LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-TR-222521

Level-1 Miletone 350

Terri Quinn

June 30, 2006

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

Table of Contents

Milestone Definition Text

Milestone (350): Provide a 100 teraflops Platform environment supporting tri-lab DSW and Campaign simulation requirements.

Level: 1 Fiscal Year: 2007

Completion Date: 12/2006

ASC nWBS Subprogram: CSSE and FOUS

Participating Sites: LLNL, LANL, and SNL

Description: This milestone is the direct result of work that started seven years ago with the planning for a 100 Teraflop platform and will be satisfied when 100 Teraflops is placed in operation and readied for Stockpile Stewardship Program simulations. The end product of this milestone, code named Purple, will be a production high-performance computing system designed to be used to solve the most demanding stockpile stewardship problems, that is, the large-scale application problems at the edge of our understanding of weapon physics. This fully functional 100 TeraOP/s system must be able to serve a diverse scientific and engineering workload. It must also have a robust code development and production environment both of which facilitate the workload requirements.

This multi-year effort includes major activities in contract management, facilities, infrastructure, system software, and user environment and support. Led by LLNL, the Tri-labs defined the statement of work for a 100 Teraflop system that resulted in a contract with IBM known as the Purple contract. LLNL worked with IBM throughout the contract period to resolve issues and collaborated with the Program to resolve contractual issues to ensure delivery of a platform that best serves the Program for a reasonable cost.

The Purple system represents a substantial increase in the classified compute resources at LLNL for NNSA. The center computer environment must be designed to accept the Purple system and to scale with the increase of compute resources to achieve required end-to-end services. Networking, archival storage, visualization servers, global file systems, and system software will all be enhanced to support Purple's size and architecture. IBM and LLNL are sharing responsibility for Purple's system software. LLNL is responsible for the scheduler, resource manager, and some code development tools. Through the Purple contract IBM is responsible for the remainder of the system software including the operating system, parallel file system, and runtime environment.

LLNL, LANL, and SNL share responsibility for the Purple user environment. Since LLNL is the host for Purple, LLNL has the greatest responsibility. LLNL will provide customer support for Purple to the Tri-labs and as such has the lead for user documentation, negotiating the Purple Usage Model, mapping of the ASC Computational Environment requirements to the Purple environment and demonstrating those requirements have been met. In addition LLNL will demonstrate important capabilities of the computing environment including full functionality of visualization tools, file transport between Purple and remote site file systems and the build environment for principle ASC codes. LANL and SNL are responsible for delivering unique capabilities in support of their users, port important applications and libraries, and demonstrating remote capabilities. The key capabilities that LANL and SNL will test are user authorization and authentication, data transfer, file system, data management, and visualization. SNL and LANL should port and run in production mode a few key applications on a substantial number of Purple nodes.

Success Criteria

- Successful completion of the associated L2 milestones
- Successful completion of the milestone deliverables as defined by the Completion Criteria below
- Completion of the milestone certification method below

Associated L2 Milestones

- ASC Purple System Final Hardware Acceptance [MS #1899]: Complete
- TSF Activation [MS # 1348]: Complete
- Deploy First Phase of I/O Infrastructure for Purple [MS # 461]: Complete
- Deploy Next-Generation Data and Visualization Capabilities for BlueGene/L and Purple Environments [MS # 1703]: Due FY06 Q3

Milestone Deliverables

- Hardware Deliveries: System hardware deliveries from vendor to site are complete, including the basic hardware to integrate "the system" as contractually defined.
- Hardware Installation: Installation of the system by the contractor on-site to the extent that is contractually required is substantially complete
- Hardware Acceptance: Hardware acceptance testing as contractually required is completed In general, contractual requirements for formal hardware acceptance have been substantially completed
- System Software: System software needed for basic operation of the system is delivered, tested, and demonstrated to be operational. LLNL is responsible for the scheduler, resource manager, and some code development tools. Through the Purple contract IBM is responsible for the remainder of the system software including the operating system, parallel file system, and runtime environment.
- Scalability Testing: Vendor has completed on-site capability scaling testing and demonstration. System is ready to begin on-site integration into local and remote computing infrastructure, and user environment integration
- Machine Accessibility and Integration: System is made available to approved capability-class projects Approved users can request accounts, gain access to, and expect availability of the system and all supporting infrastructure
- Operational Support: All system software, tools, utilities and processes to support operation and use of the machine are available and functional. Includes hardware and software monitoring, systems support and response, user support, training, documentation, and tools and policies for scheduling and managing the projected capability workload
- Usage Model: System has demonstrated an acceptable production user environment as defined in the platform Usage Model, based on tri-lab user requirements in the ASC Computing Environment (ACE) document
- User Testing: ASC applications targeted to run on the platform must be ported and made available to designers, analysts, and engineers for a reasonable time period prior to milestone completion Sufficient testing has been performed in limited access mode by users and issues impacting their use of the system have been addressed
- System Reliability: System has demonstrated acceptable reliability performance targets for capability mode usage
- Application and I/O Performance Testing: At-scale system performance tests running capability workloads will be expected to meet targets, as indicators of the system's readiness for production use, including supporting I/O infrastructure

Completion Criteria

- Hardware Deliveries, Hardware Installation, and Hardware Acceptance
    1. Demonstrate that the Purple hardware met the contract requirements
- System Software and Scalability Testing **(See attachments 1, 2, 3, 4, 6)**
    1. Successful completion of the Synthetic Workload Test as defined in the Purple Synthetic Workload Test Plan. Areas tested include:
        - MPI functionality and performance
        - GPFS functionality and performance
        - Compiler, debugging, and tuning tool functionality
    2. Scheduling (SLURM and LCRM) functionality and performance

