

Stream Processing Papers

Stream Processing

- [MillWheel: Fault-Tolerant Stream Processing at Internet Scale \(Video\)](#)
- [Photon: Fault-tolerant and Scalable Joining of Continuous Data Streams](#)
- [Scalable Distributed Stream Processing](#)
- [Distributed Operation in the Borealis Stream Processing Engine](#)
- [STREAM: The Stanford Data Stream Management System](#)
- [Fault-tolerance and high availability in data stream management systems](#)
- [Highly Available, Fault-Tolerant, Parallel Dataflows](#)
- [MapReduce Online](#)
- [Discretized Streams: Fault-Tolerant Streaming Computation at Scale](#)
- [Towards a Streaming SQL Standard](#)
- [Semantics and Implementation of Continuous Sliding Window Queries over Data Streams](#)
- [S-Store: A Streaming NewSQL System for Big Velocity Applications](#)
- [The 8 Requirements of Real-Time Stream Processing](#)
- [Models and Issues in Data Stream Systems](#)
- [S4: Distributed Stream Computing Platform](#)
- [Muppet: MapReduce-Style Processing of Fast Data](#)
- [Models and Issues in Data Stream Systems](#)

Streaming Algorithms

- [HyperLogLog: the analysis of a near-optimal cardinality estimation algorithm](#)
- [HyperLogLog in Practice: Algorithmic Engineering of a State of The Art Cardinality Estimation Algorithm](#)
- [Hyperloglog and MinHash](#)
- [Data Streams as Random Permutations: the Distinct Element Problem](#)
- [Approximately Detecting Duplicates for Streaming Data using Stable Bloom Filters](#)
- [An Improved Data Stream Summary: The Count-Min Sketch and its Applications](#)
- [Fast Incremental Maintenance of Approximate Histograms](#)
- [Effective Computation of Biased Quantiles over Data Streams](#)
- [Dynamic Histograms: Capturing Evolving Data Sets](#)
- [Incremental calculation of weighted mean and variance](#)
- [A curated collection of papers on streaming algorithms](#)
- [Probabilistic Data Structures for Web Analytics and Data Mining](#)
- [Philippe Flajolet's contribution to streaming algorithms](#)
- [Approximate Frequency Counts over Data Streams](#)
- [Methods for Finding Frequent Items in Data Streams](#)
- [The space complexity of approximating the frequency moments](#)
- [Cuckoo Filter: Practically Better Than Bloom](#)
- [Streaming/Sketching Conference from AK Tech](#)
- [Medians and Beyond: New Aggregation Techniques for Sensor Networks](#)
- [t-digest](#)
- [Count Min Sketch](#)
- [References for Data Stream Algorithms](#)
- [Data Streams - Algorithms and Applications](#)
- [Distributed Streams Algorithms for Sliding Windows](#)

Distributed Systems

- [Blazes: Coordination Analysis for Distributed Programs](#)