Thursday, January 9. 2014

Configuration hint when using F40/F80 cards

I'm using the Sun Flash Accelerator F40/F80 PCIe Card quite frequently (as i've used the F5100 and F20 in the past) in the recent time for POCs. In my experience to get best performance you really should abide the tips on this page.

When you put F40 cards into your systems, add this line into /kernel/drv/sd.conf.sd-config-list = "ATA 3E128-TS2-550B01","disksort:false, cache-nonvolatile:true, physical-block-size:4096";The F80 card works best with this line in /kernel/drv/sd.confsd-config-list = "ATA   2E256-TU2-510B00","disksort:false, cache-nonvolatile:true, physical-block-size:8192";When you use both cards you have concatenate both lines into a single one.
sd-config-list="ATA   2E256-TU2-510B00","disksort:false, cache-nonvolatile:true, physical-block-size:8192",\
"ATA   3E128-TS2-550B01","disksort:false, cache-nonvolatile:true, physical-block-size:4096";The effects of this lines are quite simple, but very important. I'm simplifying the reasoning in the next paragraphs a little bit:

disksort:false - This part deactivates the disk sort algorithm. The disk sort algorithm tries to resort disk accesses to minimise head seeks. However, with a disk without disk heads the use case is obviously somewhat limited and may even have an adverse effect (when you have read requests for block 100, 150, 200, 180, then the execution is like 100,150,180,200. When an access for block 110 is requested before the request for 15 has been send to disk, the last new requests is actually executed before the older ones are send to disk ... normally a good thing as you you can assume that you don't have to move the head that much to get block 110 when you just was at block 100 and so get the block en-passant, however the requests for blocks 150,180 and 200 take longer than you would expect as the block 110 requests jumped the queue) that isn't offset by the positive effects of sorting the requests that minimise head movements.

cache-nonvolatile:true - This part tells the system, that the device has a non-volatile cache. When you do not set this ZFS thinks it has to flush caches of the disk quite often to ensure that the data is really on disk and not just in the cache soldered to the disk. However the flash cards doesn't need this because it's ensured that the data written to the disks are persistent after a reboot. The additional unnecessary flushes cost performance ... so we can just deactivate them. As soon ZFS detects that the devices are flagged nonvolatile, ZFS automatically stops the flushes of the disk cache. You can gather some additional information at >docs.oracle.com

physical-block-size:4096 respectively 8192 - The f40 cards are optimised for accesses with a block size of 4096 bytes, the f80 card is optimised for 8192 bytes. By setting this parameter, you override the data reported out of compatibility reasons by the disks and set a different block size. The system then assumes that the device is a 4k or 8k disk and aligns it's accesses accordingly

Posted by Joerg Moellenkamp in English, Solaris at 22:28

POCs?
Something like OOCA?

Greetings from Berlin

Sven
Anonymous on Jan 10 2014, 22:53

Proof of concept
Anonymous on Jan 10 2014, 23:31

Hello Joerg,
great informations, thank you!

How can I verify that the settings are correctly in place and working?

Do you think it makes sense to make use of these settings when using SSD drives as L2ARC?

Or another usecase: Harddisks and SSDs in an X4170 M2 server on a hardware LSI RAID controller with 512MB BBWC:
We have this setup in the datacenter, 4 harddisks and 3 SSD's on that LSI controller in RAID0 mode and then in ZFS pool.

Best regards,

Bernhard

Anonymous on Jan 12 2014, 13:27

Hello,

Is a similar setting for the F20 available?
'cause the F20 is the only FA card certified for the M9000

Regards
Olaf

Anonymous on Jan 20 2014, 23:11

Hello,

Is a similar setting for the F20 available?
'cause the F20 is the only FA card certified for the M9000

Regards
Olaf

Anonymous on Jan 20 2014, 23:12

The correct line for F20 and F5100 is

```
sd-config-list="ATA MARVELL SD88SA02","throttle-max:32, disksort:false, cache-nonvolatile:true, physical-block-size:4096";
```

Anonymous on Jan 21 2014, 05:27

I have some t4-4 here and I am sad that those cards doesn't fit in it...

Anonymous on Mar 8 2014, 22:23