Virtual Disk Architecture and the S/390 Linux Environment

Scott Ledbetter – StorageTek

scott_ledbetter@storagetek.com

SHARE San Francisco-- Session 9335

August 20, 2002

Trademarks

The following are trademarks or registered trademarks of Storage Technology Corporation: StorageTek, SnapVantage, Shared Virtual Array, SVA and Iceberg.

The following are trademarks or registered trademarks of the International Business Machines Corporation: IBM, RVA, RAMAC, zSeries, S/390, VM, z/OS and z/VM.
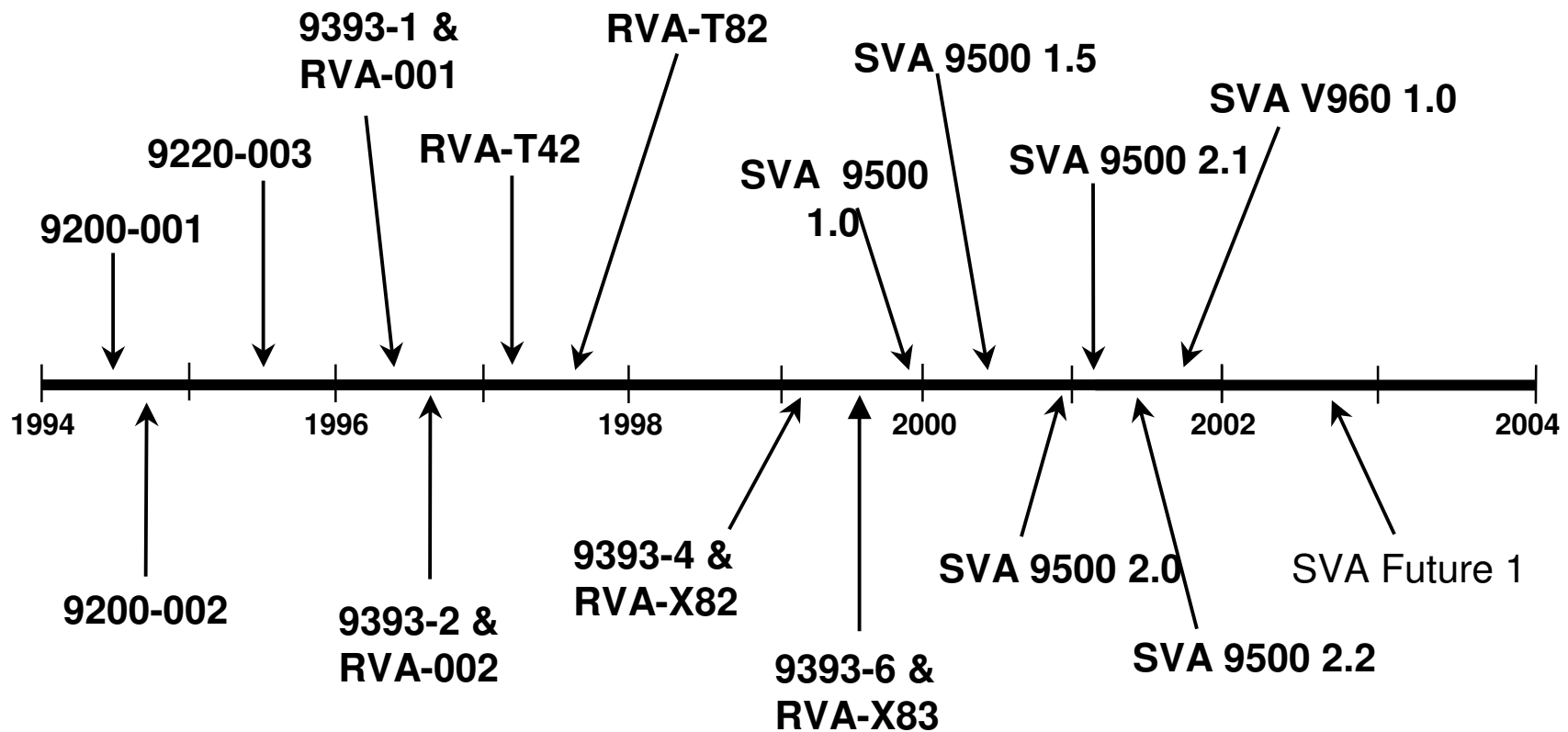
Linux is a registered trademark of Linus Torvalds.

