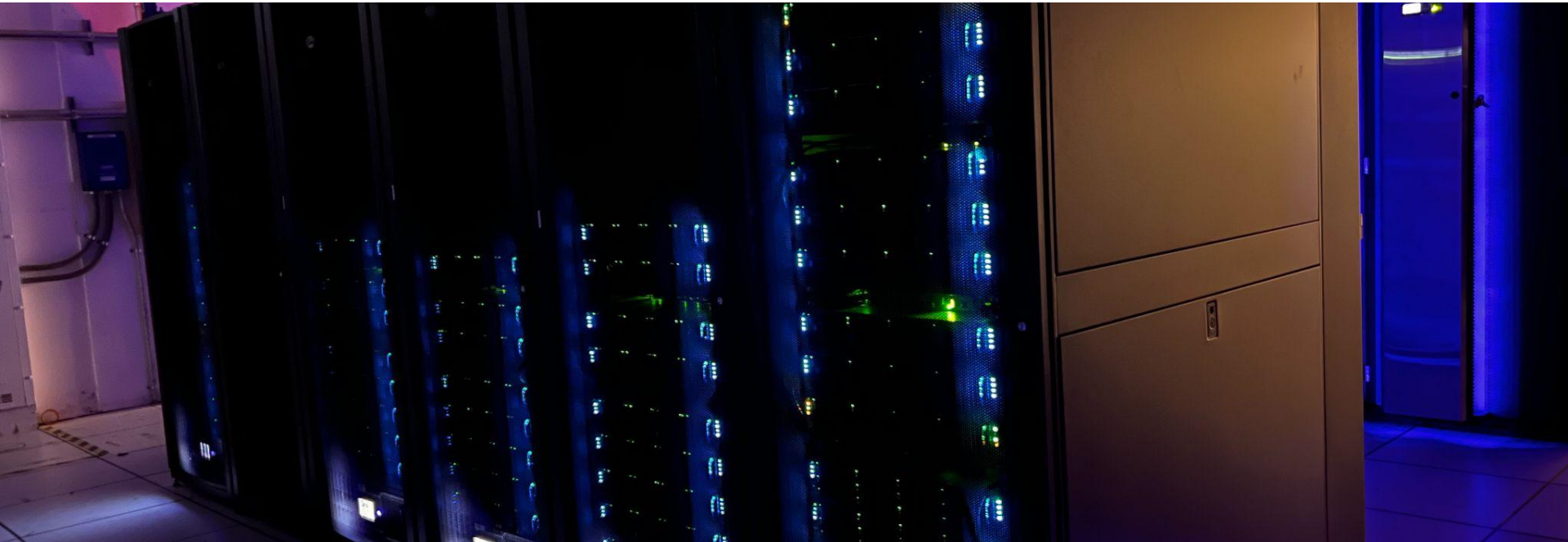
BERKELEY LAB
Bringing Science Solutions to the World

U.S. DEPARTMENT OF ENERGY
Office of Science

HPC Training

Job Submissions on Lawrencium

September 26, 2023



Agenda

Brief Overview

- General Steps: Using an MPI Example

- GNU Parallel Example
- GPU Jobs
- Python Examples
- Containerization w/ Apptainer

- Q&A

Generic Steps for Running Applications on a Cluster

Seven steps to get up and running

01

Software Packages

02

Access to data (If any)

03

Testing Interactively with SLURM

04

SLURM Batch Scripts

05

Job Monitoring

06

Completion

07

Submit more jobs

General Steps: Following an MPI Example: Hello World

01. Software Packages

Software Packages

General Steps: Following an MPI Example: Hello World

01. Software Packages

```
module avail

module avail gcc
module load gcc/12.1.0

module avail  # What's built with
              # gcc?
module avail openmpi
module load openmpi/4.1.4-gcc

module list

module remove openmpi

module purge
```

General Steps: Following an MPI Example: Hello World

01. Software Packages

```
module avail

module avail gcc
module load gcc/12.1.0

module avail  # What's built with
              # gcc?
module avail openmpi
module load openmpi/4.1.4-gcc

module list

module remove openmpi

module purge
```

General Steps: Following an MPI Example: Hello World

01. Software Packages

```
module avail

module avail gcc
module load gcc/12.1.0

module avail  # What's built with
              # gcc?
module avail openmpi
module load openmpi/4.1.4-gcc

module list

module remove openmpi

module purge
```

General Steps: Following an MPI Example: Hello World

02. Access to data

Data Management

General Steps: Following an MPI Example: Hello World

02. Access to data: If Applicable

- Please use the Dedicated Transfer Node:

```
lrc-xfer.lbl.gov
```

- Tools
  - Globus (lbnl#lrc)
  - CLI Tools like `scp` or `rsync`
  - SCP Clients like WinSCP, FileZilla, etc.

https://it.lbl.gov/resource/hpc/for-users/hpc-documentation/data-movement-and-storage/

General Steps: Following an MPI Example: Hello World

02. Access to data: Where to store data?

Storage and Backup:

Lawrencium cluster users are entitled to access the following storage systems so please get familiar with them.

