

docWizz

System Requirements



Table of Contents

- 1 docWizz3
 - 1.1 Introduction docWizz editions3
 - 1.2 Overview docWizz editions.....4
 - 1.3 The three docWizz editions in comparison.....5
 - 1.4 Additional modules/features7
- 2 docWizz requirements.....8
 - 2.1 docWizz Basic | Pro | Enterprise8
 - 2.1.1 Scope.....9
 - 2.1.2 docWizz clients9
 - 2.1.2.1 Operator workstation (dWClient, dWControlCenter)9
 - 2.1.2.2 Scanner workstation (dWScanClient)..... 11
 - 2.1.3 docWizz processing server (dWService instances) 13
 - 2.1.4 RemoteQA solution 15
 - 2.1.4.1 Overview..... 15
 - 2.1.4.2 Benefits..... 16
 - 2.1.4.3 Technical requirements 17
- 3 Environmental requirements 19
 - 3.1 Recommendations for server combinations 20
 - 3.2 Standard requirements 23
 - 3.2.1 Domain server 23
 - 3.2.2 Database server 24
 - 3.2.3 File server 26
 - 3.2.4 Firewall 27
- 4 Appendix 28
 - 4.1 Calculation 28
 - 4.2 docWizz site definitions 29
 - 4.3 Glossary..... 30
 - 4.4 Special recommendations for docWizz 7.1 or higher versions 33

1 docWizz

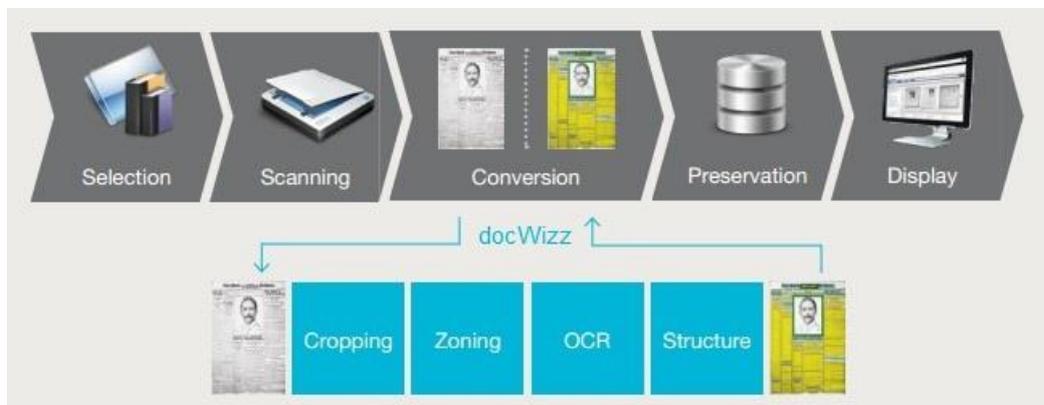
docWizz is a software program used by the most renowned libraries, publishing houses, and companies worldwide to digitize and convert their valuable library holdings and archives for easy access, searchability, and long-term preservation.

docWizz is the only software that bundles all necessary conversion steps in a single, smooth workflow.

The three docWizz editions (Basic, Pro and Enterprise) guarantee a custom-fit solution for your project, be it a small collection or a National Library.

Why docWizz?

- Single, smooth workflow with central control center
- Time savings due to streamlined and automated process
- No expensive errors from false copying or lost data shipments
- Consistent, standardized output
- Easily up-scalable, from thousands to millions of pages



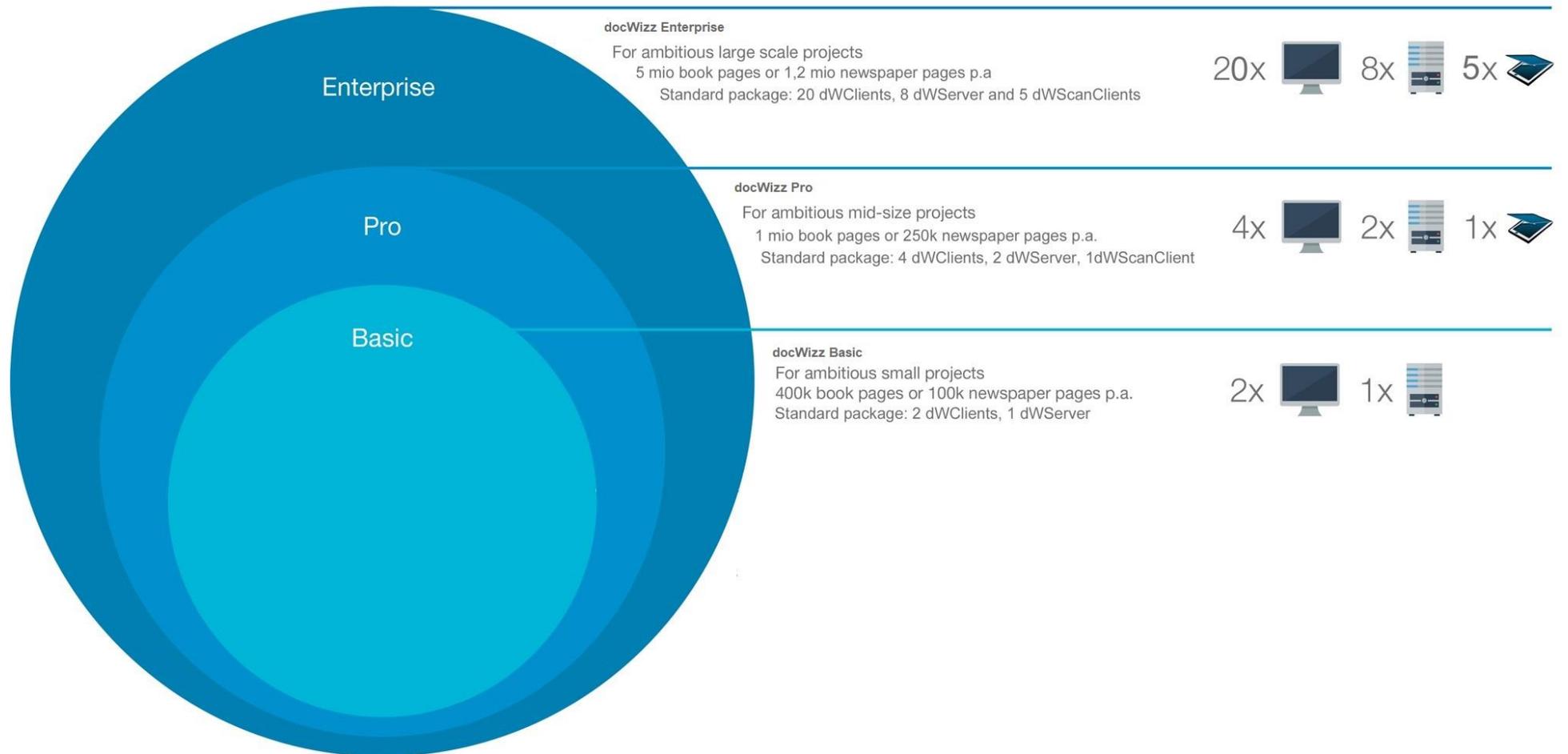
1.1 Introduction docWizz editions

docWizz comes in three different editions. Each edition is directed towards a specific field of application, from the affordable docWizz Basic edition for ambitious small digitization projects up to the flagship docWizz Enterprise edition for very large collections.

Basic, Pro and Enterprise editions are client-server-solutions. Those have more or less the same features, but differ in the dimension of their capabilities. You can scale up your workflow at any time by adding servers or client workstations. We help you to select the right edition based on your individual needs and project scope.

