Performance Parameters in IIS 6

IIS 6 contains a huge array of performance parameters—from the way you can control individual applications and their use with specific processors right down to the way application pools and worker processes are recycled. Some of this is covered in the *IIS 6 Delta Guide* book, but some of it has existed in past versions and deserves a mention.
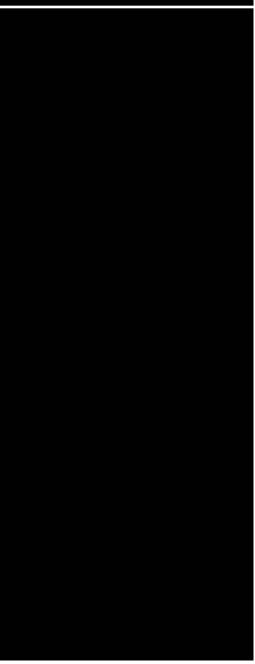
The full range of performance parameters includes

Worker processes and recycling—Because worker processes can be used to handle the individual requirements of an application or Web site, we have a wide range of control over the division of labor. With worker processes, we can assign jobs to a specific worker process or even to a specific processor. With the recycling functionality, we can control how frequently and when worker processes are recycled. From that, we can control the reliability and availability of our applications.

Idle timeouts—On a machine with many Web sites and worker processes to support them, it is a waste of both CPU and RAM to keep worker processes idling without any work to do. IIS 6 provides the capability to shut down unused worker processes after they have been idle for a reasonable period and then automatically start them up again when they are required.

Request queues—The new HTTP.sys kernel mode driver is responsible for accepting requests from clients and putting them into queues ready for a worker process to execute the request. By fine tuning the request queue limit, we can gain a balance between the horsepower of the server, its application loading, and when it classes itself too busy to accept new work. Ultimately, this affects the

perceived response time of the server—set a request queue length too high and the machine might take a long time to respond to the requests; too low, and it refuses requests even when it's not busy.

CPU monitoring—Monitoring individual CPU loading allows you to halt excessive CPU use before it gets out of control and completely overtakes your server.

Web gardens—Multi-processor machines and the new hyper-threading technology offered by newer Intel P4 CPUs means that it's possible for a single machine to handle a number of simultaneous applications without causing a significant overloading of the processors. Web gardens can make use of the multi-processing facility by enabling multiple worker processes to work within a single application pool.

Health monitoring—By monitoring worker processes to make sure they are executing correctly, haven't crashed, or aren't over using resources, IIS 6 can re-create worker processes to ensure the availability of your Web site.

CPU affinity—Allows you to assign worker processes to specific processors. This can either be used to take advantage of the better cache or other abilities of a processor, or to devote a dedicated CPU to a single application to maximize performance.

Quality of Service (QoS) parameters—These set specific parameters on the way that IIS handles requests, such as keeping connections alive, limiting the number of connections, and throttling the bandwidth on a site-by-site basis.