



# enaio® System sizing

Version 11.0

Letzter Stand: 20.12.2023

Sämtliche Softwareprodukte sowie alle Zusatzprogramme und Funktionen sind eingetragene und/oder in Gebrauch befindliche Marken der OPTIMAL SYSTEMS GmbH, Berlin oder einer ihrer Gesellschaften. Sie dürfen nur mit gültigem Lizenzvertrag benutzt werden. Die Software sowie die jeweils zugehörige Dokumentation sind nach deutschem und internationalem Recht urheberrechtlich geschützt. Das illegale Kopieren und Vertreiben der Software stellt Diebstahl geistigen Eigentums dar und wird strafrechtlich verfolgt. Alle Rechte vorbehalten, einschließlich der Wiedergabe, Übermittlung, Übersetzung sowie Speicherung mit/auf Medien aller Art. Für vorkonfigurierte Testszenarien oder Demo-Präsentationen gilt: Alle Firmennamen und Personen, die in Beispielen (Screenshots) erscheinen, sind frei erfunden. Eventuelle Ähnlichkeiten mit tatsächlich existierenden Firmen und Personen sind zufällig und unbeabsichtigt.

Bei der Zusammenstellung von Texten und Abbildungen wurde mit Sorgfalt vorgegangen. Dennoch können Fehler nicht vollständig ausgeschlossen werden. Herausgeber und Autor können weder für fehlerhafte oder unvollständige Angaben noch für deren Folgen irgendeine Haftung oder juristische Verantwortung übernehmen. Verbindlich sind ausschließlich individuelle Angebote.

In Dokumentationen enthaltene Codebeispiele dienen der Verdeutlichung der Funktionalität und können ohne Überprüfung und Anpassung nicht übernommen werden. Zur übersichtlicheren Darstellung sind Codebeispiele beispielsweise mit Zeilenumbrüchen abgebildet, die in der jeweiligen Programmiersprache nicht erlaubt sind. Aus diesem Grund kann OPTIMAL SYSTEMS keine Haftung für Codebeispiele übernehmen.

Copyright: 1992-2023

OPTIMAL SYSTEMS GmbH

Cicerostraße 26

D-10709 Berlin

# Inhaltsverzeichnis

|   |    |
|---|----|
| enaio® – System sizing  | 4  |
| Beeinflussende Faktoren   | 5  |
| Benutzer .....  | 5  |
| Objektmenge .....   | 5  |
| Verwendete Schnittstellen .....                                 | 6  |
| Messwertkriterien .....   | 7  |
| Referenzdokumente .....   | 8  |
| Best Practices  | 14 |
| Applikationsserver .....  | 14 |
| enaio® documentviewer .....                                     | 14 |
| enaio® fulltext .....   | 15 |
| Classify .....  | 16 |
| Beispielsysteme   | 18 |
| System 1 – klein bis mittel .....                               | 18 |
| System 2 – mittel, wenige Benutzer, Datenablage, Volltext ..... | 21 |
| System 3 – mittel, hauptsächlich Web, ausfallsicher .....       | 24 |
| System 4 – mittel, etwas von allem .....                        | 26 |

## enaio® – System sizing

Bei der Einführung oder beim Update eines enaio®-Gesamtsystems steht neben der Frage nach den Systemvoraussetzungen eine weitere Frage im Mittelpunkt der Planung: Welche Hard- und Software-Erfordernisse sind für meine Form der Implementierung erforderlich, um einen reibungslosen Betrieb gewährleisten zu können?

Im Folgenden möchten wir primär allen Systemintegratoren, Consultants, Administratoren oder technischen Supportmitarbeitern einen Leitfaden zur Verfügung stellen, der bei dieser Planung unterstützt und hilft, die zentrale Frage der Dimensionierung besser einschätzen und beantworten zu können.

Unter **Sizing** verstehen wir generell die Dimensionierung eines enaio®-Gesamtsystems mit allen beteiligten Komponenten unter Berücksichtigung der notwendigen Teilaspekte – immer mit dem Ziel, die bereitzustellende Hard- und Software begründen zu können.

Die auf den folgenden Seiten beschriebenen Angaben zum System sizing basieren auf Erfahrungswerten, die als Grundlage für die Dimensionierung eines Kundensystems verwendet werden können. Um spezifische Aussagen zu Ihrem System treffen zu können, werden u. a. Angaben zum Benutzer- und Mengengerüst sowie Einsatzszenarien benötigt. Bitte wenden Sie sich diesbezüglich an Ihren Ansprechpartner bei OPTIMAL SYSTEMS. Gern begrüßen wir Sie auch in einem von unseren Sizing- und Skalierungs-Workshops.

# Beeinflussende Faktoren

Es gibt verschiedene Faktoren, die das Sizing eines enaio®-Gesamtsystems beeinflussen. Die Faktoren sind in Teilbereiche untergliedert, die Sie im Einzelnen betrachten. Ziel ist es Erkenntnisse zu konkreten Zahlen, Daten, Summen und Zeiträumen zu gewinnen, die eine realistische Basis für ein Sizing darstellen.

- "Benutzer" unten
- "Objektmenge" unten
- "Verwendete Schnittstellen" auf der nächsten Seite
- "Messwertkriterien" auf Seite 7
- "Referenzdokumente" auf Seite 8

## Benutzer

Ein Benutzer als Person an seinem Arbeitsplatz ist nicht gleichzusetzen mit einer Sitzung des Benutzers am Applikationsserver, die durch verschiedenen enaio®-Komponenten ([Schnittstellen](#)) geöffnet wird. Dementsprechend ist es wichtig, im Vorfeld die möglichen Benutzer-Szenarien im Detail zu betrachten.

Schätzen Sie die Anzahl und die Art der verwendeten Clients (zum Beispiel 100 Benutzer für enaio® webclient plus Office-Integration).

