An Overview of the ECP Software Ecosystem

Todd Munson on behalf of the ECP Software Ecosystem and Delivery and SDK teams

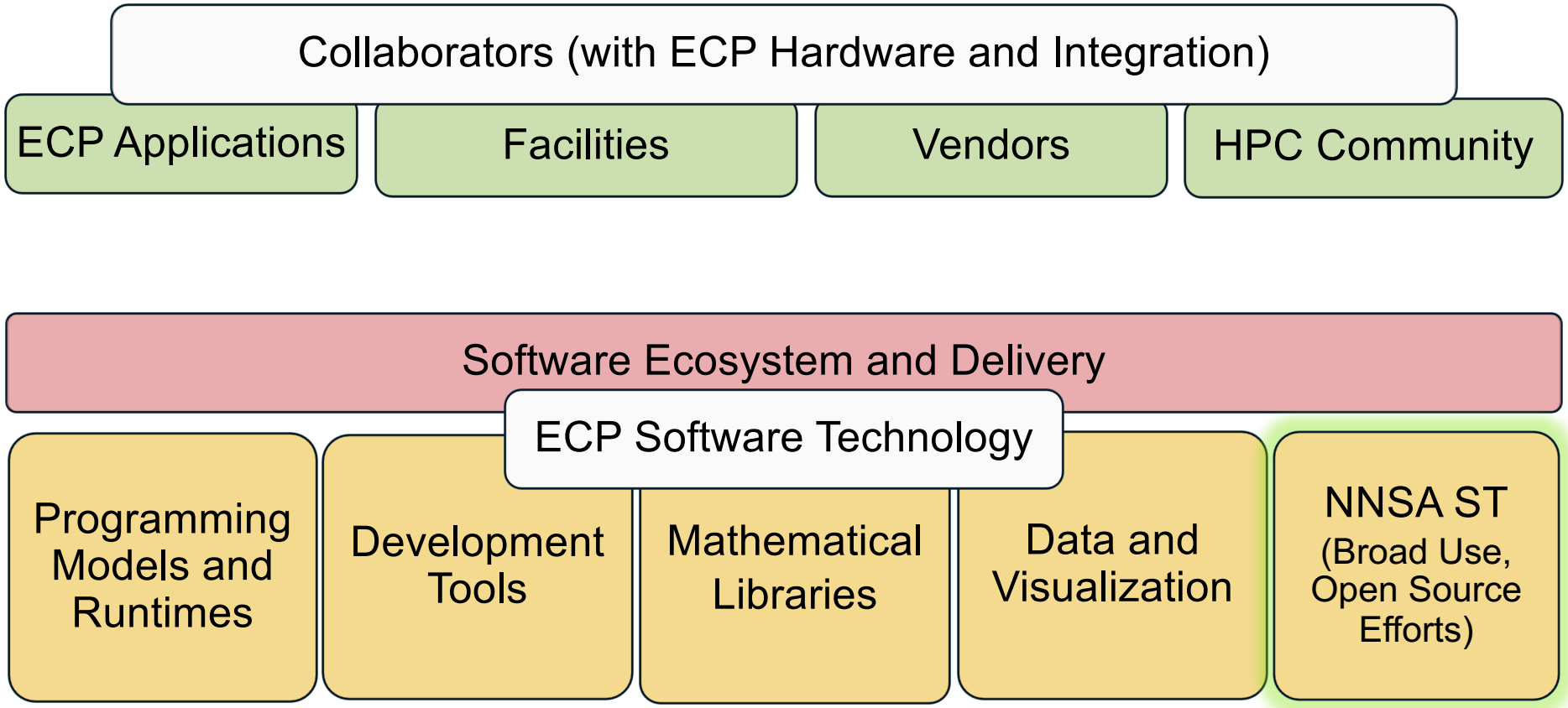
First Extreme-scale Scientific Software Stack Forum

