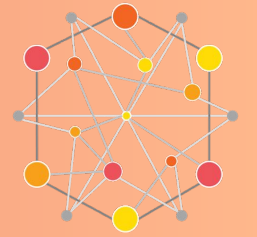
EXASCALE COMPUTING PROJECT

ExaWorks
https://exaworks.org

Tutorial: Developing Robust and Scalable Next Generation Workflows Applications and Systems

PEARC 2022



Lawrence Livermore National Laboratory · Argonne National Laboratory · Brookhaven National Laboratory · Oak Ridge National Laboratory · U.S. Department of Energy · Office of Science

Agenda

8:30 a.m. – 8:50 a.m.	Introduction and brief overviews of the tools, with Q&A	The team
8:50 a.m. – 9:00 a.m.	Logging into AWS / short break	
9:00 a.m. – 9:35 a.m.	RADICAL EnTk Exercises	Matteo Turilli
9:35 a.m. – 10:00 a.m.	Swift/T Exercises part 1	Justin Wozniak
10:00 a.m. – 10:30 a.m.	Coffee break	Ballroom B/Statler
10:30 a.m. – 10:45 a.m.	Swift/T Exercises part 2	Justin Wozniak
10:45 a.m. – 11:20 a.m.	Parsl Exercises	Kyle Chard
11:20 a.m. – 11:30 a.m.	Community outreach and wrap up	Rafael Ferreira da Silva

The Team

Dan Laney

Lawrence Livermore National Laboratory

Kyle Chard

Argonne National Laboratory

Shantenu Jha

Brookhaven National Laboratory

Rafael Ferreira da Silva

Oak Ridge National Laboratory

Todd Munson

Argonne National Laboratory

Aymen Alsaadi

Brookhaven National Laboratory

Ben Clifford

Argonne National Laboratory

James Corbett

Lawrence Livermore National Laboratory

Mihael Hategan

Argonne National Laboratory

Ketan Maheshwari

Oak Ridge National Laboratory

Andre Merzky

Brookhaven National Laboratory

Zeke Morton

Lawrence Livermore National Laboratory

Mikhail Titov

Brookhaven National Laboratory

Matteo Turilli

Brookhaven National Laboratory

Andreas Wilke

Argonne National Laboratory

Justin M. Wozniak

Argonne National Laboratory

Previous Contributors

Dong H. Ahn

Lawrence Livermore National Laboratory

Stephen Herbein

Lawrence Livermore National Laboratory

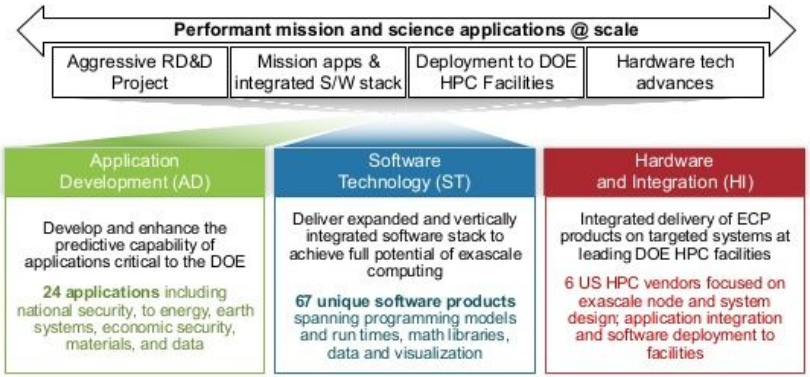
Yadu Babuji

Argonne National Laboratory

Exascale Computing Project (ECP)

Seven-year, $1.8B project that aims to accelerate R&D, acquisition, and deployment of **exascale** computing capability to DOE

Six core national laboratories are focused on software, applications, hardware, system engineering and testbed platforms



Scientific computing workflows underlie a significant number of projects in the Exascale Computing Project (ECP) portfolio

Many teams are creating infrastructures to:

