



Planning and Benchmarking New Web Applications

CMG 2001 - Session 4402
December 5, 2001
Ellen Friedman
SRM Associates, Ltd.



Agenda

- ▶ Workload Characterization for Web Applications
 - Understanding Web Log data
 - Unix performance measurements
 - Correlating Web Log data and performance data
- ▶ Methodology for Planning for new Web Applications
- ▶ Case Study: Unix Solaris
 - Environment
 - Methodology
 - Results
- ▶ Summary

www.srmassoc.com

Planning for New Web Applications

- ▶ What do we need to consider?
 - Purpose of the web site:
 - To provide information- Content
 - To sell a product
- ▶ How do we keep the customer happy?
 - Give them what they want- Content
 - Do it quickly
- ▶ Performance is critical
 - Bad Performance = Unhappy customer
 - Abandon the site
 - Go to a competitor



www.srmassoc.com

Predicting Performance for New Web Applications

- ▶ How do we anticipate future performance?
 - Modeling
 - Simulation
 - Benchmarking- Develop representative test cases and
 - Load/Stress Application and Environment
 - Our focus: Developing Representative Scripts for Load testing

www.srmassoc.com



Predicting Web Server Performance

- ▶ Collect measurements from the server environment
- ▶ Collect measurements about the network traffic
- ▶ Understand how the site is being used:
 - Capture/script user sessions at the site
 - Load test based on representative scripts
 - Develop resource profiles for types of user sessions
 - E.g., correlate resource usage and session statistics


www.smassoc.com



Scripting and Workload Characterization:

- ▶ Load testing tools
 - Mercury Interactive – Load Runner
 - Radview – Web Load
 - Empirics
 - Hyperion
 - Load testing services from Keynote: *Key Readiness-Loadpro*

www.smassoc.com



Developing the Script: Creating the workload:

- ▶ Workload Characterization:
 - Representative/typical profiles of work in terms of user behavior and usage of resources
 - Simulate user sessions- what is typical behavior
 - Develop test scripts which can be reproduced to test out different testing scenarios
 - Correlate resource usage with business workload created

www.srmassoc.com



Workload Characterization

- ▶ Customers interact through sessions.
 - What are the characteristics of the session?
 - How does the user navigate through the session?
 - What pages are hit.

www.srmassoc.com

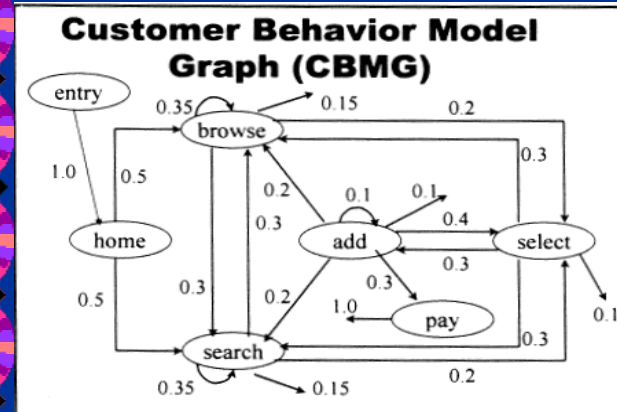
Example: User Sessions

Example: Go to Amazon.com
To buy a children's book

1. Search on children's books, refine search by age, further refine search again by type of book (picture books), select it, look at review, and buy it
2. Buy Curious George without searching

www.srmassoc.com

Customer Behavior Model Graph Example: Buy a Children's Book



www.srmassoc.com Example: User page navigation flow



Web Site Navigation and Metrics

Site Navigation

- ▶ Entry
- ▶ Home
- ▶ Browse
- ▶ Select
- ▶ Add
- ▶ Pay

Measurements

User Session:

- ▶ Page navigation
- ▶ Frequency of Pages accessed
- ▶ Time per page
- ▶ Duration of session

