

SmarterTools Inc.

# SmarterStats Deployments in Local and Distributed Networks

SmarterStats Architecture Recommendation

## Table of Contents

Introduction.....	4
Who Should Use This Document.....	4
Determining the Required Architecture.....	4
Calculating Disk Space Requirements.....	5
Understanding CPU Requirements .....	5
CPU Requirements for MRS .....	6
CPU Requirements for Collectors.....	6
Web Servers Using Remote Agents.....	6
Individual Websites (SmarterStats Free Edition) .....	7
Remote Web Server and Local SmarterStats Installation .....	7
Local Website and SmarterStats Installation .....	7
System Recommendations.....	8
Low-volume Deployment (One Web Server) .....	9
SmarterStats Professional or Enterprise Edition.....	9
System Recommendations.....	9
Medium-volume Deployments (Multiple Web Servers) .....	10
SmarterStats Professional with Shares .....	10
System Recommendations.....	10
SmarterStats Enterprise Edition.....	11
System Recommendations.....	12
High-volume and Specialized Deployments .....	13
SmarterStats Enterprise Edition with Servers Running in a Collector Role.....	13
System Recommendations.....	14
SmarterStats Enterprise Edition Hybrid .....	15
System Recommendations.....	16
Load-balanced Websites (Clustering) for High Traffic Websites .....	17
Session-based Load Balancing.....	17
Hit-based Load Balancing.....	18
Search Engine Optimization .....	19
SEO Processing .....	19
Summary .....	21

Glossary..... 21

Important Notice ..... 21

## Introduction

### Who Should Use This Document

This document is intended for use by all users of SmarterStats to help determine the most effective architecture to gather statistics on their websites and/or in hosted environments.

### Determining the Required Architecture

The authors have chosen to divide their recommendations into four categories: individual website deployments, low-volume deployments, medium-volume deployments, and high-volume and specialized deployments. For the purposes of these recommendations:

- **Individual website deployments** are those in which the Free edition of SmarterStats is utilized to gather statistics on one domain either locally or on a remote, hosted server.
- **Low-volume deployments** are those where a purchased edition of SmarterStats is reporting on up to 250 websites—based locally and/or on remote servers delivering log files to the SmarterStats server for analysis.
- **Medium-volume deployments** are those in which a purchased edition of SmarterStats is reporting on up to 2,000 websites from multiple remote Web servers and/or locally hosted websites.
- **High-volume and specialized deployments** are those in which a higher-level Enterprise edition of SmarterStats is reporting on tens of thousands of websites per Management Reporting Server (MRS) across potentially distributed networks in a variety of methods.

Additional information and recommendations on SmarterStats implementations in various environments is available at the SmarterTools website: [www.smartertools.com](http://www.smartertools.com)

If you have questions about which category may be the best one for your environment, contact [sales@smartertools.com](mailto:sales@smartertools.com) for additional information and assistance.

## Calculating Disk Space Requirements

SmarterStats uses advanced data storage techniques that enable it to store log files as relational files (called SmarterLogs) that are faster to query and much smaller than the original log files. On average, SmarterStats can store the same log data in 10% to 15% of the hard disk space required by the original log files.

SmarterStats can also export log data to the original log format, eliminating the need to keep raw log files for extended periods of time, resulting in a substantial savings in disk space usage. In most cases, customers actually free up hard drive space on their servers when using SmarterStats instead of taking additional space.

In distributed installations, the MRS is a Web reporting engine that only serves as a request/response engine. It is for this reason that an MRS can handle up to 30,000 websites in a single installation. The MRS itself does not store any imported log data (unless there are sites on the MRS being processed). Log data on distributed Web servers is contained on the remote service installations.

## Understanding CPU Requirements

SmarterStats was engineered to process statistics quickly, and to reduce CPU requirements when other processes are running. This allows SmarterStats to run on the Web server without negatively affecting the performance or delivery of the websites on the server.

If the server is dedicated to SmarterStats, or if the server is used strictly as a SmarterStats collector (described herein), the authors recommend allocating additional CPU capacity to the SmarterStats and/or collector processes to increase performance. For instructions on how to make this change, please refer to the Knowledge Base article [How To – Configure SmarterStats to Use More CPU to Process Logs](#).

Recommended CPU requirements for medium and high-volume deployments are included in this section. However, it should be noted that the following recommendations are estimates based on the average utilization by an average number of sites with an average number of hits, system administrators may need to adjust their CPU requirements based on their site and/or server needs.

### **CPU Requirements for MRS**

MRS fall into three different categories: Web servers reporting on up to 250 sites; high-volume MRS that do not act as collectors and process a maximum of 30,000 sites; and high-volume MRS acting as a collector for a maximum of 1,500 sites and reporting on a maximum of 30,000 total sites.

