Linux on zSeries
Disk and Tape Connectivity

Volker Sameske (sameske@de.ibm.com)
Linux auf zSeries Entwicklung
IBM Labor Böblingen

GSE Herbst-Tagung

Arbeitsgruppe VM / VSE

20.-22. September 2004

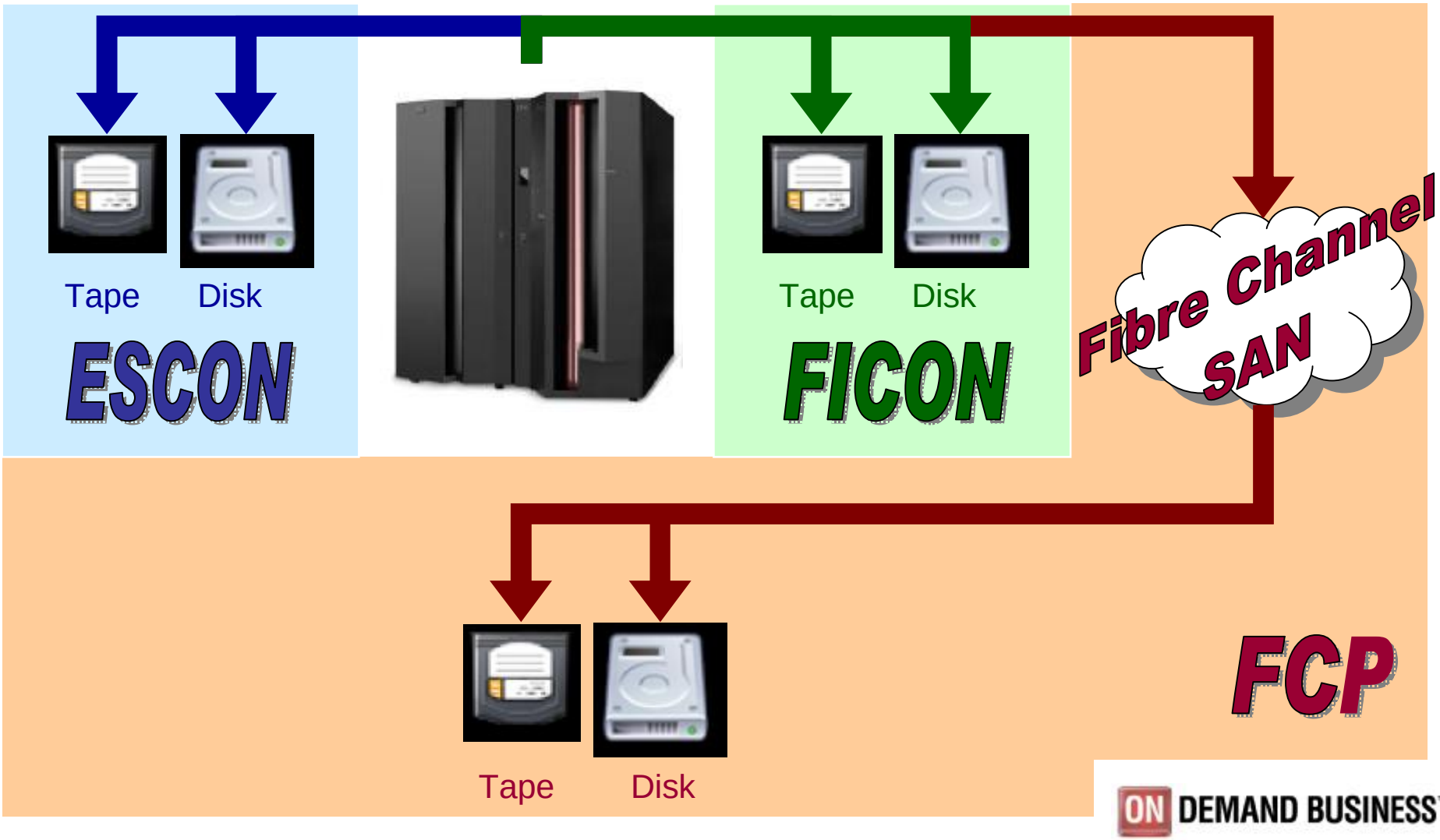
Agenda

- ESCON ⟷ FICON ⟷ FCP
- DASD ⟷ SCSI Disk
- ESCON/FICON Tape ⟷ SCSI Tape
- Performance
- DASD Konfiguration ⟷ SCSI Konfiguration
- CCW IPL ⟷ SCSI IPL

