CS 10.5 Migration to HDMS

Migration Design Document

|  |  |
| --- | --- |
| Created By/Role | Arlette Cruz |
| Last Modified | Oct 26, 2023 |

# Table of Contents

[1. General Project Information 4](#_Toc103161116)

[2. Overview 5](#_Toc103161117)

[3. Scope 5](#_Toc103161118)

[4. Current State 5](#_Toc103161119)

[4.1 Users and groups 6](#_Toc103161120)

[4.2 eB Director 6](#_Toc103161121)

[4.3 Classes of contents 6](#_Toc103161122)

[4.4 Data classifications 6](#_Toc103161123)

[4.5 File and database size 6](#_Toc103161124)

[4.6 Data Backup 7](#_Toc103161125)

[5. Future State 7](#_Toc103161126)

[5.1 Migration Conceptual Design 7](#_Toc103161127)

[5.2 Target Solution Logical Design 8](#_Toc103161128)

[5.2.1 Functional modules or components 8](#_Toc103161129)

[5.2.1.1 Modules in HDMS 8](#_Toc103161130)

[5.2.1.2 Functional modules or components in PPH 9](#_Toc103161131)

[5.2.2 Information architecture 9](#_Toc103161132)

[5.2.2.1 Information architecture for Lima document management in HDMS 9](#_Toc103161133)

[5.2.2.2 Target site in PPH 12](#_Toc103161134)

[5.2.3 Network structure 12](#_Toc103161135)

[5.2.3.1 Decision on external access to HDMS 12](#_Toc103161136)

[5.2.3.2 External user access to HDMS 12](#_Toc103161137)

[5.2.3.3 External user access to PPH 13](#_Toc103161138)

[5.2.4 Kiosk machine access 13](#_Toc103161139)

[5.2.5 SAP integration 13](#_Toc103161140)

[5.2.6 Security architecture 13](#_Toc103161141)

[5.2.6.1 AD groups 13](#_Toc103161142)

[5.2.6.2 HDMS authentication 14](#_Toc103161143)

[5.2.6.3 HDMS user groups and permissions 14](#_Toc103161144)

[5.2.6.4 PPH authentication 15](#_Toc103161145)

[5.2.6.5 PPH user groups and permissions 15](#_Toc103161146)

[5.2.7 Disaster recovery design 15](#_Toc103161147)

[5.2.8 Lima document operations and support model 15](#_Toc103161148)

[5.2.9 Administration and monitoring 15](#_Toc103161149)

[6. Design Influential Factors 16](#_Toc103161150)

[6.1 Alignment to Roadmaps 16](#_Toc103161151)

[6.2 Design Assumptions 16](#_Toc103161152)

[6.3 Exceptions/new Standards 16](#_Toc103161153)

[6.4 Legal and Regulatory Considerations 16](#_Toc103161154)

[6.5 Data Classification 16](#_Toc103161155)

[Technical Debt / Risks 16](#_Toc103161156)

|  |  |  |  |
| --- | --- | --- | --- |
| **Document Version Control** | | | |
| **Revision** | **Date** | **Description** | **Author** |
| 0.1 | Oct 26, 2023 | Draft created | Arlette Cruz |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

| **Reviewers** | | |
| --- | --- | --- |
| **Review Team** | **Name** | **Date** |
| **Operations** | Ian Mayhood |  |
|  |  |  |

# Overview

After the completion of the migration of CS 10.5 content to CDMS, it was found that several documents linking to JDE and Maximo have not been moved.

Since HDMS will eventually become the single document repository in Cenovus, it was decided to move all these documents to HDMS instead, where they will be accessible to the archive project CART.

Accenture carried out the data assessment of the full CS 10.5 instance and the development of the scripts required to migrate the identified content. This document describes the migration steps that the managed services team will have to carry out to migrate all content over the following months.

# Data Assessment

## Documentation

The full data assessment results are recorded in the following document stored in SharePoint:

[CS10.5 - Data Assessment - 20231023.xlsx](https://cenovus.sharepoint.com/:x:/r/teams/ECM_PS/ECMWIP/Content%20Server/CS10.5%20Decommission/Arlette%20Data%20Assessment%20%26%20Code/CS10.5%20-%20Data%20Assessment%20-%2020231023.xlsx?d=w471cf63b679c45bcbd3e6ff3e13f46e0&csf=1&web=1&e=Eaktt6)

This document included the following:

* Summary of contents per object type rolled up to the root folder level.
* Summary of categories used rolled up to the root folder level.
* Summary of migrated folders, listing the folders identified by Cenovus to be migrated to HDMS, including number of items per object type and categories that should come over to HDMS as well.
* Analysis of JDE links, identifying the url types that were translated and the list of folders that contain documents liking to JDE.
* Analysis of Maximo links, identifying the url types that were translated and the list of folders that contain documents liking to Maximo.
* One tab per root folder with the detail of number of items per object type rolled up to the 3rd or 4th level of subfolders within.

