



DOCUMENTS 7

System Configuration Guide

September 2013 – ReadSoft AB

Copyright ReadSoft AB (publ). All rights reserved. The contents of this document are subject to change without notice. ReadSoft is a registered trademark of ReadSoft AB. Other product and company names herein may be the trademarks or registered trademarks of their respective owners.

Questions or comments about this document may be emailed to documentation@readsoft.com.

05092013.0132

Contents

About this guide	4
Related documents	4
System setup	5
Components of a DOCUMENTS solution	5
Overall setup	6
Databases	6
Servers	7
Workstations	7
System requirements	9
DOCUMENTS application servers and workstations.....	9
Database server	13
License server	13
Web Verify server (optional)	14
Network products	16
Virtual environments	16
Scanners.....	16
Network requirements.....	17
Using switched Ethernet.....	17
Bandwidth.....	18
Network latency.....	18
Image management and distributed solutions / remote scanning	19
Running Production tasks as services	22
Thin-client requirements	23
System-wide specifications	23

About this guide

This guide is for people who design and install a DOCUMENTS 7 system. The guide helps you to plan the overall DOCUMENTS solution.

You must have in-depth knowledge of DOCUMENTS to use this guide. It contains information about:

- Hardware and software requirements
- Network bandwidth requirements
- Thin-client guidelines
- Solution design questions and answers

Related documents

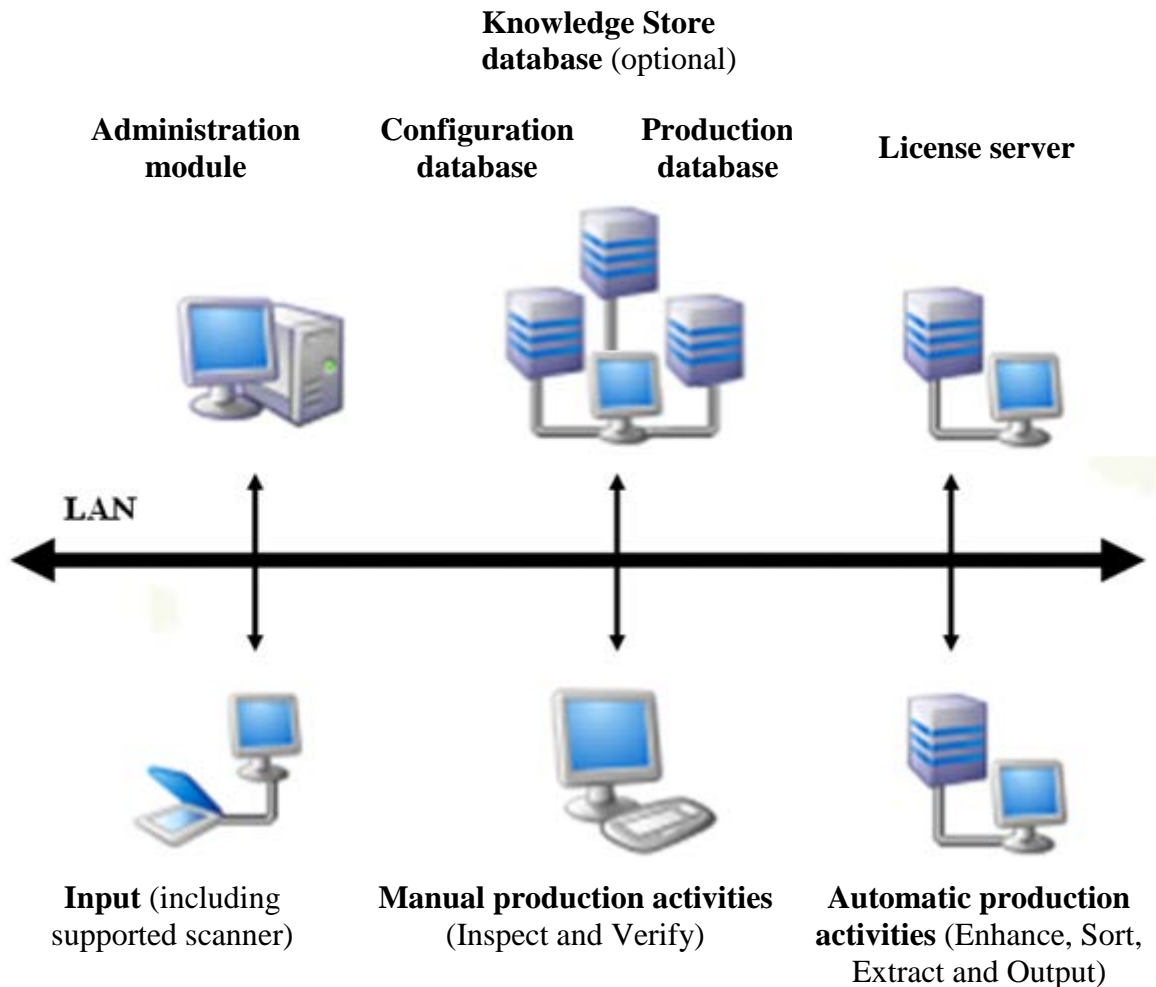
- *DOCUMENTS Installation Guide*
- *DOCUMENTS 7 Upgrade Guide*
- *Installing and Configuring Microsoft SQL Server For Use With DOCUMENTS 7*

