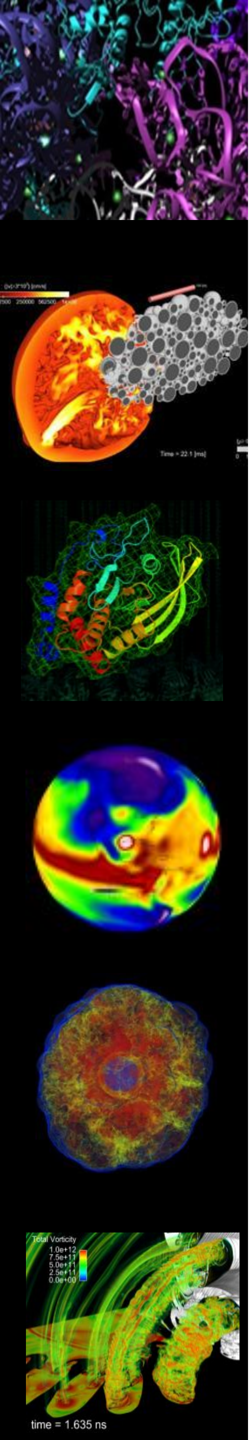



# Status on Design, Deployment, Acceptance, and Operation

*Anthony M. Agelastos*

*Jennifer K. Green*

*Kevin D. Stroup*

Presented to the *Cray Users Group Meeting* (CUG'22) - May 2022

SAND No: SAND2022-4432 C
LA-UR: 22-23479
UNCLASSIFIED

# Crossroads Supercomputer

- 3rd Advanced Technology System (ATS-3) in the Advanced Simulation and Computing (ASC) Program
- Supports:
  - National Nuclear Security Administration's (NNSA) Stockpile Stewardship Program (SSP)
  - Current and planned Stockpile Life Extension Programs activities
- The primary users of ASC platforms are designers, analysts and computational scientists
  - Los Alamos National Laboratory (LANL)
  - Lawrence Livermore National Laboratory (LLNL)
  - Sandia National Laboratories (SNL)

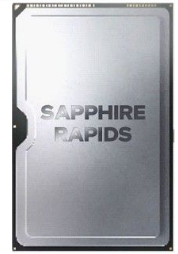
# Overview of Presentation

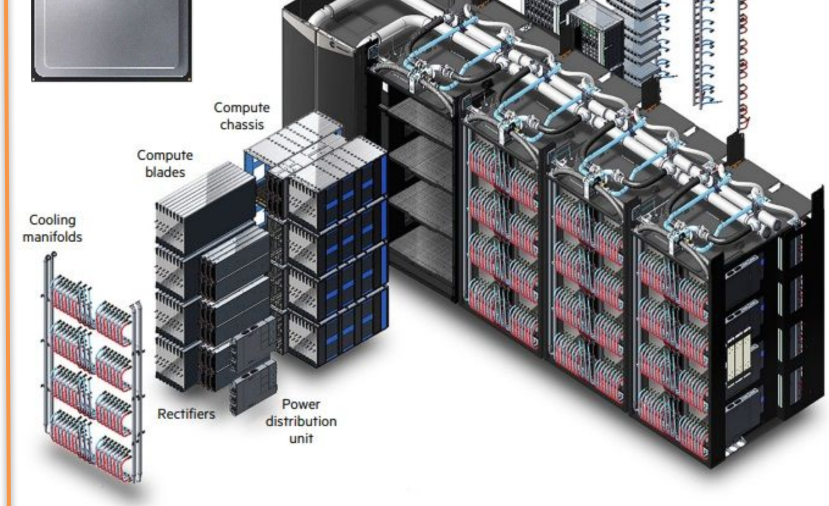
- Design of Crossroads
- Programming Environments
- Operations
- Deployment
- Acceptance
- Performance Acceptance
- Closing & Questions

# Design

# Design

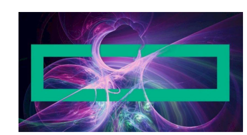
- **HPE-Cray Shasta EX Supercomputer**
  - Follow-on to Trinity, ACES current Advanced Technology System
- **Intel Sapphire Rapids** processors
- **Cray Slingshot** Next-Gen Fabric
- Final configuration is **High Bandwidth Memory**
- **DDR-5** for early deliveries
- HPE Cray "Shasta" Cabinets
  - **Mountain**
    - High density Cray blades
    - 64 compute blades per cabinet
    - 2/4/8/16 NICs per blade
  - **River**
    - Flexible support for arbitrary nodes




Mountain
"Rack"

River
"Rack"



