

Fakten statt Bauchgefühl: RAID-Mathematik für Admins

- ▶ Heinlein Professional Linux Support GmbH
 - ▶ Holger Uhlig
 - ▶ h.uhlig@heinlein-support.de

Agenda:

- ▶ Was will ich?
 - ▶ MB/s vs. IOPS
- ▶ Worauf achten?
 - ▶ Berechnung von Durchsatz und IOPS
 - ▶ Manndeckung oder Libero
 - ▶ Ein Vergleich verschiedener RAID-Level
- ▶ Ab wann ist RAID-6 sinnvoll?
 - ▶ Mean Time To Data Loss
 - ▶ Bit Error Rate

Was will ich?

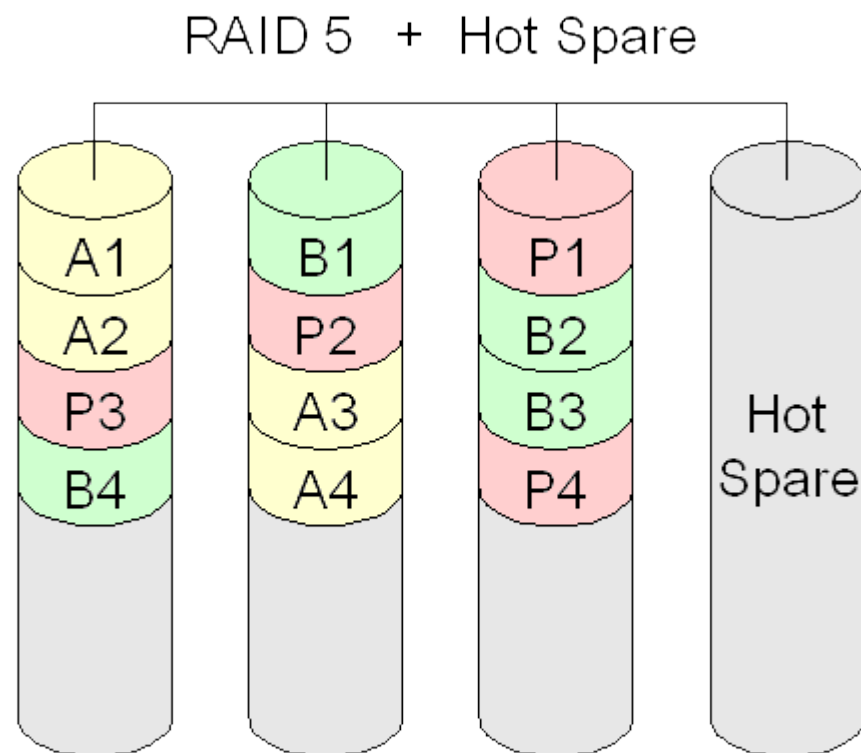
MB/s vs. IOPS:

- ▶ „Ich kann nicht beides haben.“
 - ▶ MB/s wenn der Nachteil zur Ausrichtung mehrerer Festplattenköpfe ausgeglichen ist
 - ▶ I/Os wenn ein Request im Volumen durch eine Festplatte schneller abgearbeitet werden kann
- ▶ Faustregeln:
 - ▶ Applikationen mit gegensätzlichen I/O-Anforderungen trennen
 - ▶ I/O-Arrays isolieren (entweder sequential oder random)
- ▶ Ohne Requestmerge, Controller- / Diskcache rechnen

Worauf achten?

Kalkulation MB/s:

- ▶ Abhängigkeit definiert durch:
 - ▶ IOPS:
I/O pro Sekunde aller Festplatten des RAID
 - ▶ TransferSize = Chunksize
- ▶ Chunk:
 - ▶ log. Datenblock einer Festplatte
- ▶ Stripe:
 - ▶ Summe aller Chunks einer Datengruppe



Worauf achten?

