



enaio®

System Sizing

Version 10.0

Last update: 8/1/2022

All software products and all related programs and functions are registered and/or used trademarks of OPTIMAL SYSTEMS GmbH, Berlin, or one of its companies. They may only be used with a valid license agreement. The software and its associated documentation are protected by German and international copyright law. Unauthorized duplication and sales is plagiarism and shall be subject to criminal prosecution. All rights reserved, including reproduction, transmission, translation, and storage with/on all kinds of media. For all preconfigured test scenarios or demo presentations: all company names and persons appearing in examples (screenshots) are fictitious. Any similarities to existing companies or persons is purely coincidental and unintentional.

Care has been taken in the compilation of text and illustrations. However, we cannot completely exclude errors. The publisher and author cannot accept any liability or legal responsibility for incorrect or incomplete information or for the consequences thereof. Only individual quotes are binding.

Code examples contained in documentations serve to clarify functionalities and cannot be adopted without reviewing and adjusting them. Examples of code are shown with line breaks, for example, which are not allowed in the respective programming language, for a clearer presentation. For this reason, OPTIMAL SYSTEMS does not assume any liability for examples of code.

Copyright: 1992-2022

OPTIMAL SYSTEMS GmbH

Cicerostraße 26

D-10709 Berlin

Table of Contents

enaio® – System Sizing	4
Factors that Influence Sizing	5
Users	5
Volume of Objects	5
Interfaces Used	6
Criteria for Metrics	7
Reference Documents	8
Best Practices	14
Application Server	14
enaio® documentviewer	14
enaio® fulltext	15
Classify	16
Examples of Different Systems	18
System 1 – small to medium-sized	18
System 2 – medium-sized, small number of users, data storage, full text	20
System 3 – medium-sized, primarily web, failsafe	23
System 4 – medium-sized, bit of everything	26

enaio® – System Sizing

In addition to the question of system requirements, there is another important question you need you ask yourself when planning to roll out or update a complete enaio® system: What are the hardware and software requirements for the type of implementation you plan to ensure the system works properly? On the pages to follow, we will focus primarily on providing system integrators, consultants, administrators, and technical support staff with a guide that will support them in the planning phase and help them to more effectively assess and answer the central question of system sizing.

The general definition of **sizing** we use is as follow: the sizing of an entire enaio® system including all related components, taking into account other aspects that must be factored in. The goal at all times should be to provide a rationale for why the hardware and software to be provided are needed.

The specifications on system sizing described on the pages to follow are based on experience. They can be used as the basis for sizing a customer system. To make a detailed statement about your system, information on the user and quantity structure as well as deployment scenarios will be required, among other things. Please contact your customer representative at OPTIMAL SYSTEMS for more information. We look forward to seeing you at one of our sizing and scaling workshops.

Factors that Influence Sizing

There are a number of factors that influence the sizing of the overall enaio® system. They are broken down into subcategories that you consider on their own. The objective is to obtain detailed numbers, data, totals, and time periods that form the basis for realistic sizing of the system.

- "Users" unten
- "Volume of Objects" unten
- "Interfaces Used" auf der nächsten Seite
- "Criteria for Metrics" auf Seite7
- "Reference Documents" auf Seite8

Users

A user as the person at his or her workstation is not the same as a session of the user on the application server opened by different enaio® components ([interfaces](#)). It is therefore important to carefully consider possible user scenarios beforehand.

Make an estimate of the number and type of clients used (for example, 100 users for enaio® webclient plus Office integration).

i Number of users (== real users, not technical users or interfaces) and the relevant clients used (enaio® client, enaio® webclient, Outlook/Office integration)

Volume of Objects

For optimal sizing, in addition to the actual objects in the system, consider, among other things, nested register structures in folder objects that must also be managed (document, metadata, index/control data).

Calculate the following values in the units shown in brackets.

These values provide a rough estimate of the amount of data. You can make a rough estimate or use a higher amount, for example, "20 GB of documents per year plus 100 GB of legacy data from the outset. Roughly 20 percent of the data is image objects". Please note, however, that the more

precisely you define the amount of data, the more accurately you will be able to estimate the number of objects. This could look something like this: "40 of the 120 GB/year are e-mails sent via enaio® exchange, 40 GB are from SAP, 20 GB are from daily/incoming e-mails, and the remaining 20 GB are received via other interfaces (enaio® appconnector); incoming e-mails are image files only (tiff/jpg)".

