Marius.bloggt-in-braunschweig.de

präsentiert

Der (Super)-Gau:

Wenn die Hardware versagt, und das Filesystem mit in den Abgrund zieht.

Vorwort:

"Backups sind ein wichtiger Bestandteil eines funktionierenden Computersystems. Jeder, der ohne Backups arbeitet, hat seine Daten jetzt schon verloren."

"Ein zweiter Computer im Haushalt, kann nie schaden."

Die unscheinbaren Anfänge

Es war ein lauer Sonntagmittag, der Stream des Chaosradios ergoß sich von Radio Fritz, der Frachter in EVE flog seine Route, als plötzlich ...

"can't create tempfile in /tmp/gluster-x34234.tmp"

(Super)-Gau: Hardwareversagen

. GW224

16.5312201 ata4.00: status: { DRDY] 16.5396141 ata4.00: cmd 61/04:c0:ad:?c:10/00:00:00:00/40 tag 25 ncg dma 2048 out 16.5354231 ata4.00: failed command: WRITE FPDMA QUEUED res 40/00:a0:9f:7c:10/00:00:00:00:00/40 Emask 0x10 (ATA bus error) 16.548107] ata4.00: status: { DRDY } 16.552301] ata4.00: failed command: WRITE FPDMA QUEUED 16.5564901 ata4.00: cmd 61/02:d0:b1:7c:10/00:00:00:00:00/40 tag 26 ncq dma 1024 out res 40/00:a0:9f:7c:10/00:00:00:00:00/40 Emask 0x10 (ATA bus error) 16.5564901 16.564950] ata4.00: status: { DRDY } 16.569171] ata4.00: failed command: WRITE FPDMA QUEUED 6.573353] ata4.00: cmd 61/02:d8:b3:7c:10/00:00:00:00:00/40 tag 27 ncg dma 1024 out res 40/00:a0:9f:7c:10/00:00:00:00:00/40 Emask 0x10 (ATA bus error) 6.5733531 6.581810] ata4.00: status: { DRDY } 5.5860301 ata4.00: failed command: WRITE FPDMA QUEUED 5.590212] ata4.00: cmd 61/02:e0:b5:7c:10/00:00:00:00:00/40 tag 28 ncg dma 1024 out res 40/00:a0:9f:?c:10/00:00:00:00:00/40 Emask 0x10 (ATA bus error) 5.5902121 .598670] ata4.00: status: { DRDY } .602888] ata4.00: failed command: WRITE FPDMA QUEUED .607073] ata4.00: cmd 61/01:e8:b7:7c:10/00:00:00:00:00/40 tag 29 ncg dma 512 out res 40/00:a0:9f:7c:10/00:00:00:00:00/40 Emask 0x10 (ATA bus error) .6070731 615535] ata4.00: status: { DRDY } 619760] ata4.00: failed command: WRITE FPDMA QUEUED 623947] ata4.00: cmd 61/02:f0:b8:7c:10/00:00:00:00:00/40 tag 30 ncg dma 1024 out res 40/00:a0:9f:7c:10/00:00:00:00:00/40 Emask 0x10 (ATA bus error) 6239471 532418] ata4.00: status: { DRDY } A start job is running for File System Check on /dev/mapper/luks-ffbd61f2-4c1e-4dc8-b12b-c89e t FIS failed) A start job is running for File System Check on /dev/mapper/luks-ffbd61f2-4c1e-4dc8-b12b-c89 t FIS failed) A start job is running for File System Check on /dev/mapper/luks-ffbd61f2-4c1e-4dc8-b12b-c89 : FIS failed) A start job is running for File System Check on /dev/mapper/luks-ffbd61f2-4c1e-4dc8-b12b-c89



