可怕的Full GC （转自Hbase不睡觉书）

PS：之前做项目的时候，需要做个复杂的查询，大量的查询总是导致hbase集群奔溃，最后定位到时full GC的原因。

以下转自《Hbase不睡觉书》----------

可怕的Full GC

随着内存的加大， 有一个不容忽视的问题也出现了， 那就是JVM的堆内存越大， Full GC的时间越久。 Full GC有时候可以达到好几分钟。在Full GC的时候JVM会停止响应任何的请求， 整个JVM的世界就像是停止了一样， 所以这种暂停又被叫做Stop-The-World（ STW） 。当ZooKeeper像往常一样通过心跳来检测RegionServer节点是否存活的时候， 发现已经很久没有接收到来自RegionServer的回应， 会直接把这个RegionServer标记为已经宕机。 等到这台RegionServer终于结束了Full GC后， 去查看ZooKeeper的时候会发现原来自己已经“ 被宕机” 了， 为了防止脑裂问题的发生， 它会自己停止自己。 这种场景称为RegionServer自杀， 它还有另一个美丽的名字叫朱丽叶暂停， 而且这问题还挺常见的， 早期一直困扰着HBase开发人员。 所以我们一定要设定好GC回收策略， 避免长时间的Full GC发生， 或者是尽量减小Full GC的时间。

GC回收策略优化

由于数据都是在RegionServer里面的， Master只是做一些管理操作， 所以一般内存问题都出在RegionServer上。 接下来主要用RegionServer来讲解参数配置， 如果你想调整Master的内存参数， 只需要把HBASE_REGIONSERVER_OPTS换成HBASE_MASTER_OPTS就行了。JVM提供了4种GC回收器：

- 串行回收器（ SerialGC） 。
- 并行回收器（ ParallelGC） ， 主要针对年轻带进行优化（ JDK 8默认策略） 。
- 并发回收器（ ConcMarkSweepGC， 简称CMS） ， 主要针对年老带进行优化。
- G1GC回收器， 主要针对大内存（ 32GB以上才叫大内存） 进行优化。

具体实现请参考《Hbase不睡觉书》第八章第一节。

《Hbase不睡觉书》下载 http://www.aboutyun.com/thread-25255-1-1.html

过程：

gc时间过长，超过40秒的maxSessionTimeout时间，使得zk认为regionserver已经挂掉dead；

zk返回dead region到master，master就让其他regionserver负责dead regionserver的regions；

其他regionserver会读取wal进行恢复regions，处理完的wal，会把wal文件删除；

dead regionserver的gc完成，并且恢复服务之后，找不到wal，已经产生上面截图中的报错（wal.FSHLog: Error syncing, request close of WAL）；

dead regionserver从zk得知自己dead，就关闭自己（Region server exiting，

java.lang.RuntimeException: HRegionServer Aborted）

最终原因：tickTime超时

经过上面的分析，是gc时间超过40秒的maxSessionTimeout导致的regionserver挂掉。但是，我们就很纳闷了，因为我们设置的zookeeper.session.timeout超时时间为60秒，远远超过40秒时间。非常奇怪呀！

经过hbase社区求助，以及google类似的问题，最终找到原因

详细链接，请参考：https://superuser.blog/hbase-dead-regionserver/

原来我们的HBase 并没有设置tickTime，最终hbase与zk的会话最大超时时间并不是zookeeper.session.timeout参数决定的，而是有zk的maxSessionTimeout决定。zk会根据minSessionTimeout与maxSessionTimeout两个参数重新调整最后的超时值，minSessionTimeout=2*tickTime， maxSessionTimeout=20*tickTime。我们的大数据集群，zk的tickTime设置为默认值（2000ms）2秒，因此，最终hbase 与 zk的超时时间就为40秒。

经过调整zk的tickTime为4秒，相应的zookeeper.session.timeout为80秒，最终解决regionserver 频繁挂掉的故障。

https://superuser.blog/hbase-dead-regionserver/

zookeeper.session.timeout

Default value is 90 seconds. We know that region server maintain a session with zookeeper server to remai

Minimum session timeout : 2 X tick time

Maximum session timeout : 20 x tick time

Now, no matter what session timeout you set in your client configs, if your zookeeper server timeout is