Content Services for EMC Centera 1.1 Administration Guide

neeb & partner GmbH

Bartningstraße 55
D-64289 Darmstadt

Tel.: +49 (0) 6151 - 1793 - 0
Fax: +49 (0) 6151 - 1793 - 190
e-mail: np.info@np-gmbh.com

documentum
Signature Partner

Bartningstraße 55
D-64289 Darmstadt

Tel.: +49 (0) 6151 - 1793 - 0
Fax: +49 (0) 6151 - 1793 - 190
e-mail: np.info@np-gmbh.com
Internet: www.np-gmbh.com

Table of Contents

1. Introduction

Content Services for EMC Centera (CSEC) integrates EMC Centera into Documentum. It provides a rule-based fixed content storage while all search, retrieval and access capabilities of Documentum remain.

The rules for an archiving process are defined by the familiar Documentum Query Language (DQL). Job-scheduled, the rules are executed at a given time and the corresponding documents and metadata are transferred from a Documentum Docbase into a configurable Centera cluster.

The data transfer is ensured by transaction security so that errors are detected and logged at once. After a successful archiving operation a unique C-Clip reference is saved and the content deleted from the Docbase. With the installation of the system a reference table called `np_safe_relation` is created in Documentum. It organizes all references from the Centera cluster (See appendix 7.1).

Thus, the Documentum system is cleared from this content, but the functionalities for display and inquiry of the documents remain to the full, because just the content is deleted, not the corresponding object information.

The whole operation is completely transparent for the end-user. They can still use the documents in the system – as before the archiving. If the end-user wants to open an archived document, the Documentum system automatically recognizes it for an archived document and starts a retrieval operation to make the document available to the end-user.

During the retrieval operation, the document is extracted with the help of the C-Clip reference from the Centera cluster and made available to the user. The user himself does not realize if the document was locally stored or had to be fetched from the archiving system.

To integrate CSEC into the Documentum system two independent processes are necessary:

• Archiving of the objects:	**Archiver**
• Retrieval of the objects:	**Reader**

For the integration of the modules in Documentum any number of **Docbases** can be configured. For access to a certain Docbase the corresponding access rights must be configured. The selection of the objects, that are to be archived, is controlled by the **Rules**.

The configuration of the modules is done by the central configuration file (*csec_config.xml*). It contains four main sections, where all components and their attributes are defined.

- **Docbase**
- **Archiver**
- **Reader**
- **Rules**

The following chapters will give further descriptions for each section.

2. Archiver

The Archiver is responsible for the archiving of the documents. With the help of configurable Rules this process selects the documents for the next archiving operation. As the operation needs Documentum functions, it requires the corresponding system rights of the *Documentum Installation Owner*.

At the installation of the system a Documentum Job is created, which ensures a continuous execution of the Archiver with adjustable criteria. For the execution the Archiver Process can be controlled by the following command line parameters:

Parameters:

Parameter	Description
`-d docbase`	Name of Docbase to process This parameter starts the Archiver for a certain Docbase. The Docbase must be configured in the Docbase section of the configuration file. Without this parameter every configured Docbase is processed. (See configuration section Docbase)
`-c container`	Name of container to process Containers exist to group configured Rules in a logical order. Thus, certain Rules can be assigned to a defined Archiver operation, e. g. to assign Rules to a certain Docbase (See configuration section Rules 4.4).
`-i r_object_id`	'`r_object_id`' of document to archive To archive a document with a certain `r_object_id`.
`-m count`	Maximum number of objects to process With this parameter can be defined, how many documents should be processed per archiving operation at the most. All remaining documents are processed at the next archiving operation. This setting helps to implement an intelligent load balancing at peek-periods.
`- dmcl_trace`	Turn on dmcl tracing This parameter turns on the extended log-function. All Documentum API commands are logged by the system during archiving.
`- h`	Help screen

Table 1: Archiver command line parameters

3. Reader

The Reader is responsible for the retrieval of the archived documents. If an archived document is accessed, the Reader performs the retrieval process of the content. If the content is successfully transferred, Documentum regains control to display the data in correspondence to the data format in read-only mode. With this operation, the object is not restored in Documentum, but is just made temporarily available to the user.

