

From: Derek Asirvadem <derek.asirvadem@gmail.com>
Newsgroups: sybase.public.ase.administration
Subject: Re: RAID5 vs RAID1+0
Date: 24 Feb 2009 19:52:48 -0800

Substantiation

Here are some technical articles which support my statements. As stated from the outset, the problem with this subject is there is a lot of misinformation, which is not challenged, and that keeps the misinformation in place. Since this post provides evidence, and evidence is not short, one must be willing to see each argument to the end.

This article is dedicated to the Destruction of Ignorance, as a service to the community. Those who fit into the category of "Vociferous Ignorance" should read no further:

<http://www.dbdebunk.com/page/page/3161496.htm>

(In the first item, for the words "data fundamentals in general" substitute "technology in general"; and for "relational concepts, principles and methods in particular" substitute "the laws of physics in particular" .)

RAID Overviews & Benchmarks

There are scores of good articles. And then there are many people working in technical jobs with no technical qualifications or experience, blissfully ignorant of all the facts, still "debating" a small subset of the technical facts, ten years later: there are hundreds of articles from these types. I am sure you can figure out the difference. Hence: "a little knowledge [some facts] is a dangerous thing [ignorance of the other facts]". It is particularly dangerous because the person has understandable confidence, and is unaware of the facts that are going to kill him. Read then, with a sense of holding clearly in your mind the technical facts of which you are already convinced: accept each presented technical fact only if it integrates with your facts; XOR reject each fact that stands in contradiction. In the end, at least you will be more educated about the subject, and can thus provide recommendations from an educated position [rather than quoting false gods]; at best you will k.n.o.w the difference between conclusions derived from all the evidenced facts versus conclusions based on a subset of the facts.

Here's a selection. A.n.y DBA (Oracle, DB2, Sybase) would baarf at RAID5 (and all derivatives of RAID5) for database use. Contrary to some posts, the issue was officially closed on 4 Jun 03.

This link sets the mood for the uninitiated:

<http://www.baarf.com/>

This is probably the oldest (1999 I believe) and most famous:

http://www.miracleas.com/BAARF/RAID5_versus_RAID10.txt

Another often-quoted article (03 Jan 00):

<http://www.miracleas.com/BAARF/1.Millsap2000.01.03-RAID5.pdf>

This link provides a typical benchmark:

[Performance.aspx](http://weblogs.sqlteam.com/billg/archive/2007/06/18/RAID-10-vs.-RAID-5-Performance.aspx)

The

diagrams in the links are not very good, so I have attached a simple overview which can be used for discussion here.

Recoverability

RAID1 or 1+0

There is no substitute, or alternative, that is comparable to a 100% mirrored active drive. If a drive breaks the replacement is already there, with up-to-date data sitting in the correct blocks; there is no work to do; there is no change in performance. The system continues unhindered. When the drive is replaced: configuration and enabling are instantaneous (we know where the blocks will go; we know what the blocks will contain; there is no compute time); bringing it up to speed is effortless and short.

This can sustain the loss of HALF the drives: for 16 RAID1 mirrored pairs, an outage of up to 16 disks (but not both disks in any single pair).

RAID5

This can sustain the loss of ONE drive, period. A second outage will kill the system.

I could not find any articles that suggest RAID5 is better than RAID1 or 1+0 in the recoverability department. "Until the array is repaired, every I/O call upon the failed disk will require a read or write from every surviving disk in the array. The performance degradation will last until after the failed disk has been re-placed and the replacement disk is resynchronized. The resynchronization event is extremely I/O intensive, requiring a read of every block from every disk in the array. Of course, the resynchronization reads compete for I/O service with the normal I/Os already being generated by your application."

Performance

Note that operational databases read and write random small blocks, not large sequential blocks. Disks have the only moving parts in a computer system, and they fade and die. Except by the power of magic, for any given hardware configuration, it is impossible by physical means to perform the work of RAID5:

- every write writes to [at least] one second block which must be located; the parity block must be read, re-computed, re-written

faster than RAID1+0:

- every write writes to two blocks asynchronously (no compute time; no find-the-parity-block-and-write)

- every read reads from one of the mirrored pair, which has its heads closer to the required track

(no check time)

This doesn't even get into saturation of the cache, etc.

RAID5 certainly has improved in ten years, but that is an improvement within its class (unchanged recoverability, slightly improved performance over some other RAID5 configuration), it is nowhere near the recoverability or performance of the RAID1 class.

Configuration

One important fact as evidenced in the field, is RAID1 and 1+0 is simple to implement and configure. RAID5 is difficult to configure, hence most people do not obtain the promised performance from it.

Second, once performance is recognised as a problem, RAID5 configurations are difficult to change, either to identify and correct the performance issues or just to add more resources. The corporation is left with the only option (same as poorly designed databases): sink more money into it by ADDING more disks, without changing the config that is in place. That's why you are better off to have started with 100% mirrors and flat, unaffected performance, with a config that is easy to change.

