How to safely use 8TB Drives in the Enterprise

by George Crump, Lead Analyst

After a few year hiatus higher capacity hard drives are coming to market. We expect 8TB drives to be readily available before the end of the year with 10TB drives soon to follow. And at the rate that capacity demands are increasing those drives can't get here soon enough. But, these new extremely high-capacity disk drives are being met with some trepidation. There are concerns about performance, reliability and serviceability. Can modern storage systems build enough safeguards around these products for the enterprise data center to count on them?

The Economics Advantage

A storage array populated with 8TB and 10TB drives will have a clear advantage over other hard drive based systems and certainly flash arrays. The cost per GB of these systems is going to be measured in pennies. These high capacity drives are going to be a significant step forward for the hard drive industry since flash storage has been narrowing the price per GB gap. 8TB and 10TB drives will re-widen that gap, making predictions of an all-flash data center a distant memory that never came to fruition.

It is more likely that data centers will be a mixture of high capacity systems and high performance systems. And in some ways it may make sense for those systems not to mix. This allows storage designers to surround differing systems with hardware and software optimized for these two distinctly different use cases. But we will need to see independent software emerge that can address data movement between separate, independent storage systems from a mixture of vendors.

Potentially even more valuable than the cost per GB gain are the gains that can be seen in density, meaning a need for fewer storage shelves or storage nodes in a scale-out storage system. The combination of lower price per GB and higher density will be welcome news to data centers that are creating data at an alarming rate. But, again this news will only be welcomed if the data center can deploy them with confidence that performance expectations can be met, that systems will recover quickly from failure and that the data can be reliably stored for a long period time.

The 8TB Performance Problem

Obviously 8TB drives are not going to perform like a 15K RPM low capacity drive, but most environments that are concerned about that kind of performance already have incorporated some type of flash technology – or will be. In fact, some of these capacity based storage systems can also support flash, while providing block, file and object access in an effort to truly consolidate the storage infrastructure. Whether in capacity systems or consolidated storage architectures, 8TB and 10TB drives will be used for lower access, less performance critical environments.

Is There a Reliability Problem?

Thus far there has been little indication that an 8TB drive will be any more or less reliable than its 1TB and 2TB counterparts. Plus, data protection technologies like RAID, replication and erasure coding exist to protect against failure of a given drive.

There IS Serviceability Problem

While the reliability of 8TB and 10TB drives could be debated, the serviceability cannot. A drive of this size in a RAID set is going to take a very, very long time to rebuild. A RAID5/6 rebuild has been estimated to take anywhere from 1 week to 3 weeks per drive. Storage Switzerland has predicted the death of RAID as a protection scheme for years, high capacity 8TB/10TB drives may just put the final nail in that coffin.

The Durability Problem

In addition to serviceability challenges there will also be durability challenges. While hard disk drives don't degrade in the same way that flash storage does, data on those drives can degrade and suffer from what is commonly called "bit rot". A system designed to house high capacity drives should also have the ability to verify stored data on a periodic basis and confirm that it is still readable. If an error occurs that data needs to be automatically recovered from a known good copy. This capability is especially important in the flash era, because the role of disk based systems will change from one of serving active primary data to becoming a long term repository for inactive data.

Solving the 8TB Problem

Addressing these challenges is going to require a storage system that's purpose-built for high capacity hard disk drives. In much the same way that all-flash arrays are designed specifically to provide performance, these high capacity storage systems will need to provide reliable and durable retention. Potentially, the two designs can be blended to offer a truly consolidated solution.

Solving the re-build problem

The key to solving the high capacity drive rebuild problem is for the storage system to have a granular understanding, of the data it's storing. This means that if a hard drive fails the entire drive doesn't need to be rescanned and rebuilt, just the data that needs to be recovered from that drive. Again, on traditional RAID systems with 1TB or 2TB drives, this process can take hours, an 8TB or 10TB system may take days if not weeks to rebuild back to a protected state. Another challenge with high capacity drives is that while a rebuild is happening the data protection level is decreased and another failure or two could lead to complete data loss.

These storage systems can provide data protection alternatives to traditional RAID by either replicating data across multiple nodes or using a parity based protection technique called "erasure coding". Replication is the less complicated of the two technologies, simply requiring admins to set an acceptable replication level (2, 3 or many copies of data) for each data type. If a drive fails and the data types that were on that drive fall below that replication level those data types are recreated on another known good drive. But that recreation is done via a series of copies, not CPU-intensive calculations of parity data. This makes the technique ideal for moderate capacity environments because they can leverage less expensive, less powerful storage nodes.

The downside to replication is that it requires 2X or more the original storage space. Thanks to the excessive capacity per drive and the density of data that those drives can store, the cost and data center floor space impact is greatly minimized. Much of this additional data consumption could be offset if the storage system leverages deduplication and compression. Compression by it self should deliver a 2:1 return. And for the right data set, when combined with deduplication, that ratio can grow to as high as 5:1. These ratios assume production data, but for backup data the returns could be 10:1 or higher. There is a point where even with this data efficiency offset, despite its inherent simplicity, replication could require too much data protection overhead. But thanks to high capacity drives that level is much higher than it used to be making it an acceptable strategy for the overwhelming majority of organizations.

When and if the data protection overhead of replication becomes too great, the alternative is erasure coding, a parity based data protection scheme that provides lower storage overhead than replication and even traditional RAID algorithms. It is designed to distribute parity across the high capacity storage nodes that are common in storage architectures, although it can also be implemented within a single array. Erasure coding also works on a sub-drive level so that in the event of a drive failure it only has to recover the actual data on that drive not read the drive from end to end.

Rebuilds of even high capacity drives should be very rapid since only the data actually on that drive needs to be recovered and, as is the case with replication, all the nodes can help in the rebuild process. While erasure coding does save on total storage capacity used and data center floor space, it does have a higher processor burden and may drive

up the cost per storage node. But in higher capacity environments that cost can be offset by requiring less capacity overhead.

Solving the Durability Challenge

Both replication and erasure coding can solve the durability challenge by continuously validating the quality of the data on the storage system. If the integrity of that data is suspect then it can be re-created from other copies of data. Essentially, the storage system responds to data corruption in the same way that it responds to a drive or node failure. But again, thanks to its granularity it only has to recover the specific component of data that has been corrupted. This means that a data set can be restored to good health very quickly.

Replication will typically verify data integrity by scanning the storage cluster via a CRC-like validation. Erasure coding, to some extent, has data validation built in. Erasure coding assigns a unique ID to every data segment based on a binary code of the data it contains. A data segment should always return the same binary code. If for some reason that binary code changes then it's reasonable to expect that the data has become corrupted.

Conclusion

High capacity drives, 8TB and larger, are on their way to the enterprise. Those enterprises desperately need the increased capacity and, more importantly, the density that these drives will provide. But concerns over drive recovery and data durability are legitimate. The good news is that an increasing number of storage infrastructures have at least two options available to them to address these concerns head on, making 8TB or greater hard drives economical and safe for enterprise use. The use of replication and/or erasure coding should one of the first considerations when selecting a new storage system.

Sponsored By Hedvig

Hedvig is a software defined storage system that takes a distributed systems approach to solving storage challenges. The software leverages commodity hardware to create a scale-out storage architecture that provides block, file and object storage services and complete enterprise storage capabilities. You can learn more about Hedvig in our briefing note or by visiting their web site directly.