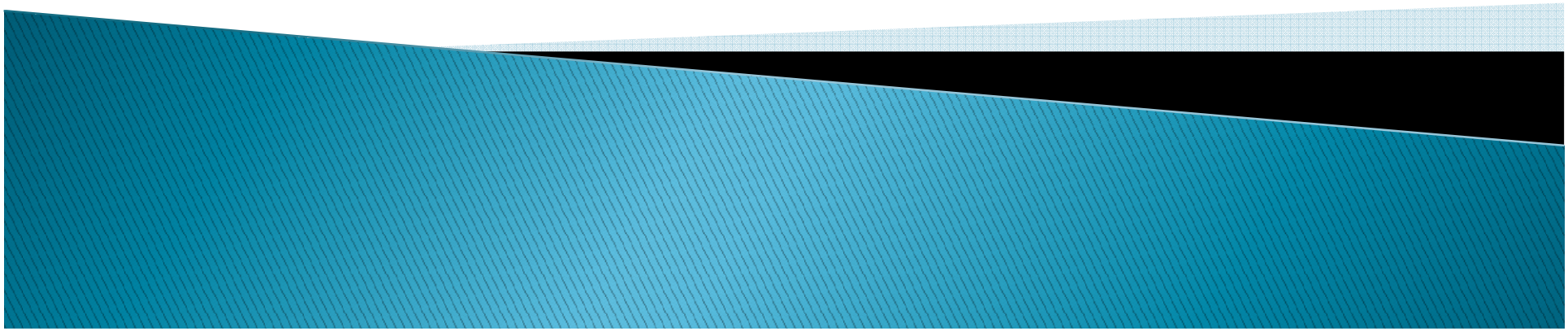
High Availability & Load Balancing @ LNF