Kalkulation MB/s:

- ▶ Formel:

$$MB/s = \frac{IOPS \times TransferSize}{1024}$$

- ▶ Beispiel für 2000 IOPS bei 64KB TransferSize

$$MB/s = \frac{2000 IOPS \times 64 KB}{1024}$$

$$MB/s = 128.000 KB / 1024$$

$$MB/s = 125$$

Worauf achten?

Kalkulation IOPS:

- ▶ Formel:

$$IOPS = 1 / [Rotation_{Latenz} + Seektime_{AVG}] * 1000$$

- ▶ Rotation – Zeit für eine halbe Umdrehung
 - ▶ Seektime – Zeit der Lesekopfausrichtung (Herstellerangabe)
-
- ▶ Beispiel einer 72GB, 15k SAS mit 3ms Seektime

$$IOPS = 1 / [2ms + 3ms] * 1000$$

$$IOPS = 1 / 5 * 1000$$

$$IOPS = 0,2 * 1000$$

$$IOPS = 200$$

Worauf achten?

Latenzzeit der Sektorsuche:

- ▶ Formel:

$$Rotation_{Latenz} = 1 / \frac{RPM}{2}$$

- ▶ Beispiel einer 72GB, 15k SAS mit 3ms Seektime

$$Rotation_{Latenz_{AVG}} = 1 / \left[\frac{15.000}{60.000ms} \right] / 2$$

$$Rotation_{Latenz_{AVG}} = 4ms / 2$$

$$Rotation_{Latenz_{AVG}} = 2ms$$

Worauf achten?

Die richtige Chunksize:

- ▶ überwiegend kleine I/Os => Chunksize etwas größer als I/O-Size
 - ▶ Ziel: Request wird nur von einer Platte gelesen
 - ▶ Vorteil: andere Platten können weitere I/O-Requests bedienen
 - ▶ Beispiel: Mailserver (viele kleine Dateien 4-7KB)

- ▶ überwiegend grosse I/Os => Chunksize so klein wie sinnvoll :-)
 - ▶ Ziel: Gemeinsame Lese/Schreibgeschwindigkeit beteiligter Platten
 - ▶ Nachteil: Positionierungszeit aller Platten
 - ▶ Beispiel: Video-/Downloadportale (grosse Dateien Iso, Flv, ...)

Worauf achten?

Die richtige Chunksize:

- ▶ Beispiel:

$$MB/s = \frac{2000 \text{ IOPS} \times 64 \text{ KB}}{1024}$$

$$MB/s = 128.000 \text{ KB} / 1024$$

$$MB/s = 125$$

$$MB/s = \frac{2000 \text{ IOPS} \times 256 \text{ KB}}{1024}$$

$$MB/s = 512.000 \text{ KB} / 1024$$

$$MB/s = 500$$

- ▶ ! Das maximale Durchsatzvolumen ist zusätzlich abhängig vom Bus und der Netzbandbreite.

Worauf achten? Manndeckung oder Libero

- ▶ Wonach ist ein RAID auszurichten?
 - ▶ Lese-Performance ist von RAID 10 : 5 : 6 annähernd gleich.
 - ▶ Ausschlaggebend ist daher die Write-Performance, d.h. der Aufwand und Performanceverlust für einen Schreibzyklus.
 - ▶ Erweitert die Priorität der Volumen- und Kosteneffizienz beachten.
 - ▶ Mirror-RAID => Speicher+Kosten < Performance
 - ▶ Parity-RAID => Speicher+Kosten > Performance