- <sup>i</sup> Anzahl der Benutzer (= echte Personen, keine technischen Benutzer oder Schnittstellen) und die jeweils verwendeten Clients (enaio® client, enaio® webclient, Outlook-/Office-Integration)

## Objektmenge

Für ein optimales Sizing betrachten Sie neben den eigentlichen Objekten im System unter anderem auch verschachtelte Registerstrukturen in Ordner-Objekten, die ebenfalls verwaltet werden müssen (Dokument-, Meta-/Index- und Steuerdaten).

Ermitteln Sie die nachfolgenden Werte mit der jeweils in Klammern dargestellten Einheit.

Diese Werte geben Auskunft über Ihre „rohe“ Datenmenge. Sie können diese grob überschlagen und größer formulieren: Zum Beispiel „120 GB Dokumentvolumen pro Jahr plus 100 GB Alt-Daten von Beginn an, ca. 20 % der Daten sind Bild-Objekte“. Beachten Sie jedoch, dass je genauer die Datenmenge pro Weg definiert ist, desto besser wird die Einschätzung über die Objektmenge. Etwa: „40 von den 120 GB im Jahr werden E-Mails aus enaio® exchange sein, weitere 40 GB kommen aus SAP, 20 GB kommen aus Tages-/Eingangspost, die restlichen 20 GB durch weitere Schnittstellen (enaio® appconnector), Eingangspost ist ausschließlich Bildmaterial (tiff/jpg)“.

- **Bestehendes Datenvolumen:** Anzahl von Objekten/Meta-/Index-/Steuerdaten/Dokumenten (Angabe in Anzahl und Größe: GB/TB)
- **Zu übernehmende Alt-Daten:** gibt es ein Alt-Archiv, das komplett importiert werden soll oder ein Filesystem-Share, das überführt wird? (GB/TB)
- **Periodisch hinzukommende Daten:** stündliche Jobs, tägliche Importe, monatliche Abrechnungen, Input aus Schnittstellen etc. Ist enaio® repository-manager oder enaio® exchange im Einsatz? (GB/TB pro Zeiteinheit)
- **Aufteilung des Dokumentvolumens:** Unterscheidung zwischen Bildmaterial (gescannt ohne hinterlegten Text) und durchsuchbarem Material. Dies kann ggf. über Stichproben und pessimistische Schätzung ermittelt werden. (Angabe in % des Datenvolumens)

## Verwendete Schnittstellen

Neben den am System arbeitenden Benutzern öffnen angebundene Schnittstellen oder enaio®-Komponenten weitere Sitzungen am Applikationsserver und haben somit Einfluss auf das richtige Sizing.

Definieren Sie die verwendeten Schnittstellen (Ja/Nein) und für diese jeweils die Anzahl (zum Beispiel zwei Portale angebunden über enaio® appconnector).

- Importe und Exporte (Daten- und Dokumenteninteraktion jeglicher Art)
- Server-API-Verwendungen (spezielle Projektlösungen oder selbst designte Schnittstellen, die Informationen direkt mit dem Server austauschen ohne auf die bekannten Schnittstellen zurückzugreifen)
- Portale (meist über enaio® appconnector angebunden, kann jedoch auch anders umgesetzt

sein)

- Server-zu-Server-Komponenten (MailArchiver, enaio® exchange, SmartFix, OSC etc..)

## Messwertkriterien

Überlegen Sie, welche der folgenden Kriterien (in welchem Maße) auf Ihre Anwendungsszenarien im enaio®-Gesamtsystem zutreffen. Unter Anwendung verstehen wir Clients, Kerndienste oder Schnittstellen. Ziel ist es, dass Sie sich unter Berücksichtigung aller genannten Kriterien ein Bild des enaio®-Gesamtsystems machen können und dieses mit Datenflüssen und Komponenten füllen. Dementsprechend sollten Ihnen ebenfalls die Mengen und/oder erwarteten/geschätzten Zeiten bekannt sein.

Hat die Anwendung direkte Benutzer-Interaktion (enaio® gateway/enaio® webclient/...)?

- Wie wird das System genutzt? Dient es nur zur Datenablage und Suche über Clients? Oder bearbeiten Benutzer Daten im System?
- Wie viele Benutzer greifen auf die Anwendung zu und starten Anfragen oder Jobs über die Anwendung?
- In welcher Frequenz (Jobs pro Minute oder Stunde) werden diese gestartet?
- Welche Geschwindigkeit wird dabei erwartet?

Muss die Anwendung in einem bestimmten Zeitfenster Dokumentvolumen verarbeiten?

- Um welches Volumen handelt es sich (GB, Anzahl, Menge/Stunde, Tag, Zeiteinheit)?
- Liegt das Volumen in einer Warteschlange und kann asynchron abgearbeitet werden oder warten Benutzer „synchron“ auf die Ergebnisse?
- Wo wird wahrscheinlich der Flaschenhals sein? (also auch: muss die Anwendung die Daten erst von einer entfernten Quelle holen, die langsam angebunden ist?)
- Gibt es Last- oder Spitzenzeiten? Muss die Anwendung zu Spitzenzeiten genauso schnell reagieren wie in Lastarmen Zeiten?

Sind Optionen in enaio® erforderlich, die die Anwendung beeinflussen?

Für welche Objekttypen in enaio® müssen

- Seitenzahlen,
- Vorschauen,
- Miniaturansichten (Thumbnails),
- Text-Extrakte

durch enaio® documentviewer erstellt werden?

Unter Berücksichtigung aller genannten Kriterien, sollte man sich nun ein Bild vom Gesamtsystem machen und dieses mit Datenflüssen und Komponenten füllen. Auch die Mengen und/oder erwarteten/geschätzten Zeiten sollten bekannt sein.

## Referenzdokumente

Die Definition der Referenzdokumente dient dazu, effektivere Lasttest-Szenarien abzubilden und vergleichbar zu gestalten. Bitte beachten Sie, dass diese Seite sich im Aufbau befindet und keinen Anspruch auf Vollständigkeit erhebt.

