

# Content Central

## Technical White Paper



---

# Table of Contents

Revision History .....	iii
Introduction .....	1
Software Fundamentals .....	2
1. System Requirements .....	2
1.1. Server(s) .....	2
Hardware .....	2
Software .....	2
1.2. Client Workstations .....	2
Hardware .....	2
Software .....	3
Optional Software .....	3
2. Content Central Concepts .....	3
2.1. Documents .....	3
2.2. Document Types .....	3
2.3. Catalogs .....	4
2.4. Capture Types .....	4
2.5. Coding Methods .....	5
3. Server Modules .....	6
3.1. Web Application Server(s) .....	6
3.2. MS SQL Database Server(s) .....	6
3.3. Catalog Service .....	7
3.4. Capture Service .....	7
3.5. Workflow Service .....	7
3.6. Document Storage Area .....	7
3.7. Search Indexes .....	8
Database .....	9
Security .....	10
1. Document Security .....	10
2. Web-Site Authentication .....	10
2.1. Permissions .....	10
2.2. Data Storage Encryption .....	11
3. SSL Encryption .....	11
Supported File Types with Existing Content .....	12
Supported Bar Code Symbologies .....	14
Supported ODBC Providers .....	15

---

# Revision History

2009-12-09	Added Revision History
2010-01-07	Updated Requirements for Server(s) and Client Workstations

---

# Introduction

This white paper outlines the requirements, operation, and maintenance of the Content Central document management software package. This white paper is technical in nature and the intended audience should have an advanced understanding of the Microsoft® Windows® platform as well as a good understanding of computer networks.

After reading this white paper, you will have learned about:

Minimum System Requirements (server and clients)  
Unique Concepts to Content Central  
Installation Considerations  
Database Definitions; and  
Security Models

## **Ademero, Inc. Contact Information**

Phone           (863) 937-0272  
Toll-Free       (888) 276-2914  
Fax              (863) 582-9438  
E-mail          [info@ademero.com](mailto:info@ademero.com) [<mailto:info@ademero.com>]  
Web Site       <http://www.ademero.com/>

---

# Software Fundamentals

## 1. System Requirements

### 1.1. Server(s)

Content Central requires at least one PC-based server platform for product installation.

#### Hardware

##### Minimum Requirements

- 2GHz Dual-Core Processor
- 2GB RAM
- 80GB Redundant Storage Space

##### Recommended Hardware

- Quad-Core Intel® Xeon® Processors
- 4GB RAM
- 250GB Redundant Storage Space

#### Software

Content Central (server software) runs only on Microsoft® Windows® operating systems.

- Microsoft® Windows® XP Professional or higher (Windows Server 2008 or higher recommended)
- Microsoft® Internet Information Services (IIS) 5.0 or higher (7.0 or higher recommended)
- Microsoft® .NET Framework 3.5 or higher (included in the Content Central installation package)
- Microsoft® SQL Server 2005 or higher (SQL Server 2008 Express Edition included in the Content Central installation package)

### 1.2. Client Workstations

Content Central does not require software to be installed on client workstations. The client should contain an operating system and a Web browser. An optional PDF viewer (such as Adobe® Reader®) can be used to view captured paper documents.

#### Hardware

##### Minimum Requirements

- 2GHz Processor
- 1GB RAM

### Recommended Hardware

- 2.5GHz Dual-Core Processor
- 2GB RAM

### Software

- Microsoft® Windows® 2000 or higher (Microsoft® Windows® XP Pro or higher recommended); or
- Mac® OS X or higher
- Microsoft® Internet Explorer 7.0 or higher; or
- Mozilla Firefox® 3.0 or higher
- Sun Microsystems® Java™ runtime environment (JRE) 6.0 or higher

### Optional Software

- Browser-based PDF viewer
- TWAIN driver for any scanner you wish to use with the DirectScan™ applet

## 2. Content Central Concepts

This section identifies the key concepts behind how this document management system operates. This information will help you throughout the rest of this document.

