

WHITE PAPER

Accelerating Semiconductor Design Pipelines with Pure Storage FlashBlade//S

Experience a higher level of performance for EDA and HPC workloads.

Contents

Introduction	3
HPC and EDA Workloads and Processes	3
FlashBlade Storage Endpoints	5
Multi-chassis Linear Performance of FlashBlade//S200 and FlashBlade//S500	9
Conclusion	13
Additional Resources	14



Introduction

Leading semiconductor companies are designing sub-10nm chips that require highly scalable infrastructure including compute, network, and storage architecture to handle a high number of front- and back-end jobs during the design process. Pure Storage® FlashBlade® has long been a preferred storage platform for companies with EDA workloads. With the design process generating a high volume of data throughout the life of their projects, the high performance, parallel architecture found in FlashBlade accelerates EDA workloads and shortens chip design life cycles.

The Pure Storage [FlashBlade//S™ platform](#) and Purity//FB 4.5 offers a higher level of performance to Electronic Design Automation (EDA) and High Performance Computing (HPC) workloads. As workloads in HPC server farms ramp up due to the increased complexity in sub-10nm chips, chip design teams require more IOPS and throughput performance from the underlying storage to meet customer demand faster.

This white paper uses the Spec 2020 benchmark with EDA workloads to show significantly increased IOPS and bandwidth performance from the new FlashBlade//S500.

HPC and EDA Workloads and Processes

Figure 1 shows the typical chip design pipeline. For more information on HPC/EDA workloads and processes refer to [Part 1](#) and [Part 2](#) of the blog “Real-world File Storage Performance with FlashBlade in Chip Design.”

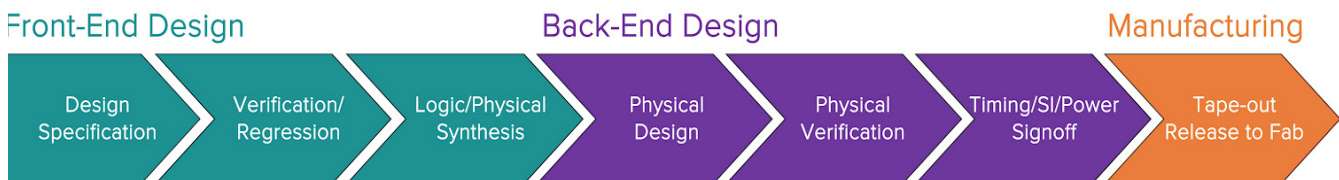


FIGURE 1 Semiconductor design workflow

It is a common best practice to grow and scale compute requirements on demand during the chip design process. Equal attention should be given to the storage infrastructure for these workflows, as often the IO contention suffers as data is accessed and processed at scale by EDA jobs, creating bottlenecks and slow response times.

In an analog or digital chip design workload, the access pattern and workload generated by many competing phases depend on file system size and directory structure. Some parts of the design phase generate massive amounts of metadata, while others require high throughput, and some require both.

Many of the top semiconductor companies use FlashBlade for chip design workflows. A deep analysis of their workloads reveals that these workloads are highly dependent on metadata performance, in addition to the read and write bandwidth that FlashBlade offers.



While partnering with these companies, the challenges they face in completing chip design jobs and workflows associated with the simulation and tapeout phases included in a modern chip design process became clear to us at Pure. The main challenges are:

- Scaling compute and storage capacity independently to handle high metadata and bandwidth operations performance in high file count environments
- Accelerating job completion time by reducing simulation and regression times during the semiconductor design process

Given our deep understanding of the nature of semiconductor jobs and workflows, we are able to accurately recreate the read, write, and metadata characteristics of typical design workflows on FlashBlade using the SpecStorage2020 benchmark and load generator tool. Three key custom workload profiles have been developed that represent the most important design workload characteristics.

This paper illustrates the simulated chip design workflow on FlashBlade using the SpecStorage2020 benchmark and load generator tool to measure and compare the first generation FlashBlade with the next generation FlashBlade//S200 and FlashBlade//S500 platform performance at scale, based on IOPS, bandwidth and latency.

[Pure Storage FlashBlade//S, a unified fast file and object \(UFFO\) storage platform](#), is designed to store, manage, and protect exponential data growth with high storage efficiency and performance at scale. The FlashBlade//S series comes in two variants: S200 (capacity optimized) and S500 (performance optimized). FlashBlade//S was purposely designed to accelerate the HPC workload demands that are characteristic of EDA tools used during chip design and tapeout. FlashBlade is deployed in some of the top semiconductor and foundry businesses in the world to help them bring innovative products to market faster and more efficiently.

