Environmental Systems Research Institute, Inc., 380 New York St., Redlands, CA 92373-8100 USA • TEL 909-793-2853 • FAX 909-307-3014

Environmental Systems Research Institute, Inc.

Dual Core CPU ArcIMS Capacity and Scalability Testing

Enterprise Systems Lab Test Report

Prepared by:

Jeff DeWeese
ESRI Implementation Services
Redlands, California

March 20, 2006

Test Objective

Evaluate performance and capacity of an ArcIMS server based on dual core Intel CPUs relative to single core Intel CPUs. Secondary objective is to perform scalability tests on the dual core server with hyper-threading enabled and disabled to model scalability for both configurations.

Test Hypothesis

Single Core vs. Dual Core Performance and Capacity Comparison Test

CPU cores are essentially CPUs and therefore a dual CPU/dual core server should perform in a similar manner as compared to an equivalent quad CPU server. Standard SPEC benchmarks, which have in the past been shown to predict ArcIMS performance fairly well, can be used to predict performance on a dual core server relative to a single core server. For the comparison to follow, a dual 3.2 GHz Xeon baseline server will be tested and compared to a test server, which is a dual CPU/dual core 2.8 GHz Xeon. Based on the published SPECint_rate2000 benchmarks and comparing CPU counts, the dual core 2.8 GHz server with a SPECint_rate2000 of 59.2 should provide roughly the same user performance and roughly twice the overall capacity relative to the 3.2 GHz baseline server with a SPECint_rate2000 of 28.2.

Dual Core Hyper-Threading Scalability Test

Past testing has shown a 20-30% increase in capacity with hyper-threading enabled and is expected to be realized here as well.

Test Hardware and Software Configuration

The description of the baseline server and test server are presented below in Figures 1 and 2.

Figure 1
Baseline Server Configuration

Configuration Item	Configuration
Make and Model	Inline
Operating System	Windows Server 2003, SP1
Number of CPUs & Type	Two Single Core Intel Xeon 3.2 GHz 1 MB L3 Cache
Memory	2 GB
SPECint_rate2000	28.2
Software	ArcIMS 9.1

Figure 2
Test Server Configuration

Configuration Item	Configuration
Make and Model	HP DL380 G4
Operating System	Windows Server 2003, SP1
Number of CPUs & Type	Two Dual Core Intel Xeon 2.8 GHz 2x2 MB L2 Cache
Memory	2 GB
SPECint_rate2000	59.2
Software	ArcIMS 9.1

Test Data and Tools

The test data descriptions are presented in Figure 3.

Figure 3
Test Data Description

Layer	Data Type	Feature Type	Features
Fire Hydrants	Shapefile Layer	Point	23,585
Schools	Shapefile Layer	Point	1,091
Roads	Shapefile Layer	Line	145,027
Train Routes	Shapefile Layer	Line	387
Freeways	Shapefile Layer	Line	275
Parcels	Shapefile Layer	Polygon	783,159
Parks	Shapefile Layer	Polygon	5,239
Lakes	Shapefile Layer	Polygon	704
Zip Codes	Shapefile Layer	Polygon	106
Raster	TIF Raster	8-bit	NA
		Total	959,573

Test Plan

Single Core vs. Dual Core Performance and Capacity Comparison Test

Execute ArcIMS Image Service web load tests (hyper threading disabled) for both vector/raster and vector-only configurations against both servers. A total of 10 test/Spatial Server threads will be used to ensure server saturation.

Dual Core Hyper-Threading Scalability Test

Execute ArcIMS Image Service web load tests from 1 to 12 Spatial Server and client batch threads for both hyper-threading enabled and disabled configurations against the dual core server. The data set for all tests will be mixed vector and raster. Hyper-threading provides two pre-stage thread areas per CPU which allows threads to effectively run quicker on the CPU and therefore with hyper-threading enabled, it appears as if there are eight CPUs.

Test Results and Analysis

Single Core vs. Dual Core Performance and Capacity Comparison Test

Figure 4 depicts measured throughput. The results show that the server reached near maximum capacity and measured web transaction throughput came in very close to the predicted values and was in fact approximately double that of the baseline server. This demonstrates how the SPEC integer benchmarks can be used for predicting ArcIMS capacity and shows that a CPU core is essentially a CPU, which means that the dual CPU dual core server is equal to a four CPU server.

Figure 4 also shows measured response times. Response times for the dual core server were approximately half that of the baseline server. This is expected since for both tests the same number of test threads were used but since the dual core server has twice the CPUs available, twice as much concurrency occurs and results in response time that are half as slow.