- Machine Accessibility and Integration **(See attachments 14, 15, 16. See also separate document – Certification of Completion of Level-2 Milestone 461)**
    1. Formal, documented procedures to request and grant an account are in place
    2. Length of time from request submittal to account in place
    3. Documented procedures for machine login from remote locations
    4. Access to file system and archival storage space in place and documented
    5. Access to viz facilities from machine are in place and documented
- Operational Support Capability **(See attachments 6, 7)**
    1. Demonstrate a robust capability to monitor the system
    2. Demonstrate a process to manage the system configuration
    3. Demonstrate system stability meets contract requirements
    4. Document Purple support model and verify model implementation with early Purple users
- Usage Model **(See attachments 8, 9)**
    1. The usage model will be negotiated with the three Lab. A draft document that contains the Purple usage model, applicable ACE requirements, and user documentation is titled, *Purple Computational Environment with Mappings to ACE Requirements*, and has been sent to SNL and LANL for their comments.
    2. Demonstrate how the ACE requirements were met
    3. Demonstrate that the Capability Compute System Scheduling Governance Model can be implemented on Purple
- User Testing **(See attachments 10, 11, 12, 13)**
    1. Applications targeted to run on the platform have been ported
    2. Demonstrate Purple specific user requirements have been met or the issues impacting their use of the system have been addressed user requirements for Purple. Examples of the users requirements are:
        a. Code developer system access
        b. Sufficient early access by users
        c. Functionality and performance of the development environment including compilation, debugging, and performance analysis
        d. Transfer of data to and from remote sites
        e. Data analysis
        f. Restart dump efficiency
        g. Problem restart efficiency

h. Application monitoring

- System Reliability **(See attachment 1)**
    1. System meets the contract requirements for stability
- App execution under machine loaded conditions **(See attachment 13)**
    1. Single App execution met expectations of users
    2. Measure and document application performance

Milestone Certification Method

1. ASC Execs establish a review panel with three external members and two members from each lab
2. Approximately 5 to 6 months before milestone completion LLNL will review with the panel the milestone scope, deliverables, and completion criteria and resolve any concerns.
3. The ASC Program office at each lab has designated a single point-of-contact for the L1 milestone
4. The panel will meet before the milestone due date to perform its review and make a recommendation to the LLNL ASC Executive.
5. HQ is responsible for rating the overall success and completion of this milestone

Attachment 1. System Reliability and SWL Testing

Prepared by Pam Hamilton

This paper describes the milestone deliverable of System Reliability that supports the ASC L1 Milestone for the100 Teraflops Computing Environment. It will also outline the criteria for success of this deliverable and the current status.

System Reliability

The high-level definition for this deliverable is: System has demonstrated acceptable reliability performance targets for capability mode usage. Therefore, the goal of the deliverable is to demonstrate that ASC Purple is stable and reliable in the presence of typical ASC programmatic usage patterns. For the purposes of this deliverable, capability mode jobs are defined as applications utilizing 75% or more of the CPUs in the Purple system compute partition. The reliability performance targets are based on requirements detailed in the ASC Purple procurement Statement of Work (SOW) contract that was negotiated with IBM. The mean time between application termination due to hardware failure (MTBAF) which IBM committed to in the SOW is 62.7 hours (2.6 days).

Criteria for Success

An obvious criterion for success is that the system meets the contract requirement for stability or MTBAF of 62.7 hours stated above. To track and calculate the MTBAF, LLNL is recording each hardware event in a problem tracking system. All events are then extracted and categorized by type of hardware failure. Hardware failures that would cause an application failure are tabulated and the MTBAF calculated.

Another proposed measurement is to use resource manager log files to track individual job statistics and graph how many machine days per month were used by capability mode jobs. For reporting purposes, we will use the definition of Capability Mode Computing as described in the Capability Compute System Scheduling Governance Model.

A third criterion is to show that stability exists if the correct results for established applications can be generated repeatedly and reliably over a specified time span. This was demonstrated by a mix of simulated code development activity and production simulations sustained for a period of five continuous days (120 hours) (called the Synthetic Workload application load Stability Test, or SWL-ST). The application codes used in the SWL included representatives of all of the major ASC application code development efforts at LLNL, with a sampling of applications from Sandia and LANL. The SWL-ST filled the machine with a large number of simulations spanning the full range of problem sizes, with run times varying from a few minutes to several hours. No single problem lasted 120 hours. Following a successful period of five days for the SWL-ST, test results were documented and archived.

Current Status

LLNL has been actively tracking the MTBAF of ASC Purple since November 2005. The data is reviewed weekly with IBM and analyzed to watch for any hardware component failure trends. A recent MTBAF calculation was 56 hours which is obviously less than

IBM's commitment. In the beginning, this data did help identify hardware and software issues. Currently, there aren't any trends to the types of failures LLNL is experiencing. The MTBAF will continue to be tracked and will be included in the L1 Milestone completion documentation.

All of the data about individual job statistics is being captured in resource manager log files currently. A tool will be developed to mine these log files and present the information in graphical form to illustrate how many machine days per month were used by capability mode jobs. The graphs will be included in the L1 Milestone completion documentation.

The final criteria demonstrated by the Synthetic Workload application load Stability Test, or SWL-ST has been completed. The test results that were documented and archived will be included in the L1 Milestone completion documentation.

Attachment 2. Development Environment- Compiler, Debugging, and Tuning Tool Functionality

Prepared by Scott Futral

Goal of this Deliverable

Effective use of the Purple system requires a robust code development environment. The goal of this milestone is to ensure that the compilers, debuggers, and performance analysis tools expected by the ASC users were available, functional, and met the unique requirements of the Purple system.

Activities to support this goal are:

- Deploy and validate the baseline tools
- Address deficiencies and problems discovered in the tools by working with the vendors and tools developers
- Develop additional capabilities to improve the tools offerings and to address unique Purple system requirements.

