# Assessment of Emerald Data for Storage Products

ITI Energy Efficiency Working Group, SNIA Emerald TWG, and ENERGY STAR® Discussion

Storage Product Stakeholders Meeting: January 19, 2017

# Agenda

- Composition of Current Dataset
- Categorization of Storage Products beyond the On-line Categories
- General Observations Regarding the Dataset and the Emerald Test Data
  - Overall Data set
  - Family Comparisons
- General Observations on Emerald Test results for Storage products and the Impact of changes in drive type and count, system architecture, resiliency levels, and technology generation on the Emerald test results.
  - Recap of November 2015 Observations
- Benefits of COMs and Impact on Emerald Results and Data Center Energy Use and Consumption
- ITI Recommendations for Storage Version 2 Requirements and Conclusions

Notes: The use of the word “score” refers to the performance/power efficiency score.
Charts 28 to 38 are additional data.

# Composition of Current Dataset

- Data collected through November of 2016: Storage Products Certified to ENERGY STAR
  - 159 Configurations: Compared to 105 Configurations in November 2015 review
  - 45 Families compared to 25 Families in November
  - 14 (9) Manufacturers have equipment in the dataset.
    - Some manufacturers are branding other manufacturer's equipment.

| On-line Category | Number of Families | Number of Configurations | Number of Manufacturers |
|---|---|---|---|
| 2 Transactional | 3 (1) | 11 (7) | 2 (1) |
| 2 Sequential | 5 (2) | 24 (9) | 2 (1) |
| 3 Transactional | 16 (11) | 48 (29) | 5 (2) |
| 3 Sequential | 4 (3) | 10 (9) | 2 (2) |
| 4 Transactional | 10 (6) | 35 (22) | 5 (3) |
| 4 Sequential | 10 (7) | 27 (28) | 5 (4) |

Note: 2016 dataset (2015 dataset)

# Categorization of Storage Products underneath Online Categories

- Workload Type: Transaction, Streaming, Capacity

- Drive Type:  HDD or SSD
  - HDD:  Drive capacity, rpm and form factor
  - SSD Drive type - SSD, Flash, Non-volatile DIMM – and capacity

- Drive Count

- Connectivity:
  - Server to Controller
  - Controller to Storage Media

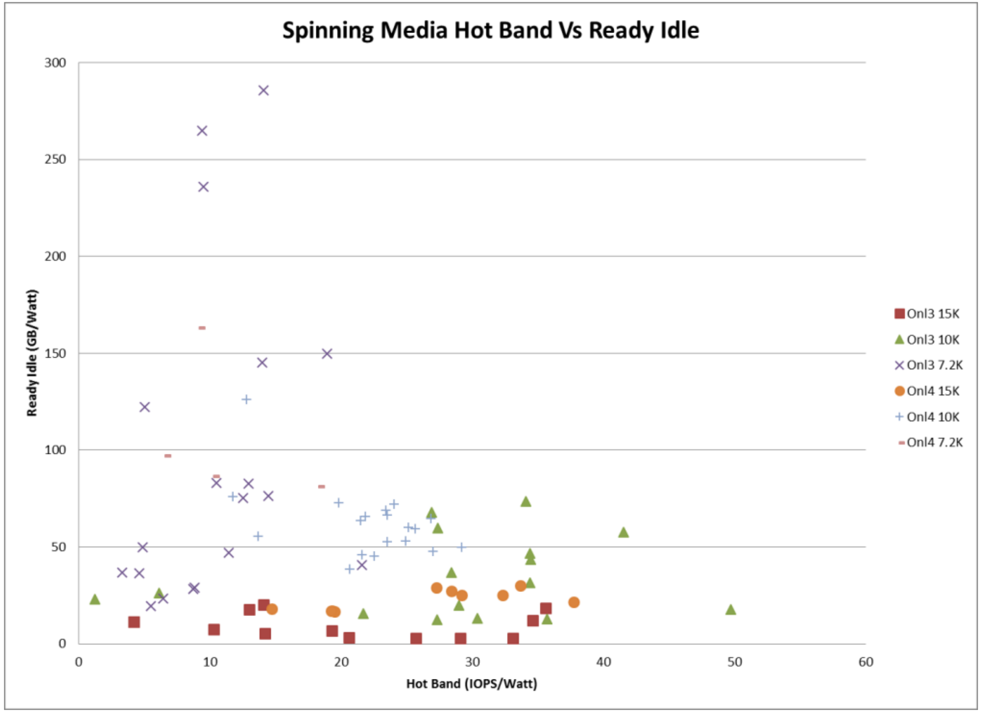
- Controller Cache Size

Note: Need these Groupings to get representative comparisons between products.

# Presentation of Data

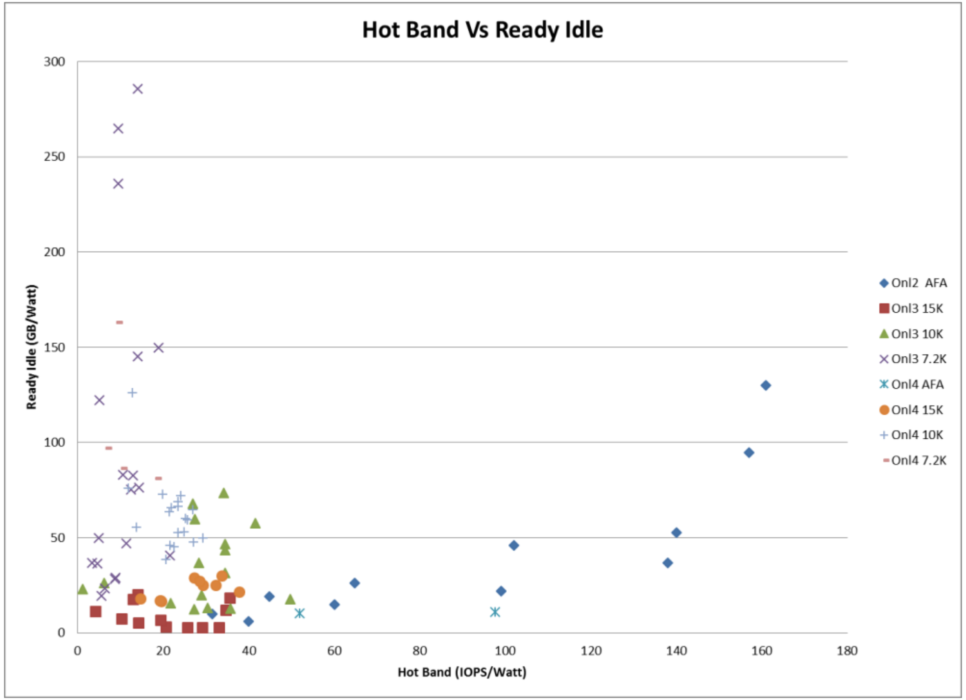
Data analysis will be based on representative data for OL-3 and OL-4 systems.  In general the observations and conclusions generated from this analysis hold for the overall dataset.

# Hot Band vs. Ready Idle for HDD systems



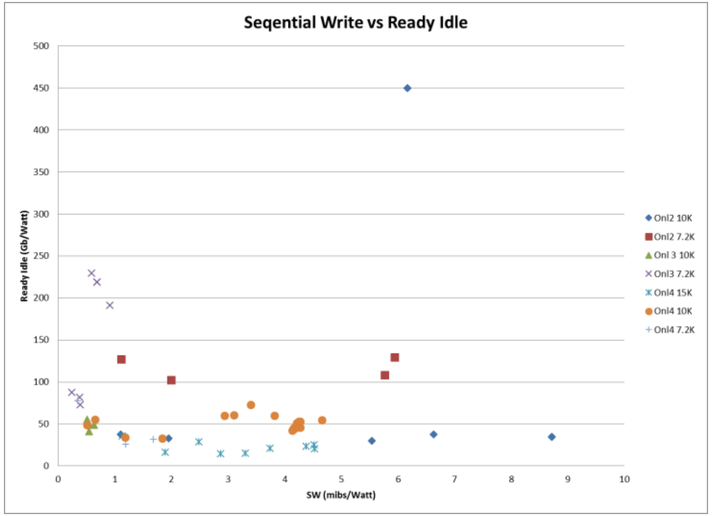

- Ready Idle (GB/W)
  - Largely segregated by On-line category and drive speed.
  - 7.2 K systems have a broad Ready Idle distribution.
- Hot Band (IOPs/W)
  - 15 K OL-3,4 and 10 K OL-3 all have broad distribution.
  - 10 K OL-3,4 systems more clustered: OL-4 is largely one manufacturer.
  - 7.2 K systems generally have the lowest IOPs/W values.

