SAND2019-10835R

SANDIA REPORT
SAND20XX-XXXX
Printed Click to enter a date

Sandia National Laboratories

Design, Installation, and Operation of the Vortex ART Platform

Nathan E. Gauntt, Kevin D. Davis, Jason J. Repik, James Brandt, Ann Gentile, Simon Hammond

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

ATS platforms are some of the largest, most complex, and most expensive computer systems installed in the United States at just a few major national laboratories. This milestone describes our recent efforts to procure, install, and test a machine called **Vortex** at Sandia National Laboratories that is compatible with the larger ATS platform **Sierra** at LLNL. In this milestone, we have **1)** configured and procured a machine with similar hardware characteristics as Sierra ATS, **2)** installed the machine, verified its physical hardware, and measured its baseline performance, and **3)** demonstrated the machine's compatibility with Sierra ATS, and capacity for useful development and testing of Sandia computer codes (such as SPARC), including uses such as nightly regression testing workloads.

ACKNOWLEDGEMENTS

We would like to acknowledge the special contributions from several organizations and individuals. We would like to thank Mike Glass and Si Hammond for assisting with configuring the procurement within budget and certifying the results. Special thanks to the SNL SPARC code team for executing end-to-end regression tests and providing valuable feedback, including Micah Howard and Sam Browne. Thanks to the SNL SIERRA code team members Rich Drake, and the SNL EMPIRE code team members Matthew Bettencourt for their testing efforts. We thank all our SNL early system user testers from the division 1400 and 1500 code teams for finding and reporting bugs.

We would like to thank all the division 9000 staff including HPC management, facilities support staff, and system administration support staff. We would also like to thank IBM vendor staff Juanice Campbell, Jeff Gerhart, Mike Stevens, Sean McCombe, and many others for providing high-quality logistic and technical support.

We would like to specially thank Lawrence Livermore National Laboratory and its HPC support staff including Scott Futral, Adam Bertsch, John Gyllenhaal, Py Watson, and Matt Legendre. For accommodating site visits, sharing code, setting up accounts, and supporting our ongoing synchronization efforts, we very much appreciate the significant effort spent to help make this tri-labs partnership a success.

CONTENTS

Abstract 3
Acknowledgements 4
Executive Summary 7
Acronyms and Definitions 9
1. Preface 10
2. introduction 11
3. Configuration and Procurement 13
4. Installation Details 15
4.1. Facilities Integration 15
4.2. Hardware Validation and Performance Testing 15
4.3. SNL and LLNL Collaboration 15
5. Results 17
5.1. Sierra ATS Compatibility and Testing Sandia Codes 17
5.2. Platform Issues and Solutions 17
6. Conclusion 19
6.1. Future Work 19
References 22
Appendix A. Implementation Plan 23
Appendix B. Evidence 24
B.1. Email Correspondence 24
B.2. Midpoint Presentation 32
B.3. Year-End Presentation 52
Distribution 63

LIST OF FIGURES

Figure 1: Timeline of Vortex platform design, procurement, and install 7
Figure 2: Vortex rack design for auxiliary infrastructure and compute cabinets matches Sierra ATS design 13

This page left blank

EXECUTIVE SUMMARY

The objectives of this ASC FOUS-led milestone were as follows:

1. Install an Application Readiness Testbed (ART) system, Vortex, consisting of hardware and a software stack conforming to the Sierra ATS sited at LLNL, including water, cooling, and power infrastructure, and integrate it into the Sandia computing environment

2. The Vortex system is generally available to Sandia code development teams and computational analysts

3. The Vortex system and its environment shall be capable of supporting development and testing of Sandia codes, including nightly regression tests (according to application readiness).

The following figure illustrates the work periods in the successful completion of this milestone. Design and procurement work begins during "FY18 prep work" in 12/2017, and the milestone is completed during "supports SPARC nightly regression tests" in 06/2019, though ongoing work continues.

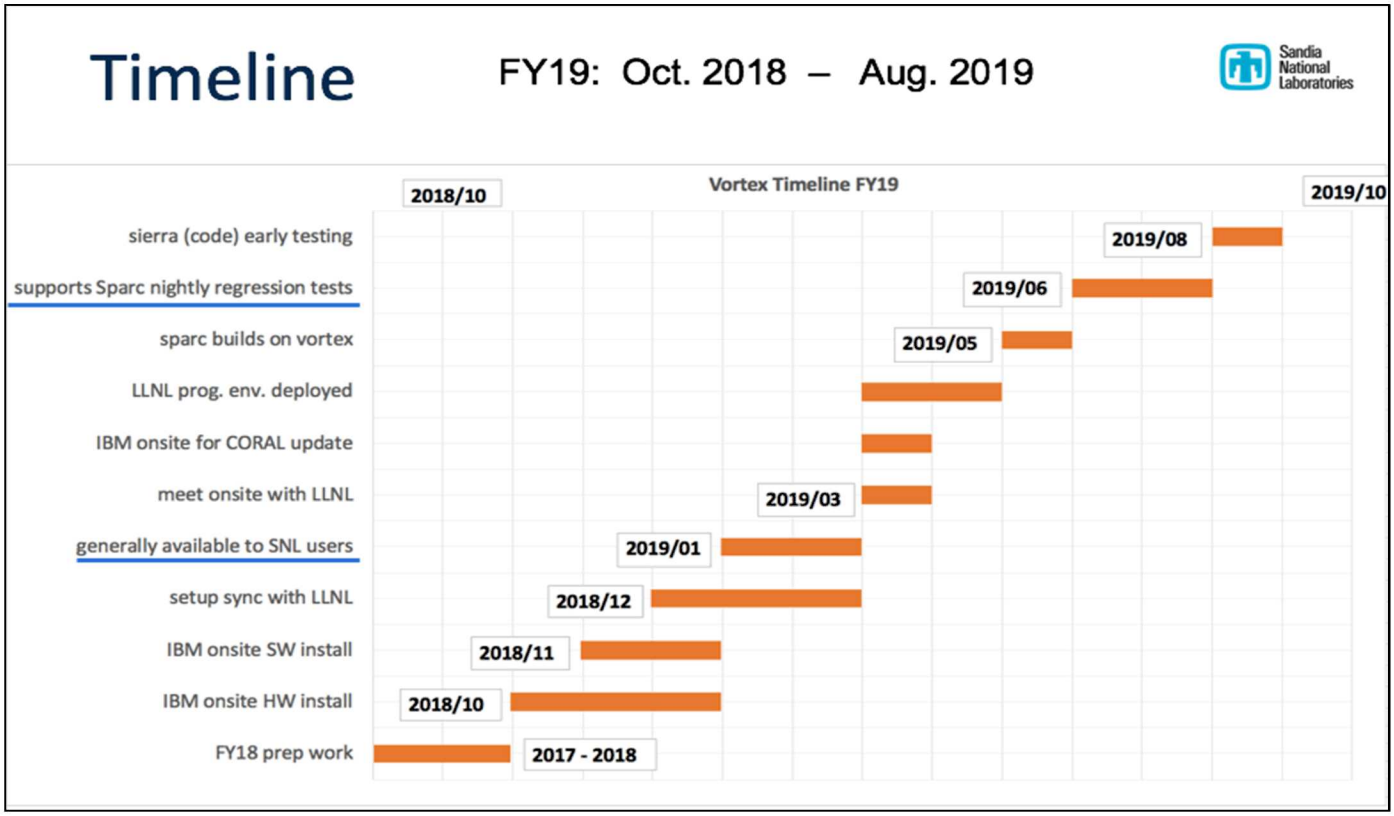


Figure 1: Timeline of Vortex platform design, procurement, and install

The following high-level conclusions were reached:

- A Sierra ATS binary-compatible machine was successfully installed at Sandia National Laboratories, including water, cooling, and power infrastructure, and integrated into the Sandia computing environment.
- The Sandia SPARC code was built on Vortex (executed nightly regression tests at SNL) and ran successfully on the RzAnsel and Sierra ATS platforms at LLNL
- The CORAL software stack and LLNL programming environments were installed, evaluated, and upgraded several times during this period. A cross-site relationship was formed with LLNL, and ongoing regular meetings and software synchronizations occur monthly
- Other SNL codes, such as EMPIRE, SALINAS, and SIERRA are being actively developed and tested on Vortex. This includes use of graphical debuggers and performance analyzers (totalview, paraview) that are not currently feasible to use from remote sites

Impacts of the milestone activities, accomplishments, and results include the following:

- Successful implementation of a complex, novel topology HPC system matching Sierra ATS at Sandia National Laboratories, including challenging facilities, operational, and technical team collaboration
- Sandia significantly grew the user and testing base for the CORAL software stack, providing a unique platform for tri-labs customers to evaluate and improve HPC vendor products
- Successful demonstration of a small-scale, Sierra ATS compatible platform that can be used by tri-labs customers

Path forward, beyond the work of this milestone, is likely to include the following:

- Support of LANL, LLNL, and other system users to validate software configurations and isolate platform issues from site-specific concerns
- Encourage better vendor response by reproducing concerns on an independent, tri-labs HPC resource
- Test of use cases like: specialized hardware optimization (NVMe), and easy access to high-speed SNL HPC filesystems (big data)

ACRONYMS AND DEFINITIONS

Abbreviation	Definition
SNL	Sandia National Laboratories
DOE	U.S. Department of Energy
NNSA	DOE's National Nuclear Security Administration
ASC	NNSA's Advanced Simulation and Computing program
FOUS	ASC's Facility Operations and User Support program element
CCE	FOUS' Common Computing Environment program element
ATS	Advanced Technology System
CTS	Commodity Technology System
TLCC	Tri-lab Linux Capacity Cluster, the predecessor of CTS
ACES	NNSA's New Mexico Alliance for Computing at Extreme Scale
POWER9	IBM PowerPC CPU
HPC	High-Performance Computing
SRN	Sandia Restricted Network
TCE	LLNL's Tri-labs Computing Environment
CORAL	Collaboration of Oak Ridge, Argonne, and Livermore
ART	Application Readiness Testbed
LLNL	Lawrence Livermore National Laboratory
LANL	Los Alamos National Laboratory

1. PREFACE

The work documented within this report showcases the Vortex ART cluster, installed and operated by the HAAPS "Heterogeneous Advanced Architecture Platforms" team, in the HPC Development department. This team provides leading-edge capabilities and services to a broad range of customers at Sandia National Laboratories.

