Value-Perspective: Quantifying Energy Efficiency Gains with BenchSEE - A Comparative Study of Optimisation Techniques

Leon Tabaro, Ahmed M. A. Sayed, Mona Jaber

School of EECS, Queen Mary University of London, UK

1. Introduction

Rapid growth in compute demand, driven by AI training and inference, large-scale analytics, and cloud services, has made server energy efficiency (EE) a primary constraint alongside traditional performance considerations. Data-centre operators increasingly face power delivery limits, cooling capacity constraints, and cost/sustainability pressures that require measurable, comparable, and repeatable efficiency metrics. In this context, EE benchmarking provides a controlled method to relate *work performed* to *energy consumed*, enabling evidence-based procurement decisions and configuration optimisation.

This report presents a controlled evaluation of BenchSEE through a set of systematic power and resource interventions applied to a system under test (SUT). The focus is to assess whether BenchSEE can objectively detect and quantify energy-efficiency improvements (and regressions) resulting from known operating-point changes, using repeatable experiments across the same workload suite.

1.1. Background: BenchSEE

The Benchmark of Server Energy Efficiency (BenchSEE) is a benchmarking framework developed by the Branch of Resource and Environment of the China National Institute of Standardisation (CNIS) for measuring and evaluating server energy efficiency. BenchSEE evaluates energy efficiency across key server subsystems; CPU, memory, and storage, using pre-configured workloads and standardized procedures. It collects performance and power data during workload execution and produces automatic reports, including workload-level metrics and aggregated component-level and overall scores. BenchSEE supports multiple architectures (e.g., ARM, x86_64) and common server operating systems.

BenchSEE reporting includes workload-level raw performance and performance–power ratio (Perf/W) views, as well as aggregated component and total metrics. The total score is derived from CPU/memory/storage results using weighted aggregation, enabling both subsystem analysis and a single overall indicator for comparison.

1.2. Objectives of this report (D2)

This report evaluates BenchSEE from a results/value perspective by applying controlled energy optimisation techniques and determining whether BenchSEE can produce a consistent and interpretable assessment of energy-efficiency change. Specifically, the objectives are:

- **O1 – Sensitivity:** Determine whether BenchSEE detects expected changes in energy-efficiency metrics under common optimisation interventions (power policy changes, frequency caps, and core-count restrictions).

- **O2 – Consistency and repeatability:** Quantify the repeatability of BenchSEE metrics under a baseline condition and confirm that intervention effects are distinguishable from baseline variability.
- **O3 – Interpretability:** Analyse changes in key outputs (raw performance, Perf/W, and EE scores) to characterise trade-offs (efficiency gains vs performance loss) and identify which techniques yield the most favourable outcomes.
- **O4 – Workload discrimination:** Confirm that BenchSEE differentiates effects across workload classes (CPU-/memory-bound vs storage/IO-bound), consistent with the mechanism of each intervention.

2. Comparison to the SUT used in previous User Report

To strengthen external validity, Deliverable 2 (D2) uses a different, higher-performance workstation than the platform used in Deliverable 1 (D1). The primary motivation is to evaluate BenchSEE on a modern heterogeneous-core CPU platform with substantially increased memory and storage capacity, and to assess whether controlled power/CPU configuration changes are reflected consistently in BenchSEE outputs.

Previous SUT (D1)

- **Model:** Dell Precision Tower 3620 (tower server/workstation)
- **CPU:** Intel Core i7-7700 (4 physical cores / 8 logical threads)
- **Memory:** 16 GB DDR4 (single DIMM)
- **OS:** Ubuntu 24.04.x LTS (Linux)

Current SUT (D2)

- **Model:** Custom-build Intel 2790 Workstation (tower server/workstation)
- **CPU:** Intel Core i9-14900K (24 physical cores / 32 logical threads; heterogeneous P/E core design)
- **Memory:** 192 GB DDR5 (4×48 GB)
- **OS:** Ubuntu 24.04.2 LTS (Linux)

2.1 Implications for benchmarking

Relative to the D1 platform, the D2 SUT introduces:

- **Greater compute headroom and parallelism** (24C/32T vs 4C/8T), which can change workload scaling behaviour.
- **Modern P/E-core heterogeneity**, which makes CPU frequency and scheduling policies especially relevant to efficiency outcomes.
- **Much larger memory capacity and bandwidth**, which can affect memory- and cache-sensitive workloads and reduce paging-related noise.

This report therefore treats the D2 platform as both a stronger performance baseline and a more realistic target for modern server-class optimisation behaviour.

3. Experimental design and methodology

3.1 Overall approach

A baseline condition was established and then compared against a set of controlled interventions (energy optimisation techniques). All scenarios were executed using the same BenchSEE workload suite and configuration to ensure comparability. Each scenario consists of multiple repeated runs to estimate variability and reduce the influence of transient system effects.

3.2 Scenarios

The evaluation comprises five scenario groups:

- **S1 – Baseline (no OS optimisations):** 10 repeated runs across standard BenchSEE workloads and load levels.
- **S2 – Power policy: power-saver profile:** 3 repeated runs to assess the impact of an OS-level low-power profile.
- **S3 – CPU max-frequency cap (80%):** 3 repeated runs using a per-core cap computed as 80% of each logical CPU's hardware maximum frequency (P/E-aware).
- **S4 – CPU max-frequency cap (60%):** 3 repeated runs using the same per-core method at 60% of each core's maximum.
- **S5 – Core-count sweeps:** 3 repeated runs for each configuration with only **N active logical CPUs**, where **N ∈ {12, 8}**, to evaluate the effect of constrained parallelism.

3.3 Metrics collected and analysis method

BenchSEE outputs were analysed at three levels:

1. **Workload level:** raw performance scores, Perf/W ratios, and EE scores per workload and load level.
2. **Component level:** CPU, memory, and storage Perf/W comparisons across runs.
3. **Overall level:** total Perf/W and aggregated results.

For baseline repeatability, variability is summarised using standard deviation and coefficient of variation (%CoV) across repeated runs, defined as %CoV = 100 × (standard deviation / mean) for each (workload, load-level). Intervention effects are expressed as percentage deltas relative to baseline, with interpretation focused on whether changes exceed baseline variability and whether trends match the expected mechanism of each optimisation.

Report structure

The remainder of this report presents: (i) baseline repeatability and reference performance, (ii) comparative results for each intervention scenario (S2–S5) including trade-off analysis, and (iii) a consolidated discussion on BenchSEE validity - highlighting sensitivity to controlled changes, workload discrimination, and practical conclusions for energy-efficiency evaluation and optimisation.

4. Baseline repeatability and reference performance (S1)

A baseline condition (S1) was established to (i) quantify BenchSEE run-to-run variability on the D2 SUT, and (ii) provide a stable reference point for evaluating the magnitude and direction of changes introduced by energy optimisation interventions (S2–S5). The baseline consists of 10 repeated runs using identical BenchSEE workload configuration and system settings. Full results breakdown can be found in **Appendix A**.

4.1 Baseline repeatability: component-level Perf/W (Appendix A.1)

BenchSEE reports component-level performance–power ratio (Perf/W) for **CPU, memory, storage**, and an aggregated **total Perf/W**. Across 10 baseline runs, average values and variability were:

- **CPU Perf/W:** mean **266.7**, SD **9.27**, **%CoV 3.48%**
- **Memory Perf/W:** mean **39.0**, SD **0.86**, **%CoV 2.22%**
- **Storage Perf/W:** mean **14118.5**, SD **970.3**, **%CoV 6.87%**
- **Total Perf/W:** mean **182.7**, SD **5.88**, **%CoV 3.22%**

Overall, CPU and memory Perf/W exhibit **moderate repeatability** (≈2–3.5% CoV), while storage shows **higher relative dispersion** (≈6.9% CoV). The elevated storage variability is consistent with typical IO sensitivity to cache state, background services, and device/controller behaviour, and it motivates interpreting storage-only deltas with wider uncertainty margins than CPU/memory deltas.

4.2 Baseline repeatability: workload-level EE scores (Appendix A.2)

At the workload level, baseline repeatability was assessed across **40 (workload, load-level)** points (CPU, memory, and storage workloads at multiple load levels). Coefficient of variation (%CoV) was computed per point to characterize the noise floor that intervention effects must exceed to be considered practically meaningful.

Across all points:

- **Mean %CoV: 4.74%**
- **Median %CoV: 4.13%**
- **Distribution: 12/40** points ≤ 3% CoV, **14/40** points in 3–5%, and **14/40** points > 5%.

This indicates that many workload points are stable, but a non-trivial subset exhibits higher dispersion. For intervention analysis later in the report, this baseline variability is used to distinguish systematic effects from ordinary run-to-run fluctuation.

Points with highest baseline variability

The most variable baseline points include:

- **SORT 50%** (~16.9% CoV)
- **OLTP 62.5%** (~11.0% CoV)
- **SORT 25%** (~10.2% CoV)
- **LU 50%** (~9.3% CoV)

These are treated as lower-confidence indicators for small intervention effects (i.e., small deltas here may be within baseline noise).

Points with strongest baseline stability

Several points were highly stable and are treated as anchor indicators:

- **CACHE (low)** (~0% CoV)
- **COMPRESS 75%** (~2.1% CoV)
- **AES 100%** (~2.2% CoV)
- **SOR 100%** (~2.3% CoV)
- **OLTP 75%** (~1.35% CoV)

4.3 Baseline drift observation (Appendix A.1)

A notable feature of the baseline runs is a **systematic uplift** in run 9–10 relative to runs 1–8 across multiple component Perf/W metrics (CPU, memory, storage, and total). This suggests the presence of a time-dependent factor (e.g. background workload differences, OS governor behaviour, or platform boosting/residency effects). Because this drift is coherent across metrics, the baseline is treated as a **distribution** rather than a single point estimate, and later intervention effects are interpreted relative to baseline variability rather than absolute single-run comparisons.

4.4 Reference performance summary

For subsequent sections, **S1 baseline mean** values serve as the reference performance/efficiency level on this SUT:

- **Total Perf/W: 182.7**
- **CPU Perf/W: 266.7**
- **Memory Perf/W: 39.0**
- **Storage Perf/W: 14118.5**

Interventions (S2–S5) are therefore reported primarily as **percentage deltas vs S1**, alongside consistency checks (directional expectations and stability across repetitions).

5. Comparative results by intervention scenario (S2–S5)

5.1 Comparison approach

Each intervention scenario (S2–S5) is compared against the baseline (S1) using repeated runs. Results are presented at:

- **Component level:** CPU, memory, storage, and total performance–power ratio (Perf/W)
- **Workload level:** raw performance scores and workload Perf/W where available
- **Trade-off lens:** efficiency gains are interpreted alongside throughput changes to separate "true efficiency improvement" from "efficiency increase caused by disproportionate performance loss."

Unless otherwise stated, percentage deltas (Δ%) are computed relative to S1 baseline means.

5.2 S2 – OS power policy: power-saver profile (Appendix B)

5.2.1 Component-level changes (Perf/W)

Power-saver produced a clear uplift in CPU efficiency and total efficiency, while storage efficiency decreased:

- **CPU Perf/W: +12.1%**
- **Memory Perf/W: ~0%** (negligible change)
- **Storage Perf/W: -6.8%**
- **Total Perf/W: +7.3%**

This indicates that the power-saver profile improves overall efficiency primarily through CPU operating-point changes, but may reduce storage-path efficiency on this platform.

5.2.2 Workload-level trade-offs

- **CPU-centric workloads:** raw performance remained ~flat (e.g., AES/OLTP close to baseline), while workload Perf/W increased substantially on several points (e.g., OLTP mid-load levels).
- **Storage workloads:** both raw performance and Perf/W declined for RANDOM/SEQUENTIAL, consistent with the component-level storage Perf/W reduction.

Interpretation: S2 demonstrates that BenchSEE can detect a plausible system-level efficiency shift and also discriminate that the improvement is not uniform across subsystems (CPU improves, storage degrades). This supports objective opinion behaviour rather than a uniformly optimistic score change.

5.3 S3 – CPU max-frequency cap at 80% (per-core, P/E-aware) (Appendix C)

5.3.1 Component-level changes (Perf/W)

Capping maximum CPU frequency at 80% produced a larger net efficiency improvement than S2:

- **CPU Perf/W: +12.4%**
- **Memory Perf/W: +25.8%**
- **Storage Perf/W: ~0%**
- **Total Perf/W: +15.5%**

The near-neutral storage response strengthens causal interpretability: the intervention targets CPU frequency, and the most pronounced effects appear in CPU/memory metrics.

6.3.2 Trade-off analysis: efficiency vs throughput

Workload-level comparisons show the expected pattern of efficiency up, throughput slightly down:

- Several CPU/memory workloads exhibited **large Perf/W gains** with **minimal raw performance change** (e.g., AES, OLTP, STREAM).
- A minority of compute-heavy workloads showed **more noticeable raw performance reductions** while still improving Perf/W (e.g., COMPRESS).

