Discussion module 3

The two different ways to correct data integrity issues are following-

1. **Check with an MD5 Checksum:**

An MD5 Checksum is essentially an algorithm that will return a hexadecimal number for the contents of a file.

* python script is:

import hashlib, sys

files = [sys.argv[1], sys.argv[2]] #these are the arguments we take

def md5(fname):

md5hash = hashlib.md5 ()

with open(fname) as handle: #opening the file one line at a time for memory considerations

for line in handle:

md5hash.update(line.encode('utf-8'))

return (md5hash.hexdigest())

print ('Comparing Files:',files[0],'and',files[1])

if md5(files [0]) == md5(files [1]):

print('Matched')

else:

print ('Not Matched')

1. **Check with the SHA1 algorithm**:

The SHA1 algorithm is hexadecimal algorithm that will convert our file contents into a string.​​​​​​​

* Python script-

import hashlib, sys

files = [sys.argv[1], sys.argv[2]] #these are the arguments we take

def sha1(fname):

sha1hash = hashlib. sha1()

with open(fname) as handle: #opening the file one line at a time for memory considerations

for line in handle:

sha1hash.update(line.encode('utf-8'))

return (sha1hash.hexdigest())

print ('Comparing Files:',files[0],'and',files[1])

if sha1(files [0]) == sha1(files [1]):

print('Matched')

else:

print ('Not Matched')

Difference between MD5 and SHA1-

