Zero-Shot Cost Estimation for Distributed Stream Processing

Roman Heinrich, Manisha Luthra,
Harald Kornmayer & Carsten Binnig
DHBW Mannheim & TU Darmstadt

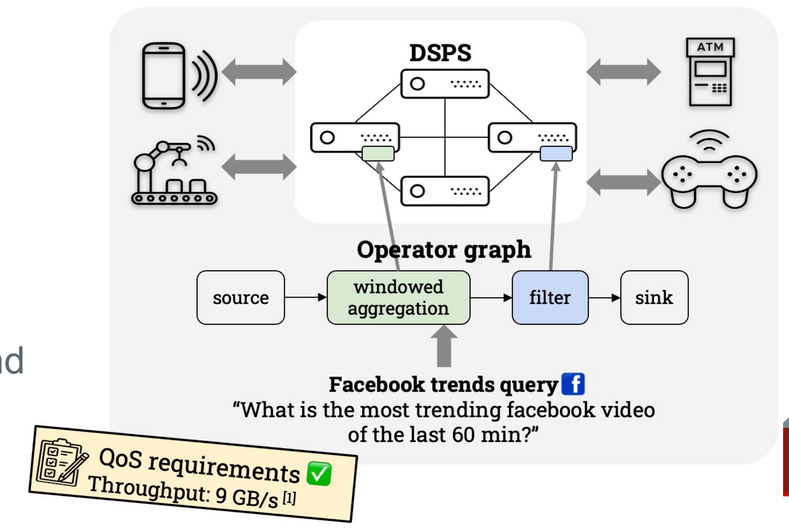
DHBW TRANSFER CONGRESS

1. Overall idea

- Providing an **AI-based model** to predict cost metrics of executing arbitrary queries in Distributed Stream Processing Systems (DSPS) [1]
- The model uses **transferable features** and **Graph Neural Networks** (GNNs)
- The model makes precise estimations for known and even unknown streaming workloads and thus is **generalizable** by the zero-shot learning approach
- The model can be used by cloud providers to help solving **optimization tasks** like finding the placement of streaming operators

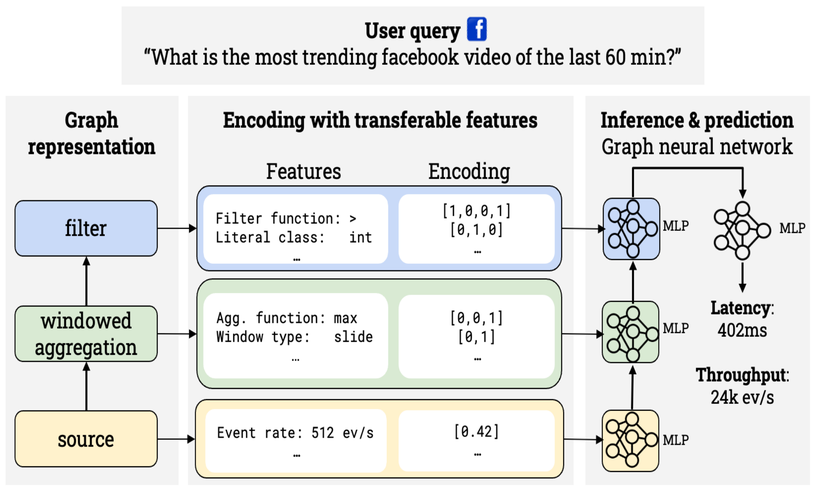
2. Foundations

- Execution of queries on **data streams** are fundamental for modern business applications
- The optimal deployment of query operators has a large impact on the overall **performance**
- These optimization tasks require precise and general cost estimation models
- Current models are **workload-driven**
  → specific for a given use-case or query



3. Our Zero-Shot approach

- In contrast to workload-driven models, we propose a generalizable learned cost model to predict the execution costs of DSPS queries
- We introduce transferable features that can be applied on any operator and data stream
- We train a GNN that combines a set of **multi-layer perceptrons** in a flexible way.
- We make use of **a broad training-dataset** that includes a variety of workloads, operators and hardware