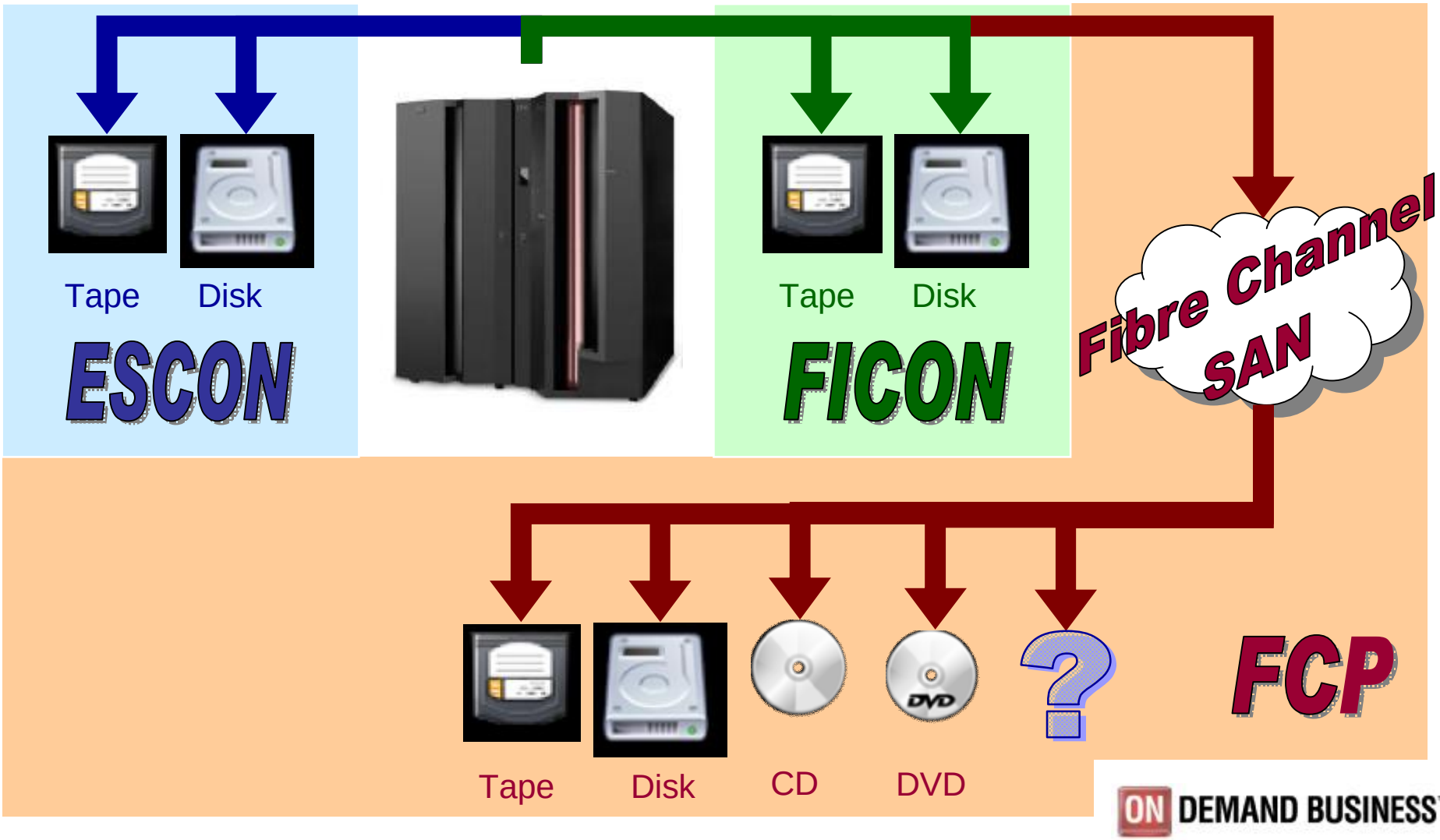
Überblick

Überblick

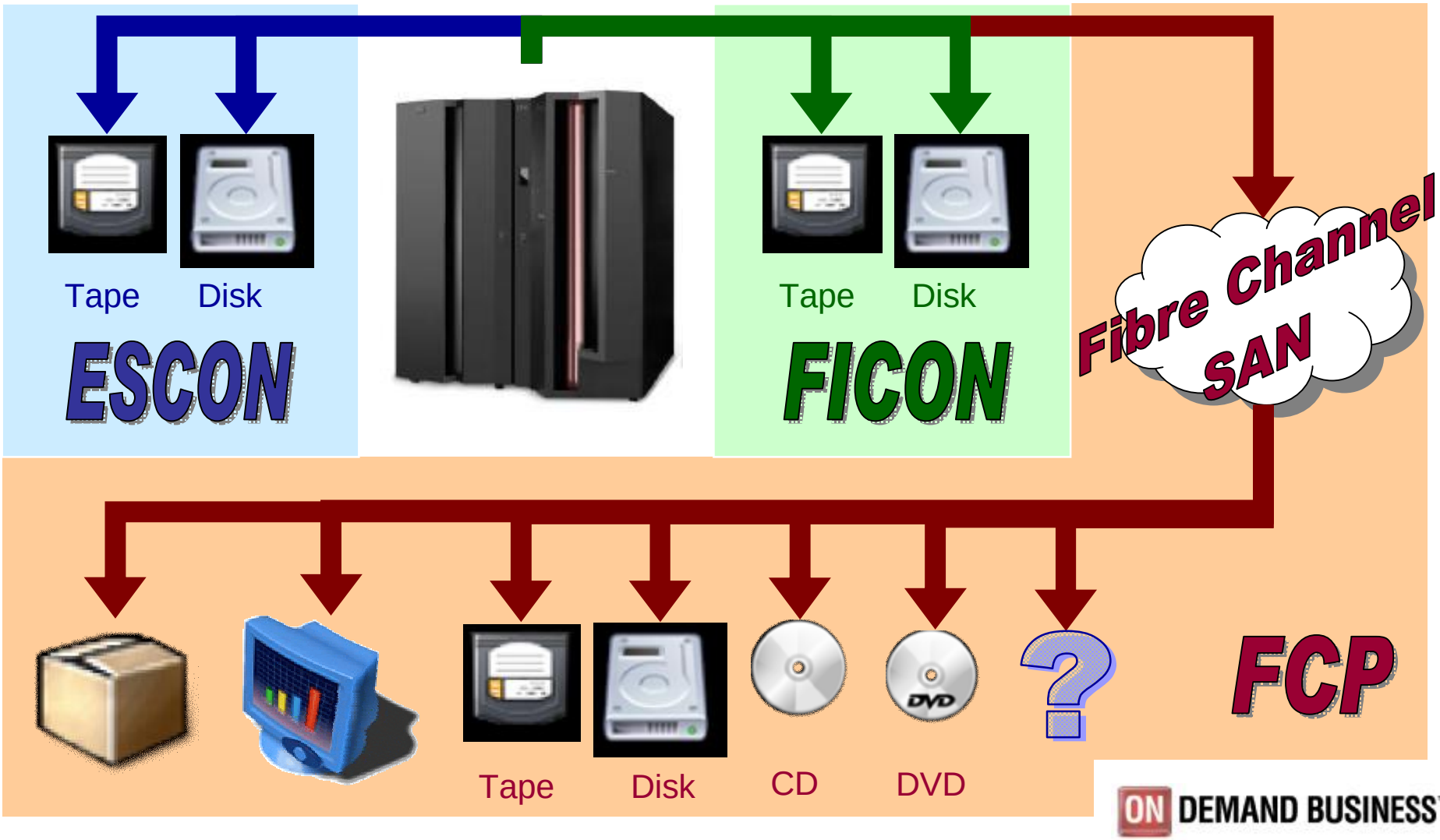
Überblick



Überblick

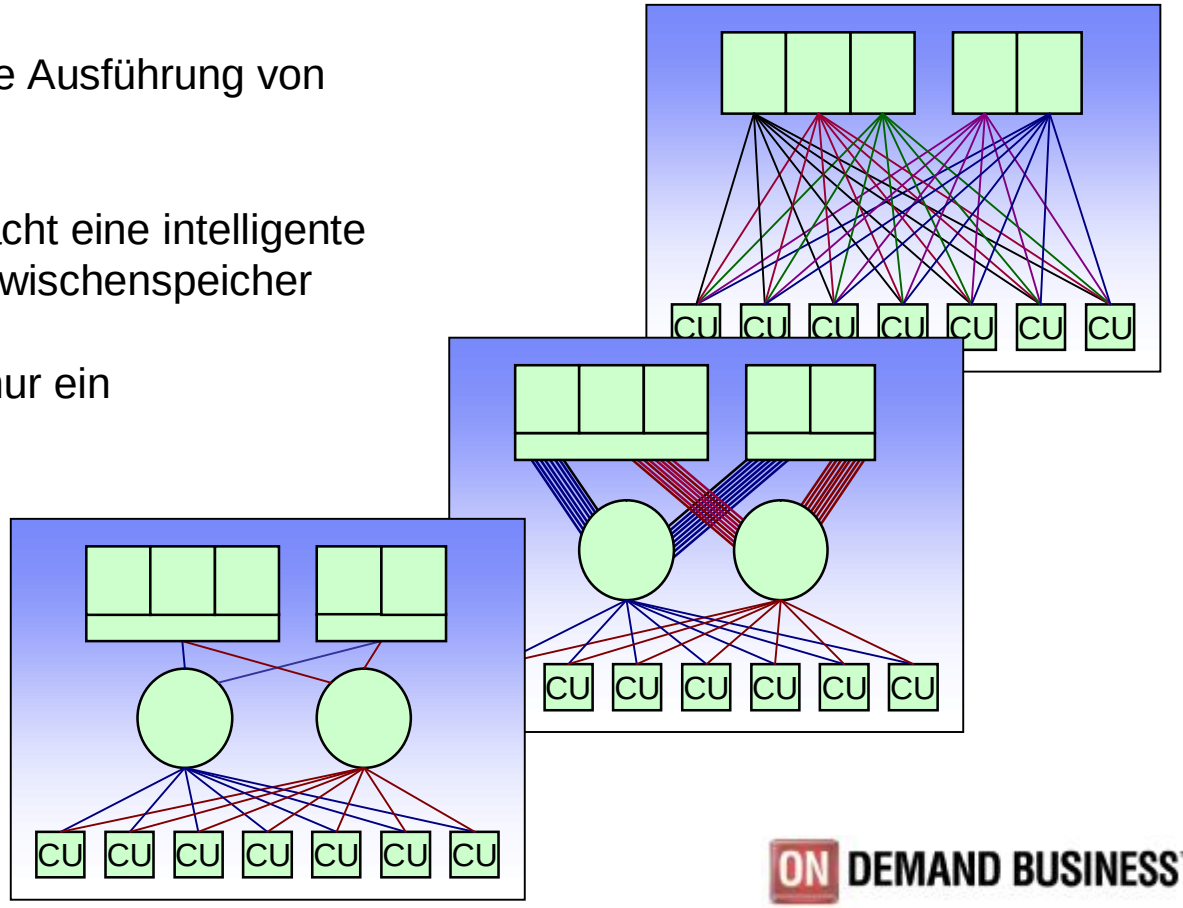


Überblick

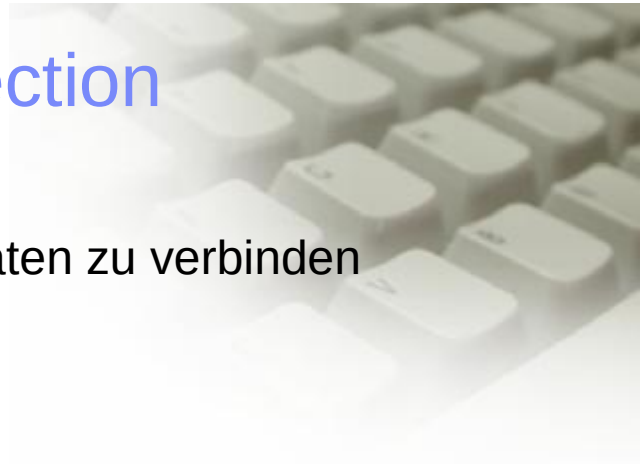


Überblick – Channel

- Ein Channel (Kanal) ist ein spezieller Prozessor, der den Dialog mit dem I/O Controller steuert.
- Grund dieses Dialoges ist die Ausführung von I/O Operationen.
- Ein Channel ist also vereinfacht eine intelligente Adapterkarte mit CPU und Zwischenspeicher
- Channels sprechen jeweils nur ein Protokoll mit dem Controller
  - 1964 – Parallel
  - 1990 – ESCON
  - 1999 – FICON
  - 2003 – FCP



ESCON – Enterprise System Connection



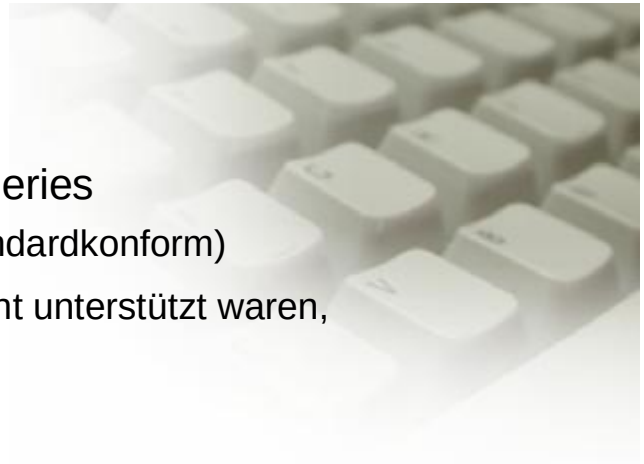
- I/O Technologie um Mainframes mit Speicher-Geräten zu verbinden
- Von IBM entwickelt
- Nur für Mainframes verfügbar
- Angekündigt im September 1990
- Ablösung der parallelen Verbindung
- Noch sehr weit verbreitet

- Daten werden bit- und byteseriell übertragen
- Wechselbetrieb: Halb-Duplex
- Control Unit antwortet auf jede einzelne Channel-Aktion
- „point-to-point connection“ – direkte Verbindung mit der Control Unit
- „switched connection“ – ESCON Director

FICON – Fibre Connection

- Reduzierte Komplexität
  - Kanal-Operationen
  - Konfiguration, Wartung
  - Infrastruktur
- Vergleich zu ESCON: Schneller, besser, weiter
  - Erweiterte Adressierung
  - Größere Entfernungen
  - Verbesserte Performance
  - Erhöhte Anzahl gleichzeitiger Verbindungen
- Protokoll
  - Fibre Channel Upper Layer Protocol (ULP)
  - Untere Protokollebenen von FICON und FCP identisch
- Reduzierte TCO (Total Cost of Ownership)
  - Weniger Kanäle und Verbindungen erforderlich (weniger Channel Extender)
  - Weniger physische Verbindungen zwischen entfernten Standorten
  - Niedrigere Leitungs-/Wartungskosten
  - ESCON-FICON Konsolidierungsverhältnis mindestens 4:1

