Pure Storage FlashBlade

Overview

The Pure Storage FlashBlade is a custom designed NAS system that scales by adding additional blades that contain both the processing function and NAND flash storage modules. With a unique design, Pure Storage has brought a NAS system to market that does not rely on commodity components. Object storage support was subsequently added to FlashBlade. Objects are stored in native format.

FlashBlade is a storage system for unstructured data, currently supporting file access using the NFSv3 and CIFS/SMB 2.1 protocol and object access with S3 protocol API. At some point in the future, protocol support for CIFS/SMB 3.0 will be added for NAS support.

The FlashBlade//S models represent the second generation of FlashBlade systems and offers two distinct models - //S200 and //S500. The //S models utilize a 5U chassis, and a more modular design than the previous generation by utilizing Pure's Direct Flash Modules.

The S200 blade has a focus on capacity and efficiency and utilizes a high level of compression. The S500 blade bypasses some of the compression to instead focus on greater performance. The two blade types cannot currently be intermixed.

As an all-flash file and object storage system, FlashBlade//S is used for acceleration of unstructured data storage. Pure Storage cites big data analytics and genomics common use cases.

Pure Storage also offers FlashBlade//E systems that focus more on large scale unstructured data repository use cases. The //E systems emphasize economic scalability with less performance focus than their //S counterparts.

Highlights

- All Flash NAS system with NFSv4 and SMB protocols
- S3 object support for high performance object access
- Up to 10 blades in a 5U rack with 1 - 4 DFMs per blade
- Clustering of up to 10 FlashBlades
- Custom interconnect between flash blades
- Compression data reduction
- Encryption
- Evergreen program for hardware updating
- Asynchronous replication

As a newer product, FlashBlade was initially missing some of the features implemented in other products, however FlashBlade has received new features and capabilities as Purity//FB continues to be updated. The Purity//FB 3.0 release added support for file and object replication, file system rollback, Kerberos authentication, and audit logging. Asynchronous replication of file and object was added in June 2020.

The //S models support up to 10 blades in a 5U chassis with a minimum of 7 blades, while the previous systems supported 15 blades in a 4U chassis. Up to 4 Direct Flash Modules can be added to the //S blades for a modular approach to scaling storage, while the older generation blades contained the flash modules themselves. 48 TB and 24 TB Direct Flash Modules are supported. Data is protected using a RAID 6 parity (called erasure coding) distributed across blades.

Each blade contains a single processor, a flash controller, DRAM, NVRAM, an FPGA used for data movement and encryption, and connects to up to 4 Direct Flash Modules. Blades are interconnected using an internal Ethernet switch called a backplane fabric running a custom protocol with aggregate bandwidth of 2.6TB/s. The switch is part of a backplane card that connects to each blade through high-density connectors. There are dual fabric manager processors connected to the switch card that manages data distribution and I/O balancing as well as providing the external network connections using TCP/IP and NFSv4.

Up to ten FlashBlade chassis may be clustered together, for a total of 100 blades and 400 Direct Flash Modules, to provide a global namespace across the systems. This multiplies the capacity and bandwidth for a single FlashBlade system. Blades may be added to the additional chassis in the cluster one blade at a time.

The internal distributed (across the blades) filesystem is implemented using a database with key-values to serve as mapping tables to files or objects. The key-value table is in DRAM with the resolution of location of files or objects retrieved from tables stored with data on NAND flash. Pure Storage refers to this as an underlying object storage implementation, which can be confused with object storage access. The database implementation for the filesystem leads to the nearly infinite scaling claim. The expected service time is 200 microseconds for access in the all flash implementation.

In June 2020 Pure Storage added Safemode to FlashBlade. Safemode is a policy-controlled immutable copy for ransomware protection.

File replication and S3 replication (called S3 Fast Copy) were added in June 2020 as well. S3 versioning, audit logs, and SNMP are supported.

Available in 4Q2020, Pure Storage offers FlashRecover which is an integration of FlashBlade and Cohesity DataProtect software executing on Cohesity scale-out compute nodes. Jointly developed as an integrated solution, FlashBlade is a target for parallel backup and restores using NFS protocol.

