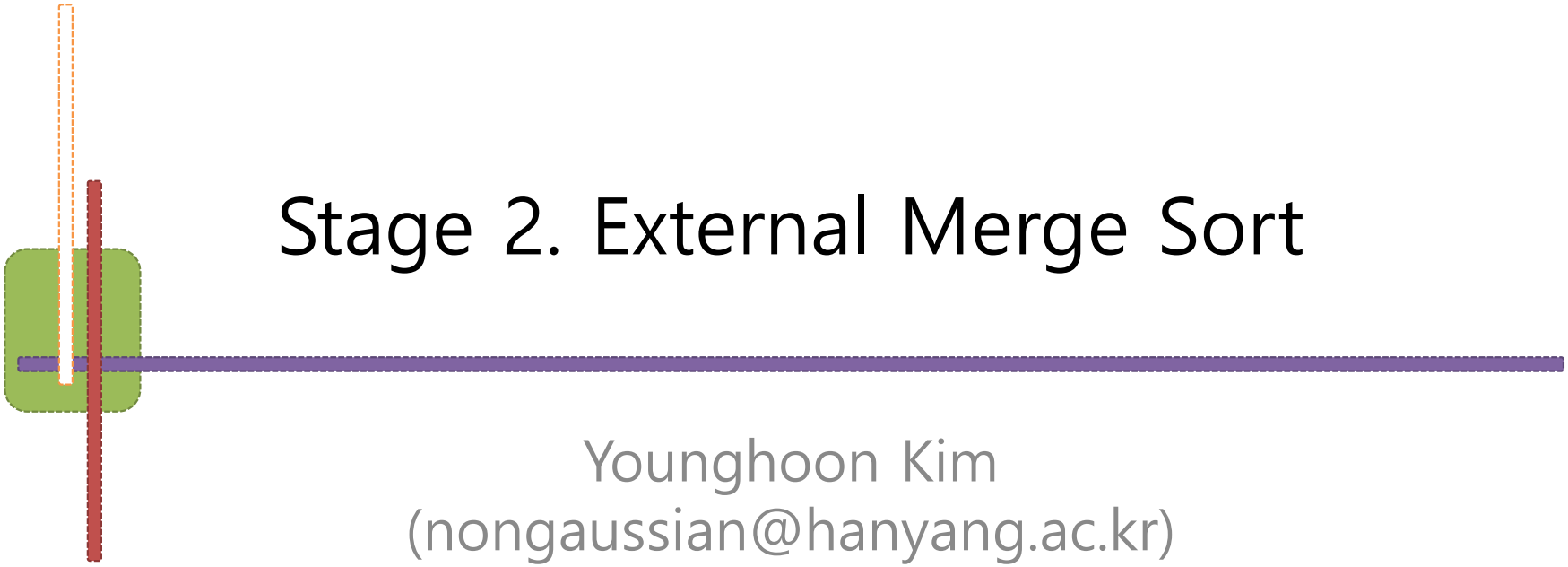

Stage 2. External Merge Sort

Younghoon Kim
(nongaussian@hanyang.ac.kr)

Goal

- Given
  - A file
    - Containing a list of triples with 3 integers (e.g., <5, 1, 2>)
    - For triples, use `org.apache.commons.lang3.tuple.MutableTriple`
- Return
  - A file
    - A list of triples sorted in the ascending order by using external merge sort
  - Sorting criteria
    - Primarily, sort by the first value
    - With tuples with an identical first value, use the second value
    - With tuples with identical first and second values, use the third value
  - Example

(4,8,4)(4,5,4)(7,9,6)(0,6,5)(6,0,3)(0,5,3)(3,1,7)(5,4,9)(4,6,6)(9,1,1)

(0,5,3)(0,6,5)(3,1,7)(4,5,4)(4,6,6)(4,8,4)(5,4,9)(6,0,3)(7,9,6)(9,1,1)

Code Template

- We provide a package of
  - A maven project created in Eclipse
- It contains
  - Template codes (`edu.hanyang.submit.TinySEExternalSort.java`)
  - TinySE framework (`lib/tinyse-0.0.1-SNAPSHOT.jar`) ← to be updated on every stage
    - Interface files (e.g., ExternalSort.java)
    - Indexer and query processer codes which will complete a search engine by connecting your submissions

To Use Code Template

- Complete *edu.hanyang.submit.TinySEExternalSort*

```
package edu.hanyang.indexer;


public interface ExternalSort {
        public void sort(String infile,  /* input file path */
                         String outfile, /* output file path */
                         String tmpdir,  /* temporary directory
                                            for creating intermediate
                                            runs */
                         int blocksize,  /* 8192 bytes */
                         int nblocks);   /* available memory size /
                                            blocksize */
}
```

Read and Write Buffers

- To implement an external sort,
  - You may need read a given size of blocks sequentially for each run
  - How do we implement it in Java?
  - → Use BufferedReader

Utility Class

- DiskIO.class
  - edu.hanyang.utils.DiskIO;

Method Summary

All Methods | **Static Methods** | **Concrete Methods**

Modifier and Type	Method and Description
static void	**append_arr**(java.io.DataOutputStream out, java.util.List<org.javatuples.Triplet<java.lang.Integer,java.lang.Integer,java.lang.Integer>> arr, int nelements) Write the data which in 'arr' from zero to 'nelements', to file.
static java.io.DataInputStream	**open_input_run**(java.lang.String filepath) Create and return DataInputStream instance.
static java.io.DataOutputStream	**open_output_run**(java.lang.String filepath) Create and return DataOutputStream instance.
static int	**read_array**(java.io.DataInputStream in, int offset, int nelements, java.util.ArrayList<org.javatuples.Triplet<java.lang.Integer,java.lang.Integer,java.lang.Integer>> arr) Read Triplet data from DataInputStream and insert into given ArrayList.
static void	**sort_arr**(java.util.List<org.javatuples.Triplet<java.lang.Integer,java.lang.Integer,java.lang.Integer>> arr, int nelements) Sort the Triplet which in given ArrayList from zero to 'nelements'.

Methods inherited from class java.lang.Object

equals, getClass, hashCode, notify, notifyAll, toString, wait, wait, wait

Constrain & Submit

- How to submit
  - Run “maven package”
  - Submit <your student ID>-0.0.1-SNAPSHOT.jar file (you can find it from <project dir>/target/)
- If any question, contact TA
  - Keonwoo Kim (김건우)
  - kdbml314@gmail.com
  - Room: 4공학관 314호
- Due date
  - Apr. 17 (11:59pm)
- Unit test
  - The maven project will be updated and provided soon
  - It will include a unit test code, test and answer files to let you test the correctness of output sorted list
  - BUT, it does not guarantee whether or not your submit works well with larger files

To Use Code Template

- Test your code