

IN&OUT AG

Hitachi Vantara VSP 5500 Performance Benchmark SUVA

Andreas Zallmann
CEO, In&Out AG



Brian Mathis
brian.mathis@suva.ch
Head of Server & Storage Plattform

«Wir sind mit der Leistung des Hitachi Vantara VSP 5500 Storages sehr zufrieden und stossen in neue Leistungsdimensionen vor, mit der die SUVA für die nächsten Jahre sehr gut aufgestellt ist.»

Dank der wertvollen Unterstützung der In&Out AG konnten wir die Storage Evaluation sehr zielgerichtet und effizient durchführen und haben ein optimales Ergebnis erzielen können, auch in finanzieller Hinsicht.»

Version: 1.0.3 Final

Datum: 11.03.2020

Klassifikation: Öffentlich

Vorbemerkung

Die Schweizerische Unfallversicherung SUVA hat in einer Ausschreibung Ihre Stagesysteme neu evaluiert und sich für zwei Hitachi Vantara VSP 5500 Systeme entschieden. Im Rahmen eines Proof of Concept wurde die Leistungsfähigkeit der Systeme gemäss den Ausschreibungsvorgaben überprüft. Die hier dargelegten Ergebnisse basieren auf dem gemeinsamen PoC mit SUVA, Hitachi Vantara und In&Out.

SUVA

Die Suva ist die grösste Unfallversicherung der Schweiz. Sie beschäftigt am Hauptsitz in Luzern, in den schweizweit 18 Agenturen und in den zwei Rehakliniken Bellikon und Sion rund 4200 Mitarbeitende. Seit mehr als 100 Jahren sorgt die Suva dafür, dass Arbeit und Freizeit in der Schweiz sicherer werden.

- Die Suva bietet den obligatorischen Versicherungsschutz für Berufstätige und Arbeitslose gegen Unfälle und Berufskrankheiten.
- Das Modell Suva vereint Prävention, Versicherung und Rehabilitation.
- Die Suva ist selbsttragend; sie erhält keine öffentlichen Gelder und gibt Gewinne in Form von tieferen Prämien an die Versicherten zurück.

In&Out

Die In&Out AG aus Zürich begleitet ihre Kunden als unabhängiges und herstellernerutrales Beratungsunternehmen seit Jahren in den Bereichen IT Infrastruktur und Datacenter mit besonderem Fokus auf Storage.

Neben Beratungsleistungen im Bereich Storage und Begleitung in Storage Ausschreibungen verfügt die In&Out AG über ausgewiesene jahrelange Erfahrung in Storage Performance Benchmarks und hat das Benchmark Tool IOgen™ entwickelt.

Hitachi Vantara & SUVA

Hitachi ist weltweit in geschäftskritischen Umgebungen fest etabliert und seit vielen Jahren auch bei SUVA als zuverlässiger Technologiepartner für die Verwaltung der Daten verantwortlich. Die SUVA Anforderungen an Leistung, Zuverlässigkeit und Katastrophenvorsorge sind sehr hoch einzustufen. Das bedeutet, SUVA muss in der IT passende und neue Technologien einsetzen. Deshalb hat SUVA den neuen Storage ausgeschrieben und die Angebote detailliert und objektiv bewertet.

Seit Ende 2019 ist das neueste Flaggschiff der Hitachi Vantara Blockstorage Familie erhältlich, die VSP 5500. Dieses Stagesystem unterstützt bis zu 12 Controller (6 Dual Controller Nodes) im Full Mesh Setup und ist wahlweise mit NVMe oder SAS SSD/FMD und HDD Storage zu bestücken. Bereits bei mittlerer Ausbaustufe hat die VSP5000 die anspruchsvollen Zielsetzungen mit Bravour gemeistert respektive sogar übertroffen. Aus den umfassenden Optionen, welche die VSP5000 hinsichtlich der Technologie-Zusammensetzung bietet hat sich SUVA für eine NVMe Konfiguration mit «Advanced Data Reduction» entschieden. Damit realisiert SUVA eine enorm hohe Leistungs- und Speicherdichte und senkt zusätzlich die Kosten. Sämtliche weiteren Informationen über Hitachi VSP5000:

<https://www.hitachivantara.com/en-us/products/storage/all-flash-hybrid-flash-storage.html>

Zielsetzung

Im Rahmen eines Proof of Concepts hat die SUVA die Leistungsfähigkeit der evaluierten Stagesysteme Hitachi Vantara VSP 5500 überprüft. Neben SUVA spezifischen Workloads wurden auch typische random und sequential Workloads gemessen, die im Storageumfeld üblich sind und gut zu Vergleichen herangezogen werden können.

Management Summary

Hitachi Vantara hat mit der VSP 5500 ein neues Top Modell im Portfolio, welches neue Leistungsmassstäbe setzt. Bereits mit 8 Controllern stösst diese in neue Dimensionen an IO Leistung vor (jeweils summiert über beide Systeme):

- 10 Mio. IOPS Frontend Random Reads 8 KB
- 3.5 Mio. IOPS Frontend Random Writes 8 KB (50% synchron gespiegelt, effektiv 5.25 Mio. IOPS)
- 200 GB/s Frontend Sequential Read
- 50 GB/s Frontend Sequential Write (50% synchron gespiegelt, effektiv 75 GB/s)
- 4 Mio. IOPS Backend Random Reads 8 KB
- 750'000 IOPS Backend Random Writes 8 KB (50% synchron gespiegelt, effektiv 1.125 Mio. IOPS)
- 100 GB/s Backend Sequential Read
- 40 GB/s Backend Sequential Write (50% synchron gespiegelt, effektiv 60 GB/s)

Dabei ist die Write Performance durch die Anzahl der Datacenter Interconnects (ISLs) limitiert und könnte möglicherweise noch gesteigert werden. Zudem ist das System noch mit einer relativ geringen Anzahl NVMe Disks ausgestattet. Last but not least kann das System noch mit 50% Controller Kapazität erweitert werden.

Die Latenz ist mit dem NVMe Backend auf einem sehr hohen Leistungsniveau:

- 8 KB Read IOPS aus dem Cache < 100µs
- 8 KB Read IOPS aus dem Backend < 300µs
- 8 KB Write IOPS in den Cache 300-350µs (Durchschnitt bei 50% synchron gespiegelten Writes). Lokale Writes bei unter 200µs, synchron gespiegelte bei unter 500µs.

Testsetup

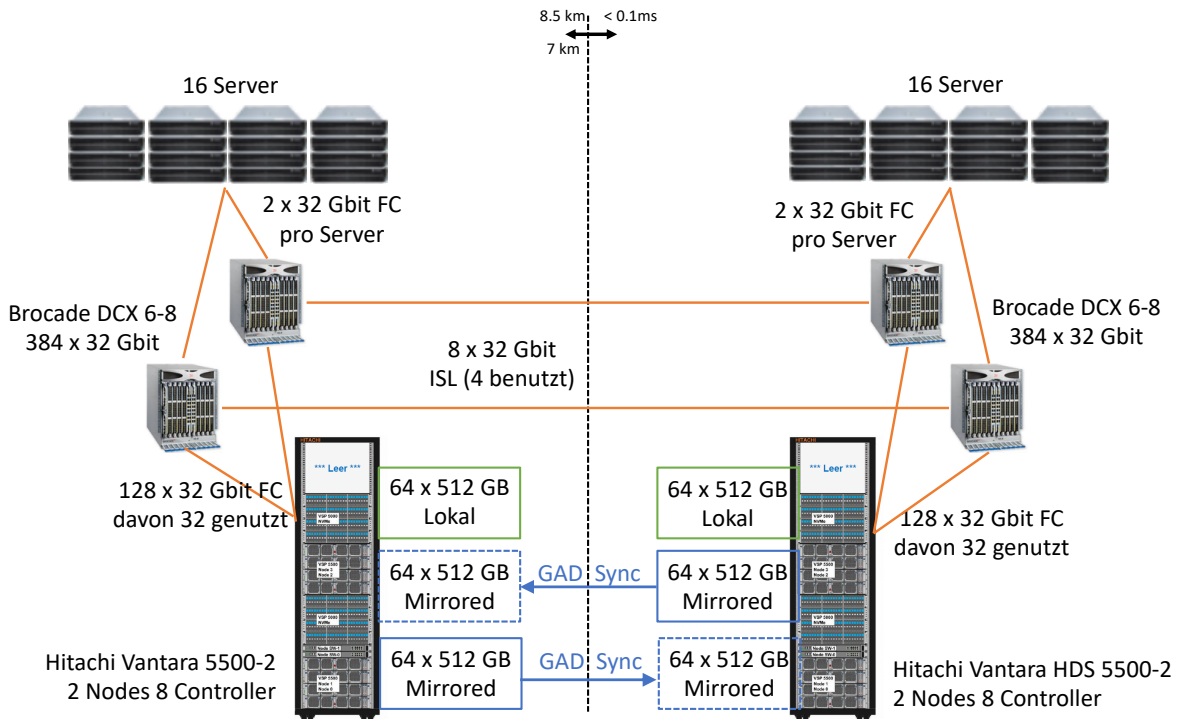


Abbildung 1 – Testsetup

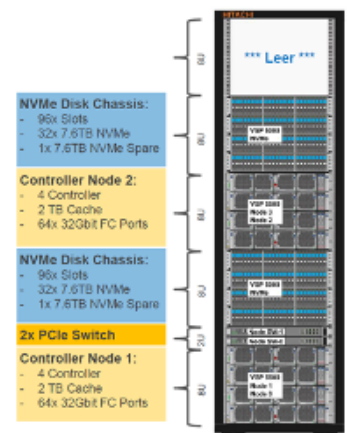
Die Stagesysteme sind auf zwei Standorte verteilt. Die RTT (Round Trip Time) zwischen den Standorten beträgt weniger als 0.1ms. Zwischen den Standorten wurden 4 ISL Leitungen mit je 32 Gbit (bidirektional) benutzt. Die Leitungslänge beträgt 8.5 respektive 7 km.