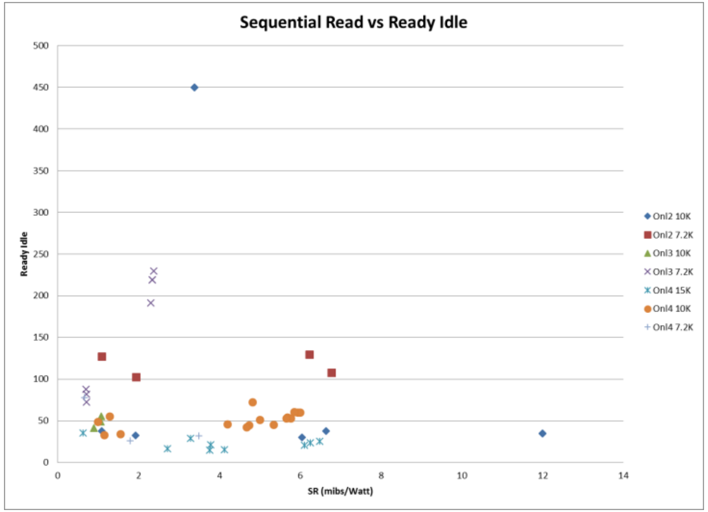
# Hot Band vs. Ready Idle for HDD and AFA Systems




- Ready Idle (GB/W)
  - Largely segregated by On-line category and drive speed.
  - 7.2 K systems have a broad Ready Idle distribution.
- Hot Band (IOPs/W)
  - 15 K OL-3,4 and 10 K OL-3 all have broad distribution.
  - 10 K OL-4 systems clustered, but largely one mfg.
  - 7.2 K systems generally have the lowest IOPs/W values.
- AFAs
  - Higher IOPs/Watt than HDDs in almost all cases.
  - Generally higher GB/W

# Sequential Read and Write Scores vs Ready Idle

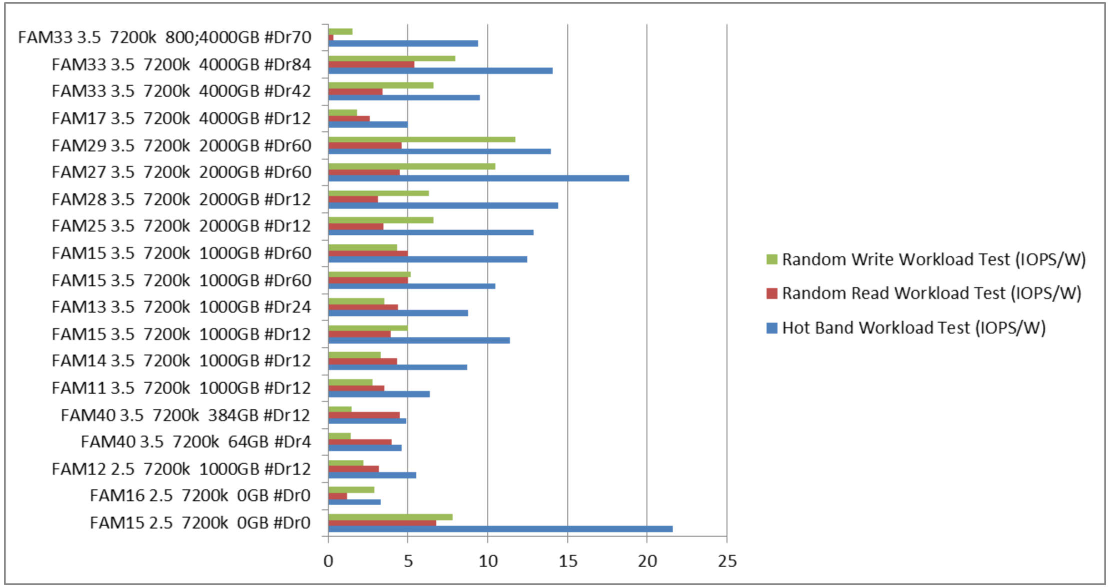



Observations:

- Ready Idle: GB/W values cluster by drive type with the exception of the OL-3 7.2 K systems.
- mibs/W: All systems have wide distribution of mibs/watt values.

# OL-3 Data Analysis: Transactional w/ 7.2 K drives

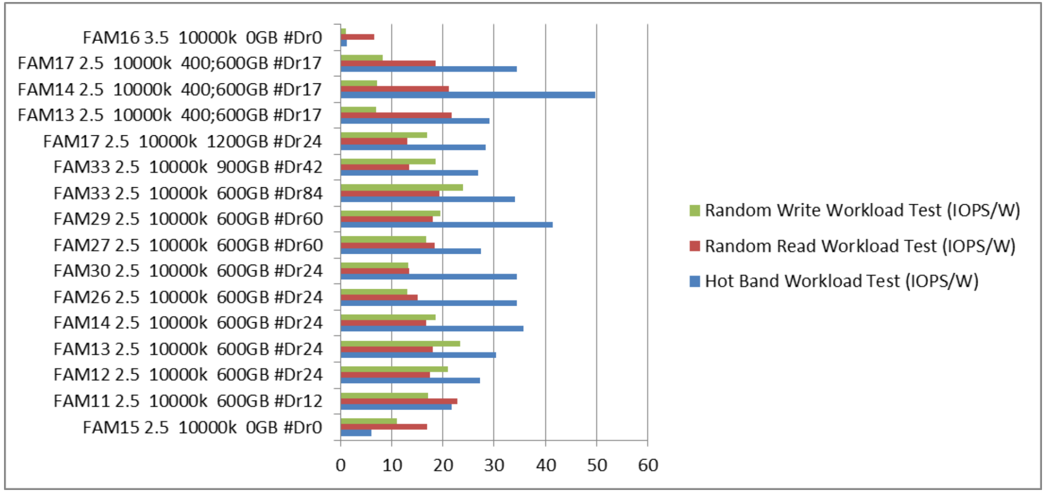



Observations:

- In general, systems with lower drive counts have lower scores than those with high drive counts due to:
    - better amortization of controller power.
    - smaller drive counts may not fully utilize the data pathways (non-optimal).
- There is differentiation between systems with same drive speed/capacity/count. Contributing factors:
    - System/controller architecture
    - Technology generation of the controller.
    - Redundant or single power supplies.

Legend Notes: Family#, FF, rpm, GB, Device count

# OL-3 Data Analysis: Transactional w/ 10 K drives

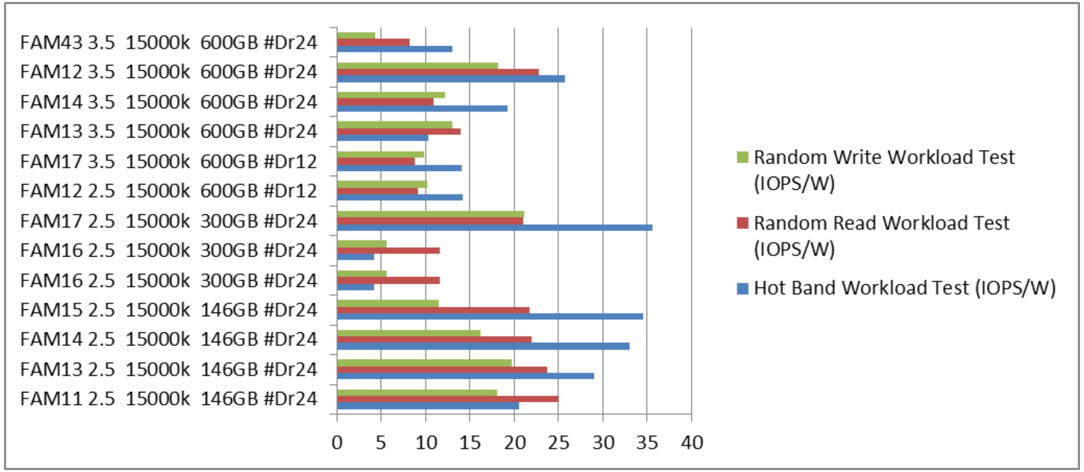



Observations:

- In general, systems with lower drive counts have lower scores than those with high drive counts due to:
  - better amortization of controller power.
  - smaller drive counts may not fully utilize the data pathways (non-optimal).
- There is differentiation between systems with same drive speed/capacity/count.
  - Architecture of the system/controller.
  - Redundant vs. single power supply
- The five families (12,13,14,26,30) with 24 count, 2.5", 10 K, 600 GB drives show differences in IOPs/W results.

Legend Notes: Family#, FF, rpm, GB, Device count

# OL-3 Data Analysis: Transactional w/ 15 K drives




Legend Notes: Family#, FF, rpm, GB, Device count

Observations:

- In general, systems with low drive counts have lower scores than those with high drive counts due to:
  - better amortization of controller power.
  - smaller drive counts may not fully utilize the data pathways (non-optimal).
- There is differentiation between systems with same drive speed/capacity/count.
  - Architecture of the system/controller.
  - Redundant vs. single power supply
  - The four families with 24 count, 2.5”, 15 K, 146 GB and 24 count, 3.5”, 15 K, 600 GB drives show differences in IOPs/W results.
- In general, 2.5 inch drives have better hot band scores than 3.5 inch drives in the same family (13,14) where drive counts are the same.

The observations for OL-3 systems are consistent with those made in evaluation the OL-4 Transactional results.