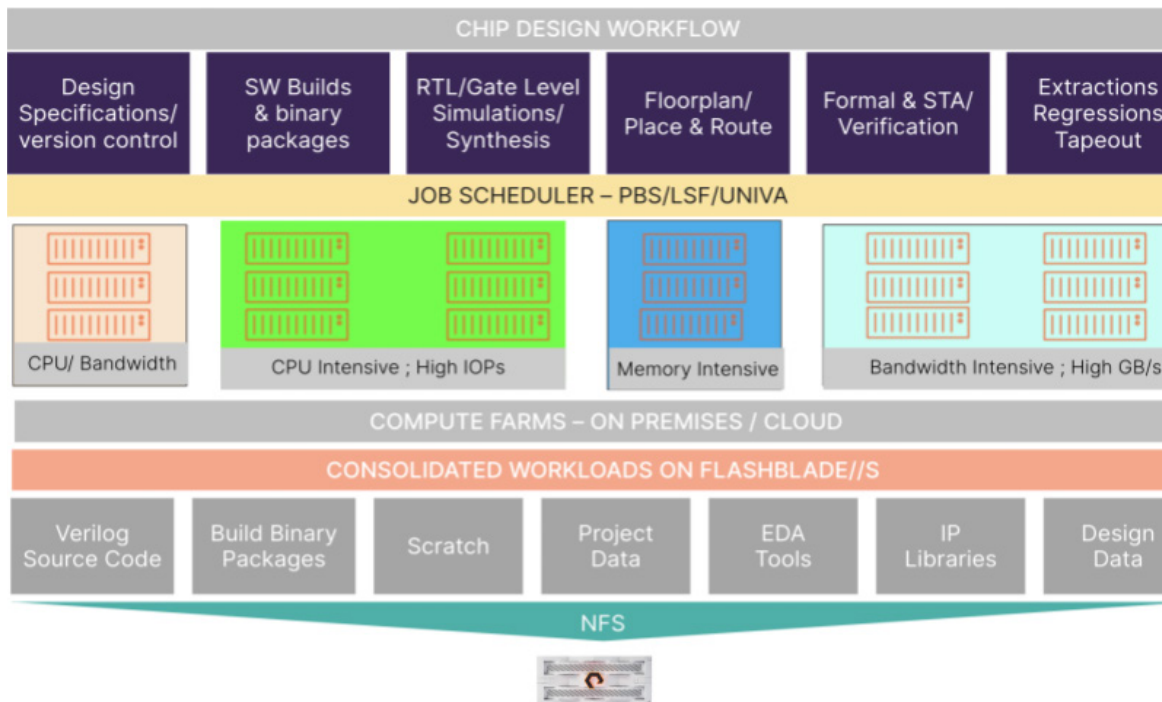


FIGURE 2 Front-end and back-end chip design workflow on FlashBlade//S



Figure 2 illustrates the front-end and back-end workloads in a typical chip design workflow for a semiconductor company with multiple cores in a compute farm with a scheduler to load balance the jobs in the queue. Server farms with tens and thousands of cores are used to run the chip design jobs.

The job schedulers look for free CPU resources for the jobs in the queue for faster turnaround time. Commonly used job schedulers in EDA workflows, like Univa Sun Grid Engine (SGE) and IBM Platform Load Sharing Facility (LSF), are not storage-aware. The IO contention to the back-end storage leads to job failures. The resubmission of the jobs to the queue is complex and time-consuming and thus slows down the design process. FlashBlade//S is purpose-built to handle the highly random nature of the job submission in the compute farm during the life cycle of a design project.

The SpecStorage2020 has by default a front-end and back-end workload profile for EDA workloads as illustrated in Figure 2 above. Based on feedback from some major semiconductor and EDA tool companies, three custom profiles were created in SpecStorage2020 for the load generation representing the following workload scenarios, in an open loop testing keeping the round trip latency under 5ms:

- The metadata-only front-end workload profile was configured with 100% non-modify metadata operations like GETATTRs, SETATTRs, and ACCESS calls with file sizes between <4k–8k. This kind of workload represents a significant part of the logical verification phase.
- The mixed metadata workload profile representing the logical simulation phase was a 75% mix of non-modifying and modifying metadata like CREATE, RENAME, DELETE, and others along with 25% READs and WRITEs. The file sizes for this workload profile were configured between 8k–16k.
- The mixed workload represents physical verification and tapeout phase along with a front-end design process. While the front-end profiles had a similar profile as described in the second profile above, the back-end profile was configured to represent high bandwidth during the tapeout phase with file sizes from 128k–512k.

FlashBlade Storage Endpoints

In an open loop testing, the SpecStorage2020 pushes the load with maximum IOPS and bandwidth defined in the custom profiles with no latency constraint with the following fully populated FlashBlade endpoints. The workloads were run against four FlashBlade hardware configurations, consisting of two FlashBlade (Gen 1) systems and two FlashBlade//S (Gen 2) systems. Notice that the FlashBlade//S products offer denser storage capacity, increased CPU count, and increased memory.

- FlashBlade 15×17TB First Generation (Gen 1)
- FlashBlade 30×17TB First Generation (Gen 1)
- FlashBlade//S200 (10 blades with 4 × 24TB Direct Flash Modules; Gen 2)
- FlashBlade//S500 (10 blades with 4 × 24TB Direct Flash Modules; Gen 2)



The initial workload generation and testing was scaled from 256 to 544 cores on Gen 1 FlashBlade platforms. An additional 352 cores were added to the test to scale the workload further on Gen 2 FlashBlade//S.

The objective of the scale testing on two generations of FlashBlades was to identify and validate the following on FlashBlade//S (Gen 2) compared to FlashBlade (Gen 1) on the sample test workload:

- Reduce the overall job completion time while increasing the number of jobs on FlashBlade//S
- Identify if FlashBlade//S can scale concurrent jobs when more servers and cores are added to the compute farm
- Measure IOPS and bandwidth improvement on FlashBlade//S with low latency response time from FlashBlade platforms while scaling the concurrent jobs with more cores in the compute farm

The following graphs show the increased performance of the FlashBlade//S with Purity//FB 4.1. The tests confirm that the read and write performance of the new generation of FlashBlade//S is superior to the already impressive Gen 1 FlashBlade. Even more revealing for EDA workloads is the substantial jump in metadata performance that is provided by the FlashBlade//S. This will be of great interest to anyone who is responsible for running HPC and EDA jobs.

