Integration MVAPICH2 into RedHat Enterprise Linux

Honggang Li
August 25, 2021
In 2009, RedHat packaged MVAPICH2 for RHEL5.3. In later major releases of RHEL, MVAPICH2 always been added as RHEL specific package.
Using MVAPICH2

- Load environment module
  - `mpi/mvapich2-psm2-x86_64` or `mpi/mvapich2-x86_64`
- Compile MPI source code with mpicc/mpicxx/mpifort, and run
- Rpm packaging special require
  - Must install the binary executables under same directory of mpicc
    - [https://bugzilla.redhat.com/show_bug.cgi?id=1958960](https://bugzilla.redhat.com/show_bug.cgi?id=1958960)
  - Workaround: explicitly requires mvapich2/mvapich2-psm/mvpaich2-psm2
    - Requires: mvapich2
  - See mpitests.spec (centos 9 stream) file for example
    - [https://gitlab.com/redhat/centos-stream/rpms/mpitests/-/blob/c9s/mpitests.spec](https://gitlab.com/redhat/centos-stream/rpms/mpitests/-/blob/c9s/mpitests.spec)
## Supported mainstream RDMA hardware

### Verbs (InfiniBand, IWARP, RoCE)

<table>
<thead>
<tr>
<th></th>
<th>cxgb3</th>
<th>cxgb4</th>
<th>ocrdma</th>
<th>bnxt_re</th>
<th>ipath</th>
<th>mthca</th>
<th>mlx4</th>
<th>mlx5</th>
<th>qedr</th>
<th>qib</th>
<th>hfi1</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHEL9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RHEL8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RHEL7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RHEL6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RHEL5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Verbs based build also available for qib and hfi1/opq
2. Only one type of mlx4 hardware will be supported for rhel9
3. usnic, vma_pvrdma, soft-rxe and soft-iwarp are not supported
4. psm3 code is available in upstream Linux kernel and libfabric, but not supported
Benchmarks

- OSU benchmarks
  - osu_bw, osu_latency, ...
- Intel MPI benchmarks
  - MPI1, EXT, IO, MT, NBC, P2P, RMA
- Only run test over **homogeneous** hardware
- Only test **GNU C/C++ compiler**, no Fortran
- Package built for **x64**, aarch64, ppc64le, but only tested x64
Upstream and downstream release schedule

<table>
<thead>
<tr>
<th>Plan</th>
<th>Feature requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Devel</td>
<td>Package/code rebase</td>
</tr>
<tr>
<td>QE</td>
<td></td>
</tr>
</tbody>
</table>

RHEL release schedule

<table>
<thead>
<tr>
<th>release</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.6</td>
<td>05/11/2021</td>
</tr>
<tr>
<td>2.3.5</td>
<td>11/30/2020</td>
</tr>
<tr>
<td>2.3.4</td>
<td>06/01/2020</td>
</tr>
</tbody>
</table>

- Upstream release interval is about 5+ months
- Steady and predictable upstream release schedule is very important
No access to MVAPICH2 commits history, it’s impossible to narrow down regression issues with ‘git bisect’. mpich and openmpi provides git repo access for public.
Problem for major RHEL release

- Because MVAPICH2 is unavailable in Fedora, MVAPICH2 always had been added as RHEL only/specific package for major release like RHEL-6.0, 7.0, 8.0 and 9.0.
  - Compilation security harden
  - SeLinux harden
  - Block CI testing

- Mpich and openmpi are inherited from corresponding Fedora releases.