

IN&OUT AG

NetApp SolidFire Performance

Andreas Zallmann
CEO, In&Out AG

Version: 1.10

Datum: 24.09.2019

Klassifikation: Öffentlich

Vorbemerkung

Das vorliegende Whitepaper wurde im Auftrag der Firma NetApp unabhängig und neutral von In&Out erstellt. Die Testumgebung wurde von NetApp bereitgestellt.

Einleitung

In&Out begleitet ihre Kunden seit Jahren in den Bereichen IT Infrastruktur und Datacenter mit besonderem Fokus auf Storage.

Neben unseren Beratungsleistungen im Bereich Storage Strategie 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.

Über NetApp

NetApp ist seit mehr als einer Dekade einer der Marktführer im Storagebereich und vor allem bekannt für seine NAS Storage-Systeme (Network Attached Storage) der FAS Serie mit dem Betriebssystem ONTAP und dem Filesystem WAFL (Write Anywhere File Layout). Diese werden bereits seit Jahren auch als FC-basierte Block-storage-Systeme als «Unified Storage» angeboten.

NetApp wird von Gartner seit Jahren im Leader Quadrant des Magic Quadrant für «General Purpose Storage» gelistet und seit 2018 sogar als Leader geführt.

Ab 2019 werden der General Purpose Storage und Full Flash Storage neu unter «Magic Quadrant for Primary Storage» zusammengefasst und NetApp ist hier von Gartner ebenfalls als Leader positioniert.



Abbildung 1 – Gartner Magic Quadrant Primary Storage (2019)

Mit dem Zukauf der Firma SolidFire im Jahr 2016 verfügt NetApp über ein alternatives Storageprodukt. Die SolidFire Produktlinie skaliert über kleine Nodes von nur 1U Höhe und 10-12 Disk Einschüben. Die Nodes unterscheiden sich durch Cache Size und die verbauten Disk-

typen. In der aktuellen Produktlinie sind Nettokapazitäten bis zu knapp 50 TB auf 1U erreichbar. Pro SolidFire Node sollen dabei bis zu 100'000 IOPS erreicht werden. Ähnlich wie beim Konzept der Hyperconverged Infrastructure soll dabei die Kapazität einfach durch Hinzufügen weiterer Nodes erhöht werden können und die Hochverfügbarkeit ebenso über die Nodes hinweg bewerkstelligt werden. Die SolidFile Storage Nodes kombiniert mit Compute Nodes ergeben die NetApp HCI Lösung, die allerdings im Unterschied zu herkömmlichen HCI Lösungen, den Storage und Compute trennt. NetApp nennt diese Lösung deshalb nicht Hyperconverged Infrastructure, sondern Hybrid Cloud Infrastructure (ebenfalls HCI, was die Nähe des Lösungsansatzes verdeutlicht).

Im Gartner Magic Quadrant Hyperconverged Infrastructure ist NetApp nicht aufgeführt, da die NetApp Hybrid Cloud Infrastructure aufgrund der Trennung von Storage und Compute für Gartner keine HCI Lösung im engeren Sinne darstellt. Hingegen führt IDC die NetApp Lösung als «Disaggregated HCI».

Zielsetzung

Das vorliegende Whitepaper hat die Block-Performance der SolidFire mit 8 Storage Nodes SF19210 im Fokus.

Management Summary

Die NetApp SolidFire Lösung mit 8 Nodes vom Typ SF19210 erreicht ebenfalls beeindruckende Leistungskennzahlen, die bereits deutlich mehr als die Hälfte der Leistung der «traditionellen» All Flash Lösung» A800 mit einem HA-Pair mit NVMe Disks ausmachen. Die Solid Fire kann von jetzt eingesetzten 8 Nodes auf 40 skalieren, was im Idealfall bei linearer Skalierung bis zu Faktor 5 höhere Performance bedeuten würde.