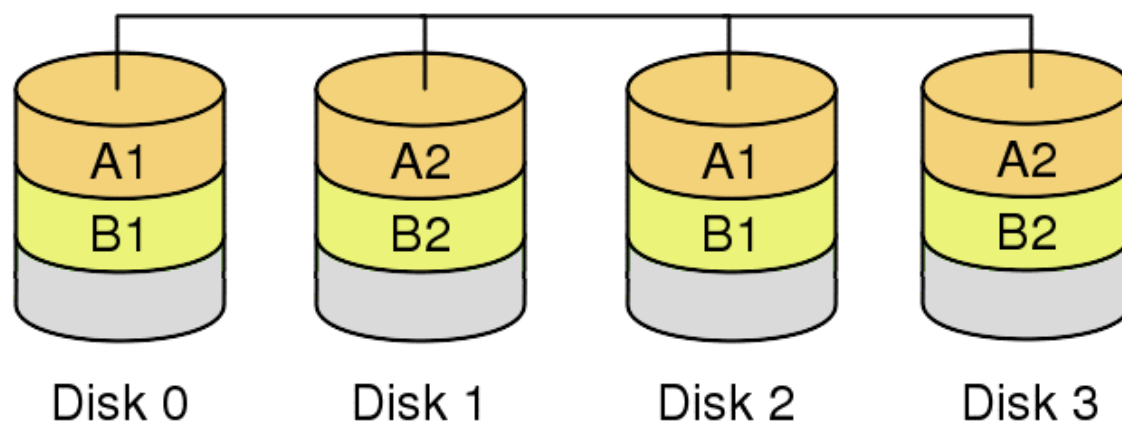
Worauf achten?

Write-Performance RAID-10:

- ▶ Wieviele Platten brauche ich für X geforderte Write-IOPS?

$$Disks_{Raid10} = (IOPS_{Ziel} / IOPS_{Disk}) \times 2$$

- ▶ 50% Writeperformance
 - ▶ logischer Write = 2 phy. Writes
 - ▶ Schreiben der neuen Daten und neuen Daten als Redundanz



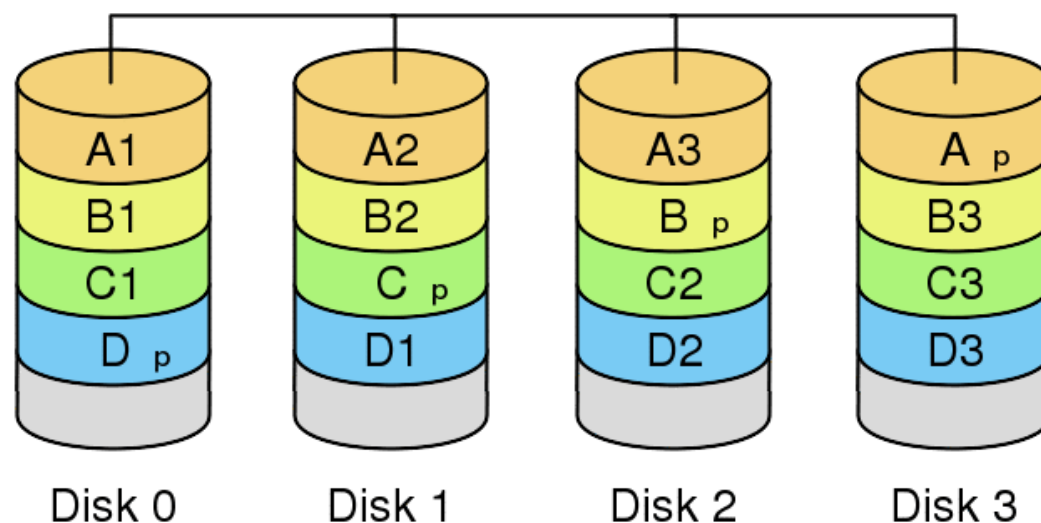
Worauf achten?

