Deploying a Steered Query Optimizer in Production at Microsoft

Wangda Zhang, Matteo Interlandi, Paul Mineiro, Shi Qiao, Nasim Ghazanfari
Karlen Lie, Marc Friedman, Rafah Hosn, Hiren Patel, Alekh Jindal

Microsoft
qo-advisor@microsoft.com

ABSTRACT

Modern analytical workloads are highly heterogeneous and massively complex, making generic query optimizers untenable for many customers and scenarios. As a result, it is important to specialize these optimizers to instances of the workloads. In this paper, we continue a recent line of work in steering a query optimizer towards better plans for a given workload, and make major strides in pushing previous research ideas to production deployment. Along the way we solve several operational challenges including, making steering actions more manageable, keeping the costs of steering within budget, and avoiding unexpected performance regressions in production. Our resulting system, `QO-Advisor`, essentially externalizes the query planner to a massive offline pipeline for better exploration and specialization. We discuss various aspects of our design and show detailed results over production SCOPE workloads at Microsoft, where the system is currently enabled by default.

CCS CONCEPTS

• **Information systems** → **Query optimization**.

KEYWORDS

query optimization, SCOPE, machine learning, contextual bandit

ACM Reference Format:
Wangda Zhang, Matteo Interlandi, Paul Mineiro, Shi Qiao, Nasim Ghazanfari, Karlen Lie, Marc Friedman, Rafah Hosn, Hiren Patel, Alekh Jindal. 2022. Deploying a Steered Query Optimizer in Production at Microsoft. In *Proceedings of the 2022 International Conference on Management of Data (SIGMOD '22), June 12–17, 2022, Philadelphia, PA, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3514221.3526052

1 INTRODUCTION

General purpose query optimization is hitting the end of the road in modern cloud-based data processing systems. These systems witness a wide variety of highly complex applications that are very hard to optimize globally [18], and yet these systems come with generic query optimizers that were built for all users, scenarios, and scales. Or worse, they are tuned for generic benchmarks like TPC-H or TPC-DS that often do not represent the real customer workloads [35]. As a result, these general purpose query optimizers are often far from optimal in their plan choices for a given customer and a given workload [25]. Therefore, there is a need to *specialize* the cloud-based query optimizers to the needs of specific users and applications at hand, and thus optimizing the workloads more effectively, also sometimes referred to as *instance optimization* [20].

Recent work on instance optimization have proposed to use machine learning to learn from a given user workload different components of a query optimizer that indirectly lead to better query plan choices. More ambitiously, Neo [27] proposed to replace the entire query optimizer with a learned one and producing the query plans directly. Given that replacing an entire query optimizer is not possible in a real system, the follow-up work, BAO [26], proposed to steer the optimizer towards better plan choices by providing rule hints to navigate the search space better for each query. While BAO considered synthetic workloads on PostgreSQL, our previous work [29] grounded the ideas in BAO to real-world workloads in industry strength cloud data processing system, SCOPE [42], illustrating both the challenges and the opportunities in an enterprise workload setting. In this paper, we continue this line of work and take the above ideas all the way to production deployment at Microsoft. Along the way, we address three sets of challenges to make such a query optimizer operational, as discussed below.

First, prior approaches to steering the query optimizer consider multiple rule hints (as many as 250 rules in the SCOPE query optimizer) at the same time to help the optimizer navigating the search space. The problem, however, is that it is hard to understand or explain what really was wrong and what helped in making the better choice. Thus, neither the system developers can learn about the quality of their query optimizer code bases, nor the service operators explain the impact of the change in case of performance regression, nor the users can build confidence on the robustness of the steering approach. Second, the experimentation costs involved in prior steering approach were non-trivial. This is because of random sampling in the space of steering hints which when executed in a large scale processing system like SCOPE leads to significant pre-production costs, making it difficult to maintain over time. And third, steering the query optimizer to newer unseen points in the search space could lead to performance regression, for newer queries and over time. This is because the estimated query costs do not necessarily lead to better plans due to inaccurate cost models [29]. These performance regressions are not acceptable for critical workloads in a running system since users expect a consistent system behavior with predictable performance.

We present `QO-Advisor` to overcome the above deployment challenges and to steer the optimizer in an explainable, cost-effective, and safe manner. Our key ideas include breaking down the steering process into smaller incremental steps that are easily explainable and reversible. We introduce a novel model pipeline configuration

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SIGMOD '22, June 12–17, 2022, Philadelphia, PA, USA
© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9249-5/22/06…$15.00
https://doi.org/10.1145/3514221.3526052

where a contextual bandit model is used to significantly cut down the pre-production experimentation costs, followed by a cloud variability model determining performance improvement or regression. Essentially, `QO-Advisor` opens up the core of a query optimizer, i.e., the query planner, which was historically seen as a black box, and externalizes it for better specialization over a customer workload. Consequently, to the best of our knowledge, `QO-Advisor` is the first production deployment of steering a query optimizer.

Contributions and Organization. We make the following major contributions in this paper.

(1) We present a background to the line of work on steering query optimizer, discussing the current state of the art, the productization challenges, and our design principles for `QO-Advisor`. (Section 2)
(2) We describe the learning principles that we implemented in `QO-Advisor` to overcome the productization challenges. These include using a contextual bandit model to recommend, and a validation model used to accept or reject suggested modification to the query plans. (Section 3)
(3) We discuss several operationalization aspects, including how we integrate with the Azure Personalizer [1] service, how we deal with plans having non-tree format, and how we featurize such plans. (Section 4)
(4) We show experimental results from production workloads and demonstrate the overall impact that customers can expect, as well as present an in depth analysis of the factors contributing to it. (Section 5)
(5) We discuss the lessons we learned over time as we implemented the ideas firstly presented in BAO from a research prototype to a production-ready system. (Section 6)

2 STEERING QO: ONE YEAR LATER

In this Section, we describe our journey in the steering query optimizer research over the last year. We first give some background on both the problem that `QO-Advisor` tries to address, and the relevant system details from the SCOPE processing engine at Microsoft (Section 2.1). Then, we discuss the various challenges we encountered while taking our initial steering optimizer implementation to production deployment (Section 2.2). Finally, we introduce `QO-Advisor` to overcome these challenges (Section 2.3), discuss the design principles we followed when implementing the `QO-Advisor` (Section 2.4), and provide an overview of the end-to-end pipeline involved in the steering process (Section 2.5).

2.1 Background

We start with a quick summary of the SCOPE [33] data processing system at Microsoft, followed by a brief overview of the prior attempt to apply steering optimizer ideas to SCOPE [29]. `QO-Advisor` builds upon that work and extends it for production readiness.

SCOPE. SCOPE is a large scale distributed data processing system [8]. It powers production workloads from a wide range of Microsoft products, processing petabytes of data on hundreds of thousands of machines every day. SCOPE uses a SQL-like scripting language that is compiled into Direct Acyclic Graphs (DAGs) of operators. SCOPE scripts, commonly referred to as *jobs*, are composed as a data flow of one or more SQL statements that are stitched together into a single DAG by the SCOPE compiler. The SCOPE optimizer is implemented similarly to the traditional cascades-style [13] query optimizer. As such, it uses a set of rules to transform a logical query plan in a top-down fashion, but it also makes all the decisions required to produce a distributed plan. Examples of such decisions are determining how to partition the inputs, or selecting the optimal amount of parallelism given the number of containers available for the job. The number of concurrent containers used by each job is referred to as number of *tokens* in SCOPE, while the total amount of tokens used for executing a job is called *vertices*. The SCOPE optimizer estimates the cost of a plan using a combination of data statistics and other heuristics tuned over the years. Once executed, several metrics are logged by the SCOPE runtime. Common metrics of interests are job *latency*, *vertices count*, and *PNhours* (i.e., the sum of the total CPU and I/O time over all vertices). Finally, SCOPE provides an A/B testing infrastructure (called *Flighting Service*) which allows to re-run jobs in a pre-production environment using different engine configurations and compare the performance of those configurations with the default one.

The importance of recurring jobs. More than 60% of SCOPE jobs are *recurring*, i.e., periodically arriving template-scripts with different input cardinalities and filter predicates [17, 19] , but same set of operators. While each execution of a template can have different metrics such as input cardinalities and selectivities, recurring jobs are overall interesting because we can use historical information on previous executions to improve future occurrences. Because of this, in the remainder of the paper we will focus on recurring jobs. We will touch upon in Section 8 how we are planning to remove such limitation in future versions of `QO-Advisor`.

SCOPE rule-based optimizer. There are 256 rules in the SCOPE optimizer. SCOPE rules are divided into 4 categories [29]: *required* (which must always be enabled to get valid plans), *off-by-default* (which are disabled by default because experimental or very sensitive to estimates), *on-by-default*, and *implementation* rules (mapping logical operators into physical ones). The default optimizer *rule configuration* may return sub-optimal plans for certain job instances or workloads, whereas an instance-optimized rule configuration could return better instance-optimal plans. This, for example, can be achieved by turning on an off-by-default rule, or conversely by turning off an on-by-default or implementation rule (and therefore restricting a part of the optimizer search space). SCOPE users are aware of this, and in fact over the years have attempted to use *optimizer hints* to override the default query optimizer behavior. This, however, is a manual task which requires both deep knowledge of SCOPE internals, as well as tedious experimentation through trials and errors. Despite this, on a daily basis we see that up to 9% of jobs have hints overriding the default configuration of the optimizer. Applying state-of-the-art techniques such as BAO [26] is non-trivial, and it will not work because the configuration space is just too big: 2^{256} in SCOPE vs just 48 considered in BAO.

