IT CONSULTING & ENGINEERING

IN&OUT AG

NetApp AFA A800 Performance

Andreas Zallmann
CEO, In&Out AG

Version:	1.10
Datum:	24.09.2019
Klassifikation:	Öffentlich

In&Out AG IT Consulting & Engineering
Seestrasse 353, CH-8038 Zürich
Phone +41 44 485 60 60

info@inout.ch, www.inout.ch

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 Storagesysteme (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 Blockstoragesysteme 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.

Figure 1. Magic Quadrant for Primary Storage

Abbildung 1 – Gartner Magic Quadrant Primary Storage (2019)

Die neuesten Generationen sind All Flash Arrays der A-Series. Das performancestärkste Modell ist dabei das hier getestete A800 System, welches über ein state-of-the-Art NVMe Backend verfügt und somit auf höchste Performance ausgelegt sind.

Zielsetzung

Das vorliegende Whitepaper hat die die Block-Storageperformance der NetApp AllFlash Lösung AFA A800 im Fokus.

Management Summary

Der NetApp All Flash Storage A800 erreicht bereits in der «Minimalkonfiguration», mit zwei Controller Nodes beeindruckende Leistungskennzahlen, die im Highend Segment liegen. Von dem hier getesteten Dual Controller Setup lassen sich bis zu 12 Controller Paare zusammen clustern.

Test	Speed	Optimum	Maximum
8 KB Read	1 Prozess	128 Prozesse	384 Prozesse
Random	11'308 IOPS	648'293 IOPS	732'767 IOPS
Backend	261 µs	590 µs	1'571 µs
8 KB Write	1 Prozess	64 Prozesse	512 Prozesse
Random	9'963 IOPS	296'158 IOPS	572'022 IOPS
Backend	297 µs	647 µs	2'833 µs
128 KB Read	1 Prozess	16 Prozesse	16 Prozesse
Sequential	2.8 GB/s	11.5 GB/s	11.5 GB/s
Backend	1'141 µs	4'419 µs	4'419 µs
128 KB Write	1 Prozess	24 Prozesse	32 Prozesse
Sequential	1.2 GB/s	10.4 GB/s	11.0 GB/s
Backend	2'446 µs	7'385 µs	9'416 µs

Tabelle 1 – Kennzahlen

Diese Leistungskennzahlen wurden mit Ethernet und iSCSI erreicht, bei Verwendung von Fiber Channel (FC) kann ggf. noch mehr Leistung erreicht werden, da FC etwas weniger Protokolloverhead aufweist. Nochmals massiv besser dürfte der Einsatz von NVMeoF sein, auf den der A800 Storage bereits vorbereitet ist.

Konfiguration NetApp A800

Abbildung 2 – NetApp AFA A800 Front mit Blende

Getestet wurde das performancestärkste Modell A800 der AFA A-Series. Das System ist mit zwei Controller Nodes (auch HA-Pair genannt) auf einer Höhe von nur 4 Rackunits (4U) mit 48 NVMe Disks zu je 1.9 TB ausgestattet. Dies entspricht einer Rohkapazität von knapp 91 TB.

Jeder der beide Controller hat 2 CPUs mit je 24 Intel Cores 2.4 Ghz (insgesamt 96 Cores im HA-Pair). Das System war mit 1'280 GB Cache Memory und 64 GB NVRAM ausgestattet. Die A800 war mit 8 x 40 GBit über iSCSI an das Netzwerk angebunden.

Die A800 Serie kann in einem Cluster auf bis zu 12 HA-Pairs mit insgesamt bis zu 24 Controllern skalieren. In unserem Test haben wir die Performance von nur einem HA-Pair mit 2 Controllern gemessen.

Abbildung 3 – NetApp AFA A800 Front ohne Blende mit 48 NVMe Disks

Es wurden insgesamt 42 LUNs mit einer Grösse von je 1 TB generiert, von denen jeweils 14 LUNs an jeden der drei Testserver gemappt wurden.

Abbildung 4 – NetApp AFA A800 Rückseite

Konfiguration Testserver

Als Lastgeneratoren wurden 3 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 drei VMs durchgeführt, die auf die drei verfügbaren SR650 Testserver verteilt wurden. Jeder der drei VMs konnte einen Testserver exklusiv nutzen.

In einem ersten Testlauf (gelbe Kurven) wurden 4 vCPUs pro VM genutzt. Der Test wurde mit einer deutlich stärkeren Konfiguration mit 48 vCPUs pro VM wieder holt (blaue Kurven).

Bereits mit der kleinen Konfiguration konnten sehr gute Ergebnisse erzielt werden, mit der Konfiguration mit 48 vCPUs wurde mit nochmals höherer Parallelität versucht, das absolute Maximum aus dem Storage herauszuholen.

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 42 TB zufällig gelesen, die Daten müssen aus dem Storagebackend gelesen werden, die Cache Hit Ratio ist nahezu Null.

