An Introduction to I/O and Storage Tuning

Randy Kreiser

Senior MTS Consulting Engineer, SGI

An Introduction to I/O and Storage Tuning

- 40/30/30 Performance Rule
  - 40% Hardware Setup
  - 30% System Software Setup
  - 30% Application Software
- Analyze the Application
  - Large/Small I/O’s
  - Sequential/Random I/O’s
  - Concurrent # of I/O’s
  - Percent of I/O mix of Reads/Writes
  - Type of I/O (direct, buffered, raw)

An Introduction to I/O and Storage Tuning

- Proper Application Analysis Yields the following:
  - Transaction I/O size and I/O type
  - Raid level to use (raid 5, raid 3, raid 1)
  - Number of disks to use in a raid lun (4+1 versus 8+1, mirror, etc.)
  - Write caching or Write buffering
  - Cache page size (transaction I/O)
  - Capacity requirements dictates:
    - Size of volume (# raid luns)
  - Performance requirements dictates:
    - Size of volume (# disks/raid luns)

An Introduction to I/O and Storage Tuning

- Proper Application Analysis Yields the following:
  - Mix of I/O, read versus writes
  - Number of concurrent I/O's
  - Network based I/O or local I/O
  - Raw, direct or buffered filesystem I/O

All of the above items plus the previous page items should dictate the raid level to be used, plus the size of the raid luns and write caching parameters used!

An Introduction to I/O and Storage Tuning

- Transaction size dictates raid level
  - Small I/O’s are classic to raid 5, raid 1 or raid 1/0 luns
  - Large I/O’s are classic to raid 3

- What is small/large I/O size
  - Small I/O < 32K
  - Large I/O > 256K
  - Grey area >= 32K and <= 256K
    - Application dependent

An Introduction to I/O and Storage Tuning

- I/O characteristics and effects:
  - The number of concurrent I/O's could impact the RAID level

- How many I/O's is too many?
  - Raid 3 is good for sequential I/O and probably not more than several concurrent I/O's. Four to eight I/O's depending on the size and layout of the volume/filesystem.

An Introduction to I/O and Storage Tuning

- I/O characteristics and effects:
  - Sequential versus random I/O
  - Raw, direct buffered filesystem I/O

- What type of I/O best for which raid level:
  - Large sequential I/O is best suited for direct I/O with a filesystem. You can achieve near raw performance and gain the benefits of having a filesystem.

An Introduction to I/O and Storage Tuning

- What type of I/O is best for which raid level:
  - Raw I/O is best suited for small transaction based I/O applications such as databases. Buffered filesystem I/O could be used here but better performance is generally found using raw I/O.
  - Random I/O is usually found to be better on RAID 5, RAID 1 and RAID 1/0.
  - High percentage read based applications generally are served better from RAID 5, RAID 1 or RAID 1/0.
  - Typically 70% better read based applications can benefit from RAID 5, RAID 1 or RAID 1/0.

An Introduction to I/O and Storage Tuning

- Other characteristics of RAID:
  - Transaction based systems require spindles to provide the I/O's to the application. Clearly the number of drives will determine the performance. Fibre allows more drives per loop thus providing more I/O's per loop/bus compared to SCSI.
  - Upto several thousand I/O's are possible per loop.
  - When partitioning raid never create a filesystem log on the same raid lun. This can cause thrashing of the disks.
  - Creating a single partition encompassing all usable space is recommended.
  - Creating more than 3 concurrently active partitions will also cause disk thrashing thus degrading performance.

An Introduction to I/O and Storage Tuning

- Other characteristics of RAID:
  - Enabling command tag queuing and setting an appropriate queue depth is very important when using multi-threaded I/O or multiple I/O's to the same filesystem.
  - CTQ depth calculation could be different between different versions of UNIX. The calculation for IRIX is as follows:
    - CTQ depth = 256 <max. queue depth supported by IRIX) / Total # luns owned by the RAID device (redundant storage processors).
  - When creating striped volumes select the appropriate stripe unit size and use lun interleaving. Always select an even stripe width I/O when possible.

An Introduction to I/O and Storage Tuning

- Other characteristics of RAID (continued):
  - Even stripe width I/O calculation:
    - Application I/O = (# of luns * stripe unit).
    - If the stripe group (# of luns in stripe) = 4 and the stripe unit is 2048 blocks. The application I/O size = 4MB.
  - Use a preallocated stripe width I/O when creating the file for better performance (if the version of UNIX your using supports this feature).

An Introduction to I/O and Storage Tuning

- Two ways for scaling I/O.
  - Use even stripe widths from the application program.
  - Use more threads of I/O from the application program.
  - Threads could be through the use of posix threads in the application program or running of more application program processes.

Bottom line creating more sustained I/O will give the best overall I/O results!

Randy Kreiser

rkreiser@sgi.com

301-572-8926