

SIGNAL65 LAB INSIGHT

Building Power Efficient AI Data Centers with Solidigm QLC SSDs

AUTHOR

Mitch Lewis

Research Analyst | Signal65

IN PARTNERSHIP WITH



DECEMBER 2024

Executive Summary

AI has become a top priority for organizations due to its vast potential for innovation. Building AI data centers, however, presents significant challenges. AI is both computationally and data intensive, involving large infrastructure requirements and consequently, large power requirements. The massive amount of power required to support new AI data centers is a key challenge, adding cost, sustainability concerns, and limiting the total infrastructure that can be deployed in a single data center.

While energy-related concerns around AI are often focused on the widespread use of GPUs, the data storage required to store large scale AI training datasets and model checkpoints additionally have a significant impact on data center power efficiency. This study focuses on the role of network-attached data storage in AI datacenters, and evaluates how different storage media can impact power efficiency.

The goal of this study was to model a new, 100 Megawatt AI datacenter and evaluate the impact that different storage devices had on the total power efficiency. The study specifically evaluated the impact of QLC SSDs, TLC SSDs, and a hybrid HDD based deployment.

Key findings of the study include:

19.5%
greater power
efficiency for
QLC SSDs
compared to
TLC SSDs

79.5%
greater power
efficiency for
QLC SSDs
compared to
hybrid TLC SSDs
and HDDs

QLC SSDs enable
more total AI
infrastructure within
the same datacenter:
1.6%
more than TLC SSDs and
26.3%
more than HDDs

AI, Energy, and Storage

Recent advancements in the field of AI have led to a renewed interest in the technology. AI applications have become a key priority across virtually all industries, and have the potential to drive significant innovation.

While the potential for innovation is large, building and deploying these new AI applications is a challenge. AI deployments are incredibly resource intensive, requiring significant compute resources and massive data capacities. Computationally, modern AI models rely on large numbers of GPUs to parallelize and expedite the training processes. When considering the data storage requirements, there are two key challenges. First, AI models require massive amounts of training data to achieve a high level of accuracy. Additionally, consistent model checkpoints need to be stored throughout the training process to save the state of the model. This results in extremely high storage capacity requirements for AI data centers.

Both the high compute and storage requirements lead to another key challenge when building an AI datacenter: power. The large power demand of AI infrastructure has become a key concern for organizations building AI data centers. The energy consumption of AI data centers can add significant cost, derail sustainability goals, and ultimately become a limiting factor for deploying AI infrastructure. In some cases, such as for large hyper-scalers, the power requirements have grown so extreme that they have begun investing in small nuclear reactors to power their AI data centers. While purchasing dedicated nuclear reactors may not be an option for the majority of data centers, AI-related power challenges remain. Organizations building AI data centers must understand the power requirements for AI and what their options are for building a more efficient data center.

The power within a data center can be broadly grouped into the following categories, compute, networking, storage, and non-IT infrastructure related power, primarily cooling. The majority of the concern around AI power requirements is often focused around compute due to the large-scale usage of GPUs. While it is true that GPUs are energy intensive, their role in AI training is irreplaceable. Training AI models is a computationally intensive and time consuming process that often cannot be achieved in a practical timeframe with reduced computational resources. Networking within an AI datacenter does not have the largest power draw, nor can it be radically changed to increase power efficiency as it is directly correlated to compute and storage scale. When considering non-IT infrastructure related power, cooling is a key area that may be altered to achieve greater power efficiency, with the use of liquid cooling. The challenge for liquid cooling, however, is that it requires entirely new infrastructure, and may add significant cost. This leaves data storage as a key area to optimize for power efficiency that can be achieved in a fairly straightforward manner.

Although GPU compute is responsible for the majority of the energy consumption in an AI data center, data storage plays a significant role as well. While the most powerful GPUs are typically necessary to accelerate training, organizations have greater flexibility when considering storage devices. Storage devices have seen an evolution over time from primarily HDD-based approaches, to solid state technology with increasing density. HDD-based approaches are typically chosen as an economic option for bulk storage, using a caching layer of SSDs to compensate for the lower performance of spinning disks. Modern all flash approaches on the other hand typically leverage either Tri-Level Cell (TLC) or Quad-Level Cell (QLC) devices, which each provide their own strengths and weaknesses. TLC devices typically offer greater performance at the expense of lower density. Alternatively, QLC devices provide ultra dense storage, with similar read performance and slightly reduced write performance, as compared to TLC devices.

The emergence of QLC has altered the typical dynamic of selecting storage devices, in which HDDs are a cost effective option for bulk storage and TLC SSDs are the clear choice for performance demanding workloads. The high density of QLC drives challenges the economic advantage of HDDs, while simultaneously offering flash performance, making them an attractive replacement for many typical HDD-based environments. The emergence of energy-related concerns has further positioned QLC as a replacement for HDDs, as SSDs are typically considered more energy-efficient than HDDs. Meanwhile, QLC also challenges the dominance of TLC devices in all-flash environments, especially for read-heavy workloads in which performance is highly competitive.

When considering data storage for AI, devices must balance performance, capacity, and energy efficiency. Training AI models typically demands petabyte-scale capacity demands, as more data helps build more accurate models, and large models must be consistently checkpointed and retained. Capacity alone, however, is not sufficient, as the storage must also meet strict performance demands to efficiently read data to the GPU servers. Layering on the requirement to optimize storage for power efficiency further complicates the task. When considering the storage requirements for AI, the mixture of high density, flash performance, and energy efficiency of QLC makes the technology an appealing choice for AI data centers.