Die pro Node in Aussicht gestellten 100'000 IOPS zu 4K wurden für die getestete Konfiguration mit 8 Nodes mit über 690k 4K IOPS fast erreicht, bis zu 8 Nodes skaliert die Lösung also fast linear.

| Test | Speed | Optimum | Maximum |
|--------------|------------|--------------|--------------|
| 4 KB Read | 1 Prozess | 128 Prozesse | 256 Prozesse |
| Random | 8'813 OPS | 605'534 IOPS | 690'677 IOPS |
| Backend | 450 µs | 844 µs | 1'371 µs |
| 4 KB Write | 1 Prozess | 96 Prozesse | 512 Prozesse |
| Random | 7'096 IOPS | 315'431 IOPS | 473'950 IOPS |
| Backend | 561 µs | 1'236 µs | 4'362 µs |
| 8 KB Read | 1 Prozess | 128 Prozesse | 192 Prozesse |
| Random | 8'137 IOPS | 523'586 IOPS | 570'825 IOPS |
| Backend | 484 µs | 975 µs | 1'371 µs |
| 8 KB Write | 1 Prozess | 96 Prozesse | 512 Prozesse |
| Random | 6'101 IOPS | 225'410 IOPS | 317'536 IOPS |
| Backend | 651 µs | 1'739 µs | 6'633 µs |
| 128 KB Read | 1 Prozess | 32 Prozesse | 48 Prozesse |
| Sequential | 0.4 GB/s | 6.9 GB/s | 7.2 GB/s |
| Backend* | 957 µs | 1'843 µs | 2'621 µs |
| 128 KB Write | 1 Prozess | 24 Prozesse | 64 Prozesse |
| Sequential | 0.35 GB/s | 4.4 GB/s | 5.2 GB/s |
| Backend | 1'479 µs | 2'880 µs | 6'601 µs |

Tabelle 1 – Kennzahlen (bei 4 VMs)

Diese Leistungskennzahlen wurden mit Ethernet und iSCSI erreicht. Zu berücksichtigen ist, dass bei der SolidFile bis zu 40 Blocks zusammengefügt werden können, also theoretisch die Performance bei linearer Skalierung um bis zu Faktor 5 höher liegen könnte.

Konfiguration SolidFire 8 Node SF19210

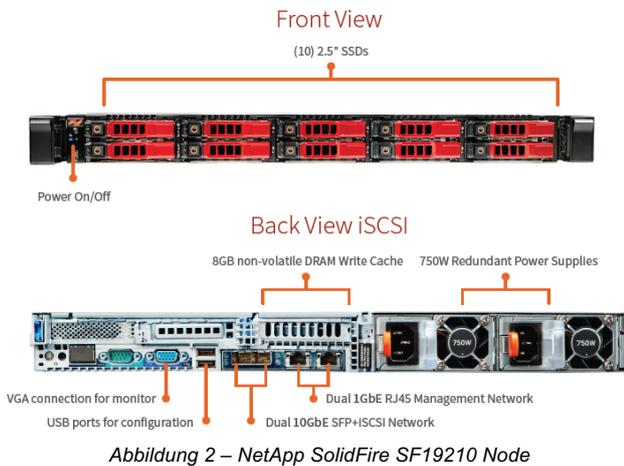


Abbildung 2 – NetApp SolidFire SF19210 Node

Getestet wurde ein Verbund aus 8 SolidFire SF19210 Nodes. Jeder Node ist mit 2 CPUs mit je 14 Cores zu 2.6 GHz und 348 GB Memory sowie 8 GB NVRAM ausgestattet. Pro SolidFire Node sind 2 x 10 GbE Interfaces vorhanden.

Insgesamt wurden pro Node 10 x 1.92 TB = 19 TB Rohkapazität verbaut, insgesamt steht somit für die 8 Nodes eine Kapazität von gut 150 TB zur Verfügung.

Aus diesem Diskpool wurden insgesamt 56 LUNs mit einer Grösse von je 1 TB generiert, von denen jeweils 14 LUNs an jeden der drei Testserver gemappt wurden.

Konfiguration Testserver

Als Lastgeneratoren wurden bis zu 4 Lenovo Server vom Typ SR650 mit 2 Intel Xeon Gold 6132 CPUs (je 14 Cores, 2.60 Ghz) und 384 GB DDR4 Memory (2666 Mhz) verwendet.

Die Server waren jeweils mit 2 x 10 Gbit Ethernet via iSCSI an den Storage angebunden.

