LLNL-TR-648169

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

High Performance Computing Operations Review Report

Kimberly C. Cupps

January 6, 2014

Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the U.S. Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344.

High Performance Computing Operations Review Report

Overview

The High Performance Computing Operations Review (HPCOR) meeting—requested by the ASC and ASCR program headquarters at DOE—was held November 5 and 6, 2013, at the Marriott Hotel in San Francisco, CA. The purpose of the review was to discuss the processes and practices for HPC integration and its related software and facilities. Experiences and lessons learned from the most recent systems deployed were covered in order to benefit the deployment of new systems.

Although the meeting continued the series of Best Practices Workshops that have been held previously, it was conducted as a DOE internal meeting to best address the issues related to ongoing procurements. In attendance at the meeting were subject experts from six DOE laboratories who are currently involved in the Trinity (LBNL/NERSC, LANL, SNL) and CORAL (ORNL, ANL, LLNL) collaborative procurements for the next generation of HPC systems. Each new generation of systems presents challenges in all aspects of deployment, including making the systems productive for the user community. Understanding the best practices of all laboratories can contribute to the successful installation of new systems.

The meeting commenced with a plenary session presentation by Bruno Van Wonterghem (LLNL) in which he described the challenges associated with the installation and operation of the National Ignition Facility (NIF) at LLNL. NIF has many aspects in common with the complex installation, power management, user scheduling, and operations of HPC systems. Next, Sue Kelly (Sandia) described Sandia's efforts to derive a use-case approach for power API requirements—a topic identified as a missing system component at an earlier Best Practices meeting.

The remainder of the meeting was devoted to discussions on eight technical topics identified by the organizing committee. These topics were discussed in parallel breakout sessions over two days and documented in breakout session reports to the full group. Each breakout session included at least one representative from each laboratory, and participants were asked to address specific questions related to the technical topic covered, the management structure of the activity, and the experiences in recent deployments. Some discussion highlights from the eight technical topics are summarized in the Technical Topics section. Appendix A is the meeting agenda and breakout session report topics. Appendix B identifies the organizing committee members and meeting attendees. Appendix C includes all breakout session reports.

Technical Topics

The following technical topics were addressed in the breakout sessions. Although each topic was independent, there was some overlap, and four related pairs of topics were identified by the organizing committee.

A. Systems (integration and the operating system)
B. Applications (benchmarking and working with code teams)
C. User Environment (development environment and user support)
D. Facility Preparation (external and internal)

Each day's breakout session technical topics are detailed below, followed by highlights from that session's discussions.

Day One Breakout Sessions

A1: System integration, early access, acceptance testing, and system shakedown prior to general availability; getting all hardware/software and file systems to work as advertised.

Acceptance testing is an iterative process, and automation is your friend. Development of controlled procedures to "break" the system is a useful process to help the system administration staff experience the symptoms of error conditions and identify any lingering system recovery issues. Integration of the new systems into configuration management systems and monitoring allows some of the acceptance test results to be easily verifiable and also allows for repeatable environments and simplifies transition to operations. Development of a representative set of acceptance tests is needed to verify system software changes before releasing them to users.

B1: Use of modeling, simulators, and benchmarking.

As DOE pursues more collaborative procurements, benchmarks need to be selected with more focus and purpose rather than a union of what each laboratory previously used. Access to early systems is vital. The benchmark suite can also play a role in identifying resiliency issues because the behavior is known; therefore, benchmarks and benchmarking teams contribute throughout the life of the system. Use of modeling does not yet have the same level of confidence as that of benchmarking. Likewise, system simulators have not yet played a significant role. A strong synergy among the HPC centers, users, and developers is needed in this area because of the significant risk taken when selecting tomorrow's systems using today's codes.

C1: Development environment preparation (parallelism support, compilers, tools).

A useful HPC development environment does not come shrink-wrapped from the vendor. There are endless activities to address, so the work must be prioritized. Each lab prioritizes differently, but every lab would like more resources devoted to addressing the development environment. Users want a feature-complete environment that is compatible with other systems. Early access to "real hardware" is important so that areas in which there have been problems on previous systems can be addressed as soon as possible. It is useful to have at least two early environments, one for early user testing and one for development testing. One laboratory makes three test environments available. It is also essential to have redundant sources for critical software, especially compilers and debuggers.

D1: Facility and utility planning impacts, demand response forecasting, etc., outside the data center.

Infrastructure planning should begin with the site master plan, updated annually. The annual updates should integrate into the long-term sustainability for the site infrastructure and should not be confined to the requirements of a single HPC system. When site infrastructure interfaces with outside utility providers, reinforce their assurances to meet (or sometimes not meet) the updated requirements with laboratory modeling, and be mindful that the modeling tools used by the utility companies for impacts of HPC—especially the swings in load—are sometimes antiquated.

Facility preparation teams and HPC operations are cooperative. For both the CORAL and Trinity procurements, the facility teams were involved early, which proved to be constructive for the projects and should be considered a best practice.

The most recently recognized risk is with water quality and water supply temperatures required by HPC vendors, often resulting in re-work, additional costs, and schedule delays to the project. Requirements set by vendors should be technically validated and practiced by the vendor as well.

Day Two Breakout Sessions

A2: Run-time operating system environment, including logging, monitoring, schedulers, and allocations.

Some activities can be handled with vendor interactions before the system arrives. NRE is critical because the vendors are not likely to meet the needs without explicit funding. Required diagnostic information is important and should be part of the system RFP. If early hardware access is available, it should include switches, interconnects, etc., as well as nodes. Learning early on about hardware failure scenarios and their impact on the operating system helps verify that redundancy designs work, but failure tests need to be simple. Configuration management with vigorous validation is needed to maintain a consistent environment for the users.

B2: Working with code teams.

Success of the newer systems depends critically on robust, well-funded, early, and active involvement with code teams. Application readiness facilitators need to be prepared to do “whatever it takes.” The appropriate people need to be involved early so that appropriate NDA presentations and workshops can be scheduled. Early hardware access is critical. NRE funding can be useful, but clear goals need to be established. Often the most key vendor interaction is identifying the individual in the vendor organization who really has insight into the hardware performance. This individual can address the most important optimizations for a given application. To avoid squandering optimization efforts, it is important to avoid over-investing in dead-end disruptive technologies or early architectures that do not match a final platform. There are, however, some code-restructuring activities that do tend to pay off across architectures (such as attention to I/O readiness.) In support of system resiliency and failure diagnostic efforts, improved monitoring and tools are needed.

C2: Usage models, user education and training, and user support.

This group should be involved early to gather requirements to inform the procurement effort and serve as an advocate for users during the procurement process. Usage Model documents have been produced for more than 10 years for the ASC systems and have been used as part of the documentation for the DOE CD process. System administrators use the document for system configuration, user support staff use it to develop training materials, and the users have an end-to-end description/functional specification. Once the new architecture is known, it is important to begin transitioning users and codes to the new architecture through user interactions and training. New systems must be integrated into a larger administrative infrastructure for accounts, reporting, and ticket systems. Early hardware access is critical to ensure that the support staff is ready to assist users. Likewise, access to other labs' systems is helpful for testing before general user access.

D2: Facilities preparation inside the building; getting building/floor space ready, platform operational requirements and tolerances (cooling temperatures, weights, etc.).

The delivery and management of the high-performance power, space, and cooling capabilities within HPC facilities must balance the long-term issues associated with the facility itself with the short-term requirements of the HPC systems that occupy the space. While the facility will have a design life that is measured in decades, the individual systems housed in these facilities will typically exist for approximately five years. These contrasting requirements create opportunities for identifying best practices that can ensure that the investment in the facilities and the systems are appropriately balanced as new systems are acquired, installed, introduced to production, and eventually decommissioned. Issues addressed within this session included the requirement for integrated project plans, long-term facility/system master plans, the collaborative roles and responsibilities of facility and systems engineers, the impact of early system assessments, innovative packaging, the intentional and early involvement of the system vendor, and the role of NRE funding to identify electrical and mechanical performance improvements, drive packaging innovation, and clarify system requirements that can reduce cost, schedule, and operational risk.

Summary

Some common themes emerged in the breakout reports.

- Most groups expressed the need to be involved very early in a procurement process, but one participant observed that having everyone involved early can create communication problems. Perhaps there should have been a track to discuss project management.
- Access to early representative systems was vital to all groups, with a caution about wasting time with dead-end prototypes.
- Close and early interaction with the major code teams was also a cross-cut theme for getting benchmarks, preparing the development environment, application readiness, and user support and training. The organization of this effort varies from laboratory to laboratory but needs focus to avoid too many requests for code team attention.
- Several groups expressed the desire to interact more frequently, possibly with site visits. A good summary observation was made by the meeting chair, Kim Cupps: "These reviews are good opportunities to talk with people who perform the same tasks differently and to learn from each other. This review was very productive."

At the conclusion of the meeting, a list of suggestions was informally submitted to DOE about topics that might be addressed in future meetings. This review report, individual breakout session reports, and additional information are available on the HPCOR website.

Appendix A. Meeting Agenda and Breakout Session Topics

Monday, November 4	
5:30-7:30 p.m.	Welcome reception; registration; meeting organization activities
Tuesday, November 5	
8:15-10:00 a.m.	Welcome; meeting overview
	Plenary talk 1: NIF operations success/best practices (Bruno Van Wonterghem, LLNL)
	Plenary talk 2: Use case approach to deriving an HPC software power API (Sue Kelly, SNL)
9:30-10:00 a.m.	Break
10:00 a.m.-12:00 p.m.	Day 1 breakout sessions
12:00-1:30 p.m.	Lunch (not provided)
1:30-2:30 p.m.	Day 1 breakout sessions (cont.)
2:30-3:00 p.m.	Break
3:00-5:00 p.m.	Day 1 breakout session reports
Wednesday, November 6	
8:15-8:30 a.m.	Remarks from HQ; questions and answers
8:30-11:30 a.m.	Day 2 breakout sessions (take break when convenient)
11:30 a.m.-1:00 p.m.	Lunch (not provided)
1:00-3:00 p.m.	Day 2 breakout session reports
3:00-3:30 p.m.	Break
3:30-4:00 p.m.	Meeting wrap-up; report instructions, dates, etc.

Breakout Session Discussion and Report Topics

The following issues should be addressed in breakout discussions and reports:

Processes

- What needs to be done at what level of effort and cost?
- What begins first: timeline for activities, e.g., before or after hardware?
- Role of early hardware access (either locally or remotely) and prototype systems
- Role of vendor partnerships.
- Role of R&D&E.
- Resiliency activities (e.g., redundancy) executed.

Organization and Management

- What is the structure of the integration and preparation teams? Who manages and oversees the different components?
- Skills for activity team.

Experiences and Lessons Learned

- Experiences/lessons learned (good and bad).
- Most productive activities.
- Resiliency experiences.
- What was highest risk? Was it a surprise or expected?

Appendix B. Organizing Committee and Attendees

The DOE sponsors overseeing the planning of this meeting were Thuc Hoang, DOE/NNSA/ASC, and Dave Goodwin, DOE/SC/ASCR. LLNL's Kim Cupps, assisted by Mary Zosel, chaired the meeting planning. The multi-lab steering committee that participated in the planning of the agenda included Susan Coghlan (ANL), Hal Armstrong (LANL), Richard Gerber (LBNL/NERSC), A. "Buddy" Bland/Jim Rogers (ORNL), and Tom Klitsner (SNL). Meeting logistics were managed by Lori McDowell and Jennifer Rose (LLNL). The following laboratory representatives attended the meeting.

Affiliation	Attendee
ANL	Coghlan, Susan Harms, Kevin Howe, Thomas Loy, Raymond Meng, Jiayuan Williams, Timothy
DOE Office of Science/ASCR	Goodwin, David Harrod, William Helland, Barbara
LANL	Armstrong, Harold Baird, Charles Green, Jennifer Kelly, Kathleen Velarde, Ron
LBNL/NERSC	Broughton, Jeffrey Cardo, Nicholas Deslippe, Jack Draney, Brent Fagnan, Kjiersten Gerber, Richard Jacobsen, Douglas Pezzaglia, Larry Shalf, John Skinner, David Srinivasan, Jay Wasserman, Harvey
LLNL	Bailey, Anna Maria Bertsch, Adam Carnes, Brian Cupps, Kimberly Futral, Scott Gyllenhaal, John Van Wonterghem, Bruno Zosel, Mary
NREL	Hammond, Steve
ORNL	Barker, Ashley Hill, Jason Joubert, Wayne Messer, Bronson Rogers, James
SAIC/DOE NNSA/ASC	Macaluso, Antoinette

Affiliation	Attendee
SNL	Balance, Robert Haskell, Karen Kelly, Suzanne Klitsner, Tom Martinez, Dave Monk, Stephen Pavlakos, Constantine Stevenson, Joel

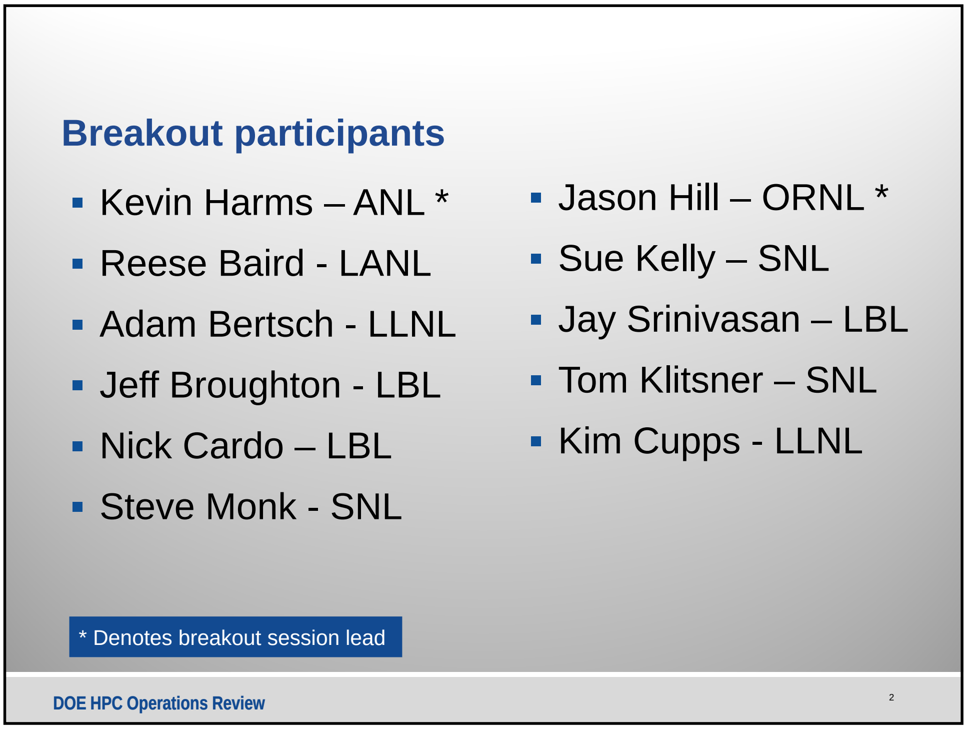
Appendix C. Breakout Session Reports

These eight breakout session reports document the technical topics discussed during each day's breakout sessions.

System Integration Breakout Session

DOE HPC Operations Review
San Francisco, November 5-6, 2013

LLNL-PRES-646612

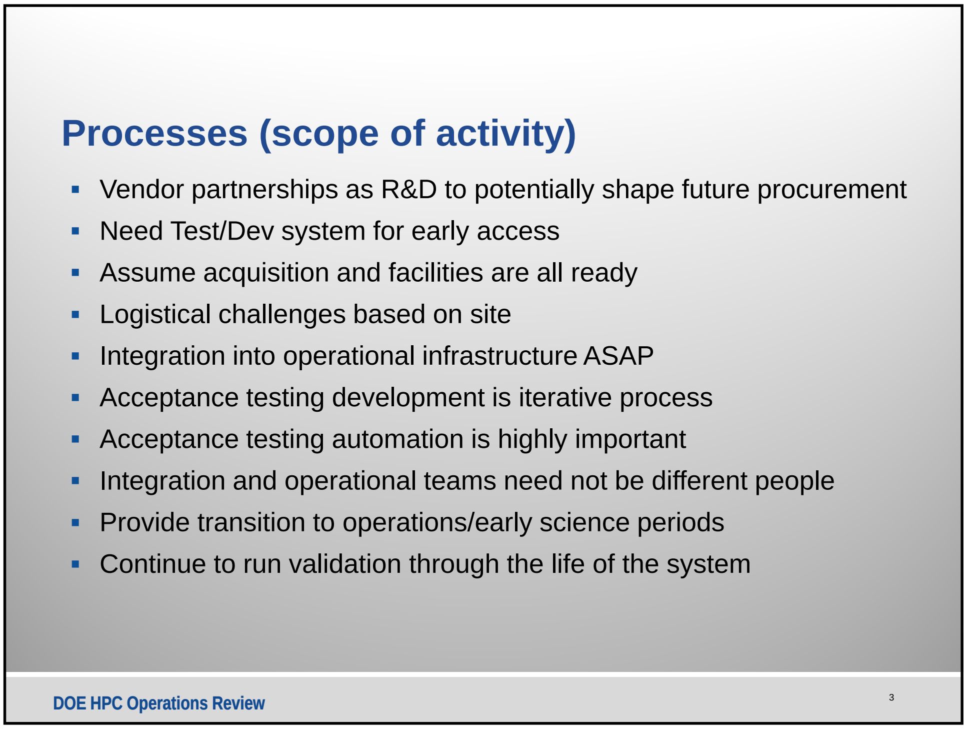


Breakout participants

- Kevin Harms – ANL *
- Reese Baird - LANL
- Adam Bertsch - LLNL
- Jeff Broughton - LBL
- Nick Cardo – LBL
- Steve Monk - SNL
- Jason Hill – ORNL *
- Sue Kelly – SNL
- Jay Srinivasan – LBL
- Tom Klitsner – SNL
- Kim Cupps - LLNL

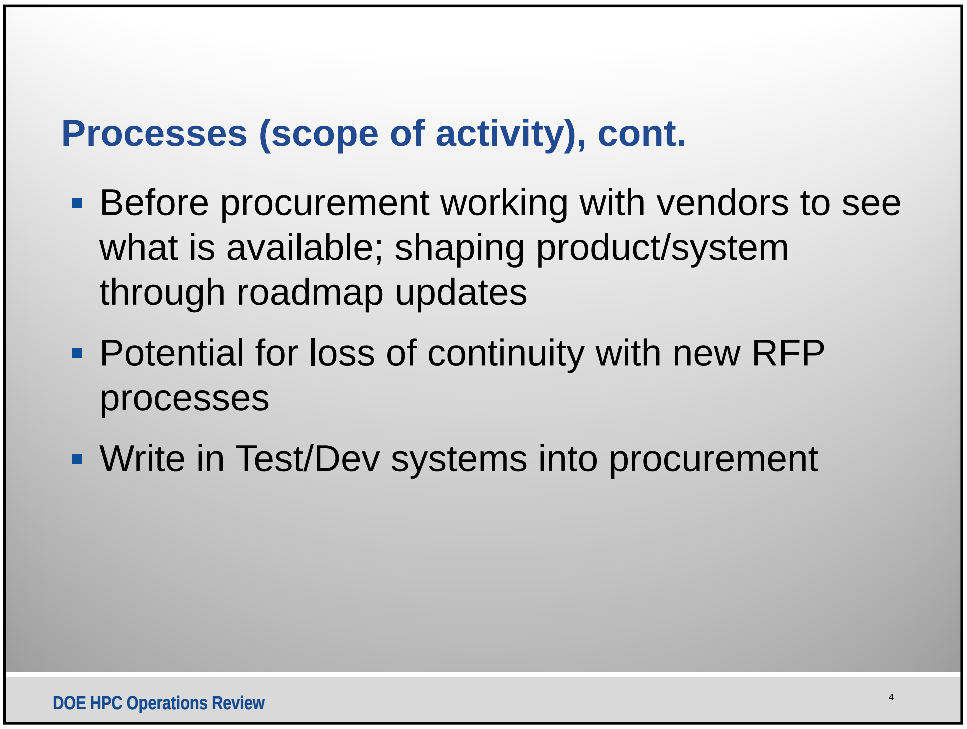
* Denotes breakout session lead

DOE HPC Operations Review 2



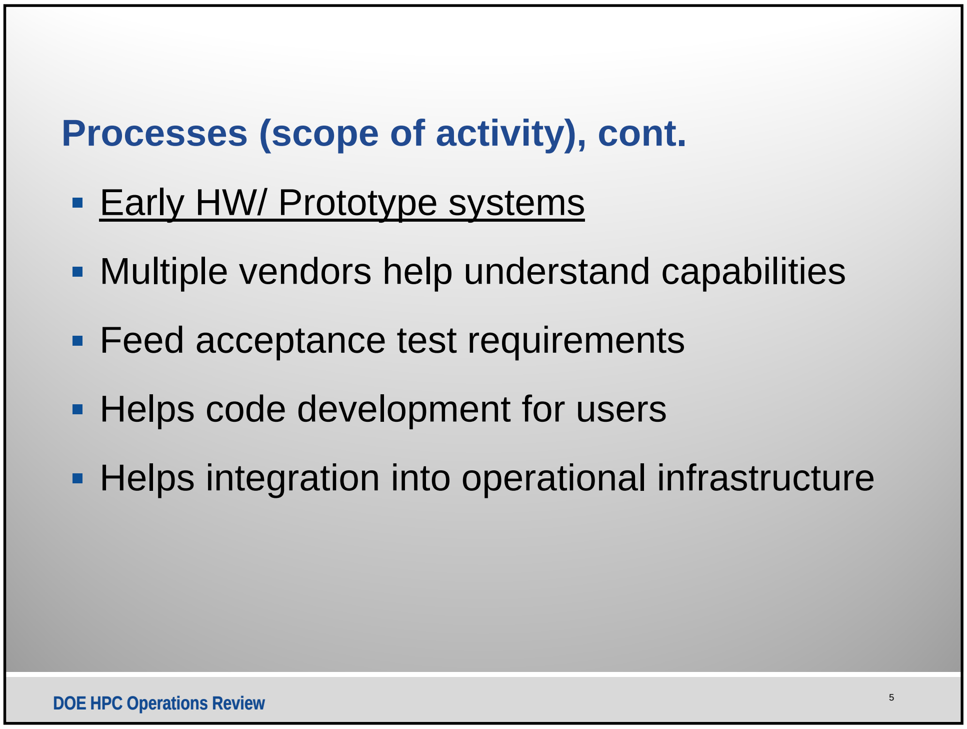
Processes (scope of activity)

- Vendor partnerships as R&D to potentially shape future procurement
- Need Test/Dev system for early access
- Assume acquisition and facilities are all ready
- Logistical challenges based on site
- Integration into operational infrastructure ASAP
- Acceptance testing development is iterative process
- Acceptance testing automation is highly important
- Integration and operational teams need not be different people
- Provide transition to operations/early science periods
- Continue to run validation through the life of the system



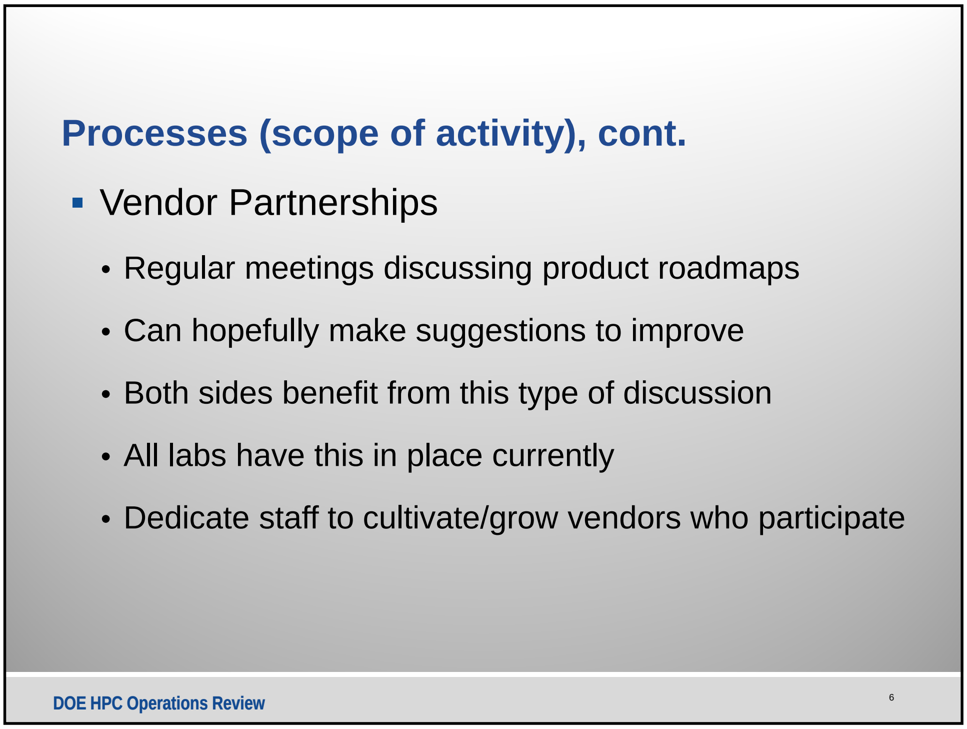
Processes (scope of activity), cont.

- Before procurement working with vendors to see what is available; shaping product/system through roadmap updates
- Potential for loss of continuity with new RFP processes
- Write in Test/Dev systems into procurement



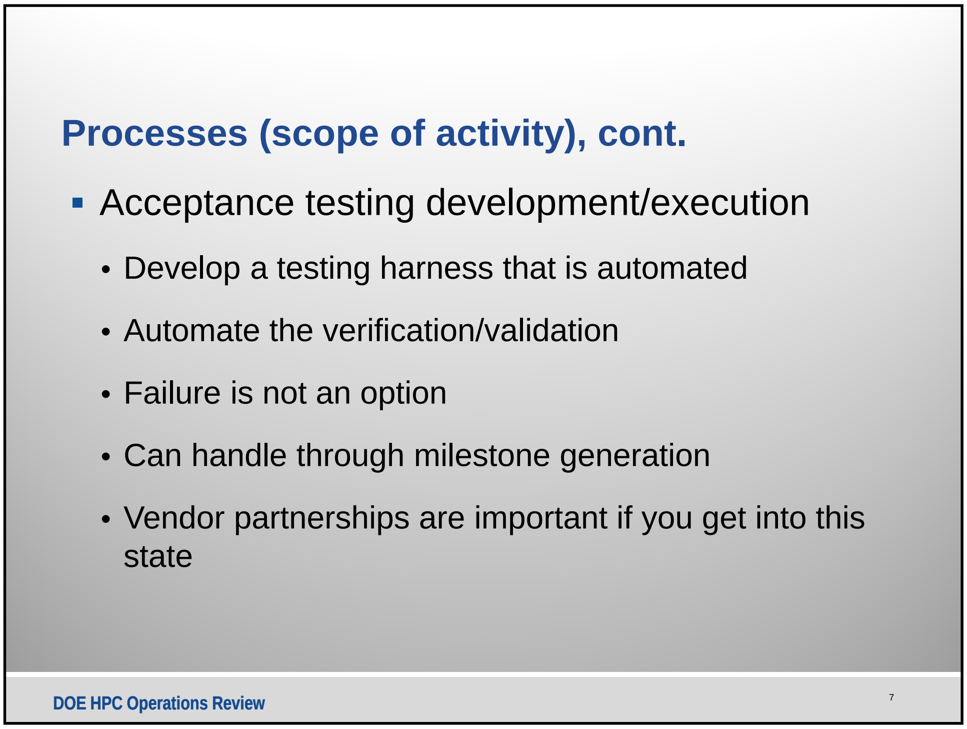
Processes (scope of activity), cont.

- Early HW/ Prototype systems
- Multiple vendors help understand capabilities
- Feed acceptance test requirements
- Helps code development for users
- Helps integration into operational infrastructure



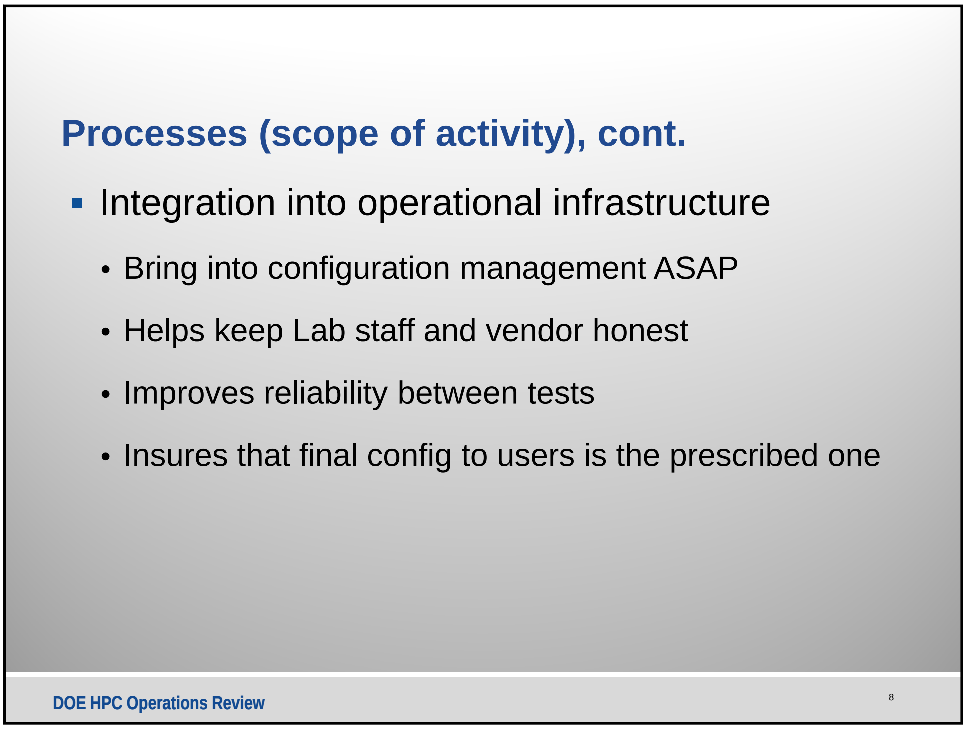
Processes (scope of activity), cont.

- Vendor Partnerships
  - Regular meetings discussing product roadmaps
  - Can hopefully make suggestions to improve
  - Both sides benefit from this type of discussion
  - All labs have this in place currently
  - Dedicate staff to cultivate/grow vendors who participate



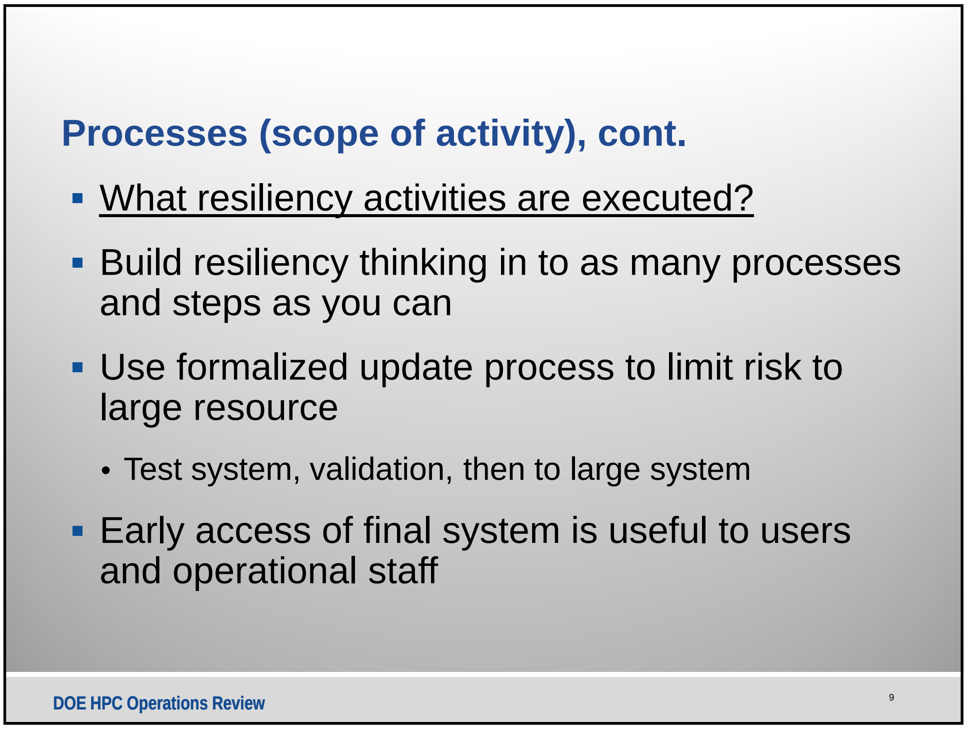
Processes (scope of activity), cont.

- Acceptance testing development/execution
  - Develop a testing harness that is automated
  - Automate the verification/validation
  - Failure is not an option
  - Can handle through milestone generation
  - Vendor partnerships are important if you get into this state



Processes (scope of activity), cont.

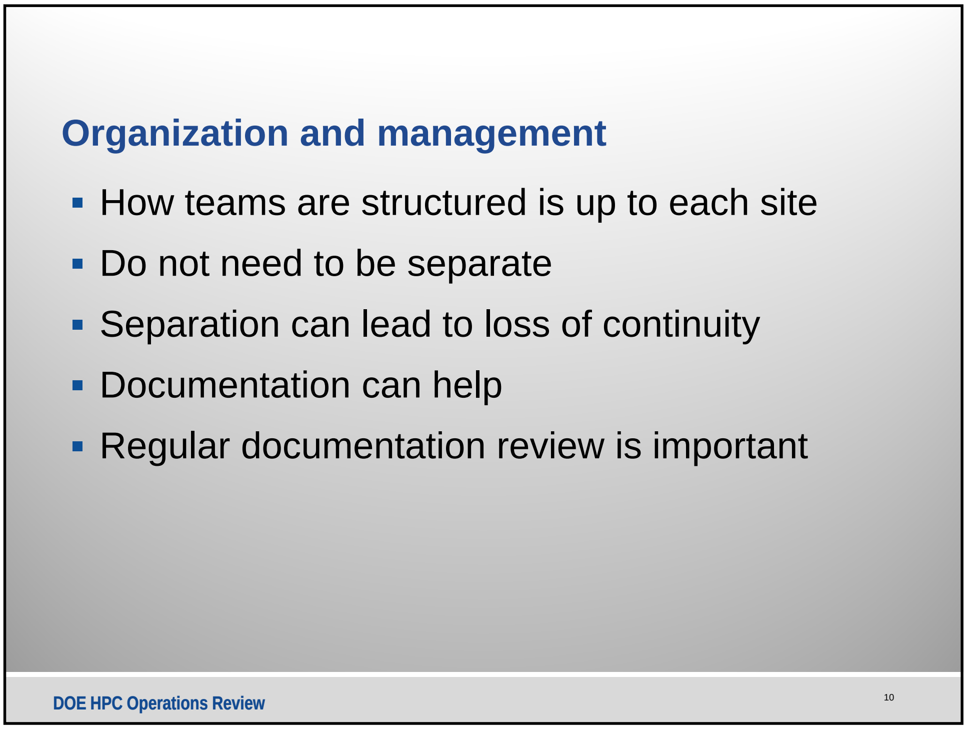
- Integration into operational infrastructure
  - Bring into configuration management ASAP
  - Helps keep Lab staff and vendor honest
  - Improves reliability between tests
  - Insures that final config to users is the prescribed one



Processes (scope of activity), cont.

- What resiliency activities are executed?
- Build resiliency thinking in to as many processes and steps as you can
- Use formalized update process to limit risk to large resource
  - Test system, validation, then to large system
- Early access of final system is useful to users and operational staff

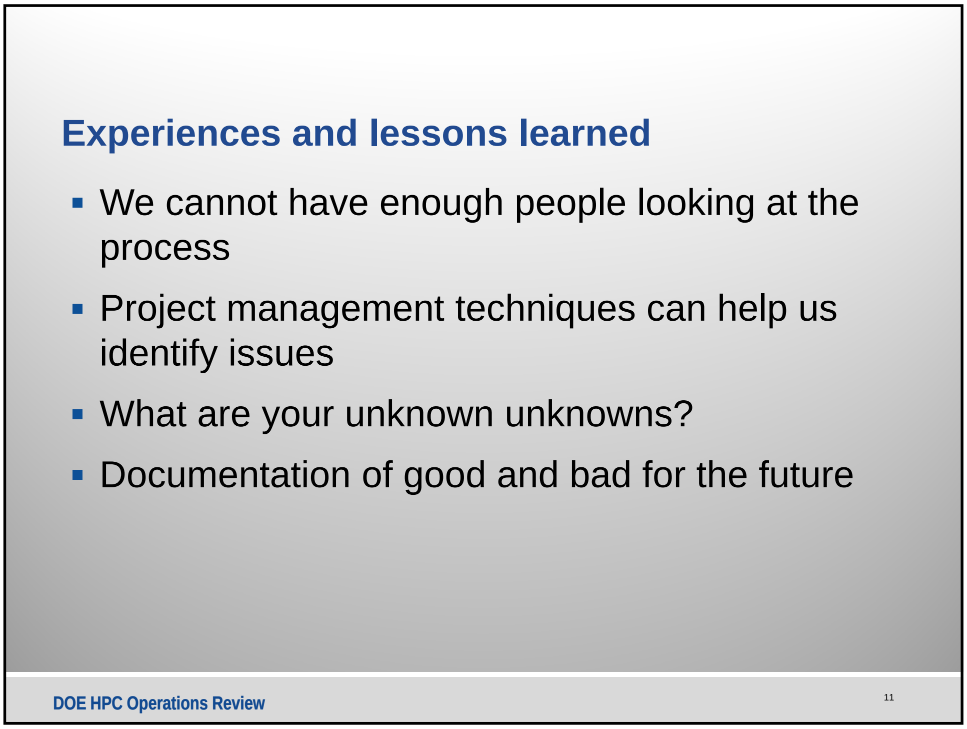
DOE HPC Operations Review 9



Organization and management

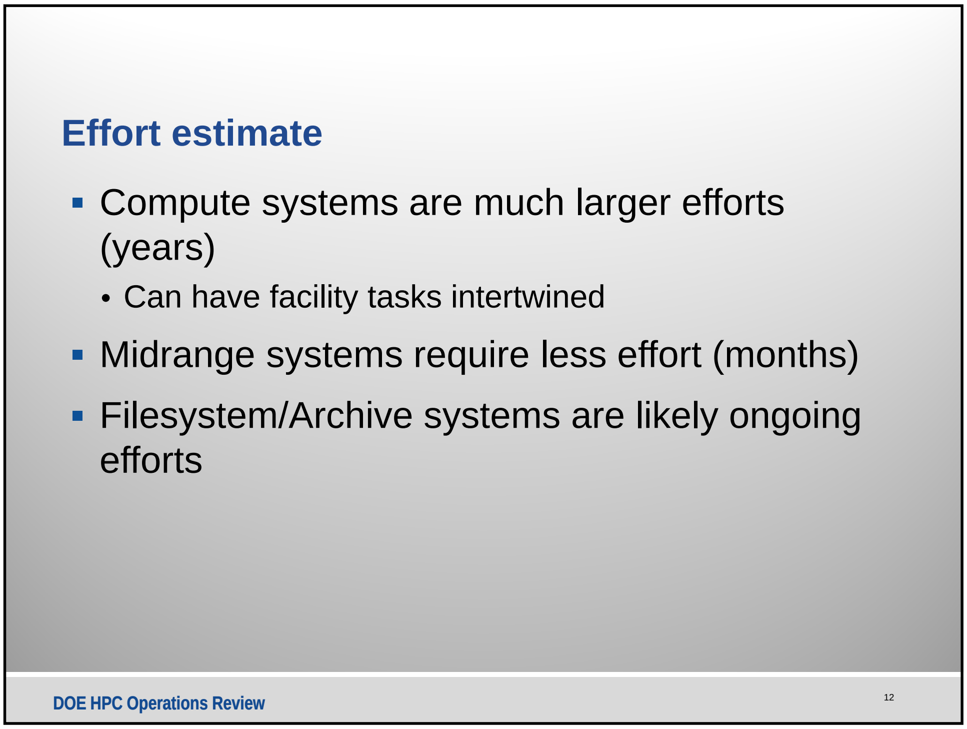
- How teams are structured is up to each site
- Do not need to be separate
- Separation can lead to loss of continuity
- Documentation can help
- Regular documentation review is important

DOE HPC Operations Review 10



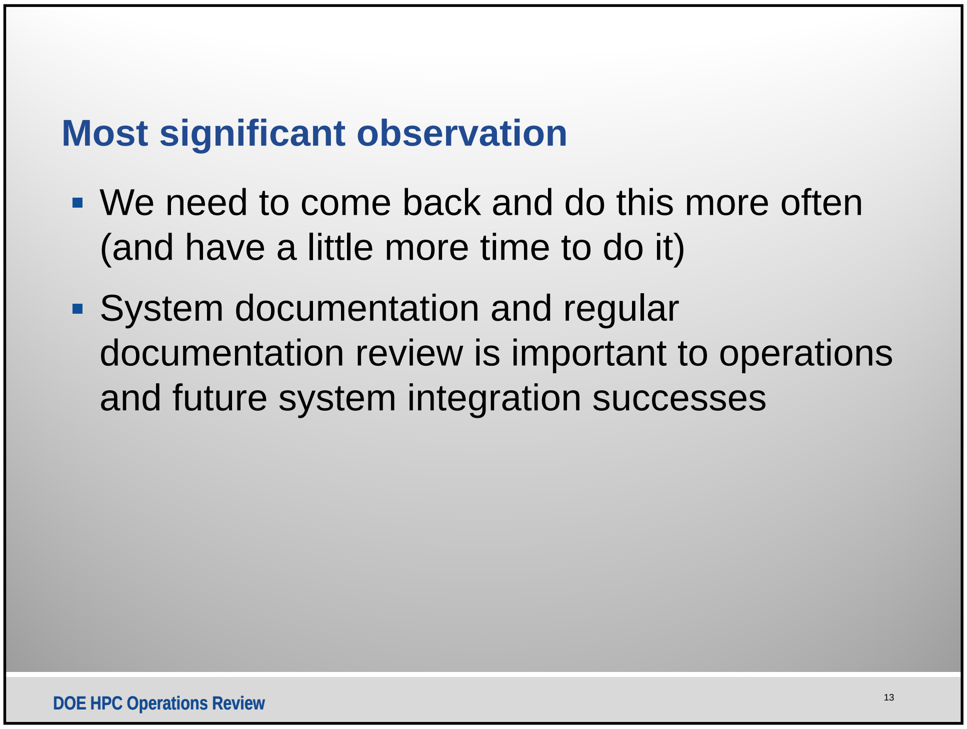
Experiences and lessons learned

- We cannot have enough people looking at the process
- Project management techniques can help us identify issues
- What are your unknown unknowns?
- Documentation of good and bad for the future



Effort estimate

- Compute systems are much larger efforts (years)
  - Can have facility tasks intertwined
- Midrange systems require less effort (months)
- Filesystem/Archive systems are likely ongoing efforts



Most significant observation

- We need to come back and do this more often (and have a little more time to do it)
- System documentation and regular documentation review is important to operations and future system integration successes

DOE HPC Operations Review

13

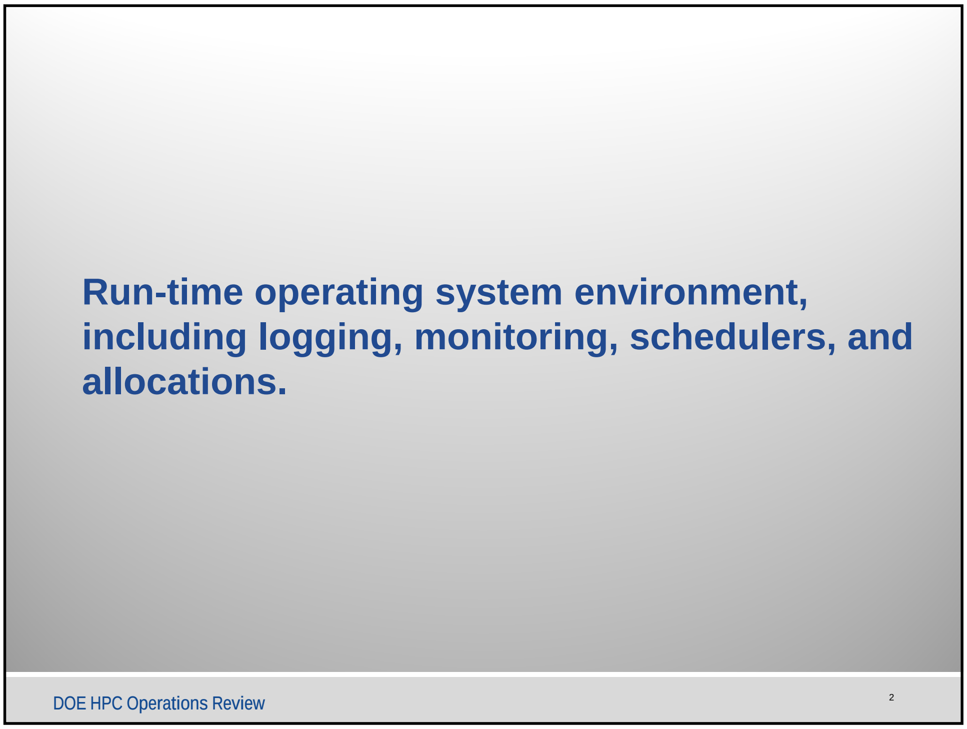
Run-time operating system environment

DOE HPC Operations Review

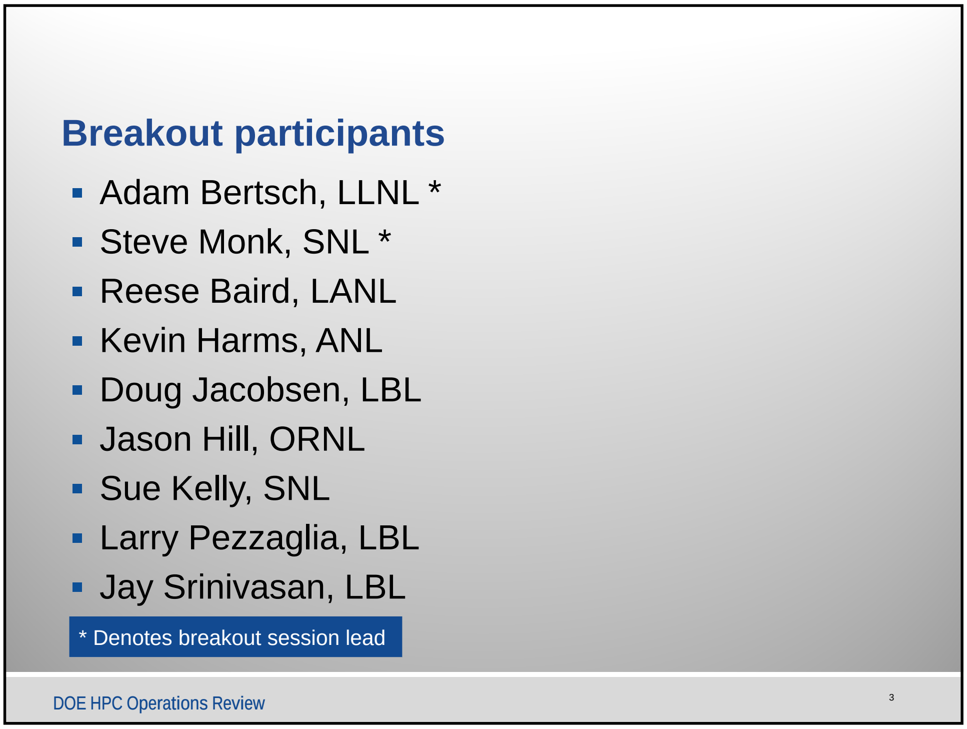
San Francisco, November 5-6, 2013

IBM
ASC
NNSA
TITAN
U.S. DEPARTMENT OF ENERGY
Office of Science

LLNL-PRES-646612



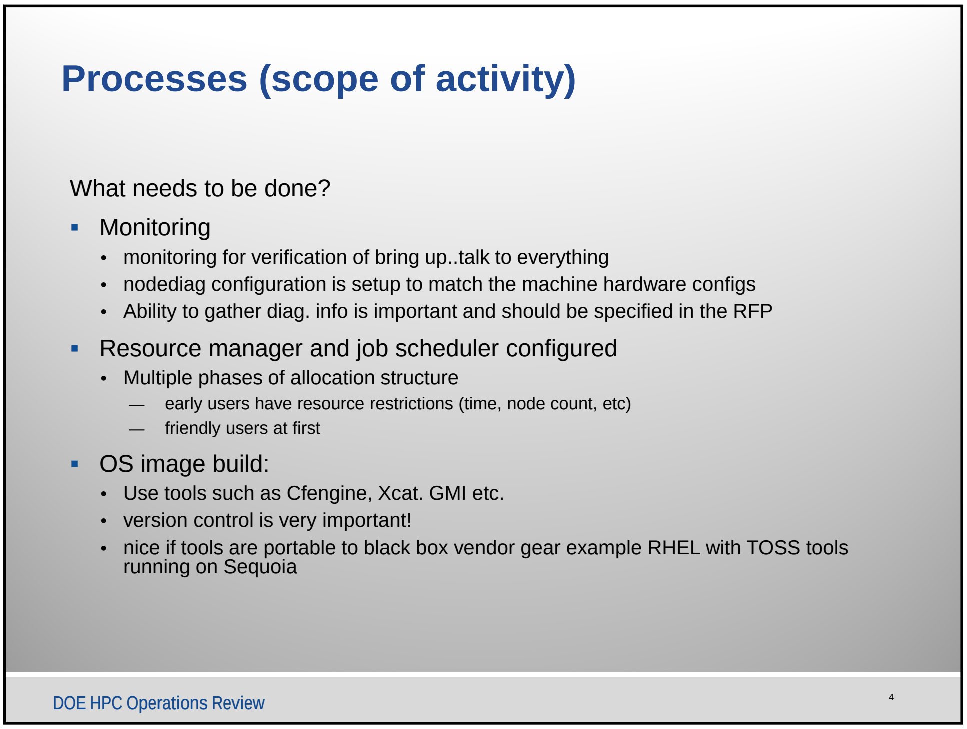
Run-time operating system environment, including logging, monitoring, schedulers, and allocations.