VSP 5500 Konfiguration

Pro Standort wurde eine VSP 5500 in folgender Konfiguration genutzt:

	Je Standort
Controller	8 (4N)
Cache	4'096 GB
FC	128 x 32 Gbit
- für Server	32 x 32 Gbit
- für Remote Copy	4 x 32 Gbit
Disks	66 x 7.6 TB NVMe SSD, davon 2 Hotspare 4 Raidgruppen Raid 6 (14+2)
Usable Capacity	376 TiB
LUN Konfiguration	64 LUNs lokal x 512 GB = 32 TB 64 LUNs sync.gespiegelt x 512 GB = 32 TB 64 TB pro Standort, 128 TB insgesamt 50% sync. gespiegelt auf anderen Standort

Tabelle 1 – Systemkonfiguration je Standort



Zu beachten ist, dass für den Proof of Concept die Disk Kapazität noch nicht voll ausgebaut war. Insgesamt wird später eine nutzbare Kapazität von 2 PiB eingesetzt.

Deduplication ist als Post Process konfiguriert und Inline Compression ist aktiviert.

Die Konfiguration wurde so gewählt, dass 50% der Kapazität synchron auf den anderen Standort gespiegelt wurden, dies entspricht dem Anforderungsprofil der SUVA. Somit werden 50% aller Writes synchron gespiegelt.

Mit den eingesetzten 32 FC x 32 GBit ist eine theoretische Maximalbandbreite von 1'024 GBit/s oder 256 GB/s pro Standort erreichbar. Mit den für Remote Copy (Spiegelung) eingesetzten 4 FC x 32 Gbit sind 128 GBit oder 16 GB/s (in beide Richtungen) übertragbar.

Test Server

Pro Standort sind 16 Server im Einsatz gewesen, die mit je 2 x 32 GBit FC an die SAN Switches angebunden waren. Pro Standort betrug die Bandbreite der Serveranbindung somit 1'024 GBit/s oder 256 GB/s.

Bei den Servern handelt es sich um Server der RHEV (RedHat Enterprise Virtualization) Farm, die während der Tests mit anderen VMs betrieben wurde. Folgende Servertypen waren im Einsatz:

- Huawei 2288 V5 mit 2 Sockets Xeon Gold 6146 mit 2 x 12 Cores, 3.2 Ghz, 768 GB RAM
- Lenovo ThinkSystem SR 650 mit 2 Sockets Xeon Gold 6146 mit 2 x 12 Cores, 3.2 Ghz, 768 GB RAM
- Lenovo SR650 G2 mit 2 Sockets Xeon Gold 6246 mit 2 x 12 Cores, 3.3 Ghz, 1.5 TB RAM

IOgen™

Die Messungen erfolgen mit IOgen™ 6.3.1 der In&Out AG. IOgen wurde dabei direkt auf dem RHEV Hypervisor der 32 Server installiert. Der Zugriff auf die LUNs erfolgte direkt als Raw Device (/dev/mapper/...).

Testläufe

Die Testläufe wurden durch IOgen™ parallel und zeitlich auf die Millisekunde synchronisiert auf 32 Servern durchgeführt. Dabei wurden pro Standort 16 Server eingesetzt, die das jeweils lokale Stagesystem genutzt haben. Die Writes werden 100% lokal geschrieben und 50% davon werden synchron auf den anderen Standort gespiegelt.

Die Parallelität (Anzahl Worker Prozesse) wurde dabei langsam von 1 Worker pro Server gesteigert. Alle Angaben bezgl. der Parallelität beziehen sich auf die Worker pro Server, müssen also mit 32 Servern multipliziert werden.

Folgende Tests wurden durchgeführt:

- FE/BE RR – Random Read IOPS 8 KB
- FE/BE RW – Random Write IOPS 8 KB
- FE/BE SR – Sequential Read IOPS 1 MB
- FE/BE SW – Sequential Write IOPS 1 MB

Frontend (FE) Tests werden gegen das Frontend und den Storage Cache gefahren. Hier werden pro LUN immer nur die ersten 1000 Blöcke verwendet, die Cache Hit Ratio steigt deshalb schnell auf 100%. Bei den **Backend (BE)** Tests werden hingegen alle Blöcke des LUNs verwendet, die Cache Hit Ratio liegt nahe bei 0%.

Bei allen Schreibtests wird das gleiche Schreibmuster verwendet. 60% des Blocks werden mit zufälligen Mustern gefüllt und 40% des Blocks sind konstant und sehr gut komprimier- und deduplizierbar. Diese Verteilung ist typisch für die SUVA. Mit dieser Verteilung ist ein Dedup/Compression (DeCo) Faktor von ca. 1.66:1 zu erwarten (100/60).

In den folgenden Auswertungen werden Frontend Tests (durchgezogene Linien) und Backend Tests (gestrichelte Linien) jeweils zusammen dargestellt.

Für Random IOs werden die IOPS (IOs pro Sekunde) und für Sequential IOs der Durchsatz in GB/s (Gigabyte pro Sekunde) dargestellt. Darüber hinaus wird immer auch die auf dem Server gemessene Latenz (Latency) in µs (Microsekunden) = 0.001 ms (Millisekunden) dargestellt.

In den Betrachtungen unterscheiden wir zwischen dem Maximalwert (orange markiert) und dem Optimalwert (blau markiert). Beim **Maximalwert** wird zwar der höchste Durchsatz erzielt, aber die Latenz kann bereits aufgrund der hohen Last sehr stark angestiegen sein. Der **Optimalwert** wird als der Wert betrachtet, bei dem der Durchsatz noch stärker steigt als die Latenz. Bis dahin bewegen wir uns sozusagen im «gesunden» Bereich.

Random Read IOPS 8 KB

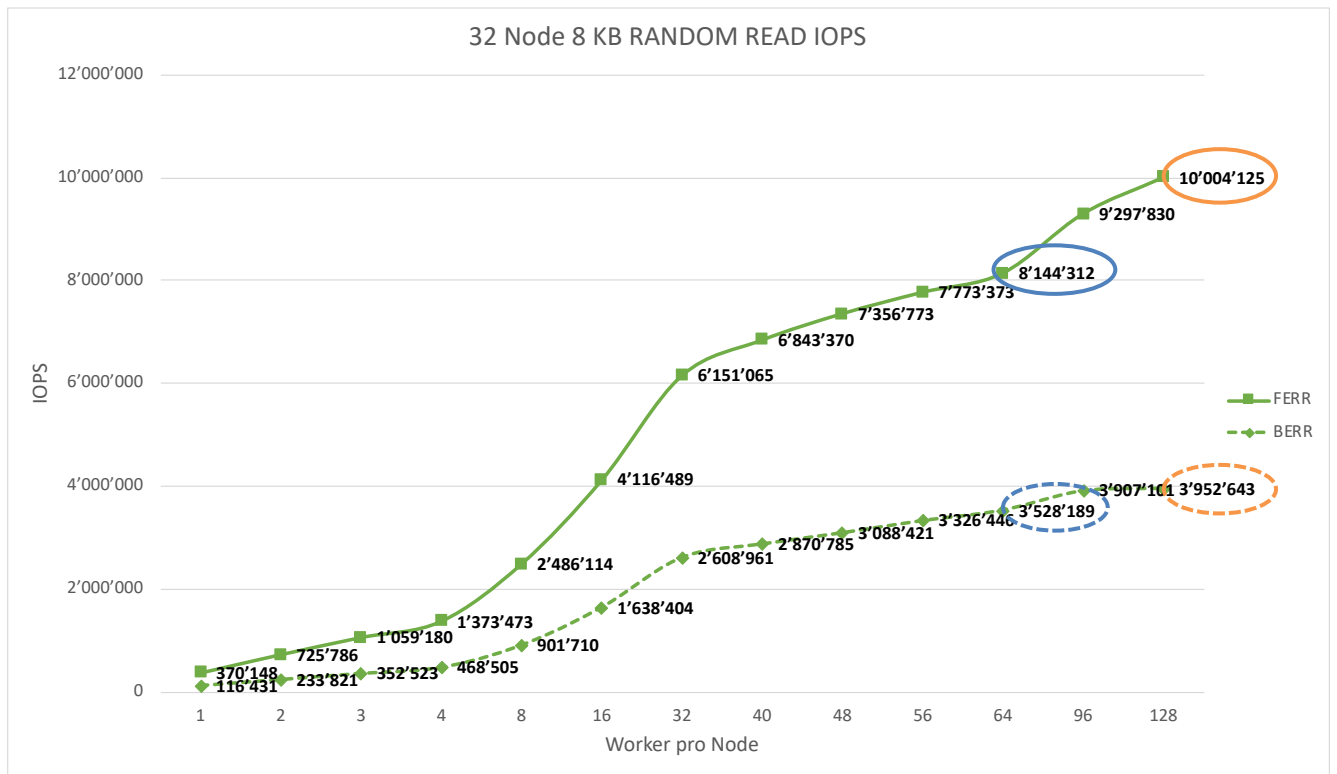


Abbildung 2 – Random Read Performance IOPS 8 KB

Im Frontend können die beiden Systeme zusammen mehr als **10 Millionen Frontend Random Read IOPS 8 KB** verarbeiten (mehr als 5 Millionen IOPS pro System). Im Backend können die Systeme im «Worst Case» bei 0% Cache Hit Ratio immer noch fast **4 Millionen Backend Random Read IOPS 8 KB** liefern, 2 Millionen pro System.

