WiSeDB: A Learning-based Workload Management Advisor for Cloud Databases (Poster Abstract)

Ryan Marcus  
Brandeis University  
rcmarcus@brandeis.edu

Olga Papaemmanouil  
Brandeis University  
olga@cs.brandeis.edu

Abstract

Workload management for cloud databases deals with the tasks of resource provisioning, query placement and query scheduling in a manner that meets the application's performance goals while minimizing the cost of using cloud resources. Existing solutions have approached these challenges in isolation and with only a particular type of performance goal in mind. In this poster, we showcase WiSeDB, a learning-based framework for generating end-to-end workload management solutions customized to application-defined performance metrics and workload characteristics. Our approach relies on decision tree learning to train offline cost-effective models for guiding query placement, scheduling, and resource provisioning decisions. These models can be used for both batch and online scheduling of incoming workloads. A unique feature of our system is that it can adapt its offline model to stricter/looser performance goals with minimal re-training. This allows us to present to the user alternative workload management solutions that address the typical performance vs cost trade-off of cloud services. Experimental results show that our approach has very low training overhead while it discovers near optimal solutions for a variety of performance goals and workload characteristics.

We argue that cloud-based databases could benefit from a workload management advisor service that allows applications to specify their workload characteristics and performance objectives and returns a set of low-cost and performance-efficient strategies for executing their workloads on a cloud infrastructure.

We believe that such a service should meet three design goals:

1. **End-to-end solutions**: the service should calculate the cloud resources to be provisioned (e.g., number/type of VMs), (b) the distribution of resources among the workload queries (e.g., which VM will execute a given query) and (c) the execution order of these queries.

2. **Cost-aware**: since cloud providers offer their resources for some cost (i.e., price/hour for renting a VM), optimizing workload schedules for this cost is vital for cloud-based applications to be effective.

3. **Custom application-level goals**: different workload schedules offer different performance vs. cost trade-offs for different performance metrics, the system should be able to discover the "best" heuristic for executing a given workload under an application-specific performance goal. Low cost workload schedules should be discovered independently of the performance metric.

This poster showcases WiSeDB ([W]orkload management [Se]rvice for cloud [DB]s), a workload management advisor for cloud databases designed to satisfy the above requirements. WiSeDB relies on a learning framework that "learns" close-to-optimal heuristics for executing incoming workloads under specific performance metrics and constraints.
WiSeDB: A Learning-based Workload Management Advisor for Cloud Databases

**Motivation**

- Learning-based approach that “learns” close-to-optimal heuristics specific to workload and SLA from extracted features
- End-to-end solution for query placement, query scheduling and resource provisioning
- Cost-aware workload manager to help cloud DBs meet custom SLAs

**Workload Management**

- Custom goals need custom solutions
  - Hand-crafting heuristics is tedious, ineffective, and non-adaptable
  - State-of-the-art deals with isolated workload management challenges
    - e.g., scheduling, placement, provisioning

- Decision trees navigate the graph step-by-step, resulting in a schedule
- Learns heuristics tailored to the user’s workload

**Adaptive Modeling**

- Enable users to explore cost/performance tradeoffs
- Take a model for one SLA and quickly create a model for a shifted SLA
- If edge weights strictly increase, we can apply Adaptive A*
- Model shift time << model train time

**Results**

- WiSeDB outperforms ideal heuristics for diverse performance goals
- WiSeDB can shift SLAs very quickly, enabling performance/cost exploration
- WiSeDB maintains its performance with heavily skewed workloads