

# HotTopics in BigData

## Homework-1

Kookmin University

Department of Computer Science

Kyungyong Lee

# Write a MapReduce Program

- Calculating an average review score of each product
- Dataset
  - `hdfs:///input/amazon/reviews_Books_5.json`
  - Dataset description - <http://jmcauley.ucsd.edu/data/amazon/>
- The output should be stored in the HDFS `/student{ID}/` folder
  - Note that this is an individual assignment and should not share the output folder
- Due – May 22<sup>nd</sup>.
- What to submit
  - Java file
  - `part-r-00000`
- Homework handling via GitHub
  - <https://classroom.github.com/assignment-invitations/b82111623914f6a65bedf3bb789f6ac7>

# Tips

- Recommend to use JSON parser
  - In the #assignment channel (java-json.jar)
- Sample output in the right

```
000100039X 4.674757281553398
0001055178 3.5555555555555554
0001473123 4.625
0001473727 5.0
0001473905 4.666666666666667
0001712772 4.666666666666667
000171287X 4.583333333333333
0001714538 3.7142857142857144
0002005395 4.071428571428571
0002006715 4.6
0002007649 2.888888888888889
0002007770 4.3524439544285185
0002008572 4.764705882352941
0002051850 4.446969696969697
0002111306 4.142857142857143
0002113570 4.852941176470588
0002117088 4.785714285714286
0002153211 4.4
0002154129 4.833333333333333
0002154129 4.8
```