# OL-3 Transactional Systems: Top 25%

| Label | Hot Band Workload Test (IOPS/W) | Random Read Workload Test (IOPS/W) | Random Write Workload Test (IOPS/W) | Ready Idle Workload Test (GB/W) |
|---|---|---|---|---|
| FAM15 2.5 7200k 0GB #Dr0 | 21.6 | 6.8 | 7.8 | 40.7 |
| FAM16 2.5 7200k 0GB #Dr0 | 3.3 | 1.2 | 2.9 | 36.8 |
| FAM12 2.5 7200k 1000GB #Dr12 | 5.5 | 3.2 | 2.2 | 19.4 |
| FAM40 3.5 7200k 64GB #Dr4 | 4.6 | 3.97 | 1.4 | 36.49 |
| FAM40 3.5 7200k 384GB #Dr12 | 4.88 | 4.49 | 1.48 | 49.9 |
| FAM11 3.5 7200k 1000GB #Dr12 | 6.4 | 3.5 | 2.8 | 23.3 |
| FAM14 3.5 7200k 1000GB #Dr12 | 8.7 | 4.3 | 3.3 | 28.1 |
| FAM15 3.5 7200k 1000GB #Dr12 | 11.4 | 3.9 | 5 | 46.9 |
| FAM13 3.5 7200k 1000GB #Dr24 | 8.8 | 4.4 | 3.5 | 29.1 |
| FAM15 3.5 7200k 1000GB #Dr60 | 10.5 | 5 | 5.2 | 83 |
| FAM15 3.5 7200k 1000GB #Dr60 | 12.5 | 5 | 4.3 | 75.1 |
| FAM25 3.5 7200k 2000GB #Dr12 | 12.91 | 3.45 | 6.6 | 82.61 |
| FAM28 3.5 7200k 2000GB #Dr12 | 14.42 | 3.14 | 6.3 | 76.38 |
| FAM27 3.5 7200k 2000GB #Dr60 | 18.9 | 4.48 | 10.47 | 149.6 |
| FAM29 3.5 7200k 2000GB #Dr60 | 13.97 | 4.6 | 11.72 | 145.25 |
| FAM17 3.5 7200k 4000GB #Dr12 | 5 | 2.6 | 1.8 | 122.3 |
| FAM33 3.5 7200k 4000GB #Dr42 | 9.5 | 3.4 | 6.6 | 236.1 |
| FAM33 3.5 7200k 4000GB #Dr84 | 14.1 | 5.4 | 8 | 285.7 |
| FAM33 3.5 7200k 800;4000GB #Dr70 | 9.4 | 0.3 | 1.5 | 264.9 |
| FAM15 2.5 10000k 0GB #Dr0 | 6.1 | 16.9 | 11 | 26.2 |
| FAM11 2.5 10000k 600GB #Dr12 | 21.7 | 22.9 | 17.1 | 15.5 |
| FAM12 2.5 10000k 600GB #Dr24 | 27.3 | 17.4 | 21 | 12.5 |
| FAM13 2.5 10000k 600GB #Dr24 | 30.4 | 18.1 | 23.3 | 13 |
| FAM14 2.5 10000k 600GB #Dr24 | 35.7 | 16.7 | 18.5 | 12.7 |
| FAM26 2.5 10000k 600GB #Dr24 | 34.41 | 15.15 | 13.04 | 46.74 |
| FAM30 2.5 10000k 600GB #Dr24 | 34.42 | 13.51 | 13.26 | 43.3 |
| FAM27 2.5 10000k 600GB #Dr60 | 27.34 | 18.32 | 16.77 | 59.71 |
| FAM29 2.5 10000k 600GB #Dr60 | 41.53 | 18.01 | 19.46 | 57.71 |
| FAM33 2.5 10000k 600GB #Dr84 | 34.1 | 19.4 | 23.9 | 73.6 |
| FAM33 2.5 10000k 900GB #Dr42 | 26.9 | 13.4 | 18.5 | 67.8 |
| FAM17 2.5 10000k 1200GB #Dr24 | 28.4 | 13.1 | 16.9 | 36.6 |
| FAM13 2.5 10000k 400;600GB #Dr17 | 29 | 21.8 | 6.9 | 19.8 |
| FAM14 2.5 10000k 400;600GB #Dr17 | 49.7 | 21.1 | 7.2 | 17.7 |
| FAM17 2.5 10000k 400;600GB #Dr17 | 34.4 | 18.5 | 8.3 | 31.5 |
| FAM16 3.5 10000k 0GB #Dr0 | 1.2 | 6.6 | 1 | 22.9 |
| FAM11 2.5 15000k 146GB #Dr24 | 20.6 | 25 | 18.1 | 3.1 |
| FAM13 2.5 15000k 146GB #Dr24 | 29.1 | 23.8 | 19.7 | 2.8 |
| FAM14 2.5 15000k 146GB #Dr24 | 33.1 | 22 | 16.2 | 2.7 |
| FAM15 2.5 15000k 146GB #Dr24 | 34.6 | 21.7 | 11.5 | 12 |
| FAM16 2.5 15000k 300GB #Dr24 | 4.2 | 11.7 | 5.6 | 11.3 |
| FAM16 2.5 15000k 300GB #Dr24 | 4.2 | 11.7 | 5.6 | 11.3 |
| FAM17 2.5 15000k 300GB #Dr24 | 35.6 | 21 | 21.2 | 18.3 |
| FAM12 2.5 15000k 600GB #Dr12 | 14.2 | 9.2 | 10.2 | 5.3 |
| FAM17 3.5 15000k 600GB #Dr12 | 14.1 | 8.8 | 9.9 | 20.3 |
| FAM13 3.5 15000k 600GB #Dr24 | 10.3 | 14 | 13.1 | 7.3 |
| FAM14 3.5 15000k 600GB #Dr24 | 19.3 | 10.9 | 12.2 | 6.8 |
| FAM12 3.5 15000k 600GB #Dr24 | 25.7 | 22.8 | 18.2 | 2.7 |
| FAM43 3.5 15000k 600GB #Dr24 | 13 | 8.2 | 4.4 | 17.5 |

<u>48 Transactional Submissions:</u>

- (19) 7.2 K rpm submissions (blue highlight)
- (16) 10 K rpm submissions (white)
- (13) 15 K rpm submissions (purple)

<u>Observations</u>:

- All but 2 of the highest Ready Idle values are on the 7.2 K rpm high capacity drives.
  - The two 10 K rpm systems with Top 25% Ready Idle values have high drive counts.
- The top 25% Hot Band scores are distributed between the 15 K and 10 K rpm drives.
  - Higher drive count configurations.
  - Lower capacity drives.
- Only 1 system is top 25% for all 4 metrics: Largest drive count.

# OL-3 Transactional Systems with 10 K HDDs

| Label | Hot Band Workload Test (IOPS/W) | Random Read Workload Test (IOPS/W) | Random Write Workload Test (IOPS/W) | Ready Idle Workload Test (GB/W) |
|---|---|---|---|---|
| FAM15 2.5 10000 0GB #Dr0 | 6.1 | 16.9 | 11 | 26.2 |
| FAM11 2.5 10000 600GB #Dr12 | 21.7 | 22.9 | 17.1 | 15.5 |
| FAM12 2.5 10000 600GB #Dr24 | 27.3 | 17.4 | 21 | 12.5 |
| FAM13 2.5 10000 600GB #Dr24 | 30.4 | 18.1 | 23.3 | 13 |
| FAM14 2.5 10000 600GB #Dr24 | 35.7 | 16.7 | 18.5 | 12.7 |
| FAM26 2.5 10000 600GB #Dr24 | 34.41 | 15.15 | 13.04 | 46.74 |
| FAM30 2.5 10000 600GB #Dr24 | 34.42 | 13.51 | 13.26 | 43.3 |
| FAM27 2.5 10000 600GB #Dr60 | 27.34 | 18.32 | 16.77 | 59.71 |
| FAM29 2.5 10000 600GB #Dr60 | 41.53 | 18.01 | 19.46 | 57.71 |
| FAM33 2.5 10000 600GB #Dr84 | 34.1 | 19.4 | 23.9 | 73.6 |
| FAM33 2.5 10000 900GB #Dr42 | 26.9 | 13.4 | 18.5 | 67.8 |
| FAM17 2.5 10000 1200GB #Dr24 | 28.4 | 13.1 | 16.9 | 36.6 |
| FAM13 2.5 10000 400;600GB #Dr17 | 29 | 21.8 | 6.9 | 19.8 |
| FAM14 2.5 10000 400;600GB #Dr17 | 49.7 | 21.1 | 7.2 | 17.7 |
| FAM17 2.5 10000 400;600GB #Dr17 | 34.4 | 18.5 | 8.3 | 31.5 |
| FAM16 3.5 10000 0GB #Dr0 | 1.2 | 6.6 | 1 | 22.9 |

Observations:

- No single configuration is in the top 25% for all 4 metrics.
- 11 of 16 systems have at least one metric in the top 25%

- Family 13 and 14 have (17) 10 K rpm HDDs and 7 600 GB SSD/AFAs.
  - Family 13 has a previous generation controller as compared to family 14.

# OL-4 Sequential & Transactional 10 K rpm Configs with Top 25% Highlighted

| Label | Sequential Read Workload Test | Sequential Write Workload Test | Ready Idle Workload Test (GB/W) |
|---|---|---|---|
| FAM4 2.5 10k 450GB #Dr48 | 0.99 | 0.51 | 48.75 |
| FAM5 2.5 10k 450GB #Dr54 | 1.15 | 1.84 | 32.95 |
| FAM5 2.5 10k 450GB #Dr88 | 1.55 | 1.18 | 34.32 |
| FAM7 2.5 10k 600GB #Dr50 | 4.2 | 4.27 | 45.8 |
| FAM8 2.5 10k 600GB #Dr50 | 5.34 | 4.16 | 45.2 |
| FAM2 2.5 10k 600GB #Dr75 | 4.73 | 4.18 | 44.7 |
| FAM7 2.5 10k 600GB #Dr75 | 5 | 4.21 | 51 |
| FAM8 2.5 10k 600GB #Dr75 | 5.68 | 4.66 | 54.4 |
| FAM24 2.5 10k 600GB #Dr75 | 4.67 | 4.13 | 42.5 |
| FAM2 2.5 10k 600GB #Dr125 | 5.77 | 4.25 | 52.8 |
| FAM8 2.5 10k 600GB #Dr125 | 5.93 | 3.82 | 60.1 |
| FAM24 2.5 10k 600GB #Dr125 | 5.65 | 4.27 | 53 |
| FAM2 2.5 10k 600GB #Dr200 | 6 | 2.94 | 60.2 |
| FAM24 2.5 10k 600GB #Dr200 | 5.85 | 3.1 | 60.8 |
| FAM7 2.5 10k 600GB #Dr250 | 4.82 | 3.4 | 72.7 |
| FAM44 2.5 10k 600GB #Dr528 | 1.28 | 0.65 | 55.4 |

OL-4 OPTIMAL SEQUENTIAL CONFIGURATIONS
Green: Best 25% for the category

Observations:

8 systems have one or more >25% metric.

3 of 4 top 25% for Read and Ready Idle test results correlate.

No correlation between Write and Ready Idle Test Results.

No system is Top 25% in all three metrics.

One system is top 25% in both Read and Write test results.

