

ARCHER2 Upgrade: User Impact

Andy Turner, EPCC, The University of Edinburgh

a.turner@epcc.ed.ac.uk

www.archer2.ac.uk



Outline

- Impacts on users
 - Updated login node host keys
 - Recompile and test software
 - Python 2 no longer available
 - Slurm changes:
 - Slurm: `--cpus-per-task=X`
 - Slurm: "socket" definition change
 - Singularity + MPI
- CSE team experience so far

ARCHER2 Partners



THE UNIVERSITY
of EDINBURGH



**Hewlett Packard
Enterprise**

User action summary

- Remove old host keys from `.ssh/known_hosts` file
- Check versions of centrally-installed software and update submission scripts and build processes as needed
- Recompile and test software
- Be aware that Python 2 is no longer available or supported
- Update job submission scripts, add:
`export SRUN_CPUS_PER_TASK=$SLURM_CPUS_PER_TASK`
- Be aware of change in Slurm placement definition: Slurm “socket” is now 4 cores (CCX) rather than 16 cores
 - Most users unaffected
 - Only impacts you if you use a cyclic process placement
- If you use Singularity containers with MPI, update job scripts to have new bind and library path settings

Reminder: ARCHER2 publications

- Please add all DOI associated with ARCHER2 to SAFE:
<https://epcced.github.io/safe-docs/safe-for-users/#regdoi>
 - All users can do this, not just project leaders
- Provides important evidence of the impact of ARCHER2
- ARCHER2 will donate £1 to Save the Children for every publication added
- DOI associated with an ARCHER2 projects/consortia can be outputted for import into ResearchFish during the submission window
- All publications added appear at:
<https://www.archer2.ac.uk/research/publications/>
- Keen to develop ARCHER2 Case Studies from work on ARCHER2 – if you are interested get in touch: support@archer2.ac.uk

Changes post-upgrade



lepcci

Updated login node host keys

- Remove old ARCHER2 login node keys from `.ssh/known_hosts`
- If you forget to do this, you will see an error similar to →

```

@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@ WARNING: POSSIBLE DNS SPOOFING DETECTED! @
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
The ECDSA host key for login.archer2.ac.uk has changed, and
the key for the corresponding IP address 193.62.216.43 has
a different value. This could either mean that DNS SPOOFING
is happening or the IP address for the host and its host
key have changed at the same time. Offending key for IP in
/Users/auser/.ssh/known_hosts:11
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@ WARNING: REMOTE HOST IDENTIFICATION HAS CHANGED! @
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
IT IS POSSIBLE THAT SOMEONE IS DOING SOMETHING NASTY!
Someone could be eavesdropping on you right now (man-in-
the-middle attack)! It is also possible that a host key has
just been changed. The fingerprint for the ECDSA key sent
by the remote host is
SHA256:UGS+LA8I46LqnD58WiWNlaUFY3uD1WFr+V8RCG09fUg. Please
contact your system administrator.

```

Recompile and test software



- Upgrade has given ARCHER2 a new operating system version and much more recent programming environment
 - OS based on SLES sp4
 - HPE Cray PE: 22.12
- Many OS-level software libraries have changed version
- All HPE Cray PE software libraries have changed version
- All centrally-provided software from ARCHER2 CSE team at EPCC has been recompiled and tested

Python 2 no longer available

- Python 2 was deprecated in 2020
- No longer available or supported on ARCHER2

Slurm: `--cpus-per-task=X`



Change in Slurm behaviour:

- The setting from the `--cpus-per-task` option to `sbatch/salloc` is no longer propagated by default to `srun` commands in the job script
- This can lead to very poor performance due to oversubscription of cores with processes/threads if job submission scripts are not updated
- Simplest workaround: add the command:
`export SRUN_CPUS_PER_TASK=$SLURM_CPUS_PER_TASK`
before any `srun` commands in the script
 - Can also explicitly use the `--cpus-per-task` option to `srun`

