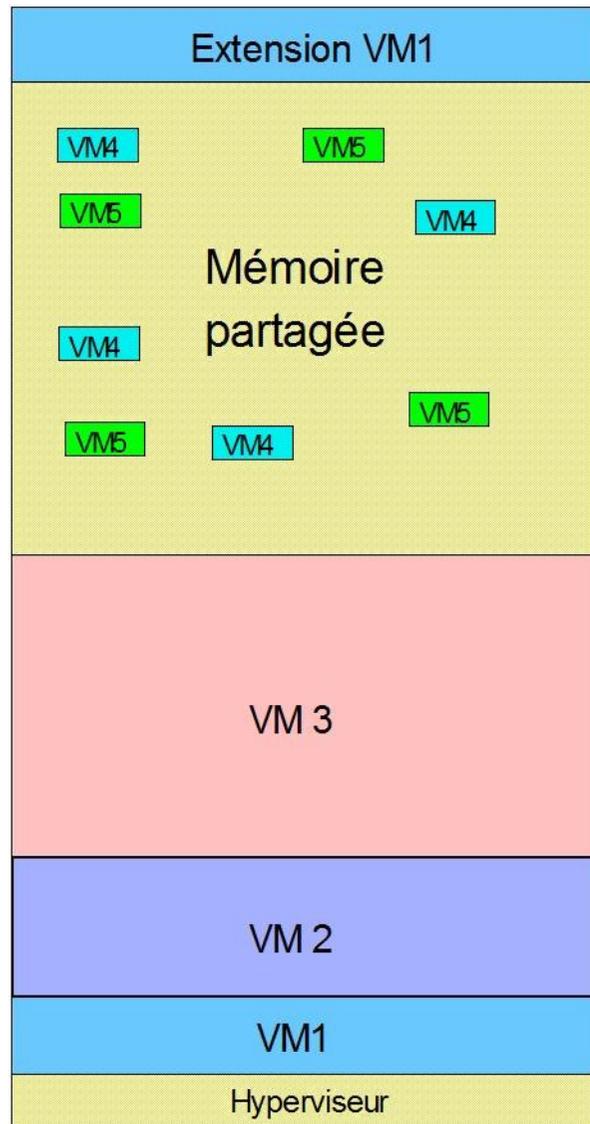
- Le système d'exploitation distribue les processus entre les processeurs logiques
- Chaque processeur dédié ou virtuel a un, deux ou quatre processeurs logiques associés suivant le nombre de processus SMT définis



Détection des instructions privilégiées

<p style="text-align: center;">Interception et émulation</p> <p>Exemples CP-67, VM/370 Bénéfices OS non modifié Inconvénients Dégradation importante</p>	<p style="text-align: center;">Traduit, intercepte et émule</p> <p>Exemples VMware, Microsoft Hyper-V Bénéfices OS modifié, traduction nécessaire Inconvénients Dégradation importante</p>
<p style="text-align: center;">Appels Hyperviseur (“Paravirtualisation”)</p> <p>Exemples POWER Hyperviseur, Xen, CMS Bénéfices Efficacité élevée Inconvénients OS modifié pour générer des Hcalls</p>	<p style="text-align: center;">Virtualisation matérielle directe</p> <p>Exemples PR/SM, z/VM (Utilise aussi des appels hyperviseur) Bénéfices Efficacité maximale dépendant du matériel OS non modifié Inconvénients Réclame un support matériel sous-jacent</p>

Mémoire dédiée et/ou partagée



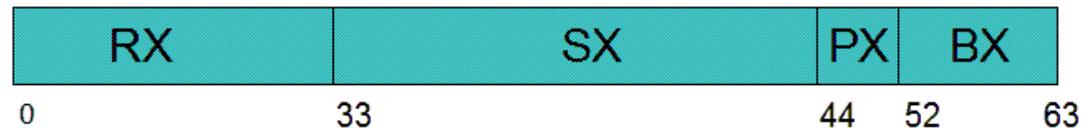
- Chaque système virtuel dispose d'une part fixe de mémoire réelle :
 - ✓ exemple : LPAR système z
- Mémoire fixe ou partagée
 - ✓ exemple : z/VM , POWERVM
- Mémoire réservée pour extensions statiques ou dynamiques

La mémoire virtuelle

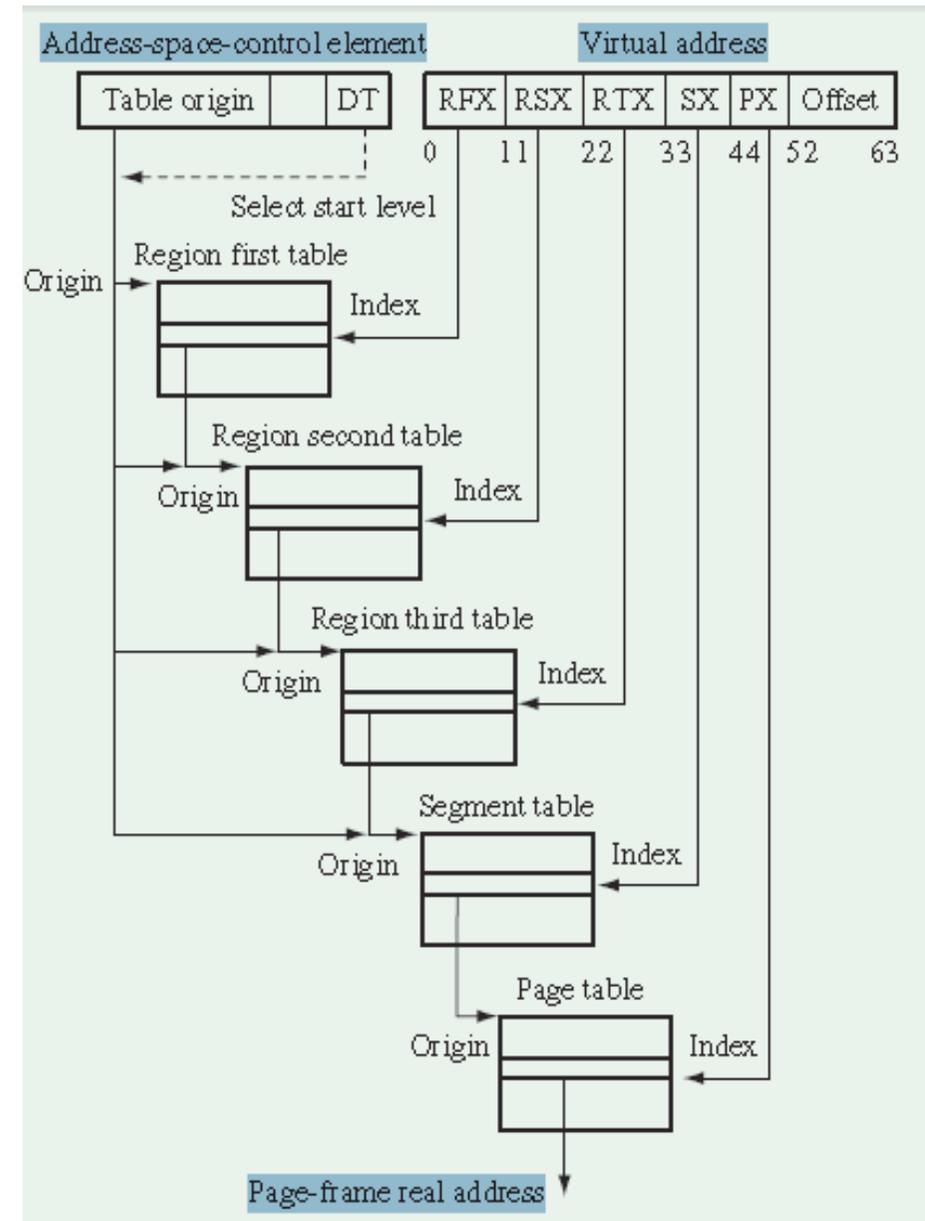
- Création pour chaque Machine Virtuelle des tables de segment et de pages correspondant à la taille de la machine virtuelle définie (limitation liée à la version de l'hyperviseur).
- Possibilité de fractionner la mémoire virtuelle pour certains hyperviseurs
- Nécessité de création de « shadow tables » pour tenir compte de la structure de la mémoire virtuelle de la Machine virtuelle
- Partage de mémoire virtuelle entre plusieurs Machines virtuelles

La mémoire virtuelle (système z)

