**Configuring and Running the CEBIT (Bulk Import) tool:**

Bulk Import Tool is installed in the following default directory:

* Windows: **C:\Program Files\FileNet\ContentEngine\tools\CEBI**
* UNIX:/opt/IBM/FileNet/ContentEngine/tools/CEBI

**Preparing to use Bulk Import Tool:**

1. Prepare transact.dat file like below

**class\_code:document\_properties:external\_index:files|+file\_name**

where:

***class\_code***

Is one of the codes that is specified in the CEBI.cfg file for a document class that is involved in the import.

***document\_properties***

Includes values for each of the properties that are specified in the IndexNames parameter in the CEBI.cfg file. The values must be listed in the same order that they are listed in the DocClassAttribute parameter in the CEBI.cfg file. If a property is set to **Required** on the Content Platform Engine server, the Bulk Import Tool fails a batch if that property is not set in the transact.dat file. Spaces are considered valid characters, so a required property that contains only spaces can be processed by Bulk Import Tool. Additionally, if there are other characters in the field, the leading and trailing spaces are not removed before they are processed by Bulk Import Tool. Setting a space or non-integer in a property field of an integer returns the error message **Value = Bad number conversion** and causes the batch to fail.

**Multi-valued entries with examples for the transact.dat file that use | as a delimiter:**

**Integer** (integer): 19:{3|1088|45},DocNameMVint::test.txt

**String** (string): 03:{This|is|test},DocNameMVstr::test.txt

**Binary** (Base64): 10:{9876|54321},DocNameMVbin::test.txt

**Date/Time** (date or date and time): 02:{03/07/2012T13:28:49.123z|09/17/2011T02:01:00.888z},CEBIMVDateTimeFile::test.txt

**Float64** (exponential) - 06:{12345678901234567890|12345678901234567890|3.21E5},CEBIMVFloat64::test.txt

**ID** (GUID): 07:{562F02F1-E172-41AB-B041-E02A6C0DA9EC|562F02F1-E172-41AB-B041-E02A6C0DA9ED},CEBIMVid::test.txt

**Boolean** (zero/one):05:{no|yes|true},CEBIMVboo::test.txt

**Choice list entries in the transact.dat file where Green, Blue, and Red are valid options from the choice list.**

Single-valued entry: 17:Green,CEBIStr::test.text

Multi-valued entry: 17:{Green|Blue|Red},CEBIStr::test.text

**Important:** To process choice lists entries by using the Bulk Import Tool, all of the entries in the choice list must be the specified data type. For example, if the data type is String, all of the entries in the choice list must be strings. The entries cannot be a combination of string, integer, or any other data type.

***external\_index***

Is a value that is recorded in the report when the batch is completed. This value is not imported as part of the document.

***files***

Specifies one or more names of the content element files that are associated with the document.

***file\_name***

Is the name of a text file that contains a list of the content element file names that are associated with the document.

A simple entry in the transact.dat file looks like the following example. The document uses one image, named cat\_picture.jpg, as its content element.

03:67890,"Hi there",3320.09:This is a test:cat\_picture.jpg

An entry that lists multiple content element file names looks like the following example. The document uses two images, named kids.tif and ocean\_surf.jpg, as its content elements.

03:67890,"Hi there",3320.09:This is a test:kids.tif,ocean\_surf.jpg

Alternatively, an entry that specifies a text file that lists the content element file names looks like the following example. The text file x.txt specifies two images, named kids.tif and ocean\_surf.jpg, as its content elements.

03:67890,"Hi there",3320.09:This is a test:+x.txt

The actual entries in the x.txt file are as follows:

kids.tif

ocean\_surf.jpg

**Configuring the Bulk Import Tool:**

1. Open the CEBI.cfg file and enter the following login information in the LogonAttribute section:

**UserName**

The user name for logging in to Content Platform Engine.

**PassWord**

The password for logging in to Content Platform Engine.

**CeUri**

The Content Platform Engine URI. Example: http://cm-vmwtcl4a:8080/wsi/FNCEWS40MTOM.

**ObjStore**

The name of the object store that contains the document classes and documents. Example: bkmOS8nomove

To specify multiple object stores, create a CEBI.cfg file for each object store. Start Bulk Import tool with the **-h**<file list of home directories> flag. Each home directory that you specify requires a CEBI.cfg that uses a specific object store.