Agenda

- StorageTek Virtual Disk Evolution
- StorageTek Virtual Disk Overview
- SnapShot and Instant Format
- PPRC
- S/390 Linux Volumes
- S/390 Linux Considerations for RVA/SVA
- SnapVantage Demo
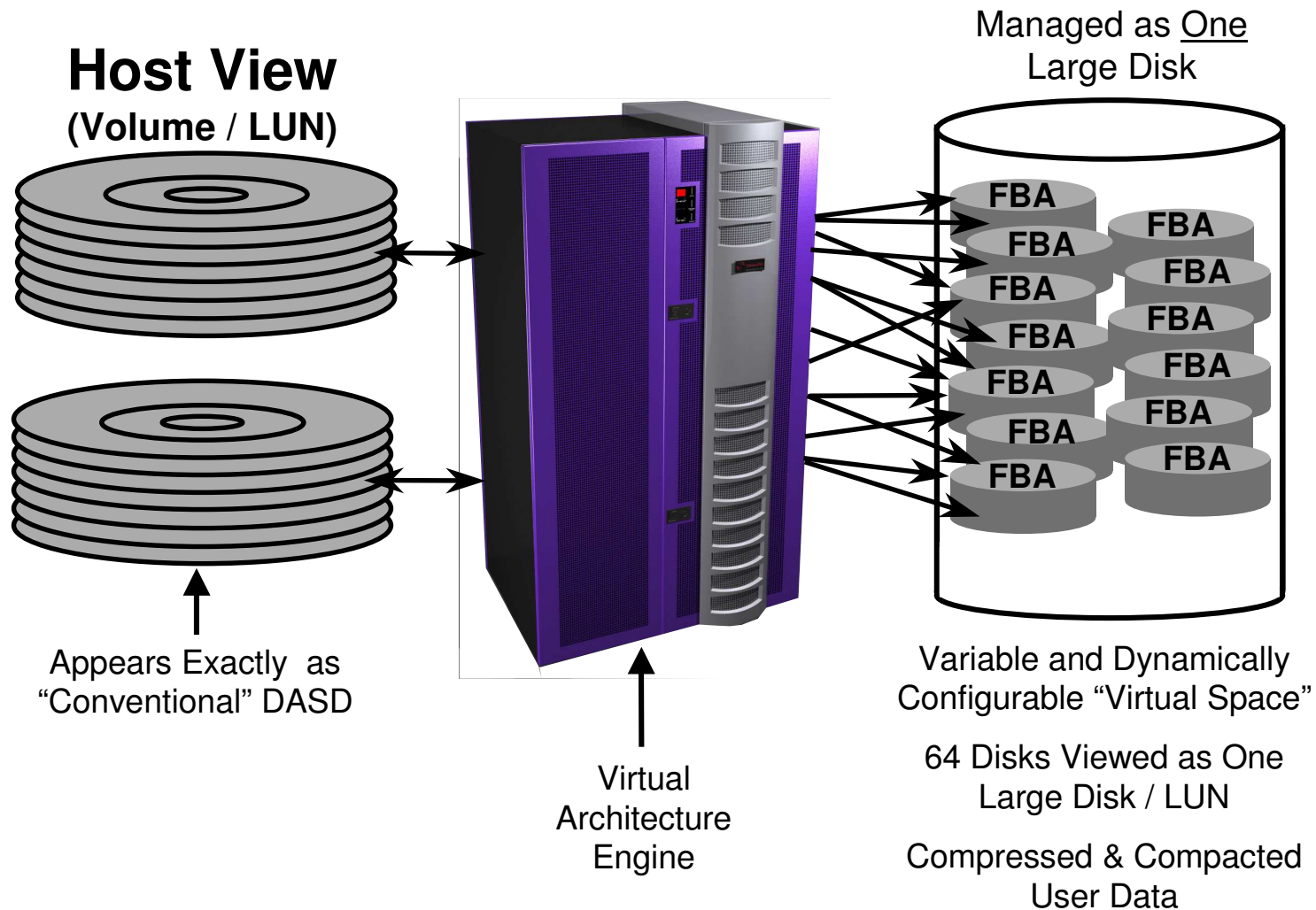- Q&A

