

Implementation of a Benchmark Suite for Strymon

Nicolas Hafner

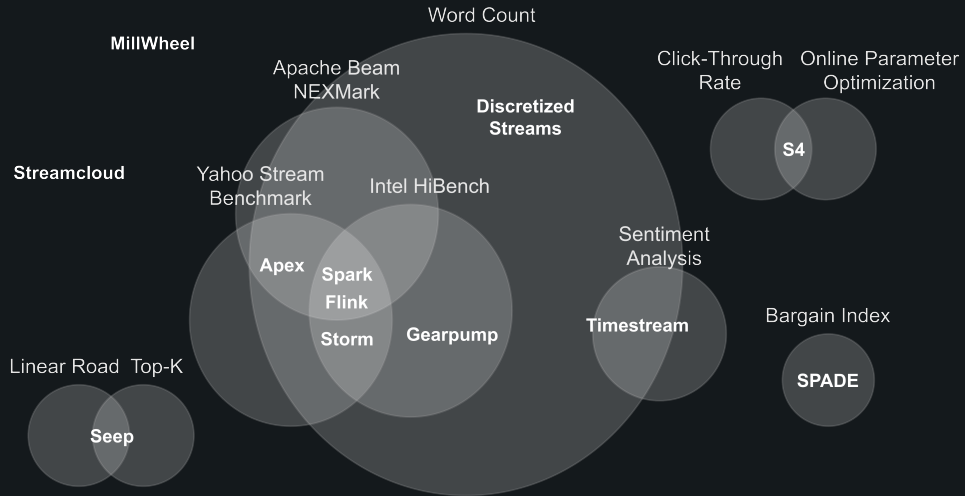
ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

inf |

Informatik
Computer Science



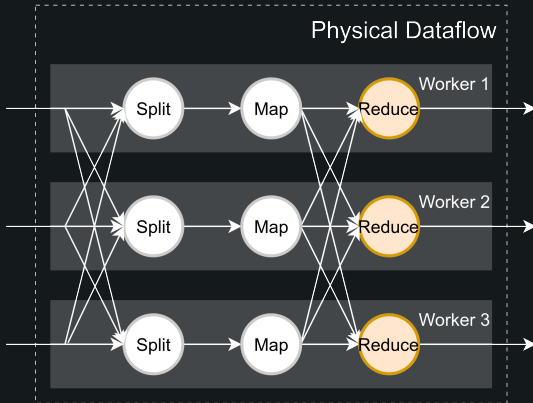
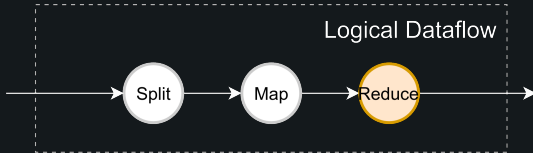


Other Works

- Investigated 9 papers for other systems
- Almost no paper used a standardised benchmark
- Often very simple benchmarks like Word Count
- Code and data often not published



Timely



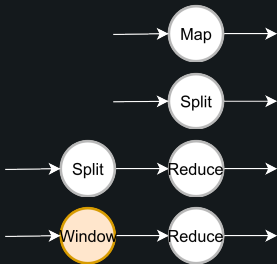
Benchmarks

- We implemented three benchmarks:
 1. Intel's HiBench
 2. Yahoo's Streaming Benchmark
 3. Apache Beam's NEXMark
- Comparable against many other systems

Intel's HiBench

- Big Data micro-benchmark
- Only four latency tests:

1. Identity
2. Repartition
3. Word Count
4. Window Reduce



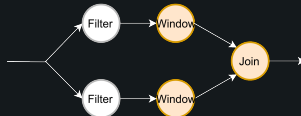
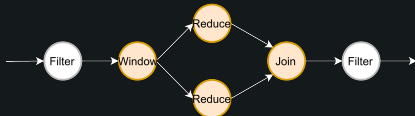
Yahoo Stream Benchmark

- Count ad views for ad campaigns
- Only one, relatively simple data flow:



Beam's NEXMark¹

- Implements an “auctioning system”
- 13 data flows in total
- Uses filter, map, reduce, join, window, session, partition
- Dataflows for Query 5 and 8:



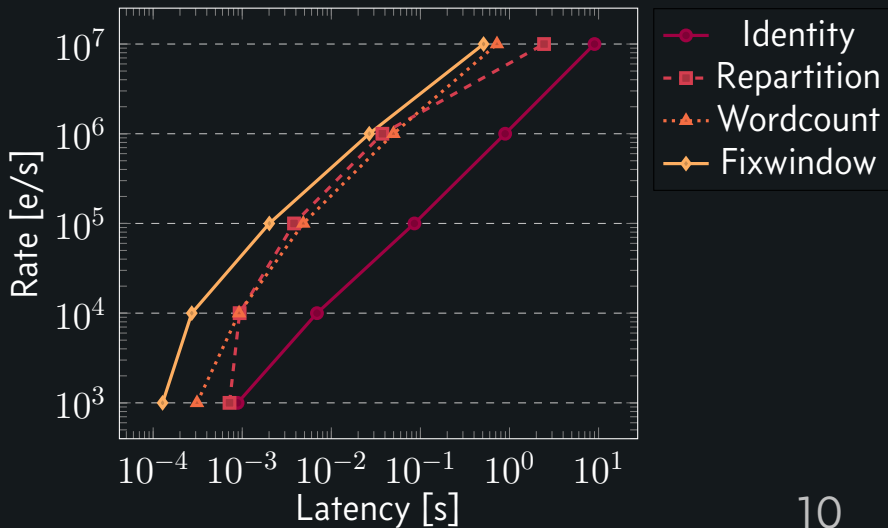
¹Pete Tucker et al. NEXMark—A Benchmark for Queries over Data Streams (DRAFT). Tech. rep. Technical report, OGI School of Science & Engineering at OHSU, Septembers, 2008.

Evaluation

- Run on sgs-r815-03 (AMD, 32 cores, 2.4GHz)
- Measured closed-loop per-epoch latency
- Data generated directly in memory
- Generation re-implemented in Rust
- No foreign systems like Kafka used
- Workload varied between 1K-10Me/s, 1-32 workers

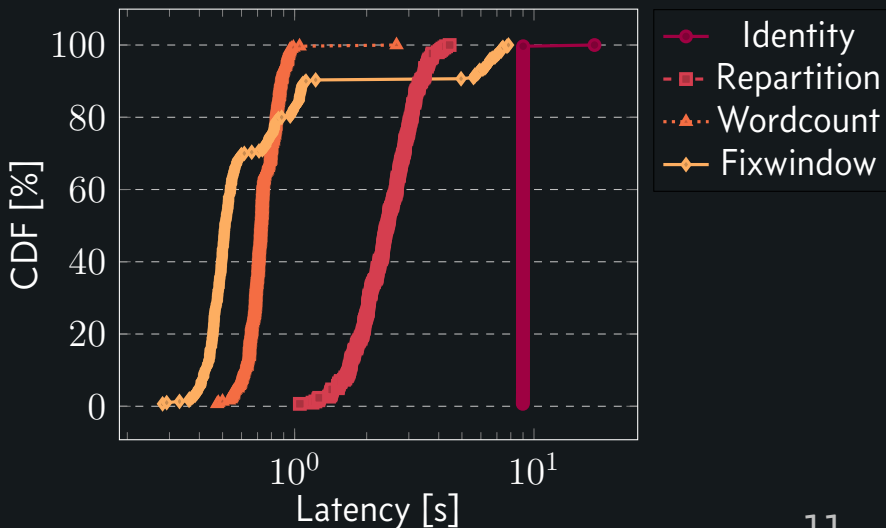
HiBench Latency Scaling

Median Latency (32 workers)



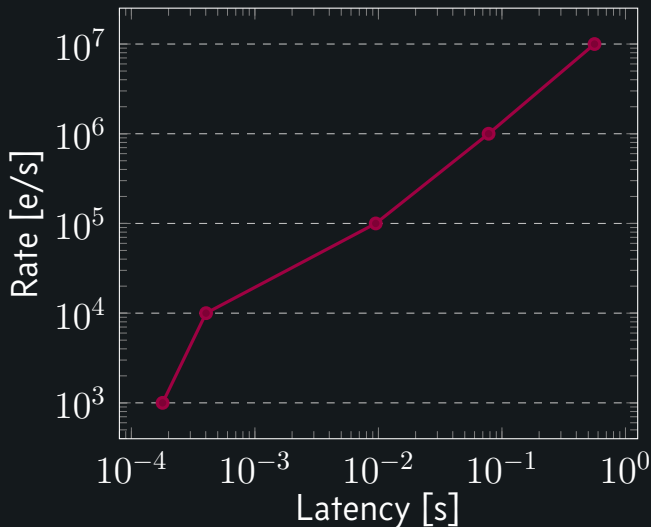
HiBench CDF

CDF (32 workers, 10'000'000 e/s)



