

Technical White Paper



Actuate 7SP1 Performance and Scalability on Windows 2003

Performance and Scalability on Windows, IBM BladeCenter

Recent benchmark tests at the IBM Innovation Center for Business Partners prove that, for the sixth time in two years, Actuate is the highest performing and most scalable Enterprise Reporting Application platform in the industry.

Among the highlights of the tests:

- Proved able to serve 900,000 named users on just 12 CPUs with an average response time of 3.3 seconds.
- Generated 97 million report pages per day on just 12 CPUs.
- Demonstrated linear scalability (>95%) for report generation and near-linear scalability for viewing.
- Proved able to generate nearly 2.2 million on-demand, fully functional Excel Spreadsheets per day.
- Surpassed all industry records for per-CPU performance of all reporting activities.



Authors

Susan Raisty-Egami, Sr. Product Manager, Actuate Corporation
Roshanak Afsarifard, Director, Engineering, Actuate Corporation
Kelvin Chu, Sr. Software Engineer, Actuate Corporation

Notice

The information in this white paper is proprietary to Actuate Corporation ("Actuate") and may not be used in any form without the prior consent of Actuate.

© 2004 by Actuate Corporation. All rights reserved.

Actuate Corporation trademarks and registered trademarks:

Actuate, e.Report, e.Reporting, e.Spreadsheet, Formula One, ReportCast and Report Encyclopedia.

IBM trademarks and registered trademarks:

IBM, IBM WebSphere, IBM BladeCenter, eServer, DB2

All other trademarks are property of their respective owners.

Version 1, April 23, 2004, © Actuate Corporation

Actuate Corporation

701 Gateway Boulevard
South San Francisco, CA 94080
Tel: (888) 422-8828
<http://www.actuate.com>

Table of Contents

Executive Summary	4
<i>Summary of Benchmark Results</i>	<i>4</i>
Introduction.....	6
The Importance of Scalability	6
<i>What Is Scalability?</i>	<i>6</i>
<i>Scalable Architectures Are Rare</i>	<i>6</i>
<i>Planning for Success Means Planning for Scalability</i>	<i>7</i>
<i>Scalability and Performance: Proof is Required.....</i>	<i>8</i>
Characteristics of Meaningful Benchmarks	8
<i>The Measurements that Matter.....</i>	<i>8</i>
<i>Active Users with Think Time</i>	<i>11</i>
Actuate 7SP1 on Windows Benchmark: Results & Analysis	11
<i>Batch Report Generation</i>	<i>11</i>
<i>Viewing</i>	<i>12</i>
<i>On-Demand Reporting: Generation & Viewing.....</i>	<i>13</i>
<i>Mixed Load: Simultaneous Viewing and Generation.....</i>	<i>14</i>
<i>On-Demand Spreadsheets</i>	<i>15</i>
<i>Overall Architecture</i>	<i>17</i>
<i>System Configuration</i>	<i>17</i>
Conclusion	18
Appendix – At a Glance Summary of Benchmark Results	19
For More Information	20

Executive Summary

With a scalable Enterprise Reporting architecture, growth can be effectively handled by adding hardware.

Today, organizations need to satisfy the reporting requirements of ever-larger user populations, both internally and externally. To achieve 100% user adoption, it is essential to deploy Enterprise Reporting solutions that are built from the ground up to handle continual growth in user populations while maintaining short response times. Only a solution that has near-linear scalability and offers industry-leading performance is capable of providing this.

Scalability is the ability to proportionally increase system capacity, in terms of additional users or transactions, by adding hardware. With the additional hardware, a scalable system can handle the additional users and transactions while maintaining consistent response times and user experiences. A scalable system utilizes additional hardware as efficiently as the hardware already existing in the system.

For applications of all sizes, Enterprise Reporting Applications with proven scalability and excellent performance mean 1) reduced cost, 2) reduced risk, 3) increased user adoption, and 4) better planning and budgeting.

Without scalability, applications will eventually **hit the wall**, where user response times continually increase to unacceptable levels—resulting in an unbearable user experience—where additional hardware does nothing to increase system capacity. The only way to improve such a situation is to rip out the non-scalable system and replace it with a truly scalable enterprise reporting application platform—a very costly endeavor.

Truly scalable solutions are extremely rare. Most vendors lack the enterprise reporting experience to understand and solve the complex technical issues in this area. With our focus on Enterprise Reporting, Actuate has hundreds of man-years of experience in the tough, in-the-trenches work that is needed to develop a truly scalable enterprise reporting architecture. Because of this, while many enterprise reporting architectures hit the wall on configurations with no more than a handful of CPUs, Actuate comfortably scales to extremely large configurations of 80 CPUs or more.

Because most vendors' solutions are not scalable, very few have attempted to prove it in published benchmarks. But scalability is rare even among the handful vendors that have published benchmarks. In fact, careful examination of many of these benchmarks indicates that the solutions do not scale at all, or in fact demonstrate *negative scalability* where adding additional hardware actually degrades performance.¹

In summary, regardless of the size of their applications, organizations should only consider reporting solutions that have been proven—in published, open, detailed benchmark that provide all the result measurements on all configurations—to be extremely scalable and have excellent performance. It is simply too risky and costly to do otherwise.

Summary of Benchmark Results

Actuate recently benchmarked the performance of Actuate 7SP1 on Windows 2003 using IBM BladeCenter hardware at the IBM Innovation Center for Business Partners in San Mateo, California. For the sixth time in two years, Actuate once again proved to be the highest performing and most scalable Enterprise Reporting Application platform in the industry.

¹ Readers of competitive benchmarks should pay careful attention to tests with zero think time but long response times, especially if throughput measurements are not provided. With such benchmarks, to artificially boost the number of users—and thereby give the *appearance* of scalability—a common tactic is to allow much longer response times on larger configurations. To get an *accurate* view of scalability, readers should calculate the system throughput according to the formula described in this paper under *Characteristics of Meaningful Benchmarks*, in the *Throughput* subsection, and ensure that throughput increases proportionately with increased hardware.

Another tactic used to hide a system's lack of scalability is to increase CPU speed as the hardware configuration gets larger, but ignoring these speed difference when reporting the results. So, a system that appears to double throughput when hardware is doubled might only be doubling throughput when hardware is quadrupled or more.

Truly scalable solutions are extremely rare... many enterprise reporting architectures hit the wall on configurations with a small number of CPUs.