Towards instance-optimized rule configurations. To overcome the above problem, in [29] we have explored a first step towards automating this manual experimentation task. We analyzed large production workloads from SCOPE and introduced two new concepts to apply the steering ideas in such a system:

Rule Signature. For each plan, we extended the SCOPE optimizer to return a bit vector specifying which rules directly contributed to it. For example, if only the first and the second rule were used during the optimization, then the rule signature will be 1100000000.

Job Span. Given a job, we compute a set containing all rules which, if enabled or disabled, can affect the final query plan. Intuitively, the span of a job contains all the rules that can lead to some change in the final optimized plan. The job span allows to narrow down the exploration of rule configurations by skipping the unworthy ones. The job span is computed heuristically, as described in [29].

Given the rule signature for a job, and its span, in [29] we showed that it is possible to generate instance-optimized rule configurations returning better estimated cost and runtime metrics (up to 90% latency improvement) than using the default rule configuration. The heuristic used in [29] for the configuration search is as follows:

(1) For each job, randomly sample 1000 configurations over that job's span following a uniform distribution;
(2) Recompile all generated samples and confront the new estimated cost returned by the optimizer against the ones from the default rule configuration;
(3) For all configurations with better cost estimates, pick the 10 more promising one, and flight them against the default configuration for validating that better cost estimates translate into better runtime metrics;
(4) Among all the flighted configurations, pick the one with the best runtime metrics improving over the default configuration (if it exists), and apply such configuration to the next occurrence of the recurring job.

In this paper, we will close the loop, and describe how we are able to achieve fully automated instance-optimized rule configurations of SCOPE jobs with QO-Advisor.

2.2 Productization Challenges

Over the past one year, we have been trying to deploy in production the heuristics of [29] for steering the SCOPE query optimizer. However, while doing this, we faced several challenges.

Difficult to debug. Instance-optimized rule configurations are generated using uniform sampling. This provides no real insight on why a configuration is better than another. Consequently, when an IcM (Incident Management ticket) is raised (e.g., because of some unexpected performance regressions) debugging is almost impossible.

Expensive to maintain. The amount of resources required to maintain this feature in production is not trivial. In fact, for each recurring job we recompile 1000 different rule configurations, as well as flight (re-run in pre-production environment) 10 of them, which can quickly become very expensive.

Performance regressions are hard to catch upfront. There is no principled way of detecting regression or bad configurations upfront. The only guard we have are the estimated costs from the SCOPE optimizer after recompilation (whose reliability is well known to be lacking [37]) and flighting (which does not consider the variance inherent in the clusters, unless we repeat each flighting run several times).

2.3 Introduction of QO-Advisor

We now give an overview of the QO-Advisor, which addresses the above challenges in SCOPE-like big data query optimizers. Similar to index recommendations by an index advisor [3, 10, 11], the goal of QO-Advisor is to recommend better search paths, via rule hints, for a query plan. The difference, however, is that while index recommendations are applied offline and appropriate indexes are created a priori, the search space recommendations are applied online during the query optimization process itself. Therefore, to keep this process lightweight, QO-Advisor does all the heavy lifting of generating the appropriate rule hints for different recurring job templates in an offline pipeline. The idea is to leverage the massive past telemetry to explore better query plans in an offline loop, and integrate the recommendations to steer the optimizer to those plans in the future. Figure 1 shows how the QO-Advisor pipeline sits next to the SCOPE engine and helps steer its search space.

QO-Advisor opens up a core query optimizer component, namely the query planner, to be explored and improved upon externally via the use of past telemetry and offline compute resources. This is a major shift from traditional query planning and allows one to think differently. For example, we can consider more expensive and more exhaustive search algorithms which one cannot afford during query optimization. We can observe the actual execution costs seen in the past as compared to relying purely on the estimated costs. We can do more validation offline to anticipate any unexpected query plan behavior and even improve upon it before actually deploying it to future queries. To top it all, QO-Advisor integrates back seamlessly with the query optimization of future queries in an automated loop, without having users to worry about or even notice the externalized query planning component. Thus, QO-Advisor introduces a new way to think about query planning that helps scale query optimizers to the complexity and challenges that we see in modern cloud workloads [18].

2.4 Design Principles

As mentioned above, the QO-Advisor is a pipeline of tasks that is recurrently triggered every day. QO-Advisor pipeline takes as input historical metadata for a given day, and return a list of job template identifiers and rule hint pairs. These pairs are consumed by the SCOPE optimizer such that, every time a job matching one of the template identifier is found, the provided rule hint is used at compile time to steer the query optimizer.

Single rule flip. The first major difference compared to previous approaches is that QO-Advisor does not provide full rule configurations, but it only amends the default SCOPE rule configuration by turning on or off a single rule at a time. We made this design decision for several reasons: (1) the search space is much smaller with single rule flips and therefore easier to manage and less expensive to maintain; (2) in case of performance regression, one single rule change is far more easier to manage/revert compared to arbitrary changes to the rule configurations; (3) this simpler approach is easier to control in production settings, therefore better for building confidence with the product teams.

From naive to informed experimental design. Uniform sampling over the configuration space generated by the job span is expensive to maintain because it requires the generation of several

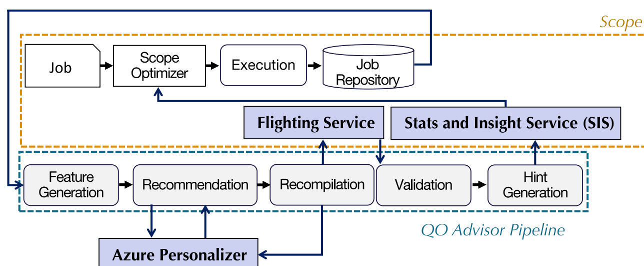

Figure 1: Overview of the QO-Advisor pipeline and its integration with SCOPE. The pipeline is composed of five tasks (Feature Generation, Recommendation, Recompilation, Validation and Hint Generation) and it uses three services: Azure Personalizer, SCOPE Flighting, and Stats and Insight Service.

configurations, the majority of which are discarded. In QO-Advisor we provide a more principled approach whereby we train a contextual bandit model over features extracted from the historical runs, as detailed in Sections 3 and 4. This allows us to decrease the number of required re-compilations, as well as flighting runs.

Learning over estimates rather than runtime metrics. Learning over query execution times, as in BAO, is hard at SCOPE scale. SCOPE daily executes 100,000s of jobs spanning 100,000s of machines, whereby: (1) re-executing even a small fraction of them can be prohibitively expensive; and (2) testing configurations without any (although imprecise) guardrail introduces the chance of executing plans with orders of magnitude worse runtime metrics, which, at SCOPE scales, means a large waste of resources. Therefore we decide to learn rule configurations over the estimated costs output by the SCOPE optimizer. While this approach is not ideal in terms of accuracy (since we are learning to improve estimated costs rather than real ones), still it allows us to run at SCOPE scale.

Models pipeline for avoiding regression. Our learned model suggests new rule configurations by predicting the eventual improvement on estimated costs. The question on how to avoid regressions is still open, since (predicted or actual) estimated costs can be off compared to the real execution runtime, as previously suggested in Section 2.2. To address this, in QO-Advisor we decide to resort to a model pipeline whereby a validation model is applied after the contextual bandit model. The goal of the second model is to make sure that the selected rule configuration will not create performance regressions after execution. To gather runtime measurements, we use flighting. However we flight only a small number of jobs, and under a tight budget. The flighting budget is proportional to how strictly we want to avoid performance regressions.

2.5 Pipeline Overview

Figure 1 provides a sketch of the QO-Advisor pipeline and its integration with SCOPE. The QO-Advisor pipeline is executed offline and daily. It takes as input historical information on job execution for a specified date, and produces a list of pairs (job template, hints) which are loaded into SCOPE's optimizer through the Stats and Insight Service (SIS) [16]. From the historical runs metadata, the pipeline first generates a set of features, and then feeds them into a Recommendation task which uses Contextual Bandit [24] to suggest up to one rule hint for each job. These rule hints are then passed to the SCOPE optimizer for a Recompilation run. Jobs are recompiled such that: (1) we can catch compilation errors due to the new rule settings upfront; and (2) we can get a new estimated cost for the plan. Jobs with worse estimates are pruned, while for the remaining jobs we run an A/B test through the Flighting Service. The output of flighting is then fed into a Validation task that applies a linear regression model. This provides a guard over performance regressions, given that it is crucial for production workloads [4]. The job templates (and related hints) that pass validation, are then written into a file by the Hint Generation task, and loaded into SIS, which, in turn, loads them into the SCOPE optimizer such that the generated hint is applied to the next occurrence of the job template.

Next we describe QO-Advisor's training tasks in the following section. Thereafter, in Section 4, we will provide insights on QO-Advisor's operationlization and integration with SCOPE.

3 LEARNING PRINCIPLES

In this section we describe how QO-Advisor learns to steer the SCOPE optimizer. Below we first give a short background to contextual bandits and then describe our formulation for steering the optimizer search space.