Write-Performance RAID-5:

- ▶ Wieviele Platten brauche ich für X geforderte Write-IOPS?

$$Disks_{Raid5} = (IOPS_{Ziel} / IOPS_{Disk}) \times 4$$

- ▶ 25% Writeperformance
 - ▶ logischer Write = 2 phy. Reads + 2 phy. Writes
 - ▶ Lesen der alten Daten und alten Parität
 - ▶ Schreiben der neuen Daten und neuen Parität



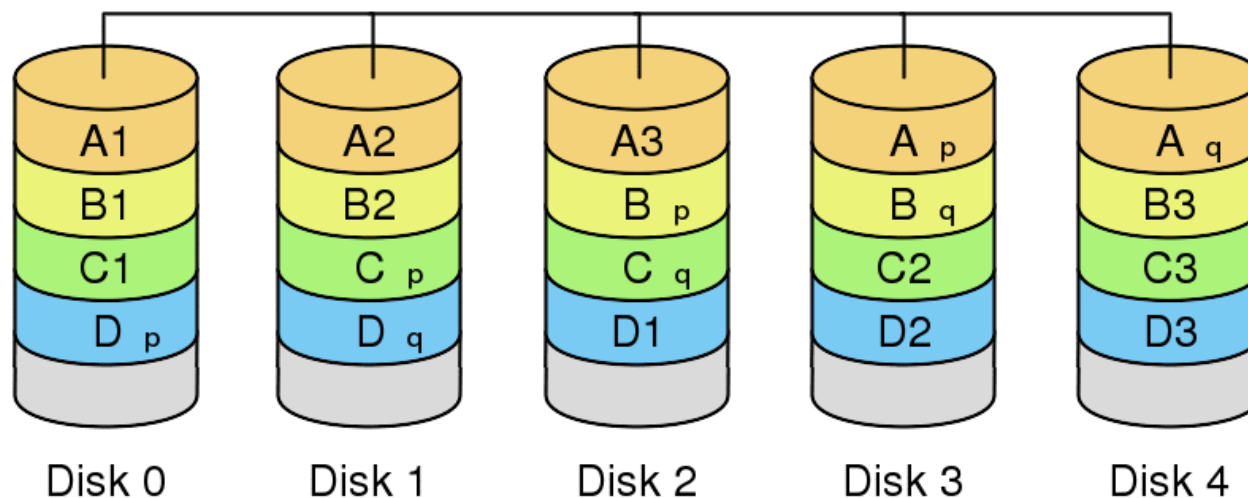
Worauf achten?

Write-Performance RAID-6:

- ▶ Wieviele Platten brauche ich für X geforderte Write-IOPS?