September 23, 2019

NNSA National Nuclear Security Administration

U.S. DEPARTMENT OF ENERGY | Office of Science

Software Ecosystem and Delivery: Context for the portfolio



Software Ecosystem and Delivery: Context for the portfolio

Vision	A sustainable, high-quality software ecosystem that is continuously improved by a robust research and development effort, is deployed on advanced computing platforms, and is broadly adopted by application teams and software developers to accelerate their science	
Mission	To deliver reliable, performant, and exascale-ready software to application development teams and the packaging technologies needed to deploy them to facilities, developers, and end users	
Objective	Provide regular releases of ECP software technologies that are tested using continuous integration capabilities and that can be built using Spack as either modules or containers	
Challenges	Managing the complexity of builds across the varied software technologies and the unique exascale computing environments on which they will be delivered	
Starting Point	Existing development efforts on Spack and container technologies, and experience developing and deploying high-quality software from the xSDK and IDEAS projects	
Portfolio Goals	**Software Ecosystem SDK**	Coordinate the software development kit activities across the software technologies and regularly release the Extreme-Scale Scientific Software Stack
	Packaging Technologies	Develop the technologies needed to deploy large software collections at the facilities and optimize and improve interoperability of container technologies

Software Ecosystem and Delivery: Content of the portfolio

E4S/Software

- Regular additions of software technologies, integration testing, and releases of E4S
- Certified Spack recipes and binaries for software technologies
- Currently includes 40 software technologies
- Broad adoption of reliable, performant, exascale-ready software

Packaging Tech/Facilities

- Turnkey deployment of large collections of software technologies at facilities
- Coordinate with and leverage facility personnel and resources, such a continuous integration capabilities
- Spack recipes for over 2,800 software packages
- Broad deployment of exascale-ready software technologies

Packaging Tech/Applications

- Optimize and improve interoperability of container technologies
- Enable entire application to be packaged into reproducible container images
- Accelerate application development and deployment workflows

Software Ecosystem and Delivery: Portfolio

Project Short Name	PI Name, Inst	Short Description/Objective
Software Ecosystem SDK	Willenbring, SNL	Coordinate the software development kit activities across the software technologies and regularly release the Extreme-Scale Scientific Software Stack
Packaging Technologies	Gamblin, LLNL	Develop the Spack technologies needed to deploy large software collections at the facilities and optimize and improve interoperability of container technologies for exascale computing environments

Software Ecosystem and Delivery: Priorities and challenges

- Priorities
  - Develop features in Spack necessary to support all software technology products in the Extreme-Scale Scientific Software Stack (E4S)
  - Develop Spack Stacks for reproducible turnkey deployment of large collections of software (i.e. facility scale, software development kits, E4S)
  - Optimize and improve interoperability of container technologies for use by application development teams
  - Regular releases and integration testing of E4S and other software development kits (SDKs) with frequent additions of new software technologies
  - Work with application development focus area for adoption of software technologies
  - Work with the hardware and integration focus area for the deployment of software technologies and leveraging of facility personnel and resources, such as their continuous integration capabilities
- Challenges
  - Complexities of multi-version, multi-product coordination become too challenging
  - ECP release processes like E4S and SDKs are not sustainable by ongoing projects and the open-source community and does not take root post-ECP

Software Ecosystem: E4S context

Vision	Curate sustainable, scalable, and high-quality software stack with customizable builds that is: continuously improved by a robust research and development effort; deployed on first exascale computers; and broadly adopted by application teams and software developers to accelerate their science
Mission	To deliver a scalable, performant, and interoperable software stack that is deployable from laptops to exascale computing platforms and is customizable for facilities, developers, and end users
Objective	Coordinate software development kit activities across the software technologies, support reproducible container and bare metal builds of custom stacks, and regularly release the entire Extreme-Scale Scientific Software Stack (E4S)
Challenges	Managing the complexity of builds across the varied software technologies, interoperatiblity and continuous integration of the components, and the unique exascale computing environments on which they will be delivered
Starting Point	Experience developing and deploying high-quality software from the xSDK and IDEAS projects
Team Leads	Chuck Atkins (Data and Visualization), Tim Haines and Barton Miller (Development Tools), Sameer Shende (Programming Models and Runtimes), Jim Willenbring (Software Ecosystem)