Auf den Testservern kam VMware ESX 6.7 U1 und auf den VMs RedHat Enterprise Linux RHEL 7.5 als Betriebssystem zum Einsatz.

Testläufe

Die Testläufe wurden jeweils parallel und durch IOgen™ zeitlich genau synchronisiert auf vier VMs durchgeführt, die auf die vier verfügbaren SR650 Testserver verteilt wurden. Jeder der vier VMs konnte einen Testserver exklusiv nutzen.

Es wurden Testläufe mit 1 VM (gelb), 2 VMs (blau), 3 VMs (grün) und 4 VMs (rot) durchgeführt. Dabei war jede VM mit 48 vCPU konfiguriert. Die VMs wurden jeweils auf einem dedizierten physischen Server gestartet.

Random Read 4 KB

Die Ergebnisse der Random Read Tests sind im Folgenden dargestellt. Dabei wurde eine Blockgrösse von 4 KB verwendet und die gesamte konfigurierte Storagekapazität von 56 TB zufällig gelesen, die Daten müssen aus dem Storagebackend gelesen werden, die Cache Hit Ratio ist nahezu Null.

Dabei wurde synchronisiert auf 1,2,3 und 4 VMs die Anzahl der Leseprozesse (IOgen Prozesse) von 1 bis maximal 512 pro VM gesteigert. Der Höchstwert wurde jeweils bei einer Parallelität von 196 oder 256 Prozessen erreicht.

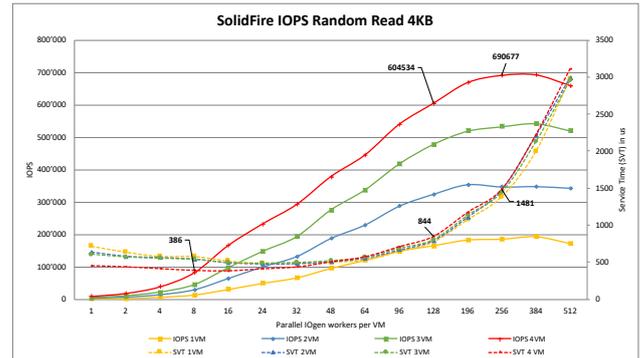


Abbildung 3 – Random Read Performance 4 KB

In der Grafik aufgeführt sind jeweils die IOPS Kurven (linke Skala, durchgezogene Linien) und die dazugehörigen Latencies (gemessen auf dem Server in Microsekunden µs, rechte Skala gestrichelte Linien).

Pro VM werden ca. 190'000 IOPS erreicht, die IOPS skalieren beinahe linear mit den VMs. Leider standen maximal 4 VMs und Server zur Verfügung, somit konnten wir nicht testen, ob die Solid Fire Storage Systeme sogar noch weiter skalieren.

Die minimale Latenz beträgt 386 µs, mit denen aus dem Storage Backend gelesen werden kann.

Die Latenz bleibt bei Erhöhung der Parallelität bis zu 64 Workern pro VM absolut konstant und steigt dann mit zunehmender Sättigung des Stages an.

Der Optimalwert von 605'534 IOPS mit 4 VMs wird bei einer Parallelität von 128 Workern pro VM mit einer Latenz von guten 844µs erreicht.

Anschliessend steigt die Latency schneller wie die IOPS. Der Maximalwert von 690'677 IOPS mit 4 VMs wird bei einer Parallelität von 256 Workern pro VM mit einer Latenz von noch immer akzeptablen 1'481µs erreicht.

Bei höheren Parallelitäten wird die IO Leistung leicht geringer, die Latencies steigen massiv. Hier befinden wir uns klar im Überlastbereich.

Random Write 4 KB

Die Ergebnisse der Random Write Tests sind im Folgenden dargestellt. Dabei wurde ebenfalls eine Blockgrösse von 4 KB verwendet und die gesamte konfigurierte Storagekapazität von 56 TB zufällig geschrieben. Die Daten werden zwar im Storage Cache zwischen gespeichert, müssen aber sehr rasch in das Backend verschoben werden, da permanent verschiedene Blöcke geschrieben werden und der Cache in kürzester Zeit gefüllt ist.

