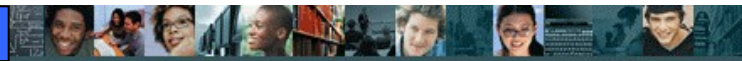
Virtualizzazione nel Mainframe

SISTEMI CENTRALI 2011

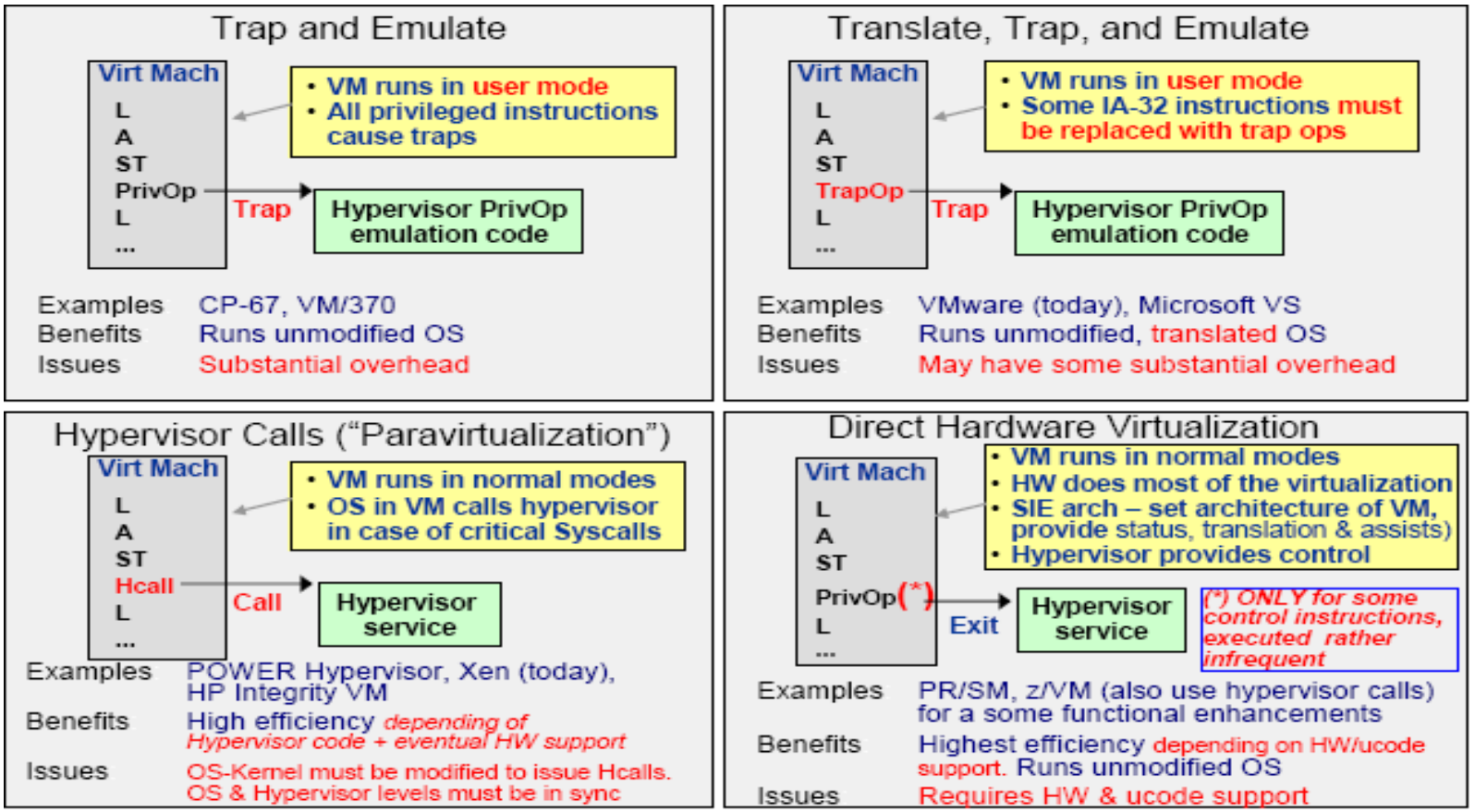
Virtualizzazione

- *PR/SM*
- *z/VM*
- *Virtualizzazione di rete*
  - *Hipersockets*
  - *VLAN*
  - *VSWITCH*

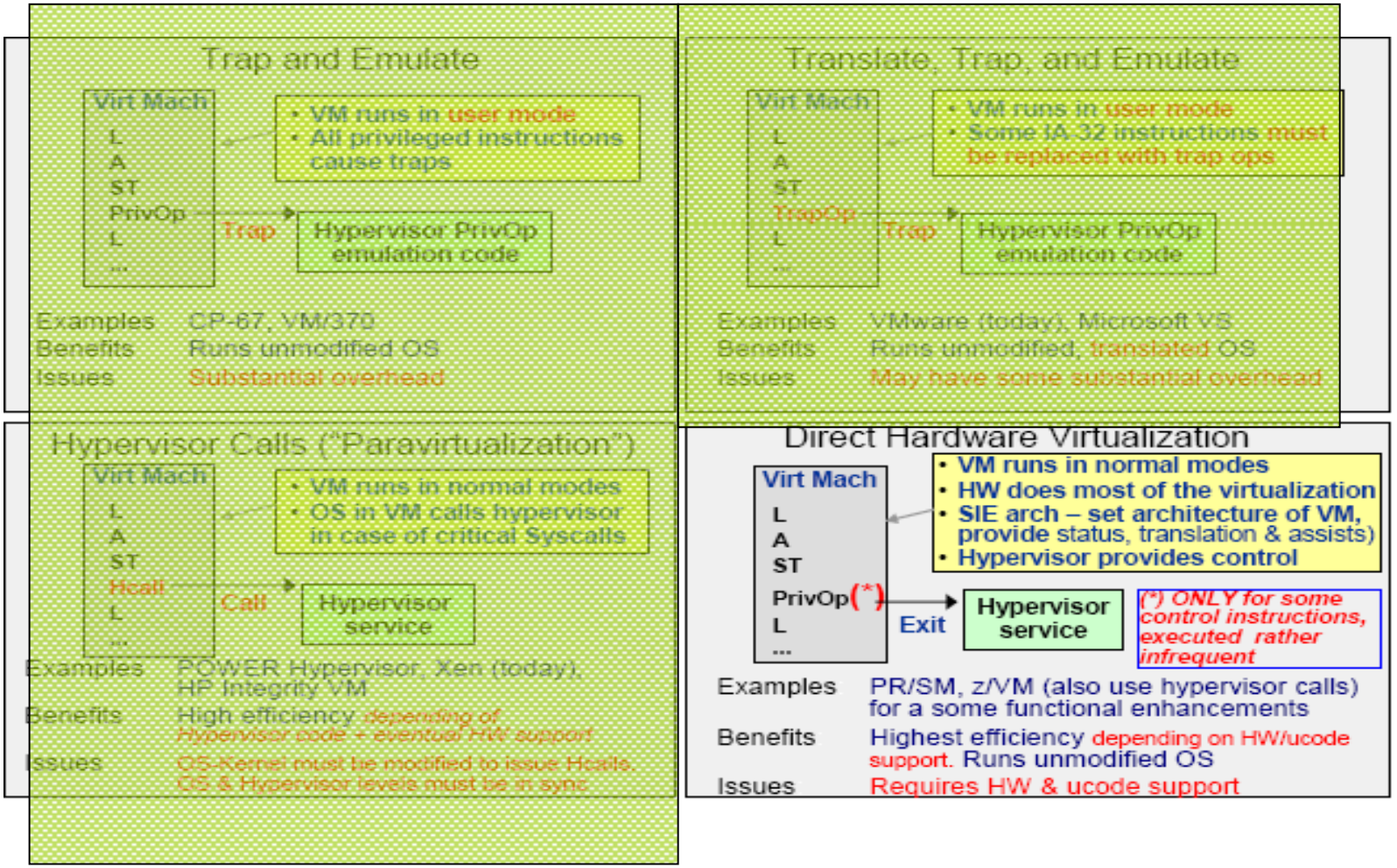
Virtualizzazione e Terminologia

- *Resource Manager = PR/SM = z/VM*
- *Hypervisor = Ipervisore= PR/SM = z/VM*
- *Virtual Machine= Macchina Virtuale = z/VM guest*
- *LPAR= Logical Partition= Partizione*
- *Immagine = S.O. che opera in una LPAR*
- *Mainframe= Central Processor Complex*
- *Server partitioning =Partizionamento*

