

SABER: Window-Based Hybrid Stream Processing for Heterogeneous Architectures^{*}

Alexandros Koliouisis[†], Matthias Weidlich^{†‡}, Raul Castro Fernandez[†],
Alexander L. Wolf[†], Paolo Costa[‡], and Peter Pietzuch[†]
{akoliouis, mweidlic, rc3011, alw, costa, prp}@imperial.ac.uk

[†]Imperial College London

[‡]Humboldt-Universität zu Berlin

[‡]Microsoft Research

Abstract

Stream processing systems found wide-spread application in domains such as credit fraud detection, urban traffic management, and click stream analytics. These systems process continuous streams of input data in an online manner, aiming at maximising processing *throughput* while staying within acceptable latency bounds. Heterogeneous architectures that combine multi-core CPUs with many-core GPGPUs have the potential to improve the performance of stream processing engines. Yet, a stream processing engine must execute streaming SQL queries with sufficient data-parallelism to fully utilise the available heterogeneous processors, and decide how to use each processor in the most effective way.

Addressing these challenges, we present SABER, a *hybrid* high-performance relational stream processing engine for CPUs and GPGPUs. It executes window-based streaming SQL queries following a hybrid execution model. Specifically, SABER incorporates the following innovations:

- It features a *hybrid stream processing model* based on query tasks, each comprising a batch of stream data and a query operator. Instead of relying on offline performance models to select the processor on which to run a query operator, SABER employs an adaptive *heterogeneous lookahead scheduling* strategy to balance the load on the different types of processors.
- It provides *window-aware task processing*, supporting sliding window semantics in the presence of fixed-sized batches. SABER ensures result correctness after the out-of-order processing of tasks by first buffering and then incrementally releasing the results as tasks finish execution.
- It exploits *pipelined stream data movement* to the GPGPU that interleaves data movement and task execution, thereby maintaining high utilisation of the PCIe bandwidth.

An experimental comparison against state-of-the-art engines shows that SABER increases processing throughput while maintaining low latency for a wide range of streaming SQL queries with both small and large window sizes.

^{*} Published as: A. Koliouisis, M. Weidlich, R. C. Fernandez, A. L. Wolf, P. Costa, and P. Pietzuch. SABER: Window-based hybrid stream processing for heterogeneous architectures. In F. Özcan, G. Koutrika, and S. Madden, editors, Proceedings of the 2016 SIGMOD Conference, San Francisco, CA, USA, June 26 - July 01, 2016, pages 555–569, ACM.