File Reader Design

Written by Sivasubramani T

Q: ***Designing generic File based reader which supports reading data and handles encoding, decryption & decompression based on user configuration.***

Take file properties as input

CharSet :

Charset with which the file data (binary) is encoded with. For example ISO-8859-1 is used to convert ISO Latin Alphabet string to binary for writing into file.

In case if this property is not specified then consider the bytes (Update: File is encoded with UTF-8) are encoded with UTF-8.

Encryption :

Encryption algorithm used for encrypting the data before writing to file. For example AES. If not specified, then there is no encryption and the data persisted as plain text.

Compression:

Compression logic used to compress the data before writing to file. Possible values are gzip, bzip, bz etc. In case not specified, data is uncompressed.

Design classes with respective methods to read files and perform the operation specified and return the plain text to the caller.

File Reader Design

Written by Sivasubramani T

***Steps & Design Plans*:**

1. **User Requirements**
2. **Library Requirements**
3. **Design**
4. **Code Flow**
5. **Logging**
6. **Unit Testing**
7. **Execution**
8. **Deployment**

1. Usage and User Requirements:

- File ***Encryption*** (Additional case implemented by myself)
- File ***Decryption***
- File ***Compression*** by ***GZIPZ/GZ*** format
- File ***Decompression*** by ***GZIP/GZ*** format (Additional case implemented by myself)
- File ***Compression*** by ***BZIP2/BZ*** format
- File ***Decompression*** by ***BZIP2/BZ*** format (Additional case implemented by myself)
- Parse as ***CSV record***

Cryption List:

ENC_AES128

DEC_AES128

Compression List:

COM_GZIP

COM_GZ

COM_BZIP2

COM_BZ2

DCOM_GZ

DCOM_GZIP

DCOM_BZIP2

DCOM_BZ2

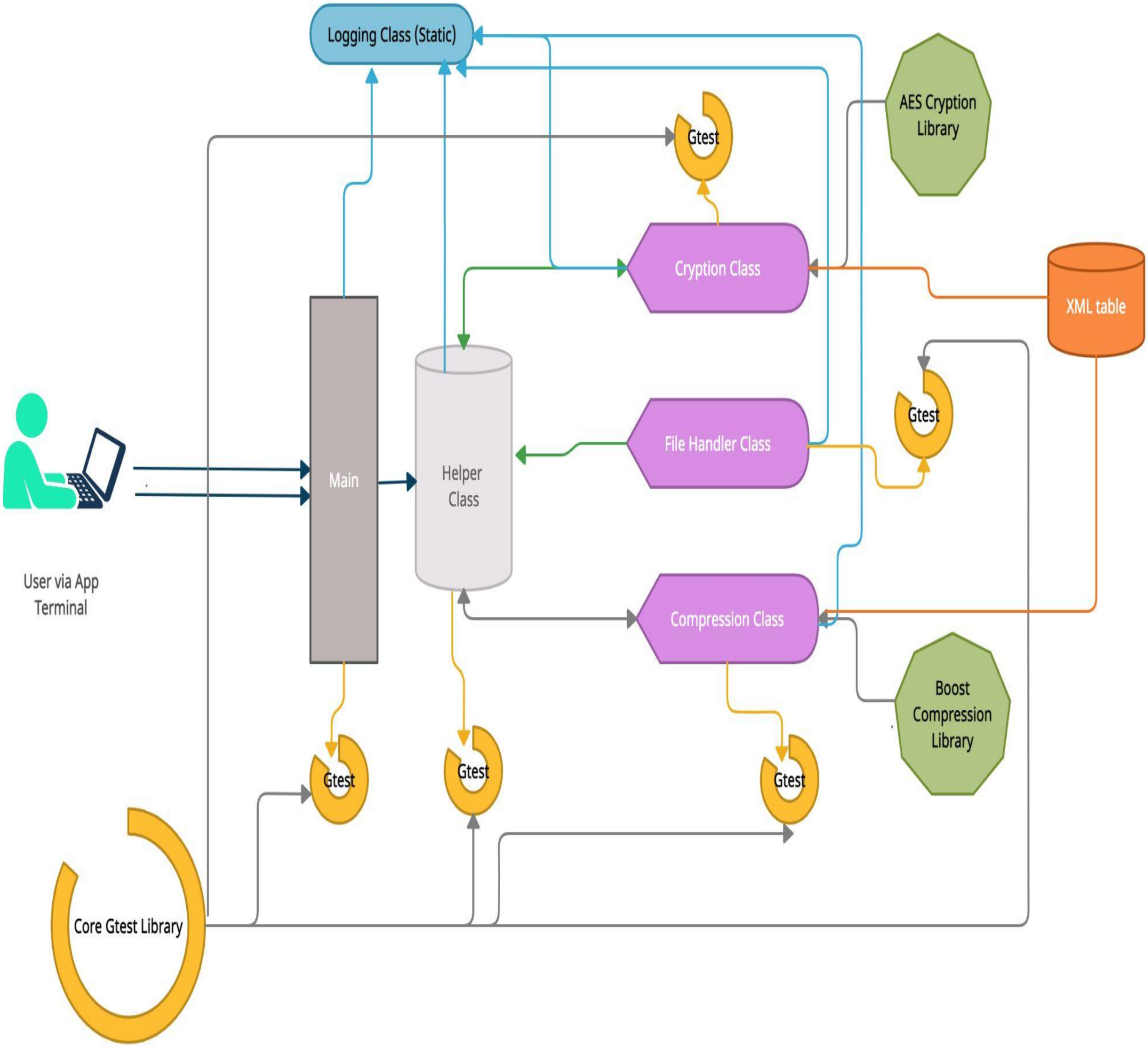
Above list of code only valids in this Application, will reject other codes.

2.Library Requirements:

I am using C++ code for this design, For Gzip/Bzip2 Compressions using Boost Library and for AES Cryption using AES library implementation in my code. You can see below advantages for using these libraries than other libraries.

- ***Boost Iostreams Library -*** *Speed of compressing is good (inbuilt C)*
- ***AES 128 Cryption Library -*** *mandatory library for cryptio in C++*
- ***Google test Library (Unit Testing) -*** *Gtest will be good for new people to understand and easy to learn and deploy.*

3. Design:


Logging Class (Static)
AES Cryption Library
Gtest
Cryption Class
XML table
Main
Helper Class
File Handler Class
Gtest
User via App Terminal
Compression Class
Boost Compression Library
Gtest
Gtest
Gtest
Core Gtest Library

XML Table used reason: Compression name and cryption name list maintained in XML which is better than list,vector,array *because compilation not need for* ***XML and we can change the xml value at run time to make program dynamic*** *and* ***we don’t need to restart application to add/delete value in application***.

