

2021.2 Administrator Guide

Contents

Administrator Guide for WorkZone PDF 2021.2	3
What's new	4
WorkZone PDF concepts	13
About WorkZone PDF	13
Conversion process	14
Configure WorkZone PDF	18
WorkZone PDF with policies	18
WorkZone PDF Engine scalability (HTTP load balancing using ARR)	26
View errors and logs	29
Conversion errors	29
Event Viewer	31
Advanced logging in WorkZone PDF Engine and WorkZone PDF Crawler	33
Tables	34
Conversion states	34
Failed PDF conversion error messages	37
Supported file types	38
Auto-updated Word fields	42
Troubleshooting and tips	44
Terms and conditions	47

Administrator Guide for WorkZone PDF 2021.2

Related product documentation

- Installation Guide for WorkZone
- WorkZone PDF Developer Guide

WorkZone links

- WorkZone documentation
- WorkZone support
- WorkZone website
- WorkZone portal

What's new

2021.1

No changes in this release.

2021.0

Serilog configuration can be used for advanced logging.

2020.3

No changes in this release.

2020.2

No changes in this release.

2020.1

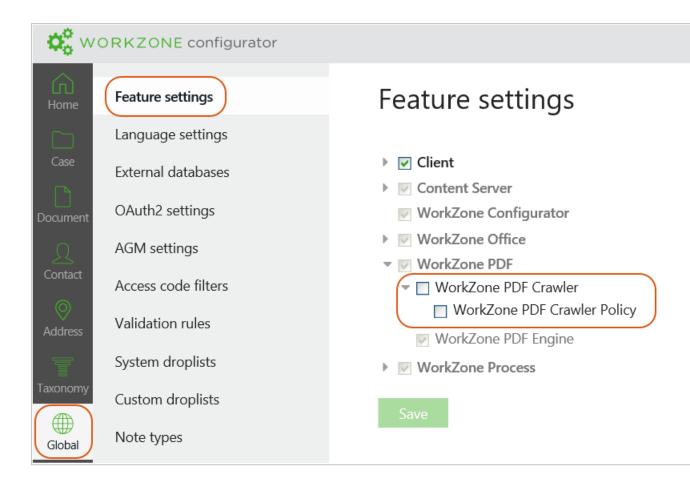
During the conversion, the date and time formats are now taken from the <u>WorkZone</u>
 <u>Configurator settings</u>. WorkZone Configurator administrators can assign specific
 date and time formats for each language. Note that this feature is only applied to date
 and time inserted during the conversion. Regular content in the documents is not
 affected.

2020.0

 The <u>View errors and logs</u> section has been improved and now contains more detailed information.

2019.3

• By default, WorkZone PDF Crawler and its policies are now disabled after the installation. You can enable them in WorkZone Configurator.



- WorkZone PDF Crawler logs have been introduced. To enable this functionality, the LogsLocation parameter must be enabled. See WorkZone PDF Crawler logs.
- Parameter ClearConfidentialInformation is renamed and its settings are extended. See <u>ActionsOnDocumentBeforeConversion</u> in WorkZone Installation Guide.

2019.2

- You can now find information about installing WorkZone PDF in a common WorkZone Installation Guide that covers installation of all WorkZone products. See WorkZone Installation Guide.
- You must enable or disable the policies functionality in WorkZone Configurator. By default, it is disabled.
- OData path has been removed from the web.config file.

2019.1

 Two new parameters were introduced to deal with missing fonts during the document conversion to PDF or PDF/A:

- FailIfMissingFonts defines whether WorkZone PDF must check availability of missing fonts. By default, the parameter is disabled, that is the missing fonts are not checked. Note that WorkZone PDF checks missing fonts on the IIS server only for particular file types. See the list of file types in the Font check column in the Supported file types table.
- FallbackFont substitutes a missing font if FaillfMissingFonts parameter is enabled.

2019.0

 In addition to converting different types of documents to PDF documents, the WorkZone PDF Engine service includes the API that extracts texts from various types of documents (including OCR).

2018.2

- WorkZone PDF supports the functionality of previews of draft versions in WorkZone Client. See more detailed information with examples in the API documentation.
- WorkZone PDF no longer installs Captia Web Client fixed lists.
- The time required for the PDF Engine to start up after being idle due to a time-out or when processing a request for the first time has been reduced. The default setting for the IIS has been changed from Terminate to Suspend, which means the process gets paged out to the disk instead of being terminated. Also, the PDF Engine application is now explicitly started by the IIS just after installation without having to wait for the first request. As a result, the time required to preview PDF documents in the WorkZone Client is also reduced.

The Application Initialization feature for the IIS 8.0 must be installed prior to installing WorkZone PDF Engine in order to take advantage.

2018.1

- New Case Notes report: A new standard report has been introduced enabling you to
 create a report from the case notes and thereafter attach the report as a
 PDF document to the case, thereby submitting any notes as attached documents to
 the case. You can select which case notes are to be included in the case Notes
 report.
- Deploy Custom Reports during installation or update: By placing your custom report templates in the Reports sub-folder of the same folder where the

KMD WorkZone PDF.exe program is located, you can automatically deploy your custom reports during installation or update when selecting the mandatory **Database**Configuration option.

- Preview Text removed: The This is a Preview watermark text on documents viewed as PDF previews has been removed to facilitate reading of preview documents.
- Preview of original PDF Documents: If the original document attached to a case is in PDF format, either as an attached document or as a document directly saved as a PDF document, previews of that document will be taken from the original or saved PDF document instead of waiting for the WorkZone PDF Engine to create a PDF rendition of the original PDF document. The original PDF document will still be converted during normal operations.