The following graphic shows you the four editions of docWizz regarding their capacity and which version could be of interest for you.

1.2 Overview docWizz editions



1.3 The three docWizz editions in comparison

	Basic	Pro	Enterprise
	For ambitious small projects	For ambitious midsize projects	For ambitious large-scale projects
Guidance value	400.000 book pages or 100.000 newspaper pages per year	1 million book pages or 250.000 newspaper pages per year	5 million book pages or 1,25 million newspaper pages per year
SOFTWARE			
MODULE LICENSES			
dWClient	2	4	20
dWServer	1	2	8
dWScanClient	*	1	5
MONITORING			
dWControlCenter	✓	✓	✓
IMPORT			
Image files	✓	✓	✓
PDF	✓	✓	✓
Batch-option	✓	✓	✓
CROPPING			
Basic cropping	✓	✓	✓
Advanced cropping	✓	✓	✓
ZONING			
Fast correction tool	✓	✓	✓
Customized zone types	✓	✓	✓
OCR			
Tesseract (incl. Gothic)	✓	✓	✓
ABBYY FineReader (excl. Gothic)	✓	✓	✓
OmniPage	*	*	*
STRUCTURE			
Document Hierarchy View	✓	✓	✓
OUTPUT			
METS/ALTO/JP2	✓	✓	✓
PDF	✓	✓	✓
ePub	✓	✓	✓

TRAINING			
Video tutorials	✓	✓	✓
Project documentation	✓	✓	✓
Training	Substantial remote training	On-site training	Extensive on-site training
MAINTENANCE			
Online support	✓	✓	✓
Software updates	✓	✓	✓
On-site support, yearly	n/a	1 day	2 days



= included

n/a = not available

* = see Additional modules/features

1.4 Additional modules/features

* ADD ONS			
Additional dWClient licence	✓	✓	✓
Additional dWServer licence	✓	✓	✓
Additional dWScanClient licence	✓	✓	✓
Catalogue connection (Z39.50)	✓	✓	✓
OmniPage OCR	✓	✓	✓
ABBYY OCR Single Core Runtime	✓	✓	✓
ABBYY OCR Dual Core Runtime	✓	✓	✓
ABBYY OCR Non-standard languages ("CJK" Chinese Traditional / Japanese / Korean or Vietnamese or Thai or Hebrew or Arabic) for Single Core Runtime	✓	✓	✓
ABBYY OCR Non-standard languages ("CJK" Chinese Traditional / Japanese / Korean or Vietnamese or Thai or Hebrew or Arabic) for Dual Core Runtime	✓	✓	✓
ABBYY Gothic A4 page-based	Pricing upon request	Pricing upon request	Pricing upon request
Special XML output	On request	On request	On request
Work-day training (travel expenses not included)	✓	✓	✓
Work-day customizing	✓	✓	✓
RemoteQA	On request	On request	On request

ADD-ONS are subject to maintenance, where applicable.

-  = available
 n/a = not available

2 docWizz requirements

This chapter provides the system requirements for a docWizz installation including hardware, operating system and additional settings.

The editions Basic, Pro and Enterprise have similar requirements - depending on the project scope - which we've combined in the chapter *docWizz Basic / Pro / Enterprise*.

The environmental requirements for the domain, file and database server are described in *Environmental requirements*, with a special chapter covering the essential firewall settings (*Firewall*).

Video tutorials are provided for initial training and CCS provides support via web platform.

2.1 docWizz Basic | Pro | Enterprise

In this chapter, the scalable docWizz editions Basic, Pro and Enterprise are outlined.

These editions can be extended at any time to meet the project or organizational needs.

[docWizz Basic - get going](#)

docWizz Basic edition is a client / server ready software with network integration within a Microsoft domain. The IT department should be involved to provide a reliable environment. docWizz Basic allows you to run dWClient instances on two computers in parallel, so two people can work on different jobs. Plus, a server is used to run all automated processes in the background 24/7. dWScanClient can be used on scanner computers to easily transfer images and import documents into docWizz processing. Additional licenses can be ordered and an upgrade to a higher edition can be done anytime. In addition to video tutorials we provide substantial introduction training and production assistance via remote access.

[docWizz Pro - scale up](#)

This edition is for professional digitization, and is typically set up if there is a multi-year digitization plan or anticipation of a large project. docWizz Pro allows the digitization of millions of pages per year. This set-up consists of five client licenses, so five people can work in parallel using docWizz. Two server licenses ensure redundant background processing 24/7 for all automated processes. CCS provides on-site training and production assistance. An update to the new release, an introduction for best usage of the new improvements and assistance for docWizz operations is given on-site annually.

[docWizz Enterprise - go big](#)

The enterprise edition is a good fit if you plan to digitize at really large scale over the next couple of years. If there is a digitization strategy and a plan to scan and convert millions of pages and many library collections, this edition of docWizz is the best option. Extensive on-site training and consulting for the setup of the workflow are included, as well as annual on-site consulting to ensure you get the most out of docWizz. This also provides an opportunity to discuss the latest trends and best practices. CCS consults with your IT department about hardware and helps with the software integration in the libraries' systems.

2.1.1 Scope

We suggest these docWizz editions to customers who aim to professionally or commercially digitize large amounts of books and newspapers.

We offer operator training, direct support and customization of the software (upon request) to help you to have the most efficient production environment.

Due to the high-end docWizz environment, there is a lot of flexibility in how the infrastructure to be set up.

The system can be scaled-up to meet any needs by adding additional workstations and server machines. Keep in mind that this could affect the overall system performance and other upgrades or changes e.g. for hardware could be necessary. For details see Recommendations for server combinations.

- For all these three editions you need a central file server, a Microsoft domain and one Microsoft database server.
- The docWizz server and docWizz Client requires access to all of these by network.
- Minimum network bandwidth is 1 Gbit/sec; for large production environments, more bandwidth might be needed for specific servers (e.g. file server).
- Each docWizz server and docWizz client is able to perform the same tasks within the environment. Due to that, no single point of failure exists as soon two or more machines of the same kind exist. There is no production data locally stored on these machines, which guarantees more security for your production.

The risk of all remaining single point of failures for network, file, domain and database server can be eliminated. Details see Recommendations for server combinations.

All three editions can work in a RemoteQA scenario. For more information, see [RemoteQA solution](#).

Note: There are server configurations possible, which could influence the performance of tasks, such as image processing and structuring.

2.1.2 docWizz clients

The following two sections describe the requirements for the clients modules operators work with:



Manual tasks, such as verification of cropping and OCR correction, will be done in the dWClient.

The interface guides you through this process. The dWScanClient provides one of the first opportunities to catch and correct mistakes on the digitization flow. It provides a clear interface which works independent to scanner manufacturers and makes it easy to train new operators.