All related documents are available in the `Documentation` folder on the installation CD.

System setup

A DOCUMENTS system is a specific setup of DOCUMENTS modules and related software, hardware, network, and servers that work together to process documents for a specific customer.

Components of a DOCUMENTS solution



Overall setup

- Administration module
- Configuration and Production area databases
- Knowledge store database (optional)
- License server, including HASP license manager
- Workstations for:
 - Input, including supported scanners
 - Manual production activities
 - Automatic production activities, run as Windows services
- System-wide configuration files
- Other common data such as output data, image and log files, audit trail data, and process statistics

Databases

ReadSoft recommends that the DOCUMENTS databases reside on a dedicated server. Since the databases are critical to the entire document-processing system, running other resource-demanding programs on the same machine can have a major impact on the performance of the DOCUMENTS solution.

See also: “Is it OK to simultaneously run other programs along with DOCUMENTS on the same workstation?” on page 7. The performance and compatibility issues discussed there also apply to the database server.

If knowledge store is used, the knowledge store database can reside on the same server as the DOCUMENTS databases, or it can reside on a different server.

Network traffic

Network traffic occurs in DOCUMENTS solutions where data is validated and exchanged as part of the process. In these solutions, the program is tightly integrated with external systems or third-party connectors.

Network traffic depends on how many Production modules you run simultaneously and how many documents you process. Consider that each module communicates with the database server continuously during production.

Servers

These servers are part of a DOCUMENTS solution recommended for solutions with medium to high processing requirements:

- Application servers – for running tasks/activities as Windows services.
- License server – for HASP, the license manager used by DOCUMENTS. The HASP license manager does not require much processing power and can often be run on the database server or the application server.
- Optional: Web Verify server.

Workstations

A workstation must be capable of running the task(s) that will be assigned to that station. A typical solution spreads the workload over several computers, having separate computers for each activity or group of activities. As a result, the different computers will have different performance requirements.

In addition to these requirements, a workstation used for scanning in the Input activity needs to have a supported scanner attached. The varying requirements for administration activities, manual activities, and automatic activities run as a service can be seen in the table in the next section.

Is it OK to simultaneously run other programs along with DOCUMENTS on the same workstation?

Two types of problems can occur when multiple programs are installed on the same workstation: problems related to performance and problems related to compatibility.

Performance issues

If other programs are running at the same time as ReadSoft software, it is important that the workstation or server hardware configuration meets the total requirements for ReadSoft software as well as all the other programs.

Typical programs that can cause performance problems are virus protection programs with online scanning of each file, backup software running in the background simultaneously with production, and entertainment programs such as web radio.

Any kind of background activity should be avoided, especially activities that require heavy usage of processing capacity and that move large amounts of data.

