Online Analytics for Complex Simulation Workflows using DataSpaces

Tong Jin, Fan Zhang, Qian Sun, Hoang Bui, Manish Parashar
Rutgers University

Hongfeng Yu
University of Nebraska

Scott Klasky, Norbert Podhorszki, Hasan Abassi
Oak Ridge National Laboratory

Motivation

- Advanced coupled simulation workflows running at extreme scale generate large amounts of data that must be managed and analyzed to get insights
- Online data analysis approaches based on in-memory data staging and in-situ/in-transit processing becomes promising
- Simulations based on dynamic formulations such as Adaptive Mesh Refinement (AMR) present data management challenges for online analytics
  - Dynamically changing volume of data
  - Imbalanced data distribution
  - Heterogeneous resource requirements

Experiments Setup

- Platform: ORNL Titan Cray XK7 system
- Application
  - Chombo-based AMR simulations
- Analysis
  - Visualization: marching cubes algorithm, the de facto standard isosurface extraction algorithm

Cross-layer Adaptations for Dynamic Data Management

- Problem: dynamic runtime behaviors of AMR-based simulation increase the complexity of managing staging resources and scheduling in-situ/in-transit data processing
- Objective: manage online data analytics using cross-layer adaptations that respond at runtime to the dynamic data management and processing requirements

Application Layer Adaptation

Figure. Comparison of in-transit cores usage between static resource allocation and adaptive resource allocation. Utilization efficiency of static allocation: 54.57%; utilization efficiency of adaptive allocation: 87.11%.

Conclusions

- Manage dynamic data processing requirements at extreme scales using coordinated algorithm, middleware and resource layer adaptations
- Accelerated the data-to-insights process by up to 75% for a large-scale AMR-based simulation-analytic workflow
- Reduced overall data movement between the AMR-based simulation and in-situ analytics by 45%