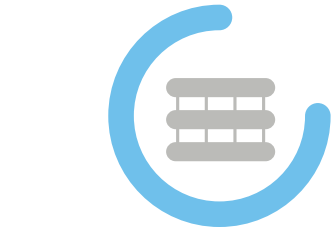


Meeting of the Technical Steering Committee (TSC) Board

Wednesday, February 10th, 2021
11:00am ET

Antitrust Policy Notice

- Linux Foundation meetings involve participation by industry competitors, and it is the intention of the Linux Foundation to conduct all of its activities in accordance with applicable antitrust and competition laws. It is therefore extremely important that attendees adhere to meeting agendas, and be aware of, and not participate in, any activities that are prohibited under applicable US state, federal or foreign antitrust and competition laws.

- Examples of types of actions that are prohibited at Linux Foundation meetings and in connection with Linux Foundation activities are described in the Linux Foundation Antitrust Policy available at http://www.linuxfoundation.org/antitrust-policy. If you have questions about these matters, please contact your company counsel, or if you are a member of the Linux Foundation, feel free to contact Andrew Updegrove of the firm of Gesmer Updegrove LLP, which provides legal counsel to the Linux Foundation.

THE LINUX FOUNDATION

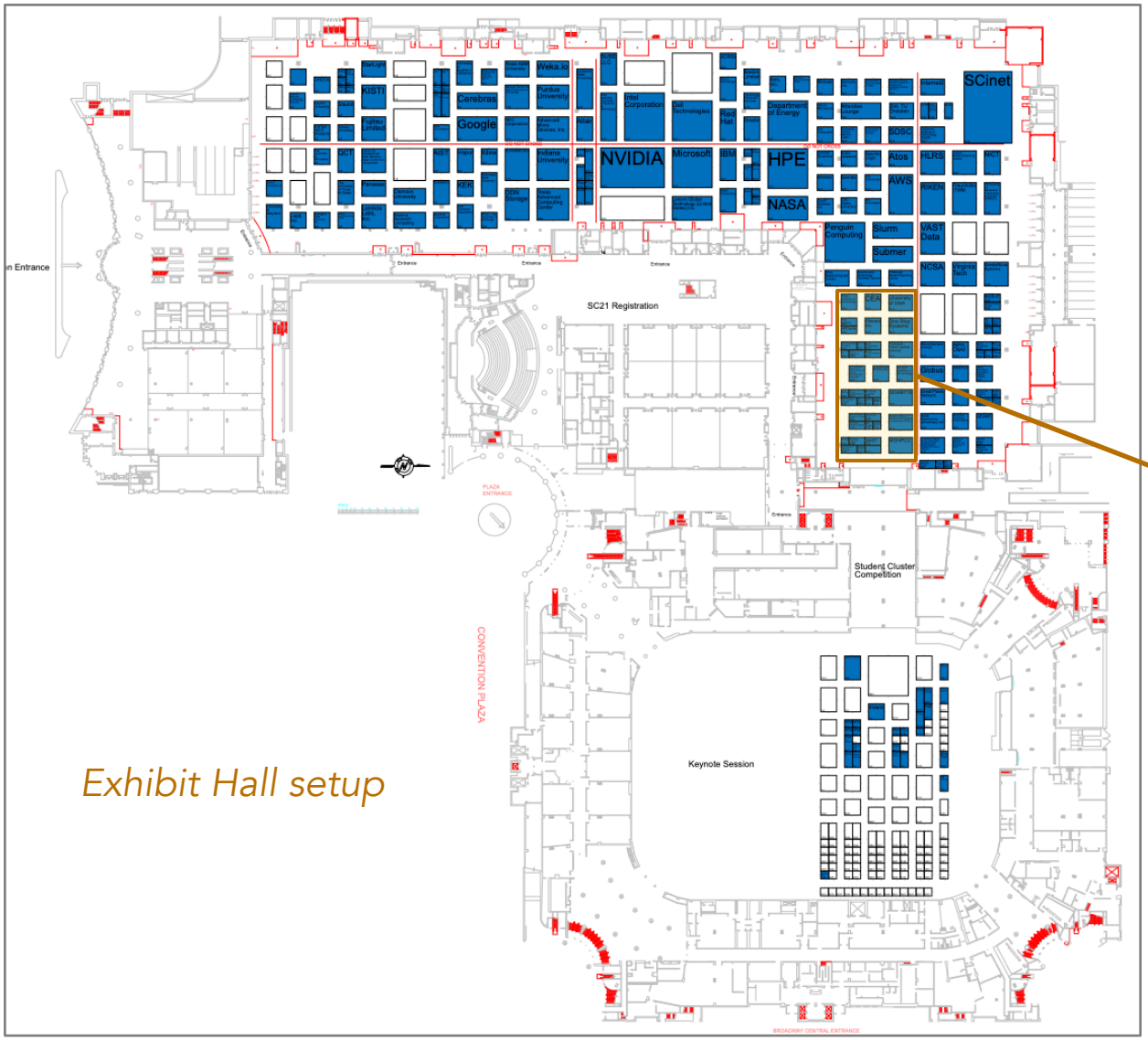
Agenda/Updates

- Announcements, Upcoming talks and deadlines
  - ISC 2021
    - Accepted BoF Sessions of ISC 2020 will be held at ISC 2021
  - PEARC'21 BoF submission – due May 9
  - SC'21 BoF submission – due August 6
  - Infrastructure outage:
    - OpenHPC rack at TACC will be relocating in their datacenter
    - CI and OBS build system will be affected
    - Hopefully only a 1-day outage: beginning next Tuesday (2/16) at 8am (US/central)
  - Interesting community tidbit regarding RHEL/CentOS8 discussion
    - one NASA group is waiting to hear how OpenHPC reacts going forward

---

- Linaro CI system updates?
- SC'21 Booth
- Cloud working group
  - PEARC'21 discussion
- 2.1 Build/Test status and changes

SC'21 – OpenHPC Booth (St. Louis, MO)



