SIGMOD 2019

AStream: Ad-hoc Shared Stream Processing

Jeyhun Karimov[1], Tilmann Rabl[2], Volker Markl[1,3]
[1]DFKI GmbH, [2]HPI, [3]TU Berlin
jeyhun.karimov@dfki.de, tilmann.rabl@hpi.de, volker.markl@tu-berlin.de

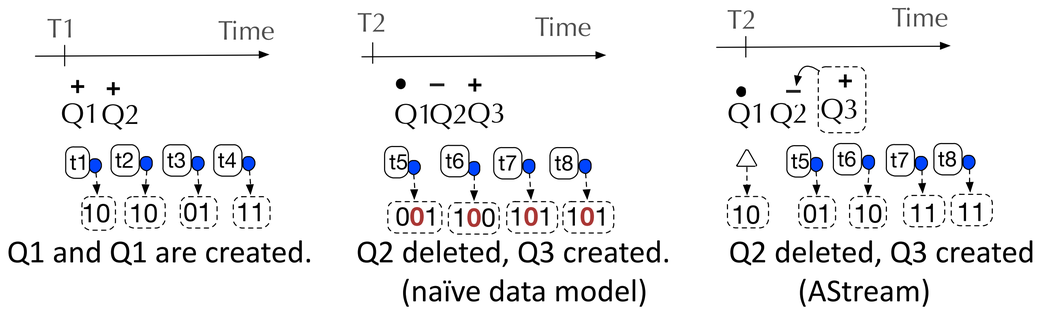
Abstract

The goal of this paper is to bridge the gap between stream processing and ad-hoc queries in SPEs by sharing computation and resources.
We define three main requirements for ad-hoc shared stream processing:

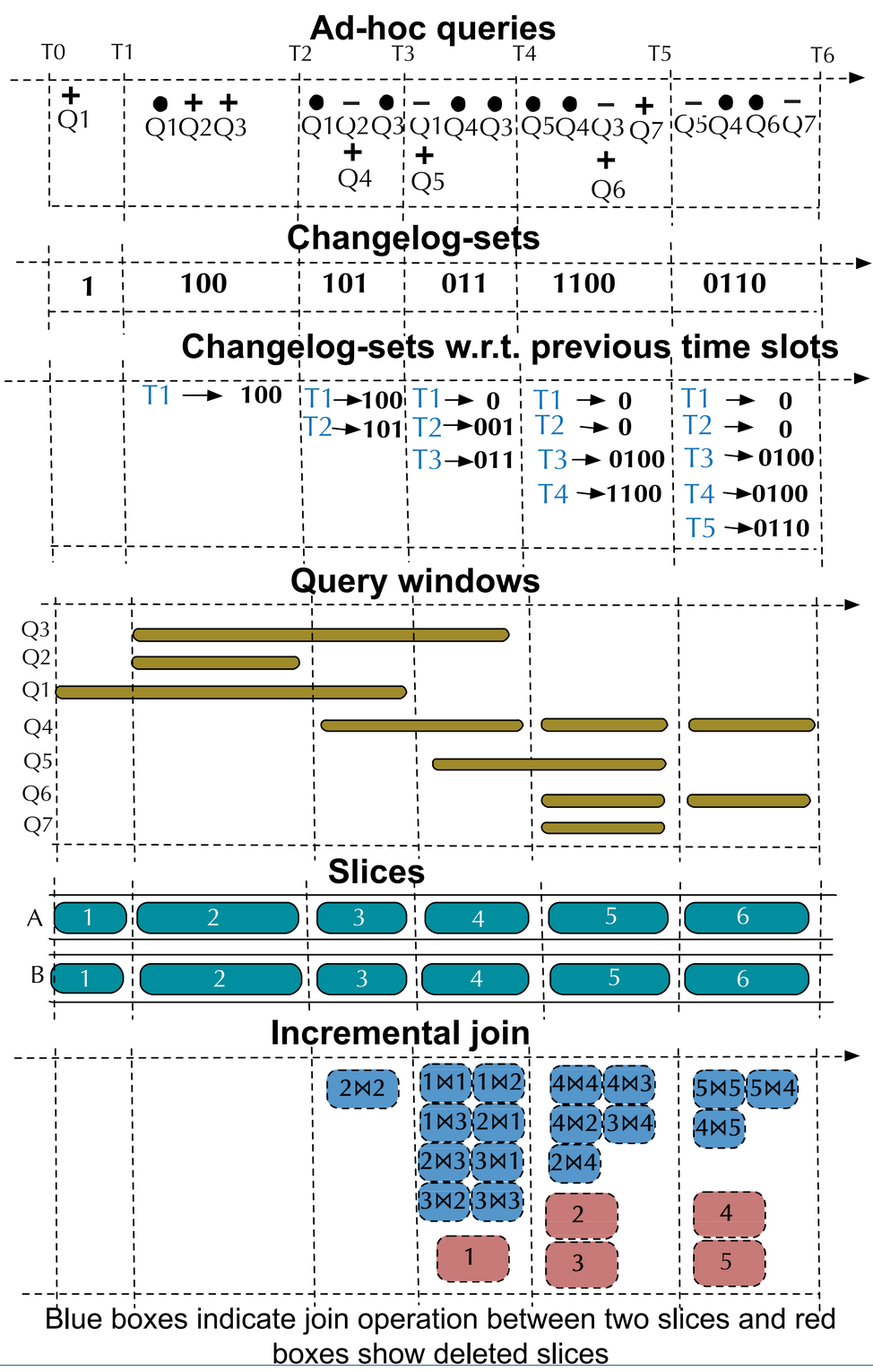
- **Integration**: Ad-hoc query processing should be a composable layer which can extend stream operators, such as join, aggregation, and window operators
- **Consistency**: Ad-hoc query creation and deletion must be performed in a consistent manner and ensure exactly-once semantics and correctness
- **Performance**: In contrast to state-of-the-art SPEs, ad-hoc SPE should not only maximize data throughput but also query throughout via incremental computation and resource sharing