FCP – Fibre Channel Protocol



- Größere Auswahl an Speicher-Lösungen für Linux auf zSeries
  - FCP und parallel SCSI Storage Controller und Geräte (standardkonform)
  - Neue Speicher-Geräte, die bisher mit ESCON / FICON nicht unterstützt waren,
    - CD / DVD
    - Optische Library
    - Jukebox
- Reduzierte TCO (Total Cost of Ownership)
  - Benutzung von existierender FICON Infrastruktur
  - Benutzung von existierenden SCSI / FCP Speicher-Geräten
- Einbindung von Linux auf zSeries in „Open SANs“ (Storage Area Network)
  - Einbindung des zSeries Servers in ein existierendes FC SAN
  - „Beliebige“ Disk-Größen
  - SCSI-basiertes I/O (nötig für SCSI-basierte Backup-Lösungen etc.)
- Linux/Unix Server Konsolidierung
  - Mehrere Linux Systeme auf einem einzigen zSeries Server
  - Weiterbenutzung der gleichen FCP/SCSI Storage-Geräte wie zuvor
- Derzeit noch ein paar wenige Einschränkung
  - LUN Masking, Zoning, LPAR SCSI Reboot, FCP Switch erforderlich

ON DEMAND BUSINESS™

Vergleich ESCON – FICON / FCP

	ESCON	FICON/FCP
max Entfernung (ohne Verstärker)	3km	10km
max Entfernung (Verstärker)	9km	100km
Max. Anzahl Kanäle	256	4 x 256
Max. Device-Adressen pro Link	4096	65536
Max. logische CU-Pfade pro Port	64	256
Device-Adressen pro Kanal	1024	16384
Link Rate	20 MB/s	200 MB/s
Erzielbare Transferrate	17 MB/s	170 MB/s
Vollduplex-Betrieb	NEIN	JA
Gleichzeitige I/O-Operationen	1	bis zu 32
CCW-Ausführung	Synchron	Asynchron (nur FICON)

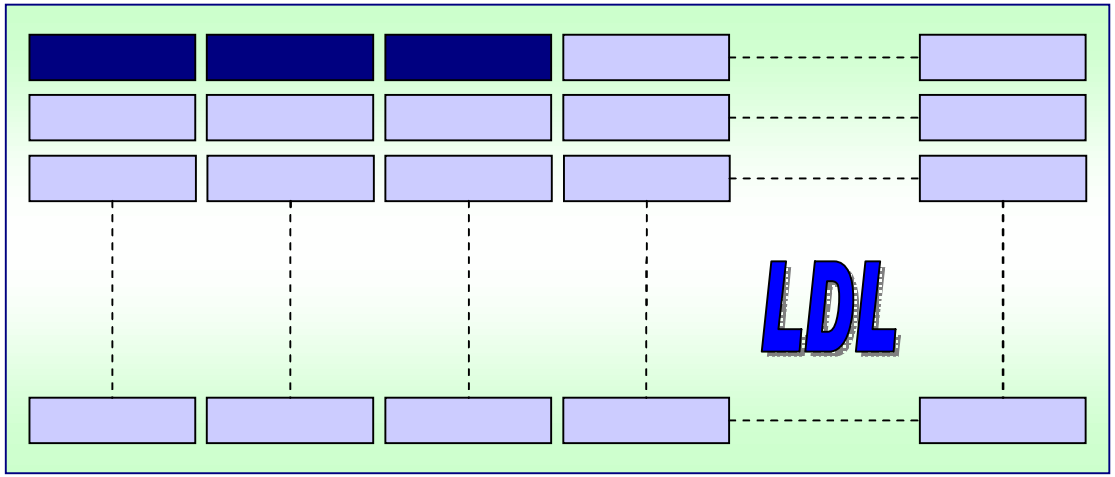
DASD (ESCON / FICON)

- DASD = Direct Access Storage Device
- Bislang Feste Größen (3390/3, 3390/9, 3390/27)
- Jetzt auch variable DASD-Größen
- DASD 3390 Model 3
  - 3339 Cylinder mit je 15 Tracks
  - Formatiert mit 4k Blockgröße ca. 2,3 GB
- ECKD Emulation durch Linux DASD Treiber

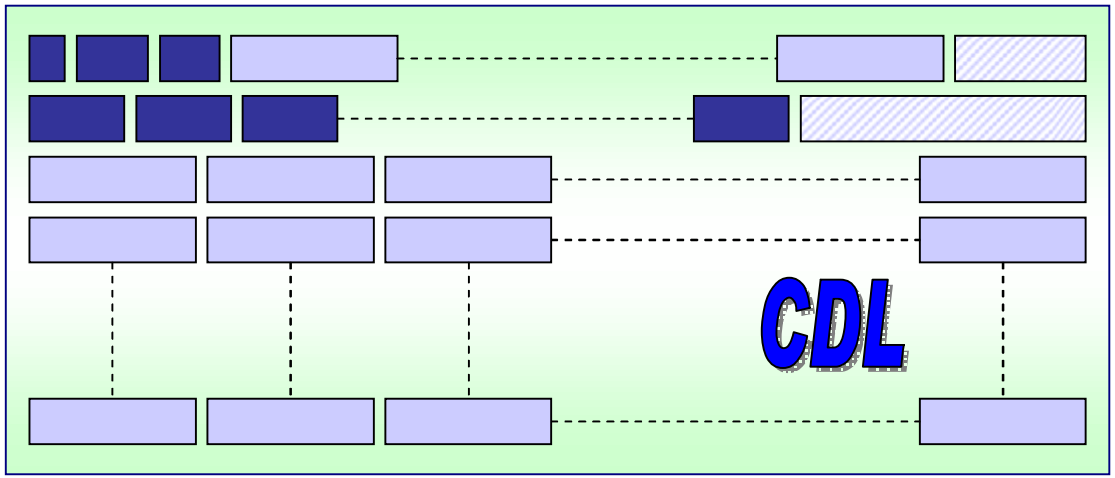
- Verschiedene DASD-Disziplinen
  - **ECKD** (Extended Count Key Data)
    - Hauptsächlich genutzte DASD Disziplin
    - Zwei verschiedene Formate (ldl und cdl)
    - Unterschiedliche Blockgröße in nicht-Linux Systemen
  - **FBA** (Fixed Block Architecture)
    - Weniger verbreitet
    - Wird wieder interessant mit z/VM 5.1
    - Feste Blockgröße (meist 512 Byte)
  - **DIAG** (DIAG-accessed)
    - Selten eingesetzt
    - Kernel 2.4 - CMS-Reserved MiniDisks
    - Kernel 2.6 - jede VM-administrierte Disk (Alternativmethode)