- PDF (Version 1.4), coded information, Dateigröße = 200 KB (3 Seiten, DIN A4, Fließtext)
- EML, Dateigröße = 100 KB, Fließtext, ohne eingebettete Bilder, ohne Anhang
- EML, Fließtext ca. 100 KB, ohne eingebettete Bilder, mit Anhängen (laut Referenzliste) max. 5 MB
- DOC/DOCX, Dateigröße 200 KB, DIN A4, Fließtext
- XSLX, Dateigröße 200 KB
- TIFF, Dateigröße 150 KB, s/w, 300 dpi
- TXT, Dateigröße 80 KB

**Hinweis:** Bitte beachten Sie, dass die maximal unterstützte Dateigröße von verschiedenen Faktoren abhängig ist. Dazu gehören: der Dateityp, die Dateiinhalte, der konkrete Anwendungsfall und der verwendete Client.

## Details zu den genannten Dateitypen

Die Testbibliothek umfasst eine Reihe verschiedener Dateitypen. Die folgende Tabelle zeigt, welche Aktionen getestet wurden. Die maximal unterstützte Dateigröße für die jeweiligen Aktionen ist von verschiedenen Faktoren abhängig. Dazu gehören: der Dateityp, die Dateiinhalte, der



konkrete Anwendungsfall und der verwendete Client.

#### Allgemeine Formate

| Beschreibung/<br>Dateiendung             | Haupttyp                               | Inhaltstyp-<br>erkennung | Text-<br>extraktion | PDF-<br>Erstellung | PDF/A-<br>1b-<br>Erstellung | Vorschau |
|--|--|--------------------------|---------------------|--------------------|-----------------------------|----------|
| reine<br>Textdateien<br>(.txt)           | W-Dokument                             | ✓                        | ✓                   | ✓                  | ✓                           | ✓        |
| E-Mail-Text<br>(.eml)                    | E-Mail                                 | ✓                        | ✓                   | ✓                  | ✓                           | ✓        |
| Portable<br>Document<br>Format (.pdf)    | W-Dokument                             | ✓                        | ✓                   |                    |                             | ✓        |
| Portable<br>Document<br>Format (.pdf)    | W-Dokument,<br>Bild<br>(SW/Grau/Farbe) | ✓                        | ✓                   | ✓                  | ✓                           | ✓        |
| Comma-<br>separated<br>values (.csv)     | W-Dokument                             | ✓                        | ✓                   | ✓                  | ✓                           | ✓        |
| Extended<br>Markup<br>Language<br>(.xml) | W-Dokument                             | ✓                        | ✓                   | ✓                  | ✓                           | ✓        |

## Microsoft-Formate

| Beschreibung/<br>Dateiendung                | Haupttyp       | Inhaltstyp-<br>erkennung | Text-<br>extraktion | PDF-<br>Erstellung | PDF/A-1b-<br>Erstellung | Vorschau |
|---|----------------|--------------------------|---------------------|--------------------|-------------------------|----------|
| MS Word 97-<br>2003 Word<br>2010 (.doc)     | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |
| MS Word 2010<br>(.docx)                     | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |
| MS Office<br>Word OOXML-<br>Format (.docx)  | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |
| MS Excel 97-<br>2003 (.xls)                 | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |
| MS Excel<br>2007<br>Arbeitsmappe<br>(.xlsx) | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |
| MS<br>PowerPoint<br>97-2003 (.ppt)          | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |
| MS<br>PowerPoint<br>2010 (.pptx)            | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |
| Microsoft<br>Exchange<br>Message<br>(.msg)  | E-Mail         | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |

| Beschreibung/<br>Dateiendung    | Haupttyp       | Inhaltstyp-<br>erkennung | Text-<br>extraktion | PDF-<br>Erstellung | PDF/A-1b-<br>Erstellung | Vorschau |
|---------------------------------|----------------|--------------------------|---------------------|--------------------|-------------------------|----------|
| MS Project<br>97-2003<br>(.mpp) | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |
| MS Project<br>2010 (.mppx)      | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |
| MS Visio 97-<br>2003 (.vsd)     | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |
| MS Visio 2010<br>(.vsdx)        | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |

## OpenOffice-Formate

| Beschreibung/<br>Dateiendung            | Haupttyp       | Inhaltstyp-<br>erkennung | Text-<br>extraktion | PDF-<br>Erstellung | PDF/A-1b-<br>Erstellung | Vorschau |
|---|----------------|--------------------------|---------------------|--------------------|-------------------------|----------|
| OpenDocumen<br>t Text (.odt)            | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |
| OpenDocumen<br>t Spreadsheet<br>(.ods)  | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |
| OpenDocumen<br>t Presentation<br>(.odp) | W-<br>Dokument | ✓                        | ✓                   | ✓                  | ✓                       | ✓        |

## Bildformate

| Beschreibung/<br>Dateiendung                | Haupttyp                               | Inhaltstyp-<br>erkennung | Text-<br>extraktion | PDF-<br>Erstellung | PDF/A-<br>1b-<br>Erstellung | Vorschau |
|---|--|--------------------------|---------------------|--------------------|-----------------------------|----------|
| Encoded TIFF (-<br>tiff/.tif)               | W-Dokument,<br>Bild (SW/Grau)          | ✓                        | ✓                   | ✓                  | ✓                           | ✓        |
| Tagged Image File<br>Format<br>(.tiff/.tif) | W-Dokument,<br>Bild (SW/Grau)          | ✓                        | ✓                   | ✓                  | ✓                           | ✓        |
| Portable Network<br>Graphics<br>(.png)      | W-Dokument,<br>Bild<br>(SW/Grau/Farbe) | ✓                        | ✓                   | ✓                  | ✓                           | ✓        |
| JPEG-Datei<br>(.jpg, .jpeg)                 | W-Dokument,<br>Bild (Farbe)            | ✓                        | ✓                   | ✓                  | ✓                           | ✓        |
| Windows Bitmap<br>(.bmp)                    | W-Dokument,<br>Bild<br>(SW/Grau/Farbe) | ✓                        | ✓                   | ✓                  | ✓                           | ✓        |

