

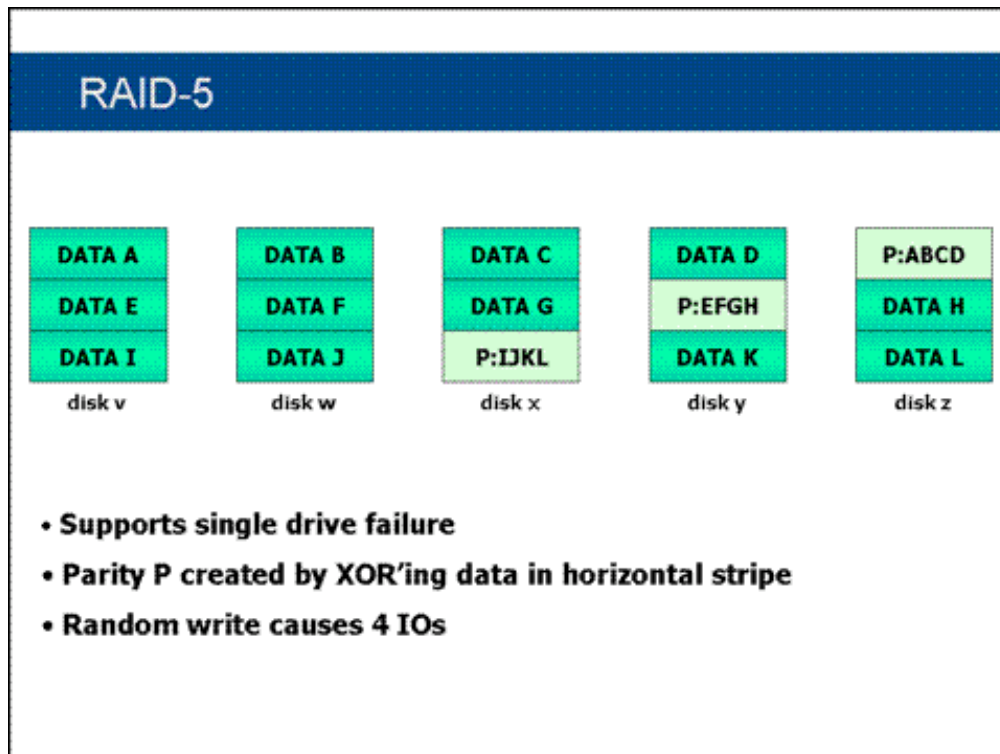
[Home](#) | [About the Storage Advisors](#) | [Adaptec Trusted Storage](#)
« [Managed offsite client backups](#)
[SAS Webinar](#) »

[A tale of multiple RAID-6s](#)

Posted in [Storage Interconnects & RAID](#), [Advisor - Tom Treadway](#) by Tom Treadway

There have been several posts recently regarding the reliability of RAID-5 and RAID-6, and how each applies to SATA and SAS drives. Now that your head is sufficiently spinning it's probably worth going back and explaining how RAID-6 is defined. Unlike the other RAID levels, RAID-6 is very vendor unique.

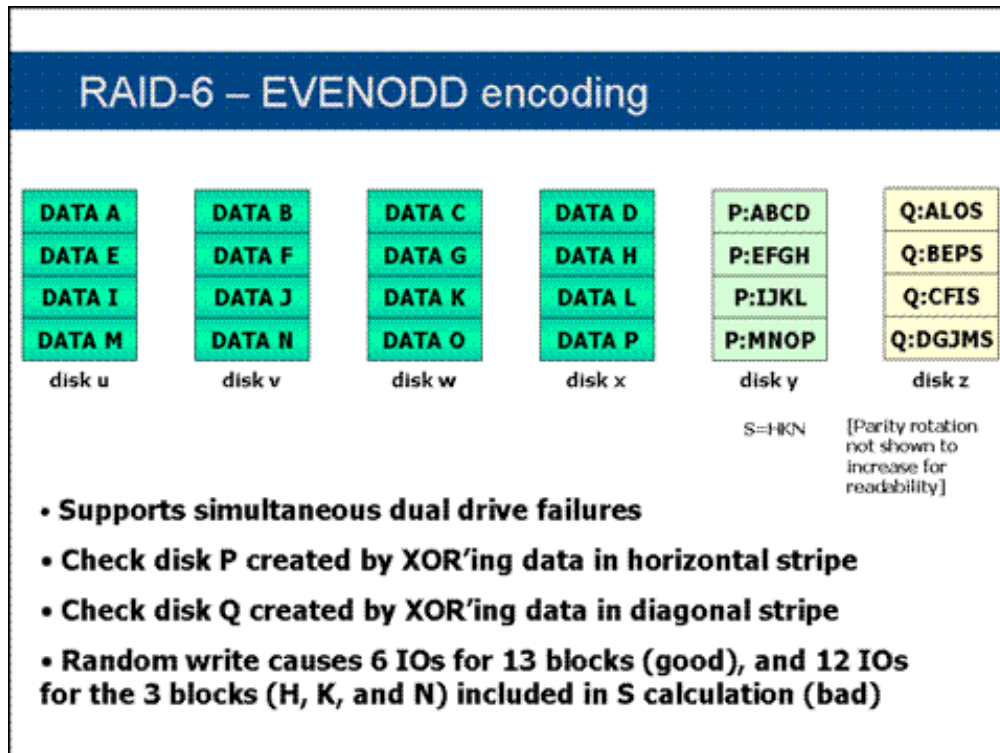
First, let's start with a diagram of how RAID-5 is laid out. With this baseline, the RAID-6 diagrams will make more sense.