Dael Maselli – Workshop INFN CCR 2010



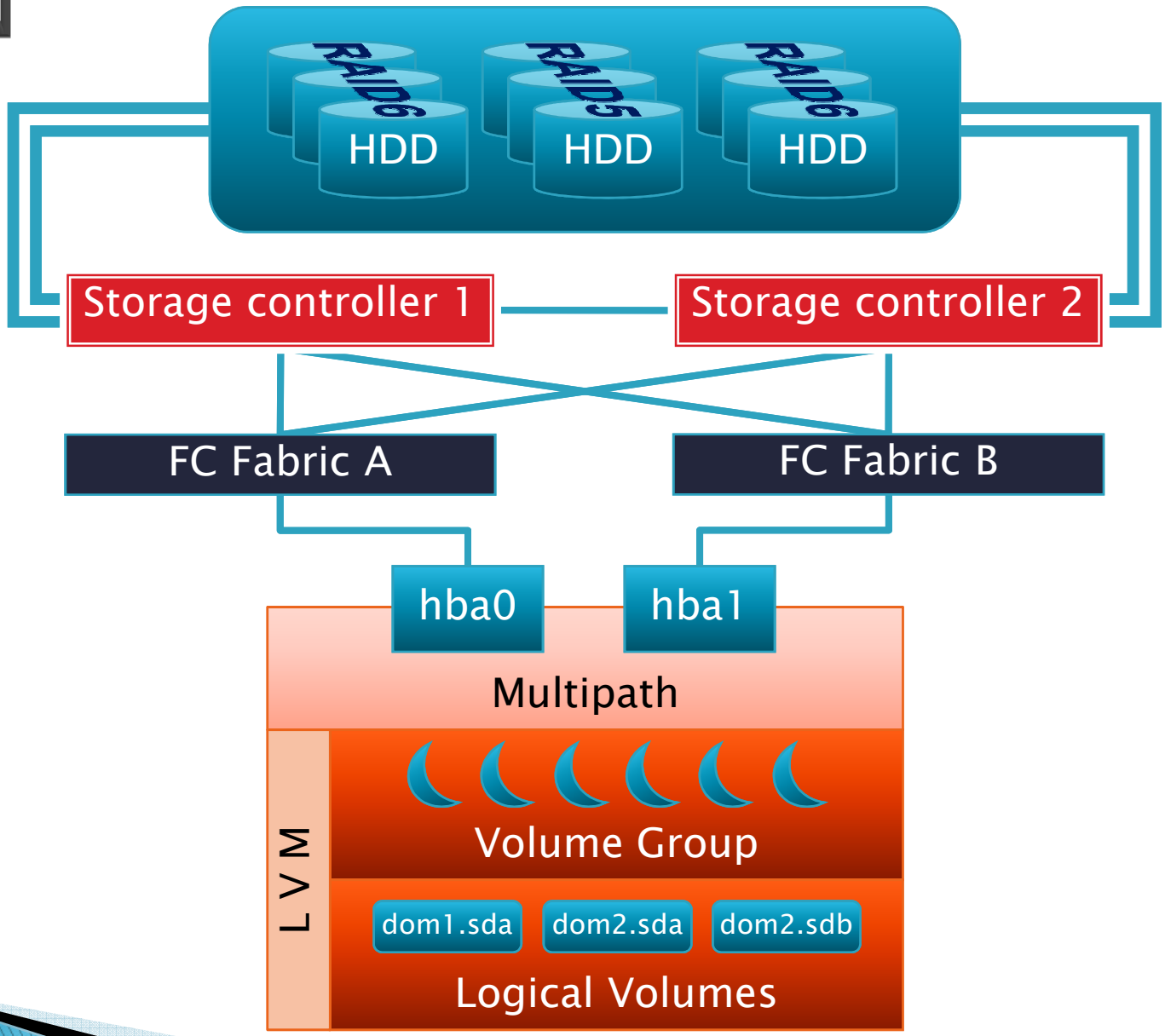
Architettura

- 3 cluster di nodi fisici attivi (2 LNF + 1 AAI)
  - Sistemi blade DELL e IBM
  - Red Hat Cluster Suite su Scientific Linux 5
- Ogni nodo è collegato a 2 switch FC
  - SAN Fibre Channel no-single-point-of-failure
- Ogni nodo è collegato a 2 switch Ethernet
  - I 2 switch sono collegati a centri stella diversi

Storage

- Tutto lo storage del cluster risiede su SAN
  - Compresi i device di boot delle macchine fisiche
- Multipathing con device-mapper-multipath nativo di Linux
- Lo spazio disco dedicato alle macchine virtuali è gestito tramite il Clustered Logical Volume Manager di Red Hat

