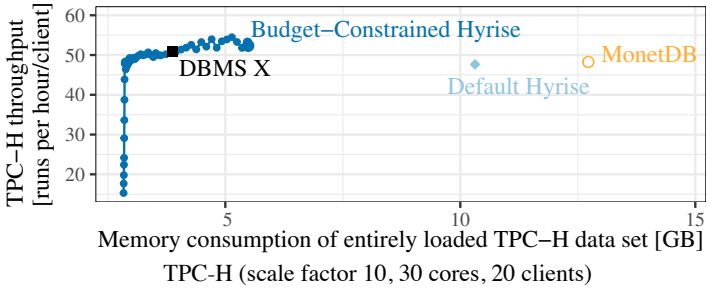
Robust and Budget-Constrained Encoding Configurations

To appear in PVLDB 15 Issue 4

Martin Boissier. Robust and Budget-Constrained Encoding Configurations for In-Memory Database Systems. PVLDB, 15(4): 780 - 793, 2022.

In-Memory Database Systems

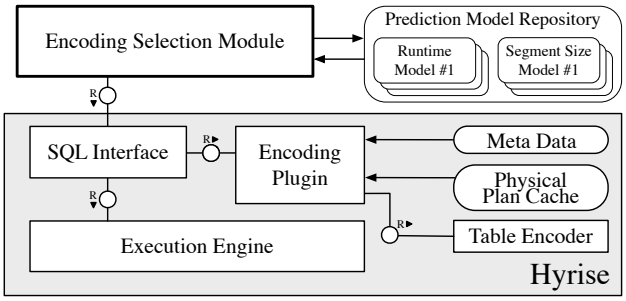
- For in-memory database systems, **memory consumption is a significant cost-driver,** especially in cloud deployments.
- Our goal is a **workload-driven selection of encoding schemes** that exploits workload patterns to reduce memory consumption and improve runtime performance.



TPC-H (scale factor 10, 30 cores, 20 clients)

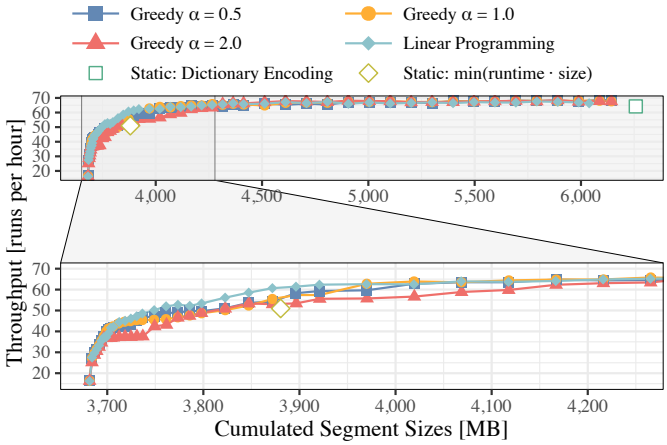
Encoding Selection

- To estimate sizes and query runtimes for a given encoding configuration, we use **linear regression models** that are robust, efficient to train, and can be applied to out-of-sample data (e.g., larger data sets).
- We propose integer linear programming to determine **optimal encoding configurations** as well as efficient heuristics for very large problems.
- We extend the linear programming model to **incorporate robustness constraints** that mitigate unexpected performance regressions and allow to avoid SLA violations.



Results

- The models accurately predict sizes and runtimes of arbitrary encoding configurations.
- For the Join Order Benchmark, TPC-H, and TPC-DS, optimized encoding configurations can **improve runtime performance** while **decreasing memory consumption** over state-of-the-art dictionary encoding.
- The proposed heuristics yield competitive performance and provide an efficient alternative without requiring solvers.



TPC-DS (scale factor 10, single-threaded)

Contact:
Martin Boissier
martin.boissier@hpi.de

Hyrise:
hyrise/hyrise
@hyrise_db