ServiceSPAN®
Making Work Centers Work Better

Prepared by:

ServiceSPAN

in cooperation with

Sun Microsystems

Application Load Test

March 18, 2007
Version 3

Introduction

ServiceSPAN, a provider of work center automation software to enterprise businesses, performed a load test of their Work Center Manager (WCM) application on November 30, 2006. WCM is a purpose-built application that provides a configurable infrastructure for the management, presentation and resolution of work items. WCM, which is based on service oriented architecture (SOA), leverages many SUN technologies including Java and JSP, Solaris 10 with Zones and SunFire servers; to achieve configuration flexibility in system-human and human-human business processes while exceeding end users' transaction response time expectations.

The purpose of the load test was to benchmark the transaction scalability of the WCM application, by measuring the response time under user requested auto-selected work (the "AUTO transaction") which is accomplished by clicking on the AUTO button in WCM's Intelligent Work Portal (IWP).

The "AUTO transaction", is the most costly transaction in the WCM application" as measured by processor and I/O resources. Both Sun Microsystems' SunFire T-2000 and the SunFire V890 were used.

The load test was performed in cooperation with resources from Sun's ISV Engineering (Tim Mac), Sun's EzQual Virtual Lab (Marion Trompeter), and Sun's Benchmarking Center (Robert Tsai).

About ServiceSPAN

ServiceSPAN provides work center automation software to improve system-to-human and human-to-human workflow. Our software is typically deployed in work centers where reducing work item handling time, distributing work efficiently, increasing information flow and simplifying user interfaces are key components of efficient operations.

Our product, Work Center Manager, is a purpose-built application that provides a configurable infrastructure for the management, presentation and resolution of work items.

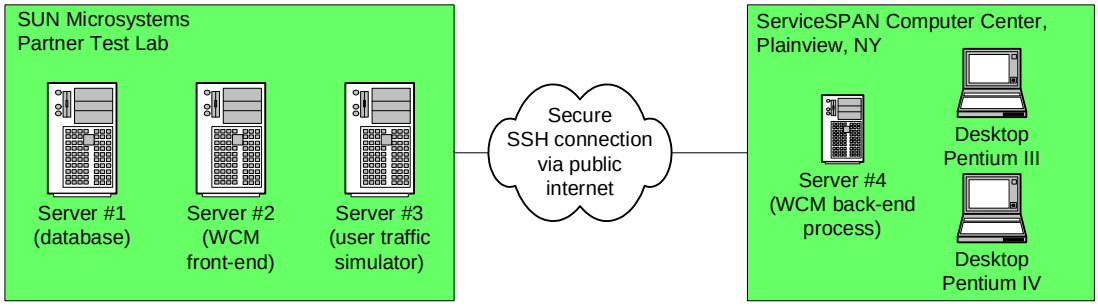
WCM creates value in a work center by focusing on sub-task automation. Work center tasks are comprised of sub-tasks such as; looking up information, cross referencing data, interacting with other organizations, updating data entry values, etc. The cost of the "task" is a product of the time and resources (labor and systems) required to perform the task. Sub-task automation, directly reduces the cost of a "task", by reducing the amount of time and resources required to perform the "task". As the percentage of sub-task automation increases so does the percentage of savings, along with WCM's value and the ability to implement Zero-Touch initiatives.

Customers have found that our value proposition of improved work center productivity, reducing operating costs, and facilitating faster service delivery have driven a strong ROI and improved their bottom line financials.

Load Test Environment

Two separate load testing environments were setup. One environment was based on the SunFire T-2000, to simulate smaller installations of WCM, while the other was based on the SunFire V890, to simulate mid-sized enterprise wide installations. Each environment used separate hardware for the WCM application and for the Oracle database, to enable specific measurement of WCM's IWP response time under load.

SunFire T-2000 Test Environment Configuration



Server #1 (*Database*):
- T-2000 8-core single 1Ghz CPU
- 16 GB RAM
- Oracle 10g set to automatic optimization (no fine-tuning)

Server #2 (*WCM*):
- T-2000 8-core single 1Ghz CPU
- 16 GB RAM
- WCM application with business rules for Telecom Switch Rejects
- Telecom Switch Reject documents

Server #3 (*Traffic Simulator*):
- T-2000 8-core single 1Ghz CPU
- 16 GB RAM
- Simulated user requesting work, making instantaneous decisions and requesting next piece of work