The following graphs represent the test results from the validations of the three EDA workload profiles with respect to IOPS and bandwidth. The IOPS in Figure 3 and bandwidth in Figure 5 indicate that all the three EDA workload profiles scale linearly for tests on a 15-blade Gen 1 FlashBlade and 30-blade Gen 1 FlashBlade compared to 10-blade Gen 2 FlashBlade//S200, and 10-lade Gen 2 FlashBlade//S500, respectively.

The Y-axis in the following graphs indicates the scale factor for each of the workloads at scale and does not represent the raw numbers.

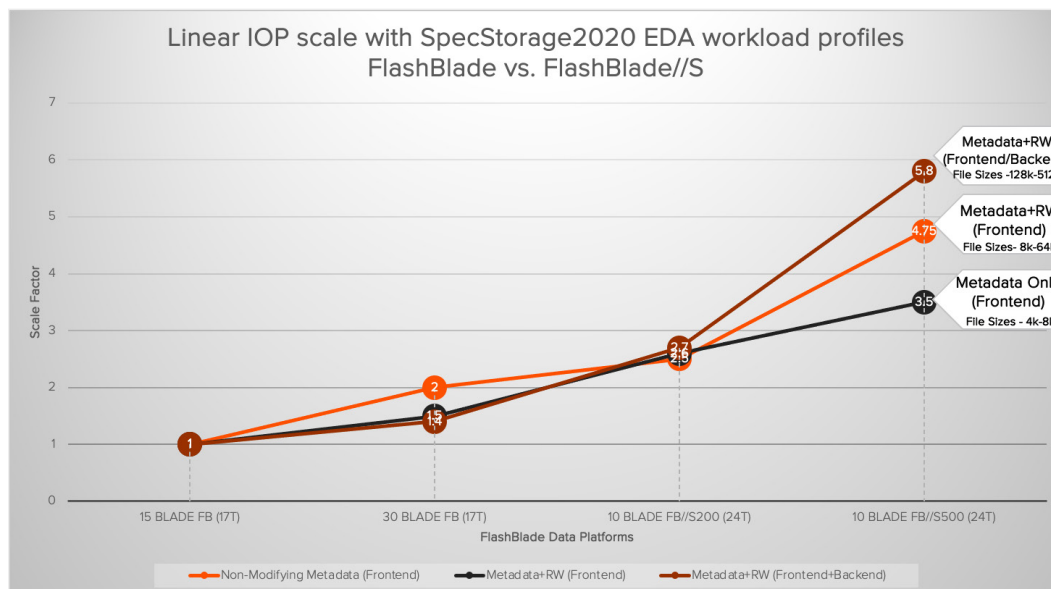


FIGURE 3 Scaling IOPS for three custom EDA profiles with storage latency under 5ms

The IOP scalability measured in above Figure 3 indicates that FlashBlade//S500 is capable of speeding up the logical verification phase in the chip design process more than two times. The improved performance for the front-end workload on the FlashBlade//S is based on the following observations:

- Reduced job completion time up to 50% with FlashBlade//S, compared to Gen 1 FlashBlade. The job completion time with FlashBlade//S500 dropped to 15 minutes, compared to 30 minutes from the Gen 1 FlashBlade.
- FlashBlade//S500 can support more than 60% core density in the compute farm to scale, which means more jobs can be in the queue. The FlashBade//S500 completed the jobs in 15 minutes. FlashBlade//S500 was able to support 896 cores without any job failure compared to Gen 1 FlashBlade that maxed out with 544 cores under acceptable latency.
- FlashBlade//S500 was still able to serve more jobs with the additional 60% cores in the compute farm, but it hit a point of diminishing returns and the completion time nearly doubled. However, there were no job failures reported during the tests with higher job completion on a fully saturated FlashBlade//S. Job failures lead to excessive completion times due to the resubmission of the jobs in the queue, further adding complexity to job management and the overall design process.

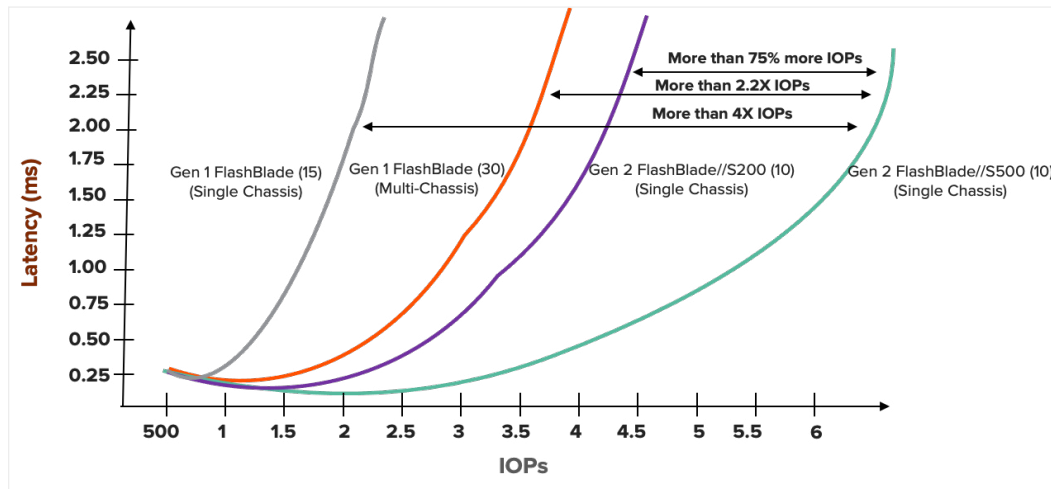


FIGURE 4 IOP vs. latency curve (relative data)

