**archive-big-data-dropbox**

Automated framework to upload large volume (TBs) of data to Dropbox.

Open this folder as workspace in Notepad++ “E:\Software\archive-big-data-dropbox-3.1”

**Introduction**

Dropbox provides [paid services](https://www.dropbox.com/business/pricing) with unlimited data storage space. This can be useful for small to medium sized businesses which generate large amount of data and do not have the resources to build their own storage servers for data management. However, uploading hundreds of terabytes of data to Dropbox servers can be challenging due to three reasons -

1. *If the number of files synchronized to* [*Dropbox exceeds 300,000*](https://help.dropbox.com/accounts-billing/space-storage/file-storage-limit) *the performance of the desktop application may decline.* To solve this problem, it is recommended to make a zip a folder containing a large number of small files into a larger single file.
2. *Dropbox is not good at synchronizing* [*TBs of data*](https://help.dropbox.com/installs-integrations/desktop/unexpected-quit) *in one shot.* To solve this problem, it is recommended to synchronize the data to Dropbox in small batches (~ 500 GB per batch).
3. *Dropbox desktop application cannot synchronize large files.* Contrary to their claim, Dropbox app cannot synchronize single files if they exceed 600 GB (This number could be smaller but I have done test on a 600 GB file.) A way to resolve this is to split the large file into smaller chunks and then upload it to Dropbox.

This package provides a solution to seamlessly compress and upload your data to Dropbox.

**Requirements**

* Windows 10
* [Dropbox application](https://www.dropbox.com/install) installed on your computer and linked to the Business account
* [7zip](https://www.7-zip.org/)
* python >= 3.10.8
  + argparse 1.1
  + numpy 1.23.4
  + pandas 1.5.2
  + tqdm 4.64.1
  + psutil 5.9.4
  + dropbox 11.36.0
  + platform 1.0.8
  + re 2.2.1
  + json 2.0.9
  + shutils

**E:\Software\archive-big-data-dropbox-3.1\app>pip install -r requirements.txt**

There is a requirements.txt file in the code and above command calls it to install all the above libraries from Argparse to shutils. Above command should be run after going to app folder

**Usage**

The process is divided into two parts - 1. data compression, and 2. data synchronization with Dropbox.

**Data compression**

1. Due to its inability to handle a large number of files, it is recommended to zip a folder containing a large number of small files. For example, the Gatan OneView electron microscope camera records a movie at 100 fps and stores every frame in a directory '001'. In this case, we will compress the files in the folder '001' and make a new large file '001.zip'.
2. Open MobaXterm.
3. PS C:\Users\dbsgbis> E:
4. PS E:\> cd .\Software\archive-big-data-dropbox-3.1\
5. PS E:\Software\archive-big-data-dropbox-3.1>
6. PS E:\Software\archive-big-data-dropbox-3.1> cd .\app\
7. PS E:\Software\archive-big-data-dropbox-3.1\app> python .\dataCompressNoSplit\_app.py -src "Z:\yingyingj\to\_dropbox" -size 50
8. –src is the source folder which you want to compress.
9. ~~Navigate to the directory './app'.~~
10. **~~python ./dataCompress\_app.py -src "source directory to be compressed" -size 50~~**
11. ~~The -size flag is optional and it corresponds to the maximum allowed folder/file size in GB. Default value is 50 GB.~~
12. A process log file is generated in the './logs' directory.

**Uploading to Dropbox**

This can be done using the Dropbox Windows desktop application, or using Dropbox's API to upload directly to Dropbox.

**Using the APP**

1. Navigate to the directory './app'.
2. **python ./dropboxAppUpload\_app.py -src "source directory to be uploaded" -dst "destination directory" -profile "Dropbox profile"**
3. A process log file is generated in the './logs' directory. Around 0.8 - 1 TB of data can be moved to the cloud in a day.

**Using the APIs**

1. TODO

**Additional Notes**

* Make sure you have the permission to read, write, and delete files. If not, contact the system administrator to give you appropriate permission.
* Before starting the data upload, make sure that there is sufficient space available on Dropbox.
* Dropbox app tends to crash frequently if the folder size exceeds 2 TB. In order to avoid this, keep moving the data to 'Online Only' mode once every 24 hours.