2. INTRODUCTION

The description as specified in the FY19 ASC Implementation Plan of Level-2 (L2) Milestone #6814 is provided below. This Milestone's full implementation plan is provided within Appendix A.

> "Sandia will complete **deployment and production availability** of an Application Readiness Testbed (**ART**) system consisting of hardware and a software stack **conforming to the Sierra ATS sited at LLNL**. The system will support the porting and testing of Sandia's codes and their applications against this new architecture. Site preparation includes **water, cooling, and power infrastructure**. The system will be **integrated into the Sandia computing environment**, including network, security, and DevOps. The milestone is complete when the system is in production and available to users."

The completion criteria states,

> "The mini-Sierra ART system is **generally available** to Sandia code development teams and computational analysts. The system and its environment shall be **capable of supporting development and testing** of Sandia codes, including **nightly regression tests** (according to application readiness)."

The mini-Sierra ART system in the milestone description above is a 3-rack IBM cluster sited at SNL called **Vortex**, logically similar to the much larger ATS system **Sierra** sited at LLNL. Vortex has identical compute nodes, network topology, filesystem, and auxiliary equipment to the larger Sierra system, but at much smaller scale. Vortex was procured in support of Sierra by the Facility Operations and User Support (FOUS) program element, which supports the Department of Energy (DOE) National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.

The **Sierra** platform is the ASC program's second Advanced Technology System (ATS-2), and as of this writing, holds the second position on the world's top500 list of the most powerful computers, and is installed at Lawrence Livermore National Laboratory. The Sierra procurement also includes several smaller, compatible systems for building and testing.

The ATS systems are defined as part of ASC's Computing Strategy and are further supported by the Common Computing Environment (CCE), a program element of FOUS. This L2 milestone focused on the installation and initial standup of the platform, including hardware validation, software installation, performance testing, and application compatibility verification with Sandia codes.

The results from this Milestone were delivered in the following ways:

1. This SAND Report, which documents the bulk of all findings
2. Two presentations to stakeholders, at midpoint and end of fiscal year
3. A 2-page executive summary, and a handoff memo to primary customer

The configuration and procurement of **Vortex** platform is described in greater detail within Section 3.

Section 4 provides additional details for how the platform was sited, including facilities integration (Section 4.1), followed by hardware validation and baseline performance testing (Section 4.2).

Section 5 describes some of the main results, showing how the machine maintains compatibility with the Sierra ATS software environment and supports development and testing of Sandia codes (Section 5.1), including regression testing workloads. Some recent platform issues were discovered (Section 5.2), and this section includes a discussion of short-term mitigations applied and long-term solutions being developed.

Section 6 concludes with high-level findings and (Section 6.1) recommends ideas for future work.

3. CONFIGURATION AND PROCUREMENT

Initial discussions between SNL and LLNL about siting a Sierra ATS test machine at SNL began December 2017, and concluded with the purchase and delivery of Vortex July 2018, comprising a configuration and procurement cycle of about 8 months. In this cycle, SNL stakeholders from the 1400 Computing Research division (Si Hammond), the 1500 Engineering Sciences division (Mike Glass), and many from the 9000 Mission Assurance division (Tom Klitsner, Ann Gentile, Jim Brandt, others) met to determine the configuration and procurement for a machine that would be logically similar to Sierra at LLNL. The final configuration targeted compatibility with RzAnsel, an LLNL sited test machine built for Sierra ATS, and used the model of the successful ACES partnership that includes the locally sited Application Readiness Testbed (ART) system named Mutrino.

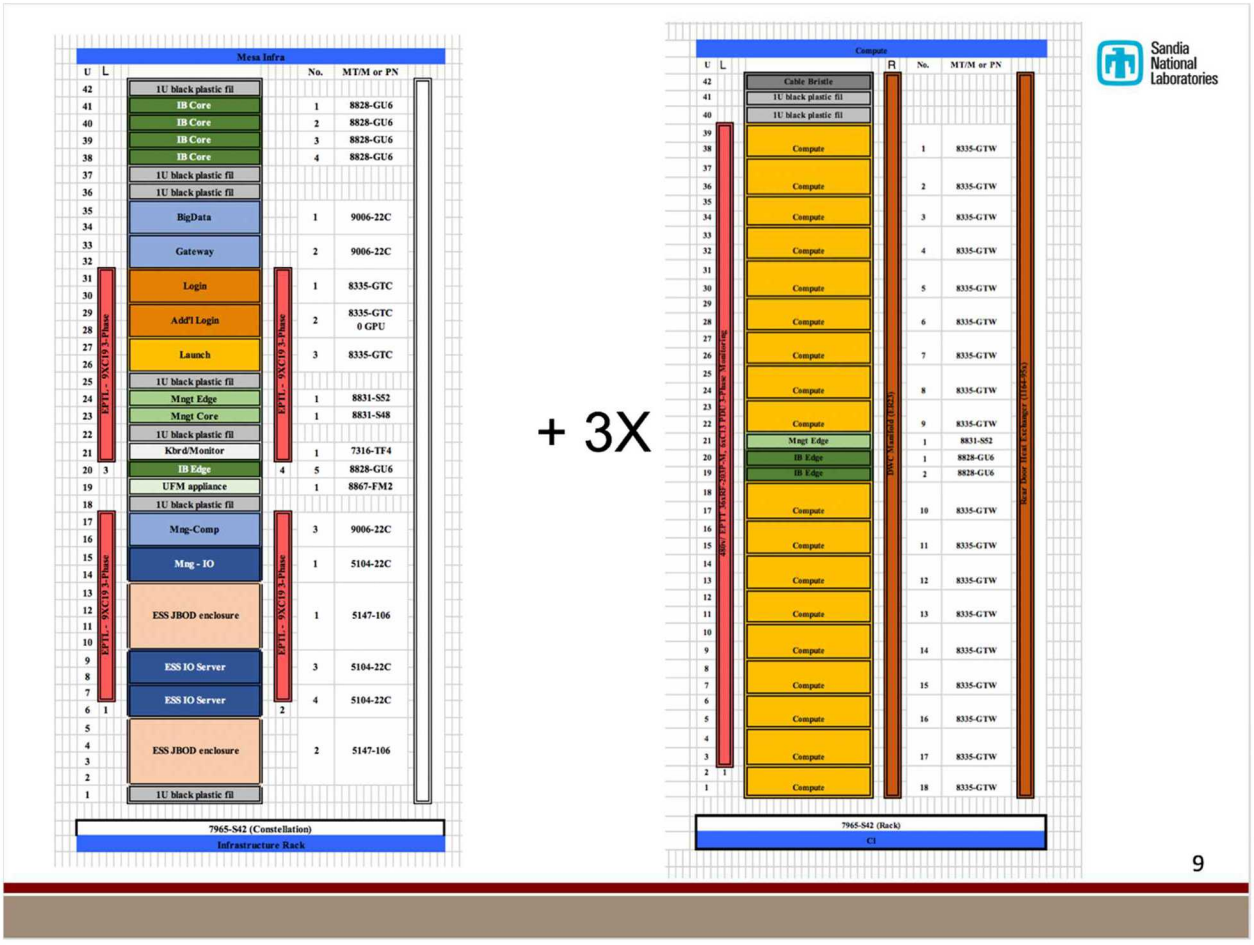


Figure 2: Vortex rack design for auxiliary infrastructure and compute cabinets matches Sierra ATS design

The design of Vortex infrastructure and compute racks matches that of RzAnsel and by proxy Sierra ATS. The Vortex infrastructure rack contains two IBM power9 "Boston" login nodes, a network filesystem Infiniband "gateway" node, a full fat tree topology high-speed EDR 100GB/s Infiniband network, an IBM Big Data log analysis server, and a 1.3 Petabyte GPFS parallel filesystem. There are

3 identical compute racks, comprising a total of 54 nodes, where each node has dual socket power9 CPUs, 22 cores/socket, 256GB system RAM, 4x Nvidia Tesla V100 GPU with a total 64GB GPU RAM, and 1.5 TB NVMe SSD local node burst buffer storage.

The Vortex racks were assembled at a technology demonstration in June 2018 before purchase at the IBM facility in Rochester, MN, where baseline functional, computational, and storage I/O tests were executed. The baseline functional tests demonstrated basic system management functions, storage functions, and network connectivity tests of all support nodes via command line diagnostics. The computational (HPL) and storage (IOR) tests were performed via the system job scheduler (LSF) after a full power cycle of the system. The computational tests consisted of running the STREAM diagnostic on both CPUs and GPUs to measure the memory bandwidth of each component and verifying no major discrepancies across components. In addition to this, Linpack (HPL) was run on both individual nodes separately, and on the system as a whole, to verify completion of computations with no residual errors or major performance discrepancies across components. The high-speed network was tested by running the Intel MPI benchmarks (IMB) to measure network throughput and latency with no major performance discrepancies across components. The storage I/O was tested by running IOR on the GPFS parallel filesystem with large object reads and writes, for N-1 and N-N mappings, using MPI I/O and POSIX to demonstrate functionality. No major I/O performance discrepancies were found across nodes.