- **Existing data volume:** Number of objects/metadata/index data/control data/documents (size and quantity in GB/TB)
- **Legacy data to be imported:** Is there a legacy archive that needs to be imported in full or a filesystem share that has to be transferred? (GB/TB)
- **Periodically added data:** Hourly jobs, daily imports, monthly statements, input from interfaces, etc. Is enaio® repository-manager or enaio® exchange in use? (GB/TB per unit of time)
- **Allocation of documents to categories:** Categorization as image material (scanned, without a defined text) or searchable materials. You can do this by taking random samples or making a conservative estimate. (in % of volume of data)

Interfaces Used

In addition to the users working in the system, connected interfaces or enaio® components open additional sessions on the application server and therefore are a factor in ensuring you correctly size your system.

Define the interfaces used (yes/no) and the number of each interface (for example, two portals connected via enaio® appconnector).

- Core services (enaio® appconnector, enaio® documentviewer, enaio® fulltext, enaio® webclient, etc.)
- Imports and exports (data and document interaction of any type)
- Server API applications (special project solutions or self-designed interfaces that exchange information directly with the server without relying on the known interfaces)
- Portals (mostly connected via enaio® appconnector, though they can also be implemented differently)
- Server-to-server components (MailArchiver, enaio® exchange, SmartFix, OSC, etc.)

Criteria for Metrics

Consider which of the following criteria apply to your application scenarios in the entire enaio® system (and to what extent they apply): We define applications as clients, core services, or interfaces. This should allow you to create a diagram of the entire enaio® system, taking all the criteria described above into account, and to fill it with data flows and components. You should therefore also know the quantities and/or expected/estimated times.

Is there direct interaction between the user and the application (enaio® gateway/enaio® webclient/...)?

- How is the system being used? Is it used solely for storing data or for client searches? Or do users edit data in the system?
- How many users access the application and start searches or jobs in it?
- How often are jobs started (per minute or hour)?
- What speeds are required?

Does the application need to process a specific volume of documents within a defined time period?

- What volume of documents does this entail (in GB, volume/hour, day, unit of time)?
- Is the volume in a queue? Is asynchronous processing possible, or do users wait their "turn" for the results?
- Where are bottlenecks expected? (In addition, does the application have to first fetch the data from a remote source which has a slow connection?)
- Are there peak load times? Are the same application response times required at both high and low loads?

Are there options required in enaio® that influence the application?

For which object types in enaio® do

- page numbers,
- previews,
- thumbnails,
- and text extracts,

need to be created by enaio® documentviewer?

Taking into account the criteria specified above, you should now create a diagram of the entire system and fill this with data flows and components. The quantities and/or times should also be known.

Reference Documents

Reference documents are defined to improve the quality of load test scenarios and make it possible to compare them. Please note that this page is currently under construction and makes no claims to being complete.

- PDF (version 1.4), coded information, file size = 200 KB (three pages, A4, continuous text)
- EML, file size = 100 KB, continuous text, without embedded images or attachments
- EML, continuous text: approx. 100 KB, without embedded images, with attachments (as per reference list) max. 5 MB
- DOC/DOCX, file size: 200 KB, A4, continuous text
- XSLX, file size: 200 KB
- TIFF, file size: 150 KB, b/w, 300 dpi
- TXT, file size: 80 KB

Hinweis: Please note that the system currently only supports file extensions that do not exceed ten characters in length. Longer file extensions are truncated automatically.

Hinweis: Please note that the maximum supported file size depends on a number of different factors. This includes the file type, the file content, the specific use case, and the client used.

Information on the Specified File Types

The test library contains a range of different file types. The table below shows which actions have been tested. The maximum supported file size for each action depends on a number of different factors. This includes the file type, the file content, the specific use case, and the client used.

