Sharing the Wealth Using Vlans on Vswitch

David Kreuter, Dave Jones
Denver Share
Session 9163

Objectives

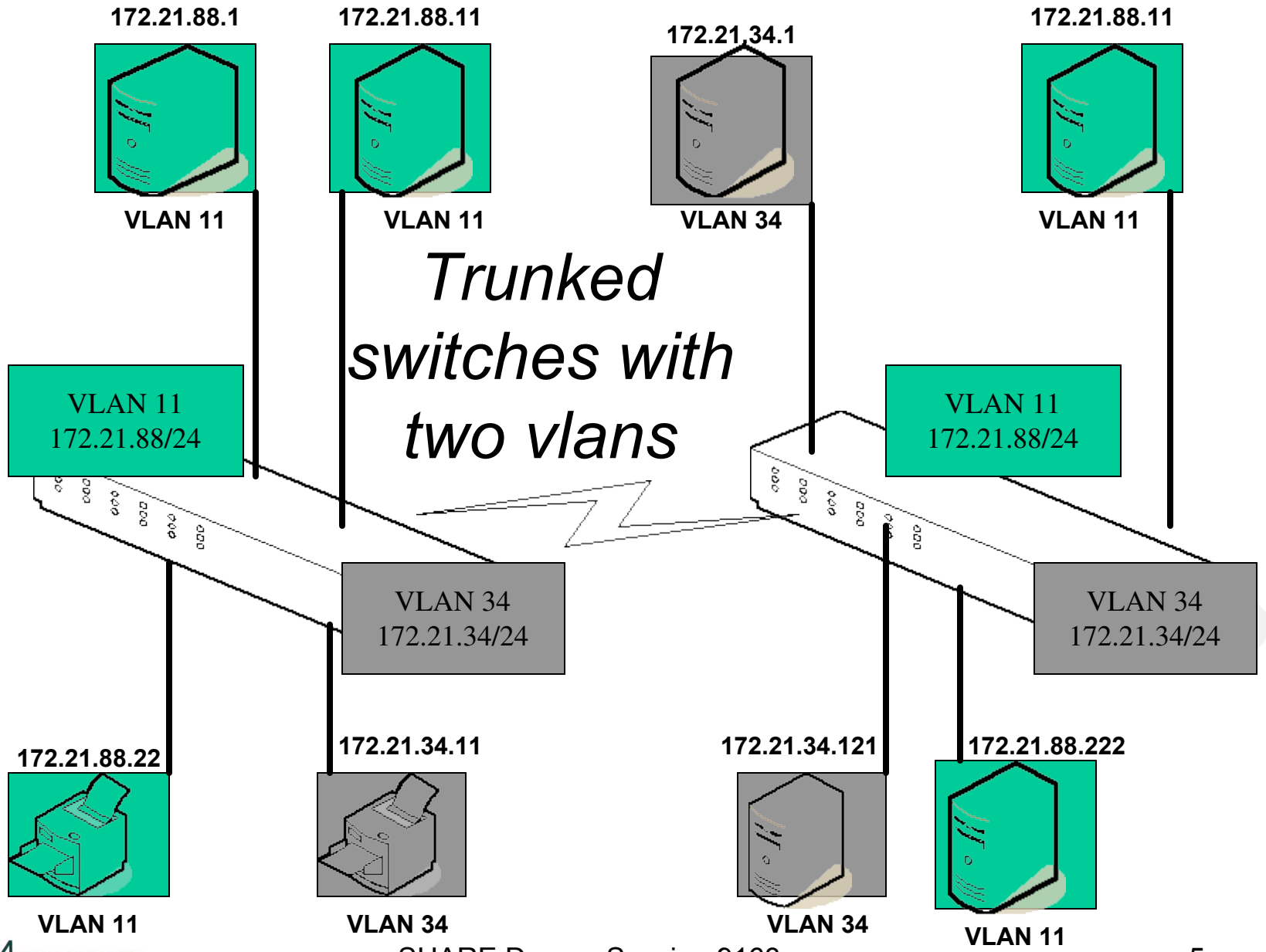
- Definition of virtual lan
- Purpose of vlan
- Hipersocket, vswitch and vlan in z/VM
- Linux virtual machine support
- Configuration and usage examples

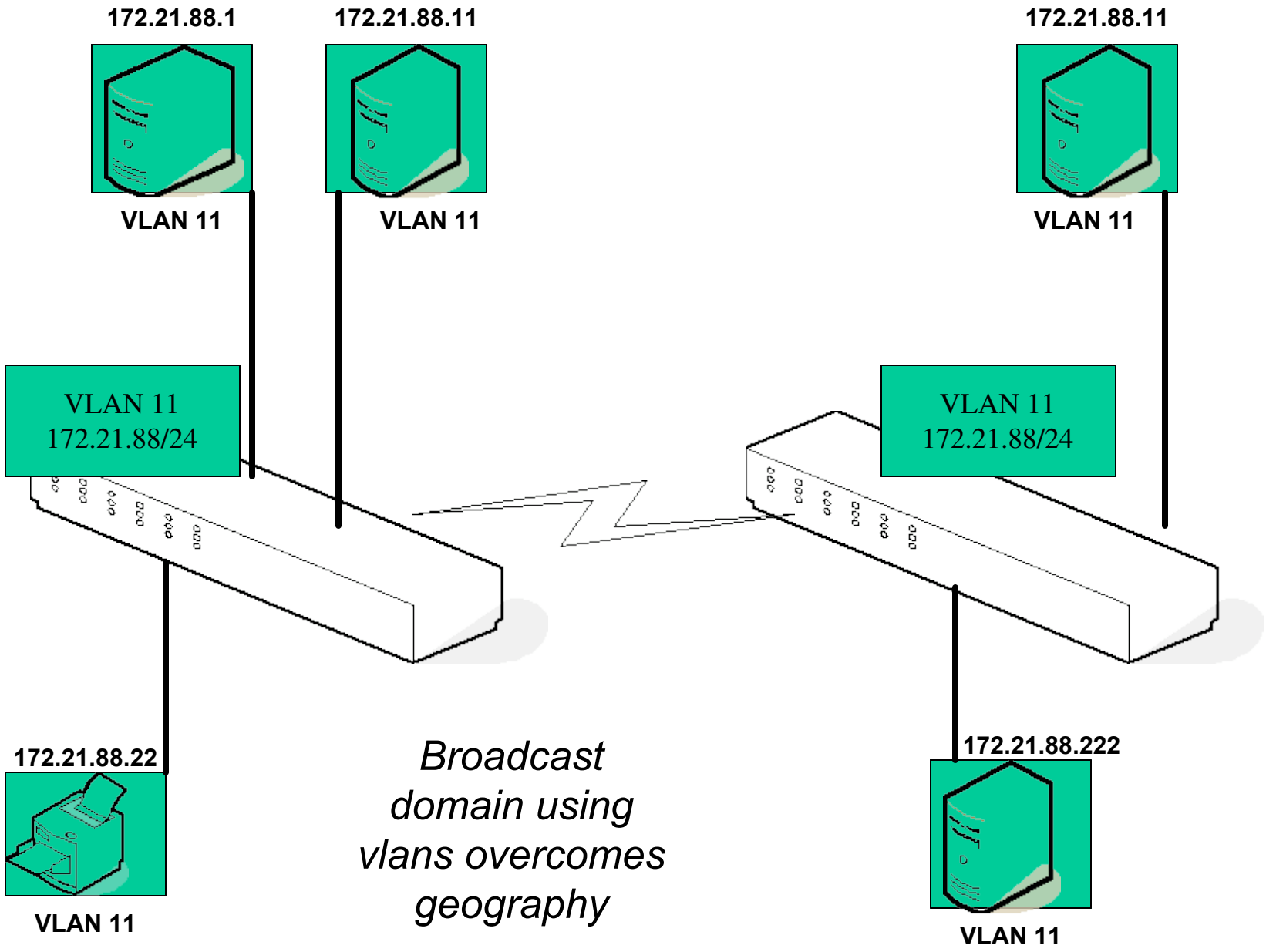
What are vlans

- Virtual lans
- Group stations by function regardless of switch location.
  - Daisy chaining switches
- Limit size of broadcast domain by organizing functionally.
- Proprietary methods and IEEE standard 802.1Q
- Software controls

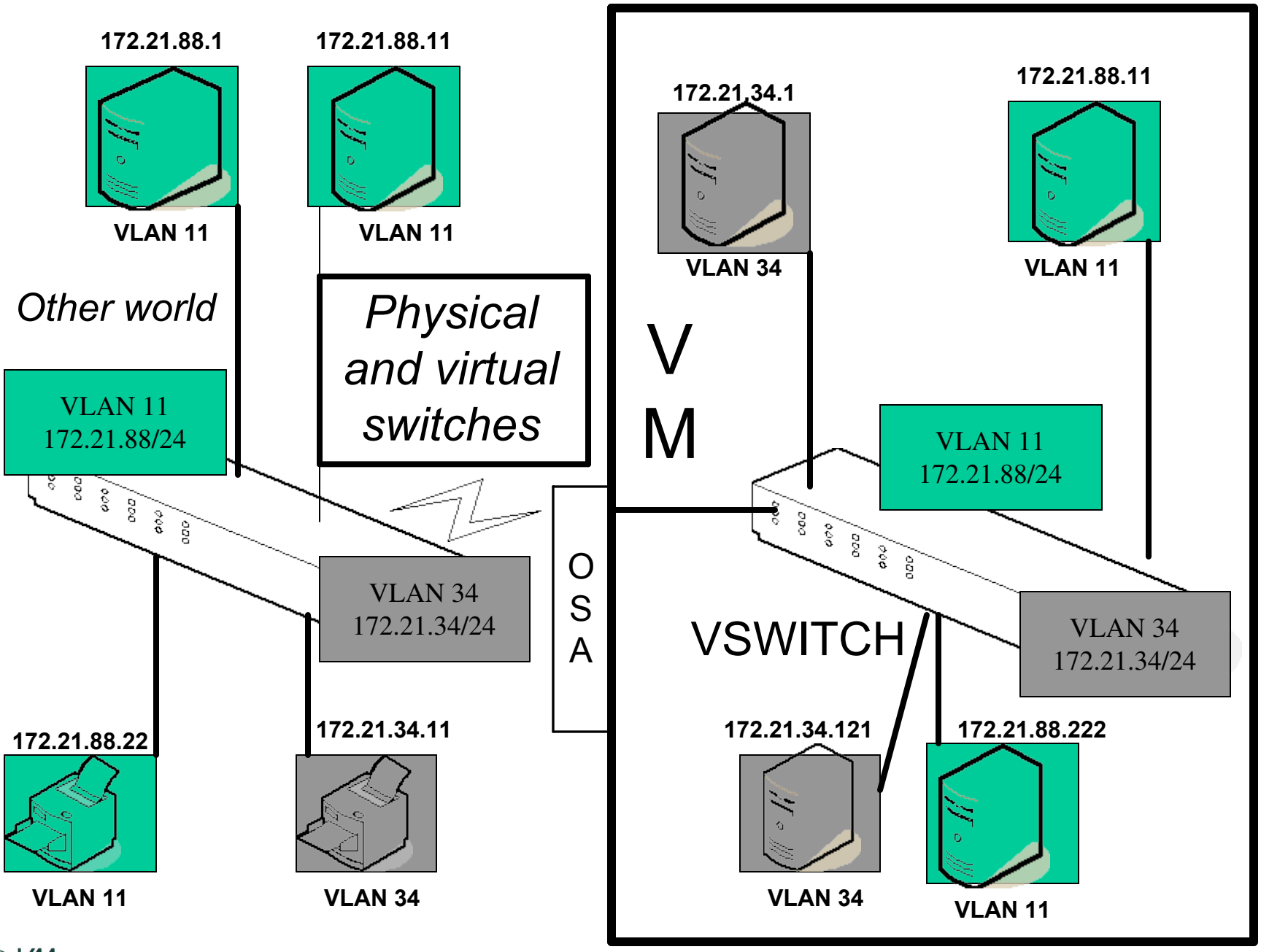
Simplified vlan mechanics

- Frame contains two extra bytes with vlan number referred to as the vlan tag.
- Stations that use **trunk** connections on switch include vlan tag in transmission.
  - VLAN aware host can also send untagged frames. Such frames are assigned the default port VLAN ID. Be careful! May not get delivered where you intended.
- Stations that use **access** connections the switch strips off the tag then delivers.
  - Presence of VLAN tag renders packet "invalid" to a non-VLAN-aware host.


