摘要

本文通过对两个私有集群和两个高性能集群的跟踪记录研究，证明此前的研究成果过拟合于Google数据集。本文还通过使用上述四个跟踪记录和Google跟踪记录来评估作业运行时预测变量的性能，证明了数据集多样性和多样性的重要性。

数据集信息

私人作业负载跟踪记录来自来自对冲基金公司Two Sigma的两个数据中心，工作负载包括处理财务数据的数据分析作业。LANL Mustang集群是LANL用于容量计算的HPC集群，研究使用的数据集由565个用户提交并由开源集群资源管理器SLURM [45]收集的210万个多节点作业的跟踪记录组成。来自LANL Trinity超级计算机的数据集由88个用户发出的25237个多节点作业组成，并由开源集群调度系统MOAB [1]收集。来自Google的数据集是一份为期29天的跟踪记录，包含672074个工作和4,800万个任务。

工作特点

图1显示了所有跟踪记录中CPU内核的作业请求的累积分布函数（CDF），x轴为对数刻度。数据显示，平均而言，TwoSigma和LANL跟踪中的作业所请求的CPU内核比Google跟踪中的作业多3到406倍。 LANL跟踪中的作业大小更均匀地分布。

图2显示了所有跟踪的作业持续时间的CDF。数据显示，Google跟踪中的中位作业比LANL或TwoSigma跟踪中的中位作业短4-5倍。但是，Google跟踪中最长的1％作业比LANL和TwoSigma跟踪中相同部分的作业长2-6倍。

工作负载异构

在图3中，显示了一天中每小时的作业调度请求数。数据显示，日间模式是普遍的。集群在白天收到更多的调度请求和较小的工作，而Google跟踪的偏差较小。

图3还显示出在集群之间调度请求的速率可能有很大差异，调度请求率在整个集群中相差最多3个数量级。为了跟上工作量，亚秒级的调度决策是必要的。

资源利用

研究发现，与Google集群不同，其他集群没有出现超额使用资源的情况。

图5显示了作业到达间隔长度的CDF，图6显示了每个跨组织作业的任务数量的CDF。数据显示，大多数的工作间隔时间是亚秒级以下。

分析发现，在LANL和TwoSigma跟踪记录中，用户资源请求比在Google跟踪记录中更具有多样性。

失败分析

图7按作业结果细分了作业总数以及所有作业消耗的CPU总时间。结果显示，Google跟踪中未成功终止的工作比其他跟踪中高1.4-6.8倍。 LANL上不成功的作业使用的CPU时间减少了34-80％。

图8显示了单个作业的作业大小的CDF（在CPU内核中）。数据显示，对于Google跟踪记录，不成功的工作往往比成功的工作需要更多的资源，而所有其他跟踪记录并不具备此特点。

图9根据作业消耗的CPU时间（大小和长度的度量）对作业进行了分组，并显示了每个组的成功率。结果显示，对于Google和TwoSigma跟踪，成功率下降的根源是占用更多CPU时间的作业，LANL跟踪则相反。

关于多元化和多样性的案例研究

这里通过对JamaisVu [51]集群调度程序的job runtime2预测程序模块JVuPredict进行案例研究，证明针对多样的跟踪记录的研究有助于确定新研究的实际敏感性并证明其普遍性。

JamaisVu [51]的运行时预测模块JVuPredict旨在使用有关过去作业特征和运行时的历史数据来预测作业提交时的运行时。与传统方法不同，它尝试检测重复的作业，即使连续运行未声明为重复也是如此。

图10显示了研究的评估结果。x轴绘制了JVuPredict的运行时估算值的预测误差，以作业实际运行时的百分比表示；y轴显示其预测属于每个存储区的作业百分比。结果表明，尽管逻辑工作名称方法在Google跟踪中很好用，但是很难以匿名形式为其他跟踪生成它们，因此它们通常不可用；在没有工作名称的情况下，除了OpenTrinity跟踪以外，还有其他字段可以替代它们并提供可比的准确性。

跟踪记录的时间跨度的重要性

图11通过将每个月不同指标的平均值与整个记录的总体平均值进行比较，来总结研究的结果。结果表明，集群工作负载每个月都显示出显著差异。

结论

本文研究和发布了来自两个私有集群和两个HPC集群的四个跟踪记录数据集。分析表明，私有集群的工作负载类似于所研究的HPC工作负载，而不是流行的Google跟踪工作负载。本文还通过案例分析证明了数据集的多样性和多样性在评估新研究中的重要性。