Automated Software Testing of Spack/E4S with buildtest

Shahzeb Siddiqui (LBNL)

Venue: SC23
BoF: Software Testing for Scientific Computing in HPC
Nov 15th, 2023

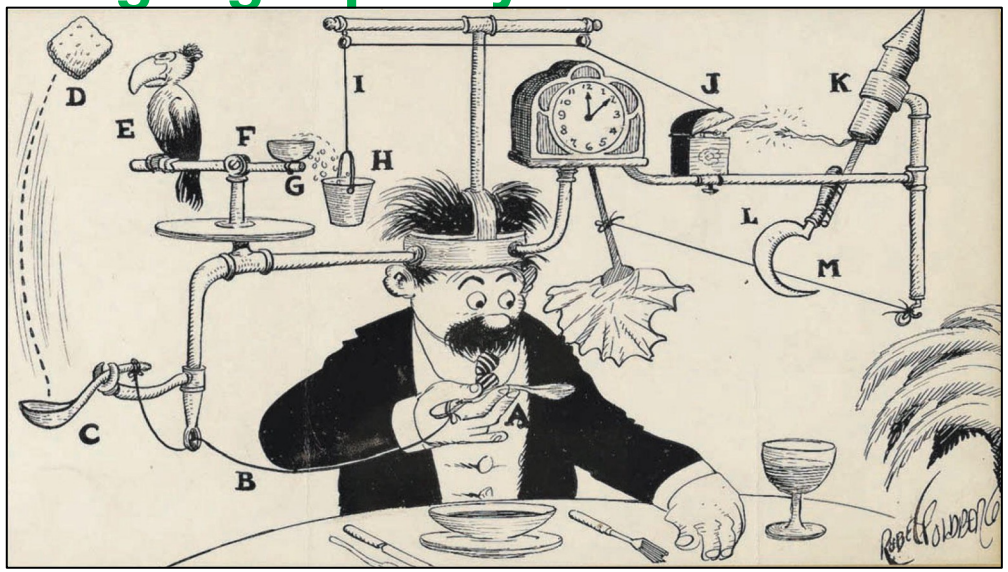
https://buildtest.rtfd.io/
https://github.com/buildtesters/buildtest
http://hpcbuildtest.slack.com/