2.1.2.1 Operator workstation (dWClient, dWControlCenter)



- The docWizz user interface applications are installed on operator workstations (dWClient & dWControlCenter for Basic edition and higher). The monitoring interface named dWControlCenter is installed on these machines to control and maintain the production of the docWizz environment
- The intermediate verification tasks within the docWizz workflow are performed by operators with the docWizz Client. The dWControlCenter monitors the docWizz background services and environment
- In general these machines only operate during office hours. No automated tasks or data transfers through the network are required here after hours.
- The parallel execution of dWClient's are limited by license keys (floating seat licenses)

- No production data is stored on these operator machines.
Only the docWizz binaries (program data) are stored locally on the machine and the local hard drives are just used for temporary data swapping

dWClient / dWControlCenter requirements

CPU	Minimum: Intel® Core™ i3 Recommended: Intel® Core™ i5
RAM	Minimum: 4 GB RAM Recommended: 8 GB RAM
NIC	Minimum: 1Gbit/s
Screen resolution	Minimum: 1280 x 720 Recommended: 1920 x 1080 or UHD(4K)
Local disk space	4 GB of local space required More than 1 GB free space for the temporary export generation swap.
Operating system *	<p>Versions: Windows 10</p> <p>Processor architecture: 64 bit recommended, 32 bit supported only for docWizz versions smaller or equal with 7.0 (deprecated)</p> <p>Professional or higher</p> <p>Version type: English**</p> <p>Language:</p>
Domain	Workstation must be member of the Microsoft domain
Permission	<p>Applications need to be executed on domain user account with permission for updating binaries on local machine in %ProgramFiles(x86)% and %ProgramFiles% folder.</p> <p>The machine / domain user needs permanent full access to all docWizz locations.</p> <p>If operator has no such domain account or has insufficient permission, the dW applications can be executed on another pre-defined domain account.</p>
Internet access	Internet access for validation of XML data and activation of licenses is required for validation purposes. If no internet access is available, the configuration needs to be changed to local validation.
Firewall	MS Firewall needs to be disabled or opened for full ODBC and UNC connection. See Firewall .

*) For installations and updates a completely updated operating system is required. All current Windows service packs have to be installed.

***) Other languages possible, but not tested.

2.1.2.2 Scanner workstation (dWScanClient)



- Scanners can be integrated into the docWizz workflow, either with the delivered scanning manufacturer software or by using the application dWScanClient
- When dWScanClient is used, it runs on the scanner computer and the user interface displays the images as they are being generated. It also stores additional metadata
- The application generates thumbnail images of the scanned images on-the-fly
- The scanner computer must be fast enough to run dWScanClient software in parallel to scanning
- dWScanClient uploads images to the IN share of docWizz after completion of the scanned item. In case of performance issue (parallel I/O on the hard drive) the upload process can be paused and performed after the scan production
- The upload task requires the dWScanClient application to be executed. The scanner computer can be kept running after scan operation and can be shutdown automatically after completion
- On Scanner computer also dWControlCenter and dWClient application are installed. The dWControlCenter can also be started here for maintenance of the environment. The dWClient is needed just temporarily in case re-scans are requested for documents being already in the docWizz processing
- The parallel execution of dWScanClient's are limited by license keys (floating seat licenses)
- No production data is stored on these scanning machines. Local hard drives only hold temporarily the scanned images. Only in case the all space on the available IN shares are occupied, dWScanClient keep the images on the scanning machine, until free space is provided again on the storage. According information are outlined to the operator on the application interface

For the final system requirements the specification of the scanner manufacturer needs to be considered.

dWScanClient requirements

CPU	Minimum: Intel® Core™ i3 Recommended: Intel® Core™ i5								
RAM	Minimum: 4 GB RAM Recommended: 8 GB RAM <i>dWScanClient application might need up to 1 GB of memory (depending on size of scanned images)</i>								
NIC	Minimum: 1 Gbit/s								
Screen resolution	Minimum: 1280 x 720 Recommended: 1920 x 1080 or UHD(4K)								
Local disk space	4 GB for docWizz binaries + minimum 1 complete item free space for scanning (the complete book / newspaper / reel needs to be kept locally before it can be uploaded in its entirety docWizz storage) ***)								
Operating system *	<table> <tr> <td>Versions:</td> <td>Windows 10</td> </tr> <tr> <td>Processor architecture:</td> <td>64 bit recommended, 32 bit supported only for docWizz versions smaller or equal with 7.0 (deprecated)</td> </tr> <tr> <td>Version type:</td> <td>Professional or higher</td> </tr> <tr> <td>Language:</td> <td>English **</td> </tr> </table>	Versions:	Windows 10	Processor architecture:	64 bit recommended, 32 bit supported only for docWizz versions smaller or equal with 7.0 (deprecated)	Version type:	Professional or higher	Language:	English **
Versions:	Windows 10								
Processor architecture:	64 bit recommended, 32 bit supported only for docWizz versions smaller or equal with 7.0 (deprecated)								
Version type:	Professional or higher								
Language:	English **								
Domain	Workstation must be member of the Microsoft domain								
Permission	<p>Applications need to be executed on domain user account with permission for updating binaries on local machine in %ProgramFiles(x86)% folder.</p> <p>The machine / domain user needs permanent full access to all docWizz locations.</p> <p>If operator has no such domain account or has insufficient permission, the dW applications can be executed on another pre-defined domain account.</p>								
Internet access	Internet access for validation of XML data and activation of licenses is required for validation purposes. If no internet access is available, the configuration needs to be changed to local validation.								
Firewall	MS Firewall needs to be disabled or opened for full ODBC and UNC connection. See Firewall .								

*) For installations and updates a completely updated operating system is required. All current Windows service packs have to be installed.

**) other languages possible, but not tested

***) book has maximum 1000 pages, each 2 MB of disk space => minimum 2 GB free space required.

2.1.3 docWizz processing server (dWService instances)



- The dWService instances are installed on processing servers; these are running 24/7 as Windows services
- The dWService instances are controlled by a separate application called "docWizz ControlCenter". This application is part of the docWizz installation and can be started from any dW server or operator workstation. The operators do not need direct access to the servers to control the services
- Four dWService instances are installed on each server. In exceptional cases, fewer dWService instances are installed
- The amount of processing servers and total amount of dWService instances are limited by license keys
- No production data is stored on these processing servers. Local hard drives only hold the installed dW binaries and are used for temporary data swapping
- For support / administrative work, internal RDC (Remote Desktop Connection) access should be enabled for all servers

For high stability demands:

- The services request all processing data on demand from the file servers. To ensure effective operation, the file servers and processing servers should be closely connected to each other (best same [stacked] switch). No production data is stored locally
- Due to operation time of 24/7 and prevention of down-times, a redundant hardware design is recommended (power supply, hard drive [RAID-1], NIC [recommended: load balanced])
- In large production environments, BladeCenter technology is recommended. By default, this already provides remote management and monitoring interfaces and can send out alerts if a problem is detected
- Additional information on security aspects is available in the introduction PPT "Concept for Production Security / Fallback Scenario"

dWServices requirements

CPU	<p>Minimum: Intel® Xeon® E3 (or similar) Recommended: Intel® Xeon® E5 (or similar)</p> <p>For higher performance when using ML improvements (available from version docWizz 7.1) please see additional recommendations on chapter Special recommendations for docWizz 7.1 or higher versions.</p> <p>For higher performance when using Tesseract as OCR engine please see as well Special recommendations for docWizz 7.1 or higher versions.</p>								
RAM	<p>Minimum: 8 GB RAM</p> <p>More RAM may be required for large-sized image processing (>DIN A2)</p>								
NIC	<p>Minimum: 1 Gbit/s</p> <p>For high-demand environments 2 to 4 Gbit/s trunks are beneficial depending on entire environment and file server speed</p>								
Screen resolution	<p>No local screens are required on these machines (except for administrative work by IT). Minimum: 1280 x 1024</p>								
Local disk space	<p>4 GB of local space required for docWizz binaries files. More than 4 GB free space for the temporary export generation swap.</p>								
Operating system *	<table> <tr> <td>Versions:</td> <td>Windows Server: Minimum 2012 Recommended 2016</td> </tr> <tr> <td>Processor architecture:</td> <td>64 bit (required to address full memory)</td> </tr> <tr> <td>Version type:</td> <td>Standard or higher**</td> </tr> <tr> <td>Language:</td> <td>English ***</td> </tr> </table>	Versions:	Windows Server: Minimum 2012 Recommended 2016	Processor architecture:	64 bit (required to address full memory)	Version type:	Standard or higher**	Language:	English ***
Versions:	Windows Server: Minimum 2012 Recommended 2016								
Processor architecture:	64 bit (required to address full memory)								
Version type:	Standard or higher**								
Language:	English ***								
Domain	Server must be member of the Microsoft domain.								
Permission	Services need to be registered on domain user. This domain user needs local administrator rights to be able to operate as a local service. The machine / domain users need permanent full access to all dW locations.								
Firewall	<p>MS Firewall needs to be disabled or opened for full ODBC and UNC connection.</p> <p>See Firewall</p>								
Graphical card	Not needed for regular processing, but recommended for higher performance when using ML improvements (available from version docWizz 7.1) please see additional recommendations on chapter Special recommendations for docWizz 7.1 or higher versions .								

