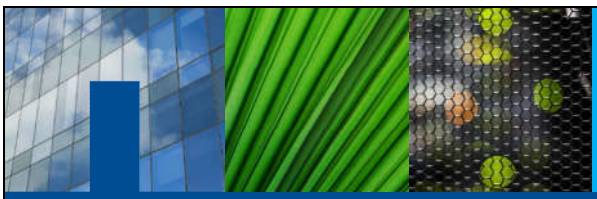
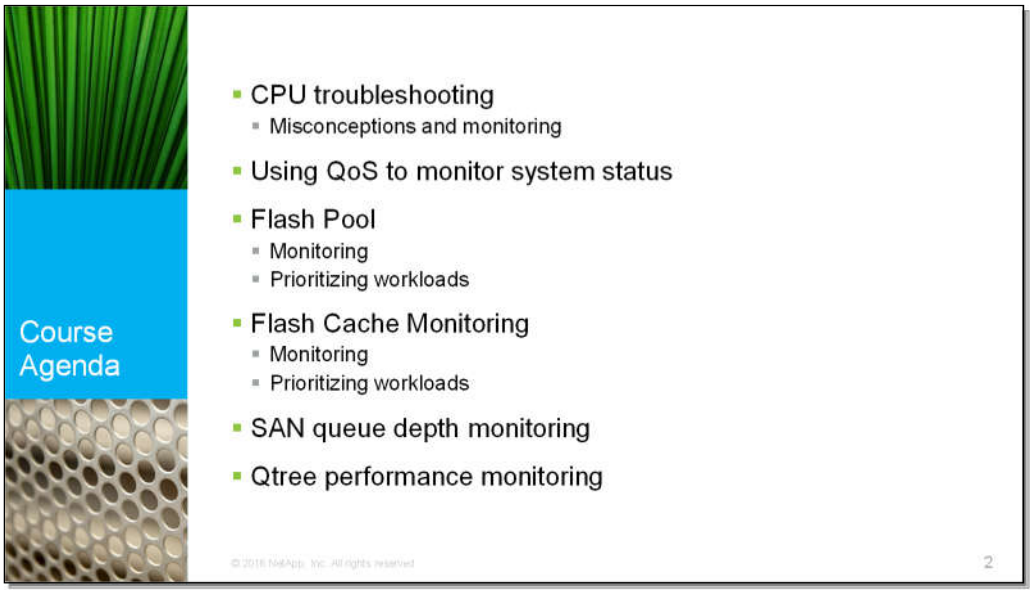


Chapter 8

More Command Line Monitoring


Course Agenda
CPU troubleshooting
Misconceptions and monitoring
Using QoS to monitor system status
Flash Pool
Monitoring
Prioritizing workloads
Flash Cache Monitoring
Monitoring
Prioritizing workloads
SAN queue depth monitoring
Qtree performance monitoring
© 2016 NetApp, Inc. All rights reserved
2

Module 1

CPU Troubleshooting

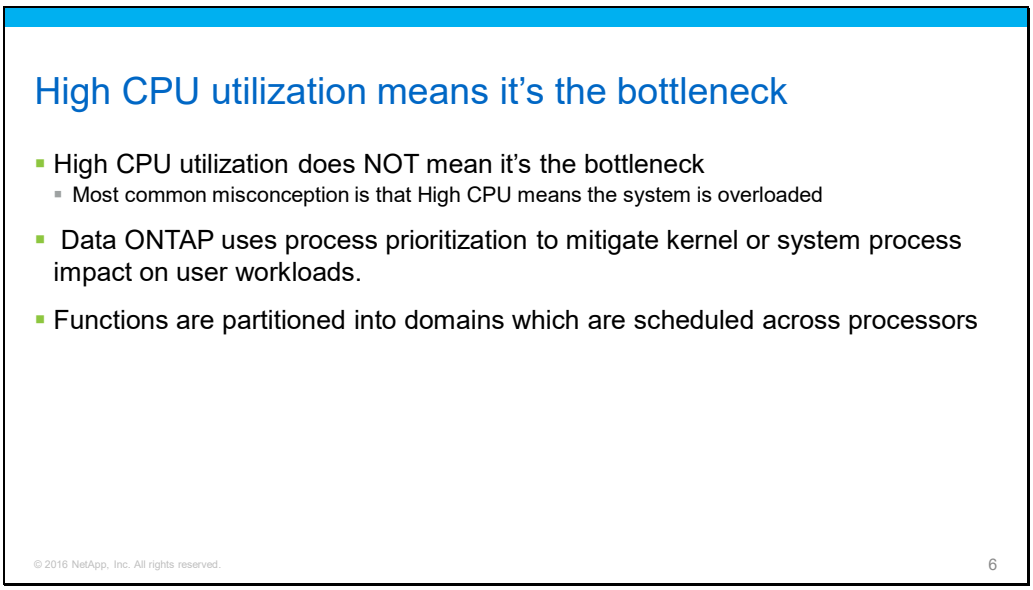
About This Module

This module focuses on the following:

- Common CPU utilization misconception

Lesson 1

Common CPU utilization misconception



Each storage controller contains one or more multi-core CPUs. These physical CPU cores are the primary compute resource available to Data ONTAP for processing work. In addition to CPU cores, Data ONTAP interacts with other physical hardware such as ethernet ports, FC ports, disks, and NVRAM. The use of all these physical resources occurs in a highly optimized and parallel manner, taking into account active requests, resource availability, and the overall activity level of the system.

As the interaction between physical resources can be complex and inter-dependent, a measure of CPU being busy (CPU utilization) does not increase linearly with an increase in incoming requests from clients, nor can it be used alone as a measure of the overall system utilization. That said, CPU resources have several unique characteristics that can be useful to understand when analyzing the system as a whole.

Processor Domains

Domain name	Typical Tasks in Domain	CPU Concurrency	Notes
nwk_exclusive	IP processing, NFS protocol processing	1	
nwk_exempt	IP processing, NFS protocol processing (7-mode and cDOT), SMB processing (cDOT)	1+	Maximum number of CPUs is dependent on controller model and Data ONTAP release
nwk_legacy	IP processing, NFS protocol processing	1	Networking code which can only run on a single CPU concurrently
Storage	Communication with disks	1+	Concurrency of 1 prior to Data ONTAP 8.2.1 or if less than 6 CPUs
raid	RAID subsystem	1	
raid_exempt	RAID subsystem	1+	Introduced in Data ONTAP 8.2
target	SCSI (FCP/iSCSI) processing	1	7-mode only
ssan_exempt	SCSI (FCP/iSCSI) processing	1+	Introduced in clustered Data ONTAP 8.2
Kahuna	Serialized WAFL and anything not in another domain	1	Exclusive with WAFL_Ex (i.e. either Kahuna can be active on 1 CPU or WAFL_Ex can be active on 1+ CPUs, but both cannot be active at the same time)
WAFL_Ex	Parallelized WAFL	1+	Exclusive with Kahuna (i.e. either Kahuna can be active on 1 CPU or WAFL_Ex can be active on 1+ CPUs, but both cannot be active at the same time)
WAFL_XCleaner	WAFL	1+	
SM_Exempt	SnapMirror	1+	
cifs	SMB protocol processing (7-mode only)	1	Initial decoding only; majority of SMB processing occurs in WAFL
exempt	General parallelized work	1+	
hostOS	Tasks owned by the BSD layer including NTP, environmental sensor monitoring, ZAPI handling, autosupport	1+	

7

Data ONTAP uses a Coarse Symmetric Multiprocessing (CSMP) design which partitions system functions into logical processing domains. Each logical processing domain has a set of rules that govern how and when the logical CSMP domain can be scheduled across physical CPU cores. These rules are designed to ensure that all processing occurs in a safe and efficient manner.

The following table describes some of the common logical processing domains, their typical tasks, and describes if the logical domain can run on one or more CPU cores concurrently, along with any specific scheduling rules:

Logical CSMP domains are scheduled to run on physical CPU cores by the Data ONTAP kernel. The scheduling logic is unique to a given Data ONTAP release and hardware platform and is tuned to maximize the overall system performance. As such, the level of parallelism seen for a given logical domain may vary based on a number of factors including the incoming workload rate, the type of work being requested, Data ONTAP OS version and more.

Domain name	Typical Tasks in Domain	CPU Concurrency	Notes
nwk_exclusive	IP processing, NFS protocol processing	1	
nwk_exempt	IP processing, NFS protocol processing (7-mode and cDOT), SMB processing (cDOT)	1+	Maximum number of CPUs is dependent on controller model and Data ONTAP release
nwk_legacy	IP processing, NFS protocol processing	1	Networking code which can only run on a single CPU concurrently
Storage	Communication with disks	1+	Concurrency of 1 prior to Data ONTAP 8.2.1 or if less than 6 CPUs
raid	RAID subsystem	1	
raid_exempt	RAID subsystem	1+	Introduced in Data ONTAP 8.2
target	SCSI (FCP/iSCSI) processing	1	7-mode only
ssan_exempt	SCSI (FCP/iSCSI) processing	1+	Introduced in clustered Data ONTAP 8.2
Kahuna	Serialized WAFL and anything not in another domain	1	Exclusive with WAFL_Ex (i.e. either Kahuna can be active on 1 CPU or WAFL_Ex can be active on 1+ CPUs, but both cannot be active at the same time)

Domain name	Typical Tasks in Domain	CPU Concurrency	Notes
WAFL_Ex	Parallelized WAFL	1+	Exclusive with Kahuna (i.e. either Kahuna can be active on 1 CPU or WAFL_Ex can be active on 1+ CPUs, but both cannot be active at the same time)
WAFL_XCleaner	WAFL	1+	
SM_Exempt	SnapMirror	1+	
cifs	SMB protocol processing (7-mode only)	1	Initial decoding only; majority of SMB processing occurs in WAFL
exempt	General parallelized work	1+	
hostOS	Tasks owned by the BSD layer including NTP, environmental sensor monitoring, ZAPI handling, autosupport	1+	

CPU bottleneck types

- Average CPU core utilization: The average measure of CPU core utilization for all cores reaches 100%.
- Logical domain bottleneck: A logical domain reaches its concurrency limit. For example if a logical domain has a concurrency of 1 CPU core and it reaches 100% utilization.
- Interactions between logical domains: Some logical domains are mutually exclusive and cannot run concurrently with another correlated logical domain. For example, `WAFL_ex` represents parallel WAFL processing while `Kahuna` represents serial WAFL processing. These two logical domains are mutually exclusive, meaning either `Kahuna` can be active on 1 CPU, or `WAFL_ex` can be active on 1+ CPUs, but both Kahuna and `WAFL_Ex` cannot be active at the same time. Depending on the workload, it is possible for `Kahuna` to limit the amount of work that can be performed by `WAFL_ex`. It is important to note that this type of bottleneck is a simple variation on the previous condition.

As mentioned earlier, CPU is just one of the physical resource types available to Data ONTAP. When analyzing system performance, it is crucial to look at the system holistically. A general strategy for analyzing the bottlenecks is to use both service metrics (protocol/volume/lun latency/workload) and component metrics (CPU, Disk IO, Network IO) to provide a complete view of the system and reduce the chance of coming to an incorrect conclusion.

Looking specifically at the CPU resource, work is classified into priorities and some types of work are identified as background or non-essential/opportunistic. This means that when background work is using one or more CPU cores, it will effectively yield to higher priority work as the requests arrive. Also, as the system load increases, it is likely that processing optimizations will result in non-linear scaling for the measure of both the physical CPU core utilization and the logical CSMP domain utilization. This is normal in a complex compute system.

Note: A bottleneck on a physical CPU core is not possible without either reaching a domain bottleneck or average CPU bottleneck. Accordingly, the monitoring of physical CPU utilization as a direct measure is not effective.

Note: Beginning with Data ONTAP 8.2.1, the algorithm for representing CPU utilization ('sysstat') has been changed now and reports the max of avg_processor_busy or the busiest domain that has a concurrency of 1.

One way to measure CPU utilization

```
netapp::> set diag

Warning: These diagnostic commands are for use by NetApp personnel only.

Do you want to continue? {y|n}: y

netapp::*> node run -node netapp-01 sysstat -M 1

   ANY1+ ANY2+ ANY3+ ANY4+ ANY5+ ANY6+ ANY7+ ANY8+ ANY9+ ANY10+ ANY11+ ANY12+ ANY13+ ANY14+
   100%  100%  100%   99%   98%   96%   94%   91%   86%    81%    76%    70%    64%    57%

   ANY15+ ANY16+  AVG CPU0 CPU1 CPU2 CPU3 CPU4 CPU5 CPU6 CPU7 CPU8 CPU9 CPU10 CPU11 CPU12
    48%    37%    81% 78%  76%  77%  83%  82%  83%  82%  82%  82%  82%   83%   84%   83%

   CPU13 CPU14 CPU15 Nwk_Excl Nwk_Ig Nwk_Exmpt Protocol Cluster Storage Raid Raid_Ex Target
    82%   83%   82%      3%     2%       450%       0%       0%      49%    2%    136%      0%

   Kahuna WAFL_Ex(Kahu) WAFL_XClean SM_Exempt Cifs Exempt SSAN_Ex Intr Host  Ops/s   CP
      4%     511%( 94%)       0%          0%     0%   112%      0%   28%   8%  47111   0%
```


9

In this example, Average CPU Utilization is 81% across the 16 cores. The busiest domains are WAFL Exempt at 511%, networking exempt at 450%, RAID exempt at 136%, and exempt at 112%. WAFL was active 98% of the sample interval, with 4% spent in serial processing and 94% in parallel processing. Because WAFL serial processing is quite low it is likely that more work could be completed by parallelized WAFL, so being 98% active within the sample interval it is not a concern without other contributing performance indicators. Overall CPU resources are getting scarce increasing the liklihood that work will queue for CPU potentially impacting client latency.

When reviewing CPU information in Data ONTAP, CPU AVG is a better indicator of overall CPU utilization compared to CPU BUSY.

Another way to look at CPU Utilization

- ANY1+ reads 100% very easily, and this is where UM and many other tools get their CPU information from
- In this example ANY1+ represents the amount of time at least one CPU was busy doing work in a second. This can give a false report of actual CPU usage.

Begin: Sun Feb 14 11:08:57 GMT 2016

ANY1+	ANY2+	ANY3+	ANY4+	ANY5+	ANY6+	ANY7+	ANY8+	AVG	CPU0	CPU1	CPU2	CPU3	CPU4	CPU5	CPU6	CPU7
88%	72%	54%	38%	25%	17%	12%	8%	40%	36%	30%	28%	28%	47%	49%	49%	49%
96%	85%	68%	51%	35%	23%	15%	10%	48%	44%	38%	36%	37%	48%	61%	61%	60%
85%	68%	50%	35%	23%	14%	9%	6%	37%	33%	26%	25%	26%	44%	46%	46%	46%
95%	85%	69%	51%	36%	24%	16%	11%	49%	45%	39%	38%	38%	48%	61%	61%	61%
89%	74%	56%	39%	26%	16%	10%	6%	40%	37%	30%	29%	29%	38%	52%	52%	52%
92%	78%	60%	44%	30%	21%	14%	10%	44%	40%	33%	32%	32%	52%	54%	54%	55%
95%	84%	67%	49%	34%	23%	15%	10%	47%	43%	37%	36%	38%	47%	59%	60%	60%
94%	82%	65%	47%	32%	21%	13%	8%	46%	40%	33%	32%	44%	55%	55%	55%	55%
92%	78%	62%	45%	32%	22%	15%	10%	45%	40%	34%	33%	33%	54%	54%	54%	54%
95%	84%	68%	50%	35%	24%	16%	11%	48%	42%	36%	35%	48%	56%	56%	57%	56%
90%	77%	60%	43%	29%	19%	12%	8%	43%	39%	32%	32%	32%	41%	55%	55%	55%
89%	74%	57%	41%	28%	19%	13%	9%	42%	38%	31%	31%	31%	50%	52%	51%	52%
93%	81%	63%	45%	31%	20%	14%	9%	45%	41%	35%	34%	34%	45%	57%	57%	57%
100%	92%	76%	55%	35%	21%	13%	9%	50%	48%	43%	41%	43%	57%	57%	58%	58%
100%	89%	72%	52%	34%	22%	14%	9%	49%	45%	40%	38%	55%	53%	53%	54%	54%
100%	89%	73%	53%	35%	22%	14%	10%	50%	47%	42%	42%	49%	54%	54%	54%	55%
100%	90%	73%	51%	32%	20%	12%	8%	49%	47%	42%	40%	41%	54%	55%	55%	55%
100%	82%	63%	43%	28%	18%	12%	8%	45%	42%	36%	35%	51%	47%	48%	49%	49%
100%	96%	84%	64%	43%	27%	18%	12%	56%	53%	48%	47%	57%	54%	62%	62%	63%

Reality Check

- Having a high ANY1+ value, without any business impact, is not a cause for alarm or concern if Data ONTAP has free CPU cycles it will find something to do with them.

Module 2

Using QoS to monitor Cluster performance

About This Module

This module focuses on the following:

- QoS usage, beyond just throttling a workload

Lesson 1

QoS Details

Rules for QoS

- Storage objects and policy groups must belong to the same Vserver. You specify the Vserver to which the policy group belongs when you create the policy group. Multiple policy groups can belong to the same Vserver.
- Nested storage objects cannot belong to policy groups

If you assign a...	Then you cannot assign...
Vserver to a policy group	Any storage objects contained by the Vserver to a policy group
Volume to a policy group	The volume's containing Vserver or any child LUNs or files to a policy group
LUN to a policy group	The LUN's containing volume or Vserver to a policy group
File to a policy group	The file's containing volume or Vserver to a policy group

QoS Statistics Commands

- `qos statistics characteristics show` - Display QoS policy group characterization. Example of with/without Policy Groups created.

```
    cluster1::> qos statistics characteristics show -iterations 100 -rows 4
Policy Group          IOPS      Throughput      Request size  Read  Concurrency
-------------------- -------- --------------- --------------- ------ -----------
-total-                 31      304.00KB/s          10041B     0%          16
_System-Best-Effort     15           0KB/s             0B     0%           0
vol1                    11       44.00KB/s           4096B     0%          40
vol2                     4      256.00KB/s          65536B     0%          14
vs1vol0                  1        4.00KB/s           4096B     0%           4
```

```
cluster1::*> qos statistics characteristics show
Policy Group          IOPS      Throughput Request size Read Concurrency
-------------------- -------- --------------- ------------ ---- -----------
-total-               3344        4.53MB/s        1419B  47%          89
User-Best-Effort      3330        4.53MB/s        1425B  47%          89
_System-Work            15        0.08KB/s           5B   0%           0
-total-               1844        3.85MB/s        2188B  44%          94
```


16

The command displays the following data:

- The QoS policy group name (Policy Group)
- Input/output operations performed per second (IOPS)
- Throughput achieved in kilobytes per second (KB/s) or megabytes per second (MB/s) as appropriate (Throughput)
- Request size in bytes (B) (Request size)
- Read percentage from total I/O (Read)
- Concurrency, which indicates the number of concurrent users generating the I/O traffic (Concurrency)

The results displayed per iteration are sorted by IOPS. Each iteration starts with a row that displays the total IOPS used across all QoS policy groups. Other columns in this row are either totals or averages.

[-iterations <integer>] - Number of Iterations
Specifies the number of times the display is refreshed before terminating. If you do not specify this parameter, the command iterates until interrupted by Ctrl-C.

[-rows <integer>] - Number of Rows in the Output
Specifies the number of busiest QoS policy groups to display. The default setting is 10. The allowed range of values is 1 to 20.

QoS Statistics Commands - Continued

- `qos statistics latency show` - Display latency breakdown data per QoS policy group. Example of with/without Policy Groups created.

```
    cluster1::> qos statistics latency show -iterations 100 -rows 3
Policy Group          Latency    Network    Cluster    Data       Disk       QoS
--------------------- ---------- ---------- ---------- ---------- ---------- ----------
-total-               110.35ms   110.02ms        0ms   327.00us        0ms        0ms
vs1vol0               167.82ms   167.22ms        0ms   603.00us        0ms        0ms
vol1                  117.76ms   117.56ms        0ms   191.00us        0ms        0ms
vol2                   44.24ms    44.05ms        0ms   190.00us        0ms        0ms
-total-                38.89ms    38.63ms        0ms   256.00us        0ms        0ms
```

```
cluster1::*> qos statistics latency show
Policy Group          Latency    Network    Cluster       Data       Disk
   QoS      NVRAM
--------------------- ---------- ---------- ---------- ---------- ---------- ----
------ ----------
-total-               36.47ms    641.00us        0ms    672.00us    35.16ms
   0ms     4.00us
User-Best-Effort      36.47ms    641.00us        0ms    672.00us    35.16ms
   0ms     4.00us
-total-               32.65ms    645.00us        0ms    829.00us    31.17ms
```


17

The command displays the following data:

- The QoS policy group name (Policy Group)
- Total latency observed per I/O operation (Latency)
- Latency observed per I/O operation in the Network subsystem (Network)
- Latency observed per I/O operation across the internally connected nodes in a Cluster (Cluster)
- Latency observed per I/O operation in the Data management subsystem (Data)
- Latency observed per I/O operation in the Storage subsystem (Disk)
- Latency observed per I/O operation in the QoS subsystem (QoS)

The results displayed per iteration are sorted by the Latency field. Each iteration starts with a row that displays the average latency, in microseconds (us) or milliseconds (ms), observed across all QoS policy groups.

QoS Statistics Commands - Continued

- `qos statistics performance show` - Display system performance data per QoS policy group. Example of with/without Policy Groups created.

```
    cluster1::> qos statistics performance show -iterations 100 -rows 4
Policy Group          IOPS      Throughput   Latency
--------------------- --------  ---------------  ----------
-total-                 79     1296.00KB/s   337.41ms
_System-Best-Effort     25           0KB/s        0ms
vol1                    24       96.00KB/s   193.72ms
vol2                    18     1152.00KB/s   750.98ms
vs1vol0                 12       48.00KB/s   707.38ms
-total-                109        1.99MB/s   133.27ms
```

```
cluster1::*> qos statistics performance show
Policy Group             IOPS      Throughput    Latency
--------------------- --------  ---------------  ----------
-total-                 2418        2.17MB/s    37.71ms
User-Best-Effort        2327        2.17MB/s    39.16ms
_System-Work              91        3.92KB/s   362.00us
-total-                 1209        2.64MB/s    77.15ms
```


The command displays the following data:

- The QoS policy group name (Policy Group)
- Input/output operations performed per second (IOPS)
- Throughput in kilobytes per second (KB/s) or megabytes per second (MB/s) as appropriate (Throughput)
- Latency observed per request in microseconds (us) or milliseconds (ms) as appropriate (Latency)

The results displayed per iteration are sorted by IOPS. Each iteration starts with a row that displays the total IOPS used across all QoS policy groups. Other columns in this row are either totals or averages.

QoS Statistics Commands - Continued

- `qos statistics resource cpu show` - Display CPU resource utilization data per QoS policy group. Example of with/without Policy Groups created.

```
cluster1::> qos statistics resource cpu show -node nodeA -iterations 100 -rows 3
Policy Group         CPU
-------------------- -----
-total- (100%)         9%
fast                   1%
slow                   3%
medium                 5%
-total- (100%)         8%
```

```
cluster1::*> qos statistics resource cpu show -node cluster1-01
Policy Group           CPU Wafl_exempt Kahuna Network Raid  Exempt Protocol
-------------------- ----- ---------- ------ ------- ----- ------ --------
-total- (200%)        166%        10%     0%     41%    1%   108%       6%
_System-Best-Effort   131%         0%     0%     21%    0%   107%       3%
User-Best-Effort       32%         8%     0%     20%    0%     1%       3%
_System-Work            1%         1%     0%      0%    0%     0%       0%
-total- (200%)        178%        16%     0%     54%    1%    96%      11%
```


19

The command displays the following data:

- The QoS policy group name (Policy Group)
- CPU utilization observed in percentage (CPU)

The results displayed per iteration are sorted by total CPU utilization. Each iteration starts with a row that displays the total CPU utilization across all QoS policy groups.

QoS Statistics Commands - Continued

- `qos statistics resource disk show` - Display disk resource utilization data per QoS policy group. Example of with/without Policy Groups created.

```
cluster1::> qos statistics resource disk show -node nodeA -iterations 100 -rows 3
Policy Group         Disk  No. of Disks
-------------------- ----- ------------
-total-               40%           27
pg1                   22%            5
slow                  10%           10
fast                   8%           12
_System_Default        7%           20
-total-               42%           27
```

```
cluster1::*> qos statistics resource disk show -node cluster1-01
Policy Group          Disk Number of HDD Disks   Disk Number of SSD Disks
--------------------- ----- ------------------- ------ -------------------
-total-                45%                  11     4%                   3
User-Best-Effort       20%                  25     2%                   6
-total-                37%                  11     3%                   3
```


The command displays the following data:

- The QoS policy group name (Policy Group)
- Disk utilization (Disk)
- The number of data disks utilized (No. of Disks)

The results displayed are sorted by total disk utilization. Each iteration starts with a row that displays the total disk utilization across all QoS policy groups.

The disk utilization shows the percentage of time spent on the disk during read and write operations. The command displays disk utilization for system-defined policy groups; however, their disk utilization is not included in the total utilization.

QoS Statistics Commands - Continued

- `qos statistics workload characteristics show` - Display QoS workload characterization

```
    cluster1::> qos statistics workload characteristics show -iterations 100 -rows 4
Workload          ID     IOPS      Throughput      Request size    Read  Concurrency
--------------- ------ -------- ---------------- --------------- ------- -----------
-total-            -       68       176.00KB/s            2650B      7%           8
vs1vol0-wid102     102     24        96.00KB/s            4096B     20%          13
_Scan_Besteff..    101     23            0KB/s               0B      0%           0
vol_1-wid103       103     20        80.00KB/s            4096B      0%          12
vol_2-wid104       104      1            0KB/s               0B      0%           0
-total-            -      157       528.00KB/s            3443B      3%           4
```

```
cluster1::*> qos statistics workload characteristics show
Workload           ID     IOPS       Throughput Request Size Read Concurrency
--------------- ------ -------- ---------------- ------------ ---- -----------
-total-             -     2551         2.81MB/s        1154B  50%         128
User-Default        2     2546         2.81MB/s        1157B  50%         128
_USERSPACE_APPS    14        5            0KB/s           0B   0%           0
-total-             -     3499         3.04MB/s         912B  49%          92
```


21

The qos statistics workload characteristics show command displays data that characterizes the behavior of QoS workloads.

The command displays the following data:

- The QoS workload name (Workload)
- The QoS workload ID (ID)
- Input/output operations performed per second (IOPS)
- Throughput achieved in kilobytes per second (KB/s) or megabytes per second (MB/s) as appropriate (Throughput)
- Request size in bytes (B) (Request size)
- Read percentage from total IOPS (Read)
- Concurrency, which indicates the number of concurrent users generating the I/O traffic (Concurrency)

The results displayed per iteration are sorted by IOPS. Each iteration starts with a row that displays the total IOPS used across all QoS workloads. Other columns in this row are either totals or averages.

QoS Statistics Commands - Continued

- `qos statistics workload latency show` - Display latency breakdown data per QoS workload

```
    cluster1::> qos statistics workload latency show -iterations 100 -rows 3
Workload          ID   Latency     Network  Cluster     Data     Disk     QoS
--------------- ------ --------   -------- --------   -------- -------- --------
-total-                110.35ms   110.02ms       0ms   327.00us       0ms      0ms
vs1vol0           111 167.82ms   167.22ms       0ms   603.00us       0ms      0ms
vol1             1234 117.76ms   117.56ms       0ms   191.00us       0ms      0ms
vol2              999  44.24ms    44.05ms       0ms   190.00us       0ms      0ms
-total-             -  38.89ms    38.63ms       0ms   256.00us       0ms      0ms
```

```
cluster1::*> qos statistics workload latency show
Workload          ID   Latency     Network    Cluster       Data       Disk        QoS       NVRAM
--------------- ------ ----------  ---------- ----------  ---------- ---------- ---------- ----------
-total-              -    26.40ms   549.00us        0ms   787.00us    25.06ms        0ms      4.00us
User-Default         2    26.40ms   549.00us        0ms   787.00us    25.06ms        0ms      4.00us
-total-              -    21.44ms   428.00us        0ms   880.00us    20.13ms        0ms      4.00us
```


The command displays the following data:

- The QoS workload name (Workload)
- The QoS workload ID (ID)
- Total latency observed per I/O operation (Latency)
- Latency observed per I/O operation in the Network subsystem (Network)
- Latency observed per I/O operation across the internally connected nodes in a Cluster (Cluster)
- Latency observed per I/O operation in the Data management subsystem (Data)
- Latency observed per I/O operation in the Storage subsystem (Disk)
- Latency observed per I/O operation in the QoS subsystem (QoS)

The results displayed per iteration are sorted by the total latency field. Each iteration starts with a row that displays the average latency, in microseconds (us) or milliseconds (ms) observed across all QoS workloads.

QoS Statistics Commands - Continued

- `qos statistics workload performance show` - Display system performance data per QoS workload

```
    cluster1::> qos statistics workload performance show -iterations 100 -rows 4
Workload          ID     IOPS      Throughput    Latency
--------------- ------ -------- ---------------- ----------
-total-              -       97         1.90MB/s   216.87ms
_Scan_Besteff..    101       31            0KB/s        0ms
vol_2-wid104       104       28         1.75MB/s   412.78ms
vol_1-wid103       103       25       100.00KB/s   169.16ms
vs1vol0-wid102     102       13        52.00KB/s   403.78ms
-total-              -       98      1276.00KB/s    89.98ms
```

```
cluster1::*> qos statistics workload performance show
Workload            ID     IOPS       Throughput    Latency
--------------- ------ -------- ---------------- ----------
-total-              -     3758         2.84MB/s    23.87ms
User-Default         2     3546         2.83MB/s    25.27ms
_USERSPACE_APPS     14      213        12.09KB/s   468.00us
-total-              -     2649         2.47MB/s    33.96ms
```


The command displays the following data:

- The QoS workload name (Workload)
- The QoS workload ID (ID)
- Input/output operations performed per second (IOPS)
- Throughput in kilobytes per second (KB/s) or megabytes per second (MB/s) as appropriate (Throughput)
- Latency observed per request in microseconds (us) or milliseconds (ms) as appropriate (Latency)

The results displayed per iteration are sorted by IOPS. Each iteration starts with a row that displays the total IOPS used across all QoS workloads. Other columns in this row are either totals or averages.