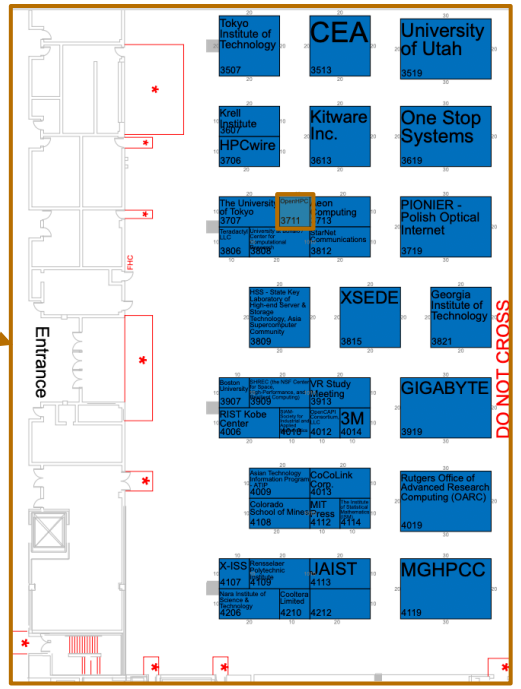


Exhibit Hall setup

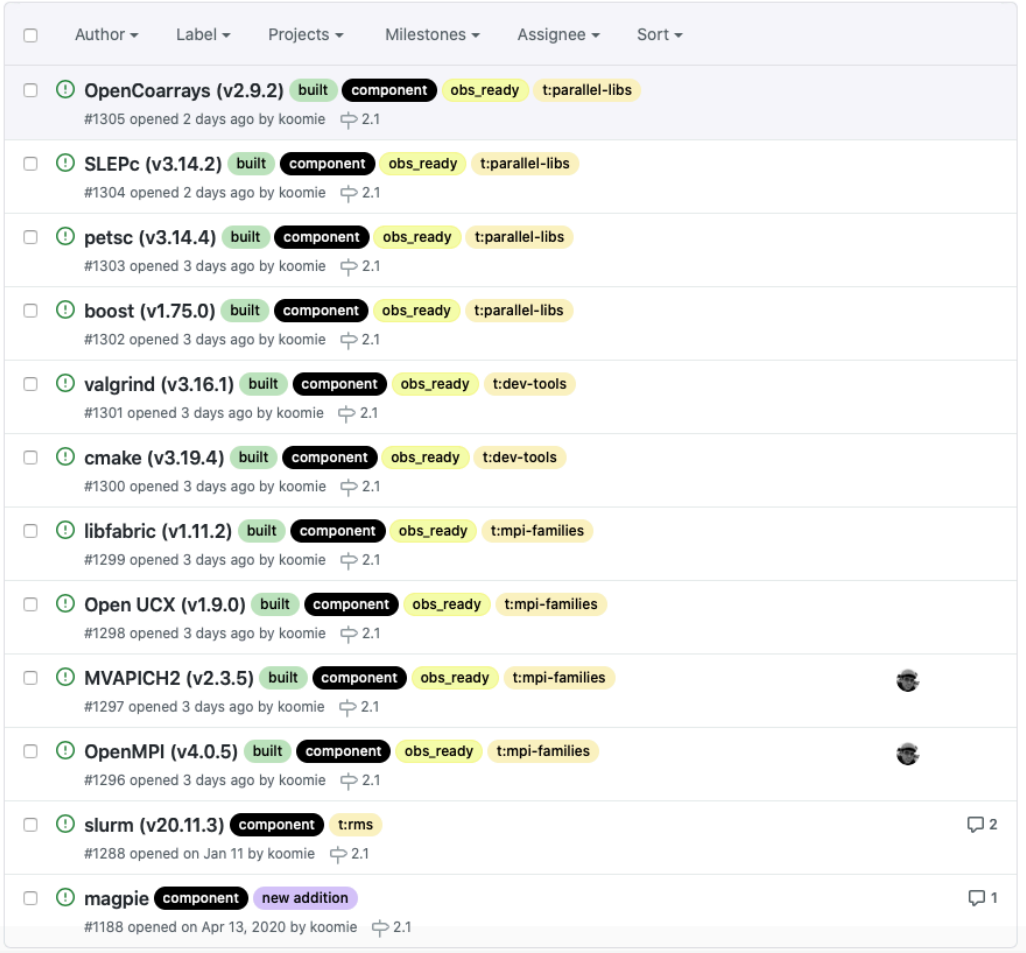
OHPC Cloud Updates

- We decided to not do a PEARC tutorial this year
- Will focus on submitting up to 2 short papers
  - Experiences with containers + MPI
  - Deploying OpenHPC at AWS
- Will also submit a BOF to discuss issues around CentOS

Additional Activities for 2021

- AWS Blog Post by Chris D. on OpenHPC at AWS
- Github-pages based documentation update