Usage

As a NAS and object storage system, FlashBlade targets high performance unstructured data usage with a system designed to exploit flash technology and a custom, scale-out architecture. File-based applications with need for performance such as data analytics, especially in real-time, and high-performance computing applications would benefit from the acceleration. Newly developed applications that use object storage may have performance requirements that would need FlashBlade.

The //E systems have less performance oriented target than the original FlashBlade or //S models. The //E systems focus on large scale unstructured data repositories while offering an all flash competitor where spinning disk systems are often used.

- Characteristics
  - Performance – The scale out system with up to 10 blades per chassis is addressing the latency time for access to file data with an expectation of 200 microseconds latency. Load balancing is performed across the blades, which could be referred to as nodes.
  - Availability – N-way scaling across the blades provides a high availability configuration. The minimum number of blades is seven, ensuring HA and data protection in a RAID 6 method. Interconnects between blades is an internal switch on the backplane with an aggregate bandwidth of 2.6 TB/s and an external bandwidth of 16 GB/s. The aggregate bandwidth across clustered FlashBlades is 75 GB/s.
  - File and object replication for DR.
  - Single file system with single namespace.
- Applications
  - The Pure Storage FlashBlade NAS system is targeted at performance demanding file applications. Called out specifically are big data analytics, genomic processing, and other engineering or scientific usages.
  - With object support using S3 protocol, the system can be used for high performance object access.
- System environments
  - The NAS support currently is for NFSv3, NFSv4 and native CIFS/SMB 2.1. Also added is additional system software to support file sharing and locking for simultaneous access from the two remote file system protocols.
  - Object access with the S3 protocol
- Deployment and Administration
  - As a new system, Pure Storage will likely participate in the installation of the system. In the future, the FlashBlade is expected to be customer installable and supportable.
  - Performance and capacity monitoring are available through Pure1 cloud-based monitoring and support.

Key Capabilities

Hardware Architecture

There are up to 10 blades in the 5U //S chassis with a minimum of 7 blades. Data is protected using a dual parity encoding distributed across blades. Each blade contains a single processor, a flash controller, DRAM, NVRAM, and an FPGA used for data movement and encryption. 1-4 Direct Flash Modules are connected to each blade. Each blade must be of the same type (S200 or S500), with the same number and capacity of DFMs. Blades are interconnected using an internal Ethernet switch running a custom protocol with aggregate bandwidth of 2.6TB/s. The switch is part of a backplane card that connects to each blade through high-density connectors. There are dual fabric manager processors connected to the switch card that manages data distribution and I/O balancing as well as providing the external network connections using TCP/IP and NFSv4.1 and CIFS/SMB 2.1.

The NVRAM is used not only to protect from failure but also for coalescing of writes to the NAND flash chips. Data written to NVRAM is mirrored to two other FlashBlades before acknowledging the operation is complete to the host.

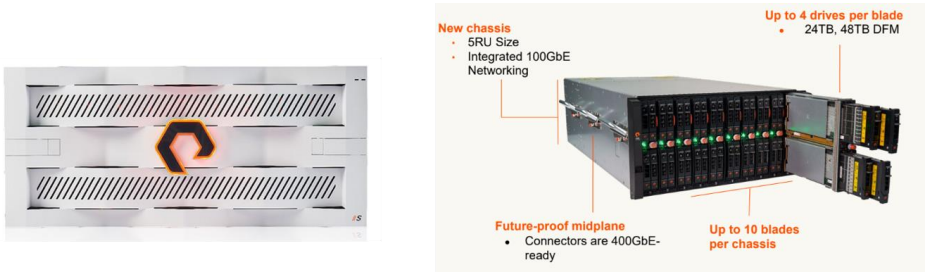

Figure 1: FlashBlade //S (source: Pure Storage)

Figure 2: FlashBlade //S Chassis (source: Pure Storage)

//E Hardware Architecture

FlashBlade//E follows a similar hardware architecture to the //S models, with the key differentiator being its scalability and multiple chassis types. The //E systems consist of 2 distinct chassis types – the EC chassis which contains both storage and compute resources and the EX chassis which contains only storage devices. The minimum configuration is 4 PB, consisting of one EC node and one EX node, and can scale in increments of 2 PB. Both EC and EX nodes can be

added, depending on whether the system requires additional storage only or both storage and compute, up to a total of 10 nodes.