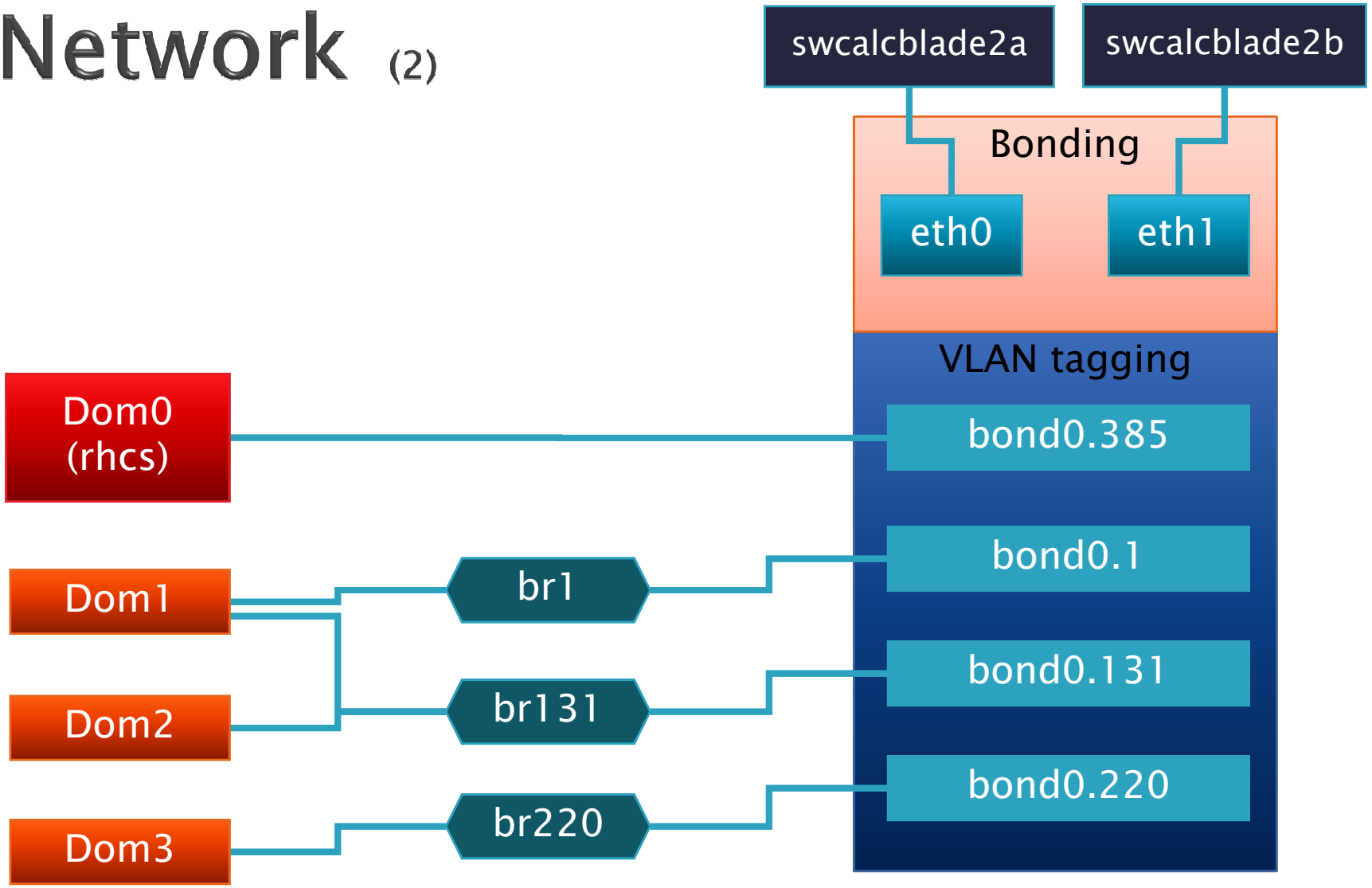
SAN


RAID6
HDD
RAID5
HDD
RAID6
HDD
Storage controller 1
Storage controller 2
FC Fabric A
FC Fabric B
hba0
hba1
Multipath
LVM
Volume Group
dom1.sda
dom2.sda
dom2.sdb
Logical Volumes

Network (1)

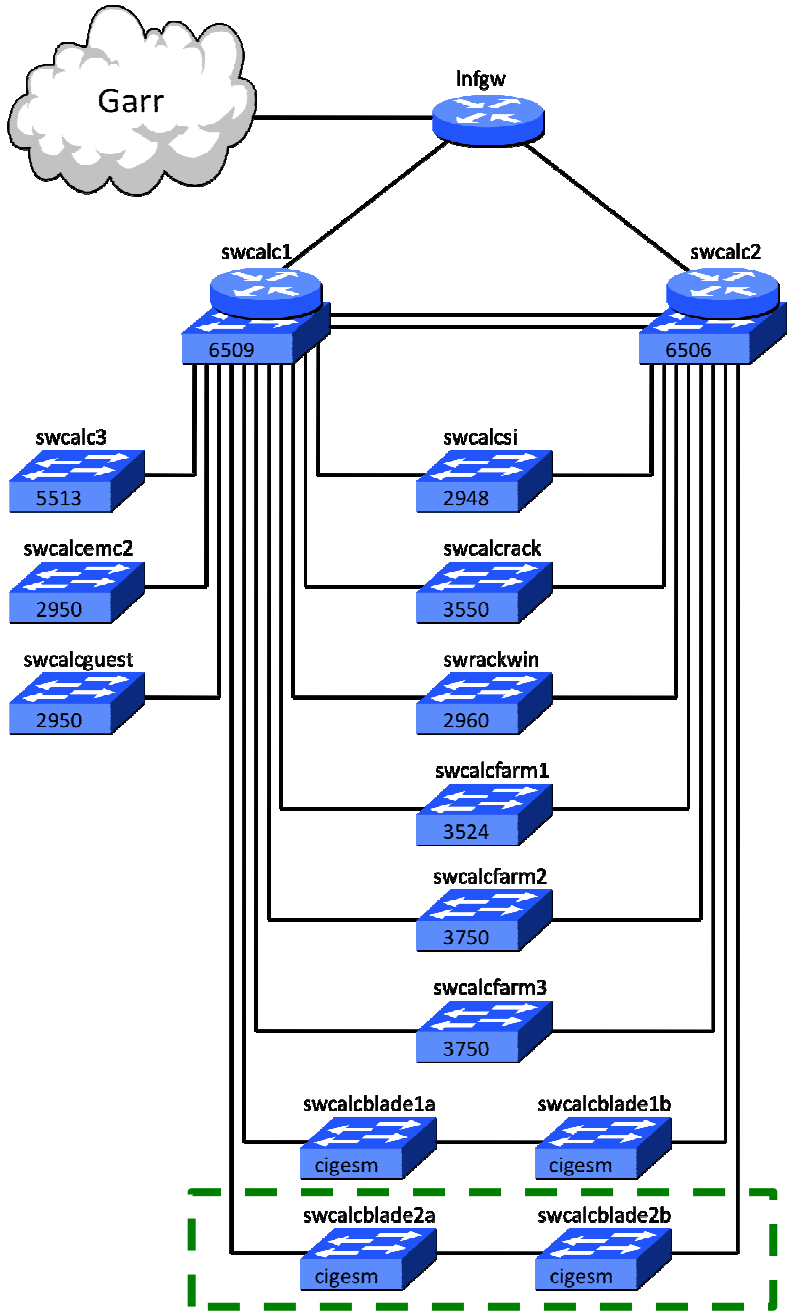
- Le 2 interfacce Ethernet sono in Bonding con modalità fault tolerant active-backup
- Per la comunicazione del cluster RHCS è necessaria una vlan dedicata
- Le macchine virtuali hanno interfacce su più VLAN differenti