Figure 4
Throughput and Response Time

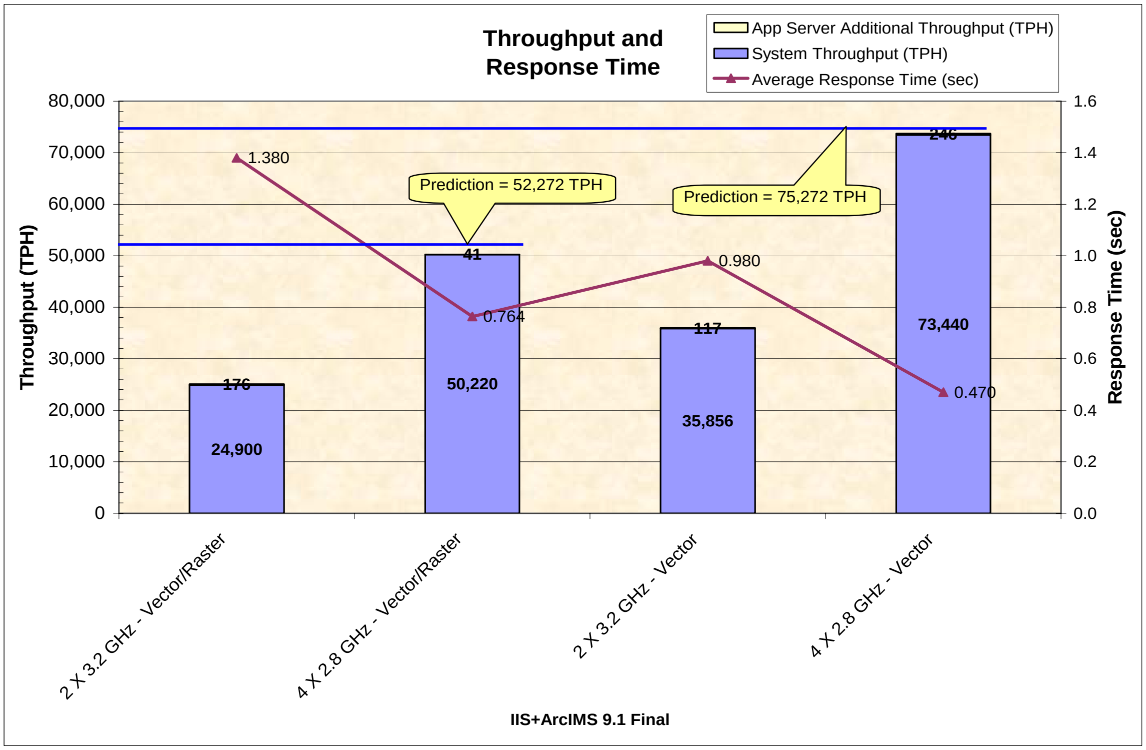


Figure 5 shows CPU utilization. Both servers were nearly maximized for all tests to demonstrate maximum throughput. No CPU bottlenecks were observed.

