Data Models and Data Analysis at Exascale

Rob Ross

Mathematics and Computer Science Division

Argonne National Laboratory

rross@mcs.anl.gov

Please, not another file system.

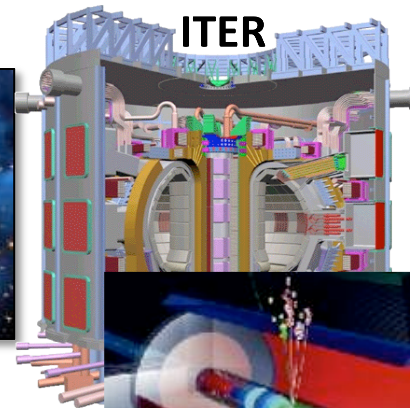
Rob Ross
Mathematics and Computer Science Division
Argonne National Laboratory
rross@mcs.anl.gov

Argonne NATIONAL LABORATORY

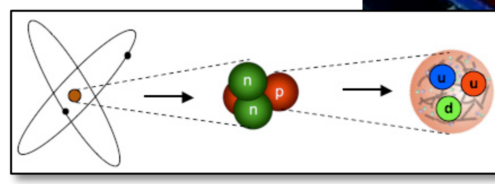
U.S. DEPARTMENT OF ENERGY

Exascale systems are instruments for scientific discovery.

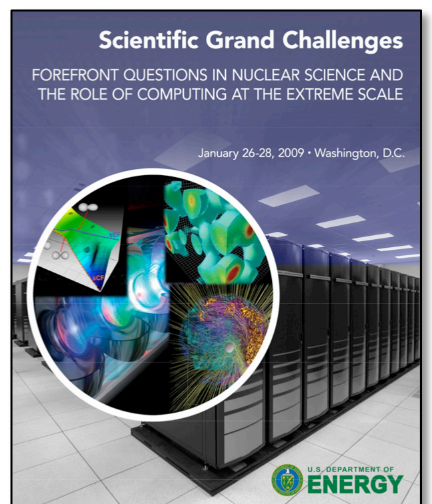
- High Energy & Nuclear Physics
  - Dark-energy and dark matter
  - Fundamentals of fission fusion reactions
- Facility and experimental design
  - Effective design of accelerators
  - Probes of dark energy and dark matter
  - ITER shot planning and device control
- Materials / Chemistry
  - Predictive multi-scale materials modeling: observation to control
  - Effective, commercial technologies in renewable energy, catalysts, batteries and combustion
- Life Sciences
  - Better biofuels
  - Sequence to structure to function

