SHARE
Technology · Connections · Results

Tending the SANity of the Flock

SAN Experiences at Nationwide

Rick Troth <trothr@nationwide.com>

August 14, 2007

SHARE 109 session 9217

Disclaimer

The content of this presentation is informational only and is not intended to be an endorsement by Nationwide Insurance. Each site is responsible for their own use of the concepts and examples presented.

Or in other words: Your mileage may vary. “It Depends.” Results not typical. Actual mileage will probably be less. Do not fold, spindle, or mutilate. Not to be taken on an empty stomach.

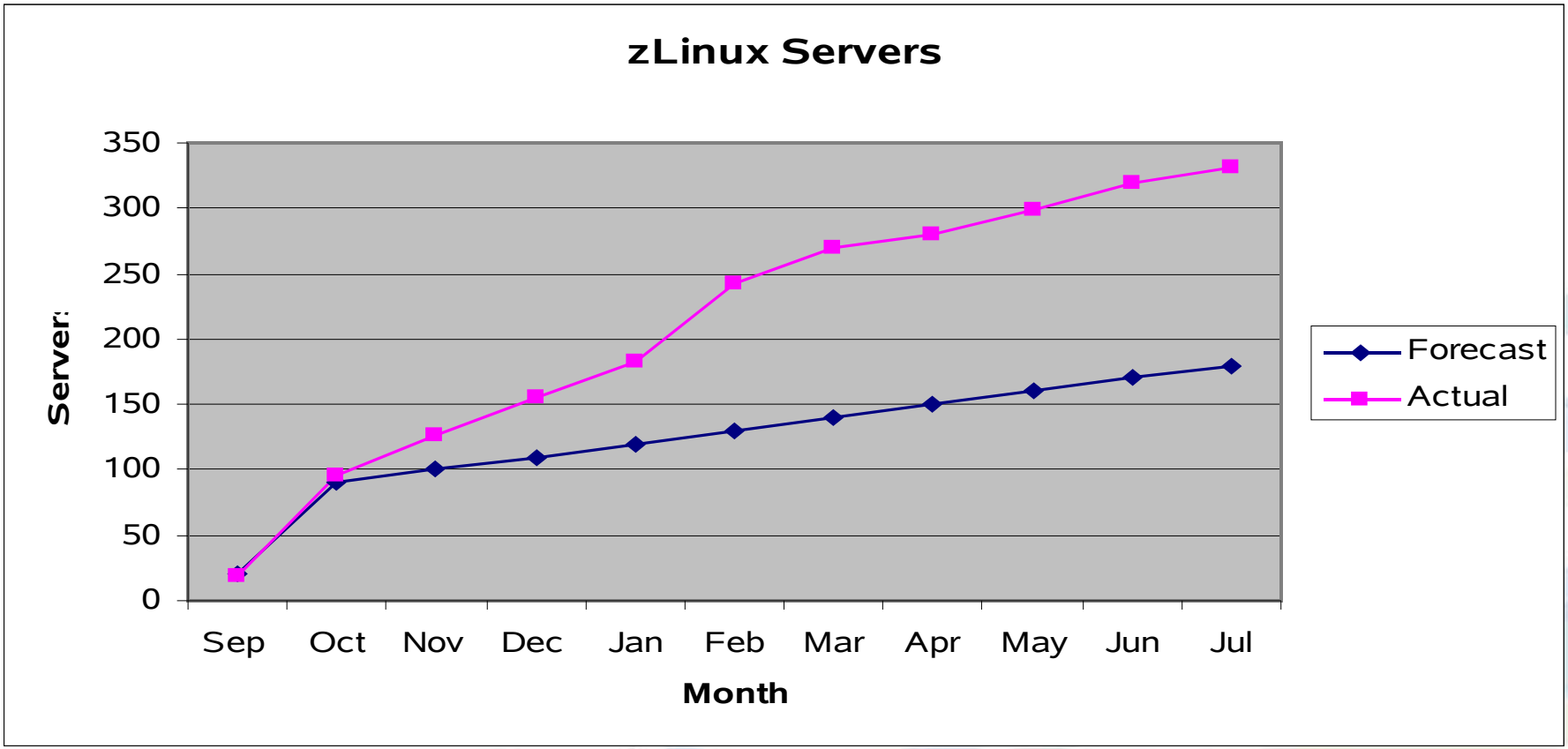
When in doubt, ask!