Interpretation: S3 provides strong validity evidence because (i) efficiency gains are large and consistent across runs, and (ii) the benchmark response is workload-dependent in a way that matches the mechanism of the intervention (CPU cap affects CPU/memory workloads strongly, storage weakly).

5.4 S4 – CPU max-frequency cap at 60% (per-core, P/E-aware) (Appendix D)

5.4.1 Component-level changes (Perf/W)

S4 produced the strongest total efficiency improvement of all optimisation scenarios:

- **CPU Perf/W: +16.1%**
- **Memory Perf/W: +52.8%**
- **Storage Perf/W: +2.1%**
- **Total Perf/W: +25.3%**

Within-scenario repeatability was extremely high - very small run-to-run spread across the three runs, indicating the intervention produces a stable and reproducible shift.

5.4.2 Trade-off analysis

S4 also makes the trade-off structure explicit:

- **Favourable cases:** Some workloads improved Perf/W strongly with modest throughput impact (e.g., AES, STREAM).
- **Costly cases:** Some compute-heavy workloads exhibited a substantial throughput reduction even though Perf/W increased (e.g., COMPRESS), indicating a performance sacrifice route to improved efficiency.

Interpretation: This is an important validity signal: BenchSEE does not merely report better under restriction; it exposes the *shape* of efficiency gains and the degree of performance cost by workload class.

5.4.3 Monotonicity observation across scenarios (S2 → S3 → S4)

Total efficiency improves monotonically as the interventions become more restrictive and energy-oriented:

- **S2 (+7.3%) → S3 (+15.5%) → S4 (+25.3%)**

This monotonic behaviour is a strong sanity check that BenchSEE responds predictably to progressively stronger operating-point constraints.

5.5 S5 – Core-count sweeps (12 logical CPUs and 8 logical CPUs) (Appendix E and F)

S5 differs from S2–S4 in that it constrains available parallel compute resources rather than adjusting power policy/frequency. As a result, it functions as a useful boundary case: reduced core availability may reduce power, but it can also reduce throughput significantly, and therefore does not necessarily improve efficiency.

5.5.1 Component-level changes (Perf/W)

Both core-restriction configurations reduced total efficiency relative to baseline:

12 logical CPUs

- **CPU Perf/W: -44.2%**
- **Memory Perf/W: -4.2%**
- **Storage Perf/W: +0.8%**
- **Total Perf/W: -32.4%**

8 logical CPUs

- **CPU Perf/W: -50.7%**
- **Memory Perf/W: -12.2%**
- **Storage Perf/W: +7.2%**
- **Total Perf/W: -39.0%**

5.5.2 Trade-off analysis

Workload-level results show that parallel CPU workloads (e.g., OLTP, LU, SORT, COMPRESS) experience large raw performance reductions, and workload Perf/W also drops substantially. This indicates that performance loss dominates any power savings achieved by reducing active cores, resulting in **lower overall efficiency**.

By contrast, a small subset of workloads, particularly STREAM and some storage metrics remained stable or improved. This reinforces that BenchSEE is workload-discriminating: it does not force every workload to move in the same direction under a given intervention.

Interpretation: S5 strengthens the validity argument by demonstrating that BenchSEE is capable of detecting *regressions* in efficiency under plausible “energy-saving” actions that are not actually efficient for parallel workloads.

5.6 Cross-scenario synthesis: which techniques improve efficiency, and at what cost?

Across interventions, the clearest pattern is that **frequency and power-policy interventions (S2–S4)** consistently improve total Perf/W, while **core restriction (S5)** substantially reduces it for this workload suite.

- **Best total efficiency improvement: S4 (60% cap)**: **+25.3%** total Perf/W vs baseline
- **Second-best: S3 (80% cap)**: **+15.5%**
- **Moderate improvement: S2 (power-saver)**: **+7.3%**
- **Efficiency regression: S5 (12/8 logical CPUs)**: **-32% to -39%**

From a trade-off perspective, S3 and S4 illustrate that meaningful efficiency gains can occur with only modest performance loss on some workloads (e.g., AES/STREAM), while other workloads (notably COMPRESS) may incur a large throughput penalty to achieve higher Perf/W. S5 demonstrates that simply reducing active CPU resources can reduce total efficiency when workloads scale with parallelism.

Table 1. Scenario summary (component and total Perf/W deltas vs baseline). (Perf/W = performance – power ratio; deltas computed vs S1 baseline means.)

Scenario	Runs (n)	CPU Perf/W	Memory Perf/W	Storage Perf/W	Total Perf/W
S1 Baseline	10	266.7	39.0	14118.5	**182.7**
S2 Power-saver	3	**+12.1%**	-0.2%	**-6.8%**	**+7.3%**
S3 CPU cap 80%	3	**+12.4%**	**+25.8%**	-0.1%	**+15.5%**
S4 CPU cap 60%	3	**+16.1%**	**+52.8%**	+2.1%	**+25.3%**
S5 Core limit: 12 logical CPUs	3	**-44.2%**	-4.2%	+0.8%	**-32.4%**
S5 Core limit: 8 logical CPUs	3	**-50.7%**	-12.2%	+7.2%	**-39.0%**

6. EE score comparative analysis (S2–S5 vs S1)

To avoid the storage workloads (RANDOM/SEQUENTIAL) dominating any aggregate (their scores are orders of magnitude larger), the table below reports **percent changes** and uses the **median delta across all 40 workload points** as the main overall indicator.

Table 2. Summary of EE-score deltas vs baseline (computed across workload points)

Scenario	Median Δ% (all 40 points)	Mean Δ% (CPU workloads)	Mean Δ% (Memory workloads)	Mean Δ% (Storage workloads)
S2 Power-saver	**+2.4%**	**+13.2%**	+0.4%	**-6.8%**
S3 CPU cap 80%	**+8.6%**	**+13.5%**	**+28.7%**	-0.1%
S4 CPU cap 60%	**+13.7%**	**+19.7%**	**+57.3%**	+2.1%
S5 Core limit: 12 logical CPUs	**-42.2%**	**-41.5%**	+5.9%	+0.8%
S5 Core limit: 8 logical CPUs	**-45.0%**	**-47.0%**	-1.1%	**+7.2%**

Interpretation: The EE-score response is monotonic across the operating point interventions (**S2 → S3 → S4**) and strongly negative for compute resource restriction (**S5**), which is exactly as expected from a *work done per energy* benchmark.

Table 3. Workload-level average EE-score change (Δ% vs baseline)

Workload	S2 Δ%	S3 Δ%	S4 Δ%	S5 (12) Δ%	S5 (8) Δ%
AES	+12.8%	+17.4%	**+28.3%**	-17.4%	-23.3%
COMPRESS	+12.0%	+12.1%	+14.2%	**-52.1%**	**-62.7%**
LU	+7.9%	+3.0%	+14.7%	-40.1%	-51.4%
OLTP	+14.5%	+17.9%	**+29.6%**	**-49.0%**	-44.4%
SHA256	+8.0%	+10.2%	+10.9%	**-51.2%**	**-62.0%**
SOR	+19.1%	+24.5%	**+44.5%**	-8.3%	-18.7%
SORT	+11.1%	+13.0%	+9.8%	**-59.4%**	**-65.5%**
CACHE	-1.1%	+22.7%	**+53.8%**	-17.6%	-34.1%
STREAM	+1.3%	**+30.7%**	**+56.0%**	+22.7%	+32.0%
RANDOM	-4.7%	+0.8%	+3.2%	+0.6%	**+9.7%**
SEQUENTIAL	-8.4%	-1.0%	+1.2%	+0.9%	+4.6%

Each workload value is the **mean of its load levels** (e.g., OLTP averaged across 100%...12.5%; RANDOM across 100% and 50%). Values below are **percent deltas vs S1**.

7. Discussion: BenchSEE validity and practical implications

7.1 Does BenchSEE provide an objective opinion on efficiency improvement?

Across S2–S4, BenchSEE reports consistent, directional improvements in both (i) component/total Perf/W and (ii) workload EE scores, with a clear monotonic progression as the CPU operating point is increasingly constrained:

- Total Perf/W: S2 (+7.3%) → S3 (+15.5%) → S4 (+25.3%)
- Median EE-score delta across workload points: S2 (+2.4%) → S3 (+8.6%) → S4 (+13.7%)

This alignment between *high-level aggregates* (total Perf/W) and *benchmark-native judgement* (EE scores) supports the conclusion that BenchSEE is not simply capturing noise or incidental variance, but is responding to controlled interventions in a coherent way.

Importantly, BenchSEE also detects efficiency regressions where an intervention plausibly reduces power but harms throughput more than it helps energy use. Under S5 (12 and 8 logical CPUs), total Perf/W drops sharply (-32% to -39%) and the median EE-score delta is strongly negative (-42% to -45%). This is valuable evidence of objectivity.

7.2 Sensitivity and repeatability: separating effects from baseline noise

Baseline repeatability analysis established a practical "noise floor", with moderate variability in CPU/memory and higher relative dispersion in storage. Despite this, the intervention effects observed in S2–S5 are large in magnitude and consistent in direction across repeats, and therefore exceed baseline run-to-run variation.

A baseline drift effect (uplift in runs 9–10 relative to earlier runs) was observed, indicating that environmental or platform state factors can influence results over time. However, the intervention deltas reported here are substantially larger than the baseline drift and were consistent across repeated runs, suggesting that the benchmark remains usable for comparative evaluation provided baseline replication and careful experimental control are maintained.

7.3 Workload discrimination and mechanism-consistency

A key validity criterion for an EE benchmark is that it should be mechanistically interpretable: interventions should affect workloads in ways consistent with the subsystem being changed.

This is observed clearly in the results:

- **CPU frequency capping (S3/S4)** produces broad improvements on CPU/memory-centric workloads (e.g., OLTP, AES, STREAM, SOR), while storage workloads change only marginally. This matches expectations: the intervention targets the CPU operating point, so CPU/memory performance-per-watt improves most.
- **Power-saver (S2)** improves CPU-side efficiency but degrades storage metrics more noticeably (RANDOM/SEQUENTIAL). This suggests that OS-level power policies can change platform behaviour beyond the CPU alone (e.g., IO path and residency effects), and BenchSEE correctly reflects that non-uniform subsystem impact rather than reporting a uniform gain.
- **Core-count restriction (S5)** strongly penalises parallel CPU workloads (OLTP, LU, SORT, COMPRESS) where throughput scales with available compute. Meanwhile, STREAM and some storage-related metrics remain stable or even improve, indicating that BenchSEE is sensitive to workload class and does not force all results to move in the same direction.

Overall, the workload-level EE scores and Perf/W ratios provide a consistent narrative: BenchSEE is discriminating between workload types and capturing expected “where the efficiency comes from” behaviour.

7.4 Trade-off visibility: efficiency gains versus performance sacrifice

An additional strength of the BenchSEE output set is that it enables transparent trade-off analysis via the combined availability of:

- raw performance scores,
- workload Perf/W,
- and workload EE scores.

This makes it possible to separate:

- **favourable efficiency improvements** (efficiency increases with minimal performance loss), from
- **costly efficiency improvements** (efficiency increases but performance decreases substantially).

This distinction is important for value perspective decision-making, since the optimal configuration depends on whether the primary objective is absolute throughput, energy efficiency, or a balanced operating point. In this evaluation, stronger caps (e.g., S4) improved overall efficiency most, but some workloads (notably COMPRESS) exhibit a greater throughput penalty than others. BenchSEE therefore supports not only ranking configurations by efficiency but also identifying which workloads are most sensitive to particular optimisation choices.

7.5 Practical conclusions for using BenchSEE in optimisation studies

Based on the observed behaviour, BenchSEE is suitable for comparative EE evaluation provided the following practices are applied:

- **Use repeated baselines and replicate interventions** to account for drift and variability, particularly in storage metrics.
- **Report both EE scores and Perf/W**, and always pair efficiency gains with raw performance deltas to make trade-offs explicit.
- **Interpret storage-heavy results with wider uncertainty margins**, and treat IO-sensitive points as less reliable for small deltas.
- **Prefer scenario ranking by both (i) total outcome and (ii) workload-class impact**, since interventions can improve CPU/memory efficiency while leaving storage unchanged or slightly worse.

