SuperGlue: Standardizing Glue Components for HPC Workflows

Jay Lofstead, Alexis Champsaur, Jai Dayal, Matthew Wolf, Greg Eisenhauer

gflofst@sandia.gov, alexis.champsaur@gmail.com, mwolf@ornl.gov, jdayal3@gatech.edu, eisen@cc.gatech.edu

Example Architecture

Addressed Scenarios
- Simulation writes in a unique format
- Analysis tools cannot read unique format
- Unique connector or “glue” code need for each pair (sim->analysis, analysis->analysis/viz, etc.)
- Maintain metadata as long and completely as possible
- Take the place of a Unix/Linux pipe operator
- IO bandwidth inadequate for desired data volume (NVM helps, but is not enough)
- Workflow engines manage processes rather than connections and data flows
- Handle parallel decomposed data structures

Example Applications
- LAMMPS – Molecular Dynamics
- GTCP – Particle in Cell fusion reactor simulator
- Very different data structures (list of atoms and positions in 3-D space vs. cells in toroidal space

Examined Connection Tech

FlexPath and ADIOS
- DataFlow Management between components
- Monitor flows to detect bottlenecks
- Re-balance deployment for optimal throughput
- Fault detection and mitigation (e.g., write to disk rather than next component)

Select
- First step to reduce output to just desired variables (that may need further adjustment)
- Include a header in stream to describe data format
- Slice off struct components

Dim-Reduce
- “Fold” one dimension into another
- May require memory rearrangement
- Maintain metadata/dimension units for downstream meaning

Magnitude
- Input as a 2-d set
- Output as a list of magnitudes

Histogram
- Configure component with a bin count, n
- Input is a list of values
- Output is a list n values that are the counts

Plotter
- Determine the axis information from metadata
- Input 2-d data and output an image stream

Dumper
- One of a family
- Accept input in some format
- Write to storage in the desired format

Future Work
- Investigate Data Analytics flows
- Examine additional “glue” component needs

Example Applications
- LAMMPS – Molecular Dynamics
- GTCP – Particle in Cell fusion reactor simulator
- Very different data structures (list of atoms and positions in 3-D space vs. cells in toroidal space

Example LAMMPS Performance

Select Time-step Time (Cycling Speed) in LAMMPS Workflow

- Example LAMMPS performance showing the improvement in time-step time when using SuperGlue components.