Criteria for Success

- Successful use of the tools by the ASC code teams in deploying, debugging, and tuning their codes is the ultimate success criteria.
- Provide a report on problems encountered with the tools and environment, and the resolution or status
- Demonstrate and report new capabilities that have been incorporated into the development environment tools for Purple.

Status

- The development environment is deployed and largely robust. ASC applications teams are successfully using the development environment on Purple. As a third-generation IBM ASC system, this is as expected.
- There were some challenges for the tools from the Purple environment. New capabilities are available in the compilers (XL and GNU) to address the 64-bit default build environment, to address the “large page” memory model for AIX, and to address some C++ incompatibilities between compilers.
- The TotalView debugger has new features added to enhance performance and capability on Purple, and new memory debugging features.
- Performance analysis and tuning capabilities are provided and enhanced, and additional tools capabilities expected to be delivered prior to completion of the L1 Milestone.

Attachment 3. MPI Functionality and Performance

Prepared by Terry Jones

Goal

The requirements for MPI are designed to ensure that applications which depend on MPI will have a robust, functionally complete, and high performance MPI. We specifically targeted three categories of MPI validation: robustness, functionally complete, and high performance:

> **Robustness:** It doesn't matter how fast you arrive at an answer if the answer is wrong. Since any new flagship machine for the DOE complex will have pushed the envelope for scale, ensuring no unwanted behavior at scale is an important activity.
>
> **Functionally complete:** MPI functionality concerns usually deal more with coverage than concerns over correctness (no doubt a result of the maturity of the specification). Here we ensure the desired interfaces are present and there operation proceeds as expected.
>
> **High perfomance:** For a software stack to be considered "high performance" it must efficiently deliver the capabilities of the underlying hardware and provide levels of performance in keeping with the leading machines of the time.

Criteria for Success

LLNL established separate items for each of the three component areas of robustness, functionally complete, and high performance. Included in ***functionality*** was a demonstration of scaling to 8192 tasks, a demonstration of scalable memory usage, acceptable documentation, and full MPI-2 minus dynamic tasking. The ***robustness*** element for MPI was addressed separately via full MPI application mtbf in the Synthetic Workload (SWL). The following table outlines the ***performance*** measurements:

Performance Highlights		
Description	**Target**	**Actual**
1tpn Interconnect link bandwidth (any number sources to sinks via striping)	4.46 GB/sec bi (57% of 8 GB/sec) 3.23 GB/sec uni (84% of 4 GB/sec)	5.68 GB/sec bi-directional 3.31 GB/sec uni-directional
8tpn Interconnect link bandwidth (8 sources to 8 sinks)	4.8 GB/sec bi (60% of 8 GB/sec) 3.2 GB/sec uni (85% of 4 GB/sec)	5.85 GB/sec bi-directional 3.77 GB/sec uni-directional
1tpn Interconnect link latency (1 source to 1 sink)	5.5 us ping-poing (msg + ack)	5.01 us
8tpn Interconnect link latency (8 sources to 8 sinks)	8.0 us ping-poing (msg + ack)	6.00 us
Aggregate machine bi-section BW, worse case pairing (all communication across 3rd stage)	45% efficient (NumNodes * .45 * 8 GB/sec)	47.19% random-random ~70% typical 98.8% nearest neighbor
Collective Operation Scaling (10,000 allreduces with "crunch cycle")	Verify scales as Log2(ntask).	passed with co-scheduler

Current Status

All MPI related Statement of Work (SOW) target performance objectives have been met. Both MPI-only and Hybrid-MPI codes have successfully met scaling expectations on Purple (including ale3d, yf3d, and other classified applications).

The final criteria demonstrated by the Synthetic Workload application load Stability Test, or SWL-ST has been completed. The test results that were documented and archived will be included in the L1 Milestone completion documentation.

Special Challenges

During the testing of MPI, it was discovered that interference or noise introduced by the operating system was impacting performance. In particular, if a histogram is generated for a tight loop in which each iteration consumes exactly 1.00 milliseconds, then the result for millions of samples reveal significant binning above the expected 1.00 milliseconds. An investigation determined that this interference was primarily due to issues within the hypervisor. A new version of the hypervisor (IBM GA7) removes almost all of the interference.

Of all of the MPI requirements, the most problematic performance metric during the time of testing was bi-section bandwidth. Eventually this metric was achieved, but it was in doubt when Purple was first delivered to LLNL. LLNL and IBM undertook an effort to understand the extent of impact for various levels of shortfall on ASC applications while other efforts continued in parallel to bring up the metric up to the target of 45% efficiency for worse case pairings. By using environment tuning, we were able to achieve 47% efficiency for worse case (most pairings actually perform much higher).

Attachment 4 GPFS Functionality and Performance

Prepared by Bill Loewe

Deliverable

The GPFS deliverable for the Purple system requires the functionality and performance for ASC I/O needs. The functionality includes POSIX and MPIIO compatibility, and multi-TB file capability across the entire machine. The bandwidth performance required is 122.15 GB/s to a single shared file, as necessary for productive and defensive I/O requirements, and the metadata performance requirement is 5,000 file stats per second.

Criteria

To determine success for this deliverable, several tools are employed. For functionality testing of POSIX, 10TB-files, and 1024-node capability, the parallel file system bandwidth performance test IOR is used. The MPIIO functionality is tested with the MPIIO test suite from the MPICH library. Bandwidth performance is tested using IOR for the required 122.15 GB/s sustained write. Metadata performance is tested after "aging" the file system with 80% data block usage and 20% inode usage. The fdtree metadata test is expected to create/remove a large directory/file structure in under 20 minutes time, akin to interactive metadata usage. Multiple (10) instances of "ls -lR" concurrently on different large directory is used to demonstrate 5,000 stats/sec.

