ZFS RAID

Übersicht ZFS-RAID

- RAID-1 (mirror) *Empfohlen z.B. für reines Proxmox-System ohne VMs*
- RAID-Z1 (1 Parity-Bit, ~RAID 5)
- RAID-Z2 (2 Parity-Bits, ~RAID 6)
- RAID-Z3 (3 Parity-Bits)

RaidZ(2,3)

- Sinnvoll für Datenablage und gelegentlichen Zugriff
- Oft in Kombination mit harter Kompression im NAS Betrieb
- Bei Erweiterung müssen alle Datenträger im Verbund getauscht werden
- Bei PVE als Datastore dringend 16k Blocksize für neue Systeme hinterlegen, da sonst extremer Overhead (siehe Warnung)
- x-y Verlust an Platz für Redundanz

Raid Z+1

- Eventuelle Kette aus mehreren RaidZ denkbar
- Besserer Kompromiss aus Platzbedarf und Redundanz
- Für VMs mit leichtem Anspruch akzeptabel
- 2x x-y Verlust an Platz für Redundanz

Raid 1+0

- Absolut notwendig für VM Einsatz mit drehenden HDDs, sonst Latenz schlecht
- Weitere Spiegel sind jederzeit anfügbar
- Einzelne Spiegel können während des Betriebs ohne Redundanzverlust getauscht werden (wenn Slots frei)
- Tausch zweier Datenträger führt zur Erweiterung der Kapazitäten (`zpool set autoresize=on <rpool>`)
- Jedoch 50% Platzverlust für Redundanz

Raid5+0

- Kompromiss

Solidstate Disks

- Bei ZFS immer zu bevorzugen, da Latenz ca 2x langsamer als ohne ZFS
- Für HDD Systeme immer Cache Device einsetzen, dieses muss nicht redundant oder hotswappable sein. Ausfall unterbricht nicht den Betrieb
  - **SATA**
    - nur 6GBit/s halb-duplex (ca. 600 Megabyte/s)
    - bei onBoard Controller darauf achten, mit wie vielen PCIe Lanes und mit welcher PCIe Generation angebunden, ggf. Flaschenhals
    - günstig
    - Samsung Pro, Intel mit krummen Zahlen (960GB, 1920GB), z.B. DC-S4510
    - nicht immer Hot-Swap möglich
    - SSDs im Formfaktor 2,5 Zoll und M.2 gängig
  - **SAS**
    - 12GBit/s und voll-duplex (ca. 1,2 Gigabyte/s)
    - teurer, aber weniger Müll am Markt
    - Hersteller noch unklar
    - Hot-Swap
  - **NVMe**
    - Protokoll setzt nativ auf PCIe auf, Latenz und Overhead gegenüber den klassischen Protokollen (SATA, SAS) sehr gering
    - Vollduplex
    - Bandbreite abhängig von PCIe-Generation und Anzahl der angebundenen Lanes (z.B. mit PCIe 4.0 x4 sind etwa 7 Gigabyte/s möglich, mit PCIe 3.0 x4 etwa 3,5 Gigabyte/s)
    - **M.2 Formfaktor**
      - nicht verwechseln mit M.2 SATA SSDs, PCIe / NVMe muss in den Specs stehen
      - Sinnvoll für Cache oder LOG
      - Wenn durch Bios bootfähig kann man sich ein Notsystem darauf bauen
      - günstig
      - belegen keinen Slot
    - **U.2 Formfaktor (SFF-8639)**
      - Bekannte irische Band LOL
      - Einsatzbereich Server und Storage
      - Wenig Erfahrung mit ZFS
      - Hot-Swap
      - PCIe 3.0 x4 Anbindung, Abwärtskompatibel zu SAS und SATA, Mischbetrieb über ein Kabel mit entsprechender Backplane möglich
      - teuer
    - **U.3 Formfaktor (SFF-TA-1001)**
      - Nachfolger der U.2 Schnittstelle
      - PCIe4.0 Unterstützung

- weitere Eigenschaften identisch zur U.2 Schnittstelle

RAIDZ-BUG

NAME	VOLBLOCK	LREFER	REFER
pool1/test5	8K	16.7G	35.7G (213% overhead)
pool1/test2	16K	89.4G	95.6G (6,93% overhead
pool1/test3	32K	89.9G	96.0G (6,78% overhead)
pool1/test4	64K	50.1G	53.5G (6,78% overhead)
pool1/test	128K	60.7G	60.7G (0% overhead)

Version #2
Erstellt: 2023-01-04 21:42:30 UTC von Admin
Zuletzt aktualisiert: 2023-09-13 12:54:42 UTC von Björn Roters