Sandia National Laboratories

Exceptional service in the national interest

TRACKING TRILINOS PERFORMANCE

Chris Siefert, Sandia National Laboratories

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

U.S. DEPARTMENT OF ENERGY

NNSA

MOTIVATION

- Motivation: Taking ownership of Trilinos performance
  - Avoid (negative) surprises by the app teams saving them time.
  - Reducing time for initial setup for apps / developers on new-to-them systems / updated libs.
  - Work with Sandia Application Performance Team (APT) to stay at the most recent "good" version of libraries.
  - Understand behavior across time and across different systems.
  - Provide a way for app teams to know if it is safe to update Trilinos.
  - *Coordinate with Kokkos/KokkosKernels Teams to monitor performance.*

- FY26 Team: Chris Siefert, Luc Berger-Vergiat, Dane Camacho, Tim Fuller, Christian Glusa, Jonathan Hu, Brian Kelley.

WHAT DO WE MEAN BY “RUNS EVERYWHERE” ?

CPU
Eclipse (Broadwell)
Amber (Sapphire Rapids/DDR)
Tachi/Rocinante (Sapphire Rapids/HBM)

NVIDIA GPU
Perlmutter (A100)
Hops (H100)
Cronus (H200)

AMD GPU
Frontier (MI250)
Tioga (MI250)
El Dorado (MI300)

Intel GPU
Aurora (Intel Max)

WHAT DO WE MEAN BY “RUNS EVERYWHERE” ?

CPU
Eclipse (Broadwell)
Amber (Sapphire Rapids/DDR)
Tachi/Rocinante (Sapphire Rapids/HBM)

NVIDIA GPU
Perlmutter (A100)
Hops (H100)
Cronus (H200)

NEW!

AMD GPU
Frontier (MI250)
Tioga (MI250)
El Dorado (MI300)

Intel GPU
Aurora (Intel Max)

SO WHAT'S NEW ON PERFORMANCE?

- Soon-to-be open-source Warden replacement for Watchr
  - **Watch Dane Camacho's talk in the next slot!**
- Kokkos/KokkosKernels performance testing.
- HPSF/Benchmarking WG inspired improvements.
- Perf testing on SNL's Cronus (CTS2+H200) system.
- New performance testing in Belos/Zoltan2.
- Refinement to various existing tests (Mostly Tpetra & MueLu).

SO WHAT'S NEW ON PERFORMANCE?