As shown in the graph in Figure 4, both Gen 1 and Gen 2 FlashBlades can push more IOPS with higher storage latency. High storage latency leads to longer job completion time. The overall storage latency is reduced by 50% for metadata based workloads on Gen 2 FlashBlade//S data platforms compared to Gen 1 FlashBlade.



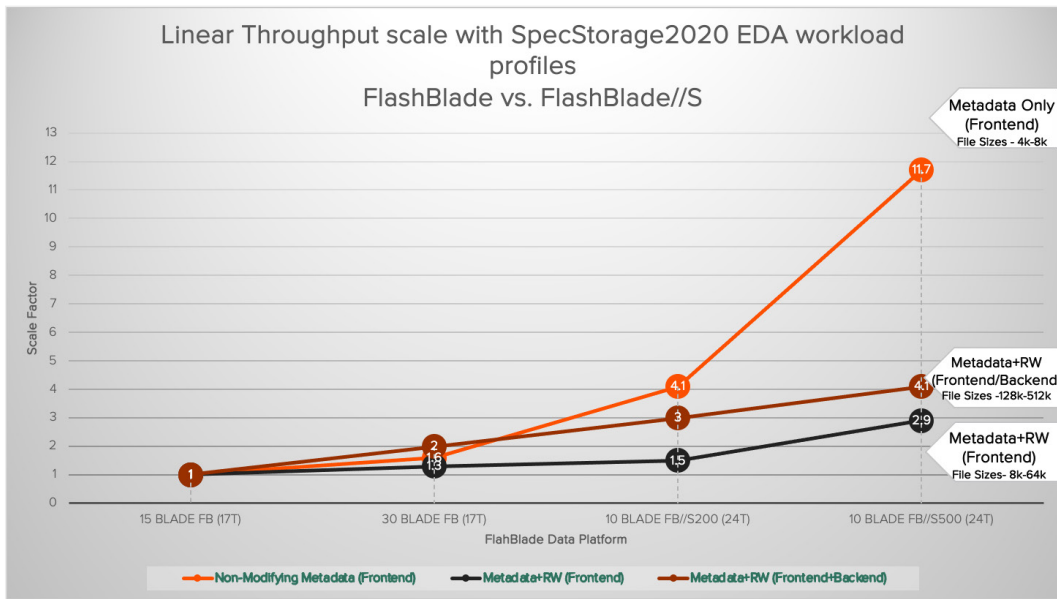


FIGURE 5 Scaling bandwidth for three custom EDA profiles with storage latency under 5ms

The linear scalability in bandwidth, shown in Figure 5, indicates that the regression time during tapeout can be reduced by 50% on FlashBlade//S. The improved bandwidth from the backend workload during tapeout is based on the following observations:

- With 50% more cores in the compute farm, the FlashBlade//S was able to cut the job completion time by 50% compared to Gen 1 FlashBlade for large GDSII files during tapeout. The FlashBlade//S is able to support more cores in the compute farm for reduced completion times.
- There is more than 75% linear improvement in the aggregated bandwidth for the backend workload on the sample test data with FlashBlade//S500 compared to Gen 1 FlashBlade. The graph in Figure 6 indicates the FlashBlade//S provides better throughput with nconnect enabled on the compute hosts under low latencies.

NOTE: The “nconnect” NFS mount option to support read and write operations to FlashBlade is available in commercial versions of RHEL8.4/Ubuntu20.04/SLES15 and later. The “nconnect” on the linux hosts and FlashBlade provides a significant improvement to bandwidth and latency. It is recommended not to mix different nconnect options (nconnect=8 & nconnect=16) for the same data-VIF while mounting the filesystem. The nconnect picks up the first setting and ignores the second one. You should also use different data-VIF on the FlashBlade while mounting a different filesystem using nconnect.