1. Enter the document class and index information in the DocClassAttribute section. You can enter multiple DocClassAttributes sections. Each DocClassAttributes section must describe only one document class. However, you can enter duplicate DocClassAttributes sections for the same class if the sections have different class codes.

For example, you can add documents that have specific sets of properties that are configured for each DocClassAttributes section to the same document class.

To generate files that contain your document class information for the specified object store, enter the following command:

java -cpclasspath;BulkImport.jar bulkImport.BI\_Start -h /home path -G

Enter your class path information and the Bulk Import tool directory for home path. The files DocClassAttributes.txt and DocClassAttributes.xml are created in the home directory that you specified in the command. Use the information in either of these files to update your CEBI.cfg file to specify the following parameters:

**ClassName**

The symbolic name of the class in the object store where you want to import your documents.

**ClassCode**

The class code. The class code can be a number 1 - 999. This number is generated by the -G option in the Bulk Import Tool. It does not come from the Content Platform Engine server. You can manually set this code to any valid number that is unique in the class codes that are listed in the CEBI.cfg file.

**IndexName**

The symbolic name of the Content Platform Engine class property that is associated with that Content Platform Engine class. You can add as many **IndexName** parameter entries as you need. However, you must include the properties that are set as required on Content Platform Engine. These required properties must have a valid entry in the transact.dat file for the batch to process correctly.

The value entries for multi-valued properties must be contained in braces and separated by the multi-valued delimiter. For example, {multi\_value1|multi\_value2|multi\_value3}. If you specify 1 value for a multi-valued property, that value must still be contained in braces, such as {multi\_value}.

Property fields that are of the type **Date** must provide a valid date mask that follows the property name. The date mask must follow this format:

* + CEBI.cfg format example: IndexName=CEBIMVdatetime:"MM/dd/yyy'T'HH:mm:ss.SSS'z'"
  + transact.dat format example that uses the same date/time formatting for multi-valued properties with two values:02;{03/07/2012T13:28:49.123z|09/17/2011T02:01:00:888z},CEBIMVDateTimeFile;;test.txt

**Restriction:** If milliseconds are specified, they are rounded to the nearest second when the document is imported to Content Platform Engine.

You can modify the date format, for example yyyy/MM/dd or dd/MM/yyyy. But your format must match the date format that is used in the .dat file. 'T' must be uppercase and 'z' can be either case if it matches the format in the .dat file.

The order of the **IndexName** entries is the expected order of the values that are found in the transact.dat file.

You can use multi-valued properties and choice lists.

1. Enter the following batch information:

**Important:** When you specify a **Date/Time** property, you cannot use a colon or a forward slash character as a delimiter because colons are used in the **Time** property and forward slashes are used in the **Date** property. This restriction applies when you set **FieldDelimiters**, **ItemDelimiters**, and **MultiDelimiters**. You can use another character as a delimiter, such as a semi-colon.

**MaxSubBatchSize**

The maximum size of a subbatch in 1 K units. The MaxSubBatchSize setting controls the size of network packets. The default is 1024 bytes.

**MaxDocPerSubBatch**

The maximum number of documents of a subbatch. The default is 999.

**WorkingDirectory**

The working directory where Bulk Import tool searches for the batches of documents that you want to import. The directory can be on a local drive or a shared network drive. This setting is required.

**JournalDirectory**

The directory where the Bulk Import tool log files are written. If not specified, the journals directory is created in the WorkingDirectory.

**FieldDelimiter**

Sets the major delimiter in the transact.dat file. The default is colon. Special characters can be used: ASCII 0x01 to 0xFF, except for 0x20 (space). To be able to use the extended ASCII codes from 0x80 to 0xFF, the CEBI.cfg and transact.dat files must be created with ANSI encoding.

**ItemDelimiter**

Sets the minor delimiter in the transact.dat file. The default is comma. Special characters can be used: ASCII 0x01 to 0xFF, except for 0x20 (space). To use the extended ASCII code from 0x80 to 0xFF, you must create the CEBI.cfg and transact.dat file files with ANSI encoding.