Adressage 64-bit



- RX Region index (une entrée représente 2Go)
- SX Segment index (une entrée représente 1Mo)
- PX Page index (une entrée représente 4Ko)
- BX Byte index (déplacement dans la page)



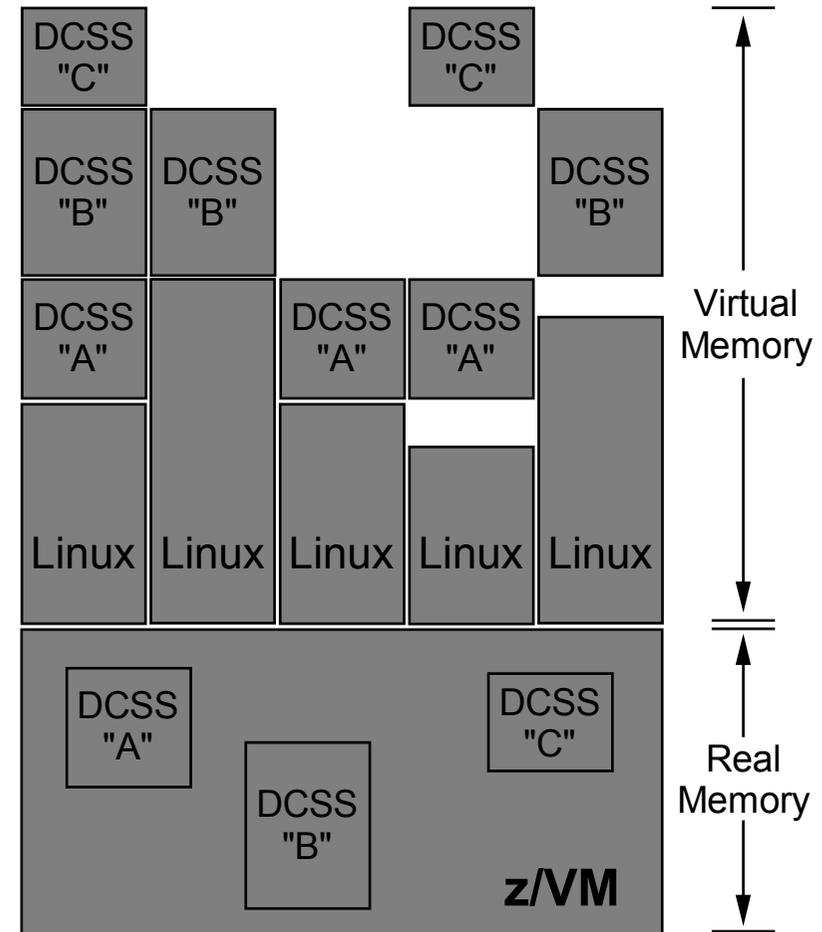
Mémoire partagée

- Espaces mémoire partagés par plusieurs systèmes (Share Segments – Deduplication)
- Connivence CP – Linux (Cooperative Memory Management)
- Possibilité de définir des pages à conserver en mémoire centrale (fixed, pinned, locked, wired)
- Des systèmes virtuels peuvent initier des opérations d'entrée/sorties avec des adresses mémoire réelles (applications réseaux)

Mémoire partagée Linux avec z/VM

Exploitation du support z/VM pour les segments partagés (DCSS)

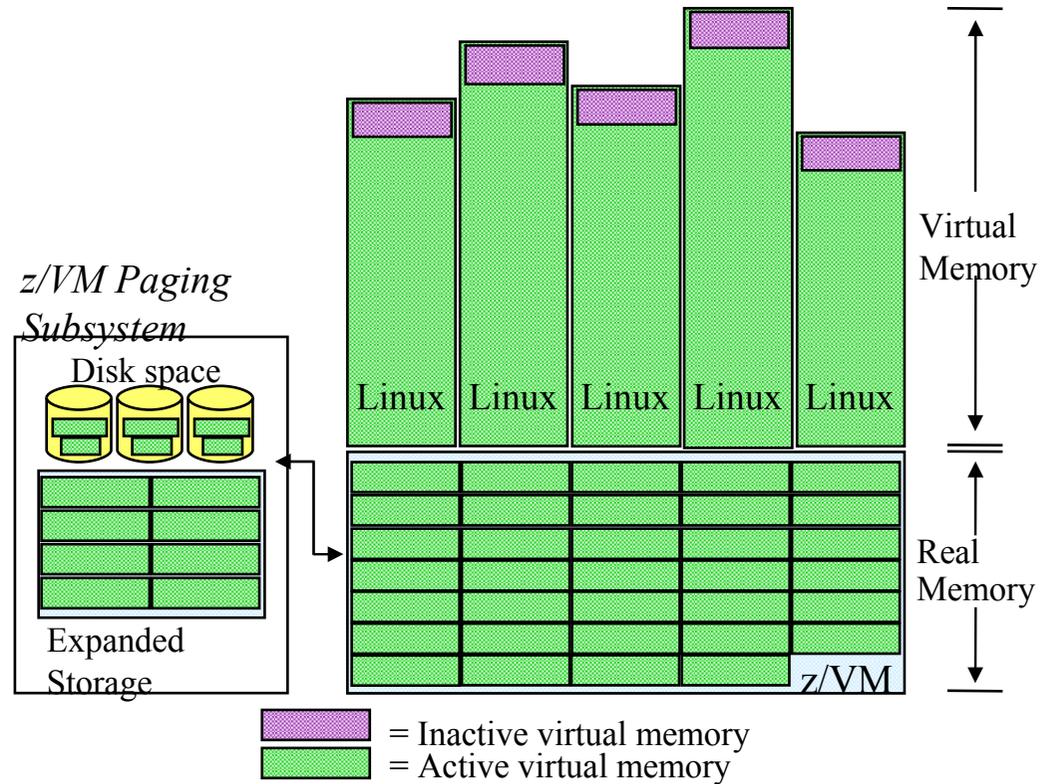
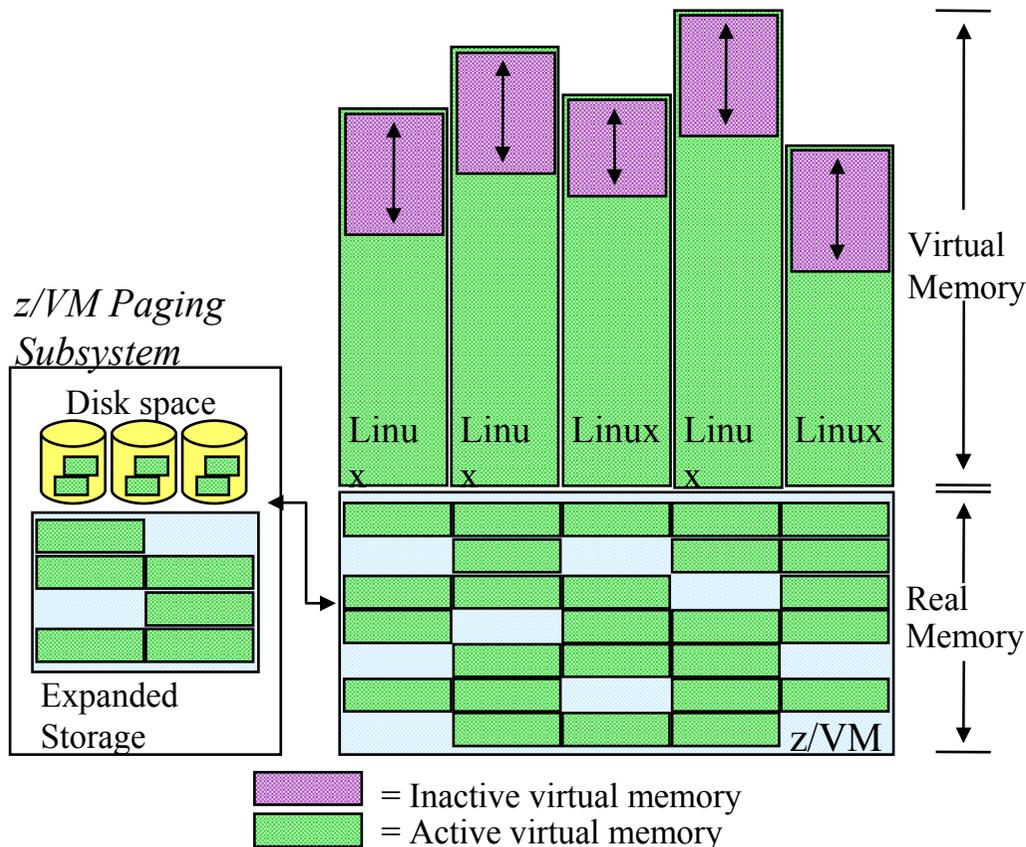
- Exclusivité de z/VM
 - Partage une unique zone mémoire réelle entre des machines virtuelles multiples
 - Performance optimum des accès mémoire
 - Aide à réduire l'utilisation de mémoire réelle
- Exploitation par Linux disponible
 - Système de fichiers Execute-in-place (xip2)
 - Les segments partagés peuvent résider au delà de la taille mémoire définie pour la machine
 - L'accès au système de fichiers se fait aux vitesses mémoire; les exécutables sont appelés directement en dehors du système de fichiers (pas de déplacement de données requis)
 - Évite la duplication de mémoire virtuelle et de données enregistrées sur disque
 - Les images Linux bénéficient de meilleures performances et d'augmentation de leur nombre dans une même configuration.



