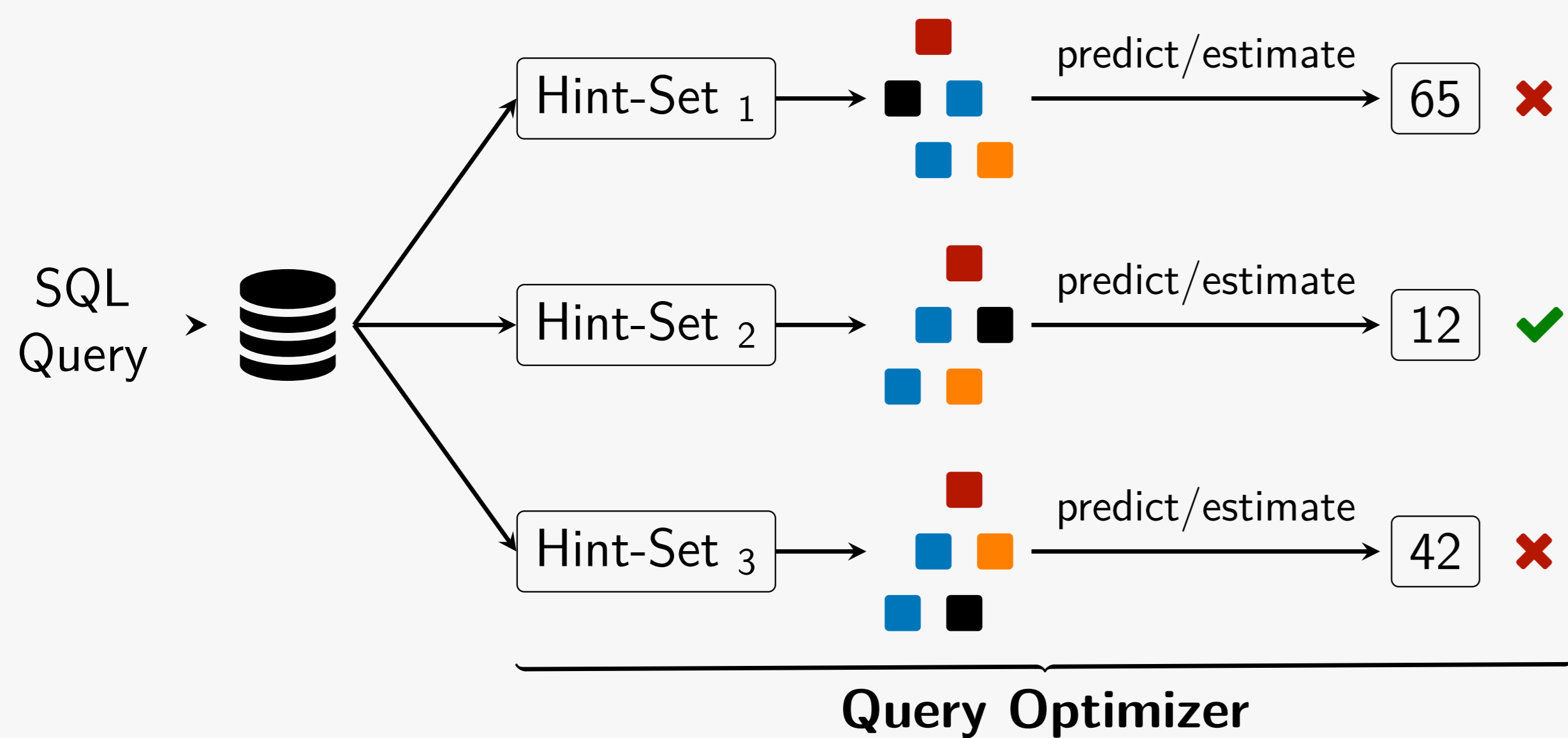


QO-Insight – Inspecting Steered Query Optimizers

Christoph Anneser[†] Mario Petrucci[†] Nesime Tatbul^{°‡} David Cohen[°]
 Zhenggang Xu[∞] Prithviraj Pandian[∞] Nikolay Laptev[∞] Ryan Marcus[◇] Alfons Kemper[†]
[†]Technical University of Munich [°]Intel [‡]MIT [∞]Meta [◇]University of Pennsylvania

Background – Steered Query Optimizers

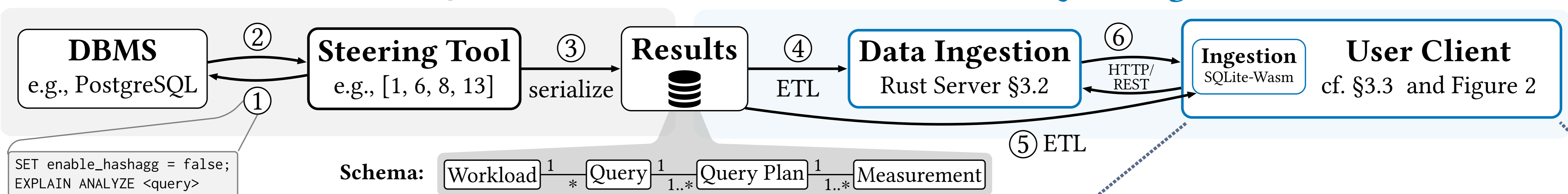


- ▶ Database systems **expose knobs** that can be used to steer query execution. For example, PostgreSQL has knobs to disable nested loop joins or index scans.
- ▶ **Hint-sets** can combine multiple knobs. For example: `{indexscan:false, nestloop:false}`.
- ▶ Recent work on steered query optimizers either adaptively selects [1], predefines [6] or randomly chooses [8, 13] multiple hint-sets, which are used to **generate alternative query plans**.
- ▶ A **deep neural network predicts** the execution time of each plan.

