

# Documentation of results

The structure of the project is as follows:

The **Record.java** file offers a way of parsing the raw data file. Using the constructor, all necessary information for this exercise is initialized.

This time mappers and reducers are existing twice:

- **RecordMapperNaive** and **MeanTemperatureReducerNaive** for Task 1
- **RecordMapperCombine** and **MeanTemperatureReducerCombine** for Task 2

In both Mappers, custom counters are incremented if respective conditions are met.

For Task 2 **MeanTemperatureCombiner** was implemented. A new class acting as data transfer object was introduced as well (**SumCountWritable**). This class is then used in input/output values of mapper, reducer and combiner. For testing the combiner, every input/output data had to be the same:

- **Mapper Output:** `<Text, SumCountWritable>`
- **Combiner Input:** `<Text, SumCountWritable>`
- **Combiner Output:** `<Text, SumCountWritable>`
- **Reducer Input:** `<Text, SumCountWritable>`

Therefore, **RecordMapperCombine** had to be implemented (which is also using **SumCountWritable** in signature)

Tests are provided in the **MeanTemperatureTestNaive** class. While Mapper and Reducer are tested individually and in combination, another test also assures that only valid records are processed. In addition, counter-logic is tested as well.

✓	MeanTemperatureTestNaive (org.ex 1 sec 868 ms)
✓	testMapper 1 sec 535 ms
✓	testFaultyTemperatureCounter 65 ms
✓	testInvalidRecordFiltered 123 ms
✓	testQualityCodeCounter 48 ms
✓	testMapReduce 56 ms
✓	testReducer 41 ms

In **MeanTemperatureTestCombine** the combiner is tested as well.

MeanTemperatureTestCombine (org 1 sec 250ms)	
✓ testReducerFinal	1 sec 22 ms
✓ testCombiner	36 ms
✓ testFaultyTemperatureCounter	103 ms
✓ testMapReduceWithCombiner	42 ms
✓ testInvalidRecordFiltered	29 ms
✓ testQualityCodeCounter	18 ms

The results are:

```
QualityCounter
    1=13129
    9=1
Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
File Input Format Counters
    Bytes Read=1790298
File Output Format Counters
    Bytes Written=44
org.example.mean_temperature.RecordMapperCombine$FaultyTemperatureCounter
    MISSING=1
bash-4.2$ cat /home/hadoop/result
1901    4.6698507007922
1902    2.165955826351866
bash-4.2$ 
```

Rewriting and using the .sh script provided in exercise 1 i shows that result is valid:

```
→ Exercise_2 git:(main) ✘ ./calc_solutions.sh
1901    46,6985
1902    21,6596
→ Exercise_2 git:(main) ✘ 
```