Note: Look at distribution of measurements- min, max, percentiles

www.srmassoc.com

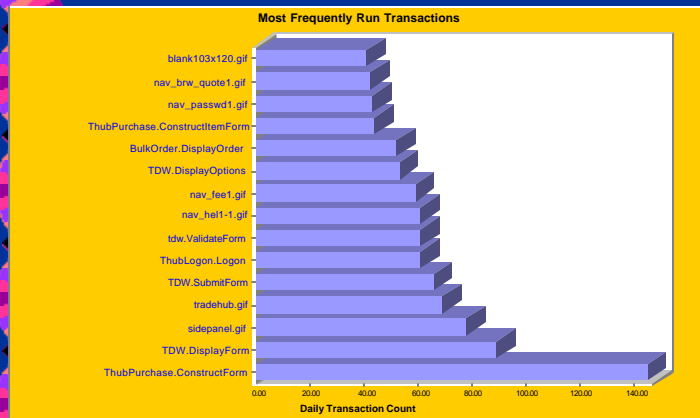


What do we want to find out?

- ▶ What pages/files are being viewed by the user
- ▶ How did the user get to your site- referring URL
- ▶ Where is the user coming from
- ▶ How long is user staying at the site
- ▶ **Can We Get this Information ???**
- ▶ **Analyze WEB Logs:**
 - **almost all Web servers can record this activity**
 - **the information is stored in a plain text format called an access log file or WEB LOG**

www.srmassoc.com

Sample Web Log Analysis



www.srmassoc.com

Workload Characterization

- ▶ Collect Web Log data
- ▶ Filter down to e-business requests
- ▶ Set up into sessions/users (if this is available)
- ▶ Use clustering algorithms to classify sessions into groups (i.e., CBMG pathing)
- ▶ Should be set up by class, ie. Heavy user, occasional, etc.

www.srmassoc.com



Data Available- in web log

- ▶ IP Address
- ▶ Host name
- ▶ Date/Time
- ▶ Request Method
- ▶ Request Path
- ▶ Protocol (http)
- ▶ Return Code
- ▶ Bytes sent/received
- ▶ User Agent
- ▶ Cookie values
- ▶ Browser string

```
195.238.161.136 - - [06/Nov/1998: 14:54:33 +0000]  
"GET/img/navigation/top_nav/jamba_dips_stat.gif HTTP/1.0" 200 743  
"/navigation/top_nav/jamba_dips_stat.html  
HTTP/1.0" "Mozilla/4.05 [en] (Win95; I)" "Cookie data here
```

www.srmassoc.com



Identifying Users or Sessions

- ▶ Can we use session Id or User name- if we add it to the web log- (assumes you have user sign-in /password)
- ▶ IP Address: Proxy servers will resolve to the same IP address- or coming thru a firewall of some type- or AOL etc.
- ▶ Cookies- Best way but some users turn them off and also can come in without it.
- ▶ Caching – presents a problem : page may be cached on clients browser or at proxy server (UUNET, AOL etc.)

www.srmassoc.com



Characterizing Usage

- ▶ Options1: Ignore session characteristics because:
 1. Can't uniquely identify users anyway
 2. Can't identify when user leaves- and where they go, can't necessarily characterize length of session due to some pages being cached.
 3. Just measure page hits and file hits to determine web site usage (useful to correlate with system data for tuning)
- ▶ Options2: Use cookies where they exist or unique users (if password required).
 - Filter data so only those records with cookies/users are analyzed.
 - Use clustering or other graphic techniques to find groups of users

www.srmassoc.com



Examining Web Log Data

- ▶ Identifying users: Use NSLookup facility to resolve IP address
 - Where the users are located
- ▶ Looking at Referring URL- to see whether links are effective or not, advertising is effective or not.

www.srmassoc.com



May need to instrument application/log files

- **Application logs can give us information:**
 - regarding usage of application caches.
 - Regarding transaction types- or pages accessed/user groupings
 - time stamp information
- **Web Logs can be customized:**
 - Different web log formats:
 - Common Log Format (CLF)
 - Extended Common Log Format
 - W3C Standard
 - Other formats may be product-specific or extensions of the CLF/ Extended CLF formats

www.srmassoc.com



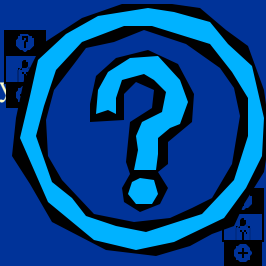
Instrumenting/benchmarking the web application

