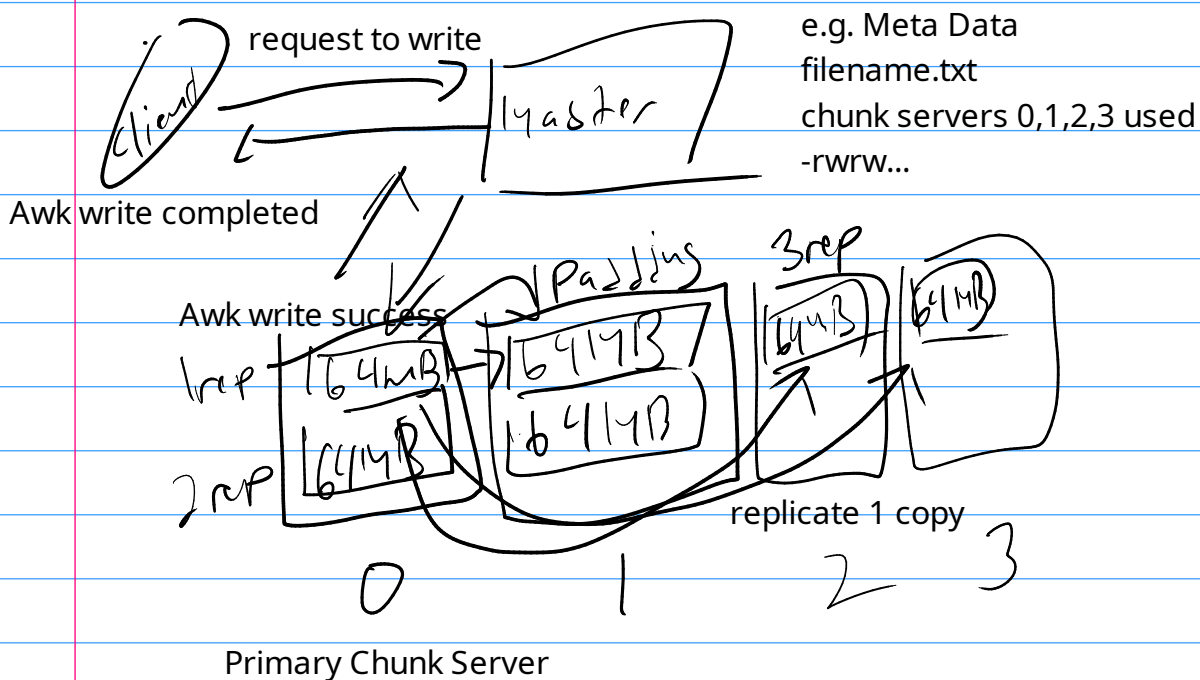


1. Tony Siu CIS 5517 Quiz 3



First the client makes a request to the master server for the master server. The client data is divided into two 64 mb chunks where the last chunk is padded. is padded. The master server sends an awk request due proximity to a chunk server where the chunk server becomes the primary chunk server. The chunked client data is sent to the primary chunk server and the primary chunk server sends the other chunk to a nearby chunk server. The primary chunk server makes duplicates of the client data in the same chunks. These chunks are distributed to other chunk servers due proximity of the primary chunk server. This is illustrated above. Once this process is done, the primary chunk server sends an awk to the master server to confirm the write process is completed. The master server records the metadata of the client data chunks, its written location and access permissions. Finally, the master server sends a response back to the client saying that the write process is finished.

2.

Step 1

Mapper 1

to process ->"We compare PageRank to a random web user"

we,1

compare,1

pagerank,1

to,1

a,1

random,1

web,1

user,1

Mapper 2

to process ->"We show how to compute PageRank for a number of pages"

we,1

show,1

how,1

to,1

compute,1

pagerank,1

for,1

a,1

number,1

of,1

pages,1

Mapper 3

to process ->"We show how to apply PageRank to search"

we,1

show,1

how,1

to,1

apply,1

pagerank,1

to,1

search,1

number,1

of,1

pages,1

apply,1

search,1

Step 2 Grouping step

we,(1,1,1)

pagerank,(1,1,1)

compare,(1)

to,(1,1,1,1)

a,(1,1)

random(1)

web,(1)

user,(1)

show,(1,1)

how,(1,1)

compute,(1)

for,(1)

number,(1)

of,(1)

pages,(1)

apply,(1)

search,(1)

Step 3 one reducer

we,3

pagerank,3,

compare,1

to,4

a,2

random,1

web,1

user,1

show,2

how,2

compute,1

for,1

3. B,D

4.B

5.1 True

5.2 True, The namednode may send the request to the datablock to send the data to the client but the namednode itself never stores or has direct access to the stored client data in the datanodes. Depends on what "retrieve" means.

5.3 False, typically mapping tasks are requires more resources for map the key value pairs, so therefore its the other way around. Typically more mapper than reducers.