About the Study

To address the emerging interest in AI alongside the increasing concern around the power consumption of AI infrastructure, this study evaluated the impact of storage devices on the power efficiency of an AI data center. To do so, Signal65 and Solidigm collaborated to model a realistic implementation of a 100 Megawatt AI data center to measure the impact that different storage devices have on overall power efficiency. Storage devices as a variable were chosen as they provide a practical option for IT decision makers to optimize power efficiency. Modelling was done with three distinct storage system configurations, a hybrid HDD-based solution, an all TLC SSD solution, and an all QLC SSD solution. Devices chosen for the evaluation included Solidigm high capacity QLC SSDs, competitive TLC SSDs, and competitive HDDs combined with a caching layer of Solidigm TLC SSDs.

Configuration Details

To isolate the impact of storage devices, all other variables were kept consistent across each environment. Each storage configuration was modeled around the following parameters:

- All infrastructure is contained within a 100MW AI data center.
- Compute is provided by NVIDIA DGX H100 clusters, with 4 servers per rack.
- AI training data is stored on network attached storage. Storage was modeled after a software defined storage solution with support for either all-flash or HDD based configurations. A 2x or 3x redundancy architecture was chosen based on the recommended configuration for the specific media¹.
- Storage software was deployed on commodity servers with connected JBODs or JBOFs, depending on the storage devices.

¹2x replication was selected for SSD environments and 3x replication was selected for HDD environments. Redundancy architecture chosen based on guidelines for Ceph. (Source: [Red Hat](#))

- Each storage configuration was modeled based on the total number of storage management server and JBOF or JBOD pairings required to meet pre-determined capacity points when populated with the various devices. In the hybrid HDD environment, 10% of the total capacity was reserved for the TLC SSD caching layer, with the remaining 90% deployed as HDDs in JBODs. For the SSD environments, the total number of devices required was divided between the storage management servers and JBOFs based on a ratio of the pairing's maximum capacity and the total required capacity.
- QLC SSD Environment
 - Solidigm D5-P5336 122.88 TB High Capacity QLCs SSDs used for all storage.
 - Software defined storage deployed on commodity servers with support for up to 12 drives.
 - JBOFs with up to 32 drives.
 - 2x redundancy architecture
- TLC SSD Environment
 - 61.44 TB TLC SSDs used for all storage
 - Software defined storage deployed on commodity servers with support for up to 12 drives.
 - JBOFs with 32 drives
 - 2x redundancy architecture
- Hybrid HDD Environment
 - Combination of 24 TB HDDs with 7.68 TB TLC SSDs used as a caching tier
 - Storage management servers with up to 12 SSDs
 - JBODs with support for up to 32 HDDs
 - 10% of total capacity stored in SSD caching tier
 - 3x redundancy architecture

Methodology and Key Assumptions

To evaluate the three storage environments in a realistic and fair way, key assumptions must be made and upheld throughout the modeling process. These assumptions provide a framework for the model and have been based on external research and industry knowledge; however, it should be noted that these assumptions may not hold for all real-world deployments, and results may vary. The key assumptions used in the modeling process include the following:

Workload Assumptions and Device Power Calculations

This evaluation focused on AI model training, as it is typically a more resource intensive process than inferencing. To accurately represent an AI training workload, it was assumed that the workload consisted of a 90% read, 10% write split. The high percentage of reads is representative of the requirement to consistently read training data to the GPU servers, while the 10% of writes account for model checkpointing requirements.

Power consumption metrics of all storage devices evaluated were taken directly from vendor specifications. To balance the power consumption of reads and writes for the specific workload, a weighted average was calculated. In the case of the competitive TLC SSDs and HDDs, the vendors do not supply separate metrics for read and write power consumption, therefore a single active power metric was used. Below are the power consumption values used for each device:

122.88 TB Solidigm High Capacity QLC SSD

Average Read Power	13.44 W
Average Write Power	22.08 W
90/10 Weighted Average	14.3 W
Idle Power	5 W

Competitive TLC SSD

Active Power	20 W
Idle Power	5 W

Solidigm TLC SSD (cache only)

Average Read Power	18 W
Average Write Power	18 W
90/10 Weighted Average	18 W
Idle Power	5 W

Competitive HDD

Active Power	8.2 W
Idle Power	6.5 W

To further consider the balance of active and idle time, a duty cycle calculation was used, based on the performance and density of each drive. Due to the time it takes for HDDs to transition from idle to active, and the high throughput requirements of the workload, it was assumed that the HDDs, and therefore their associated SSD caching layer, were active 100% of the time. The duty cycles of the competitive QLC and TLC drives were then derived by calculating the relative percentage of time required to achieve the same output as the HDDs using the performance and capacity metrics for each drive. Similar to the calculation of power metrics, performance was calculated as a weighted average of the read and write throughput values of each drive to maintain consistency with the read-heavy AI workload.

The total power requirement associated with each drive type was calculated by taking a weighted average of the drive's active and idle power consumption, utilizing the calculated duty cycles. This value was then multiplied by the total number of devices required, both in the storage servers and the JBODs or JBOFs, to support a given effective capacity. In addition, each storage management server was assumed to consume 560 W, which were attributed to the compute power required for running the software defined storage.

