Efficient Extraction of Regional Subsets from Massive Climate Datasets using Parallel IO

J. Daily, K. Schuchardt, B.J. Palmer

IN41A-1360

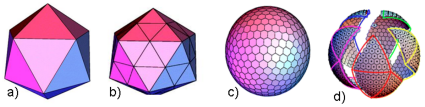
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Problem Statement

- Size of datasets now petabytes in scale
- Parallel tools are required
- Most current tools are serial and at the scale of workstations
- Most tools specialize to regularly gridded data
- No current tools correctly handle geodesic grids
- IO is the single greatest bottleneck

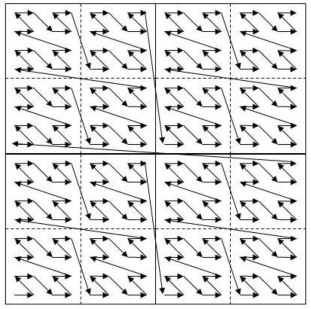
Creating the Geodesic Grid

- a) start with a regular icosahedron
- b) bisect and join the midpoint of each edge to produce four new triangles
- c) project each new vertex onto surface of sphere and repeat until desired resolution acquired; each vertex represents the center of a hexagon
- d) original triangles are paired to form logically structured square blocks



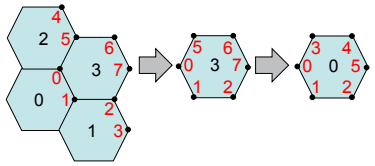
Data Layout using Morton-ordering

- Data uses Morton-ordering for each block
- Much easier to guarantee contiguous reads/writes



Subsetting Geodesic Grids

- Grid coordinates are not monotonic
- Explicit topology is required
- Cells have two unique corners and three unique edges



- (Left) Each cell provides two unique corners (red)
- (Middle) A cell is extracted with all associated corners
- (Right) Cells and their corners must be reindexed to a zero-based, monotonic scheme, since the NetCDF arrays in which they are stored are zero-based

Efficient Extraction of Subsets

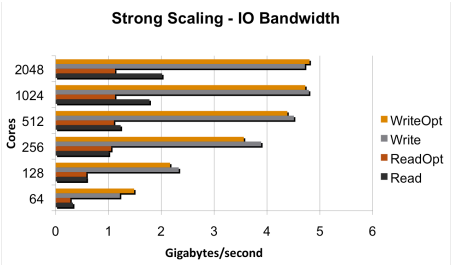
- Each process computes the subset for their local data and redistributes the subset
- Prior to reading from disk, each process can determine a priori whether its data is part of a subset
- Otherwise, data would be read from disk and immediately discarded
- Non-subsetting processes still participate in collective IO operations, but don't request data
- Morton-ordering helps to keep logically adjacent cells adjacent in memory and creates candidates for these selective subset IO operations

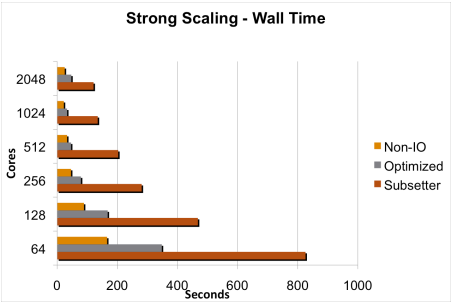
Experimental Setup

- Strong scaling test performed for subsetting operations over the "MJO" region 20N,20S,160E,90E, roughly 6.5% of the global data
- Tests performed on the franklin Cray-XT4 at NERSC
- Two versions of subsetter program tested, the original and one using the optimization described above
- Data for one time step for the largest variable was over 12 GB; 24 time steps were used
- Bandwidth and wall clock time were measured

Experimental Results

- Write bandwidth nearly 5 Gigabytes per second
- Bandwidth for writes did not benefit from optimization
- Bandwidth for optimized reads was less at higher scales due to less data requested from IO system
- Benefit most noticeable in wall clock time, causing IO to become a less dominant factor





Ongoing Development

- Parallel Analysis of GeOdesic Data (pagoda)
  - Data-parallel API
  - Parallel command-line tools mimicking the NetCDF Operators
  - Operators curently implemented: pgra, pgea, pgbo, pgsub
  - Supports both regular and geodesic grids
- Further IO optimizations using the latest Parallel NetCDF enhancements
- Grid interoperability and semi-structured grid standards

About Pacific Northwest National Laboratory

The Pacific Northwest National Laboratory, located in southeastern Washington State, is a U.S. Department of Energy Office of Science laboratory that solves complex problems in energy, national security and the environment, and advances scientific frontiers in the chemical, biological, materials, environmental and computational sciences. The Laboratory employs 4,000 staff members, has a $760 million annual budget, and has been managed by Ohio-based Battelle since 1965.

For more information about the science you see here, please contact:

Jeff Daily
Pacific Northwest National Laboratory
P.O. Box 999, MS-K7-90
Richland, WA 99352
(509) 372-6548
jeff.daily@pnl.gov

Software Website

https://svn.pnl.gov/gcrm/wiki/pagoda

Collaborators

Colorado State University