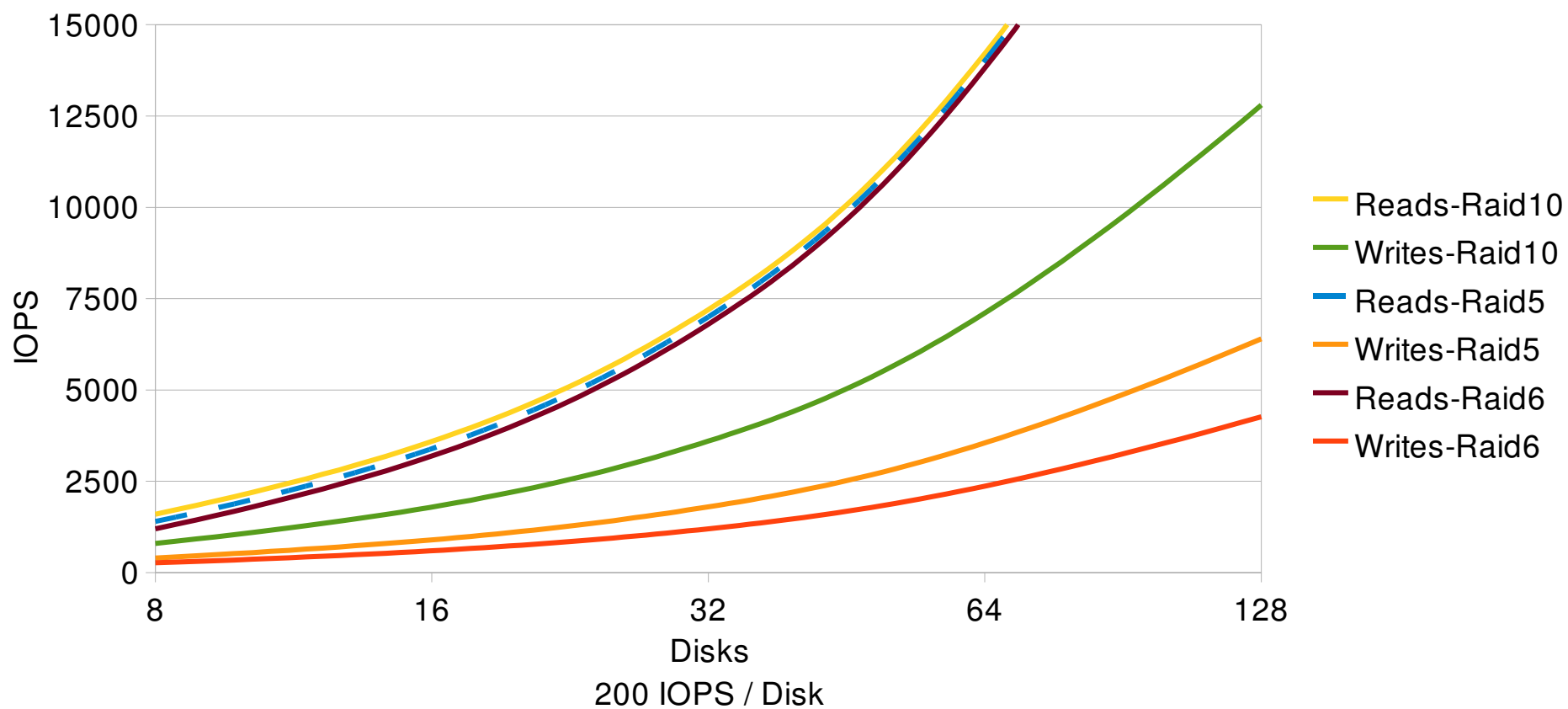
$$Disks_{Raid6} = (IOPS_{Ziel} / IOPS_{Disk}) \times 6$$

- ▶ 17% Writeperformance
 - ▶ logischer Write = 3 phy. Reads + 3 phy. Writes
 - ▶ Lesen der alten Daten und alten Paritäten
 - ▶ Schreiben der neuen Daten und neuen Paritäten