- **Easier to do with Intranet application**
- **Provides control over what is executed and therefore simplifies analysis**
 - **Not all of the measurement data can be correlated.**
 - **Scripting the application yields**
 - an understanding of application flow and helps to correlate with resource usage

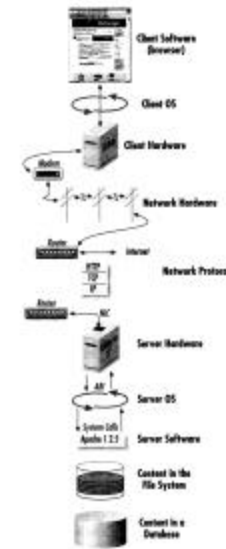
www.srmassoc.com

Setting up the Load Test- What Do We Need To Consider?

- ▶ Hardware
- ▶ Software
- ▶ Infrastructure
- ▶ Network Topology
- ▶ Tools
 - Scripting
 - Load simulation
 - Measurement
 - Analysis
- ▶ Staff
- ▶ Information



www.smassoc.com



Topology Overview

- Server size & Number
- Memory
- Disk layout
- Database
- Network Topology
- Bandwidth

www.smassoc.com



Analyzing Web Server Performance

- ▶ Resource and Business Information must be Correlated:
 - OS level: Server Resource Usage
 - e.g., Unix performance measurements- sar, vmstat, etc.
 - Active Web site monitors:
 - Measures of client response time- e.g., Mercury's Topaz, Keynote
 - Web Logs
 - Log Analyzer tools such as Web Trends Log Analyzer
 - Application instrumentation
 - Application log files

www.srmassoc.com



Measurement Sources-Server

- ▶ UNIX Measurement
 - sar, ps, iostat, vmstat, netstat
 - HP Measureware
 - BMC/Patrol- Best/1
- ▶ NT Measurement
 - Perfmon, NT/SMF, Perfman
- ▶ Z/OS
 - Websphere measurements, RMF/SMF

Need to correlate server resource usage with session metrics

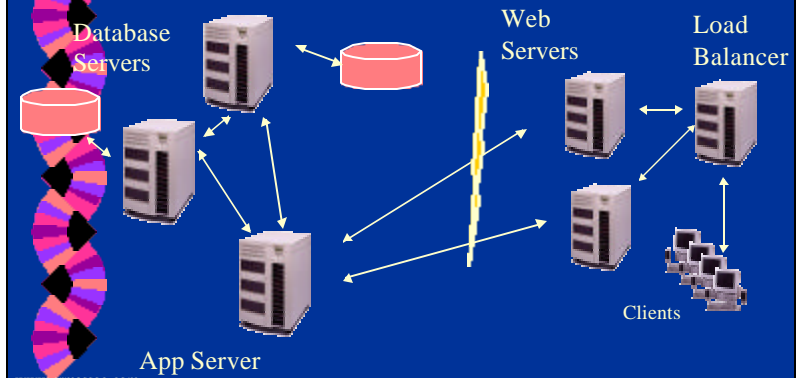
www.srmassoc.com

Questions to Ask

- ▶ How many hits per second?
- ▶ What does the web site do?
- ▶ Have the log files been analyzed?
- ▶ How tolerant are the users/ SLAs?
- ▶ What is average transfer size?
- ▶ Availability?
- ▶ Content and other work?

