Contribution ID: **32** Type: **Presentation**

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

*Wednesday, October 11, 2017 5:00 PM (30 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. All these have already been implemented as examples.

These complex algorithms can then be run on very large inputs using a distributed computing cluster. Among the main design goals of Thrill is to lose very little performance when composing primitives such that small data types are well supported. Thrill thus raises the questions of a) how to design algorithms using the scalable primitives, b) whether additional primitives should be added, and c) if one can improve the existing ones using new ideas to reduce communication volume and latency.

Our aim is to provide a high-performance platform for next-generation Big Data algorithms, which is both faster and easier to use the Hadoop, Spark, or other current technology.

More information on Thrill is available at http://project-thrill.org/

## Track

BDAHM

**Author:** Mr BINGMANN, Timo (KIT)

**Presenter:** Mr BINGMANN, Timo (KIT)

**Session Classification:** Platforms