QoS Statistics Commands - Continued

- `qos statistics workload resource cpu show` - Display CPU resource utilization data per QoS workload

```
cluster1::> qos statistics workload resource cpu show -node nodeA -iterations 100 -rows 3
Workload        ID    CPU
--------------- ----- -----
--total- (100%)     -    9%
vs0-wid-102       102    5%
file-bigvmdk-..   121    2%
vs2_vol0-wid-..   212    2%
-total- (100%)      -    8%

cluster1::*> qos statistics workload resource cpu show -node cluster1-01
Workload           ID   CPU  Wafl_exempt  Kahuna  Network  Raid  Exempt  Protocol
---------------  -----  -----  -----------  ------  -------  -----  ------  --------
-total- (200%)       -  118%          8%      0%      35%    1%     69%        5%
System-Default       1   85%          0%      0%      14%    0%     68%        3%
User-Default         2   30%          6%      0%      21%    0%      1%        2%
_WAFL                7    1%          1%      0%       0%    0%      0%        0%
-total- (200%)       -  163%         14%      0%      54%    0%     83%       12%
```


The command displays the following data:

- The QoS workload name (Workload)
- The QoS workload ID (ID)
- CPU utilization observed in percentage (CPU)

The results displayed per iteration are sorted by total CPU utilization. Each iteration starts with a row that displays the total CPU utilization across all QoS workloads.

QoS Statistics Commands - Continued

- `qos statistics workload resource disk show` - Display disk resource utilization data per QoS workload

```
cluster1::> qos statistics workload resource disk show -node nodeB -iterations 100 -rows 3
Workload          ID   Disk  No. of Disks
--------------- ------ ----- ------------
-total- (100%)       -   30%            4
_RAID                -   20%            4
vs0-wid101         101   12%            2
file-1-wid121      121   10%            1
vol0-wid1002      1002    8%            1
_WAFL                -    7%            3
-total- (100%)       -   30%            4
```

```
cluster1::*> qos statistics workload resource disk show -node cluster1-01
Workload            ID   Disk Number of HDD Disks   Disk Number of SSD Disks
--------------- ------ ------ ------------------- ------ -------------------
-total-              -    39%                  11     2%                   3
User-Default         2    43%                  10     2%                   3
-total-              -    37%                  11     2%                   3
```


25

The command displays the following data:

- The QoS workload name (Workload)
- The QoS workload ID (ID)
- Disk utilization (Disk)
- The number of data disks utilized (No. of Disks)

The results displayed are sorted by total disk utilization. Each iteration starts with a row that displays the total disk utilization across all QoS workloads.

The disk utilization shows the percentage of time spent on the disk during read and write operations. The command displays disk utilization for system-defined workloads; however, their disk utilization is not included in the total utilization.

QoS Statistics Commands - Continued

- `qos statistics volume latency show` - Display latency breakdown data per volume

```
    cluster1::> qos statistics volume latency show -iterations 100 -rows 3
Workload           ID  Latency    Network  Cluster      Data      Disk      QoS      NVRAM
--------------- ------ --------   -------- --------   -------- -------- --------  ----------
-total-                110.35ms   110.02ms      0ms   327.00us      0ms      0ms         0ms
vs1vol0            111 167.82ms   167.22ms      0ms   603.00us      0ms      0ms         0ms
vol1              1234 117.76ms   117.56ms      0ms   191.00us      0ms      0ms         0ms
vol2               999  44.24ms    44.05ms      0ms   190.00us      0ms      0ms         0ms
-total-              -  38.89ms    38.63ms      0ms   256.00us      0ms      0ms         0ms

cluster1::*> qos statistics volume latency show
Workload           ID    Latency    Network    Cluster      Data       Disk        QoS        NVRAM
--------------- ------ ----------  ---------- ---------- ---------- ---------- ---------- ----------
-total-                   34.04ms  1011.00us         0ms     2.25ms    30.78ms         0ms         0ms
vs1_vol02-wid..   9808    406.54ms   789.00us         0ms    51.81ms   353.94ms         0ms     3.00us
vs2_vol01-wid..  27641    335.25ms     2.45ms         0ms   732.00us   332.06ms         0ms         0ms
vs2_vol02-wid..  27670    288.06ms     4.70ms         0ms    39.07ms   244.30ms         0ms         0ms
vs1_vol01-wid..  24580    137.12ms     6.21ms         0ms  1090.00us   129.81ms         0ms         0ms
lun_vsiscsi_1..  19266      2.70ms   466.00us         0ms   206.00us     2.03ms         0ms         0ms
-total-                   25.95ms   498.00us         0ms  1065.00us    24.39ms         0ms         0ms
```


The command displays the following data:

- The QoS volume name (Workload)
- The QoS workload ID (ID)
- Total latency observed per I/O operation (Latency)
- Latency observed per I/O operation in the Network subsystem (Network)
- Latency observed per I/O operation across the internally connected nodes in a Cluster (Cluster)
- Latency observed per I/O operation in the Data management subsystem (Data)
- Latency observed per I/O operation in the Storage subsystem (Disk)
- Latency observed per I/O operation in the QoS subsystem (QoS)
- Latency observed per I/O operation for NVRAM transfer (NVRAM)

The results displayed per iteration are sorted by the total latency field. Each iteration starts with a row that displays the average latency, in microseconds (us) or milliseconds (ms) observed across all volumes.

Module 3

Flash Pool Monitoring and Prioritization

About This Module

This module focuses on the following:

- Monitoring Flash Pool utilization
- Prioritizing Flash Pool workloads
- How to determine Flash Pool Candidacy

Lesson 1

Flash Pool Monitoring

Flash Pool monitoring from the CLI

- `statistics cache flash-pool show` - Flash pool throughput metrics

```
cluster1::*> statistics cache flash-pool show

cluster1 : 3/18/2017 21:15:03
                         Read Write
                          Hit   Hit *Cache   Read  Write
Aggregate Vserver  Volume  (%)   (%)   Used Blocks Blocks Rejects
--------- ------- ------- ---- ----- ------ ------ ------ -------
n01_aggr2       - -total-   64     0 241996 130140 106332       0
n02_aggr2       - -total-    0     0    787      0      0       0
n01_aggr2     vs2 vs2_vol02(1)
                          67     0 118821  65252  53569       0
n01_aggr2     vs1 vs1_vol02
                          60     0 117651  64888  52763       0
```

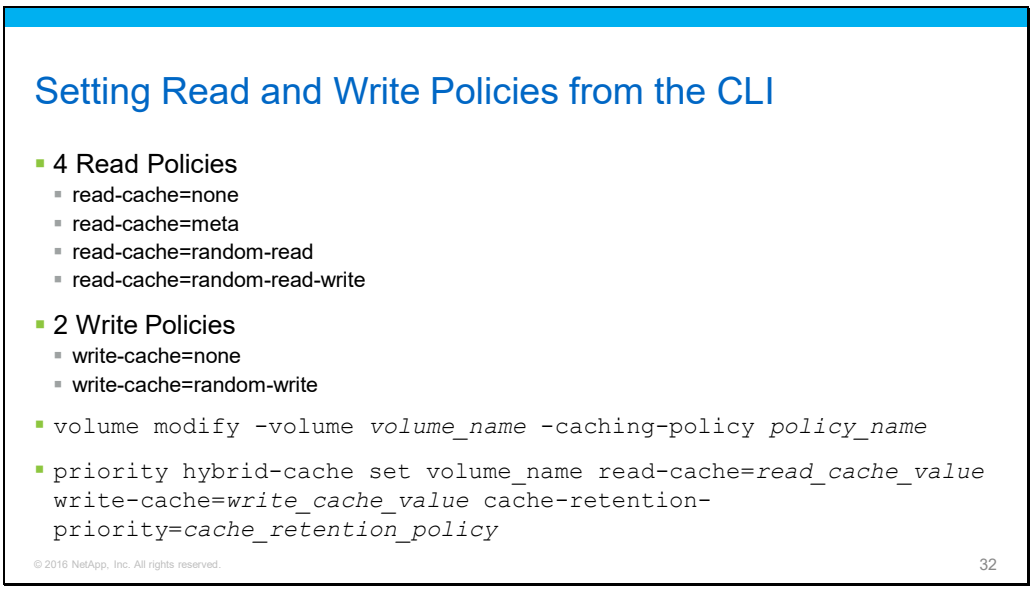

30

This command continuously displays performance data for flash pool caches at a regular interval. The command output displays data in the following columns:

- Aggregate - aggregate name.
- Vserver - vserver name.
- Volume - volume name.
- Read Hit - percent of IOs serviced from a cache level.
- Write Hit - percent of IOs serviced from a cache level.
- Cache Used - percent of cache used.
- Read Blocks - read blocks.
- Write Blocks - write blocks.
- Rejects - cache rejects.

Lesson 2

Flash Pool workload prioritization



1. read-cache = none. If this policy is set on a particular volume, no data from that volume will be placed on SSDs for any host read operations from this volume thereafter. You may want to do that if you know data on this volume is infrequently used, or fast read latency from this volume is not a concern.

2. read-cache = meta. Metadata is data structures or entities that the controller uses to keep track of user data. This policy means only metadata (and no user data) will be placed on SSDs for host read operations. In general, the size of metadata is a fraction of the user data size. If the user datasets are too large to be effectively cached on SSDs, you may choose to use this policy.

3. read-cache = random-read. This policy will cause both metadata and randomly read data being placed on SSDs for accelerating subsequent host random read operations. This means if the host reads are sequential in nature, those sequential reads will not be cached on SSDs. This is the default policy, as it is likely the common use case.

4. read-cache = random-read-write. This policy really makes it interesting. What it does is that it'll let you cache metadata, random reads and random writes on SSDs for accelerating subsequent host random read operations. It is categorized as a read policy. But you wonder why. Why should a read caching policy deal with random writes? I think we can probably come up with a case where a host needs to read the data soon after it has been written to the hard drives.

5. write-cache = none. If this policy is set on a particular volume, no random writes to this volume will be placed on SSDs thereafter. This means effectively, write caching for this volume is disabled. All random writes will go to hard disk drive (HDD) as they would on a normal aggregate. If your workload does not do random overwrites frequently, then this policy should be a good fit.

6. write-cache = random-write. This policy will cause frequent random overwrites being written on SSDs rather than HDDs. This provides a new and fast destination for those overwrites, thus reducing the I/O load on HDDs. The data in the write cache could also satisfy host reads if they happen to be cache hits, thus further reducing the load on HDDs. This is the default policy, as it is a target use case of Flash Pool.

Modifying caching policies of Flash Pool aggregates

You should modify the caching policy of a volume only if a different caching policy is expected to provide better performance. You can modify the caching policy of a volume on a Flash Pool aggregate.

Before you begin
You must determine whether you want to modify your caching policy.

About this task
In most cases, the default caching policy of auto is the best caching policy that you can use. The caching policy should be changed only if a different policy provides better performance for your workload. Configuring the wrong caching policy can severely degrade volume performance; the performance degradation could increase gradually over time. You should use caution when modifying caching policies. If you experience performance issues with a volume for which the caching policy has been changed, you should return the caching policy to auto.

Step
Modify the volume's caching policy:
`volume modify -volume` ***`volume_name`*** **`-caching-policy`** ***`policy_name`***

Example
The following example modifies the caching policy of a volume named "vol2" to the policy none:
```
volume modify -volume vol2 -caching-policy none
```

Setting the cache-retention policy for Flash Pool aggregates

Beginning with ONTAP 9.0, you can assign cache-retention policies to volumes in Flash Pool aggregates. Data in volumes with a high cache-retention policy remains in cache longer and data in volumes with a low cache-retention policy is removed sooner. This increases performance of your critical workloads by making high priority information accessible at a faster rate for a longer period of time.

Before you begin
You should know whether your system has any conditions that might prevent the cache-retention policy from having an impact on how long your data remains in cache.

About this task
The task must be done in advanced privilege mode.

Steps

Verify the volume's cache-retention policy:
By default the cache retention policy is normal.

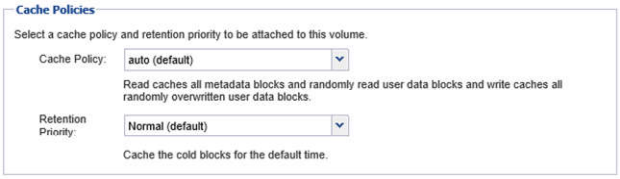
Set the cache-retention policy:
`priority hybrid-cache set` ***`volume_name`*** **`read-cache=`** ***`read_cache_value`*** **`write-cache=`** ***`write_cache_value`*** **`cache-retention-priority=`** ***`cache_retention_policy`***

Set *cache_retention_policy* to high for data that you want to remain in cache longer. Set *cache_retention_policy* to low for data that you want to remove from cache sooner.

Verify that the volume's cache-retention policy is changed to the option you selected.

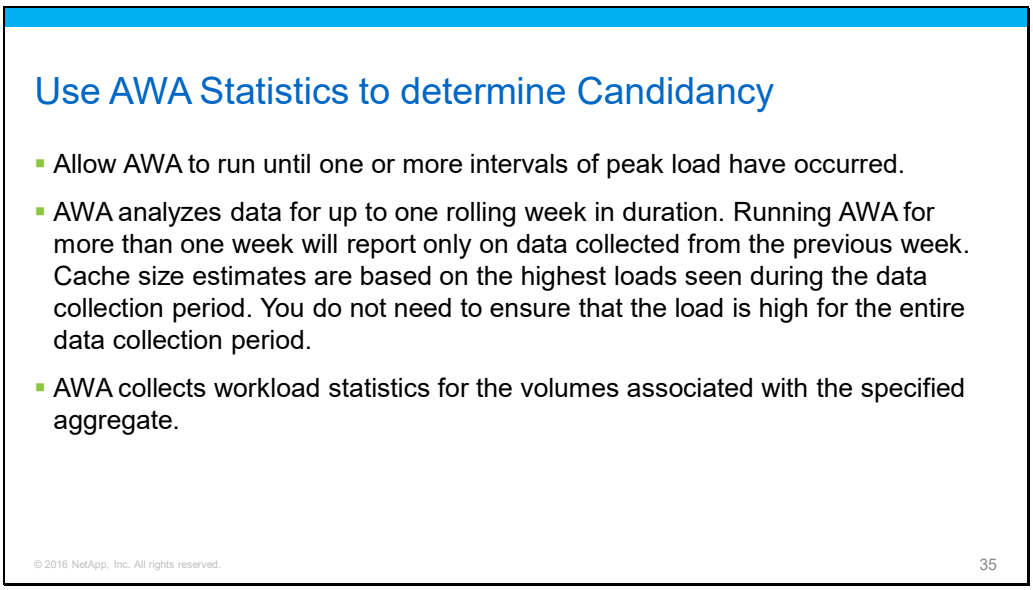
Setting Read and Write Policies from the GUI

- In the volume properties click the advanced tab. If your volume is in a flash pool you can set the cache policies



Lesson 3

How to determine Flash Pool Candidacy



Before you begin

You should know approximately when the aggregate you are analyzing experiences its peak load.

Steps

1. Enter advanced mode:

   set advanced

2. If you need to determine whether an existing aggregate would be a good candidate for conversion to a Flash Pool aggregate, determine how busy the disks in the aggregate are during a period of peak load, and how that is affecting latency:

   statistics show-periodic -object disk:raid_group -instance *raid_group_name* -counter disk_busy|user_read_latency -interval 1 -iterations 60

   You can decide whether reducing latency by adding Flash Pool cache makes sense for this aggregate.

   Example

   The following command shows the statistics for the first RAID group of the aggregate "aggr1":**statistics show-periodic -object disk:raid_group -instance /aggr1/plex0/rg0 -counter disk_busy|user_read_latency -interval 1 -iterations 60**

3. Start AWA:**system node run -node *node_name* wafl awa start *aggr_name***

   AWA begins collecting workload data for the volumes associated with the specified aggregate.

4. Exit advanced mode:**set admin**

5. Allow AWA to run until one or more intervals of peak load have occurred.

   AWA analyzes data for up to one rolling week in duration. Running AWA for more than one week will report only on data collected from the previous week. Cache size estimates are based on the highest loads seen during the data collection period. You do not need to ensure that the load is high for the entire data collection period.

   AWA collects workload statistics for the volumes associated with the specified aggregate.

6. Enter advanced mode:

   set advanced

7. Display the workload analysis:

   system node run -node *node_name* wafl awa print

   AWA displays the workload statistics and optimal Flash Pool cache size.

8. Stop AWA:

9. **system node run -node *node_name* wafl awa stop**

   All workload data is flushed and is no longer available for analysis.

10. Exit advanced mode:**set admin**

AWA Example

- In the example, AWA was run on aggregate "aggr1".
- Here is the output of the awa print command after AWA had been running for about 3 days (442 10-minute intervals):

```
### FP AWA Stats ###

Basic Information

                Aggregate aggr1
           Current-time Mon Jul 28 16:02:21 CEST 2014
             Start-time Thu Jul 31 12:07:07 CEST 2014
     Total runtime (sec) 264682
  Interval length (sec) 600
        Total intervals 442
      In-core Intervals 1024

Summary of the past 442 intervals
                                    max
         Read Throughput       39.695 MB/s
        Write Throughput       17.581 MB/s
      Cacheable Read (%)           92 %
     Cacheable Write (%)           83 %
Max Projected Cache Size         114 GiB
   Projected Read Offload          82 %
  Projected Write Offload          82 %

Summary Cache Hit Rate vs. Cache Size

      Size       20%       40%       60%       80%      100%
    Read Hit      34        51        66         7%        82
   Write Hit      35        44        53        62         82

The entire results and output of Automated Workload Analyzer (AWA) are
estimates. The format, syntax, CLI, results and output of AWA may
change in future Data ONTAP releases. AWA reports the projected cache
size in capacity. It does not make recommendations regarding the
number of data SSDs required. Please follow the guidelines for
configuring and deploying Flash Pool; that are provided in tools and
collateral documents. These include verifying the platform cache size
maximums and minimum number and maximum number of data SSDs.


### FP AWA Stats End ###
```


The results provide the following pieces of information:

- Read Throughput and Write Throughput

  The throughput measurements can help you identify an aggregate that is receiving a higher amount of traffic. Note that these numbers do not indicate whether that aggregate is I/O bound.

- Max Projected Cache Size

  The size at which the SSD cache would hold every eligible data block that was requested from disk during the AWA run. Note that this does not guarantee a hit for all future I/O operations, because they might request data that is not in the cache. However, if the workload during the AWA run was a typical one, and if your budget allows for it, this would be an ideal size for your Flash Pool cache.

- Projected Read Offload and Projected Write Offload

  The approximate percentages of read and write operations that would have been handled by a Flash Pool cache of the optimal size rather than going to disk (projected cache hit rate). Note that this number is related to the performance increase you would see by converting the aggregate to a Flash Pool aggregate, but not an exact prediction.

- Summary Cache Hit Rate vs. Cache Size

  This table can help you predict the performance impact of decreasing the size of the SSD cache from Max Projected Cache Size. These values are highly impacted by your workload. Depending on whether data that was aged out of the cache was ever accessed again, the impact of decreasing the size of the cache might be large or almost nonexistent. You can use this table to find the "sweet spot" of cost versus performance for your workload and budget.

Module 4

Flash Cache Monitoring and Prioritization

About This Module

This module focuses on the following:

- Monitoring Flash Cache utilization
- Prioritizing Flash Cache workloads

Lesson 1

Flash Cache Monitoring

Flash Cache monitoring from the CLI

- `show-periodic -object ext_cache_obj -instance ec0 -counter hit_percent -interval 2 -iterations 5`

```
cluster1: ext_cache_obj.ec0: 10/21/2015 04:44:00
hit
percent
-------
23%
6%
21%
22%
22%
cluster1: ext_cache_obj.ec0: 10/21/2015 04:44:10
hit
percent
-------
Minimums:
6%
Averages for 5 samples:
18%
Maximums:
23%
```

Lesson 2

Flash Cache prioritization

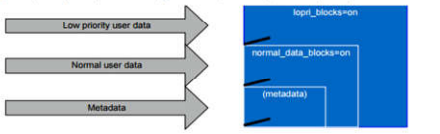
Flash Cache prioritization from the CLI

- To prioritize use the node shell to set flexscale options on a per node basis
  - nodeA> options flexscale
    - flexscale.enable on
    - flexscale.lopri_blocks off
    - flexscale.normal_data_blocks on

The FlexScale options settings to enable low-priority mode are:

```
flexscale.enable                on
flexscale.lopri_blocks          on
flexscale.normal_data_blocks on
```

Figure 8) Low-priority data caching; metadata, user data, and low-priority data are cached.



- To improve performance in the case of a failover try using
  - `system node external-cache modify -is-rewarm-enabled true`

[-is-rewarm-enabled {true|false}] - Is rewarm on

Specifies whether an external-cache module should attempt to preserve data across reboots. Valid values for this option are true and false. This option applies only to cache hardware with persistent media. It does not apply to Predictive Cache Statistics (PCS). Enabling this option will marginally increase the duration of system boot and shutdown, but it will reduce or eliminate the time required for cache warming.

The default value for this option is determined by the cache hardware type. The option is disabled by default.

Module 4

SAN Queue Depth

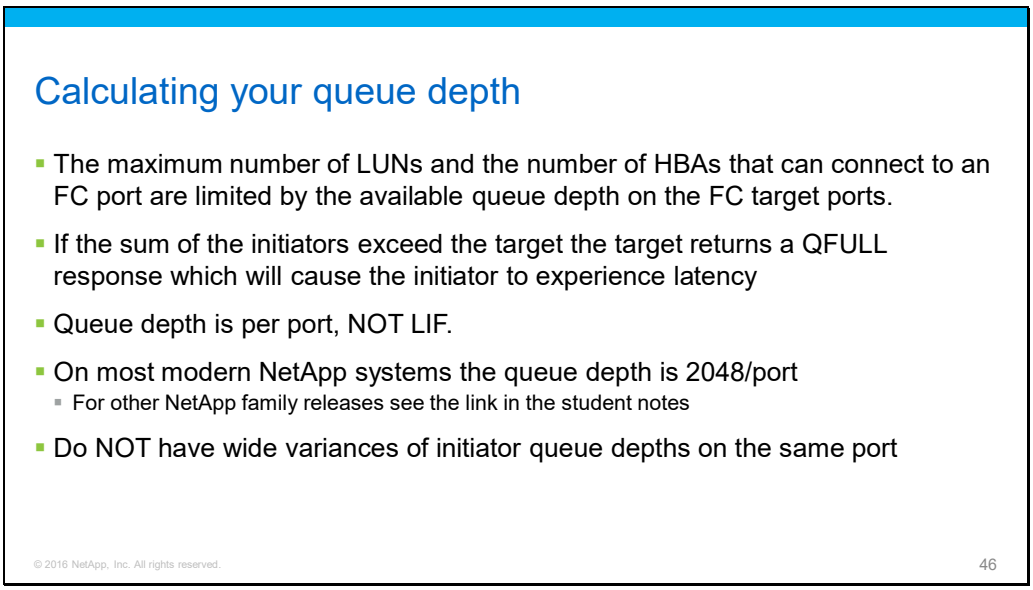
About This Module

This module focuses on the following:

- Monitoring SAN queue depth

Lesson 1

How to fine tune your FC queue depth.



Queue depth is the number of I/O requests (SCSI commands) that can be queued at one time on a storage controller. Each I/O request from the host's initiator HBA to the storage controller's target adapter consumes a queue entry. Typically, a higher queue depth equates to better performance. However, if the storage controller's maximum queue depth is reached, that storage controller rejects incoming commands by returning a QFULL response to them. If a large number of hosts are accessing a storage controller, plan carefully to avoid QFULL conditions, which significantly degrade system performance and can lead to errors on some systems.

https://library.netapp.com/ecmdocs/ECMP1196793/html/GUID-A055B184-0876-4376-9C75-35FE8C9BE832.html

https://kb.netapp.com/support/s/article/ka21A0000000YzpQAE/what-are-the-maximum-fibre-channel-fc-queue-depth-settings-per-port-for-each-single-controller-active-active-configuration

Try not to have wide variances of initiator queue depths on the same port. Even if the target queue depth is not exceeded it is still possible for an initiator with a deep queue depth to bully a client with a shallow queue depth.

Client side queue depth configuration

- For Microsoft SQL server deeper queue depths can typically lead to better performance. A depth of 64 is normally the minimum to start, with 255 recommended in some instances.
  - See link in notes for more info.
- For Oracle servers the same holds true, normally a deeper queue depth will see greater performance, at lower latencies.
  - If you search google you will see people setting queue depths as great as 512 with 255 recommended in the link in the student notes.
- For Microsoft Exchange the recommendation from Microsoft is 128
  - See link in the student notes
- For Vmware the current recommendation is 64.
  - See the student notes for more info on this

Microsoft SQL Server
https://blogs.msdn.microsoft.com/joesack/2009/01/28/sql-server-and-hba-queue-depth-mashup/
http://longwhiteclouds.com/2016/03/14/performance-testing-with-microsoft-diskspd/

Oracle
http://www.oracle.com/technetwork/server-storage/san-storage/documentation/wp_fs1_vsphere6_20160401_en.pdf

Microsoft Exchange
https://technet.microsoft.com/en-us/library/aa997725(v=exchg.80).aspx

Vmware
https://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1267
https://communities.vmware.com/docs/DOC-6490

Lesson 2

How to monitor your queue depth

View current value of the Target Port queue Depth

For FlexArray and V-Series

- From the advanced priv set run the following command to show the current value of the target port queue depth
  - `storage array show -instance`

```
cluster1::*> storage array show -instance

                              Name: NETAPP_VD_1
                            Prefix: NET-1
                            Vendor: NETAPP
                             Model: VD
                           options:
                     Serial Number: 7668611100000000
          Target Port Queue Depth: 512
                  LUN Queue Depth: 32
              Optimization Policy: symmetric
                         Affinity: none
                       Error Text: -
                  Upgrade Pending: false
        Path Failover Time (sec): 0
Extend All Path Failure Event (secs): 0
```

Port queue depth setting

For FlexArray and V-Series

- From the advanced priv set run the following command to show the queue depth setting for each port on the storage array.
  - `storage array port show -fields max-queue-depth`

```
cluster1::*> storage array port show -fields max-queue-depth
name        wwnn             wwpn             max-queue-depth
----------- ---------------- ---------------- ---------------
NETAPP_VD_1 0021281100000000 0000000000000000 -
NETAPP_VD_1 0021501100000000 0000000000000000 -
NETAPP_VD_1 0021891100000000 0000000000000000 -
NETAPP_VD_1 0022671100000000 0000000000000000 -
```

- A value of '-' for Max Queue Depth indicates that the port does not have a specific max queue depth setting and is using the value set at the storage array level.

Port Performance information

For FlexArray and V-Series

- Use the storage array port show command to display performance information about storage array target ports.

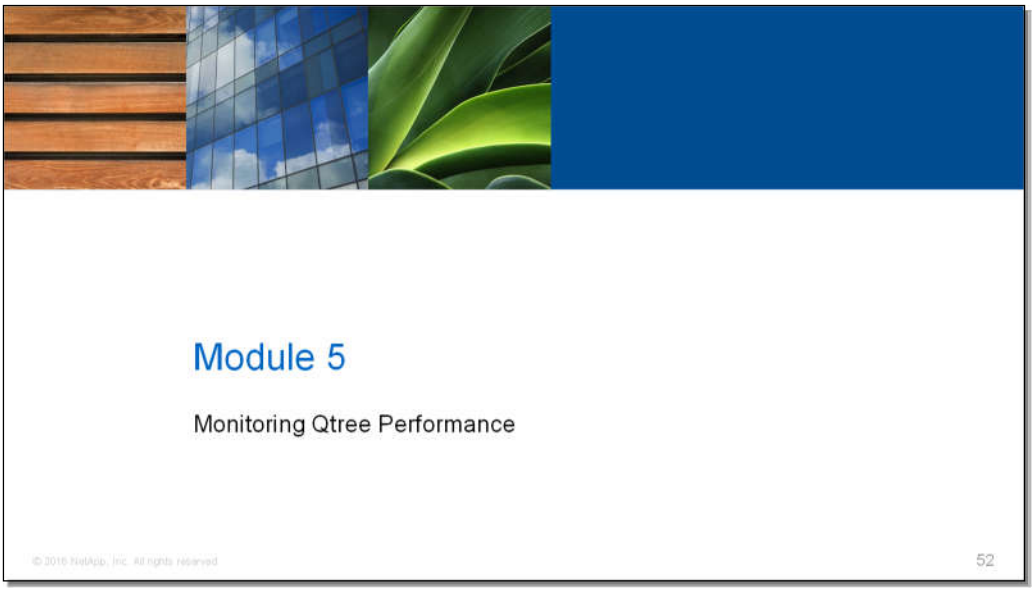
```
vgv3070f51::*> storage array port show

Array Name: HP2
WWNN: 50014380025d1500
WWPN: 50014380025d1508
Connection Type: fabric
Switch Port: vgbr300s70:9
Link Speed: 4   GB/s
Max Queue Depth: -
                                LUN                                    Link
Node       Initiator  Count   IOPS    KB/s    %busy    %waiting   Errs
---------  ---------  -----  ------  ------  -------  ----------  ----
vgv51-02       0a     21       2      53       0           0     0
vgv51-01       0a     21       2      48       1           0     0
```


The results of this command helps you determine if there are performance problems related to the ports. The %busy and %waiting values provide a high-level view of the performance on a port. If these values show a high percentage of requests waiting to be processed or show that the port is busy for a great percentage of time, then you might want to investigate further into the state of the port.

You can obtain more detailed information about ports using the storage array port show -fields command with the average-latency-per-iop, average-pending, average-waiting, max-pending, or max-waiting fields

To estimate the queue depth needed to achieve a certain I/O per second throughput, use this formula. Needed queue depth = (Number of I/O per second) x (Response time)

For example, if you need 40,000 I/O per second with a response time of 3 milliseconds, the needed queue depth = 40,000 x (.003) = 120.

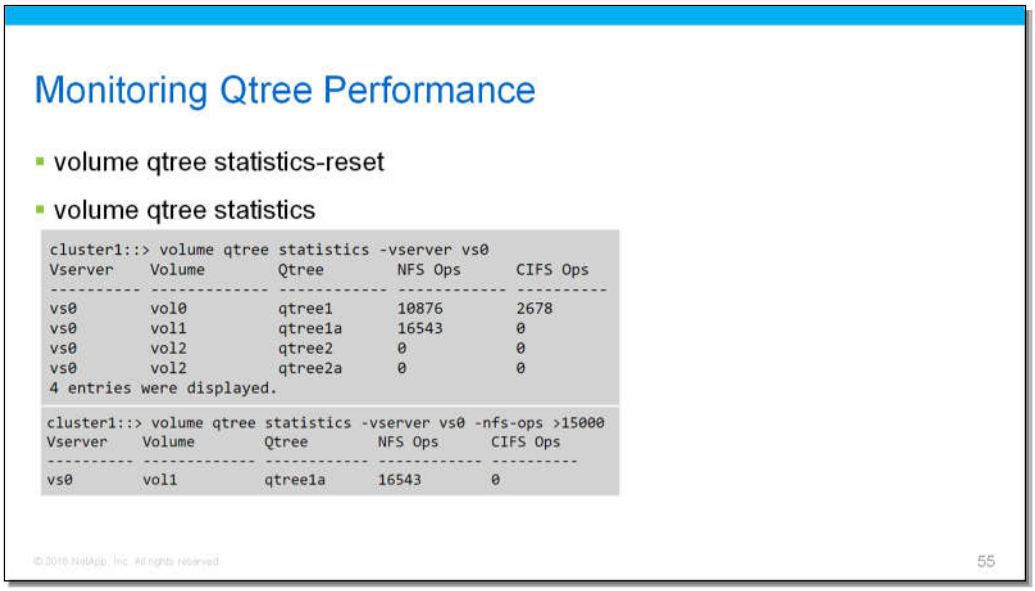

Module 5
Monitoring Qtree Performance
© 2016 NetApp, Inc. All rights reserved
52

About This Module

This module focuses on the following:

- Monitoring Qtree performance


Lesson 1
Monitoring Qtree Performance
© 2016 NetApp, Inc. All rights reserved
54



This command displays NFS and CIFS operations statistics for qtrees. Note that qtree statistics are available only when the volume containing the qtree is online. This command is not supported on Infinite Volumes.

