# OAK RIDGE LEADERSHIP COMPUTING FACILITY PROJECT (OLCF-3)

## Lessons Learned Report

**DRAFT**

**September 2013**

# 1. LESSONS LEARNED

## 1.1 OVERVIEW

During the course of the OLCF-3 Project, innovations and actions in response to unplanned events were tracked and are summarized here as lessons learned.

## 1.2 KEY LESSONS LEARNED

The specific lessons listed below were identified by OLCF staff and other important stakeholders during the OLCF-3 Project. Many lessons learned in past projects have been transformed into best practices for the OLCF staff.

The lessons below have been organized by Project Execution, Business Management, and Project Management, with a notation of whether the lesson was learned from a project success (*Success*) or an observation of needed improvement *(Area for Improvement).*

***Project Execution***

**Facilities**

- Weekly project meetings attended by the facility operations, computational customer, the 3rd party building owners, and the contractors performing work are key to keeping open communication channels. Frequent project focused meetings allowed the various stakeholders on the project to discuss project schedule, scope changes, impacts to operations, and construction quality. (*Success*)

- Specifications for acquired major equipment must identify dimensional limitations required to allow the equipment to be moved into and out of facilities. Using data supplied by with vendor submittals, verify that equipment meets these dimensional limitations prior to signing purchase or lease documents and follow up with visual inspection of equipment after it is delivered to ensure that something was not missed in this process.

A large transformer was purchased for this project. ORNL engineering personnel questioned the design build contractor and the vendor that would supply the transformer regarding its dimensions. Their response indicated that the transformer would fit down a narrow corridor it had to go through to reach a retrofitted transformer room. The dimensions given by the vendor were for the cores with the enclosure removed but didn't include the base that transformer set on. Inspection of the transformer after it was delivered to the site for storage revealed that the base was too wide. This required that the base be cut around the perimeter to provide the necessary clearance then reassembled after the transformer was set in place. The reassembly of the base returned the transformer to a condition equivalent to what it was structurally before the based was cut. *(Area for Improvement)*

- Observe new contract workers, especially those in lead or supervisory positions, to ascertain that they understand and are following ORNL safety and quality assurance procedures and take quick action with contractor to correct deficiencies or make personnel adjustment when it becomes clear that the performance is below par and he or she will not make appropriate effort to correct their deficiencies.

A contractor's superintendent, new to ORNL assignment, neglected to follow Lab safety and quality assurance guidelines. Eventually this was discovered, but not in time to avoid significant rework on

some of the medium voltage cable supplying the new transformer. The new superintendent was presented by the contractor as a Master Electrician with 25+ years' experience at this position and very qualified for the job. This Superintendent's poor performance had been noted prior to work on the Titan project but it took some time for contractor to replace him. A more rigorous inspection by ORNL of construction work activities managed by this superintendent could have mitigated some of the impact and provided a better basis for the contractor to expedite his actions relative to the superintendent. *(Area for Improvement)*

- Require unbalanced load or primary current injection testing of power transformer secondary breaker ground fault interrupters to verify correct polarity of neutral circuit current transformers (CT).

The new power transformer and its close coupled secondary breaker have a CT on the neutral circuit that runs through the breaker enclosure (not the breaker) back to the transformer neutral and ground with the breaker for this installation having phase current CT's inside the breaker. On two previous installations that were similar to this one but occurred prior to this and on other computer systems it was discovered that because the load current went through the transformer secondary breaker from bottom to top, the neutral CT in the breaker enclosure was wired backwards resulting in the phase and neutral current vector sums being about double what the phase current was. The vector sum of these values should be zero at all times. Left uncorrected, this would have eventually resulted in the transformer secondary breaker tripping off. The new installation for Titan was not checked for this problem during installation because the load current went through the breaker from top to bottom, which is the conventional direction. After the transformer was put in service and a small load applied, the metering equipment on the breaker was checked to verify absence of this problem and a significant amount of ground current which being displayed by the meters indicating a CT wiring problem. A factory representative was called in and the system was de-energized to correct this problem. *(Area for Improvement)*

- Delays in the installation of a critical piece of equipment (Unit Substation dry type transformer) led to a need to temporarily store the equipment outside of 5600. The equipment was packaged for shipping with weather proof wrapping. The wrapping was opened by the contractor upon arrival to take field measurements. By opening the wrapping the contractor inadvertently provided a potential path for water to enter the unit. The contractor attempted to repair the wrapping but heavy rain did allow a small amount of water to enter the unit. The unit was dried out thoroughly prior to operation, but ideally the unit should have been moved to an interior location for storage if unable to be immediately installed. *(Area for Improvement)*

**Acceptance Testing and Commissioning**

- The massive effort put forth during a lengthy system acceptance process can cause team member fatigue (staff burnout). Labor resource management is required during this time should be part of the up-front project planning. *(Area for Improvement)*

- Better diagnostic tools are needed to assist in troubleshooting and locating problematic and/or suspect hardware. Development efforts related to such tools need to be performed upstream of actual acceptance. *(Area for Improvement)*

**Application Readiness**

- Development of good application performance metrics takes lots of time working with the actual application developers. This must be done well ahead of actual acceptance. (*Area for Improvement*)

- Application readiness is critical for overall project success. Select the applications to be used in the acceptance process (for functionality, stability, and performance) well ahead of the actual acceptance period and thereafter hold regular meetings with the science teams to track the ability of each application to utilize and stress the target and/or representative HPC platform effectively. Use this period to perform a detailed performance analysis of each application and develop a predictive performance model of that application on the planned hardware. Bring in additional support for this effort when required. (*Success*)

***Business Management***

**Contracts**

- The process of negotiating and contracting with the OEM and leasing companies should be coordinated to ensure that the negotiation and approval of both agreements are in sync. Use 3rd party leasing companies when beneficial to the project and DOE, but include sufficient time to for the process to be followed properly.

