IBM System z

IBM z/VSE 3.1 and SCSI Performance Considerations

Ingo Franzki – ifranzki@de.ibm.com

March 12, 2009

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and / or other counties.

CICS*
DB2*
DB2 Connect
DB2 Universal Database
e-business logo*
Enterprise Storage Server
HiperSockets
IBM*
IBM logo*
IMS
Intelligent Miner
Multiprise*
MQSeries*
OS/390*
S/390*
SNAP/SHOT*
Virtual Image Facility
VM/ESA*
VSE/ESA
VisualAge*
VTAM*
WebSphere*
xSeries
z/Architecture
z/VM
z/VSE
zSeries

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

LINUX is a registered trademark of Linus Torvalds

Tivoli is a trademark of Tivoli Systems Inc.

Java and all Java-related trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries

UNIX is a registered trademark of The Open Group in the United States and other countries.

Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.

SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.

Intel is a registered trademark of Intel Corporation.

Disclaimer

The information contained in this document has not been submitted to any formal IBM test and is distributed on an "AS IS" basis without any warranty either express or implied. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk.

In this document, any references made to an IBM licensed program are not intended to state or imply that only IBM's licensed program may be used; any functionally equivalent program may be used instead.

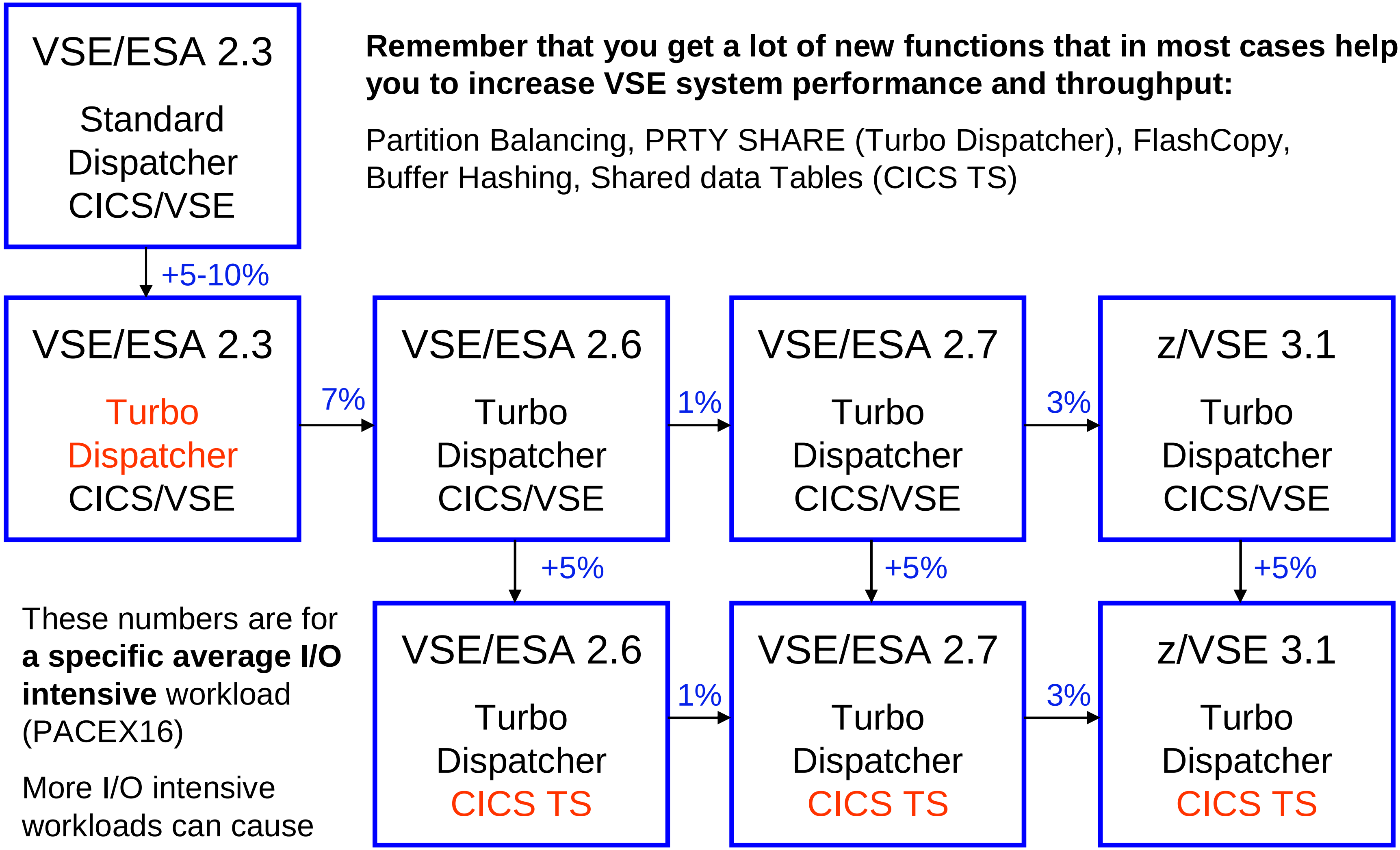
Any performance data contained in this document was determined in a controlled environment and, therefore, the results which may be obtained in other operating environments may vary significantly. Users of this document should verify the applicable data for their specific environments.

It is possible that this material may contain reference to, or information about, IBM products (machines and programs), programming, or services that are not announced in your country. Such references or information must not be construed to mean that IBM intends to announce such IBM products, programming or services in your country.

Overhead Deltas for VSE Releases

Remember that you get a lot of new functions that in most cases helps you to increase VSE system performance and throughput:

Partition Balancing, PRTY SHARE (Turbo Dispatcher), FlashCopy, Buffer Hashing, Shared data Tables (CICS TS)



VSE/ESA 2.3
Standard Dispatcher CICS/VSE

+5-10%

VSE/ESA 2.3
Turbo Dispatcher CICS/VSE

7%

VSE/ESA 2.6
Turbo Dispatcher CICS/VSE

1%

VSE/ESA 2.7
Turbo Dispatcher CICS/VSE

3%

z/VSE 3.1
Turbo Dispatcher CICS/VSE

+5%

+5%

+5%

VSE/ESA 2.6
Turbo Dispatcher CICS TS

1%

VSE/ESA 2.7
Turbo Dispatcher CICS TS

3%

z/VSE 3.1
Turbo Dispatcher CICS TS

These numbers are for **a specific average I/O intensive** workload (PACEX16)

More I/O intensive workloads can cause more overhead

New VSE CPU Monitor Tool

- **Intended to help customers to measure the CPU utilization of their VSE system over a period of time.**
- **When you plan for a processor upgrade it is very important to know the CPU utilization of your VSE system over a day or a week.**
  - Helps you to estimate the size of the new processor.
- **The VSE CPU Monitor Tool is not intended to replace any existing monitoring product provided by partners.**
- **It provides only very basic monitoring capabilities on an overall VSE system level.**
- **No details about CPU usage of certain applications are provided**
- **Download**
  - http://www.ibm.com/servers/eserver/zseries/zvse/downloads/tools.html
  - ‘As is’, no official support, e-mail to zvse@us.ibm.com

New VSE CPU Monitor Tool

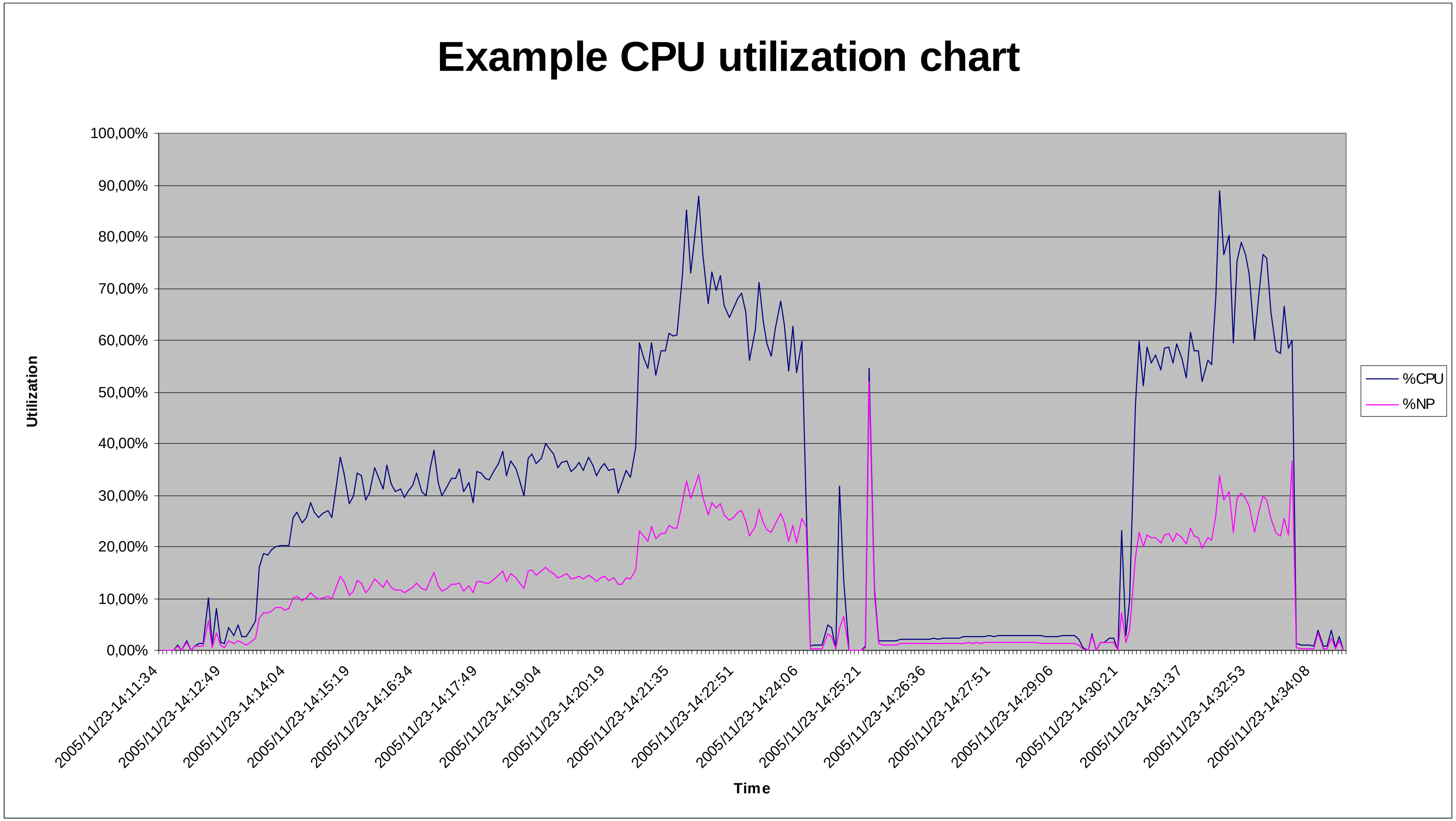
- CPUMON periodically issues a TDSERV FUNC=TDINFO macro to get performance relevant data.
- The data provided by the macro is the same as command QUERY TD shows.
- The data from each measurement interval is printed to SYSLST in a comma separated format.
- Later on this data can be imported into a spreadsheet (EXCEL)
- CPUMON runs in a VSE partition (dynamic or static).
- CPUMON is started using:

```
// EXEC DTRIATTN,PARM='SYSDEF TD,RESETCNT‘
/*
// EXEC CPUMON,PARM='nn‘  nn = interval in seconds
/*
```

- The tool can be stopped by entering the following command:

```
MSG xx,DATA=EXIT          xx = partition id
```