Statistics are cumulative values from the time the volume is brought online or when the statistics have been reset by using the "**`volume qtree statistics-reset`**" command.

The command output depends on the parameters specified with the command. If no parameters are specified, the command displays the following statistics information about all qtrees:

Vserver name
Volume name
Qtree name
NFS operations
CIFS operations

Note:

Qtree statistics are not persistent. If you restart a node, if a storage takeover and giveback occurs, or if the containing volume is set to offline and then online, qtree statistics are set to zero.

This command is particularly useful if you have made home directories in a volume (as qtrees) and want to figure out which user is contributing to the overall volume activity.

Summary

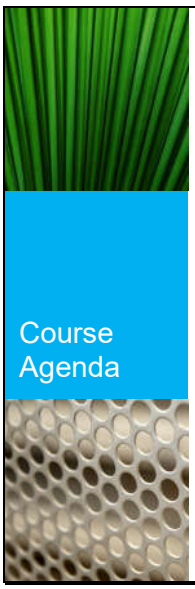
- CPU Troubleshooting
  - Common CPU Utilization Misconceptions
- Using QoS to Monitor Cluster Performance
  - QoS Command Details
- Flash Pool Monitoring and Prioritization
  - Monitoring Flash Pool Utilization
  - Prioritizing Flash Pool Workloads
  - How to determine Flash Pool Candidacy
- Flash Cache Monitoring and Prioritization
  - Monitoring Flash Cache Utilization
  - Prioritizing Flash Cache workloads
- SAN Queue Depth
  - Monitoring SAN Queue Depth
- Qtree monitoring
  - Monitoring qtree activity

Exercise

- The Lab exercise should take around 120 Minutes

Chapter 9

OnCommand System Manager
OnCommand Unified Manager
OnCommand Performance Manager
Harvest



Course Agenda

- System Manager Cluster Dashboards
- System Manager Predictable Performance
- OCUM/OPM Simplified operations management
- Unified Manager 7.0 & 7.1 features NEW
- Performance Manager 7.0 Product Overview
- Performance Manager 7.0 & 7.1 features NEW
- OCUM/OPM Installation and Deployment
- Harvest

Module 1

Oncommand System Manager

About This Module

This module focuses on the following:

- Using OnCommand System Manager to Monitor Performance

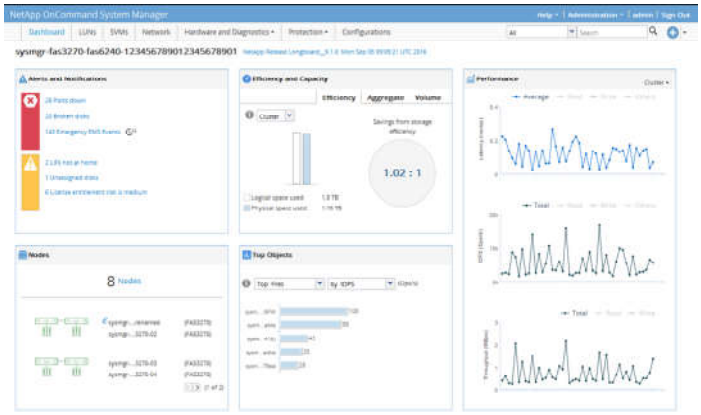
Lesson 1

OCSM 9.1 Cluster Dashboard

Cluster Dashboard

Landing page

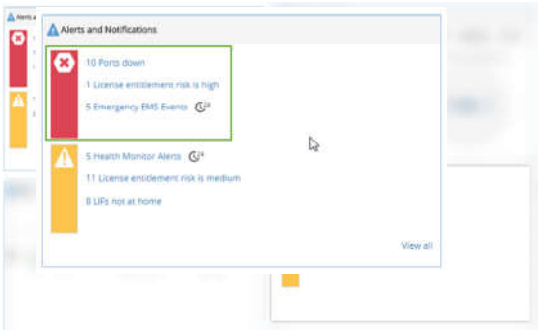
- Dashboard displays the current health and configuration details of a cluster which includes:
  - Alerts and Notifications
  - Efficiency and capacity
  - Performance
  - Nodes
  - Top Objects
  - Links to other pages for easier navigation



Cluster Dashboard

Alerts and Notifications portlet

- Alerts and Notifications includes:
  - Emergency EMS events
  - Health monitor alerts
- Indicates that the alerts and notifications are captured within last 24 hours
- Typically shows 3 messages in each section for better visibility.
- Click on “View all" link to open a pop-up to display all the messages

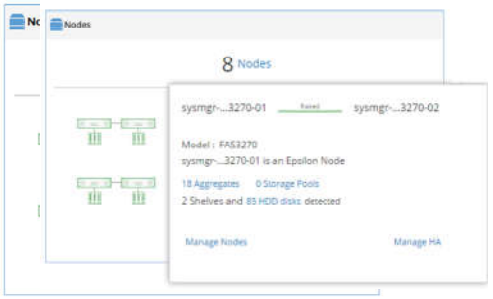


7

Cluster Dashboard

Nodes portlet

- Displays the number of nodes, name and a color coded pictorial representation of the status of the nodes in the cluster.
- Clicking on the number of Nodes will redirect the user to the 'Hardware and Diagnostics > Nodes' page.
- There is a horizontal scroll button available when there are more than 4 nodes in the cluster
- A pop-up appears with brief details of the node(s), and a link to manage the node(s) when clicking on the node(s)



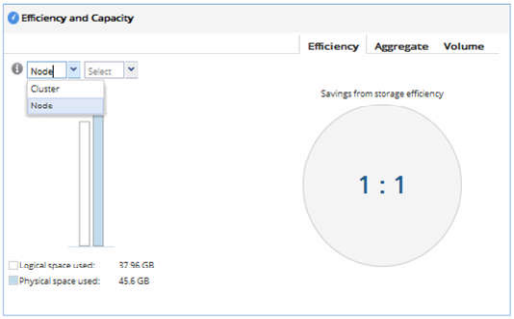
- ε - denotes the Epsilon node.
- ϟ - denotes All Flash FAS (AFF) node.

8

Cluster Dashboard

Efficiency and Capacity portlet

- Improvements include storage efficiency savings with enhancements to inline deduplication, inline compression, and object-compaction feature
- Savings are represented in the industry-standard format of ratio, which can be viewed for a cluster or per-node basis
- The refresh interval is 15 minutes
- Efficiency tab is available on FAS and All-Flash FAS (AFF) optimized clusters



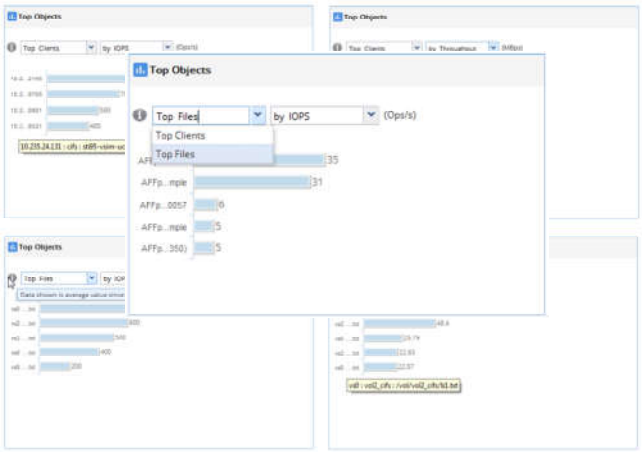
9

In OCSM 9.0, storage efficiency savings were visible in Efficiency tab on All Flash FAS (AFF) systems only. For FAS systems, the portlet was called "Capacity" and contained tabs for Aggregate and Volume only. With OCSM 9.1, we now display Efficiency savings for both FAS and AFF systems.

Cluster Dashboard

Top Objects

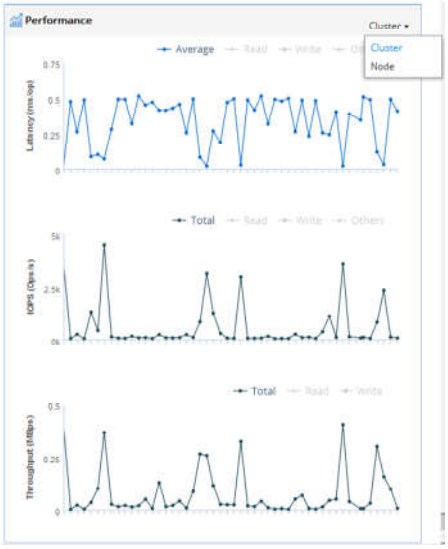
- Display Top-Clients and Top-Files information by IOPS or Throughput
  - Hovering over specific locations retrieves appropriate information in tooltips
  - The Sampling Interval is 1 minute
  - The Duration of samples expiration is set to 5 minutes.



Cluster Dashboard

Performance portlet

- You can view the performance for the cluster or per node basis
- You can toggle to enable/ disable viewing the graph for:
  - Average / Total
  - Read
  - Write
- Hovering over a point in the graph displays the performance value at that point in time

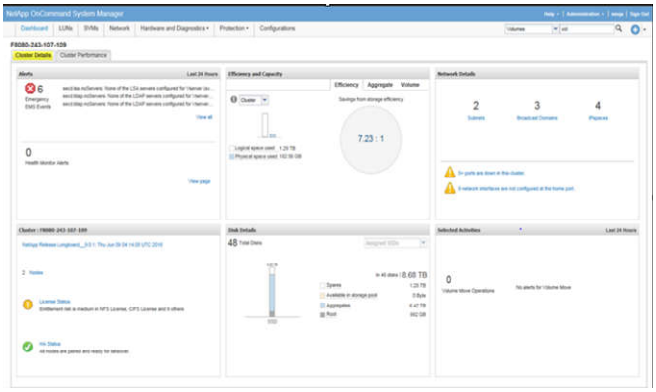


11

Dashboard

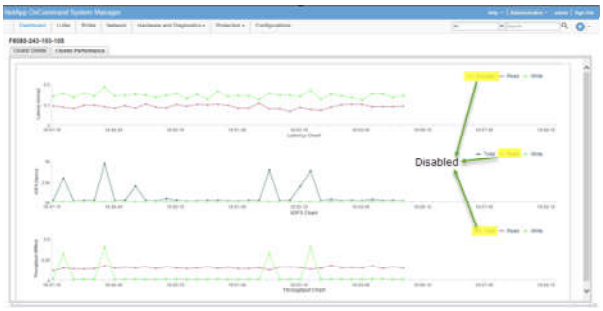
Cluster details

- Dashboard displays the current health and configuration details of a cluster which includes:
  - Alerts for the last 24 hours
    - Emergency EMS events
    - Health monitor alerts
  - Efficiency and capacity
  - Network details
  - Cluster details
  - Disk details
  - Links to other pages for easier navigation



Dashboard
Cluster performance

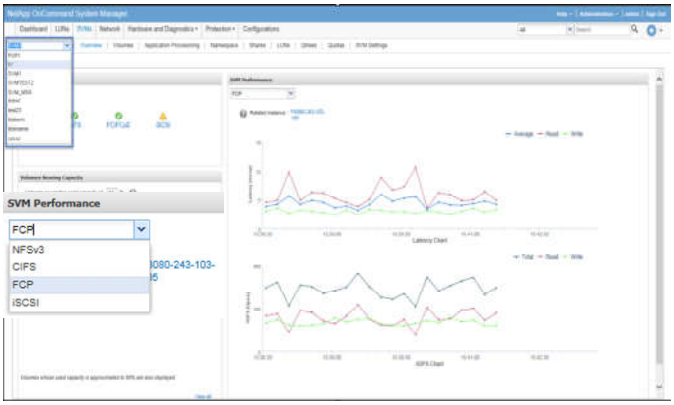
- Displays performance graphs for:
  - Latency (ms/op)
  - IOPS (Ops/s)
  - Throughput (MBps)
- Hovering over a point in the graph displays the performance value at that point in time
- You can toggle to enable/ disable viewing the graph for:
  - Average / Total
  - Read
  - Write



SVM Dashboard

Dashboard for Storage Virtual Machine (SVM)

- Displays current health and configuration details of an SVM.
- Consists of three portlets:
  - Protocol status
  - Volumes nearing capacity
  - SVM performance graphs for enabled protocols selected from filter displayed in :
    - Latency (ms/op)
    - IOPS (Ops/s)
- Drop-down menu option to filter SVM
- The statistics are polled every 15 seconds.

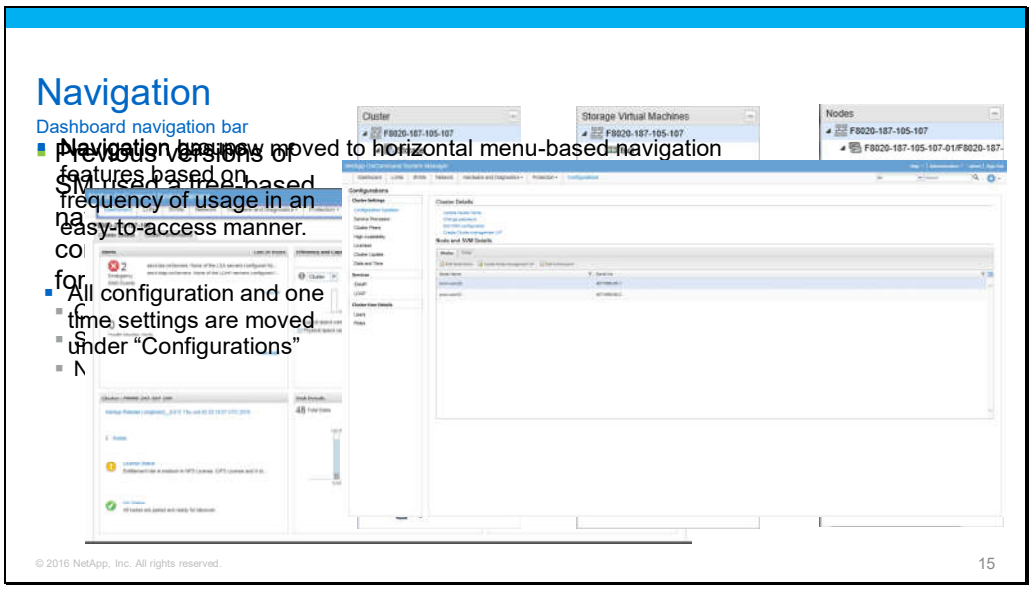


14

SVMs > Select the SVM > Click ‘Manage’ > SVM Dashboard

The SVM dashboard displays a warning message if:
1. SVM is stopped (Displays “The selected Storage Virtual Machine is in the Stopped state.“)
2. SVM root volume is corrupt or deleted.
3. SVM attributes are edited and the details are not refreshed before accessing the said SVM.

The ‘Overview’ page is not valid for SVMs with Infinite Volume and will be redirected to the ‘Volumes’ page.



System Manager classifies the features into three categories viz. Cluster, SVM & Node and the UI presents them in a tree format. Under each main category also there are sub groups. First time user has difficulty in finding a feature as they need to expand the nodes and search for them. Repeated uses also need to expand all the nodes to access it. Navigation simplicity project is trying to solve this issue.

Cluster - All cluster level features are grouped under cluster node where there sub groups like services, storage, configuration etc.

Vserver - List of SVMs and related objects and their functionality is covered under Vserver tree.

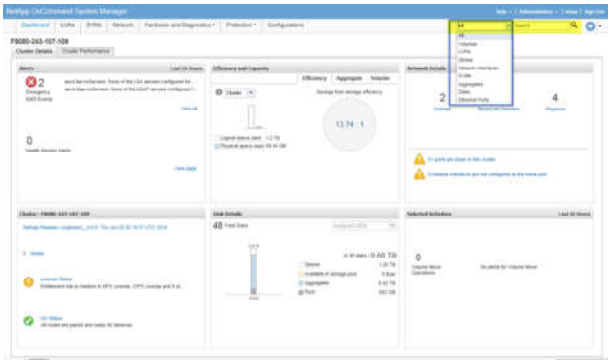
Node - Node inventory and few storage objects and Autosupport, Flashpool statistics is covered under Node tree.

All cluster-wide configuration is grouped under Configurations

Search

Search bar

- Search for various object types in the storage system
- Search text allows all special characters
- Search queries are case sensitive; only actual name search will provide correct results
- Further filter your search by using the Search filter



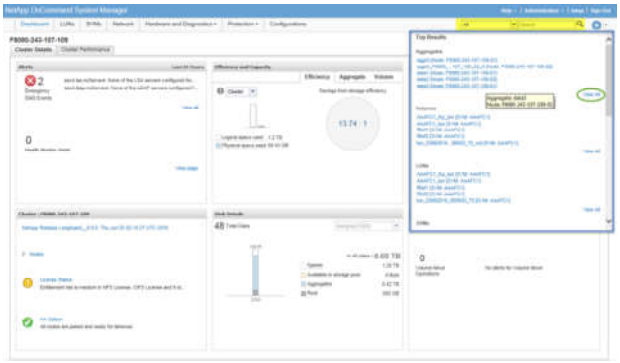
16

The newer and simpler search bar in the topmost toolbar allows to search for certain object types as specified in the filter such as:

1. All
2. Volumes
3. LUNs
4. Qtrees
5. Network Interfaces
6. SVMs
7. Aggregates
8. Disks
9. Ethernet Ports

Search

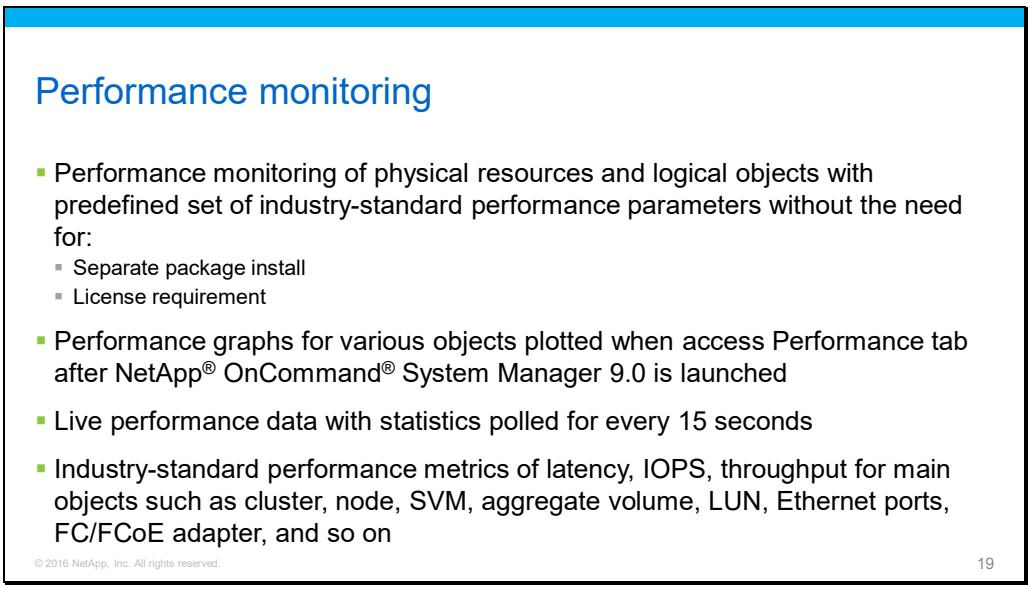
Search results



- Search results are not displayed in any particular order.
- Search results consists of:
  - Object type
  - Hyperlinked searched object
  - Tooltip with more information about the returned object.
  - View All to navigate to the actual page of object
- Clicking the search result hyperlink takes you to the page for that searched object

Lesson 2

OCSM 9.x Predictable Performance



Industry Standard Performance Matrices

►**Throughput (MB/s)** = **IOPS(IO/s)** X **IO Block Size(KB)**

If the application block size changes to 64k from 8k (i.e X8), the throughput would be same, even if the IOPS dropped by 1/8th. Hence an *idea* about the IO Block Size can be *estimated* from Throughput and IOPS.

► **Little's Law:**
Average Queue-depth = **Average IOPS (IO/s)** X **Avg. Response Time (usec)**

Intuitively Little's law gives the average queue-depth at the different queuing point in the IO path. Basically per second, how many IO the resource/object can handle and the average latency (round-trip time) of each IO. So the cross-product gives an estimate of the IO queue buildup at each device.

►**Concurrency**. The **Average Queue-depth** of the device is indicative for the

Concurrency of the device - how many concurrent IO the device can process. For example: in VMware world, a LUN has a default queue-depth of 32 means it can handle 32 IO in parallel. However for HBA it is pretty high (=1024).

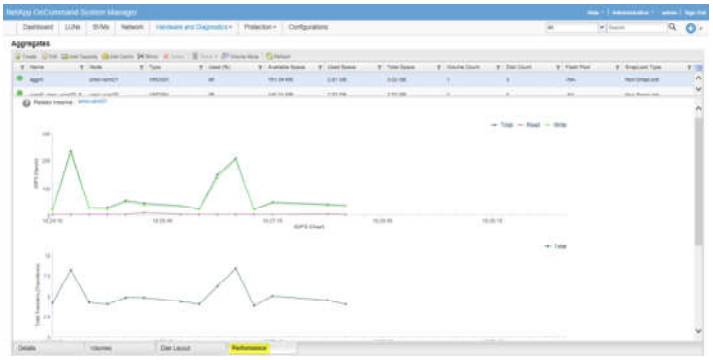
►**Utilization**: Also, intuitively, the Little's Law can give it, when looking from the device/object perspective.

Hence with the "Industry Standard Performance Matrices" of (Data-rate, IOPS and Response Time) triplet is fair enough for getting a black box analytical understanding of the device in hand.

Performance monitoring

Aggregate

- Parameters displayed:
  - IOPS (Ops/s): Number of blocks read/write/total per second on the aggregate
  - Total Transfer (Transfers/s): Total transfers per second serviced by the aggregate with IOPS by HDD/SSD - (HDD/SSD/Total)
  - Workload Impact (%): Impact of (non-read) workload on the aggregate

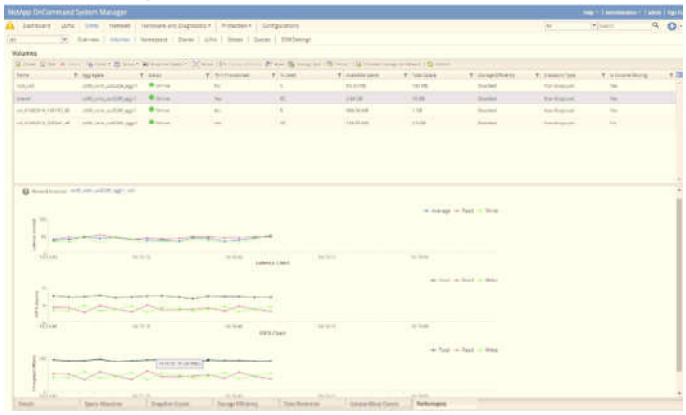


User workload can come in one of two forms, modify ops and non-modify ops. Modify ops are those such as write, set attribute, rename, remove, clone, etc. Modify ops by nature update or create data which needs to be written to disk. Prior to the data making it to disk, the file-system journal(nvlog) and in-core dirty buffers are consumed by this data. This chart shows the impact of such non-read portion of the workload on the aggregate,classified by journal(nvlog) usage % and dirty buffer cache %.

Performance monitoring

Volume

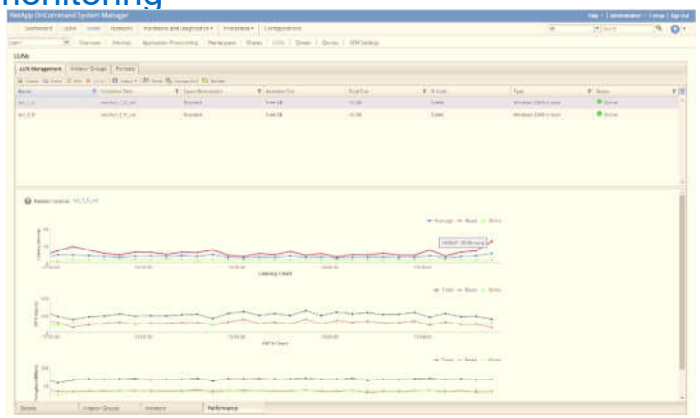
- Parameters displayed:
  - Latency (ms/op)
  - IOPS (Ops/s)
  - Throughput (MBps)



Performance monitoring

LUN

- Parameters displayed:
  - Latency (ms/op)
  - IOPS (Ops/s)
  - Throughput (MBps)



Module 2

Unified Manager 7.1 Integration with Performance Manager 7.1

About This Module

This module focuses on the following:

- How to use Oncommand Unified Manager and Oncommand Performance Manager
- Monitoring capabilities in OCUM and OPM

Lesson 1

Simplified Operations Management

Simplified Operations Management

Unified Manager 6.4 and later releases

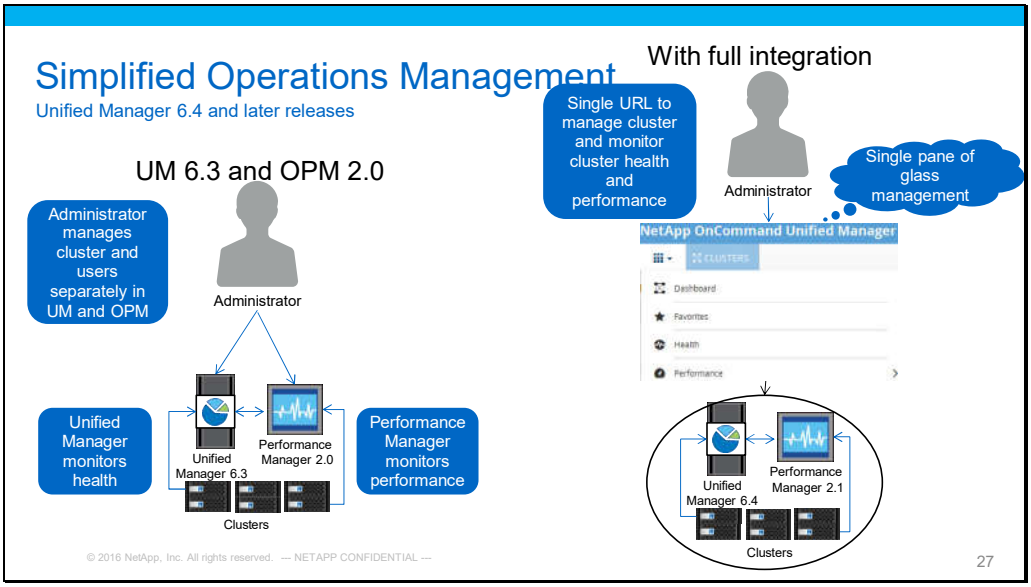
- No more redundant (multiple) logins
  - Single URL to manage NetApp storage
  - One login for Unified Manager 6.4,Performance Manager 2.1 and later releases
- Access everything in one place
  - View cluster health and performance attributes through central dashboard
  - Single location to configure clusters, users and authentication attributes
- Faster access to your most frequently managed storage objects
  - New “Favorites dashboard”

26

No more redundant (multiple) logins
- Single URL to manage NetApp storage
- One login for Unified Manager 6.4 and Performance Manager 2.1

Access everything in one place
- View cluster health and performance attributes through central dashboard
- Single location to configure clusters, users and authentication attributes

Faster access to your most frequently managed storage objects
- New “Favorites dashboard



Today an administrator manages cluster and users separately in UM and OPM. This means he needs to manage each cluster object(s) twice for measuring performance and health. This leads to a lot of management overhead.

With Unified Manager 6.4, an administrator can monitor and manage cluster health , performance, capacity , protection and reporting from a single URL and single location.

Lesson 2

Unified Manager 7.0 New Features

OnCommand Unified Manager 7.0

Unified Manager 7.0 New Features

Feature	Content Overview
ONTAP 9.0 Feature Support	▪ Support for SnapLock volumes and aggregates
	▪ Supports RAID Triple Parity
	▪ Flash monitoring improvements – Identifies whether a Flash Pool volume is priority volume
	▪ i18n Compatibility for Qtree monitoring
ONTAP EMS **Monitoring**	Unified Manager 7.0 can directly subscribe to ONTAP events and showcase these as regular events in the Unified Manager events window from ONTAP event notifications.

OnCommand Unified Manager 7.0

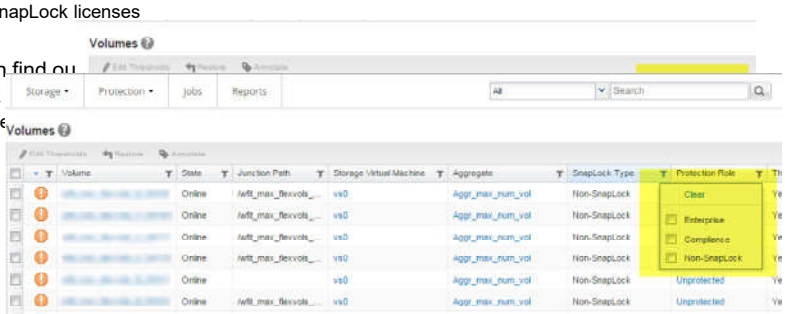
Unified Manager 7.0 New Features Contd.

Feature	Content Overview
New Platform support	▪ Monitoring support for ONTAP Select and ONTAP Cloud Systems. Note: There will be no IBM branded version support of OnCommand Unified Manager 7.0.
Enhanced support for Importing Custom Reports	▪ User can import a custom report into their Unified Manager 7.0 instance without the need to edit user credentials and Unified Manager URL in the report file.