Dabei wurden synchronisiert auf 1,2,3 und 4 VMs die Anzahl der Schreibprozesse (IOgen Prozesse) von 1 bis maximal 512 pro VM gesteigert. Selbst bei dieser hohen Parallelität steigen die IOPS immer noch an, allerdings weit weniger stark als die Latenz.

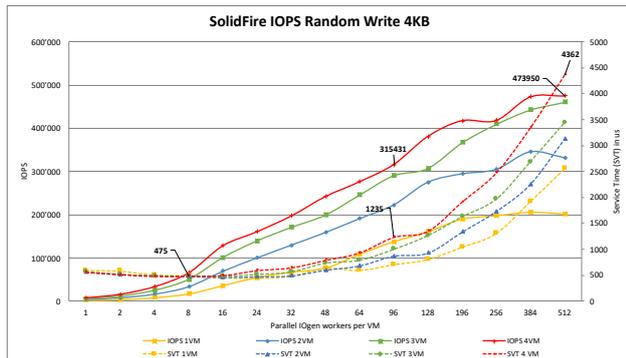


Abbildung 4 – Random Write Performance 8 KB

Pro VM werden ca. 200'000 IOPS erreicht, die IOPS nehmen allerdings nicht linear mit der Anzahl der VMs zu. Die Werte mit 4 VMs sind nur marginal höher als mit 3 VMs. Die VMs können das Storage-System hier auslasten.

Die minimale Latenz beträgt 475μs, mit denen auf den Storage geschrieben werden kann. Diese Latenz wird auf dem Server gemessen, nicht auf dem Storage. Dies ist die optimale Latenz, wenn der Storage freie Write Cache Kapazität verfügbar hat und den IO in den Cache schreibt.

Die Latenz bleibt bei Erhöhung der Parallelität bis zu 64 Workern pro VM einigermaßen konstant und steigt dann mit zunehmender Sättigung des Storage stärker an.

Der Optimalwert von 315'431 IOPS mit 4 VMs wird bei einer Parallelität von 96 Workern pro VM mit einer Latenz von guten 1'235μs erreicht.

Anschliessend steigt die Latency deutlich schneller wie die IOPS. Der Maximalwert von 473'950 IOPS mit 4 VMs wird bei einer Parallelität von 512 Workern pro VM mit einer Latenz von 4'362μs erreicht. Dabei ist der Storage bereits klar im Überlastbereich.

Random Read 8 KB

Die Ergebnisse der Random Read Tests sind im Folgenden dargestellt. Dabei wurde eine Blockgrösse von 8 KB verwendet und die gesamte konfigurierte Storagekapazität von 56 TB zufällig gelesen, die Daten müssen aus

dem Storagebackend gelesen werden, die Cache Hit Ratio ist nahezu Null.

Dabei wurde synchronisiert auf 1,2,3 und 4 VMs die Anzahl der Leseprozesse (IOgen Prozesse) von 1 bis maximal 512 pro VM gesteigert. Der Höchstwert wurde jeweils bei einer Parallelität von 196 oder 256 Prozessen erreicht.

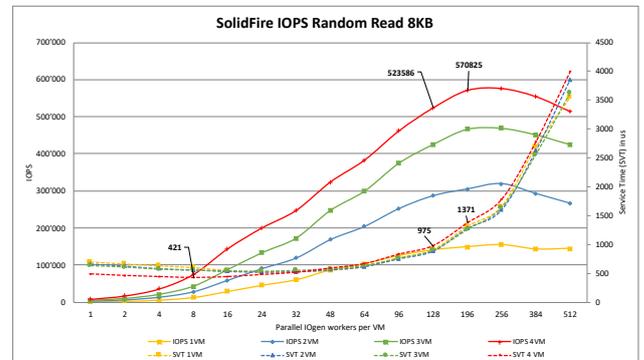


Abbildung 5 – Random Read Performance 8 KB

Pro VM werden ca. 150'000 IOPS erreicht, die IOPS skalieren beinahe linear mit den VMs. Leider standen maximal 4 VMs und Server zur Verfügung, somit konnten wir nicht testen, ob die Solid Fire Storage Systeme sogar noch weiter skalieren.