4. INSTALLATION DETAILS

4.1. Facilities Integration

After initial verification of Vortex hardware at the factory, site preparation at SNL began July 2018, with final installation success October 2018. This work included initial requirements discussions between SNL facilities and IBM, chiefly addressing power and cooling concerns. An existing chiller unit was used to cool the Vortex water-cooled racks; however, extra water-cooled exhaust doors for compute racks had to be added to the procurement to keep the ambient air heat load within acceptable facilities parameters. Also, the compute racks required a total of 3x new 100Amp power connectors installed, which necessitated installing a new power service panel dedicated to this machine. By the end of October 2018, the facilities electrical work and plumbing was complete, and IBM came onsite to complete the hardware installation, including wiring the high-speed network and basic power and cooling tests.

4.2. Hardware Validation and Performance Testing

With the hardware and facilities integration of Vortex complete, IBM came onsite in November 2018 to install and test the CORAL HPC software stack used at LLNL and ORNL, including upgrading and testing the GPFS filesystem software. After the CORAL HPC software was installed, IBM executed full-scale optimized Linpack runs (HPL) to validate the configuration, resulting in a gross combined CPU and GPU performance of 1.06 Petaflop, which was the highest measured per-cabinet computing density of HPC machines installed in the data center.

The system was configured to operate with SNL enterprise infrastructure, including network and security configuration, and connected to SNL network user filesystems. In January 2019, the system was released to friendly users for initial testing using the IBM provided programming environment. In March 2019, the CORAL HPC software stack was updated to the latest tested by IBM, and initial tests of the burst buffer infrastructure were successful.

4.3. SNL and LLNL Collaboration

In March 2019, a contingent of SNL staff visited LLNL to meet with their HPC staff and leadership, to be briefed on LLNL customization of the system software and programming environment, to setup collaboration on reporting bugs to vendors, and to agree on methods and schedule for synchronizing the SNL and LLNL CORAL HPC environments. Regular monthly calls to synchronize the systems were started in June 2019. After meeting with LLNL staff, SNL staff were able to clone and begin testing of LLNL's customizations to the operating and programming environment of the CORAL HPC stack. In March 2019, SNL began testing the LLNL changes to the HPC job scheduler interface,

including verifying that commands such as lalloc, lrun, and others worked to the same specifications as listed in the official LLNL HPC new user's documentation. SNL regularly attends the Sierra ATS user group meetings to discuss system and performance issues, and we look forward to our continuing and close collaboration.

5. RESULTS

5.1. Sierra ATS Compatibility and Testing Sandia Codes

By May 2019, with the help and feedback of SNL SIERRA code team members (Sam Browne), a Sandia developed NNSA computer code SPARC was built on Vortex and successfully running regression tests after being copied to the LLNL machine RzAnsel, the unclassified Sierra ATS test machine. By June 2019, Vortex was successfully running nightly regression testing of the SPARC code, demonstrating technical milestone completion.

In July 2019, SNL stakeholders Si Hammond and Micah Howard report SPARC runs successfully on the classified network Sierra ATS, and are pleased with SPARC progress enabled by debugging and testing on the Vortex machine. In late July 2019, SIERRA code team member Rich Drake has verified he can run initial SIERRA code tests successfully on Vortex. In August 2019, another synchronization of the CORAL programming environment was completed, giving SNL teams access to the latest beta compilers and tools available on RzAnsel and LLNL Sierra ATS systems.

5.2. Platform Issues and Solutions

Some recent platform issues were discovered during the deployment and testing of Vortex. As background for this section, the CORAL software stack is comprised of a base RedHat Linux Operating System with many customizations. Such customizations include, for example, special nVidia GPU drivers, a custom kernel from RedHat, and special versions of Infiniband high-speed network drivers. None of these components are generally available to the public or supported through the normal corporate channels. Instead, members of the CORAL project can request special access and support from the individual vendors.

These customizations address at-scale performance and reliability issues generally only seen on huge machines like Sierra ATS. One at-scale customization is the use of a unique "launch node", which separates the normal two functions of the login node into two distinct hardware nodes, for: 1) compiling code and 2) starting multi-process jobs. This was necessary on Sierra ATS, as both compiling code and starting multi-process jobs on the login nodes rendered the login nodes unusable from the system load incurred. Another at-scale customization is the use of a sub-scheduler to coordinate compute nodes participating in an allocation as a scheduled job is started. This is accomplished by the CSM daemon, a custom piece of system code written by IBM to reduce the jitter and scale issues seen when LSF (the default IBM scheduler) attempted to spawn jobs directly to the Sierra ATS compute nodes. These two, and perhaps other, at-scale solutions for Sierra ATS disrupt the normal workflow of LLNL users whose codes and scripts are accustomed to a more traditional HPC topology. Therefore, LLNL has written a comprehensive array of wrapper scripts to give their HPC users an experience more like traditional HPC machines, with the scripts doing special work needed to allocate jobs and tune performance efficiently.

On Vortex, we have undertaken to educate our HPC users on the use of this new IBM topology and the new IBM commands and LLNL wrapper scripts that are necessary to use at SNL to maintain Sierra ATS compatibility. Additionally, some care is needed when synchronizing the CORAL software stack and LLNL programming environment, as both are under active development by IBM and LLNL. Especially, care is needed synchronizing the LLNL wrapper scripts, as these can unintentionally break the runtime environment and user scripts of SNL Vortex users.

In summary, we saw some unexpected platform issues arise when synchronizing Vortex to the unique topology and runtime environment of Sierra ATS. Most of the issues seen are related to the active development of Sierra ATS and related patching by IBM and LLNL to address at-scale and performance issues as they are arising. These kinds of issues will be addressed by implementing a more frequent patching schedule on Vortex. For the wrapper script issues in particular, we have developed a better testing regime at Sandia to catch these potential user-facing bugs earlier, and we will continue to use the process we developed with LLNL to push Sandia changes upstream, improving the portability and reliability of these wrapper scripts.

6. CONCLUSION

This section, we provide some high-level conclusions from the results of the previous section. A machine called Vortex was designed, procured, and successfully installed at Sandia National Laboratories. The CORAL software stack and LLNL programming environments were installed, tested, and upgraded several times during this milestone period. A Sandia code SPARC was demonstrated on Vortex by its successful compilation and execution of nightly regression tests. Compatibility with Sierra ATS was demonstrated by running and executing these same SPARC binaries and test scripts on RzAnsel and Sierra ATS. Additionally, other code teams at SNL such as EMPIRE, SALINAS, and SIERRA are actively developing and testing on Vortex, and we look forward to more interaction with them as they continue to use the platform.

For SNL stakeholders, there are unique benefits to a locally sited machine. As shown in this report, disruptive use cases such as nightly regression testing that requires use of large numbers of nodes are allowed on Vortex. The usage policies of LLNL on RzAnsel are more restrictive, as it also serves as a critical test and development machine for Sierra ATS shared among many user groups. Another local use we see is the use of graphical debuggers and performance analyzers, such as totalview and paraview. These applications have been usefully deployed on Vortex, and would be impractical or impossible to use remotely, such as to LANL and LLNL machines, whose internet gateways restrict such traffic and for which performance is extremely poor over wide-area networks.

These milestone activities have also been useful exercises for the platform teams at SNL. We have formed a close partnership with LLNL HPC staff, similar to the one we have with LANL to run the small-scale Trinity test machine (Mutrino) at SNL. Some unique challenges have been highlighted, including low latency patching and synchronization of runtime scripting needed between SNL and LLNL to maintain similar environments, despite frequent improvements and changes to the Sierra ATS software stack. We believe it has been beneficial for Sandia to be another national laboratory customer testing the CORAL software stack, as this helps grow the user and product testing base for CORAL, which otherwise would be limited to only a few customers.

6.1. Future Work

HPC machine users from SNL, LLNL and LANL can all request accounts on Vortex at SNL. For our tri-labs partners, this could help them diagnose platform issues, in order to rule out site-specific issues such as internal networking or enterprise service problems. There is more leverage to spur vendor action when multiple labs are having similar problems.

Other potentially disruptive machine uses we allow and hope to further investigate include: continuous integration services for code teams, dedicated system time for whole-machine runs, root-level access for experiments, dedicated login nodes and queues for critical code teams, and easy access to high-speed SNL HPC filesystems (for big data applications). We hope that the flexibility of the Vortex platform and the locality of the SNL code teams encourages diverse future uses that, while currently

unknown, serve to further and uniquely enhance the capability of Sierra ATS and related tri-labs efforts.

This page left blank

REFERENCES

[1] SAND2017-9933, “Application of Performance Analysis Tools on SNL ASC Codes”, Agelastos, A.; Pase, D; et. al; Sandia National Laboratories, 2017

[2] powerpoint presentation, "Vortex L2 Midyear Checkpoint", Repik, J; Gauntt, N; Hammond, S; Brandt, J; Gentile, A; Sandia National Laboratories; May 20, 2019

[3] powerpoint presentation, "Vortex L2 Final Checkpoint", Gauntt, N; Davis, K; Hammond, S; Gentile, A; Brandt, J; Sandia National Laboratories, August 19, 2019

APPENDIX A. IMPLEMENTATION PLAN

A scanned copy of this Milestone's Implementation Plan is provided in Figure A.1.

Milestone (ID#6814 ~~TBD~~): Site Preparation, Integration, and Operation of Mini-Sierra ART System		
Level: 2	**Fiscal Year**: FY19	**DOE Area/Campaign**: ASC
Completion Date: 9/30/19		
ASC nWBS Subprogram: FOUS		
Participating Sites: SNL		
Participating Programs/Campaigns: ASC		
Program Plan Target: Appendix J, target FOUS-1		
Description: Sandia will complete deployment and production availability of an Application Readiness Testbed (ART) system consisting of hardware and a software stack conforming to the Sierra ATS sited at LLNL. The system will support the porting and testing of Sandia's codes and their applications against this new architecture. Site preparation includes water cooling and power infrastructure. The system will be integrated into the Sandia computing environment, including network, security, and DevOps. The milestone is complete when the system is in production and available to users.		
Completion Criteria: The mini-Sierra ART system is generally available to Sandia code development teams and computational analysts. The system and its environment shall be capable of supporting development and testing of Sandia codes, including nightly regression tests (according to application readiness).		
Customer: ASC program and Sierra ATS platform users		
Milestone Certification Method A report is prepared as a record of milestone completion. The "handoff" of the developed capability (product) to a nuclear weapons stockpile customer is documented.		
Supporting Resources: IBM (platform vendor), Sandia Facilities, CSSE, ATDM, IC, Sandia DevOps		

Figure A-1. Scanned Implementation Plan for this Milestone that includes its description, completion criteria, and milestone certification methods.