OnCommand Unified Manager 7.0 Feature

ONTAP 9 - SnapLock Monitoring

- Unified Manager can -
  - Monitor SnapLock volumes and aggregates (Compliance & Enterprise)
  - Report SnapLock licenses
- A user can find ou SnapLock
  - Aggregate
  - Volume

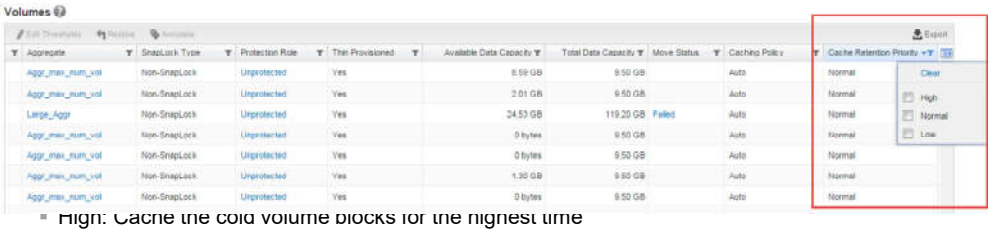


31

OnCommand Unified Manager 7.0 Feature

Flash Enhancements

- With Unified Manager 7.0, you can monitor flash pool volumes and

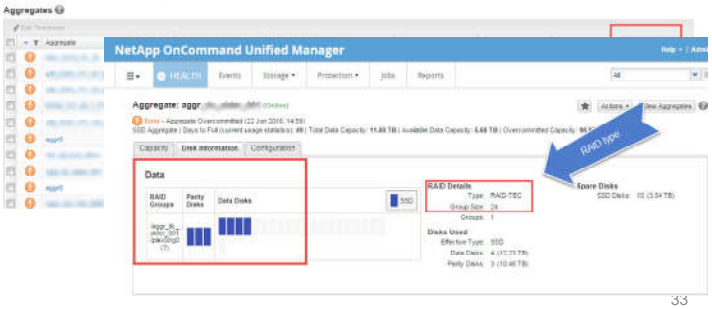


  - High: Cache the cold volume blocks for the highest time

OnCommand Unified Manager 7.0 Feature

ONTAP 9 - RAID Triple Parity

- Unified Manager 7.0 allows you to:
  - Monitor RAID triple parity aggregates introduced in ONTAP 9.0
  - Display the RAID layout clearly marking triple parity disks
  - View available capacity based upon the triple parity disks
- You can view this in t
  - In the aggregate grid or
  - In the aggregate detaile

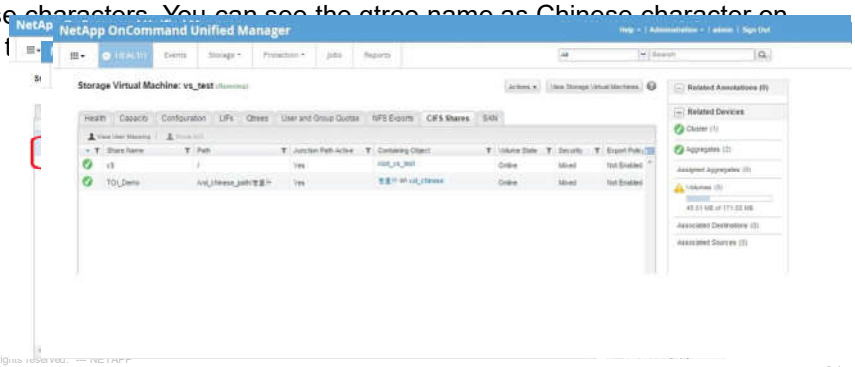


33

OnCommand Unified Manager 7.0 Feature

i18n compatibility

- With Unified Manager 7.0 you can monitor qtree names scribed in Chinese or Japanese characters. You can see the qtree name as Chinese character on different t



OnCommand Unified Manager 7.0 Feature

EMS Notifications

- Unified Manage
  showcase these
  has support for -
  - Forwarding of EM
  - Alert configuratior
  - Alert script for EM
- To perform this y
  console; access
  commands in th
  a) To display all ev
  b) To display all ev
  c) To display all ev
  d) To display the de
  messagename C
  e) To display the action for an event, using the message name from a or b output: clus::> event route show -messagename CR.Data.File.Inaccessible -fields messagename,severity,action

```
:> event route show -severity ERROR

                                                                  Freq     Time
Message                          Severity       Destinations     Threshd  Threshd
-------------------------------- -------------- --------------- -------- --------
LUN.create.nvfailed.volume       ERROR          -               0        0
LUN.inconsistent.blocks          ERROR          -               0        0
LUN.inplace.dumpRestoreFail      ERROR          -               0        0
LUN.inplace.restore.failed       ERROR          -               0        0
LUN.metadb.version.mismatch      ERROR          -               0        0
LUN.nvfail.vol.proc.complete     ERROR          -               0        0
LUN.nvfail.vol.proc.started      ERROR          -               0        0
LUN.oovc.version.mismatch        ERROR          -               0        0
LUN.op.nvfail.offline            ERROR          -               0        0
LUN.outplace.dumpRestoreFail     ERROR          -               0        0
LUN.provisioning.failed          ERROR          -               0        0
LUN.transition.fail.no.i2p       ERROR          -               0        0
LUN.transition.fail.no.space     ERROR          -               0        0
LUN.transition.failed            ERROR          -               0        0
LUN.transition.incons.attrs      ERROR          -               0        0
LUN.transition.unsupp.clone      ERROR          -               0        0
LUN.vdisk.cnt.exceeded           ERROR          -               0        0
LUN.vdisk.prov.cnt.exceeded      ERROR          -               0        0
LUN.vol.no.space                 ERROR          -               0        0
LUN.vol.proc.fail.no.space       ERROR          -               0        0
LUN.vproc.version.mismatch       ERROR          -               0        0
```


35

OnComm

EMS Notifications

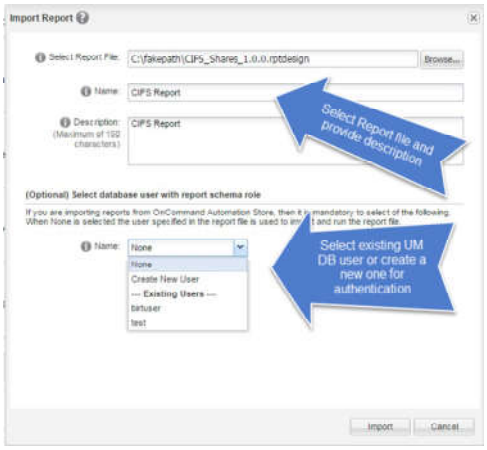
- Once you h
  OnComma

OnCommand Unified Manager 7.0 Feature

Importing Custom Reports - Simplified

- Starting OnCommand Unified Manager 7.0, users can import a custom report without the complexity of editing details in report file.
- Theory of operation
  - While importing a Custom report, a user is provided the “Select database user with report schema role” option to key in a new (or) select from an existing DB user with report schema role and password.
  - This will replace the user credentials and Unified Manager URL in the report file with the current URL the user is in and data that is supplied

Note: Credentials are saved in base64 encoding format under the UM Database



Lesson 3

Unified Manager 7.1 New Features

OnCommand Unified Manager 7.1

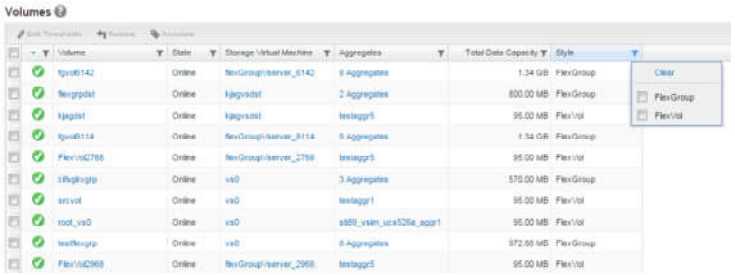
Unified Manager 7.1 New Features

Feature	Content Overview
ONTAP 9.0, 9.1RC1 Feature Support	▪ Support for FlexGroup volumes
New Platform Support	▪ Support for ONTAP Select, ONTAP Cloud, all FAS Platforms that are supported in ONTAP 9.0, 9.1 RC1 and earlier. **Note:** There will be no IBM branded version support of Unified Manager 7.1

OnCommand Unified Manager 7.1 Features

Monitoring FlexGroup volumes

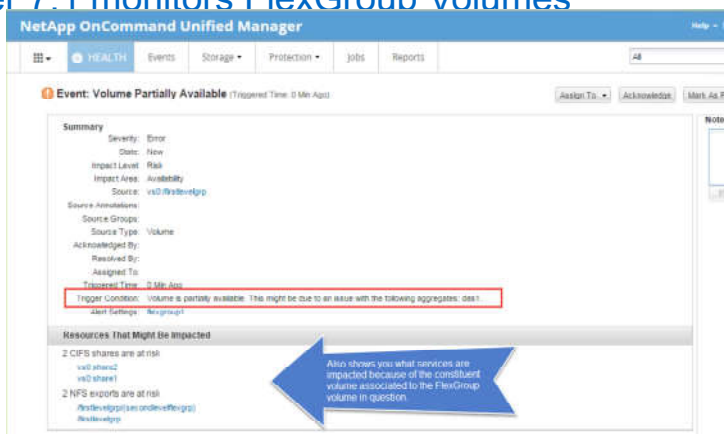
- Unified Manager is able to monitor and manage FlexGroup volumes.
- Manage and monitor:
  - Availability
  - Capacity



Unified Manager 7.1 monitors FlexGroup Volumes

Availability Monitoring

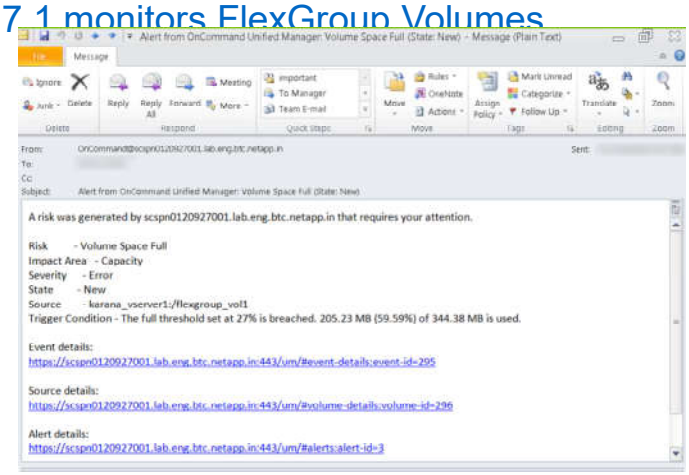
- Create alerts for FlexGroup volumes to be viewed as
  - Email (or) Trap
  - Under the events section in Unified Manager 7.1
- Get recommendations to perform root cause analysis
- Get to know the services that might be impacted due to the unavailability.



Unified Manager 7.1 monitors FlexGroup Volumes

Capacity Monitoring

- Configure capacity alerts for FlexGroup volumes and/or the constituent volumes that breaches the threshold limits defined on the them.



Restrictions when monitoring FlexGroup volumes

Unified Manager 7.1RC1 has the following restrictions when monitoring FlexGroup volumes

- Unified Manager does not support the discovery/monitoring of protection aspects of FlexGroup volumes, so the Protection operations are disabled for FlexGroup volumes
- Annotations and Groups are not supported
- When monitoring FlexGroup volumes, the FlexGroup constituents are not exported when you select the Export button from the Volumes page
- The event “Thin Provisioned Volume At Risk” is not raised for thin provisioned FlexGroup volumes
- You cannot calculate the amount of disk space that can be reclaimed if one or more Snapshot copies are deleted
- When using the Unified Manager CLI, the command “um volume list” lists only FlexVol volumes when sorted based on aggregates or nodes. The “um aggr list” and “um cluster node list” commands also have issues when sorting for a single FlexGroup volume
- FlexGroup functionality has not been integrated with the Reports feature. Some reports that include volumes in their output may include FlexGroup constituent volumes, while other reports may not

Lesson 4

Performance Manager(OPM) 7.0 Product Overview

OnCommand Performance Manager 7.0

What's new ?

- Performance capacity used and free (%) metric
- Performance capacity advanced chart
  - breakdown by user protocol workload and background activity
- Available IOPS for node and aggregate
- Data continuity collection poll
- Flash cache performance statistics
- Failover planning
- User defined threshold based on Performance capacity used (%) counters
- Download third party dependencies for installation using Red hat subscription

OnCommand Performance Manager 7.0

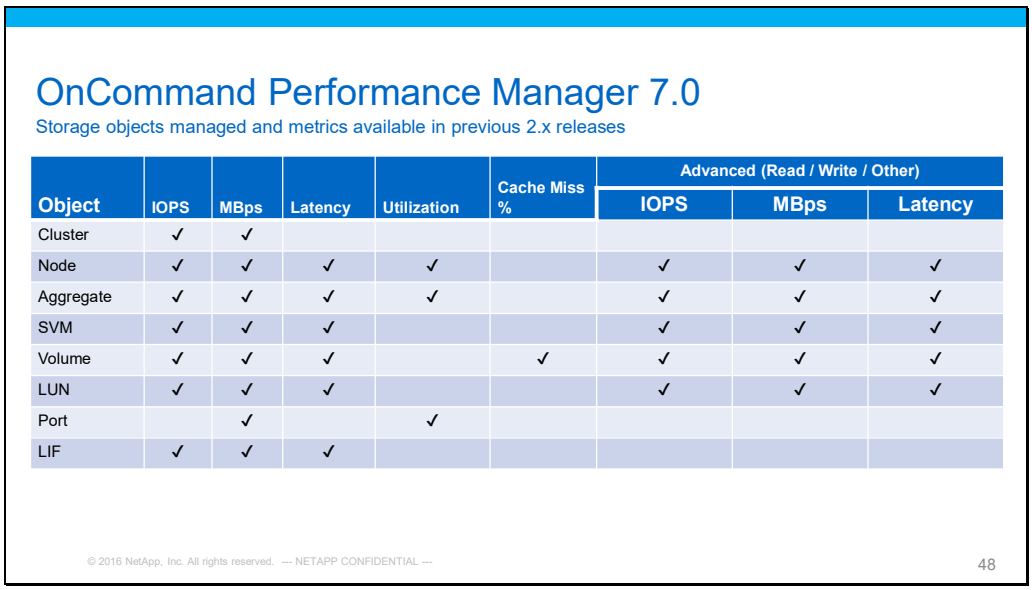
What's the same ?

- Unified Manager and Performance Manager full integration configuration mode
  - Central dashboard to manage cluster health and performance
- Access to your critical storage objects via Favorites dashboard
- 15 minute configuration change interval
- 5 or 10 or 15 minute granularity cluster performance collection cycle
- Global search
- External data provider connector

OnCommand Performance Manager 7.0

What's the same ?

- Bully-victim contention analysis
- Automated baseline and dynamic threshold establishment
- System-defined and user defined thresholds
- Workload cluster component latency contribution
- Efficient data collection
- Choice of VMware or RHEL install



OnCommand Performance Manager 7.0

Storage objects managed and metrics available in previous 2.x releases

Object	IOPS	MBps	Latency	Utilization	Cache Miss %	Advanced (Read / Write / Other)		
						IOPS	MBps	Latency
Cluster	✓	✓						
Node	✓	✓	✓	✓		✓	✓	✓
Aggregate	✓	✓	✓	✓		✓	✓	✓
SVM	✓	✓	✓			✓	✓	✓
Volume	✓	✓	✓		✓	✓	✓	✓
LUN	✓	✓	✓			✓	✓	✓
Port		✓		✓				
LIF	✓	✓	✓					

48

The OPM 2.0 release is primarily about managing performance while maintaining the overall philosophy to simplify operating product. All metrics are presented using standard industry accepted terminology. **Every effort was made to fill critical functionality gaps against Performance Advisor**. The objective is to stop exposing ONTAP internals to the end user and let OPM utilize that information on behalf of the user to simplify operation.

Thus OPM 2.0 made every effort to:

- monitor and display the utilization of objects that are shared with limited resource capacity
- describe performance with easily consumable and consistent set of metrics

OnCommand Performance Manager 7.0

Additional new metrics

Object	Flash Cache Reads	Performance Capacity Used (%)	Advanced Charts		
			Performance Capacity Advanced	Available IOPS	Failover Planning
Cluster					
Node	✓	✓	✓	✓	✓
Aggregate		✓	✓	✓	
SVM					
Volume					
LUN					
Port					
LIF					

49

The OPM 2.0 release is primarily about managing performance while maintaining the overall philosophy to simplify operating product. All metrics are presented using standard industry accepted terminology. **Every effort was made to fill critical functionality gaps against Performance Advisor**. The objective is to stop exposing ONTAP internals to the end user and let OPM utilize that information on behalf of the user to simplify operation.

Thus OPM 2.0 made every effort to:

- monitor and display the utilization of objects that are shared with limited resource capacity
- describe performance with easily consumable and consistent set of metrics

Performance Manager 7.0

Dashboard, Monitoring, Events

- Dashboard
  - What needs my attention now?
  - What is the performance status of my storage systems?
- Performance visualization
  - How is my storage system performing in detail?
  - Are there trends that may cause future issues?
  - Is there a correlation between resource utilization and service time?
- Events and alerts
  - Am I meeting all critical service level objectives and agreements?
  - What is abnormal and why?

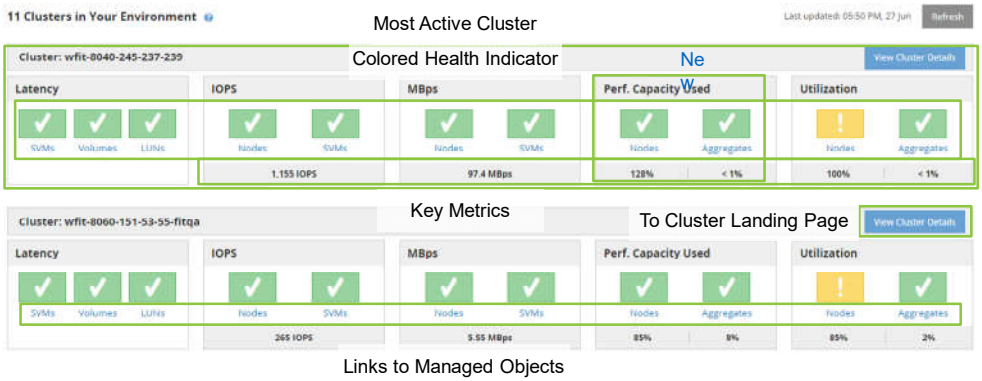
Performance Manager 7.0

Dashboard

- Clusters sorted by level of importance
  - Clusters with performance issues displayed in order of severity
  - Updated automatically at each five-minute collection period
- Shows cluster performance status and health
  - Key performance metrics – Latency, IOPS, MBps, Utilization,
  - Performance capacity used (%) -NEW
  - Highlights active resources in cluster
  - Color-coded health indicators
- Launching point for further analysis and navigation of storage objects
  - Click view cluster details
  - Click object icons to display inventory page for the object

Performance Manager 7.0

Dashboard User Interface

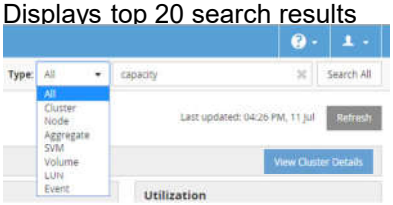


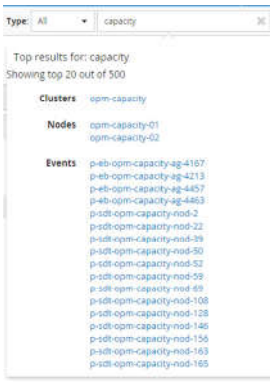
52

Performance Manager 7.0

Global Search

- Quickly access a specific object or event
- Search by object or across all objects
- Displays top 20 search results
-





53

Performance Manager 7.0

Beyond the dashboard

- Cluster Summary page
  - Every cluster has a landing page
  - Summarizes performance and events over previous 3 days
  - View and analyze trends
- Top performers page – Enhanced in 7.0 release
  - Top performing storage objects based on selected performance metric – New
  - Highest or lowest performance value in the selected time range
- Cluster information page – New
  - View Physical and logical attributes of cluster
- Access to performance explorer

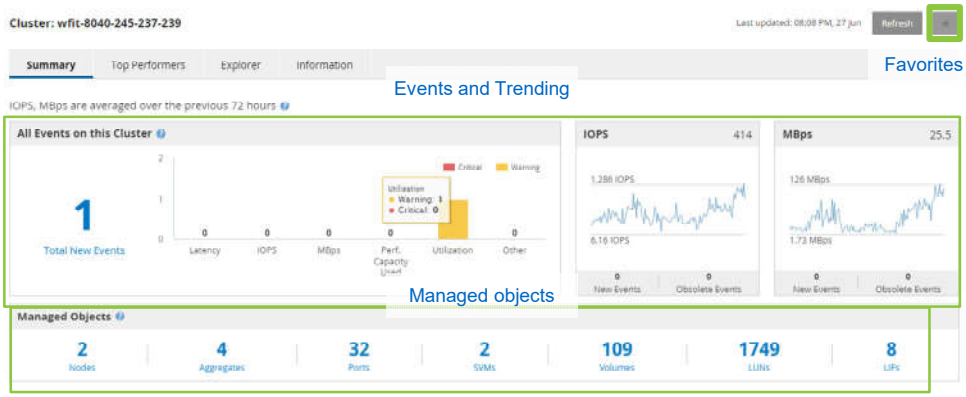
Performance Manager 7.0

Cluster landing page

- Favorites button – Introduced with Performance Manager 2.1
  - Available when Performance Manager fully integrated with Unified Manager
- Summary page
  - Cluster events pane
    - All events on this cluster – New performance events for the preceding 72 hours
  - IOPS counter panel - Averaged over the previous 72 hours
  - MBps counter panel – Averaged over the previous 72 hours
  - Managed objects pane
    - Managed objects count – Nodes, Aggregates, Ports, SVMs, Volumes, LUNs, LIFs
    - New objects discovered at 15 minute interval

Performance Manager 7.0

Cluster landing page



56

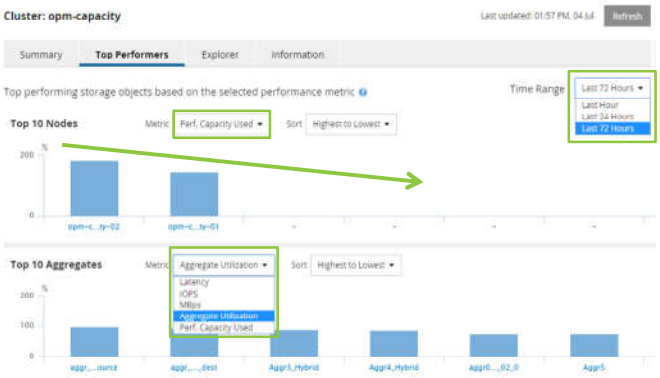
Performance Manager 7.0

Cluster landing page

- Top Performers page – Enhanced in 7.0 release
  - Displays top performance storage objects based on selected performance metric
  - Highest or lowest performance value in the selected time range
  - Top 10 storage objects
    - Nodes
    - Aggregates
    - SVMs
    - Volumes
    - LUNs
  - Performance metric
    - Latency
    - IOPS
    - MBps
    - Performance Capacity Used (for nodes and aggregates) - NEW
    - Utilization (for nodes and aggregates)
  - Last hour, Last 24 hours ,Last 72 hours time range for viewing top performers

Performance Manager 7.0

Top Performers page

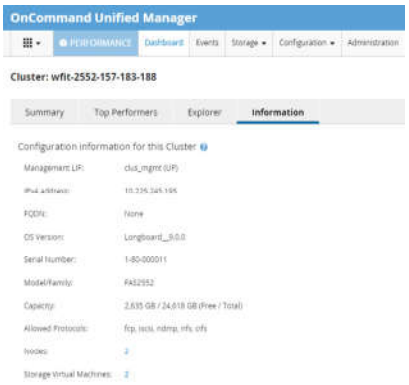


58

Performance Manager 7.0

Cluster Information page

- View Physical and logical attributes of cluster
  - Management LIF
  - IP Address
  - FQDN
  - OS Version
  - Serial Number
  - Model / Family
  - Capacity(Free/Total)
  - Available Protocols
  - Nodes
  - SVMs
  - LIFs



Performance Manager 7.0

Beyond the dashboard and cluster landing page

- Object landing page
  - Every object has a landing page
  - Summarizes performance and events over previous 3 days
  - View and analyze trends
- Performance explorer page
  - Modular component of all landing pages (cluster and objects)
  - Compare related objects (eg. parent, sibling, and child)
  - Select any time range
  - Launch into detailed views
  - Performance capacity used(%) charts – nodes and aggregates
  - Available IOPS – nodes and aggregates
  - Failover Planning

Performance Manager 7.0

Object landing page – Node Summary page

61

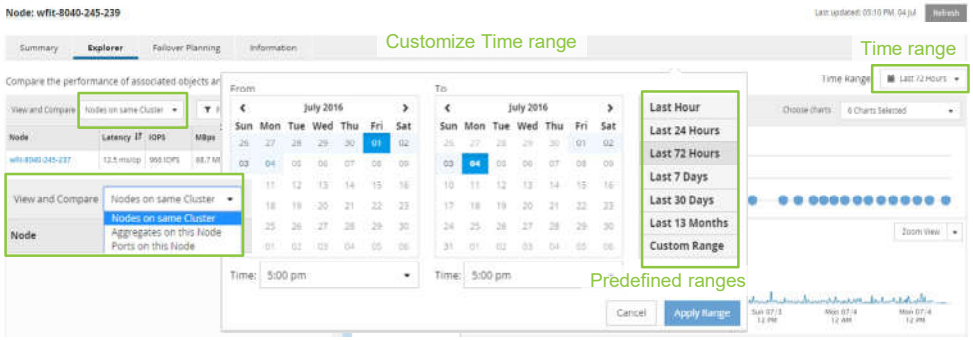
Performance Manager 7.0

Performance Explorer

- Use for trending and trouble-shooting
- Consistent user interface across all managed objects
  - Time ranges (one hour granularity)
  - Object comparison to parent, child, and siblings
  - Event correlation
  - Powerful filtering
  - Add and remove metric charts
  - Launch into detailed views

Performance Manager 7.0

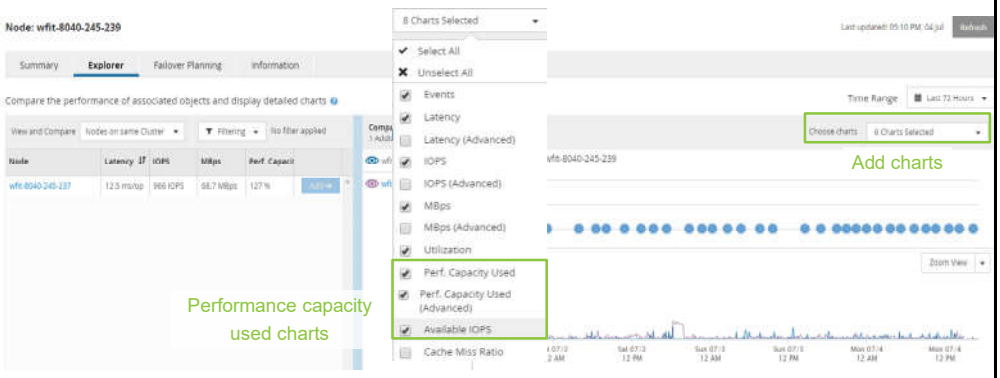
Customize Time range ,view & compare related objects



63

Performance Manager 7.0

Performance Explorer – Add associated object performance chart & New Performance capacity charts



64

Performance Manager 7.0

ONTAP 9 New Performance charts and features

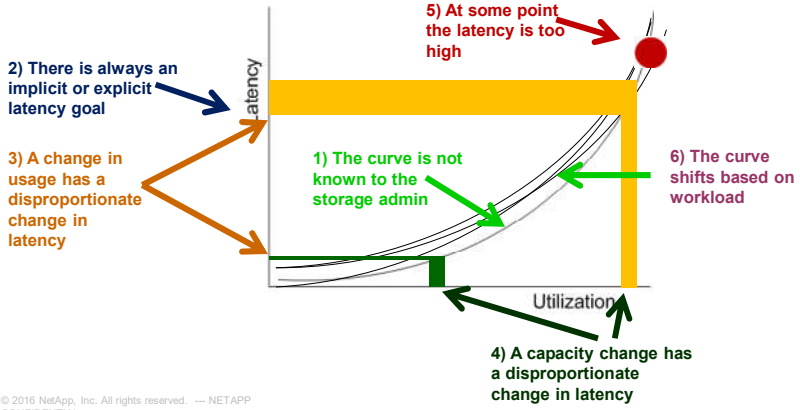
- Features available in Performance Manager with ONTAP 9
  - Performance capacity used (%)
    - Available for node and aggregate storage objects
  - Performance capacity used (%) – Advanced chart
    - Breakdown of user protocol workload and background activity
  - Available IOPS
    - IOPS available at the node and aggregate
  - Node Failover planning helps in the below scenarios
    - Understand implications of node failover to surviving node in HA pair
    - Understand workload patterns to plan maintenance activity on nodes

Lesson 5

Performance Manager 7.0 Performance Capacity metric

OnCommand Performance Manager 7.0

Understanding Latency versus Utilization



67

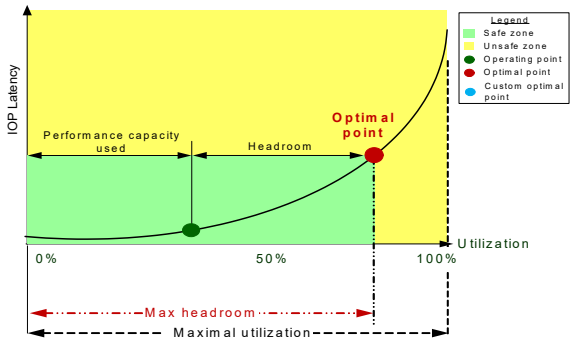
Performance Manager 7.0

Performance Capacity Used and Utilization

- Performance Capacity Used incorporates both utilization and latency
- Perf. Capacity Used can go to 100% and that is alright
- Perf. Capacity Used can exceed 100%
  - Past diminishing returns
  - A little more work results in much more latency
- Utilization is useful for a **resource** that is not concerned about latency

Performance Manager 7.0

Headroom and Performance capacity used

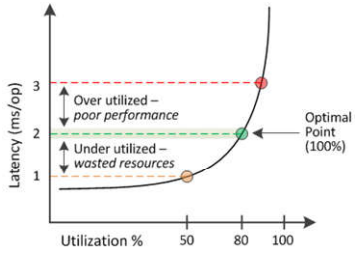


- Optimal point
  - A little increase in utilization beyond optimal point has a bigger increase in latency
- Headroom – ONTAP 9
  - Metric used in ONTAP 9
  - Depends on workload mix
  - A resource remaining useful capacity when measured from optimal point
- Performance Capacity Used
  - Metric used in Performance Manager 7.0
  - Optimal point minus headroom
  - Performance metric for node and aggregate