### 2.1. Documents

A document is represented as a file on the file system. What sets a document apart from other files is its use. documents usually contain text or other data that can be searched for retrieval. Common document formats include PDF [Definition: Portable Document Format], Microsoft® Word®, and Microsoft® Excel®. Some document formats contain content and/or metadata that may be inherited when imported. For a full list of these file types, see [Supported File Types with Existing Content](#).

Document properties, also known as index fields, tags, or metadata, provide a classification system that helps you find documents more quickly and accurately in Content Central.



#### Note

Ademero strongly recommends the use of document properties when capturing information.

When paper images have been captured from a scanner or other input device, Content Central converts them to PDF files. The PDF format conveniently stores images, text, and document properties in one file.

### 2.2. Document Types

Each document is described by a document type, which serves as a template for the document. Unique security permissions, fields for document properties, and more can be defined at the document-type level. Each document inherits these settings when captured.

## 2.3. Catalogs

A catalog contains information about a related set of documents in the system. You can create as many catalogs as needed. Catalogs usually take the form of an existing business department or business process. The information stored in a catalog is as follows:

- Document types
  - User & Group Permissions
  - Fields (metadata)
  - Text-Recognition Zones (Zonal OCR)
  - Barcode-Recognition Zones (Zonal Barcode)
  - Field-Lookup Integration
  - Approval Processes [*Enterprise Edition*]
  - Workflow Rules [*Enterprise Edition*]
  - Message Templates [*Enterprise Edition*]
  - Folder & File Building
  - Capture Forms
  - Retention Policy
  - Search & Results Display Fields
- Documents
  - Document Name (file name)
  - Document Location (file path)
  - Document properties (metadata)
  - Document Text (full text of a text-supported file format or OCR [Definition: Optical Character Recognition] from a captured image of a paper document)

Catalogs are created and managed by Content Central and are stored within the SQL database designated for Content Central.

## 2.4. Capture Types

DirectScan™ (Browser)

This Java™ applet allows users to scan paper into the Web browser using a TWAIN-compliant scanner.

QCard™ (Browser, Monitored Folder, E-mail)

QCards™ contain barcodes used to identify individual paper documents in a batch. Users create and print QCards™ from their Web browser. QCard™-attached documents can be scanned to the Web browser using the DirectScan™ applet, saved to a monitored folder, or sent to a monitored e-mail address.

CustomBarcode (Browser, Monitored Folder, E-mail)	Pages containing barcodes may be used to identify the beginning of paper documents in a batch. The information provided within the barcodes may be assigned to field data automatically. The paper documents can be scanned to the Web browser using the DirectScan™ applet, saved to a monitored folder, or sent to a monitored e-mail address.
PDF Form (AcroForm/XFA) (Browser)	New documents can be generated by filling out a PDF form based on a template in the system.
Drag & Drop Upload (Browser)	Users can drag one or more folders and files to one of two Java™ applets in the Web browser.
Single Upload (Browser)	Users can browse to a single file, classify it if desired, and upload it using the Web browser.
Multi Upload (Browser)	Users can browse to one or more folders and files and upload them using the Web browser.
Electronic (Monitored Folder, E-mail)	Files can be saved to a monitored folder or sent to a monitored e-mail address. They will be left in their native format and can be routed to either the <i>Coding Queue</i> or directly to a catalog.
Image-Only (Monitored Folder, E-mail)	Scanned images (TIFF, PDF, JPEG, BMP, PNG, GIF) can be saved to a monitored folder or sent to a monitored e-mail address. They will be converted to fully-searchable PDF files and can be routed to either the <i>Coding Queue</i> or directly to a catalog.
XML (Monitored Folder, E-mail)	Files can be identified with XML descriptors and saved to a monitored folder or sent to a monitored e-mail address. The XML file can contain document properties used to classify the document when captured. The target catalog and document type can also be defined in the XML file. For more information on the XML descriptor, visit <a href="http://www.ademero.com/XmlSchemas/ContentCentral/XmlCaptureDescriptor/">http://www.ademero.com/XmlSchemas/ContentCentral/XmlCaptureDescriptor/</a> .

