Engineer

RAID

☰ Mehrere Datenträger als Einheit, meist mit Ausfallschutz

Was es ist

RAID betreibt mehrere Datenträger als ein einziges Blockgerät, damit die Gruppe schneller, größer oder ausfallsicherer sein kann als jedes einzelne Mitglied. Der Name stammt aus einem technischen Bericht der University of California, Berkeley, von 1987, verfasst von David Patterson, Garth Gibson und Randy Katz: *Ein Plädoyer für redundante Verbünde preiswerter Datenträger (RAID)*. Linux schreibt die Stufen klein und ohne Leerzeichen, raid1 oder raid5, wie sein md-Treiber sie meldet.

Der Bericht beginnt mit einem Missverhältnis. Prozessoren und Arbeitsspeicher wurden schneller. Der Abstract warnt, dass dieser Gewinn verloren geht, wenn Ein- und Ausgabe nicht mithalten. Er stellt den einzelnen großen teuren Datenträger, Single Large Expensive Disk oder SLED, dessen Kapazität rasch, dessen Leistung aber nur mäßig gewachsen war, Verbünden aus preiswerten PC-Datenträgern gegenüber. Von diesen erwartet er gegenüber dem großen Datenträger ungefähr zehnfache Verbesserungen bei Geschwindigkeit, Zuverlässigkeit, Strombedarf und Erweiterbarkeit. Der Vergleich ist konkret: eine IBM 3380 für Großrechner, eine Fujitsu Super Eagle für Minicomputer und die Conner CP 3100 mit 100 MB für Personal Computer.

Dann folgt der Haken, in einem von den Autoren als schlechte Nachricht betitelten Abschnitt. Fallen Datenträger unabhängig und mit konstanter Rate aus, was laut Bericht die Annahme der Hersteller bei Angaben zur mittleren Zeit bis zum Ausfall (MTTF) ist, entspricht die MTTF eines Verbunds der eines Datenträgers geteilt durch die Zahl der Datenträger. Hundert CP 3100 mit jeweils 30.000 angegebenen Stunden ergeben 300 Stunden, weniger als zwei Wochen; tausend ergeben 30 Stunden. Das Fazit ist deutlich: „Ohne Fehlertoleranz sind große Verbünde preiswerter Datenträger zu unzuverlässig, um nützlich zu sein.“

Die Antwort ist Redundanz. Die Datenträger werden in Zuverlässigkeitsgruppen aufgeteilt, jeweils mit zusätzlichen Prüflaufwerken, die redundante Informationen speichern. Fällt ein Datenträger aus, wird er ersetzt und sein Inhalt anhand dieser Informationen auf dem neuen wiederhergestellt. Die dafür benötigte Zeit ist die mittlere Reparaturdauer. Hot-Spare-Laufwerke, die beim Ausfall elektronisch zugeschaltet werden, verkürzen sie. Der Bericht beschreibt dann fünf Organisationsformen und nennt jede eine Stufe. Stufe 1 spiegelt jeden Datenträger. Stufe 2 korrigiert Fehler mit einem Hamming-Code, wie es Speichersysteme taten. Stufe 3 braucht nur ein Prüflaufwerk je Gruppe. Stufe 4 erlaubt kleine Lese- und Schreibvorgänge gleichzeitig auf getrennten Datenträgern. Stufe 5 verteilt Daten und Prüfinformationen auf alle Datenträger, damit kein einzelnes Prüflaufwerk Schreibvorgänge aufhält.

Die Praxis hat die Liste erweitert. mdadm, das Linux-Software-RAID verwaltet, akzeptiert unter anderem linear, raid0 (oder stripe), raid1, raid4, raid5, raid6 und raid10. Striping ohne Redundanz, heute Stufe 0 genannt, steht nicht in der Klassifikation von 1987, die mit gespiegelten Datenträgern beginnt. Der Bericht ließ offen, ob ein Verbund in Hardware oder Software aufgebaut wird, und bezeichnete das allein als Kosten-Nutzen-Frage. Linux md ist die Softwarevariante; eine Controllerkarte, die dem Betriebssystem einen großen Datenträger zeigt, ist die Hardwarevariante.