Mémoire partagée Linux avec z/VM

Gestion mémoire collaborative

- **Problème:** l'utilisation de mémoire virtuelle excède fortement la taille de la mémoire réelle
- La pagination du Control Program de z/VM devient excessive
- La performance globale est dégradée

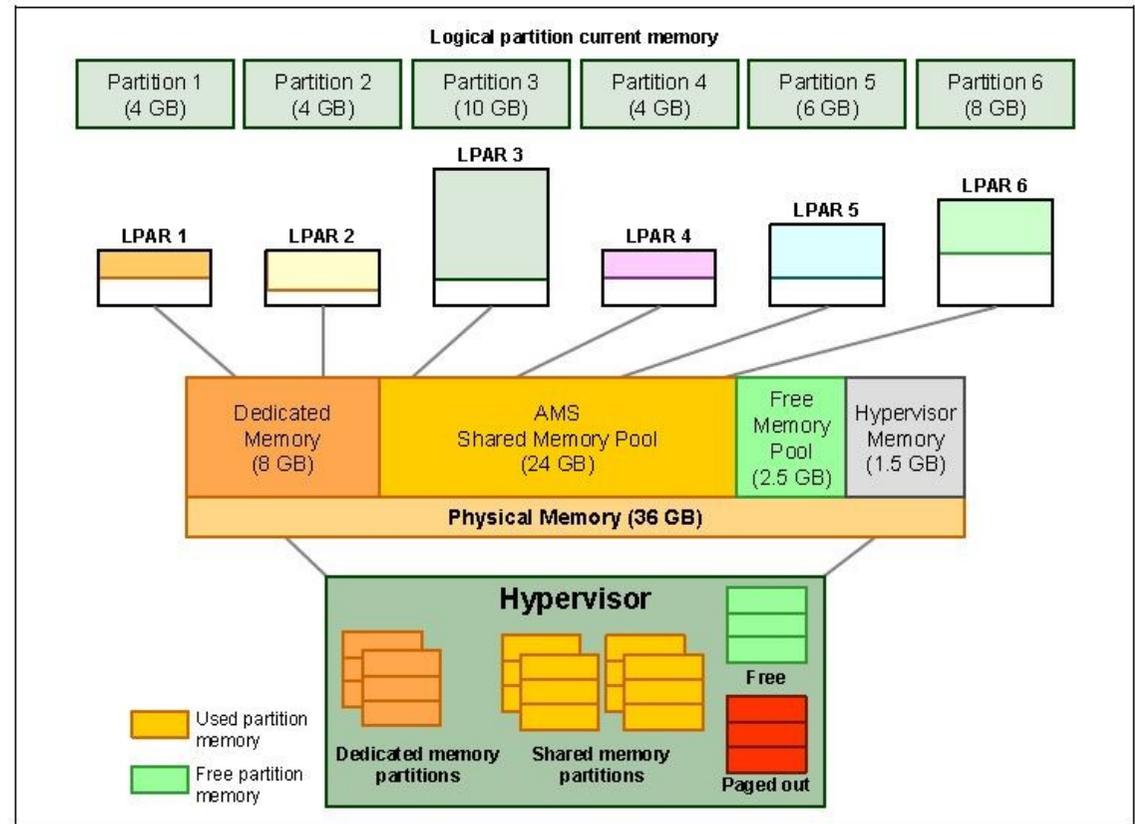
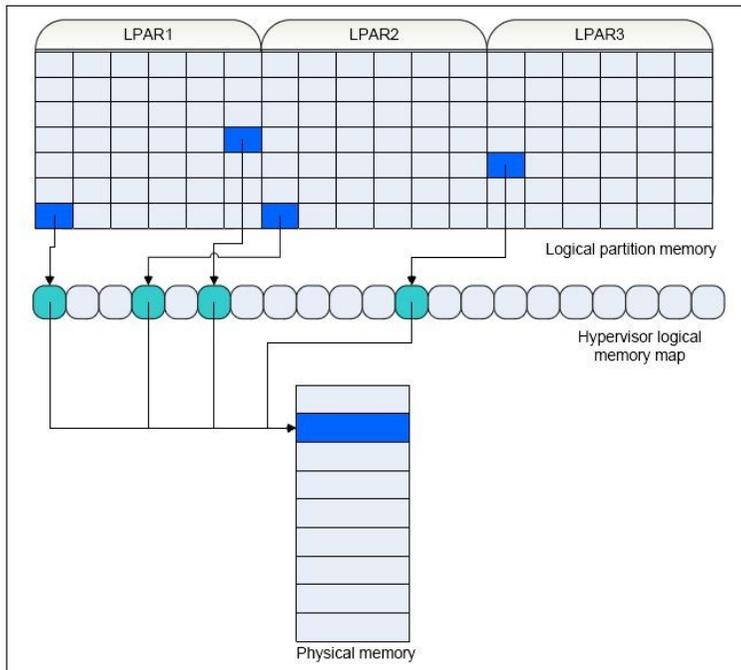


- **Solution:** Détection de la contrainte mémoire et signal envoyé aux images Linux pour réduire leur utilisation de mémoire
- Les pages de mémoire Linux sont rendues
- Les besoins en mémoire réelle et la pagination sont diminués
- Performance système et des images améliorée

Mémoire partagée PowerVM (deduplication)

Applicable seulement aux partitions qui partagent un « pool » de mémoire

Les pages des différentes partitions avec un contenu identique n'occupent qu'une page en mémoire physique (code réentrant ou données en lecture seule)



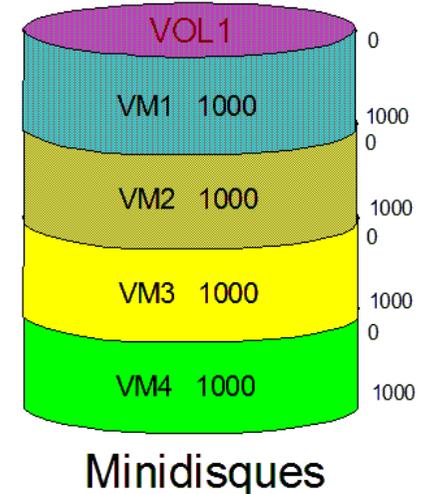
Unités externes

- **Dédiées ou attachées**
 - ✓ La machine virtuelle a l'usage exclusif de l'unité réelle
- **Virtualisée**
 - ✓ Part ou totalité d'une unité réelle pouvant être partagée par d'autres machine virtuelle
- **Simulée**
 - ✓ unité sans contrepartie réelle
 - Canaux, disques, LAN virtuels, unités spoulées
- **Emulée**
 - ✓ Unité virtuelle d'un type différent de l'unité réelle