Additional Power Considerations:

- The power requirements of each NVIDIA DGX H100 server was assumed to be 10,200 W, as found in NVIDIA DGX H100 documentation. In addition, each server was assumed to require additional networking and fabric management resources, such as those used in NVIDIA DGX H100 SuperPOD configurations, accounting for an additional 912 W of power. In total, each NVIDIA DGX H100 server was assumed to require 11,112 W, with each rack of 4 servers requiring 44,448 W.
- For each rack of storage infrastructure, a 400GbE switch with a 1,500 W power consumption was assumed to be required. This assumption was based on specifications from a leading switch manufacturer.
- Titanium rated PSUs were assumed to be utilized for all storage racks, providing 96% power efficiency. The remaining 4% of loss was then added to the power consumption of the solution.
- To consider the impact of non-IT infrastructure power consumption within a data center, a Power Usage Effectiveness (PUE) ratio of 1.3 was chosen. The amount of non-IT infrastructure power utilized in each environment was calculated as the amount of power required to maintain a PUE of 1.3, given the total power consumed by all IT equipment included in the model.

Capacity Assumptions

A key challenge in accurately modeling an AI data center is determining the required storage capacity. While it is generally accepted that AI requires large amounts of data, there is still significant variance between the storage requirements of different AI applications. For this study, the required capacity directly effects the resulting power metrics of interest, and therefore it may be misleading to build the model entirely around a single capacity point.

To build a model that is broadly applicable to various AI environments, three distinct capacity points were chosen, representing different possible use cases. These capacity points were determined based on extensive research into existing AI data centers and evaluating several reference architectures for NVIDIA DGX H100 deployments. The following details the three capacity points selected and outlines environments in which they may be applicable.

Low Capacity – 1 PB of storage per rack of GPU servers

In general, AI often requires PB-scale data storage, with around 1PB per rack commonly cited in various references, often as a starting point that can be scaled up further. Environments may fit in this low capacity range for a variety of reasons. Natural language models, for example, typically require a relatively small capacity of training data, as it is primarily text based. Model checkpointing, however, still requires significant storage, which will vary based on the size of the model and total amount of training required. Training of smaller models with fewer parameters can lower capacity requirements, as can leveraging pre-trained foundational models with processes like transfer learning or fine tuning to reduce the overall training requirements.

Medium Capacity – 5 PB of storage per rack of GPU servers

This capacity point accommodates for larger storage capacities than what has been defined as "Low Capacity", however it is still lower than in some very large AI deployments. Capacity requirements for AI rise due to both larger training data sets and greater checkpoint requirements. Larger training data sets may stem from greater data collection to achieve more accurate models or from utilization of larger data types, such as for multi-modal models. Checkpoint requirements grow both from larger models, which increase the size of each checkpoint, as well as longer training periods, which increase the total number of checkpoints created.

High Capacity – 10 PB of storage per rack of GPU servers

This capacity represents AI deployments that are considered to have very high storage requirements. It should be noted that 10 PB of storage per rack is not a limit, and many AI environments with even larger capacities may exist. Large capacities may be required for models trained on large quantities of image or video data, such as for autonomous vehicles or medical imaging use cases. Very large models, such as foundational models, with large numbers of parameters and long training times may also contribute to higher capacity requirements.

The chosen capacity points represent a range of plausible storage deployments, based on several references, to provide a more nuanced understanding of data storage's impact on AI power consumption. It should be noted, that many AI deployments may fall outside or in between these specific capacity points. In general, however, future capacity requirements for AI are likely to increase, as models continue to get bigger and more training data is compiled.

Results

When evaluating the power efficiency of storage devices, it is important to analyze not only the relative power efficiency between various devices, but also their impact on the data center as a whole. To create a comprehensive understanding of the impact of each storage device modeled, the power efficiency findings were evaluated across several calculations.

First, the storage only power consumption was evaluated, isolating the power consumption required for only the network attached storage required to support a single rack of GPU servers at the chosen capacity point. To understand the importance of storage power efficiency alongside compute power, the total power consumption of compute and storage was also calculated. This was measured as the total power consumption of a single rack of GPU servers, and all associated storage to meet the capacity requirements. Finally, to understand the broader impact of storage throughout an entire AI data center, the total infrastructure supported within a 100 MW data center was measured, as well as breakdown of the total percentage of power attributable to data storage.

Total Power Consumption – Storage Only

A first step in evaluating the power efficiency impact of various storage devices within a data center is to understand the total power consumed directly by storage in each configuration. This was achieved by isolating the power that is attributable to the network attached storage required to support a single, fully populated, rack of GPU servers at each of the "High", "Medium", and "Low" capacity points established. Power calculations were modeled for storage configurations with each of the three devices under evaluation, as can be seen in Figure 1.