Figure 5
CPU Utilization

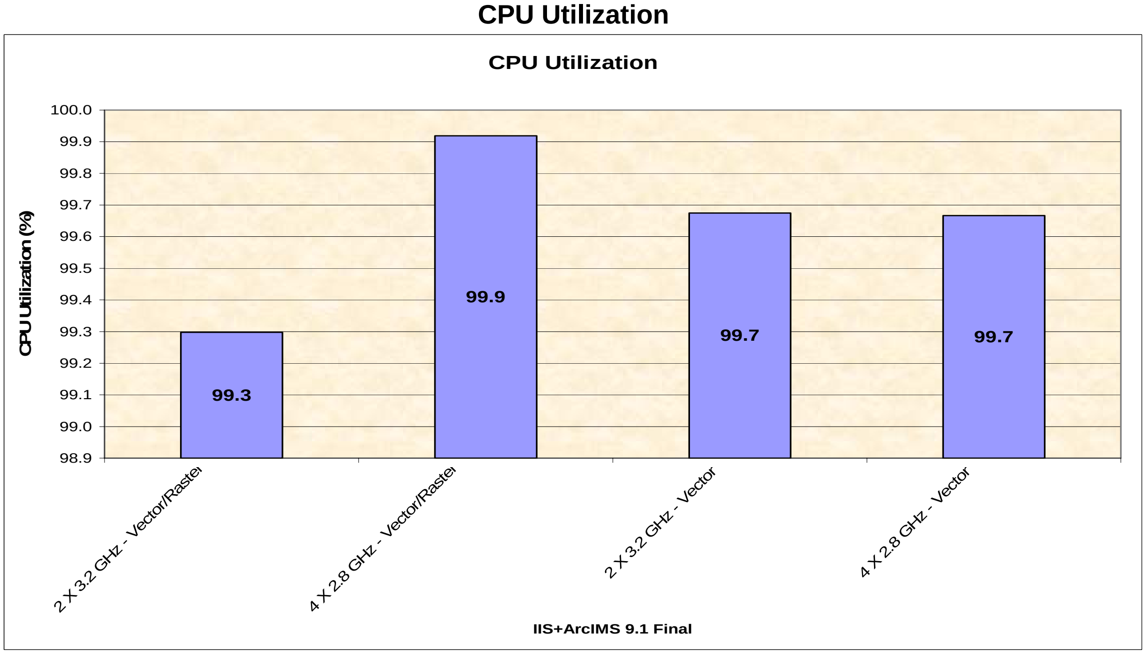
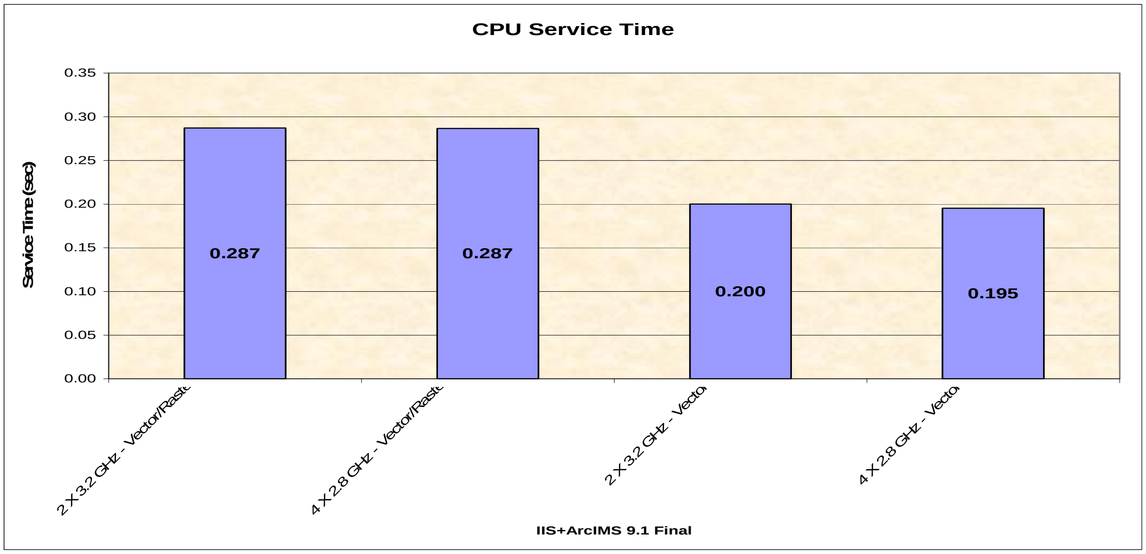


Figure 6 depicts CPU service time. Service times were nearly identical for both test type comparisons validating that the CPU performance of the 3.2 GHz processor is on par with the 2.8 GHz processor. The 2.8 GHz processor, though slower in terms of GHz, is based on newer Intel technology and is performing as fast as the older 3.2 GHz version.

Figure 6
CPU Service Time



Dual Core Hyper-Threading Scalability Test

Figures 7 and 8 show throughput and response time for the scale up tests for both hyper-threading disabled (HTD) and enabled (HTE), respectively.