For Web servers reporting on up to 250 sites, the authors recommend a single-core 1.5 GHz processor.

For high-volume MRS that do not act as collectors and process a maximum of 30,000 sites, the authors recommend a single-core 1.5 GHz processor.

The authors also recommend a dual-core 2.5 GHz processor for high-volume MRS that act as a collector for up to 1,500 sites and report on a maximum of 30,000 total sites.

### **CPU Requirements for Collectors**

The authors recommend a single-core 1.5 GHz processor for collector servers that perform no other tasks besides processing log files from a maximum of 2,000 sites.

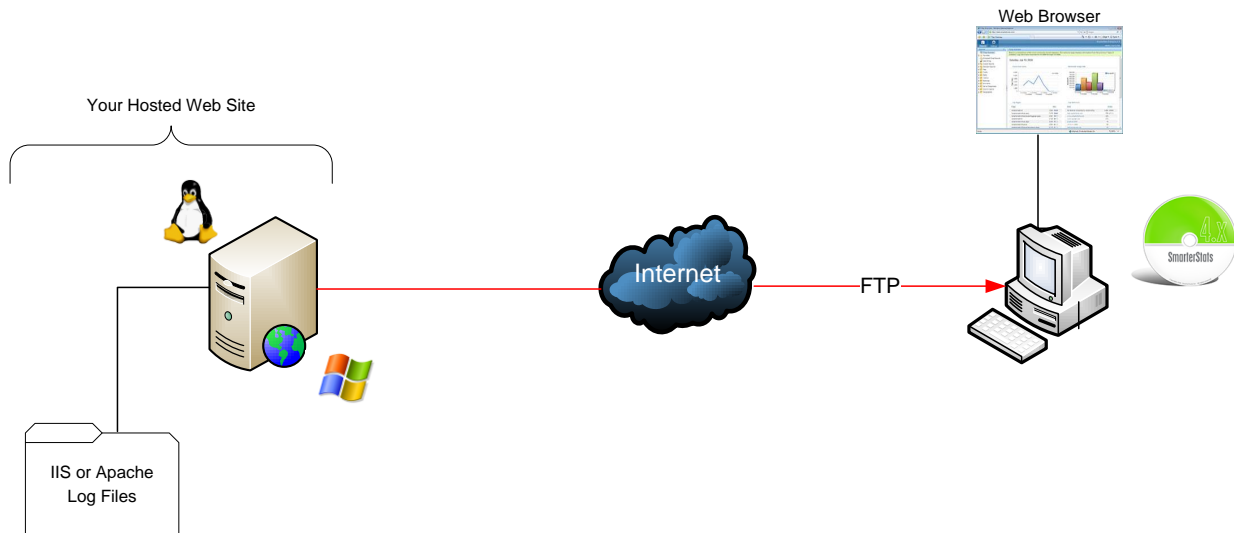
### **Web Servers Using Remote Agents**

The authors recommend a dual-core 2.5 GHz processor for individual Web servers running remote agents for up to 1,000 websites.

## Individual Websites (SmarterStats Free Edition)

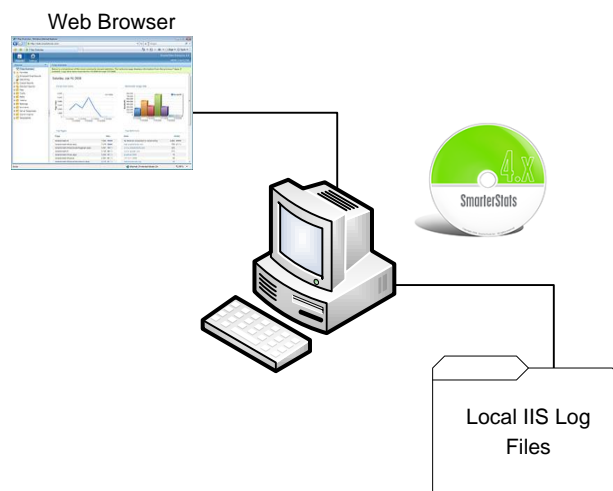
The free edition of SmarterStats can be used to collect statistical data for a single website. In this case, SmarterStats needs to be installed on a user's PC. In the scenario below, the user has a hosted website so the website's log files need to be downloaded via FTP and then imported into SmarterStats. Once the importing process is complete and the SmarterLog files are created, website traffic reports can be generated and viewed via the PC's Web browser.

## Remote Web Server and Local SmarterStats Installation



## Local Website and SmarterStats Installation

In this scenario, SmarterStats is installed on the same PC or server as the website. This eliminates the need to download the log file via FTP since the website log files are stored locally.