Compatibility issues

Every DOCUMENTS installation must be tested to detect potential compatibility conflicts with other programs. Some rules to follow:

- Avoid installing other programs that have the same or similar functionality as the DOCUMENTS modules installed on the computer. For example, other scanning software must not be installed on the workstation where documents are scanned using DOCUMENTS.
- Avoid running other programs at the same time as DOCUMENTS.

System requirements

DOCUMENTS application servers and workstations

DOCUMENTS is normally installed in a local area network (LAN) environment. On a LAN, individual activities or groups of activities are run on separate workstations or on application servers.

The Administration and Production modules can run as programs on PC workstations. In addition to meeting workstation specifications, scanning in the Input activity requires that a supported scanner is connected to the workstation.

Please note that the numbers and figures for workstations and application servers found on the following pages are the *minimum* recommended requirements for the platform. Fast processors with more RAM result in improved performance and are therefore recommended.

To assure production quality and speed, the requirements listed below need to be met. ReadSoft recommends using dedicated databases and servers (that is, no other programs should be used while DOCUMENTS is running). If (contrary to the recommendation) other programs are run at the same time on the same computer, the requirements for each additional product must be added to the requirements listed here.

Operating systems

DOCUMENTS can run on:

- Windows 8 (32-bit & 64-bit)
- Windows Server 2012
- Microsoft Windows 7 SP1 (32-bit & 64-bit)
- Microsoft Windows XP Professional (32-bit & 64-bit)
- Microsoft Windows Server 2003 & 2003 R2 (32-bit & 64-bit)
- Microsoft Windows Server 2008 & 2008 R2 (64-bit)

The operating system should be updated to the latest ReadSoft-approved service pack, which is listed in the DOCUMENTS 7 Fact Sheet.

Important

- Microsoft .NET Framework 3.5 SP1 is required. See page 12 for details.
- In order to install DOCUMENTS and its databases you need to be logged on as an administrator.

Because of its built-in firewall, Windows 7 and Windows XP may prevent some ReadSoft products, or products with which they are integrated, from working properly. If you encounter problems, please see *Using Windows XP Service Pack 2 with ReadSoft Products*, available in the `Documentation` folder on the DOCUMENTS installation CD.

DOCUMENTS application servers

Minimum recommended requirements for an application server (Normally used for automatic activities such as Enhance, Sort, Extract, and Output)	
CPU	Quad core. Any brand PC with an Intel or AMD processor is acceptable. Hyper threading is supported.
RAM	Min. 1 GB per core (min. 4 GB per application server).
Disk space	4 GB, with an additional 1GB per extra core for running large operations on multicore servers.
USB port	The server running the license service needs to have one free USB port if a hardware key is used. If a software key is used, this is unnecessary.

DOCUMENTS workstations

Minimum recommended requirements for workstations (Normally used for manual activities such as system administration, Scan, Inspect, and Verify)	
CPU	Dual core. Any brand PC with an Intel or AMD processor is acceptable. Hyper threading is supported.
Display	19" display with 1600 x 1200 resolution or better.
RAM	2 GB
Disk space	2 GB

Minimum recommended requirements for workstations

(Normally used for manual activities such as system administration, Scan, Inspect, and Verify)

Scanner interfaces

Each workstation where documents are to be scanned requires one of the following scanner interfaces, depending on which scanner is chosen:

Kofax and SCSI cards

- Please refer to the DOCUMENTS 7 Fact Sheet, available from ReadSoft Marketing, for the latest list of supported cards and drivers.

FireWire

- Certain scanners require FireWire, for example Kodak scanners in the i200 and i600 series.

USB 2.0

- Certain scanners have an additional USB port, for example Kodak i30/i40, and Canon 2580C.

For information about installing scanner interfaces, see the instructions supplied with the scanner.

Drivers for scanner interfaces

Each workstation where documents are to be scanned requires drivers for the scanner interface that will be used:

Kofax

- Image Controls 3.65 or later
- VRS 4.2 or later. See the help topic “Using VRS with DOCUMENTS” in *DOCUMENTS Help*.

