Workflow Case Study: EPSi Data Processing Workflow

Jong Choi (ORNL), Tahsin Kurc (SUNY Stonybrook), Scott Klasky (ORNL)

1 Background

Edge Physics Simulation (EPSi), an advanced fusion simulation software to provide insight into edge plasma physics in magnetic fusion devices, is one the largest science simulations running on Leadership Computing Facilities (LCFs) in the USA. Numerical models used in EPSi and simulation codes have been enhanced in recent years. These advances led to greater data generation capabilities that already exceed the file system and disk-based storage capacities of current LCF. For the last several years, we have been researching and developing systems to support such challenging workflow scenarios in EPSi through our ADIOS framework. We extended ADIOS to support seamless integration of EPSi workflows with data transformation [1], monitoring systems [2], and in-situ visualization [3, 4]. We continue to focus on providing cutting edge workflow technologies to fulfill EPSi's requirements for the coming exa-scale era.

2 Network and Data Architecture

Local networks. Output data are stored on parallel file systems. Using staging nodes in near clusters or streaming outputs to remote facilities are also being considered. For simulation validation and verification purpose, EPSi is working on sending and receiving experimental data through wide-area networks in a near real-time fashion.

3 Collaborators

C.S. Chang in Princeton Plasma Physics Lab (PPPL), who is leading the EPSi project.

4 Instruments and Facilities

Present: Edge Physics Simulation (EPSi) is an advanced fusion simulation software and runs on Leadership Computing Facilities (LCFs) in the USA. EPSi saves checkpoint information and reduced analysis data to the file systems, which is about a few peta bytes per day. A set of in-situ visualization and analysis routines are developed to use with extra staging nodes in order to bypass IO overheads.

Next 2-5 years: Next generation machines such as Summit at ORNL and Aurora at ANL that will come online in the next 2-3 years are 5x-18x more powerful than current Leadership machines. These machines will enable EPSi to simulate numerical models much faster and produce simulation datasets that are 5x-10x larger in size. Data movement can occur within the cluster nodes but also can happen between near clusters or remote resources through wide are networks. Developing algorithms or methods to maintain low network latency and bandwidth will be an import issue.

Beyond 5 years: We expect more increased power in computing and will have a variety of resources for data management. This will add more complexity in EPSi workflow management.

5 Process of Science

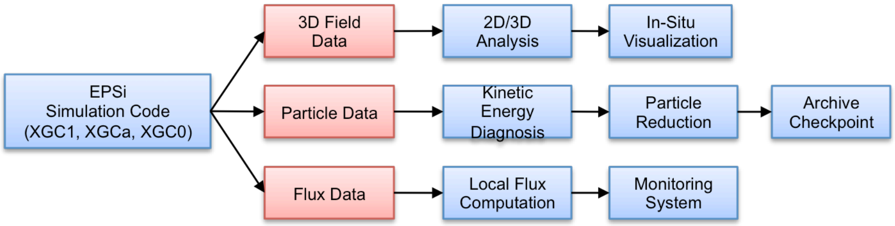

Figure 1. EPSi data analysis workflow.

Present: EPSi simulations generate extreme scales of data in the range of tens of petabytes per hour and are expected to generate multi-petabyte scale datasets on next generation leadership machines. The volume, rate, and variety of output data pose tremendous challenges in storing the datasets and carrying our post-simulation analysis. In order to keep up with data volumes and rates within current storage limits, output from an EPSi simulation needs to be processed in-situ and in-transit via complicated workflows for meaningful data reduction (before being written to storage) and for scientific data analysis and visualization (see Figure 1). However, managing and organizing such workflows with in-situ resources is a daunting task for scientists.

Another challenging scenario in EPSi is coupling of simulations. The suite of codes in EPSi includes two simulation codes, called XGC1 and XGCa, designed to simulate a large number of plasma ion and electron particles in a magnetic field but with different characteristics. While XGC1 focuses more on turbulence behavior near edges, XGCa is designed to model macro-scale fluctuations of plasma. Coupled execution of XGC1 and XGCa with tightly coordinated data sharing during their concurrent executions is needed to study different levels of plasma behavior and fusion reactions in a tokamak device. The workflow in a coupled execution of XGC1 and XCGa is illustrated in Figure 2. Execution of this workflow on extreme scale machines and for large-scale simulations introduces challenging research problems. Large-scale data sharing (for example, exchanging trillions of particle information) between the two simulations has to be supported with the requirements of minimum latencies between non-uniform processes in HPC environments. Additionally, given the amounts of data to be processed it is imperative that data analysis, reduction and visualization be performed seamlessly in various portions of the workflow. System support should take into account and leverage multi-level data storage hierarchies and high performance network interconnections as well as enable data reduction, compression, or indexing, in order to efficiently move data between tasks during execution.