## Migrated Objects

The following objects will be migrated:

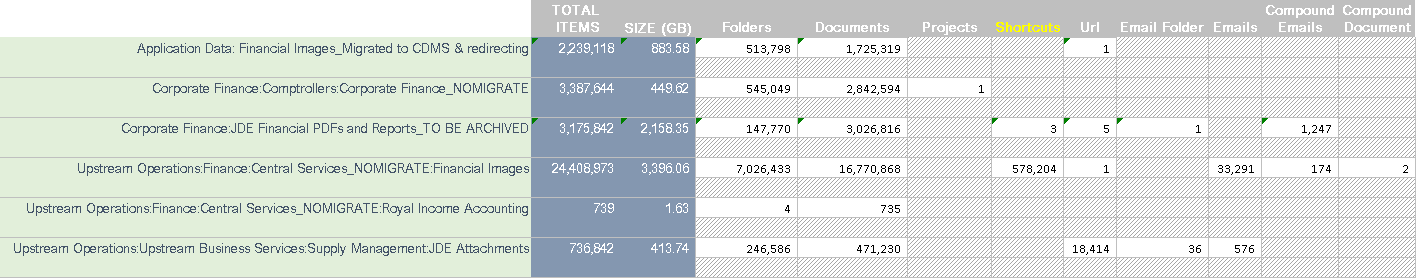
* Container object types:
* Folders
* Email folders
* Compound Documents
* Compound Emails
* Content object types:
* Documents
* Emails
* Urls
* Shortcuts that will be migrated as Urls

## Migrated Folders

The following table shows the CS 10.5 folders that will be migrated to the corresponding target folders in HDMS:

|  |  |
| --- | --- |
| **CS 10.5 Folder** | **HDMS Folder** |
| Application Data: **Financial Images\_Migrated to CDMS & redirecting** | Enterprise:Department Administration:Finance:z-Archives:**(Application Data) Financial Images** |
| Corporate Finance:Comptrollers:Corporate Finance\_NOMIGRATE | Enterprise:Department Administration:Finance:z-Archives:**(Corporate Finance) Comptrollers Corporate Finance** |
| Corporate Finance:JDE Financial PDFs and Reports\_TO BE ARCHIVED | Enterprise:Department Administration:Finance:z-Archives:**(Corporate Finance) JDE Financial PDFs** |
| Upstream Operations:Finance:Central Services\_NOMIGRATE:Financial Images | Enterprise:Department Administration:Finance:z-Archives:**(Upstream Operations) Central Services Financial Images** |
| Upstream Operations:Finance:Central Services\_NOMIGRATE:Royal Income Accounting | Enterprise:Department Administration:Finance:z-Archives:**(Upstream Operations) Central Services Royal Income Accounting** |
| Upstream Operations:Upstream Business Services:Supply Management:JDE Attachments | Enterprise:Department Administration:Finance:z-Archives:**(Upstream Operations) Supply Mgmt JDE Attachments** |

The contents to be migrated are shown below:



## Migrated Categories

2 categories will be migrated:

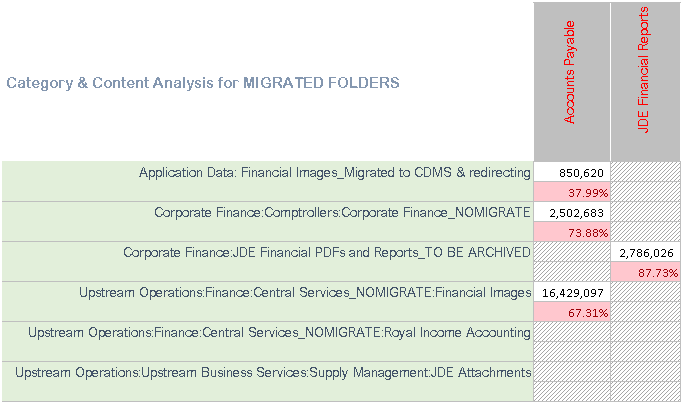
* Accounts Payable
* JDE Financial Reports

Additionally, the Migration System Properties category will be populated with the legacy metadata. Date Modified, Created By, NickName and Migration Date will be populated in the Metadata field.

|  |  |
| --- | --- |
| Accounts Payable |  |
| JDE Financial Reports |  |
| Migration System Properties |  |

The Migration System Properties category will be applied to all objects (container and content types).

