Title: Fighting Latency
Author: Thomas Cormen

Abstract

Some data accesses take much longer than others. Accessing data on disk can take five to six orders of magnitude longer than accessing data in RAM. Computations that work with data that must be accessed externally can run very slowly if they are not designed and implemented with high-latency accesses in mind. In this talk, I will summarize two approaches to mitigating the performance penalties incurred by high-latency accesses. One approach focuses on algorithms that minimize the number of high-latency accesses. The other allows us to implement these algorithms efficiently by providing a software layer that makes it easy to overlap high-latency operations with the rest of the computation.