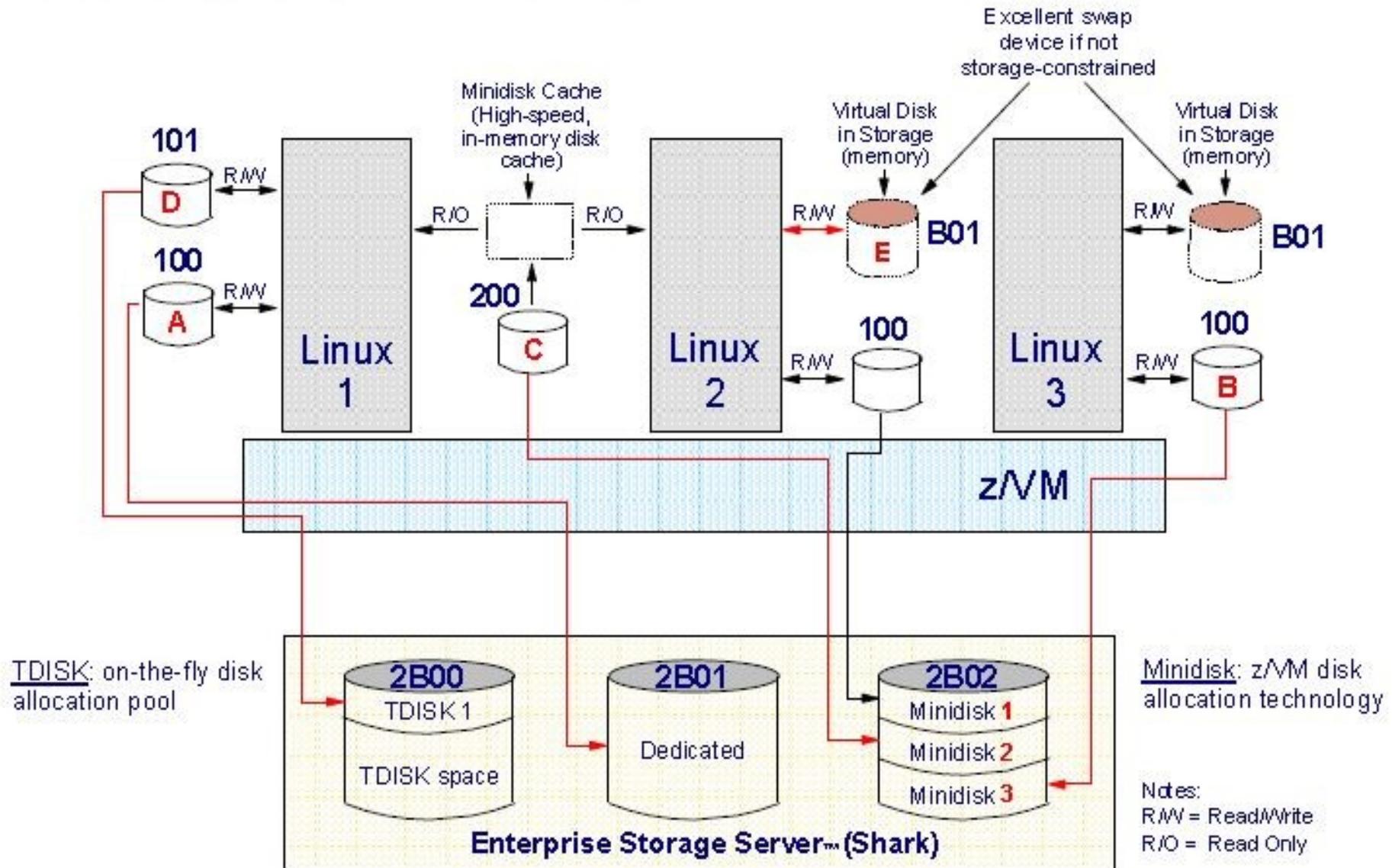
Les unités de stockage disque

Disques virtuels

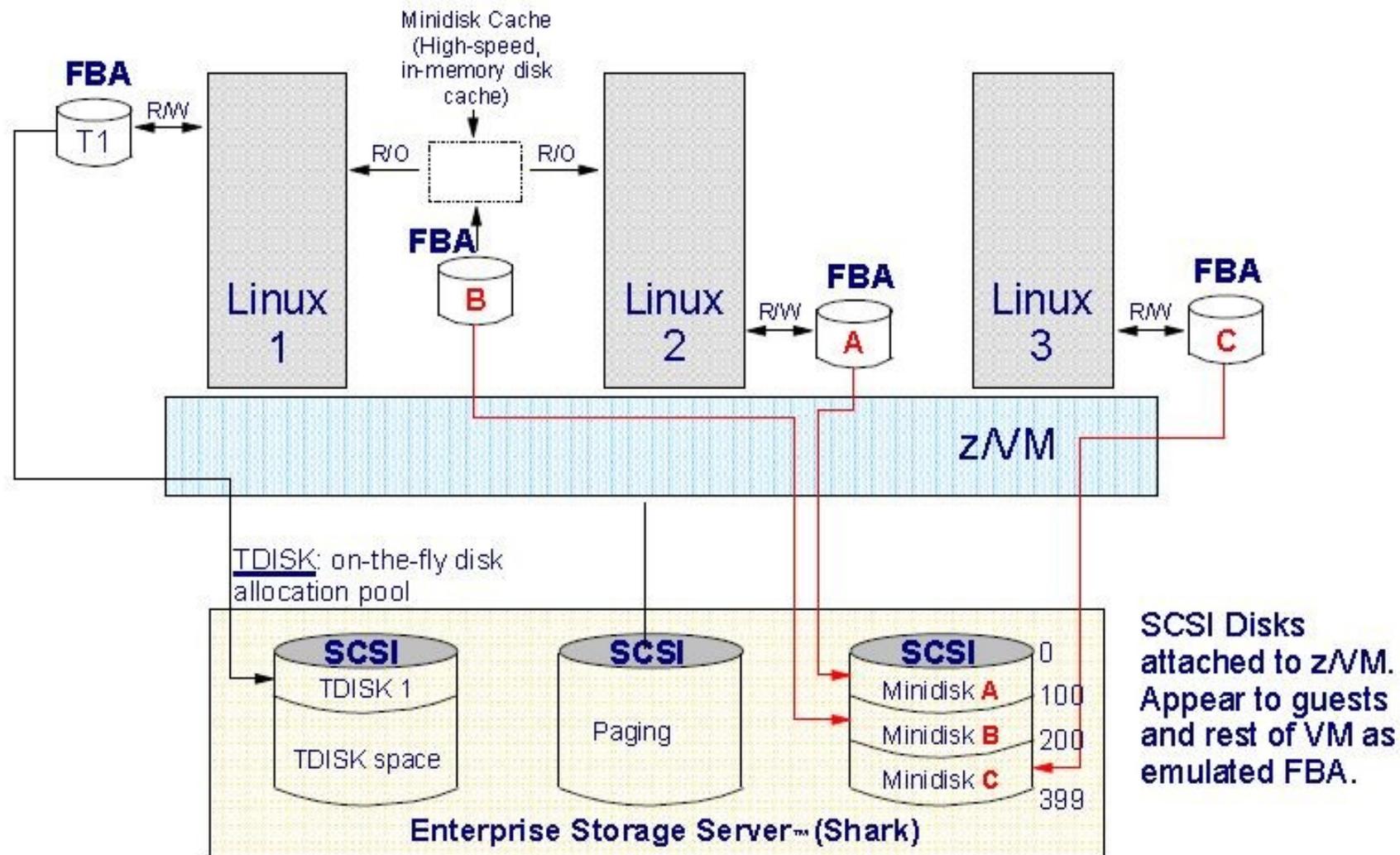
- Affectation de volumes à la machine virtuelle
- Disque virtuel sous forme de fichier .vhd (Microsoft) ou .vmdk (VMWare)
- Zone partielle ou entière de disque réel (minidisques) 
 - ✓ Cache en mémoire de CP paramétrable individuellement
- Espace disques groupant plusieurs unités (volumes logiques, filepool)
- Disques réseaux
- Disques virtuels résidant en mémoire (temporaires)
- Disques temporaires (espace d'un disque réel)
- Partage de disques entre plusieurs machines virtuelles