## Container-Formate

| Beschreibung/<br>Dateiendung        | Haupttyp       | Inhaltstyp-<br>erkennung | Text-<br>extraktion | PDF-<br>Erstellung | PDF/A-1b-<br>Erstellung | Vorschau |
|-------------------------------------|----------------|--------------------------|---------------------|--------------------|-------------------------|----------|
| ISO-Image<br>(.iso)                 | W-<br>Dokument | ✓                        |                     |                    |                         |          |
| ZIP-<br>Dateiformat<br>(.zip)       | Container      | ✓                        |                     |                    |                         |          |
| RAR-<br>Dateiformat                 | W-<br>Dokument | ✓                        |                     |                    |                         |          |
| Audio Video<br>Interleave<br>(.avi) | Movie          | ✓                        |                     |                    |                         | ✓        |

# Best Practices

Zur Unterstützung stellen wir Ihnen im Folgenden von uns gemachte Erfahrungen als Best-Practice-Beispiele zur Verfügung.

- "Applikationsserver" unten
- "enaio® documentviewer" unten
- "enaio® fulltext" auf der nächsten Seite
- "Classify" auf Seite 16

## Applikationsserver

| System                              | Anzahl enaio®-Benutzer  |
|-------------------------------------|---|
| 32-Bit-Server: enaio® 8.00 bis 9.10 | max. 250 angemeldete Clients je Server, davon 50 gleichzeitig aktive Benutzer     |
| 64-Bit-Server: ab enaio® 9.00       | 400-500 angemeldete Clients je Server, davon 150-200 gleichzeitig aktive Benutzer |

Diese Angaben sind vom Verhalten der Benutzer abhängig und davon dass ein Server durch andere Aktivitäten nicht weiter belastet wird (wie z. B. Importe o. Ä.).

Weitere Applikationsserver können für technische Aufgaben, wie Importe, Archivierung, Workflow-Abarbeitung etc. eingesetzt werden. Dies kann der Trennung der System- von der Benutzerlast dienen. Bei der Wahl der Anzahl an Servern ist auch die Ausfallsicherheit zu berücksichtigen, sodass die verbleibenden Server die Gesamtanzahl der Benutzer tragen können.

Die genannte Anzahl an Benutzern ist analog für enaio® gateway, enaio® appconnector und enaio® webclient anzuwenden.

## enaio® documentviewer

Sie ermitteln die Größe des von enaio® documentviewer benötigten Cache-Bereichs nach folgender Formel:

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

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

Laut Systemvoraussetzungen wird pro Dokument (8 Seiten) 1 MB im Cache des enaio® documentviewer benötigt. Dementsprechend erlaubt dies eine Hochrechnung basierend auf der Anzahl von Dokumenten oder Seiten.

## Empfehlung

enaio® documentviewer sollte auf einem separaten Host bereitgestellt werden. Es muss zudem geprüft werden, ob das Dokumentenvolumen pro Stunde/Tag/Woche mit den Angaben (1 CPU / 512 MB RAM) zu schaffen ist. In den meisten Fällen, in denen periodische Importe laufen und die Benutzer die Vorschauen „sofort“ erwarten, ist dies nicht der Fall. In diesem Fall kann ein weiterer enaio® documentviewer eingesetzt werden, der die Vorschauauslieferung und Ad-hoc-Erzeugung für Benutzer durchführt.

Alle Datenbanken, temp-Verzeichnisse und Konverter liegen immer lokal und führen daher viele Lese- und Schreiboperationen auf der Festplatte aus. Sind mehrere enaio® documentviewer im Einsatz, wird der Cache von allen geteilt und kann im Netzwerk (auf einem \\UNC\-Pfad) liegen.

## enaio® fulltext

enaio® fulltext basiert auf der Zusatzsoftware Elasticsearch. Sie rechnen die Größe der Volltext-Datenbank wie folgt hoch:

**Pro ca. 1 Mio.** Objekte im Datenbestand werden **25 bis 30 GB** HDD benötigt (je nach Dokumentinhalt und -diversität). Pro Replika werden **weitere 25 bis 30 GB** benötigt. Replikas müssen bei Cluster-Installationen zum Einsatz kommen (mehr als eine Elasticsearch-Node) und dienen der Datenverfügbarkeit, Performance und somit Skalierbarkeit von enaio® fulltext.

Die Datenbank von Elasticsearch darf weder auf \\UNC-Pfaden\, noch auf angebundenen Netzlaufwerken liegen, sondern muss **lokal** liegen. Elasticsearch selbst versteht sich immer als Cluster mit min. einer Node. Bei Skalierung von Elasticsearch wird immer eine **ungerade** Anzahl von Nodes empfohlen.

Eine Skalierung wird empfohlen, sobald die Datenbankgröße auf HDD (ohne Replikas) den Elasticsearch zur Verfügung stehen den RAM um ein Zehnfaches übersteigt. Dann sollte der Cluster erweitert werden mit max. 64 GB RAM pro Elasticsearch-Node.

Die Arbeiten in der Datenbank auf HDD sind relativ ressourcenintensiv. Es muss jedoch nicht zwingend ein SSD-Premium-Speicher verwendet werden. Weiterhin gilt die Faustregel: Je weniger RAM, desto schneller muss die HDD sein, um Inhalte aus der Datenbank „nachzureichen“.