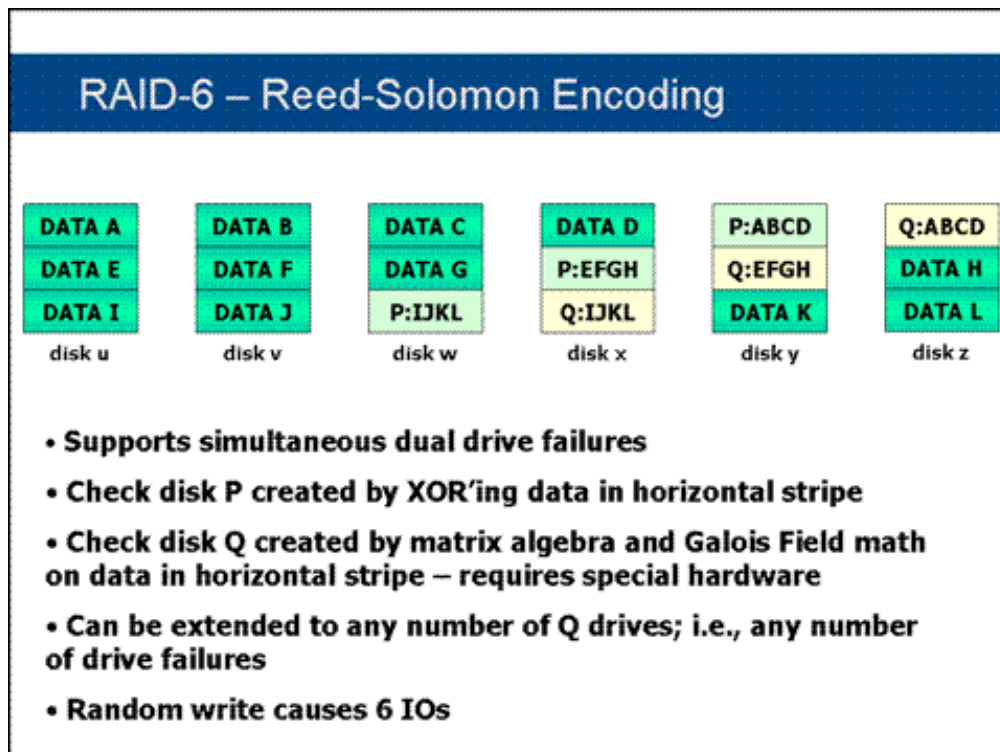
As you can see, the parity is built by XOR'ing the data stripes on the same horizontal row. Note that RAID-5 can have either right-to-left (as shown) or left-to-right slanting parity. Also, there are various methods for the data layout from one horizontal row to the next. Regardless of these details all RAID-5's have the same basic reliability and performance characteristics.

One of the first commercially available versions of RAID-6, created by IBM Almaden, is called EvenOdd. There is rumor that a similar scheme is used by NetApp. The advantage of this scheme is that it's based on XOR. The disadvantage is that it has a few hot spots in certain diagonal blocks that cause very poor short write