Les disques virtuels CKD



Les disques virtuels SCSI



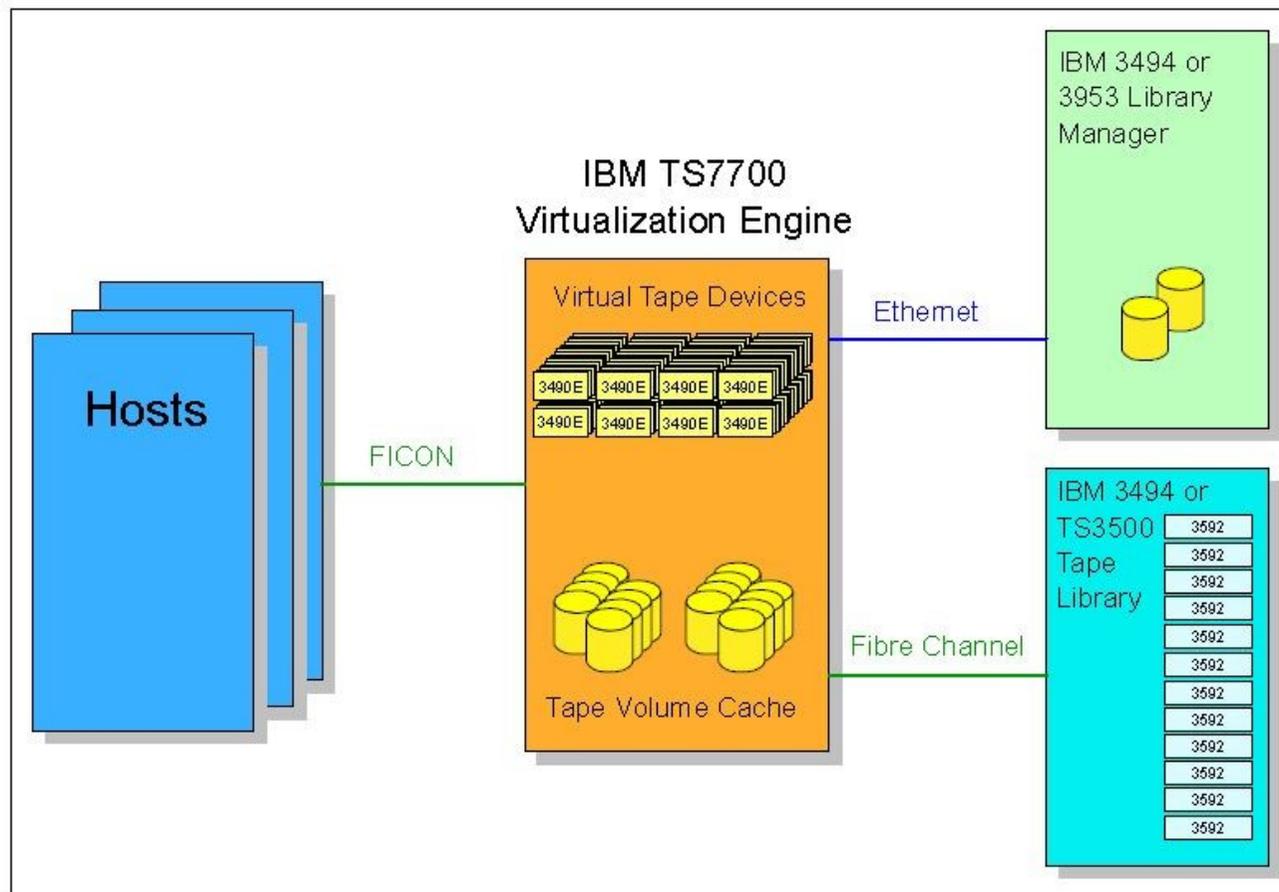
Les unités de stockage bande

Lecteurs physiques partagés

- Partage successif sauf en présence d'un logiciel de gestion spécifique
- Unités réelles affectées par un gestionnaire (VTAPE)

Lecteurs virtuels

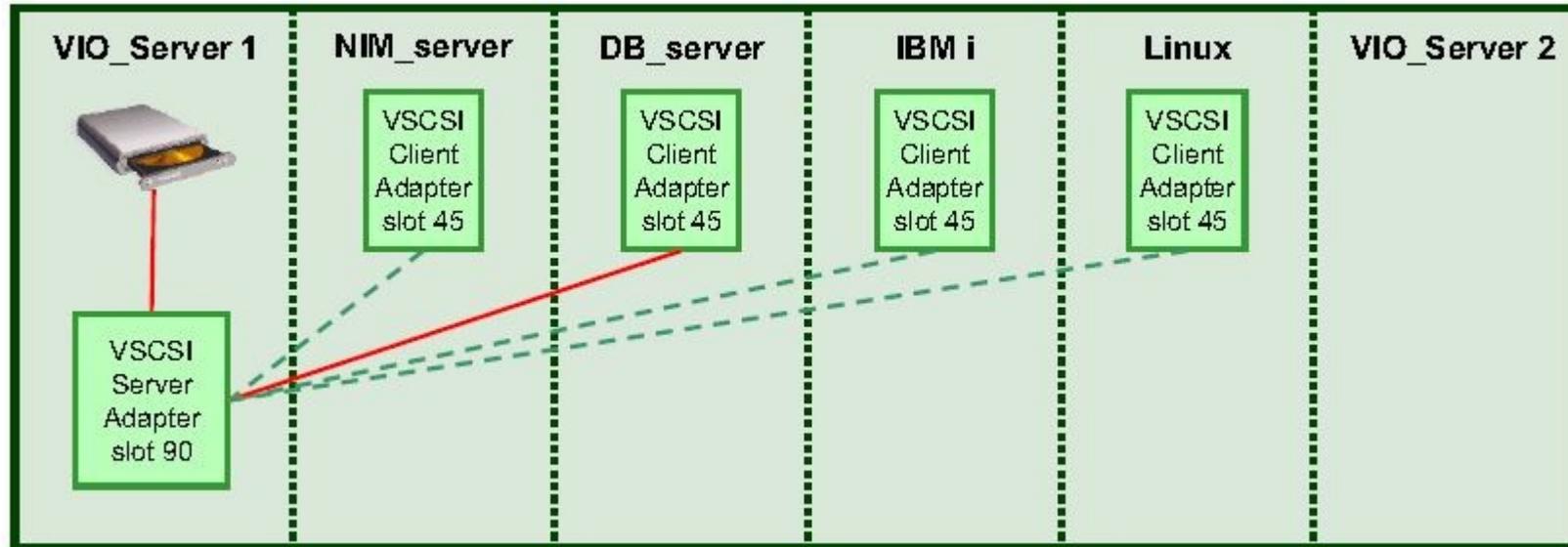
- Application simulant des lecteurs réels et stockage sur disque local ou réseau
- Unité externe simulant des lecteurs (**Virtual Tape Server**)



Les unités de lecture optique

Lecteurs physiques partagés

- accès par une seule partition à la fois



La console système

Nécessité d'émuler :

- ✓ La fonction de « boot » du système hôte dans une machine virtuelle
- ✓ Réinitialisation des interfaces
- ✓ Le choix de l'architecture de la MV
- ✓ La configuration dynamique des unités
- ✓ La reconfiguration des ressources partagées

Solution

- TTY virtuel
- 3270 émulé
- Menus ou interface web

Les accès au système

- A l'origine terminaux dédiés à une machine virtuelle
- Aujourd'hui connectés à travers le réseau en Telnet/SSH ou par des émulateurs
 - TN3270
 - Xterm, rxvt
 - Konsole
 - Tmux
- Une Infrastructure de bureau virtuel (VDI) généralement présente pour faciliter l'accès aux applications (Intel)

Les interfaces de communication zVM

Communication internes

- Canal à canal virtuels
- VMCF, IUCV
- VM Guest LAN
- Hipersockets

Communication externes

- Canaux FICON, Fiber Channel
- Canal à canal
- Cartes OSA Express

Les interfaces de communication réseau

Contraintes

- Partages des interfaces entre les machines virtuelles
 - Cartes réseau spécialisées
 - Gestion des adresses MAC (statiques ou dynamiques)
- Interruptions à très haute fréquence
 - Mode de communication n'impliquant pas l'hyperviseur
 - Partage direct de la mémoire entre la carte et la machine virtuelle
- Switch virtuel intégré à la carte réseau, interne à l'hyperviseur ou géré par une machine virtuelle
 - requiert un gestionnaire

Solutions

- QDIO
- VSWITCH, Hypersockets (système z)
- VLAN

Interfaces de communication Wintel

Réseau

- Partage des interfaces réseau entre plusieurs machines virtuelles
 - ✓ Cartes réseau et BIOS adaptés
 - ✓ Assistance matérielle pour réduire la charge CPU