Network (2)


swcalcblade2a
swcalcblade2b
Bonding
eth0
eth1
VLAN tagging
Dom0
(rhcs)
bond0.385
bond0.1
Dom1
br1
bond0.131
Dom2
br131
bond0.220
Dom3
br220

Network (3)

- Schema della rete della sala macchine LNF
- Le ultime due coppie rappresentano gli switch interni agli chassis dei blade delle dom0



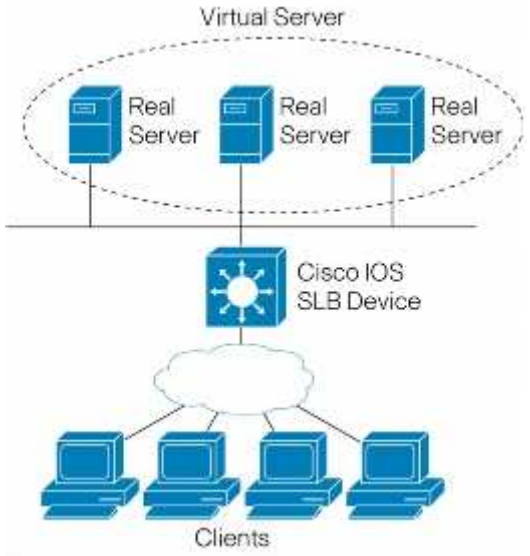
Cluster per Virtualizzazione

- Red Hat Cluster Suite si occupa di avviare, controllare e riallocare, in caso di fault, le macchine virtuali sui nodi del cluster
- La virtualizzazione è Xen (paravirtualized)
  - Con i pacchetti originali della distribuzione

Servizi

Alta Affidabilità e Load Balancing a livello 4

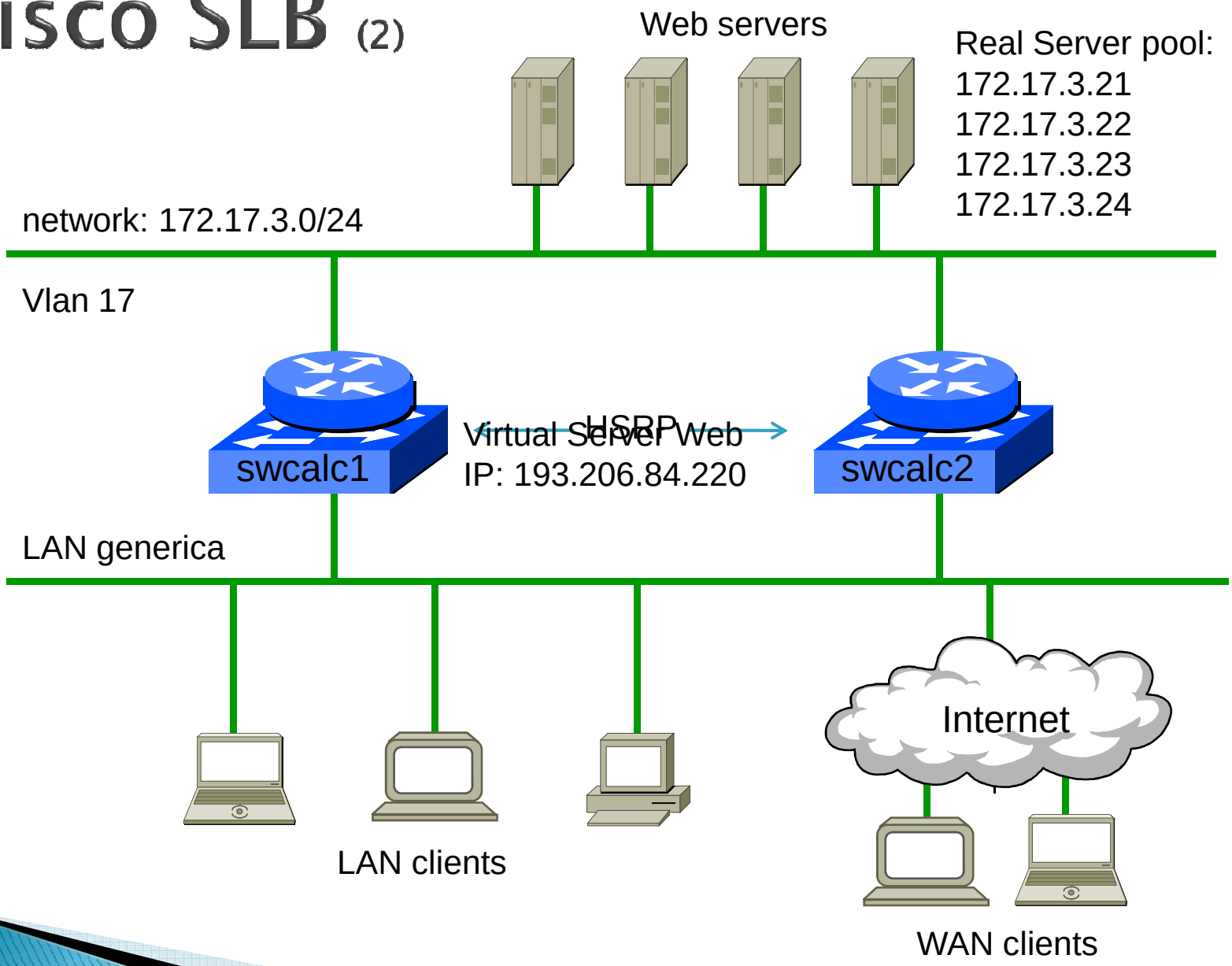
- Cisco SLB (Server Load Balancing)
- Necessario uno Switch Cisco Layer3 con IOS
  - *Policy Feature Card*