3.1 Introduction to Contextual Bandits

Contextual bandits (CBs) are an extension of supervised learning where the feedback is limited to the actions made by the learning algorithm [24]. For optimization problems where enumeration is impractical, CB is a useful abstraction because they limit the evaluation burden to the actions made by the learning algorithm. Specifically, in our problem setting, the burden of computing the quality of all plausible rule configurations implies supervised learning has a very high overhead, but CB learning has low overhead as only the rule configuration selected by the learning algorithm will need to be evaluated.

CBs achieve statistical guarantees similar to those of supervised learning by constructing an experimental design and then reducing it to supervised learning [2, 12]. However, the use of supervised learning as a subroutine implies CB learning inherits some of the issues of supervised learning. In particular, query plan representations can have a large impact on performance. We discuss issues of representation later in this section.

The experimental designs induced by CB algorithms are randomized; informally, good actions become increasingly more likely under the experiment design as more data accumulates, but other actions still have some likelihood despite their poor historical performance. Specifically to our setting, a CB algorithm will initially choose rule configurations uniformly-at-random, but as historical data accumulates it will play (what the supervised learning model indicates is) the best rules configuration with increasing frequency. Although the guarantees of CB learning have the same dependence on data volume as supervised learning, the effective number of data points is proportional, in the worst case, to the number of possible actions per example. This motivates our strategy for limiting the number of actions per example discussed later in this section.

These aspects of CB learning (limited evaluation requirements and informed randomization) are the basis for the improvements over our previous approach [29].

3.2 Contextual Bandit Formulation

In CB, the learning algorithm repeatedly receives a *context* and *action-set* pair $(x, \{a_i\})$, it chooses an action a_i, and then receives *reward* r. To formulate query optimization as a CB problem, we must specify the context, the actions, and the rewards.

Recommendation. The Recommendation task consumes information available after the default compilation and optionally recommends an alternative compilation, as shown in Figure 1. For this component we use the fractional reduction in (optimizer) estimated cost as the reward. Rewards with extreme dynamic ranges can cause problems for existing learning algorithms, so we use clipping to mitigate the effects of outliers when training.

The specification of the action set $\{a_i\}$ leverages the concepts of rule signature and job span introduced in Section 2.1. The main statistical issue is to limit the number of actions per example in order to control the amount of data required for learning to converge. With B possible bits in the rule signature, there are naïvely 2^B possible actions corresponding to each rule signature. This is intractably large so we instead only consider the rule signatures which differ in the job span. If there are S bits in the job span for a query, then there are 2^S possible actions corresponding to possible settings of the span bits. For this work, we limited the action space to single bit deviations from the default query plan (both to reduce data requirements for learning and for the reasons in Section 2.4). Thus the number of possible actions is $(1 + S)$, corresponding to either changing nothing (1) or flipping a single bit in the span (S). Empirically, S is on average 10 but with a long tail distribution [29], so having actions scale linearly with S ensures tractability.

For the context, any information which is known after the initial default compilation can in principle be utilized. We initially attempted to compute features directly from the logical query plan via properties of the DAG, but this was not effective. Instead we found representing the logical query by the job span itself to be more effective. In other words, the complete set of bit positions in the job span provides valuable and concise information about which bits can be flipped to improve the result, especially when interacted to create second and third order co-occurrence indicators. Beyond these features, we found representing some properties of the input data streams (e.g., row count) provided marginal improvement. Further details are provided in Section 4.2.

Validation. The Validation component consumes information available after a single flighting run of the query plan and decides whether to accept or reject the modified plan. Because (1) there are only two actions, and (2) the reward of reject is known (relative change is 0), we treat this a supervised learning problem and utilize a model trained on a fixed dataset. Further details are in Section 4.3.

4 OPERATIONALIZATION

In this section we will provide insights on the QO-Advisor implementation. We have implemented the QO-Advisor pipeline entirely in Python. Additionally, the QO-Advisor interface with external components, namely SCOPE, Flighting Service, SIS and Azure Personalizer [1], use a mix of Python and Powershell commands.

We run the QO-Advisor pipeline daily from a set of Windows machines. QO-Advisor is currently deployed for some of SCOPE workloads running on multiple SCOPE clusters, where it is enabled by default, i.e., every job submitted within those workloads gets search space recommendations produced by the QO-Advisor pipeline. We are currently investigating further integration with automation services such as Azure Data Factory (ADF) [5].

QO-Advisor takes as input a denormalized *view* of the workload aggregating all compile time as well as execution information of jobs run on a given date [16]. This view file is generated automatically by an ADF pipeline and is shared across different learned components in SCOPE [16, 17, 36, 37]. It contains information such job identifier, name, a description of the job plan, as well as optimizer (e.g., estimated cardinalities, estimated cost) and runtime statistics (e.g., latency, PNhours, actual row counts). We defer to Section 4.1 below for a detailed description on the features available from the above view file, as well as how they are used by our machine learning models in QO-Advisor. This denormalized view for a given date is typically available on COSMOS [33] (the big data platform at Microsoft) within 3 days. This means that QO-Advisor is triggered over jobs executing not earlier than 3 days from the current date.

QO-Advisor pipeline is composed of five tasks, as depicted in Figure 1, namely Feature Generation, rule Recommendation with Recompilation, plan Validation, and the final Hint Generation. Below, we provide implementation details for each of these tasks.

4.1 Feature Generation

In the first step of the QO-Advisor pipeline, starting from the denormalized workload view, we run a sequence of SCOPE jobs to (1) generate the span for all jobs, (2) return all features necessary for the training of the downstream models.

Job span generation. We use the same algorithm illustrated in [29] without any further modification. Briefly, the span algorithm implements an heuristics search whereby only new rules having effect on the final plan can be discovered. Specifically, for each job we start from the original rule configuration, and we turn on all the off-by-default rules, while we turn off all the on-by-default and implementation rules that appear in the original rule signature. We then pass this new rule configuration to the SCOPE optimizer for a recompilation pass. Since some rules that were previously used (and on) are now off, and some other rules that instead were previously off are now available to the optimizer, after recompilation we may have a rule signature with new used rules. We then again turn off all the newly used (on-by-default, off-by-default or implementation rules) and run the new configuration through the optimizer. This process is repeated until we reach a fix-point (i.e., no new rule is added to the signature, or the recompilation fails). All jobs that have an empty span (i.e., the heuristics cannot find any modification to the default configuration) are not further considered.

Feature aggregation. An interesting aspect of SCOPE is that each job executes a *script* which can contain one or more queries. Therefore, for each job, the optimized plan is a DAG of operators (instead of a single tree, as is common for relational database plans), with one or more output nodes, one for each resulting dataset. Each output node can be seen as the root of a tree, whereby the SCOPE optimizer and runtime generates some statistics (features) per tree, while other statistics are generated per script. Examples of these

Table 1: Features generated for each job, how they are aggregated, and their source. For each feature we also highlight whether it is a job-level (J) or a query-level (Q) feature. Note that for job-level features we always use MIN as aggregate function, since all queries within the same job have the same value for those features.

Feature	Aggregation	Source	Level
Normalized Job Name	MIN	Job Metadata	J
Rule Signature	MIN	Optimizer	J
Latency	MIN	Runtime Statistics	J
Estimated Cost	MIN	Optimizer	J
Query Template	MIN	Job Metadata	Q
Total Number of Vertices	MIN	Runtime Statistics	J
Estimated Cardinalities	SUM	Optimizer	Q
Bytes Read	SUM	Runtime Statistics	Q
Maximum Memory Used	MIN	Runtime Statistics	J
Average Memory Used	MIN	Runtime Statistics	J
Average Row Length	AVG	Optimizer	Q
Row Count	SUM	Optimizer	Q
PNHours	MIN	Runtime Statistics	J

per-tree features are row counts, estimated cardinalities and average row length. Examples of job-level statistics are PNhours, latency and total number of used containers. Another interesting aspect of having DAGs instead of trees is that each job could be part of different templates, since template information are generated per-query tree and not per-job script. Finally, since `QO-Advisor` executes at the job granularity (and similarly hints can be provided to the SCOPE optimizer at the job level) we found that there is a disconnect between how the features are generated by SCOPE and the format that is needed for `QO-Advisor` to operate properly.

To solve the above challenge, from the raw features and plans contained in the view file, we transform DAGs into trees by generating a *super root node*. Super root nodes aggregate all the features and information contained in the sub-query trees composing a job. Features are aggregated in two different ways: (1) following their semantics, e.g., we do SUM over estimated cardinalities, and AVG over average row length; (2) for plan-level features as well as for template-level features, we just get the MIN. Table 1 summarizes all the features generated at this step of the pipeline, and the related aggregate function. After this step, all features are job level and properly aggregated. We then attach the span information and feed the result to the Recommend task, which we describe next.

4.2 Rule Recommendation

As formulated in Section 3.2, we use a CB model to recommend *rule flips* (i.e., turn off an on-rule or turn on an off-rule) for a specific job. More specifically, each action flips a rule in the job span. We featurize the actions using the rules id and rules category information as introduced in Section 2.1. The context includes the features of Table 1 generated from the previous step of the `QO-Advisor` pipeline, plus the job span. The reward is the relative change in the optimizer's estimated cost, i.e. the ratio between estimated cost in the default setting over the estimated cost after re-compilation with a recommended rule flip. The contextual bandit optimization is to maximize the reward: if a rule flip leads to a higher reward, the estimated cost after recompilation becomes smaller, so that a plan with lower estimated cost is found. We clip the range of the ratio so that extreme values do not overly skew the recommendation model. Currently we heuristically clip any range greater than 2.0, i.e., we clip any plan that is more than 2× the baseline. The clipped delta is then fed back to the contextual bandit model as the reward.