Main and Helper Class: Main is just an interface between command and application. Helper class helps the main function to perform action. Helper class holds all activities of command from the user and manages compression class, Cryption class and File handling class. Gtest also done for each test case.
File handler Class: It holds all activities of File copying, writing parsing and reading . Gtest also done for each test case.

Cryption Class: It holds all activities of Encryption/Decryption and is connected to logging and fetching data from xml while instance creation. Gtest also done for each test case .

Compression Class: It holds all activities of Compression/Decompression and is connected to logging and fetching data from xml while instance creation. Gtest also done for each test case.

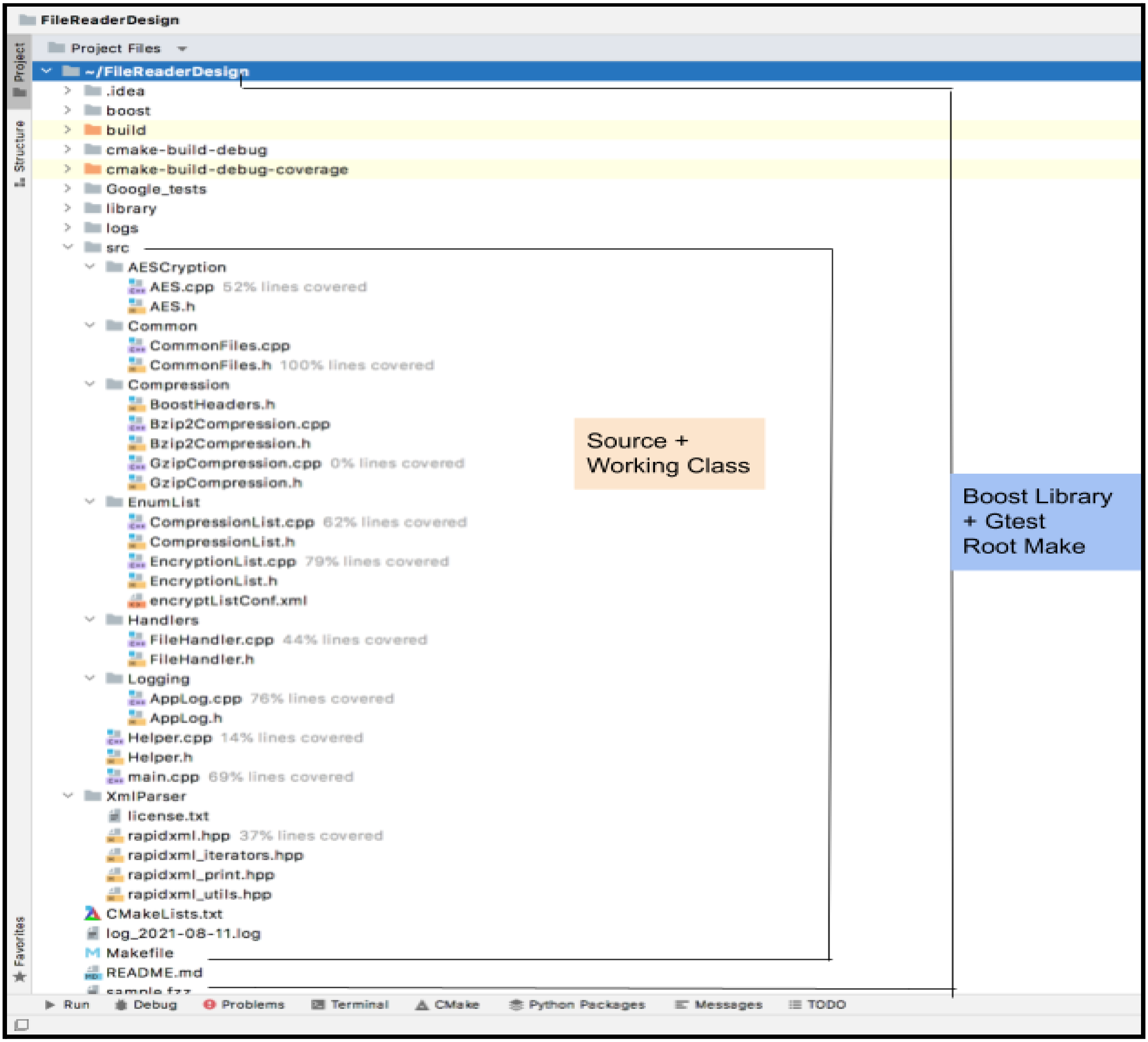
Logging: This class methods are static and invoked in all modules. Please see a separate topic below for logging.

Gtest: Gtest (unit test case) done for all modules.

File Reader Design

Written by Sivasubramani T

4. Code Flow:


FileReaderDesign
Project Files
Project
Structure
Favorites
Source +
Working Class
Boost Library
+ Gtest
Root Make
Run
Debug
Problems
Terminal
CMake
Python Packages
Messages
TODO

```
~/FileReaderDesign
  > .idea
  > boost
  > build
  > cmake-build-debug
  > cmake-build-debug-coverage
  > Google_tests
  > library
  > logs
  v src
      v AESCryption
          AES.cpp 52% lines covered
          AES.h
      v Common
          CommonFiles.cpp
          CommonFiles.h 100% lines covered
      v Compression
          BoostHeaders.h
          Bzip2Compression.cpp
          Bzip2Compression.h
          GzipCompression.cpp 0% lines covered
          GzipCompression.h
      v EnumList
          CompressionList.cpp 62% lines covered
          CompressionList.h
          EncryptionList.cpp 79% lines covered
          EncryptionList.h
          encryptListConf.xml
      v Handlers
          FileHandler.cpp 44% lines covered
          FileHandler.h
      v Logging
          AppLog.cpp 76% lines covered
          AppLog.h
        Helper.cpp 14% lines covered
        Helper.h
        main.cpp 69% lines covered
  v XmlParser
        license.txt
        rapidxml.hpp 37% lines covered
        rapidxml_iterators.hpp
        rapidxml_print.hpp
        rapidxml_utils.hpp
    CMakeLists.txt
    log_2021-08-11.log
    Makefile
    README.md
    sample.fzz
```

File Reader Design

Written by Sivasubramani T

5.Logging:

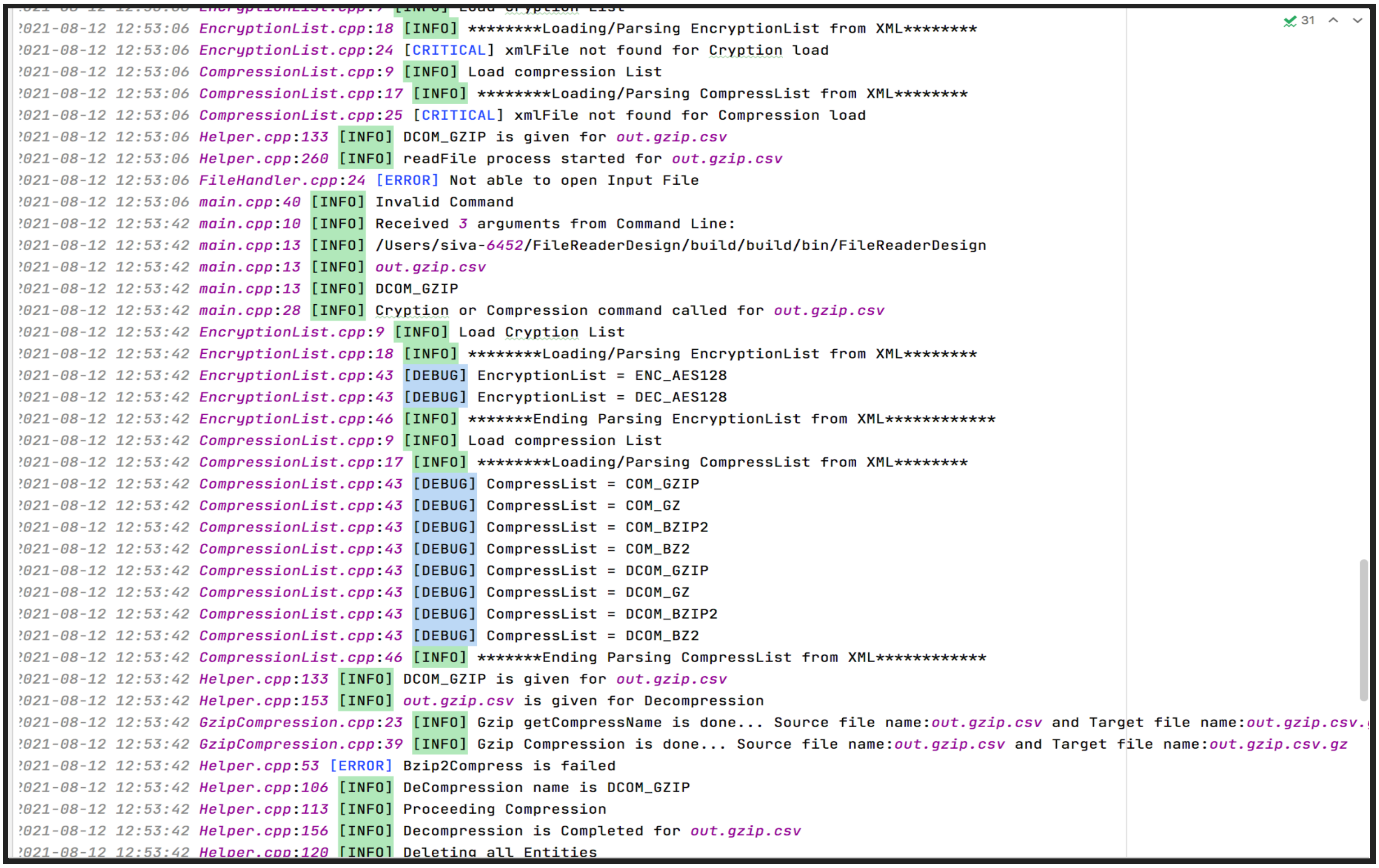
Log is very important for any kind of system design to detect the issues, monitor etc so I deployed monitoring for this design.

Used Static member in class because log is common for all types of usage.

Implemented Log level of 5 types with time ,date and file name :

INFO,WARING,CRITICAL,ERROR,DEBUG

Example:



```
2021-08-12 12:53:06 EncryptionList.cpp:9 [INFO] Load Cryption List
2021-08-12 12:53:06 EncryptionList.cpp:18 [INFO] ********Loading/Parsing EncryptionList from XML********
2021-08-12 12:53:06 EncryptionList.cpp:24 [CRITICAL] xmlFile not found for Cryption load
2021-08-12 12:53:06 CompressionList.cpp:9 [INFO] Load compression List
2021-08-12 12:53:06 CompressionList.cpp:17 [INFO] ********Loading/Parsing CompressList from XML********
2021-08-12 12:53:06 CompressionList.cpp:25 [CRITICAL] xmlFile not found for Compression load
2021-08-12 12:53:06 Helper.cpp:133 [INFO] DCOM_GZIP is given for out.gzip.csv
2021-08-12 12:53:06 Helper.cpp:260 [INFO] readFile process started for out.gzip.csv
2021-08-12 12:53:06 FileHandler.cpp:24 [ERROR] Not able to open Input File
2021-08-12 12:53:06 main.cpp:40 [INFO] Invalid Command
2021-08-12 12:53:42 main.cpp:10 [INFO] Received 3 arguments from Command Line:
2021-08-12 12:53:42 main.cpp:13 [INFO] /Users/siva-6452/FileReaderDesign/build/build/bin/FileReaderDesign
2021-08-12 12:53:42 main.cpp:13 [INFO] out.gzip.csv
2021-08-12 12:53:42 main.cpp:13 [INFO] DCOM_GZIP
2021-08-12 12:53:42 main.cpp:28 [INFO] Cryption or Compression command called for out.gzip.csv
2021-08-12 12:53:42 EncryptionList.cpp:9 [INFO] Load Cryption List
2021-08-12 12:53:42 EncryptionList.cpp:18 [INFO] ********Loading/Parsing EncryptionList from XML********
2021-08-12 12:53:42 EncryptionList.cpp:43 [DEBUG] EncryptionList = ENC_AES128
2021-08-12 12:53:42 EncryptionList.cpp:43 [DEBUG] EncryptionList = DEC_AES128
2021-08-12 12:53:42 EncryptionList.cpp:46 [INFO] *******Ending Parsing EncryptionList from XML************
2021-08-12 12:53:42 CompressionList.cpp:9 [INFO] Load compression List
2021-08-12 12:53:42 CompressionList.cpp:17 [INFO] ********Loading/Parsing CompressList from XML********
2021-08-12 12:53:42 CompressionList.cpp:43 [DEBUG] CompressList = COM_GZIP
2021-08-12 12:53:42 CompressionList.cpp:43 [DEBUG] CompressList = COM_GZ
2021-08-12 12:53:42 CompressionList.cpp:43 [DEBUG] CompressList = COM_BZIP2
2021-08-12 12:53:42 CompressionList.cpp:43 [DEBUG] CompressList = COM_BZ2
2021-08-12 12:53:42 CompressionList.cpp:43 [DEBUG] CompressList = DCOM_GZIP
2021-08-12 12:53:42 CompressionList.cpp:43 [DEBUG] CompressList = DCOM_GZ
2021-08-12 12:53:42 CompressionList.cpp:43 [DEBUG] CompressList = DCOM_BZIP2
2021-08-12 12:53:42 CompressionList.cpp:43 [DEBUG] CompressList = DCOM_BZ2
2021-08-12 12:53:42 CompressionList.cpp:46 [INFO] *******Ending Parsing CompressList from XML************
2021-08-12 12:53:42 Helper.cpp:133 [INFO] DCOM_GZIP is given for out.gzip.csv
2021-08-12 12:53:42 Helper.cpp:153 [INFO] out.gzip.csv is given for Decompression
2021-08-12 12:53:42 GzipCompression.cpp:23 [INFO] Gzip getCompressName is done... Source file name:out.gzip.csv and Target file name:out.gzip.csv.
2021-08-12 12:53:42 GzipCompression.cpp:39 [INFO] Gzip Compression is done... Source file name:out.gzip.csv and Target file name:out.gzip.csv.gz
2021-08-12 12:53:42 Helper.cpp:53 [ERROR] Bzip2Compress is failed
2021-08-12 12:53:42 Helper.cpp:106 [INFO] DeCompression name is DCOM_GZIP
2021-08-12 12:53:42 Helper.cpp:113 [INFO] Proceeding Compression
2021-08-12 12:53:42 Helper.cpp:156 [INFO] Decompression is Completed for out.gzip.csv
2021-08-12 12:53:42 Helper.cpp:120 [INFO] Deleting all Entities
```

