**Selecta: Heterogeneous Cloud Storage Configuration for Data Analytics**

（Selecta：一种用于数据分析的异构云存储配置）

Ana Klimovic Stanford University，Heiner Litz UC Santa Cruz，Christos Kozyrakis Stanford University

姓名：许佳豪 学号：M201873224 班级：M1807 论文号：54

# 摘要

数据分析是公共云服务上一类重要的数据密集型工作负载。但是，为这些应用选择合适的计算与存储配置十分困难，因为可以选择的范围很广，每个选项之间交互也十分复杂。除此之外，不同类型的存储设备可以更好的服务工作负载中不同的数据流。

本文提出了一种Selecta工具，这是一种为数据分析工作推荐最佳云计算和存储资源配置的工具。 Selecta基于过往收集的系数数据，使用潜在因子协作过滤来预测应用程序如何在不同的配置下执行。我们用超过一百个Spark SQL和ML应用程序评估Selecta，表明Selecta选择了接近最优的性能配置(波动控制在10%之内)的概率为94％，成本最优的概率为80％。我们还使用Selecta来绘制包括NVMe闪存设备的性能-成本效率，云存储的需求以及对端到端存储优化的支持。

1. 领域进展

1.1研究现状

由于大量公司将其工作负载 转移到Amazon AWS, Google Cloud Platform以及Microsoft Azure等平台上导致了公共云市场正在经历前所未有的增长。公共云除了提供高弹性以外，还要提供较低的成本以及租户之间的资源共享。但是，要实现性能和成本效率，需要为每个给定的应用选择合适的配置。但是大量的实例类型和配置选项使得为应用程序选择合适的资源很困难。

存储的选择通常是必不可少的，特别是对于数据密集分析的云部署。云供应商提供各种各样的存储选项，包括对象、文件和块存储。块存储可以包括硬盘（HDD），固态驱动器（SSD）或高带宽，低延迟NVMe闪存设备（NVMe）。设备可以是运行应用程序或远程（r）的云实例的本地（l）。仅这些选项就导致存储配置选项在吞吐量，延迟和每比特成本方面可能有几个数量级的差异。

选择正确的云存储配置对性能和成本都至关重要。

我们工作的一个关键贡献是我们对云存储系统的分析以及分析工作负载的使用，这些工作分析了重要的一些重要内容。我们除了提供最佳性能外，基于NVMe的配置还可为各种应用提供低执行成本。我们观察到云存储选项的需求，这些选项支持容量和带宽的细粒度分配，类似于无服务器云服务提供的计算和内存资源的细粒度分配[7]。分解的NVMe Flash可以提供基础设施，可以提供灵活的存储功能。最后，我们展示了对云存储的端到端优化的需求，包括应用程序框架，操作系统和云服务，因为由于存储堆栈的低效，多个存储配置无法满足其潜力。

1.2主要挑战

A.性能成本目标

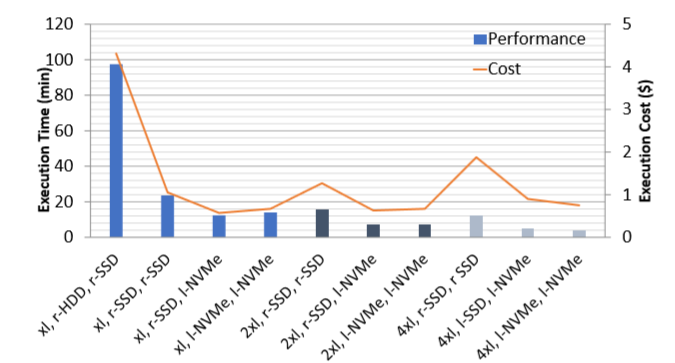
虽然为应用程序配置尽可能多的CPU内核、内存，会使得存储速度得到很大的提升，但是这样的资源配置会导致成本的大量提高。如何使用较低的成本提供一个具高的性能存储系统（例如，预测最大化性能或最小化成本的配置）是一个巨大的挑战。

B.异构的应用程序数据

确定分析应用程序的正确云配置具有挑战性。即使我们将自己局限于单一实例类型并专注于优化性能，对特定应用程序的存储配置的选择仍然是非常重要的。我们需要分析工作负载访问的多个数据流，包括输入和输出文件，日志和中间数据（shuffle和广播）。这些数据流通常具有不同的数据特征：访问频率、数据生命周期、访问类型（随机与顺序）和I / O大小。例如，输入/输出数据通常是长期的并且顺序访问，而中间数据是短暂的并且大多数访问是随机的。这些特征使得不同的数据流需要不同类型的存储设备。