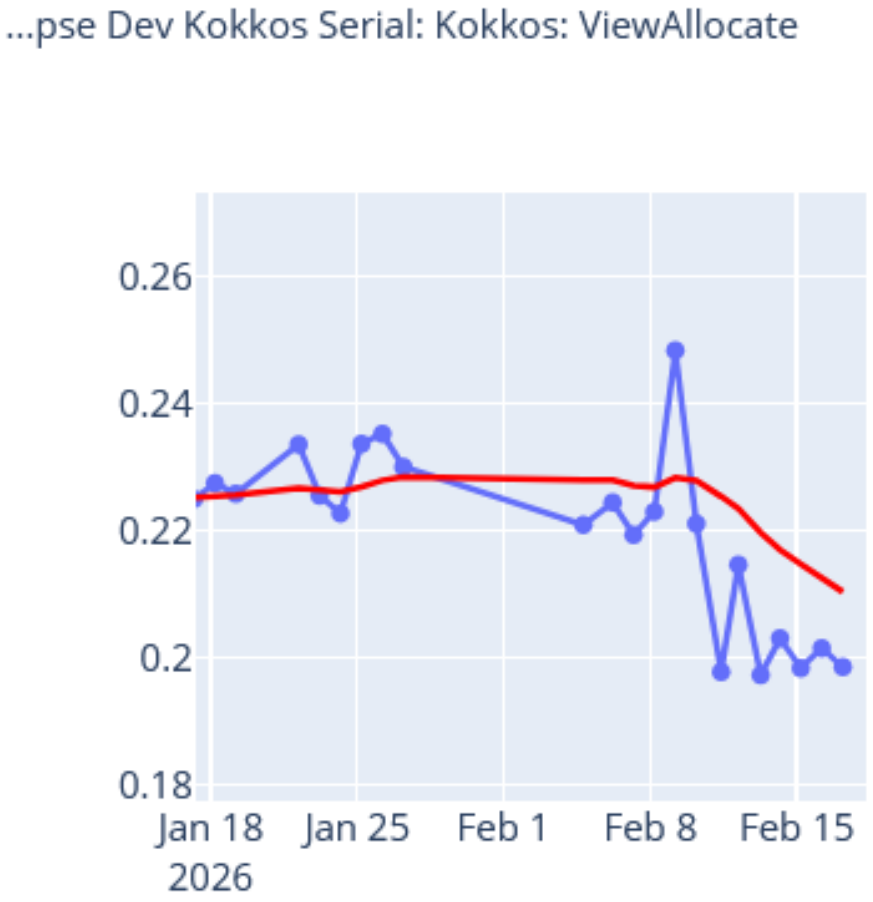
- Soon-to-be open-source Warden replacement for Watchr.
  - **Watch Dane Camacho's talk in the next slot!**
- Kokkos/KokkosKernels performance testing.
- HPSF/Benchmarking WG inspired improvements.
- Perf testing on SNL's Cronus (CTS2+H200) system.
- New performance testing in Belos/Zoltan2.
- Refinement to various existing tests (Mostly Tpetra & MueLu).

IMPROVING TESTING WITH KOKKOS/KOKKOSKERNELS

- Developed a Warden converter to handle gtest style timers (and allow similar tests to be nested for viz).

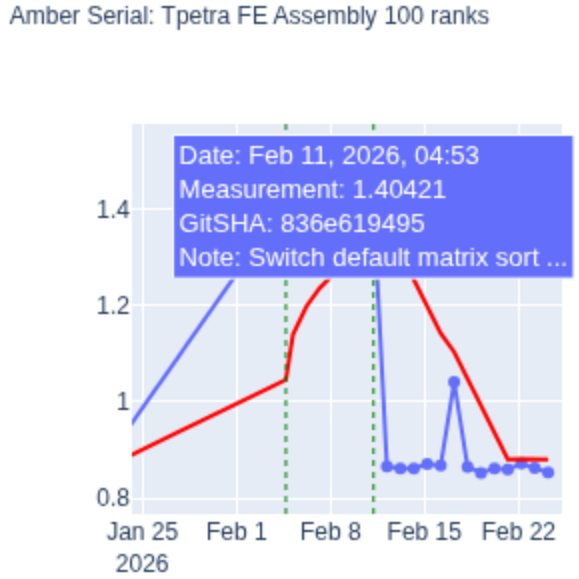
- Stood up Trilinos+ VOTD Kokkos/KokkosKernels on 3 systems
  - Eclipse (CPU)
  - Hops (NVIDIA GPU)
  - El Dorado (AMD GPU)

- Can now detect performance regressions *before* Kokkos promotions.