The Reader process is automatically started by Documentum if an archived document is requested and does not need to be provided with different command line parameters. The configuration of the Reader is done in the central configuration file (*csec_config.xml*) (See configuration section Reader 4.3).

4. Configuration File

The configuration of the system is done with the configuration file ***csec_config.xml***. At the installation, this file is provided with standard values and is located in the root directory of the Documentum installation (environment variable `DOCUMENTUM`).

The configuration file itself is a XML document an can be edited with any XML or text editor. The following subchapters give a closer analysis of the separate sections of the configuration file.

4.1. Docbase Config

In the section `<docbase_config>` the configuration of a Docbase is defined. Any number of `<Docbase>` sections can be defined within the `<docbase_config>` section. Example for two different Docbases:

```
<docbase_config>
   <docbase>
      <name>docbase1</name>
      <login_user>dbuser1</login_user>
      <login_pwd>dbpassword1</login_pwd>
      <name_of_filestore>np_safe_store</name_of_filestore>
      <centera_pool_address>192.168.100.10:3218</centera_pool_address>
      <centera_streaming>true</centera_streaming>
   </docbase>
   <docbase>
      <name>docbase2</name>
      <login_user>dbuser2</login_user>
      <login_pwd>dbpassword2</login_pwd>
      <name_of_filestore>np_safe_store</name_of_filestore>
      <centera_pool_address>192.168.100.20:3218</centera_pool_address>
      <centera_streaming>true</centera_streaming>
   </docbase>
</docbase_config>
```

The configuration parameters of a Docbase have the following meaning:

Parameter	Description
`<name>`	Name of the Docbase
`<login_user>`	Documentum Installation Owner
`<login_pwd>`	Password of the Documentum Installation Owner
`<Name_of_filestore>`	Name of the external filestore (Default: „np_safe_store“)
`<centera_pool_address>`	IP-Address and Port of the Centera cluster. At this point a range of IP-addresses of the Centera cluster can be provided, separated by comma. Example: 192.168.100.20:3218, 192.168.100.21:3218, …
`<centera_streaming>`	This parameter regulates the way the content is transferred from the Centera cluster. Possible values: true Content is transferred in blocks of 16 kb and made available to Documentum. false Content is completely transferred at first, temporarily stored and then made available to Documentum.

Tabelle 2: Docbase configuration

> *It is absolutely necessary to check the connection to the Centera cluster with CenteraVerify before the configuration.*

If the Archiver process is started with a certain Docbase as parameter **-d docbase** (See section 2 Archiver), the Archiver searches the corresponding configuration parameters, which were given as `<name>`, in the configuration file.

4.2. Archiver Config

The section `<archiver_config>` in the central configuration file *csec_config.xml* defines the parameters for the control of the Archiver process.

The configuration parameters have the following meaning:

Variable	Description
`<archiver_log_path>`	Path for logfiles Please give the directory into which the generated logfiles shall be written. For a better readability each day a new logfile with the date and name of the process is generated. (archiver_mmddyy.log)
`<archiver_temp_path>`	Path for temporary files During operation the Archiver process creates temporary files that are deleted afterwards. Please name a directory where temporary files can be stored.
`<archiver_log_level>`	Log level for all actions Depending on how detailed the entries in the logfiles are required to be, you can set different values (See the appendix 7.2 for a further description).
`<archiver_max_retry>`	Max. number of archiving trials at an archiving operation
`<archiver_max_objects>`	Max. number of objects to be archived at an archiving operation
`<archiver_update_sysobject>`	Change of `a_storage_type` in the Documentum object `dm_sysobject`. Possible values: true false

Table 3: Archiver configuration

The parameter `<archiver_update_sysobject>` states if the attribute `a_storage_type` of the `dm_sysobject` is to be changed in Documentum at the archiving process. If this parameter is set to *true*, the user sees with the help of the attribute `a_storage_type` that the object is in the archive (`a_storage_type` = np_safe_store). If the value of the parameter is *false*, the user cannot see if the object is archived (`a_storage_type` = filestore01).