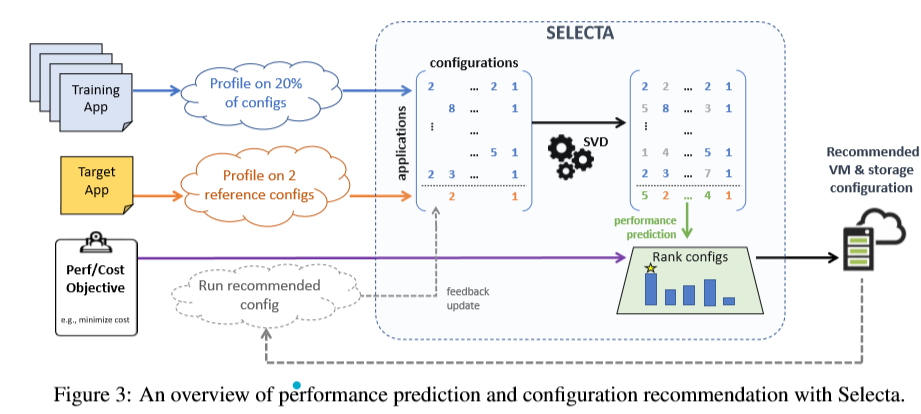
C.存储决策复杂.

为作业选择正确的配置可显着减少执行时间和成本，如图2所示，它比较了不同配置下8个节点的执行时间以及运行成本。可以看出成本差异最高相差7.5倍

****

1. Selecta方法与设计

2.1模型概述



Selecta作为一种可以自动预测目标应用程序在一组候选配置上的性能的工具。其工作流程如图3所示。

Selecta的输入为：

1. 一组训练应用程序在几个配置上的执行时间；
2. 目标应用程序参考配置的执行时间
3. 目标应用程序的性能成本目标。

配置由节点数（VM实例），CPU核心和每个节点的内存，以及用于输入/输出数据和中间数据的存储类型和容量来定义。 Selecta使用潜在因子协作过滤来预测目标应用程序在剩余（非参考）候选配置上的性能。通过这些性能预测和各种VM实例和存储选项的单位时间成本，Selecta可以为用户的性能成本目标推荐正确的配置。例如，Selecta可以推荐最小化执行时间，最小化成本或最小化特定预算内执行时间的配置。

随着新应用程序的不断启动，之前测试过的应用程序性能逐渐成为训练集的一部分，提高Selecta的准确性。我们还在Selecta推荐的配置上运行目标应用程序后反馈性能测量， 这有助于降低测量噪声并提高准确性。

2.2预测性能

A.预测方法

Selecta使用协作过滤来预测目标应用程序在候选配置上的性能。选择协作过滤，因为它与所使用的数据分析框架的细节无关（例如Spark与Storm），它允许利用跨应用程序和配置收集的稀疏训练数据。

Selecta的目标是利用多个应用程序提供的训练数据，只需两次目标运行即可收敛到准确的推荐应用。我们讨论协作过滤的替代方案来解释我们的选择。

基于内容的方法，例如线性回归，随机森林和神经网络模型，根据应用特征（例如，数据读取/写入数据的GB）和配置特性（例如，I / O带宽或特征）等特征构建模型。每个VM的核心数量）。我们发现除非输入功能（如目标应用程序在目标配置上的平均CPU利用率）在模型中使用，否则基于内容的预测器没有足够的信息来学习应用程序的计算和I / O要求并实现低精度。需要在所有候选配置上运行目标应用程序来收集特征数据的方法是不切实际的。

另一种方法是基于分析框架的结构构建性能预测模型，例如地图的特定Spark中的减少阶段。这导致框架特定模型，并且可能需要在框架实现发展时进行重新调整甚至重新构建（例如，随着CPU的序列化操作的改进而提高）。

B.性能-成本目标定义

在预测所有配置的应用程序性能之后，Selecta建议基于用户定义的排名功能进行配置。例如，为了最小化运行时成本，排名函数为min（运行时间×成本/小时）。在选择存储技术（SSD与NVMe Flash）时，Selecta还必须考虑应用程序的存储容量要求。 Selecta利用统计数据来改变闪存监控日志，以确定中间（shuffle）数据和输入/输出数据容量。