8. Conclusion

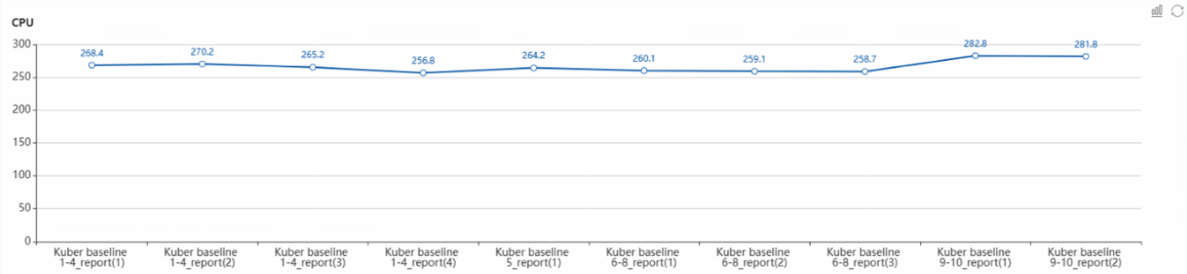
This report evaluated BenchSEE on a modern x86_64 workstation (Intel i9-14900K, 192 GB DDR5) using a controlled set of energy optimisation techniques and resource constraints. The results support a positive validity assessment:

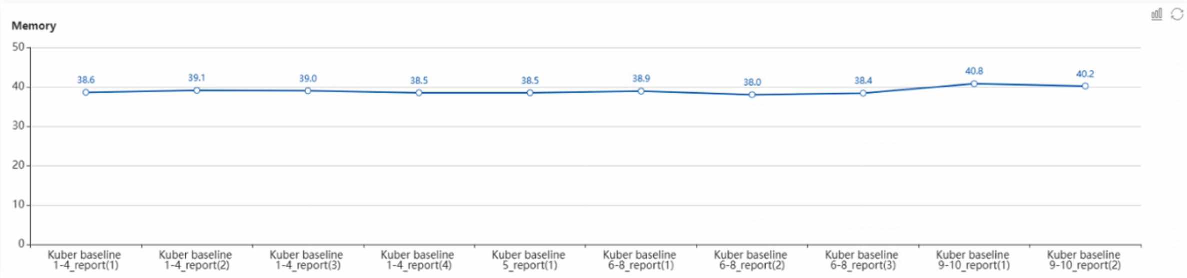
1. **Sensitivity:** BenchSEE detects efficiency improvements under power-policy and frequency-capping interventions (S2–S4), with monotonic improvements in both total Perf/W and workload EE-score aggregates.
2. **Objectivity:** BenchSEE reports efficiency regressions under core-count restriction (S5), demonstrating that it does not automatically reward energy-saving settings when they reduce work performed disproportionately.
3. **Workload discrimination:** The benchmark differentiates subsystem effects across workload classes in a mechanistically interpretable way, aligning CPU-targeted interventions with CPU/memory workload improvements and leaving storage relatively stable (except under OS-wide power-saver policies).
4. **Decision usefulness:** The combined reporting of EE scores, Perf/W, and raw performance enables explicit trade-off analysis, supporting practical configuration decisions based on efficiency goals and workload priorities.

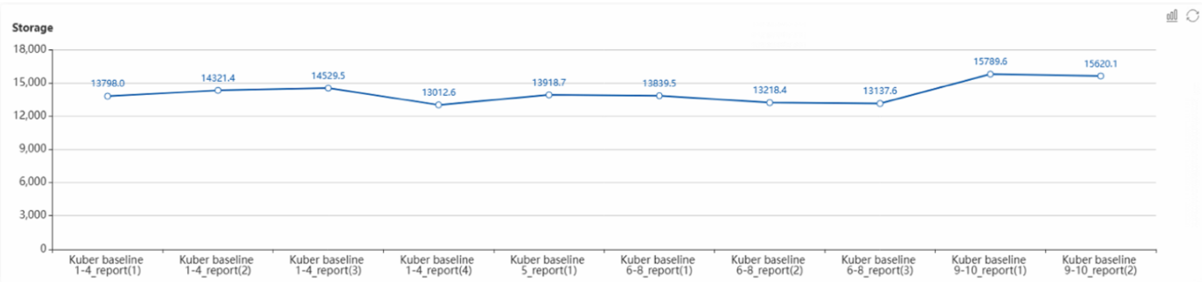
Overall, BenchSEE provides a consistent and interpretable basis for evaluating server energy efficiency changes under controlled configuration interventions, and is appropriate for comparative optimisation studies when paired with replication, baseline control, and workload-class-aware interpretation.

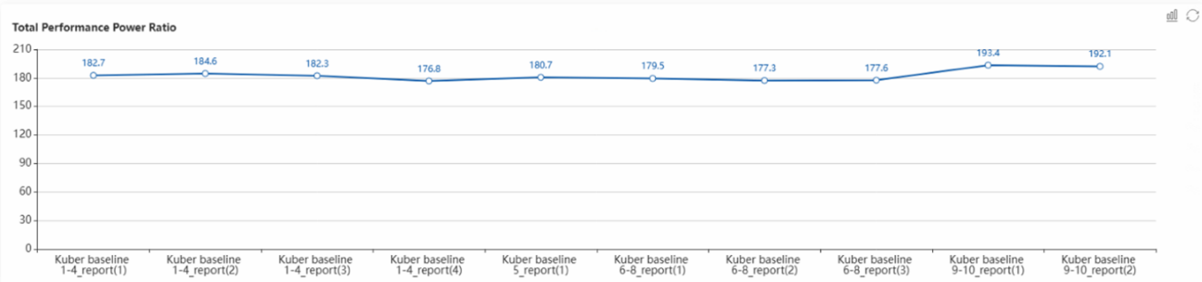
Appendix A.1: Baseline results: Performance Power Ratio Comparison of Each Component

Component	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10
CPU	268.4	270.2	265.2	256.8	264.2	260.1	259.1	258.7	**282.8**	281.8
Memory	38.6	39.1	39.0	38.5	38.5	38.9	38.0	38.4	**40.8**	40.2
Storage	13798.0	14321.4	14529.5	13012.6	13918.7	13839.5	13218.4	13137.6	**15789.6**	15620.1
Total Performance Power Ratio	182.7	184.6	182.3	176.8	180.7	179.5	177.3	177.6	**193.4**	192.1

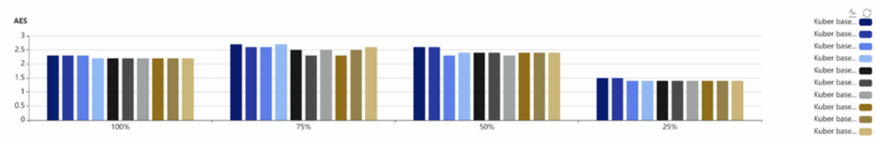


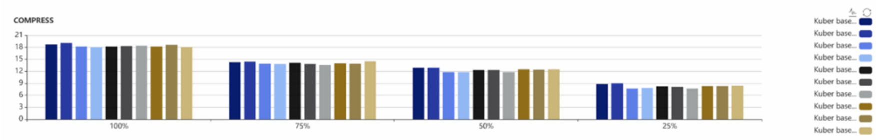


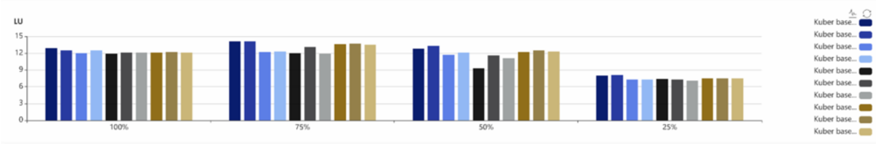


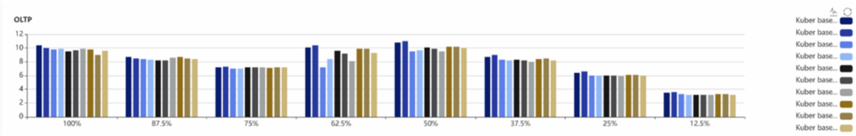


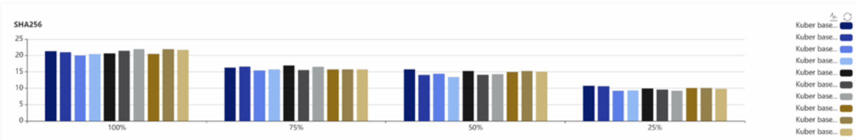
Appendix A.2: Baseline Results: Energy Efficiency Score Comparison of Each Workload

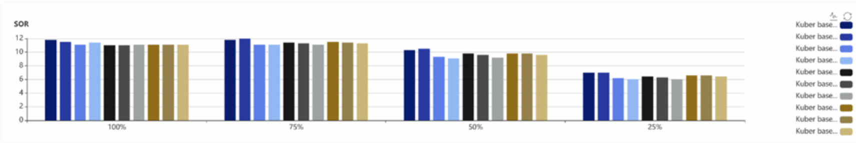


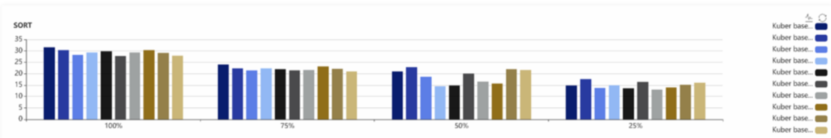


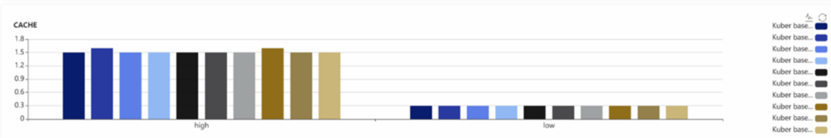


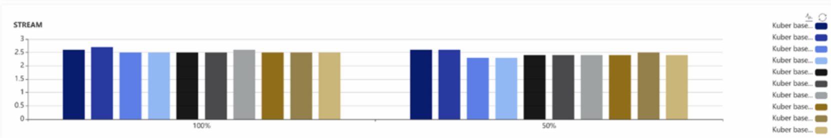


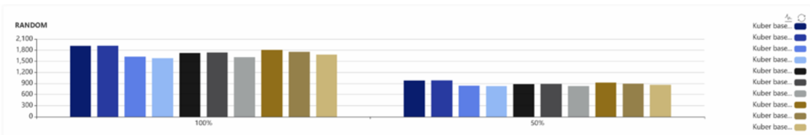


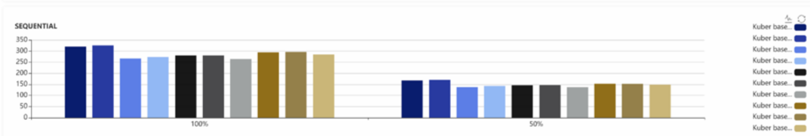












Appendix A.2 – Baseline results: Energy Efficiency Score Comparison of Each Workload

Workload	Workload Level	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10
AES	100 %	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.3	2.3	2.3
	75 %	2.6	2.5	2.3	2.5	2.3	2.5	2.7	2.6	2.6	2.7
	50 %	2.4	2.4	2.4	2.3	2.4	2.4	2.4	2.3	2.6	2.6
	25 %	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.5	1.5
COMPRESS	100 %	18.1	18.6	18.2	18.4	18.3	18.2	18.0	18.2	19.1	18.7
	75 %	14.5	13.9	14.0	13.6	13.8	14.1	13.8	13.9	14.4	14.3
	50 %	12.5	12.4	12.5	11.8	12.3	12.3	11.8	11.8	12.9	12.9
	25 %	8.4	8.3	8.3	7.7	8.1	8.2	7.8	7.7	9.0	8.8
LU	100 %	12.1	12.2	12.1	12.1	12.1	11.9	12.5	12.0	12.5	12.9
	75 %	13.5	13.7	13.6	11.9	13.1	12.0	12.3	12.2	14.1	14.1
	50 %	12.3	12.5	12.2	11.1	11.6	9.3	12.1	11.7	13.3	12.8
	25 %	7.5	7.5	7.5	7.1	7.3	7.4	7.3	7.3	8.1	8.0
OLTP	100 %	9.6	9.0	9.8	9.9	9.7	9.5	9.9	9.8	10.0	10.4
	87.5 %	8.4	8.5	8.7	8.6	8.2	8.2	8.3	8.4	8.5	8.7
	75 %	7.2	7.2	7.1	7.2	7.2	7.2	7.0	7.0	7.3	7.2
	62.5 %	9.3	9.9	9.9	8.1	9.2	9.6	8.4	7.2	10.4	10.1
	50 %	10.0	10.2	10.2	9.5	9.9	9.5	9.7	9.5	11.0	10.8
	37.5 %	8.2	8.5	8.4	8.0	8.2	8.3	8.2	8.3	9.0	8.7
	25 %	6.0	6.1	6.1	5.9	6.0	6.0	6.0	6.0	6.6	6.4
	12.5 %	3.2	3.3	3.3	3.2	3.2	3.2	3.2	3.3	3.6	3.5
SHA256	100 %	21.7	21.9	20.5	21.9	21.4	20.6	20.4	20.0	21.0	21.3
	75 %	15.7	15.8	15.7	16.5	15.5	16.9	15.7	15.4	16.6	16.3
	50 %	15.0	15.2	14.9	14.3	14.1	15.2	13.4	14.4	14.0	15.8
	25 %	9.8	10.1	10.0	9.2	9.6	9.9	9.3	9.2	10.6	10.7
SOR	100 %	11.1	11.1	11.1	11.1	11.0	11.0	11.4	11.1	11.5	11.8
	75 %	11.3	11.4	11.5	11.1	11.3	11.4	11.1	11.1	12.0	11.8
	50 %	9.6	9.8	9.8	9.2	9.6	9.8	9.1	9.3	10.5	10.3
	25 %	6.4	6.6	6.6	6.0	6.3	6.4	6.0	6.2	7.0	7.0
SORT	100 %	27.9	29.1	30.3	29.3	29.3	27.8	29.8	29.3	30.3	31.6
	75 %	21.0	22.1	23.2	21.6	21.5	22.0	22.3	21.4	22.3	24.0
	50 %	21.6	22.0	15.7	16.5	20.1	14.8	14.5	18.7	22.9	21.0
	25 %	16.0	15.1	13.9	13.0	13.0	16.4	13.6	14.9	17.6	14.8
CACHE	high	1.5	1.5	1.6	1.5	1.5	1.5	1.5	1.5	1.6	1.5
	low	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
STREAM	100 %	2.5	2.5	2.5	2.5	2.6	2.5	2.5	2.5	2.7	2.6
	50 %	2.4	2.5	2.4	2.4	2.4	2.4	2.3	2.6	2.6	2.6
RANDOM	100 %	1677.9	1755.8	1808.4	1609.4	1734.4	1719.9	1586.2	1627.2	1918.7	1913.7
	50 %	860.7	892.6	920.7	829.1	886.1	877.4	827.0	837.1	983.8	975.7
SEQUENTIAL	100 %	284.4	296.0	294.7	264.4	279.9	280.2	273.7	266.8	325.5	320.5
	50 %	148.2	152.3	152.5	136.5	145.7	145.7	142.8	137.6	169.9	167.1