2018.0

- New standard meeting report: WorkZone Client now includes a standard meeting
 report with a cover page for the meeting, displaying the meeting details and meeting
 participants as well as displaying the meeting agenda. The meeting report itself contains the meeting agenda and all attached agenda documents.
- The DVS_RENDER_INFO table no longer contains the OK status in order to improve search performance when finding unconverted documents. All entries in the DVS_RENDER_INFO table with the OK status will be deleted upon installation and when updating to newer WorkZone PDF versions.
- The OutputPdfFormat parameter has been renamed: The OutputPdfFormat
 WorkZone PDF Engine custom and installation parameter has been renamed to OutputPdfArchiving and the values changes from Default/Archive to True/False where the False parameter value sets the PDF output format for generic usage and True sets the PDF output format for long-term storage.
- Improved Error logging messages: The error logging in the Windows Event Log has been improved to give a better overview of potential errors encountered during PDF conversions.
- The OutOfMemory error no longer stops the WorkZone PDF Crawler. The WorkZone PDF Crawler will restart 1 minute after an OutOfMemory error is logged. The document causing the error will be marked as Failed and the conversion process will continue.
- The procedure for updating previous versions of WorkZone PDF (before WorkZone 2018) has been changed. You must uninstall all previous versions of WorkZone

PDF before installing any of the WorkZone PDF modules (WorkZone PDF Engine and/or WorkZone PDF Crawler). Remember to back up your WorkZone PDF Engine Web.config and WorkZone PDF Crawler application configuration files prior to uninstalling.

- All WorkZone PDF Engine and WorkZone PDF Crawler custom parameter settings have been moved to the WorkZone Configurator and the WorkZone PDF Crawler configuration file is now obsolete and will no longer be used. The WorkZone PDF Engine configuration file can still be used. As a result of this, you can now define parameters for each instance of the WorkZone PDF Engine and WorkZone PDF Crawler.
- Custom texts in headers and footers: When creating or editing custom reports or converting documents to PDF, you can add custom text to any header or footer placeholders, for example enabling you to create Page of XX footers where XX is the page number. The custom text will not be displayed if the header or footer placeholder does not contain data.
- Additional fields available for headers and footers: You can use field values from custom fields as well as standard fields from expanded tables in the report JSON in custom header or footer placeholders.
- The option to specify redirect options during the WorkZone PDF Engine installation procedure has been removed. You can still set up redirection rules for the WorkZone PDF Engine after the installation by using the standard Microsoft Internet Information Server redirection rules.
- The WorkZone PDF Crawler parameter IntervalBetweenIterations has been removed, optimizing performance and removing the need for a delay period between WorkZone PDF Crawler iterations.
- The WorkZone PDF Crawler parameter RecordsMaxLoads has been removed as the parameter no longer is relevant due to performance improvements and optimizations made to the WorkZone PDF Crawler.
- The WorkZone PDF Crawler now uses the WorkZone OData service to connect to WorkZone. Connection parameters to the Workzone OData service must be specified in the Database Configuration installation option.
- Headers, footers, and table of contents are now automatically updated: The table of contents in Microsoft Word documents are automatically updated when creating PDF

- documents. All document fields in the headers and footers are also automatically updated except the **Filename** field. The source word document is not updated.
- The installation procedure has been expanded and improved with the following changes:
 - The installation of WorkZone PDF Crawler now contains an embedded WorkZone PDF Engine, removing the need for the PDF Crawler to utilize a PDF Engine web service.
 - You can specify custom parameter values for WorkZone PDF Crawler and WorkZone PDF Engine during the installation procedure. Since custom parameter values are now stored in the database, this is done in the new Database Configuration installation option.
 - During the Database Configuration installation, you must specify connection parameters (user, password and location) for the WorkZone database and the WorkZone OData service.
 - You can specify standard (Real-time conversions) or customized (Custom) installation of the WorkZone PDF Engine during installation, defining only instance name in the standard installation.
 - The WorkZone PDF Crawler can only be installed in the standard version, but custom parameters can still be set in the database using the **Database** Configuration installation option or normal database operations (Create, Read, Update, Delete)
 - You can install WorkZone PDF Crawler up to 99 times per WorkZone server and can have up to 99 services running, each service handling one database.
- Reduced Backward compatibility: Note that WorkZone PDF 2018 only supports WorkZone products from 2017 and newer.

2017 SP2

 Microsoft Power BI: WorkZone PDF includes six Microsoft Power BI standard reports which can assist you to quickly analyze progress and trends. The Power BI reports require WorkZone Client 2017 SP1 as well as Microsoft Power BI July 2017 (version 2.48.4792.721) or newer. The new power BI reports are found in the **Power** BI sub-menu in the **Reports** menu of the WorkZone Client.

- Separate cover pages for PDF documents: Each document in a PDF document can
 be separated with its own cover page to better differentiate between the individual
 documents. The cover page can be configured for all reports (as a configuration parameter) or for the individual report (in the report request).
- Global page numbering for PDF documents: You can number all pages in a PDF document. Each page will receive its own page number corresponding to its placement in the PDF document. The optional page numbering of the original document is still retained. You can specify the syntax, location and alignment of the global page numbering.
- Page numbering removed from standard reports: All page numbering functionality
 has been removed from the standard report templates as page numbering now can
 be defined explicitly.
- Windows Server 2016 support: The WorkZone PDF Engine and WorkZone PDF Crawler applications can be installed and run on Windows Server 2016.
- Custom header and/or footer information: You can specify custom header and/or footer information in the report such as page number, page number total, document title and date. Microsoft Word field codes are used for the custom header and footers texts.
- New parameter BypassBoundaryCheckForFiles: You can specify a list of document types (by file extension) which the out-of-bounds check then will bypass when the PDF Engine or PDF Crawler converts documents to PDF. The out-of-bounds check will return an error if the document to be converted contains one or more objects which exceed the defined document boundaries.
- Preview Error messages: When a document could not be displayed in preview, the
 error message is simplified and short. A detailed error messages can be displayed
 by setting the **Debug** parameter to True in the Internet Information Server (IIS) settings. By default, this parameter is set to False.
- Improved error messages: Improved error messages. Error messages for unconverted files from the WorkZone PDF Engine and WorkZone PDF Crawler have been overhauled, reducing the amount of non-essential and potentially confusing information presented to the users.

Documents that are Locked (Document status: UL), Closed (Document status: AFS)
or Archived (Document status: ARK) and contain content that is out of bounds will
not be checked for out of bounds content when converted to PDF, even if the value
of the PDF Engine parameter FaillfOutOfBounds is set to True.

