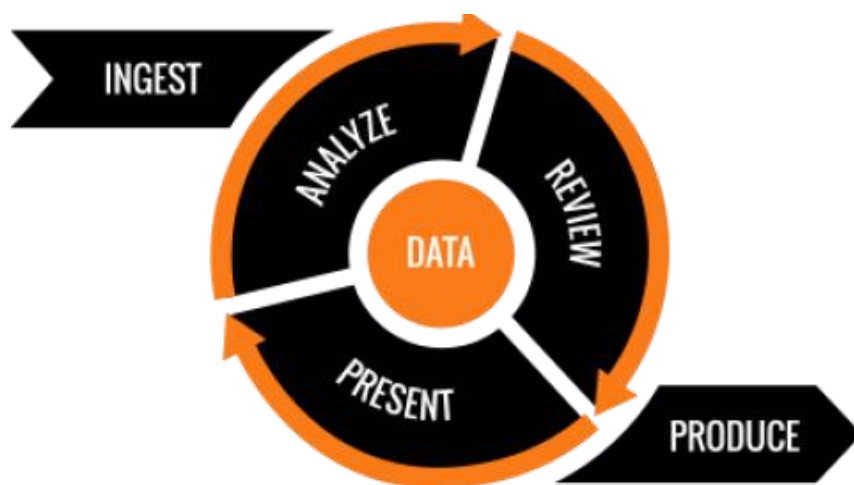




Discovery Simplified.

# System Requirements

EDT 6.0



# Contents

- Introduction..... 3
- 1 Components, Modules & Data Repositories ..... 3
- 2 Infrastructure Options..... 5
  - 2.1 Scenario 1 - EDT Portable or Server..... 5
  - 2.2 Scenario 2 - EDT Site ..... 5
  - 2.3 Scenario 3 - EDT Site with Distributed Processing ..... 5
- 3 Software Requirements ..... 6
  - 3.1 Client Computers ..... 6
  - 3.2 EDT Servers ..... 6
- 4 Hardware Requirements – Standalone machines ..... 7
- 5 Hardware Requirements – Distributed environments ..... 8
- 6 Optical Character Recognition ..... 9
- 7 Virus Protection..... 9

# Introduction

EDT is a modular, server based application with central data stores. It is designed to work with standard Microsoft technologies (Microsoft Windows Server, Microsoft SQL Server, Microsoft IIS, and the .NET framework) in order to leverage the power and features that they provide. Reliance on external third party tools is kept at a minimum. This reduces the complexity of the system, it's technical footprint and administrative overhead.

Please arrange a discussion with one of our technical team before implementing EDT to ensure your environment is configured effectively to meet your needs.

## 1 Components, Modules & Data Repositories

The components, modules and data repositories of an EDT implementation are as follows:

- **Components**

These are architectural items in an EDT deployment that the application relies upon to function correctly. Components are not developed by EDT Software. Components may be provided by one or more physical or virtual server. The underlying Components in an EDT implementation are:

  - Application Server – Microsoft Windows Server
  - Database Server – Microsoft SQL Server
  - File Server – Provides access to source and temporary data
  - PrizmDoc Server – Provides near native rendering of documents
  - Web Server – Microsoft IIS
  
- **EDT Modules**

These are desktop applications, web applications and windows services that form the EDT product. EDT Modules are developed, maintained and supported by EDT Software. The EDT Modules that are deployed in an EDT implementation are:

  - EDT Web Application
    - Analyst – for performing Early Case Assessment of ingested data
    - Reviewer – the document review interface
    - Administration – configuration of Case settings, perform tasks (i.e. ingest OCR text, perform document exports, etc.)
  - Desktop
    - Importer – for ingesting structured load files
    - Loader – for ingesting native electronic documents
    - QA Manager – for triage of processing exceptions
  - Windows Services
    - Agent – performs document retrieval, rendering and export tasks
    - EDT Server – performs database management functions
    - Web Manager – maintains temporary files in EDT Web Application
  
- **Data Repositories**

These are the locations provided by the File Server required for the various file operations that are performed by the EDT Modules. Like Components, they may be provided by one or more File Servers. The Data Repositories in an EDT implementation are:

- Source Documents
- SQL Case Databases
- Common File Store
- Temporary Directory
- Export Path

Each Component and EDT Module can be installed on a single server as shown in Figure 1 below.



Figure 1 - EDT Server Implementation

Alternatively, it is possible to install the Components and EDT Modules on multiple servers (shown in Figure 2 below) to cater for factors such as the volume of data you anticipate to ingest, or the number of reviewers that will be accessing the system.

