

Estimate performance and capacity requirements for   
Visio Services in SharePoint Server 2010

This document is provided “as-is”. Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it.

Some examples depicted herein are provided for illustration only and are fictitious. No real association or connection is intended or should be inferred.

This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.

© 2010 Microsoft Corporation. All rights reserved.

Contents

[Estimate performance and capacity requirements for Visio Services 3](#_Toc261872909)

[Introduction 3](#_Toc261872910)

[Test farm characteristics 3](#_Toc261872911)

[Test definition 3](#_Toc261872912)

[Dataset 4](#_Toc261872913)

[Recommended and maximum configurations 4](#_Toc261872914)

[Hardware setting and topology 5](#_Toc261872915)

[Lab hardware 5](#_Toc261872916)

[Lab software 5](#_Toc261872917)

[Topology 6](#_Toc261872918)

[Test results 7](#_Toc261872919)

[Overall scale 8](#_Toc261872920)

[Hardware cost per transaction 8](#_Toc261872921)

[Recommendations 9](#_Toc261872922)

[Hardware recommendations 9](#_Toc261872923)

[Scaled-up and scaled-out topologies 10](#_Toc261872924)

[Optimizations 10](#_Toc261872925)

[Settings optimizations 10](#_Toc261872926)

[Common bottlenecks and their causes 12](#_Toc261872927)

[Performance monitoring 12](#_Toc261872928)

[Visio Services performance counters 12](#_Toc261872929)

[Visio Web Access performance counters 12](#_Toc261872930)

[See Also 13](#_Toc261872931)

|  |
| --- |
|  |
| Estimate performance and capacity requirements for Visio Services in SharePoint Server 2010 |
|  |

This performance and capacity planning document provides guidance on the footprint that usage of Visio® Services in Microsoft® SharePoint® 2010 has on topologies running SharePoint Server 2010.

For general information about how to plan and run your capacity planning for SharePoint Server 2010, see [Performance and Capacity Management](http://technet.microsoft.com/en-us/library/cc262971(office.14).aspx).

## Introduction

Visio Services enables users to view Visio Web drawings in SharePoint Server 2010. It also enables users, when Web drawings are connected to external data sources, to update the visual appearance of Web drawings, based on underlying data changes. For more information about Visio Services, see the [Visio Services overview (SharePoint Server 2010)](http://technet.microsoft.com/en-us/library/ee663485(office.14).aspx) TechNet article.

This paper addresses the effect of topology on the overall service **latency** (how long it takes to render a drawing) and **throughput** (how many drawings can be processed in a second) when a typical mix of drawings is rendered. The test outlined in this paper measures performance under recommended user load and maximum user load on an instance of Visio Services that is configured with default settings.

With this information, you can better scale your deployments based on your latency and throughput requirements. It is important to note that the specific capacity and performance figures that are presented in this article are different from the figures in real-world environments. The figures presented are intended to provide a starting point for the design of an appropriately scaled environment. After you have completed your initial system design, test the configuration to determine whether your system supports the factors in your environment.

## Test farm characteristics

There are several possible configurations that can be used when you are setting up a test environment for capacity planning. Based on empirical evidence and on best practices, the Visio team settled on what it considers to be a set of common characteristics for hardware, topology and datasets. This section describes what those are.

### Test definition

The following test procedure was used for each performance scenario. Note that this test assumes a farm that is dedicated to Visio Services (no other SharePoint tests running) and that only one user is using the service.

|  |  |  |
| --- | --- | --- |
| Test ID | Test name | Test description |
| #1 | Rendering a Typical Mix of Visio Web drawings | 1. From the browser, a simulated user requests Visio Services to render a drawing. 2. Visio Services renders the drawing by using the Portable Network Graphics (PNG) output format and returns it to the user via the browser. If the drawing is data connected, the drawing is refreshed before it rendered. 3. Visio Services caches the drawing to reduce latency the next time it is requested. 4. Steps 1, 2, and 3 are repeated with files randomly picked from the [dataset](#_Dataset) with no think time between requests. |

