

**Pure Storage** uses its own product, the FlashBlade™ storage array, as an important tool in enhancing its software development. By mounting log data directly onto the FlashBlade platform, the team removes a time-consuming and costly operation from the software-development pipeline. The FlashBlade storage platform has delivered significant benefits: the flexibility to allocate resources with optimal efficiency, better performance with low latency and high throughput, simplicity by eliminating time-consuming storage-management tasks, and faster time-to-market with a single high-performance storage platform.

**BUSINESS TRANSFORMATION**

Pure Storage is deploying its FlashBlade storage platform to improve the efficiency and productivity of its software development process.

**GEO**

North America

**INDUSTRY**

Technology

“Using our FlashBlade system as a single centralized storage platform, we can add, delete or modify services easily.”

Evan Driscoll,  
*vice president of engineering*

**HOW PURE STORAGE ACHIEVES SOFTWARE DEVELOPMENT INNOVATION WITH ELASTIC SCALE-OUT STORAGE**

Establishing more efficient processes for developing, testing and deploying software is growing not only in importance, but also in complexity. That’s because development teams are constantly balancing the imperatives of software quality and stability against the pressure to accelerate time to market.

The software development team at Pure Storage, the market’s leading independent all-flash data platform vendor, can attest to this reality.

In an ongoing effort to improve the software-development process — not only at Pure Storage, but across the industry — numerous technologies and techniques are being employed, including containerization, cloud computing, big-data analytics, open-source, Jenkins and Spark, to name just a few.

As a company whose software is purpose-built to exploit the full potential of flash for the cloud era, Pure Storage realizes the importance of applying many advanced tools to support software development. Pure employs nearly 400 software engineers whose development efforts provide many of the company’s most valuable competitive advantages. This includes de-duplication and data compression, Web-based product management, S3 protocol support, hardware-accelerated distributed transactions, non-disruptive upgrades and more.

**FLASHBLADE PUT TO THE TEST**

Recently, Pure Storage began using one of its own products, the FlashBlade storage array, as an important tool in enhancing its software development. FlashBlade employs a blade architecture that offers an elastic scale-out system that delivers ultra-high performance to multi-petabyte-scale data sets. It combines commodity hardware components with unique software and ultra-dense packaging to achieve extreme performance, scale, density and efficiency.

While software development at Pure Storage has many similarities to best practices at other companies, there is a significant difference: the Pure software is written solely to perform in conjunction with its massive storage arrays.

“We aren’t really software development, we are infrastructure development,” observed Evan Driscoll, vice president of engineering at Pure Storage. “Software development at Pure is an ongoing iterative process in which hardware is always involved in test

**COMPANY:**

Pure Storage  
[www.purestorage.com](http://www.purestorage.com)

**USE CASE:**

- Software Development: containers, cloud computing, big-data analytics, Jenkins and Spark

**CHALLENGES:**

- Need to establish more efficient processes for developing, testing and deploying software.
- The close interaction between hardware and software requires a longer and more complex review cycle.
- Ultra-high availability is essential.

**IT TRANSFORMATION:**

- Flexibility to allocate storage and compute resources independently allows resources to be used with optimal efficiency.
- Low latency and high throughput mean storage is never a bottleneck.
- Time-consuming storage-management tasks are eliminated in favor of an NFS storage platform that scales, without interruption, almost infinitely.
- A single storage platform eliminates contention for resources, reducing friction that can slow the development process and delay release.

activities. Sometimes we can test in a virtual environment using Docker containers, but eventually we always have to involve hardware.”

The close interaction between hardware and software, Driscoll noted, “means we have a longer and more complex review cycle. And because our products are used in settings where ultra-high availability is essential, we must test for a level of reliability that’s not expected for, say, Web software.”

Pure Storage takes a radically different approach — in both technology and business model — to meeting the needs of enterprise-level mass storage. So, it is not surprising that it thinks creatively about applying its own advanced storage technology in a software development environment.

“Traditionally, the focus of IT infrastructure for software development has been primarily on the compute side, making sure there is enough processing power for coders to do their work. Storage is often closely tied to servers in ways that impede productivity — for instance, by making it complex and time-consuming to spin up new development environments, or requiring hours-long waits to refresh or replicate databases,” Driscoll noted.

Adding to the trends of containerization and virtualization, Pure Storage has introduced a third key technology — centralized all-flash storage. Together, these technologies are breaking rigid bonds between compute and storage, adding enormous flexibility to the software development process while simultaneously simplifying the task of storage management. In addition, FlashBlade has opened the door to a new way of analyzing log data from software development.

