

# Managing and Monitoring Scientific Workflows through Dashboards

Roselyne Barreto<sup>1</sup>, Terence Critchlow<sup>2</sup>, Ayla Khan<sup>3</sup>, Scott Klasky<sup>1</sup>, Leena Kora<sup>3</sup>, Jeffrey Ligon<sup>4</sup>, Pierre Moullem<sup>4</sup>, Meiyappan Nagappan<sup>4</sup>, Norbert Podhorski<sup>5</sup>, Mladen Vouk<sup>4</sup>

## Introduction

- Dashboards provide multi-view interfaces that allow research scientists to see condensed information about computations in progress, and computations that have been submitted to the batch systems.
- Research scientists want to monitor and interact with simulations in progress. They want feedback on the state of the run, feedback on the trust they can place into data produced, feedback on the state of the machines they are running on, etc.
- A dashboard displays the most relevant/interesting information about the domain data (simulation), workflow progress (processes), and environment status (e.g., machine queues, and similar), and perhaps some other information of useful to the end-user.
- We present the dashboard system developed by the DOE SDM Center in conjunction with the Center for Plasma Edge Simulation (CPES).
- The CPES project couples multiple coupled codes from multiple machines to use first principle techniques to explore the complex physics on the edge of the plasma in a tokamak.

## Machine Monitoring

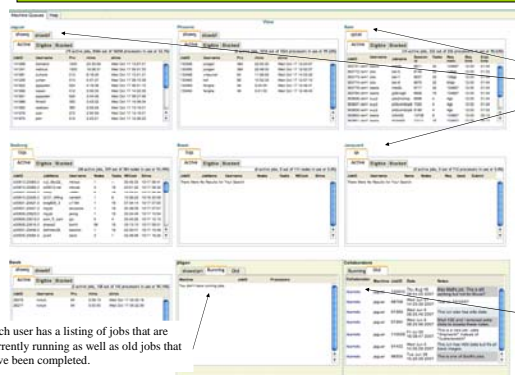


Figure 1. Machine Monitoring of supercomputing resources.



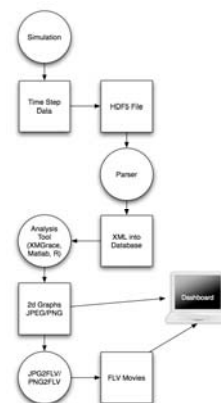
Figure 2. Mouseover information of machine status.

## Technology

- This dashboard subsystem is part of the general environment that supports Kepler. It is LAMP-based and its principal components are, in addition to computational resources (supercomputers), the Kepler orchestration engine, an authentication and authorization system, provenance database with APIs that allow standardized input and output and the dashboard.
- Technologies used to construct and feed the dashboard include AJAX programming techniques; Adobe Flash technology; the Kepler Workflow System; PHP; Python; and web services.
- A more general version of the dashboard with ability to display provenance information on previous simulation runs and drill-down into individual data-sets is being developed as part of the Kepler Scientific Workflow support suite.

## Flow

- A Simulation creates a large amount of data at each time step.
- Some of the data is appended to HDF5 and netcdf files.
- A parser creates a listing of the variables available in the netcdf and HDF5 files and stores the list in a database.
- A process runs through every variable listed in the database and uses XMGrace to create xy plots of the data from the netcdf data, and contour plots using the HDF5 data.
- The 2d graphs are now compiled into FLV movies for viewing on the Dashboard, or are loaded into the Dashboard as simple images as the Dashboard user drags them into the Visualization window.



## Simulation Monitoring

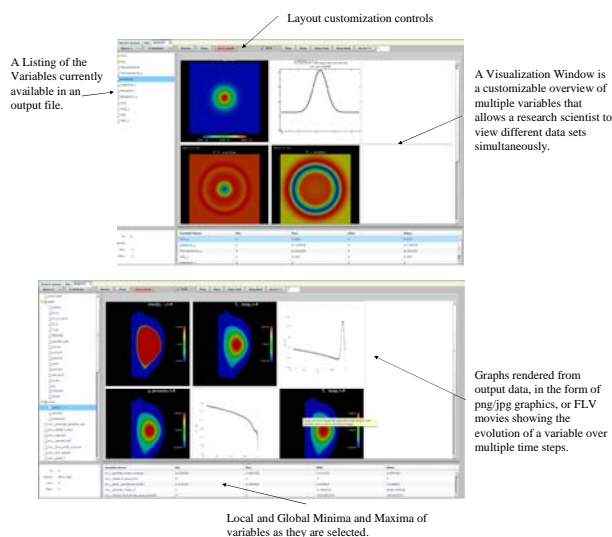


Figure 3. Snapshot of a production run

## Scope

Our current production-level implementation concentrates on:

- monitoring large supercomputing resources and clusters;
- monitoring of jobs on these large resources;
- submitting jobs, editing input files, and interacting with remote resources;
- organizing and displaying simulation runs on these resources, including the capturing of annotations to describe the simulation;
- monitoring the simulation itself in real-time and for later post processing;
- displaying scientific information on the dashboard;
- methods for interacting with running simulations and how this interacts with specialized workflows for controlling simulations.

A **key feature** is that we allow **recording of annotations** in a database, and capturing of the provenance. This gives scientists the ability to track all elements of a simulation. Information processing includes: graphs of data (xy, contour, 3D) along with the time related data, annotations of these graphs, comparison of the graphs and data within the run and with external data and similar