Status

In November, 2005, the Purple acceptance test was performed on a 1024-node system with 3 metadata servers and 101 I/O servers. The functionality testing was completed successfully. For the metadata performance, the "ls -lR" performance was 9500 stats/sec in aggregate. The fdtree test completed in under 20 minutes as well. In addition, application testing was performed under load using UMT2K on 850 nodes (3400 procs) writing a 1163GB-file per timestep while IOR was repeatedly writing 40TB-files from 200 nodes (200 procs). The average of 10 timesteps for UMT2K was 18 GB/s, ranging from 9 to 60 GB/s while IOR was averaging 74 GB/s to the same file system.

In April, 2006 on the full Purple file system with 3 metadata servers and 125 I/O servers, IOR showed file-per-process performance rates of 73 GB/s for write and 115 GB/s for read on 512 nodes to 16GB-files. The single-shared-file performance was 129 GB/s (W) and 153 GB/s (R) for 1024 nodes to an 8TB-file.

Upon replacement of existing hard drives on Purple's GPFS, future testing will include bandwidth, metadata, data integrity, and application I/O results.

Special Issues

Of utmost importance to the Purple file system is the integrity of the data. Testing is performed with IOR using the data write-checking option. After IOR completes a write of a file with this option enabled, the file is then read back (on a different node to prevent any cache effect) and compared against the known, written data pattern. In cases where data is written and stored correctly but intermittently read back incorrectly, there is an option to write data and then reread multiple times to find errors.

Attachment 5 Resource Management – LCRM and SLURM

Prepared by Don Lapari

Goal – Run SLURM on Purple Instead of IBM's LoadLeveler

- LoadLeveler's performance is very poor on White and would have been much worse on Purple.
- LoadLeveler licenses are very expensive.
- SLURM has become a very successful, open-source alternative to LoadLeveler. SLURM has been downloaded by sites around the world. Many developers outside of LLNL contribute to the SLURM code base and provide bug fixes.

Goal – Modify LCRM to Support SLURM Running on an AIX Platform

- Existing version of LCRM supported LoadLeveler on AIX and SLURM on Linux, but SLURM on AIX required substantial modifications to the LCRM code.

Challenges - To Extend the SLURM Resource Manager to the AIX Platform

- Adapt SLURM's Linux-based build system to a new architecture with AIX tools
- Develop a new library for SLURM to be used instead of LoadLeveler by IBM's parallel job launch utility, *poe* (essentially make *poe* think it was talking to LoadLeveler).
- Translate LoadLeveler constructs to SLURM (e.g., LoadLeveler's pools to SLURM's partitions)
- Develop a form of remote job launch for AIX
- Create a new driver for IBM's Federation switch
- Develop an AIX kernel extension to track all processes created by a job
- Implement the Checkpoint / Restart capability into SLURM

Challenges – To Extend LCRM to Support an AIX version of SLURM

- Thousands of lines of code need to be reviewed and modified for this support.
- All LCRM enhancements must provide a user interface that conforms to existing conventions and modes of operation.

Criteria for Success

- Purple runs jobs without LoadLeveler.
- LCRM / SLURM collect and report computing resource usage statistics for every job run on Purple.
- LCRM / SLURM enforce the computing resource usage policy defined by the CCCs.
- Users invoke the same commands and follow the same conventions for running LCRM / SLURM on Purple as they do for all other machines in the Center.
- Job launch times are under a minute or two (excluding factors outside the control of LCRM / SLURM).
- Jobs terminate cleanly (without leaving orphaned processes) in less than one or two minutes (again, excluding factors outside the control of LCRM / SLURM).

Attachment 6. Purple Operational Support

Prepared by Pam Hamilton

Goal of Deliverable

This deliverable is to address the requirements, criteria of success, and current status of the Operational Support area of the Purple Level 1 Milestone. The requirements are:

1. A capability to monitor the hardware and software subsystems and provide a support and response mechanism that maximizes the availability of Purple.
2. A process to manage the configuration and environment of Purple.
3. A suite of tests to stress and measure system stability.
4.

Criteria for Success

Purple is composed of four major hardware subsystems that require constant monitoring, these are, the nodes of the system, the high performance switch interconnect, the external networking, and the disk subsystem. Purple is also composed of numerous software subsystems, but this item will only relate to the monitoring requirements of the GPFS file system, the resource scheduler, and software involved in accessing the system. For the system to be maximally available problems with any of these systems need to be identified and fixed as soon as possible. This requirement will be successful when a set of tools and procedures is in place that can test for subsystem status, report problems to operational staff, and procedures are in place to fix or work around issues.

The Purple system has 1532 nodes, 512 HPS switch boards, 506 RAID controllers, 65 hardware management controllers, and one management server. All these devices need consistent configuration and tunables for the system to operate correctly. This requirement will be successful when a set of tools are available to install consistent configuration files, activate tunables, and check for correctness on all devices.

For Purple to be useful to the program, it must be stable and deliver reliable results under a typical ASC work load during a given time span. This requirement will be successful when a suite of test codes and a process to run the suite and to verify results are available.

Current Status

Livermore Computing has a common system to monitor all the production platforms, file and application servers, and raid controllers that have potential impact to the users. This system is call Host Monitoring (HM) and is based on SNMP daemons reporting status to a central server, and web pages for status display. Each host or cluster can be customized to report status on specific subsystems beyond the common items like CPUs on line, percent free in the file system, etc. The HM system is used by system administrators and off hours operations staff to verify systems are operating as expected (green status), identify potential problems or conditions that might become problems (yellow status), and identify hardware failures and critical issues (red status). The HM system has been ported to the Purple environment and is now in use. The operations staff has procedures to fix problems or call out on-call system administrators. Additional work needs to be done to monitor the new disk subsystem that will be installed on Purple this summer. Also, we have recently experienced 10GE link problems on Purple and plan to add more

monitoring on end-to-end performance of the networks on Purple that connect to the archive and the tri-lab.

