Kostengünstige Virtuelle Speicher-Pools im Petabyte-Bereich

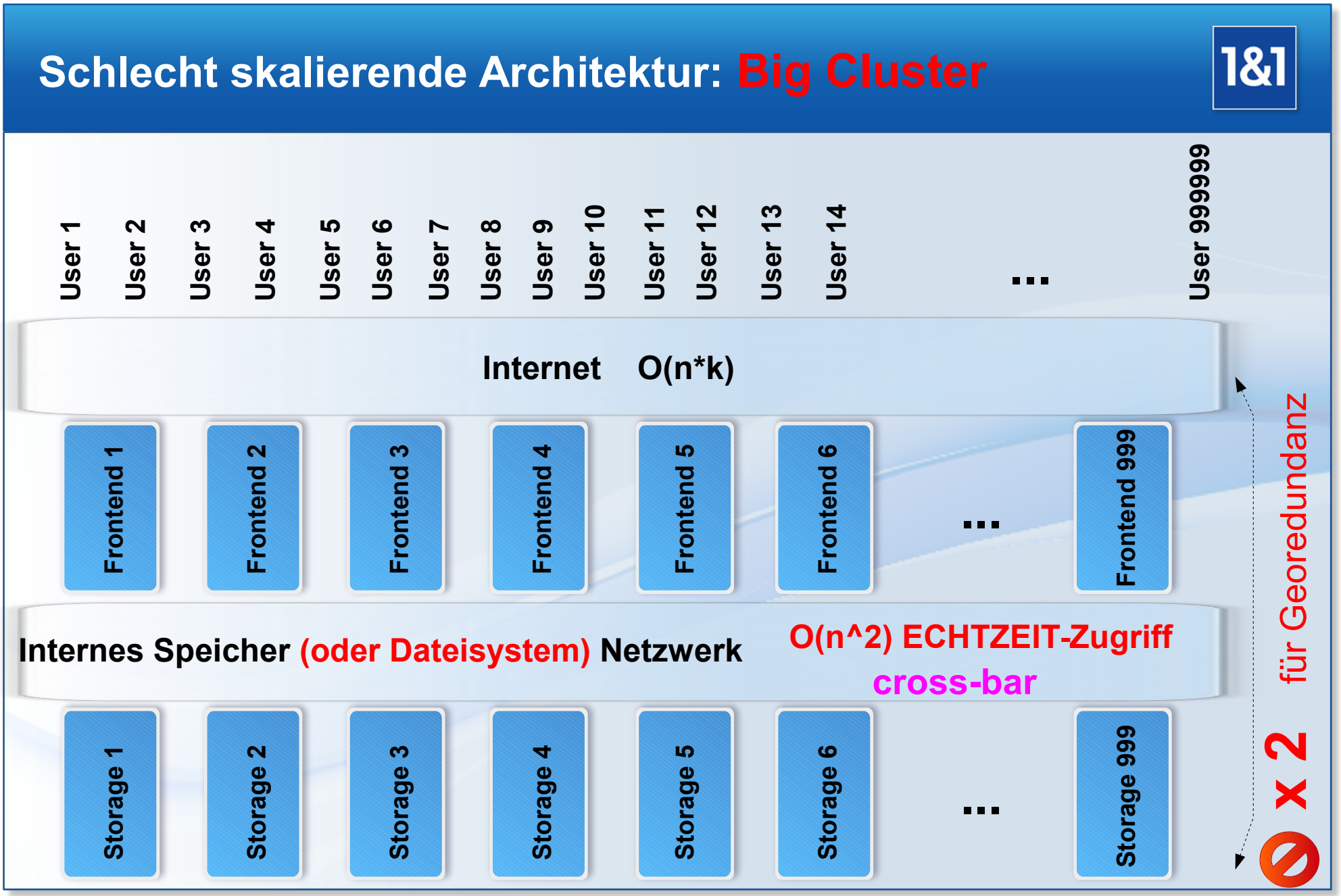
mithilfe von MARS

GUUG 2017 Vortrag von Thomas Schöbel-Theuer

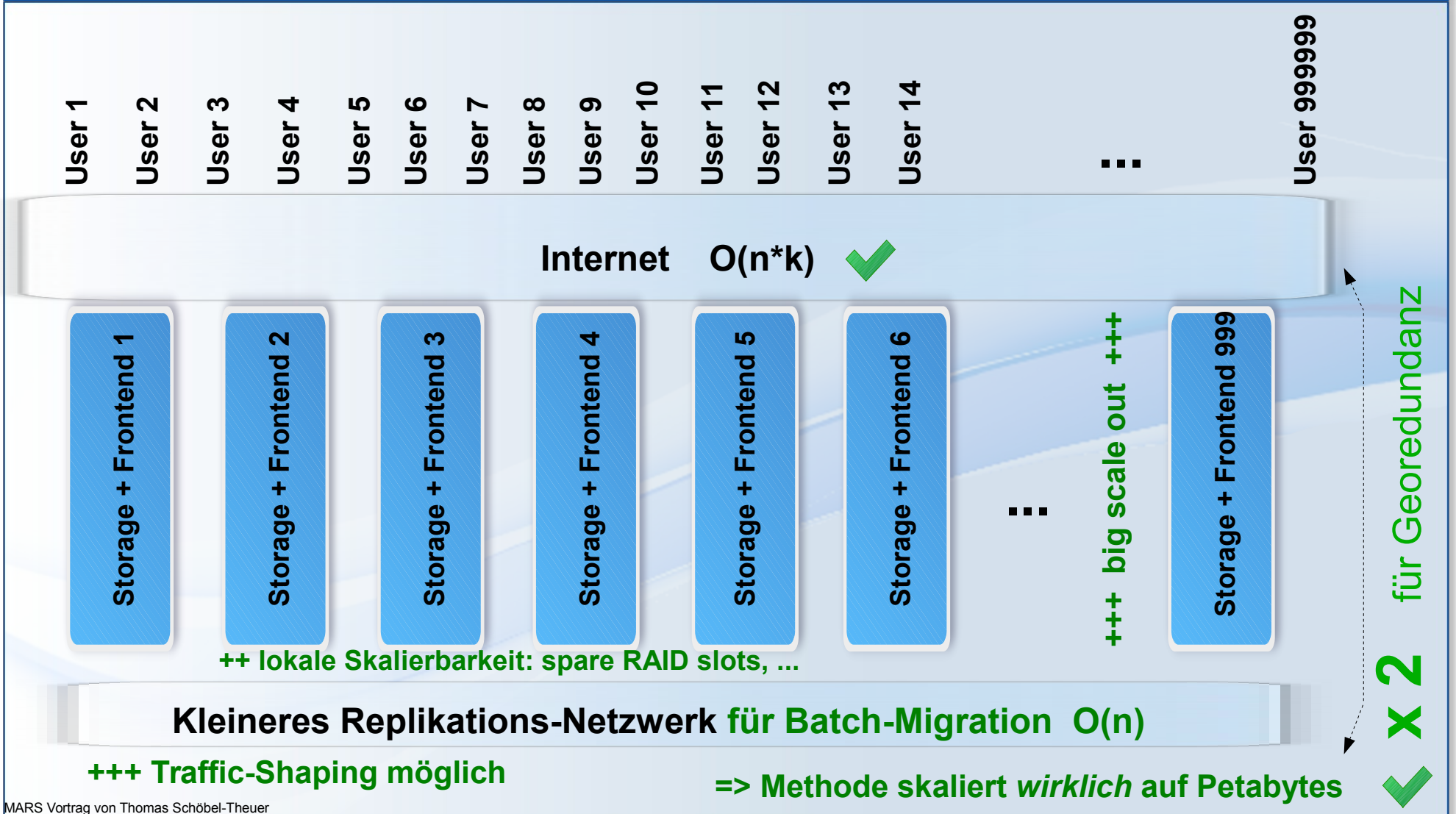
Virtuelle Speicher-Pools: Agenda

- Skalierungs-Eigenschaften von Speicher-Architekturen
- Motivation: Kosten
- Flexibles MARS Sharding + Cluster-on-Demand
- Last-Balancierung mittels Hintergrund-Daten-Migration
- Aktueller Status / Zukunfts-Pläne

Schlecht skalierende Architektur: Big Cluster



Gut skalierende Architektur: Sharding



Kosten (nicht-georedundante Variante)

- Big Cluster:
  - Typisch ≈RAID-10 zur Ausfall-Kompensation
- Platten: > 200%
- Zusätzliche CPUs und RAM für Storage-Knoten
- Zusätzlicher Strom-Verbrauch
- Zusätzliche HE

