

## Tutorial Numerical Question

11. If you have a file of size 100 GB stored in HDFS with a replication factor of 3, how much total storage space is used to store this file?
12. A file of size 256 MB is stored in HDFS with a block size of 128 MB. How many blocks will this file be divided into?
13. Suppose you have 5 DataNodes in your Hadoop cluster, and each DataNode has 10 TB of storage. If you set the replication factor to 2, what is the maximum amount of unique data (in TB) that your cluster can store?
14. Consider a MapReduce job that processes 1,000,000 lines of text, with each line containing an average of 10 words. If the map function emits a key-value pair for each word, how many key-value pairs will the map phase produce?
15. A MapReduce job is configured with 5 reducers. If each reducer processes an average of 200,000 key-value pairs, how many key-value pairs are processed in total?
16. In a word count MapReduce job, if using a combiner reduces the number of key-value pairs by 60% before they reach the reducer, how many key-value pairs will be sent to the reducers if the map phase originally produced 500,000 key-value pairs?
17. If a Hive query takes 15 minutes to process 1 TB of data, estimate how long it would take to process 3 TB of data, assuming the processing time scales linearly with the data size.
18. A Hive table is partitioned by date, and each partition contains 100 GB of data. If the table contains 365 partitions, what is the total size of the table?
19. In a Hive query, if one reducer is handling 40% of the total data and the remaining reducers handle the rest equally, what percentage of the total data is handled by each of the remaining reducers if there are 4 reducers in total?
20. Suppose you create an index on a column of a Hive table that reduces the query time by 50% and the index occupies 10% of the table's size. If the table is 1 TB in size, how much storage space is used by the index?