Programming systems for data centric computing

Moderators:
Dan Quinlan - LLNL
John Wu - LBNL

Participants:

Dan Quinlan - LLNL
John Wu - LBNL
Alice Koniges - LBL
Pei-Hung Lin - LLNL
Joseph Manzano - PNNL
Nathan Tallent - PNNL
Josh Suetterlein - PNNL
Barbara Chapman - BNL and Stony Brook University
Andrew Lumsdaine - Indiana University
Ray Chen - University of Maryland
Kevin Huck - University of Oregon
Paul Woodward -- UMN
Josh Milthorpe - IBM Research

Charge

Observational and experimental facilities are creating increasingly massive and complex data sets about important chemical, physical and biological processes; many of these experiments are expecting **real-time or near real-time responses** to **control the experiment or steer the observations**. In other cases, these observations are **compared** with the even large output produced by large and increasingly realistic numerical simulations of the same processes. Many of these data analysis tasks also require data from many different sources and formats. How to effectively support these real-time, near-real-time, distributed data management and analysis workloads? What are the new capabilities that allow a closer coupling of experimental, observational and simulation processes that will drive new scientific insights? What **programming support** (languages, compilers, runtime systems, OS, tools) is needed to create these new capabilities? Which elements of the **software stack** need to change the most and why?

Themes:

Software stack for large-scale bursty streaming data, satisfying some constraints (e.g., near-real-time),

- Language and system software support for data-centric research.
- Hardware research: identify appropriate architecture for data-centric computations

Motivating Examples

DOE office of science experimental facilities:
ALS, SNS, APS, LCLS, NIF, HPX, STAR, ATLAS, ITER(as a partner)

Application examples

- Smart Grid, Smart Cities, Smart Mobility (outside of office of science, but of great interest)
- BNL nanotube optimization
- Magnetic fusion experimental data management/analysis
- Compact Laser Accelerator such as BELLA
- NNSA HED physics facility such as NIF
- Astrophysics simulation

Research Topics

- Language support, system software support for data-centric research: better languages/better Compilers/DSL
- Scheduler to address the dynamic behavior of data/system
- Portability on HPC/Cloud
- Constraint-based software stack: real-time, interactive,
- Runtime system appropriate for an application specific requirement such as reproducibility, resilience, durability - in collaboration with workflow engines
- Computational Steering (introspect and adapt): terminate experiment, change parameters, feedback
- Metric/QOS: productivity of programmer/application scientists, expressiveness