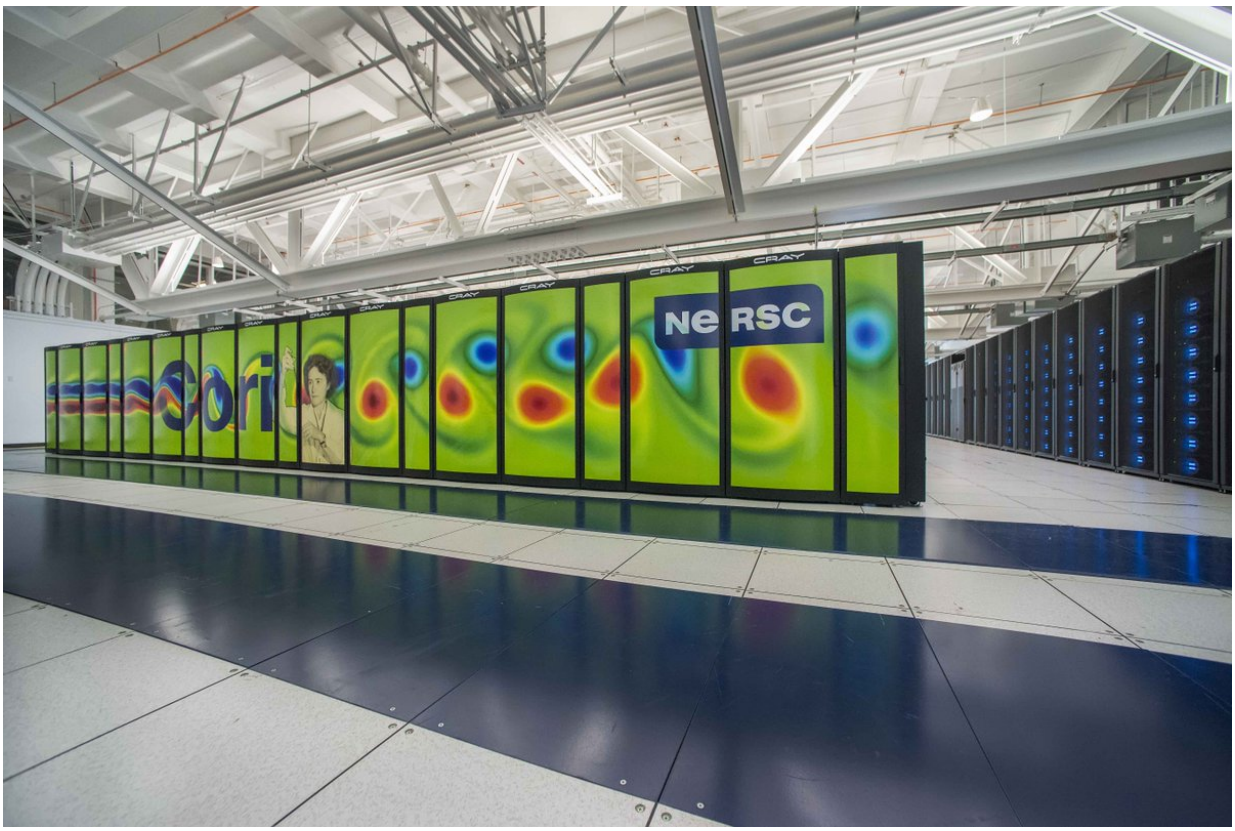
NERSC staff, users readying for delivery of Cori phase 2 Knights landing-based system in July

June 20 2016, by Jon Bashor



Picture of NERSC's Cori system. When fully installed, Cori will be the largest system for open science based on Knights Landing processors. Credit: Lawrence Berkeley National Laboratory

For the past year, staff at the Department of Energy's National Energy Research Scientific Computing Center (NERSC) have been preparing users of 20 leading science applications for the arrival of the second phase of its newest supercomputer, Cori, which consists of more than 9,300 nodes containing Intel's Xeon Phi Knights Landing processor - which was officially unveiled June 20 at the International Supercomputer Conference in Germany. The first compute cabinets are scheduled to arrive in July.

When fully installed, Cori will be the largest system for open science based on Knights Landing processors. The Knights Landing nodes will comprise phase two of the Cori system.