SCSI cards

- ISIS or ISIS-compatible driver

FireWire and USB 2.0

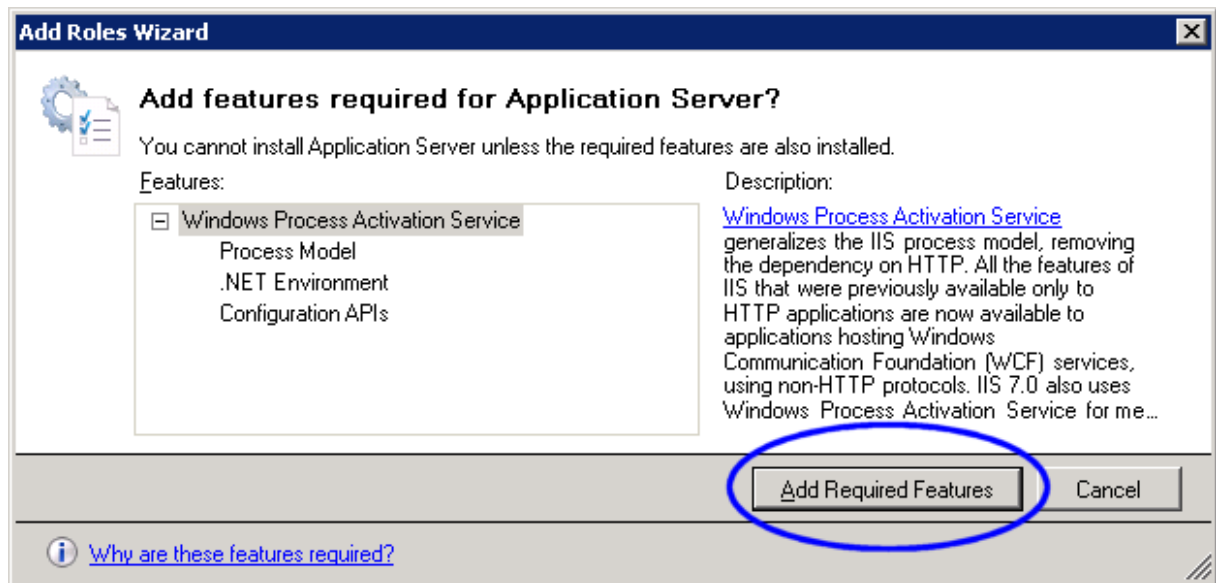
- ISIS or ISIS-compatible driver. The drivers that are included with scanner manufacturers’ standard installation are supported.

The workstation used for scanning must not have any other scanning software installed.

.NET Framework 3.5 SP1 is required

On Windows XP, Windows 7, and Windows 2003 .NET Framework 3.5 Service Pack 1 is installed automatically. On Windows Server 2008 or Windows Server 2008 R2, use this procedure:

1. Select **Start** menu > **Programs** > **Administrative Tools** > **Server Manager**.
2. Select **Add Roles** > Select **Application Server**.



3. Click **Add Required Features**.
4. Click **Next** several times, until the wizard is completed.

Database server

Recommended minimum configuration for the database server	
Database server	<p>The configuration data, production data, and knowledge store data for DOCUMENTS can reside in:</p> <ul style="list-style-type: none"> ▪ Microsoft SQL Server 2012 Standard Edition ▪ Microsoft SQL Server 2008 R2 ▪ Microsoft SQL Server 2008¹ <p>The operating system should be updated to the latest ReadSoft-approved service pack, listed in the DOCUMENTS 7 Fact Sheet.</p>
Operating system	Refer to the recommendations from the database server's manufacturer.
Processor speed	Refer to the recommendations from the database server's manufacturer.
Display	Refer to the recommendations from the database server's manufacturer.
RAM	Refer to the recommendations from the database server's manufacturer.
Disk space	<p>This is determined by a number of factors, including the amount of information that needs to be stored and available at any given time, and how many black and white (ca 60 kB per image) versus color images (500 kB-1MB per image) are stored.</p> <p>Standard hard drive capacities are sufficient unless storing many documents containing a large amount of data.</p>

License server

The server running the license service needs to have one free USB port if a hardware key is used. This is a service that can run anywhere on the network. If a software key is used, this is unnecessary.