**MultiDelimiter**

Sets the Multi\_Value delimiter in the transact.dat file to separate multi-valued property lists. The default is a vertical bar (|). Special characters can be used: ASCII 0x01to 0xFF except for 0x20 (space). To use the extended ASCII code from 0x80 to 0xFF, you must create the CEBI.cfg and transact.dat files with ANSI coding.

**SleepInterval**

The total number of seconds that Bulk Import tool waits from the initial search for available batchname.eob files to the next search for available batchname.eob files. This interval is not affected by how long it takes Bulk Import tool to process its current work unless the current work exceeds the sleep interval time setting. If the work exceeds the sleep interval, then another sleep interval is appended to the end of the last sleep interval before Bulk Import tool does another search for an available batchname.eob file. The default and suggested setting is 60 seconds. The maximum setting is 3600 seconds, 1 hour.

**DelayProcess**

Creates a delay in processing between the physical time stamp on a batchname.eob and when Bulk Import tool recognizes the batchname.eob file for processing. Setting this keyword to 300 causes a 5-minute delay between the time that the batchname.eob file is created and when Bulk Import tool reads the file for processing. This setting is often referred to as the "batchname.eob age". Set this parameter to zero seconds. The maximum setting is 3600 seconds, 1 hour. The default is 30 seconds.

**OSNice**

Sets Bulk Import tool to “sleep” mode between batch processing. Using this setting slows down Bulk Import tool and allows other processes to get processing time. This parameter is useful when you run Bulk Import tool in normal production time. The default is zero seconds and the maximum is 21600 seconds, 6 hours.

**Timing**

Log output that is used in performance analysis. The default value is FALSE. A setting of TRUE provides extra logging details. You can use the timer.awk program with the added log information for performance analysis.

UNIX systems and Windows systems with UNIX tools for DOS loaded: Some newer UNIX systems might require the use of **nawk** instead of **awk**. Example:

* + **awk –f timer.awk journals/imp20020328**
  + **nawk –f timer.awk journals/imp20040129**

**ExternalPassDirectory**

The directory where a copy of the batchname.pass file is written. This directory is usually an externally mounted directory that is visible to third-party programs.

**ExternalErrorDirectory**

The directory where a copy of the batchname.err file is written. This directory is usually an externally mounted directory that is visible to third-party programs.

**ExternalRptDirectory**

The directory where a copy of the batchname.rpt file is written. This directory is usually an externally mounted directory that is visible to third-party programs.

**PassCopyDirectory**

The directory where a copy of the batchname.pass file is written is usually a local directory.

Creating a .eob file

Last Updated: 2023-02-09

In the working directory, you must create a *batchname*.eob file, which is used to signal Bulk Import Tool that a batch is ready for import.

About this task

The *batchname*.eob file name is arbitrary but the .eob file extension is necessary in order for Bulk Import Tool to identify the file and process the work.

Create a *batchname*.eob file where entries in the file have the following format:

*path* *#docs* *#objects* *transact.dat\_name*

where:

***path***

The directory location of the transact.dat file.

You must specify a path in the .eob file. The path can be an absolute path or a relative path, such as ../location/MVDateTime1x1 1 1 transactCustom1.dat. The path ./ 1 1 is the minimum that you can use to run a batch on Bulk Import Tool.

***#docs***

The number of documents to be imported.

**Tip:** If a batch is too large and fails to complete processing, all of the items that were imported from that batch before processing failed are removed from FileNet® P8. You must run the entire batch again. As a guide, you can try running batches sizes of 5,000 items or less.

***#objects***

The total number of content elements to be imported.

***transact.dat\_name***

(optional) To specify a different name for the transact.dat file.

Procedure

To create a .eob file:

Using the following example format, create a file that contains the information for your system: /cebulkimportdata/EOB\_data 2 4

/cebulkimportdata/EOB\_data is the directory that contains the transact.dat file. The number two is the number of entries in the transact.dat file. The number four is the total pages in the batch. In this example, each entry has two content elements, two entries times two content elements each, equals four.

This example illustrates the transact.dat file.

03:67890,Document Title,3320.09:Comments here:+x.txt

x.txt is a file that lists two images that are to be used for that entry.