Figure 7
Throughput and Response Time - HTD

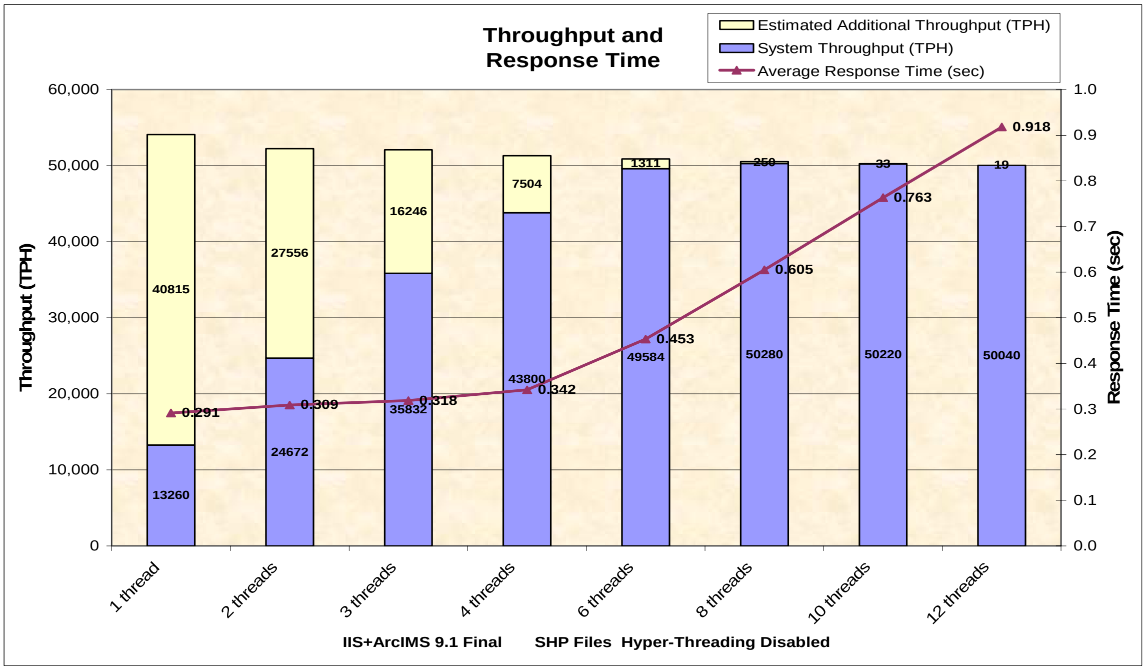
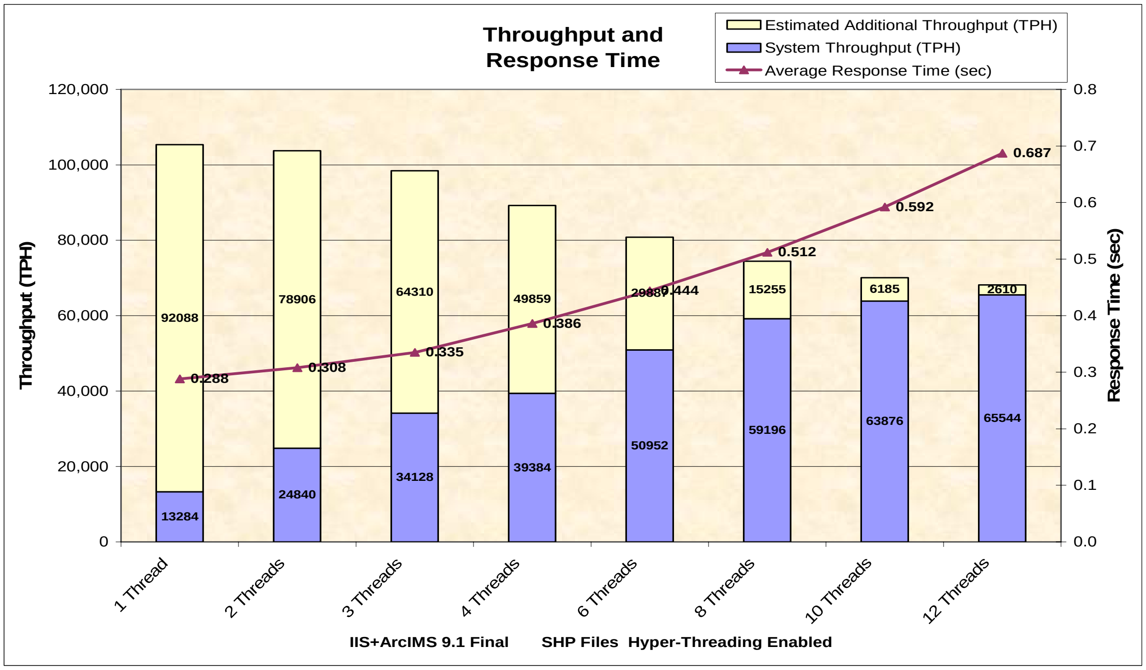


