Cluster Building Challenge

Cluster Building Challenge

1. VM setup
2. Module
3. HPL & HPCG
4. MPI Benchmark
5. Distributed File System
6. What should I submit?

All these tasks need familiarity with Linux operations. For all the candidates, task 1&2 is a must. **For candidates who feel like to be in charge of the team's cluster management, please complete at least one of the remaining tasks.**

1. VM setup

Please setup at least **2** virtual machines with at least 70GB disk capacity.

OS: Latest Ubuntu Server / **CentOS** (PREFERRED) LTS version

Compilers:

- [Required] GNU Compiler Collection 5/6/7, please compile the gcc from source code with preset compiler provided by OS (usually also gcc). Don't forget to enable the language support for C, C++ and **Fortran**.
- [Optional] LLVM with Clang, choose a version you prefer and compile it from source code as mentioned above.
- [Required] Intel Parallel Studio XE 2018/2019

Libraries:

- [Required] MPIs: OpenMPI 1/2/3/4, MPICH 3

  It's both acceptable to install from package manager or compile from source code. If you select the hard way please choose the latest stable release for each major version, for example OpenMPI 1.10.7 / 2.1.6.
- [Required] CUDA Toolkit 9/10
- [Optional] OpenCV, please compile from source code and enable GPU support.

Applications:

- [Required] CMake 3
- [Required] Environment module (details below)

2. Module

Environment module is an application to control the environment variables of the current session of shell. For example, by loading the latest gcc module, directory of the latest gcc binary will be prepend to `PATH` environment variable and command `gcc` will be overrided by the latest one. In other cases, some compile-time-required libraries should be located in the `LD_LIBRARY_PATH`, which can be easily handled with module.

Task: Create some *modulefiles* to allow all non-root user load the compilers & libraries & applications mentioned above.

* Keep in mind that this is the very infrastructure for a cluster, especially for high performance computation environment which demand heavy and complex compilation tasks.

* Environment module has lots of variant implement in different languages. We strongly recommend the version provided by system package manager (usually written in *TCL*), still you can play with others like *Lmod*.

* Hint: use *env2* to convert the environment script of intel psxe.

3. HPL & HPCG

Task: Run the Intel MPI version of HPL & HPCG in intel psxe.

[Optional] Tune the parameters to gain the best performance.

4. MPI Benchmark

Run *OSU* and Intel MPI benchmark. Test the bi-direction bandwidth and communication latency of two virtual machines.

5. Distributed File System

NFS & Gluster

Task: Configure an NFS and a Gluster file system in one of the VMs (Server). Mount the file system on the other VM (Client).

[Optional] Install the compilers & libraries & applications in one of these file systems. Make it possible to load it from the client using module.

6. What should I submit?

A simple report should be submit.

For 1&2 you should leave some screenshots to show that you have already complete the task. For example, use module to switch from each MPI libraries and run `which mpirun` to prove you have already set up both module and mpi.

If you want more credit for the recruitment, or want to be the SysOp of the cluster, just leave any proof on the report for other tasks. We will judge and ask you questions during interview.