HGST Delivers **Rebuild Assist Mode** for Faster RAID Recovery

Contents

Introduction

Organizations choose RAID configurations to reliably protect their data, while minimizing down time and maintaining performance. Naturally, this is why businesses invest in enterprise-class HGST drives, which offer best-in-class reliability, sustained throughput, and capacity levels. However, even the world's most reliable drives can't eliminate the risk of data loss. RAID systems go a very long way toward mitigating these risks, but the inevitable RAID rebuild process that initiates as a result of a failing drive which must be replaced will eventually pose its own time, performance, and data safety challenges — challenges that HGST now addresses with new Rebuild Assist support.

The Root of RAID Rebuilds

No storage technology is immune from error. Disk and flash alike can fall prey to fault-generating conditions. HGST employs a range of on-drive diagnostics to assess possible error conditions and report to the RAID controller for further analysis and possible corrective action.

As evidenced by 2M hour MTBF ratings, HGST drives are designed to minimize error conditions. Minor errors can often be repaired or reallocated, resulting in no or minimal data loss. However, a drive within a large population may eventually exhibit significant error conditions, causing it to be identified as a failing drive. In this case, the RAID controller must take action: the failing drive will be taken offline and the array will be placed into degraded mode to reduce the possibility of additional data loss resulting from ongoing failure. The RAID controller continues to provide access to stored data while the data is reconstructed to a newly installed drive, or an available hot spare, using parity data from the remaining drives in the array through a rebuild process known as XORing.

The upside to this process is that it works. The downside is that rebuilding the contents of a drive can take a very long time. While the array is operating in a degraded condition, it must reconstruct the data and at the same time continue responding to data requests. From a performance perspective, RAID configurations are often used as shared storage, which means degraded performance caused by rebuilds can impact the productivity of many workers.

RAID rebuilds happen. RAID users know this, but not all understand the increased risk of data loss that can occur if a secondary failure happens during the rebuild process. Reducing the rebuild time reduces this risk.

Why Rebuild Times are Rising

In the days when arrays spanned from megabytes into gigabytes, RAID rebuild times were reasonable. However, the practical impact of the move into terabytes becomes quickly apparent when considering the basic formula for estimating rebuild times:

Capacity (MB) / Sustained Transfer Rate (MB/s) = Rebuild time (s)

In the case of a 3TB volume rebuild on a typical 3.5-inch hard drive averaging a 110MB/s sustained transfer rate, the rebuild would complete in 19,108 seconds (5.3 hours). This would be typical of a simple mirrored array (RAID 1). If this weren't bad enough, recall that more complex arrays operate in a "degraded mode" when one of their drives is offline. This reduces volume performance and increases rebuild time even more. Real world rebuild rates on a RAID 5 can easily plummet into the 10MB/s range, requiring rebuild times to take days – unacceptable to most data center managers who must deliver on service level agreements (SLAs) for their end customers.

As drive capacities grow, rebuild time grows and so does data risk. For example, using a rebuild time estimator for a 12-drive RAID 5 with one hot spare comprised of 3TB drives vs. 5TB drives and a generous rebuild speed of 40MB/s, the annual odds of data loss are 54% greater for the higher-capacity 5TB-drive array, primarily because the rebuild time would take about 13 hours longer*. If a second drive fails in a RAID lacking additional redundancy, the volume contents are gone, often permanently. RAID 6 is often used to help mitigate the risk of a second failure.

Bottom line, there is a direct relationship between drive capacity and rebuild time, or more specifically, the higher the capacity, the longer the rebuild time under degraded array conditions. To address this challenge, HGST supports an industry standard feature called Rebuild Assist on its newest capacity enterprise drives. Rebuild Assist not only slashes array rebuild times, but also reduces the risk of data loss by quickly returning the array to full performance.

Rebuild Assist: How it Works

In a traditional rebuild scenario (Figure 1), typically a large portion of the data can still be accessed but the system treats the entire drive as if it were unrecoverable and takes it offline. The system must now reconstruct all the data to a replacement drive using the remaining healthy drives. As noted earlier, this can take many hours to days to complete.

Rebuild Assist (Figure 2) allows the system to first copy the good data (i.e., "copy what you can") at normal operational speed to the hot spare, leaving a smaller portion of unrecoverable data to be reconstructed. This copying of good data at normal operational speed is the key principle underlying the Rebuild Assist feature now found on HGST's latest high-capacity hard drives, Ultrastar® He8 and Ultrastar 7K6000.