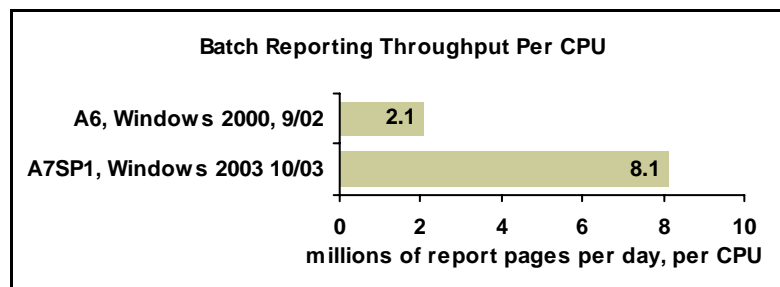
The following table provides a snapshot of the benchmark results. The results all correspond to a 12-CPU Actuate iServer cluster, consisting of 6 IBM HS20 BladeCenter blade servers. Each HS20 BladeCenter blade server had two 2.4 GHz CPUs and 2 GB of RAM. The results all point to Actuate's ability to handle millions of users and hundreds of millions of reporting transactions—all while maintaining a high quality, quick-response experience for users.

Test Scenario	Test Results – Snapshot
Viewing of Cached Reports	<ul style="list-style-type: none"> 9,000 active users Supports user populations of 90,000 to 9 million users
Batch Report Generation	<ul style="list-style-type: none"> 1,128 pages/second Equivalent of 97 million pages per day
On-Demand Report ² Generation & Viewing	<ul style="list-style-type: none"> 1,800 active users Supports user populations of 18,000 to 1.8 million users Produce and view over 4.4 million on-demand reports per day
Mixed Load Test	<ul style="list-style-type: none"> Support 7,500 active viewing users, while simultaneously batch generating nearly 16 million pages per day.
On-Demand e.Spreadsheets	<ul style="list-style-type: none"> 800 active users Supports user populations of 8,000 to 800,000 Generate and view nearly 2.2 million e.Spreadsheets per day

Highlights of the benchmark tests also include the following:

- **Unparalleled performance and near-linear scalability**, for all major reporting activities, from a 1-CPU configuration to a 6-node, 12-CPU configuration.
- **Quadrupling of own industry record for batch reporting per CPU.**³ Compared to Actuate's previous industry-best benchmark on Windows, throughput per CPU increased dramatically, from 2.1 million report pages per day per CPU to a new mark of 8.1 million. Adjusted for CPU speed, this indicates a 25% improvement in generation throughput.

These benchmark tests did not push Actuate 7SP1 to anywhere near its limit.



- **Performance limits have not been approached.** These benchmark tests, on 12 CPUs, did not push Actuate 7SP1 to anywhere near its limit. Earlier published benchmarks proved that Actuate is able to scale across 80 CPUs (over 10 cluster nodes), a huge number of CPUs that can provide enough capacity for millions of users and hundreds of millions of report transactions per day.

² "On-Demand Reports" are reports that are generated on demand and immediately viewed without being stored in the report repository. On-Demand Reports are also called transient reports or express content reports.

³ Actuate 6 Performance on Windows 2003, October 2002. Test configuration was 24 CPUs at 900 MHz.

Introduction

Today, organizations need to satisfy the reporting requirements of ever-larger user populations, both internally and externally. To achieve 100% user adoption, it is essential for organizations to deploy Enterprise Reporting Application platforms that are built from the ground up to handle continual growth in user populations with short, predictable response times. Only a solution that has near-linear scalability and offers industry-leading performance is capable of providing this.

As recent performance tests conducted with the assistance of IBM demonstrate, Actuate 7 delivers unparalleled performance and scalability. Actuate 7 is the Enterprise Reporting Application platform of choice for organizations that want to deliver millions of reports per day to millions of users, as well as for organizations that want to start small and grow as their information demands and user populations increase.

This white paper details the Actuate 7SP1 on Windows 2003 benchmark, conducted in October 2003 at the IBM Innovation Center for Business Partners in San Mateo, California, on leading-edge IBM BladeCenter hardware.

The Importance of Scalability

Benchmarks are useful for illustrating a product's performance and scalability, both of which are key requirements for any Enterprise Reporting Application, big or small. Before delving into benchmark results, it is first essential to understand scalability and why it is important to all organizations, regardless of the size of their applications.

What Is Scalability?

Scalability is the ability to proportionally increase system capacity, in terms of additional users or transactions, by adding hardware. With the additional hardware, a scalable system can handle the additional users and transactions while maintaining consistent response times and user experiences. In a system with perfect linear scalability, doubling the processing power and other hardware resources doubles the system's capacity, all while maintaining consistent response times.

A truly scalable Enterprise Reporting Application platform is specifically designed so that user operations and requests do not interfere with each other. Response time should remain at acceptable levels—less than 5 to 10 seconds—even when under loads that are quite heavy relative to the available hardware. With a scalable, well-tuned Enterprise Reporting platform, performance issues can be effectively solved with additional hardware, *all* of which is efficiently used.

Scalable Architectures Are Rare

While scalability is critical for—and often claimed by—most enterprise software applications, truly scalable architectures are quite rare. This is because scalable architectures 1) are extremely difficult to design and develop, 2) require working knowledge of some of the most advanced concepts in computer science, and 3) entail tremendous upfront cost and time to develop product capabilities that, by definition, are meant to be invisible to the end user and therefore do not “demo” well.

Performance and scalability must be first and foremost considerations when a system architecture is first conceived. Just as a solid foundation cannot be added to a house after the first walls are constructed, scalability and performance cannot be added to a system architecture in a later product release.

Most vendors lack the enterprise reporting experience to understand and solve the complex technical issues in this area. With our focus on Enterprise Reporting, Actuate has hundreds

of man-years of experience in the tough, in-the-trenches work needed to develop a truly scalable enterprise reporting architecture. Because of this, while many enterprise reporting architectures hit the wall on configurations with no more than a handful of CPUs, Actuate comfortably scales to extremely large configurations of 80 CPUs or more.

Planning for Success Means Planning for Scalability

Regardless of the size of their applications, enterprises should only consider reporting solutions that have been proven—in published, open benchmarks—to be extremely scalable and have excellent performance.

Without scalability, applications will eventually fall prey to their own successes. As the numbers of users on a non-scalable system increase, response times will continually increase to unacceptable levels. Worse, there will be a point where additional users cannot be added to the system, regardless of how much additional hardware is deployed. In fact, some Enterprise Reporting Application platforms actually demonstrate worse performance as hardware is added, a phenomenon known as “negative scalability.”

