ZIM
Zentrum für Informations- und Medientechnologie

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

GPFS - Clusterfilesystems

Philipp Helo Rehs
Dr. *rer. nat.* Stephan Raub

Kompetenzzentrum für
wissenschaftliches Rechnen und Speichern
(WiRe/S)

Universitätsstr. 1
Raum 25.41.00.51
40225 Düsseldorf

+49-211-8113-911

stephan.raub@hhu.de

http://www.xing.com/profile/stephan_raub

http://de-de.facebook.com/stephan.raub.7

ZIM Zentrum für Informations- und Medientechnologie

Cluster-Filesysteme

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Klassisches NFS

- Alle Clients teilen sich den Server
- Keine Redundanz beim Server
- Übernahme bei Ausfall schwierig
  - Locks gehen verloren

ZIM Zentrum für Informations- und Medientechnologie

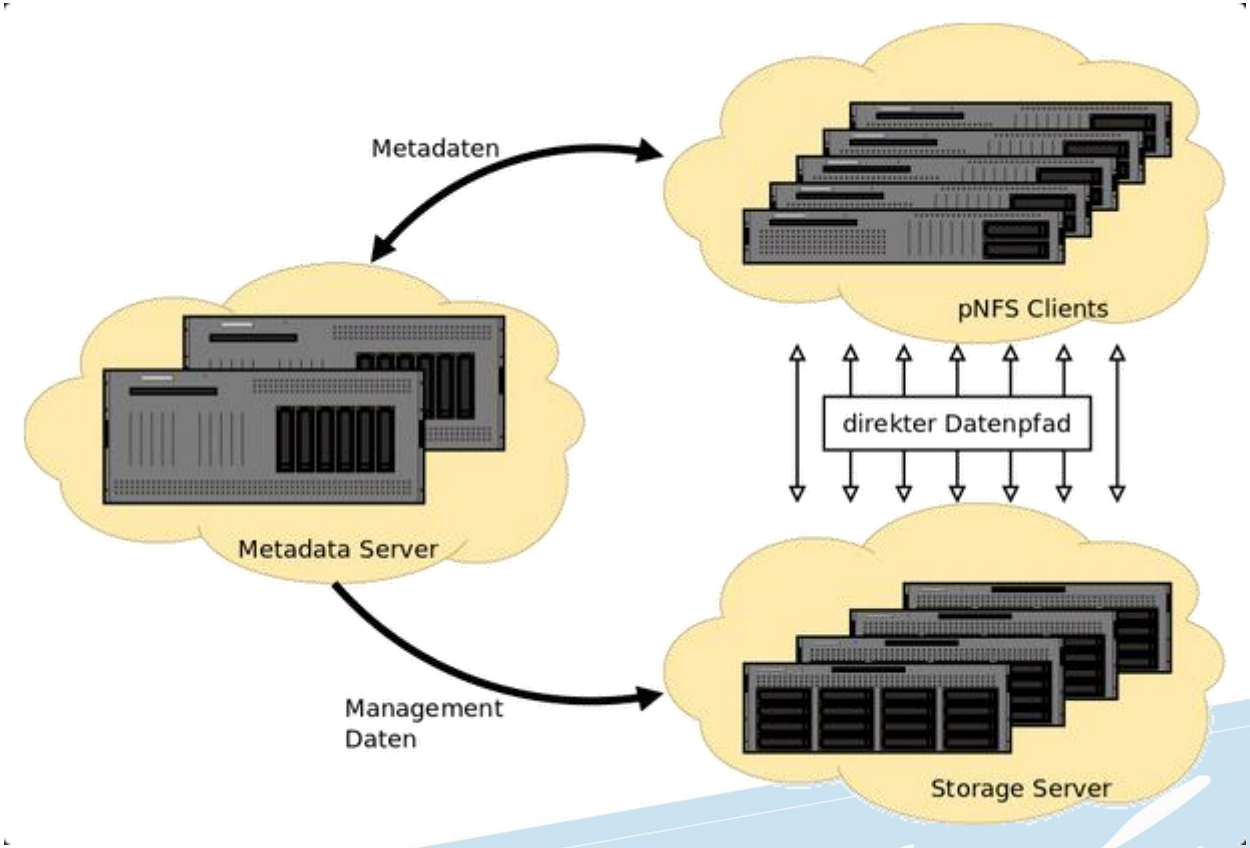
Cluster-Filesysteme

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Klassisches NFS

- Performantes Storage / JBOD / RAID steigert die Bandbreite
- Single-Point-Of-Failure bleibt bestehen