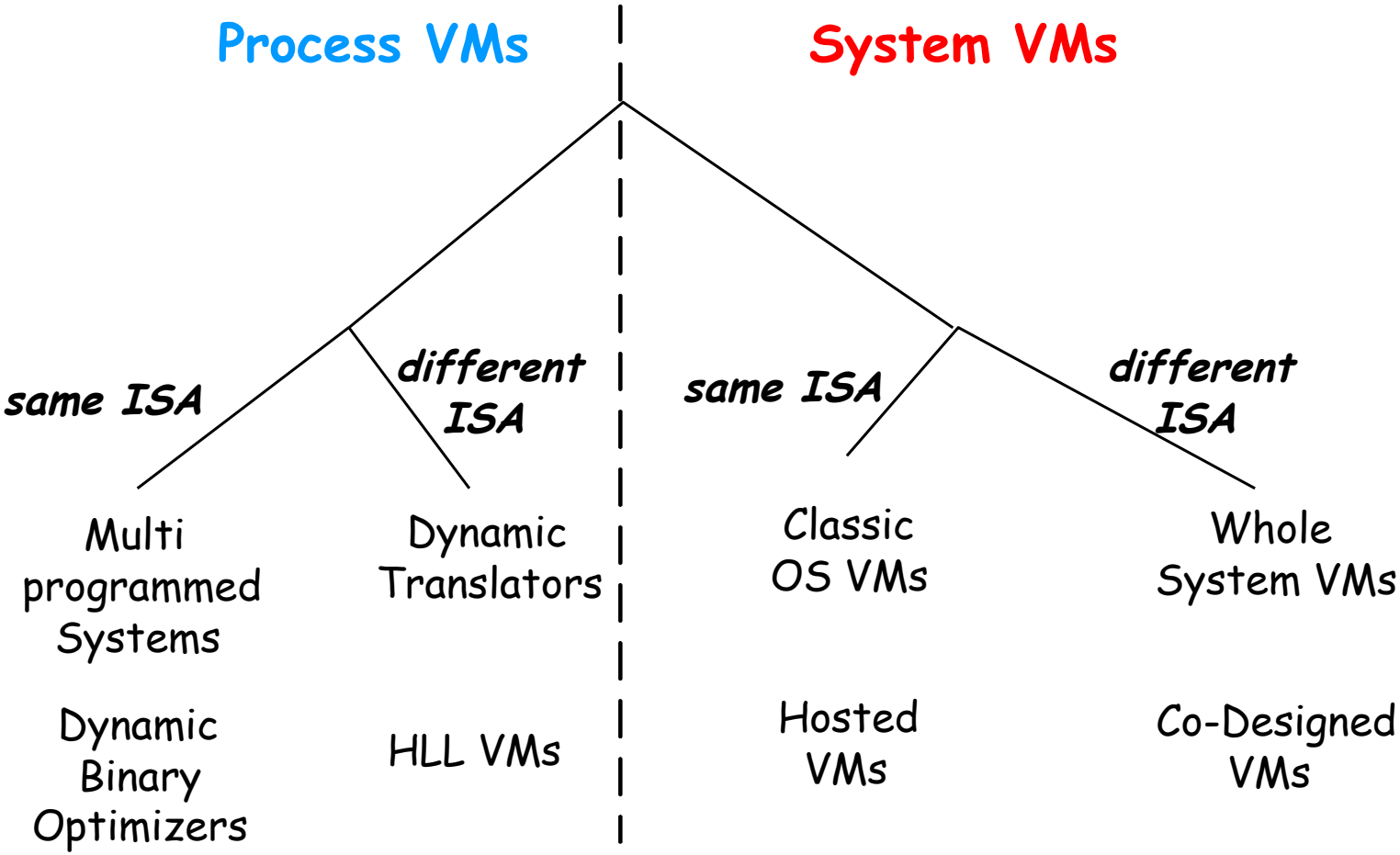
Virtualizzazione e Mainframe



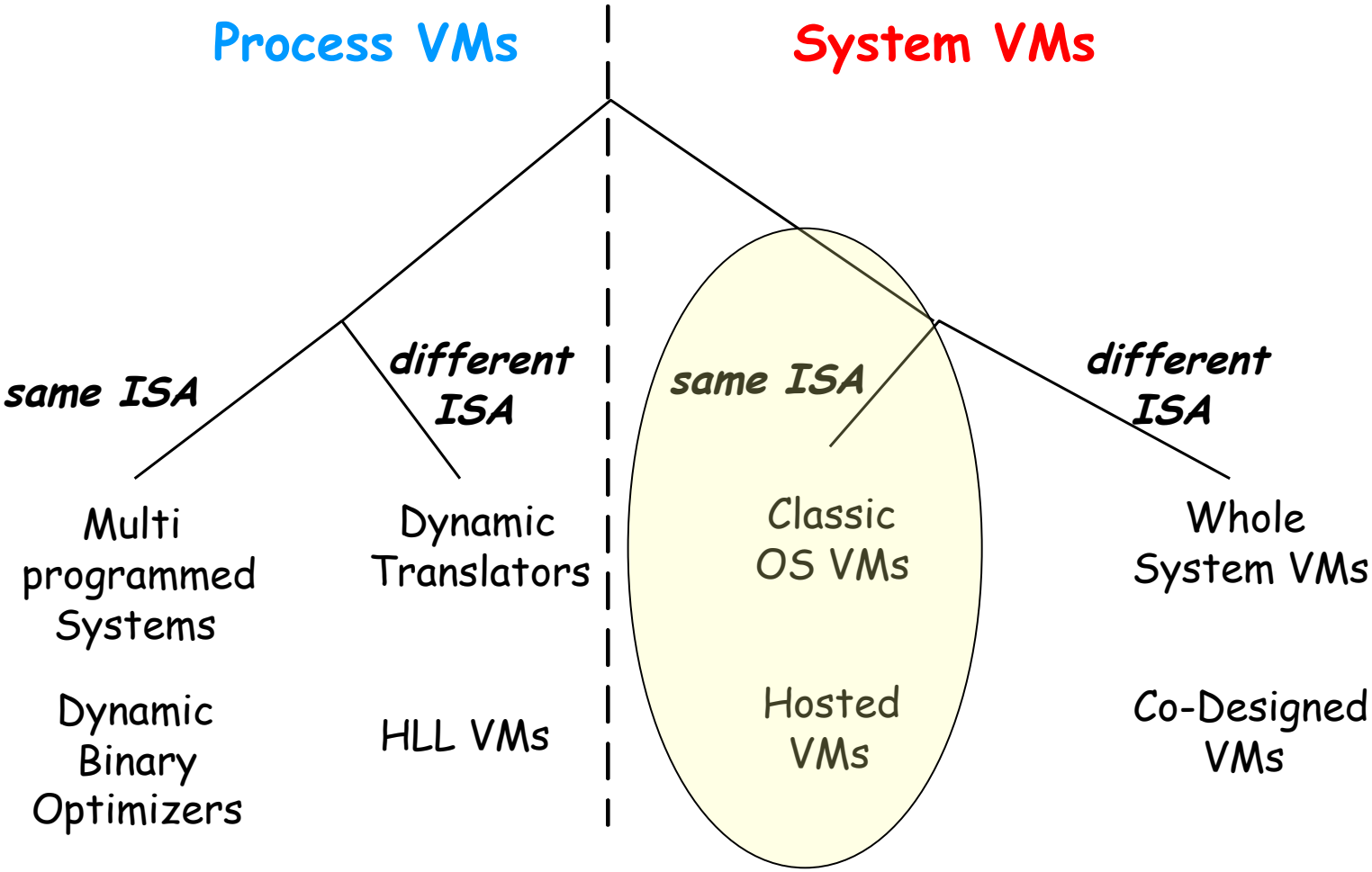
Virtualizzazione e Mainframe



Architetture a Macchine Virtuali

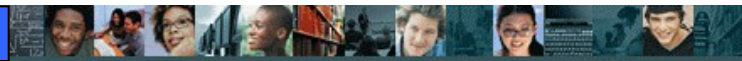


Architetture a Macchine Virtuali



Virtualizzazione (wikipedia)

- Per virtualizzazione si intende la creazione di una versione virtuale di una risorsa normalmente fornita fisicamente. Qualunque risorsa hardware o software può essere virtualizzata: sistemi operativi, server, memoria, spazio disco, sottosistemi. Un tipico esempio di virtualizzazione è la divisione di un disco fisso in partizioni logiche.

