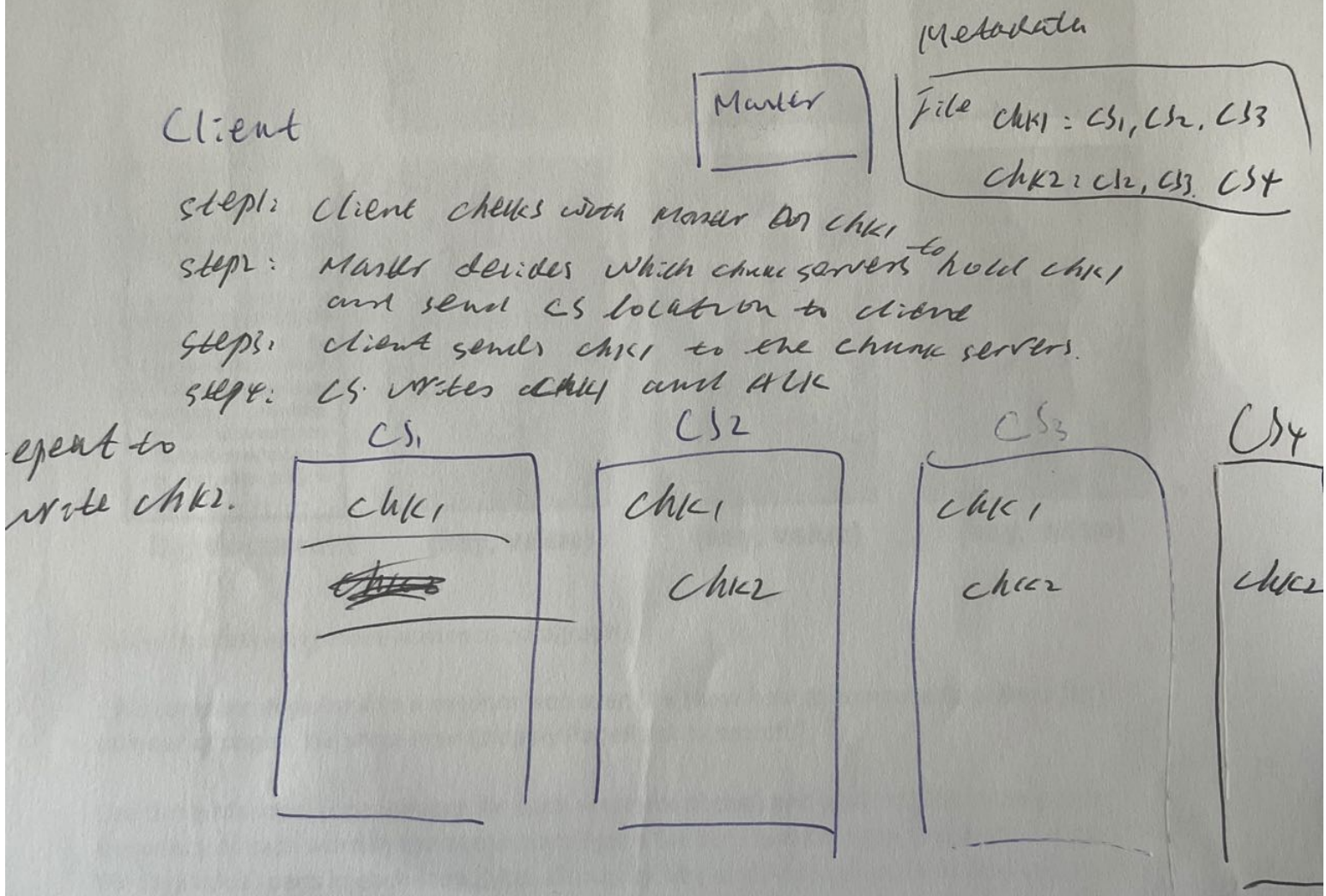
CIS4517/5517 Quiz 3 Name:

1. (10 pts) Given a GFS cluster consisting of 1 master and 4 chunk servers with a default configuration (chunk size: 64 MB, each chunk with 3 replicas), explain step by step how a client writes a file with the size of 100MB to this GFS cluster. Show the final outcome of chunk servers and the metadata in the master node.