Note: of the 27 Sequential Submissions, 16 had 10 K drives.

| Label | Hot Band Workload Test (IOPS/W) | Random Read Workload Test (IOPS/W) | Random Write Workload Test (IOPS/W) | Ready Idle Workload Test (GB/W) |
|---|---|---|---|---|
| FAM34 2.5 10000 1000GB #Dr24 | 29.2 | 14.9 | 9.5 | 49.9 |
| FAM37 2.5 10000 1000GB #Dr24 | 27 | 13.2 | 8.9 | 47.7 |
| FAM7 2.5 10000 600GB #Dr50 | 21.6 | 3.85 | 6.9 | 45.8 |
| FAM9 2.5 10000 600GB #Dr50 | 22.5 | 6.21 | 8.13 | 45.2 |
| FAM2 2.5 10000 600GB #Dr125 | 23.48 | 12.48 | 5.67 | 52.8 |
| FAM7 2.5 10000 600GB #Dr125 | 21.83 | 2.91 | 5.11 | 65.6 |
| FAM9 2.5 10000 600GB #Dr125 | 25.12 | 13.1 | 7.04 | 60.1 |
| FAM24 2.5 10000 600GB #Dr125 | 24.91 | 12.3 | 10.03 | 53 |
| FAM19 2.5 10000 600GB #Dr150 | 11.75 | 7.85 | 14.54 | 75.98 |
| FAM2 2.5 10000 600GB #Dr175 | 25.62 | 13.81 | 5.46 | 59.3 |
| FAM7 2.5 10000 600GB #Dr250 | 19.8 | 2.25 | 4.12 | 72.7 |
| FAM24 2.5 10000 600GB #Dr250 | 26.84 | 14.63 | 7.96 | 64.6 |
| FAM9 2.5 10000 600GB #Dr275 | 23.38 | 7.9 | 5.39 | 69 |
| FAM2 2.5 10000 600GB #Dr325 | 23.48 | 9.82 | 5.85 | 66.4 |
| FAM24 2.5 10000 600GB #Dr500 | 24.03 | 16.04 | 6.89 | 71.9 |
| FAM44 2.5 10000 600GB #Dr528 | 13.67 | 10.52 | 4.5 | 55.4 |
| FAM42 2.5 10000 600GB #Dr576 | 21.49 | 16.21 | 5.71 | 63.6 |
| FAM18 2.5 10000 900GB #Dr192 | 20.63 | 19.63 | 12.39 | 38.6 |
| FAM35 3.5 10000 0GB #Dr84 | 12.8 | 15.3 | 6.3 | 126.2 |

OL-4 OPTIMAL TRANSACTIONAL CONFIGURATIONS
Green: Top 25%

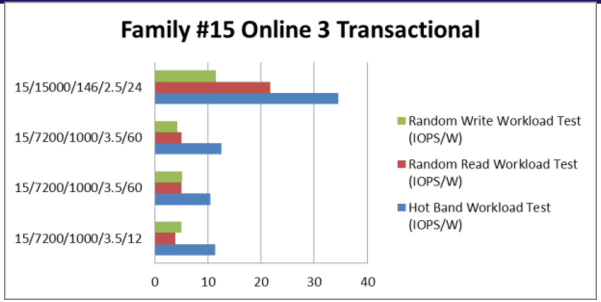
Observations:

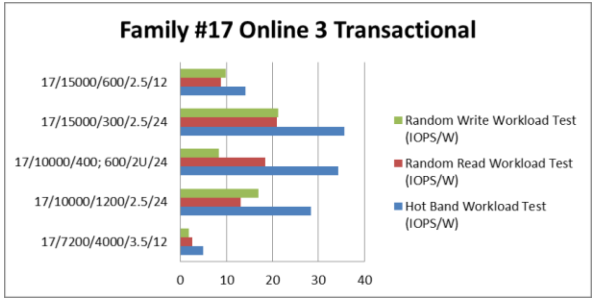
11 Systems with one or more >25% metric and there are disjointed metric correlations.

No correlation between top 25% Hot Band IOPs/W and ready idle GB/W.

Of the 35 transactional submissions, 19 had 10 K drives.

# OL-3 Family Analysis: Transactional



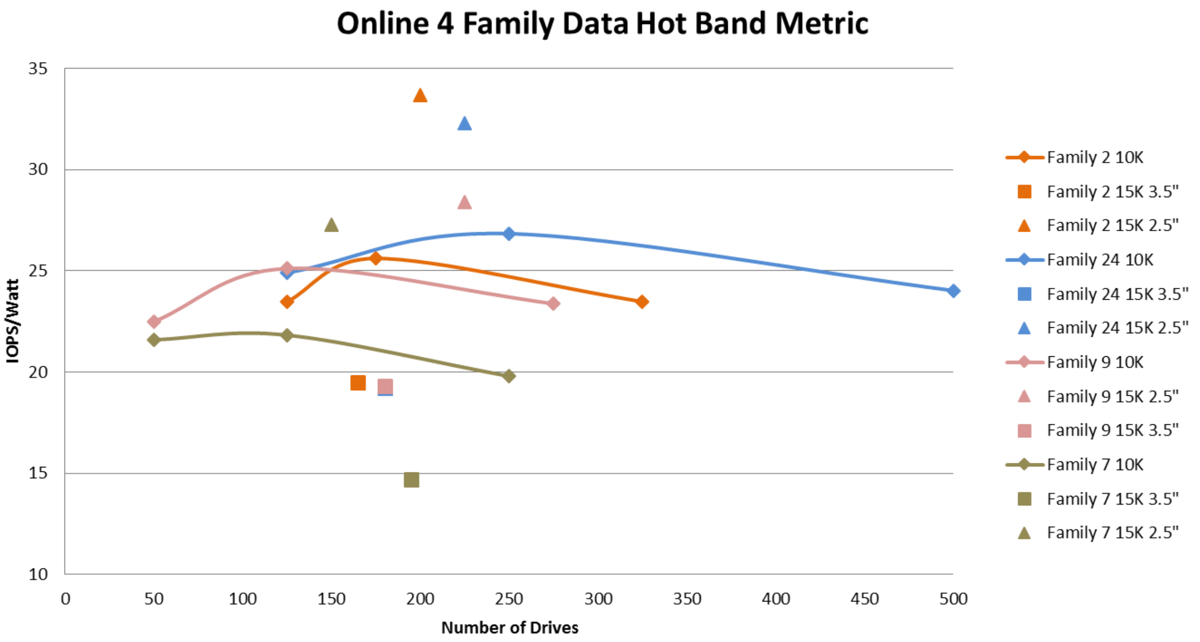




Legend Notes: Family#, rpm, FF, GB, Device count

Observations:

- Family #15: 7.2 K drives with count of 12 and 60 shows HB w/i 20%.
  - The two 60 device systems use different variations of the controller.
  - High variation likely caused by smaller number of drives to cover controller power. The other 12 drive count configuration shows the same low score.
- Family #17: 15 K drives with count of 12 and 24 shows HB outside of 20%
- The center configuration for family #17 has 7 SSD and 17 10K HDD devices.
  - The configurations with mixed SSD/HDD devices had better scores than HDD only configurations with same components. There are several other OL-3 configs with mixed SSD/HDD drives that show the same behavior.
  - The matching configuration with 24 2.5” 10 K drives is below the HDD/SSD mix.

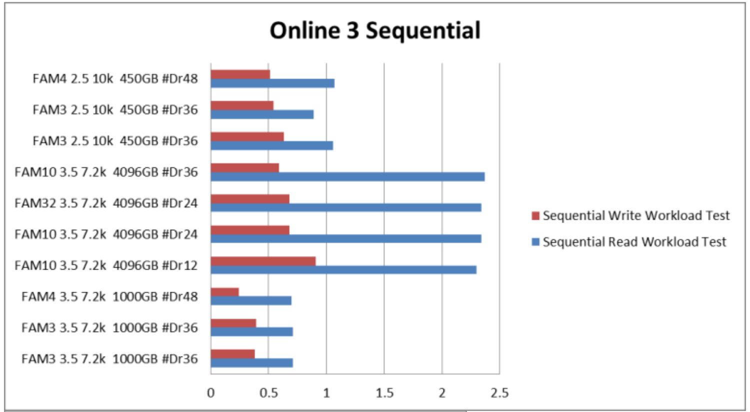
# Online 4 Family Hot Band Metric Data

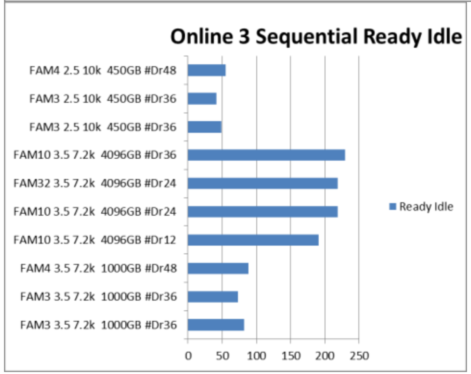



Observations:

- For the 4 families with min/opt/max configurations, it can be argued that the optimum configuration provides a reasonable assessment of the overall system performance.
  - IOPs/Watt values on either side of optimum are within 10% of the optimum value.
  - Negates need to test minimum and maximum configurations.

# OL-3 Data Analysis: Sequential



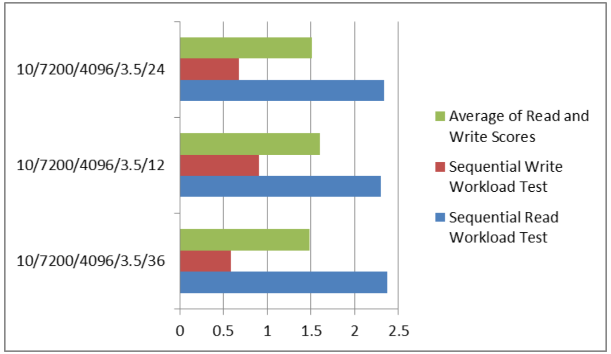




Observations:

- For family #3, the configuration with the higher scores for both device types has a single controller, the other configuration has a dual controller.
- Family #4 has a single controller for both device types.
- 3.5” configurations have better ready idle (GB/W) characteristics than 2.5” characteristics.
- Higher capacity drives have better ready idle characteristics and mibs/W results than lower capacity drives.