www.smassoc.com

Case Study: Web Portal Application





Application Background and Scope

1. Portal Application: Dynamic Pages containing specific information on requested topics
 - E.g., examining stock portfolio for a particular client
 - Stock information vs. bond vs. market trend and performance
2. To examine and understand the resource requirements of the portal application.
 - To facilitate projections of the computing resources necessary to support production requirements
 - To determine application scalability

www.srmassoc.com



Test Lab Configuration

<u>Machine Type</u>	<u>Hardware</u>	<u>Software</u>
<u>Load Balancer:</u>	Sun E420 with (1) x 450 MHz CPU, 256 RAM	Solaris 2.6 (5/98)
<u>Web Servers (each):</u>	Sun E420 with (2) x 450 MHz CPU, 2 GB	Solaris 2.6 (5/98)
<u>Application Server:</u>	Sun E420 with (4) x 450 MHz CPU, 1 GB RAM	Solaris 2.6 (5/98)
<u>Database Servers (each):</u>	Sun E420 with (4) x 450 MHz CPU, 2 GB RAM	Solaris 2.6 (5/98), Sybase 11.0.3

www.srmassoc.com



Test Laboratory Facility

Script execution from 50 to 250 users (limited initially based on licensing of load test software)

Test environment and application was optimized over several test executions

- E.g., additional web servers with load balancing added
- Application cache structure was enhanced
- Scripts were modified to reflect changes in application

Application was instrumented to facilitate the logging of session navigation and application and data caching

- Required for measurement of web server cache effectiveness rates, and to assess its impact on scalability

www.srmassoc.com



Testing Analysis Focus

• Determine application scalability and identify potential performance bottlenecks under production loadings

- Server utilization
- Web server balance
- Page load time
- Cache hit rates
- Network activity
- Sybase database activity
- Disk I/O (kbytes/second, utilization)

www.srmassoc.com



Data Sources analyzed

1. Web log data to provide http hits per page, session metrics: page navigation flow and page load time
2. Application log data to provide application cache effectiveness rates
3. Load Runner scenario results- Response time and #hits per page
4. Best/1- CPU metrics, System+User, Disk metrics, network traffic for each server
5. Sybase metrics- Cache effectiveness, Transaction volume and I/U/D to Database

www.srmassoc.com



Example of Web Log Data

```
198.75.82.117 - - [25/Feb/2001:20:00:06 -0500] "GET /ct/ HTTP/1.1" 302 0 "Mozilla/4.0 (compatible; MSIE 5.0; Win32)"
198.75.82.102 - demo [25/Feb/2001:20:00:03 -0500] "GET /ct/img/cp_back.gif HTTP/1.1" 200 105
"Mozilla/4.0 (compatible; MSIE 5.0; Win32)"
198.75.82.102 - demo [25/Feb/2001:20:00:03 -0500] "GET /portal/img/Capital_Raising_up.gif HTTP/1.1"
200 280 "Mozilla/4.0 (compatible; MSIE 5.0; Win32)"
198.75.82.102 - demo [25/Feb/2001:20:00:03 -0500] "GET /portal/img/bc_arrow.gif HTTP/1.1" 200 62
"Mozilla/4.0 (compatible; MSIE 5.0; Win32)"
198.75.82.114 - lferron [25/Feb/2001:20:00:04 -0500] "GET
/portal/process_login.jsp?requestedURI=42Fportal42Fmain.jsp HTTP/1.1" 302 - "Mozilla/4.0
(compatible; MSIE 5.0; Win32)"
198.75.82.114 - - [25/Feb/2001:20:00:07 -0500] "GET /ct/main.jsp?layout=100 HTTP/1.1" 302 0
"Mozilla/4.0 (compatible; MSIE 5.0; Win32)"
198.75.82.104 - schellew [25/Feb/2001:20:00:14 -0500] "GET /ct/main.jsp?layout=100 HTTP/1.1" 200 -
"Mozilla/4.0 (compatible; MSIE 5.0; Win32)"
198.75.82.104 - - [25/Feb/2001:20:00:15 -0500] "GET /ct/main.jsp?layout=3600 HTTP/1.1" 302 0
"Mozilla/4.0 (compatible; MSIE 5.0; Win32)"
198.75.82.115 - lferron [25/Feb/2001:20:00:05 -0500] "GET /portal/main.jsp HTTP/1.1" 302 -
"Mozilla/4.0 (compatible; MSIE 5.0; Win32)"
198.75.82.125 - joelmei [25/Feb/2001:20:00:14 -0500] "GET /ct/main.jsp?layout=100 HTTP/1.1" 200 -
"Mozilla/4.0 (compatible; MSIE 5.0; Win32)"
```