A New Iceberg

- The issue: ECKD constrained
- The solution: put some content on SAN
- The implementation
- The results …

Linux has Grown Fast!!

*And I thought we were busy **before** we got Linux!*
Rick Barlow, Aug 1, 2006



Using up our ECKD Space

ECKD constrained

- Cost (we could buy more, but …)
- Interconnection / Interoperability
- Different Granularity (than other Linux)

So … put “user files” onto SAN

Stretching the Shared Disk Envelope

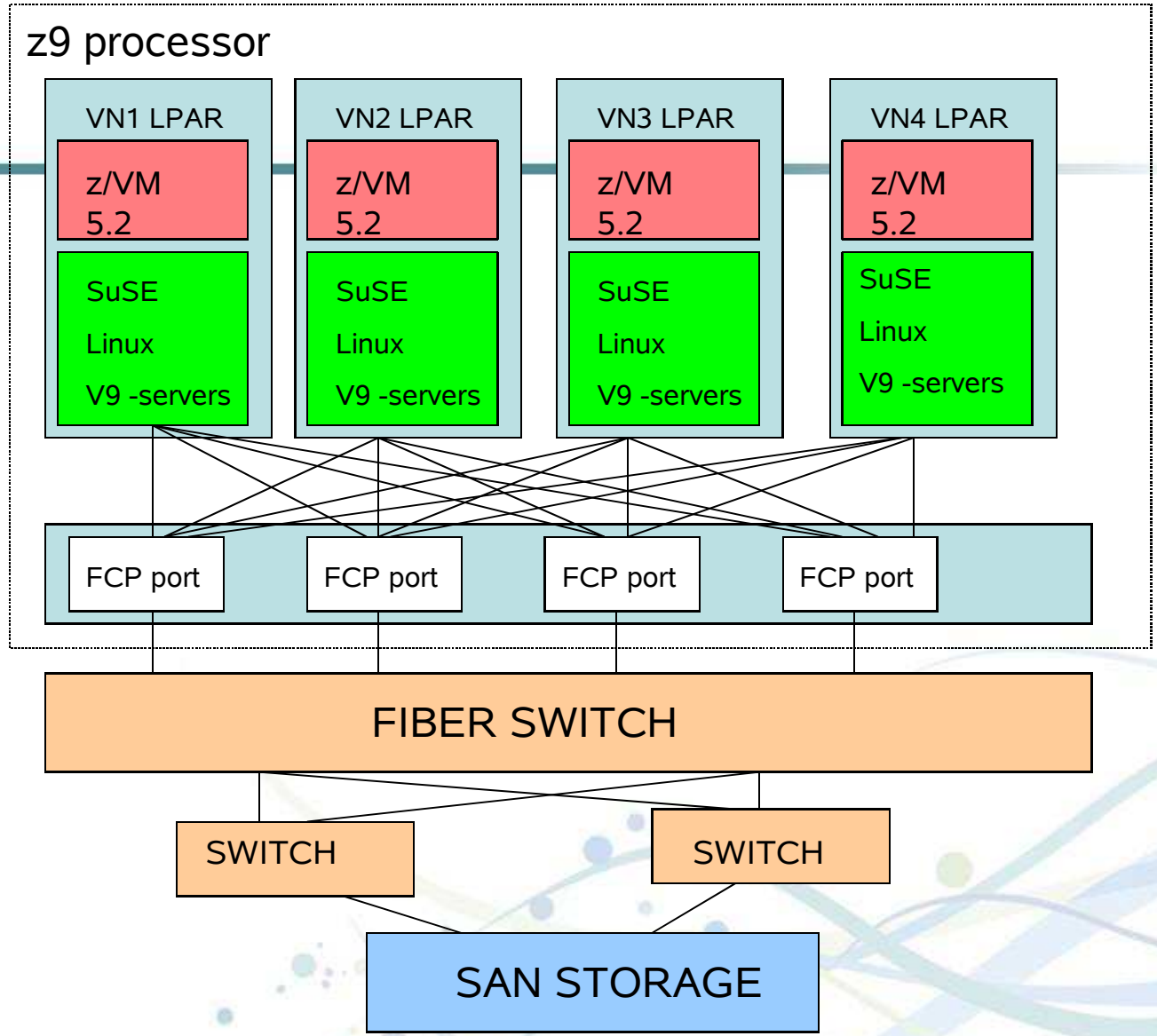
- Can we share SAN volumes? simultaneously? across unlike systems?
- Will discuss shared filesystems more, and especially read-only root, later in the week (session 9216)

Storage Area Network

- common disk interface for all large systems, not just IBM System z
- opportunity to share disk-resident content across platforms
- common skills and work for disk management staff
- potential for more cost effective data storage options (but why?)