## Classify

Zur Dokumentenklassifizierung wird von enaio® auf die Komponente „smart FIX“ von Insiders zurückgegriffen. Neben den vom Hersteller gesetzten Einschränkungen und Vorgaben haben sich folgende Empfehlungen bewährt:

smart FIX benötigt weitere Datenbanken auf einem MSSQL-Server, je nach eingehender Dokumentenmenge pro Tag ist das Sizing für diese unterschiedlich. Mindestens jedoch sollten 1 GB RAM und 1 CPU auf dem MSSQL-Server dafür vorgehalten werden. Die Klassifizierungskomponente selbst wird auf einem dedizierten Host eingesetzt und benötigt ebenso je nach Dokumentmenge Ressourcen. Als Richtwert kann angenommen werden, dass je DIN A4-Seite (150 dpi) bei einer 3GHz CPU 20 bis 30 Sekunden Extraktionszeit benötigt werden. Dies ist eine Worst-Case-Annahme, begründet durch das zu verarbeitende Scangut, dessen Qualität und der übrigen Auslastung der Infrastruktur. Meist erfolgt die Verarbeitung schneller. Der RAM-Verbrauch wird mit 1 GB pro CPU-Kern angegeben.

### Beispiel:

~20.000 Seiten im Monat könnten von 1 CPU verarbeitet werden.

Hintergrund: 20.000 Seiten / 20 Arbeitstage / 8 Stunden pro Arbeitstag / 60 Minuten ergibt 2,08 Seiten pro Minute, diese sind durch einen CPU-Kern zu bewältigen.

Obiges Beispiel setzt einen sehr gleichmäßigen Eingang an Dokumenten voraus. Sollten diese in Wochen- oder Monatsperioden stoßweise in das System gelangen ändern sich die Anforderungen je nach Anspruch an die Verarbeitungsgeschwindigkeit. Als weiteres Beispiel sei angenommen, die Belege eines Monats (20.000) sollen an einem Tag (8 h) verarbeitet werden.



Also: 20000 Seiten / 8 Stunden / 60 Minuten ergibt 41,7 Seiten pro Minute. Dies wiederum bedingt 20 CPU-Kerne für die Verarbeitung durch die Klassifizierung.

Der Schlüssel ist also die gewünschte Verarbeitungsgeschwindigkeit unter Berücksichtigung der Periodizität der Dokumenteingänge. Daraufhin muss die Skalierung erfolgen.

# Beispielsysteme

Anhand von vier Systemen stellen wir Ihnen im Folgenden beispielhaft Informationen zu konkreten Systemanforderungen, einem dazu passenden Sizing und Berechnungen sowie Überlegungen zum jeweiligen System zur Verfügung. Diese Informationen geben Ihnen Auskunft über grobe Eckdaten, auf deren Basis Sie sich einen Eindruck über mögliche Dimensionierungen machen können. Die Methodik zur Ermittlung des eigenen Sizings anhand bereitgestellter Eckdaten und weitere Details vermitteln wir Ihnen gerne in einem Sizing- und Skalierungs-Workshop.

- [System 1 – klein bis mittel](#)
- [System 2 – mittel, weniger Benutzer, Datenablage, Volltext](#)
- [System 3 – mittel, hauptsächlich Web, ausfallsicher](#)
- [System 4 – mittel, etwas von allem](#)

## System 1 – klein bis mittel

### System 1 – Anforderungen

#### Anforderungen

- MSSQL auf dem System
- Capture-Importe für 5 Gesellschaften
  - Je Gesellschaft 400 Rechnungen (im Schnitt: 2 Seiten) pro Monat, 100% OCR notwendig
- CSV-Importe, Buchungsdaten für die Rechnungen. Einmal am Tag.
- Gesamtvolumen: Aktuell 20 GB, 10 GB/Jahr Zuwachs (WORK-Statistik)
  - 180.000 Objekte im System
  - 1,8 Mio. Historien-Einträge
- Maximal 15 parallele enaio® clients, im Schnitt 8, 3 Power User

- keine Office-Integration
- dezentrales Rechtesystem, ca. 75 Objekttypen
- viele Klauseln
- enaio® webclient wird gewünscht
- Übersichtliche Objektdefinition: > 30 Felder auf Indexdatenmasken kommt 4 Mal vor, sonst weniger.
- Rechnungsworkflow: 75 Prozesse am Tag, Folge aus Importen. (aktuell zwei Gesellschaften, werden 5)
- Keine Ausfallsicherheit
- Keine Langzeitarchivierung
- Keine E-Mail Integration, kein enaio® exchange
- Inkl. Dokumentenvorschau für alle Objekttypen
- Volltext: viel im Einsatz, schnelle Reaktionszeiten erwartet
- Weiteres Volumen: E-Mails per Drag & Drop, 300 im Monat
- Primär Importe per Capture, CSV-Importe und Buchhalter, die Suchen auf dem System recherchieren.

## System 1 – Sizing

### Sizing

#### Zentrale Server / Dienste

- Server 1: Fileserver für WORK, CACHE und DV-Cache  
1:1 Dokumentenmenge + 100 GB

#### enaio®-Applikationen

- Server 1: MSSQL, Appserver, enaio® gateway, enaio® appconnector  
6 CPU, 8 GB RAM, HDD: E:\ 60 GB F:\ 50 GB

- Server 2: enaio® documentviewer, enaio® service-manager (Search, Index, OCR, OSWEB, License), Elasticsearch  
6 CPU, 14 GB RAM, HDD: E:\ 100 GB F:\ 50 GB