Das Konzept für eine Entscheidung nutzen

Nutzen Sie „Was ist RAID?“ für eine konkrete Entscheidung. Wir können helfen, das Konzept mit dem System in Verbindung zu bringen, das Sie entwickeln oder einkaufen. Sicherungen, deren Funktion geprüft und nicht vorausgesetzt wird, und ein schriftlicher Plan für den Morgen, an dem ein Server nicht wieder startet.

Unsere Leistungen entdecken Kontakt aufnehmen →

Die Annahmen hinter dem Begriff prüfen

Nutzen Sie die Erklärung und die verlinkten Arbeiten, um Fragen für Ihr Projekt zu formulieren. Wir können die Annahmen gemeinsam mit Ihnen prüfen, bevor ein Werkzeug ausgewählt wird. 100 wichtige Sicherungen mit Barman automatisiert — 100 automatisierte geschäftskritische Datenbanksicherungen

100 wichtige Sicherungen mit Barman automatisiert Kontakt aufnehmen →

Warum es wichtig ist

Ein Datenträger verschleißt. Bei einem Server mit nur einem Datenträger stoppt dessen Ausfall den Rechner und überlässt die Daten der vorhandenen Sicherung, falls es eine gibt. Genau damit beginnt der Zuverlässigkeitsabschnitt des Berichts: Datenträger sind unzuverlässig genug, dass Administratoren jede Nacht alles sichern. Mit Redundanz wird aus einem defekten Datenträger eine Reparatur statt eines Ausfalls. Der Verbund arbeitet mit den verbleibenden

Mitgliedern weiter, in dem von Linux degraded genannten Zustand, und erreicht wieder seine volle Funktionsfähigkeit, sobald ein Ersatz hinzugefügt und dessen Datenbestand wiederaufgebaut wurde.

Dieser Schutz hält nur, solange die verbleibenden Datenträger halten. Die Rechnung des Berichts betrachtet Daten als verloren, wenn in derselben Gruppe ein zweiter Datenträger ausfällt, bevor der erste repariert ist. Die Dauer des Reparaturfensters zählt deshalb ebenso wie die Zahl der Prüflaufwerke. Mit Hot-Spare-Laufwerken und kurzen Reparaturen werden die Zahlen eindrucksvoll: Für einen Verbund der Stufe 5 mit 1.000 Datenträgern in Zehnergruppen und 5 Prozent Reservelaufwerken berechnet der Bericht noch eine MTTF von 90 Jahren. Anderswo warnt er, dass ein Erdbeben oder eine Überspannung die Datenträger eines Verbunds gemeinsam ausfallen lassen kann. Seine Formeln setzen voraus, dass dies nicht geschieht.

Die Wahl einer Stufe wägt Kapazität, Geschwindigkeit und Kosten gegeneinander ab. Das Fazit des Berichts lautet, dass gespiegelte Datenträger zur Transaktionsverarbeitung passen, wenn höchstens die Hälfte des Speichers genutzt wird. Stufe 5 erscheint günstiger, wenn mehr Kapazität benötigt wird, für Supercomputerarbeit oder für eine Mischung daraus und aus Transaktionsverarbeitung.

Ist die Redundanz aufgebraucht, bleibt Datenrettung. Der Portfolioeintrag zur Datenrettung von SD-Karten, Festplatten, SSDs und RAID behandelt einen ausgefallenen RAID-Verbund als eigenes Problem, getrennt von einer toten SSD oder einem defekten Plattenlaufwerk. Die passende Vorgehensweise hängt vom Medium ab.