	1	82.9544941 sd	3:0:0:0: rejecting	1/0 to offline d	evice	
	I.	82.9545491 sd	3:0:0:0: rejecting	1/0 to offline d	evice	
	i i	82 9545641 sd	3:0:0:0: rejecting	Lol to offline d	euice	
1	i i	82 9545961 ed	3:0:0:0: no jecting	I d to offline d		
		02.3313001 80	3.0.0.0. rejecting		EVICE	
		02.7545771 Sd	3:0:0:0: rejecting	1/U to offline a	evice	
		82.9546101 sd	3:0:0:0: rejecting	1/0 to offline d	evice	
		82.9546201 sd	3:0:0:0: rejecting	I/O to offline do	evice	
	L.	82.9546311 sd	3:0:0:0: rejecting	I/O to offline de	evice	
	L.	82.9546441 sd	3:0:0:0: rejecting	I/O to offline de	evice	
	ι,	82.9546551 sd	3:0:0:0: rejecting	I/O to offline de	vice	
	۲.	82.9546671 sd	3:0:0:0: rejecting	1/0 to offline de	vice	
	L.	82.9546781 sd	3:0:0:0: rejecting	1/0 to offline de	vice	
	I.	82.9547571 sd	3:0:0:0: rejecting	1/0 to offline de	vice	
	٦ I	82.9547641 sd	3:0:0:0: rejecting	1/0 to offline de	vice	
	I.	82.9547761 sd	3:0:0:0: rejecting	1/0 to offline de	uice	
	1	82.9547861 sd	3:0:0:0: rejecting	1/0 to offline de	ulce	
	I.	82.9574211 sd	3:8:8:8: rejecting	1/0 to offline de		
	L.	82.9578901 sd	3:0:0:0: rejecting	1/0 to offline de		
	L.	82.9579051 sd	3:0:0:0: rejecting	1/0 to offline de		
	L.	82.9579141 sd	3:0:0:8: rejecting	1/0 to offline de	VICE	
	I.	82.9579231 sd	3:0:0:8: rejecting	1/0 to offline de	VICE	
	I.	82.9579311 sd	3:0:0:8: rejecting	1/0 to offline de		
	1	82.9579391 sd	3:8:8:8: rejecting	1/0 to offline de		
	ſ	82.9579481 sd	1 3:8:8:8: rejecting	1/0 to offline de		
	I	82.9579561 sd	1 3:0:0:0: rejecting	1/0 to offline de		
	1	82.9579641 sd	1 3:0:0:0: rejecting	1/0 to offline de	vice	
	I	82.9581391 sd	1 3:0:0:0: rejecting	1/0 to offline de	utce	
	L I	82.9581521 sd	1 3:0:0:0: rejecting	1/0 to offline de	uice	
	L L	82.9581601 sd	1 3:0:0:0: rejecting	1/0 to offline de	tice	
	L.	82.9581681 sd	d 3:0:0:0: rejecting	1/0 to offline des	lice	
	1	82.9581761 st	d 3:0:0:0: rejecting	1/0 to offline dev	lice	
		82.9581841 se	d 3:0:0:0: rejecting	1/0 to offline dev	lice	
3		82.9581921 sc	d 3:0:0:0: rejecting	1/0 to offline dev	vice	
3/		02.958200J SC	d 3:0:0:0: rejecting	1/0 to offline dev	vice	
B		02.730210J SC	d 3:0:0:0: rejecting	1/0 to offline dev	vice	
er Blue		82 9592261 a	a 3:0:0:0: rejecting	1/0 to offline dev	lice	
1	ī	82.9582341	a 3.0.0.0: rejecting	1/0 to offline dev	lce	
V.	i i	82.9582421 st	d 3:0:0:0: rejecting	1/0 to offline dev	lce	
	C	82.9582501 st	d 3:0:0:0: rejecting	1/0 to offline dev	ice	
	I I	82.9582581 st	d 3:0:0:0: rejecting	1/0 to orrline dev	ice	
	, I	82.9582651 st	d 3:0:0:0: rejecting	10 to offline dev	ice	
	L L	OK 1 Stopped	/dev/disk/bu-uuid/ef	redbbe-atae 4242 b	ice	
		UK 1 Stopped	/dev/disk/by-id/dm-u	id-CRYPT-LINS1_CC	Ca5-1249d6r2dr67.	
		J Stopped	/dev/disk/by-id/dm-n	me-luks-ffbd6162-	Acte-44c9 1421 - 00	69c9fa-luks-ffbd61f2-4c1e-4
		J Stopped	/dev/dm-1.	11 000116-	1010-100-0120-089e9f69c	9fa.
		I Stopped	/sys/devices/virtual.	block/dm-1.		
	STREET, STREET	a ocoppea	target Encounted Hat			

Die schwierigen Folgen

Beim Rebooten startete das System nicht mehr mit dem Hinweis, daß die Superblöcke der Partionen/Filesysteme beschädigt wären.

/var/log/messages-20170402:Apr 2 15:05:32 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:08:10 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:08:30 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:08:30 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:08:34 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:08:44 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:08:47 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:08:49 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:09:46 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:09:58 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:09:58 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:09:58 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:09:58 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:21:39 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:21:39 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!
/var/log/messages-20170402:Apr 2 15:25:56 eve kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!

Ein Mounten der Bootpartition war nicht mehr möglich, weil der Superblock defekt war.

