STORAGE DATA ANALYTICS

Technology Kick Off 2018

BEST PRACTICE | KNOW-HOW | EXPERTENTREFFEN | NETWORKING

Herzlich Willkommen

Benchmarking und Key Performance Indikatoren von SSD-Storage-Systemen

Partner

HUAWEI

INFONIQA

Andreas Zallmann
CEO, In&Out AG

www.inout.ch
andreas.zallmann@inout.ch

STORAGE DATA ANALYTICS Fachkongress 2018

Leica Geosystems IN&OUT

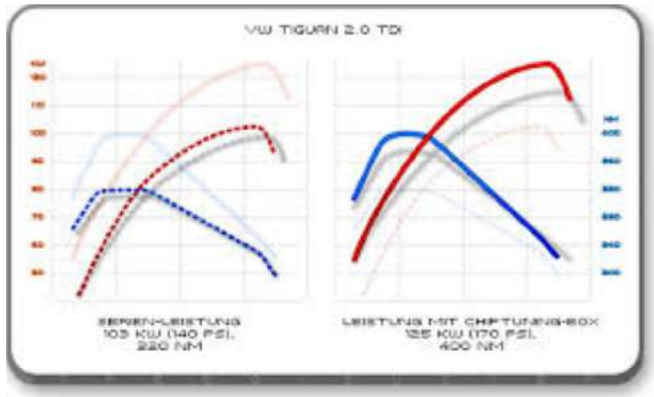
Inhaltsverzeichnis

Storage Benchmarks
Messen heisst Wissen

- Realitätsnahe, vergleichbare und reproduzierbare Kennzahlen in Ihrer Umgebung

- Maximal-Performance von Storagesystemen
  - Durchsatz (maximale Bandbreite)
  - Geschwindigkeit (Latenz)
  - Verifikation und Einordnung von Herstellerangaben
  - Proof of Concepts bei Ausschreibungen
  - Auswirkungen von Setups (z.B. Raid Level), Features (Dedup, Compression, Thin Provisioning), Datenschutz (Snapshots, Remote Mirrors, etc).

- Nutzbare Performance von Ihren IT-Plattformen
  - Identifikation von Engpässen
  - Optimierung und Tuning der Systemperformance
  - Sicherstellen der «Operational Readiness»



STORAGE DATA ANALYTICS
Technology Kick Off 2018

Leica Geosystems IN&OUT

Storage Benchmarks
IOgen®

PARAMETRIER- & AUTOMATISIERBAR

- Testreihen und Testpakete in beliebiger Abfolge voll automatisierbar

- Parallelität / Auto-Scale
- Read / Write / Mixed
- Sequential / Random / Mixed
- Frontend / Backend
- Blockgrössen
- Read und Write Cache Hit Ratio
- Pattern (Random, Pseudorandom, Change, Constant, Zero) für Featues wie Thin Provisioning / Dedup / Compression
- Synchron / Asynchron
- Synchronisierte Benchmarks von verschiedenen Servern

EINSATZ

- DWH, Big-Data, Cloud, OLTP, etc.
- Highend, Midrange, All-Flash, SDS, etc.
- Block Fibre-Channel und File Ethernet (IP)
- Verfügbar für alle gängigen Betriebssysteme (Windows, Linux, AIX, Solaris, HP-UX, zLinux, etc.)

PERFORMANCE KENNZAHLEN

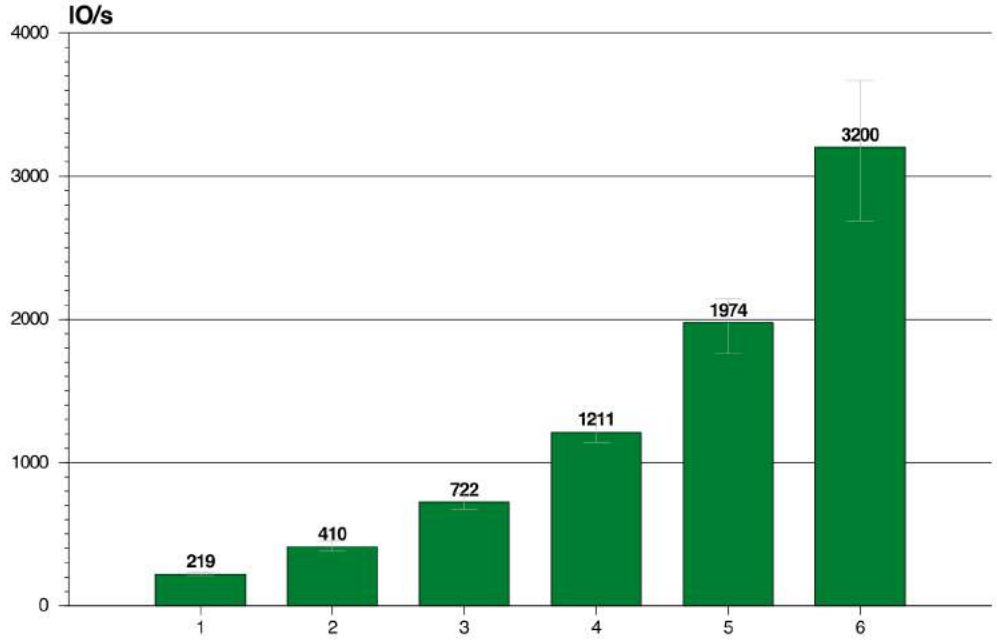
- IOPS bzw. MB/s
- IO-Latency (µs), auch Verteilung
- CPU Load
- Testgüte

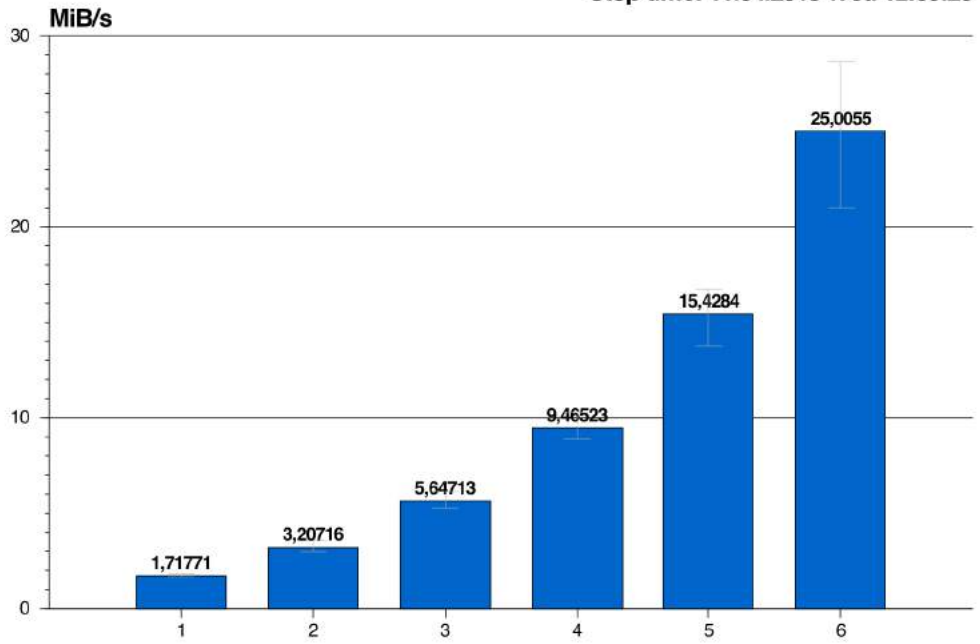
ERGEBNISSE

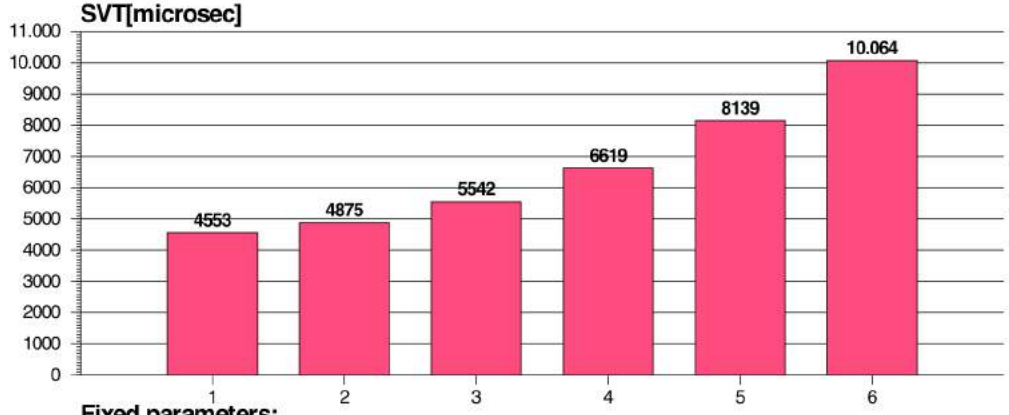
- Automatische Grafiken
- Excel Exports

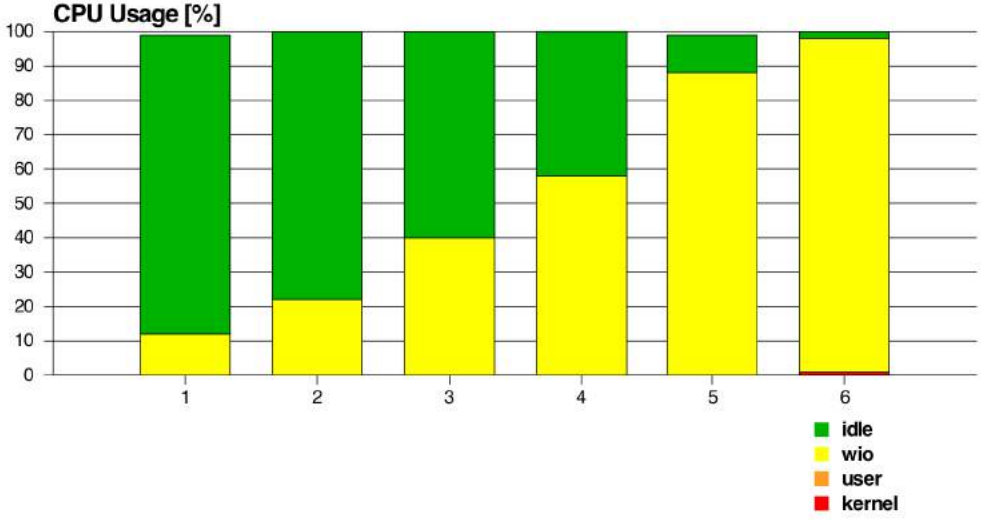
Result "'Backend 8K Random Read"'

Hostname: andreass-imac-pro-1.home
Start time: 11.04.2018 Wed 12:02:19
Stop time: 11.04.2018 Wed 12:08:23









Fixed parameters:

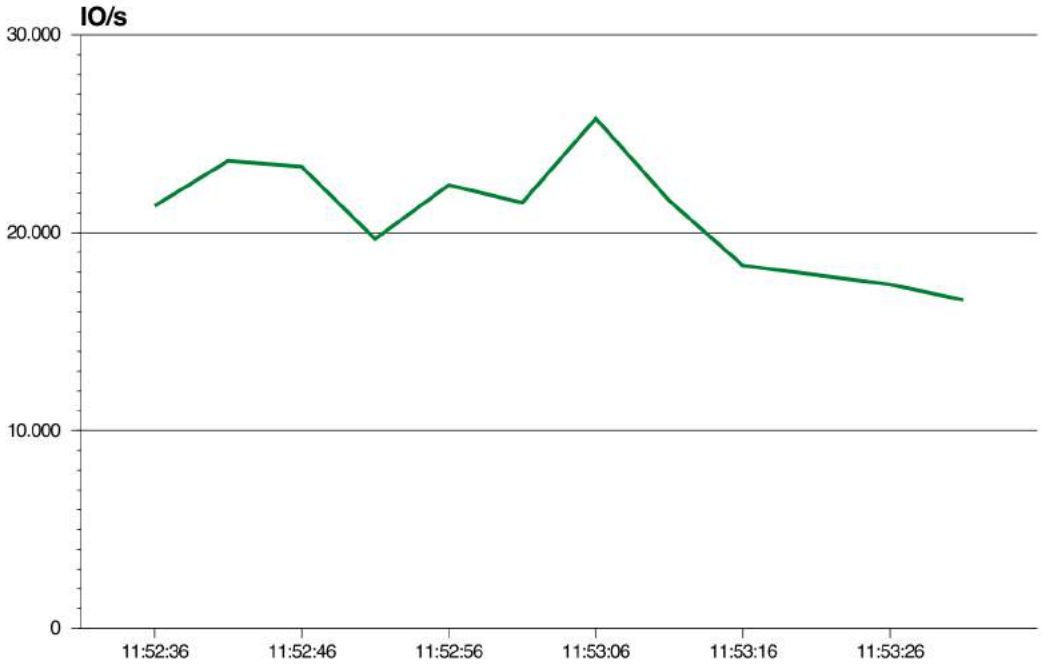
Device /dev/sdb [100.000 GiB-25.000%],
Device /dev/sdc [101.000 GiB-25.000%],
Device /dev/sdd [102.000 GiB-25.000%],
Device /dev/sde [103.000 GiB-25.000%],
Write 0% / Read 100%, Random 100% / Sequential 0%, CHANGE 100%,
BS 8.000 KiB [100.000%], CB 1024, CR 0%, CW 0%, Sync IO, Seq IO 1000,
Total test time 60 sec, Interval time 5 sec,