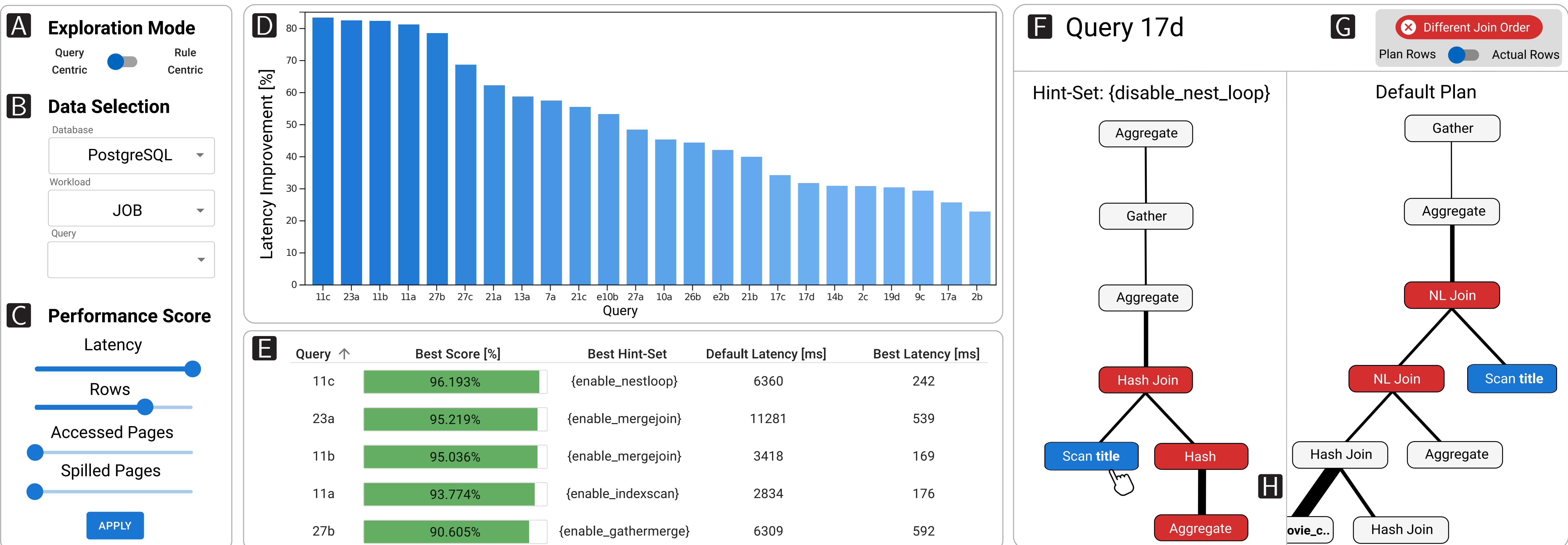
[1] Anneser et al.: “AutoSteer: Learned Query Optimization for Any SQL Database” (VLDB’23)
 [6] Marcus et al.: “Bao: Making Learned Query Optimization Practical” (SIGMOD’21)
 [8] Negi et al.: “Steering Query Optimizers: A Practical Take on Big Data Workloads” (SIGMOD’21)
 [13] Zhang et al.: “Deploying a Steered Query Optimizer in Production at Microsoft” (SIGMOD’22)

Data Generation §3.1

QO-Insight



QO-Insight



A Users can switch between the query-centric mode for database admins and the rule-centric mode for query optimization experts.

B The data selection component allows users to define what data they want to explore.

C QO-Insight supports user-defined performance scores to evaluate query plans.

D A bar chart visualizes the performance improvements of the selected data according to the performance score.

E The table shows more detailed information than the bar chart, is adaptive, and enables the users to export the data.

F Users can select two query plans by clicking on the bars, opening the widget in

full-screen mode and showing the two query plans side-by-side. A custom algorithm finds the matching nodes and highlights the difference between the two query plans.

G Show actual or estimated cardinalities.

I Upload custom trace files and process them entirely within the browser.

Scenario I: Database Admins

Goal: Tune the database for a custom workload or query

- ▶ Switch to **query-centric** mode in **A**
- ▶ Adapt the performance metric in **C**
- ▶ Sort the table in **E** by the performance score
- ▶ Select the best hint-set for the query and apply it

Scenario II: Query Optimization Experts

Goal: Improve the query optimizer’s implementation

- ▶ Switch to **rule-centric** mode in **A**
- ▶ Click on two bars in the bar chart and open **F**
- ▶ Compare the two query plans
- ▶ Improve the query optimizer implementation