Outline

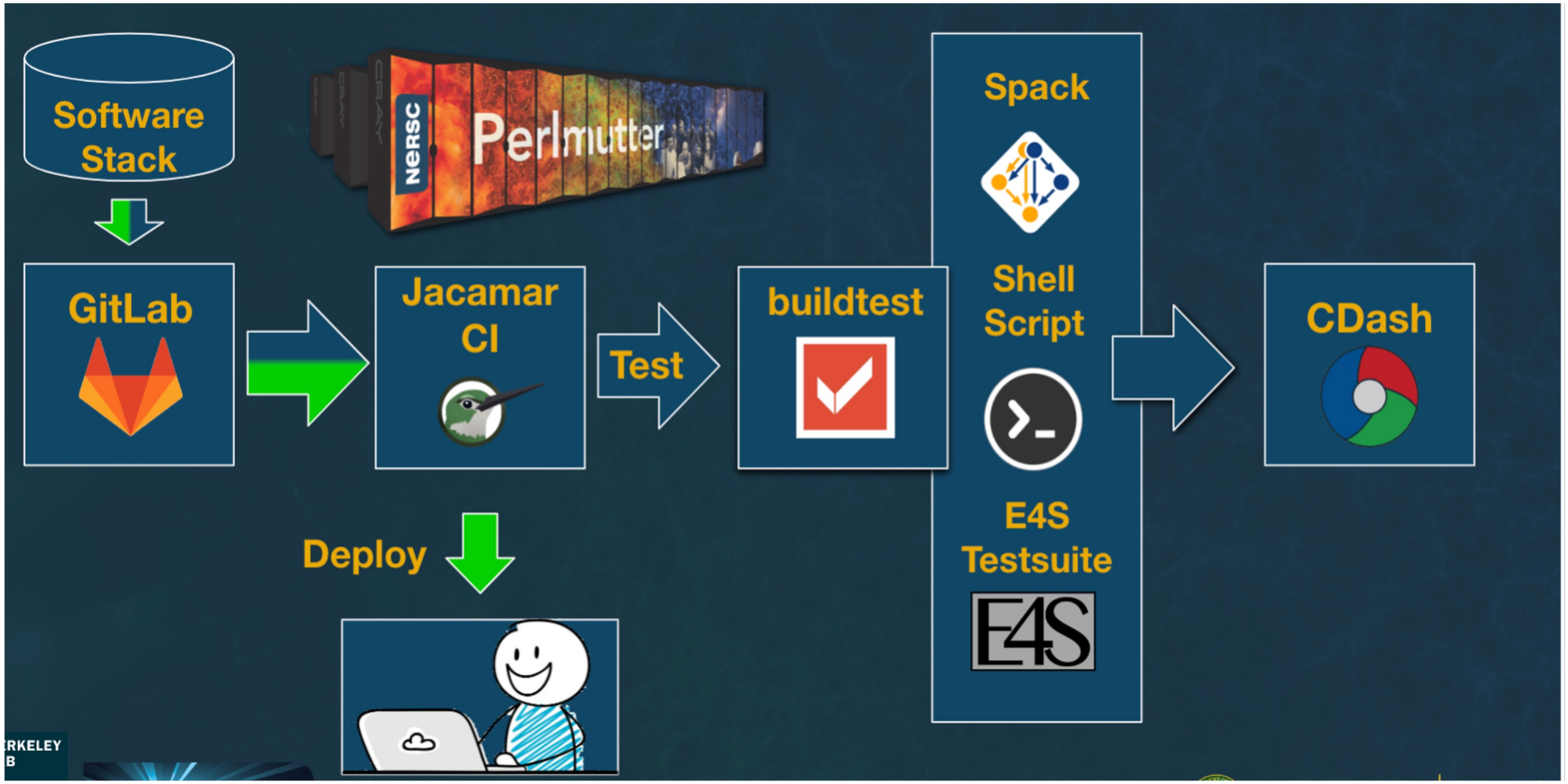
- Deploying Software in HPC is complex
- Workflow for Automated Testing
- Buildtest
- Challenges we faced
- Key Takeaways

Deploying Software in HPC is complex

- Complex dependencies between software packages and the environment
- Incompatibility issues
- Lots of dependencies
- Testing is vital for providing high quality software

