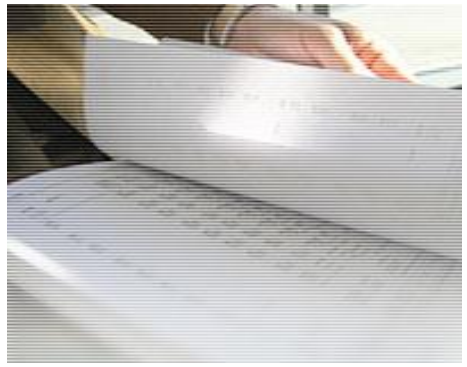




Large Scale Video Server PoC
ein ESS und CES Erfahrungsbericht

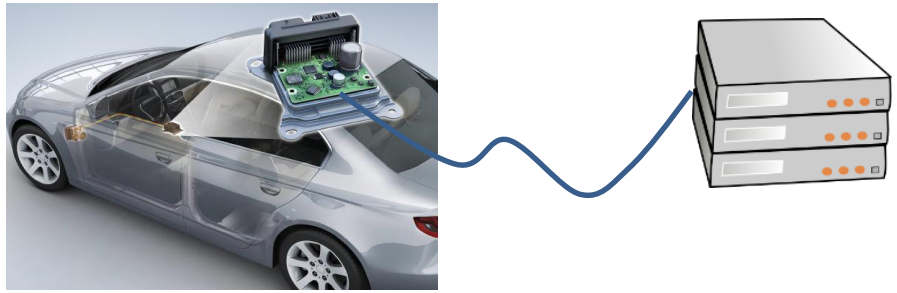
jochen.zeller@sva.de

Agenda

- PoC Hintergrund
- PoC Aufbau - ESS
- PoC Aufbau - CES
- PoC Ergebnis

Das Projektumfeld

- Entwicklungssystem für autonomes Fahren
- Aufzeichnung von Video- und Sensor Daten mit Spezialfahrzeugen, weltweit
- 30 Gigabyte und 5 Minuten pro Sequenz / Videofile
- Video Files werden manuell getagged
- Sogenannte H.I.L. Stationen, Echtzeit Systeme auf Basis von Windows Server mit angeschlossenem Steuergerät „fahren“ die Video Files wieder und wieder ab, und durchlaufen eine Prüfung, ob die neu Entwickelten oder angepasste Software Module korrekt auf das aufgezeichnete Umfeld reagieren.



Das Projektumfeld - Setup

- Entwicklungssystem für autonomes Fahren

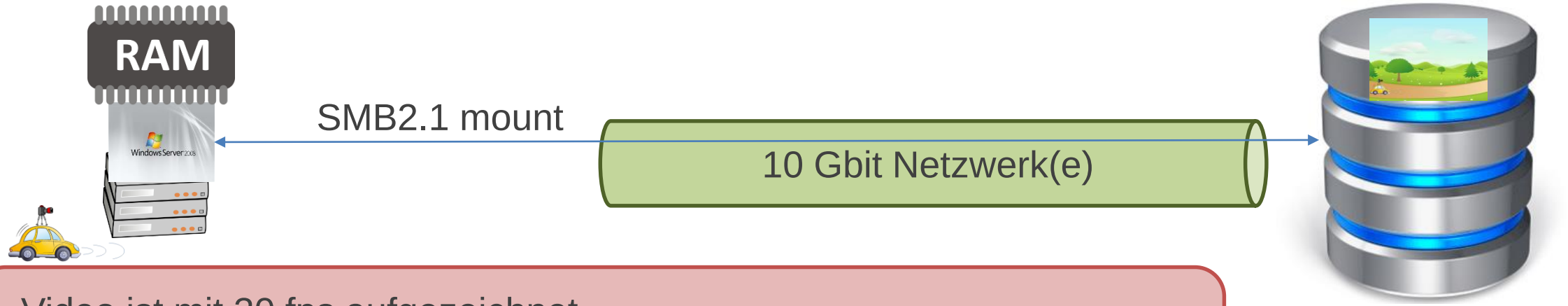
100 H.I.L. Stationen mit angeschlossenen Steuergeräten