## System 1 – Anmerkungen, Gedanken und Berechnung

### Anmerkungen, Gedanken und Berechnung

- MSSQL-Server kommt auf den Applikationsserver. enaio® erhält eine eigene Datenbank, Schema, Login für diese. Die angegebenen Ressourcen stehen auf der Instanz noch zur Verfügung.  
MSSQL-Daten und Logfiles liegen am Applikationsserver auf F:\
- WORK liegt zentral – Platzbedarf hier 1:1 das anfallende Objektivolumen der Importe.  
Renditioncache ebenso zentral, hier mit 50 GB Cachegröße berechnet, bei 10 GB/Jahr sind dies 2 Jahre Vorhaltezeit.
- Die Größe der C:\-Partition darf auf allen Servern frei gewählt werden.
- enaio®-Komponenten bekommen eine eigene Partition (hier: E:\ auf allen Servern).
- Auf dem Server2 wird eine Installation von Microsoft Office benötigt (enaio® documentviewer).
- Fulltext: Je Million Objekte 25 bis 30 GB. Aktuell ca. 300.000 Objekte im System. Lange ausreichend.
- Server1-CPU-Rechnung:  
1 Windows, 2 MSSQL, 1 Appserver, 0,5 enaio® gateway, 0,5 enaio® appconnector, 1 Puffer
- Server1-RAM-Rechnung:  
1,5 Windows, 3 MSSQL, 1,5 Appserver, 0,25 enaio® gateway, 0,25 enaio® appconnector, 1 Puffer
- Server2-CPU-Rechnung:  
1 Windows, 2 enaio® documentviewer, 1 enaio® service-manager, 1 Elasticsearch, 1 Puffer
- Server2-RAM-Rechnung:  
1,5 Windows, 3 enaio® documentviewer, 3 enaio® service-manager, 6 Elasticsearch, 0,5 Puffer

Nicht alle Instanzen von enaio® capture sollten zeitgleich importieren (da nur ein Applikationsserver).

## System 2 – mittel, wenige Benutzer, Datenablage, Volltext

### System 2 – Anforderungen

#### Anforderungen

- Fulltext mit OCR
- Kein Workflow
- Keine Ausfallsicherheit
- Keine Langzeitarchivierung
- Importe – ca. 60.000 Seiten bereits durchsuchbare PDFs pro Monat
- SAP – ca. 300 Belege am Tag, nicht durchsuchbar
- Inkl. Dokumentenvorschau, primär aus SAP heraus
- Volltext: viel im Einsatz, schnelle Reaktionszeiten erwartet
- Kein Capture
- Benutzer: ca. 100 parallel in enaio® client, davon 10 Power User, kein enaio® webclient
  - Moderates Rechtesystem, ca. 20 Objekttypen
  - Übersichtliche Objektdefinition: 30 Felder auf Indexdatenmasken kommt 1 oder 2 Mal vor, sonst weniger.
- Primär Datensenke für SAP. Suche über enaio® client in den sprechenden Schränken. Benutzer arbeiten viel mit der Volltext-Suche, diese muss daher schnell sein.
- Weiteres Volumen: ca. 150 MS-Office-Dokumente werden pro Tag bearbeitet/erstellt.

### System 2 – Sizing

#### Sizing

## Zentrale Server / Dienste

- Server 1: RDBMS (MSSQL)  
2 CPU, 8 GB RAM, HDD: 50 GB
- Server 2: Fileserver für WORK, CACHE und documentviewer-Cache  
1:1 Dokumentenmenge + 100 GB

## enaio®-Applikationen

- Server 1: Appserver, 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

## System 2 – Anmerkungen, Gedanken und Berechnung

### Anmerkungen, Gedanken und Berechnung

- MSSQL-Server ist bereits in der Infrastruktur vorhanden. enaio® erhält eine eigene Datenbank, Schema, Login für diese. Die angegebenen Ressourcen stehen auf der Instanz noch zur Verfügung.
- WORK liegt zentral – Platzbedarf hier 1:1 das anfallende Objektvolumen der Importe/Belege aus SAP.  
  
Renditioncache ebenso zentral, hier mit 50 GB Cachegröße berechnet, ~ 8.000 Seiten pro Tag  
== ~ 3 bis 4 Monate Cache-Vorhaltezeit. 50 GB als Puffer hinzu.
- Die Größe der C:\-Partition darf auf allen Servern frei gewählt werden.
- enaio®-Komponenten bekommen eine eigene Partition (hier: E:\ auf allen Servern).
- Auf dem Server3 wird eine Installation von Microsoft Office benötigt (enaio® documentviewer).

- Die Dienste werden in etwa gleichberechtigt verteilt. OSR3 und Appserver sind auf demselben Host, um hier gegen Localhost arbeiten zu können.
- Elasticsearch mit einem Replika, je auf Server 1 und 2. enaio® documentviewer und OCR separat aufgrund hoher Zahl an Lese- und Schreiboperationen, separate LUN (F:\) für Fulltext-Datenbanken.
- Server1-CPU-Rechnung:  
1 Windows, 2 Appserver, 1 OSR3, 1 Elasticsearch, 1 enaio® service-manager, 0 Puffer
- Server1-RAM-Rechnung:  
1,5 Windows, 3 Appserver, 5 Elasticsearch, 2 enaio® service-manager, 0,5 Puffer
- Server2-CPU-Rechnung:  
1 Windows, 0,5 enaio® gateway, 0,5 enaio® appconnector, 1 Elasticsearch, 1 enaio® service-manager, 0 Puffer
- Server2-RAM-Rechnung:  
1,5 Windows, 0,5 enaio® gateway, 0,5 enaio® appconnector, 5 Elasticsearch, 2 enaio® service-manager, 0,5 Puffer
- Server3-CPU-Rechnung:  
1 Windows, 4 enaio® documentviewer, 0,5 enaio® service-manager, 0,5 Puffer
- Server3-RAM-Rechnung:  
1,5 Windows, 1,5 enaio® documentviewer, 0,5 enaio® service-manager, 0,5 Puffer

Bis auf die erhöhten Anforderungen für enaio® fulltext ist dies ein übliches Szenario. Ein bereits bestehender SQL-Server wird mitverwendet. Es gibt einen zentralen Bereich für Ablagen, moderates Dokumentvolumen und keine besonderen Anforderungen.

Sollten für die SAP-Schnittstelle nicht Wochen- oder Monatsläufe anstehen (also die gesamte Anzahl an Dokumenten stoßweise kommen), so ist erst nach 3 bis 5 Jahren mit einer Neubetrachtung zu rechnen.