¹ Workstations running on Windows Server 2008 cannot connect to a Microsoft SQL Server 2008 database server. When you try to connect to the named instance of Microsoft SQL Server, the connection fails and you receive an error such as "Specified SQL server not found" or "Error Locating Server/Instance Specified." For more information and suggested workarounds, please see [Microsoft Knowledge Base article 944390](#).

Web Verify server (optional)

Benefit

Web Verify is an alternative to using Invoice Verify to verify invoices. Web Verify makes it less expensive to roll out and maintain Verify clients. Larger IT departments normally use deployment tools to manage roll-out and maintenance of thick clients. However, maintaining this type of a solution (for example a Citrix solution) is usually more costly.

Tip: *DOCUMENTS Installation Guide* contains a technical overview of Web Verify.

What the customer needs to know

Customers who intend to use DOCUMENTS for invoice processing must be informed about the correct way of using Web Verify:

- *Web Verify does not replace the thick Verify client.* It does not have all the functionality or exactly the same functionality. It is a complement to the thick client and is best suited for simpler verification tasks such as:
 - Verification of header fields
 - Verification of line items for learned suppliers
- It is suited to provide extra verification in peak periods. Extra users simply log on to the server and start processing.
- Thick clients are still needed to manage new suppliers, for advanced optimization of suppliers, and to reroute invoices that are not connected to the correct profile.

Known issues and limitations

Several issues have been identified in connection with the current Web Verify architecture:

- Limitation in the number of concurrent users on one Web Verify server. At best, 20 users may concurrently use a Web Verify client. With Windows Vista and Windows 7, you are limited to 10 concurrent users.
 - Verify Inbox configuration is affected by this limitation, because Web Verify uses the Inbox. You will have less flexibility compared to the Inbox in a thick-client setup. Recommendations can be found in the *DOCUMENTS Help* topic “Using Web Verify to verify invoices.”
 - This limits cost effectiveness of the solution, since more servers must be added when invoice volume is higher and more verifiers are needed. On the other hand, 20 users is a lot of Verify operators. They should be able to verify up to 1 million invoices per year.

- Concurrency in itself is noticeable:
 - Invoices become locked when users access them, and it is not always clear to other users what is happening. For this reason, **Keep batch together** is not recommended (since an open invoice in one batch locks all invoices in that batch). Consequently, *Inspect* is done *once*, and invoices are not rerouted back for re-separation.

Since DOCUMENTS is a complex and highly configurable product, not all scenarios are tested or analyzed. Additional locking issues will likely be discovered.
 - Each action request by a user is queued in the Web Verify server. When many users are working at the same time, some actions take several seconds or even longer. This can contribute to a negative user experience.

- Web Verify does not contain all the functionality of Invoice Verify:
 - Web Verify events are not sent to REPORTER. Therefore, systems that include REPORTER normally do not include Web Verify.
 - It is not possible to handle exceptions, such as **Set last page** and **Reroute**. The invoice must be rejected so it can be handled using a full Invoice Verify client.
 - You can edit field values, and field input is validated, but a full validation of each invoice is not done until you click **OK**.
 - Image operations that are performed locally, such as rotating an image, are not saved to the database or to the output.
 - The Web Verify invoice list cannot display more than 1024 invoices at a time. If there are more invoices than that, you will not be able to see them.
 - At the time of this writing, you cannot “learn” table field values automatically as described in the *DOCUMENTS Help* topic “Generating table field values in Verify.”
 - In the table view in Web Verify, there are no extra tabs for tax tables and bank account tables. You need a thick Verify client to verify these types of tables.
 - Views are currently not configurable. Field groupings are fixed.

Supported operating systems

The Web Verify server runs on these operating systems:

Windows Vista*

Windows Server 2008

Windows 7*

Windows Server 2008 R2

*Max. 10 concurrent users.

Network products

DOCUMENTS can be installed on Microsoft Windows compatible networks.

Virtual environments

DOCUMENTS can be run in Citrix, Microsoft Hyper-V, and VMware ESX virtual environments.