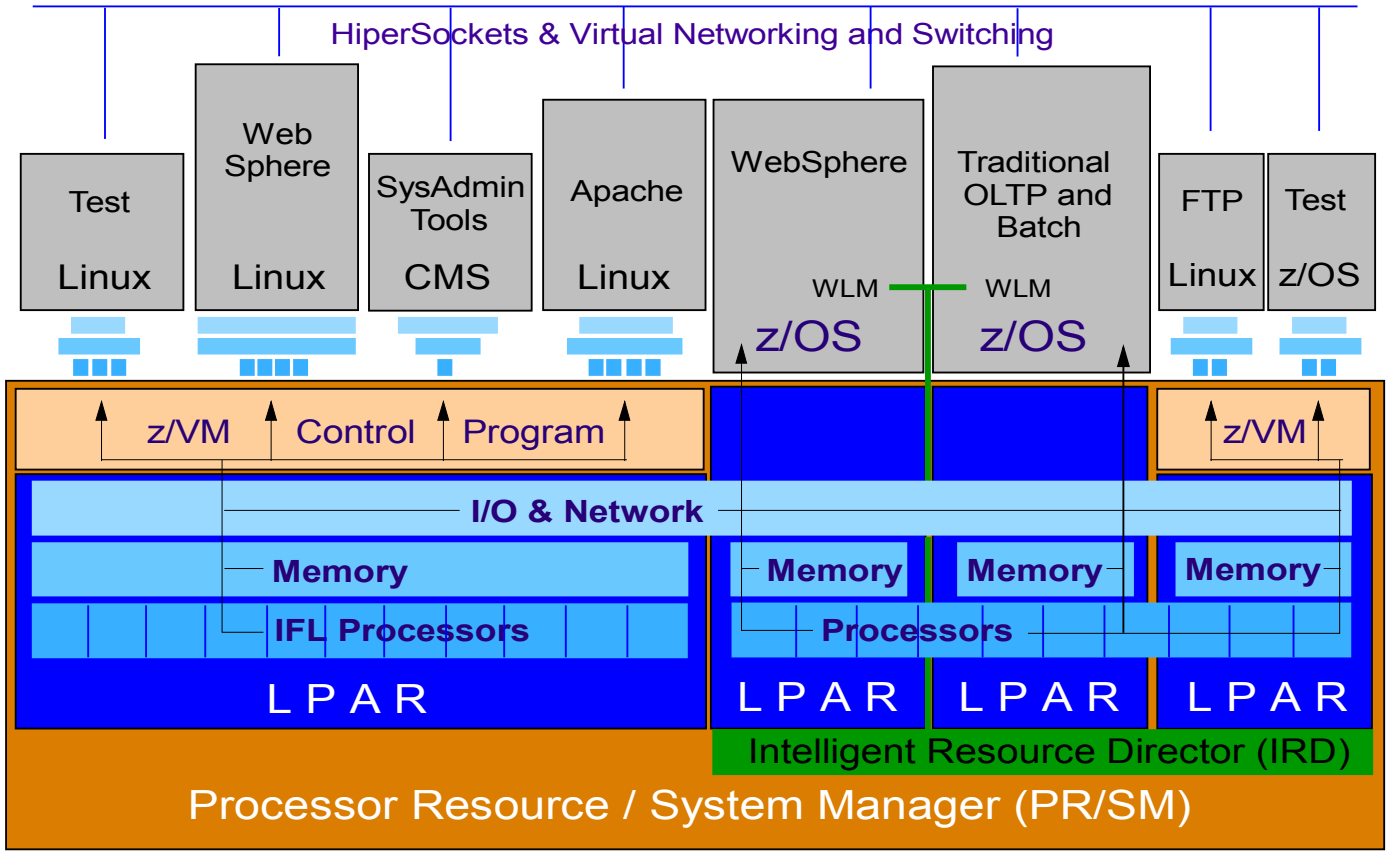



Virtualizzazione e Partizionamento

- **La Virtualizzazione è l**a capacità di creare, attraverso una componente hardware o software, immagini di risorse informatiche virtuali sulla base delle risorse fisiche realmente disponibili delle quali rappresentano un sottoinsieme logico.

- Il **Partizionamento** e' una suddivisione fisica o logica delle risorse di un singolo server in sistemi isolati e indipendenti su cui possono operare S.O. diversi

Virtualizzazione e Mainframe



Virtualizzazione PR/SM e z/VM

- Hardware: sudddivisione fisica del server anche detta Physical partitioning . Non piu' usata dal 1990
- Software : z/VM, commercialmente disponibile dal 1972, supporta Macchine virtuali con tutti sistemi operativi che supportano la z/Architecture (z/OS, zVSE, z/VM, zLinux,zTPF) con sovraccarico di ipervisore minimo e affidabilità ad alto livello
- Logica: Logical Partitioning tramite PR/SM , introdotta nel 1988, che permette il partizionamento di server per un utilizzo ottimale e condiviso delle risorse disponibili

Sia z/VM che PR/SM impiegano tecnologie Hardware e firmware sviluppate negli anni che fanno della virtualizzazione una delle componenti di base del mainframe

PR/SM e Partizionamento di CPU

Il PR/SM e' un ipervisore che opera su mainframe permettendo il suo partizionamento e/o la condivisione di

- CPU
- Memoria
- I/O

Ogni partizione logica (**LPAR**) ha:

- Uno o piu' processori logici dedicati o condivisi
- Una quantità di memoria reale dedicata
- Una connettivita' verso i dispositivi di I/O dedicata o condivisa

Es. LPAR MVS3

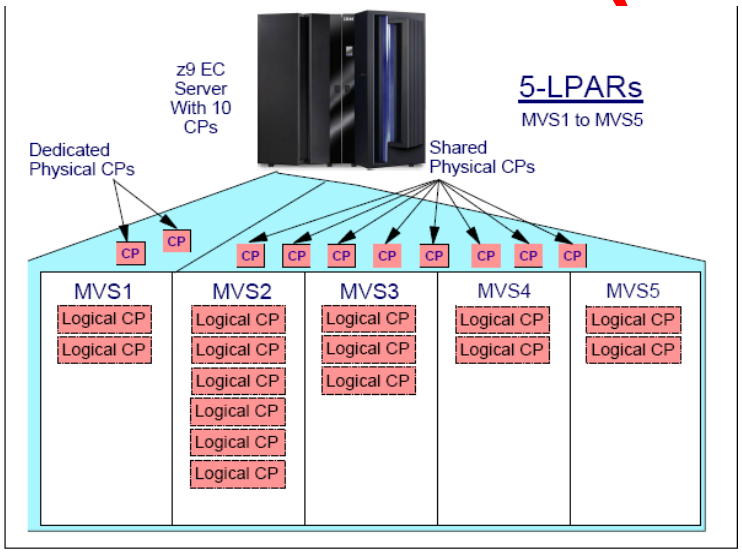
3 Processori shared

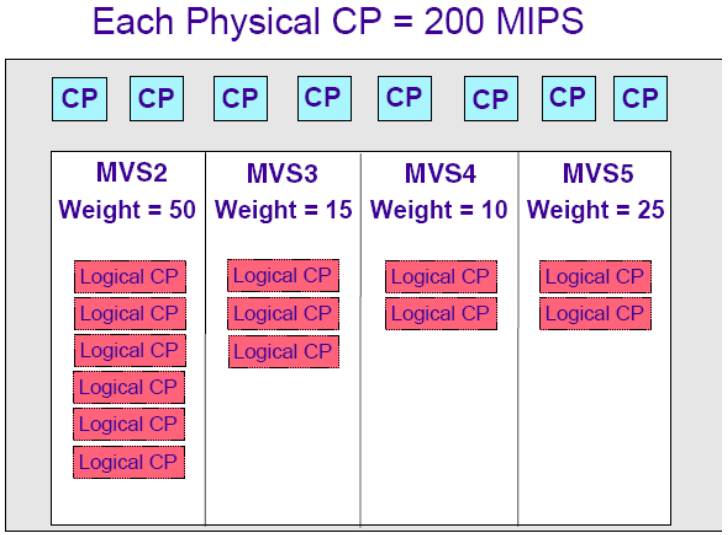
Peso relativo 15%

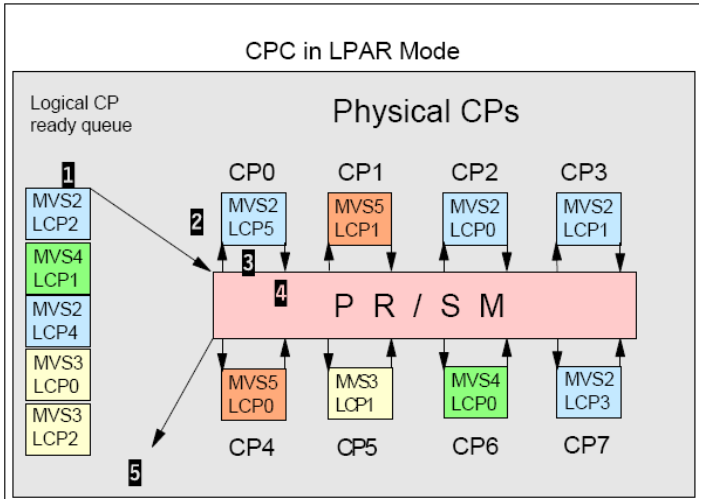
2 GB di Memoria Reale

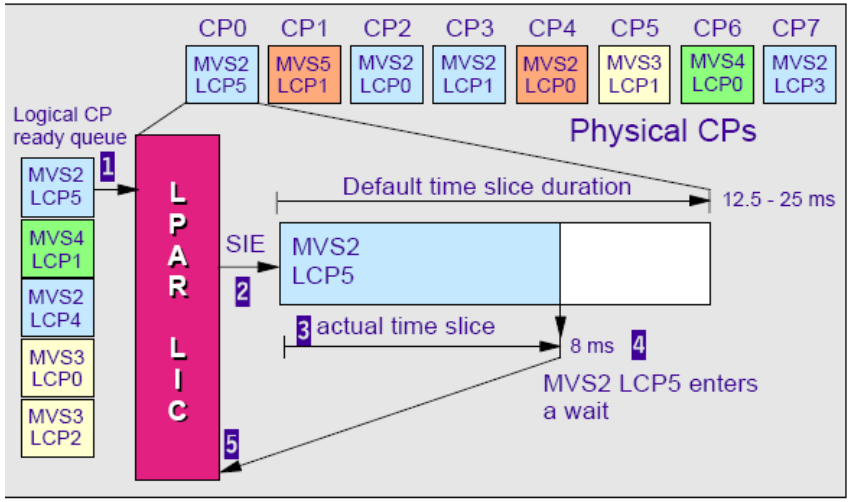
Dispositivi I/O

Virtualizzazione hardware (PR/SM)