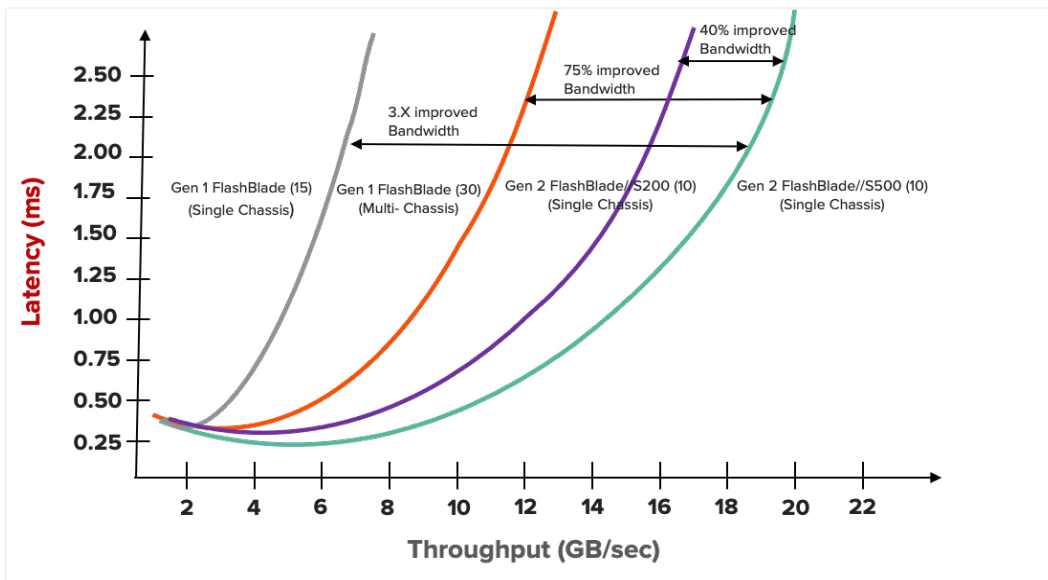


FIGURE 6 Throughput vs. latency curve (relative data)

- The tests indicated that the data reduction was 15% more for the sample test data on FlashBlade//S compared to Gen 1 FlashBlade. The high space savings for small and large files during the design process indicates that more workloads can run on FlashBlade//S500 with a smaller data footprint.
- The write latency for the large files during tapeout was reduced by 60% on FlashBlade//S500 compared to Gen 1 FlashBlade. The faster writes with low latencies indicate that FlashBlade//S500 can perform more throughput at scale.

Many semiconductor design customers are using single-chassis Gen 1 FlashBlade and are satisfied with the performance in scratch space. More customers have expanded to multi-chassis Gen 1 FlashBlade and are achieving powerful results. With the addition of the Gen 2 FlashBlade//S200 and FlashBlade//S500 platforms, semiconductor business owners have two more powerful options to consider.

The FlashBlade//S200 achieves better performance than the current 30-blade Gen 1 FlashBlade. A 30% improvement in data reduction on an average design dataset over the Gen1 FlashBlade indicates that the Gen 2 platforms are capable of reducing the data footprint further. Customers who want a balance of performance and cost will choose this option. The above test results and charts support this statement.

Multi-chassis Linear Performance of FlashBlade//S200 and FlashBlade//S500

Semiconductor design and production data centers have massive amounts of CPU cores that can overrun any storage platform. The single chassis can scale performance to a certain number of IOPS and throughput under lower latencies beyond which the storage hits a saturation point where the EDA job completion time or the turnaround time (TaT) is longer, thus slowing down the chip design workflow. Pure Storage has the ability to scale from a single to a multi-chassis configuration to scale performance and capacity linearly. The second generation FlashBlade//S multi-chassis model, which can be configured for up to 10 chassis, allows EDA workloads to scale with the number of design and tapeout jobs under low latencies. Apart from reducing simulation and regression time by boosting IOPS and throughput at low latencies, FlashBlade//S provides improved cost efficiency with higher data reduction and lower operational costs with multi-chassis.



SpecStorage2020 was used to benchmark the EDA workload with three FlashBlade//S chassis to measure the maximum IOPS and throughput under sustained low latencies. The workloads are run against two multi-chassis FlashBlade//S (Gen 2) systems. The objective of the scale testing on FlashBlade//S multi-chassis was to validate and measure identify and validate the TaT for the EDA jobs with higher number of CPU cores.

- FlashBlade//S200 (30 blades with 4 × 24TB Direct Flash Modules; Gen 2)
- FlashBlade//S500 (30 blades with 4 × 24TB Direct Flash Modules; Gen 2)

During workload generation and testing, a total of 1152 client cores were used to stress the Gen 2 FlashBlade//S multi-chassis platforms. The 30% increase in cores during the multi-chassis testing was a limitation in the test scenario.

The graphical representation in Figure 7 depicts the consistent increase in performance for the FlashBlade//S, aligning with the incremental growth in chassis numbers with Purity//FB 4.3.4 These findings hold significant appeal for those overseeing high-performance computing (HPC) and electronic design automation (EDA) jobs.