## 2.5. Coding Methods

Pre-Capture Coding	Users select a catalog and document type and provide document properties <i>before</i> the capture process. Content Central automatically converts and routes these documents to their storage areas without the need for further user intervention.
Post-Capture Coding	Users designate the catalog, document type, and document properties <i>after</i> the capture process. Content Central routes these documents to the user's <i>Coding Queue</i> where they await catalog and document type selection and document properties coding. After a user codes a document in the <i>Coding Queue</i> , the document is routed to the storage area. Users who capture with this method will find the documents in their personal <i>Coding Queue</i> . The nature of a Post-Capture Coding QCard™ allows the user to reuse the same QCard™.
Versatile Coding	Users select a catalog, document type and <i>Coding Queue Destination</i> to which the document should be sent, and are given the option to provide document properties before <i>and/or</i> after capture. Documents

will be routed to the *Coding Queue* for review and additional document properties before being routed to the storage area.

OCR Only / Capture Only

Users select a catalog and document type before capture, but document properties will *not* be added to these documents. Only the OCR process (images) or capture process (electronic) will be performed, and each document will be routed to its storage area using the filename from the original image or electronic file.

## 3. Server Modules

The complete Content Central application consists of seven server modules. Each server module can reside on separate, physical servers or be combined on one server. In high-volume environments, e.g., more than 100 active users and/or more than 10,000 captured pages per day, it may be wise to provide a server for each of the modules. This will greatly improve performance. In low to medium-volume environments all modules can typically run on the same physical server without any performance degradation. When more than one server will be used, each should be connected to the same local network.

### Content Central Server Modules

- Web Application Server(s)
- MS SQL Database Server(s)
- Catalog Service
- Capture Service (includes the Configuration Manager application)
- Workflow Service
- Document Storage Area (includes *Coding Queue* Documents, Deleted Content, Unprocessed Content)
- Search Indexes (includes the Catalog Manager application)

#### Important

Search Indexes and the Catalog Service should exist on the same machine for the best performance.

### 3.1. Web Application Server(s)

This module delivers the content that each user of Content Central interacts with on a regular basis. As a browser-based document management system, most of the administrative and user tasks will be performed within this module.

Requirements: IIS 5.0 or higher; Microsoft® .NET 3.5 Framework

### 3.2. MS SQL Database Server(s)

Content Central uses Microsoft® SQL Server to store all information related to the application, including user accounts, system configuration settings, document records, logging, and notifications.

Requirements: Microsoft® SQL Server 2005 or higher

### 3.3. Catalog Service

Both the Catalog Service and Capture Service are Windows® services, each of which can be dedicated to a physical server. The Catalog Service is responsible for updating catalogs with information about new, modified, or deleted documents. This service also removes documents from the system when their specified retention period has expired.

### 3.4. Capture Service

The Capture Service performs Optical Character Recognition (OCR) on captured images to provide full-text search capabilities and then converts those images to PDF documents. It also handles the capture process for electronic files obtained from monitored folders. Zonal recognition operations and data-source field lookups are also handled in this service.

### 3.5. Workflow Service

The Workflow Service performs automated operations based on live events and scheduled processes.