APPENDIX B. EVIDENCE

This section contains hard-copy of relevant email and other documents demonstrating the claims listed in this report.

B.1. Email Correspondence

This section contains a copy of relevant email from stakeholders providing critical feedback concerning the L2 milestone.

Email Summary:

1. **12/5/2018** - Correspondence refers to HPL runs executed on 12/5/2018 that achieved gross FLOPS rate of 1.062 petaflops.

2. **1/17/2019** - Vortex platform generally available to all SNL users with approved WebCARS account requests

3. **3/21/2019** - Vortex updated with major revision (3/1) of the CORAL OS software stack

4. **4/11/2019** - Vortex updated with initial sync of the LLNL "module" programming environment

5. **6/7/2019** - Vortex HPL performance data submitted to top500.org supercomputer registry to appear Nov. 2019

6. **7/25/2019** - Correspondence refers to in-person interview with Sam Browne on 7/25/19, who reported SPARC code successfully runs nightly regression tests on Vortex.

7. **7/26/2019** - Si Hammond reports Micah Howard is happy with SPARC progress on Vortex

8. **7/29/2019** - Rich Drake, initial SIERRA (code) tests are running on Vortex

9. **8/16/2019** - Vortex updated with second sync of LLNL programming environment for 2019

Full Email Record Below:

> 1. **12/5/2018** - Correspondence refers to HPL runs executed on 12/5/2018 that achieved gross FLOPS rate of 1.062 petaflops.
>
> Note: this email has been shortened for clarity by removing redundant debug traces. Please email the author for an original copy.
> ===

From: "Repik, Jason J" <jjrepik@sandia.gov>
To: Vortex-Admin <Vortex-Admin@sandia.gov>
Subject: FW: Vortex HPL
Date: Thu, 6 Jun 2019 12:34:12 -0600

In case you can't find it.
-----Original Message-----
From: Repik, Jason J
Sent: Wednesday, December 05, 2018 12:32 PM
To: Repik, Jason J <jjrepik@sandia.gov>
Subject: Vortex HPL
[root@vortex60 RUN_216_531]# cat out.hpl.531
RANK 89 on host vortex3 PID 47154
RANK 125 on host vortex38 PID 47943
RANK 91 on host vortex3 PID 47156

*** Note: some debug traces removed for clarity ***

Applications clocks set to "(MEM 877, SM 1312)" for GPU 00000035:04:00.0
All done.
==
==============
HPLinpack 2.1 -- High-Performance Linpack benchmark -- October 26, 2012
Written by A. Petitet and R. Clint Whaley, Innovative Computing Laboratory, UTK
Modified by Piotr Luszczek, Innovative Computing Laboratory, UTK
Modified by Julien Langou, University of Colorado Denver
==
==============
An explanation of the input/output parameters follows:
T/V : Wall time / encoded variant.
N : The order of the coefficient matrix A.
NB : The partitioning blocking factor.
P : The number of process rows.
Q : The number of process columns.
Time : Time in seconds to solve the linear system.
Gflops : Rate of execution for solving the linear system.
The following parameter values will be used:
N : 1099520
NB : 1280
PMAP : Row-major process mapping
P : 9
Q : 24
PFACT : Right
NBMIN : 4
NDIV : 2
RFACT : Left
BCAST : 2ringM
DEPTH : 0
SWAP : Spread-roll (long)
L1 : no-transposed form
U : transposed form
EQUIL : no
ALIGN : 8 double precision words

- The matrix A is randomly generated for each test.
- The following scaled residual check will be computed:
||Ax-b||_oo / (eps * (|| x ||_oo * || A ||_oo + || b ||_oo) * N)

- The relative machine precision (eps) is taken to be 1.110223e-16
- Computational tests pass if scaled residuals are less than 16.0
trsm_cutoff from environment variable 9000000
gpu_dgemm_split from environment variable 1.000
monitor_gpu from environment variable 1
gpu_temp_warning from environment variable 70
gpu_clock_warning from environment variable 1312
gpu_power_warning from environment variable 350
chunk_size from environment variable 5120
ichunk_size from environment variable 1536
max_h2d_ms from environment variable 15
max_d2h_ms from environment variable 15
gpu_pcie_gen_warning from environment variable 3
gpu_pcie_width_warning from environment variable 2
test_system from environment variable 1
num_work_buf from environment variable 8
ranks_per_socket from environment variable 2
ranks_per_node from environment variable 4
print_scale from environment variable 1
sort_ranks from environment variable 0
fact_gemm from environment variable 1
fact_gemm_min from environment variable 256

*** Note: some debug traces removed for clarity ***

Prog= 99.14% N_left= 225280 Time= 777.59 Time_left= 6.75 iGF= 380508.59 GF= 1129843.86 iGF_per= 1761.61 GF_per= 5230.76
696 ** FACT GEMM: 0.00446686 s 1056.35 GF (without delay: 0.00441965 s 1067.64 GF 23040 320 320)
696 ** FACT GEMM: 0.00947627 s 1991.75 GF (without delay: 0.0094208 s 2003.48 GF 23040 640 640)
696 ** FACT GEMM: 0.00445027 s 1060.29 GF (without delay: 0.00439808 s 1072.88 GF 23040 320 320)
720 ** FACT GEMM: 0.00385408 s 1071.27 GF (without delay: 0.00380621 s 1084.75 GF 20160 320 320)
720 ** FACT GEMM: 0.00812096 s 2001.36 GF (without delay: 0.00806502 s 2015.24 GF 19840 640 640)
720 ** FACT GEMM: 0.00379529 s 1053.33 GF (without delay: 0.00374272 s 1068.13 GF 19520 320 320)
744 ** FACT GEMM: 0.00339001 s 1005.27 GF (without delay: 0.00334438 s 1018.98 GF 16640 320 320)
744 ** FACT GEMM: 0.00714816 s 1906.99 GF (without delay: 0.00709837 s 1920.37 GF 16640 640 640)
744 ** FACT GEMM: 0.00339423 s 1004.02 GF (without delay: 0.00334643 s 1018.36 GF 16640 320 320)
768 ** FACT GEMM: 0.0026823 s 977.31 GF (without delay: 0.00263674 s 994.199 GF 12800 320 320)
768 ** FACT GEMM: 0.00568218 s 1845.38 GF (without delay: 0.00563504 s 1860.81 GF 12800 640 640)
768 ** FACT GEMM: 0.00265849 s 986.063 GF (without delay: 0.00260608 s 1005.89 GF 12800 320 320)
Prog= 99.88% N_left= 115200 Time= 809.67 Time_left= 0.93 iGF= 205805.28 GF= 1093228.59 iGF_per= 952.80 GF_per= 5061.24
792 ** FACT GEMM: 0.0022109 s 918.908 GF (without delay: 0.00216371 s 938.949 GF 9920 320 320)
792 ** FACT GEMM: 0.00462524 s 1700.31 GF (without delay: 0.00457011 s 1720.82 GF 9600 640 640)
792 ** FACT GEMM: 0.00209884 s 905.52 GF (without delay: 0.00205107 s 926.61 GF 9280 320 320)
816 ** FACT GEMM: 0.00158922 s 824.755 GF (without delay: 0.00154726 s 847.121 GF 6400 320 320)
816 ** FACT GEMM: 0.0034166 s 1534.53 GF (without delay: 0.00336896 s 1556.23 GF 6400 640 640)
816 ** FACT GEMM: 0.00155677 s 841.949 GF (without delay: 0.0015104 s 867.797 GF 6400 320 320)
840 ** FACT GEMM: 0.000973263 s 538.691 GF (without delay: 0.000932864 s 562.02 GF 2560 320 320)
840 ** FACT GEMM: 0.00247357 s 847.823 GF (without delay: 0.0024361 s 860.866 GF 2560 640 640)
840 ** FACT GEMM: 0.000802936 s 652.964 GF (without delay: 0.000746496 s 702.332 GF 2560 320 320)
Prog= 100.00% N_left= 5120 Time= 832.94 Time_left= 0.00 iGF= 43794.53 GF= 1063909.54 iGF_per= 202.75 GF_per= 4925.51
2018-12-05 12:21:11.481
==
T/V N NB P Q Time Gflops

```
WR03L2R4 1099520 1280 9 24 834.74 1.062e+06
--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV-
Max aggregated wall time rfact . . . : 22.29
+ Max aggregated wall time pfact . . : 12.34
+ Max aggregated wall time mxswp . . : 8.85
Max aggregated wall time update . . : 304.01
+ Max aggregated wall time laswp . . : 279.68
+ Max aggregated wall time spread . : 154.42
+ Max aggregated wall time roll . . : 108.29
Max aggregated wall time bcast . . . : 624.39
Max aggregated wall time up tr sv . : 1.23
----------------------------------------------------------------------------
||Ax-b||_oo/(eps*(||A||_oo*||x||_oo+||b||_oo)*N)= 0.0001628476 ..... PASSED
============================================================================
===============
Finished 1 tests with the following results:
1 tests completed and passed residual checks,
0 tests completed and failed residual checks,
0 tests skipped because of illegal input values.
----------------------------------------------------------------------------
End of Tests.
============================================================================
===============
===
```

