

ZFS Optimierung und Todsünden

ZFS Funktionen für SSDs

- `zpool get autotrim <poolname>`: Auslesen ob „autotrim“ eingeschaltet ist
- `zpool trim <poolname>`: Pool manuell trimmen
- `zpool status -t <poolname>`: Auslesen TRIM status
- `zpool set autotrim=on <poolname>`: generell trimmen (autotrim einschalten, wenn manuelles trimmen erfolgreich)
- `zfs add rpool cache nvme0e1`: füge nvme0e1 als cache Device in rpool

ZFS mit Spindeln

- Je mehr Spindeln desto schneller
- Gespiegelte Disks ggf. auf mehrere SAS Lanes verteilen (dmesg beobachten)
- SAS ist schneller als SATA!
- Maximaler Speed ist oft durch alte HBA limitiert (6/12G)
 - G ist Gigabit, also $6/8 = 750\text{MB/sek}$ max

Todsünden

- HW-Raid Controller nutzen und dort jede Disk als Volume für ZFS SW Raid (knallt)
- Deduplizierung aktivieren (geht nicht mehr aus pro Pool)
- ZFS_ARC_MAX zu hoch/niedrig (ca 1GB pro TB Richtlinie)
- Pool über 80% Füllgrad laufen lassen
- Snapshotbereinigung / Management aus dem Blick verlieren (Empfehlung: 8×1/4h, 48h, 10T, 4W, 3M)
- Bei ZFS-Auto-Snapshot auf Ziel nicht deaktiviert (`zfs set com.sun:auto-snapshot=false rpool/replica`)
- Anfügesnapshot durch schlechtes Management verlieren (Beispiel: nur houlry replizieren, nach 49h kein Anfügen mehr möglich)
- Druckbetankung (kann bei schlechten Systemen zum Stillstand führen, gerade wenn OS auf Datenpool)

- PVE auf selbem Raid wie Systeme (kann bei hoher Last zum Stillstand führen)
- FreeNAS auf USB Sticks (Die Zeit ist vorbei, bitte SSDs nehmen mit USB Adapter, als Beispiel)
- Virtuelle Maschinen mit ZFS - Abgesehen von der Speicherplatzverschwendung, kann es zu unkontrollierten Mounts aus ZVOLS kommen

Update 2024 zum Thema Volblocksize

root@pve83:~# zfs create -ovolblocksize=8k rpool/8k -V 1GWarning: volblocksize (8192) is less than the default minimum block size (16384).To reduce wasted space a volblocksize of 16384 is recommended.

rpool/16k	1.11G	15.2T	81.4K	-
rpool/8k	1.49G	15.2T	81.4K	-

NAME	STATE	READ	WRITE	CKSUM			
rpool	ONLINE	0	0	0			
raidz1-0	ONLINE	0	0	0			
scsi-201000000000000005cd2e475ae4b5651-part3	ONLINE	0	0	0			
scsi-201000000000000005cd2e4f1bd4e5651-part3	ONLINE	0	0	0			
scsi-201000000000000005cd2e4ef8f4b5651-part3	ONLINE	0	0	0			
scsi-201000000000000005cd2e4889d4b5651-part3	ONLINE	0	0	0			

Bei RaidZ2 wäre der Overhead über 200%

Inzwischen warnt ZFS, jedoch sieht das keine in der PVE GUI!