(Super)-Gau: Aufbau einer Festplatte

Modell Samsung SSD 840 EVO 120GB (EXTOCB6Q) Größe 120 GB (120.034.123.776 Bytes)

- Partitionierung Partitionssektor
- Seriennummer S1D5NSAFB00974P

Einschätzung Das Laufwerk ist in Ordnung (32° C / 90° F)

Datenträger



Partitionierung einer Festplatte



Datenträger



LUKS Festplattenverschlüsselung



Wie behebt man ein defektes Filesystem ?

Zunächst einmal, solange das System noch hochfährt, und das defekte System nicht ROOT ist, kann man es einfach aushängen und mit *fsck /dev/sdxX* prüfen lassen.

Wenn das System nicht mehr bootet, benutzt man am besten einen USB-Stick und bootet den Rechner damit.

Der Stick ist auch nötig, wenn man LUKS aktiviert hat, weil man ansonsten nicht an die Partition kommt.

"Das Filesystem meldet Unsinn, weil der Superblock beschädigt ist"

kernel: EXT4-fs (sdc1): filesystem has both journal and inode journals!

Wer den Fehler sieht, hat verloren! Eine Reparatur des Filesystems ist ohne Verluste unwahrscheinlich.

Mounten der Partition ohne CHECKs

mount -t ext4 -o check=none /dev/sdc1 /mnt

Das kann klappen und dann heißt es Datensicherung betreiben, also alles auf einen Stick ziehen, was noch zu retten ist.

Mounten der Partition mit einem anderen Superblock

mount -t ext4 -o sb=131072 /dev/sdc1 /mnt

Das kann klappen und dann heißt es Datensicherung betreiben, also alles auf einen Stick ziehen, was noch zu retten ist.

Wie bekommt man raus, wo der Reserve-Superblock ist?

dumpe2fs /dev/sdc1 | grep Super dumpe2fs 1.43.3 (04-Sep-2016) Primär Superblock in 0, Gruppendeskriptoren in 1-1 Datensicherung Superblock in 32768, Gruppendeskriptoren in 32769-32769 Datensicherung Superblock in 98304, Gruppendeskriptoren in 98305-98305

Superblock Nummer != Superblock Position

Datensicherung Superblock in **32768**, Gruppendeskriptoren in 32769-32769

Aber: mount -t ext4 -o sb=**131072** /dev/sdc1 /mnt

Weil: Ein Block 4KB lang ist, ist der Superblock auf der Partition bei 32768*4 = 131072.

WARNUNG: Wer andere Blockgrößen verwendet, muß das berücksichtigen!

Wie bekommt man raus, wo der Reserve-Superblock ist?

Wenn dumpe2fs nicht zur Verfügung steht und die Partition um die es geht, nicht gemountet ist, kann man durch eine Simulation eines neuen Filesystems den Superblock bestimmen:

mke2fs -n /dev/sdxX

Wobei "-n" verhindert, daß tatsächlich was geschrieben wird.

Wie man das Filesystem mit dem Reserve-Superblock repariert:

fsck **-s 131072** /dev/sdc1

Man kann den Reserveblock als Quelle für den Filesystemcheck benutzen. Und wenn man den auch verhunzt hat, kann man den nächsten Superblock benutzen.

Deswegen verteilt das Filesystem so viele Kopien vom Superblock auf der Partition, in der Hoffnung, daß nicht alle beschädigt wurden.

Wer Glück hatte, ist hier fertig.

Wenn keine Dateien gelöscht wurden, kann man versuchen das System zu booten.

Aber, mit dem Fehler: "filesystem has both journal and inode journals!"

ist das aber fast ausgeschlossen, weil sich das nicht durch einen FSCK beheben läßt. Das müßte man mit *debugfs* das Filesystem bearbeiten, was außerhalb dies Vortrags ist.

Deswegen ...

Der letzte Ausweg

... ist oftmals nur eine neue Partition bzw. ein Reformatieren der Partition der letzte Ausweg.

Da das Formatieren die UUIDs einer Partition ändert, stimmen die Daten auf der Systemplatte nicht mehr.

Diese müssen erst angeglichen werden.

Wie man eine neue Partition anlegt

Eine leere Partition bekommt man mit den mkfs Befehl. Diese läßt sich auch wieder normal mounten. Was vorher auf der Platte war, ist weg, deswegen Backup machen, solange es noch geht (also JETZT ;))

mkfs.ext4 /dev/sdc1

Für diesen Vortrag war es die Bootpartition mit den Kernels drauf, ohne Kernels **bootet** das System jetzt todsicher **nicht mehr**.

"Es bootet nicht mehr…!"

Don't Panic!

Wie man eine neue Bootpartition anlegt

Alles halb so schlimm, wir bauen einfach eine neue Bootpartition auf.