HPSF/BWG INSPIRED IMPROVEMENTS

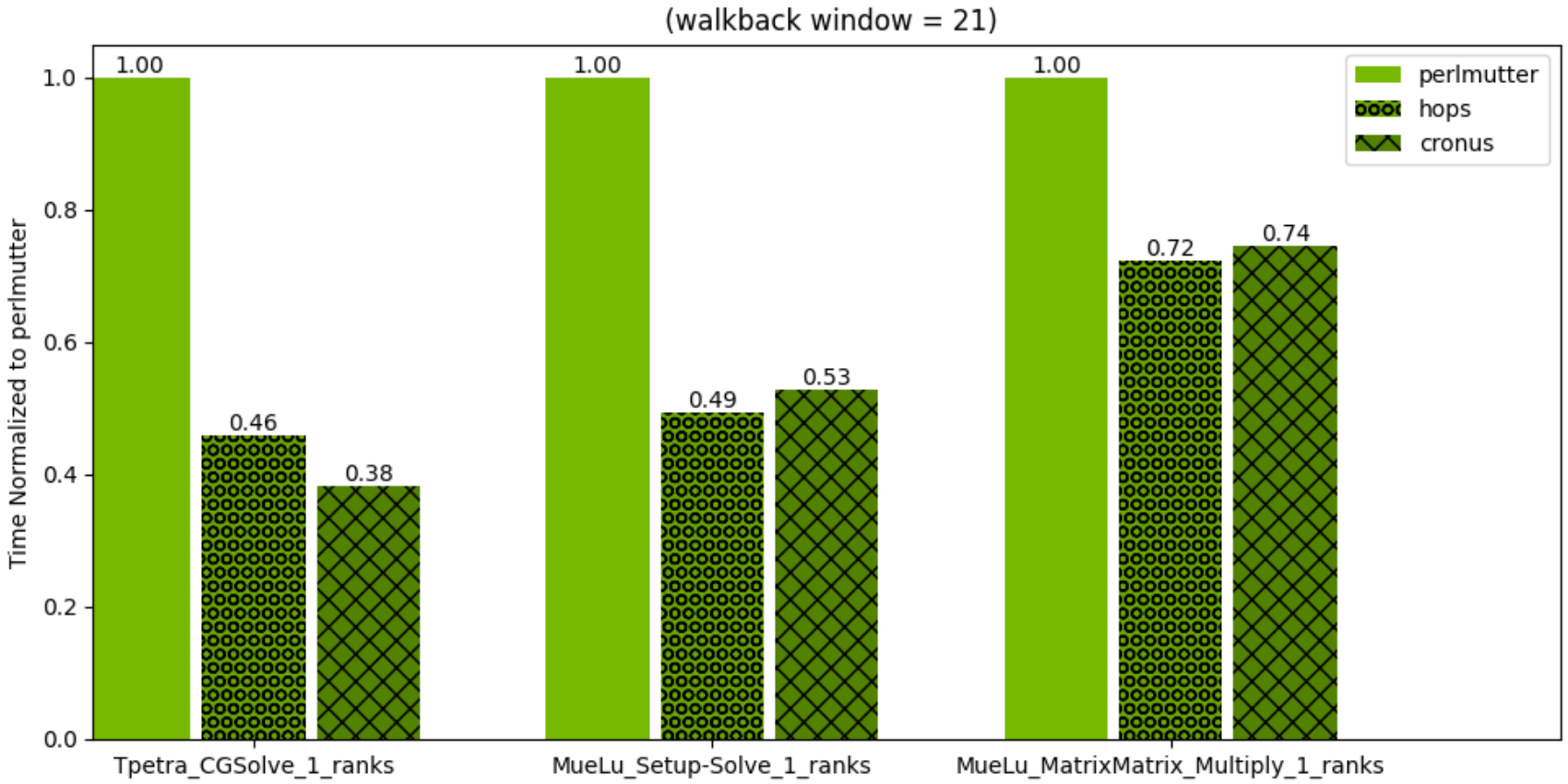
- Annotations
  - Provide a way to record changes in code/machine state.