Wie es in der Praxis aussieht

Nehmen wir einen Server mit je einer Partition auf zwei Datenträgern, `/dev/sda1` und `/dev/sdb1`, die für ein Datenvolume gespiegelt werden sollen. Unter Linux md bauen drei Befehle den Verbund auf, erzeugen darauf ein Dateisystem und halten ihn in der Konfiguration fest:

```
mdadm --create /dev/md0 --level=1 --raid-devices=2 /dev/sda1 /dev/sdb1
mke2fs -t ext4 -L data /dev/md0
mdadm --detail --scan >> /etc/mdadm/mdadm.conf
```

Der erste erzeugt `/dev/md0`, einen RAID1-Verbund aus beiden Partitionen. Der Kernel beginnt wie nach jeder Neuanlage einen Resync und berechnet dabei die Redundanz neu. Der zweite formatiert den Verbund genau wie einen einzelnen Datenträger; das Dateisystem sieht ein Gerät. Der dritte hängt eine Beschreibung des Verbunds an die mdadm-Konfiguration unter `/etc/mdadm.conf` oder `/etc/mdadm/mdadm.conf` an. Das Handbuch verlangt, die Ausgabe vor der Verwendung zu prüfen, da sie unerwünschte Details enthalten kann.

Drei Dateien zeigen den Zustand des Verbunds:

```
cat /proc/mdstat
cat /sys/block/md0/md/degraded
cat /sys/block/md0/md/sync_action
```

`/proc/mdstat` listet jedes aktive md-Gerät mit seinem Zustand auf. mdadm liest die Datei, um Verbünde zu finden und den Wiederaufbau zu verfolgen. Jeder Verbund hat außerdem ein Verzeichnis unter `/sys/block/`. Darin enthält `degraded` die Zahl fehlender Mitglieder: 0 bei einem vollständigen Verbund, 1 bei einem fehlenden Datenträger. `sync_action` zeigt `idle`, wenn nichts geschieht, `resync` nach der Neuanlage oder einem unsauberen Herunterfahren, `recover`, während ein Hot-Spare-Laufwerk den Inhalt eines verlorenen Mitglieds übernimmt, sowie `check` oder `repair` während einer vollständigen Überprüfung. `check` hineinzuschreiben startet diese Prüfung, die jeden Block des Verbunds liest:

```
echo check > /sys/block/md0/md/sync_action
cat /sys/block/md0/md/mismatch_cnt
```

Danach gibt `mismatch_cnt` die Zahl der neu geschriebenen Sektoren an oder derjenigen, die neu geschrieben worden wären. Laut Kerneldokumentation arbeiten die meisten Stufen mit Seiten statt Sektoren. Die Zahl kann deshalb größer als die tatsächliche Fehlerzahl sein.

Von einem Ausfall muss auch jemand erfahren. `mdadm --monitor` fragt die Verbünde ab und macht aus seinen Beobachtungen Ereignisse, die es als E-Mail an eine Adresse senden oder einem Programm übergeben kann. Auf einem von systemd betriebenen System nennt das Handbuch `mdmonitor.service` als Hauptweg und `MAILADDR` in `mdadm.conf` als einzutragende E-Mail-Einstellung.

Nur fünf Ereignisse senden E-Mail: `Fail`, `FailSpare`, `DegradedArray`, `SparesMissing` und `TestMessage`. Das letzte soll belegen, dass der Meldeweg funktioniert:

```
MAILADDR ops@example.com
```

```
mdadm --monitor --scan --oneshot --test
mdadm --detail --test /dev/md0
```

Der erste Befehl sendet für jeden gefundenen Verbund eine Testnachricht und endet. Der zweite eignet sich für Skripte: Mit `--test` liefert er den Exitstatus 0 für einen gesunden Verbund, 1 bei mindestens einem ausgefallenen Gerät, 2 bei so vielen Ausfällen, dass der Verbund unbrauchbar ist, und 4, wenn mdadm beim Sammeln der Verbundinformationen einen Fehler feststellt.