For CB learning we use an off-policy learning approach [40], where we gather reward information using the uniform-at-random policy, but for the subsequent steps we act using the learned contextual bandit policy. This accelerates learning by inducing a maximally informative training dataset, at the cost of doubling the number of rule configurations test compilations. Since job recompilations are relatively inexpensive this is an acceptable trade-off.

In our implementation, we use the Azure Personalizer service [1] to generate rule recommendations. Azure Personalizer uses the contextual bandit approach introduced in Section 3.1 to predict and learn under a specific problem scenario. Azure Personalizer allows for better development in our scenario compared to ad-hoc solution, thanks to three main advantages: (1) Azure Personalizer automatically handles all aspects of model management, fault-tolerance and high availability; (2) it logs with high fidelity so that we can counter-factually evaluate policies; and (3) it adheres to all Azure compliance and security settings.

4.3 Validation

In the previous rule Recommendation task, we evaluated the rule flips (actions) generated by the CB model. However, we found that using estimated costs alone could sometimes lead to performance regressions. This is mainly due to the variability inherent in the cloud computing system which makes cost estimation incredibly hard [37]. Interestingly though, we found that the variability problem is more severe for the job latency metric and lesser so for the PNhours metric. While we will further discuss this finding in Section 5.1, where we run A/A tests to evaluate the variance in SCOPE clusters, in summary we found that since PNhours is computed as the sum between CPU time and I/O time, the variablity of I/O time across A/A runs is bounded as data read and data written remain constant. From the above observations we made two conclusions:

(1) The variance in the cluster is so high that it makes hard to infer improvements in the runtime reliably and programmatically, without having to execute each job several times.
(2) The total amount of data read and data written are good predictors for whether a rule flip introduces improvement or not. Intuitively, if with the new configuration a job read and write less data, this will likely translate into better runtime.

In order to gather information about data read, data written and other indicators, we use the SCOPE Flighting Service to test the rule flips in a pre-production environment. Next we describe how we use the flight results to validate the recommended actions that are output from the contextual bandit model.

Flighting. Given a list of jobs and their related recommended actions, we use the estimated costs (over the recompiled plans with the rule configurations embedding the flips suggested by the model) to heuristically select which jobs to flight. Fighting jobs is in fact not just expensive, but it is the larger source of resource and time consumption for `QO-Advisor`. Because of this, we have

a limited budget of machines that can be concurrently used for flighting, and we set thresholds on (1) the maximum flighting time for each job (24 hours); (2) the total time budget for flighting; and (3) the delta between the estimated cost from recompilation, and the default one. Additionally, we do not flight all jobs that reach the Recompilation step, but instead we flight one representative job per template (picked randomly). The intuition here is that jobs with the same template are composed by the same queries, and therefore it is not necessary to flight all of them because they have the same plans and eventually the same rule flip. Finally, we flight jobs with lower estimated costs first, such that if we finish the total time budget, we are still able to learn and provide some suggestion even if we are not able to complete the flighting process in full.

Given our limited budget on the number of jobs that can be concurrently flighted, we interact with the SCOPE Flighting Service through a job queue of fixed size. For each job being flighted, the service can return several different outcomes: (1) failure (e.g., the job information or the input data expired); (2) timeout; (3) filtered (e.g., if the job belongs to certain classes of jobs that are not supported by the Flighting Service); (4) success. Only the jobs/templates that are flighted with success are passed to the next step in the pipeline. While flighting introduces some overhead cost for QO-Advisor, this cost is one-time for each job, and amortized over time. The cost is also bound by the available budget, which can be tuned based on how aggressive we want to be in production. In Section 8 we will touch on how we are planning to improve the flighting process for increasing the throughput.

Validation model. The flighting results go over a Validation step for catching possible regressions before running the new rule configuration in production, as suggested in Section 3. In this step we run a linear regression model that learns the PNhours delta given the per job total *DataRead* and *DataWritten features* returned by flighting. Again, the intuition here is that if a job reads and writes less data, then it is likely that the PNhours will be reduced (despite the variance inherent in the cluster). In order to train this supervised validation model, we flight a random subset of the jobs over a period of 14 days to gather a data set of flighting results. The data points are indexed by their timestamps, so that we can split the dataset by date to generate a training set (e.g., data in week0) and a testing set (data in week1), in order to test whether the trained model can generalize to other dates temporally.

The output of the Validation model is then compared against a pre-determined safety threshold such that only when the PNhours delta is below this threshold we are actually confident with running the chosen action in production, without causing significant regressions. The threshold can be increased or decreased based on how much aggressive we want be. At the moment, for the workloads we are currently running in production, this threshold is set to -0.1, meaning that, for a particular job, if the expected reduction in PNhours is at least 10%, the job passes the validation.

4.4 Hint Generation

In this final step, we gather the validated (job templates, new rule configuration) pairs and then explode them by applying the same configuration to all jobs belonging to the same template. The output is saved to a file in the Stats and Insight Service (SIS) pre-defined format, and this file is uploaded to the SIS. SIS makes deploying models and configurations in SCOPE easier as it manages versioning and validates the format before installing them in the SCOPE optimizer [16]. In our case, the SIS picks the file containing jobs and related rule configurations, and applies them as hints to the SCOPE optimizer every time a new instance of the same job template is submitted in the future.

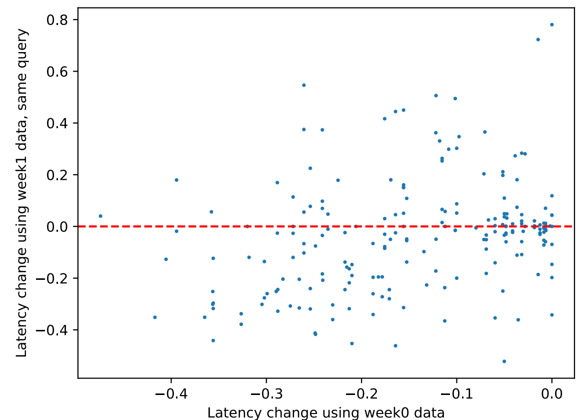

Figure 2: Recurring job stability: latency improvements over recurring jobs for week0 cannot always be repeated in week1.

5 EXPERIMENTS

In this Section, we present several results and insights from one of the production SCOPE workloads, in a pre-production environment. Based on these results, the QO-Advisor was enabled by default for this workload beginning November 2021. Throughout this section, we focus on answering the following questions that made the above deployment possible:

(1) Which runtime metrics are best suited for measuring success? (Section 5.1)
(2) Can we reliably use compile-time-only information to predict runtime outcomes? (Section 5.2)
(3) Can we reliably predict future runtime outcomes from a single flight run? (Section 5.3)
(4) Do we improve over default query plans in aggregate? (Section 5.4)
(5) What is the impact on query performance in the worst case? (Section 5.5)
(6) Is our biased search using contextual bandit more effective than a uniformly-at-random baseline? (Section 5.6)

5.1 Metrics

Which runtime metrics are best suited for measuring success?

When we first tested QO-Advisor in pre-production over recurring jobs, we found that even if we can find savings in latency (or PNhours) by A/B testing against the default configuration, when we run the same recurring job again on a different week, the savings cannot always be repeated, as shown in Figure 2. For example, for a job that we found in week0 with 25% reduction in latency ($x = -0.25$), the same recurring jobs running in week1 does not have any latency reduction, and in fact the latency became larger.

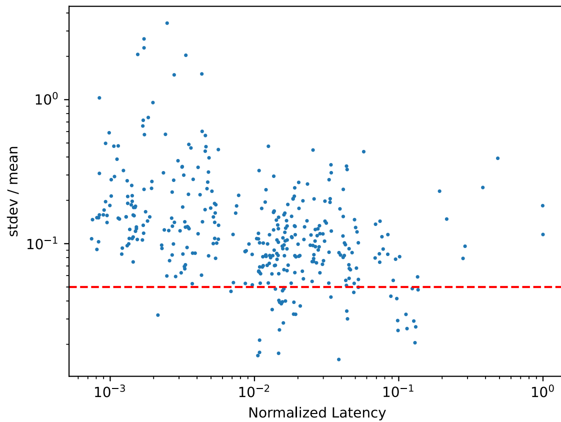

Figure 3: Variance of latency plotted over jobs with different execution time (normalized). Red line is 5% variance. The majority of jobs have high-variance.

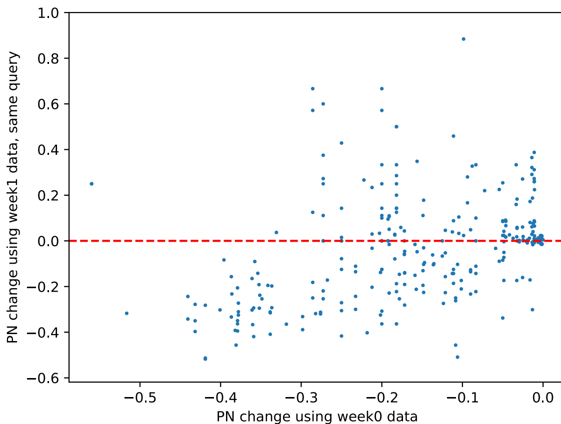

Figure 4: Recurring job stability: savings in PNhours over recurring jobs for week0 cannot always be repeated in week1.