2017 SP1

- The Unconverted list now only displays documents that have failed to convert with the REVIEW state. To display all documents that have failed conversion regardless of state (REVIEW, FAILED, and SKIPPED), use the All unconverted list. The All unconverted - total list is unchanged.
- A new Responsible OU field has been added to the Classification Scheme standard report and the FacetType column has been removed. The File_Class filter has been defined as the default sorting order of the report. When selecting rows in the report, the filter FacetPlan <> "J" is automatically applied.
- The EmailConvertWithAttachments custom parameter has been removed and replaced with the new ConvertWithAttachments parameter. The ConvertWithAttachments parameter defines whether or not document attachments are converted to PDF with the main document and can be applied to API calls, Web.config files or during installation.
- Microsoft Excel custom report templates: You can now create custom reports based on Excel templates. Like Word custom report templates, Excel custom report templates utilize Content Controls in the template to populate the report.

2017

- You can now install WorkZone PDF Engine and WorkZone PDF Crawler using one installation wizard. See .
- The number of custom settings has been extended to include header, footer, and watermark. You can specify custom settings in the installation wizard, in the Web.config file, and as part of API requests.
- In policy based conversions, locked documents are not checked for out of bound content even if the FaillfOutOfBounds parameter is set to True.
- Documents that fail to convert to PDF because the file type is not supported now get the state SKIPPED in the State field in the DVS_RENDER_INFO table. See Conversion states.

- After upgrading WorkZone PDF Crawler now re-converts one previously failed document at a time and only when WorkZone PDF Crawler is idle. See Conversion states.
- You can now configure the maximum number of documents and policies that WorkZone PDF Crawler processes at a time using the two new WorkZone PDF Crawler configuration options RecordsMaxLoad and PoliciesMaxLoad.
- You can schedule when policy-based conversions are to take place, for example to avoid load on the server during work hours. See Schedule a policy.
- Two new WorkZone PDF Engine APIs are now available:
 - Create MSG files
 - Convert MSG files to EML files and vice versa.
- API documentation is now part of the WorkZone PDF installation. After installation, you can enter http://<wrkzone URL>/Render/help in a browser to open the WorkZone PDF API documentation, for example http://db01/Render/help.

WorkZone PDF concepts

About WorkZone PDF

Products overview

WorkZone PDF is a WorkZone product that is used to convert existing documents to PDF. WorkZone PDF converts documents to either the default PDF format (PDF v1.7) or to the PDF/A format (PDF/A-3b). You can specify which PDF format to convert to during the WorkZone PDF installation. WorkZone PDF can be used as a standalone application or in combination with other WorkZone products.

WorkZone PDF consists of two modules:

- WorkZone PDF Engine A stateless Web Application that performs real-time conversions of files to PDF documents, that generates reports, conducts remerging, text extraction, etc.
- WorkZone PDF Crawler A Windows Service that is used for deferred, asynchronous conversions of selected documents, either manually or with policies.

If an organization uses WorkZone PDF with policies, WorkZone PDF Crawler searches for documents that have not been converted, and that match one of the defined policy. When a document that matches a policy is found, it is converted to PDF document and then saved in the WorkZone Content Server. See WorkZone PDF with policies.

Important:

A separate license is required to use the WorkZone PDF Crawler with policies.

WorkZone PDF usage by other products

WorkZone Content Server

WorkZone PDF Engine contains a .Net library that is used in a WorkZone environment for extracting texts from various types of documents (including OCR). The texts are then used for the WorkZone search functionality.

WorkZone Client

WorkZone PDF Engine is used to preview documents in WorkZone Client and WorkZone PDF Crawler is used to create high-quality PDF versions of documents. WorkZone Client also uses WorkZone PDF Engine to print case documents, reports, and case notes.

WorkZone Process

WorkZone Process uses WorkZone PDF Engine to print tasks and process history.

Smart Post uses the WorkZone PDF Engine to create the letters and remerge them.

WorkZone Mobile

WorkZone Mobile opens documents in PDF format created by the WorkZone PDF Crawler program.

Supported file types

See which Supported file types are supported by the WorkZone PDF.

WorkZone PDF Engine APIs

In addition to converting different types of documents to PDF documents, the WorkZone PDF Engine service includes the following APIs:

- Creates server-side emails (*.msg files) that can be streamed to clients. Users can then open the MSG files in Microsoft Outlook and add more information.
- Converts MSG files to EML files and vice versa.
- Extracts texts from various types of documents (including OCR).

For more information, see the API documentation located at http://<Server->/render/help.

Conversion process

About conversion

WorkZone PDF Engine converts one or more documents into a PDF document. A document must pass the following stages to be converted into a PDF document:

1. Getting the source file.

Files can be passed into the service in the following ways:

- Through URI (Uniform Resource Identifier) A URI to the file must be specified.
- By stream The file is provided directly to the service.
- Using JSON with a list of URIs to the files. Documents in the database can be referenced by providing a valid OData URI to the document.
- 2. Converting the file into a PDF document.
- 3. If Microsoft Office Word documents are converted, update the Table of Contents (TOC).

The table of contents in the source Word document will be automatically updated and used as the table of contents in the PDF document in order to ensure a correct TOC. The source Word document is not updated.

All document fields contained in the headers and footers are automatically updated. The **Filename** field will not be updated in order to maintain the source document filename.

4. Applying bookmarks.

In a merged document, a bookmark can be added for every source file. The bookmark name in the document is the link to the file.

5. Applying watermarks (optional).

A watermark is added to the pages of the converted PDF document.

6. Applying headers and footers (optional).

Headers and footers are added to the pages of the converted PDF document.

7. Merging.

Two or more documents will be merged into a single PDF document.

- 8. Applying optimization and compression.
- 9. Providing the PDF output file.

Conversion priorities

The WorkZone PDF Crawler constantly polls the documents database for documents to convert. Each iteration of the PDF Crawler service checks for any potential documents to convert and will thereafter convert these documents.

If there are no documents to be converted, the PDF Crawler service will wait for a number of seconds before starting a new search iteration.

The length of the wait period between PDF Crawler iterations depends on the origin of the document conversion. User requests and edited documents have a higher priority than documents that have failed previously and therefore the PDF Crawler iterations are shorter with regards to user-requested and changed documents.

The table below provides an overview of the priority of the conversion by request origin as well as how often the PDF Crawler starts its search for new documents to convert (the Iteration Start column).

The Priority column indicates which documents will be processed first by the PDF Crawler in an iteration if multiple documents are found.