10 Gbit Netzwerk(e)

Videospeicher mit tausenden von Video Files

Das Projektumfeld - Ablauf

- Ablauf Video Processing



- Video ist mit 30 fps aufgezeichnet
- Die Toleranz des Steuergerätes endet bei 3 Frames Verlust!
- Deswegen erfolgt das Video Processing aus einer RAM Disk heraus

Agenda

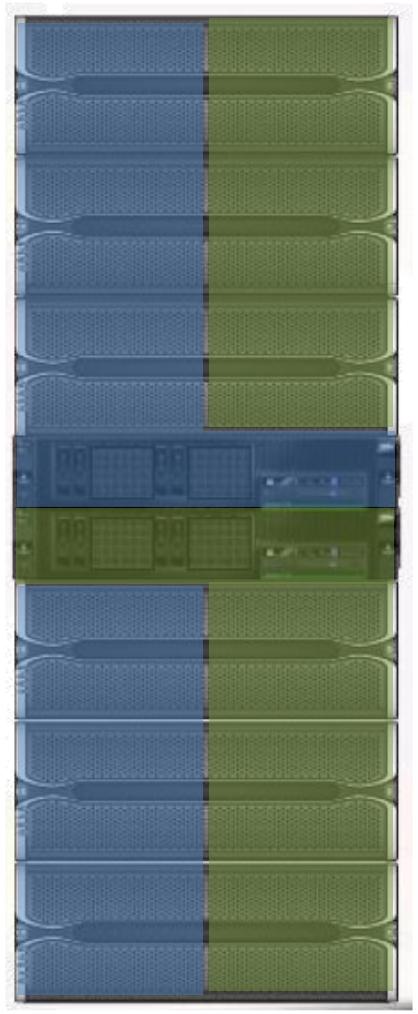
- PoC Hintergrund
- PoC Aufbau - ESS
- PoC Aufbau - CES
- PoC Ergebnis

Der PoC Aufbau

Protocol Nodes 2 x POWER 812L	
Processor	1 x 3,42 GHz 10 Core
Memory	128 GB
Netzwerk ESS	2 x Infiniband FDR
Netzwerk H.I.L.	4 x 10 Gbit Ethernet

IBM ESS GL6	
Netto Kapazität	1.499 TB / 1.363 TiB
Leistung	+15Gbyte/sec
Netzwerk Protocol Nodes	8 x Infiniband FDR
Rackspace	28 HE + 4 HE Mgmt.
Stromaufnahme	max. 9,8 KW

Der PoC Aufbau - ESS Konfiguration



Recovery Groups:

Recovery Group RG1 – 174 x 6TB

Recovery Group RG2 – 174 x 6TB

Declustered Arrays (Default bis 4.0):

DA1 DA2 DA3

DA1 DA2 DA3

Declustered Arrays (Default ab 4.0):

DA1

DA1

```
# mmcrrecoverygroup rg_ess01 -F rg1.cfg
--servers ess01,ess02 --version LATEST
```

```
# cat rg1.txt

...
%da: daName=DA1
  spares=2
  vcdSpares=2
  replaceThreshold=1
  scrubDuration=14
  nspdEnable=no
...
%pdisk: pdiskName=e1d1s01
        device=/dev/sdxr
        da=DA1
        nPathActive=2
        nPathTotal=4
...
```

ESS Konfiguration – RG / pdisk / DA / vdisk

(aktualisiert auf ESS Version 4.0)