PR/SM e Partizionamento di CPU

- Un CP logico continua a eseguire su di un CP fisico fino a che accade uno dei seguenti eventi :

1. Finisce il il suo "Time Slice" (12.5-25ms)
2. Esso entra in uno stato di "CPU wait"
3. Quando consuma CPU al di sopra del suo "peso", ad es .esso e' interrompibile da un I/O per un altro CP logico che sta lavorando meno del peso attribuito
4. Nel caso di z/OS allorche' un processo inizia a stare in "spin waiting" per un evento (es.lock)

- la durata di un Time slice e'
  - » Determinata da opzioni di utente, oppure
  - » Determinata dinamicamente dalla LPAR

$$\text{time slice} = \frac{25\ \text{ms} * \text{number of shared CPs}}{\text{number of LPs started}}$$

Interpretive Execution

- Sia PR/SM che z/VM usano l'istruzione privilegiata START INTERPRETIVE EXECUTION (SIE) per eseguire partizioni logiche o macchine virtuali
- Il programma che esegue la SIE e' chiamato Host
- Il programma che lavora sotto il controllo della SIE e' chiamato guest o sistema ospite
- L'operando dell' istruzione SIE e' un descrittore di stato che descrive il sistema ospite (L'istruzione SIE e' eseguita dall'Host)

  PSWs
  CPU timer e Clock comparator
  Control registers
  General registers 14 e 15
  Prefix register

- In z/VM che gira in una partizione e' supportato un livello di nesting (SIE sotto SIE)

Responsabilità del programma ospite

Compito del programma ospite e:'

- All'esecuzione della SIE (SIE entry)
  - Caricare i General register 0-13 del sistemea ospite (guest)
  - Caricare i FP registers e i Control FP registers
  - Caricare agli access register
- Alla Fine della SIE (SIE exit)
  - Gestire l'intercettazione (intercept)

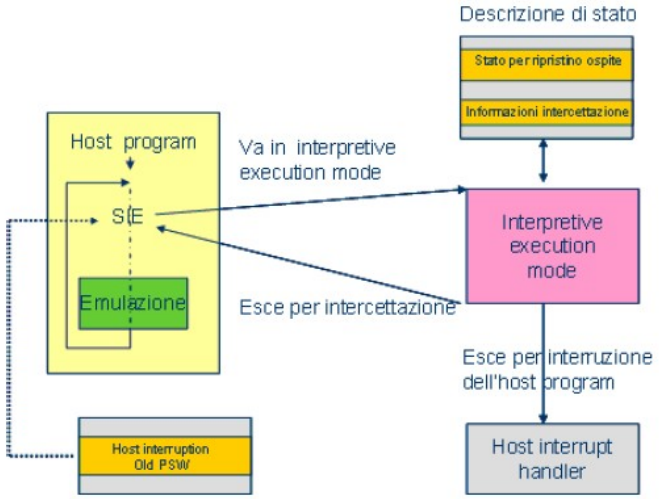
Condizioni di Uscita della SIE

- Una intercettazione

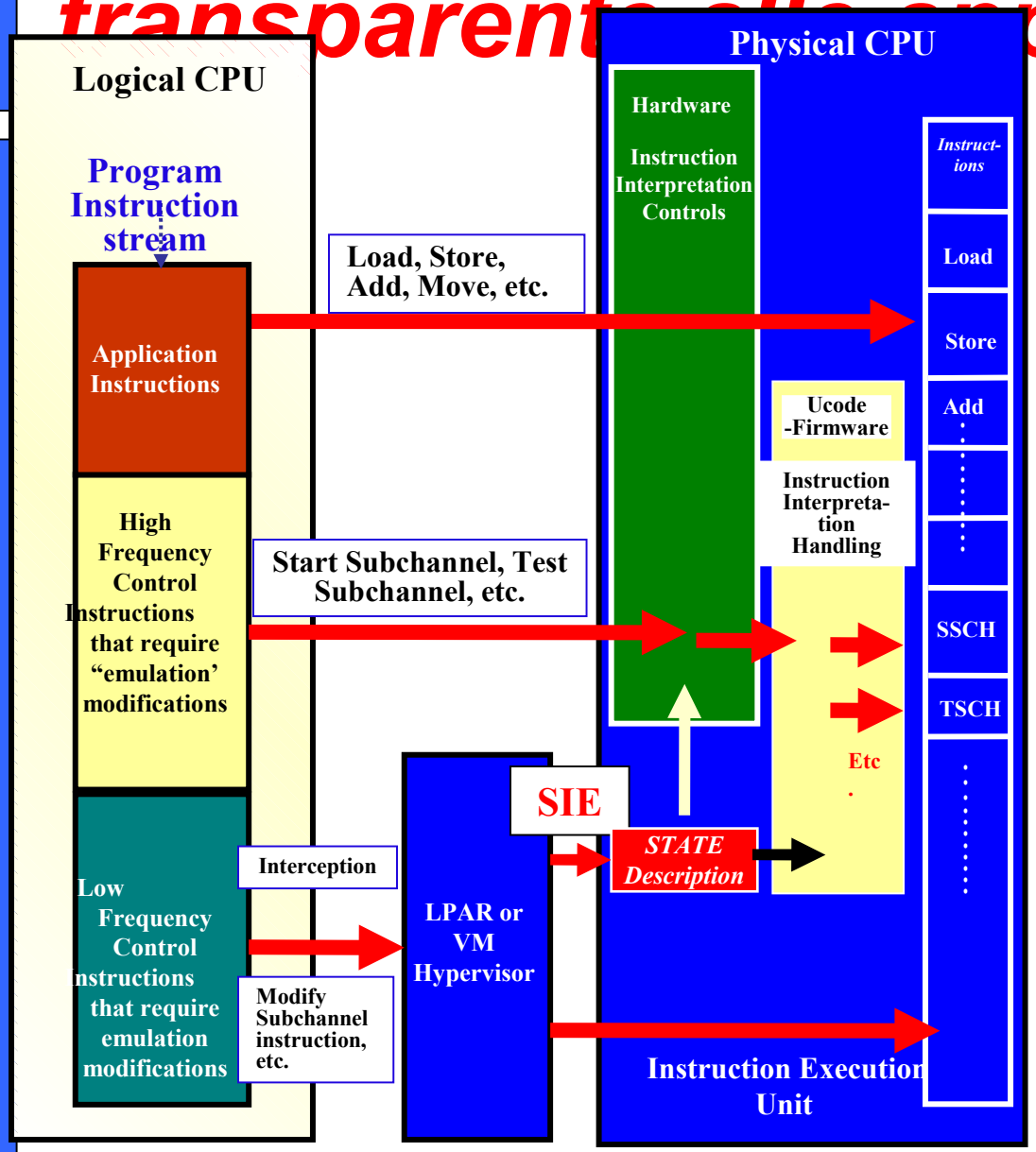
  Il descrittore di stato è aggiornato , un codice di intercettazione viene memorizzato e l'Host program PR/SM o z/VM riassume il controllo dall'istruzione successiva alla SIE

- Un Interruzione dell Host

  Ad Es. Un "external Interrupt" o un "I/O interrupt" o un "translation exception" In questo caso l'intera operazione e' annullata la old PSW punta alla istruzione SIE . Il codice di interruzione è memorizzato



Direct HW virtualization – *transparente alle applicazioni/OS*



System z with SIE

(Start Interpretive Instruction Execution)

SIE e Memoria

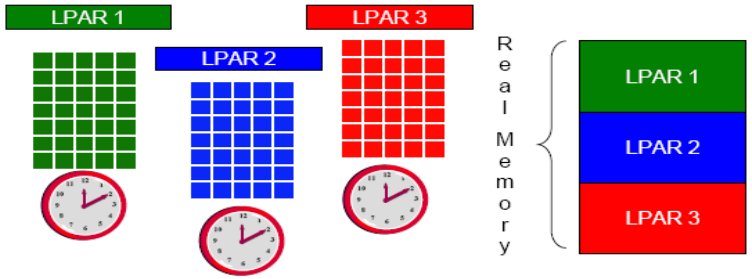
La SIE ha due possibili modalità di trattamento della memoria

1. Memoria fissa e contigua
2. Memoria Virtuale con Dynamic Address Translation

Il PR/SM si comporta nei confronti dei suoi ospiti (LPARs) con la prima modalità

Interpretive Execution Facility - Logical partitions

- PR/SM does for LPARs what CP does for virtual machines, but storage is fixed and contiguous



SIE e Zone Relocation

Zone relocation

- SIE fornisce regioni multiple di memoria con origine "0" in un sistema.