2. **1/17/2019** - Vortex platform generally available to all SNL users with approved WebCARS account requests

```
===
From: "Gauntt, Nathan E" <negaunt@sandia.gov>
To: Vortex-Users <Vortex-Users@sandia.gov>
CC: Vortex-Admin <Vortex-Admin@sandia.gov>
Subject: Vortex open to users
Date: Thu, 17 Jan 2019 13:05:00 -0700

Users,
The initial programming environment for the power9 Vortex cluster has been verified and is now open and
available to users.
To login:
> ssh vortex.sandia.gov
Please see /opt/VORTEX_INTRO for usage. Please email any issue to vortex-help@sandia.gov. Good hunting!
Thank you,
-Nate
===
```

3. **3/21/2019** - Vortex updated with major revision (3/1) of the CORAL OS software stack

```
===
From: "Gauntt, Nathan E" <negaunt@sandia.gov>
To: Vortex-Users <Vortex-Users@sandia.gov>
CC: Vortex-Admin <Vortex-Admin@sandia.gov>
Subject: Vortex DST over
Date: Thu, 21 Mar 2019 17:08:01 -0600
Users,
Vortex cluster is back up and returned to service. We have upgraded the software environment to match the
3/1/2019 release of the Coral software stack (LLNL Sierra). Please email any issues to vortex-help@sandia.gov.
```

Thank you,
-Nate
===

4. **4/11/2019** - Vortex updated with initial sync of the LLNL "module" programming environment

===
From: "Gauntt, Nathan E" <negaunt@sandia.gov>
To: Vortex-Users <Vortex-Users@sandia.gov>
CC: Vortex-Admin <Vortex-Admin@sandia.gov>
Subject: new modules
Date: Thu, 11 Apr 2019 17:35:47 -0600
Users,
A new module environment is available by default on Vortex login, launch, and computes. This environment should closely match the environment available on LLNL's Sierra platform and related systems. A few modules, such as xl, cuda, and spectrum-mpi are loaded by default, as on Sierra-like platforms.
Use the familiar 'module' commands as before, but note the implementation has changed as Lmod is the new underlying module system.
Please let us know if you have any questions. Email any issues to vortex-help@sandia.gov.
Thank you,
-Nate

vortex # module list
Currently Loaded Modules:
1) xl/2019.02.07 3) cuda/9.2.148
2) spectrum-mpi/rolling-release 4) StdEnv

vortex2 # module av | head -n 10
---------- /usr/tce/modulefiles/Compiler/xl/2019.02.07 ----------
spectrum-mpi/rolling-release (L,D) spectrum-mpi/2018.08.30
spectrum-mpi/2018.04.27 spectrum-mpi/2018.10.10
spectrum-mpi/2018.06.01 spectrum-mpi/2018.11.14
spectrum-mpi/2018.06.07 spectrum-mpi/2018.12.14
spectrum-mpi/2018.07.12 spectrum-mpi/2019.01.18
spectrum-mpi/2018.08.02 spectrum-mpi/2019.01.22
spectrum-mpi/2018.08.13 spectrum-mpi/2019.01.30
===

5. **6/7/2019** - Vortex HPL performance data submitted to top500.org supercomputer registry to appear Nov. 2019

===
From: "Gauntt, Nathan E" <negaunt@sandia.gov>
To: "Gentile, Ann" <gentile@sandia.gov>, "Hammond, Simon David" <sdhammo@sandia.gov>, "Repik, Jason J" <jjrepik@sandia.gov>
CC: Vortex-Admin <Vortex-Admin@sandia.gov>
Subject: Re: Vortex HPL
Date: Fri, 7 Jun 2019 15:33:47 -0600

> This message is in MIME format. Since your mail reader does not understand
this format, some or all of this message may not be legible.
--B_3650135185_1427748715
Content-type: text/plain;
charset="UTF-8"
Content-transfer-encoding: quoted-printable

Hi All,
Si helped me submit Vortex top500 numbers today. It may not make the list=
until November 19. See attached!
Thanks,
-Nate

-----Original Message-----
From: "Gentile, Ann" <gentile@sandia.gov>
Date: Thursday, June 6, 2019 at 9:13 PM
To: Si Hammond <sdhammo@sandia.gov>, "Gauntt, Nathan E" <negaunt@sandia.gov=
>, "Repik, Jason J" <jjrepik@sandia.gov>
Cc: Vortex-Admin <Vortex-Admin@sandia.gov>
Subject: Re: Vortex HPL
This is a very cool idea!=20
Good luck!
=20
Ann

_____
From: Hammond, Simon David
Sent: Thursday, June 6, 2019 9:08 PM
To: Gauntt, Nathan E; Repik, Jason J
Cc: Vortex-Admin
Subject: Re: Vortex HPL
=20
I can help you - then next time you'll know what to do. We may have mis=
sed the deadlines but let's see what happens.
=20
Are you in tomorrow?
=20
=20
--
Si Hammond
Scalable Computer Architectures
Sandia National Laboratories, NM, USA
[Sent from remote connection, excuse typos]
=20
=20
On 6/6/19, 3:36 PM, "Gauntt, Nathan E" <negaunt@sandia.gov> wrote:
=20
Sure, are there instructions somewhere?
=20
Thanks,
-Nate
=20
-----Original Message-----
From: Si Hammond <sdhammo@sandia.gov>
Date: Thursday, June 6, 2019 at 12:36 PM
To: "Repik, Jason J" <jjrepik@sandia.gov>
Cc: Vortex-Admin <Vortex-Admin@sandia.gov>
Subject: Re: Vortex HPL
=20
Thanks Jason.
=20
Nate - do you want to coordinate the submission? Its probably g=
ood experience.
=20
=20

=E2=80=94
Si Hammond
Scalable Computer Architectures
Sandia National Laboratories, NM
=20
===

6. **7/25/2019** - Correspondence refers to in-person interview with Sam Browne on 7/25/19, who reported SPARC code successfully runs nightly regression tests on Vortex.

===
From: "Gauntt, Nathan E" <negaunt@sandia.gov>
To: "Gentile, Ann" <gentile@sandia.gov>
CC: Vortex-Admin <Vortex-Admin@sandia.gov>
Subject: L2 update
Date: Thu, 25 Jul 2019 17:04:31 -0600
Hi Ann,
I've interviewed Sam from sierra about user/code/compatibility experience on Vortex. The summary I have so far, is that 100% of build and runtime tests of Sparc are running nightly on Vortex. The process is, at 6pm each night, builds are wiped clean, pulled from git, rebuilt, runtime tested, and then pushed to rzansel/llnl. Their builds are finishing in a reasonable amount of time, before morning.
The story with sierra (code) is less far along. He can get sierra to manually build on Vortex right now, but automated build/regression testing are not in place yet; however, this activity is expected to be ramping up in the next several weeks.
Before talking about any of this, he suggested we loop in Rich Drake his team lead, and for the Sparc info, Micah Howard. So this is all preliminary info, please do not officially publicize.
Thanks,
-Nate
===

7. **7/26/2019** - Si Hammond reports Micah Howard is happy with SPARC progress on Vortex

===
From: "Hammond, Simon David" <sdhammo@sandia.gov>
To: "Gauntt, Nathan E" <negaunt@sandia.gov>
CC: "Gentile, Ann" <gentile@sandia.gov>, "Brandt, James M"
<brandt@sandia.gov>, Vortex-Admin <Vortex-Admin@sandia.gov>
Subject: Re: Vortex L2 feedback
Date: Fri, 26 Jul 2019 10:53:40 -0600
Nate,
I think things are progressing fairly well on the Micah/SPARC front. He has been pretty happy with the system in the conversations I have had with him.
There is a persistent anxiety of how closely we are tracking the LLNL updates, particular RZAnzel in 1400. Because we are tracking Lassen (on LLNL's recommendation) everyone is worried we will diverge from Anzel where most of our code work needs to take place because of Export Control rules.
In the review, I think we should surface this issue - we are following recommended practice from LLNL because they believe this will give the most stable variant of the system. The period of difference between the platforms should be small and is, by design in LLNL, to give them the ability to roll back.
Overall, I'd say everyone thinks Vortex is going well.
S.
--
Si Hammond
Scalable Computer Architectures
Sandia National Laboratories, NM, USA
[Sent from remote connection, excuse typos]
On 7/25/19, 12:08 PM, "Gauntt, Nathan E" <negaunt@sandia.gov> wrote:

Hi Si,
August 19 is our final L2 checkpoint meeting for Vortex. Ann had suggested you may know if Micah Howard or Mike Glass were having issues or success with Vortex builds running on Sierra (platform). I will be talking to other Sierra folk this week (Sam, Mark, Rich) to get their feedback on Vortex relative to the lower-level details for the non-classified LLNL systems.

Do you know of any higher-level issues or successes Micah and Mike are having with Vortex?

And is there any 1400 Vortex feedback you would like us to know about?