*) For installations and updates a completely updated operating system is required. All current Windows service packs have to be installed.

**) The Essentials edition of Windows 2012 R2 server is designed as a small business server and might cause problems, if used in a network with several other servers, running docWizz services. Hence, we do not recommend the Essentials edition, but rather go for the Standard Edition of Windows 2012 R2 or higher.

***) Other languages possible, but not tested

2.1.4 RemoteQA solution



The RemoteQA solution is an option to allow quality control and manual correction of digitized material to be separated from processing site.

For additional information and details about the functionality and benefits, see separate PPT introduction "Environmental structure".

The differences in requirements within a RemoteQA solution, additional aspects and requirements to be considered can be found in the following chapters.

2.1.4.1 Overview

The RemoteQA solution is, generally speaking, a combination of individual docWizz environments. Each of the docWizz sites is setup and exists independently.

At any time, the sites can be bound in a RemoteQA solution. At this moment, the roles for the sites are defined.

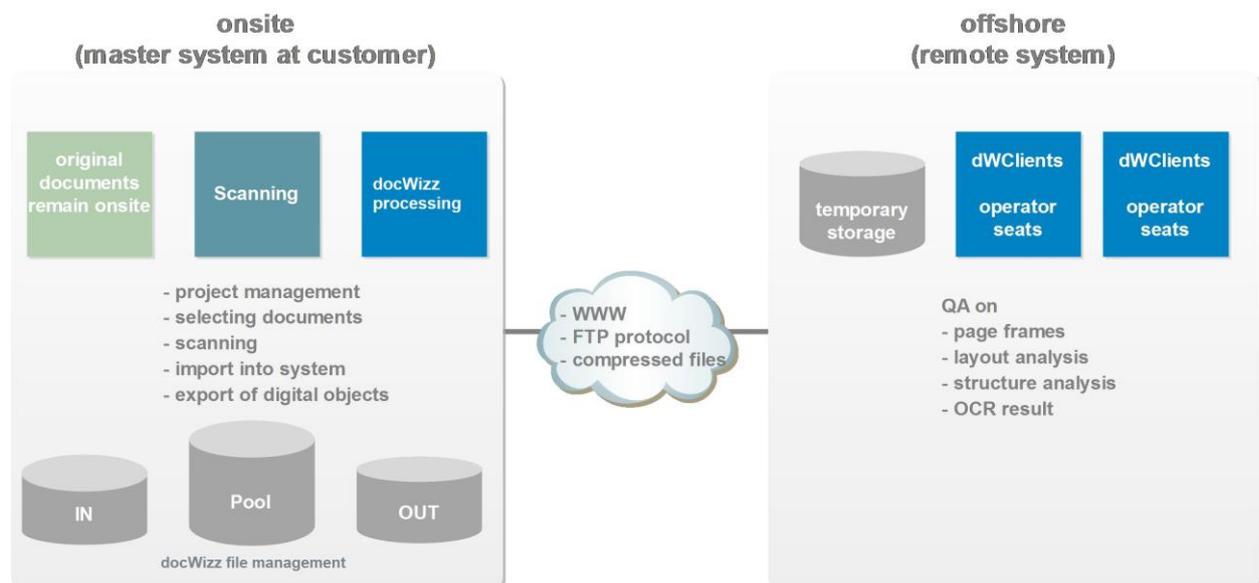
One site becomes the "Manager". The "Manager" is the site where the original images exist or are scanned. On this site, all automated processing, like the OCR has to take place on the high quality / original scanned images. Also, the final output for the digital objects can only be generated here.

The other site becomes the "Loader". The "Loader" is the site where the operators are available for the manual quality control and correction of the digitized material.

Only highly compressed images from "Manager" get transmitted to the "Loader". These images will be optimized so that they are good enough for the operators tasks, but as small as possible for fast transfer rates.

Due to that no processing of image operation, layout analysis or OCR could be performed on "Loader" site since the highest quality results could not be achieved.

Here is a diagram of this RemoteQA solution:



2.1.4.2 Benefits

The activation for the RemoteQA workflow and for which steps / tasks is done project based.

Since no master images are available on "Loader" site, the storage capacity demands are much less and IN and OUT locations are not used there.

On each site, one additional service needs to be installed (RemoteQAManager resp. RemoteQALoader). This service manages the document status and sets the documents to "active" or "inactive" to ensure the document is only able to be opened on one site at a time.

The dWFTPClient runs on the "Manager" or "Loader" site as a service. This service immediately uploads and downloads all data packages to/from server side (push & pull).

Any regular FTP Server should be installed on the opposite site. No special FTP server is required.

Finally, the RemoteQA solution is not limited to a 1:1 relation of docWizz sites. It can be extended to an 1:n or even could be extended to n:n relationship.

A production site ("Manager") where scanning is taking place could share the documents with several remote sites ("Loader") for manual correction.

Also, a service provider operating on one project as "Loader" to correct documents for a library could operate on another project as "Manager" where he transmit documents to a subsidiary in another country.

Within one project, one docWizz site can only operate as a "Manager" or as a "Loader". Due to this, also a forwarding of documents to a 3rd site received from any other "Manager" is not possible!

Benefits:

- Minimized data transfer through the internet (quality level configurable) makes hard drive shipping with all kind of related problems obsolete (corrupt disk, corrupt images, tax issues, ...)
- Immediate data transfer allows for correction of scan mistakes. If a document has errors, pages can be highlighted and transferred between "Loader" and "Manager" to allow operators to resolve the problem.
- Automatic monitoring and highlighting of orphan documents.
- Automatic transmission of data 24/7 including all interruption handling (file based checksum handling).
- RemoteQA option combinable with local operations (e.g. keep final QA locally).
- Share the risk for milestone achievement by using multiple "Loader" sites on same project.
- Remote QA option possible for each project by specific rules (percentage, titles).

The system requirements for docWizz RemoteQA solution are built out of the standard requirements plus some additional demands, which are described in following chapter.

2.1.4.3 Technical requirements

For the production site, the standard requirements are valid.