## System Recommendations

The authors recommend the following hardware for SmarterStats local machines (PCs):

- 512 MB RAM
- 100 MB hard disk space (only allows for SmarterStats, .NET Framework and small website logs)
- IIS 6\* or higher with the Microsoft .NET 4.0 Framework (with all applicable service packs and/or patches)
- Windows Operating System:
  - Windows XP (Home or Professional, SP2 recommended)
  - Windows 2003 or higher
  - Windows Vista
- The Web interface supports the following browsers:
  - Internet Explorer 5.5 and higher
  - Firefox 1.5 and higher
  - Opera 8.5 and higher
  - Mozilla 1.7 and higher
  - Google Chrome

\*SmarterStats includes a basic Web server so that the product is fully functional upon installation—even without the existence of IIS or other Web servers. Although not required in single-site and low-volume installations, it is recommended to install IIS 6 or higher in place of the SmarterStats Web server for increased performance and security.

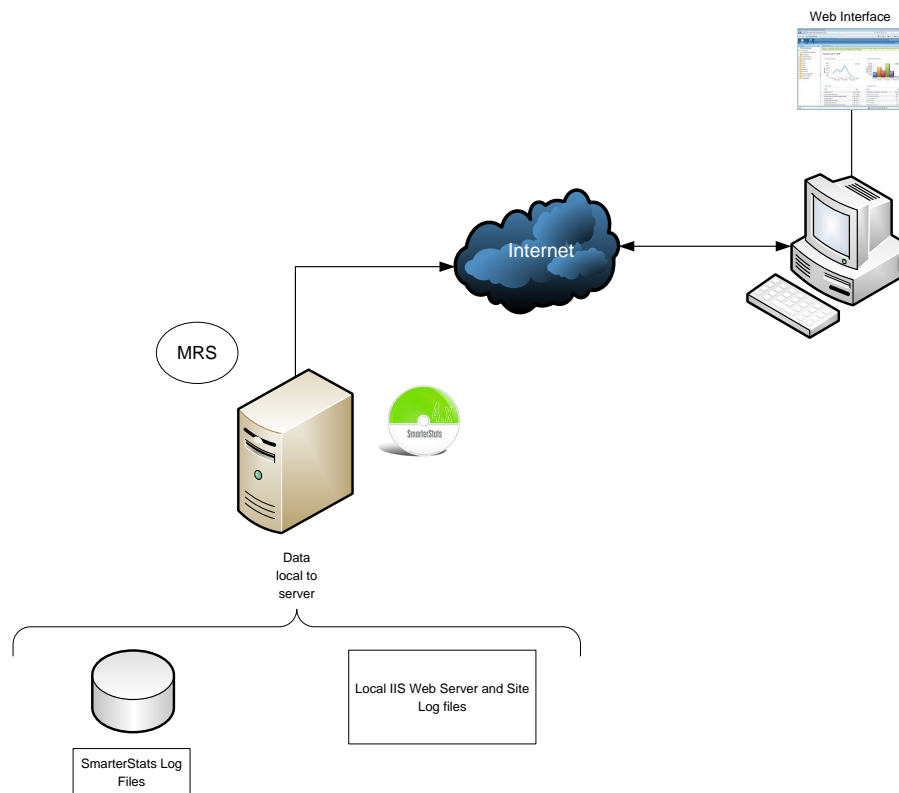


## Low-volume Deployment (One Web Server)

Low-volume deployments are those in which a purchased edition of SmarterStats is reporting on up to 250 websites—based locally and/or on remote servers delivering log files to the SmarterStats server for analysis.

### SmarterStats Professional or Enterprise Edition

SmarterStats is installed directly on the Web server where it can import the locally stored IIS log files.



### System Recommendations

The authors recommend the following hardware for SmarterStats local servers:

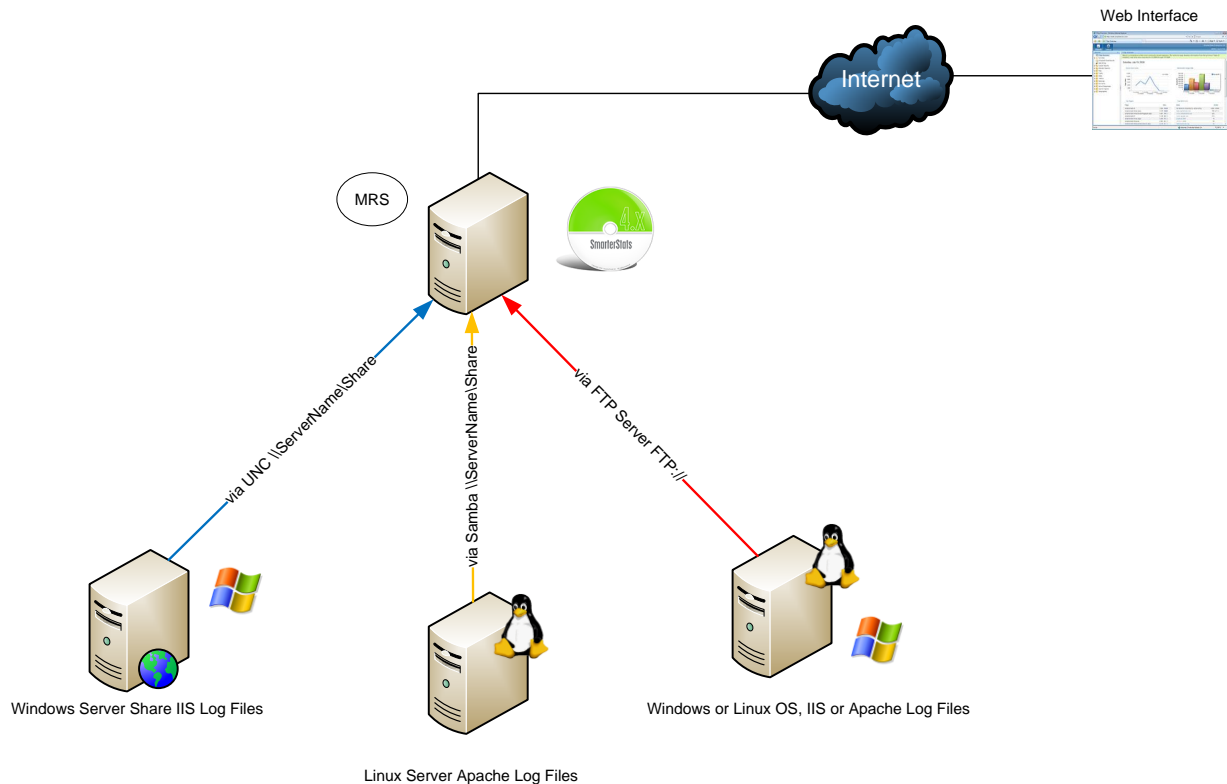
- Windows 2003 or higher
- 512 MB RAM
- 1/5 of the total disk space that would be required to store the raw LOG formats
- IIS 6 or higher

## Medium-volume Deployments (Multiple Web Servers)

Medium-volume deployments are those in which a purchased edition of SmarterStats is reporting on up to 2,000 websites from multiple remote Web servers and/or locally hosted websites.

### SmarterStats Professional with Shares

SmarterStats Professional can import website logs files from Windows IIS Web servers via UNC Shares or FTP server. In addition, SmarterStats can import log files from Linux Apache servers via Samba Shares and FTP server.



### System Recommendations

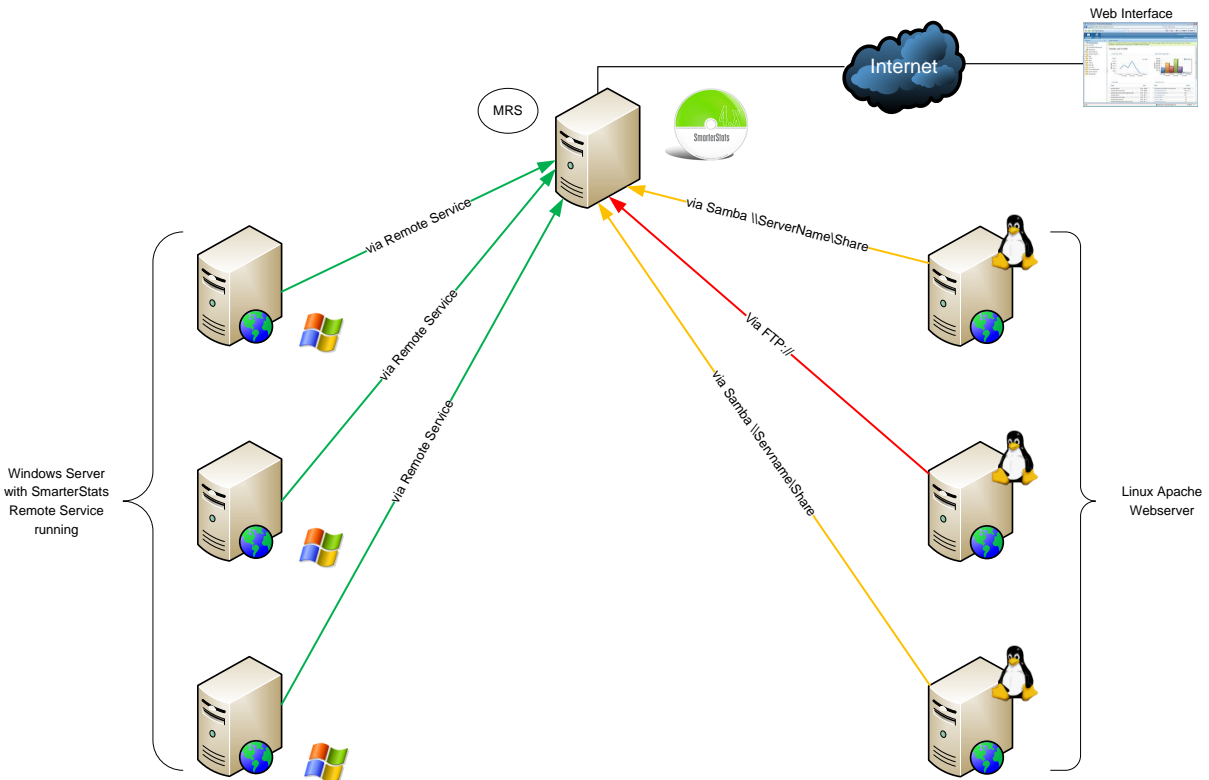
The authors recommend the following hardware for SmarterStats medium-volume distributed networks:

- Windows 2003 or higher
- 1 GB RAM
- 1/5 of the total disk space that would be required to store the raw LOG formats
- IIS 6 or higher

## SmarterStats Enterprise Edition

This configuration shows SmarterStats Enterprise edition's flexibility. The MRS imports log files from Linux Apache Web servers via Samba Shares and FTP servers. In addition, SmarterStats receives reports from the SmarterStats Remote Service running on Windows Web servers.

Running the SmarterStats Remote Service on the Windows Web server is the most efficient configuration in terms of speed and system utilization.



## System Recommendations

The authors recommend the following hardware for SmarterStats Enterprise medium-volume distributed networks:

### *MRS*

- Windows 2003 or higher
- 1 GB RAM
- 200 MB hard disk space
- IIS 6 or higher

### *Web Servers Using Remote Agents*

- 2 GB RAM
- 1/5 of the total disk space that would be required to store the raw LOG formats
- Microsoft .NET 4.0 Framework (including all applicable service packs and/or patches)
- Windows Operating System:
  - Windows XP
  - Windows 2003 or higher
  - Windows Vista

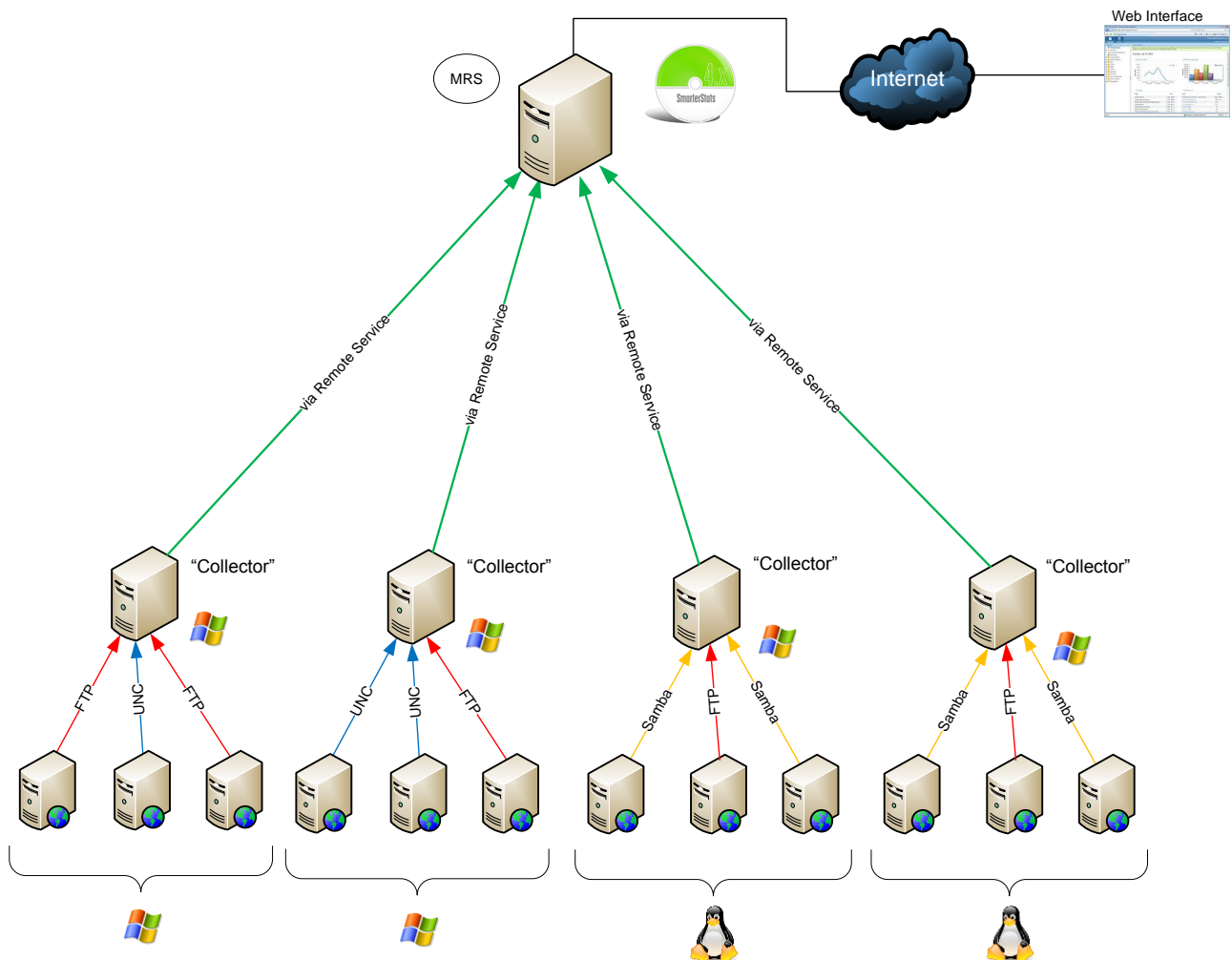
*Note: For ease of installation on Web servers and collectors, an automatable setup package containing just the remote service is available as a separate download.*