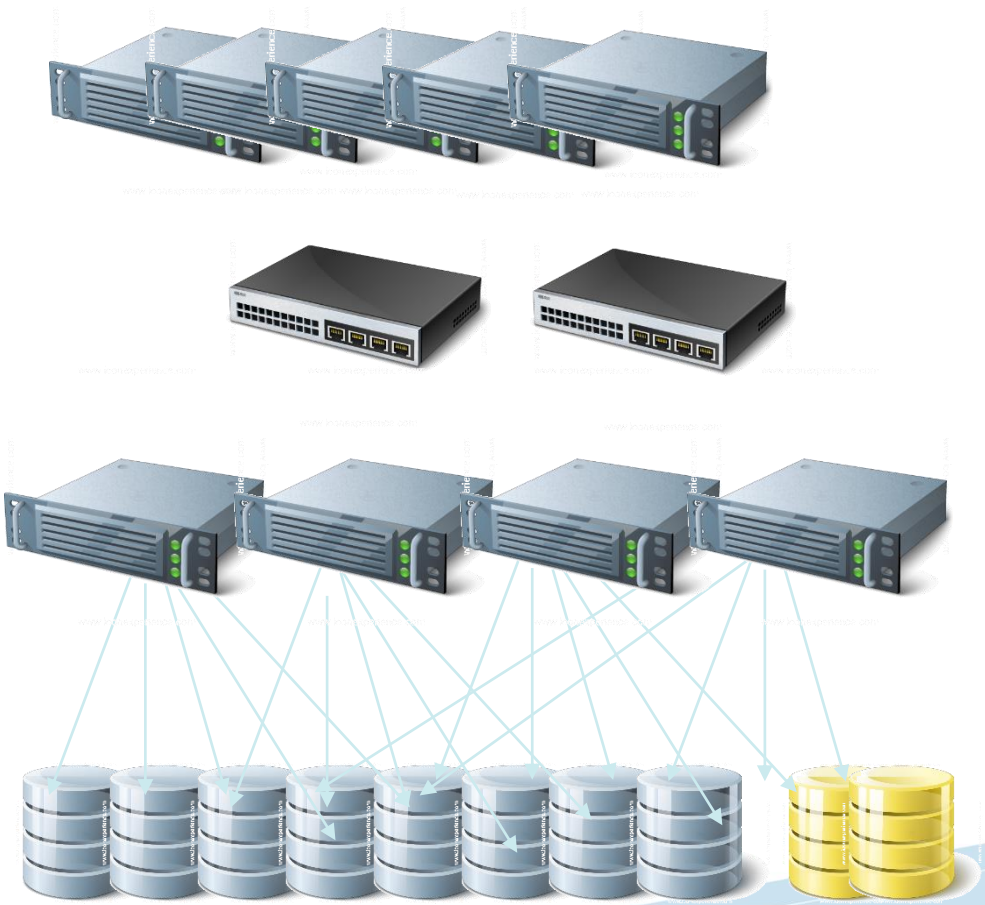
Ja, es gibt auch sowas wie pNFS...