4. Experimental evaluation

- We executed 15.000 Queries on 10 clusters using DSPS Apache Storm[2]:
- By reporting the Q-error:
  $q(c,\hat{c}) = \max(c/\hat{c},\ \hat{c}/c)$
  we evaluated the ability of the model to:
  - A) interpolate for an **unseen test-set**
  - B) extrapolate for **unseen benchmark** queries from DSPBench[3]
  - C) extrapolate for unseen and **extreme operator properties** and streaming workloads
  - D) unseen query structures (see paper)

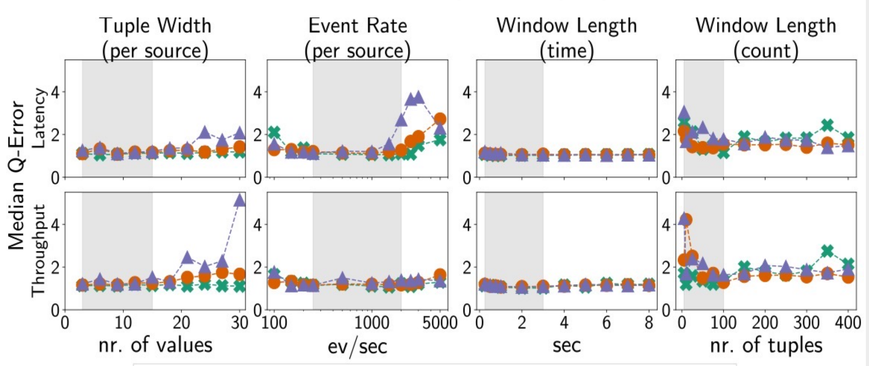
A) Prediction accuracy for unseen test-set

Latency		Throughput	
median	95th	median	95th
1.13	3.19	**1.16**	3.50

B) Prediction accuracy for unseen benchmark

Benchmark	Latency		Throughput	
	median	95th	median	95th
Advertisement (clicks)	**1.51**	1.53	**1.38**	1.39
Advertisement (imp.)	**1.51**	1.52	**1.38**	1.39
Advertisement (join)	**1.99**	2.06	**1.55**	2.16
Spike Detection	**1.01**	1.04	**1.73**	1.94
Smart Grid (local)	**1.21**	1.23	**1.92**	1.92
Smart Grid (global)	**1.20**	1.66	**1.91**	1.91

C) Extrapolation for unseen data streams & operator properties



Results

- We deliver an AI-based model to **predict cost metrics** of executing DSPS queries
- We propose **transferable features** to describe streaming queries
- We applied our model on Apache Storm
- We evaluated our zero-shot model which:
  - provides **very accurate estimates** for queries within the training range
  - predicts cost for queries **beyond the training range** with reasonable precision
  - shows a one-time training effort
- First contact with cloud providers established

Outlook

- **Extend** the model towards inclusion of real heterogeneous hardware, user-defined operators or operator parallelism
- **Apply** the model in optimization tasks like operator placement by combining it with a reinforcement learning approach

Cooperative partners

TECHNISCHE UNIVERSITÄT DARMSTADT

DHBW Duale Hochschule Baden-Württemberg Mannheim — 4C Cloud Computing Competence Center

References

[1] Roman Heinrich, Manisha Luthra, Harald Kornmayer, and Carsten Binnig. 2022. „Zero-shot cost models for distributed stream processing“. In ACM DEBS, 2022, p. 85–90.

[2] A. Toshniwal, S. Taneja, A. Shukla, K. Ramasamy, J. M. Patel, S. Kulkarni, J. Jackson, K. Gade, M. Fu, J. Donham, N. Bhagat, S. Mittal, and D. Ryaboy, “Storm@twitter,” in ACM SIGMOD, 2014, p. 147–156.

[3] M. V. Bordin, D. Griebler, G. Mencagli, C. F. R. Geyer, and L. G. L. Fernandes, “DSP-bench: A suite of benchmark applications for distributed data stream processing systems,” IEEE Access, vol. 8, pp. 222 900–222 917, 2020.

Kontakt

Duale Hochschule Baden-Württemberg

Roman Heinrich
Coblitzallee 9, 68163 Mannheim
roman.heinrich@dhbw-mannheim.de

Alle Informationen finden Sie unter: **www.dhbw.de**