Thanks,
-Nate
===

8. **7/29/2019** - Rich Drake, initial SIERRA (code) tests are running on Vortex

===
From: "Drake, Richard R" <rrdrake@sandia.gov>
To: "Gauntt, Nathan E" <negaunt@sandia.gov>
Subject: Re: Vortex/Sierra status
Date: Mon, 29 Jul 2019 16:53:44 -0600
Hi Nate,
Thanks for reaching out. We can meet up whenever. I just tried compiling and running tests on Vortex and tests are running. The team probably has more experience than I do, at least a few of them.
Does the queueing system have a high latency? That is my only concern at this point (looking forward to running the plethora of out tests __.
-rich
-----Original Message-----
From: "Gauntt, Nathan E" <negaunt@sandia.gov>
Date: Friday, July 26, 2019 at 12:53 PM
To: Richard Drake <rrdrake@sandia.gov>
Subject: Vortex/Sierra status
Hi Rich,
We're getting ready for the final Vortex L2 milestone report on 8/19, and I'm curious how well you and your team are running on Vortex.
Are you having any immediate issues/concerns building or running sierra on Vortex?
Are there any near-term issues you expect in the next several weeks?
Kevin Davis and I have been working closely with Sam Browne and crew to rapidly address smaller issues (compilers, debuggers, editors, sierra mounts, etc.), and Sam has said this has been very helpful in his ability to make good progress on the platform. We're planning another LLNL sync the week of 8/5 to bring in some compiler updates currently available on rzAnsel.
I'm out for vacation next week, but will be back on 8/5. If you're available after 8/5, I'd like to setup a meeting with you and our team and get your feedback on the platform. Please let me know if you're available that week, and I'll set something up.
Thanks,
-Nate
(aka. vortex-admin@sandia.gov)
===

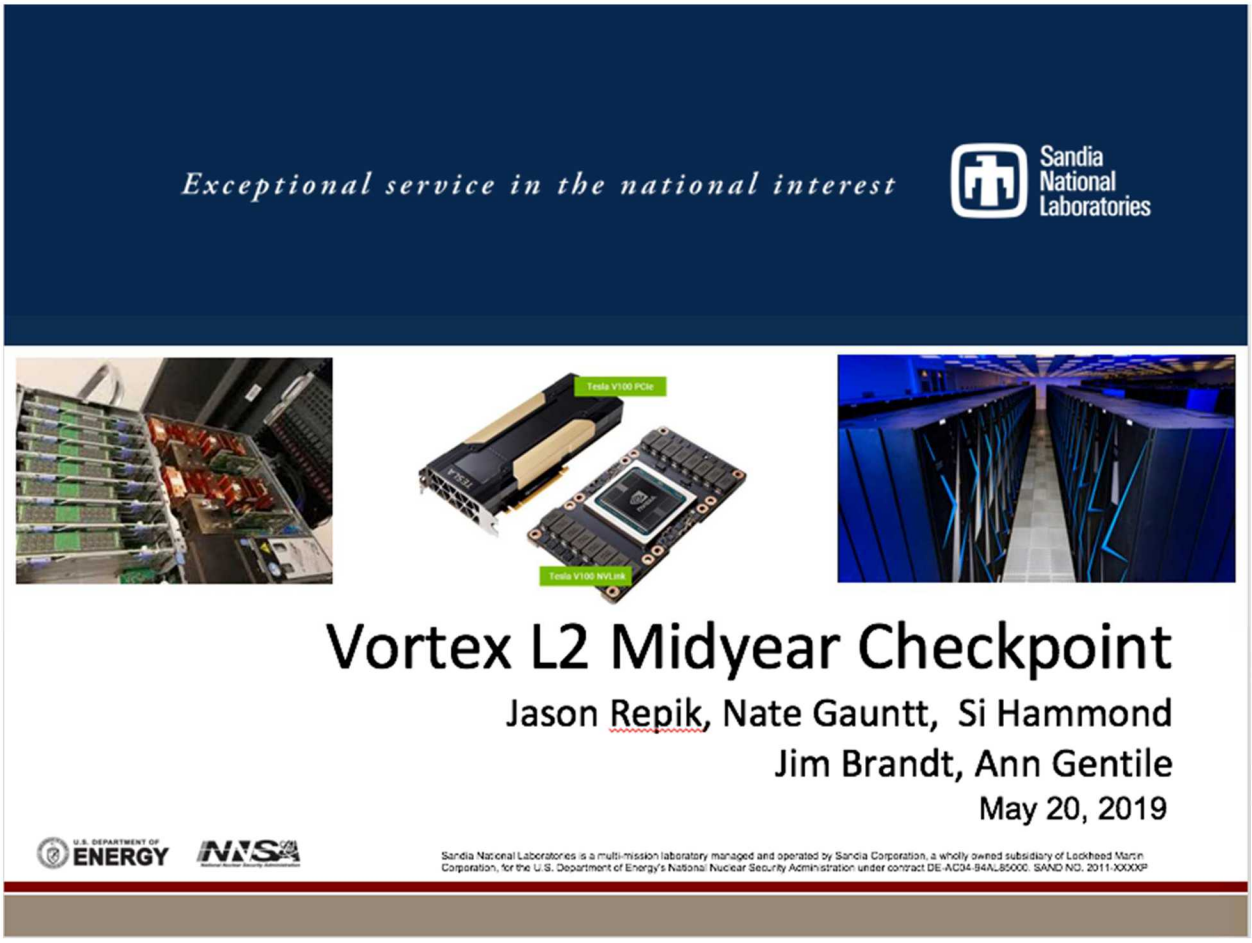
9. **8/16/2019** - Vortex updated with second sync of LLNL programming environment for 2019

===
From: "Gauntt, Nathan E" <negaunt@sandia.gov>
To: Vortex-Users <Vortex-Users@sandia.gov>
CC: Vortex-Admin <Vortex-Admin@sandia.gov>
Subject: new Vortex prog env
Date: Fri, 16 Aug 2019 17:05:32 -0600

Users,
Vortex has been updated with the latest module programming environment today. This environment matches what is currently available on rzansel and Lassen, including many new compiler versions.
The default module set loaded has not changed, to use these new compilers, please use 'module load' as per usual.
We have several nodes still down for triage. As a result, you will only be able to use 51 nodes for the near-term.
Please email any issues to vortex-help@sandia.gov.
Thank you,
-Nate
===

B.2. Midpoint Presentation

This section contains a copy of the "Vortex L2 Midyear Checkpoint", provided to the 9328, 1400, and 1500 stakeholders on 5/20/2019.



Overview:

Sandia National Laboratories

- **Milestone Description**
- **Timeline**
- **Current Progress**
  - **System Definition, Procurement, Demonstration**
  - **Site Prep**
  - **Delivery and Standup**
  - **Acceptance Testing**
  - **Release to Users**
  - **Software Upgrades**
  - **LLNL Coordination**
  - **Software Stack**
- **Future Work**

Milestone Description (#6814)

Sandia National Laboratories

- **Description:** Sandia will complete **deployment and production availability** of an Application Readiness Testbed (**ART**) system consisting of hardware and a software stack **conforming to the Sierra ATS sited at LLNL**. The system will support the porting and testing of Sandia's codes and their applications against this new architecture. Site preparation includes **water, cooling, and power infrastructure**. The system will be **integrated into the Sandia computing environment**, including network, security, and DevOps. The milestone is complete when the system is in production and available to users.

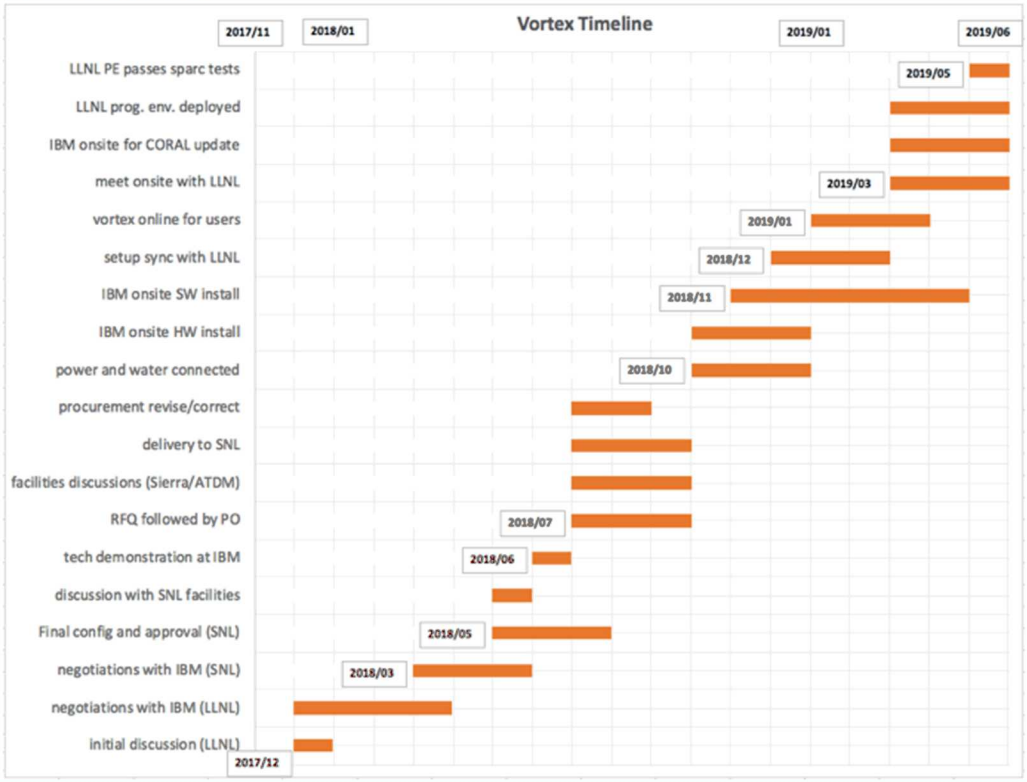
- **Completion Criteria:** The mini-Sierra ART system is **generally available** to Sandia code development teams and computational analysts. The system and its environment shall be **capable of supporting development and testing** of Sandia codes, including **nightly regression tests** (according to application readiness).