  - Detailed discussion of design considerations
  - Released with the blog post and 2.1 release
- New implementation of cloud infrastructure using terraform
- CI/CD infrastructure to support cloud releases
- Aim to have our first formal release/recipe for SC

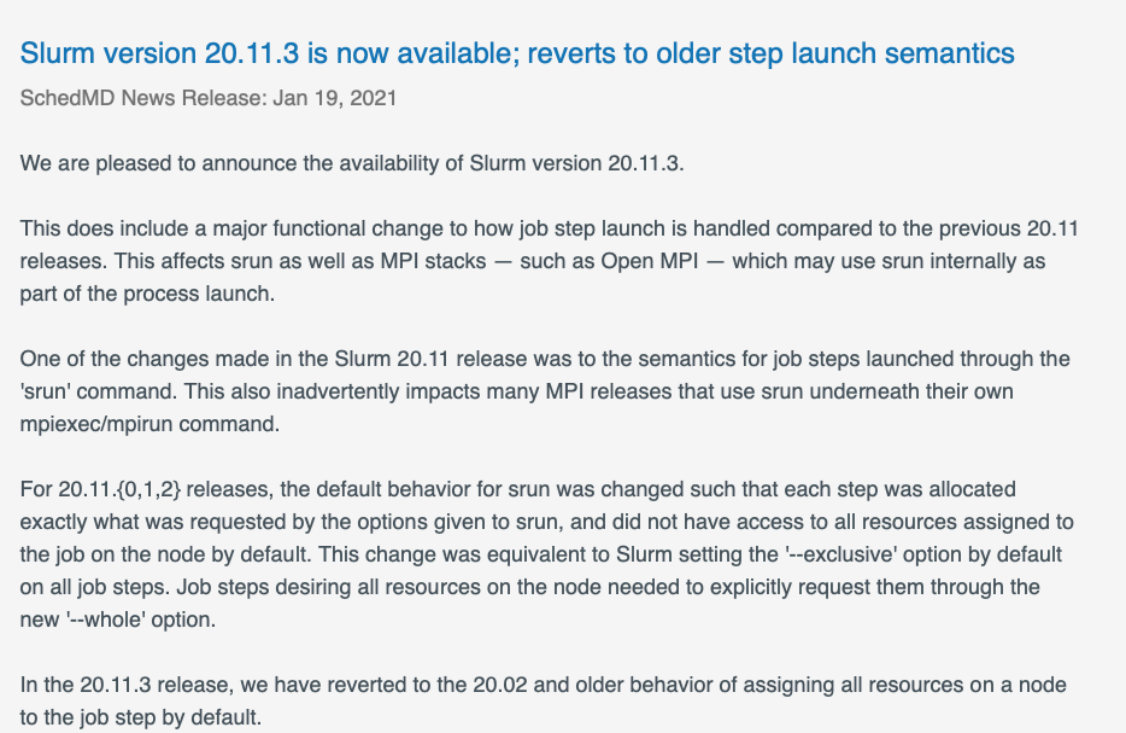
v2.1 Work in Progress



- scanned upstream for a several package updates for 2.1 (in addition to testing against CentOS 8.3)
  - package versions highlighted here
  - majority are built in our OBS system