Overall, we found that more than 40% of the jobs will regress when re-run one week apart.

After further investigation we found that the above result is due to the variance inherent in the large-scale distributed data processing systems like SCOPE. Figure 3 shows an A/A test for the same jobs of Figure 2, where we run each of them 10 times using the default rule configurations. We plot the job variance over the normalized job execution time. The red line marks 5% variance that is typically expected by the product teams. As we can see from the Figure, more than 90% of the jobs have more than 5% variance in latency, with few jobs having over 100% variance. This variance creates problems for both learning and evaluation. For learning, datasets with such high variance contains too much noise. For evaluation, we cannot rely on a single A/B test to determine the performance improvements, and we cannot afford to always run it multiple times, and then take the mean for example. While this variance is intrinsic in distributed processing over a cluster of machines, it is exacerbated by how SCOPE manages resources [7].

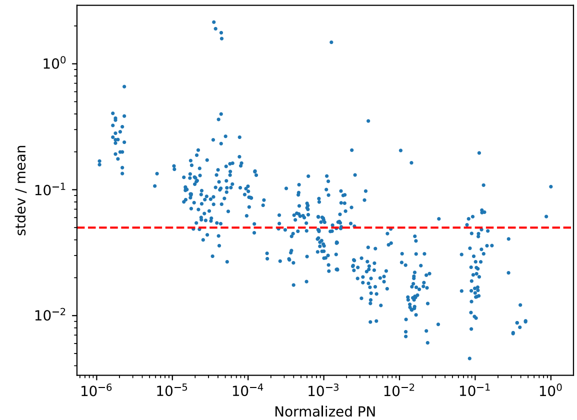

Figure 5: Variance of PNhours plotted over jobs with different normalized execution time. The red line is 5% variance. Compared to the latency plot of Figure 3, PNhours is more stable, with more than 50% jobs incurring <5% variance.

A similar conclusion can be drawn for PNhours, as shown in Figures 4 and 5. Simply relying on PNhours found previously in week0 will lead to more than 40% regression. The variance here looks more contained however, with less than 50% of jobs having a variance greater than 5%. Since the PNhours metric appears to have less variance, and there is no clear signal between estimated cost and job latency (as we shall explain next), in the current deployment of `QO-Advisor` we decided to focus on optimizing for PNhours.

5.2 Estimated Cost vs. Real Performance

Can we reliably use compile-time-only information to predict run-time outcomes?

Although our initial design did not include a validation component, the results in this section convinced us that some run-time information was necessary for reliable improvement. Specifically, we tested the rule flips leading to lower estimated costs in an A/B testing using the Flighting Service.

Figure 6 plots the estimated cost delta (relative difference between the estimated cost output of the recompilation with the new hints minus the original estimated cost output of the default rule configuration), and the delta of the latency of 950 jobs run over 5 days on one of the SCOPE clusters. Our expectation was that as the estimated cost delta decreases (i.e., the cost decreases), the latency delta would decrease as well. However as we can see from the figure, there is no real correlation between improvements in the estimated costs and latency. For example, the left-side of the figure contains the jobs with large improvements on the estimated costs, but over 40% of them show actually performance regression.

Nevertheless, estimated cost information are still important for the success of the approach. For instance, we ran an experiment where we disabled any use of the estimated cost coming from the SCOPE optimizer. Specifically, in the Recommendation step we disabled all filters based on estimated costs, and we randomly flip one rule in the span, instead of using the CB model. What we found

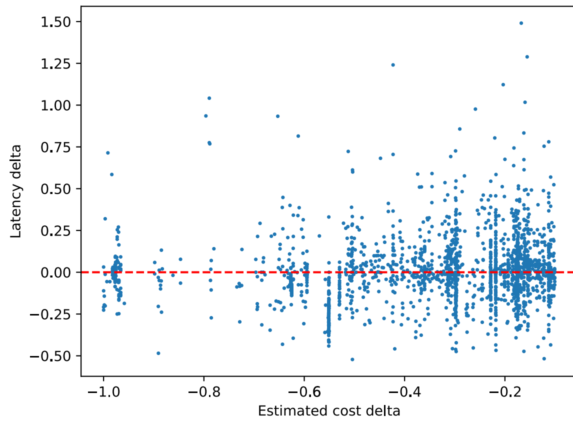

Figure 6: Delta (new value / old value - 1) in estimated cost versus latency. Improvement in estimated cost after recompile do not often translate in improvement in latency.

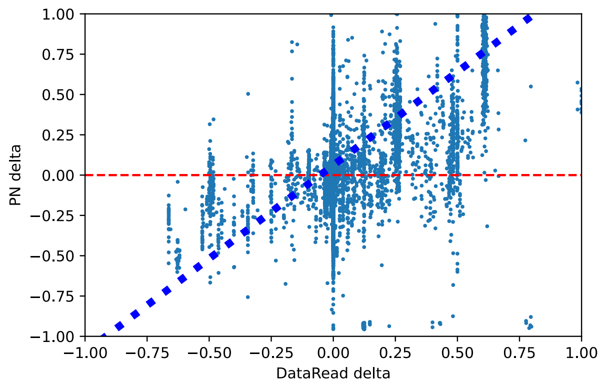

Figure 7: Correlation between DataRead and PNHours. Blue dotted line highlights the trend.

is that, after three days, `QO-Advisor` was not able to complete flighting (a task that usually is completed within half a day). This is because, without using estimated costs to heuristically (1) pick rules, and (2) filter which jobs to flight, plans leading to job latencies orders of magnitude worse than the baseline can be introduced into the pipeline, eventually making fighting impractical.

5.3 Validation Model Performance

Can we reliably predict future runtime outcomes from a single runtime outcome?

As discussed in Section 4.3, we use a Validation step to verify that the recommended rule flip can indeed improve PNhours, without causing significant regression on other metrics. Since there is still some variance in the PNhours metric, we cannot trust a single run of the A/B testing for recurring jobs (Figure 4), and, instead, we have to build a model to predict future PNhours delta.

Through experimentation we found that in addition to the PNhours metric itself, DataRead and DataWritten deltas are good

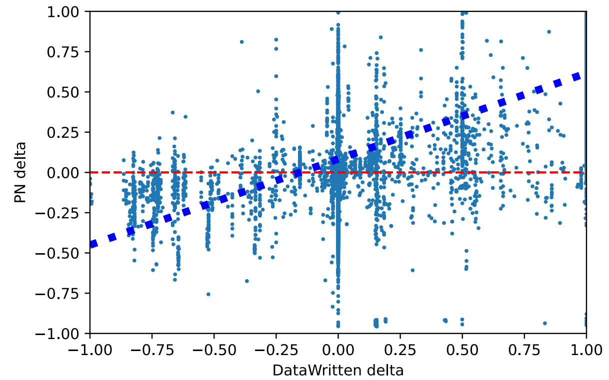

Figure 8: Correlation between DataWritten and PNHours. Blue dotted line highlights the trend.

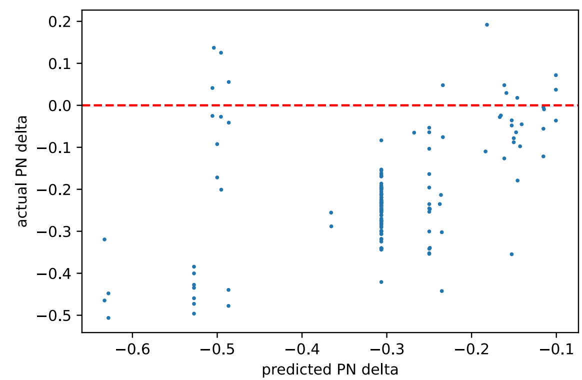

Figure 9: Predicted PNhours delta vs. Actual PNhours delta.

indicators of how PNhours delta will change. Figure 7 shows that there is a correlation between DataRead delta and PNhours delta in the historical data we gathered. The dotted blue line demonstrates the trend by a one-dimensional polynomial fit. This result suggests that if we have seen less data read in the A/B testing, then it is likely that the PNhours will also be reduced. Similar observations can be found in Figure 8 for DataWritten. These results corroborate the intuition in Section 4.3 that by reducing I/O during job executions, we can reduce the PNhours and achieve better performance.

Using the Validation model described in Section 4.3, we can compute the predicted delta of PN hours. Given a certain threshold (e.g., $delta < -0.1$), we select the jobs with predicted PNhours delta qualified for this threshold, and run them in A/B testing to validate our prediction and check the results on other performance metrics. To compensate for the variance in PNhours metric, we usually use a threshold smaller than zero, so that we can avoid some regressions introduced by the variability in the cluster, as discussed in Section 5.1. Figure 9 demonstrates the accuracy of the Validation model (trained on historical data) on 150 jobs in one day of the test data. For the jobs we predicted with PNhours delta smaller than -0.1, 85% of them have their actual PNhours deltas smaller than -0.1, and 91% of the jobs have their actual PNhours deltas smaller than 0.0 as indicated by the red line.