172.21.88.1
172.21.88.11
172.21.34.1
172.21.88.11
VLAN 11
VLAN 11
VLAN 34
VLAN 11
Trunked switches with two vlans
VLAN 11
172.21.88/24
VLAN 11
172.21.88/24
VLAN 34
172.21.34/24
VLAN 34
172.21.34/24
172.21.88.22
172.21.34.11
172.21.34.121
172.21.88.222
VLAN 11
VLAN 34
VLAN 34
VLAN 11



Broadcast domain using vlans overcomes geography


172.21.88.1
172.21.88.11
VLAN 11
VLAN 11
Other world
Physical and virtual switches
VLAN 11
172.21.88/24
VLAN 34
172.21.34/24
OSA
172.21.88.22
172.21.34.11
VLAN 11
VLAN 34
172.21.34.1
VLAN 34
172.21.88.11
VLAN 11
V
M
VLAN 11
172.21.88/24
VSWITCH
VLAN 34
172.21.34/24
172.21.34.121
172.21.88.222
VLAN 34
VLAN 11

Vlans and z/VM

- Multiple vlans on top of same network
  - Use on hipersocket to connect lpars
- Each vlan assigned to unique network
  1 Disconnected vswitch to collect SNMP and monitor data
  2 Mapping successfully to physical switches via OSA QDIO connectivity

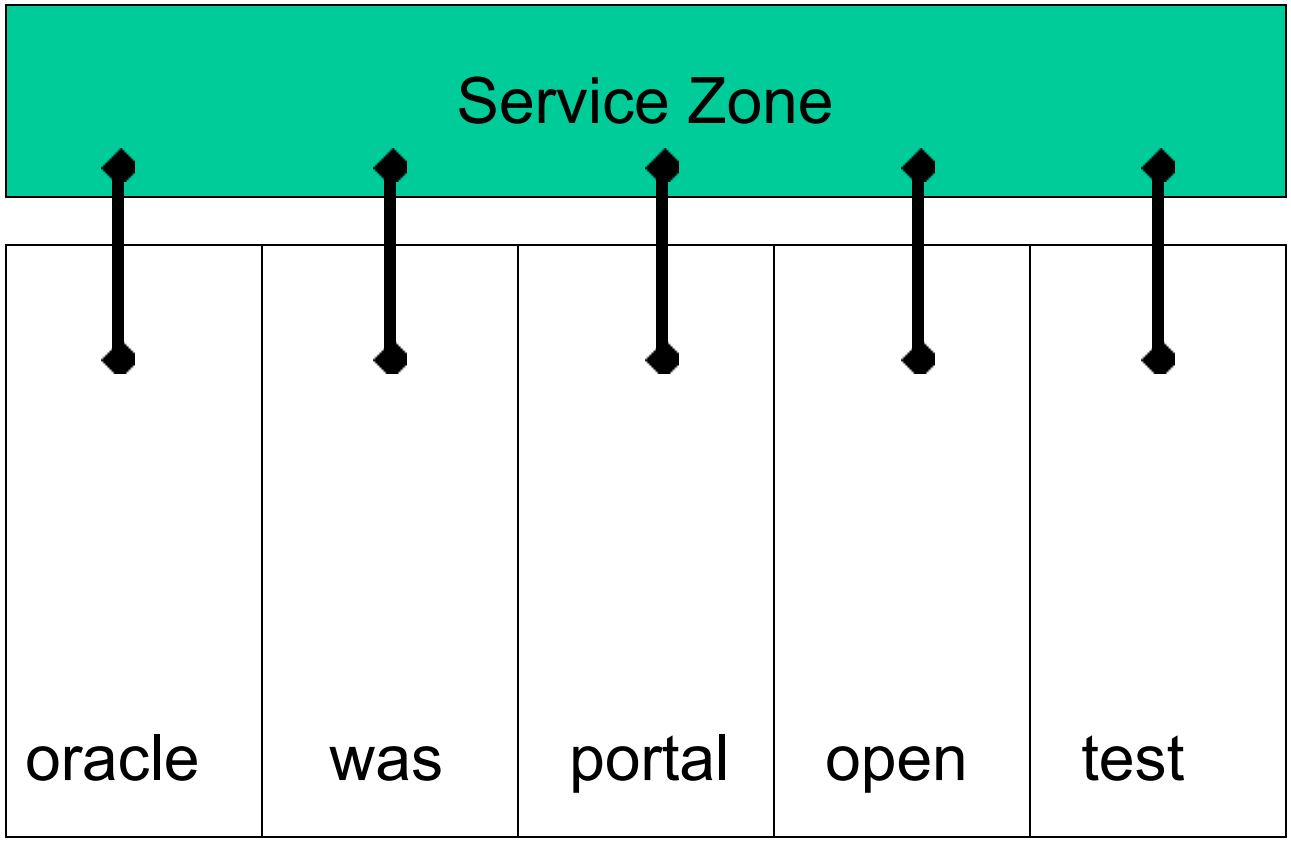
All are useful and work great!

Sysprog Predicament

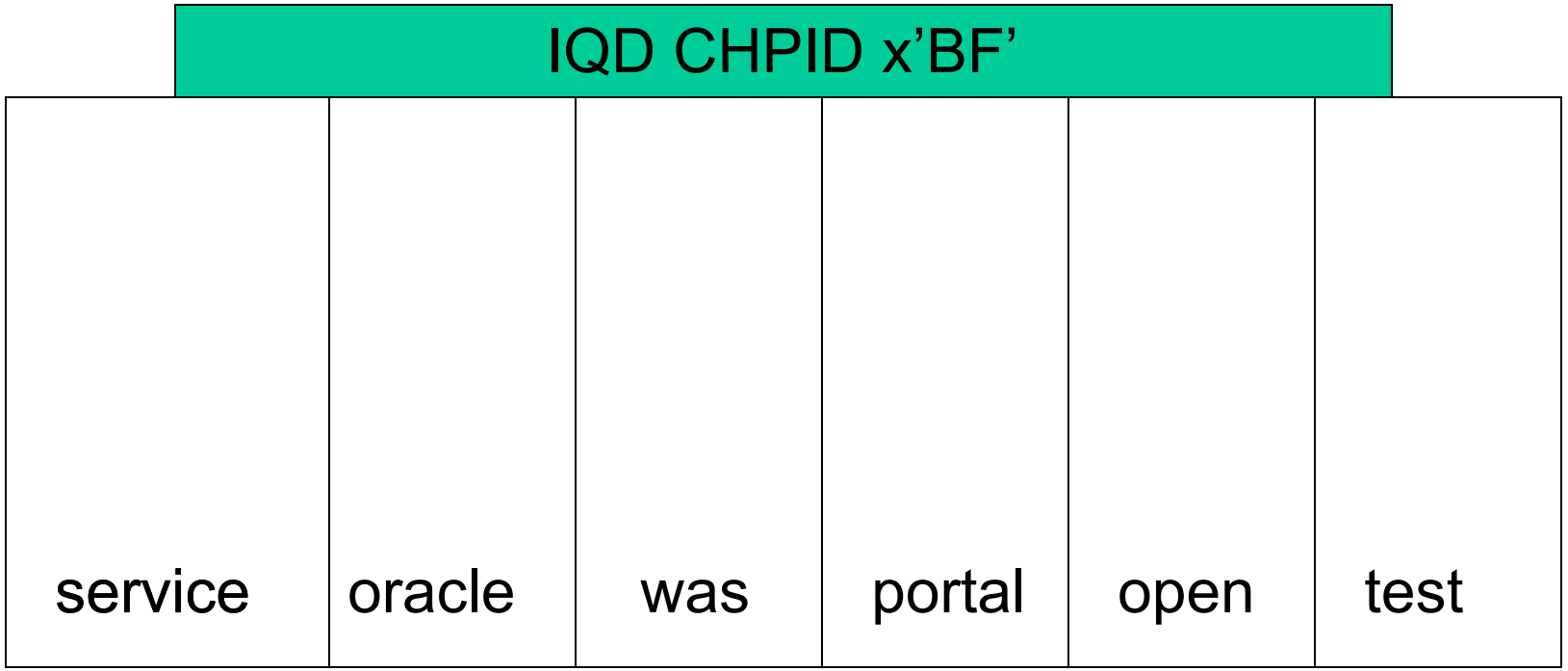
- Multiple lpars need to be connected from the service zone lpar, but the other lpars cannot see each other.
- Connection needed for systems programming needs: ftp, rscs, and telnet, remote dirmaint.