In addition to the standard requirements, the following additions should be considered:

Production site "Manager"

<p>Storage capacity</p>	<p>The highly compressed images that have been transmitted to the "Loader" site are also stored in the POOL of the "Manager" during life-time of the document.</p> <p>This additional storage capacity is only recognized in very large scale productions (millions of pages) and with long turn-around times.</p> <p>In average, the additional space can be calculated as 10% of the original source image (uncompressed).</p>
<p>Removed components</p>	<p>If all quality control is done on the Remote site, only a few dWClient are needed for production managers to manage and review documents and verify individual issues.</p>
<p>Additional work tasks</p>	<p>Additional work tasks can be executed on "Manager" site to</p> <ul style="list-style-type: none"> • prepare RemoteQA data packages for upload • restore the adapted data from "Loader" side in "Manager" pool • monitor and detect orphan documents
<p>dWRQAManager service</p>  <p>dWRQAManager</p> <p>Management service which controls the document status changes between "Manager" and "Loader" site.</p>	<p>This service is added to one of the docWizz servers; no additional requirements.</p>
<p>dWFTPClient Service (by default on Manager site)</p>  <p>dWFTPClient</p> <p>Special FTP client which automatically transmits the RemoteQA data via FTP forth and back (push & pull).</p>	<p>This service is added to one of the docWizz servers.</p> <p>This service requires internet connection to establish the FTP connection to remote location (Loader site). Firewall settings might need to be adapted.</p> <p>The FTP port can be set to any preferred control port. Active and passive mode are also possible.</p> <p>Hint:</p> <p>In case of security or firewall issues, the FTPClient can be installed on "Loader" site and the FTPServer could be installed on "Manager" instead.</p> <p>See requirements for this below.</p>

For the remote site, the standard requirements are valid.

In addition to the standard requirements, the following additions should be considered:

Remote site "Loader"

<p>Storage capacity</p>	<p>For the storage, less capacity is required because no "Manager" images will be stored in the IN share and only highly compressed images will be placed in the POOL.</p> <p>For a generic estimation, the following calculation can be used:</p> <p>Total capacity = $\langle \text{SourceImage size} \rangle \times 0,2 \times \langle \text{numberOfPagesPerDay} \rangle \times \langle \text{turnAroundTimeForQA (days)} \rangle$</p>
<p>Additional work tasks</p>	<p>Additional work tasks are executed on remote site to</p> <ul style="list-style-type: none"> • unpack RemoteQA data packages on retrieval • pack the corrected documents from "Loader" for transfer to "Manager" • monitor and detect orphan documents
<p>dWRRemoteQALoader</p>  <p>Management service as corresponding partner to the dWRRemoteQAManager (managing document status changes)</p>	<p>This service is added to one of the docWizz servers; no additional requirements.</p>
<p>"Regular" FTPServer</p>  <p>3rd party software for FTP connection retrieval.</p>	<p>This software is NOT part of docWizz. CCS, by default uses open-source FTP Server like "FileZilla Server", if no other exists on customer side.</p> <p>The third-party software will dictate its own system requirements.</p>

Both sites ("Manager" and "Loader")

For the data transfer through internet the external network bandwidth needs to be calculated. Based on production experience, the following factors recommended:

$\langle \text{SourceImage size (uncompressed)} \rangle \times 0,1 \times \langle \text{numberOfPages} \rangle / \langle \text{period for quality control} \rangle$

CCS experience:

color: $500 \text{ kB} / \text{page} \times \langle \text{numberOfPages} \rangle / \langle \text{period for quality control} \rangle$

gray: $150 \text{ kB} / \text{page} \times \langle \text{numberOfPages} \rangle / \langle \text{period for quality control} \rangle$

ATTENTION: It has to be considered that the upload stream on "Manager" site is equal the download stream on Loader site. As the images are only transmitted to "Loader" (not backwards to "Manager") and the data volume of the analysis files in comparison can be ignored, the bandwidth to be approved against the contract with ISP (internet service provider) are:

on "Manager": upload bandwidth

on "Loader": download bandwidth

3 Environmental requirements

docWizz was developed for Microsoft environments as server- / client- application.

The central criteria of this solution:

- Store all processing data so that it is centralized and accessible from any service or client.
- One central configuration for all modules to operate with the same settings.

By this, a parallel processing, common access to stored data and monitoring of the operations is possible.

Also, this allows the docWizz environment to scale from one server with one client up to a processing farm with more than 25 servers and hundreds of docWizz clients for a real mass digitization workflow.

From software point of view there are (almost) no limits. What needs to be considered on the scaling are the increasing environmental requirements for network, storage, database and domain services.

For example the demand on domain authorization, database operations and file access on the file server caused by two processing servers is that low, that the docWizz processing servers can also operate in the role as database and file server in addition and an existing domain of the organization could just be used.

This would not be possible for a large scale environment, for example with 20 servers and 50 clients and more.

The amount of clients has a secondary impact on the environmental requirements as these are not operating 24/7 and are not as demanding as services are on resources from database and file server(s). They also do not create that much network traffic compared with the services.

The real impact is the amount of processing servers. These services cause permanent traffic on the network, on the file server for load and save operations, and on the database for read and write access.

The amount of parallel operations, especially on file server and database, can bring an environment to a total "breakdown".

Due to that, we outline in the following chapter the best-practice recommendations based on the experience of CCS with all of our clients.

3.1 Recommendations for server combinations

In this chapter, for simplification the environmental requirements were grouped in the major roles that the servers have:

- Domain server – all basic domain functionality, including DHCP, DNS, NTP, etc.
- Database server – database server with all docWizz databases (3 resp. 4 databases). All docWizz modules are constantly connected to this database server.
- File server – the server which provides the UNC paths for the docWizz required shares:
 - dWShare – central configuration for whole docWizz environment
 - IN – source image location
 - POOL – intermediate processing results (during docWizz process)
 - OUT – location where the resulting digital output is created

Generally, all roles (domain, database and file server) can be separated. This is not outlined separately. Also, the file server can be split into independent servers. For example POOL and dWShare (configuration) could be placed on a faster file server (small FC or SAS disks with high RPM / SSD disks for permanent random access) while IN and OUT share are setup on a regular file server (big SATA disks).

Details for possible separations and best practices regarding data security can be found in the introductory presentation 04-dW_SecurityScenarios.

In the following table, we display the possible combinations of server roles:

Sub-combinations which are not outlined here are also possible, e.g. just separating the database to use an existing domain server. The explanation for consequences and criteria to be considered are outlined in the comments column.

If you have any questions, you might find answers in the FAQ's - or - do not hesitate to ask dWSupport. Our technical consultants and system engineers are pleased to provide the required information.

The combinations are based on the standard amount of clients / services. In each edition, additional processing servers / client machines can be added. The recommendations do not change on adding one or two servers, but when multiple are added, the environment needs to be carefully thought out and monitored to ensure there will be no performance issues or lag in response times.

■ combination, where specific conditions needed, weak point is outlined

■ standard combination

■ combination, where specific conditions needed

	Domain server	Database server	File server	Processing server	Comments
General	x	(x)	(x)	x	dWServices need to be executed with local admin rights. Since domain servers do not have local users, the services would need to be executed on a domain admin account. This is possible, but not preferred since the credentials need to be shared with CCS for support. We prefer to set docWizz up with local admin rights to prevent this situation.
dW Basic a)	x	x	x		For Basic, with two servers, the load on network interfaces is not that high; the second server would not get into access issues.
dW Basic b)		x	x	x	On using an existing server, the combination with processing server is no problem. Also from performance, the combination with Domain server role would be possible, but not recommended (see General).
dW Pro a)	x	x			Due to amount of processing servers the network traffic and I/O load on disk, the file server is increasing a lot. The separation of the file server to independent NAS (network attached storage) is recommended.
dW Pro b)	x		x		see Pro a) As long a network trunk of 4 Gb/s is setup and the machine CPU is very good, this is also a possible combination.
dW Pro c)		x	x		
dW Pro d)		x		x	The combination of processing server with database server is no real problem from the resource point of view. It is more about the fallback concept point of view. The one processing server would become a single-point of failure. There the mixture with the processing server role is not recommended.