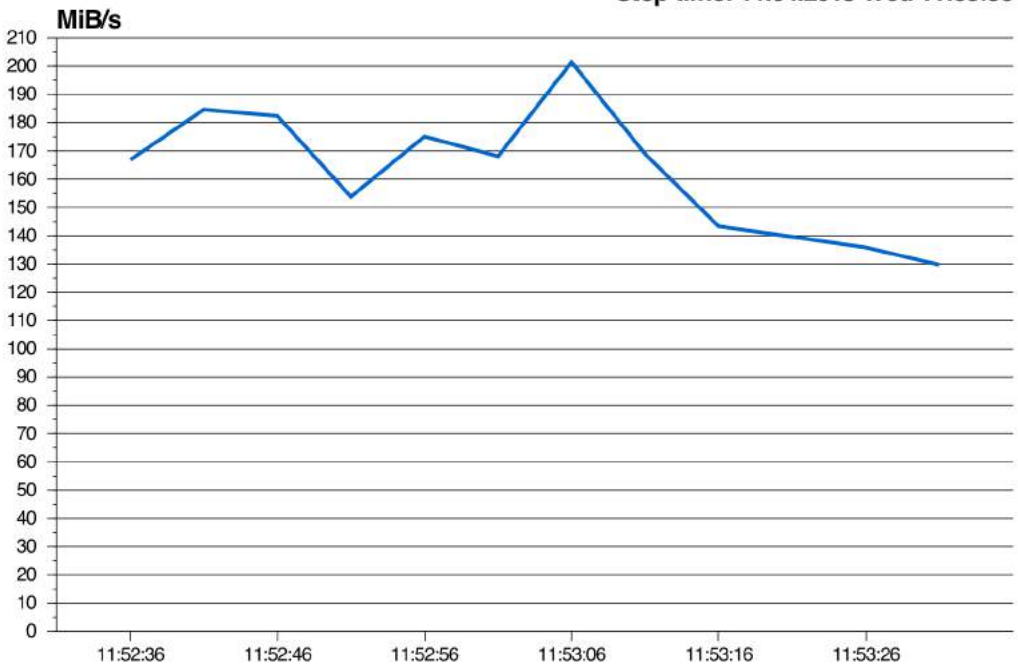
Variable parameters:

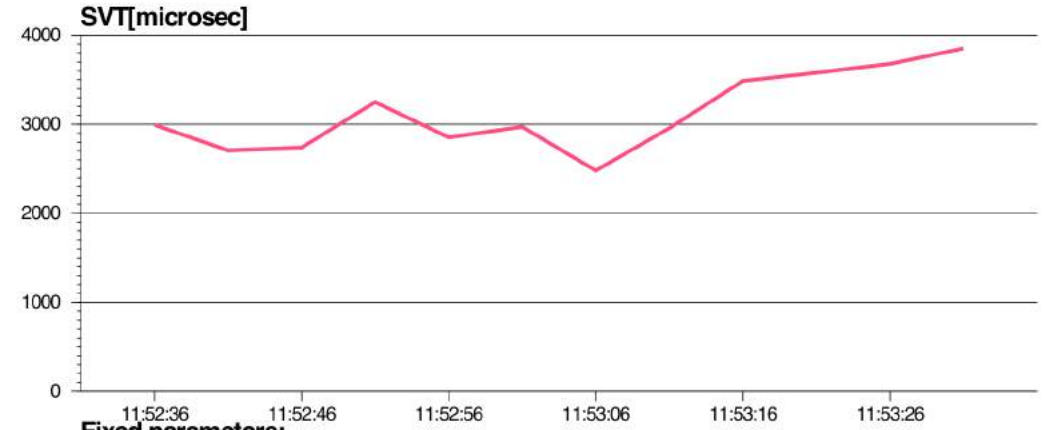
Test 1: Parallel processes 1,
Test 2: Parallel processes 2,
Test 3: Parallel processes 4,
Test 4: Parallel processes 8,
Test 5: Parallel processes 16,
Test 6: Parallel processes 32,

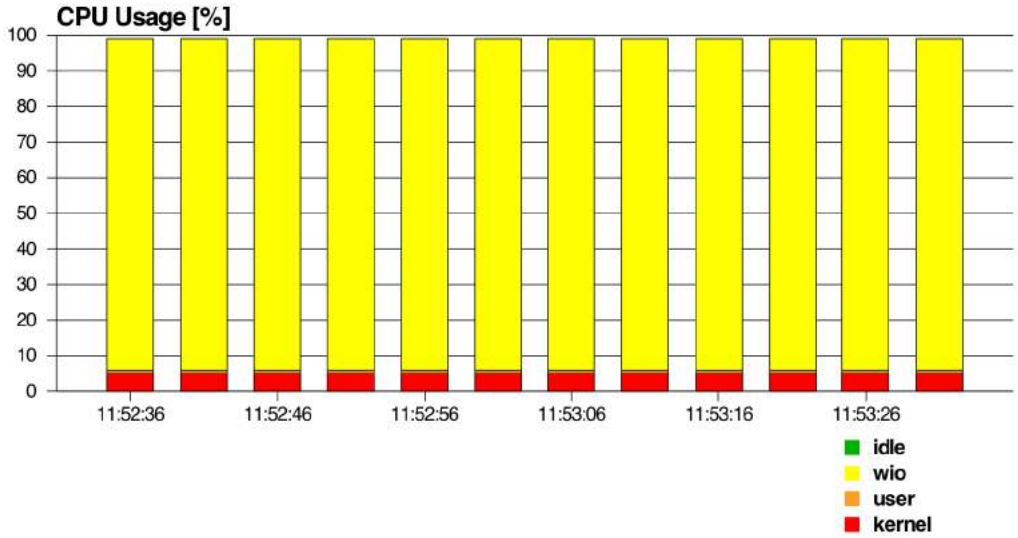
Result 'Frontend8KRandomWrite-7'

Hostname: andreass-imac-pro-1.home
Start time: 11.04.2018 Wed 11:52:36
Stop time: 11.04.2018 Wed 11:53:36







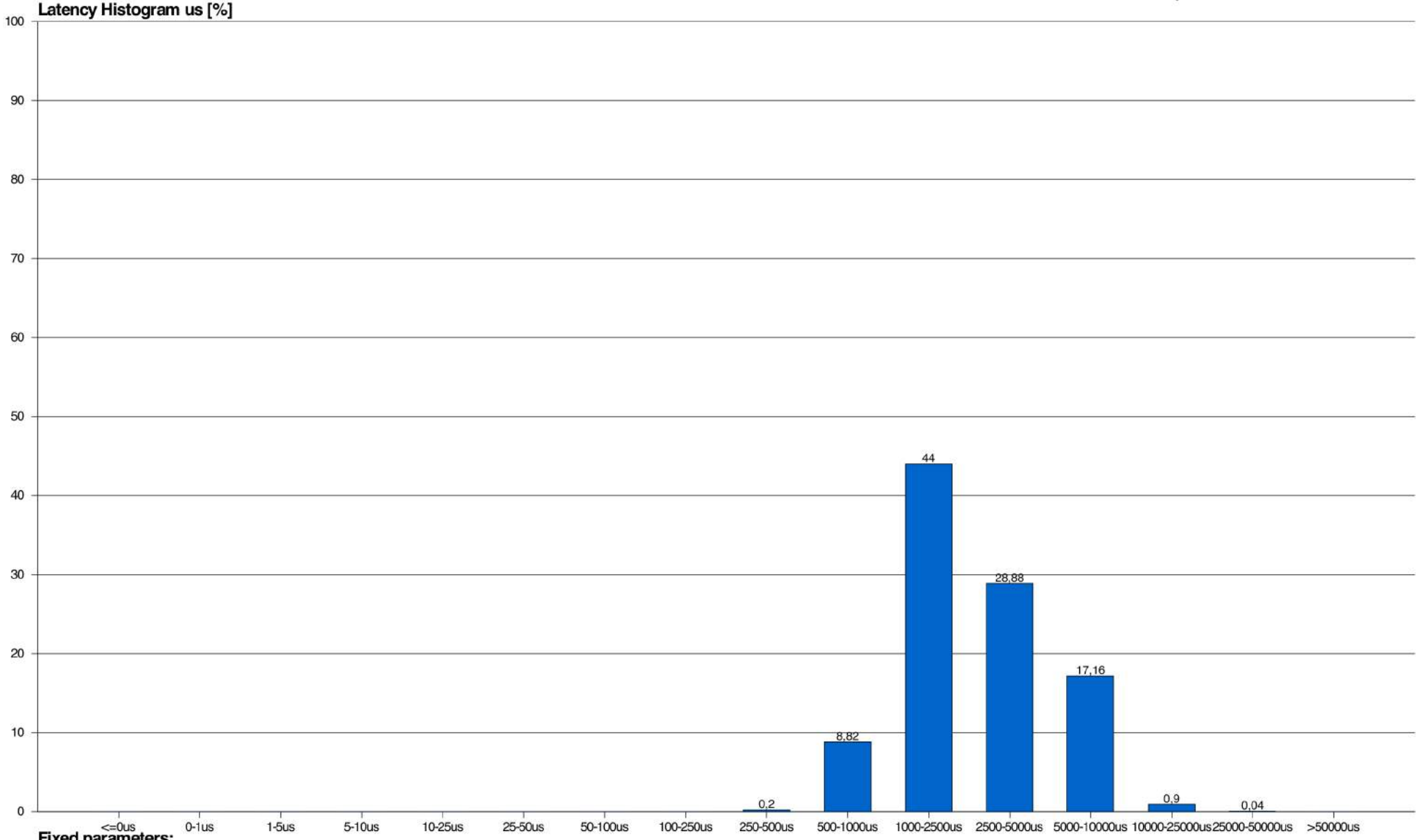


Fixed parameters:

Device /dev/sdb [100.000 GiB-25.000%],
Device /dev/sdc [101.000 GiB-25.000%],
Device /dev/sdd [102.000 GiB-25.000%],
Device /dev/sde [103.000 GiB-25.000%],
Write 100% / Read 0%, Random 100% / Sequential 0%, CHANGE 100%,
BS 8.000 KiB [100.000%], CB 102400, CR 100%, CW 100%,
Sync IO, Seq IO 1000, Parallel processes 64, Total test time 60 sec, Interval time 5 sec

Result 'Frontend8KRandomWrite-7'

Hostname: andreass-imac-pro-1.home
Start time: 11.04.2018 Wed 11:52:36
Stop time: 11.04.2018 Wed 11:53:36



Fixed parameters:

Device /dev/sdb [100.000 GiB-25.000%],
Device /dev/sdc [101.000 GiB-25.000%],
Device /dev/sdd [102.000 GiB-25.000%],
Device /dev/sde [103.000 GiB-25.000%],
Write 100% / Read 0%, Random 100% / Sequential 0%, CHANGE 100%,
BS 8.000 KiB [100.000%], CB 102400, CR 100%, CW 100%,
Sync IO, Seq IO 1000, Parallel processes 64, Total test time 60 sec, Interval time 5 sec

Benchmarks & Assessments
Referenzen (Auszug)

Benchmarks & Assessments

Storagesysteme (Auszug aktuelle Systeme)

NetApp

NetApp FAS 9000 & 8020

IBM

FlashSystem 840

HITACHI DATA SYSTEMS

HDS G1000

PURESTORAGE

FlashSystem //M20

DELL EMC

VMAX 450F

Hewlett Packard Enterprise

3PAR 20800

HUAWEI INFONIQA

Dorado 5000 V3

INFONIQA HUAWEI

OceanStor 6800 V5

Inhaltsverzeichnis

Key Performance Indikatoren Storage

RANDOM IO

- Relativ kleine Blöcke 4, 8, 16 KB
- Z.B. DB OLTP, Filezugriff, Indexzugriff, Metadaten, etc.
- Entscheidend ist Latenz (µs) und IOPS

SEQUENTIAL IO

- Grössere IOPS (256 KB - 1 MB)
- Lesen von Files, DB Full Table Scans
- Storage kann Read-Ahead durchführen, Full Row Raid Calculation
- Weniger aufwändig als Random
- Entscheidend ist Latenz (µs) und MB/s

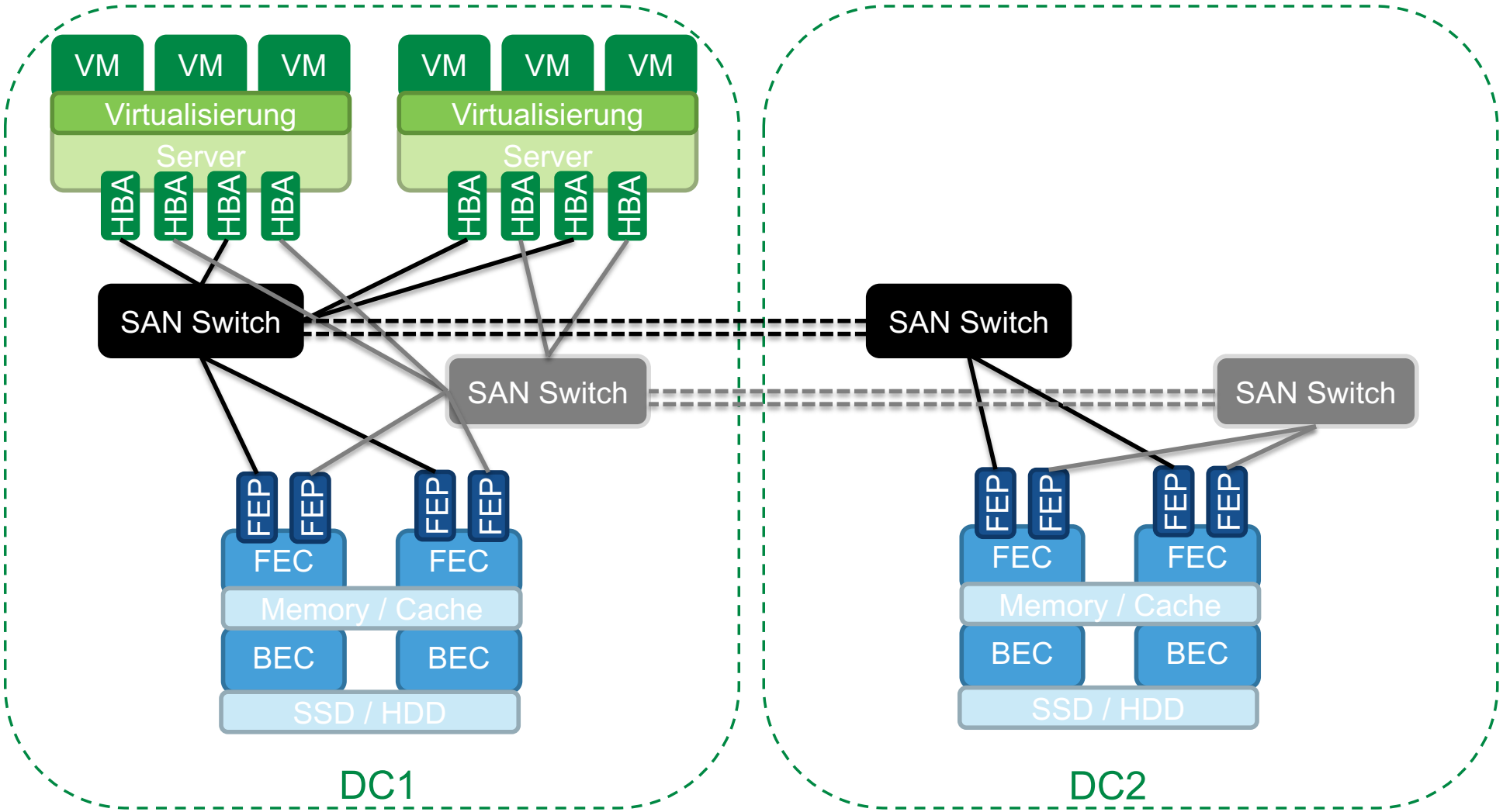
TESTARTEN (BEST CASE, WORST CASE)

