NetApp FlexGroup: One More Layer Doesn't Make C-Mode Scale Out

800% growth in data volume is expected over the next five years. [80% of that volume is unstructured data, mostly generated by machines](#), and rich with insight waiting to be unlocked. Trends of such magnitude occur once-in-a-generation, and it has legacy storage vendors scrambling for answers. Last week, NetApp announced another upgrade to ONTAP C-Mode to try and address these big data challenges, adding yet another layer of virtualization: FlexGroup.

NetApp continues to struggle with the transition to C-Mode (Cluster Mode). It's becoming increasingly obvious that next-generation analytic "big data" workloads require true scale-out storage, which unfortunately C-Mode is not. C-Mode simply delivers federated management of individual appliances, leading to performance hot-spots and complexity. A key building block for FlexGroup is FlexVol, which allows a volume to be tied to a physical aggregate and node. By stitching multiple

FlexVol volumes, FlexGroup acts like a single large volume to clients.

While the announcement is interesting, it certainly has me scratching my head. If FlexVol is already widely deployed in their customer base, why not leverage it for unstructured data workloads?

FlexVol is simply not BIG enough, FAST enough, nor SIMPLE enough for unstructured data workloads, according to the NetApp blog. It is limited to 100 TB of data, or up to 2B files. Performance suffers with increasing metadata operations as data scales. FlexVol management also grows in complexity as data scales, as more nodes and aggregates are needed. As their blog states, "If you have a 10-node cluster, each with multiple aggregates, you might not be getting the most bang for your buck."

BIG, FAST, and SIMPLE: The Ethos of FlashBlade

Here at Pure, we agree that in this new era of unstructured data, customers need a data platform that is FAST, BIG, and SIMPLE, and that's exactly why we built FlashBlade. FlashBlade is the world's first AFA purpose-built for unstructured data, unleashing the power of modern data analytics. It is engineered with these three design principles at its core:

- BIG: 1.6 PB of usable storage in a 4U, size of a microwave (at an assumed 3-to-1 data reduction and including all overhead for resiliency, flash management, and metadata).

- FAST: Delivers up to 17 GB/sec and 1M IOPS per chassis with ultra-fast compute for metadata, with true linear scale via a bladed architecture

- SIMPLE: Integrated networking coupled with Pure1 management software- what more needs to be said?

Big data presents unique challenges for the storage industry. Predictable performance needs to scale seamlessly from small to large files – without changes in architecture, deployment model, or tuning. Metadata performance is key for many workloads. Ultra-efficient density and power is a must as data volume continues to grow exponentially.

For these reasons, Pure engineers realized early that an ideal solution cannot be patchworked, or retrofitted, using legacy approaches and technologies. With a combination of a clean sheet of paper and three years of arduous engineering work, FlashBlade was born and unveiled to the world.

NetApp FlexGroup: More Complexity to the Rescue?

We applaud NetApp for attempting to address the inherent limitations of FlexVol. However, in our view, adding yet another layer of retrofit complexity is a step in the wrong direction. Here are some reasons why FlexGroup falls short.

- *Capacity Limitations*: Each directory is linked to a FlexVol, which is tied to a physical aggregate and node with a limitation of 100TB and 2B files. This means that a directory has a forced limitation in size and number of files.

- *Performance Hotspot for Files*: Some analytics algorithms, like Monte Carlo simulations for portfolio risk analysis, require accessing the same dataset over and over again. Since FlexGroup lacks file striping, if a large file is hot, then the controller where the volume resides will inevitably be a performance limiter.

- *Performance Hotspot for Directories:* As mentioned above, a directory must be tied to a physical aggregate/node via FlexVol. This awkward constraint means if a directory is hot, then application performance inevitably suffers as a single controller is the bottleneck where the volume resides. In fact, if a hotspot develops in one of the FlexVols, there's no way to mitigate that by redistributing the data.

- *Perfect Scenario Required:* It seems the ideal situation for FlexGroup to deliver predictable performance is a workload distributed equally with flat utilization across all the nodes. Unfortunately, nothing in life is that perfect, especially in the world of unstructured data.

- *Complex Under the Hood*: FlexGroup is a complex plumbing of FlexVols glued together with junctions. Nested junctions need to be managed, particularly with export policies between lower level and higher level junctions.

- *Complex Above the Hood*: While FlexGroup may be simpler to manage than FlexVol, it still requires nearly 100 pages in its [Implementation Guide](#). This level of complexity just doesn't work in [the age of generalist IT](#).

FlashBlade is built from the ground-up with big data in mind, with a modern architecture that delivers performance natively without the limitations of legacy technologies. With a highly parallelized, distributed load balancing design at its core, it delivers scale-out performance from small to big files without any hotspots, management or tuning. FlashBlade brings real intelligence to the exploding world of unstructured data – so you can focus on getting the results you need, and not on the storage.

Get Started with FlashBlade

We agree. The right data platform for unstructured data needs to be BIG, FAST, and SIMPLE. Solutions originally built on legacy hardware (i.e. spinning disks) and software (complex layers-on-layers built over 20 years) are ultimately designed for prior generation workloads. We believe that for modern workloads that transform big data into actionable intelligence, a purpose-built data platform like FlashBlade is required.

FlashBlade is the only data platform in the world built for big data analytics powered by all-flash. Today customers are accelerating discoveries and insights in genomics, seismic processing, machine learning, and deep learning, to name a few. If you are interested in learning more, visit us at FlashBlade website for more details.

We can't wait to hear what exciting possibilities open up for you as you deploy FlashBlade.