



Energy Efficient

# Performance Based TCO

## Redefining TCO for Enterprise Storage and Hyperscale Data Centers with the MaxLinear Panther Accelerator



Lower TCO

Improved Performance

WHITE PAPER

### Authors:

**Nader Salessi**  
Advisor

**Pinaki Chanda**  
Senior Director of Software Engineering

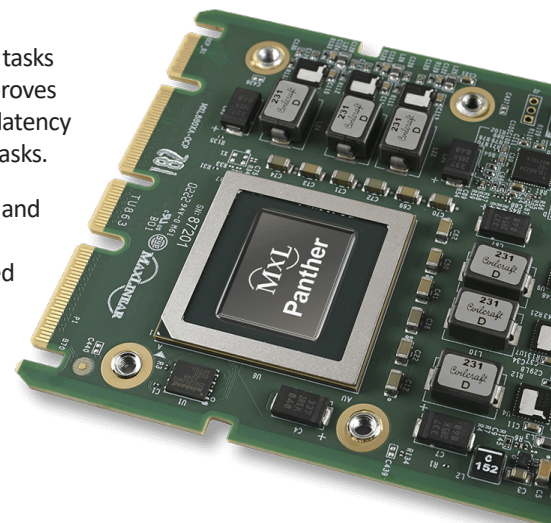
### Introduction

With data volumes growing exponentially, enterprise storage and hyperscale data centers are under increasing pressure to improve performance, energy efficiency, and reduce total cost of ownership (TCO). Traditional architectures that rely solely on general-purpose CPUs can no longer meet the compute and energy demands of modern workloads—especially those involving compression algorithms like GZIP, ZLIB, and XP10. MaxLinear’s Panther hardware accelerator redefines the TCO equation by delivering significant performance and efficiency gains through offloaded, high-throughput data compression.

### Introduction to Panther for AI/Inference Workloads

As AI-driven applications and inference workloads continue to grow, the need for scalable, high-performance storage infrastructure becomes increasingly critical. Offloading compute-intensive tasks such as compression, security, and dedupe hash computation to an accelerator like Panther improves overall system performance and efficiency. This is especially beneficial in high-throughput, low-latency environments, where it allows CPU cores or other accelerators to remain focused on inference tasks.

For example, Retrieval-Augmented Generation (RAG) systems rely on large-scale vector databases and document repositories to deliver relevant context for language model inference, making efficient storage a key enabler of scalability and performance. Storage accelerators with hardware-optimized compression and decompression can rapidly unpack compressed data during query execution, significantly reducing retrieval latency. This approach not only lowers storage costs and relieves memory bandwidth pressure during LLM inference but also maximizes the throughput capabilities of NVMe storage. The result is faster document retrieval, improved resource efficiency, and scalable, high-performance RAG deployments for enterprise environments.



To evaluate Panther’s effectiveness, three scenarios were examined:

- No compression vs. compression with Panther
- Software-based light compression (e.g., LZ4) vs. Panther
- Software-based compression vs. Panther

In all cases, Panther significantly reduced workload processing time while improving energy efficiency and usable storage capacity. Additionally, Panther freed up CPU cores to run revenue-generating customer applications, including but not limited to AI inference for various AI use cases.

The following sections detail the results of a case study comparing a software-based GZIP compression algorithm with Panther-accelerated GZIP compression.

### A New TCO Model: Performance-Driven Efficiency

Traditional TCO models have primarily focused on hardware costs and power consumption. However, a modernized TCO framework must also account for Time to Completion (TTC) and overall system utilization, critical factors that directly impact infrastructure efficiency and operational costs.

Panther excels in both areas by significantly accelerating compression workloads and freeing up host CPU resources.