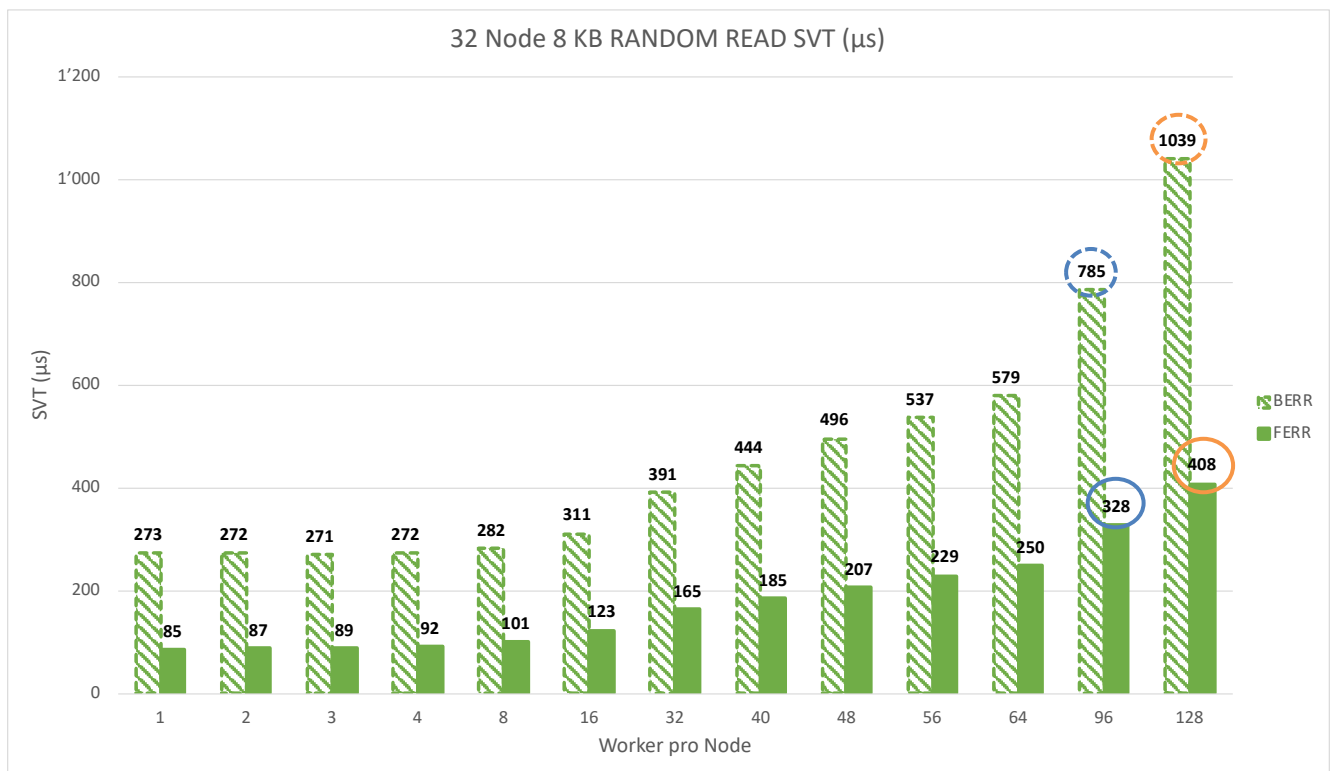


Abbildung 3 – Random Read Latenz 8 KB

Im Frontend beträgt die Latenz bei geringer Parallelität weniger als 100µs und steigt erst mit zunehmender Parallelität auf einen immer noch sehr guten Wert von 400µs (0.4 ms). Bei dieser sehr geringen Latenz von 400µs werden 10 Mio. IOPS verarbeitet. Der Optimalwert von über 8 Mio. Frontend Random Reads wird bei einer Latenz von 250µs erreicht.

Im Backend ist die Latenz bei geringer Parallelität bei ca. 270µs, dies entspricht also der Zeitspanne um einen 8 KB Block, der sich nicht im Cache befindet, von den Backend NVMe SSDs zu lesen und an den Server zu senden. Diese Latenz ist sehr niedrig und bleibt auch sehr lange konstant. Erst mit sehr hoher Parallelität steigt Latenz langsam an. Bei 4 Mio. IOPS als Maximalwert beträgt die Latenz immer noch sehr gute 1'039µs, als optimalen Wert würden wir 3.5 Mio. IOPS betrachten, die mit einer Latenz von 579µs gelesen werden.

Exemplarisch ist für einen Frontend Random Read Tests ist mit 64 Prozessen in der folgenden Grafik der Testverlauf über 10 Minuten in Sekundengenauigkeit dargestellt. Es ist zu sehen, dass die IOPS und die Service Time absolut konstant bleiben.

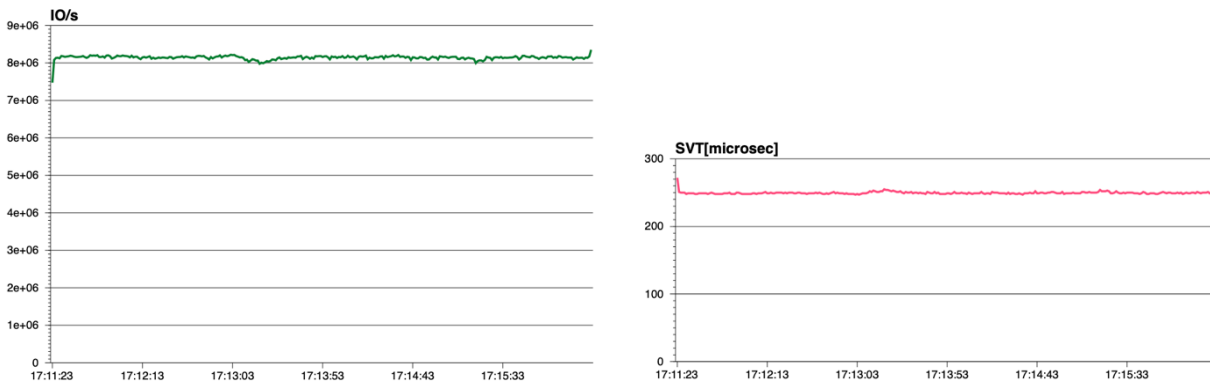


Abbildung 4 – Frontend Random Read Testverlauf (64 Prozesse)

Die CPU Belastung der Storage Controller erreicht sowohl bei den Frontend wie bei den Backend Tests Werte von über 70%.