5.4 Aggregate Performance

Do we improve over default query plans in aggregate?

As introduced in Section 2.1, metrics of interests in our experiments include PNhours, latency, and vertices count. As just discussed in the Section 5.1, in QO-Advisor our primary focus is optimizing PNhours, which measures the execution cost of a SCOPE job in terms of total resources. But our goal is also to avoid significant regressions on job latency and vertices count as well.

Table 2 presents the aggregate pre-production results for one SCOPE workload. This workload contains 70 jobs on a single day matching the hints generated by the QO-Advisor pipeline. Typically, we expect that around 5% of the unique jobs of the workloads with QO-Advisor enabled can find matches in the hints recommendation. In aggregate, comparing to the default plans generated by the SCOPE optimizer, on this particular workload we observed 14.3% savings in total PNhours, 8.9% savings in total job latency, and 52.8% savings in vertices count, after using the recommended hints for these jobs. Note that we set the threshold of predicted PN delta to -0.1, and the PNhours reductions is roughly in this range.

Table 2: Pre-production results. % of reduction for QO-Advisor compared to the default rules configuration.

Metric	%Reduction
PNhours	-14.3%
Latency	-8.9%
Vertices	-52.8%

5.5 Distribution of Performance Metrics

What is the impact on query performance in the worst case?

In this section we drill down on the aggregated results presented in the previous Section. Figures 10, 11, and 12 present the performance changes for the 70 jobs of Section 5.4 on three metrics (PNhours, latency, and vertices, respectively) compared with the default query plan. Note that in each figure, the jobs are ordered by the change in that specific metric, and a delta greater than 0.0 indicates a regression.

In Figure 10, we see that over 80% of the jobs have smaller PNhours, indicating savings in resource usage. In the best case on the left part of the figure, the PNhours of a job is reduced by almost 50%. Conversely, in the worst case, the PNhours of a job is increased by at most 15%—a considerable improvement over the 100% worst case regression of our previous work that uses randomly chosen rule flips without any validation [29].

Regarding job latency, in Figure 11 we also observe that about 80% of jobs have reduced their latency, by a maximum of over 90%. We also see that 20% of jobs have latency increased, but compared with the results in Figure 6 (using estimated cost alone) or Figure 4 (using recurring jobs without validation), the regression rate is reduced from more than 40% to 20%. However, since in the Validation step we primarily focused on optimizing for PNhours, the regression on latency is less contained.

Finally, for vertices count (Figure 12), we see that only two jobs used 10% more vertices. This effect appears to be explainable by our approach of optimizing PNhours and reducing the data that is read and written during job execution, since the I/O reduction might be a natural result of fewer vertices used for job execution which consequentially requires less data transmission.

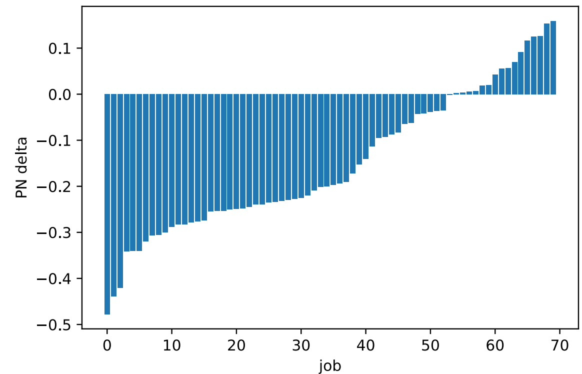

Figure 10: Pre-production results drill down: PNhours delta. In the best case, QO-Advisor improves PNhours over the default rule configuration by approximately 50%. In the worst case PNhours increases by 15%.

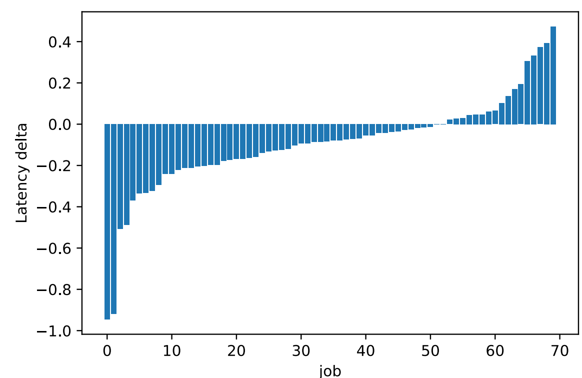

Figure 11: Pre-production results drill down: latency delta. In the best case QO-Advisor is able to improve latency by 90%. In the worst case, QO-Advisor introduces a regression of about 45%. The larger regression compared to PNhours (Figure 10) is because QO-Advisor is tuned over PNhours.

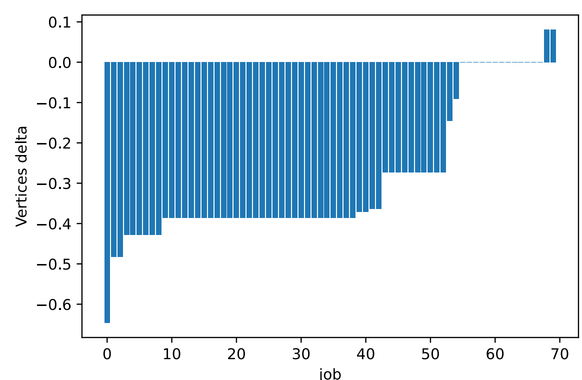

Figure 12: Pre-production results drill down: vertices delta. In this case QO-Advisor introduces a regression of 10% in the worst case, while in the best case it can improve vertices utilization by more than 60%.

Table 3: Comparison between random and CB rule flips.

Number of jobs	Random	Random %	CB	CB%
Lower cost	377	10.6%	1226	34.5%
Equal cost	1257	35.4%	1141	32.1%
Higher cost	1280	36.0%	693	19.5%
Recompile failures	638	18.0%	492	13.9%

So given the small fraction of regressions and the contained impact of the regressed jobs, we consider that QO-Advisor is working well on production workloads and overall we can reduce the PNhours, job latency and vertices for the recurring jobs matching our recommendations.

5.6 Biased Randomization

Is our biased search using contextual bandit more effective than a uniformly-at-random baseline?

Conceivably, the Validation step of the QO-Advisor pipeline might be strong enough to ensure aggregate improvement, even if the biased randomization in the Recommendation component provides no improvement. Therefore, we evaluated the effectiveness of the rule Recommendation step (Section 4.2) using CB against a baseline that chooses a rule uniformly at random in the action set to flip. For this experiment, we check how often these two approaches can find better query plans in terms of optimizer estimated costs, which we use as the reward to train the reinforcement learning policy.

For the workload tier we are currently running QO-Advisor on, from one single day after feature generation (Section 4.1) around 66% of the jobs have non-empty job span (i.e. the action set). For these jobs, the random baseline flips one rule out of the span at uniform random, while our approach chooses one rule with the guidance of the contextual bandit policy. Note that for the contextual bandit policy, the selected rule flip does not necessarily lead to a better plan than the default rule configuration. Thus we always recompile with the CB's selected rule flip and short-circuit subsequent processing if there is no estimated cost improvement.

For the same set of jobs in the workload, we recompile with the rule flips to check whether the estimated costs of the jobs become lower, higher, or remain the same compared to the default rule configurations. Table 3 shows the number of jobs in each category of the cost changes. We observe that using CB increases the number of jobs with lower costs by 3×, and reduces the number of jobs with higher costs by almost 2×. It also leads to fewer recompilation failures. Overall, the total estimated cost of this particular workload is reduced from $1.7e11$ using random rule flips to $1.0e09$ using the CB rule flips—an improvement of over 100×.

6 LESSONS LEARNED

From research to practice: trade-offs. Applying academic research ideas into production always involves trade-off, and going from the BAO's ideas to QO-Advisor was no exception. For example, already in [29] we realized that generating new rules configurations using runtime information at SCOPE scale requires a large amount of resources. Not just because SCOPE workloads are orders of magnitude larger than the one considered in [26], but also because SCOPE workloads are more complex, and only focusing on few rules (as in BAO) would not work for all use-cases. In our initial investigation, in fact, we found that there is no subset of rules that captures all the benefits, whereas all the rules need to be considered to properly optimize SCOPE workloads. In QO-Advisor we therefore decided to focus on all the 256 rules available in SCOPE, and to learn over estimated costs (which can be quickly returned by the SCOPE optimizer) rather than actual runtime metrics.

Simplicity first. Similarly, we found that suggesting arbitrary complex rule configurations such as in [29] was not ideal for our product partners because in case of regression it would be almost impossible to debug. We therefore decided to stage the problem with only one rule difference from the default rule configuration in the first version. This allowed us to build trust with the product team where this feature works, and also explain better which rules are really moving the needle, something the team could further investigate and corroborate based on their domain knowledge.

Do not reinvent the wheel. In the first prototype of QO-Advisor, we used Vowpal Wabbit [34] for the contextual bandit model. However, as we hardened the system, we realized that maintaining the state over pipeline runs in a reliable way is non-trivial, especially if, for explainability, we want to trace how the model evolves and learns over time. Instead of building our own infrastructure, we decided to integrate QO-Advisor with Azure Personalizer [1], which manages the log and state, as well as provides all the enterprise-grade features we needed for our scenario.