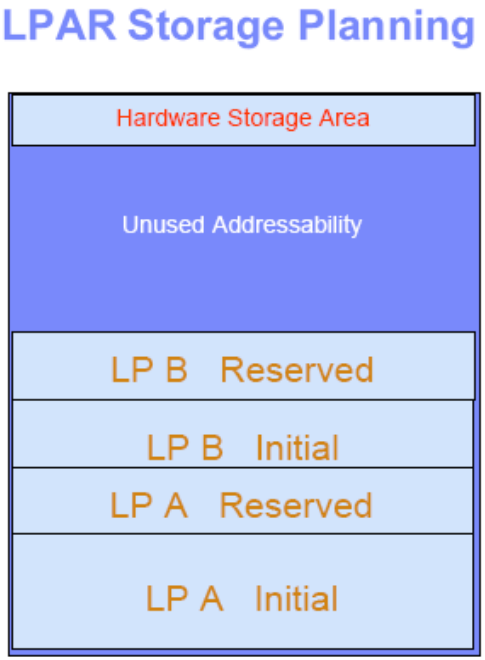
- La definizione della partizione stabilisce il limite inferiore e superiore della memoria fisica ad essa assegnata (zone). Su tale zona viene rimappata con un registro dedicato per creare una memoria reale con origine "0" (Zone Relocation)

- Abilita il Sottosistema di I/O (Canale) ad accedere direttamente la memoria della partizione senza richiesta dell'Ipervisore

PR/SM e Partizionamento di Memoria

Memoria della Partizione

- Allocazione Iniziale e Allocazione riservata
- Configurabile dinamicamente
- Alcune complessità/rigidità
- Memoria non condivisibile



PR/SM e partizionamento di I/O

- Canali Riconfigurabili I Canali dedicati ad una partizione possono essere riconfigurati ad un'altra partizione con dei comandi

- Il MIF (Multi Image Facility)
  - Canali di I/O possono essere definiti e condivisi tra varie partizioni (ESCON , FICON, OSA; FCP;...)
  - Dispositivi di I/O su canali condivisi possono essere acceduti contemporaneamente dalle partizioni che li condividono (sharing partition)
  - Dispositivi di I/O su canali condivisi possono essere confinati all'uso di un sottosistema di partizioni che li condividono acceduti contemporaneamente dalle partizioni che li condividono (sharing partition)

- Partizionamento dei Dispositivi di I/O Dispositivi di I/O possono essere ripartiti tra varie partizioni (es. zoning, ..)

PR/SM e partizionamento di I/O

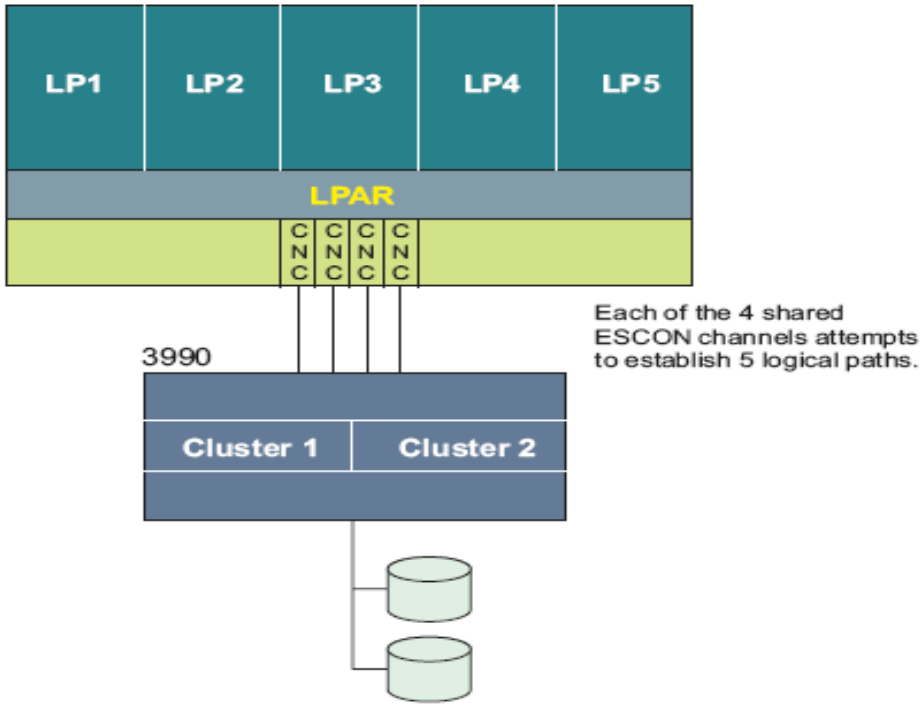


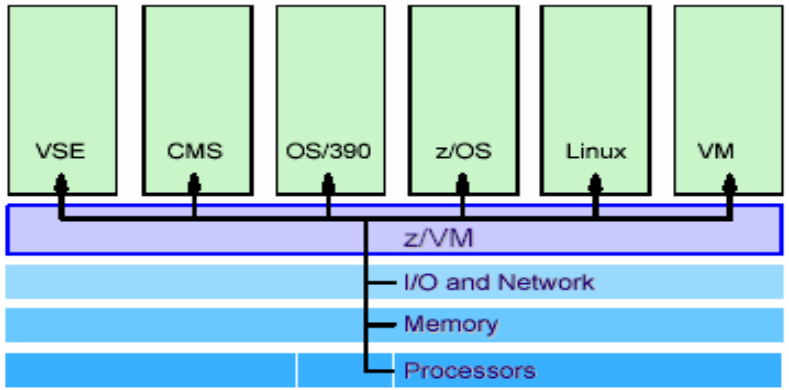
Figure 2-2. A Shared ESCON Configuration that Can Benefit from Better Logical Path Management

L'Ipervisore z/VM

- z/VM (z/Virtual Machine) consente di gestire un sistema a macchine virtuali utilizzando risorse hardware e virtualizzandole

- z/VM supporta contemporaneamente molte macchine virtuali diverse, ognuna delle quali viene eseguita nel proprio ambiente operativo(OS "guest") protetto da funzionalità di sicurezza e isolamento.

- z/VM può sovraimpegnare le risorse reali e consentire agli utenti di creare un set di macchine virtuali le cui risorse eccedono in larga misura le potenzialitàdell'hardware reale



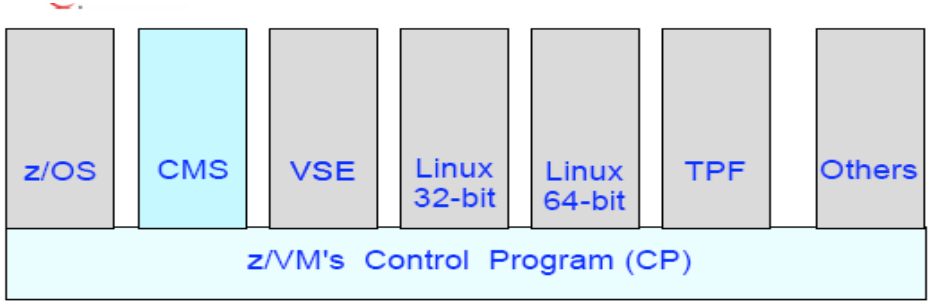
Le componenti di base del z/VM

- **Control Program (CP)** ovvero il cuore del z/VM'.

Esso rappresenta un supervisore di tutte le attività di controllo e gestione delle attività del z/VM (resource-sharing, device management, dispatching, virtual storage management, ed altro) con particolare attenzione alla creazione e gestione delle macchine virtuali.