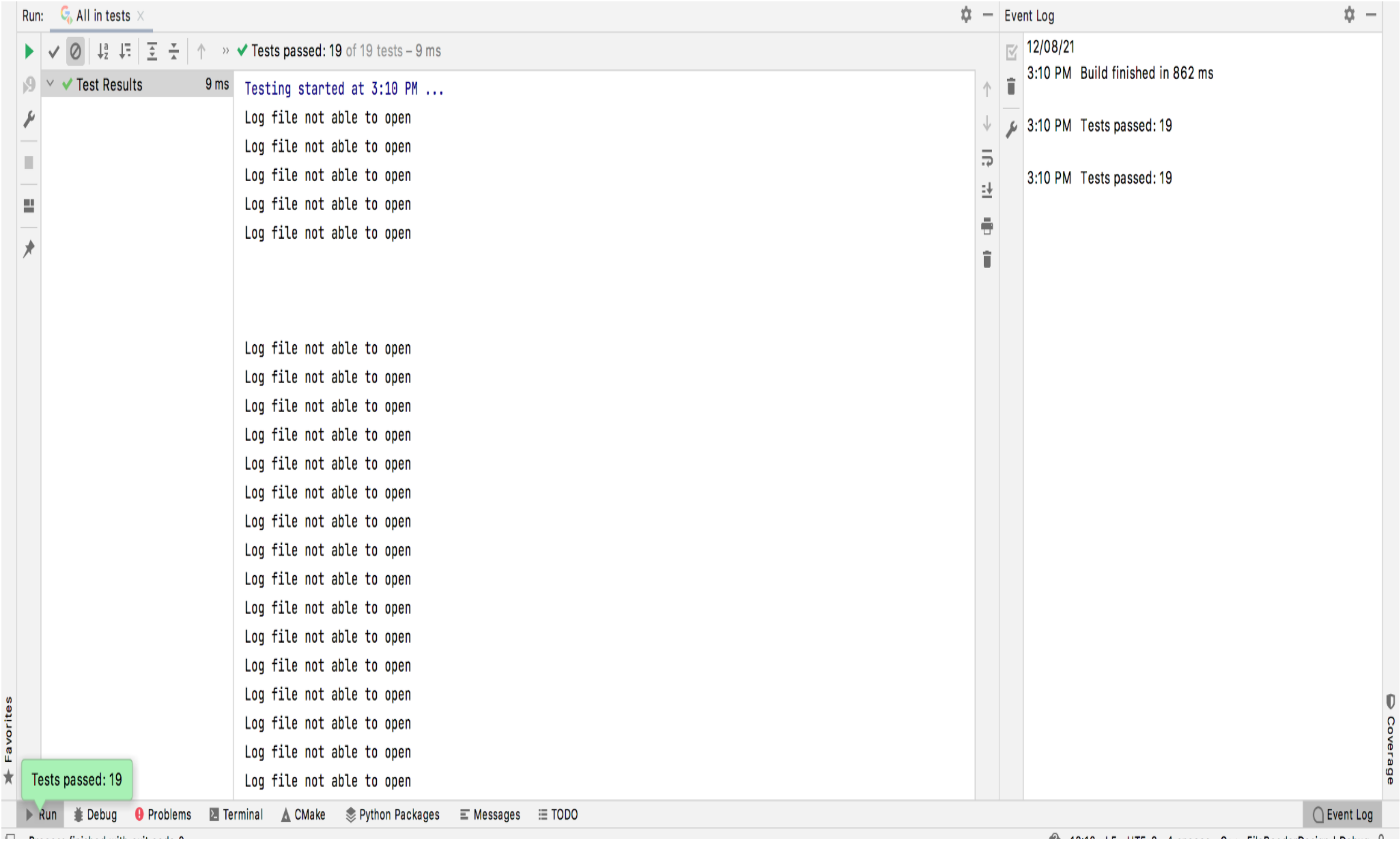
6.Unit Testing:

Unit test is another important feature to detect and debug the issue before going into production. I used Google Test which is easy to use and learn than other libraries. Unit test is linked with root cmake by linking src directory as library.

Each module tested by Gtest:

Total Test function =***19 Passed (Out of 19)***

Total Test cases =***123 Passed (Out of 123)***


Run:
All in tests
Event Log
Favorites
Coverage
Run
Debug
Problems
Terminal
CMake
Python Packages
Messages
TODO
Event Log

Tests passed: 19 of 19 tests – 9 ms

Test Results 9 ms

```
Testing started at 3:10 PM ...
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open



Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
Log file not able to open
```

```
12/08/21
3:10 PM  Build finished in 862 ms

3:10 PM  Tests passed: 19

3:10 PM  Tests passed: 19
```

Tests passed: 19

7. Execution:

For this project, I used :

IDE : ***Clion (Jetbrains)***

Make file: **CMake (clion) and Make (in Mac terminal)**

Compiled Library : ***MacOS (.dylib)***

Info: *Please compile and use* ***.a for Windows*** *and* ***.so for Linux.***

Changing cmake into Make (if needed):

```
install:
all:
    echo Compiling $(CLION_EXE_DIR)/$@ ...
    g++ src/Helper.cpp src/Helper.h -o $(CLION_EXE_DIR)/FileReaderDesignMake
```

9. Deployment:

A.Clion Deployment:

Clone the repo ***https://github.com/Sivasubramani/FileReaderDesign***

Set Directory path

Run Command in Clion

B.Commands to install in mac:

Download build from below path :

https://github.com/Sivasubramani/FileReaderDesign/blob/main/build.zip

And do below:

```
unzip build.zip in
```

File Reader Design

Written by Sivasubramani T

```
cd build
make install
cd build/bin/
```

```
Use case 1:
./FileReaderDesign filename.xxx
ex: ./FileReaderDesign sample.fzz

Use case 2a:
./FileReaderDesign filename.xxx encryption_format
ex: ./FileReaderDesign sample.fzz ENC_AES128

Use case 3:
./FileReaderDesign filename.xxx compression_format
ex1: ./FileReaderDesign sample.fzz COM_GZIP
ex2: ./FileReaderDesign sample.fzz COM_BZIP2

Use case 4:
./FileReaderDesign filename.xxx compression_format encryption_format
ex1: ./FileReaderDesign sample.fzz COM_GZIP ENC_AES128
ex2: ./FileReaderDesign sample.fzz COM_BZIP2 ENC_AES128

Use case 5:
./FileReaderDesign filename.xxx decryption_format
ex: ./FileReaderDesign sample.fzz DEC_AES128

Use case 6:
./FileReaderDesign filename.xxx decompression_format
ex1: ./FileReaderDesign sample.fzz DCOM_GZIP
ex2: ./FileReaderDesign sample.fzz DCOM_BZIP2
```

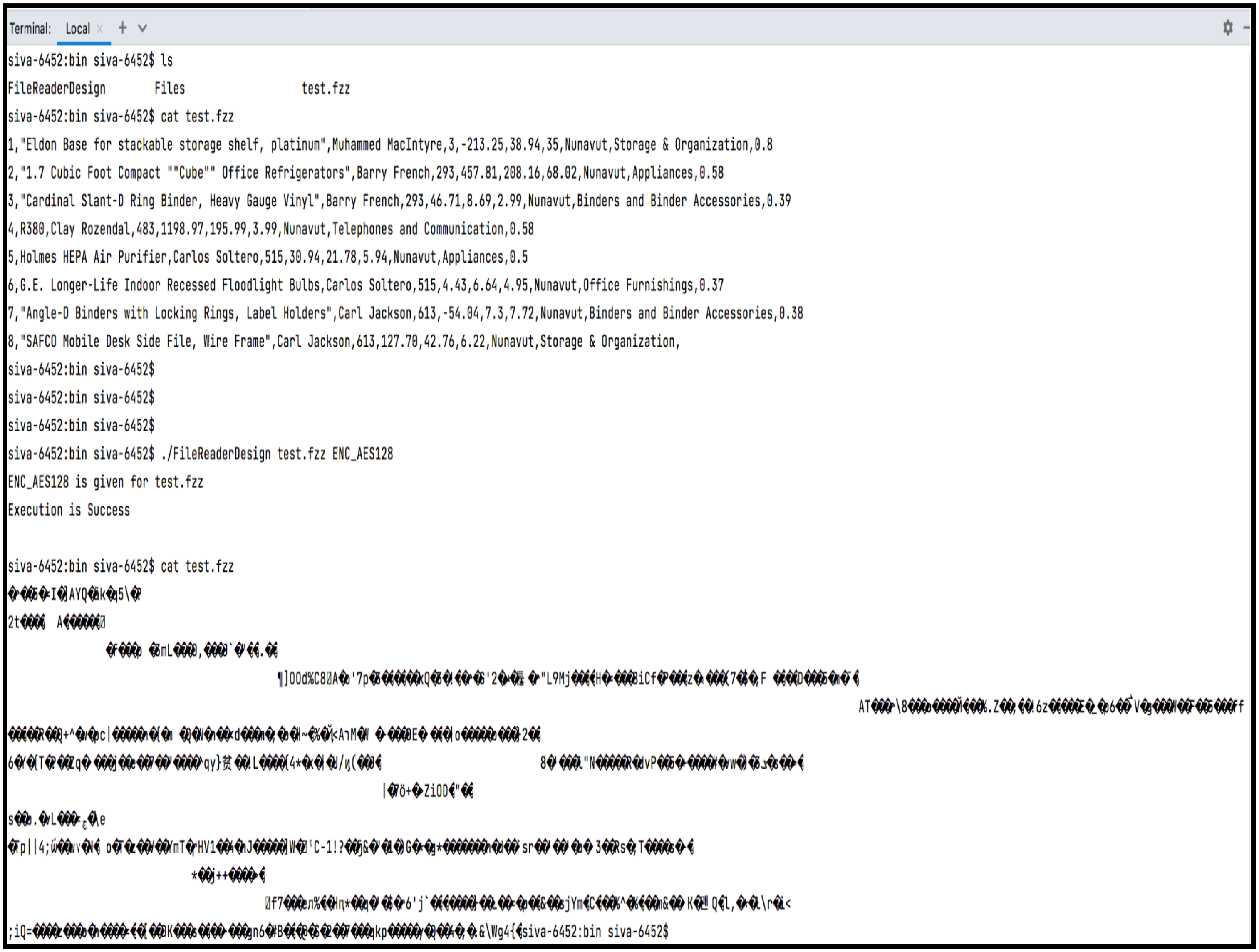
File Reader Design

Written by Sivasubramani T

```
Use case 7:
./FileReaderDesign filename.xxx decompression_format
decryption_format
ex1: ./FileReaderDesign sample.fzz DCOM_GZIP DEC_AES128
ex2: ./FileReaderDesign sample.fzz DCOM_BZIP2 DEC_AES128
```

Samples and Working Model:

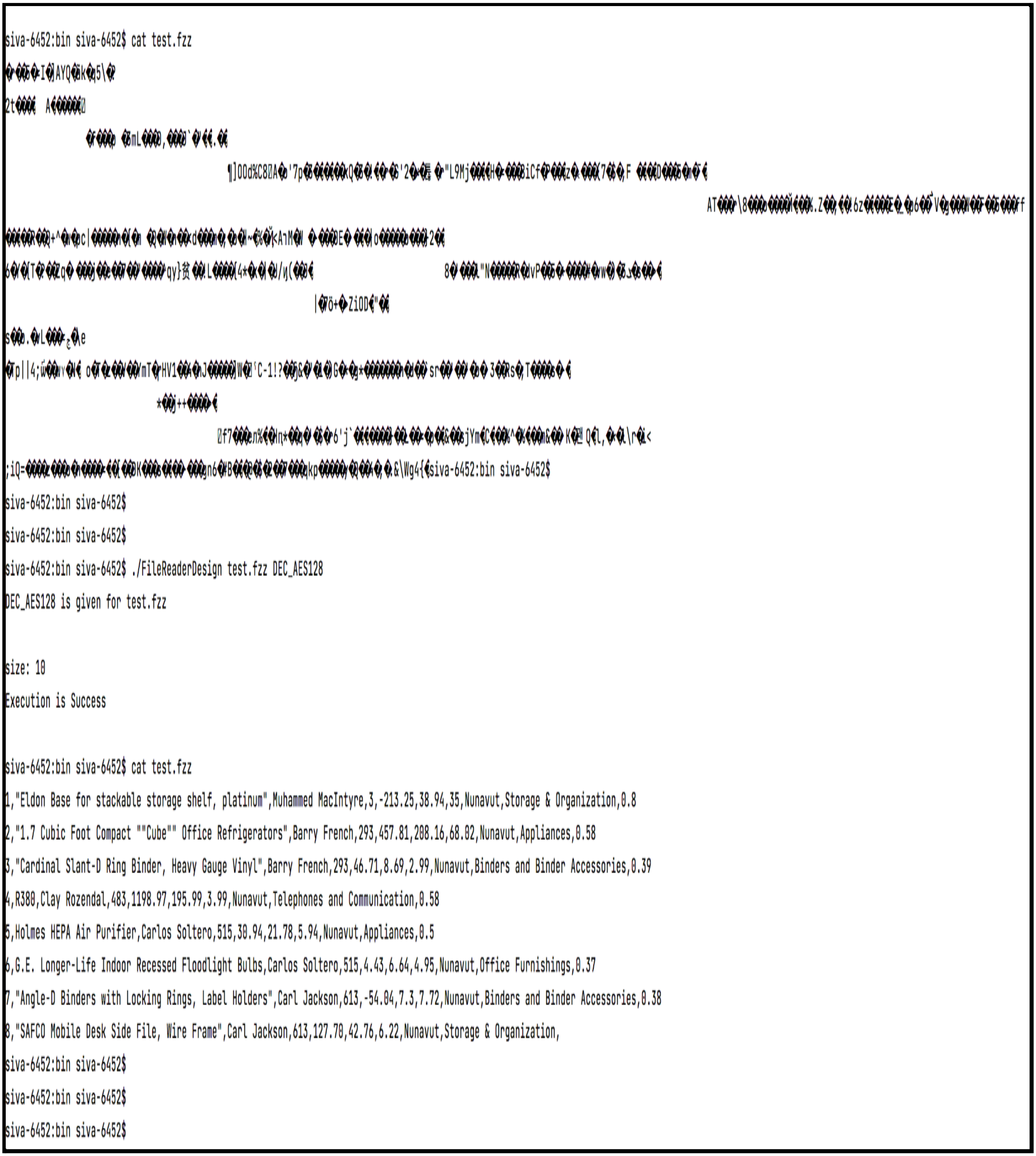
1.*ENC & DEC OF AES128:*



```
siva-6452:bin siva-6452$ ls
FileReaderDesign       Files                   test.fzz
siva-6452:bin siva-6452$ cat test.fzz
1,"Eldon Base for stackable storage shelf, platinum",Muhammed MacIntyre,3,-213.25,38.94,35,Nunavut,Storage & Organization,0.8
2,"1.7 Cubic Foot Compact ""Cube"" Office Refrigerators",Barry French,293,457.81,208.16,68.02,Nunavut,Appliances,0.58
3,"Cardinal Slant-D Ring Binder, Heavy Gauge Vinyl",Barry French,293,46.71,8.69,2.99,Nunavut,Binders and Binder Accessories,0.39
4,R380,Clay Rozendal,483,1198.97,195.99,3.99,Nunavut,Telephones and Communication,0.58
5,Holmes HEPA Air Purifier,Carlos Soltero,515,30.94,21.78,5.94,Nunavut,Appliances,0.5
6,G.E. Longer-Life Indoor Recessed Floodlight Bulbs,Carlos Soltero,515,4.43,6.64,4.95,Nunavut,Office Furnishings,0.37
7,"Angle-D Binders with Locking Rings, Label Holders",Carl Jackson,613,-54.04,7.3,7.72,Nunavut,Binders and Binder Accessories,0.38
8,"SAFCO Mobile Desk Side File, Wire Frame",Carl Jackson,613,127.70,42.76,6.22,Nunavut,Storage & Organization,
siva-6452:bin siva-6452$
siva-6452:bin siva-6452$
siva-6452:bin siva-6452$
siva-6452:bin siva-6452$ ./FileReaderDesign test.fzz ENC_AES128
ENC_AES128 is given for test.fzz
Execution is Success

siva-6452:bin siva-6452$ cat test.fzz
�r��5�I�]AYQ�ãk�q5\�
2t���   A������
               �r���p �3mL���,���`�'��.��
                                           ¶]00d%C8A�o'7p�3������kQ�3�.��r�6'2�����r"L9Mj����H�����3iCf�P���z�.���(7�$�,F ����D���5�m�¯�
                                                                                                                                  AT���r\8���a����Й���%.Z���,���.6z������E�_�p6���V�g����W���F���5����ff
����R���+^�w�pc|�����n�{�m ��W�n���d����m�,�b�H~�%�Й<A¬M�W ����E� ���|o������b����2��
6�/�{T�.��Zq� ���j���e��7�� ����^qy}贫 ��.L����{4*�x�|��/ŋ(���                      8� ���L"N�����R�dvP��5�����#�w�)�3ء�s���
                                                                       |�7ö+�ZiOD�"��
s��b.�wL���-ƹ�\e
�Tp||4;ŵ���wγ�N� o�T�z����/��YmT�rHV1���4�nJ�����]W�'C-1!?���j&��'�1�)G�*�g*��������n�d�� sr�� �� �b� 3��Rs�,T����s��
                                     *��j++�����
                                                    f7���eл%���Hη*���q� �$�r6'j`�������� ��L��-�p��&��sjYm�C���%^�%����m&�� K�ᄈ Q�l,�*�l\r�i<
;iQ=����z����b�n�����-���[��DK���s���� ����gn6�#B���@�$��2��7���qkp�����y�Q���4�,�.&\Wg4{�siva-6452:bin siva-6452$
```

