

# DIY: Enabling large-scale data-parallel analysis

## Main Ideas and Objectives

- Decouple analysis technique (user) from data-intensive parallelism (DIY)
- Enable large-scale data-parallel analysis (visual and numerical) on all HPC machines (IBM & Cray leadership machines)
- Provide internode scalable data movement
- For scientists, visualization researchers, tool builders
- For in situ, coprocessing, postprocessing

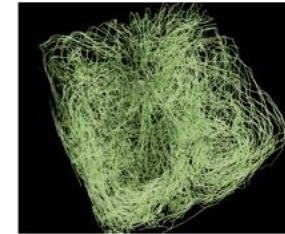
## Features

- Parallel I/O to/from storage
- Domain decomposition
- Network communication
- Written in C++, with C-style bindings, can be called from Fortran, C, C++
- Autoconf build system
- Lightweight: libdiy.a 800KB
- Maintainable: ~15K LOC
- MPI + openmp hybrid parallel model

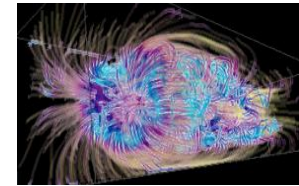
## Benefits

- Researchers can focus on their own work, not on parallel infrastructure
- Analysis applications can be custom
- Reuse core components and algorithms for performance and productivity

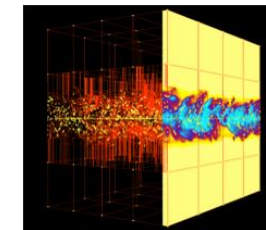
## Applications



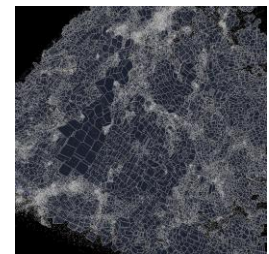
Particle tracing in thermal hydraulics



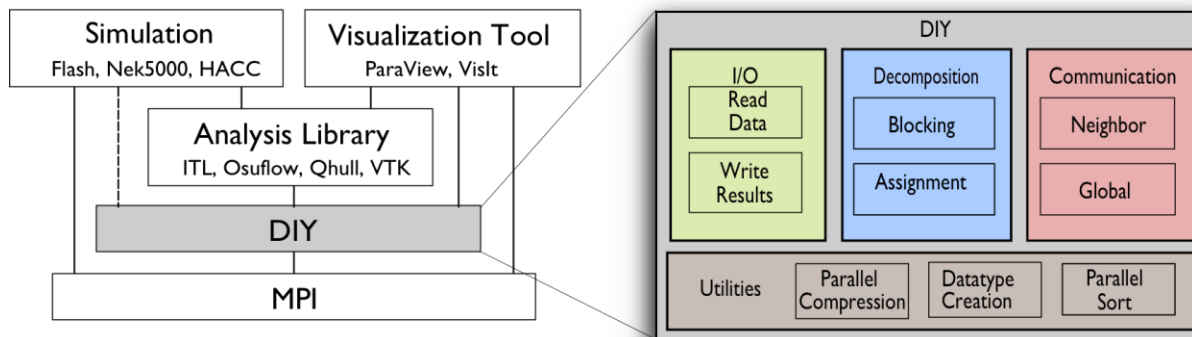
Information entropy in astrophysics



Topology in combustion



Computational geometry in cosmology



DIY usage and library organization