Die minimale Latenz beträgt 421 μs, mit denen aus dem Storage Backend gelesen werden kann.

Die Latenz bleibt bei Erhöhung der Parallelität bis zu 64 Workern pro VM absolut konstant und steigt dann mit zunehmender Sättigung des Storage an.

Der Optimalwert von 523'586 IOPS mit 4 VMs wird bei einer Parallelität von 128 Workern pro VM mit einer Latenz von guten 975μs erreicht.

Anschliessend steigt die Latency schneller wie die IOPS. Der Maximalwert von 570'825 IOPS mit 4 VMs wird bei einer Parallelität von 196 Workern pro VM mit einer Latenz von noch immer akzeptablen 1'371μs erreicht.

Bei höheren Parallelitäten wird die IO Leistung leicht geringer, die Latencies steigen massiv. Hier befinden wir uns klar im Überlastbereich.

Random Write 8 KB

Die Ergebnisse der Random Write Tests sind im Folgenden dargestellt. Dabei wurde ebenfalls eine Blockgrösse von 8 KB verwendet und die gesamte konfigurierte Storagekapazität von 56 TB zufällig geschrieben. Die Daten werden zwar im Storage Cache zwischen gespeichert, müssen aber sehr rasch in das Backend verschoben werden, da permanent verschiedene Blöcke geschrieben werden und der Cache in kürzester Zeit gefüllt ist.

Dabei wurden synchronisiert auf 1,2,3 und 4 VMs die Anzahl der Schreibprozesse (IOgen Prozesse) von 1 bis maximal 512 pro VM gesteigert. Selbst bei dieser hohen Parallelität steigen die IOPS immer noch an, allerdings weit weniger stark als die Latenz.

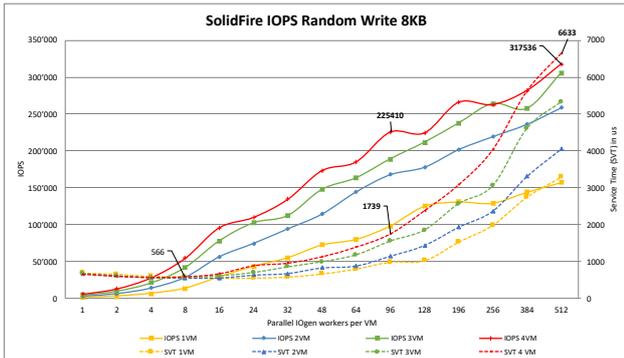


Abbildung 6 – Random Write Performance 8 KB

Pro VM werden ca. 150'000 IOPS erreicht, die IOPS nehmen allerdings nicht linear mit der Anzahl der VMs zu. Die Werte mit 4 VMs sind nur marginal höher als mit 3 VMs. Die VMs können das Storage-System hier auslasten.

Die minimale Latenz beträgt 566µs, mit denen auf den Storage geschrieben werden kann. Diese Latenz wird auf dem Server gemessen, nicht auf dem Storage. Dies ist die optimale Latenz, wenn der Storage freie Write Cache Kapazität verfügbar hat und den IO in den Cache schreibt.

Die Latenz bleibt bei Erhöhung der Parallelität bis zu 64 Workern pro VM einigermaßen konstant und steigt dann mit zunehmender Sättigung des Storages stärker an.

Der Optimalwert von 225'410 IOPS mit 4 VMs wird bei einer Parallelität von 96 Workern pro VM mit einer Latenz von guten 1'739µs erreicht.

Anschließend steigt die Latency deutlich schneller wie die IOPS. Der Maximalwert von 315'537 IOPS mit 4 VMs wird bei einer Parallelität von 512 Workern pro VM mit einer Latenz von 6'633µs erreicht. Dabei ist der Storage bereits klar im Überlastbereich.

Sequential Read 128 KB

Die Ergebnisse der Sequential Read Tests sind im Folgenden dargestellt. Dabei wurde eine Blockgrösse von 128 KB verwendet und die Storagekapazität pro Worker an verschiedenen Stellen sequentiell gelesen. Die Daten befinden sich mit grosser Wahrscheinlichkeit nicht im Cache, allerdings kann der Storage das sequentielle Pattern erkennen und theoretisch einen Read Ahead durchführen. Im Zeitalter von SSD Disks spielt dieser Effekt allerdings kaum noch eine Rolle.