Appendix A.3 – Baseline Results: Raw Performance Score Comparison of Each Workload

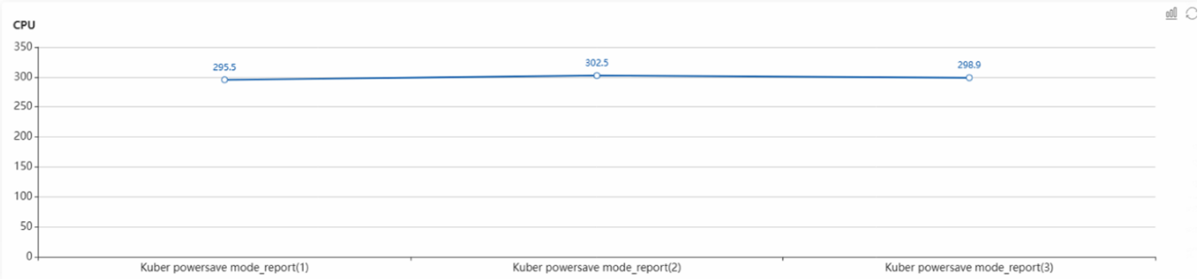
Workload	Workload Level	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10
AES	100%	480470.9	479726.9	480445.5	480555.7	480615.0	480280.9	479829.5	480528.8	480414.3	480675.2
	75%	360341.3	359778.2	360221.8	360537.9	360519.5	360182.0	359849.6	360366.8	360378.8	360533.5
	50%	240232.0	239881.3	240217.7	240331.1	240342.2	240125.3	239899.1	240274.6	240225.1	240406.2
	25%	120146.7	119901.6	120124.3	120174.0	120116.3	120074.5	119946.3	120134.7	120139.9	120215.3
COMPRESS	100%	19080.3	19369.0	19379.1	19101.6	19342.3	19091.9	19203.1	19344.2	19335.0	19358.9
	75%	14308.6	14536.9	14569.8	14330.7	14519.7	14345.7	14429.3	14517.9	14529.9	14546.7
	50%	9562.3	9709.8	9704.9	9585.6	9694.1	9573.2	9621.2	9702.7	9685.6	9703.0
	25%	4794.7	4873.1	4854.5	4786.1	4840.0	4787.7	4828.0	4852.7	4850.9	4855.5
LU	100%	532506.5	532010.9	532572.5	531750.7	533225.3	533239.1	532806.3	534496.4	533081.2	532091.1
	75%	399427.1	398971.0	399444.8	398865.2	399921.5	399866.4	399535.3	400928.3	399849.8	399023.5
	50%	266277.0	266039.3	266217.1	265851.8	266635.1	266688.8	266425.5	267306.9	266459.0	266042.0
	25%	133130.3	133037.0	133208.1	132895.4	133366.6	133298.0	133216.4	133658.4	133259.2	133003.5
OLTP	100%	2538852.3	2538369.6	2532885.3	2544660.8	2524750.6	2552715.2	2541920.1	2549450.0	2525695.1	2536360.4
	87.5%	2230005.3	2226924.2	2225263.2	2236522.9	2229714.5	2236756.0	2228943.4	2236008.9	2227281.2	2225782.8
	75%	1911344.9	1908718.7	1907491.0	1916748.6	1911200.2	1917036.0	1910592.0	1916178.6	1908984.1	1908050.2
	62.5%	1592699.5	1590696.5	1589592.2	1597414.6	1592469.7	1597547.7	1591812.2	1596918.4	1590525.0	1589820.0
	50%	1274123.9	1272616.5	1271679.5	1277916.0	1274218.8	1277939.0	1273643.4	1277662.8	1272562.3	1271977.9
	37.5%	955723.0	954196.7	953543.9	958289.4	955526.3	958483.5	955252.8	958301.3	954412.0	954048.8
	25%	636944.2	636270.5	635826.1	638997.2	636977.7	639010.1	636660.5	638804.2	636427.9	636029.1
	12.5%	318590.7	318149.3	317946.9	319517.1	318547.5	319567.4	318412.3	319453.1	318054.4	318039.9
SHA256	100%	62532.1	62169.5	61688.3	61815.9	61624.6	62662.7	62386.7	60753.1	62593.2	62268.4
	75%	46927.3	46656.4	46272.0	46353.4	46200.7	46991.8	46808.1	45562.1	46993.9	46695.5
	50%	31328.6	31101.1	30860.9	30896.5	30849.0	31343.0	31209.5	30388.0	31338.4	31173.2
	25%	15647.7	15574.9	15432.3	15504.5	15413.4	15697.0	15609.2	15218.6	15696.4	15588.3
SOR	100%	2385.6	2384.4	2380.2	2384.3	2379.2	2383.1	2382.7	2381.4	2383.9	2381.7
	75%	1800.8	1794.5	1803.2	1805.4	1798.7	1809.4	1808.6	1803.4	1795.3	1802.3
	50%	1210.0	1201.3	1203.5	1207.0	1211.6	1198.4	1202.2	1203.2	1209.6	1204.4
	25%	612.4	609.8	614.6	608.7	610.5	607.5	612.7	615.5	608.8	614.4
SORT	100%	2915126.7	3063013.5	3202695.7	2994478.6	3021720.8	3031854.7	3091720.1	2971205.1	2993095.2	3234445.8
	75%	2186654.1	2297370.8	2402189.4	2246183.9	2266515.4	2274283.4	2319137.4	2228662.4	2245066.9	2425956.3
	50%	1457486.0	1531541.9	1601406.9	1497253.9	1510977.0	1516232.5	1545772.7	1485660.2	1496568.2	1617372.7
	25%	728611.9	765779.4	800665.7	748566.9	755381.6	758078.3	772969.1	742796.9	748182.5	808699.4
CACHE	high	712801.4	718126.5	712613.4	712663.0	718243.3	716700.1	713199.4	715056.0	714308.7	716989.0
	low	142872.5	143457.6	142787.9	142637.7	143663.9	143423.9	143048.7	143173.8	143235.6	143618.4
STREAM	100%	5594.6	5590.8	5596.2	5592.5	5586.9	5585.9	5590.6	5589.1	5586.4	5589.6
	50%	2797.3	2795.4	2798.1	2796.3	2793.5	2793.0	2795.3	2794.6	2793.2	2794.8
RANDOM	100%	57278.6	57975.2	59860.0	59744.3	59843.9	59401.1	59627.3	59346.9	57688.7	57733.1
	50%	28639.6	28987.7	29930.3	29872.2	29922.2	29700.9	29814.0	29673.9	28844.7	28866.7
SEQUENTIAL	100%	22506.7	22567.8	22603.8	22394.3	22536.4	22298.6	22553.0	22221.2	22765.6	22565.9
	50%	11253.4	11284.0	11302.0	11197.2	11268.3	11149.3	11276.6	11110.7	11382.9	11283.0

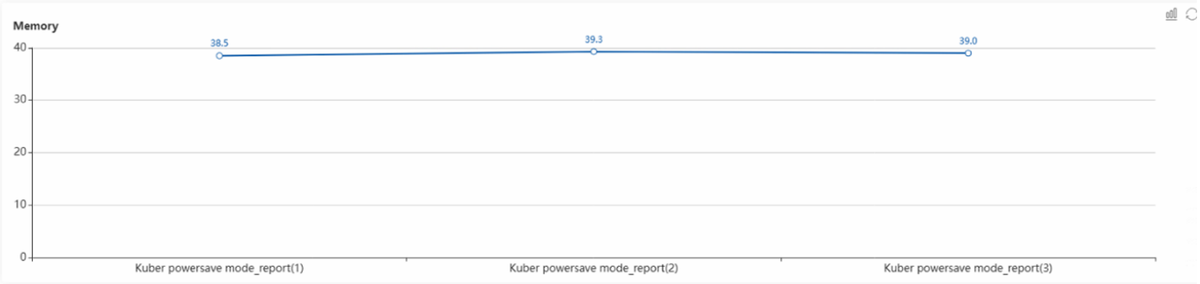
Appendix A.4 – Baseline Results: Raw Performance Power Ratio Comparison

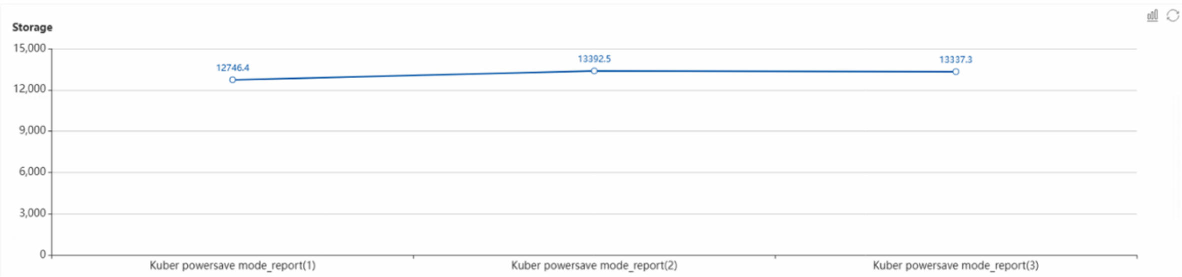
Workload	Workload Level	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10
AES	100%	60.9	61.7	61.8	61.8	60.7	60.7	62.2	63.2	63.7	63.6
	75%	70.9	68.3	63.7	68.1	64.6	70.3	75.6	73.0	70.9	74.5
	50%	67.4	67.3	67.2	63.7	67.1	66.8	67.7	65.2	72.4	72.4
	25%	39.4	39.3	39.4	38.6	39.1	39.3	39.2	38.8	43.0	42.7
COMPRESS	100%	501.5	516.9	504.9	512.0	507.5	505.9	500.6	504.3	530.8	520.4
	75%	403.8	385.4	387.8	377.9	384.6	390.6	383.0	385.5	399.3	397.6
	50%	346.9	345.8	346.5	328.4	341.9	341.7	328.5	326.9	359.5	357.4
	25%	233.7	231.3	230.2	214.7	224.2	226.4	217.3	214.1	250.2	243.8
LU	100%	337.1	337.9	336.2	337.5	335.8	331.1	347.5	333.3	346.2	358.6
	75%	375.4	381.2	377.3	330.1	364.2	333.1	340.4	338.4	391.9	392.6
	50%	342.7	347.3	338.3	309.2	322.7	258.3	335.7	325.4	369.3	356.1
	25%	207.2	207.4	208.3	197.1	203.8	204.2	204.1	202.9	226.2	222.7
OLTP	100%	321.2	301.6	326.1	328.4	323.7	316.2	329.2	327.4	332.5	346.0
	87.5%	281.1	284.3	290.6	287.8	274.1	274.3	275.2	281.4	283.4	289.6
	75%	238.4	241.1	237.3	241.3	240.1	238.6	233.8	234.6	244.7	241.1
	62.5%	310.7	330.9	330.0	271.6	307.4	320.4	278.9	241.5	346.4	335.8
	50%	334.5	341.5	341.7	317.0	330.9	337.4	323.3	317.2	366.7	358.7
	37.5%	272.1	282.5	280.9	265.6	272.1	276.9	273.6	276.6	300.5	289.9
	25%	199.3	202.4	202.6	195.6	198.4	200.1	200.0	198.6	218.5	213.8
	12.5%	107.2	108.9	108.7	107.9	106.6	107.4	108.3	108.5	118.6	117.1
SHA256	100%	603.4	609.3	569.5	608.5	594.1	572.3	568.1	556.0	583.6	591.2
	75%	436.9	437.8	435.2	459.4	431.6	469.9	437.0	427.4	460.0	453.3
	50%	415.3	421.1	414.9	396.3	390.7	421.2	371.5	400.2	387.9	440.0
	25%	272.1	279.7	276.5	255.5	266.9	275.6	257.2	256.0	293.4	296.3
SOR	100%	307.9	308.5	307.4	308.3	306.6	306.8	317.9	308.8	318.9	326.6
	75%	314.3	317.7	319.8	307.8	313.9	317.7	307.9	309.0	333.1	329.1
	50%	268.0	272.3	270.9	257.0	265.3	270.9	252.8	259.4	292.2	285.9
	25%	177.5	182.7	182.6	166.8	175.4	178.3	165.8	173.0	195.5	193.7
SORT	100%	776.0	807.7	842.3	813.4	821.1	827.1	814.1	783.4	842.7	878.7
	75%	583.4	612.8	644.0	598.7	597.6	610.1	618.2	595.5	619.3	666.1
	50%	599.7	611.9	436.3	458.6	558.3	411.1	403.4	520.8	636.0	582.7
	25%	444.1	420.2	384.9	361.5	454.2	377.7	412.9	381.9	488.9	412.0
CACHE	high	46.1	46.8	48.6	46.1	46.6	46.4	46.1	46.6	48.0	48.3
	low	9.4	9.5	9.2	9.1	9.5	9.5	9.2	9.2	9.4	9.5
STREAM	100%	72.4	72.5	72.4	75.0	70.8	72.4	72.5	73.0	78.2	74.8
	50%	71.1	72.7	71.1	69.8	70.2	71.9	67.6	69.3	78.0	76.2
RANDOM	100%	46609.4	48771.1	50233.3	44707.4	48179.4	47775.3	44060.4	45201.8	53296.9	53159.6
	50%	23909.3	24795.0	25575.8	23030.4	24612.9	24373.5	22971.3	23253.3	27327.0	27101.4
SEQUENTIAL	100%	7899.6	8222.8	8186.9	7343.4	7775.9	7783.1	7602.9	7412.5	9040.6	8902.0
	50%	4117.4	4230.5	4237.1	3792.2	4070.2	4047.7	3967.3	3823.5	4720.5	4641.7