Breakout participants

- Adam Bertsch, LLNL *
- Steve Monk, SNL *
- Reese Baird, LANL
- Kevin Harms, ANL
- Doug Jacobsen, LBL
- Jason Hill, ORNL
- Sue Kelly, SNL
- Larry Pezzaglia, LBL
- Jay Srinivasan, LBL

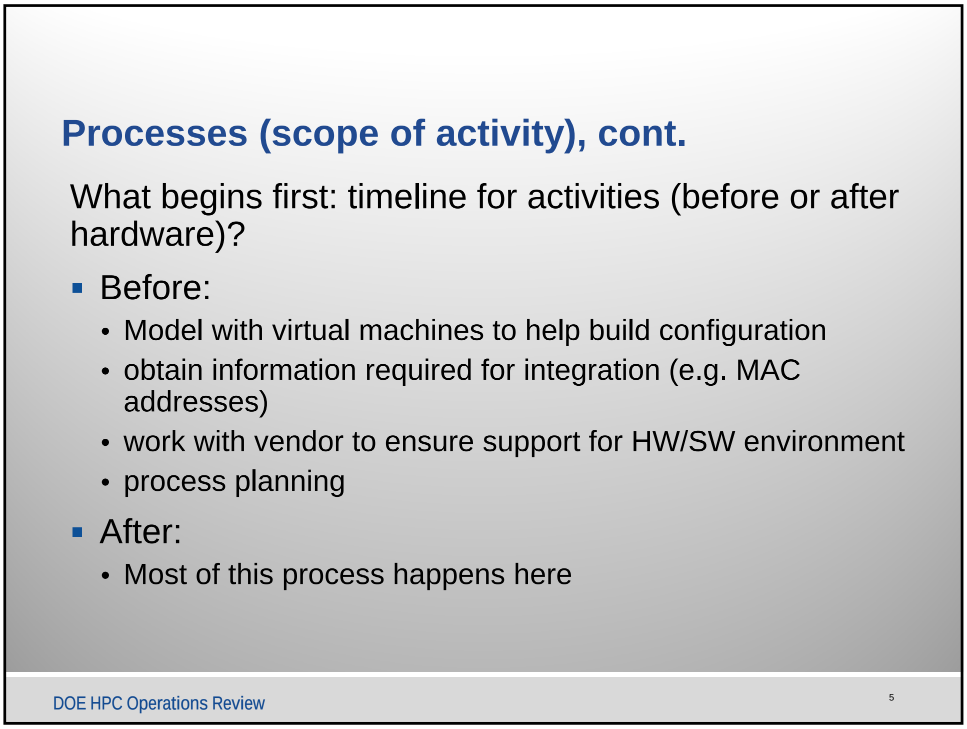
* Denotes breakout session lead



Processes (scope of activity)

What needs to be done?

- Monitoring
  - monitoring for verification of bring up..talk to everything
  - nodediag configuration is setup to match the machine hardware configs
  - Ability to gather diag. info is important and should be specified in the RFP
- Resource manager and job scheduler configured
  - Multiple phases of allocation structure
    - early users have resource restrictions (time, node count, etc)
    - friendly users at first
- OS image build:
  - Use tools such as Cfengine, Xcat. GMI etc.
  - version control is very important!
  - nice if tools are portable to black box vendor gear example RHEL with TOSS tools running on Sequoia

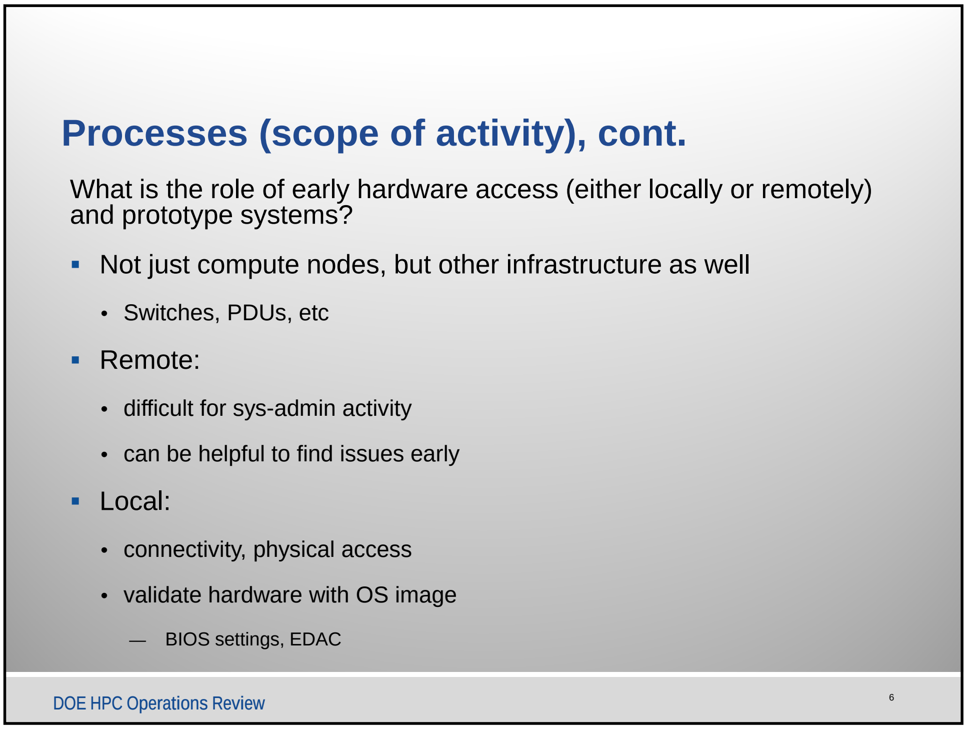


Processes (scope of activity), cont.

What begins first: timeline for activities (before or after hardware)?

- Before:
  - Model with virtual machines to help build configuration
  - obtain information required for integration (e.g. MAC addresses)
  - work with vendor to ensure support for HW/SW environment
  - process planning
- After:
  - Most of this process happens here

DOE HPC Operations Review 5

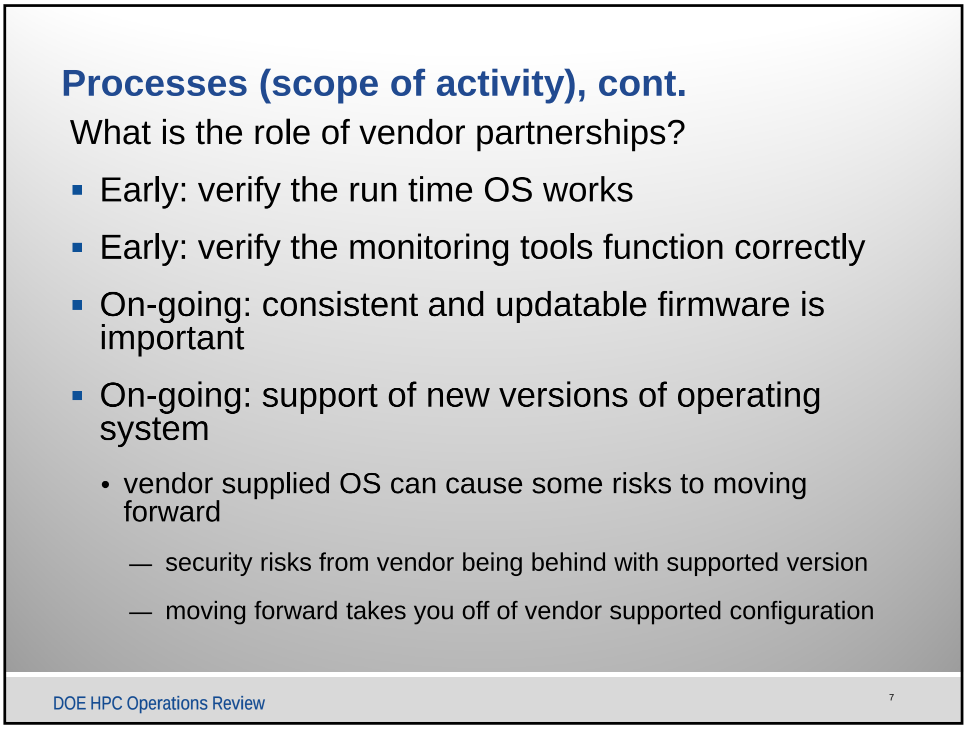


Processes (scope of activity), cont.

What is the role of early hardware access (either locally or remotely) and prototype systems?

- Not just compute nodes, but other infrastructure as well
  - Switches, PDUs, etc
- Remote:
  - difficult for sys-admin activity
  - can be helpful to find issues early
- Local:
  - connectivity, physical access
  - validate hardware with OS image
    - — BIOS settings, EDAC

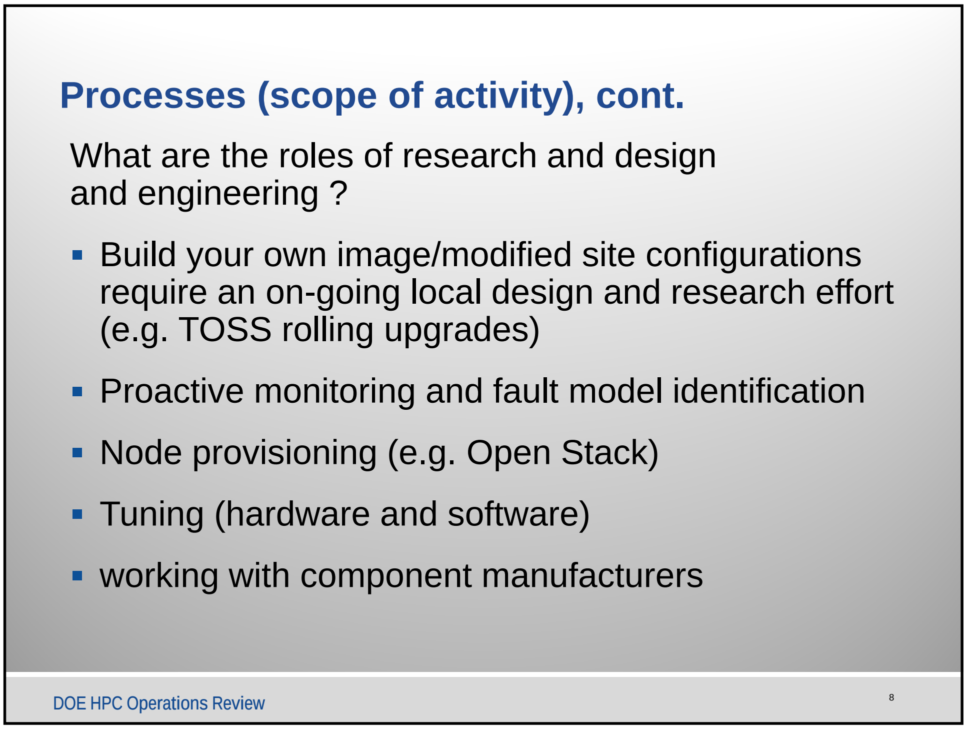
DOE HPC Operations Review 6



Processes (scope of activity), cont.

What is the role of vendor partnerships?

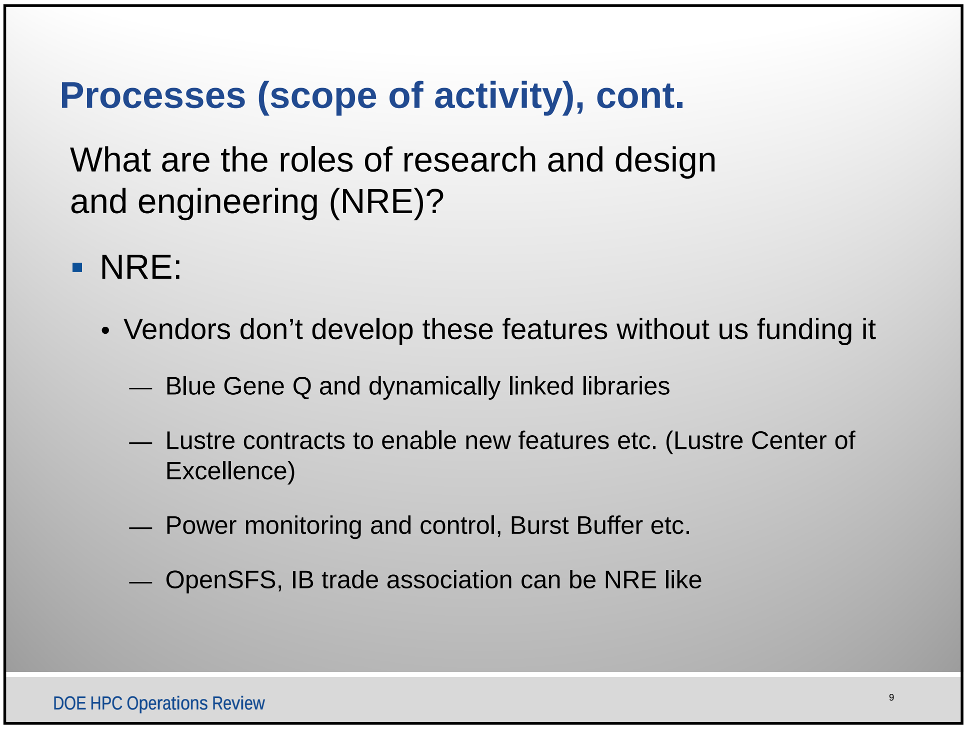
- Early: verify the run time OS works
- Early: verify the monitoring tools function correctly
- On-going: consistent and updatable firmware is important
- On-going: support of new versions of operating system
  - vendor supplied OS can cause some risks to moving forward
    - security risks from vendor being behind with supported version
    - moving forward takes you off of vendor supported configuration



Processes (scope of activity), cont.

What are the roles of research and design
and engineering ?

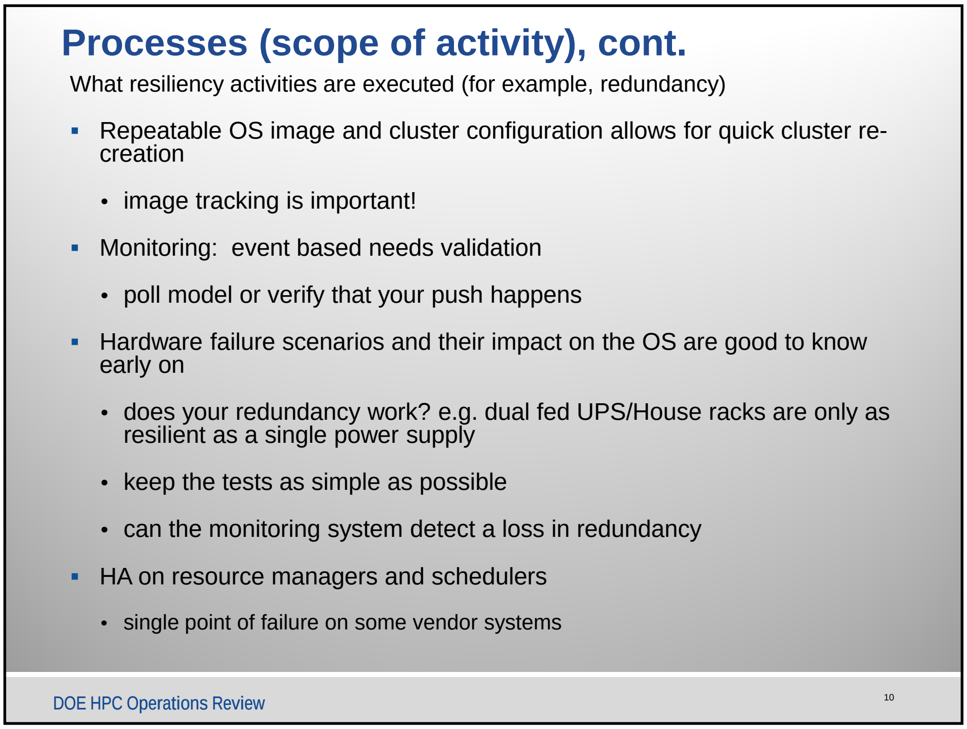
- Build your own image/modified site configurations require an on-going local design and research effort (e.g. TOSS rolling upgrades)
- Proactive monitoring and fault model identification
- Node provisioning (e.g. Open Stack)
- Tuning (hardware and software)
- working with component manufacturers



Processes (scope of activity), cont.

What are the roles of research and design and engineering (NRE)?

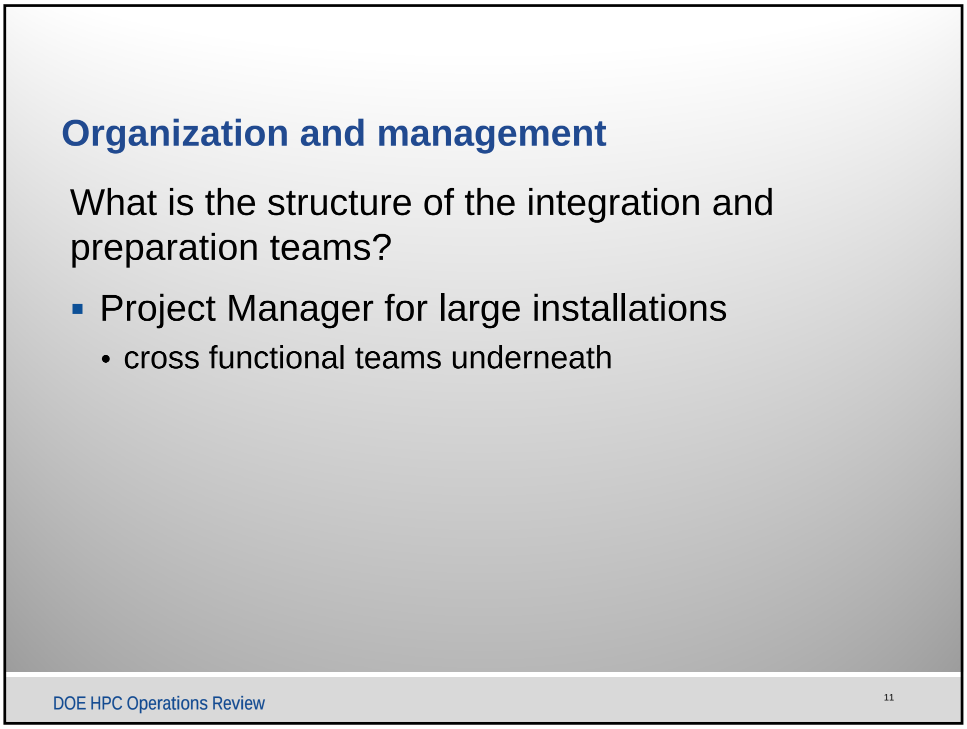
- NRE:
  - Vendors don’t develop these features without us funding it
    - Blue Gene Q and dynamically linked libraries
    - Lustre contracts to enable new features etc. (Lustre Center of Excellence)
    - Power monitoring and control, Burst Buffer etc.
    - OpenSFS, IB trade association can be NRE like



Processes (scope of activity), cont.

What resiliency activities are executed (for example, redundancy)

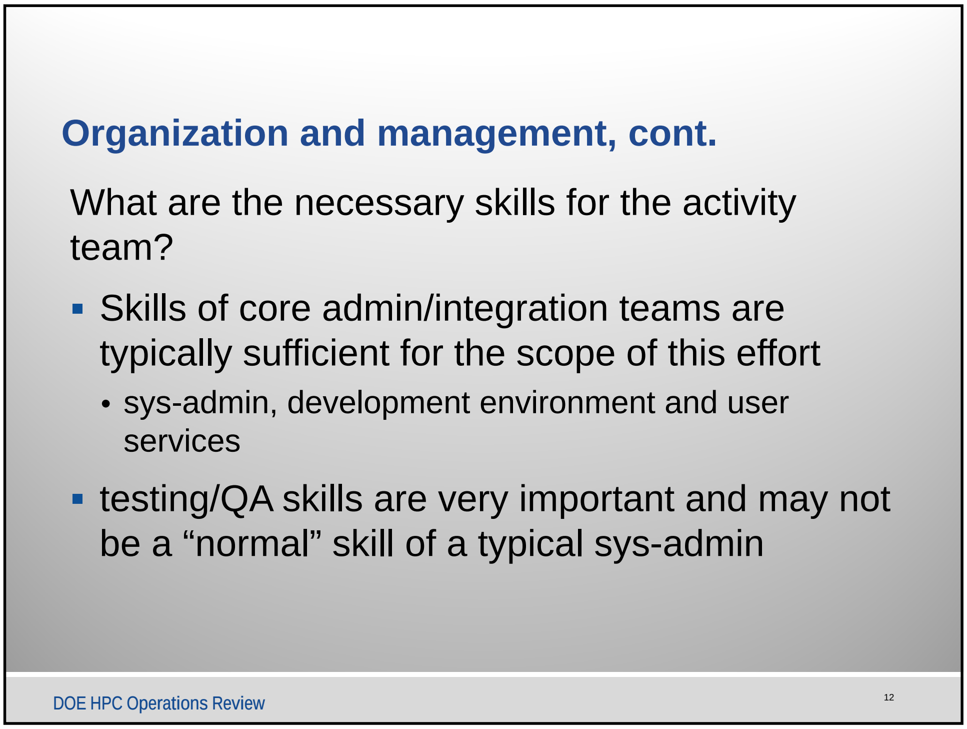
- Repeatable OS image and cluster configuration allows for quick cluster re-creation
  - image tracking is important!
- Monitoring: event based needs validation
  - poll model or verify that your push happens
- Hardware failure scenarios and their impact on the OS are good to know early on
  - does your redundancy work? e.g. dual fed UPS/House racks are only as resilient as a single power supply
  - keep the tests as simple as possible
  - can the monitoring system detect a loss in redundancy
- HA on resource managers and schedulers
  - single point of failure on some vendor systems



Organization and management

What is the structure of the integration and preparation teams?

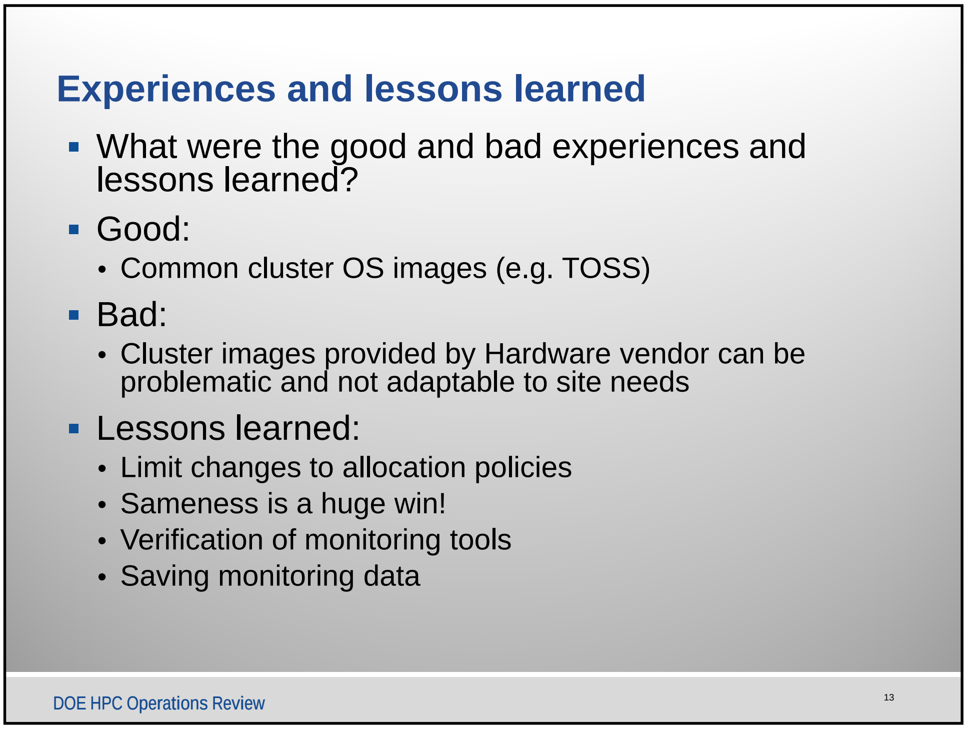
- Project Manager for large installations
  - cross functional teams underneath



Organization and management, cont.

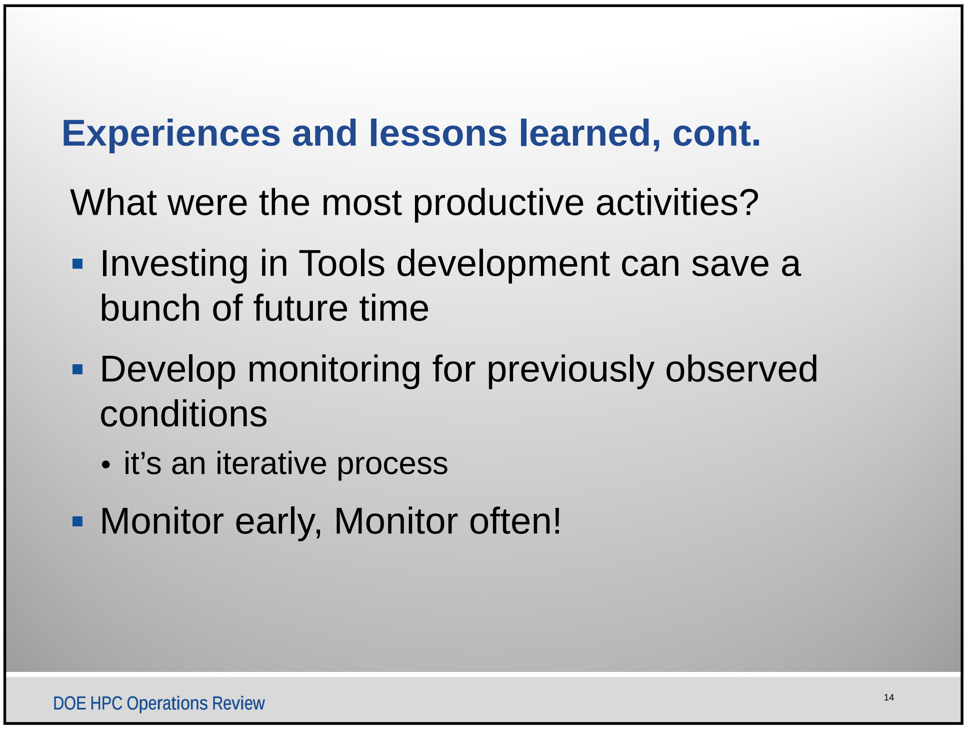
What are the necessary skills for the activity team?

- Skills of core admin/integration teams are typically sufficient for the scope of this effort
  - sys-admin, development environment and user services
- testing/QA skills are very important and may not be a “normal” skill of a typical sys-admin



Experiences and lessons learned

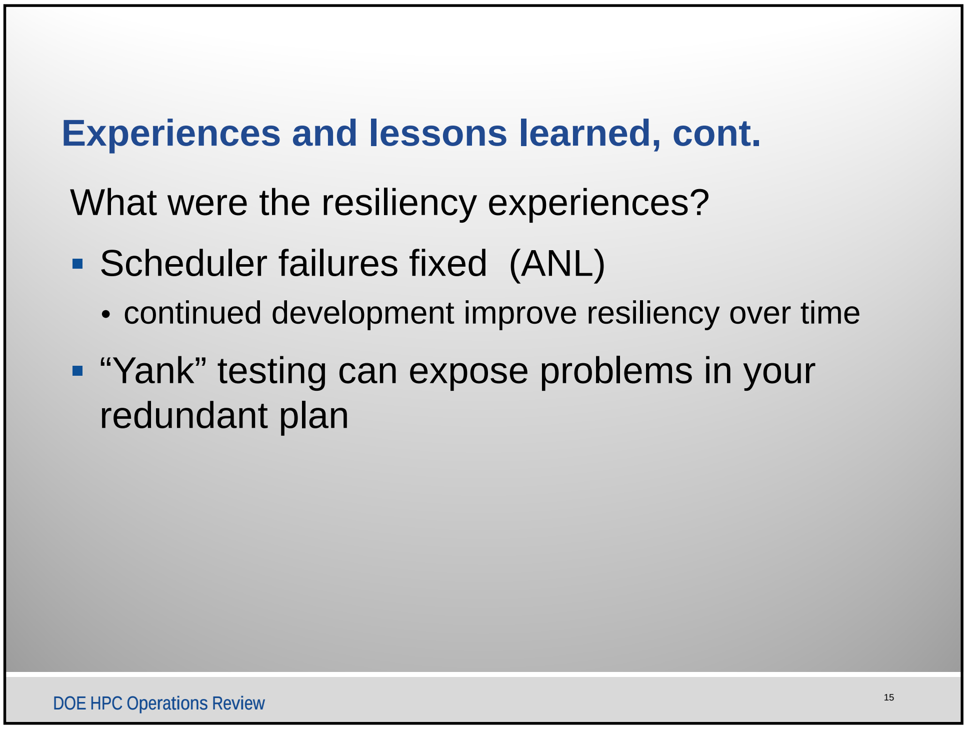
- What were the good and bad experiences and lessons learned?
- Good:
  - Common cluster OS images (e.g. TOSS)
- Bad:
  - Cluster images provided by Hardware vendor can be problematic and not adaptable to site needs
- Lessons learned:
  - Limit changes to allocation policies
  - Sameness is a huge win!
  - Verification of monitoring tools
  - Saving monitoring data



Experiences and lessons learned, cont.

What were the most productive activities?

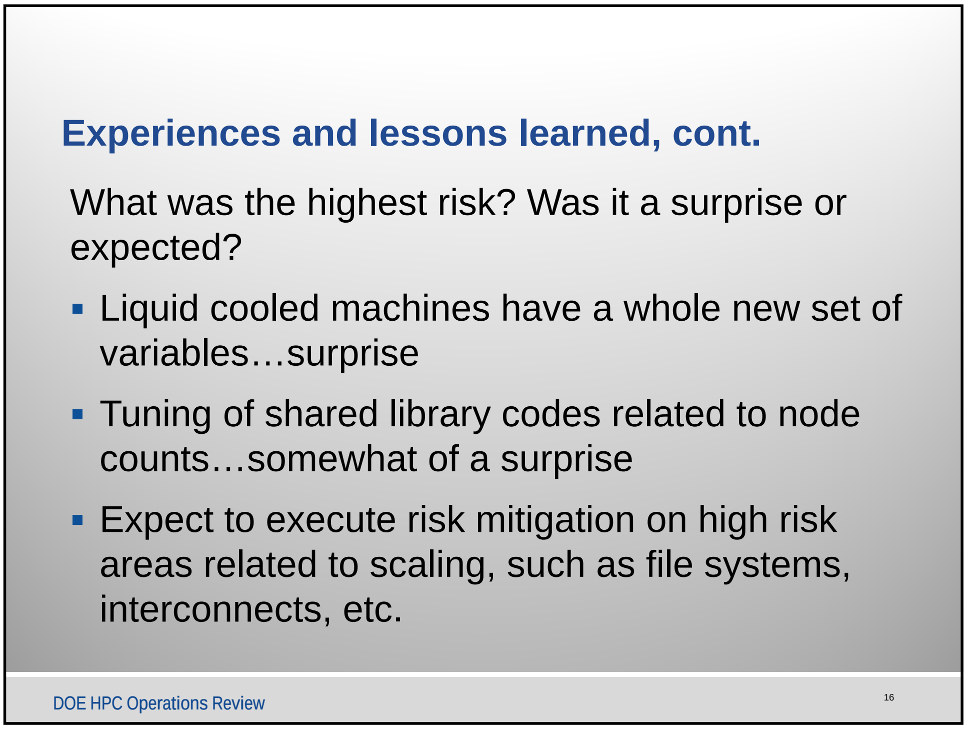
- Investing in Tools development can save a bunch of future time
- Develop monitoring for previously observed conditions
  - it’s an iterative process
- Monitor early, Monitor often!



Experiences and lessons learned, cont.

What were the resiliency experiences?

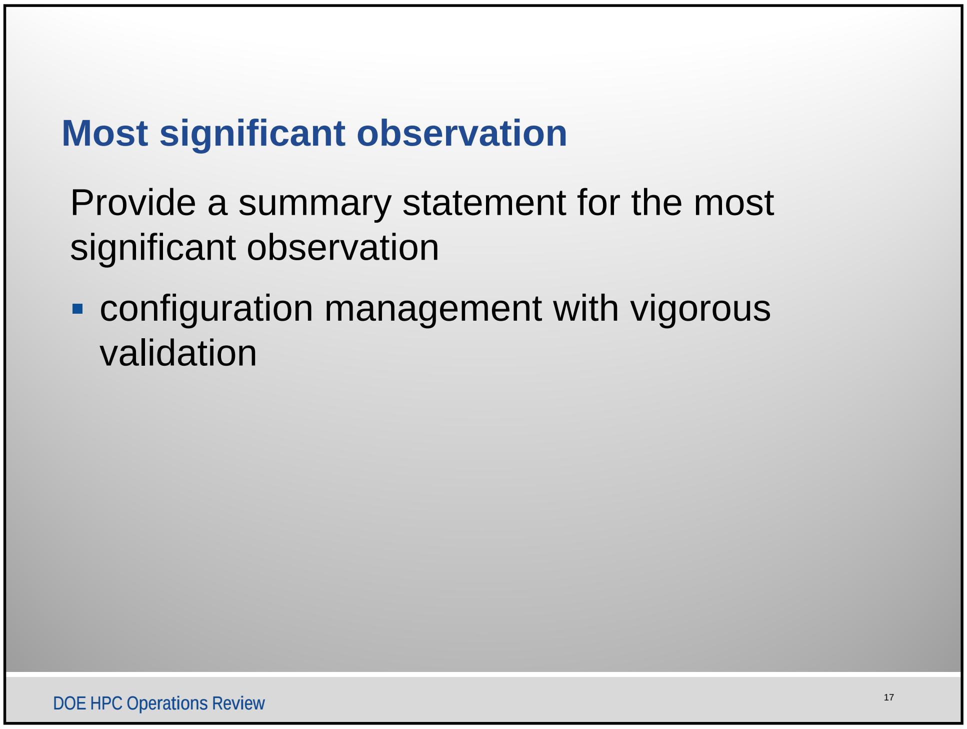
- Scheduler failures fixed (ANL)
  - continued development improve resiliency over time
- “Yank” testing can expose problems in your redundant plan



Experiences and lessons learned, cont.

What was the highest risk? Was it a surprise or expected?

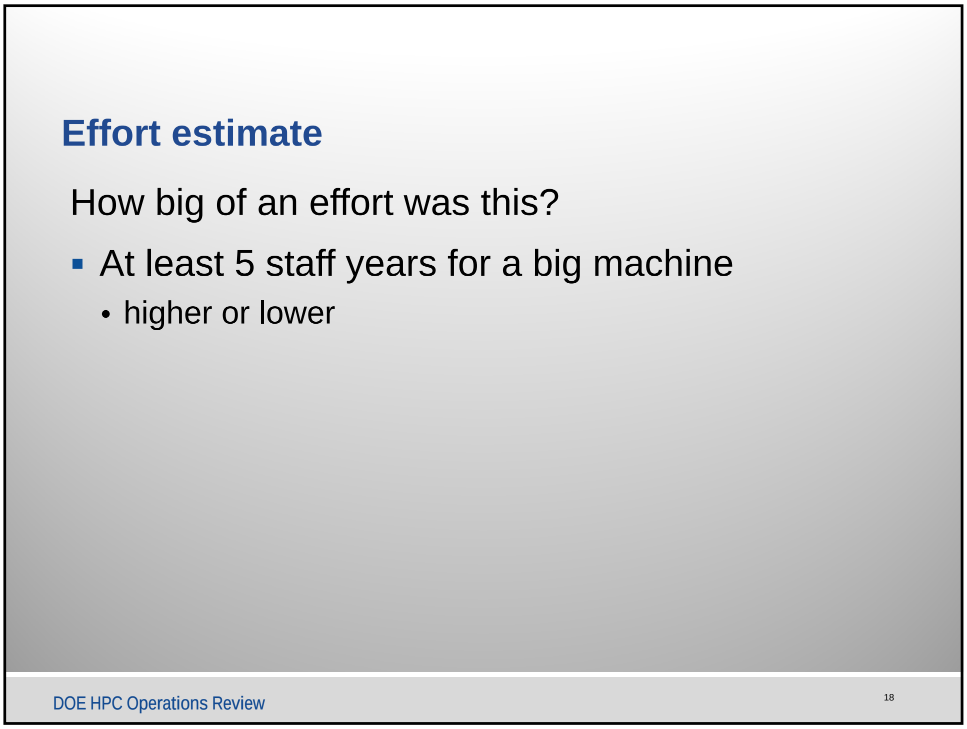
- Liquid cooled machines have a whole new set of variables…surprise
- Tuning of shared library codes related to node counts…somewhat of a surprise
- Expect to execute risk mitigation on high risk areas related to scaling, such as file systems, interconnects, etc.



Most significant observation

Provide a summary statement for the most significant observation

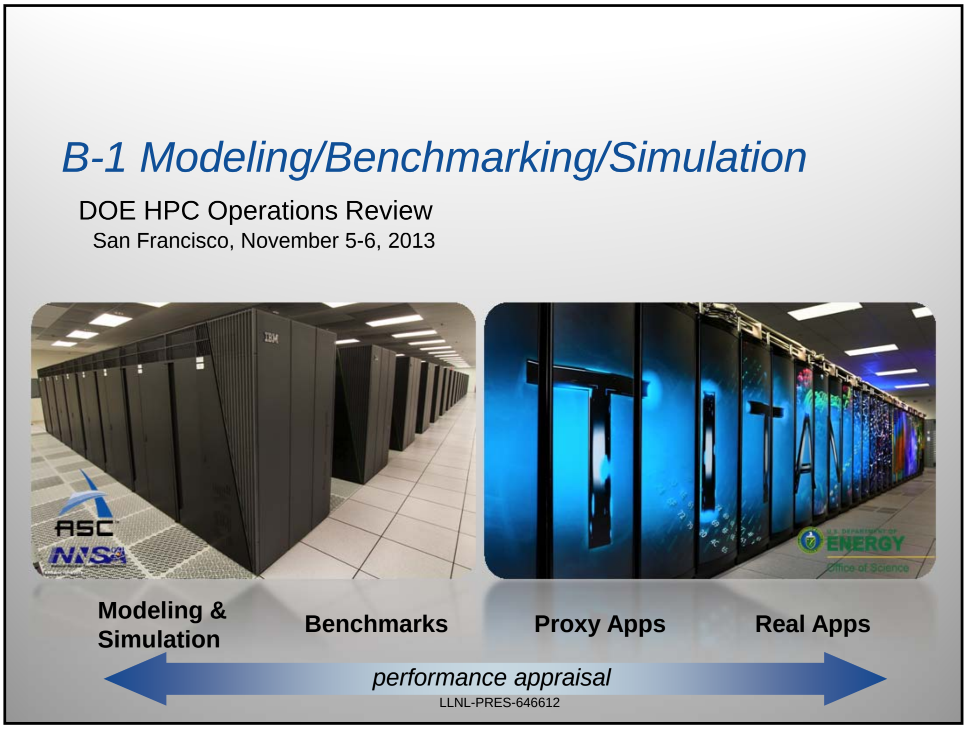
- configuration management with vigorous validation



Effort estimate

How big of an effort was this?