Disk Einschub: 58 NL-SAS Drives in 5 Schubladen á 12 Drives

6 Einschübe á 58 Disks in einem GL6 = 348 NL-SAS Drives

Recovery Group:
max. 512 pdisks
1-16 Declustered Arrays
1-64 vdisks

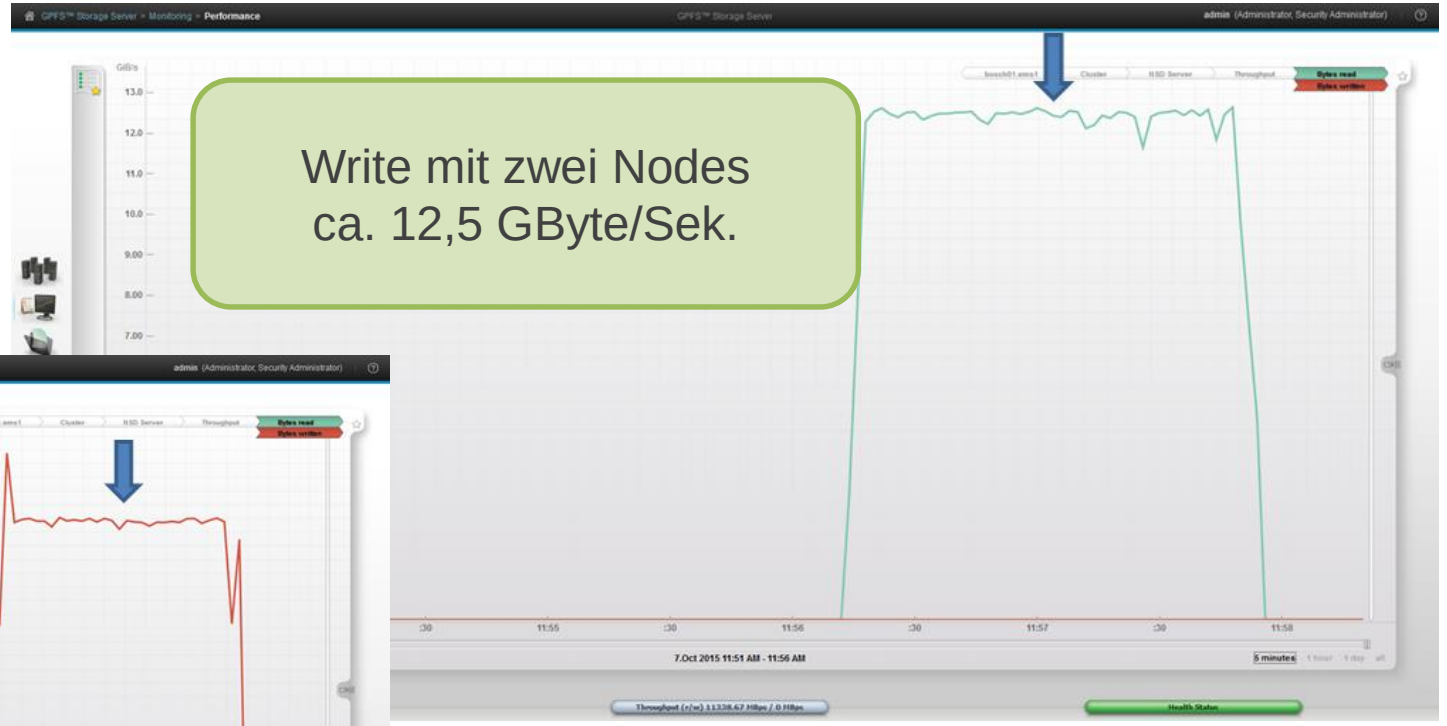
Declustered Array:
max. 512 pdisks
min. pdisk bei 8+2P = 11, besser 12
min. pdisk bei 8+3P = 12, besser 13

8+2P GNR Verteil Algorythmus:
max. 1 Block pro Disk
max. 1 Block pro Drawer
max. 2 Blöcke pro Enclosure

