Data Stream Algorithms

S. Muthukrishnan

Google Inc
`muthu@google.com`
`www.cs.rutgers.edu/~muthu`

Abstract

How does one deal with massive data sets that is available for analyses? We will describe the classical *data stream model* in which we make one pass over the data and with sublinear resources perform much of the data analyses we care about, such as frequent items, summaries, compressed sensing, clustering and others. We will present the basic algorithmic techniques used to build the sophisticated analyses above. In addition, we will present extensions to other problems (graph, matrix, statistics) and to other models (probabilistic models, parallel models such as Google's MapReduce). One of the reasons this area of research thrives is its immediate application to a number of scenarios, which we will describe.

1 Introduction

See [1].

References

[1] S. Muthukrishnan. Data Streams: Algorithms and Applications. *Foundations and Trends in Theoretical Computer Science*, Vol 1, Issue 2, August 2005.