- Frontend Read = Block wird im Storage Cache getroffen
- Frontend Write = Kein Downstaging auf Disks, beschränktes Working Set
  Keine Raid Calculations, kein Downstaging
- Backend Read = Block wird nicht im Cache getroffen (sehr häufig wenn serverseitiges Caching), lesen von Disks / SSD
- Backend Write = Cache gefüllt, Downstaging, Raid Calculation notwendig, Downstaging Geschwindigkeit limitierend

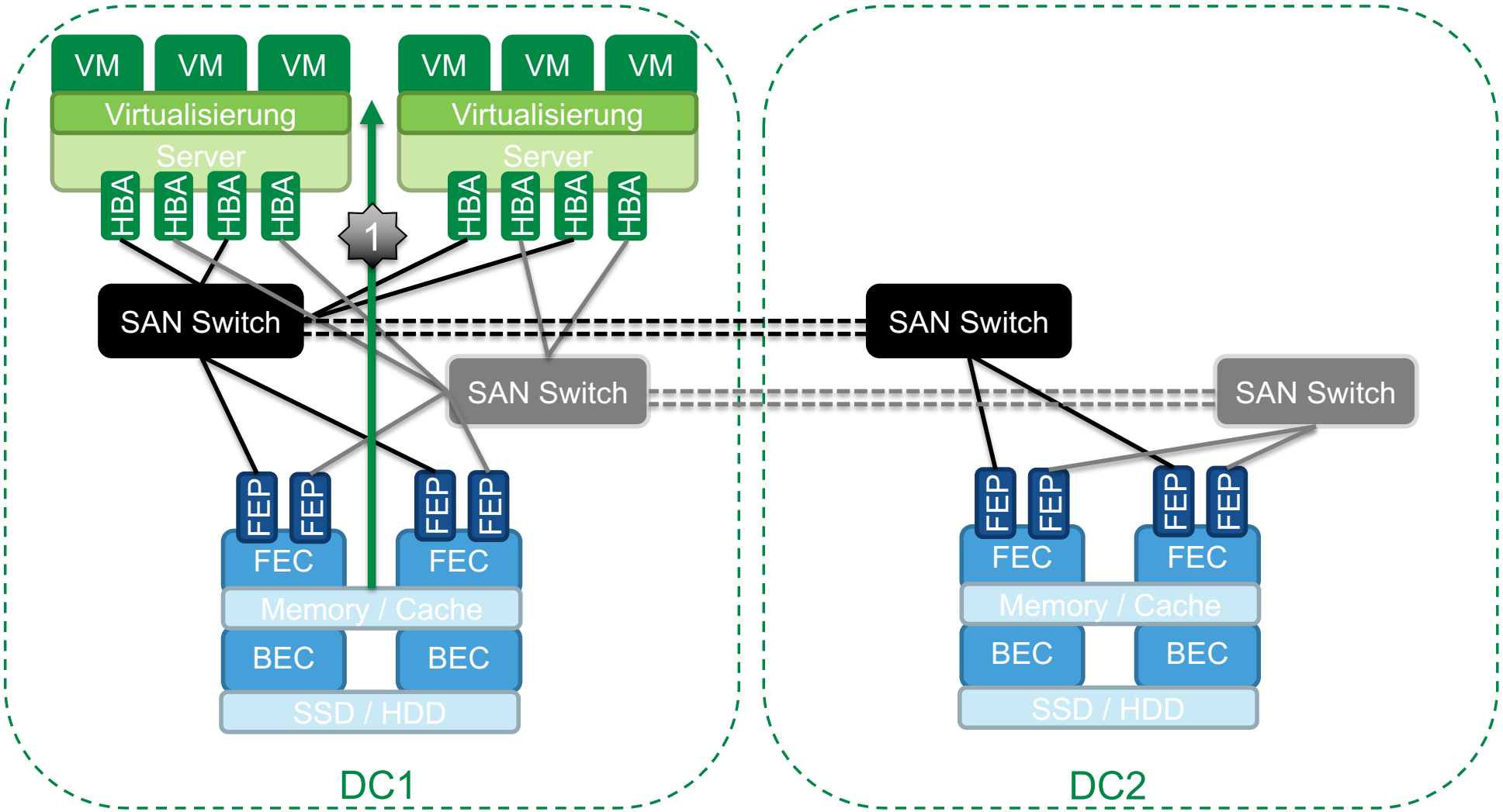
SPEED: PERFORMANCE VON 1 PROZESS

