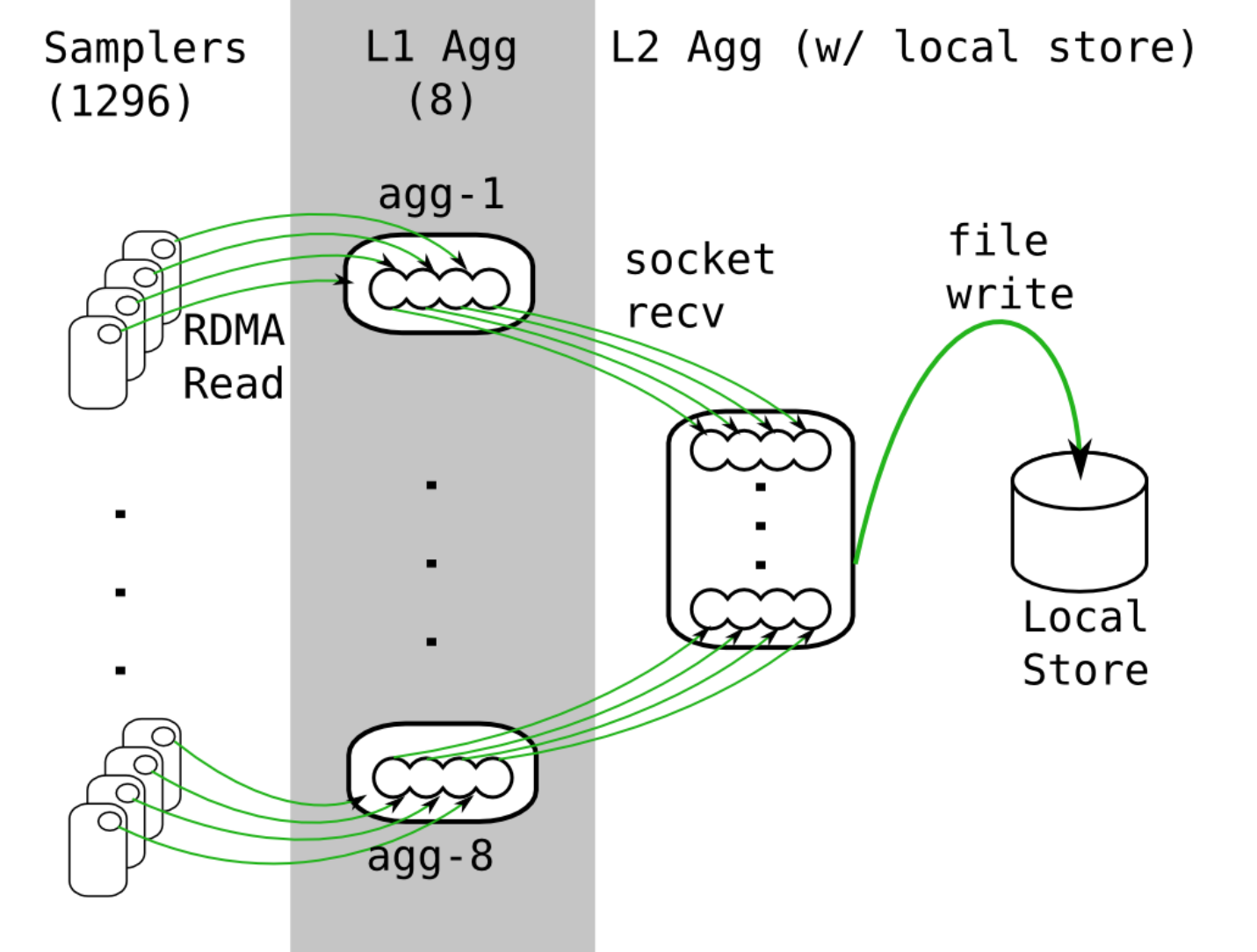
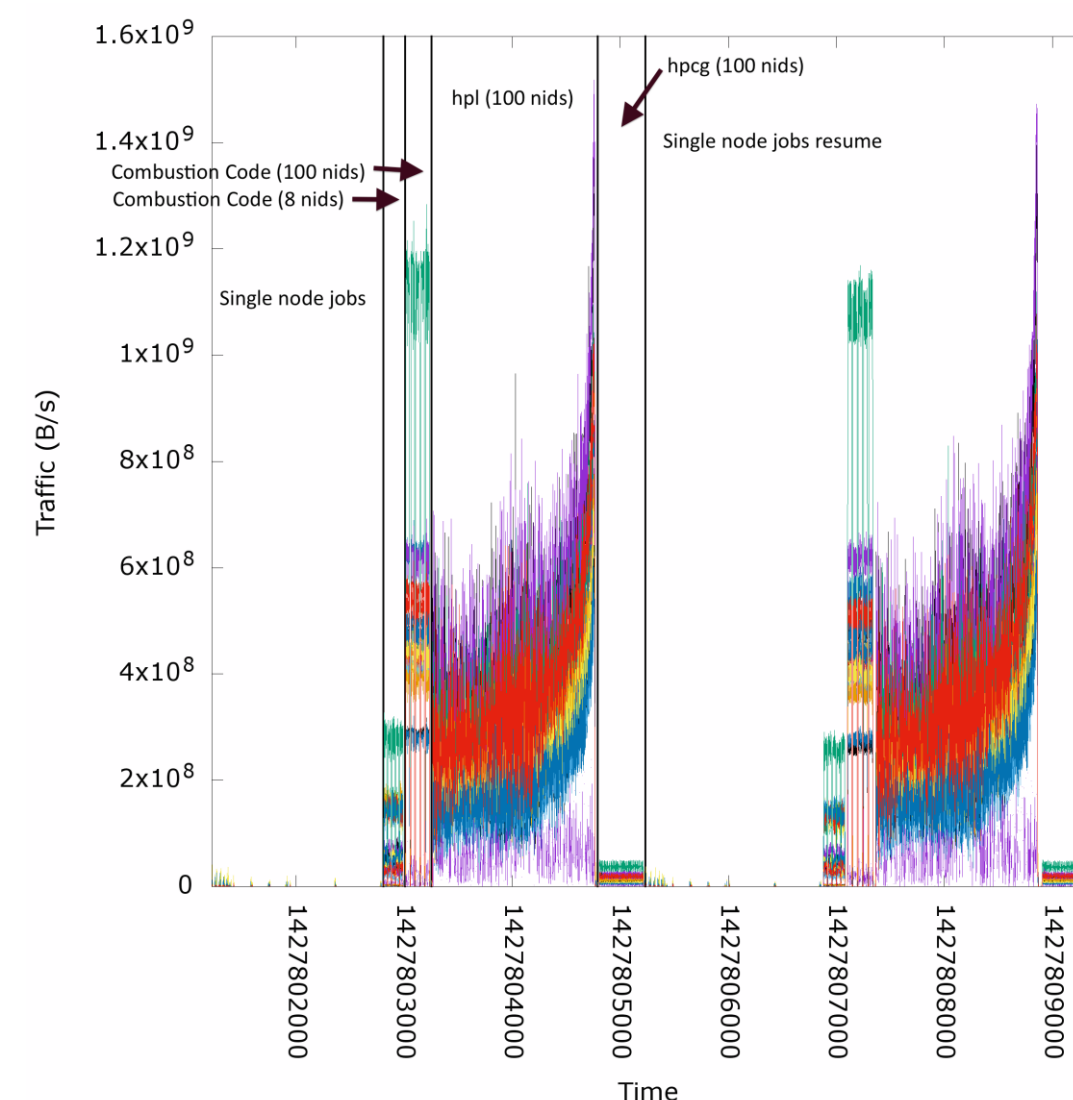