- Couple multiple applications
- Manage jobs, sometimes dynamically
- Orchestrate compute/analysis and manage data

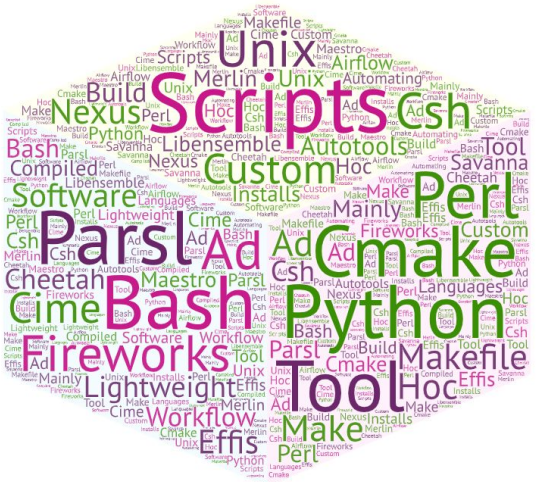
There is **duplication of effort** in these infrastructures

These customized workflows incur **significant costs** to port, maintain and scale

These tools do not always interface with facilities smoothly

The costs could be minimized by creating a reliable, scalable, portable **software development kit (SDK) for workflows**

ExaWorks Survey in 2020:
responses from 15/31 ECP application teams highlight the ad hoc workflows landscape



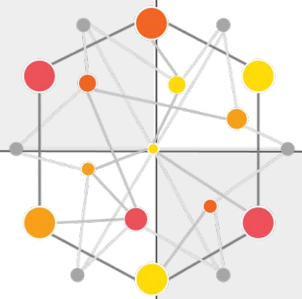
Our approach will ensure exascale readiness of a wide range of ECP workflows and improve their long-term sustainability

ECP Engagement

Partner with ECP AD and other teams for co-design and adoption of **ExaWorks SDK** to address their workflow problems

Facilities Engagement

Engage with DOE **compute facilities** to support deployment and use of workflows at scale; contribute requirements for next-generation systems

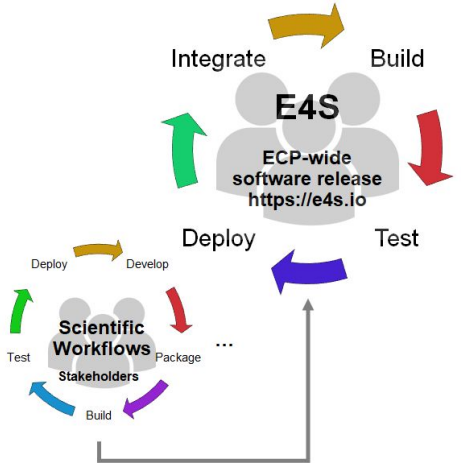


SDK Dev and CI/CD

Curate community SDK to enable robust, scalable, portable, performant workflows; progressively increase the availability of composable workflow components

Community Engagement

Lead the workflows community towards interoperability and reuse; build the case for future standardization for long-term sustainability



The ExaWorks SDK is packaged, deployed, and tested using E4S and ECP CI infrastructure

ExaWorks is *not* funded to build another workflow system

We are funded to provide a production-grade Software Development Kit (SDK) for exascale workflows