THROUGHPUT: MAXIMALER DURCHSATZ

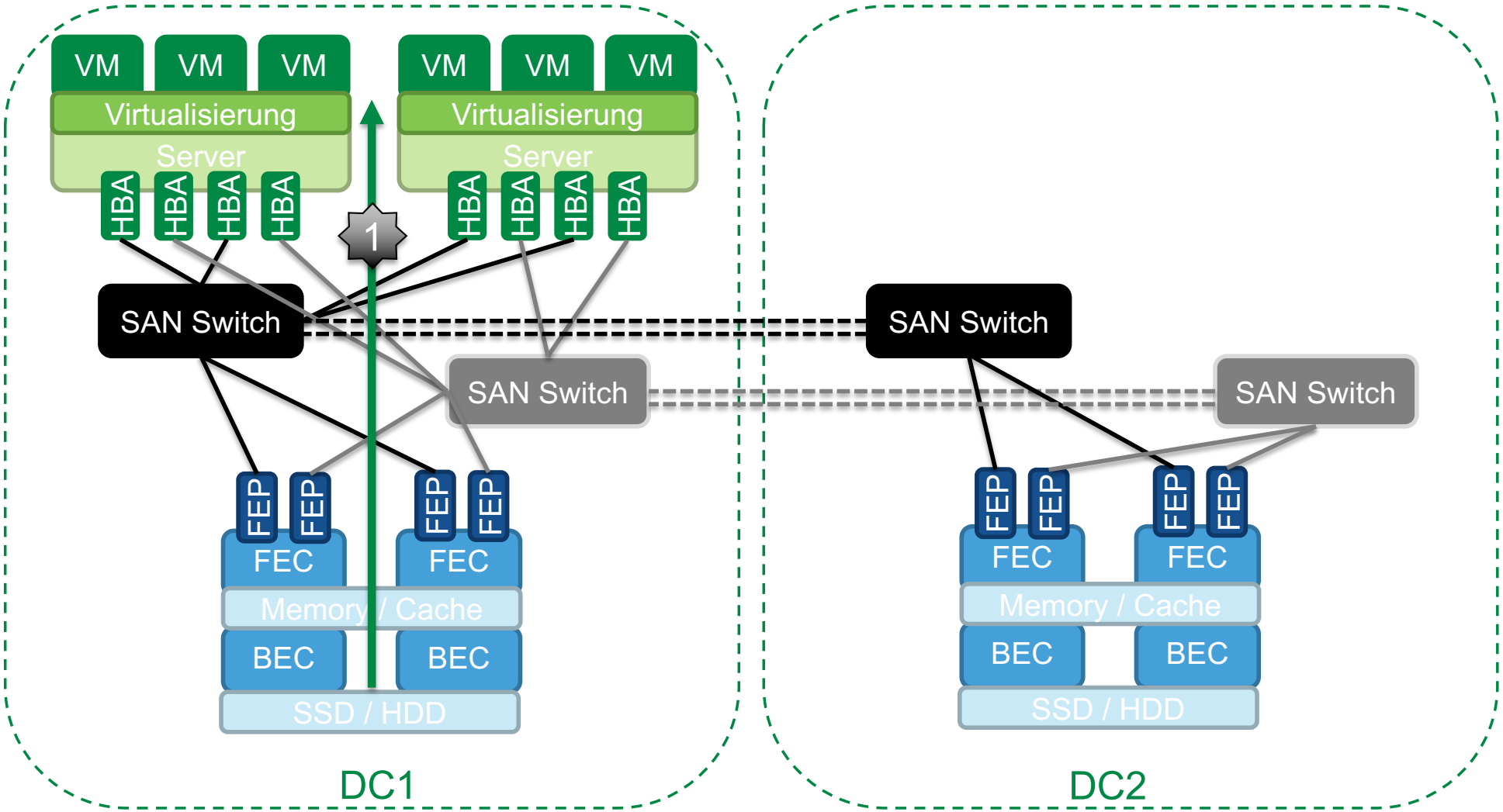
Frontend / Backend



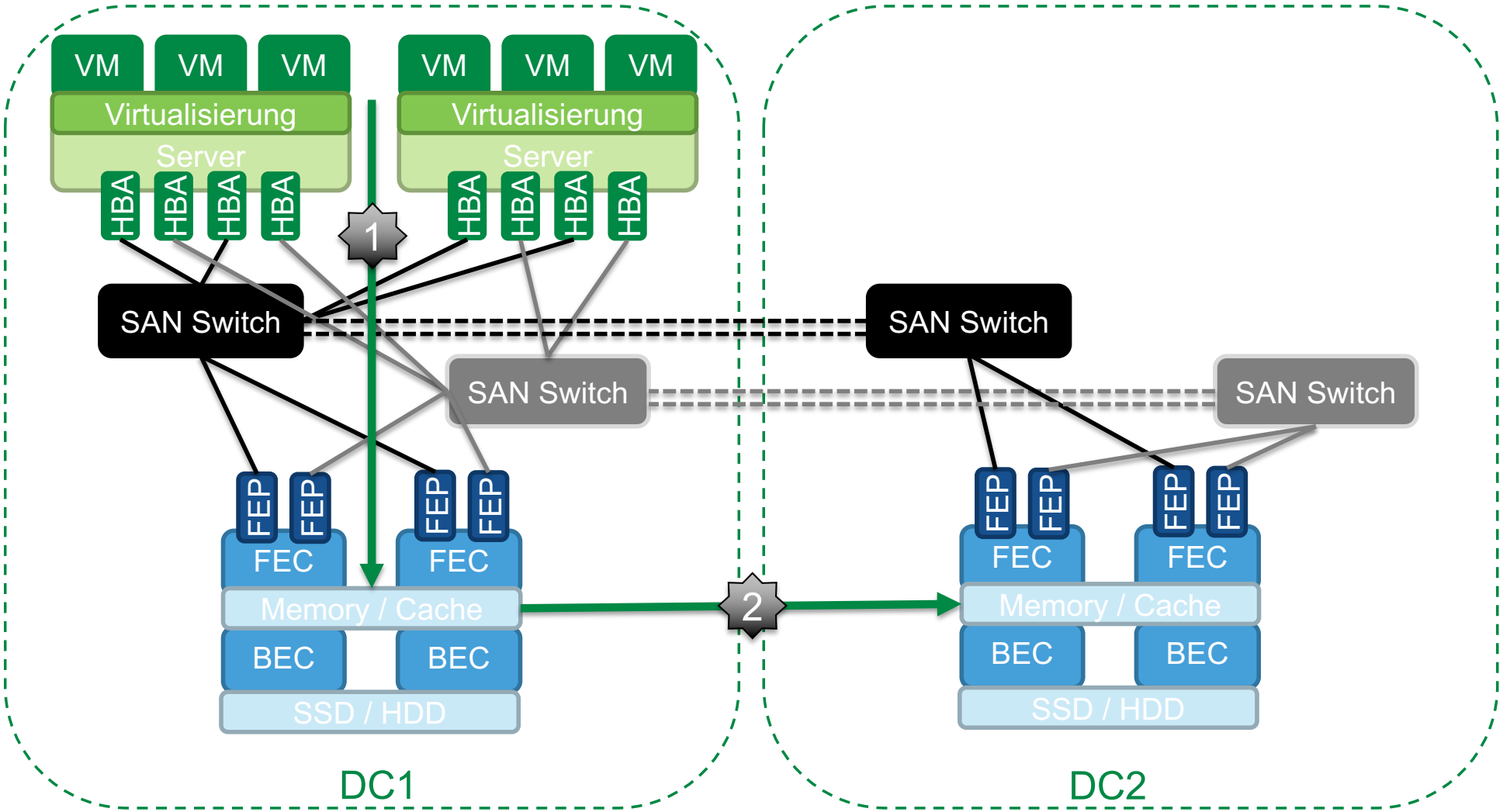
Frontend Read



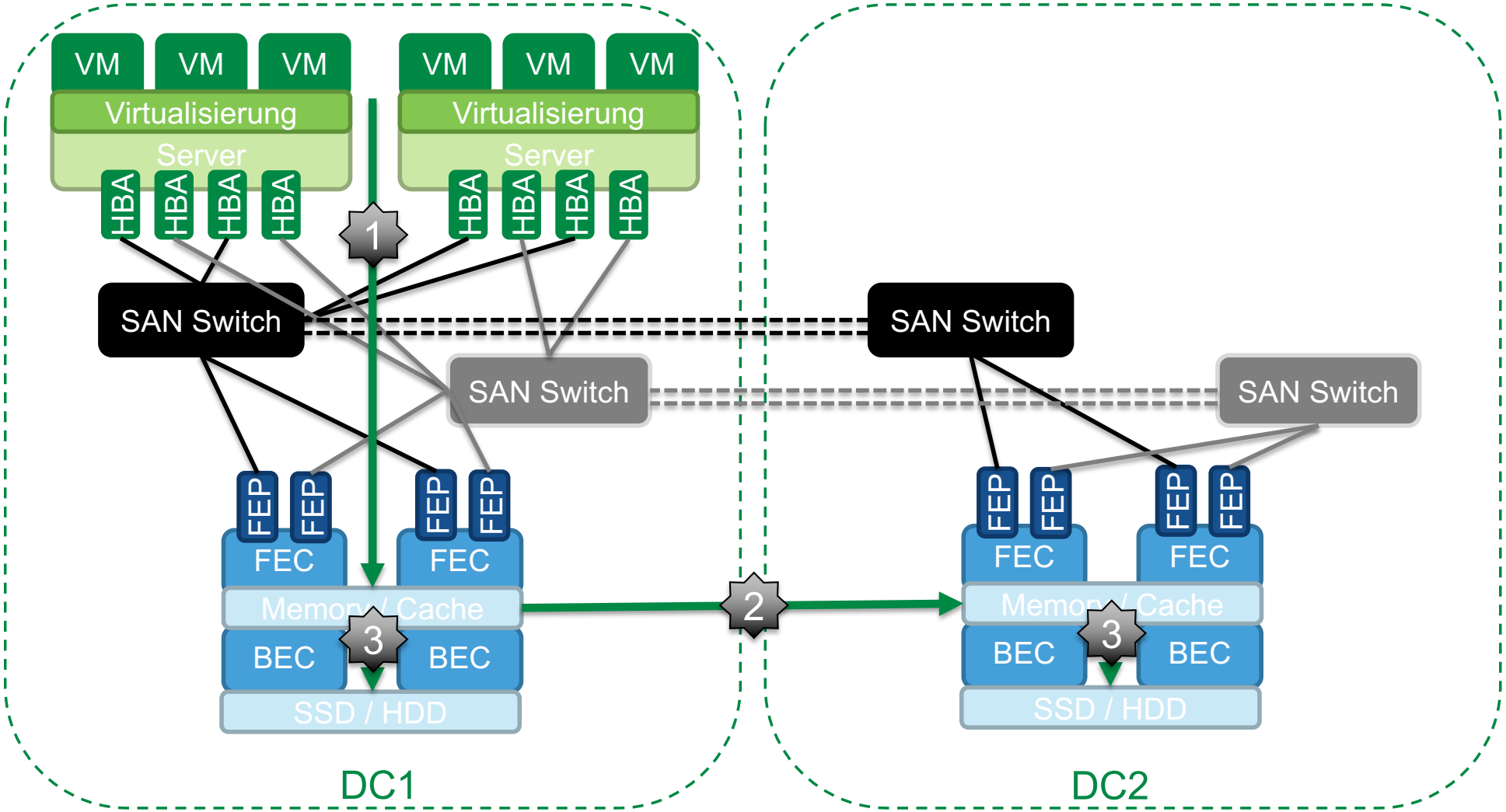
Backend Read



Frontend Write

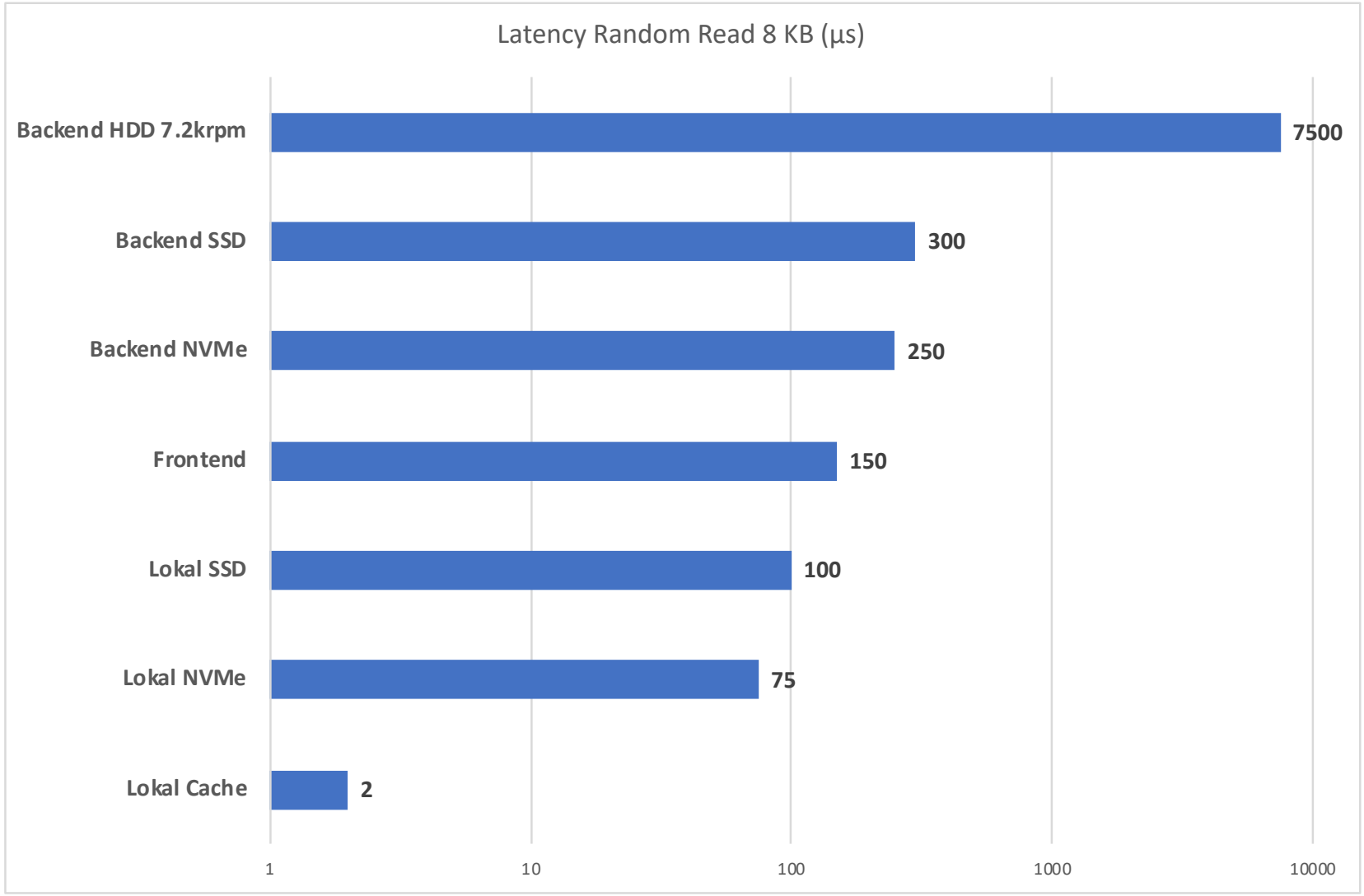


Backend Write



Key Performance Indikatoren Storage

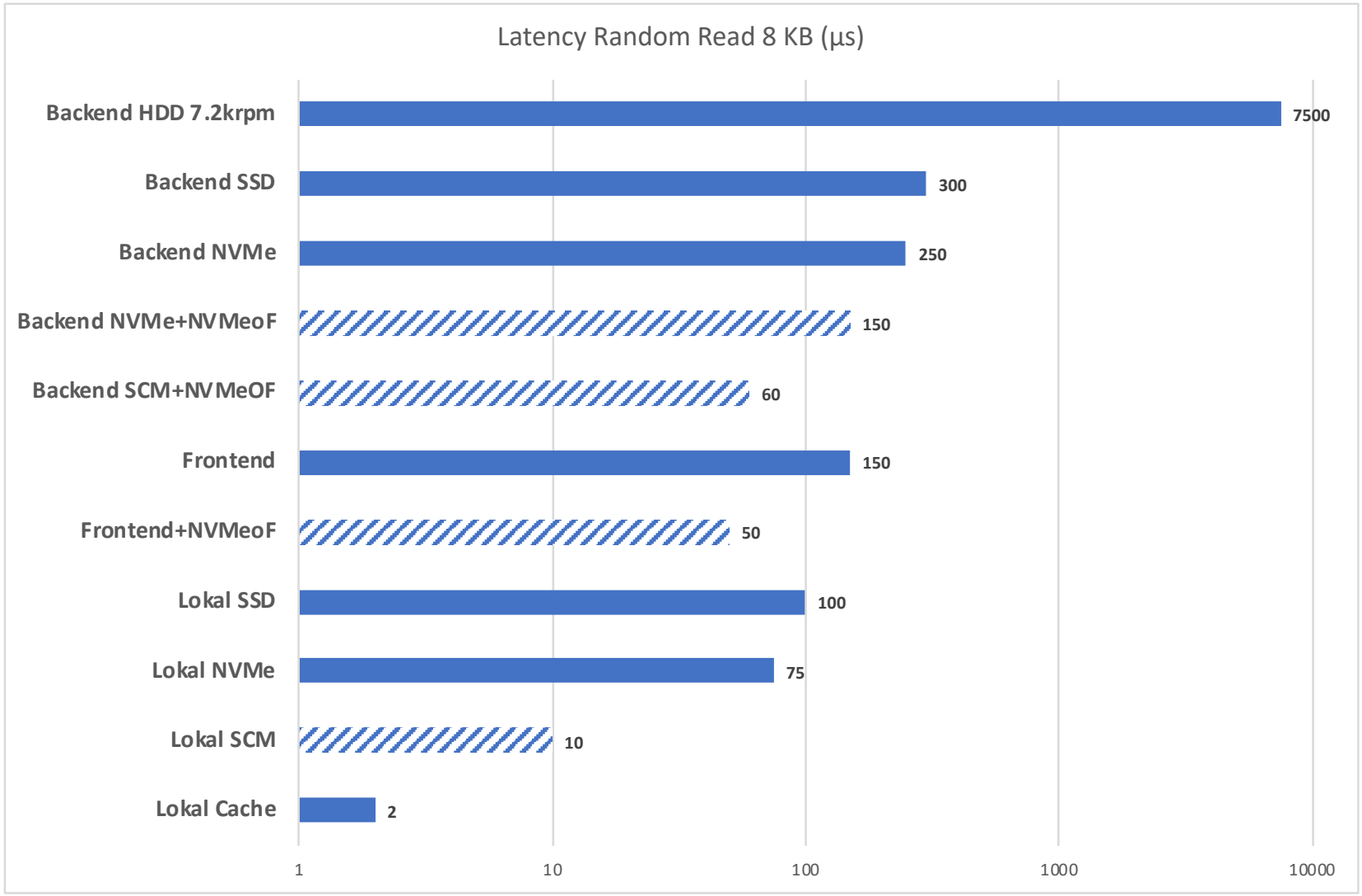
Latenz in µs nach Technologie (8 KB IO)



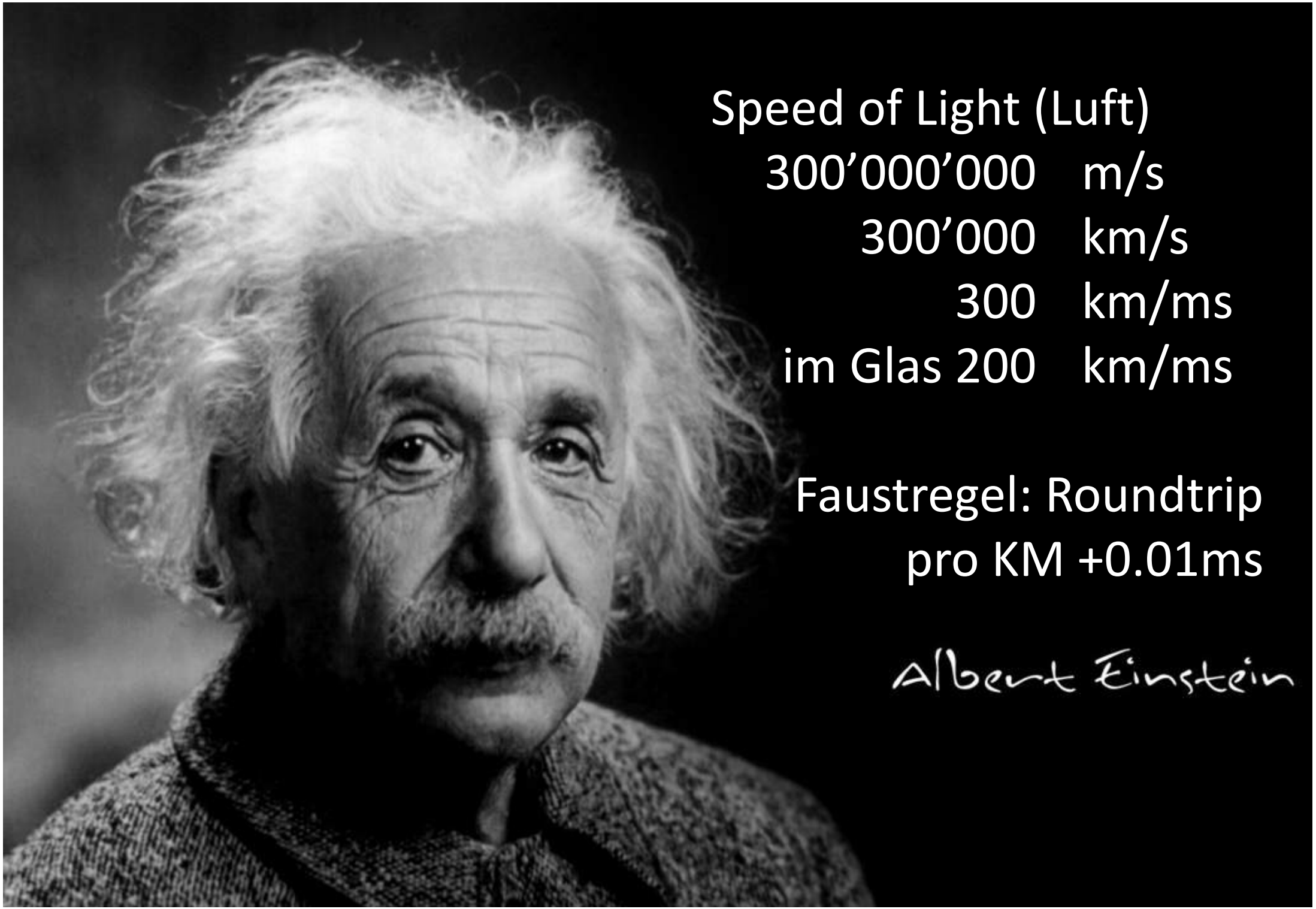
Skala logarithmisch!