3

Timeline

Dec. 2017 – May 2019

Sandia National Laboratories



Timeline (detail)

Sandia National Laboratories

- 2017/12 – initial discussion with LLNL
- 2017/12 – 2018/03 – negotiations with IBM through LLNL
- 2018/03 – 2018/06 – negotiations with IBM directly
- 2018/05 – Final configuration determination 932X+Glass+Hammond. Approval to move ahead from Klitsner+Collis
- 2018/05 – Initial discussion with SNL facilities & IBM hardware
- 2018/06 – Technology Demonstration at IBM
- 2018/07 – RFQ followed by PO
- 2018/07 – Priority setting with Sierra and ATMD (Glass and Hammond) re other facility work that would delay Vortex stand up
- 2018/07 – Delivery to SNL
- 2018/07-09 – Procurement Revisions, missing switch, missing fat-tree network config
- 2018/10 – Power & water connected to IBM racks
- 2018/10 – IBM onsite for Hardware install
- 2018/11 – IBM onsite for initial software install/setup
- 2018/12 – Negotiations with LLNL for system sync process
- 2019/01 – Vortex online with initial IBM software available for users
- 2019/03 – on-site meeting with LLNL for initial system sync coordination
- 2019/03 – IBM onsite for first Coral software stack update
- 2019/03 – Initial deployment of LLNL programming environment
- 2019/05 – LLNL environment successfully tested with Sparc regression pass

6

Current Progress

Sandia National Laboratories

- System Definition
  - Requirements discussion with Si Hammond and Mike Glass
    - Design is for small-scale LLNL compatibility (RZAnsel)
    - Similar proportion as (Mutrino : Trinity) ratio
  - 4 cabinets
    - Infrastructure rack: 1 / 4
      - full fat tree EDR 100 Gb/s Infiniband High-Speed Network
      - Big Data log analysis box
      - IB gateway, User login, Sierra login, GPFS (1.3 PB) nodes
    - Compute racks: 3 / 4
      - 54 nodes, power9, dual socket, 22 cores/socket
      - 318 GB RAM, 1.5 TB NVMe (BB), 1.3 PB Disk (GPFS)
    - High-Speed Network
      - full fat tree EDR 100 Gb/s Infiniband High-Speed Network

Current Progress

Sandia National Laboratories

- Procurement Negotiations
  - Initial interactions with IBM and LLNL to see if we could get the CORAL pricing (we could not)
  - Various rounds of negotiating/assessing:
    - 1, 2, or 3 racks since SNL budget was unconfirmed
    - GPU pricing variation with time
    - CORAL RHEL pricing vs site license
    - Explicit schematic diagram including roles of all components of the system
    - Support coverage and Warranty
    - Stability of pricing given gaps in IBM provided materials
  - ORNL asked for acquisitions costs for DOE ASCR (7/2018)

Current Progress

Sandia National Laboratories

- Pre-Purchase Technology Demonstration
  - Visit IBM to review platform functionality
    - 1. Baseline Functional Tests
    - 2. Baseline Computational Tests
    - 3. Baseline I/O and Storage Tests
  - The following 3 Tests were given to IBM as part of acceptance criteria

Current Progress

Sandia National Laboratories

- Pre-Purchase Technology Demonstration
  - 1. Purpose of Baseline Functional Tests:
    - Demonstrate basic system management functionality.
    - Demonstrate basic functionality, storage functionality, and network connectivity of all support nodes. This will be done by logging into the node and exercising the functionality via command line or simple diagnostic.
    - All computational and storage tests will be performed via the system scheduler.
    - A full system shutdown/power off, followed by full system bring up to the point of running jobs

12

Current Progress

Sandia National Laboratories

- Pre-Purchase Technology Demonstration
  - 2. Purpose of Baseline Computational Tests:
    - Run STREAM on both nodes and GPUs (memory bandwidth performance test) to measure individual node memory bandwidth. Benchmarks must run to completion with no major discrepancies across components.
    - Run Linpack (HPL) in two modes:
      - » single node performance (Flops) test across all nodes and GPUs separately.
      - » multi-node performance test across all or a substantial percentage of the system.

      The benchmarks must run to completion and pass the residual test. No major performance discrepancies across components in individual and multi-node performance test.
    - Run Intel MPI benchmarks (IMB) and measure network throughput and latency speeds. Benchmarks must run to completion with no major performance discrepancies across components. This will likely run across the cluster in small groupings of nodes that are to be determined..

13

Current Progress

Sandia National Laboratories

- Pre-Purchase Technology Demonstration
  - 3. Purpose of Baseline I/O and Storage Tests:
    - Run IOR with large object writes and reads, for N-1 and N-N using MPI I/O and POSIX to demonstrate functionality. No major performance discrepancies across nodes.

14

Current Progress

Sandia National Laboratories

- Pre-Purchase Technology Demonstration
  - Demonstrate basic features of the system are functional 2018/06
- Large STREAM Benchmarking
  - Apprximately 85% of system memory must be used
  - Demonstrate all DIMMs are functional, installed correctly and the processor can maintain high-performance timing for access
  - All nodes reported successful STREAM numbers (approx. 240GB/s)
- HPL Benchmarking on Processors and GPUs
  - Ran a relatively small problem on CPUs and GPUs to ensure fast(er) testing (approximately 15 – 30 minutes)
  - Sufficient to demonstrate thermal load
  - Run in parallel to show MPI connectivity is functional
- IMB (MPI) Benchmarking
  - Performs large range of MPI tests across various permutations of ranks
  - Checks basic Infiniband connectivity and MPI software stack
  - Showed bug in MPI_Gather for IBM Spectrum
  - All other operations successful and showed good host-NIC performance

15

Current Progress

Sandia National Laboratories

- Site Prep
  - Facilities planning and work 2018/07 – 2018/10
    - Arranged initial requirements discussion with SNL facilities and IBM
    - 3x 100A connections required for IBM compute racks
      - New power panel in X50 required for 3x 100A connections
    - Existing CDU would be sufficient for new IBM system
    - Extra cooling doors added based on heat load to meet facilities requirements
    - 4x Network connections required for new IBM system

17

Current Progress

Sandia National Laboratories

- System Delivery and Initial Standup
  - System delivered to SNL 2018/07
    - All racks staged in holding area (site/facilities delay for Astra)
    - Electrical work & plumbing completed
    - IBM system connected to facilities power & water
    - IBM onsite to complete hardware installation
  - IBM onsite for software install 2018/10
    - Latest LLNL Coral stack download
    - IBM installed Coral stack/update
    - IBM configure GPFS filesystem

18

Current Progress

Sandia National Laboratories

- Acceptance Testing
  - See: Acceptance testing done in Rochester
  - SNL validation on-site with Sean McCombe (IBM) 2018/11
    - 12/4: full-machine HPL results: **1.06** PFlop
    - comparison with Astra: **1.52** PFlop
    - note: > 75% Vortex Flops from GPU

19

Current Progress

Sandia National Laboratories

- Release to Users
  - System received SNL localizations
    - Vortex Admin team configures (authentication, network, etc.)
    - Vortex mounts configured for /home & /projects
      - Shared NFS space from CapViz centralized NFS servers
  - Early release to friendly users 2019/01

Current Progress

Sandia National Laboratories

- System Software Upgrades
  - Coral software stack update 2019/03
    - IBM site visit to observe and help with update
    - Vortex Admin team completes update
    - Initial IBM burst buffer implementation installed

Current Progress

Sandia National Laboratories

- LLNL Coordination
  - Trip to LLNL 2019/03
    - Meet with LLNL Admin team: Scott, Adam, Py, Matt, John
    - Get accounts on LLNL development systems
    - Briefings on LLNL software modifications
  - Agreements
    - Collaboration on software updates, bug reporting, admin time
    - Regular monthly synchronization meetings with LLNL, starting 6/19

23

Current Progress

Sandia National Laboratories

- Vortex/LLNL Software Stack
  - LLNL scripting working, lalloc, lrun, etc. 2019/03
    - verified on Vortex using LLNL's official user guide
  - LLNL TCE modules working
    - verified by sierra (SNL) sparc builds on rzansel 2019/05
  - Nvidia drivers: SNL has direct access now to Coral drivers
  - SNL config vs. LLNL
    - LLNL: BlueOS = IBM RHEL + TCE + TOSS built for PPC
    - SNL: BlueOS' = IBM RHEL + TCE + minimal TOSS pkgs.
    - working on licensing, adding SNL to LLNL's build farm
  - Flux – initial discussions and collaboration