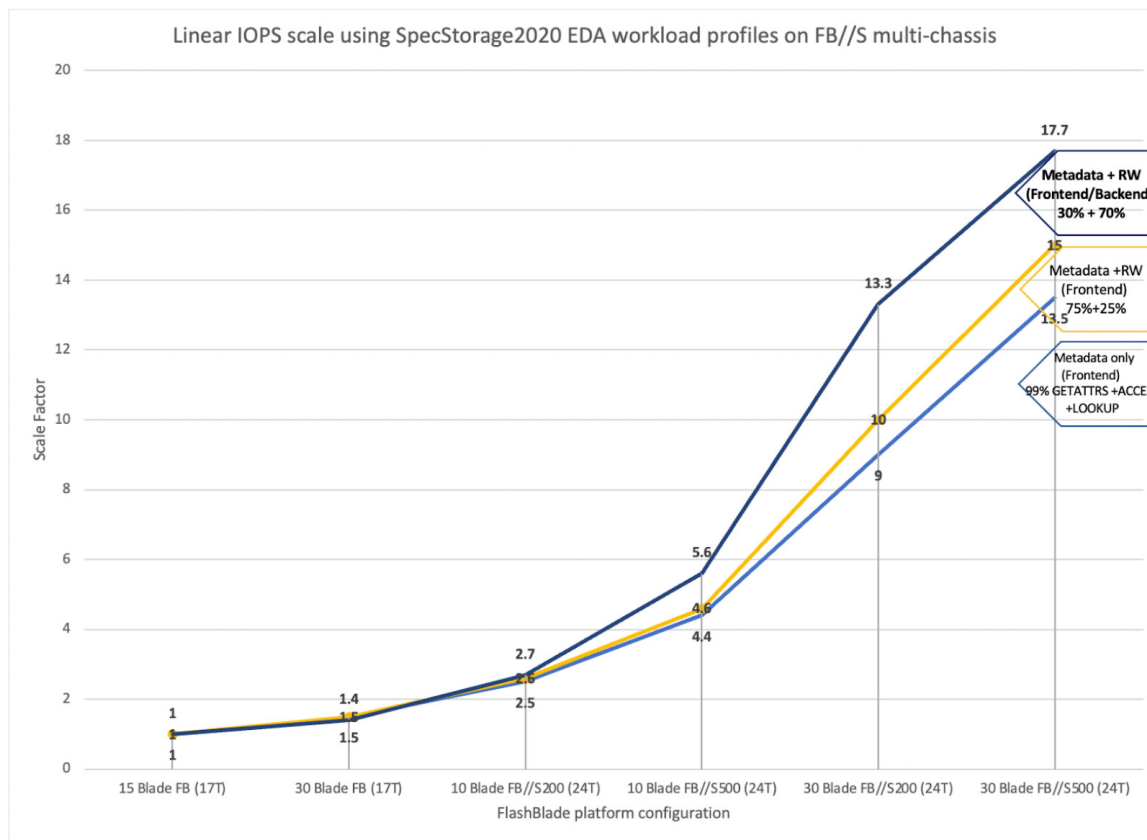


FIGURE 7 Scaling IOPS for three custom EDA profiles with storage latency under 5ms on three FlashBlade//S multi-chassis units

The above graph represents the test results from the three EDA workload profiles identical to the tests on the single chassis. The IOPS in Figure 7 and bandwidth in Figure 9 indicate that all the three EDA workload profiles have significant linear scalability with FlashBlade//S multi-chassis compared to Gen 1 FlashBlade. Based on the scalability observed from the multi-chassis tests, it is highly recommended to upgrade to FlashBlade//S from the end-of-life (EOL) Gen 1 FlashBlade.



IOPS Performance for EDA workloads better by 6.5X

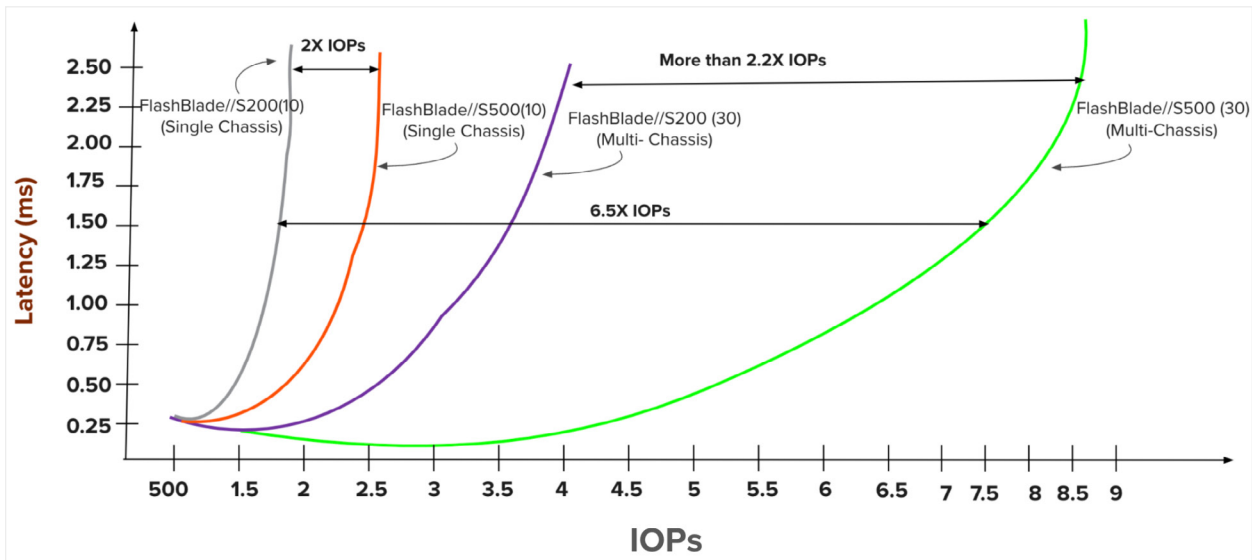


FIGURE 8 Latency vs. IOPS comparison between FlashBlade//S single and multi-chassis

Figure 8 shows the latency and IOPS curve for various single- and multi-chassis FlashBlade//S configurations. Scaling the IOPS is not limited to a single chassis; rather multi-chassis also demonstrates much better linearity as the IOPS scales with a higher number of EDA jobs.