Der Verbund deckt ein Ereignis ab, den Ausfall eines Datenträgers, und sonst nichts. Der Portfolioeintrag über automatisierte Sicherungen und Wiederherstellungstests beschreibt einen getrennten Wiederherstellungsentwurf: verschlüsselte, deduplizierte restic-Snapshots bei einem anderen Speicheranbieter über SFTP, systemd-Zeitgeber und einen monatlichen Ablauf, der eine bekannte Datei an einen temporären Ort wiederherstellt und ihre Bytes vergleicht. Das ist ein Wiederherstellungstest im Entwurf, kein Beleg seiner Ausführung. Der Implementierungsnachweis vom 2026-10-27 hält fest, dass die Zeitgeber ohne konfiguriertes Repository deaktiviert waren und keine abgeschlossene echte Wiederherstellung verzeichnet wurde.

Wo es schiefgeht

Ein Verbund ist für ein Ereignis gebaut, den Ausfall eines Mitgliedsdatenträgers. Die meisten Arten, auf die er Daten verliert, liegen außerhalb dieses Falls.

Der Verbund wird für ein Backup gehalten. Ein Spiegel schreibt jede Änderung gleichzeitig auf beide Datenträger. Dazu gehören versehentlich gelöschte Dateien, durch eine fehlerhafte Migration entfernte Tabellen und durch Ransomware verschlüsselte Volumes. Jedes Mitglied enthält dann denselben

Schaden. Die Korrektur ist eine separate Sicherung außerhalb des Hosts mit regelmäßig getesteten Wiederherstellungen. So schützt der Verbund gegen Hardwarefehler und das Backup gegen alles andere.

Ein degradierter Verbund bleibt unbemerkt. Ein degradierter Verbund arbeitet weiter und sieht bis zum nächsten Datenträgerausfall nicht anders aus. mdadm sendet ohne Adresse oder Programm keine E-Mail. Mit `--scan` und ohne beides überwacht es laut Handbuch nichts. `DegradedArray` betrifft zudem nur Verbünde, die beim ersten Auffinden bereits degradiert sind. Ein später ausfallender Datenträger erzeugt ein `Fail` -Ereignis. Die Korrektur: `MAILADDR` setzen, den Betrieb von `mdmonitor.service` sicherstellen und mit einer `--test` -Nachricht belegen, dass die E-Mail ankommt.

Der zweite Datenträger fällt beim Wiederaufbau aus. Der Wiederaufbau rekonstruiert den fehlenden Datenträger aus allen anderen. Das Modell des Berichts verliert die Daten genau dann, wenn ein weiterer Datenträger der Gruppe innerhalb dieser Reparaturdauer ausfällt. Eine Woche lang nicht ersetzte defekte Hardware verlängert das Fenster auf eine Woche. Verkürzen Sie es: Halten Sie ein Hot-Spare-Laufwerk im Verbund bereit, das mdadm bei der Neuanlage mit `--spare-devices` einrichten kann, ersetzen Sie defekte Datenträger umgehend und führen Sie einen geplanten `check` aus. So werden problematische Datenträger gefunden, solange der Verbund die Fehler noch korrigieren kann.

Die Datenträger sind nicht unabhängig. Die MTTF-Rechnung nimmt einzelne, unabhängige Ausfälle an. Der Bericht selbst nennt Erdbeben und Überspannungen als Gegenbeispiele. Zwei Datenträger im selben Gehäuse teilen Stromversorgung, Controller, Raum und Gebäude. Betrachten Sie die berechnete Zahl als Obergrenze und decken Sie gemeinsame Ursachen durch etwas außerhalb des Gehäuses ab, etwa eine Kopie an einem anderen Standort.