4.3. Reader Config

The configuration parameters for the Reader in the file *csec_config.xml* are:

Variable	Description
`<reader_log_path>`	Path for logfiles Please give the directory into which the generated logfiles shall be written. For a better readability each day a new logfile with the date and name of the process is generated. (reader_mmddyy.log)
`<reader_temp_path>`	Path for temporary files During operation the Archiver process creates temporary files that are deleted afterwards. Please name a directory where temporary files can be stored.
`<reader_log_level>`	Log level for all actions Depending on how detailed the entries in the logfiles are required to be, you can set different values (See the appendix 7.2 for a further description).
`<reader_max_retry>`	Max. number of retrieval trials
`<reader_fail_content_path>`	Path for the failcontent

Table 4: Reader configuration

Should an archived object be unavailable (e. g. because the Centera system cannot be reached or an error has occurred), a failcontent is redelivered. The selection of the locally stored failcontents is made with the help of the file type. The administrator can individually design the failcontents. Standard failcontent files are part of the installation package.

4.4. Archiving Rules

In the section `<rules>` of the configuration file the rules of the archiving operation are defined. Any number of `<rule>` sections can be defined within the `<rules>` section. Example for two different rules:

```
<rules>
      <rule>
            <description>First rule</description>
            <container>Con1</container>
            <archiving_acl>pre_archive_acl</archiving_acl>
            <archived_acl>post_archive_acl</archived_acl>
            <dql>
                  SELECT r_object_id FROM dm_document (all) WHERE
                  FOLDER('/Centera_Store')
            </dql>
      </rule>
      <rule>
            <description>Second rule</description>
            <container> Con2</container>
            <archiving_acl>pre_archive_acl</archiving_acl>
            <archived_acl>post_archive_acl</archived_acl>
            <dql>
                  SELECT r_object_id FROM dm_document
                  WHERE FOLDER('/Test_Store')
            </dql>
      </rule>
</rules>
```

An archiving rule consists of the following parts:

- DQL statement (select ...) *What* to archive?
- ACL archiving (ACL 1) *How*? Access rights **during** the archiving.
- ACL archived (ACL 2) *How*? Access rights **after** the archiving.
- Container *What*? Attribute of a rule (free text)
  If a container is given as parameter of the archiving process, only the rules with this container are performed.

Caution: Please note that the DQL statements should only select objects you really want to archive. A selection of system objects can lead to problems with the Documentum system.
All defined rules must include the column `r_object_id` in the result. (See examples for DQL statements)

The access rights to the archived documents are set with the parameter `<archived_acl>` of the archive rules. At the most it should only be allowed to create a new version of an archived document. With the installation of CSEC two System ACL's are created in Documentum:

- pre_archive_acl
- post_archive_acl

The pre-configured ACL's should be adjusted to your needs after the installation. Additional System ACL's in the Documentum system can be created for other rules.

Examples for DQL queries:

```
SELECT r_object_id FROM dm_document WHERE FOLDER('/XYZ', descend)
```

(Archive all objects of a given cabinet/folder)

```
SELECT   r_object_id   DROM   dm_document   WHERE   r_object_type   =
'document'
```

(Archive all objects of a given document type)

```
SELECT r_object_id FROM dm_document (ALL) WHERE ...
```

(Archive all versions of any objects / without “(all)” only the recent version is archived)

```
SELECT r_object_id FROM dm_document WHERE group_name = 'XYZ'
```

(Archive all objects of a given usergroup)

```
SELECT r_object_id FROM dm_document WHERE r_modify_date > ' ... '
```

(Archive all objects from a certain date)

5. Scheduling

With the installation of the system a Documentum Job „*archive_job*" is created that executes the created method „*archive_method*" at regular intervals. This method starts the Archiver to initiate an archiving operation.

After the installation the created Job „*archive_job*" and method „*archive_method*" should be adjusted to your needs and the Job must be activated. To execute the Archiver process at different times with different parameters it is possible to create further Documentum Jobs and methods. Further information to create Jobs and methods can be found at the *Server Admin Guide* of Documentum.

6. FAQ – Frequently Asked Questions

How to detect if a given document is archived?