New VSE CPU Monitor Tool

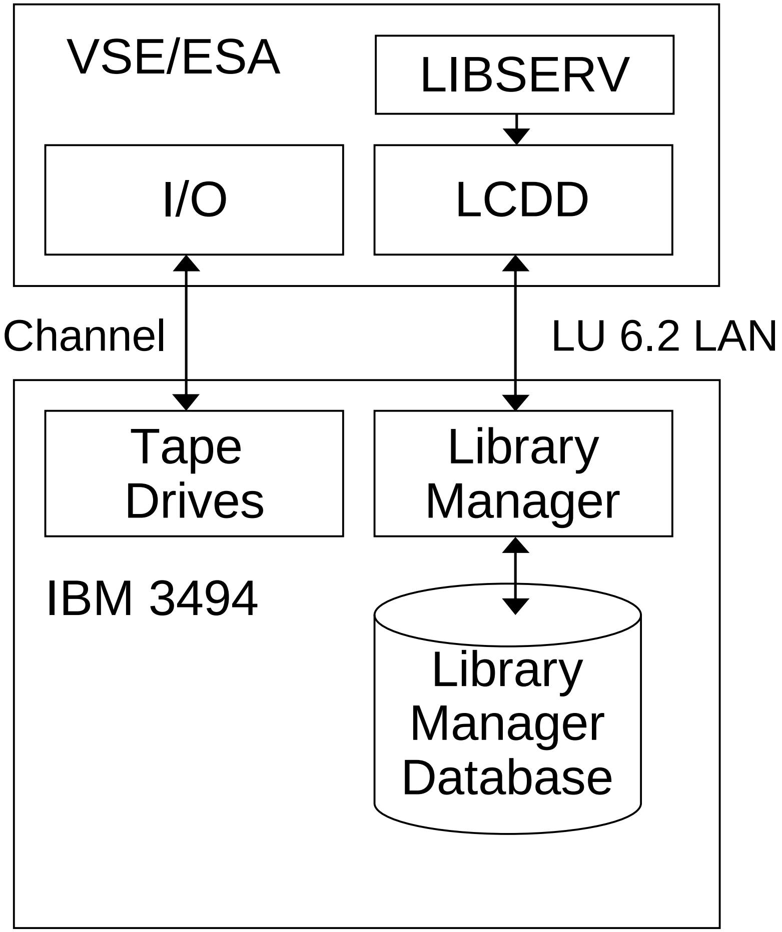


Native Tape Library Support (TLS)

- **z /VSE supports**
  - IBM TotalStorage 3494 Enterprise Tape Library
  - IBM TotalStorage 3584 UltraScalable Tape Library
  - IBM TotalStorage Virtual Tape Server

Native Tape Library Support (TLS)

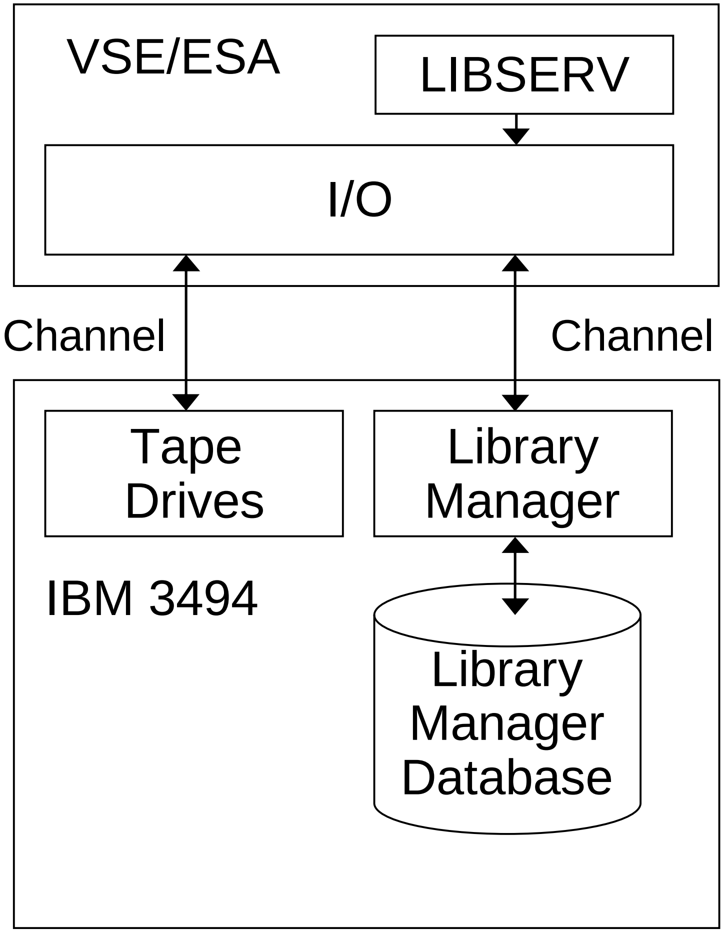
- **Before z/VSE 3.1**
  - LCDD (library control device driver)
  - Running in a VSE partition
  - via VTAM LU6.2
  - difficult VTAM – LAN setup
  - XPCC connection



Native Tape Library Support (TLS)

- **New with z/VSE 3.1**
  - Via S/390 channel command interface
  - No XPCC
  - No VTAM LU6.2
- **LBSERV macro**
- **LIBSERV JCL / AR command interface**
- **Example :**

```
// LIBSERV MOUNT,VOL=123456/w,UNIT=480
// LIBSERV RELEASE,UNIT=480
```



Native Tape Library Support (TLS)

§ New LIBSERV JCL commands

- LIBSERV AQUERY,VOL=123456
  - query all libraries for a specified volume
- LIBSERV CMOUNT,UNIT=480,SRCCAT=SCRATCH01
  - mount from category
- LIBSERV CQUERY,LIB=TAPELIB1,SRCCAT=SCRATCH01
  - query count of volumes
- LIBSERV DQUERY,UNIT=480
  - query device
- LIBSERV IQUERY,LIB=TAPELIB1,SRCCAT=SCRATCH01
  - query inventory of library
- LIBSERV LQUERY,LIB=TAPELIB1
  - query library
- LIBSERV MINVENT,MEMNAME=ALL,TGTCAT=SCRATCH01
  - manage inventory
- LIBSERV SETVCAT,VOL=123456,TGTCAT=SCRATCH01
  - change category
- LIBSERV SQUERY,VOL=123456,LIB=TAPELIB1

Native Tape Library Support (TLS)

- **z/VSE TLS support customization**
  - SYS ATL=TLS | VSE | VM
- **TLSDEF skeleton in ICCF LIB 59**
  - define library name and corresponding tape drives

```
// LIBDEF *,CATALOG=IJSYSRS.SYSLIB
// EXEC LIBR,PARM='MSHP‘
   ACCESS S=IJSYSRS.SYSLIB
   CATALOG TLSDEF.PROC REPLACE=YES
   LIBRARY_ID TAPELIB1 SCRDEF=SCRATCH00 INSERT=SCRATCH00
   LIBRARY_ID TAPELIB2              * SECOND LIB DEF
   DEVICE_LIST TAPELIB1 460:463     *  DRIVES  460 TO 463
   DEVICE_LIST TAPELIB2 580:582     *  DRIVES  580 TO 582
   QUERY_INV_LISTS LIB=TLSINV       * MASTER INVENTORY FILES
   MANAGE_INV_LISTS LIB=TLSMAN      * MANAGE FROM MASTER
/+
```

VTAM 31 Bit I/O Buffer Support

- **VSE/VTAM is providing new 31 bit IO buffer support via PTF**
  - Removes the 24 bit I/O buffer restriction.
  - Moves the I/O buffer pool and all I/O CTC packing buffers above the 24 bit line
  - Support I/O operations in 31 bit mode (using format 1 CCWs).
  - Allow z/VSE customers to grow their communications workloads associated with their business critical applications
- **PTFs**
  - VTAM support - APAR DY46471 - PTF UD52964
  - z/VSE support - APAR DY46396 - PTF UD52873 (AF Base) or UD52874 (Generation Feature)

VTAM 31 Bit I/O Buffer Support

- **To enable the 31 Bit support:**
  - Make sure both PTFs (VTAM and AF) are applied
  - Set VTAM start option IOBUF31=YES (default is NO)
- **Display settings**

```
d net,vtamopts,opt=iobuf31
AR 0015 1C39I COMMAND PASSED TO ACF/VTAM
F3 0003 IST097I DISPLAY ACCEPTED
F3 0003 IST1188I ACF/VTAM V4R2 STARTED AT 17:11:29 ON 11/11/05
F3 0003 IST1349I COMPONENT ID IS 5686-06501-FE6
F3 0003 IST1348I VTAM STARTED AS INTERCHANGE NODE
F3 0003 IST1497I VTAM FUNCTIONAL SUPPORT LEVEL IS INTERENTERPRISE
F3 0003 IST1189I IOBUF31 = YES
F3 0003 IST314I END
```

Hardware and software requirements for SCSI

- **IBM eServer zSeries 800, 900, 890, 990, z9 or z10**
- **IBM zSeries FCP Adapter**
  - Microcode Level:
    - z800 und z900: J11233.015 or higher
    - z890 und z990: J13471.004 or higher
- **FCP Switch (e.g. IBM 2109)**
- **IBM TotalStorage Enterprise Storage Server (ESS)**
  - Microcode Level: 2.3.1 or higher
- **IBM TotalStorage DS6000 or DS8000**
- **z/VSE Version 3 Release 1**
- **z/VM 4.4. or higher (only if VSE runs under VM)**

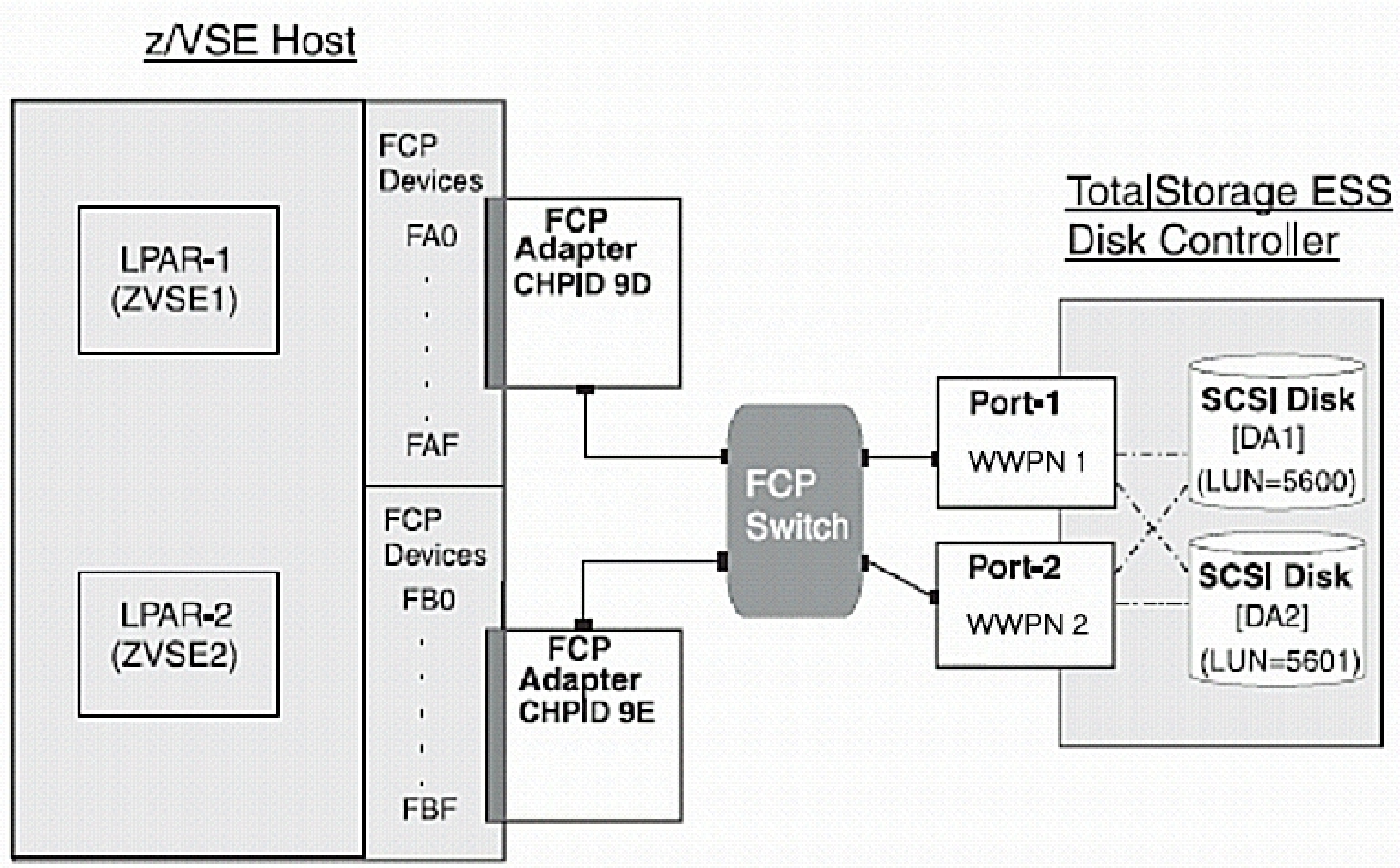
Hardware and software requirements for SCSI (2)