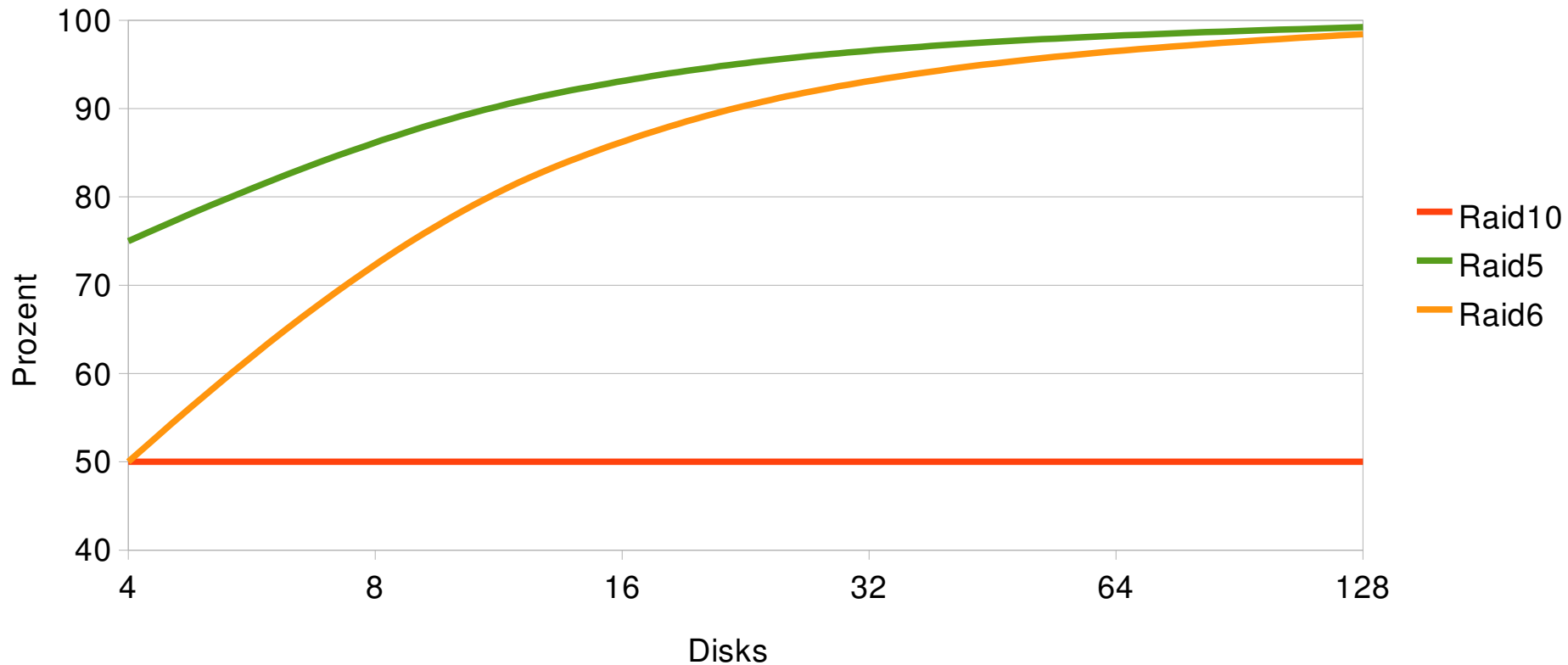
Worauf achten? Ein Vergleich:

IO Durchsatz
RAID 10 : 5 : 6



Worauf achten? Ein Vergleich:

Speichereffizienz
Raid 10 : 5 : 6



Ausfallwahrscheinlichkeit

Die zu beachtenden Faktoren:

- ▶ **MTTF:** vom Hersteller definierte Lebenszeit der Festplatten
 - ▶ i.d.R ein Wert zwischen 250.000 – 1.000.000 Stunden
- ▶ **MTTR:** durchschnittliche Zeitspanne zur Reparatur
 - ▶ der Zeitraum von Ausfall bis zum abgeschlossenen Rebuild
- ▶ **N:** Anzahl der Festplatten
- ▶ **G:** Anzahl Festplatten der ParityGroup

Ausfallwahrscheinlichkeit Mean Time To Data Loss:

- ▶ Berechnung für RAID-5 bei Ausfall von zwei Festplatten
 - ▶ Double Disk Failure

$$DDF = \frac{MTTF^2}{N \times (G - 1) \times MTTR}$$

- ▶ Berechnung für RAID-6 bei Ausfall von drei Festplatten
 - ▶ Triple Disk Failure

$$TDF = \frac{MTTF^3}{N \times (G - 1) \times (G - 2) \times MTTR^2}$$

Ausfallwahrscheinlichkeit Mean Time To Data Loss:

▶ Lebenszeit von RAID-5 und RAID-6

Disks**	4	6	8	10	20	40
RAID-5*	24.841	9.937	5.323	3.312	784	191
RAID-6*	129.381.945	25.876.389	9.241.568	4.312.732	453.972	52.381

