



# DOE Workshop on Petascale System Integration into Large Scale Facilities

**William T.C Kramer**  
**NERSC General Manager**  
kramer@nersc.gov  
510-486-7577

Ernest Orlando Lawrence  
Berkeley National Laboratory



This work was supported by the Director, Office of Science, Division of Mathematical, Information, and Computational Sciences of the U.S. Department of Energy under contract number DE-AC03-76SF00098.

# Workshop Goals

- **Identify challenges and issues involved in the installation and deployment of large scale HPC systems**
- **Identify best practices for installing large-scale HPC systems into scientific petascale facilities**
- **Identify methods to assure system performance and function continue after initial testing and deployment**
- **Identify systematic and research issues for vendors, sites and facilities that would improve the timeliness and quality of deployment**
- **Share tools and methods that are helpful in expediting the installation, testing and configuration of HPC systems**
- **Establish communication paths for technical staff at multiple sites that might make HPC installations more effective**
- **Make recommendations to DOE and other stakeholders to improve the process of HPC system deployment.**



# Workshop Structure

- **The agenda is meant to encourage conversation and information exchange**
  - Limited number of presentations
  - Much opportunity for dialogue
- **Large enough group to cover the range of issues,**
- **Small enough group to go into depth**
- **Some motivational examples**
- **A set of breakouts**
- **Reports from the breakouts and opportunity to discuss**
- **Panels to cover specific topics**



# Workshop Structure

- **May 15, 2007 – 8 am start**
  - **First Session - Plenary**
  - **Introduction and Logistics – Bill Kramer/Yeen Mankin**
  - **Welcome – Dan Hitchcock**
  - **Motivation for the Workshop – Bill Kramer**
    - **System Integration at LLNL – Mark Seager**
    - **System Integration at NCAR – Tom Bettge**
- **Break**
- **Second Session – Breakouts (more later)**
- **Third Session – Plenary**
  - **Reports from breakouts**
  - **Panel – The Vendor Side of Deployments**
    - **TBD – Jeff Beckelhimer - Cray, Chulho Kim - IBM, Wayne Vieira - Sun, Dave Sundestrom - Linux Networkx**



# Workshop Structure

- **Working Dinner**
- **Panel of the Whole –**
  - **If only I had known! –the biggest blunders/mistakes and humorous experiences in large system deployments – All**
- **Day 2 - May 16, 2007 – 8 am start**
- **Fourth Session – Breakouts**
- **Fifth Session – Plenary**
  - **Reports from breakouts**
  - **Panel Session – How will Petascale systems change what we have been doing?**
    - **Ray Bair (ANL), Patricia Kovatch (SDSC), Brad Comes (DODMod), Bob Ciotti (NASA)**
- **Sixth Session – Plenary**
  - **Report Summary**
- **Conclusion**



# Workshop Overview

## Why are we here?



# What Scientists Want from an HPC System

- **Performance**
  - How fast will a system process their work if everything is perfect
- **Effectiveness**
  - What is the likelihood they can get the system to do their work at the performance they expect
- **Reliability**
  - The system is available to do work and operates correctly all the time
- **Consistency/Variability**
  - How often will the system process their work as fast as it can
- **Usability**
  - How easy is it for scientists to get the system to go as fast as possible

## PERCU



# The Steps of Getting HPC to Science

1. Identify Science Requirements
2. Develop system/procurement requirements
  - Specifications
  - Tests
3. Do Market Assessment
4. Procurement (or just build)
  - Many different methods
  - Based on the site specific goals, mission, expertise, culture, budget, history
5. Vendors Respond
6. Negotiate and Award
7. Vendors build a system and maybe test what they build
8. Site Preparations



# The Steps of Getting HPC to Science

## 9. System Arrives and Installed

10. Set the system up the way we want to use it

11. Site testing of the system

12. Acceptance of some form

- Test, evaluate, mitigate, negotiate

13. Transition to Operations

14. Hopefully a long and useful life for the system



# The Steps of Getting HPC to Science

- **Steps 1-8 are well understood**
  - Many rules and regulations for procurements
  - Many system engineering classes for requirements definition
  - Enormous body of work on performance benchmarks
  - Workshops of HPC procurements
    - ISC 2006
    - SC 06
- **Steps 5, 7 and 9 are well understood by vendors (we hope)**
- **Step 14 is well defined and many metrics of successful operation**



# The Steps of Getting HPC to Science

- At the end of Step 9 we are filled with excitement and optimism
- At the beginning of step 14, we are typically exhausted and maybe disappointed
- Why?
- How can we do better?



# The Steps of Getting HPC to Science

- **Steps 10-13 are really a “black art”**
  - We all do it
  - It is the time when we evaluate whether the systems are meeting expectations and what to do where they are not.
- It is where we make key decisions and compromises
- It is also where we make the biggest mistakes that we have to live with
- This is potentially very expensive for vendors and sites
- **Questions**
  - How do we do this?
  - Do we do it as well as we could?
  - How can we do it better?
  - Do we do it often enough?