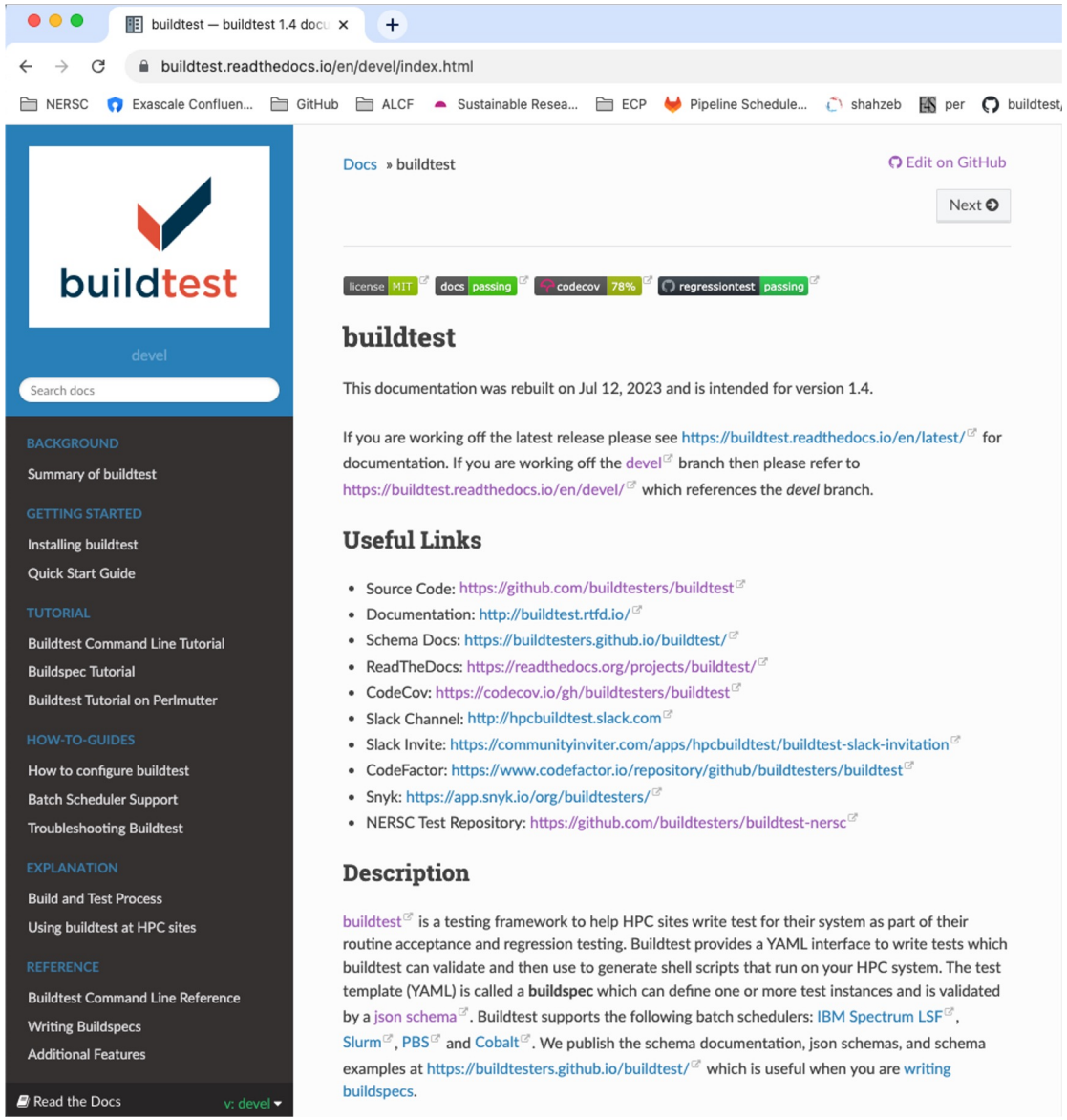


Workflow for Automated Testing



Buildtest

- **Buildtest** is a testing framework that builds and executes tests on your HPC systems
- Intended for **HPC staff**, **developers,** and **end-users**.
- Tests are written in YAML
- Integration with Batch Scheduler, modules, spack, CrayPE



https://github.com/buildtesters/buildtest

Challenges we faced

- Bugs in tests, CI job or buildtest
- High CPU utilization and long runtimes (~45min)
- Gitlab runner availability
- Non Deterministic output due to concurrency limit in CI jobs
- Failures in spack test can’t be fixed in production deployment
- User Environment (shell, startup configuration, modules) can impact test behavior

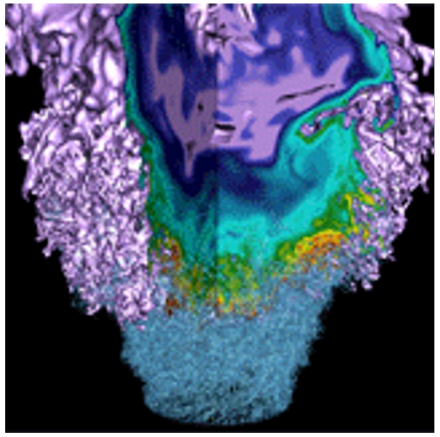
Key Takeaways

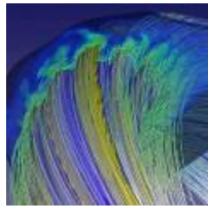
- Use a HPC testing framework (reFrame, buildtest, pavilion2) to conduct site-local testing
- Infrastructure issues are bound to happen. Communication amongst groups is essential
- Spack for automated testing has its issues
  - Single Repository: Not feasible for facilities to fix test issues that are site-specific
  - Not compatible with batch schedulers for testing on compute nodes

