Exascale Computing Project 2021 Annual meeting summary

Fri@CEESD - April 23, 2021

Luke Olson, Matthias Diener, Matt Smith, and Mike Campbell

Software development, environment, and testing

- E4S: Extreme-Scale Scientific Software Stack
  - Pantheon - workflow management
  - DOE Code
  - ECP CI
- IDEAS Team: Better Scientific Software through better practices
- Summary and next steps

E4S: Extreme-Scale Scientific Software Stack

Community and infrastructure for developing and deploying software in HPC environments.

- Heavily spack-based
- Supercontainers (e.g. Docker, Singularity, Shifter)
- Turnkey environments
- Standardization of HPC software deployment
- Collection of build caches
- ECP stacks, tested, deployed on HPC platforms (ECP CI)

- Extreme-Scale Scientific Software Development Kits xSDK
  - Math Libraries
  - Data and Visualization
  - Scientific Workflows (Pantheon)
  - Software Ecosystem (DOE Code, ECP CI)

https::/e4s-project.github.io

E4S for CEESD

- Use E4S:
  - *MIRGE* in an E4S turnkey environment
  - Simplify deployment on emerging platforms
  - Accelerate builds and runtime (E4S Build Cache, RAM cache)
- Join E4S:
  - Requirements (https://e4s-project.github.io/policies.html)
  - Leverage E4S validation test suite (automated deployment!)
  - Resources (a throat to choke)
  - Increased visibility/exposure

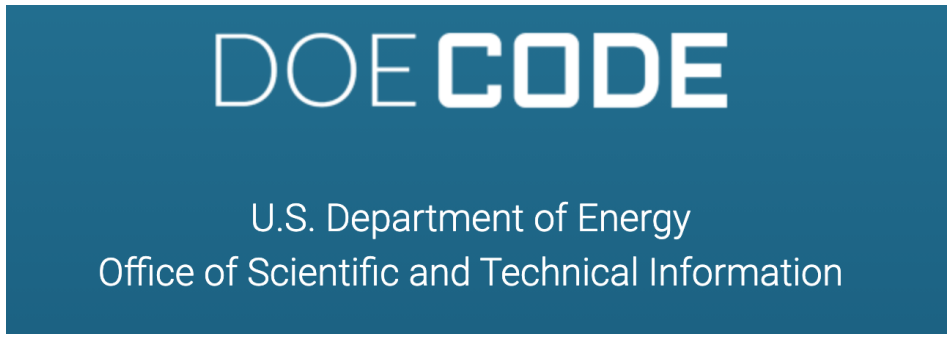
https://www.exascaleproject.org/research-project/e4s-and-sdk-efforts/

Pantheon: Reproducible workflows for HPC

- What is it?
  - Open standard for workflow specification (https://pantheonscience.org/standards)
  - YAML-based (install, run, postprocess, validate)
  - What's under-the-covers is your business (e.g. Parsl, conda)
  - Unification of job design for all supported platforms (includes all lab platforms)
- How can we use it at CEESD?
  - Publish *MIRGE* production workflows
  - Manage/re-imagine multi-platform *MIRGE* performance monitoring/benchmarking

https://pantheonscience.org
https://github.com/cinemascienceworkflows

DOE Code



(https://osti.gov/doecode)

- DOE establishes DOI for the code
- Maintain/propagate legacy of the DOE-funded code
- Reduces the big messy pile of DOE-funded code
- Helps resolve ongoing licensing debacle

ECP Continuous Integration

- CI servers at most major supercomputing sites
  - Operates from site-hosted gitlab instances
  - All currently operating lab-based machines (emerging supported quickly)
  - Regularly exercises ECP codes w/ E4S stacks
- Two ways CEESD could leverage ECP CI:
  - Independent DOE-funded code mirror
  - E4S stack inclusion

General ECP-CI Info, documentation: (https://ecp-ci.gitlab.io)
LLNL/LC (including Lassen): (https://lc.llnl.gov/gitlab)

ECP - IDEAS Team

1 Customize and curate methodologies

- Target scientific software productivity and sustainability
- Use workflow for best practices content development

3 Establish software communities

- Determine community policies to improve software quality and compatibility
- Create Software Development Kits (SDKs) to facilitate the combined use of complementary libraries and tools

2 Incrementally and iteratively improve software practices

- Determine high-priority topics for improvement and track progress
- *Productivity and Sustainability Improvement Planning (PSIP)*

4 Engage in community outreach

- Broad community partnerships
- Collaboration with computing facilities
- Webinars, tutorials, events
- *WhatIs* and *HowTo* docs
- Better Scientific Software site (https://bssw.io)

For more about our work see this report:
https://doi.org/10.2172/1606662

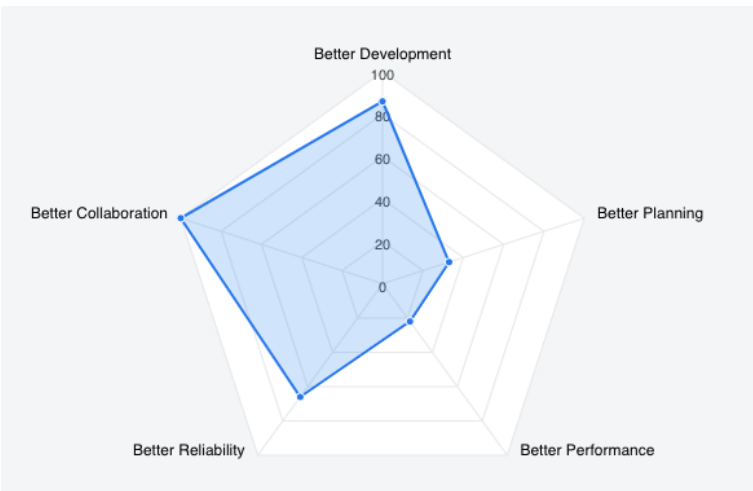
IDEAS productivity

ECP EXASCALE COMPUTING PROJECT

(https://ideas-productivity.org/ideas-ecp)

Software process and practice

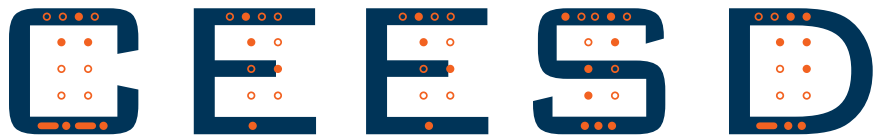
- Better Scientific Software (https://bssw.io)
  - Best practices - development and project management
  - Training materials and resources
- Some observations questions, and recommendations:
  - Your software will live longer than you expect, so plan for it now (software lifecycle)
  - Consider stand-alone, separate testing suites
  - Restart testing is highly valuable.
  - Resource adaptation is highly valuable.
  - There is a deep interest in performance tracking over time (Kitware supported)
  - Testing tax is mostly imagined. The cost of defects is extremely high.
  - Onboarding and offboarding procedures are highly valuable.



Summary and next steps

DOE is providing resources for helping teams step up development, deployment, and project management strategies with emphasis on sustainability. They think it is important. We should take advantage of this.

- Now: Revive spack-based install
- Now: Identify an E4S stack that will work for us, and install to it
- Now: Mirror *MIRGE* to gitlab@lc and start kicking the tires on ECP CI
- Now: Engage with IDEAS Team (invite Greg Watson@ORNL for Fri@CEESD), more to follow
- Iterative: Better up our software planning, performance testing, on/offboarding
- Longer term: Join E4S/become an E4S/ECP package



Questions?

This material is based in part upon work supported by the Department of Energy, National Nuclear Security Administration, under Award Number DE-NA0003963.