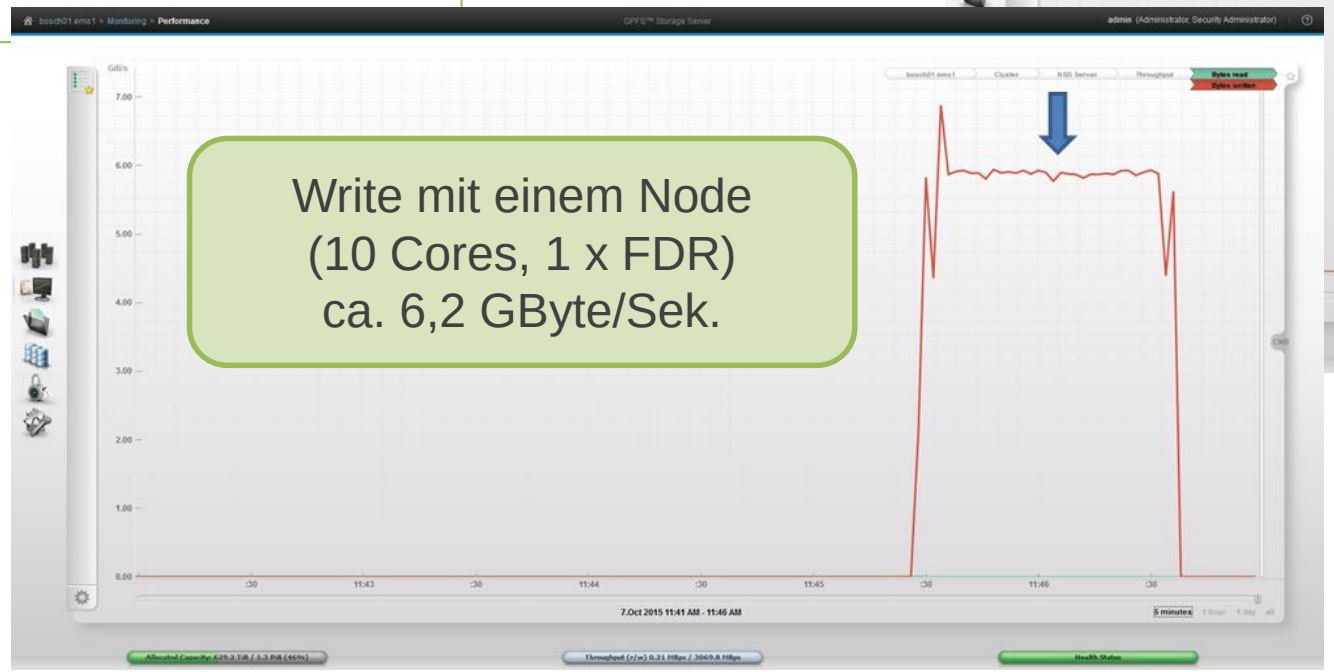
Drawer / Enclosure loss protection
GL6: Enclosure + Drawer loss protection mit 8+2P und 8+3P
GL4: Drawer loss protection mit 8+2P und 8+3P, Enclosure loss protection mit 8+3P

Der PoC Aufbau – ESS und Filesystem Performance

Filesystem Layout
6 DA á 58 pdisks
6 vdisks (8+2P) á 200TB
FS-Blocksize: 8 MB



Write mit zwei Nodes
ca. 12,5 GByte/Sek.



Write mit einem Node
(10 Cores, 1 x FDR)
ca. 6,2 GByte/Sek.

Agenda

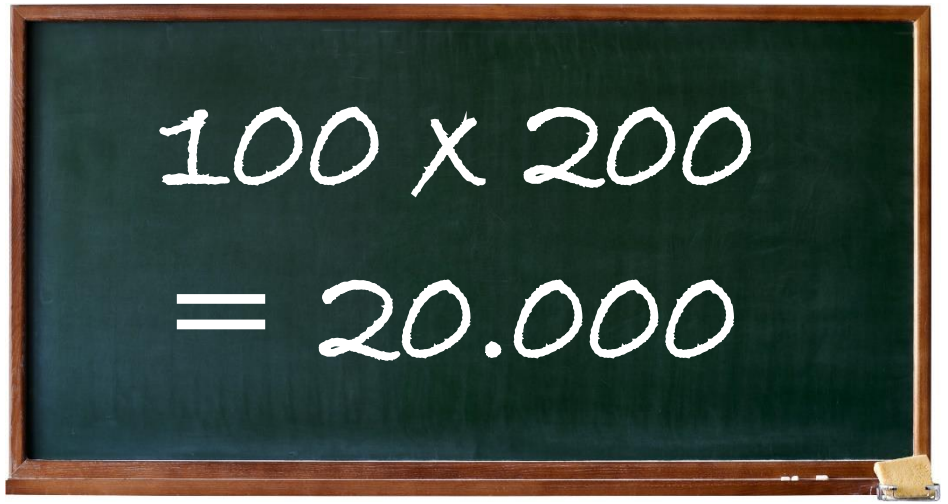
- PoC Hintergrund
- PoC Aufbau - ESS
- PoC Aufbau - CES
- PoC Ergebnis

Der PoC Aufbau - CES

CES – Cluster Export Services – Filer oder Protocol Nodes für Spectrum Scale

Projektanforderung:

- SMB 2.1 (Windows 2008R2) und SMB3.0 (Windows 2012R2)
- NFSv3 und NFSv4
- Gemeinsamer Zugriff auf NFS und SMB Shares → einheitliches ID Mapping
- Leistungsanforderungen:
    - Single Stream SMB 2.1 Leistung: min. 200 MB/Sek