NAME	LOCATION	QUOTA	BACKUP	ALLOCATION	DESCRIPTION
HOME	/global/home/users/$USER	20GB	Yes	Per User	HOME directory for permanent data storage
GROUP-SW	/global/home/groups-sw/$GROUP	200GB	Yes	Per Group	GROUP directory for software and data sharing with backup
GROUP	/global/home/groups/$GROUP	400GB	No	Per Group	GROUP directory for data sharing without backup
SCRATCH	/global/scratch/users/$USER	none	No	Per User	SCRATCH directory with Lustre high performance parallel file system

General Steps: Following an MPI Example: Hello World

02. Access to data: Special Storage

- Do you have special storage requirements?
- Please reach out to hpcshelp@lbl.gov to start a conversation regarding your storage needs
  - `/global/scratch/projects/$GROUP`
  - `/clusterfs/$GROUP`

General Steps: Following an MPI Example: Hello World

03. Testing Interactively with SLURM

Testing Interactively with SLURM

The faster it fails the better.

General Steps: Following an MPI Example: Hello World

03. Testing Interactively with SLURM: The Goal

Quick Sneak Peek:

```
salloc --account scs   --partition lr6   --qos lr_normal   --time
0:10:0   --nodes 1
```

```
[n0000 rec]$ salloc --account scs --partition lr6 --qos lr_normal --time 10 --no
des 1
salloc: Pending job allocation 64308719
salloc: job 64308719 queued and waiting for resources
salloc: job 64308719 has been allocated resources
salloc: Granted job allocation 64308719
salloc: Waiting for resource configuration
salloc: Nodes n0025.lr6 are ready for job
[elam3@n0000 rec]$
```

General Steps: Following an MPI Example: Hello World

03. Testing Interactively with SLURM: What are SLURM Associations?

Choose one below, they're all similar:

- `sacctmgr show assoc -p user=$USER`
-
```
sacctmgr -P show assoc user=$USER format="Account,Partition,QOS" \
| (sed -u 1q; sort -k1,2) \
| column -t -s\|
```
- `check_slurm_assoc`

General Steps: Following an MPI Example: Hello World

03. Testing Interactively with SLURM: What are SLURM Associations?

Three essential parts:

- Account
  - Choose an account to charge against
- Partition
  - Choose a generation of hardware to use: the older the less it costs
- QOS (Quality of Service)
  - Typically three types:
    - normal, debug, lowprio

Pick one line-item to use,

and one QOS within the line-item.

```
[n0000 rec]$ check_slurm_assoc
Account  Partition    QOS
dirac1   dirac1       normal
nano     etna_bigmem  normal
nano     etna_gpu     normal
nano     etna         normal
nano     etna-shared  normal
nano     nano1        nano_debug,normal
scs      alsacc       normal
scs      cf1          cf_debug,cf_normal
scs      cm1          cm1_debug,cm1_normal
scs      es1          es_debug,es_lowprio,es_normal
scs      lr2          lr_debug,lr_lowprio,lr_normal
scs      lr3          lr_debug,lr_lowprio,lr_normal
scs      lr4          lr_debug,lr_lowprio,lr_normal
scs      lr5          lr_debug,lr_lowprio,lr_normal
scs      lr6          lr6_lowprio,lr_debug,lr_normal
scs      lr7          lr_debug,lr_normal
scs      lr_bigmem    lr_debug,lr_normal
scs      xmas         normal
[n0000 rec]$
```

General Steps: Following an MPI Example: Hello World

03. Testing Interactively with SLURM: Choosing a Partition

Preferences:

- By Cost?
  - LR3 is free to use!
  - Cost Table
- By Hardware?
  - LR6 is newer than LR5, so on and so forth.
- By Availability?
  - Which partition has idle nodes to use?

```
PARTITION  AVAIL  TIMELIMIT  NODES  STATE
lr3           up   infinite      1  drain
lr3           up   infinite     72   idle
lr4           up   infinite      1  drain
lr5           up   infinite      2  drain
lr5           up   infinite     67   idle
lr6           up   infinite      2   comp
lr6           up   infinite      3  drain
lr6           up   infinite     95   idle
[n0000 rec]$
```

General Steps: Following an MPI Example: Hello World

03. Testing Interactively with SLURM: Choosing a Partition

Preferences:

- By Cost?
  - LR3 is free to use!
  - Cost Table
- By Hardware?
  - LR6 is newer than LR5, so on and so forth.
- By Availability?
  - Which partition has idle nodes to use?