Priority	Origin	Description	Iteration Start (seconds)
1	User-requested documents	Documents a user requests to be displayed as a PDF document, for example in the WorkZone Client	5
2	Changed doc- uments	Documents that have been edited or changed since the last conversion.	5
3	Archived doc- uments	Documents that have been archived since the last conversion.	60
4	Policy documents	Documents identified for conversion through PDF Crawler policies.	60
5	Pending doc- uments	Documents that are in the process of being converted to PDF.	60
6	Retry documents	Documents where the conversion process has	60

Priority	Origin	Description	Iteration Start (seconds)
		failed and the conversion status is set to Retry.	
7	Failed documents	Documents where the conversion process has failed and the conversion status is set to Failed.	60

See Also

Supported file types.

Remerging documents

Configure WorkZone PDF

WorkZone PDF with policies

Prerequisite:

- A separate license is required to use the WorkZone PDF Crawler with policies.
- You must enable the policies functionality in WorkZone Configurator.

About policies

After installing WorkZone PDF Crawler, you can configure policies to define which documents are to be converted and in which priority.

You can specify which documents to convert based on the following criteria:

- Document state (for example, Draft, Locked, Archived)
- Document type (for example, Incoming, Outgoing)
- Document group (for example, Decision, Hearing, Complaint, Minutes)
- Case class
- Date of last change

You create and maintain policies in the **dvs_policy** table that you access through the **ScanSQL** program, adjusting the policies to the requirements and work flow of your organization, for example, setting up Crawler policies to convert documents based on their date:

- · Convert all documents for the last year
- Convert all older documents meaning documents older than one year.

You can also set up policies to convert documents based on:

The on-going production of documents, for example the changed date >
 installation date so that users do not have to wait for the PDFs that they have

requested.

 All archived documents (State=ARK|AFS) as a batch process for the purpose of delivery.

Policy priority

The highest priority is 0. You can assign as many priority values to the policies as you need, for example, 0, 1, 2, ... 10, 11 and so on.

Setting policy priorities enables converting of important (high priority) policies first and relegates converting of less important (lower priority) policies later, when there may be vacant machine resources to convert documents of these policies - for example, certain document types (WorkZone Document types: DOK, EMBED, I and such) could be assigned a lower priority policy to ensure there are sufficient resources to convert more important documents from a higher prioritized policy.

Create a policy

To create a new policy, execute SQL similar to the following on the database that you want the policy to cover:

Example:

cutoff date, renderer, priority, case class)

The policy converts all documents of the type DOKUMENT and with the document group REF that are not personal drafts and have been changed since May 4, 2017. insert into dvs_policy (policy_key, name, enabled, state, record_type, record_grp,

values(1, 'FixedPDFPolicy', 'J', 'UÅ|UL|ARK|AFS', 'DOKUMENT', 'REF', '>20170504', 'PDF', '0', ")

If the single quotes are empty ", there will be no restrictions as to which documents will be converted:

Example:

insert into dvs_policy (policy_key, name, enabled, state, record_type, record_grp, cutoff_date, renderer, priority, case_class)

values(1, 'FixedPDFPolicy', 'J', ", ", ", ">20110504', 'PDF', '0', ")

Example:

The policy results in rendition of all documents after a specific date, for example the last 3 years. For example, today's date is January 1, 2015 and the policy results in rendition of all documents 3 years back, that is January 1, 2012.

insert into dvs_policy (policy_key, name, enabled, state, record_type, record_grp, cutoff_date, renderer, priority, case_class)

values(1, 'FixedPDFPolicy', 'J', ", ", ", '>20120101', 'PDF', '0', ")

Field descriptions

Field	Description	
policy_key	A key for the policy, for example, 1, 2, 3, and so on. The $\texttt{policy_key}$ must be unique.	
name	Name of policy, for example, PDF_of_Everything.	
enabled	Enable or disable the policy. Possible values are ${ t J}$ (yes) and ${ t N}$ (no).	
state	Indicates the state of the documents to convert. Possible values are UP, UÅ, UL, ARK, and AFS.	
record_type	Defines the type of documents to convert. Possible values depend on the database setup, but typically include, for example, DOKUMENT (Document), I (Incoming), U (Outgoing), and so on.	
record_grp	Defines which documents from which document groups to convert. Possible values depend on the database setup but typically include, for example, AFG (Decision), $HØR$ (Hearing), KLA (Complaint), REF (Minutes), and so on.	

Field	Description
	Note: To include only documents where record_grp has not been filled out, set the record_grp field to `=""', that is an equal sign and two double quotes.
cutoff_date	Date and time expression to limit the amount of documents covered in a policy based on the date the document last was updated. The format is YYYYMMDD where
	YYYY: YearMM: MonthDD: Day
	To cut off at May 4th 2015, use >20150504 (8 digits) as the cut-off or $>=20150504$ if you want to include the date. If you only want to convert the newest documents, insert today's date.
renderer	Indicates the format you want to convert to. Possible value is PDF.
priority	Defines which renditions to prioritize first.
case_class	Defines which documents from which case class (Danish: "sagsgruppe") to convert. Possible values are the classes defined in the case classification scheme (Danish: "Journalplan").
	The table below gives an overview of how you can use the <code>case_class</code> field.
days	Defines on which weekdays to run a policy. You specify weekdays by a list of numbers (Monday = 1,, Sunday = 7) separated by a pipe " ", for example:
	1 2 3 4 5 - Monday - Friday 6 7 - Weekend
hours	Defines the time to run a policy. You specify start and end time in 24 hours format with or without a leading zero and separated by "-". Allowed values are: [0-23] or [1-24]. You cannot set values to "0-0", or "24-24", and the value on the "right" must be greater than the value on the "left".

Field	Description
	Example: Working Hours 08-17 - Working hours 00-05 - Night hours 12-13 - Lunch
	Note: Start and end times are exclusive. In the Working hours example above, the policy will start at 8:00 and finish at 16:59.

You can add several values for state, record_type, record_grp, and case_class fields using the "|" symbol, for example, "DO|I".

Error-checking

In the current version, there is no error-checking at the SQL level, which means that errors in, for example, the search criteria will only be revealed at run-time. Furthermore, there is no error checking on the stated requirements. For more information, see Troubleshooting policies and Troubleshooting and tips.

Change a policy

You can update a policy through SQL. For example, you can disable the 'FixedPDFPolicy' policy by executing the following SQL:

update dvs policy set enabled='N' where name='FixedPDFPolicy'

Schedule a policy

You can schedule when to run policy-based conversions by adding weekday and time filters to a policy, see the <code>days</code> and <code>hours</code> in the Field descriptions. For example you may want to run policy-based conversions outside work hours to avoid extra load on the servers.