General Formats

Description/ file extension	Main type	Content type recognitio n	Text extractio n	PDF generatio n	PDF/A-1b generatio n	Preview
Plain text files (.txt)	W-Document	✓	✓	✓	✓	✓
E-mail message text (.eml)	E-mail	✓	✓	✓	✓	✓
Portable Document Format (.pdf)	W-Document	✓	✓			✓
Portable Document Format (.pdf)	W-Document, image (B&W/gray/col or)	✓	✓	✓	✓	✓
Comma- separated values (.csv)	W-Document	✓	✓	✓	✓	✓
Extended Markup Language (.xml)	W-Document	✓	✓	✓	✓	✓

Microsoft Formats

Description/file extension	Main type	Content type recognition	Text extraction	PDF generation	PDF/A-1b generation	Preview
MS Word 97–2003, Word 2010 (.doc)	W-Document	✓	✓	✓	✓	✓
MS Word 2010 (.docx)	W-Document	✓	✓	✓	✓	✓
Microsoft Office Word OOXML format (.docx)	W-Document	✓	✓	✓	✓	✓
MS Excel 97–2003 (.xls)	W-Document	✓	✓	✓	✓	✓
MS Excel 2007 spreadsheet (.xlsx)	W-Document	✓	✓	✓	✓	✓
Microsoft PowerPoint 97–2003 (.ppt)	W-Document	✓	✓	✓	✓	✓
Microsoft PowerPoint 2010 (.pptx)	W-Document	✓	✓	✓	✓	✓
Microsoft Exchange Message (.msg)	E-mail	✓	✓	✓	✓	✓

Description/file extension	Main type	Content type recognition	Text extraction	PDF generation	PDF/A-1b generation	Preview
Microsoft Project 97–2003 (.mpp)	W-Document	✓	✓	✓	✓	✓
Microsoft Project 2010 (.mppx)	W-Document	✓	✓	✓	✓	✓
Microsoft Visio 97–2003 (.vsd)	W-Document	✓	✓	✓	✓	✓
Microsoft Visio 2010 (.vsdx)	W-Document	✓	✓	✓	✓	✓

OpenOffice Formats

Description/file extension	Main type	Content type recognition	Text extraction	PDF generation	PDF/A-1b generation	Preview
OpenDocument Text (.odt)	W-Document	✓	✓	✓	✓	✓
OpenDocument Spreadsheet (.ods)	W-Document	✓	✓	✓	✓	✓
OpenDocument Presentation (.odp)	W-Document	✓	✓	✓	✓	✓

Image Formats

Description/file extension	Main type	Content type recognition	Text extraction	PDF generation	PDF/A-1b generation	Preview
Encoded TIFF (-tiff/.tif)	W-Document, image (B&W/gray/)	✓	✓	✓	✓	✓
Tagged Image File Format (.tiff/.tif)	W-Document, image (B&W/gray/)	✓	✓	✓	✓	✓
Portable Network Graphics (.png)	W-Document, image (B&W/gray/color)	✓	✓	✓	✓	✓
JPEG file (.jpg, .jpeg)	W-Document, image (color)	✓	✓	✓	✓	✓
Windows Bitmap (.bmp)	W-Document, image (B&W/gray/color)	✓	✓	✓	✓	✓

Container Formats

Description/file extension	Main type	Content type recognition	Text extraction	PDF generation	PDF/A-1b generation	Preview
ISO image (.iso)	W-Document	✓				
ZIP file format (.zip)	Container	✓				
RAR file format	W-Document	✓				
Audio Video Interleave (.avi)	Movie	✓				✓

Best Practices

To help make your work easier, we draw on past experience to provide you with examples of best practices.

- "Application Server" unten
- "enaio® documentviewer" unten
- "enaio® fulltext" auf der nächsten Seite
- "Classify" auf Seite16

Application Server

System	Number of enaio® users
32-bit server: enaio® 8.00 to 9.10	max. 250 logged-in clients per server with 50 concurrent users
64-bit server: from enaio® 9.00	max. 400-500 logged-in clients per server with 150-200 concurrent users

The actual figures depend on the user behavior and on whether the server is performing other activities (such as imports or the like).

Additional application servers can be used for technical tasks such as imports, archiving, workflow processing, etc. This can be used to separate the system load from the user load. When choosing the number of servers, you also need to consider system stability, meaning the remaining servers must be able to support the total number of users.