Note: Layout=page 100

www.srmassoc.com



Building the Test Script

Mercury Load Runner was used to Load Test the Client Web application

- Scripts were built based on 80% current production usage of initial application and 20% random access

Analysis of Web and Application logs

- Application log: session identifier, time stamp, IP address and user name.
- Web log data: the number of hits (or gets) to particular page
 - Multiple gets may be required to complete a page e.g., multiple gifs on the page
 - Time stamp for each request and the user (or IP address) making the request

www.smassoc.com



Other Measurement Sources (not initially used in case study)

- Network Monitoring tools: Decomposition of Response time -
 - Server time, network latency, latency through the firewall
- Topaz, Other networking tools
- Keynote: provides measures of external response time

www.smassoc.com

Application Log Analysis

1. Scripts to analyze application and web log data which provided reports summarized for analysis timeframes and included:
 - a. Average, Min, Max server time per page and for each page component
 - b. Count of number of hits per page
 - c. Count of number of hits for each Server- e.g., number of hits to Web02/Web03 and Apps Server summarized by time

Example of Page Load Times

```
[Times are in seconds, down to the millisecond]
Page:Module
-----
Layout >> 100 < === Home
beginningPart      0.174 0.002 0.066 951 0.099 0.100 0.101 0.102
cache-ctrl.inc    0.249 0.001 0.004 951 0.001 0.002 0.011 0.012
pageStart.inc     0.015 0.000 0.001 951 0.001 0.001 0.001 0.001
navigation.inc    0.848 0.002 0.003 951 0.002 0.003 0.003 0.004
* Market Rates    0.294 0.003 0.015 939 0.006 0.016 0.031 0.057 0%
* Custom Charts   0.242 0.006 0.032 939 0.021 0.038 0.081 0.100 1%
* Market Charts   0.277 0.025 0.067 939 0.059 0.086 0.107 0.127 0%
* Market Update   0.629 0.030 0.171 939 0.167 0.192 0.213 0.252 1%
* Research Highlights 0.615 0.015 0.156 939 0.151 0.177 0.206 0.238 95%
* Sector Research 0.593 0.004 0.139 939 0.123 0.139 0.173 0.240 96%
* News Headlines  0.609 0.012 0.147 939 0.124 0.143 0.240 0.249 92%
modules.inc       4.639 0.211 0.379 951 0.117 0.416 0.526 0.609
endingPart        0.016 0.000 0.000 951 0.000 0.000 0.001 0.001
totalServerTime   4.662 0.219 0.454 951 0.408 0.496 0.602 0.684

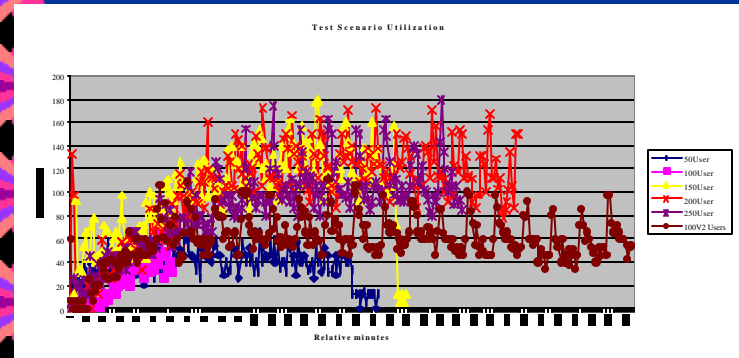
Layout >> 200 < === Capital Raising US High Grade
beginningPart      0.133 0.002 0.065 369 0.099 0.100 0.101 0.101
cache-ctrl.inc    0.056 0.001 0.003 369 0.001 0.002 0.011 0.012
pageStart.inc     0.012 0.000 0.001 369 0.001 0.001 0.001 0.001
navigation.inc    0.025 0.002 0.003 369 0.002 0.003 0.003 0.003
* US HG Spread History 0.078 0.006 0.020 362 0.019 0.025 0.033 0.039 0%
* US HG Issuance Volume 0.429 0.001 0.018 362 0.014 0.024 0.031 0.038 1%
* US HG Syndicate Calendar 22.939 0.178 0.922 362 0.239 0.367 1.467 4.448 0%
* US HG Benchmark Trading Le 13.334 0.002 0.493 362 0.008 0.024 1.742 2.977 1%
* US HG iPricer    23.003 0.136 1.314 362 0.321 1.083 3.889 5.062 84%
modules.inc       23.122 0.034 1.392 369 0.410 1.154 3.991 5.120
endingPart        0.023 0.000 0.000 369 0.000 0.000 0.001 0.001
totalServerTime   23.236 0.063 1.464 369 0.504 1.232 4.051 5.224
```