Key Performance Indikatoren Storage

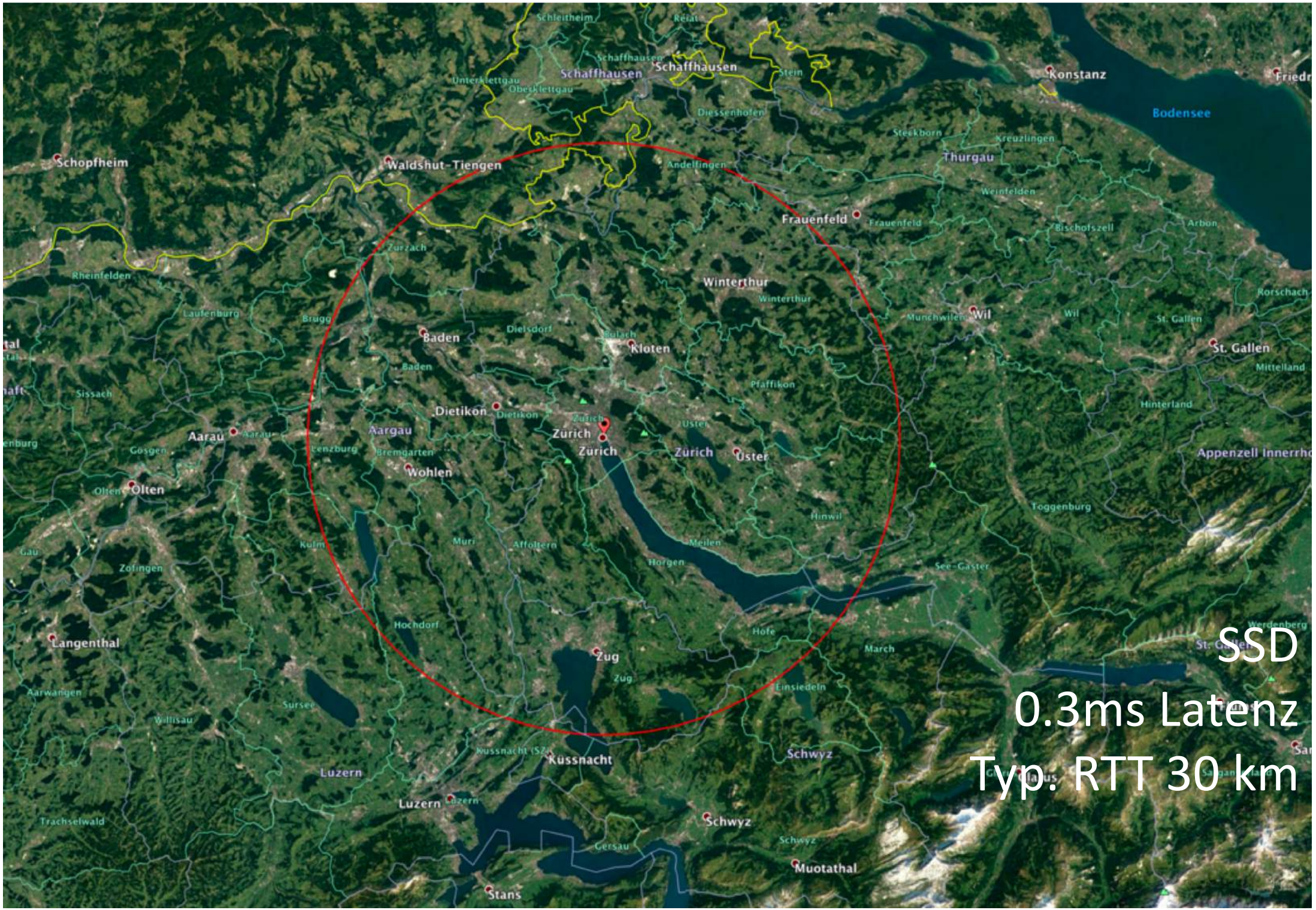
Latenz in µs nach Technologie (8 KB IO) - Zukunft

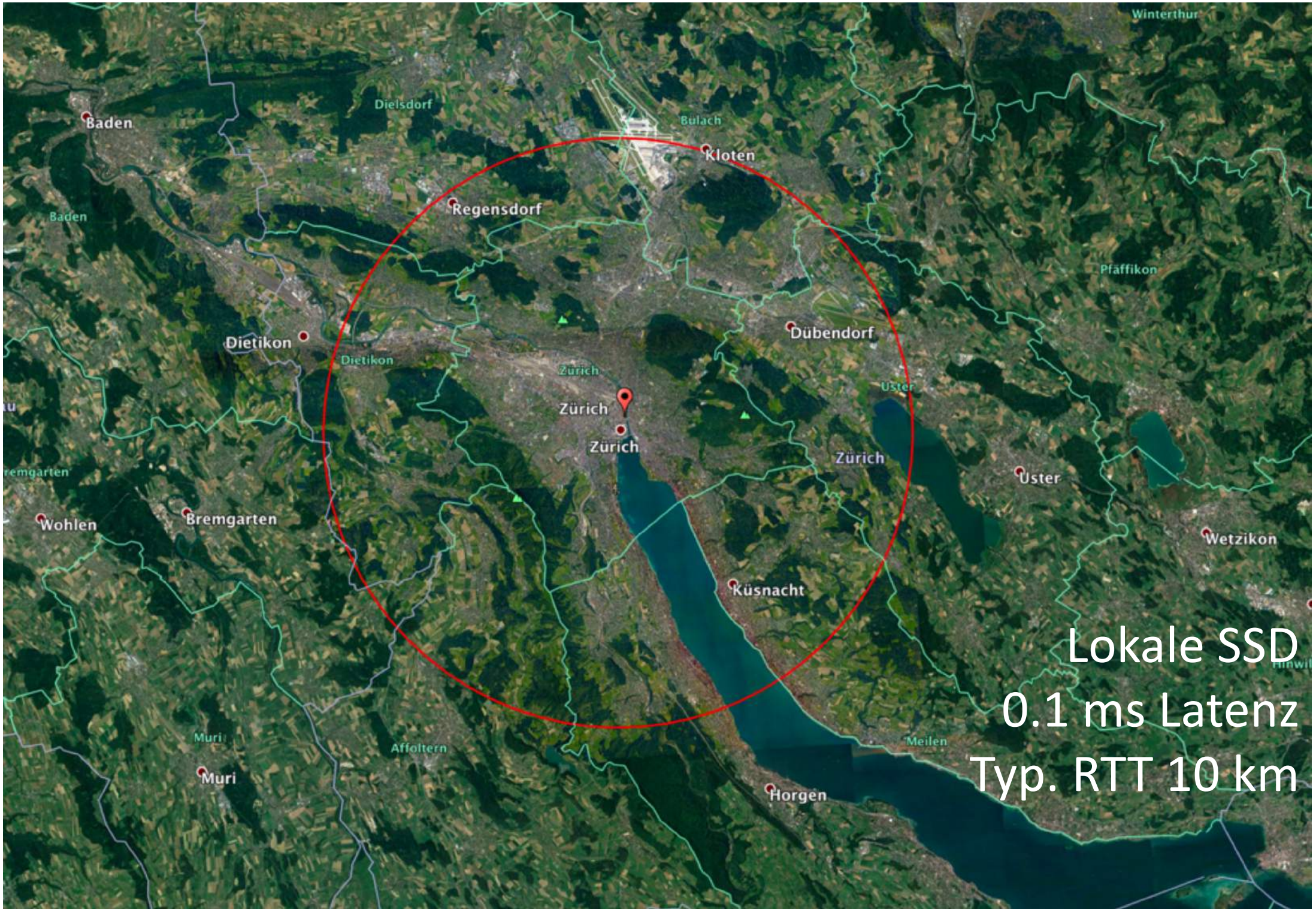


Skala logarithmisch!


Speed of Light (Luft)
300’000’000 m/s
300’000 km/s
300 km/ms
im Glas 200 km/ms
Faustregel: Roundtrip
pro KM +0.01ms
Albert Einstein

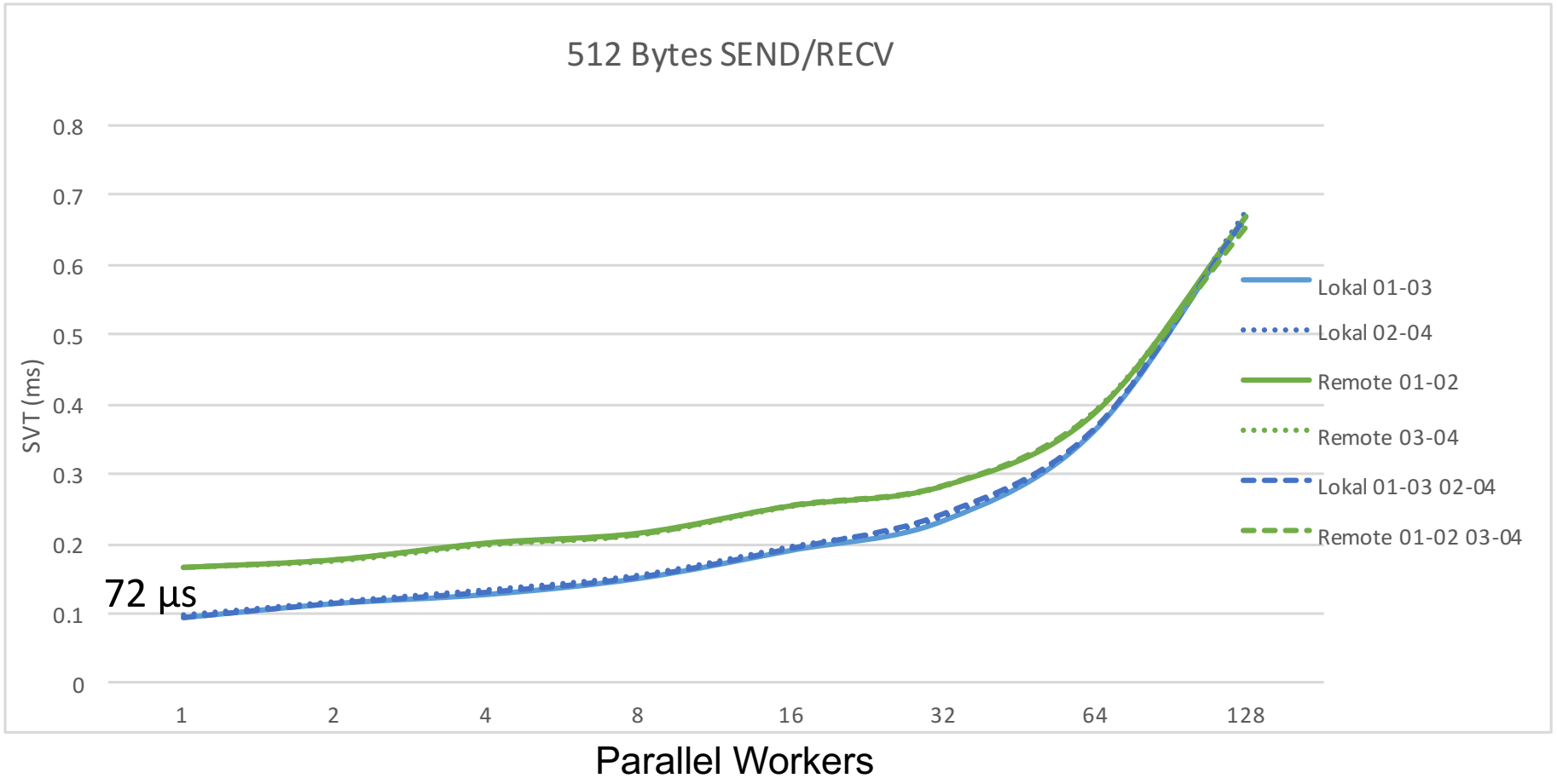

7.2krpm HDD
7.5 ms Latenz
Typ. RTT 750 km


SSD
0.3ms Latenz
Typ. RTT 30 km


Winterthur
Dielsdorf
Baden
Bulach
Kloten
Regensdorf
Baden
Pfaffikon
Dietikon
Dietikon
Dübendorf
Zürich
Uster
Zürich
Zürich
Zürich
Uster
Wohlen
Bremgarten
Wetzikon
Küsnacht
Lokale SSD
0.1 ms Latenz
Typ. RTT 10 km
Muri
Affoltern
Meilen
Muri
Horgen

