SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY — INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Information Systems

Visualization and statistical analysis of performance measurements of database systems

Julian Macias De La Rosa

SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY — INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Information Systems

Visualization and statistical analysis of performance measurements of database systems

Visualisierung und statistische Aufbereitung von Performance Messungen von Datenbanksystemen

Author:	Julian Macias De La Rosa
Supervisor:	Supervisor
Advisor:	Maximilian Bandle
Submission Date:	20.08.2023

I confirm that this master's thesis is my own work and I have documented all sources and material used.

Munich, 20.08.2023 Julian Macias De La Rosa

Acknowledgments

Abstract

Contents

1 Introduction

Interactive performance visualization is a powerfull skill and plays a vital role for the demonstration of meaningful data insights in the context of performance measurements. Our goal is to use this powerfull skill properly to enable potential optimization possibilities for compiling database systems.

1.1 Background and motivation

1.2 Existing Visualization of Performance Data of Umbra //PDF

1.3 Research objectives

1.4 Scope and contribution of the thesis

1.5 Thesis structure

2 Related Work

In this chapter, we give an overview about the existing work in the domain of the visualization of database performance profiling. We will investigate the importance of optimizing query executions in database systems and the role of visualizations in identifying potential improvements. As performant measurement and analysis play a crucial role in developing and optimizing database systems, it is essential to examine the state-of-the-art techniques and tools that have been used in this domain. We will also cover a visualization tool closely associated with this thesis, as its key feature is integrated into the Benchy Viewer.

2.1 Database Performance Profiling

Performance profiling in database systems is crucial for optimizing their execution regarding achieving optimal hardware utilization and query efficiency. Profiling the performance of database systems involves collecting and analyzing various performance metrics during query execution.
Besides profilers presenting results at the instruction and function granularity, a paper on "Profiling Dataflow Systems on Multiple Abstraction Levels" [Bei+21] proposes a solution that tracks the code generation process and aggregates profiling data to higher abstraction levels. This approach helps bridging the semantic gap between low-level profiles and high-level constructs, making it easier for developers to interpret profiling results and identify bottlenecks and hotspots in the system. The paper introduces the concept of Tailored Profiling, which extends the compilation steps to annotate the generated code with metadata. This enables the mapping of profiling results back to desired abstraction levels and provides more understandable profiling data. Building on the insights from this work the opportunity arises to create more meaningful visualizations regarding the dataflow in system performance profiling.
An essential concept of this thesis is to build upon the concepts of tailored profiling to gain a deeper understanding of the system's performance and support the location of potential optimization possibilities. Thus, we integrate an intuitive and interactive query plan visualization feature that is able to break down complex queries into their constituent operators and pipelines. We clarify further details about the query plan in section 2.2.1 and in chapter X (Implementation) Todo: Chapter linking.

2.2 Related visualization tools

This section explores related visualization tools that aid developers analyse their database system queries, with a specific focus on performance visualization. We will go through the Query Plan Difference Visualizer and the Umbra Profiler Todo: Zitat, which are both tools, that are strongly related to the Benchy Viewer.

2.2.1 Query Plan Difference Visualizer

The Query Plan Difference Visualizer is a web application that compares and visualizes physical query execution plans from different relational database systems, as shown in Figure 2.1.

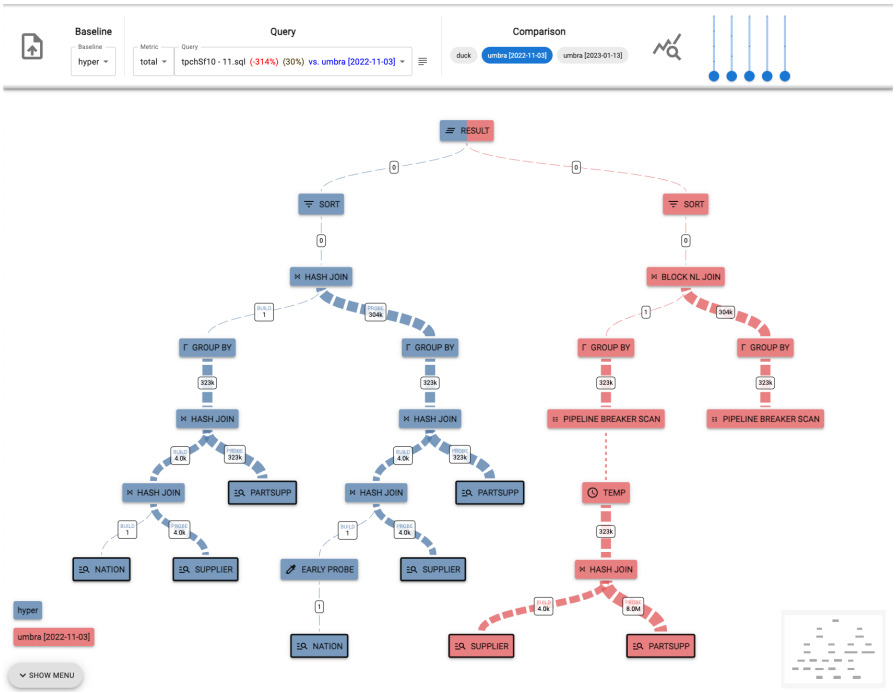

Figure 2.1: Query Plan Difference Visualizer

The comparative tool is designed for database developers who want to inspect the correlation between variations in query execution speed and the respective query plans. Through enhanced hierarchical differencing algorithms with semantic information about query plans, the tool is able to capture and presents interacively the difference between query plans.
This is particularly useful for comparing different database systems or different versions

of the same system, when varying queryplans are used by the systems under test to tackle the given query.
This tool is strongly related to this thesis since the Benchy Viewer builds upon the Query Plan Difference Visualizer by integrating its core features and providing the ability to compare queryplans. We will dive deeper into the integration in Chapter X.
Todo: Chapter linking