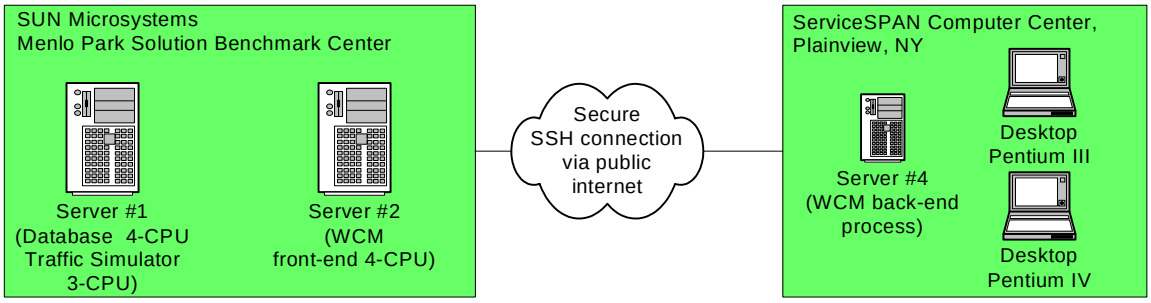
Server #4 (*Data Source*):
- SunFire V240
- 2 processor
- 4 GB RAM
- Generated (+/- 10%) 2,000 exception documents per hour

Desktop #1 (*User*):
- Pentium IV 3.2 Ghz
- 1 GB RAM
- Microsoft Windows XP
- Used to manually perform transactions to confirm response times

Desktop #2 (*User*):
- Pentium III 800 Mhz
- 512 MB RAM
- Microsoft Windows XP
- Used to manually perform transactions to confirm response times on a slower desktop

Secure SSH Connection:
- Used between the Sun lab and ServiceSPAN New York facility
- Encrypted SSH tunnel established over the public internet
- Connection to the public internet = 3MB/sec.

SunFire V890 Test Environment Configuration



Server #1 Zone 1 (*Database*):
- V890 (4) CPU
- 24 GB RAM
- Oracle 10g set to automatic optimization (no fine-tuning)
- Zone 0 configured with (1) CPU and not used

Server #1 Zone 2 (*Traffic Simulator*):
- V890 (3) CPU
- 8 GB RAM
- Simulated user requesting work, making instantaneous decisions and requesting next piece of work
- Zone 0 configured with (1) CPU and not used

Server #2 Zone 1 (*WCM*):
- V890 (4) CPU
- 16 GB RAM
- WCM application with business rules for Telecom Switch Rejects
- Telecom Switch Reject documents
- Zone 0 configured with (1) CPU and not used

Server #4 (*Data Source*):
- SunFire V240
- 2 processor
- 4 GB RAM
- Generated (+/- 10%) 10,000 exception documents per hour

Desktop #1 (*User*):
- Pentium IV 3.2 Ghz
- 1 GB RAM
- Microsoft Windows XP
- Used to manually perform transactions to confirm response times

Desktop #2 (*User*):
- Pentium III 800 Mhz
- 512 MB RAM
- Microsoft Windows XP
- Used to manually perform transactions to confirm response times on a slower desktop

Secure SSH Connection:
- Used between the Sun lab and ServiceSPAN New York facility
- Encrypted SSH tunnel established over the public internet
- Connection to the public internet was less than 1.5 MB/sec.

Test Methodology

The methodology used for the load test was to simulate end-users performing complex, multi-host, multi-frame, manual web based transactions in a pure Java-based environment while the application continued to perform dynamic prioritization and routing.

At each step of the simulation the current time was logged to facilitate the calculation of ñTotal Transaction Timeò. To increase load on each successive test run, additional scripts were run simultaneously.

To insure the user traffic response times were valid, the database was simultaneously loaded with back-end traffic in the form of processing the receipt of 2,000 documents/hr for the T-2000 tests; and 10,000 documents/hr in the V-890 tests, while the user traffic simulator scripts were running.

Test Script

A complete User Traffic Simulator script consists of the following steps:

1. Logging into the application (and receiving the various software components through the network to the desktop browser)
2. Executing an ñAUTO Transactionò
3. Executing 10 complete transactions in rapid succession, without delay
4. Clicking on tabs to look for information such as history (research)
5. Executing a Cancel, Transfer or Complete action for the work item
6. Logging out, and repeating at Step 1 without delay

Transactions

The desktop computer of each WCM user generates the maximum load on the server at the moment the Intelligent Work Portal [AUTO] button is clicked. At this moment the application:

1. chooses the highest priority work item that meets the userôs work assignment and skills
2. locks that record to the user
3. executes rules against each element of the selected work item to determine what should be rendered to the userôs Intelligent Work Portal (IWP).
4. generates a stream of data back to the desktop that will provide the user with a ñtask orientedòcomposite presentation of a work item.

To measure ñTransactions Per Minuteòwe calculate the Total Transaction Time as shown below:

A. **“AUTO transaction” Time =** Time from Auto button click till all data sent to desktop
B. **User Click Time** = Time to process user clicks to review additional information in IWP (e.g., research)
C. **Wrap Up Time** = Time to process Cancel, Transfer or Complete request
D. **Desktop Rendering Time** = Time required by Internet Explorer to render HTML to user.