Create a policy that converts documents of the record type UÅ and the record group DOK. The policy runs on Saturdays and Sundays from 8:00 to 24:00.

insert into dvs_policy (policy_key, name, enabled, renderer, state, record_type, record_grp, days, hours) values('1', 'Policy 1', 'J', 'PDF', 'UÅ', 'DOK', 'DO', '6|7', '8-24')

Update a policy named policy_1 to run on weekdays from 8:00 to 14:00.

update dvs_policy set days='1|2|3|4|5', hours='8-14' where name='Policy_1'

Create policies that ensure conversion during the night and early morning the next day, for example Sunday night and Monday early morning. It requires two policies.

The first policy ensures conversion from Sunday evening till midnight.

insert into dvs_policy (policy_key, name, enabled, renderer, state, record_type, record_grp, days, hours) values('2', 'NightPolicy 1', 'J', 'PDF', 'UÅ', 'DOK', 'DO', '7', '19-24')

The second policy ensures conversion from midnight till early Monday morning.

insert into dvs_policy (policy_key, name, enabled, renderer, state, record_type, record_grp, days, hours) values('3', 'NightPolicy_2', 'J', 'PDF', 'UÅ', 'DOK', 'DO', '1', '0-6')

Policy types

There are two types of policies:

Event-driven

Policies are aimed at the daily production requirements.

Batch-driven

Policies are aimed at catching up in the database where you want to impose new rules and produce new renditions. Batch-driven policies are also convenient when you want to start using WorkZone PDF in a running installation in the middle of a filing period and you want to prepare the existing document archive for transfer or distribution.

Event-driven policies

Event-driven policies convert all documents to PDF as they are produced to facilitate distribution.

Field	Value
Document State	All
Record Type	All
Record Groups	All except DOPE (Agenda Module)

Field	Value
Class	AII
Priority	2
Cut-off date	15th September 2011

SQL:

```
insert into dvs_policy (policy_key, name, enabled, state, record_type, record_
grp, case_class, cutoff_date, renderer, priority)
values(1, 'PDF-Event', 'J', '', '<>DOPE', '', '>20110915', 'PDF', '2')
```

Batch-driven policies

Batch-driven policies convert all existing documents to PDF to facilitate distribution.

Field	Value
Document State	All
Record Type	All
Record Groups	All except DOPE
Class	All
Priority	4
Cut-off date	None

SQL:

```
insert into dvs_policy (policy_key, name, enabled, state, record_type, record_
grp, case_class, cutoff_date, renderer, priority)
values(1, 'PDF-Batch', 'J', '', '<>DOPE', '','', 'PDF', '4')
```

Converts documents of the BIB type from the case class 10 to PDF.

Field	Value
Document State	Empty
Record Type	BIB
Record Groups	All

Field	Value
Class	10
Priority	3
Cut-off date	4th May 2011

SQL:

```
insert into dvs_policy (policy_key, name, enabled, state, record_type,
record_grp, cutoff_date, renderer, priority, case_class)
values(1, 'p1', 'J', 'UÅ', 'BIB', '=""", '>20110504', 'PDF', '3', '10')
```

Converts documents of the BIB type from any class.

Field	Value
Document State	Empty
Record Type	BIB
Record Groups	All
Class	All
Priority	3
Cut-off date	4th May 2011

SQL:

```
insert into dvs_policy (policy_key, name, enabled, state, record_type,
record_grp, cutoff_date, renderer, priority, case_class)
values(1, 'p1', 'J', 'UÅ', 'BIB', '=""', '>20110504', 'PDF', '3', '')
```

Troubleshooting policies

The current version of WorkZone PDF does not have a user interface for entering and maintaining policies, and there is only little error handling when editing policies in SQL. Errors will be handled only at runtime, when the WorkZone PDF Crawler tries to fetch and interpret the policies.

There are two possible types of errors:

 One or more of the fields in a policy are not in a format which the WorkZone PDF Crawler can interpret.

If you write $M\&M^*$ as a criterion, it matches all documents with an M in the corresponding field because & is a SOM operator. This is probably not what you intended.

 The WorkZone PDF Crawler interprets the policy but not in the way you intended it to

The last type of error is revealed only by the number of documents that are converted. You may experience that too few or too many documents are converted.

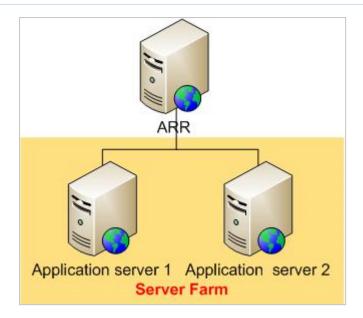
See issues and workarounds in Troubleshooting and tips.

WorkZone PDF Engine scalability (HTTP load balancing using ARR)

You can configure Application Request Routing (ARR) to load balance HTTP requests to achieve high availability and scalability. This topic describes some of the features that Application Request Routing uses to monitor the health of the content servers and collect requests from a client to a content server.

About HTTP load balancing

The purpose is to load balance HTTP requests across several content servers using Application Request Routing as displayed below:



Each application server hosts WorkZone PDF Engine whereas ARR is a web farm controller.

Prerequisites

To configure the Application Request Routing module to load balance HTTP requests, the following requirements must be met.

- IIS 7.0 or newer, on Windows 2008 or newer.
- Microsoft Application Request Routing Version 1 and dependent modules.
- A minimum of two content servers with working sites and applications.

If the Application Request Routing module version 1 is not already installed, you must install it.

- Microsoft Application Request Routing Version 1 for IIS 7 (x86) click <u>here</u> to download.
- Microsoft Application Request Routing Version 1 for IIS 7 (x64) click <u>here</u> to download.

Follow the steps outlined in the <u>Install Application Request Routing</u> document to install Application Request Routing.

Another prerequisite is that the reader has defined and configured a server farm using the steps outlined in <u>Define and Configure an Application Request Routing (ARR) Server Group</u>.

For more information, see <u>Application Request Routing</u> (External link) and <u>Using the Application Request Routing Module</u> (External link)

Server Web Farms

Prerequisite:

You must install the Application Request Routing module before you can create a server web farm.