LRC Partition Cost Table

PARTITION	NODES	SU TO CORE CPU HOUR RATIO	EFFECTIVE RECHARGE RATE
lr3	332	free	free
lr4	141	0.50	$0.005 per Core CPU Hour
lr5	192	0.75	$0.0075 per Core CPU Hour
lr6	290	1.00	$0.0100 per Core CPU Hour
lr7	60	1.00	$0.0100 per Core CPU Hour
cf1	72	0.40	$0.0040 per Core CPU Hour
lr_bigmem	2	1.50	$0.0150 per Core CPU Hour
es1	47	1.00	$0.0100 per Core CPU Hour
cm1	14	0.75	$0.00750 per Core CPU Hour
cm2	3	1.00	$0.0100 per Core CPU Hour
ood_inter	5	1.00	$0.0100 per Core CPU Hour

NOTE: The usage calculation is based on the resource that is allocated to the job instead of the actual usage of the job.

https://it.lbl.gov/resource/hpc/lawrencium/

General Steps: Following an MPI Example: Hello World

03. Testing Interactively with SLURM: Choosing a Partition

Preferences:

- By Cost?
  - LR3 is free to use!
  - Cost Table
- By Hardware?
  - LR6 is newer than LR5, so on and so forth.
- By Availability?
  - Which partition has idle nodes to use?

```
es1  CPUS=16  Feature=es1_v100,es1,c16      Gres=gpu:V100:2        RealMemory=192093M
es1  CPUS=64  Feature=es1_a40,es1           Gres=gpu:A40:4         RealMemory=515865M
es1  CPUS=64  Feature=es1_a40,es1           Gres=gpu:A40:4         RealMemory=515865M
es1  CPUS=8   Feature=es1_2080ti,es1        Gres=gpu:GRTX2080TI:4  RealMemory=191996M
es1  CPUS=8   Feature=es1_2080ti,es1        Gres=gpu:GRTX2080TI:4  RealMemory=191996M
es1  CPUS=8   Feature=es1_2080ti,es1        Gres=gpu:GRTX2080TI:4  RealMemory=95228M
es1  CPUS=8   Feature=es1_2080ti,es1        Gres=gpu:GRTX2080TI:4  RealMemory=95228M
es1  CPUS=8   Feature=es1_2080ti,es1        Gres=gpu:GRTX2080TI:4  RealMemory=96236M
es1  CPUS=8   Feature=es1_v100,es1          Gres=gpu:V100:2        RealMemory=192086M
es1  CPUS=8   Feature=es1_v100,es1          Gres=gpu:V100:2        RealMemory=192094M
es1  CPUS=8   Feature=es1_v100,es1          Gres=gpu:V100:2        RealMemory=64318M
lr3  CPUs=16  Feature=lr3_c16,lr3           RealMemory=64344M
lr3  CPUs=16  Feature=lr3_c16,lr3           RealMemory=64347M
lr3  CPUs=16  Feature=lr3_c16,lr3           RealMemory=64378M
lr3  CPUs=20  Feature=lr3_c20,lr3           RealMemory=64346M
lr3  CPUs=20  Feature=lr3_c20,lr3           RealMemory=64378M
lr4  CPUs=24  Feature=lr4                   RealMemory=56266M
lr4  CPUs=24  Feature=lr4                   RealMemory=64318M
lr4  CPUs=24  Feature=lr4                   RealMemory=64327M
lr5  CPUs=20  Feature=lr5_c20,lr5           RealMemory=128820M
lr5  CPUs=28  Feature=lr5_c28,lr5           RealMemory=64333M
lr5  CPUs=28  Feature=lr5_c28,lr5           RealMemory=64333M
lr6  CPUs=32  Feature=lr6,lr6_cas           RealMemory=95120M
lr6  CPUs=32  Feature=lr6,lr6_cas           RealMemory=95291M
lr6  CPUs=32  Feature=lr6,lr6_m192          RealMemory=192021M
lr6  CPUs=32  Feature=lr6,lr6_sky,lr6_m192  RealMemory=192000M
lr6  CPUs=32  Feature=lr6,lr6_sky,lr6_m192  RealMemory=192098M
lr6  CPUs=32  Feature=lr6,lr6_sky           RealMemory=192098M
lr6  CPUs=32  Feature=lr6,lr6_sky           RealMemory=192098M
lr6  CPUs=32  Feature=lr_bigmem             RealMemory=1546810M
lr6  CPUs=40  Feature=lr6,lr6_cas,lr6_m192  RealMemory=192029M
lr6  CPUs=40  Feature=lr6,lr6_cas           RealMemory=95290M
lr6  CPUs=40  Feature=lr6,lr6_cas           RealMemory=95290M
lr6  CPUs=40  Feature=lr6,lr6_cas           RealMemory=95297M
lr6  CPUs=40  Feature=lr6,lr6_cas           RealMemory=96205M
lr6  CPUs=40  Feature=lr6,lr6_cas           RealMemory=96205M
lr6  CPUs=40  Feature=lr6,lr6_m192,lr6_cas  RealMemory=192021M
lr6  CPUs=40  Feature=lr6,lr6_m192,lr6_cas  RealMemory=192021M
lr6  CPUs=40  Feature=lr6,lr6_m192,lr6_cas  RealMemory=192029M
lr6  CPUs=40  Feature=lr6,lr6_m192,lr6_cas  RealMemory=192029M
lr6  CPUs=40  Feature=lr6,lr6_m192,lr6_cas  RealMemory=192058M
```

