AWS 95: Performance & Fault Tolerance Drives (5/93)

Article Created: 24 May 1993

TOPIC ---

The Apple Workgroup Server (AWS) 95 will initially be available with internal hard disk drive capacities 230 MB, 500 MB, and 1000 MB. For customers especially concerned with performance and fault tolerance there will be several solutions available from third parties implementing various levels of RAID technology.

DISCUSSION ---

RAID

Since servers need very fast, very large, and very reliable or fault tolerant disk subsystems the AWS 95 will support a variety of RAID devices. RAID stands for Redundant Array of Inexpensive Disks. RAID devices provide a level of disk mirroring or data redundancy, along with higher performance than conventional disk drives, or SLEDs (Single Large Expensive Disk). RAID devices come in several configurations, numbered 0 through 5. The number indicates the level of striping, or distribution of data across multiple disks.

Below is a brief description of RAID configurations:

- RAID 0 describes a single disk that uses striping redundancy techniques, but is still a single disk.

- RAID 1 describes simple disk mirroring. Disk mirroring means having two separate disk drive mechanisms that appear to the system as a single volume, but all data is written to both disks at the same time. Therefore there is always a backup. In the event that one disk fails, the other disk will contain all of the data and the system can continue to run until the failed disk is replaced and rebuilt. Some implementations allow the CPU to read from the disks in parallel and therefore reduce seek time, enhancing overall performance.

- RAID 2 uses a large number of disks (10 or more) and uses 30% - 40% of them for checking or parity information. This way when each block is written to the system a checksum (interleaved Hamming code) is generated

and stored on one of the check disks. In the event of a failure the data on the failed disk can be reconstructed from the data on the check disks. Because of the large number of disks required and its optimization for large blocks of data RAID 2 is relatively impractical for PC or workgroup solutions. RAID 2 is found primarily on mainframes.

- RAID 3 employs a single check (or parity) disk for several data disks. When a block is written to the volume the data is distributed across the data disks, then a checksum (XOR) is written to the parity disk. Because data is written and read in parallel from many disks, performance is highest for large blocks of data. RAID 3 implementations are not as efficient with lots of small bits of data. However, RAID 3 uses proportionally more disk space for data and less for parity (~20%) than a RAID 1 system.

  It is important to remember that the majority (70-80%) of time taken in reading or writing is seek time. Having multiple disk mechanisms means the drives can seek to different places at the same time. While one drive is writing, other drives can be seeking so when the system gets to the next drive it is ready to read or write immediately. This is true for all multiple disk RAID implementations.

- RAID 4 systems are best for frequent small reads. Here there is only one parity disk for multiple data disks. Check information is written to the parity disk for each sector that is written to a data disk. The data is spread on the data disks a full sector at a time. Since there is only one disk for check data writing is slower than reading, but more disk space is available for data.

- RAID 5 is the most efficient, highest performance, and most reliable implementation. This system spreads the check data across all disks in the array allowing check data to be written in parallel with other data on other disks. RAID 5 also allows "hot-swapping" meaning that a failed disk can be removed and replaced in the array while the system is still running. The data on the failed drive can be rebuilt on the new disk by recombining the check bits on the other disks. This system is completely redundant and extremely fault tolerant. It is also best for reading and writing both large and small blocks of information.

Keywords: <None>

==========

This information is from the Apple Technical Information Library.

19960215 11:05:19.00