Total Transaction Time = A + B + C

End User Wait Time = A + D

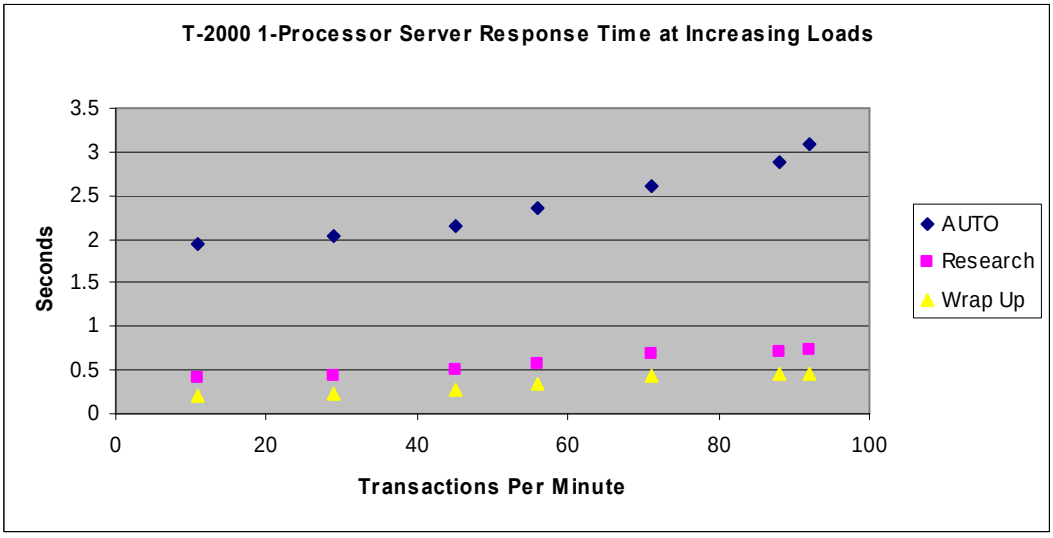
Test Results

Key response time area to watch is the AUTO Transaction Time below, which is the response time of the server to the user's AUTO request under increasing server load.

The ñAUTO transactionò was used because it is the most costly transaction (as measured by processor and I/O resources) in the WCM application.

T-2000 Server Results

Transactions Per Minute	(A) AUTO Transaction Time (sec.)	(B) User Click Time (sec.)	(C) Wrap Up Time (sec.)	(D) Desktop Rendering Time P4 / P3 (sec.)	A + B + C Total Transaction Time (sec.)	A + D End User Wait Time P4 / P3 (sec.)
11	1.952	0.406	0.201	1.0 / 1.7	**2.559**	3.0 / 3.7
29	2.046	0.445	0.240	1.0 / 1.7	**2.731**	3.0 / 3.7
45	2.153	0.514	0.278	1.0 / 1.7	**2.945**	3.2 / 3.9
56	2.355	0.582	0.344	1.0 / 1.7	**3.281**	3.4 / 4.0
71	2.598	0.688	0.431	1.0 / 1.7	**3.717**	3.6 / 4.3
88	2.886	0.709	0.461	1.0 / 1.7	**4.056**	3.9 / 4.6
92	3.079	0.736	0.453	1.0 / 1.7	**4.268**	4.1 / 4.8