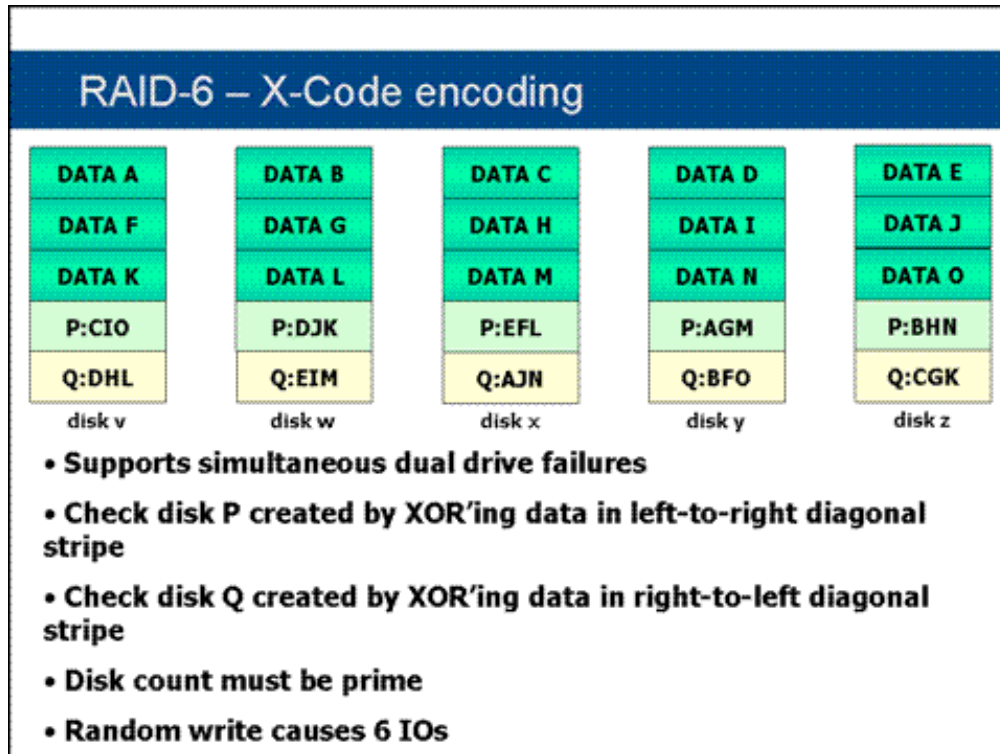
performance.



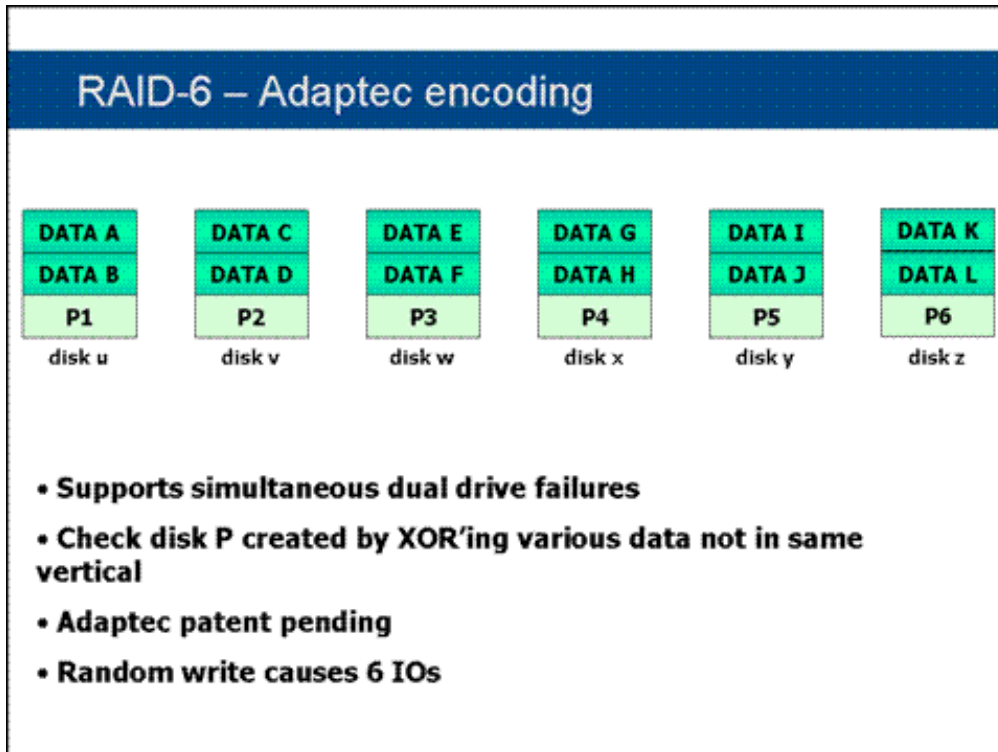
A more common version of RAID-6 is based on Reed-Solomon encoding. The math is rather complex and requires a Galois Field lookup which, due to performance issues, isn't feasible for a typical RAID IOP. Therefore R-S requires specialized hardware. Also, this encoding scheme is rumored to be similar to the ADG scheme used by HP.



The following RAID-6 scheme called X-Code is interesting in that (a) it's based on XOR, and (b) it interleaves rows of parity within rows of data. Unfortunately this scheme only works on a prime number of drives.