The legacy categories will be applied to the migrated folders as follows:



# Migration Process

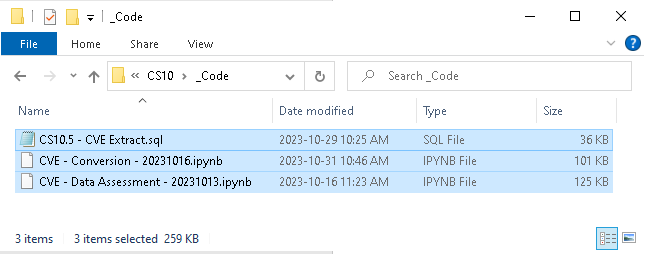
The migration is done using Object Importer for the creation of objects and ingestion of documents into HDMS. The files are copied from the CS 10.5 file share to a share in the wsidm009PD server using bat files, from where object importer uploads them into HDMS.

## Location of Execution and Program files

All code and execution files are stored in this share: **\\wsidm009pd\CS10**

### Python Code and SQL Code

The code to generate the bat and OI files is stored under this folder: **\\wsidm009pd\CS10\\_Code**

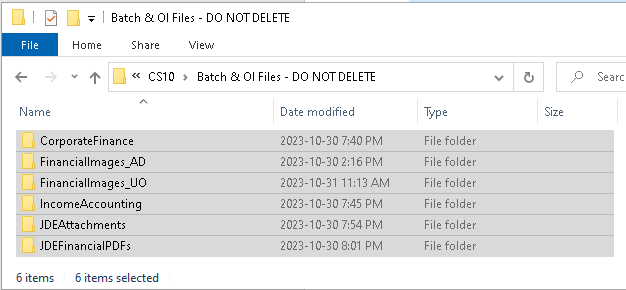


The bat files and Object Importer XML files were generated using a python program that extracted the metadata from the database and created the files. The code was written using jupyter notebook, which allows the execution of python code in a console fashion, similar to SQL developer for Oracle. The jupyter notebook file (.ipynb) stores the code as well as the results of the execution and can be opened in any text editor if needed, but won’t be executable unless opened in jupyter notebook, or the code portions are copied and saved separately and executed in a python console.

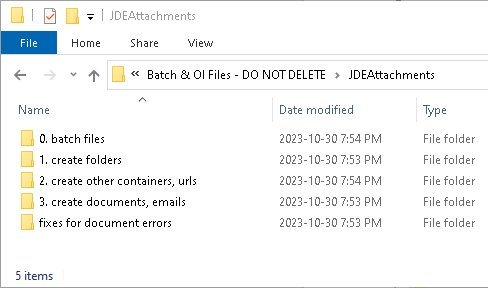
A copy of the SQL statements used for the data assessment and by the python code to generate the PI XMLs are are saved in a SQL file and can be executed directly on the database.

### Batch Files and Object Importer files

Underneath this folder in the share: **\\wsidm009pd\CS10\Batch & OI Files - DO NOT DELETE,** there are 6 folders, each one corresponding to each of the 6 CS10.5 root locations that will be migrated:



Underneath each one of these folders there is a set of subfolders containing the bat files and object importer files needed to perform the migration:

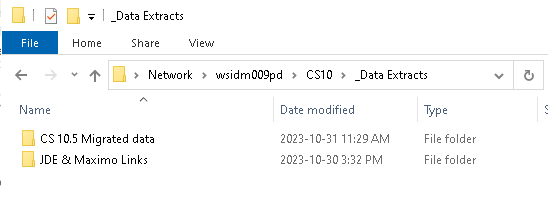


0.batch files, contains the bat files used to copy documents and emails from the CS 10.5 EFS to the wsidm009pd/CS10 files share where OI will pick them up to upload them to HDMS. All other folders contain object importer files, and they must be executed in the order of the folders, this is, folders are created first, then other types of containers and at the end document and emails.