Random Write IOPS 8 KB

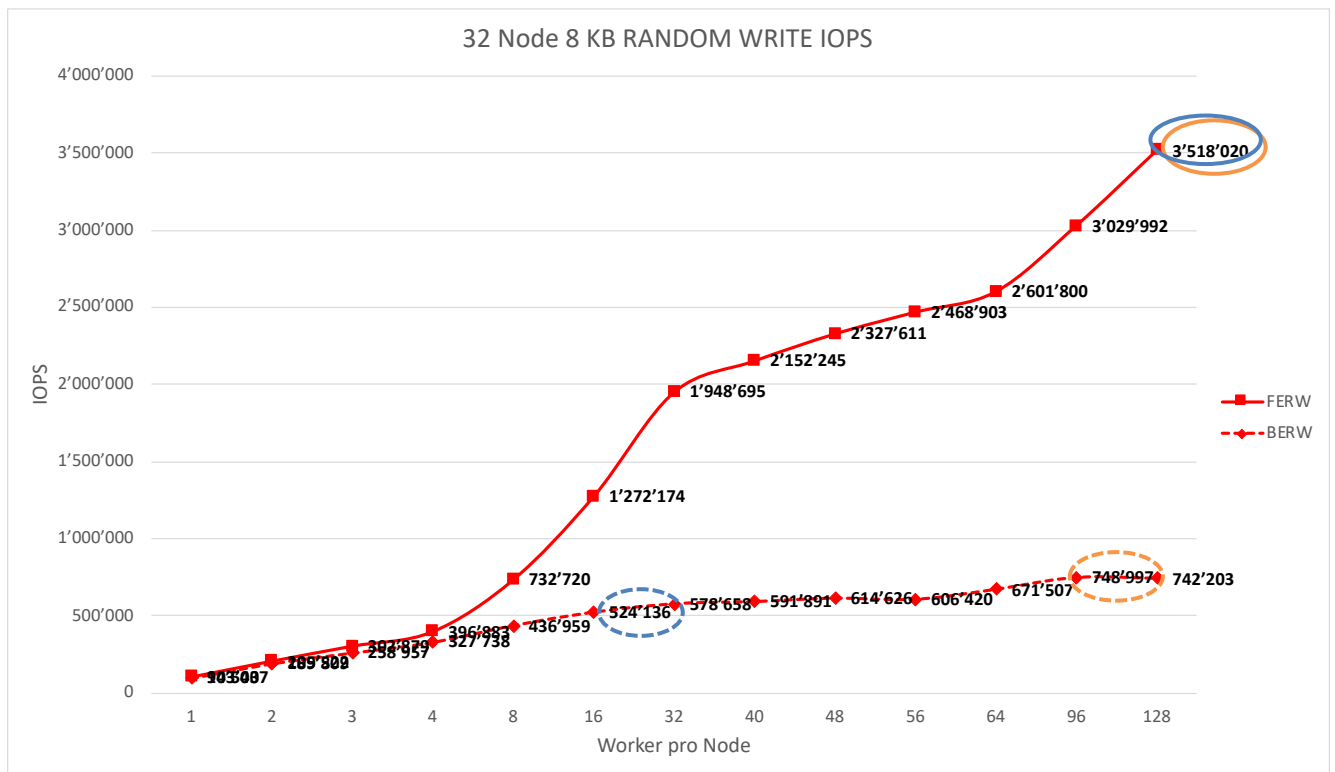


Abbildung 5 – Random Write Performance 8 KB

Im Frontend können die beiden Systeme zusammen mehr als **3.5 Millionen Frontend Random Write IOPS 8 KB** verarbeiten (mehr als 1.75 Millionen IOPS pro System). Von den 3.5 Mio. Writes sind 50% lokal und 50% werden synchron auf den anderen Storage gespiegelt. Pro Storage werden somit 1.75 Mio. Writes lokal und zusätzlich 875'000 Writes vom anderen System abgewickelt, somit werden pro Storage 2.625 Mio. Writes oder über beide Systeme **5.25 Mio. Frontend Random Write IOPS 8 KB** verarbeitet.

Im Backend können die Systeme im «Worst Case» bei 0% Cache Hit Ratio knapp **750'000 Backend Random Write IOPS 8 KB** verarbeiten, oder 375'000 pro System. Effektiv sind es durch die Spiegelung **1.125 Mio. Backend Random Write IOPS**. Auf diese Rate fällt die Verarbeitungsleistung zurück, wenn der Write Cache vollständig gefüllt ist, ein äusserst untypisches Szenario. Der Wert entspricht der Downstaging Rate auf die NVMe SSDs. Diese würde mit mehr NVMe SSD vermutlich noch steigen.

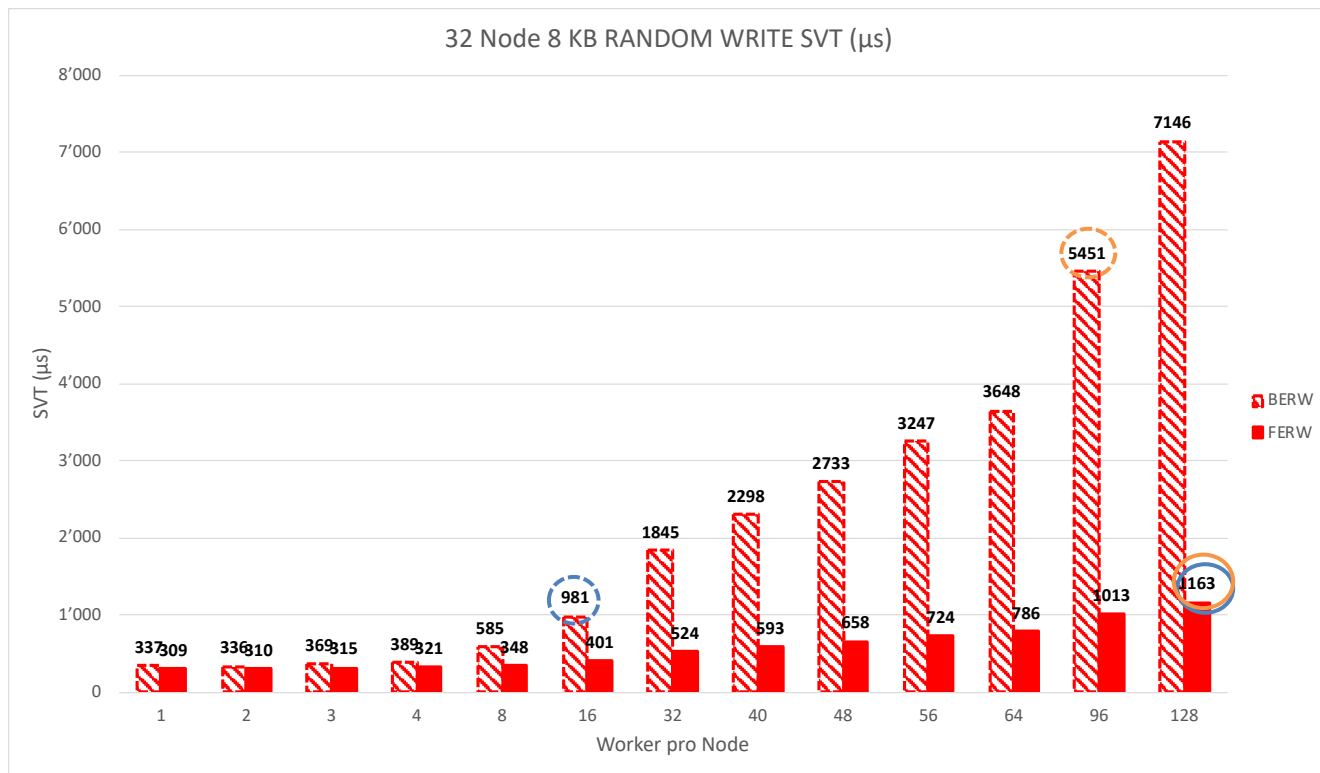


Abbildung 6 – Random Write Latenz 8 KB

Im Frontend beträgt die Latenz bei geringer Parallelität 309µs und bleibt lange sehr konstant. Bei der maximalen Parallelität beträgt die Latenz 1'163µs, bei über 3.5 Mio. Random Writes (effektiv sogar aufgrund der Spiegelung über 5 Mio. Random Writes).

Es ist zu berücksichtigen, dass die 50% gespiegelten Writes auf den Storage am anderen Standort geschrieben und von diesem Storage bestätigt werden müssen, bevor der Server den Write Acknowledge erhält. Vor diesem Hintergrund sind die geringen Latenzen umso bemerkenswerter.

Im Backend liegt die Latenz mit geringer Parallelität bei 337µs. Im zugehörigen Latency Diagramm unten sieht man dabei die Verteilung der Latenzen. 51.77% der Writes werden mit einer Latenz von 100-250 µs abgewickelt. Dabei handelt es sich um die ungespiegelten Writes. Die weiteren 50% gespiegelte Writes werden aufgrund des Transports auf den anderen Standort mit einer Roundtripzeit von ca. 100µs mit einer Latenz von mehr als 250µs abgewickelt. Dabei liegen 22.31% bei weniger als 500µs und 24.89% bei weniger als 1'000µs.

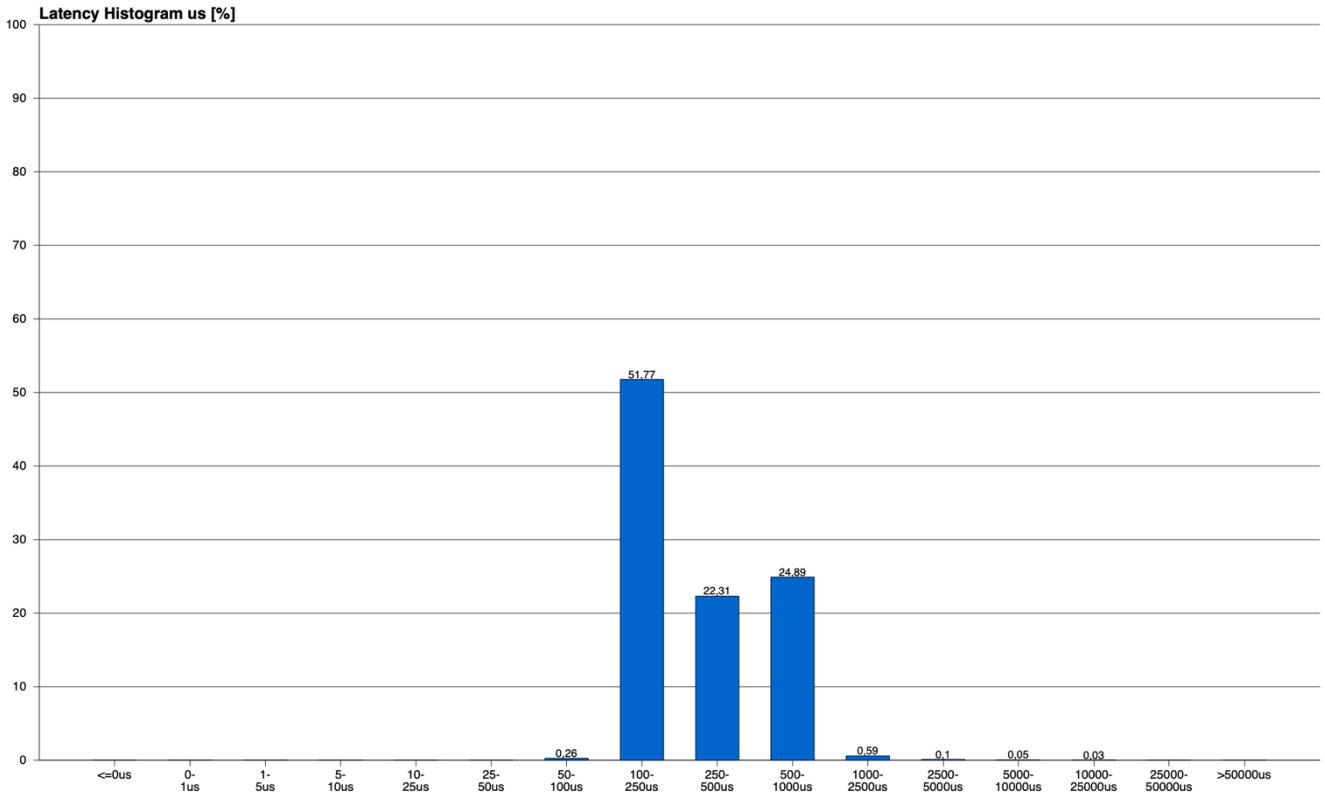


Abbildung 7 – Latency Histogram Backend Random Write (1 Prozess)