### 3.6. Document Storage Area

System	<p>The subfolders beneath the <code>System</code> root are necessary for Content Central to run properly. They can each grow in size, and may need to be checked periodically.</p> <ul style="list-style-type: none"><li>• <code>CodingQueue</code>: This subfolder will hold documents that are awaiting user coding and have not been committed to their appropriate storage areas.</li><li>• <code>DeletedContent</code>: This subfolder will contain documents that have been removed from the Content Central database by user action or by an enforced retention policy.</li><li>• <code>Indexes</code>: This subfolder will contain a subfolder for each catalog. These subfolders store the Index information used to provide quick search results. For more information, see <a href="#">Section 3.7, “Search Indexes”</a>.</li><li>• <code>Unprocessed</code>: This subfolder will hold documents that have not been successfully captured by the Capture Service.</li></ul>
Incoming	<p>The <code>Incoming</code> folder will contain one or more subfolders for each catalog and document type. Each of these subfolders are monitored by the Capture Service to import images and other content.</p> <ul style="list-style-type: none"><li>• <code>IncomingQCard</code>: This subfolder is the drop point for image files acquired from a scanning device using QCards™. The images will be converted into searchable PDF files.</li><li>• <code>IncomingImage</code>: This subfolder is the drop point for image files acquired from a scanning device <i>without using</i> QCards™. The image files will be converted into searchable PDF files.</li><li>• <code>IncomingElectronic</code>: This subfolder is the drop point for electronic files. Files dropped in this folder will be captured as-is.</li><li>• <code>IncomingXML</code>: This subfolder is the drop point for XML files that describe other files dropped in the same folder. The XML file can define document boundaries,</li></ul>

document properties, and more. For more information, visit <http://www.ademero.com/XmlSchemas/ContentCentral/XmlCaptureDescriptor/>.

**Documents** The `Documents` folder will contain a subfolder for each catalog. These subfolders are the root storage location for documents and other content. This storage space should be fully redundant and backed up on a regular basis for data security and integrity. The space required will vary by organization. At least 80 gigabytes of storage space is recommended for even the smallest operation.

## 3.7. Search Indexes

Each catalog, containing document types describing documents, lives within the Content Central SQL database. A search Index is also generated (as a flat file on the file system) for each catalog. Storage-space requirements should be taken into consideration. A typical Index will require an additional 10 to 20 percent of the amount of space the documents within a catalog require. For example: An estimated 100-gigabyte catalog of documents will require an additional 10 to 20 gigabytes of storage space for the Index.

---

# Database

Content Central makes use of a relational database model, which employs SQL.

Content Central generates and uses one SQL relational database for storage of several key production elements stored in separate tables. Some key elements are as follows:

- User Accounts & Preferences
- Group Accounts
- Catalog Definitions & Permissions
- Document Records (file-system location, associated metadata, associated catalog)
- Document Version information
- QCard™ Definitions (for batch-capturing scanned images)
- Zonal-Recognition Definitions
- Approval Processes
- Workflow Rules
- Message Templates
- Retention Policies
- Event Logging

Content Central requires a new or existing Microsoft® SQL Server database engine. The installation of Content Central provides Microsoft® SQL Server 2008 Express Edition, a license-free version of the database engine. The SQL server may run on the same machine as other modules of Content Central, or it may reside on a dedicated machine.

The Content Central SQL database, as of version 6.0.x, requires approximately 15 percent the storage space of the document repository<sup>1</sup>. This percentage could be as low as 10 percent or as high as 20 percent depending on the contents of each document. The Microsoft® SQL Server Express Edition has a 4-gigabyte limitation on database size, allowing approximately 700,000 total pages based on the above document size. High-volume environments with 100 or more active users and/or 2,500+ captured pages per day (5 days per week) should consider installing the *workgroup* version or higher of Microsoft® SQL Server after the first year of use to move beyond the 4GB limit.

<sup>1</sup>Approximate requirement is based on an average document size of ten pages containing three metadata fields and around 3,500 total words of OCR text.

## **Example 1. Storage Space Requirements for the Content Central SQL Database.**

An estimated 100-gigabyte (100GB) document storage area will require an additional 15-gigabyte (15GB) minimum of storage space for the database.

---

# Security

Content Central makes use of existing Windows® file security mechanisms as well as its own security model for authentication to the system. An optional SSL certificate/keypair may be added to the Web-site module to encrypt and secure both internal and external communications.