### Dataset

Visio Services capacity and performance depends on, among other things, the makeup of the Web drawings that are hosted on the service. File-related factors that affect performance are listed below:

|  |  |
| --- | --- |
| Factor | Qualitative effect on performance |
| Drawing disk size | Larger files sizes increase latency on the SharePoint network. |
| Drawing complexity | Drawings with many complex shapes increase rendering latency and resource usage on the application servers. |
| Drawing data-connectivity | Visio Web Drawings can be either static or data-connected.  For static drawings, Visio Services loads drawings and renders directly to the browser.  For data-connected drawings, Visio Services performs the extra step of polling the underlying data sources for fresh data and updating the drawing before rendering it to the browser. This extra step increases latency and resource usage on application servers. |

To simulate the effect of these factors, files were categorized into the following buckets:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Small | Medium | Large |
| Number of shapes | 20 | 50 | 100 |
| Typical File Size | 300 KB | 600 KB | 900 KB |
|  |  |  |  |
| If data connected\*: |  |  |  |
| Number of refreshable elements | 10 | 120 | 180 |
| Number of rows imported | 10 | 40 | 80 |
| Percent of data change | 50% | 50% | 50% |
| Frequency of refresh | 100% | 100% | 100% |

The different file buckets were then assembled into a test dataset by using the following distribution:

|  |  |
| --- | --- |
| Types of file | Percentage |
| Small static | 49% |
| Small data-connected | 21% |
| Medium static | 14% |
| Medium data-connected | 6% |
| Large static | 7% |
| Large data-connected | 3% |

\*Note that data-connected files were connected to a SQL Server data source via Secure Store Service Authentication. This paper does not cover the performance effect of drawings that are connected to other external data sources.

### Recommended and maximum configurations

For each topology configuration a green zone and red zone user load was determined before running throughput tests. These configurations are defined below:

|  |  |
| --- | --- |
| Configuration | Definition |
| Recommended (Green Zone): | The user load at which the test being run consumes approximately **half** the bottlenecking resource. In the case of Visio Services this is the CPU usage on the front-end Web server (WFE).  You should expect the Green Zone throughput to be sustained for long periods of time in real world deployments. |
| Maximum (Red Zone): | The user load at which maximum throughput was reached for a topology while latency is at minimum. After this point, throughput typically remained stable while latency increased.  You should expect the Red Zone throughput to be tolerated for a short period of time on a farm, but it should be avoided. |

## Hardware setting and topology

### Lab hardware

To provide trend information, several farm configurations were used for testing with increasing computing capacity. Farm configurations ranged from two to five Web servers plus a single database server running Microsoft SQL Server® 2008. Testing was performed with one client computer generating all requests. All Web server computers and the database servers were 64-bit.

The following table lists the specific hardware that was used for testing.

|  |  |  |  |
| --- | --- | --- | --- |
| Machine Model | DELL PE 2950 | DELL PE 2950 | DELL PE R900 |
| Role | WFE | Application server | SQL Server-based server |
| Processor(CPUs) | 2pX4  (Xeon L5420@2.5GHz) | 2pX4  (Xeon L5420@2.5GHz) | 4pX4  (Xeon E7330@2.4GHz) |
| RAM(GB) | 16 | 16 | 32 |
| Operating System | Windows Server® 2008 R2 Enterprise | Windows Server 2008 R2 Enterprise | Windows Server 2008 R2 Datacenter |
| Authentication | NTLM | NTLM | NTLM |
| Storage: Operating System | 4x 146 GB, 10 K RPM, RAID 0 | 4x 146 GB, 10 K RPM, RAID 0 | 2x 146 GB, 15 K RPM, RAID 1 |
| Storage: Backups | -- | -- | 3x 300 GB, 15 K RPM, RAID 5 |
| Storage: SQL Server data | -- | -- | 9x 300 GB, 15 K RPM, RAID 5 |
| Storage: SQL Server logs | -- | -- | 6x 300 GB, 15 K RPM, RAID 5 |
| Number of instances of SQL Server | 0 | 0 | 1, SQL Server 2008 SP1 CU6 |
| Number of NICs | 1 | 1 | 4 |
| NIC speed | 1 GB | 1 GB | 1 GB |
| Load balancer type | NLB | NA | NA |
| ULS Logging level | Medium | Medium | Medium |
| Antivirus Settings | Microsoft Forefront | Microsoft Forefront | Microsoft Forefront |

