

kosteneffizienter Speicher  
a.k.a  
Basisspeicherserver  
[...]

Max-Planck-Institut für Gravitationsphysik  
(Albert-Einstein-Institut)  
Leibniz Universität  
Hannover

2019

# ÜBERSICHT

Motivation

Beispielspezifikation

Vor- und Nachteile

Zukunft

## DISCLAIMER:

WEDER UNSER VORSCHLAG NOCH EINE SEITE DIESES VORTRAGS ENTHIELT, ENTHÄLT ODER WIRD EINEN DATEISERVER ENTHALTEN, DER ALLE PROBLEME FÜR WENIG GELD LÖSEN WIRD!

## DISCLAIMER:

WEDER UNSER VORSCHLAG NOCH EINE SEITE DIESES VORTRAGS ENTHIELT, ENTHÄLT ODER WIRD EINEN DATEISERVER ENTHALTEN, DER ALLE PROBLEME FÜR WENIG GELD LÖSEN WIRD!

Einige Probleme können schon für wenig Geld gelöst werden.

# MOTIVATION

Warum überhaupt ein "Standard" Speichersystem?

# BAR VORTRAG OKTOBER 2015

- ▶ Auswertung von Speichersystem-Anträgen von 2011 – 2015
- ▶ günstigster BAR Antrag: €96/TB
- ▶ teuerster BAR Antrag: €2100/TB
- ▶ Selten: sehr spezielle Anforderungen
- ▶ Häufig: große Preisspanne bei ähnlichem Bedarf

(und ja, Preis pro TB ist/kann/darf nicht das **einzige** Kriterium sein, bleibt aber dennoch ein sehr gewichtiges!)

# MOTIVATION

Häufig ähnliche Anforderungen:

- ▶ Bandbreite, Kapazität, I/O-Raten
- ▶ Ausfallsicherheit, Redundanz, geringe Ausfallzeiten
- ▶ Wartbarkeit (wenige Personenstunden Aufwand)

Aber:

- ▶ Günstige Lösungen sind (zu) billig?
- ▶ Zu wenig Personal und/oder Fachkenntnisse
- ▶ Bestandssystem lief immer zuverlässig — warum jetzt wechseln?

## EIN BEISPIEL AUS DEN LETZTEN JAHREN

- ▶ etwa 1,2 PB Kapazität
- ▶ Gesamtkosten etwa €240 000 netto:
  - ▶ €83 000 für Hardware
  - ▶ €88 000 für Softwarelizenz
  - ▶ €66 000 für Wartung & Installation
  - ▶ (plus "Kleinteile")
- ▶ etwa €200/TB und 1,5 Jahre Projektstelle beim Hersteller bezahlt

Immer zu stellende Fragen:

- ▶ Benötigt man wirklich eine Enterprise Storage Lösung?
- ▶ Warum? Technisch? Administrativ? Bestandswissen?
- ▶ Geht man Vendor Lock-In ein? Exit Strategie?



## DAHER: IDEE EINES BASISPEICHERSYSTEMS

- ▶ günstig
- ▶ stabil
- ▶ flexibel
- ▶ erweiterbar/skalierbar
- ▶ einfach gehalten
- ▶ dokumentiert

“Weiche” Ziele:

- ▶ Sollte viele, typischer Anforderungsfälle erfüllen können
- ▶ “Angst” nehmen:  
Setup, Betrieb in Eigenverantwortung, ...

# ÜBERSICHT

Motivation

Beispielspezifikation

Vor- und Nachteile

Zukunft

# BASISSPEICHERSERVER (BEISPIEL)

## Supermicro basierender Server

- ▶ 24 x Toshiba 12 TB SATA HDD (7200 rpm): 288 TB
- ▶ 1 x NVMe 1,6 TB: read and/or write cache
- ▶ 2 x SSDs 240 GB: mirrored OS
- ▶ Intel E5-1650v4 6 Kerne @ 3,6 GHz
- ▶ 2 x 10 Gbit Ethernet
- ▶ 128 GB RAM
- ▶ Gesamtkosten unter €10 000: €9451
- ▶ €33/TB (Nettopreis/Bruttokapazität)

# SKALIERUNGEN

Mantra: Es ist nur ein ganz normaler Server

Daher:

<b>Problem</b>	<b>mögliche Lösung</b>
Speicherplatz	→ JBOD anhängen → Chassis mit 44 HDD/4U
lokales Processing zu langsam	→ anderes Board mit mehr oder schnelleren CPU Kernen
Mehr IOPS	→ mehr Spindeln, mehr NVMe Flash → all-flash mit U.2

# WICHTIGSTES DETAIL DES BEISPIELSERVERS?

# WICHTIGSTES DETAIL DES BEISPIELSERVERS:

## USB-Stick auf dem Mainboard

Basierend auf subjektiven und nicht repräsentativen Eindrücken während der Demosession

# ÜBERSICHT

Motivation

Beispielspezifikation

**Vor- und Nachteile**

Zukunft

# NACHTEILE/NÖTIGES WISSEN

- ▶ *lokales* Wissen wird benötigt!
- ▶ *keine* vorgefertigten GUIs! Interaktion mit CLI!
- ▶ Linux (Installation, Administration, Wartung)
- ▶ ZFS (Aufsetzen, Wartung, Monitoring)
- ▶ Hardwareüberwachung (SMART, Logs, ...)
- ▶ Hardwareaustausch
- ▶ Netzwerktechnik (Ethernet, IB, FC, ...)



# VORTEILE

- ▶ kostengünstig
  - ▶ mehr Speicher für das gleiche Geld
  - ▶ mehr Performance für das gleiche Geld
- ▶ keine komplizierten Verhandlungen mit Verkaufsprofis
- ▶ einfache Wartungsverträge
- ▶ flexibel aufsetzbar (IO Raten versus Kapazität)
- ▶ weniger "Fingerpointing" zwischen großen Firmen
- ▶ Wissen & Erfahrung  $\Rightarrow$  Vereinfachtes Debugging!
- ▶ Skalierbarkeit

# ÜBERSICHT

Motivation

Beispielspezifikation

Vor- und Nachteile

**Zukunft**

## WAS TUN WIR JETZT DAMIT?

- ▶ **keine** Lösung für alles
- ▶ aber hoffentlich Lösungsansatz für viele **Basis-Probleme!**
- ▶ out of the box:
  - ▶ NAS direkt im Netzwerk oder am Experiment — NFS, CIFS
  - ▶ SAN block storage (iSCSI, FC, IB, ...)
- ▶ Wofür gut geeignet?
  - ▶ scratch space im Netz
  - ▶ lokales Postprocessing von Projekten
  - ▶ direkt am Experiment/Gerät zum Zwischenspeichern
  - ▶ bei uns: iSCSI Block Storage für HSM-/home

## WAS WIR IN DEN RING WERFEN:

- ▶ Dokumentation dieses Basissystems (WIP)
- ▶ Pflege dieser Dokumentation (WIP)
- ▶ anfänglicher Support — **ABER** lokale Expertise unabdingbar (zumindest rudimentär)
- ▶ 3 Server zum Testen beschafft
- ▶ Schaffung eines grundlegenden Wissensnetzwerk

# INTERESSE?

Wer Interesse hat, bitte spricht uns/sprechen Sie uns an!

`atlas_admin@aei.mpg.de`

Kosten:

bis auf Fracht (etwa €25) initial keine

Dauer:

3 Monate testen, danach Rücksendung oder Kauf

## ABER:

Wir kennen viele MPI-lokale Probleme nicht!  
Daher benötigen wir Hilfe/Beispiele/Zuarbeit/"Success Stories"

- ▶ CIFS + LDAP/AD
- ▶ NFSv4 + Kerberos
- ▶ CephFS/BeeGFS/...Storage Bricks ↔ GWDG(?!)
- ▶ Nextcloud Appliance? Wird ohne lokale Expertise scheitern!

## FIO BEISPIELERGEBNISSE

Sets	Setup			Stream	random	Kapazität	
	d	+	p	write [MB/s]	read [MB/s]	read [IO/s]	[TB]
1	21	+	3	790	1900	252	201(???)
4	5	+	1	1390	2070	696	201
2	10	+	2	1090	1950	456	192
6	3	+	1	1300	1640	624	184
8	2	+	1	1840	1690	1104	168
3	6	+	2	1300	1910	624	180
4	4	+	2	1510	2460	768	168
12	1	+	1	1500	2640	1680	126

(je 96 parallele Zugriffe auf 1 GB Dateien)

**Danke für die Aufmerksamkeit!**

Bitte den Sitznachbarn wieder wecken.