ON DEMAND BUSINESS™

ECKD LDL und CDL DASD Disk Layout



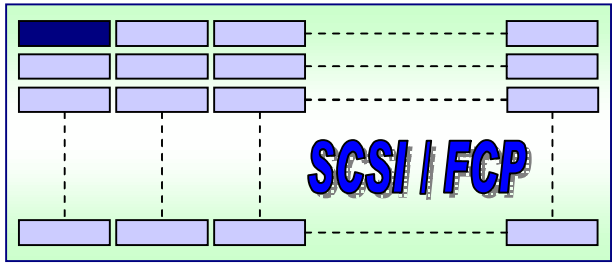
- ECKD Linux Disk Layout
  - Nur eine einzige Partition
  - Volume Label muss nicht vorhanden sein (LNX1)
  - Partition, als auch ganze DASD nutzbar
  - Problematisch mit z/VM oder z/OS



- ECKD Compatible Disk Layout
  - Bis zu 3 Partitionen möglich, mindestens 1 erforderlich
  - Volume Label ist immer vorhanden (VOL1)
  - Nur Partitionen nutzbar, nicht die ganze DASD
  - Werden von z/VM und z/OS erkannt

ON DEMAND BUSINESS™

SCSI Disk



- Feste Blockgröße von 512 Bytes
- „Beliebige“ Disk Größen
- Kaum noch „echte“ 3390 DASDs im Einsatz

- Konfiguration hauptsächlich im Linux
  - Hinzufügen von SCSI-Platten ohne IOCDS-Änderung
  - Nur der FCP Adapter wird noch im IOCDS konfiguriert

- Schneller als ECKD (Durchsatz)
  - Asynchrone, und dadurch viele parallel laufende SCSI I/O Requests
  - Kein Emulations-Overhead wie bei ECKD

- Partitionierung
  - Partitions-Tabelle im Master Boot Record
  - Bis zu 16 Partitionen möglich

ECKD DASD oder SCSI Disk

	ECKD DASD	SCSI Disk
Konfiguration	IOCDS/VM (Operator)	IOCDS/VM & Linux (Operator & Linux Admin)
Zugriffsmethode	SSCH	QDIO
Block Größe (Byte)	512, 1K, 2K, 4K	512
Disk Größe	3390 Model 3/9/27 Jetzt auch variabel	beliebig
Formatierung (low level)	dasdfmt	nicht notwendig
Partitionierung	fdasd	fdisk
Dateisystem	mke2fs (oder andere)	
Zugriff	mount	

ESCON/FICON Tape

- Tape Treiber entwickelt von IBM
- Identisch für ESCON/FICON – transparent
- Unterstützt sind 3480, 3490 und 3590 Tapes
- Keine Changer oder Tape Libraries
- Verschiedene Disziplinen tape_34xx oder tape_3590 (oco)
- Bestandteil von SLES8/9 und RHEL3 (nur tape_34xx)
- Problematisch ohne udev/hotplug

```
# lstape
TapeNo  BusID     CuType/Model  DevType/DevMod  BlkSize  State     Op   MedState
0       0.0.01a1  3490/10       3490/40         auto     UNUSED    ---  UNLOADED
1       0.0.01a0  3480/01       3480/04         auto     UNUSED    ---  UNLOADED
2       0.0.0172  3590/50       3590/11         auto     IN_USE    ---  LOADED
N/A     0.0.01ac  3490/10       3490/40         N/A      OFFLINE   ---  N/A
```

SCSI/FCP Tape

- Tape Devices:
  - IBM TotalStorage Enterprise Tape System 3590.
  - IBM TotalStorage Enterprise Tape Drive 3592.
  - IBM TotalStorage Enterprise Tape Library 3494.
  - IBM TotalStorage UltraScalable Tape Library 3582, 3583 and 3584 w/ Ultrium 2 Fibre Channel Tape Drives.
- IBMtape and IBMtapeutil packages required
  - /lib/modules/(Your system's kernel name)/kernel/drivers/scsi/IBMtape.o
  - /usr/bin/IBMtapeconfig
  - /usr/bin/IBMtaped
  - /usr/bin/IBMtapeutil

