

Project Statement:

Client is getting multiple files in a shared folder, client is invested in AWS S3, he wants to stage the files in S3 and wants to automate and organize the file landing process with proper logs in mail.

Requirement 1:

Multiple files with various file types (csv, xls, xlsx, pdf, .doc,.zip) in a commo shared folder, he wats to segregate them properly in a folder so that he can streamline the process of uploads in s3

i.e.

file csv should go in csv folder

file xls and xlsx should go in xlx folder

..

..

File pdf should go in pdf folder

Requirement 2:

Validate the and csv files for loads for number of columns, file format. Xls files should be converted to csv for load

Requirement 3:

Proper logging should be done for all the files,

Logs to cover details as:

- 1) How many files processed
- 2) Start and end time of Python code
- 3) Log at each file when it is processed
- 4) Log at each file if its and excel or csv file, how many number of columns and rows it has
- 5) Where files are getting uploaded in S3 and upload duration.
- 6) If the code fails, log should cover the detail of error in log

Requirement 4:

S3 will have as Archive folder where all raw files need to be uploaded for backup, All other files will go in stagging folder with sub folders as csv, pdf, doc...

Requirement 4:

At the end of the script, send the log files in a mail to python.batch20182022@gmail.com