Exceptional service in the national interest



# Understanding HPC System and Application Behaviors

**Intern:** Andrew De La Cruz, UC Berkeley, B.A Applied Mathematics, est. May 2017

**Mentors:** Sophia Lefantzi, Org. 8954, Quantitative Modeling & Analysis, Sandia National Laboratories, Livermore, CA  
Benjamin A. Allan, Org. 9328, Scientific Computing Systems, Sandia National Laboratories, Albuquerque, NM

**Abstract:** Rapid detection and diagnosis of new performance problems or software failures on large high performance computing systems is critical to maintaining the high productivity of these national assets. Early in the diagnostic process, we need integrated timeline visualization of hardware performance metrics and events logged by application and system software to narrow the scope of the search for causes. Across each of thousands of nodes hundreds of hardware counters are collected via the Lightweight Distributed Metric Service (LDMS) [1]. Large streams of free-form log messages about system or application events are also gathered. These data flows are on the order of terabytes per day and higher. We are developing scalable tools to automate extraction and visualization of arbitrary performance metrics and messages. Common pain points in managing HPC applications include memory usage and parallel file system performance during data output. Our test case is to visualize events and performance metrics from CLSVOF (Coupled Level-Set/Volume-of-Fluid) [2], a parallel spray modeling code.

## Introduction:

We are constructing joint visualizations of system collected and production application logged data. Our approach is to build on open-source numerical and graphical tools in the Python scripting language. We generate static images suitable for embedding in web-based data exploration interfaces.

## Motivation:

The existing process for filtering and co-visualizing performance data and log events from production systems is tedious and manual, typically performed for an explanatory presentation days after the issue being explored has been resolved. We want to automate this process sufficiently that visualization becomes a practical screening tool in problem diagnosis and cause detection. The signal streams (both performance counters and events) are often highly correlated in normal operations, but not always consistently. Thus, visual prescreening and reduction of the data may be necessary before machine-learning based approaches to correlating event and performance data can be fruitful.