- **IPL from SCSI**
  - CPU Feature Code 9904
  - z800 and z900:
    - Microcode Level EC-Number J12811 or higher
  - z890 and z990:
    - Microcode Level EC-Number J12221 or higher
- **IPL from SCSI under z/VM 4.4**
  - z/VM Service Level:
    - UM31181 (English)
    - UM31180 (German)
    - UM31179 (Kanji)
- **Emulated FBA Disks:**
  - z/VM 5.1

SCSI disk characteristics

- **Size of a SCSI Disk**
  - Minimum 8 MB
  - Maximum about. 24 GB
  - 4 MB are used internally from z/VSE
  - Usable size = size – 4 MB
  - VSAM can only use the first 16 GB
- **Model**
  - SCSI Disks are defined as FBA Devices
    - 9336 Model 20
- **Block sizes**
  - z/VSE supports only SCSI Disks with a block size of 512 Bytes
- **Standards**
  - SCSI Disks must support ANSI SCSI Version 3

SCSI components



SCSI setup for z/VSE

- **FCP Devices:**
  - ADD 4A7:4A9,FCP
- **FBA Devices:**
  - ADD 608:61B,FBA
- **Define SCSI:**
  - DEF SCSI,FBA=608,FCP=4A7,WWPN=5005076300CA9A76,LUN=5710
  - DEF SCSI,FBA=609,FCP=4A7,WWPN=5005076300CA9A76,LUN=5711
  - …
  - DEF SCSI,FBA=60D,FCP=4A8,WWPN=5005076300CA9A76,LUN=5715
  - DEF SCSI,FBA=60E,FCP=4A8,WWPN=5005076300CA9A76,LUN=5716
  - …
- **IPL from SCSI (VM)**
  - Minimum 32M Memory
  - SET LOADDEV PORT 50050763 00CE9A76 LUN 57350000 00000000
  - IPL 4B8   (IPL from FCP device)

SCSI commands

§SCSI definitions during IPL:

```
DEF SCSI,FBA=cuu,FCP=cuu,WWPN=nnnnnnnnnnnnnnnn,LUN=nnnn
```

§SCSI definitions online:

```
SYSDEF SCSI,FBA=cuu,FCP=cuu,WWPN=nnnnnnnnnnnnnnnn,LUN=nnnn
```

– FBA Device and FCP Device must have been already defined during IPL

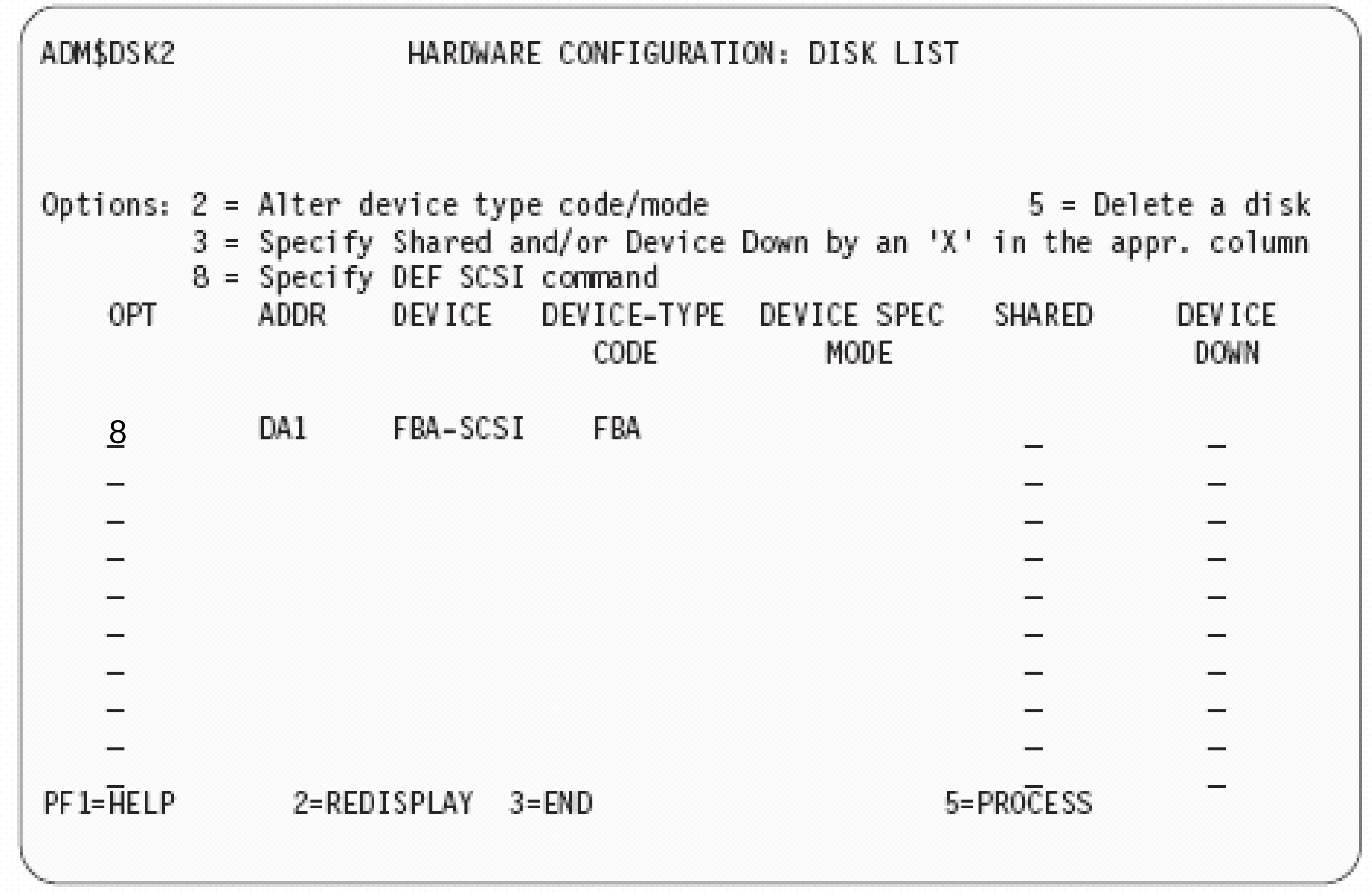
§Delete SCSI definitions:

```
SYSDEF SCSI,DELETE,FBA=cuu,FCP=cuu,WWPN=nnnnnnnnnnnnnnnn,LUN=nnnn
```

§Display SCSI definitions:

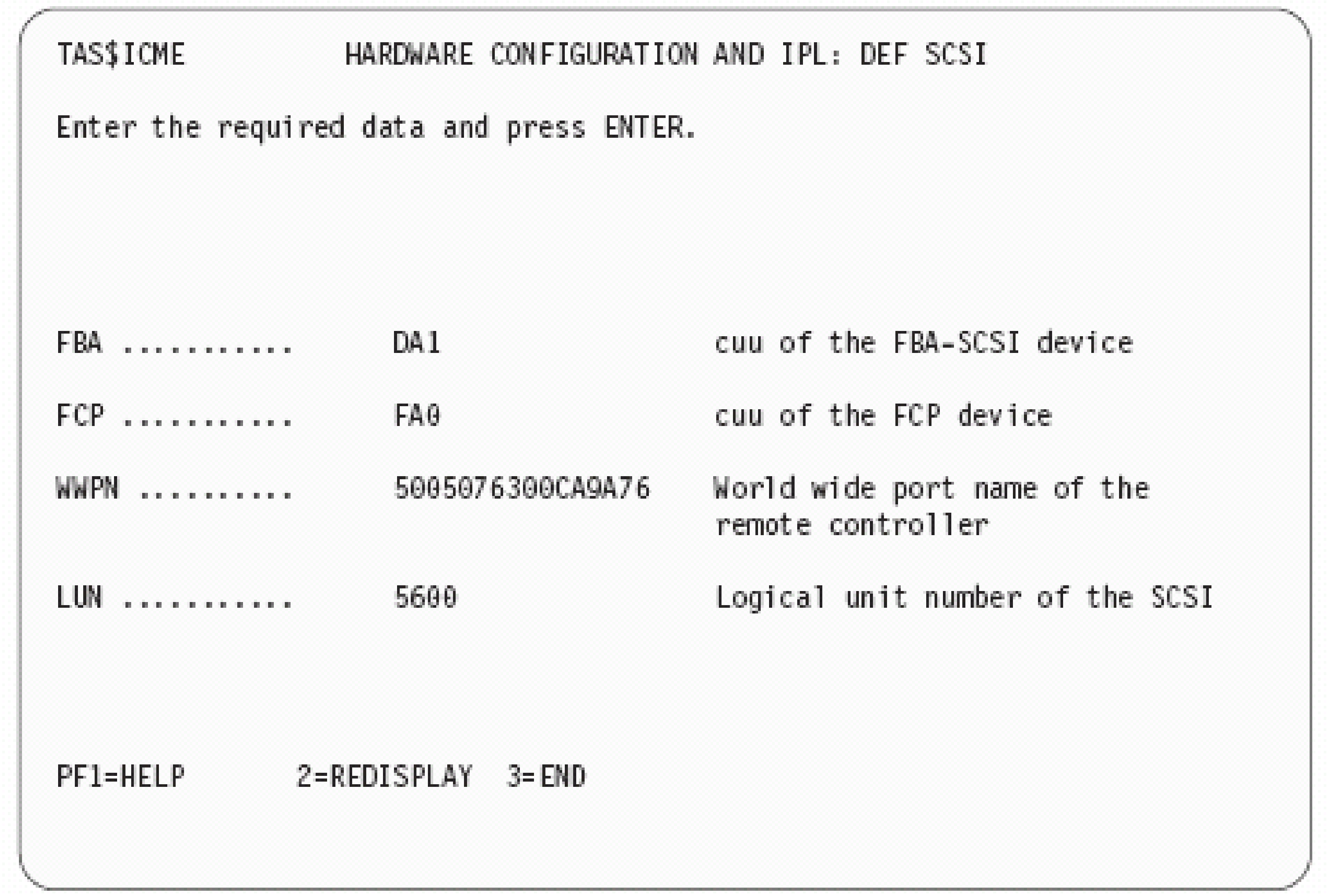
```
QUERY SCSI

AR 0015 FBA-CUU   FCP-CUU   WORLDWIDE PORTNAME   LOGICAL UNIT NUMBER
AR 0015     608       4A7     5005076300CA9A76     5710000000000000
AR 0015     609       4A7     5005076300CA9A76     5711000000000000
AR 0015     60A       4A7     5005076300CA9A76     5712000000000000
AR 0015     60B       4A7     5005076300CA9A76     5713000000000000
AR 0015     60C       4A7     5005076300CA9A76     5714000000000000
AR 0015     60D       4A8     5005076300CA9A76     5715000000000000

QUERY SCSI,608    (FBA Device)

AR 0015 FBA-CUU   FCP-CUU   WORLDWIDE PORTNAME   LOGICAL UNIT NUMBER
AR 0015     608       4A7     5005076300CA9A76     5710000000000000
```