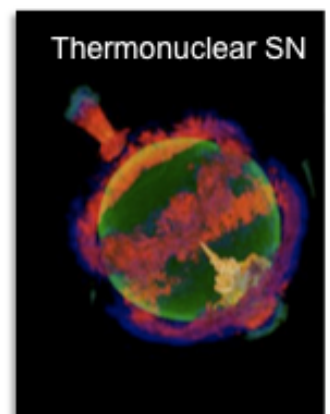


Hubble image of lensing



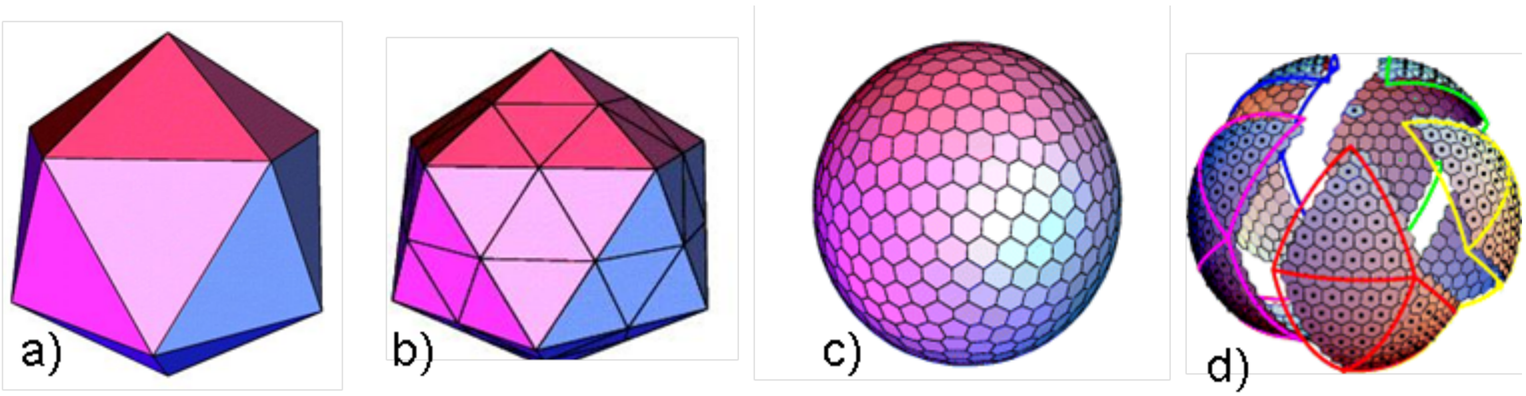
Structure of nucleons





Slide compliments R. Stevens and A. White.

Computational science codes have complex data models and structures.



Geodesic grid used in global climate resolving model (GCRM) at increasing levels of refinement.

From B. Palmer, A. Koontz, and K. Schuchardt, "An IO API for a Global Cloud Resolving Model". Environmental Modelling & Software. (submitted)

Exascale storage should support (i.e., be aware of) science data models.

Motif	Data Model/Structure	Examples
Dense Linear Algebra	Multidimensional Arrays	ScaLAPACK, S3D
Sparse Linear Algebra	Sparse Matrix	OSKI, SuperLU
Spectral Methods	Multidimensional Arrays	Nek5000
N-Body Methods	Trees, Unstructured Meshes	Molecular Dynamics
Structured Grids (+ AMR)	Multidimensional Arrays	FLASH, Chombo-based
Unstructured Grids (+AMR)	Unstructured Meshes	UNIC, Phasta
Graph Traversal	Sparse Matrix, DAG	Decision Trees (e.g., C4.5)

List derived from Berkeley Seven Dwarfs, modifications and additions by A. Choudhary, N. Samatova, Q. Koziol, T. Tautges, R. Latham, W. Liao, and R. Ross.

The data model characteristics are already captured, but in other tools and in limited forms.

```
netcdf r6_geopotential_19010101_100000 {
dimensions:
        time = UNLIMITED ; // (3 currently)
        cells = 40962 ;
        interfaces = 26 ;
variables:
        double time(time) ;
                ...
        float grid_center_lat(cells) ;
                ...
        float grid_center_lon(cells) ;
                ...
        float area(cells) ;
                ...
        float geopotential(time, cells, interfaces) ;
                geopotential:long_name = "Geo Potential" ;
                geopotential:units = "m**2/sec**2" ;
                geopotential:coordinates = "grid_center_lat
grid_center_lon" ;
        ...

// global attributes:
                :history = "Fri Jan 23 15:41:48 2009: ncks -v
grid_corner_lat,grid_corner_lon wind_19010101_100000.nc;
}
```

File systems don’t know anything about data models and structure.

```
0000000: 4344 4602 0000 0003  CDF.....
0000008: 0000 000a 0000 0005  ........
0000010: 0000 0004 7469 6d65  ....time
0000018: 0000 0000 0000 0005  ........
0000020: 6365 6c6c 7300 0000  cells...
0000028: 0000 a002 0000 000d  ........
0000030: 6365 6c6c 6e65 6967  cellneig
0000038: 6862 6f72 7300 0000  hbors...
0000040: 0000 0006 0000 000a  ........
0000048: 696e 7465 7266 6163  interfac
0000050: 6573 0000 0000 001a  es......
0000058: 0000 000b 6365 6c6c  ....cell
0000060: 636f 726e 6572 7300  corners.
0000068: 0000 0006 0000 000c  ........
0000070: 0000 0001 0000 0007  ........
0000078: 6869 7374 6f72 7900  history.
0000080: 0000 0002 0000 007d  .......}
0000088: 4672 6920 4a61 6e20  Fri Jan 
0000090: 3233 2031 353a 3431  23 15:41
0000098: 3a34 3820 3230 3039  :48 2009
00000a0: 3a20 6e63 6b73 202d  : ncks -
```