StorageTek Virtual Disk Evolution

9393-1 &
RVA-001

RVA-T82

SVA 9500 1.5

9220-003

RVA-T42

SVA V960 1.0

SVA 9500 2.1

9200-001

SVA 9500
1.0

1994   1996   1998   2000   2002   2004

9200-002

9393-4 &
RVA-X82

SVA 9500 2.0

SVA Future 1

9393-2 &
RVA-002

9393-6 &
RVA-X83

SVA 9500 2.2

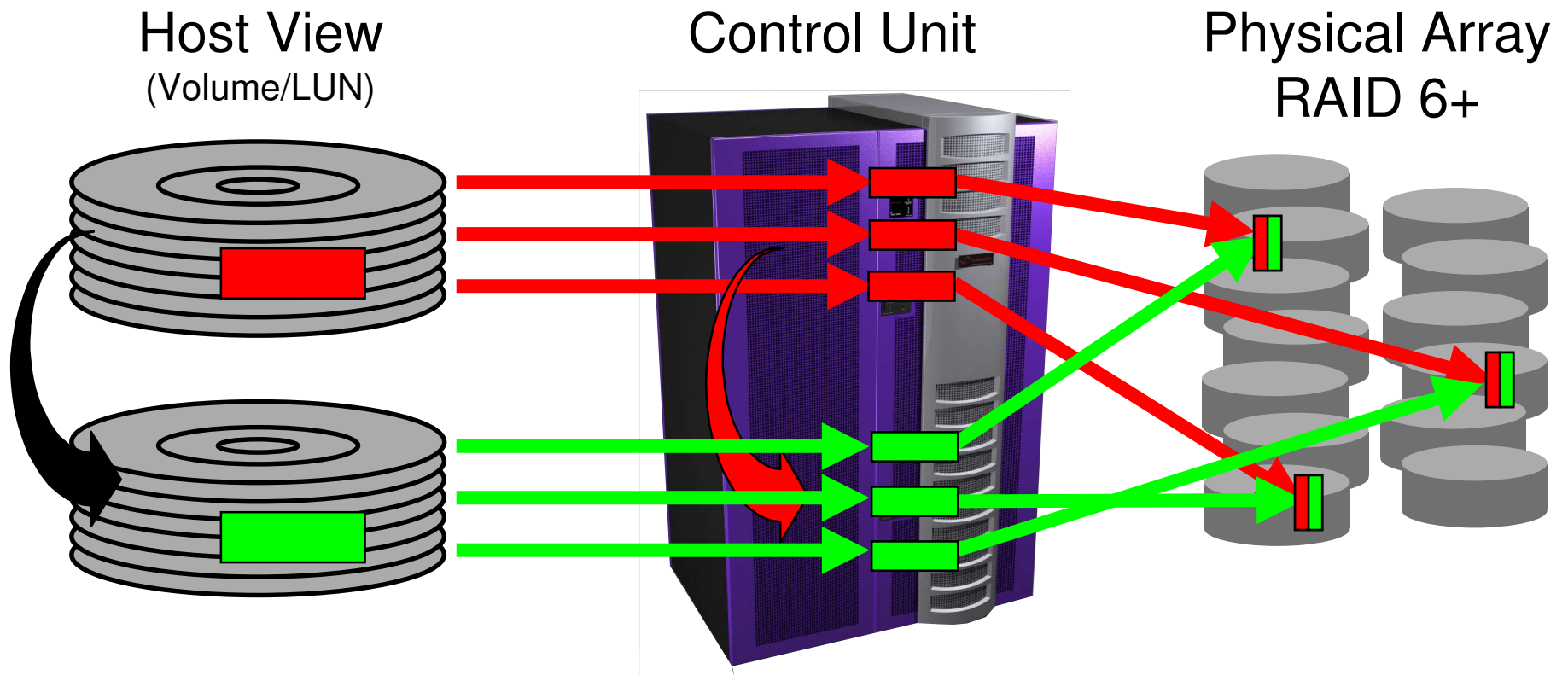*There are over 12,000 virtual disk subsystems installed world wide*

- STK Virtual Disk is based on a very refined Log-Structured File(LSF) Architecture on top of RAID 6 arrays.

- Controller Function in Microcode:

  – 3990-3+ Emulation (ESCON)

  – SCSI (Fibre Channel)

- Device Emulation in Microcode(3380/3390/SCSI LUN)

- Hardware data compression engines on I/O interface.