Slingshot
Technology

CROSSROADS

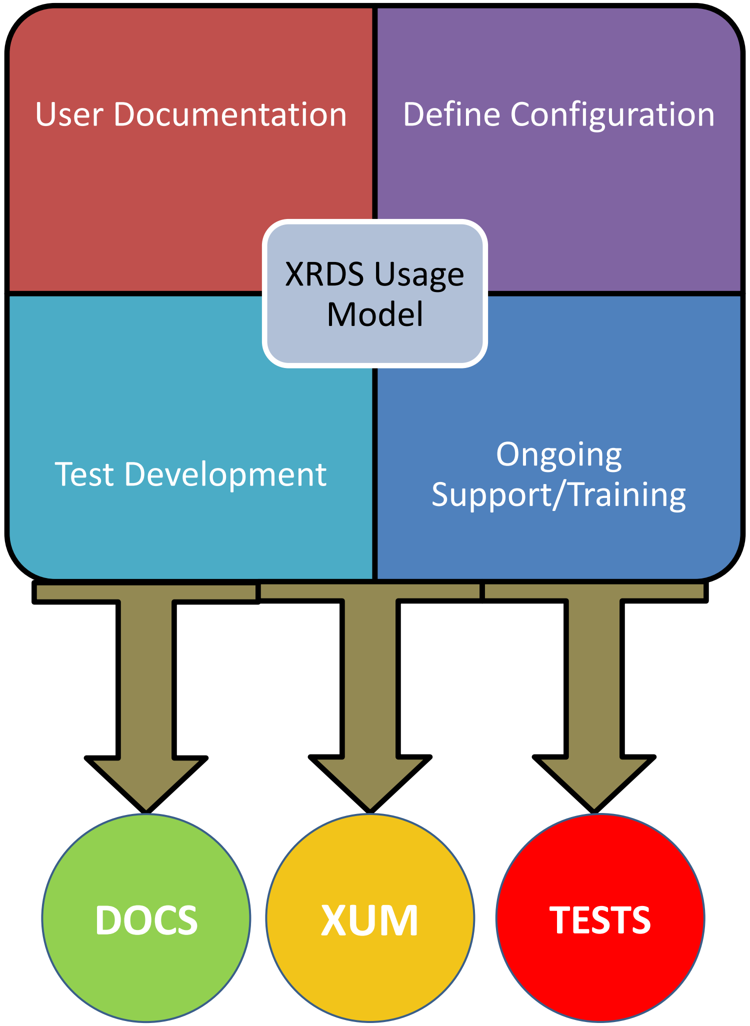
# Programming Environment

# Programming Environments on Crossroads

- Software Environment
    - CPE - Cray Programming Environment Support
        - Cray Supplied Software Environment
        - GCC/Cray/Intel Compilers
        - Cray MPICH2 tuned to SlingShot interconnect
        - Libraries, tools, and utilities supporting HPC workloads
    - Trilab Computing Environment (TCE)
        - Spack supplied software stack
        - NNSA Tri-lab collaboration
- Development Environment
    - Container Support
    - DevOps Support (remote via RCE)
    - Code Development Tools
- Filesystems and Scheduling Interfaces
- Data Science/Analysis Support
    - Visualization Support

# Programming Environment Working Group

- Visualization Support
- Spack Support and TCE
- Scheduler Environment
- PE Functional & Performance Testing
- Filesystems User Interfaces
- DevOps Support
- Development Environment Testbeds
- Cray PE Support
- Container Support
- Compilers and MPI
- Code Development Tools Support




CROSSROADS

# Programming Environment Software

Hewlett Packard Enterprise

HPE Cray Programming Environment

Spack

MVAPICH

ECP EXASCALE COMPUTING PROJECT

oneAPI

GCC

CRAY

LATEX

ParaView

MPI

CI/CD

Charliecloud

arm FORGE PRO

LLVM COMPILER INFRASTRUCTURE

FFmpeg

julia

mercurial

visit

intel

splunk>

OpenMP

MPICH

vnc2hpc

Grafana

RICE

ANACONDA

EnSight

Stack Trace Analysis Tool

CROSSROADS

CMake

Lmod

# Operations

CROSSROADS

# Administrative Management System

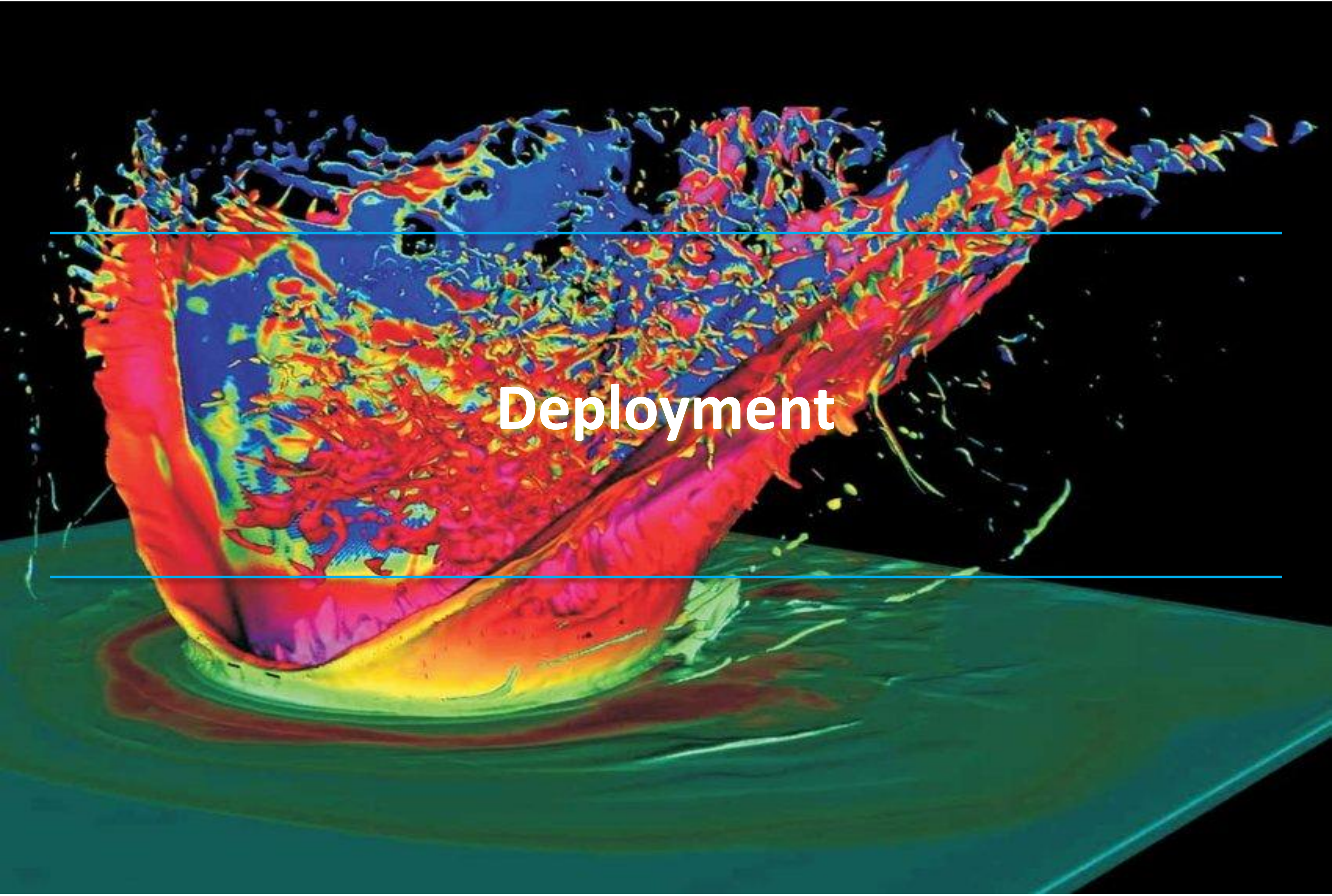
- Cray System Management (CSM)
  - Cray Operating System (COS)
  - User Access Nodes (UAN)
  - Image Management
- Networks
  - Hardware Management Network (Service)
  - Node Management Network (NMN)
  - High Speed Network (HSN)
  - Customer Access Network (CAN)

CROSSROADS

# User Environment Administration

- Cray Programming Environment (CPE)
  - Content Projection Service (CPS)
  - PE Image Orchestration
  - Environment Modulefiles (LMOD/TMOD)
- SchedMD Slurm Workload Manager
  - Allocates access to compute resources to users for some duration of time so they can perform work
  - Framework for starting, executing, and monitoring work
  - Arbitrates contention for resources by managing a queue of pending work

