Data Integrity

An Honest Look at Why Data Accuracy
and Consistency Matters.

What is Critical

The core functionality of a storage solution can be boiled down to one simple statement. **The solution must efficiently and effectively store and fetch data upon request.** How the solution does this—how it chunks up and treats the data through this input/output pipeline—makes or breaks the overall utility and viability of the solution. For performance-intensive work flows such as those carried out in post-production shops, data treatment and integrity are vital to achieving the speed necessary to support client workstation requests.

66 With OpenDrives, we don't have to worry about the quality or accuracy of our data being compromised. 99

Ben Boese, IT Director The Steadman Clinic

Our Unique Data Treatment for NAS

At its simplest and highest level, a storage solution has a few key tasks it must accomplish. It must take data generated by client workstations and store that data on disk until requested. Upon request, it must then fetch the data and provide it to any requesting client workstation. These input/output operations, or IOPs, are the core functions of any storage system. If it can carry these tasks out efficiently and effectively, then the storage solution probably has a certain amount of market viability.

What separates most viable storage solutions from each other is the way it carries out these IOP activities. The most attractive solutions for high-performance environments are those which are optimized at the operating-system level to treat that data—segmenting it, writing it to disk, and reading it from disk to render to requesting workstations—with the least amount of transactional latency and the maximum amount of data integrity. Storage is inherently transactional, and the faster and more accurately it can carry out transactions, the more the solution can be differentiated from the pack.

OpenDrives was created with the singular goal of creating the fastest and most efficient NAS storage solutions on the market. The base functionality within the proprietary operating system has been intentionally designed to transact data in ways that other solutions simply can't accomplish. Along the way, we have overcome a lot of the obstacles from which other solutions still suffer, things like static and inefficient block sizes, ineffective caching and fetching, and inadequate data integrity and compression. Let's take a look at how OpenDrives carries out these functions and brings new value to high-performance environments.
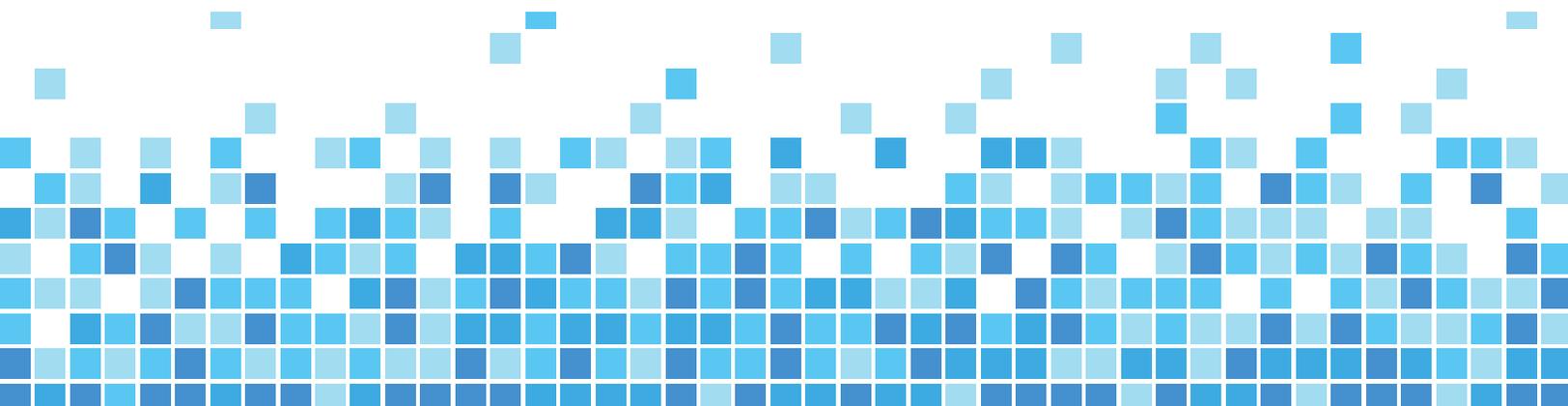
Guaranteed Data Integrity

Variable Block Size

The **block size** for a storage solution equates to the preset allocation of space for any given re-cord. Generally speaking, the smaller the block size, the more IOPS the storage solution must execute to write the record to disk. The larger the block size, the fewer IOPS. With more tradition-al storage solutions, setting this block size threshold is a serious consideration. If a block size is set to accommodate larger "chunks" of data, more than likely space will be wasted when records smaller than the block size must be written to disk. The extra reserved space will be filled up with zeros, so to speak. No matter how you look at it, most storage solutions introduce waste and ineffi-ciencies with block size configurations.

OpenDrives circumvents all this with intelligent variable block sizes. Our variable block size func-tion enables our solutions to use precisely what is needed, so a 4K bit of data receives a 4K block, while a 1MB bit of data gets that size block allocation. For larger records, our operating system concatenates the smaller individual blocks in order to reduce the IOPS requirements (and there-fore increase speed and performance) while still "right-sizing" the block-size allocations for space efficiencies. At the end of the day, it's about increasing speed while using only the amount of disk space truly needed to store the data. And it's all assigned variably on the fly.