69

Performance Manager 7.0

Performance Capacity Used and Optimal Point



- Performance Capacity used (%)
  - Incorporates resource utilization and latency (response time) into a single performance metric
- Optimal point
  - Point at which output of a resource is optimal within desired latency limits
  - Depends on resource physical hardware capabilities
    - Nodes with higher cores have optimal point higher than nodes with lower cores
    - Workload mix
  - 100% performance capacity used is called the " optimal point"
  - Below 100% means nodes or aggregates are underutilized
  - Above 100% means latency is becoming undesirable

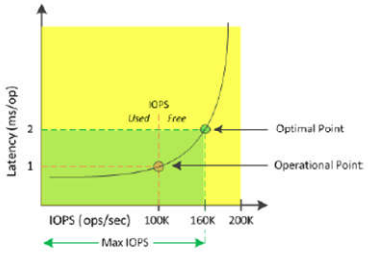
Performance Manager 7.0

Performance Capacity Used Chart - Color banding

- Performance capacity used color banding in charts
- Green band indicates object (node or aggregate) under utilized
- Yellow band indicates object optimal range
- Red band indicates object over utilized

Performance Manager 7.0

Available IOPS



- Available IOPS is the estimated amount of remaining average IOPS for a resource
- Node
  - Total IOPS based on physical characteristics of the node
  - number of CPUs, CPU speed , amount of RAM
- Aggregate
  - Total IOPS based on physical properties of disk
  - SATA, SAS, SSD disk
- Calculation based on FAS platform
  - For e.g. Performance capacity used value of 60% means
    - Available IOPS 10000 on FAS 8080
    - Available IOPS 1000 on FAS 2520

Performance Manager 7.0

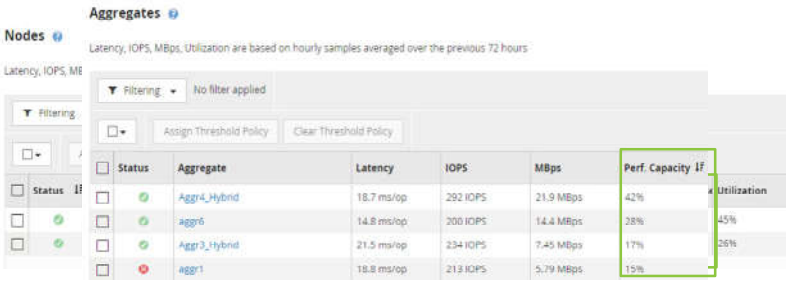
Performance capacity – Use cases

- User or storage administrators look at node or aggregate utilization metric to determine storage utilization with a conservative outlook
  - Storage resource utilization is maintained well within safety margins
  - Cluster is not optimally utilized
- Performance capacity helps with
  - Provisioning and balancing workloads
  - Preventing overloading of a node or aggregate
  - Determining where additional storage equipment might be needed
  - reduces the need to troubleshoot performance issues in the future

Performance Capacity Use case

Provisioning & Balancing Workloads, Preventing overloading of nodes and aggregates use cases

- Node and Aggregate Inventory page
  - Performance Manager provides
    - hourly samples averaged over the previous 72 hours in node and aggregate inventory page

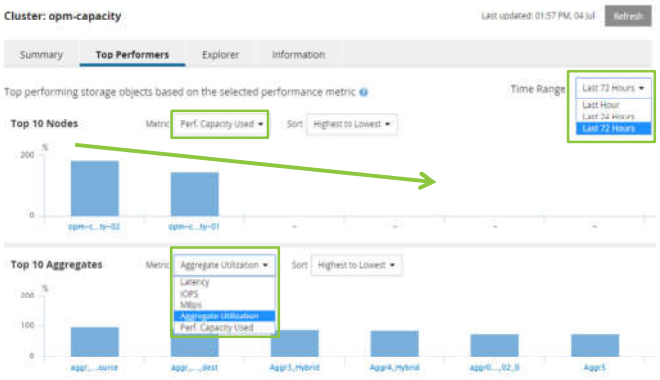


74

Performance Capacity Use case

Provisioning & Balancing Workloads, Preventing overloading of nodes and aggregates use cases

- Performance Manager provides
  - Nodes and aggregates with Top and least performance capacity used metric
    - Maximum performance capacity used value over the selected time range
    - Node 2 with more than 150 % performance capacity used will result in increased latencies on workloads
    - With node 1 being close to 100% optimal performance capacity used
    - Detailed performance charts will inform if node 2 is consistently over-utilized

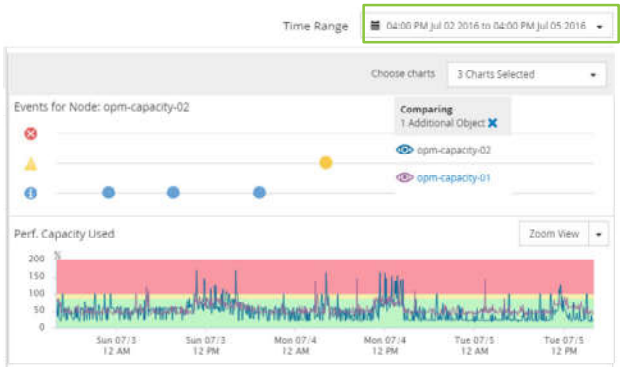


75

Performance Capacity Use case

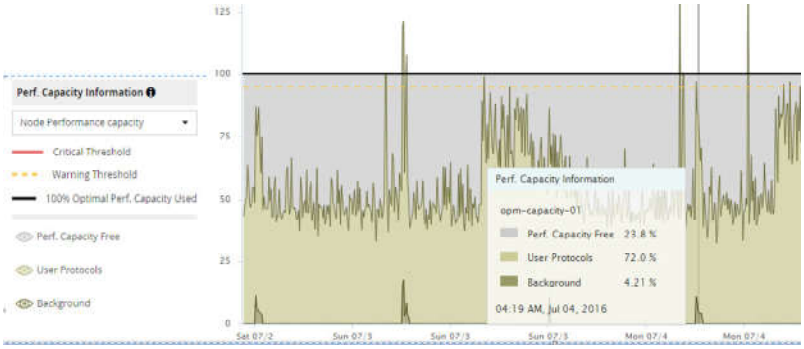
Provisioning & Balancing Workloads, Preventing overloading of nodes use cases

- Performance Capacity Used (%) charts
  - Observe and understand performance capacity used on node for selected time range of week or month
    - Co-relate with performance events for the selected time range
    - Compare and view performance charts with other nodes in the cluster
    - Both nodes operating in safe zone with performance capacity used almost similar on nodes 1 and 2
    - Performance capacity used on both nodes on Sunday has increased
    - Use Performance capacity Advanced charts to understand



Performance Capacity Use case

Provisioning & Balancing Workloads, Preventing overloading of nodes use cases



77

Performance Capacity Use case

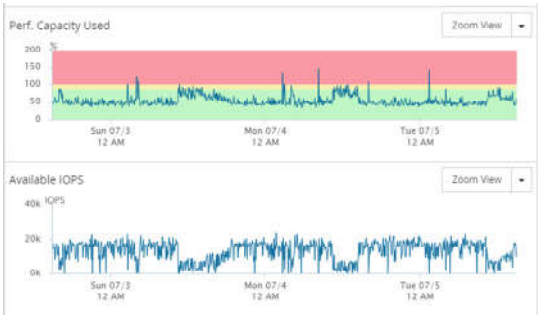
Provisioning & Balancing Workloads, Preventing overloading of nodes use cases

- Performance capacity advanced chart provides breakdown of user protocol workload and background activity
  - Background or internal system processes includes storage efficiency, data replication and system health
  - Performance capacity user defined thresholds are applied to Advanced charts
  - Performance capacity free is also displayed
  - Advanced chart displays background and user protocol breakdown for the node
- From Advanced chart on slide 49, performance capacity used on node is for user protocol workload with minimal background activity
- Further analysis can look at if user protocol workload increasing for Sunday is one time or regular occurrence

Performance Capacity Use case

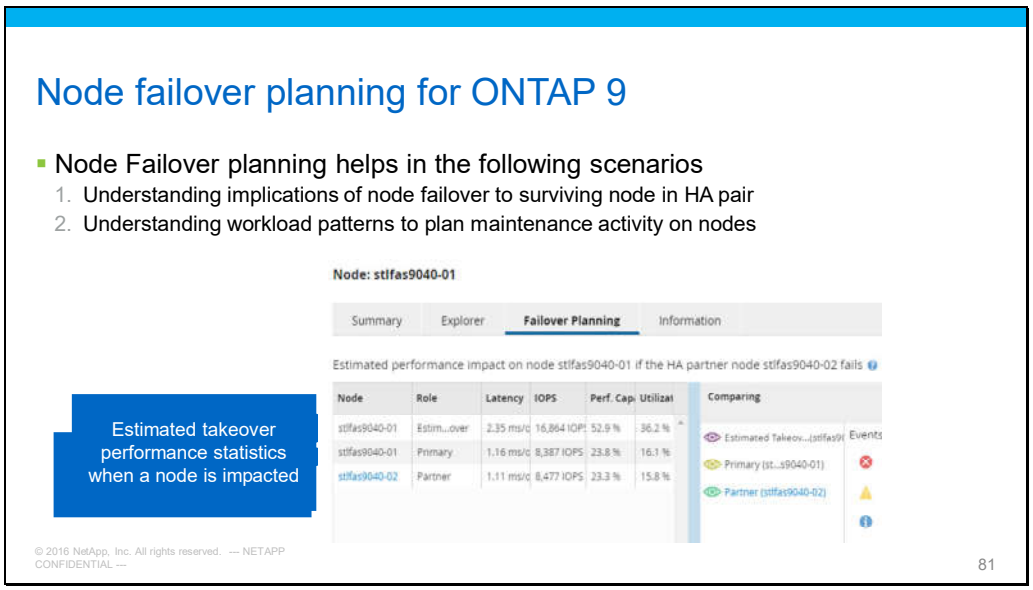
Provisioning & Balancing Workloads, Preventing overloading of nodes use cases

- Available IOPS on nodes and aggregates is an important aspect for the above use cases
  - Indicates IOPS left in node and aggregate to be used
    - New provisioning or increasing load in the current workload
    - Available IOPS is displayed only if user protocol workload is running on node or aggregate



Lesson 6

Performance Manager 7.0 Failover Planning



Performance Manager 7.0 introduces Node Failover Planning feature for ONTAP 9 systems.

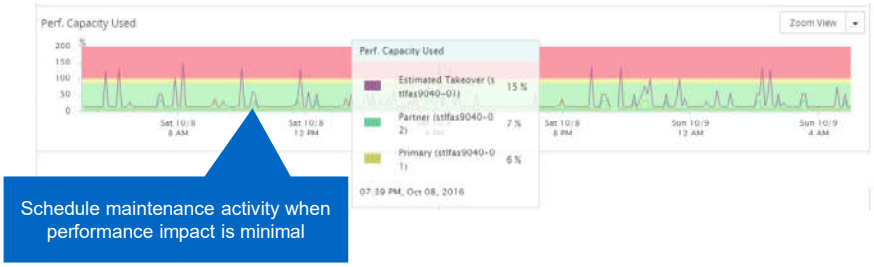
The **Node Failover Planning** page estimates the performance impact on a node if its HA partner node fails. Performance Manager bases the estimates on the historical performance of the nodes in the HA pair.
Estimating the performance impact of a failover helps you plan in the following scenarios:

If a failover consistently degrades the takeover node's estimated performance to an unacceptable level, consider taking corrective actions to reduce the performance impact due to a failover.

Before initiating a manual failover to perform hardware maintenance tasks, you can estimate how the failover affects the performance of the takeover node to determine the best time to perform the task.

Node failover planning for ONTAP 9

- Understand workload patterns to plan maintenance activity on nodes
  - Performance Capacity Used charts of Primary, Partner and Estimated Takeover impact

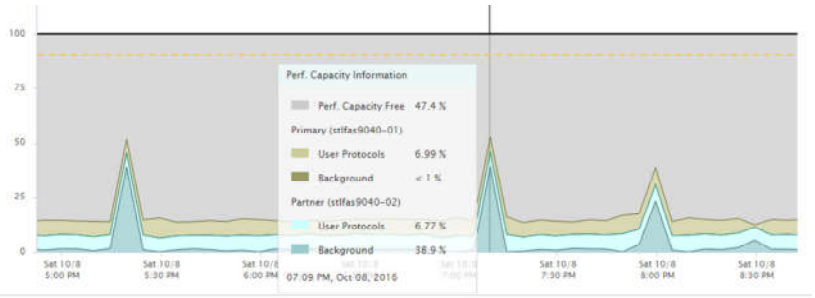


82

Set time range for the last 7 days and Use performance capacity chart to look for a window when maintenance activity can be performed, performance impact should be minimal during maintenance activity. NetApp ensures non-disruptive client operations during maintenance activity.

Node failover planning for ONTAP 9

- Performance Capacity Used advanced charts
  - Understand User protocol workload and background system processes breakdown for both nodes
  - Understand background activity on nodes before scheduling for maintenance activity



83

Performance capacity advanced chart helps understand if performance capacity used on both the nodes are caused by user protocol workload or background system processes. SnapMirror transfers, RAID re-construct, disk scrub and storage efficiency processes are grouped under background system processes.

Performance capacity advanced chart helps understand if there are any system processes running in the background and which could be impacted if a maintenance window is scheduled during this time period.

Node failover planning for ONTAP 9

- Understand the implications of node failover to surviving node in HA pair
  - Failover consistently degrades the takeover node's estimated performance to an unacceptable level
  - Take corrective actions to reduce performance impact due to failover

Example: Non-disruptive volume move or add nodes to cluster

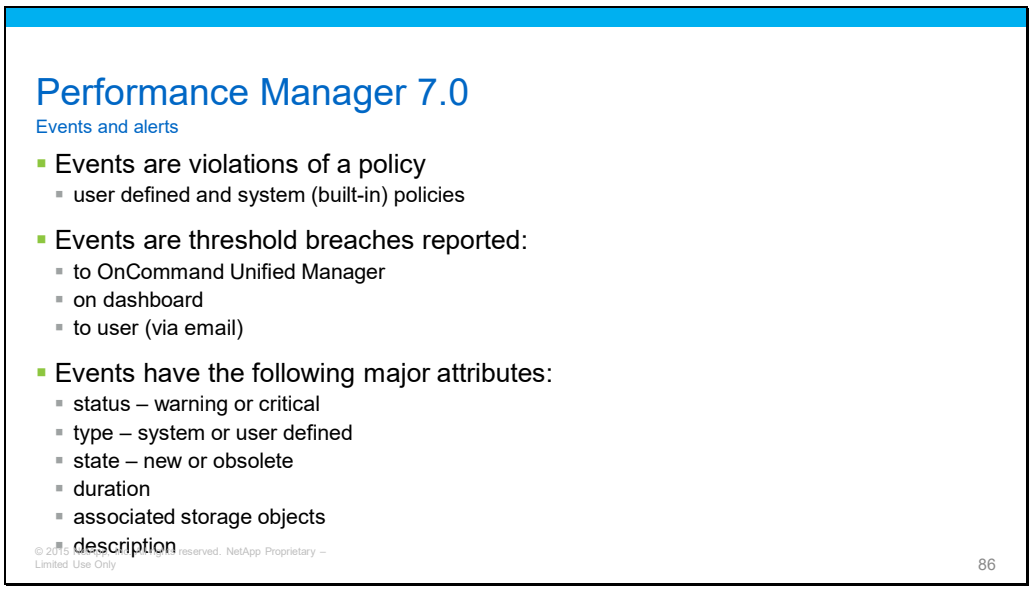
84

If a failover consistently degrades the takeover node's estimated performance to an unacceptable level, consider taking corrective actions to reduce the performance impact due to a failover.

For example, to reduce the estimated performance impact of a failover, you can move some volumes from a node in this HA pair to other nodes in the cluster. Doing so ensures that the primary node can continue to deliver acceptable performance after a failover. Also, if the cluster does not consist of the maximum number of nodes, you can add more nodes to the cluster to lessen the workload on this HA pair.

Lesson 7

Performance Manager 7.0 Events and Thresholds



Policies are established when business agreements are made based on performance requirements of the storage system. These policies are encoded into OPM. OPM constantly analyzes performance data from cDOT storage systems to ensure a policy is not violated. When a violation occurs and event is recorded and communicated to the user immediately.

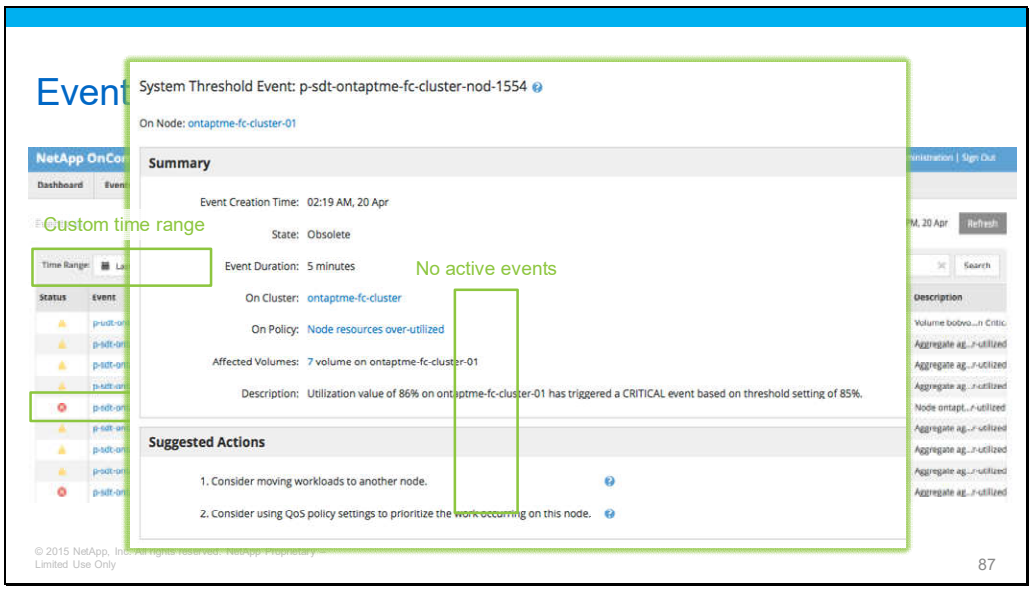
For example, a service level can be contractually agreed upon and closely aligned with a given workload and one or more performance metrics. OPM analyzes performance metrics every 5 minutes for any policy violations.
If the service level agreement is not met then OPM generates an event and an incident is logged.

System defined policies are those automatically implemented by OPM by default. These include policies that monitor node CPU utilization and aggregate disk utilization.

Event attributes:

status – An event status can be **warning** or **critical** depending on the severity. The description provides an explanation when necessary (this will become more clear in some of the examples).
state - An event is **new** when it is still active. When the condition terminates the event becomes **obsolete**.
description – Contains the details of the the event, such as as the severity of the event, the objects involved, and the causes for the event.

The remaining attributes are self explanatory.



The above shows the events landing page. In this example, there are no active events at this time (all indicate obsolete). Events can be filtered to customized time ranges. Hyper-links are provided to every event.

This is an example of a **system defined** event where the node utilization reached 86% exceeding a pre-defined threshold of 85%.

Performance Manager 7.0

User-defined Threshold Policies

- Events are triggered by breaching conditions defined in a threshold policy
- A threshold policy is user created
- A threshold policy is a template
  - must be applied to objects
  - can be applied in bulk
- A threshold policy has the following attributes:
  - object type and counter
  - warning and critical threshold levels
  - length of time the threshold must be breached to trigger an event

88

The Threshold Policy is the primary construct implementing user defined thresholds. The user creates a template that must be applied to one ore more objects. OPM makes it simple to apply policies in bulk to many objects at once. Policies are specific to an object type, supporting early warning and critical level thresholds, and duration that will trigger an alert appearing on the dashboard, OCUM, and email.

Complex policies can be defined for volume and LUN objects when used in combination with latency metrics. This helps minimize false latency alerts when op rate is unusually low.

Performance Manager 7.0

Threshold Policies – Enhancements in this release

- System-defined threshold policies
  - Monitor performance and generate events automatically
  - Enabled by default
- System-defined threshold policies enhanced in this release
  - Node resources over-utilized
  - Node HA pair over-utilized
- User-defined threshold policies
  - User can set threshold policies on storage objects
  - Generate event and send information on issues impacting cluster performance
  - New Performance capacity used (%) performance counters to set threshold policies on storage objects
- Performance Capacity Used counters can be used in combination threshold policies

Performance Manager 7.0

System-defined Node Threshold Policies - Enhancements

- Node resources over-utilized
  - Warning event
  - For nodes installed with ONTAP 8.3.x and earlier
    - Looks for nodes using more than 85% CPU and RAM resources for more than 30 minutes
  - For nodes installed with ONTAP 9
    - Looks for nodes using more than 100% of their performance capacity for more than 30 minutes
- Node HA pair over-utilized
  - Information event
  - For nodes installed with ONTAP 8.3.x and earlier
    - Looks for CPU and RAM usage for the two nodes in the HA pair
    - Looks out for combined utilization exceeding 140% for more than one hour
  - For nodes installed with ONTAP 9
    - Looks at combined performance capacity used value for the two nodes in the HA pair
    - Looks out for combined performance capacity used value exceeding 200% for more than one hour

Threshold Policies

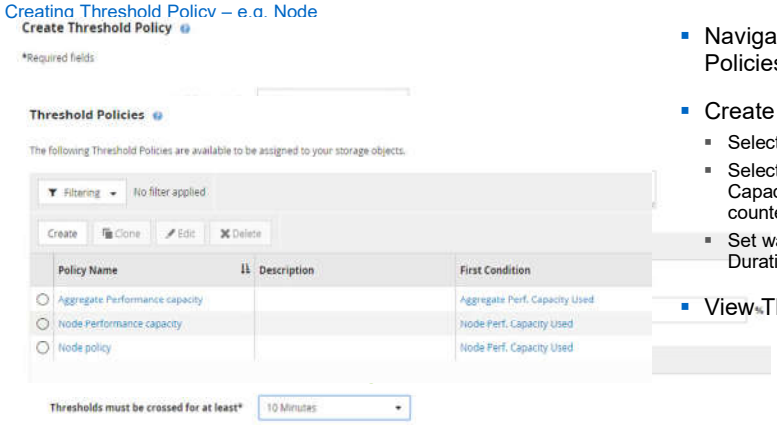
User-defined Threshold Policies – New Performance Counters

- User-defined threshold policies
  - New Performance capacity used (%) performance counters to set threshold policies on storage objects
  - Performance capacity used (%) performance counters can be set on node and aggregate
  - Refer to OPM 7.0 User Guide for Available performance counters to set thresholds

Storage Object	Performance Counter	Description
Node	Performance Capacity Used	Average percentage of performance capacity consumed by the node
	Performance Capacity Used - Takeover	Average percentage of performance capacity consumed by the node and its partner node in the HA pair
Aggregate	Performance Capacity Used	Average percentage of performance capacity consumed by the aggregate

Threshold Policies

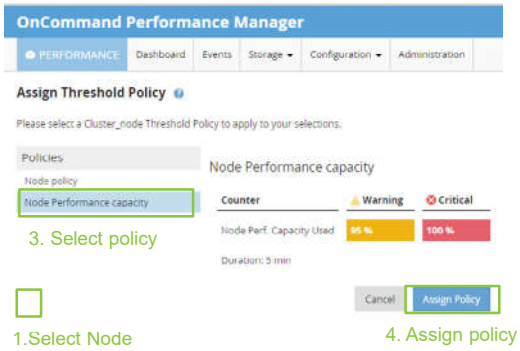
Creating Threshold Policy – e.g. Node



- Navigate to Threshold Policies Page
- Create Threshold Policy
  - Select Node as Object Type
  - Select Performance Capacity used object counter
  - Set warning, critical & Duration values
- View Threshold Policy

Threshold Policies

Assigning Threshold Policy – Node



- Navigate to Node Inventory Page
- Select Node and Assign Threshold Policy

Threshold Policies

Combination Threshold Policies – Enhancements in this release

- Combination threshold policies
  - Only some performance counters can be used together in combination policies
  - Both performance counters must exceed their maximum limits before event generated
  - Performance capacity used counter based combination policies new in this release

Primary storage object and counter	Secondary storage object and counter
Volume: Average Latency	Aggregate: Performance Capacity Used
	Node: Performance Capacity Used
	Node: Performance Capacity Used takeover
LUN: Average Latency	Aggregate: Performance Capacity Used
	Node: Performance Capacity Used
	Node: Performance Capacity Used takeover

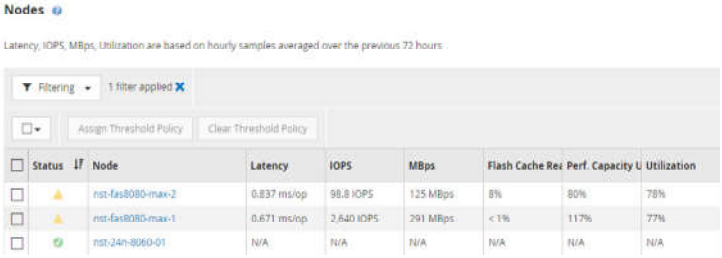
Lesson 8

Performance Manager 7.0 Other Features

Performance Manager 7.0 features

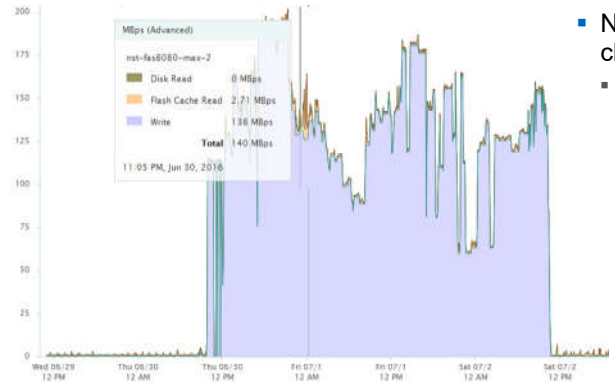
Flash Cache Statistics

- Flash Cache data displayed for nodes
- Data displayed when flash cache module installed in node
- Displays Flash cache read throughput percentage on Node inventory page



Performance Manager 7.0 features

Flash Cache Statistics – Node Explorer Page



- Node Explorer MBps advanced chart
  - Breakdown of Flash cache read throughput and disk read throughput statistics

Performance Manager 7.0 Features

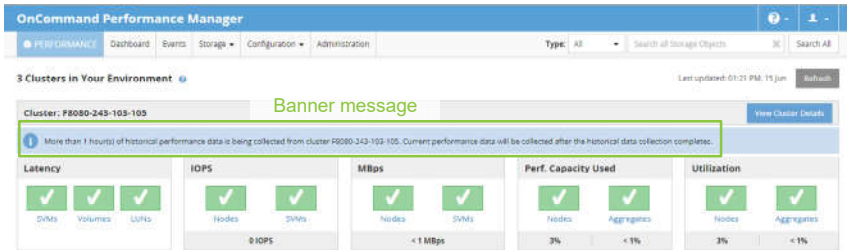
Data Continuity Collection

- Retrieves performance data outside of real-time cluster performance collection cycle
  - Collects a maximum of 24 hours of historical data
- Enables Performance Manager to fill in gaps of statistical data
  - Occur when unable to collect real-time data
- OPM shut down for upgrade to 7.0, enables up to 24 hours of historical data collection
- Supported on ONTAP 8.3.1 and later software versions
- Data continuity collection polls of historical performance data occurs for the following events
  - A cluster in initially added to Performance Manager
  - Current performance data collection cycle does not finish on time
  - Performance Manager inaccessible and then comes online
  - Cluster inaccessible and then comes online

Performance Manager 7.0 Features

Data Continuity Collection

- Data continuity and real-time collection cycle cannot run concurrently
  - Data continuity collection cycle must complete before real-time collection initiated
- Banner message displayed in dashboard if more than one hour of historical data needs to be collected



Performance Manager 7.0 Features

Data Continuity Collection Poll

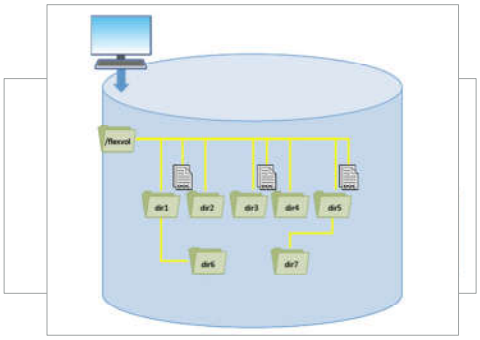
- Cluster initially added to Performance Manager
  - Historical performance data collected for previous 24 hours
  - View a full day of historical performance data for a cluster
  - System-defined threshold events reported from previous 24 hours
- Current performance data collection cycle does not finish on time
  - Data continuity collection cycle initiated to gather missing information
- Performance Manager inaccessible and then comes online
  - Restarted, network outage, shutdown during software upgrade or creating backup file
- Cluster inaccessible and then comes online
  - Network outage, slower WAN delayed performance data collection

Lesson 9

Performance Manager 7.1 Features

FlexGroup introduced in ONTAP 9.1

- Scale-out file system constructed from a series of flexible volumes
- Automatically and transparently work together to share a traffic load
- FlexGroup maximum capacity is 20PB



Performance Manager 7.1 monitors performance of FlexGroup

Steps to take and how it works

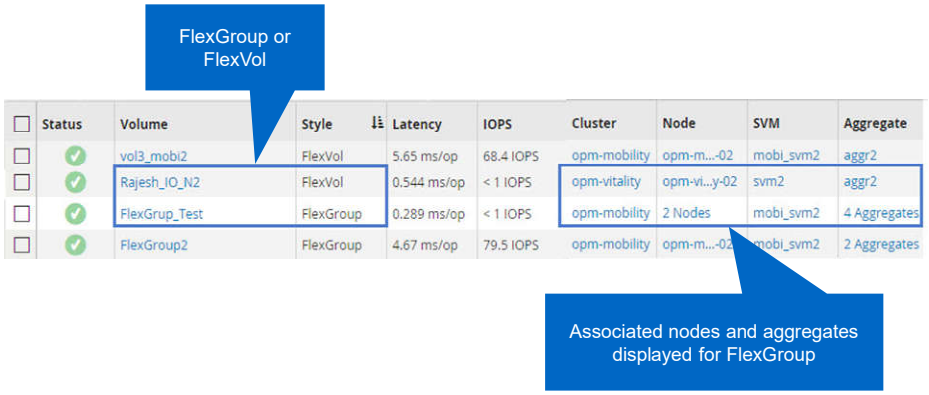
- User creates FlexGroup using FlexGroup Deploy template from CLI or System Manager
- Unified Manager and Performance Manager in Full Integration mode
  - Add cluster to Unified Manager 7.1 or if cluster monitored from Unified Manager 7.1, discovers FlexGroup
- Performance Manager in stand-alone mode
  - Add cluster to Performance Manager 7.1 or if cluster monitored from Performance Manager 7.1, discovers FlexGroup
- Use OPM volume Inventory page to view discovered FlexGroup

Performance Manager 7.1 monitors performance of FlexGroup

Steps to take and how it works

- Load performance explorer page to view the FlexGroup constituents and its related objects performance metrics
- Assign threshold policy to a FlexGroup
  - Use alert and find the specific constituent that contributes the most to the threshold breach

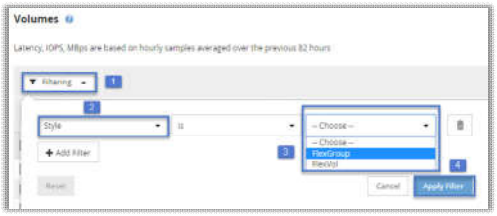
Monitor FlexGroup Volumes



105

Filter volumes by FlexGroup Style

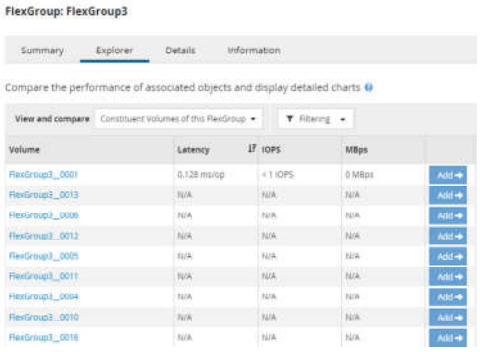
- Filter FlexGroup volume constituents in volume inventory page
- Choose filtering style as FlexGroup



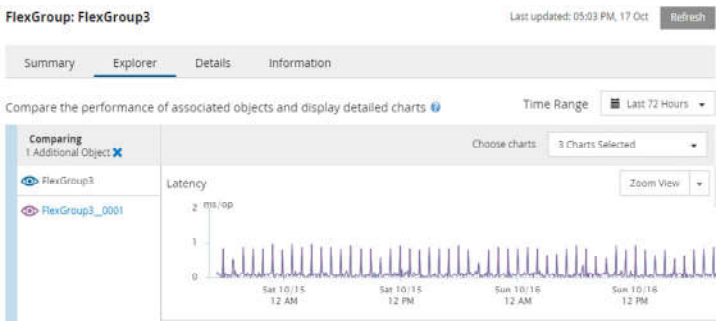
Additional filters available to filter FlexGroup volume constituents in volume inventory page. Choose filtering style as FlexGroup.