Both//E chassis are 5U and support up to 10 blades, each supporting either 2 or 4 QLC Direct Flash Modules. The two chassis are networked together using two 1U external fabric modules (XFM).

Figure 3: FlashBlade//E EC Chassis (top) and EX Chassis (bottom) (source: Pure Storage)

Software Architecture

The Pure Storage Purity//FB operating environment is custom embedded software to perform the storage function. An instance of Purity//FB executes on each of the FlashBlades with communication between the blades using the backplane switch.

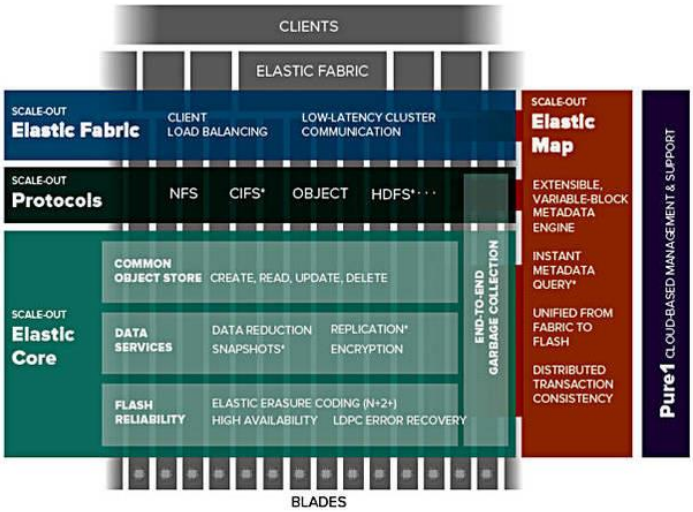

Figure 3: FlashBlade Software (source: Pure Storage)

Data is distributed across FlashBlades in stripes. Files are broken into 16MB chunks for the striping operation. If more FlashBlades are added to the chassis, the stripe size is increased for new writes to be across all FlashBlades.

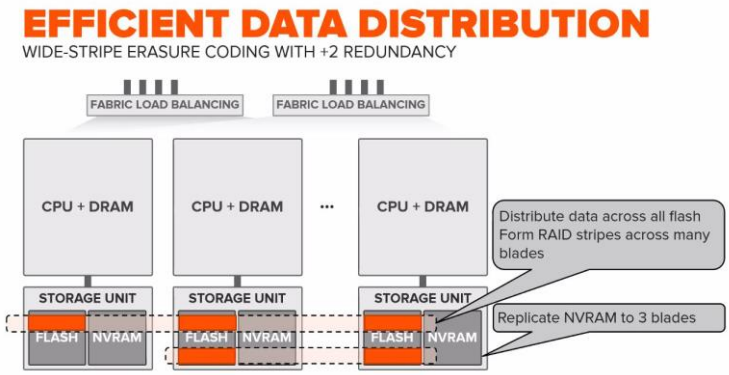

Figure 4: Elasticity Data Striping (source: Pure Storage)

A virtual IP address is presented to hosts for mounting of a filesystem. With data transfer, the IP address is rewritten in the packet as an inline process. With the virtual IP address and the internal switch, the system does load balancing across the FlashBlades.

Data Reduction

FlashBlade includes inline compression for all data stored on an individual blade done in software. Pure Storage expects a 3:1 reduction across most file data with exceptions in some vertical market data types. Compression is handled differently between the S200 blade and the S500 blade. The S500 blade reduces compression to reduce computational overhead and increase performance.

Failure Protection

RAID 6 (erasure encoding) protection is used across the FlashBlades in a chassis. Automatic rebuild and rebalance is implemented as part the RAID protection.

Replication

FlashBlade added functionality for asynchronous replication with the Purity//FB 3.0 update. Replication is supported between FlashBlade systems for both file and object data. Object data can also be replicated to Amazon S3 in native format.