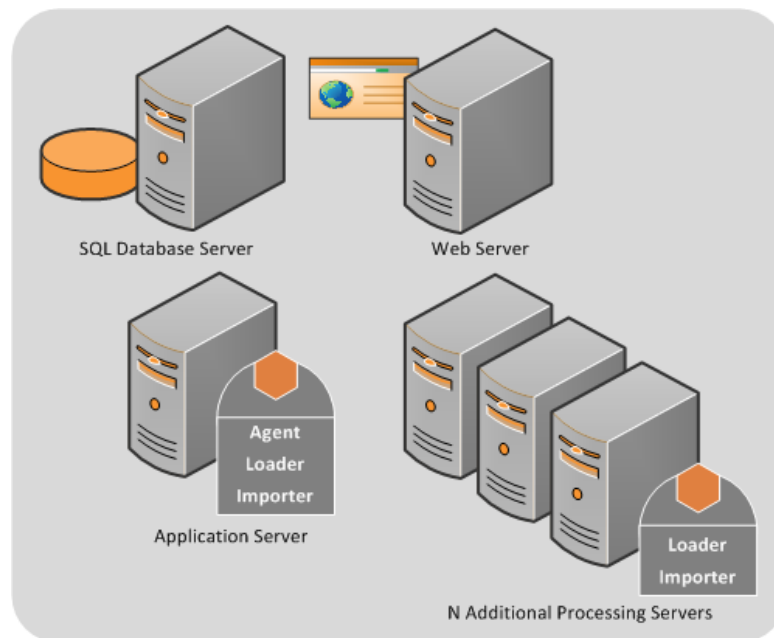


Figure 2 - EDT Site Implementation

Further, more detailed information to assist in planning for your EDT deployment can be found in the supplementary guide "EDT 5.0 Components, Modules, Data Repositories and Workflow".

## 2 Infrastructure Options

EDT's flexibility can accommodate almost any infrastructure requirements, whether they be an onsite single instance or a multi-client distributed processing and review service. Our multithreaded, concurrent architecture enables infrastructure to scale with demand, something essential to growing an eDiscovery capacity or when fully leveraging Cloud based solutions and supporting Infrastructure and Software as a Service models.

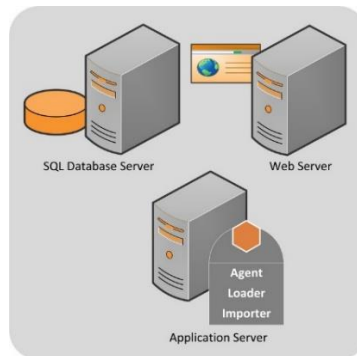
### 2.1 Scenario 1 - EDT Portable or Server

All Components and EDT Modules installed a single laptop or server.



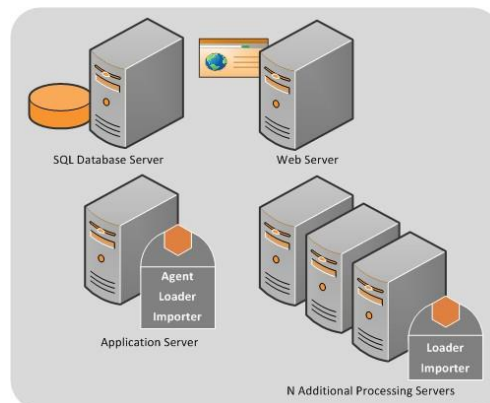
### 2.2 Scenario 2 - EDT Site

Traditional distribution and segregation of Components and EDT Modules across multiple servers within a single site.



### 2.3 Scenario 3 - EDT Site with Distributed Processing

An expansion of Scenario 2 to allow deployment of Agent, Importer and Loader EDT Modules across multiple machines to increase data ingestion and production throughput.



## 3 Software Requirements

### 3.1 Client Computers

End user computers accessing the Analyst or Reviewer require the following software.

Software	Analyst	Reviewer
Windows 7 or later	■	■
Microsoft Internet Explorer 11 with JavaScript enabled.	■	■
Adobe Acrobat Reader 8+		□

■ Required    □ Optional

### 3.2 EDT Servers

The table below lists the required software for the server and administrator applications.

Software	Web Application (IIS) Server Web App (Analyst & Reviewer) Web Manager	SQL Database Server EDT Server	PrizmDoc Server	Loader	Agent Service	Importer	QA Manager
Operating System: Windows Server 2008 R2 (x64) or 2012 R2	■	■	■	■	■	■	■
IIS 7+ (Internet Information Services)	■						
SQL Server 2012 or SQL Server 2014 <sup>1</sup>		■					
Ghostscript 9.05+					■		
PDF Printer. One of the following: • Adobe® PDF Printer • Bullzip PDF Printer • bioPDF PDF Writer					■		
IBM® Lotus Notes® Client 8.5 (Standalone, Messaging) <sup>2</sup>				■	■		■
Mount Image Pro v6 <sup>3</sup>				□	□		□
Microsoft Access Database Engine 2010 (64 bit)					■		
Microsoft .NET Framework 4.6.1	■	■	■	■	■	■	■

■ Required    □ Optional

<sup>1</sup> Similar Documents functionality requires that Semantic Search is installed and configured in SQL Server 2012 or above.

<sup>2</sup> The IBM® Lotus Notes® Client is required to load .NSF files.

<sup>3</sup> Mount Image Pro is required to load Forensic Image files

## 4 Hardware Requirements – Standalone machines

This information is provided by way of a guideline only as there are many different ways the solution can be implemented to service different client environments and requirements. Detailed discussions should take place with an EDT technical consultant prior to infrastructure procurement and implementation to ensure capacity will meet client needs.

