

# It's your space. You can use all of it.

HOW QUMULO ALLOWS 100 PERCENT  
STORAGE UTILIZATION—AND WHY  
ISILON AND NETAPP CAN'T

# CONTENTS

- Introduction** 1
- Why legacy scale-out storage solutions are built for wasted capacity** 2
  - Traditional data protection: From grossly inefficient to slightly less inefficient 2
  - Small file storage: Another way legacy vendors under-deliver on usable capacity 4
  - Rebuild operations: The hidden cost of disk recovery 4
- How Qumulo is different: Usable capacity means usable capacity** 5
  - Smarter block-level data protection 5
  - High-efficiency small-file storage 5
  - Rebuild operations that don't take a bite out of usable capacity 6
- Peak performance at 100 percent utilization** 6
- What Qumulo means for your data** 7

## INTRODUCTION

When Qumulo tells you the usable capacity of your storage, we mean exactly that: this is the capacity you can use to store files. It seems simple, yet this is a statement that even industry giants like Isilon and NetApp can't make. In fact, taking into account the inefficiencies of traditional data protection methods and the performance problems that can come with full utilization, most storage vendors make you leave as much as 30 percent of your capacity unused. In a world where you need all of your data at your fingertips, that's a major deficit.

In this white paper, we explain how Qumulo makes it possible to rely on all of your usable capacity for files—even at petabyte scale—without sacrificing performance or data protection. That's true no matter how many files you store or how large or small they may be. In fact, you can store billions of small files just as efficiently as large ones. It's your storage—you can use it any way your business demands, and you can use all of it. After all, storage management can be challenging enough without having to wonder whether “usable capacity” really means what it's supposed to.

## WHY LEGACY SCALE-OUT STORAGE SOLUTIONS ARE BUILT FOR WASTED CAPACITY

The difference between Qumulo and traditional storage vendors is deeply rooted, resulting from fundamental differences in their approaches to data protection, small file storage, and rebuild operations. We'll discuss these one by one.

### Traditional data protection:

#### From grossly inefficient to slightly less inefficient

Data protection is clearly non-negotiable. All enterprise-grade file storage systems are designed to prevent data loss if disks fail, and all rely on some form of redundancy or duplication of information across storage devices. The approach used, however, makes a tremendous difference in data protection efficiency, defined as the amount of data stored divided by the total disk capacity used.

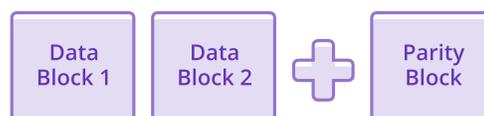
**Mirroring**, the most rudimentary form of data protection, is based on the creation of two or more full copies of the data being protected. Each copy resides on a different disk so that it is recoverable if one of the disks fails. This is effective in terms of recovery, but it's grossly inefficient, cutting in half the amount of capacity available for file storage.

Double mirroring, which keeps three copies of data for protection against up to two simultaneous drive failures, is that much more effective for recovery purposes—but also that much more inefficient, leaving two-thirds of “usable” capacity unavailable for files. In this case, mirroring for two-drive protection requires 3TB of raw capacity to store 1TB of file data.

At petabyte scale, it's obviously preferable to avoid mirroring as much as possible to avoid wasting two-thirds of your budget on storage you can't use for files.

**Erasur coding (EC)** is the best-known alternative for data protection that's more efficient than mirroring, as well as faster and more configurable. A key advantage of EC is the flexibility it offers. Admins can decide how to strike the right balance of performance, recovery time in the case of failed physical media, and the number of allowable simultaneous failures.

Working at the block level rather than the file level, EC makes it possible to protect data effectively without having to create a one-to-one copy of the entire data volume. Instead, block data is encoded into partially redundant segments that are stored across separate physical media. In the simplest example, known as (3, 2) encoding, three storage blocks are used to safely encode two blocks of user data; the third block, known as a “parity block,” is used for recovery.



The contents of the parity block are calculated by the erasure coding algorithm. Even this simple scheme is more efficient than mirroring—you are only writing one parity block for every two data blocks. In a (3, 2) encoding, if the disk containing any one of the three blocks fails, the user data in blocks 1 and 2 is safe.

Here's how it works. If data block 1 is available, then you simply read it. The same is true for data block 2. However, if data block 1 has been lost, the EC system reads data block 2 and the parity block, and then reconstructs the value of data block 1. Similarly, if data block 2 resides on the failed disk, the systems read data block 1 and the parity block.

A (3, 2) encoding has an efficiency of 67 percent—in other words, two-thirds of your available storage can be used for user data, while the remaining third is used for data protection.

Adding disks can improve the protection level. For example, a (6, 4) encoding, which has the same 67 percent efficiency as (3, 2), can tolerate two disk failures instead of just one. In other words, even if two disks fail at the same time, the system can still operate without downtime or loss of data. The extra protection with no reduction of efficiency isn't a free lunch—the process of recovering the (6, 4) encoded data requires more work than in the case of (3, 2) encoding, which means that the rebuild time is longer.