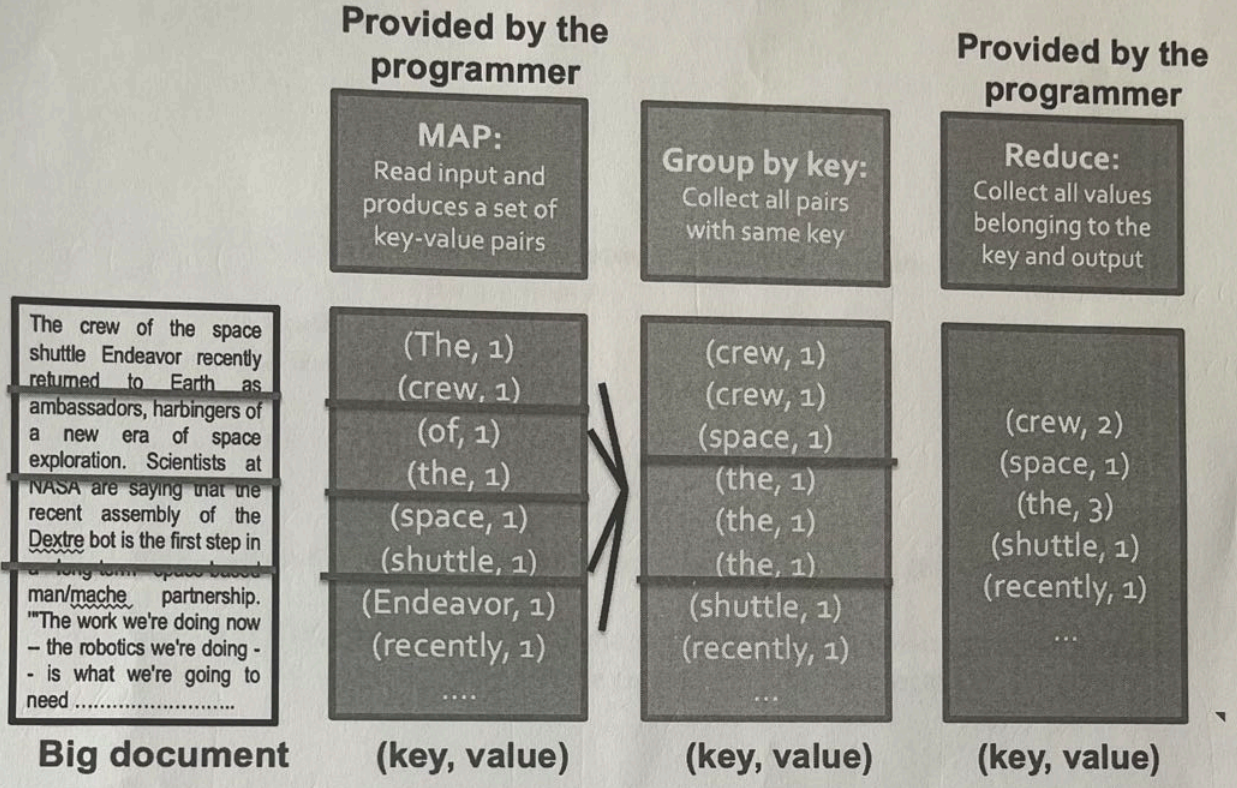

repeat to write chk2.

File: 100MB: 2 chunks
Each chunk has 3 replicas.

| chk1 | chk2 |
|---|---|

Flip over →

2. (10 points) We discussed how to use Map/Reduce to count the frequency of words in a text document as shown below.




Given the following three-sentence paragraph,

*"We compare PageRank to a random web user. We show how to compute PageRank for a number of pages. We show how to apply PageRank to search."*

Use three Mappers (one mapper for each sentence above) and one reducer to count the frequency of each word in the above para (igore the punctuation signs and white spaces). Show the (key value) pairs in each step (Map, Group by Key, and Reduce) and the final output of the reducer.

Map 1
(we, 1)
(compare, 1)
(pageRank, 1)

Map 2
(we, 1)
(show, 1)

Map 3
(we, 1)
(show, 1)

Group by key
(we, 1)
(we, 1)
(we, 1)

reduce
(we, 3)
(show, 2)
(PageRank,

3. (2 points) Which **two** of the following operations are new operations designed in GFS (compared to traditional file systems)?

   A: Random writes
   B: Snapshots
   C: Sequential writes
   D: Record appends

   B, D

4. (2 points) In HDFS, communications between the name node and data nodes are conducted by which following scheme:

   A: Queues
   B: Heartbeat messages
   C: RPC
   D: Pipes

   B

5. (6 points) True or false (Provide a brief explanation if you think the statement is false.)

   1) In default GFS configurations, the master node stores chunk-related metadata and maps the 64-bit chunk labels to the corresponding chunk locations.

      T

   2) In HDFS, when a client accesses a data block, it first checks with the name node to get the block location, and then the name node will retrieve the data block from the corresponding data node and transfer the corresponding data block to the client.

      False, data doesn't flow through name node.

   3) In a typical map-reduce task, the number of mappers (M) is ~~smaller~~ greater than the number of reducers (R) to be efficient.

      False,