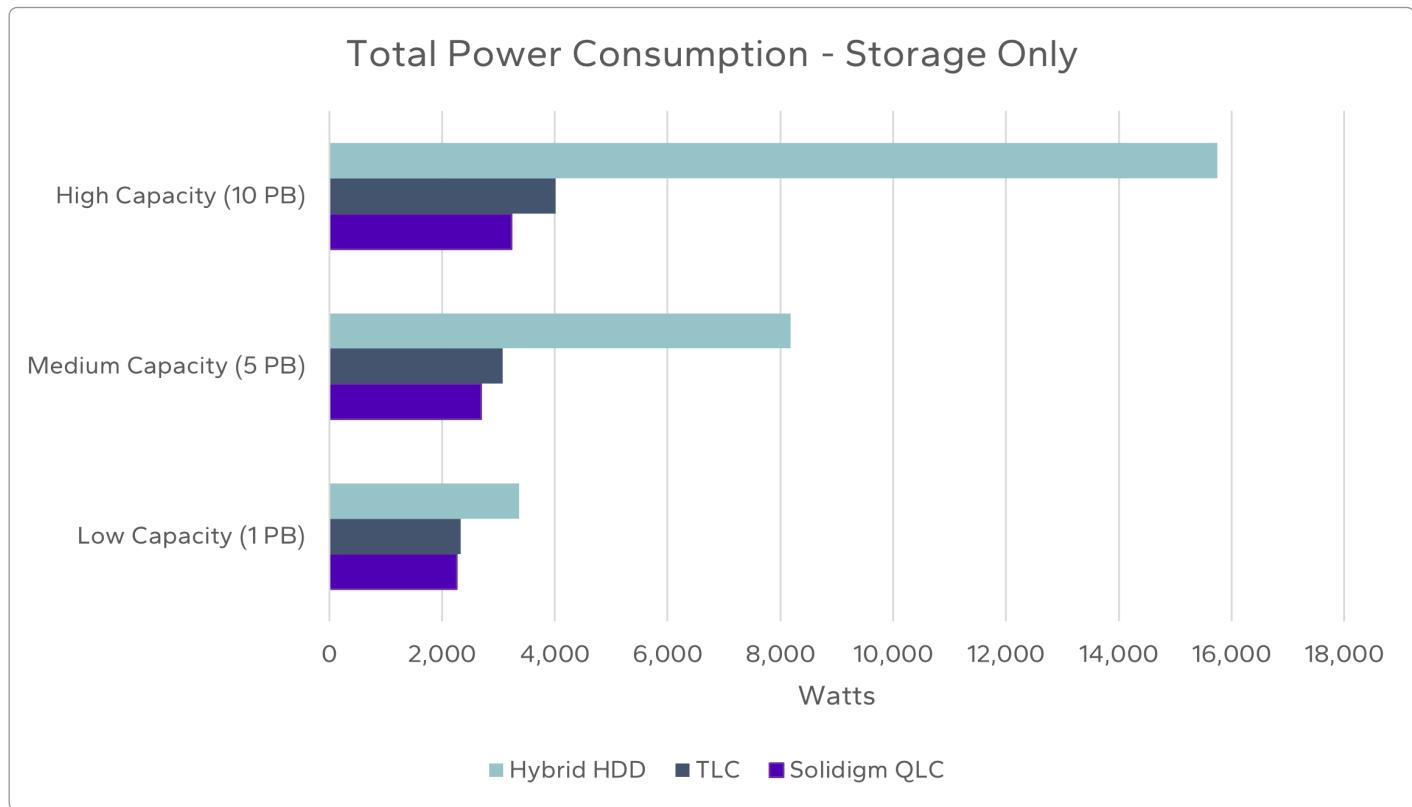


Figure 1: Total Power Consumption – Storage only

When comparing the power consumption of the three storage configurations, the model shows that the QLC SSDs are more efficient than both the TLC SSDs and the HDD configuration at each capacity point. Further, the power efficiency advantage of QLC grows as the capacity increases. Compared to TLC, QLC was found to provide a power efficiency advantage between 3.3% to 19.5% as the capacity increased from 1 PB to 10 PB. Compared to the Hybrid HDD environment, the advantage was even larger, ranging from 32.9% at the low capacity level, to 79.5% at the high capacity level.

When directly evaluating storage, it becomes apparent that QLC devices can provide significantly greater power efficiency than either TLC SSDs or HDDs. Although the three devices modeled have varying active power consumption values, with HDDs measuring the lowest for an individual drive, the high density QLC devices achieve the same total capacities with far fewer drives, resulting in lower overall power consumption. The impact of device density is particularly apparent compared to HDDs, which offer lower density than either SSD type.

The difference in device density, and its impact on power consumption, between QLC SSDs and TLC SSDs becomes increasingly apparent as the capacity requirements for the environment grow. While QLC was found to have a moderate power efficiency advantage of 3.3% over TLC at the low capacity point, that advantage grows to 12.7% and 19.5% at the medium and high levels. This not only demonstrates a power efficiency advantage for QLC, but demonstrates why high density drives are so impactful for big data challenges.

Total Power Consumption – Compute and Storage

While isolating the power consumption of storage within the data center is useful for understanding the various power efficiencies of different device types, it should also be considered alongside the broader context of the data center. For AI data centers, power consumption is typically heavily weighted by compute due to the usage of GPU servers. To understand the impact of storage devices on the overall power efficiency of AI infrastructure, the power consumption of the GPU servers and its associated storage was calculated. Each calculation accounted for a single rack of GPU servers, while modifying the device type and total capacity required, as can be seen in Figure 2.

