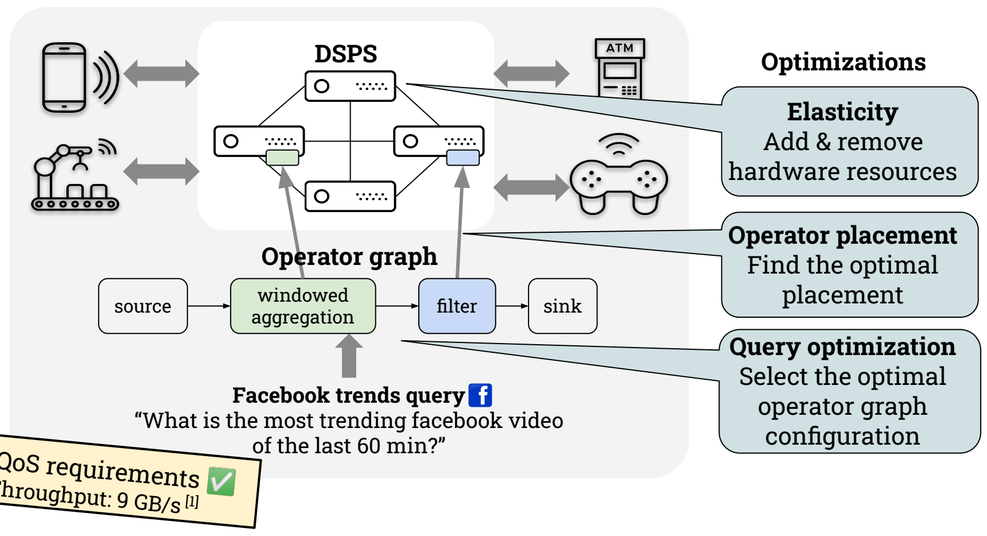
Zero-Shot Cost Models for Distributed Stream Processing

Roman Heinrich[1], Manisha Luthra[2], Harald Kornmayer[1], Carsten Binnig[2]

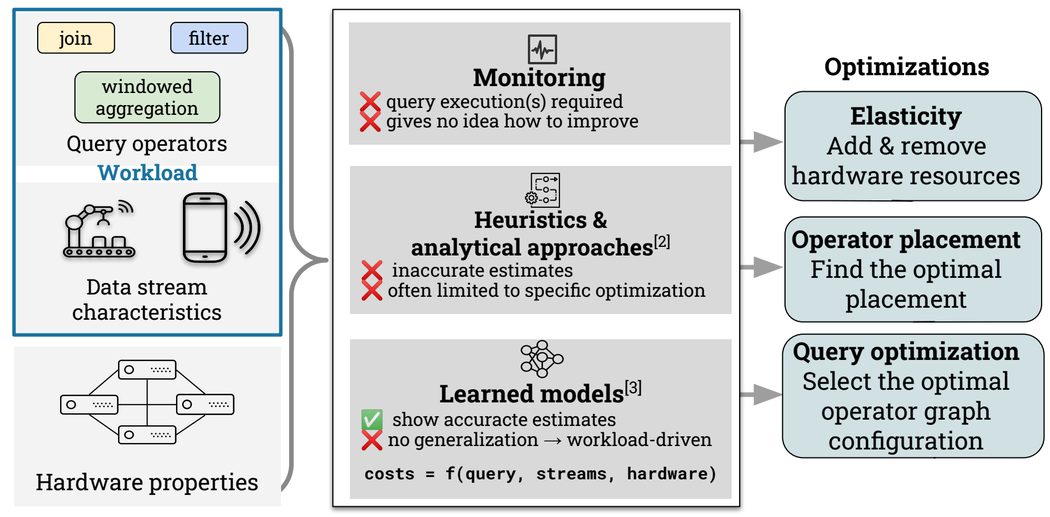
[1] DHBW Mannheim, [2] TU Darmstadt

Distributed and Event-based Systems (DEBS 2022)