https://it.lbl.gov/resource/hpc/lawrencium/

General Steps: Following an MPI Example: Hello World

03. Testing Interactively with SLURM: Choosing a Partition

Preferences:

- By Cost?
  - LR3 is free to use!
  - Cost Table
- By Hardware?
  - LR6 is newer than LR5, so on and so forth.
- By Availability?
  - Which partition has idle nodes to use?

```
sinfo
```

```
sinfo -p lr6
```

```
sinfo --state=idle -p lr3,lr4,lr5,lr6 --format="%10P %.5a %.10l %.6D %.6t"
```

```
PARTITION  AVAIL  TIMELIMIT  NODES  STATE
lr3           up   infinite      1  drain
lr3           up   infinite     72   idle
lr4           up   infinite      1  drain
lr5           up   infinite      2  drain
lr5           up   infinite     67   idle
lr6           up   infinite      2   comp
lr6           up   infinite      3  drain
lr6           up   infinite     95   idle
[n0000 rec]$
```

General Steps: Following an MPI Example: Hello World

03. Testing Interactively with SLURM: Choosing a Partition

Shared vs Exclusive Partitions:

- Exclusive Partitions
  - Most partitions are exclusive partitions, with the exception of the two partitions below
  - When you use an exclusive partition, all the resources like cpus and memory is available to you.

- Shared Partitions
  - ES1
  - LR7
  - A node can be running jobs for multiple users
  - Resources are shared

If you'd prefer not to share resources with other jobs, you can add the --exclusive flag when requesting a shared partition

e.g.

```
#SBATCH --partition lr7
#SBATCH --exclusive
```

General Steps: Following an MPI Example: Hello World

03. Testing Interactively with SLURM: Ask SLURM to Allocate a Compute Node

Putting it all together:

```
salloc --account scs --partition  lr6 --qos lr_normal --time 0:10:0 --nodes 1
```

```
[n0000 rec]$ salloc --account scs --partition lr6 --qos lr_normal --time 10 --no
des 1
salloc: Pending job allocation 64308719
salloc: job 64308719 queued and waiting for resources
salloc: job 64308719 has been allocated resources
salloc: Granted job allocation 64308719
salloc: Waiting for resource configuration
salloc: Nodes n0025.lr6 are ready for job
[elam3@n0000 rec]$
```

General Steps: Following an MPI Example: Hello World

03. Testing Interactively with SLURM: Ask SLURM to Allocate a Compute Node

Putting it all together:

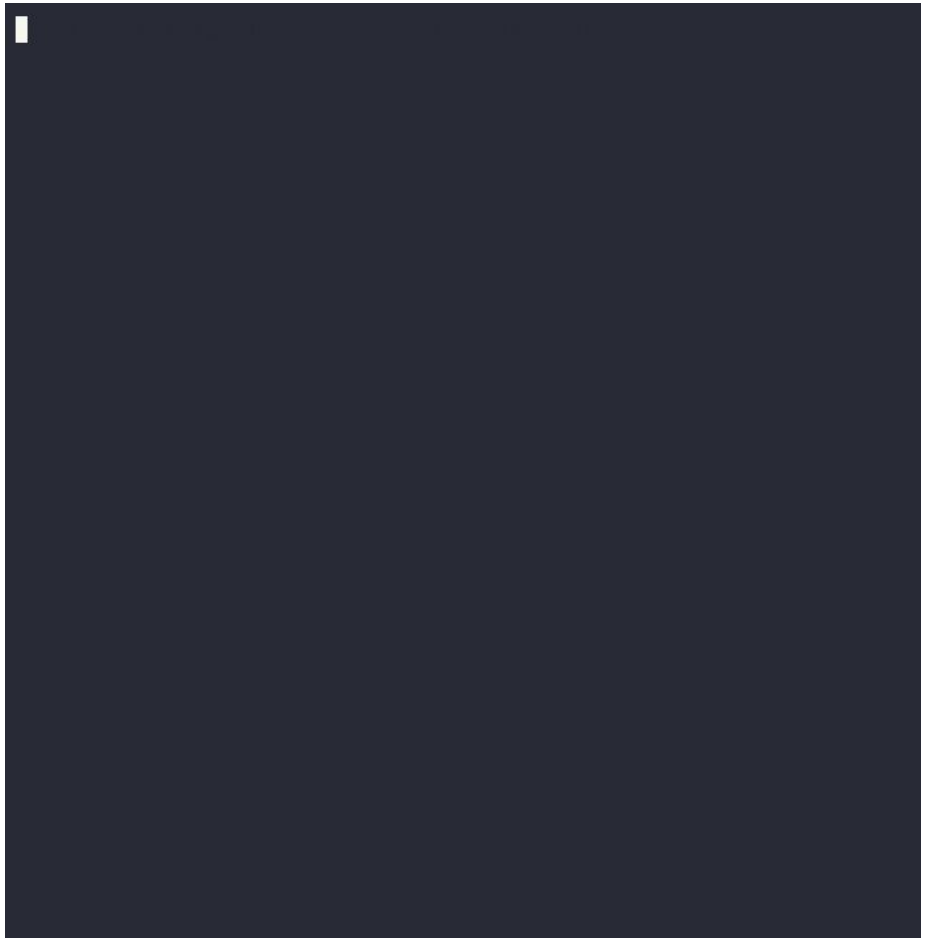
```
salloc --account scs --partition  lr6 --qos lr_normal --time 0:10:0 --nodes 1

salloc  -A       scs  -p          lr6 --qos lr_normal  -t    0:10:0   -N 1
```

```
[n0000 rec]$ salloc --account scs --partition lr6 --qos lr_normal --time 10 --no
des 1
salloc: Pending job allocation 64308719
salloc: job 64308719 queued and waiting for resources
salloc: job 64308719 has been allocated resources
salloc: Granted job allocation 64308719
salloc: Waiting for resource configuration
salloc: Nodes n0025.lr6 are ready for job
[elam3@n0000 rec]$
```

General Steps: Following an MPI Example: Hello World

03. Testing Interactively with SLURM: Take Good Notes

```
$ salloc --account scs --partition lr6 --qos
lr_normal --time 0:10:0:0 --nodes 1
$ ssh my_allocated_node
$ cd /global/scratch/users/elam3/mpi_hello_world/
$ module load gcc/12.1.0 openmpi/4.1.4-gcc
$ rm a.out
$ vim mpi_hello_world.c
$ mpicc mpi_hello_world.c
$ mpirun ./a.out
```