File Reader Design

Written by Sivasubramani T

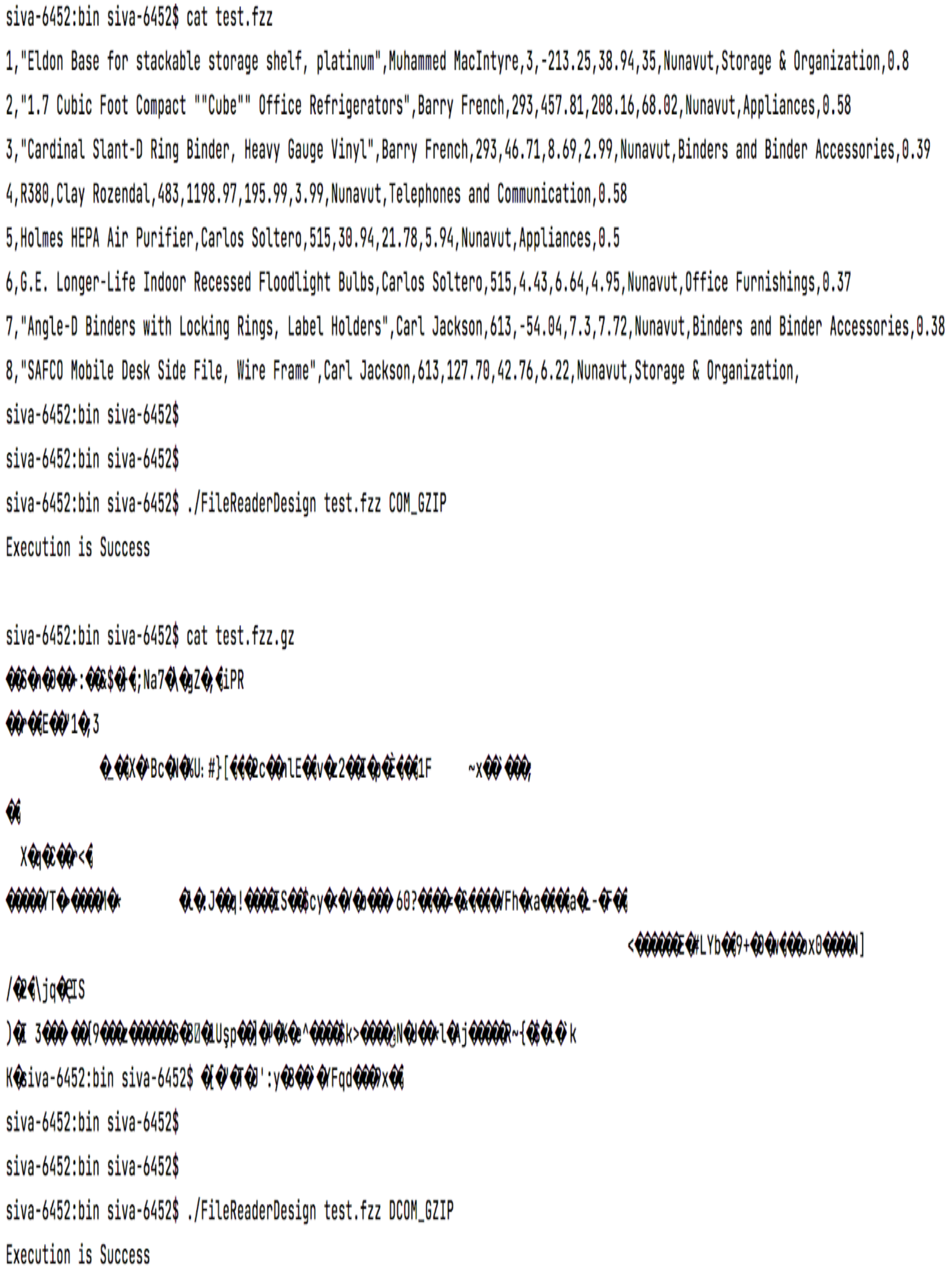


```
siva-6452:bin siva-6452$ cat test.fzz
�r��5�=I�]AYQ�ãk�q5\�?
2t���  A������
                �r���p �3mL���0,���`�'��.��
                                            ¶]OOd%C8A�b'7p�3������xQ�3�.��r�S'2�+��� �r"L9Mj����H�=���8iCf�P���z�.���(7�$�,F ����D���5�m�F�
                                                                                                                                          AT���r\8���b�����й���%.Z��,��.6z�����E�_�p6��̃V�g���Ж�T��5���f
������R���+^�W�pc|�����n�{�n ��W�n���xd����m�,�b�H~�%�ǰ<A¬M�W �����E� ���|o������b����2��
6�r�{T�?��Zq� ���j��e��7��'����qy}贫��.L����{4*�x��|�J/ŋ(���                           8� ���L"N�����R��dvP��5�-�����#�w�)�3J�s���
                                                              |�7ö+�ZiOD€"��
s��b.�vL���-ꞔ�\e
�Tp||4;ŵ���ᵧ�N� o�T�Z���V��YmT�rHV1��4�nJ�����]W�Ɱ'C-1!?���j&�'��1�)G�x�g*������n�d�� sr�� �� �b� 3��Rs�,T����s��
                              *��j++����
                                          ⍰f7���eл%���Hη*���q� ��$�r6'j`����������L���-�p���&���jYm�C����^��%����m&�� K�뺄Q�l,�x��\r��<
;iQ=����Z����n����=��[��DK���s���� ���gn6�#B��@��$�Z��7���qkp�����y�Q���4�,�.&\Wg4{�siva-6452:bin siva-6452$
siva-6452:bin siva-6452$
siva-6452:bin siva-6452$
siva-6452:bin siva-6452$ ./FileReaderDesign test.fzz DEC_AES128
DEC_AES128 is given for test.fzz

size: 10
Execution is Success

siva-6452:bin siva-6452$ cat test.fzz
1,"Eldon Base for stackable storage shelf, platinum",Muhammed MacIntyre,3,-213.25,38.94,35,Nunavut,Storage & Organization,0.8
2,"1.7 Cubic Foot Compact ""Cube"" Office Refrigerators",Barry French,293,457.81,208.16,68.02,Nunavut,Appliances,0.58
3,"Cardinal Slant-D Ring Binder, Heavy Gauge Vinyl",Barry French,293,46.71,8.69,2.99,Nunavut,Binders and Binder Accessories,0.39
4,R380,Clay Rozendal,483,1198.97,195.99,3.99,Nunavut,Telephones and Communication,0.58
5,Holmes HEPA Air Purifier,Carlos Soltero,515,30.94,21.78,5.94,Nunavut,Appliances,0.5
6,G.E. Longer-Life Indoor Recessed Floodlight Bulbs,Carlos Soltero,515,4.43,6.64,4.95,Nunavut,Office Furnishings,0.37
7,"Angle-D Binders with Locking Rings, Label Holders",Carl Jackson,613,-54.04,7.3,7.72,Nunavut,Binders and Binder Accessories,0.38
8,"SAFCO Mobile Desk Side File, Wire Frame",Carl Jackson,613,127.70,42.76,6.22,Nunavut,Storage & Organization,
siva-6452:bin siva-6452$
siva-6452:bin siva-6452$
siva-6452:bin siva-6452$
```