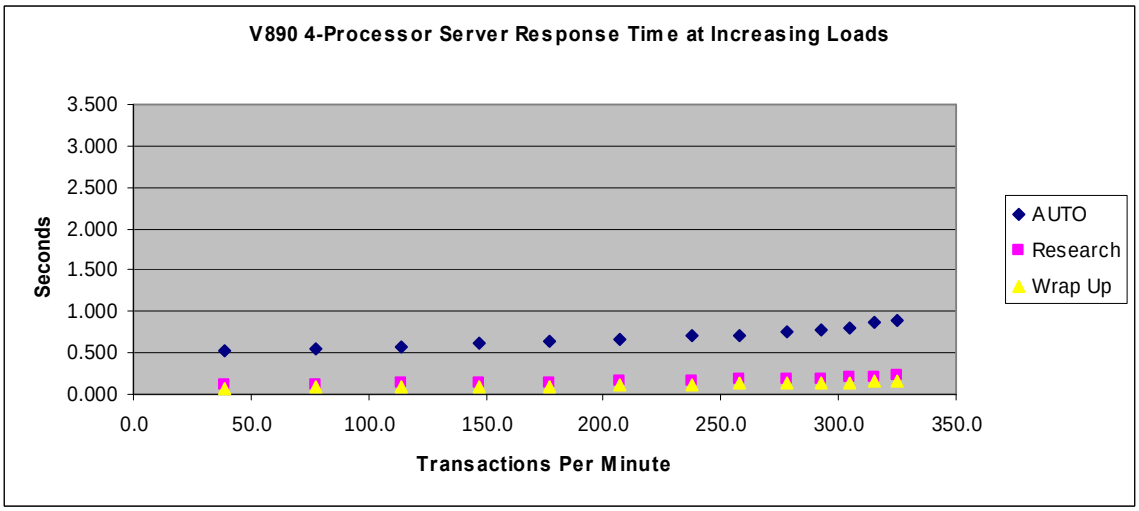


Note the linear performance in the chart, especially between **29** and **88** transactions per minute.

V890 Server Results

Key response time area to watch is the AUTO Transaction Time below, which is the response time of the server to the user's AUTO request under increasing server load.

Transactions Per Minute	(A) AUTO Transaction Time (sec.)	(B) User Click Time (sec.)	(C) Wrap Up Time (sec.)	(D) Desktop Rendering Time P4 / P3 (sec.)	A + B + C Total Transaction Time (sec.)	A + D End User Wait Time P4 / P3 (sec.)	Oracle Database **ZONE** CPU Idle Time (sec.)	WCM Application **ZONE** CPU Idle Time (sec.)
39	0.532	0.116	0.080	1.0 / 1.7	**0.728**	1.5 / 2.2	99	94
78	0.550	0.119	0.083	1.0 / 1.7	**0.752**	1.6 / 2.3	98	89
114	0.579	0.128	0.090	1.0 / 1.7	**0.797**	1.6 / 2.3	97	84
147	0.610	0.139	0.098	1.0 / 1.7	**0.847**	1.6 / 2.3	96	80
177	0.644	0.147	0.100	1.0 / 1.7	**0.891**	1.6 / 2.3	95	76
207	0.668	0.154	0.110	1.0 / 1.7	**0.932**	1.7 / 2.4	95	72
238	0.698	0.165	0.121	1.0 / 1.7	**0.984**	1.7 / 2.4	94	68
258	0.718	0.172	0.128	1.0 / 1.7	**1.018**	1.7 / 2.4	94	65
278	0.752	0.182	0.132	1.0 / 1.7	**1.066**	1.8 / 2.5	93	62
293	0.779	0.184	0.134	1.0 / 1.7	**1.097**	1.8 / 2.5	93	60
305	0.807	0.201	0.144	1.0 / 1.7	**1.152**	1.8 / 2.5	93	59
315	0.861	0.209	0.156	1.0 / 1.7	**1.226**	1.9 / 2.6	92	58
325	0.898	0.231	0.163	1.0 / 1.7	**1.292**	1.9 / 2.6	92	56



Note the linear performance in the chart, especially between **45** and **325** transactions per minute.

Conclusion

SUN T-2000 SunFire Server Platform (1-processor)

- On a T-2000 platform, at the 21 transactions[1] per minute average volume required for users to complete 10,000 transactions[1] over a period of 8 hours, the end users response time on a Pentium IV computer is 3 seconds.

- The T-2000 serverôs performance was well-behaved and linear up through 80 transactions per minute, which could support as many as 38,000 transactions over the same 8 hour period.

SUN V-890 SunFire Server Platform (4-processor configuration)

- On a V-890 platform, even at a much higher 325 transactions[2] per minute, the end userôs response time on a Pentium IV computer was 50% faster, at 2 seconds.

- 325 transactions per minute represents more than 20,000 manual transactions[2] per hour, which could support 5,000 users simultaneously performing manual transactions with a duration of 15 minutes each.

- Our database server reported a 93% idle time, indicating that many additional front-end servers could be inexpensively added to more than triple the number of transactions to 60,000 and users to 12,000, on the same size database server.

  Supporting 60,000 transactions/hour would require four (4) V490 or V890 servers with 4-processors each, or two (2) V890 servers with 8-processors each.

[1]Transaction = One user transaction from start to finish, all inclusive of distributing the highest priority work item that meets the userôs work assignment and skills, executing configurable rules against each element of the work item to provide task oriented rendering of work item and additional supporting data, streaming data back to the user desktop, and the processing of the final action user takes, while the system is simultaneously receiving new work items at a rate of 2,000 per hour in the background.

[2]Transaction = Same as transaction[1] above, except for the V-890 server the load test was run while the system is simultaneously receiving new work items at a rate of 10,000 per hour in the background.

For additional information about this report or Work Center Manager:

ServiceSPAN
101 Sunnyside Blvd.
Plainview, NY 11803
(516) 576-8000
www.servicespan.net