

# A Unified Data-Driven Approach for Programming *In Situ* Analysis and Visualization



Pat McCormick, PI  
Christine Sweeney



Janine Bennett, Co-I  
Ali Pinar



Alex Aiken



Charles Hansen  
Valerio Pascucci

## Overview

Amid today's challenges of constrained power budgets and rapidly changing processor and system architecture designs, this project aims to study a **unified data-driven approach for programming applications and *in situ* analysis and visualization**. Our studies are driven by the needs of common operations as well as those of three application areas.

## Themes Addressed

This project focuses on studying the impact of the following, in the context of a unified data-driven approach:

- Supporting effective *in situ* data management, analysis, and visualization.
- Providing a foundation for building efficient and effective workflow management.
- Enabling an interactive *in situ* user environment on the underlying runtime software design.
- Understanding the impact on existing applications, infrastructures and tools.

## Mission Problems

- Constrained power budgets for supercomputing.
- Increased quantities of data generated by scientific computing.
- Limits on placement and movement of data due to performance and energy constraints at extreme scale.
- Programming scientific applications and data analysis and visualization amid rapidly changing processor and system architecture designs.

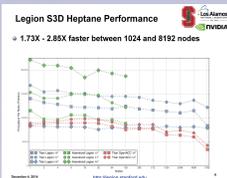
## Collaborations

- Application areas of combustion, climate and cosmology, with scientists located at Sandia, Los Alamos and Argonne.
- Exact Co-Design Center.
- Industry partners from DOE FastForward and DesignForward activities (AMD and NVIDIA).
- Two by invitation workshops that will include the external community.

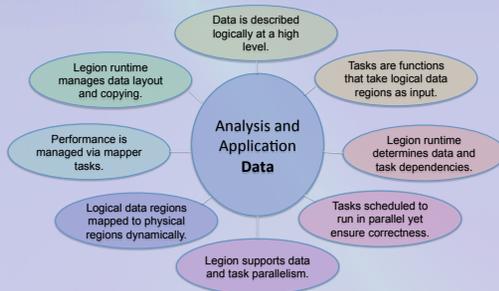
## Legion: Data-Aware Programming Model

This project is based on the Legion Programming Model and Runtime System. (<http://legion.stanford.edu>).

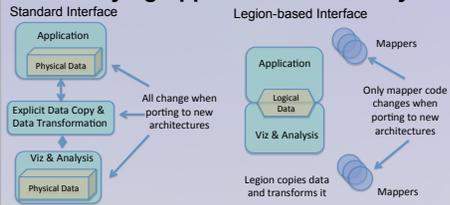
Legion has scaled to over 8,000 nodes with a new implementation of the S3D combustion code on Titan at OLCF.



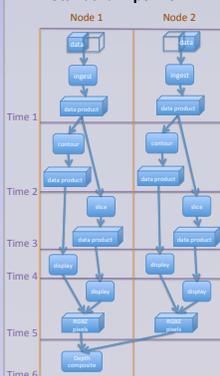
## Legion Programming Model



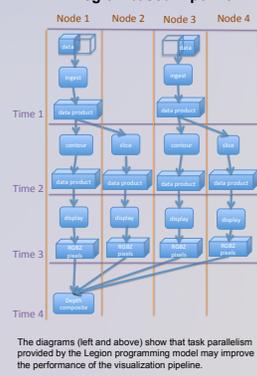
## Unifying Applications and Analytics



## Standard Pipeline



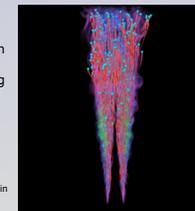
## Legion-based Pipeline



## Co-Design Applications And Architectures

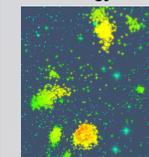
### Combustion

We are focused on several *in situ* analytics and visualization tools used to identify regions where a flame is extinguishing or auto-igniting. These will all utilize the Legion-based implementation of S3D.



At right: Millions of particles in this combustion simulation track autoignition in a lifted jet flame. Image courtesy Jacqueline Chen et al.

### Cosmology



The visualization of large datasets generated by cosmology simulations (sometimes more than a trillion particles) at full resolution poses a major challenge. We will focus on integrating data-driven *in situ* halo-finding capabilities and other filters into HACC, an N-body cosmology code designed for extreme scaling.

Image courtesy Patrick McCormick et al.

**Dark energy** is a hypothetical form of energy that permeates all of space and accelerates the expansion of the universe. **Halos** are dark matter clumps in which gas condensation, star formation and galaxy formation occur.

### Climate

In an effort to understand ocean heat uptake, sea level rise and the ocean's role in global heat transport, climate scientists use Lagrangian tracers as an important diagnostic tool. We will work with climate scientists on analyzing large numbers of tracers within the Model for Prediction Across Scales (MPAS) ocean model.

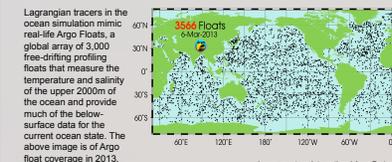
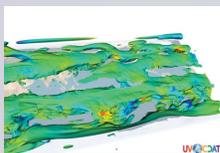


Image courtesy International Argo Project.

## Implementing Analysis and Visualization Algorithms Using Legion

### Data and Execution Model Abstractions for VTK

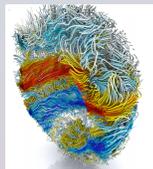
Explore the impact on VTK's existing algorithms and execution models (e.g. push and pull).



Color mapped isosurface of wind velocity produced with VTK inside DV3D inside of UVCDAT. Image courtesy of Kitware.

### Ray Tracing

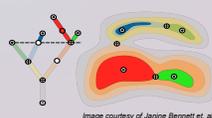
Understand the impact on image generation through ray tracing as both a rendering and analysis infrastructure.



The above ray-traced image is the magnetic reversal of a stellar-like object. Image courtesy of Chuck Hansen.

### Data Transformations

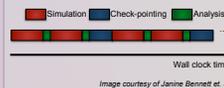
Assess the impact on topological and statistical data transformation techniques that reduce data while maintaining properties of interest for exploratory post-processing, analysis and archival storage.



Merge trees represent the merging of contours as a function is lowered through its range. Merge trees are segmented by thresholding. Image courtesy of Janine Bennett et al.

### Sublinear and Streaming Techniques

Assess how sublinear and streaming techniques can be deployed to provide efficient, adaptive, data-driven control-flow frameworks in a unified framework. These methods access a tiny fraction of the domain, while quantifying the uncertainty introduced due to using only a sample of the data.



Sublinear variants of expensive diagnostics can be used to determine the frequency at which I/O and analysis should be performed *in situ*. Image courtesy of Janine Bennett et al.

### Architecture Co-Design

Our primary hardware interests are in tracking features related to the memory hierarchy, including storage systems: *processor-in-memory* (PIM) and *non-volatile memory* (NVRAM) configurations and their algorithmic and programmability aspects.



Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

LA-UR-15-20103



U.S. DEPARTMENT OF ENERGY

Office of Science