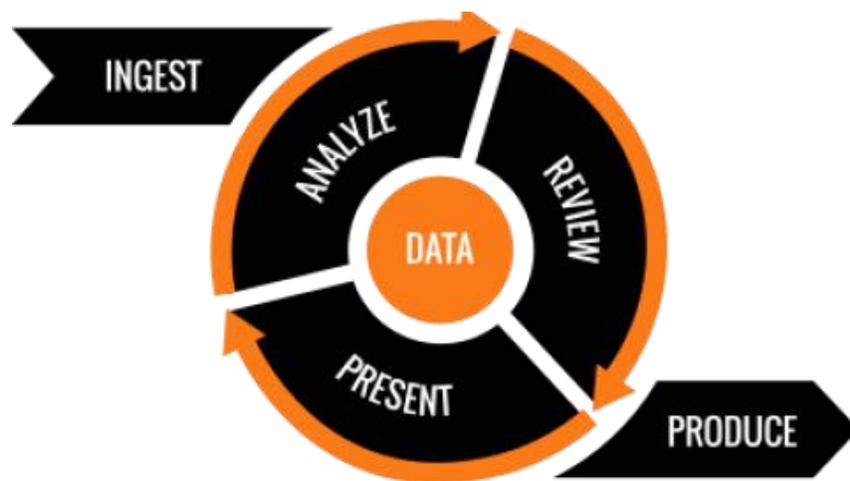




Discovery Simplified.

# System Requirements

EDT



# 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
- Annexure 1 – Supported File Formats ..... 10
- Annexure 2 – Data Ingestion and Data Production Rates ..... 13
  - EDT’s Architecture ..... 13
  - The EDT Workflow ..... 13
  - Data Ingestion Rates ..... 13
  - Data Storage Volumes ..... 13
  - Searching and Analytics ..... 14
  - Concurrent Users and Cases ..... 14
  - Data Production ..... 14

# 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.

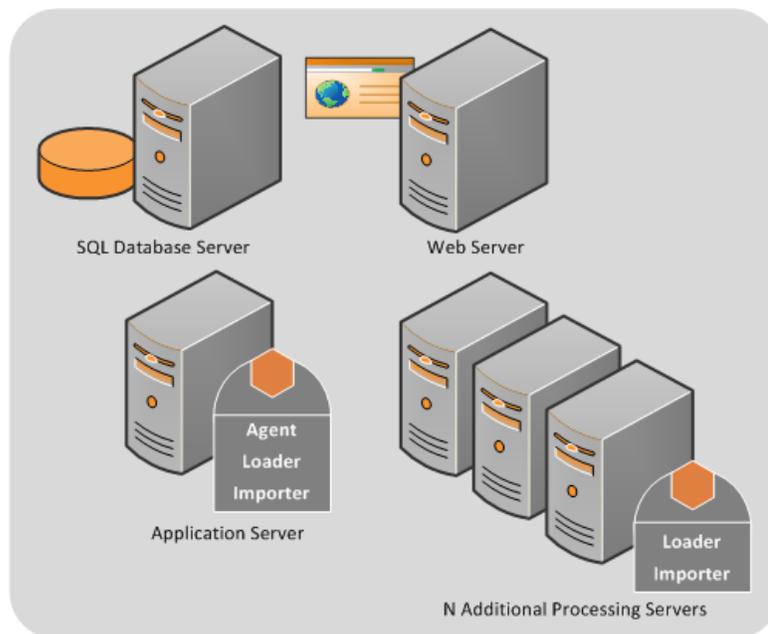


Figure 2 - EDT Site Implementation

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.

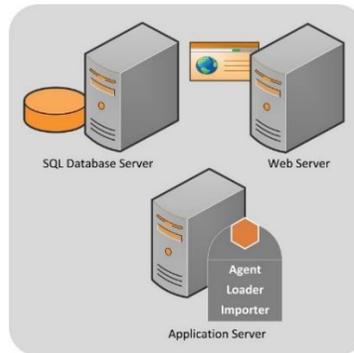
### 2.1 Scenario 1 - EDT Portable or Server

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



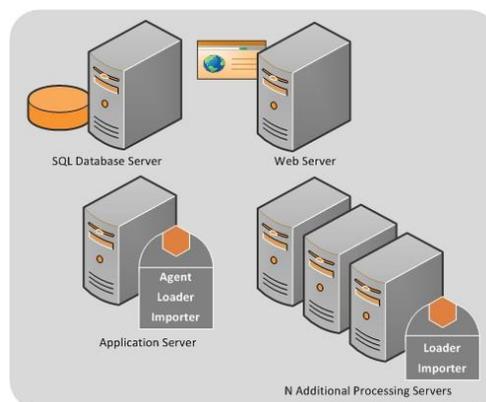
### 2.2 Scenario 2 - EDT Site

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: <ul style="list-style-type: none"> <li>• Adobe® PDF Printer</li> <li>• Bullzip PDF Printer</li> <li>• bioPDF PDF Writer</li> </ul>					■		
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.

