

---

# WEEKLY REPORT

---

May 28, 2022

Xin Wen  
Nanchang University  
Information Engineering School  
Generic Operational and Optimal Data Group

# Contents

1	Highlights of the Week . . . . .	2
2	Study contents . . . . .	2
2.1	目标 . . . . .	2
2.2	文献调研 . . . . .	2
2.3	计划 . . . . .	3

## 1 Highlights of the Week

1. 重新思考科研问题，开始在 ATC, SC, TCC 上调研服务切换相关文献。
2. 练习 CUDA, HIP 编程。
3. 维修服务器、更新实验室网站。

## 2 Study contents

### 2.1 目标

我们的 idea 是整合异构 \*aaS 之间的差异性因素，针对一种大规模持续性负载实现一个快速的服务切换接口或系统，从而实现混合 \*aaS 对该负载的优化，可能主要体现在标准化 Ops/s/\$、p99 延迟等指标上。

### 2.2 文献调研

目前最有参考价值的文献是发表在 ATC'20 的 OptimusCloud [1]，这项工作实现了一个在线重新配置系统，在云上综合考虑了分布式数据库的不同配置和虚拟机实例的不同配置，使用马尔可夫链预测模型和一些成熟的预测模型实现工作负载预测，使用 D-optimal [2] 优化离线数据收集过程，最终实现了随着工作负载行为的变化而动态调整数据库和虚拟机实例配置，相比之前的 CherryPick [3], Selecta [4], SOPHIA [5] 提供了更好的性能（标准化 Ops/s/\$、p99 延迟），并接近理想最佳离线工作负载情形。

该项工作 [1] 主要优化了分布式数据服务的大型查询负载性能，共使用了三种大型工作负载作为基准，包括：

1. MG-RAST [6]。全球规模的宏基因组学门户网站，是同类中最大的，它允许许多用户同时将他们的宏基因组数据上传到存储库，应用计算密集型流程的管道，并可选地将结果提交回存储库供共享使用。其工作量没有任何明显的每日或每周模式，因为请求来自全球各地，其工作量可能在几分钟内发生巨大变化。总共分析了 80 天的真实查询跟踪，60 天用于训练，20 天用于测试。
2. Bus-Tracking [7]。其中读取、扫描、插入和更新操作被提交到后端数据库。该数据具有很强的每日和每周模式。此工作负载的切换频率较低，中位数为 60 次/天。对于此工作负载，分析了应用程序 60 天的实际查询跟踪 (40 天用于训练，20 天用于测试)。

3. 数据分析作业队列 [8]。根据提交给真实的微软 Cosmos HPC 计算集群的数据分析请求对作业进行建模。每项工作都分为几个阶段，工作的工作量特征会随着每个阶段的变化而变化。工作规模是一个随机变量  $U(200, 100K)$  次运算。工作负载切换的中位数为 780/天，并发级别为 10 个作业。通过观察队列中的作业并利用大部分作业模式重复出现的事实，可以在较长的时间范围内 (大约一个小时) 预测工作负载。

虽然这项工作 [1] 成功实现了针对大规模持续性分布式数据库查询负载的优化，但与我们讨论的混合 \*aaS 还是存在差距。毕竟我们的轻量级 FaaS 具有无状态属性，难以支持批量数据库查询，这也是一个问题。其次，该方案使用的预测模型虽然有类似实验室 TCC'22 BW-Raft 的间隔时间段  $T$  进行一次计算和预测的过程，但是在切换配置上存在一定问题，这是因为数据库的配置或虚拟机实例配置一旦要修改，其重新分配资源以及数据库节点更换、重启的成本非常高，论文中提及其更改虚拟机实例配置需要 90 秒，更改 Cassandra 配置则需要 30 秒，这与我们 idea 的关键点“快速切换”相悖。

## 2.3 计划

经过一些不算充实的调研，我的实现计划 (to do list) 如下：

1. 核心观点还是什么样的服务支持在线快速切换 (秒级或毫秒级)，这需要进一步的调研，或许我需要阅读一些涉及到大型 FaaS 工作负载相关工作。
2. 对于混合 \*aaS 配置空间的建模，这需要大量的测试，以确定某些性能上的优劣比较。一个初步的构想还是使用无成本的开源 FaaS 平台，我的选择是轻量级的 Fn project、重量级的 OpenWhisk 以及基于 K8s 的 OpenFaaS。
3. 在预测模型上，我还是坚持 MAB 的在线预测。而 MDP 在大多数方案中有效，我或许可以验证不同预测模型的性能和其时间复杂度上限。一些文献中的离线学习方案似乎并不生效。

# REFERENCE

- [1] Ashraf Mahgoub, Alexander Michaelson Medoff, Rakesh Kumar, Subrata Mitra, Ana Klimovic, Somali Chaterji, and Saurabh Bagchi. {OPTIMUSCLOUD}: Heterogeneous configuration optimization for distributed databases in the cloud. In *2020 USENIX Annual Technical Conference (USENIX ATC 20)*, pages 189–203, 2020. 2.2, 2.2
- [2] Of standards, n. i., and (nist), t. d-optimal designs. 2.2
- [3] Omid Alipourfard, Hongqiang Harry Liu, Jianshu Chen, Shivaram Venkataraman, Minlan Yu, and Ming Zhang. {CherryPick}: Adaptively unearthing the best cloud configurations for big data analytics. In *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17)*, pages 469–482, 2017. 2.2
- [4] Ana Klimovic, Heiner Litz, and Christos Kozyrakis. Selecta: Heterogeneous cloud storage configuration for data analytics. In *2018 USENIX Annual Technical Conference (USENIX ATC 18)*, pages 759–773, 2018. 2.2
- [5] Ashraf Mahgoub, Paul Wood, Alexander Medoff, Subrata Mitra, Folker Meyer, Somali Chaterji, and Saurabh Bagchi. {SOPHIA}: Online reconfiguration of clustered {NoSQL} databases for {Time-Varying} workloads. In *2019 USENIX Annual Technical Conference (USENIX ATC 19)*, pages 223–240, 2019. 2.2
- [6] Argonne national laboratory. 1
- [7] Lin Ma, Dana Van Aken, Ahmed Hefny, Gustavo Mezerhane, Andrew Pavlo, and Geoffrey J Gordon. Query-based workload forecasting for self-driving database management systems. In *Proceedings of the 2018 International Conference on Management of Data*, pages 631–645, 2018. 2
- [8] Andrew D Ferguson, Peter Bodik, Srikanth Kandula, Eric Boutin, and Rodrigo Fonseca. Jockey: guaranteed job latency in data parallel clusters. In *Proceedings of the 7th ACM european conference on Computer Systems*, pages 99–112, 2012. 3