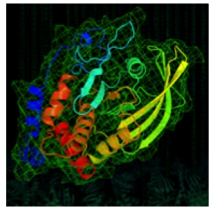
Acknowledgement

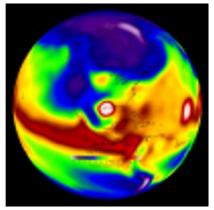
This research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of two U.S. Department of Energy organizations (Office of Science and the National Nuclear Security Administration) responsible for the planning and preparation of a capable exascale ecosystem, including software, applications, hardware, advanced system engineering and early testbed platforms, in support of the nation's exascale computing imperative.

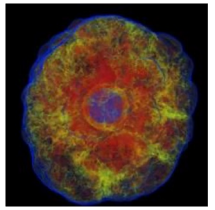
Reference Slides

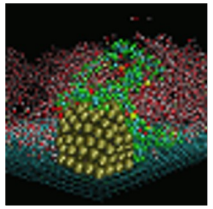












Slides: https://docs.google.com/presentation/d/15xW4tA2eoMatMZs1cqX2fxhbuQB2bWdv/edit?usp=sharing&ouid=102126664227037583807&rtpof=true&sd=true

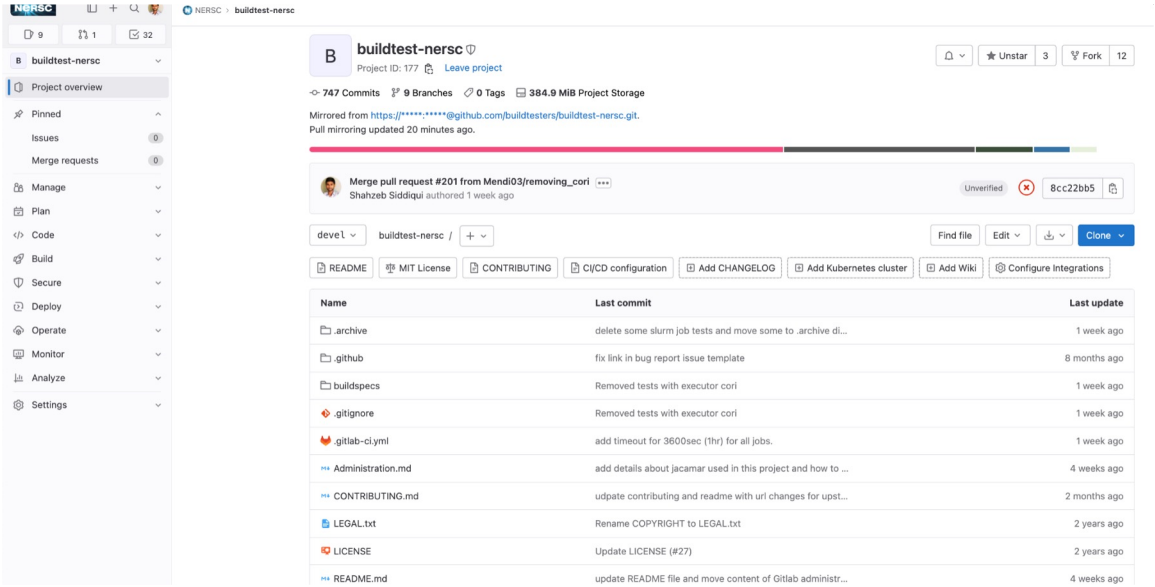
Buildtest-NERSC: Post-Deployment Testing and Automation