You can create and set up a server web farm in the Internet Information Services (IIS) Manager on the ARR machine, adding any required servers as necessary. You can add additional servers to the server farm later if the need arises.

For more information, see

- About server farms
- Build a Web Farm with IIS Servers
- Build a Web Farm with IIS Servers

Manage a Web Farm

In the Internet Information Services (IIS) Manager, you can configure, track, and manage your server farm.

For more information about configuring web farm, see Configure a Web Farm with IIS Servers.

Health check monitoring

You can configure health test monitoring in the Internet Information Services (IIS) Manager.

The Application Request Routing module monitors the health of the content servers in two ways:

- Via the live traffic
- Via an explicit URL testing

Live traffic testing

Live traffic testing is performed automatically by default when requests are made to Application Request Routing.

Explicit URL testing

The explicit URL testing is an additional test that you can use with the live traffic testing.

Tip:

- WorkZone PDF does not require caching by the Application Request Routing module and you can disable caching in the Internet Information Services (IIS) Manager.
- To avoid application server time-outs, it is recommended that you increase the time-out parameter value in the Internet Information Services (IIS) Manager.

View errors and logs

You can view WorkZone PDF errors in the following places:

- The DVS_RENDER_INFO table in WorkZone Content Server. It contains errors
 triggered by the document conversions. Users can view these errors in WorkZone Client. Administrators can view the errors in the ScanSQL program.
- The Event Viewer. It contains generic errors, warnings, or information regarding WorkZone PDF Engine and WorkZone PDF Crawler.
- The WorkZone PDF Crawler logs. Three logs were designed to track WorkZone PDF Crawler processes. Besides the errors, the logs also contain run-time type information.

Conversion errors

Conversion errors are stored in the **DVS_RENDER_INFO** table in WorkZone Content Server. Users can view these errors in WorkZone Client. Administrators can view the errors in the ScanSQL program.

View unconverted documents in WorkZone Client

The WorkZone PDF installation creates the following lists in WorkZone Client: **Unconverted**, **All unconverted** and **All unconverted - total**. Users can open these lists from the navigation pane on the Dashboard. The **Unconverted** list is displayed in the navigation pane by default. Users can add the other two lists as needed. See how to <u>add a list to the navigation pane</u> in WorkZone Client.

Prerequisite:

Users must have the LISTCONF access code.

Unconverted (Danish: Ikke-konverterede)

The list displays changed by the current user documents that failed to convert due to some issues. The documents have the status REVIEW. It implies that the user can correct the issues and the documents will be converted later.

All unconverted (Danish: Alle ikke-konverterede)

The list displays changed by the current user documents that cannot be converted. The documents have a status REVIEW, SKIPPED, or FAILED. In other words, the list contains documents that the user can correct and the documents that cannot be corrected.

All unconverted - total (Danish: Alle ikke-konverterede - total)

The list displays all documents that cannot be converted, even if the current user hasn't worked with them. The documents have a PDF status REVIEW, SKIPPED, or FAILED.

The lists contain the following PDF conversion specific fields:

- PDF status (Danish: PDF-tilstand) It can obtain three states REVIEW, SKIPPED, or FAILED - depending on the type of error. See Conversion states.
- PDF error message (Danish: PDF-fejlbeskrivelse) It shows a description of the error.

Handle conversion errors

All documents must be in a state where they can be converted or have a copy that can be converted. Depending on the state of the document and on the type of error, users can perform actions to correct the content and then have the document re-converted.

The way to correct documents that cause unsuccessful conversions depends on whether the document is a draft or not.

Drafts

If a document is in a draft state, for example, UP or UÅ, the user can correct the document, save it, and it will be re-converted.

Locked documents

If a document is locked and thus has one of the other states, for example, UL, ARK, AFS, and O, the user cannot change it and, therefore, it will never be converted successfully. Depending on the type of error, the user may need to create a new copy of the document, correct the errors, save the document, and it will then be converted.

Note: Even if the new version of the document is converted successfully, the old version will still be on your list of failed documents until your organization delivers documents to the Danish National Archives and starts a new filing period.

See also

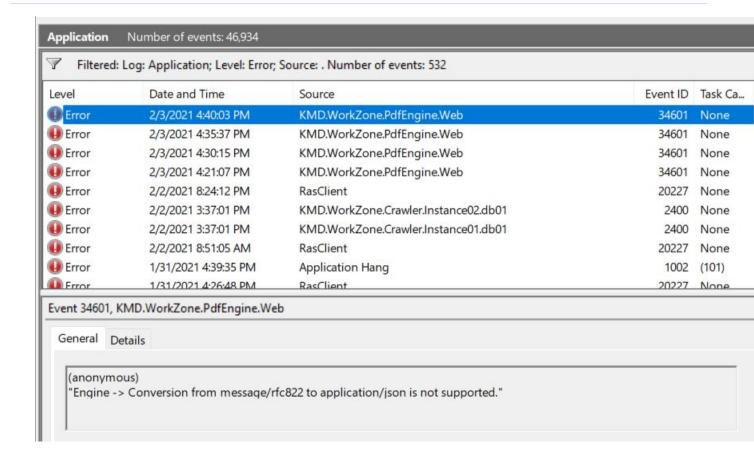
Conversion states

Failed PDF conversion error messages

Event Viewer

You can view errors and warnings triggered by WorkZone PDF Engine and WorkZone PDF Crawler in the **Event Viewer** form under Windows logs. The **Event Viewer** form provides information on relevant User ID, Record ID, and error message.

See an example of the Event Viewer form:

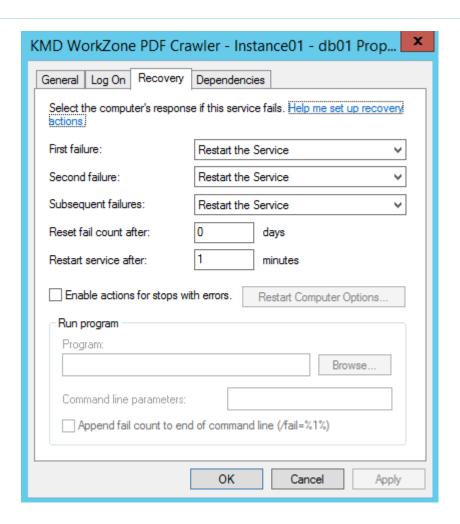


The OutOfMemory exception

If an OutOfMemory exception is thrown, the event will be registered in the Windows event log and the WorkZone PDF Crawler will restart after 1 minute. The document that caused the OutOfMemory exception obtains status Retry and will be included later (around 1 hour) in the new pass of the WorkZone PDF Crawler.