The disks can be split horizontally and used for other purposes, thus nothing is lost, but that is an advanced subject which I will not go into here. However, that can only be contemplated in RAID1 or 1+0; it is unthinkable in RAID5.

Business

I have already posted (go directly to Post 12):

http://groups.google.com.au/group/sybase.public.ase.administration/browse_frm/thread/f62e08645a76aea9?q=derek+asirvadem+business+clerk

My

take on it as an IT Professional is, it is my job to communicate ALL (not a subset) the pros and cons of the RAID options to them; they make the decision and live with the pros and cons of the option they chose. Otherwise I am not doing the job they hired me to do; I am a highly qualified clerk. Just try explaining why it has been days since you replaced the broken disk drive, but you still have not been able to get the RAID5 array back to operating status, and every user is complaining about performance.

The common reason for the "business" choosing RAID5 over RAID1+0 is the initial cost appears to be cheap (they do not want to "lose half their disk space"). It is not "losing half", it is gaining a mirrored whole. The information needs to be presented honestly, citing RAID5 maintenance and reconfiguration costs.

Clarity

If, after reading these articles, you are anything less than 100% certain about:

- RAID1+0 being far superior to RAID5 for databases, in both the performance and recoverability categories, OR

- any of the technical facts [of the many] which obviously support my statements (original post)

OR

please post a specific question.

--

Cheers

Derek

Senior Sybase DBA / Information Architect

Copyright © 2008 Software Gems Pty Ltd

--

Quality Standards = Zero Maintenance + Zero Surprises

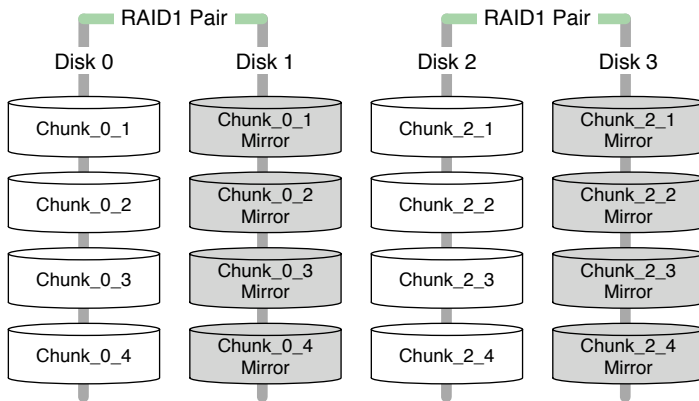
Performance Standards = Predictability + Scaleability

RAID Comparison

RAID Comparison

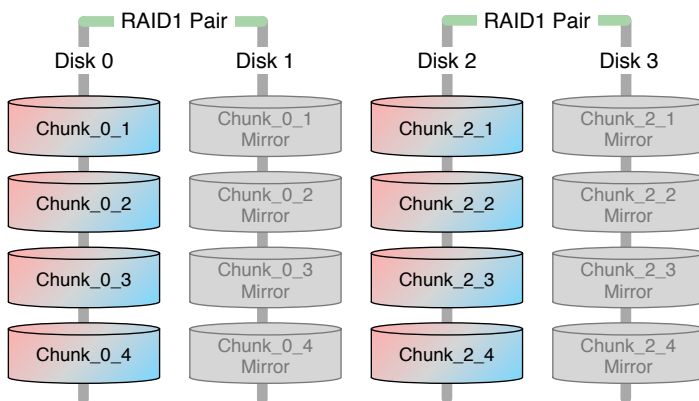
This depicts the two RAID configurations, on the same set of 4 disks, for comparison purposes. Refer to the accompanying article for details. The disks have been divided into Subdisks or 'chunks', as shown.

RAID1: Mirrored



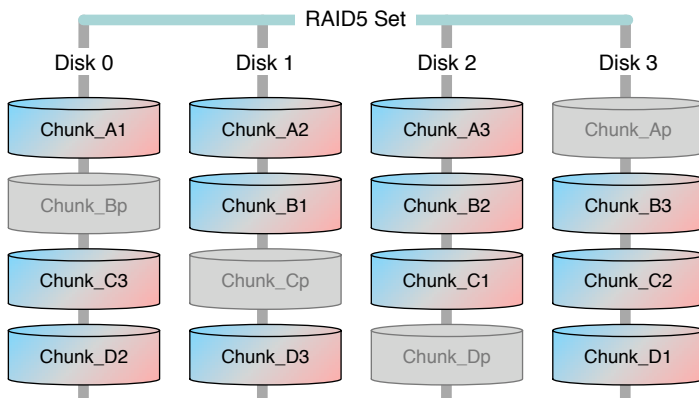
Disk = 143gb x 4
Mirrored = 143gb x 2 = 286gb

RAID1+0: Mirrored, then Striped



Disk = 143gb x 4
Mirrored = 143gb x 2 = 286gb

RAID5: Striped



Disk = 143gb x 4
Avail = 143gb x 3 = 429gb