# Deployment



# Deployment Timeline

**SDT**
Systems Deployment Testbed

**ARS**
Application Regression System

**DDR5**
XRDS Phase I

**HBM**
XRDS Phase II

May 2022

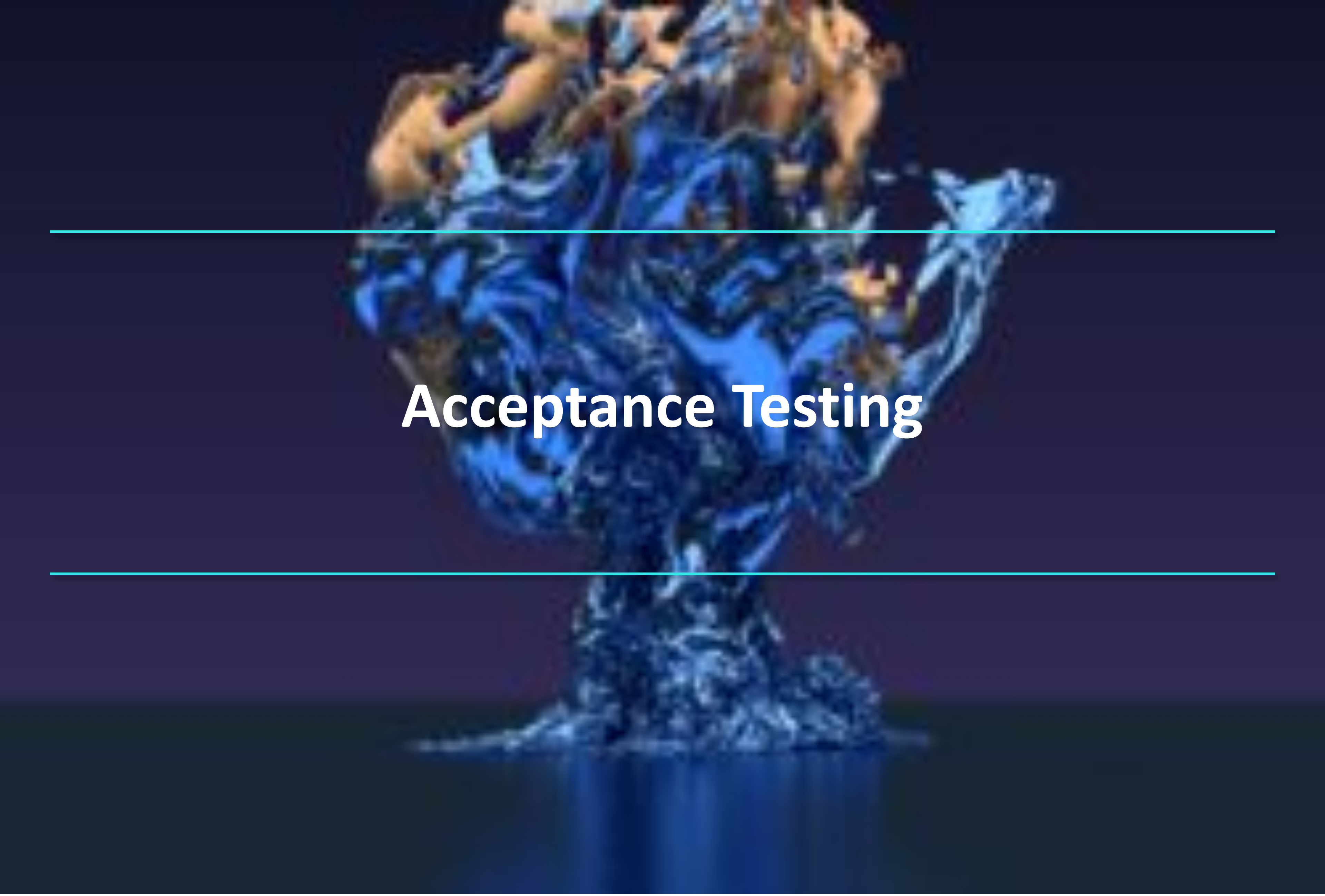
June 2022

July 2022

December 2022

CROSSROADS

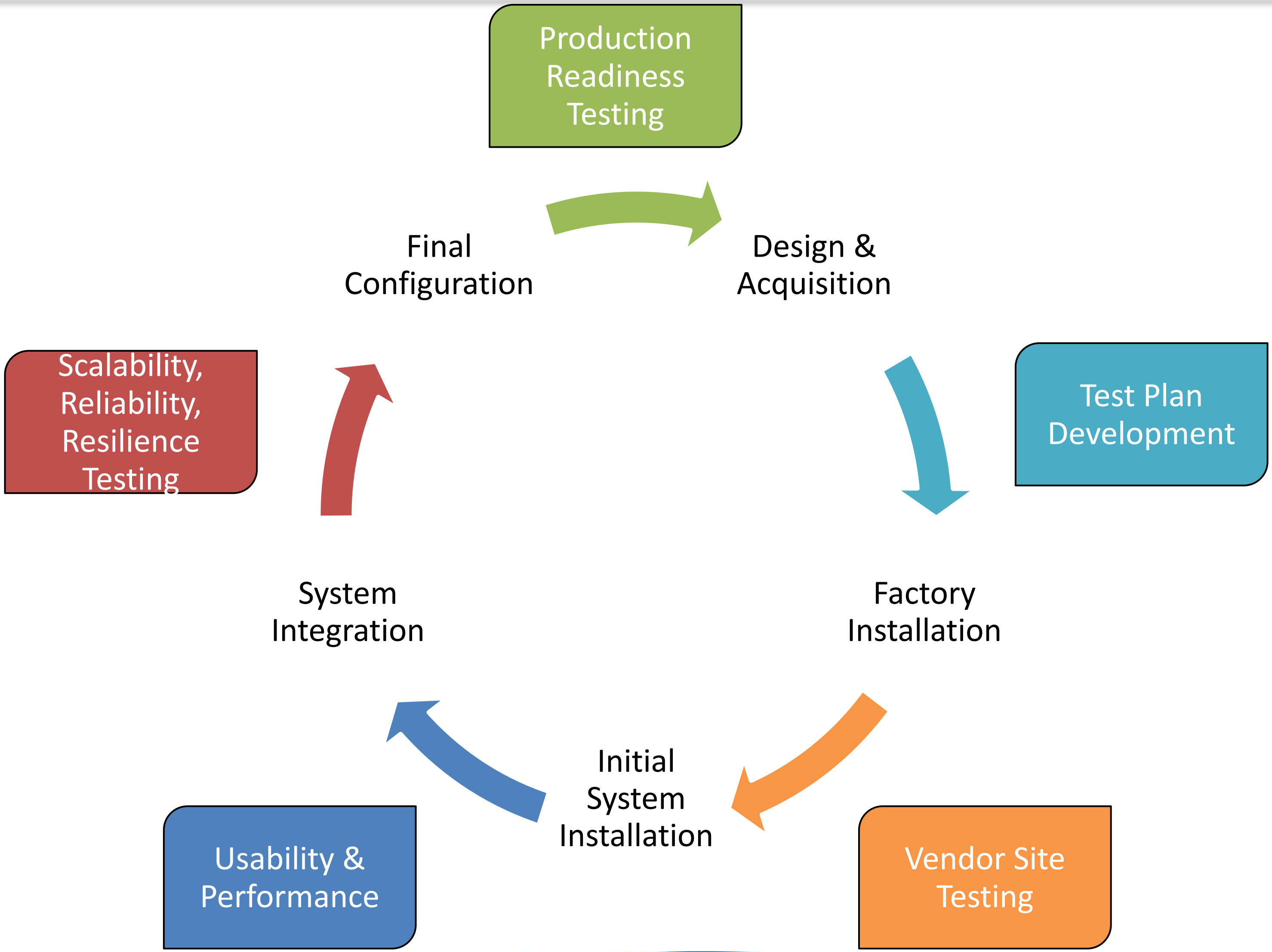
