DIAS: Data-Intensive Applications and Systems Laboratory
School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne
Building BC, Station 14
CH-1015 Lausanne
URL: http://dias.epfl.ch/

Real-time monitoring of stream processing

Keywords: stream processing, runtime monitoring, real-time analytics

Problem: From social network analytics to gaming, fraud detection, and stock trading, streaming applications require the real-time processing of high-throughput, in-motion data. As streaming applications run continuously, they usually utilize a monitoring component that collects at runtime statistics about the query's performance and health. These components are used by data engineers to keep track of performance metrics as well as by the system itself to optimize its components.

Project: The first goal of this project would be to create a monitoring component for stream processing that monitors not only performance statistics but also information about the input workload and data. For example, it would monitor the current data distribution, and frequent keys. Such a monitoring component can be used to adaptively optimize streaming engines, which is a topic our lab is actively working on [1].
As monitoring can become a bottleneck since it involves updating and storing a set of statistics for every input tuple, the second part of the project will be dedicated to exploring some approximation techniques that make calculating statistics faster (e.g. sketches, sampling).

Plan:

1. Familiarize oneself with background work on stream processing.
2. Implement a monitoring component.
3. Extend the monitoring component to utilize approximation techniques.

Available for: Master students that want to pursue a semester project or master thesis.

References:

1. Eleni Zapridou, Ioannis Mytilinis, and Anastasia Ailamaki. 2022. Dalton: Learned Partitioning for Distributed Data Streams. Proc. VLDB Endow. 16, 3 (November 2022), 491–504. https://doi.org/10.14778/3570690.3570699

Supervisor: Prof. Anastasia Ailamaki, anastasia.ailamaki@epfl.ch
Responsible collaborator(s): Eleni Zapridou