Cisco SLB (1)

- Il servizio SLB risponde ad un indirizzo IP virtuale su una certa porta per ogni servizio
- Inoltra le connessioni ai server reali tramite routing diretto o NAT
- Se SLB non vede passare pacchetti di risposta da un real server lo esclude temporaneamente
  - Ciò comporta che il traffico debba passare sempre per il router SLB, ossia che i real server siano su una network IP diversa dai client

Cisco SLB (2)


Web servers
Real Server pool:
172.17.3.21
172.17.3.22
172.17.3.23
172.17.3.24
network: 172.17.3.0/24
Vlan 17
swcalc1
HSRP
Virtual Server Web
IP: 193.206.84.220
swcalc2
LAN generica
LAN clients
Internet
WAN clients

Cisco SLB (3)

- Connessioni *sticky*
  - Entro un certo timeout dall'ultima connessione il singolo client contatterà sempre lo stesso real server
- Possibilità di ridondanza tramite HSRP
- Sustained Throughput Performance
  - Catalyst 6509 con Supervisor 2 (acquistata nel 2003)
  - SLB in NAT
  - iperf TCP su Gigabit
  - impatto: -6% (930 ➔ 870 Mbit/s)
  - CPU del Catalyst: +50% (nat)

Cisco SLB (4)

- ApacheBench
  - 10000 richieste dallo stesso client
  - 10 concorrenti
  - pagina statica solo html
  - Requests per second: 1184.14 [#/sec] (mean)
  - Time per request: 8.445 [ms] (mean)
  - Transfer rate: 661.34 [Kbytes/sec]
  - CPU del Catalyst: +10%
- *Il sistema SLB in questo caso non comporta alcun peggioramento delle prestazioni*

Cisco SLB (conf)

```
ip slb serverfarm WWW
 nat server
 real 172.17.3.21
  inservice
 !
 real 172.17.3.22
  inservice
 !
 real 172.17.3.23
  inservice
 !
 real 172.17.3.24
  inservice
!
```

```
ip slb vserver WWW-HTTP
 virtual 193.206.84.220 tcp www
 serverfarm WWW
 sticky 360 group 81
 inservice
!
ip slb vserver WWW-HTTPS
 virtual 193.206.84.220 tcp https
 serverfarm WWW
 sticky 360 group 81
 inservice
!
```

HA livello Applicazione

- Il Cisco SLB controlla il funzionamento del real server a livello di trasporto
- A volte le applicazioni possono continuare a rispondere alle connessioni ma non servire i contenuti correttamente
- La soluzione, semplice ma efficace, è di controllarne la funzionalità con uno script a crontab e in caso di errore spegnere il servizio
  - in tal modo il server non risponderà alle connessioni e SLB farà il resto

F I N E

…domande?