DuraCloud Chunker Tool

Introduction

The Chunker Tool is a utility which was created in order to provide a simple way to copy files from a local file system to DuraCloud in a "one-off" manner. Actually, although the common case is to use this tool to copy one or more files to DuraCloud, it may also be run to copy files to another location on the local file system.

The Sync Tool is a better choice for transferring files to DuraCloud. The Chunker Tool can be build from DuraCloud source, but is no longer provided as a binary distribution.

Operational notes

- If you want to jump directly into using the tool, download it from the link above and run the following command

```
java -jar chunk-{version}-driver.jar
```

    The resulting usage statement (detailed below) should be enough to help you get started.
- The Chunker Tool allows you to copy multiple local files and directories into a single space within DuraCloud. The names of the objects which are added to DuraCloud will contain all of the directory elements in the path starting from the first element below the base directory down to the individual file names.

Prerequisites

As of DuraCloud version 4.0.0, the Chunker Tool requires Java 8 to run. The latest version of Java can be downloaded from here.

Using the Chunker Tool

- To run the Chunker Tool, open a terminal or command prompt and navigate to the directory where the Chunker Tool is located and run the above command.
- The following options are available when running the Chunker Tool

Short Option	Long Option	Arguments	Description
-a	--add	<f t s>	add content from directory:<f> to space or directory:<t> of maximum chunk size:<s>, where the chunk size must have a unit suffix of K,M, or G — If the -c option is provided, the destination space <t> will be interpreted as the name of a space in the DuraCloud account found at the host:port provided in the -c option, otherwise the destination space will be interpreted as a directory on the local file system.
-c	--cloud-store	<host:port>	use cloud store found at <host>:<port> as content destination
-d	--dir-filter	<l>	limit processed directories to those listed in file-list:<l> — If the -d option is not used, all directories under the base source directory provided in the -a option will be included. The file specified by this option is expected to contain a list of directory names each on there own line. The list is converted to an OrFileFilter from Apache Commons IO
-f	--file-filter	<l>	limit processed files to those listed in file-list:<l> – The file specified by this option is expected to contain a list of file names each on there own line. The list is converted to an OrFileFilter from Apache Commons IO
-g	--generate	<outFile numBytes>	generate test data to <outFile> of <size> bytes — This option does not copy any files, it only generates test data files of the size specified in the give argument.
-i	--ignore-large-files	no args	if this option is set, files over the chunk size specified in the 'add' option will be ignored.
-p	--password	<password>	password of duracloud instance
-u	--username	<username>	username of duracloud instance
-x	--exclude-chunk-md5s	no args	if this option is set, chunk MD5s will NOT be preserved in the manifest — It is expected that this option is rarely used, but in certain situations where the MD5s of the segments of a file that needed to be chunked because the parent file was larger than the limit set in the -a option, not generating these MD5s improves performance.

Creating your own Chunks

If you are interested in creating chunked files in DuraCloud using your own tools, you may do so by adhering to the XML schema used by DuraCloud to create chunks.

Download the Chunker XSD from the Downloads page