Legend Notes: Family#, FF, rpm, GB, Device count

# OL-3 Family Analysis: Sequential




Family #10 On-Line Sequential

- Read test is w/i 20% for performance scores, write tests are not.
- Average Read/Write score decreases from 12 to 36 drives:
  - 12 drives is optimal
  - 24 and 36 drives are 2 maximum points, both within 20% of the optimal score.
    - 24 drives are within 7.7%
    - 36 drives are within 7.8%

Legend Notes: Family#, rpm, FF, GB, Device count

# General Observations

- Idle is not a good indicator of operational energy efficiency.
  - Tested systems range 2% to 20% difference between maximum and idle power where that data is available.
  - Impact of drive spinning (HDDs) and data clean-up protocols (SSD and HDD) keep idle power demand high.

- Assessment by taxonomy and workload type is biased:
  - to the higher performance drives for transactional workloads,
  - to the higher capacity drives for sequential workloads.

- 7.2 K, high capacity drives are the most energy efficient but have lower performance for transactional workloads.
  - Performance per watt will be lower than faster drives, but the capacity per watt will be up to an order of magnitude higher.
  - The mechanics of the test, which require full prep of all the drives, makes testing large 7.2 K high capacity systems uneconomic – it takes over a week to do drive prep on large drive count systems.

# General Observations

- The number of manufacturers and configurations by category is relatively small.
  - Difficult to evaluate relationships between different component types within the database.

- The relative product capabilities and configuration size of OL-3 to OL-4 and OL-4 to OL-5 products necessitate engineering judgement and interpretation to categorize a product.

- When comparing systems with the same drive size and speed, the following system attributes will influence the magnitude of the score:
  - Quantity of working memory and cache on the controller
    - In some cases, a manufacturer will combine cache and working memory,
    - In some cases, a manufacturer may use drives for cache, and
    - In other cases, they are managed separately but perhaps with overlap.
  - Total number of drives
  - Number of servers pushing the data and number of front-end pipes
  - The connection types between the servers/controllers/drives
  - Controller architecture (technology generation, cpu number and types, data movement capabilities, back-ends)

# General Observations (cont.)

- Controller power and the number of supported drives affect scores:
  - Higher controller functionality increases power demand and may impact scores depending on the number and type of drives attached to the controller.
  - OL- 2 and small OL-3 partial or single drawer systems carry heavy power debt
  - Impact of controller power will be reduced on a scale up system tested with multiple drawers.

- Drive counts below and above the optimal point serve different purposes:
  - Up to the Optimal point, you are assessing the product on performance per watt
    - You will configure a storage product to enable growth to the optimum point and beyond.
  - Beyond the Optimal point, you are assessing on capacity per watt.
    - As you move beyond the optimal point, you are assessing capacity and TCO to determine when to add an additional storage product.
  - In general there is reasonably comparable performance/watt values across the min/opt/max configuration set in a family.

# General Observations (cont.)

- Architectural and Configuration Differences can make a significant difference in scores:
  - Data discussion shows significantly different IOPs/W scores for similar configurations driven by controller technology generation or architectural differences.
  - Different architectures are used to address specific customer needs; may require separate categorization.

- Some Flash/SSD drives currently have a detailed, high priority clean-up cycle which may limit performance at higher capacities and drive counts and increase idle power. May not be universal, but is a significant impact on some systems.
  - This illustrates the impact of general data housekeeping on all storage systems.

# General Observations: ENERGY STAR Testing

- Testing is complicated, time consuming, and expensive:
  - Identification of the optimal point is difficult
    - Optimization is as much an art as it is a science.
    - Performance scores can vary by a factor of 2 to 4 based on the “excellence” of the tech setting up the system.
  - Loading of storage registers on large systems takes days
  - Single tests on large systems are taking a month or more.

- Storage testing is inherently difficult due to:
  - The inherent complexity of the systems.
  - The dependency of the performance on the choice and set-up of the system:
    - Matching storage allocation to servers (resetting configurations with different drive counts and workloads)
    - I/O capacity of servers and switches in test rig
    - Identifying optimal thread count on servers running workload

- Emerald is a workable energy efficiency test: TGG/SNIA recommends that only the optimal configuration be tested and reported.
  - The optimal configuration is sufficiently representative of the overall product capabilities.

# General Observations: Transaction Tests

- All transaction tests were optimized for Hot Band workload
- Read and Write tests are not optimized and not representative of an “optimum configuration”
  - Reported Read and Write scores should not be evaluated
- Hot Band (HB) scores are highest on the higher speed/lower capacity drives in the same Form Factor (FF)
  - 2.5” FF will have better scores than the 3.5” FF
    - Power scales with rotational speed and size
    - At all speed and capacity points, 2.5” drives have preferable performance/power characteristics as compared to 3.5” drives.
  - Drive capacity has variable affects depending on design and technology.
  - SSDs offer significantly better IOPs/Watt based on available data, but lower ready idle readings (GB/W)
- Reporting of Cache capacity, VDBench version, and configuration in the ENERGY STAR database was inconsistent, making it difficult to analyze differences in the HB data.

# General Observations: Sequential Tests

- The Sequential Test is Optimized across the read/write functionality
  - The optimized ENERGY STAR score is the average of the read and write scores.
  - Optimum configuration set-up is balanced to maximize the two scores.
  - Getting the balance of the average R/W score to get min/max performance/power within 20% of optimal is very difficult/time consuming.

- Operations are moving to more reads than writes with time.
  - SNIA will watch the market to determine how storage use is changing.
  - As we continue to collect data, SNIA can evaluate different weightings using the data.

# COMS Discussion

Herb Tanzer and Chuck Paridon

# V2 Storage Recommendations

# Review of V1 Challenges

- Data center storage systems, particularly OL-3 and OL-4 systems, are highly configurable
  - Adequate characterization across the spectrum of configuration variables is expensive in terms of equipment, time and human resources
  - V1-specified test configurations do not reflect the actual systems purchased by our customers

- Costs of meeting V1 testing result in
  - Qualifications of only a subset of systems sold
  - Smaller vendors opting out of testing due to cost-benefit concerns
  - Larger vendors opting out of testing for newer technologies because of cost and limited availability, poor ROI

- The data discussion shows that setting thresholds presents challenges due to the product technical complexity and small data sets available in each OL category.

# Criteria and Testing Recommendations

- Recommended ENERGY STAR V2 Certification Criteria
  - PSUs must meet “GOLD” or better efficiency criteria
  - Require Reporting Emerald Test results for workload(s) of choice.
  - The number of COMs required in category OL-3 and OL-4 should increase by 1
  - Inlet temperature reporting should be mandatory
  - Timestamping on temperature and power readings should be mandatory
  - Products should be qualified to ASHRAE A2 or higher

- Require only 1 Emerald data set per test workload: optimal point for 1 drive type
  - Data shows the performance/watt characteristics of drives are predictable and can be derived from a single type
  - Data shows that system performance/watt scales up as drive count reaches optimal; drop-off beyond optimal stays within envelope identified in V1

# Conclusions

- Storage Products are complex – active efficiency and ready idle thresholds will be extremely difficult, if not impossible, to implement.

- The use of COMs will increase the ready idle metric by reducing the physical footprint of storage system and/or drive count which lowers data center energy consumption, but:
  - The COMs activity reduces the overall I/O rate of the storage system, reducing the active efficiency metric.
  - Have to balance the benefits of a reduced product footprint versus the “efficiency” of the individual storage product.

- IOPs/W and mibs/W scores are highly dependent on:
  - On-line category: resiliency and scale
  - Controller and power supply redundancy
  - Drive type (HDD or SSD/AFA), Form Factor, speed and count.
  - Controller Architecture
  - Technology Generation

- Idle and Ready Idle Values, by themselves, are a poor indicator of the energy efficiency of a on-line storage product.

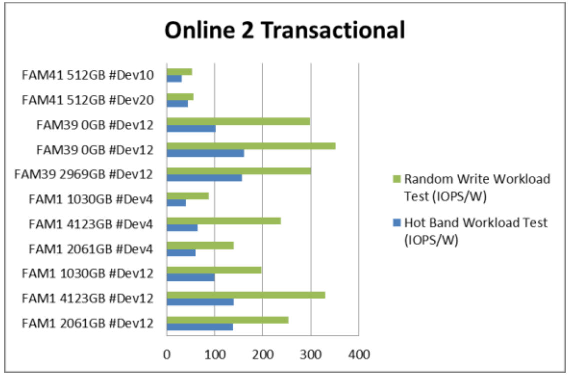
- Customer needs drive the diversity of configurations and capabilities. It is important not to limit the available architectures/configurations

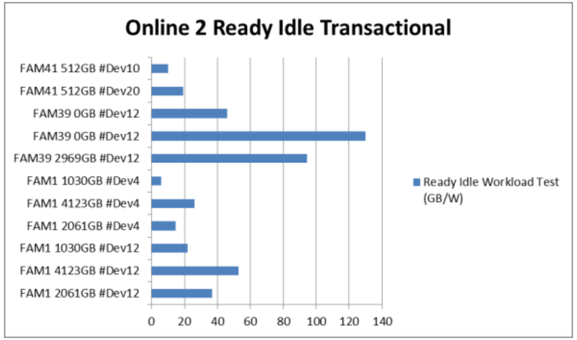
# Additional Data Analysis

All configurations shown are Optimal except for the family charts.

These data charts largely correlate the results presented from slides 7 to 19 and support the general observations detailed in slides 20 to 26.