ZIM
Zentrum für Informations- und Medientechnologie

Cluster-Filesysteme

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF



ZIM Zentrum für Informations- und Medientechnologie

Cluster-Filesysteme

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

- Parallelität auf allen Ebenen
  - Block-Storage, Server, Netzwerk, Software
- Clients kennen jeden Server
- Server haben Multipath zum Storage (Block / LUN)
- Direkter Zugriff von Client auf den Block-Storage
- Voll Paralleles Lesen und Schreiben bis auf Block Level
- Verteilte Metadaten, es gibt keinen zentralen Metadaten-Server
- Quorum um Konsistenz zu gewährleisten
- Automatische Behandlung von Komponenten-Ausfällen

ZIM Zentrum für Informations- und Medientechnologie

Abgrenzung CEPH

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Ceph Clusters in CERN IT

- Standard Ceph node architecture
  - 16-core Xeon / 64-128GB RAM
  - 24x 6TB HDDs
  - 4x 240GB SSDs (journal/rocksdb)

CERN | ceph

ZIM Zentrum für Informations- und Medientechnologie

Vergleich GPFS

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

- Beispiel-Aufbau auf Grund von IBM ESS
  - Server für Metadaten und Zugriff
    - 2 Power 8 / 128 GB Memory
    - 2 – 4 x Infiniband EDR / 100GbE-Ethernet
  - JBODS mit 84 Platten zum skalieren

- Bis zu 502 Nearline-SAS Festplatten an einem Controller
  - 6 x 5 U für Festplatten
  - 2 x 2 U für Server
  - 5 PB Raw
  - 2 – 4 x Infiniband EDR

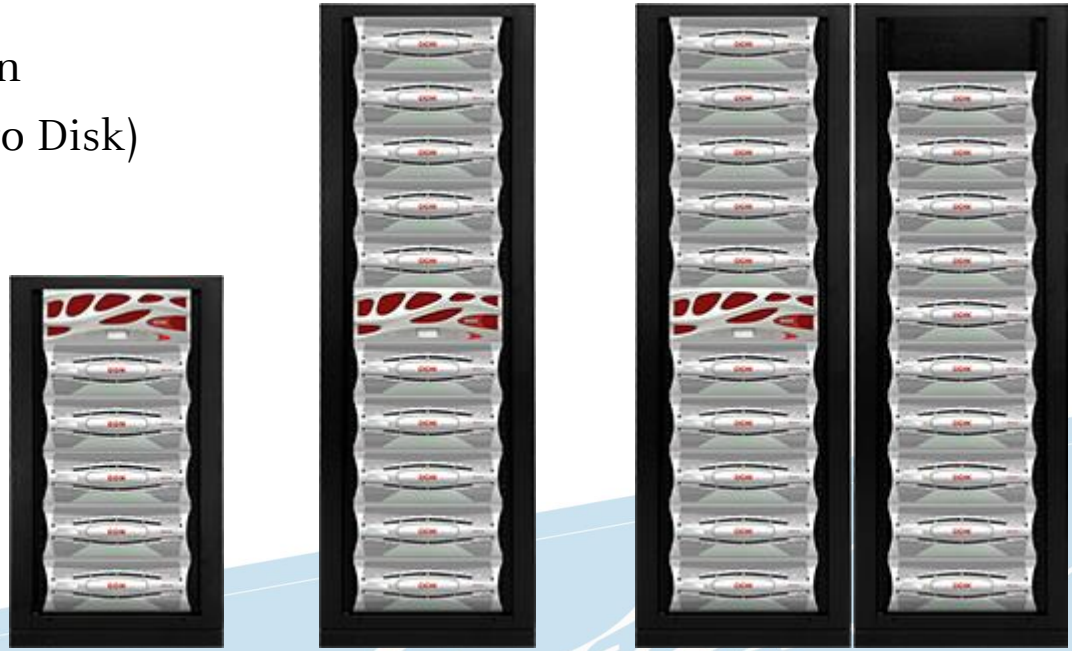
ZIM Zentrum für Informations- und Medientechnologie

Vergleich GPFS DDN HHU

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

- DDN GS14KXE
  - 2 embedded Controller
  - 4 x Infiniband EDR
  - 5 x 4 U für Chassis
  - 4 U für Controller + SSD
  - 5 PB Raw
  - 480 HDDs in RAID-Verbünden
  - 8 SSDs für Metadaten (2TB pro Disk)

IO500-Score: 13,7



522 Disk / 24U 972 Disk / 44U 1872 Disk / 84U

ZIM
Zentrum für Informations- und Medientechnologie

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

raub@uni-duesseldorf.de