User Population Growth is Inevitable

If an Enterprise Reporting Application is successful, the number of users will inevitably continue to increase well beyond original user population estimates. Once 100% user adoption is attained within the enterprise, most organizations next extend parts of the application to partners and customers. Also, as the value and success of the reporting application is recognized, more Enterprise Reporting Applications are developed and hosted on the same reporting infrastructure, bringing large groups of users onto the infrastructure at once. Finally, today’s volatile business climate, with many acquisitions and mergers, also contributes to the inevitable growth of user populations.

Population Growth + Non-Scalable Reporting = Disaster

Continual user population growth is the hallmark of every successful reporting application, however all this growth can be extremely problematic if the application is built on a non-scalable reporting platform.

In a non-scalable architecture, user requests interfere with each other. As the number of users increase, the number of “collisions” between user requests and jobs exponentially increases, resulting in dramatically longer user response times, inconsistent service levels, and poor user experiences. Additional hardware can alleviate the problem to a small extent, but the collisions continue to increase exponentially, resulting in each additional CPU being used far less efficiently than the last. Finally, at some point, adding more hardware makes no difference to user response times and additional users cannot be supported—***the system has hit the wall.***

When the reporting solution hits the wall,

- User response time skyrockets
- The user experience becomes unbearable
- User productivity plummets
- Service Level Agreements are broken, and reparations often need to be made
- The percent of total users using the system drops off – perhaps dramatically and permanently

At this point, the only remaining course of action is to rip out the non-scalable reporting platform and replace it with a true Enterprise Reporting Application platform that is able to achieve near-perfect scalability—a ***very costly*** endeavor in terms of ***wasted effort and money, lost time, and damaged reputations.***

It is a situation that could have been avoided if a truly scalable Enterprise Reporting Application platform was used from the very beginning.

As numbers of users increase on a non-scalable system, response times quickly rise to unacceptable levels.

Scalability and Performance: Proof is Required

Because scalability is the ability to proportionally increase system capacity by adding hardware resources, a benchmark **cannot** prove scalability unless it compares performance of several different hardware configurations.

Unfortunately for customers, the “scalability benchmarks” published many vendors do not actually prove scalability because only one or two hardware configurations are tested.

For a benchmark to definitively *prove* system scalability:

- The benchmark must publish server capacity, in terms of system throughput, for several hardware configurations with varying amounts of processing power.
- The solution needs to prove its ability to “scale up” over multiple CPUs within a single server as well as its ability to “scale out” over multiple servers in a cluster configuration.
- The acceptable response time of the system must remain constant from configuration to configuration in order to uniformly evaluate the results of one test against another.
- CPU speed, a critical component of any benchmark, must be openly disclosed and discussed. When comparing results across different CPU speeds, the results must be specifically adjusted to take these differing speeds into account.

Recognizing that customers need substantiated, transparent proof of scalability and performance, Actuate has published the results of six benchmarks within two years, all of which were conducted at the world’s leading third-party performance laboratories. All the benchmark tests explicitly demonstrate the scalability of the product in scenarios that cover all the core functionality of an Enterprise Reporting Application platform: content generation and delivering content to users in both batch and on-demand modes. In addition, all the benchmarks are completely open, with complete detail on each scenario, including the detailed breakdown of activities and full descriptions of all reports. All tests conform to a uniform, published, repeatable testing procedure and all tests are held to the same, strict, pass/fail acceptance criteria.

Characteristics of Meaningful Benchmarks

All benchmark studies conducted by Actuate are specifically designed to provide customers with as much meaningful information and analysis as possible. The hope is that customers can then apply this knowledge to their own applications and configurations.

The Measurements that Matter

Enterprise Reporting Applications have some unique characteristics, so certain types of benchmark tests and results measurements are needed to properly depict a product’s performance and scalability. To that end, for each configuration, the Actuate Benchmark directly measures:

- 1) **Throughput** rates for both report generation and content viewing
- 2) **Scaling efficiency** across configurations of varying sizes
- 3) The **number of active users** supported for interactive tests (viewing and on-demand reporting).
- 4) **Average user response times** for interactive tests,

Throughput is the most precise measure of system capacity.

Throughput

Throughput is the most critical benchmark measurement for both types of loads because it is the most precise measure of system capacity⁴. It is a measure of how much content can be generated per second, or how much content can be served to users per second. Scalability can only be proven by measuring how a system's throughput increases as hardware resources increase.

For Enterprise Reporting Applications, there are two types of load that require the vast majority of processing power: generating reports and handling user requests to view reports.

For batch generation tests, the server throughput is measured in pages per second or reports per second.

For viewing and on-demand reporting tests, throughput is measured in user requests per second. For Actuate, when a user makes a "user request" to view a report, usually just one URL is triggered. So, Actuate viewing throughput is measured in URLs per second. On-demand reporting throughput is measured in reports/sec.

While the Actuate Benchmark directly measures throughput for all tests, some vendors do not provide throughput figures but instead provide only numbers of users and average response time—figures which alone cannot adequately describe the capacity of the system or its scalability because the number of users can always be dramatically increased by sacrificing acceptable response time. Fortunately, however, throughput can be derived using the below formula. With this formula, organizations can compare throughput of various configurations to determine scalability, or make cross-vendor comparisons.

$$\text{Throughput_in_User_Requests_Per_Second} = \frac{\text{NumberActiveUsers}}{(\text{AvgThinkTime}^5 + \text{AvgResponseTime})}$$

Scaling Efficiency

System scalability is measured in terms of scaling efficiency. In a system that scales in a perfectly linear fashion, the scaling efficiency is 100%. In common usage, systems are referred to as being "X% scalable" if their scaling efficiencies are X%. Scaling efficiency is calculated according to the following formula.

$$\text{Scaling_Efficiency} = \text{Actual_Throughput} / \text{Theoretical_Throughput}$$

where

$$\text{Theoretical_Throughput} = \text{Throughput_for_1_CPU} * \text{Number_CPUs}$$

If scaling efficiency is greater than 95%, systems are typically referred to as having perfect linear scalability. If scaling efficiency is less than 70%, the system is not scalable and should not be considered for enterprise deployments.

Number of Active Users

An important measurement for Enterprise Reporting Application platforms is the number of active users that can be supported on a given configuration. This metric is important for any test that measures the capacity of the server to handle interactive user requests, such as viewing of cached reports or on-demand report generation and viewing.