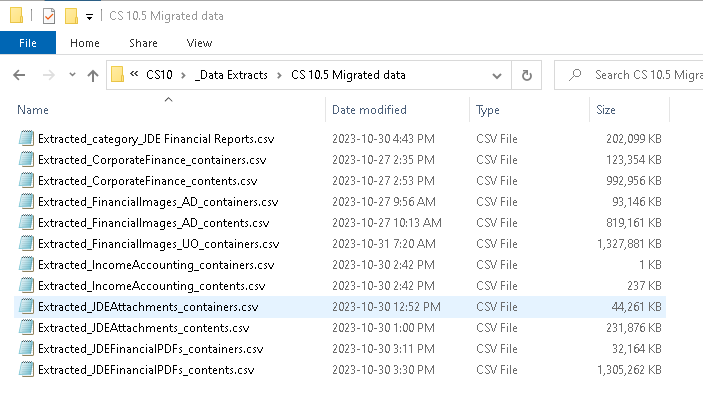
Folders and other containers, do not require from the physical files copied by the batch files, therefore these OI files can be executed before or concurrently with the batch files. Document and Emails can only been imported once the corresponding files copy with the bat file has completed.

All files are split in batches of 20,000 items each.

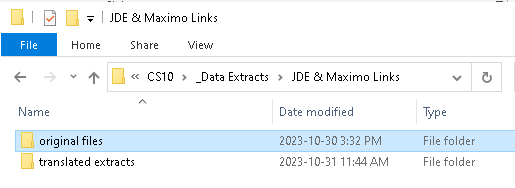
### Extracted Data Files

All data extracted from CS 10.5, as well as the parsed JDE and Maximo links are stored in this folder: **\\wsidm009pd\CS10\\_Data Extracts**

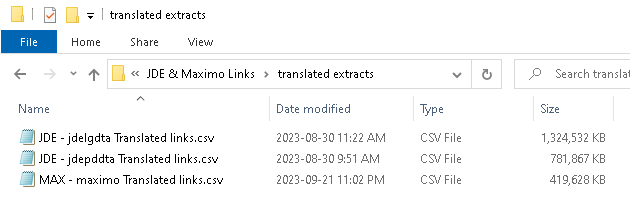
The **CS 10.5 Migrated data** folder contains csv files with all the data extracted from CS 10.5, there are 2 files per each one of the 6 root folders. One file contains all the container objects, the other one all the urls, documents and emails. These files can be used as the data source for the generation of OI files, without requiring access to the CS 10.5 database , since they are flat file dumps of the database data used to generate the OI files with the python program.



The **JDE & Maximo Links** folder contains 2 folders:



The firs folder contains the original files provided by Cenovus with the extract of links in JDE and Maximo. The second folder has 3 CSV files including the original link (column2) and the dataid (column4) parsed from the url. These files will be used to generate the mapping between the old link and the new link that will be provided to Cenovus to replace the corresponding links in JDE and Maximo.



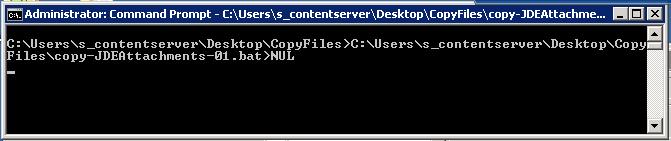
Please note that not all the urls extracted from JDE or Maximo linked to a Content Server document. These urls could link to other Cenovus systems as well. The last column in these files (type), has a value indicates whether the url was identified as a potential match for a CS object. The first column in these files can be disregard as it corresponds to an internal identifier generated by the python program used to parse and translate the urls into dataids.