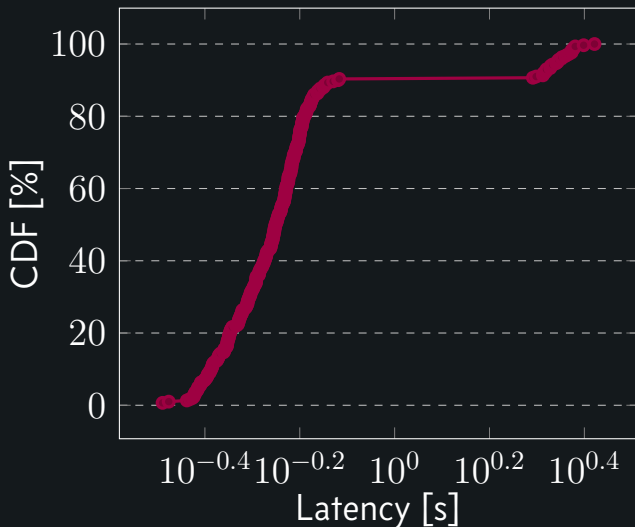
YSB Latency Scaling

Median Latency (32 workers)



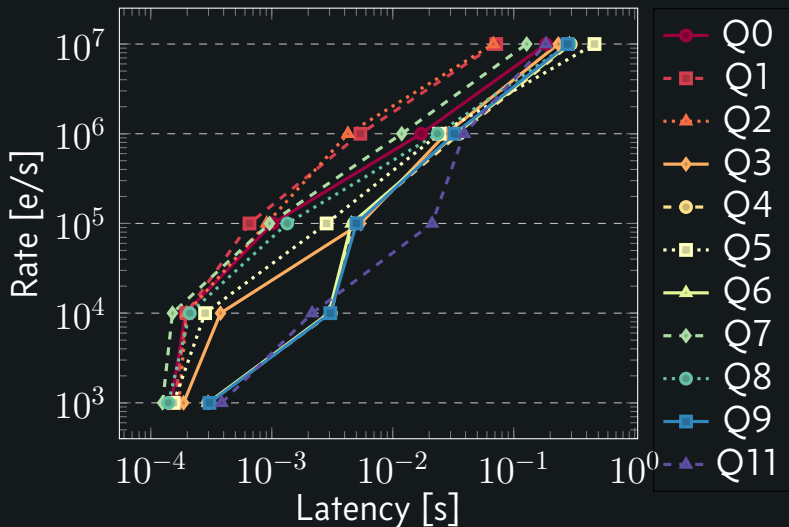
YSB CDF

CDF (32 workers, 10'000'000 e/s)



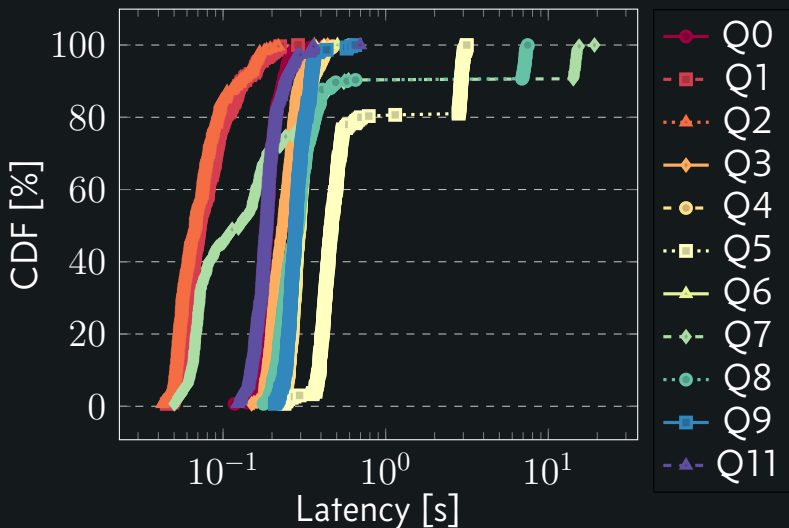
NEXMark Latency Scaling

Median Latency (32 workers)