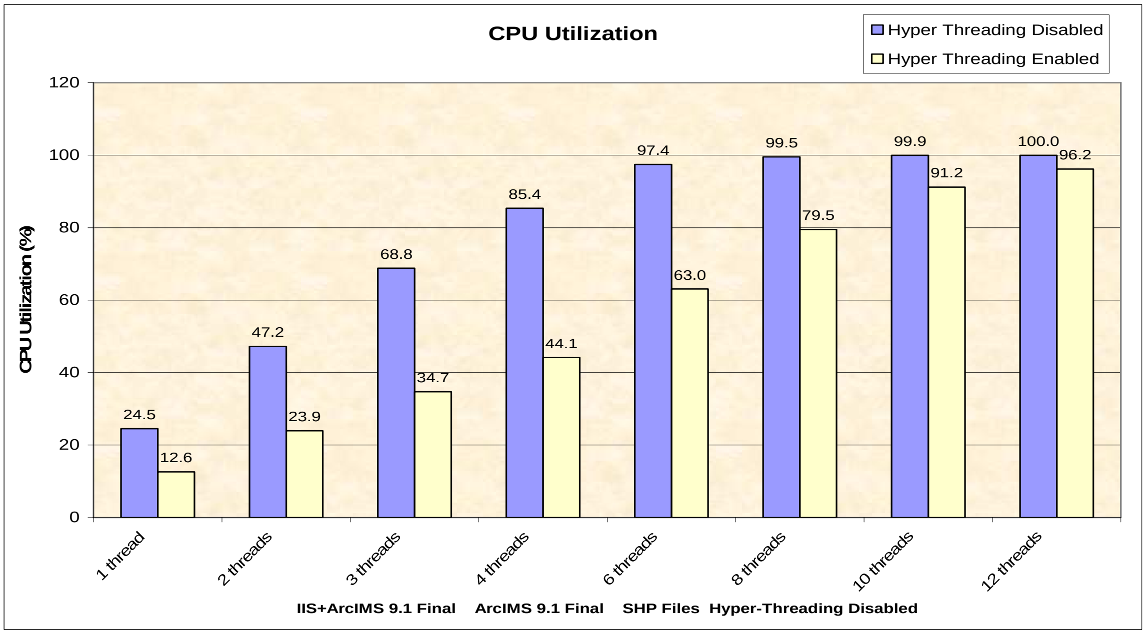
Figure 8
Throughput and Response Time - HTE



For HTD, scalability progressed as expected and peaked at two threads per CPU and at 50,280 TPH. As expected, response time remained fairly flat out to the point where the number of test threads equaled the number of available CPUs. The results for HTE are somewhat ambiguous but are mostly explainable. Throughput increased gradually as expected and peaked out at three threads per CPU and at 65,544 TPH. This is a 30.4% increase in throughput over the HTD configuration, which is consistent with previous hyper-threading comparison testing. The noticeable drop in estimated additional throughput as threads were increased is due to the way in which CPU utilization is reported when hyper-threading is enabled. CPU utilization measurements are inaccurate for HTE since they are based on the assumption by the operating system that eight CPUs are present when in reality only four CPU cores exist. The estimated additional throughput is calculated from CPU service time which is based on measure CPU utilization. In regards to response time, it unexpectedly increased gradually and somewhat more quickly than the HTD configuration, even when the number of test threads was equal to or less than the number of available CPU cores. The curve did not remain fairly flat then kick up at six threads as with the HTD configuration. However, after four threads, the response times were better than the HTD configuration as expected.

Figure 9 shows CPU utilization for both HTD and HTE and demonstrates the CPU measurement problem. While under lighter loads and where the number of test threads was equal to or less than the number of available CPU cores, measured CPU for HTE was half that of HTD. However, CPU utilization became more comparable as the server became CPU saturated. Again, this is just how the operating system reports CPU utilization for HTE since it believes there are actually eight CPUs available.

Figure 9
CPU Utilization



Conclusions

The single core vs. dual core results showed that the SPEC integer benchmarks continue to be a good estimator of expected performance when used to extrapolate performance from a known baseline. The benchmark results predicted that the dual core server would provide twice the capacity (throughput) and this was in fact the case. Individual performance is based on the speed of a single CPU. The CPU service time comparison showed that the newer 2.8 GHz CPU performance is on par with the older 3.2 GHz CPU. Also, the testing showed that a core is essentially a CPU providing twice the capacity due to twice the number of available CPUs.

The dual core hyper-threading scalability results showed that the dual core server scales as expected from light to heavy loads and that there was a 30.4% increase in throughput with hyper-threading enabled. The response times for the HTE tests were not quite as expected but did ultimately show improvement over the HTD configuration.

Configuration Support

It is important to realize that for enterprise configurations such as the one described in this report, ESRI Support will typically be limited to supporting only the ESRI software components (ArcIMS, ArcSDE, etc.). ESRI Support strives to provide the best assistance possible, but problems or questions regarding third-party applications and components may require you to contact the support services provided by the respective vendor.

ESRI offers ongoing, dedicated assistance with the design and/or implementation of an enterprise configuration through our Professional Services Department. Professional Services staff has real-world experience with enterprise configurations, as well as expert knowledge of ESRI resources. To learn more about what ESRI Professional Services can offer your organization, please see the contact information posted at http://www.esri.com/consulting/contact.html.