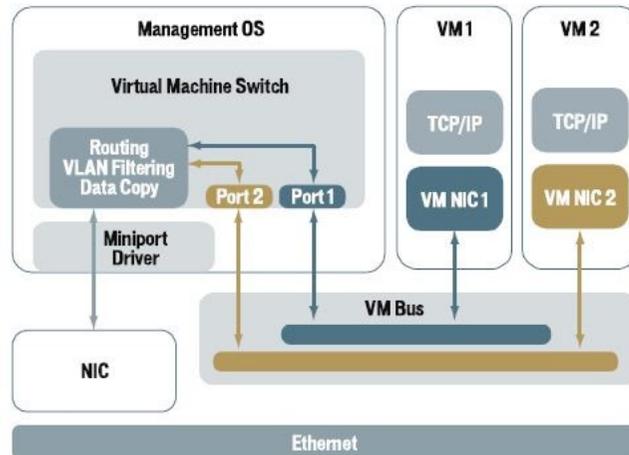
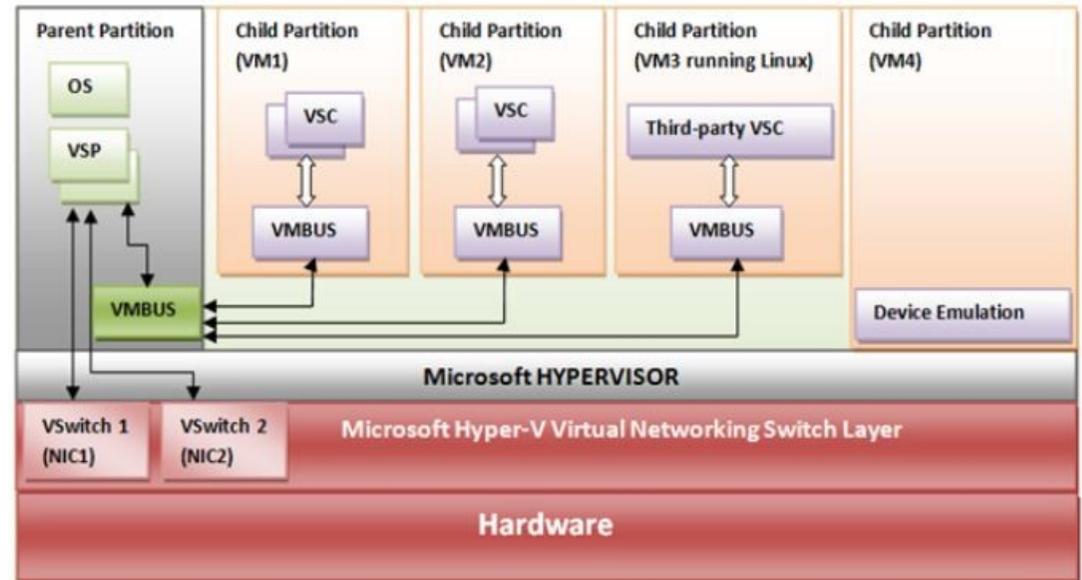


Figure 1. In virtual environments today, the hypervisor's virtual switch manages network I/O. All data passes through the virtual switch in the Management OS.

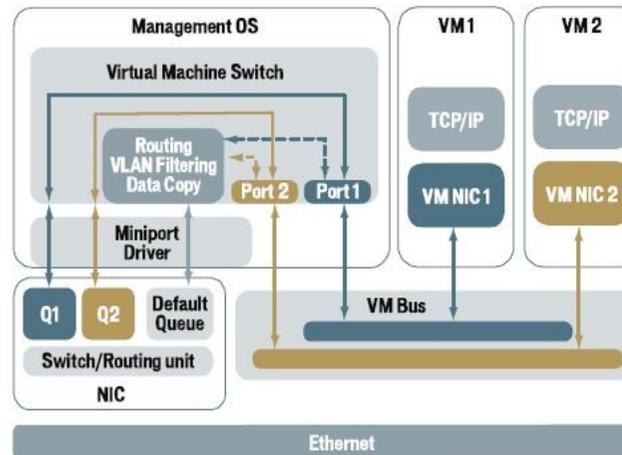


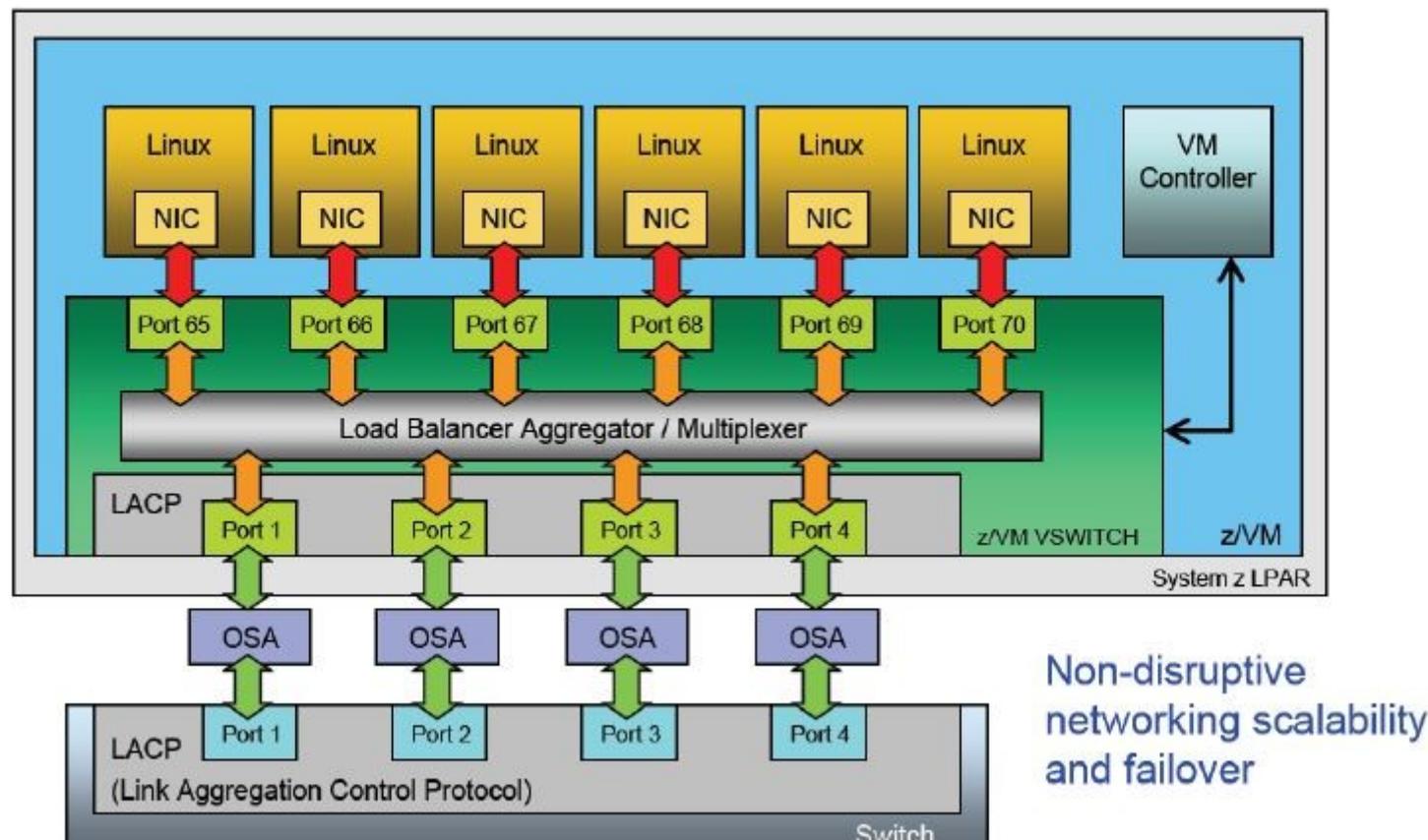
Figure 2. With VMQ and VMDq technologies, data packets are delivered directly to the VM without data copy and bypassing the virtual switch.

Interfaces de communication IBM - VSWITCH

Partage des interfaces réseau entre plusieurs machines virtuelles

- ✓ Cartes réseau OSA standard
- ✓ Transmission directe des paquets IP en second niveau (I/O assist + QDIO)
- ✓ Connexion au réseau local