Solution: hipersocket network with vlan
Each lpar with a unique vlan number

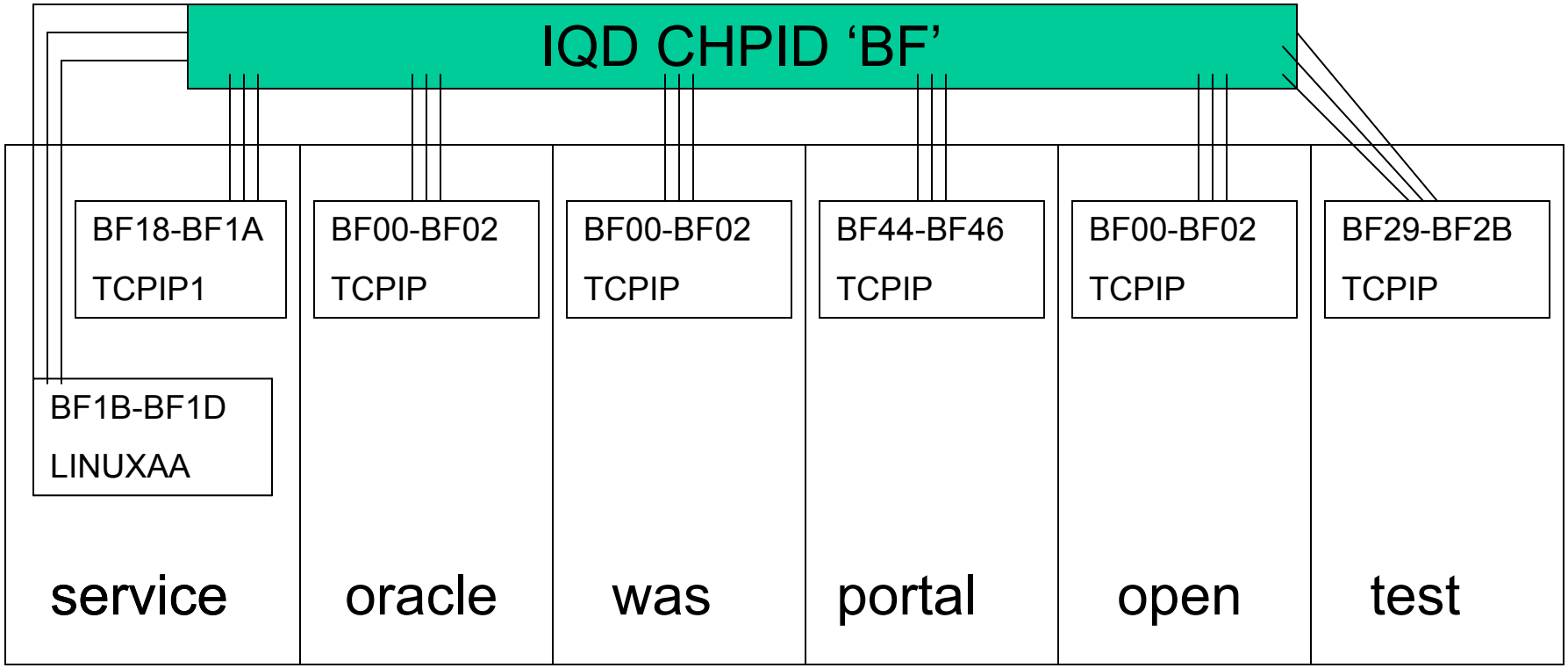
Conceptual View: Permitted connections



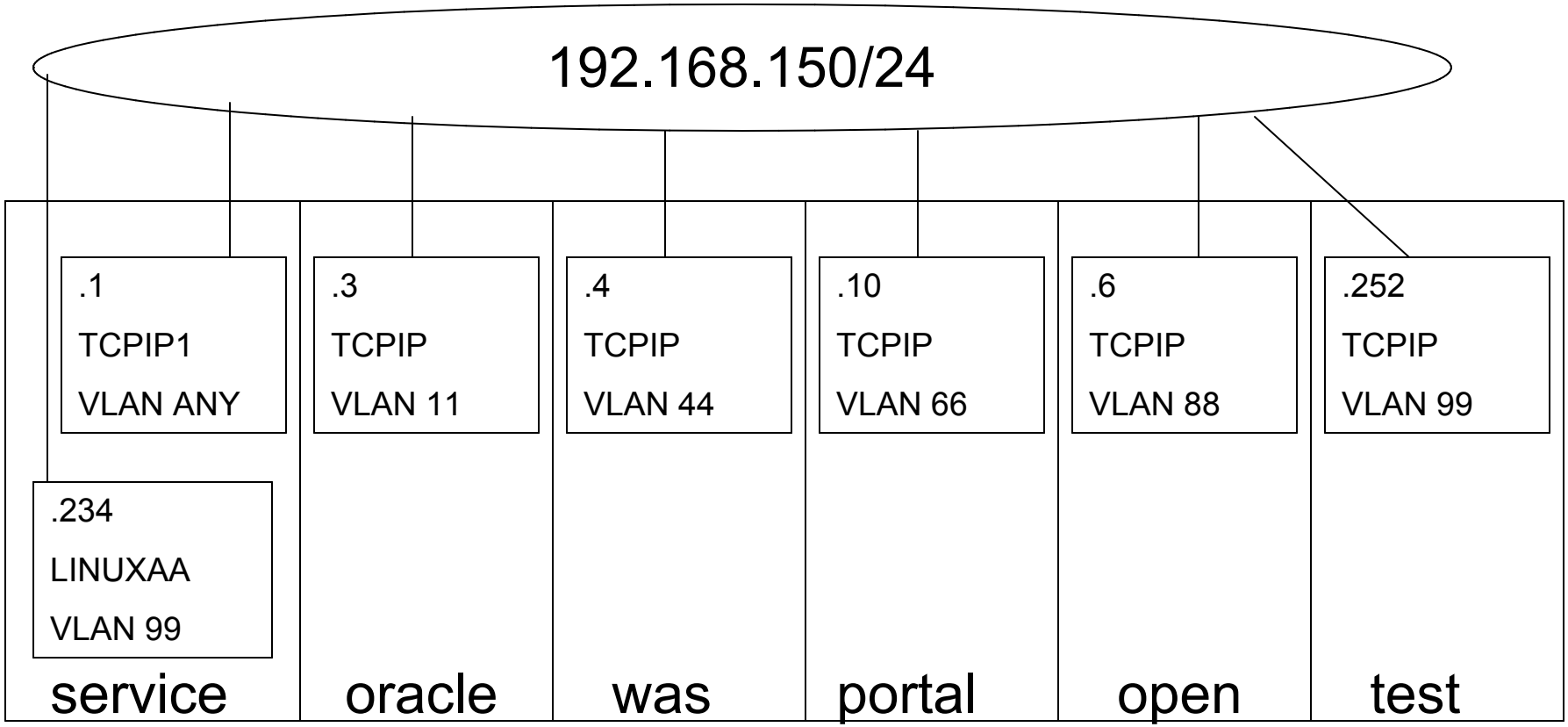
Hipersocket Hardware



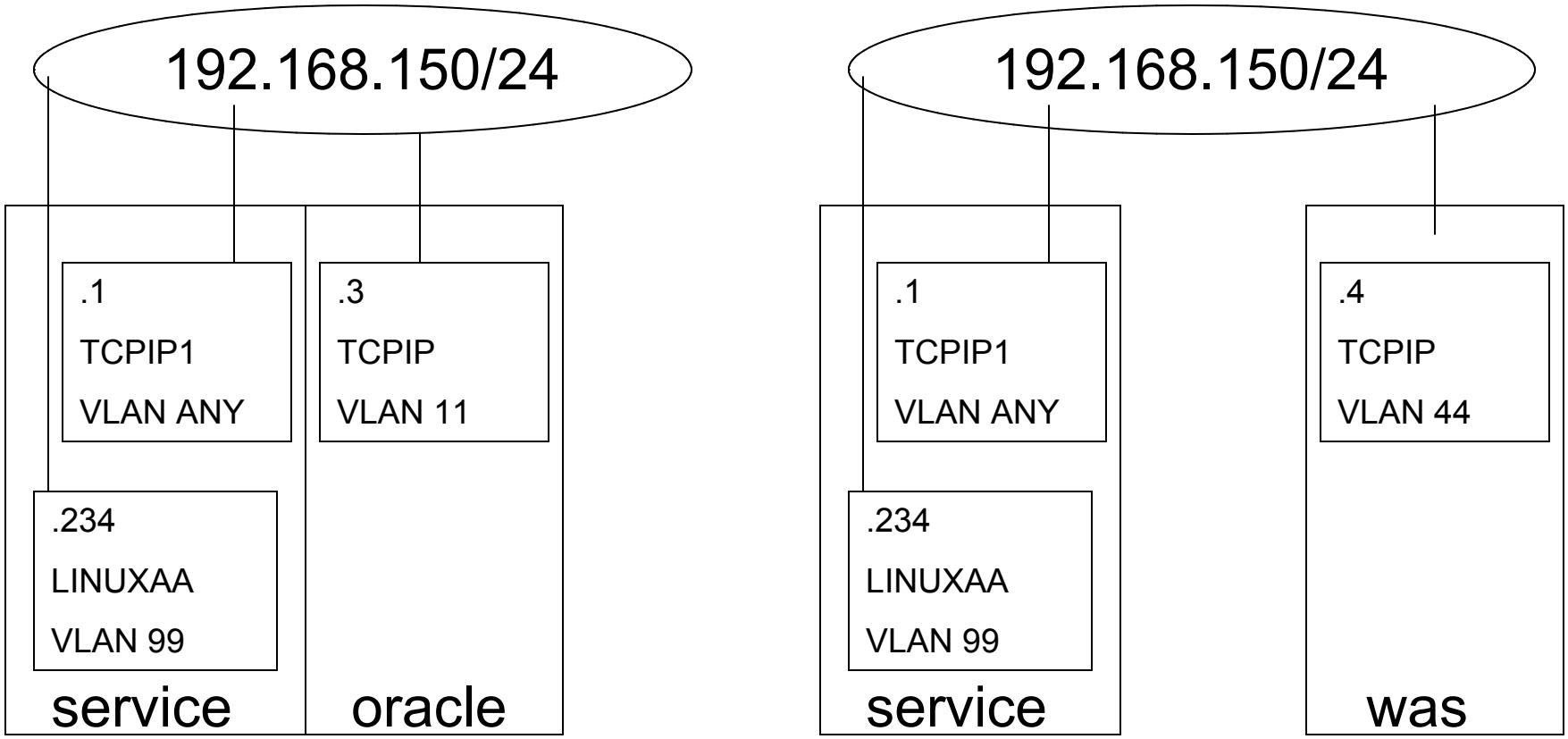
Hipersocket hardware: triplets



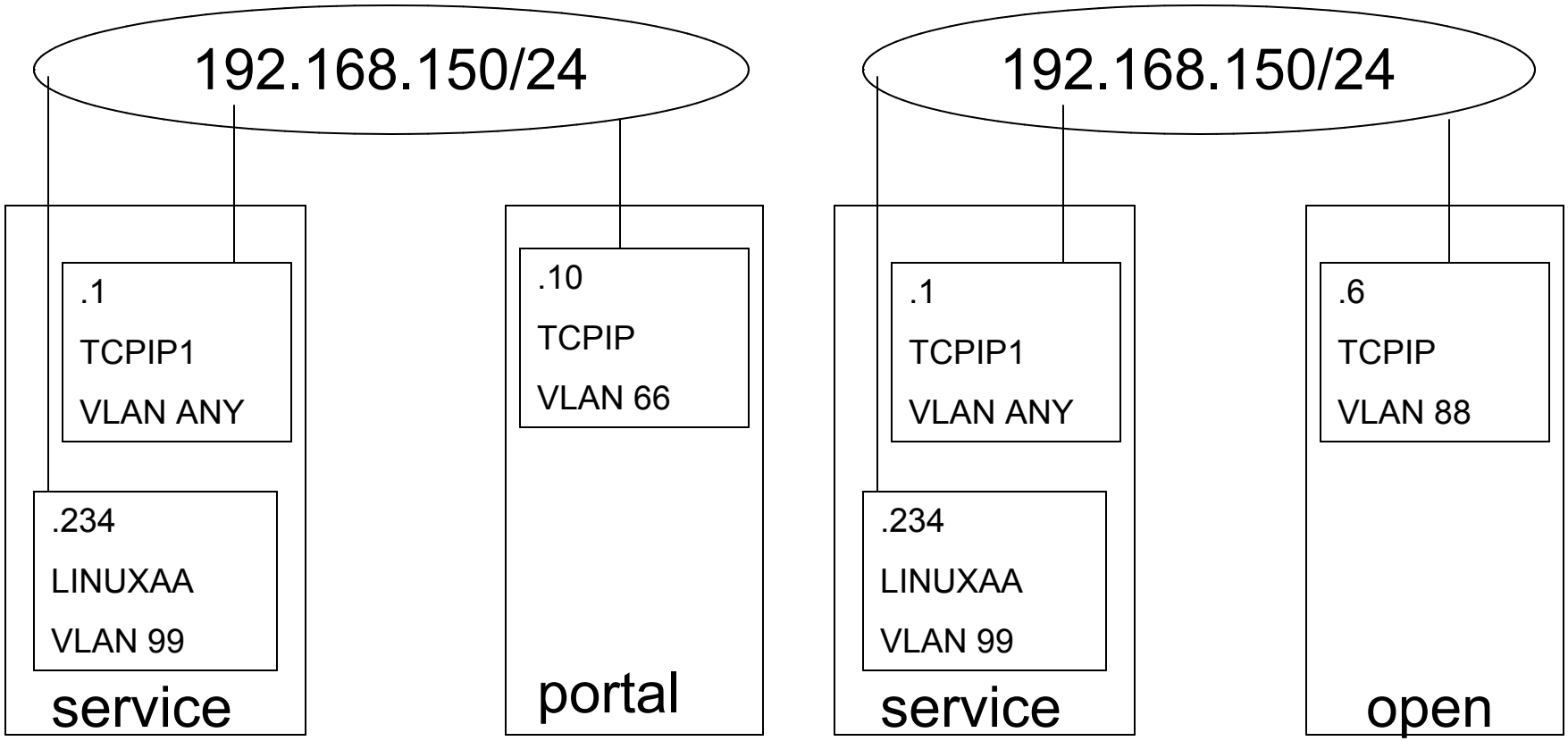
Hipersocket network: IP addresses and vlan assignment



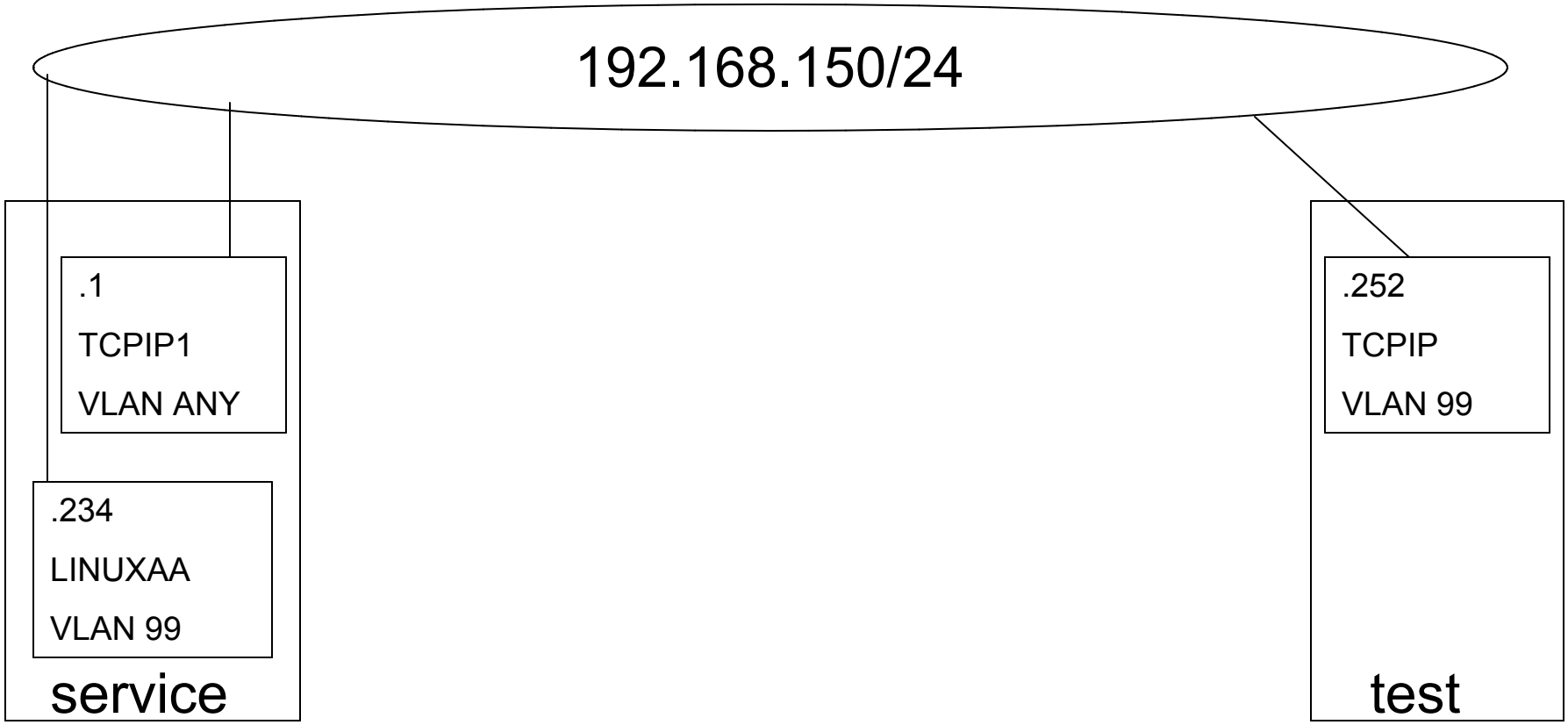
Vlan on hipersocket networks



Vlan on hipersocket network three



Hipersocket network: IP addresses

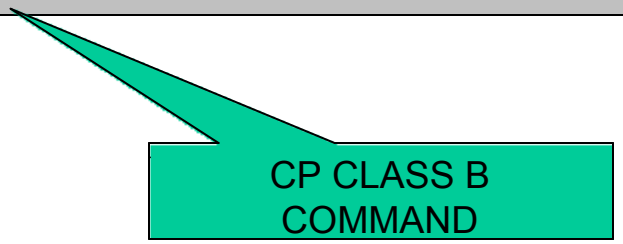


Will show …

- Confguration
  - CP
  - TCPIP
- Dynamically adding Linux virtual machine with a vlan device at ip address 192.168.150.234
- queries

Hipersocket CP Configuration

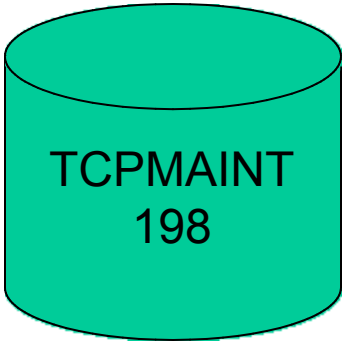
```
query osa
OSA  BF18 ATTACHED TO TCPIP1   BF18 DEVTYPE HIPER      CHPID BF IQD
OSA  BF19 ATTACHED TO TCPIP1   BF19 DEVTYPE HIPER      CHPID BF IQD
OSA  BF1A ATTACHED TO TCPIP1   BF1A DEVTYPE HIPER      CHPID BF IQD
:
```




```
query cplevel

z/VM Version 5 Release 3.0, service level 0802 (64-bit)
```

TCPIP1 Configuration: SYSTEM DTCPARMS

:NICK.TCPIP1 :TYPE.SERVER :CLASS.STACK
:ATTACH.BF18-BF1A



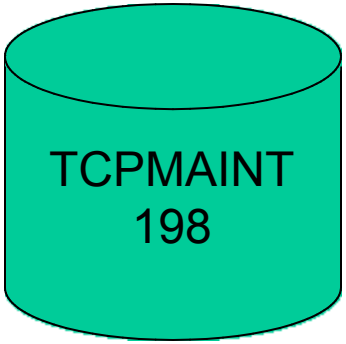
netstat level

VM TCP/IP Netstat Level 530

IBM 2094; z/VM Version 5 Release 3.0, service level 0802 (64-bit), VM TCP/IP Level 530; RSU 0802 running TCPIP MODULE E2 dated 09/27/08 at 00:26

TCPIP1 Configuration: TCPIP1 TCPIP

```
:
DEVICE HIPER1 HIPERS BF18
LINK HIPER QDIOIP HIPER1 NOFWD MTU 0 VLAN ANY

HOME
192.168.150.1   255.255.255.0 HIPER

GATEWAY
DEFAULTNET 192.168.150.1 HIPER 4000


START HIPER1
```