Tip:

You can set the custom restart period in the **Restart service after** field on the **Recovery** tab of the KMDWorkZone PDF Crawler service properties. To open, click **Control Panel** > **Administrative tools** > **Services** form in the server.



Advanced logging in WorkZone PDF Engine and WorkZone PDF Crawler

To enable custom logging, you can use Serilog configuration. See Serilog documentation.

Tables

Conversion states

In the DVS_RENDER_INFO table, you can view the status of files that have not successfully been converted by WorkZone PDF Crawler. The table can be accessed through the ScanSQL program.

Note: The status information in **DVS_RENDER_INFO** table is related to user-requested and policy-based conversions only (not files preview in WorkZone Client).

Field	Descrip	otion		
record_key	The uni	The unique record key (aktnr).		
render_type	PDF	PDF		
state	May contain one of the following states: PENDING, FAILED, REVIEW, RETRY, or SKIPPED.			
		PENDING	A requested conversion is pending.	
		RETRY	The document conversion has failed but the WorkZone PDF Crawler will wait at least an hour in order to ensure there are free system resources and then attempt to convert the document again when WorkZone PDF Crawler is idle.	
			By default, WorkZone PDF Crawler retries the conversion 3 times. You can set the number of retries in the WorkZone PDF Crawler parameters. When the specified number of retries has been reached, the state changes to FAILED.	
			For more information, see Configure	

Field	Description	
		WorkZone PDF Crawler.
	FAILED	The document conversion has failed.
		A conversion process can fail for a number of reasons. For example, the conversion takes more time than permitted or conversion requires more system resources than available
		After an upgrade of WorkZone PDF Crawler, it attempts to convert documents with the state FAILED again. It re-converts one previously failed document at time and only when WorkZone PDF Crawler is idle. The reconversion does not affect the user requested and policy based conversions as they always have higher priority.
		If the conversion is successful, entry in the DVS_RENDER_INFO table is deleted. If it fails again, the Updated and the Message fields are updated with new information.
	REVIEW	The document has not been converted.
		The user or the administrator needs to perform actions on the document for the document to convert successfully. The Message field contains a description of the issue that must be corrected to make the document convert successfully. For example, the document is password-protected or the document content exceeds the bounds of the document.
	SKIPPED	The document file type is not supported and, therefore, the document cannot be converted. See Supported file types.

Field	Description	
	The Message field will display the file type of the document that cannot be converted.	
	WorkZone PDF Crawler will not try to re-convert documents with the state SKIPPED until the specific file type is supported.	
message	The latest error message that relates to this conversion.	
	Note: The actual localized messages are located in the table DVS_ RENDER_MESSAGE.	
exception	Full stack trace of the exception in case the conversion has failed.	
updated	The time that this conversion was last acted on.	
retries	The number of retries if a temporary failure occurs (see RETRY above).	
server	The name of the server on which a document was converted.	
For example		

For example

Example:

RECORD_KEY: 20

RENDER_TYPE: PDF

STATE: FAILED

MESSAGE: HttpRequestException: ServerError 500: Internal Server Error

EXCEPTION: System.Net.Http.HttpRequestException: ServerError 500: Internal Server Error at KMD.WorkZone.ArchiveCrawler.HttpResponse.ReadContentAsStream() at KMD.WorkZone.ArchiveCrawler.RecordProcessor.ProcessRecord(String recordKey,

IPolicy policy)

UPDATED: 2018-12-12 14:21:53

COLLECTION:

RETRIES:

SERVER: SJORA01 18.2.16077.1/

Failed PDF conversion error messages

Error Message	Description
The conversion was not completed due to insufficient memory on the server.	The conversion process requires more memory on the server than it was originally assigned. The conversion process will attempt to convert the document later.
The conversion was not completed within the specified time	Each document is allotted a specific amount of time for conversion. This is done to ensure that a large document doesn't stall the conversion of many other smaller documents.
interval.	For one reason or another, the document was not converted during the allotted time. The conversion process will attempt to convert the document later.
The document(s) can not be converted to PDF format.	- For reasons other than the available error messages can display, the document in question was not converted to PDF format.
The document is password-protected.	The PDF conversion process cannot convert documents that are password-protected.
The document is unsupported.	The PDF conversion can only convert certain document types to PDF. See <u>Supported file types</u> .
The content(s) exceeds the bounds of the document.	The document contains content (images, text, or other types of content) which exceeds the boundaries of the document itself.
Re-merge failed for document [Title]. Reason: [Reason]	WorkZone PDF could not perform the remerge action/operation.

Supported file types

- File type or program File types supported by WorkZone PDF. If WorkZone PDF encounters file types that are not supported, they will be skipped and the action will not be completed.
- File extension A list of file extensions supported by WorkZone PDF.
- Font check If the <u>FailIfMissingFonts</u> parameter is enabled, WorkZone PDF checks availability of fonts on the IIS server for the file types marked with ♥. The checking happens during the document conversion to PDF or PDF/A format.

Tip:

If the font is missing and you still want to convert the document with another font, you can set the alternative font in the FallbackFont parameter.

File type or program	File extension	Font check
AutoCAD	DWG	
Email	EML	
	VCF	
Fixed Layout	PDF	
Images	ВМР	
	EMF	
	GIF	
	ICO	
	JPG	
	PNG	
	TIFF	

File type or program	File extension	Font check
Text	CSV	
	LOG	
	SQL	
	TXT	
Microsoft Excel	XLA	
	XLAM	
	XLC	
	XLM	₹
	XLS	₹
	XLSB	
	XLSM	
	XLSW	•
	XLSX	₹
	XLT	₹
	XLTM	
	XLTX	
Microsoft Outlook	MSG	
	OFT	

File type or program	File extension	Font check
Microsoft PowerPoint	POT	
	POTM	
	POTX	•
	PPA	•
	PPAM	•
	PPS	•
	PPSM	•
	PPSX	•
	PPT	•
	PPTM	•
	PPTX	
Microsoft Project	MPP	
	MPT	•
	MPX	₩

File type or program	File extension	Font check
Microsoft Visio	VDW	
	VDX	
	VSD	
	VSDX	
	VSS	₽
	VST	₽
	VSX	₽
	VTX	
Microsoft Word	DOC	•
	DOCM	
	DOCX	•
	DOT	
	DOTM	
	DOTX	
	RTF	•
Open Office	ODP	•
	ODS	•
	ODT	•
	OTS	•
	ОТТ	•

File type or program	File extension	Font check
Others	ZIP	
	ZIPX	
WordprocessingML	XML	
Web	HTM	
	HTML	•

Auto-updated Word fields

When WorkZone PDF converts documents to PDF, it automatically triggers specific Word fields to update. This is to generate a PDF document that contains the most recent information. The Word fields are updated during the conversion in the whole document, including **Header** and **Footer**.