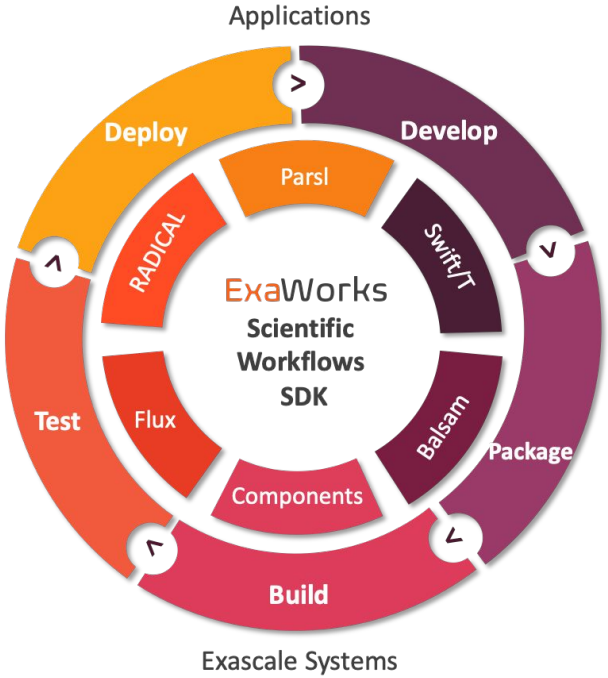
SDK democratizes access to hardened, scalable, and interoperable workflow management technologies and components

Implemented via community-based approach at progressively integrated levels

- Level 0: Technologies are packaged together
- Level 1: Component interfaces or pairwise integrations
- Level 2: Community developed and supported APIs

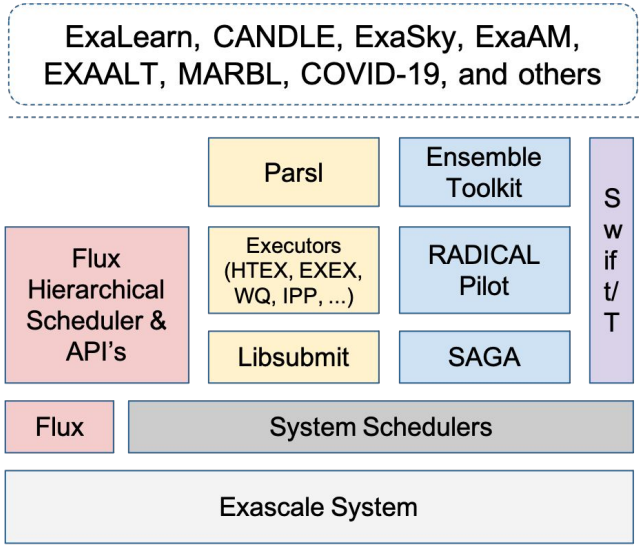
Approach

- Community policies for software quality (based on E4S)
- Open community-based design and implementation process
- Ensure scalability of components on Exascale Systems
- Standard packaging and testing



ExaWorks SDK brings together five seed technologies currently impacting ECP applications

- Scientific workflows SDK includes four seed technologies
  - **Flux** – hierarchical resource and job management software
  - **Parsl** – flexible and scalable parallel programming library for Python
  - **RADICAL** – component-based workflow middleware
  - **Swift/T** – high performance dataflow computing



PSI/J: Portable Submission Interface for Jobs

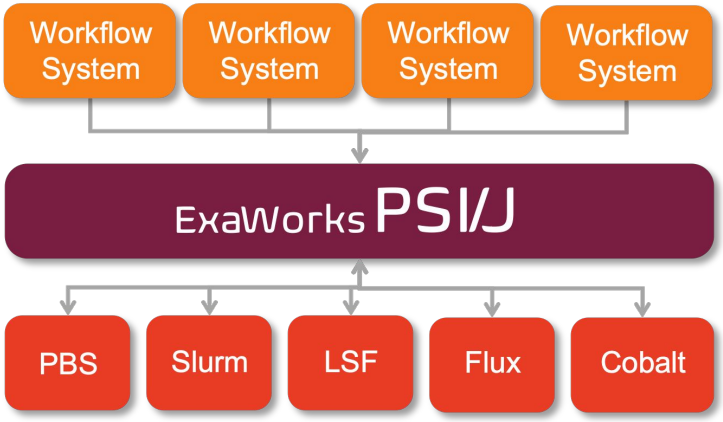
A set of interfaces that allow the specification and management of “jobs”

