

理解 MapReduce 思想

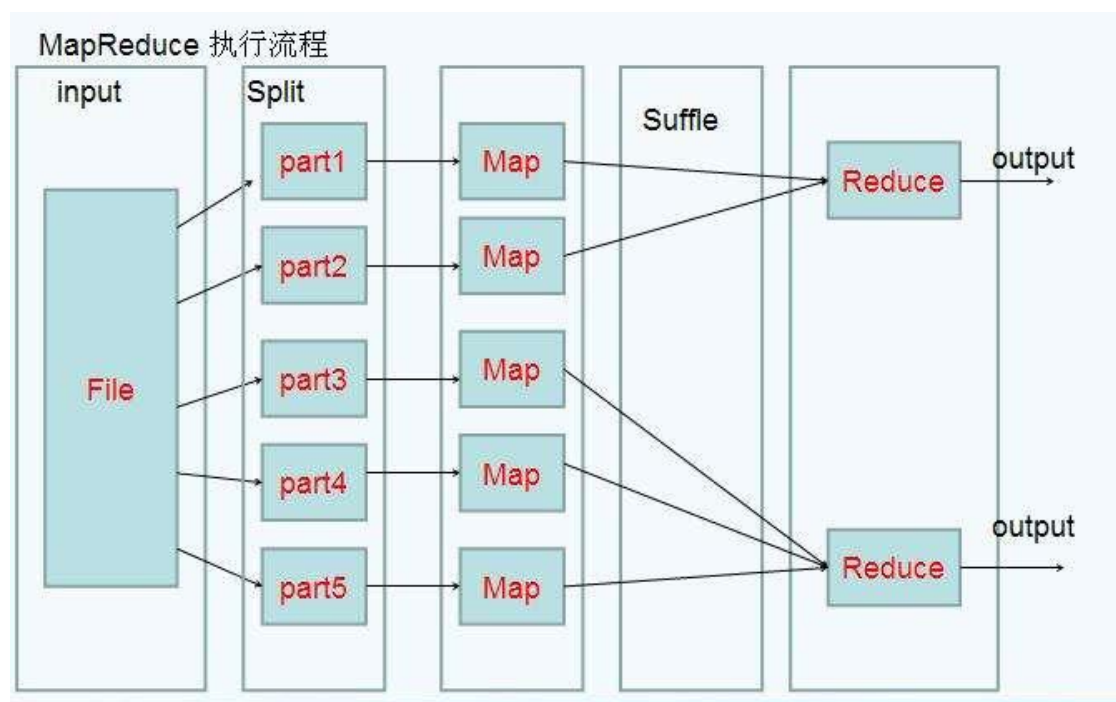
MapReduce 思想在生活中处处可见。或多或少都曾接触过这种思想。

MapReduce 的思想核心是“分而治之”，适用于大量复杂的任务处理场景（大规模数据处理场景）。即使是发布过论文实现分布式计算的谷歌也只是实现了这种思想，而不是自己原创。

Map 负责“分”，即把复杂的任务分解为若干个“简单的任务”来并行处理。可以进行拆分的前提是这些小任务可以并行计算，彼此间几乎没有依赖关系。

Reduce 负责“合”，即对 map 阶段的结果进行全局汇总。

这两个阶段合起来正是 MapReduce 思想的体现。



图：MapReduce 思想模型

还有一个比较形象的语言解释 MapReduce：

我们要数图书馆中的所有书。你数 1 号书架，我数 2 号书架。这就是“Map”。

我们人越多，数书就更快。

现在我们到一起，把所有人的统计数加在一起。这就是“**Reduce**”。