- There is a 6.5X improvement in the number of IOPS generated from a single chassis S200 compared to a multi-chassis S500.
- A 30% additional cores were used in the test scenario to scale the load on the multi-chassis keeping the job completion time around 15 mins. This does not mean that the 3 X FlashBlade//S multi-chassis is only limited to 1152 cores. The FlashBlade//S500 multi-chassis had a lot of CPU headroom available to service more jobs at scale. The scalability was limited by the number of cores available in the test environment. Enhancing IOPS for metadata-only workloads by more than 2.5 times and subsequently reducing overall job completion time.
- The FlashBlade//S500 multi-chassis was 2.2X faster compared to FlashBlade//S200 multi-chassis.
- This achieved a significant reduction in overall latency by 50%.
- Apart from scaling with higher compute cores by 30% to handle high metadata and bandwidth operations performance in high file count environments, the FlashBlade//S multi-chassis improved ROI by reducing storage and operational costs.
- Increasing data reduction by more than 30% on an average for various EDA datasets provides improved cost efficiencies with FlashBlade//S multi-chassis platforms.



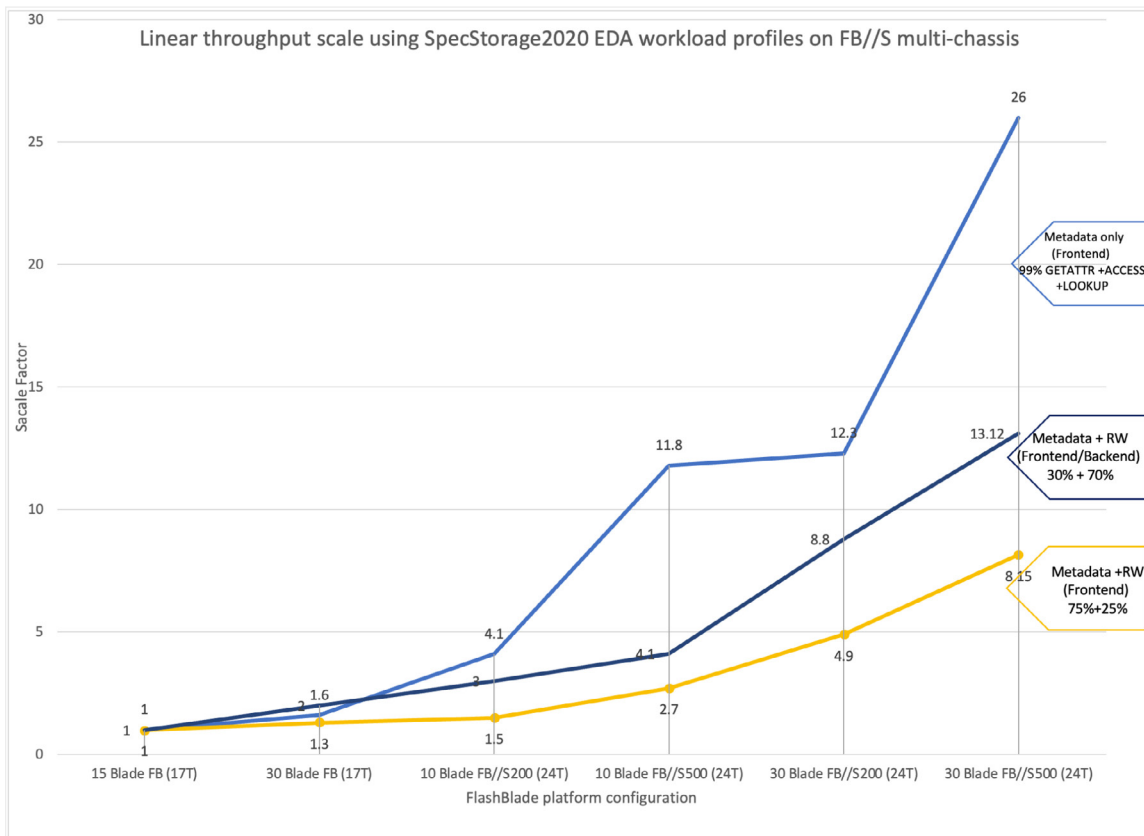


FIGURE 9 Scaling throughput for three custom EDA profiles with storage latency under 5ms on three FlashBlade//S multi-chassis

The demonstrated increase in bandwidth, as depicted in Figure 9, attributed to the following observations:

- A single-chassis FlashBlade//S500 generates similar number throughput compared to 3X S200 multi-chassis. This means the rack density is further reduced when using a single S500 over 3X S200 for high metadata workloads.
- The other two EDA profiles with mixed workloads show linear throughput improvement from FlashBlade//S single chassis to multi-chassis under low write latencies consistently thus providing a faster TaT.
- As mentioned earlier in this section, each of the blades in FlashBlade//S had four Data Fabric Modules (DFM). Tests indicated that the maximum throughput can also achieved from 3 DFMs. The 4th DFM can be used to add more capacity. This allow the business owners to avoid over subscribing capacity when it is not needed and thus reducing costs without compromising performance.
- The modular architecture of the FlashBlade//S platform allows business owners to customize the storage based on the workloads for a better total cost of ownership (TCO). Adding blades to a single or multi-chassis and DFMs to a blade provides the ability to scale IOPS and throughput linearly without any disruption to the workflow The CPU cores in the compute farm and the storage performance and capacity can scale disaggregate to optimize cost and storage efficiencies.