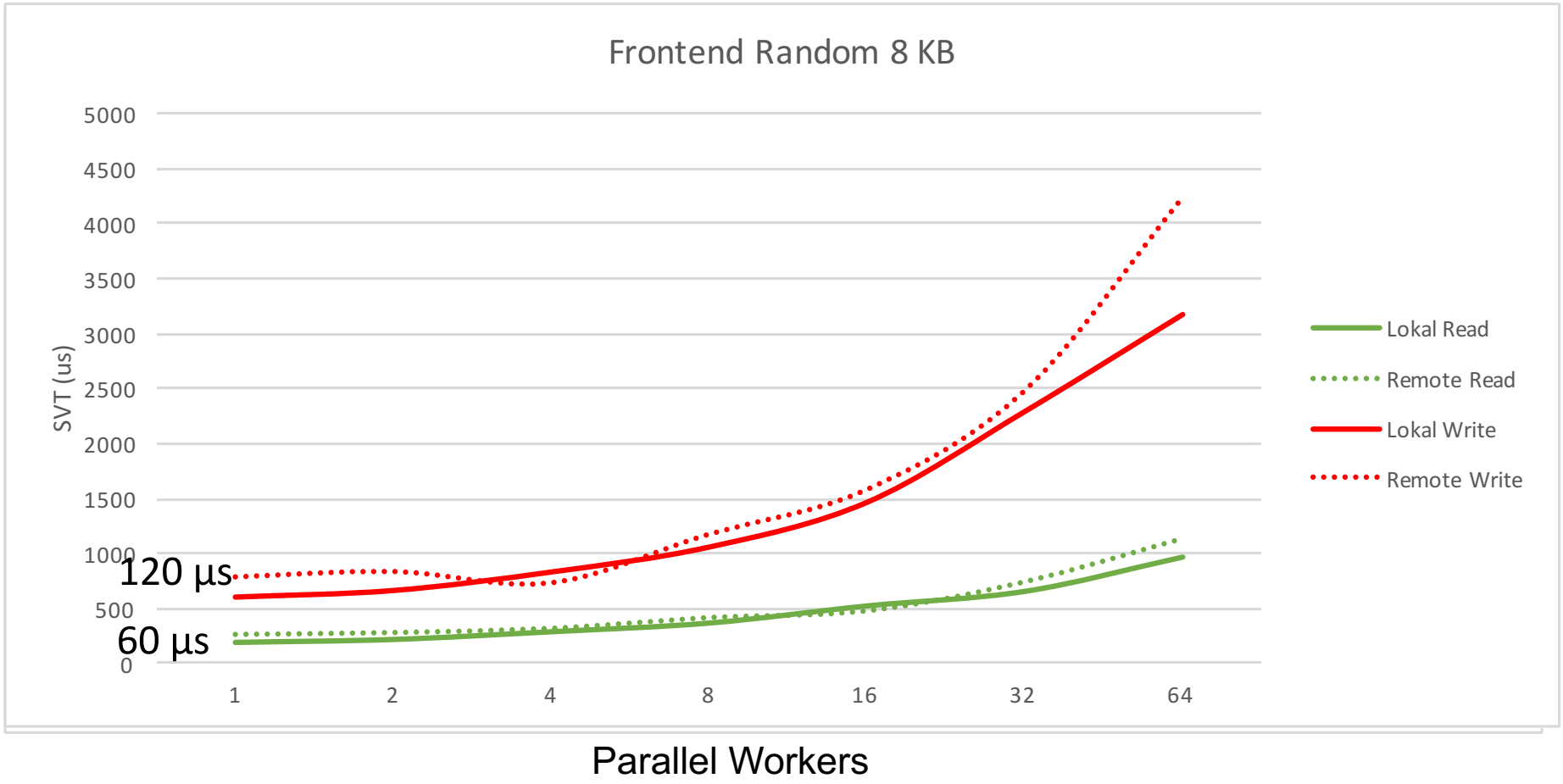

Storage Class Memory
0.01 ms Latenz
Typ. RTT 1 km

Beispielmessung Netzwerk Latency, 6.5 km



Real 72 µs Roundtripzeit (Faustregel 65 µs)

Beispielmessung SAN, 8km



Remote Read +60 µs, Remote Write +120 µs (Faustregel 65 µs)

Inhaltsverzeichnis

Storage Benchmark OceanStor 6800 V5 @ Leica
Konfiguration

STORAGESETUP
- 2 x OceanStor 6800 V5 (Distanz ca. 100m) je
- Quad Controller 0A-0D
- 2'048 GB Cache (512 GB pro Controller)
- 2 Quad 16 Gbit FC Module, 8 x 16 Gbit FC
- 1 Quad 10 GbE SFP+
- 3.84 TB SAS-SSD

TESTSERVER
- 3 HDS CR220 HM (2xBlock, 1xFile)
- 2 Sockets, 8 Cores, 16 Cores Total, 2.9 GHz
- HyperThreading on, 32 Threads
- 64 GB RAM
- 4 x 8 GBIT FC
- 2 x 10 GBE 802.3ad (Link Aggregation)

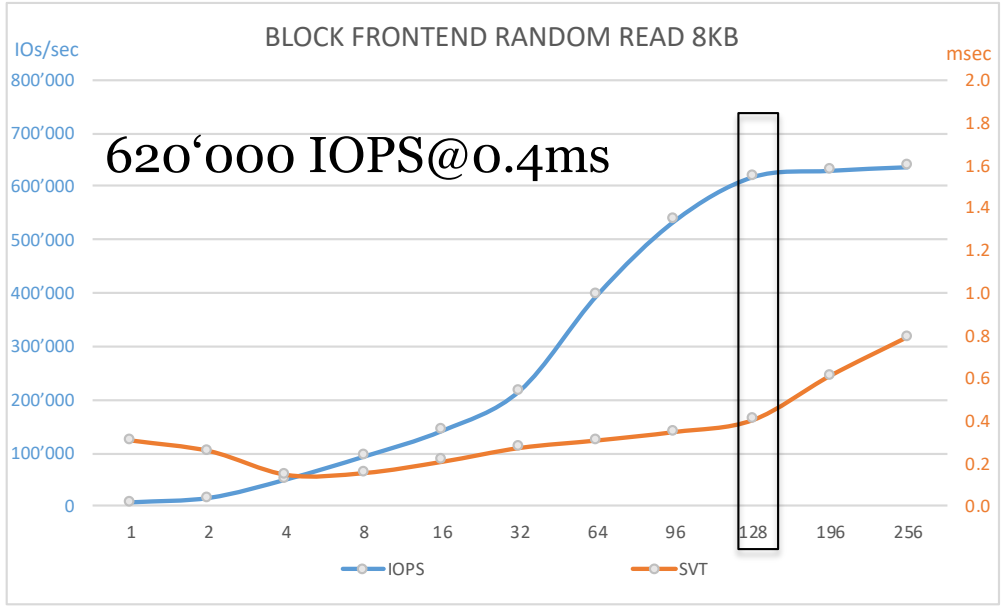
DISK SETUP BLOCK
- Disk Domain: 102 x 3.84 TB SSD
- Storage Pool: Raid 5 (13+1), 310 TB Usable

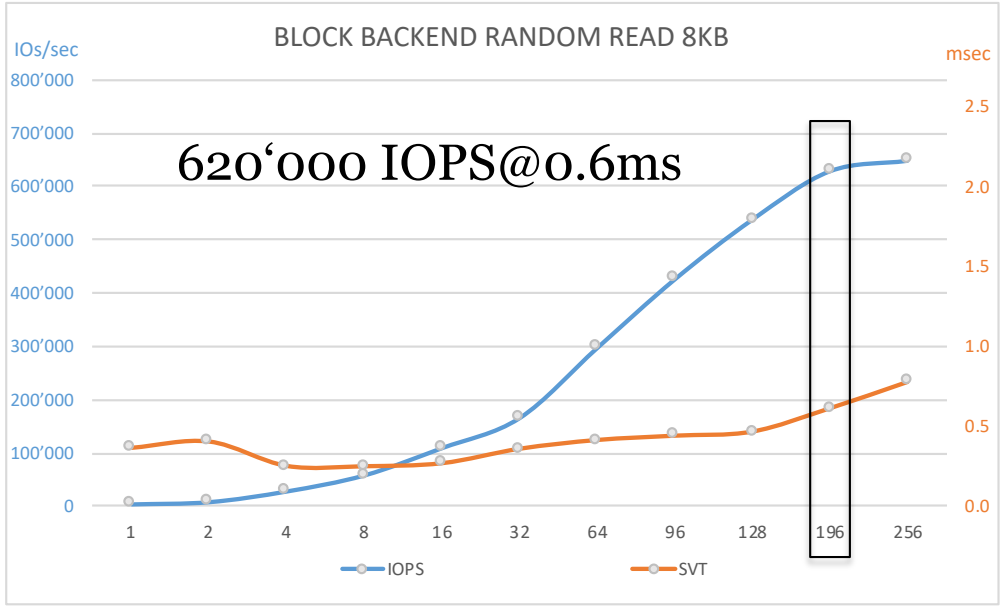
- IOGEN 4.1.0

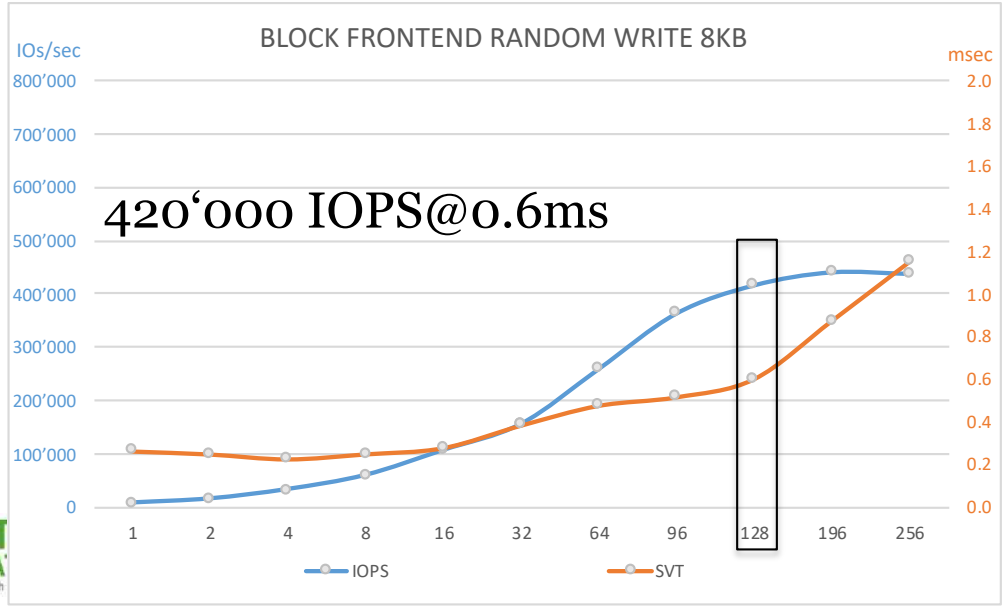
DISK SETUP FILE
- Disk Domain: 36 x 3.84 TB SSD
- Storage Pool File: Raid 5 (13+1), 106 TB Usable

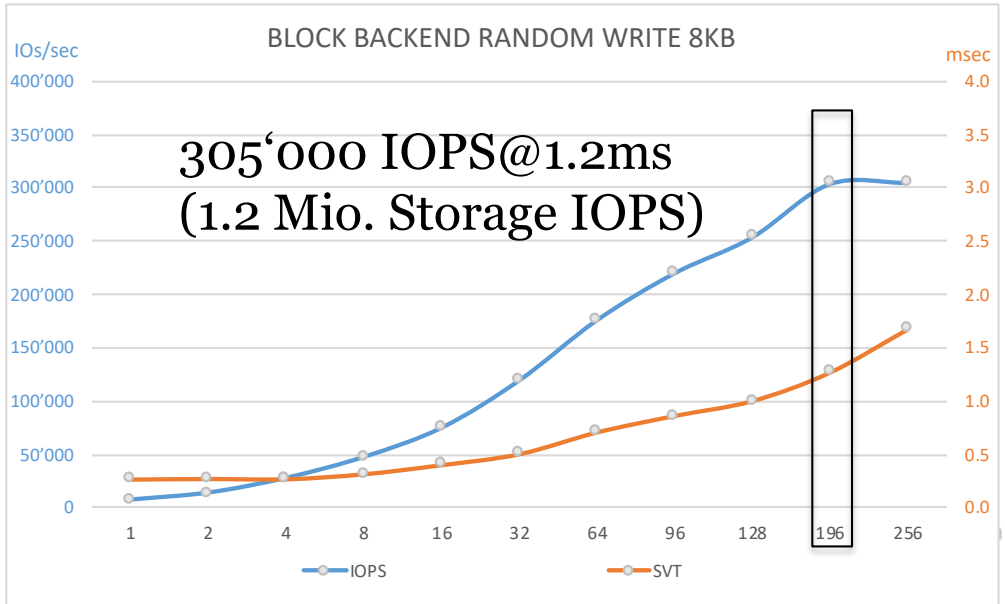
Storage Benchmark OceanStor 6800 V5 @ Leica

8KB Random IO Block Storage (gleichzeitig mit File)