Der PoC Aufbau - CES

CES Installation:

- yum repositories für Ganesha und SMB erstellen

```
[Ganesha-Repository]
name=Ganesha
baseurl=file:///usr/lpp/mmfs/4.1.1.2/ganesha_rpms
enabled=1
gpgcheck=0
```

```
# yum install /usr/lpp/mmfs/4.1.1.2/ganesha_rpms/nfs-ganesha*.rpm
# yum install /usr/lpp/mmfs/4.1.1.2/smb_rpms/gpfs.smb*.rpm
```

CES Konfiguration:

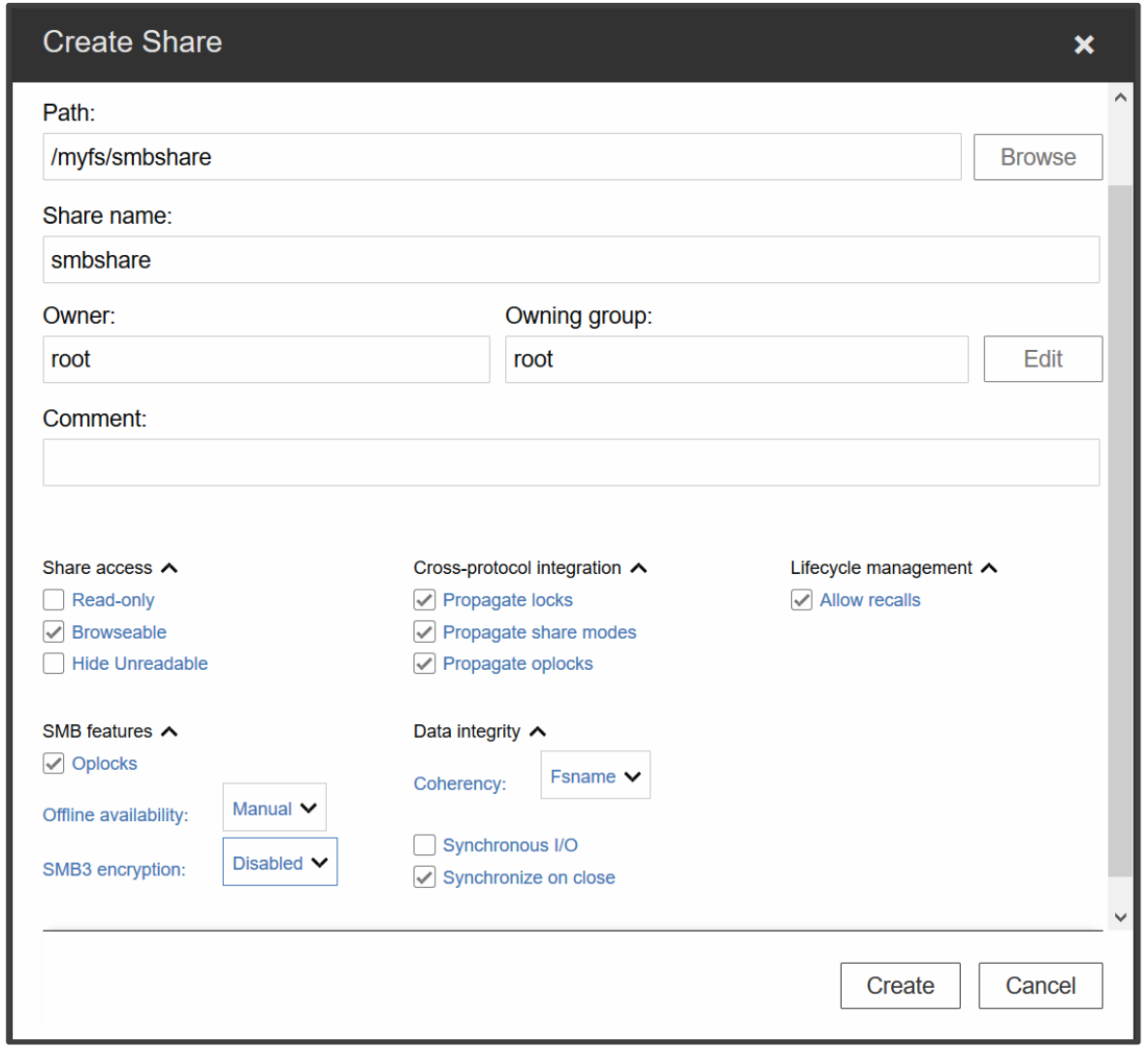
Am besten eigenes Filesystem für cesSharedRoot, weil CES Nodes, sobald aktiv, auf dieses Verzeichnis zugreifen → kein sauberer umount möglich

```
# mmchconfig cesSharedRoot=/ces/.sharedroot
# mmchnode --ces-group=ces -N node01,node02
# mmchnode --ces-enable -N ces
# mmces service enable nfs ; mmces service enable smb
# mmces service start nfs -N ces ; mmces service start smb -N ces
# mmces address add --ces-node node01 --ces-ip 192.168.200.10
# mmces address add --ces-node node02 --ces-ip 192.168.200.20
# mmsmb export add smbshare /shares/smb01
# mmsmb export change smbshare --option "smb encrypt = disabled"
```