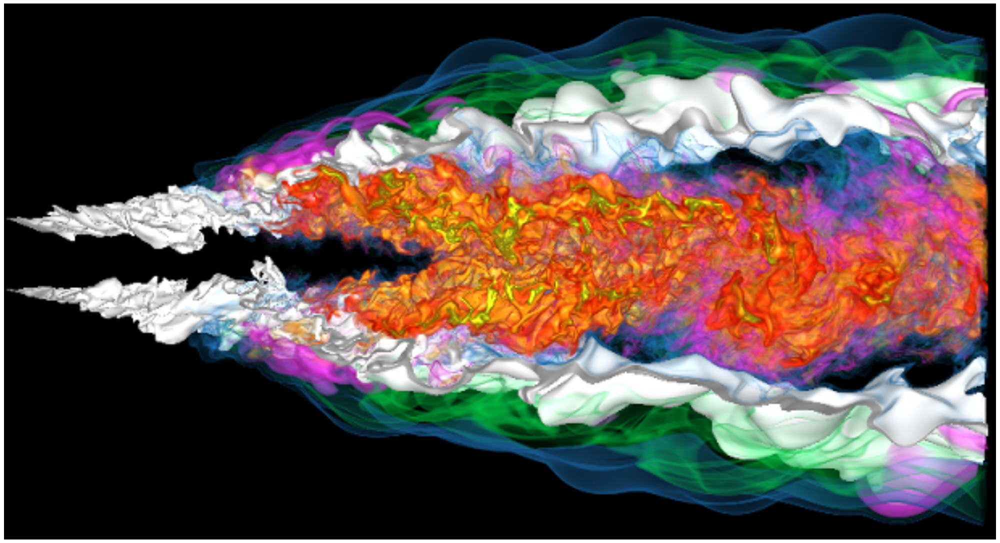
“od” dump of geodesic grid dataset. FS cannot interpret this.

Science discoveries come from analysis.

In situ analysis incorporates analysis routines into the simulation. This technique allows analysis routines to operate on data while it is still in memory. With help from the application scientist to identify features of interest, we can compress data of less interest to the scientist, **reducing I/O demands** during simulation and further analysis steps.

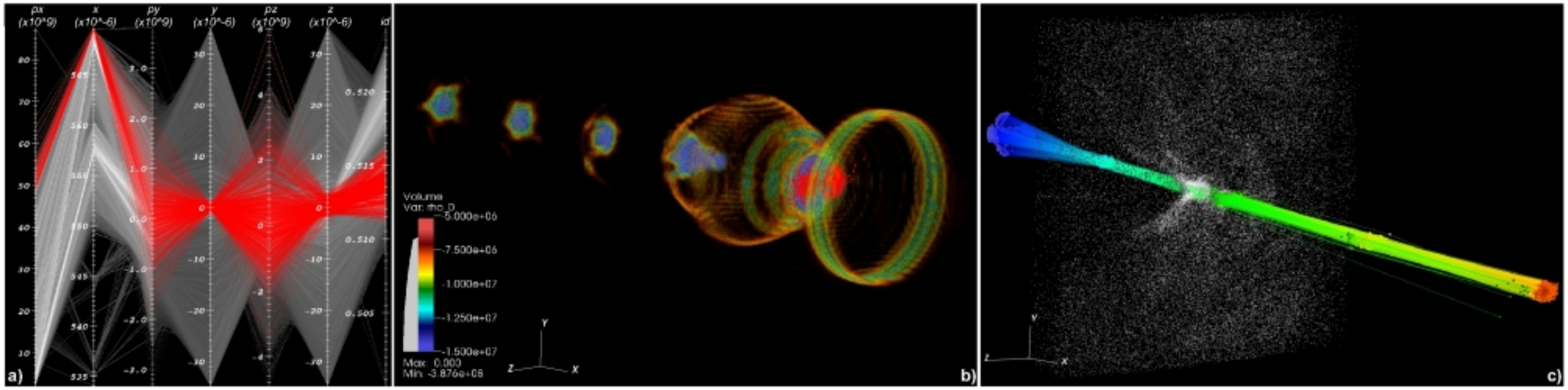
The feature of interest in this case is the mixture fraction with an iso value of 0.2 (white surface). Colored regions are a volume rendering of the HO2 variable (data courtesy J. Chen (SNL)).

