Analytical Processing using Cloud Functions

Keywords: OLAP, cloud function, query execution, serverless.

Problem: Cloud function is a kind of serverless execution environment on the cloud. Popular examples are AWS Lambda, Google Cloud Functions, and Azure Functions etc. Compared to cloud virtual machines, cloud functions are more suitable for running small, fine granularity, and highly scalable computation tasks. It can greatly reduce the cost for many bursty computation workloads.

At the meanwhile, analytical workloads are often bursty or low volume[1], and require huge amounts of computation and I/O resources during query execution time. Using physical or virtual machines would leave users paying for resources even when underutilized. Due to its better elasticity and a more granular pricing model, cloud function becomes an attractive execution environment for analytical workloads.

This project aims to explore, analyze and implement analytical engines using cloud functions. Existing work like Starling [1] can be used as the baseline.

Project: In this project, the student will implement and evaluate common query operators, including join, scan, projection, and aggregations using cloud functions (i.e. AWS Lambda). The student will also participate in the design and implementation of a new analytical engine based on cloud functions.

Plan:
1. Implement the baseline query operators in [1].
2. Performance evaluation for the baseline operators.
3. Find the possible bottlenecks and design optimized solutions.

Supervisor: Prof. Anastasia Ailamaki, anastasia.ailamaki@epfl.ch

Responsible collaborator(s): Haoqiong Bian

Duration: 6 months

References:
[1] Starling: A Scalable Query Engine on Cloud Functions, Matthew Perron et.al., SIGMOD ’20