1. Why do we need to specify the number of reducers, but not the number of mappers in Hadoop?

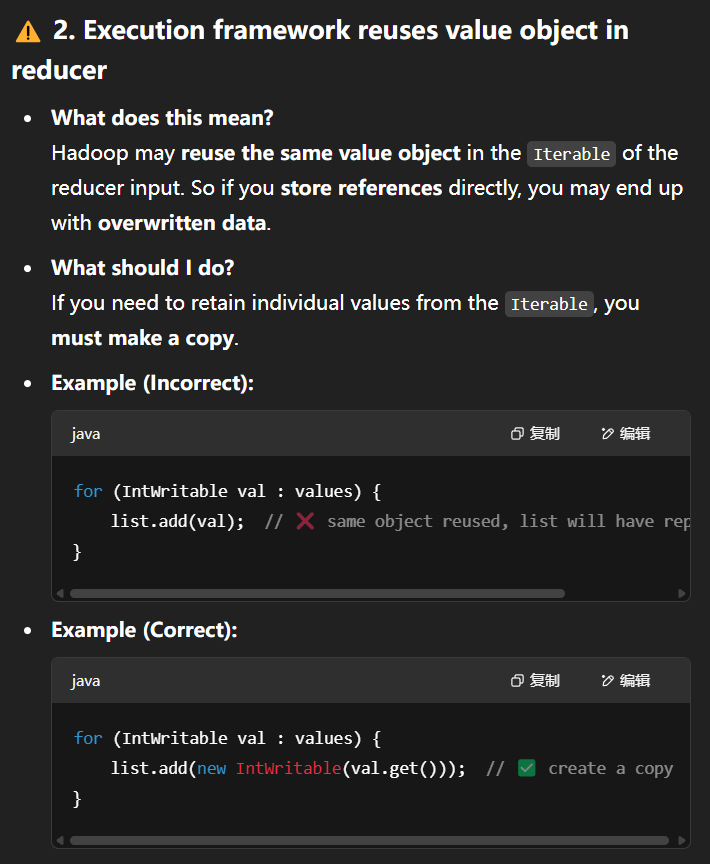
The number of mappers is automatically decided by how the input data is split.

Input data is divided into splits (usually based on HDFS block size, e.g., 128MB).

