WHITE PAPER

How TrueNAS Leverages OpenZFS

CONTENTS

1 Executive Summary
2 History of iXsystems
3 Overeview of TrueNAS
4 OpenZFS
    4.1. History
    4.2 Technical Overview
        4.2.1 Checksumming
        4.2.2 Snapshots and Replication
        4.2.3 Variable Block Size
        4.2.4 Compression
        4.2.5 Deduplication
        4.2.6 Read and Write Acceleration
        4.2.7 Using OpenZFS to create Hybrid Storage
5 Conclusion

1 Executive Summary

Organizations need to deploy more applications with fewer resources. TrueNAS is a unified, enterprise-grade array that leverages the OpenZFS file system. OpenZFS allows TrueNAS to segment reads from writes and to continuously verify data integrity, keeping data safe. OpenZFS overcomes many of the limitations of traditional file systems and volume managers. The result is faster, more efficient, more flexible storage at significantly lower cost using high-performance technologies, such as DRAM and flash, alongside traditional capacity-driven technologies like hard drives.

Whether as a replication target for business continuity, storage for VMs, or video/audio storage, streaming, or editing, TrueNAS provides a reliable, fast, and easy-to-manage solution. You don't have to rely on multiple spindles of power-hungry or short-stroked disks to get the IOPS for complex physical or virtual applications. TrueNAS provides high availability, variable block sizes, storage optimization, snapshots, replication, flash I/O acceleration, and caching, making it the right storage array for any use case.

2 History of iXsystems

Nestled in the heart of Silicon Valley, we have been committed to serving technology needs, with a focus on Open Source and enterprise hardware, since our beginning in 1996. From our inception onward, we've been perfecting our craft of making quality storage solutions and custom-built servers, and our dedication to a superior customer experience has been our guiding principle. See what our customers say about us at https://www.vendop.com/vendor/ixsystems-inc/reviews/.

3 Overview of TrueNAS

Many enterprises dedicate storage to a single server or application. By supporting multiple file and block shares, TrueNAS consolidates storage into one array, so you can deploy one TrueNAS storage array for multiple servers, physical, and virtual applications.

Instead of requiring you to acquire hardware and install storage software, we built TrueNAS as an integrated hardware and software solution. TrueNAS is powered by OpenZFS and supported by iXsystems. This gives you one place to get your support rather than dealing with multiple vendors. To learn more about TrueNAS, visit truenas.com

4 OpenZFS

4.1 History

The ZFS file system was originally developed by Sun for the OpenSolaris operating system and released under the CDDL Open Source software license. When Oracle acquired Sun, development of the Open Source version of ZFS ceased at "v28". Other operating systems were using ZFS, including FreeBSD, OpenIndiana, Mac OS X, and Linux.

In order to differentiate Open Source distributions of ZFS from the now proprietary Oracle version, the naming method for Open Source versions of ZFS moved from numbers to feature flags. In 2013, the main groups developing Open Source ZFS formed the OpenZFS project in order to promote the Open Source version of ZFS, called OpenZFS. TrueNAS CORE and TrueNAS use the latest FreeBSD version of OpenZFS.

4.2 Technical Overview

OpenZFS was designed for enterprise storage workloads and has many advanced features to prevent data loss and increase performance. Multiple levels of protection - checksums, snapshots, and replication - are all focused on protecting data and ensuring corrupted data is never returned.

OpenZFS is an advanced, modern file system specifically designed to provide features not available in traditional UNIX file systems and to simplify volume management. When OpenZFS writes new data, the blocks containing the old data are retained, allowing the previous version of the file system to be maintained. This is known as copy-on-write (COW). Since OpenZFS snapshots use this copy-on-write capability, they are created instantaneously, since all data composing the snapshot is already stored. Snapshots are also space efficient, since any unchanged data is shared among the file system and its snapshots.

OpenZFS was designed to be scalable, as seen in the following table.

Maximum Volume Size	256 quadrillion zebibytes (2^{128} bytes)
Maximum File Size	16 wexbibytes (2^{64} bytes)
Maximum Number of Files	2^{48} per directory and unlimited per file system
Maximum Filename Length	255 ASCII characters
Maximum Number of Snapshots	64 exabytes (2^{64})

4.2.1 Checksumming

Every block of data is accompanied by a checksum, stored with the block pointer pointing to it. This means that every block can be checked using separately stored metadata. Because this also applies to every metadata block in the file system, the entire file system can be cryptographically checked for correctness from top to bottom. Every time a block is read, it is checked for correctness. If the check fails, OpenZFS will rebuild from parity data if possible. If it's impossible, the read will fail. OpenZFS will never return corrupted data.

