Parallel Data Storage, Analysis, and Visualization of a Trillion Particles

Suren Byna
J. Chou, O. Rübel, Prabhat, H. Karimabadi, W. S. Daughton, V. Roytershteynz, E. W. Bethel, M. Howison, K.-J. Hsu, K.-W. Lin, A. Shoshani, A. Uselton, and K. Wu

Lawrence Berkeley National Laboratory
Tsinghua University, Taiwan
University of California - San Diego
Los Alamos National Laboratory
Brown University
Modern scientific discoveries are driven by data

- By 2020, climate data is expected to be hundreds of exabytes or more.
- LHC experiments produce petabytes of data per year.
- Light source experiments at LCLS, ALS, SNS, etc. produce tens of TB/day.
- 1 Exabyte per a day (10 petabytes every hour).

- Storing, analyzing, and visualizing large data are big challenges.
- VPIC is a simulation that pushes the limits of data management tools on large supercomputers.
Vector Particle-in-Cell (VPIC) Simulation

✧ A state-of-the-art 3D electromagnetic relativistic PIC plasma physics simulation
✧ It is an exascale problem and scales well on large systems
✧ An open boundary VPIC simulation of magnetic reconnection
✧ NERSC Hopper Supercomputer
  o 6,384 compute nodes; 2 twelve-core AMD 'MagnyCours' 2.1-GHz processors per node; 32 GB DDR3 1333-MHz memory per node; Interconnect with a 3D torus topology
  o Lustre parallel file system with 156 OSTs at a peak BW of 35 GB/s
20,000 MPI processes using 120,000 cores

- Each MPI process writes ~51 Million (±15%) particles
  - Non-uniform number of particles
- Lustre-aware MPI-IO implementation
  - MPI collective buffer size is equal to the stripe size
  - Number of MPI aggregators is equal to the stripe count
- Each particle has 8 variables
- Particle dataset size per time step varies (30TB to 39TB)
- Collected a total of 400 TB data for 11 time steps
Data Challenges

✧ What is a scalable I/O strategy for storing massive particle data output?
  ○ In situ analysis works well when analysis tasks are known *a priori*
  ○ Many scientific applications require to store data for exploratory analysis

✧ What is a scalable strategy for conducting analysis on these datasets?
  ○ Sift through large amounts of data looking for useful information

✧ What is the visualization strategy for examining the datasets?
  ○ Display information that makes sense
Our Tools and Techniques

❖ Scalable I/O strategy for storing particle data
  o H5Part: A simple API on top of HDF5 to read/write particle data
  o Search for Lustre striping optimizations

❖ Scalable strategy for conducting analysis on these datasets
  o FastBit: Bitmap index generation and querying software
  o Hybrid Parallel FastQuery
    ✓ API to generate bitmap indexes
    ✓ API to query indexed or data from different data formats (HDF5, NetCDF, and ADIOS-BP)

❖ Visualization strategy for examining the datasets
  o Query-driven visualization using VisIt
Performance of Writing and analyzing

- Reached I/O peak rate in writing each variable
- Amortized I/O rate of 26 GB/s on the Lustre parallel file system with 35 GB/s peak bandwidth
  - 10 minutes to index 30TB data and 3 seconds to query highly energy particles with FastQuery
Query-driven Visualization

- Reduced the number of particles before rendering by down-selecting the scientifically interesting features
  - Highly energetic particles in this case
- New feature: Cross-Mesh Field Evaluation (CMFE)
  - Correlate particle data with the underlying magnetic field data

Diagram:

- FastBit & FastQuery
- Query-driven Visualization Pipeline
  - Data + Index
  - Query
  - Vis/Analysis
  - Display
Query-driven Visualization

- Showing all the particles with ‘Energy > 1.3’ in gray and those with ‘Energy > 1.5’ in color
- 164 million particles with Energy > 1.3 and 424,000 particles with Energy > 1.5
A science principle visualized for the first time

- The X-line, where magnetic reconnection happens
- Particle distribution of $U_{\perp 1}$ vs. $U_{\perp 2}$ in the vicinity of X-line
- The lack of cylindrical symmetry about the local magnetic field, called Agyrotropy
- This confirms the expected signature of the reconnection site in collisionless plasma
Conclusions

❖ Addressed the data management and analysis challenges posed by a highly scalable plasma physics simulation
  ○ Storage: 26 GB/s
  ○ Indexing: 10 minutes to index 30TB data file
  ○ Querying: ~3 seconds

❖ Demonstrated that exploratory analysis can handle challenges posed by large data

❖ Using query-driven visualization approach, application scientists explored and gained insights from massive particle datasets for the first time
  ○ Several of the phenomena visualized in this study have been conjectured about, but the capabilities developed here can unlock the scientific insights in unprecedented data volume
Thanks!