

Algorithm Term Project Proposal

Streaming algorithms are generally required to operate within limited space and time complexity as they are required to analyze infinite data streams. As the data stream is infinite for streaming applications traditional data structures do not work well for them. Count Distinct elements/ Cardinality Estimation is a popular streaming problem in which we need to find the count of unique elements in the stream. The solution to this problem can be applied to Networking / Traffic Monitoring, Data mining, etc.

HyperLogLog is one of the state of the art algorithmic solution for this problem. We are proposing to study and implement different implementations of HyperLogLog and their performance on different datasets. Currently, we are planning to use below reference papers from ACM and IEEE publications for this project.

- [1] HyperLogLog in Practice: Algorithmic Engineering of a State of The Art Cardinality Estimation Algorithm, Stefan Heule, Marc Nunkesser, Alex Hall Proceedings of the EDBT 2013 Conference, ACM, Genoa, Italy
- [2] Remove Minimum (RM): An Error-Tolerant Scheme for Cardinality Estimate by HyperLogLog, IEEE Transactions on Dependable and Secure Computing, August 2020
- [3] Fast Updates for Line-Rate HyperLogLog-Based Cardinality Estimation, IEEE Communications Letters, Pages 2737 - 2741, 20 August 2020