4.2.2 Snapshots and Replication

The copy-on-write model makes it easy to create and maintain in-place snapshots of file system state. A snapshot records all the data and metadata blocks that comprise the file system at the time of the snapshot. If those blocks are later changed or deleted from the live file system, the on-disk blocks are maintained for reference by the snapshot. Snapshots are read-only but can be cloned for use as a live file system without damaging the later file system or the original snapshot.

Additionally, the snapshot method makes it easy to replicate data remotely, incrementally sending only the blocks changed since the last replicated snapshot.

4.2.3 Variable Block Size

One of the performance-focused features of OpenZFS is its support for variable block sizes. In order to avoid thrashing storage media with many small write operations, data is banded together into larger blocks that are written to disk in a single transaction. This increases write performance when dealing with large amounts of data.

4.2.4 Compression

OpenZFS supports inline compression of data on a per file system basis. The OpenZFS implementation used by TrueNAS defaults to the LZ4 high-performance compression algorithm that is designed for multi-core CPUs and suitable for use in almost all scenarios. Traditional compression algorithms usually present a performance penalty when compressing mostly non-compressible data. LZ4 instead determines whether or not compression will result in sufficient savings, minimizing the impact for all workloads. Because less data has to be written and read from disk later, compression usually increases storage performance due to the reduced I/O load.

