**学习日志**

姓名：李奥 学号： 1812998 日期： 6.26

**学习内容安排：**

**MapReduce特性**

**学习反馈：**

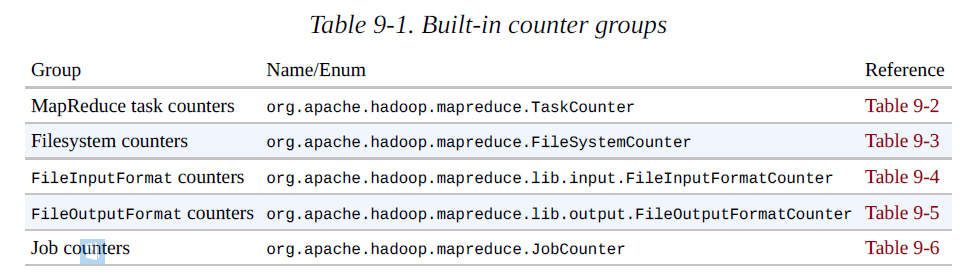
****Counters****

Counter是一个比较好的途径，用于收集job的统计信息，无论是质量控制，还是应用层面的数据统计。同时它还有助于问题的诊断。如果你很想在map task或reduce task中放入一条日志信息，比较好的做法是看能否可以用一个Counter来代替，因为，对于一个大型的分布式job来说，获取Counter的值要比获取日志的输出容易。另外，要得到一个record在某种条件下发生的次数，从日志文件中获得要做更多的工作。

****Built-in Counters****

Hadoop为每个Job维护着一些计数器，它们记录着不同方面的内容。例如，统计已处理的字节数和record数，可由此确定预期消耗的输入量和预期产生的输出量。

Counters被分成了不同的组，下表为几个常用的内置counter组：



每个组要么包括task counter（用于更新task的进度），要么包括job counters（用于更新job的进度）。

****Task counters****

Task counter用于收集task整个执行过程信息，其结果为job中所有task的汇总。例如，MAP\_INPUT\_RECORDS 计数器，用于统计每个map task读取到的record，并对job中的所有map task做汇总，所以，最后的结果为整个job输入record的数量。

Task counters还维护着每个task的尝试数，并定期的发送给application master，所以它可以做全局聚合，另外，task counter每次发送都是全量的，而不是仅发送自上次以来的统计，因为，这可以预防因消息的丢失而引起的错误，并且，在job运行期间，如果task失败了，counter可能也会停止。

Counter的值是在job成功完成时才确定的，然而，一些counter会提供一些有用的诊断信息作为一个task的进展，并且可以通过Web UI来监控它们。

****Job counters****

Job counters是由application master维护的，所以它需要通过网络发送。它们属于job级别的统计，在task运行时它们的值是不会改变的。例如，TOTAL\_LAUNCHED\_MAPS 计数器用于统计job的整个过程启动的task的数量。