

quiz 2

wbg231

January 2023

1 Introduction

question 1

- a split in map reduce can either be smaller or larger than an hdfs block
- a split in map reduce is pretty much just file size over number of mappers so it is dependent on input file size

question 2

- Combiners in map-reduce improve efficiency by decreasing the number of intermediate keys prior to the reduce phase.
- [false](#) Combiners reduce the number of key value pairs not the number of keys

question 3:

- the replication factor of an HDFS file can be modified after the file is created
- [true](#) hdfs does not care if you change the replication size it would just have to create or delete replicated block data

• question 4

- the data nodes in hdfs are responsible for file name and location data
- [false](#) that is what the name nodes document

question 5

- question 5 is cursed i'm going to skip for now. it is like writing the psudo code for a map reduce function