dW Pro e)	x	x	x	<p>Based on CCS experience, this combination is not economic any more.</p> <p>A 10 Gb/s interface would be required for this server; a multi-CPU machine with extra-ordinary RAM is required.</p> <p>The increased costs are not in a balance to the benefit. Also, in case of a failure there is no possibility for workaround to temporarily change a role from one existing server to another.</p>
dW Enterprise a)	x	x		<p>Due to amount of processing servers the network traffic and I/O load on disk, the file server is increasing a lot. The separation of the file server to independent NAS (network attached storage) is recommended.</p>
dW Enterprise b)	x		x	<p>For enterprise environments, the load on file servers becomes very high.</p> <p>Here it is typical to split the shares across multiple file servers and even to have multiple shares for IN and POOL.</p>
dW Enterprise c)		x	x	<p>Also, normally professional NAS storages are utilized here with safety concepts for high reliability and best performance.</p> <p>Due to that, domain and file server or database and file server are seldom combined.</p>
dW Enterprise d)	x	x	x	<p>Based on CCS experience, this combination is not economic any more.</p> <p>A 10 Gb/s interface would be required for this server; a multi-CPU machine with extra-ordinary RAM is required.</p> <p>The increased costs are not in a balance to the benefit. Also, in case of a failure there is no possibility for workaround to temporarily change a role from one existing server to another.</p> <p>For Enterprise environments, the normal storage capacity needs exceed the dimension that Microsoft file server can efficiently manage.</p> <p>The maximum size of file systems that is technically possible (defined by amount of spindles in one partition) has been reached.</p>

3.2 Standard requirements

Finally the requirements for the individual roles and specific hints to be considered within a docWizz environment are outlined in the following subchapters.

3.2.1 Domain server

The domain server is the administrative server running all Microsoft domain services.

Herewith it is considered of all basic domain functionality including DHCP, DNS, NTP, etc.

(DNS or DHCP services can also be operated by network routers of course.)

Following docWizz production considerations:

- For small environments, the usage of existing domains is recommended (prevention of administrative overhead). For large environments the amount of accounts and production separation often make sense – best practice recommendation: define a sub-domain of existing organization domain.
- For operator monitoring (docWizz log) and access permission control, each operator should get his personalized domain account.
- Also, local redundancy is recommended to prevent loss of the domain controller (power supply, local hard drive (RAID-1)), NIC (recommended: load balanced).
- Due to operation time of 24/7 and prevention of down-times a second domain controller is recommended or running on virtual machine with redundant hosts.

Further information for security aspects is available in PPT introduction "Concept for Production Security / Fallback Scenario".

The following overview of requirements are just for an initial general idea.

The concrete requirements for the version used should be checked on Microsoft website.

Domain server requirements

CPU	Minimum: Intel® Xeon® E3 (or similar) Recommended: Intel® Xeon® E5 (or similar)
RAM	Minimum: 4 GB RAM required (by Microsoft)
NIC	Minimum: 1 Gbit/s
Screen	No local screens are required on these machines (except for administrative work).
OS *	Versions: Windows Server: 2012 recommended 2016 Processor architecture: 64 bit Version type: Standard or higher Language: only English OS supported
Domain	Usage of separate subdomain / independent domain for dW production recommended (dW Pro and Enterprise).

*) For installations and updates a completely updated operating system is required. All current Windows service packs have to be installed.

3.2.2 Database server

The SQL server is the heart of each docWizz environment. The SQL server just hosts administrative data (list of all documents in pool, history of processing tasks, etc.). The SQL server does not contain any image data or the whole description of each documents. It just contains the main meta information such as the document ID, project name, document type and title. All further information is stored on file system in the POOL share.

Without the databases no production is possible (single point of failure). Without the SQL server you cannot access any document or get any information about the document processing history.

While for any document access and for all history reporting the SQL server is needed, a high performant SQL Server response for the dWClients and servers is mandatory for a smooth operation.

Following docWizz production considerations:

- Due to operation time of 24/7 and prevention of down-times, the SQL server should be clustered or running on virtual machine with redundant hosts.
- Also hardware redundancy is recommended to prevent any loss of the SQL server (power supply, hard drive [RAID-1], NIC [recommended: load balanced]).
- Due to similar requirements for redundancy / fallback of the SQL server to the Domain Controller, it makes sense to set it up together on the domain controller.
- For large scale environments SQL Server Express might be not sufficient due to CPU limitations and the amount of parallel database requests. (details available on Microsoft site or see FAQ on docWizz global site)
- In small and medium scale environments, the SQL databases are not under high demand and could be located on existing SQL server. But due to easy monitoring reasons and prevention of performance interferences with other databases it is easier to install a separate SQL server Express Edition.
- Starting with version 7.1, docWizz supports also TLS 1.2 and it is recommended to use it for SQL Server connection

For further information for security aspects see introduction PPT "Concept for Production Security / Fallback Scenario"

The following overview of requirements is just for an initial, general idea.

The concrete requirements for the version used should be checked on Microsoft website.

Database server requirements

CPU	Minimum: Intel® Xeon® E3 (or similar) Recommended: Intel® Xeon® E5 (or similar)
RAM	12 GB RAM recommended for combined server (8 GB for dW processing + 4 GB for domain services and database services) 500 MB RAM required for large dW databases.
NIC	Minimum: 1 Gbit/s
Screen	No local screens are required on these machines (except for administrative work).
OS *	<p>Versions: Windows Server Minimum: 2012 Recommended: 2016</p> <p>Processor architecture: 64 bit</p> <p>Version type: Standard or higher</p> <p>Language: English</p>
Domain	SQL Server must be member of the Microsoft domain
SQL server	<p>MS SQL server version: Minimum 2012 Recommended 2014 / 2016</p> <p>Version type: For dW edition Basic / Pro : Express Edition (for free) For dW edition Enterprise: Standard Edition or higher required (see comment above)</p>
Firewall	MS Firewall needs to be disabled or opened for full ODBC and UNC connection access
dW environment	All machines / domain user need permanent access to this SQL server

*) For installations and updates a completely updated operating system is required. All current Windows service packs have to be installed.

3.2.3 File server

As File server(s) we consider any server hosting the volumes for the docWizz shares.

This can be a Windows server with shared folders on UNC paths up to a professional NAS (network attached storage) for high-performance operations.

Following docWizz production considerations:

All shares need to be accessible from any docWizz machine (Servers and operator machines) at any time.

Please note: SAN storage, with assignment of storage to individual processing servers is not applicable (see requirement above).

If SAN is used, the server having assigned the storage capacities of the SAN will operate as file server and share these locations to all processing servers / docWizz client machines.

The required storage capacities depend on the following facts:

- Scanned dimensions, resolution, color depth, image format (compressed / uncompressed)
- Daily incoming data / scanned pages
- Turn-around-time for docWizz processing / holding time of the scanned / processing data

In case help for storage calculation is required, the above mentioned criteria needs to be provided / estimated as good as possible.

The speed of file servers depends mainly on the type and number of discs. For docWizz operations, we can differentiate between two categories:

- IN & OUT share: less traffic and more sequential writing (-> less fragmentation)
- dWShare & POOL: more parallel traffic and high fragmentation

Based on these categories, different kinds of storage can be used and costs can be reduced for large scale environments. High-end discs are not required for all storage.

See detailed comments in the table below.

Due to operation time of 24/7 and prevention of down-times, the file server(s) should fulfill highest uptime requirements.