Case Study Analysis: Unix and Sybase Measurements

1. Best/1 reports:
 - a. CPU usage: System +User
 - b. Disk metrics
 - c. Network traffic
 - d. Response time per server component
 - e. Process statistics showing CPU usage per process (e.g., Java CPU usage, Sybase CPU usage)
2. Sybase metrics- Cache effectiveness, Transaction volume and I/U/D to Database

www.smassoc.com

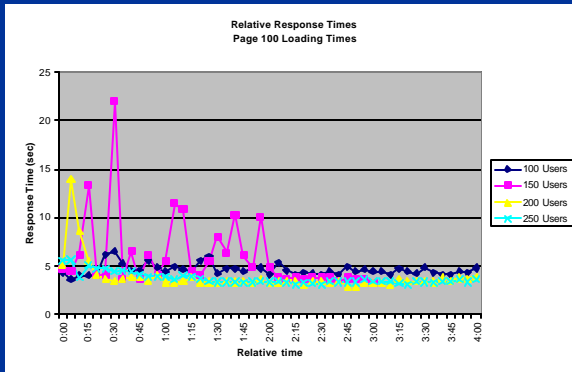
Database CPU Utilization



Note: Ramp up period until 1.5 - 2 hours into test

www.smassoc.com

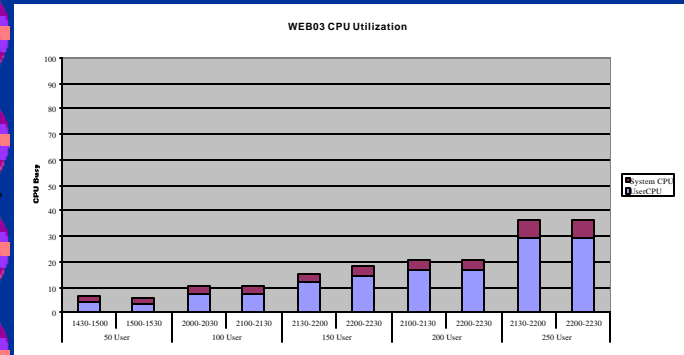
Relative Response Times Page 100 Load timings