SMB Signing (Encryption) ist per Default eingeschaltet, reduziert die Leistung auf etwa 50 MB/Sek.

Der PoC Aufbau - CES

Oder einfacher:



Der PoC Aufbau - CES

CES Authentifizierung

- Einfachste Variante: auto ID Mapping
  - CES speichert eine selbst generierte UID/GID zu jeder SID im CCR

```
# mmuserauth service create --data-access-method file --type ad --servers ad.net --user-name
administrator --password XXXXXXXX --netbios-name videosrv --idmap-role master
```

  - Achtung: Kein einheitliches mapping über Cluster hinaus
  - Achtung: GPFS AD Clients werden nicht über CES gemapped, sondern über GPFS

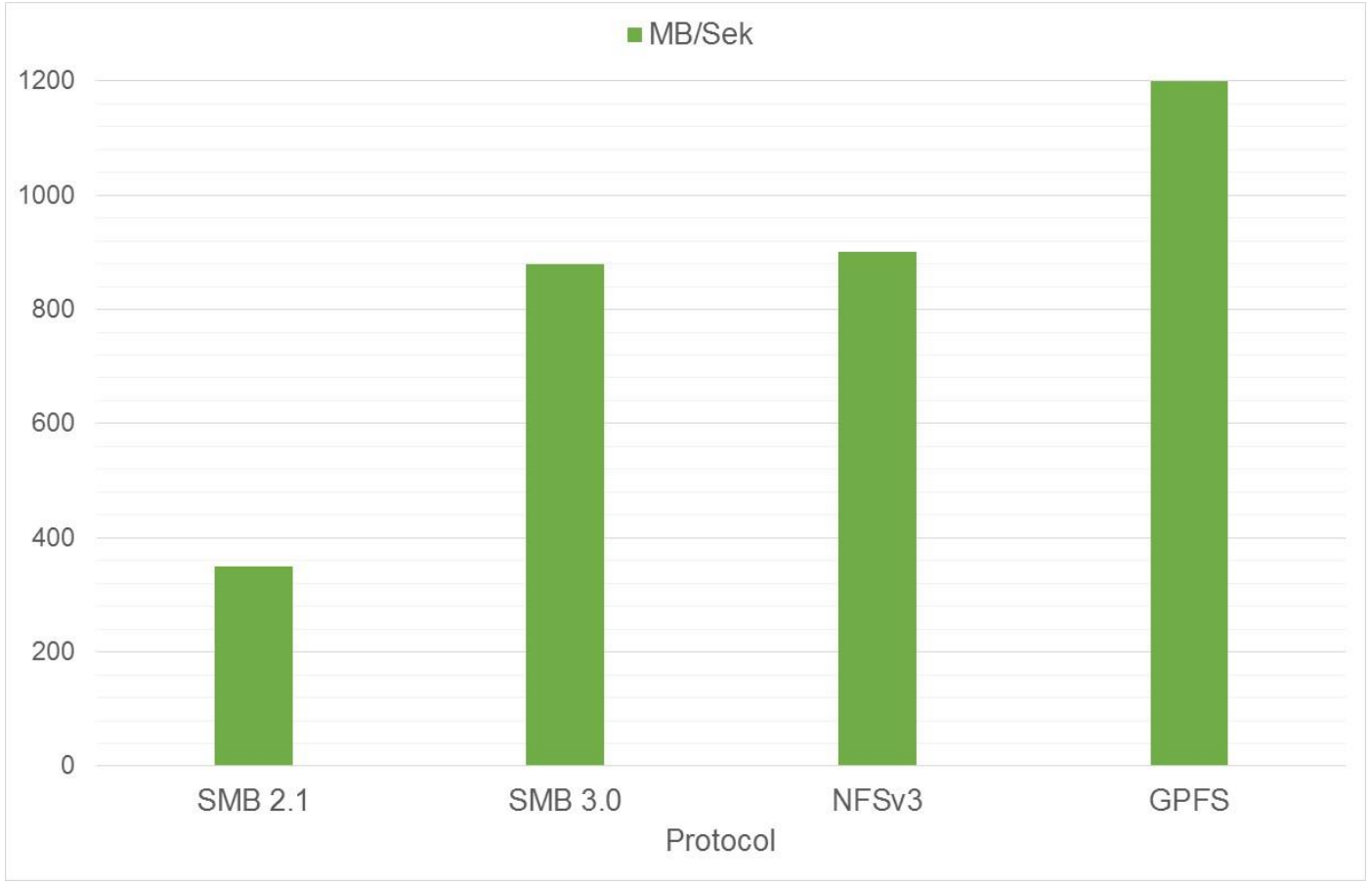
- Besser: UID/GID in AD oder LDAP ID mapping

```
# mmuserauth service create --data-access-method file --type ad --servers ad.net --user-name
administrator --password XXXXXXXX --netbios-name videosrv --idmap-role master
--enable-nfs-kerberos --unixmap-domains "ad.net(1000000-2000000)"
```

