

Using K2 : The (very incomplete) getting started guide:

Access:

As always, only text ssh access through the IUAC NKN gateway is possible. First: ssh <username>@gate.iuac.res.in . Next: ssh k2 (or ssh vidya if you are used to that!)

Compiling code:

The default compiler is the Intel Composer XE 2013 suite, which provides the icc and ifort compilers. The compilers are installed in the /opt/intel/bin directory, but the path should be defined by default and just icc / ifort should work.

The default MPI is the Intel MPI, which also provides the compiler wrappers mpiicc and mpiifort. The MPI and wrapper compilers are in the directory /opt/intel/impi/4.1.3.048/intel64/bin/; again, default paths should work and the full path definition is normally not necessary.

You could change this to the GCC suite and/or OpenMPI; both are installed, but have not been tested. Must emphasize: have not yet tested the gcc/OpenMPI and gcc/IntelMPI-compiled-with-gcc combinations. Safer, and significantly faster, to stay with the Intel compilers.

Running MPI programs:

As usual, jobs should be submitted only through the scheduler. The scheduler for K2 is now SGE (not SLURM as in the previous incarnation). The usual SGE commands (qsub, qstat, qdel) should work. So should their man pages. If they don't, paths have not been defined: please source /etc/profile.

A sample qsub file is on the next page; please tweak this to suit your requirements. This works for programs that work with mpirun. For applications that use their own mpi-like transports (GAMESS, WIEN2K), please look up the application documentation to see how they work with SGE.

Each node has 16 cores (2 8-core CPUs), and users are requested to keep the <no. of cores> to a multiple of 16.

Environment variables:

Please add the following lines to the ~/.bash_profile:

```
export I_MPI_FABRICS=shm:tmi
export TMI_CONFIG=/opt/intel/impi/4.1.3.048/etc64/tmi.conf
```

These variables enable the use of Intel MPI with Infiniband and the fast Qlogic Infiniband PSM framework ("faster" than the IPoIB Infiniband usage, which can be turned on by changing I_MPI_FABRICS to shm:dapl; and certainly faster than Ethernet, which is the default if I_MPI_FABRICS is not specified, and can be turned on explicitly by I_MPI_FABRICS=shm:tcp).

The shm part enables the use of local memory transport for cores within the same node. To use IPoIB instead of PSM, I_MPI_DAPL_PROVIDER=ofa-v2-ib0 must also be exported besides I_MPI_FABRIC=shm:dapl; however, have not tested this out after the new install.

Sample qsub script:

```
#!/bin/bash
#$ -S /bin/bash
#$ -cwd
#$ -N parcas14
#$ -pe mpich 32
#$ -e $JOB_ID.err
#$ -o $JOB_ID.out
#$ -q all.q
#$ -V

# In the above, the lines starting with #$ are directives to SGE
#     parcas14 is the job name given to SGE
#     -pe asks for a SGE parallel environment; only mpich available
#     32 is number of processor cores requested; change this as you need
#     $JOB_ID is the job number returned by SGE when you qsub
#     all.q is SGE queue asked for; only one configured for now for MPI programs
#     -cwd looks for input, and puts output and err, in current working directory
#     -S specifies shell; if there is good reason not to use /bin/bash, change /etc/profile line as well

# The following line should not be needed
# If it is, a setting on a node is wrong; please let me know: Sumit
#ulimit -l unlimited

# If your binary is on a path defined in /etc/profile, leave the following line uncommented
# For instance, system-wide applications kept in /opt
source /etc/profile

# The actual command line you would run without the SGE
# parcas is the name of my compiled binary
mpirun ./parcas

# Generic command line should work
# mpirun /full/path/to/execfile < /full/path/to/inputfile > /full/path/to/outputfile

# To submit, qsub script_file (this being the script_file)
```

Statutory warning:

I know we are too civilized to do all this, but still ... Please don't use mpirun directly to try programs out. Ever. It works. It also confuses SGE, slows other users' jobs, and sometimes leaves running MPD processes behind. Please don't log in to compute nodes to run programs, or put your files away. With SGE, unlike with the earlier SLURM install, you can. Same effect. And gate.iuac.res.in is just that: a gateway system. Please don't leave your data there, or compile or run programs. Will bring the system to its knees.