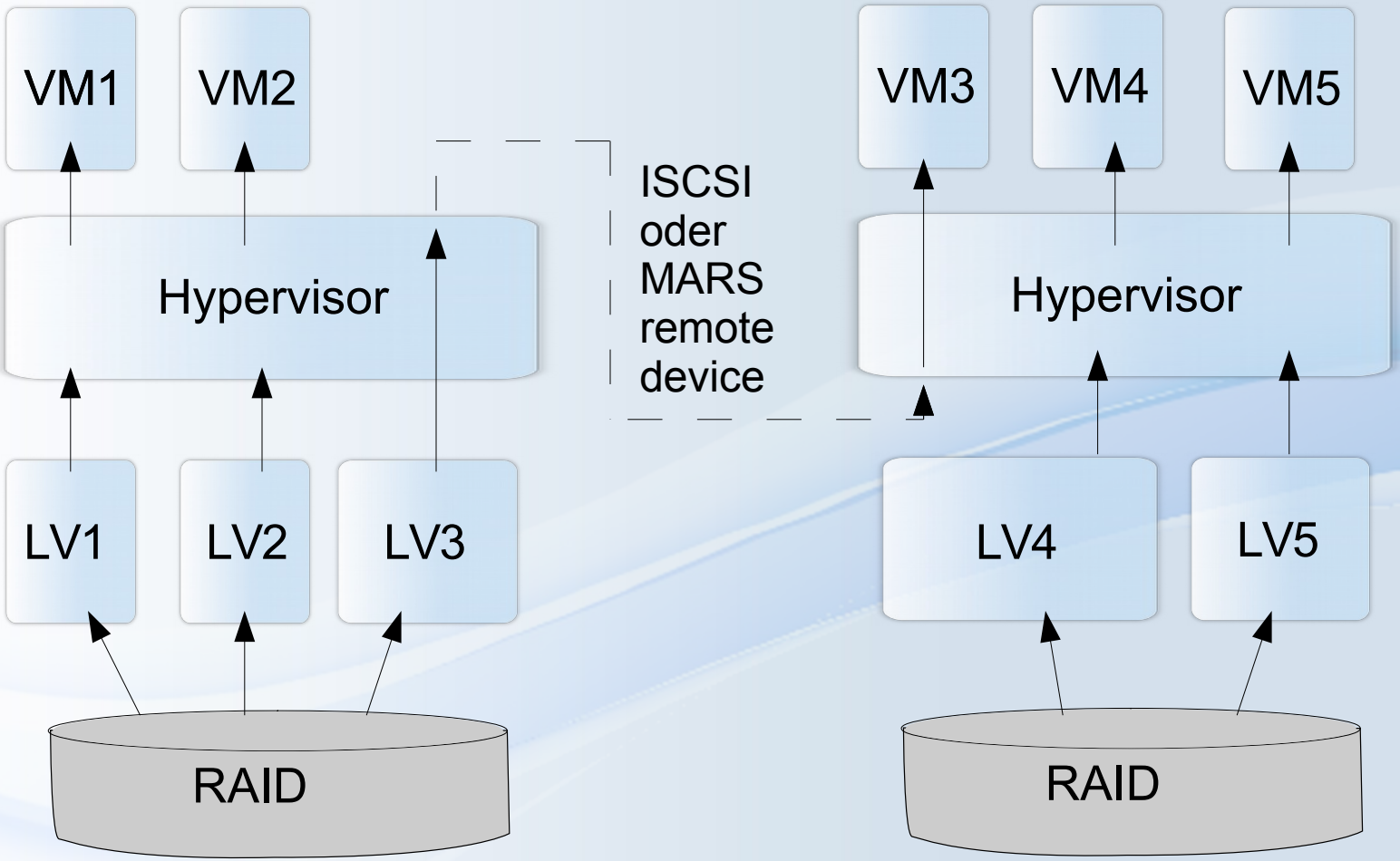
- Sharding:
  - oft reicht lokales RAID-6
- Platten: < 120%
- Hardware RAID Controller mit BBU-Cache auf 1 Karte
- Weniger Strom
- Weniger HE

Kosten (georedundant => mit Langstrecke möglich)

- Big Cluster:
  - 2x ≈RAID-10 zur Ausfall- Kompensation
    (kleiner geht nicht wegen längeren RZ-Ausfall-Szenarien)
- Platten: > 400%
- Zusätzliche CPUs und RAM für Storage-Knoten
- Zusätzlicher Strom-Verbrauch
- Zusätzliche HE