Always use the specified number of users for enaio® gateway, enaio® appconnector, and enaio® webclient.

enaio® documentviewer

You can determine the size of the cache area required by enaio® documentviewer using the following formula:

i „Erwartetes **Dokumentenvolumen** pro Monat“ (Umrechnung in Größe in GB) mal „Erwartete sofortverfügbare Vorschauen im Client in Monatsalter“ mal 2 = Empfehlung für die minimale HDD-Größe des vorzuhaltenden Cache für enaio® documentviewer.

Also: $\text{Volumen}_{(\text{GB} / \text{Monat})} * \text{Vorschaualter}_{(\text{Anzahl Monate})} * 2 = \text{Cachegröße}_{(\text{GB})}$

According to the system requirements, 1 MB is required per document (eight pages) in the enaio® documentviewer cache. This makes it possible to make a projection based on the number of documents or pages.

Recommendation

enaio® documentviewer should be allocated to a separate host. You also need to check that the document volume per hour/day/week can be handled with the given specifications (1 CPU/512 MB RAM). This is not the case in most instances where periodic imports are run and users do not expect to receive the previews immediately. It is possible to deploy an additional enaio® documentviewer that generates and supplies previews ad hoc for users in such cases.

All databases, temp directories, and converters are always local and therefore run a large number of read and write operations on the hard drive. If several enaio® documentviewer are in use, the cache is shared by them all and can be located in the network (on a \\UNC\path).

enaio® fulltext

enaio® fulltext is based on the add-on software Elasticsearch. You estimate the size of the full-text database as follows:

25 to 30 GB HDD will be needed **for roughly one million objects** in the database (depending on document content and diversity). **An additional 25 to 30 GB** is required per replica. Replicas must be used in cluster installations (more than one Elasticsearch node). They are used to ensure data availability as well as the performance and hence scalability of enaio® fulltext.

The Elasticsearch database cannot be on \\UNC paths\ or on connected network drives. It must be located **locally**. Elasticsearch is always a cluster with at least one node. It is recommended that you always select an **odd** number of nodes when scaling Elasticsearch.

Scaling is recommended once the size of the database on the HDD (without replicas) exceeds the RAM available to Elasticsearch by a factor of ten. At this point, the cluster should be extended with max. 64 GB RAM per Elasticsearch node.

Working in the database places relatively high demands of HDD resources. It is not necessary to use SSD premium storage memory. Follow this general rule of thumb: The less RAM is available, the faster the HDD needs to be in order to keep pace with contents from the database.

Classify

enaio® uses the smart FIX component from Insiders to classify documents. Aside from the restrictions and guidelines set out by the software maker, the following recommendations are proven best practices:

smart FIX requires additional databases located on an MSSQL server. System sizing will vary depending on the daily volume of incoming data. At least 1 GB RAM and one CPU on the MS SQL server should be reserved for smart FIX. The classification component is deployed on a dedicated host and requires resources in line with the volume of documents. As a rough guide, it can be assumed that 20 to 30 seconds will be required to extract one A4 page (150 dpi) if you are using a 3 GHz CPU. This is a worst-case scenario, which is a reasonable assumption considering the scanned originals to be processed, the quality of the scanned originals, and other use of infrastructure. Processing typically takes less time. The specified RAM usage is 1 GB per CPU core.

Beispiel:

It is possible for one CPU to process roughly 20,000 pages per month.

Here's why: $20,000 \text{ page} / 20 \text{ working days} / \text{eight hours per working day} / 60 \text{ minutes} = 2.08 \text{ pages/minute}$; this can be handled by a single CPU core.

The example above assumes a highly constant flow of incoming documents. If large batches of documents are received by the system during certain weeks or months, the requirements will change based on the demands in terms of processing speed. As a further example, it is assumed that the documents generated during a month (20,000) need to be processed in a single day (eight hours).

In short: $20,000 \text{ pages} / \text{eight hours} / 60 \text{ minutes} = 41.7 \text{ pages/minute}$. 20 CPU cores will in turn be needed for processing by the classification system.

The key variable is thus the desired processing speed, taking into account the intermittency of incoming documents. Scaling must be carried out on this basis.