At Pure Storage, a 15-blade, 52TB FlashBlade system supports the work of about 100 engineers who develop and test code used in the FlashBlade product. In more than a year of use in production environments, the FlashBlade array has delivered significant value in three areas:

- Applying analytics to log data
- Moving more services from stateful to ephemeral
- Decoupling storage and compute functions

**APPLYING ANALYTICS TO LOG DATA**

Each day, the software development activities at Pure Storage generate about 5TB of log data which, if properly analyzed, can yield valuable insights that will improve the quality of software and make the development process more efficient.

It’s possible that a quality assurance engineer can look at log data and identify failures in coding. But processing that much data can be a challenge. So, engineers at Pure Storage are working on a project to automate the process of identifying errors from log data. Initially, they used Elasticsearch (a search engine based on Lucene) as a tool to identify failed states, but this required significant pre-processing of data. As an alternative, the project team mounted the log data directly on the FlashBlade system, using Spark to develop analytic tools.

By mounting log data directly onto the FlashBlade platform, the team removes a time-consuming and costly operation from the software-development pipeline. Developers can access data from shared storage directly through NFS instead of using specialized software that adds cumbersome steps to the development and testing process. The flexibility to quickly spawn new clusters of analytics infrastructure — without having to copy or move data — enables data scientists to freely experiment with different scenarios and iterate more quickly through alternative approaches.

In addition, using FlashBlade lowers costs in two ways. First, it is unique in hosting both the log data and the analytics, eliminating the need for separate hardware platforms. Second, skipping the pre-processing step means less compute power is required to perform the log analysis.

### MOVING MORE SERVICES FROM STATEFUL TO EPHEMERAL

Driscoll describes the software-development process at Pure Storage as “a lot of distributed systems working together — scheduling, resource reservation, metrics, etc. — most of them micro-services,” said Driscoll. “One of our objectives is to move as many of these services from stateful to ephemeral as possible. That means we don’t have to worry about which machine a service is running on, and how much storage that machine needs.”

Driscoll added, “Using our FlashBlade system as a single centralized storage platform, we can add, delete or modify services easily. Let’s say, for instance, that we have a Jenkins machine that suddenly becomes much more active and needs more CPUs. All we need to do is turn on a bigger machine; the data is already there on the FlashBlade. I don’t have to migrate it from one machine to another.”

According to Driscoll, a high-performance centralized storage platform “unlocks a whole new class of ephemeral services. Having an ephemeral test environment improves our reliability and accuracy by protecting against outside state/environment changes. And having ephemeral services makes our testing more scalable and resilient against local machine failures.”

### DECOUPLING STORAGE AND COMPUTE FUNCTIONS

A single FlashBlade platform, providing a central NFS store, can accommodate huge amounts of data (8TB or 52TB per blade) and provides ultra-high performance (up to 17 GBs bandwidth and 1 million IOPS per 4 rack units). “You can be much more flexible with your compute when you have a fast, centralized storage system,” Driscoll noted. “For example, you can spin up compute on the fly in Docker containers across many machines that all point back to a single storage location. You can scale your compute, up or down, to meet demand without worrying about matching storage to whatever job is running on each machine.”

Driscoll reported that the performance and scalability of a FlashBlade platform allows greater flexibility. “For example, we have deployed ephemeral Jenkins instances to several VMs, all of which tap centralized storage on the FlashBlade platform. And that same storage system supports all of our analytics work.”

### CONCLUSION

The FlashBlade storage platform has delivered significant benefits in four areas:

1. **Flexibility:** The ability to allocate storage and compute resources independently allows resources to be used with optimal efficiency, and encourages more iterative creativity among the development team.
2. **Performance:** Low latency and high throughput mean storage is never a bottleneck, and multiple simultaneous workloads can be handled easily.
3. **Simplicity:** Time-consuming storage-management tasks are eliminated in favor of an NFS storage platform that scales, without interruption, almost infinitely.
4. **Faster time-to-market:** A single high-performance storage platform eliminates contention for resources, reducing friction that can slow the development process and delay release.

“An ephemeral test environment improves our reliability and accuracy by protecting against outside state/environment changes. And having ephemeral services makes our testing more scalable and resilient against local machine failures.”

Evan Driscoll,  
vice president of engineering



**info@purestorage.com**  
www.purestorage.com/customers