Note: 150 user test exhibited greater variability

www.smassoc.com

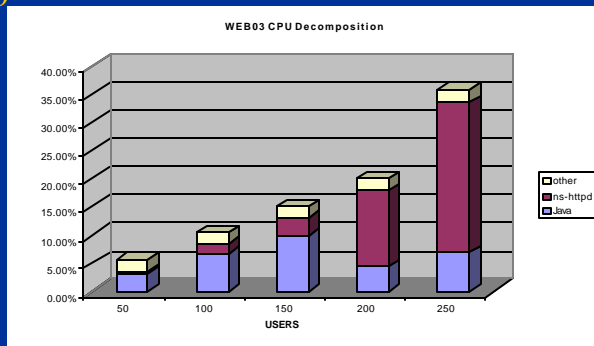
CPU Utilization for Web Server



Note: Imbalance existed between Web2 and Web3

www.smassoc.com

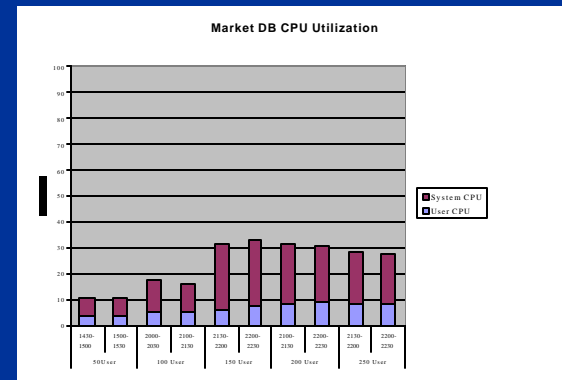
Decomposition of CPU Utilization for Web Server



Note: Scaling of total CPU but shift in workload composition

www.srmassoc.com

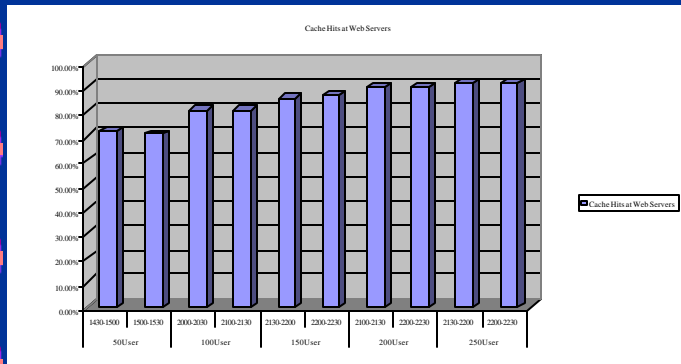
Market Database CPU Utilization



- CPU Usage increased from the 50 to 100 user test from 10 to 15%.
- The 150 and 200 user tests are relatively the same at 30% usage.

www.srmassoc.com

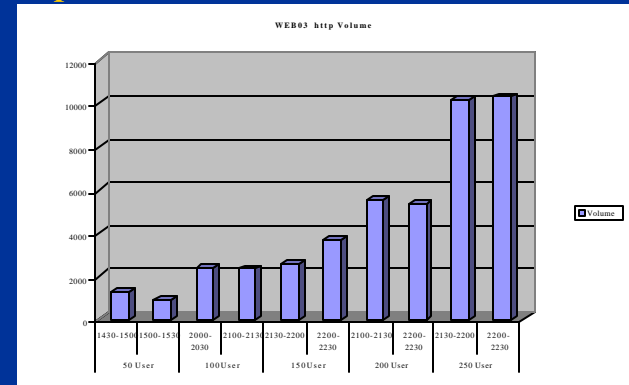
Web Server Cache Hit Rates



- The cache hit rate stabilized after 150 users
- Cache hits for Web servers obtained from Application log

