

Learned Index for Stream Joins

Keywords: Learned Index, Index Structure, Stream Joins, Parallelism

Problem: Current research indicates that index structures can significantly improve stream window joins; however, existing approaches often lack sensitivity to data distributions. Leveraging data distribution knowledge could notably enhance the performance of joins, as suggested by current learned index research. Therefore, we aim to explore the integration of learned indexes to optimize stream joins, focusing on their potential to dynamically adapt to varying data distributions and efficiently manage multiple, concurrent stream joins. By doing so, we anticipate advancements in both the efficiency and speed of processing real-time data streams.

Project: This research project aims to implement and evaluate learned indexes specifically tailored for stream joins in high-performance computing environments. By integrating learned index structures, the project seeks to significantly reduce the computational overhead associated with traditional indexing methods during stream processing. This approach is expected to enhance the capabilities of stream systems to perform joins over continuous data streams by optimizing both insertion and search operations dynamically.

Plan:

1. *Review and Implement Existing Stream Joins Approaches:* We comprehensively review existing methodologies, such as handshake joins and PIM-trees, and implement these techniques within our testing framework.
2. *Develop and Implement Learned Index for Stream Joins:* Design and implement a learned index specifically tailored for stream joins.
3. *Performance Evaluation and Analysis:* Conduct a thorough performance evaluation of the implemented learned index approach.

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

Responsible collaborator(s): Liang Liang, liang.liang@epfl.ch

Duration: 6-8 months