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

Master's Project Summer Term 2021

Building an Elastic Query Engine on Serverless Cloud Infrastructure

Description

Analytics workloads are often both *interactive*, serving user-facing applications, and *bursty*, leaving pre-provisioned resources idle much of the time. They require database systems to be elastically scalable to achieve sufficient performance and cost efficiency. Serverless cloud infrastructure, such as Function-as-a-Service platforms or object storage services, promises ultimate elasticity with its fine-grained resource allocation and billing.

The EPIC research group is building the Skyrise cloud-based database system on serverless infrastructure components. 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.

In this project, we aim to build out the Skyrise query engine to better cover the capabilities needed to run the widely used TPC-H benchmark for comparing analytical database systems. The project goals include:

- Extend our benchmark framework with operator microbenchmarks.
- Improve the memory management and concurrency in operator implementations.
- Analyze the TPC-H benchmark and identify missing operator capabilities.
- Extend the exchange operator to more topologies and partitioning schemes.
- Grouping, sorting, joins..

Above goals may be addressed largely independently. We will select goals depending on the number of student participants, 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 the AWS public cloud. 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
- Improve your research methodology and academic writing
- Gain hands-on experience with modern cloud infrastructure
- Learn to design and implement efficient and scalable cloud-based software systems
- Deepen your database knowledge

Prerequisites

Prior knowledge of the internals of database systems and the C++ programming language is required. Experience using public cloud infrastructure is beneficial. Amongst others, the following courses are relevant:

- Datenbanksysteme II 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
- Perron et al. Starling: A Scalable Query Engine on Cloud Functions, SIGMOD 2020
- Boncz et al. TPC-H Analyzed: Hidden Messages and Lessons Learned from an Influential Benchmark, TPCTC 2013
- AWS Lambda Developer Guide
- AWS SDK for C++ Developer Guide

Organization

- Master's programs: ITSE, DE
- Extent: 8 SWS / 12 ECTS
- Content:
    - Group work
    - Programming project
    - Research report
    - Final presentation

Contact

You are welcome to contact us via email.

- Dr. Michael Perscheid
- Thomas Bodner