## High-volume and Specialized Deployments

### SmarterStats Enterprise Edition with Servers Running in a Collector Role

In this configuration, all collectors are Windows-based servers running the SmarterStats Remote Service. Each collector imports data from other Windows IIS Web servers via UNC Shares or FTP servers. Other collectors can import logs from Linux Apache Web servers via Samba Shares or FTP Server.

This method is highly efficient and all SmarterStats log files are created and stored locally. The MRS is only used as the customer user interface when reports are requested from the collectors and then displayed.



## System Recommendations

The authors recommend the following hardware for SmarterStats Enterprise high-volume collector networks:

### *MRS*

- Windows 2003 or higher
- 1 GB RAM
- 200 MB hard disk space
- IIS 6 or higher

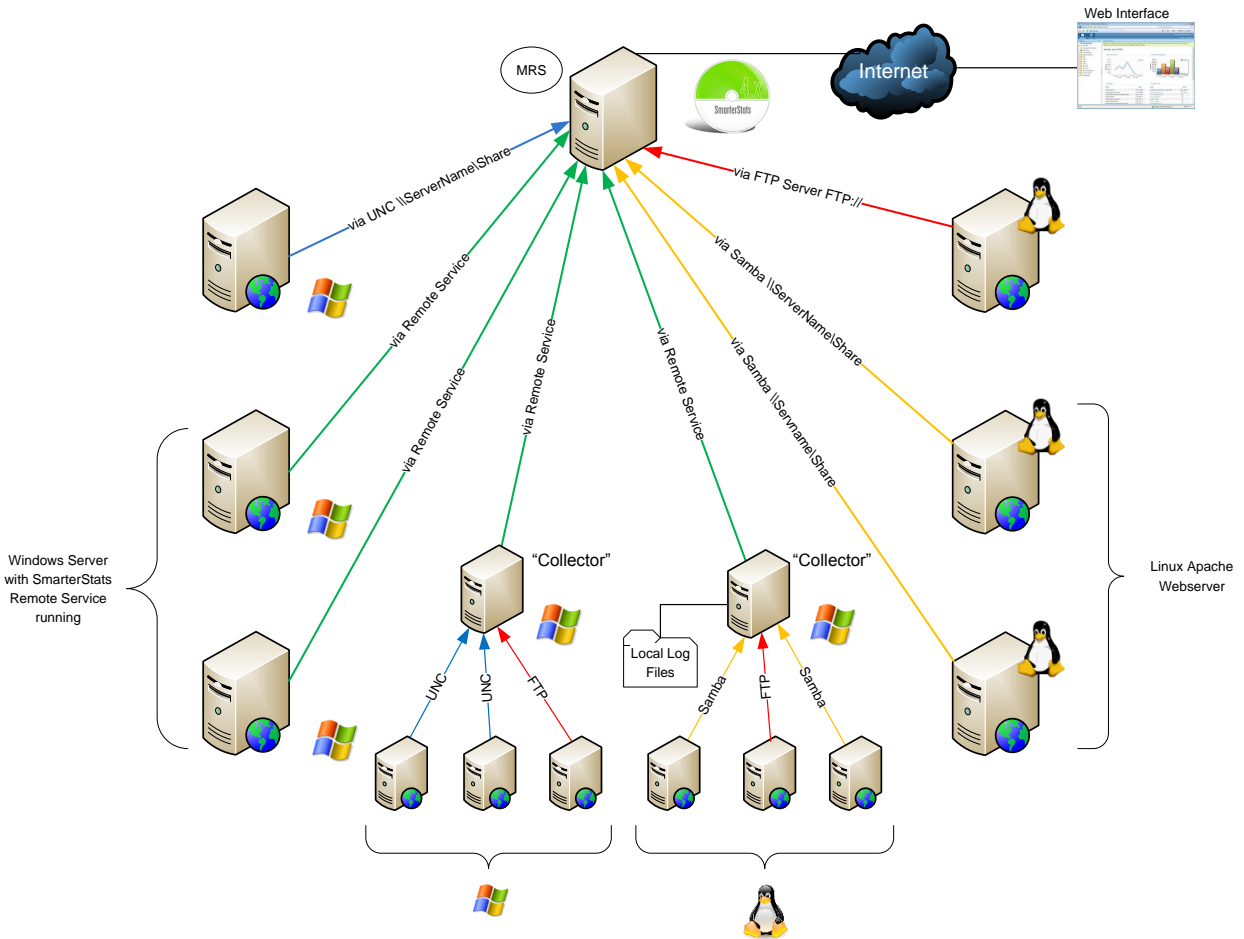
### *Collector Servers*

- 512 MB RAM
- 1/5 of the total disk space that would be required to store the raw LOG formats
- Microsoft .NET 4.0 Framework (including all applicable service packs and/or patches)
- Windows Operating System:
  - Windows XP
  - Windows 2003 or higher
  - Windows Vista

**Note:** For ease of installation on Web servers and collectors, an automatable setup package containing just the remote service is available as a separate download.

## SmarterStats Enterprise Edition Hybrid

This configuration demonstrates the power and flexibility of SmarterStats in its ability to import and provide website statistics from many different sources.



## System Recommendations

The authors recommend the following hardware for SmarterStats Enterprise high-volume networks with multiple data collection methods:

### *MRS*

- Windows 2003 or higher
- 1 GB RAM
- 200 MB hard disk space
- IIS 6 or higher

### *Collector Servers*

- 512 MB RAM
- 1/5 of the total disk space that would be required to store the raw LOG formats
- Microsoft .NET 4.0 Framework (including all applicable service packs and/or patches)
- Windows Operating System:
  - Windows XP
  - Windows 2003 or higher
  - Windows Vista

### *Web Servers Using Remote Agents*

- 2 GB RAM
- 1/5 of the total disk space that would be required to store the raw LOG formats
- Microsoft .NET 4.0 Framework (including all applicable service packs and/or patches)
- Windows Operating System:
  - Windows XP
  - Windows 2003 or higher
  - Windows Vista