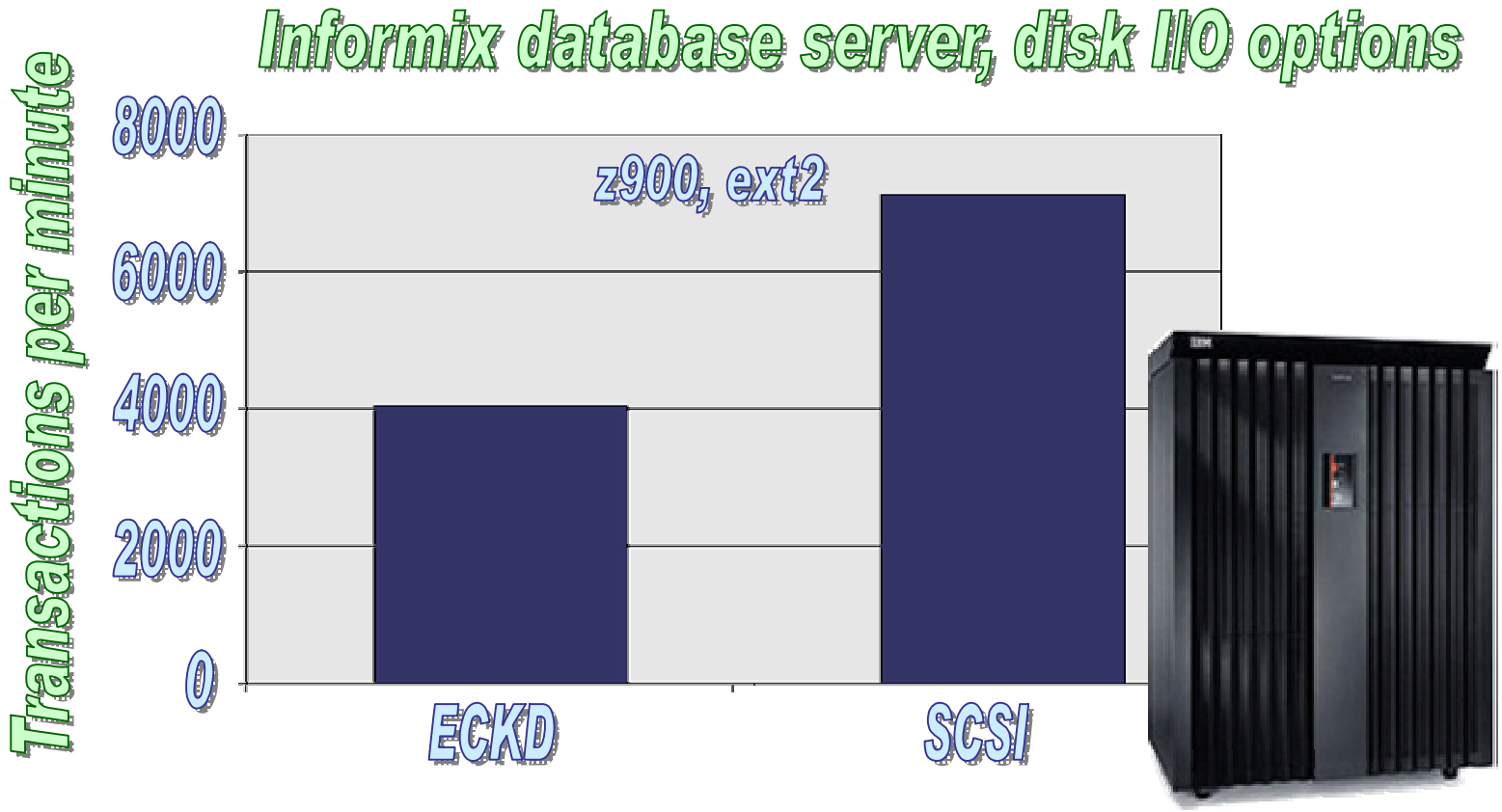
SCSI/FCP Tape

- IBMtape special files (created by IBMtapeconfig):
  - /dev/IBMtape0
  - /dev/IBMtape0n
  - /dev/IBMchanger0
- Tape utility program (IBMtapeutil):

```
# Mount cartridge from slot 3
IBMtapeutil -f /dev/IBMchanger0 mount 3

# Backup myfile.tar to tape
IBMtapeutil -f /dev/IBMtape0 write -s myfile.tar
```

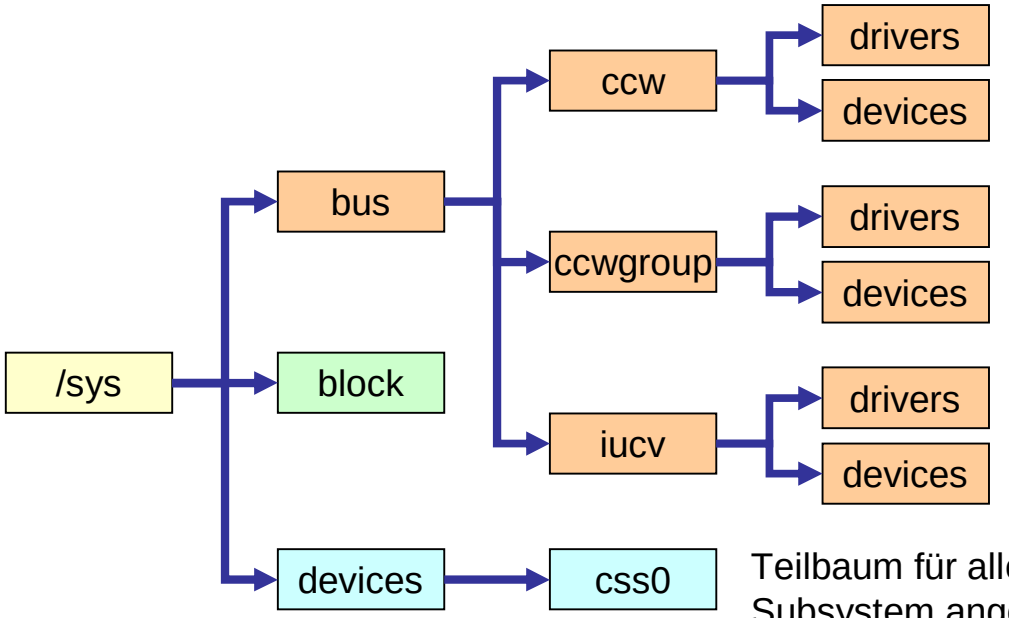
OLTP Workload Informix – I/O Options



ON DEMAND BUSINESS™

SysFS

- Neues Dateisystem mit Linux Kernel 2.6
- Enthält alle Geräte-Treiber- und Geräte-spezifischen Informationen
- KEIN Ersatz für /proc Dateisystem
- Wird benutzt für die Geräte-Konfiguration benutzt



Teilbaum für Channel Subsystem Geräte-Treiber und Geräte

Teilbaum für Group Geräte-Treiber und Geräte

Teilbaum für IUCV Geräte-Treiber und Geräte

Teilbaum für alle ans Channel Subsystem angeschlossenen Geräte und CHPIDs

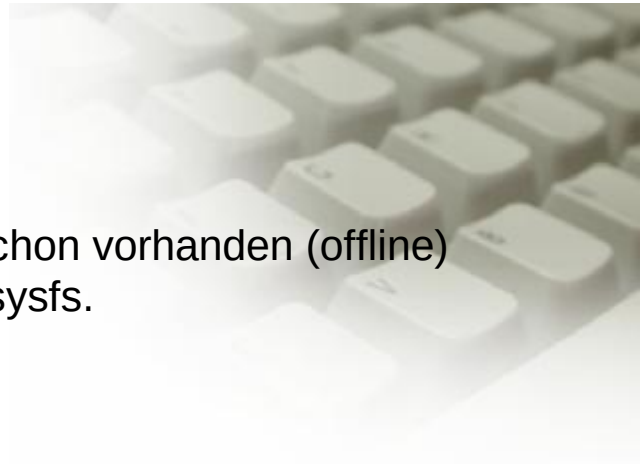
Linux 2.6 Konfiguration – DASD

- Alle vom System erkannten DASDs sind im sysfs schon vorhanden (offline)
- Jede DASD hat ein eigenes Unterverzeichnis im sysfs.
- Alle Device-Nummern im Format „0.0.xxxx“
- Einfaches aktivieren über „online“-Attribut
- Neues „lsdasd“ Tool als Ersatz für /proc/dasd/devices
- chccwdev-Script vereinfacht Konfiguration

```
# lsdasd
0.0.5ca9(ECKD) at ( 94:  0) is dasda      : active at blocksize 4096, 601020 blocks, 2347 MB
# cd /sys/bus/ccw/drivers/dasd-eckd/
# ls
.  ..  0.0.0190  0.0.0191  0.0.019d  0.0.019e  0.0.0592  0.0.5ca8  0.0.5ca9
# cd 0.0.5ca8/
# ls
availability  ...  detach_state  devtype  discipline  online  readonly  use_diag
# cat online
0
# echo 1 > online
# lsdasd
0.0.5ca9(ECKD) at ( 94:  0) is dasda      : active at blocksize 4096, 601020 blocks, 2347 MB
0.0.5ca8(ECKD) at ( 94:  4) is dasdb      : active at blocksize 4096, 601020 blocks, 2347 MB
```