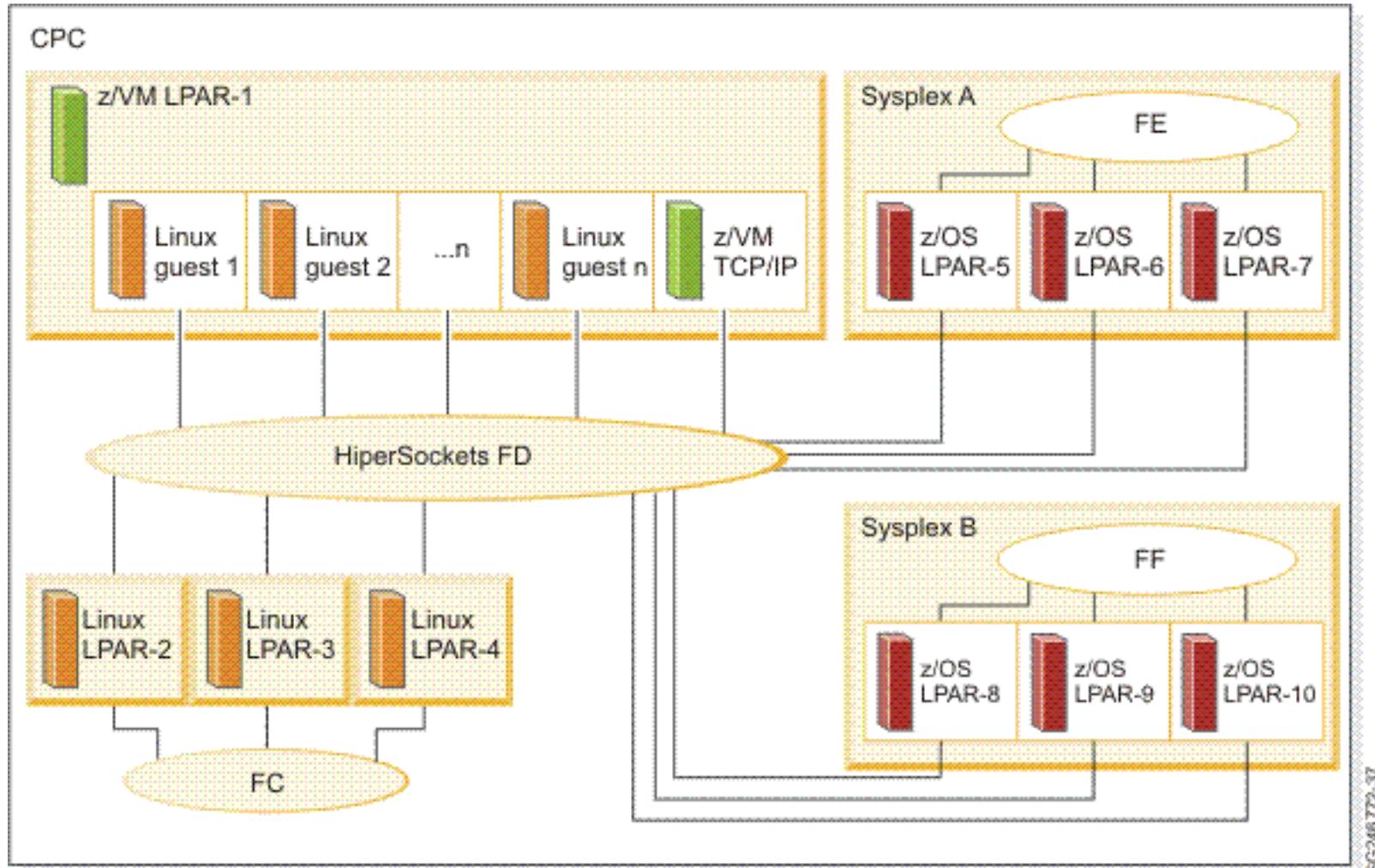
VSWITCH



Interfaces de communication IBM - Hipersockets

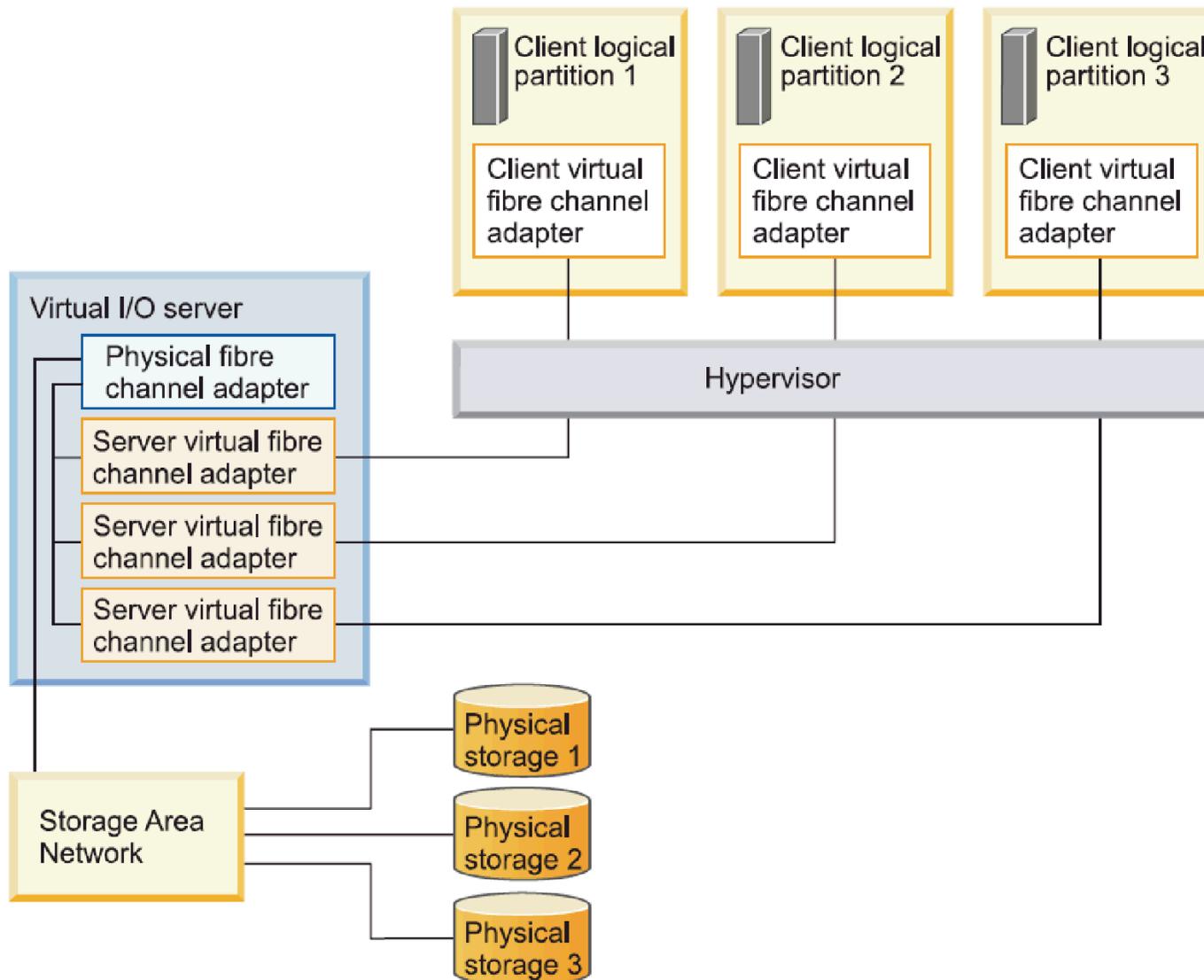
LAN interne au niveau LPAR

- ✓ Pas de carte réseau – équivalent au VM LAN
- ✓ Communication exclusive entre partitions



Interfaces de communication POWER FC

Gestion par un ou deux **Virtual I/O Server**



Les différents timers et horloge

- Horloge (TOD) virtuelle propre à chaque système virtuel
- Gestion de l'ensemble des demandes d'interruption liées au temps

Les systèmes sauvés

Le système d'exploitation sauvegardé est rechargé en mémoire (cas de CMS, GCS)

- ✓ Copie unique pour tous les systèmes utilisant ce système d'exploitation
- ✓ Chargement ultra performant (inférieur à 1 seconde)
- ✓ Utilisation de programmes et de données communs à tous les systèmes identiques