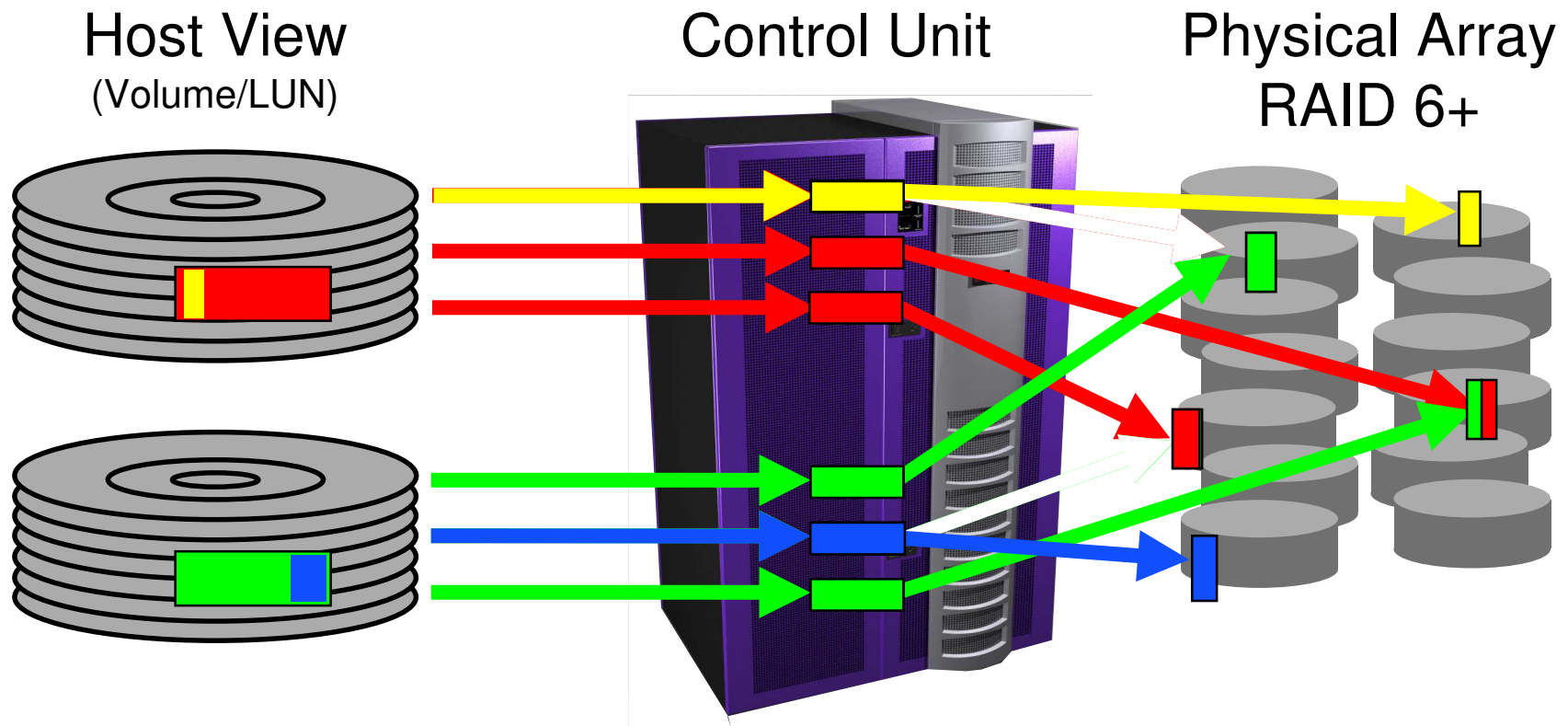- Dynamic device definition/reconfiguration.
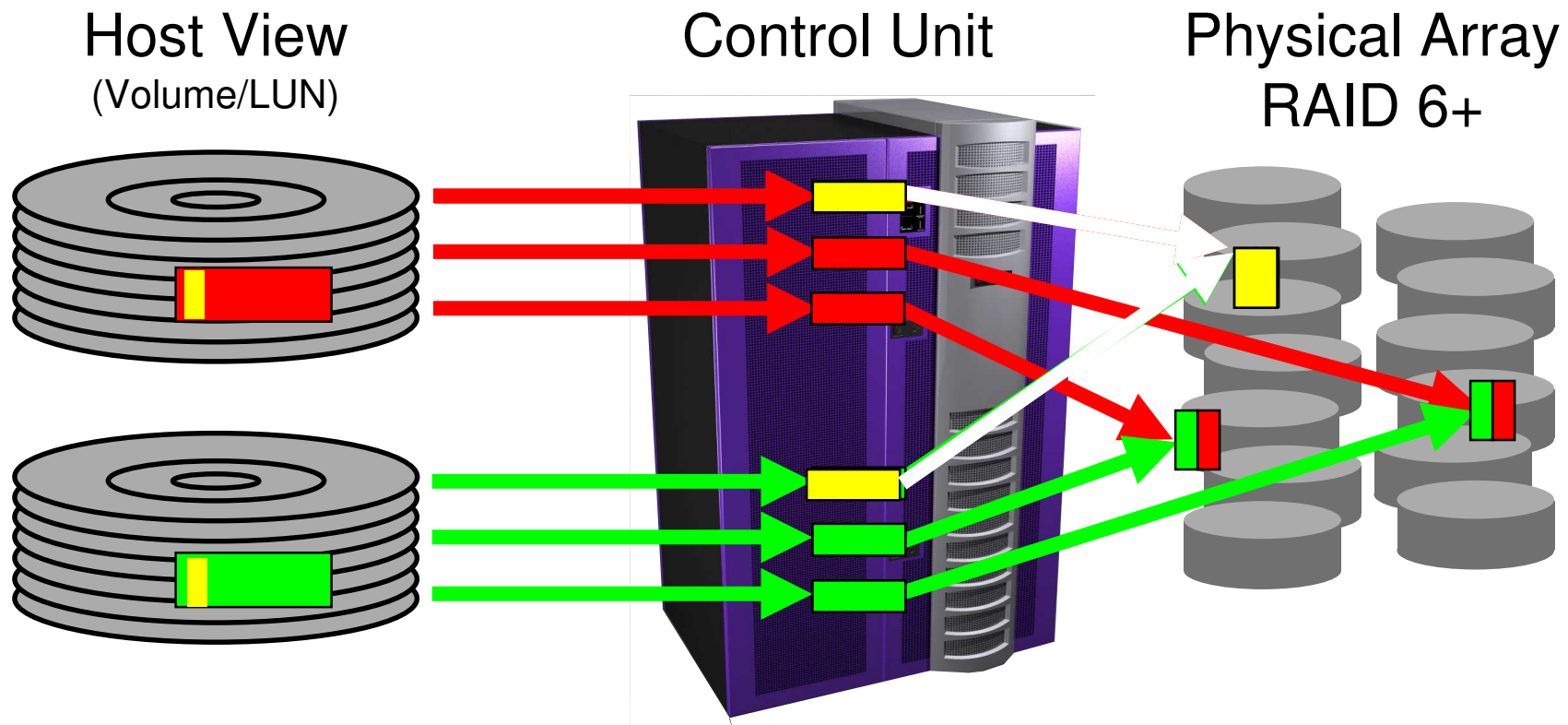
StorageTek Virtual Disk Overview