## 1. Document Security

As [previously discussed](#), a document is nothing more than an individual file on the Windows® file system. The administration of Content Central should take into consideration normal security practices when it relates to these files. A best practice is to secure each `Documents` data folder created to the level that best suits the environment of the organization. Typically, non-administrative users on the Windows® network should not be given direct access to any of these folders and files. Data folders (and their files), at a minimum, should be provided with read and write access for Administrators and the Windows® account used to authenticate with Content Central (the *Network Service* account, the *ASPNET* account, or an impersonation account used with ASP.NET). If this practice is followed, the rights-management of each document is passed on to Content Central; permission to view and modify documents will be controlled from within the Web-site security model.



### Note

Allowing other Windows® groups and users to access documents at the file-system level is not recommended; however, this may be desired in particular configurations and will not influence Content Central in a negative way.

## 2. Web-Site Authentication

All users of Content Central must pass through an authentication gateway to gain access to the software and its functions. A username and password, matching a provisioned user account in Content Central or a Windows® Active Directory user account, must be entered on the Content Central login page. Once authentication has been established, users may navigate within the Content Central software and perform actions matching their allowed permissions. These permissions will have been defined by an administrator when the user account is created.

Content Central utilizes the ASP.NET session state to initiate and maintain each logged session. Each individual session will timeout when there is no activity for the duration specified in the session-timeout setting within the ASP.NET application settings of the Content Central Web site/virtual directory in IIS.

Content Central also makes use of several cryptography technologies for storing data such as passwords.

### 2.1. Permissions

The Content Central permissions to be allowed or disallowed per user are as follows:

Document Searching	Documents within a given document type may be queried.
Document Viewing	Following a successful search query, documents in the results listing may be viewed, saved, or e-mailed.
Document Editing	Documents may be checked out for revision, appended to, or replaced.

Document Adding	Documents may be captured from a scanning device (paper) or uploaded from a local machine (electronic).
Document Deleting	Documents may be (soft) deleted from the corresponding catalog. Deleted documents are routed to the DeletedContent system folder.
Properties Editing	Document properties (metadata) may be modified.
Approval-Process Assignment	Documents may be assigned to an approval-process path based on any active approval processes in the document type.
Work-Queue Assignment	Documents may be assigned to the <i>Work Queue</i> of one or more users.
Approval-Process Administrator	Documents on any approval process for the given document type can be managed in the <i>Admin Queue</i> of the <i>Approval Queue</i> .
Work-Queue Administrator	Documents in the <i>Work Queue</i> for the given document type can be managed in the <i>Admin Queue</i> of the <i>Work Queue</i> .
Retention Policy Override	Custom retention policies may be defined for one or more documents of the document type.
Document-Type Administrator	Users granted this permission can manage settings within the document type. Document-type membership and permissions may be managed by these users based on a system setting.

## 2.2. Data Storage Encryption

Each user's password is stored within the SQL database as a SHA1 hash. This hash is non-reversible. When the software attempts to authenticate a user, the user's entered password is converted to a hash, and this hash is compared to the hash stored in the database for the username provided. If both hashes match, the software authenticates the user. If it does not match, the user is denied access.

## 3. SSL Encryption

An optional SSL server certificate may be added to the Content Central Web site/virtual directory in the IIS application. This added security layer will secure communications between the server and each client. This is a best practice for environments that allow remote (internet) access and fall within HIPAA or similar requirements. The encryption type and strength will vary from site to site.

---

# Supported File Types with Existing Content

As of version 6.0.x these are the supported file types that the Catalog Service will recognize for *existing* content and metadata.