Scanners

For a list of certified scanners and information about setup for specific scanners, please visit the ReadSoft Customer Web or Partner Web. ReadSoft employees can access the scanner list [here](#).

Network requirements

This section provides information about the network requirements for DOCUMENTS solutions that are installed in a LAN environment.

Using switched Ethernet

All DOCUMENTS workstations share one small, crowded pathway to the server. Since they all wait for a response from the server many times, the wait must be minimized.

In an unswitched network, all packets go to all destinations. In a switched network, packets go to only their specified destination. The return on investment greatly outweighs the cost of a switched network. Collisions no longer occur since packets have a regulated path to their destination.

Ethernet principles

An Ethernet workstation can send data packets only when no other packets are traveling on the network. If multiple stations sense an opening and start sending at the same time, a "collision" occurs. Each station then waits a random amount of time and then tries to resend its packet. After 16 consecutive failed attempts, an application trying to send a packet must start again.

In the case of DOCUMENTS, the application reports that contact has been interrupted and that the program must be restarted. As the number of collisions increases, errors and subsequent retransmissions grow quickly, compounding the problem.

Fast Ethernet communication means that the greater bandwidth allows for more and larger packets that can travel faster, but collisions still occur. When more than 50% of the network is used, congestion begins. Too many collisions slow the network and packets are lost. When more than 60% of the network is used, the network can grind to a halt. The solution is to use switched Ethernet.

Tips for configuration

- Connect servers and workstations to the same switch to prevent bottlenecks and errors across switch-to-switch links.
- Use high-quality hardware and cables.

Bandwidth

DOCUMENTS runs best on a dedicated network with a minimum bandwidth of 10 Mbps (megabits per second). A slower network or additional network traffic can have a significant negative impact on performance, while a bandwidth of up to 1 Gbps can greatly increase overall performance. This bandwidth is for the entire DOCUMENTS solution, database server, and image archiving. The greatest amount of bandwidth is used during input and output.

If bandwidth is an issue, there are various steps that can be taken to reduce network traffic, depending on your setup and possibilities. These include:

- Copying add-ons (extensions, or plug-ins) to client PCs, with logging done locally.
- Adjusting how to store images, as described in the *DOCUMENTS Help* topic “Specifying how to store images.”
- Adjusting image quality and file size, as described in the *DOCUMENTS Help* topic “Setting the image quality and file size of scanned images.”
- Adjusting threshold settings for color images, as described in the *DOCUMENTS Help* topic “Specifying threshold settings for color images.”
- Using small display images during production, as described in the *DOCUMENTS Help* topic “Creating display images for Production operators.”

Network latency

Latency is the time it takes for a packet of data to go from a program in one network node to a program in another node. Latency is not a measure of the network bandwidth; it is a measure of the delay time in transporting data from one node to another.

Latency is normally low in a LAN environment. Latency is normally high, or very high, when a wide area network (WAN) or the internet is included in the network.

The network latency in a DOCUMENTS solution must be low enough to allow the solution components to interact with each other without introducing unacceptable delays.

In general, you should not attempt to install a DOCUMENTS solution when network latency is higher than that of a normal LAN.

Customers who require a wide geographical distribution of DOCUMENTS components should consider other solution designs such as those using thin clients.

If you are having network latency problems, such as database communication timeouts for example, you can adjust the database timeout settings in the **Runtime systems** global policy.

Image management and distributed solutions / remote scanning

As the entire production process starts with and revolves around image files, image management is an important consideration when setting up your system and planning your solutions. A number of settings and features in DOCUMENTS allow you to optimize image management, reducing the amount of image data sent across your network.

For information about image location and image quality, please refer to the *DOCUMENTS Help* topic “Specifying how to store images.” Another related topic: “Remote scanning, or scanning from different workstations.”