The definition of "active users" requires some explanation. Any application will support a population of "named users." Named users are all the distinct users that are able to log in to use the server. At any time, some of these named users are *active users*, meaning they are

⁴ As explained in more detail in subsequent sections, the number of users and response times are not true measures of system capacity. This is because the number of users can often be substantially increased by increasing the acceptable response time. The only way to "normalize" benchmarks where both the number of users and the response times change from configuration to configuration is to use these figures to derive the system throughput. The formula for doing so is provided later in this white paper.

⁵ The concept of *think time* is explained in more detail in the following sub-sections.

logged in and actively submitting report viewing and generation requests every X seconds on average, where X is referred to as the “think time.”

In an Enterprise Reporting architecture, the number of active users relative to the number of named users depends on the application. In Actuate’s experience, inside-the-firewall Enterprise Reporting Applications might be up to 10% of the total named user population actively submitting viewing and generation requests during peak usage times. For extranet applications, active users are often only 1% of the total named user population at any one time. For Internet applications, the active user percentage might be a few tenths of a percent.

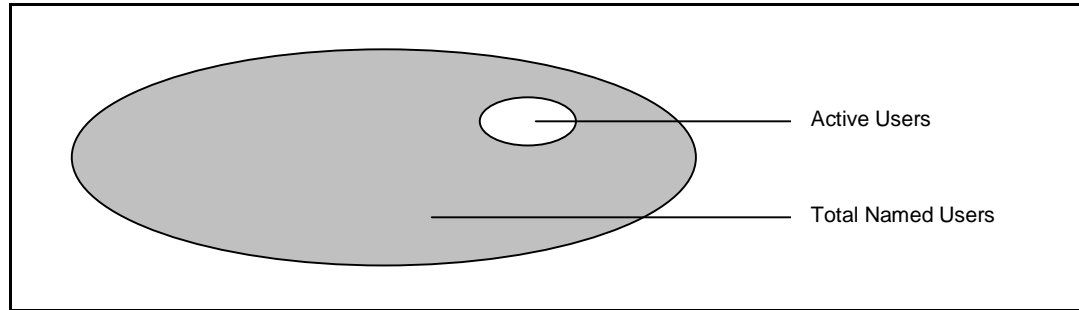


Figure 1 – The total named user population that can be supported is a multiple of the number of active users. Active users are logged in and actively submitting requests on a regular basis.

Response Time

Throughput, user counts, and scalability metrics (described above) are meaningless and can be manipulated unless the test has strict limitations on the acceptable response time. These limitations are needed because an application is simply unusable if response times are too long. 100% adoption simply cannot be attained unless response times are tightly limited.

According to Jakob Nielsen, renowned web usability expert, “10 seconds is about the limit for keeping the user’s attention focused on the dialogue. For longer delays, users will want to perform other tasks while waiting for the computer to finish.”⁶ In accordance with this philosophy, and recognizing that short response times are essential to quality user experiences and 100% user adoption, all published Actuate benchmarks must meet the strictest response time criteria in the industry. If the Actuate response time requirements are not met, the results of the test are not considered valid and are not published in any form⁷.

For the Viewing of cached reports, the following response time criteria must be met for the test to be valid.

- Average and median response time must be under 5 seconds
- 90% of all requests must be serviced within 5 seconds
- 100% of all requests must complete within 120 seconds

The following response time criteria must be met for the On-Demand Reporting and On-Demand e.Spreadsheet benchmark tests to be valid, although it should be noted that most tests demonstrate much shorter response times:

- Average and median response time (time for first page to be generated and returned for viewing) must be under 15 seconds
- 90% of requests must be serviced within 15 seconds

⁶ Refer to Jakob Nielsen’s website at <http://www.useit.com/papers/responsetime.html>

⁷ Note, however, that if customers are willing to accept slower response times, they should be able to support significantly greater numbers of active users on the same hardware.

All benchmarks published by Actuate meet the strictest response time criteria in the industry.

- 100% of requests must be successfully serviced within 120 seconds

Only the results of valid tests, that meet the above criteria, are published within this white paper.

Active Users with Think Time

Meaningful benchmark tests mimic real-world behavior as closely as possible. In the real world, large numbers of users actively use the server. However, at any one instant, only a portion of the users have a request that is being processed by the server while the rest of the users are “thinking,” meaning that they are looking at and thinking about the content returned by their previous request. Also, in the real world, each user spends a different amount of time “thinking.” The result is a non-constant load on the server with periods of activity where the server is much more overloaded than average.

In a meaningful benchmark, therefore,

- 1) The actions of the desired number of users are mimicked as accurately as possible, meaning the benchmark simulates the actual actions of active users with think time.
- 2) The think time is randomly determined, but centered around a system average.

Actuate 7SP1 on Windows Benchmark: Results & Analysis

Actuate conducted cluster and standalone benchmark tests using five types of test scenarios:

1. Batch Report Generation
2. Viewing
3. On-Demand Report Generation & Viewing
4. Mixed Load: Simultaneous Viewing and Batch Reporting
5. On-Demand Spreadsheets

For most test scenarios, the Actuate iServer was tested in both cluster and standalone configurations ranging from 1 to 12 CPUs, with 1 to 6 blade server nodes.

Batch Report Generation

Description

The Batch Report Generation test simulates the simultaneous generation of multiple reports, where each report was scheduled to start running at the same time. Reports are batch generated and stored in the encyclopedia; users do not wait to view the resultant reports. This test determines the average number of report pages that the Actuate iServer can generate per second over a prolonged period of time.

Four variations of this test were used. They differed according to the report used and the data source:

1. 51-ctrl is a 500 page pivot table report that contains an average of 51 distinct elements, also known as “controls”⁸, per page, plus one static image per page.
2. 101-ctrl is a 500 page report that contains an average of 101 controls, plus one static image per page.
3. 101-ctrl-DB2 is the same as 101-ctrl, except that it retrieves from a DB2 database.
4. 72-ctrl-Detail-DB2 is a 270 page complex detail report with 4 levels of grouping, an average of 72 controls per page, and over 270 aggregations.

⁸ A control is a distinct element on a report page, including data fields, labels, lines, etc. A 10 by 10 table contains 100 controls.

Results

On just 12 CPUs, Actuate 7 was able to generate 1,128 pages/second, which is the equivalent of 97 million report pages per day. Scalability was near-perfect; from 1 to 12 CPUs, the scaling efficiency was over 97%. The iServer was able to attain an extremely high CPU utilization, which is essential for makers of truly scalable software.