- At least 5 staff years for a big machine
  - higher or lower



B-1 Modeling/Benchmarking/Simulation

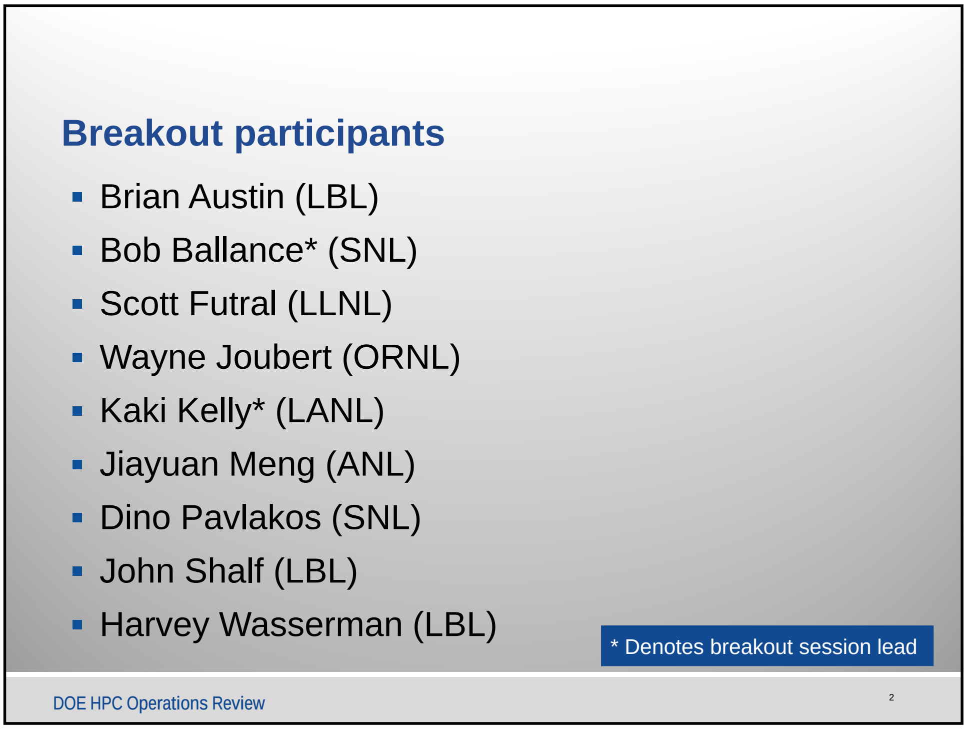
DOE HPC Operations Review

San Francisco, November 5-6, 2013

Modeling & Simulation **Benchmarks** **Proxy Apps** **Real Apps**

performance appraisal

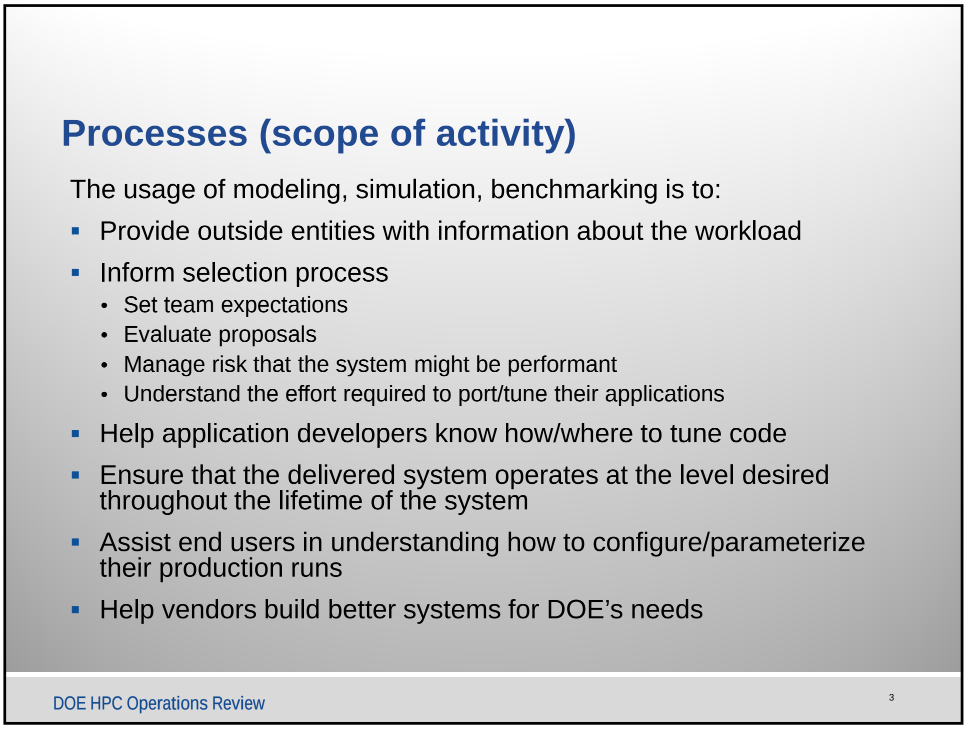
LLNL-PRES-646612



Breakout participants

- Brian Austin (LBL)
- Bob Ballance* (SNL)
- Scott Futral (LLNL)
- Wayne Joubert (ORNL)
- Kaki Kelly* (LANL)
- Jiayuan Meng (ANL)
- Dino Pavlakos (SNL)
- John Shalf (LBL)
- Harvey Wasserman (LBL)

* Denotes breakout session lead

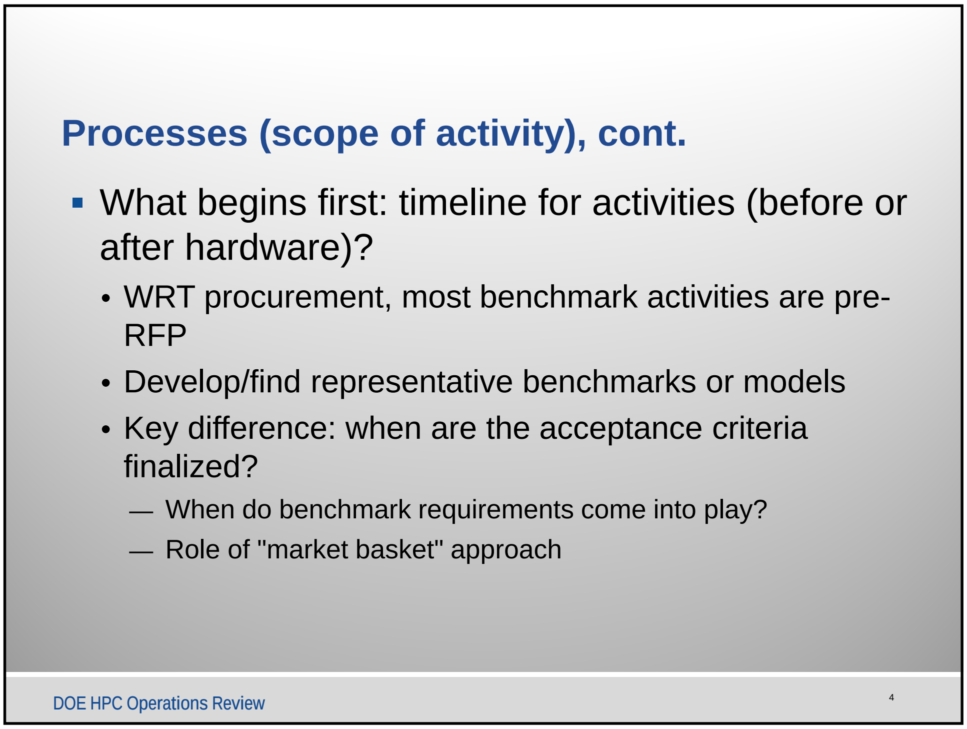


Processes (scope of activity)

The usage of modeling, simulation, benchmarking is to:

- Provide outside entities with information about the workload
- Inform selection process
  - Set team expectations
  - Evaluate proposals
  - Manage risk that the system might be performant
  - Understand the effort required to port/tune their applications
- Help application developers know how/where to tune code
- Ensure that the delivered system operates at the level desired throughout the lifetime of the system
- Assist end users in understanding how to configure/parameterize their production runs
- Help vendors build better systems for DOE’s needs

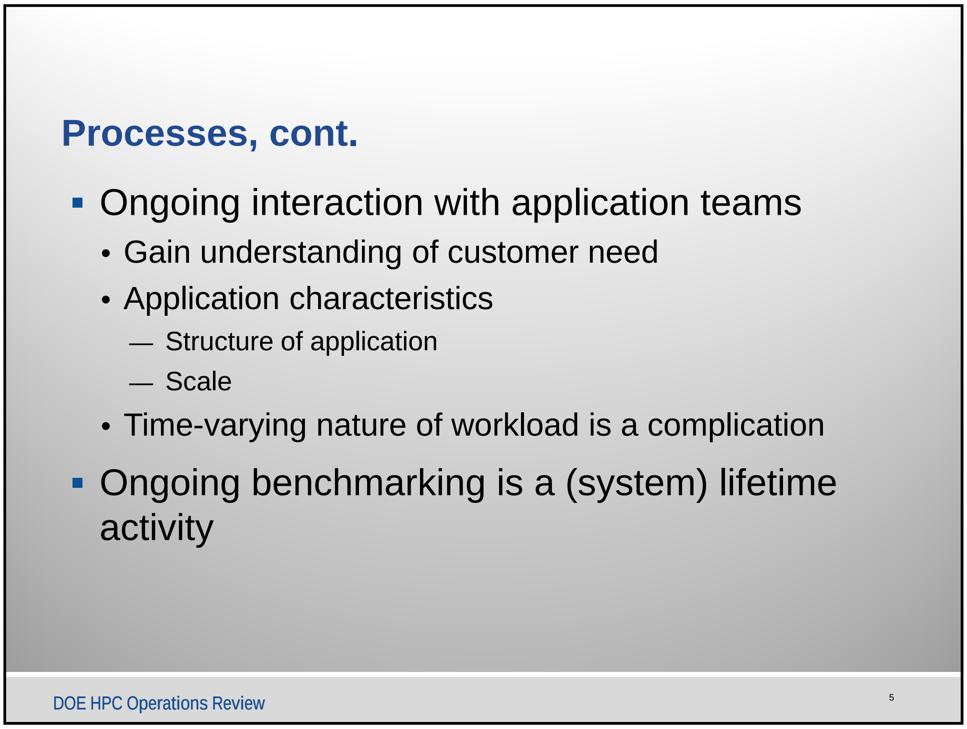
DOE HPC Operations Review 3



Processes (scope of activity), cont.

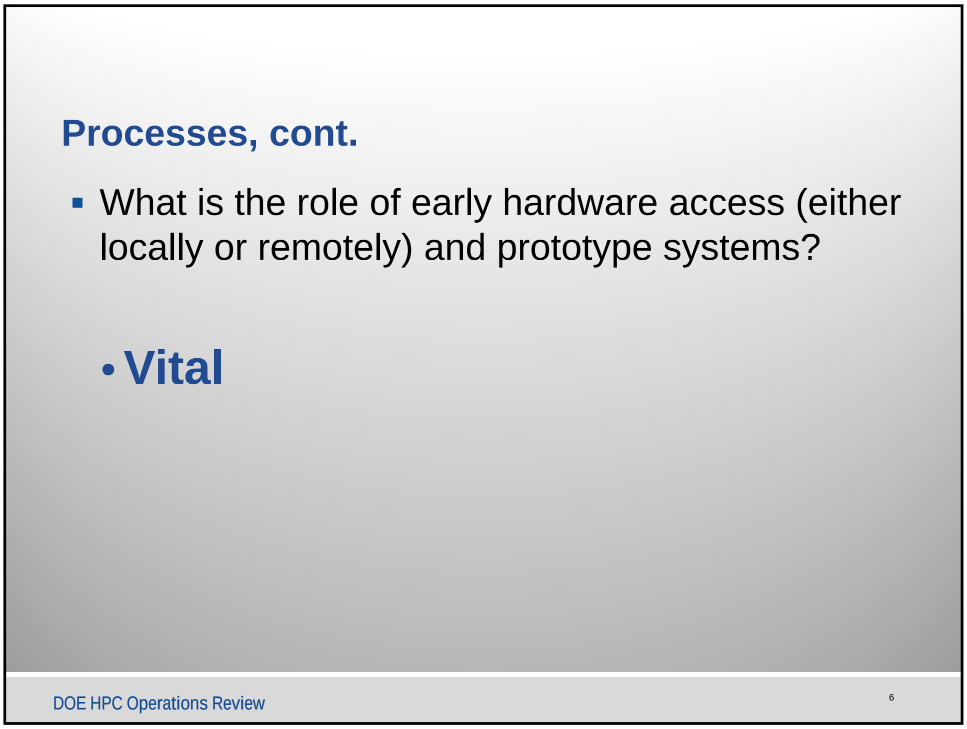
- What begins first: timeline for activities (before or after hardware)?
  - WRT procurement, most benchmark activities are pre-RFP
  - Develop/find representative benchmarks or models
  - Key difference: when are the acceptance criteria finalized?
    - When do benchmark requirements come into play?
    - Role of "market basket" approach

DOE HPC Operations Review 4



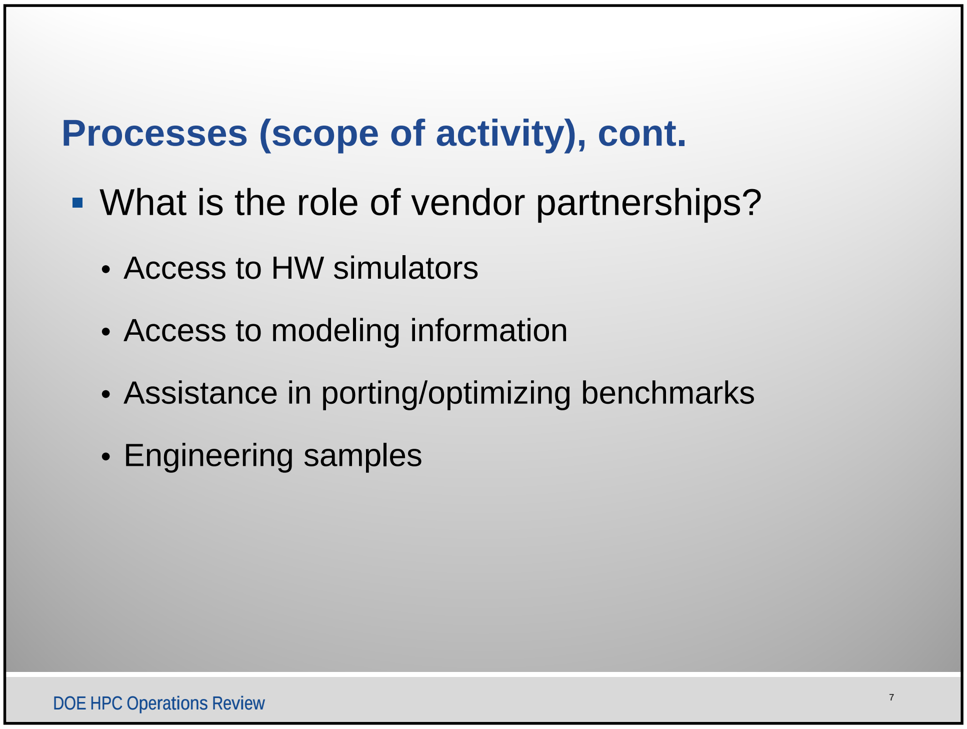
Processes, cont.

- Ongoing interaction with application teams
  - Gain understanding of customer need
  - Application characteristics
    - — Structure of application
    - — Scale
  - Time-varying nature of workload is a complication
- Ongoing benchmarking is a (system) lifetime activity



Processes, cont.

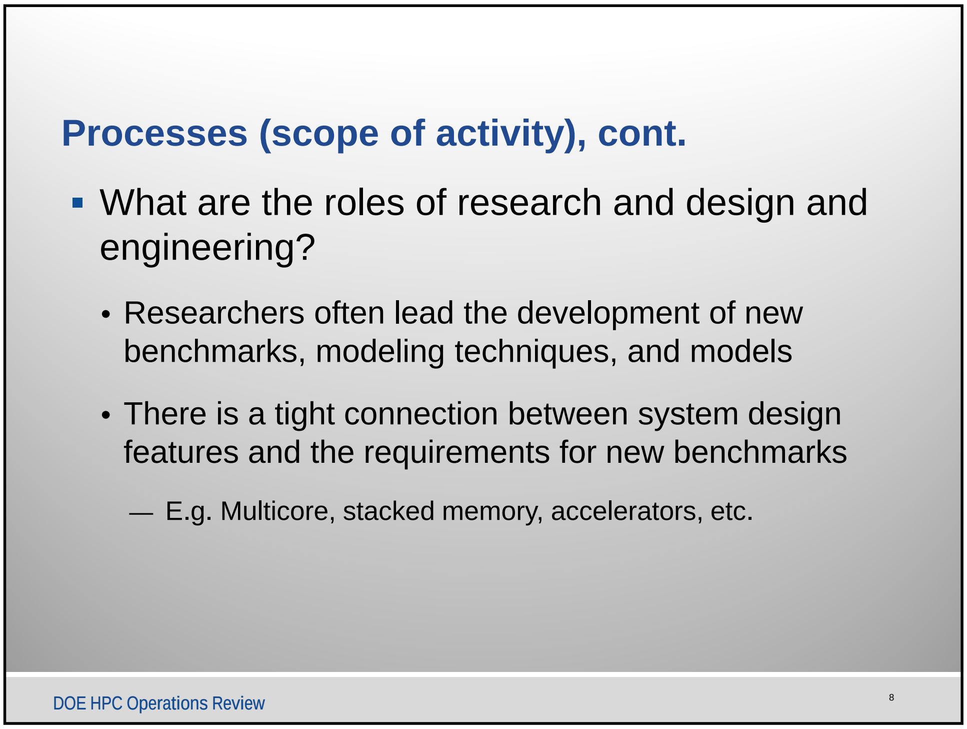
- What is the role of early hardware access (either locally or remotely) and prototype systems?
  - **Vital**



Processes (scope of activity), cont.

- What is the role of vendor partnerships?
  - Access to HW simulators
  - Access to modeling information
  - Assistance in porting/optimizing benchmarks
  - Engineering samples

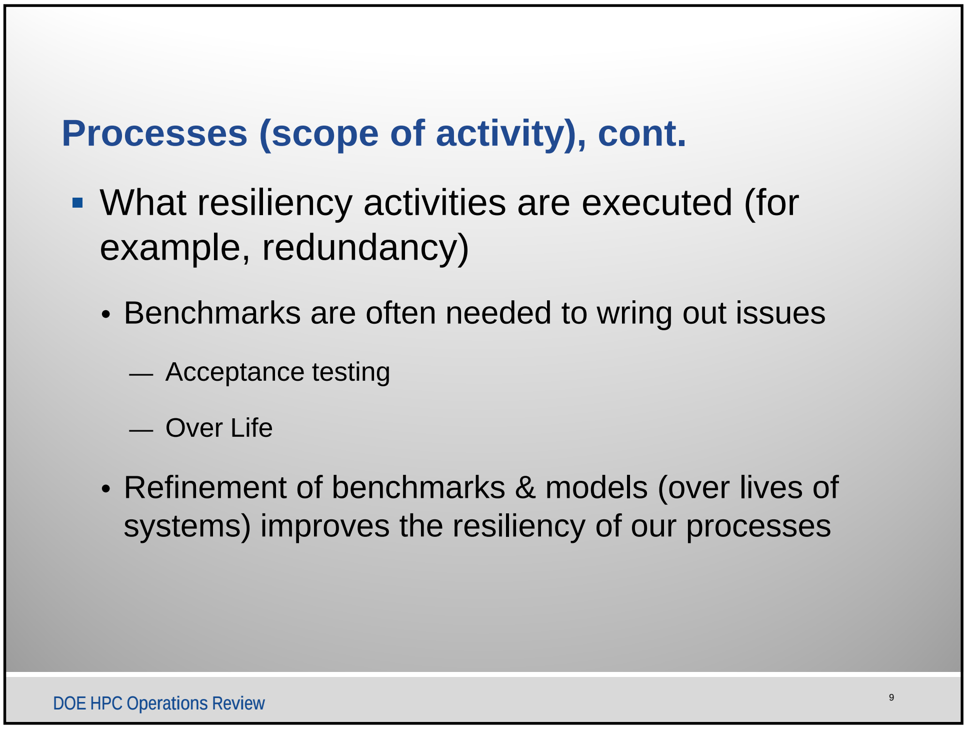
DOE HPC Operations Review 7



Processes (scope of activity), cont.

- What are the roles of research and design and engineering?
  - Researchers often lead the development of new benchmarks, modeling techniques, and models
  - There is a tight connection between system design features and the requirements for new benchmarks
    - — E.g. Multicore, stacked memory, accelerators, etc.

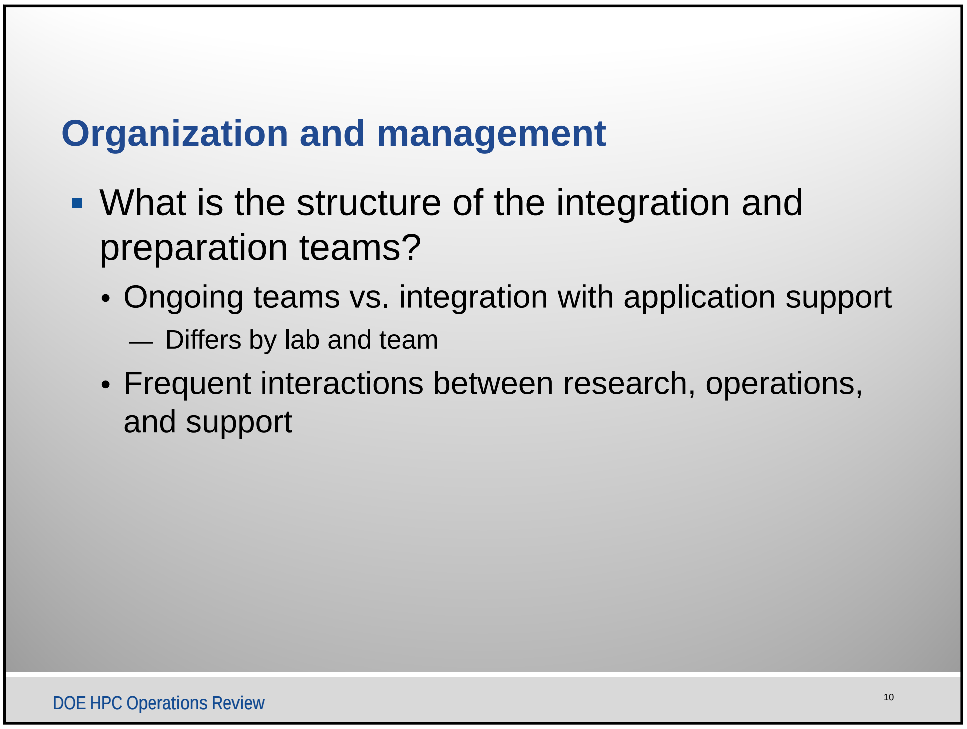
DOE HPC Operations Review 8



Processes (scope of activity), cont.

- What resiliency activities are executed (for example, redundancy)
  - Benchmarks are often needed to wring out issues
    - — Acceptance testing
    - — Over Life
  - Refinement of benchmarks & models (over lives of systems) improves the resiliency of our processes

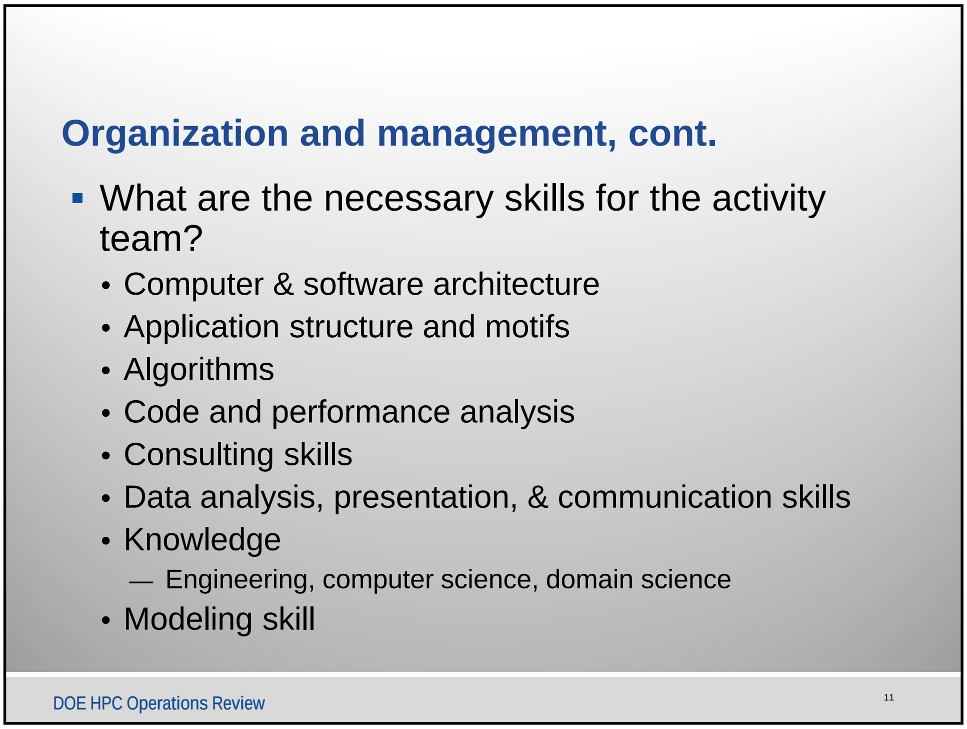
DOE HPC Operations Review 9



Organization and management

- What is the structure of the integration and preparation teams?
  - Ongoing teams vs. integration with application support
    - — Differs by lab and team
  - Frequent interactions between research, operations, and support

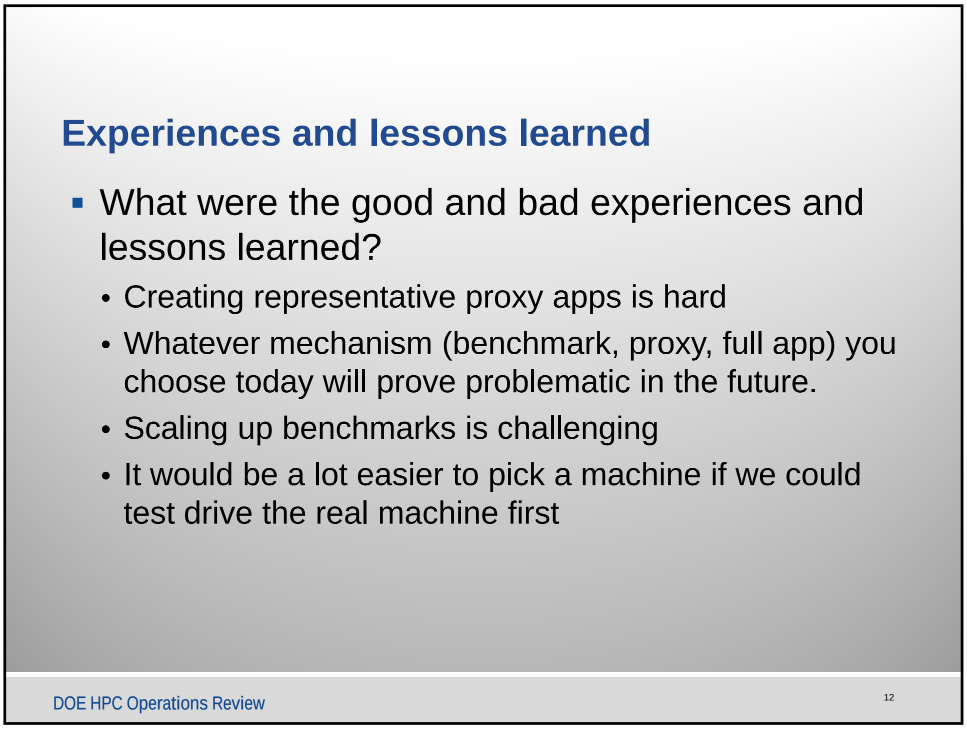
DOE HPC Operations Review 10



Organization and management, cont.

- What are the necessary skills for the activity team?
  - Computer & software architecture
  - Application structure and motifs
  - Algorithms
  - Code and performance analysis
  - Consulting skills
  - Data analysis, presentation, & communication skills
  - Knowledge
    - Engineering, computer science, domain science
  - Modeling skill

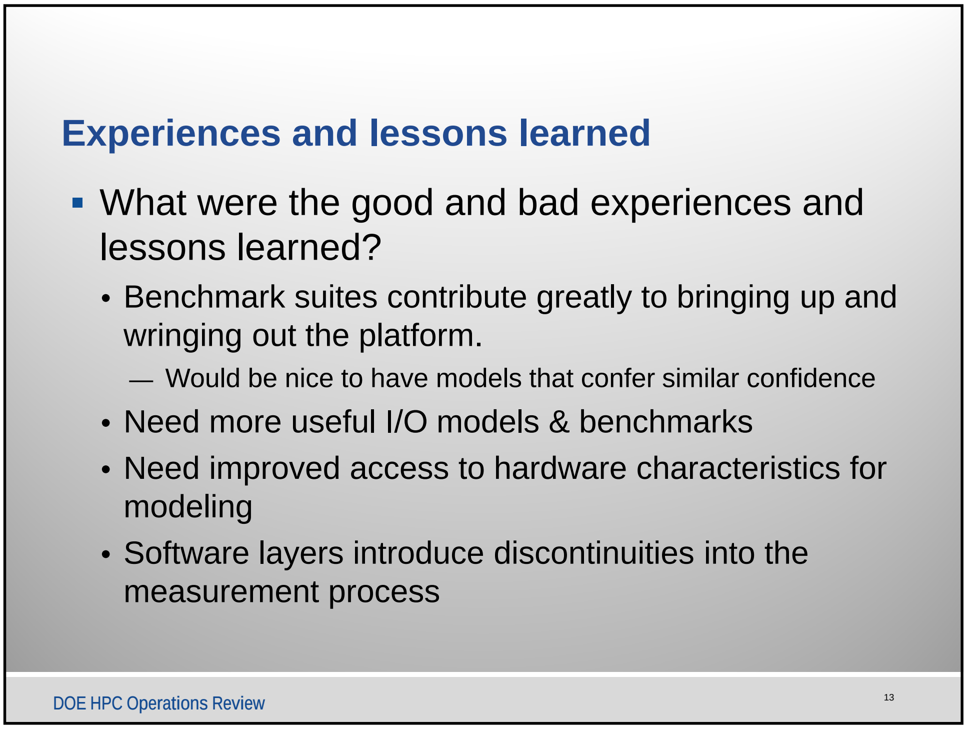
DOE HPC Operations Review 11



Experiences and lessons learned

- What were the good and bad experiences and lessons learned?
  - Creating representative proxy apps is hard
  - Whatever mechanism (benchmark, proxy, full app) you choose today will prove problematic in the future.
  - Scaling up benchmarks is challenging
  - It would be a lot easier to pick a machine if we could test drive the real machine first

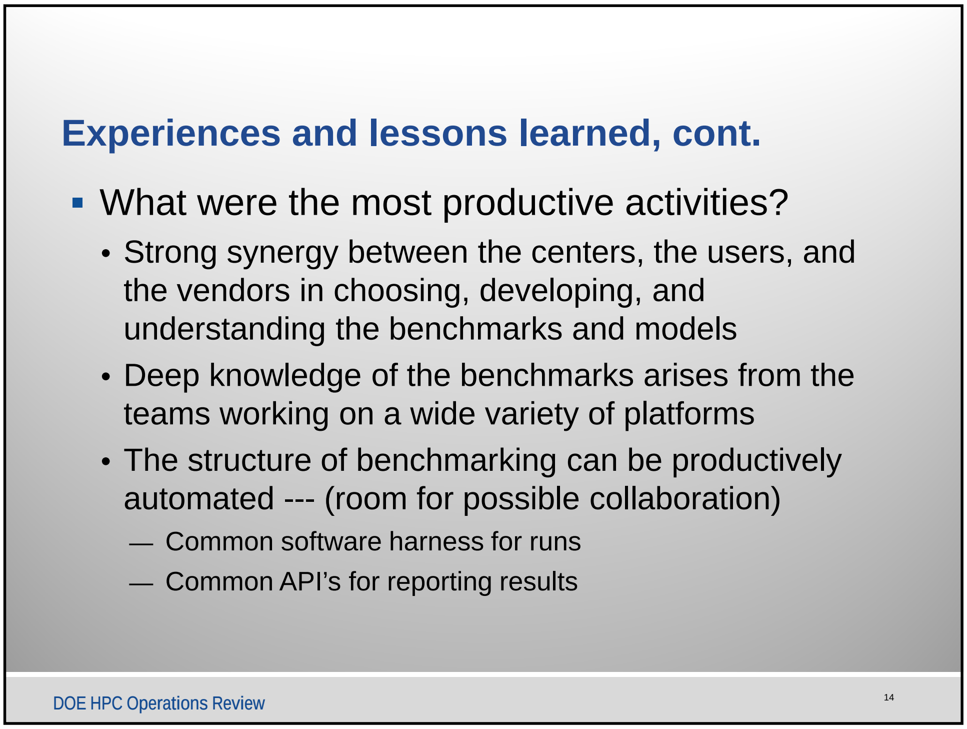
DOE HPC Operations Review 12



Experiences and lessons learned

- What were the good and bad experiences and lessons learned?
  - Benchmark suites contribute greatly to bringing up and wringing out the platform.
    - — Would be nice to have models that confer similar confidence
  - Need more useful I/O models & benchmarks
  - Need improved access to hardware characteristics for modeling
  - Software layers introduce discontinuities into the measurement process

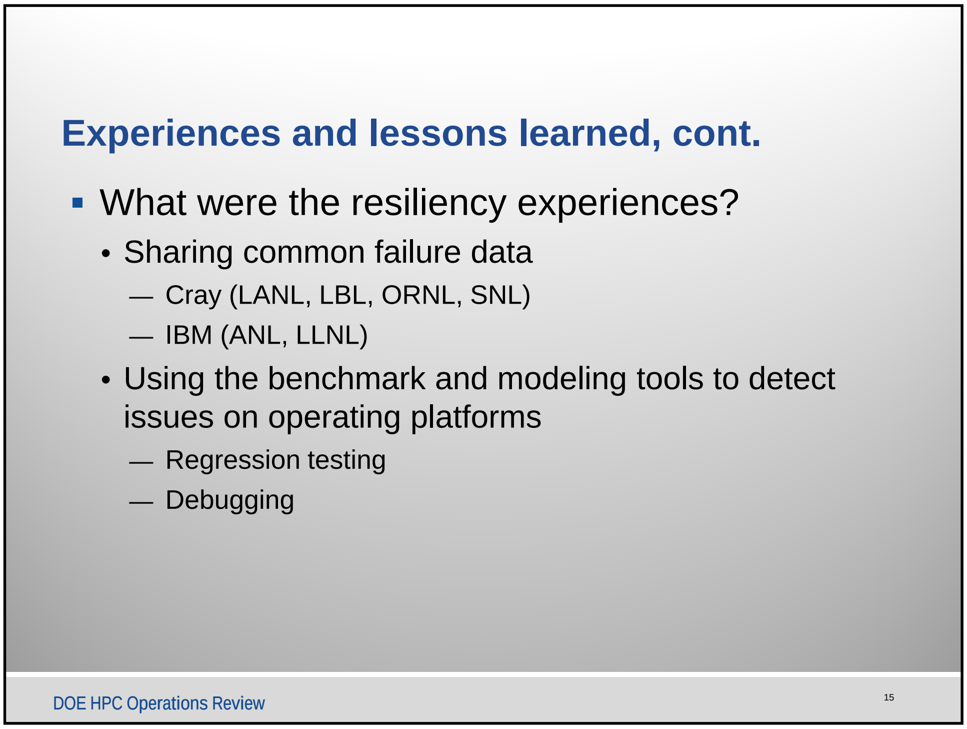
DOE HPC Operations Review 13



Experiences and lessons learned, cont.

- What were the most productive activities?
  - Strong synergy between the centers, the users, and the vendors in choosing, developing, and understanding the benchmarks and models
  - Deep knowledge of the benchmarks arises from the teams working on a wide variety of platforms
  - The structure of benchmarking can be productively automated --- (room for possible collaboration)
    - — Common software harness for runs
    - — Common API's for reporting results

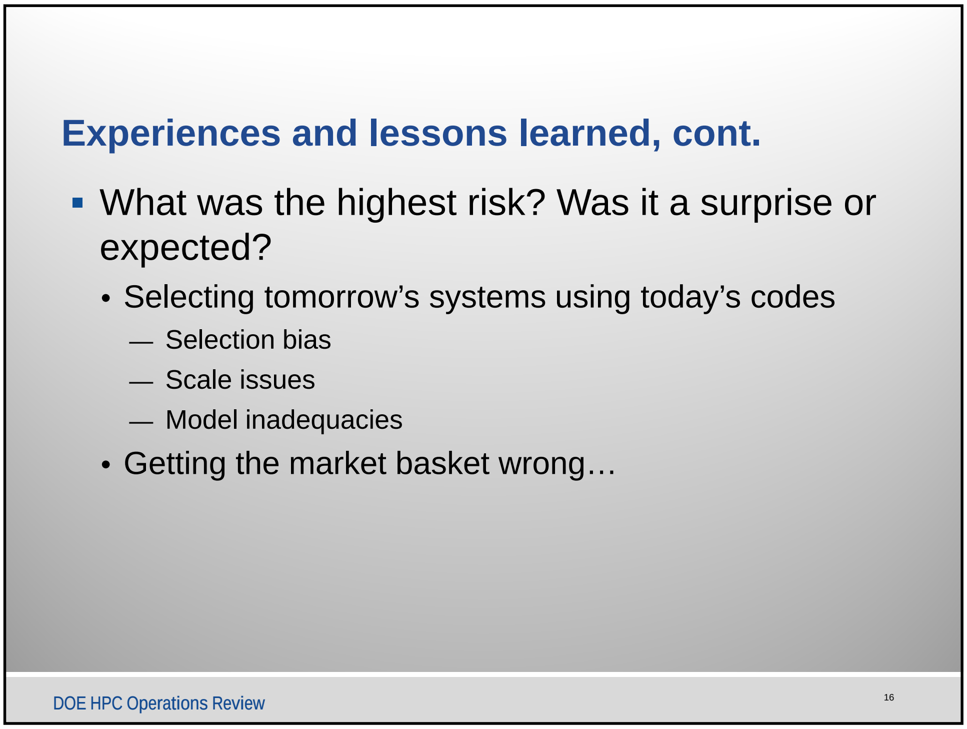
DOE HPC Operations Review 14



Experiences and lessons learned, cont.

- What were the resiliency experiences?
  - Sharing common failure data
    - Cray (LANL, LBL, ORNL, SNL)
    - IBM (ANL, LLNL)
  - Using the benchmark and modeling tools to detect issues on operating platforms
    - Regression testing
    - Debugging

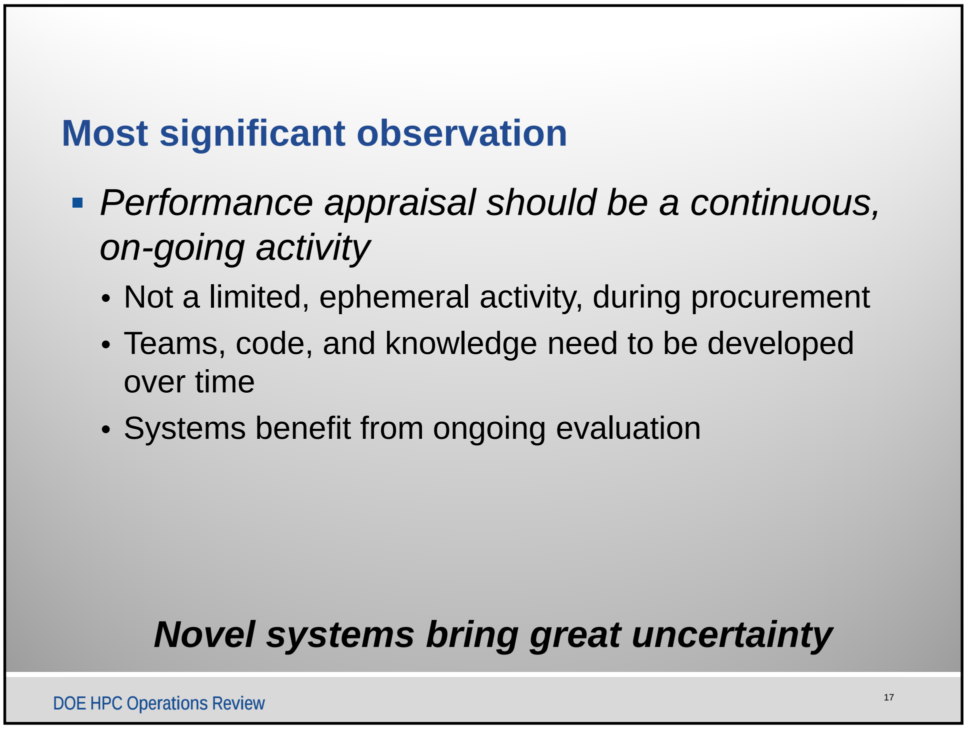
DOE HPC Operations Review 15



Experiences and lessons learned, cont.

- What was the highest risk? Was it a surprise or expected?
  - Selecting tomorrow’s systems using today’s codes
    - Selection bias
    - Scale issues
    - Model inadequacies
  - Getting the market basket wrong…

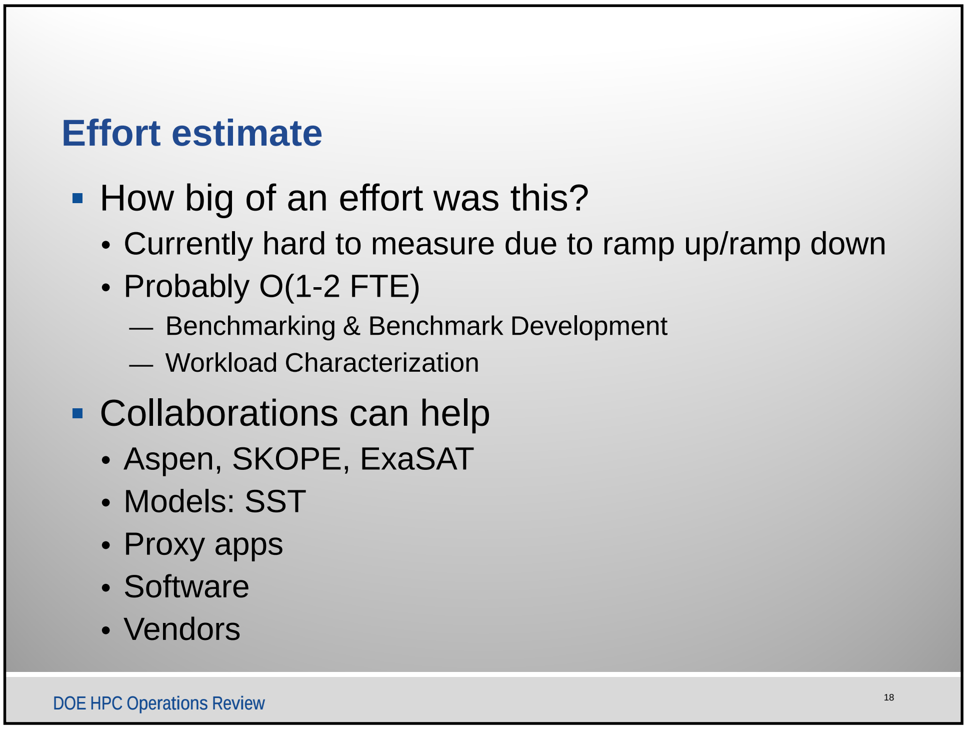
DOE HPC Operations Review 16



Most significant observation

- *Performance appraisal should be a continuous, on-going activity*
  - Not a limited, ephemeral activity, during procurement
  - Teams, code, and knowledge need to be developed over time
  - Systems benefit from ongoing evaluation

Novel systems bring great uncertainty



Effort estimate

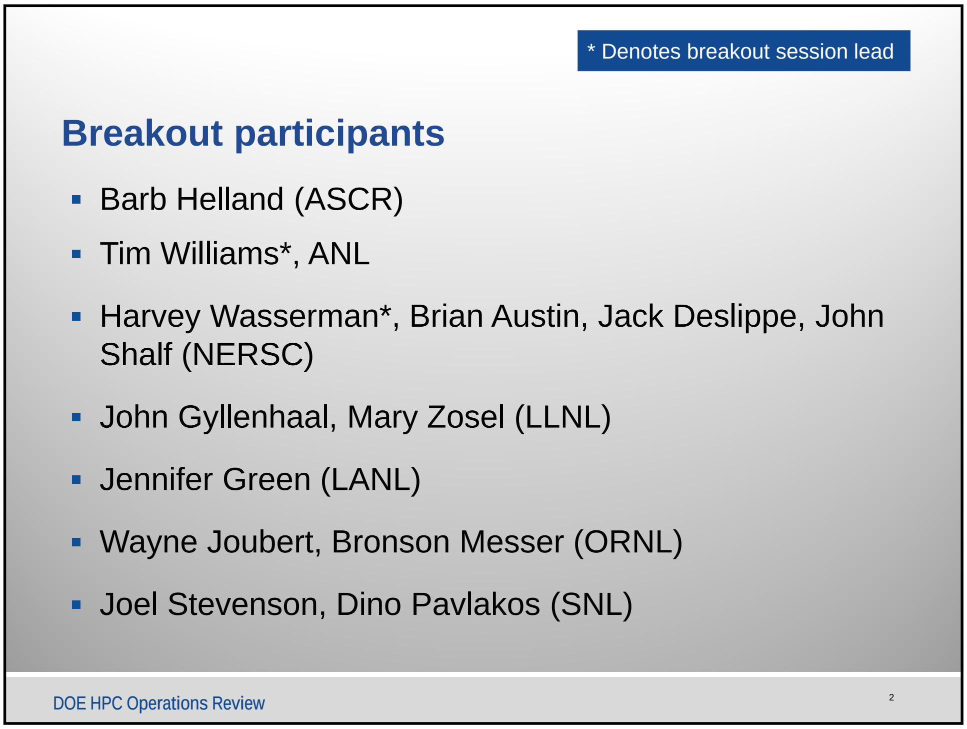
- How big of an effort was this?
  - Currently hard to measure due to ramp up/ramp down
  - Probably O(1-2 FTE)
    - — Benchmarking & Benchmark Development
    - — Workload Characterization
- Collaborations can help
  - Aspen, SKOPE, ExaSAT
  - Models: SST
  - Proxy apps
  - Software
  - Vendors

B/2: Working With Code Teams

(What are the models and how are the applied math and algorithm scaling addressed?)