* MTTDL in Jahren

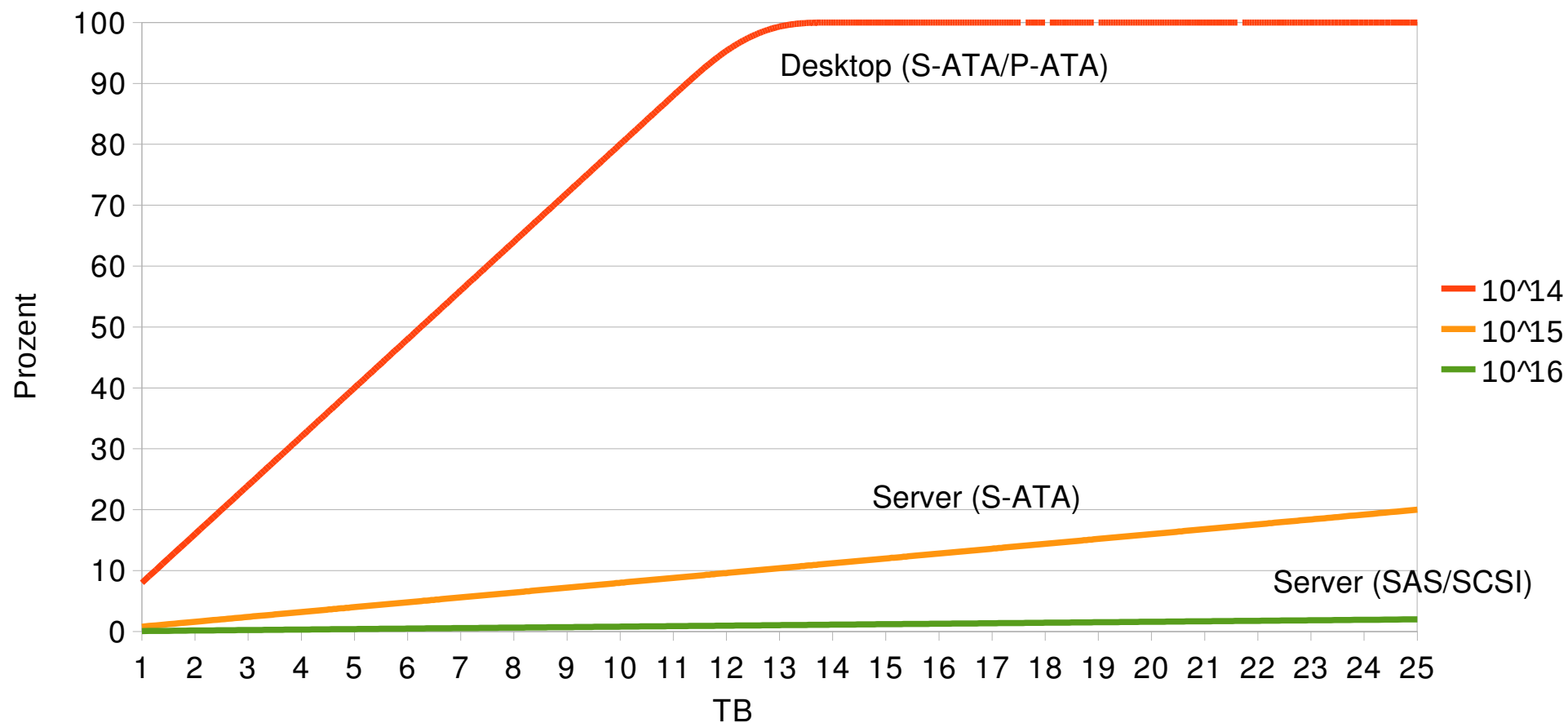
** MTTF 250.000h

MTTR 24h

- ▶ Warum macht RAID-6 dann Sinn?
 - ▶ Bei Betrachtung der Bitfehlerwahrscheinlichkeit.

Ausfallwahrscheinlichkeit Bitfehlerwahrscheinlichkeit:

Abhängig zur Gesamtkapazität



Ausfallwahrscheinlichkeit

Mean Time To Data Loss mit Bit Error Rate:

- ▶ P Ist die Wahrscheinlichkeit, alle Sektoren erfolgreich auf einer Festplatte fehlerfrei lesen zu können.