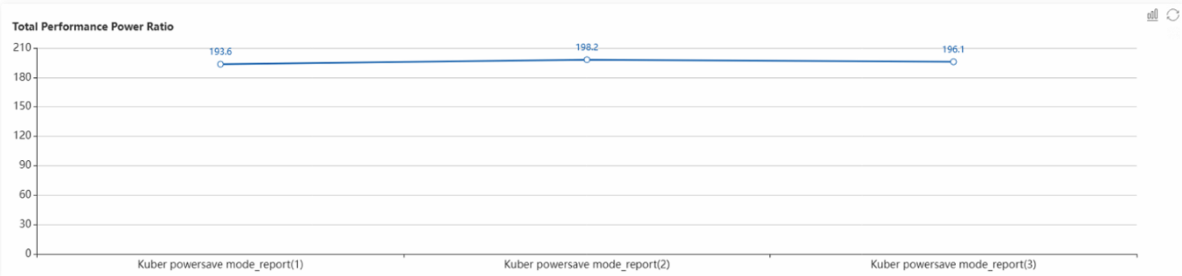
Appendix B.1: Power-save mode: Performance Power Ratio Comparison of Each Component

Component	Test 1	Test 2	Test 3
CPU	295.5	**302.5**	298.9
Memory	38.5	**39.3**	39.0
Storage	12746.4	**13392.5**	13337.3
Total Performance Power Ratio	193.6	**198.2**	196.1









Appendix B.2 – Power-save mode: Energy Efficiency Score Comparison of Each Workload

Workload	Workload Level	Test 1	Test 2	Test 3
AES	100%	2.2	2.2	2.2
	75%	3.5	3.5	3.5
	50%	2.6	2.6	2.6
	25%	1.4	1.4	1.4
COMPRESS	100%	18.1	18.0	18.0
	75%	14.2	15.3	14.4
	50%	15.3	15.8	15.9
	25%	10.9	11.0	11.0
LU	100%	12.5	12.1	12.0
	75%	14.7	16.0	15.7
	50%	12.9	13.1	13.0
	25%	7.5	7.6	7.5
OLTP	100%	10.0	9.8	9.8
	87.5%	8.6	8.4	8.8
	75%	10.9	12.3	12.1
	62.5%	11.8	12.4	12.3
	50%	10.7	11.0	11.0
	37.5%	8.6	8.8	8.8
	25%	6.1	6.2	6.2
	12.5%	3.2	3.3	3.3
SHA256	100%	20.3	20.3	19.6
	75%	15.7	16.6	15.4
	50%	17.8	19.3	18.0
	25%	12.0	12.3	12.2
SOR	100%	11.3	11.1	11.1
	75%	14.1	14.8	15.2
	50%	12.7	12.8	12.9
	25%	7.5	7.5	7.5
SORT	100%	27.4	29.9	28.8
	75%	23.1	23.9	22.0
	50%	26.0	26.0	25.7
	25%	16.5	17.4	17.4
CACHE	high	1.5	1.5	1.5
	low	0.3	0.3	0.3
STREAM	100%	2.5	2.6	2.5
	50%	2.5	2.6	2.5
RANDOM	100%	1643.8	1663.7	1672.4
	50%	832.8	841.4	846.6
SEQUENTIAL	100%	251.7	274.6	271.0
	50%	128.7	140.6	138.5

Appendix B.3 – Power-save mode: Raw Performance Score Comparison of Each Workload

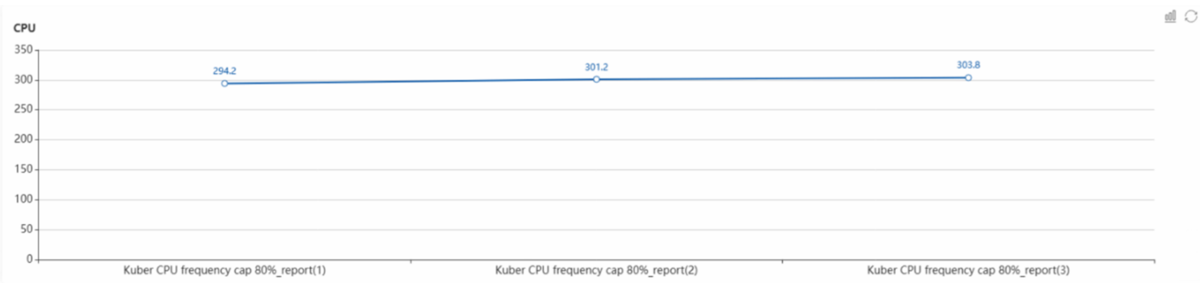
Workload	Level	Test 1	Test 2	Test 3
AES	100%	479981.9	479691.1	479711.1
	75%	360045.3	359907.4	359739.3
	50%	239973.8	239887.1	239874.2
	25%	119995.3	119971.5	119940.1
COMPRESS	100%	19339.3	19340.5	19310.1
	75%	14497.0	14527.7	14494.3
	50%	9678.7	9684.1	9664.3
	25%	4848.7	4853.2	4837.5
LU	100%	532586.1	532437.1	530867.0
	75%	399475.4	399394.2	398207.3
	50%	266319.7	266169.7	265489.2
	25%	133179.0	133151.3	132676.9
OLTP	100%	2544737.9	2528812.2	2543797.2
	87.5%	2232544.0	2228681.1	2234540.5
	75%	1913798.2	1910027.9	1915232.1
	62.5%	1594623.1	1591889.0	1596063.5
	50%	1275576.2	1273527.2	1276845.9
	37.5%	956679.9	955030.1	957678.7
	25%	637893.2	636702.3	638462.9
	12.5%	318883.9	318379.1	319208.4
SHA256	100%	60941.6	62029.1	61192.0
	75%	45739.9	46547.5	45900.5
	50%	30500.3	31018.3	30620.8
	25%	15261.7	15536.7	15321.4
SOR	100%	2384.1	2381.8	2382.5
	75%	1806.9	1791.4	1794.7
	50%	1214.1	1208.5	1212.5
	25%	614.7	611.4	610.1
SORT	100%	2943865.0	3029275.3	3031407.4
	75%	2208277.7	2271998.0	2273336.9
	50%	1471888.7	1514823.7	1516028.2
	25%	735925.0	757238.1	757825.5
CACHE	high	714881.4	713006.1	706076.5
	low	142984.2	142905.4	141551.6
STREAM	100%	5596.0	5592.2	5595.4
	50%	2798.0	2796.1	2797.7
RANDOM	100%	55050.2	54622.1	55139.5
	50%	27525.1	27311.4	27570.0
SEQUENTIAL	100%	19263.4	20735.1	20505.5
	50%	9631.8	10367.6	10252.9

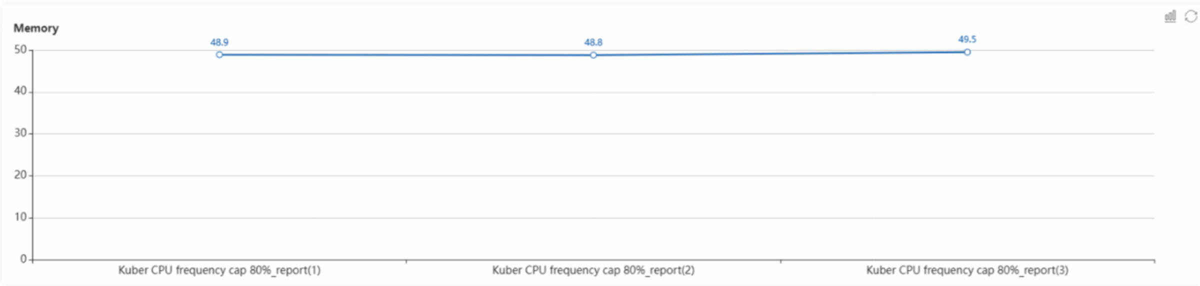
Appendix B.4 – Power-save mode: Raw Performance Power Ratio Comparison

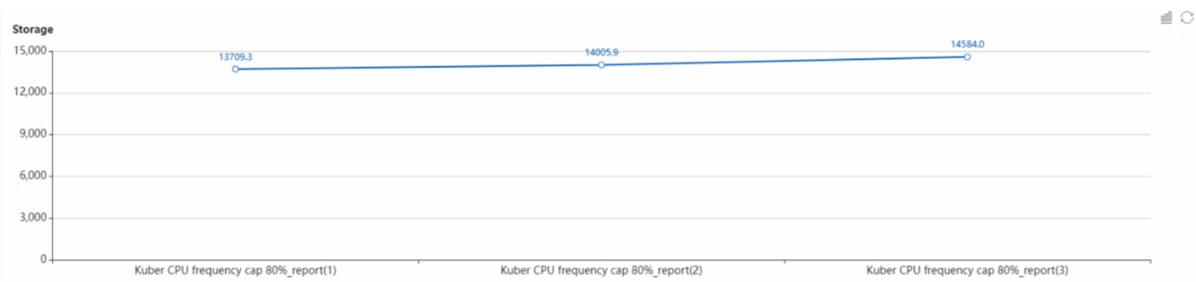
Workload	Level	Test 1	Test 2	Test 3
AES	100%	61.5	61.8	61.5
	75%	96.2	97.5	97.9
	50%	71.6	71.8	71.9
	25%	39.4	39.4	39.4
COMPRESS	100%	502.8	500.5	501.2
	75%	394.2	423.9	401.1
	50%	425.0	440.0	441.1
	25%	302.3	306.4	304.7
LU	100%	348.5	336.9	334.8
	75%	409.7	445.6	436.2
	50%	359.0	362.6	360.3
	25%	209.7	210.0	209.4
OLTP	100%	332.7	325.6	326.3
	87.5%	286.5	279.3	294.5
	75%	363.3	409.7	404.5
	62.5%	393.8	413.1	411.5
	50%	356.7	366.1	366.6
	37.5%	287.6	291.8	292.8
	25%	203.9	205.9	206.6
	12.5%	107.7	108.4	108.7
SHA256	100%	563.7	563.8	544.0
	75%	437.3	460.0	428.9
	50%	495.3	535.3	501.3
	25%	332.4	342.5	338.3
SOR	100%	314.7	307.1	307.1
	75%	393.0	411.5	421.9
	50%	352.3	354.4	359.7
	25%	208.2	208.9	209.6
SORT	100%	760.4	830.9	800.8
	75%	642.4	662.9	611.7
	50%	721.4	722.3	714.5
	25%	457.1	482.5	483.2
CACHE	high	46.5	46.2	48.1
	low	9.1	9.1	9.1
STREAM	100%	71.7	75.0	72.1
	50%	72.3	75.8	73.1
RANDOM	100%	45662.7	46213.8	46455.8
	50%	23133.5	23371.8	23518.4
SEQUENTIAL	100%	6992.2	7628.1	7526.8
	50%	3573.8	3904.5	3847.9