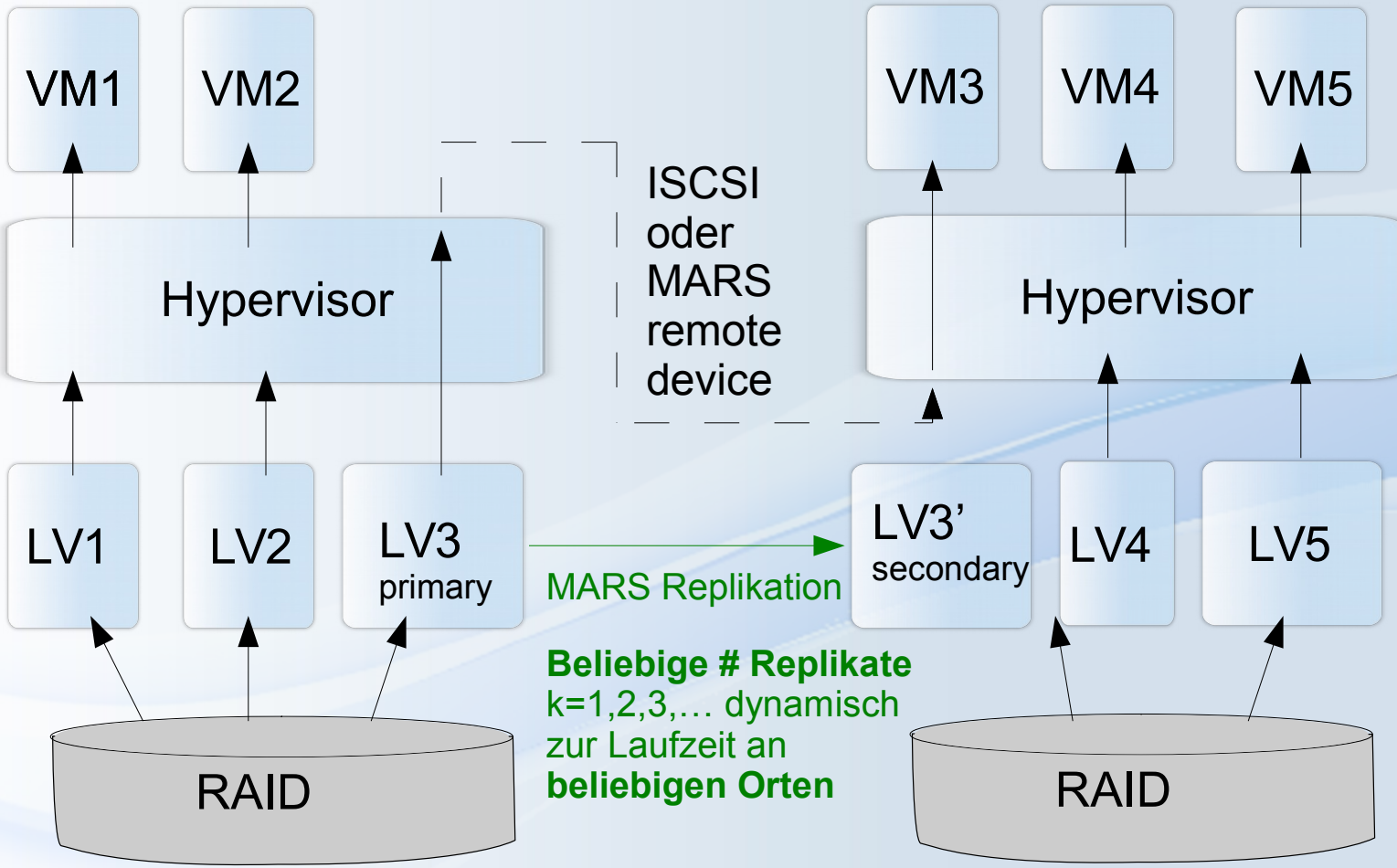
- Sharding:
  - 2 x lokales RAID-6
- Platten: < 240%
- Hardware RAID Controller mit BBU-Cache
- Weniger Strom
- Weniger HE

Flexibles MARS Sharding + Cluster-on-Demand



Die Hypervisoren können sowohl in Client- als auch Server-Rollen und bevorzugt auch **lokal** arbeiten

Flexible MARS Hintergrund-Migration



Beliebige # Replikate
k=1,2,3,… dynamisch
zur Laufzeit an
beliebigen Orten

=> ein Hypervisor kann gleichzeitig Quelle oder Ziel verschiedener LV-Replikationen sein

MARS aktueller Status

- MARS Sourcecode unter GPL + Docu:
  `github.com/schoebel/mars`
  `mars-manual.pdf` ~ 100 Seiten
- mars0.1stable produktiv auf Kundendaten seit 02/2014
- Rückgrat des 1&1 Geo-Redundanz Features
- MARS Status Feb 2017:
  > 2000 Server (Shared Hosting + Datenbanken)

  > 2x8 Petabyte brutto

  ~ 10 Milliarden inodes in > 3000 xfs Instanzen

  > 25 Millionen Betriebsstunden
- Neues internes Efficiency Projekt
  - Höhere interne LXC Container Verdichtung pro 1 Hypervisor
  - Neuer öffentlicher Branch mars0.1b mit vielen neuen Features, z.B. massen-skalierbares Clustering, Socket Bundling, Remote Device, etc
  - mars0.1b aktuell im ALPHA-Stadium

MARS Zukunfts-Pläne

Automatische Last-Balancierung	TBD Separate Implementation oder libvirt / Openstack Plugins … ?
Virtuelle LVM-artige Speicher + VM Pools	WIP 1&1 clustermanager `cm3` und/oder libvirt Plugin … ?
Physisch ge-shardete Pools	Erledigt: MARS anstatt DRBD

Kollaboration gesucht

=> Gelegenheiten für andere OpenSource-Projekte!