MPI Jobs

MPI jobs run a single program across multiple processes, potentially spanning several nodes. Use the **Tasks** and **Tasks per node** fields to configure the MPI layout.

Fields to set

Field	sbatch equivalent	When to use
Nodes	`--nodes`	Number of nodes to allocate.
Tasks	`--ntasks`	Total MPI process count across all nodes.
Tasks per node	`--ntasks-per-node`	MPI processes per node. Use instead of Tasks when you want even distribution.
CPUs per task	`--cpus-per-task`	Threads per MPI process. Set to 1 for pure MPI; higher for hybrid MPI+OpenMP.

Example — pure MPI, 4 nodes × 32 tasks

```
module load openmpi/openmpi-5.0.7-rocky9-slurm
cd $HOME/myproject


mpirun ./simulate --config config.yaml
```

Set **Nodes** to `4`, **Tasks per node** to `32`, **CPUs per task** to `1`. Do not pass `-n` to `mpirun` — Slurm passes the process count automatically via PMI.

Example — hybrid MPI + OpenMP

```
module load openmpi/openmpi-5.0.7-rocky9-slurm
export OMP_NUM_THREADS=$SLURM_CPUS_PER_TASK
cd $HOME/myproject

mpirun ./simulate --config config.yaml
```

Set **Tasks per node** to the number of MPI ranks per node, and **CPUs per task** to the number of OpenMP threads. The product of the two should not exceed the node's core count.

Choosing a partition

MPI jobs that span nodes require a partition with inter-node connectivity. Check with HPC support if you are unsure which partition to use for multi-node jobs.

Created 2026-06-25 08:34:29 UTC by levk
Updated 2026-07-08 15:37:53 UTC by levk