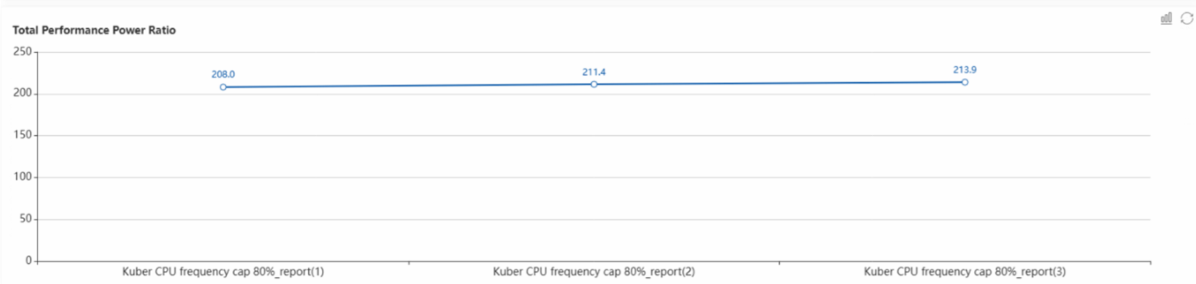
Appendix C.1 – CPU 80% Frequency: Performance Power Ratio Comparison of Each Component

Component	Test 1	Test 3	Test 3
CPU	294.2	301.2	**303.8**
Memory	48.9	48.8	**49.5**
Storage	13709.3	14005.9	**14584.0**
Total Performance Power Ratio	208.0	211.4	**213.9**









Appendix C.2 – CPU 80% Frequency: Energy Efficiency Score Comparison of Each Workload

Workload	Workload Level	Test 1	Test 2	Test 3
AES	100%	3.0	3.0	3.0
	75%	3.2	3.3	3.2
	50%	2.4	2.5	2.5
	25%	1.4	1.4	1.4
COMPRESS	100%	19.6	19.7	19.7
	75%	16.6	16.9	16.9
	50%	13.7	14.2	14.1
	25%	8.7	9.0	9.0
LU	100%	13.8	13.8	13.8
	75%	12.1	13.0	13.6
	50%	11.1	12.3	12.2
	25%	7.2	7.6	7.6
OLTP	100%	13.4	13.4	13.4
	87.5%	12.0	12.2	12.1
	75%	10.7	10.8	10.8
	62.5%	9.5	10.1	10.0
	50%	9.6	10.0	10.3
	37.5%	8.0	8.4	8.5
	25%	5.8	6.0	6.0
	12.5%	3.2	3.3	3.3
SHA256	100%	22.9	22.9	23.1
	75%	19.1	19.2	19.6
	50%	15.7	15.9	15.5
	25%	9.6	9.9	10.1
SOR	100%	15.0	15.2	15.0
	75%	14.6	14.9	15.0
	50%	11.3	11.4	11.7
	25%	6.7	6.8	7.2
SORT	100%	32.3	32.0	32.3
	75%	26.4	26.2	26.9
	50%	21.2	22.0	24.6
	25%	14.5	15.3	15.1
CACHE	high	1.8	1.8	1.9
	low	0.4	0.4	0.4
STREAM	100%	3.7	3.7	3.7
	50%	2.8	2.8	2.9
RANDOM	100%	1696.8	1732.5	1828.5
	50%	868.9	884.0	926.7
SEQUENTIAL	100%	277.4	284.3	295.1
	50%	145.1	148.4	151.9

Appendix C.3 – CPU 80% Frequency: Raw Performance Score Comparison of Each Workload

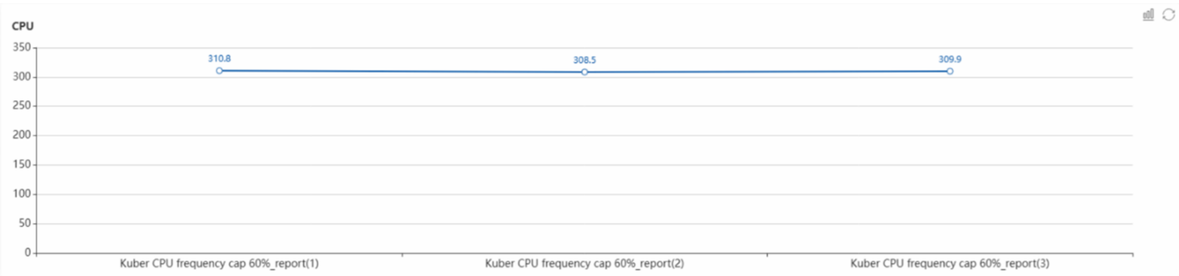
Workload	Workload Level	Test 1	Test 2	Test 3
AES	100%	478770.0	479428.1	479547.8
	75%	359116.9	359543.7	359716.5
	50%	239343.5	239711.7	239781.9
	25%	119688.7	119970.8	119944.1
COMPRESS	100%	17457.5	17837.9	17652.6
	75%	13106.2	13416.1	13247.9
	50%	8760.0	8936.4	8839.7
	25%	4372.2	4477.0	4422.1
LU	100%	536053.8	536623.7	536513.9
	75%	402168.5	402630.3	402489.5
	50%	267998.7	268344.5	268364.2
	25%	133955.5	134186.9	134172.6
OLTP	100%	2502064.8	2494861.9	2493289.9
	87.5%	2200512.0	2193606.4	2192774.6
	75%	1886025.2	1880256.9	1879690.8
	62.5%	1571735.1	1566753.5	1566309.7
	50%	1257270.1	1253632.3	1253013.2
	37.5%	943076.4	940038.1	939605.4
	25%	628679.3	626711.5	626525.7
	12.5%	314447.8	313322.0	313308.9
SHA256	100%	55950.4	55464.7	56865.1
	75%	41985.8	41606.5	42653.8
	50%	28029.3	27734.5	28428.0
	25%	14009.7	13878.7	14236.8
SOR	100%	2380.0	2379.0	2379.9
	75%	1806.0	1802.3	1798.0
	50%	1209.1	1202.2	1202.8
	25%	611.3	604.7	613.3
SORT	100%	2884436.1	2841725.0	2889178.2
	75%	2163473.4	2131284.5	2167269.5
	50%	1442162.3	1420845.4	1444642.9
	25%	720985.5	710416.5	722417.2
CACHE	high	706937.7	704158.2	704546.9
	low	143082.2	142513.4	142822.0
STREAM	100%	5600.7	5599.6	5601.1
	50%	2800.4	2799.8	2800.5
RANDOM	100%	59574.0	58755.5	58774.7
	50%	29787.2	29378.1	29387.6
SEQUENTIAL	100%	22510.6	22498.5	22061.6
	50%	11255.4	11249.4	11030.8

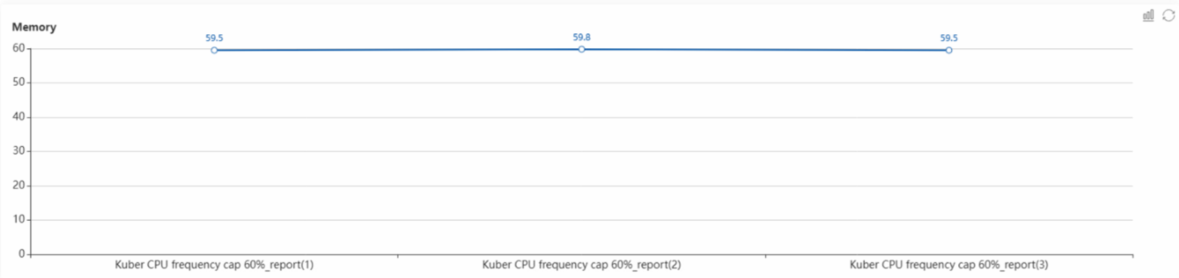
Appendix C.4 – CPU 80% Frequency: Raw Performance Power Ratio Comparison

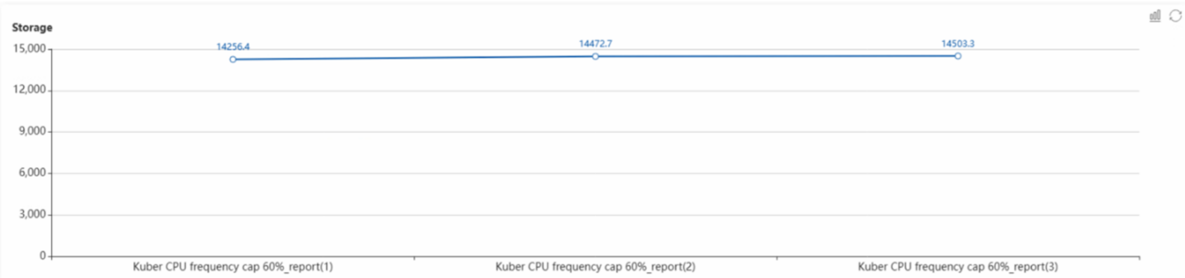
Workload	Workload Level	Test 1	Test 2	Test 3
AES	100%	84.0	82.9	83.1
	75%	88.6	91.4	88.0
	50%	67.7	68.8	68.5
	25%	39.2	40.0	40.1
COMPRESS	100%	543.2	546.2	546.5
	75%	459.9	468.5	468.9
	50%	380.8	395.6	391.2
	25%	241.2	249.7	250.2
LU	100%	382.7	383.7	383.3
	75%	335.9	360.4	378.2
	50%	309.0	341.7	339.9
	25%	201.2	211.8	212.4
OLTP	100%	447.0	447.9	447.1
	87.5%	399.0	406.3	404.2
	75%	355.8	358.5	359.1
	62.5%	317.3	336.3	332.2
	50%	319.3	334.2	342.6
	37.5%	268.1	280.8	282.4
	25%	194.8	201.4	200.6
	12.5%	107.6	108.4	108.6
SHA256	100%	637.4	637.0	641.9
	75%	530.6	533.3	543.8
	50%	435.6	442.5	429.9
	25%	267.4	274.3	281.9
SOR	100%	415.5	421.6	417.1
	75%	406.0	413.6	417.7
	50%	313.1	316.9	326.1
	25%	186.9	189.0	198.9
SORT	100%	897.4	889.1	896.4
	75%	733.0	727.4	746.7
	50%	587.8	611.1	682.3
	25%	404.0	426.2	418.3
CACHE	high	56.8	56.5	57.1
	low	11.2	11.2	11.3
STREAM	100%	107.4	107.3	107.4
	50%	83.4	83.8	86.6
RANDOM	100%	47135.1	48124.9	50792.1
	50%	24135.2	24557.0	25742.1
SEQUENTIAL	100%	7706.3	7896.4	8197.5
	50%	4029.3	4123.5	4220.6

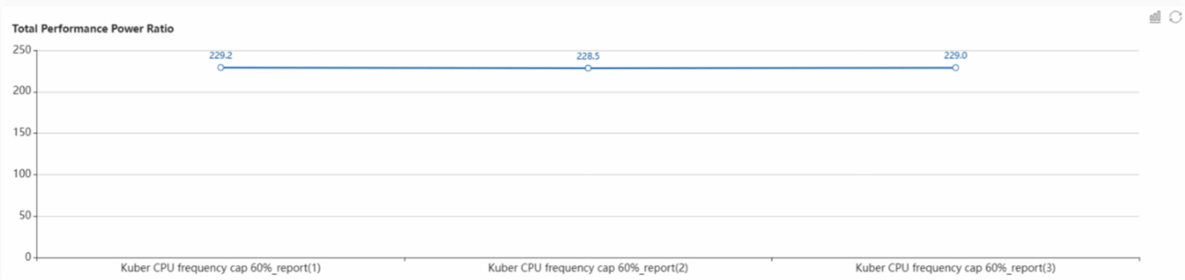
Appendix D.1 – CPU 60% Frequency: Performance Power Ratio Comparison of Each Component

Component	Test 1	Test 2	Test 3
CPU	**310.8**	308.5	309.9
Memory	59.5	**59.8**	59.5
Storage	14256.4	14472.7	**14503.3**
Total Performance Power Ratio	**229.2**	228.5	229.0









Appendix D.2 – CPU 60% Frequency: Energy Efficiency Score Comparison of Each Workload

Workload	Workload Level	Test 1	Test 2	Test 3
AES	100%	3.8	3.8	3.8
	75%	3.4	3.4	3.4
	50%	2.5	2.4	2.4
	25%	1.4	1.4	1.4
COMPRESS	100%	21.6	21.7	21.6
	75%	17.6	17.9	17.8
	50%	13.2	13.5	13.4
	25%	7.6	7.8	7.7
LU	100%	17.7	17.7	17.7
	75%	14.5	14.6	14.6
	50%	11.9	12.0	12.0
	25%	7.0	7.0	7.0
OLTP	100%	16.3	16.2	16.3
	87.5%	14.5	14.5	14.6
	75%	12.8	12.8	12.8
	62.5%	11.0	11.0	11.0
	50%	9.7	9.7	9.7
	37.5%	8.0	7.9	8.0
	25%	5.7	5.5	5.6
	12.5%	3.0	3.0	3.0
SHA256	100%	24.8	25.1	24.7
	75%	20.0	20.3	20.0
	50%	14.8	15.1	14.9
	25%	8.3	8.4	8.3
SOR	100%	19.3	19.3	19.3
	75%	16.9	17.0	16.9
	50%	12.5	12.7	12.6
	25%	7.2	7.2	7.2
SORT	100%	35.0	33.1	34.2
	75%	28.1	26.5	27.5
	50%	21.4	19.4	20.5
	25%	12.3	11.1	11.7
CACHE	high	2.3	2.3	2.3
	low	0.5	0.5	0.5
STREAM	100%	4.7	4.7	4.7
	50%	3.1	3.1	3.1
RANDOM	100%	1775.0	1805.8	1805.5
	50%	901.4	919.1	916.6
SEQUENTIAL	100%	289.6	293.1	294.7
	50%	149.7	151.5	152.4