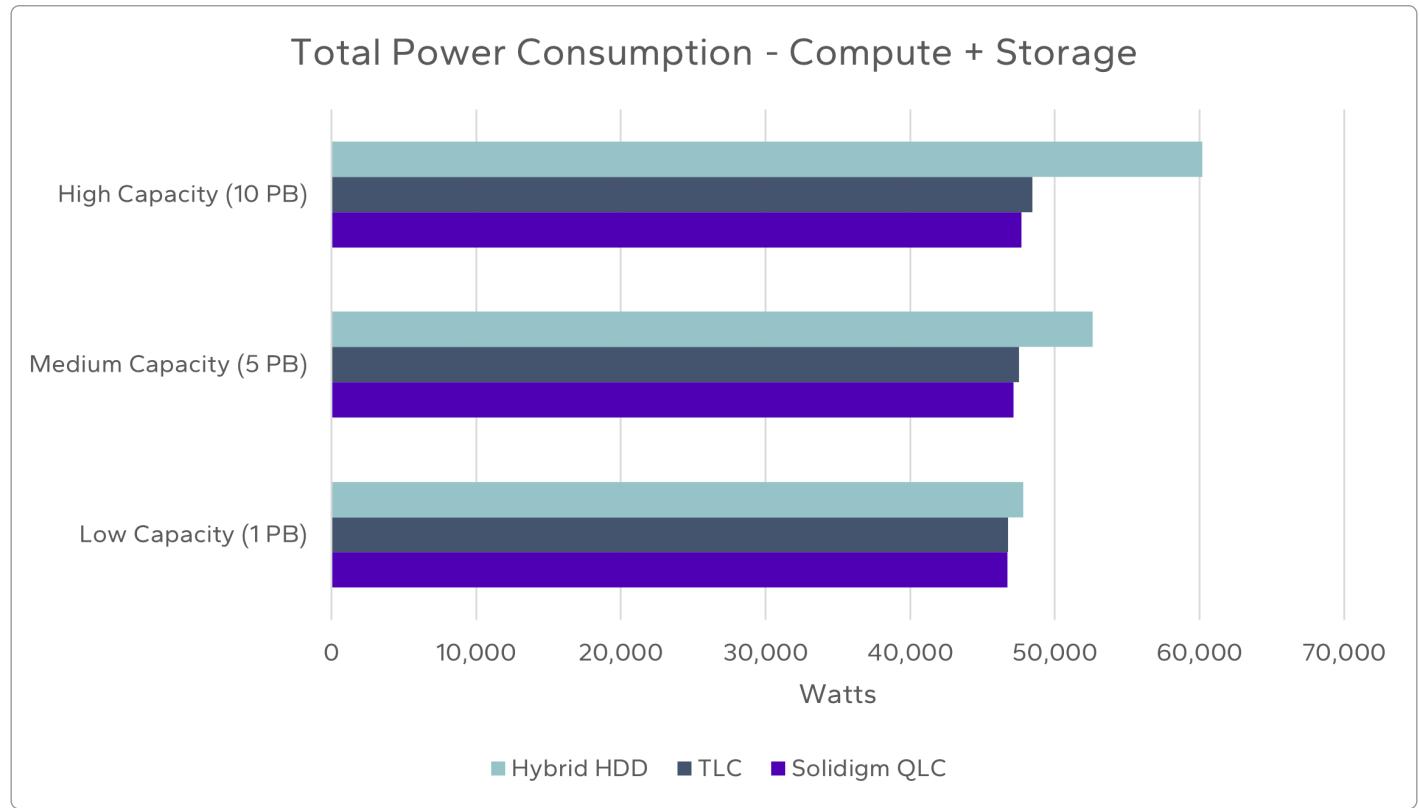


Figure 2: Total Power Consumption – Compute + Storage

As with the storage only comparison, this calculation shows a power efficiency advantage for QLC SSDs. Since the power consumption of the GPU servers is consistent across all three environments, and QLC was shown to provide more power efficient storage, this is to be expected. What this evaluation does show, however, is that using more power efficient storage impacts the power efficiency of the total AI infrastructure, even though the majority of power is consumed by compute resources.

At the low capacity level, the power efficiency advantage of the QLC environment is modest, with a 0.16% advantage over the TLC environment and a 2.32% advantage over the HDD environment. As capacity is grown, however, the impact of storage on the overall power efficiency increases. At the medium capacity point, the QLC environment was calculated to be 1% more efficient than the TLC environment and 10% more efficient than the HDD environment. At the high capacity point, the power efficiency advantage for QLC grows to 1.6% compared to TLC and 20.8% compared to HDDs.

This trend shows that the power efficiency achieved by high capacity QLC storage devices becomes increasingly impactful as AI storage requirements continue to grow. The impact of these savings can be further understood by evaluating how this power advantage can be leveraged to deploy additional AI infrastructure.

Total AI Infrastructure per 100 MW Data Center

The key goal of this study was to understand the impact on total AI infrastructure that could be supported within a data center with a set amount of power. Specifically, the study evaluated the amount of GPU server racks that could be supported within a 100 MW data center. Each rack added to the model was accompanied by the required storage infrastructure, and non-IT infrastructure power usage was calculated by assuming a constant PUE value of 1.3. The resulting number of GPU racks supported at each capacity and drive combination can be visualized in Figure 3.