The Purple configuration is managed by a combination of IBM provided features and locally developed tools. The AIX operating system package management system has been enhanced, based on recommendations by LLNL and the SP-XXL group, to track non-official patches to the system, and track dependencies that are required for the official fix. This system is used along with the normal package management to track the state of system files and ensure the configuration on all nodes in the cluster is identical. The locally written tools are based on the UNIX rdist command, perl wrapper scripts, and the LLNL developed Genders cluster management package that is common on LC compute platforms. All system tunables and configuration files on Purple are managed by this system, updates are run daily and incorrect configuration listed in daily reports. Additional work will need to be done to manage the new disk subsystem that will be installed this summer.

To test the stability and consistency of results on Purple the Synthetic Workload (SWL) was developed. The SWL is composed of three parts: functionality tests, performance tests, and stability tests. The functionality and performance tests were developed to test specific requirements in the Purple contract. The stability test is a suite of benchmarks and application codes that are representative of ASC work loads. The SWL was run on 1280 nodes of the Purple system in the fall of 2005 as part of the acceptance. The SWL suite of tests has also been used and will continue to be used to test stability and consistency of results after major updates to the system.

Attachment 7. Operational Support Capability

Prepared by Jean Shuler

Milestone deliverable: Operational Support Capability: All system software, tools, utilities and processes to support operation and use of the machine are available and functional. Includes hardware and software monitoring, systems support and response, user support, training, documentation, and tools and policies for scheduling and managing the projected capability workload.

Completion Criteria:
Document Purple support model and verify model implementation with early Purple users.

Metrics of Success:

- User support hotline in place and operational as evidenced by Remedy reports with metrics about Purple tickets by numerous criteria.
- Purple classes are presented at LLNL and Tri-Lab locations. Numbers of classes, numbers and affiliations of students, and ratings will be documented.
- Complete web suite of Purple information available on the classified and unclassified networks.

Status:

- Personal help: The LC hotline is open 7:30 AM to 4:45 PM and currently responds to questions and problems of users of Purple. Operational support is available 24x7 with an operator responding to email and phone calls. The REMEDY trouble ticket system is used by Operations and the LC hotline to record, track, assign and resolve problems, issues, requests, and information. The REMEDY system also provides detailed metrics concerning the type of problems reported. Chat rooms are also available for quick response between the LC hotline, Operations staff and system administrators,
- Documentation: Machine-specific user documentation is available in the /usr/local/docs directory on Purple (this includes basic instructions for running on Purple, FAQs, "gotchas", "dos-and-don'ts"). Man pages are available on the machine and via the web. MOTDs and news items are kept up-to-date and there are specific email lists for sending status to users. Four thousand pages of Purple-relevant online manuals such as EZACCESS, EZFILES, EZJOBCONTROL, EZSTORAGE are available on both the classified and unclassified web sites at http://www.llnl.gov/computing/hpc/documentation/index.html#manuals . Tri-Lab information is available on the unclassified web page at http://hpc.llnl.gov and on the classified web page at http://www.llnl.gov/trilab .
- Status information: The machine status page for hardware monitoring is on the classified web page at http://www.llnl.gov/computing. Network availability information (NETMON) is at http://lc.llnl.gov/discom showing the availability and bandwidth data of the Tri-Lab networks.
- Training: The purple tutorial is at http://www.llnl.gov/computing/tutorials/purple/ The first Purple workshop was held at LLNL on June 29. Tri-Lab users attended the class. Two "Using ASC Purple" workshops are scheduled at SNL, NM in July. LANL and SNL users have registered.

Challenges:
Provide an unauthenticated classified web server so Tri-Lab users could access web information without using a token. A new web server is in place and specific information is being modified so it is viewable without a token.

Attachment 8. Purple Usage Model

Prepared by Jean Shuler

Milestone deliverable:
Usage Model: System has demonstrated an acceptable production user environment as defined in the platform Usage Model, based on tri-lab user requirements in the ASC Computing Environment (ACE) document.

Completion Criteria:

- The usage model will be negotiated with the three Labs.
- Testing demonstrates how the ACE requirements were met.

Metric of Success:

- The usage model is successfully negotiated with the three labs. The ACE requirements reflect the high performance computing requirements for the General Availability User Environment capabilities of the ASC community.
- The set of ACE requirements implemented on Purple exceed the minimal expectation of what the major Tri-Lab applications require to run. A written memo stating the usage model is acceptable is sent to the ASC execs by the Tri-Lab POCs.
- Tri-Lab user testing and documentation demonstrate that the agreed upon ACE requirements were met satisfying the criteria for Purple milestone success from each Lab.

Status:
Version 1.1 of the draft document that contains the Purple usage model, applicable negotiated ACE requirements, and user documentation was sent to Tri-Lab POCs for comments. This document is titled "Purple Computational Environment with Mappings to ACE Requirements" and has been reviewed and updated twice. Each ACE requirement is noted in the appropriate section of the document. For each section of this document, a description of ACE requirements met and not met is provided. A section is devoted to addressing the requirement and documenting the current status.
LLNL staff have already been assigned to each requirement and are working on the issues. SNL and LANL specified their "Criteria for running successfully" on Purple in a document that was sent to LLNL in January, 2006. Identified LC staff are addressing these requirements and criteria to ensure the criteria are met in order to meet the L1 milestone.

Challenges:
There is current discussion on what is an acceptable time to start or restart a job on Purple.

Attachment 9. Capability Compute System Scheduling Governance Model

Prepared by Brian Carnes

Goal

In the fall of 2005, the ASC Program appointed a team to formulate a new governance model for allocating resources and scheduling the stockpile stewardship workload on ASC Capability Systems. The objectives of this initiative are:

- To ensure that the capability system resources are allocated on a priority-driven basis according to the Program requirements.
- To utilize ASC Capability Systems for the large capability jobs for which they were designed and procured.