Examples of Different Systems

You will find detailed system requirements, sizing information, and calculations for each system, as well as other notes on the respective system below. A total of four systems are described here. This information provides basic parameters which will give you a general idea of the sizing options available. We would be happy to demonstrate the method we use to size your system based on the given parameters and provide additional information at a sizing and scaling workshop.

- [System 1 – small to medium-sized](#)
- [System 2 – medium-sized, small number of users, data storage, full text](#)
- [System 3 – medium-sized, primarily web, failsafe](#)
- [System 4 – medium-sized, bit of everything](#)

System 1 – small to medium-sized

Requirements

- MSSQL on the system
- Capture imports for five companies
 - 400 invoices per company (two pages on average) per month, 100% OCR required
- CSV imports, posting data for invoices. Once a day.
- Total volume: currently 20 GB, 10 GB added per year (WORK statistics)
 - 180,000 objects in the system
 - 1.8 million history entries
- Max. 15 parallel enaio® clients, on average eight, three power users
 - No Office integration
 - Decentralized rights system, approx. 75 object types
 - Large number of clauses

- enaio® webclient is requested
- Simple object definition: > 30 fields on index data forms in four instances, otherwise fewer fields.
- Invoice workflow: 75 processes a day, sequence based on imports. (currently two companies, will be five in future)
- No system stability
- No long-term archiving
- No e-mail integration or enaio® exchange
- Incl. document preview for all object types:
- Full text: in common use, fast response times expected
- Additional data volume: e-mails via drag-and-drop, 300/month
- Primarily imports via Capture, CSV imports, and accountants who search the system.

Sizing

Central Servers/Services

- Server 1: file server for WORK, CACHE, and DV cache
1:1 document volume + 100 GB

enaio® Applications

- Server 1: MSSQL, app server, enaio® gateway, enaio® appconnector, enaio® webclient
6 CPU, 8 GB RAM, HDD: E:\ 60 GB F:\ 50 GB
- Server 2: enaio® documentviewer, enaio® service-manager (search, index, OCR, license), Elasticsearch
6 CPU, 14 GB RAM, HDD: E:\ 100 GB F:\ 50 GB

Notes, Thoughts, and Calculation

- MS SQL Server is installed on the application server. enaio® receives its own database, schema, and login for this. The specified resources are still available on the instance.

MSSQL data and log files are located on the application server on F:\.

- WORK is centrally located – space requirement 1:1 in line with the generated object volume of imports.

The rendition cache is also located centrally, calculated here as a cache size of 50 GB, two-year cache retention time at 10 GB/year.

- The size of the C:\ partition may be freely selected on all servers.
- enaio® components receive their own partition (here: E:\ on all servers).
- One Microsoft Office instance must be installed on server 2 (enaio® documentviewer).
- Full text: 25 to 30 GB per one million objects. Approx. 300,000 objects currently in the system. Enough capacity to last a long time.
- Server 1 – CPU calculation:
1 Windows, 2 MSSQL, 1 app server, 0.5 enaio® gateway, 0.5 enaio® appconnector, 0.5 enaio® webclient, 0.5 buffer
- Server 1 – RAM calculation:
1.5 Windows, 3 MSSQL, 1.5 app server, 0.25 enaio® gateway, 0.25 enaio® appconnector, 0.5 enaio® webclient, 1 buffer
- Server 2 – CPU calculation:
1 Windows, 2 enaio® documentviewer, 1 enaio® service-manager, 1 Elasticsearch, 1 buffer
- Server 2 – RAM calculation:
1.5 Windows, 3 enaio® documentviewer, 3 enaio® service-manager, 6 Elasticsearch, 0.5 buffer

Not all instances of enaio® capture should import at the same time (since there is only one application server).

System 2 – medium-sized, small number of users, data storage, full text

Requirements

- Full text, with OCR
- No workflow
- No system stability
- No long-term archiving
- Imports – approx. 60,000 pages of already searchable PDFs per month
- SAP – approx. 300 receipts/day, non-searchable
- Incl. document preview, primarily from SAP
- Full text: in common use, fast response times expected
- No Capture
- Users: approx. 100 parallel users in enaio® client, of which ten power users, no enaio® webclient
 - Average-sized rights system, approx. 20 object types
 - Simple object definition: 30 fields on index data forms in one or two instances, otherwise fewer fields.
- Primary data sink for SAP. Search via enaio® client in the non-technical cabinets. Users use full-text search a lot; it must therefore be fast.
- Additional volume: approx. 150 MS Office documents are edited/created per day.