ON DEMAND BUSINESS™

Linux 2.6 Konfiguration – SCSI Disk



- Alle vom System erkannten FCP Adapter sind im sysfs schon vorhanden (offline)
- Jeder FCP Adapter hat ein eigenes Unterverzeichnis im sysfs.
- Einfaches aktivieren über „online“-Attribut
- Jede Menge Attribute zu Adapter, Port und Unit

```
[root: root]# cd /sys/bus/ccw/drivers/zfcp/
[root: zfcp]# ls
0.0.5588 loglevel_cio … loglevel_scsi version
[root: zfcp]# cd 0.0.5588/
[root: 0.0.5588]# ls
availability card_version … online … port_add … wwpn
[root: 0.0.5588]# cat online
0
[root: 0.0.5588]# echo 1 > online
[root: 0.0.5588]# cat online
1
```

ON DEMAND BUSINESS™

Linux 2.6 Konfiguration – SCSI Disk

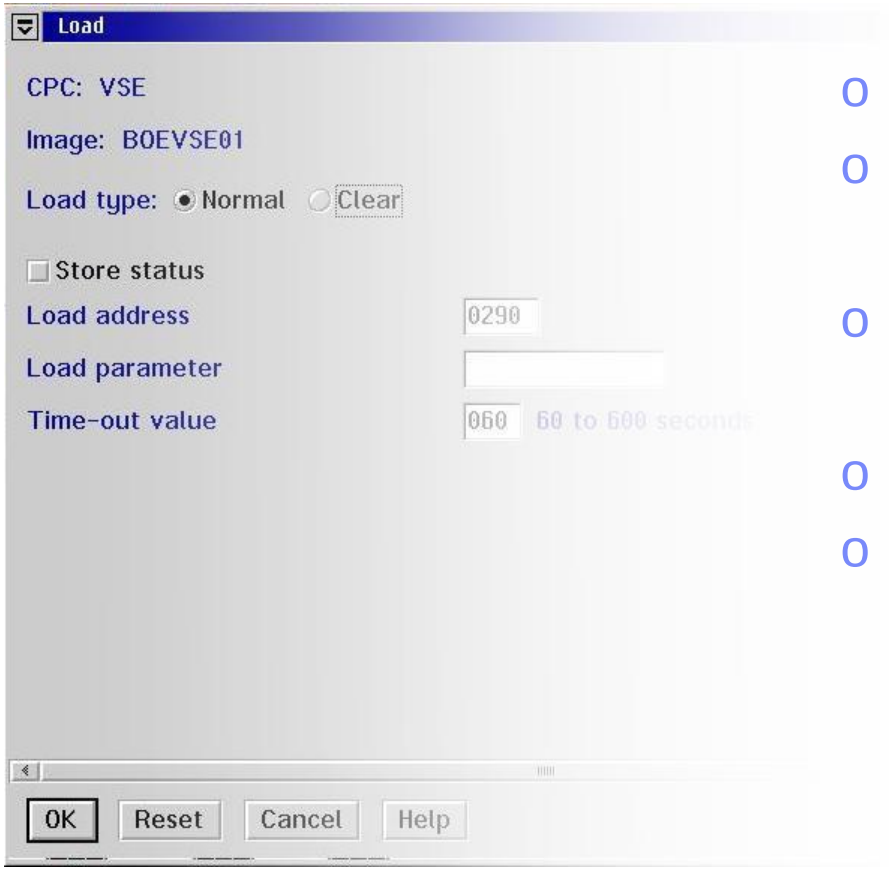
- Konfiguration von Port (WWPN) und Unit (LUN)
- Neues „lsscsi“ Tool als Ersatz für /proc/scsi/scsi

```
[root: 0.0.5588]# ls
availability card_version … online … port_add … wwpn
[root: 0.0.5588]# echo 0x5005076300c693cb > port_add
[root: 0.0.5588]# ls
0x5005076300c693cb availability … wwpn
[root: 0.0.5588]# cd 0x5005076300c693cb
[root: 0x5005076300c693cb]# ls
d_id … unit_add unit_remove wwnn
[root: 0x5005076300c693cb]# echo 0x5125000000000000 > unit_add
[root: 0x5005076300c693cb]# ls
0x5125000000000000  d_id  …  unit_add  unit_remove wwnn
[root: 0x5005076300c693cb]# cd 0x5125000000000000/
[root: 0x5125000000000000]# ls
detach_state  failed  in_recovery  scsi_lun  status
[root: 0x5125000000000000]# lsscsi
[0:0:1:0]  disk   IBM   2105F20    .693  /dev/sda
```

ON DEMAND BUSINESS™

Standard IPL (ESCON/FICON)

- IPL = Initial Program Load
- Laden und starten eines Betriebssystems



- CCW = Channel command word
- Unterstützt sind nur CCW basierte I/O Geräte
- I/O wird von Channel-Programmen gesteuert
- Eine “device number” (2 Byte)
- Konfiguration im IOCDS (I/O configuration data set)

