Runtime system for I/O staging in support of in-situ processing of extreme scale data

ORNL: **Scott Klasky** (klasky@ornl.gov), Norbert Podhorszki, Nagiza Samatova, Matthew Wolf; **LBNL**: Arie Shoshani, John K. Wu, **Georgia Tech**: Karsten Schwan, Greg Eisenhauer

Project summary

As we approach the extreme scale in computing, we must realize new strategies to deal with the daunting challenge of managing and exploiting the massive quantities of complex data produced by scientific simulations. The challenge is exacerbated by the fact that I/O and memory systems have not seen increases in performance at the same rate as those observed for computational elements. This not only leads to unfavorable tradeoffs concerning machine power consumption for I/O and memory vs. computation, but it also means that the time scientists will spend on analyzing and visualizing the results produced by their simulations will greatly slow down the knowledge discovery process**. Our research will create and evaluate an I/O infrastructure and tools for extreme-scale applications and facilities so that they can reduce the time to discovery at small cost in machine resources and consequent power consumption**.

New tools must be highly scalable, portable, and easy-to-use, so that scientists can gain control of their science and concentrate on producing important scientific discovery in their own domain. **Accelerating the rate of insight and scientific productivity**, therefore, demands new solutions to managing the avalanche of data expected at extreme scale.

Partnering with many application teams and working on petascale machines, our team has developed an approach and **delivered proven technology** that accelerates I/O and the knowledge discovery process by reducing, analyzing, and indexing the data produced by a simulation while it is still in memory (referred to as **"in-situ" processing of data**). These technologies include the Adaptable I/O system **(ADIOS), FastBit indexing, and Parallel R**. For the proposed project, we will leverage those technologies and integrate them to **create a runtime system** that will allow scientists **to create easy-to-use scientific workflows** that will run in situ on select nodes of the extreme scale machine. This will not only **accelerate simulation and I/O**, but it will also provide scientists with immediate and valuable insights into data with online methods that **pre-analyze, index, visualize, and reduce the overall amount of data** produced by their simulations**.**

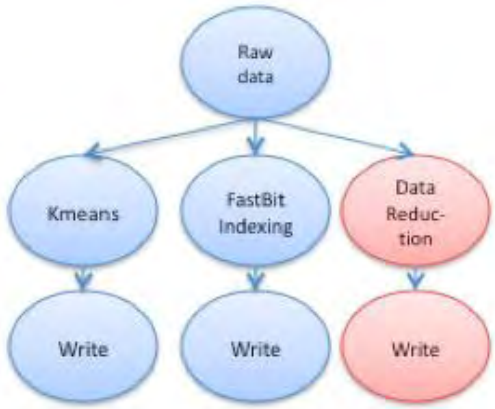
Goals:

1. **Minimize the need for scientists to have detailed knowledge of system hardware and operating systems.**
   a. We plan to enhance ADIOS. All of the applications that we target already use the best practices for I/O provided by ADIOS. We have demonstrated its advantages of simplicity and increased I/O efficiency, using multiple file formats including ADIOS-BP, HDF5, and parallel NetCDF.
   b. We plan to supply scalable parallel data analytics to the user, which is executed *in situ* using what we term the 'data staging' area. Specifically, we plan to take advantage of the advanced multi-/many-core staging area nodes and also investigate the use of accelerators. We will create services in the staging area that will be built from libraries, so that when the libraries are enhanced with accelerator technology, our analytical routines will also be enhanced.
   c. We plan to generate FastBit indexes as the data is generated. This will allow the scientist to access indexed data in post-simulation exploration. We will enable this technology to be used on parallel architectures.
2. **Provide scalable data triage, summarization, and analysis methods, including adaptive, power-aware algorithms and software for in-situ data reduction and/or analysis of massive multivariate data sets.**
   a. Existing ADIOS methods already permits summarization of massive data volumes, and these capabilities provide an excellent example of how in-situ processing can be performed efficiently.
   b. Novel and better *in-situ* data analysis and reduction techniques in the staging area are being researched and developed in this project. These techniques range from data compression to new techniques to improving performance by reducing data flow.
   c. FastBit (which will pre-compute the indexes in the staging area) will enable scientists to efficiently visualize the data

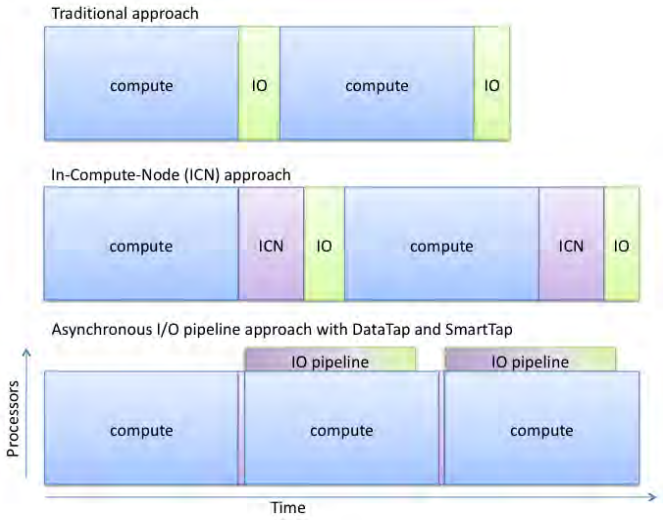
quickly after the simulation has completed. FastBit indexes are especially effective for finding patterns in multivariate datasets.

Key Implementation Features

1. In-situ pipelines enabling fast and efficient categorization of data before writing to minimize later resource consumption.

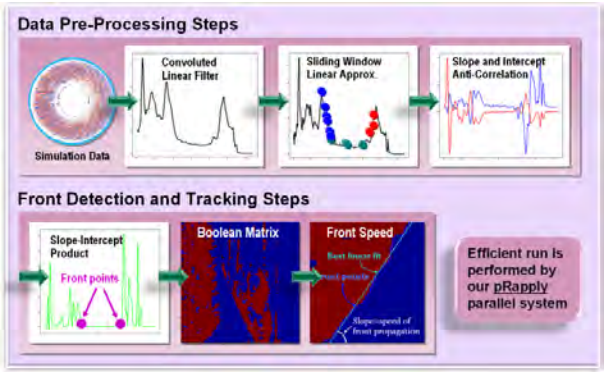


2. Staging methods for overlapping asynchronous I/O with in situ I/O pipelines to enable real-time indexing, analysis, reductions, and visualization.

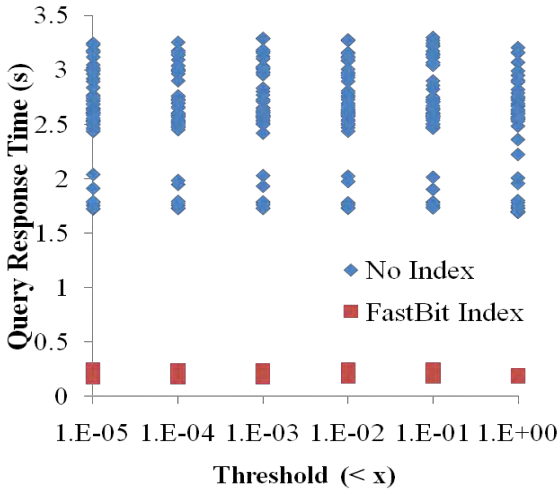


Examples and Demonstration Results

1. Turbulent front detection and tracking process as part of the in situ data analysis



2. Using FastBit indexes speeds up searches on Pixie3D data by an average of 13x.



3. *In situ* Fusion simulation analysis pipeline, demonstrated at SC 2009.