### Lab software

The following table lists the specific software that was installed on the lab machines during testing.

|  |  |
| --- | --- |
|  | Description |
| Operating System | Windows Server 2008 R2 Enterprise Version 6.1.7600 |
| SQL Server Version | SQL Server 2008 Version 10.0.2531.0 |
| IIS | Version 7.5.7600.16385 |
| SharePoint Server | Microsoft SharePoint Server Build 14.0.4758.1000  **Note:** this is a pre-release version of SharePoint Server 2010 release (it is, however, very close to the final version of the product). |

### Topology

The following topologies were used to detect scale-out performance trends.

**1 WFE x 1 application server x 1 SQL Server data source (1x1x1)**



**1 WFE x 2 application servers x 1 SQL Server data source (2x1x1)**



**2 WFE x 2 application servers x 1 SQL Server data source (2x2x1)**



**3 WFE x 2 application servers x 1 SQL Server data source (3x2x1)**



## Test results

The following sections show the test results of Visio Services in SharePoint Server 2010.

After a calibration run that is used to determine the Green Zone and Red Zone user loads, the Rendering a Typical Mix of Visio Web Drawings test was run repeatedly, while changing only the topology to show its progressive effect on farm performance. The RPS that is reported in this document is the average RPS of a constant user load test.

Note that all the tests reported on in this article were conducted without think time, which is a natural delay between consecutive operations. In a real-world environment, each operation is followed by a delay as the user performs the next step in the task. By contrast, in this testing, each operation was immediately followed by the next operation, which resulted in a continual load on the farm. This load introduced database contention and other factors that can adversely affect performance.

For information about bottlenecks in Visio Services in SharePoint Server 2010, see the [Common bottlenecks and their causes](#bottlenecks) section later in this article.

### Overall scale

The following table summarizes the effect of adding WFE servers and application servers to the throughput of Visio Service.

|  |  |  |
| --- | --- | --- |
|  | Recommended (Green Zone)  Requests per second | Maximum (Red Zone)  Requests per second |
| 1x1x1 | 97.5 | 126 |
| 2x1x1 | 172.5 | 192 |
| 2x2x1 | 195 | 218 |
| 3x2x1 | 242 | 266 |

The following graph shows that the addition of WFE servers and application servers increases both Green Zone and Red Zone throughput linearly. By slope comparison, the graph of these values below also shows that the addition of a WFE server has a larger positive effect on the throughput than the addition of application servers. This indicates that WFE servers can be a bottleneck for Visio Services deployments

**Throughput increases more dramatically when WFE servers are added.**

**As topologies grow, throughput increases**

## Hardware cost per transaction

The following table highlights the hardware cost of running the Rendering a Typical Mix of Visio Web Drawings test across topologies in the Green Zone.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scorecard Dashboard | 1x1x1 | 2x1x1 | 2x2x1 | 3x2x1 |
| Average RPS  (requests/sec) | 97.5 | 172.5 | 195 | 242 |
| Average WFE server CPU  (%) | 59.5 | 54.5 | 59.8 | 50.2 |
| Average application server CPU  (%) | 16.4 | 27.9 | 17.25 | 21.5 |
| Failure rate | 0.006 | 0.006 | 0.01 | 0.02 |
| 90% QoS availability | 1 | 1 | 1 | 1 |
| 25th percentile latency  (seconds) | 0.05 | 0.05 | 0.05 | 0.06 |
| Percent time in garbage collector | WFE: 3.55  App: 6.6x10-7 | WFE: 3.11  App: 0.00014 | WFE: 3.27  App: 0.68 | WFE: 3.13  App: 0.125 |
| Number of WFE server crashes | 0 | 0 | 0 | 0 |
| Average memory used  (bytes) | WFE: 210,728,784  App: 210,172,592 | WFE: 664,374,336  App: 810,444,288 | WFE: 203,343,584  App: 229,627,536 | WFE: 199,905,600  App: 259,536,336 |
| Maximum memory used  (bytes) | WFE: 212,664,320  App: 211,578,880 | WFE: 719,638,528  App: 1,309,839,360 | WFE: 204,537,856  App: 229,969,920 | WFE: 200,081,408  App: 262,713,344 |