Per Kopie von einem Laptop
 von einer LIVEDISK

Wie man eine neue Bootpartition anlegt Per Kopie von einem Laptop

Der Weg ist einfach:

USB Stick ins Laptop
Bootpartition zusammenpacken
USB Stick in den Rechner ohne Bootkernels
Bootpartitionsinhalt auf die leere Partition kippen
UUIDs & BLKIDs anpassen
Bootroot anpassen (Falls nötig)
/etc/fstab anpassen

Fertig.

Wie man eine neue Bootpartition anlegt

Probleme:

Die Kernels auf dem Laptop waren andere als auf dem Desktoprechner.

Die Festplattenanordnung war anders

Die Partitionsanordnung war anders

Diese Fehler kann man alle in Verlauf der Reparatur anpassen.

(Super)-Gau: Aufbau einer Festplatte

Modell Samsung SSD 840 EVO 120GB (EXTOCB6Q) Größe 120 GB (120.034.123.776 Bytes)

- Partitionierung Partitionssektor
- Seriennummer S1D5NSAFB00974P

Einschätzung Das Laufwerk ist in Ordnung (32° C / 90° F)

Datenträger



Partitionierung einer Festplatte



Datenträger



LUKS Festplattenverschlüsselung



von BLKIDs und UUIDs

\$ lsblk						
NAME	MAJ:MIN	RM	SIZE	RO	TYPE	MOUNTPOINT
sr0	11:0	1	1024M	0	rom	
sdc	8:32	0	111,8G	0	disk	
⊢sdc2	8:34	Θ	103,7G	0	part	
└─luks-ffbd61f2-4c1e-4dc8-b12b-c89e9f69c9fa	253:0	0	103,7G	0	crypt	/
—sdc3	8:35	Θ	7,6G	0	part	
└─luks-ca1f019d-39ce-4cf8-b522-f6d3e63ebe2a	253:1	0	7,6G	0	crypt	[SWAP]
L_sdc1	8:33	Θ	525,5M	0	part	/boot
sda	8:0	0	1,8T	Θ	disk	
⊢sda4	8:4	Θ	1K	Θ	part	
—sda2	8:2	Θ	78,1G	0	part	
└─luks-6d45b281-c56c-40f0-acf7-c3058d4d6913	253:3	Θ	78,1G	Θ	crypt	
—sda5	8:5	Θ	1,8T	0	part	
└─luks-7881f602-9462-497d-810a-7d6111ad6085	253:2	0	1,8T	0	crypt	/sata_home
—sda3	8:3	Θ	7,8G	Θ	part	
└─luks-384d6b27-6263-4d53-bfce-e7e5bcd221b9 └─sda1	253:4	0	7,8G	0	crypt	- Ma

von BLKIDs und UUIDs

blkid

/dev/sda1: UUID="aee1b027-ebd7-4ad9-a0ea-0fc881193708" TYPE="ext4" PARTUUID="000c4469-01"

/dev/sda2: UUID="6d45b281-c56c-40f0-acf7-c3058d4d6913" TYPE="crypto_LUKS" PARTUUID="000c4469-02"
/dev/sda3: UUID="384d6b27-6263-4d53-bfce-e7e5bcd221b9" TYPE="crypto_LUKS" PARTUUID="000c4469-03"
/dev/sda5: UUID="7881f602-9462-497d-810a-7d6111ad6085" TYPE="crypto_LUKS" PARTUUID="000c4469-05"
/dev/sdc1: LABEL="Boot" UUID="221608f2-5914-4619-9ef7-6dfddf233fd4" TYPE="ext4"

PARTUUID="0000a3dd-01"

/dev/sdc2: UUID="ffbd61f2-4c1e-4dc8-b12b-c89e9f69c9fa" TYPE="crypto_LUKS" PARTUUID="0000a3dd-02"
/dev/sdc3: UUID="ca1f019d-39ce-4cf8-b522-f6d3e63ebe2a" TYPE="crypto_LUKS" PARTUUID="0000a3dd-03"
/dev/mapper/luks-ffbd61f2-4c1e-4dc8-b12b-c89e9f69c9fa: UUID="e62edbbe-a1ae-4242-bca5-1249d6f2df67"
TYPE="ext4"

/dev/mapper/luks-ca1f019d-39ce-4cf8-b522-f6d3e63ebe2a: UUID="46da0d80-21fb-45b7-8567-ba047de66cb6" TYPE="swap"

/dev/mapper/luks-7881f602-9462-497d-810a-7d6111ad6085: UUID="196a4455-7ccb-40e4-bc71-7b2929f29225" TYPE="ext4"