RAID configurations that use parity for data protection divide array volumes into a series of "stripes." When an array containing drives with Rebuild Assist technology initializes a spare to replace a failing drive, all blocks in the failing drive's good stripes get sequentially copied to the new spare. Only stripes containing bad blocks get XORed and reconstructed in the traditional way.

One might think of RAID rebuilding like conducting an archaeological dig for an artifact known to be ten feet underground. Copying is like using a big shovel; it's fast and moves a lot of data all at once. XORing is like using a pick and brushes; it's painstakingly slow and not good for moving a lot of data. Conventional rebuilding is like doing an entire dig with a pick and brushes. Rebuild Assist gives you a shovel for most of the job.

The pivotal component in Rebuild Assist mode is copying good data at speeds much faster than reconstructing. Other actions, however, also contribute to acceleration. Rebuild Assist will take the following actions:

1. Disable long error recovery procedures when reading from the drive.
2. Write-protect the drive.
3. Reduce unnecessary background tasks.

Disk size (GB)	Rebuild	RAID 5 with 1 hot spare		
		Array size (TB)	Data loss odds per year	
3000 (3TB)	20 hours, 50 minutes	30	1 in 127.9	
5000 (5TB)	34 hours	50	1 in 80.4	← 54% higher risk of data loss

*Example modeled using rebuild time calculator available at https://www.memset.com/tools/raid-calculator with a 40MB/s sustained transfer rate, 12 drives per array (including parity and hot spare), 5-year warranty and allowing 7 days for a drive to be replaced.

Figure 1. Traditional rebuild process disregards any good data on the failing drive and reconstructs the entire contents to the hot spare by using data from the remaining healthy drives.

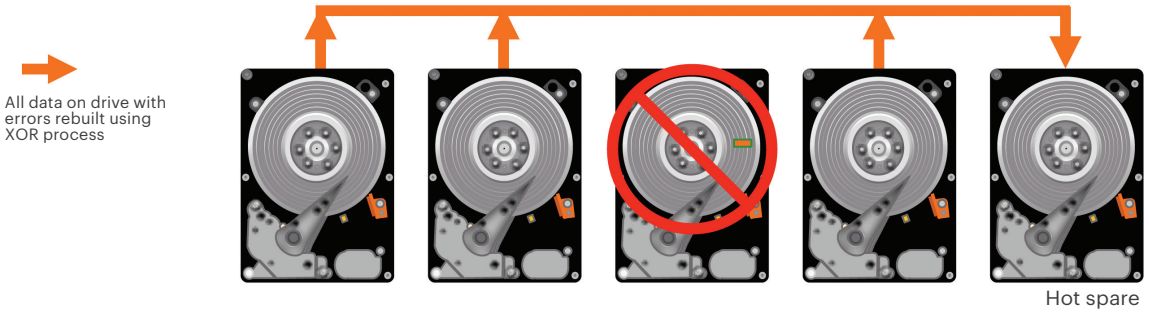
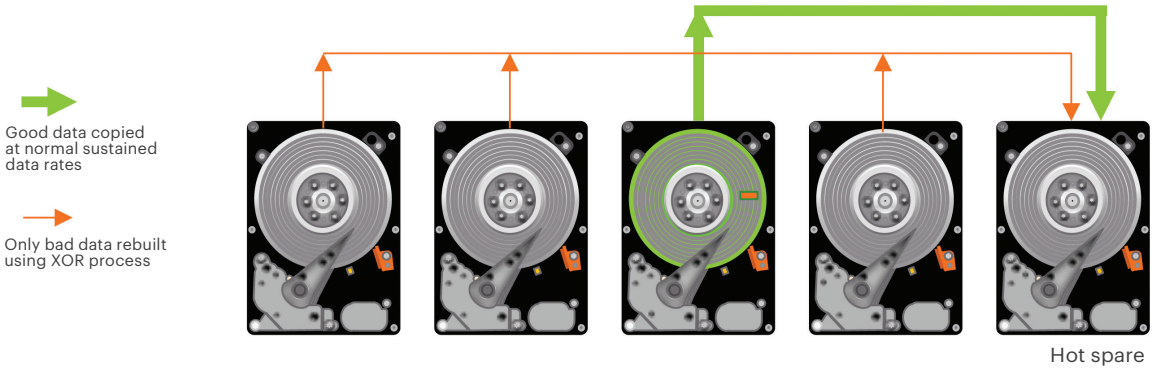


Figure 2. Rebuild Assist quickly copies the good data from the drive with errors. It then reconstructs the missing data using the remaining healthy drives, significantly reducing rebuild times.