## System 3 – mittel, hauptsächlich Web, ausfallsicher

### System 3 – Anforderungen

#### Anforderungen

- Dezentrales Arbeiten, zu 90% über enaio® webclient. Gesamt ca. 1000 Benutzer. Wenige Power User (ca. 30) über enaio® webclient, diese zentral
- Ausfallsicherheit, komplette Webclient-Funktionalität ist ein „Muss“, Volltext und Vorschauen wären verzichtbar.
- Langzeitarchivierung vorhanden
- Bestand: ca. 10 Mio. Objekte im System, hauptsächlich Read-only/PDF /TIFF, kaum durchsuchbare Objekte.
- Zentrale Capture-Installation, ca. 600 Seiten/Tag
- Zudem: über REST-Schnittstelle von einem Portal ca. 1200 Seiten/Tag  
Daraus resultierende Workflows: ca. 300 Prozesse am Tag
- Ausgefeiltes Rechtesystem, ca. 50 Objekttypen, jedoch viele Klauseln durch Mandantentrennung und viele Abteilungen.
- Reaktionszeiten für Webanwender stehen im Vordergrund. Der Fokus liegt auf Schnelligkeit und Stabilität.

### System 3 – Sizing

#### Sizing

##### Zentrale Server / Dienste

- Server 1: RDBMS (MSSQL-Cluster)  
4 CPU, 12 GB RAM, HDD: 50 GB
- Server 2: Fileserver für WORK, CACHE und DV-Cache  
1:1 Dokumentenmenge + 150 GB



## enaio®-Applikationen

- Server 1: Appserver, enaio® gateway, enaio® appconnector  
8 CPU, 8 GB RAM, HDD: E:\ 60 GB
- Server 2: Appserver, enaio® gateway, enaio® appconnector, enaio® service-manager (OSWEB, License)  
8 CPU, 8 GB RAM, HDD: E:\ 60 GB
- Server 3: enaio® documentviewer, enaio® service-manager (Search, Index, OCR), enaio® fulltext  
12 CPU, 16 GB RAM, HDD: E:\ 60 GB F:\ 500 GB

## System 3 – Anmerkungen, Gedanken und Berechnung

### Anmerkungen, Gedanken und Berechnung

- Ein MSSQL-Cluster ist bereits in der Infrastruktur vorhanden. enaio® erhält eine eigene Datenbank, Schema, Login für diese. Die angegebenen Ressourcen stehen auf den Instanzen noch zur Verfügung.  
4 CPU durch komplexeres Rechtesystem, 12 GB durch 10 Mio. Objekte.
- WORK liegt zentral – Platzbedarf hier 1:1 das anfallende Objektvolumen aus Capture und durch REST-Schnittstelle.  
Renditioncache liegt ebenso zentral. Hier mit 100 GB Cachegröße berechnet. Hohe Cache-Vorhaltezeit für schnelle Auslieferung an enaio® webclient.
- Die Größe der C:\-Partition darf auf allen Servern frei gewählt werden.
- enaio®-Komponenten bekommen eine eigene Partition (hier: E:\ auf allen Servern).
- Auf dem Server3 wird eine Installation von Microsoft Office benötigt (enaio® documentviewer).
- Aufgrund des anspruchsvollen Rechtesystems, der Zahl der Workflows und der Langzeitarchivierung wird ein Appserver mit 8 CPUs benötigt.
- Keine Besonderheiten für enaio® documentviewer und enaio® fulltext. Mit Datenbestand hochskaliert, noch über je eine Instanz zu erledigen. Laufwerk F:\ hat für die Fulltext-Datenbank mit 500 GB noch Puffer bis ca. 12 bis 15 Mio. Dokumente (bei steigender Dokumentzahl, muss der RAM ebenso mitskaliert werden, andernfalls sinkt die Antwortgeschwindigkeit).

- Durch den hohen OCR-Bedarf wird allein mit vier Kernen für OCR gerechnet. Potentiell werden 1800 Seiten in einem 8-Stunden-Arbeitstag verarbeitet (225 Seiten pro Stunde bzw. 3,75 Seiten pro Minute).
- Server1-CPU-Rechnung:  
1 Windows, 4 Appserver, 0,5 enaio® gateway, 2 enaio® appconnector, 0,5 Puffer
- Server1-RAM-Rechnung:  
1,5 Windows, 4 Appserver, 0,5 enaio® gateway, 1 enaio® appconnector, 1 Puffer
- Server2-CPU-Rechnung:  
1 Windows, 3 Appserver, 0,5 enaio® gateway, 2 enaio® appconnector, 1 enaio® service-manager, 0,5 Puffer
- Server2-RAM-Rechnung:  
1,5 Windows, 4 Appserver, 0,5 enaio® gateway, 1 enaio® appconnector, 0,5 enaio® service-manager, 0,5 Puffer
- Server3-CPU-Rechnung:  
1 Windows, 4 enaio® documentviewer, 6,5 enaio® service-manager (*bestehend aus: 4 OCR, 1 Search und Index, 1,5 restliche Services*) , 0,5 Puffer
- Server3-RAM-Rechnung:  
1,5 Windows, 1,5 enaio® documentviewer, 3 enaio® service-manager, 10 Elasticsearch, 0 Puffer

## System 4 – mittel, etwas von allem

### System 4 – Anforderungen

#### Anforderungen

- Importe im Einsatz, divers, REST (Droptargets), XML, CSV, Volumen: ca. 1000 Seiten pro Tag, alles bereits PDF und mit Textinformationen
- Stammdaten-Exporte für Drittsysteme als CSV und XML
- enaio® exchange für Journaling und 100 Benutzer, gesamt ca. 2500 E-Mails pro Tag
  - enaio® search für die betroffenen Benutzer im Einsatz

- OSR3, 1 Repository zu SAP, ca. 400 Seiten pro Tag, OCR notwendig.
- Frequente AD-Sync für ca. 500 Benutzer in ca. 20 Gruppen
- enaio® capture mit Kofax und OCR
- enaio® webclient, 200 Benutzer
- enaio® client, 100 Benutzer
- Volltext-Durchsuchbarkeit insbesondere für E-Mails (inkl. Anhänge, Tiefe: 2) gewünscht.
- Clientseitige Office und Outlook-Integration (NG-Addins)
- Arbeitsweise: Benutzer arbeiten viel mit Trefferlisten und der Dokumentvorschau bzw. der Suche darin. Trefferlisten werden entweder über gespeicherte Suchen (enaio® webclient) oder über die Volltextsuche und dynamische Anfragen (enaio® client) abgerufen.

