The main function of this software is contained in main.py file.

This main.py starts to read airport and flight file with details , and then process according to the mapreduce algorithm, so it results in the organized outcomes.

There used 10 threads for data process and every thread will process following the same algorithm and procedure.

In general, the main function in python is declared as follows.

If \_\_name\_\_ == “\_\_main\_\_”:

main()

Looking through the code in main.py, it will be clear about that.

For more details, please refer to python documentations and tutorials.

This part in code is on line 174.

Then, all procedures will be implemented in function “main”.

The function in python is declare and defined as follows.

def main():

*Content of function*

When looking through the lines 160-165, it will be find out how the threads are created and linked to the core of the thread.

In python, thread is created and linked to the function as follows.

At first, a package for thread has to be imported.

Import threading

Then,

thread = threading.Thread(target = process\_fun, args = (arg1, arg2,))

Here, *process\_fun* contains the content to be run in thread and *args* represents the arguments to be passed to the *process\_fun*.

After creating thread, it must be started. It is as follows.

thread.start()

Please go to the definition of process function.

Here it reads the data from csv files.

It is implemented as follows in code.

Please look mapper.py.

It can be find the functions “read\_flight” and “read\_airport”.

They read the flights and airports from files, respectively.

In python, it is easy to read csv file and as follows.

f = open(“file.csv”, “r”)

While reading files, it is to fix errors in files.

It is implemented in function “fix\_error”.

In this function, it takes every character ad ascii code and compares whether that value is between the range desired.

If not so, it calls function “findCorrect”.

This function keeps all airport data and find the matched airport name.

The mapreduce algorithm is implemented in lines 101-127 in main.py

Here uses dictionary in python. That is,

flight\_from\_airport = {}

All informations can be saved in this dictionary type variable.

For example,

flight\_from\_airport = {“DEN”:2, “JFK”:2}

After every thread processes his data, it will collected into one and summed up according to the mapreduce algorithm.

It is implemented in function “reduceAndSum “ in main.py.

Then, all informations are organized rightly.

I wrote summary of the software.

So if you have any questions here, pls have a free chance to ask.