AStream is **the first system** that supports distributed ad-hoc stream processing.
We design AStream based on the requirements listed above.

AStream vs Naïve Data Model

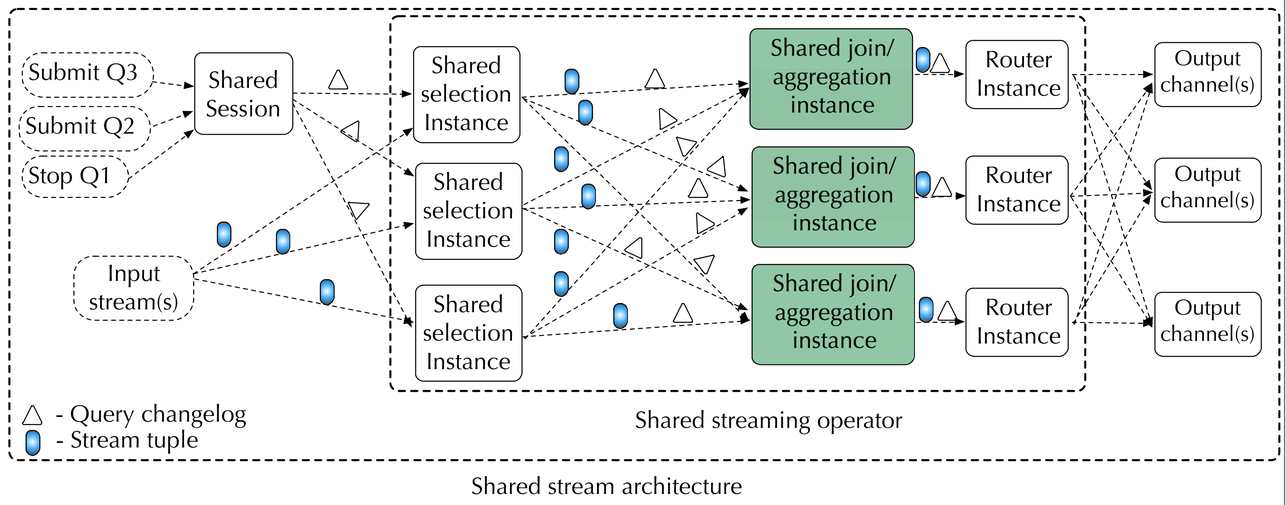


End-to-end ad-hoc query processing



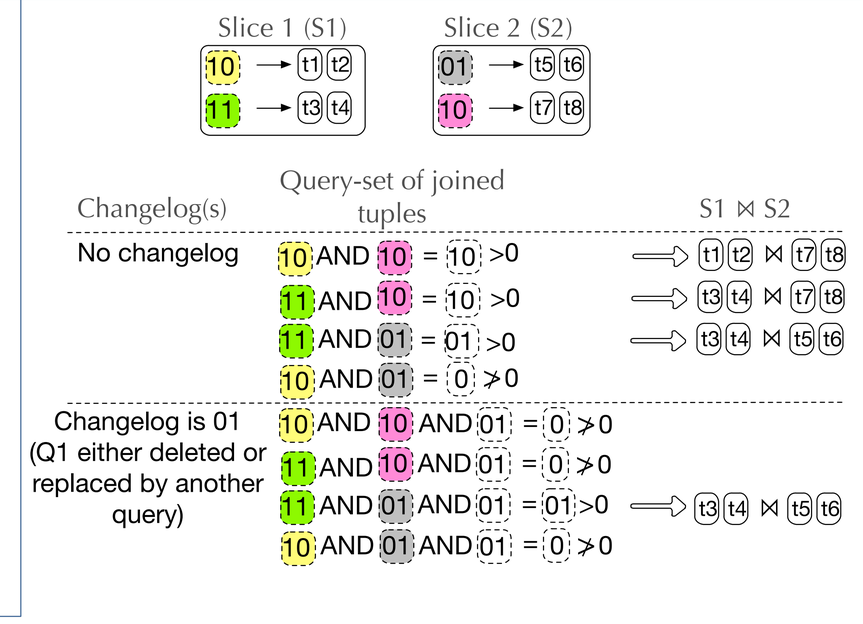
Blue boxes indicate join operation between two slices and red boxes show deleted slices