This test also sheds light on the factors that drive iServer performance. In general, as the number of controls doubled, the throughput of the system halved. Report complexity, such as aggregations and deep grouper, has lesser impact. This test also shows that the choice of data source, be it a flat file or a relational database, has little or no effect on iServer throughput if the data source itself is not a bottleneck.

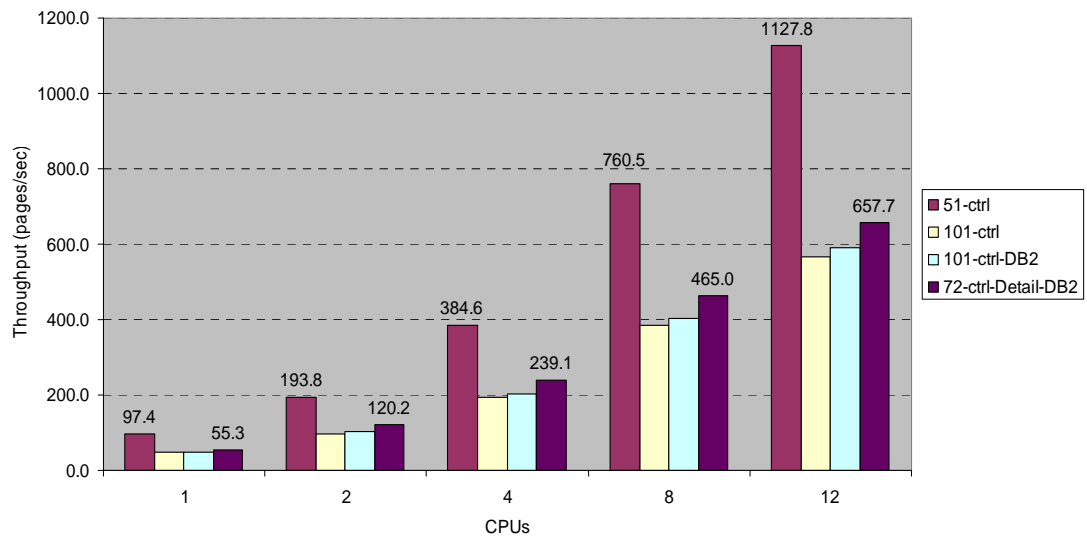


Figure 2 – Batch Generation: Near-perfect linear (> 95%) scalability across 6 nodes / 12 CPUs. The Actuate iServer sustained a generation rate of 1127.8 pages/second for the report called 51-ctrl. This is the equivalent of 97 million pages per day.

Viewing

Description

The Viewing test simulates multiple users simultaneously logging in and conducting different types of viewing requests on randomly selected pages of pre-generated reports. To ensure that this test is as demanding as possible and thoroughly exercises the iServer, *all* users actively view reports; no users are relegated to tasks like portal navigation that require less CPU power.

The main outputs from this test are 1) server throughput, meaning the maximum number of user viewing requests (URLs) the Actuate iServer supports per second, and 2) the maximum number of active users that can be simultaneously logged in and conducting viewing requests while meeting the 5-second response time criteria.

Results

On just 12 CPUs, the Actuate iServer proved able to support 9,000 active users. This translates into support for user populations ranging from 90,000 to 9 million.⁹ These users

⁹ Through years of experience with customers' high-volume reporting applications, Actuate has learned that for internal-to-the-Enterprise applications, typically only 1 in 10 authorized users are actively using the system at any one time. For extranet applications, 1 active user for every 100 authorized named users is a common ratio. For Internet applications, the ratio even smaller: 1 active user for every 1000 authorized named users or less.

collectively submitted 270.6 user Viewing Requests (URLs) per second to the iServer, the equivalent of 23 million viewing requests per day.

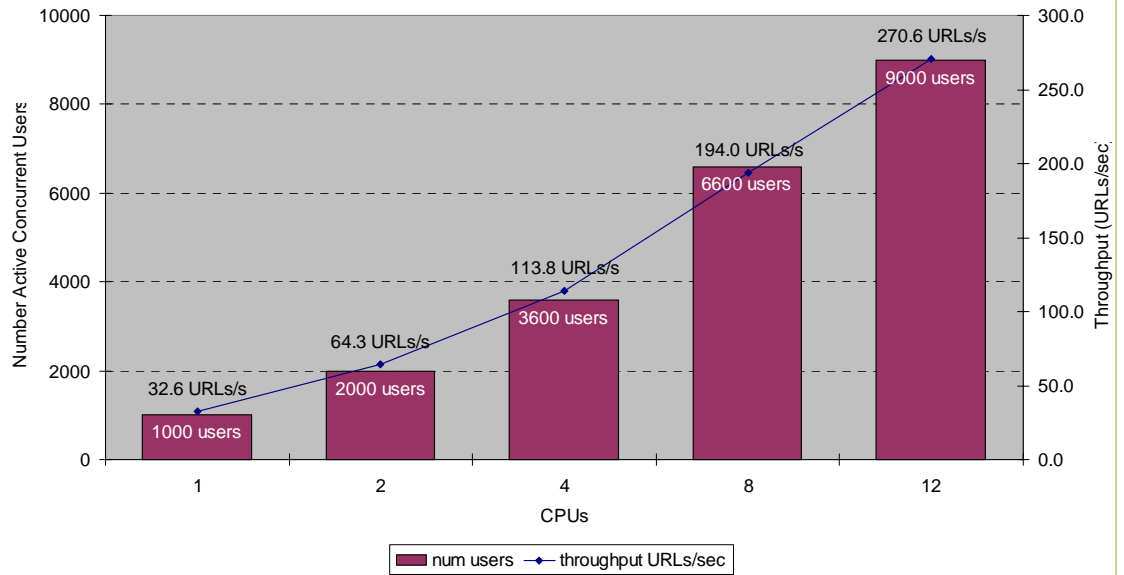


Figure 3 – Viewing Cached Reports. On just 12 CPUs, Actuate supported 9,000 users, or 750 users per CPU.

Average Response times throughout the entire test were less than 3.3 seconds for all configurations, well under the maximum allowed 5 seconds for valid tests, and well within the maximum of 10 seconds required for a web application to provide an acceptable interactive user experience.

	1 CPU	2 CPUs	4 CPUs	8 CPUs	12 CPUs
Response Time (seconds)	0.53	1.41	2.00	2.10	3.30
Active Users	1,000	2,000	3,600	6,600	9,000

Figure 4 – Average Response Times for Viewing. Response times averaged less than 3.3 seconds for all test configurations.