Slurm: "socket" definition change

- Only affects users who use a cyclic placement scheme
- The Slurm definition of a "socket" has changed
 - Previous setting: socket = 16 cores (all share a DRAM memory controller)
- Updated setting:
 - Socket = 4 cores (corresponding to a CCX - Core Complex)
 - Each CCX shares 16 MB L3 Cache

```
Node summary for 1 nodes:
Node 0, hostname nid000005, mpi 32, omp 1,
executable xthi_mpi_mp
MPI summary: 32 ranks
Node 0, rank 0, thread 0, (affinity = 0)
Node 0, rank 1, thread 0, (affinity = 4)
Node 0, rank 2, thread 0, (affinity = 8)
Node 0, rank 3, thread 0, (affinity = 12)
Node 0, rank 4, thread 0, (affinity = 64)
Node 0, rank 5, thread 0, (affinity = 68)
Node 0, rank 6, thread 0, (affinity = 72)
Node 0, rank 7, thread 0, (affinity = 76)
```

New placement

```
Node summary for 1 nodes:
Node 0, hostname nid000005, mpi 32, omp 1, executable
xthi_mpi_mp
MPI summary: 32 ranks
Node 0, rank 0, thread 0, (affinity = 0)
Node 0, rank 1, thread 0, (affinity = 16)
Node 0, rank 2, thread 0, (affinity = 32)
Node 0, rank 3, thread 0, (affinity = 48)
Node 0, rank 4, thread 0, (affinity = 64)
Node 0, rank 5, thread 0, (affinity = 80)
Node 0, rank 6, thread 0, (affinity = 96)
Node 0, rank 7, thread 0, (affinity = 112)
```

Old placement

Singularity containers and MPI



- Due to changes in software libraries and their location, users that use Singularity with MPI need to update their container launch configuration:
 - Update the locations mounted from the host system within the containers
 - Update LD_LIBRARY_PATH to point to the correct host libraries
- Instructions on how to build containers that incorporate MPI updated
- See the documentation for the updated settings and build instructions.

<https://docs.archer2.ac.uk/user-guide/containers/#using-singularity-with-mpi-on-archer2>

CSE team experience so far



Experiences

- Caveat: not had much time on the full system
 - CSE got access to full system for final testing this morning
 - Have been able to test small scale on Test and Development System (8 nodes)
- No big issues found
- Performance (at small scale) in line with that seem pre-upgrade
- Large scale jobs seem to work on main system
 - Run up to 4096-node LAMMPS calculations (this morning!)
 - Performance still to be evaluated

Other notes



lepcc |

User action summary

- Remove old host keys from `.ssh/known_hosts` file
- Check versions of centrally-installed software and update submission scripts and build processes as needed
- Recompile and test software
- Update job submission scripts, add:
`export SRUN_CPUS_PER_TASK=$SLURM_CPUS_PER_TASK`
- Be aware of change in Slurm placement definition: Slurm “socket” is now 4 cores (CCX) rather than 16 cores
 - Most users unaffected
 - Only impacts you if you use a cyclic process placement
- If you use Singularity containers with MPI, update job scripts to have new bind and library path settings

<https://docs.archer2.ac.uk/faq/upgrade-2023/>

CSE Supported Software: Applications



Software	Versions
CASTEP	22.11, <u>23.11</u>
Code_Saturne	7.0.1
ChemShell/PyChemShell	3.7.1/21.0.3
CP2K	2023.1
FHI-aims	221103
GROMACS	2022.4
LAMMPS	17_FEB_2023
NAMD	2.14
Nektar++	5.2.0
NWChem	7.0.2 (more recent versions have issues)
ONETEP	6.9.1.0
OpenFOAM	v10.20230119 (.org), v2212 (.com)
Quantum Espresso	<u>6.8</u> , 7.1
VASP	5.4.4.pl2, 6.3.2, 6.4.1-vtst, <u>6.4.1</u>

CSE Supported Software: Libraries



Software	Versions
AOCL	3.1, <u>4.0</u>
Boost	1.81.0
Eigen	??
GSL	2.7
HYPRE	2.18.0, <u>2.25.0</u> (default)
METIS/ParMETIS	5.1.0/4.0.3
MUMPS	5.3.5, <u>5.5.1</u>
PETSc	13.14.2, <u>13.18.5</u>
PT/Scotch	6.1.0, <u>7.0.3</u>
SLEPC	13.14.1, <u>13.18.3</u>
SuperLU/SuperLU_Dist	5.2.2 / 6.4.0, <u>8.1.2</u>
Trilinos	12.18.1