

Spark sql解析异常java.lang.StackOverflowError处理

如果你在Spark SQL中运行的SQL语句过长的话，会出现 java.lang.StackOverflowError 异常：

```
java.lang.StackOverflowError
  at org.apache.spark.sql.hive.HiveQl$$anonfun$22.apply(HiveQl.scala:924)
  at org.apache.spark.sql.hive.HiveQl$$anonfun$22.apply(HiveQl.scala:924)
  at scala.collection.TraversableLike$$anonfun$map$1.apply(TraversableLike.scala:244)
  at scala.collection.TraversableLike$$anonfun$map$1.apply(TraversableLike.scala:244)
  at scala.collection.immutable.List.foreach(List.scala:318)
  at scala.collection.TraversableLike$class.map(TraversableLike.scala:244)
  at scala.collection.AbstractTraversable.map(Traversable.scala:105)
  at org.apache.spark.sql.hive.HiveQl$.nodeToPlan(HiveQl.scala:924)
  at org.apache.spark.sql.hive.HiveQl$.nodeToPlan(HiveQl.scala:1214)
  at org.apache.spark.sql.hive.HiveQl$.nodeToRelation(HiveQl.scala:1225)
  at org.apache.spark.sql.hive.HiveQl$$anonfun$22.apply(HiveQl.scala:966)
  at org.apache.spark.sql.hive.HiveQl$$anonfun$22.apply(HiveQl.scala:924)
```

这是因为程序运行的时候 Stack 大小大于 JVM 的设置大小。我们可以通过在启动 Spark-sql 的时候加上 --driver-java-options "-Xss10m" 选项解决这个问题。

-Xss 是 JVM 为每个线程分配的堆栈大小。JDK 1.5 之前每个线程堆栈默认大小为256K；JDK 1.5 以后每个线程堆栈默认大小为1M。默认的大小设置查看可以通过运行下面的命令获取：

```
[iteblog@www.iteblog.com ~]$ java -XX:+PrintFlagsFinal -version | grep ThreadStackSize
```

```
intx CompilerThreadStackSize          = 0                {pd product}
intx ThreadStackSize                  = 1024             {pd product}
intx VMThreadStackSize                 = 1024             {pd product}
java version "1.8.0_77"
Java(TM) SE Runtime Environment (build 1.8.0_77-b03)
Java HotSpot(TM) 64-Bit Server VM (build 25.77-b03, mixed mode)
```

具体大小的设置需要根据实际的应用程序来决定，不能随便设置。

本博客文章除特别声明，全部都是原创！
原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。

本文链接: [【】 \(\)](#)