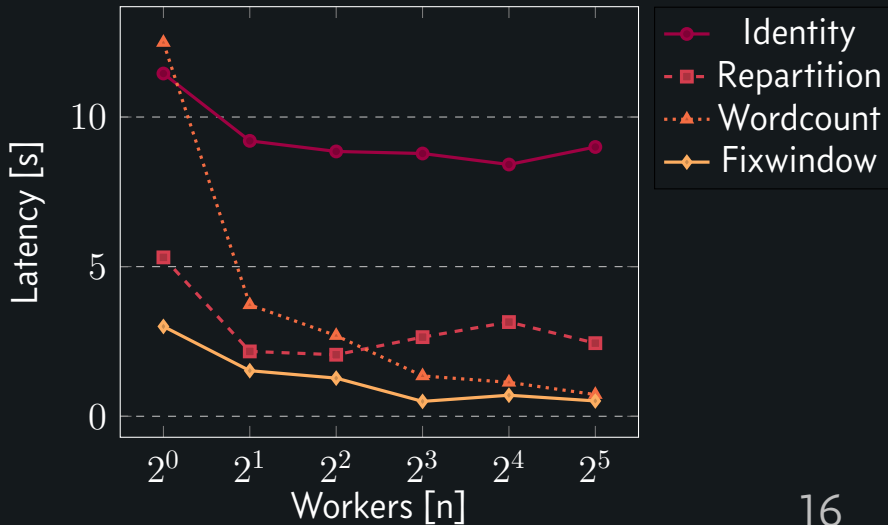
NEXMark CDF

CDF (32 workers, 10'000'000 e/s)



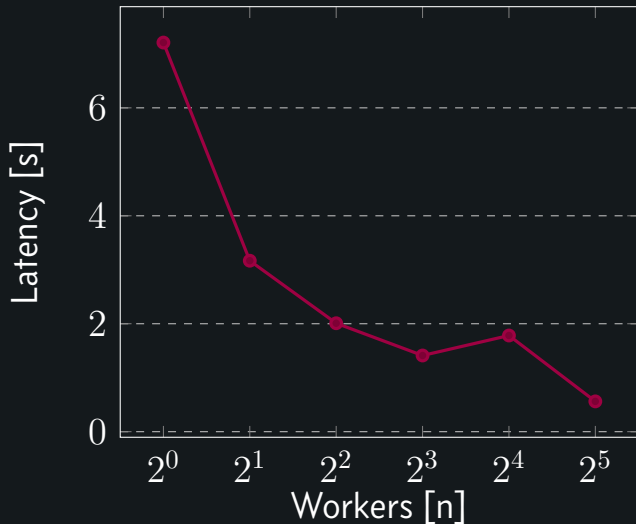
HiBench Worker Scaling

Scaling (10'000'000 e/s)



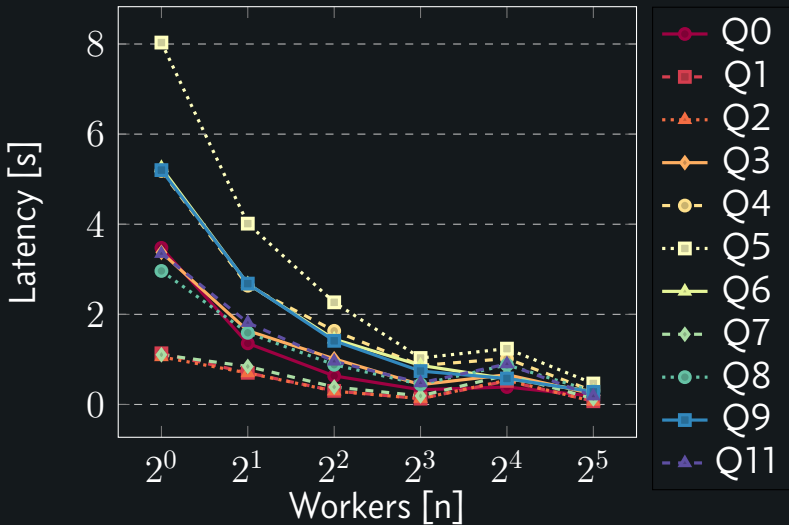
YSB Worker Scaling

Scaling (10'000'000 e/s)



