Abstract Submitted
for the DPP20 Meeting of
The American Physical Society

Moving fusion energy into the big and fast data lane RALPH KUBE, MICHAEL CHURCHILL, Princeton Plasma Physics Laboratory, JONG YOUL CHOI, RUONAN WAND, Oak Ridge National Laboratory, CS CHANG, Princeton Plasma Physics Laboratory, SCOTT KLASKY, Oak Ridge National Laboratory — A wide array of diagnostics are routinely used to measure pulsed plasma discharges in tokamaks.Diagnostics with the highest spatial and temporal resolutions readily produce data streams upward of 1 GByte/s. Reducing such large volume, high-velocity high-dimensional data time-series into analysis results available to scientists in near real-time allows to accelerate scientific discovery. Here we present Delta, a novel framework that leverages computational resources of remote high-performance compute facilities to analyze data streams from plasma diagnostics in near real-time. As a demonstration, we use Delta to calculate a suite of spectral analysis routines using data from the KSTAR ECE diagnostic on Cori, a Cray XC-40 supercomputer operated at NERSC. The ECE diagnostic samples Te fluctuations on a 24 by 8 pixel grid at two toroidal locations with a rate of about 1 MHz. Our experiments show that we can consistently stream the entire 5GB large dataset with up to 500 MB/sec from KSTAR to Cori and perform the entire analysis suite in about 5 minutes. A web-based live dashboard visualizes the analysis in near real-time. Finally, we discussing ongoing efforts to incorporate variational auto encoders in Delta to compress the data stream and perform outlier detection.

Ralph Kube
Princeton Plasma Physics Laboratory

Date submitted: 26 Jun 2020 Electronic form version 1.4