## Execution of bat files: Copy of documents and emails

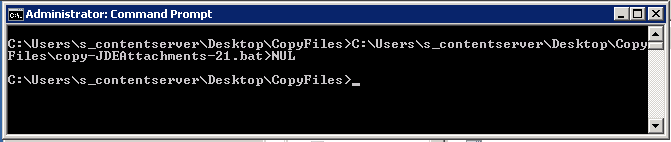
|  |
| --- |
| **Step Notes** |
| * Only the last version of documents and emails will be migrated. * CS 10.5 does not have archive server implemented; therefore, documents can be accessed directly from the files share where they are stored. * Bat files with XCopy commands are used to copy the documents. The command renames the original file name (dataid.dat) to the following naming convention to avoid overwriting documents with the same file names: (dataid-versionnum) – filename.extension * Each bat file contains 20,000 documents or emails * A log file with the same name as the bat file and .log extension is created in the same folder where the command prompt is located at the time of the execution of the bat file. * The log file has a start time at the beginning of the file and records the end time at the end. * Before each XCopy command is executed, the bat file will write the dataid – filename of the document to the log file * If the document is copied successfully the message **1 File(s) copied** will be written to the log file. * If the document is not copied the message **does not exist** with be written to the log file. * Several command prompts may be opened at the same time to execute multiple bat files concurrently. Performance testing was done with 10 separate command windows executing 10 batch files concurrently without issues. |

### Steps and Troubleshooting

* Remote into the **CGYAP57925** server with your 99 account.
* Copy the bat files from the following share **\\wsidm009pd\CS10\Batch & OI Files - DO NOT DELETE\<CS 10 FolderName>\0. batch files** to a local folder in the **CGYAP57925** server. **This is important, since executing the file from the remote share will slow down the execution.**
* Open a cmd window, drag and drop the name of the **local** bat file to the window. This will copy the name of the file to the command line. At the end of the line type the following: **>NUL** and hit enter. This will prevent the bat file output from being written to the console which will speed up the execution.



* Once the file is done executing, you will see the command window display again the prompt for a new command.



* Navigate to the share where the documents were copied to and verify that there are 20,000 files in each folder
* If there are less, open the log file and search for “not exist”, to identify the files that were not copied.
* With the dataid in the log file, query the database as follows: SELECT \* FROM ProviderData WHERE ProviderId IN (SELECT ProviderID FROM DVersData WHERE DocId = **<dataid from log file>**)
* Inspect the **ProviderData** column, if you see something like this: *A<1,?,'providerInfo'=***'ZeroByte Provider Info',***'Providername'='ZeroByte','storageProviderName'='ZeroByte','subProviderName'='ZeroByte'>*, it is a valid error since there isn’t a file in the CS 10.5 file share. Otherwise, look for the file in the path described in the ‘providerinfo’ tag.

## Execution of Object Importer Files

|  |
| --- |
| **Step Notes** |
| * Only one object importer process can run per server. Object importer is configured to scan the upload folder every 5 minutes and once it pick a file, it will not pick the next one until if is finished processing the first one. * Multiple server can execute Object Importer simultaneously. Performance testing was done with up to 3 servers without observing database performance degradation, but adding up to 5 servers may be possible depending on testing. * 3 files are created in the log file during the execution: .log, \_uncreated and \_unprocessed. The \_unprocessed file will be deleted once the execution is complete. The \_log file will report the start and end time of the execution, the number of items processed, ingested and omitted as well as all errors from the execution. The \_uncreated file will have a copy of the XML to create the object that failed. * The execution of the Object Import files that create folders must be the first step. * Folders are created explicitly to apply the migration and legacy category and preserve the CS 10.5 dataid that will be needed to generate JDE and Maximo links after the migration. * Folders are created separately so issues may be troubleshot before the import of all its contents, to avoid cascading issues or the creation of folder without categories. * The creation of folders, or other containers (OI files in folders 1. And 2.) can be done in parallel with the execution of batch files that copy the documents. * Multiple create\_folder files can be run concurrently in separate servers. There is however a small risk of having children folders being created before the parent folder, if the parent folder and the children folders are not in the same OI file. This error will not prevent the creation of the folders, since OI is configured to create the folder path. However, the parent folder will be created without the migration category. If this issue occurs, the OI log may report the parent folder was not created because a folder with the same name already exists. If this is the case, navigate to the folder created by OI automatically and apply the missing category data from the \_uncreated file. * The execution of OI files in the **3. Documents and emails** folder must be done only once the corresponding bat file has completed and all files in the 1., and 2. folders are executed and errors re-processed. |

