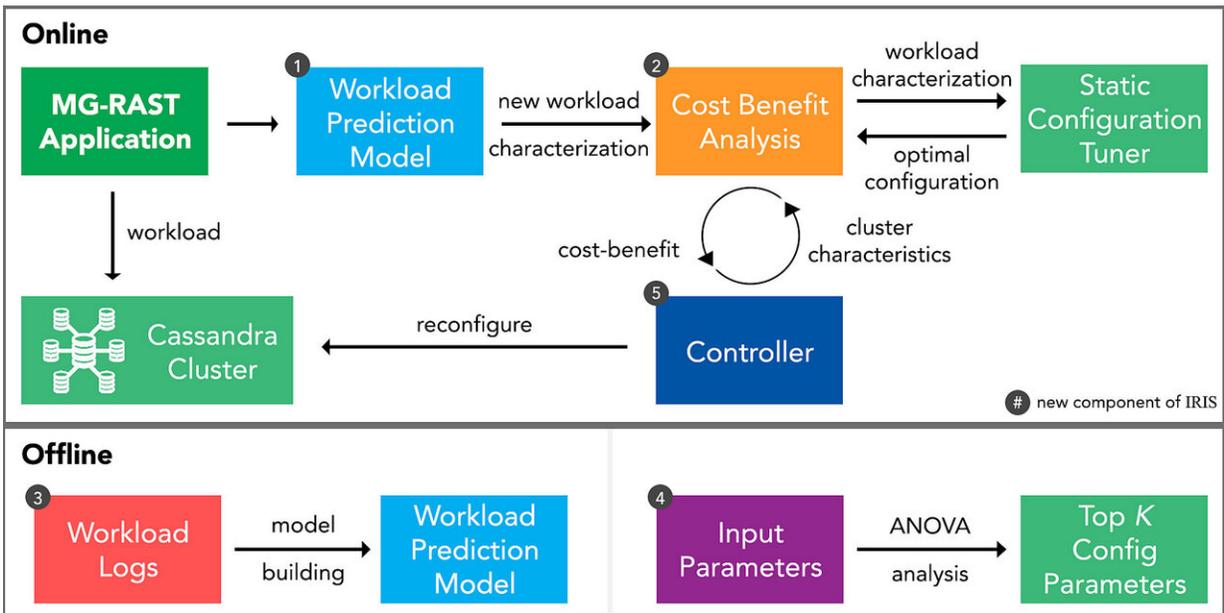


# System designed to improve database performance for healthcare, IoT

May 6 2020



Workflow of SOPHIA with offline model building and the online operation. It also demonstrates the interactions of the noSQL cluster and a static configuration tuner called Rafiki. Credit: Purdue University

Sometimes it is best to work smarter and not harder. The same holds true when it comes to peak performance for databases.

One of the big challenges for using databases—whether for health care, Internet of Things or other data-intensive applications—is that higher

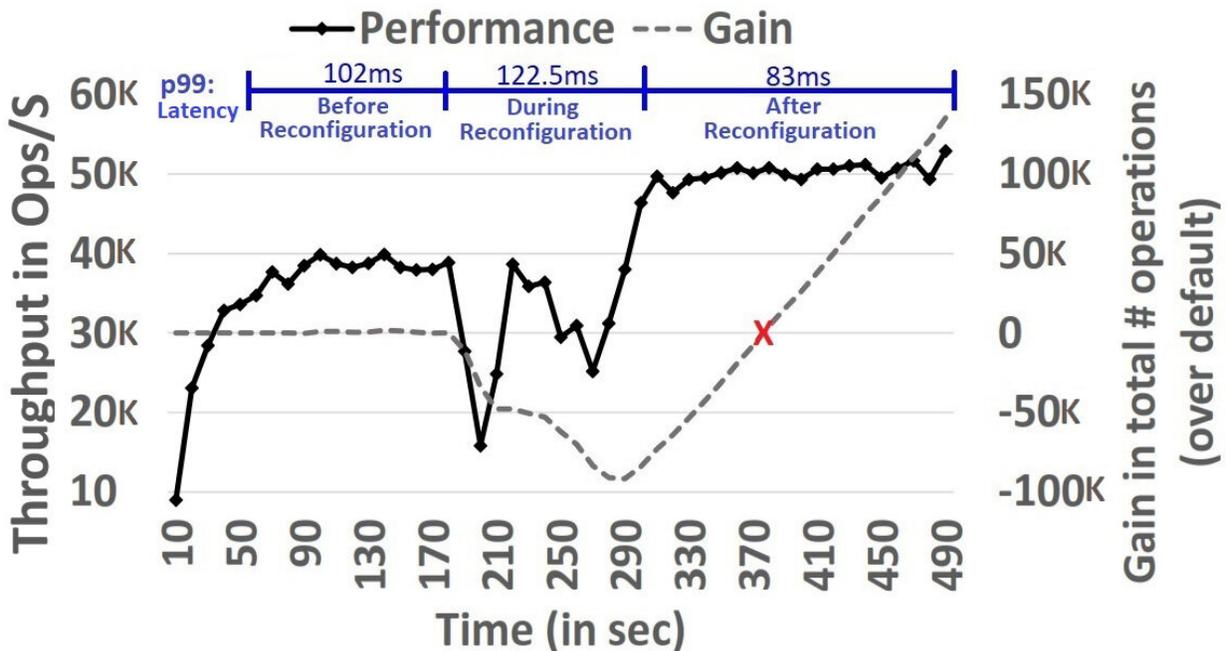
speeds come at a cost of higher operating costs, leading to over-provisioning of data centers for high data availability and database performance.

