Online Data Analysis and Reduction at the Exascale

By 2024, computers are expected to compute at 10^{18} operations per second but write to disk only at 1012 bytes/sec: a compute-to-output ratio 200 times worse than on the first petascale systems. In this new world, applications must increasingly perform online data analysis and reduction—tasks that introduce algorithmic, implementation, and programming model challenges that are unfamiliar to many scientists and that have major implications for the design of various elements of exascale systems.

The goal of the Center for Online Data Analysis and Reduction at the Exascale **(CODAR)** is to produce infrastructure for online data analysis and reduction; provide valuable abstractions for implementing and customizing data analysis and reduction methods; import, integrate, and develop essential libraries implemented using these abstractions; incorporate the libraries into scientific applications and quantify accuracy and performance; release software artifacts; construct application-oriented case studies; document success stories and the process applied to obtain them; and report on codesign trade-off investigations.

PI: Ian Foster, Argonne National Laboratory

Collaborators: Argonne National Laboratory, Brookhaven National Laboratory, Oak Ridge National Laboratory, Brown University, Rutgers University, State University of New York – Stonybrook, University of Oregon

Particle-Based Applications

Particle-based simulation approaches are ubiquitous in computational science and engineering. The "particles" may represent, for example, the atomic nuclei of quantum and classical molecular dynamics methods or gravitationally interacting bodies or tracer particles in N-body simulations. In each case, every particle interacts with its environment by direct particle-particle interactions at shorter ranges and/or the particle-mesh interactions between a particle and a local field that is set up by longer range effects.

The Co-design Center for Particle Applications **(CoPA)** is addressing the challenges for particle-based applications to run on upcoming exascale computing architectures by partnering with ECP applications teams to create co-designed numerical recipes for particle-based methods: short-range particle–particle interactions (e.g., those which often dominate molecular dynamics methods), long-range particle-particle (e.g., electrostatic and gravitational) interactions, particle-in-cell methods, and linear-scaling electronic structure and quantum molecular dynamics algorithms.

PI: Susan Mniszewski, Los Alamos National Laboratory

Collaborators: Los Alamos National Laboratory, Argonne National Laboratory, Lawrence Livermore National Laboratory, Oak Ridge National Laboratory, Princeton Plasma Physics Laboratory, Sandia National Laboratories

Efficient Implementation of Key Graph Algorithms

Fundamental changes in the way electric power is generated, transmitted, and consumed today have resulted in an unprecedented need for computation to solve problems related to the design, planning, and operation of the power grid. This endeavor and the need to computationally design and model chemicals, materials, and biosystems at a molecular level are fundamental to accomplishing DOE's energy, environment, and national security missions. Recent advances in systems-based approaches coupled with ever-increasing data volumes require extreme-scale computing.

Combinatorial algorithms in general and graph algorithms in particular play a critical enabling role in numerous scientic applications. The irregular memory access nature of these algorithms makes them some of the most difficult algorithmic kernels to implement on parallel systems. Efficient implementation of key combinatorial (graph) algorithms chosen from four exascale application domains—smart grids, computational biology, computational chemistry, and climate science—will be captured in a unified soware framework, **ExaGraph**, that targets a diverse set of current and future extreme-scale architectures.

PI: Mahantesh Halappanavar, Pacific Northwest National Laboratory

Collaborators: Pacific Northwest National Laboratory, Argonne National Laboratory, Lawrence Berkeley National Laboratory, Purdue University, Sandia National Laboratories

Exascale Machine Learning Technologies

The **ExaLearn** Co-design Center leverages the revolution in what is variously termed machine learning, statistical learning, computational learning, and articial intelligence. New machine learning technologies can have profound implications for computational and experimental science and engineering and thus for the exascale computing systems that DOE is developing to support those disciplines.

ExaLearn will identify the fundamental machine learning challenges associated with ECP applications and concentrate efforts on the development of scalable machine learning technologies for the analysis of data generated by exascale applications and DOE user facilities as well as to guide the optimal selection and steering of (1) complex computer simulations (e.g., current exascale application projects) and (2) experiments (e.g., light sources, NIF, accelerators). The key to success in this endeavor is a deliberate focus on verification and validation and uncertainty quantification with a solid determination of generalization errors.A unifying principle is that of using exascale machine learning to improve the efficiency and effectiveness both of DOE computing resources and experimental facilities.

PI: Francis Alexander, Brookhaven National Laboratory

Collaborators: Brookhaven National Laboratory, Argonne National Laboratory, Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, Oak Ridge National Laboratory, Pacific Northwest National Laboratory, Sandia National Laboratories