Overview

- [Buildtest-NERSC](#) is a collection of tests (mostly Spack/E4S) that are performed after a software deployment to ensure its quality. It utilizes [buildtest](#) to manage and run these tests smoothly

Operational Mechanism

- Utilizes Gitlab CI for automated test execution
- Scheduled cron runs for consistent testing
- Test results shared via CDASH
- Gitlab runners on Perlmutter and Muller to enable CI workloads
- All test are run using the default branch (devel)



https://github.com/buildtesters/buildtest-nersc

buildtest + Spack

```yaml
buildspecs:
  spack_test_amrex_e4s_22.05:
    type: spack
    executor: '(perlmutter|muller).local.sh'
    description: "Test AMReX for e4s/22.05 stack for Perlmutter via spack test"
    tags: e4s
    pre_cmds: |
      module swap PrgEnv-nvidia PrgEnv-gnu
      module load e4s/22.05
    spack:
      root: /global/common/software/spackecp/perlmutter/e4s-22.05/spack/
      verify_spack: false
      test:
        run:
          specs: ['amrex']
        results:
          option: '-l'
          specs: ['amrex']
```

- Spack is a package manager for supercomputers
- Used to deploy our software stack
- Spack supports stand-alone testing which buildtest can leverage

buildtest + E4S Testsuite

```yaml
#This buildtest was generated from e4s-template.yaml
buildspecs:
  hdf5_e4s_testsuite_22.05:
    type: script
    executor: perlmutter.slurm.regular
    description: Run E4S Testsuite hdf5 test for e4s/22.05 stack
    tags: [e4s]
    sbatch: ["-t 30", "-N 1", "-G 1", "-A [redacted]", "-C gpu"]
    run: |
      module load e4s/22.05
      spack load --first cmake
      git clone https://github.com/E4S-Project/testsuite
      cd testsuite

      sh test-all.sh --color-off validation_tests/hdf5 --print-logs
maintainers:
  - shahzebsiddiqui
  - wspear
```

- E4S Testsuite includes over 100 tests, with a goal of full test coverage for the E4S ecosystem
- Test code largely extracted from package sources
- Uses a simple, extensible bash-script test driver

What does a typical build look like

```
(buildtest) ~/ buildtest build -b $BUILDTEST_ROOT/tutorials/hello_world.yml
                                    buildtest summary

User:               siddiq90
Hostname:           DOE-7086392
Platform:           Darwin
Current Time:       2023/02/02 11:31:37
buildtest path:     /Users/siddiq90/Documents/github/buildtest/bin/buildtest
buildtest version:  1.1
python path:        /Users/siddiq90/.local/share/virtualenvs/buildtest-Ir4AdrfC/bin/python3
python version:     3.7.3
Configuration File: /Users/siddiq90/Documents/github/buildtest/buildtest/settings/config.yml
Test Directory:     /Users/siddiq90/Documents/github/buildtest/var/tests
Report File:        /Users/siddiq90/Documents/github/buildtest/var/report.json
Command:            /Users/siddiq90/Documents/github/buildtest/bin/buildtest build -b /Users/siddiq90/Documents/github/buildtest/tutorials/hello_world.yml

                                 Discovering Buildspecs
                        Discovered buildspecs

| buildspec                                                |
| /Users/siddiq90/Documents/github/buildtest/tutorials/hello_world.yml |

Total Discovered Buildspecs:  1
Total Excluded Buildspecs:  0
Detected Buildspecs after exclusion:  1
```

```
                                    Running Tests
Spawning 1 processes for processing builders
                                     Iteration 1
hello_world/947d2c9b does not have any dependencies adding test to queue
In this iteration we are going to run the following tests: [hello_world/947d2c9b]
hello_world/947d2c9b: Current Working Directory : /Users/siddiq90/Documents/github/buildtest/var/tests/generic.local.bash/hello_world/hello_world/947d2c9b/stage
hello_world/947d2c9b: Running Test via command: bash --norc --noprofile -eo pipefail hello_world_build.sh
hello_world/947d2c9b: Test completed in 0.084696 seconds
hello_world/947d2c9b: Test completed with returncode: 0
hello_world/947d2c9b: Writing output file -  /Users/siddiq90/Documents/github/buildtest/var/tests/generic.local.bash/hello_world/hello_world/947d2c9b/hello_world.out
hello_world/947d2c9b: Writing error file - /Users/siddiq90/Documents/github/buildtest/var/tests/generic.local.bash/hello_world/hello_world/947d2c9b/hello_world.err
                                      Test Summary

| builder | executor | status | checks (ReturnCode, Regex, Runtime) | returncode | runtime |
| hello_world/947d2c9b | generic.local.bash | PASS | N/A N/A N/A | 0 | 0.084696 |

Passed Tests: 1/1 Percentage: 100.000%
Failed Tests: 0/1 Percentage: 0.000%

Adding 1 test results to /Users/siddiq90/Documents/github/buildtest/var/report.json
Writing Logfile to: /Users/siddiq90/Documents/github/buildtest/var/logs/buildtest_o8vabck6.log
```