Appendix D.3 – CPU 60% Frequency: Raw Performance Score Comparison of Each Workload

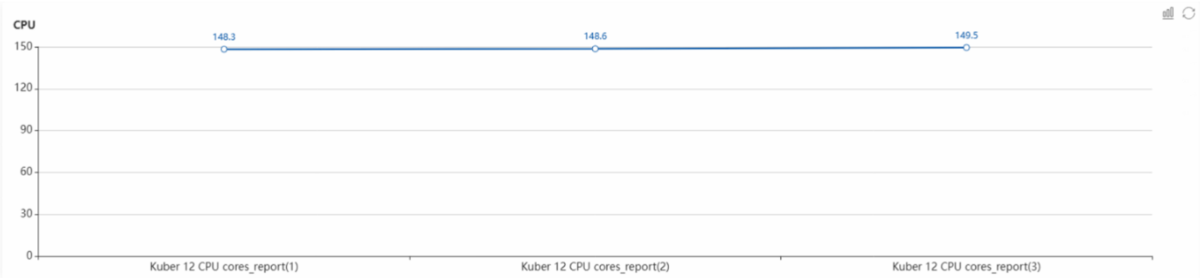
Workload	Workload Level	Test 1	Test 2	Test 3
AES	100%	474004.1	473107.0	474061.7
	75%	355537.7	354878.0	355578.4
	50%	237198.1	236597.7	236993.9
	25%	118561.2	118241.8	118476.3
COMPRESS	100%	13248.9	13385.1	13265.4
	75%	9952.5	10053.7	9963.6
	50%	6632.2	6698.8	6651.7
	25%	3340.0	3366.4	3338.4
LU	100%	493746.9	494023.6	494082.1
	75%	370261.2	370463.8	370554.1
	50%	246885.1	247088.4	247143.2
	25%	123504.6	123533.3	123532.5
OLTP	100%	2350362.4	2342013.4	2339541.6
	87.5%	2063352.2	2059925.9	2055708.4
	75%	1768744.6	1765676.3	1761901.7
	62.5%	1473808.1	1471266.9	1468111.2
	50%	1179000.8	1177200.1	1174649.9
	37.5%	884415.0	882855.1	881007.0
	25%	589680.5	588510.9	587242.4
	12.5%	294791.3	294240.3	293600.4
SHA256	100%	41970.5	42572.0	41946.7
	75%	31536.3	31925.9	31474.6
	50%	21016.8	21306.2	21014.1
	25%	10526.0	10655.3	10509.9
SOR	100%	2366.6	2366.8	2361.3
	75%	1785.1	1796.3	1781.9
	50%	1192.9	1202.1	1199.3
	25%	607.2	606.3	600.7
SORT	100%	2107251.9	1981832.3	2053871.9
	75%	1580531.5	1486489.2	1540316.6
	50%	1053776.7	991212.1	1026848.9
	25%	526858.4	495549.5	513487.9
CACHE	high	655445.4	654663.6	653318.3
	low	135493.5	134947.9	134792.9
STREAM	100%	5600.7	5600.9	5603.2
	50%	2800.4	2800.5	2801.6
RANDOM	100%	58831.3	59434.6	59520.7
	50%	29416.0	29717.6	29760.8
SEQUENTIAL	100%	22181.9	22241.8	22459.7
	50%	11091.1	11121.0	11229.9

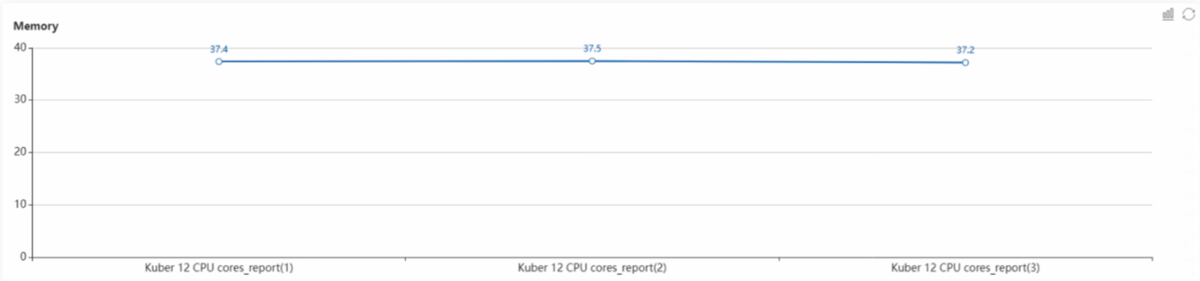
Appendix D.4 – CPU 60% Frequency: Raw Performance Power Ratio Comparison

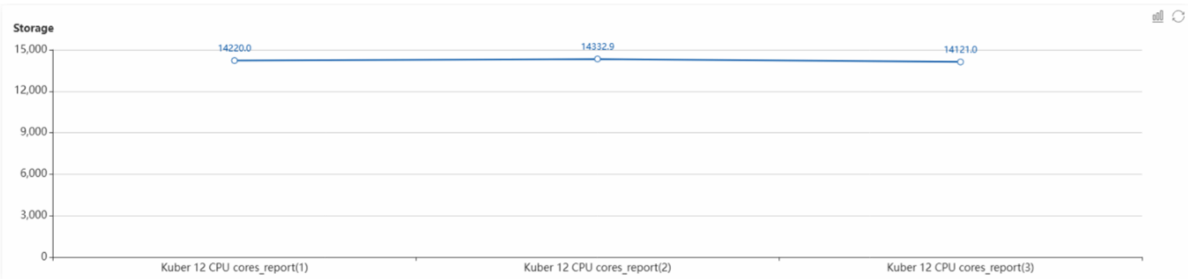
Workload	Workload Level	Test 1	Test 2	Test 3
AES	100%	106.9	106.9	106.8
	75%	93.5	94.3	94.8
	50%	68.3	67.6	67.5
	25%	38.8	38.4	38.9
COMPRESS	100%	600.0	604.2	601.0
	75%	490.0	496.4	493.4
	50%	368.0	374.5	372.7
	25%	212.3	215.9	214.2
LU	100%	491.6	492.4	492.2
	75%	403.0	406.3	405.8
	50%	330.9	334.6	333.0
	25%	193.5	193.1	193.7
OLTP	100%	541.9	541.7	543.7
	87.5%	484.2	485.0	486.3
	75%	425.7	426.1	426.4
	62.5%	366.6	367.2	366.3
	50%	324.2	323.3	323.1
	37.5%	266.4	263.0	266.1
	25%	188.7	183.0	185.7
	12.5%	100.1	99.0	100.2
SHA256	100%	687.6	696.5	687.0
	75%	555.4	562.9	556.0
	50%	411.0	418.3	413.7
	25%	230.9	233.1	231.5
SOR	100%	535.4	535.7	535.7
	75%	469.9	472.3	470.1
	50%	347.8	351.4	351.1
	25%	199.7	200.5	198.7
SORT	100%	971.5	920.2	950.0
	75%	779.7	736.0	763.0
	50%	593.4	537.7	568.8
	25%	341.6	308.3	326.0
CACHE	high	70.6	70.9	70.7
	low	14.3	14.3	14.3
STREAM	100%	136.3	136.6	136.6
	50%	90.9	91.9	90.9
RANDOM	100%	49307.0	50160.9	50153.2
	50%	25040.3	25530.4	25462.5
SEQUENTIAL	100%	8043.6	8141.2	8186.7
	50%	4159.4	4208.1	4232.2

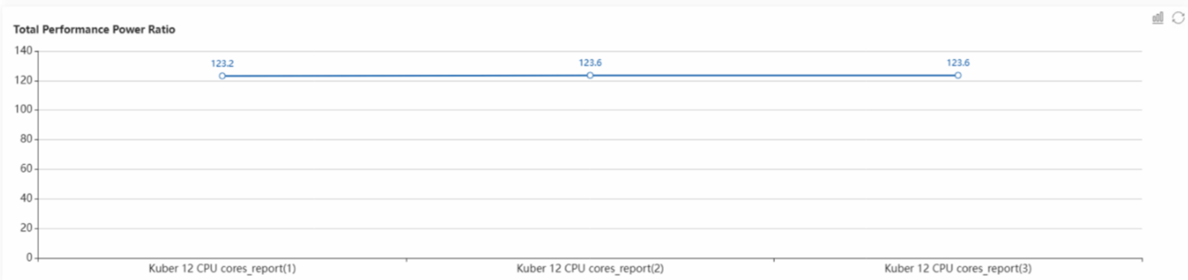
Appendix E.1 – 12 CPU Cores: Performance Power Ratio Comparison of Each Component

Component	Test 1	Test 2	Test 3
CPU	148.3	148.6	**149.5**
Memory	37.4	**37.5**	37.2
Storage	14220.0	**14332.9**	14121.0
Total Performance Power Ratio	123.2	**123.6**	**123.6**









Appendix E.2 – 12 CPU Cores: Energy Efficiency Score Comparison of Each Workload

Workload	Workload Level	Test 1	Test 2	Test 3
AES	100%	2.6	2.6	2.6
	75%	1.9	1.9	1.9
	50%	1.4	1.4	1.4
	25%	1.2	1.2	1.2
COMPRESS	100%	9.4	9.4	9.4
	75%	7.2	7.3	7.3
	50%	5.2	5.2	5.1
	25%	3.5	3.6	3.5
LU	100%	9.8	9.9	9.8
	75%	7.4	7.4	7.4
	50%	5.2	5.2	5.2
	25%	4.3	4.5	4.2
OLTP	100%	6.8	6.6	6.8
	87.5%	5.9	5.8	5.9
	75%	5.2	5.0	5.1
	62.5%	4.1	4.1	4.2
	50%	3.5	3.5	3.5
	37.5%	2.8	2.8	2.8
	25%	2.1	2.1	2.1
	12.5%	1.6	1.6	1.5
SHA256	100%	11.4	11.3	11.3
	75%	8.5	8.4	8.5
	50%	6.0	5.9	6.0
	25%	4.3	4.3	4.3
SOR	100%	12.3	12.4	12.6
	75%	10.1	10.1	10.1
	50%	7.7	7.7	7.7
	25%	5.3	5.4	5.3
SORT	100%	13.6	13.7	14.5
	75%	10.1	10.1	10.8
	50%	6.6	6.6	7.1
	25%	3.5	3.5	3.7
CACHE	high	1.2	1.2	1.2
	low	0.3	0.3	0.3
STREAM	100%	3.6	3.6	3.6
	50%	2.5	2.6	2.5
RANDOM	100%	1740.9	1773.4	1714.1
	50%	880.8	900.6	913.8
SEQUENTIAL	100%	295.1	293.2	287.6
	50%	151.7	151.4	148.2

Appendix E.3 – 12 CPU Cores: Raw Performance Score Comparison of Each Workload

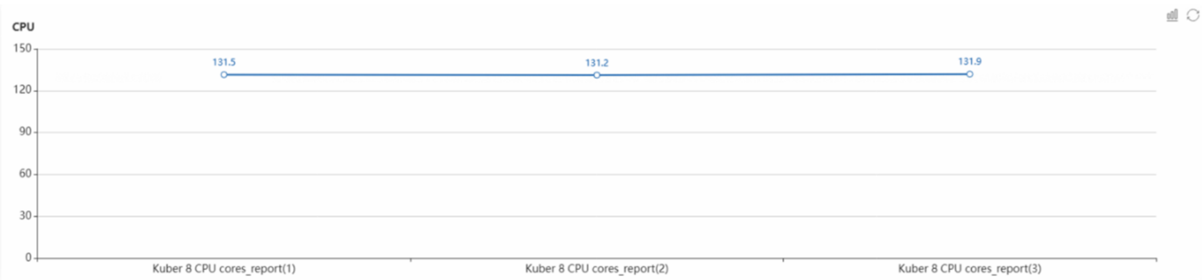
Workload	Workload Level	Test 1	Test 2	Test 3
AES	100%	451790.9	451614.2	452310.6
	75%	338914.2	338766.0	339323.2
	50%	225984.1	225856.9	226242.0
	25%	112979.9	112959.7	113108.9
COMPRESS	100%	7512.0	7502.2	7509.8
	75%	5655.0	5635.2	5656.8
	50%	3782.9	3763.5	3775.3
	25%	1894.0	1894.6	1898.3
LU	100%	339696.0	339853.0	339140.2
	75%	254831.7	254818.5	254412.6
	50%	169936.0	169923.8	169555.5
	25%	84935.9	85013.7	84778.1
OLTP	100%	1412043.4	1378153.4	1410076.0
	87.5%	1240024.8	1214080.3	1240606.4
	75%	1063007.4	1040764.7	1063272.0
	62.5%	885736.1	867346.1	886054.7
	50%	708492.5	693876.6	708735.8
	37.5%	531493.5	520306.0	531655.8
	25%	354307.2	346885.8	354413.4
	12.5%	177219.3	173482.5	177248.6
SHA256	100%	25450.3	25430.8	25493.0
	75%	19119.0	19087.2	19111.1
	50%	12735.6	12737.9	12757.6
	25%	6386.6	6367.1	6389.6
SOR	100%	2147.3	2146.9	2149.1
	75%	1630.4	1630.7	1614.5
	50%	1091.5	1091.7	1088.8
	25%	548.1	548.8	554.2
SORT	100%	1058818.7	1057014.6	1147675.6
	75%	794236.7	792787.6	860755.1
	50%	529383.0	528453.9	573834.8
	25%	264762.0	264158.1	287040.4
CACHE	high	433522.6	432944.4	433624.8
	low	89150.2	89174.9	89362.4
STREAM	100%	6182.1	6181.9	6174.8
	50%	3091.0	3090.9	3087.4
RANDOM	100%	60759.6	61371.4	63777.9
	50%	30380.3	30685.9	31889.2
SEQUENTIAL	100%	23801.4	23538.6	23507.6
	50%	11900.7	11769.3	11753.9