With higher data volumes, databases may queue workloads, such as reads and writes, and not be able to yield stable and predictable performance, which may be a deal-breaker for critical autonomous systems in smart cities or in the military.

A team of computer scientists from Purdue University has created a system, called SOPHIA, designed to help users reconfigure databases for optimal performance with time-varying workloads and for diverse applications ranging from metagenomics to high-performance computing (HPC) to IoT, where high-throughput, resilient databases are critical.

The Purdue team presented the SOPHIA technology at the 2019 USENIX Annual Technical Conference.

"You have to look before you leap when it comes to databases," said Somali Chaterji, a Purdue assistant professor of agricultural and biological engineering, who directs the Innovatory for Cells and Neural Machines [ICAN] and led the paper. "You don't want to be a systems administrator who constantly changes the database's configuration parameters, naïvely, with a parameter space of more than 50 performance-sensitive and often interdependent parameters, because there is a performance cost to the reconfiguration step. That is where SOPHIA's cost-benefit analyzer comes into play, as it performs reconfiguration of noSQL databases only when the benefit outweighs the cost of the reconfiguration."



The effect of reconfiguration on the performance of the system. SOPHIA uses the workload duration information to estimate the cost and benefit of each reconfiguration step and generates plans that are globally beneficial. Credit: Purdue University

Purdue's SOPHIA system has three components: a workload predictor, a cost-benefit analyzer, and a decentralized reconfiguration protocol that is aware of the data availability requirements of the organization.

"Our three components work together to understand the workload for a database and then performs a cost-benefit analysis to achieve optimized performance in the face of dynamic workloads that are changing frequently," said Saurabh Bagchi, a Purdue professor of electrical and computer engineering and computer science (by courtesy). "The final component then takes all of that information to determine the best times to reconfigure the database parameters to achieve maximum success."

The Purdue team benchmarked the technology using Cassandra and Redis, two well-known noSQL databases, a major class of databases that is widely used to support application areas such as social networks and streaming audio-video content.

"Redis is a special class of noSQL databases in that it is an in-memory key-value data structure store, albeit with hard disk persistence for durability," Chaterji said. "So, with Redis, SOPHIA can serve as a way to bring back the deprecated virtual memory feature of Redis, which will allow for data volumes bigger than the machine's RAM."

The lead developer on the project is Ashraf Mahgoub, a Ph.D. student in computer science. This summer he will go back for an internship with Microsoft Research, and when he returns this fall, he will continue to work on more optimization techniques for cloud-hosted databases.

The Purdue team's testing showed that SOPHIA achieved significant benefit over both default and static-optimized database configurations. This benefit stays even when there is significant uncertainty in predicting the exact job characteristics.

The work also showed that Cassandra could be used in preference to the recent popular drop-in ScyllaDB, an auto-tuning [database](#), with higher throughput across the entire range of [workload](#) types, as long as a dynamic tuner, such as SOPHIA, is overlaid on top of Cassandra.

SOPHIA was tested with MG-RAST, a metagenomics platform for microbiome data; high-performance computing workloads; and IoT workloads for digital agriculture and self-driving cars.

Provided by Purdue University

Citation: System designed to improve database performance for healthcare, IoT (2020, May 6)  
retrieved 19 September 2024 from

<https://techxplore.com/news/2020-05-database-healthcare-iot.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.