DOE HPC Operations Review
San Francisco, November 5-6, 2013

LLNL-PRES-646612

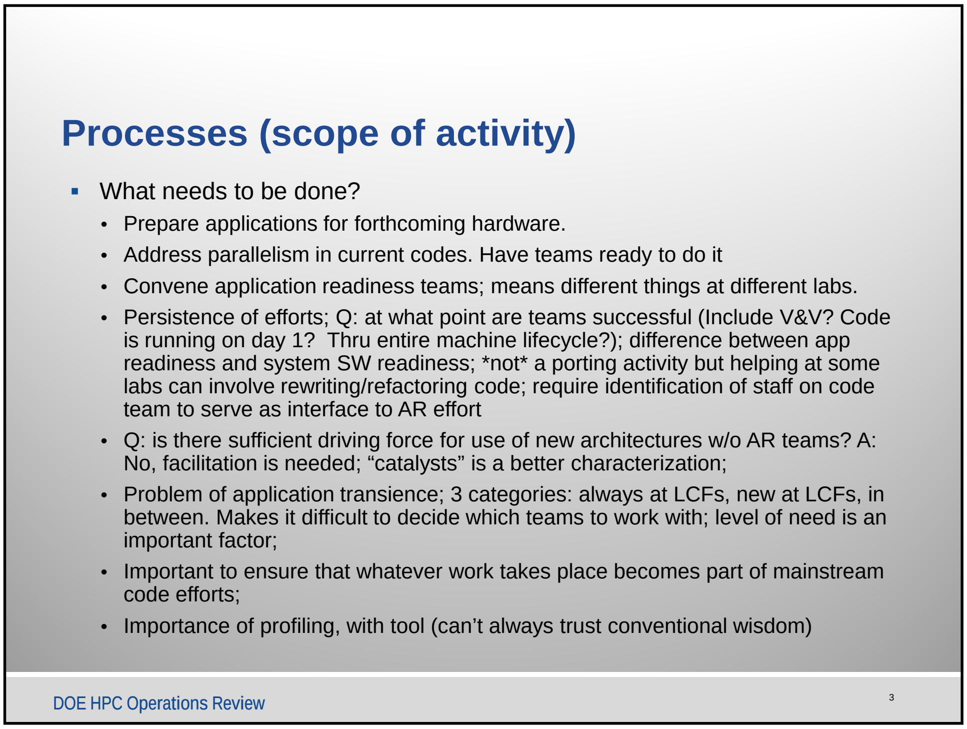


* Denotes breakout session lead

Breakout participants

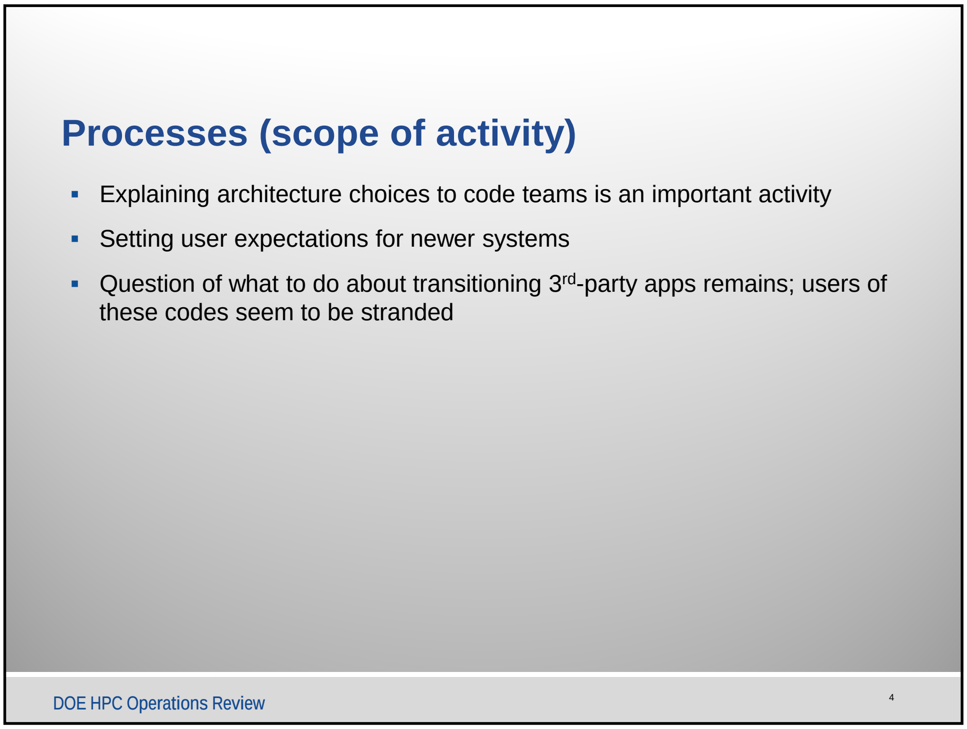
- Barb Helland (ASCR)
- Tim Williams*, ANL
- Harvey Wasserman*, Brian Austin, Jack Deslippe, John Shalf (NERSC)
- John Gyllenhaal, Mary Zosel (LLNL)
- Jennifer Green (LANL)
- Wayne Joubert, Bronson Messer (ORNL)
- Joel Stevenson, Dino Pavlakos (SNL)

DOE HPC Operations Review 2



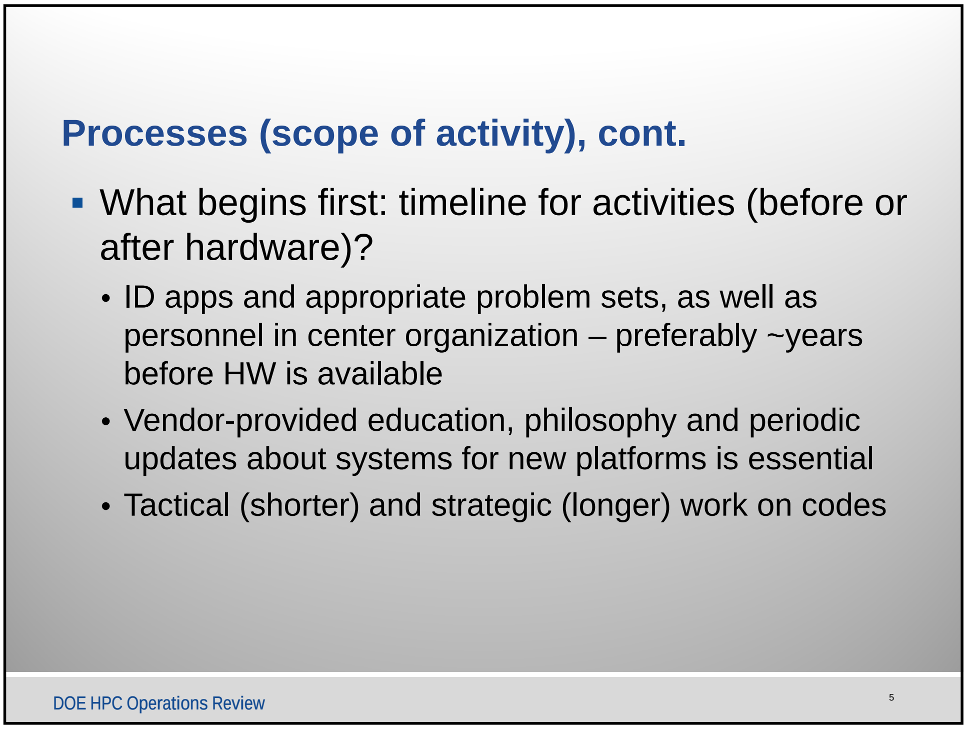
Processes (scope of activity)

- What needs to be done?
  - Prepare applications for forthcoming hardware.
  - Address parallelism in current codes. Have teams ready to do it
  - Convene application readiness teams; means different things at different labs.
  - Persistence of efforts; Q: at what point are teams successful (Include V&V? Code is running on day 1? Thru entire machine lifecycle?); difference between app readiness and system SW readiness; *not* a porting activity but helping at some labs can involve rewriting/refactoring code; require identification of staff on code team to serve as interface to AR effort
  - Q: is there sufficient driving force for use of new architectures w/o AR teams? A: No, facilitation is needed; “catalysts” is a better characterization;
  - Problem of application transience; 3 categories: always at LCFs, new at LCFs, in between. Makes it difficult to decide which teams to work with; level of need is an important factor;
  - Important to ensure that whatever work takes place becomes part of mainstream code efforts;
  - Importance of profiling, with tool (can’t always trust conventional wisdom)



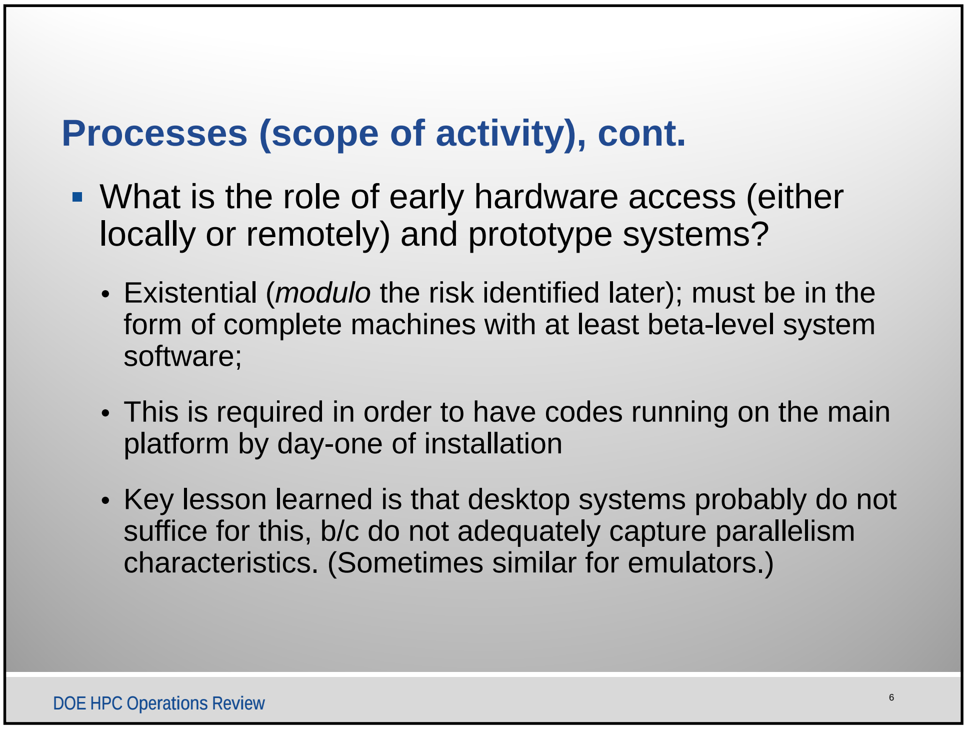
Processes (scope of activity)

- Explaining architecture choices to code teams is an important activity
- Setting user expectations for newer systems
- Question of what to do about transitioning 3rd-party apps remains; users of these codes seem to be stranded



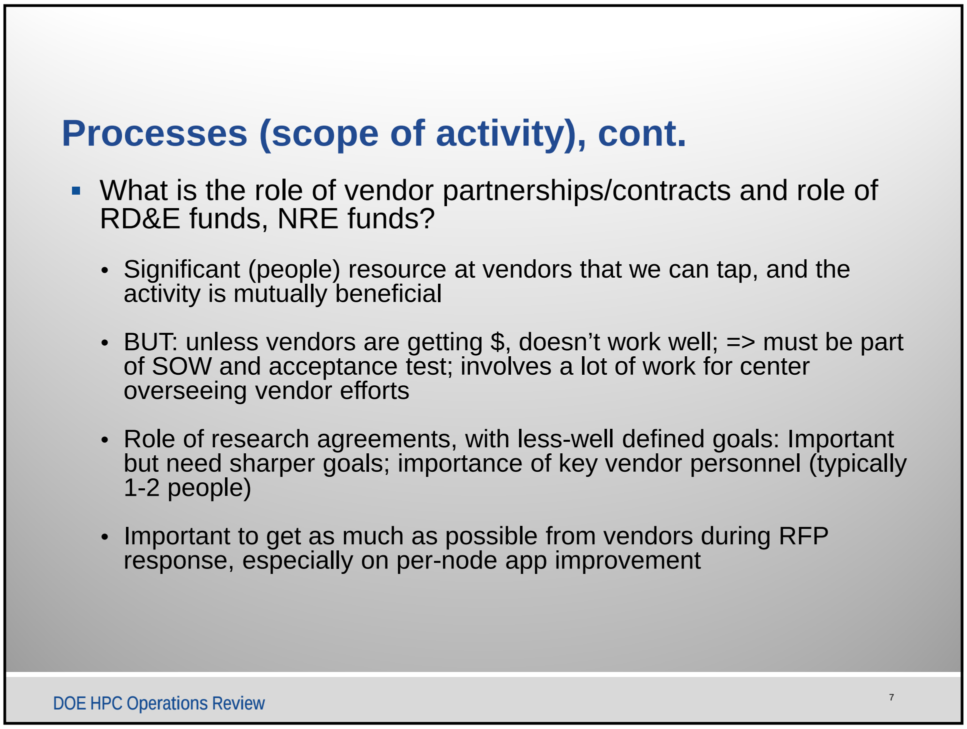
Processes (scope of activity), cont.

- What begins first: timeline for activities (before or after hardware)?
  - ID apps and appropriate problem sets, as well as personnel in center organization – preferably ~years before HW is available
  - Vendor-provided education, philosophy and periodic updates about systems for new platforms is essential
  - Tactical (shorter) and strategic (longer) work on codes



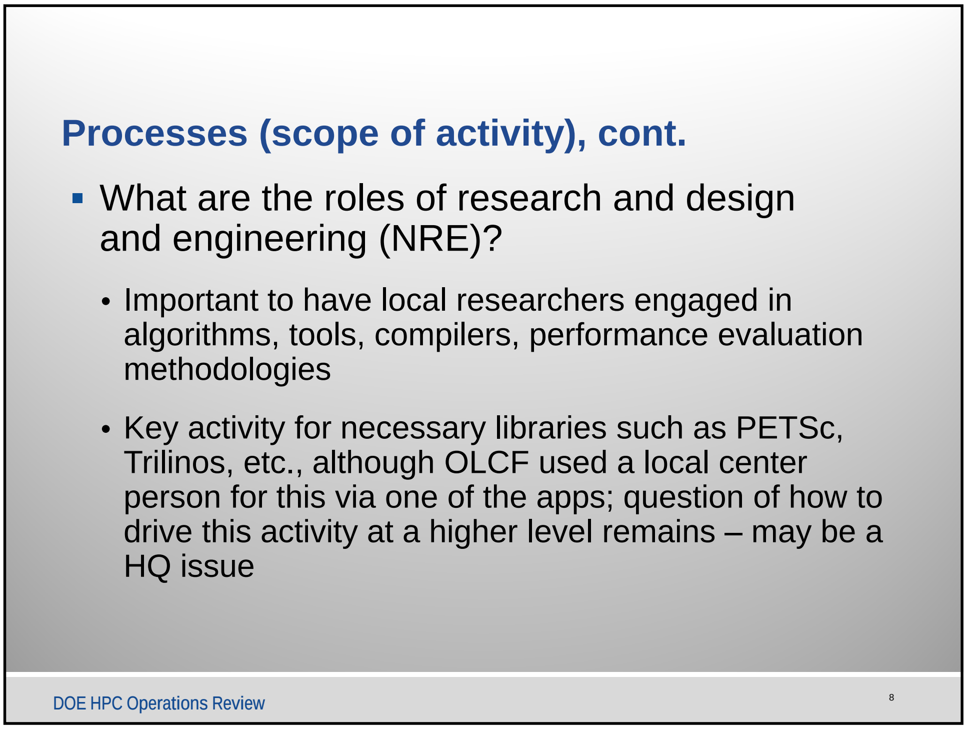
Processes (scope of activity), cont.

- What is the role of early hardware access (either locally or remotely) and prototype systems?
  - Existential (*modulo* the risk identified later); must be in the form of complete machines with at least beta-level system software;
  - This is required in order to have codes running on the main platform by day-one of installation
  - Key lesson learned is that desktop systems probably do not suffice for this, b/c do not adequately capture parallelism characteristics. (Sometimes similar for emulators.)



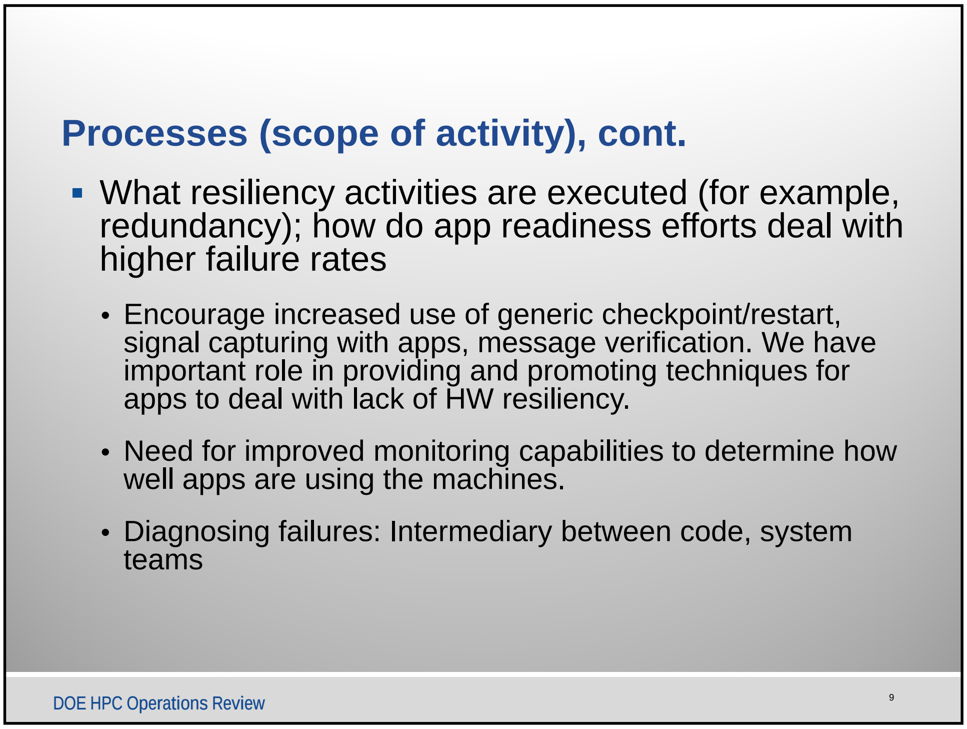
Processes (scope of activity), cont.

- What is the role of vendor partnerships/contracts and role of RD&E funds, NRE funds?
  - Significant (people) resource at vendors that we can tap, and the activity is mutually beneficial
  - BUT: unless vendors are getting $, doesn't work well; => must be part of SOW and acceptance test; involves a lot of work for center overseeing vendor efforts
  - Role of research agreements, with less-well defined goals: Important but need sharper goals; importance of key vendor personnel (typically 1-2 people)
  - Important to get as much as possible from vendors during RFP response, especially on per-node app improvement



Processes (scope of activity), cont.

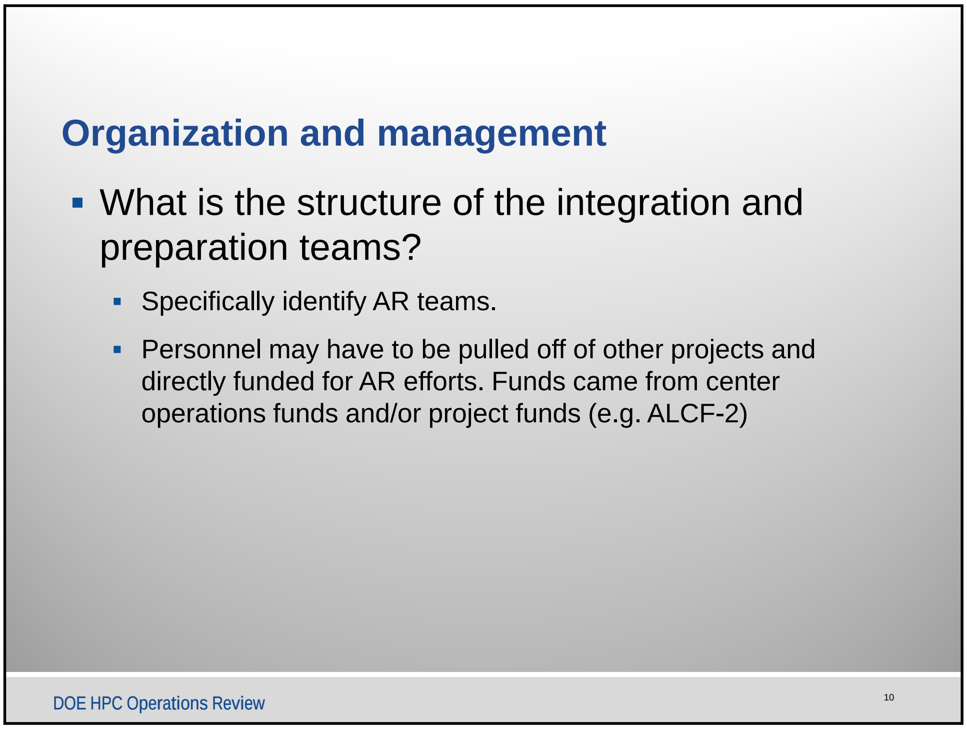
- What are the roles of research and design and engineering (NRE)?
  - Important to have local researchers engaged in algorithms, tools, compilers, performance evaluation methodologies
  - Key activity for necessary libraries such as PETSc, Trilinos, etc., although OLCF used a local center person for this via one of the apps; question of how to drive this activity at a higher level remains – may be a HQ issue



Processes (scope of activity), cont.

- What resiliency activities are executed (for example, redundancy); how do app readiness efforts deal with higher failure rates
  - Encourage increased use of generic checkpoint/restart, signal capturing with apps, message verification. We have important role in providing and promoting techniques for apps to deal with lack of HW resiliency.
  - Need for improved monitoring capabilities to determine how well apps are using the machines.
  - Diagnosing failures: Intermediary between code, system teams

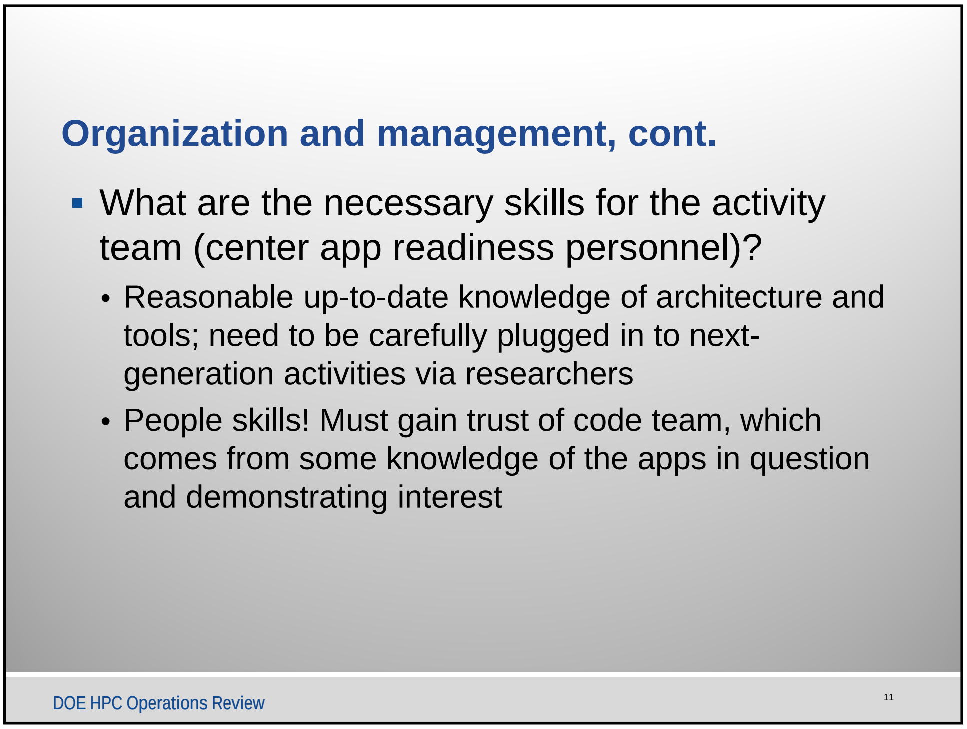
DOE HPC Operations Review 9



Organization and management

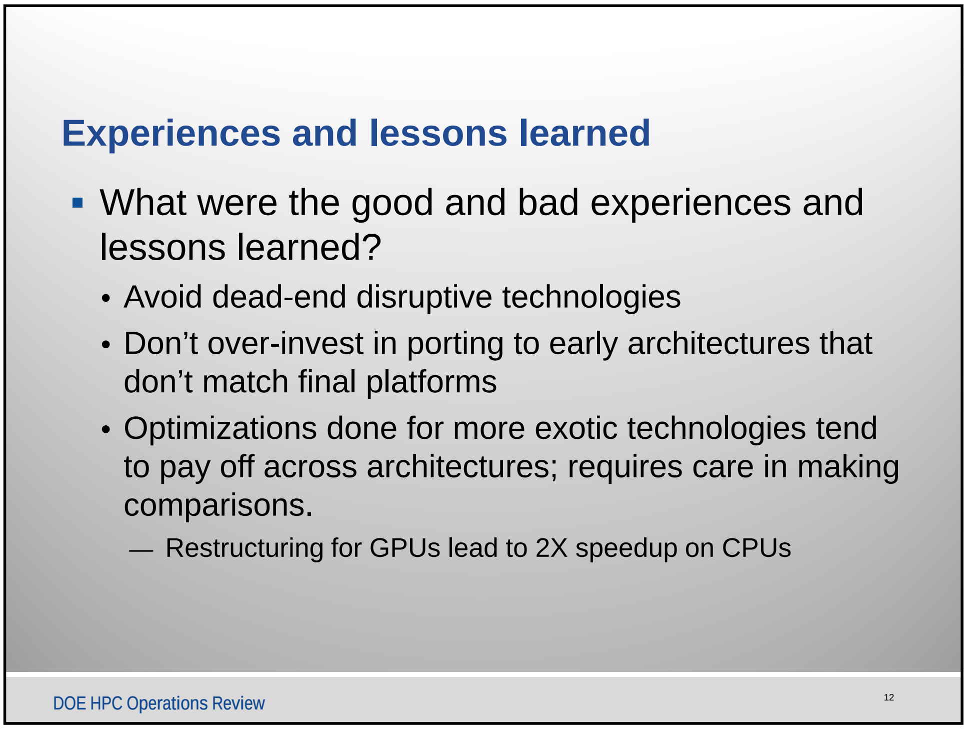
- What is the structure of the integration and preparation teams?
  - Specifically identify AR teams.
  - Personnel may have to be pulled off of other projects and directly funded for AR efforts. Funds came from center operations funds and/or project funds (e.g. ALCF-2)

DOE HPC Operations Review 10



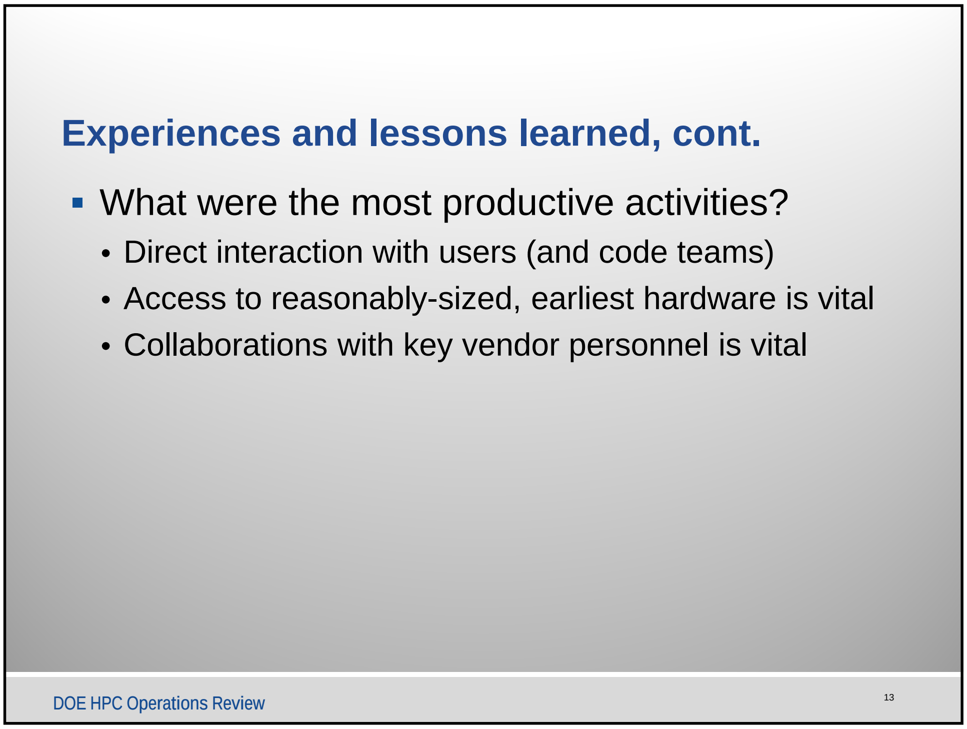
Organization and management, cont.

- What are the necessary skills for the activity team (center app readiness personnel)?
  - Reasonable up-to-date knowledge of architecture and tools; need to be carefully plugged in to next-generation activities via researchers
  - People skills! Must gain trust of code team, which comes from some knowledge of the apps in question and demonstrating interest



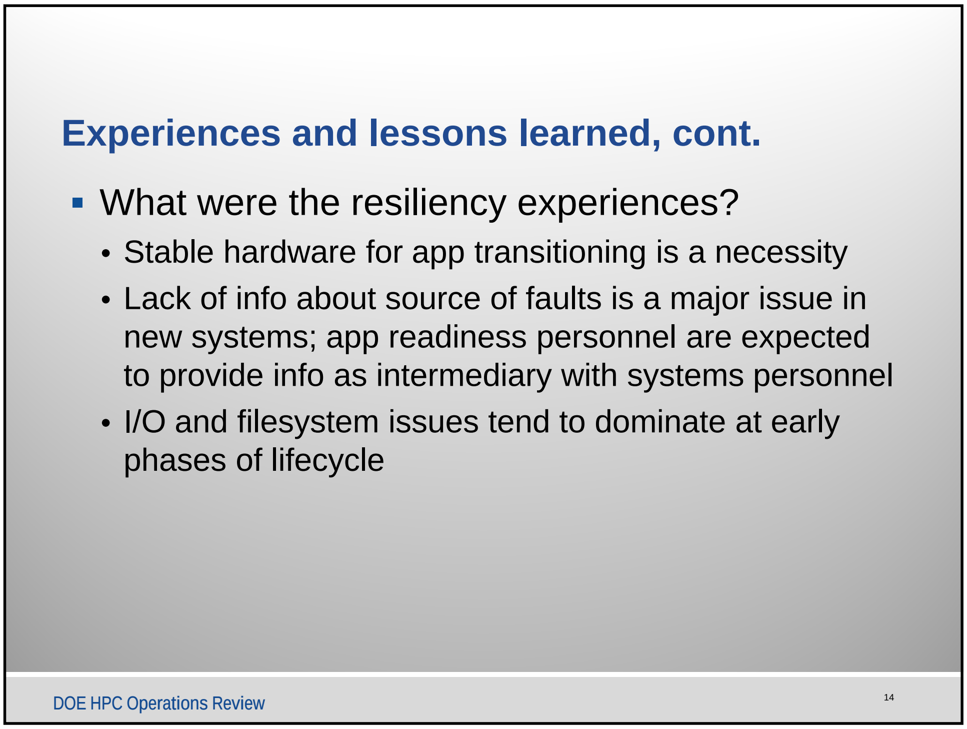
Experiences and lessons learned

- What were the good and bad experiences and lessons learned?
  - Avoid dead-end disruptive technologies
  - Don’t over-invest in porting to early architectures that don’t match final platforms
  - Optimizations done for more exotic technologies tend to pay off across architectures; requires care in making comparisons.
    - — Restructuring for GPUs lead to 2X speedup on CPUs



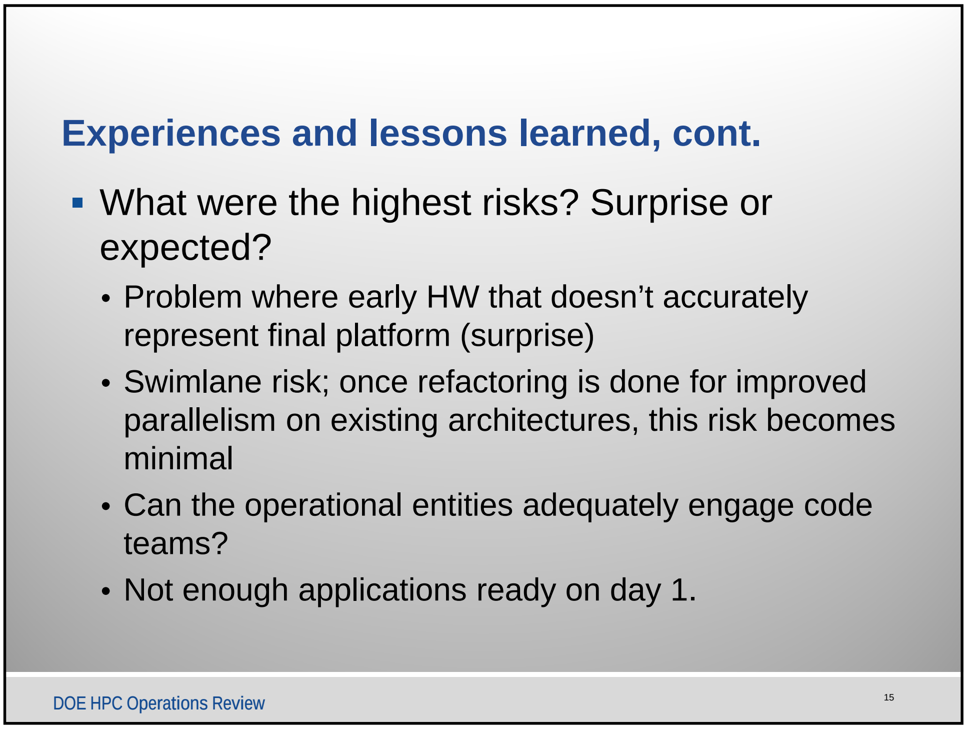
Experiences and lessons learned, cont.

- What were the most productive activities?
  - Direct interaction with users (and code teams)
  - Access to reasonably-sized, earliest hardware is vital
  - Collaborations with key vendor personnel is vital



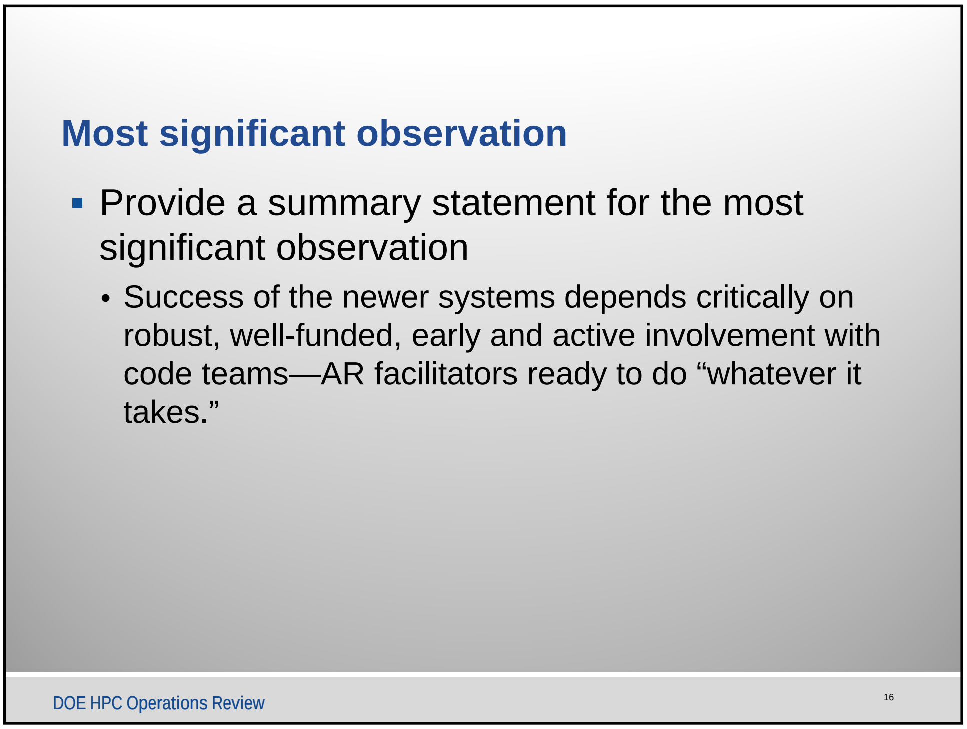
Experiences and lessons learned, cont.

- What were the resiliency experiences?
  - Stable hardware for app transitioning is a necessity
  - Lack of info about source of faults is a major issue in new systems; app readiness personnel are expected to provide info as intermediary with systems personnel
  - I/O and filesystem issues tend to dominate at early phases of lifecycle



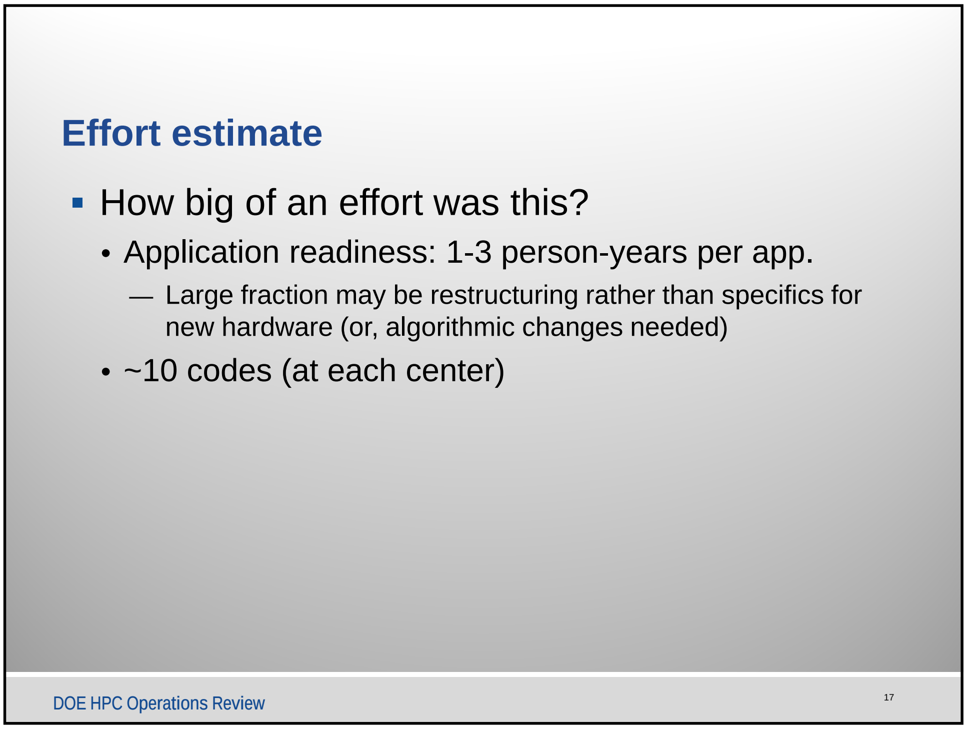
Experiences and lessons learned, cont.

- What were the highest risks? Surprise or expected?
  - Problem where early HW that doesn’t accurately represent final platform (surprise)
  - Swimlane risk; once refactoring is done for improved parallelism on existing architectures, this risk becomes minimal
  - Can the operational entities adequately engage code teams?
  - Not enough applications ready on day 1.



Most significant observation

- Provide a summary statement for the most significant observation
  - Success of the newer systems depends critically on robust, well-funded, early and active involvement with code teams—AR facilitators ready to do “whatever it takes.”



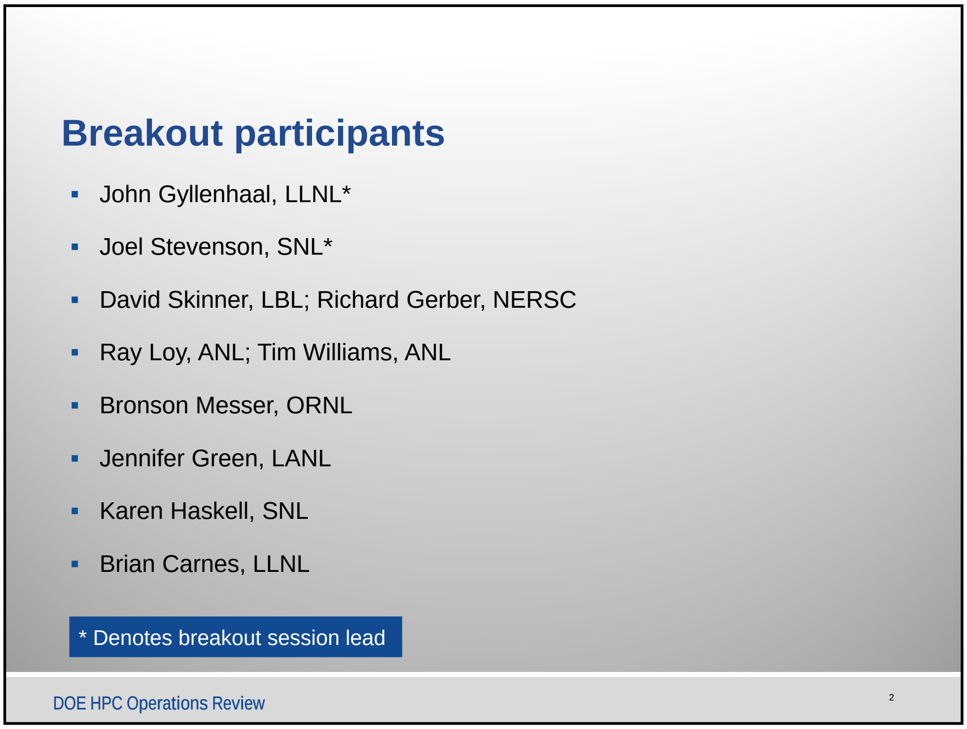
Effort estimate

- How big of an effort was this?
  - Application readiness: 1-3 person-years per app.
    - — Large fraction may be restructuring rather than specifics for new hardware (or, algorithmic changes needed)
  - ~10 codes (at each center)

Development Environment Preparation (parallelism support, compilers, tools)

DOE HPC Operations Review
San Francisco, November 5-6, 2013

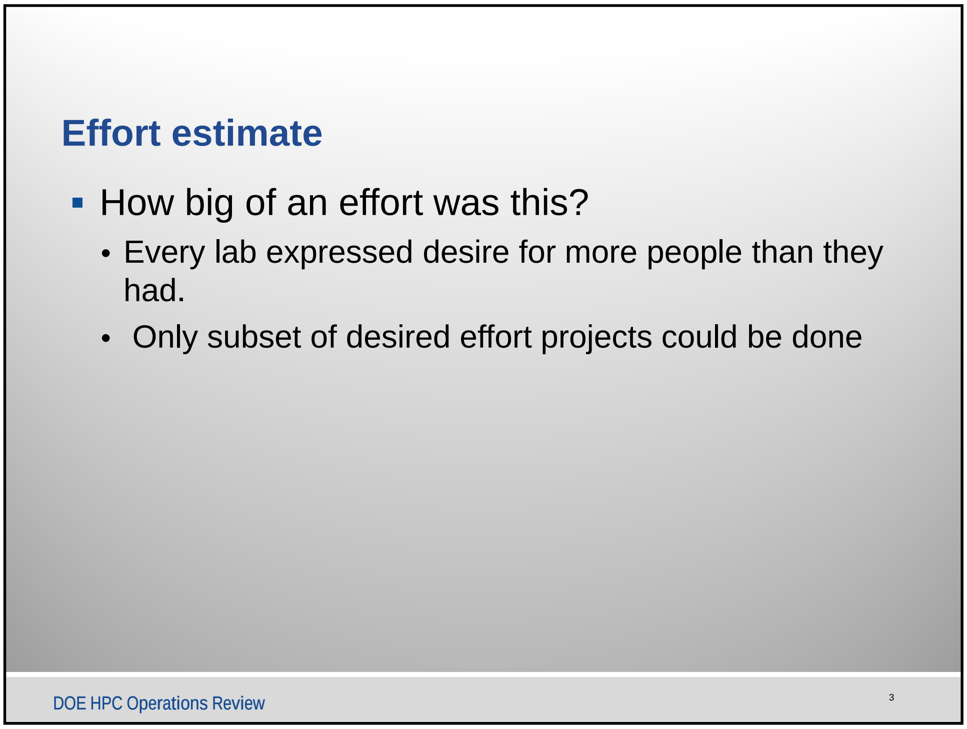
LLNL-PRES-646612



Breakout participants

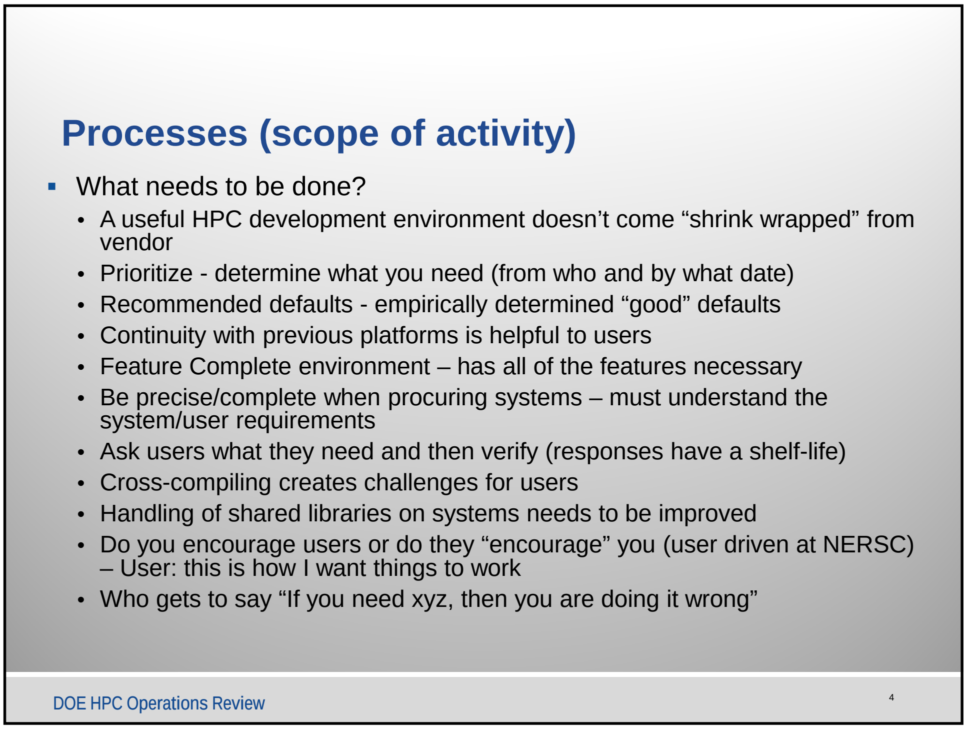
- John Gyllenhaal, LLNL*
- Joel Stevenson, SNL*
- David Skinner, LBL; Richard Gerber, NERSC
- Ray Loy, ANL; Tim Williams, ANL
- Bronson Messer, ORNL
- Jennifer Green, LANL
- Karen Haskell, SNL
- Brian Carnes, LLNL

* Denotes breakout session lead



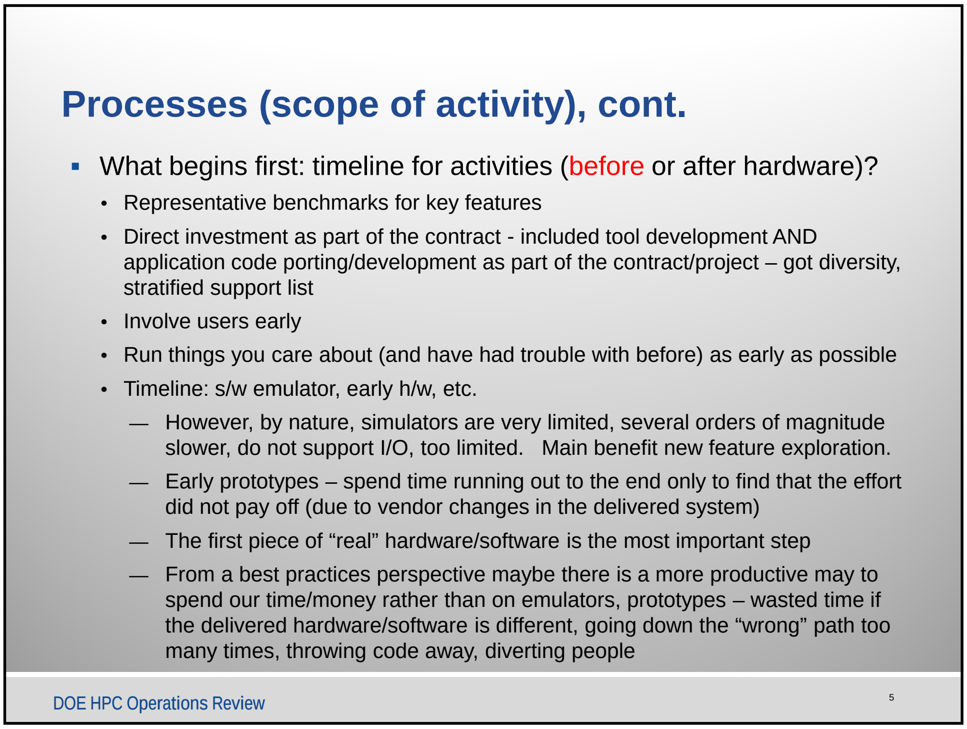
Effort estimate

- How big of an effort was this?
  - Every lab expressed desire for more people than they had.
  - Only subset of desired effort projects could be done



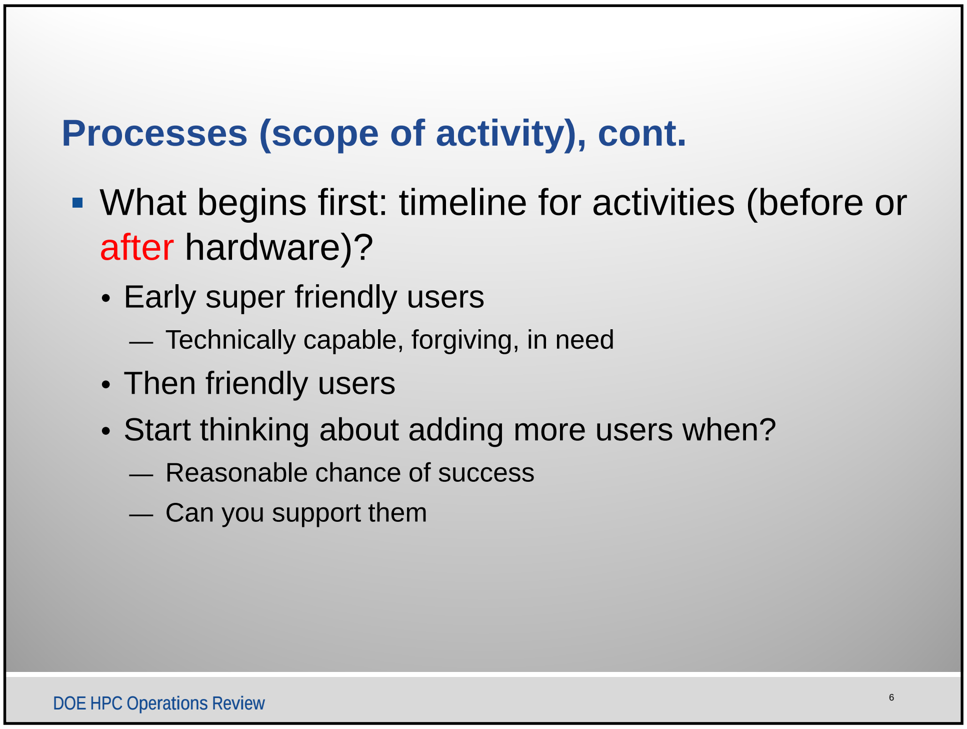
Processes (scope of activity)

- What needs to be done?
  - A useful HPC development environment doesn’t come “shrink wrapped” from vendor
  - Prioritize - determine what you need (from who and by what date)
  - Recommended defaults - empirically determined “good” defaults
  - Continuity with previous platforms is helpful to users
  - Feature Complete environment – has all of the features necessary
  - Be precise/complete when procuring systems – must understand the system/user requirements
  - Ask users what they need and then verify (responses have a shelf-life)
  - Cross-compiling creates challenges for users
  - Handling of shared libraries on systems needs to be improved
  - Do you encourage users or do they “encourage” you (user driven at NERSC) – User: this is how I want things to work
  - Who gets to say “If you need xyz, then you are doing it wrong”



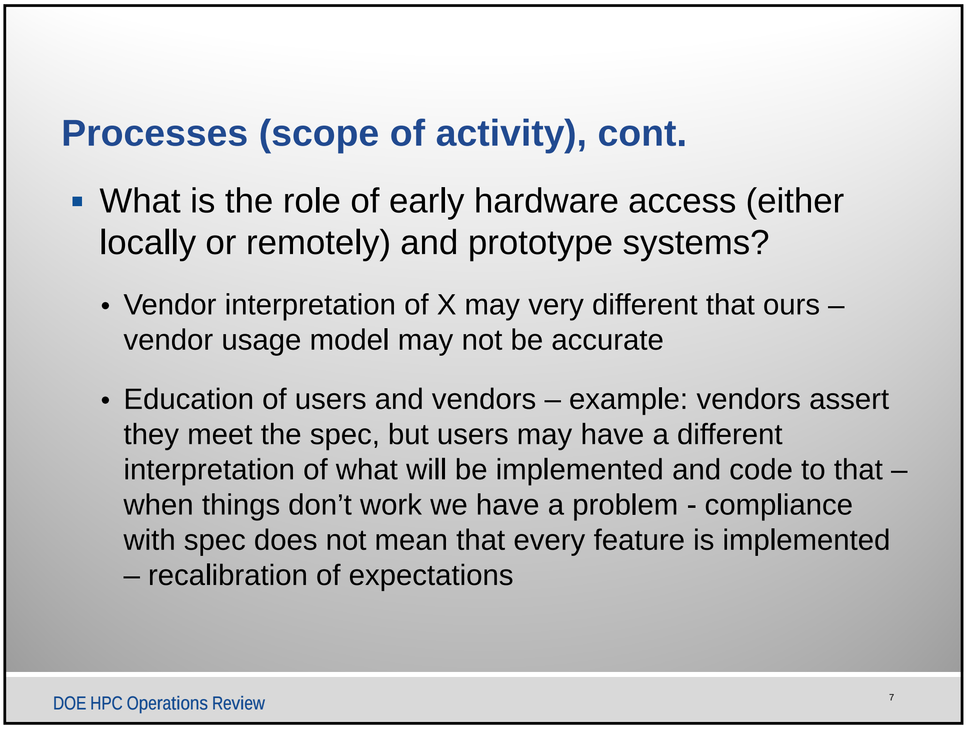
Processes (scope of activity), cont.