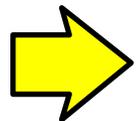
Situation actuelle

Définitions et types de virtualisation

Les premiers systèmes virtuels

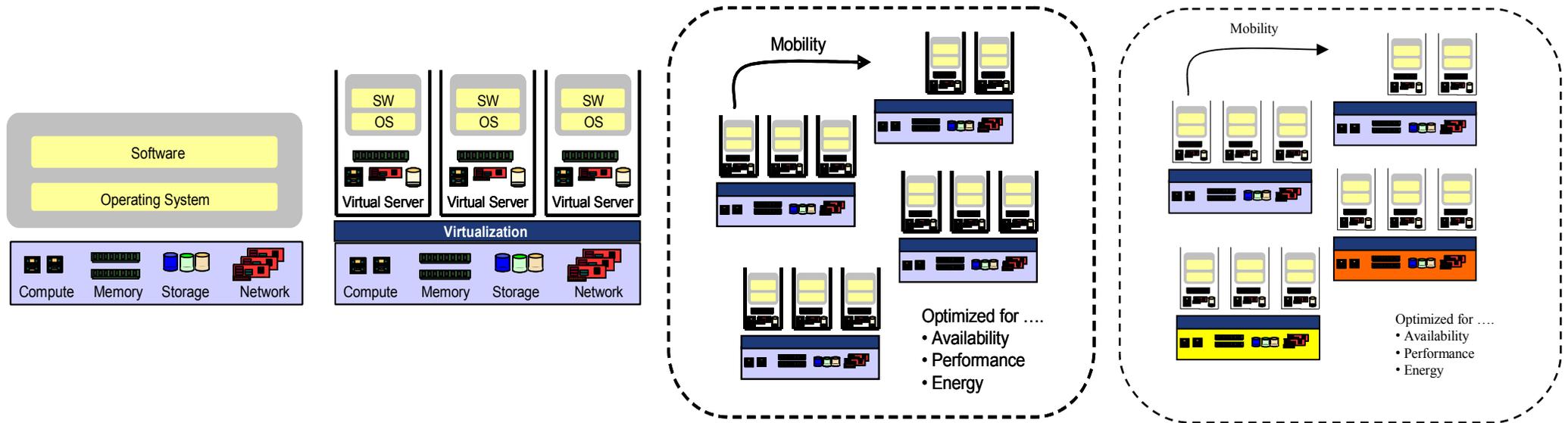
Les besoins

Techniques mises en oeuvre



Situation actuelle

L'évolution de l'architecture des serveurs

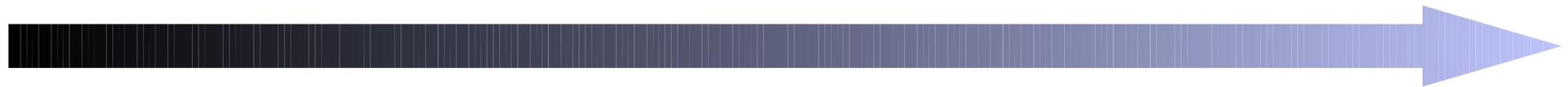


Server

Virtualized Server

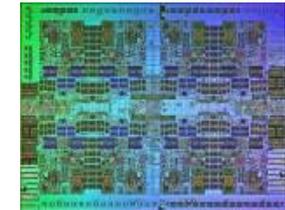
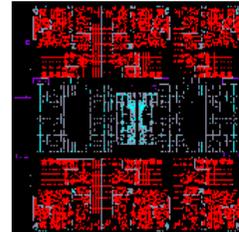
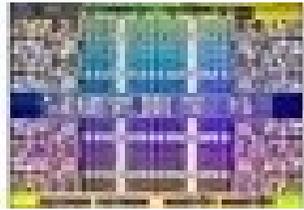
Virtualized Cluster

Virtualized Heterogeneous Cluster



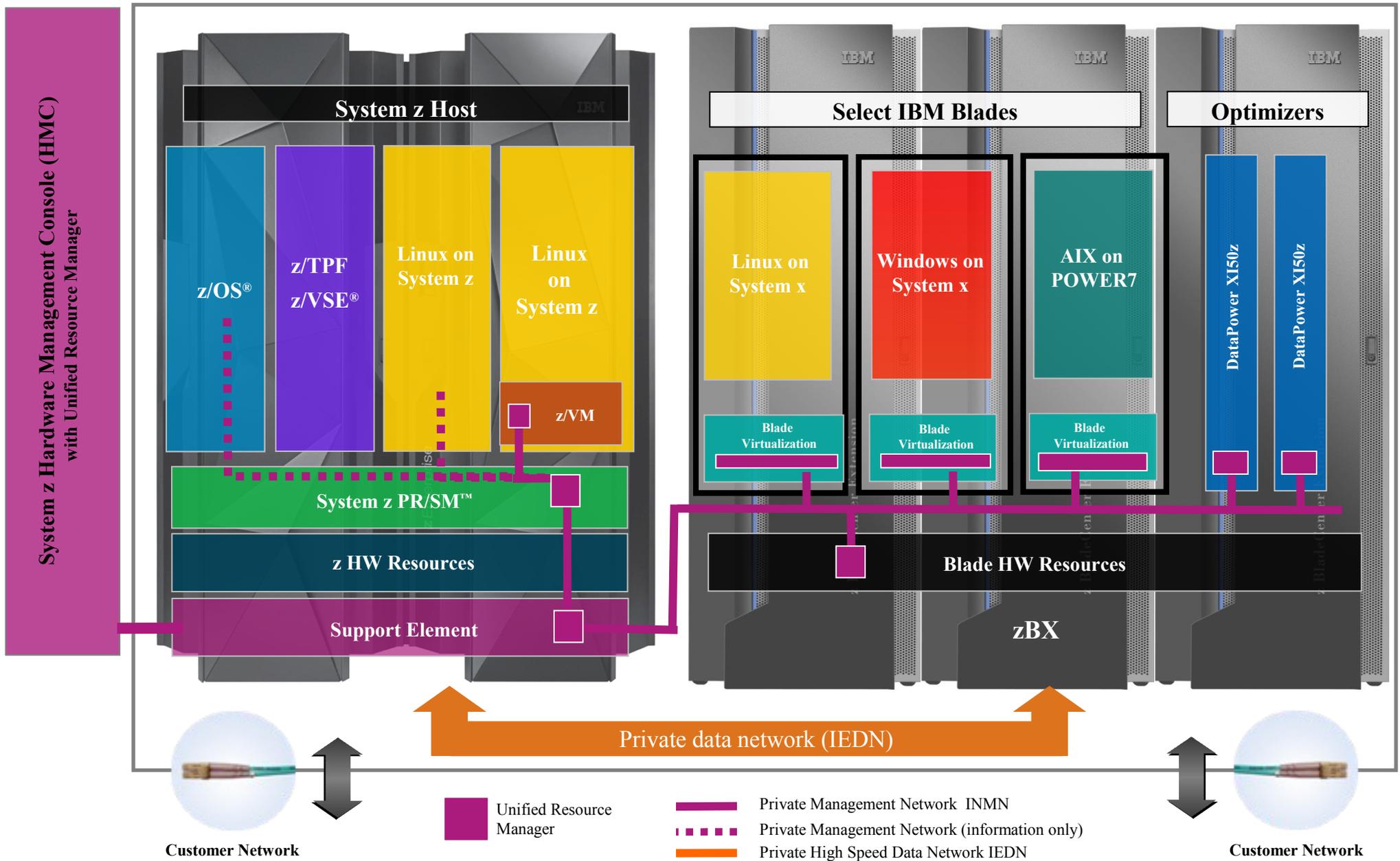
Advanced Management of Virtualized IT Infrastructure

Processeurs



	Intel Westmere EX	IBM zEC12	IBM POWER7
Taille	513 mm ²	598 mm ²	567 mm ²
Transistors	2.6 billion	2.75 billion	1.2 billion
Coeurs	4 / 6 / 8 / 10	6	4 / 6 / 8
Threads par Coeur	2	1	4
Fréquence maximale	3.46 GHz	5.5 GHz	4.25 GHz
Cache L3	24 MB SRAM	48 MB eDRAM	32 MB eDRAM
Scalability	8 Sockets	20 Sockets	32 Sockets

Systeme zEnterprise (hybride)



Hyperviseurs

<i>Facteurs d'évolutivité</i>	VMware ESXi 5 (EMC)	Hyper-V (Microsoft)	XEN (Citrix)	PowerVM (IBM)	z/VM (IBM)
Processeur	Intel	Intel	Intel	Power	zProcessor
Hyperviseur Type	1	2	1	1	1
CPUs virtuels par VM	32	64	16	256	64
Mémoire par VM	96 GB	1 TB	128 GB	8192 GB	16 EB
VMs par server	512	384	512	1000	Nom 8 chars > 1000000
Mémoire par server	2 TB	4 TB	1 TB	8 TB	3 TB
Système d'exploitation	Windows, Linux	Windows, Linux	Windows, Linux, Solaris	AIX, IBM I, Linux	z/OS, z/VM, z/VSE, z/TPF, Linux

Merci. Des questions ?

Dank u
Dutch

Merci
French

Спасибо
Russian

Gracias
Spanish

شكراً
Arabic

감사합니다
Korean

Tack så mycket
Swedish

धन्यवाद
Hindi

תודה רבה
Hebrew

Obrigado
Brazilian
Portuguese

谢谢
Chinese

Dankon
Esperanto

Thank You

ありがとうございます
Japanese

Trugarez
Breton

Danke
German

Tak
Danish

Grazie
Italian

நன்றி
Tamil

děkuji
Czech

ขอบคุณ
Thai

go raibh maith agat
Gaelic