### Benchmark Results

#### Energy Consumption

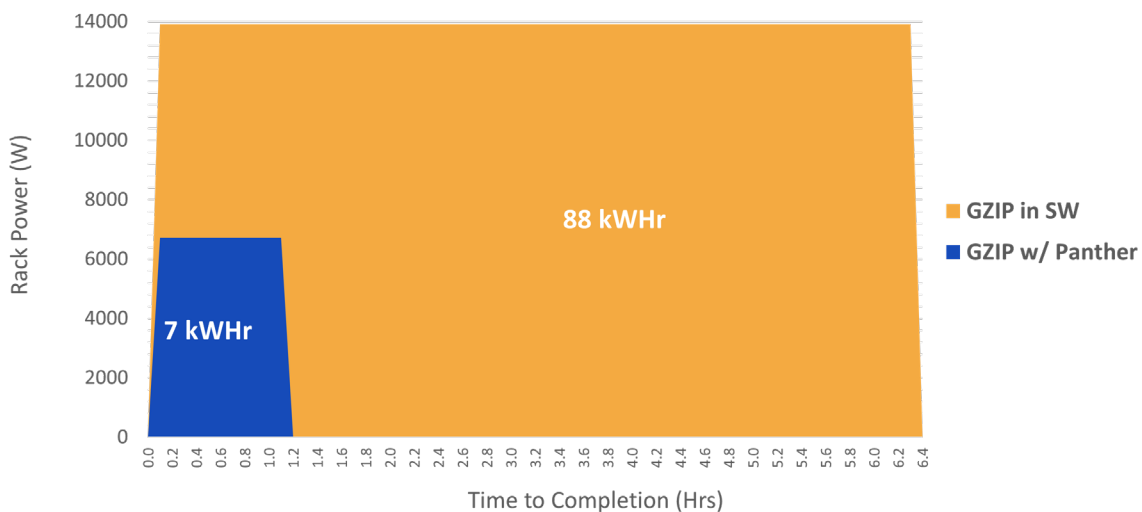
In benchmark testing of a 100TB GZIP compression workload within a rack:

- The rack with Panther Accelerators completed the task 5.7 times faster (1.11 hour versus 6.35 hours)
- The software-based rack consumed 88 KWHr, while the Panther-enabled rack consumed only 7kWhr resulting in 82% reduction in energy consumption.

**Panther accelerates workload by 5.7X and cuts energy use by 82%**

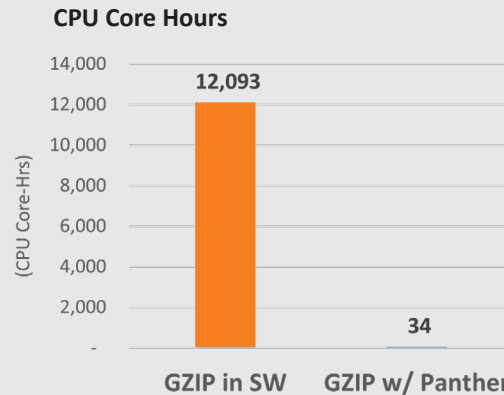
These results highlight Panther’s ability to deliver dramatically improved performance gains while minimizing resource consumption.

Energy Consumption for Time to Completion of 100TB

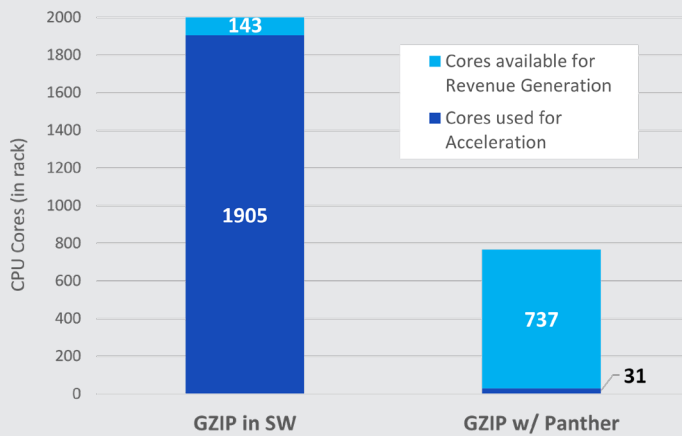


### Exceptional CPU Efficiency

Even more striking, total CPU core-hours dropped from 12,093 to just 34, illustrating Panther's ability to handle compute-intensive workloads with minimal CPU involvement.



### Core Utilization (Rack Level)

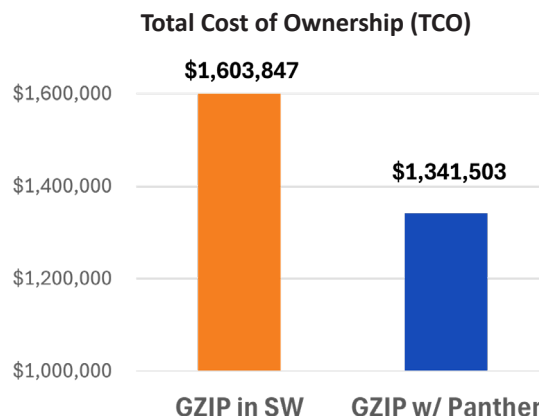


Panther utilized only 4% of rack CPU cores (31 cores), compared to 94% (1905 cores) used by software-based solutions. This left 5 times (5X) more cores available for other revenue-generating tasks (737 vs. 143). These improvements translate into better infrastructure utilization, reduced latency, and lower operating costs—core drivers of a next-generation TCO model.