Sizing

Central Servers/Services

- Server 1: RDBMS (MSSQL)
2 CPUs, 8 GB RAM, HDD: 50 GB
- Server 2: file server for WORK, CACHE, and documentviewer cache
1:1 document volume + 100 GB

enaio® Applications

- Server 1: app server, OSR3, ServiceManager1 (search, index), Elasticsearch
6 CPU, 12 GB RAM, HDD: E:\ 60 GB F:\ 100 GB

- Server 2: enaio® gateway, enaio® appconnector, ServiceManager2 (search, index), Elasticsearch
4 CPU, 10 GB RAM, HDD: E:\ 60 GB F:\ 100 GB
- Server 3: enaio® gateway, ServiceManager3 (OCR)
6 CPU, 4 GB RAM, HDD: E:\ 60 GB

Notes, Thoughts, and Calculation

- MSSQL server is already available in the infrastructure. enaio® receives its own database, schema, and login for this. The specified resources are still available on the instance.
- WORK is centrally located – space requirement 1:1 in line with the generated object volume of imports/receipts.

The rendition cache is also centrally located, calculated here as a cache size of 50 GB, approx. 8,000 pages per day == approx. three to four month cache retention time. Additional 50 GB as buffer.
- The size of the C:\ partition may be freely selected on all servers.
- enaio® components receive their own partition (here: E:\ on all servers).
- One Microsoft Office instance must be installed on server 3 (enaio® documentviewer).
- The services are distributed roughly evenly. OSR3 and app server on the same host to make it possible to work against localhost.
- Elasticsearch with one replica on servers 1 and 2. enaio® documentviewer and OCR separate due to high number of read/write operations, separate LUN (F:\) for full-text databases.
- Server 1 – CPU calculation:
1 Windows, 2 app servers, 1 OSR3, 1 Elasticsearch, 1 enaio® service-manager, 0 buffer
- Server 1 – RAM calculation:
1.5 Windows, 3 app servers, 5 Elasticsearch, 2 enaio® service-manager, 0.5 buffer
- Server 2 – CPU calculation:
1 Windows, 0.5 enaio® gateway, 0.5 enaio® appconnector, 1 Elasticsearch, 1 enaio® service-manager, 0 buffer

- Server 2 – RAM calculation:
1.5 Windows, 0.5 enaio® gateway, 0.5 enaio® appconnector, 5 Elasticsearch, 2 enaio® service-manager, 0.5 buffer
- Server 3 – CPU calculation:
1 Windows, 4 enaio® documentviewer, 0.5 enaio® service-manager, 0.5 buffer
- Server 3 – RAM calculation:
1.5 Windows, 1.5 enaio® documentviewer, 0.5 enaio® service-manager, 0.5 buffer

Except for the more extensive requirements for enaio® fulltext, this is a standard scenario. An existing SQL server is shared. There is a central domain for filing, moderate document volume, and no special requirements.

If the SAP interface is not scheduled for weekly or monthly runs (i.e., the total amount of documents comes in several batches), expectations are that a reassessment will not be carried out until after three to five years.

System 3 – medium-sized, primarily web, failsafe

Requirements

- Remote work, 90 percent via enaio® webclient. Total of approx. 1,000 users. Small number of power users (~30) via enaio® webclient; these are central
- System stability, complete Webclient functionality is essential, full text and previews are not essential.
- Long-term archiving available
- Inventory: approx. 10 million objects in the system, primarily read-only/PDF /TIFF, tiny number of searchable objects.
- Central Capture installation, ~600 pages/day
- In addition: via REST interface from a portal ~1,200 pages/day
Resulting workflows: ~300 processes per day
- Complex rights system, approx. 50 object types, but many clauses due to tenant separation

and large number of departments.

- Response times for web users are the primary focus. The focus is on speed and stability.

