Lecture 6 - MPI and Singularity

DSE 512

Drew Schmidt
2022-02-10

From Last Time

- No homework yet --- soon
- Questions about ISAAC?
- Questions about anything else?

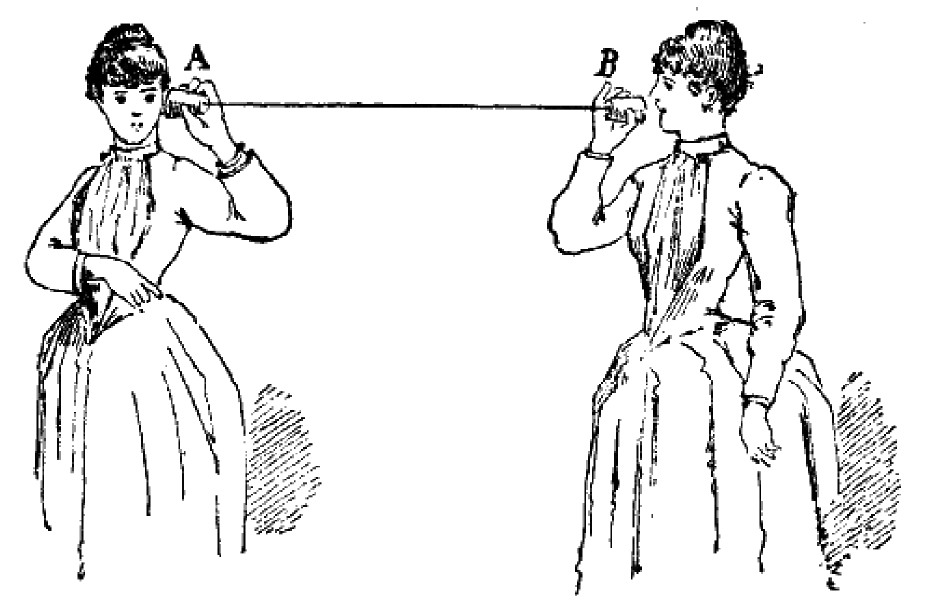
Running Multi-Node Programs

Multi-node Programming

- How do we use multiple nodes at once?
- How do we use 2 computers at the same time to solve a problem?
- Two sets of skills
  - Writing a multi-node program
  - Running a multi-node program
- Writing comes later…

MPI

- Message Passing Interface
- Distributed programming standard
- Implementations
  - OpenMPI
  - MPICH
  - MPT
  - Spectrum



MPI

- Programs pass *messages* to each other
    - Send
    - Recv
- From simple message passing, lots of abstractions
    - Reduce
    - Gather
    - MANY MORE

HLL Environments

Packages

- Python: mpi4py
- R: pbdMPI, rmpi
- Towards the end of the course
    - Distributed math/statistics
    - Deep learning

Package Installation

```
module unload PE-intel
module load gcc
module load openmpi
module load Python

pip install mpi4py

# need these later
module load cmake
pip install tensorflow
pip install torch
pip install horovod
```

Some Friendly Advice

READ THE WARNINGS

```
WARNING: The script horovodrun is installed in '/nfs/home/mschmid3/.local/bin' which is not on PATH.
Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-scr
```

Running MPI Programs

- Assume coding/compiling/linking done properly
- Launched with a special launcher
  - stock: `mpirun`
  - slurm: `srun`
  - other systems use other launchers
- ISAAC is only configured for `srun`
- *Only works in batch*

MPI Hello World

https://github.com/wrathematics/mpi-hello

- Step 1: Download on ISAAC (`git clone`, `wget`, whatever)
- Step 2: Extract the archive as necessary
- Step 3: Compile it (do `less README.md` for instructions)
- Step 4: Do a quick test with `./mpi-hello`

Job Files

See https://oit.utk.edu/hpsc/isaac-open-enclave-new-kpb/running-jobs-new-cluster-kpb/

```
#!/bin/bash
#SBATCH --account ACF-UTK0011
#SBATCH --job-name=MPI_hello_world
#SBATCH --nodes=2
#SBATCH --ntasks=4
#SBATCH --ntasks-per-node=2
#SBATCH --time=00:00:30
#SBATCH --partition=campus
#SBATCH --qos=campus

cd ~/mpi-hello
srun ./mpi-hello
```

mpi4py Hello World

```python
from mpi4py import MPI

comm = MPI.COMM_WORLD
rank = comm.Get_rank()
size = comm.Get_size()

comm_localrank = MPI.Comm.Split_type(comm, MPI.COMM_TYPE_SHARED, 0)
rank_local = comm_localrank.Get_rank()
size_local = comm_localrank.Get_size()

for p in range(0, size):
    if p == rank:
        print("Hello from rank ", end="")
        print(str(rank) + "/" + str(size) + " global ", end="")
        print(str(rank_local) + "/" + str(size_local) + " local")

    comm.Barrier()

mpi4py.MPI.Finalize()
```

pbdR Hello World

```r
suppressMessages(library(pbdMPI))

rank = comm.rank()
size = comm.size()
rank_local = comm.localrank()

hostname = system("uname -n", intern=TRUE)
hostnames = allgather(hostname) |> unlist |> table
size_local = hostnames[hostname] |> unname

msg = paste0("Hello from rank ", rank, "/", size, " global ", rank_local, " local\n")
comm.cat(msg, all.rank=TRUE, quiet=TRUE)
```

Singularity

Singularity

- Containers for HPC
- Not docker
- Pros
    - Not docker
    - Can consumes docker images
    - Containers are FILES
    - Works great with NVIDIA
- Cons
    - Not docker
    - Monolithic builds
    - Singularity Hub
    - MPI can cause trouble

Creating Singularity Images

Recommended Workflow

1. Create Docker container (see lecture 4 and docker handout)
2. Convert to Singularity image with `docker2singularity`
3. Move Singularity image to ISAAC (`sftp`, `scp`, ...)

Environments

- Your local Linux box
- An EC2 instance
- Your local non-Linux box ???

docker2singularity

```sh
#!/bin/sh

MY_CONTAINER="ubuntu:20.04"

mkdir -p /tmp/d2s
docker run \
  -v /var/run/docker.sock:/var/run/docker.sock \
  -v /tmp/d2s:/output \
  --privileged -t --rm \
  singularityware/docker2singularity ${MY_CONTAINER}
```

Live Demo