Basic Storage Requirements

- Replication (more than failover)
- Multipath (failover within storage space)
- Backup (multiple points of recovery)
- Security / Isolation


z9 processor
VN1 LPAR
z/VM 5.2
SuSE Linux V9 -servers
VN2 LPAR
z/VM 5.2
SuSE Linux V9 -servers
VN3 LPAR
z/VM 5.2
SuSE Linux V9 -servers
VN4 LPAR
z/VM 5.2
SuSE Linux V9 -servers
FCP port
FCP port
FCP port
FCP port
FIBER SWITCH
SWITCH
SWITCH
SAN STORAGE

Storage Area Network

What SAN is not …

- SAN is not NAS
- SAN is not a networked filesystem
  - Unix NFS protocol
- SAN is not “mapped drives”
  - Windows SMB protocol

Storage Area Network

What SAN is …

- External Storage with Long Wires
- Talks like SCSI Disk
- Works like Mainframe Disk (sort of)
- Separate from other networks

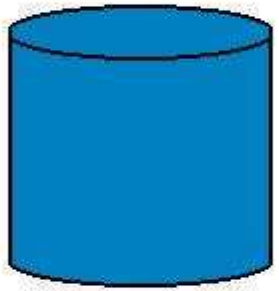
Storage Area Network

Picking out Furniture …

- Point to Point
- Arbitrated Loop
- Switched Fabric ←

ECKD mainframe disk

- z/VM (CP)
- z/VM (CMS)
- Linux
- VSE
- z/OS



ECKD

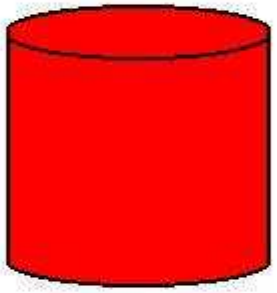
extended count/key/data

(and tracks & records)

ECKD traffic includes non-data

FBA mainframe disk

- **z/VM (CP)**
- **z/VM (CMS)**
- **Linux**
- **VSE**
- z/OS



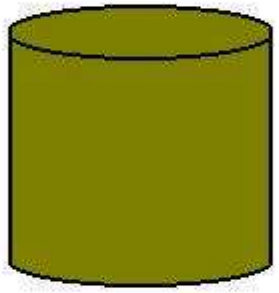
FBA

fixed blocks / just data

z/OS cannot use FBA disks

SAN disk or SCSI disk

- z/VM (CP)
- z/VM (CMS)
- Linux
- VSE
- **Solaris, AIX, HP-UX**
- **Windows**



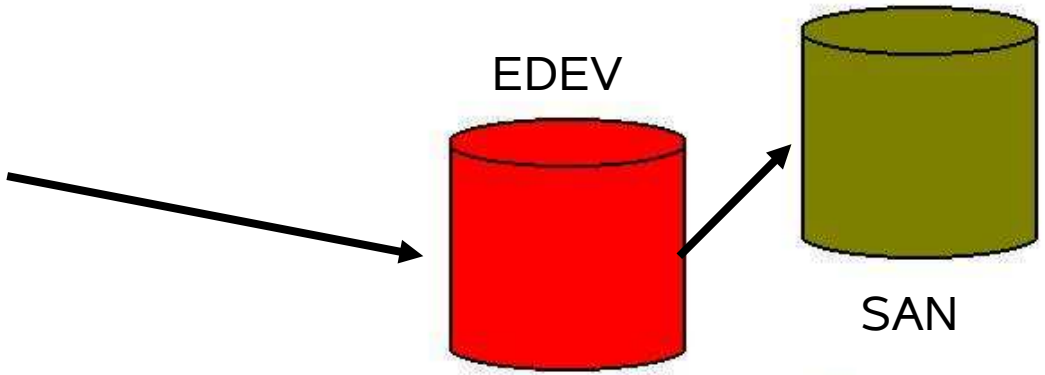
SAN

fixed blocks / just data

SAN disk presented same as SCSI

SAN is FBA … sort of

- **z/VM (CP)**
- **z/VM (CMS)**
- **Linux**
- **VSE**



EDEV makes SAN look like FBA (9336)

They have the same structure.