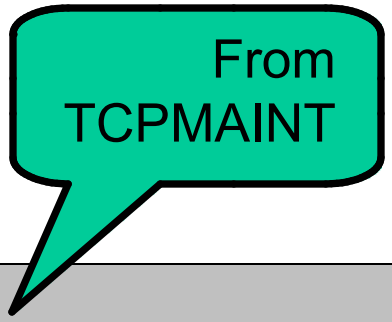


Packets from/to any vlan on the hipersocket device

TCPIP1 Console log

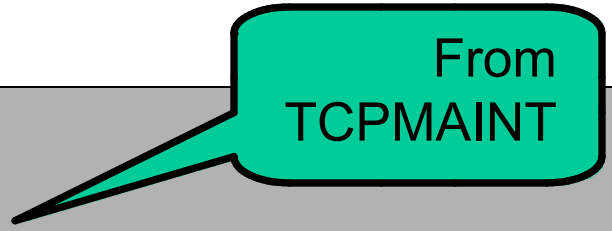
```
08:04:04 DTCOSD080I HIPERS initializing:
08:04:04 DTCPRI385I  Device HIPER1:
08:04:04 DTCPRI386I     Type: HIPERS, Status: Not started
08:04:04 DTCPRI387I     Envelope queue size: 0
08:04:04 DTCPRI388I     Address: BF18
:
08:04:05 DTCQDI001I QDIO device HIPER1 device number BF1A:
08:04:05 DTCQDI007I   Enabled for QDIO data transfers
08:04:05 DTCOSD238I ToOsd: IPv4 multicast support enabled for HIPER1
08:04:05 DTCOSD246I HIPERS device HIPER1: Assigned IPv4 address 192.168.150.1
```

TCPIP1 Queries of the Dynamic Nature




```
ifconfig hiper
HIPER    inet addr: 192.168.150.1 mask: 255.255.255.0
         UP BROADCAST MULTICAST MTU: 4000
         vdev: BF18 rdev: BF18 type: HIPERS
         ipv6: DISABLED
         cpu: 0 forwarding: DISABLED
         RX bytes: 2441368403 TX bytes: 501172660
```

NETSTAT HOME




```
netstat home
VM TCP/IP Netstat Level 530

IPv4 Home address entries:

Address        Subnet Mask     Link          VSWITCH
-------        -----------     ------        -------
192.168.150.1  255.255.255.0   HIPER         <none>

IPv6 Home address entries: None

Ready; T=0.01/0.01 17:29:23
```

NETSTAT DEVLINKS from TCPMAINT

```
netstat devlinks
VM TCP/IP Netstat Level 530
Device HIPER1              Type: HIPERS       Status: Ready
 Queue size: 0     CPU: 0     Address: BF18       Port name: UNASSIGNED
 IPv4 Router Type: NonRouter  Arp Query Support: No
  Link HIPER               Type: QDIOIP       Net number: 0
    BytesIn: 2441369810      BytesOut: 501176551
    Forwarding: Disabled     MTU: 4000          IPv6: Disabled
    Maximum Frame Size  : 24576
    Broadcast Capability: Yes
    Multicast Capability: Yes
    Group                        Members
    -----                        -------
    224.0.0.1                      1
```

Adding to a Linux critter dynamically

- Attach devices
- Create group devices structure
- Configure interface

SuSE and Linux levels

SuSE SLES 10 Service Pack 2

 # cat /proc/version

Linux version 2.6.16.60-0.21-default (geeko@buildhost) (gcc version 4.1.2 20070115 (SUSE Linux)) #1 SMP Tue May 6 12:41:02 UTC 2008

 # cat /etc/SuSE*

SUSE Linux Enterprise Server 10 (s390x)

VERSION = 10

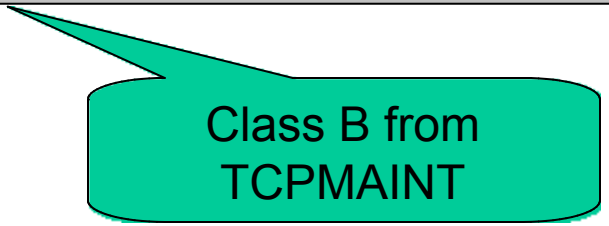
PATCHLEVEL = 2

cat: /etc/SuSEconfig: Is a directory

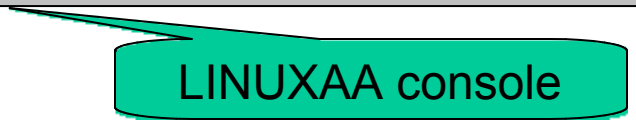
Attach devices

attach bf1b-bf1d LINUXAA

BF1B-BF1D ATTACHED TO LINUXAA

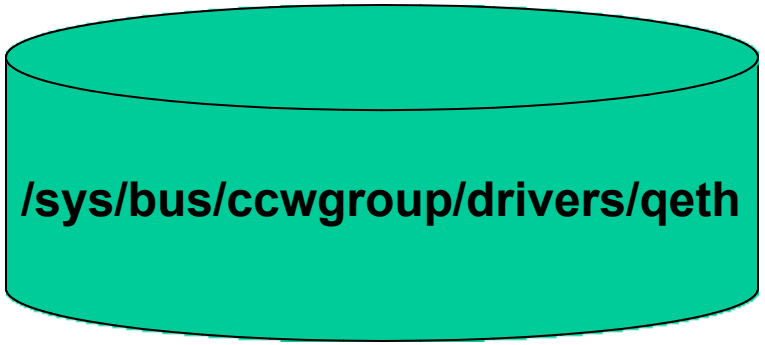


BF1B-BF1D ATTACHED TO LINUXAA



Create group devices structure dynamically

In memory file system



From root in LINUXAA

/sys/bus/ccwgroup/drivers/qeth

```
# echo 0.0.bf1b,0.0.bf1c,0.0.bf1d > group


# cd /sys/bus/ccwgroup/devices/0.0.bf1b
```

In /sys/bus/ccwgroup/devices/0.0.1bfb place the devices online

 # echo 1 > online

qeth: Device 0.0.bf1b/0.0.bf1c/0.0.bf1d is a HiperSockets card (level: () with link type HiperSockets.

qeth: set adapter parameters not supported on device 0.0.bf1b.

qeth: Hardware IP fragmentation not supported on hsi0

qeth: VLAN enabled

qeth: Multicast enabled

qeth: IPv6 not supported on hsi0

qeth: Broadcast enabled

qeth: Using SW checksumming on hsi0.

qeth: Outbound TSO not supported on hsi0

From root in LINUXAA

Load the vlan support, configure the physical device, and add the vlan device.

```
# modprobe 8021q

# ifconfig hsi0 0.0.0.0

# vconfig add hsi0 99
Added VLAN with VID == 99 to IF -:hsi0:-
```

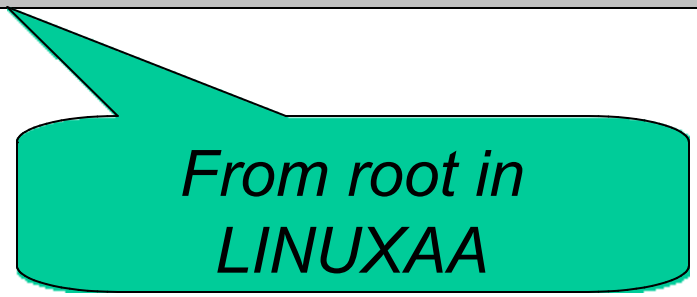
Load the vlan support module

From root in LINUXAA

Add vlan device on vlan 99 on the hsi0 interface

Display of vlan device before configuring it

```
#ifconfig hsi0.99
hsi0.99   Link encap:Ethernet  HWaddr 00:00:00:00:00:00
          BROADCAST NOARP MULTICAST  MTU:16384  Metric:1
          RX packets:0 errors:0 dropped:0 overruns:0 frame:0
          TX packets:0 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:0 (0.0 b)  TX bytes:0 (0.0 b)
```



Configure vlan device

```
# ifconfig hsi0.99 192.168.150.234


# ifconfig hsi0.99
hsi0.99   Link encap:Ethernet  HWaddr 00:00:00:00:00:00
          inet addr:192.168.150.234 Bcast:192.168.150.255  Mask:255.255.255.0
          inet6 addr: fe80::200:ff:fe00:0/64 Scope:Link
          UP BROADCAST RUNNING NOARP MULTICAST  MTU:16384  Metric:1
          RX packets:0 errors:0 dropped:0 overruns:0 frame:0
          TX packets:8 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:0 (0.0 b)  TX bytes:588 (588.0 b)
```

From root in LINUXAA

2 pings: #1 on same vlan (works) #2 on different vlan (fails as expected)

```
# ping 192.168.150.252 -c 2
PING 192.168.150.252 (192.168.150.252) 56(84) bytes of data.
64 bytes from 192.168.150.252: icmp_seq=1 ttl=60 time=0.175 ms
64 bytes from 192.168.150.252: icmp_seq=2 ttl=60 time=0.178 ms
--- 192.168.150.252 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 999ms
rtt min/avg/max/mdev = 0.175/0.176/0.178/0.013 ms


# ping 192.168.150.6 -c 2
PING 192.168.150.6 (192.168.150.6) 56(84) bytes of data.
--- 192.168.150.6 ping statistics ---
2 packets transmitted, 0 received, 100% packet loss, time 1007ms
```

From root in LINUXAA

Uses for the hipersocket vlan network

- RSCS
- Telnet
- ftp
- Remote DIRMAINT
- *Show rscs*

RSCS network query from the service zone: all nodes

```
smsg rscs query links connect show defparm name
 Link
 Name     Default Parm Text
 ORACLE   TCPID=TCPIP1 HOST=192.168.150.3
 WAS      TCPID=TCPIP1 HOST=192.168.150.4
 OPEN     TCPID=TCPIP1 HOST=192.168.150.6
 TEST     TCPID=TCPIP1 HOST=192.168.150.252
 PORTAL   TCPID=TCPIP1 HOST=192.168.150.10
 5 links found
```

RSCS network query from the ORACLE zone: only the service zone

```
sm rscs cmd ORACLE query links connect show defparm name
Ready; T=0.01/0.01 14:26:17
 From ORACLE: Link
 From ORACLE: Name     Default Parm Text
 From ORACLE: SERVICE TCPID=TCPIP1  HOST=192.168.150.1
 From ORACLE: 1 link found
```

Command issued from service zone to RSCS on the ORACLE zone

Usefulness of disconnected vswitch with lan

- Gathering SNMP data for Velocity product.
- All Linux machines produce this data.
  - Linux machines on different zones.
- With a disconnected vswitch vlan membership can be RACF protected.
  - *Cannot be protected in a guest lan*

RACF and vlans

- Vswitch membership protected as well.
  - Base profile
- Vlan membership protected in the **vmlan** resource class.
  - Vlan qualified control.
- Does not do any packet investigation.
- Will not allow communication with vlans that you are not a member of even if you configure in Linux.

rac rlist vmlan system.perfmon all

```
CLASS     NAME
-----     ----
VMLAN     SYSTEM.PERFMON
:
ALTER COUNT   CONTROL COUNT   UPDATE COUNT   READ COUNT
-----------   -------------   ------------   ----------
 000000       000000          002197         000000

USER     ACCESS   ACCESS COUNT
----     ------   ------ -----
LINUXAA  UPDATE   000975
TCP1PERF UPDATE   000967
GROUPAA  UPDATE   000252
GROUPBB  UPDATE   000252
:
```

Shows racf groups and user profiles in the resource list (vswitch)

The vswitch itself is a protected resource in a base profile.

rac rlist vmlan system.perfmon.0048 all

```
CLASS     NAME
-----     ----
VMLAN     SYSTEM.perfmon.0048
:
ALTER COUNT   CONTROL COUNT   UPDATE COUNT   READ COUNT
-----------   ------------    ------------   ----------
  000002        000000         000324        000000

USER      ACCESS   ACCESS COUNT
----      ------   ------ -----
LINUXAA   UPDATE     000975
TCP1PERF  UPDATE     000967
GROUPAA   UPDATE     000252
:
```

Shows racf groups and user profiles in the resource list (vlan).

Vlan membership is a protected resource! Nice – eh?

rac rlist vmlan permit system.perfmon.0048 class(vmlan) id(maint)

```
rac permit system.perfmon.0048 class(vmlan) id(maint) access(update)


rac rlist vmlan system.perfmon.0048 all
Ready; T=0.01/0.01 14:18:47
 CLASS     NAME
 -----     ----
 VMLAN     SYSTEM.PERFMON.0048
 :
 USER     ACCESS   ACCESS COUNT
 ----     ------   ------ -----
 LINUXAA  UPDATE     000975
 TCP1PERF UPDATE     000967
 GROUPAA  UPDATE     000252
 MAINT    UPDATE     000000
:
```