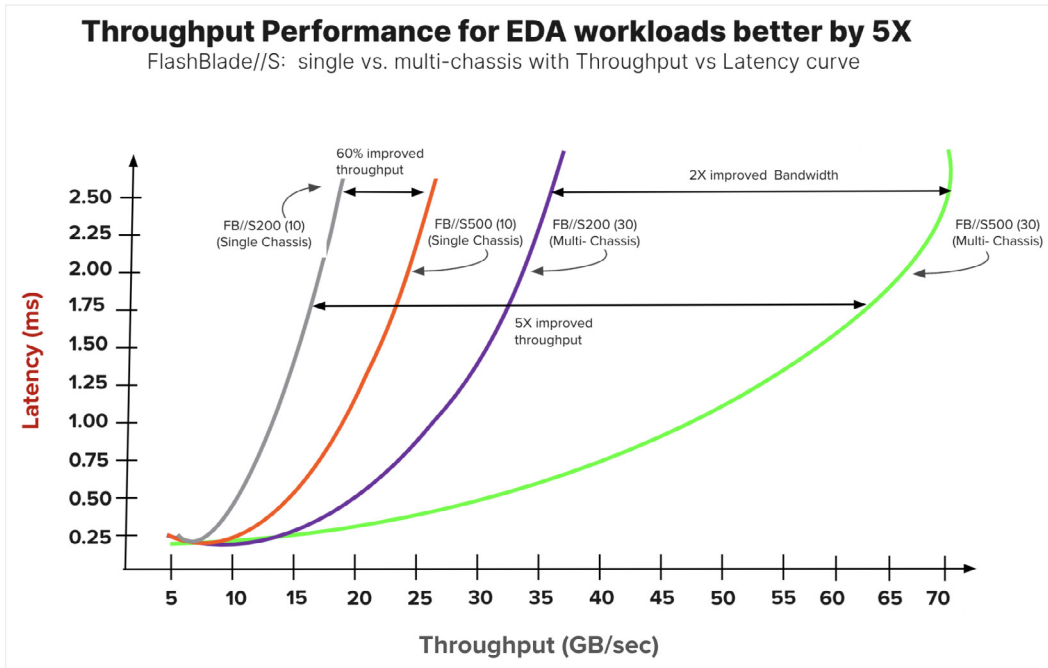


FIGURE 10 Latency vs. IOPS comparison between FlashBlade//S single and multi-chassis

- Based on the latency vs. throughput curve in Figure 10), scaling tapeout workloads that require high throughput can be 2X faster with S500 multi-chassis compared to S200 and 5X faster compared to a single chassis S200. While the latency stayed below 5ms for the most part of the EDA mixed workloads, jobs continued to complete at high speed even though the latencies went higher at maximum throughput levels. Apart from throughput scalability, the FlashBlade//S500 multi-chassis provides reduced power consumption with more compute power per watt for storage in the data center.

Conclusion

The FlashBlade//S platform offers a new level of performance for EDA and HPC workloads that are run with large server farms. As workloads in these farms increase due to the increased complexity in sub-10nm chips, the chip design teams require more IOPS and throughput performance from the underlying storage on demand. The IOPS and bandwidth performance from the single chassis FlashBlade//S500 is more than 2.2X and 75% greater respectively, compared to Gen 1 30 blade FlashBlade.

The metadata workload for small files are improved with more powerful CPUs on each blade, while the throughput for reads and writes are highly parallelized to the back-end DFMs compared to the Gen 1 FlashBlade. This means that the chip design business owners do not have to provision more storage to improve the metadata performance where additional capacity is never used. This reduces the overall storage footprint in the data center without impacting the growing performance requirement, helping EDA companies to meet their environmental, sustainability, and governance (ESG) goals, while improving performance for ever expanding EDA workload requirements.



The linear performance and capacity scalability is not limited to a single chassis. The multi-chassis FlashBlade//S200 and FlashBlade//S500 showed IOPS and throughput results that scale with more blades and DFMs. As the EDA workloads scale with more CPU cores for short and long running jobs in the compute farm, FlashBlade//S multi-chassis has the ability to process the jobs at high speed for a faster TaT. The improved data reduction capability in FlashBlade//S platform further helps to store large volumes of data during the life of an active project.

Faster storage performance has a direct impact on the job completion time. Faster completion time leads to more digital and analog design projects, thus optimizing EDA tool license costs and maximizing business value. Faster deletes using native FlashBlade technology and host-based utility like [Rapid File Toolkit](#) for quickly removing directories/files in parallel to support critical business functions.

In addition to on-premises CAPEX options, Pure Storage also provides Evergreen® [subscription](#)-based OpEx models to benefit chip design businesses with storage that can be consumed on-demand and aligned with chip design life cycles as a flexible alternative to standard CAPEX procurement. This, along with [Connected Cloud solutions with Microsoft Azure](#), provide IT teams the greatest performance, cost efficiency, and agility to meet their dynamic manufacturing needs.

Additional Resources

For more information on FlashBlade for EDA and HPC workloads:

- [Electronic Design Automation Solutions](#)
- [FlashBlade//S Fast File and Object Storage Platform](#)
- [Electronic Design Automation Solution Brief](#)
- [High Performance Computing Solution Brief](#)

purestorage.com

800.379.PURE