Sizing

Central Servers/Services

- Server 1: RDBMS (MSSQL cluster)
4 CPUs, 12 GB RAM, HDD: 50 GB
- Server 2: file server for WORK, CACHE, and DV cache
1:1 document volume + 150 GB

enaio® Applications

- Server 1: app server, enaio® gateway, enaio® appconnector, enaio® webclient, enaio® service-manager (license)
8 CPUs, 8 GB RAM, HDD: E:\ 60 GB
- Server 2: app server, enaio® gateway, enaio® appconnector, enaio® webclient, enaio® service-manager (license)
8 CPUs, 8 GB RAM, HDD: E:\ 60 GB
- Server 3: enaio® documentviewer, enaio® service-manager (search, index, OCR), enaio® fulltext
12 CPUs, 16 GB RAM, HDD: E:\ 60 GB F:\ 500 GB

Notes, Thoughts, and Calculation

- An MSSQL server is already available in the infrastructure. enaio® receives its own database, schema, and login for this. The specified resources are still available on the instances. Four CPUs with more complex rights system, 12 GB resulting from ten million objects.
- WORK is centrally located – space requirement 1:1 in line with the object volume generated from Capture and by the REST interface.

The rendition cache is also located centrally. The calculated cache size here is 100 GB. Long cache retention time for rapid dispatch to enaio® webclient.

- The size of the C:\ partition may be freely selected on all servers.

- enaio® components receive their own partition (here: E:\ on all servers).
- One Microsoft Office instance must be installed on server 3 (enaio® documentviewer).
- An app server with eight CPUs is required due to the complex rights system, the number of workflows, and for long-term archiving.
- All services in the chain required for the enaio® webclient are duplicated or clustered.

Browser > Load balancer > Gateway > WebClient > AppConnector > Appserver > DB

- An external load balancer is required, which distributes 50 percent of the load to each of the two gateways.
- No special features with regards to enaio® documentviewer and enaio® fulltext. Scaled up in line with database, still to be performed individually for each instance. Drive F:\ has a 500 GB buffer for up to approx. 12 to 15 million documents for the full-text database (if the number of documents increases, the RAM must also be scaled up accordingly; if not, the response speed decreases).
- Due to high demand for OCR, four cores will be needed just for OCR. 1,800 pages could potentially be processed in an 8-hour workday (225 pages per hour, or 3.75 pages per minute).
- Server 1 – CPU calculation:
1 Windows, 3 app servers, 0.5 enaio® gateway, 2 enaio® appconnector, 1 enaio® webclient, 0.25 enaio® service-manager, 0.25 buffer
- Server 1 – RAM calculation:
1.5 Windows, 4 app servers, 0.5 enaio® gateway, 1 enaio® appconnector, 0.5 enaio® webclient, 0.5 buffer
- Server 2 – CPU calculation:
1 Windows, 3 app servers, 0.5 enaio® gateway, 2 enaio® appconnector, 1 enaio® webclient, 0.25 enaio® webclient, 0.25 buffer
- Server 2 – RAM calculation:
1.5 Windows, 4 app servers, 0.5 enaio® gateway, 1 enaio® appconnector, 0.5 enaio® webclient, 0.5 buffer

- Server 3 – CPU calculation:
1 Windows, 4 enaio® documentviewer, 6.5 enaio® webclient (*consisting of: 4 OCRs, 1 search and index, 1.5 other services*), 0.5 buffer
- Server 3 – RAM calculation:
1.5 Windows, 1.5 enaio® documentviewer, 3 enaio® webclient, 10 Elasticsearch, 0 buffer

System 4 – medium-sized, bit of everything

Requirements

- Imports in use, miscellaneous, REST (drop targets), XML, CSV, volume: ~1,000 pages per day, all are already available as PDFs with text information
- Master data exports for external systems as CSV or XML
- enaio® exchange for journaling and 100 users, total of ~2,500 e-mails per day
 - enaio® search in use for the users in question
- OSR3, one repository to SAP, ~400 pages per day, OCR required.
- Frequent AD sync for approx. 500 users in 20 or so groups
- enaio® capture with Kofax and OCR
- enaio® webclient, 200 users
- enaio® client, 100 users
- Full-text searchability, esp. for e-mails (incl. attachments, depth: 2) requested.
- Office and Outlook integration on client (NG add-ins)
- Usage: users make extensive use of hit lists, document preview, and searches within them. Hit lists are retrieved either via saved searches (enaio® webclient) or via the full-text search and dynamic searches (enaio® client).