www.srmassoc.com

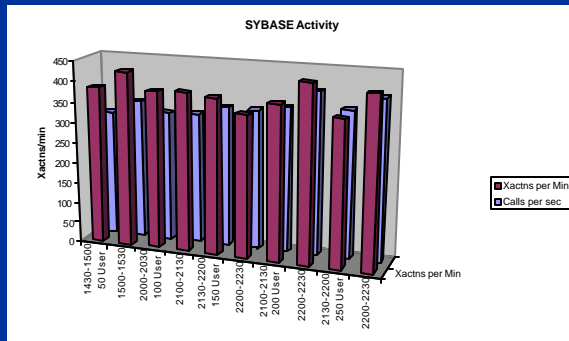
http Volume to Web Server



http volume for Web server obtained by analyzing Web and Application logs

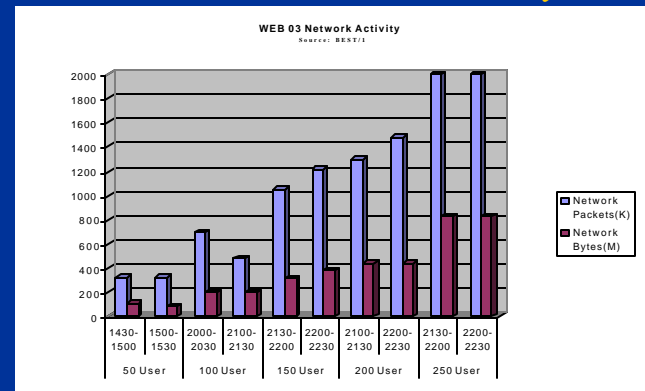
www.srmassoc.com

Sybase Database Activity



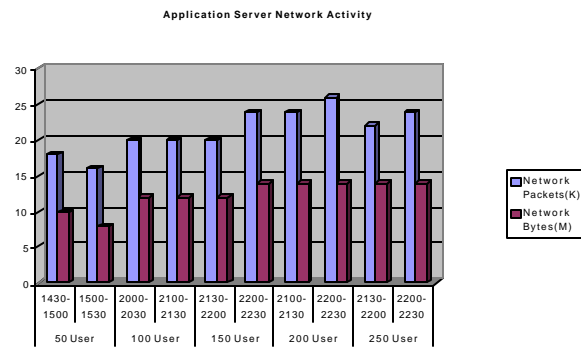
SPMON Data analyzed: Shows similar profile beyond 150 users

Web Server Network Activity



Network traffic increased linearly

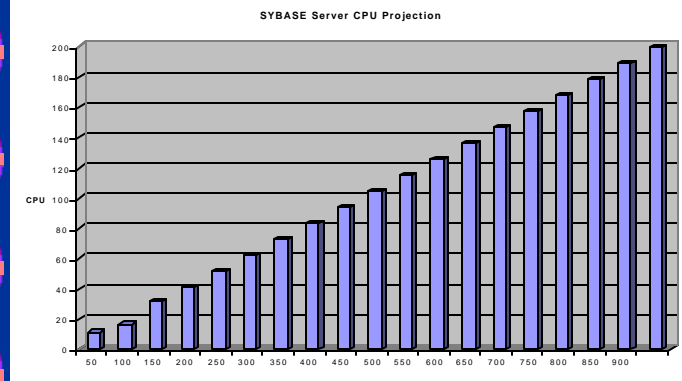
Application Server Network Activity



Application server did not exhibit an increase in network traffic due to caching

www.srmassoc.com

SYBASE Server CPU Projections



Server projections based on the 50:150 cases is conservative
Need to do more testing

www.srmassoc.com



Test Findings

- ▶ Sybase Server can support up to 350 users
- ▶ Each Web Server can support 300-350 users
- ▶ Application server had minimal CPU usage
- ▶ Recommendations: Additional stress testing is required to validate Sybase projections
 - Verify application scalability at higher usage levels

www.smassoc.com



The Testing Process: Checklists

- ▶ Definition
- ▶ Planning
- ▶ Preparation
- ▶ Execution
- ▶ Reporting

www.smassoc.com



Test Definition

- ▶ Who is participating? Interests? Involvement?
- ▶ System configuration to be tested?
- ▶ What to measure? How?
- ▶ Workload to be used?
- ▶ Success criteria?
- ▶ SCOPE !!

www.smassoc.com



Benchmarking and Planning Summary

- ▶ Develop a Prototype
- ▶ Establish a target system (H/W, S/W, Wkld)
- ▶ Test measurement tools
- ▶ Develop tools
 - Develop application scripts
 - Instrument application/log files
- ▶ Execute controlled benchmark
- ▶ Analyze data & repeat if required
- ▶ Extrapolate
- ▶ Document all findings and results

www.smassoc.com



Planning for New Web Applications

- ▶ Determine user profiles – how will users navigate through the application (analyze similar applications that are already built).
 - *Develop scripts to run application in test environment*
- ▶ Execute controlled benchmarks
 - *Analyze resource measurements: web servers, apps servers and database servers.*
- ▶ Instrumentation/Customization of Web and Application logs will be necessary
- ▶ Run the tests , analyze and report results

www.srmassoc.com



Questions?????

Ellen Friedman
SRM Associates, Ltd.
ellen@srmassoc.com
(516) 433-1817

www.srmassoc.com