Bis zu 16 Prozessen x 32 Nodes (512 Prozesse) steigt die Verarbeitungsleistung stärker als die Latenz. Der Optimalwert liegt bei 436'959 Backend Random Writes 8 KB bei einer Latenz von 981µs. Der Maximalwert von 750'000 Backend Random Writes wird dann aber nur mit einer stark angestiegenen Latenz von 7ms erreicht. Das System ist hier nicht mehr im optimalen und kann die IOPS nicht mehr optimal schnell übertragen und auf die SSD downstagen.

Exemplarisch ist für einen Backend Random Write Tests in der folgenden Grafik der Testverlauf mit 8 Prozessen über 10 Minuten in Sekundengenauigkeit dargestellt. Zu Beginn ist die Verarbeitungsleistung höher, die Servicetime geringer, da hier der Cache noch nicht vollständig gefüllt ist. Anschliessend stabilisiert sich der Wert bei gut 400'000 IOPS und einer Servicezeit (Latenz) von 600µs.

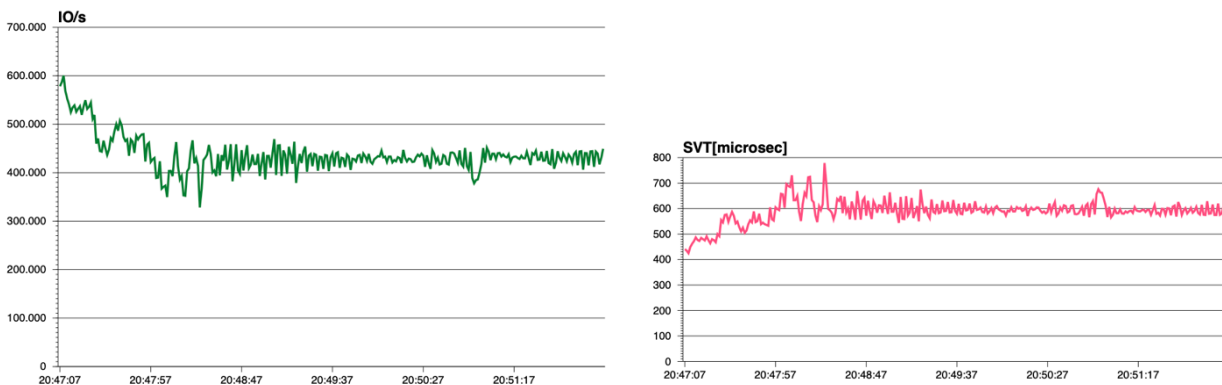


Abbildung 8 – Random Read Testverlauf (64 Prozesse)

Die CPU Belastung der Storage Controller beträgt bei den Frontend Tests 70% und bei den Backend Tests fast 90%.

Sequential Read 1 MB

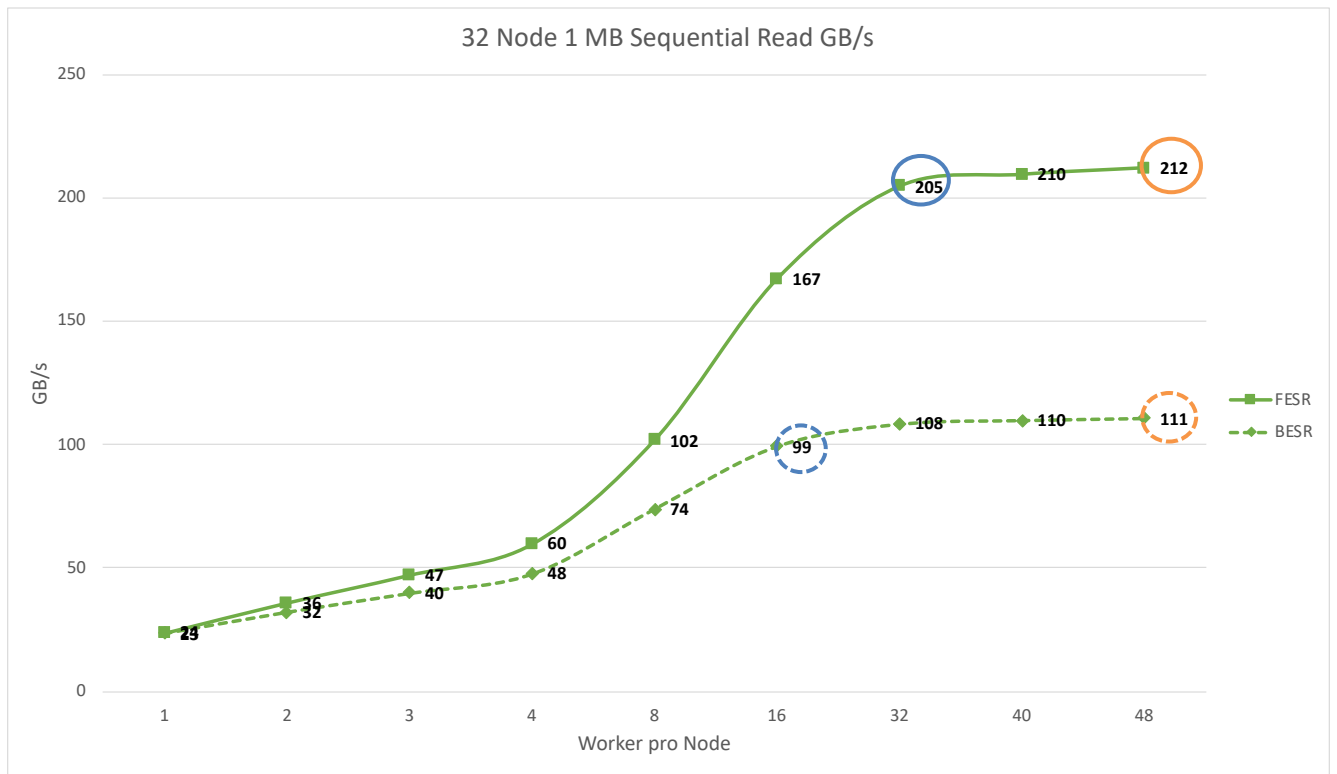


Abbildung 9 – Sequential Read Performance 128 KB

Im Frontend können die beiden Systeme zusammen **212 GB/s im Frontend** verarbeiten (mehr als 100 GB/s pro System). Die Werte sind in Gigabyte pro Sekunde nicht in Gigabit pro Sekunde angegeben. Im **Backend** können die Systeme im «Worst Case» bei 0% Cache Hit Ratio immer noch **111 GB/s** verarbeiten oder 55 GB/s pro System.