- What begins first: timeline for activities (before or after hardware)?
  - Representative benchmarks for key features
  - Direct investment as part of the contract - included tool development AND application code porting/development as part of the contract/project – got diversity, stratified support list
  - Involve users early
  - Run things you care about (and have had trouble with before) as early as possible
  - Timeline: s/w emulator, early h/w, etc.
    - However, by nature, simulators are very limited, several orders of magnitude slower, do not support I/O, too limited.   Main benefit new feature exploration.
    - Early prototypes – spend time running out to the end only to find that the effort did not pay off (due to vendor changes in the delivered system)
    - The first piece of "real" hardware/software is the most important step
    - From a best practices perspective maybe there is a more productive may to spend our time/money rather than on emulators, prototypes – wasted time if the delivered hardware/software is different, going down the "wrong" path too many times, throwing code away, diverting people



Processes (scope of activity), cont.

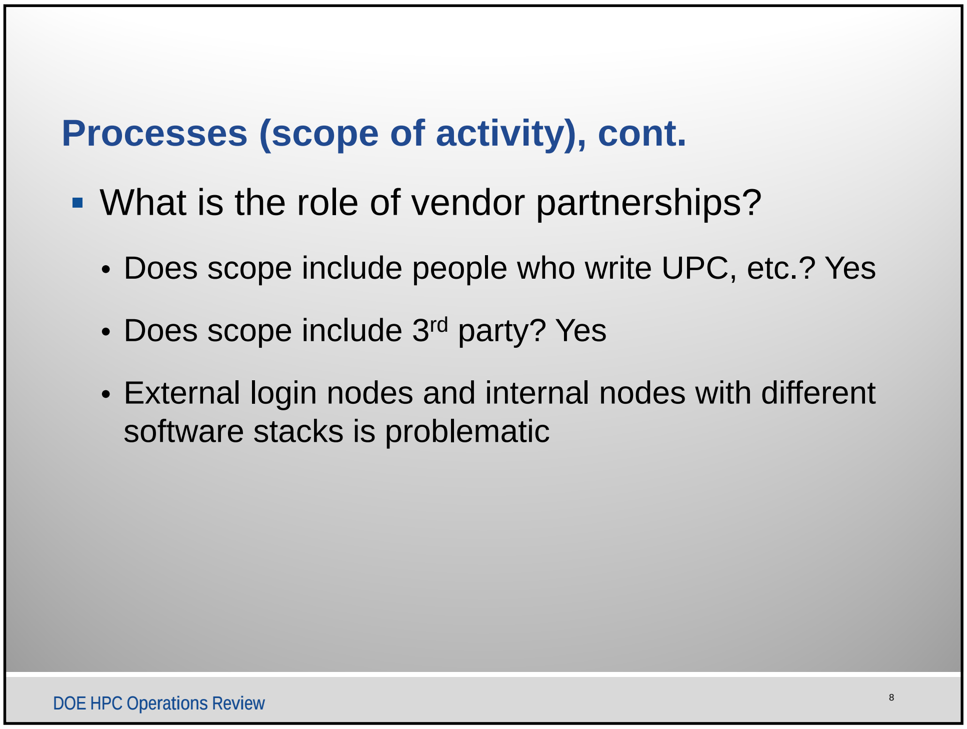
- What begins first: timeline for activities (before or after hardware)?
  - Early super friendly users
    - Technically capable, forgiving, in need
  - Then friendly users
  - Start thinking about adding more users when?
    - Reasonable chance of success
    - Can you support them



Processes (scope of activity), cont.

- What is the role of early hardware access (either locally or remotely) and prototype systems?
  - Vendor interpretation of X may very different that ours – vendor usage model may not be accurate
  - Education of users and vendors – example: vendors assert they meet the spec, but users may have a different interpretation of what will be implemented and code to that – when things don't work we have a problem - compliance with spec does not mean that every feature is implemented – recalibration of expectations

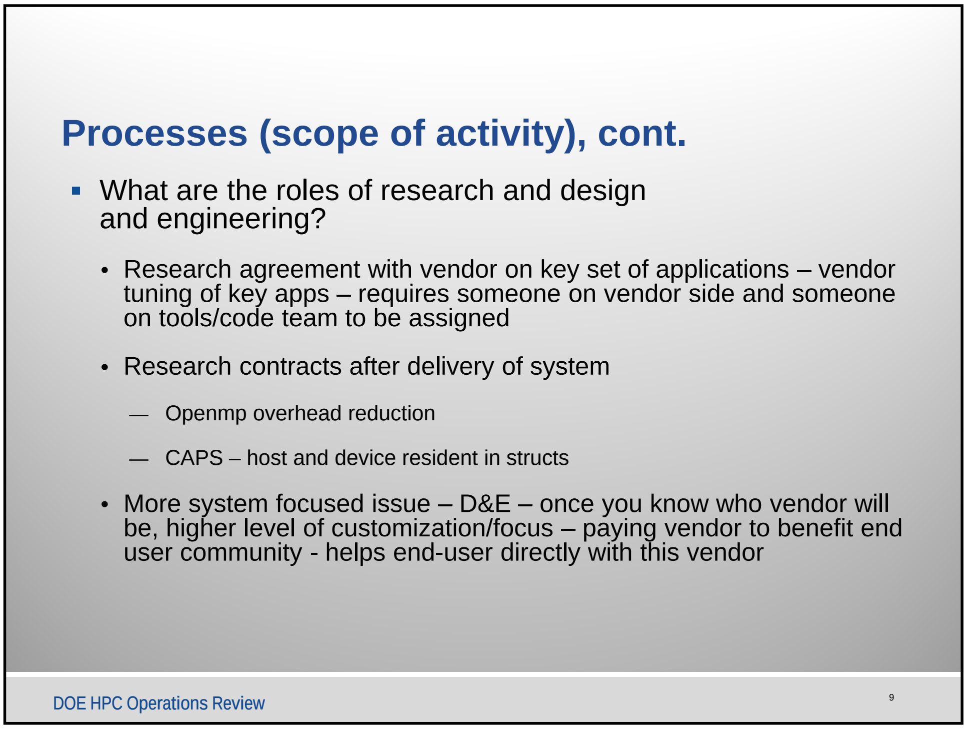
DOE HPC Operations Review 7



Processes (scope of activity), cont.

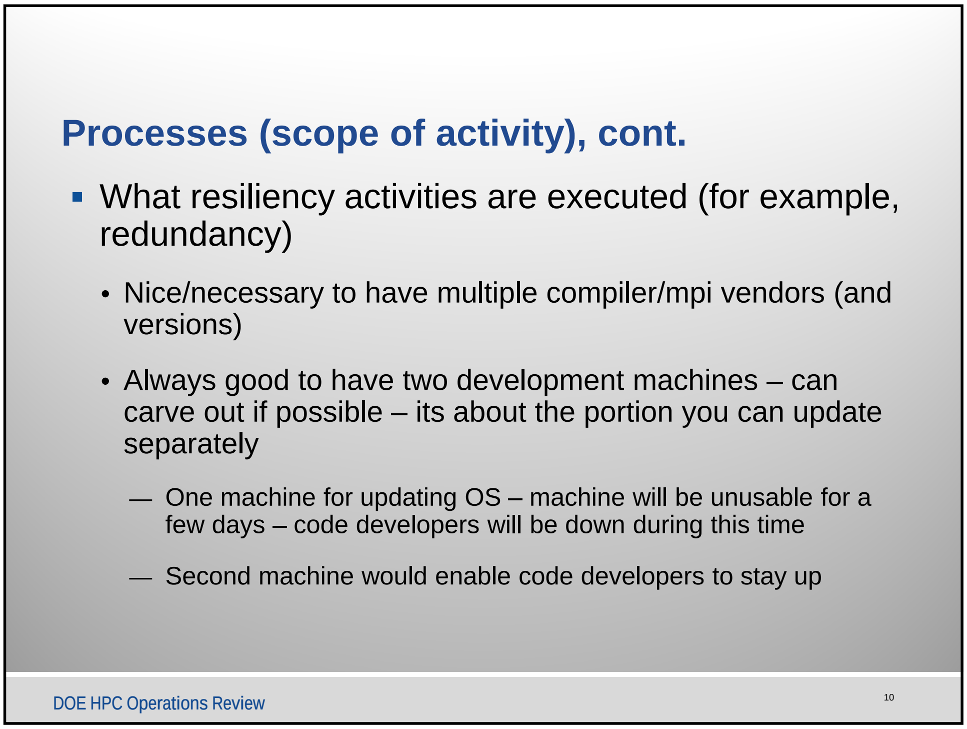
- What is the role of vendor partnerships?
  - Does scope include people who write UPC, etc.? Yes
  - Does scope include 3rd party? Yes
  - External login nodes and internal nodes with different software stacks is problematic

DOE HPC Operations Review 8



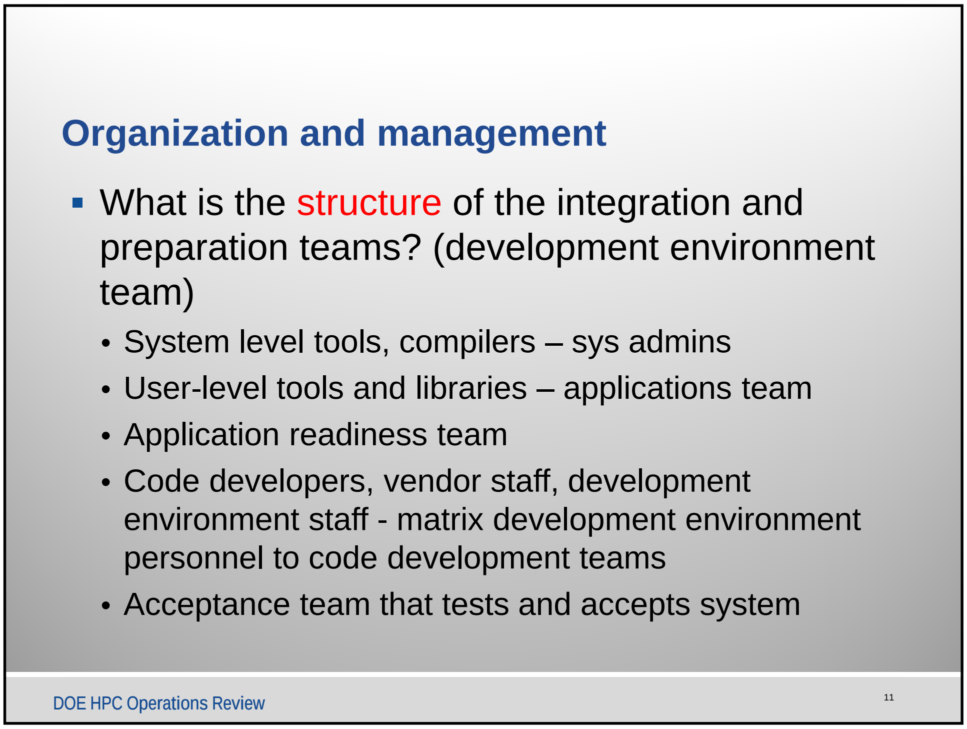
Processes (scope of activity), cont.

- What are the roles of research and design and engineering?
  - Research agreement with vendor on key set of applications – vendor tuning of key apps – requires someone on vendor side and someone on tools/code team to be assigned
  - Research contracts after delivery of system
    - — Openmp overhead reduction
    - — CAPS – host and device resident in structs
  - More system focused issue – D&E – once you know who vendor will be, higher level of customization/focus – paying vendor to benefit end user community - helps end-user directly with this vendor



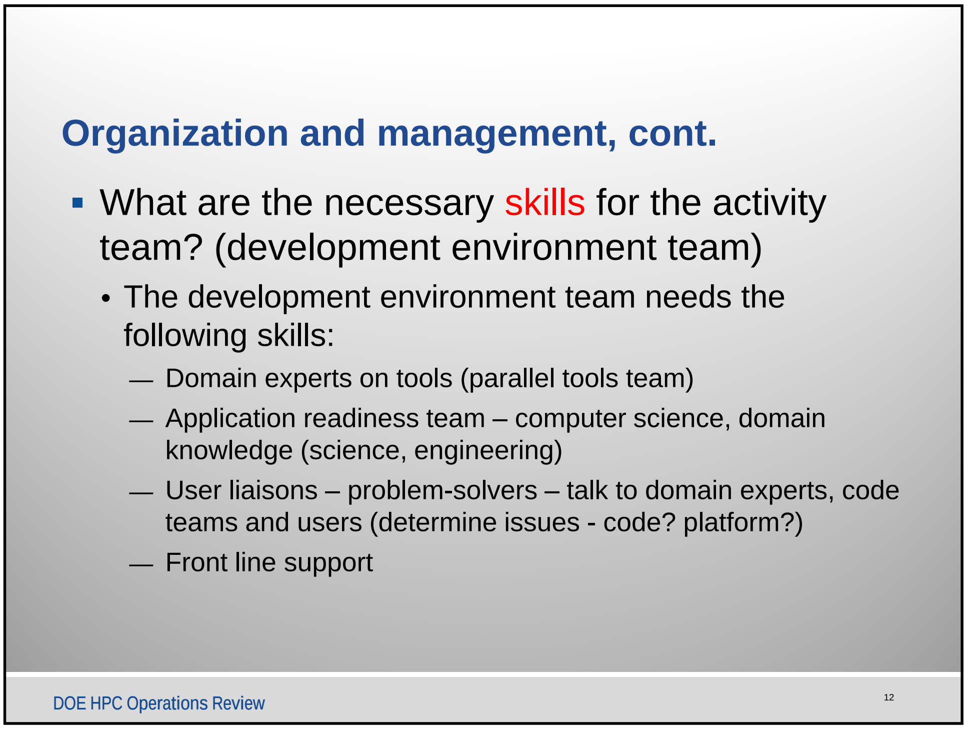
Processes (scope of activity), cont.

- What resiliency activities are executed (for example, redundancy)
  - Nice/necessary to have multiple compiler/mpi vendors (and versions)
  - Always good to have two development machines – can carve out if possible – its about the portion you can update separately
    - — One machine for updating OS – machine will be unusable for a few days – code developers will be down during this time
    - — Second machine would enable code developers to stay up



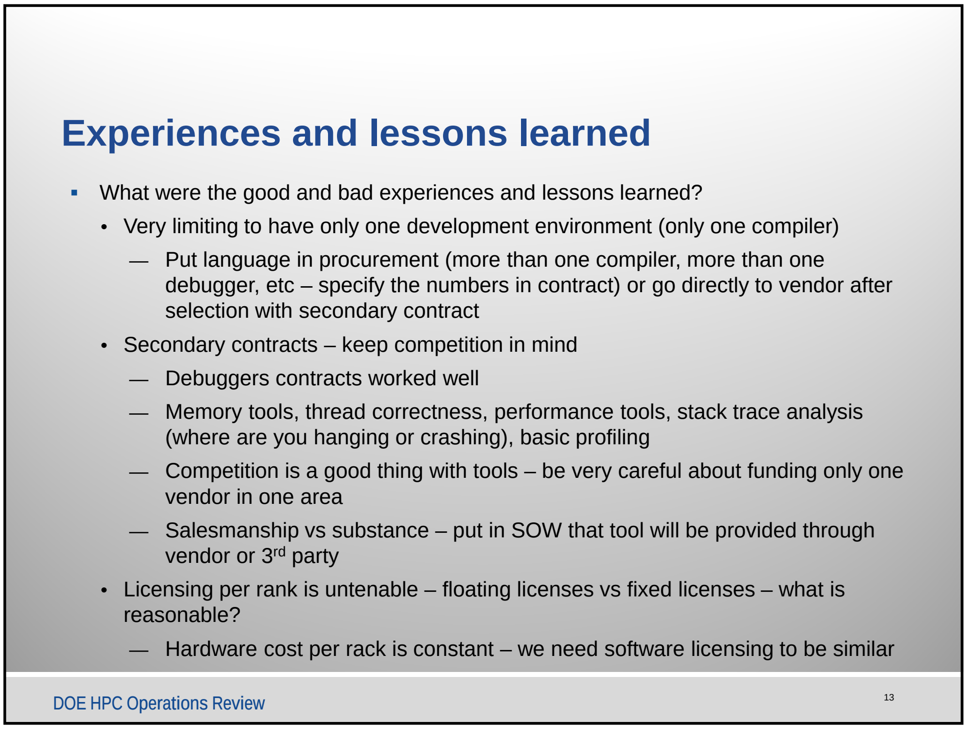
Organization and management

- What is the structure of the integration and preparation teams? (development environment team)
  - System level tools, compilers – sys admins
  - User-level tools and libraries – applications team
  - Application readiness team
  - Code developers, vendor staff, development environment staff - matrix development environment personnel to code development teams
  - Acceptance team that tests and accepts system



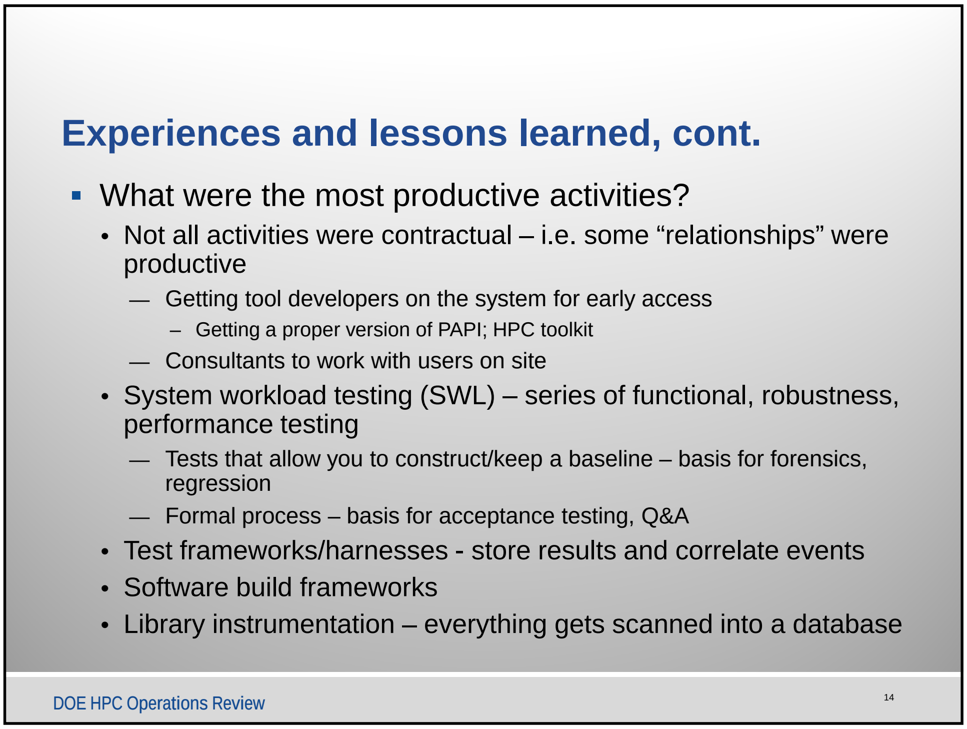
Organization and management, cont.

- What are the necessary skills for the activity team? (development environment team)
  - The development environment team needs the following skills:
    - — Domain experts on tools (parallel tools team)
    - — Application readiness team – computer science, domain knowledge (science, engineering)
    - — User liaisons – problem-solvers – talk to domain experts, code teams and users (determine issues - code? platform?)
    - — Front line support



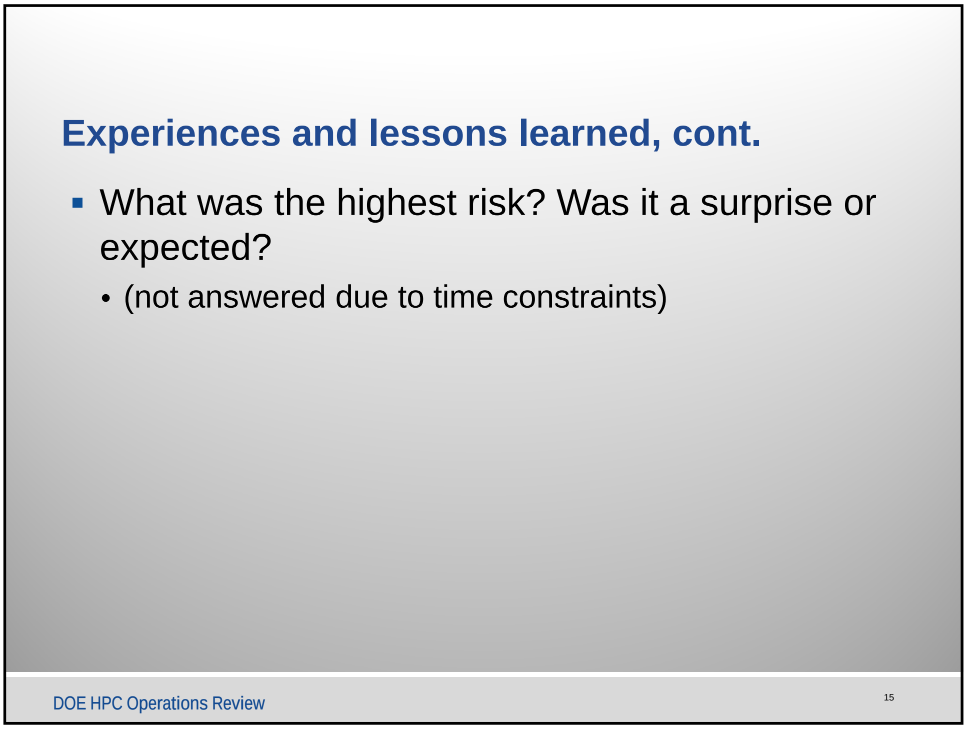
Experiences and lessons learned

- What were the good and bad experiences and lessons learned?
  - Very limiting to have only one development environment (only one compiler)
    - Put language in procurement (more than one compiler, more than one debugger, etc – specify the numbers in contract) or go directly to vendor after selection with secondary contract
  - Secondary contracts – keep competition in mind
    - Debuggers contracts worked well
    - Memory tools, thread correctness, performance tools, stack trace analysis (where are you hanging or crashing), basic profiling
    - Competition is a good thing with tools – be very careful about funding only one vendor in one area
    - Salesmanship vs substance – put in SOW that tool will be provided through vendor or 3rd party
  - Licensing per rank is untenable – floating licenses vs fixed licenses – what is reasonable?
    - Hardware cost per rack is constant – we need software licensing to be similar



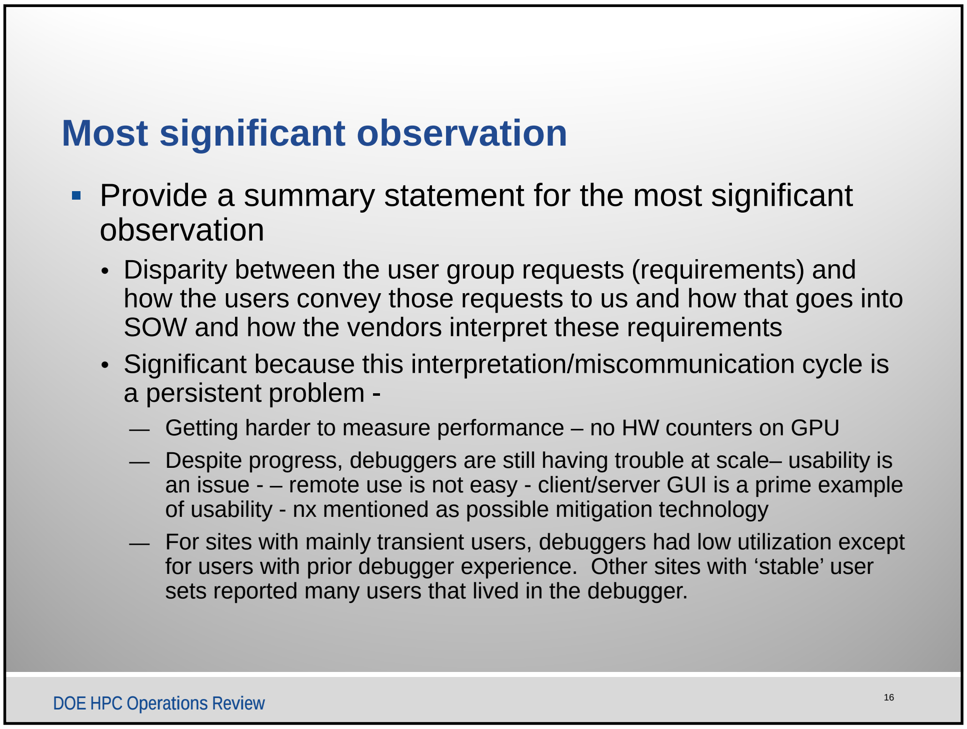
Experiences and lessons learned, cont.

- What were the most productive activities?
  - Not all activities were contractual – i.e. some “relationships” were productive
    - Getting tool developers on the system for early access
      - Getting a proper version of PAPI; HPC toolkit
    - Consultants to work with users on site
  - System workload testing (SWL) – series of functional, robustness, performance testing
    - Tests that allow you to construct/keep a baseline – basis for forensics, regression
    - Formal process – basis for acceptance testing, Q&A
  - Test frameworks/harnesses - store results and correlate events
  - Software build frameworks
  - Library instrumentation – everything gets scanned into a database



Experiences and lessons learned, cont.

- What was the highest risk? Was it a surprise or expected?
  - (not answered due to time constraints)



Most significant observation

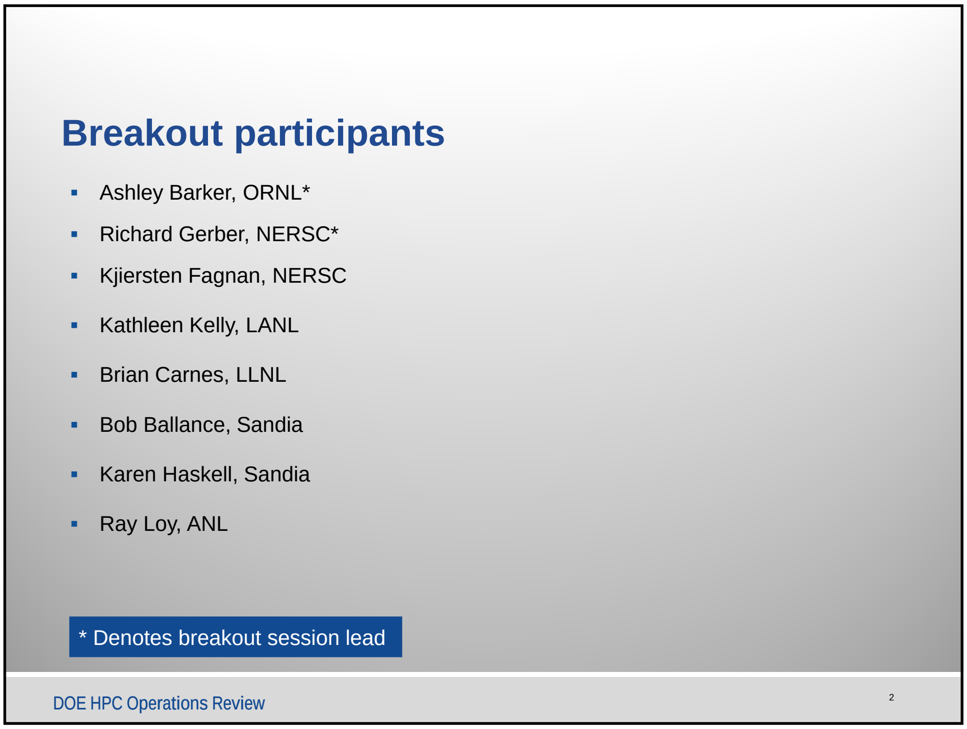
- Provide a summary statement for the most significant observation
  - Disparity between the user group requests (requirements) and how the users convey those requests to us and how that goes into SOW and how the vendors interpret these requirements
  - Significant because this interpretation/miscommunication cycle is a persistent problem -
    - Getting harder to measure performance – no HW counters on GPU
    - Despite progress, debuggers are still having trouble at scale– usability is an issue - – remote use is not easy - client/server GUI is a prime example of usability - nx mentioned as possible mitigation technology
    - For sites with mainly transient users, debuggers had low utilization except for users with prior debugger experience. Other sites with ‘stable’ user sets reported many users that lived in the debugger.

Usage Models, Training, Education, and User Support

DOE HPC Operations Review

San Francisco, November 5-6, 2013

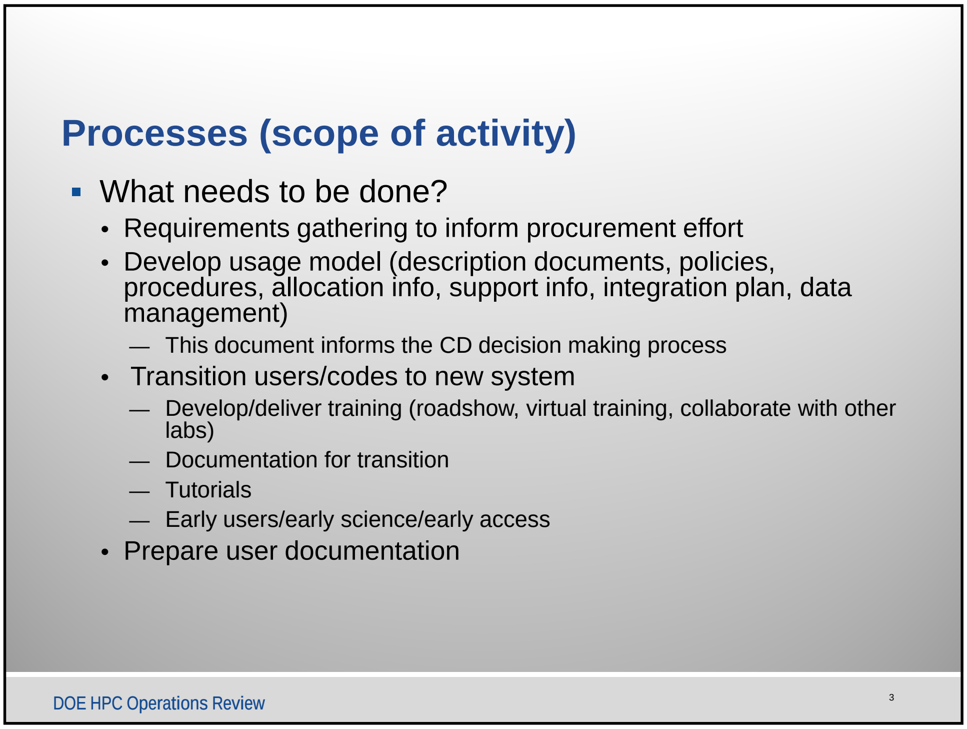
LLNL-PRES-646612



Breakout participants

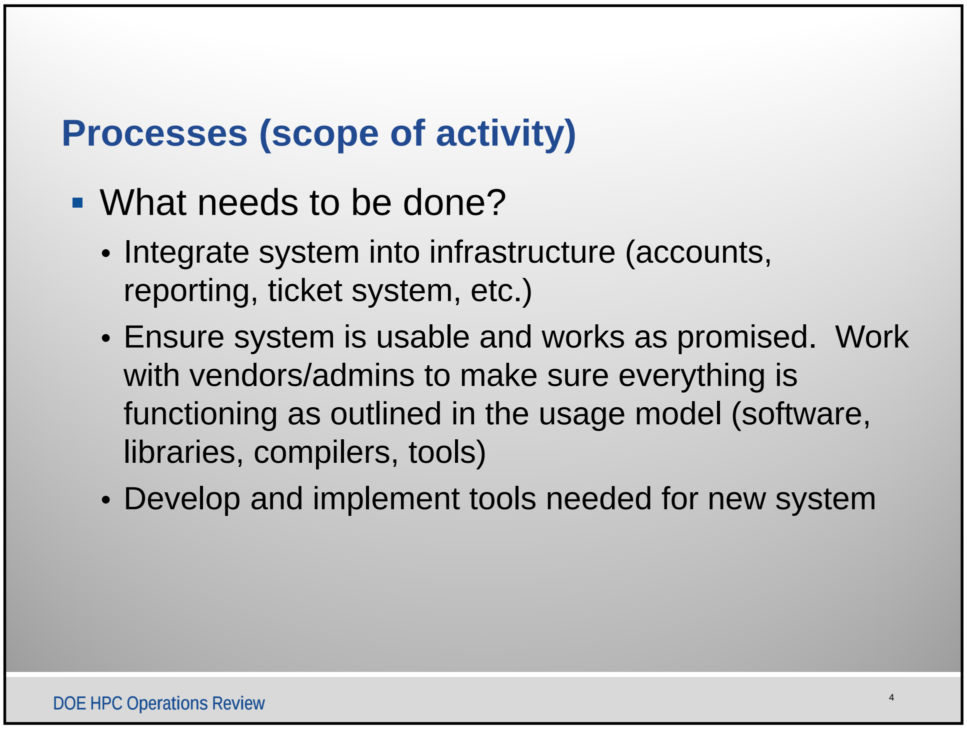
- Ashley Barker, ORNL*
- Richard Gerber, NERSC*
- Kjiersten Fagnan, NERSC
- Kathleen Kelly, LANL
- Brian Carnes, LLNL
- Bob Ballance, Sandia
- Karen Haskell, Sandia
- Ray Loy, ANL

* Denotes breakout session lead



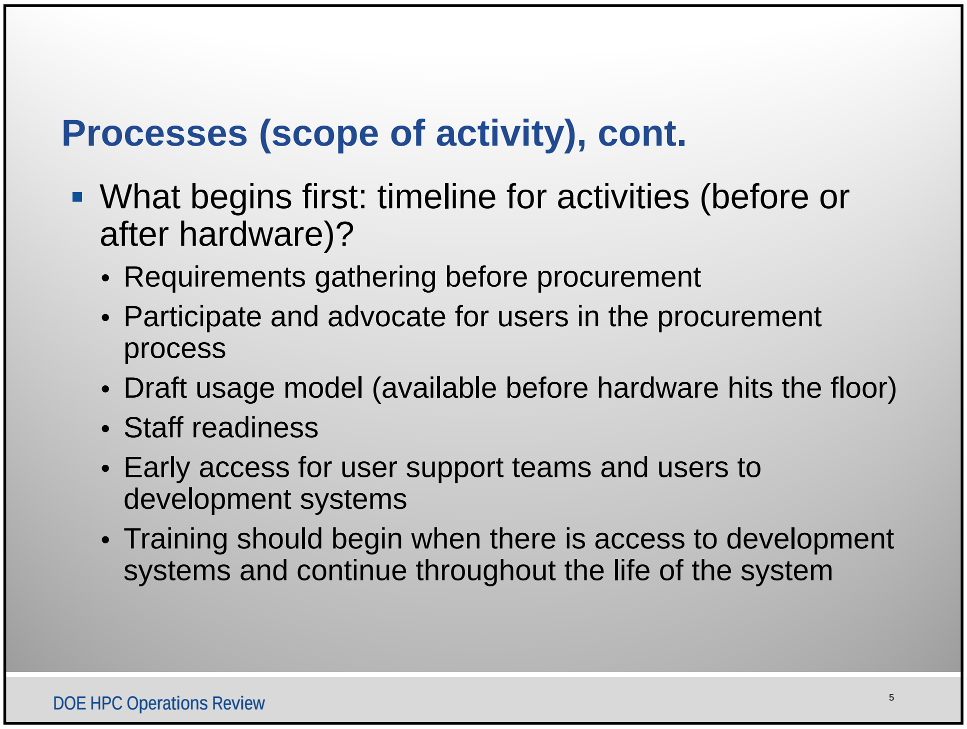
Processes (scope of activity)

- What needs to be done?
  - Requirements gathering to inform procurement effort
  - Develop usage model (description documents, policies, procedures, allocation info, support info, integration plan, data management)
    - This document informs the CD decision making process
  - Transition users/codes to new system
    - Develop/deliver training (roadshow, virtual training, collaborate with other labs)
    - Documentation for transition
    - Tutorials
    - Early users/early science/early access
  - Prepare user documentation



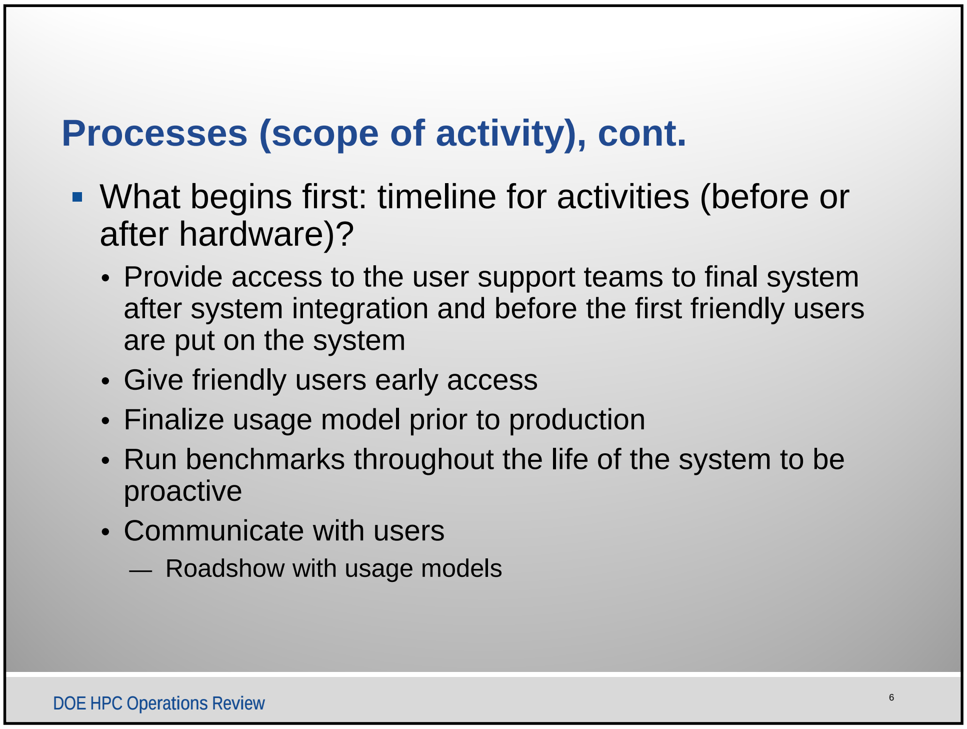
Processes (scope of activity)

- What needs to be done?
  - Integrate system into infrastructure (accounts, reporting, ticket system, etc.)
  - Ensure system is usable and works as promised. Work with vendors/admins to make sure everything is functioning as outlined in the usage model (software, libraries, compilers, tools)
  - Develop and implement tools needed for new system



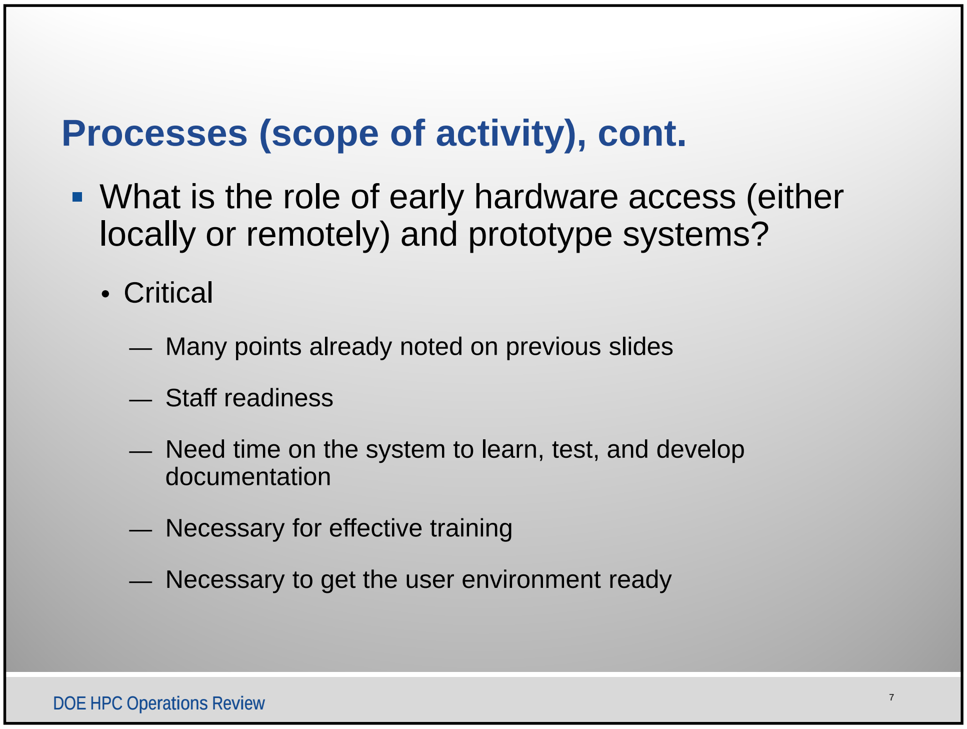
Processes (scope of activity), cont.

- What begins first: timeline for activities (before or after hardware)?
  - Requirements gathering before procurement
  - Participate and advocate for users in the procurement process
  - Draft usage model (available before hardware hits the floor)
  - Staff readiness
  - Early access for user support teams and users to development systems
  - Training should begin when there is access to development systems and continue throughout the life of the system



Processes (scope of activity), cont.

