

Implementation of a Benchmark Suite for Strymon

Nicolas Hafner



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich



Informatik
Computer Science

Motivation

Other Works

Timely

Benchmarks

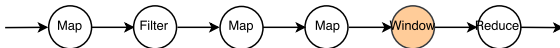
- We implemented three benchmarks:
 1. Intel's HiBench
 2. Yahoo's Streaming Benchmark
 3. Apache Beam's NEXMark

HiBench

- General Big Data benchmark
- Includes a section for streaming systems
- Only four latency tests:
 1. Identity
 2. Repartition
 3. Word Count
 4. Window Reduce

YSB

- Initially developed internally at Yahoo:
- Count ad views for ad campaigns
- Original setup requires Redis table lookup
- Only one, relatively simple data flow:



NEXMark

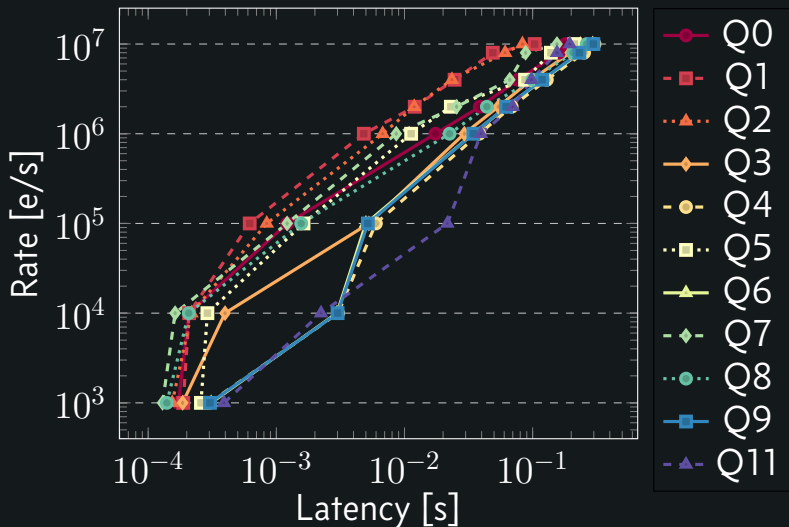
- Based on a paper by Tucker et al.
- An adaptation of XMark for streaming
- Implements an “auctioning system”
- 13 data flows in total
- Uses filter, map, reduce, join, window, session, partition

Evaluation

- Run on sgs-r815-03 (32 cores, 2.4GHz)
- Measured closed-loop per-epoch latency
- Data generated directly in memory
- Timely performs very well!

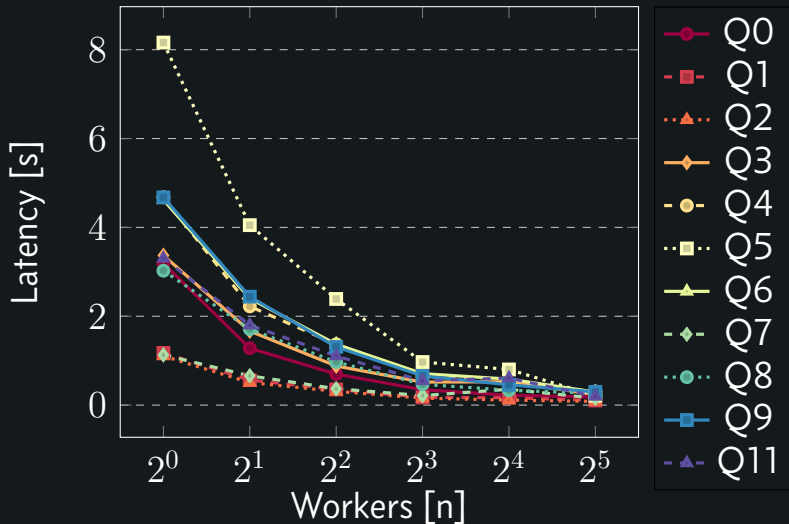
NEXMark Latency Scaling

Median Latency (32 workers)



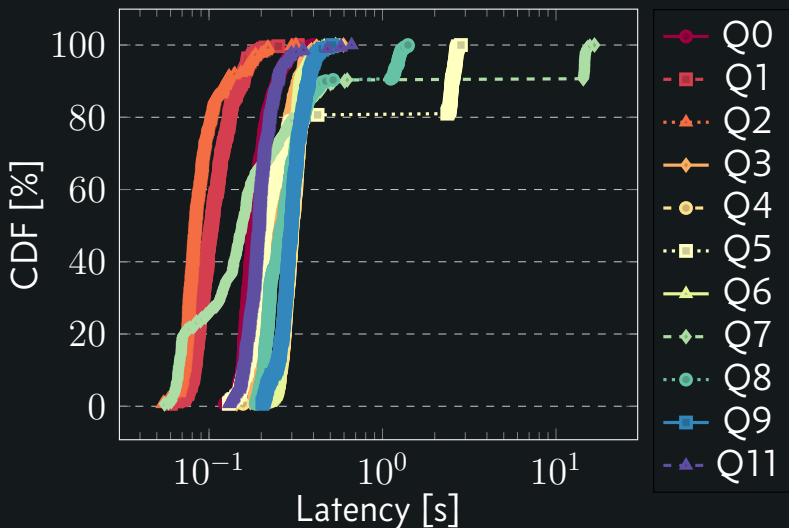
NEXMark Worker Scaling

Scaling (10'000'000 e/s)



NEXMark CDF

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



Conclusion