Interactive Interface Dialog



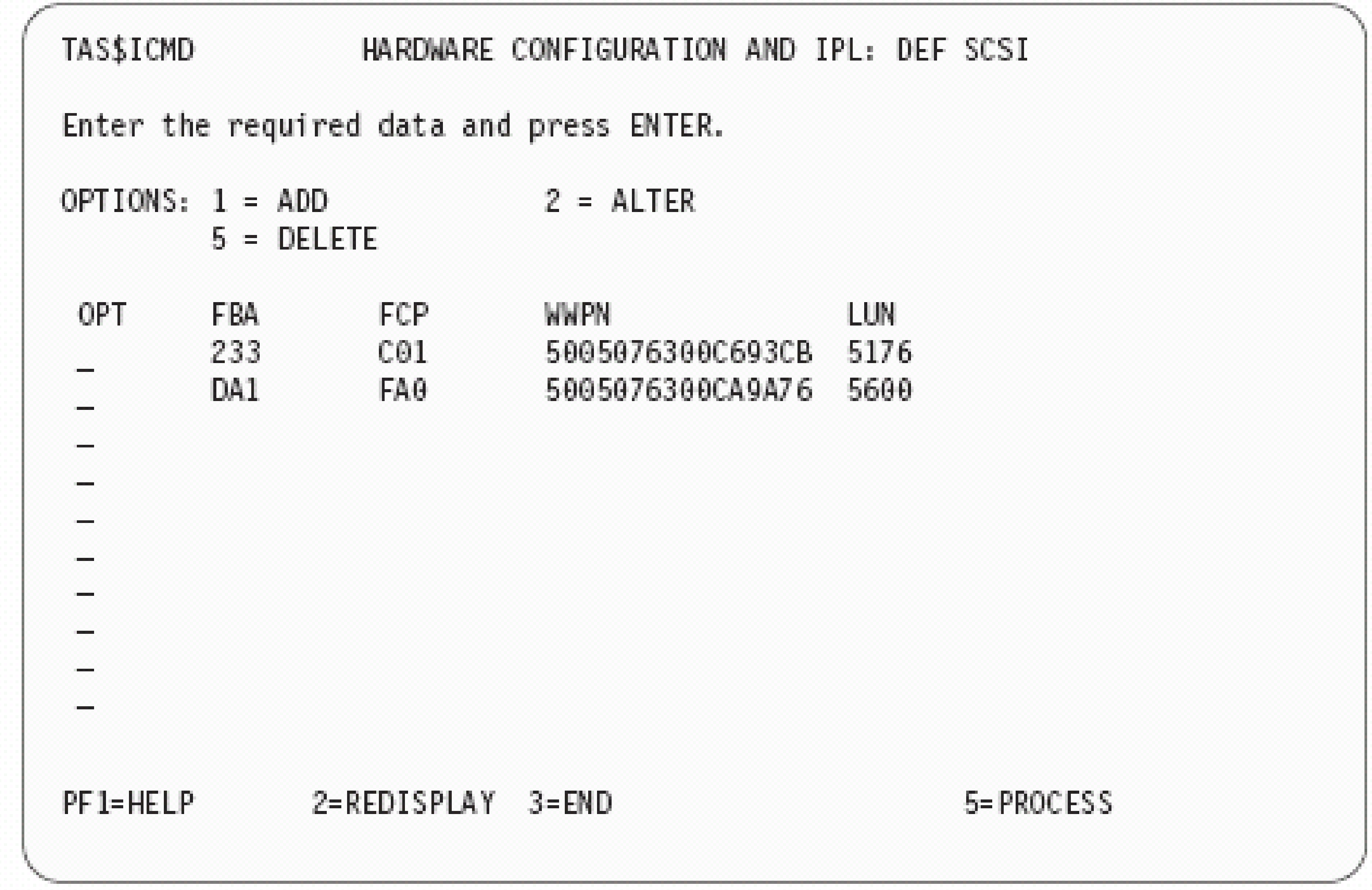
```
ADM$DSK2                 HARDWARE CONFIGURATION: DISK LIST



Options: 2 = Alter device type code/mode                      5 = Delete a disk
         3 = Specify Shared and/or Device Down by an 'X' in the appr. column
         8 = Specify DEF SCSI command
     OPT     ADDR    DEVICE    DEVICE-TYPE  DEVICE SPEC    SHARED      DEVICE
                                  CODE         MODE                     DOWN

     8       DA1     FBA-SCSI     FBA                        _           _
     _                                                       _           _
     _                                                       _           _
     _                                                       _           _
     _                                                       _           _
     _                                                       _           _
     _                                                       _           _
     _                                                       _           _
     _                                                       _           _
     _                                                       _           _
PF1=HELP      2=REDISPLAY   3=END                      5=PROCESS
```

Interactive Interface Dialog



```
TAS$ICME        HARDWARE CONFIGURATION AND IPL: DEF SCSI

Enter the required data and press ENTER.




FBA ..........      DA1                 cuu of the FBA-SCSI device

FCP ..........      FA0                 cuu of the FCP device

WWPN .........      5005076300CA9A76    World wide port name of the
                                        remote controller

LUN ..........      5600                Logical unit number of the SCSI



PF1=HELP      2=REDISPLAY   3=END
```

Interactive Interface Dialog



```
TAS$ICMD           HARDWARE CONFIGURATION AND IPL: DEF SCSI

Enter the required data and press ENTER.

OPTIONS: 1 = ADD                  2 = ALTER
         5 = DELETE

 OPT     FBA        FCP        WWPN                LUN
 _       233        C01        5005076300C693CB    5176
 _       DA1        FA0        5005076300CA9A76    5600
 _
 _
 _
 _
 _
 _
 _
 _

PF1=HELP       2=REDISPLAY   3=END                      5=PROCESS
```

SCSI messages

- **AR 0033 0S45I SCSI DEVICE 618 CONSISTS OF 03906304 BLOCKS, 03897432 BLOCKS ARE AVAILABLE, 680 BLOCKS ARE UNUSED**
  - Multiple of 777 Blocks
    - Internal Model: 1 „Cylinder“ = 777 Blocks
- **AR 0033 0S40I SCSI PROCESSING EVENT: REASON=0060 FUNCTION=INIT-SCSI FBA=609 FCP=4A7 WWPN=5005076300CA9A76 LUN=5711000000000000**
  - Message description shows the reason based on the Reason Codes
  - SCSI I/O errors are mapped in S/390 I/O errors for user programs (e.g. Unit check)
  - In addition, message 0S40I message is issued, to inform about the exact reason

SCSI multipathing

- **One or more alternative paths to the same SCSI Disk**
  - Increases the availability
  - NOT: Workload-Balancing
- **Each path must be defined over a different FCP adapter**
  - One FCP card can contain multiple FCP adapters (CHPID)
  - To increase availability, you should use different FCP adapters on different physical FCP cards
- **As best, even over different switches and/or ports**
- **Example:**

```
DEF SCSI,FBA=DA1,FCP=FA0,WWPN=5005076300CA9A76,LUN=5600
DEF SCSI,FBA=DA1,FCP=FB0,WWPN=5005076300C29A76,LUN=5600
```

- **QUERY SCSI**

```
AR 0015 FBA-CUU FCP-CUU WORLDWIDE PORTNAME LOGICAL UNIT NUMBER
AR 0015 DA1     FA0     5005076300CA9A76   5600000000000000
AR 0015 DA1     FB0     5005076300C29A76   5600000000000000
```

  - The first path is currently used to access the SCSI disk

Sharing SCSI disks

- **ADD cuu,FBA,SHR (FBA Device)**
- **Lockfile is used (see DLF)**
  - Using Reserve/Release SCSI Command (internal)
    - Reserve is based on FCP Adapter
    - Release must be done from same FCP Adapter
  - Possible problem:
    - Hardwait during disk is reserved
      - Disk stays reserved
    - The system tries to release the disk during hardwait processing, but this may fail
- **Restrictions**
  - Lockfile can not reside on DOSRES/SYSWK1 (only for SCSI)
  - No Multipathing possible for Lockfile-Disks
  - Each VSE System must access the Lockfile using its own FCK CHPID
    - If you have enabled NPIV mode (N_Port ID Virtualization) you can use the same FCP CHPID.
- **Suggestion**
  - Use separate Disk for Lockfile
  - Enable NPIC (N_Port ID Virtualization) if applicable
    - Available on IBM System z9/z10 processors and Ficon Express2 adapters

Base installation on SCSI disks

- **Only base installation possible**
  - FSU from ECKD- to SCSI-Disks not possible
- **Automatic installation**
  - IPL from Tape

```
BG 0000 SI70D IF YOU WANT TO USE SCSI DEVICES SPECIFY YES, ELSE NO
0 YES
BG 0000 SI75I ENTER SCSI COMMAND FOR DOSRES
BG 0000 SA80D SCSI,FBA=CUU,FCP=CUU,WWPN=PORTNAME,LUN=LUN
0 SCSI,FBA=608,FCP=C00,WWPN=5005076300C69A76,LUN=5745
AR 0033 0S45I SCSI DEVICE 608 CONSISTS OF 09765632 AVAILABLE, 651 BLOCKS ARE UNUSED
BG 0000 SA76I ENTER SCSI COMMAND FOR SYSWK1
BG 0000 SA80D SCSI,FBA=CUU,FCP=CUU,WWPN=PORTNAME,LUN=LUN
0 SCSI,FBA=609,FCP=D00,WWPN=5005076300C29A76,LUN=5746
AR 0033 0S45I SCSI DEVICE 609 CONSISTS OF 09765632 AVAILABLE, 651 BLOCKS ARE UNUSED
BG 0000 SI08I DOSRES IS 608, DEVICE TYPE FBA
BG 0000 SI09I SYSWK1 IS 609, DEVICE TYPE FBA
```

- **Hardware Configuration Dialog**
  - Press PF5 to catalog the IPLPROC
    - Otherwise the next IPL will fail because it does not find SYSWK1