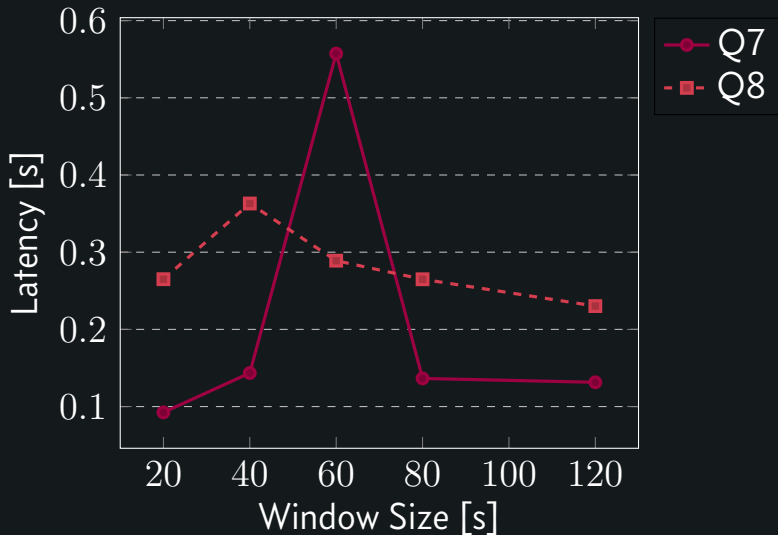
NEXMark Worker Scaling

Scaling (10'000'000 e/s)



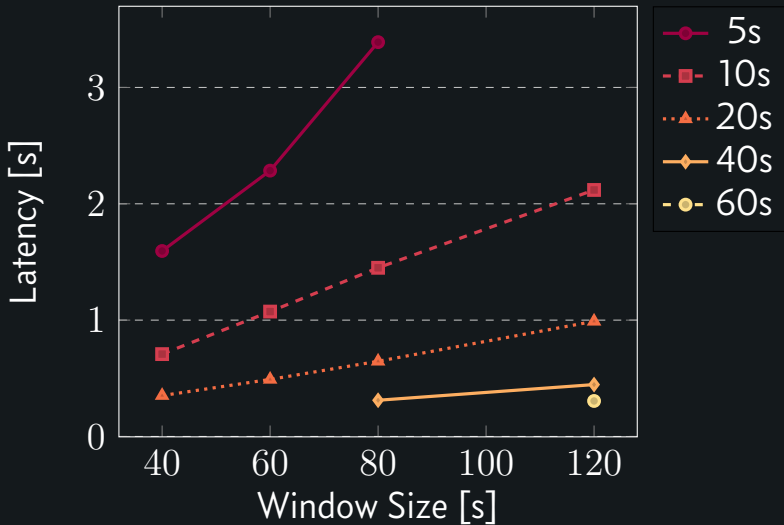
NEXMark Window Scaling

Windowing (32 workers, 10'000'000 e/s)



NEXMark Slide Scaling

Window Slides Q5 (32 workers, 10'000'000 e/s)



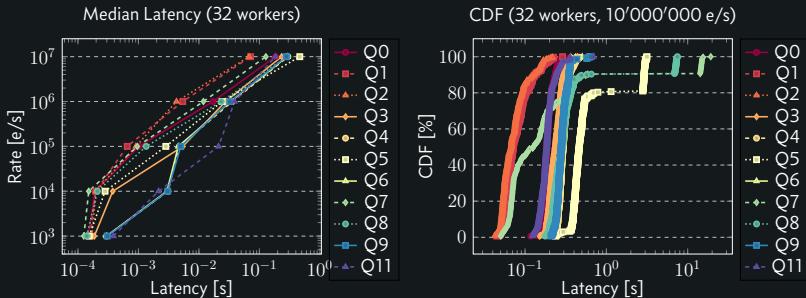
Important Findings

- Benchmarks are underspecified
- No result verification
- External systems compound complexity
- Focus is only on latency

Benchmark Suggestions

- Abstract model definitions for data flows
- Various short and long data flows
- Deterministically generated workloads
- Correctness verification tools

Conclusion



Timely can keep up with 10 million events per second.