SLURM update

- was holding off jumping to newer slurm based on comments from others who were burned with MPI job launch for openmpi
  - https://www.schedmd.com/news.php?id=244#OPT_244
  - https://bugs.schedmd.com/show_bug.cgi?id=10383
  - sites probably want to avoid v20.11.{0,1,2} of SLURM
- Previous job launch functionality restored in v20.11.3 (we will try to go with this version; revert to 20.02.6 if still problems)



Slurm version 20.11.3 is now available; reverts to older step launch semantics

SchedMD News Release: Jan 19, 2021

We are pleased to announce the availability of Slurm version 20.11.3.

This does include a major functional change to how job step launch is handled compared to the previous 20.11 releases. This affects srun as well as MPI stacks — such as Open MPI — which may use srun internally as part of the process launch.

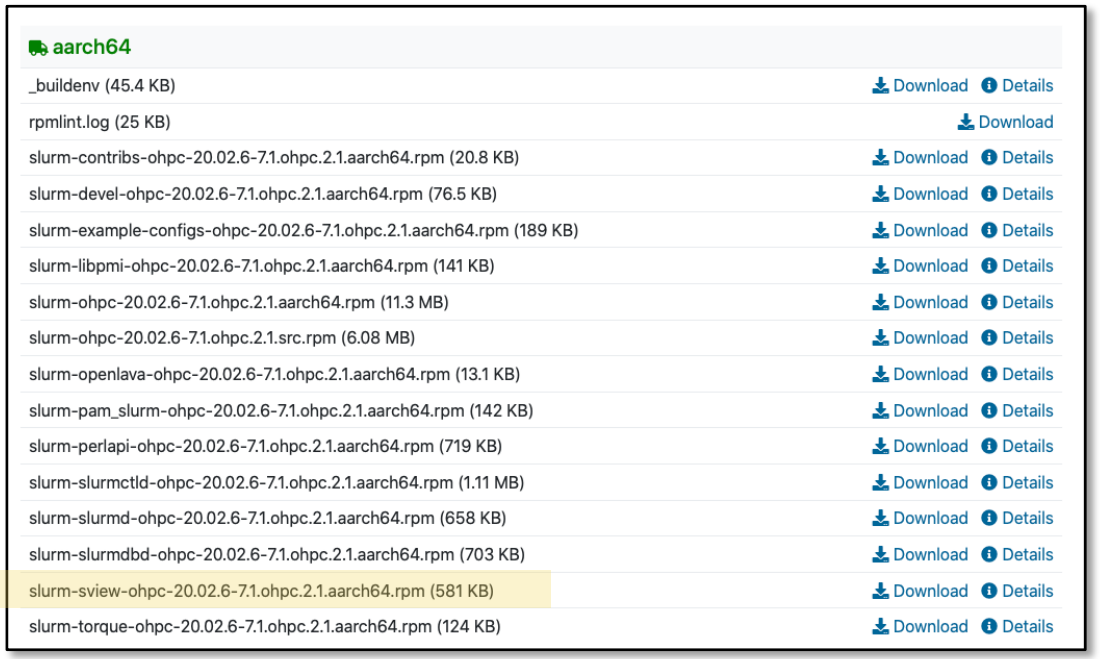
One of the changes made in the Slurm 20.11 release was to the semantics for job steps launched through the 'srun' command. This also inadvertently impacts many MPI releases that use srun underneath their own mpiexec/mpirun command.

For 20.11.{0,1,2} releases, the default behavior for srun was changed such that each step was allocated exactly what was requested by the options given to srun, and did not have access to all resources assigned to the job on the node by default. This change was equivalent to Slurm setting the '--exclusive' option by default on all job steps. Job steps desiring all resources on the node needed to explicitly request them through the new '--whole' option.

In the 20.11.3 release, we have reverted to the 20.02 and older behavior of assigning all resources on a node to the job step by default.