Host View

(Volume / LUN)

Appears Exactly as
"Conventional" DASD

Virtual
Architecture
Engine

Managed as <u>One</u>
Large Disk

FBA
FBA
FBA
FBA
FBA
FBA
FBA

FBA
FBA
FBA
FBA
FBA
FBA

Variable and Dynamically
Configurable "Virtual Space"

64 Disks Viewed as One
Large Disk / LUN

Compressed & Compacted
User Data

SnapShot

Host View
(Volume/LUN)

Control Unit

Physical Array
RAID 6+

SnapShot – Data Update

Host View
(Volume/LUN)

Control Unit

Physical Array
RAID 6+

Only updated tracks use additional capacity…

SnapShot – Data Delete

Host View
(Volume/LUN)

Control Unit

Physical Array
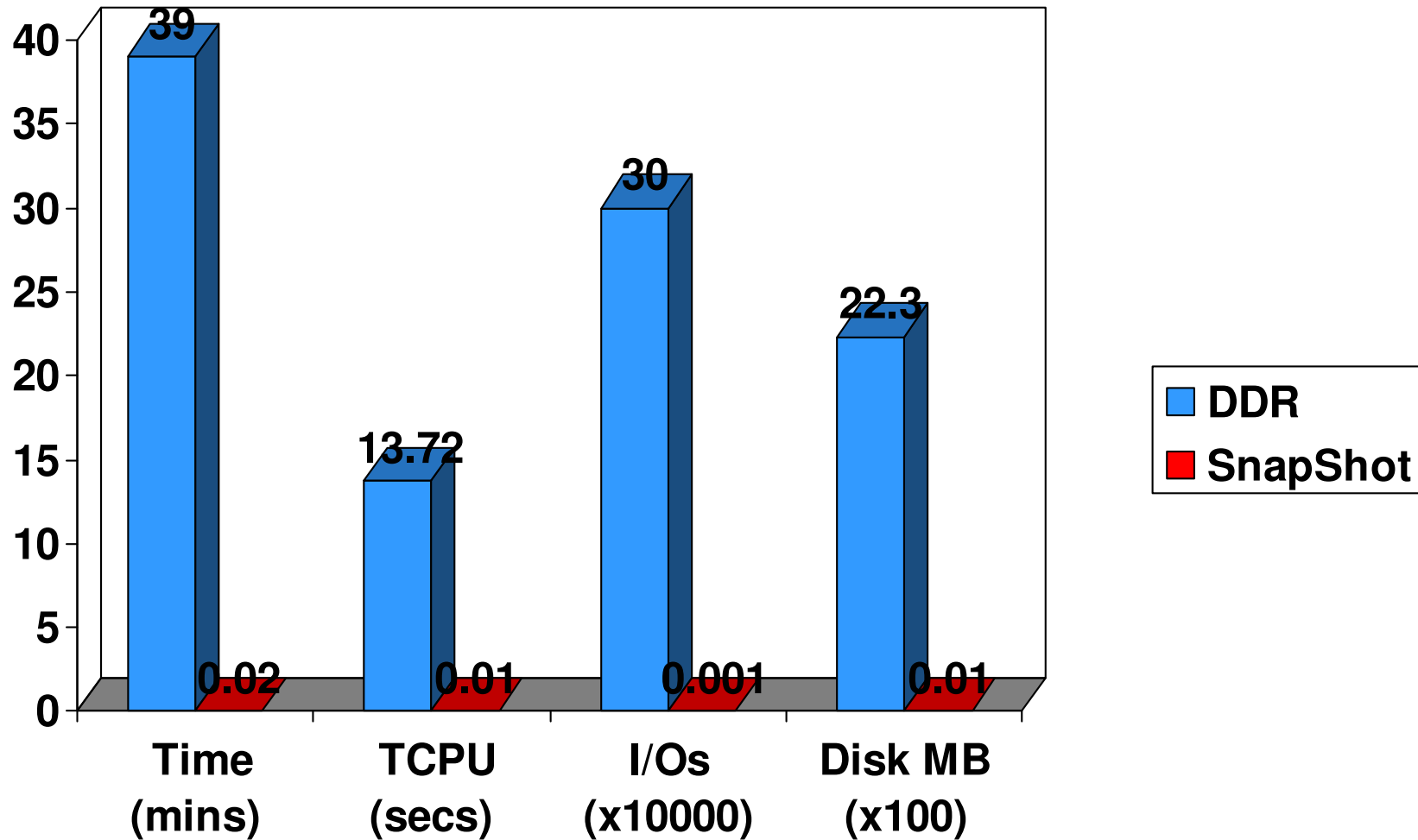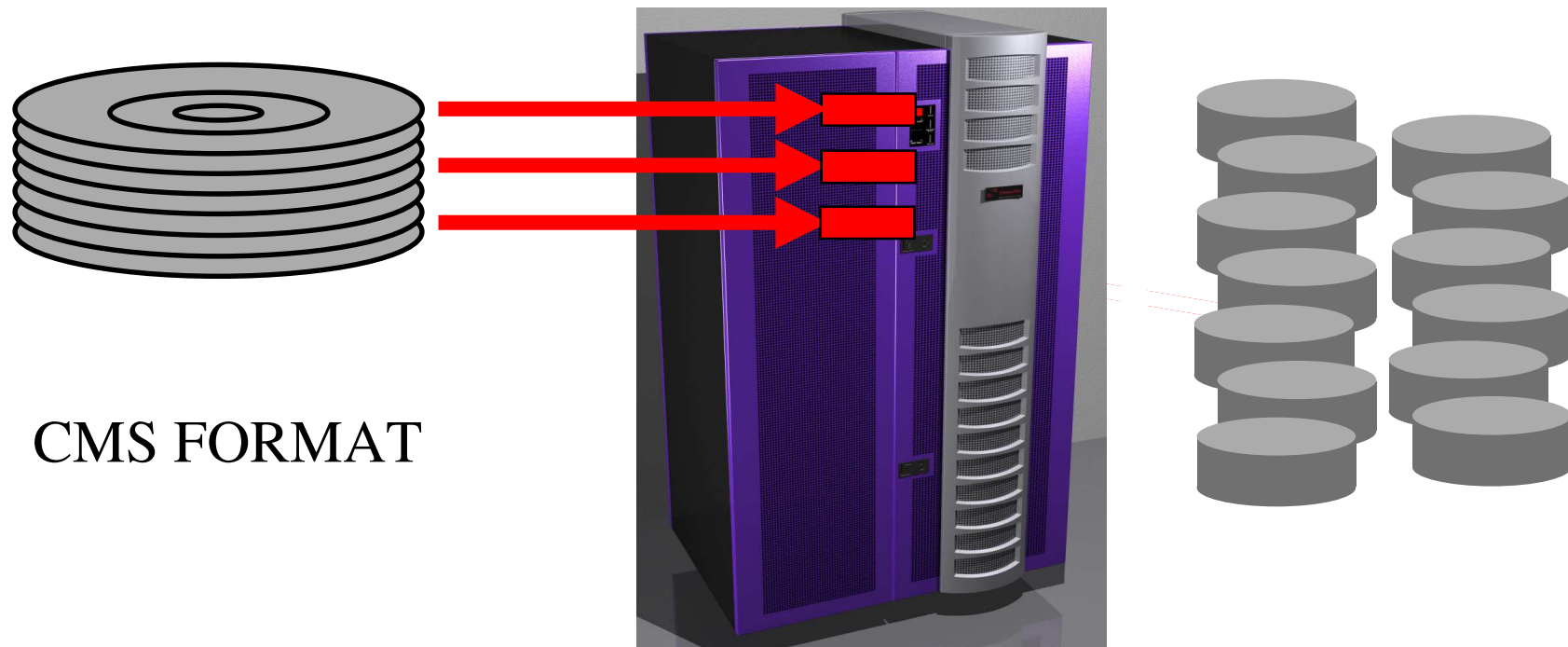RAID 6+