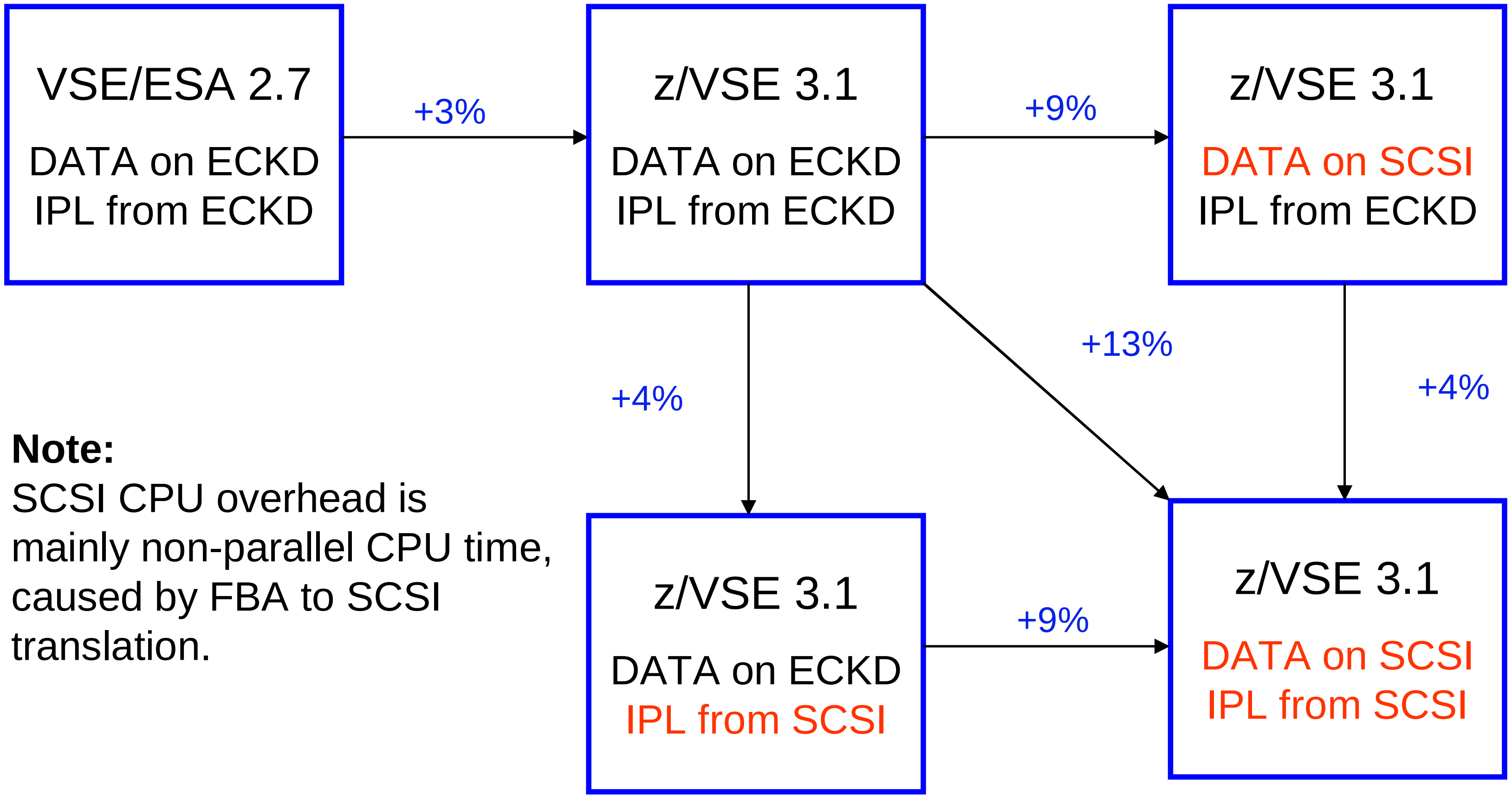
IPL from SCSI

- **Uses the „Machine Loader“**
  - Platform independent Hardware-Tool
- **Native or LPAR**
  - Perform a Load using Hardware Management Console (HMC)
    - Load Address = FCP Device
    - WWPN
    - LUN number
- **Under z/VM**
  - SET LOADDEV PORTNAME 50050763 00C29A76 LUN 56010000 00000000
  - IPL cuu (FCP Device)

Migration from ECKD to SCSI

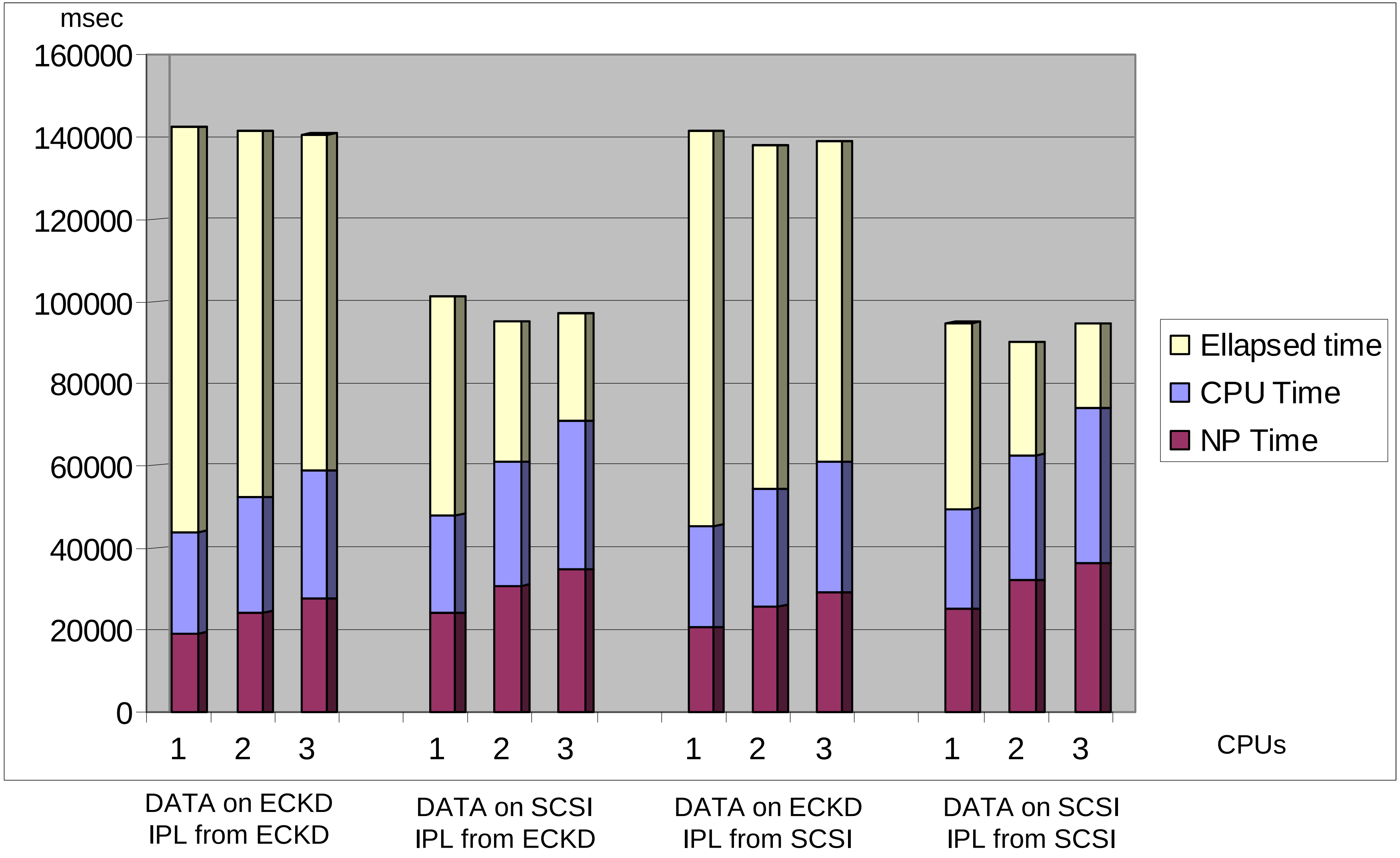
- **FSU from ECKD to SCSI not possible**
  - Only base installation
- **File allocations must be adapted**
  - Tracks/Cylinder into Blocks (for 3390):
    - 1 Track = about 112 Blocks
    - 1 Cylinder = about 1680 Blocks
  - VSAM Space
    - Hint: Specify cluster sizes in RECORDS, not Tracks
  - Sequential files
  - VSE Libraries
    - Hint: 1 LIBR Block = 1024 Bytes = 2 SCSI Blocks
- **Programs must be able to work with FBA disks**
  - As best, implement it device independent

Overhead Deltas for SCSI



Note:
SCSI CPU overhead is mainly non-parallel CPU time, caused by FBA to SCSI translation.

SCSI Overhead



SCSI I/O count considerations

- **For PACEX16 workload**
  - ECKD:
    - 18000 ECKD I/Os per disk
  - SCSI:
    - 20000 FBA I/Os per disk
  - ECKD → FBA:
    - 11% increased I/O counts
- **In general 1 FBA I/O is translated into 1 SCSI I/O**
- **Except for**
  - Long CCW chains
  - Overlapping addresses (e.g. PHASE loading)

Comparison ESCON – FICON/FCP

	ESCON	**FICON/FCP**
Max # of channels	256	4 x 256
Max # device addresses per link	4096	65536
Max # logical CU-paths per port	64	256
Device addresses per channel	1024	16384
Link rate	20 MB/sec	200 MB/sec
Max achievable transfer rate	17 MB/sec	170 MB/sec
Full duplex	No	Yes
Concurrent I/O operations	1	Up to 32
CCW execution	Synchrony	Asynchrony (FICON)

When should I (not) use SCSI?

When should I use SCSI?

- When enough CPU power is available to handle the additional SCSI overhead.
  - SCSI overhead is mostly non-parallel code.

When should I not use SCSI?

- When you are already CPU constraint.
  - If you are today running at 80% CPU utilization, SCSI would fill up your CPU up to 100 %

z/VSE 3.1 Hardware support

- **z/VSE 3.1 runs on the following machines**
  - IBM System z9 or z10
  - zSeries: z800, z900, z990, z890
  - 9672 Parallel Enterprise Server (G5/G6)
  - Multiprice 3000 (7060)
  - equivalent emulators (Flex-ES)
- **z/VSE 3.1 is based on the hardware instruction set described in the manual 'ESA/390 Principles of Operation' (SA22-7201).**
  - It is assumed that all the ESA/390 instructions and facilities described in that manual can be used.

z/VSE 3.1 Hardware support - continuation

- **z/VSE 3.1 is designed to support:**
  - IBM eServer zSeries 890, 990, IBM System z9 or z10
  - SCSI disks attached to zSeries FCP channels
  - OSA-Express2 and FICON Express2 adapters
  - Crypto Express2 and CP Assist for Cryptographic Function (CPACF)
  - IBM TotalStorage 3494 Virtual Tape Server
  - improved support for IBM 3494 Tape Library
  - IBM TotalStorage DS8000 and DS6000 series Storage Servers
  - IBM TotalStorage Enterprise Storage Server (ESS)

Supported VSE Releases

VSE Release	Available	End of Marketing	End of Service
z/VSE 4.2	10/17/2008		
z/VSE 4.1	03/16/2007	10/17/2008	04/30/2010
z/VSE 3.1	03/04/2005	05/31/2008	07/31/2009
VSE/ESA 2.7	03/14/2003	09/30/2005	02/28/2007 (out of service)
VSE/ESA 2.6	12/14/2001	03/14/2003	03/31/2006 (out of service)
VSE/ESA 2.5	09/29/2000	12/14/2001	12/31/2003 (out of service)
VSE/ESA 2.4	06/25/1999	09/29/2000	06/30/2002 (out of service)
VSE/ESA 2.3	07/12/1997	06/30/2000	12/31/2001 (out of service)

VSE Server Support

IBM Server	z/VSE 4.2	z/VSE 4.1	z/VSE 3.1	VSE/ESA 2.7 (*)	VSE/ESA 2.6 (*)	VSE/ESA 2.5 (*)	VSE/ESA 2.4/2.3 (*)
IBM System z10 BC/EC	Yes	Yes (PTF required)	Yes (PTF required)	Yes (PTF required)	Yes (PTF required)	Yes (PTF required)	No
IBM System z9 BC/EC (z9-109)	Yes	Yes	Yes	Yes	Yes (PTF required)	Yes (PTF required)	No
zSeries 890, 990	Yes	Yes	Yes	Yes	Yes (PTF required)	Yes (PTF required)	No
zSeries 800, 900	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S/390 Parallel Enterprise Server G5/G6	No	No	Yes	Yes	Yes	Yes	Yes
S/390 Multiprise 3000	No	No	Yes	Yes	Yes	Yes	Yes
S/390 Parallel Enterprise Server G3/G4	No	No	No	No	Yes	Yes	Yes
S/390 Multiprise 2000	No	No	No	No	Yes	Yes	Yes
S/390 Integrated Server	No	No	No	No	Yes	Yes	Yes
S/390 Parallel Enterprise Server G2 / G1 (out of Service)	No	No	No	No	Yes	Yes	Yes
ES/9000 – 9221, 9121, 9021 (out of Service)	No	No	No	No	Yes	Yes	Yes
P/390 and R/390 (out of Service)	No	No	No	No	Yes	Yes	Yes

(*) Note: Although VSE/ESA 2.7 or earlier releases technically run on selected servers, these releases are Out-of-Service anyway.