If you must partition …

disk type	*driver*	*format with*	*partition with*
ECKD	`dasd`	`dasdfmt`	`fdasd`
FBA	`dasd`		`fdisk`
SAN	`zfcp+scsi`		`fdisk`
EDEV	`dasd`		`fdisk`

“The Nucleus Got Bigger”

- 31-bit support dropped
  CP Nucleus should have gotten smaller
- SAN support (EDEV) added
  CP Nucleus actually got *larger*

IBM leveraged AIX driver code – cool!

FCP attached to VM

```
/* make a SAN volume work like an FBA disk */
'CP SET EDEV FF02 TYPE FBA ATTR SCSI' ,
   'FCP_DEV 010A WWPN 50060482D52CC7F2 LUN 0002000000000000' ,
   'FCP_DEV 020A WWPN 50060482D52CC7FD LUN 0002000000000000'
'CP VARY ON FF02'

/* how does it look to CP? */
'CP Q DASD DETAILS FF02'
'CP Q 10A 20A'
...
FF02  CUTYPE = 6310-80, DEVTYPE = 9336-10, VOLSER = SAN002,
      CYLS = 91003, BLKS = 70709760
FCP  010A ATTACHED TO SYSTEM   0000 CHPID 50
FCP  020A ATTACHED TO SYSTEM   0000 CHPID 54
```

FCP attached to guest

- Requires Multipath Support in Linux
  - Two or more FCP “channels” per guest
- Demands Multipath Management
- Some Loss of Control (of guest storage)
- Coarse Grained Allocations
  - Sometimes okay with LVM
- Should it be N-port Virtualized?

FCP attached to guest

```
FCP   0100 ON FCP   0304 CHPID D1 SUBCHANNEL = 0018
      0100 DEVTYPE FCP          CHPID D1 FCP
      0100 QDIO ACTIVE          QIOASSIST ACTIVE
 …
      WWPN C05076FC7D000D90

FCP   0200 ON FCP   0404 CHPID D5 SUBCHANNEL = 0019
      0200 DEVTYPE FCP          CHPID D5 FCP
      0200 QDIO ACTIVE          QIOASSIST ACTIVE
 …
      WWPN C05076FC7D001110
```

```
cd /sys/bus/ccw/drivers/zfcp
echo 1 > $HBA/online
echo $WWPN > $HBA/port_add
echo $LUN > $HBA/$WWPN/unit_add

ls -l $HBA/$WWPN/$LUN/.
```

Storage Area Network

Picking out Appliances …

- EVMS
- MPIO+LVM2 ←

Can You Say “coalesce”?

- Combined Two Paths into One PV
  - `http://www.webster.com/dictionary/coalesce`
- “logical volume” in a different sense
  - Physical PV represents an I/O path
  - Logical PV is fed to LVM
- Modify `/etc/lvm/lvm.conf` accordingly

FCP attached to guest

Define paths manually or via YaST, then …

```
/etc/init.d/boot.multipath start
/etc/init.d/multipathd start
pvcreate /dev/mapper/3600604800001901006305330304538 32
vgcreate sanvg1 \
         /dev/mapper/3600604800001901006305330304538 32
lvcreate -L 4G -n sanlv1 sanvg1
```

Avoid gratuitous partition tables

- Common partitioning: zero, 1, 2, or 3
- Understood by either driver (scsi or dasd)
- Use PC “primary partitions”

But don’t!

- Partitioned requires double layer admin
- Non-partitioned gives simpler LVM admin
- Non-partitioned makes sharing easier

```
# cat /proc/partitions
  ...
    8     0    35354880 sda
    8    16    35354880 sdb
  253     0    35354880 dm-0
    8    32    35354880 sdc
    8    48    35354880 sdd
  253     1    35354880 dm-1
```

