1. hadoop方面？
2. 关键节点放在不同机架
3. 若存在大量小文件时？set hive.merge.mapfiles=true同时可以通过手动设置split最小输入大小避免申请大量资源浪费
4. container重用的开启
5. 配置reduce copy阶段内存占用比例、堆外内存的使用
6. 数据块平衡时的限流，提高period和带宽和线程数
7. 配置hadoop写磁盘策略为可用空间策略，防止datanode的磁盘大小不一导致写满
8. 配置磁盘剩余空间，目的是为了mr任务的正常运行
9. 配置磁盘损坏个数的容忍
10. hive方面？
11. 数据倾斜的解决办法
12. 如果使用ORC格式，需要打开hive.optimize.index.filter= true，使用ORC的索引
13. spark方面？
14. mysql方面？
15. kakfa方面？