By compressing data more aggressively the further it is from this surface, we can attain a compression ratio of 20-30x while still retaining full fidelity in the vicinity of the surface.



C. Wang, H. Yu, and K.-L. Ma, "Application-driven compression for visualizing large-scale time-varying volume data", IEEE Computer Graphics and Applications, 2009. Application is S3D combustion code, from J. Chen et al.

Query plays a key role in many analysis operations.



Parallel coordinates display | Particle density | Fast particles over time

SDM and VACET centers collaborated to use FastBit indices for data selection in analysis of lawer wakefield accelerator simulation data.

- VORPAL code produces 2D and 3D simulations of particles in laser wakefield
- Finding and tracking particles with large momentum is key to design the accelerator
- Brute-force algorithm is super-linear over all particles (taking 5 min on 0.5 mil particles), FastBit time is linear in the number of results (takes 0.3 s)

Slide compliments A. Shoshani; app. contacts: C.G.R. Geddes, E. Cormier-Michel (LBNL), P. Messmer (Tech-X).

Exascale storage must be co-designed.

Co-design incorporates the needs and perspectives of multiple groups into a product (e.g., hardware designers, system software specialists, and application scientists).

- We don't need a byte stream data model.
- We don't need strong default consistency semantics that we must disable.
- We don't need to be able to generate "files" at an astounding rate.
- We don't need to continue to access a file that we deleted.
- We will probably boot off some other storage volume.
- We don't need mknod.

We don't need an exascale file system, we need an exascale <u>storage</u> system.

Exascale storage should be specialized to support science access patterns, needs, and scales.

- Exascale systems must be used efficiently.
  - Exascale storage must ingest data quickly, must always be available.
- Scientists need to know where results came from.
  - Provenance capture should be assisted by the storage system.
- Scientists want to look for patterns in their data.
  - Exascale storage should enable search.
- Some data will be moved other places or archived.
  - There will be a need for serialization/pickling of some data.
- Scientists routinely analysis large time series datasets.
  - Data should be organized on storage to accelerate analysis.
  - Storage software should complement analysis/visualization advances.

Lee's talk

Blake's talk (too).

Never waste the opportunities offered by a good crisis.

– Niccolo Machiavelli

Acknowledgments

- DOE Office of Advanced Scientific Computing Research (ASCR)
- NSF Office of Cyberinfrastructure
- SciDAC Scientific Data Management Center
- SciDAC Institute for Ultra-Scale Visualization
- SciDAC Center for Scalable Application Development Software

- Pete Beckman, Phil Carns, Jason Cope, Kevin Harms, Kamil Iskra, Dries Kimpe, Sam Lang, Rob Latham, Rusty Lusk, Mike Papka, Tom Peterka, Katherine Riley, Seung Woo Son, Rajeev Thakur, Venkat Vishwanath, Justin Wozniak (ANL)
- Alok Choudhary, Kui Gao, Wei-keng Liao, Arifa Nisar (NWU)
- Arie Shoshani (LBNL)
- Kwan-Liu Ma, Hongfeng Yu (UC Davis)
- Lee Ward (SNL)
- Gary Grider, James Nunez, John Bent (LANL)
- Steve Poole, Terry Jones, Nagiza Samatova (ORNL)
- Chaoli Wang (Michigan Tech)
- Quincey Koziol (The HDF Group)
- Yutaka Ishikawa, Kazuki Ohta (University of Tokyo)