# Acceptance Testing



# Acceptance Testing Phases

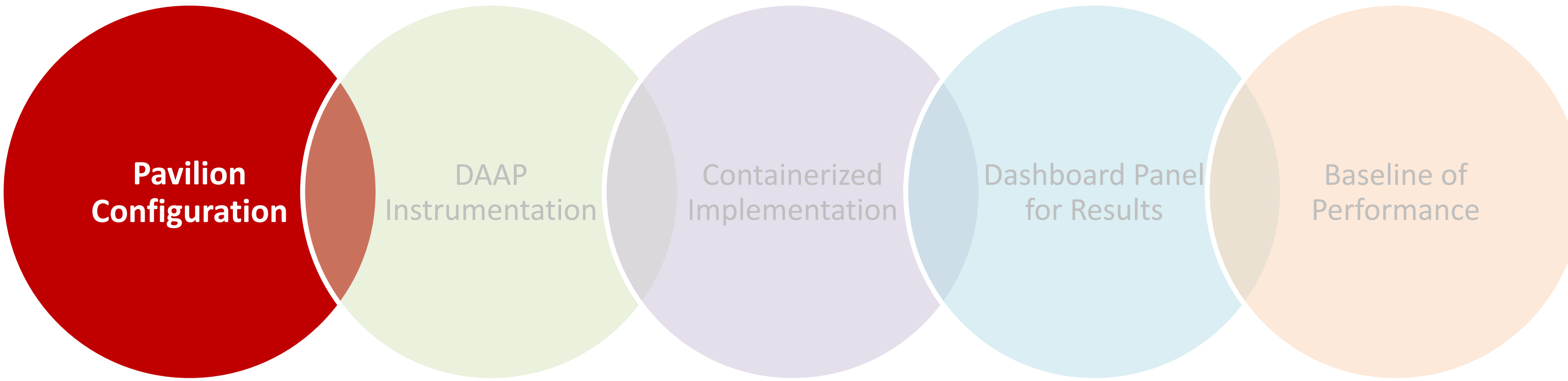
- System Requirements Testing
  - Scalability
  - System Software & Runtime
  - Software Tools and Programming Environment
  - Parallel Storage System
  - Application Performance Requirements
  - Resilience, Reliability & Availability
  - System Operations