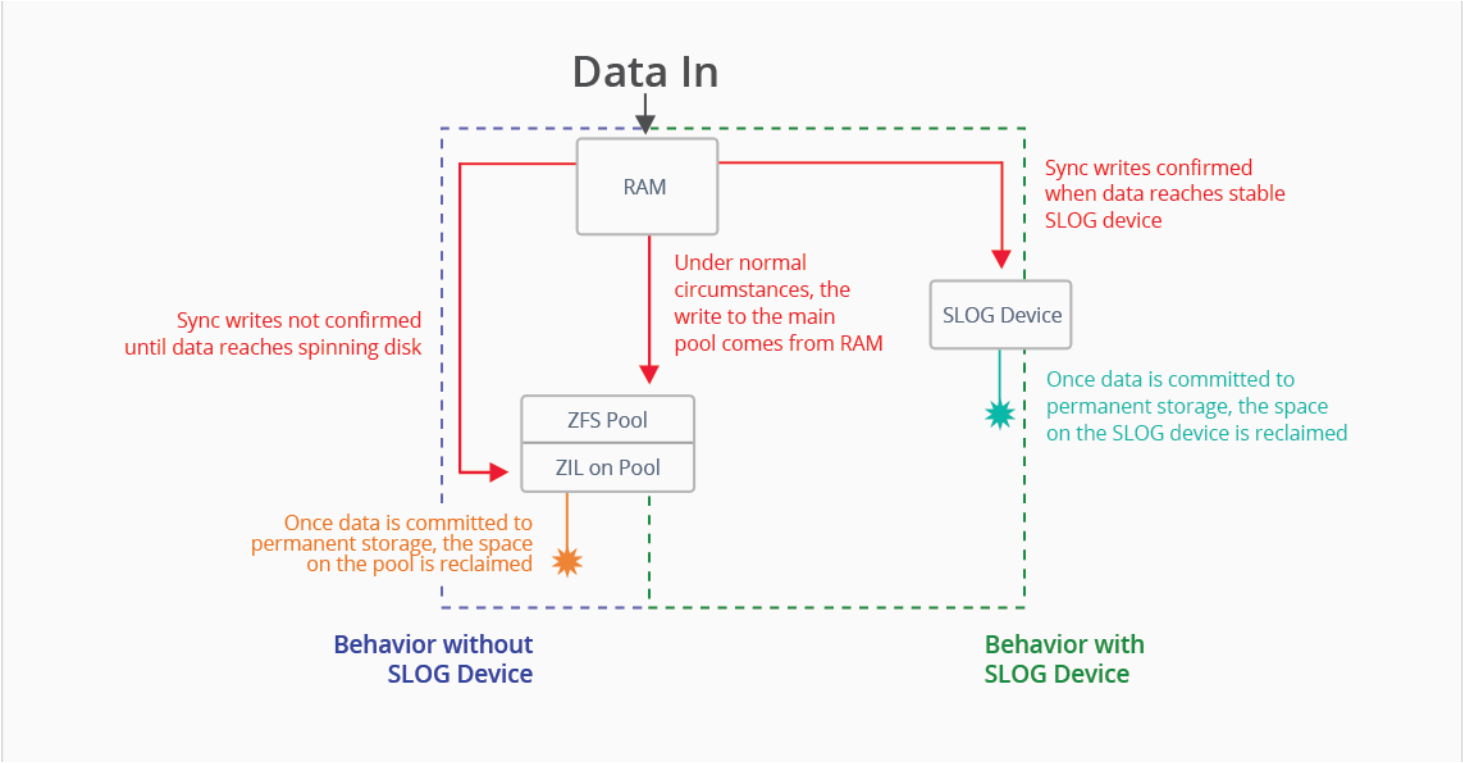
4.2.5 Deduplication

OpenZFS supports in-line deduplication for reduction of redundant data. Blocks written while deduplication is enabled are deduplicated across the file system. Due to the performance reduction of deduplication, it is not usually recommended for use in frontline systems. Systems can be purpose-built to leverage OpenZFS deduplication, by configuring with additional RAM, CPU and adding SSDs to act as Metadata vdevs.

Deduplication is part of the Intelligent Storage Optimization provided by TrueNAS.

4.2.6 Read and Write Acceleration

The core performance-enhancing feature of OpenZFS is the ability to use system memory (DRAM) and solid state drives (SSD) to accelerate read and write performance. By default, OpenZFS will use system memory to cache recently and often-used data using the Adaptive Replacement Cache (ARC) algorithm. For working datasets smaller than the amount of system memory used, this means incredibly high performance with very low latency, since RAM is significantly faster than flash or a spinning hard disk.



For larger datasets, OpenZFS also supports dedicated devices to cache data evicted from ARC to the the Level 2 ARC (L2ARC), expanding the ability of the storage system to provide high-performance access to larger sets of data.

Other storage vendors encounter performance bottlenecks if the application requires frequent small synchronous writes. This is not an issue with OpenZFS. Databases and virtual machines can afford no data loss, even in the case of power failure, so they utilize synchronous writes. OpenZFS can provide synchronous writes with its ZIL (ZFS Intent Log), which can be augmented by adding a Secondary LOG (SLOG). If the SLOG is on a high performance and persistent device, such as an SSD, write performance will be vastly improved. Also, if power is ever removed from the system, the ZIL is automatically replayed to ensure that no in-flight writes are lost.

4.2.7 Using OpenZFS to create Hybrid Storage

The ARC, L2ARC, and ZIL features of OpenZFS enable an OpenZFS-based array-like TrueNAS to implement a Hybrid Storage model. In the Hybrid Storage model, performance is achieved by reading from and writing to flash storage instead of spinning disks as often as possible.

By comparison, non-hybrid storage must use 10K and 15K SAS drives to provide high performance. The downside to that model is that high-speed SAS drives are power-hungry, have lower capacity than other disks, and are orders of magnitude slower than flash. This ends up costing more since more disks are needed to achieve the performance equivalent of a few flash drives.

With hybrid storage, high-speed SAS drives are less important as lower-cost and fewer Nearline SAS (NL-SAS) drives can be used instead. NL-SAS uses the disk technology of SATA with a completely compatible SAS interface. This allows NL-SAS drives to fit seamlessly into a SAS fabric, with the higher capacity and lower power consumption per gigabyte of SATA drives.

With NL-SAS providing the capacity and flash drives providing the performance, Hybrid Storage provides better performance than large SAS arrays, and often the same performance as all-flash arrays, with far higher capacity and much lower acquisition, maintenance, power, and cooling costs.

Conclusion

This paper described how TrueNAS uses OpenZFS and leverages its self-healing, data reliability, and performance-enhancing features. TrueNAS also includes high availability, performance tuning, certification by major virtualization vendors, and a modular field-service-friendly hardware design for enhanced reliability. The capacity scales to nearly 9PB and its Intelligent Storage Optimization delivers over two and a half times more effective capacity for the Hybrid versions and up to ten times more effective capacity for TrueFlash.

Unlike other storage vendors, you can use the enterprise features of TrueNAS without having to pay per feature -- it's all included. TrueNAS is a unified storage array that includes snapshots, storage optimization, multiple protocols, and replication for less than most competitors' basic model price. There's no question that TrueNAS is hands-down the best value in storage. Don't take our word for it, see why DCIG, a leading analyst site, ranked TrueNAS "Excellent" with "Best" in Class hardware. Learn more by calling one of our friendly, consultative Storage Architects at 1.855.GREP.4.IX, visiting our web page at TrueNAS.com, or emailing us at info@ixsystems.com. We look forward to hearing from you.