Monitor Constituents of FlexGroup volumes

- View FlexGroup constituents from Performance Explorer page
- Compare performance of FlexGroup with associated objects and display detailed charts
- View and Compare with
  - Constituent volumes of FlexGroup
  - SVM of this FlexGroup
  - Aggregates of this FlexGroup

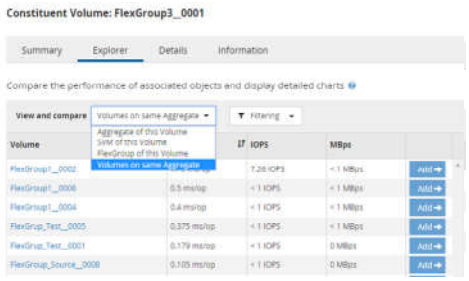


Compare Performance statistics of FlexGroup with constituents



Constituent FlexGroup Volume Explorer Page

- Compare performance of associated objects and display detailed charts
- Constituent FlexGroup volume Explorer Page
- Compare with
  - SVM of this volume
  - FlexGroup of this volume
  - Aggregate of this volume
  - Volumes on same aggregate



Applying threshold policies to FlexGroup

- FlexGroup volume threshold counters that can be applied
  - Average latency
  - Average IOPS
  - Average MBps
  - %Cache miss ratio
- Volume combo threshold policies that can be applied

Primary Object and Counter	Secondary object and counter
FlexGroup Average Latency	FlexGroup Average IOPS
FlexGroup Average Latency	FlexGroup Average MBps

Restrictions when monitoring FlexGroup volumes

- Certain combination threshold policies cannot be applied to FlexGroup volumes
- FlexGroup volumes and FlexGroup constituent volumes considered as victims when dynamic thresholds are identified
  - Event generated for the FlexGroup volume only; not for individual constituent volumes
- When monitoring FlexGroup volumes, no bullies are identified
- When sending performance data to an External data provider server (such as Graphite), FlexGroup and FlexVol volume data is sent
  - No data for individual FlexGroup constituent volumes are sent
- FlexGroup constituent volumes cannot be added to Favorites list
- Refer to Performance Manager 7.1 RC1 Release Notes for detailed information

111

- Certain combination threshold policies cannot be applied to FlexGroup volumes
- FlexGroup volumes and FlexGroup constituent volumes considered as victims when dynamic thresholds are identified
- Event generated for the FlexGroup volume only; not for individual constituent volumes
- When monitoring FlexGroup volumes, no bullies are identified
- When sending performance data to an External data provider server (such as Graphite), FlexGroup and FlexVol volume data is sent
  - No data for individual FlexGroup constituent volumes are sent
- FlexGroup constituent volumes cannot be added to Favorites list

Refer to Performance Manager 7.1 RC1 Release Notes for detailed information

OnCommand Performance Manager

Key Takeaways

- Dashboard - provides a comprehensive view of all storage performance
  - what's important now?
  - what's hot now?
- Monitor - view common storage object performance metrics
  - what's the typical performance profile over the last few months?
  - has resource demand changed significantly?
  - Are my nodes and aggregates optimally utilized ?
  - FlexGroup and its constituent volumes NEW
- Manage – define custom threshold policies for early warning and critical alerting
  - am I meeting business critical service objectives?
- Failover planning
  - What's the implications of node failover to surviving node in HA pair ?
  - What is the right time to plan maintenance activity on nodes ?

Lesson 10

Unified Manager 7.0,Performance Manager 7.0 and later releases Installation and Deployment

Unified Manager 7.1 and Performance Manager 7.1

Important changes in installation on Red Hat Enterprise Linux

- Additional Java Software Compatibility
  - Unified Manager and Performance Manager 7.1 fully supports both the Oracle JDK and the OpenJDK. Support includes any release in the 1.8.0 family, but version1.8.0.101 or later is recommended for security considerations.
- Changes in Third-Party Software Bundle
  - No third-party dependencies bundle is included with Unified Manager 7.1 or Performance Manager 7.1 for a Red Hat Enterprise Linux installation. See the **OnCommand Unified Manager Installation and Setup Guide for Red Hat Enterprise Linux** and **OnCommand Performance Manager Installation and Setup Guide for Red Hat Enterprise Linux** for complete installation and configuration information.

Unified Manager 7.1 and Performance Manager 7.1

Important changes in installation on Red Hat Enterprise Linux

- Relaxed Third-Party dependencies
  - From Unified Manager 7.0 GA and Performance Manager 7.0 GA, we relaxed the third-party dependencies specified in the Linux RHEL package to allow customers to self-patch their Linux systems for security and other minor patches. Applying patches (such as for Java or MySQL) does not now always require patching Unified Manager and Performance Manager first, simplifying ongoing maintenance of Unified Manager and Performance Manager.
- Refer to Unified Manager 7.1 Release Notes and RHEL Install and Setup Guide for more information
- Refer to Performance Manager 7.1 Release Notes and RHEL Install and Setup Guide for more information

Supported browsers and NetApp ONTAP versions

- Always refer to NetApp Interoperability matrix tool for latest information

Browser	Google Chrome version 52, 53	Microsoft Internet Explorer 11	Mozilla Firefox ESR 38 and ESR 45
Unified Manager 7.1	✓	✓	✓
Performance Manager 7.1	✓	✓	✓

NetApp ONTAP version	ONTAP 8.2.x	ONTAP 8.3.2, 8.3.1, 8.3.0	ONTAP 9.0, 9.1RC1
Unified Manager 7.1	✓	✓	✓
Performance Manager 7.1	✓	✓	✓

Supported virtual appliance and operating systems

- Refer to Unified Manager 7.1 and Performance Manager 7.1 Installation and Setup Guide for details

Virtual appliance requirements	VMware ESXi 5.5 and updates	VMware ESXi 6.0, 6.0U1, 6.0U2
Unified Manager 7.1	✓	✓
Performance Manager 7.1	✓	✓

Operating System supported	Red Hat Enterprise Linux version 6.6, 6.7, 6.8, 7.0, 7.1 and 7.2 (64-bit)	Microsoft Windows Server 2008 SE(SP2) / EE(SP2) / R2 SE /R2 EE, 2012 SE /DCE/ R2SE / R2 DCE
Unified Manager 7.1	✓	✓
Performance Manager 7.1	✓	×

NetApp

Supported Virtual appliance and operating systems

- Full Integration of Unified Manager 7.1 and Performance Manager 7.1 does not require that they have to be installed on the same host operating system
- Any combination of supported host operating systems are allowed
- Performance Manager 7.1 installed on Red Hat Enterprise Linux can be integrated with Unified Manager 7.1 installed on Windows

Performance Manager 7.1		Unified Manager 7.1 Deployment platform		
		Windows	RHEL	Virtual Appliance
Deployment platform	Virtual Appliance	✓	✓	✓
	RHEL	✓	✓	✓

NetApp

Integration of the two products does not require that they are installed on the same host operating system. Any combination of host operating systems is allowed. For example, Performance Manager 7.0 installed on Red Hat Enterprise Linux can be integrated with Unified Manager 7.0 installed on Windows.

Hardware configuration requirements

	Hardware Configuration			High Availability Support
	Platform Support	CPU Core(s)	RAM (in GB)	
Unified Manager 7.1	Virtual appliance	4 vCPU	12	Native VMware HA
	Windows	4	12	Microsoft Cluster Service
	RHEL	4	12	Veritas Cluster Server
Performance Manager 7.1	Virtual appliance	4 vCPU	12	Native VMware HA
	RHEL	4	12	Not Available

Module 3

NetApp Harvest

About This Module

This module focuses on the following:

- What NetApp Harvest is and what it can do
- How to get it
- How to get the appliance
- Appliance configurations
- Appliance Default dashboards

Lesson 1

What it is

What is NetApp Harvest

- Flexible Data collector for 7-mode, Clustered ONTAP, and OCUM 6.x
- Near Real time (60 seconds by default) monitoring

	OPM 2.0	NetApp Harvest
Polling interval	every 5 minutes, cannot be modified	default every 1 minute, can be modified
Counter inventory	3 presets to choose from, most detailed has ~115 counters, cannot be modified	Flexible, default has ~500 counters, can be modified
Dashboards provided	No	19 Grafana dashboards, 500+ graph and singlestat panels
Data ONTAP 7-mode	No	Yes
Storage capacity information	No	Yes, for cDOT via OCUM 6.x
Scaling	OPM scaling constraints	cDOT scaling constraints
Ease of use	Configure via menu from OPM console; no separate installation required	Configure via text file; separate installation required
Downloadable from	NetApp Support Site	NetApp Support ToolChest Site
Supported by	NetApp Support Center	Written in Perl with Community Support

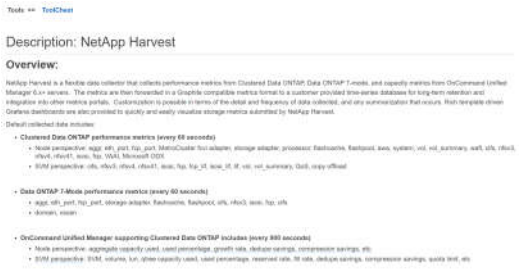
123

Lesson 2

Where to get it

Where to get Harvest

- NetApp Support site in the toolchest
- Also need Graphite and Grafana
- Setup and configuration is difficult


Description: NetApp Harvest

Overview:

http://mysupport.netapp.com/tools/info/ECMLP2314554I.html?productID=61924

Where to get the Appliance

- The Pre-Built appliance is available for download
- Easily deployed .ova
- After its deployed just download harvest and the NMSDK from NetApp and add to the ova.
- Setup directions available on download site
- Average deployment time is about 5 minutes

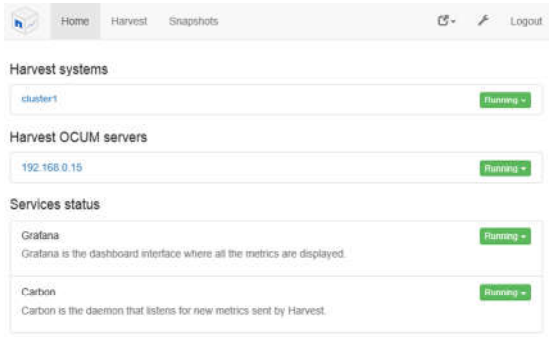
https://nabox.tynsoe.org/downloads/

Lesson 3

Appliance Configuration

Appliance Administration

- Accessible by going to https://*nabox*/admin and logging in



128

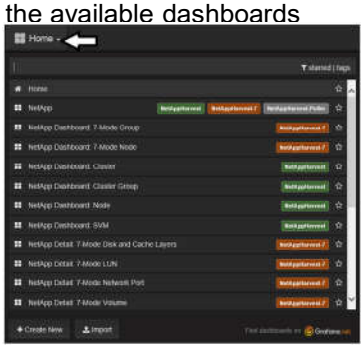
The Harvest tab allows you to add storage systems and Unified Manager systems. You can also modify systems and change the collection interval from the default of 60 seconds.

Lesson 4

Appliance Default Dashboards

Grafana home page

- All data is stored in Grafite, and displayed in Grafana. To get to grafana browse to `https://nabox`
- Click the Home drop down to display all of the available dashboards
- Green is Cluster mode
- Orange is 7-mode



cDOT Dashboard Map

- Cluster Group - Overall capacity and performance across clusters; cluster views
- Cluster - Overall cluster capacity and performance; node and SVM views
- Node - Individual node capacity and performance; node view
- SVM - SVM capacity and performance; SVM view
- Headroom - Headroom Analysis; Node and Aggregate view (EXPERIMENTAL)

NetApp cDOT Dashboard Map

The table below provides a map to available dashboards depending on your area of interest

Dashboard Name	Description	When to Use
Cluster Group	Overall capacity and performance across clusters; cluster views	Analyze workload balance across many clusters, identify top resources in use, identify top SVM workloads, analyze capacity and performance trends, etc.
Cluster	Overall cluster capacity and performance; node and SVM views	Analyze workload balance across nodes in the cluster, identify top node resources in use, identify top SVM workloads, analyze capacity and performance trends, etc.
Node	Individual node capacity and performance; node view	Analyze node performance and resource utilization, analyze top resources within a node, aggr capacity utilization, etc.
SVM	SVM capacity and performance; SVM view	Useful to view SVM performance, identify busiest volumes, capacity utilization, etc.
Headroom	Headroom Analysis; Node and Aggregate view (EXPERIMENTAL)	Useful to view headroom of node and aggregate resources. Requires ONTAP 9 or later.

NetApp Dashboard: Cluster Group

NetApp Dashboard: Cluster

NetApp Dashboard: Node

NetApp Dashboard: SVM

cDOT Detail Map

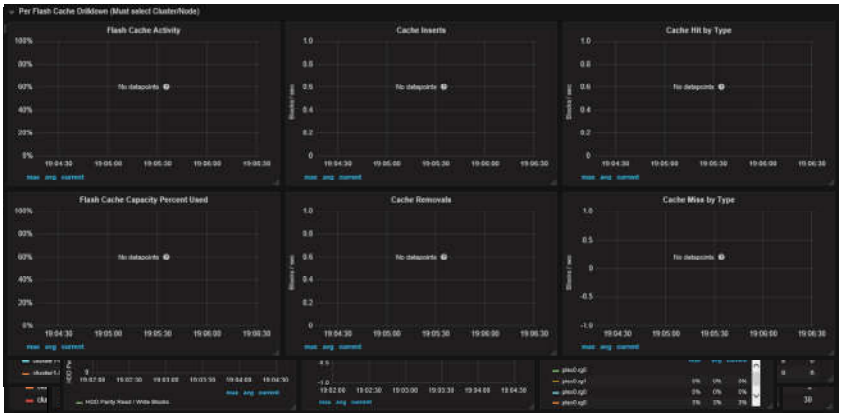
- Disk and Cache Layers - Top disk aggregates, disk IO behavior, aggregate capacity, Flash Pool, and Flash Cache; node and aggregate views
- Network Port - Top Ports with send/receive data and link utilization for Ethernet, send/receive data for FCP; Network Port views
- Network LIF - Top LIFs with send/receive data, IOPs and latency for SAN LIFs; SVM LIF views

NetApp cDOT Detail Map

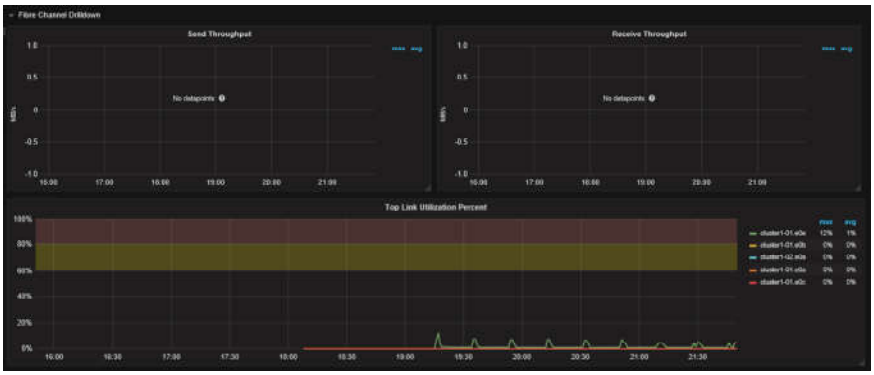
The table below provides a map to available detail views depending on your area of interest

Dashboard Name	Description	When to Use
Disk and Cache Layers	Top disk aggregates, disk IO behavior, aggregate capacity, Flash Pool, and Flash Cache; node and aggregate views	Analyze disk and caching layer performance and capacity, efficiency from Flash Cache and Flash Pool, Flash Pool sizing, troubleshoot data layout issue, etc.
Network Port	Top Ports with send/receive data and link utilization for Ethernet, send/receive data for FCP; Network Port views	Analyze for busiest Ports, trending of usage, etc
Network LIF	Top LIFs with send/receive data, IOPs and latency for SAN LIFs; SVM LIF views	Analyze for busiest LIFs, SAN LIFs with high latency, trending of LIF usage, etc
Volume	Top FlexVols with throughput, IOP, and latency. Latency breakdown within cluster. Reads from cache/disk, capacity view, etc; Volume views	Analyze for busiest volumes, view detailed latency breakdown for a volume, trend usage, etc
Lun	Top LUNs with throughput, IOP, and latency. vStorage offload, IOP sizes, etc. Capacity view including LUN fill rate; LUN and Volume views	Analyze for busiest LUNs, view details of LUN performance, latency breakdown for parent volume, trend usage, etc
MetroCluster	View of MetroCluster specific resources like FCVI (NVRAM) Mirroring, views of plexes	Analyze performance and latency of MetroCluster mirroring

NetApp Detail: Disk and Cache Layers



NetApp Detail: Network Port



NetApp Detail: Network LIF

cDOT Detail Map - Continued

- Volume - Top FlexVols with throughput, IOP, and latency. Latency breakdown within cluster. Reads from cache/disk, capacity view, etc; Volume views
- LUN - Top LUNs with throughput, IOP, and latency. vStorage offload, IOP sizes, etc. Capacity view including LUN fill rate; LUN and Volume views
- MetroCluster - View of MetroCluster specific resources like FCVI (NVRAM) Mirroring, views of plexes

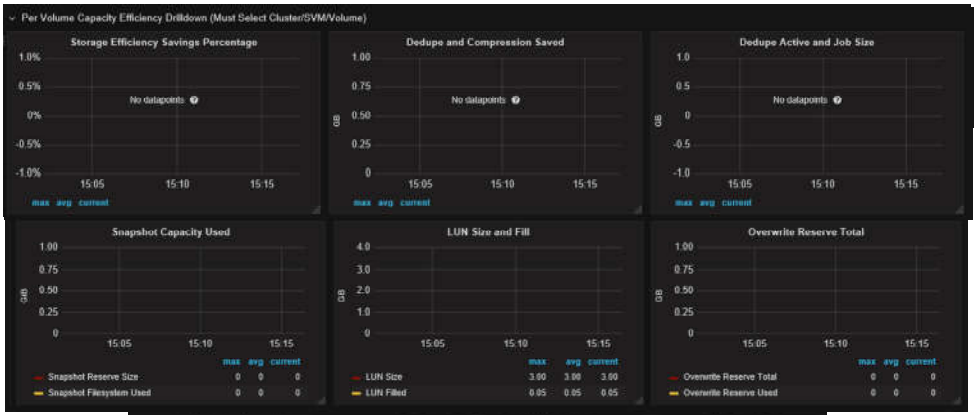
NetApp cDOT Detail Map

The table below provides a map to available detail views depending on your area of interest

Dashboard Name	Description	When to Use
Disk and Cache Layers	Top disk aggregates, disk IO behavior, aggregate capacity, Flash Pool, and Flash Cache; node and aggregate views	Analyze disk and caching layer performance and capacity, efficiency from Flash Cache and Flash Pool, Flash Pool sizing, troubleshoot data layout issue, etc.
Network Port	Top Ports with send/receive data and link utilization for Ethernet, send/receive data for FCP; Network Port views	Analyze for busiest Ports, trending of usage, etc
Network LIF	Top LIFs with send/receive data, IOPs and latency for SAN LIFs; SVM LIF views	Analyze for busiest LIFs, SAN LIFs with high latency, trending of LIF usage, etc
Volume	Top FlexVols with throughput, IOP, and latency. Latency breakdown within cluster. Reads from cache/disk, capacity view, etc; Volume views	Analyze for busiest volumes, view detailed latency breakdown for a volume, trend usage, etc
Lun	Top LUNs with throughput, IOP, and latency. vStorage offload, IOP sizes, etc. Capacity view including LUN fill rate; LUN and Volume views	Analyze for busiest LUNs, view details of LUN performance, latency breakdown for parent volume, trend usage, etc
MetroCluster	View of MetroCluster specific resources like FCVI (NVRAM) Mirroring, views of plexes	Analyze performance and latency of MetroCluster mirroring

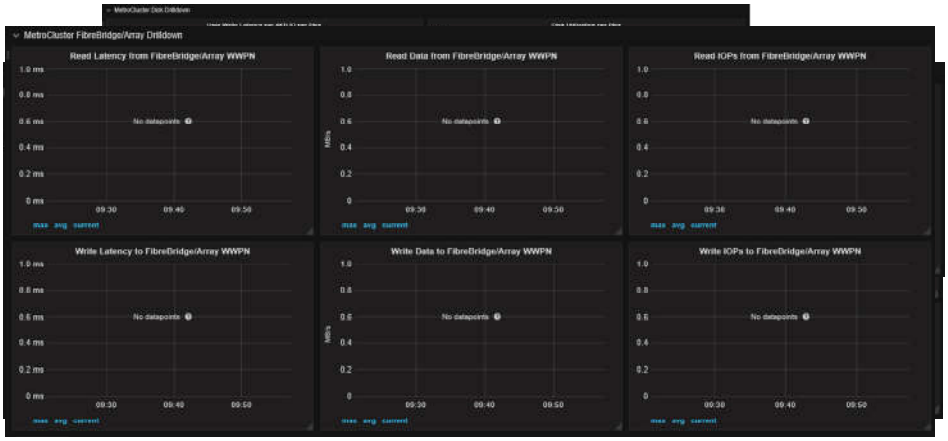
NetApp Detail: Volume

NetApp Detail: LUN



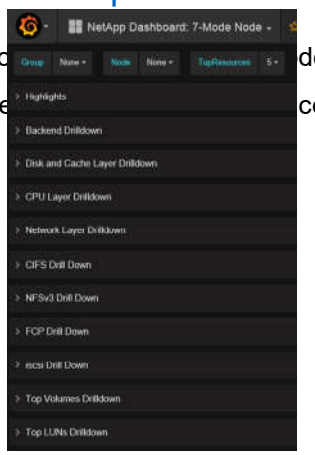
142

NetApp Detail: MetroCluster



7-Mode Dashboard Map

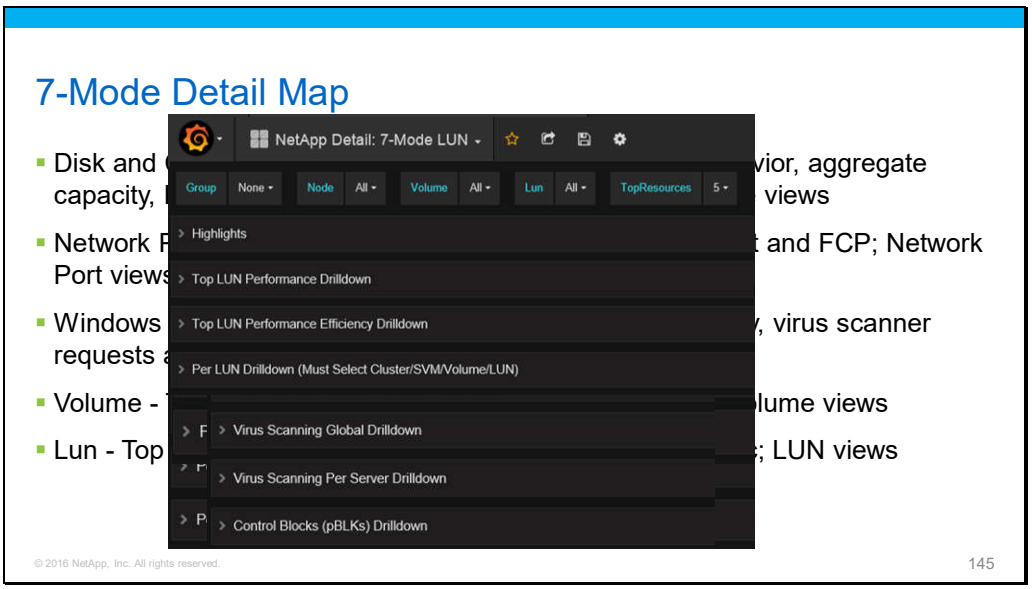
- Node Group - Group c… des view
- Node - Individual node… ce; node view



NetApp 7-Mode Dashboard Map

The table below provides a map to available dashboards depending on your area of interest

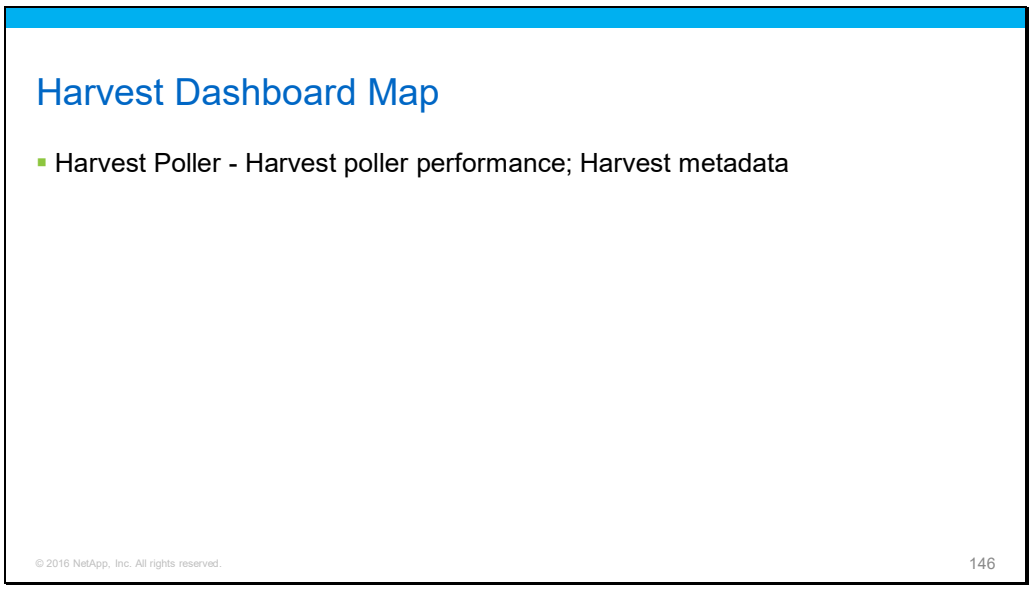
Dashboard Name	Description	When to Use
Node Group	Group of nodes performance; nodes view	Analyze nodes performance and resource utilization, analyze top resources across a set of nodes, for example an HA-Pair or all nodes in a group.
Node	Individual node capacity and performance; node view	Analyze node performance and resource utilization, analyze top resources within a node, aggr capacity utilization, etc.



