Analysis Challenges
Experience with Turbulence Dataset

Michael E. Papka
Argonne National Laboratory and The University of Chicago
Visualization and Analytics Team
Brad Gallagher, Chad Glendenin, and Randy Hudson
Outline

- Overview of data
- Data transfer
- Data validation
- Visualizations
- Future
- Summary and Challenges
Data for Visualization and Analysis

- 759 plotfiles (13.3TB)
  - 32,768 files per timestep
  - 928 x 928 x 928 volumes (29 x 29 x 29)
  - $\text{Velocity}_{x,y,z}$ density, pressure, enstrophy

- 1381 particle files (0.62TB)
  - 32,768 files per timestep
  - 16M particles per timestep
  - Timesteps < 1040
    - Particle id (float), time, evolved time, position$_{x,y,z}$, velocity$_{x,y,z}$
  - Timesteps > 1040
    - Particle id (int), time, evolved time, position$_{x,y,z}$, velocity$_{x,y,z}$
## Data Transfer

- Used GridFTP from LLNL to UofC
  - Created tar files directly from HPSS store
  - Transferred to scratch space
  - Expect scripts managed tarring
  - Python scripts managed GridFTP transfers
- Transfers peaked at 20MB/s
- Average transfer rates were ~6MB/s
- 28 days to transfer entire set of plotfiles
- Weeks to transfer particles files (~21 hours)
- Issues
  - Disk space at Chicago
  - VPN into LLNL had a 12 hour timeout
  - Machine stability on nodes used at LLNL
Data Validation

- Confirm the integrity of the data
Particle Reconstruction

10 million of the 16 million particles did not have unique ids
Two Common Analysis Workflows

1. Interactive (Planning and initial analysis)
   - Requires: Availability, fast access to data, possible large memory footprint
   - Enables: Interactive analysis and discovery, analysis artifacts

2. Batch (Full analysis)
   - Requires: Availability
   - Enables: Production of analysis artifacts

Visualization Pipeline

Data → Filter → Mapper → Display

Compression
Remote Visualization (Interactive/Batch)

- Method of choice

LLNL VisIt
Local Visualization (Interactive)

- Easy to work with?
Local Visualization (Interactive)
Local Visualization (Interactive)
Local Visualization (Batch)
Future

• Complete visualizations and aid in analysis
• Investigation of integrating analysis and visualization into simulation (SciDAC II)
• Help in making dataset available to community
  • Developing a plan for distribution
  • Developing tools for interaction with the data
Summary and Challenges

• Summary
  • Transfer of data
  • Development of a suite of tools for managing data
  • Aided in development of structure function analysis
  • Developed method of particle recovery
  • Initial visualization of plot and particle files
  • Investigation of data sharing

• Challenges
  • Data management
    • Storage
    • Transfer
    • Organization
  • Tools
    • Scalability
    • Stability