Data space converts to freespace when all pointers are removed.
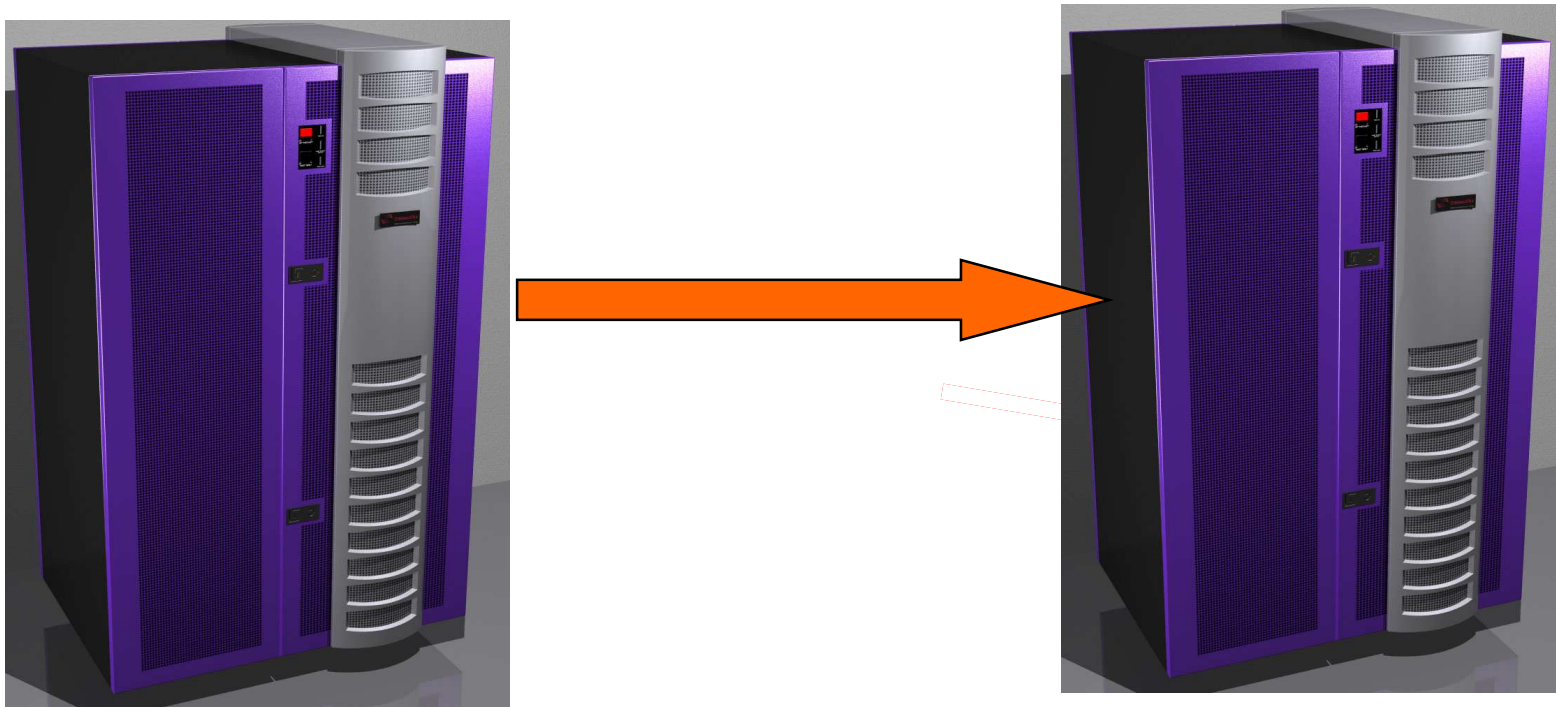
Minidisk Copy - 10,000 Cylinders

CMS FORMAT

"Pre-formatted" track definitions are inserted into the mapping tables. No back-end space is used.

PPRC (Peer-to-Peer Remote Copy)

- Compressed track images are transmitted, saving bandwidth

- Invoked at a DASD device level using ICKDSF under VM.

- Control unit feature, minimal host software involvement

- Synchronous. I/O delay depends on distance.

Using STK Virtual Disk With S/390 Linux

- S/390 Linux (currently) talks to 'standard' ECKD devices. (FC next year).

- The IBM-written DASD driver provides the interface between the Linux VFS and FS, and the DASD layout.

- Devices can be full volumes or VM minidisks. Volumes can be 'Linux' format (no MVS VTOC), or 'CDL' format (an MVS VTOC is used).

- Full volumes must be used if running Linux in an LPAR. They are optional if running under VM.
  - CDL format writes a VTOC so that the volume can be varied online to MVS for dump processing. A CDL volume can have up to three partitions, which are reflected in the VTOC as MVS datasets.

```
sssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssss
DSLIST – Data Sets on volume 0X0152                              Row 1 of 3
Command ===>                                                 Scroll ===> CSR


Command – Enter "/" to select action                 Message          Volume
---------------------------------------------------------------------------
        LINUX.V0X0152.PART0000.NATIVE                                  0X0152
        LINUX.V0X0152.PART0001.NATIVE                                  0X0152
        LINUX.V0X0152.PART0002.NATIVE                                  0X0152
**************************** End of Data Set list ***************************
```

Using STK Virtual Disk With S/390 Linux – Full Volumes (cont.)