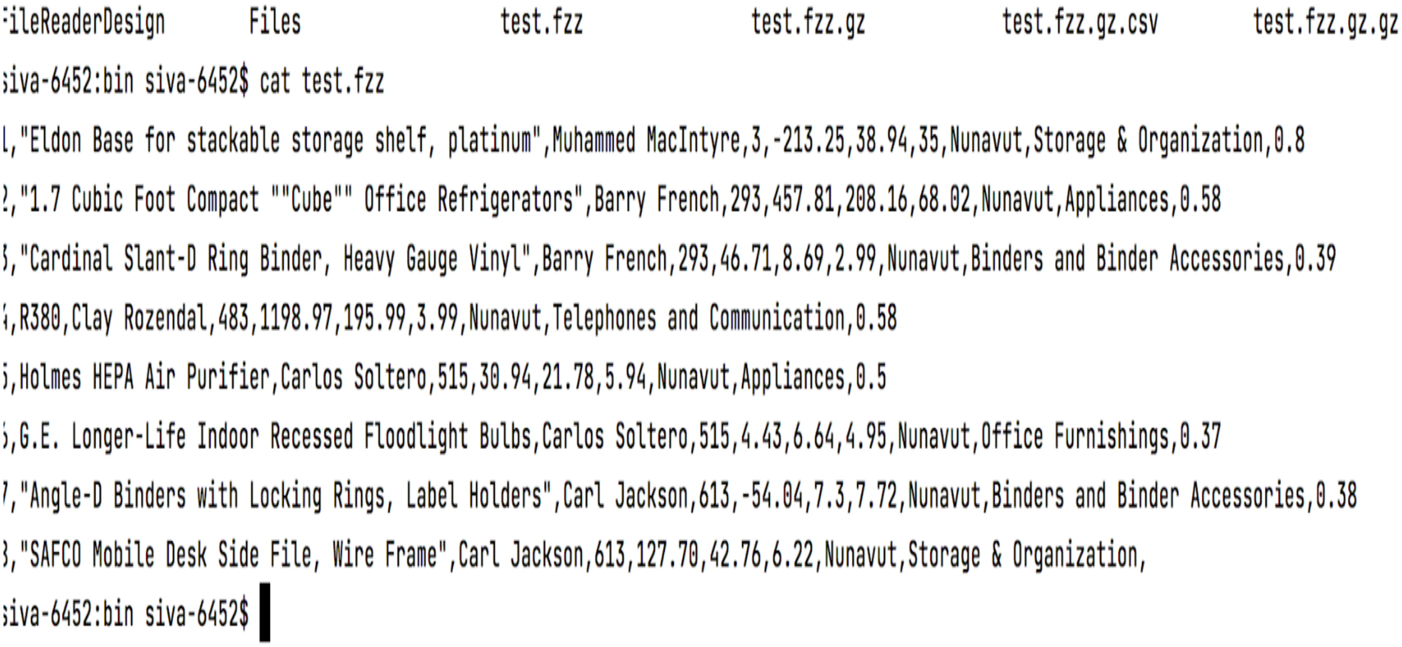
File Reader Design

Written by Sivasubramani T

2. COM & DCOM OF GZIP/BZIP:



```
siva-6452:bin siva-6452$ cat test.fzz
1,"Eldon Base for stackable storage shelf, platinum",Muhammed MacIntyre,3,-213.25,38.94,35,Nunavut,Storage & Organization,0.8
2,"1.7 Cubic Foot Compact ""Cube"" Office Refrigerators",Barry French,293,457.81,208.16,68.02,Nunavut,Appliances,0.58
3,"Cardinal Slant-D Ring Binder, Heavy Gauge Vinyl",Barry French,293,46.71,8.69,2.99,Nunavut,Binders and Binder Accessories,0.39
4,R380,Clay Rozendal,483,1198.97,195.99,3.99,Nunavut,Telephones and Communication,0.58
5,Holmes HEPA Air Purifier,Carlos Soltero,515,30.94,21.78,5.94,Nunavut,Appliances,0.5
6,G.E. Longer-Life Indoor Recessed Floodlight Bulbs,Carlos Soltero,515,4.43,6.64,4.95,Nunavut,Office Furnishings,0.37
7,"Angle-D Binders with Locking Rings, Label Holders",Carl Jackson,613,-54.04,7.3,7.72,Nunavut,Binders and Binder Accessories,0.38
8,"SAFCO Mobile Desk Side File, Wire Frame",Carl Jackson,613,127.70,42.76,6.22,Nunavut,Storage & Organization,
siva-6452:bin siva-6452$
siva-6452:bin siva-6452$
siva-6452:bin siva-6452$ ./FileReaderDesign test.fzz COM_GZIP
Execution is Success


siva-6452:bin siva-6452$ cat test.fzz.gz
��S�n����:��&$��;Na7�\�gZ�,�iPR
���E��'1�,3
            �_��X�^Bc�N�%U:#}[���2c��nlE��v�z2��I�p�È���1F     ~x��`���,
��
  X�q�C���<�
�����YT�����M�*        �L�.J��q!����IS���cy�*�/�b��� 60?���-�&����/Fh�ka����a�L-�F��
                                                                                   <������E�#LYb��9+�D�w����x0�����N]
/�2�\jq�CIS
)�I 3��� ��{9���z������6�3��lUşp��]�"�%�e^�����k>����¸N�J��*l�Aj�����R~{�$�L�`k
K�siva-6452:bin siva-6452$ �[�'�T�J':y�3��`�/Fqd����x��
siva-6452:bin siva-6452$
siva-6452:bin siva-6452$
siva-6452:bin siva-6452$ ./FileReaderDesign test.fzz DCOM_GZIP
Execution is Success
```



```
ileReaderDesign       Files                 test.fzz             test.fzz.gz           test.fzz.gz.csv        test.fzz.gz.gz
iva-6452:bin siva-6452$ cat test.fzz
,"Eldon Base for stackable storage shelf, platinum",Muhammed MacIntyre,3,-213.25,38.94,35,Nunavut,Storage & Organization,0.8
,"1.7 Cubic Foot Compact ""Cube"" Office Refrigerators",Barry French,293,457.81,208.16,68.02,Nunavut,Appliances,0.58
,"Cardinal Slant-D Ring Binder, Heavy Gauge Vinyl",Barry French,293,46.71,8.69,2.99,Nunavut,Binders and Binder Accessories,0.39
,R380,Clay Rozendal,483,1198.97,195.99,3.99,Nunavut,Telephones and Communication,0.58
,Holmes HEPA Air Purifier,Carlos Soltero,515,30.94,21.78,5.94,Nunavut,Appliances,0.5
,G.E. Longer-Life Indoor Recessed Floodlight Bulbs,Carlos Soltero,515,4.43,6.64,4.95,Nunavut,Office Furnishings,0.37
,"Angle-D Binders with Locking Rings, Label Holders",Carl Jackson,613,-54.04,7.3,7.72,Nunavut,Binders and Binder Accessories,0.38
,"SAFCO Mobile Desk Side File, Wire Frame",Carl Jackson,613,127.70,42.76,6.22,Nunavut,Storage & Organization,
iva-6452:bin siva-6452$ █
```