2.2.2 Umbra Profiler

The Umbra Profiler is a tool that simplifies the analysis of performance profiles of the query runtime compiled in the database system Umbra.
It offers multiple perspectives, including a runtime dashboard, memory dashboard and an instruction dashboard, each depicting distinct information, as illustrated in Figure 2.2.

Figure 2.2: Umbra-Profiler: Tool for analyzing and profiling Umbra's compiling queries. From left-to-right: Runtime dashboard, memory dashboard, and instruction dashboard.

Similar to the Benchy Viewer, the goal is to support database engineers in optimizing query execution by providing an interactive user interface, enabling an effective in-depth analysis process.
The Umbra Profiler is designed based on the innovative Tailored Profiling approach [Bei+21], where the connection between query plans and compiled code is maintained. This technique was previously unadressed by standard profilers and is now used in both the Umbra Profiler and the Benchy Viewer.
Mehr über Funktionen von Umbra Profiler beschreiben, z.B welche Diagramme gibt es, Metriken, etc. / Was gibt es, was es beim Benchy Viewer nicht gibt...
However, unlike the Benchy Viewer, the Umbra Profiler is constrained to operate exclusively with the database system Umbra. In contrast, the Benchy Viewer has the versatility to function with multiple database systems or multiple instances of a single database system. Ist der Satz inhaltlich korrekt?
In broad terms, the Umbra Profiler is primarily designed for in-depth analysis of query performance within a single database system, while the Benchy Viewer is

oriented towards its comparative function, enabling the comparison of queries executed by different instances. This comparative approach enhances the understanding of differences between instances.

3 Theoretical Foundations

3.1 Technical Background

3.1.1 React

3.1.2 Redux

3.1.3 React Sweet State//

3.1.4 Plotly

3.1.5 React-Flow

3.1.6 Material UI

4 Example

4.1 Section

Citation test [Lam94].

Acronyms must be added in `main.tex` and are referenced using macros. The first occurrence is automatically replaced with the long version of the acronym, while all subsequent usages use the abbreviation.

E.g. `\ac{TUM}`, `\ac{TUM}` ⇒ Technical University of Munich (TUM), TUM

For more details, see the documentation of the `acronym` package[1].

4.1.1 Subsection

See Table 4.1, Figure 4.1, Figure 4.2, Figure 4.3.

Table 4.1: An example for a simple table.

A	B	C	D
1	2	1	2
2	3	2	3

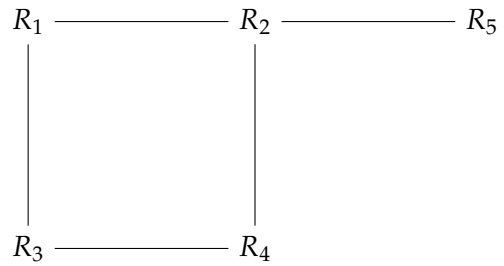

Figure 4.1: An example for a simple drawing.

[1] `https://ctan.org/pkg/acronym`

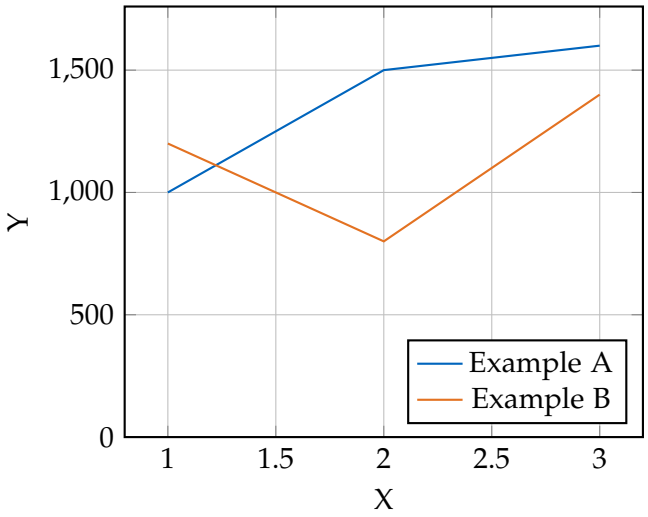

Figure 4.2: An example for a simple plot.

```
SELECT * FROM tbl WHERE tbl.str = "str"
```

Figure 4.3: An example for a source code listing.

Abbreviations

TUM Technical University of Munich

List of Figures

List of Tables

Bibliography

[Bei+21] A. Beischl, T. Kersten, M. Bandle, J. Giceva, and T. Neumann. "Profiling dataflow systems on multiple abstraction levels." In: Apr. 2021, pp. 474–489. DOI: 10.1145/3447786.3456254.

[Lam94] L. Lamport. *LaTeX : A Documentation Preparation System User's Guide and Reference Manual*. Addison-Wesley Professional, 1994.