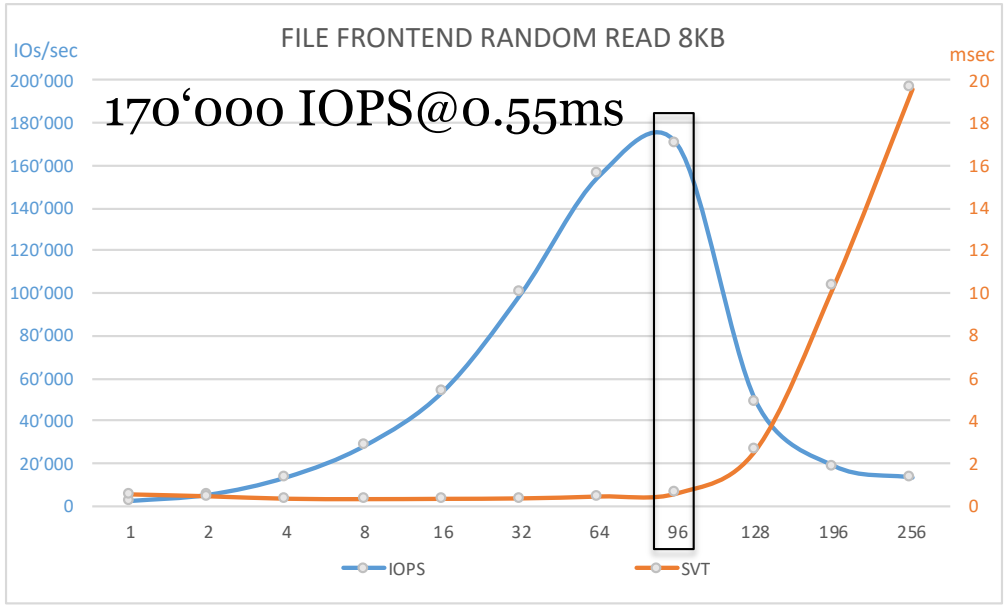


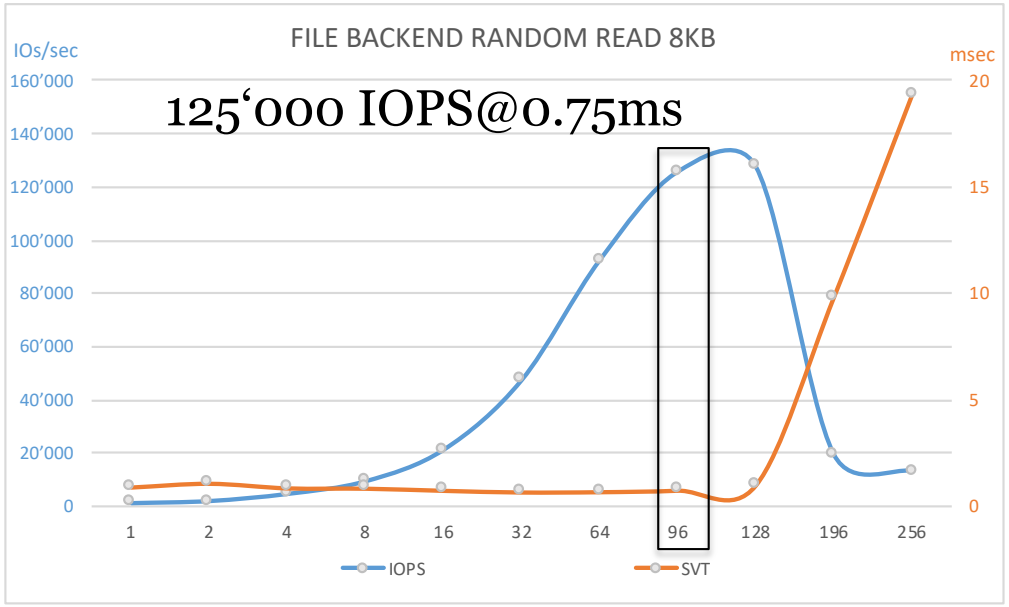


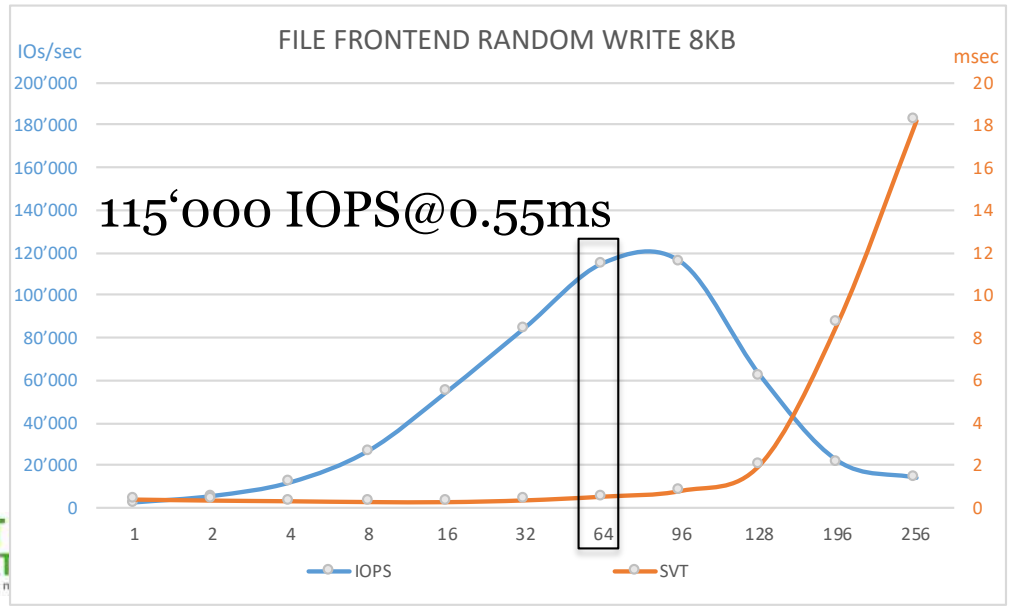


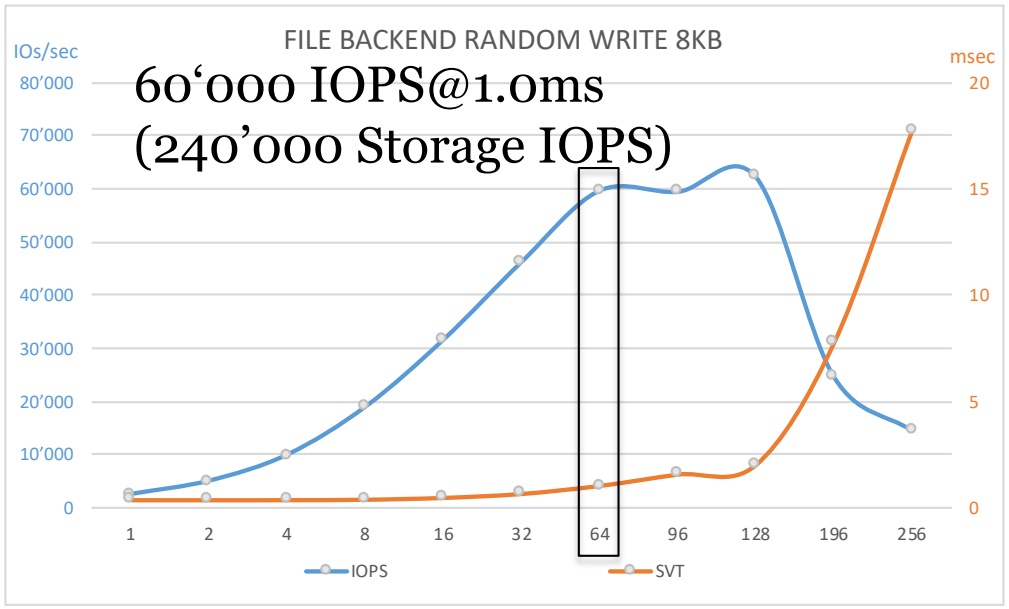
Storage Benchmark OceanStor 6800 V5 @ Leica

8KB Random IO File Storage (gleichzeitig mit Block)