# Annexure 1 – Supported File Formats

## EDT Supported File Formats (Viewer)

EDT supports the following file types either natively or via our OEM integration with Accusoft's Prizm HTML5 Viewer. Most formats are detected automatically except where the signature is poor quality or non-existent in which case file extension is used.

Format	File Extension	Auto Detection
<b>Document Formats</b>		
Adobe Portable Document format	.pdf	Yes
Microsoft Excel format	.xls .xlt	Yes
Microsoft Excel Open XML format	.xlsx .xlsm .xltx .xltn	Yes
Microsoft PowerPoint format	.ppt .pot .pps	Yes
Microsoft PowerPoint Open XML format	.pptx .pptm .potx .potm .ppsx .ppsm	Yes
Microsoft Visio Drawing format	.vsd	Yes
Microsoft Visio XML Drawing format	.vsdx .vsdm .vdx	Yes
Microsoft Word format	.doc .dot	Yes
Microsoft Word Open XML format	.docx .docm .dotx .dotm	Yes
OpenDocument Drawing format	.odg .otg .fodg	Yes
OpenDocument Math Formula format	.odf	Yes
OpenDocument Presentation format	.odp .otp .fodp	Yes
OpenDocument Spreadsheet format	.ods .ots .fods	Yes
OpenDocument Text format	.odt .ott .fodt	Yes
Rich Text format	.rtf	Yes
<b>CAD Formats</b>		
AutoDesk AutoCAD format (version 2.5 through 2014)	.dwg .dxf	Yes
AutoDesk Design Web format	.dwf	Yes
MicroStation Drawing format (V7 and V8)	.dgn	Yes
<b>Web Formats</b>		
Extensible HyperText Markup Language format	.xhtml .xhtm	Yes
HyperText Markup Language format	.html .htm	No
Scalable Vector Graphics	.svg	Yes
<b>Email Formats</b>		
Lotus Notes format	.nsf	Yes
Microsoft Outlook format	.msg	Yes
Outlook Express Email format	.eml	No
<b>Text Formats</b>		
Comma-Separated Values format	.csv	No
Text format	.txt	No
<b>Image Formats</b>		
Adobe Photoshop format	.psd .psb	Yes
Continuous Acquisition and Life-cycle Support format	.cal .cals	Yes
Dr. Halo format	.cut	No
Enhanced Metafile format (Windows only)	.emf	Yes

Graphics Interchange Format	.gif	Yes
Icon Resource format	.ico	Yes
JPEG 2000 File Format and Code Stream	.jp2 .jpc	Yes
JPEG File Interchange Format	.jpg .jpeg	Yes
Kodak Photo CD format	.pcd	Yes
Macintosh Metafile format	.pct .pic .pict	Yes
Microsoft Windows Bitmap format	.bmp .dib	Yes
NCR Image format	.ncr	Yes
Portable Bitmap format	.pbm	Yes
Portable Graymap format	.pgm	Yes
Portable Network Graphics format	.png	Yes
Portable Pixmap format	.ppm	Yes
Scitex Color Tone format	.sct	Yes
Silicon Graphics Image format	.sgi	Yes
Sun Raster Data format	.ras	Yes
Tagged Image File Format	.tif .tiff	Yes
Truevision Targa format	.tga .tpic	Yes
Windows Cursor format	.cur	Yes
Windows Metafile format (Windows only)	.wmf	Yes
Wireless Bitmap format	.wbmp	Yes
WordPerfect Graphics Metafile format	.wpg	Yes
X Bitmap format	.xbm	Yes
X Window Dump format	.xwd	Yes
Xerox 9700 Graphic format	.img	Yes
ZSoft Paintbrush DCX format	.dcx	Yes
ZSoft Paintbrush PCX format	.pcx	Yes

### EDT Supported File Formats (Ingestion)

EDT supports the following file types for data ingestion natively or via integration with Mount Image Pro™.

Format	File Extension	Auto Detection
Animated raster graphics file	.gif	No
Apple Text File	.emlx	No
Bitmap	.bmp	No
Comma Separated Values	.csv	Yes
Enhanced MetaFile	.emf	No
Excel file	.xltn	Yes
Excel Macro	.xlm	Yes
Excel Open XML Spreadsheet Template file.	.xltx	Yes
Excel Spreadsheet	.xls	Yes
Excel Spreadsheet	.xlsm	Yes
Excel Template	.xlt	Yes
Excel XML	.xlsx	Yes
Extensible Markup Language	.xml	Yes
Graphics file used with Microsoft Windows	.wmf	Yes

