

HP XC How To



Using MPICH on XC

Version 1.0 October, 2005

© 2005 Hewlett-Packard Development Company, L.P.

The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Linux is a U.S. registered trademark of Linus Torvalds.

MPICH is an open source implementation of MPI, the standard for message-passing libraries.

Contents

Introduction

Procedures

Building, testing and installing MPICH	6
Allocating nodes to MPICH	7
Using the MPICH script with SLURM and LSF	7

Revision history

Revision table

Table 1 Revisions

Date	Edition	Revision
Oct 2005	V1.0	First Edition

Introduction

MPICH is an open source implementation of MPI, which is a standard for message-passing libraries. The current version of MPICH is 1.2.7, released in June, 2005. The version of MPICH tested on HP XC and described in this XC HowTo is **MPICH version 1.2.6**.

You can obtain the MPICH source from the Argonne National Laboratory web site. The download link at the time of publication is at the following URL:

<http://www-unix.mcs.anl.gov/mpi/mpich/downloads/mpich.tar.gz>.

For other locations, refer to the MPICH home Web page at the following URL:

<http://www-unix.mcs.anl.gov/mpi/mpich/>

Installation and end-user documentation is provided in the download kit. The MPICH product is not otherwise distributed or supported by HP.

Procedures

The procedures in this XC HowTo describe:

- How to build and test the MPICH kit.
- A wrapper script to allocate cluster resources.
- Using MPICH with SLURM and with LSF.

You must be familiar with the procedure for installing standard LSF, as documented in the HP XC user manuals. You must also be familiar with the use of SLURM and LSF commands.

Building, testing and installing MPICH

The build takes approximately two hours on a 900MHz XC6000 system. The following test results are expected:

- Two Fortran tests that are not 64-bit clean will fail.
- Tests using `ADIOI_Set_lock()` might fail on some server platforms. The reason for this on HP XC is unknown at this time, but there is discussion of this test failure on various Web-based MPICH discussion forums.

Use the following procedure to build and test MPICH on HP XC:

1. Identify a working file system with at least 80 MB of free disk space.
2. Following the kit instruction provided with MPICH, unpack the `gz` file into a directory named `mpich-1.2.6`.
3. Change default to the directory `mpich-1.2.6` directory.
4. Enter the following commands to build MPICH for XC:

```
% /bin/env RSHCOMMAND=ssh ./configure --prefix=/opt/mpich --with-device=ch_p4
% make
```

(If you prefer an installation directory other than `/opt/mpich`, specify it as the argument to the `--prefix` option.)

5. Use the following command to test MPICH:

```
% make testing
```

6. Use the following command to install MPICH:

```
% make install
```

Allocating cluster resources to MPICH

To avoid running MPICH jobs on nodes that are allocated to other tasks, you must configure MPICH jobs as follows:

- Jobs must request node allocation using either SLURM or LSF.
- Jobs are restricted to using only the resources that are granted.

The following examples provide a simple wrapper script to specify resource allocation and launch an MPICH task. These examples are not intended as full solutions for integrating MPICH with XC.

1. Using a text editor, cut and paste the following script to a file named `wrapper` in your scripts directory:

```
% cat wrapper
#!/bin/csh
srun csh -c 'echo `hostname`:2' | sort | uniq > machinelist
set hostname = `head -1 machinelist | awk -F: '{print $1}`
ssh $hostname /opt/mpich/bin/mpirun options... -machinefile machinelist a.out
```

2. Modify the script according to your cluster configuration and operational requirements. The script makes the following assumptions:
 - a. Each node in the cluster contains two CPUs
 - b. The current working directory is available on all nodes where an MPICH job might run.
 - c. SSH between nodes is enabled.
3. When using the script, ensure that you match the node and processor counts in the SLURM `srun` and LSF `bsub` commands below with the appropriate MPICH options in the wrapper script.

Using the MPICH script with SLURM and LSF

Use the wrapper script as follows:

- To use SLURM allocation, run an `srun` command similar to the following:

```
% srun -A <slurm_options>
% ./wrapper
% exit
```

- To use LSF allocation, run a `bsub` command similar to the following:

```
% bsub -I <lsf_options> wrapper
```