C.适应变化

重复的工作及其输入数据集可能会发生变化。为了检测可能影响最佳配置选择的应用程序特征的变化，Selecta依赖于来自应用程序实例和后续执行轮次的CPU利用率信息。当应用程序首次引入系统时，Selecta会为存储应用程序指定一个唯一的ID来识别信息的使用情况。每当重新执行一个应用程序时，Selecta会将当前的iowait时间与存储的配置进行比较。根据iowait时间的不同，Selecta将根据可用的测量结果进行计算，或者将工作负荷视为新的应用程序，从而启动新的计算机运行。

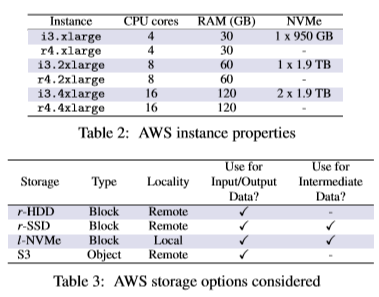
D.云中噪声处理

推荐最佳配置的另一个挑战是公共云平台上的噪声，这是由于与其他人，硬件异构或其他来源的干扰而产生的。为了解决噪声问题，Selecta依赖于性能和CPU利用率测量的反馈。最初，由于噪声影响，Selecta的性能预测会受到噪声的影响。随着更多测量结果被输入系统，Selecta平均了性能和CPU利用率，并使用储层采样来避免异常值的高度偏斜。 Selecta为应用程序配置矩阵（例如，三个）中的每个条目保留可配置数量的样本点，以检测如上所述的应用程序中的变化。如果特定运行受到噪声的严重影响，使得计算和I / O瓶颈与以前的运行显着不同，则Selecta检测应用程序变化的机制会识别异常值。

1. 模型的选择

3.1实现原理

我们在Amazon EC2上部署Selecta，并考虑表2和表3中所示的实例和存储选项的配置。



A.实验方法：

我们在每个应用程序上运行每个应用程序，以确定每个应用程序的地面真实性能和最佳配置选择。为了解释云中的噪声，我们运行每个实验（即每个候选配置的每个应用程序）三次，并使用不等于运行的平均值。两次运行是连续的，一次运行是在一天中的不同时间。我们还通过使用来自一次运行的数据作为Selecta的输入以及跨运行的平均性能作为基本事实来验证我们的结果。为了训练和测试Selecta，我们使用留一法交叉验证[58]，这意味着一次一个工作负载作为目标应用程序，而剩余工作负载用于培训。我们假设所有候选配置都有训练应用，除了§4.4中的灵敏度分析，我们调查训练矩阵密度要求以进行准确预测

B.度量标准：

我们使用两个指标来衡量Selecta预测的质量。首先，我们报告相对均方根误差（RMSE），这是推荐系统的通用指标。 Selecta的第二个更相关的指标是制定准确配置建议的概率。如果配置在该应用的真实最佳配置的阈值T内满足用户的成本 - 性能目标，则我们认为推荐是准确的。例如，对于T = 10％的最低成本目标，准确预测的概率是Selecta建议（在所有测试应用中）的百分比，其真实成本在真实最优成本配置的10％之内。使用阈值对噪声更稳健，并且允许我们对Selecta的准确性做出更有意义的结论，因为第二最佳配置可能具有与最佳配置相似或显着更差的性能。我们的绩效指标是执行时间，成本以美元计算。

3.2预测准确度

我们提供了一个204行的矩阵作为Selecta的输入，其中一行（应用程序）被指定为每个测试轮次中的目标应用程序。我们运行Selecta 204次，每次都考虑使用不同的应用程序作为目标。目前，我们假设矩阵中所有剩余的训练数据行都是密集的，这意味着用户在所有候选配置上都有专业的训练应用程序。单个目标应用程序行是稀疏的，仅包含两个条目，其中一个是参考配置的预防措施。