Within the constraints of meeting the two primary objectives, this model maximizes effective use of the machine both by minimizing idle cycles and by enhancing the probability of productive and useful capability calculations. A capability class system is similar in value and uniqueness to a large experimental facility. Major programmatic computing efforts will be organized as computing work packages and will be reviewed and prioritized for relevance, importance and technical rationale. Each proposed work package, called a *Capability Computing Campaign* (CCC), consists of at least one major calculation needing a significant proportion of an ASC capability system, together with related supporting jobs of smaller sizes. Allocation of capability resources will be achieved using a two-step process. First, a Capability Planning Advisory Committee (CPAC) will review the proposed CCCs and make a recommendation to the Capability Executive Committee (CEC). Second, the CEC will review this recommendation and make the final allocations. The portfolio of approved CCCs will be will be managed by LCRM on the Purple System. Usage will be tracked and reports will be generated and distributed on a monthly basis to insure time is used as allocated and job sizes meet proposal requirements.

Criteria for Success

The CCC concept will be declared a success if we can fairly and efficiently implement the objectives outlined above by the Milestone completion date. We will document the following:

- Prioritized list of CEC approved campaigns and allocations awarded.
- Monthly reports to projects and management on utilization, job size, idle and down time.

This will prove effective enforcement of allocations according to established priorities.

Status

We are now in early Tri-Lab access mode for Purple. CPAC representation from each Lab has been announced; however the formal CCC concept has yet to be initiated. The call for proposals is being developed as well as the process for review and implementation.

Special Challenges

The CCC concept has never before been used to allocate time for Tri-Lab ASC capability systems, Purple will be the first system to utilize this new governance model.

Attachment 10. LANL Input for Purple Milestone Success

Prepared by Rand Rheinheimer, LANL

Using the SAGE code, run an asteroid impact problem as an acceptance test.

- This simulation will be run on a minimum of 8000 PE's, and will initially contain at least one billion cells. Demonstrate simulation results are consistent with simulation results obtained on other platforms.
- Achieve 24 cumulative (i.e. with restarts) simulation wall clock hours within a 30-hour wall clock window. The simulation time is to be measured by the SAGE application, with total computational wall time also recorded.
- Demonstrate an ability to write a problem restart dump in 3 minutes. This dump file is expected to have a size of roughly 500 GB. Alternatively, (if the file size differs) demonstrate the ability to write a restart dump at a rate of 3 GB/s.
- Demonstrate an ability to restart the simulation from a dump file within 10 minutes.
- Demonstrate an ability to effectively monitor, from LANL, the progress of the simulation at LLNL.
- Demonstrate an ability to view at LANL the visualization files generated by the simulation.
- Demonstrate an ability to archive and delete problem dump/graphics files during simulation. See notes for projected requirements.

Demonstrate full LLNL Hotline support for LANL users by reporting the number of open and closed LANL-related tickets generated by October 31, 2006.

After completion of the SAGE acceptance testing, do a 3D problem production run with a Crestone Project code. This problem should achieve at least 24 hours of simulation wall time.

To confirm general availability, run a 3D problem with a code from the Silverton project and analyze the results. This problem should achieve at least 24 hours of simulation wall time. Also begin a 2-D high-fidelity physics simulation.

Attachment 11. LLNL NIF Capsule Simulation[1]

Prepared by Marty Marinak, Nathan Barton, Rob Neely, and Steve Langer

We will test purple by running a high-resolution simulation of the implosion of a beryllium ablator capsule at the National Ignition Facility. Simulations of this sort are used to assess the effects of the grain structure in beryllium on the quality of the implosion. The National Ignition Campaign will decide whether to use a capsule with a beryllium or a diamond ablator at the end of 2007. The results of this simulation will provide important input to that decision and will help guide target fabrication efforts between now and the first ignition experiments in 2010.

The beryllium is initially heated by absorbing x-rays (causing partial melting) and then is completely melted by a strong shock. The grain structure leads to non-uniform expansion during the initial heating phase and the resulting perturbations may be amplified by hydrodynamic instabilities during the implosion.

The first step of the simulation will be to use ALE3D to model the interval between the start of the experiment and the point at which the first shock breaks into the fuel (frozen DT). The temperature, density, and velocity fields from the ALE3D run will then be coupled to HYDRA. HYDRA will model the implosion, including the growth of the grain-scale perturbations calculated by ALE3D.

The grains in beryllium are less than a micron wide while the capsule is a millimeter in radius. The ALE3D simulation will require roughly a billion zones and the HYDRA simulation will require roughly 100 million zones. The ALE3D simulation should take roughly 48 hours on 8192 processors of purple. It should take a few hours for Overlink to convert the ALE3D results into input for HYDRA. The HYDRA simulation should take about 48 hours on 8192 processors of purple. These simulations will have roughly 3X larger solid angle than previous simulations and will allow us to investigate the coupling between the short wavelength (l~1000) modes seeded by the grain structure and the l~30 modes which have the greatest impact on capsule performance (l is the number of wavelengths that fit around the circumference of the capsule). The total request is for roughly 100 hours on 8192 processors.

Both simulations will be run on most of purple, making this a good test of the ability of purple to run “Grand Challenge” simulations. These simulations will also verify the ability of these codes to effectively use nearly 10,000 processors to deliver a timely result of a high-resolution simulation.

[1] This attachment released separately as UCRL – ABS - 222476

Attachment 12. SNL Completion Criteria

Prepared by Judy Sturtevant, SNL

SNL Demonstration Applications Summary
Sandia has two demonstration applications, Salinas and ALEGRA.

The Salinas code has been ported to all other Tri-Lab ASC platforms, except BG/L, and was the SNL demonstration application for White and Q. To date, it has demonstrated almost perfect scaling to 1000 processors. The problem that will be run on Purple includes a series of computations on increasingly finer meshes to establish the effect of mesh resolution on the accuracy of the solution. The finest mesh resolutions will require thousands of processors.