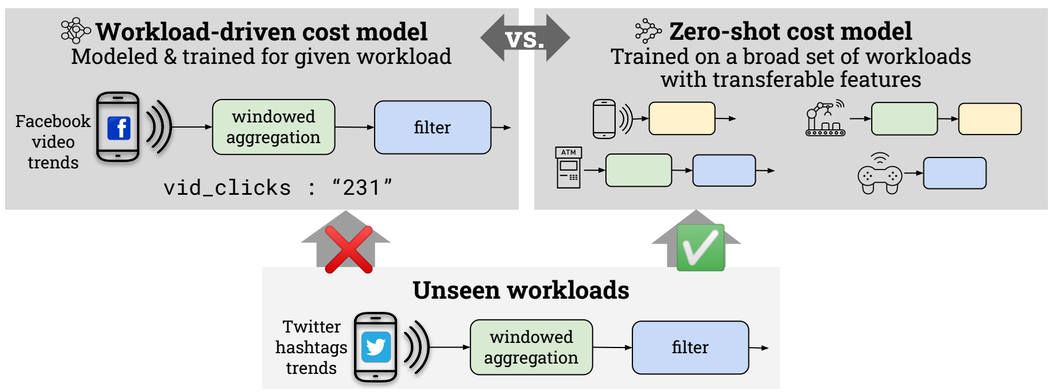
Cost models for DSPS



Limitations of existing approaches



Workload-driven vs. zero-shot models



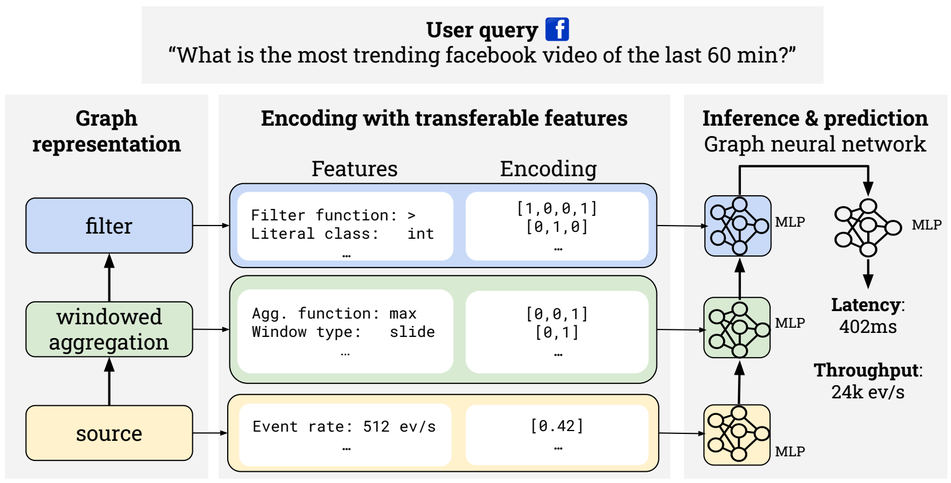
❌ No generalization possible

❌ Costly re-training for a new workload is required

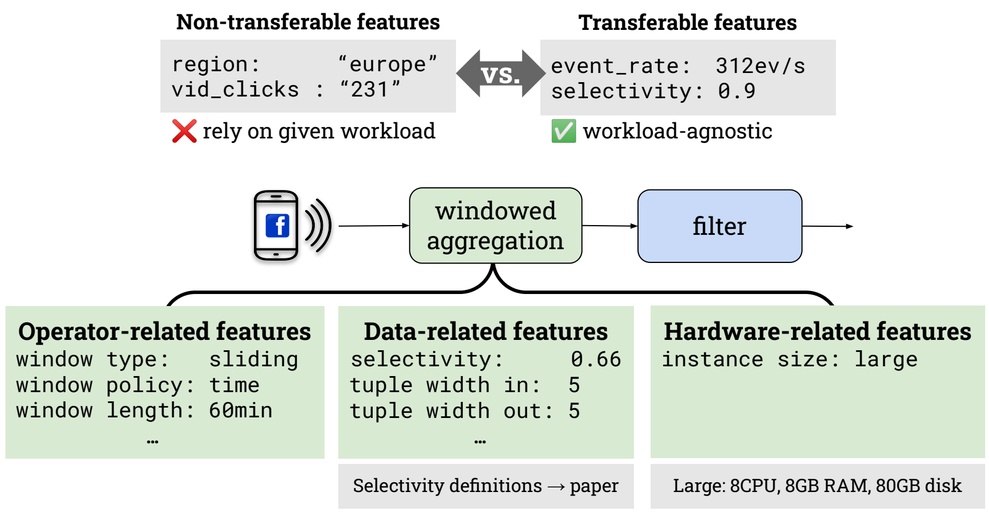
✅ Generalization to unseen streams, queries and hardware