DOCUMENTS can create several variations of each image, and access only the one needed for a certain activity. Using several images, each created for a specific purpose, increases production efficiency, particularly when sending images over a network with limited bandwidth. The different types of images are:

- **The original file** – the scanned image, which is very large if in color.
- **Thumbnails** – automatically created for every image during Input. Used during Inspect and Verify.
- **Threshold image** – a black-and-white image, created from a color original used to aid classification and extraction during Sort and Extract.
- **Color plane image** – created from the different colors in a color original, used in Sort and Extract.
- **Display image** – a compressed version of the original seen by production operators in Inspect and Verify.
- **Barcode image** – an image that is optimized for barcode recognition.

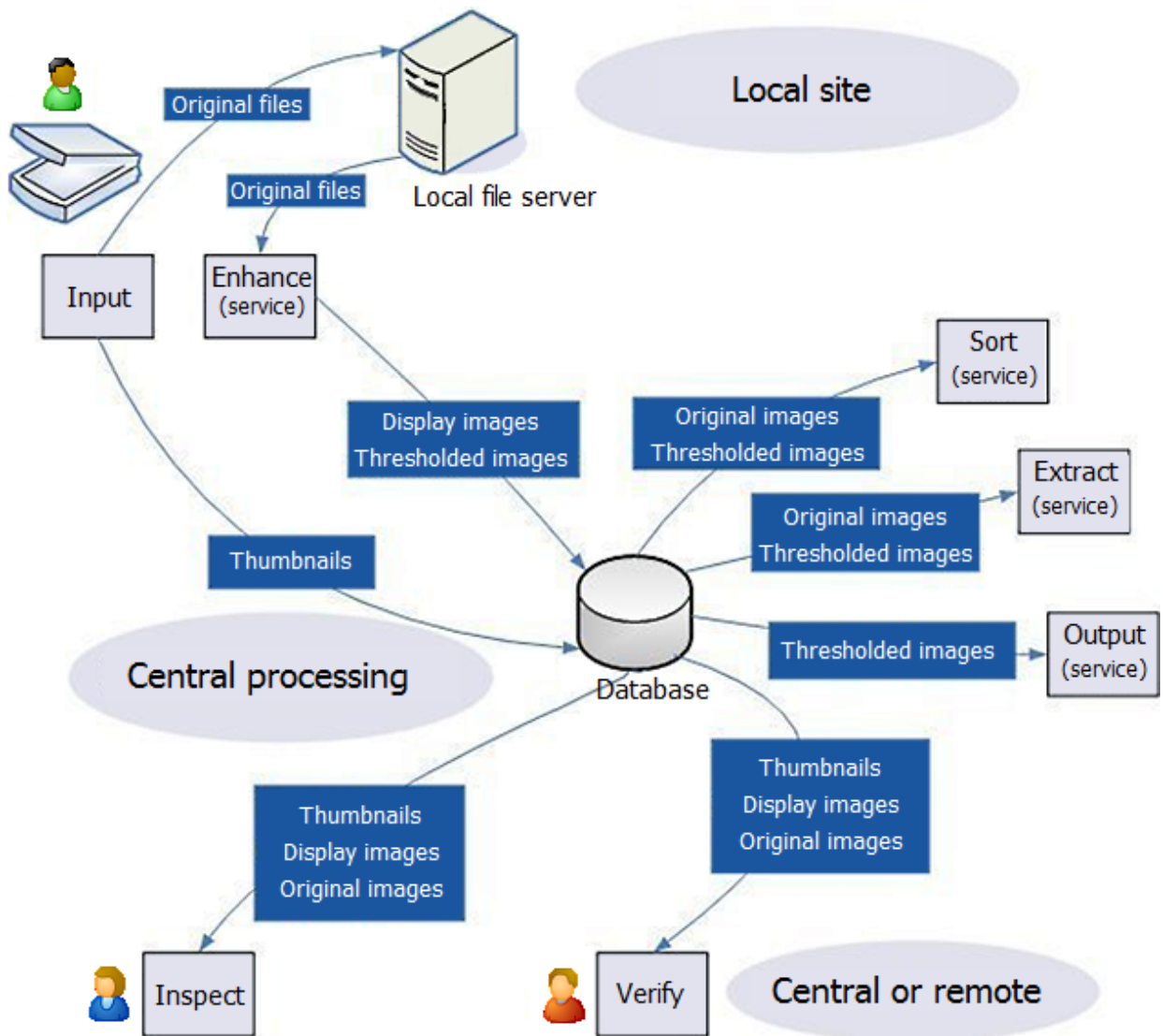
You can specify thresholding, display image, and other general image settings in production tasks containing the Enhance activity.