Important:

Word fields in the archived documents (ARK, UL, or AFS) are not updated during the conversion.

In the following table, you can see the list of Word fields that are updated during the conversion. For more detailed information on each of them, see List of field codes in Word.

Description
Table of content.
Table of authorities.
Number of the page.
The total number of pages in the document.
The total number of words in the document.
The number of characters in the document.
The paragraphs' numbers.

Word fields	Description
AutoNumLgl	The paragraphs' numbers for legal and technical publications.
AutoNumOut	The paragraphs' numbers in outline style.
ListNum	A set of numbers anywhere in a paragraph.
PageRef	Page number of a bookmark for a cross-reference.
Sequence	Sequence numbers of chapters, tables, figures, and other items in a document.
Section	The number of the current section.
SectionPages	The total number of pages in a section.

Troubleshooting and tips

Click an issue below to see the solution or workaround.

Time for converting large documents is not sufficient

By default, ASP.NET grants 110 seconds to convert a document. However, in some cases, this time may not be sufficient. If this is the case, you can extend the timeout to any number you want. In this example, we'll show how to extend timeout to 600 seconds (10 minutes).

- Open Web.config file in the WorkZone PDF Engine directory. Its default location is the following: C:\Program Files
 (x86)\KMD\WorkZone\IIS\WorkZone\Render\Web.config.
- 2. Add the following parameter executionTimeout="600" to httpRuntime, for example:

3. Save the changes.

Error message "This access control list is not in canonical form and therefore cannot be modified"

If you see this message during the WorkZone PDF Engine installation, try one of the suggested solutions in articles: 1 and 2.

Invalid DSN during installation

If the installation fails with the message "DSN is invalid. Installation cannot be performed", you must ensure that the ODBC Data Source Name has the same name as the Transparent Network Substrate (TNS) Service name of the WorkZone Content Server Oracle database. For a solution to this problem, see The installation uses different names for DSN and TNS.

Note:

All ODBC DSN names are listed in the registry under: HKLM\SOFTWARE\
(Wow3264Node\)ODBC\ODBC.INI.

The installation uses different names for DSN and TNS

If your installation does not use the same name for ODBC (DSN name) and Oracle (TNS name) access to your database, you need to create a duplicate TNS name which is identical to the DNS name before running the WorkZone PDF Crawler installation. This extra TNS name is required only during installation, and can be removed after installation.

If your database uses the TNS name <TNS>, you can expose it as <DSN> from the database server by following the steps below:

- 2. Find the definition of the TNS name of your database:

```
<TNS> = (DESCRIPTION = (ADDRESS = (PROTOCOL = TCP) (HOST = scjor-
a04.scanjour.dk) (PORT = 1521))
(CONNECT_DATA = (SERVER = DEDICATED)
(SERVICE_NAME = UV006.scanjour.dk)))
```

Where <TNS> is the Oracle TNS name of your database.

- 3. Copy the lines above.
- 4. In the copy, change the Oracle TNS name to be the same as the DSN name you use with ODBC. The new entry will look like this:

```
<DSN> = (DESCRIPTION = (ADDRESS = (PROTOCOL = TCP) (HOST = scjor-
a04.scanjour.dk) (PORT = 1521))
(CONNECT_DATA = (SERVER = DEDICATED)
(SERVICE_NAME = UV006.scanjour.dk))
```

Where <DSN> is the new name.

Do not change the old entry.

- 5. Save and exit thshames.ora.
- 6. Verify that both TNS names work. Open a command prompt on the agent server and enter:

```
tnsping <TSN>
tnsping <DSN>
```

In both cases, the command should reply with a copy of the entry in tnsnames.ora followed by OK (20 msec) or similar.

The sent date and time in PDF version of an email may not reflect the time when the email was sent

This is due to the fact that the sent date and time of an email is the date and time of the Exchange Server time zone. So, if the sender of the email is in a different time zone than that of the Exchange Server, the sent date and time of the email is not the date and time where the sender sent the email. This is Outlook default behavior. The conversion of an email to a PDF document does not affect the date and time.

The format of the sent date in the PDF document is the default format, which is the format defined on the server. The timestamp format is UTC with time zone offset.

Note that UTC does not observe daylight saving time.

Documents with custom fonts do not convert correctly

When converting documents with custom fonts to PDF, the documents will be converted but they may not use the custom font and therefore not display correctly.

To make the documents convert correctly, install the custom font on the server where WorkZone PDF Engine is running.

Note:

After installing custom fonts on the server, recycle the PDF Engine App Pool.

Alternatively, set up the fallback font, which will be used instead of missing custom font.

The date is not displayed in the expected format

Dates are displayed in the format that is defined in WorkZone Configurator by default, unless you have defined another date format in the document. If the document does not display the expected date format, you can manually define the desired date format in the document, which will then be used instead of the standard date format from WorkZone Configurator.

See also

View errors and logs

Troubleshooting in the Installation Guide.

Terms and conditions

Intellectual property rights

This document is the property of KMD. The data contained herein, in whole or in part, may not be duplicated, used or disclosed outside the recipient for any purpose other than to conduct business and technical evaluation provided that this is approved by KMD according to the agreement between KMD and the recipient. This restriction does not limit the recipient's right to use information contained in the data if it is obtained from another source without restriction set out in the agreement between KMD and the recipient or by law.

Disclaimer

This document is intended for informational purposes only. Any information herein is believed to be reliable. However, KMD assumes no responsibility for the accuracy of the information. KMD reserves the right to change the document and the products described without notice. KMD and the authors disclaim any and all liabilities.

Copyright © KMD A/S 2021. All rights reserved.