

Hitachi Dynamic Tiering WebTech Series- Session 2 of 3– July 20, 2011

- 1. Can we use SSD at a 7+1 configuration and SAS at 6+2?**
Yes. Some of my comments in the presentation require a few subnotes. Raid 6+2 is now our general recommendation for all storage, except SSD (where the size and rebuild speed makes the recommendation less relevant).
- 2. Is it recommended to NOT create a pool across 2 different external arrays?**
I would not recommend it for maximum performance of a production pool receiving heavy I/O. For a temporary, archive or development area then it may be appropriate. The key to successful pool design is balance. This is because pools and tiers tend to slow down to the slowest device. It will be very hard to build 2 external devices with identical performance (but not impossible).
- 3. What should we do with the remaining capacity after 2.99TB for 600GB Raid-5 raid groups?**
As mentioned in the presentation, place **all** the capacity into the pool or tier. Never use it for another purpose. The size difference is not a problem for HDP or HDT. However, modern microcodes now permit splitting into equal size LDEVs which is “neater”.
- 4. External Tier 6: So if you have an external storage externalized behind VSP – external LUNs – it automatically gets the lowest tier?**
Absolutely. There is no way to “tell” the VSP it is RAID10 SSD or RAID6 SATA. The latter is assumed to be what most users will want. Other methods are being considered.
- 5. What is the rule of thumb for sizing the CLPR?**
Size the CLPR for the whole pool, not a tier. Separate CLPR per tier do not isolate the load of slower tiers. Use a CLPR for the DP-VOLS. Sizing for HDT is little different to sizing HDP or static volumes. Your sales team would be best to make such a recommendation. This should take account of workload and capacity. I have seen 1 gigabyte per usable terabyte used as a rule of thumb.
- 6. So you do not recommend striping with HDP?**
It is not nearly as definitive as that. Single unstriped LUNs often give performance bottlenecks associated with command tag queuing, internal OS queuing and latency effects. Huge LUNs are best avoided.

Moderate disk striping permits a workload to be distributed across multiple LUNs. With many OS and applications, having multiple LUNs means that multiple transactions can be processed in parallel. This should increase total throughput (although each individual transaction could be a little slower). But there can be a downside: with small stripe units, a single block of data could be split across multiple LUNs, ruining locality. This is most seen on random data. That said, sequential loads benefit from small stripe units and several LUNs.

Over striping over many (I've seen hundreds) of LUNs is very likely to lead to problems. Just managing the Q-depth budget will be a challenge. So, my striping recommendation is: moderate number of LUNs, fine striping on sequential, coarse striping on random. Ideally separate random and sequential loads onto different LUNs.

Although not directly related to striping, virtualization presents a similar issue with potentially multiple users of a single LUN. Again, moderation is the key. If you place several hosts running heavy SQL Server loads on one LUN, expect issues.

7. Does VSP support virtualization of AMS 2500 with 15K 300GB disks?

Yes, but it will be the lowest tier. If combined in a pool with an internal SATA tier the performance will be “odd”. It will still work but will avoid the 15K devices. Effectively Tiers 2 and 3 will work as if they were SATA speed, wasting the expensive 15K devices. This is not recommended.

8. Can you elaborate more on using Oracle ASM with HDP? Is it recommended?

It is recommended. Look at the section: one DP-VOL or many. The ideal is “a few” devices, not a huge number, not one huge LUN. Coarse striping of tablespaces and fine striping of redo is recommended – this is the Oracle ASM default. We recommend separate DATA and REDO diskgroups. (Often: BACKUP and TEMP too)

I understand the caution. When I first considered ASM and other software striping systems I thought the “load spreading” aspect of ASM (VZFS VxVM) would defeat the HDT algorithm. But on analysis this is wrong. An ASM diskgroup containing multiple disks still has lots of different data on it. This data will still exhibit locality: the hot table, hot index, redo will probably be hot (and compact: ideal for HDT).

9. How are disks added? What is the minimum number of disks in a Parity Group?

Like HDP on the VSP or USP V, you add parity groups, not disks (like AMS). These should all be the same configuration. Typically a parity group has 4 or 8 disks in it.

10. Are the tools you mentioned part of a professional services engagement or can we use them ourselves?

The tools are internal use only. They are available through a GSS engagement or for limited requests through your customer technical consultant.

11. Is there a way to integrate external storage in an HDT pool which was created before V02 microcode?

I believe a 2 tier V01 pool can have an external tier added. There are differences between V01 and V02 format pools (q.v. mixable) and they cannot be upgraded. It does not affect this issue.

12. Is there a LUN adding sequence of large parity group LUNs into a pool?

No, any order.

13. Did he just say that HDT does NOT support RAID10?

Yes, I did. Only Raid 5 and Raid 6 are available.

14. Do we have tools that will reveal the use lines you just showed?

Out of context, I’m not sure what lines you refer to. It is most likely the slide about planning: where I showed various loads: all tiers operating inside specification, then various tiers seeing overloads. There are 2 reasons to consider this analysis: planning and design (what the slide was about) and operational monitor andreplan:

For planning, the internal HDT planning tool uses this analysis and a report shows the lines. Remember for planning purposes we only care if the lines are inside or outside the box. For real-time monitor, then no, that exact display is not immediately available but the general data is. The lines aren’t there directly because load is shifted to prevent overload. In fact it is done before overload to prevent problems at peak times. (An automatic tuning system does just that). If you go back to the first presentation and look at the tier properties graphs and data panel (both are needed), you can see the same information (but upside down and logarithmic). In this case however it is “where you are now”.

15. What happens to an HDT pool containing internal and external disks when external disk array become inaccessible?

The whole pool will blockade and servers will lose access. A pool is as reliable as its weakest component: hence my strong emphasis on external reliability factors.