The diagram below shows the location of the local file server and scanning station, the database, and the workstations used for automatic tasks (which run as Windows services) and manual tasks, in a typical distributed solution. The process is as follows:

1. In the **Persistence policy** dialog, select **Scan to file and store in database from Enhance**.

2. During Input, documents are scanned to a local file server / the Scan client.
3. Create an Enhance task, run as a service at the same location.
 - For documents in black and white, use the default **Display image** settings.
 - For documents in color:
 - a) Specify the **Display image settings**, creating small display images by selecting **Use 8-bit palette image** and **Delete original image after creating smaller display image**.
 - b) Use the default **Thresholding** settings.

After Enhance, these images are removed from the local file server and saved in the database for central processing.



4. Configure services to run Sort, Extract and Output. (The diagram shows which images these tasks use.)
 5. Specify Inspect and Verify tasks to be run manually by production operators at workstations that have database access.
- Note:** In an invoice solution using Invoice modules, thresholded images are saved temporarily on a different file server. DOCUMENTS Help contains a separate diagram that shows this – see the topic called “Image location in an invoice solution”.

Important when documents are scanned at multiple sites

When documents are scanned at more than one local site, you need a way to ensure that the Enhance activity at one site does not try to access documents scanned at another site. When documents are scanned, the image file is created at the local site while document information (including the location of the image) is sent to the database where central processing takes place. If the images are not where Enhance expects them to be when it checks the database, an error occurs.

The easiest way to keep things running smoothly when scanning from multiple sites is by using queues. For example, for a solution where scanning is done at two local sites, one Scan site scans to Queue 1, and the other to Queue 2. Two separate production tasks, one at each site, are set up to enhance images from their respective queues only. Production tasks used for central processing are then set up to process documents from both queues.

Running Production tasks as services

Using Windows services to run automatic production tasks is convenient, the services start automatically, and they process tasks without any user interface. ReadSoft recommends following these best practices when setting up and running services:

- Normally, automatic tasks are separated into three or four services; one task running the Enhance and/or Sort activities, one running Extract, and one containing the Output activity.
- Do not specify the auto restart time (**Wait time before restart**) in a production task's **Performance** settings. Instead, configure services to use the `/B` command-line parameter to specify a specific number of batches for Production to process. You can also specify in what sequence tasks are processed, followed by the `/R` command-line parameter to state how many times the sequence is repeated before the Production module is restarted.
- For optimal performance, the number of parallel instances in the behavior file, which controls the services, should be no greater than the number of CPUs on the server. If hyper-threading is available, it is possible to have twice as many parallel instances per CPU.
- Use the multiple command lines to optimize CPU usage on the server. See the “Production as a service: Overview” topic in *DOCUMENTS Help*.
- When setting up a service, start by having the logging = “1” parameter switched on, and check the service's status using the Windows Event Viewer.
- Change the service properties, on the **Log On** tab. Because the local system account may not always be allowed the write permissions in all necessary folders, a special account should be used to run the service (with an account and password specified).

Thin-client requirements

DOCUMENTS solutions can be run in thin-client environments such as Citrix MetaFrame XP and Microsoft Terminal Server. However, ReadSoft cannot provide design guidance for these environments. Please contact Microsoft or a certified Citrix partner for system design and implementation.

System-wide specifications

After installation, some global policies are already in place to help you to get started quickly. However, some policies and settings are specific to your system and must be defined using the Administration module. Check and (if necessary) adjust the settings on the **Configuration** tab, under **System specifications**.

For list of things to check (with links to instructions), please refer to the *DOCUMENTS Help* topic “Setting up the system after installation.” This includes:

- Configure add-ons (if any).
- Set up logging.
- Enable audit trails, and specify what information to include in them.
- Select which country and language specifications to use when processing documents.
- Specify environment variables.
- Adjust user management settings.