General Steps: Following an MPI Example: Hello World

04. SLURM Batch Scripts

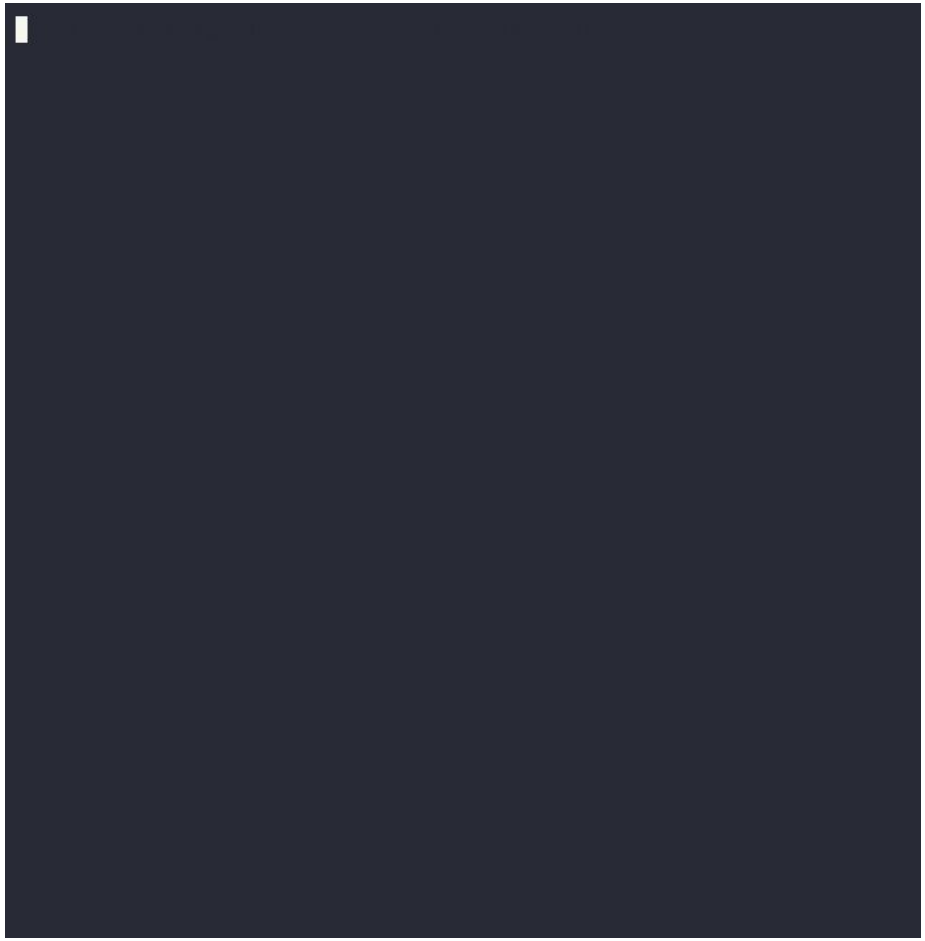
SLURM Batch Scripts

We were (almost) done before we even started.

General Steps: Following an MPI Example: Hello World

04. SLURM Batch Scripts

```
$ salloc --account scs --partition lr6 --qos
lr_normal --time 0:10:0 --nodes 1
$ ssh my_allocated_node
$ cd /global/scratch/users/elam3/mpi_hello_world/
$ module load gcc/12.1.0 openmpi/4.1.4-gcc
$ rm a.out
$ vim mpi_hello_world.c
$ mpicc mpi_hello_world.c
$ mpirun ./a.out
```



General Steps: Following an MPI Example: Hello World

04. SLURM Batch Scripts

```
salloc --account scs --partition lr6 --qos lr_normal --time 0:10:0
--nodes 1
ssh my_allocated_node
cd /global/scratch/users/elam3/mpi_hello_world/
module load gcc/12.1.0 openmpi/4.1.4-gcc
rm a.out
vim mpi_hello_world.c
mpicc mpi_hello_world.c
mpirun ./a.out
```

General Steps: Following an MPI Example: Hello World

04. SLURM Batch Scripts

```
#!/bin/bash
salloc --account scs --partition lr6 --qos lr_normal --time 0:10:0
--nodes 1
ssh my_allocated_node
cd /global/scratch/users/elam3/mpi_hello_world/
module load gcc/12.1.0 openmpi/4.1.4-gcc
rm a.out
vim mpi_hello_world.c
mpicc mpi_hello_world.c
mpirun ./a.out
```

General Steps: Following an MPI Example: Hello World

04. SLURM Batch Scripts

```
#!/bin/bash
#SBATCH --account scs --partition lr6 --qos lr_normal --time 0:10:0
--nodes 1
ssh my_allocated_node
cd /global/scratch/users/elam3/mpi_hello_world/
module load gcc/12.1.0 openmpi/4.1.4-gcc
rm a.out
vim mpi_hello_world.c
mpicc mpi_hello_world.c
mpirun ./a.out
```

General Steps: Following an MPI Example: Hello World

04. SLURM Batch Scripts

```
#!/bin/bash
#SBATCH --account scs --partition lr6 --qos lr_normal --time 0:10:0
--nodes 1
~~ssh my_allocated_node~~
cd /global/scratch/users/elam3/mpi_hello_world/
module load gcc/12.1.0 openmpi/4.1.4-gcc
rm a.out
vim mpi_hello_world.c
mpicc mpi_hello_world.c
mpirun ./a.out
```

General Steps: Following an MPI Example: Hello World

04. SLURM Batch Scripts

```
#!/bin/bash
#SBATCH --account scs --partition lr6 --qos lr_normal --time 0:10:0
--nodes 1

cd /global/scratch/users/elam3/mpi_hello_world/
module load gcc/12.1.0 openmpi/4.1.4-gcc
rm a.out
vim mpi_hello_world.c
mpicc mpi_hello_world.c
mpirun ./a.out
```