Support for Slurm, LSF, Cobalt, Flux, PBS

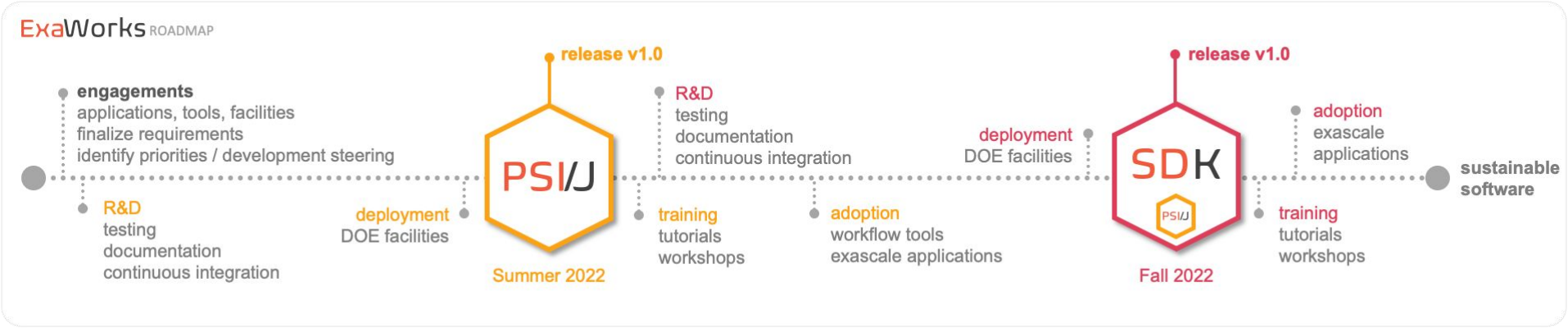
Open document to define a language-independent specification

Community specification

http://exaworks.org/job-api-spec/specification.html



ExaWorks RoadMap



Exascale Workflows|Community

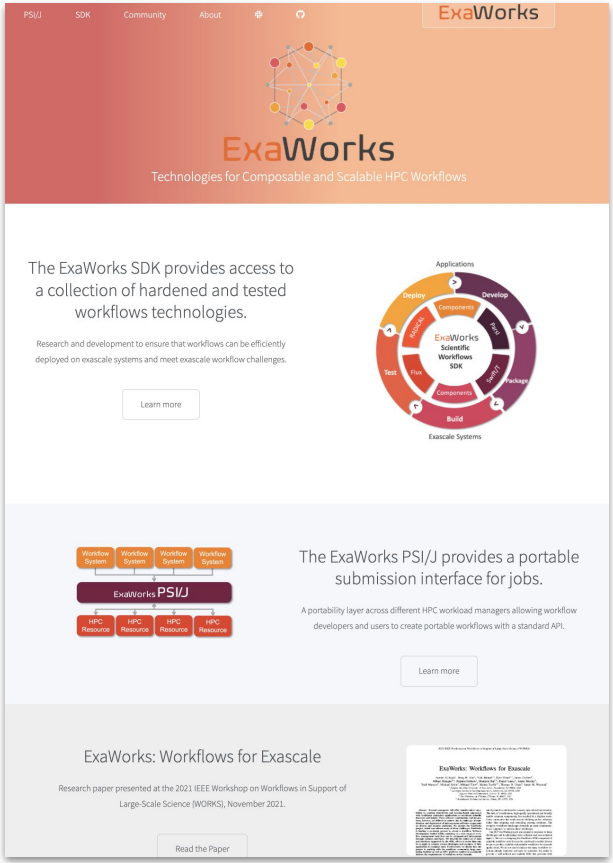
Learn more…

https://exaworks.org

- Join our Slack Channel
- Read the documentation

Engagements

- Get in touch to discuss how ExaWorks components can benefit your project



ExaWorks

https://exaworks.org

https://tinyurl.com/exaworks

Thank you!

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC.

Lawrence Livermore National Laboratory · Argonne National Laboratory · Brookhaven National Laboratory · Oak Ridge National Laboratory · U.S. Department of Energy · Office of Science