ALEGRA is a full science simulation is multi-physics code. On Purple, the team will first try several scaling studies for single physics in ideal cases. If these don't scale well, there is a problem in the application code that must first be solved. Each of the five different physics scaling studies will require about one week of low usage on a few hundred processors, then one or two days of dedicated time on whole machine. Meanwhile the team will start to develop a full science calculation, requiring tens of smaller simulations, on tens to hundreds of processors for about a week each. Depending on code performance in scaling studies and science run simulation, the team can then determine full computing requirements for capability-sized runs.

SNL Completion Criteria

1. Access Purple user documentation at LLNL. Glenn Machin, Karen Haskell
2. Complete SNL user authorization, authentication, and access tasks. Glenn Machin
3. Provide SNL user documentation for remote use of Purple. Karen Haskell, Joel Stevenson, and Jon Goldman
4. Establish SNL user defined groups for file sharing. Karen Haskell, Joel Stevenson
5. Test archival storage for SNL needs. Tim Preston
6. Complete tests of moving data between SNL and LLNL. Tim Preston, Tom Pratt, John Naegle, Brian Kellogg, Joel Stevenson
7. Test effective use of the file systems and I/O libraries. Marty Barnaby, Greg Sjaardema, Eric Illescas
8. Complete application development environment porting:
    a. Port SNL SEACAS pre- and post-processing tools; port data format and I/O libraries Greg Sjaardema
    b. Port SNL SIERRA pre- and post-processing tools; port data format and I/O libraries Greg Sjaardema, Eric Illescas
    c. Port Trilinos library. Karen Haskell
9. Test data management environment for SNL needs. Greg Sjaardema
10. Test scientific visualization environment for SNL needs. (Requires running parallel servers of both ParaView and EnSight, connecting to clients at SNL.) Lisa Ice, John Greenfield, Jon Goldman, David Karelitz, David McCutcheon

11. Port and run SNL Salinas and ALEGRA application codes in production mode:
    Demonstration applications and users:
    Salinas application developers – Garth Reese, Christopher Riley Wilson, Joel Miller
    Salinas end users – Wil Holzmann, Angel Urbina, Luba Kmetyk
    ALEGRA application developers – Randy Summers, Sue Carroll, Rich Drake
    ALEGRA end users – Tom Mehlhorn, Tom Brunner, Chris Garasi, Joe Crepeau
    Additional application help from Mahesh Rajan, Hal Meyer
    Additional visualization help from Lisa Ice, John Greenfield, Jon Goldman, David Karelitz, David McCutcheon
    a. Execute application pre-processing in interactive mode.
    b. Execute application in interactive debugging mode.
    c. Execute application in batch mode.
    d. Monitor the application as it executes.
    e. Write restart files as needed and restart from them.
    f. Post process in parallel for data analysis.
    g. Visualize data in parallel, using both EnSight and ParaView, with the servers on Purple and the client(s) at SNL.
    h. Write data to LLNL HPSS system.
    i. Move desired results back to SNL systems.

Attachment 13. Application Execution Performance

Prepared by Scott Futral

Goal of this Deliverable:

The execution performance of applications is always an area of interest with new systems. The goal of this milestone is to ensure that the expected performance has been achieved for well understood applications, and appropriate measures are taken to analyze and tune Purple to improve run time performance for the more general set of ASC applications.

Activities to support this goal are:

Demonstrate Performance of Marquee Codes

Perform HPCC Benchmarks

The 'SCOPE' scaling and performance study

Criteria for Success

- Successful run of applications using sufficient memory for sufficient length of time that achieve Purple SOW requirements.
- Report HPCC Benchmark results
- Present results of the SCOPE study, and application team generated performance and validations results for at least three ASC applications.
- Demonstrate capture of performance and scaling data for a representative group of Tri-Lab applications.

Status

- Purple SOW requirements were satisfied for the Marquee Codes.
- Outstanding Linpack and HPCC results were achieved.
- The SCOPE effort has analyzed a large number of critical applications with the involvement of Tri-Lab users. Several significant issues regarding usage were uncovered and either fixed or a significant workaround put in place. 1 large source of noise and application performance degradation has been uncovered and a potential fix has been tested with positive results. There are still some run-to-run variations that may be related and need to be better understand. Recommendations for executing on Purple are to be documented and made available to the user community.

Attachment 14. Machine Accessibility and Integration

Prepared by Jean Shuler

Deliverable Goal:
System is made available to approved capability-class projects. Approved users can request accounts, gain access to, and expect availability of the system and all supporting infrastructure.

Completion Criteria:

- Formal, documented procedures to request and grant an account are in place.
- Length of time from request submittal to account in place is minimal.
- Documented procedures for machine login from remote locations
- Access to file system and archival storage space in place and documented.
- Access to viz facilities from machine are in place and documented.

Metric of Success:

- URL for web pages documenting the procedures for requesting and granting an account and submitting a CCC request to run on Purple is available on the classified and unclassified web pages. The CCC governance model is linked from the CCC form web page.
- From the time the completed account form arrives at LC-Support an existing LLNL ASC user will have an account the next day. New accounts will take up to two working days.
- Fully descriptive web pages and user manuals documenting the procedures for machine login from remote locations are available on the classified and unclassified networks.
- Fully descriptive web pages and user manuals documenting access to file system and archival storage space are available on the classified and unclassified networks.
- Web pages and information are available in the /usr/local/doc directory on Purple to document access to the viz facilities and describe methods for using the 64-node "viz" pool.