1. Add user profile maint to the vlan resource.

2. MAINT is in the access list

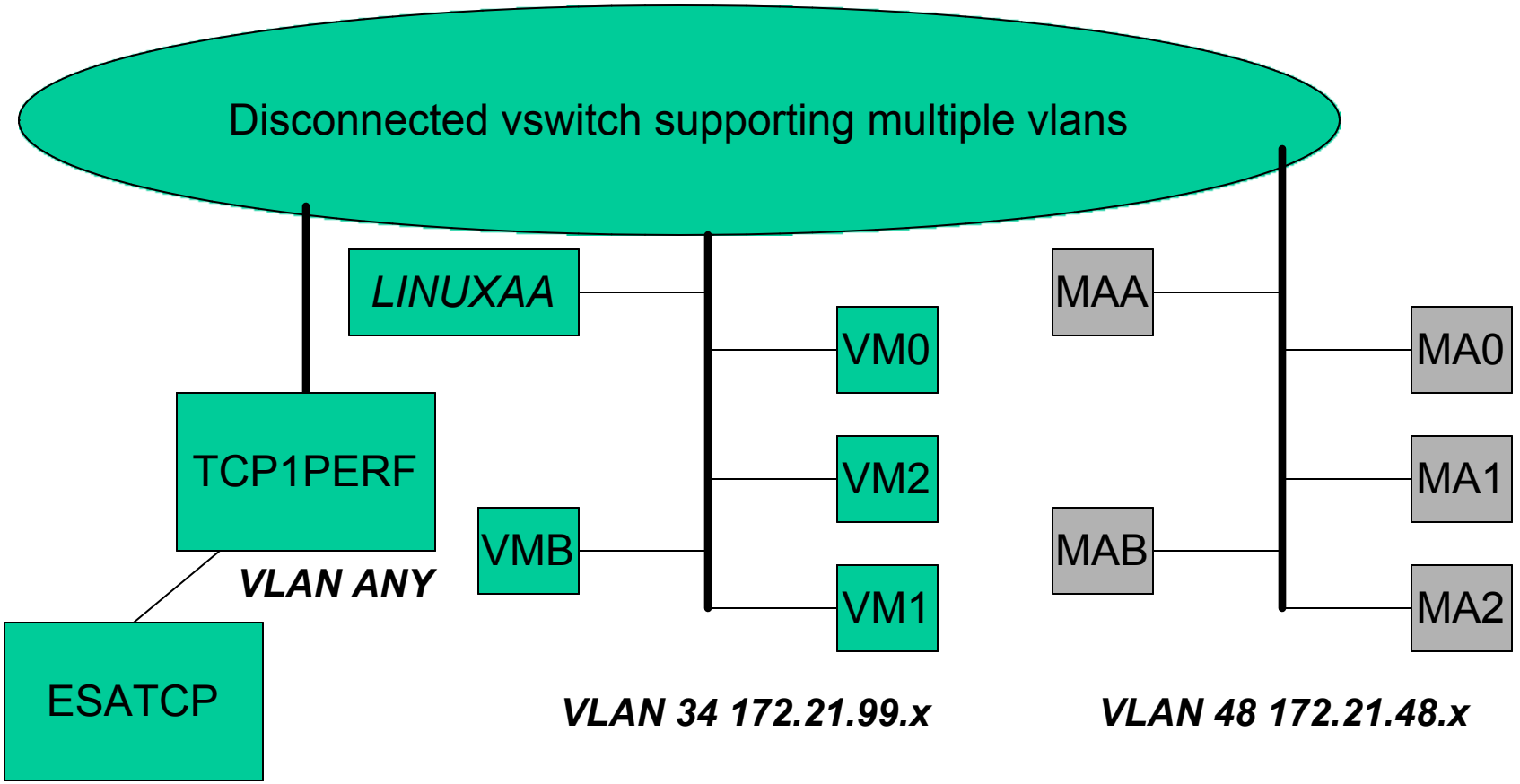
rac rlist vmlan permit system.perfmon.0048 delete

```
rac pe system.perfmon.0048 class(vmlan) id(maint) delete


rac rl vmlan system.perfmon.0048 all
Ready; T=0.01/0.01 14:22:02
 CLASS     NAME
 -----     ----
 VMLAN     SYSTEM.PERFMON.0048


:
USER     ACCESS   ACCESS COUNT
 ----     ------   ------ -----
 LINUXAA  UPDATE     000975
 TCP1PERF UPDATE     000967
 GROUPAA  UPDATE     000252
```

3. Some mechanics – remove maint

4. Maint not in list anymore

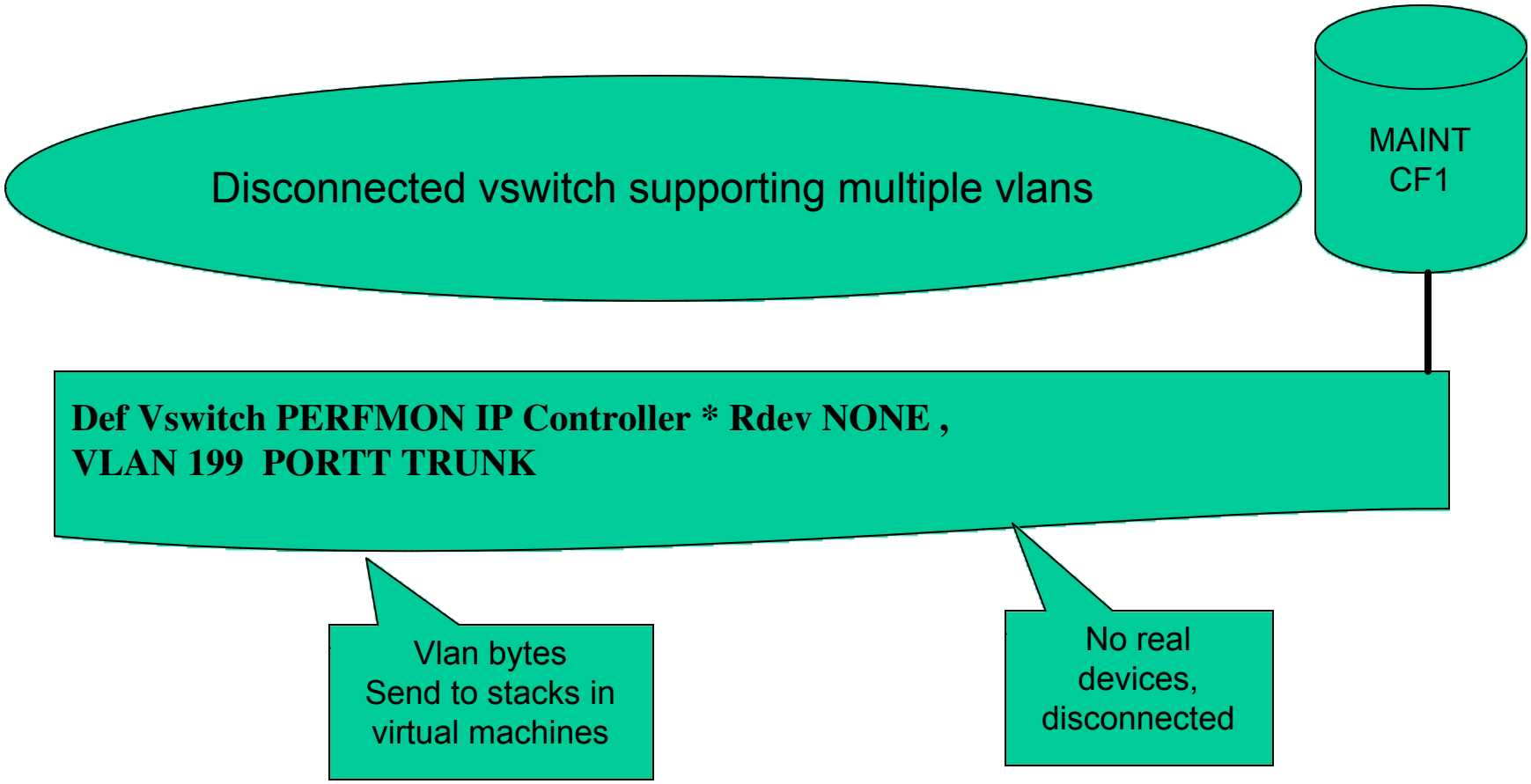
Network design for SNMP data collection



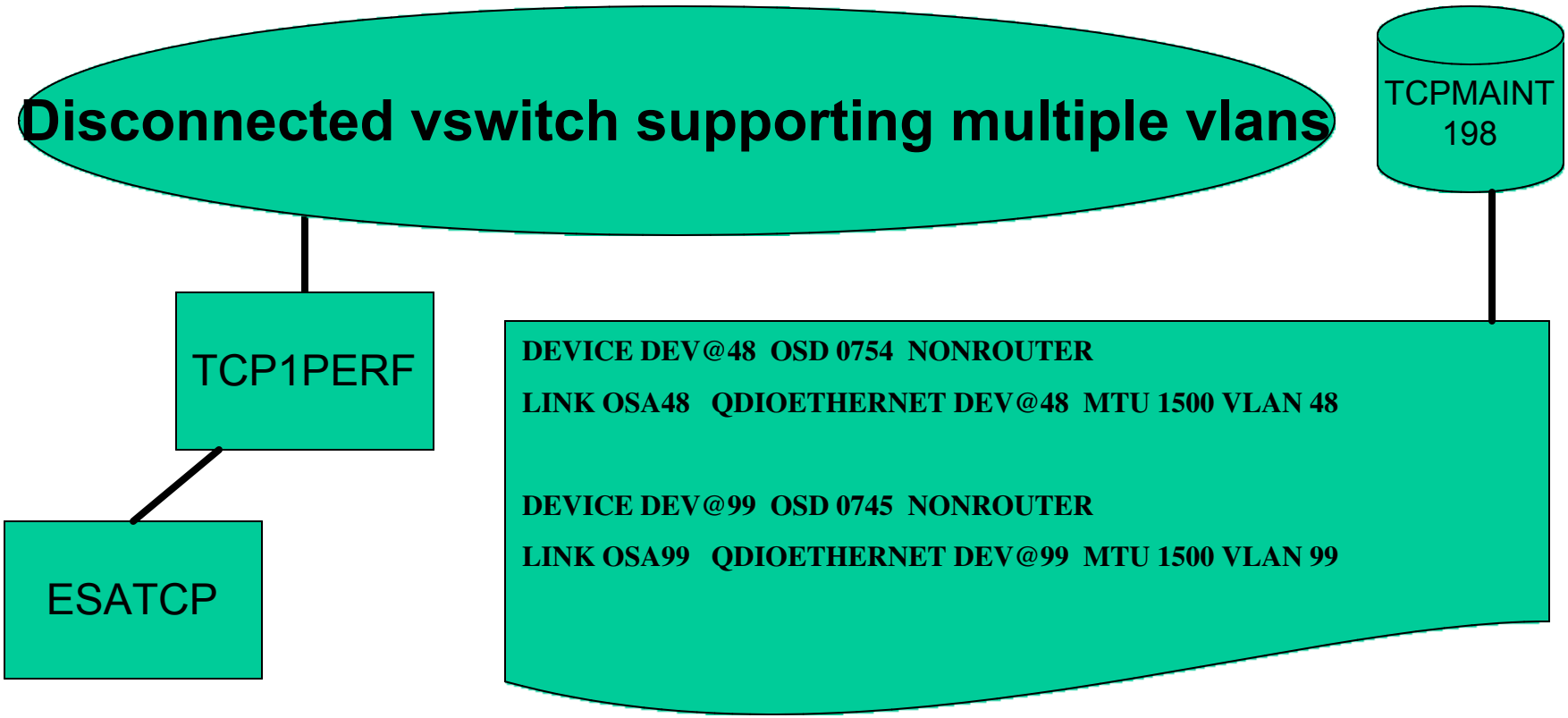
Configuration

- Vswitch defined in SYSTEM CONFIG with vlan
- VM TCP stack configued in the TCP1PERF virtual machine supporting vlan.
  - TCPMAINT 198
- Velocity collection configured in the ESATCP 191.
- Linux configured in /etc/sysconfig/network support vlan.
  - Virtual machine *LINUXAA* exemplified.

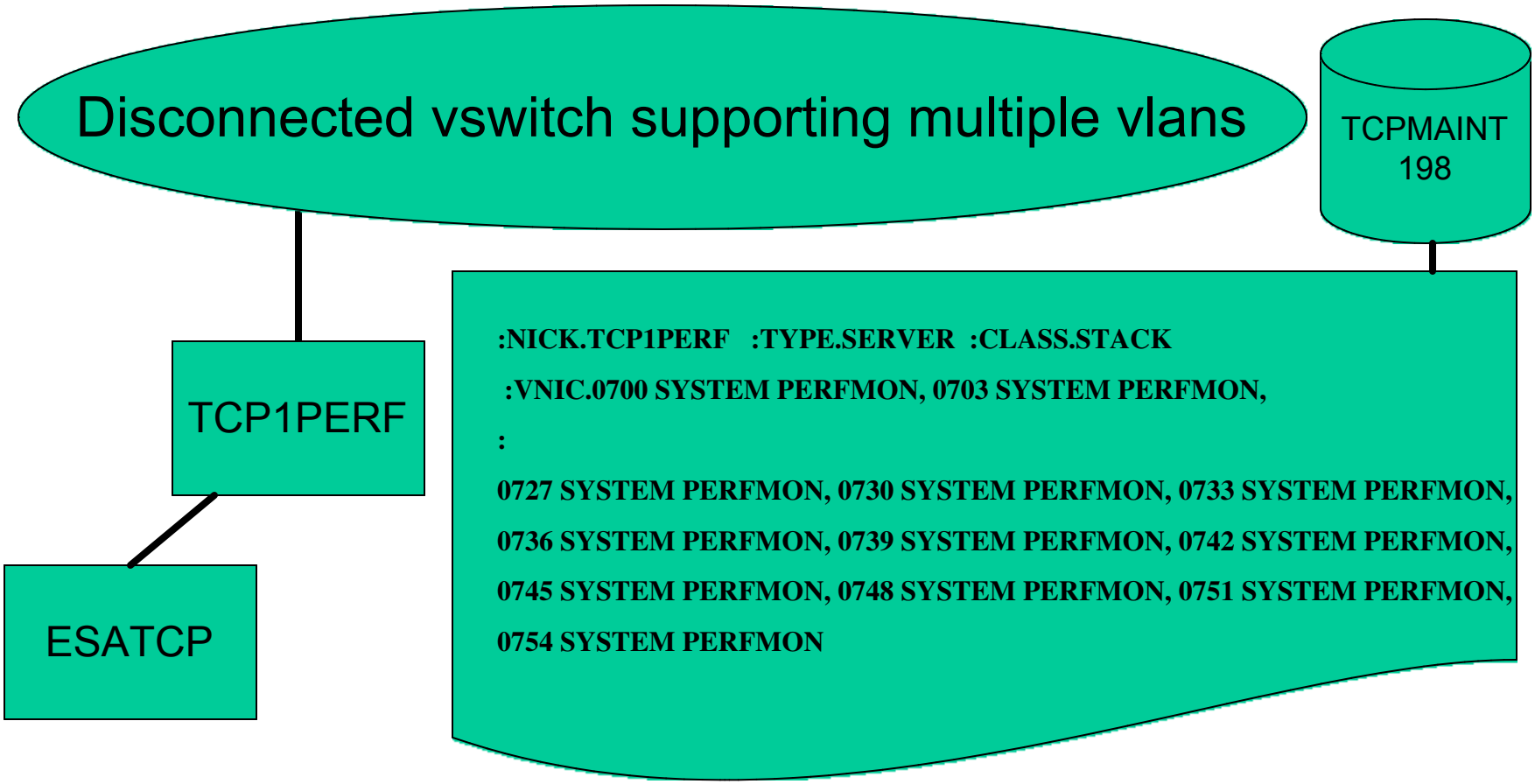
PERFMON Vswitch Defined in SYSTEM CONFIG