- Record machine state
  - ENV vars (CUDA, ROCM, MPI, etc).
  - LDD output
  - Nvidia-smi / rocm-smi / sycl-ls
  - Lscpu
  - Modules
  - sensors

```
14  <metadata key="OMPI_MCA_btl" value="^openib,vader,self"/>
15  <metadata key="OMPI_MCA_btl_openib_allow_ib" value="true"/>
16  <metadata key="OMPI_MCA_btl_openib_ib_retry_count" value="7"/>
17  <metadata key="OMPI_MCA_btl_openib_ib_timeout" value="21"/>
18  <metadata key="OMPI_MCA_ess" value="pmi"/>
19  <metadata key="OMPI_MCA_hwloc_base_binding_policy" value="none"/>
20  <metadata key="OMPI_MCA_orte_ess_num_procs" value="1"/>
21  <metadata key="OMPI_MCA_orte_externally_bound" value="1"/>
22  <metadata key="OMPI_MCA_orte_precondition_transports" value="000000c90000001a-0000001
23  <metadata key="OMPI_MCA_osc" value="ucx"/>
24  <metadata key="OMPI_MCA_plm_slurm_args" value="--external-launcher"/>
25  <metadata key="OMPI_MCA_pml" value="ucx"/>
26  <metadata key="TPETRA_ASSUME_GPU_AWARE_MPI" value="1"/>
27  <metadata key="TPETRA_USE_NEW_COPY_AND_PERMUTE" value="ON"/>
28  <metadata key="ldd" value="   linux-vdso.so.1 (0x0000155555551000)
```

PERLMUTTER (A100) VS. HOPS (H100) VS. CRONUS (H200)



Cronus performance is similar to Hops on this small set of tests we got running.

PERFORMANCE-DRIVEN IMPROVEMENTS TO TPETRA

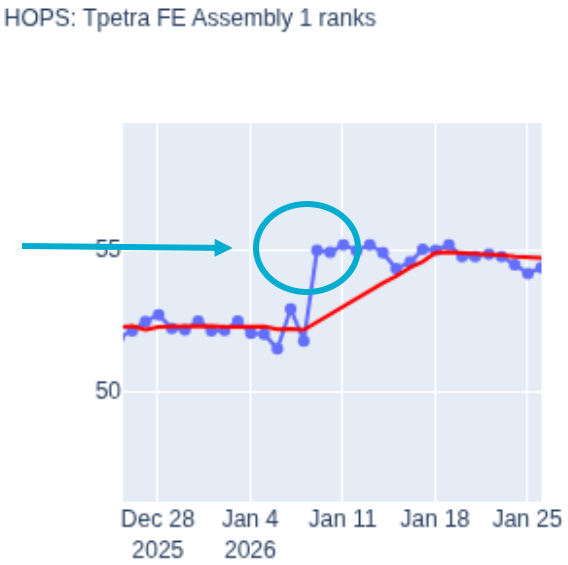
- Sorting was a bit of a mess.
  - Multiple different algorithms in different places (CrsMatrix, CrsGraph, Import_Util).
  - KokkosKernels was not consistently utilized.
  - Some tests did not reflect real work applications.
  - KokkosKernels only had a radix sort, which did not perform well on CPUs.

- Christian Glusa….
  - Refactored Tpetra to have consistent sorting.
  - Pushed a shell sort to KokkosKernels.
  - Modified the testing to be more realistic.

- Most importantly, these changes were carefully timed to allow Christian regular feedback from the performance system.

PERFORMANCE DRIVEN IMPROVEMENTS TO KOKKOS

Kokkos 5 integration



- Kokkos 5 slowed down subview creation, which the FE Assembly test does a lot of.
- To quote Trott: "BasicView ctor [created] .... mdspan to pass to submdspan, and then constructed a BasicView from an mdspan. For reference counted Views that amounts to three calls to increment and two to decrement."
- Issue was reported by us and remedied by Trott.

WHERE ARE WE GOING FOR THE REST OF THE YEAR?

- Continue migration to gitlab runners over Jenkins / Cron where we can.
  - Allow team members to (partially) performance test a candidate branch with a few clicks.
- Get a Kokkos/KokkosKernels only Warden dashboard for those teams.
- Improving automatic attribution
  - Better connection of SHAs and commits.
  - Automatic detection & annotation of changes in system state.
- Continued incremental improvement of GPU performance.