The ability of the project to use a 3rd party leasing company(s) provided the project financial flexibility to pursue lower overall project costs. Lease agreements require a significant amount of review by both the laboratory and DOE and this time should be incorporated into the project baseline schedule. (*Success*)

- Allow flexible terms in initial contracts for these state of the art computer purchases, but keep contracts current with updated decisions about computer design and configuration.

The initial contract for state-of-the-art computers may need to use flexible terms to accommodate industry roadmap changes (i.e. incorporation of the latest versions of components, changes to hardware and/or to quantities to meet performance specifications), but these contracts need to be updated regularly to reflect the specific deliverables as these deliverables are determined. (*Success*)

- System components can be purchased through a single vendor source if the risk associated with system integration is high. Weigh the costs and benefits of having system integration performed through the main system vendor to insure that the mitigation of the risk is worth the additional cost. (*Success*)

- Ensure that contracts are written to include all important aspects of acceptance criteria to ensure that proper emphasis is placed during contract execution by the vendor. Include acceptance test language in the portion of the contract that represents the acceptance test applications required. If the feature is not embodied within the contractual terms, it won't appear in the system. (*Success*)

- There is a lack of depth in the Contract Services with regard to awarding and managing leasing agreements. This applies to all leases, including those for large computer systems. Additional staff in Contract Services must be trained in the execution and management of leasing agreements. *(Area for Improvement)*

***Project Management***

**Opportunity Costs**

- Careful consideration of impacts to the user community should be considered when looking at major upgrades.

The upgrade of Jaguar to Titan saved $25M in up-front costs, which were rolled back into the project upgrade, allowing Titan to be larger. This was a big win in that it gave us a system large enough to attract users to go to the effort of converting their codes to run on a new hybrid CPU/GPU architecture. The down side is that the schedule showed that the users would be without a system for about 3 months while the initial upgrade from XT5 to XK6 was completed, and then for several weeks when the GPUs were installed and accepted. While this was probably overly optimistic based on a best case scenario, something that rarely if ever happens when installing serial #1 systems, it also didn't account for the possibility of something going wrong. In this case, we ended up having to repair all of the node boards in the system, twice. The resulting down time to users was likely more than $25M in lost productivity and time on the system. So in the future, careful consideration of impacts to the user community should be considered when looking at major upgrades. If there are many new technologies, assume that one or more may be immature and result in significant debug and repair time. Also assume that it will not be apparent until you run the system at scale, so testing at the scale of 25% of the final system may not identify the problem.

**Risk Mitigation**

- When coupled with a long-term partnership such as OLCF's with Allinea and CAPS, placing support staff on location at the center is a best practice that enhances the partnership, leads to improvements in the partner's products, and facilitates center operations.

Allinea Software, Ltd. is the developer of the distributed debugging tool (DDT), and CAPS Enterprise is the developer of the HMPP compiler suite for GPUs. We have implemented subcontracts with Allinea and CAPS to enhance and extend their products to meet the scale and functionality needs of OLCF users. In 2012 we expanded these partnerships to include the full-time placement of partner employees on-site at ORNL to provide in-depth support for these tools to OLCF staff and users. On-site CAPS and Allinea staff actively work with application teams to address problems that go beyond typical tool support, often assisting in use of the tools in challenging situations. For example, Allinea on-site staff member Dirk Schubert recently helped an application team use DDT to isolate and fix a bug that occurred in its application only when it was run at full scale on Titan. (*Success*)

**Training and Support**

- It would be advantageous to have a platform available to users throughout the duration of the training activities. For part of the project, we made TitanDev available to users and that was a useful training platform. It would have been more beneficial to have made a resource like TitanDev available throughout the duration of the training activities.

Users commented that they would like to participate in more hands on training activities to help reinforce and apply what they learned from the training. In order to offer hands on activities, it is useful to have a training platform that is available not only during the class but later as users start to put into practice what they learned in the training. *(Area for Improvement)*

- Training for staff and users on new technology must be timed should that the technology selected is stable enough to create a meaningful training program. Plan the timing of training so that (1) the lessons presented represent the actual technology to be employed and (2) there are training platforms available for hands-on training exercises.

OLCF attempted to prepare for training too early in the project. The technological picture was unsettled causing several rewrites to lesson plans and some confusion for users as things shook out. In a related matter, training users long before a development platform is available to put the training into practice is liable to result in users forgetting much of the training. *(Area for Improvement)*