- **Conversational Monitor System (CMS)**

E' un Sistema Operativo Monoutente utilizzato come ambiente interattivo dotato di un suo file system, di servizi di programmazione, accesso ai dispositivi di I/O ,è dotato di una interfaccia verso il CP per dare comandi diretti



Le componenti di base del z/VM

- **z/VM System Directory.**

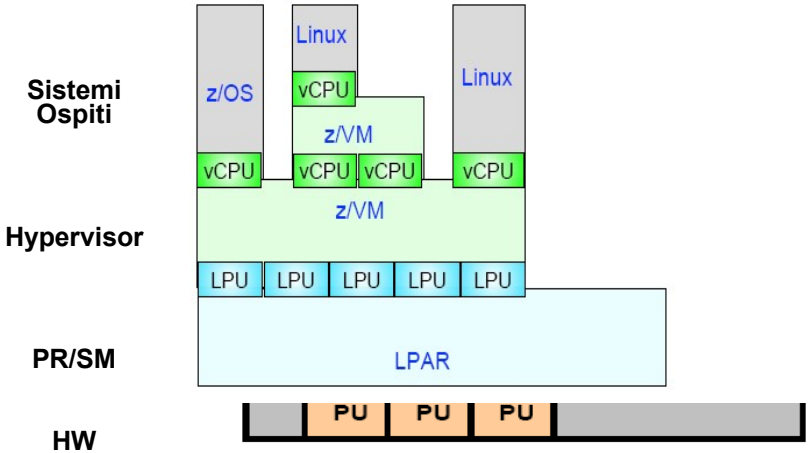
Tabella che ospita le definizioni delle Macchine Virtuali . Essa include:

- Virtual machine identifier (user ID) e password
- POSIX user ID (UID) e group ID (gid)
- Classi di privilegio assegnate
- Dimensioni di memoria virtuale (virtual storage) iniziale e massima
- Dispositivi di I/O reali dedicati, simulati, virtuali (minidisk)
- Permessi all uso di funzioni speciali di Control Program

```
USER LINUX01       MYPASS 128M 128M  G
MACHINE ESA
IPL 190 PARM AUTOCR
CONSOLE 01F 3270 A
SPOOL   00C 2540 READER *
SPOOL   00D 2540 PUNCH  A
SPOOL   00E 1403 A
MDISK   191 3390 012 001 ONEBIT MW
MDISK   200 3390 050 100 TWOBIT MR
LINK    MAINT 190 190 RR
LINK    MAINT 19D 19D RR
LINK    MAINT 19E 19E RR
```

z/VM e Partizionamento di CPU

- Virtual Machine CPU anche detti Processori Virtuali (Virtual Processor)
- Da 1 a 64 processori virtuali dedicati o condivisi
- La condivisione (virtual CPU sharing) può essere assoluta o relativa eventualmente con hard o soft Capping
- I processori virtuali possono sovrabbondare i processori Logici (non consigliato)



VCPU = CPU Virtuale ovvero cio' che la macchina virtuale intende per processore

LPU = Processor Unit Logica (virtuale) ovvero cio' che il sistema operativo (di primo livello) intende per processore

PU = Processor Unit Reale ovvero il processore reale

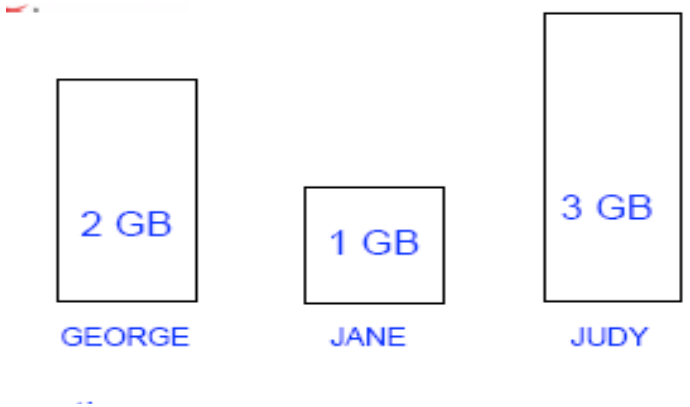
Gestione dei processori

- VM
  - Scheduler determina le priorità in base alle caratteristiche di condivisione prescelte
  - Dispatcher esegue un processore su un processore reale (logico) per un certo time slice
  - I processori virtuali sono in competizione per un time slice a meno che non siano dedicati
- PR/SM
  - Lo scheduler usa i Pesi per eseguire un processore logico su un processore fisico
  - LPARs possono avere Processori fisici dedicati

z/VM e Memoria

Nella CP directory di z/VM definisco la quantità si memoria "reale" di ogni macchina virtuale

```
USER LINUX01       MYPASS 128M 128M  G
MACHINE ESA
IPL 190 PARM AUTOCR
CONSOLE 01F 3270 A
SPOOL   00C 2540 READER *
SPOOL   00D 2540 PUNCH  A
SPOOL   00E 1403 A
MDISK   191 3390 012 001 ONEBIT MW
MDISK   200 3390 050 100 TWOBIT MR
LINK    MAINT 190 190 RR
LINK    MAINT 19D 19D RR
LINK    MAINT 19E 19E RR
```



SIE e Memoria

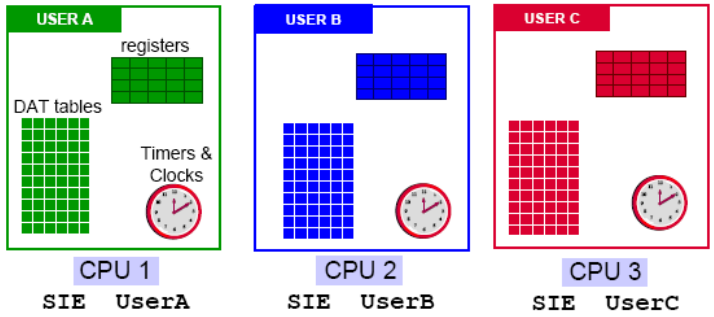
La SIE ha due possibili modalità di trattamento della memoria

1. Memoria utilizza memoria fissa e contigua
2. Memoria Virtuale con Dynamic Address Translation

z/VM si comporta nei confronti dei suoi ospiti (macchine virtuali) con la seconda modalità

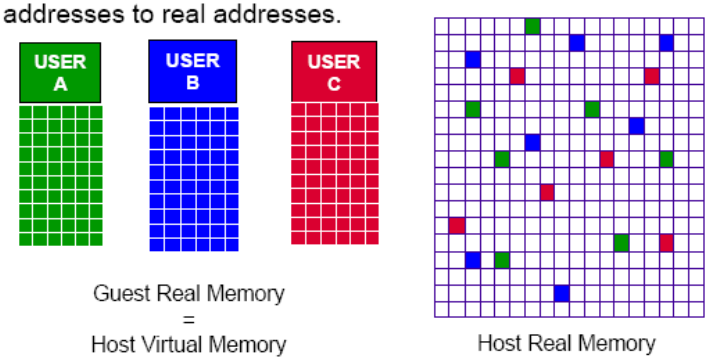
Interpretive Execution Facility – Virtual machines

- CP maintains for each virtual CPU a description of the execution environment for that user



Virtual Memory

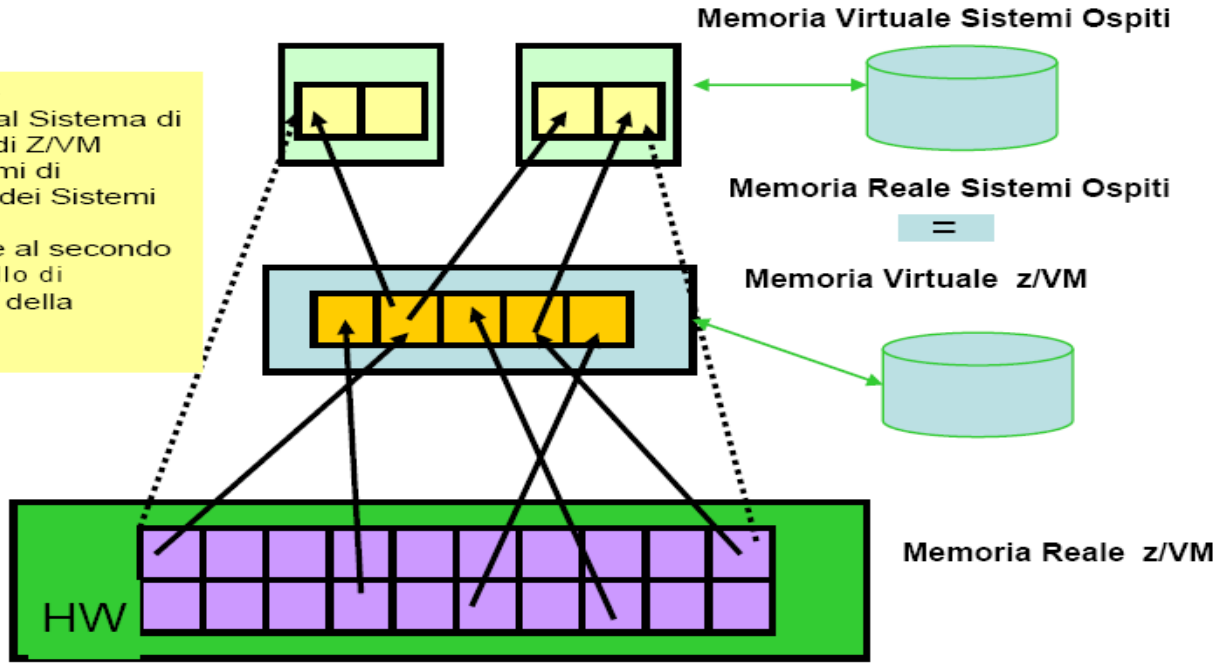
- SIE uses dynamic address translation to convert virtual addresses to real addresses.



z/VM e Memoria

La memoria in ambiente virtuale necessita di molti livelli di traduzione degli indirizzi

La memoria Reale infatti viene virtualizzata una prima volta dal Sistema di Paginazione eVirtualizzazione di Z/VM (Gestito dal CP)e poi dai Sistemi di Paginazione e Virtualizzazionedei Sistemi Ospiti.
Se il Z/VM viene attivato anche al secondo livello ci sara' un ulteriore livello di Paginazione e Virtualizzazione della Memoria



Shared Segments

- La tecnica “Shared segments”

  - permette a macchine virtuali multiple di condividere la stessa area di memoria per gestire piu efficacemente la sua memoria riducendo drammaticamente il numero di page frames duplicate

  - possono essere read-only o blocchi read-write di memoria contenenti codice o dati di molte macchine virtuali. Ad esempio il kernel code readonly di zLinux .