**Note:** For ease of installation on Web servers and collectors, an automatable setup package containing just the remote service is available as a separate download.



## Load-balanced Websites (Clustering) for High Traffic Websites

A load-balanced (clustered) website is one in which a single website is distributed across more than one server. Another common term for this type of structure is "Web farm." This technique is utilized by very high-volume websites, websites that have particularly high up-time requirements, and/or other mission critical functions that require fail-over resilience exceeding the norm.

There are two common methods of load-balancing: session-based and hit-based. Each method requires following different steps to ensure SmarterStats analyzes the data correctly. *Note: The load-balanced deployment is recommended for users of SmarterStats 4.x and earlier. Because the latest versions of SmarterStats allow importing from multiple log sources, this deployment solution does not apply to users of SmarterStats 5.x and later.*

### Session-based Load Balancing

Session-based load balancers attempt to keep all traffic for a specific visitor on the same server. This is known as "persistence" or "stickiness."

The following steps explain how to copy and rename logs from per-session load-balanced servers to a single server so that SmarterStats can import them. You can alter or adjust these steps to accommodate particular requirements as needed.

On each of the load-balanced servers, do the following:

- Make a new directory on the C: drive called C:\SmarterStatsLogFileMerge
- Create a new batch file using notepad called DoLogFileMerge.bat that contains the code below (all one line).

```
XCOPY "C:\WINDOWS\system32\Logfiles\W3SVC1\*.LOG" " \\SmarterStats\Logs\Site1\*.LOG_A " /D /Y
```

*Note 1: This code assumes that your original logs on this server are contained at C:\WINDOWS\system32\Logfiles\W3SVC1, that SmarterStats is running on a remote server called file://smarterstats/, and that the share \Logs\Site1 is where the logs should be copied to on that server.*

*Note 2: The file mask \*.LOG\_A below should be different for each server. For example, other servers may be \*.LOG\_B, \*.LOG\_C, etc.*

- Save the file and exit Notepad.
- Go to Start -> Programs -> Accessories -> System Tools -> Scheduled Tasks
- Add a new scheduled task.
- Click Browse and click the newly created batch file.
- Choose the frequency of the copy you want. Most will want to set this to Daily, starting around 1:00 AM.
- Enter a username and password for a user that has permission to copy to the share you made.
- Save the scheduled task.

- Test the scheduled task by right-clicking it and choosing Run. After it runs, ensure that the logs were successfully copied.