# OL-2 Data Analysis: Transaction



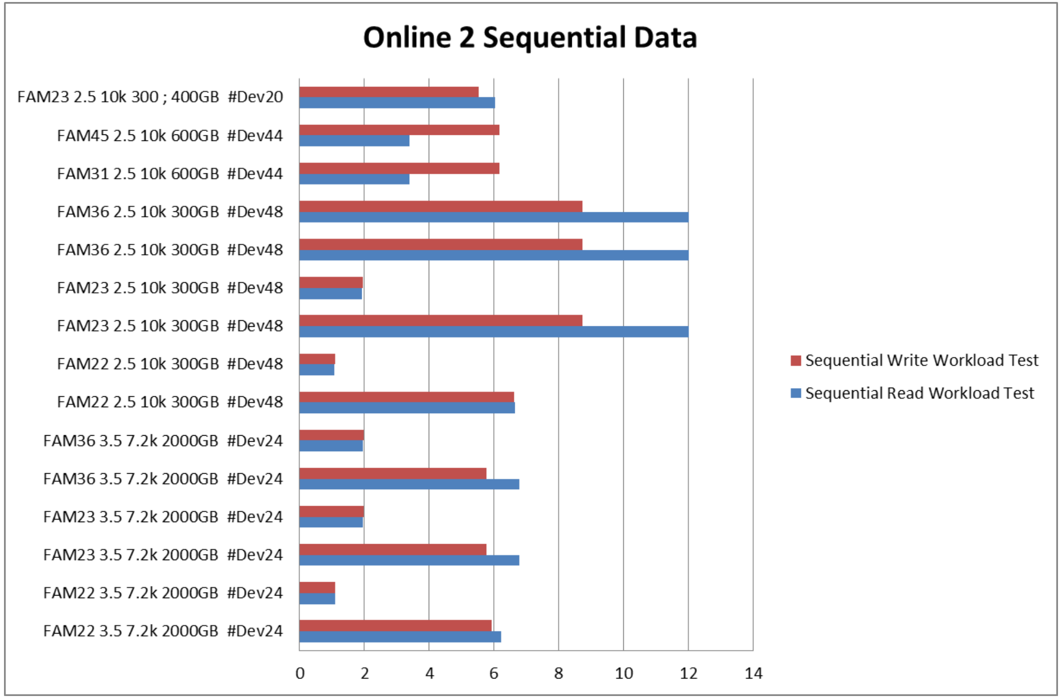

Observations:

- For flash drives, the HB and Random Write IOPS/W score improve with increased Flash capacity.
    - The increase from IOPS/W increase is large from 1 – 2 TB, smaller from 2-4 TB
    - At 12 drives Random Read scores improve from 1-2 TB and degrade to 4 GB when compared to both 1 and 2 TB scores.
    - Low drive counts have lower IOPS/W scores due to larger impact on controller power on the per watt score.

- IOPS/W scores on products with higher drive counts on higher capacity drives may be slightly lower (2 TB vs. 4 TB) because of “garbage collection” or clean-up that occurs on a milli-second cycle time.
- Quantity of Flash/SSD data is insufficient to draw broad conclusions, though Solid State devices performed better than uncached spinning devices.

Notes:
Random Read Data excluded to improve clarity of the HB score differences.
Legend is Family#/Device GB/Device Count
All configurations are Flash/SSD products.

# OL-2 Data Analysis: Sequential

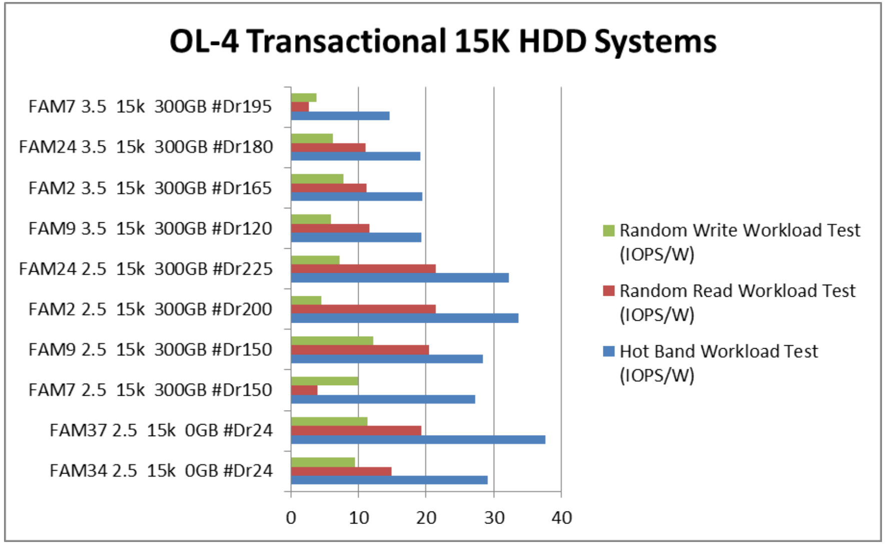


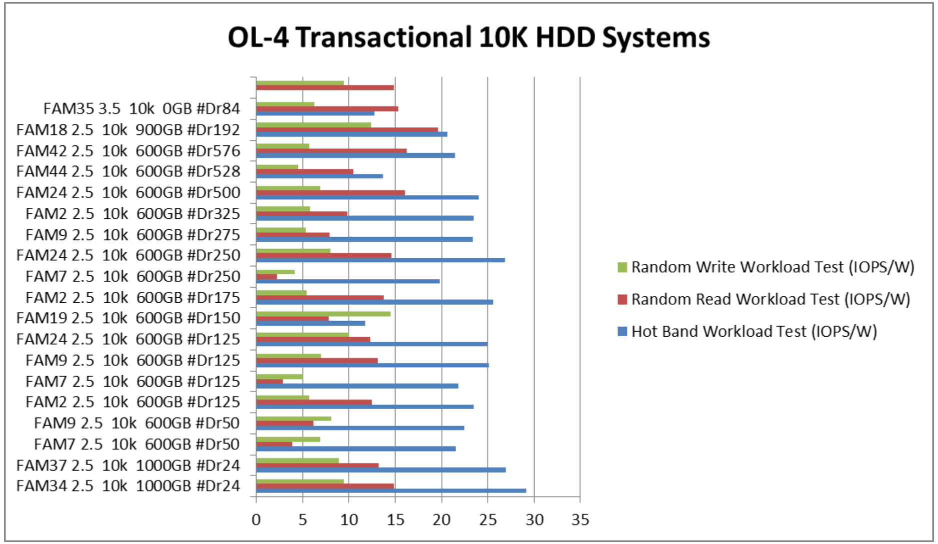

Legend Notes: Family#, FF, rpm, GB, Device count
All scores are IOPs/Watt

Observations:

- IOPs per watt scores are higher on systems with higher disk counts due to better amoritization of power.
- For all families, the higher scoring systems where configurations are comparable have a newer generation controller.
- The differences between the 3.5”/7.2K and 2.5”/10K configs are driven in part by architectural and technology generation differences.

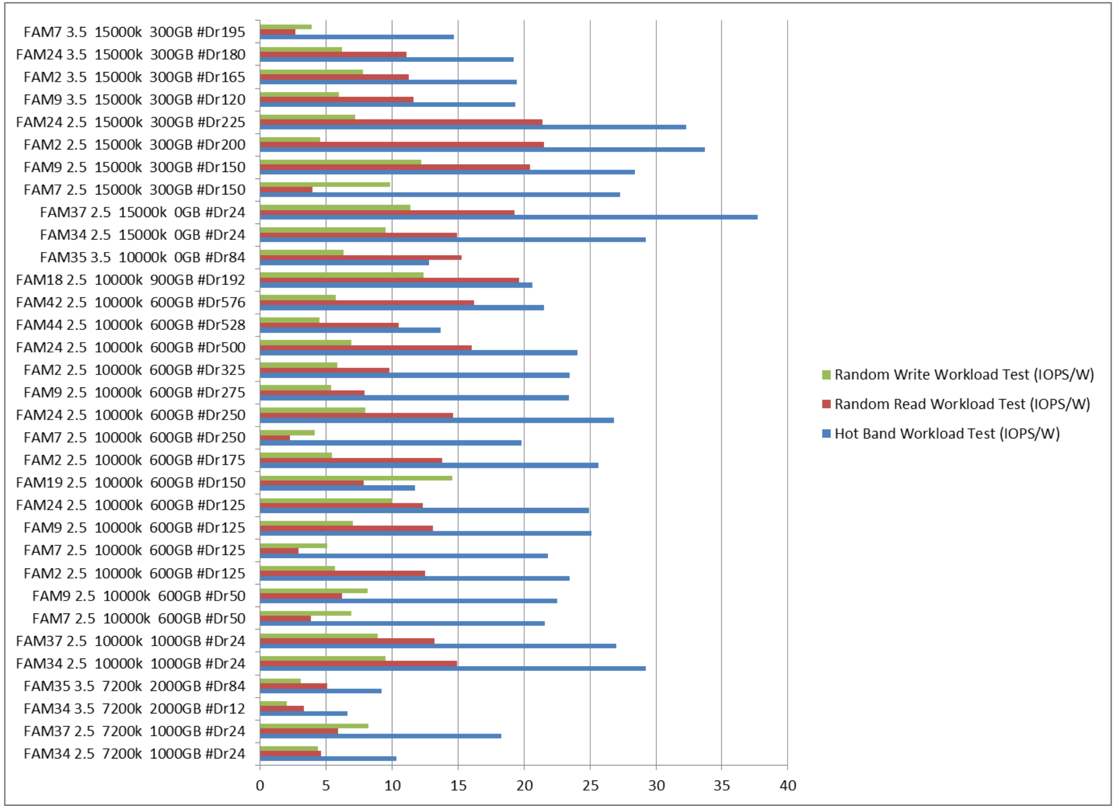
# OL-4 Data Analysis: Transactional



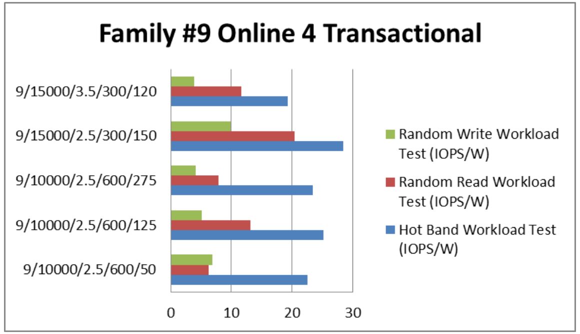

- 2.5” 15 K HDD has higher Hot Band efficiency scores than a 3.5” 15K HDD due to requiring less power for same data movement.
- Higher rpm drives within the same from factor (FF) have better transactional scores. Due to faster transfer rates for same power draw.

