DuraCloud Sync Tool

- Introduction
- Download and Install
- Command Line Interface
- Metadata
- Troubleshooting

Introduction

The Sync Tool is a utility which was created in order to provide a simple way to move files from a local file system to DuraCloud. To get started with the Sync Tool, read on, or watch videos of the process here.

Download and Install

Download installers for Mac OSX, Windows, or Linux from the DuraCloud Downloads page.
The Sync Tool requires that Java version 8 or above be installed on your system in order to run. The installers for your operating system will check to make sure that you have the correct version of Java and will prompt you to download and install Java if necessary. If you would like to update Java directly, it can be downloaded from here.

The Sync Tool defaults to a graphical user interface.

- The Sync Tool provides a web-browser-based application user interface which begins with a configuration wizard, then provides a dashboard display showing the current status of the sync process. This interface is the default and is started by running selecting any of the shortcuts created by the installer.
- Once running, this interface will be continually available at this address: http://localhost:8888/sync.

Operational Notes

- Closing the browser window will not stop the Sync Tool. It will continue to run and transfer files.
- Getting back to the Sync Tool
  - Once the Sync Tool is started, it will continue to run in the background, even if you close your browser. You can get back to the UI by pointing your browser to: http://localhost:8888/sync (hint: bookmark this page).
  - Selecting any of the shortcuts created by the installer will bring up the Sync Tool UI.
- Stopping the Sync Tool
  - Within the UI there are options to stop and start the sync. This will allow you to stop syncing for a time, and start it up again later.
- Stopping the Sync Tool process
  - If you would like to completely shut down the Sync Tool process, such that the UI is no longer available:
    - On Windows: Look for a DuraCloud Sync icon in the task tray, right click on it, select Exit
    - On Mac: Look for a DuraCloud Sync icon in the menu bar, right click on it, select Exit
    - On Ubuntu: Look for a DuraCloud Sync icon in the task bar, right click on it, select Exit
- Work Directory
  - The work directory is named duracloud-sync-work, and can be found under your home directory (C:\Users\[username] on Windows, /Users/[username] on Mac, /home/[username] on Linux)
  - In the work directory you will find:
    - A configuration file which includes the data you entered when configuring the tool
    - A logs directory with log files containing runtime status information of the Sync Tool. These can be helpful when diagnosing a problem the tool may have had.
- Jump Start
  - The Jump Start option available in the SyncTool is designed to streamline the transfer of new file sets to DuraCloud. This is accomplished by removing the checks that the SyncTool traditionally performs before uploading a file. These checks normally try to determine if a file already exists in DuraCloud. With the Jump Start option enabled, the SyncTool assumes that all files are new and need to be moved to DuraCloud. This is option is ideal for the initial data transfer into DuraCloud, when all selected data needs to be transferred. The Jump Start option should be turned off when running the SyncTool over a data set that is already in DuraCloud (in order to discover and transfer any new files), so that unnecessary content transfers can be avoided.
- Transfer Rate Optimization
  - When performing a transfer of files to DuraCloud, the goal is often to get those files moved as quickly as possible. To assist with this the Sync Tool allows you to adjust the number of simultaneous transfers (a.k.a "threads") on the Configuration tab. One caveat here is that the higher the number of threads, the more system resources will be consumed. Additionally as the number of simultaneous transfers increases, more network bandwidth will be consumed. So what is the best number of threads? It depends on the characteristics of your machine and the network to which it is attached. We've added **a handy new feature** that will automatically determine the optimal number of threads for your system. In the "Transfer Rate" section of the configuration page you'll notice an "Optimize Automatically" button. Click it to automatically discover and set the optimal number of threads.
  - Note that "optimal", in this context, means the number of threads that will allow content to be transferred as quickly as possible.
  - The determination of "optimal" thread count is based on testing actual timed transfers, which is why the test may take a while to run when resources such as upload bandwidth or CPU or memory capacity are constrained. This also means that the optimal thread count given will reflect the capability of the machine while the test is running. If other tasks on the machine are consuming significant system resources, this will affect the results of the test.
  - If the machine being used for the transfer of content is not primarily dedicated to this one task (at least while the SyncTool is running), then you may want to set the thread count lower than the determined "optimal" setting. This will, of course, reduce the transfer rate, but would allow the machine to have capacity for other activities.