Also, local redundancy is recommended to prevent loss of the file server(s) (power supply, hard drive [RAID-1], NIC [recommended: load balanced]).

File server requirements

	Basic	Pro	Enterprise
Storage type	Microsoft File server Also possible Linux based storage		Recommended: High availability storage
Disc types	Due to less parallel I/O non-high-speed storages are sufficient	dW Share and pool shares have high random I/O demands. Here high-speed storages are required. IN and OUT share(s) have more sequential write/read operations and are less frequented.	
NIC	Minimum: 1 Gbit/s Recommended: network trunk of 2 Gbit/s	Minimum: network trunk of 2 Gbit/s Recommended: trunks with 4 Gbit/s	Minimum: 4 Gbit/s Recommended: 10 Gbit/s
dW environment	CIFS share required (no.delayed write, case-insensitivity) NTP (network time protocol) required		

3.2.4 Firewall

As this is an essential requirement for all docWizz processing servers, docWizz client machines, file servers and SQL servers, we have summarized the requirements for the firewall settings in this separate chapter.

CCS recommends having firewalls enabled for security reasons, but certain rules need to be configured according to the following table:

	Port	Source	Destination
File share access	TCP/445, TCP/139	dW processing server dW operator machine	<file server(s)>
RemoteOCR	TCP/445	dW processing server / dW operator machine	<dW processing servers with OCR engine>
MS SQL server	TCP/1433-1434 TCP/2382 (SQL Browser)	dW processing server dW operator machine	<SQL server>
Z39.50 *)	TCP/7090, TCP/7091	dW processing server, dW operator machine	<catalog server>

*) optional, if Z39.50 metadata interface is used

4 Appendix

4.1 Calculation

To provide a figure on which formats / image formats can be processed now by docWizz, here is a calculation for uncompressed images sizes, according to "DIN A" formats. docWizz can process images with the maximum size of 600 MB (marked in green color).

		DIN format	A4	A3	A2	A1	A0
		height x width (cm)	29,8 x 21	42 x 29,8	59,6 x 42	84 x 59,6	119,2 x 82
		height x width (inch)	11,73 x 8,27	16,54 x 11,73	23,46 x 16,54	33,07 x 23,46	46,93 x 33,07
resolution 300 dpi	color depth		ca. 8525,32 MP	ca. 17050,64 MP	ca. 34101,28 MP	ca. 68202,55 MP	ca. 136405,12 MP
b/w	2		2,08 Mb	4,16 Mb	8,33 Mb	16,65 Mb	33,30 Mb
gray	8		8,33 Mb	16,65 Mb	33,30 Mb	66,60 Mb	133,21 Mb
color	24		24,98 Mb	49,95 Mb	99,91 Mb	199,81 Mb	399,62 Mb
color (BMP) *	32		33,30 Mb	66,60 Mb	133,21 Mb	266,42 Mb	532,83 Mb
			ca. 15156,13 MP	ca. 30312,25 MP	ca. 60624,45 MP	ca. 121248,99 MP	ca. 242497,98 MP
resolution 400 dpi	color depth		ca. 15156,13 MP	ca. 30312,25 MP	ca. 60624,45 MP	ca. 121248,99 MP	ca. 242497,98 MP
b/w	2		3,70 Mb	7,40 Mb	14,80 Mb	29,60 Mb	59,20 Mb
gray	8		14,80 Mb	29,60 Mb	59,20 Mb	118,41 Mb	236,81 Mb
color	24		44,40 Mb	88,81 Mb	177,61 Mb	355,22 Mb	710,44 Mb
color (BMP) *	32		59,20 Mb	118,41 Mb	236,81 Mb	473,63 Mb	947,26 Mb
		ca.	34101,28 MP	68202,56 MP	136405,12 MP	272810,23 MP	ca. 545620,47 MP
resolution 600 dpi	color depth	ca.	34101,28 MP	68202,56 MP	136405,12 MP	272810,23 MP	ca. 545620,47 MP
b/w	2		8,33 Mb	16,65 Mb	33,30 Mb	66,60 Mb	133,21 Mb
gray	8		33,30 Mb	66,60 Mb	133,21 Mb	266,42 Mb	532,83 Mb
color	24		99,91 Mb	199,81 Mb	399,62 Mb	799,25 Mb	1598,50 Mb
color (BMP) *	32		133,21 Mb	266,42 Mb	532,83 Mb	1065,66 Mb	2131,33 Mb

*) BMP has 8 bit reserved per pixel. This is used by MS operating system - and due to that - is also needed in some cases in docWizz (e.g. GUI)

4.2 docWizz site definitions

Single Site versus Multi-Site	<p>A site means a "Single-Location" installation in the terminology of CCS. A multi-site means a "Multi-Location" installation in the terminology of CCS.</p>	
Definition "Single-Location"	<p>One single docWizz configuration share <dWShare>, which contains the central configuration:</p> <p>Configuration files: ..\config Project config files: ..\proj-cfg TCL script files: ..\scripts and others</p> <p>A "Single-Location" can still contain multiple file shares pool shares: Pool1, Pool2 IN shares: IN-1, IN-2 OUT shares: OUT-1, OUT-2 and others</p>	
Definition "Multi-Location"	<p>For each site a docWizz configuration share is existing, each having separate central configuration:</p> <p><dWShare> one e.g. \\Server1\dWshare Configuration files: ..\config Project config files: ..\proj-cfg TCL script files: ..\scripts and others</p> <p><dWShare> two e.g. \\Server2\dWshare Configuration files: ..\config Project config files: ..\proj-cfg TCL script files: ..\scripts and others</p>	<p>Each site can have multiple IN, OUT and POOL shares, but they need to be physically separated. Means Site 1 and Site 2 cannot use the same shares.</p> <p>For more details about pro and cons of multiple environments see presentation 35-dW_MultipleEnvironments</p>
Why and when are "Multi-Location" setups needed?	<p>There are different geographical locations involved. There are different LAN (Local Area Network) environments involved Both options infer a slow connection between the dWClient and services and the docWizz share. This means at least a decrease of performance, if not even a disruption and unreliability of the system.</p>	
What are the side effects of "Multi-Location" setups?	<p>Since the detailed layout and scope cannot be put in a simple template for calculation purposes, a project consultation for "Multi-Location" is highly recommended, even required. It always will be a customized project pricing, which needs to be planned and executed in the project plan.</p>	

4.3 Glossary

	Functionality / Description
Domain server(s)	<p>Administrative server running Microsoft domain services.</p> <p>docWizz needs to operate in Microsoft domains. For this, all basic domain services need to be provided (domain controller, DNS, [DHCP], NTP service, ...). Microsoft workgroup is not supported.</p> <p>It is not required to set up an isolated domain for docWizz operation.</p> <p>An existing Domain can be used or a subdomain can be created.</p> <p>docWizz needs to operate through the network and access other UNC path locations (e.g. \\Servername\Path). For this, at least one explicit domain user is required for the dWServices.</p> <p>All processing servers and operator workstations need to be members of the same domain.</p> <p>Single exception possible is an authenticated subdomain, where domain users can get access permission across the domain borders.</p> <p>NTP (network time protocol) needs to distribute the time to ensure validated file server access from any machine.</p>
Database server	<p>SQL server hosting administrative production data.</p> <p>docWizz operates on multiple SQL databases for fast status verification, reporting and monitoring.</p> <p>In SQL databases, no images or other content data is stored. Only administrative information is recorded here.</p> <p>SQL server needs to be authorized in Microsoft domain and accessible from any docWizz machine (network access).</p> <p>The database is accessed either by SQL user (one fixed user for all – recommended) or on the domain account the docWizz applications are executed.</p>
File server / storage	<p>Storing configuration files and processing data.</p> <p>The concept of docWizz is to have all source images and intermediate data accessible by any processing server and workstation. This ensures there can be no single point of failure on individual computers and provides high flexibility in production management and prioritizing.</p> <p>Access to the storages has to be provided by UNC (Universal Naming Convention) paths. Mapped network drives and substitutions are not supported. For Non-Microsoft storages the shares need to be provided as CIFS shares (Common Internet File System) with "non- case sensitive" file system.</p> <p>The main data shares are dWShare (dW configuration), IN share(s) (source images), POOL share(s) (intermediate analysis data), OUT share(s) (final output). It is recommended to split these four areas into individual shares. For medium and large environments this is required due to parallel I/Os on the storage. For large environments, multiple shares for IN or POOL are generally required.</p> <p>The required capacities for the shares need to be calculated based on the individual digitization throughput.</p>