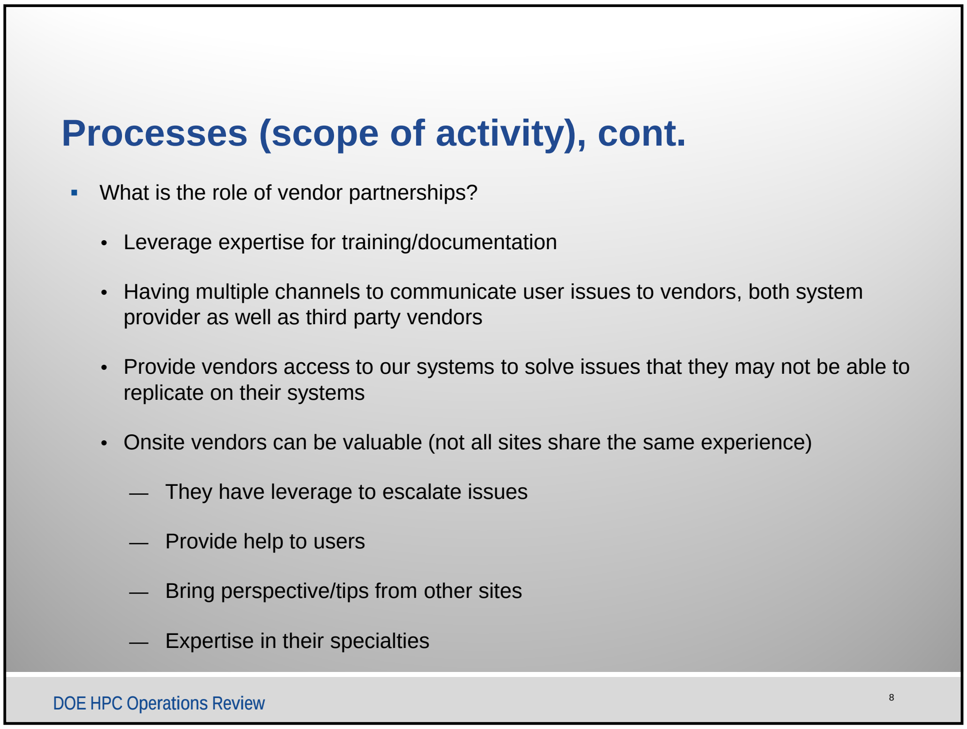
- What begins first: timeline for activities (before or after hardware)?
  - Provide access to the user support teams to final system after system integration and before the first friendly users are put on the system
  - Give friendly users early access
  - Finalize usage model prior to production
  - Run benchmarks throughout the life of the system to be proactive
  - Communicate with users
    - — Roadshow with usage models



Processes (scope of activity), cont.

- What is the role of early hardware access (either locally or remotely) and prototype systems?
  - Critical
    - — Many points already noted on previous slides
    - — Staff readiness
    - — Need time on the system to learn, test, and develop documentation
    - — Necessary for effective training
    - — Necessary to get the user environment ready

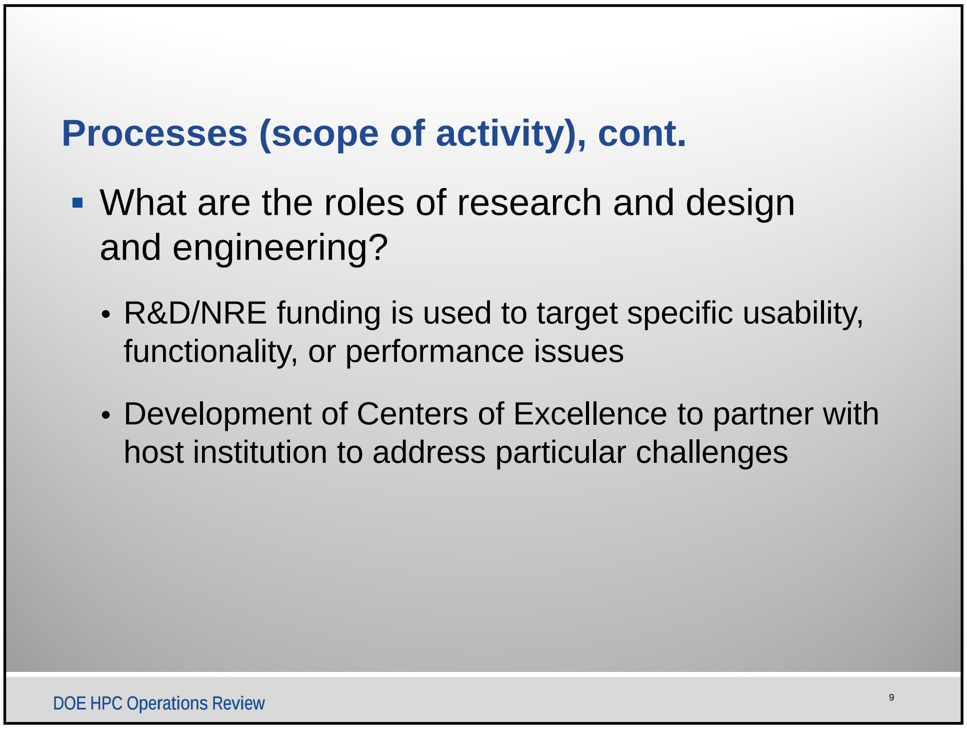
DOE HPC Operations Review 7



Processes (scope of activity), cont.

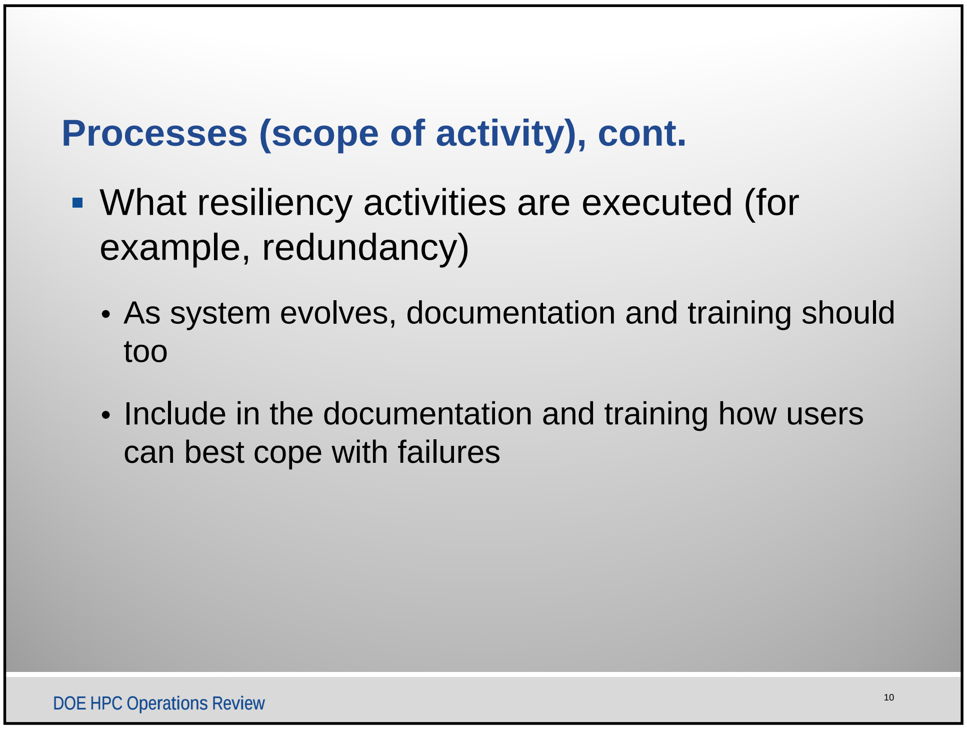
- What is the role of vendor partnerships?
  - Leverage expertise for training/documentation
  - Having multiple channels to communicate user issues to vendors, both system provider as well as third party vendors
  - Provide vendors access to our systems to solve issues that they may not be able to replicate on their systems
  - Onsite vendors can be valuable (not all sites share the same experience)
    - — They have leverage to escalate issues
    - — Provide help to users
    - — Bring perspective/tips from other sites
    - — Expertise in their specialties

DOE HPC Operations Review 8



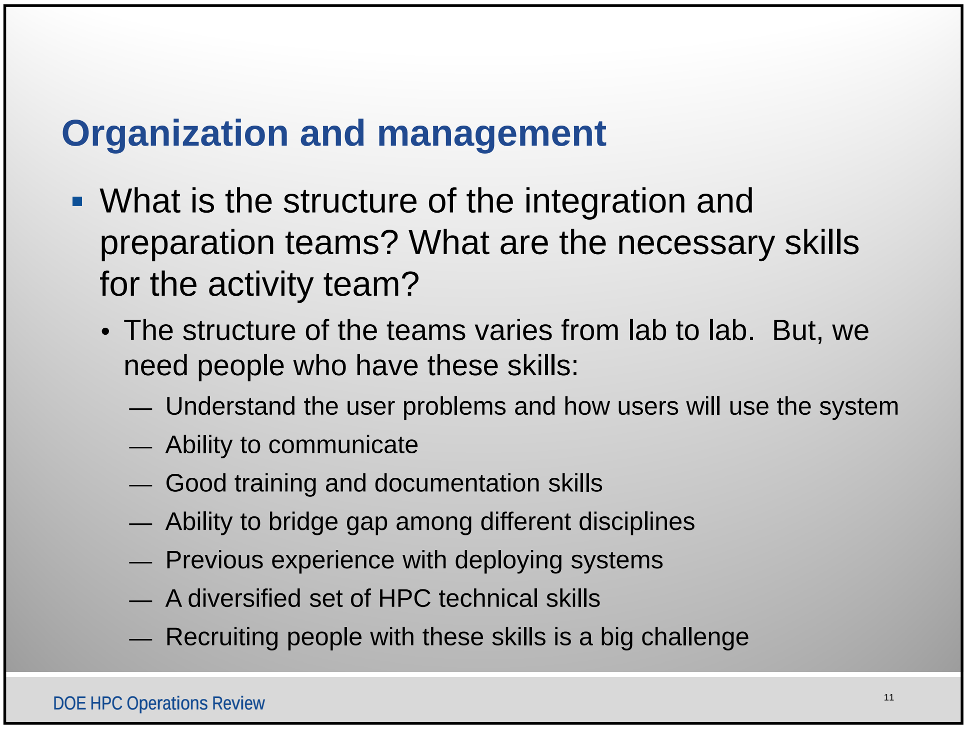
Processes (scope of activity), cont.

- What are the roles of research and design and engineering?
  - R&D/NRE funding is used to target specific usability, functionality, or performance issues
  - Development of Centers of Excellence to partner with host institution to address particular challenges



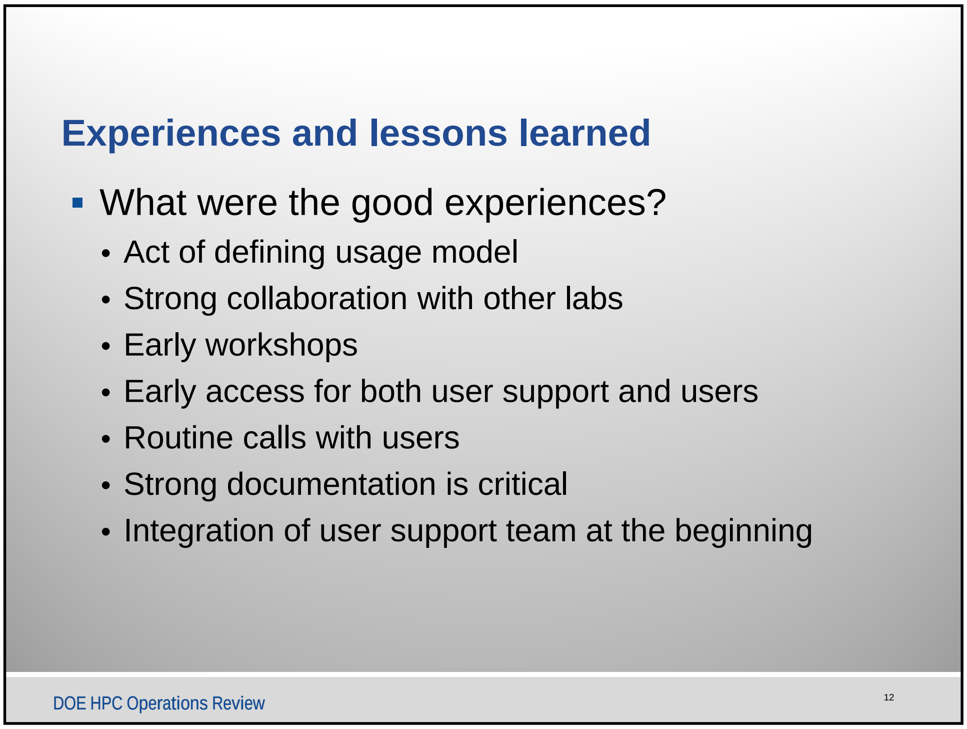
Processes (scope of activity), cont.

- What resiliency activities are executed (for example, redundancy)
  - As system evolves, documentation and training should too
  - Include in the documentation and training how users can best cope with failures



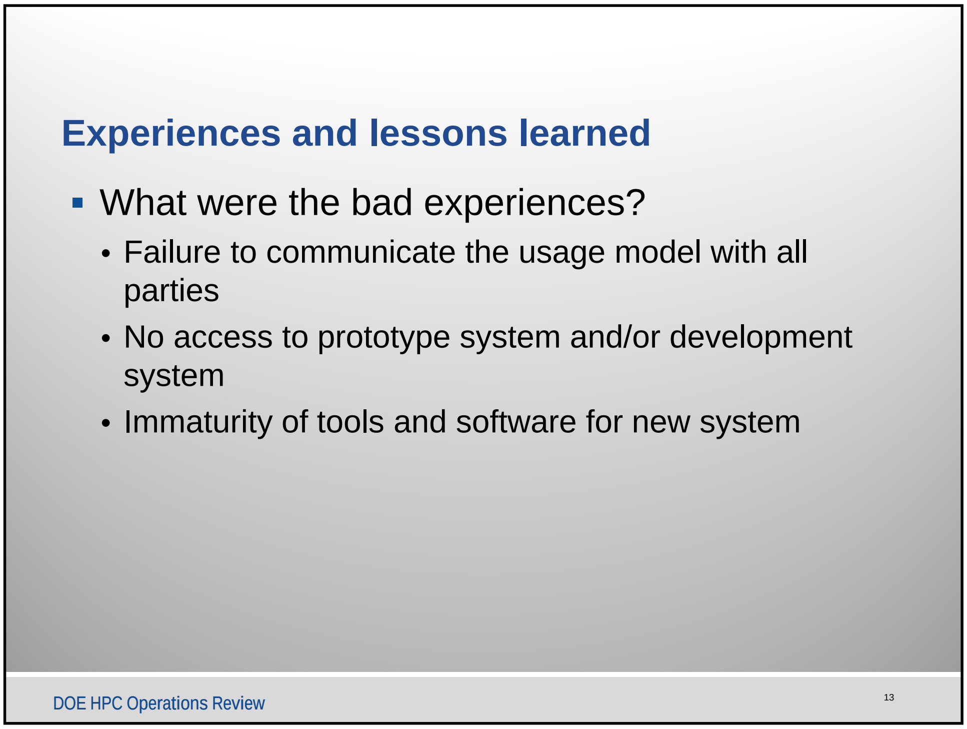
Organization and management

- What is the structure of the integration and preparation teams? What are the necessary skills for the activity team?
  - The structure of the teams varies from lab to lab. But, we need people who have these skills:
    - — Understand the user problems and how users will use the system
    - — Ability to communicate
    - — Good training and documentation skills
    - — Ability to bridge gap among different disciplines
    - — Previous experience with deploying systems
    - — A diversified set of HPC technical skills
    - — Recruiting people with these skills is a big challenge



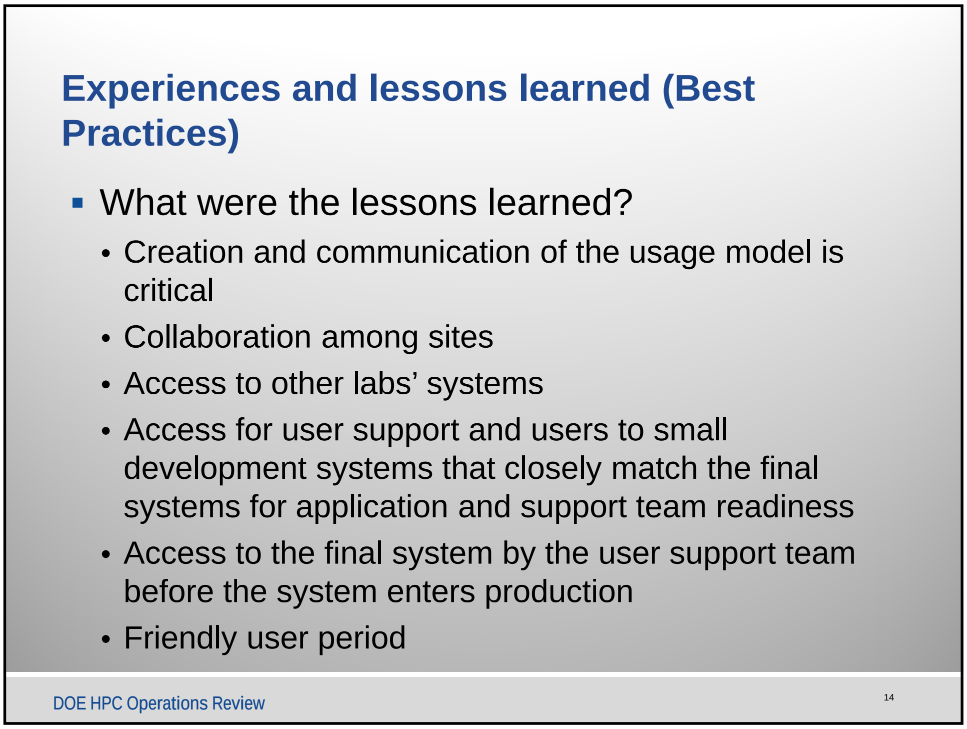
Experiences and lessons learned

- What were the good experiences?
  - Act of defining usage model
  - Strong collaboration with other labs
  - Early workshops
  - Early access for both user support and users
  - Routine calls with users
  - Strong documentation is critical
  - Integration of user support team at the beginning



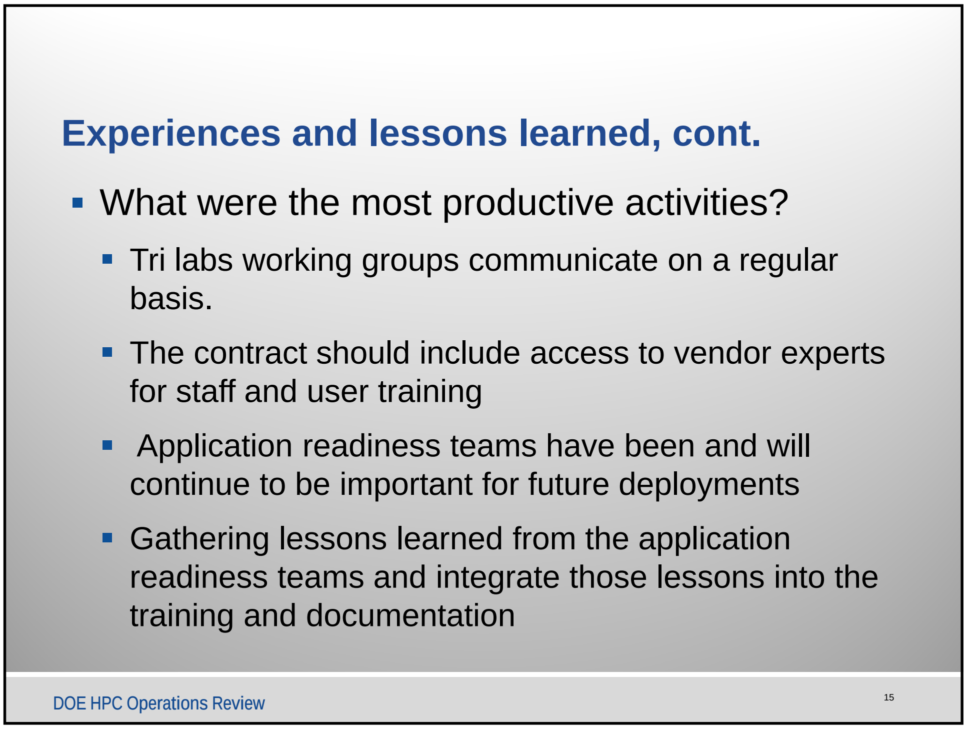
Experiences and lessons learned

- What were the bad experiences?
  - Failure to communicate the usage model with all parties
  - No access to prototype system and/or development system
  - Immaturity of tools and software for new system



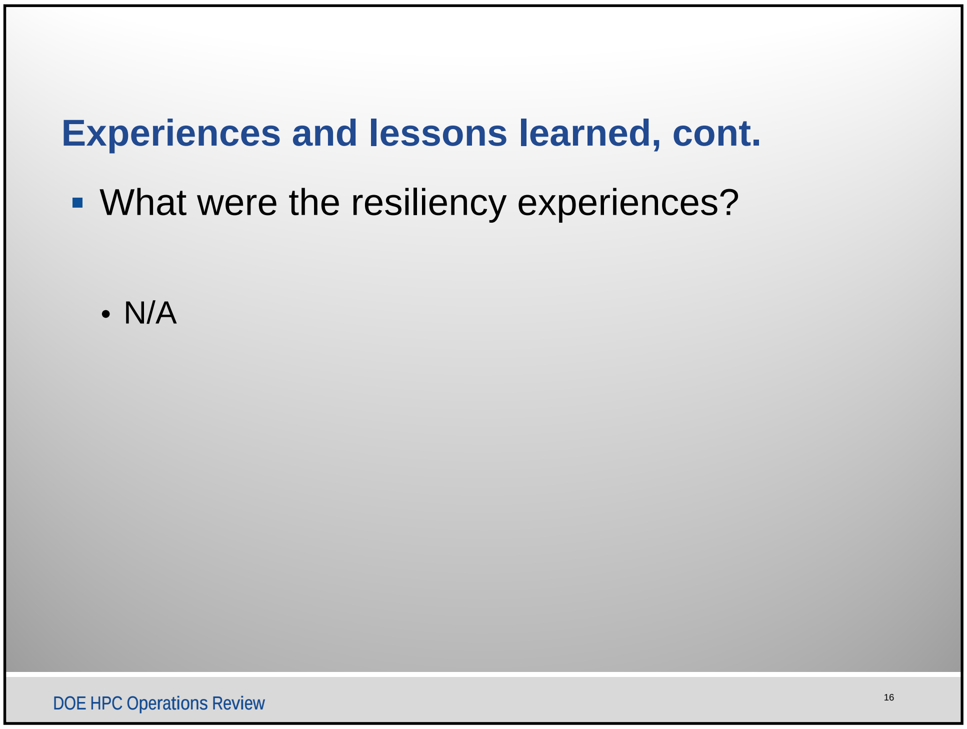
Experiences and lessons learned (Best Practices)

- What were the lessons learned?
  - Creation and communication of the usage model is critical
  - Collaboration among sites
  - Access to other labs' systems
  - Access for user support and users to small development systems that closely match the final systems for application and support team readiness
  - Access to the final system by the user support team before the system enters production
  - Friendly user period



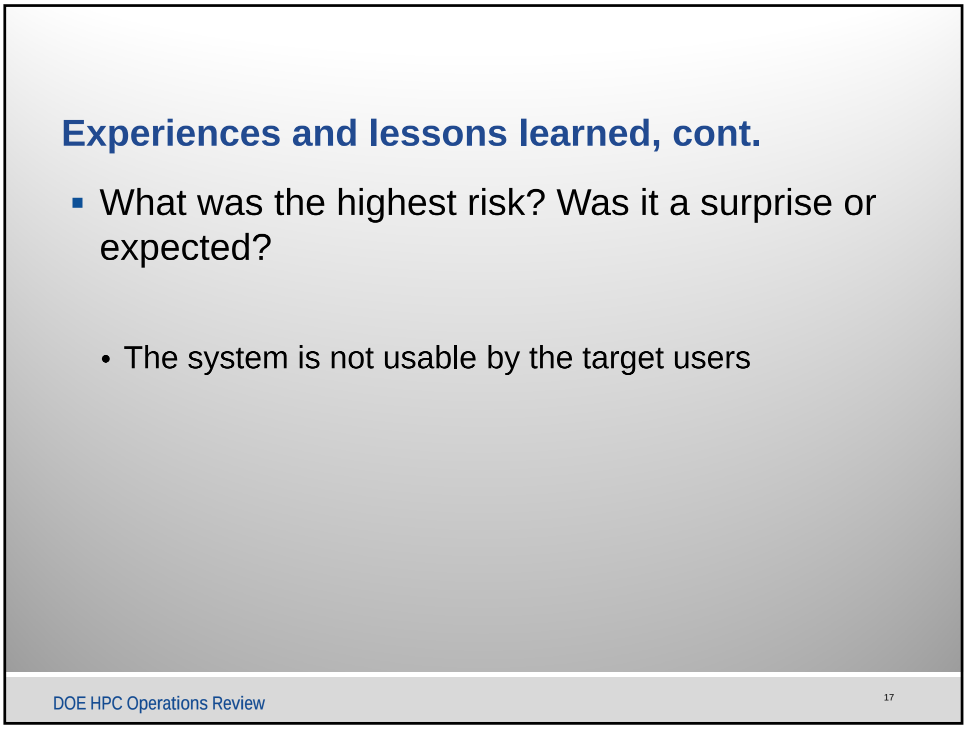
Experiences and lessons learned, cont.

- What were the most productive activities?
  - Tri labs working groups communicate on a regular basis.
  - The contract should include access to vendor experts for staff and user training
  - Application readiness teams have been and will continue to be important for future deployments
  - Gathering lessons learned from the application readiness teams and integrate those lessons into the training and documentation



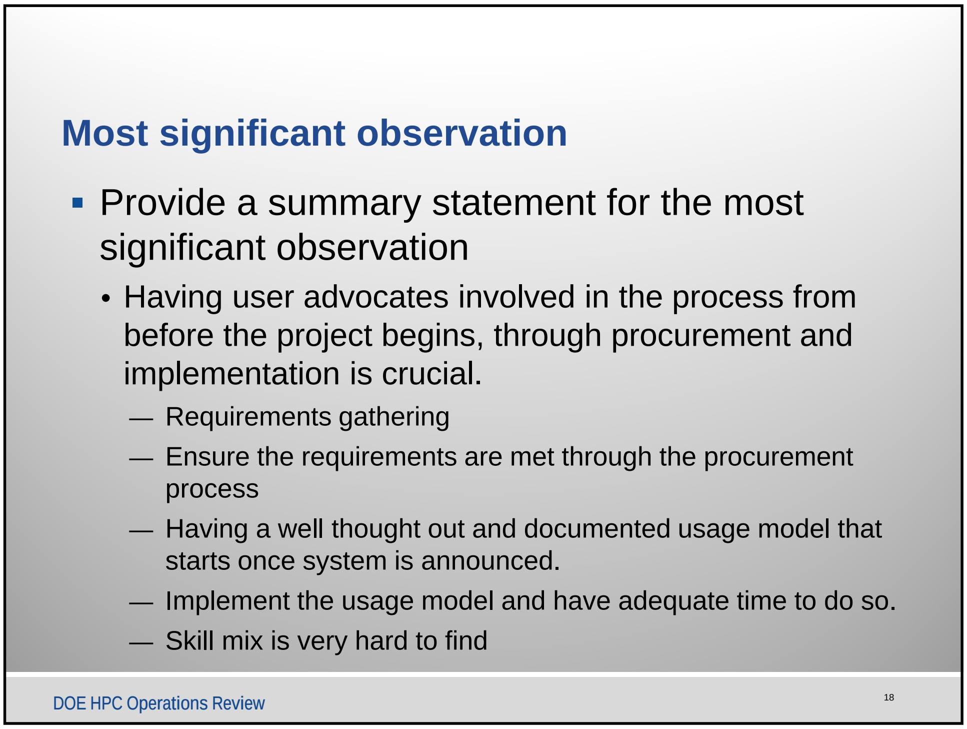
Experiences and lessons learned, cont.

- What were the resiliency experiences?
  - N/A



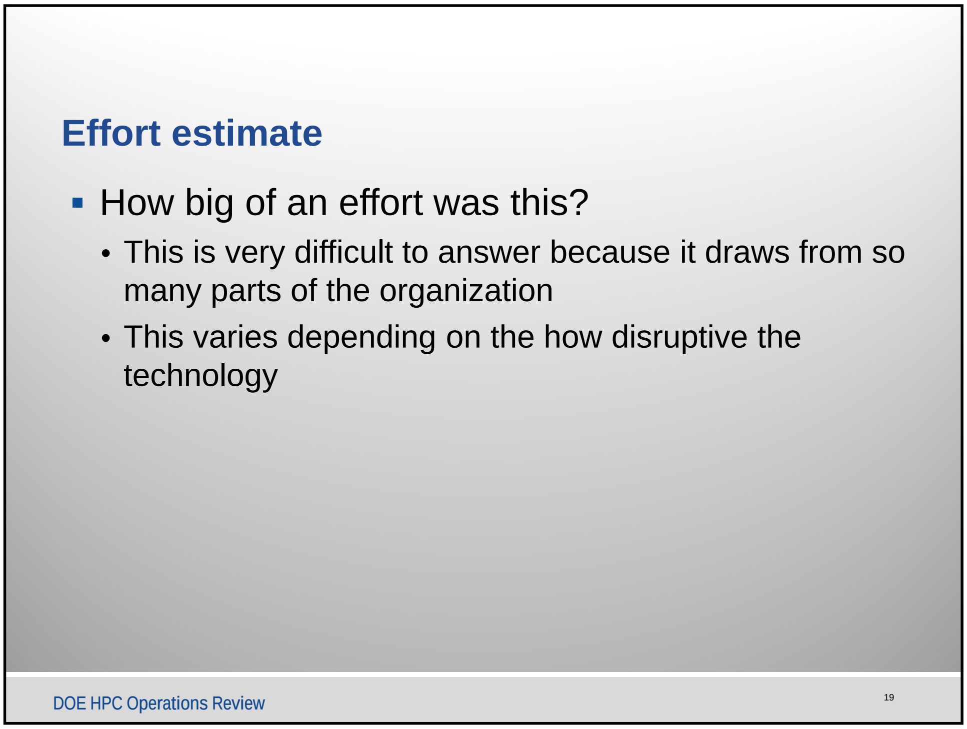
Experiences and lessons learned, cont.

- What was the highest risk? Was it a surprise or expected?
  - The system is not usable by the target users



Most significant observation

- Provide a summary statement for the most significant observation
  - Having user advocates involved in the process from before the project begins, through procurement and implementation is crucial.
    - — Requirements gathering
    - — Ensure the requirements are met through the procurement process
    - — Having a well thought out and documented usage model that starts once system is announced.
    - — Implement the usage model and have adequate time to do so.
    - — Skill mix is very hard to find



Effort estimate

- How big of an effort was this?
  - This is very difficult to answer because it draws from so many parts of the organization
  - This varies depending on the how disruptive the technology

Facility and Utility Planning Impacts, Demand Response Forecasting and Additional External Data Center Issues

DOE HPC Operations Review
San Francisco, November 5-6, 2013

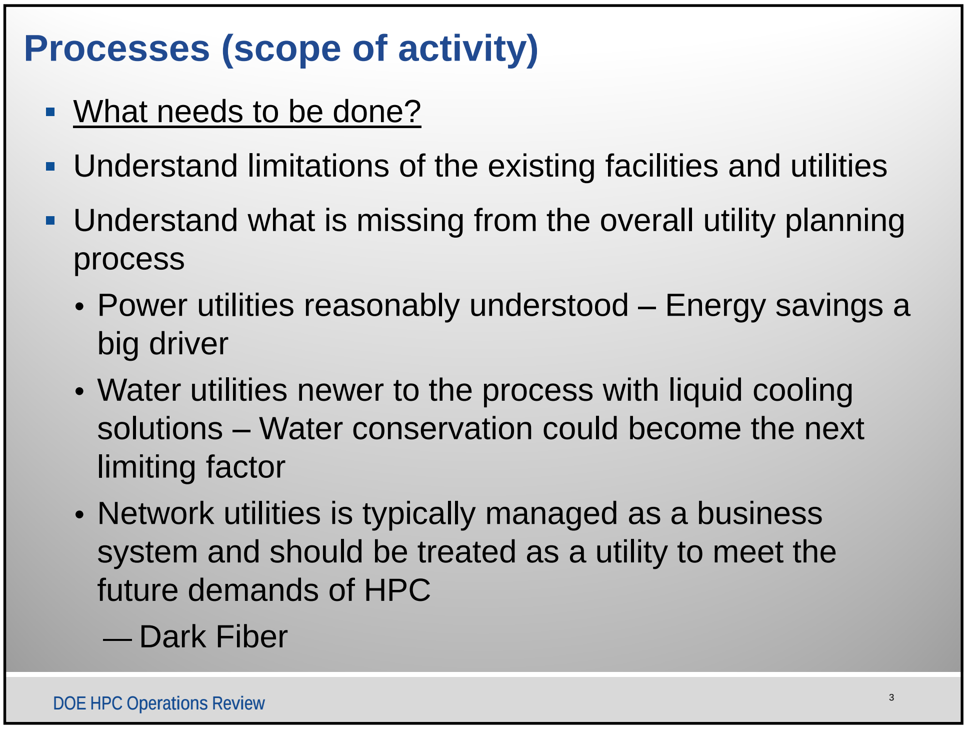
LLNL-PRES-646612

Breakout participants

- Anna Maria Bailey - LLNL*
- Thomas Howe – ANL *
- Hal Armstrong – LANL
- Susan Coghlan - ANL
- Thomas Davis – LBL NERSC
- Brent Draney – LBL NERSC
- Dave Goodwin – DOE
- Steve Hammond – NREL
- Dave Martinez – SNL
- Jim Rogers – ORNL
- Ron Velarde – LANL

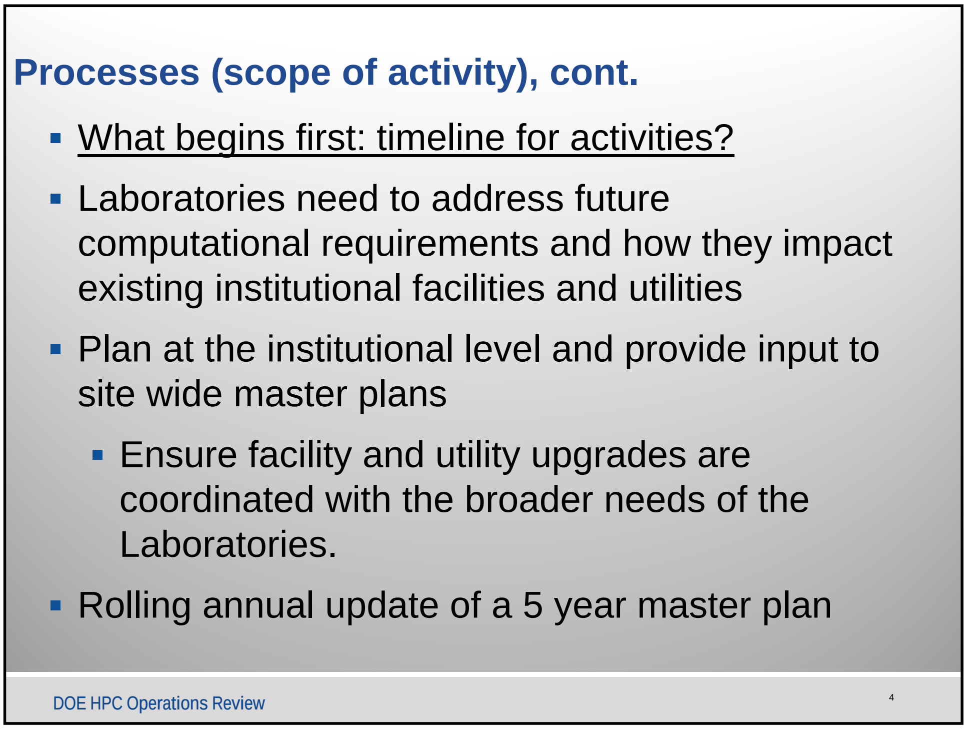
* Denotes breakout session leads

DOE HPC Operations Review 2



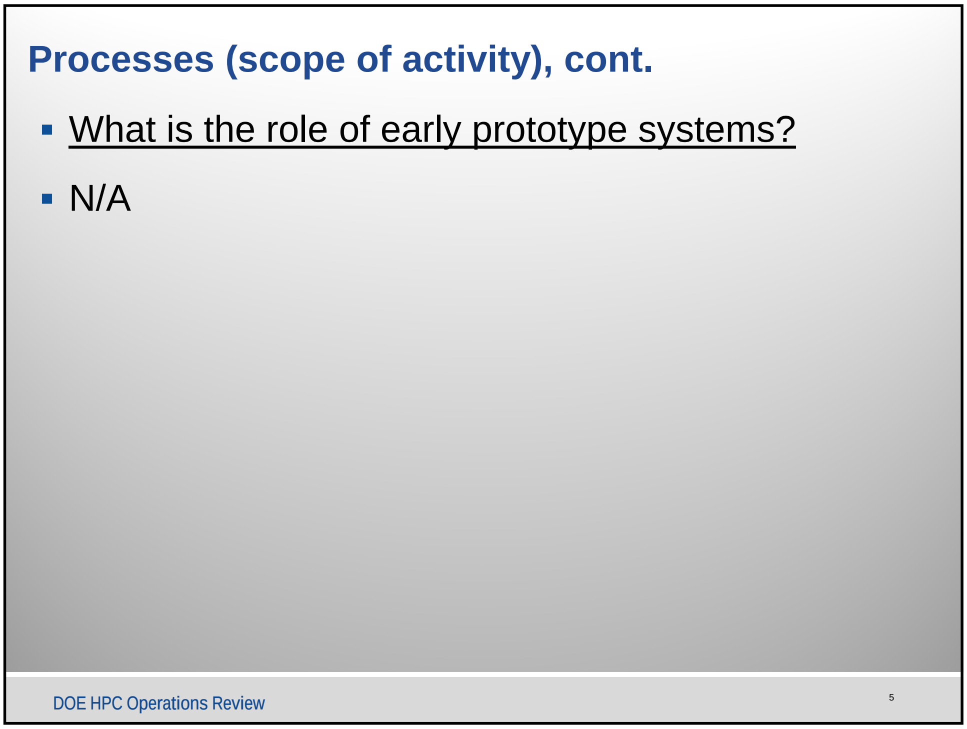
Processes (scope of activity)

- What needs to be done?
- Understand limitations of the existing facilities and utilities
- Understand what is missing from the overall utility planning process
  - Power utilities reasonably understood – Energy savings a big driver
  - Water utilities newer to the process with liquid cooling solutions – Water conservation could become the next limiting factor
  - Network utilities is typically managed as a business system and should be treated as a utility to meet the future demands of HPC
    - — Dark Fiber



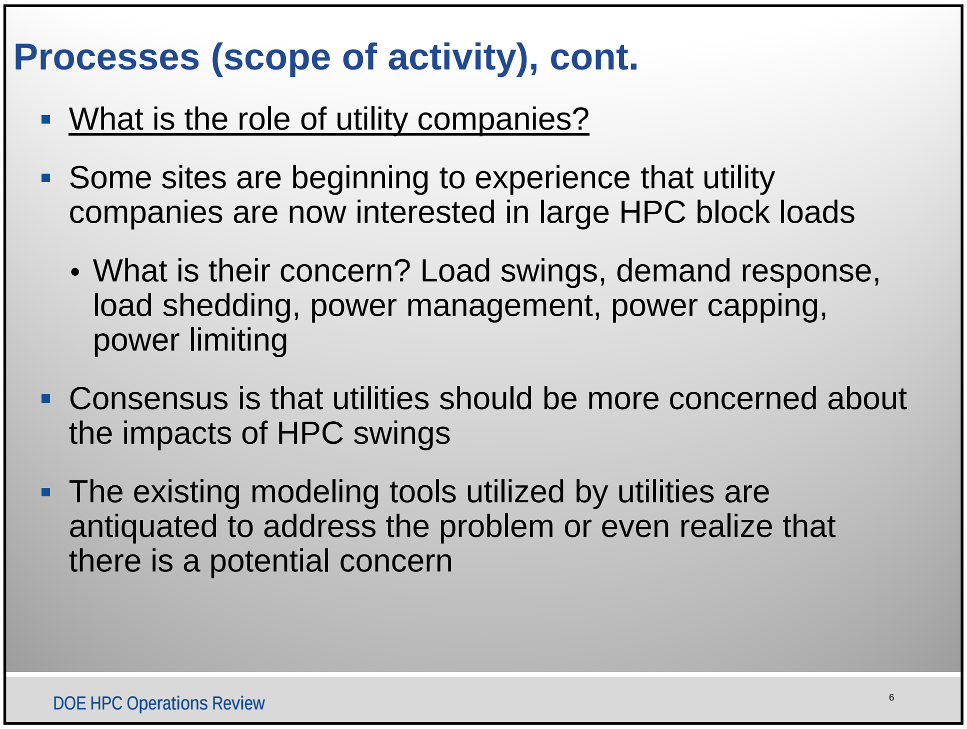
Processes (scope of activity), cont.

- What begins first: timeline for activities?
- Laboratories need to address future computational requirements and how they impact existing institutional facilities and utilities
- Plan at the institutional level and provide input to site wide master plans
  - Ensure facility and utility upgrades are coordinated with the broader needs of the Laboratories.
- Rolling annual update of a 5 year master plan



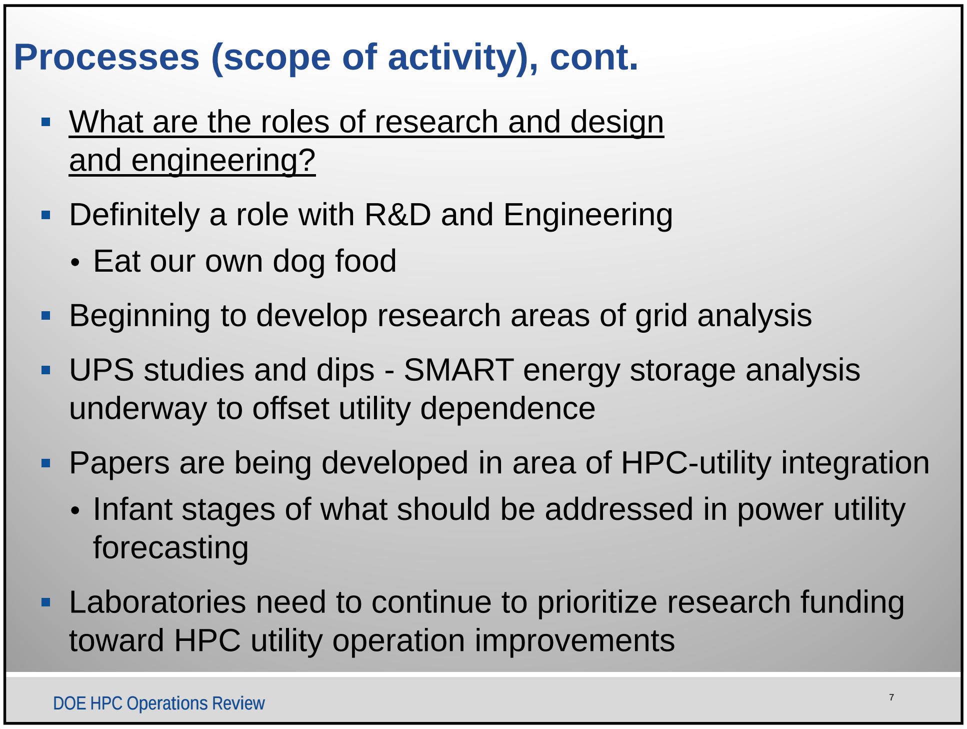
Processes (scope of activity), cont.

- What is the role of early prototype systems?
- N/A



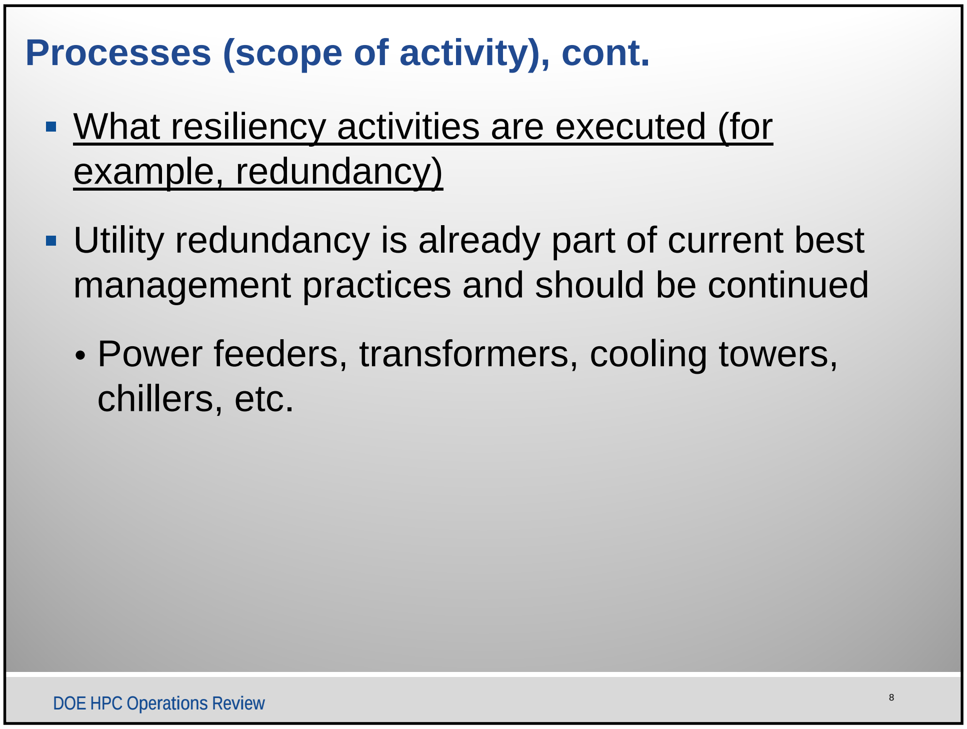
Processes (scope of activity), cont.