The last scheme is an Adaptec proprietary scheme similar to X-Code because it uses XOR and has row parity, but it works on any drive count. Patents are still pending, so the actual relationship of the data to the parity is not shown.



The advantage of this scheme is that it does not require specialized hardware and is backwards compatible to older hardware. And the performance characteristics are very good.

So as you can see, all RAID-6 is not alike. There are pros and cons to each, but they all support two simultaneous drive failures.

TT

This entry was posted on Monday, November 7th, 2005 and is filed under [Storage Interconnects & RAID](#), [Advisor - Tom Treadway](#). You can follow any responses to this entry through the [RSS 2.0](#) feed. You can skip to the end and leave a response. Pinging is currently not allowed.

5 Responses to “A tale of multiple RAID-6s”

1. [Jon Toigo](#) Says:
[November 14th, 2005 at 7:40 am](#)

Great post. I am cross-linking from DrunkenData.com.

2. [Jon Toigo](#) Says:
[November 14th, 2005 at 7:53 am](#)

Tom,

What are the practical impacts of each RAID 6 variant? Does application I/O run faster or slower over one or another? Does RAID set restore happen faster or slower over one or another?

Please advise.

3. [Tom](#) Says:
[November 16th, 2005 at 5:25 am](#)

Jon,

In general, if implemented correctly, you're going to see roughly the same performance in all the versions of RAID-6. I'd like to say that Adaptec's version is faster, but I can't. We've been able to maintain 90% of the long sequential write performance in going from RAID-5 to RAID-6, but that's more due to internal optimizations than due to the algorithm itself. The advantage of our algorithm is that it works on all legacy hardware with a simple XOR engine.

On short random writes all the algorithms should give a 50% hit in performance, except for EvenOdd if you hit blocks on the diagonal which will cause a 80-90% hit in performance. Stay away from EvenOdd.

Reads should be identical between all the algorithms, unless someone screwed something up in the implementation. The performance of RAID-6 reads should equal that of RAID-5.

Lastly, don't even think about running Reed-Solomon in software RAID, such as Linux's DMRAID. The data conversion involved in the Galois Field math has tremendous overhead that can bring a powerful x86 to its knees. If you run performance numbers you might think, hmm, this is pretty dang good. But if check the CPU utilization you'll find that it's pegged at 100%, leaving nothing for your applications.

Regarding the restore (rebuild) time - good question. But that's a little trickier since it heavily depends on the actual implementation. For example, everyone pretty much does RAID-5 the same way, but you'll notice huge differences in restore time. On paper, the RAID-6 varieties should be pretty similar however I expect to see big differences when we finally get a chance to run all the competitive products.

I'll try to post some real life data as it becomes available.

TT

4. [Joe Fagan](#) Says:
[December 2nd, 2005 at 8:37 pm](#)

Tom,

Nice RAID-6 discussion. I can't follow the argument that the need for hardware assist for Reed-Solomon encoding is a disadvantage - On the basis that it's more flexible (protects m drive failures for m 'parity' drives), more efficient than some alternatives when n is odd, and works at a stripe level, why don't the RAID vendors recognise it as an opportunity to differentiate themselves!

RAID-6 is not just a "fad-du-jour" - it will be around forever! It can't be improved upon (space-efficiency wise) - so why not start making some serious hardware and forget backward compatibility. Anyway, compatibility is needed only if you're trying to support multiple raid cores or codes. Pick the best one and stick some hardware down.

Besides, once polynomial arithmetic is down it opens the possibility for maybe hardware assisted encryption and/or compression and lots more.

Hardware I say - Hardware!!!

Joe

5. [Tom](#) Says:
[December 3rd, 2005 at 6:16 am](#)

Joe,

The reason I say that hardware assist is a disadvantage for R-S is that most chips don't have it yet. I'm contrasting this to XOR-based schemes that can effectively do RAID-6 with practically any hardware produced in the last 10 years. Of course this is only a temporary disadvantage as R-S becomes more and more common. But we're not there yet!

And I do certainly agree that R-S is more flexible for supporting >2 drives. I think we're still pretty far away from needing that, but I've often imagined a box full (dozens and dozens) of 2.5" or maybe 1" drives delivering an insane number of IOPS. 😊 Tolerating more than 2 drive failures would be a necessary feature. Of course you would need a lot of cache to try and hide that nasty RAID-6 RMW overhead on random writes.

Viva la RAID-6!

TT

Leave a Reply

Name (required)

Mail (will not be published) (required)

Website

By submitting a comment, you understand that if your post violates any federal, state or local law, or contains libel, slander, defamation, product disparagement, harassment, obscenity, or indecency, it will be edited or deleted.

Comments on this blog are moderated. Editorial privilege on comments may be exercised to protect the identity and security of the commenter, blogger, company, employees, partners, customers and competitors.