Selecta预测性能相对于应用程序的平均相对RMSE为36％。为了理解Selecta的性能预测如何转化为建议，我们在图4中绘制了性能，成本和成本\*性能目标的准确度。该图显示了作为阈值T的函数的接近最优推荐的概率，其定义了从最佳值开始的足够近的百分比。在搜索性能最佳的配置时，Selecta的最大概率为94％，其中包含最佳的10％的可信度。对于最低成本目标，Selecta有80％的可能性在最佳值的10％范围内推荐配置。预测成本\*性能更具挑战性，因为Selecta对候选配置中的应用程序的相对执行时间预测是平方的：成本\*性能=（执行时间）2 \*确定每小时成本。

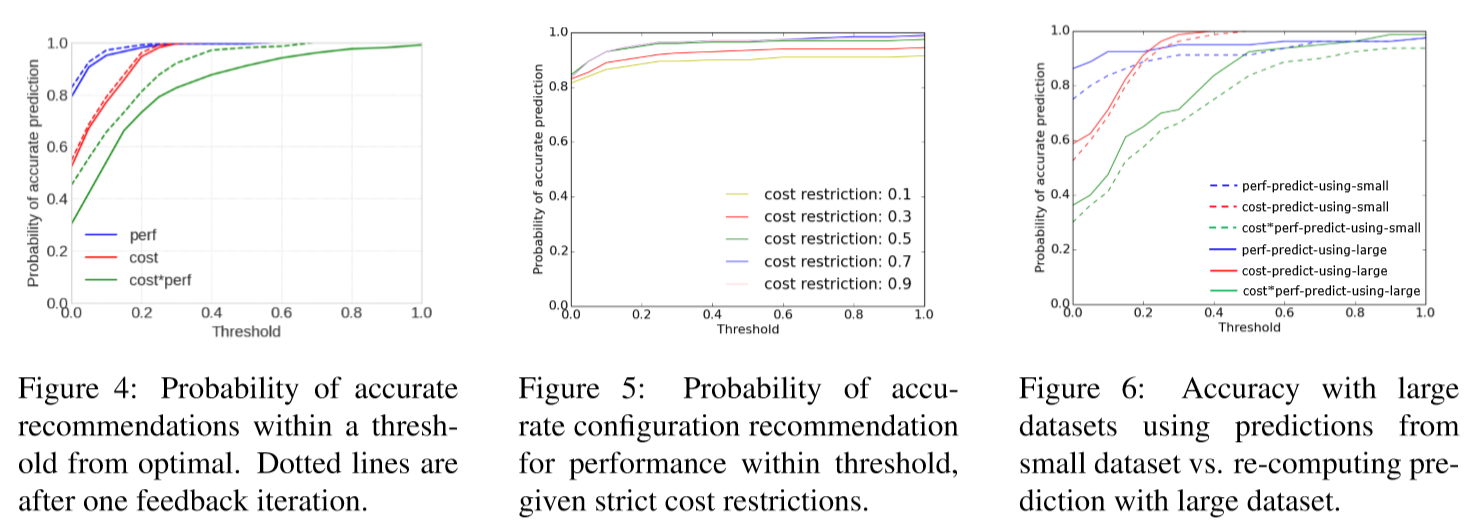
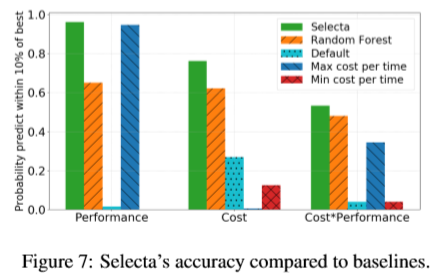


图4中的虚线显示了单次反馈回合后精度的提高情况。在这里，我们假设目标应用程序在反馈回合中具有相同的数据集。这为目标应用程序行提供了额外的培训输入（如果推荐的配置不是参考配置，则为新条目;如果推荐的配置为参考配置，则为新现有数据的平均值）。对于成本\*性能目标，近似最优推荐的概率最显着地增加，从反馈后的52％到65％，T = 10％。

图5显示了对于“选择具有固定成本限制C的最佳性能配置”形式的目标的准确建议的可能性。对于此目标，如果Selecta的成本小于或等于预算并且其性能在阈值内，我们认为Selecta的建议是准确的。目标的真正最佳配置。假设T = 10％，Selecta在图5中的成本限制达到了83％到94％的准确度。长尾是由于性能预测错误导致选择性地低估了小配置百分比的执行成本（即，Selecta建议实际超出预算的配置的情况）。