### Capital and Operational Cost Reduction

Panther delivers substantial savings in both capital and operational expenditure resulting in lower total cost of ownership:

- Total Cost of Ownership (TCO): By accelerating workload completion and reducing CPU dependency, data centers can achieve the same throughput with fewer servers or fewer cores per server. Resulting in over \$260k in TCO savings.
- Capital Expenditure (CapEx): Racks with Panther accelerator cards reduce Capex by 13%, thanks to the use of less expensive servers with lower core counts.
- Operational Expenditure (OpEx): Panther also cuts power and cooling requirements, leading to a 36% reduction in OpEx.



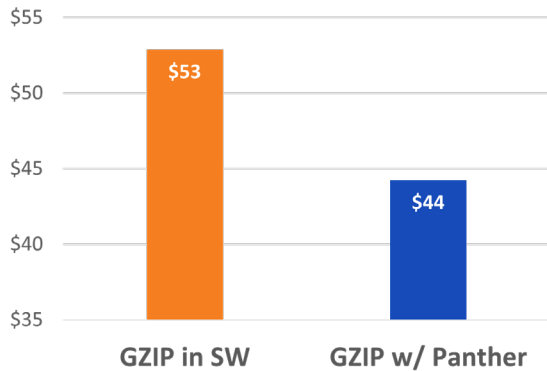
**Panther**  
Saves **\$262k** in TCO

Combined, these savings contribute to a more than 16% reduction in total TCO per rack, making Panther a highly cost-effective solution for an economically compelling option for large-scale deployments.

### TCO per TBe: A Strategic Metric

For service providers and hyperscalers, TCO per terabyte effective (TBe) is a critical metric. Panther reduces this by 16%, enabling more cost-effective deployment of storage-as-a-service platforms and optimized handling of cold and warm data tiers. This not only boosts ROI but also allows for more competitive service pricing.

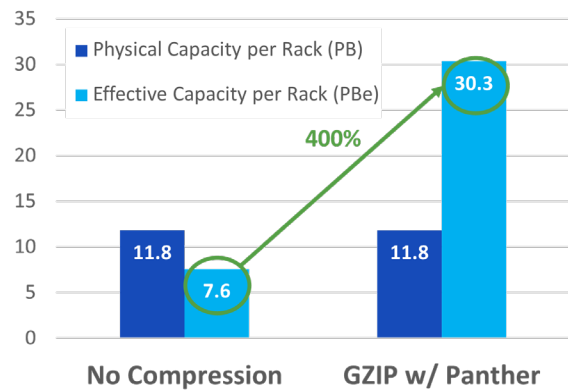
TCO (OpEx + CapEx) per TBe



### Maximizing Rack-Level Storage Efficiency

Panther's benefits extend beyond compute performance; they also transform storage efficiency. Its real-time hardware-based compression enables up to a 400% increase in usable storage capacity, allowing a single rack to support up to 30PB of effective data. This dramatically reduces physical footprint and infrastructure complexity, which is especially valuable for AI/ML, analytics, and backup workloads in hyperscale environments.

Capacity (Physical & Effective) per Rack



<p><b>Panther Transforms Storage Efficiency</b></p>	<p><b>400% more usable storage capacity</b> Store up to <b>30PB</b> of effective data per rack</p>
---	--

### Conclusion

MaxLinear's Panther accelerator redefines enterprise storage and hyperscale data center economics by delivering faster workload completion, superior energy efficiency, and greater storage density. By integrating performance metrics like TTC and system utilization into the TCO model, Panther demonstrates that hardware acceleration is essential for modern, scalable infrastructure. For operators aiming to reduce costs while improving scalability and sustainability, Panther offers a clear and compelling strategic advantage.



**MaxLinear, Inc.**  
 5966 La Place Court, Suite 100  
 Carlsbad, CA 92008  
 Tel.: +1 (760) 692-0711  
 Fax: +1 (760) 444-8598  
[www.maxlinear.com](http://www.maxlinear.com)

The content of this document is furnished for informational use only, is subject to change without notice, and should not be construed as a commitment by MaxLinear, Inc. MaxLinear, Inc. assumes no responsibility or liability for any errors or inaccuracies that may appear in the informational content contained in this document.

Reproduction, in part or whole, without the prior written consent of MaxLinear, Inc. is prohibited. MaxLinear, the MaxLinear logo, and any other MaxLinear trademarks (including but not limited to MXL, Full-Spectrum Capture, FSC, AirPHY, Puma, and AnyWAN, VectorBoost, MXL WARE, and Panther) are all property of MaxLinear, Inc. or one of MaxLinear's subsidiaries in the U.S.A. and other countries. All rights reserved. All third-party marks and logos are trademarks™ or registered® trademarks of their respective holders/owners.

© 2025 MaxLinear, Inc. All rights reserved.