SLURM build update

- in an older previous v1.3 release, we tweaked the slurm build to enable building of the `sview` binary
  - X-based utility for slurm
  - requires lot's of dependencies like GTK, so we separated out into a separate RPM
- This tweak got lost at some point during a slurm upgrade;
  - user on the mailing list asked if we could build it, so resurrecting the build and separate packaging for v2.1



MVAPICH2 changes/issues

- after bumping build to latest MVAPICH2 release (v2.3.5), we started to encounter CI failures for basic job launch
  - dlopen errors

```
mv2_ibv_dlopen_init returned -1
mv2_mad_dlopen_init returned -1
mv2_umad_dlopen_init returned -1
mv2_rdma_dlopen_init returned -1
INTERNAL ERROR: invalid error code ffffffff (Ring Index
out of range) in MPIDI_CH3_Init:413
[unset]: aborting job:
Fatal error in MPI_Init:
Other MPI error, error stack:
MPIR_Init_thread(490):
MPID_Init(397)…….: channel initialization failed
MPIDI_CH3_Init(413)..:

[unset]: write_line error; fd=-1 buf=:cmd=abort
exitcode=1096207
:
```

- After poking at the problem, noticed this comment in changelog
  - part of the issue is that thru use of dlopen, rpm packaging does not see requirement for a number of required OFED packages
  - job launch can be resolved by installing packages by hand, but also need devel version (e.g. `rdma-core-devel` on CentOS, whereas `rdma-core` was sufficient previously)

•NEW Remove dependency on underlying libibverbs, libibmad, libibumad, and librdmacm libraries using dlopen

MVAPICH2 changes/issues

- a new configure option is available to disable the new dlopen approach

  `--disable-ibv-dlopen`

- Have updated our latest build to use this option to restore standard RPM .so dependencies

Requires

Symbol	Provided by
libgfortran.so.5(GFORTRAN_9)(64bit)(ohpc)	gnu9-compilers-ohpc
libibmad.so.5()(64bit)	infiniband-diags
libibmad.so.5(IBMAD_1.3)(64bit)	infiniband-diags
libibumad.so.3()(64bit)	libibumad
libibumad.so.3(IBUMAD_1.0)(64bit)	libibumad
libibverbs.so.1()(64bit)	libibverbs
libibverbs.so.1(IBVERBS_1.0)(64bit)	libibverbs
libibverbs.so.1(IBVERBS_1.1)(64bit)	libibverbs
libm.so.6()(64bit)	glibc
libm.so.6(GLIBC_2.2.5)(64bit)	glibc

MVAPICH2 changes/issues

- With the fix, still seeing a few regression test failures with latest MVAPICH2
  - Not sure yet what is going on here
- No issues detected with other MPI stacks thus far
- TODOs
  - resolve a few builds
    - updated SLURM
    - warewulf-vnfs on CentOS
  - lustre client for CentOS 8.3
  - remaining MVAPICH2 failures
  - aarch64 testing?
    - what to do without CI hardware?

CentOS8/openpbs/IB

Test Result

8 failures

1,458 tests
Took 13 min.
add description

All Failed Tests

Test Name	Duration	Age
UserLevelTests.MiniFE.[Apps/miniFE] run miniFE on multi nodes under resource manager (openpbs/gnu9/mvapich2)	0 ms	1
UserLevelTests.mpi4py.[dev-tools/py3-mpi4py] python hello world (openpbs/gnu9/mvapich2)	1.1 sec	1
UserLevelTests.FFTW.[libs/FFTW] MPI C binary runs under resource manager (openpbs/gnu9/mvapich2)	0 ms	1
UserLevelTests.Mfem.[libs/MFEM] p_laplace_perf MPI binary runs under resource manager (openpbs/gnu9/mvapich2)	0 ms	1
UserLevelTests.Mfem.[libs/MFEM] p_diffusion MPI binary runs under resource manager (openpbs/gnu9/mvapich2)	0 ms	1
UserLevelTests.PTScotch.[libs/PTScotch] dgraph_redist binary runs under resource manager (openpbs/gnu9/mvapich2)	0.98 sec	1
UserLevelTests.PTScotch.[libs/PTScotch] dgord binary runs under resource manager (openpbs/gnu9/mvapich2)	1 sec	1
UserLevelTests.Extrae.[Extrae] Parallel trace merge runs under resource manager (openpbs/gnu9/mvapich2)	1.2 sec	1

All Tests

Package	Duration	Fail	(diff)	Skip	(diff)	Pass	(diff)	Total	(diff)
InstallTests	0.54 sec	0		0		7	+7	7	+7
RootLevelTests	18 sec	0		0		36	+36	36	+36
UserLevelTests	12 min	8	+8	0		1407	+1407	1415	+1415