Der PoC Aufbau - CES

CES single stream Performance (10Gbit Ethernet):

CES single stream	
SMB 2.1	ca. 350 MB / Sek
SMB 3.0	ca. 880 MB / Sek.
NFSv3	ca. 900 MB / Sek.
Spectrum Scale (Windows 2012R2)	ca. 1.200 MB / Sek.



Agenda

- PoC Hintergrund
- PoC Aufbau - ESS
- PoC Aufbau - CES
- PoC Ergebnis

Das PoC Ergebnis

- Zur Erinnerung: Ablauf Video Processing

Das PoC Ergebnis

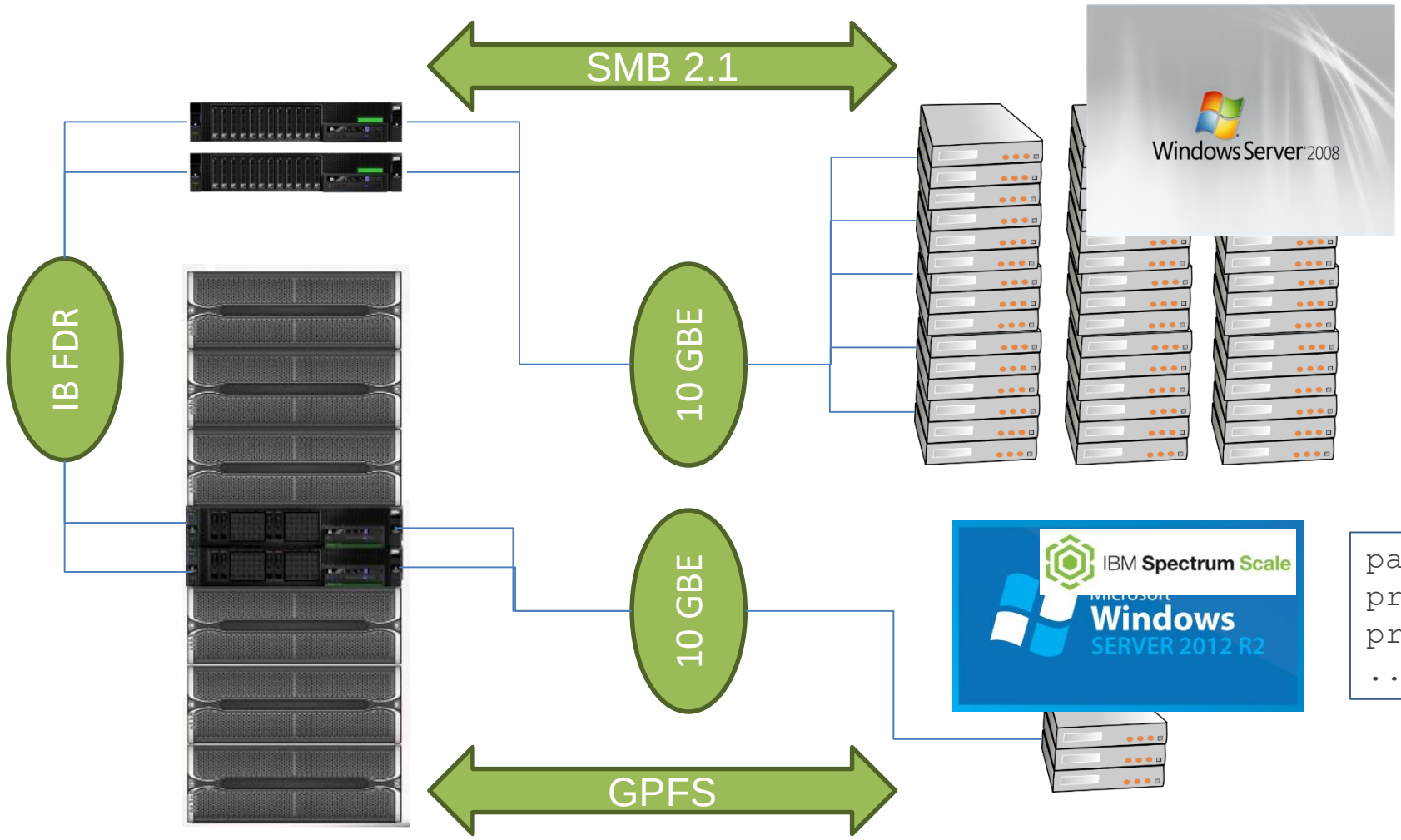
System	Protokoll	Single File Copy	Single File Processing	Single File Gesamt	1000 Files	Differenz zum Bestand in Std.
Bestand	SMB 2.1	3 Minuten	5 Minuten	8 Minuten	8.000 Minuten	0
ESS / CES	SMB 2.1	2 Minuten	5 Minuten	7 Minuten	7.000 Minuten	16 Stunden
ESS / CES	SMB 3.0	1 Minuten	5 Minuten	6 Minuten	6.000 Minuten	33 Stunden

Das PoC Ergebnis

- Ablauf Video Processing – ohne Zwischenkopie?

- Video ist mit 30 fps aufgezeichnet
- Die Toleranz des Steuergerätes endet bei 3 Frames Verlust!

Das PoC Ergebnis - Ass im Ärmel




```
pagepool = 24G
prefetchPct = 60