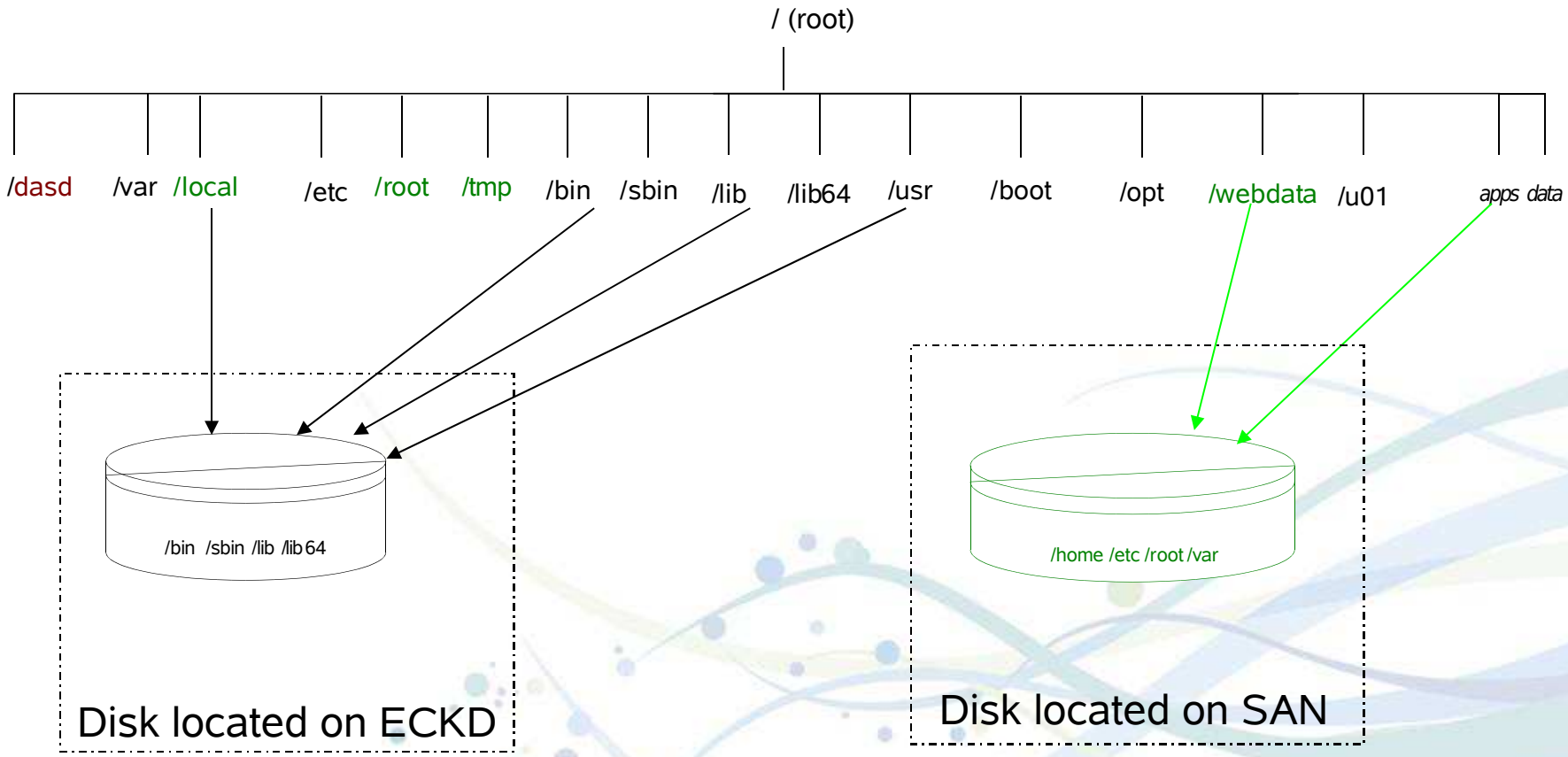
Multipath Management

```
:vmid.NZVJT002  :node.VS2
:chpid.51  :realwwpn.50050764016208c5
:dev.0304  :virtwwpn.c05076fc7d800c10
:sanframe.1822  :sandev.EE0
:targwwpn.50060482d52e4fa3  :lun.0027000000000000
:size.36G  :uuid.360060480000190101822533030454530
```

SAN is Seamless

- Operating systems stay on ECKD for now
- No Change of application file access
- Open-Ended storage capacity
- Secured at the hardware level

Mixed Media Methodology



Speaking of Security …

- LUNs are zoned and masked
- NPIV enabled for the fabric
- Without NPIV
  - One WWPN for CHPID
- With NPIV
  - Unique vWWPN per device

Stretching the Brand X Envelope

- It's all about interoperability …

SAN does for disk

what z/VM does for systems … sort of

Thank You!!

Richard Troth

Senior VM Systems Programmer

Nationwide Services Co., LLC

One Nationwide Plaza, MB-02-201
Columbus, OH 43215-2220
Voice: 1-614-249-7642
Cell: 1-614-849-8255
trothr@nationwide.com