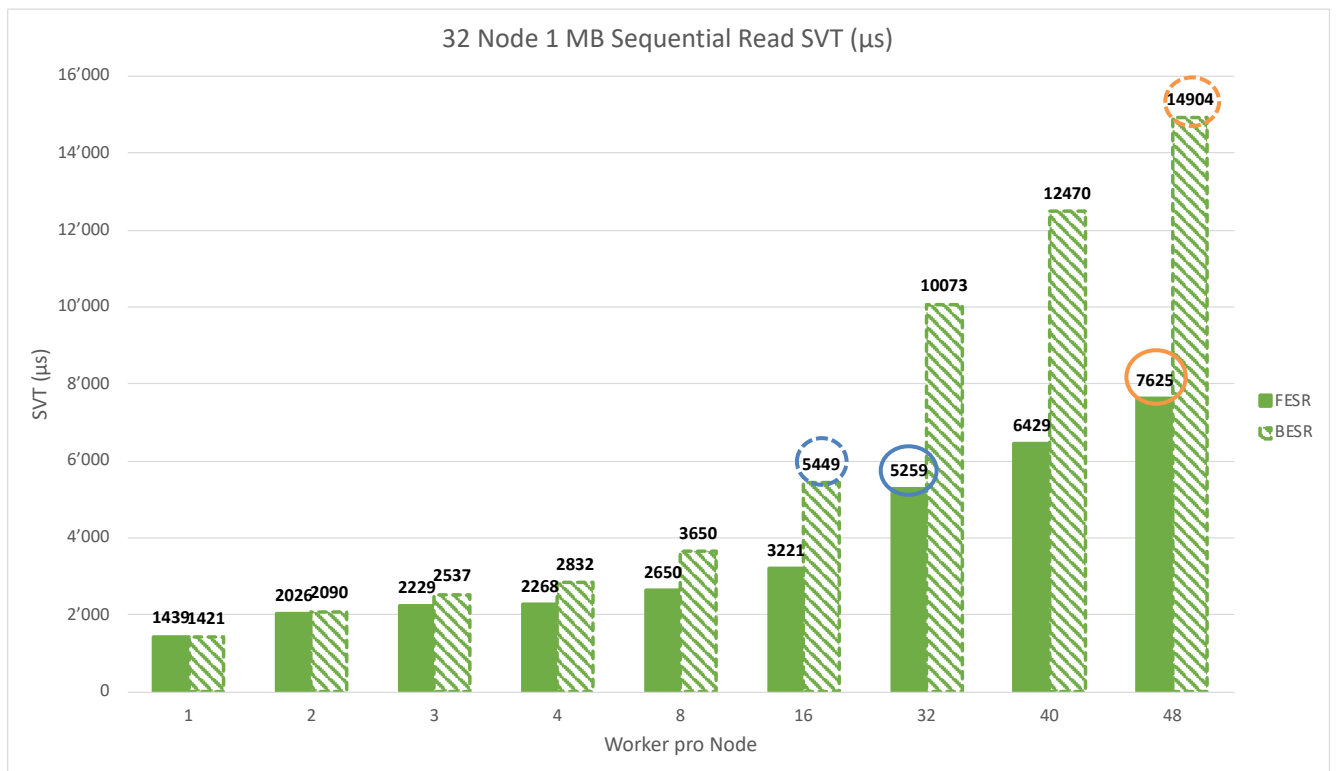


Abbildung 10 – Sequential Read Latenz 1 MB

Im Frontend und Backend beträgt die Latenz zum Lesen eines 1 MB Blocks bei geringer Parallelität ca. 1'400μs.

Im Frontend wird der optimale Durchsatz von 205 GB/s bei einer Latenz 5ms bei einer Parallelität von 32 Prozessen pro Server erreicht. Danach steigt nur noch die Latenz und der Durchsatz nur unwesentlich auf 212 GB/s.

Im Backend wird der optimale Durchsatz von 99 GB/s bei einer Latenz 5ms bei einer Parallelität von 16 Prozessen pro Server erreicht. Danach steigt nur noch die Latenz und der Durchsatz nur unwesentlich auf 111 GB/s.

Exemplarisch ist für einen Frontend Sequential Read Test in der folgenden Grafik der Testverlauf mit 16 Prozessen über 10 Minuten in Sekundengenauigkeit dargestellt. Es ist zu sehen, dass die Verarbeitungsleistung konstant bei 160 GB/s und die Service Time bei 3ms liegt.

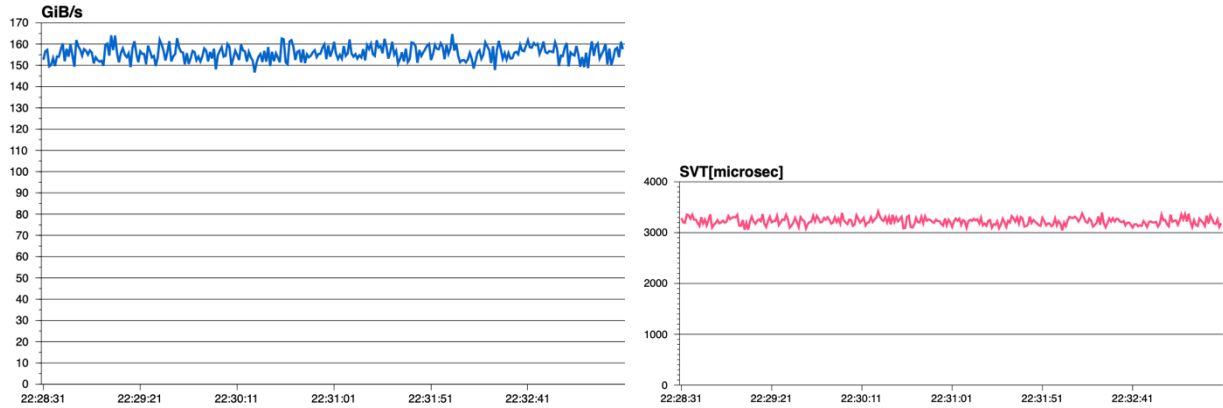


Abbildung 11 – Frontend Sequential Read Testverlauf (16 Prozesse)

Die CPU Belastung der Storage Controller ist aufgrund der viel geringeren Anzahl der IOPS deutlich geringer als bei den Random 8 KB Tests. Sie liegt beim Frontend Sequential Read bei maximal 50% und beim Backend Sequential Read bei maximal 65%.

Sequential Write 1 MB

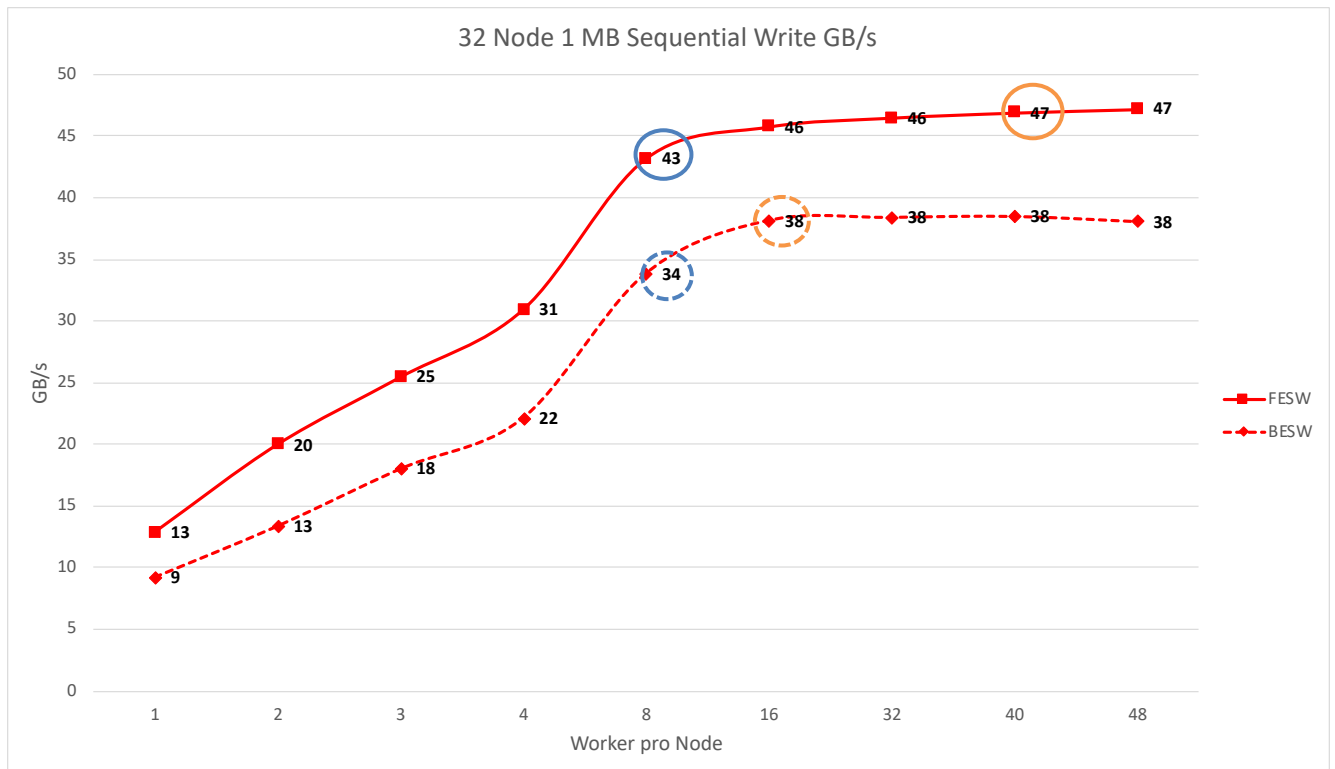


Abbildung 12 – Sequential Write Performance 1 MB

Im Frontend können die beiden Systeme zusammen knapp **47 GB/s** verarbeiten (ca. 23 GB/s pro System). Von den 47 GB/s Sequential Writes sind 50% lokal und 50% werden synchron auf den anderen Storage gespiegelt. Pro Storage werden somit 23 GB/s lokal und zusätzlich 11.5 GB/s vom anderen System abgewickelt, somit werden pro Storage 35 GB/s respektive über beide Systeme **70 GB/s** verarbeitet.

Im Backend können die Systeme im «Worst Case» bei 0% Cache Hit Ratio **38 GB/s** verarbeiten, oder 19 GB/s pro System. Der Backend Wert liegt nicht wesentlich tiefer als der Frontend Wert.

Entscheidende Beschränkung für den Write Durchsatz sind die Links zwischen den Rechenzentren, die mit 4 Links x 32 Gbit = 128 Gbit = 16 GB/s pro Richtung übertragen können. Wir erreichen im Frontend pro Richtung jeweils eine Übertragungsrate von ca. 11.5 GB/s. Die Inter-switch Links (ISL) zwischen den Standorten sind damit zu mehr als 70% ausgelastet. Bei Einsatz von mehr ISLs würden möglicherweise beim Schreiben noch höhere Durchsätze erreicht.