General Steps: Following an MPI Example: Hello World

04. SLURM Batch Scripts

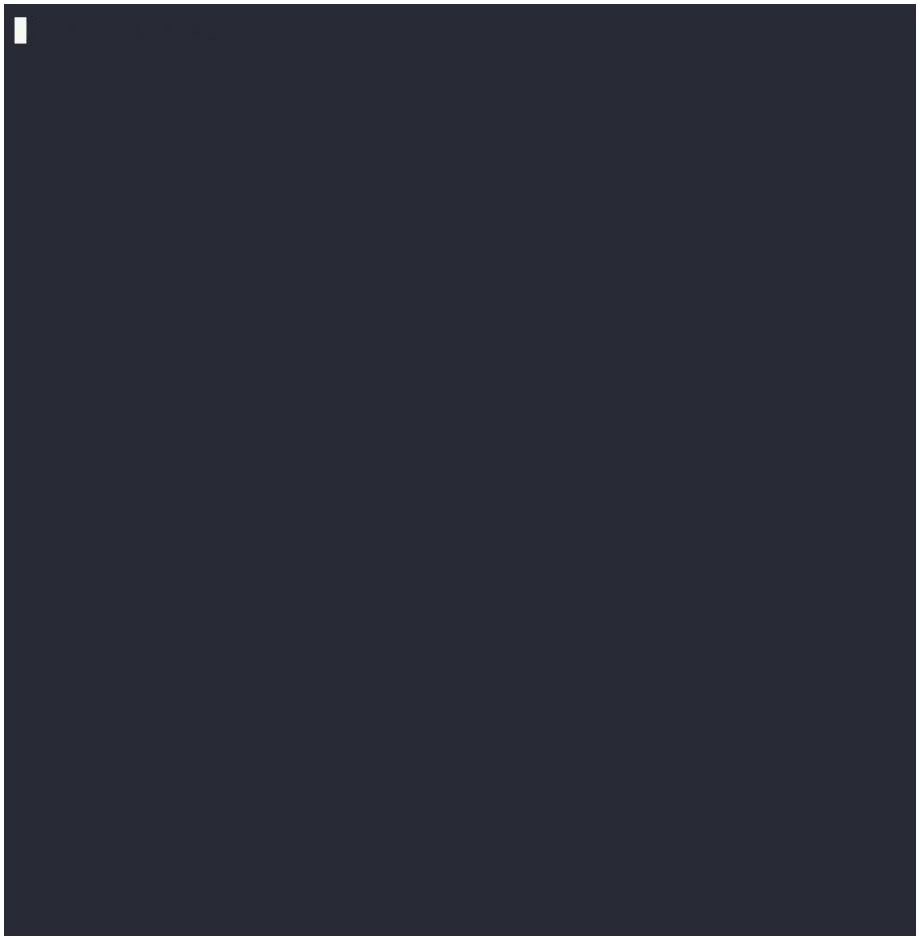
```
#!/bin/bash
#SBATCH --account scs --partition lr6 --qos lr_normal --time 0:10:0
--nodes 1

cd /global/scratch/users/elam3/mpi_hello_world/
module load gcc/12.1.0 openmpi/4.1.4-gcc



mpirun ./a.out
```

General Steps: Following an MPI Example: Hello World

04. SLURM Batch Scripts

```
#!/bin/bash
#SBATCH --account scs --partition lr6 --qos
lr_normal --time 0:10:0 --nodes 1

cd /global/scratch/users/elam3/mpi_hello_world/
module load gcc/12.1.0 openmpi/4.1.4-gcc


mpirun ./a.out
```

```
$ sbatch submit.sh
```



General Steps: Following an MPI Example: Hello World

05. Job Monitoring

Job Monitoring

Is it done yet?

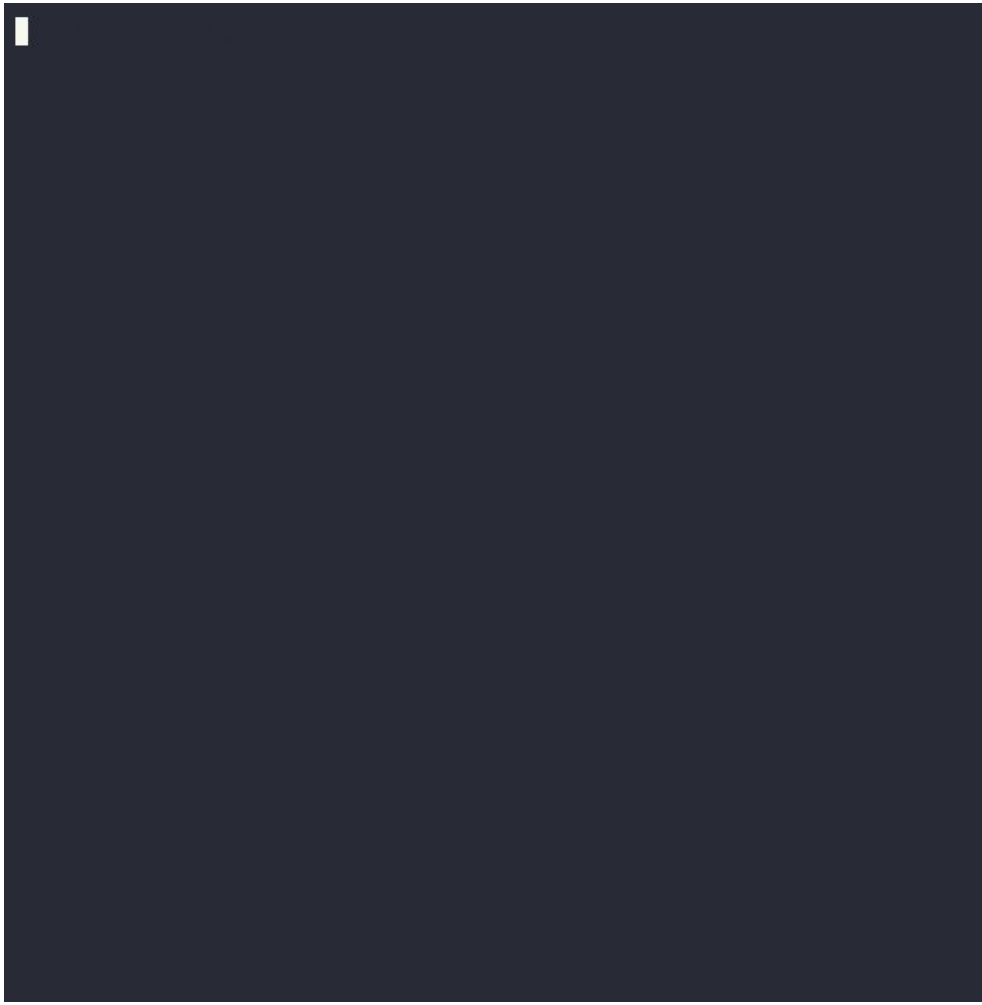
General Steps: Following an MPI Example: Hello World