VSE Hardware Support

VSE Release	HiperSockets	OSA Express (QDIO mode)	Hardware Crypto
z/VSE 4.2	Yes	Yes	Yes (PCICA, CEX2C, CEX2A, CPACF)
z/VSE 4.1	Yes	Yes	Yes (PCICA, CEX2C, CEX2A, CPACF)
z/VSE 3.1	Yes	Yes	Yes (PCICA, CEX2C, CEX2A, CPACF)
VSE/ESA 2.7	Yes	Yes	Yes (PCICA, CPACF)
VSE/ESA 2.6	No	Yes	No
VSE/ESA 2.5 or earlier	No	No	No

Crypto Card	z800	z900	z890	z990	z9 BC/EC	z10 BC/EC
PCICA	No	Yes	Yes	Yes	No	No
CEX2C	No	No	Yes	Yes	Yes	Yes
CPACF	No	No	Yes	Yes	Yes	Yes
CEX2A	No	No	No	No	Yes	Yes

IBM Processor Capacity Reference for zSeries (zPCR)

- **The zPCR tool was released for customer use on October 25, 2005**
  - **http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS1381**
  - **'As is', no official support, e-mail to zpcr@us.ibm.com**
- **PC-based productivity tool under Windows**
- **It is designed to provide capacity planning insight for IBM System z9/z10 and eServer zSeries processors running various workload environments**
- **Capacity results are based on IBM's LSPR data supporting all IBM System z9/z10 and eServer zSeries processors**
  - Large System Performance Reference: http://www.ibm.com/servers/eserver/zseries/lspr/
- **For VSE use z/VSE workloads Batch, Online or Mixed**

zSeries Remarks – Split cache

- **Prior to zSeries there is one cache for data and instructions**
- **zSeries has split data and instruction cache**
- **Performance implications:**
  - If program variables and code that updates these program variables are in the same cache line (256 byte)
    - Update of program variable invalidates instruction cache
    - Performance decrease if update is done in a loop
  - See APAR PQ66981 for FORTRAN compiler

zSeries Remarks – Split cache - example

Killer example:

```
*    prepare length
BCTR    R2,0    ADJUST FOR SS-INSTR.
STC     R2,*+5
MVC     RECEIVER(*-*),SENDER
```

STC instruction modifies the next instruction to set the length.

Better code:

```
*    prepare length
BCTR    R2,0  ADJUST FOR SS-INSTR.
EX      R2,MVC01
...
MVC01   MVC    RECEIVER(*-*),SENDER
```

Use EXECUTE instruction instead.

zSeries Performance: Processor Design Considerations:
http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/FLASH10208

zSeries Remarks – Split cache - example

Not causing a problem:

```
         LA    R1,PHASNAME   POINT AT PHASE NAME
         CDDELETE (1)
+*       SUPERVISOR - CDDELETE - 5686-032-06
+        CNOP  0,4
+        BAL   15,*+8
+        DC    A(B'00010010')
+        L     15,0(,15)
+        SVC   65            ISSUE SVC FOR CDDELETE
         DS    0H
```

CDDELETE uses an inline flag byte, but does not modify it

Can cause a problem:

```
         WTO TEXT=DATA
+        CNOP  0,
+        BAL   1,IHB0003A   BRANCH AROUND MESSAGE
+        DC    AL2(8)                TEXT LENGTH
+        DC    B'0000000000010000'   MCSFLAGS
+        DC    AL4(0)               MESSAGE TEXT ADDR
         ...
+IHB0003A DS   0H
+        LR    14,1         FIRST BYTE OF PARM LIST
+        SR    15,15        CLEAR REGISTER 15
+        AH    15,0(1,0)    ADD LENGTH OF TEXT + 4
+        AR    14,15        FIRST BYTE AFTER TEXT
+        LA    15,DATA      LOAD TEXT VALUE
+        ST    15,4(0,1)    STORE ADDR INTO PLIST
+*       SUPERVISOR - SIMSVC - 5686-032
         ...
+        SVC   35           ISSUE SVC 35
 @GE00016 DS   0H
```

WTO uses an inline parameter list, but modifies the parameter list

Note: WTO can be coded with an external parameter list: WTO …,MF=(E,addr)

z890, z990, z9 and z10 Considerations

- **The z890, z990, z9 and z10 are LPAR-only machines**
  - No basic mode any more
  - Even if you run just one VSE system, it now runs in an LPAR
  - Running z/VSE systems under z/VM means
    - running z/VSE in z/VM in an LPAR
  - No I/O Assist in LPARs
    - Only available if z/VM runs in basic mode, but no basic mode available on z890, z990, z9 and z10

z/VM V5 considerations

- **z/VM V5 no longer supports V=R and V=F guests**
- **z/VM V5 no longer support I/O Assist**
  - If you currently run with preferred guests, you will need to estimate and plan for a likely increase in processor requirements as those preferred guests become V=V guests as part of the migration.
  - Refer to Preferred Guest Migration Considerations at http://www.vm.ibm.com/perf/tips/z890.html for assistance and background information
- **How to size the impact (on your current system)**
  - **Loss of I/O Assist:** Run your workload with CP SET IOASSIST OFF and measure the increase
  - **Loss of V=R/F:** Run your workload with V=V and use the CP Monitor to watch for increased CPU consumption
- **How to tune**
  - **Dedicated processors:** CP SET SHARE ABSOLUTE
  - **Dedicated memory:** CP SET RESERVED
  - **I/O Assist:** Use minidisks, turn minidisk caching on (MDC)

Possible performance issues with PPRC

- **Issue may occur if**
  - PPRC is used
  - VSE runs in native or in LPAR
  - Not all devices that are defined in IOCP are also defined in VSE ADD statements
- **In case there is an PPRC state change, interrupts are sent to all LPARs where the related device are defined in IOCP.**
  - If the device is defined in VSE ADD, no problem occurs: VSE will process the interrupt correctly.
  - If the device is NOT defined in VSE ADD, the interrupt is ignored by VSE and the interrupt is resent very quickly to that LPAR
    - Results in very high channel activity (up to 100%)
- **Solution:**
  - Define ALL devices in VSE ADD that are defined in IOCP

VSE/POWER POFFLOAD Performance Issues

- **Caused by incompatibility between VSE/POWER tape format and new tape drives**
- **3490F empties cache for FSF used by POFFLOAD LOAD**
  - Install DY46164/DY46245 for VSE/ESA 2.7/2.6
- **3590 synchronizes cache with tape for each WTM**
  - Install microcode FC0520 on A60 controller + VSE/AF APAR DY45817 + AR command TAPE WTM=NOSYNC
  - Unfortunately controller A50 is to small to install FC0520

IBM TotalStorage DS6000

- **Designed and priced to lower the total cost of ownership**
- **For medium and large enterprises**
- **Open systems and mainframe host attachment**
- **Advanced copy services**
  - equivalent to and interoperable with DS8000 series and ESS 800 and 750 systems
- **Includes the IBM TotalStorage DS Storage Manager**
  - GUI interface and Express Configuration wizards
- **Using modular, 3U, 16 disk drive, rack-mountable enclosures**
- **Up to 67.2TB physical storage**
- **See:**
  - http://www.ibm.com/servers/storage/disk/ds6000/index.html

IBM TotalStorage DS8000

- **Robust, flexible and cost-effective disk storage for mission-critical workloads**
- **IBM's first implementation of storage system logical partitions (LPARs) using IBM Virtualization Engine technology**
- **Up to 192TB of physical storage**
- **Supports extensive connectivity:**
  - Fibre Channel
  - FICON
  - ESCON
- **Support storage sharing for a wide variety of servers**
  - zSeries
  - pSeries, eServer p5
  - iSeries, eServer i5
  - xSeries and other Intel based servers,
  - Sun and Hewlett-Packard
- **See: http://www.ibm.com/servers/storage/disk/ds8000/index.html**

3494 Virtual Tape Server

- **Can help reduce real tape mounts, because many mount requests are satisfied from the Tape Volume Cache (TVC)**
- **Can reduce physical tape cartridges required because of higher utilization of cartridge capacity**
- **Can help reduce operating costs such as power, maintenance, operations and support staff**
- **Can help reduce floor space required to support the tape process, as a result of fewer physical resources**
- **Can help to improve performance due to the elimination of most of the physical movement of tape**
- **Can be upgraded to Peer-to-Peer configuration to support business continuance**
- **See:**
  - http://www.ibm.com/servers/storage/tape/3494vts/index.html

3494 Tape Library

- **Designed to provide reliable, scalable tape automation**
- **Provides multiplatform connectivity**
- **Supports 3592 rewritable and WORM cartridges**
- **Supports multiple IBM tape drive models**
- **Supports IBM Virtual Tape**
- **By combining the various models of the 3494 Tape Library, you can create an automated tape library of up to 16 library frames that can contain over 6000 tape cartridges and up to 5.6PB of stored data.**
- **See:**
  - http://www.ibm.com/servers/storage/tape/3494/index.html

Turbo Dispatcher - Overview

- **Turbo Dispatcher**
  - available since 1995
  - VSE/ESA 2.1-2.3 Standard and Turbo Dispatcher
  - since VSE/ESA 2.4 only Turbo Dispatcher
  - last changes:
    - VSE/ESA 2.6.2 (APAR DY45869)
    - VSE/ESA 2.7.0 (APAR DY45926)
  - Supports basic (native), LPAR and VM mode
  - Runs on Uni- and n-Way-procerssors
    - CPUs have "equal" rights
    - more than 3 CPUs are not recommended

Turbo Dispatcher - Overview (2)

- **IPL is done on 1 CPU only**
  - after IPL other CPUs can be started
  - CPUs can be started or stopped without re-IPL
  - at least 1 CPU (IPL CPU) must always be active

```
SYSDEF TD,START=n|ALL
SYSDEF TD,STOP=n|ALL
SYSDEF TD,STOPQ=n|ALL
QUERY TD
```

Turbo Dispatcher - Quiesced CPUs

- **SYSDEF TD,STOPQ=n to set a CPU in quiesced mode**
  - Implemented for z/VM guest systems
    - Not started guest CPUs stop IOASSIST
    - STOPQ remains IOASSIST active, and avoids TD Overhead, (CPU will no longer participate in work unit selection)
    - quiesced CPUs will not process any work units
    - quiesced CPUs will not handle any interrupt
    - quiesced CPUs can be started with SYSDEF TD,START

Turbo Dispatcher - Design

- **TD dynamically assigns partitions to CPUs**
  - Work unit = from assignment to one CPU until next interrupt/SVC
  - If one task (subtask) of a partition is active, no other task of the same partition will be selected
  - TD dispatches on partition-basis, not on task-basis
  - A job running in a partition is processed in several work units.

Turbo Dispatcher - Design (2)

- **parallel work units**
  - application code (CICS, Batch)
  - may run on any CPU concurrently with other parallel or non-parallel work units.
- **non-parallel work units**
  - system code (Services, VTAM, Vendor code)
  - As long as one non-parallel work unit is active on one CPU, no other non-parallel work unit can execute on any other CPU.