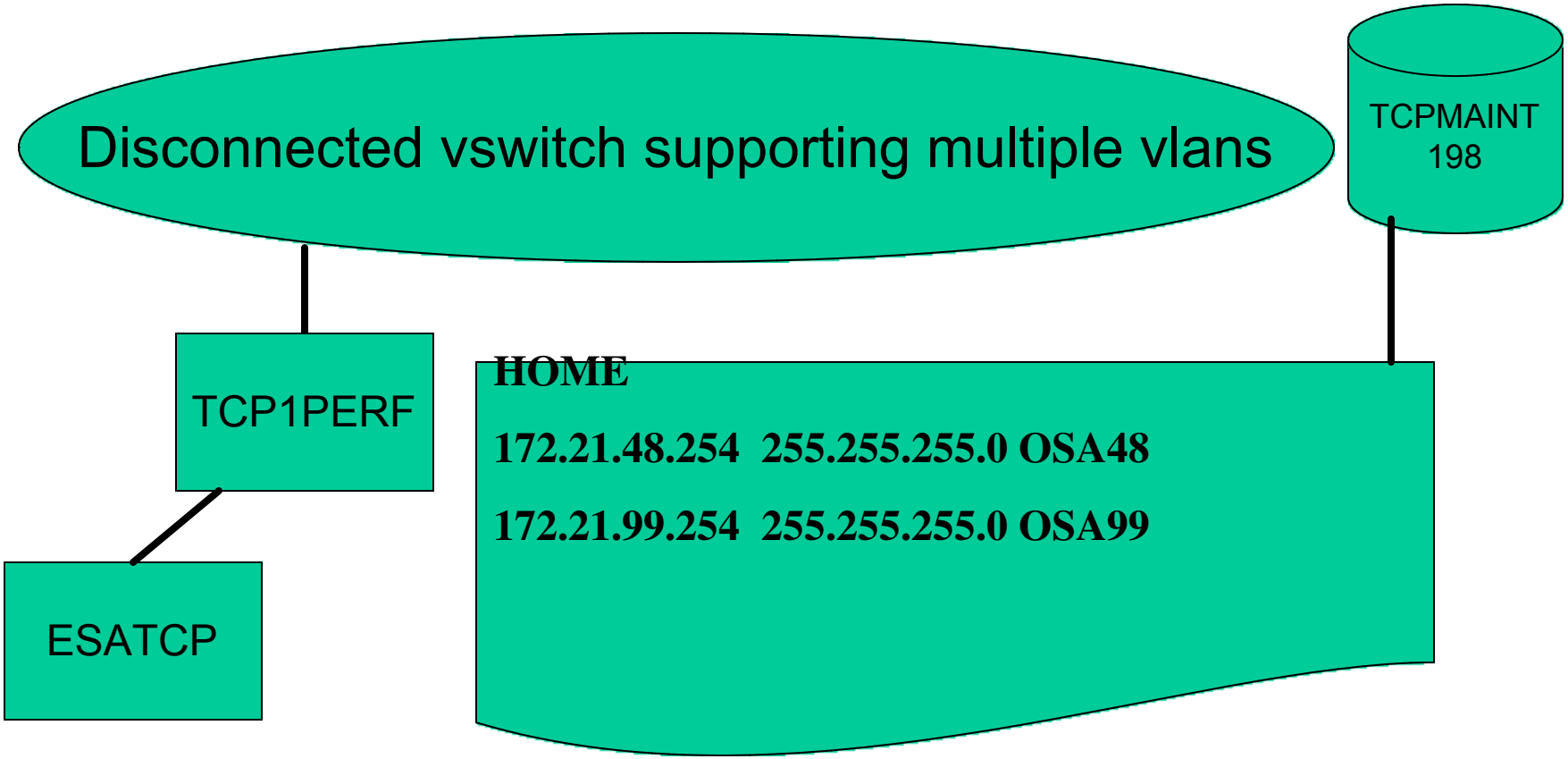
The TCP1PERF TCPIP configuration profile: DEVICE/LINK combo



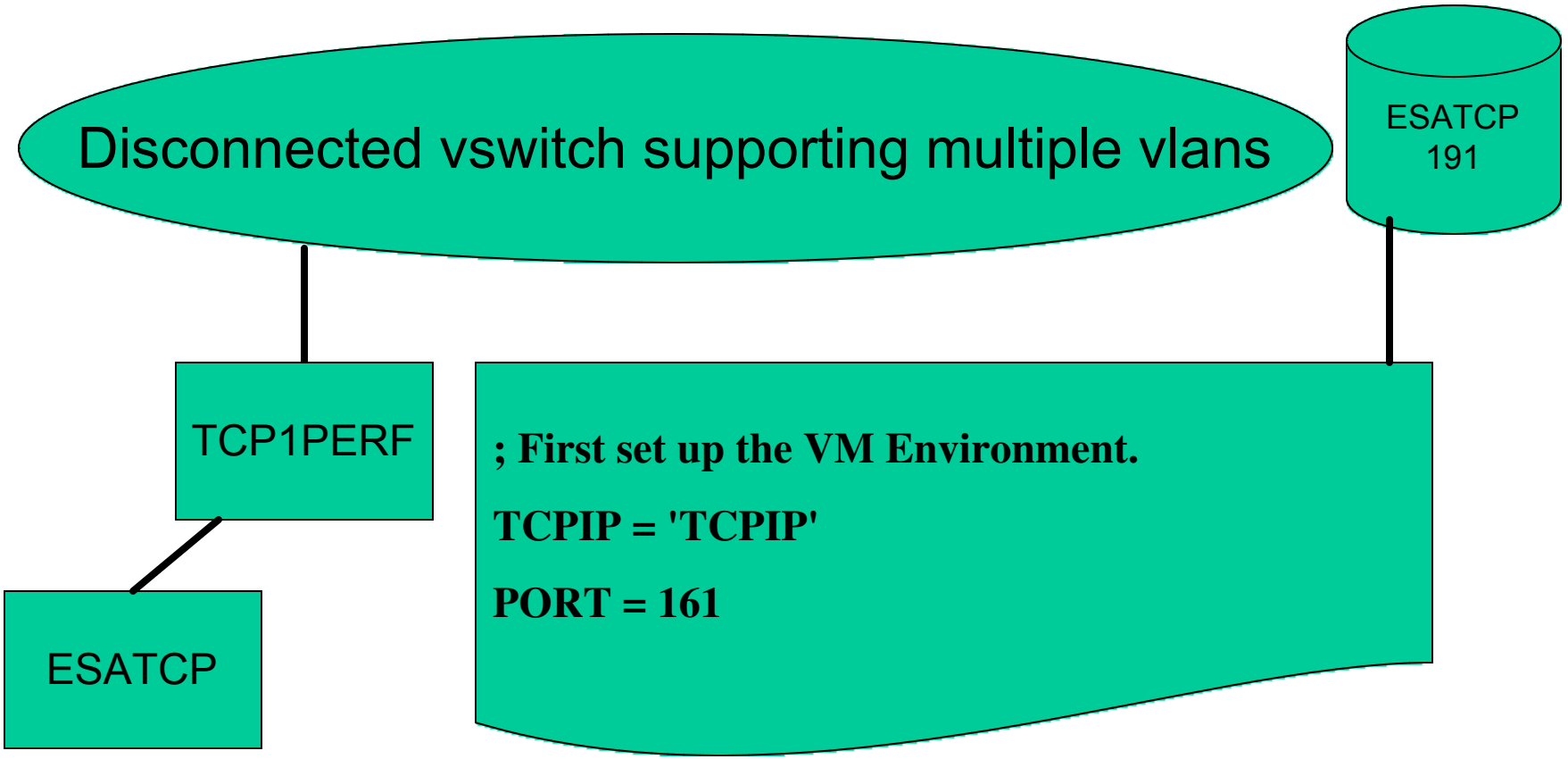
The SYSTEM DTCPARMS configuration profile



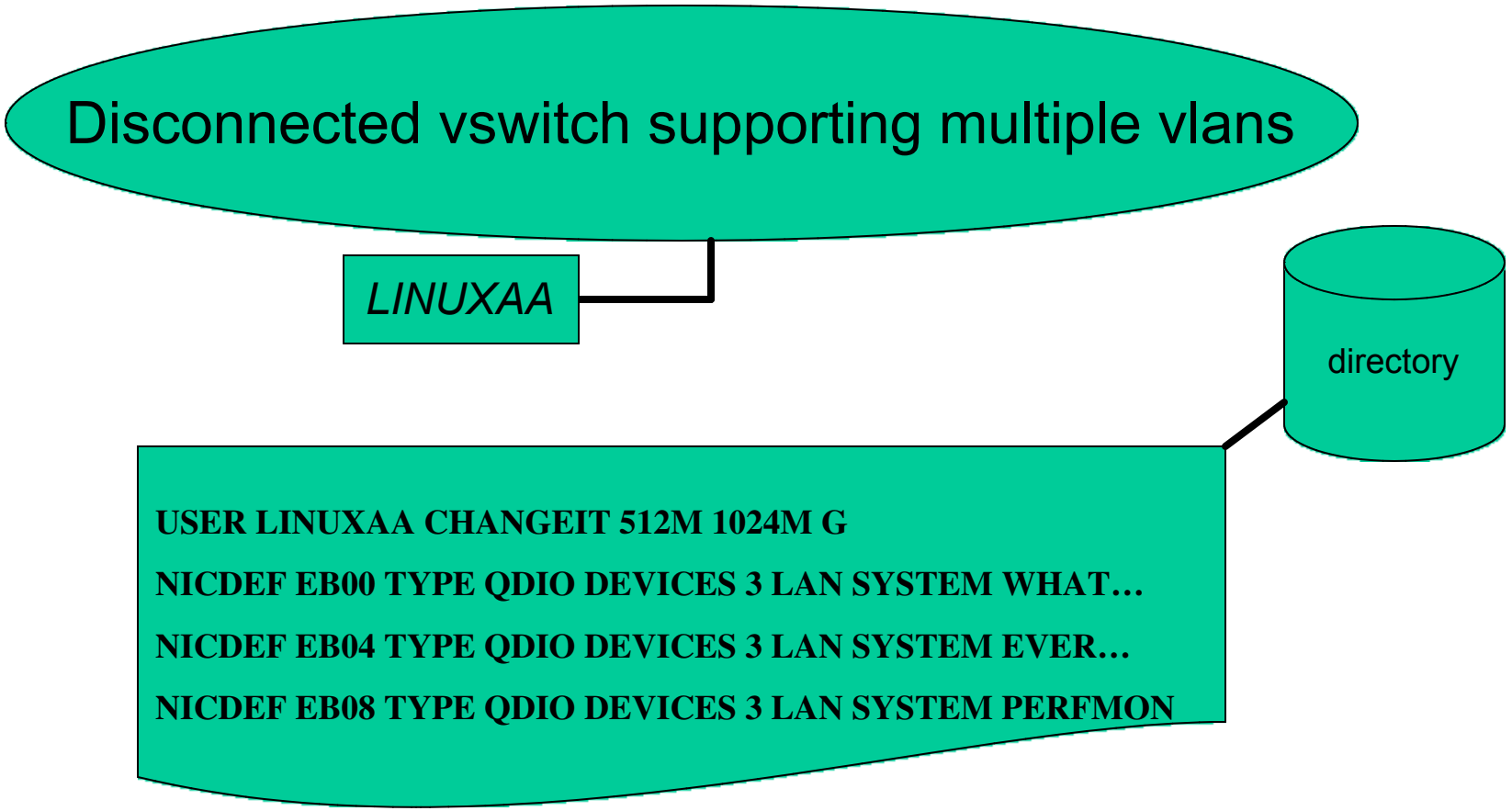
The TCP1PERF TCPIP configuration profile: HOME statements



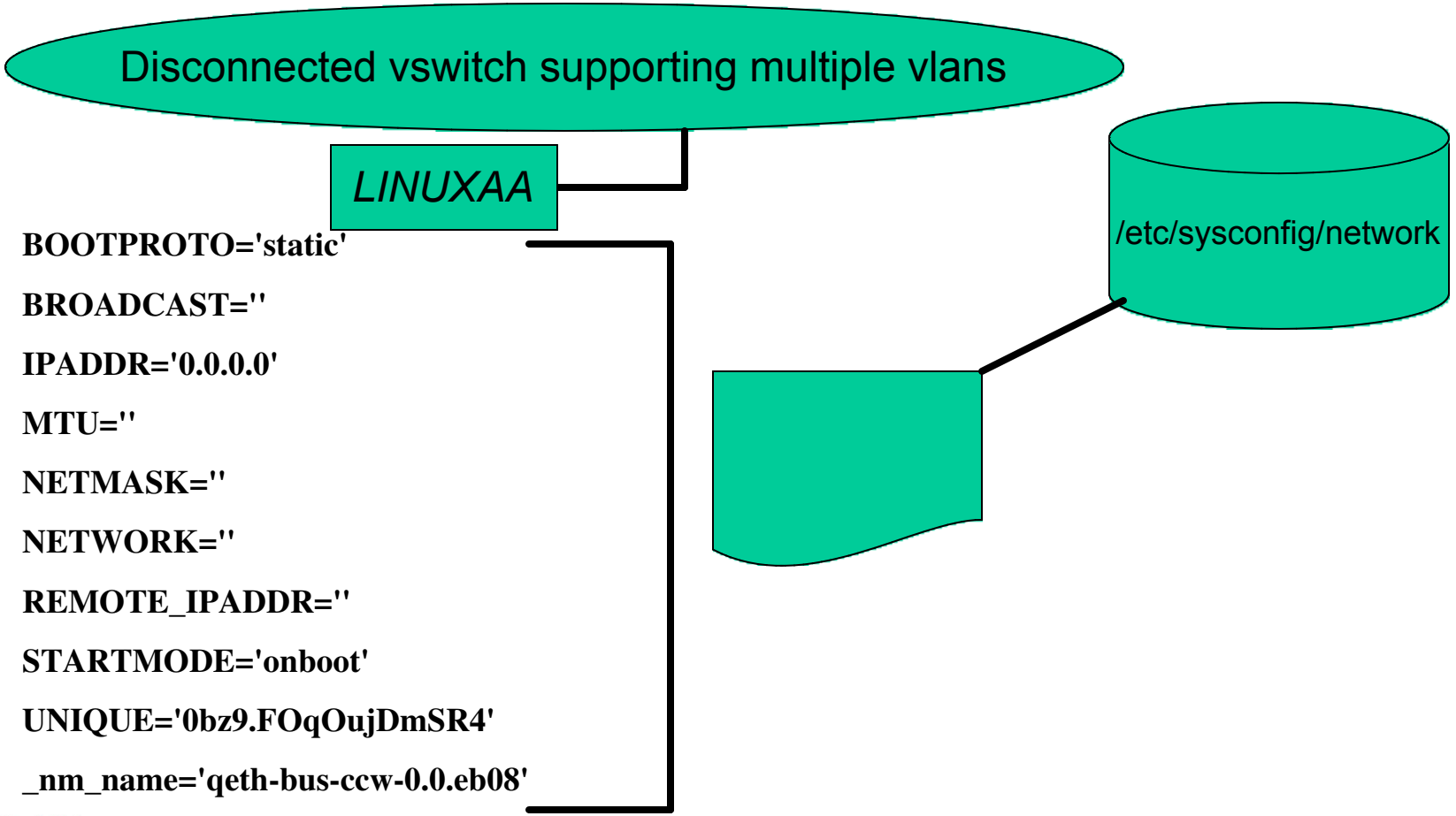
The ESATCP PARM configuration profile for ESATCP