Since the objective of a rebuild is to recreate the data onto the hot spare, the optimal approach is to copy good data from the bad drive (i.e., "copy what you can"). This also reduces the load on the remaining healthy drives to maintain their performance for ongoing use. After the good data has been copied, the missing data will be reconstructed (by XORing) from the healthy drives. When Rebuild Assist completes, performance of the array returns to normal and the drive with errors can be retired.

Overall, Rebuild Assist technology delivers faster RAID recovery and system-level performance during RAID recovery similar to the performance of an unaffected array. HGST observed cases where priority was given to rebuilding the array over maintaining performance. The array was able to sustain transfer speeds in excess of 180 MB/s with Rebuild Assist mode actively rebuilding a drive. Without Rebuild Assist, throughput cratered to a mere 3.5 MB/s.

Engineers compared the rebuild performance of healthy drives, failing drives, and failing drives with Rebuild Assist mode. They found negligible performance differences between a healthy array and a failing array operating in Rebuild Assist mode. In terms of time and IOPS performance, Rebuild Assist essentially erases the typical penalties of traditional array recovery.

RAID Rebuilding At-a-Glance

RAID rebuilding boils down to a few essential steps. Here is a quick before-and-after view of what has changed with Rebuild Assist.

Without Rebuild Assist	With Rebuild Assist
1. The hard drive self-detects errors and notifies the host.	1. The hard drive self-detects errors and notifies the host.
2. If significant error conditions are detected, the host flags the drive as unhealthy and puts the array into degraded mode.	2. If significant error conditions are detected, the host flags the drive as unhealthy and puts the array into degraded mode. The unhealthy drive does a self-test and provides a list of known bad block ranges to the host. The host can then quickly copy the mass of "good data" from the unhealthy drive to a hot spare. Any unpredicted errors that occur during the copying of good data are quickly flagged as "bad" without spending time in error recovery, thus speeding the copy process.
3. The RAID manager initiates *a lengthy XOR rebuilding process for all of the unhealthy drive's data* from the remaining healthy drives in the array. This data gets rebuilt to a spare/replacement drive.	3. The host XORs data from the healthy drives to recreate the small amount of missing data to the hot spare.
4. The spare becomes an active array member, seamlessly replacing the original unhealthy drive.	4. The spare becomes an active array member, seamlessly replacing the original unhealthy drive *in far less time than without Rebuild Assist.*

Start Investing in Rebuild Assist Now

Rebuild times have grown to unacceptable levels in recent years as drive capacities grew to terabyte dimensions. In May 2012, with 3.5-inch hard drive capacities reaching 4TB, the T10 working group approved including Rebuild Assist in the SCSI SBC-3 specification. In August 2013, the SATA-IO committee adopted TPR-045 (Rebuild Assist) as part of the SATA 3.2 specification. Today, with hard drives edging past 8TB, the need for Rebuild Assist has only grown more imperative.

Rebuild Assist is not an all-or-nothing proposition. Businesses can begin deploying it as opportunity allows, either when adding new capacity or replacing retired drives. If a RAID controller supporting Rebuild Assist finds Rebuild Assist supported on a failing drive which must be replaced, the feature will be used. If the drive does not support Rebuild Assist, the rebuild will be done in the traditional, all-XORed fashion. There are some less-common situations where Rebuild Assist will not work, such as in the case of a dead drive interface or multiple head failures, but most of the time, a drive can determine that it has errors, ignore them, and let Rebuild Assist work its magic.

The first HGST hard drives to deliver RAID rebuild assist functionality are the 8TB HelioSeal® Ultrastar He8 and the 6TB Ultrastar 7K6000, both available in 6GB/s SATA and 12Gb/s SAS models. Equipped with these enterprise-grade drives, businesses will not only enjoy the best-of-breed speed and reliability typical of HGST storage products but also the added productivity and security that comes with slashing RAID rebuild times to unprecedented lows. Rebuild Assist functionality can also be leveraged in distributed file systems that use erasure coding to significantly reduce the impact of rebuilding failed drives.

For more information about HGST enterprise storage solutions, visit http://www.hgst.com.