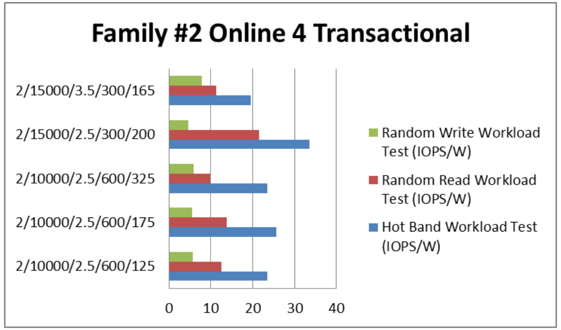
# OL-4 Transactional Data

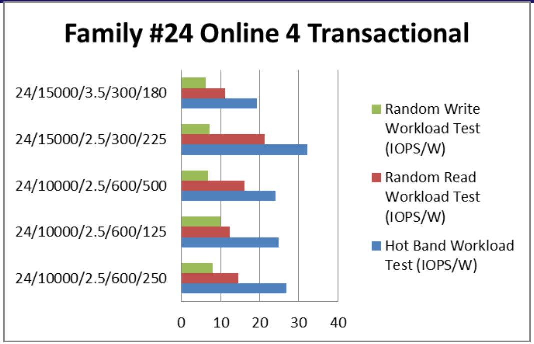


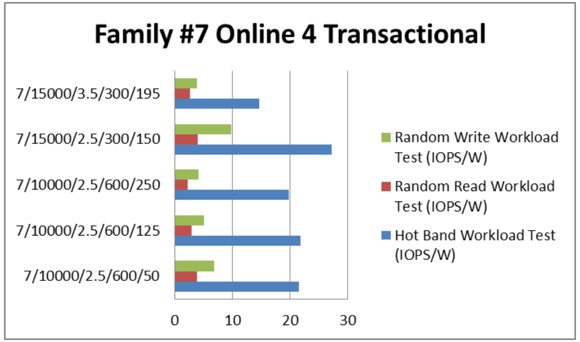

- In general, 15 K rpm HDDs have higher IOPs/W scores as compared to 10 K rpm HDDs and both have higher scores than the 7.2 K rpm HDDs.
  - It is important to know which systems are optimum and which are minimum and maximum as scores will differ.
- 2.5” HDD has higher transactional efficiency scores than a 3.5” HDD due to requiring less power for same data movement.
- Higher rpm drives within the same from factor (FF) have better transactional scores due to faster transfer rates for same power draw.

# OL-4 Family Analysis: Transactional



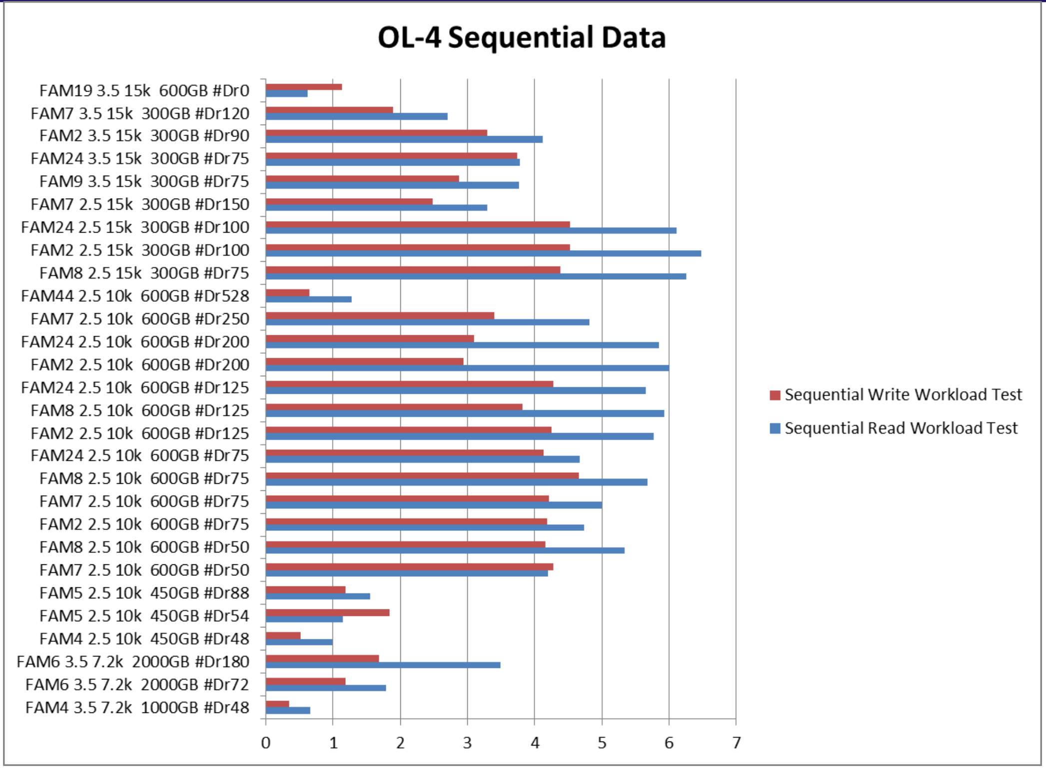




- For the 4 families, the HB min and max configs are within 20% of optimum.
- For the 4 families, the HB scores are reasonably close, but there is some differentiation.
- These 4 families, by design, span a large portion of the OL-4 space. The goal is to focus on different customer cost points based on total capacity and maximizing total performance for that capacity.
- Maximizing the qualified range gives customers the best value but requires at least 3x the effort to find the min and max points that maximize the qualified drive count range.

Legend Notes: Family#, rpm, FF, GB, Device count
Min/Opt/Max families are systems 5 to 3 on each graphic.

# OL-4 Data Analysis: Sequential

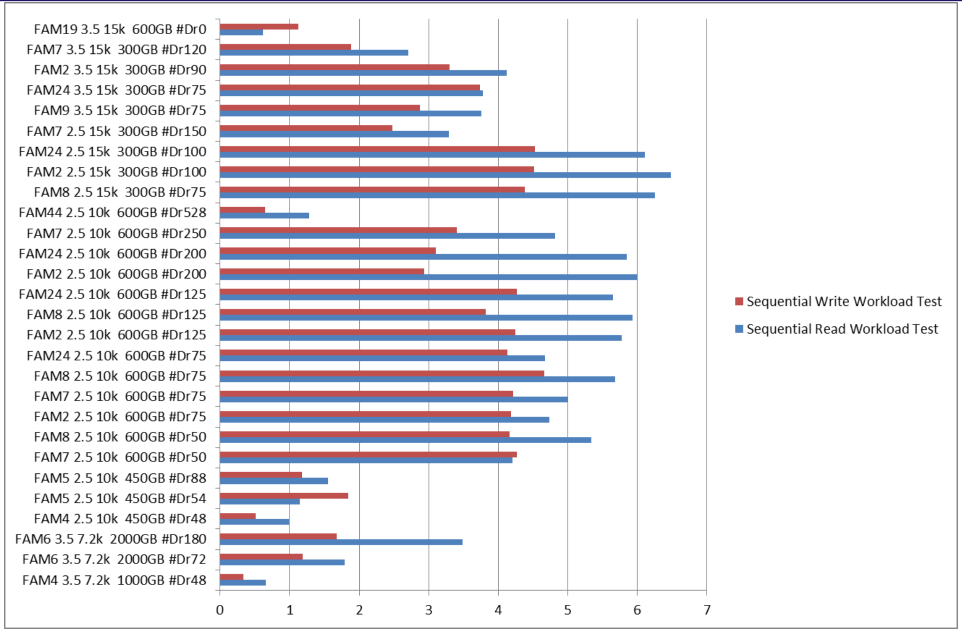


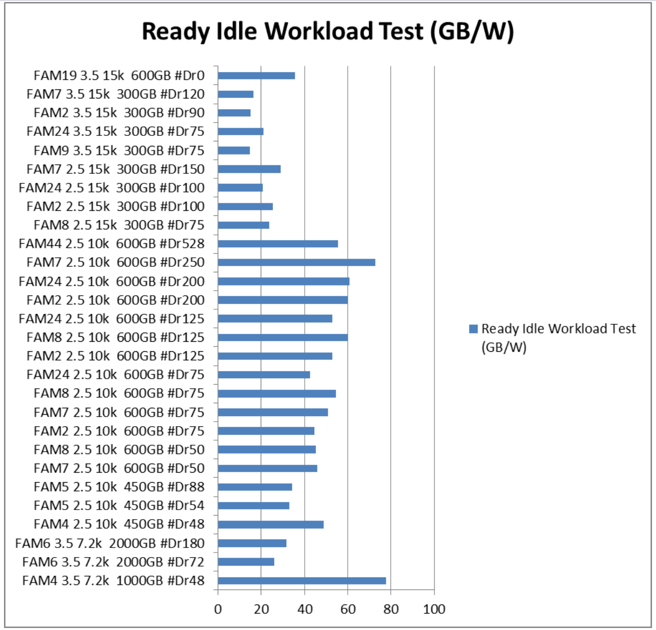

Legend Notes: Family#, FF, rpm, GB, Device count

Observations:

- For 2.5” drives, 15 K drives give better scores than 10 K drives
- 10 K 2.5 drives have two distinct groupings of scores. Likely due to technology generations.