- The archiving operation can be reproduced with the help of the logfiles if the Archiver process operates with the parameter log-level 3 (workflow).

- With the help of the `r_object_id` of a Documentum object the following DQL statement can check on which store the object resides:

```
SELECT name FROM dm_store WHERE r_object_id IN
(SELECT  storage_id  FROM  dmr_content  WHERE  ANY  parent_id=
   '<r_object_id>')
```

- With the help of the following DQL statement all archived objects can be displayed:

```
SELECT f.object_name AS foldername,d.object_name, r.status FROM
   dm_document (ALL) d, np_safe_relation r, dm_folder f WHERE
   d.r_object_id = r.oid AND ANY d.i_folder_id = f.r_object_id
   ORDER BY f.object_name, d.object_name
```

An object was not archived even though it is part of the DQL statement result. Under which circumstances can this happen?

- Content of the object is connected with another object.
- Repetition trials have reached the permitted maximum.
- Lost Connection to the Centera cluster.
- Check current ACL of the object for write access for the Installation Owner.
- Check configuration parameters.

The error "Can't read configuration file" appears. Why did this happen?

- The configuration file *csec_config.xml* is not accessible. (E. g. wrong directory or corrupt)
- The needed environment variable (DOCUMENTUM) does not exist.

The Archiver process is operating normally but no document is archived. Why?

- The syntactic of the DQL statement is correct but the result is useless (no `r_object_id` in the result).
- The syntactic of the DQL statement is wrong (Can't execute DQL query).
- All selected objects are already archived.
- Parameter `max_objects` is too low.
- Parameter `max_retry` is too low and there are connection problems.

An archived document should be displayed, but the screen only shows the failcontent document. What has to be checked?

- Check the availability of the Centera cluster (CenteraVerify).
- Check all configuration parameters.

An archived document should be displayed, but the screen only shows an empty document. Why?

- The failcontent path in the configuration is wrong and the retrieval of the document failed. (Take a look at the last question)

7. Appendix

7.1. Reference Table

The reference table `np_safe_relation` lists all documents that are to be archived or that already are in the archiving system.

The reference table has the following structure:

OID	SOID	AI	ARCHIV_ID	CONTAINER	STATUS	STATUS_DATE	RETRIES	TEMP

Table 5: Structure of the `np_safe_relation` table

With the help of the table entries archived objects and their state can be determined (See table 6 State values). The table columns have the following meaning:

- `oid`: *r_object_id*
  Unique id of the Documentum object.
- `soid`: *string r_object_id*
  A default character string and the `r_object_id`. This value is not changed by the Documentum dump & load mechanism.
- `ai`: *Archive Instance*
  Archive instance of the archived object
  (Default: CSEC_DEFAULT_INSTANCE)
- `archive_id`: Unique C-Clip reference. This references the object at the archive system.
- `target`: This maps to the attribute “container” of a rule. To group configured Rules in a logical order.
- `status`: Current state of the object. (See table 6)
- `status_date`: Date of the last change of state. If the object is already archived, this date gives the time of archiving.
- `retries`: This value shows how many archive trials had been necessary before the object was successfully archived (e.g. in case of connection timeout).
- `temp`: Reserved.

The entries of the `np_safe_relation` table can have different state values. The state value gives information about the state of the object in relation to the archive operation. The following states are defined:

State	Description
1	Marked for archiving
2	During archiving
3	Archived
10	Can't get metadata
11	Can't assign archiving ACL
12	Can't assign archived ACL
13	Can't change filestore
14	Can't set Archive Primary Key / Status
15	Can't archive Object

Table 6: State values `np_safe_relation`

7.2. Log-level Definition

Entry	Log-level	Description
0	`NONE`	Turn off the log function (only the copyright message).
1	`ERROR`	Only error messages are logged.
2	`WARNING`	In addition to error messages warning messages are logged. Warning messages result from discrepancies in the system but do not stop the operation.
3	`WORKFLOW`	Every operation is logged.
4	`DETAIL`	In addition to the operations all used values are logged.
5	`DEBUG`	Full description of the operation with the intention of finding any error without the help of a debugger program. Use this level for testing purposes only.

Table 7: Log-level definition