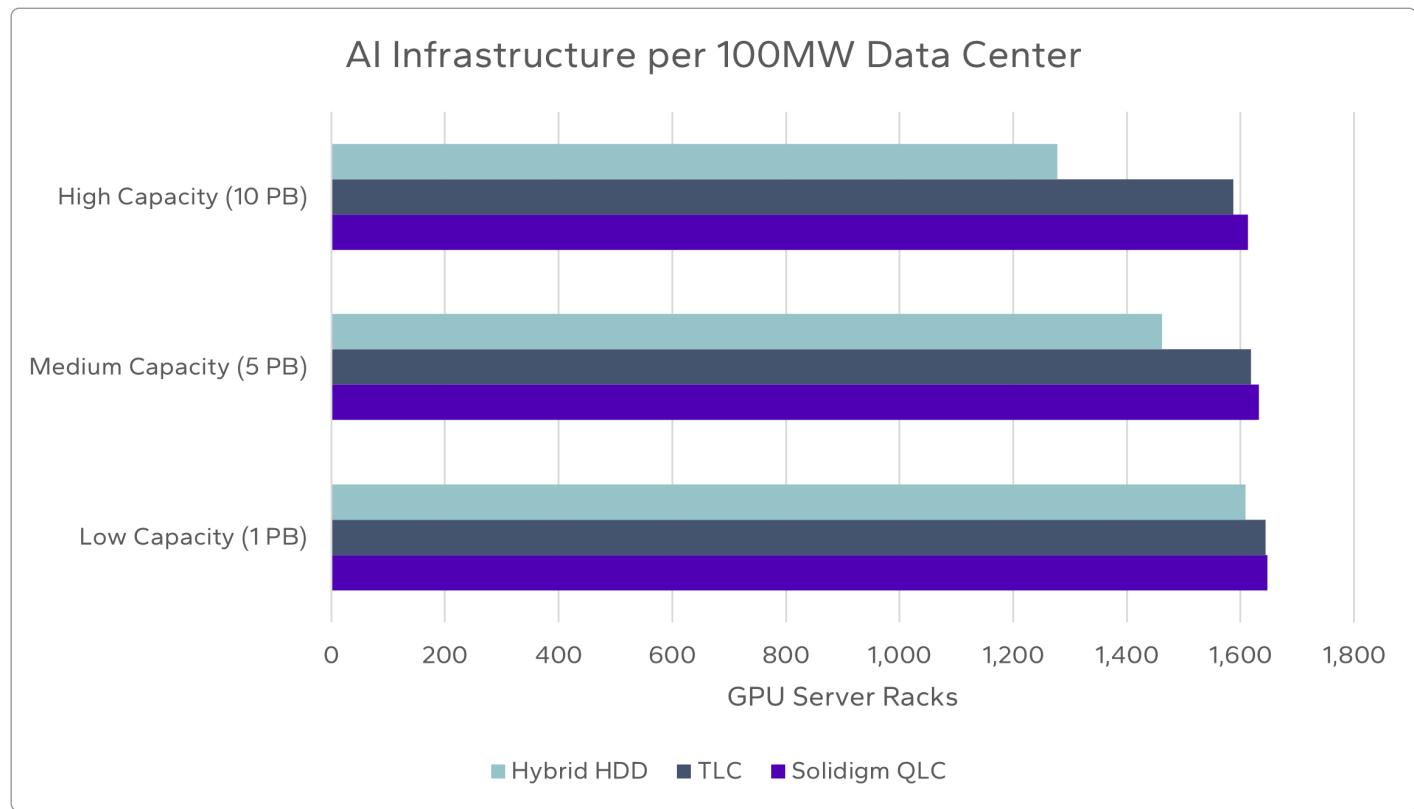


Figure 3: AI Infrastructure in 100 MW Data Center

The results of this calculation once again show a similar trend, with QLC providing an advantage which increases with higher capacities. In this case however, the power efficiency advantage gained by utilizing dense QLC SSDs is quantified by the total amount of infrastructure that can be deployed within a data center. At the high capacity level, the power efficiency gains of QLC SSDs enable an additional 26 racks of GPU servers compared to TLC SSDs, and an additional 335 racks of GPU servers when compared to an HDD-based deployment.

For AI data centers, support for such extra GPU compute can be incredibly impactful. GPUs are a key enabler for AI innovation, however, the large energy consumption of AI workloads can limit the total deployable infrastructure. The results of this evaluation demonstrate that the choice of storage media can impact the total amount of AI infrastructure supported.

Total Percentage of Power

Finally, the importance of data storage power in a data center can be understood by evaluating the total percentage of data center power that was consumed by data storage.