Regressions. We learned that performance regressions are important to catch upfront. At SCOPE scale, if even a 1% of the jobs introduce regression, the number of customer incidents could easily overwhelm the product team [15]. At the same time, we learned that SCOPE clusters have non-negligible variability in performance, making it practically impossible to effectively catching regression without extensive A/A and A/B testing. Our trade-off here is to try to catch regression in a best effort way, with the limited budged allowed for flighting, and by using a machine learning model to validate the output of flighting before uploading the actions for online production consumption. In general, live online experimentation is expensive, difficult, or even not allowed to be done in SCOPE-like production systems that run critical customer workloads. We use counter-factual evaluations where we can rely on past telemetry offline to improve learning parameters and to tune the model.

The surprising effectiveness of span features. Interestingly, we found that complex featurizations of query plans, as suggested in previous works (e.g., [26, 27]) were mostly ineffective, whereas the use of the complete job span as context features was critical to our success. In particular second and third order coocurrence indicator features over the job span. Intuitively, the complete job span concisely represents the query plan by identifying the set of controls which affect the optimizer output *on this instance*. This strategy is plausibly reusable in other scenarios where a learning system is attempting to steer the result of an optimization subroutine in an instance-dependent fashion.

7 RELATED WORK

Steering query optimizers. BAO [26] generates 48 rule configurations in PostgreSQL that affect the optimizer behaviour for choosing

scan operators, join operators, and join orders. BAO treats each rule configuration as an arm in a multi-armed bandit problem and given a new query, it learns to choose one of the 48 arms. BAO models a reinforcement learning problem in which it sees a sequence of queries, and over time learns to make the correct decisions. BAO was evaluated on a custom dataset with queries ranging from a few to several hundred seconds. Also, a core component of BAO is a tree convolutional neural network which learns a cost model for tree structured PostgreSQL query plans by executing the plans.

In our recent work [29], we extended the steering optimizer research by considering how an approach similar to BAO can be ported over real-world workloads in industry strength cloud data processing system, SCOPE [8]. We show that indeed it is possible to improve SCOPE plans by steering its query optimizer, although this requires substantial experimentation and resources since SCOPE optimizer rule set is quite large. Another problem that we faced is that, while plans can be improved, a large portion of the explored plans actually introduce regressions. In this work we address both these concerns, and show, through production results, that is possible to steer SCOPE optimizer towards better plans, without necessarily introduce too much regressions.

Learning-based optimizations in SCOPE. Apart from the scale and complexity, one of the other major challenges in SCOPE is efficiency [30]. As a result, the Peregrine infrastructure has been built over the recent years to introduce workload optimizations in SCOPE to optimize multiple jobs and reduce the overall costs [16]. Subsequently, several efforts have built on top of Peregrine to introduce learned components in SCOPE. Examples include learned models to automatically choose the number of concurrent containers a job should use [6, 32, 36], learned cardinality estimation tailored to the past SCOPE workloads [41], a learning-based approach for cost models [37], and even learning-based checkpointing [43].

Instance-optimized database components. Several recent works focus on improving database components [9, 21, 22, 31, 39] or generating efficient query plans [23, 27, 28] by exploiting learning over actual runtimes or the output of an optimizer's cost model. While instance optimization is very appealing for modern cloud customers, there are concerns about its end to end integration, impact on actual performance, and also potential performance regressions [4, 11].

Contextual Bandit, Reinforcement Learning and their applications. Applications and systems constantly face decisions that require picking from a set of actions based on contextual information. Reinforcement-based learning algorithms such as contextual bandits [24] can be very effective in these settings. The Azure Personalizer service [1] implements several contextual bandits algorithms and is deployed to many content recommendation applications such as MSN news recommendation and XBox website personalization. For systems, deployed instances of contextual bandit approaches include Azure Compute that wants to contextually optimize the waiting time with non-responsive virtual machines in a cloud service, and tuning parameters of components in Skype [14].

8 LIMITATIONS AND FUTURE WORK

In its current version, `QO-Advisor` has several limitations that we plan to improve as we build deployment experience and confidence in the pipeline. The first limitation is that `QO-Advisor` currently suggests only one single rule flip per job. While the simplicity of this solution helps with debugging and with building trust with the product partners, it limits how many plans can be improved. In future work we will propose multiple rule flips, e.g., by utilizing techniques from combinatorial contextual bandits or short-horizon episodic reinforcement learning [38].

Another limitation is the inefficiency of the Validation stage of the pipeline, which requires comparing the runtime metrics of the proposed rule change to the default configuration. Even for recurrent queries, where this additional query execution cost can be amortized over the lifetime of the query, the overhead limits the absolute number of proposed rule changes we can accept per day due to capacity constraints. In future work we will attempt to optimistically accept proposed query plans and detect regressions from subsequent runtime metrics.

Other limitations include: (1) we are currently using heuristics to pick which jobs to flight, and (2) each pipeline is run on a given date independently from the others. We are exploring how to make `QO-Advisor` stateful such that jobs already explored on a previous date can be skipped, when unmodified, in successive dates. In the long term, we are planning to make flighting part of the model, such that we can chose which job to flight based on what part of the space the model thinks is better to explore. Eventually, we would like to go beyond recurrent jobs and provide hints for any SCOPE jobs. This will require a tighter integration between the SCOPE optimizer and Azure Personalizer, in an online learning setting assisted by an offline pipeline for exploring new configurations.

Finally, `QO-Advisor` is just one of many learned components that are currently being explored in production to improve SCOPE performance. It is not clear how these learned components synergize with each other, and this is an exciting area requiring further exploration. For instance, the quality of query plans is directly related to the quality of the estimated costs and cardinalities, but these estimates can be off by a large margin in practice. `QO-Advisor` avoids this pitfall by learning how to directly steer the query optimizer towards better plans. Will better learned cardinalities improve the plans generated by `QO-Advisor`, and vice-versa, in a virtuous cycle? Or as we add more learned components, will the improvements coming from each feature compound with each other?

9 CONCLUSION

We provide an update on our journey towards making steering query optimizers a reality in production for SCOPE users. We introduce `QO-Advisor` to address several shortcomings of previous approaches in a novel way. Specifically, we use a pipeline of models to (1) generate new rule configurations 1-edit distance from the default configuration; and (2) validate A/B testing runs to catch possible regressions before production deployment. As of November 2021, `QO-Advisor` pipeline runs offline periodically (once a day) and is currently deployed in production over different SCOPE clusters.

ACKNOWLEDGMENTS

We would like to thank Carlo Curino, John Langford, Raghu Ramakrishnan, and Siddhartha Sen for their insightful feedback, as well as GT Ni for the work during early stages of the development.

REFERENCES

[1] Alekh Agarwal, Sarah Bird, Markus Cozowicz, Luong Hoang, John Langford, Stephen Lee, Jiaji Li, Dan Melamed, Gal Oshri, Oswaldo Ribas, et al. 2016. Making contextual decisions with low technical debt. *arXiv preprint arXiv:1606.03966* (2016).

[2] Alekh Agarwal, Daniel Hsu, Satyen Kale, John Langford, Lihong Li, and Robert Schapire. 2014. Taming the monster: A fast and simple algorithm for contextual bandits. In *International Conference on Machine Learning*. PMLR, 1638–1646.

[3] Sanjay Agrawal, Surajit Chaudhuri, Lubor Kollar, Arun Marathe, Vivek Narasayya, and Manoj Syamala. 2005. Database tuning advisor for microsoft sql server 2005. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*. 930–932.

[4] Remmelt Ammerlaan, Gilbert Antonius, Marc Friedman, HM Sajjad Hossain, Alekh Jindal, Peter Orenberg, Hiren Patel, Shi Qiao, Vijay Ramani, Lucas Rosenblatt, et al. 2021. PerfGuard: deploying ML-for-systems without performance regressions, almost! *Proceedings of the VLDB Endowment* 14, 13 (2021), 3362–3375.

[5] Microsoft Azure. [n.d.]. Azure Data Factory. https://azure.microsoft.com/en-us/services/data-factory/#overview.

[6] Malay Bag, Alekh Jindal, and Hiren Patel. 2020. Towards plan-aware resource allocation in serverless query processing. In *12th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 20)*.

[7] Eric Boutin, Jaliya Ekanayake, Wei Lin, Bing Shi, Jingren Zhou, Zhengping Qian, Ming Wu, and Lidong Zhou. 2014. Apollo: Scalable and Coordinated Scheduling for Cloud-Scale Computing. In *11th USENIX Symposium on Operating Systems Design and Implementation (OSDI 14)*. 285–300.

[8] Ronnie Chaiken, Bob Jenkins, Per-Åke Larson, Bill Ramsey, Darren Shakib, Simon Weaver, and Jingren Zhou. 2008. Scope: easy and efficient parallel processing of massive data sets. *Proceedings of the VLDB Endowment* 1, 2 (2008), 1265–1276.

[9] Carlo Curino, Neha Godwal, Brian Kroth, Sergiy Kuryata, Greg Lapinski, Siqi Liu, Slava Oks, Olga Poppe, Adam Smiechowski, Ed Thayer, et al. 2020. MLOS: An Infrastructure for Automated Software Performance Engineering. In *Proceedings of the Fourth International Workshop on Data Management for End-to-End Machine Learning*. 1–5.