- What is the role of utility companies?
- Some sites are beginning to experience that utility companies are now interested in large HPC block loads
  - What is their concern? Load swings, demand response, load shedding, power management, power capping, power limiting
- Consensus is that utilities should be more concerned about the impacts of HPC swings
- The existing modeling tools utilized by utilities are antiquated to address the problem or even realize that there is a potential concern



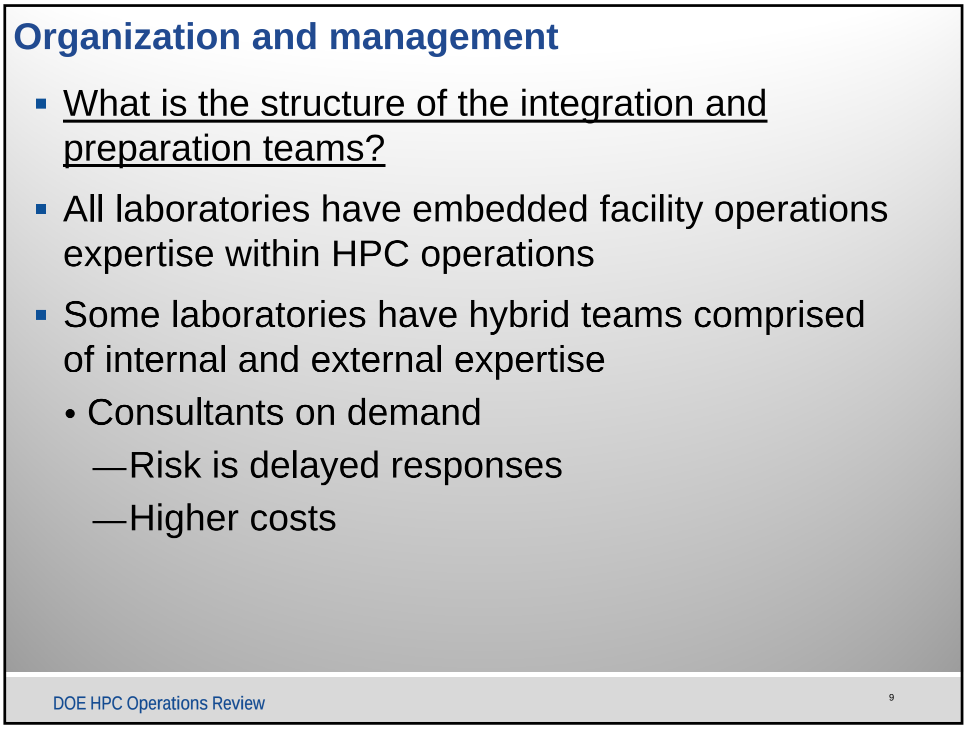
Processes (scope of activity), cont.

- What are the roles of research and design and engineering?
- Definitely a role with R&D and Engineering
  - Eat our own dog food
- Beginning to develop research areas of grid analysis
- UPS studies and dips - SMART energy storage analysis underway to offset utility dependence
- Papers are being developed in area of HPC-utility integration
  - Infant stages of what should be addressed in power utility forecasting
- Laboratories need to continue to prioritize research funding toward HPC utility operation improvements



Processes (scope of activity), cont.

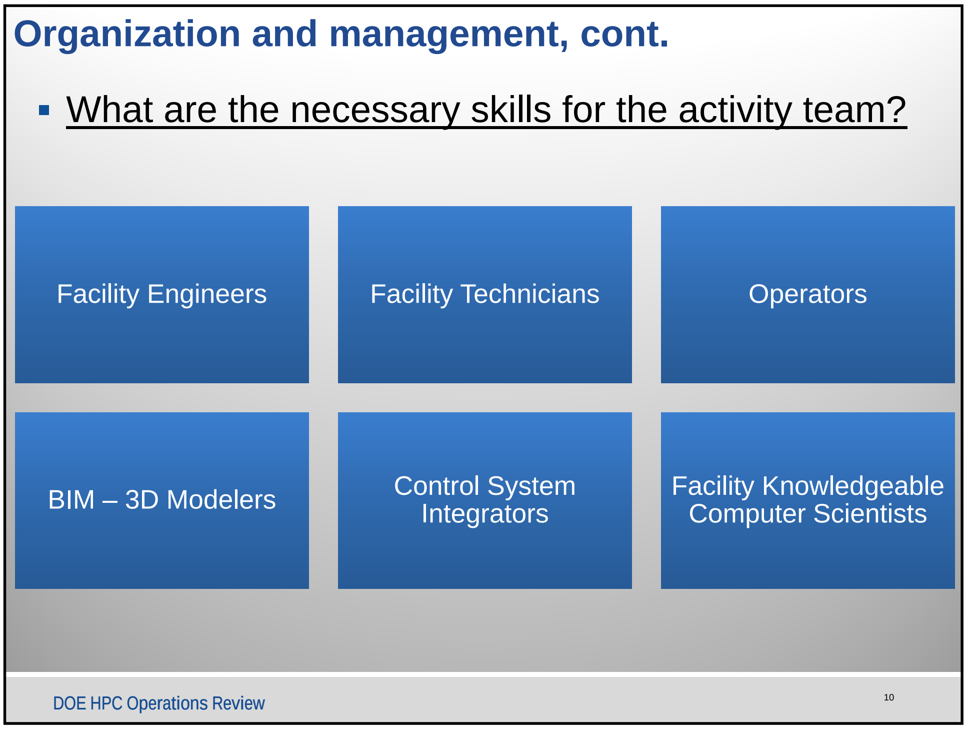
- What resiliency activities are executed (for example, redundancy)
- Utility redundancy is already part of current best management practices and should be continued
  - Power feeders, transformers, cooling towers, chillers, etc.



Organization and management

- What is the structure of the integration and preparation teams?
- All laboratories have embedded facility operations expertise within HPC operations
- Some laboratories have hybrid teams comprised of internal and external expertise
  - Consultants on demand
    - —Risk is delayed responses
    - —Higher costs

DOE HPC Operations Review 9



Organization and management, cont.

- What are the necessary skills for the activity team?

Facility Engineers

Facility Technicians

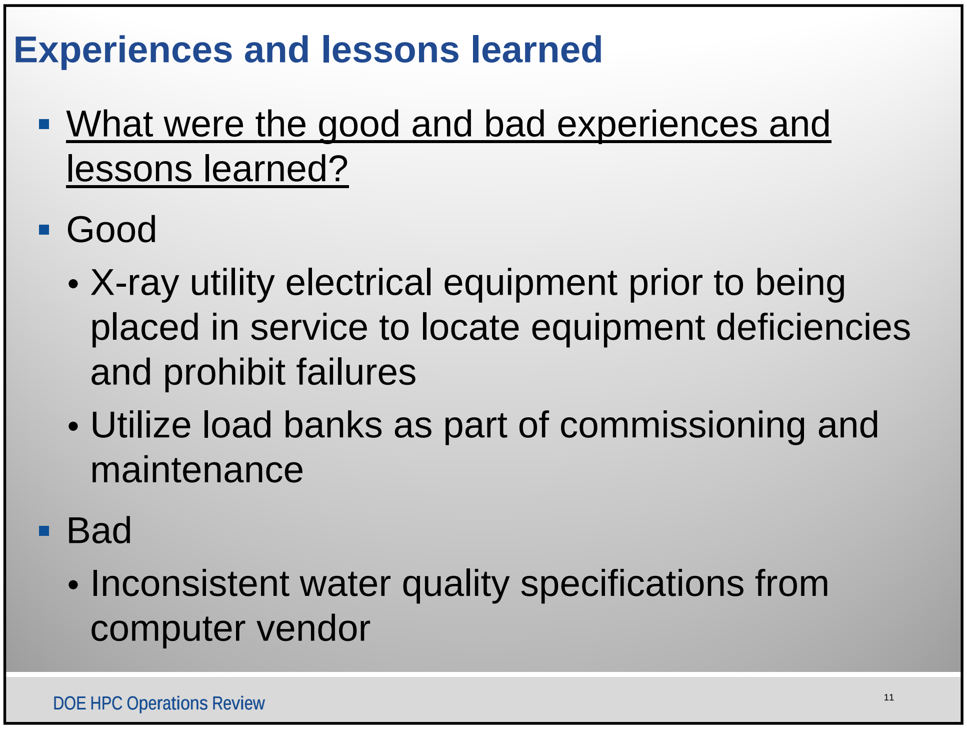
Operators

BIM – 3D Modelers

Control System Integrators

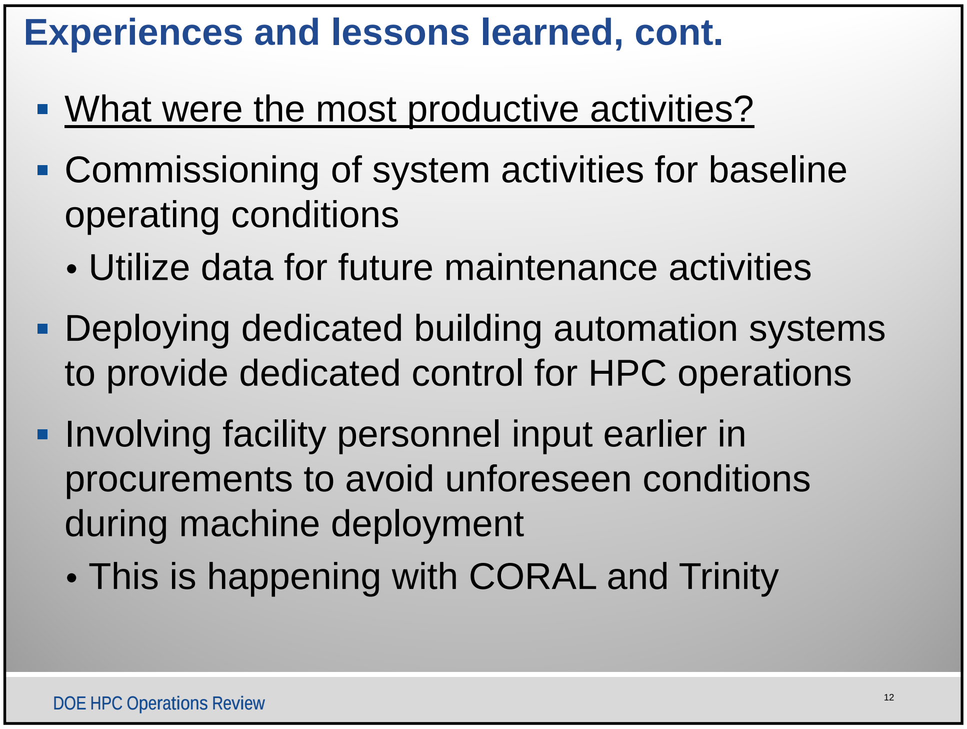
Facility Knowledgeable Computer Scientists

DOE HPC Operations Review 10



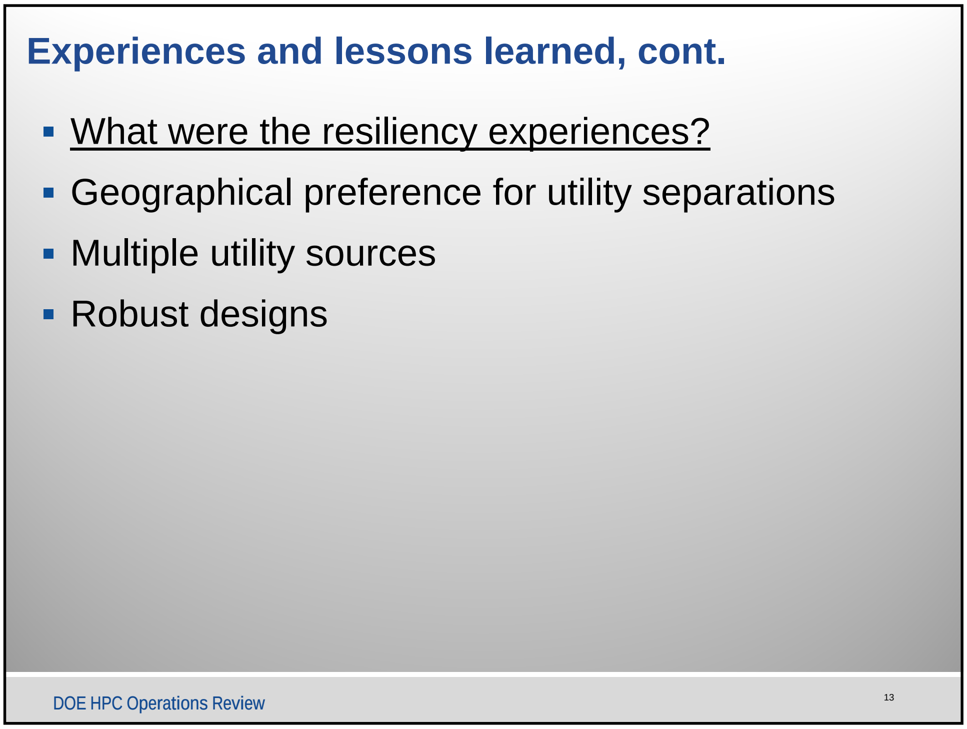
Experiences and lessons learned

- What were the good and bad experiences and lessons learned?
- Good
  - X-ray utility electrical equipment prior to being placed in service to locate equipment deficiencies and prohibit failures
  - Utilize load banks as part of commissioning and maintenance
- Bad
  - Inconsistent water quality specifications from computer vendor



Experiences and lessons learned, cont.

- What were the most productive activities?
- Commissioning of system activities for baseline operating conditions
  - Utilize data for future maintenance activities
- Deploying dedicated building automation systems to provide dedicated control for HPC operations
- Involving facility personnel input earlier in procurements to avoid unforeseen conditions during machine deployment
  - This is happening with CORAL and Trinity

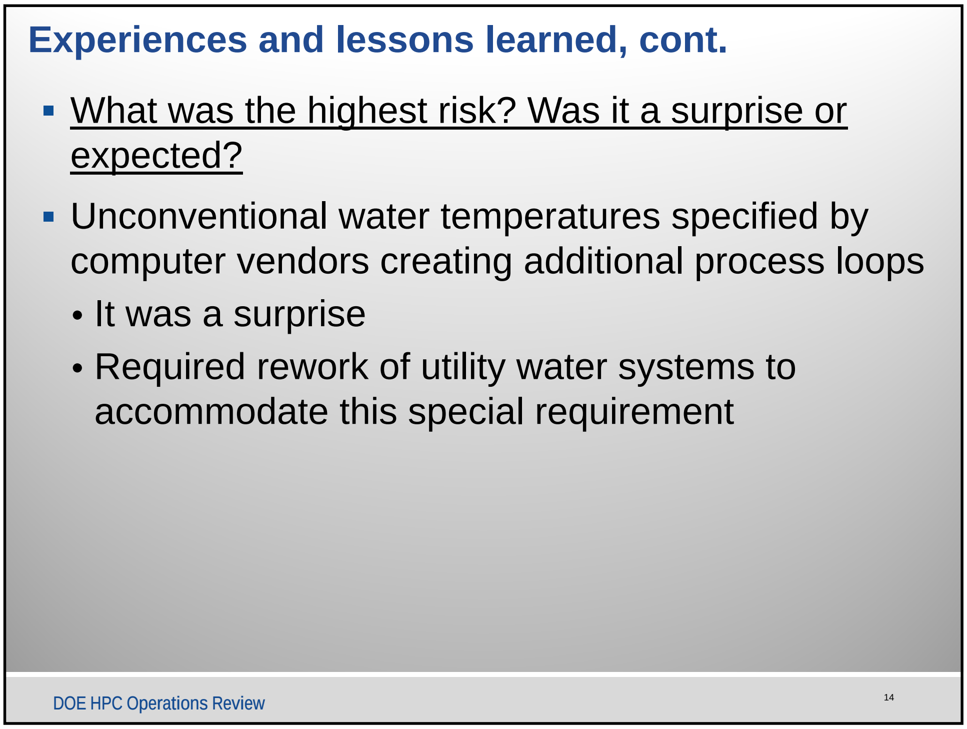


Experiences and lessons learned, cont.

- What were the resiliency experiences?
- Geographical preference for utility separations
- Multiple utility sources
- Robust designs

DOE HPC Operations Review

13

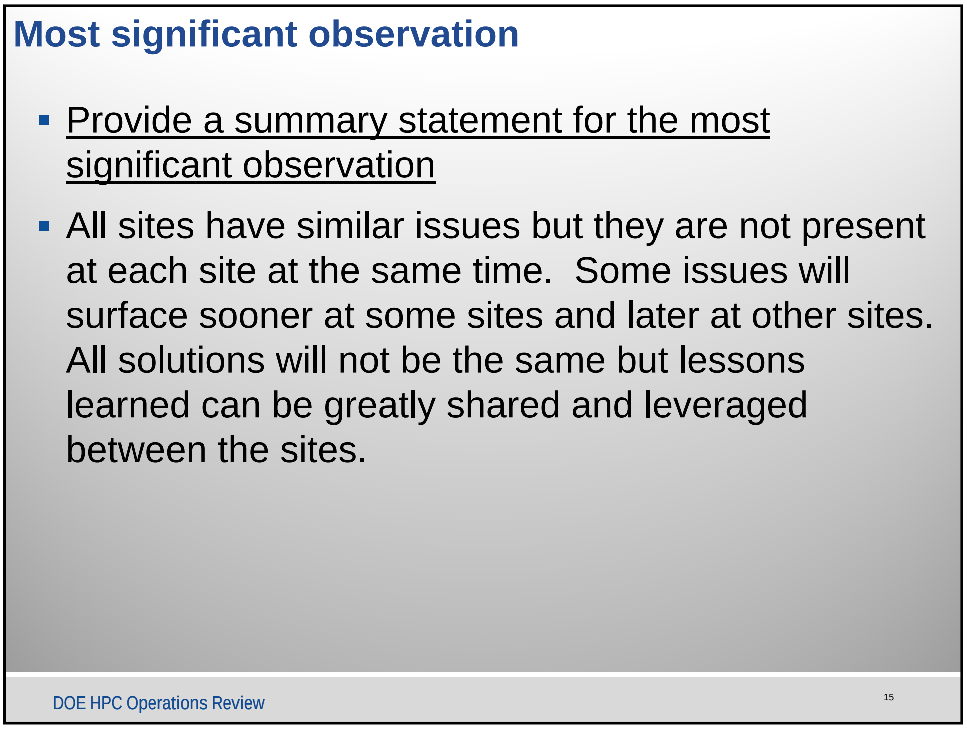


Experiences and lessons learned, cont.

- What was the highest risk? Was it a surprise or expected?
- Unconventional water temperatures specified by computer vendors creating additional process loops
  - It was a surprise
  - Required rework of utility water systems to accommodate this special requirement

DOE HPC Operations Review

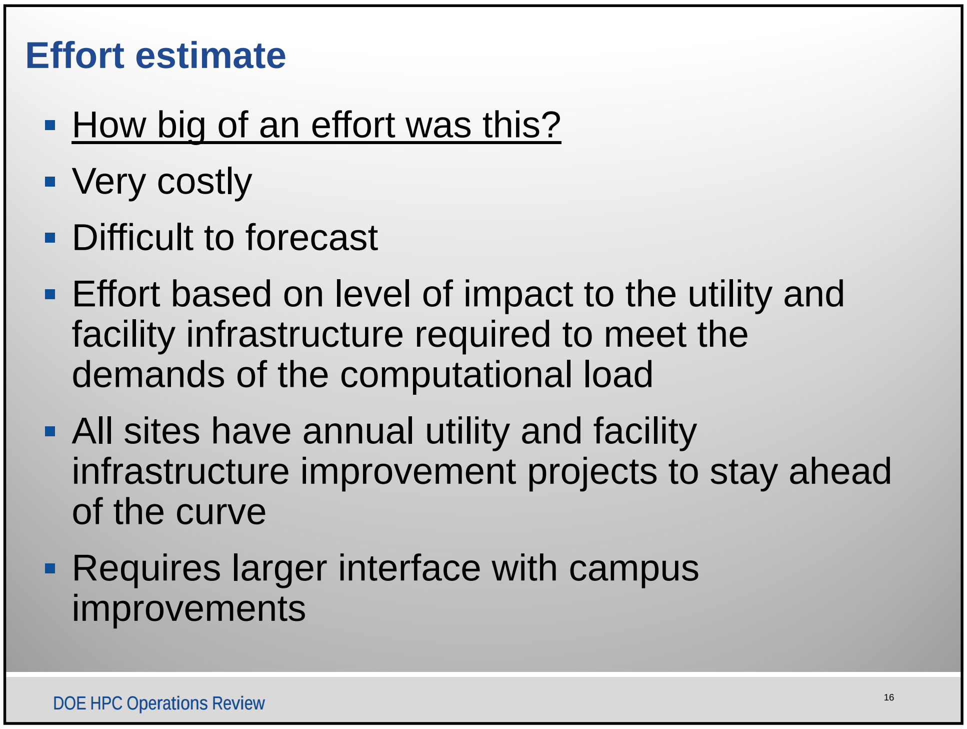
14



Most significant observation

- <u>Provide a summary statement for the most significant observation</u>
- All sites have similar issues but they are not present at each site at the same time. Some issues will surface sooner at some sites and later at other sites. All solutions will not be the same but lessons learned can be greatly shared and leveraged between the sites.

DOE HPC Operations Review 15



Effort estimate

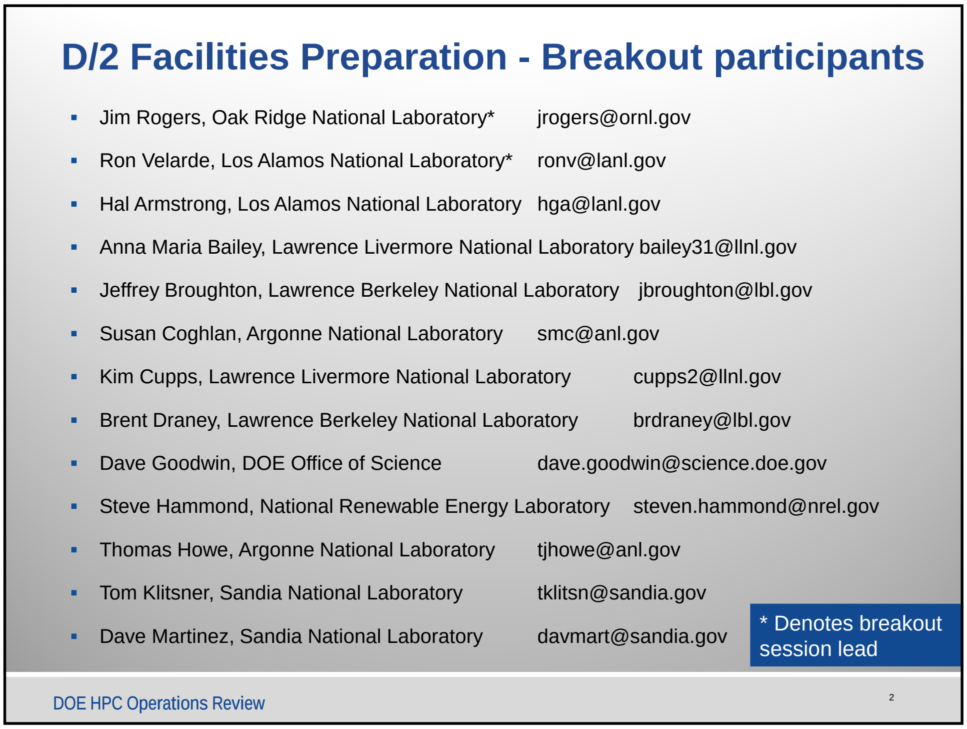
- <u>How big of an effort was this?</u>
- Very costly
- Difficult to forecast
- Effort based on level of impact to the utility and facility infrastructure required to meet the demands of the computational load
- All sites have annual utility and facility infrastructure improvement projects to stay ahead of the curve
- Requires larger interface with campus improvements

DOE HPC Operations Review 16

D/2: Facilities Preparation Inside the Building

DOE HPC Operations Review
San Francisco, November 5-6, 2013

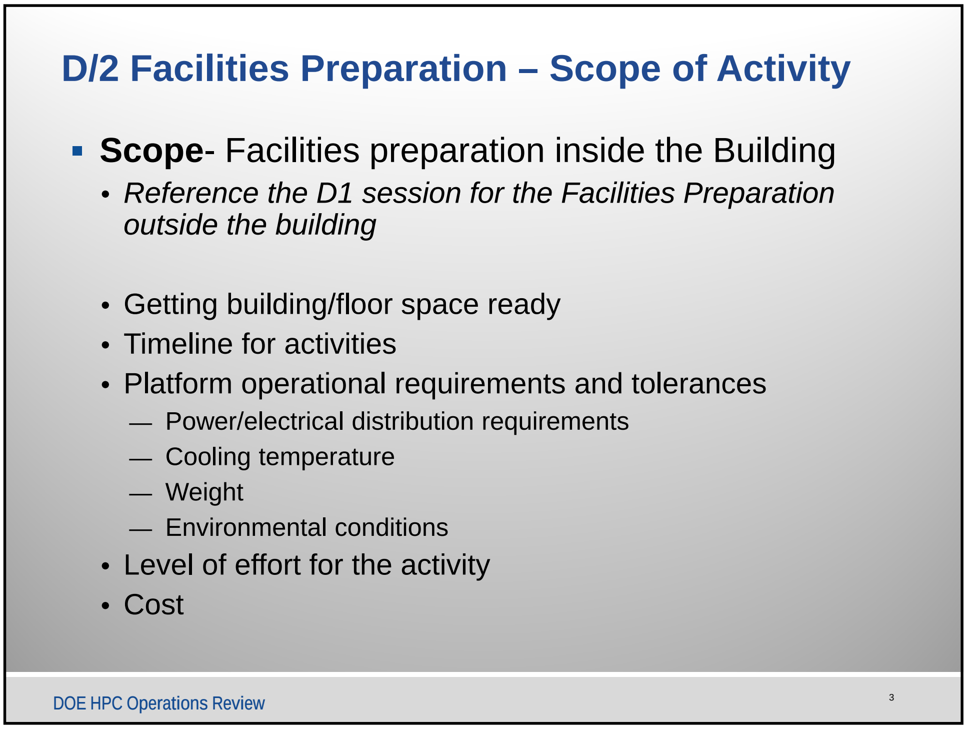
LLNL-PRES-646612



D/2 Facilities Preparation - Breakout participants

Jim Rogers, Oak Ridge National Laboratory*	jrogers@ornl.gov
Ron Velarde, Los Alamos National Laboratory*	ronv@lanl.gov
Hal Armstrong, Los Alamos National Laboratory	hga@lanl.gov
Anna Maria Bailey, Lawrence Livermore National Laboratory	bailey31@llnl.gov
Jeffrey Broughton, Lawrence Berkeley National Laboratory	jbroughton@lbl.gov
Susan Coghlan, Argonne National Laboratory	smc@anl.gov
Kim Cupps, Lawrence Livermore National Laboratory	cupps2@llnl.gov
Brent Draney, Lawrence Berkeley National Laboratory	brdraney@lbl.gov
Dave Goodwin, DOE Office of Science	dave.goodwin@science.doe.gov
Steve Hammond, National Renewable Energy Laboratory	steven.hammond@nrel.gov
Thomas Howe, Argonne National Laboratory	tjhowe@anl.gov
Tom Klitsner, Sandia National Laboratory	tklitsn@sandia.gov
Dave Martinez, Sandia National Laboratory	davmart@sandia.gov

* Denotes breakout session lead

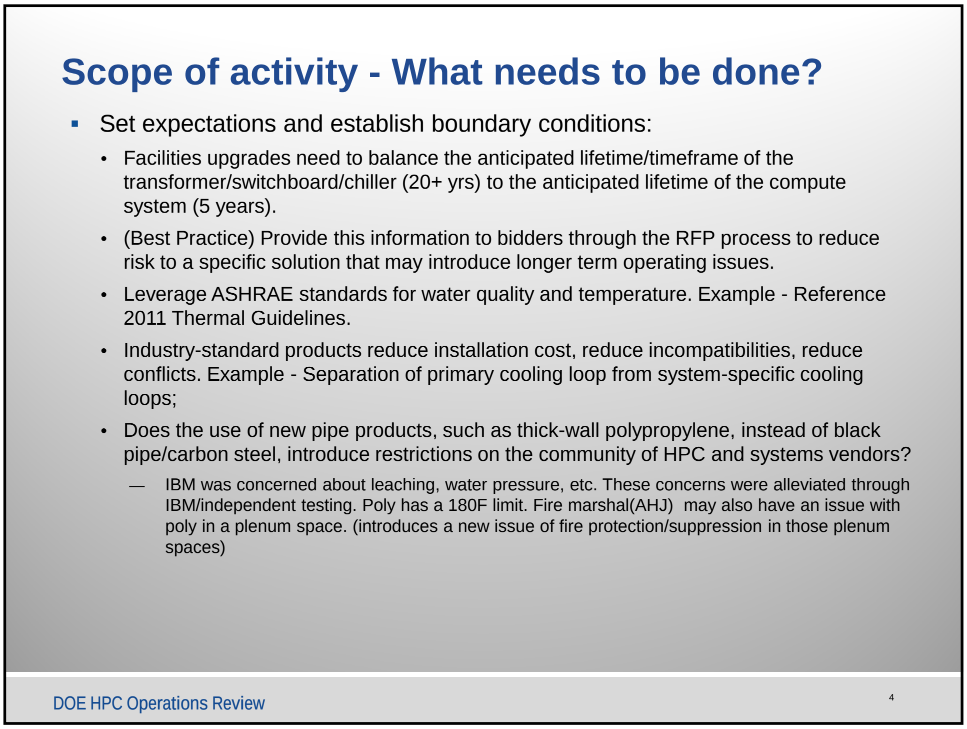


D/2 Facilities Preparation – Scope of Activity

- **Scope**- Facilities preparation inside the Building
  - *Reference the D1 session for the Facilities Preparation outside the building*

  - Getting building/floor space ready
  - Timeline for activities
  - Platform operational requirements and tolerances
    - — Power/electrical distribution requirements
    - — Cooling temperature
    - — Weight
    - — Environmental conditions
  - Level of effort for the activity
  - Cost

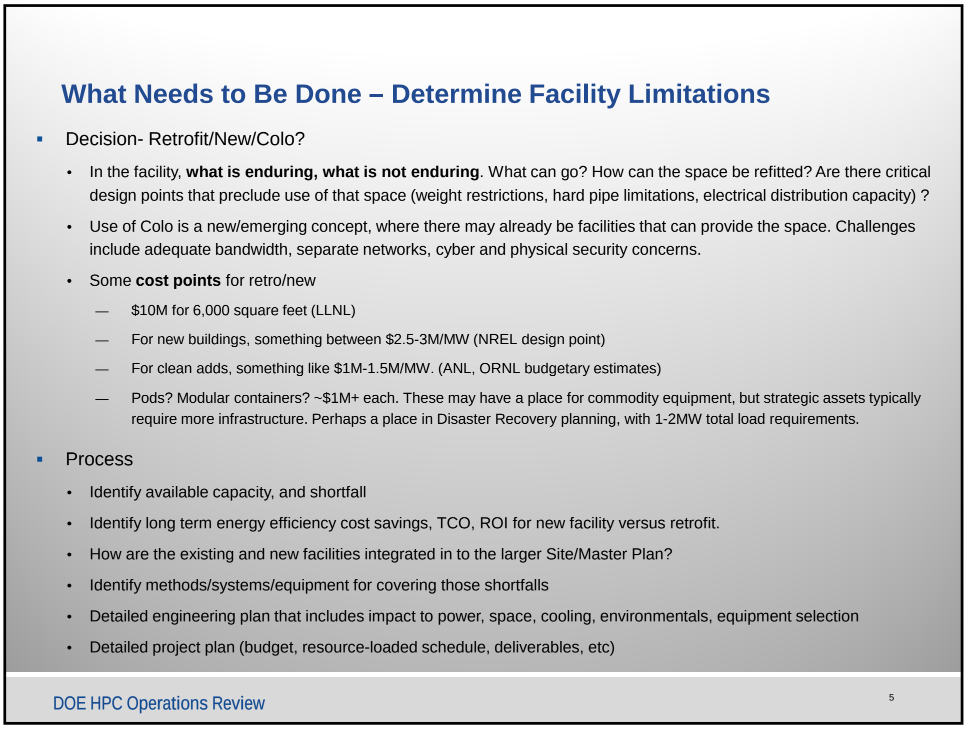
DOE HPC Operations Review 3



Scope of activity - What needs to be done?

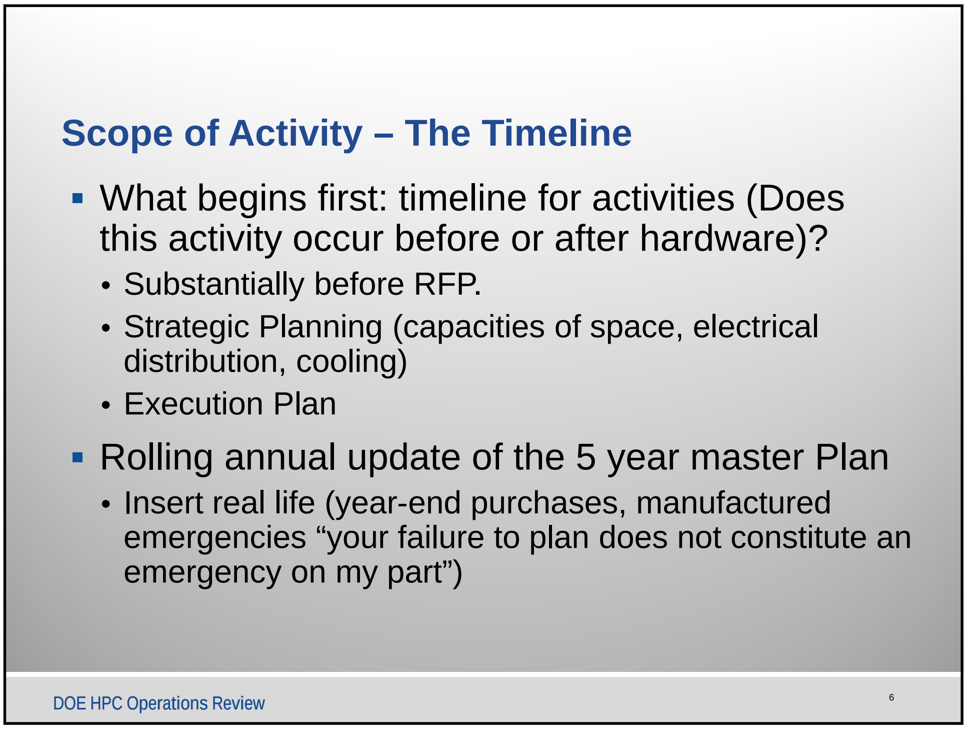
- Set expectations and establish boundary conditions:
  - Facilities upgrades need to balance the anticipated lifetime/timeframe of the transformer/switchboard/chiller (20+ yrs) to the anticipated lifetime of the compute system (5 years).
  - (Best Practice) Provide this information to bidders through the RFP process to reduce risk to a specific solution that may introduce longer term operating issues.
  - Leverage ASHRAE standards for water quality and temperature. Example - Reference 2011 Thermal Guidelines.
  - Industry-standard products reduce installation cost, reduce incompatibilities, reduce conflicts. Example - Separation of primary cooling loop from system-specific cooling loops;
  - Does the use of new pipe products, such as thick-wall polypropylene, instead of black pipe/carbon steel, introduce restrictions on the community of HPC and systems vendors?
    - — IBM was concerned about leaching, water pressure, etc. These concerns were alleviated through IBM/independent testing. Poly has a 180F limit. Fire marshal(AHJ) may also have an issue with poly in a plenum space. (introduces a new issue of fire protection/suppression in those plenum spaces)

DOE HPC Operations Review 4



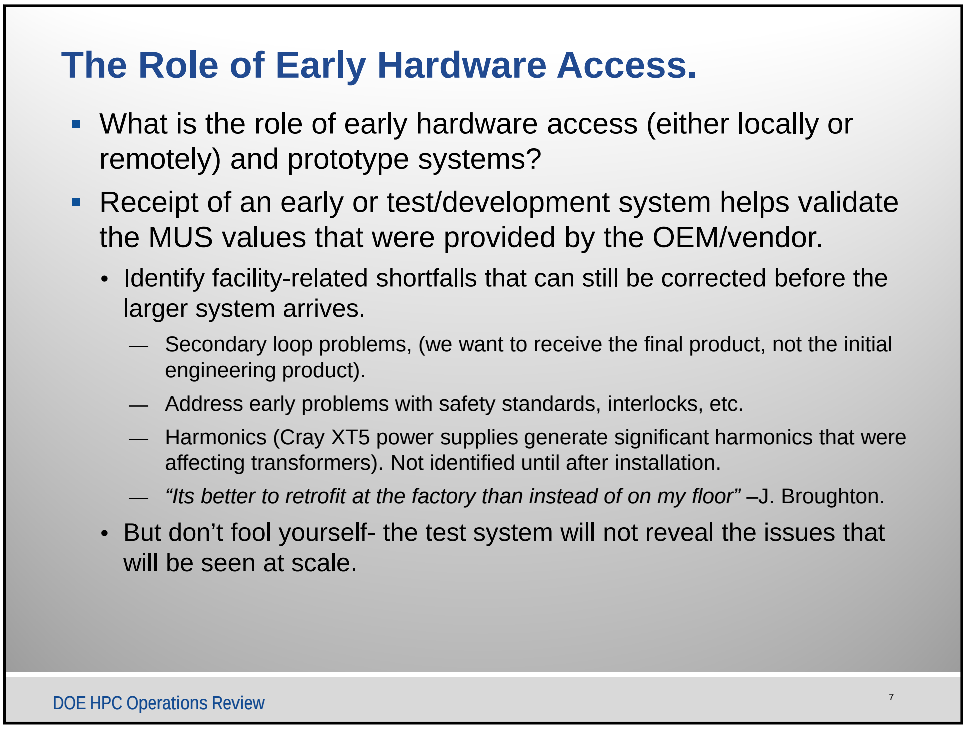
What Needs to Be Done – Determine Facility Limitations

- Decision- Retrofit/New/Colo?
  - In the facility, **what is enduring, what is not enduring**. What can go? How can the space be refitted? Are there critical design points that preclude use of that space (weight restrictions, hard pipe limitations, electrical distribution capacity) ?
  - Use of Colo is a new/emerging concept, where there may already be facilities that can provide the space. Challenges include adequate bandwidth, separate networks, cyber and physical security concerns.
  - Some **cost points** for retro/new
    - — $10M for 6,000 square feet (LLNL)
    - — For new buildings, something between $2.5-3M/MW (NREL design point)
    - — For clean adds, something like $1M-1.5M/MW. (ANL, ORNL budgetary estimates)
    - — Pods? Modular containers? ~$1M+ each. These may have a place for commodity equipment, but strategic assets typically require more infrastructure. Perhaps a place in Disaster Recovery planning, with 1-2MW total load requirements.
- Process
  - Identify available capacity, and shortfall
  - Identify long term energy efficiency cost savings, TCO, ROI for new facility versus retrofit.
  - How are the existing and new facilities integrated in to the larger Site/Master Plan?
  - Identify methods/systems/equipment for covering those shortfalls
  - Detailed engineering plan that includes impact to power, space, cooling, environmentals, equipment selection
  - Detailed project plan (budget, resource-loaded schedule, deliverables, etc)



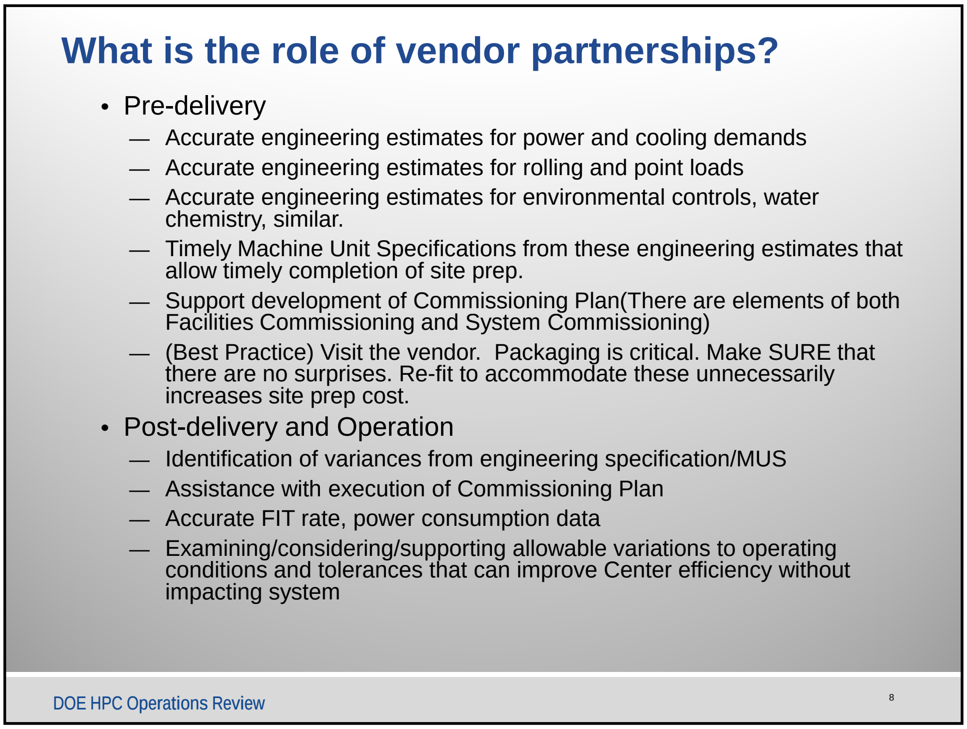
Scope of Activity – The Timeline

- What begins first: timeline for activities (Does this activity occur before or after hardware)?
  - Substantially before RFP.
  - Strategic Planning (capacities of space, electrical distribution, cooling)
  - Execution Plan
- Rolling annual update of the 5 year master Plan
  - Insert real life (year-end purchases, manufactured emergencies “your failure to plan does not constitute an emergency on my part”)



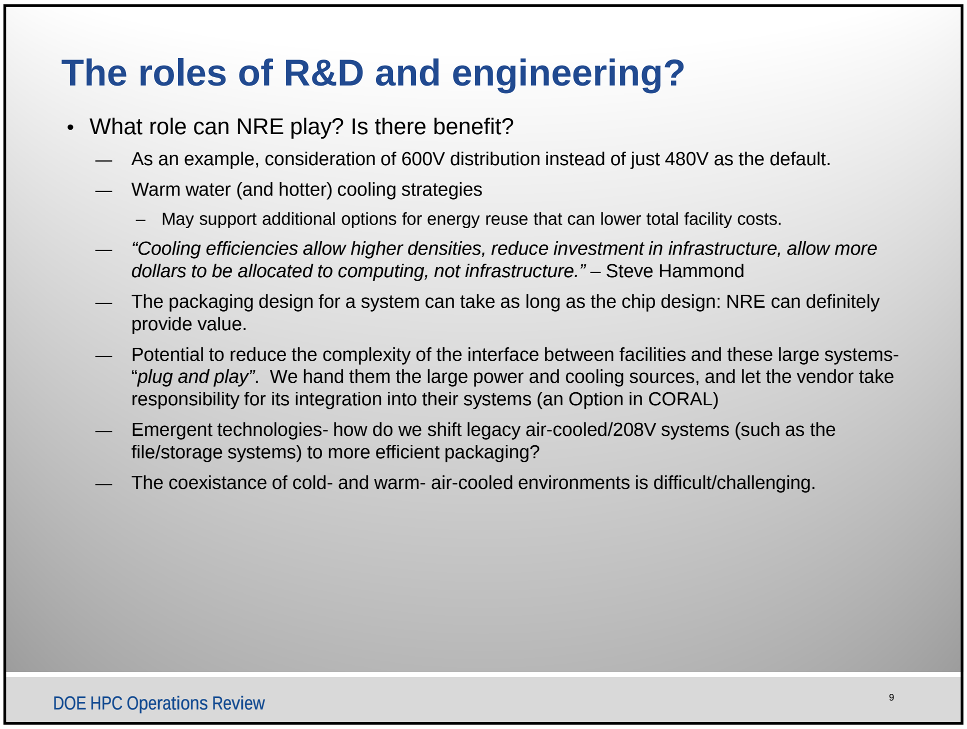
The Role of Early Hardware Access.

- What is the role of early hardware access (either locally or remotely) and prototype systems?
- Receipt of an early or test/development system helps validate the MUS values that were provided by the OEM/vendor.
  - Identify facility-related shortfalls that can still be corrected before the larger system arrives.
    - — Secondary loop problems, (we want to receive the final product, not the initial engineering product).
    - — Address early problems with safety standards, interlocks, etc.
    - — Harmonics (Cray XT5 power supplies generate significant harmonics that were affecting transformers). Not identified until after installation.
    - — *"Its better to retrofit at the factory than instead of on my floor"* –J. Broughton.
  - But don't fool yourself- the test system will not reveal the issues that will be seen at scale.



What is the role of vendor partnerships?

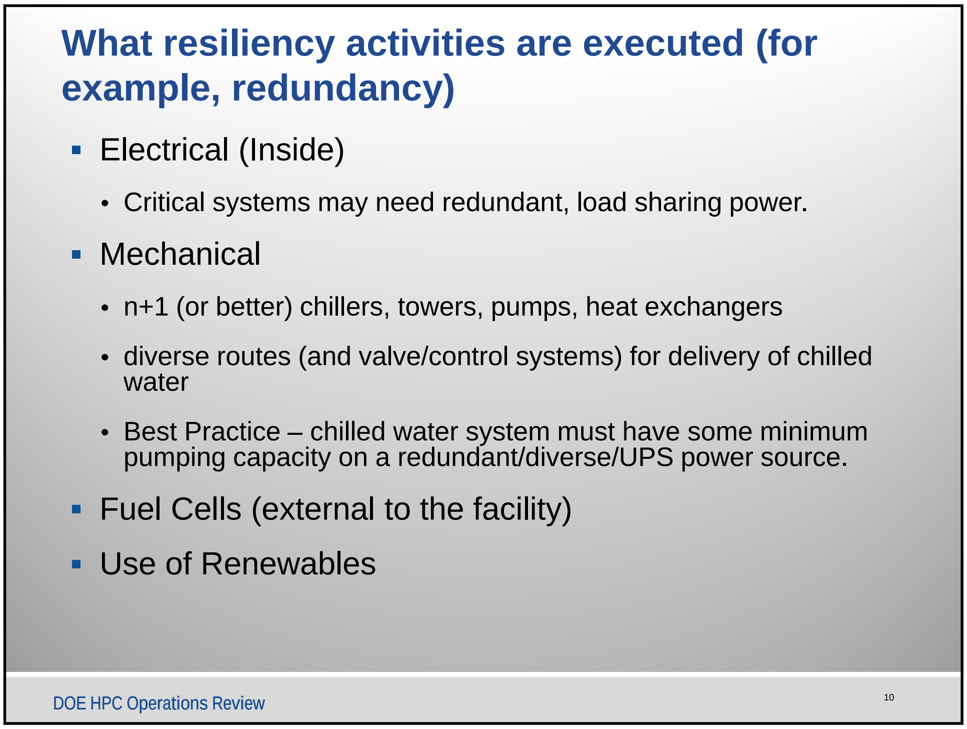
- Pre-delivery
  - — Accurate engineering estimates for power and cooling demands
  - — Accurate engineering estimates for rolling and point loads
  - — Accurate engineering estimates for environmental controls, water chemistry, similar.
  - — Timely Machine Unit Specifications from these engineering estimates that allow timely completion of site prep.
  - — Support development of Commissioning Plan(There are elements of both Facilities Commissioning and System Commissioning)
  - — (Best Practice) Visit the vendor. Packaging is critical. Make SURE that there are no surprises. Re-fit to accommodate these unnecessarily increases site prep cost.
- Post-delivery and Operation
  - — Identification of variances from engineering specification/MUS
  - — Assistance with execution of Commissioning Plan
  - — Accurate FIT rate, power consumption data
  - — Examining/considering/supporting allowable variations to operating conditions and tolerances that can improve Center efficiency without impacting system



The roles of R&D and engineering?