# OL-4 Sequential Read/Write and Ready Idle scores

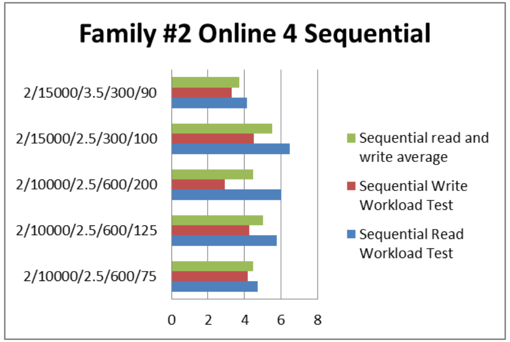


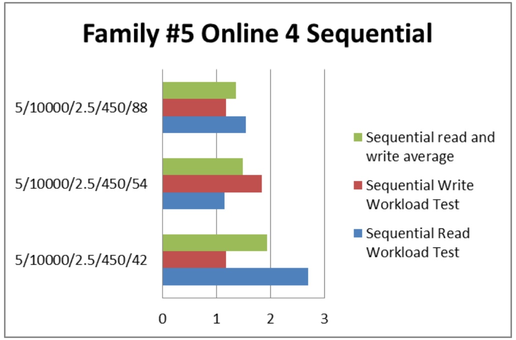



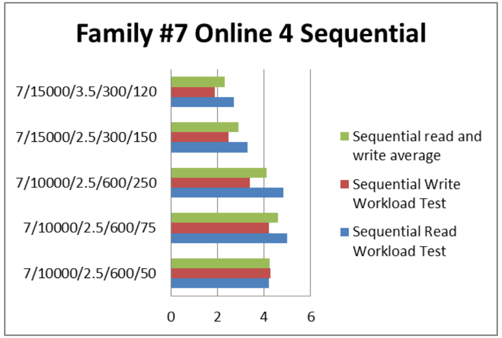

Observations:

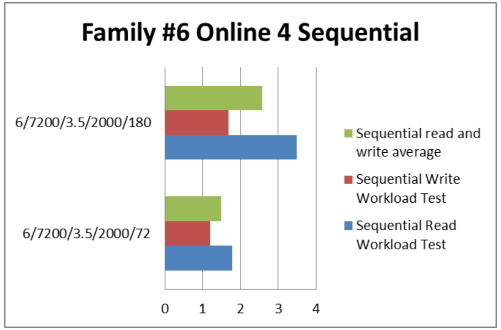
- Higher drive systems can increase the GB/W, even at lower capacity drives, due to better amorization of power use.

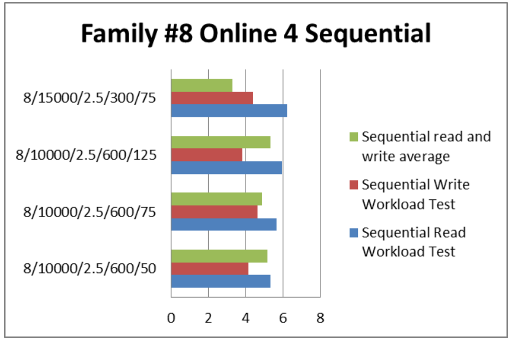
# OL-4 Family Analysis: Sequential

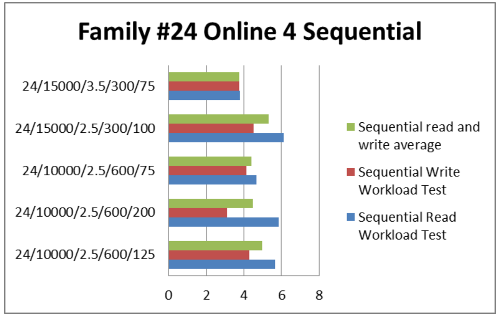


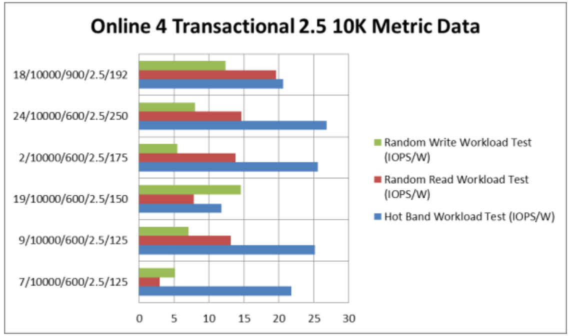



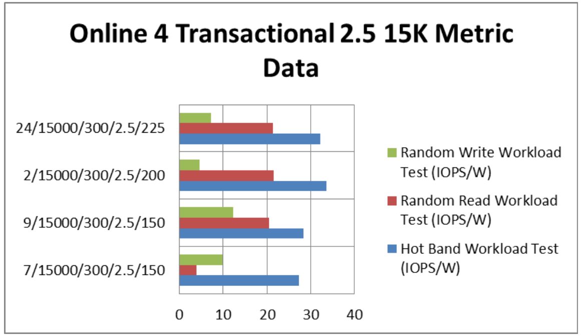

Observations:

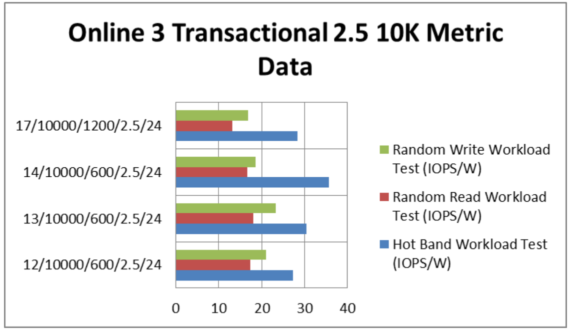
- Family #6 is a scale out with a controller for every 12 drives.
- In general, conclusions for sequential systems match transaction systems.

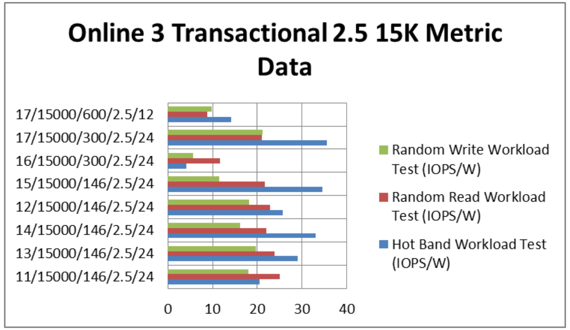
Legend Notes: Family#, rpm, FF, GB, Device count

# OL-3 to OL-4 Comparison: Transactional



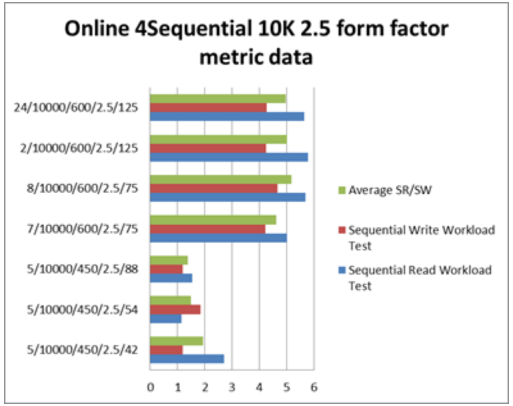



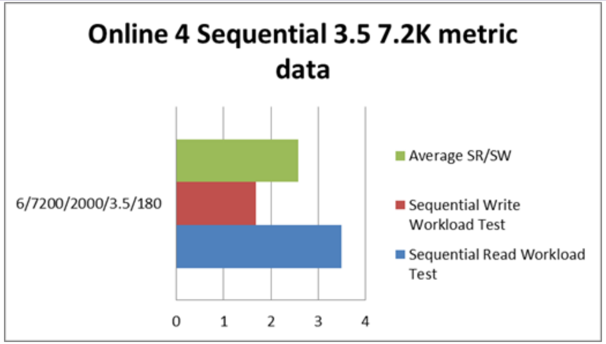



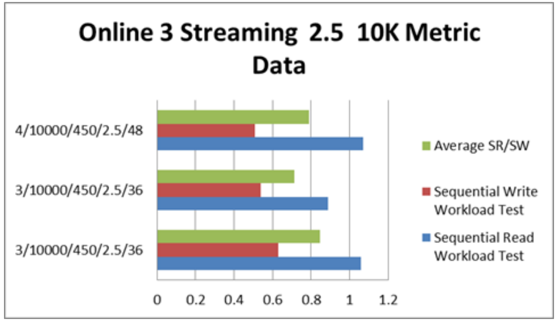

- Many OL-3 systems are single controller, while all OL-4 are dual
  - OL-4s draw more controller power
- 15k rpm drives consistently deliver more iops/watt than 10k drives in a form factor regardless of OL type
- For larger data sets and/or improved resilience, OL-4 systems provide more total performance for a small increase in power.
- 1 OL-4 system can handle as much storage as 8 OL-3 systems, resulting in lower overall costs and power use.
- Many applications cannot spread data across multiple storage systems, necessitating use of larger systems or systems that distribute the data for the application.

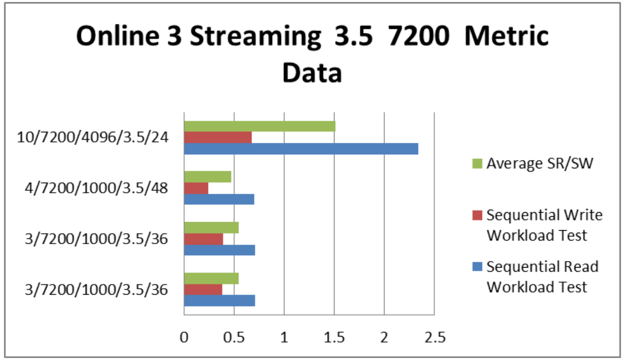
Legend Notes: Family#, rpm, FF, GB, Device count

# OL-3 and OL-4 Comparison: Sequential




The OL-4 systems are showing better GBS/W performance compared to OL-3 systems with comparable drive counts. This could be caused by:

- more back-end buses for parallel data delivery
- stronger data movement and CPU controller components
- more front-end pipes

Legend Notes: Family#, rpm, FF, GB, Device count

# References

- Link to SNIA Emerald training page
- Emerald Training Introduction to COMs.pdf
- Storage Considerations in Data Center Design
- Green Storage Technologies 2 by Alan Yoder