✅ No retraining required

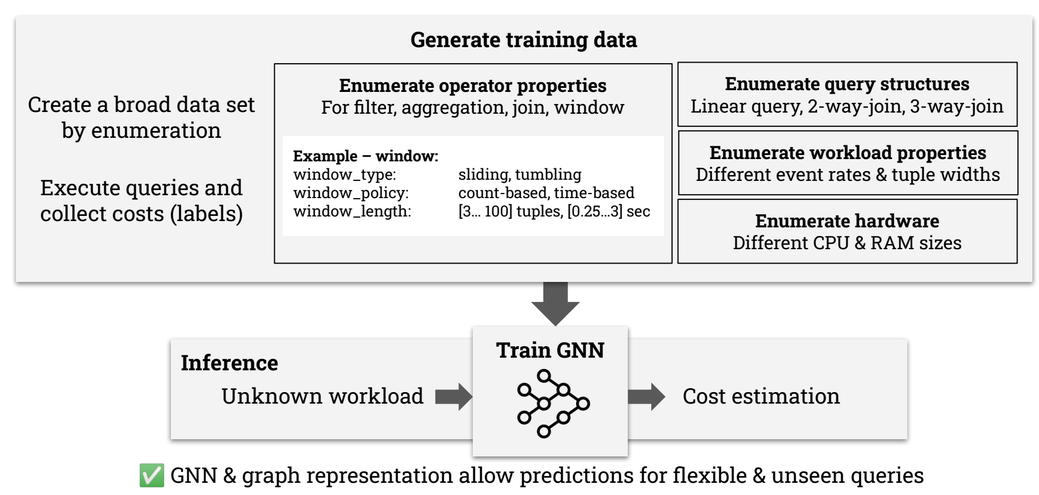
Zero-shot model architecture



Transferable features



Training & inference methodology



✅ GNN & graph representation allow predictions for flexible & unseen queries

Evaluation on zero-shot models

- **Set-up:** 10 clusters (each 10 nodes) with Apache Storm v2.2.0
- **Cost metrics**: end-to-end latency & throughput
- **Metric**: q-error
  `q(c,ĉ)=max(c/ĉ, ĉ/c)`
  - reporting median and 95-percentile
  - q=1: perfect estimate

Interpolation for workloads & placements

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

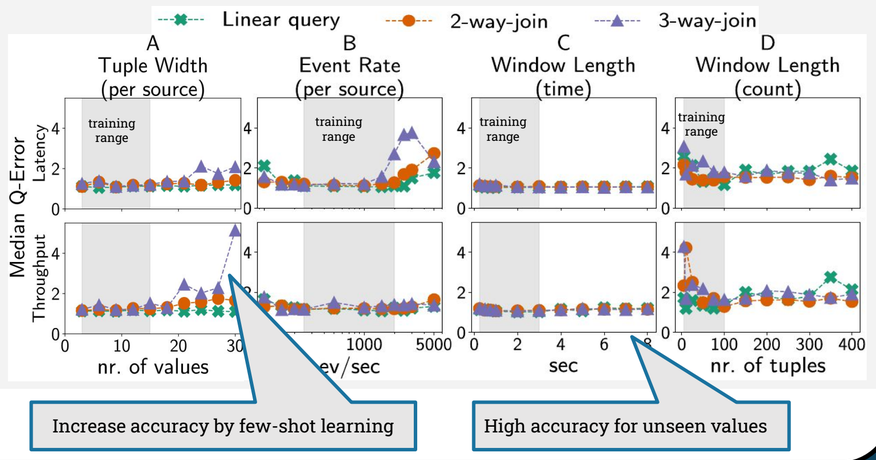
Extrapolation for **unseen benchmarks** (DSPBench[4])

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

Extrapolation for **unseen query structures**

	Latency		Throughput	
	median	95th	median	95th
2-filter chain	**1.14**	2.41	**1.59**	3.65
3-filter chain	**2.67**	46.34	**2.82**	27.78
4-filter chain	**7.33**	54.68	**3.94**	59.73
4-way joins	**1.95**	24.30	**1.33**	20.79
5-way joins	**1.91**	26.67	**1.35**	21.87

Extrapolation for **unseen data streams & operator properties**



Conclusion & outlook

Our zero-shot cost model...

- ...is generalizable and workload independent
- ...requires an one-time training effort
- ...predicts accurately and robustly for seen & unseen workloads

✅ can be used as a main building block in DSPS optimization tasks

Open questions on zero-shot models:

- How to model hardware properties more precisely?
- How to featurize co-location of operators?
- How to make use of the cost model in specific optimization tasks like providing elasticity?

[1] Z. Shao, "Real-time analytics at facebook," XLDB, 2011.
[2] L. Eskandari, J. Mair, Z. Huang, and D. Eyers, "I-scheduler: Iterative scheduling for distributed stream processing systems," Future Generation Computing Systems, vol. 117, pp. 219–233, 2021.
[3] T. Li, Z. Xu, J. Tang, and Y. Wang, "Model-free control for distributed stream data processing using deep reinforcement learning," PVLDB, vol. 11, no. 6, p. 705–718, 2018.
[4] M. V. Bordin, D. Griebler, G. Mencagli, C. F. R. Geyer, and L. G. L. Fernandes, DSPbench: A suite of benchmark applications for distributed data stream processing systems", IEEE Access, vol. 8, pp. 222 900–222 917, 2020.