Dabei wurde synchronisiert auf 3 VMs die Anzahl der Leseprozesse (IOgen Prozesse) von 1 bis maximal 512 pro VM gesteigert. Bei der Konfiguration mit 4 vCPUs (gelb) war die Sättigung bei 192 Prozessen pro VM erreicht, bei 48 vCPUs (blau) konnte noch minimal weiter skaliert werden.

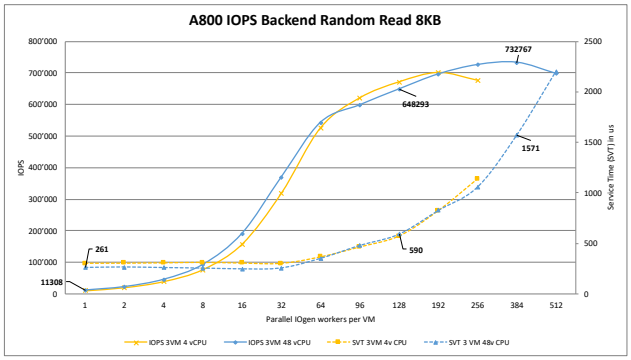

Abbildung 5 – Random Read Performance 8 KB

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 Skale gestrichelte Linien).

Random Reads können bei geringer Parallelität mit äusserst geringer Latenz von 260µs oder 0.26ms aus dem Storage Backend gelesen werden.

Die Latenz bleibt bei Erhöhung der Parallelität bis zu 32 Workern pro VM absolut konstant und steigt dann mit zunehmender Sättigung des Storages an.

Der Optimalwert von 648'293 IOPS wird bei einer Parallelität von 128 Workern pro VM mit einer Latenz von sehr guten 590 µs erreicht.

Danach steigt die IO Leistung geringer an und die Latenz nimmt stärker zu. Das Maximum wird bei einer Parallelität von 384 Prozessen pro VM mit 732'767 IOPS erreicht. Dabei wird eine immer noch akzeptable Latenz von 1'571µs gemessen.

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 42 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 3 VMs die Anzahl der Schreibprozesse (IOgen Prozesse) von 1 bis maximal 256 (bei 4 vCPUs) bzw. 512 (bei 48 vCPUs) pro VM gesteigert. Selbst bei dieser hohen Parallelität steigen die

IOPS immer noch an, allerdings weit weniger stark als die Latenz.

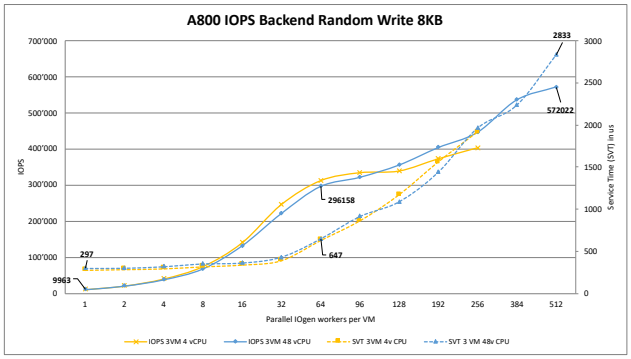

Abbildung 6 – Random Write Performance 8 KB

Random Writes können bei geringer Parallelität mit sehr geringer Latenz von 297µs oder 0.3ms zum Storage übermittelt werden. 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 32 Workern pro VM absolut konstant und steigt dann mit zunehmender Sättigung des Storages an.

Der Optimalwert von 296’158 IOPS wird bei einer Parallelität von 64 Workern pro VM mit einer Latenz von sehr guten 647 µs erreicht.Danach steigt die IO Leistung geringer an und die Latenz nimmt stärker zu. Das Maximum wird bei einer Parallelität von 512 Prozessen pro VM mit 572’022 IOPS erreicht. Dabei wird eine Latenz von 2'833µs gemessen.

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 ultraschnellen NVMe Disks spielt dieser Effekt allerdings kaum noch eine Rolle.

Dabei wurden synchronisiert auf 3 VMs die Anzahl der Leseprozesse (IOgen Prozesse) von 1 bis maximal 32 pro VM gesteigert. Dabei wird unabhängig von der Anzahl der vCPUs bei 16 Prozessen und einer Lesebandbreite von 11.5 GB/s die Sättigung erreicht.

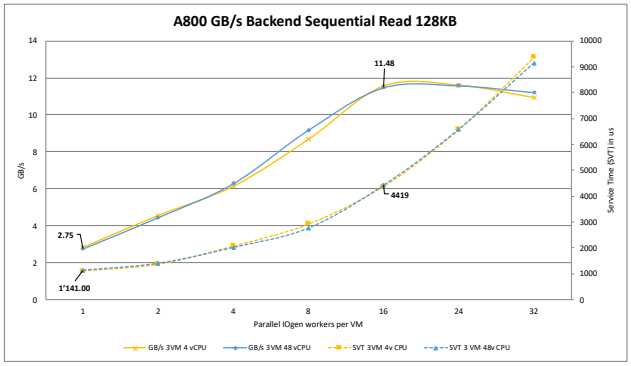

Abbildung 7 – Sequential Read Performance 128 KB

Sequential Reads können bei geringer Parallelität mit einer Latenz von 1’141µs zum Storage übermittelt werden. Diese Latenz ist sehr gering für eine Blockgrösse von 128 KB.