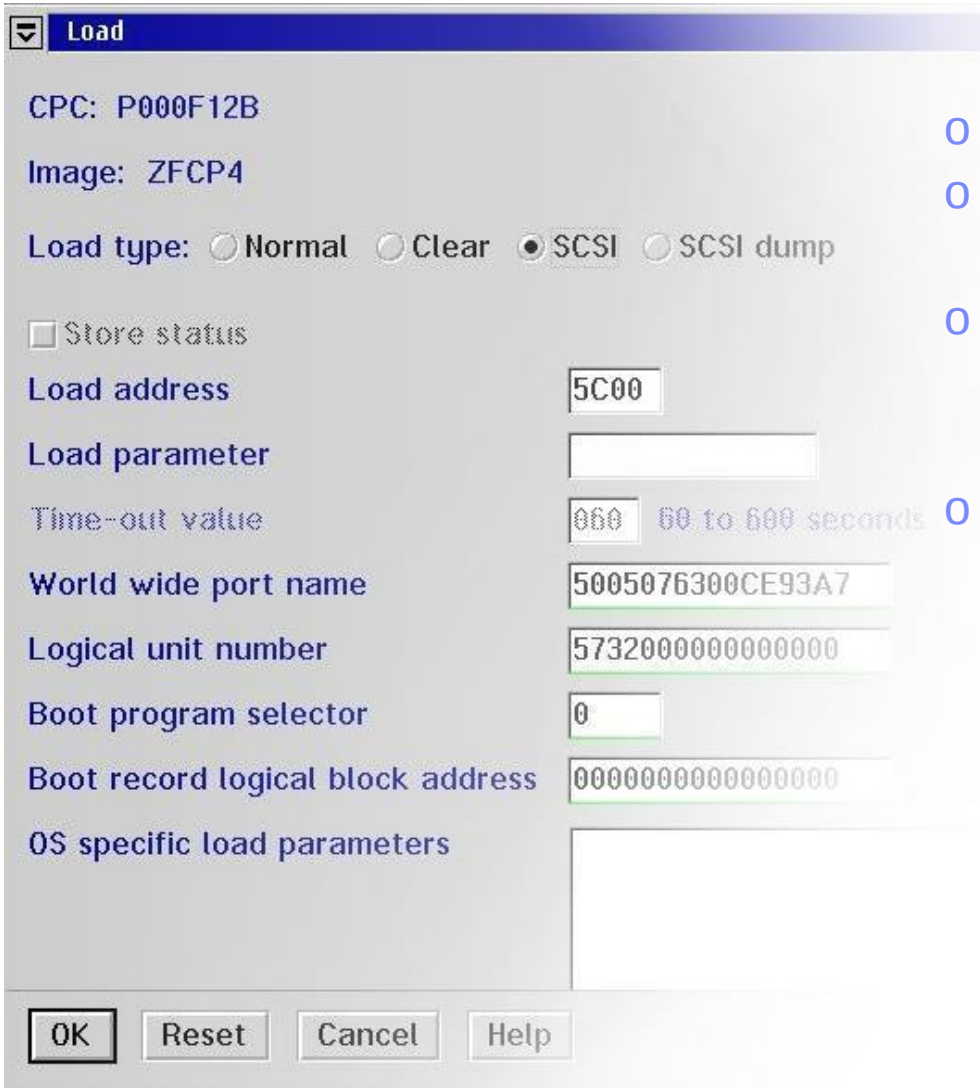
SCSI IPL (FCP)

- SCSI IPL von FCP attached SCSI Disks
  - CCW basiertes IPL funktioniert nicht für FCP Geräte
- Erlaubt Linux Installationen auf SCSI Disks
  - Inklusive IPL und Dump
  - IPL von LPARs und z/VM Gästen unterstützt
- Erweiterung der Palette an zSeries I/O-Geräten
  - SCSI Disks sind nicht mehr nur reine Daten-Geräte
- Standalone LPAR Dump wird unterstützt
  - Dump Programm wird von der SCSI Disk gelesen
  - Dump Daten werden auf die SCSI Disk geschrieben

- **Linux läuft komplett auf SCSI Disks**
  - Unabhängig davon, dass z/VM 4.x noch ECKD DASDs benötigt

ON DEMAND BUSINESS™

SCSI IPL (FCP)



- Neue zusätzliche IPL Parameter
- SCSI Disks werden mit Linux „zipl“ Tool präpariert
- Bis zu 31 Boot-Konfigurationen möglich

- Requirements
  - IBM zSeries Server 800, 890, 900 oder 990
  - Freischaltung mit Feature Code FC9904
  - FCP Channels/Disks
  - z/VM 4.4 (PTF UM30989)

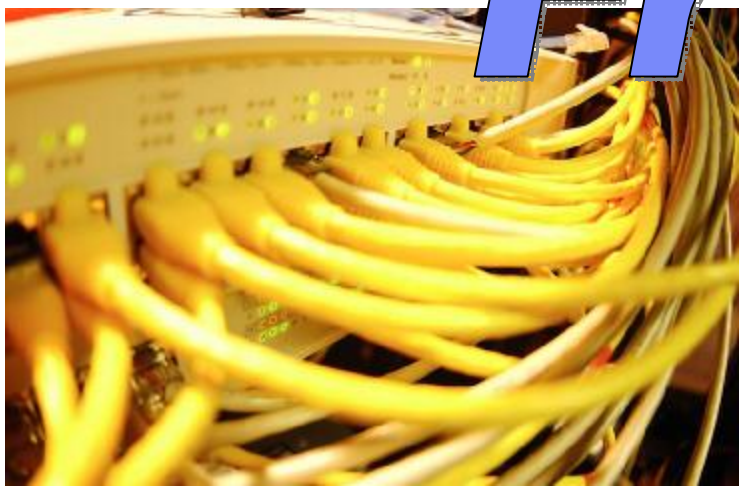
Kurzfassung

- ESCON
  - Traditionelle Verbindungsart
  - Riesiger Fortschritt zur parallelen Verbindung
- FICON
  - Schneller als ESCON
  - Besserer Durchsatz weil: Vollduplex, asynchron, gleichzeitige Übertragung, bedeutend weniger “handshakes” (CU Rückmeldungen)
  - Entfernungen haben kaum Effekt bzgl. Durchsatz (wenig CU Rückmeldungen)
  - Löst das Channel-Limitations-Problem bei großen Installationen
- FCP
  - Schneller und flexibler als FICON
    - Keine ECKD Emulation mehr nötig
    - Konfiguration hauptsächlich im Linux und nicht mehr im IOCDS
  - Ermöglicht Zugriff auf existierende Infrastruktur
    - FICON (Express) Adapter, Kabel und Switches
    - FC Storage Area Network
  - Riesige Bandbreite an FCP und SCSI Speicher-Geräten

Nützliche Links

- I/O Connectivity on IBM zSeries mainframe servers
  - http://www-1.ibm.com/servers/eserver/zseries/connectivity/
- Getting Started with zSeries Fibre Channel Protocol, IBM Redpaper
  - http://www.redbooks.ibm.com/redpapers/pdfs/redp0205.pdf
- z/VM Version 4 Release 4
  - Version 4.4: http://www.vm.ibm.com/zvm440/
  - Version 5.1: http://www.vm.ibm.com/zvm510/
- ABCs of z/OS System Programming Volume 10 (Connectivity Kapitel)
  - http://publib-b.boulder.ibm.com/abstracts/sg246990.html
- Linux auf zSeries and S/390
  - Kernel 2.4: http://oss.software.ibm.com/linux390/june2003_recommended.shtml
  - Kernel 2.6: http://oss.software.ibm.com/linux390/april2004_recommended.shtml
- Linux Device Drivers and Installation Commands
  - Kernel 2.4: http://oss.software.ibm.com/linux390/docu/lx24jun03dd02.pdf
  - Kernel 2.6: http://oss.software.ibm.com/linux390/docu/lx26apr04dd00.pdf
- IBM TotalStorage Tape Device Drivers – Installation and User‘s Guide
  - ftp://ftp.software.ibm.com/storage/devdrvr/Doc/
- ESS Fibre Channel Attachment White Paper
  - http://www.storage.ibm.com/disk/ess/support/essfcwp.pdf

Fragen



Fragen ?