25

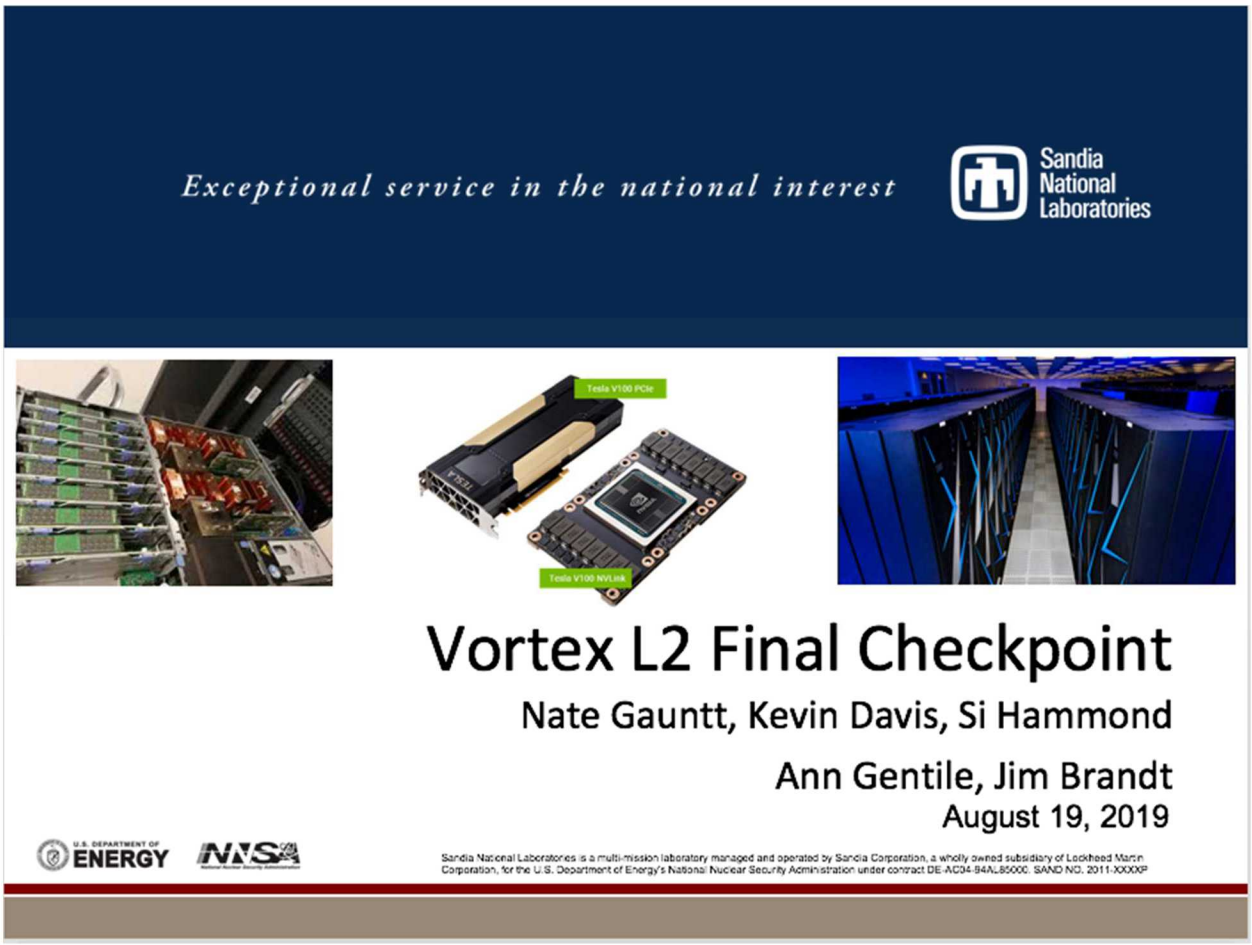
Future Work and Conclusions

Sandia National Laboratories

- Continuous work on the LLNL programming environment (PE)
- Full scale evaluation of scheduler (LSF) and user environment
  - LSF scheduler slow, CSM tuning issues
- Vortex is open to all users
  - Direct interaction with Sparc resolving LLNL/SNL PE discrepancies
  - SNL Sierra developers (Sam Browne) actively using and testing LLNL PE
  - Sparc regression tests passing on LLNL (RzAnsel) 2019/05

B.3. Year-End Presentation

This section contains a copy of the "Vortex L2 Final Checkpoint", provided to the 9328, 1400, and 1500 stakeholders on 8/19/2019.



Overview:

Sandia National Laboratories

- **Milestone Description**
- **Timeline**
- **Results**
- **Known Issues**
- **LLNL Collaboration Details**
- **Conclusions and Future Work**

Milestone Description (#6814)

Sandia National Laboratories

- **Description:** Sandia will complete **deployment and production availability** of an Application Readiness Testbed (**ART**) system consisting of hardware and a software stack **conforming to the Sierra ATS sited at LLNL**. The system will support the porting and testing of Sandia's codes and their applications against this new architecture. Site preparation includes **water, cooling, and power infrastructure**. The system will be **integrated into the Sandia computing environment**, including network, security, and DevOps. The milestone is complete when the system is in production and available to users.

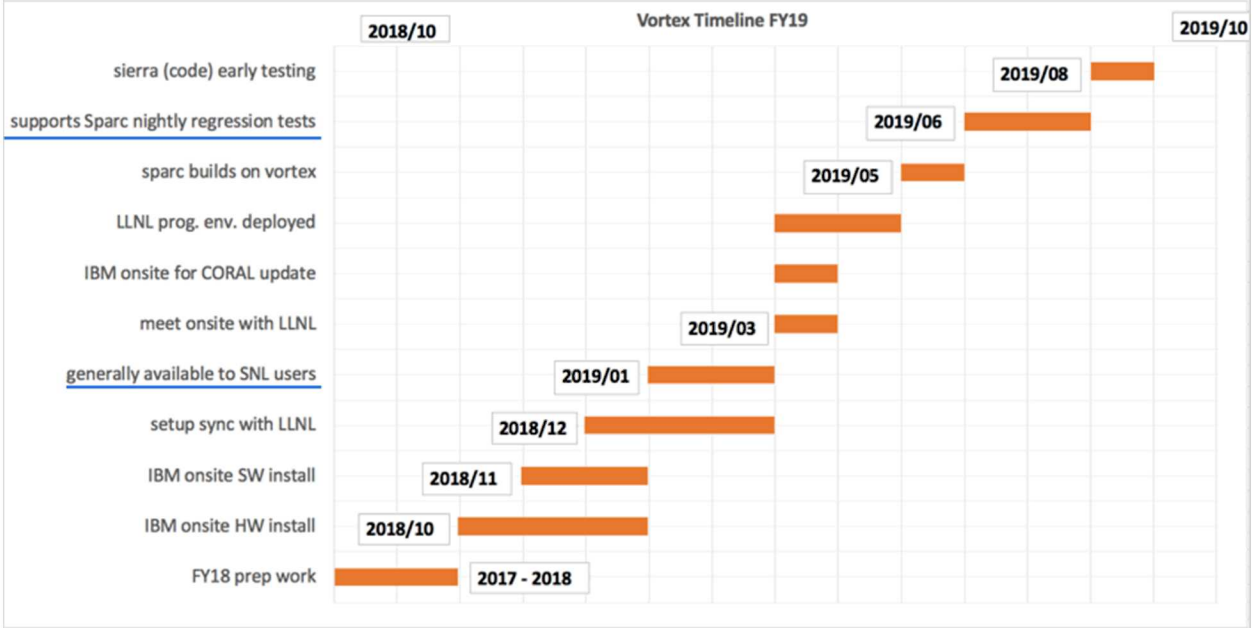
- **Completion Criteria:** The mini-Sierra ART system is **generally available** to Sandia code development teams and computational analysts. The system and its environment shall be **capable of supporting development and testing** of Sandia codes, including **nightly regression tests** (according to application readiness).

3

Timeline

FY19: Oct. 2018 – Aug. 2019

Sandia National Laboratories



Timeline (detail)

Sandia National Laboratories

- 2017 - 2018 – FY18 prep work
- 2018/10 – IBM onsite for Hardware install
- 2018/11 – IBM onsite for initial software install/setup
- 2018/12 – Negotiations with LLNL for system sync process
- 2019/01 – **generally available** to SNL users
- 2019/03 – on-site meeting with LLNL for initial system sync coordination
- 2019/03 – IBM onsite for first Coral software stack update
- 2019/03 – Initial deployment of LLNL programming environment
- 2019/05 – LLNL environment successfully tested with Sparc builds
- 2019/06 – **supports nightly regression tests** of Sparc
- 2019/08 – Early Sierra (code) testing on Vortex
- 2019/08 – Sync latest LLNL programming environment

5

Results

Sandia National Laboratories

- Milestone Criteria met
  - 2019/01 – **generally available** to SNL users
  - 2019/06 – **supports nightly regression tests** of Sparc

- Supporting Evidence
  - Sierra ATS compatibility
    - latest LLNL PE/compilers available on Vortex
    - Sparc built on Vortex runs on rzAnsel and Sierra ATS
  - User feedback
    - 7/25: Sam Browne: Sparc git pull, compile, nightly tests running
    - 7/26: Si Hammond: Micah Howard happy with Sparc progress
    - 7/29: Rich Drake: can compile/run initial sierra (code) tests

Known Issues

Sandia National Laboratories

- Scheduler job startup is slow
  - 90 sec job startup, automated testing worries
  - fixed in latest OS patches (in CSM)
- Cuda10 support
  - experimental at LLNL
  - available for small-scale at SNL (White, Ride)
  - will be supported on Vortex in FY19
- My Sierra ATS scripts don't work at SNL…
  - binary compatibility, but …
  - many LLNL wrappers for native IBM tools (scheduler, MPI)
  - performance tuning, scale issues, LLNL user expectations
  - updates may require SNL users to rewrite scripts

LLNL Collaboration Details

Sandia National Laboratories

- LLNL Programming Environment: TCE
  - Tri-labs Computing Environment: compilers, runtime, wrappers
  - 295 GB, backwards compatible, can be rsync'd
  - first available at SNL 2019/04, updated 2019/08
  - sync schedule: compilers vs. wrappers

- LLNL Compute Environment: BlueOS
  - BlueOS: IBM RHEL, drivers + LLNL built TOSS rpms (PE, tools)
  - short term: give SNL rpms, images built at SNL
  - long term: build image at LLNL, accept SNL changes (plan of record)
  - sync schedule: 2018/12, 2019/03, (TBD 2019/09), follows LLNL
  - both at 2019/03 base, LLNL ahead on at-scale patches

8

LLNL Collaboration Details

Sandia National Laboratories

- Sync to Lassen or rzAnsel?
  - rzAnsel: latest compilers, small system, may have bugs
  - Lassen: working compilers, larger system, closer to Sierra ATS
  - currently syncing with rzAnsel
  - Only difference is PE, OS is identical
- Other LLNL Interactions
  - regular monthly sync call with SNL
  - SNL changes accepted upstream in wrappers
  - Sierra ATS user group: performance, tuning, system updates

9

Conclusions and Future Work

Sandia National Laboratories

- Benefits of SNL-sited Sierra ATS clone
  - regression testing, nightly builds, continuous integration practical
  - disruptive use allowed (DSTs, system software experiments)
  - root level access allowed
  - dedicated login node and queues for critical code teams
  - easy access to SNL HPC filesystems (big-data)
  - **excellent and friendly local customer support!!! ;)**
  - local interactive debugging (totalview)
  - local visualization (paraview)
  - flexible scheduler policy, large-scale runs
  - shorter queue times
- Thank You!

This page left blank

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
O. Erik Strack	1150	oestrac@sandia.gov
S. Scott Collis	1400	sscoll@sandia.gov
Micheal Glass	1545	mwglass@sandia.gov
Tom Klitsner	9320	tklitsn@sandia.gov
Technical Library	01177	libref@sandia.gov

Email—External (encrypt for OUO)

Name	Company Email Address	Company Name

This page left blank

This page left blank

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.