/dev/mapper/luks-6d45b281-c56c-40f0-acf7-c3058d4d6913: UUID="0fd1b33c-5a2e-421a-a872-7bfc784ea2cf" TYPE="ext4"

/dev/mapper/luks-384d6b27-6263-4d53-bfce-e7e5bcd221b9: UUID="0bb5a502-f854-4e40-a3bd-08c3a2e6bf22" TYPE="swap"

Der GRUB Bootentry

menuentry 'Fedora (4.10.6-200.fc25.x86_64) 25 (Twenty Five)' --class fedora --class gnu-linux --class gnu --class os --unrestricted \$menuentry_id_option 'gnulinux-4.10.6-200.fc25.x86_64advanced-e62edbbe-a1ae-4242-bca5-1249d6f2df67' { load video set gfxpayload=keep insmod gzio insmod part_msdos insmod ext2 set root='hd2,msdos1' if [x\$feature_platform_search_hint = xy]; then search --no-floppy --fs-uuid --set=root --hint-bios=hd2,msdos1 --hint-efi=hd2,msdos1 --hint-baremetal=ahci2, msdos1 **221608f2-5914-4619-9ef7-6dfddf233fd4** else search --no-floppy --fs-uuid --set=root 221608f2-5914-4619-9ef7-6dfddf233fd4 fi **linux** /vmlinuz-4.10.6-200.fc25.x86 64 **root=UUID=e62edbbe-a1ae-4242-bca5-1249d6f2df67** ro vconsole.font=latarcyrheb-sun16 rd.luks.uuid=luks-ca1f019d-39ce-4cf8-b522-f6d3e63ebe2a rd.luks.uuid=luks-ffbd61f2-4c1e-4dc8-b12b-c89e9f69c9fa rhqb quiet splash audit=0 rd.driver.blacklist=nouveau nouveau.modeset=0 **initrd** /initramfs-4.10.6-200.fc25.x86 64.img

/etc/fstab

\$ cat /etc/fstab							
#							
# /etc/fstab							
# Created by anaconda on Tue Jul 1 16:05:13 2014							
#							
# Accessible filesystems, by reference, are maintained under '/dev/disk'							
# See man pages fstab(5), findfs(8), mount(8) and/or blkid(8) for more info							
+							
/dev/mapper/luks-ffbd61f2-4c1e-4dc8-b12b-c89e9f69c9fa /	ext4	defaults,x-					
systemd.device-timeout=0 1 1							
UUID=221608f2-5914-4619-9ef7-6dfddf233fd4 /boot	ext4	defaults	12				
/dev/mapper/luks-7881f602-9462-497d-810a-7d6111ad6085 /sata_home	ext4	defaults,x-					
systemd.device-timeout=0 1 2							
/dev/mapper/luks-calf019d-39ce-4cf8-b522-f6d3e63ebe2a swap	swap	defau⊥ts,x-					
systemd.device-timeout=0 0 0							

Systempartition mounten
DEV und PROC in die Systempartition mounten BOOT in die Systemplatte mounten
CHROOT in die Systemplatte machen
Bisherige Kernel finden
Kernel neu installieren Kernel mit der Livedisk installieren Systempartition mounten:

1. normal ohne LUKS

mount /dev/sdc2 /media

2. mit LUKS vorher die LUKS Partition entschlüsseln

mount /dev/mapper-luks-uuid... /media

DEV + PROC + BOOT mounten:

mount -t devtempfs devtempfs /media/dev

mount -t procfs procfs /media/proc

mount -t ext4 /dev/sdc1 /media/boot



Ins System wecheln mit chroot

chroot /media/

Jetzt kann man mit seinem alten System weiterarbeiten und hat alle Tools zur Verfügung, die installiert sind.

Bisherige Kernel finden: rpm -qa | grep kernel-core Kernel neu installieren : dnf reinstall kernel-core-*

GRUB Bootloader fixen

Es müssen noch der Bootloader neu geschrieben werden *grub2-install /dev/sdc*

Und ggf. ist es ratsam einmal die Grubkonfig neu erzeugen zu lassen, falls das nicht beim Reinstallieren der Kernel passiert ist:

grub2-mkconfig -o /boot/grub2/grub.cfg

GRUB Bootloader fixen

TIP: wer jüngst ein OS Upgrade hatte, dem seine Grubkonfig könnte defekt sein. In ...

/etc/grub.d/10_linux

... "sixteenbit="16"" mit "sixteenbit=""" ersetzen und dann die Grub Konfig neu bauen lassen.

Glückwunsch!

Sie haben es überstanden!