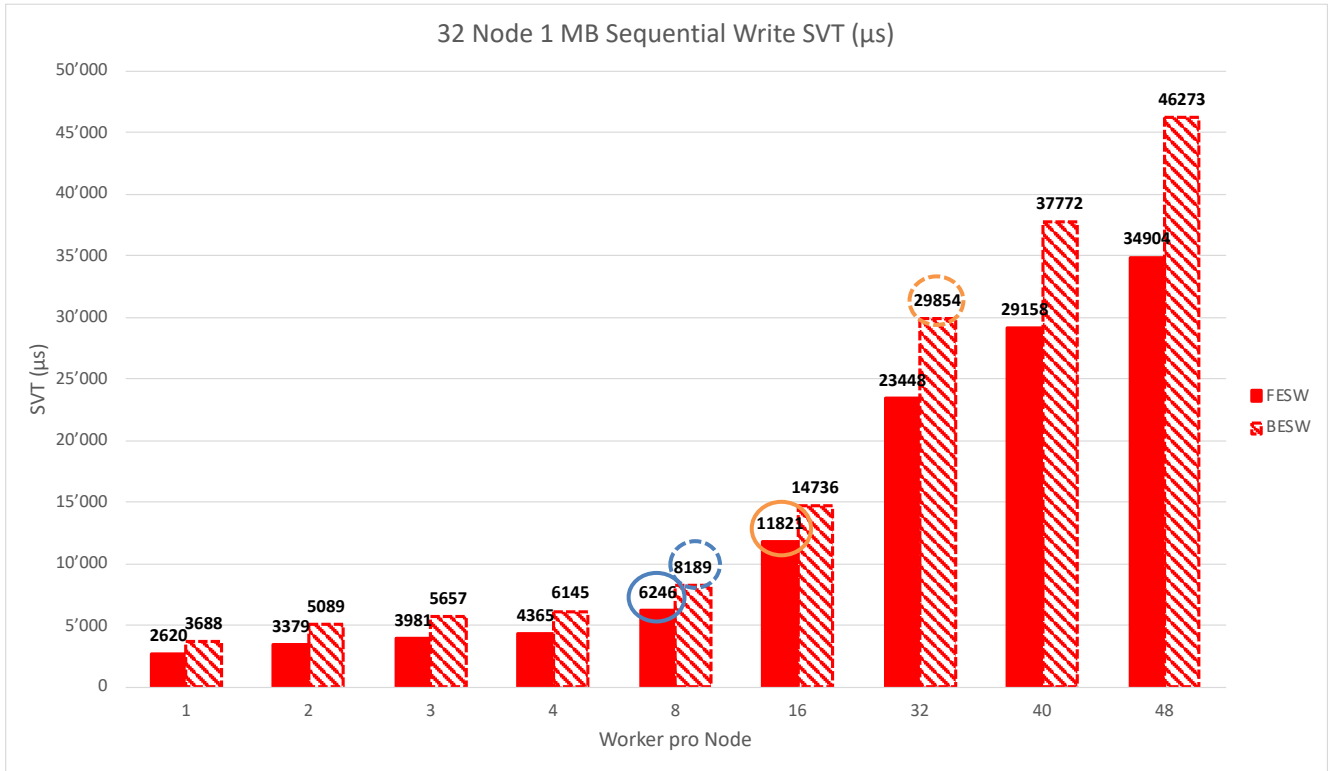


Abbildung 13 – Sequential Write Latenz 1 MB

Im Frontend beträgt die Latenz bei geringer Parallelität 2.6ms, im Backend sind es 3.7ms. Die Verarbeitungsgeschwindigkeit steigt im Frontend rasch auf 200 GB/s bei einer Latenz von 6ms und im Backend auf 100 GB/s bei einer Latenz von 8ms. Danach ist nur noch eine unwesentliche Steigerung des Durchsatzes aber ein starkes Ansteigen der Latenz zu beobachten, was im Wesentlichen an den stark ausgelasteten ISLs zwischen den Rechenzentren liegt.

Die CPU Belastung der Storage Controller erreicht bei den Frontend Tests bis zu 40% und bei den Backend Tests in der Spitze 65%.

VSP 5500 IO Analyse

Mit Hilfe des Hitachi Vantara IOportals wurden die Messwerte von IOgen mit den Kennzahlen der Storage-Systeme verglichen. Die folgende Grafik zeigt wesentliche IO Kennzahlen eines Storage-Systems.

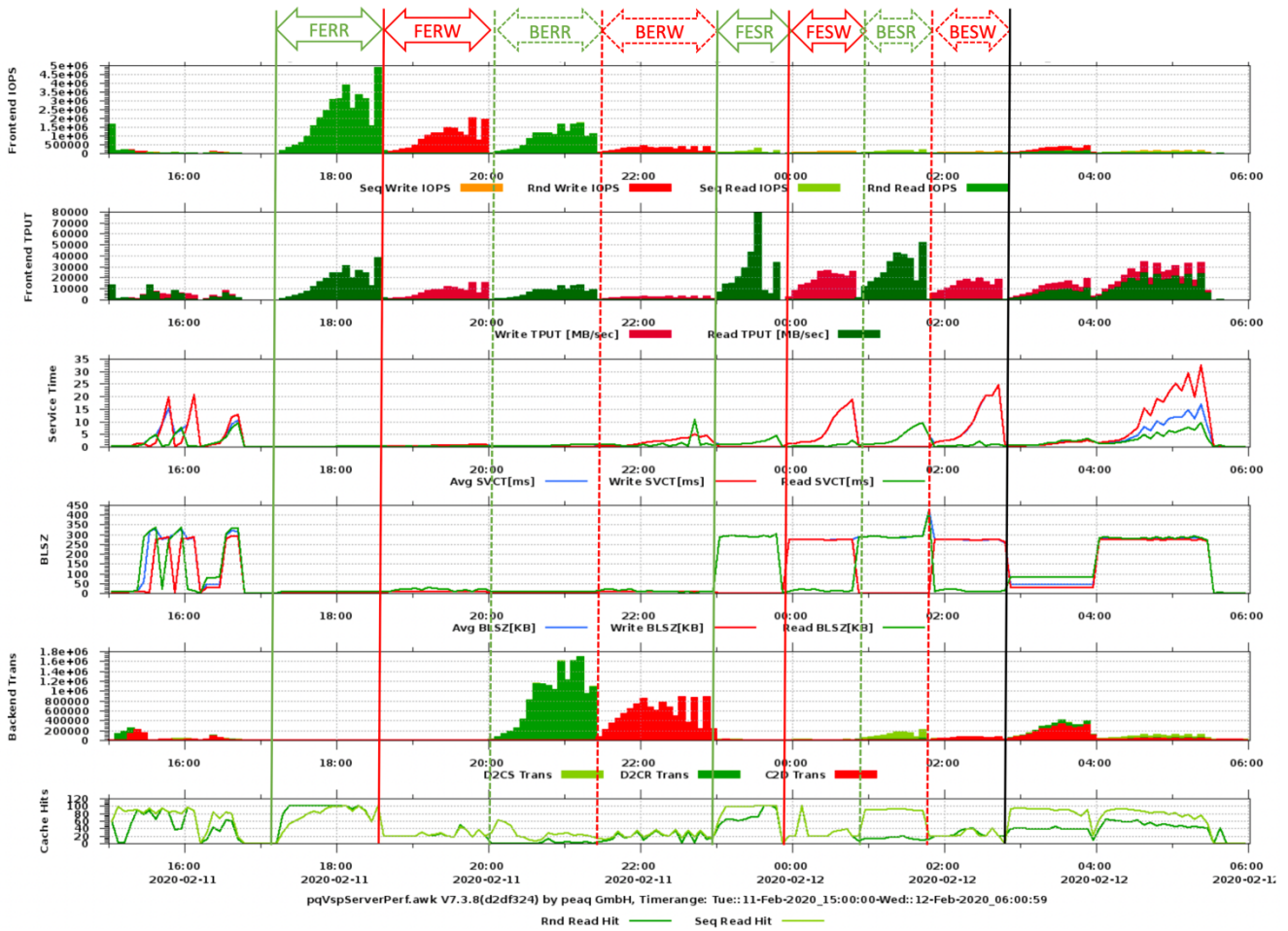


Abbildung 14 – Kennzahlen aus dem IOportal

Man erkennt die 5 Mio. IOPS während dem Frontend Random Read Test und die 1.75 Mio. IOPS während dem Backend Random Read. Ebenso korrespondieren die IOPS für die Frontend Random Write mit 1.75 Mio. IOPS und dem Backend Random Write mit 400'000 IOPS. Da dies nur Werte für ein System sind, sind diese für beide getesteten Systeme zu verdoppeln.

Weiterhin ist zu erkennen, dass die Blocksize (BLSZ) für die Random Tests sehr klein ist (8 KB) und dass nur bei den Backend Tests Transfers aus dem Backend (grün) bzw. ins Backend (rot) erfolgen.

Für die Frontend Sequential Reads sieht man bis zu 80 GB/s (real sind es etwa 100 GB/s pro System oder 200 GB/s gesamt, die Werte im IO Portal sind über 5 Minuten gemittelt) für Backend Sequential Reads 50 GB/s pro System (100 GB/s summiert). Die Zahlen für Frontend Sequential Write korrespondieren ebenfalls mit den gemessenen Zahlen.

Die Blocksize in den Sequential Tests ist auch laut IOportal 256 KB, die IOPS werden offensichtlich hostseitig in etwas kleinere IOPS unterteilt (Blocksize von IOgen war ursprünglich 1 MB). Bei den Frontend Tests finden ebenfalls keine Backend Transfers statt, dies ist nur bei den Backend Tests der Fall. Insgesamt stimmen die per IOgen gemessenen Werte mit den auf den Storage-Systemen aufgezeichneten Kennzahlen überein.