### Steps and Troubleshooting

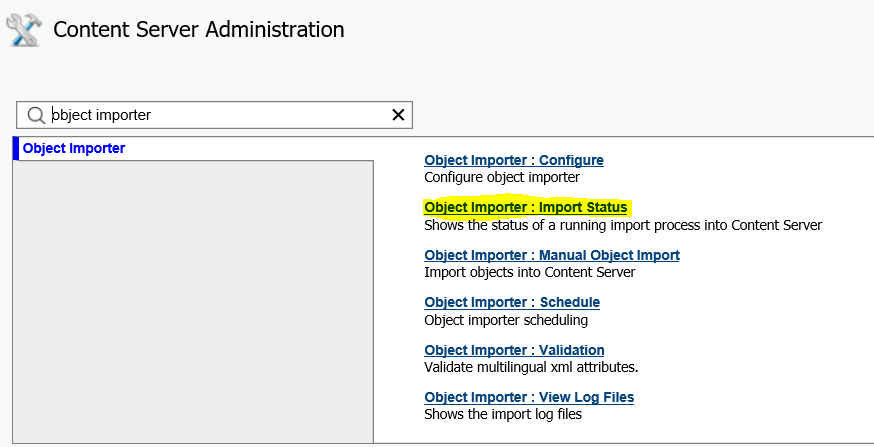
* Copy the OI XML files from the folder structure under this location **\\wsidm009pd\cs10\Batch & OI Files - DO NOT DELETE**:, into the Object Importer queue of any of the following servers:

[\\wsidm009pd\ObjectImporter\Upload\ICF](file:///\\wsidm009pd\ObjectImporter\Upload\ICF)

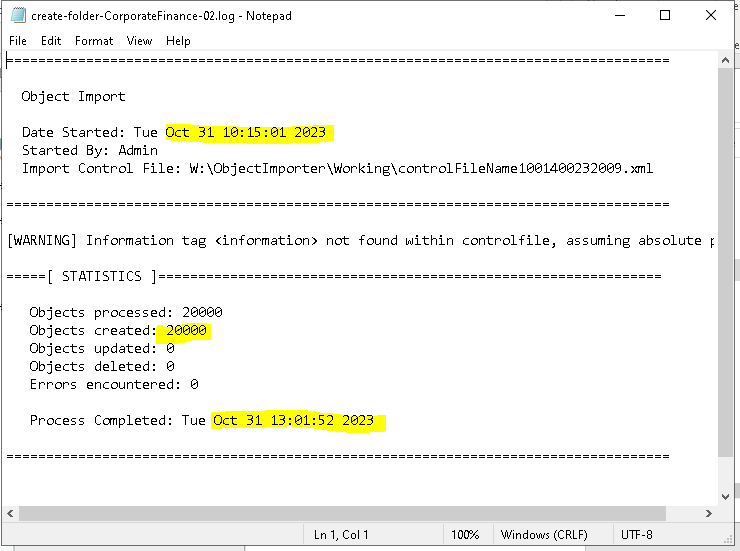
[\\wsidm005pd\ObjectImporter\Upload\ICF](file:///\\wsidm005pd\ObjectImporter\Upload\ICF)

[\\wsidm006pd\ObjectImporter\Upload\ICF](file:///\\wsidm006pd\ObjectImporter\Upload\ICF)

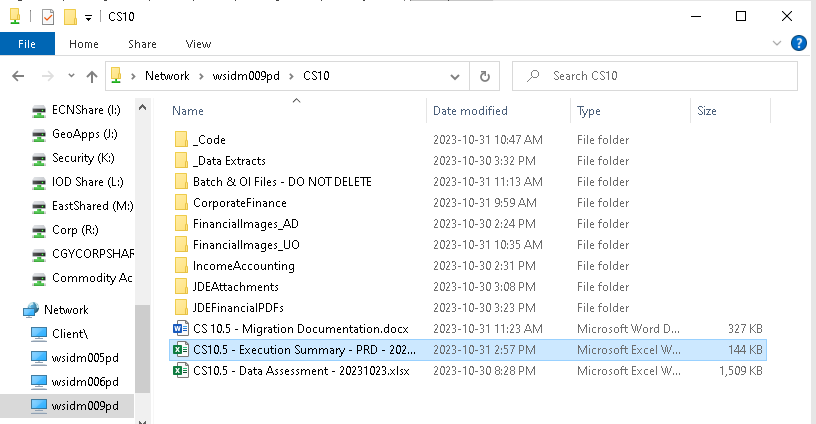
* The files should be picked up within the next 5 minutes. When the file is grabbed, it will deleted from the folders above
* Navigate to the corresponding log folders to confirm the log, uncreated and unprocessed files have been created:
  + [\\wsidm009pd\ObjectImporter\logs](file:///\\wsidm009pd\ObjectImporter\logs)
  + [\\wsidm006pd\ObjectImporter\Logs](file:///\\wsidm006pd\ObjectImporter\Logs)
  + [\\wsidm005pd\ObjectImporter\Logs](file:///\\wsidm005pd\ObjectImporter\Logs)
* The execution can also be monitored from the object import status window: [Import Status (huskyenergy.com)](https://hdms.huskyenergy.com/OTCS/llisapi.dll?func=oi.ImportStatus)