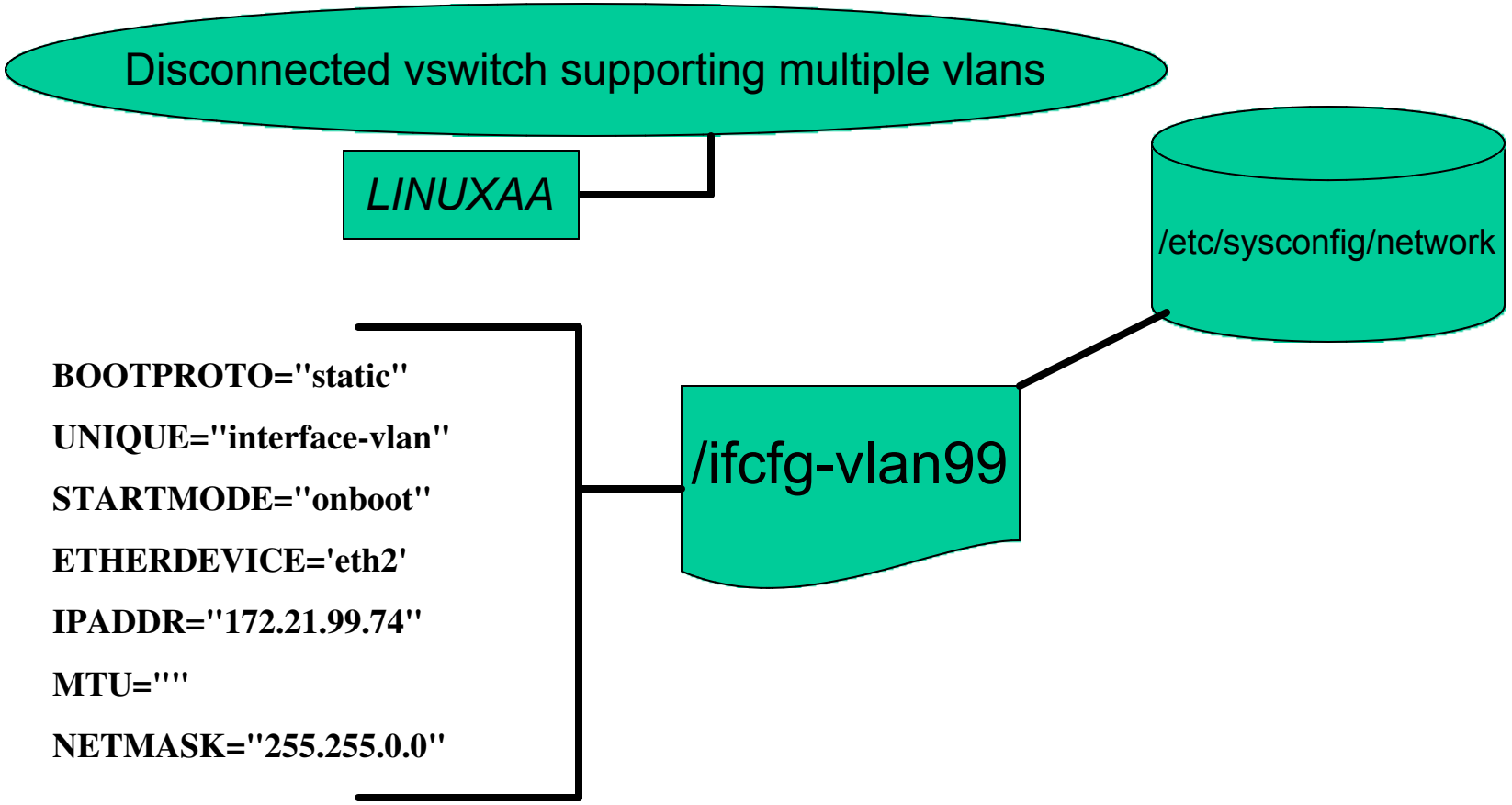
Nicdefs defined in the directory of *LINUXAA*



Network definition for ifcfg-qeth-bus-ccw-0.0.eb08



Network definition for ifcfg-vlan99



dmesg Showing the start up messages

Linux version 2.6.16.46-0.12-default (geeko@buildhost) (gcc version 4.1.2 20070115 (prerelease) (SUSE Linux)) #1 SMP Thu May 17 14:00:097

qeth: Device 0.0.eb08/0.0.eb09/0.0.eb0a is a Guest LAN QDIO card (level: V534)

with link type GuestLAN QDIO (portname: suseport)

qeth: Hardware IP fragmentation not supported on eth0

qeth: VLAN enabled

qeth: Multicast enabled

qeth: IPV6 enabled

qeth: Broadcast enabled

qeth: Using SW checksumming on eth0.

qeth: Outbound TSO not supported on eth0

802.1Q VLAN Support v1.8 Ben Greear <greearb@candelatech.com>

All bugs added by David S. Miller <davem@redhat.com>

vlan99: add 33:33:00:00:00:01 mcast address to master interface

vlan99: add 33:33:ff:00:00:2b mcast address to master interface

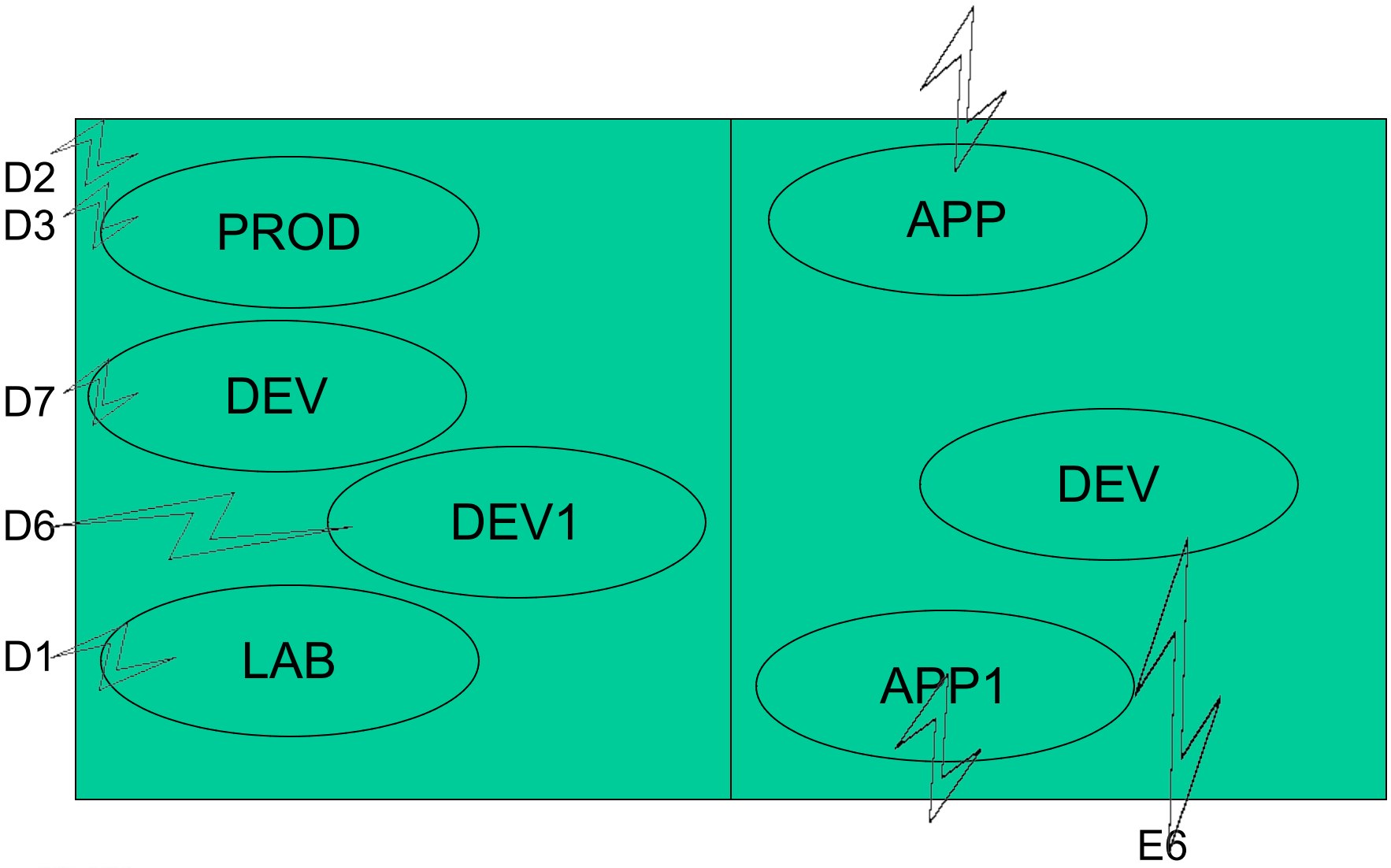
vlan99: add 01:00:5e:00:00:01 mcast address to master interface

Vlan99 interface will be automatically started by the network script. The script will load the 8021q module required for vlan support.

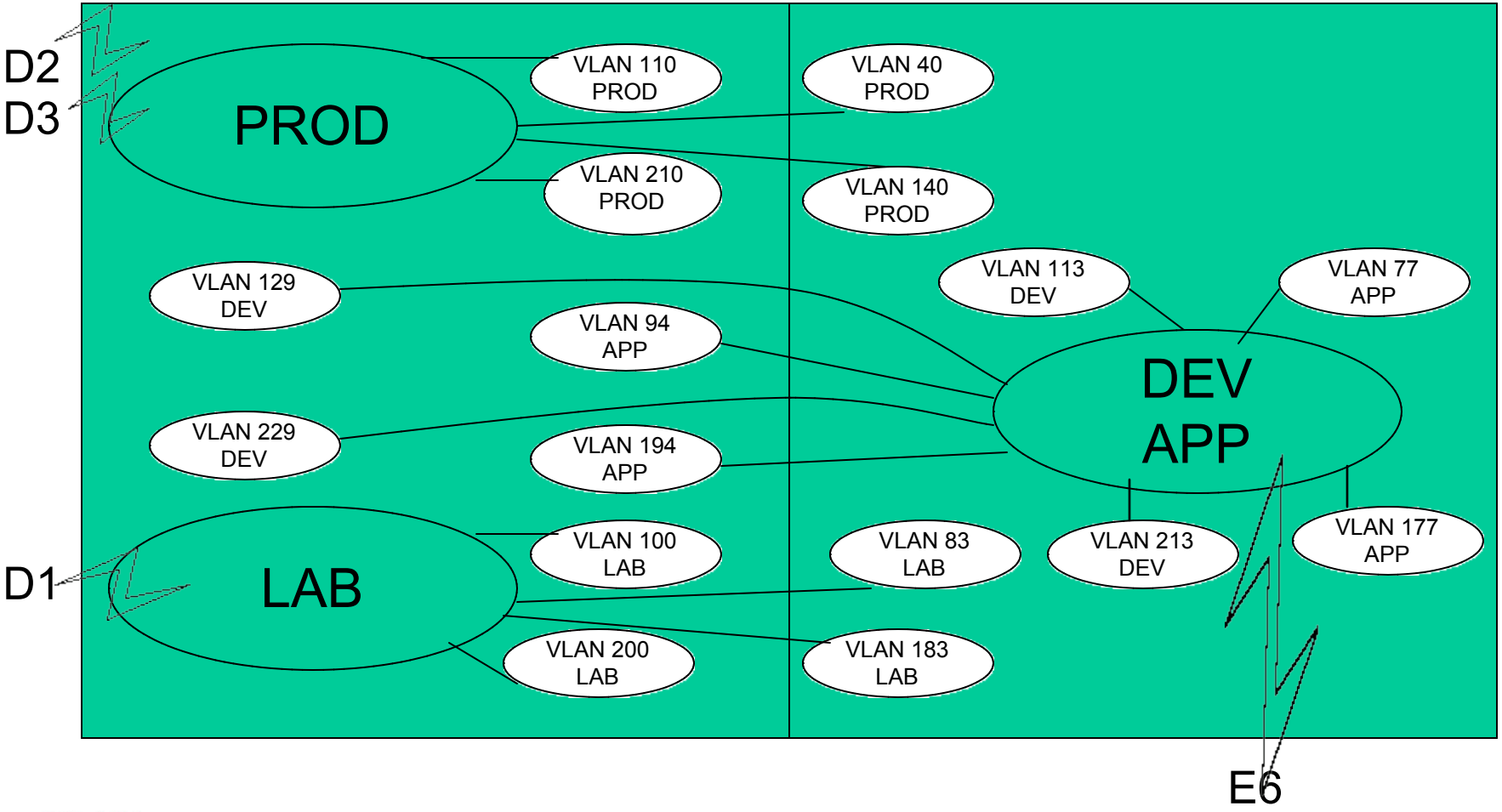
Velocity Report Based on SNMP Data

```
Screen: ESAHST1  Customer LPAR1              ESAMON 3.7.0 02/20 15:14-15:15
1 of 1  Linux HOST Software Analysis Report    NODE * LIMIT 500    2094 4FFEF

            <--Software Program----------> <CPU Seconds> CPU  Storage(K)
Time    Node    Name     ID   Type    Status  Total Intrval Pct   Current
-------- -------- -------- ----- ------- ------- ----- ------- ----- ---------
15:15:00 LINUXAA  java    31551 Applic  ResWait  3292    6.70 11.24   2065422
            snmpd    18449 Applic  Running  1218    0.08  0.13     7019
            httpd     3409 Applic  ResWait   127    0.05  0.09    236035
            pdflush     98 Applic  ResWait    10    0.03  0.04         0
            events/0     4 Applic  ResWait   145    0.03  0.04         0
            *Totals*     0 Unknown Unknown     0    6.88  0.23   2736872
      LinuxZZ  java    11216 Unknown ResWait    97    0.16  0.26    475155
            httpd     7448 Applic  ResWait    62    0.02  0.04    237467
            httpd     7446 Applic  ResWait    62    0.02  0.04    237507
snmpd    2298 Applic  Running  1153    0.02  0.04     6779
            *Totals*     0 Unknown Unknown     0    0.27  0.01   1371845
```