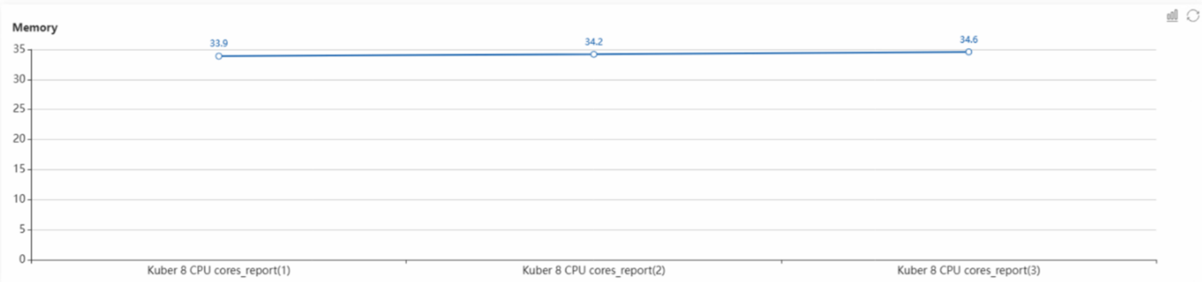
Appendix E.4 – 12 CPU Cores: Raw Performance Power Ratio Comparison

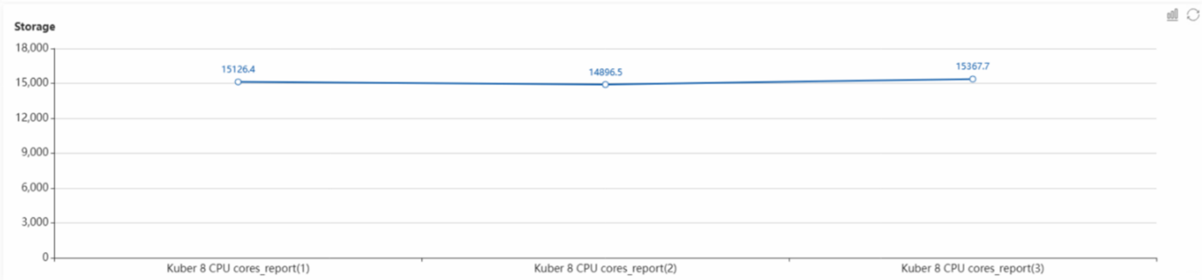
Workload	Workload Level	Test 1	Test 2	Test 3
AES	100%	71.2	71.5	73.1
	75%	53.3	53.2	53.8
	50%	39.5	39.6	39.4
	25%	32.4	34.0	32.0
COMPRESS	100%	262.0	260.3	261.2
	75%	200.9	201.4	201.5
	50%	144.4	145.3	142.7
	25%	96.1	99.0	96.0
LU	100%	271.7	274.1	273.0
	75%	204.6	204.8	205.7
	50%	145.7	144.5	143.1
	25%	119.3	125.5	117.6
OLTP	100%	225.4	219.9	227.4
	87.5%	197.9	193.2	197.1
	75%	173.5	166.9	170.3
	62.5%	138.3	136.4	139.7
	50%	117.0	115.1	116.7
	37.5%	94.6	93.1	94.2
	25%	70.5	69.1	69.5
	12.5%	54.8	54.6	50.5
SHA256	100%	317.8	312.8	312.7
	75%	235.7	234.0	236.6
	50%	166.0	165.0	165.6
	25%	119.8	119.2	118.4
SOR	100%	343.0	344.2	351.2
	75%	281.8	281.2	279.8
	50%	213.8	215.1	214.4
	25%	146.9	149.1	147.4
SORT	100%	378.2	379.5	403.9
	75%	280.7	280.4	300.9
	50%	184.1	184.4	198.5
	25%	96.1	96.0	103.2
CACHE	high	34.8	34.9	35.1
	low	7.3	7.3	7.2
STREAM	100%	103.0	103.4	102.8
	50%	74.2	75.8	73.9
RANDOM	100%	48359.4	49261.3	47613.2
	50%	24468.3	25017.7	25382.7
SEQUENTIAL	100%	8198.5	8144.0	7989.9
	50%	4214.9	4204.7	4117.7

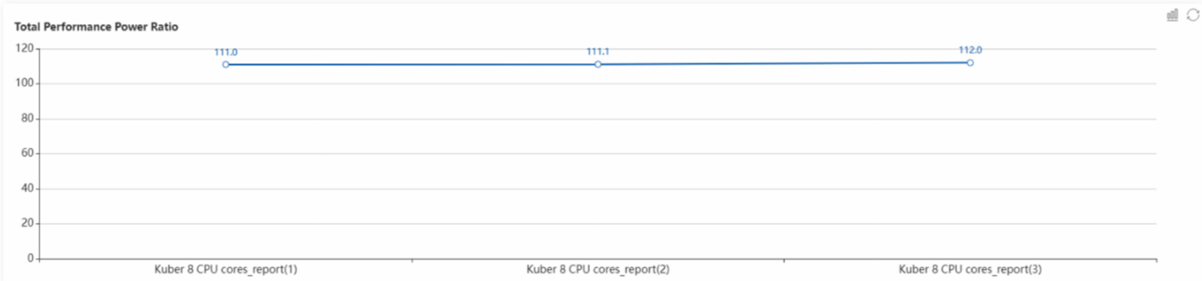
Appendix F.1 – 8 CPU Cores: Performance Power Ratio Comparison of Each Component

Component	Test 1	Test 2	Test 3
CPU	131.5	131.2	**131.9**
Memory	33.9	34.2	**34.6**
Storage	15126.4	14896.5	**15367.7**
Total Performance Power Ratio	111.0	111.1	**112.0**









Appendix F.2 – 8 CPU Cores: Energy Efficiency Score Comparison of Each Workload

Workload	Workload Level	Test 1	Test 2	Test 3
AES	100%	2.3	2.3	2.3
	75%	1.9	1.8	1.8
	50%	1.4	1.4	1.4
	25%	1.1	1.0	1.1
COMPRESS	100%	7.1	7.0	7.0
	75%	5.7	5.6	5.7
	50%	4.3	4.3	4.3
	25%	2.8	2.7	2.8
LU	100%	7.5	7.6	7.6
	75%	5.9	6.0	6.0
	50%	4.6	4.6	4.6
	25%	3.6	3.6	3.6
OLTP	100%	6.7	6.7	6.8
	87.5%	6.1	6.0	6.2
	75%	5.5	5.4	5.5
	62.5%	4.8	4.7	4.9
	50%	4.1	4.0	4.1
	37.5%	3.3	3.2	3.3
	25%	2.7	2.4	2.7
	12.5%	1.7	1.6	1.7
SHA256	100%	8.5	8.4	8.4
	75%	6.7	6.7	6.7
	50%	5.0	5.0	5.0
	25%	3.3	3.2	3.3
SOR	100%	10.9	10.8	10.9
	75%	9.1	9.0	9.2
	50%	7.1	6.9	7.0
	25%	4.6	4.5	4.6
SORT	100%	10.1	11.0	10.3
	75%	8.1	8.8	8.1
	50%	6.0	6.6	6.1
	25%	4.3	4.5	4.3
CACHE	high	1.0	1.0	1.0
	low	0.2	0.2	0.2
STREAM	100%	4.0	4.0	4.0
	50%	2.6	2.6	2.6
RANDOM	100%	1907.2	1881.5	1921.7
	50%	967.7	964.5	997.2
SEQUENTIAL	100%	303.4	295.5	307.2
	50%	157.0	154.2	159.1

Appendix F.3 – 8 CPU Cores: Raw Performance Score Comparison of Each Workload

Workload	Workload Level	Test 1	Test 2	Test 3
AES	100%	380987.4	381313.5	378220.1
	75%	285733.0	286027.7	283605.6
	50%	190511.1	190655.7	189075.9
	25%	95296.3	95294.7	94605.4
COMPRESS	100%	5327.1	5335.1	5336.4
	75%	3997.0	3996.6	4010.4
	50%	2666.9	2671.3	2680.2
	25%	1336.6	1340.7	1339.9
LU	100%	245365.1	249587.5	245582.5
	75%	184060.6	187247.7	184148.7
	50%	122682.3	124807.5	122851.3
	25%	61302.4	62404.6	61399.0
OLTP	100%	1239257.0	1240689.2	1248601.1
	87.5%	1087569.9	1089468.3	1096346.8
	75%	932245.4	933836.7	939536.5
	62.5%	776824.5	778113.3	783020.6
	50%	621500.8	622565.1	626522.4
	37.5%	465980.2	466897.3	469761.6
	25%	310685.6	311286.3	313129.7
	12.5%	155365.5	155634.7	156590.6
SHA256	100%	17961.9	17942.5	17868.4
	75%	13496.8	13464.8	13396.0
	50%	8991.0	8956.3	8942.8
	25%	4490.3	4480.4	4467.7
SOR	100%	1803.0	1790.8	1790.5
	75%	1358.9	1349.7	1345.5
	50%	906.3	900.2	899.1
	25%	452.0	455.0	450.5
SORT	100%	746376.6	832394.0	765262.0
	75%	559827.7	624484.7	573933.4
	50%	373156.0	416264.3	382788.0
	25%	186579.6	208020.1	191360.9
CACHE	high	321084.4	326804.2	325040.2
	low	67361.9	68492.0	68275.0
STREAM	100%	5808.7	5809.7	5808.9
	50%	2904.4	2904.9	2904.5
RANDOM	100%	65619.3	66006.3	65774.1
	50%	32810.0	33003.5	32887.6
SEQUENTIAL	100%	24338.2	24459.7	24464.7
	50%	12169.2	12229.9	12232.5

Appendix F.4 – 8 CPU Cores: Raw Performance Power Ratio Comparison

Workload	Workload Level	Test 1	Test 2	Test 3
AES	100%	64.1	63.4	62.9
	75%	51.4	50.8	50.9
	50%	39.6	39.1	39.2
	25%	30.4	29.0	30.1
COMPRESS	100%	196.9	194.1	194.8
	75%	159.6	154.9	159.5
	50%	119.6	118.2	120.3
	25%	78.1	75.2	78.5
LU	100%	208.4	212.4	211.2
	75%	165.3	167.9	168.0
	50%	126.6	127.4	127.2
	25%	99.4	98.9	100.2
OLTP	100%	223.3	222.3	225.9
	87.5%	203.3	201.4	205.8
	75%	182.5	180.3	184.0
	62.5%	160.5	157.9	161.9
	50%	136.1	133.7	137.5
	37.5%	109.0	107.0	109.8
	25%	89.5	80.9	90.8
	12.5%	55.4	54.0	55.7
SHA256	100%	235.8	232.4	234.1
	75%	187.1	184.9	186.9
	50%	139.1	137.6	139.5
	25%	91.6	89.3	91.8
SOR	100%	303.5	300.7	301.9
	75%	251.6	250.1	255.5
	50%	197.7	192.1	195.7
	25%	128.9	126.1	128.4
SORT	100%	279.5	306.2	286.5
	75%	223.6	243.6	226.4
	50%	166.7	182.4	169.9
	25%	118.8	125.1	119.4
CACHE	high	27.5	27.9	28.5
	low	5.7	5.8	5.9
STREAM	100%	113.5	113.5	113.8
	50%	74.7	73.6	75.8
RANDOM	100%	52979.1	52263.3	53379.2
	50%	26881.0	26792.6	27701.0
SEQUENTIAL	100%	8428.2	8208.0	8532.7
	50%	4361.7	4284.4	4420.6