    - You can use the SyncOptimize Tool to perform these tests if you would prefer to run them on the command line, have more control over the parameters used, and see more details about the testing process.
- Destination Prefix
    - Using the prefix option, the content IDs that are created for the files being moved to DuraCloud by the SyncTool can be made to begin with a consistent text value. There are several reasons this might be useful, such as to include the name of a top-level directory in the path, or to be able to run the Sync from a new sub-directory, but still maintain the full path included on all existing stored content. Suppos e the path to a local file (found within the watch directory) is "dir1/file.txt" and you would like the resulting content stored in DuraCloud to be 'a/b/c/dir1/file.txt. To achieve that result, the destination prefix of "a/b/c/" would need to be set.

      Adding or changing a prefix for content that has already been transferred to DuraCloud will result in those files being duplicated in DuraCloud storage with the new prefix. Removing the duplicate files can be done by using the "sync deletes" option, but this will cause all content in the destination space which does not include the prefix to be deleted (along with any content that is not found in the local watch directories.) Be cautious when using this feature if you have already uploaded content to your DuraCloud space.
      If you use a prefix to include a file path (such as a top level directory name), remember to include the "/" character at the end of your prefix. For example, using the prefix "dir1/" with file "file.txt", your final content ID will be "dir1/file.txt". If you were to forget the slash, your prefix would be "dir1", which would lead to a content ID of "dir1file.txt", which is likely not what you want.
- Run Modes
    - You may run your synchronization operations in one of two modes: continuous or single pass. In continuous mode, DuraCloudSync will continue indefinitely to watch for additions, updates, and deletions to the file system after adding all the files in your watched directories when the sync starts. In the single pass mode, the application will not continue to watch for changes after making the initial pass of your configured directories and/or files.

Command Line Interface

The Sync Tool provides a command line interface which can be executed directly, used in scripts, or used for scheduling sync activities (such as within a cron job.) The command line interface provides access to all feature of the Sync Tool, some of which are not available (yet) in the graphical interface.

- More information about using the command line interface can be found here.

Metadata

As the Sync Tool transfers files to DuraCloud, it will attempt to capture certain types of metadata about each file, and include that information as part of the content item added to DuraCloud. The list below describes the metadata that is captured automatically. You have the option to add, update, or delete the properties of each file after it has been transferred to DuraCloud.

- Mime Type
    - The content type of the file.
    - As the Sync Tool transfers your files to DuraCloud, it attempts to determine the mime type of each file based on the file's extension. If it cannot determine a mime type for a given file, that file's type is set to "application/octet-stream", which is a generic mime type for binary data. Select the "Edit" button on the DuraCloud web interface to change a file's mime type.
    - If you find that files with certain extensions are not being mapped as you would prefer, you can always change the value on uploaded files from within DuraCloud. If you would like to make sure that files with a given extension are given your preferred mime type during upload, you simply need to update the mapping file. The mapping of file extension to mime type is determined by a file included in your Java installation called content-types.properties. This file is usually located in the "lib" folder under your Java runtime installation directory. After making a copy of the original file as a backup, simply update it following the formatting conventions used throughout the file to include the mappings you prefer, then save the file. After making changes, you will need to re-start the Sync Tool to ensure that the changes are picked up properly.
- Space
    - The space in which a content item is stored. This field cannot be edited.
- Size
    - The size of a content item. This field cannot be edited.
- Modified
    - The date on which the file was added to DuraCloud. This value is updated when a file is added or updated.
- Checksum
    - The MD5 checksum of the file. This field cannot be edited.
- Creator
    - The creator is the DuraCloud user who transferred the file into DuraCloud storage.
- Content file path
    - The full path of the file in its original storage location
- Content file created
    - The date when the file was created, as determined by the originating file system. This information may not be available from all file systems.
- Content file modified
    - The date when the file was last modified, as determined by the originating file system. This information may not be available from all file systems.
- Content file last accessed
    - The date when the file was last accessed, as determined by the originating file system. This information may not be available from all file systems.

Troubleshooting

If you encounter an error when running the Sync Tool, please first consult the list of error messages and suggested fixes below. If the error you are experiencing is not included in the list below, please visit the support system and submit a ticket with a detailed description of the issue you are experiencing (and include screenshots when available).

- Out of Memory Error or Java Heap Space Error