Audit Logging

Audit logging functionality is supported by FlashBlade to provide visibility and compliance.

Safemode

In June 2020 Pure Storage released Safemode, which is a feature to counter ransomware. By policy control, snapshot copies of data can be made immutable in Safemode. Administrators and access through APIs cannot delete a Safemode copy. Access to the copy must be done by a named designee with Pure support.

Object Storage

The object storage implementation for FlashBlade is a custom implementation and uses the same underlying key-value database and tables as the NAS implementation. FlashBlade translates a bucket ID into and internal ID and distributes the data across the system (blades and chassis). A hash code is used for the internal information and which partition the object is stored into. The partition has a data structure to make to the data segments. This is the key-value database.

Currently, the S3 API is supported as the protocol over HTTP/S.

S3 versioning support is included. Statistics for S3 buckets can be displayed through the GUI.

S3 Fast Copy is replication of objects to another FlashBlade or to S3.

FlashRecover

Available in the fourth quarter of 2020, Pure Storage offers FlashRecover which is an integration of FlashBlade and Cohesity DataProtect software executing on Cohesity scale-out compute nodes. Jointly developed as an integrated solution, FlashBlade is a target for parallel backup and restores using NFS protocol.

Pure1 Cloud Support

Clouds-based monitoring and reporting developed by Pure Storage for FlashArray is offered with FlashBlade as well. Pure1 provides monitoring from any internet attached device for customers. Analysis and reporting information is processed as a SaaS offering.

Significant Announcements

- February 2023 – Purity//FB release 4.1: support for S3 Object Lock, object bucket quotas, scaling of up to 10,000 file systems.
- March 2023 – FlasheBlade//E announced. //E models are focused on scalability and economic efficiency, with less performance optimization than //S systems.

Futurum Group EvaluScale – Data Protection Software

The Futurum Group product review methodology "EvaluScale" assesses each product within a specific technology area. The evaluation of each product is based on its capabilities, with capabilities for each technology segment grouped into distinct categories. For the NAS and SAN EvaluScales, products are evaluated based on the following 4 criteria categories:

- Performance / Capacity
- Basic Functionality
- Advanced Capabilities
- Ability to Execute

The full NAS EvaluScale can be found here.

The full Object Storage EvaluScale can be found here.

The Futurum Group Opinion and Outlook for Pure Storage FlashBlade

Pure Storage has delivered a unique solution with FlashBlade, even though in its early version, there are many capabilities slated in follow-on releases. Use of flash for file and object access is inevitable as the price of flash continues to drop and data reduction becomes prevalent. Pure has been the first to showcase a new system design specifically with flash for storage. The system currently is a NAS solution with NFSv4 and SMB2.1 protocol support and object with S3 protocol. While initially missing some features found in other NAS and Object Storage systems, FlashBlade has seen continuous development.

Performance and cost would appear to make this an exceptional system and will become a competitive target. Scale out is addressed with the addition of blades in a chassis and clustering of up to ten chassis. With the updates to //S, FlashBlade has added functionality to scale and upgrade in a modular fashion, as well as provide two separate models for organizations to focus on either capacity or performance.

//E systems offer a compelling alternative to disk based file and object systems. Pure has leveraged the improving economics of flash storage, such as dense QLC devices. With the ability to scale by adding storage only EX chassis, Pure can offer an efficient, economic solution – competing with disk based solutions – where performance is not prioritized, while maintaining the advantageous of all-flash solutions.

The expectation for Alletra Storage MP is for it to become the primary block and file offering for HPE, and a major component of its GreenLake offerings. Likely, the Alletra Storage MP block solution will eventually replace the previous Nimble and Primera based versions. A notable omission from the Alletra Storage MP platform is support for object storage. While the traditional VAST Data solution supports combined file and object, the version offered from HPE is file only. The Futurum Group would expect HPE to make some type of object-based addition to the shared Alletra Storage MP platform in the future. An inclusion of object storage, whether internally developed or provided from an existing partner such as VAST or Scality, would round out the concept of deploying any type of block, file, or object software on a shared hardware platform.