Answers (/community/s/group/0F90L0000001NDBSA2/) — jerdavis__ (/community/s/profile/0050L0000093AaAQAU) (Customer) asked a question.
August 11, 2014 at 11:33 PM (/community/s/question/0D50L00006BIuAJSA1/raid0-vs-jbod-vs-write-cache-etc-again)

RAID0 vs JBOD vs Write Cache etc (again)

In a follow up to this question:

http://answers.mapr.com/questions/4563/nitty-gritty-single-drive-per-volume-raid-0-vs-jbod (http://answers.mapr.com/questions/4563/nitty-gritty-single-drive-per-volume-raid-0-vs-jbod) (https://community.mapr.com/external-link.jspa?url=http%3A%2F%2Fanswers.mapr.com%2Fquestions%2F4563%2Fnitty-gritty-single-drive-per-volume-raid-0-vs-jbod (https://community.mapr.com/external-link.jspa?url=http%3A%2F%2Fanswers.mapr.com%2Fquestions%2F4563%2Fnitty-gritty-single-drive-per-volume-raid-0-vs-jbod))

I'm evaluating my next hardware configuration and I'm looking for advice on what to use for a HBA.

Specifically the link mentions the PERC H710/LSI card, and having to use a RAID0 for each drive, vs JBOD.

The PERC 710 doesn't support JBOD, but does have a write cache. The 310 does JBOD, but doesn't have a write cache.

Seems like a write cache is a good thing. Trying to answer:

Is the 710 RAID0 a real bummer (vs JBOD)?

Is having a write cache a big deal for typical Hadoop loads?

So much so that I should look at other HBA cards that don't have these problems?

Upvote Answer Share 4 answers 55 views

Top Rated Answers

MC Srivas (/community/s/profile/0050L0000093EJHQA2) (Customer)
4 years ago

Why do you think the RAID-card write-cache is a good thing? The "host" has a lot more memory than the measely 1G or 2G a typical RAID-card has. For example, if your box has 64G of DRAM, and you are running M3 or M5, MapR will use about 12.8G of memory for its cache. (with M7, it will be 22G). That is an order of magnitude more than the RAID card can cache, so when MapR actually does a issue write to the disk, the random writes have already been well gathered together at the MapR level. MapR will issue 1M or 512K writes per device, so the RAID-card's cheaper processors which transfer data back/forth to its on-board memory usually get flooded and overwhelmed.

That's why we recommend running the card in JBOD pass-through mode, so that it does very little and can keep up.

On the flip side, running the write cache messes with recovery and fault-tolerance, since the write-cache ends up falsely returning a success on a write when it hasn't yet written it out to disk.

Net, it doesn't help in recovery, and interferes with performance.

Selected as Best Upvote

All Answers

MC Srivas (/community/s/profile/0050L0000093EJHQA2) (Customer)
4 years ago

Why do you think the RAID-card write-cache is a good thing? The "host" has a lot more memory than the measely 1G or 2G a typical RAID-card has. For example, if your box has 64G of DRAM, and you are running M3 or M5, MapR will use about 12.8G of memory for its cache. (with M7, it will be 22G). That is an order of magnitude more than the RAID card can cache, so when MapR actually does a issue write to the disk, the random writes have already been well gathered together at the MapR level. MapR will issue 1M or 512K writes per device, so the RAID-card's cheaper processors which transfer data back/forth to its on-board memory usually get flooded and overwhelmed.

That's why we recommend running the card in JBOD pass-through mode, so that it does very little and can keep up.

On the flip side, running the write cache messes with recovery and fault-tolerance, since the write-cache ends up falsely returning a success on a write when it hasn't yet written it out to disk.

Net, it doesn't help in recovery, and interferes with performance.

Selected as Best Upvote Reply

jerdavis__ (/community/s/profile/0050L0000093AaAQAU) (Customer)
4 years ago

Thank You!

Upvote Reply

MC Srivas (/community/s/profile/0050L0000093EJHQA2) (Customer)
4 years ago

Yes. When you run "ps -ef | grep mfs" the "-m" argument shows how many megabytes MFS was told to use by warden

Upvote Reply

jerdavis_ (/community/s/profile/0050L0000093AaAQAU) (Customer)
4 years ago

Also the memory that MapR uses for it's cache: Is this the memory I see used by 'mfs' ?

Upvote Reply

Login to answer this question