To ensure that a significant number of its 6,000 users could make the most effective use of this new manycore architecture, NERSC staff selected 20 leading applications for the NERSC Exascale Scientific Applications Program (NESAP), a collaborative effort that partners NERSC, Intel and Cray experts with code teams across the U.S. Lessons learned from working with the 20 NESAP codes are being used to develop an optimization strategy that the rest of the user base can quickly adopt.

"Application readiness efforts are critical for enabling ground-breaking science on our HPC systems as we move toward exascale. For the past year we've been working with these 20 teams to optimize their codes for Cori, so that when the machine arrives, they are ready to take advantage of the many capabilities the new hardware offers," said Jack Deslippe, acting head of NERSC's Application Performance Group. "As the primary computing center for DOE's Office of Science, we have an understanding of a broad user base utilizing over 600 apps at NERSC, as well as strong working relationships with Cray and Intel. This puts us in a unique position to provide a venue for computational scientists to engage industry experts around application optimization and to come up with

optimization strategies that scale to the wider HPC community."

Under NESAP, a member of NERSC's Application Readiness team assists the application teams with code profiling and optimization. Team members have also held a series of "dungeon sessions" with Intel and Cray engineers to optimize the codes. The resulting optimizations have been tested using nine Xeon Phi processor nodes installed at NERSC.

"Optimization is not always a straightforward process, so we've set up a system to help keep users from getting lost in the weeds," Deslippe said. "A number of the applications are ready now and we're making progress on the others. This process we've set up can be used by nearly all of the 600 projects running at NERSC."

NERSC frames the optimization process around the roofline performance model developed at Berkeley Lab. This sets expectations for what performance a developer can expect from their algorithm and which features of the Xeon Phi processor they should target:

- Thread/process scaling across the 68 cores and 272 hardware threads available on each KNL processor
- Vector parallelism which enables 32 flops per every cycle across two VPUs (vector processing units) on each core
- 16GB of on chip MCDRAM (volatile memory) that can be used to accelerate memory bandwidth sensitive applications

Over the last year, NERSC has been gaining experience exploiting the Xeon Phi hardware. NERSC is a central point for collating this experience and sharing it with the community. NERSC staff's expertise, other resources available to the community include:

- Training sessions that include multi-day hands-on sessions as well as shorter online presentations

- A series of case studies for the codes being optimized and ported
- A library of tools to support optimization and porting

Different applications tend to require different optimization approaches. Though NERSC staff only recently gained access to Knight's Landing processors, a number of optimized applications are already projected to perform well on the Cori system (the QCD code Chroma for example is expected to have perform twice as well on the Cori Phase 2 compared to the Haswell-based Phase 1 system). NERSC's repository of case-studies for application optimization on KNL is available at: http://www.nersc.gov/users/computational-systems/cori/application-porting-and-performance/application-case-studies/).

Among the applications that have been optimized for the Knights Landing architecture are:

WARP: One example where NESAP efforts have paid off is in the accelerator modeling application, WARP, which uses a Particle In Cell (PIC) approach utilizing the PICsar mini-app/library. WARP developers at Berkeley Lab worked closely with NERSC staff as well as engineers at Cray and Intel for over a year to optimize the application targeting the Cori Phase 2 system. Improvements included algorithmic changes to increase data-reuse in cache and MCDRAM on Knights Landing (KNL). Activity peaked in a "dungeon session" aimed at improving the effective use of the large KNL vector units within PICsar and has been accelerated due to efforts by NESAP post-doc Mathieu Lobet. Without optimization, PICsar performance on Cori Phase 2 nodes was expected to lag behind the Intel Haswell-based Phase 1 nodes by a factor of 3. With optimization, parity between these node types was achieved, with more optimization in the works.

MFDn: The MFDn (Many Fermion Dynamics) application, led by James Vary of Iowa State University, is used for computing the nuclear

structure for a number isotopes of interest and predicting nuclear reaction rates and cross sections. The targeted use case for MFDN requires using all available memory on the utilized nodes (expected to be a significant fraction of the Cori system). This requires either the use of the KNL MCDRAM as a cache or the explicit management of data on the MCDRAM via "FASTMEM" directives. With the help of NERSC staff member Brandon Cook and Cray and Intel staff, MFDN developers were able to beat the performance of KNL cache and achieve a 60 percent performance advantage on Cori Phase 2 compared to Haswell based Cori Phase 1 for critical sparse matrix math steps in their algorithm.

For complete details and a number of other Cori Phase 2 optimization examples, read more optimization case studies.

Provided by Lawrence Berkeley National Laboratory