System Architecture

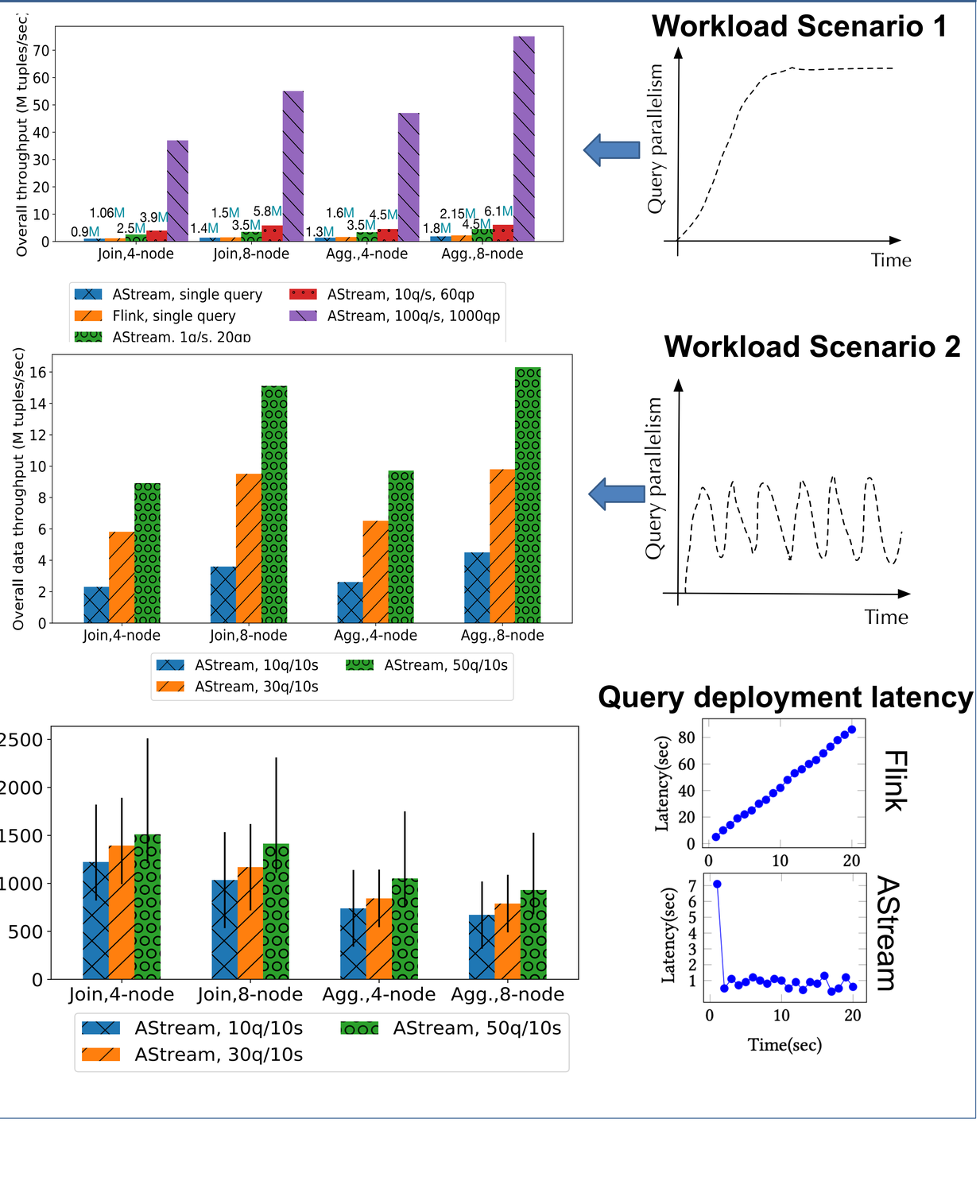


Optimizations

- Incremental query processing
- Memory efficient dynamic slice data structure
- Reduced data copy and shuffling



Experiments



German Research Center for Artificial Intelligence
Hasso Plattner Institut
Technische Universität Berlin
SIGMOD/PODS 2019 AMSTERDAM