Software Ecosystem: E4S context

- Mission
  - To deliver a scalable, performant, and interoperable software stack that is deployable from laptops to exascale computing platforms and is customizable for facilities, developers, and end users
- Challenges
  - Greatly varying software quality and related practices
  - Establishing and maintaining common versions of software dependencies
  - Establishing workflows that allow autonomy for individual teams, yet maintain an interoperable ecosystem
  - Sustaining interoperability without established processes and mature continuous integration tools
  - Development for exascale platforms

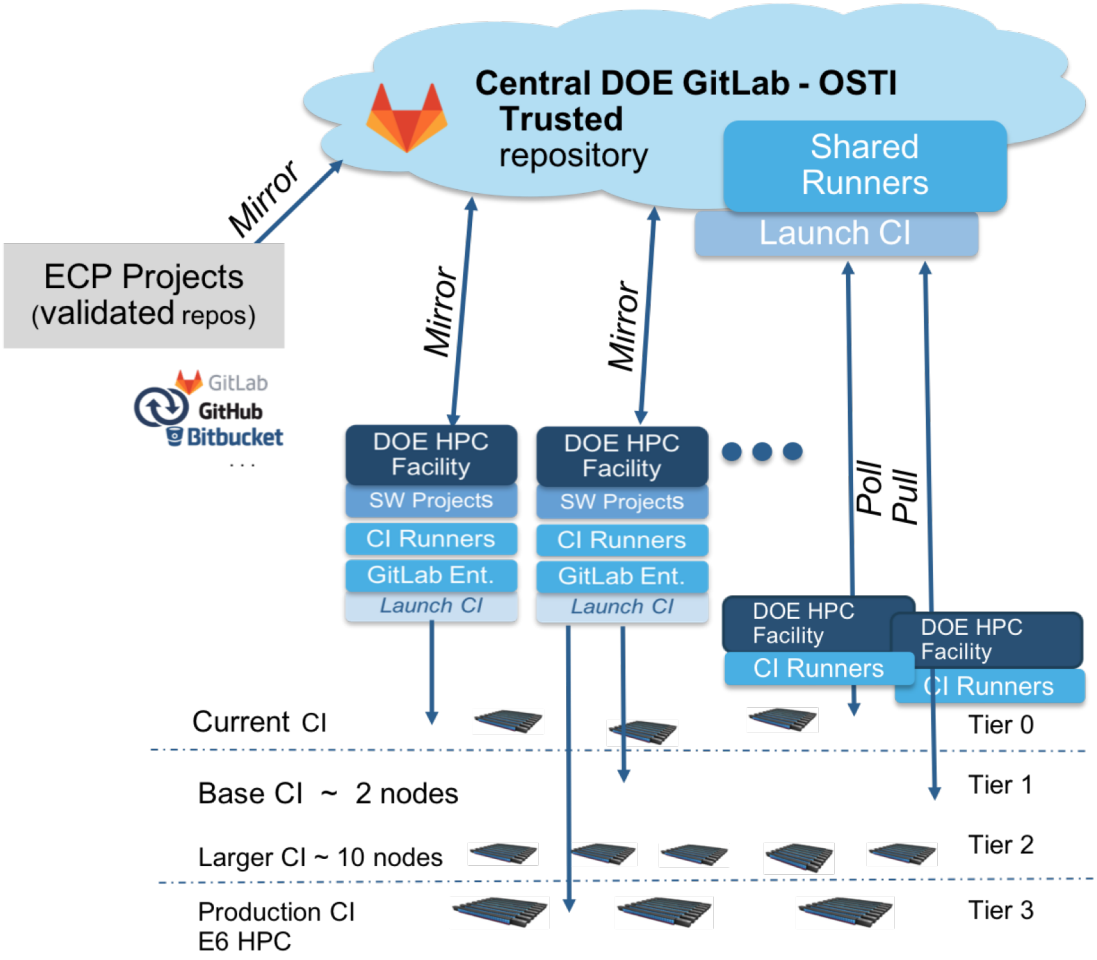
Software Ecosystem: E4S highlight

- E4S is released twice a year; two versions have been released to date
- E4S version 0.2 supports:
  - 40 full release ECP ST products
  - Limited access to 10 additional ECP ST products
  - Docker
  - Singularity
  - Shifter
  - Charliecloud
  - Inception
- Planning E4S version 0.3 for release at SC19

Software Ecosystem: E4S continuous integration

In conjunction with the Hardware and Integration (HI) area, the SDKs are principal participants in the Continuous Integration testing workflow for ECP

1. Mirror from GitHub, GitLab, etc. to project OSTI GitLab
2. Authenticate OSTI runner at lab using OnyxPoint and Auth provider (OAuth, Shibboleth, etc.)
3. Build/run at any facility on N nodes
4. Build artifacts are preserved for faster future builds
5. Push results back to OSTI



1st Workshop on NSF and DOE High Performance Computing Tools

- https://oaciss.uoregon.edu/NSFDOE19
- July 10-11, 2019 (Wed-Thu), 312 Lillis, University of Oregon, Eugene, OR 97403