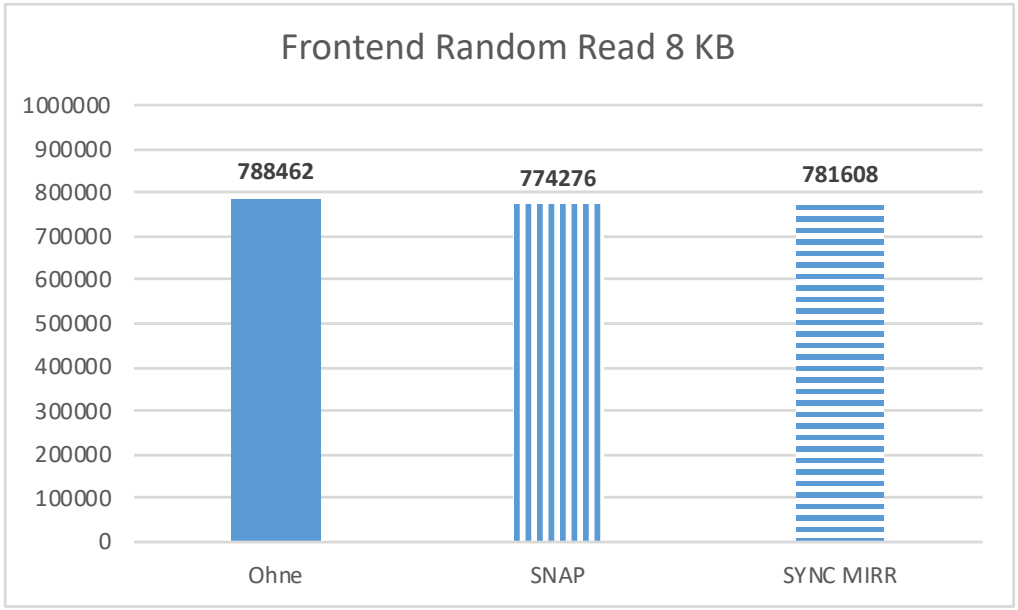


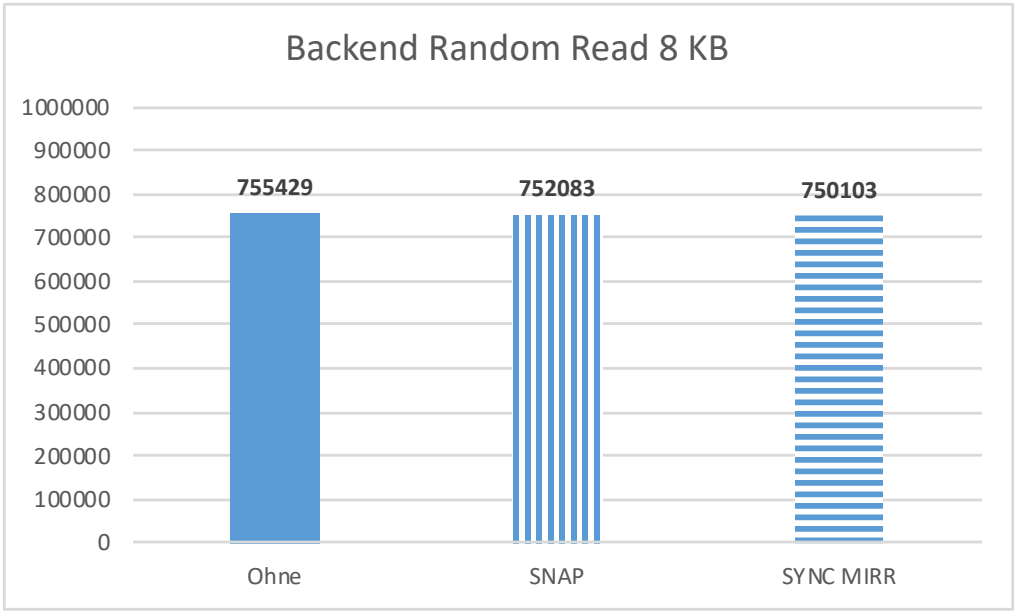


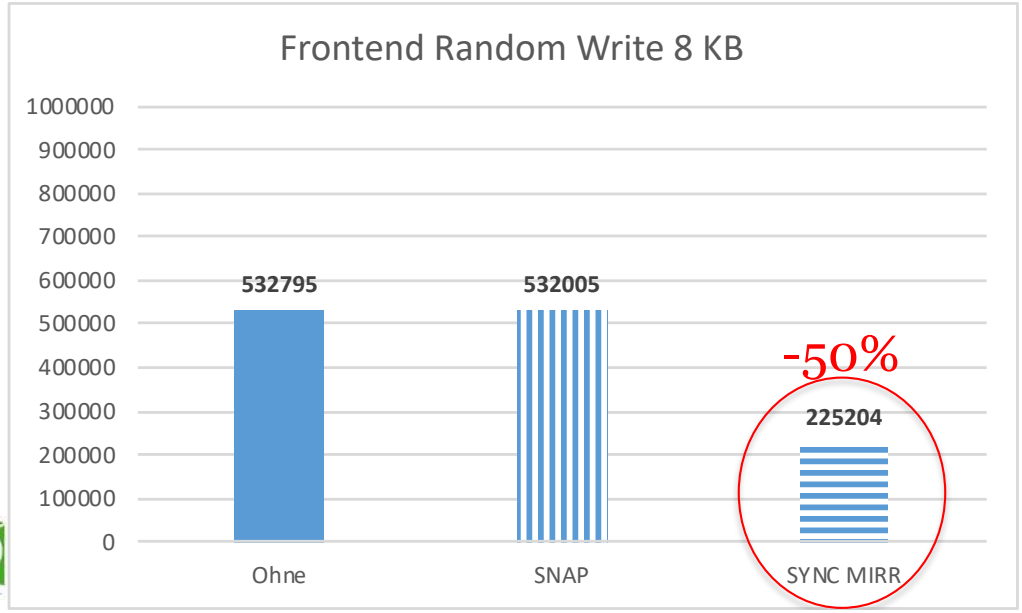


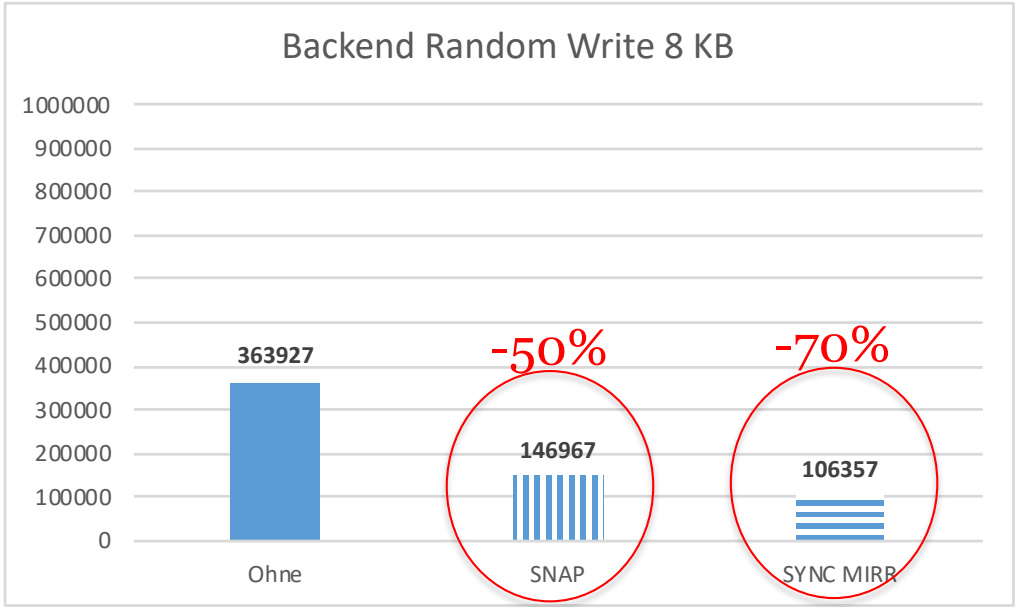
Storage Benchmark OceanStor 6800 V5 @ Leica

Total IOPS, Auswirkung Snap und Sync HyperMetro Mirror









Storage Benchmark OceanStor 6800 V5 @ Leica
Zusammenfassung

SPEED / LATENCY
- 150 µs Frontend Read
- 250 µs Backend Read
- 230 µs Frontend Write
- 270 µs Backend Write

THROUGHPUT
- 780’000 IOPS Frontend Read @ 0.5 ms
- 750’000 IOPS Backend Read @ 0.65 ms
- 530’000 IOPS Frontend Write @ 0.6 ms
- 360’000 IOPS Backend Write @ 1.2 m

IMPACT SNAP UND MIRROR
- Leseperformance Frontend und Backend ohne Impact
- Schreibimpact SNAP
  - Kein Impact im Frontend
  - 50% Impact im Backend (Copy on Write, COW, statt Redirect on Write RoW)
- Schreibimpact MIRROR
  - 50% Impact im Frontend
  - 70% Impact im Backend

FAZIT
- Sehr schneller und skalierbarer General Purpose Storage
- Gleichzeitiger Block und Filezugriff
- Impact SNAP im Backend Write und Sync Mirror Write ist jedoch signifikant
- Die definierten Anforderungen von Leica wurden klar erfüllt

Inhaltsverzeichnis

Storage Benchmark Dorado 5000 V3
Konfiguration

STORAGESETUP

- Dorado 5000 V3 NVMe
- 2 Engines mit je 2 Controllern, verbunden mit 2 PCIe Switches
- 1 TB Cache, Read Cache disabled
- 16 x 16 Gbit FC
- 2 TB NVMe-SSD (PCIe 3.0)

DISK SETUP BLOCK

- 50 x 2 TB NVMe-SSD
- Storage Pool:
  2 x 25 Disks Raid 6 (23+2)
- LUNs:
  je 12 LUNs 500 GB in beiden Storage Pools
  Total 24 LUNs, 12 TB Storage usable
- 8 LUNs per Server gemappt

TESTSERVER

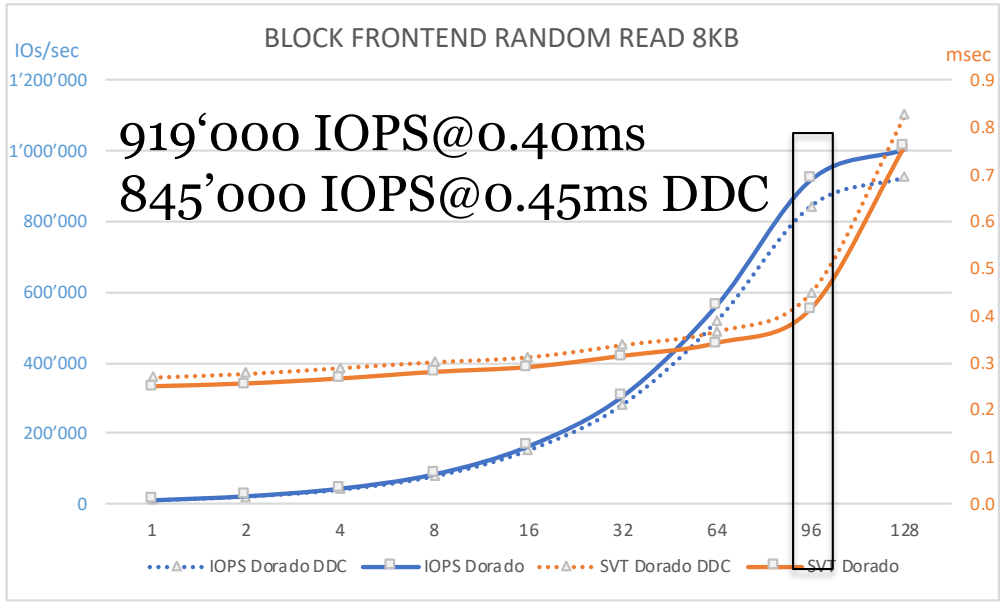
- 3 Huawei RH2288H V3
- 2 Sockets, E5-2690 v3
  8 Cores, 16 Cores Total, 2.9 GHz
- 128 GB RAM
- 4 x 16 GBIT FC

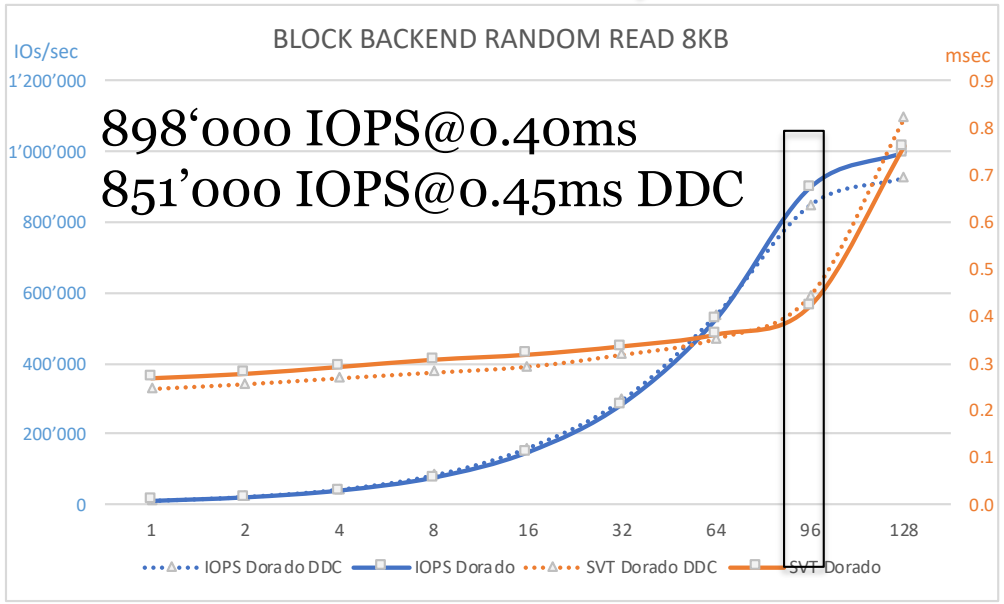
- IOGEN 4.3.0

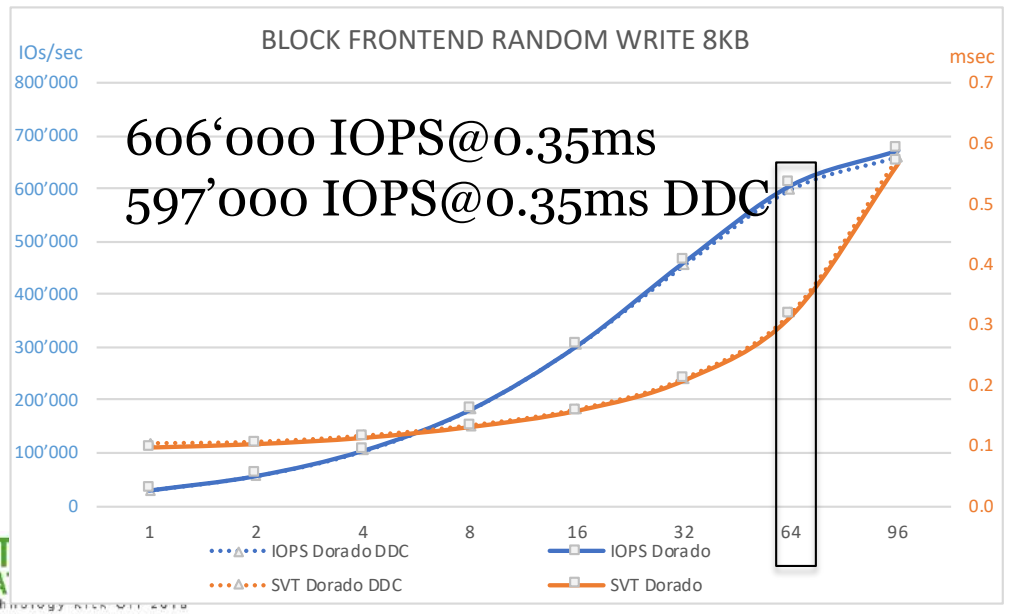
Storage Benchmark Dorado 5000 V3

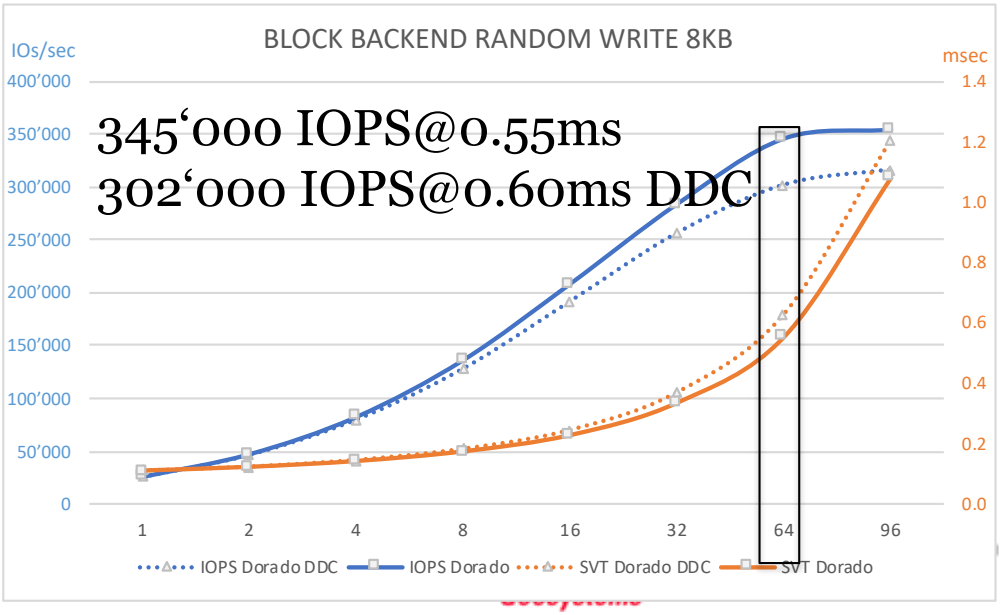
8KB Random IO Block Storage

1 MIO IOPS









Storage Benchmark Dorado 5000 V3
Zusammenfassung

SPEED / LATENCY

- 250 µs Frontend Read (OS 150 µs)
- 250 µs Backend Read (OS 250 µs)
- 100 µs Frontend Write (OS 230 µs)
- 110 µs Backend Write (OS 270 µs)
- Unabhängig von Dedup und Compression (bei OceanStor signifikanter Impact)

FAZIT

- Sehr schneller Full Flash Storage
- Sehr geringe Latenz auch unter maximaler Last (< 0.5ms in der Regel)
- Erreicht 1'000'000 Random Reads
- Höherer Throughput als OceanStor
- Nur minimaler Einfluss von Dedup und Compression

THROUGHPUT

- 919'000 IOPS Frontend Read @ 0.4 ms (OS 780'000 @ 0.5 ms)
- 898'000 IOPS Backend Read @ 0.4 ms (OS 750'000 @ 0.65 ms)
- 606'000 IOPS Frontend Write @ 0.35 ms (OS 530'000 @ 0.6 ms)
- 345'000 IOPS Backend Write @ 0.55 ms (OS 360'000 @ 1.2 ms)
- Geringer Einfluss von Dedup und Compression (bei OceanStor signifikanter Impact)

Fragen

Andreas Zallmann
CEO, In&Out AG

www.inout.ch
andreas.zallmann@inout.ch