NetApp 7-Mode Detail Map

The table below provides a map to available detail views depending on your area of interest

Dashboard Name	Description	When to Use
Disk and Cache Layers	Top disk aggregates, disk IO behavior, aggregate capacity, Flash Pool, and Flash Cache; node and aggregate views	Analyze disk and caching layer performance and capacity, efficiency from Flash Cache and Flash Pool, Flash Pool sizing, troubleshoot data layout issue, etc.
Network Port	Top Ports with send/receive data for Ethernet and FCP; Network Port views	Analyze for busiest Ports, trending of usage, etc
Windows File Services	Top windows authentication latency, virus scanner requests and latency, pBlks; Node and vFiler views	Analyze supporting infrastructure performance and utilization such as Windows AD and Virus scanning servers
Volume	Top FlexVols with throughput, IOP, and latency; Volume views	Analyze for busiest volumes, trend usage, etc
Lun	Top LUNs with throughput, IOP, latency, IOP sizes, etc; LUN views	Analyze for busiest LUNs, view details of LUN performance, trend usage, etc

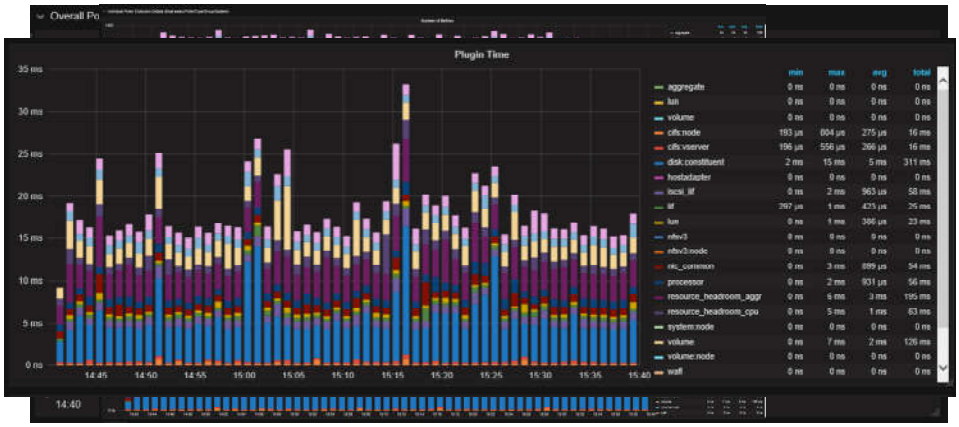


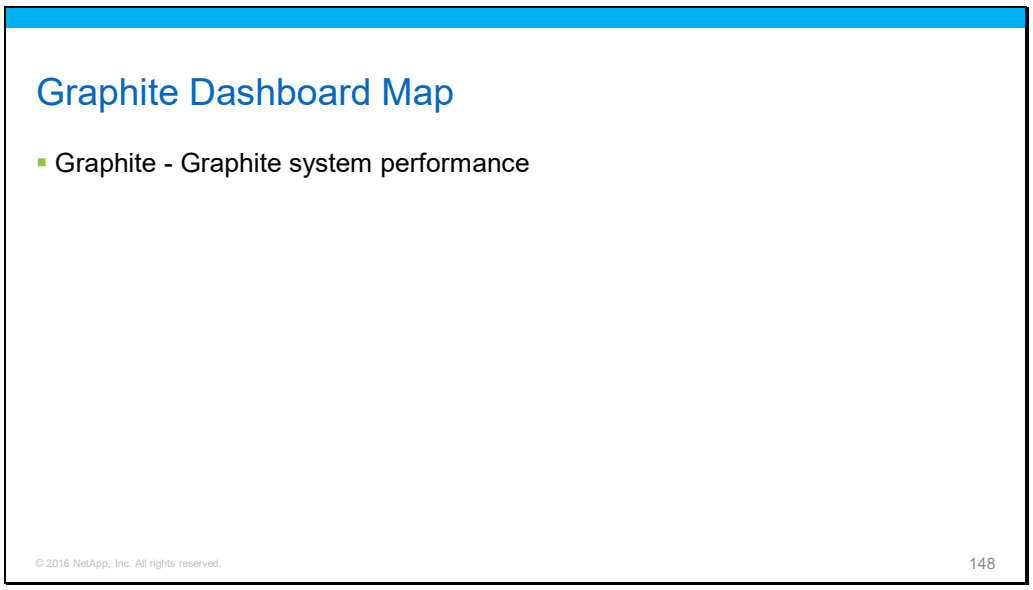
NetApp Harvest Dashboard Map

The table below provides a map to available dashboards depending on your area of interest

Dashboard Name	Description	When to Use
Harvest Poller	Harvest poller performance; Harvest metadata	Useful to identify object types that have long API response times, view which object types or monitored systems submit the most metrics, etc

NetApp Detail: Harvest Poller





Graphite Dashboard Map

The table below provides a map to available dashboards depending on your area of interest

Dashboard Name	Description	When to Use
Graphite	Graphite system performance	Useful to show Graphite process activity, performance, verify metrics are arriving, etc

NetApp Detail: Graphite Server

Summary

Things covered in this chapter

- System Manager Cluster Dashboards
- System Manager Predictable Performance
- OCUM/OPM Simplified operations management
- Unified Manager 7.0 & 7.1 features NEW
- Performance Manager 7.0 Product Overview
- Performance Manager 7.0 & 7.1 features NEW
- OCUM/OPM Installation and Deployment
- Harvest

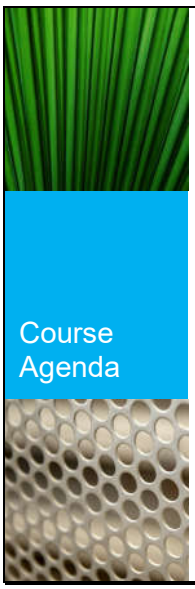
Exercise

- This exercise should take about 4 hours

Chapter 10

Basic Monitoring and Preventative maintenance

NetApp



Course Agenda

- Setting up basic monitoring
- Identifying the most active instance of an object in a cluster
- Identify and resolve performance issues
- Aggregate and Volume optimizations

Module 1

Setting up Basic Monitoring

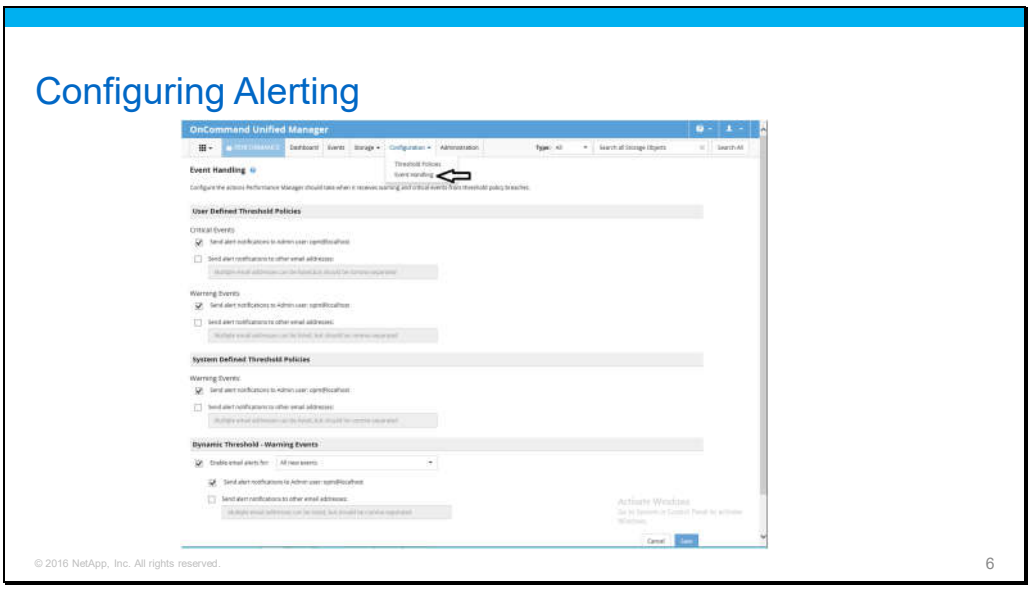
About This Module

This module focuses on the following:

- How to configure alerting
- How to configure basic monitoring

Lesson 1

How to configure alerting



Configuring alert settings

You can specify which events from Performance Manager trigger alerts, the email recipients for those alerts, and whether the events should be reported to Unified Manager.

Before you begin

You must have the OnCommand Administrator role.

About this task

You can configure unique alert settings for the following types of performance events:
- Critical events triggered by breaches of user-defined thresholds
- Warning events triggered by breaches of user-defined thresholds, system-defined thresholds, or dynamic thresholds

By default, email alerts are sent to Performance Manager Admin users for all new events. You can have email alerts sent to other users by adding those users' email addresses.
You can choose to send the alerts to Unified Manager as Critical, Error, Warning, or Information events. If you have configured Unified Manager to send alert emails when it receives performance events, the email recipients might receive notifications from both Performance Manager and Unified Manager.

Note: To disable alerts from being sent for certain types of events, clear all of the check boxes in an event category. This action does not stop events from appearing in the Performance Manager user interface.

Steps

- From the Performance Manager navigation bar, select Configuration > Event Handling. The Event Handling page is displayed.

- In the Event Handling page, configure the appropriate settings for each of the event types. To have email alerts sent to multiple users, enter a comma between each email address.
- Click Save.

Lesson 2

Basic Monitoring

Setting up Monitoring

- Set up Basic Monitoring Tasks
- Perform Daily monitoring
- Check for daily, weekly and monthly performance trends
- Create thresholds to alert you to problems

Example

You want to prevent your Microsoft Exchange Server from crashing due to average volume latency exceeding 20 milliseconds. The following example displays the Warning threshold set to 12 milliseconds and the Critical threshold to 15 milliseconds.

Setting up basic monitoring tasks
You can monitor your systems for performance issues by checking the systems daily, thereby establishing weekly and monthly performance trends. You can also create thresholds to receive notifications about potential performance issues to prevent critical performance issues.

Performing daily monitoring
You can perform daily monitoring to ensure that you do not have any immediate performance issues that require attention.

Steps
- From the Performance Manager UI, go to the Event Inventory page and view all current and obsolete events.
- Click on the new Critical or Warning events and determine what action is required.

Using weekly and monthly performance trends to identify performance issues
Identifying performance trends can assist you in identifying whether the cluster is being overused or underused by analyzing volume latency. You can use similar steps to identify CPU, network, or other system bottlenecks.

Steps
1. Locate the volume you suspect is being underused or overused.
2. On the **Details** tab, click **30 days** to display the historical data.
3. In the "Break down data by" drop-down menu, select **Latency** and click **Submit**.
4. Deselect Aggregate in the Compare the Cluster Components chart and compare with the Latency chart.
5. Select Aggregate and deselect all other components in the Compare the Cluster Components chart and compare with the Latency chart.
6. Compare the reads/writes latency chart to the Latency chart.
7. Determine if client application loads have caused a workload contention and rebalance workloads as needed.
8. Determine if the aggregate is overused and causing contention and rebalance workloads as needed.

Preventing performance issues
You can set user-defined thresholds to prevent performance issues from being critical. For example, if you have a Microsoft Exchange Server and you know that it will crash if volume latency goes above 20 milliseconds, you can set warning and critical thresholds to keep the server from crashing.

Steps
1. Create the Warning and Critical event thresholds:
   a. Select **Configuration > Threshold Policies**.
   b. Click **Create**.
   c. Select the object type and specify a name and description of the policy.
   d. Select the object counter condition and specify the limit values that define Warning and Critical events.
   e. Select the duration of time that the limit values must be breached for an event to be sent and click **Save**.
2. Assign the threshold policy to the storage object.
   a. Go to the Inventory page for the same cluster object type that you previously selected.
   b. Select the object to which you want to assign the threshold policy and click **Assign Threshold Policy**.
   c. Select the policy you previously created and click **Assign Policy**.

Identify most active instance of an object using CLI

- `statistics top client show`

```
cluster1::> statistics top client show

cluster1 : 3/19/2017 00:53:29
*Estimated
     Total
      IOPS Protocol        Node Vserver Client
---------- -------- ----------- ------- -------------
       116     cifs cluster1-01     vs1 192.168.0.11
        26      nfs cluster1-01     vs1 192.168.0.10
```

- `statistics top file show`

```
cluster1::> statistics top file show

cluster1 : 3/19/2017 00:56:53
*Estimated
     Total
      IOPS          Node Vserver    Volume File
---------- ----------- ------- --------- ---------
        89 cluster1-01     vs2 vs2_vol01 /iobw.tst
        86 cluster1-01     vs1 vs1_vol01 /iobw.tst
        34 cluster1-01     vs1 vs1_vol02 /iobw.tst
        17 cluster1-01     vs1 vs1_vol02 /300mfile
        13 cluster1-01     vs2 vs2_vol02 /iobw.tst
        10 cluster1-01     vs2 vs2_vol02 /300mfile
```

- Use the `-sort-key` to sort by the type of desired metric, use `-interval` to change the sampling default from 30 seconds.
  - read_ops, write_ops, other_ops, total_ops, read, write, total are a few different metrics

Module 2

Identifying and resolfing Performance Issues

About This Module

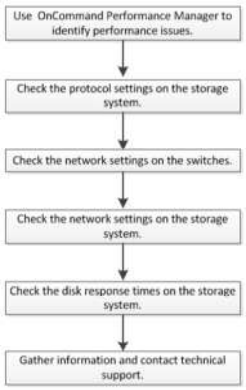
This module focuses on the following:

- Identifying remaining performance capacity
- Using OnCommand Performance Manager to identify performance issues
- Measuring latency and throughput between nodes
- Storage QoS workflow
- Where to find additional information

Lesson 1

Performance Workflow

Performance Workflow



Lesson 2

Identifying remaining performance capacity

Node CPU Headroom

- `set -privilege diag`
- `statistics start -object resource_headroom_cpu`
- `statistics show -object resource_headroom_cpu`

```
sti2520-2131454963690::*> stat show -obj resource_headroom_cpu -raw -counter ewma_hourly
  (statistics show)

Object: resource_headroom_cpu
Instance: CPU_sti2520-213
Start-time: 2/9/2016 16:06:27
End-time: 2/9/2016 16:06:27
Scope: sti2520-213

    Counter                                     Value
    -------------------------------- --------------------------------
    ewma_hourly                                 -
                                ops                              4376
                            latency                             37719
                        utilization                                86
                 optimal_point_ops                               2573
             optimal_point_latency                               3589
         optimal_point_utilization                                 72
    optimal_point_confidence_factor                                 1

Object: resource_headroom_cpu
Instance: CPU_sti2520-214
Start-time: 2/9/2016 16:06:27
End-time: 2/9/2016 16:06:27
Scope: sti2520-214

    Counter                                     Value
    -------------------------------- --------------------------------
    ewma_hourly                                 -
                                ops                                 0
                            latency                                 0
                        utilization                                 0
                 optimal_point_ops                                  0
             optimal_point_latency                                  0
         optimal_point_utilization                                 71
    optimal_point_confidence_factor                                 1
2 entries were displayed.
```

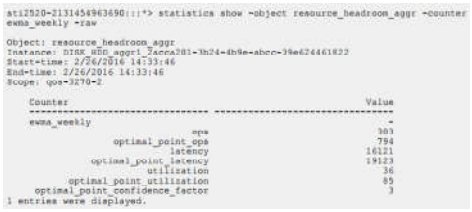

You compute the available performance capacity by subtracting the optimal_point_counter from the current_counter. In this example, the utilization capacity for CPU_sti2520-213 is -14% (72%-86%). This suggests that the node's CPU has been overutilized on average for the past one hour.

Additionally, you could have specified ewma_daily, ewma_weekly, or ewma_monthly to get the same information, but averaged over a longer period of time.

Note: The resource_headroom_cpu Counter Manager (CM) object in this example represents the entire node (all CPUs collectively).

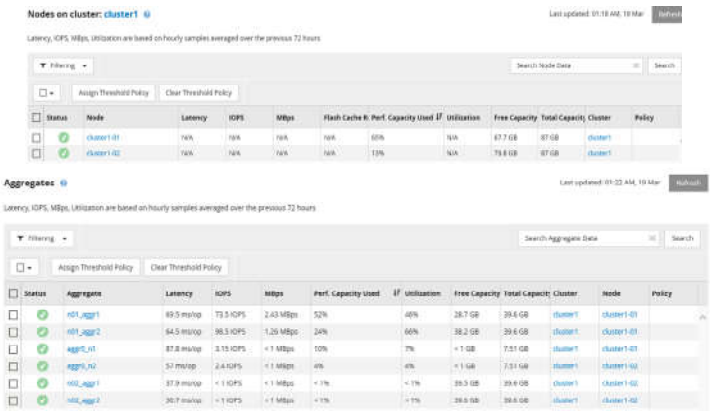
Aggr Headroom

- `set -privilege advanced`
- `statistics start -object resource_headroom_aggr`
- `statistics show -object resource_headroom`



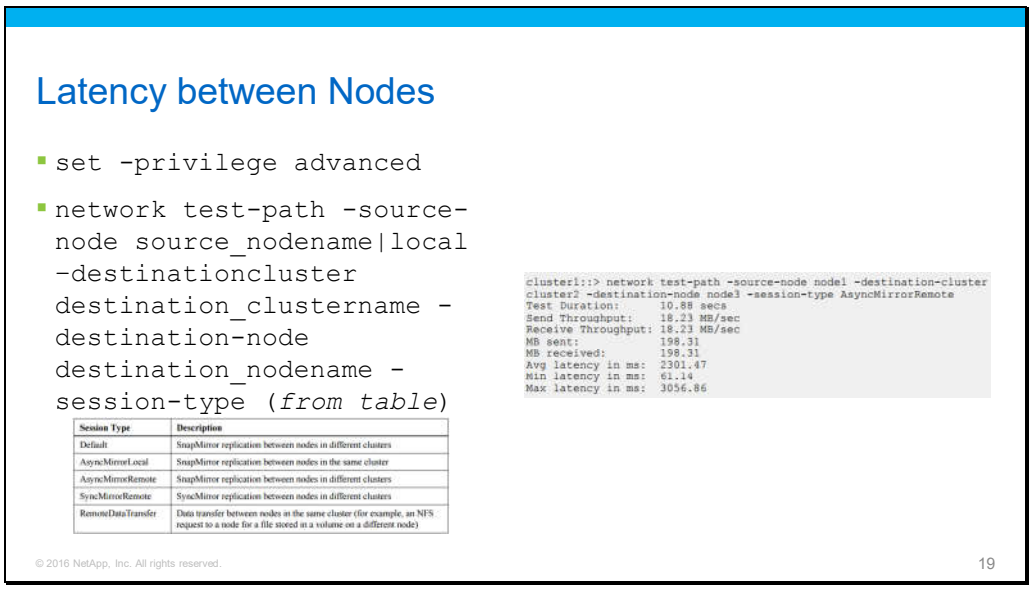
```
sti2520-2131454963690::*> statistics show -object resource_headroom_aggr -counter
ewma_weekly -raw

Object: resource_headroom_aggr
Instance: DISK_HDD_aggr1_7acca281-1b24-4b9e-abcc-39e624461822
Start-time: 2/26/2016 14:33:46
End-time: 2/26/2016 14:33:46
Scope: qos-3270-2

    Counter                                                     Value
    -------------------------------- --------------------------------
    ewma_weekly                                                     -
                                 ops                              303
                 optimal_point_ops                                794
                             latency                            16121
             optimal_point_latency                              19123
                         utilization                               36
          optimal_point_utilization                                85
    optimal_point_confidence_factor                                 3
1 entries were displayed.
```

Same information from OPM



Lesson 3

Network Testing



You can use the network test-path command to identify network bottlenecks, or to prequalify network paths between nodes. You can run the command between intercluster nodes or intracluster nodes.

Before you begin

• You must be a cluster administrator to perform this task.
• Advanced privilege level commands are required for this task.
• For an intercluster path, the source and destination clusters must be peered.

About this task

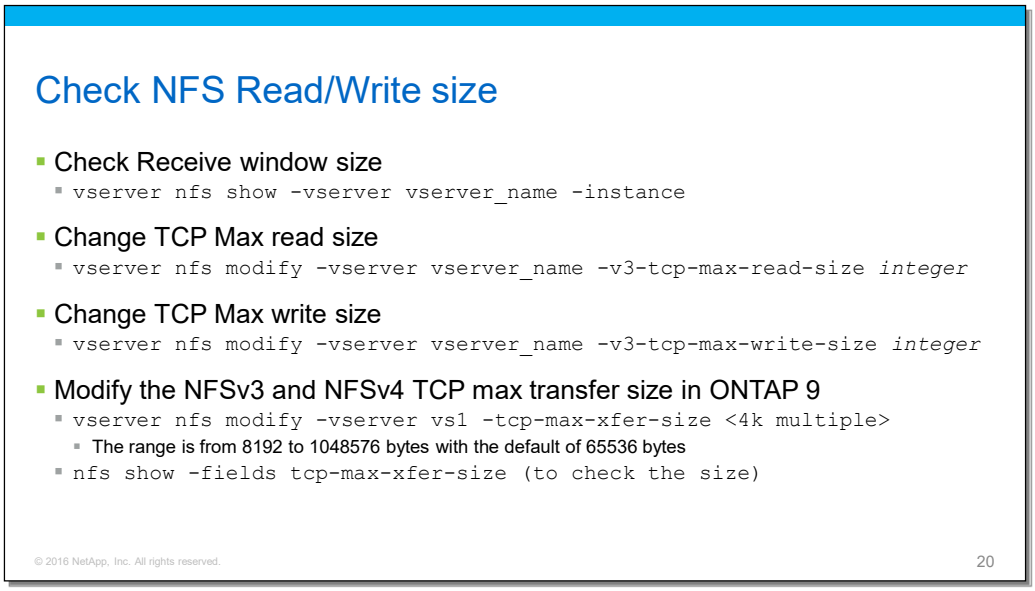
Occasionally, network performance between nodes may not meet expectations for your path configuration. A 1 Gbps transmission rate for the kind of large data transfers seen in SnapMirror replication operations, for example, would not be consistent with a 10 GbE link between the source and destination clusters.

You can use the network test-path command to measure latency and throughput between nodes. You can run the command between intercluster nodes or intracluster nodes.

Note: The test saturates the network path with data, so you should run the command when the system is not busy and when network traffic between nodes is not excessive. The test times out after ten seconds. The command can be run only between ONTAP 9 nodes.

The session-type option identifies the type of operation you are running over the network path— for example, “AsyncMirrorRemote” for SnapMirror replication to a remote destination. The type dictates the amount of data used in the test.

If performance does not meet expectations for the path configuration, you should check node performance statistics, use available tools to isolate the problem in the network, check switch settings, and so forth.



For NFS, you can check the TCP read/write size to determine if the size setting is creating a performance issue. If the size is the source of an issue, you can correct it.

ou might be able to improve the performance of NFSv3 clients connecting to storage systems over a high-latency network by modifying the TCP maximum read and write size.

When clients access storage systems over a high-latency network, such as a wide area network (WAN) or metro area network (MAN) with a latency over 10 milliseconds, you might be able to improve the connection performance by modifying the TCP maximum read and write size. Clients accessing storage systems in a low-latency network, such as a local area network (LAN), can expect little to no benefit from modifying these parameters. If the throughput improvement does not outweigh the latency impact, you should not use these parameters.

To determine whether your storage environment would benefit from modifying these parameters, you should first conduct a comprehensive performance evaluation of a poorly performing NFS client. Review whether the low performance is due to excessive round trip latency and small request on the client. Under these conditions, the client and server cannot fully use the available bandwidth because they spend the majority of their duty cycles waiting for small requests and responses to be transmitted over the connection.

By increasing the NFSv3 request size, the client and server can use the available bandwidth more effectively to move more data per unit time, therefore increasing the overall efficiency of the connection.

Keep in mind that the configuration between the storage system and the client might vary. If you configure the storage system to support 1 MB maximum read size but the client only supports 64 KB, then the mount read size is limited to 64 KB or less.

Before modifying these parameters, you must be aware that it results in additional memory consumption on the storage system for the period of time necessary to assemble and transmit a large response. The more high-latency connections to the storage system, the higher the additional memory consumption. Storage systems with high memory capacity might experience very little effect from this change. Storage systems with low memory capacity might experience noticeable performance degradation.

The successful use of these parameter relies on the ability to retrieve data from multiple nodes of a cluster. The inherent latency of the cluster network might increase the overall latency of the response. Overall latency tends to increase when using these parameters. As a result, latency sensitive workloads might show negative impact.

Beginning in ONTAP 9, the v3-tcp-max-read-size and v3-tcp-max-write-size options are obsolete. You must use the -tcp-max-xfer-size option instead.

You can modify the -tcp-max-xfer-size option to configure maximum transfer sizes for all TCP connections using the NFSv3 and NFSv4.x protocols.

You can modify these options individually for each Storage Virtual Machine (SVM).

Steps

Set the privilege level to advanced:

```
set -privilege advanced
```

If you want to modify the NFSv3 or NFSv4 TCP maximum transfer size enter the following command

```
vserver nfs modify -vserver vserver_name -tcp-max-xfer-size
integer_max_xfer_size
```

Option	Range	Default
`-tcp-max-xfer-size`	8192 to 1048576 bytes	65536 bytes

Note: The maximum transfer size that you enter must be a multiple of 4 KB (4096 bytes). Requests that are not properly aligned negatively affect performance.

Use the `vserver nfs show -fields tcp-max-xfer-size` command to verify the changes.
If any clients use static mounts, unmount and remount for the new parameter size to take effect.

Example

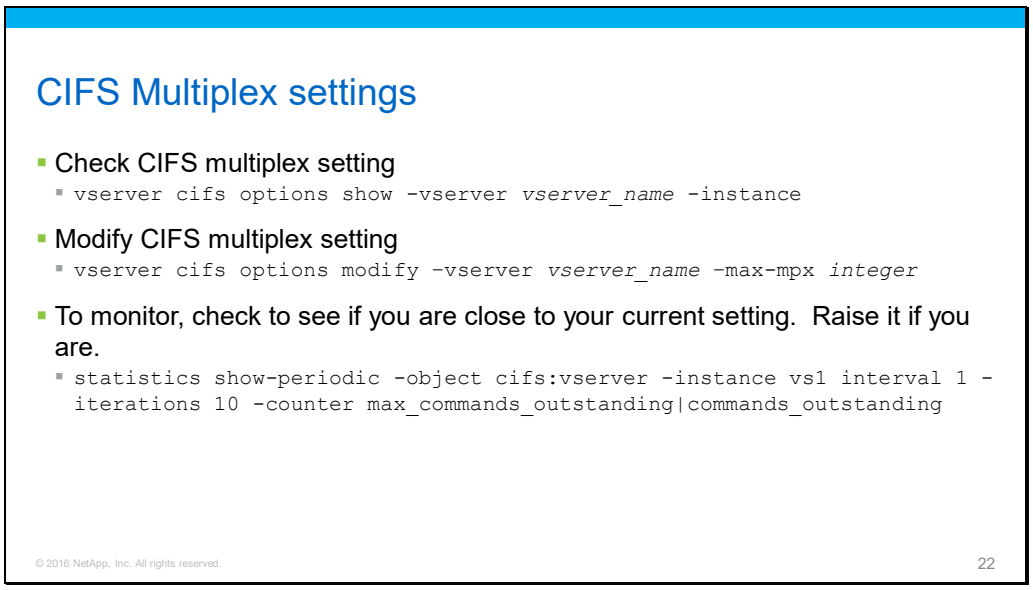
The following command sets the NFSv3 and NFSv4.x TCP maximum transfer size to 1048576 bytes on the SVM named vs1:

```
vs1::> vserver nfs modify -vserver vs1 -tcp-max-xfer-size 1048576
```

Check iSCSI TCP Read/Write size

- Check TCP window size
  - `vserver iscsi show -vserver vserver_name -instance`
- Modify TCP windows size
  - `vserver iscsi modify -vserver vserver_name -tcp-window-size integer`

Increasing the TCP window size to its maximum setting on both the storage system and the client can improve performance for large transfers. The TCP window size controls the number of TCP messages that can transmitted between the storage system and client at one time.



In order to operate more efficiently, some applications such as Citrix and IIS combine many CIFS sessions into the same TCP connection. These connections are known as multiplexed connections. The -max-mpx option determines how many outstanding SMB operations ONTAP can support per single TCP connection. When that value is reached, the client should stop sending further SMB commands to ONTAP.

The value associated with the -max-mpx option should never be set below the Windows NT default of 50. If CIFS statistical output shows the Max Multiplex to be greater than 32, then this number should be raised.

The only legal values are 50, 126, 253, 255* and 1124. Do not set this to any other values without an explicit conversation with NetApp customer service. It should be noted that automatically setting this option to a high value is not recommended. High values consume a lot of resources on the clients. Setting this to an unapproved value will cause errors on some clients.
Note: As of Data ONTAP 7.3.6 & 8.0.2, the default value for the option cifs.max_mpx has been changed to 255.

A similar value may be set on your windows clients to improve performance
https://technet.microsoft.com/en-us/library/cc960259.aspx

If using IIS see this article
https://technet.microsoft.com/en-us/library/cc960259.aspx

For terminal servers see article
https://kb.netapp.com/support/s/article/ka21A0000000WQJQA2/what-is-the-recommended-tuning-for-windows-terminal-server-wts-cifs-clients

FC Adapter port speed

- Try not to use Auto for your port speed. Set the adapter to the speed of the device to which it connects.
- Auto can cause reconnection time to go up during takeover and give back, as well as other interruptions.

The adapter target port speed should match the speed of the device to which it connects, to optimize performance. If the port is set to autonegotiation, it can take longer to reconnect after a takeover and giveback or other interruption.

