DuraCloud-facing Clients Thread

DuraCloud-facing Clients

This thread includes any client expected to work through the DuraCloud-provided interfaces noting some clients may use more than one interface.

Researcher System Monitors

Third Party tool connections

It is known that some researchers are currently making use of commercial tools that provide the ability to keep data between multiple systems backed up and synchronized. Two of the major players in this space are Box and DropBox. While the DfR system may provide some of these same features, it seems most appropriate to continue to allow researchers to use tools which benefit their workflow and which they are comfortable using. DfR should be able to connect to these existing systems to collect the data which is being created by the research team.

Below is the result of a quick review of a few of the available tools.

JungleDisk

JungleDisk provides a deamon monitor to detect changes in folders that you specify. This is the method for transferring content to the backing storage location (Amazon S3 in this case.) The local client tool also provides the ability to mount your backup storage area as a local disk. I enabled this mount point, and pointed the DuraCloud sync tool at a directory of files.

The sync tool ran well, though the logs indicated frequent errors in accessing files via jungledisk, these issues were resolved through multiple attempts. Unfortunately, the JungleDisk software became overwhelmed and crashed after the sync tool had transferred about 1600 files to DuraCloud. Even though the files were no longer available, the sync tool continued to run and act as if it were transferring files successfully, though no additional files were added to DuraCloud. This behavior was captured in DURACLOUD-609.

DropBox

DropBox requires that you create a "dropbox folder" on your local file system, then place files into it that you would like DropBox to replicate. This dropbox folder is simply a directory on your local file system which the DropBox deamon monitors. Thus, it is not surprising that the sync tool ran flawlessly in moving the files from DropBox into DuraCloud (a little over 2000 files).

DropBox readme

Box

Box provides a free level account which allows for the storage of content through their website. Of more interest is the Box Sync utility available only to Business and Enterprise accounts. Box Sync allows you to choose a file system directory under which it creates a folder named "My Box Files". When files or folders are created under this new directory, they are synced to the Box account. Folders can also be created in the Box account online, and files added via their web interface. Folders created on the website can be synced to the local directory, but they do not have to be. This general setup, which requires local files to be copied or moved into a special folder which is watched by a client, is very similar to the interaction with DropBox. As with DropBox, since the files actually reside on the local file system, the DuraCloud sync tool has no problem moving these files to DuraCloud.

Box Sync readme

Final Comments

Based on these tests, some simple conclusions can be drawn:

1. If the system managing the sync/backup for the researcher does so by actually storing all of the synced files on each synced file system, then the DuraCloud sync tool can access it as it does any other file system directory, and there should be no problems with the transfer.
2. If the system managing the sync/backup provides a mounted "virtual" view into the stored files and retrieves them to the local system on demand, then the stability of the software transfer to DuraCloud will depend on the stability of the third party software, which may not prove to be acceptable.