Ein RAID-5- oder RAID-6-Verbund ist zugleich unsauber und degradiert. Stoppt der Rechner unsauber bei fehlendem Mitglied, ist die Parität nicht verlässlich und manche Blöcke lassen sich nicht wiederaufbauen. Die Kerneldokumentation warnt daher vor unentdeckbarer Beschädigung; md weigert sich, den Verbund zu starten. Das Erzwingen mit `mdadm --assemble --force` oder bei einem Rootdateisystem mit dem Modulparameter `start_dirty_degraded` nimmt dieses Risiko in Kauf. Wählen Sie vorher eine Konsistenzstrategie: Für die Stufen 4, 5 und 6 kann md Schreibvorgänge auf einem Journalgerät protokollieren. Für Stufe 5 bietet es ein partielles Paritätslog, das laut

Dokumentation die Schreiblücke schließt. Bleibt das Erzwingen der einzige Zugang, behandeln Sie die Daten als verdächtig und gleichen Sie sie mit einer Sicherung ab.

RAID 0 wird als Schutz angesehen. md führt raid0 und linear als Verbünde ohne Redundanzinformationen. mdadms Monitor meldet sie mit `Wrong-Level`, weil sie keine der von ihm beobachteten Ausfall-, Reserve- und Resync-Operationen unterstützen. Es gibt nichts, woraus ein fehlendes Mitglied rekonstruiert werden könnte. Verwenden Sie Stufe 0 nur für wegwerfbare oder neu erzeugbare Daten und rechnen Sie sie niemals als RAID im Sinne des Berichts von 1987 an.

Ein Hardwarecontroller verbirgt die Datenträger. Hinter einer Controllerkarte sieht das Betriebssystem ein logisches Gerät. Eine Anfrage nach dem Laufwerkszustand erreicht damit den Controller statt der Laufwerke. smartctl benötigt zum Beispiel `-d megaraid,N`, um Laufwerk N hinter einem MegaRAID-Controller anzusprechen. Laut Handbuch gilt dies auch für Dell-PERC-Karten. Überwachen Sie jeden physischen Datenträger über die eigene Schnittstelle des Controllers, damit ein defektes Mitglied bemerkt wird, solange der Verbund noch vollständig ist.

Weiterführende Lektüre

- Ein Plädoyer für redundante Verbünde preiswerter Datenträger (RAID) — die Seite des Berkeley-Berichts mit Abstract und PDF.
- UCB/CSD-87-391, vollständiger Bericht (PDF) — MTTF-Rechnung, Zuverlässigkeitsgruppen und die fünf Stufen.
- RAID-Verbünde in der Linux-Kerneldokumentation — md-Dateien unter `/sys`, unsaubere degradierte Verbünde und Konsistenzstrategien.
- mdadm(8) — Verbünde anlegen, Überwachungsereignisse, E-Mail und die Exitcodes von `--detail --test`.
- smartctl(8), smartmontools — Laufwerke hinter RAID-Controllern mit der Option `-d` erreichen.

Glossar

Für Ihre tatsächlichen Rahmenbedingungen auswählen

Beschreiben Sie den Bereitstellungsweg, die Betriebsrisiken und die Bereitschaftsverantwortlichen. Wir können eine Änderung mit Prüfungen, einem Wiederherstellungsweg und klaren Zuständigkeiten abgrenzen.

Preise ansehen Kontakt aufnehmen →

Die Begriffe praktisch nutzbar machen

Fragen Sie, wie „Was ist RAID?“ auf Ihre Organisation zutrifft. Ein konkretes Beispiel aus Ihrem bestehenden System reicht aus, um das Gespräch zu beginnen.

Unsere Leistungen entdecken

welcome@engineer.company +45 50 10 61 84 Kopenhagen, Dänemark

Nehmen Sie diese Seite mit

Kontaktkarte

Mit Sorgfalt und Liebe gestaltet von D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

Copyright © 2025 – heute · Engineer ApS · Alle Rechte vorbehalten.

Gedruckt von https://engineer.company/de/glossary/raid/