### Hit-based Load Balancing

A hit-based load balancer distributes all requests evenly across its websites. For example, a visitor requesting a Web page may receive the HTML from one server and images from one or more other servers. This is the best performing style of load balancing, but it results in stats that are difficult to track because of time differences and delays in logging the data.

If your website is load balanced and separate log storage locations appear on different servers, you will need to manually combine the logs so that SmarterStats can treat them as one website. To do this, you will need a third-party log combining tool, of which many are available and can be found by searching Google.

Additional information regarding this process and aggregation tools may be available within the Knowledge Base and/or Community Forums accessible through the [SmarterTools Portal](#).

## Search Engine Optimization

Users of SmarterStats 5.x and later benefit from the software's search engine optimization (SEO) tools, which help SEO analysts and webmasters monitor SEO campaigns and enhance optimization efforts. SEO processing is performed by the SmarterStats service and each service is limited in the amount of processing it can complete each day.

### SEO Processing

In order to retrieve SEO statistics, SmarterStats needs to retrieve search pages from the supported search engines. In order to avoid getting blocked by search engines, page retrievals are limited to one page every 10 seconds. For this reason, the number of page retrievals SmarterStats can do for a specific search engine is limited to six per minute (or 360 per hour or 8,640 per day).

In addition, the number of page retrievals is different for each type of SEO statistic:

- Each keyword added for a site requires 10 page retrievals.
- Each competitor added for a site requires two page retrievals for visibility statistics and one page retrieval for the Google PageRank statistics, resulting in three total page retrievals per competitor.
- Each SEO collection also tracks the user's site, resulting in two page retrievals for visibility statistics and one page retrieval for the Google PageRank statistics (or a total of three page retrievals).

For example, let's assume a SmarterStats site administrator adds two SEO campaigns. The first campaign has four keywords and three competitor sites. The second campaign has six keywords and two competitor sites. The number of page retrievals required is calculated as follows:

#### Campaign #1:

4 keywords	X	10 page retrievals per keyword	=	40 keyword page retrievals
3 competitor sites	X	3 page retrievals per competitor	=	9 competitor page retrievals
1 local site	X	3 page retrievals	=	3 local site page retrievals
				<hr/>
				52 total page retrievals

At 10 seconds per page retrieval, it would take 8 minutes and 40 seconds to process the SEO statistics for this campaign.

**Campaign #2:**

6 keywords	X	10 page retrievals per keyword	=	60 keyword page retrievals
2 competitor sites	X	3 page retrievals per competitor	=	6 competitor page retrievals
1 local site	X	3 page retrievals	=	3 local site page retrievals
				<hr/>
				69 total page retrievals

At 10 seconds per page retrieval, it would take 11 minutes and 30 seconds to process the SEO statistics for this campaign.

In this example, it would take 20 minutes and 10 seconds to process the SEO statistics for the site.

By default, sites are limited to five competitors and five keywords, as a large number of keywords and competitors for a single site could severely limit the processing time available for other sites. SmarterStats installations with a large number of sites should keep these limits low. However, installations with only a few sites may wish to greatly increase these limits.

As a final note to SmarterStats Enterprise users: Because the processing is done by the SmarterStats service (not the website), an Enterprise version of SmarterStats with 30 remote servers can support 30 times as much SEO processing as a single SmarterStats server.

## Summary

The proper configuration and system architecture outlined in this document will provide a solid, reliable foundation. Because variations exist due to different volumes and client needs, the authors suggest starting with these recommendations and then adjusting server proportions, limits, and specifications based on the usage patterns that result.

## Glossary

The following is a list of acronyms used in this document:

<b>BIND</b>	Berkeley Internet Name Domain	<b>MB</b>	Megabytes
<b>CPU</b>	Central Processing Unit	<b>MHz</b>	Megahertz
<b>FTP</b>	File Transfer Protocol	<b>MRS</b>	Management Reporting Server
<b>GB</b>	Gigabyte	<b>RAID</b>	Redundant Array of Independent Disks
<b>IIS</b>	Internet Information Services (Server)	<b>RAM</b>	Random Access Memory
<b>IP</b>	Internet Protocol	<b>SAMBA</b>	Server Message Block (also: SaMBa)
<b>ISP</b>	Internet Service Provider	<b>SATA</b>	Serial Advanced Technology Attachment
<b>PC</b>	Personal Computer	<b>UNC</b>	Universal Naming Convention

## Important Notice

The recommendations in this document reflect the opinions of its author(s) only and are based upon their knowledge and experience. No warrantee or guarantee is expressed or implied as to the efficacy or viability of the information in this document for a particular environment or application.

SmarterStats, SmarterTools, and the respective logos are trademarks of SmarterTools Inc. All other trade-names and/or trademarks in this document are the property of their respective owners.

Comments and suggestions regarding the contents of this document should be sent to [sales@smartertools.com](mailto:sales@smartertools.com).