Status:

- The web page for submitting a CCC proposal is being tested (https://cmg-rr.llnl.gov/lcforms/CCC_proposal_form.html). The Governance model is on the web at https://cmg-rr.llnl.gov/lcforms/policies/ccc_policy.html. The web page for requesting an account is at http://www.llnl.gov/computing/hpc/accounts/#account. Tri-lab users will submit an account request using SARAPE (http://sarape.sandia.gov). This request initiates the process and provides additional verification and approval for guest requests. SARAPE does not replace the internal process established at each site to process access requests. The allocation of machine time will be determined by the CCC proposal process.
- Account requestors who have existing accounts can be immediately activated if it is high priority once the completed paperwork has been received. Typically an existing account activation takes a day. New accounts will take up to two working days to be in place. This has been tested since June 12 when the initial 14% tri-lab allocations were activated.
- Fully descriptive web page and user manuals are available on the classified and unclassified networks at http://www.llnl.gov/computing/hpc/access/ . We have tested the ability for users to follow instructions and log in successfully and Tri-Lab users are running on the Purple machine. This information is documented in the purple tutorial as well as http://www.llnl.gov/computing/hpc/access/ssh_outside.html and the EZACCESS online manual.
- Fully descriptive web pages and user manuals are available on the classified and unclassified networks. The IBM file system structure, archival storage and access information is documented

in the EZFILES online manual http://www.llnl.gov/LCdocs/ezfiles/ and archival storage information is at http://www.llnl.gov/LCdocs/ezstorage/ , the HPSS user guide is at http://www.llnl.gov/LCdocs/hpss/ and the HTAR manual is at http://www.llnl.gov/LCdocs/htar/ . Information on the IBM file systems and archival storage is also documented in the online Purple tutorial http://www.llnl.gov/computing/tutorials/purple/

- There is a 64 node visualization pool on Purple that has been used successfully. Information on using this "viz" pool is in the /usr/local/doc/purple.basics directory on Purple. Web pages at http://www.llnl.gov/computing/hpc/jobs/ describe visualization facilities and how to get help on data visualization and analysis.

Challenges:
Documented, scripted, coordinated tests by participants from tri labs will be done during weekly Tri-Lab User Support telecon calls before the machine is GA.

Attachment 15. Data Management and visualization functionality for Purple

Prepared by Steve Louis

Goal
Users will be provided measurably enhanced data management and visualization display functionality, sized to meet the exceedingly powerful Purple environment. Specifically:

1) Formal release(s) of Hopper and Telepath that incorporate new user requested or required data and resource management functionality.
2) Successful installation, integration, deployment, and use of new TSF collaborative space displays, and Purple generated visualization data.

Criteria for Success
Successfully use new higher performance Hopper data management tool for Purple files.

Successfully use the Telepath display resource tool and the new TSF collaborative space displays to visualize and analyze Purple simulation data.

Status
Hopper Release 1.3.4 (released November 2005).
Hopper Release 1.3.5 (released January 2006).
Hopper Release 1.4.0 (released June 2006).

Telepath Release 2.0 (released March 2006).

Tilden Visualization Theater (activated August 2005).

Level 2 milestone completion certification review held on June 27, 2006.

Special Challenges
None.

Milestone (ID #1703): Deploy Next-generation Data and Visualization Capabilities for BlueGene/L and Purple Environments (LL-CSSE-06-02)
Level: 2 (L2 tasks supporting the Purple L1 Milestone are highlighted below)
Fiscal Year: FY06
DOE Area/Campaign: ASC
Completion Date: FY06-Q3
ASC nWBS Subprogram: Computational Systems and Software Environment
Description: The unprecedented capabilities of BlueGene/L and Purple, as well as their unique system characteristics, require next-generation data and visualization tools, running on next-generation server and image delivery infrastructures (existing data and visualization capabilities are commensurate with present capability and capacity machines). As a result of meeting this milestone, users will be provided measurably enhanced data management and visualization functionality, sized to meet the exceedingly powerful BlueGene/L and Purple environments. Milestone deliverables include: deployment of a large 256-node cluster-based visualization server utilizing next-generation interconnects, processors, and graphics cards; hardware-accelerated parallel rendering technologies supported within the VisIt tool; **new releases to extend user features of the Hopper file management tool and Telepath resource session orchestration tools; high-resolution digital image delivery to new collaborative-use areas.** Successful deployment of next-generation capabilities requires close and ongoing interaction with key customers to help ensure that delivered functionality meets user requirements for increasingly productive environments. Enhanced data management and visualization for BlueGene/L and Purple data requires determination of necessary scaling, specific improvements, and newly needed capabilities based on the size and power of BlueGene/L and Purple. Success also requires close interaction and integration with other product areas of Computational Systems and User Environments, as well as Facility Operations and User Support.
Completion Criteria: Successful installation, integration, deployment, and customer use of the new Gauss visualization cluster in support of BlueGene/L. Formal release(s) of VisIt that incorporate hardware-accelerated parallel rendering, together with measurements that demonstrate higher rendering rates than previous tools. **Formal release(s) of Hopper and Telepath that incorporate new user requested or BlueGene/L- or Purple- required data and resource management functionality. Successful installation, integration, deployment, and customer use of new collaborative space displays, using digital image delivery technology, and BlueGene/L or Purple generated visualization data.**

Attachment 16 Visualization and Data Analysis Tools

Prepared by Steve Louis

Goal of this deliverable

Parallel visualization and data analysis tools are critical to understanding the simulation data generated on Purple. This milestone will ensure that parallel visualization tools are installed and operational for visualizing and analyzing data from ASC simulation codes. Specifically:

1) Install and test that CEI Ensight runs on Purple.
2) Install and test that ParaView runs on Purple.
3) Install and test that VisIt runs on Purple.

Criteria for Success

Use the tool to visualize and analyze data in parallel on Purple. Each tool will have a specific set of tests that test common operations for that tool. There should also be documentation in place that specifies how to use the tools on Purple.

Status

Ensight and ParaView are already installed and running on architectures similar to Purple but have not been installed and tested on purple.

VisIt 1.5.2 was been installed and tested on Purple. Furthermore it has been run 690 times by 22 different users between April 19 and June 23, 2006.

Special Challenges

None.