# System Procurement Cycle




CROSSROADS

# Test Development Strategy

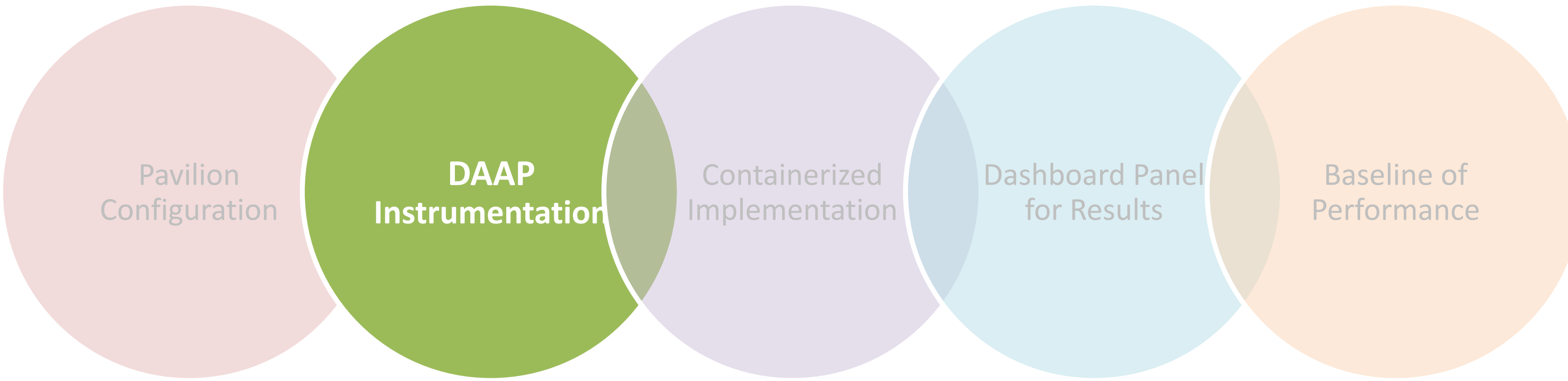


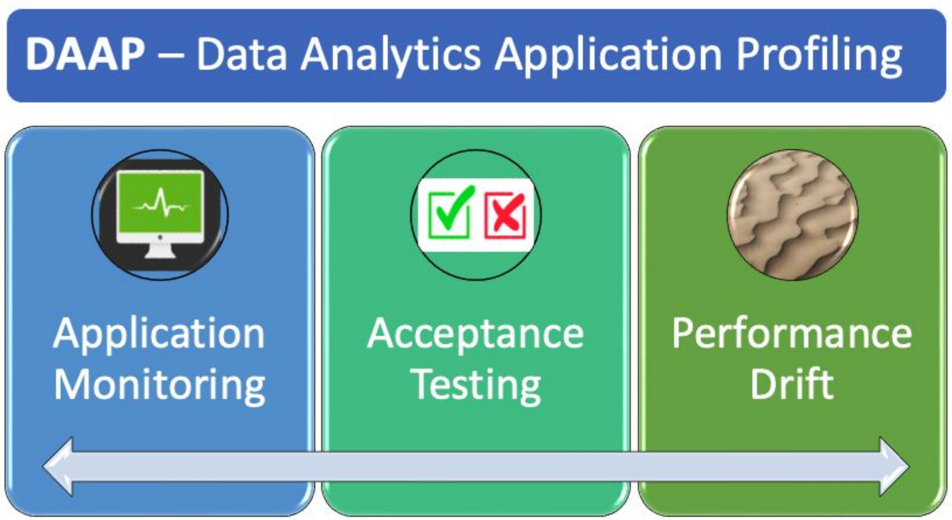

PAVILION
HPC TEST HARNESS

- Implement tests in Pavilion abstractions
  - Eases porting
  - Iterates over software dependencies
  - Permutes inputs
  - Extracts key outputs
  - Feeds analysis tools
  - Enforces uniformity

CROSSROADS

# Test Development Strategy




- ❖ Application Monitoring
- ❖ Acceptance Test Monitoring
- ❖ Machine Performance Regression

# Test Development Strategy

Pavilion Configuration

DAAP Instrumentation

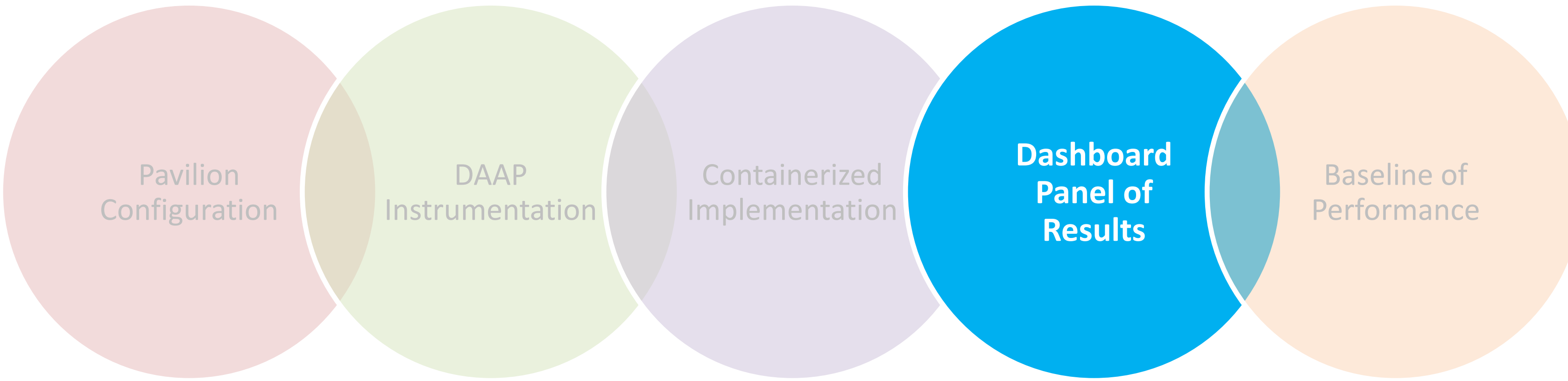
**Containerize Tests**

Dashboard Panel for Results

Baseline of Performance

- Portability
- Shareability
- Evaluates container support capability
- Informs on performance
- Basis for determining best practices for container support