[10] Debabrata Dash, Neoklis Polyzotis, and Anastasia Ailamaki. 2011. CoPhy: A Scalable, Portable, and Interactive Index Advisor for Large Workloads. *Proceedings of the VLDB Endowment* 4, 6 (2011).

[11] Bailu Ding, Sudipto Das, Ryan Marcus, Wentao Wu, Surajit Chaudhuri, and Vivek R Narasayya. 2019. Ai meets ai: Leveraging query executions to improve index recommendations. In *Proceedings of the 2019 International Conference on Management of Data*. 1241–1258.

[12] Dylan J Foster, Claudio Gentile, Mehryar Mohri, and Julian Zimmert. 2020. Adapting to misspecification in contextual bandits. *Advances in Neural Information Processing Systems* 33 (2020), 11478–11489.

[13] Goetz Graefe. 1995. The cascades framework for query optimization. *IEEE Data Eng. Bull.* 18, 3 (1995), 19–29.

[14] Jayant Gupchup, Ashkan Aazami, Yaran Fan, Senja Filipi, Tom Finley, Scott Inglis, Marcus Asteborg, Luke Caroll, Rajan Chari, Markus Cozowicz, Vishak Gopal, Vinod Prakash, Sasikanth Bendapudi, Jack Gerrits, Eric Lau, Huazhou Liu, Marco Rossi, Dima Slobodianyk, Dmitri Birjukov, Matty Cooper, Nilesh Javar, Dmitriy Perednya, Sriram Srinivasan, John Langford, Ross Cutler, and Johannes Gehrke. 2020. Resonance: Replacing Software Constants with Context-Aware Models in Real-time Communication. In *NeurIPS 2020*.

[15] Alekh Jindal and Matteo Interlandi. 2021. Machine Learning for Cloud Data Systems: the Promise, the Progress, and the Path Forward. *Proc. VLDB Endow.* 14, 12 (2021), 3202–3205. http://www.vldb.org/pvldb/vol14/p3202-jindal.pdf

[16] Alekh Jindal, Hiren Patel, Abhishek Roy, Shi Qiao, Zhicheng Yin, Rathijit Sen, and Subru Krishnan. 2019. Peregrine: Workload optimization for cloud query engines. In *Proceedings of the ACM Symposium on Cloud Computing*. 416–427.

[17] Alekh Jindal, Shi Qiao, Hiren Patel, Zhicheng Yin, Jieming Di, Malay Bag, Marc Friedman, Yifung Lin, Konstantinos Karanasos, and Sriram Rao. 2018. Computation reuse in analytics job service at microsoft. In *Proceedings of the 2018 International Conference on Management of Data*. 191–203.

[18] Alekh Jindal, Shi Qiao, Rathijit Sen, and Hiren Patel. 2021. Microlearner: A fine-grained Learning Optimizer for Big Data Workloads at Microsoft. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*. IEEE, 2423–2434.

[19] Sangeetha Abdu Jyothi, Carlo Curino, Ishai Menache, Shravan Matthur Narayanamurthy, Alexey Tumanov, Jonathan Yaniv, Ruslan Mavlyutov, Inigo Goiri, Subru Krishnan, Janardhan Kulkarni, et al. 2016. Morpheus: Towards Automated {SLOs} for Enterprise Clusters. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. 117–134.

[20] Tim Kraska. 2021. Towards instance-optimized data systems. *Proceedings of the VLDB Endowment* 14, 12 (2021).

[21] Tim Kraska, Mohammad Alizadeh, Alex Beutel, H Chi, Jialin Ding, Ani Kristo, Guillaume Leclerc, Samuel Madden, Hongzi Mao, and Vikram Nathan. 2019. SageDB: A Learned Database System. In *CIDR*.

[22] Tim Kraska, Alex Beutel, Ed H Chi, Jeffrey Dean, and Neoklis Polyzotis. 2018. The case for learned index structures. In *Proceedings of the 2018 international conference on management of data*. 489–504.

[23] Sanjay Krishnan, Zongheng Yang, Ken Goldberg, Joseph Hellerstein, and Ion Stoica. 2018. Learning to optimize join queries with deep reinforcement learning. *arXiv preprint arXiv:1808.03196* (2018).

[24] John Langford and Tong Zhang. 2007. The epoch-greedy algorithm for multi-armed bandits with side information. *Advances in neural information processing systems* 20 (2007).

[25] Viktor Leis, Andrey Gubichev, Atanas Mirchev, Peter Boncz, Alfons Kemper, and Thomas Neumann. 2015. How good are query optimizers, really? *Proceedings of the VLDB Endowment* 9, 3 (2015), 204–215.

[26] Ryan Marcus, Parimarjan Negi, Hongzi Mao, Nesime Tatbul, Mohammad Alizadeh, and Tim Kraska. 2021. Bao: Making learned query optimization practical. In *Proceedings of the 2021 International Conference on Management of Data*. 1275–1288.

[27] Ryan Marcus, Parimarjan Negi, Hongzi Mao, Chi Zhang, Mohammad Alizadeh, Tim Kraska, Olga Papaemmanouil, and Nesime Tatbul. 2019. Neo: a learned query optimizer. *Proceedings of the VLDB Endowment* 12, 11 (2019), 1705–1718.

[28] Ryan Marcus and Olga Papaemmanouil. 2018. Deep reinforcement learning for join order enumeration. In *Proceedings of the First International Workshop on Exploiting Artificial Intelligence Techniques for Data Management*. 1–4.

[29] Parimarjan Negi, Matteo Interlandi, Ryan Marcus, Mohammad Alizadeh, Tim Kraska, Marc Friedman, and Alekh Jindal. 2021. Steering query optimizers: A practical take on big data workloads. In *Proceedings of the 2021 International Conference on Management of Data*. 2557–2569.

[30] Hiren Patel, Alekh Jindal, and Clemens Szyperski. 2019. Big Data Processing at Microsoft: Hyper Scale, Massive Complexity, and Minimal Cost. In *Proceedings of the ACM Symposium on Cloud Computing*. 490–490.

[31] Andrew Pavlo, Gustavo Angulo, Joy Arulraj, Haibin Lin, Jiexi Lin, Lin Ma, Prashanth Menon, Todd C Mowry, Matthew Perron, Ian Quah, et al. 2017. Self-Driving Database Management Systems.. In *CIDR*, Vol. 4. 1.

[32] Anish Pimpley, Shuo Li, Anubha Srivastava, Vishal Rohra, Yi Zhu, Soundararajan Srinivasan, Alekh Jindal, Hiren Patel, Shi Qiao, and Rathijit Sen. 2021. Optimal resource allocation for serverless queries. *arXiv preprint arXiv:2107.08594* (2021).

[33] Conor Power, Hiren Patel, Alekh Jindal, Jyoti Leeka, Bob Jenkins, Michael Rys, Ed Triou, Dexin Zhu, Lucky Katahanas, Chakrapani Bhat Talapady, et al. 2021. The cosmos big data platform at Microsoft: over a decade of progress and a decade to look forward. *Proceedings of the VLDB Endowment* 14, 12 (2021), 3148–3161.

[34] Yahoo! Research. [n.d.]. Vowpal Wabbit. https://vowpalwabbit.org/research.html.

[35] Abhishek Roy, Alekh Jindal, Priyanka Gomatam, Xiating Ouyang, Ashit Gosalia, Nishkam Ravi, Swinky Mann, and Prakhar Jain. 2021. SparkCruise: workload optimization in managed spark clusters at Microsoft. *Proceedings of the VLDB Endowment* 14, 12 (2021), 3122–3134.

[36] Rathijit Sen, Alekh Jindal, Hiren Patel, and Shi Qiao. 2020. Autotoken: Predicting peak parallelism for big data analytics at microsoft. *Proceedings of the VLDB Endowment* 13, 12 (2020), 3326–3339.

[37] Tarique Siddiqui, Alekh Jindal, Shi Qiao, Hiren Patel, and Wangchao Le. 2020. Cost models for big data query processing: Learning, retrofitting, and our findings. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. 99–113.

[38] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.

[39] Dana Van Aken, Andrew Pavlo, Geoffrey J Gordon, and Bohan Zhang. 2017. Automatic database management system tuning through large-scale machine learning. In *Proceedings of the 2017 ACM international conference on management of data*. 1009–1024.

[40] Yu-Xiang Wang, Alekh Agarwal, and Miroslav Dudık. 2017. Optimal and adaptive off-policy evaluation in contextual bandits. In *International Conference on Machine Learning*. PMLR, 3589–3597.

[41] Chenggang Wu, Alekh Jindal, Saeed Amizadeh, Hiren Patel, Wangchao Le, Shi Qiao, and Sriram Rao. 2018. Towards a learning optimizer for shared clouds. *Proceedings of the VLDB Endowment* 12, 3 (2018), 210–222.

[42] Ming-Chuan Wu, Jingren Zhou, Nicolas Bruno, Yu Zhang, and Jon Fowler. 2012. Scope playback: self-validation in the cloud. In *Proceedings of the Fifth International Workshop on Testing Database Systems*. 1–6.

[43] Yiwen Zhu, Matteo Interlandi, Abhishek Roy, Krishnadhan Das, Hiren Patel, Malay Bag, Hitesh Sharma, and Alekh Jindal. 2021. Phoebe: a learning-based checkpoint optimizer. *Proceedings of the VLDB Endowment* 14, 11 (2021), 2505–2518.