[July] [2020]

Moving fusion research into the big-and-fast data lane

A new data-analysis framework couples fusion experiments to supercomputers and performs routine analyses over 100x faster.

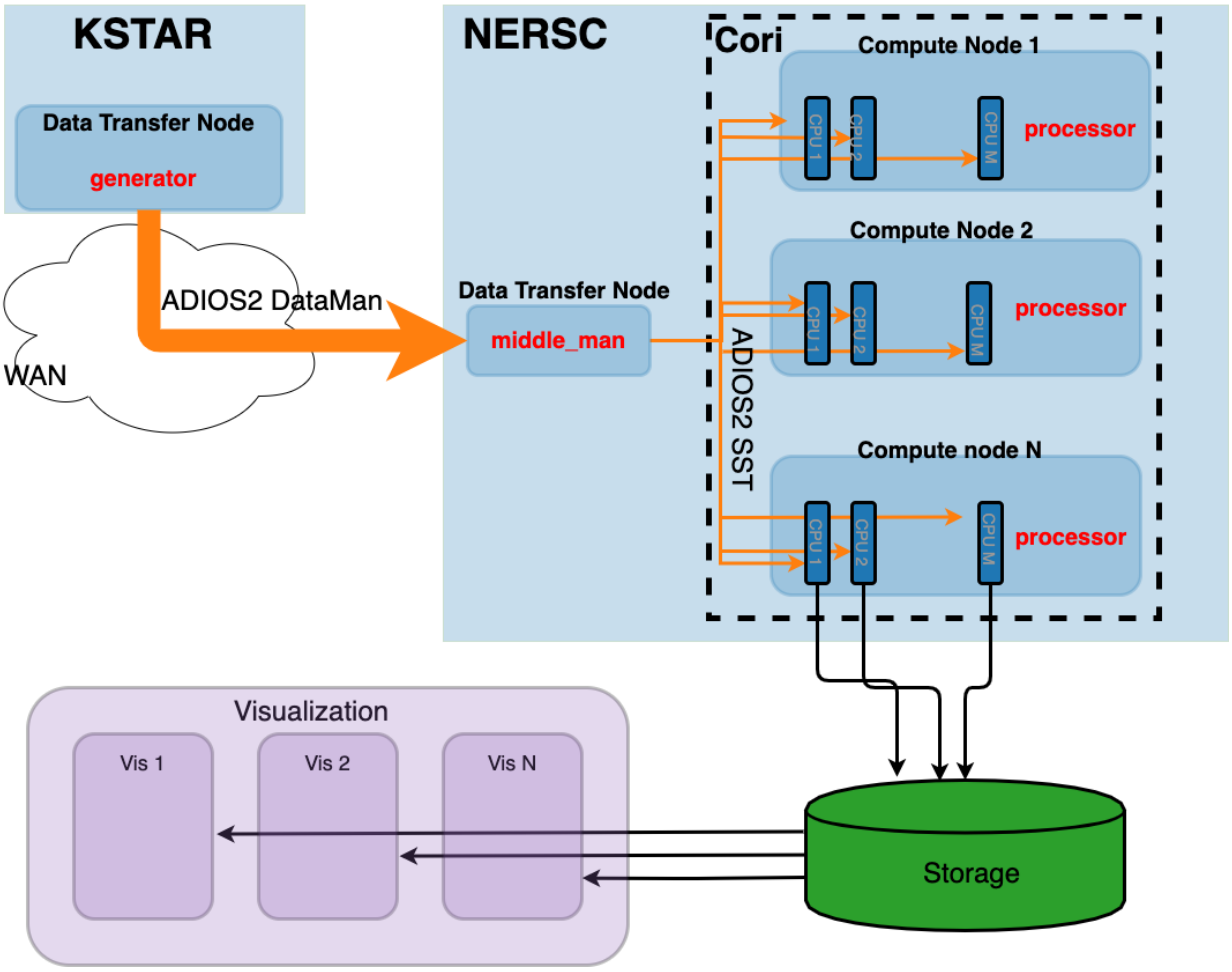

Image credit: Ralph Kube

The components of the Delta framework, shown in red, stream measurements from the KSTAR fusion experiment to the Cori Supercomputer at NERSC, where the data is automatically analyzed and visualized on a web dashboard.

The Science

In experimental fusion campaigns, 20 to 30 experiments may be performed on a single day. Sensitive diagnostics probe the plasmas in these experiments with ever increasing precision and accuracy and generate large data sets that are often too large for immediate analyses. This can delay progress. To allow rapid analyses of each experiment, scientists at PPPL have developed a novel data framework called Delta that automatically performs analyses of large datasets on supercomputers using machine learning to identify the features that are most important to sustain advanced operational performance.

This allows scientists to study the data in real time, which in turn enables more informed decisions to be made for the following experiment. This accelerates scientific discoveries towards the realization of fusion energy.

The Impact

Experimental data analysis can now be performed in minutes rather than days. Delta can perform the analysis of, for example, data from a high-frame rate camera, which takes 5 hundred thousand images every second, and which measures the crucially important plasma temperature. By streaming the data directly to a supercomputer, this now takes minutes. Previously, using a workstation, this took hours.

Summary

Fusion experiments generate data at an ever increasing rate. Next-generation facilities such as ITER will generate even ten times more data on a given day than today's experiments. To guide scientists through this data jungle, data analysis and visualization methods are essential tools.

By using state-of-the art streaming algorithms implemented in the adaptive input/output library for supercomputers known as ADIOS2, which was funded by DOE, Delta can sustain 500MBytes-per-second data streams from the Korean superconducting tokamak, KSTAR, to the Cori Supercomputer in California, which is operated by the National Energy Research Scientific Compute Center (NERSC). Delta makes effective use of Cori's distributed architecture of Cori and results are visualized using a web-based dashboard. This connects U.S. scientists to an experiment on the other side of the world.

Delta is not just fast; it is clever. So much data is obtained in fusion-relevant plasma experiments that it is a challenge to know which data needs careful examination, and which can be ignored to save time and effort. This is where machine learning comes in. Delta uses state-of-the-art machine learning algorithms to identify, for example, transient fluctuations of the magnetic field and automatically performs detailed, fine-tuned image analyses of the corresponding temperature fluctuations. Immediately after an experiment has concluded, the scientists are presented with a detailed picture of the plasma motions.

Contact

Ralph Kube
Princeton Plasma Physics Laboratory
rkube@pppl.gov

R. M. Churchill
Princeton Plasma Physics Laboratory
rmchurchi@pppl.gov

Funding

This project is funded through PPPL projects "Tokamak Theory" 3104*****NUL, LDRD 7000R085*NUL, and 3201*****NUL.

Publications

R. Kube *et al.*, "Leading magnetic fusion energy science into the big-and-fast data lane" *Proceedings of the 19th Python in Science Conference*. [DOI: 10.25080/Majora-342d178e-013]