Dabei wurden synchronisiert auf 1,2,3 und 4 VMs die Anzahl der Leseprozesse (IOgen Prozesse) von 1 bis maximal 64 pro VM gesteigert. Dabei wird unabhängig von der Anzahl der VMs bei 32 Prozessen die Sättigung erreicht.

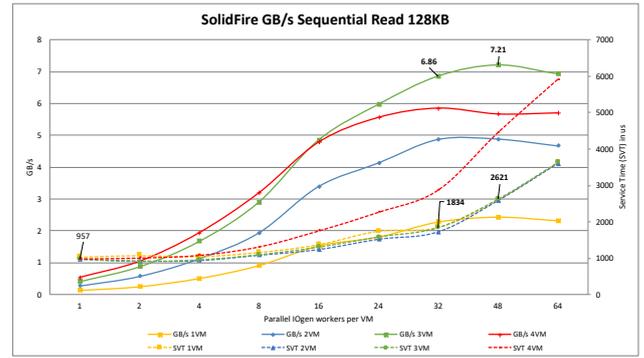


Abbildung 7 – Sequential Read Performance 128 KB

Pro VM werden ca. 2.5 GB/s erreicht, der Durchsatz steigert sich allerdings nicht linear mit der Anzahl der VMs. Die Werte mit 4 VMs sind sogar geringer als die mit 3 VMs. Die 3 VMs können das Storage-System hier auslasten.

Sequential Reads können bei geringer Parallelität mit einer Latenz von 957µs vom Storage gelesen werden. Diese Latenz ist sehr gering für eine Blockgrösse von 128 KB.

Die Latenz steigt bei Erhöhung der Parallelität bis zu 32 Workern pro VM langsam an. Danach kann keine wesentliche Steigerung des Durchsatzes mehr erreicht werden. Der Optimalwert liegt bei 6.9 GB/s bei einer Latenz von 1'834µs.

Der Maximalwert von 7.2 GB/s wird bei einer Parallelität von 48 Workern pro VM bei 3 VMs mit einer Latenz von 2'621µs erreicht.

Sequential Write 128 KB

Die Ergebnisse der Sequential Write Tests sind im Folgenden dargestellt. Dabei wurde ebenfalls eine Blockgrösse von 128 KB verwendet und die Storagekapazität pro Worker an verschiedenen Stellen sequentiell geschrieben. Die Daten werden zwar im Storage Cache zwischen gespeichert, müssen aber sehr rasch in Backend verschoben werden, da immer neue Blöcke geschrieben werden.

Dabei wurden synchronisiert auf 1,2,3 und 4 VMs die Anzahl der Schreibprozesse (IOgen Prozesse) von 1 bis maximal 64 gesteigert. Bei 1 VM wird ein maximaler Durchsatz von ca. 3.5 GB/s erreicht, der aber nicht linear beim Einsatz mehrerer VMs steigt. Bei 4 VMs ist der Durchsatz nur leicht höher als mit 3 VMs, der Storage kann somit mit den VMs voll ausgelastet werden.

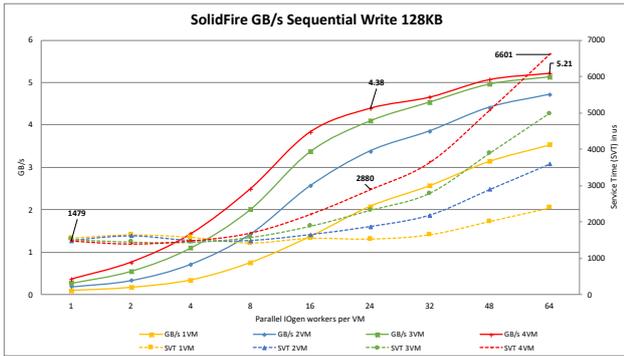


Abbildung 8 – Sequential Write Performance 128 KB

Sequential Writes können bei geringer Parallelität mit einer Latenz von 1.5ms zum Storage übermittelt werden. Dies ist die optimale Latenz, wenn der Storage freie Write Cache Kapazität verfügbar hat und den IO in den Cache schreibt.