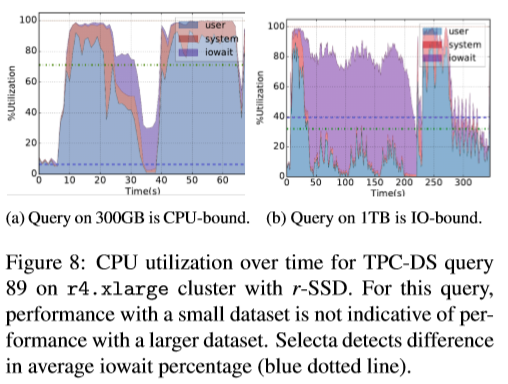


在图7中，我们将Selecta的准确度与四个基线进行比较。第一个基线是随机森林预测器，类似于PARIS [71]使用的方法。我们使用以下功能：配置提供的CPU核心数，磁盘IOPS和磁盘MB / s，应用程序的中间和输入/输出数据容量，以及运行时测量的CPU利用率，性能和总磁盘吞吐量应用于两个参考配置中的每一个。虽然随机森林预测器比Selecta具有更多功能，但它的准确性较低。协作过滤对于训练数据的稀疏性质来说是更好的选择。我们发现随机森林模型中最重要的特征都与I / O有关（例如，在参考配置上运行应用程序时测量的I / O吞吐量和用于中间数据的存储支持的读/写IOPS） ，这强调了选择正确存储的重要性。

图7中的第二个基线（标记为“默认”）使用Databricks工程博客文章中记录的推荐默认配置：l-NVMe用于中间数据，S3用于输入/输出数据[19,21,20]。 “每次最大成本”基准使用简单的启发式方法，即每单位时间总是选择最昂贵的实例。 “每次最低成本”基准选择每单位时间最便宜的实例。 Selecta优于所有这些启发式策略，确认需要一种工具来自动化配置选择。

3.3数据集

我们使用§3.1中描述的小型和大型数据集比例研究数据集大小对应用程序性能和Selecta预测的影响。我们使用所有102个具有小数据集的工作负载来训练Selecta，然后使用大量的数据集来评估Selecta的预测准确性。如图6所示，当为具有大型数据集的应用程序推荐配置文件时，该数据集的精确度仅基于具有较小数据集的应用程序的专业运行。当使用大型数据集进行预测时，选择性应用程序具有良好的准确性。对于大约8％的应用，具有小数据集的专业人员不具备与大型数据集相关的有效指标。



我们发现，如果使用小型数据集的性能不能指示大型数据集的性能，则计算和应用程序的I / O强度之间的关系会受到影响。如§3.3所述，Selecta通过比较小型和大型数据集运行的CPU利用率统计信息来检测这些情况。图8显示了小数据集性能并不表示具有更大数据集的性能的示例性工作负载。我们使用英特尔性能分析工具记录和绘制CPU利用率[34]。当热力学变换的平均爱荷华百分比基本上与大型和小型非传导性配置相关时，通常最好将该应用分配给参考配置和通用性应用。

3.4灵敏度分析

我们执行灵敏度分析以确定准确预测的输入矩阵密度要求。我们同时研究矩阵行的密度（即，训练应用的候选配置百分比）和矩阵列的密度（即所用训练应用的数量）。我们还讨论了对参考配置的选择的敏感性。

实验结果显示，每个应用的配置选择与其他应用相关性不大。

1. 总结展望

云服务上大量且不断增加的存储和计算选项使得配置数据分析集群的性能和成本效率都很高。我们介绍了Selecta，这是一种基于跨应用程序和候选配置收集的稀疏训练数据来学习计算和存储资源的近乎最佳配置的工具。 Selecta仅需要两次目标应用程序的精确运行，就可以预测接近最优的性能配置，成功率为94％，近似最优成本配置概率为80％。

此外，Selecta允许我们分析用于数据分析的云存储选项，并揭示：包括NVMe闪存存储的成本优势，云中存储容量和带宽的最佳增益分配需求以及跨层的需求存储优化。

随着数据密集型工作负载的复杂性和计算和存储的云选项的增加，像Selecta这样的工具对于最终用户，系统研究人员和安全云提供商（例如，用于调度“无服务器”应用程序代码）将变得越来越有用。

1. 本文不足

本文提出的Selecta工具过于理想化，不具有很高的普适性，虽然理念非常新颖，但是缺乏理论以及数据公式的支撑。