

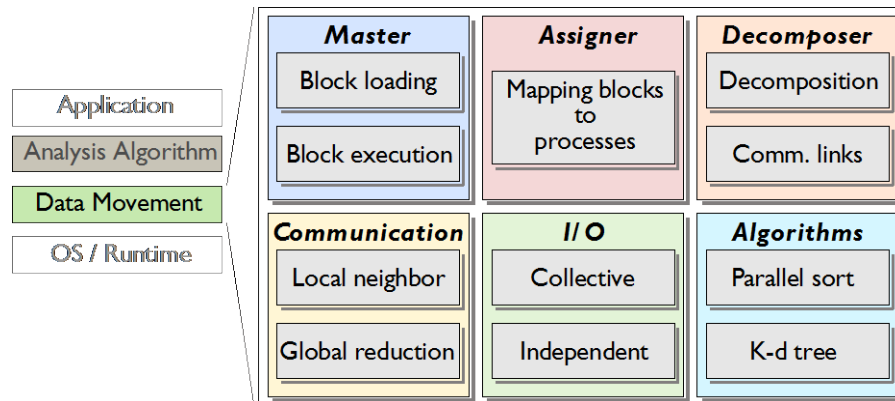
DIY Block-Parallel Data Analysis

Objectives

- DIY is a programming model and runtime for block-parallel analytics on DOE leadership machines.
- Its main abstraction is block parallelism: all parallel operations and communications are expressed in terms of blocks, not processors.
- This enables the same program to run in- and out-of-core with single or multiple threads.

Impact

- DIY enabled Delaunay and Voronoi tessellation of cosmology dark matter particles to 128K processes and improved performance by 50X [2].
- DIY enabled ptychographic phase retrieval of synchrotron X-ray images on 128 GPUs in real time. [3].
- Honorable mention paper at LDAH 2016 [1].



Components of DIY and its place in the software stack are designed to address the data movement challenge in extreme-scale data analysis.

Current Activities

- Enabling VTK-m by DIY-ing various VTK distributed-memory filters: parallel resampling, multipart dataset redistribution, and stream tracing.
- Ongoing development to prepare for exascale: relaxing synchronization, using deeper memory hierarchy, compatibility with many-core thread models.

[1] Morozov and Peterka, Block-Parallel Data Analysis with DIY2, LDAH 2016.

[2] Morozov and Peterka, Efficient Delaunay Tessellation through K-D Tree Decomposition, SC16.

[3] Nashed et al., Parallel Ptychographic Reconstruction, Optics Express 2014.