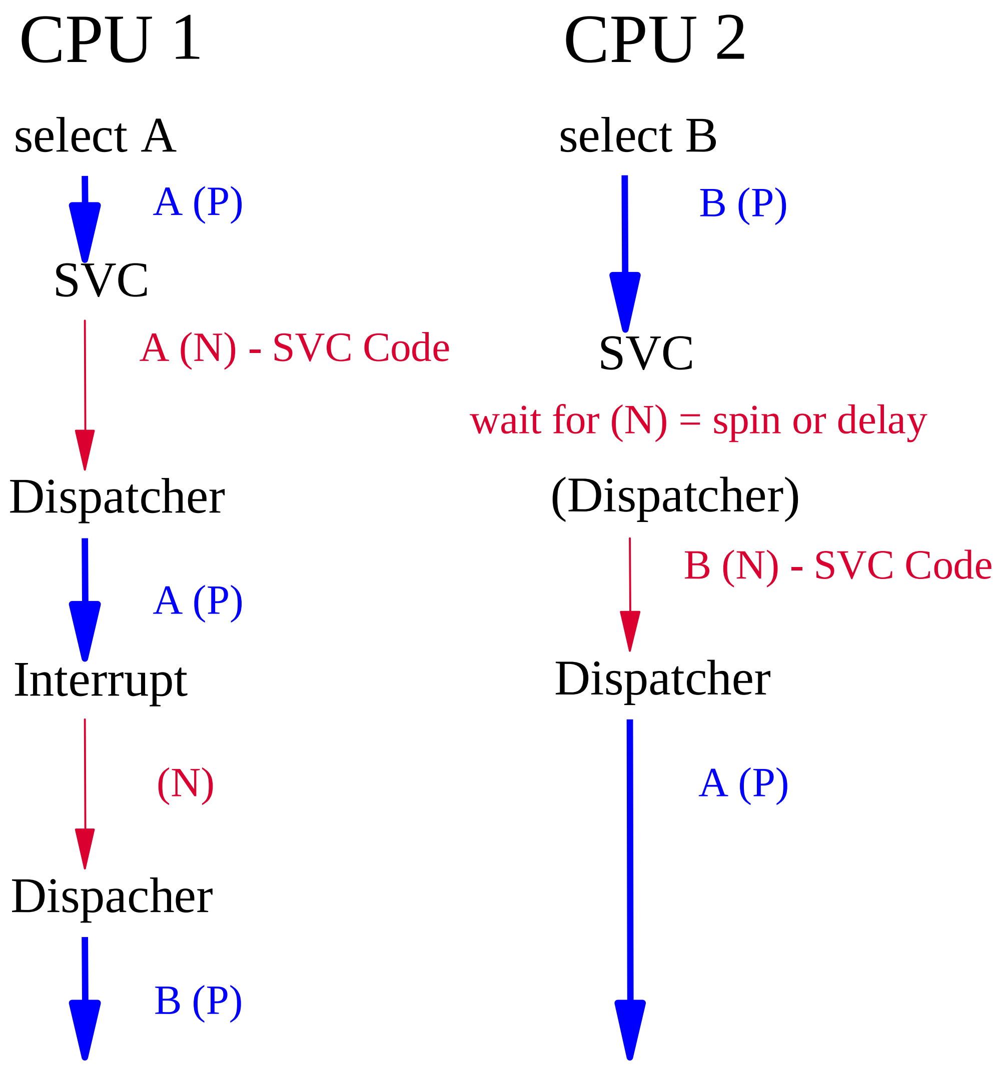
Turbo Dispatcher - Design - Example 1

Job A	A1 (N)	A2 (N)	A3 (P)
Job B	B1 (P)	B2 (N)	B3 (P)
Job C	C1 (P)	C2 (P)	C3 (P)

Ax, Bx, Cx = workunits of job A, B, C
(N) = non-parallel work unit
(P) = parallel work unit

	CPU 1	CPU 2
Step 1	A1 (N)	B1 (P)
Step 2	C1 (P)	A2 (N)
Step 3	B2 (N)	A3 (P)
Step 4	C2 (P)	B3 (P)
Step 5		C3 (P)

Turbo Dispatcher - Design - Example 2



CPU 1

select A

A (P)

SVC

A (N) - SVC Code

Dispatcher

A (P)

Interrupt

(N)

Dispacher

B (P)

CPU 2

select B

B (P)

SVC

wait for (N) = spin or delay

(Dispatcher)

B (N) - SVC Code

Dispatcher

A (P)

Turbo Dispatcher - Exploitation

- **Uni-Processor**
  - new Partition Balancing Concept
    - Helps to set priorities of partitions
  - Determination of non-parallel share, to find out if a 2. or 3. CPU would be of use
- **n-Way Processors (2-3 CPUs)**
  - System tuning required for exploitation
  - Increased Capacity (dependent on workload)
    - Exploitation increases by reduction of non-parallel work units

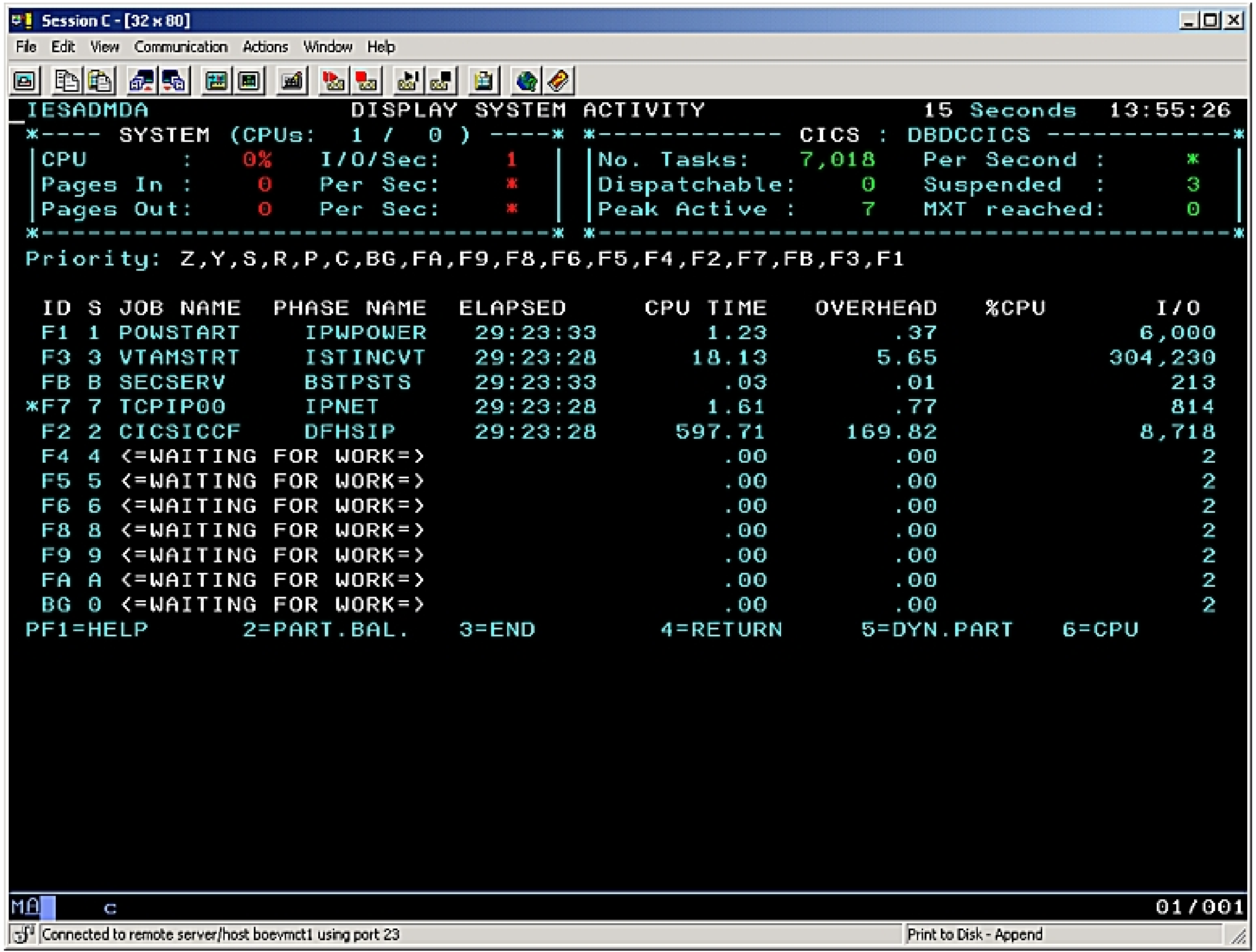
Turbo Dispatcher - CPU time measurement

- **CPU time measurement (overall system)**
  - SYSDEF TD,RESETCNT
  - Workload (e.g. run a job)
  - QUERY TD (QUERY TD,INTERNAL)

```
 CPU   STATUS    SPIN_TIME    NP_TIME TOTAL_TIME NP/TOT
  00   ACTIVE            0     237100     416698  0.568
  01   ACTIVE            0     157556     415229  0.379
  02   QUIESCED          0          0          0  *.***
  03   INACTIVE
                ---------------------------------------
TOTAL                    0     394656     831927  0.474

              NP/TOT: 0.474     SPIN/(SPIN+TOT): 0.000
 OVERALL UTILIZATION: 179%        NP UTILIZATION:  85%

 ELAPSED TIME SINCE LAST RESET:       463433
```

NP/TOT = non-paralell share (NPS)
SPIN_TIME = CPU time waiting for NP

Display System Activity Dialog


Session C - [32 x 80]
File Edit View Communication Actions Window Help
MA c 01/001
Connected to remote server/host boevmct1 using port 23
Print to Disk - Append

```
 IESADMDA                    DISPLAY SYSTEM ACTIVITY          15 Seconds   13:55:26
 *---- SYSTEM (CPUs:  1 /  0 ) ----* *------------ CICS : DBDCCICS ------------*
 |CPU       :   0%   I/O/Sec:     1 | |No. Tasks:   7,018   Per Second :      * |
 |Pages In  :    0   Per Sec:     * | |Dispatchable:     0   Suspended  :      3 |
 |Pages Out:     0   Per Sec:     * | |Peak Active :     7   MXT reached:      0 |
 *----------------------------------* *----------------------------------------*
 Priority: Z,Y,S,R,P,C,BG,FA,F9,F8,F6,F5,F4,F2,F7,FB,F3,F1

  ID S JOB NAME  PHASE NAME  ELAPSED      CPU TIME   OVERHEAD   %CPU       I/O
  F1 1 POWSTART    IPWPOWER   29:23:33        1.23        .37           6,000
  F3 3 VTAMSTRT    ISTINCVT   29:23:28       18.13       5.65         304,230
  FB B SECSERV     BSTPSTS    29:23:33         .03        .01             213
 *F7 7 TCPIP00     IPNET      29:23:28        1.61        .77             814
  F2 2 CICSICCF    DFHSIP     29:23:28      597.71     169.82           8,718
  F4 4 <=WAITING FOR WORK=>                    .00        .00               2
  F5 5 <=WAITING FOR WORK=>                    .00        .00               2
  F6 6 <=WAITING FOR WORK=>                    .00        .00               2
  F8 8 <=WAITING FOR WORK=>                    .00        .00               2
  F9 9 <=WAITING FOR WORK=>                    .00        .00               2
  FA A <=WAITING FOR WORK=>                    .00        .00               2
  BG 0 <=WAITING FOR WORK=>                    .00        .00               2
 PF1=HELP      2=PART.BAL.    3=END        4=RETURN     5=DYN.PART    6=CPU
```