Die Latenz steigt bei Erhöhung der Parallelität bis zu 16 Workern pro VM langsam an. Danach kann keine Steigerung des Durchsatzes mehr erreicht werden.

Der Optimalwert von 11.5 GB/s wird bei einer Parallelität von 16 Workern pro VM mit einer Latenz von 4’419 µs 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 neu Blöcke geschrieben werden und quasi keine Cache Blöcke mehrfach beschreiben werden.

Dabei wurden synchronisiert auf 3 VMs die Anzahl der Schreibprozesse (IOgen Prozesse) von 1 bis maximal 32 pro VM gesteigert. Dabei wird unabhängig von der Anzahl der vCPUs bis zu 32 Prozessen eine Steigerung des Durchsatzes bis zu 11.0 GB/s erreicht.

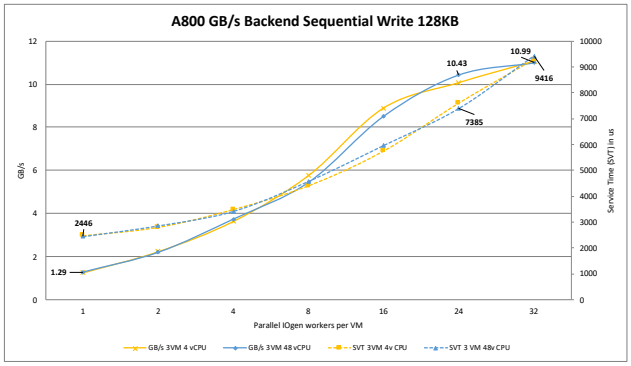

Abbildung 8 – Sequential Write Performance 128 KB

Sequential Writes können bei geringer Parallelität mit einer Latenz von 2.5ms 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 32 Workern kontinuierlich an, aber auch der Durchsatz steigt kontinuierlich an. Bei 32 Workern sieht man Anzeichen einer Sättigung.

Der Optimalwert von 10.4 GB/s wird bei einer Parallelität von 24 Workern pro VM mit einer Latenz von 7.4ms erreicht.

Danach steigt die IO Leistung geringer an und die Latenz nimmt stärker zu. Der maximale Durchsatz wird bei einer Parallelität von 32 Prozessen pro VM mit 11 GB/s erreicht. Dabei wird eine Latenz von 9.4ms 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.

Test	Speed	Optimum	Maximum
8 KB Read	1 Prozess	128 Prozesse	384 Prozesse
Random	11'308 IOPS	648'293 IOPS	732'767 IOPS
Backend	261 µs	590 µs	1'571 µs
8 KB Write	1 Prozess	64 Prozesse	512 Prozesse
Random	9'963 IOPS	296'158 IOPS	572'022 IOPS
Backend	297 µs	647 µs	2'833 µs
128 KB Read	1 Prozess	16 Prozesse	16 Prozesse
Sequential	2.8 GB/s	11.5 GB/s	11.5 GB/s
Backend	1'141 µs	4'419 µs	4'419 µs
128 KB Write	1 Prozess	24 Prozesse	32 Prozesse
Sequential	1.2 GB/s	10.4 GB/s	11.0 GB/s
Backend	2'446 µs	7'385 µs	9'416 µs

Tabelle 2 – Kennzahlen

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

Fazit

Der NetApp All Flash Storage A800 erreicht bereits in der «Minimalkonfiguration», mit zwei Controller Nodes beeindruckende Leistungskennzahlen, die im Highend Segment liegen.

So können gegen 730k Random Reads und 570k Random Writes zu 8 KB und ein Durchsatz von mehr als 10 GB/s beim Lesen und Schreiben erreicht werden. Bei geringer Last werden Latenzen von weniger als 0.3ms bei 8 KB Blöcken erreicht.

Zu berücksichtigen ist, dass diese Leistung mit Ethernet und iSCSI Anbindung erreicht wurde, die im Vergleich zu Fiber Channel (FC) etwas mehr Overhead generiert. Eine Anbindung per FC Native ist im A800 ebenfalls möglich und könnte die Performance nochmals steigern.

Sehr interessant wird dann insbesondere die Anbindung per NVMeoF (NVMe over Fabric). Bei Einsatz dieses Protokolls wird im Vergleich zu FC nochmals eine erhebliche Reduktion der Latenz und Steigerung der Bandbreite erwartet. Der A800 Storage ist auf diese zukünftige Technologie bereits vorbereitet.

Zu berücksichtigen ist, dass bis zu 12 Controller Paare zu einem A800 Cluster zusammengefügt werden können, der dann nochmals Faktoren höhere Leistungen erzielen dürfte.

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.