On-Demand Reporting: Generation & Viewing

Description

The On-Demand Reporting test simulates multiple users logging in, where each generates a report on demand and then views the report. The resultant report is not stored in the Report Encyclopedia. To make the test as demanding as possible, all users actively generate and view reports; none are relegated to just portal navigation or other less demanding tasks that are incidental to on-demand reporting.

Like Viewing, this test determines maximum throughput in URLs/second and the maximum number of active users. From the URLs/second, the number of reports per second is calculated, taking into account that an On-Demand report request triggers both generation and viewing URLs.

Results

The Actuate iServer proved able to support 1,800 active users of on-demand reports. These users collectively generated and viewed 50.9 reports per second, or 4.4 million reports per day. Total user populations of 18,000 to 1.8 million can be supported, depending on the characteristics of the application.

The iServer again demonstrated near-perfect linear scalability, with a scaling efficiency of greater than 90%, from 1 to 12 CPUs.

The test was conducted with two different reports, 101-ctrl-rpt and 101-ctrl-DB2-rpt that only differed in the data source used. The test proved that the choice of data source, whether a relational database or flat file, has no impact on iServer performance, provided that the data source itself is not a performance bottleneck.

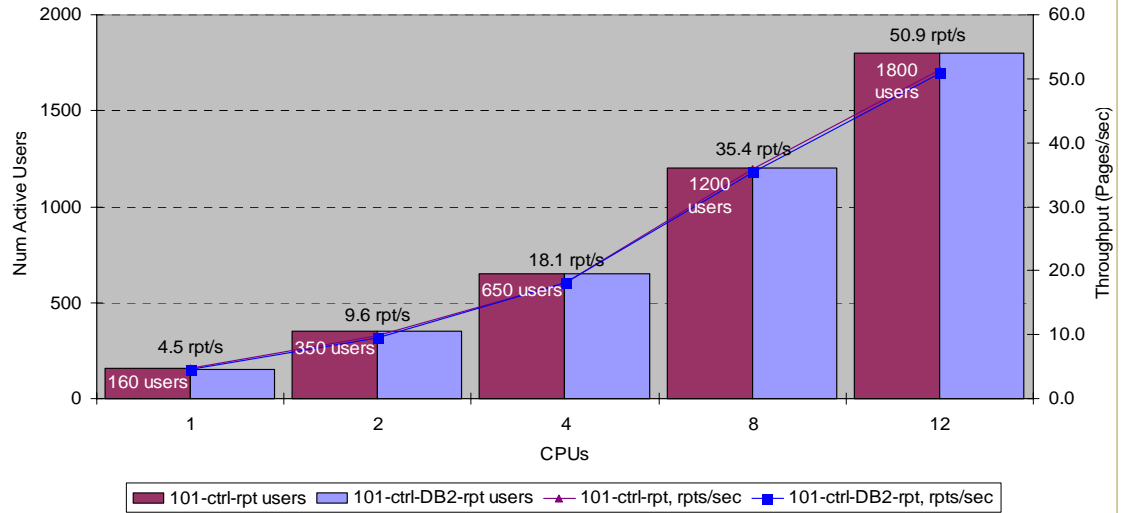


Figure 5 – On-Demand Reporting. *On just 12 CPUs, the Actuate iServer was able to support 1800 active users of on-demand reports.*

Average response times remained under 6.1 seconds for all configurations, well under the maximum 15 seconds limit allowed for valid tests, and well under the 10 seconds required for a very high quality user experience.

	1 CPU	2 CPUs	4 CPUs	8 CPUs	12 CPUs
Response Time (seconds)	2.58	5.13	5.63	6.10	4.40
Active Users	160	350	650	1,200	1,800

Figure 6 – On-Demand Reporting Response Time. *Even under extreme load, Actuate 7SP1 was able to generate and render reports in 6.1 seconds or less.*

Mixed Load: Simultaneous Viewing and Generation

Description

To serve most organizations' needs, Enterprise Reporting Application platform needs to provide excellent performance when dedicated either to viewing or generation of batch reports and on-demand reports, but it also needs to perform well when viewing and generation occur simultaneously.

To simulate such a real-life environment more closely, this benchmark measured Actuate 7SP1 while conducting both viewing and generation simultaneously.

Results

On 12 CPUs, the Actuate iServer supported 7,500 active viewing users while simultaneously generating 185 pages/second of batch reports, or nearly 16 million pages per day.

Also, Actuate 7SP1 easily achieved 100% CPU utilization, which means that the hardware was used efficiently—no processing power was wasted.

On-Demand Spreadsheets

Description

The On-Demand Spreadsheet test simulates multiple users logging in and each submitting multiple on-demand requests to generate Excel Spreadsheets from e.Spreadsheet reports. Users immediately download and view the resulting Excel Spreadsheet within their web browser. The resultant Excel Spreadsheet is not stored in the Actuate Encyclopedia.

As with the other test scenarios, all users actively generate and view reports; none are relegated to just portal navigation or other less demanding tasks that are incidental to on-demand reporting.

Two reports were used:

1. Crosstab – An e.Spreadsheet report with 1 worksheet and 2200 cells.
2. Crosstab-DB2 – The same as Crosstab, except data is from a DB2 database.

Results

The Actuate iServer was able to support 800 on-demand active users generating and downloading Excel Spreadsheet reports. These users generated and downloaded reports at the rate of 25 e.Spreadsheet reports per second, the equivalent of nearly 2.2 million Spreadsheets per day.

The test results prove that the iServer has near-linear scalability when generating e.Spreadsheets, and that the choice of data source—whether a flat file or relational database—has no impact on performance provided that the data source is not the bottleneck.