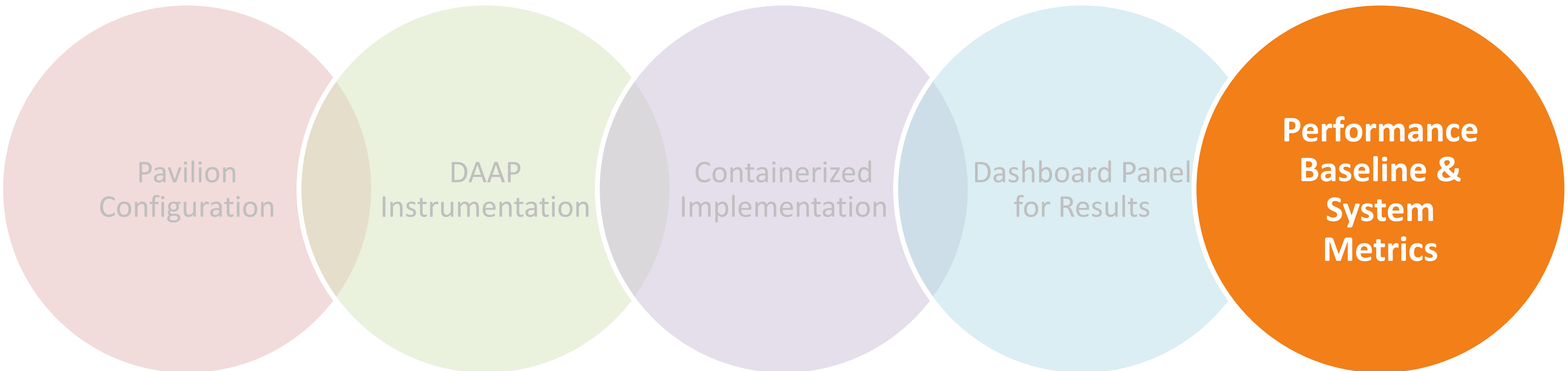
# Test Development Strategy



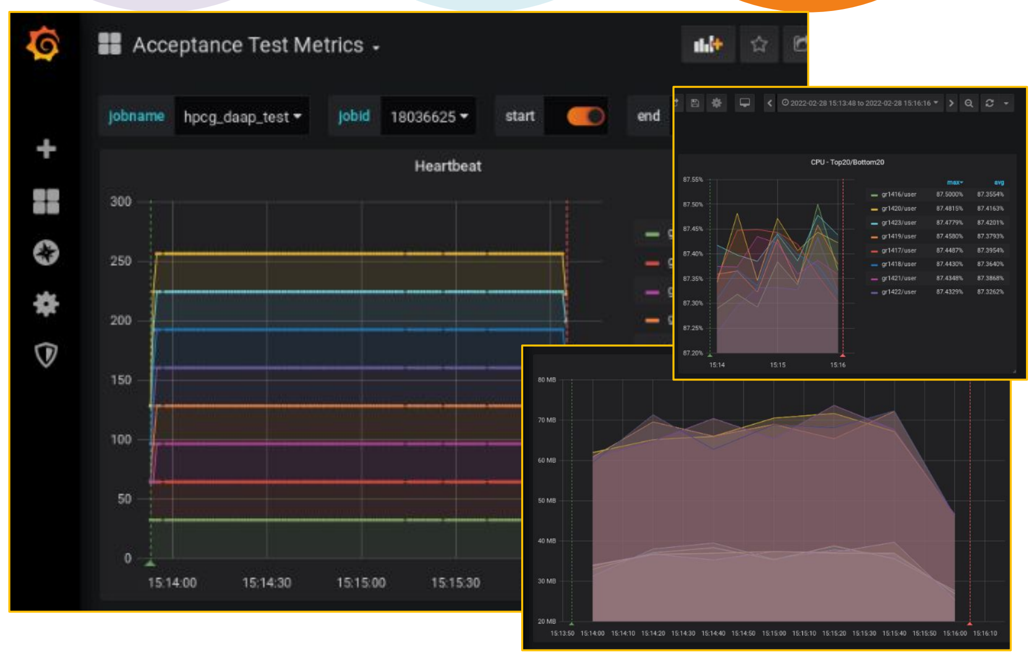


- ❖ Data feed to auto-analysis engine
- ❖ Instant feedback to test team and management
- ❖ Capture system to enforce consistency

# Test Development Strategy

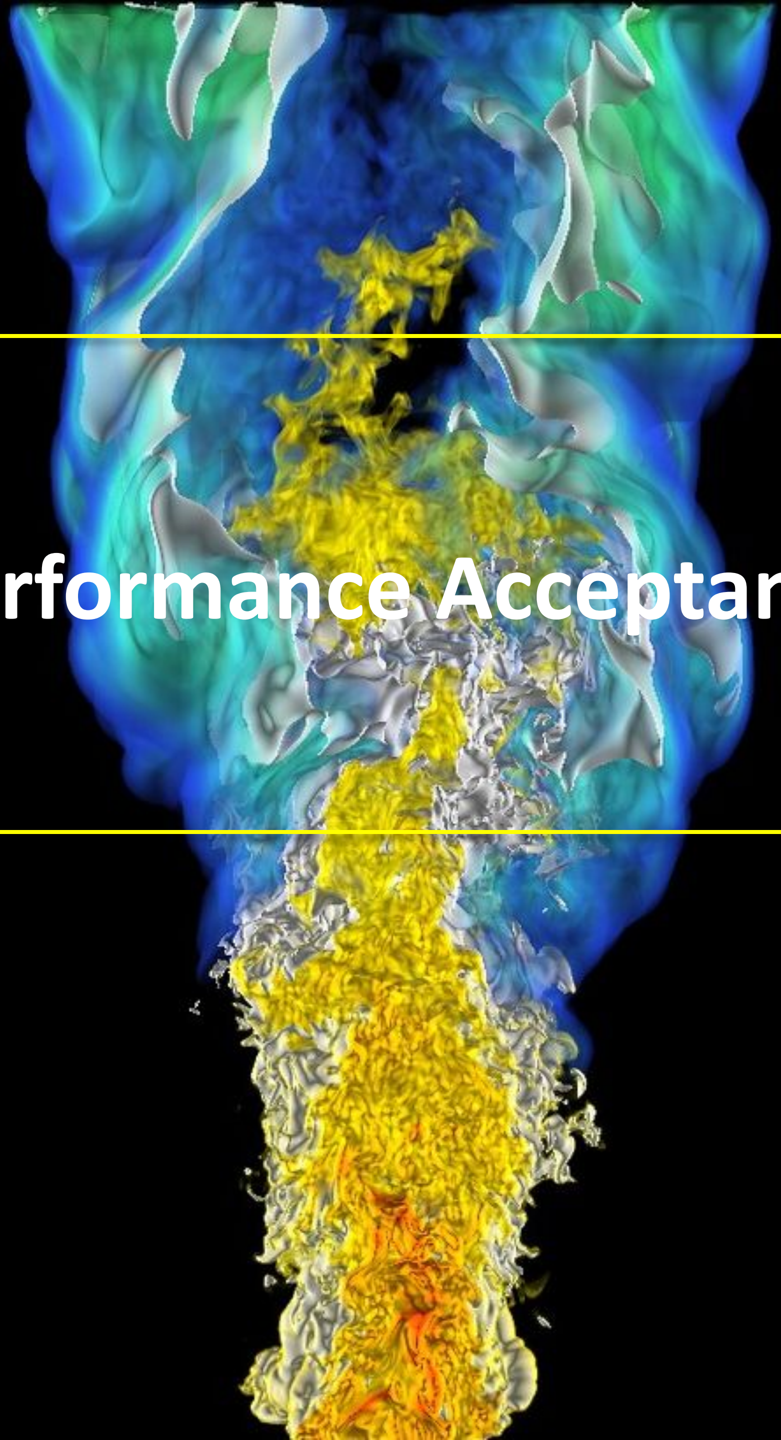



- ❖ Application progress monitoring
- ❖ CPU usage per host
- ❖ MEM usage per host
- ❖ Infiniband (IB) usage per host
- ❖ IB errors per host
- ❖ Reliability data collection, analysis, and reporting




CROSSROADS

# Performance Acceptance



# Performance Acceptance Subgroup Roster

- The **current** roster (alphabetical by lab) for the subgroup is:

### Los Alamos National Laboratory (LANL)

- Christopher DeJager
- Charles Ferenbaugh
- Paul Ferrell
- Timothy Goetsch
- Adam Good
- Jennifer Green
- Hugh Greenberg
- Francine Lapid
- Alex Long
- Daniel Magee
- William Nystrom
- Jordan Ogas
- Howard Pritchard
- Charles Shereda
- Kevin Sheridan
- David Shrader
- Nicholas Sly
- Alfred Torrez

### Sandia National Laboratories (SNL)

- Omar Aaziz
- Anthony Agelastos
- Sam Browne
- Simon Hammond
- Erik Illescas
- Douglas Pase
- Joel Stevenson
- Vanessa Surjadidjaja
- Courtenay Vaughan

This is a team effort!