The following table highlights the hardware cost of running the Rendering a Typical Mix of Visio Web Drawings test across topologies in the Red Zone.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scorecard Dashboard | 1x1x1 | 2x1x1 | 2x2x1 | 3x2x1 |
| Average RPS  (requests/sec) | 124 | 190 | 216 | 264 |
| Average WFE CPU  (%) | 73.8 | 64 | 71.05 | 59.9 |
| Average application server CPU  (%) | 18.9 | 31 | 18.35 | 23.0 |
| Failure rate | 0.006 | 0.009 | 0.009 | 0.01 |
| 90% QoS availability | 1 | 1 | 1 | 1 |
| 25th percentile latency  (seconds) | 0.06 | 0.06 | 0.07 | 0.06 |
| Percent time in garbage collector | WFE: 0.000036  App: 0.000074 | WFE: 0.00036  App: 0.00014 | WFE: 1.54  App: 0.805 | WFE: 1.15  App: 0.4 |
| Number of WFE crashes | 0 | 0 | 0 | 0 |
| Average memory used  (bytes) | WFE: 631,852,288  App: 820,075,648 | WFE: 748,467,200  App: 884,640,512 | WFE: 659,872,256  App: 511,670,960 | WFE: 730,737,301  App: 827,111,104 |
| Maximum memory used  (bytes) | WFE: 708,333,568  App: 1,446,760,448 | WFE: 787,783,680  App: 1,350,569,984 | WFE: 709,833,600  App: 1,319,833,600 | WFE: 1,070,150,997  App: 1,450,207,232 |

## Recommendations

This section provides general performance and capacity recommendations. Use these recommendations to determine which hardware, topology and configuration settings best meet your particular capacity and performance needs.

### Hardware recommendations

The basic hardware requirements for Visio Services are the same as those for SharePoint Server 2010, which can be found [here](http://technet.microsoft.com/en-us/library/cc262485(office.14).aspx).

### Scaled-up and scaled-out topologies

To increase the capacity and performance of one of the starting-point topologies, you can do one of two things. You can either scale up by increasing the capacity of your existing server computers or scale out by adding additional servers to the topology.

As a general rule, Visio Services is especially sensitive to scaling-out. When planning to scale-out, use the following general rules:

1. When scaling out, using the configurations described earlier, favor increasing WFE servers to increasing application servers. If your computers have performance specifications that are comparable to the ones in this article, a ratio of 3 WFE servers to 1 application server is recommended.
2. Remember that scaling out enables both an increase in throughput and an increase in user load that Visio Services can serve. Use the following graph to gauge the expected throughput on your topology at a particular user load. Notice that throughput tapers off at a certain user load, after which requests are served with increase latency.

## Optimizations

The following sections discuss methods for improving farm performance by optimizing Service settings and working around common system bottlenecks.