VSP 5500 Controller Auslastung

In der folgenden Grafik ist exemplarisch für eines der Storage-Systeme die CPU Belastung der einzelnen Controller für die einzelnen Testphasen dargestellt.

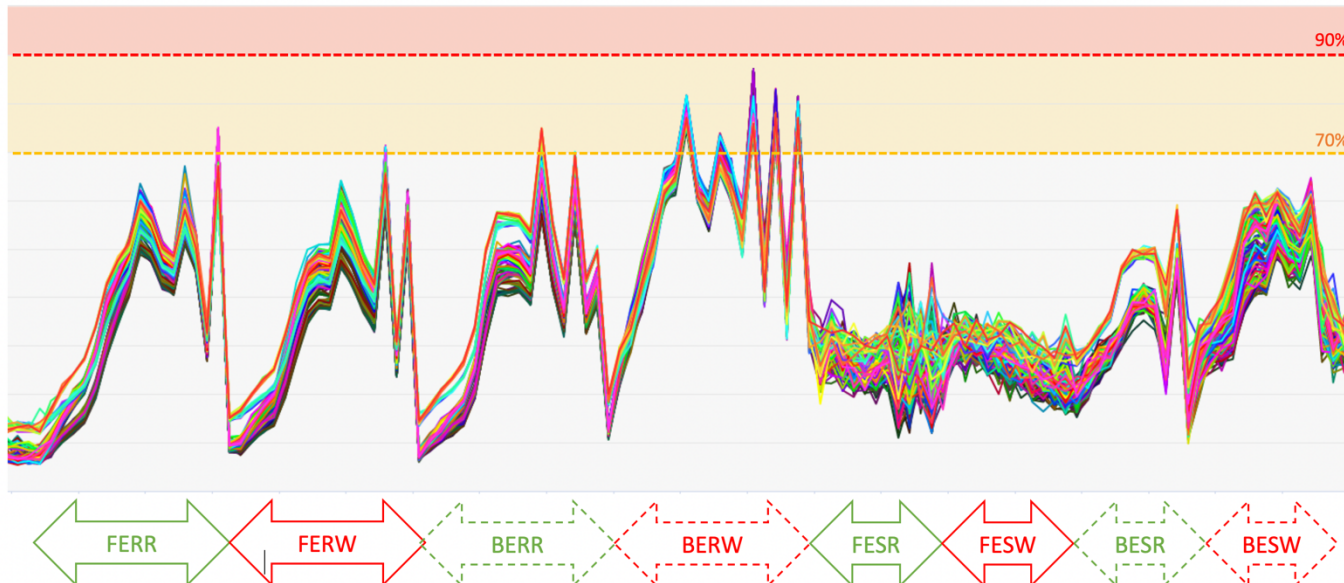


Abbildung 15 – Controllerauslastung während der Tests

Es ist gut erkennbar, dass das Storage-System die Belastung der einzelnen Controller perfekt balanciert, alle Controller sind immer gleich ausgelastet. Die Controllerauslastungen betragen bei den Random Tests mit sehr hoher Anzahl IOPS in der Spitze 70 bis 90%, die höchste Belastung weist der Backend Random Write (BERW) Test aus. Die Controller Belastung während der Sequential IO Tests mit deutlich grösseren Blocksizes ist vergleichsweise moderat und bleibt immer signifikant unter 70%.

VSP 5500 DeCo Rate

Die Deduplication und Compression Ratio hat das System nach Abschluss der Tests mit 1.85:1 angegeben. Dies liegt sogar etwas über dem erwarteten Faktor von 1.66:1.

Zusammenfassung

Die folgende Tabelle fasst die IOPS für Random IO, bzw. den Durchsatz in GB/s für Sequential IO und die Latenz zusammen. Dabei wird in der Spalte «Speed» der Wert für jeweils einen Worker pro Node angegeben (32 Worker insgesamt), in der Spalte «Optimum» der optimale Durchsatz (solange die Leistung stärker steigt als die Latenz) und in der Spalte «Maximum» der maximalen Durchsatz.

Test	Speed	Optimum	Maximum
8 KB Frontend	1 Worker	64 Worker	128 Worker
Random Read	370'148 IOPS @ 85 µs	8'144'312 IOPS @ 250 µs	10'004'125 IOPS @ 408 µs
8 KB Backend	1 Worker	64 Worker	128 Worker
Random Read	116'431 IOPS @ 273 µs	3'528'129 IOPS @ 571 µs	3'952'643 IOPS @ 1'039 µs
8 KB Frontend	1 Worker	128 Worker	128 Worker
Random Write	97'411 IOPS @ 309 µs	3'518'020 IOPS @ 1'163 µs	3'518'020 IOPS @ 1'163 µs
8 KB Backend	1 Worker	16 Worker	128 Worker
Random Write	95'543 IOPS @ 337 µs	524'136 IOPS @ 981 µs	748'997 IOPS @ 7'146 µs
1 MB Frontend	1 Prozess	32 Prozesse	48 Prozesse
Sequential Read	23.5 GB/s @ 1'439 µs	205 GB/s @ 5'259 µs	212 GB/s @ 70625 µs
1 MB Backend	1 Prozess	16 Prozesse	48 Prozesse
Sequential Read	23.7 GB/s @ 1'421 µs	99 GB/s @ 5'449 µs	111 GB/s @ 14'904 µs
1 MB Frontend	1 Prozess	8 Prozesse	40 Prozesse
Sequential Write	13 GB/s @ 2'620 µs	43 GB/s @ 6'248 µs	48 GB/s @ 29'158 µs
1 MB Backend	1 Prozess	8 Prozesse	16 Prozesse
Sequential Write	9 GB/s @ 3'688 µs	34 GB/s @ 8'189 µs	38 GB/s @ 14'763 µs

Tabelle 2 – Kennzahlen (Parallelität = Worker x 32 Server)

Die IOPS und GB/s Werte sind summiert für beide Storage-Systeme angegeben. Die hostseitigen Writes werden zu 50% auf den anderen Storage synchron gespiegelt. Die effektiv zu verarbeitenden Writes sind somit um Faktor 1.5 höher.

Fazit

Die Hitachi Vantara VSP 5500 mit 8 Controller stösst in neue Dimensionen an IO Leistung vor (jeweils summiert über beide Systeme):

- 10 Mio. Frontend Random Read IOPS 8 KB
- 3.5 Mio. Frontend Random Write IOPS 8 KB (50% synchron gespiegelt, effektiv 5.25 Mio. IOPS)
- 200 GB/s Frontend Sequential Read
- 50 GB/s Frontend Sequential Write (50% synchron gespiegelt, effektiv 75 GB/s)
- 4 Mio. Backend Random Read IOPS 8 KB
- 750'000 Backend Random Write IOPS 8 KB (50% synchron gespiegelt, effektiv 1.125 Mio. IOPS)
- 100 GB/s Backend Sequential Read
- 40 GB/s Backend Sequential Write (50% synchron gespiegelt, effektiv 60 GB/s)

Dabei ist die Write Performance durch die Anzahl der Datacenter Interconnects (ISLs) limitiert und könnte möglicherweise noch gesteigert werden. Zudem ist das System noch mit einer relativ geringen Anzahl NVMe Disks ausgestattet. Last but not least kann das System noch mit 50% Controller Kapazität erweitert werden.

Die Latenz ist mit dem NVMe Backend auf einem sehr hohen Leistungsniveau:

- 8 KB Reads aus dem Cache < 100µs
- 8 KB Reads aus dem Backend < 300µs
- 8 KB Writes in den Cache 300-350µs (Durchschnitt bei 50% synchron gespiegelten Writes).
Lokale Writes bei unter 200µs, synchron gespiegelte bei unter 500µs.

Hitachi Vantara hat hier ein neues Top Modell im Portfolio, dass neue Leistungsmassstäbe setzt. Beeindruckend war zudem, dass die Performance direkt und ohne grosses Tuning abrufbar war.

Kundenstimmen



Brian Mathis
brian.mathis@suva.ch
Head of Server & Storage Plattform

«Wir sind mit der Leistung des Hitachi Vantara VSP 5500 Storages sehr zufrieden und stossen in neue Leistungsdimensionen vor, mit der die SUVA für die nächsten Jahre sehr gut aufgestellt ist.»

Dank der wertvollen Unterstützung der In&Out AG konnten wir die Storage Evaluation sehr zielgerichtet und effizient durchführen und haben ein optimales Ergebnis erzielen können, auch in finanzieller Hinsicht.»

Über den Autor



Andreas Zallmann,
andreas.zallmann@inout.ch
CEO In&Out AG,
Seestrasse 353, 8038 Zürich
www.inout.ch

Andreas Zallmann hat Informatik an der Universität Karlsruhe studiert und ist seit dem Jahr 2000 bei der In&Out AG. Er ist verantwortlich für den Geschäftsbereich Technology und seit 2016 CEO der In&Out AG.

Die In&Out verfügt über jahrelange Praxis-Erfahrung in Architektur, Konzeption, Benchmarking und Tuning von Storage- und Systemplattformen insbesondere für Core Applikationen für Banken und Versicherungen.

Andreas Zallmann ist der Entwickler des In&Out Performance Benchmarking Tool IOgen™ und hat in den letzten Jahren sehr viele Kunden- und Hersteller-Benchmarks durchgeführt.