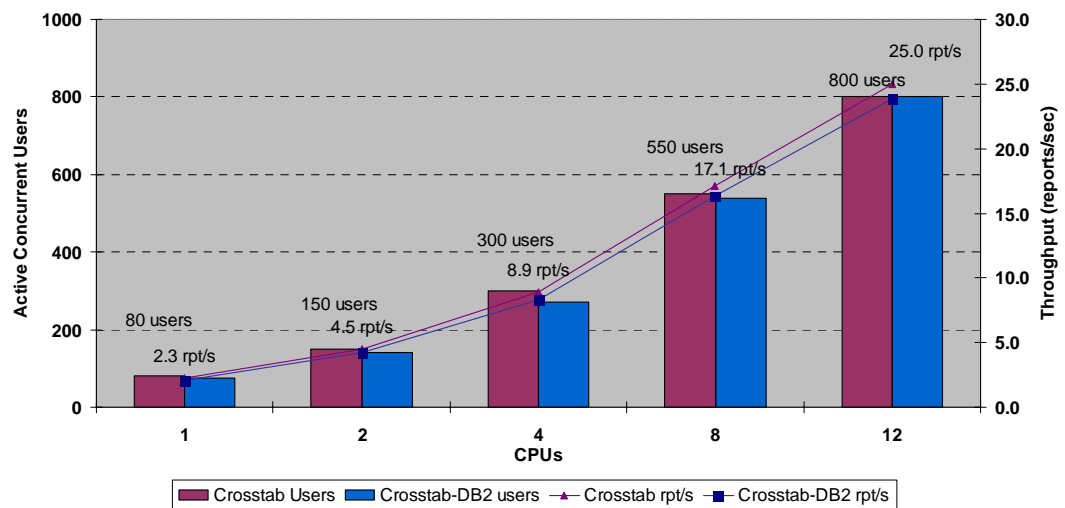


Figure 7 – On-Demand Spreadsheets. On just 12 CPUs, Actuate TSP1 can support 800 active users generating 25 reports per second.

The measured response times included the time to completely generate each Excel Spreadsheet and download it in its entirety. For all test configurations, average response time was consistently under 6 seconds, well under the maximum 15 seconds allowed by the test, and well under the 10 seconds required for a high-quality interactive user experience.

	1 CPU	2 CPUs	4 CPUs	8 CPUs	12 CPUs
Response Time (seconds)	5.60	3.34	3.22	1.72	1.65
Active Users	80	150	300	550	800

Figure 8 – On-Demand Spreadsheets Response Time. Response times for the users to completely generate and download Excel spreadsheets are consistently under six seconds.

Actuate 7 Performance Improvements over Actuate 6

Actuate has long held the industry record for Enterprise Reporting performance on all operating system platforms, including Windows. Actuate 7SP1 shattered the previous Windows industry record, set by Actuate 6 in 9/2002.

Batch reporting throughput per CPU nearly quadrupled from 2.1 million pages per day to a new mark of 8.1 million pages per day per CPU. On-demand reporting performance per CPU tripled, from 50 to 150 active users.

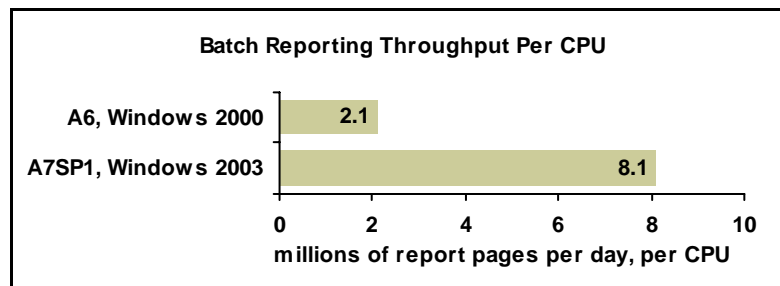


Figure 9 – Actuate 7SP1 vs. Actuate 6, per CPU performance of batch report generation

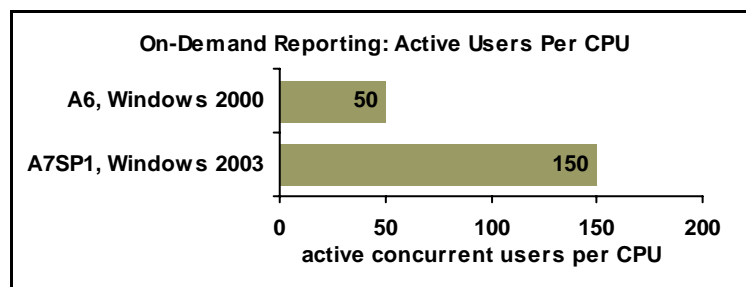


Figure 10 – Actuate 7SP1 vs. Actuate 6, per CPU performance of on-demand reporting

These results prove that Actuate 7SP1 is able to take full advantage of the higher speed CPUs used in this most current benchmark¹⁰, yet another hallmark of a scalability. The ability to take full advantage of today's high-speed CPUs, exceeding 2 GHz, is difficult to achieve. It is much easier to design software that fully utilizes the available processing power on CPUs with lower CPU speeds. Truly scalable architectures are able to take full advantage of additional hardware resources, whether in the form of additional processors or higher speed processors.

¹⁰ The Actuate 6 test used 24 CPUs clocked at 900 MHz.

Furthermore, the results indicate that Actuate 7SP1 still has a 25% improvement in generation throughput over Actuate 6 when the results are adjusted to account for the different processor speeds used in the two benchmarks. This 25% improvement is not specific to the Windows platform; it has been borne out in numerous internal tests and customer deployments across all the operating system platforms that Actuate supports.

Overall Architecture

While some companies optimize their performance test results for publication by directly accessing their server, the Actuate Benchmark configuration is more conservative. It more accurately simulates actual customer configurations by including all the components necessary for a complete information delivery solution in the test environment.

At the IBM Innovation Center, Actuate mimicked real-life configurations as closely as possible. A load simulator created actual user loads by simultaneously submitting multiple URL requests to Active Portal via the IBM WebSphere Application Server. Active Portal in turn relayed requests to the Actuate iServer, which in turn might generate queries against the DB2 database.

Figure 11 illustrates the overall architecture.

