Enterprise Platform and Integration Concepts
Research Group of Prof. Dr. h.c. Hasso Plattner

Master's Project Summer Term 2020

Performance Engineering for Cloud-based Database Systems

Description

Database systems provide applications with easy, reliable, and high-performance access to their data. They are complex software systems and as such require developers to have a deep understanding of their inner workings and performance. Performance profiling approaches for local and single box database systems already are quite sophisticated. In cloud environments, database systems profiling becomes a real challenge.

To benefit from the cost efficiency and elasticity of public cloud infrastructures, DBMSs are deployed in a virtualized and distributed fashion. This complicates database component instrumentalization and the collection of resulting performance-relevant data for analysis.

As a sister project to our established in-memory research DBMS Hyrise, the EPIC group now builds the Skyrise cloud-based database system. For the beginning of the summer term, we expect to have an early version of its query engine with basic execution operators in place. The query operators are implemented as cloud functions to be run in function services, such as AWS Lambda or Microsoft Azure Functions.

In this project, we aim to identify and alleviate bottlenecks in Skyrise's query engine and to inform its ongoing design. Project goals include:
- Extension of our microbenchmark framework for measuring the performance of low-level, database-related operations, e.g., reads and writes to remote cloud storage. These operations form the foundation for our query operator implementations.
- Build a component for automatic metrics collection and storage. Public cloud providers offer monitoring services with a rich set of APIs to produce interesting metrics that we could not get from within cloud functions, e.g., to understand the functions' life cycle and concurrency behavior.
- Analyze metrics collected from query execution runs to identify existing bottlenecks.
- Improve Skyrise's query operator implementations based on our findings.

Above goals may be addressed largely independently. We will select goals depending on the number of students, their interests, and our progress during the project.

To facilitate the development of Skyrise, we have a tool chain for both local code execution on your notebook and remote execution in AWS. We further offer you continuous support by the Skyrise development team.

After this project, there will be research opportunities to dive deeper into identified issues in the form of Master's theses.

Learning Goals

Through successful completion of this project, you will:
- Improve your programming and teamwork skills
- Gain hands-on experience with modern cloud infrastructure
- Learn to observe complex and distributed cloud-based software systems
- Learn to pinpoint and remove performance bottlenecks in this setting
- Deepen your database knowledge

Prerequisites

Prior knowledge of the fundamentals of database systems and the C++ programming language is beneficial but not required. Amongst others, the following courses are relevant.
- Datenbanksysteme Lecture
- Big Data Systems Lecture
- Develop your own Database Seminar
- Methods of Cloud Computing Lecture

Resources

- Hellerstein et al. Serverless Computing: One Step Forward, Two Steps Back, CIDR 2019
- Wang et al. Peeking Behind the Curtains of Serverless Platforms, USENIX ATC 2018
- Pu et al. Shuffling, Fast and Slow: Scalable Analytics on Serverless Infrastructure, USENIX NSDI 2019
- AWS Lambda Developer Guide
- AWS SDK for C++ Developer Guide

Organization

- Master's programs ITSE and DE
- Programming project including design, implementation, test and documentation
- Group work
- Midterm and final presentations

Contact

You are welcome to contact us via mail or visit us in the villa.
- Dr. Matthias Uflacker
- Thomas Bodner