prefetchAggressiveness = 0
...
```

Das PoC Ergebnis

System	Protokoll	Single File Copy	Single File Processing	Single File Gesamt	1000 Files	Differenz zum Bestand in Std.
Bestand	SMB 2.1	3 Minuten	5 Minuten	8 Minuten	8.000 Minuten	0
ESS / CES	SMB 2.1	2,5 Minuten	5 Minuten	7,5 Minuten	7.500 Minuten	8 Stunden
ESS / CES	SMB 3.0	1 Minuten	5 Minuten	6 Minuten	6.000 Minuten	33 Stunden
ESS / CES	GPFS	-	5 Minuten	5 Minuten	5.000 Minuten	50 Stunden

Das PoC Ergebnis - Fazit

Inbetriebnahme:	- Funktionalität in 2 Tagen hergestellt
Leistung:	- Leistungssteigerung in Simulationen pro Stunde für die bestehende H.I.L. Infrastruktur mit Win2008R2 (SMB 2.1)
	- Nächste Generation H.I.L. kommt ohne Copy Prozess aus, Video Files können via GPFS direkt gestreamt werden
Bedienbarkeit:	- ESS und CES Dank GUI einfach zu bedienen

Jochen Zeller
System Architekt

SVA System Vertrieb Alexander GmbH
Borsigstraße 14
65205 Wiesbaden-Nordenstadt

Mobil: +49 (0)151 / 180 256 77

E-Mail: jochen.zeller@sva.de
http://www.sva.de