

Scenario 1: You are working on configuring a Hadoop cluster for a project that deals with large files. The cluster has a total storage capacity of 100 TB. The average file size in your project is 1 GB. What would be an appropriate block size for your Hadoop cluster, and why?

Scenario 2: In your Hadoop cluster, you have set the replication factor to 3. However, during a routine check, you notice that some blocks have only one replica. What could be the possible causes of under-replication in Hadoop, and how would you address this issue?

Scenario 3: You have a Hadoop cluster with a replication factor of 2, and you want to increase the replication factor to 3 for all the existing data blocks. What steps would you take to achieve this target replication in Hadoop?

Scenario 4: You are setting up a new Hadoop cluster, and you want to optimize the performance by adjusting the block size. The cluster has a total storage capacity of 50 TB, and you anticipate having a large number of small files (around 100 KB each). What would be an appropriate block size for this scenario, and how would it impact the cluster's performance?

Scenario 5: You have a Hadoop cluster with a replication factor of 3. Due to hardware limitations, you can no longer maintain the desired replication factor for all the data blocks. How would you prioritize the replication of data blocks to ensure data availability and fault tolerance in such a situation?

Scenario 6: You have a Hadoop cluster with a replication factor of 3. One of the data nodes in the cluster fails, and you need to replace it with a new node. How would you ensure that the data blocks stored on the failed node are replicated to maintain the desired replication factor?

Scenario 7: In your Hadoop cluster, you have set the block size to 128 MB. However, you notice that some of the files in your cluster are very small, around 10 KB. How does the small file problem impact storage efficiency and cluster performance, and what strategies could you employ to mitigate this issue?

Scenario 8 : You are configuring a Hadoop cluster and need to decide on an appropriate replication factor. What factors would you consider when determining the replication factor, and how would you strike a balance between data availability, fault tolerance, and storage overhead?

Scenario 9: You have a Hadoop cluster where the replication factor is set to 2. However, due to a hardware failure, one of the data nodes becomes permanently unavailable, resulting in under-replicated blocks. How would you identify and resolve the under-replication issue in your Hadoop cluster?

Scenario 10: You are working on a Hadoop project with strict data durability requirements. How would you ensure that the data stored in Hadoop is adequately replicated and protected against data loss or corruption?

Scenario 11: You have a Hadoop cluster with a default block size of 64 MB. However, you are dealing with large files that are typically 2 GB in size. How would you determine an appropriate block size for your large files, and what are the implications of choosing a larger block size?

Scenario 12: In your Hadoop cluster, the replication factor is set to 2. However, you realize that some of the data blocks have more than two replicas. What could be the possible causes of over-replication in Hadoop, and how would you address this issue?

Scenario 13: You are tasked with optimizing the storage efficiency of your Hadoop cluster. Currently, you have a replication factor of 3, but you want to reduce it without compromising data availability and fault tolerance. How would you determine the optimal replication factor for your cluster?

Scenario 14: In your Hadoop cluster, you observe that some data blocks have become under-replicated due to node failures. How would you identify these under-replicated blocks, and what steps would you take to restore the desired replication factor?

Scenario 15: You have a Hadoop cluster with a replication factor of 3, but you want to increase it to 4 for enhanced data durability. What considerations would you take into account when deciding to increase the replication factor, and what steps would you follow to achieve the new target replication?

Scenario 16: You are working on a Hadoop project with a large dataset consisting of both small and large files. How would you choose an appropriate block size to optimize the storage efficiency and processing performance for such a mixed workload?

Scenario 17: In your Hadoop cluster, the replication factor is set to 3. Due to a sudden increase in data volume, you notice that the cluster is running out of storage space. How would you handle this situation to accommodate the growing data while maintaining the desired replication factor?

Scenario 18: You have a Hadoop cluster with a replication factor of 2. However, you want to increase the replication factor to 3 for specific directories that contain critical data. How would you selectively set a higher replication factor for those directories while keeping the default replication factor for the rest of the cluster?

Scenario 19: You are troubleshooting an issue in your Hadoop cluster and suspect that there might be under-replicated blocks causing data inconsistencies. How would you identify and resolve under-replication in Hadoop to ensure data integrity and fault tolerance?

Scenario 20: You have a Hadoop cluster with a target replication factor of 2. However, you notice that some of the data blocks have more than two replicas, leading to unnecessary storage overhead. How would you identify and address the over-replication issue in your Hadoop cluster?