  - Piu’ di una segment table di macchine virtuali punterà alle stesse page frames in memoria reale

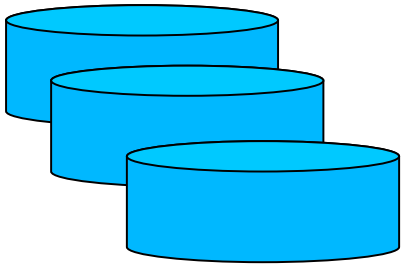
  - Il CP offre una protezione della memoria per non far alterare senza il suo assenso i segmenti salvati dichiarati in read-only.

  - z/VM permette ad una sola macchina virtuale di definire e salvare shared-segments

z/VM e I/O

Nella CP directory di z/VM definisco le risorse di I/O di ogni macchina virtuale

```
USER LINUX01       MYPASS 128M 128M  G
MACHINE ESA
IPL 190 PARM AUTOCR
CONSOLE 01F 3270 A
SPOOL   00C 2540 READER *
SPOOL   00D 2540 PUNCH  A
SPOOL   00E 1403 A
MDISK   191 3390 012 001 ONEBIT MW
MDISK   200 3390 050 100 TWOBIT MR
LINK    MAINT 190 190 RR
LINK    MAINT 19D 19D RR
LINK    MAINT 19E 19E RR
```



z/VM e I/O

- Dispositivi (device) dedicati o condivisi
- Riconfigura a livello device
- I/O throttling (tecnica per evitare di "strozzare" il sistema diminuisce la elaborazione di I/O in memoria)
- Device Virtuali
- Minidisk cache
- Nastri Virtuali in memoria
- Spooled Device

Virtual Device

I device reali e virtuali possono essere

- condivisi (ad Dischi) da piu' macchine virtuali contemporaneamente
- esclusivi ad esempio Nastri (ad uso esclusivo)

- Il CP intercetta l'operazione di I/O verso device virtuali

- Fa una assegnazione dei device virtuali agli opportuni device reali
-
- fa operazioni di traduzione (i comandi CCW devono avere indirizzi reali) e validazione

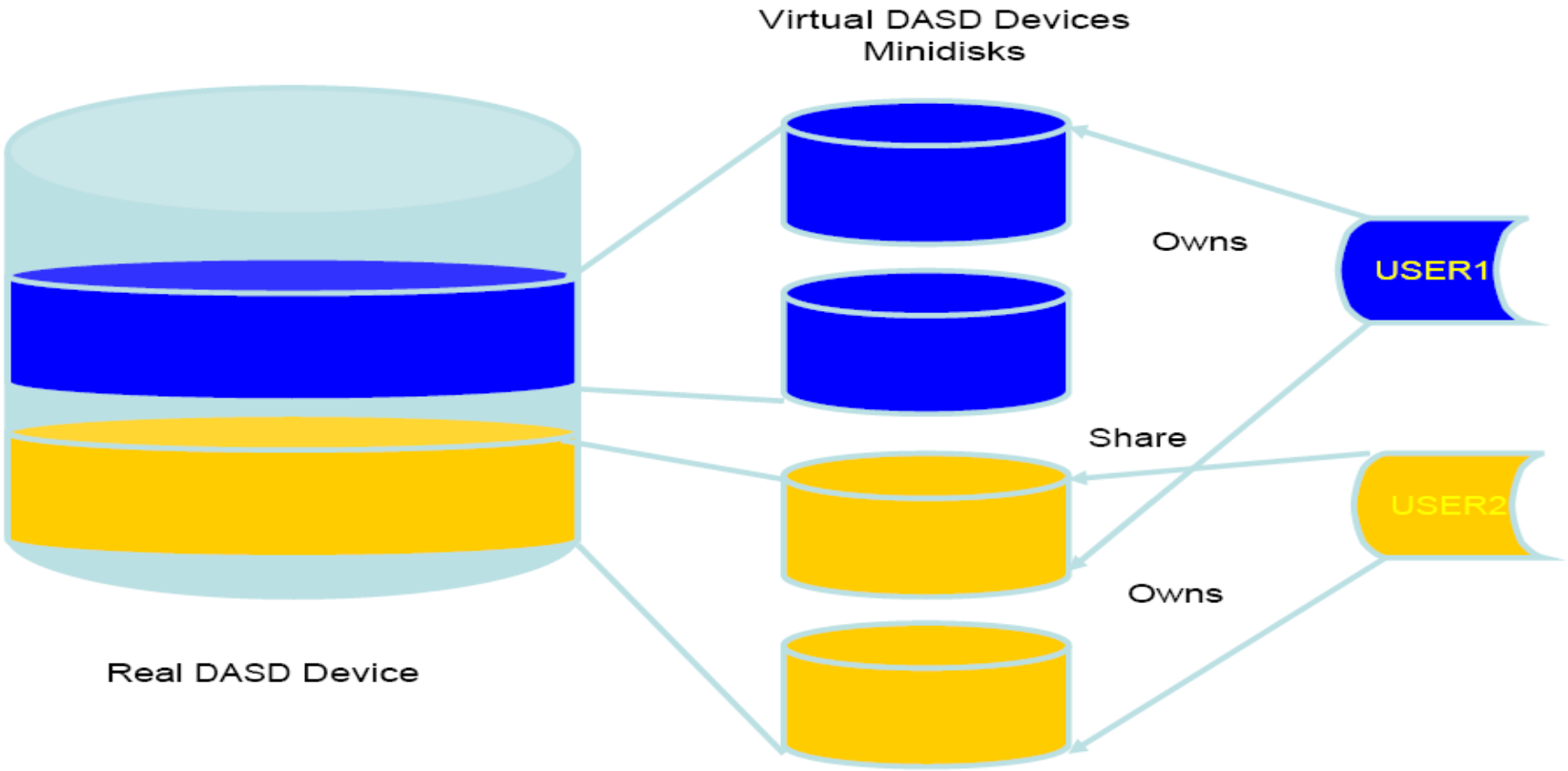
- Dopo la validazione dell'operazione il CP inizia lancia l'operazione di I/O al posto della macchina virtuale

- Se il device è condiviso il CP aggiungera alla richiesta di I/O con controlli addizionali per ulteriori limiti ad essa

- In altri casi sara dato il controllo di accesso al device in read only, nel qual caso il CP inserira i comandi nella I/O request che disabiliteranno tutti leoperazioni write-type.

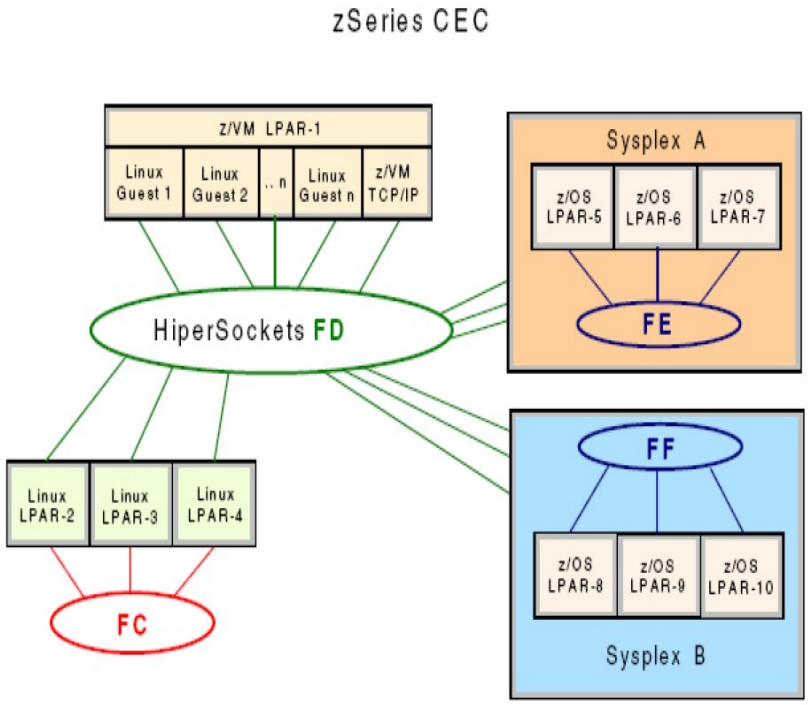
- In questa maniera verrà mantenuta la integrita dei dati e la privacy

z/VM DASD Management - Minidisks



HiperSockets

- Hipersockets e' una tecnica di comunicazione interna all Hardware gestita dall' ipervisore (PR/SM o z/VM) ed operata attraverso la memoria centrale.
- Essa fornisce un canale comunicativo simile ad una LAN all interno di un calcolatore della z/Architecture.
- Supporta il protocollo TCP/IP
- Supporta un protocollo di canle particolare denominato QDIO particolarmente veloce ed efficiente.
- HiperSockets possono essere usati per connettere tra di loro partizioni LPAR del PRSM o Macchine Virtuali gestite da z/VM .
- Si possono definire piu' HiperSockets fisicamente separati.