### Settings optimizations

One way to control the performance characteristics of Visio Services is to change the values of its performance-related service settings. This following table discusses the qualitative effect on farm performance of varying these settings.

|  |  |  |  |
| --- | --- | --- | --- |
| Setting Location | Parameter | Description | Qualitative Impact on Performance |
| Central Administration | **Maximum Diagram Size** | The maximum size in MB of a Web drawing that can be rendered. | A larger size limit can lead to reduced throughput and increased latency.  A smaller limit can prevent more complex Web drawings from being rendered but has the opposite effect on performance. |
| **Minimum Cache Age** | The minimum number of minutes that a Web drawing is cached in memory.  This value is per user per drawing. The interval begins when a user views a Web drawing. That user cannot refresh that Web drawing until the interval expires. The interval begins for other users when they first view the Web drawing. | Smaller values allow for more frequent data refresh operations for users but increase CPU and memory usage on application servers, which reduces throughput and increases latency.  A larger value increases the refresh window but has the opposite effect on performance metrics. |
| **Maximum Cache Age** | The number of minutes after which cached Web drawings are purged. | Larger values decrease file I/O and CPU usage on application servers but increase memory usage on the server. Large values reduce latency for drawings that are rendered often.  A smaller value has the opposite effect on performance. |
| **Maximum Recalc Duration** | The number of seconds before data refresh operations time out. This applies only to data connected Web drawings. | Longer timeouts allow for more complex data-connected Web drawings to be recalculated but use more processing power, reduce throughput, and increase latency.  A shorter timeout decreases the complexity of the drawings that can be rendered but has the opposite effect of performance. |
| Web Part | **Force Raster** | Forces the Visio Web Access Web Part to render the drawing as a PNG even though Silverlight® might be installed. | Choosing to render all drawings using a raster format reduces drawing visual fidelity but increases throughput slightly.  Choosing to render all drawings in XAML increases visual fidelity but reduces throughput slightly. |

### Common bottlenecks and their causes

During performance testing, the following bottleneck was revealed. A bottleneck is a condition in which the capacity of a particular constituent of a farm is reached. This causes a plateau or decrease in farm throughput.

|  |  |  |
| --- | --- | --- |
| Bottleneck | Cause | Resolution |
| WFE CPU Utilization | Due to caching on the application server layer, the throughput of the Visio Services applications servers is larger than the front-end Web servers; this causes the WFE layer to be the bottleneck of the system. | Add more front-end Web servers to help reduce the negative effect of the WFE bottleneck on the overall throughput of Visio Services. |

### Performance monitoring

To help you determine when you have to scale your Visio Services deployment, you can use performance counters to monitor its health. In addition to the performance counters that are shipped with Windows Server and that measure overall server health, you can use the following list of performance counters to gain a deeper understanding of the specific performance behaviors of Visio Services.

### Visio Services performance counters

The following table lists performance counters that monitor key Visio Services application server metrics. All these counters are in the Visio Server: Visio Graphics Service performance counter category.

|  |  |  |
| --- | --- | --- |
| Logical Category | Counter | Measurement |
| Aggregate counters | Requests Received per Second | Count of requests received per second. |
| Request Processing time | Average processing time of a rendering request in milliseconds. |
| Drawing rendering pipeline detail | File Retrieval and Parsing Time | Average time, in milliseconds, to retrieve a drawing from the content database and to parse it. |
| Server Data Refresh Time | Average time, in milliseconds, to retrieve all external data for a data-connected drawing. |
| Server Data Binding Time | Average time, in milliseconds, to update a data-connected drawing. |
| Text Data Binding Time | Average time, in milliseconds, to update text within a data-connected drawing. |
| Server Rasterization time | Average processing, in milliseconds, to create a PNG representation of a drawing. |

### Visio Web Access performance counters

The following table lists performance counters that monitor key Visio Web Access server (WFE) metrics. All these counters are in the Visio Server: Visio Web Access performance counter category.

|  |  |  |
| --- | --- | --- |
| Logical Category | Counter | Measurement |
| Aggregate counters | Request Processing Time | Average time, in milliseconds, to process a drawing render request from arrival to delivery. |
| Requests Received per second | Count of requests received per second. |
| Requests with Errors per second | Average number of requests that are returned with errors per second. |
| Average Content Initial Transmission Time | Average transmission time, in milliseconds, of initial HTML plus Java script to client. |