Lesson 4

Check Disk Response Times

Disk response times

- `statistics disk show -sort-key operation_latency`

```
cluster1::*> statistics disk show -sort-key operation_latency

cluster1 : 3/19/2017 02:03:18

                    Busy Total Read Write   Read Write *Latency
    Disk        Node  (%)   Ops  Ops   Ops  (Bps) (Bps)     (us)
-------- ---------- ---- ----- ---- ----- ------ ----- --------
NET-1.49 cluster1-01  102    28   28     0 237568     0    56833
NET-1.17 cluster1-01  100    28   28     0 439910     0    56817
NET-1.51 cluster1-01  102    28   28     0 137625     0    56759
NET-1.52 cluster1-01  102    28   28     0 165478     0    56757
NET-1.50 cluster1-01  102    28   28     0 136806     0    56754
NET-2.14 cluster1-02    5     1    0     1    819 81100    54736
```

Lesson 5

Cluster object statistics

How to view statistics on any object

- `statistics` *`object`* `show`
  - Try using the `-sort-key` switch

```
cluster1::*> statistics aggregate  show

cluster1 : 3/19/2017 02:07:19

                       *Total Read Write   Read  Write Latency
Aggregate         Node    Ops  Ops   Ops  (Bps)  (Bps)    (us)
--------- ----------- ------ ---- ----- ------ ------ -------
n01_aggr2 cluster1-01    154   53    94 256000 698368   55656
n01_aggr1 cluster1-01    119  108     6 2309120
                                              975872   50535
 aggr0_n1 cluster1-01      6    1     4  14336 305152   64614
n02_aggr2 cluster1-02      0    0     0      0      0       0
n02_aggr1 cluster1-02      0    0     0      0      0       0
 aggr0_n2 cluster1-02      0    0     0      0      0       0
```

```
cluster1::*> statistics lif show

cluster1 : 3/19/2017 02:09:50
                                   *Recv                  Sent
                           Recv     Data   Recv   Sent    Data   Sent Current
              LIF Vserver Packet   (Bps) Errors Packet   (Bps) Errors    Port
----------------- ------- ------ ------- ------ ------ ------- ------ -------
vs1_cifs_nfs_lif2     vs1    201  367634      0    163  631476      0     e0e
vs2_cifs_nfs_lif2     vs2    139  308257      0    114  301584      0     e0e
cluster1-01_clus2 Cluster      5     696      0      5     490      0     e0b
cluster1-01_clus1 Cluster      5     648      0      5     442      0     e0a
cluster1-02_clus1 Cluster      5     490      0      5     696      0     e0a
cluster1-02_clus2 Cluster      5     442      0      5     648      0     e0b
vs2_cifs_nfs_lif1     vs2      0       0      0      0       0      0     e0d
vs1_cifs_nfs_lif1     vs1      0       0      0      0       0      0     e0d
     cluster_mgmt cluster1
```

```
cluster1::*> statistics volume show

cluster1 : 3/19/2017 02:06:37

                                      *Total Read Write Other   Read  Write Latency
          Volume Vserver Aggregate      Ops  Ops   Ops   Ops  (Bps)  (Bps)    (us)
---------------- ------- --------- ------ ---- ----- ----- ------ ------ -------
lun_vsISCSI_1_vol vsISCSI n01_aggr1   3171 1558  1607     5 798080 822912      32
      vs2_vol01      vs2 n01_aggr1     86   44    37     5 180224 153600  184938
      vs1_vol01      vs1 n01_aggr1     78   40    32     5 164864 134144  179116
           vol0 cluster1-02
                            aggr0_n2     56   18    19    13   2734    406      31
           vol0 cluster1-01
                            aggr0_n1      8    0     0     5   3876   1055      29
    vsISCSI_root vsISCSI n01_aggr1      0    0     0     0      0      0       0
      vs2_vol02      vs2 n01_aggr2      0    0     0     0      0      0       0
       vs2_root      vs2 n02_aggr1      0    0     0     0      0      0       0
      vs1_vol02      vs1 n01_aggr2      0    0     0     0      0      0       0
       vs1_root      vs1 n01_aggr1      0    0     0     0      0      0       0
```

```
cluster1::*> statistics workload show

cluster1 : 3/19/2017 02:11:53

                 *Total Read Write Other    Read    Write Latency
     Workload       Ops  Ops   Ops   Ops   (Bps)    (Bps)    (us)
-------------- ------ ---- ----- ----- ------- -------- -------
User-Default     2938 1474  1464     0 1027968 1011200   11998
_USERSPACE_APPS
                   82   18    19    44    3458     2702     238
   _WAFL_SCAN       0    0     0     0       0        0       0
     _WAFL_CP       0    0     0     0       0        0       0
        _WAFL       0    0     0     0       0        0       0
 _VOLUME-MOVE       0    0     0     0       0        0       0
 _Tape-Backup       0    0     0     0       0        0       0
     _STORAGE       0    0     0     0       0        0       0
      _SPARSE       0    0     0     0       0        0       0
_SNAPMIRROR_SLO_STANDARD
                    0    0     0     0       0        0       0
_SNAPMIRROR_SLO_PREMIUM
```


27

Narrow it down to a specific object

- `statistics volume show -volume` *`volume_name`* `-vserver` *`SVM_name`* `-interval` *`interval`* `-iterations` *`iterations`* `-max` *`maximum_instances`*

```
cluster1::*> statistics volume show -volume vs1_vol02 -interval 5

cluster1 : 3/19/2017 02:16:20

                          *Total Read Write Other   Read Write Latency
   Volume Vserver Aggregate    Ops  Ops   Ops   Ops  (Bps) (Bps)    (us)
--------- ------- --------- ------ ---- ----- ----- ------- ----- -------
vs1_vol02     vs1 n01_aggr2     39   23    16     0 249856 65536  465880
```

Sorting statistics

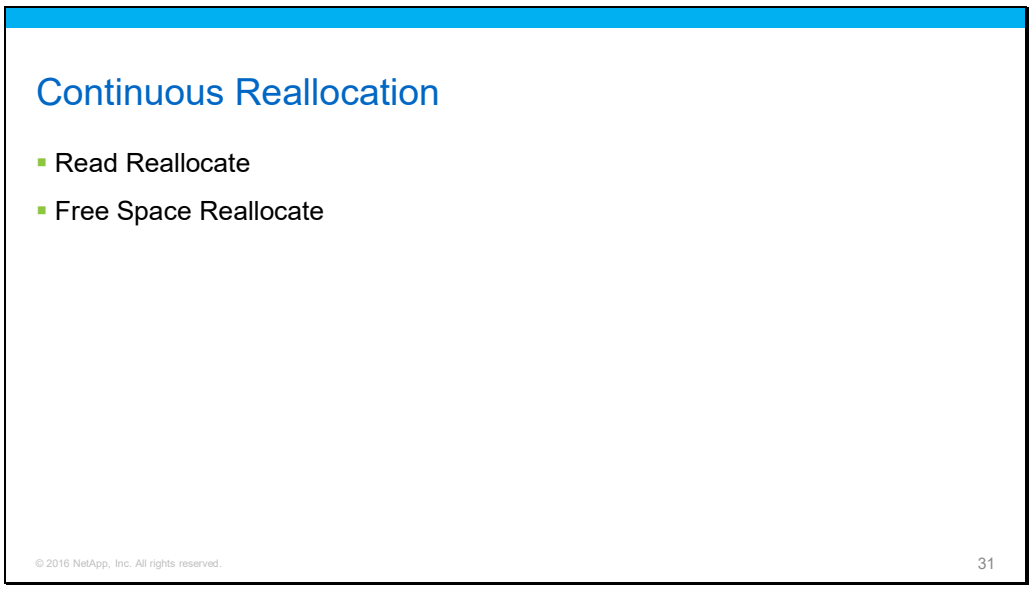
- `statistics volume show -volume` *`volume_name`* `-vserver` *`SVM_name`* `-sort-key` *`sort_counter`* `-interval` *`interval`* `-iterations` *`iterations`* `–max` *`maximum_instances`*

```
cluster1::*> statistics volume show -volume * -sort-key avg_latency -interval 5 -iterations 1 -max 4
cluster1 : 3/19/2017 02:20:37

                          Total Read Write Other   Read  Write *Latency
   Volume Vserver Aggregate  Ops  Ops   Ops   Ops  (Bps)  (Bps)     (us)
--------- ------- --------- ----- ---- ----- ----- ------ ------ --------
vs2_vol02     vs2 n01_aggr2    73   37    36     0 182272 148480   310892
vs2_vol01     vs2 n01_aggr1    94   48    46     0 274432 188416   307271
vsi_vol02     vsi n01_aggr2    86   33    52     0 137216 232448   221676
vsi_vol01     vsi n01_aggr1    86   41    45     0 168960 184320   158374
```


You can sort performance statistics by any counter to diagnose a performance issue or identify a hot spot. You might want to collect volume statistics and sort by total operations to get a list of the most active volumes.

Lesson 6

Performance Optimizations



Read reallocate is a volume option that performs opportunistic reallocation on data to improve performance. Read reallocation uses the normal workload reads along with the read-ahead engine to determine the current layout optimization. If the read was less than optimal, the data will be reallocated to improve the next read of this data. Read reallocate offers both the traditional and physical reallocation methods associated with the reallocate command. Also, because read reallocate uses the existing read workload, it does not require additional scanning or scheduling.

Read reallocate is a volume option that is enabled by using one of the following CLI commands:

7-Mode:
```
toaster> vol options volX read_realloc [on / space_optimized]
```

Cluster-Mode:
```
Cluster::> volume modify –vserver vs1 –volume volX –read_realloc [on /space-optimized]
```

Simply enabling read reallocation by using the “on” function tells Data ONTAP to use the traditional reallocation method. This has the same side effects as traditional reallocate. The `space_optimized` option is synonymous with the physical reallocation method. It does not introduce the Snapshot space 7 Reallocate Best Practices requirements associated with the traditional reallocation method. The `space_optimized` method does, however, introduce the Snapshot read performance impact, similar to physical reallocation.

Reallocate and read reallocation are complementary. Both features can be employed together to accomplish the same goal of improving spatial layout. Read reallocate can help reallocate by reducing the amount of work that reallocate needs to do in each iteration, depending on the workload applied to the system.

An example of when to employ both is in a database system with weekly large table scans. Enabling read reallocate maintains an optimal layout of frequently accessed data, while a scheduled reallocate execution prior to the large table scan might help improve the table scan speed.

Free space reallocation is an aggregate option, introduced in Data ONTAP 8.1.1, which performs opportunistic free space reallocation to maintain an optimal free space layout. When enabled, if Data ONTAP detects that free space is not optimal, it will physically move data to produce areas of contiguous free space. Optimized free space improves the efficiency of WAFL® (Write Anywhere File Layout) and can reduce overall disk utilization.

Some additional CPU utilization should be expected since Data ONTAP will be doing additional work to manage the movement of data. If the performance of the storage system is limited by CPU, enabling free space reallocation is not recommended.

Free space reallocation can be enabled on a per-aggregate basis using the following commands:

7-Mode:
```
toaster> aggr options aggr1 free_space_realloc on
```

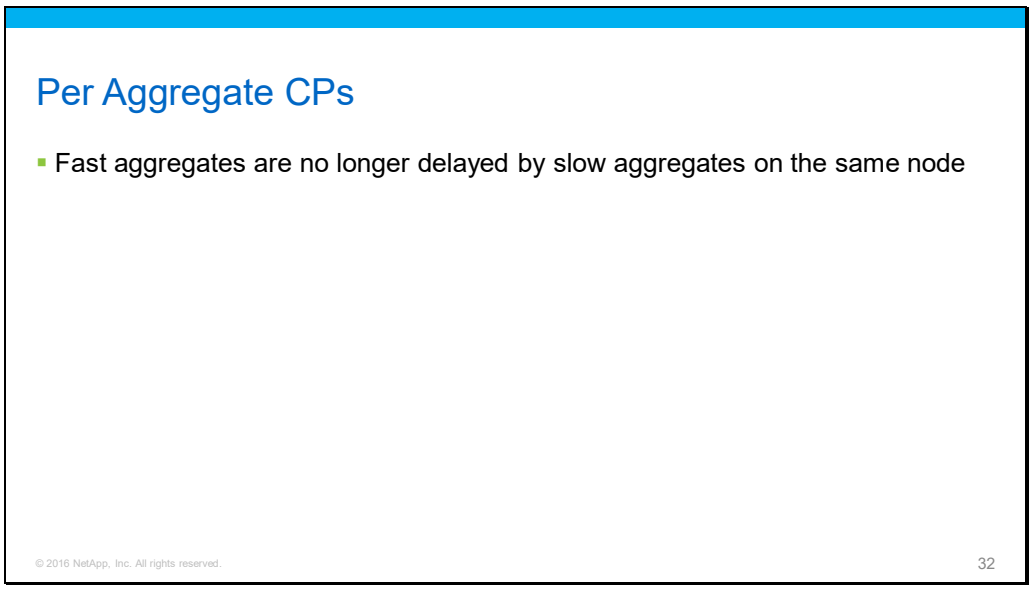
Cluster-Mode:
```
Cluster::> storage aggregate modify –aggregate aggr1 –free-spacerealloc on
```

For best results, the option should be enabled when creating a new aggregate. If enabling the option on an existing aggregate with data already stored on it, there might be a period of time when Data ONTAP will be performing additional work to optimize free space and might affect system performance temporarily.

Similar to physical reallocate, aggregate reallocation, and `space-optimized` read reallocate, free space reallocation will not move blocks that are kept in aggregate Snapshot copies. This does not apply to volume Snapshot copies; free space reallocate will still move blocks stored in volume Snapshot copies. When enabling free space reallocation on an aggregate, also consider enabling read reallocate `space_optimized` for the volumes in the aggregate. The two are complementary technologies that help maintain optimal layout. Read reallocate will optimize the system for sequential reads on the fly, while free space reallocate will optimize for writes.

Read Reallocate
https://library.netapp.com/ecmdocs/ECMP1368862/html/GUID-CDF57F21-38A5-4103-89C5-575C7DED57E1.html

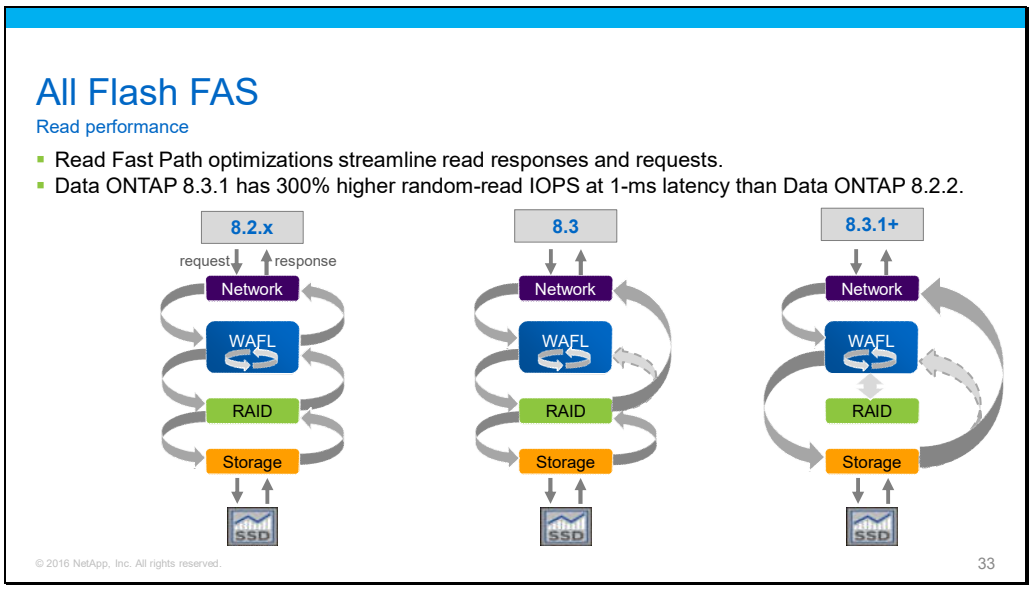
Free Space Reallocation
https://library.netapp.com/ecmdocs/ECMP1368862/html/GUID-C8D2657F-BDF9-4FDB-B417-A53867A7B5BD.html



In earlier versions of Data ONTAP it was possible to get Back-to-Back CPs because a write completion was delayed by a slow aggregate.

Under normal operation all of the data queued to be written by a CP had to be written before the write was completed. Even if the data to be written was in two different aggregates, and one was all SSD and one was all SATA. The SSD aggregate data should be written much faster than data going to the SATA aggregate. However, the CP was dependent on all of the data being written before it could be marked as complete and NVRAM buffers marked as available for overwrite.

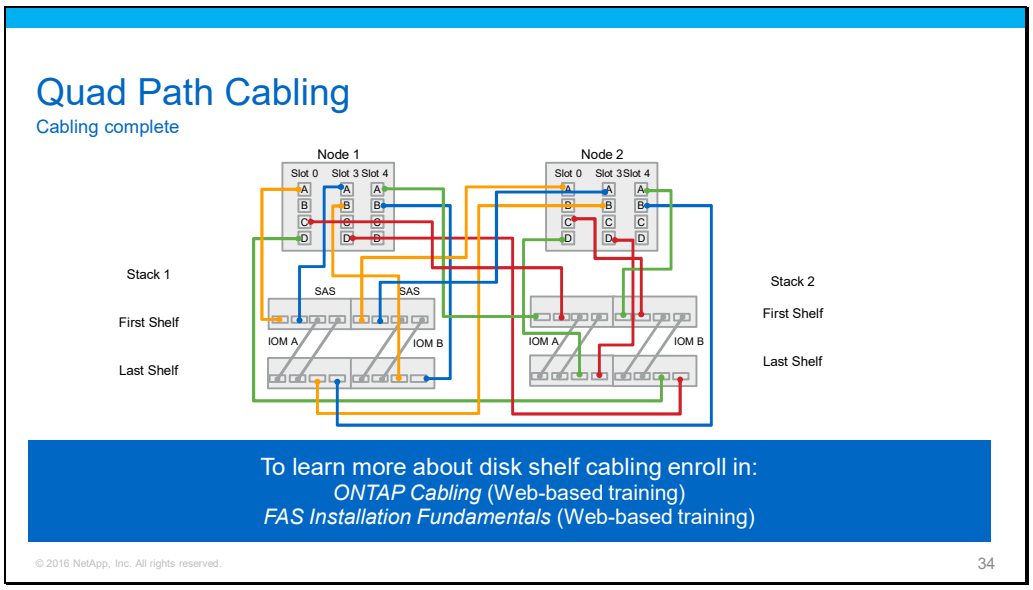
In 9.0 Per-Aggregate CPs were added. Now data is queued by aggregate and the faster SSD data is written and flagged for overwrite while the slower SATA data is still being written. This process speeds up the rate at which data can enter and be processed in Memory and in NVRAM. This improves overall write performance for the storage systems.



In Data ONTAP 8.3, if no errors are detected, read responses are passed directly from RAID to the network, bypassing WAFL. When an error is detected, the WAFL code must get involved.

Read optimization in Data ONTAP 8.3.1:

- Read requests are passed directly from WAFL to storage, bypassing RAID.
- Unless error-recovery is required, read responses are passed directly from storage to the network, bypassing both RAID and WAFL.



This slide shows the cabling completed.

SAS-3 capable disk shelves can be cabled in dual port MPHA in the same way as SAS-2 ports, but there is an option to use quad path cabling as shown. Only two stacks of shelves are connected in the illustration.

When the quad path cabling to two stacks is completed, you should have 4 paths connected at the top and 4 at the bottom of each stack.

If you follow the port selection algorithm, you should be able to quickly select the correct ports to provide continuous connectivity to all disks.

Throughput of a stack may be limited by the cabling.

IOM3 can do approximately 12Gb/s of throughput per path, IOM6 can do approximately 24Gb/s and IOM12 can do approximately 48Gb/s.

Given a standard dual path MPHA design, and IOM6 cabling, a stack can do approximately 4.8GB/s of total throughput. With that limit it would take about 1.5 shelves of SSDs to saturate the stack.

FAS3170 Comparative Drive Count for Peak Controller Throughput			
Workload	**Peak Throughput**	**Number of 15k rpm HDDs**	**Number of SSDs**
4k FCP random read	64,000 IOPS	215	8
64kB FCP sequential read	1,000MB/sec	20	8
32kB NFS sequential write	450MB/sec	12	8
OLTP	47,000 IOPS	98	11

Lesson 7

Summary

Things to keep in Mind

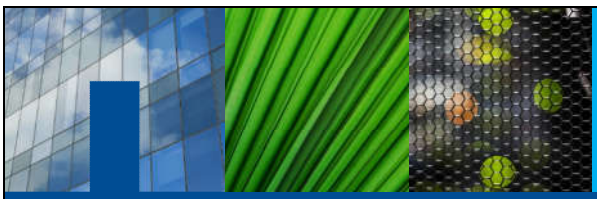
- Design
  - Shelf cabling
    - When using SSD's consider quad path cabling
  - Disk type
  - Raid Group layout
  - Network Layout
- Troubleshooting
  - Data path
    - Front end network
    - Back end network (indirect I/O)
  - CPU
  - Cluster Activity
    - Vol moves
    - Snapmirror (intra/inter cluster)

Summary

- Setting up basic monitoring
- Identifying the most active instance of an object in a cluster
- Identify and resolve performance issues
- Aggregate and Volume optimizations

Exercise

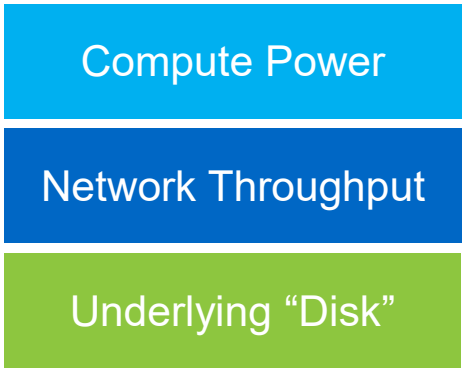
- This exercise should take 60 minutes



Chapter 11

ONTAP Cloud for AWS – Performance Factors

Factors impacting ONTAP Cloud Performance in AWS



- Compute ‘horsepower’ is tied to the AWS EC2 Instance
  - Controlling the amount of virtual CPU and Memory
- The network bandwidth is also tied to the AWS EC2 instance
  - And can typically be the bottleneck for performance
- The “disk” type, size, and number of disks
  - Deterministic factors for ensuring your back-end data I/O performance factors are met

Key contributors to the overall performance of an ONTAP Cloud deployment

2

Compute & Network Factor – AWS EC2 Instance Types

Instance type	Max. bandwidth (Mbps)*	Expected throughput (MB/s)**	Max. IOPS (16 KB size)**	License
c4.8xlarge	4,000	500	32,000	Premium
c4.4xlarge	2,000	250	16,000	
r4.2xlarge	1600	200	12,000	
r3.2xlarge	1,000	125	8,000	
c4.2xlarge	1,000	125	8,000	Standard
m4.2xlarge	1,000	125	8,000	
m3.2xlarge	1,000	125	8,000	
r4.xlarge	800	100	6,000	
r3.xlarge	500	62.5	4,000	
m4.xlarge	750	93.75	6,000	Explore
m3.xlarge	500	62.5	4,000	

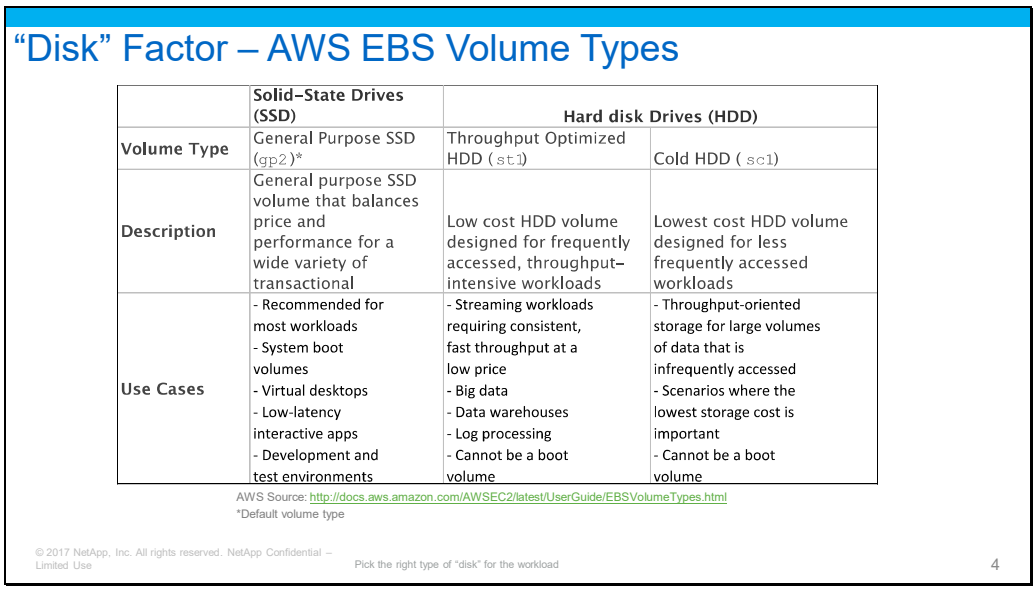
AWS Source: http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSOptimized.html

Size matters. Pick a larger EC2 Instance Type and increase your compute power and throughput

3

EC2 = Elastic Cloud Compute in AWS speak

Information collected from AWS online documentation:
http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSOptimized.html



EBS – Elastic Block Storage in AWS speak. This is their version of "block" disk available in AWS.

Information collected from AWS online documentation:
http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html

*Default volume type
**gp2/io1 based on 16KiB I/O size, st1/sc1 based on 1 MiB I/O size
† To achieve this throughput, you must have an instance that supports it, such as r3.8xlarge or x1.32xlarge.

“Disk” Factor – AWS EBS Volume Types

	Solid-State Drives (SSD)	Hard disk Drives (HDD)	
Volume Type	General Purpose SSD (gp2)*	Throughput Optimized HDD (st1)	Cold HDD (sc1)
Volume Size	1 GiB – 16 TiB	500 GiB – 16 TiB	500 GiB – 16 TiB
Max. IOPS/Volume**	10,000	500	250
Max. Throughput/Volume†	160 MiB/s	500 MiB/s	250 MiB/s
Max. IOPS/Instance	65,000	65,000	65,000
Max. Throughput/Instance	1,250 MiB/s	1,250 MiB/s	1,250 MiB/s
Dominant PerformanceAttribute	IOPS	MiB/s	MiB/s

AWS Source: http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html

*Default volume type
**gp2/io1 based on 16KiB I/O size, st1/sc1 based on 1 MiB I/O size
† To achieve this throughput, you must have an instance that supports it, such as r3.8xlarge or x1.32xlarge.

Pick the right type of “disk” for the workload

5

EBS – Elastic Block Storage in AWS speak. This is their version of “block” disk available in AWS.

Information collected from AWS online documentation:
http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html

ONTAP Cloud Sizing Exercise

- EC2
  - Customer wants to move an OLTP DB to the cloud
  - The Requirement is for peak workloads of ~11k IOPS *
  - From the chart on the right, the customer should use the r4.2xlarge, or larger instance types with an ONTAP Cloud Premium license
- EBS
  - Based on the chart to the right, the customer will need to use multiple EBS GP2 volumes for the data aggregate
  - Rules:
    - You receive 3 IOPS per GB, so for 1000GB (~ 1TB) that's 3k IOPS
    - You can have a max of 10k IOPS per EBS GP2 Volume
  - Based on the sizes of ONTAP 'disks', the customer could use:
    - 4 * 1TB EBS gp2 = 4TB ; with 4 * 3000 IOPS = 12k IOPS
    - 2 * 2TB EBS gp2 = 4TB ; with 2 * 6000 IOPS = 12k IOPS
    - 2 * 3TB EBS gp2 = 6TB; with 2 * 9000 IOPS = 18k IOPS

Instance type	Max. bandwidth (Mbps)*	Expected throughput (MB/s)**	Max. IOPS (16 KB size)**	License
c4.8xlarge	4,000	500	32,000	Premium
c4.4xlarge	2,000	250	16,000	
r4.2xlarge	1600	200	12,000	
r3.2xlarge	1,000	125	8,000	
c4.2xlarge	1,000	125	8,000	Standard
m4.2xlarge	1,000	125	8,000	
m3.2xlarge	1,000	125	8,000	

AWS Source: http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSOptimized.html

	Solid-State Drives (SSD)	Hard disk Drives (HDD)	
Volume Type	General Purpose SSD (gp2)*	Throughput Optimized HDD (st1)	Cold HDD (sc1)
Volume Size	1 GiB - 16 TiB	500 GiB - 16 TiB	500 GiB - 16 TiB
Max. IOPS**/Volume	10,000	500	250
Max. Throughput/Volume†	160 MiB/s	500 MiB/s	250 MiB/s

AWS Source: http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html

ONTAP Cloud Disk Sizes			
Volume Type	General Purpose SSD (gp2)^	Throughput Optimized HDD	Cold HDD (sc1)
Size in TB	.5, 1, 2, 4, 8	1, 2, 4, 8	2, 4, 8

Example

* based on 16KiB I/O size

6

NOTES:

Please understand that this is a very basic and high level (yes... high level) exercise for understanding what AWS resources will be required to support the workload a customer wants to run on an ONTAP Cloud system.

There are other major factors of overall throughput and is why the throughput expectation numbers are listed.

ONTAP Cloud Re-Sizing

- EC2

  Scenario: The customer wants to change the size of AWS system they're using for ONTAP Cloud… can they?

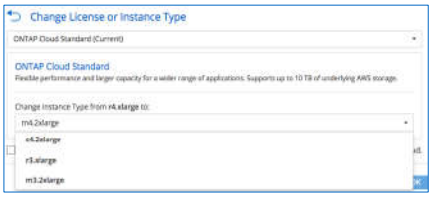
  Response: Absolutely. Cloud Manager makes it easy to adjust the AWS EC2 Instance type and ONTAP Cloud license being used

- EBS

  Scenario: The customer wants to change the number of AWS EBS 'disks', or the type of EBS they use with ONTAP Cloud… can they?

  Response: Yes again. Cloud Manager makes it easy to adjust the AWS EBS 'disks' an ONTAP Cloud system is using

  - A customer can choose to add additional disk of the same size and type
  - Or add a new aggregate containing disk of different EBS sizes and/or types.
    - Cloud Manager also makes it easy to move the volume to the new aggregate, while it is actively being used



What happens if you want to change the AWS Resources being used?

7

NOTES:

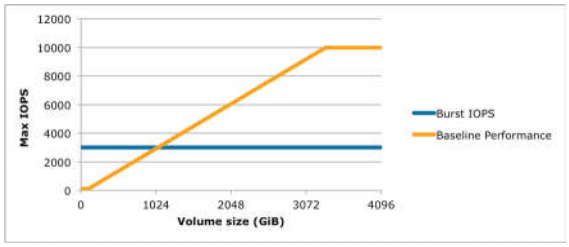
Cloud Manager enables customers to change the ONTAP Cloud license (if they want to use more / less underlying capacity), or to change the type of EC2 instance being used for ONTAP Cloud

Cloud Manager makes it easy for a customer to add additional EBS disks to their aggregates under 'Advanced Allocation'. Or if a customer wants to move to an aggregate of a different EBS type and size, they simply add that new aggregate... and once that new aggregate is available, they can use "change underlying disk type" to move the active volume from one AWS disk type to another. The ONTAP 'vol move' technology is being used to complete this action. Once the old aggregate has been vacated, it can be deleted to reclaim/release unused AWS resources.

Additional information for AWS EBS Volumes

“Disk” Factor – AWS EBS Volume Types

General Purpose SSD (gp2) volumes offer cost-effective storage that is ideal for a broad range of workloads. These volumes deliver single-digit millisecond latencies and the ability to burst to 3,000 IOPS for extended periods of time. Between a minimum of 100 IOPS (at 33.33 GiB and below) and a maximum of 10,000 IOPS (at 3,334 GiB and above), baseline performance scales linearly at 3 IOPS per GiB of volume size.



AWS Source: http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html

AWS EBS General Purpose SSD (gp2) volume type

9

EBS – Elastic Block Storage in AWS speak. This is their version of “block” disk available in AWS.

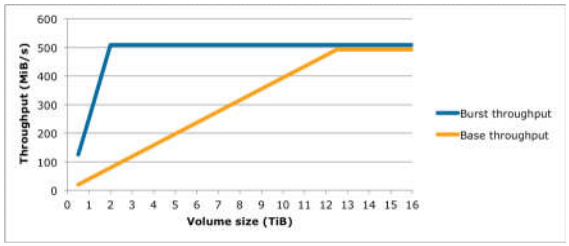
Information collected from AWS online documentation:
http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html

“Disk” Factor – AWS EBS Volume Types

Throughput Optimized HDD (st1) volumes provide low-cost magnetic storage that defines performance in terms of throughput rather than IOPS. This volume type is a good fit for large, sequential workloads.

Note

This volume type is optimized for workloads involving large, sequential I/O, and we recommend that customers with workloads performing small, random I/O use gp2.



AWS Source: http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html

AWS EBS Throughput Optimized HDD (st1) volume type

10

EBS – Elastic Block Storage in AWS speak. This is their version of “block” disk available in AWS.

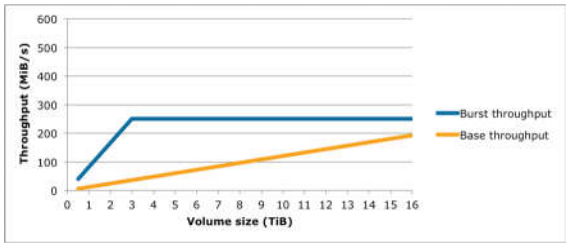
Information collected from AWS online documentation:
http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html

“Disk” Factor – AWS EBS Volume Types

Cold HDD (sc1) volumes provide low-cost magnetic storage that defines performance in terms of throughput rather than IOPS. With a lower throughput limit than st1, sc1 is a good fit ideal for large, sequential cold-data workloads. If you require infrequent access to your data and are looking to save costs, sc1 provides inexpensive block storage.

Note

This volume type is optimized for workloads involving large, sequential I/O, and we recommend that customers with workloads performing small, random I/O use gp2.



AWS Source: http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html

AWS EBS Cold HDD (sc1) volume type

11

EBS – Elastic Block Storage in AWS speak. This is their version of “block” disk available in AWS.

Information collected from AWS online documentation:
http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html

Summary

- Factors impacting Performance
- Instance Types
- Volume Types
- Disk Types

Exercise

- There is no exercise for this chapter