- This workshop will equip application developers and supercomputing center staff to better understand the DOE ECP E4S technology and develop a two-way collaboration between NSF and DOE HPC centers to expand and standardize on the HPC software development and deployment ecosystem using Spack. NSF and DOE application developers will benefit from use of this common software delivery platform, allowing these tools and applications to be easily installed at the DOE HPC sites through Spack
  - Spack tutorial with hands-on exercises
  - E4S containers, software deployment at DOE facilities, HPC container runtimes
  - Hands-on sessions with help for creating Spack recipes

Packaging Technologies: Supercontainers context

Vision	Curate sustainable, scalable, and high-quality containerized software stacks, continuously improved by a robust research and development effort, deployed on first exascale computers, that are broadly adopted by application teams and software developers to accelerate their science
Mission	To deliver scalable, performant, and interoperable container images and runtimes that are deployable from laptops to exascale computing platforms and are available to facilities, developers, and end users
Objective	Employ a multi-level approach to accelerate the adoption of container technologies for exascale computing environments, ensuring that container runtimes will be scalable, interoperable, and integrated at DOE Facilities
Challenges	Managing the complexity of container builds across the varied software technologies and the unique exascale computing environments on which they will be delivered
Starting Point	Several container runtime initiatives along with interspersed container deployment efforts at DOE Facilities
Team Leads	Todd Gamblin (Spack) and Andrew Younge (Supercontainers)