- ▶ RAID-5

$$DF + BER = \frac{MTTF}{N \times (1 - p_{disk}^{(G-1)})}$$

- ▶ RAID-6

$$DDF + BER = \frac{MTTF^2}{N \times (G-1) \times (1 - p_{disk}^{(G-2)}) \times MTTR}$$

Ausfallwahrscheinlichkeit

Mean Time To Data Loss mit Bit Error Rate:

▶ Lebenszeit von RAID-5 und RAID-6 mit BER

Disks**	4	6	8	10	20	40
RAID-5*	62	26	14	9	3	1
RAID-6*	316.854	65.956	24.503	11.888	1.507	242

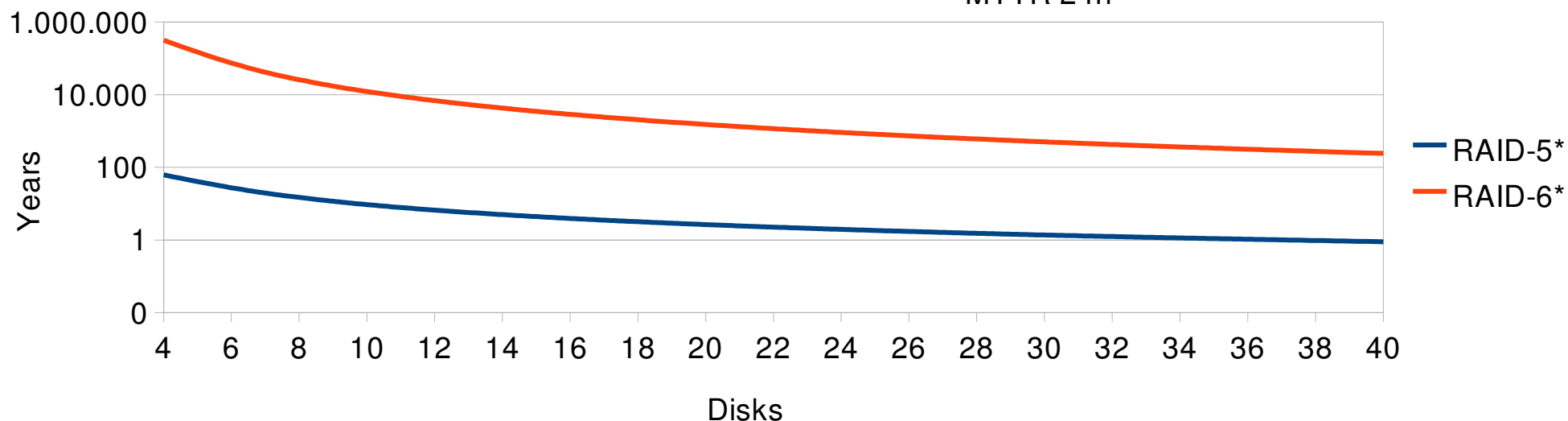
* MTTDL in Jahren

** MTTF 250.000h

* P 96 %

MTTR 24h

RAID-5 : RAID-6



Heinlein Professional Linux Support GmbH:

▶ AKADEMIE

- ▶ Von Profis für Profis: Wir vermitteln die oberen 10% Wissen. Geballtes Wissen und umfangreiche Praxiserfahrung aus erster Hand.

▶ SUPPORT

- ▶ Wir sind das Backup für Ihre Linux-Administration: LPIC-2-Profis lösen im Heinlein CompetenceCall Notfälle, auf Wunsch auch in SLAs mit 24/7-Verfügbarkeiten.

▶ HOSTING

- ▶ Wenn Hosting kein Massengeschäft sein darf: Individuelles Business-Hosting mit perfekter Maintenance durch unsere Linux-Profis. Sicherheit und Verfügbarkeit werden bei uns groß geschrieben.

Und nun...

- ▶ Herzlichen Dank für Ihren Besuch...

wir wünschen interessante weitere Gespräche...

bleiben Sie doch noch ein bißchen...

- ▶ Auf jeden Fall aber: Beehren Sie uns bald wieder!