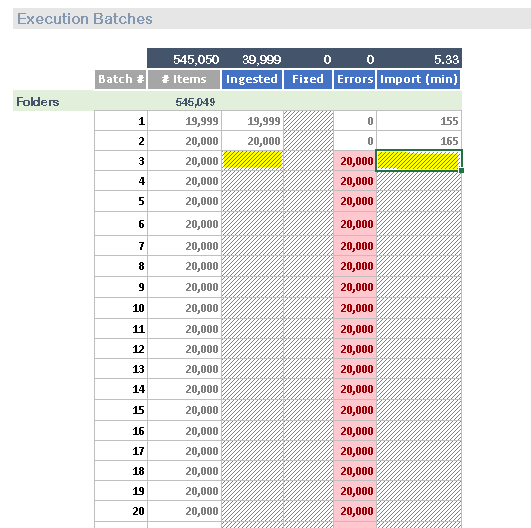
* Once the execution is complete, open the log file to collect the execution info: start and end time to calculate the execution duration, and # of items imported:



* Record the information in the Execution Summary spreadsheet located here: **\\wsidm009pd\CS10**



* Pick the appropriate tab corresponding to the root source folder that is being processed and populate columns P and S



* If there are errors in the log file, copy the \_uncreated.xml file from the log folder and rename if to filename-b.xml.
* Edit the file in notepad++ and add the following 2 lines at the beginning

<?xml version="1.0" ?>

<import>

* Add the following like at the end:

</import>

* Fix the errors reported in the log file, for example, a common error during document creation is that the name of the filein the XML does not match the name of the file copied by the bat. This error is caused by non-ascii characters in the file name that translated to a different character by the OS when the file is copied. To fix this issue do the following:
  + Open the \_uncreated file and look at the value of the <file> tag.
  + Copy the first part of the file name that corresponds to (dataid-version)
  + Navigate to the folder indicted in this tag and search for the (dataid-version).
  + If a file is found, compare the file name and correct it in the OS so it matches the OI file
  + If a file is not found, go back to the corresponding bat file, search for the same value and look for the file in the source folder to confirm whether it exists or if it is a zero byte file.
* Once all errors in the \_uncreated file are fixed, drop this file into the Object Importer queue so it gets reprocessed and confirm there are no errors in the log file as described above.

## Post Migration Tasks

Once the migration is completed, the following tasks and outputs must be created:

### Mapping document between old and new JDE and Maximo links

* A csv file containing the old JDE url pointing to CS 10.5 and the new url pointing to the document in HDMS must be provided to Cenovus so they may replace the old links with the new ones in JDE and Maximo

1. Using a SQL script get an extract of the migrated document **dataid** and the value stored in the **SystemId** attribute of the **Migration System Properties** category
2. Use the files in this folder: **\\wsidm009pd\cs10\\_Data Extracts\JDE & Maximo Links\translated extracts,** to map the **cve\_dataid** column to the value extracted from the migration category in the previous step
3. Use the HDMS dataid extracted on step (a), to generate the new url of the document.
4. Produce an extract with the url column from the files in step(b) and the url generated from the previous step.

### Clean up Z table from Data Id’s moved to HDMS

* Some of the documents migrated to HDMS had also been migrated to CDMS, therefore, the Z table in CS 10.5 must be cleaned up to avoid the accidental redirection of CS 10.5 links to CDMS

1. Use the first column (dataid) in the CSV files stored in this folder: **\\wsidm009pd\cs10\\_Data Extracts\CS 10.5 Migrated data**, to get the CS 10.5 dataid’s of the documents that need to be removed from the Z table.

### Clean up of CDMS

* The documents that were migrated to CDMS and are now also in HDMS, should be deleted from CDMS:

1. Use the first column (dataid) in the CSV files stored in this folder: **\\wsidm009pd\cs10\\_Data Extracts\CS 10.5 Migrated data**, to identify the CS 10.5 .
2. Match these dataid’s to the values in the **LegacyID** attribute from the **Migration Category** in CDMS