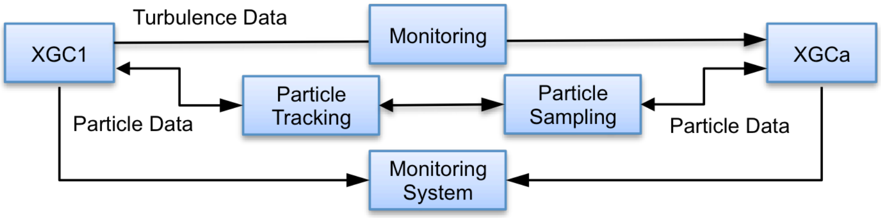

Figure 2. EPSi XGC1-XGCa coupling workflow.

Next 2-5 years: We anticipate that EPSi workflows will generate 5x-10x larger data (about 10-100 PBs) due to the increased computing powers. EPSi workflows will be integrated with more stream-based data analysis routines for on-line monitoring and in-situ analysis. EPSi will need a sustainable workflow system backed by strong network infrastructures in order to manage a large number of analysis routines and incorporate with various local and external resources for data management.

Beyond 5 years: We expect increased complexity in EPSi workflows. More in-situ/in-memory analysis routines and on-line visualization will be developed and integrated with. Strong software and network support will be required for EPSi workflows.

6 Remote Science Activities

Datasets generated by EPSi can be analyzed and visualized by collaborating teams at remote sites through the workflow defined in Section 5. Collaboration with remote fusion experimental facilities, such as KSTAR in Korea, is also being considered in order to receive experiment profiles as input and verify and validate (V&V) EPSi simulation results.

7 Software Infrastructure

Present: EPSi (for simulation runs) and ADIOS (for data storage and management). Most analysis workflows are static and performed manually.

Next 2-5 years: We expect rapid changes in computing infrastructures in the next 5 years. Next generation machines, such as Summit and Aurora, will provide more computing powers and heterogeneous subsystems for data management, which will require more sophisticated workflow system development. EPSi will need software infrastructure to support ad-hoc analysis integration and workflow automation. ADIOS will be used not only for basic I/O operations (storing data as files), but also for streaming data to staging areas or further sending data through wide area networks.

8 Outstanding Issues

EPSi simulations generate extreme scales of data in the range of tens of petabytes per hour and are expected to generate multi-petabyte scale datasets on next generation leadership machines.

The volume, rate, and variety of output data pose tremendous challenges in storing the datasets and carrying our post-simulation analysis.

[1] D. A. Boyuka, S. Lakshminarasimham, X. Zou, Z. Gong, J. Jenkins, E. R. Schendel, N. Podhorszki, Q. Liu, S. Klasky, and N. F. Samatova. Transparent in situ data transformations in adios. In Cluster, Cloud and Grid Computing (CCGrid), 2014 14th IEEE/ACM International Symposium on, pages 256--266. IEEE, 2014.

[2] R. Tchoua, J. Choi, S. Klasky, Q. Liu, J. Logan, K. Moreland, J. Mu, M. Parashar, N. Podhorszki, D. Pugmire, et al. Adios visualization schema: A first step towards improving interdisciplinary collaboration in high performance computing. In eScience (eScience), 2013 IEEE 9th International Conference on, pages 27--34. IEEE, 2013.

[3] Q. Liu, J. Logan, Y. Tian, H. Abbasi, N. Podhorszki, J. Y. Choi, S. Klasky, R. Tchoua, J. Lofstead, R. Oldfield, et al. Hello adios: the challenges and lessons of developing leadership class i/o frameworks. Concurrency and Computation: Practice and Experience, 26(7):1453--1473, 2014.

[4] **Towards Scalable Visualization Plugins for Data Staging Workflows**, David Pugmire, James Kress, Jeremy Meredith, Norbert Podhorszki, Jong Choi, Scott Klasky, 5th International Workshop on Big Data Analytics: Challenges, and Opportunities (BDAC-14)