<p>Processing servers (dWServices)</p>	<p>Performing the automated processing steps – operational 24/7</p> <p>The dWServices are installed on processing servers; run 24/7 as a Windows service.</p> <p>The dWServices are monitored by a separate application called "docWizz ControlCenter". This application is part of the docWizz installation and can be started from any dW server or operator workstation.</p> <p>On each server a maximum of four dWServices can operate in parallel. Only in exceptional cases fewer dWServices are installed.</p> <p>The amount of servers and total amount of dWServices are limited by license keys.</p> <p>No production data is stored on these processing servers – local hard drives only hold the installed dW binaries and are used for temporary data swapping.</p>
<p>Operator workstations (dWClient)</p>	<p>Processed data is verified here (quality assurance and manual correction)</p> <p>The docWizz user interface applications are installed on operator workstations (dWClient, dWControlCenter, dWScanClient).</p> <p>These machines are only operating during office hours in general. No automated tasks or data transfer through the network is required here in out-of-office hours.</p>
<p>Scanner computers (dWScanClient)</p>	<p>Scanners can be integrated into the docWizz workflow by using the application "dWScanClient"</p> <p>On integration of scanning operations into docWizz workflow the user interface "dWScanClient" runs on the scanner computers, takes over the images and store additional metadata.</p> <p>dWScanClient performs the upload of the scanned images in the background after completion of the current scanned item.</p> <p>Upload tasks require a ScanClient application to be running. The scanner computer can be kept running after scan operation and shut down automatically after completion.</p>
<p>OCR</p>	<p>OCR=Optical Character Recognition</p> <p>OCR enables searching of large quantities of full-text data, but it is never 100% accurate. The level of accuracy depends on the print quality of the original book, its condition at the time of microfilming or scanning, and the level of detail captured by the microfilm scanner. Books with poor quality paper, small print, mixed fonts, multiple column layouts, or damaged pages may have poor OCR accuracy.</p>
<p>Z39.50</p>	<p>docWizz can be configured to import bibliographic metadata from your library catalog (via Z.39.50 interface).</p>
<p>dW RemoteOCR service</p>	<p>RemoteOCR services are sharing ABBYY FineReader licenses with servers and clients which do not have a local OCR license available.</p> <p>For further information and details about the functionality and benefits, see PPT introduction "RemoteOCR-solution".</p> <p>Remote OCR services run parallel to dWServices on all processing servers where a FineReader license is installed. The services do not have extra hardware requirements and, due to this, are not specifically outlined in system requirements.</p>
<p>dW RemoteQA services</p>	<p>RemoteQA solution is an option to allow quality control and manual correction of digitized material to be separated from processing site.</p> <p>For further information and details about the functionality and benefits, see PPT introduction "Environmental structure".</p> <p>The differences in requirements within a RemoteQA solution, additional aspects and requirements to be considered can be found in chapter RemoteQA solution.</p>

Virtual environments / machines (VMs)	CCS technology can also be used with VMs. However, since VM technology is comprised of a vast array of different types and technical options, the compatibility of CCS software with each productive environment option cannot be fully tested and assured. If our customers use VMs, it is their responsibility to set up and administrate the VMs, while CCS is responsible only for setting up the docWizz environment on these machines. In addition, CCS reserves the right to refrain from supporting operation problems obviously caused by an improper setup of a VM environment.
--	---

4.4 Special recommendations for docWizz 7.1 or higher versions

In case you would like to use special settings for layout analyze improvements using machine learning approach is recommended to use for docWizz services machines more powerful processors or even better graphical cards for GPU processing (even processing works also with low resources, the differences in processing time are very big). The recommendation for more powerful processors is valid also when using Tesseract 4.1 as OCR engine

The machine learning prediction can run in four modes, depending on available processor/graphical card.

Processor dependencies:

- We support following types of processors:

- a. „No AVX“
- b. „With AVX“
- c. „With AVX2“

AVX stands for „Advanced Vector Extensions“

(https://en.wikipedia.org/wiki/Advanced_Vector_Extensions):

The ML prediction will run with the greatest version of AVX available (automatically selected).

The slowest type of prediction is made on processors with „no AVX“, while the fastest prediction on processor is made on processors „with AVX2“.

If an NVIDIA graphical card is available and CUDA v9.0 or higher installed on that machine then prediction will run on GPU and the processing time is considerably smaller compared with prediction made on CPU, even if the CPU has AVX2 capabilities.

Our recommendations for graphical cards would be the following:

- Cost-efficient but expensive: RTX 2070
- Cost-efficient and cheap: RTX 2060, GTX 1060 (6GB).
- Low budget: GTX 1060 (6GB)
- Very low budget: GTX 1050 Ti (4GB)
- Expensive and not recommended due to poor benefits/cost ratio: Any Tesla card; any Quadro card; any Founders Edition card; Titan RTX, Titan V, Titan XP

Tesseract engine is built in two versions (“AVX”, “AVX2”), the proper version is selected at runtime. AVX2 version, when processor allows it, offer better performance. Processors without AVX instructions set are not supported by Tesseract 4.1 engine. docWizz client/docWizz processing services can still run on machines without AVX, but local Tesseract OCR will not be available on these machines.

Copyright © 2022 CCS Content Conversion Specialists GmbH. All rights reserved.

No part of this publication may be reproduced, stored in databases, or transferred in any form (electronically, photo-mechanically, chemically, manually, or otherwise) without the express written permission of CCS Content Conversion Specialists GmbH. The software described in this manual is licensed software that may be used only in compliance with the licensing terms and conditions. CCS GmbH reserves the right to make changes to the content of this manual without notice. CCS GmbH makes no guarantee regarding the accuracy of the information provided in this manual. Microsoft, and Windows are registered trademarks of the Microsoft Corporation.

Product or company names that are mentioned may be trademarks or registered trademarks of the respective company. CCS GmbH uses these names and trademarks in the following manual merely for explanatory purposes and for the benefit of the respective user, and such use does not imply trademark infringement.

Under this software license, you are only permitted to reproduce materials that are not protected by copyright laws. This excludes only materials where you hold the copyright and/or legal permission to reproduce copyrighted materials. If you are uncertain about the copyright status of certain materials then please seek legal counsel. CCS GmbH holds no liability over copyright violations resulting from the use of this software.

Last updated: 01/03/2022

CCS Content Conversion Specialists GmbH

Weidestraße 134

22083 Hamburg, Germany

Phone: +49-(0)40-228582990

Fax: +49-(0)40-228582998

E-Mail: info@content-conversion.com

Website: www.content-conversion.com