In enterprise-grade storage, EC can provide very high efficiencies. For example, (16, 14) encoding has an efficiency of about 85 percent, and still allows for up to two simultaneous drive failures without loss of data.

At this point, that 85 percent storage efficiency might look pretty good, especially compared with the 33 percent efficiency of two-drive protection using mirroring. If you need to store about 1PB of files, 1.2PB of raw capacity should cover it, right? Not necessarily. Once again, the reality behind the numbers is less clear than it might seem.

## Small file storage:

### Another way legacy vendors under-deliver on usable capacity

Although your storage vendor might report usable capacity as everything that's left over after allowing for erasure coding parity bits, don't assume that you can actually use all this space. It turns out that legacy scale-out storage systems don't do a very good job when it comes to small files. By small, we mean anything under 128KB.

There's a simple reason for this. Legacy storage systems are based on a decades-old design that forces them to mirror (or double-mirror, or even triple-mirror) files smaller than 128KB. We've already discussed the inefficiencies of mirroring—now it turns out they can be an issue even with EC data protection. Here's the worst part: the space needed for this mirroring is deducted from the usable capacity reported by the vendor. It's like buying a sandwich, then discovering when you unwrap it that there's a big bite missing.

How big is that missing bite? That's another problem: you have no way of knowing. You'd have to determine in advance the exact size of each file that you plan to write to see how many fall below that 128KB threshold, and there's no way to predict that. As a result, it's impossible to know how much usable capacity you actually have—or when you'll run out. Instead, you'll have to overprovision to make sure you're covered. That means you're actually wasting money in two ways here: one, for the “usable” capacity you're losing to the small file storage quirk—and two, for the additional capacity you're buying as a cushion.

That's no way to run a data-intensive business.

### Rebuild operations: The hidden cost of disk recovery

Always full of surprises, your legacy storage vendor might have one more way to steal back your promised usable capacity. Many systems consume storage capacity for rebuild operations while recovering from a disk failure—and if there isn't enough capacity available for this, the system will struggle to complete the recovery. For this reason, most vendors recommend that you limit your utilization to 80 percent of the usable capacity they've promised. Again, this calls into question the vendor's definition of the word “usable.”

## HOW QUMULO IS DIFFERENT: USABLE CAPACITY MEANS USABLE CAPACITY

Qumulo is a different kind of file storage company. We believe that usable capacity means just that—the amount of space you can rely upon to store files. With Qumulo File Fabric (QF2), you can use 100 percent of usable capacity for files. Here’s why.

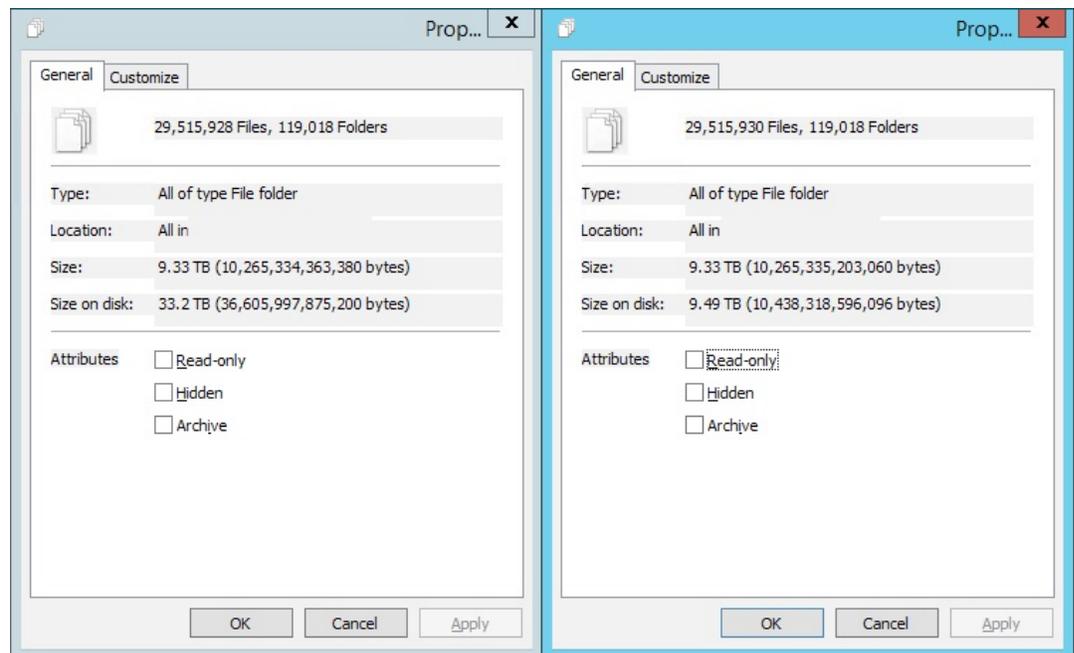
### Smarter block-level data protection

While legacy storage vendors focus on incremental improvements in efficiency, Qumulo has disrupted the industry with a fundamentally different approach. Instead of protecting data at the file level like others do, QF2 protects at the block level. With legacy file-based storage, data protection must be implemented on a file-by-file basis as well as within the metadata layer itself, a complex approach which limits the ability of these systems to optimize distributed transactions for their directory data structures and metadata. By enabling greater storage efficiency, QF2 enables typical gains of 20 percent in usage capacity for large files. And that figure doubles when small files come into the picture.