CDASH - Web Based Dashboard

5 passed, 3 failed, 0 not run, 3 missing.

Name	Status	Time	Details	Labels	History	Summary	compiler	description	endtime	hostname	starttime	user
e4s_hdf5	Failed	1h 46m 32s 940ms		e4s	Broken	Broken		Run hdf5 test from E4S Testsuite. This test is expected to fail	2022/10/18 08:50:44	cori02	2022/10/18 07:04:11	e4s
e4s_testsuite_e4s_22.02_gnu	Failed	2h 4m 6s 160ms		e4s	Broken	Broken		Run E4S Testsuite for e4s/22.02, gnu stack	2022/10/18 09:08:17	cori02	2022/10/18 07:04:11	e4s
e4s_testsuite_e4s_22.02_intel	Failed	2h 4m 36s 360ms		e4s	Broken	Broken		Run E4S Testsuite for e4s/22.02, intel stack	2022/10/18 09:08:47	cori02	2022/10/18 07:04:11	e4s
ascent_e4stestsuite_e4s_21.05	Missing											
e4s_adios2	Missing											
e4s_superlu-dist	Missing											
e4s_22_02_spack	Passed	6s 660ms		e4s spack	Unstable	Stable		Check e4s/21.05 spack instance	2022/10/18 07:04:17	cori02	2022/10/18 07:04:11	e4s
showquota	Passed	2s 170ms			Stable	Stable		run the showquota utility	2022/10/18 07:04:13	cori02	2022/10/18 07:04:11	e4s
showquota_check_m3503	Passed	830ms		e4s	Stable	Stable		run the showquota check on m3503 project directory	2022/10/18 07:04:12	cori02	2022/10/18 07:04:11	e4s
showquota_check_spackecp	Passed	680ms		e4s	Stable	Stable		run showquota on /global/common/software/spackecp	2022/10/18 07:04:11	cori02	2022/10/18 07:04:11	e4s
spack_e4s_directory_checks	Passed	80ms		e4s	Stable	Stable		Check for production directory paths for spack e4s stack	2022/10/18 07:04:11	cori02	2022/10/18 07:04:11	e4s

https://my.cdash.org/index.php?project=buildtest-nersc

- CDASH produces color coded charts, timelines (track progress over time), and links to test output
- Comprehensive presentation of test results posted to the web for users and developers
- Buildtest can upload test results to CDASH