Die Latenz steigt bei Erhöhung der Parallelität bis zu 64 Workern kontinuierlich an, aber auch der Durchsatz steigt kontinuierlich an. Ab 24 Workern sieht man Anzeichen einer Sättigung.

Der Optimalwert von 4.4 GB/s wird bei einer Parallelität von 24 Workern pro VM mit einer Latenz von 2.8ms erreicht.

Danach steigt die IO Leistung geringer an und die Latenz nimmt stärker zu. Der maximale Durchsatz wird bei einer Parallelität von 64 Prozessen pro VM mit 5.2 GB/s erreicht. Dabei wird eine Latenz von 6.6ms gemessen.

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 der Wert jeweils für einen Worker pro VM (Speed), den optimalen Durchsatz (solange die Leistung stärker steigt als die Latenz) und den maximalen Durchsatz aufgeführt.

Die pro Node in Aussicht gestellten 100'000 IOPS zu 4K wurden für die getestete Konfiguration mit 8 Nodes mit über 690k 4K IOPS fast erreicht, bis zu 8 Nodes skaliert die Lösung also fast linear.

| Test | Speed | Optimum | Maximum |
|-------------|------------|--------------|--------------|
| 4 KB Read | 1 Prozess | 128 Prozesse | 256 Prozesse |
| Random | 8'813 OPS | 605'534 IOPS | 690'677 IOPS |
| Backend | 450 µs | 844 µs | 1'371 µs |
| 4 KB Write | 1 Prozess | 96 Prozesse | 512 Prozesse |
| Random | 7'096 IOPS | 315'431 IOPS | 473'950 IOPS |
| Backend | 561 µs | 1'236 µs | 4'362 µs |
| 8 KB Read | 1 Prozess | 128 Prozesse | 192 Prozesse |
| Random | 8'137 IOPS | 523'586 IOPS | 570'825 IOPS |
| Backend | 484 µs | 975 µs | 1'371 µs |
| 8 KB Write | 1 Prozess | 96 Prozesse | 512 Prozesse |
| Random | 6'101 IOPS | 225'410 IOPS | 317'536 IOPS |
| Backend | 651 µs | 1'739 µs | 6'633 µs |
| 128 KB Read | 1 Prozess | 32 Prozesse | 48 Prozesse |
| Sequential | 0.4 GB/s | 6.9 GB/s | 7.2 GB/s |
| Backend* | 957 µs | 1'843 µs | 2'621 µs |

| | | | |
|--------------|-----------|-------------|-------------|
| 128 KB Write | 1 Prozess | 24 Prozesse | 64 Prozesse |
| Sequential | 0.35 GB/s | 4.4 GB/s | 5.2 GB/s |
| Backend | 1'479 µs | 2'880 µs | 6'601 µs |

Tabelle 2 – Kennzahlen (bei bis zu 4 VMs)

Es ist zu beachten, dass die Anzahl der Prozesse pro VM angegeben ist und immer bis zu vier VMs parallel gearbeitet haben. Die erreichten IOPS und GB/s Werte sind immer summiert für alle VMs angegeben.

Fazit

Der SolidFire Storage erreicht in der getesteten Konfiguration von 8 Nodes sehr gute Werte, die vor allem im Random Read gar nicht weit weg vom Highend All Flash Storage A800 in einer Dual Controller Konfiguration liegen. Die 8 Solid Fire Nodes erreichen 570k 8KB Random Reads (A800 730k).

Die 8K Random Writes liegen bei 317k (A800 570k). Der Durchsatz beim Lesen beträgt bis zu 6.6 GB/s (A800 10 GB/s) und beim Schreiben 4.3 GB/s (A800 10 GB/s).

Die Latencies sind auf der SolidFire bei kleinen Blöcken etwas höher als auf der A800, aber bei grösseren Blöcken sogar geringer.

Zu berücksichtigen ist, dass bei der SolidFile bis zu 40 Nodes zusammengefügt werden können, also theoretisch die Performance bei linearer Skalierung um bis zu Faktor 5 höher liegen könnte.

Über den Autor



Andreas Zallmann,
andreas.zallmann@inout.ch
 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.