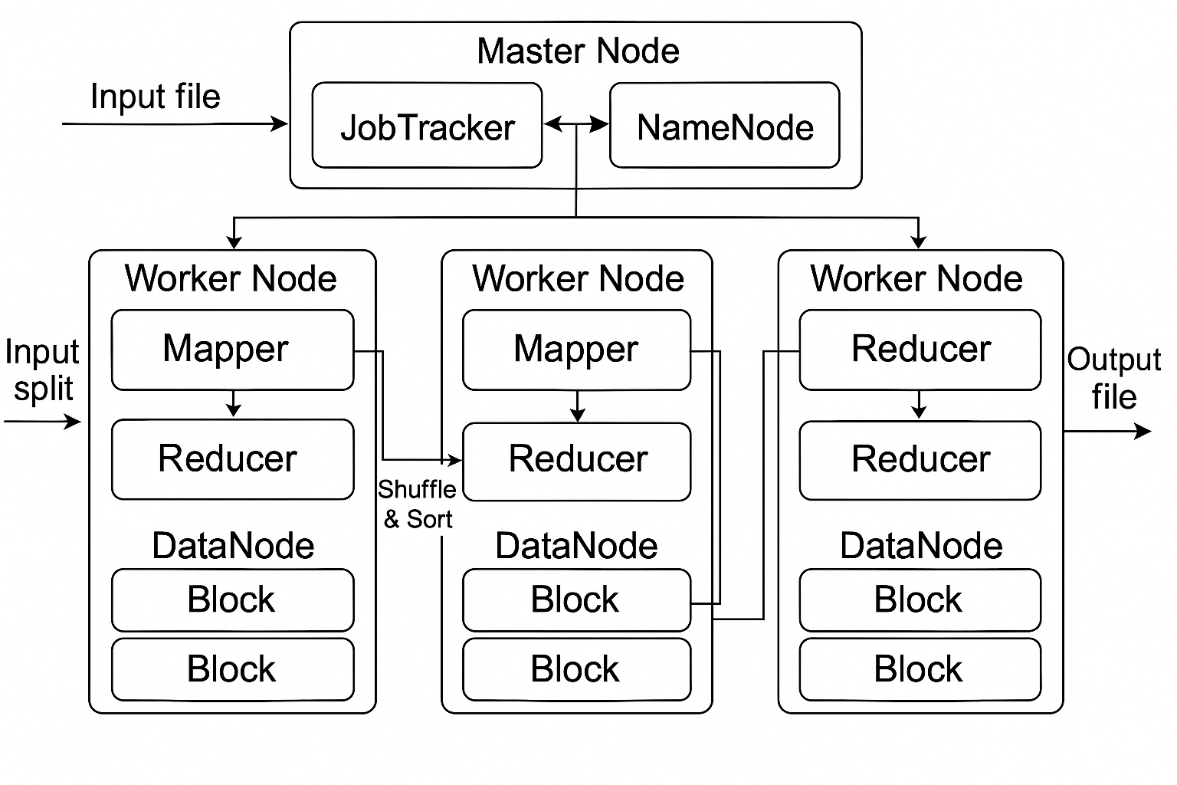
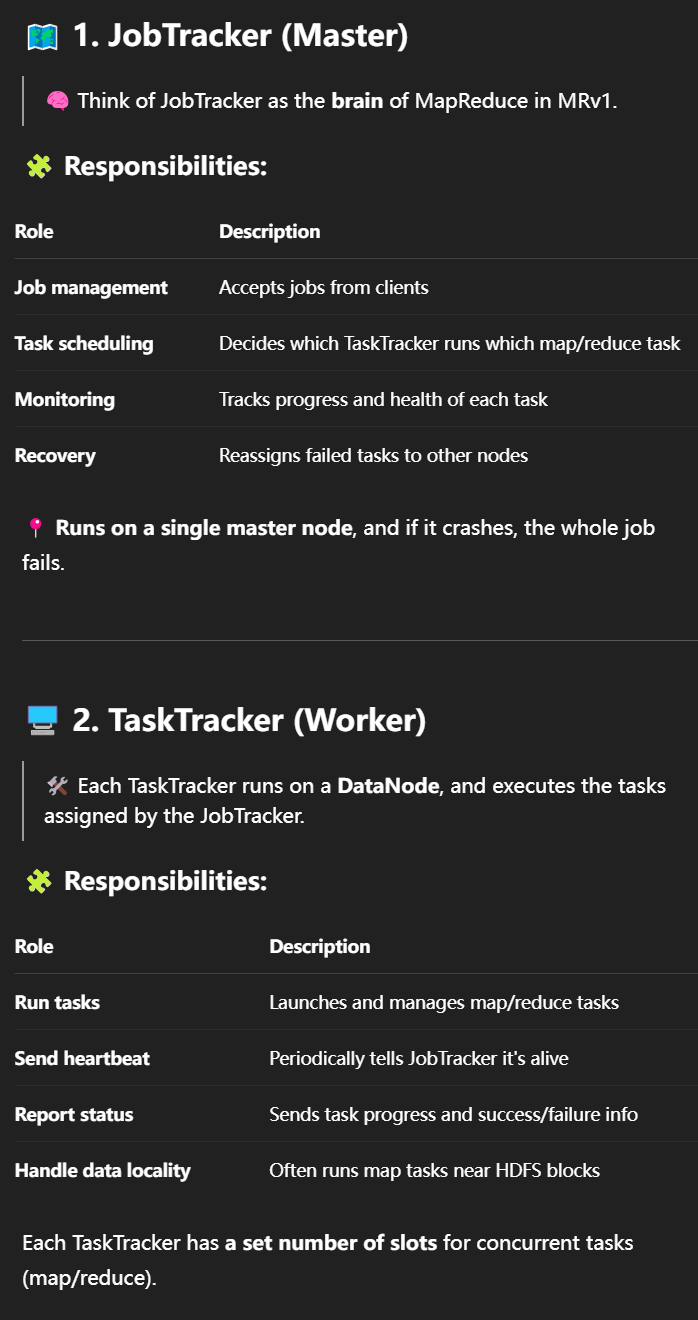
1. If too many key-value pairs (KVPs) with the same key are transferred to the same reduce node, how does Hadoop handle load balancing in such a scenario?

文本

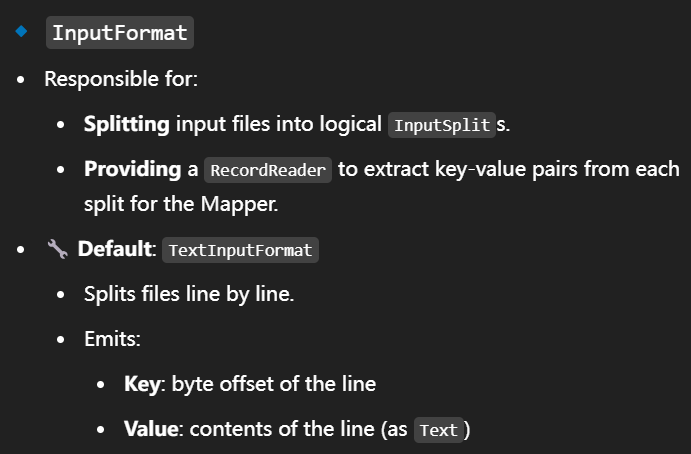
AI 生成的内容可能不正确。



1. What are the roles of JobTracker and TaskTracker in the classic Hadoop MapReduce architecture?



1. InputSplit → Mapper → Combiner (optional) → Partitioner → Shuffle & Sort → Reducer → Output
2. What are InputFormat and OutputFormat in Hadoop MapReduce, and could you provide an example to illustrate their usage?



图形用户界面, 文本, 应用程序

AI 生成的内容可能不正确。