Smart Caching

Another unique aspect of OpenDrives solutions is our intelligent use of **caching** in on-board mem-ory. When our solutions write files to storage, that data lands first in memory before being com-mitted to disk. However, that data is retained in memory in case it is immediately requested, so the disks work less in order to get that data to the requesting client. This is all happening instanta-neously underneath the hood, so to speak, and all our storage solutions benefit from this feature.

When data is read from disk, our software puts that data into memory at the moment it's read, and it stays there. That way, if your application requests that data again, it's read directly from the much faster memory cache. We also use predictive algorithms to carry out intelligent and adaptive "**pre-fetching**" of data. When a file sequence is being read, our software preloads the next file into the faster tier so that it's ready even before the requesting application requests it. The oper-ating system is constantly evaluating the files in memory, and as the files you're using become less active over time, they eventually are retired from cache to make room for new pre-fetched data. All of this activity is embedded at the operating system level, is completely transparent to the user, and does not depend on a specific type of drive (flash or HDD). What you get here is automatic accel-eration of performance with no effort or configuration needed.
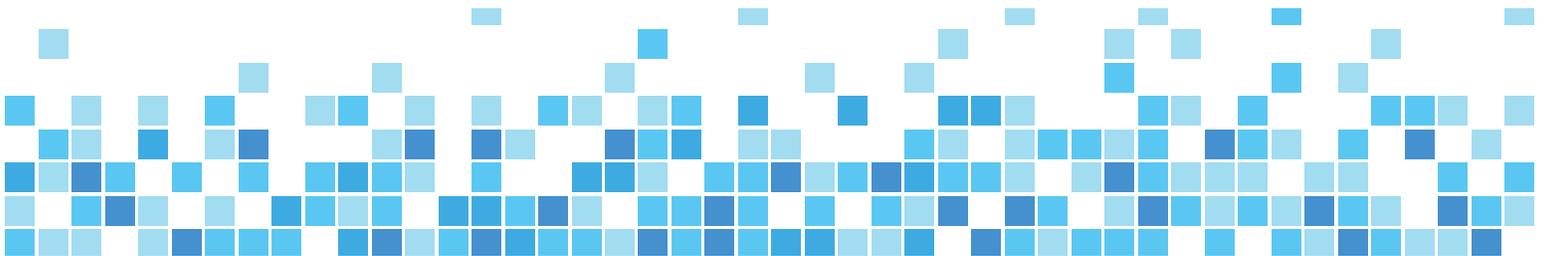
Fully Checksummed

A storage solution is pretty much worthless if it cannot ensure a very high level of data integrity. Speed and performance cannot compensate for inaccurate or corrupt data. This is the reason that OpenDrives solutions calculate a data integrity value, which is called a **checksum**, for every block of data when it's committed to disk. When that data is subsequently accessed and read, our solutions validate the checksum. This approach enables us to preserve the integrity of the data because, if even a small change in the data occurs down to a single bit, then the checksum is completely different. This makes it very easy for our solutions to determine if anything unexpected has changed or corrupted the data.

As a further guarantee of data integrity, OpenDrives stores data redundantly, both physically and logically. If the checksum for a given block does not match, the parity data is used to repair the bad block of data. What this means is that our solutions rewrite the incorrect (or corrupt) block and checksum with correct data and values. This means that OpenDrives storage solutions are self-healing due to this proprietary logic within our operating system.

Compression

Data compression is an important aspect of a storage solution in order to maintain efficiencies in disk usage. However, OpenDrives storage solutions avoid wasteful attempts to **compress data** that's already compressed. Quite a lot of media types are already compressed or have random bits that don't actually benefit from compression. Our operating system makes an analysis as to whether it's worth it to compress the data. If it determines that it's not, we skip it. What this means is that our storage solution works smartly and efficiently in order to use space judiciously while preserving performance by not wasting processing cycles.

Structured Data

Let's define the different types of data your organization produces. Some organizational data resides in very rigidly defined formats, typically through database or data warehousing solutions. We call this data **structured data**.

Unstructured Data

On the opposite end of structured data is **unstructured data**. This type of data does not have a predefined format (as does structured data, again because of the software overlay managing structured data). As such, it is often text-based and found in individual file formats. Unstructured data can present a variety of challenges for operations such as compression and de-duplication (removal of exact duplicate versions of files), because not much information about it (known as meta data) can be used by storage solutions to treat it efficiently.

OpenDrives creates shared storage solutions that are fast and efficient. However, data protection is key to preserving your data properly, so we've built into our software platform all the intelligence and efficiencies needed to preserve data integrity and work with data in the most effective manner possible. We strike a balance between speed and performance on one side, and the need for data integrity on the other. Maintaining this balance is what our storage solutions do automatically, so you don't have to.

We want to hear from you! We would love the opportunity to chat more about your data and how our OpenDrives network attached shared storage solutions can effectively underpin your storage infrastructure. We can show you how our solutions work, and we can explain all the concepts along the way.

www.opendrives.com

opendrives.