VLAN e VSWITCH

- z/VM supporta una rete di adapters virtuali per connettere tra loro delle macchine virtuali:

- **Virtual adapters**

CP offre un' Interfaccia di Rete Virtuale (NIC) che simula
- un Hipersockets
- una OSA-Express QDIO.

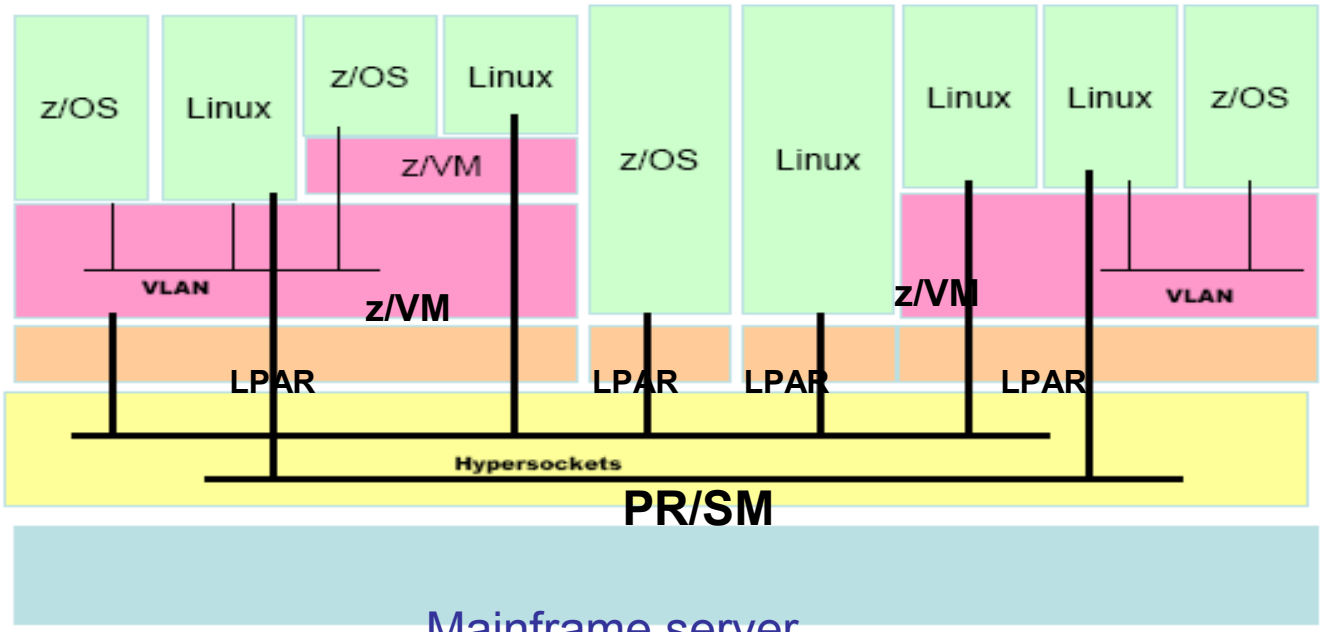
- **Virtual LAN (VLAN)**

CP fornisce comandi per connettere adapter virtuali di macchine virtuali ad una LAN (guest LAN) virtuale guest in un z/VM ad una rete virtuale

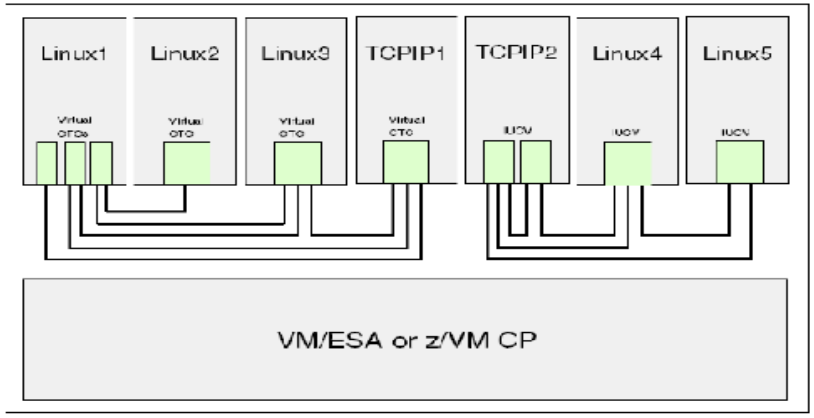
- **Virtual Switch (VSWITCH)**

Una LAN può essere "bridged" attraverso z/VM Virtual Internet Protocol (IP) switch, conosciuto come VSWITCH, ad una rete locale connessa ad un adapter OSA-Express QDIO A VSWITCH permette connettività ad una LAN esterna senza richiedere un router.

VLAN e Hipersochets

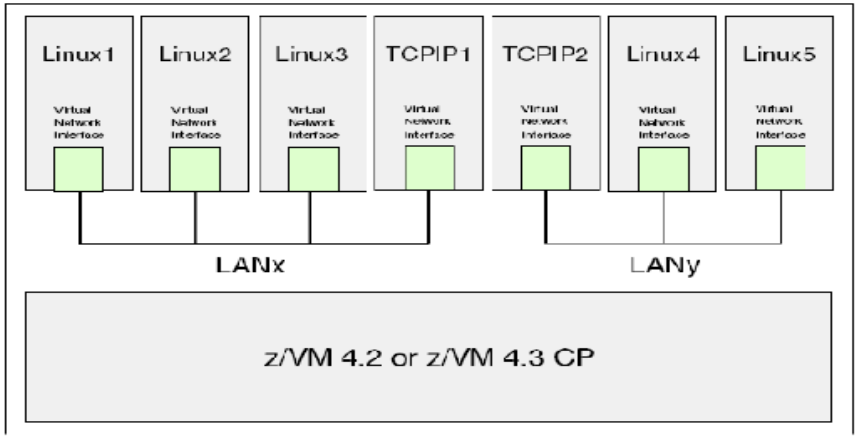


CTC e VLAN



VM point-to-point connections

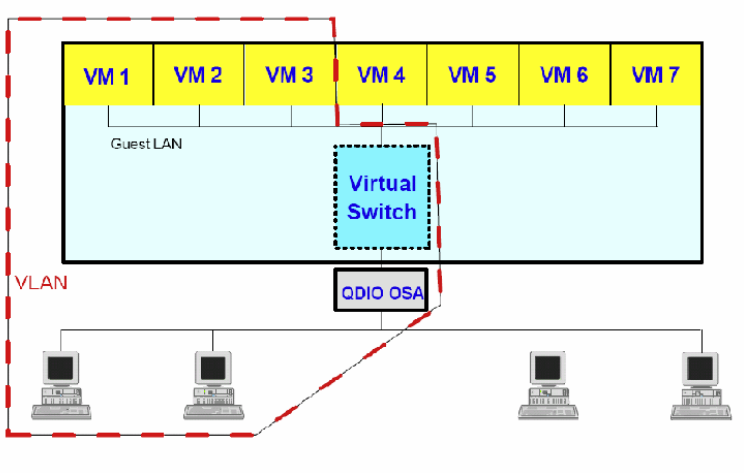
CTC point-to point

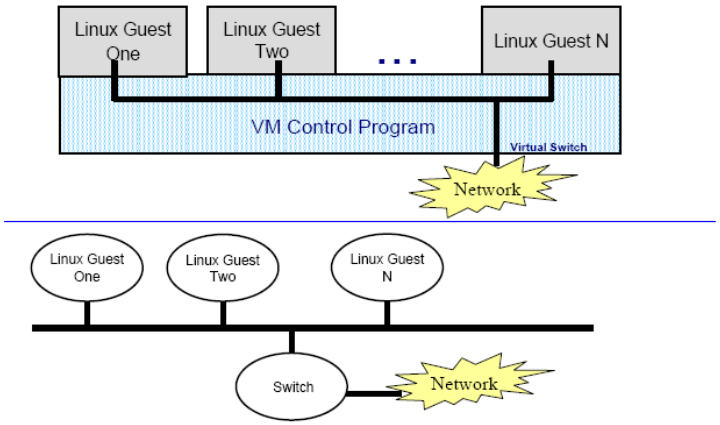


z/VM Guest LAN

VLAN

VSWITCH


VM 1
VM 2
VM 3
VM 4
VM 5
VM 6
VM 7
Guest LAN
Virtual Switch
VLAN
QDIO OSA


Linux Guest One
Linux Guest Two
Linux Guest N
VM Control Program
Virtual Switch
Network
Linux Guest One
Linux Guest Two
Linux Guest N
Switch
Network

Virtualizzazione nel Mainframe

Fine