Sizing

RECOMMENDATION 1

Central Servers/Services

- Server 1: RDBMS (MSSQL cluster)
2 CPUs, 6 GB RAM, HDD: 50 GB
- Server 2: file server for WORK, CACHE, and DV cache
1:1 document volume (~3 GB/day) + 500 GB (DV cache)

enaio® Applications

- Server 1: app server, enaio® appconnector, enaio® webclient, enaio® gateway, enaio® documentviewer(clients)
6 CPUs, 6 GB RAM, HDD: 80 GB
- Server 2: app server, enaio® appconnector, OSR3, WebServices, EEX, enaio® service-manager (EMS, license, extraction)
5.5 CPUs, 5 GB RAM, HDD: 80 GB
- Server 3: enaio® documentviewer (CPB), enaio® service-manager (OCR)
6 CPUs, 6 GB RAM, HDD: 120 GB
- Server 4: enaio® service-manager, (search, index, Elasticsearch)
2 CPUs, 6 GB RAM, HDD: 100 GB

RECOMMENDATION 2 (Lower Number of Windows Servers)

Central Servers/Services

- Server 1: RDBMS (MSSQL cluster)
2 CPUs, 6 GB RAM, HDD: 50 GB
- Server 2: file server for WORK, CACHE, and DV cache
1:1 document volume (~3 GB/day) + 500 GB (DV cache)

enaio® Applications

- Server 1: app server, enaio® webclient, enaio® gateway, enaio® documentviewer(client)
7.5 CPUs, 7.5 GB RAM, HDD: 80 GB
- Server 2: enaio® appconnector, EEX, enaio® service-manager(EMS, OCR, search, index, license, extraction), Elasticsearch
6 CPUs, 11 GB RAM, HDD: 80 GB (enaio® service-manager), 100 GB (VTX)
- Server 3: enaio® documentviewer(CPB), OSR3, WebServices
6 CPUs, 8 GB RAM, HDD: 120 GB

The app server could become a bottleneck. It is therefore necessary to increase/optimize number of threads and configuration (registry, periodic jobs, executors).

Notes, Thoughts, and Calculation

- Capture volume is missing: assumed volume: B/W scan with 1,000 pages/day
- Calculated size: $(1,000 \text{ [captures]} + 1,000 \text{ [imports]} + 400)/8 = 300 \text{ MB} + 2.5 \text{ GB}$ of e-mails per day → 2.8 GB of data per day
- DV load calculation: $(2,400 \text{ [imports]} + 10,000 \text{ [e-mails]})/\text{day (8 h)} \rightarrow 26 \text{ pages per minute}$
Plus OCR: 400/day plus 10% of e-mails (=1,000) 1,400 OCR pages per day → three pages per minute → one OCR core
- A second OCR core (license) for enaio® capture is required. OCR on the client workstation.
- An MSSQL server is already available in the infrastructure. enaio® receives its own database, schema, and login for this. The specified resources are still available on the instances.
No information on legacy data, complexity of rights system, or scripts; for that reason, two CPUs each with 6 GB RAM.
- WORK is located centrally. Space requirement 1:1 in line with the object volume generated from enaio® capture and by enaio® exchange.

The rendition cache is also locally centrally, calculated here with a cache size of 500 GB, long cache retention time for rapid dispatch to enaio® webclient and space required for e-mails from enaio® exchange.

- The size of the C:\ partition may be freely selected on all servers.

- enaio® components receive their own partition (here: E:\ on all servers).
- One Microsoft Office instance must be installed on servers 1 and 3 (enaio® documentviewer).
- No special features with regards to enaio® documentviewer and enaio® fulltext. Scaled up in line with database. Drive F:\ still has a buffer of 100 GB for up to approx. 12 to 15 million documents for the full-text database.

The first [recommendation](#) is the preferred option. Here, the load is more effectively distributed across the Windows servers. In addition, a separate application server is provided for the core services.



OPTIMAL SYSTEMS
A KYOCERA GROUP COMPANY

Head office

Cicerostraße 26

10709 Berlin

Phone: +49 30 895708-0

kontakt@optimal-systems.de