CROSSROADS

# Performance Benchmarking Applications

## Micro-Benchmarks

1. **DGEMM**: Measures the floating-point capabilities of a single node.
2. **IOR**: Measures parallel file system performance.
3. **mdtest**: Measures the metadata performance of a file system.
4. **STREAM**: Measures memory bandwidth.
5. **MPI Benchmarks**: Measures MPI and high-speed network (HSN) performance.

## Production Applications

1. **PARTISN (LANL)**: Provides neutron transport solutions on orthogonal meshes in 1, 2, and 3 dimensions using a multi-group energy treatment w/ the Sn angular approximation.
2. **Mercury (LLNL)**: Tests performance of Monte Carlo Particle Transport methods.
3. **SPARC (SNL)**: SPARC (Sandia Parallel Aerodynamics and Reentry Code) simulates the aerodynamic environment for atmospheric flight vehicles from subsonic to hypersonic speeds.

## SSI Apps (Mini and Production)

1. **SNAP**: A proxy for modern discrete ordinates neutral particle transport.
2. **HPCG**: A conjugate gradient benchmark.
3. **PENNANT**: A proxy for 2D, unstructured, finite element mesh (FEM) w/ arbitrary polygons.
4. **MiniPIC**: A particle-in-cell (PIC) proxy that solves the discrete Boltzman equation in an electrostatic field within an arbitrary domain w/ reflective walls.
5. **UMT**: A proxy that performs 3D, nonlinear, radiation transport calculations using deterministic (Sn) methods.
6. **VPIC**: A 3D, relativistic, electromagnetic PIC plasma simulation code.
7. **Branson**: A proxy for the Implicit Monte Carlo method to model the exchange of radiation w/ material at high temperatures.

lanl.gov/projects/crossroads/benchmarks-performance-analysis.php

CROSSROADS

# Performance Benchmarking Assessment

- SOW for Crossroads Phase 1 and Phase 2 is still being finalized; the **actual requirements will not be discussed until this occurs**.
- Improvements are relative to ATS-1/Trinity Phase 1 (Intel Haswell).
- **Micro-Benchmarks**: The improvements are are application-specific.
- **SSI Apps**: The improvement(s) with these mini- and production-applications are handled as the Scalable System Improvement (SSI) benchmarking metric (see next).
- **Production Apps**: The improvement(s) with these have historically been handled in aggregate, e.g., with an arithmetic mean of improvement over the baseline.

# Scalable System Improvement (SSI) Metric

$$SSI = \left( \prod_{i=1}^{M} (c_i U_i S_i)^{w_i} \right)^{\frac{1}{\sum_{i=1}^{M} w_i}}$$

- $M$: total # of applications
- $c$: capability scaling factor
- $U$: utilization factor $= \frac{n_{\text{ref}}}{n} \times \frac{N}{N_{\text{ref}}}$
  - $n$: total number of nodes used for the application
  - $N$: total number of nodes in the respective platform
  - $ref$: refers to the reference (i.e., baseline) system
- $S$: application speedup $= \frac{t_{\text{ref}}}{t}$ or $\frac{FOM}{FOM_{\text{ref}}}$
- $w$: weighting factor

# Programming Environment (PE) Focus

Intel oneAPI → Intel oneAPI C/C++ & Intel Classic Fortran → Intel Classic → Cray Programming Environment (CPE) → GNU Compiler Collection (GCC)

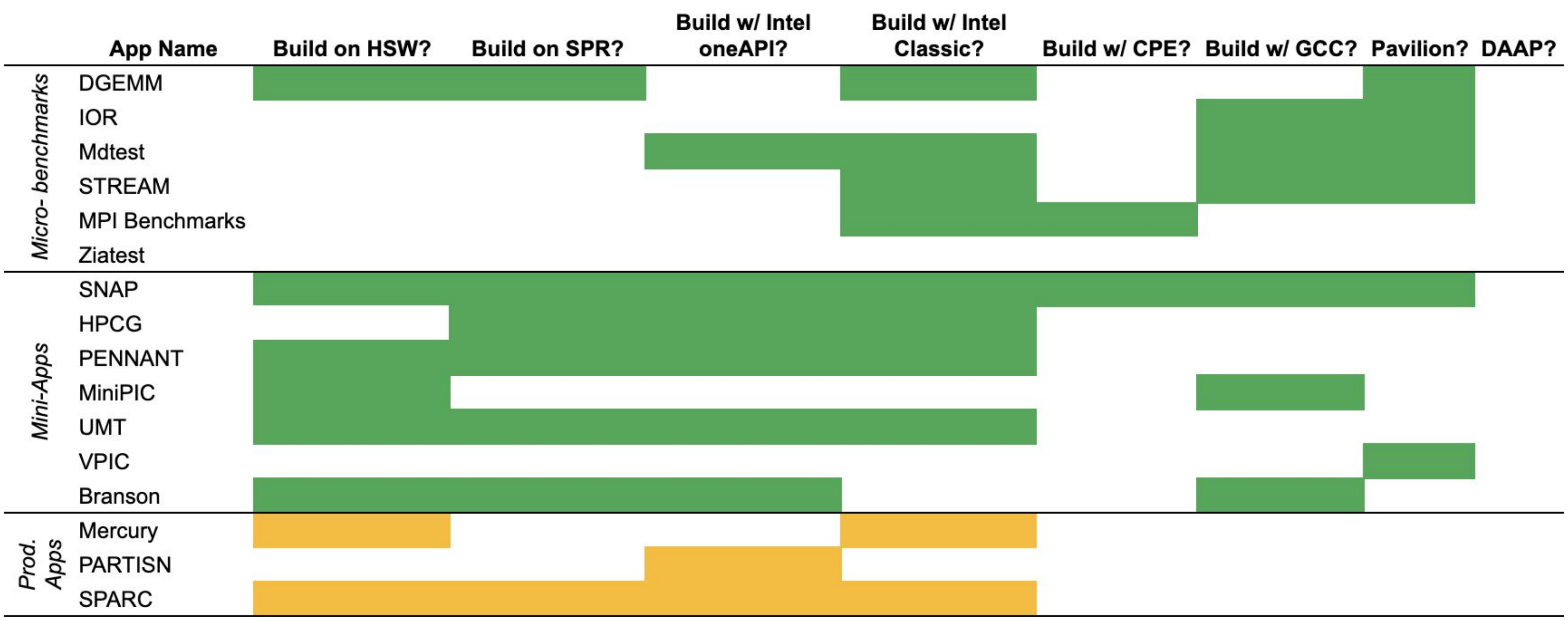
- For each of these, the goals are to:
  - Port application to latest version of PE
    - **Challenge**: Application snapshots are quite old
    - **Challenge**: Intel oneAPI is quite new and some of its components (e.g., Fortran) are not quite ready to replace Intel Classic in all cases
  - Communicate issues/successes to upstream vendors
- The order of preference above stems from generalized NNSA Tri-labs application teams' focus for Crossroads system
  - All PEs will, ultimately, be used by various teams on Crossroads
  - If performance goals are met and time remains, work will still commence until all of these PEs have been investigated

