

Last HOMEWORK

Welcome to your last homework, here are the following are the **rules** of submitting homework:

- You have to submit the code/report in **PDF**, compiled from **Markdown**. Any report written in word document or so will be discarded and score 0.
 - You have to submit your homework via Email to ncuhomework@outlook.com.
 - Your Email subject shall be **Course Z6110X0035 CC+Homework #+{Your Name}+{Your ID}**. For example, if I submit a homework, my Email subject is Course Z6110X0035 CC Homework 5 Zichen Xu 1234556678.
-

Assignment

Storage for Data Intensive Services

1. HDFS is implemented as a user-level file system vs an in-kernel file-system. (a) What is the advantage of this in the context of Hadoop?
2. The output of a Mapper is written into the local filesystem instead of the global filesystem. Why? Your answer should explain both why writing into the global file system would be undesirable as well as why it would be of minimal benefit.
3. Why does Hadoop sort records en route to a Reducer? How would it affect things if these records were processed by the Reducer in the order in which they were received from the various Mappers?
4. How is the failure of a Mapper or Reduce managed?
5. In a typical Map-Reduce graph algorithm, what data structure is used to represent the graph? Why?
6. In a typical Map-Reduce graph algorithm, how many Map-Reduce phases are typically necessary before the graph can be traversed? Why?

HINTS:

- Check our former course slides on Hadoop and MapReduce;
 - Combine this homework with your lab this week, maybe try it in your lab finishes.
-