Hyper Text Markup	.htm	Yes
Hyper Text Markup	.html	Yes
jpeg image file	.jpeg	No
jpeg image file	.jpg	No
Lotus Notes	.nsf	No
Macro-enabled Microsoft PowerPoint template	.potm	Yes
Microsoft Exchange Database	.edb	Yes
Microsoft Outlook	.pst	Yes
Microsoft Outlook Email Message	.msg	Yes
Microsoft Outlook offline	.ost	Yes
Microsoft PowerPoint Open XML Macro	.pptm	Yes
Microsoft PowerPoint Open XML Macro-Enabled Slide Show file	.ppsm	Yes
Microsoft PowerPoint Open XML Slide Show file	.ppsx	Yes
Microsoft Word	.doc	Yes
Microsoft Word Macros	.docm	Yes
Microsoft Word Templates	.dot .dotx	Yes
Microsoft Word XML	.docx	Yes
Outlook Express Email Message	.eml	Yes
Portable Document Format	.pdf	Yes
Portable Network Graphics	.png	No
Portable object template	.pot	Yes
PowerPoint Open XML Presentation file	.pptx	Yes
PowerPoint Template	.potx	Yes
Presentation file format used by Microsoft PowerPoint	.ppt	Yes
Real Text Format	.rtf	Yes
Slide show presentation on Microsoft Powerpoint	.pps	Yes
Tag Image File Format	.tiff	No
Tag Image File Format	.tif	No
Text file	.txt	Yes
Web page archive file format	.mhtml	No
Web page archive file format	.mht	No
Zip File Compression	.dotm	No
Forensic image and related formats supported via Mount Image Pro™ integration		
access data .ad1		
apple dmg		
encase .e01 ex01 .l01 lx01		
forensic file format .aff		
iso (cd and dvd images)		
microsoft vhd		
nuix mfs01		
prodiscover		
smart		
unix/linux dd		
raw images		
vmware xways container file.		

## Annexure 2 – Processing Throughput & Capacity

### EDT's Architecture

An EDT implementation is comprised of a central Database Server, and one or more Application Servers, File Servers and a Web Server. The Microsoft SQL Database Server is the central data repository, containing a master database and a database for each Case. Each Case database contains the metadata for each document ingested, full-text indices and work product (annotations, redactions, tagging and field values). The Application Server is used to ingest data (either native electronic files, or structured load files from third parties) and render documents either for review or production. Web Servers to present the EDT user interface.

EDT may be deployed with all components and modules installed on a single server, or distributed across multiple servers (i.e. 1 x Database Server, 2 x Web Servers, 5 x Application Servers).

### The EDT Workflow

The architecture of EDT is designed to enable Attorneys to access the documents they need, fast. This means a single environment from data ingestion through to production, without having to create unnecessary, time consuming load files or images to move documents between platforms. **Through EDT's modular architecture and the use of Microsoft SQL Server it is possible to have Attorneys or Investigators reviewing documents within minutes of receiving data from a client.** Simply kick off dataset ingestion, commence analysis (even before ingestion is complete) and promote batches for review (which can be done natively via our HTML5 viewer without generating images). **It is important to remember this is possible because often the question of processing throughput rates is asked because there is an assumption that clients need to wait until processing has finished before users can start looking at the data. With EDT there is no delay at all because data can be analyzed or reviewed while it is being processed, if necessary.**

### Data Ingestion Rates

Datasets are ingested into EDT either in the form of native electronic files (office documents, PST/OST files, forensic images, etc.) received directly from clients, or as structured load files (Concordance, Summation, Relativity, etc.) received from scanning bureaus or other litigants. These datasets can be ingested using the Loader module (for native electronic files) or the Importer module (for structured load files). Each of these modules can be installed on one or more application servers within an EDT implementation. Data ingestion includes metadata and text extraction and the creation of full-text indexes for searching. Data ingestion rates vary between datasets and environments. However, it is common to **ingest between 150 – 250 GB per day for a single server deployment, or over 1 TB per day for multi-server deployments.**

### Data Storage Volumes

Metadata and text from Native electronic files is stored in a separate SQL database for each case. The original binary data remains on the filesystem and is accessed as needed (e.g. for review or image production). Microsoft SQL Server is very scalable and can contain millions of records in each database. **EDT clients routinely host multiple cases, each with excess of 5 million documents.**

## Searching and Analytics

EDT uses Microsoft SQL Server for full-text indexing, searching and analytics. The indexes within the database schema support complex **searching and analytics of large datasets comprised of millions of documents**.

## Concurrent Users and Cases

The main impact of concurrent users on an EDT implementation relates to HTML5 near-native document viewing and complex querying of the underlying SQL databases. The HTML5 near-native document rendering supports up to 15 document conversions per minute per server – multiple servers may exist in a deployment. Microsoft SQL Server is designed to handle large, complex queries across numerous databases – some clients have **upwards of 200 Case databases and over 80 concurrent reviewers** accessing their EDT implementation. Additionally, it is possible to deploy multiple web servers in an EDT implementation, either manually balanced or through the use of a load balancing server that supports session state.

## Data Production

EDT is routinely used to render and produce large datasets. EDT is flexible and can be used in a variety of ways, including as a processing engine alongside third party review platforms. It's modular nature, with automated distributed image generation and production, allows clients to deploy as many application servers as their requirements demand.

Some of our clients, operating EDT in AWS Cloud, routinely produce large volumes of documents to PDF along with structured load files. The largest production we are aware of was **1.4 million documents to searchable PDF in under 48 hours** through the deployment of a few dozen temporary virtual application servers. This involved PDF rendering which is not necessary in many exchange protocols so it would have been much faster if Tiff or Native files were produced. Furthermore, productions may be paused automatically when higher priority productions are submitted, or specific application servers can be allocated to a production for finer control of resources.