# Hive及Mapreduce优化长期总结

|  |  |  |
| --- | --- | --- |
| **标题** | **网址** | **备注** |
| Hive数据倾斜（大表join大表）【优化】 | <http://www.aboutyun.com/thread-13077-1-1.html> | 两个GB表join优化 |
| hive 优化 tips | <http://blog.csdn.net/kntao/article/details/12838197> | 多项优化 |
| Map数增加 |  |  |
| Reduce数增加 | 默认5个 |  |
| 内存增大 |  |  |
| MapJoin | 默认10M,调大值200M |  |
| Jvm重用 | 当小文件过多时，可以减少小文件的数量 |  |
| 文件压缩 | Map前lzo切片，过程中快速落盘，snnapy，计算后用gzip2，可以压缩很大空间 |  |
| 调大缓冲区 |  |  |
| 分区分桶 |  |  |
| Merge |  |  |
| MapOnly |  |  |