### High-efficiency small-file storage

When managing small files, block-level protection delivers storage efficiency up to 40 percent beyond file-based protection. This is especially valuable in an age of machine-generated data, which usually comes in the form of a large number of small files.

Here’s an example from an actual enterprise customer (before and after they came to Qumulo).



This customer migrated about 30 million small files from a legacy storage cluster to a QF2 cluster. The dialog box on the left shows the amount of space these files took up on the legacy vendor's system, which mirrors small files. The box on the right shows the amount of space the files take up on the QF2 cluster. As you can see, the legacy vendor's system needed more than three times as much space to store the same files—a full 33.2TB of usable capacity for 9.33TB of file data. On the QF2 cluster, it took only 9.49TB.

That's more like it.

In fact, with QF2, there's no difference in storage efficiency between large and small files. That makes it much simpler to estimate how much storage you'll need. Instead of wrestling with complex estimations of the mix of large and small files in your workloads and hoping they're not too far off the mark, you can just look at the web UI to see how much space is available. Your stored files will take the same amount of space regardless of how many are large or small.

### Rebuild operations that don't take a bite out of usable capacity

With QF2, there's no need to set aside usable capacity for administrative tasks such as rebuilds. Instead, the system sets aside the space it needs before it reports usable capacity. That means you can recover from drive failures even if the system is 100 percent full—and without having to monitor free space. QF2 also provides faster rebuilds than traditional RAID, and doesn't introduce performance hotspots after drive failure.

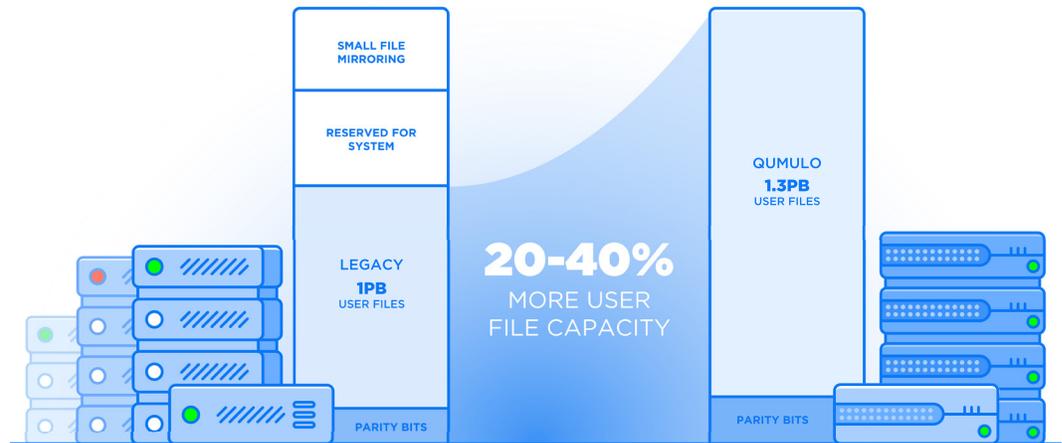
## PEAK PERFORMANCE AT 100 PERCENT UTILIZATION

The tradeoff between utilization and performance is all too familiar to storage admins. Many scale-up systems, RAID-based systems, and some of the more popular open-source file systems experience performance degradation as the file system fills. To avoid performance problems, you're supposed to stay under 70 percent of usable capacity. You shouldn't have to choose between utilization and performance—but that's the position these vendors put you in.

Unlike some other systems, QF2's performance doesn't degrade as your system fills. Instead of keeping 30 percent of your capacity in reserve, you can go ahead and use 100 percent of it, storing billions of files with no impact on performance.

## WHAT QUMULO MEANS FOR YOUR DATA

All told, the combined efficiency benefits mean a typical Qumulo customer can store the same amount of user data with 25 percent less raw capacity than other file systems.



That high efficiency is complemented by the kind of benefits that matter to data-intensive businesses:

- Fast rebuild times in case of a failed disk drive
- The ability to continue normal file operations during rebuild operations
- No performance degradation due to contention between normal file writes and rebuild writes
- Equal storage efficiency for small files and large files
- Accurate reporting of usable space
- Efficient transactions that allow QF2 clusters to scale to many hundreds of nodes
- Built-in tiering of hot/cold data that gives flash performance at archive prices.

Your data is too important to leave to outdated legacy storage methods—or to vendors with a loose grip on the concept of “usable capacity.” With QF2, Qumulo delivers the transparency, predictability, and performance you need for digital era data operations.

To learn more about Qumulo and QF2, please visit <https://qumulo.com/>.

