

# 简化 MapReduce 命题报告

CCSP 分区赛技术委员会 2019

# 题目概况

- 编程语言
  - C++ 17
  - Java
- 题目形式
  - 提供编程框架，选手需自行实现接口函数具体内容
- 评分方式
  - 正确性分数 30%
  - 性能分数 70% （各赛区分别评比，按与赛区最优者的性能比例给分）

# 测例说明

- 测例 1
  - 一个简单的 WordCount, 数据量很小
- 测例 2
  - 一个数据量较大的 PageRank
- 测例 3
  - 一个数据量很大、有很多额外操作的 WordCount
- 测例 4
  - 一个没有什么实际意义的测例, 考察整体的性能

# 参考优化

- 多线程
  - 评测机有 4 个逻辑处理器，正确实现多线程有望得到性能提升。
- 实现 Lazy Evaluation
  - 在对象很多时能显著减少内存使用。
- 尝试更多实现方式
  - 举例：group\_by\_key 函数中的 group 操作可以采取哈希、平衡二叉搜索树、排序等方法，其中哈希在我们设计的部分测例中能够得到较优性能。