## Test case:

**Aim:** Assess relationships among events and application performance features.

## Technical Approach:

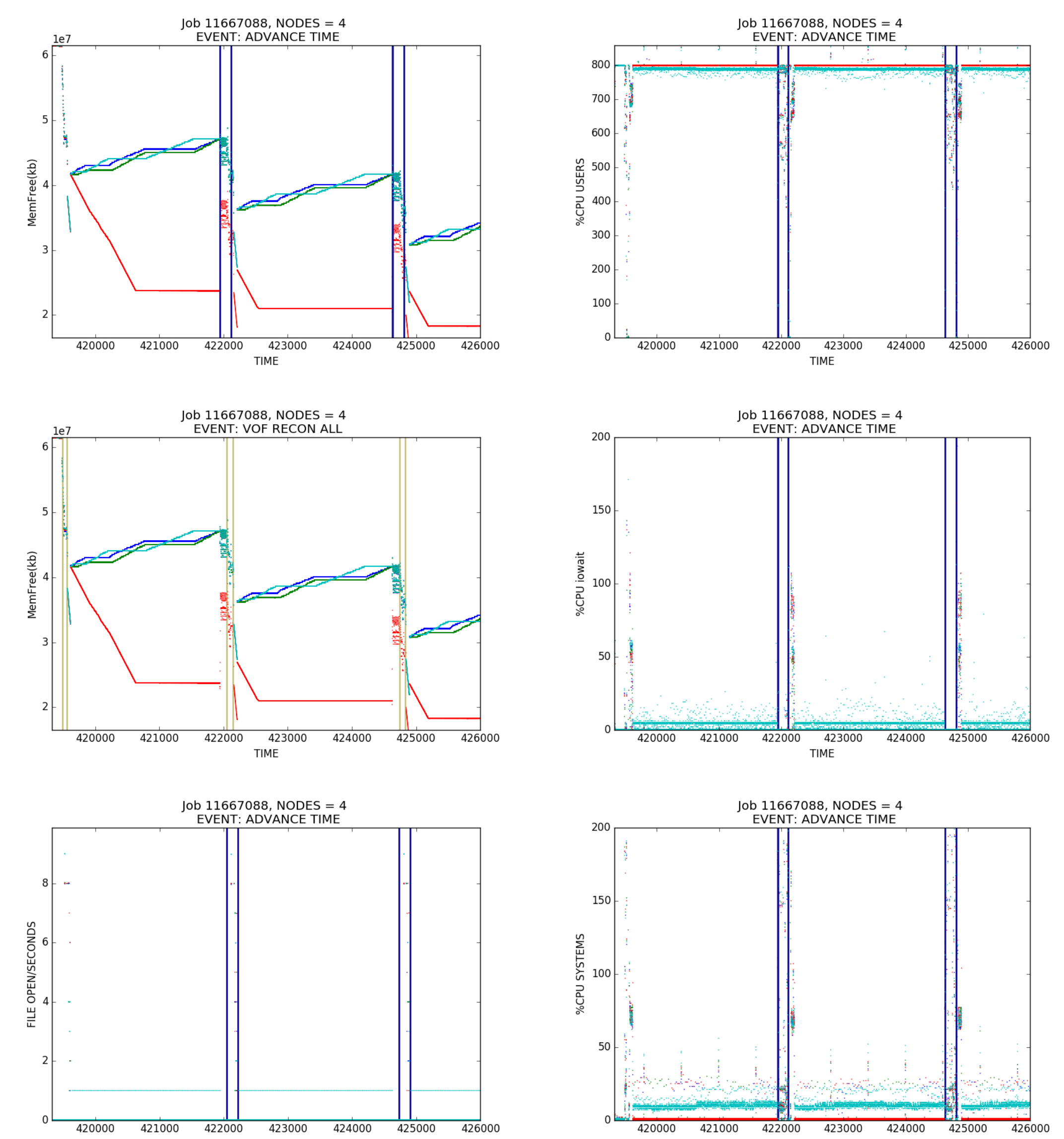
- Gather application log and LDMS data from runs.
- Identify key application log events.
- Script combining requested event and performance metric pairs graphically.
- Demonstrate our technique on CPU and RAM usage and compute storage bandwidth usage paired with numerical solution and output events.

## Application details:

Predicting the effect of liquid fuel injection in engines is the goal for the CLSVOF code. The simulation cycles through stages of grid decomposition, computation, and file output on nodes with 16 cores and 64 GB RAM.

**Results:** Annotated performance plots. Application use of the available CPU time falls during output where IO wait and system intrude. The “advance” phase

MemFree behavior can be decomposed further, showing the volume-of-fluid recon event.



## Acknowledgements:

Andrew De La Cruz was supported by NNSA.

Marco Arienti (8253) provided the spray simulation code (CLSVOF) for this study.

## References:

[1] [https://ovis.ca.sandia.gov/mediawiki/index.php/Publications\\_and\\_presentations](https://ovis.ca.sandia.gov/mediawiki/index.php/Publications_and_presentations)

[2] A High-Fidelity Study of High-Pressure Diesel Injection, M. Arienti and M. Sussman, SAE Technical Paper 2015-01-1853, 2015.

