

Thrill: High-Performance Algorithmic Distributed Batch Data Processing with C++

Friday, September 1, 2017 9:00 AM (40 minutes)

We present on-going work on a new distributed Big Data processing framework called Thrill. It is a C++ framework consisting of a set of basic scalable algorithmic primitives like mapping, reducing, sorting, merging, joining, and additional MPI-like collectives. This set of primitives goes beyond traditional Map/Reduce and can be combined into larger more complex algorithms, such as WordCount, PageRank, k-means clustering, and suffix sorting. These complex algorithms can then be run on very large inputs using a distributed computing cluster.

Presenter: BINGMANN, Timo (KIT)

Session Classification: Plenary