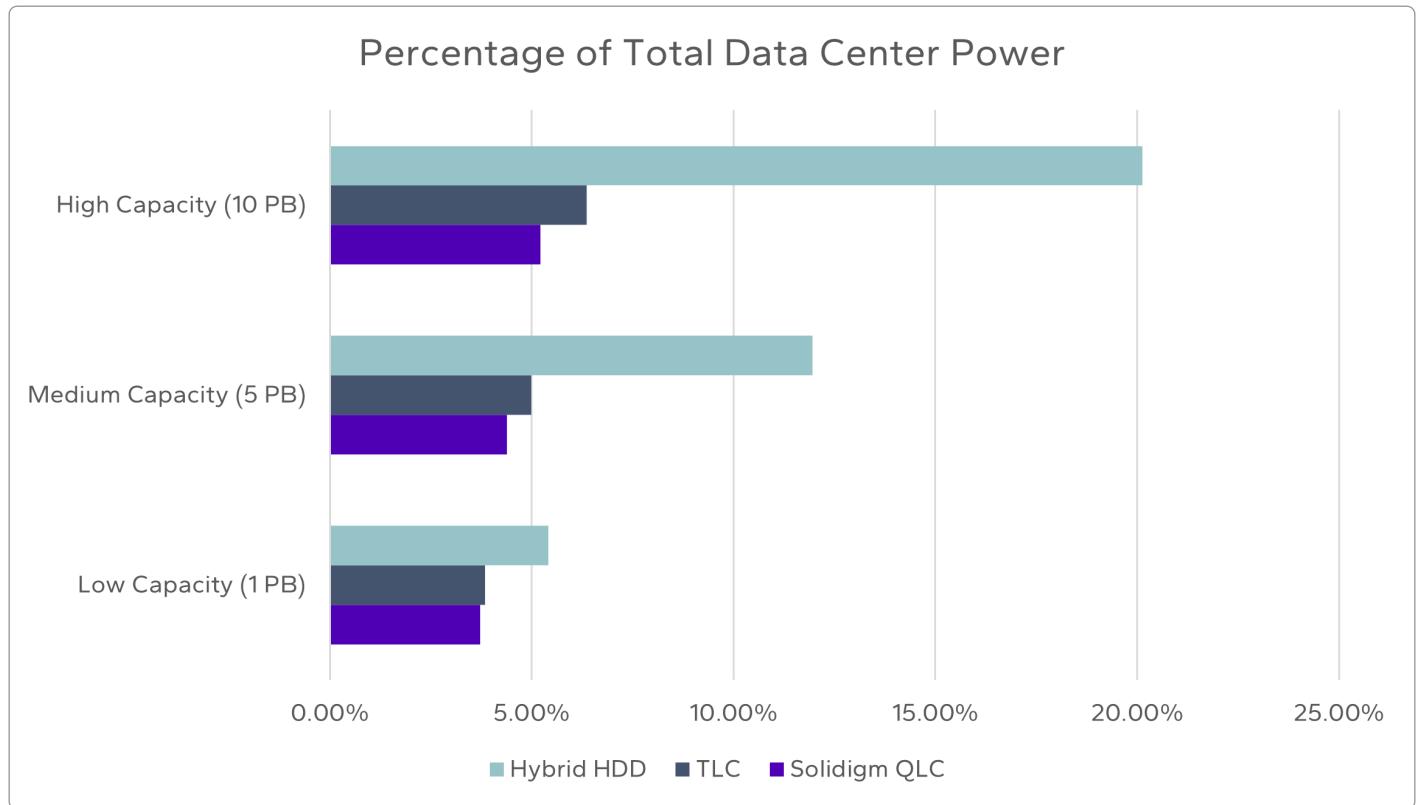


Figure 4: Percentage of Total Data Center Power

When utilizing QLC SSDs, data storage was found to account for between 3.72% and 5.21% of all data center power. Comparatively, storage configured with TLC SSDs was found to require between 3.84% and 6.37% of data center power, and the HDD configuration accounted for between 5% and 20.1% of data center power. While this shows that data storage is not the leading source of power consumption, it additionally shows that the total consumption of data storage is not negligible. When considering an AI data center with a large total power supply, such as 100 MW, even a small percentage of the total power is significant, in the range of Megawatts. While many factors within the data center cannot be altered to significantly reduce power consumption, this study demonstrates that the selection of storage devices can make a tangible impact on the total power consumption of data storage.

QLC SSDs: Key to Optimizing AI Data Centers

The results of this study highlight the significance of data storage within AI data centers. Data storage accounts for a significant portion of AI data center power consumption, and becomes increasingly impactful as storage capacities grow. Although there are several other components which contribute significantly to data center power consumption, such as compute resources and cooling requirements, data storage presents a practical avenue to optimize power efficiency. QLC technology has dramatically changed the landscape of storage devices, and has emerged as an ideal foundation for AI data storage.

QLC SSDs offer higher density than TLC SSDs, higher performance and density than HDDs, and as shown in this study, greater power efficiency than either competitive technology. These characteristics are well suited to AI data centers which not only demand both high performance and high capacity, but are increasingly limited by their overall power consumption.

This study demonstrates the value of high density QLC drives in capacity demanding environments, such as AI. To achieve such capacity levels with either HDDs or TLC SSDs, far more devices are required, reducing data center floorspace available to compute and increasing storage-related power consumption – both of which could instead be leveraged to deploy additional AI compute resources. As shown with the various capacity points used in this study, this dynamic becomes increasingly impactful as capacity requirements grow.

With the renewed interest in AI technology, the overall data requirements are likely to increase, further highlighting the need for power efficient storage at high capacity points. Future AI capacity demands are likely to be driven by larger training data sets, as well as increasingly large models. With increasingly large capacity demands, high density QLC devices become an ideal choice for efficient data storage.

Through modeling the power requirements of a 100 MW AI data center, this study showed that Solidigm High Density QLC SSDs can achieve significant power efficiency advantages over both competitive TLC SSDs and competitive HDDs. When isolating the power consumption of data storage, Solidigm's QLC SSDs were shown to be up to 19.5% more efficient than TLC SSDs and up to 79.5% more efficient than HDDs. It was additionally found that the power efficiency gained by utilizing Solidigm QLC SSDs can enable data centers to deploy more total infrastructure. When evaluating infrastructure within a 100 MW data center, utilizing QLC devices unlocked up to 1.6% more AI infrastructure compared to TLC SSDs and up to 26.3% compared to HDDs. These results demonstrate that high density QLC storage can help organizations overcome data center power limitations, enabling larger AI compute clusters, faster results, and greater overall innovation within the field of AI.