# Conflicting pressures

- **How fast can scientists get on the system**
- **How well must the system work**
- **Funding constraints and timing**
- **Technology limitations**
- **How well the system was specified**
- **How good are the tests**
- **How much pain can our users tolerate**
- **Publicity**
- **Project schedule vs best system**
- **Promises vs. reality**
- **Rapid learning curve for all**
- **Vendor business considerations**



# Do we have sufficient Guidance

- **Traditional Project Management may not be sufficient**
  - Many procurement issues overlap
  - Assessing progress is hard from the outside
  - How long should things take?
- **Some items for Integration are set based on the acquisition, but many are not**
- **What is the right thing to do**
  - **Situational Morality?**



# Why look at this now?

- **Petascale is upon us**
  - Exascale is coming
- **Some see a new crisis**
  - Technology changes
  - System Scale
- **Others see business as usual**
  - Is fielding a petascale system in today more difficult than fielding a gigascale system 20 years ago?
    - If yes, how so?
- **Either way, this discussion is needed**
  - If a crisis, what do we need and how to overcome
  - If business as usual than we have been doing this for a long time without a manual



# This workshop looks at Steps 10-13

- **There is no single “right way” to do this**
  - **Many dependencies**
    - How the acquisition was done
    - Agreements
    - Timing
    - System Mission
- **There are better and worse ways to do it**
  - **Some problems from sites that shall remain nameless**
    - Declare success and immediately go down
    - Declare success and then have some many issues the user community migrates somewhere else
    - Incremental compromises leads to a total capitulation
- **We are in search of the better ways so we can all use them**



# Breakouts



# Charge for Break Out Session

1. What are the major challenges in this area?
  2. What methods and technology are currently being used and how do we use them
  3. What methods and technology work and which ones do not
  4. What tools and technology do we wish have – particularly for Petascale systems.
  5. Other observations/suggestions/issues
- Breakout Leaders are free to facilitate the meeting
    - Short presentations, discussions, twiddling thumbs – all are ok
    - The goal is a in depth exploration
    - Add, expand or replace the questions
  - Remember the overall goals. Please address them for each subject area
  - Note that is a session fully explores there are, it is also perfectly fine for the session to move on to other topics



# Workshop Goals

- **Identify challenges and issues involved in the installation and deployment of large scale HPC systems**
- **Identify best practices for installing large-scale HPC systems into scientific petascale facilities**
- **Identify methods to assure system performance and function continue after initial testing and deployment**
- **Identify systematic and research issues for vendors, sites and facilities that would improve the timeliness and quality of deployment**
- **Share tools and methods that are helpful in expediting the installation, testing and configuration of HPC systems**
- **Establish communication paths for technical staff at multiple sites that might make HPC installations more effective**
- **Make recommendations to DOE and other stakeholders to improve the process of HPC system deployment.**



# Break Out Session 1

- **Integration Issues for Facilities**
  - Petascale systems are pushing the limits of facilities in terms of space, power, cooling and even weight.
  - There are many complex issues to deal with when integrating large scale systems and these will get more challenging with Petascale systems.
  - While we all hope technology will reverse these trends, can we count on it?
  - Besides building large facilities (at Moore's law rates) how we be better optimize facilities?
  - How can the lead times and costs for site preparation be reduced?
  - Can real time adjustments be made rather than over design?
- **Breakout Leaders - Howard Walter (NERSC), Gary New (NCAR)**



## Break Out Session 2

- **Performance Assessment of Systems**
  - There are many tools and benchmarks that help assess performance of systems, ranging from single performance kernels to full applications.
  - Performance tests can be kernels, specific performance probes and composite assessments.
  - What are the most effective tools?
  - What scale tests are needed to set system performance expectations and to assure system performance?
  - What are the best combinations of tools and tests?
- **Breakout Leaders – Tom Engle (NCAR), Rob Pennington (NCSA)**



# Break Out Session 3

- **Methods of Testing and Integration**
  - There are a range of methods for fielding large scale systems, ranging from self integration, cooperative development, factory testing, and on-site acceptance testing.
  - Each site and system has different goals and selects from the range of methods.
  - When are different methods appropriate?
  - What is the right balance between the different approaches?
  - Are there better combinations than others?
- **Breakout Leaders - Brad Comes (DOD), Buddy Bland (ORNL)**



# Break Out Session 4

- **Systems and User Environment Integration Issues**
  - Breakout session #2 looked at performance and benchmarking tools.
  - While performance is one element of successful systems, so are effective resource management, reliability, consistency and usability, to name a few.
  - Other than performance, what other areas are critical to successful integration?
  - How are these evaluated?
- **Breakout Leaders - Mike McCraney (MHPCC), Jim Rogers (ORNL)**



# Break Out Session 5

- **Break out 5 - Early Warning signs of problems – detecting and handling them**
  - Fielding large scale systems is a major project in its own right, and takes cooperation between site staff, stakeholders, users, vendors, third party contributors and many more.
  - How can early warning signs of problems be detected?
  - When they are detected, what should be done about them?
  - How can they be best handled to have the highest and quickest success?
  - How do we insure long term success vs the pressure of quick milestone accomplishment?
  - Will the current focus on formal project management methods help or hinder?
- **Breakout Leaders - Bob Tomlinson (LANL), Jim Kasdorf (PSC)**



# Break Out Session 6

- **How to keep systems running up to expectations**
  - Once systems are integrated and accepted, is the job done?
  - If systems pass a set of tests, will they continue to perform at the level they start at?
  - How can we assure systems continue to deliver what is expected?
  - What levels and types of continuous testing are appropriate?
- **Breakout Leaders - Dave Skinner (NERSC), Kevin Regimbal (PNNL)**