- What role can NRE play? Is there benefit?
  - As an example, consideration of 600V distribution instead of just 480V as the default.
  - Warm water (and hotter) cooling strategies
    - May support additional options for energy reuse that can lower total facility costs.
  - *"Cooling efficiencies allow higher densities, reduce investment in infrastructure, allow more dollars to be allocated to computing, not infrastructure."* – Steve Hammond
  - The packaging design for a system can take as long as the chip design: NRE can definitely provide value.
  - Potential to reduce the complexity of the interface between facilities and these large systems- "*plug and play*". We hand them the large power and cooling sources, and let the vendor take responsibility for its integration into their systems (an Option in CORAL)
  - Emergent technologies- how do we shift legacy air-cooled/208V systems (such as the file/storage systems) to more efficient packaging?
  - The coexistance of cold- and warm- air-cooled environments is difficult/challenging.

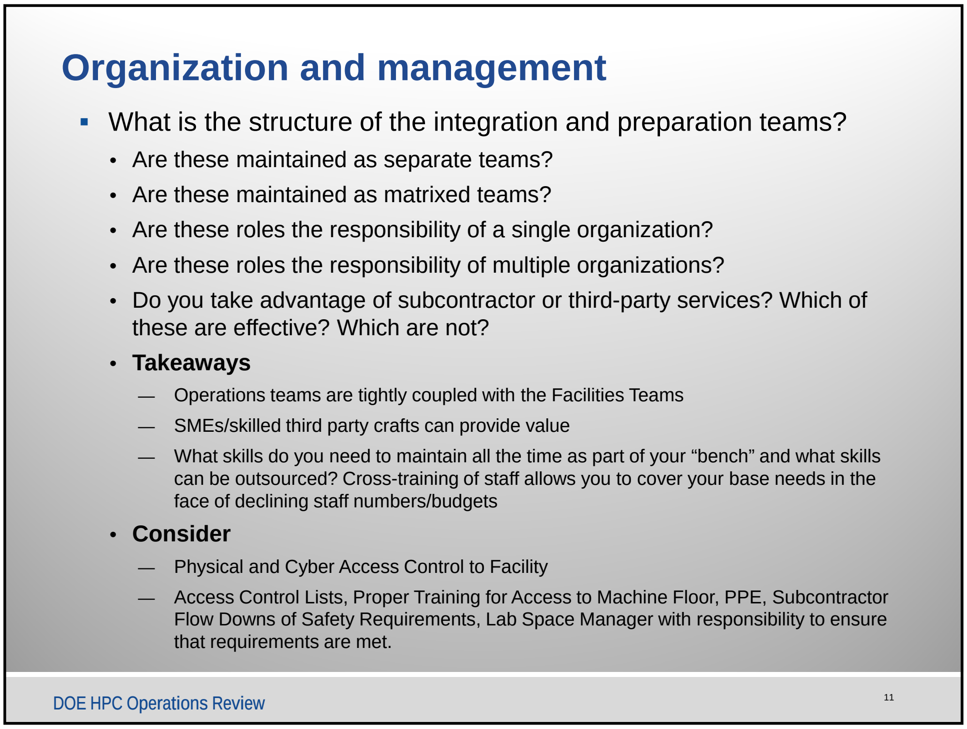
DOE HPC Operations Review 9



What resiliency activities are executed (for example, redundancy)

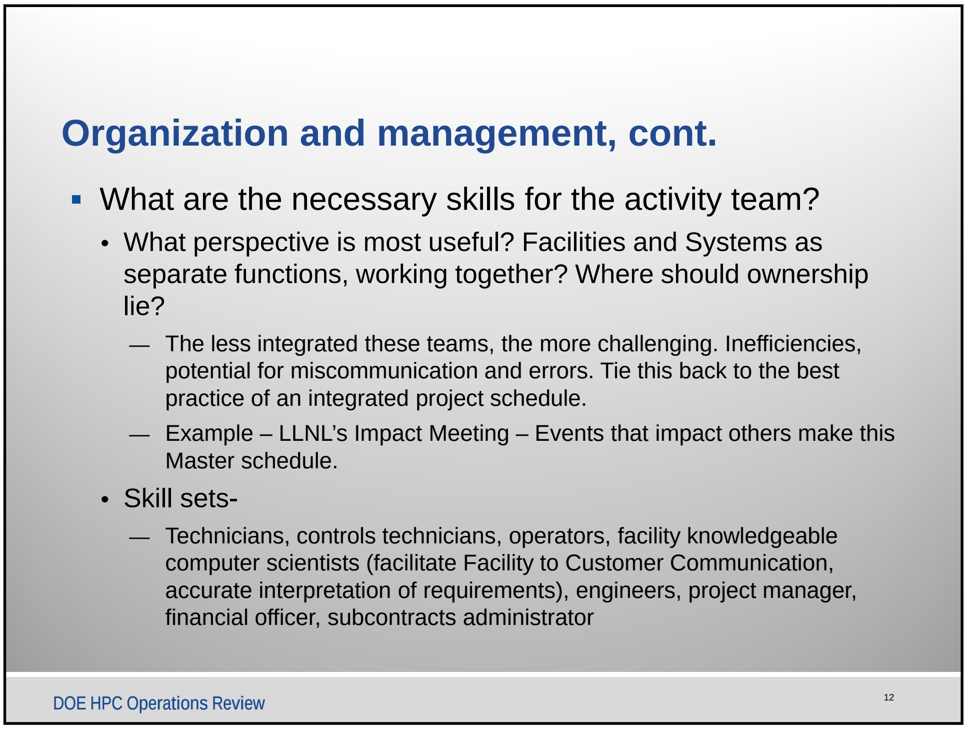
- Electrical (Inside)
  - Critical systems may need redundant, load sharing power.
- Mechanical
  - n+1 (or better) chillers, towers, pumps, heat exchangers
  - diverse routes (and valve/control systems) for delivery of chilled water
  - Best Practice – chilled water system must have some minimum pumping capacity on a redundant/diverse/UPS power source.
- Fuel Cells (external to the facility)
- Use of Renewables

DOE HPC Operations Review 10



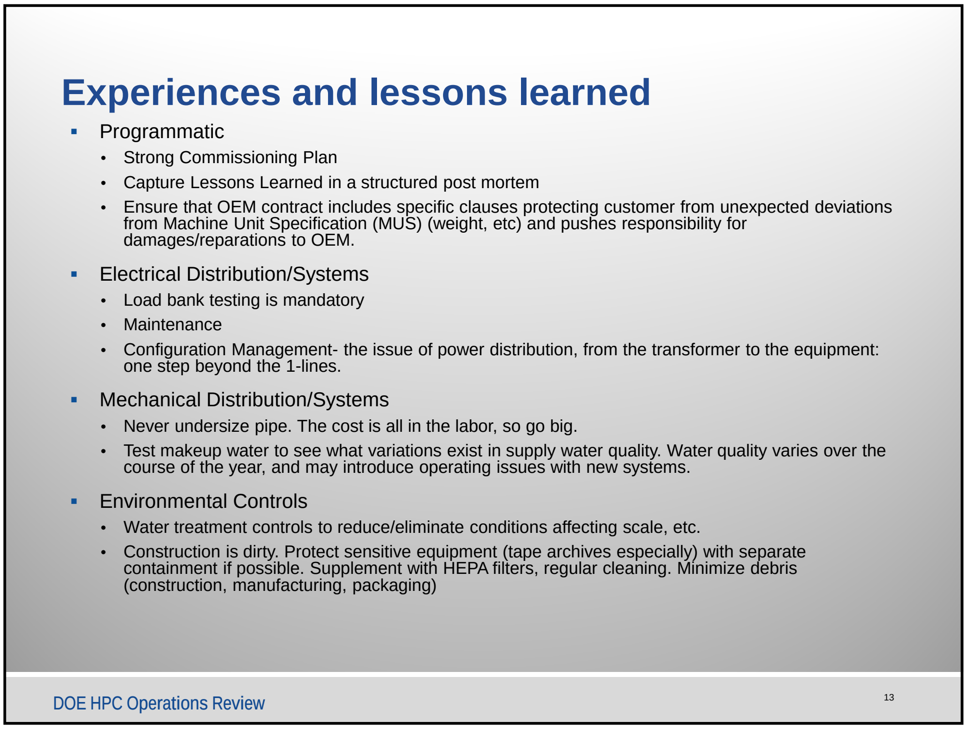
Organization and management

- What is the structure of the integration and preparation teams?
  - Are these maintained as separate teams?
  - Are these maintained as matrixed teams?
  - Are these roles the responsibility of a single organization?
  - Are these roles the responsibility of multiple organizations?
  - Do you take advantage of subcontractor or third-party services? Which of these are effective? Which are not?
  - **Takeaways**
    - Operations teams are tightly coupled with the Facilities Teams
    - SMEs/skilled third party crafts can provide value
    - What skills do you need to maintain all the time as part of your “bench” and what skills can be outsourced? Cross-training of staff allows you to cover your base needs in the face of declining staff numbers/budgets
  - **Consider**
    - Physical and Cyber Access Control to Facility
    - Access Control Lists, Proper Training for Access to Machine Floor, PPE, Subcontractor Flow Downs of Safety Requirements, Lab Space Manager with responsibility to ensure that requirements are met.



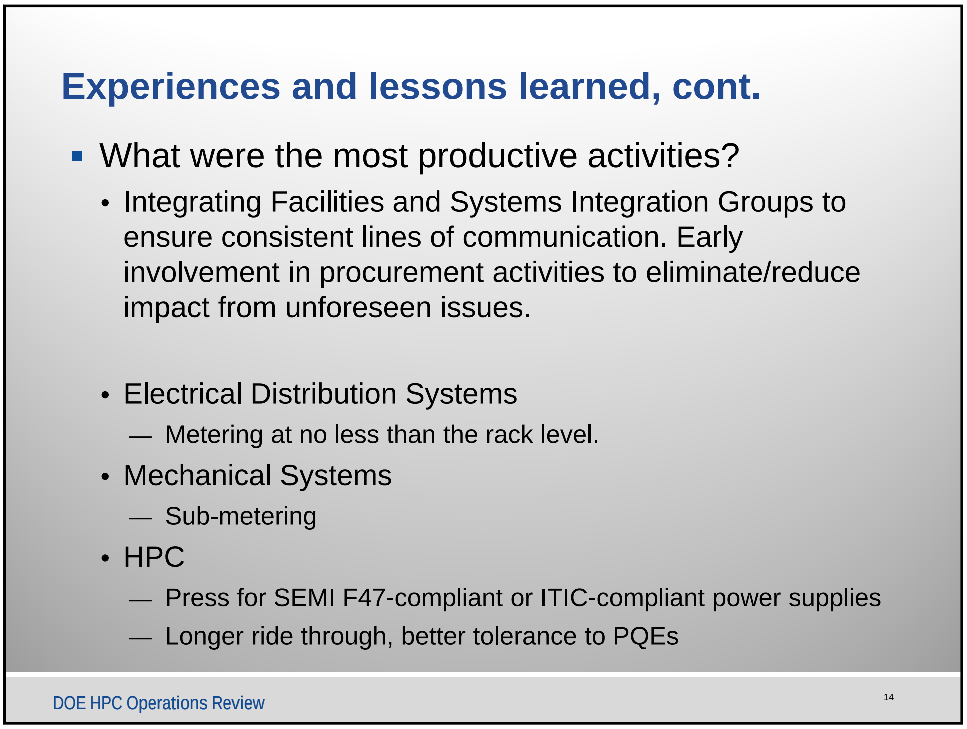
Organization and management, cont.

- What are the necessary skills for the activity team?
  - What perspective is most useful? Facilities and Systems as separate functions, working together? Where should ownership lie?
    - The less integrated these teams, the more challenging. Inefficiencies, potential for miscommunication and errors. Tie this back to the best practice of an integrated project schedule.
    - Example – LLNL’s Impact Meeting – Events that impact others make this Master schedule.
  - Skill sets-
    - Technicians, controls technicians, operators, facility knowledgeable computer scientists (facilitate Facility to Customer Communication, accurate interpretation of requirements), engineers, project manager, financial officer, subcontracts administrator



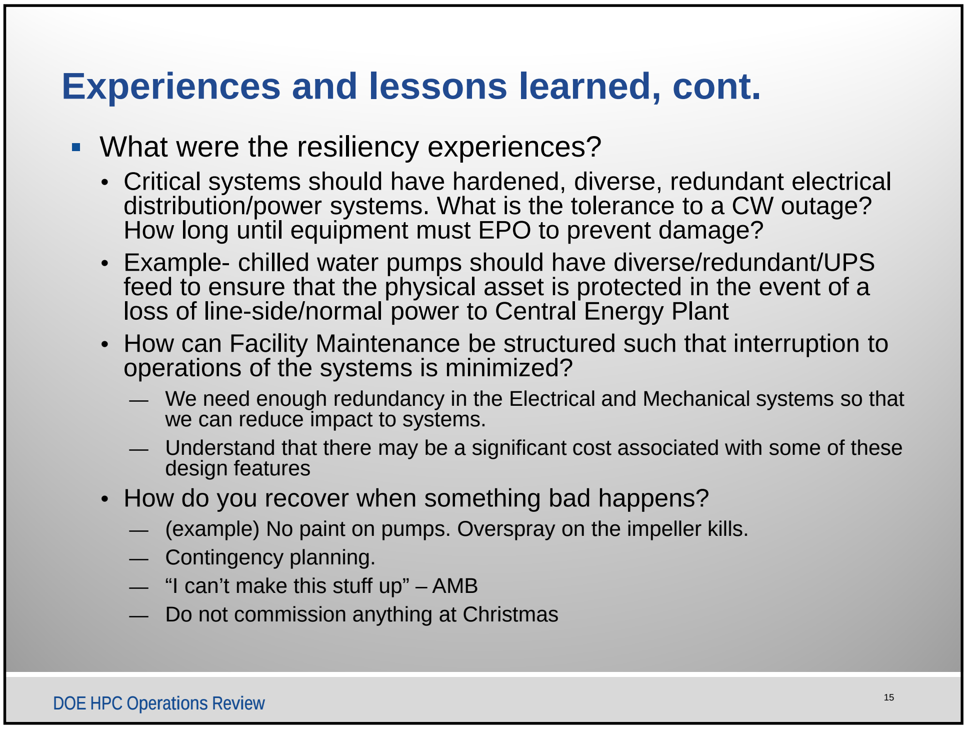
Experiences and lessons learned

- Programmatic
  - Strong Commissioning Plan
  - Capture Lessons Learned in a structured post mortem
  - Ensure that OEM contract includes specific clauses protecting customer from unexpected deviations from Machine Unit Specification (MUS) (weight, etc) and pushes responsibility for damages/reparations to OEM.
- Electrical Distribution/Systems
  - Load bank testing is mandatory
  - Maintenance
  - Configuration Management- the issue of power distribution, from the transformer to the equipment: one step beyond the 1-lines.
- Mechanical Distribution/Systems
  - Never undersize pipe. The cost is all in the labor, so go big.
  - Test makeup water to see what variations exist in supply water quality. Water quality varies over the course of the year, and may introduce operating issues with new systems.
- Environmental Controls
  - Water treatment controls to reduce/eliminate conditions affecting scale, etc.
  - Construction is dirty. Protect sensitive equipment (tape archives especially) with separate containment if possible. Supplement with HEPA filters, regular cleaning. Minimize debris (construction, manufacturing, packaging)



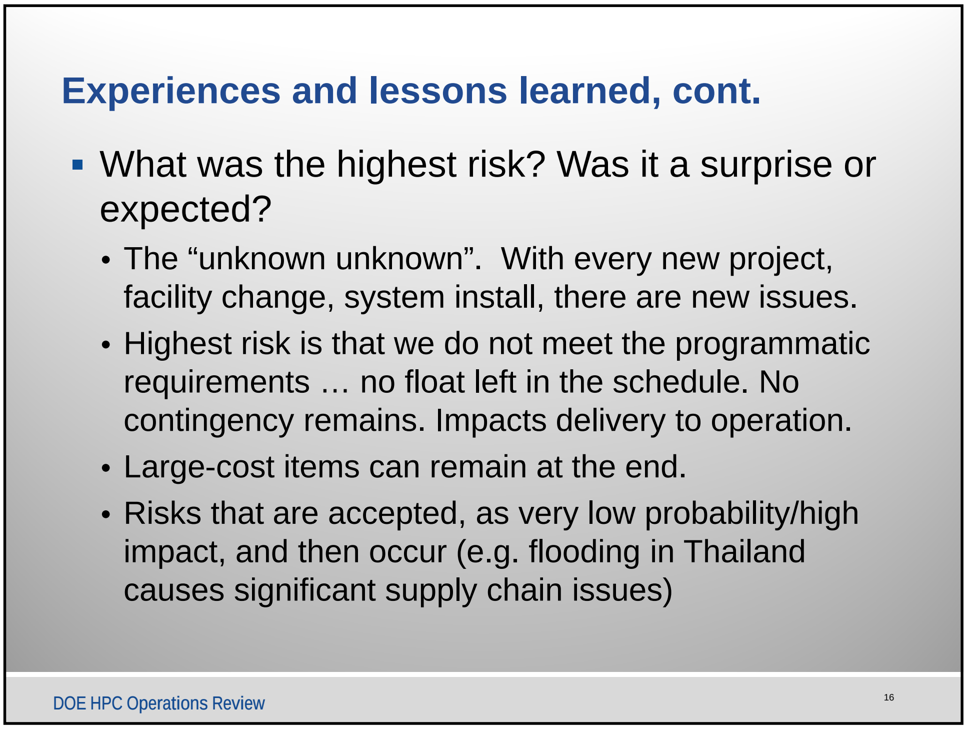
Experiences and lessons learned, cont.

- What were the most productive activities?
  - Integrating Facilities and Systems Integration Groups to ensure consistent lines of communication. Early involvement in procurement activities to eliminate/reduce impact from unforeseen issues.
  - Electrical Distribution Systems
    - Metering at no less than the rack level.
  - Mechanical Systems
    - Sub-metering
  - HPC
    - Press for SEMI F47-compliant or ITIC-compliant power supplies
    - Longer ride through, better tolerance to PQEs



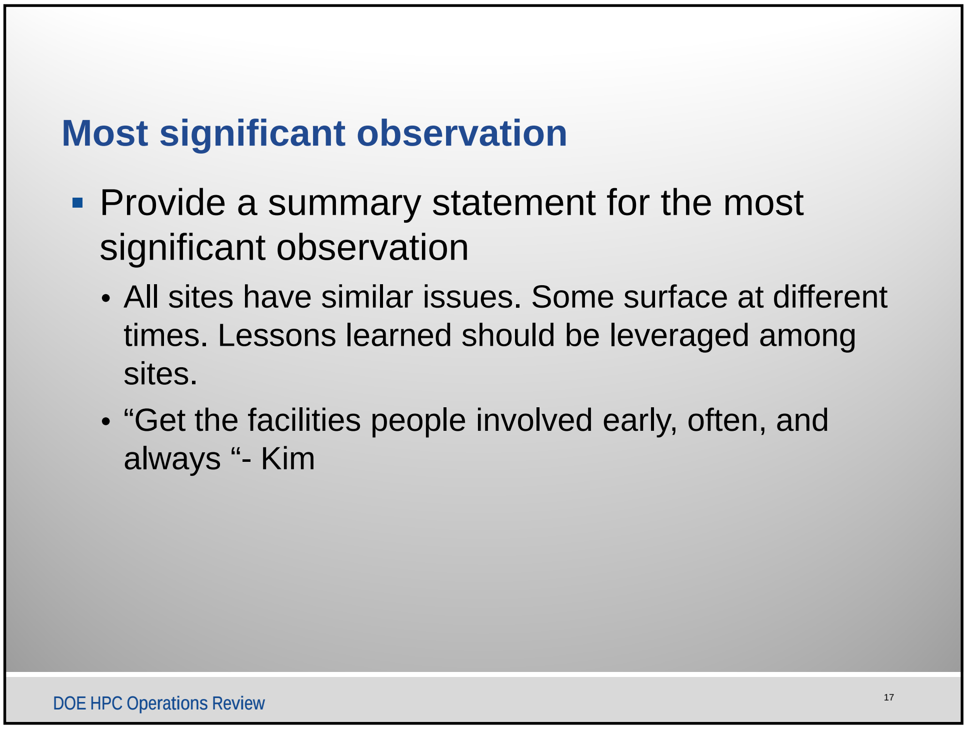
Experiences and lessons learned, cont.

- What were the resiliency experiences?
  - Critical systems should have hardened, diverse, redundant electrical distribution/power systems. What is the tolerance to a CW outage? How long until equipment must EPO to prevent damage?
  - Example- chilled water pumps should have diverse/redundant/UPS feed to ensure that the physical asset is protected in the event of a loss of line-side/normal power to Central Energy Plant
  - How can Facility Maintenance be structured such that interruption to operations of the systems is minimized?
    - We need enough redundancy in the Electrical and Mechanical systems so that we can reduce impact to systems.
    - Understand that there may be a significant cost associated with some of these design features
  - How do you recover when something bad happens?
    - (example) No paint on pumps. Overspray on the impeller kills.
    - Contingency planning.
    - “I can’t make this stuff up” – AMB
    - Do not commission anything at Christmas



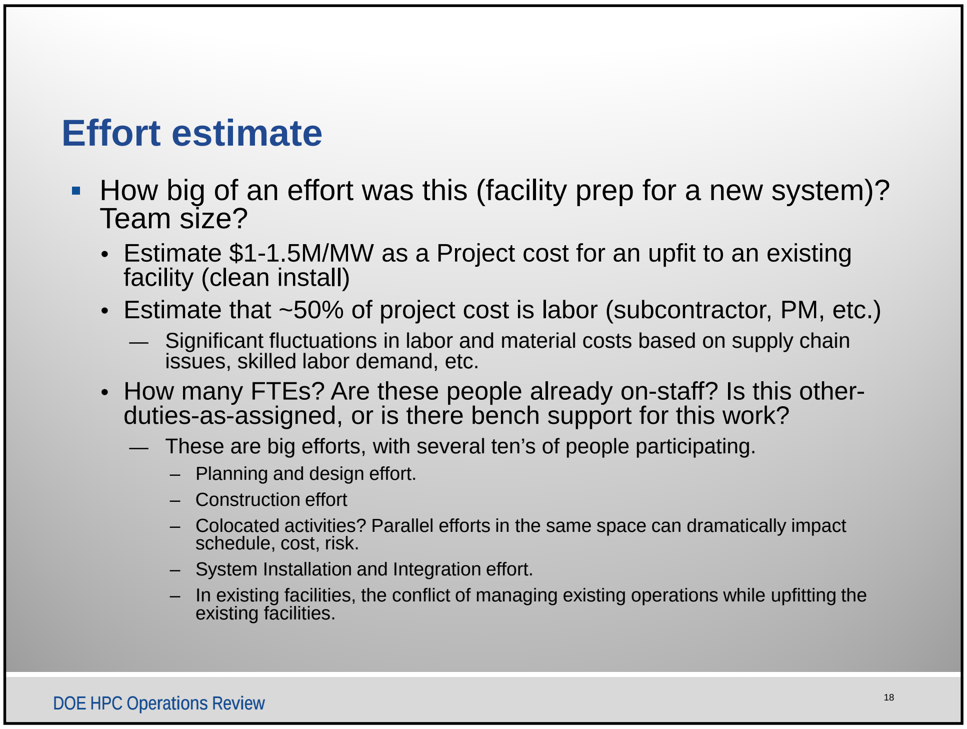
Experiences and lessons learned, cont.

- What was the highest risk? Was it a surprise or expected?
  - The “unknown unknown”. With every new project, facility change, system install, there are new issues.
  - Highest risk is that we do not meet the programmatic requirements … no float left in the schedule. No contingency remains. Impacts delivery to operation.
  - Large-cost items can remain at the end.
  - Risks that are accepted, as very low probability/high impact, and then occur (e.g. flooding in Thailand causes significant supply chain issues)



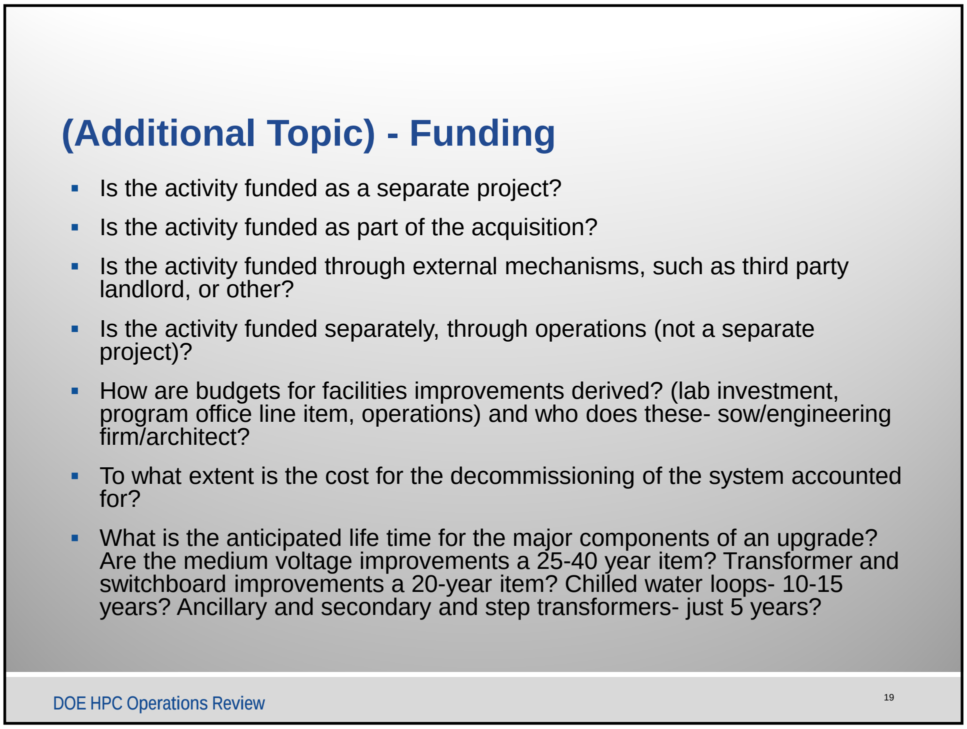
Most significant observation

- Provide a summary statement for the most significant observation
  - All sites have similar issues. Some surface at different times. Lessons learned should be leveraged among sites.
  - “Get the facilities people involved early, often, and always “- Kim



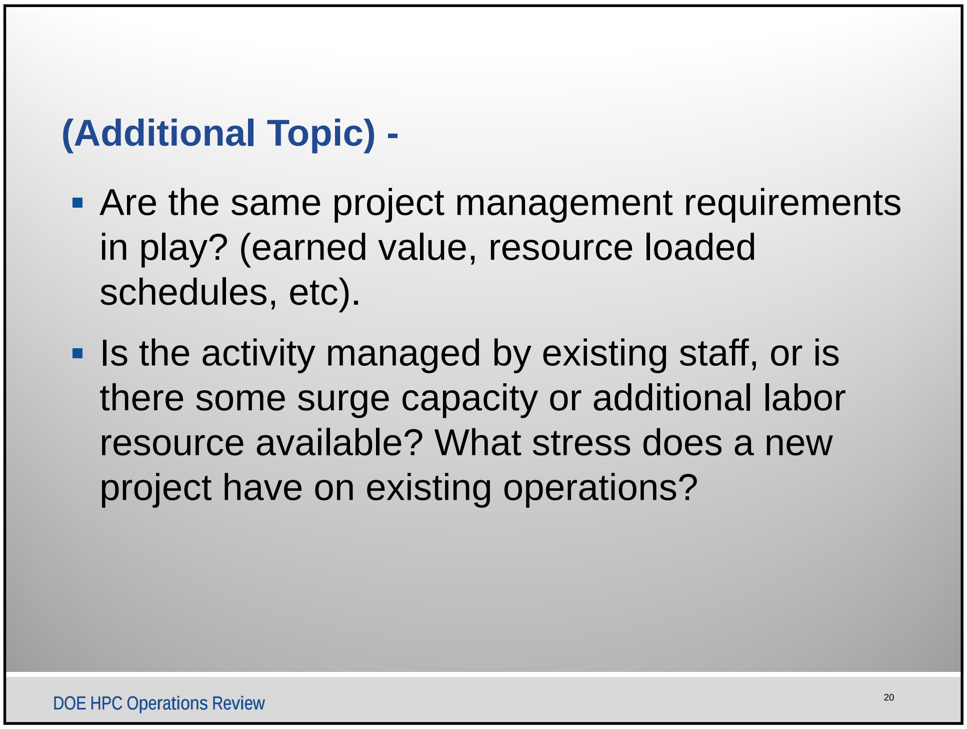
Effort estimate

- How big of an effort was this (facility prep for a new system)? Team size?
  - Estimate $1-1.5M/MW as a Project cost for an upfit to an existing facility (clean install)
  - Estimate that ~50% of project cost is labor (subcontractor, PM, etc.)
    - Significant fluctuations in labor and material costs based on supply chain issues, skilled labor demand, etc.
  - How many FTEs? Are these people already on-staff? Is this other-duties-as-assigned, or is there bench support for this work?
    - These are big efforts, with several ten’s of people participating.
      - Planning and design effort.
      - Construction effort
      - Colocated activities? Parallel efforts in the same space can dramatically impact schedule, cost, risk.
      - System Installation and Integration effort.
      - In existing facilities, the conflict of managing existing operations while upfitting the existing facilities.



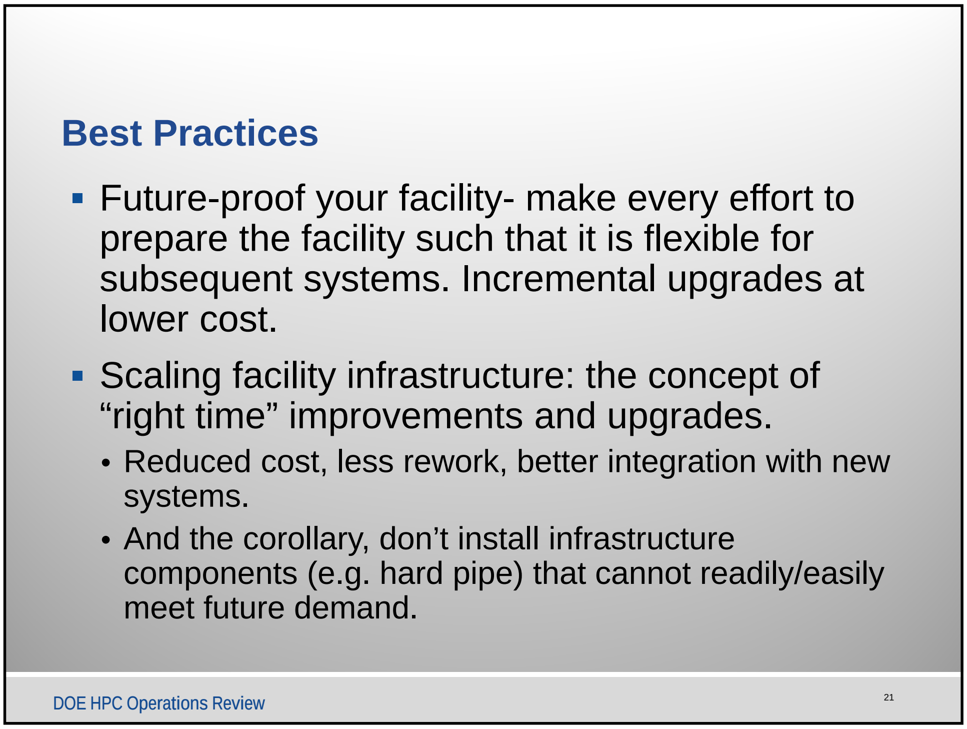
(Additional Topic) - Funding

- Is the activity funded as a separate project?
- Is the activity funded as part of the acquisition?
- Is the activity funded through external mechanisms, such as third party landlord, or other?
- Is the activity funded separately, through operations (not a separate project)?
- How are budgets for facilities improvements derived? (lab investment, program office line item, operations) and who does these- sow/engineering firm/architect?
- To what extent is the cost for the decommissioning of the system accounted for?
- What is the anticipated life time for the major components of an upgrade? Are the medium voltage improvements a 25-40 year item? Transformer and switchboard improvements a 20-year item? Chilled water loops- 10-15 years? Ancillary and secondary and step transformers- just 5 years?



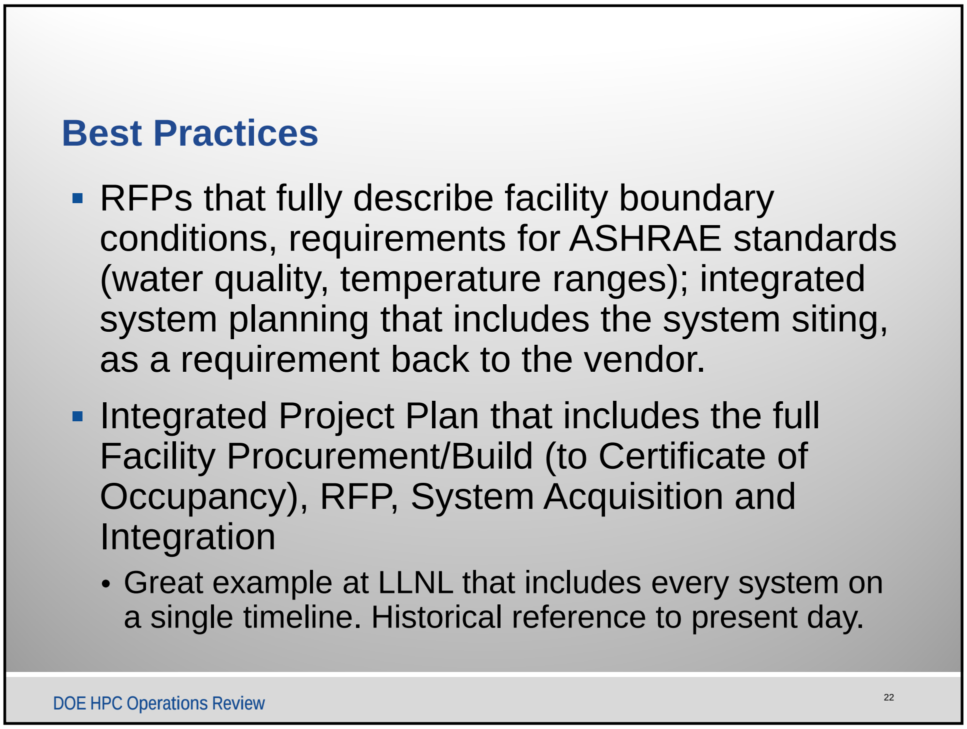
(Additional Topic) -

- Are the same project management requirements in play? (earned value, resource loaded schedules, etc).
- Is the activity managed by existing staff, or is there some surge capacity or additional labor resource available? What stress does a new project have on existing operations?



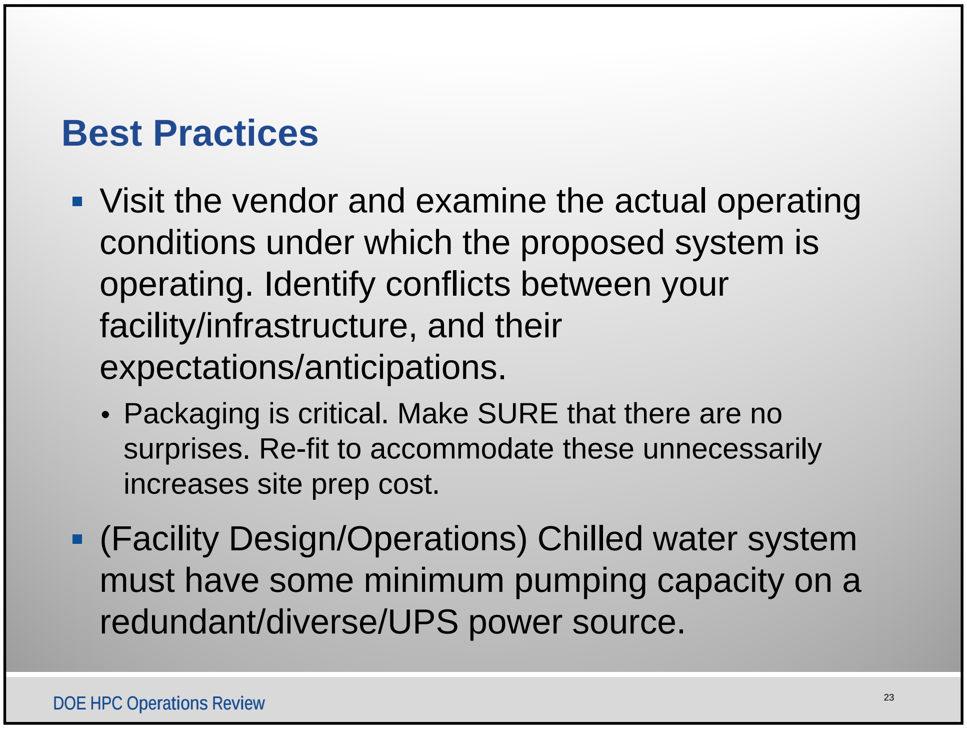
Best Practices

- Future-proof your facility- make every effort to prepare the facility such that it is flexible for subsequent systems. Incremental upgrades at lower cost.
- Scaling facility infrastructure: the concept of "right time" improvements and upgrades.
  - Reduced cost, less rework, better integration with new systems.
  - And the corollary, don't install infrastructure components (e.g. hard pipe) that cannot readily/easily meet future demand.



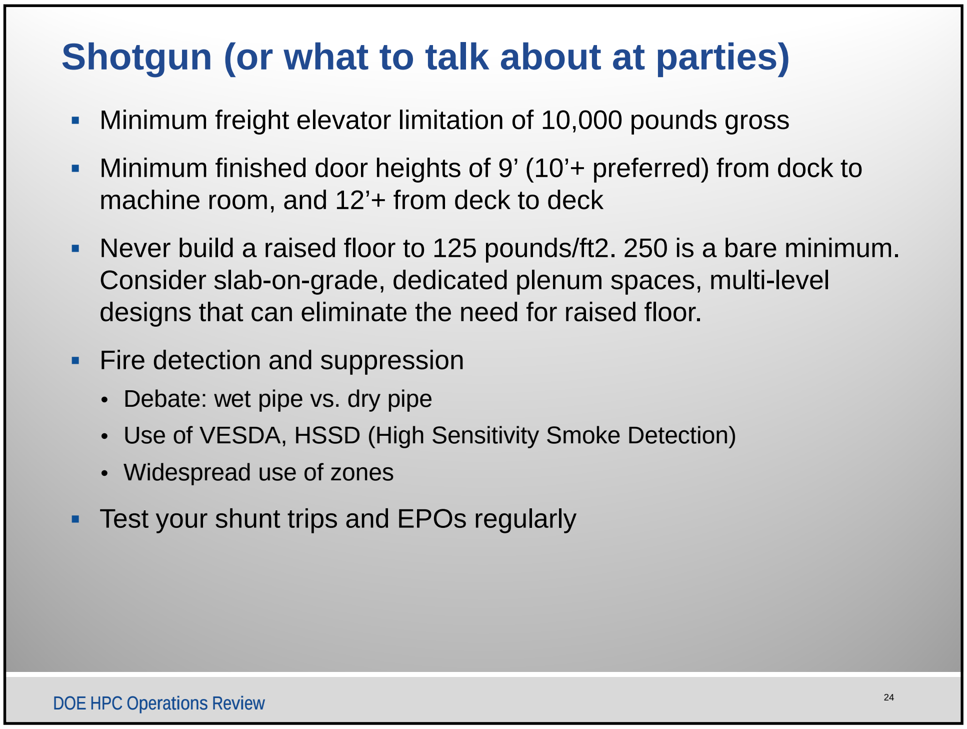
Best Practices

- RFPs that fully describe facility boundary conditions, requirements for ASHRAE standards (water quality, temperature ranges); integrated system planning that includes the system siting, as a requirement back to the vendor.
- Integrated Project Plan that includes the full Facility Procurement/Build (to Certificate of Occupancy), RFP, System Acquisition and Integration
  - Great example at LLNL that includes every system on a single timeline. Historical reference to present day.



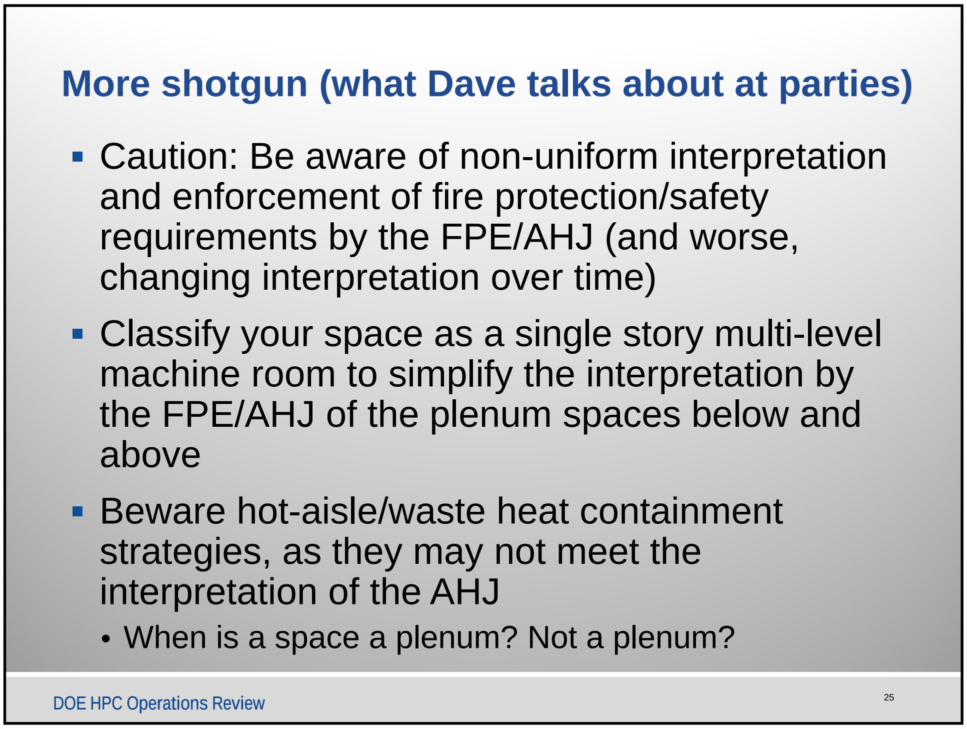
Best Practices

- Visit the vendor and examine the actual operating conditions under which the proposed system is operating. Identify conflicts between your facility/infrastructure, and their expectations/anticipations.
  - Packaging is critical. Make SURE that there are no surprises. Re-fit to accommodate these unnecessarily increases site prep cost.
- (Facility Design/Operations) Chilled water system must have some minimum pumping capacity on a redundant/diverse/UPS power source.



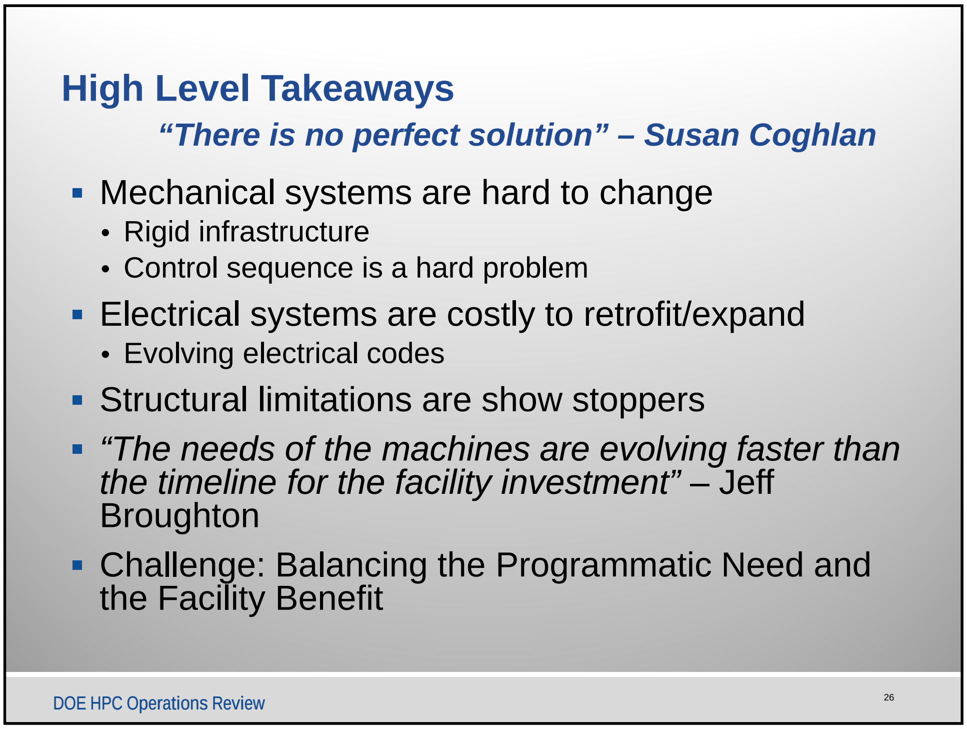
Shotgun (or what to talk about at parties)

- Minimum freight elevator limitation of 10,000 pounds gross
- Minimum finished door heights of 9’ (10’+ preferred) from dock to machine room, and 12’+ from deck to deck
- Never build a raised floor to 125 pounds/ft2. 250 is a bare minimum. Consider slab-on-grade, dedicated plenum spaces, multi-level designs that can eliminate the need for raised floor.
- Fire detection and suppression
  - Debate: wet pipe vs. dry pipe
  - Use of VESDA, HSSD (High Sensitivity Smoke Detection)
  - Widespread use of zones
- Test your shunt trips and EPOs regularly



More shotgun (what Dave talks about at parties)

- Caution: Be aware of non-uniform interpretation and enforcement of fire protection/safety requirements by the FPE/AHJ (and worse, changing interpretation over time)
- Classify your space as a single story multi-level machine room to simplify the interpretation by the FPE/AHJ of the plenum spaces below and above
- Beware hot-aisle/waste heat containment strategies, as they may not meet the interpretation of the AHJ
  - When is a space a plenum? Not a plenum?



High Level Takeaways

"There is no perfect solution" – Susan Coghlan

- Mechanical systems are hard to change
  - Rigid infrastructure
  - Control sequence is a hard problem
- Electrical systems are costly to retrofit/expand
  - Evolving electrical codes
- Structural limitations are show stoppers
- *"The needs of the machines are evolving faster than the timeline for the facility investment"* – Jeff Broughton
- Challenge: Balancing the Programmatic Need and the Facility Benefit