In addition to the results found in this study, the energy efficiency advantages of leveraging Solidigm High Density QLC SSDs have additionally been acknowledged by other leading technology vendors when discussing the growing energy challenges of AI:

“ As AI models become more sophisticated, growing energy demands must be addressed if we are going to fully harness the potential of AI. A holistic approach encompassing compute, storage and networking is key to optimizing infrastructure for AI workloads, and the pervasive Arm compute platform is enabling this from cloud to edge. Solidigm's new 122TB storage solution, powered by Arm's high-performance, power-efficient technology, will help tackle these power challenges, delivering more efficient and scalable data center designs.

– Chloe Ma, VP China GTM, IoT Line of Business, Arm

As AI continues to evolve, data storage will continue to play a key role. Solidigm high density QLC devices are positioned as an ideal choice for AI data centers, offering dense, all flash storage, capable of meeting the performance and capacity demands of AI, as well as optimizing power efficiency. The results of this study highlight the important role that data storage plays in AI power consumption, and demonstrates how QLC devices can help organizations increase their power efficiency and achieve their AI goals.

Appendix

The following charts display the full results of the study. Figure 5 shows the power consumption of storage infrastructure required for a single rack of GPU servers. Figure 6 shows the combined power consumption of a single rack of GPU servers and its associated storage. Figure 7 shows the total number of GPU server racks and associated storage that can be supported by a 100 MW data center. Figure 8 shows the total percentage of power attributed to data storage in a 100 MW data center for each configuration measured.

Total Power – Storage Only

	Solidigm QLC	TLC	Advantage	Hybrid HDD	Advantage
Low Capacity (1 PB)	2,258 W	2,335 W	3.30%	3,368 W	32.97%
Medium Capacity (5 PB)	2,690 W	3,080 W	12.65%	8,179 W	67.11%
High Capacity (10 PB)	3,230 W	4,011 W	19.46%	15,749 W	79.49%

Figure 5: Total Power – Storage Only

Total Power – Compute + Storage

	Solidigm QLC	TLC	Advantage	Hybrid HDD	Advantage
Low Capacity (1 PB)	46,706 W	46,783 W	0.16%	47,816 W	2.32%
Medium Capacity (5 PB)	47,138 W	47,528 W	0.82%	52,627 W	10.43%
High Capacity (10 PB)	47,678 W	48,459 W	1.61%	60,197 W	20.80%

Figure 6: Total Power – Compute + Storage

Total Infrastructure – Racks of GPU Compute + Storage

	Solidigm QLC	TLC	Advantage	Hybrid HDD	Advantage
Low Capacity (1 PB)	1,647	1,644	0.17%	1,609	2.38%
Medium Capacity (5 PB)	1,632	1,618	0.83%	1,462	11.64%
High Capacity (10 PB)	1,613	1,587	1.64%	1,278	26.26%

Figure 7: AI Infrastructure Supported in 100 MW Data Center

Percentage of Power Consumption Attributed to Data Storage

	Solidigm QLC	TLC	Hybrid HDD
Low Capacity (1 PB)	3.72%	3.84%	5.42%
Medium Capacity (5 PB)	4.39%	4.98%	11.96%
High Capacity (10 PB)	5.21%	6.37%	20.12%

Figure 8: Percentage of Power Consumption

Duty Cycle Calculation

The following formula was utilized to calculate the duty cycles of each device modeled:

$$\text{Duty Cycle of SSD} = \text{Duty Cycle of HDD} * (\text{Capacity of SSD} / \text{Capacity of HDD}) * (\text{performance of HDD} / \text{performance of SSD})$$

In which the Duty Cycle of HDD was assumed to be 100%

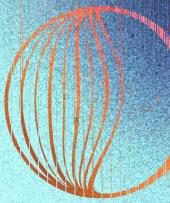
Devices

Figure 9 displays information regarding the devices used in this study:

	Solidigm QLC SSD	TLC SSD	HDD
Capacity	122.88 TB	61.44 TB	24 TB
Read Performance	7,462 MB/s	12,000 MB/s	285 MB/s
Write Performance	3,250 MB/s	5,000 MB/s	285 MB/s
Active Read Power	13.44 W	20 W*	8.2 W*
Active Write Power	22.08 W	20 W*	8.2 W*
Idle Power	5 W	5 W	6.5

Figure 9: Device Specifications

*TLC SSD and HDD power metrics were not broken down by read and write. A mixed read/write power metric, as provided in the devices specifications, was utilized for both values.



Important Information About this Report

CONTRIBUTORS

Mitch Lewis

Performance Analyst | Signal65

PUBLISHER

Ryan Shrout

President and GM | Signal65

INQUIRIES

Contact us if you would like to discuss this report and Signal65 will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "Signal65." Non-press and non-analysts must receive prior written permission by Signal65 for any citations.

LICENSING

This document, including any supporting materials, is owned by Signal65. This publication may not be reproduced, distributed, or shared in any form without the prior written permission of Signal65.

DISCLOSURES

Signal65 provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

IN PARTNERSHIP WITH



ABOUT SIGNAL65

Signal65 is an independent research, analysis, and advisory firm, focused on digital innovation and market-disrupting technologies and trends. Every day our analysts, researchers, and advisors help business leaders from around the world anticipate tectonic shifts in their industries and leverage disruptive innovation to either gain or maintain a competitive advantage in their markets.



CONTACT INFORMATION
Signal65 | signal65.com