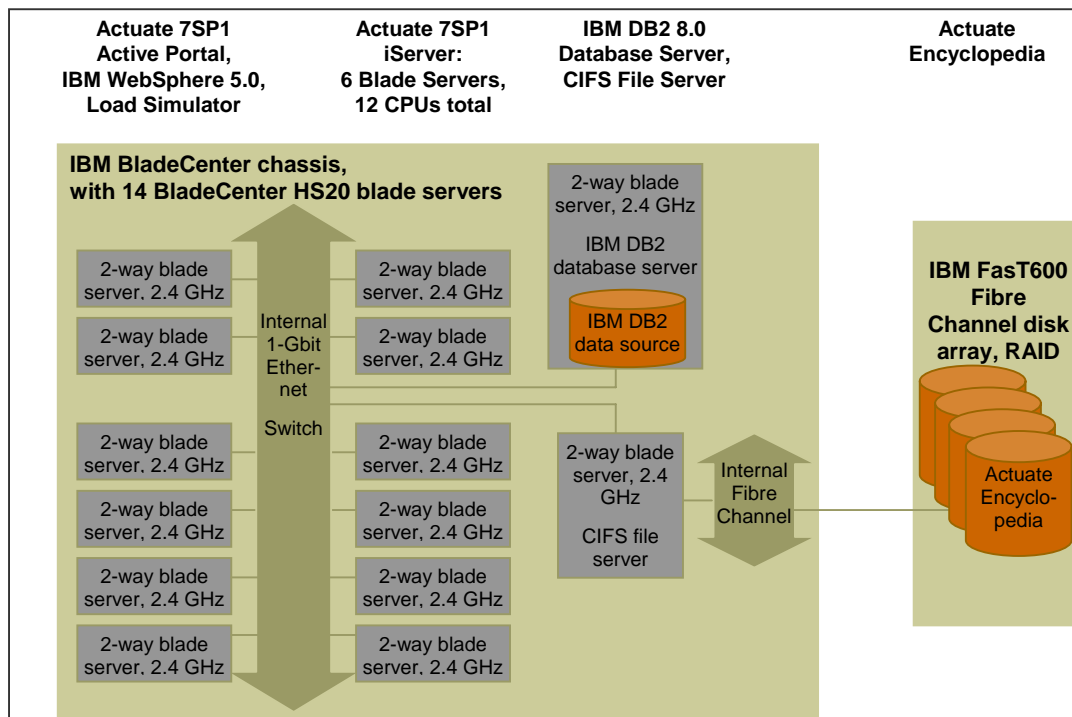


Figure 11 – Benchmark environment. Actuate 7SP1 simulates virtual users in an environment closely resembling a real-life configuration, with an application server layer, database, and file server.

System Configuration

All servers in the end-to-end system configuration—including the iServer, WebSphere, DB2 database server, and file server—were IBM HS20 BladeCenter blade servers with 2 CPUs each. All 14 HS20 Blade Servers were housed within a single IBM BladeCenter Chassis.

The IBM BladeCenter architecture is particularly well suited for Actuate cluster deployments. All the components needed to support several extremely large Enterprise Reporting Applications—including database servers, file servers, application servers and the Actuate iServer—can fit within one 12" high box that fits into a standard server rack. Customers can

start off with a small number of blade servers within the chassis and then incrementally—and cheaply—add more as their application grows.

The IBM BladeCenter is also well suited for the high availability demands of large-scale Enterprise Reporting applications and well complements Actuate’s formidable high availability features. Virtually all components within the BladeCenter chassis—including blade servers (“blades”), backplanes, internal network switches, power supplies, cooling fans, etc.—are redundant and hot-swappable. The Actuate iServer takes care of clustering the blade servers and detecting and recovering from failures.

The Actuate report encyclopedia was stored on an IBM FasT600 Fibre Channel storage array, configured as RAID 10 for both high performance (striping) and high availability (mirroring). All Actuate server nodes communicated with the FasT600 via a dedicated CIFS file server. While IBM FasT storage arrays are suitable for customers with large storage requirements who require 24x7 reliability or prefer a Storage Area Network configuration for the Actuate encyclopedia, it is possible to configure the Actuate server in a low-cost, less reliable configuration.

Component	Software	Number of nodes	Machine Model	Number CPUs	CPU Speed	RAM	O/S
iServer nodes	Actuate 7SP1 iServer	6 nodes	IBM eServer BladeCenter Chassis with 14 IBM HS20 BladeCenter blade servers	2 in each Blade, 12 total	2.4 GHz	2.5 GB each blade	Windows 2003 Advanced Server
Active Portal Application Servers Load Simulator	Actuate 7SP1 Active Portal IBM WebSphere Application Server 5.0	6 nodes		2 in each blade, 12 total			
DB2 database Server	IBM DB2 8.0	1 node		2 CPUs			
File Server	Microsoft Windows 2003 (CIFS)	1 node		2 CPUs			
Actuate Encyc-lopedia	IBM FasT600 storage array, RAID 10						

Figure 12 – Hardware configuration for the Solaris Actuate Benchmark.

Where possible, the configuration was deliberately designed so that the iServer itself was the limiting factor. By providing more than enough processing power to all other system components and by using a network and storage devices that are more than fast enough, we were able to test the true limitations of the iServer.

Conclusion

Actuate 7SP1 was designed to provide enterprises with a very high performance and highly scalable Enterprise Reporting solution. These goals have been met, and surpassed. Actuate 7SP1 demonstrates unprecedented performance and scalability, far exceeding any other solution on the market, all while meeting the strictest application response time criteria in the industry.

If organizations intend to deploy successful Enterprise Reporting Applications that are adopted by 100% of the users, then they *must* deploy with a product with proven near-linear scalability, like Actuate. Unless failure is an option, organizations cannot afford the high cost, lost time, and reputation damage that eventually results from deploying applications on a non-scalable platform.

The IBM BladeCenter architecture is particularly well suited for Actuate cluster deployments.

Appendix – At a Glance Summary of Benchmark Results

	Throughput	Number of Users	Scaling Efficiency	Avg. Response Time
Viewing	271 viewing requests / sec	9000 active users	70%	3.3 sec
Batch Generation	1128 pgs/sec	n/a	97%	n/a
On-Demand Reporting	50.9 rpts/sec	1800 users	94.3%	4.4 sec
On-Demand Spreadsheets	25.0 rpts/sec	800 active users	91%	1.7 sec
Mixed Load (Viewing + Batch Generation)	185.0 batch report pages/sec 222.5 viewing requests/sec	7500 active viewing users	99.7% (83.3% viewing, 16.4% batch)	3.3 sec

For More Information

For more information on Actuate performance and scalability, please refer to these additional white papers and information resources, available on the Actuate public website at <http://www.actuate.com>:

- *Actuate 7SP2 and SAP BW 3.1: Performance on Sun Solaris*
- *Actuate 7 Performance on Sun Solaris*
- *Actuate 6 SP1 Performance on Hewlett Packard HP-UX*
- *Actuate 6 Performance on Windows 2000*
- *Actuate 6 Performance on IBM AIX: Cluster Configuration*
- *Actuate 6 Performance on IBM AIX: Standalone Server Configuration*
- *Meeting Scalability Requirements for Web-Based Information Delivery Solutions*
- *Benchmark Load Simulation Tools Yield Equivalent Results*