Packaging Technologies: Supercontainers starting point

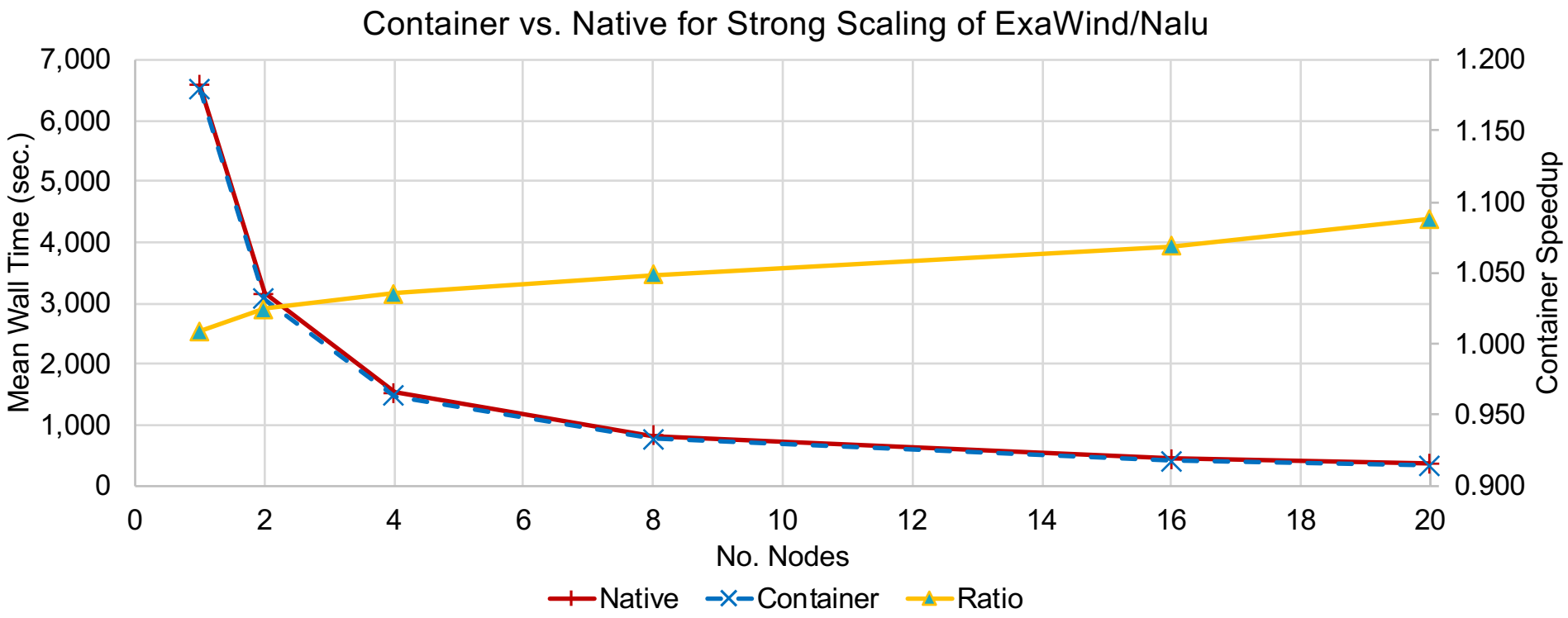
- Containers can provide greater software flexibility, reliability, ease of deployment, and portability
- Containers have gained significant interest in the ECP
  - Interest largely originates from industry efforts around Docker

- Several container runtimes exist for HPC today
  - Shifter, Singularity, Charliecloud
  - Diversity is good, but needs reconciliation
- Challenges to using containers at exascale
  - Scalability and performance
  - Resource management
  - Interoperability
  - Security
  - Integration with HPC (batch jobs, Lustre, etc)

Packaging Technologies: Supercontainers context

- Mission
  - To deliver scalable, performant, and interoperable container runtimes that are deployable on laptops to exascale computing platforms and are available to facilities, developers, and end users
- Objectives
  - Development for exascale platforms
  - Collaboration with software technologies and application development teams
  - Training, education, and support
  - Integration with facilities

Packaging Technologies: Supercontainers highlight



Software Ecosystem and Delivery: Future

- By the end of the Exascale Computing Project, we envision that
  - Software technology teams will regularly contribute to and support high-quality software development practices and will utilize continuous integration capabilities provided by facilities
  - Application development teams will adopt ECP software technologies that accelerate their development efforts because of the assurance provide by E4S and SDKs
  - Facilities, developers, and users will embrace our mature packaging technologies as their deployment mechanism
  - Facilities will continuously build and provide E4S/SDK-certified software, and it will be available for turnkey deployment at facilities, in development environments, and by users on exascale platforms and beyond
- Beyond the Exascale Computing Project, we envision a sustainable, high-quality software ecosystem that is continuously improved by a robust research and development effort, is deployed on advanced computing platforms, and is broadly adopted by application teams and software developers to accelerate their science

Questions?