Description	Hardware Example
Laptop for small, portable document processing and review cases with a single reviewer	Intel Core i7 16 GB RAM 256 GB Solid State Drive
Server for small document processing, review and export cases with up to 5 reviewers	8 Cores 128 GB RAM 128 GB SSD – Operating System 2 TB SSD – Source Data 1 TB SSD – SQL Database Files 256 GB SSD – SQL Transaction Log Files 256 GB SSD – EDT CFS & Export Path
Server for small document processing, review and export cases with up to 15 reviewers	16 Cores 256 GB RAM 128 GB SSD – Operating System 2 TB SSD – Source Data 1 TB SSD – SQL Database Files 256 GB SSD – SQL Transaction Log Files 256 GB SSD – EDT CFS & Export Path

## 5 Hardware Requirements – Distributed environments

This information is provided by way of a guideline only as there are many different ways the solution can be implemented to service different client environments and requirements. Detailed discussions should take place with an EDT technical consultant prior to infrastructure procurement and implementation to ensure capacity will meet client needs.

Assumptions	Moderately Sized Matters	Larger Matters	Very Large Matters
Concurrent Reviewers	1 – 20	20 – 50	50 +
Ingestion	100 GB per day	500 GB per day	Up to 1TB per day (Metadata only, filtering on load, and multiple case loading)
Expected Case Size	500 GB (~5,000,000 docs)	1 TB (~10,000,000 docs)	4 TB (~40,000,000 docs)
Collective Size of Cases	10 TB	20 TB	100 TB
Native Documents Reviewed per Minute	10	30	60
Production	20,000 docs per day	100,000 docs per day	250,000 docs per day

Hardware	Moderately Sized Matters	Larger Matters	Very Large Matters
Agent, Loader & Importer server(s)	4 Cores 32 GB RAM	2 x Servers 8 Cores 32 GB RAM	5 x Servers 8 Cores 32 GB of RAM
Web server(s)	4 Cores 8 GB RAM	2 x servers – load balanced 4 Cores 16 GB RAM	4 x servers – load balanced 8 Cores 16 GB RAM
PCC Server	8 Cores 64 GB RAM	2 x Servers 20 Cores 128 GB RAM	4 x Servers 20 Cores 128 GB RAM
SQL Server(s)	4 Cores 64 GB RAM	8 Cores 256 GB RAM	16 Cores 1 TB RAM
Storage	16+ TB 10 TB Source 4 TB SQL DB 1 TB SQL TX 1 TB EDT CFS / Export	30+ TB 20 TB Source 7 TB SQL DB 2 TB SQL TX 1 TB EDT CFS / Export	150+ TB 100 TB Source 40 TB SQL DB 5 TB SQL TX 5 TB EDT CFS / Export

General hardware recommendations:

- Obtain the fastest CPUs available within your budget.
- Some case-level database processes are CPU intensive. Additional memory on the SQL Server will improve the execution time of case-level operations. More memory allows SQL Server to cache more database content, thereby increasing performance.
- Follow other SQL Server best practices. For example, distribute database-related files/logs/etc across dedicated LUNs/spindles/RAID arrays.

General virtualisation recommendations:



- Dedicate sockets, memory and separate hard disks to each virtual machine i.e. do not share or over allocate physical resources

General storage recommendations:

- Use fast drives such as 15K RPM SAS and Solid State Disk (SSD) drives with, where appropriate, RAID configurations to maximum disk I/O performance.
- As a rule of thumb the storage space required for each case is double the original source data size. The storage space should initially be distributed among the Source data, the Common File Store and the database server. Additional storage is required by the web server for file caching and by the Agent for the export destination.
- Segregate Data Repositories by physical hard drive spindles. The data storage repositories include the Source data for each case, Common File Store data, Export destination, and the databases. Segregating the data reduces competition for storage resources.
- Additional storage may be required for external Optical Character Recognition (OCR) applications.

## 6 Optical Character Recognition

EDT supports both an internal OCR workflow (via the Tesseract OCR engine) and an external OCR workflow (via the use of third party products). Users can select which workflow to use during data processing:

- Internal OCR – the internal workflow utilizes the Tesseract OCR engine founded by Hewlett Packard in 1985 and development sponsored by Google since 2006. There are no additional fees for the use of this workflow. Documents requiring OCR may be processed using a selected Agent. Extracted text is added to the database and is available for searching and/or export. Text searchable PDF documents are not produced for documents processed using this workflow.
- External OCR – users may utilize the OCR tool of their preference using the external workflow. Documents requiring OCR are exported from EDT, processed for OCR by the user, and then ingested back into EDT as text files or text searchable PDF documents. We recommend the purchase of ABBYY Recognition Server (running on a dedicated workstation on the same network as EDT) for the external workflow.

## 7 Virus Protection

It is recommended that all data to be ingested into EDT be scanned for viruses prior and other malware. This is often done on a dedicated workstation prior to the loading of data. Real-time anti-virus software operating on production servers will impact the performance and throughput of EDT.