Adobe Acrobat (\*.pdf)  
Ami Pro (\*.sam)  
Ansi Text (\*.txt)  
ASCII Text  
ASF media files (metadata only) (\*.asf)  
CSV (Comma-separated values) (\*.csv)  
DBF (\*.dbf)  
EBCDIC  
EML files (emails saved by Outlook Express) (\*.eml)  
Enhanced Metafile Format (\*.emf)  
Eudora MBX message files (\*.mbx)  
GZIP (\*.gz)  
HTML (\*.htm, \*.html)  
JPG (\*.jpg)  
Lotus 1-2-3 (\*.wk?, \*.123)  
MBOX email archives (including Thunderbird) (\*.mbx)  
MHT archives (HTML archives saved by Internet Explorer) (\*.mht)  
MIME messages  
MSG files (emails saved by Outlook) (\*.msg)  
Microsoft Access MDB files (\*.mdb)  
Microsoft Document Imaging (\*.mdi)  
Microsoft Excel (\*.xls)  
Microsoft Excel 2003 XML (\*.xml)  
Microsoft Excel 2007 (\*.xlsx)  
Microsoft Outlook/Exchange  
Microsoft Outlook Express 5 and 6 (\*.dbx) message stores  
Microsoft PowerPoint (\*.ppt)  
Microsoft PowerPoint 2007 (\*.pptx)  
Microsoft Rich Text Format (\*.rtf)  
Microsoft Searchable Tiff (\*.tiff)  
Microsoft Word for DOS (\*.doc)  
Microsoft Word (\*.doc)  
Microsoft Word 2003 XML (\*.xml)  
Microsoft Word 2007 (\*.docx)  
Microsoft Works (\*.wks)  
MP3 (metadata only) (\*.mp3)  
Multimate Advantage II (\*.dox)  
Multimate version 4 (\*.doc)  
OpenOffice 2.x and 1.x documents, spreadsheets, and presentations (\*.sxc, \*.sxd, \*.sxi, \*.sxw, \*.sxcg, \*.stc, \*.sti, \*.stw, \*.stm, \*.odt, \*.ott, \*.odg, \*.otg, \*.odp, \*.otp, \*.ods, \*.ots, \*.odf) (includes OASIS Open Document Format for Office Applications)  
Quattro Pro (\*.wb1, \*.wb2, \*.wb3, \*.qpw)  
TAR (\*.tar)  
TIF (\*.tif)

Supported File Types  
with Existing Content

---

TNEF (winmail.dat)  
Treepad HJT files (\*.hjt)  
Unicode (UCS16, Mac or Windows byte order, or UTF-8)  
Windows Metafile Format (\*.wmf)  
WMA media files (metadata only) (\*.wma)  
WMV video files (metadata only) (\*.wmv)  
WordPerfect 4.2 (\*.wpd, \*.wpf)  
WordPerfect (5.0 and later) (\*.wpd, \*.wpf)  
WordStar versions 1, 2, 3 (\*.ws)  
WordStar versions 4, 5, 6 (\*.ws)  
WordStar 2000  
Write (\*.wri)  
XBase (including FoxPro, dBase, and other XBase-compatible formats) (\*.dbf)  
XML (\*.xml)  
XML Paper Specification (\*.xps)  
XSL  
XyWrite  
ZIP (\*.zip)

---

# Supported Bar Code Symbologies

As of version 6.0.x these are the supported bar code symbologies that the Capture Service will recognize when one or more recognition zones have been defined in a document type.

Codabar  
Code 11  
Code 128  
Code 3 of 9  
EAN-13  
EAN-8  
Industrial 2 of 5 (Code 25)  
Interleaved 2 of 5  
Matrix 2 of 5  
Plessey  
UPC-A  
UPC-E

---

# Supported ODBC Providers

As of version 6.0.x these are the supported ODBC providers that Content Central can use in [External Data Sources](#).

FoxPro  
Microsoft® Access®  
Microsoft® Excel®  
Microsoft® SQL Server  
MySQL  
Oracle  
Pervasive  
ProvideX (Sage)  
Quickbooks QODBC (FLEXquarters)