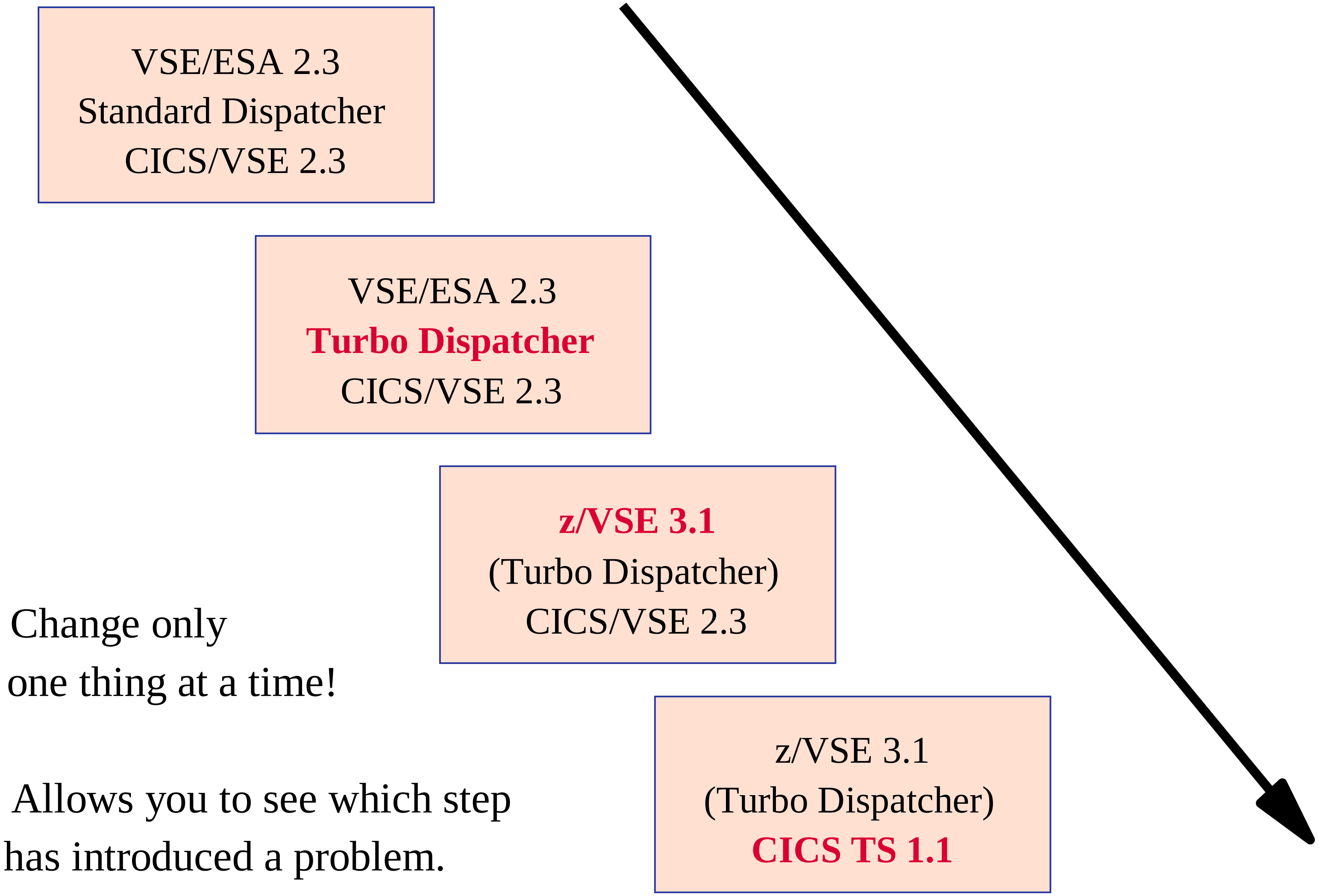
Migration aspects

- **Consider hard-/software requirements:**
  - Does my largest partition still fit into a single CPU of the target processor?
    - Note: a partition can only run on 1 CPU at a time!
  - Is the processor capacity and speed still sufficient to run the workload?
  - Does multiprocessing help to run the workload?
    - What about non-parallel share (on 1-Way)?
    - Are there many parallel batch jobs?
      - A large CICS partition does not benefit of a 2. CPU

Migration overhead

- **Uni-Processor**
  - increased overhead because of
    - Release migration (VSE/ESA 2.6/2.7 vs. z/VSE 3.1)
    - TD overhead (Standard Dispatcher vs. TD)
    - CICS/VSE vs. CICS TS
- **N-Way Processor**
  - CPU time increases when migrating from uni to n-Way Processor (for the same workload)
    - For PACEX Workload: Factor 1.4 (2 CPUs)
    - TD overhead for multiprocessor exploitation
    - z/VM Overhead

Migration path



VSE/ESA 2.3
Standard Dispatcher
CICS/VSE 2.3

VSE/ESA 2.3
Turbo Dispatcher
CICS/VSE 2.3

z/VSE 3.1
(Turbo Dispatcher)
CICS/VSE 2.3

z/VSE 3.1
(Turbo Dispatcher)
CICS TS 1.1

Change only
one thing at a time!

Allows you to see which step
has introduced a problem.

Performance Tips

- **A partition can only exploit 1 CPU at a time**
  - 2 CPUs do not have any benefit for a CICS partition
  - Use as many partitions as required for selected n-way
- **Use/define only as many CPUs as really needed**
  - additional CPUs create more overhead, but no benefit
- **Partitions setup**
  - Set up more batch and/or (independent) CICS partitions
  - Split CICS production partitions into multiple partitions

Performance Tips (2)

- **1 CPU must be able to handle all non-parallel workload**
- **Non-parallel code limits the n-Way exploitation**
  - QUERY TD: NP/TOT = NPS (non parallel share)
  - Measure NPS before migration
  - max CPUs = 0.8 - 0.9 / NPS

NPS	#CPUs	NPS	#CPUs
0.20	4.0-4.5 (4)	0.45	1.8-2.0 (2)
0.25	3.2-3.6 (3)	0.50	1.6-1.8 (2)
0.30	2.7-3.0 (3)	0.55	1.5-1.6 (2)
0.35	2.3-2.6 (2)	0.60	1.3-1.5 (1)
0.40	2.0-2.2 (2)	0.65	1.2-1.4 (1)

Performance Tips (3)

- **Non-parallel code limits the maximum MP exploitation**
- **System code (Key 0) increases non-parallel share**
  - Vendor code can have significant impact
- **Overhead increases when NP code limits throughput**
- **Data In Memory (DIM) reduces non-parallel code**
  - less system calls (I/Os)
  - may increase throughput
- **Change VSE/POWER startup to WORKUNIT=PA**
- **In general ONE faster CPU is better than multiple slower ones**
  - Even if sum of slower CPUs is higher than one faster CPU

CICS Implications

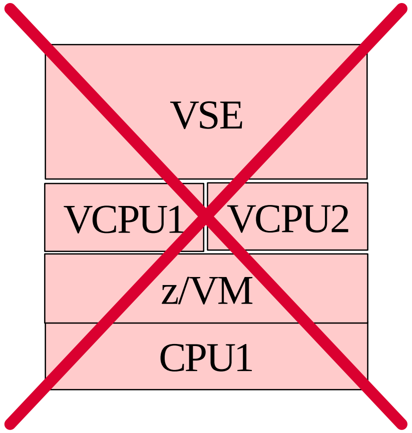
- **Single CICS**
  - Can consume processing power of one CPU only
  - parallel batch jobs may exploit 2. CPU
- **Multiple CICS partitions**
  - Number of CPUs depends on non-parallel share (NPS)
  - Function shipping and Transaction routing
    - AOR, TOR, FOR

Partition Balancing

- **Balanced Group is defined with PRTY:**
  - PRTY BG,C=F5=F8,F2,F3,F1
  - Each partition/class of the group has a default-SHARE (100)
  - Dynamic partitions gets the SHARE of its class
- **To set a SHARE (1-1999)**
  - PRTY SHARE,F5=50
  - SHARE = 0 means the lowest priority within the group

```
PRTY
AR 0015 PRTY BG,C=F5=F8,F2,F3,F1
AR 0015
AR 0015 SHARE F5=  50, F8= 100,  C= 100
MSECS
AR 0015 MSECS   976      <---- influences task selection
```

Do's and Don't Do's



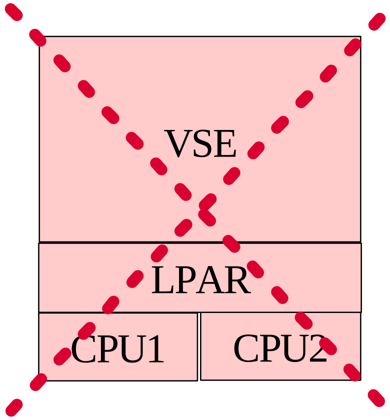
VSE

VCPU1 VCPU2

z/VM

CPU1

no virtual CPUs!
(creates overhead)

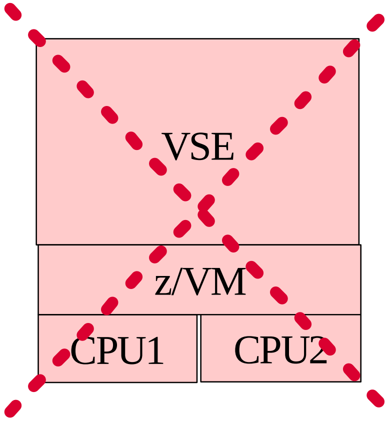


VSE

LPAR

CPU1 CPU2

only if NPS < 4.5

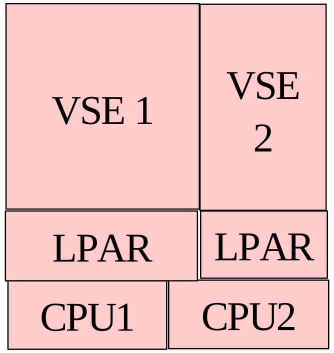


VSE

z/VM

CPU1 CPU2

only if NPS < 4.5

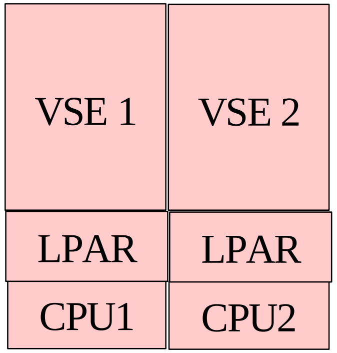


VSE 1 | VSE 2

LPAR | LPAR

CPU1 | CPU2

VSE 1 = Production
VSE 2 = Test

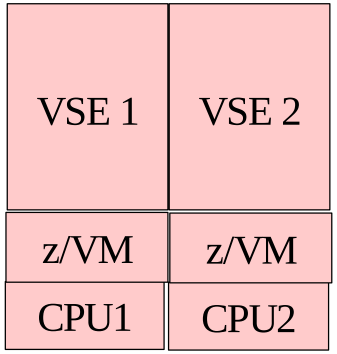


VSE 1 | VSE 2

LPAR | LPAR

CPU1 | CPU2

dedicated CPU
per VSE



VSE 1 | VSE 2

z/VM | z/VM

CPU1 | CPU2

dedicated CPU
per VSE

Do's and Don't Do's (2)

The fastest uni-processor is (almost always *) the best processor

(*) from a single VSE-image point o view

VSE Health Check

- **Goals**
  - Recognize actual/upcoming problems
  - Optimize the system for new/current workload
- **A-B-C analysis**
  - A - concentrate on the essentials
    - 20 % work for 80 % results
  - B - more detailed analysis
    - 30 % work for 15 % results
  - C - analyze all details
    - 50 % work for 5 % results
- **A-B analysis takes about 2 days**
- **C analysis takes about 1 week**
- **Should be done about once a year**

VSE Health Check - continued

- **What should be checked?**
  - Processor (utilization, dispatching, z/VM, ...)
  - DASD, Tapes (I/O rate, cache, ...)
  - Network (network load, misrouted packets, ...)
  - System software
    - Turbo Dispatcher (PRTY, PRTY SHARE, ...)
    - VSAM (CA/CI sizes, share options, buffers, ...)
    - CICS (MXT, DSA/EDSA sizes, SOS, ...)
    - Storage Layout (GETVIS 24, SVA, partitions, DSPACE, ...)
    - VTAM (buffer pool)
    - POWER (DBLK, DBLKGP, ...)
    - LE runtime options (Heap size, ...)
  - Application software
  - New Tool: VSE Health Checker
    http://www.ibm.com/servers/eserver/zseries/zvse/downloads/#healthchecker

Hints and Tips for Performance

- **Try to exploit Turbo Dispatcher functions**
  - Priority settings
  - Partition balancing
  - Partition balancing groups
- **Use as much data in memory (DIM) as possible**
  - CICS Shared Data Tables
  - Large/many VSAM Buffers (with buffer hashing)
  - Virtual Disks
- **Switch tracing/DEBUG off for production**

Hints and Tips for Connector and TCP/IP-Performance

- **Reduce amount of data transferred**
  - Transfer only data that is needed
  - Issue only requests that are needed
- **Use connection pooling**
  - Reduce overhead of connection establishment
- **Performance of connectors depends on**
  - Network performance
  - Performance of "server"
  - Performance of "client" or middle tier
- **Reduce misrouted packets**
- **Use a packet filter**
  - Unwanted packets increases TCP/IP and CPU load

Documentation

- **z/VSE homepage:**
  - http://www.ibm.com/servers/eserver/zseries/zvse/
- **VSE Performance:**
  - http://www.ibm.com/servers/eserver/zseries/zvse/documentation/performance.html
- **z/VM homepage:**
  - http://www.ibm.com/vm
- **z/VM 5.1 Preferred Guest Migration Considerations**
  - http://www.vm.ibm.com/perf/tips/z890.html
- **IBM TotalStorage DS8000 and DS 6000:**
  - http://www.ibm.com/servers/storage/disk/ds8000/index.html
  - http://www.ibm.com/servers/storage/disk/ds6000/index.html
- **IBM TotalStorage 3494 Virtual Tape Server:**
  - http://www.ibm.com/servers/storage/tape/3494vts/index.html
- **IBM 3494 Tape Library:**
  - http://www.ibm.com/servers/storage/tape/3494/index.html