Consolidating to fewer OSAs by using vlans



Consolidating to fewer OSAs by using vlans

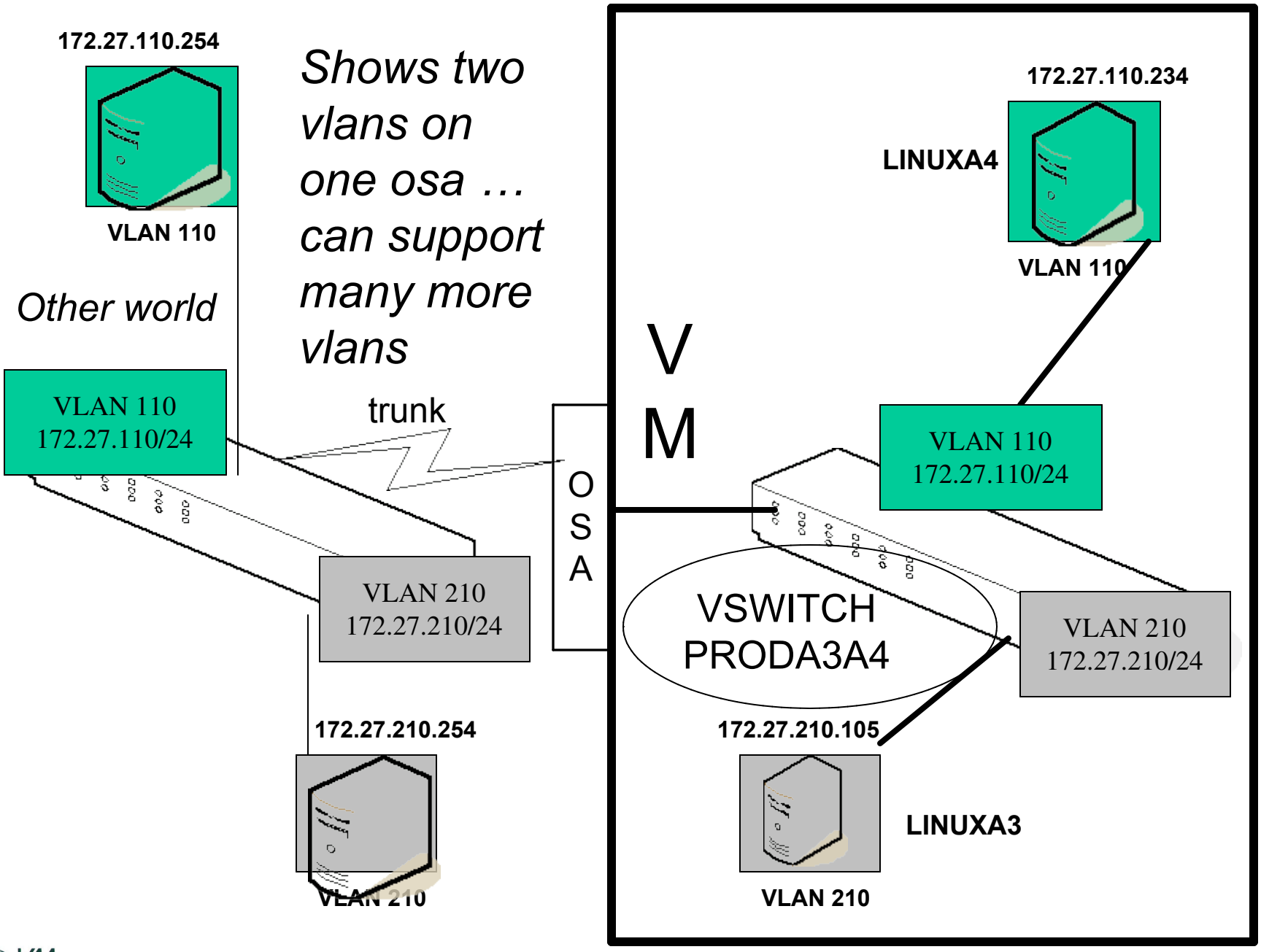


Condensed OSAs through vlan usage

- OSA is trunked to physical switch
- Supports multiple vlans in agreement with physical switch configuration
- Showing:
  - Vswitch definition
  - 2 Linuxen on same vswitch with different vlans
  - Isolated
  - *Vswitch is defined as layer 2 (vlan works either in layer 2 or 3 …)*

Notes

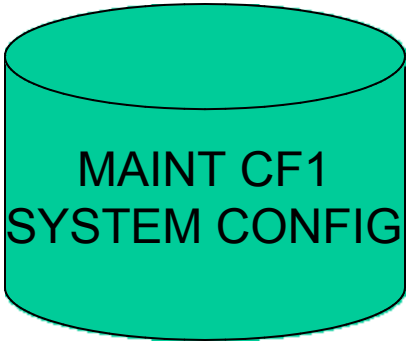
- Virtual machines connect to vswitch as trunk
  - *But in some configurations not required (including this one)*
  - *will change … maybe*
  - *Definitions on vswitch describe how the virtual machines connect to vswitch, not how vswitch connects to OSA*
- Virtual machine connecting to vswitch as trunk must deal with tags
- If only supporting one ip address per interface use access connection to vswitch


172.27.110.254
VLAN 110
Other world
Shows two vlans on one osa … can support many more vlans
VLAN 110
172.27.110/24
trunk
VLAN 210
172.27.210/24
172.27.210.254
VLAN 210
OSA
V
M
172.27.110.234
LINUXA4
VLAN 110
VLAN 110
172.27.110/24
VSWITCH
PRODA3A4
VLAN 210
172.27.210/24
172.27.210.105
LINUXA3
VLAN 210

Define Vswitch PRODA3A4 Ethernet nogroup Rdev D200 D300 Controller

```
VSWITCH SYSTEM PRODA3A4  Type: VSWITCH Connected: 8    Maxconn:
INFINITE
  PERSISTENT  RESTRICTED    ETHERNET              Accounting: ON
  VLAN Aware  Default VLAN: 0999    Default Porttype: Trunk   GVRP: Enabled
          Native  VLAN: 0999
  MAC address: 02-01-0A-00-00-03
  State: Ready
  IPTimeout: 5        QueueStorage: 8
  Isolation Status: OFF
  RDEV: D200.P00 VDEV: D200 Controller: DTCVSW2
    VSWITCH Connection:
     RX Packets: 38138851   Discarded: 3        Errors: 0
     TX Packets: 37995827   Discarded: 0        Errors: 0
     RX Bytes: 15122277122        TX Bytes: 8681966948
     Device: D202  Unit: 002   Role: DATA      vPort: 0001  Index: 0001
       Unicast IP Addresses:
172.27.110.254      MAC: 00-03-BA-5B-24-81 Remote
  RDEV: D300.P00 VDEV: D300 Controller: DTCVSW1  BACKUP
```

QUERY VSWITCH DETAILS



Query vswitch details

Adapter Owner: LINUXA3 NIC: EB00.P00 Name: suseport

Porttype: Trunk

RX Packets: 8096888 Discarded: 0 Errors: 0

TX Packets: 152976 Discarded: 8 Errors: 0

RX Bytes: 7382934749 TX Bytes: 16211529

Device: EB02 Unit: 002 Role: DATA vPort: 0069 Index: 0006

VLAN: 0210

Options: Ethernet Broadcast

Unicast MAC Addresses:

02-01-0A-00-00-1D IP: 172.27.210.105

Multicast MAC Addresses:

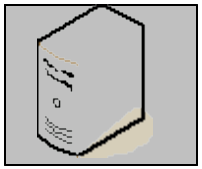
01-00-5E-00-00-01 IP: 224.0.0.1

33-33-00-00-00-01 IP: FF02::1

33-33-FF-00-00-1D IP: FF02::FF00:1D

QUERY VSWITCH DETAILS

172.27.210.105



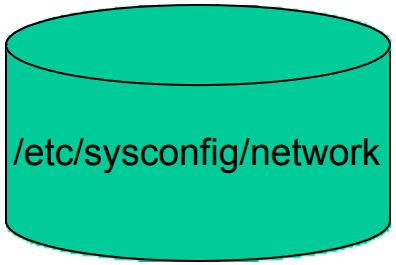
LINUXA3

VLAN 210

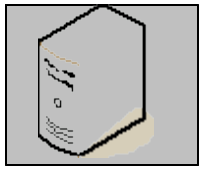
/etc/sysconfig/network/ifcfg-qeth-bus-ccw-0.0.eb00

```
BOOTPROTO='static'
BROADCAST=''
IPADDR='0.0.0.0'
MTU=''
NETMASK=''
NETWORK=''
REMOTE_IPADDR=''
STARTMODE='onboot'
UNIQUE='0bz92%RFZc#OmSR4'
nm_name='qeth-bus-ccw-0.0.eb00'
```

Dummy it up



172.27.210.105



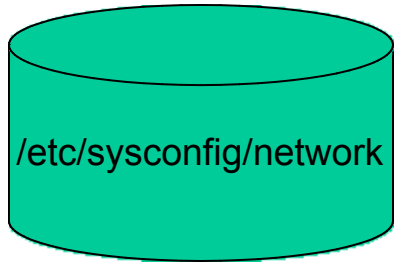
LINUXA3

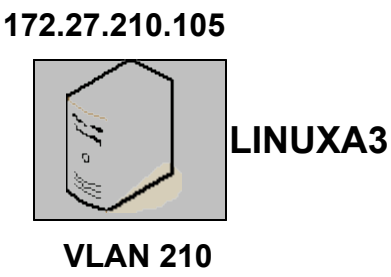
VLAN 210

LINUXA4 configuration

/etc/sysconfig/network/ifcfg-vlan210

```
ETHERDEVICE='eth0'
MTU=''
IPADDR='172.27.210.105'
NETMASK='255.255.254.0'
NETWORK='172.27.210.0'
STARTMODE=onboot
VLAN='YES'
```



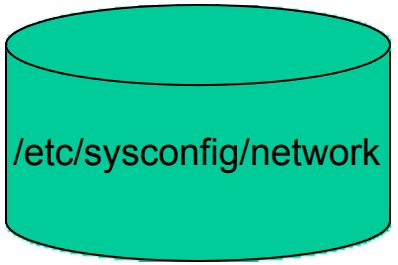


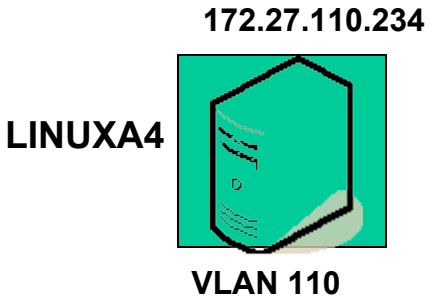
LINUXA3 configuration

/etc/sysconfig/network/ifcfg-qeth-bus-ccw-0.0.eb00

```
BOOTPROTO='static'
BROADCAST=''
IPADDR='0.0.0.0'
MTU=''
NETMASK=''
NETWORK=''
REMOTE_IPADDR=''
STARTMODE='onboot'
UNIQUE='45%?SDdd%FR_.FOmSR9'
_nm_name='qeth-bus-ccw-0.0.eb00'
```

Dummy it up

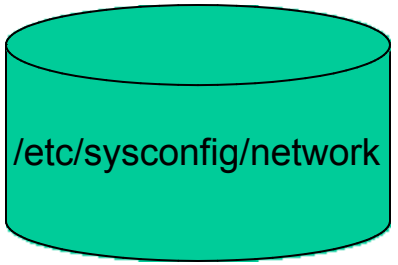


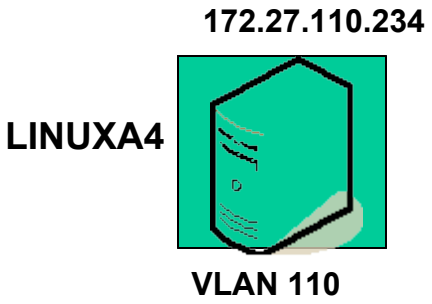


LINUXA4 configuration

/etc/sysconfig/network/ifcfg-vlan110

```
ETHERDEVICE='eth0'
MTU=''
IPADDR='172.27.110.234'
NETMASK='255.255.248.0'
NETWORK='172.27.210.0'
STARTMODE=onboot
VLAN='YES'
```





LINUXA4 configuration

Summary

- Z/VM, vswitch and vlan
- Different uses and implementations:
  - Hipersocket
  - Disconnected vswitch
  - Connecting over qdio osa devices

Thank you to …

- Alan Altmark
- Dave Jones
- Dominic Coulombe