Test Results and metadata captured in CDASH

Test: spack_test_gasnet_e4s_22.05 **(Passed)**
Build: e4s (perlmutter) on 2023-08-03 15:20:22
Labels: e4s

compiler	
description	Test gasnet build with for e4s/22.05 stack on Perlmutter via spack test
endtime	2023/08/03 08:21:54
hostname	login07
starttime	2023/08/03 08:20:22
user	bdtest
Return Code	0
Test ID	234fd2a0-4d4c-4493-9706-c4ccc4c468f3
build_script	/global/cfs/cdirs/m3503/buildtest/runs/perlmutter_scheduled_test/2023-08-03/perlmutter.local.sh/gasnet/spack_test_gasnet_e4s_22.05/234fd2a0/spack_test_gasnet_e4s_22.05_build.sh
buildenv	/global/cfs/cdirs/m3503/buildtest/runs/perlmutter_scheduled_test/2023-08-03/perlmutter.local.sh/gasnet/spack_test_gasnet_e4s_22.05/234fd2a0/build-env.txt
command	sh --norc --noprofile -eo pipefail spack_test_gasnet_e4s_22.05_build.sh
errfile	/global/cfs/cdirs/m3503/buildtest/runs/perlmutter_scheduled_test/2023-08-03/perlmutter.local.sh/gasnet/spack_test_gasnet_e4s_22.05/234fd2a0/spack_test_gasnet_e4s_22.05.err
executor	perlmutter.local.sh
id	234fd2a0
logpath	/global/cfs/cdirs/m3503/buildtest-ci-data-dir/siddiq90/builds/vKz_1yCX/001/NERSC/buildtest-nersc/buildtest/var/logs/buildtest_tg1sm8l6.log
outfile	/global/cfs/cdirs/m3503/buildtest/runs/perlmutter_scheduled_test/2023-08-03/perlmutter.local.sh/gasnet/spack_test_gasnet_e4s_22.05/234fd2a0/spack_test_gasnet_e4s_22.05.out
schemafile	spack.schema.json
stagedir	/global/cfs/cdirs/m3503/buildtest/runs/perlmutter_scheduled_test/2023-08-03/perlmutter.local.sh/gasnet/spack_test_gasnet_e4s_22.05/234fd2a0/stage
tags	e4s
testpath	/global/cfs/cdirs/m3503/buildtest/runs/perlmutter_scheduled_test/2023-08-03/perlmutter.local.sh/gasnet/spack_test_gasnet_e4s_22.05/234fd2a0/spack_test_gasnet_e4s_22.05.sh
testroot	/global/cfs/cdirs/m3503/buildtest/runs/perlmutter_scheduled_test/2023-08-03/perlmutter.local.sh/gasnet/spack_test_gasnet_e4s_22.05/234fd2a0

View GitLab CI results

Test Content

```
#!/bin/bash


######## START OF PRE COMMANDS ########
module load e4s/22.05
spack env activate -V gcc
spack load --first gasnet

######## END OF PRE COMMANDS   ########


source /global/common/software/spackecp/perlmutter/e4s-22.05/78535/spack/share/spack/setup-env.sh
spack test run --alias 6790f3ef-d99e-4d59-a5bf-c8c2f43bca27 gasnet
spack test results -l -- gasnet
```

Test output

```
==> Spack test 6790f3ef-d99e-4d59-a5bf-c8c2f43bca27
==> Testing package gasnet-2022.3.0-bwmz3br
============================ 1 passed of 1 specs ============================
==> Results for test suite '06835aa3-7b90-4347-8691-f8247e0759db', spec matching 'gas
==> test specs:
==>   gasnet-2022.3.0-bwmz3br PASSED
==> Testing package gasnet-2022.3.0-bwmz3br
==> [2023-05-28-19:04:12.181135] Running testtools
==> [2023-05-28-19:04:12.181522] '/global/common/software/spackecp/perlmutter/e4s-22.
=====> testtools config=RELEASE=2.022.3.0,SPEC=1.19,PTR=64bit,nodebug,PAR,timers_nativ
hostname is: login31 (pid=242354)
testtools running...
gasnett_exe_name()='/global/common/software/spackecp/perlmutter/e4s-22.05/78535/spack
Running testtools with 100 iterations and 10 threads
System page size is 2^12 == 4096
CPU count estimated to be: 256
Cache line size estimated to be: 64
Physical memory size estimated to be: 503 GB
A: Testing high-performance timers and sleep...
B: Testing zero-byte counting...
C: Testing local membars...
```