## System 4 – Sizing

### Sizing

#### VORSCHLAG 1

##### Zentrale Server / Dienste

- Server 1: RDBMS (MSSQL-Cluster)  
2 CPU, 6 GB RAM, HDD: 50 GB
- Server 2: Fileserver für WORK, CACHE und DV-Cache  
1:1 Dokumentenmenge (ca. 3GB / Tag) + 500 GB (DV-Cache)

##### enaio®-Applikationen

- Server 1: Appserver, enaio® appconnector, enaio® gateway, enaio® documentviewer (Client)  
6 CPU, 6 GB RAM, HDD: 80 GB
- Server 2: Appserver, enaio® appconnector, OSR3, EEX, enaio® service-manager (OSWEB, EMS, License, Extraction)  
5,5 CPU, 5 GB RAM, HDD: 80 GB

- Server 3: enaio® documentviewer (CPB), enaio® service-manager (OCR)  
6 CPU, 6 GB RAM, HDD: 120 GB
- Server 4: enaio® service-manager (Search, Index), Elasticsearch  
2 CPU, 6 GB RAM, HDD: 100 GB

## VORSCHLAG 2 (weniger Windows-Server)

### Zentrale Server / Dienste

- Server 1: RDBMS (MSSQL-Cluster)  
2 CPU, 6 GB RAM, HDD: 50 GB
- Server 2: Fileserver für WORK, CACHE und DV-Cache  
1:1 Dokumentenmenge (ca. 3GB/Tag) + 500 GB (DV-Cache)

### enaio®-Applikationen

- Server 1: Appserver, enaio® gateway, enaio® documentviewer (Client)  
7,5 CPU, 7,5 GB RAM, HDD: 80 GB
- Server 2: enaio® appconnector, EEX, enaio® service-manager (OSWEB, EMS, OCR, Search, Index, License, Extraction), Elasticsearch  
6 CPU, 11 GB RAM, HDD: 80 GB (enaio® service-manager), 100 GB (VTX)
- Server 3: enaio® documentviewer(CPB), OSR3  
6 CPU, 8 GB RAM, HDD: 120 GB

Der Appserver kann zum Flaschenhals werden. Daher ist eine Erhöhung bzw. Optimierung der Thread-Anzahlen und Konfiguration (Registry, periodische Jobs, Exekutoren) nötig.

## System 4 – Anmerkungen, Gedanken und Berechnung

### Anmerkungen, Gedanken und Berechnung

- Capture-Mengenangabe fehlt: Annahme: S/W-Scan mit 1000 Seiten/Tag
- Größenberechnung:  $(1000 \text{ [Capture]} + 1000 \text{ [Importe]} + 400) / 8 = 300 \text{ MB} + 2,5 \text{ GB E-Mails pro Tag} \rightarrow 2,8 \text{ GB Daten/Tag}$
- DV-Last-Berechnung:  $(2400 \text{ [importe]} + 10000 \text{ [E-Mails]}) / \text{Tag (8 h)} \rightarrow 26 \text{ Seiten pro Minute}$   
Plus OCR:  $400/\text{Tag plus } 10 \% \text{ der E-Mails (=1000)} \rightarrow 1400 \text{ OCR-Seiten pro Tag} \rightarrow 3 \text{ Seiten pro Minute} \rightarrow 1 \text{ OCR-Kern}$
- 2. OCR-Kern (License) für enaio® capture wird benötigt. OCR am Client-Arbeitsplatz.
- Ein MSSQL-Cluster ist bereits in der Infrastruktur vorhanden. enaio® erhält eine eigene Datenbank, Schema, Login für diese. Die angegebenen Ressourcen stehen auf den Instanzen noch zur Verfügung.  
Keine Angabe zu Altdaten, Komplexität des Rechtssystems oder Skripten, daher 2 CPUs mit je 6 GB RAM.
- WORK liegt zentral. Platzbedarf hier 1:1 das anfallende Objektvolumen aus enaio® capture und durch enaio® exchange.  
Renditioncache ebenso zentral, hier mit 500 GB Cachegröße berechnet, hohe Cache-Vorhaltezeit für schnelle Auslieferung an enaio® webclient und Platzbedarf für E-Mails aus enaio® exchange.
- Die Größe der C:\-Partition darf auf allen Servern frei gewählt werden.
- enaio®-Komponenten bekommen eine eigene Partition (hier: E:\ auf allen Servern).
- Auf den Servern 1 und 3 wird eine Installation von Microsoft Office benötigt (enaio® documentviewer).
- Keine Besonderheiten für enaio® documentviewer und enaio® fulltext. Mit Datenbestand hochskaliert. Laufwerk F:\ hat für die Fulltext-Datenbank mit 100 GB noch Puffer bis ca. 12 bis 15 Mio. Dokumente.  
Vorziehen ist der erste [Vorschlag](#). Hier wird die Last besser auf die Windows-Server verteilt. Zudem wird für die Kerndienste ein separater Applikationsserver bereitgestellt.





**OPTIMAL SYSTEMS**  
A KYOCERA GROUP COMPANY

Unternehmenszentrale

Cicerostraße 26

10709 Berlin

Telefon: +49 30 895708-0

[kontakt@optimal-systems.de](mailto:kontakt@optimal-systems.de)