Teaming with vendors enables a healthy ecosystem

# Looking to the Future: Testing

- As the applications are ported and test cases are developed, researchers are integrating them into Pavilion
- This will assist with downstream testing activities extending beyond Acceptance (e.g., platform update testing)
- This will also assist with easy transitioning of test cases from the developers to the testers (team member load balancing)

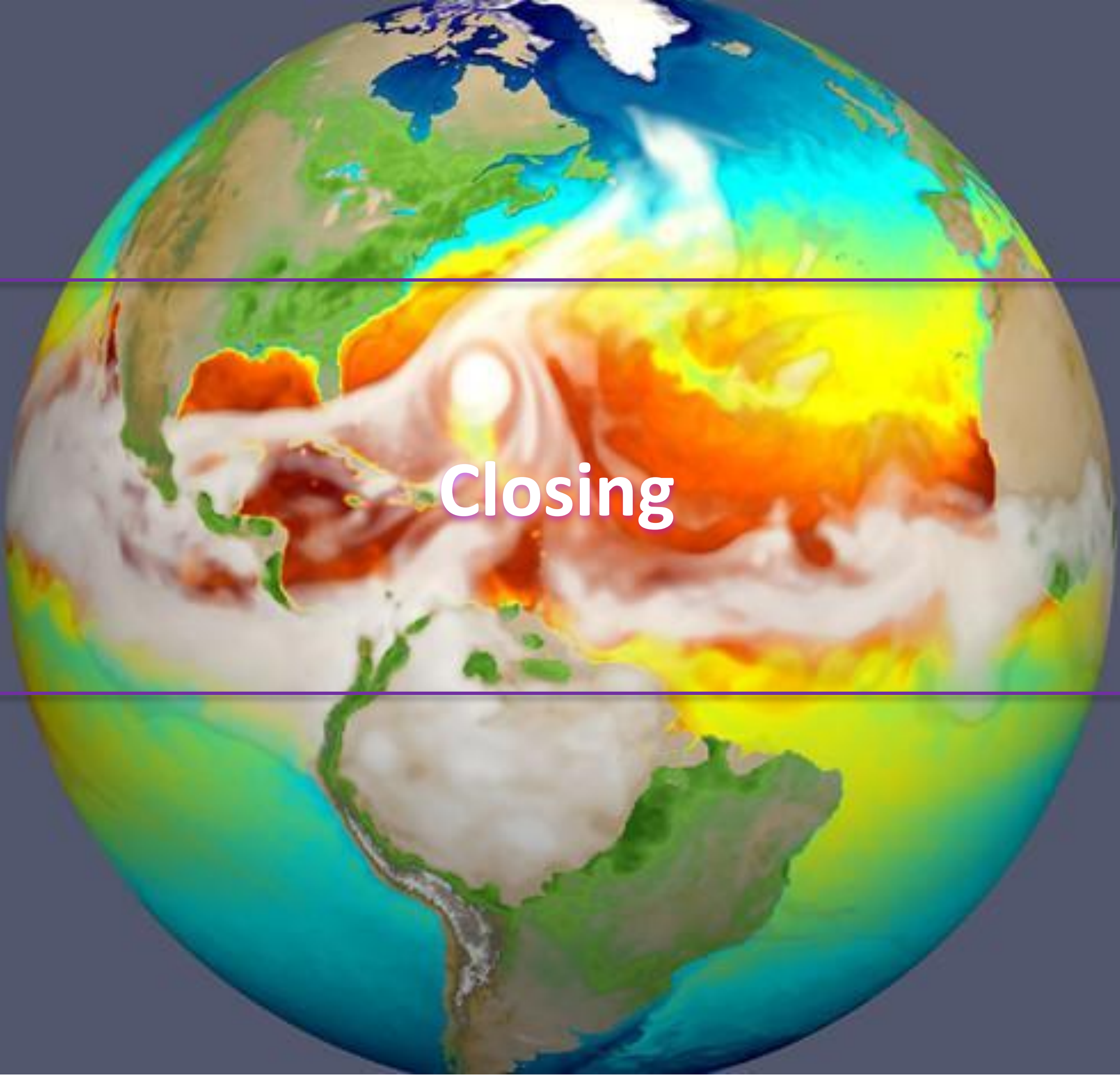
# Porting Status




Good early progress

CROSSROADS

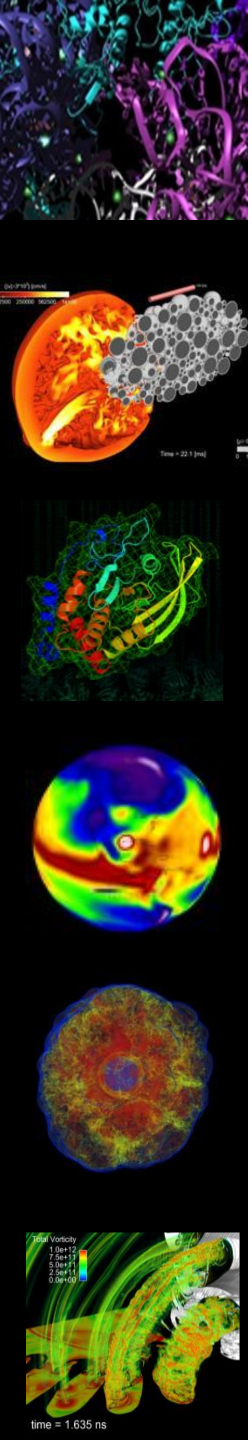
# Closing



CROSSROADS

# Conclusions/Future Work

- Finalized SOW will drive adjustments
- Test development and integration efforts underway
- Functional & Integration testing working with Performance Testing results
  - Feeds production support teams for operation
- Operational test comparisons against baselines
  - Monitor health of the machine
  - Informs next procurement design choices



# Questions?

- Email: xrds-acceptance-testing@lanl.gov
- https://www.lanl.gov/projects/crossroads/benchmarks-performance-analysis.php
- https://www.energy.gov/nnsa/national-nuclear-security-administration
- https://hpc.sandia.gov/aces/
- https://pavilion2.readthedocs.io/