SHARE
Technology · Connections · Results

- LNX format does not write an MVS VTOC on the volume.

- The volume cannot be varied on to MVS.

- The volume can be dumped from MVS using utilities that have offline volume support. Innovation's FDR is one such utility.

Using STK Virtual Disk With S/390 Linux – VM Minidisks

- VM minidisks provide maximum flexibility if
  1. Running under VM.
  2. There is no requirement for varying a volume online to a 'native' MVS.
- Minidisk pooling (VMSECUR, DIRMAINT) is especially useful for very dynamic system environments.
- SnapShot can be used to instantaneously copy minidisks. This is a major benefit with RVA/SVA under VM.
- There are no restrictions as to pool placement within VCU(LSS) for SnapShot as with some other architectures. Any volume or minidisk on an RVA/SVA can be Snapped to any other volume or minidisk of like geometry anywhere on the control unit.

- The SVAA SNAP MDISK command can be used to instantly copy minidisks under VM.

  - Cloning (example: SnapVantage)
  - Backup
    - As with any Unix type OS/filesystem, care must be taken when using a 'physical' disk backup or copy mechanism. This is due to the *nix use of system memory to cache data and metadata.
    - The Linux 'sync' command can be used to flush buffers. Unmount or mounting the filesystem R/O is much safer.
    - With minidisks, there is a one-to-one relationship between the minidisk and the Linux partition, making it easier to control access and 'keep things straight'.

- (Fuzzy) Minidisk backup example using SnapShot:

  –**From Linux command line (root):**

  ```
  #*#: echo $HOSTNAME
  ctlnx04
  #*#: sync
  #*#:
  ```

  DANGER WILL ROBINSON! THERE IS NO GUARANTEE THAT A WRITE WILL NOT TAKE PLACE IMMEDIATELY AFTER THE SYNC!!!!!!!!!! THIS IS A "FUZZY" COPY! MANY USERS ISSUE 'sync' THREE TIMES AS A MATTER OF COURSE.

  - **From VM command line :**
  CP LINK CTLNX04 151 4151 RR
  DASD 4151 LINKED R/O; R/W BY CTLNX04

  CP Q V 151
  DASD 0151 3390 SC43C3 R/W 3000 CYL 0001 RELN ON DASD 43C3 SUBCHANNEL = 0003

  CP Q V 4151
  DASD 4151 3390 SC43C6 R/O 3000 CYL 5601 RELN ON DASD 43C6 SUBCHANNEL = 000D

  sibvmrva snap 4151 151   <This command runs in the blink of an eye!
  Ready;

SnapVantage

- **Minimum set up**

    –Reduces Linux operating system deployment time to seconds

- **Create VM Machine Dynamically**

    –VM, VMSECURE, DIRMAINT and SuSE v7.0+Linux kernel 2.2.16, SuSE v7.2+Linux kernel 2.4 or Redhat v7.2+Linux kernel 2.4

- **Creates VM Minidisk Dynamically with SnapShot**

    –Does not require any pre-allocated space or setup time

    –SnapShot created with instant format

- **Creates the LINUX system and swap disk with SnapShot**

- **Zero to Linux server in seconds** (automatic boot)

Using SnapShot for Cloning Linux: SnapVantage

Using SnapShot for Cloning Linux: SnapVantage Demo

Special Considerations for RVA/SVA

- Data written into an RVA/SVA is not deleted until the control unit is told the data is to be deleted. The granularity of deletion is one track.
- On MVS, there is a Deleted Data Space Release function that can dynamically notify the control unit to delete tracks. There is a similar function for CMS minidisks.
- Open systems present a problem. There is no easy way to map freespace in an open filesystem to the track architecture. Also, some of the newer journaling file systems naturally 'wrap' the entire disk partition.
- For a non-journalling FS, compressible data can be written to free areas of the filesystem. This can be accomplished via a 'dd' command.

Heads Up

- Session 9270, 8:00 Friday: **Cloning Linux Images on VM; Practical Tools and Techniques**
Ever heard the saying 'There's more than one way to skin a cat?' Well, it turns out that there's more than one way to clone a penguin, too. Each of the speakers has used a unique method or tool to clone large numbers of Linux images running under VM. Come hear how images are cloned on the Linux Community Development System, and from a developer of the new SnapVantage Tool, to gather ideas of how you might perform this important task! **Richard Lewis, IBM and Don Johnson, StorageTek**

- This session qualifies for the Linux on S/390 and the VM System programming certificate tracks.

Heads Up

- See and hear more about Linux at the StorageTek booth in the vendor area. Booth #716 in the Hilton.

Additional Information

- StorageTek Website: http://www.storagetek.com

- StorageTek SVA V960 Website:
http://www.storagetek.com/prodserv/products/disk/V960/

- StorageTek SnapVantage Website:
http://www.storagetek.com/prodserv/products/software/svan/