CASE STUDY: DEEPDYVE™

Protecting data with ZFS Snapshots & Replication on **TrueNAS**®

> "*Losing this data to corruption or otherwise would spell the end of our company, and TrueNAS gives me peace of mind. I have 100% faith that this machine is never going to lose my data.*"

- James Satterfield

THE STORAGE CHALLENGE

DeepDyve serves up millions of documents from scholarly and professional publishers to its worldwide user base on a 24/7 basis. Typically a costly service, DeepDyve offers an affordable means for its users to access leading journals on a timed, read-only basis. In order to do this, they must convert the raw data from publishers into metadata for users to search their database, as well as PNG and HTML5 formats for users to view via a web interface.

DeepDyve needed a solution that would be able to store the data and keep up with their ever-increasing publication archives for over 50 million published documents. Putting together new NFS servers every time storage needs expanded was a costly solution which also required valuable time from IT administration. A solution that was easy to scale and that didn't require dedicated IT staff to maintain would be a substantial improvement and a considerably better investment over time.

Due the critical nature of this data to their business, they needed to ensure that their data was as secure as possible. Additionally, restoring the amount of data they have on hand can potentially take weeks or even months under some methods. DeepDyve needed a solution they could be confident would protect their data and provide for timely restoration when needed.

TrueNAS

COMMERCIAL GRADE EASE OF MANAGEMENT AND SCALABILITY WITH TRUENAS

DeepDyve had known they wanted a ZFS-based solution for its feature set, reliability, and protection from data corruption, and had utilized FreeBSD as an operating system on their NFS storage servers previously. This previous method provided snapshot capability and ZFS's provided built-in data integrity features, but it also came with the hazards of deploying and maintaining their own hardware as well as requiring them to keep a FreeBSD/ZFS expert on staff to manage their data and to take care of any issues that might arise. Expertise of that sort is difficult to come by, not to mention costly, so it left them in a precarious situation should their expert leave.

Having worked with iXsystems in the past, DeepDyve decided to give TrueNAS a chance. It came with all the storage functionality of FreeBSD and data integrity features of ZFS, all administered from a comprehensive Web Interface. This allowed DeepDyve to simply utilize their existing IT staff for basic administration duties rather than having to bring in a dedicated administrator. It also meant that hardware was supported, qualified, and guaranteed to be compatible with the operating system, eliminating trial and error in builds and substantially reducing any hardware-based issues that needed to be dealt with.

TrueNAS storage expansion shelves also made upgrading their storage painless. Starting off with appliances, DeepDyve also has implemented expansion shelves, giving them the ability to expand their storage capacity by nearly 200TB with 45 hot swap bays.

iX systems

"We serve data to thousands of customers across the globe and have never run into an instance with our TrueNAS systems where we didn't have enough performance."

- James Satterfield

Another primary factor in selecting TrueNAS was full Commercial Support offered for both the hardware and software. This means that if a problem arises, they have experts available around the clock to work on their problem. It also means that DeepDyve's IT staff can stay focused on their top priorities, instead of managing and troubleshooting storage on their own.

Deploying TrueNAS, also helped DeepDyve to reduce administration time due to its enclosure management feature. Without enclosure management, finding the physical location of a failed drive can be a big headache. The storage experts at iXsystems eliminate this problem by setting up TrueNAS to properly associate the hard drives from the software side with where they are physically located in the storage array. This means that if a hard drive fails, it's just a matter of looking for the LED indicator on your system.

TRUENAS AS BACK END STORAGE OVER NFS

DeepDyve uses their TrueNAS with expansion as backend storage over NFS to 62 virtual filesystems for the document repository for all their customers, which currently comprises over 50 million individual published documents, 900 million files and about 25TB of usable storage capacity, replicated and RAIDed data. The data is then accessed in read-only form from a web-based viewer. DeepDyve utilizes a Hadoop Storage cluster to convert publisher data (PDFs) to PNG and HTML5 formats, and those PNG and HTML5 files are stored in the TrueNAS system.

Their secondary TrueNAS appliance serves as replicated storage for their published documents as well as primary storage for their internal workgroups, comprising about 90TB of raw storage between both appliances. The systems both utilize 4 x 1GB network ports with LACP aggregation, over bonded Force10 S50N switches, which gives each TrueNAS appliance two interfaces it can use for network transfers.

REPLICATED STORAGE WITH TRUENAS

Beyond creating system rollback images on the local storage server, snapshots can be used to replicate data on local and remote hardware. Because DeepDyve is in the content delivery business, this means that their content is their business. If a system goes down, their ability to serve their customers goes down as well. By setting up replication, DeepDyve ensured that if there was ever a problem with their primary storage, they would just need to point their virtual machines to their secondary storage to get back on.

ZFS snapshots are a powerful feature that allow the user to easily back up their system incrementally on a regular basis without having to store multiple copies of the same files. The TrueNAS web interface allows administrators to schedule system snapshots on a periodic basis, giving granular options for the exact days and times. One of the biggest reasons that DeepDyve invested in TrueNAS was to get ZFS snapshots and replication in a Commercial Grade package. DeepDyve utilizes one TrueNAS with 44TB of raw storage with an expansion shelf, which replicates to their second TrueNAS with expansion shelf every 6 hours, ensuring minimal downtime in the case of any interruptions of primary storage.

In addition, ZFS snapshots allow for faster data restore times than other methods. Some methods like rsync can take weeks or even months once you get to the scale of data that DeepDyve stores. ZFS snapshots saves them substantial valuable storage space as well, since each snapshot only saves the information that has changed since the previous one. This feature also makes the snapshot process less resource intensive, meaning it won't disrupt overall performance.

CONCLUSION

By utilizing TrueNAS Unified Storage, DeepDyve has found a solution that meets all their storage needs. With users around the globe needing real-time access to hundreds of millions of files, TrueNAS provides the critical storage infrastructure & data security needed for DeepDyve to provide for their customers. It has also provided their IT staff with a solution that's easy to manage and easy to expand, allowing them to focus on their development priorities. In addition, ZFS features such as snapshots and replication allow their business to keep detailed system snapshots to roll back to prior system states & recover data quickly.

Based on open standards and engineered for the enterprise, TrueNAS delivers a robust feature set for nearly any storage application. TrueNAS appliances are designed to be easily scalable with rackmount expansion shelves. The Web Interface gives users a full set of storage administration tools available at the click of a mouse. Combined with iXsystems professional support, TrueNAS provides its users with the peace of mind they need to be confident that their data is in safe hands.

COMPANY INFORMATION

Founded by two bioinformatics who were frustrated with the inconvenience and cost of accessing peer-reviewed journal articles, DeepDyve aims to provide information professionals around the globe with simple and affordable access to authoritative research. Gaining access to that research is typically a costly and time consuming process, especially for those not affiliated with an academic library. DeepDyve provides a consolidated, affordable means of finding and reviewing the latest research, serving up millions of documents to its users around the globe.

ABOUT IXSYSTEMS

By leveraging decades of expertise in hardware design, its contributions to many Open Source software communities, and corporate stewardship of leading Open Source projects (FreeNAS and TrueOS), iXsystems has become an industry leader in building innovative storage solutions and superior enterprise servers for a global marketplace that relies on open technology.

Thousands of companies, universities, and government organizations have come to rely on iXsystems' storage, servers, and consultative approach to doing business. Headquartered in the heart of Silicon Valley since its founding in 1996, the dedication to white-glove customer service, industry-leading support, and transparent technological contributions has never wavered and continues to help lay the foundation for a new era powered by open technology.