05. Job Monitoring

```
$ squeue -u $USER
```

```
$ wwall -j 64259977
```

```
$ wwall -j 64259977 -t
```



General Steps: Following an MPI Example: Hello World

06. Completion

Completion

Did it run successfully to the end?

General Steps: Following an MPI Example: Hello World

06. Completion: When things go wrong

- Check the SLURM output files
- Job timed out
- Running out of memory
- Cancel a running job
- Unable to submit to queue
- Ask for help!

Typically the output files will have a filename structure like this:

slurm-64514127.out

where the number is the slurm job ID.

This file will contain both stdout and stderr.

General Steps: Following an MPI Example: Hello World

06. Completion: When things go wrong

- Check the SLURM output files
- Job timed out
- Running out of memory
- Cancel a running job
- Unable to submit to queue
- Ask for help!

Did the job run out of time?

Try increasing the --time value as needed.

--time Days-Hours:Minutes:Seconds

e.g.

--time 2-12:30:00

For 2 Days, 12 hours, and 30 minutes

General Steps: Following an MPI Example: Hello World

06. Completion: When things go wrong

- Check the SLURM output files
- Job timed out
- Running out of memory
- Cancel a running job
- Unable to submit to queue
- Ask for help!

For jobs that are running out of memory,

you may try increasing the value for

--cpus-per-task

By default, SLURM splits available memory on a node and divides it evenly to each cpu core.

So by increasing the cpus allocated to each task, you can increase the amount of available memory to each task.

e.g.
LR6 cpus=32 mem=192G is ~6G/cpu
LR6 cpus=40 mem=192G is ~4.8G/cpu

General Steps: Following an MPI Example: Hello World

06. Completion: When things go wrong

- Check the SLURM output files
- Job timed out
- Running out of memory
- Cancel a running job
- Unable to submit to queue
- Ask for help!

Get a list of jobs that you have in queue:
$ squeue -u $USER

Select a jobid from the list and run the scancel command:
$ scancel *jobid*

General Steps: Following an MPI Example: Hello World

06. Completion: When things go wrong

- Check the SLURM output files
- Job timed out
- Running out of memory
- Cancel a running job
- Unable to submit to queue
- Ask for help!

Typically, this happens when there is a scheduled maintenance coming up, and the job you are submitting is not able to finish running before the start of the maintenance period.

General Steps: Following an MPI Example: Hello World

06. Completion: When things go wrong

- Check the SLURM output files
- Job timed out
- Running out of memory
- Cancel a running job
- Unable to submit to queue
- Ask for help!

Please do not hesitate to reach out to us for help at hpcshelp@lbl.gov

GNU Parallel

GNU Parallel

Brief Description

GNU Parallel is great for tasks that can run independently without the need to communicate or coordinate with other tasks

Full GNU Parallel Tutorial

https://it.lbl.gov/resource/hpc/for-users/hpc-documentation/running-jobs/gnu-parallel/

GNU Parallel

Tedious Version

```
bash HelloApp.sh Lucy 'Pareja City'
bash HelloApp.sh Dale 'Talitay City'
bash HelloApp.sh Caroline 'Jefferson City'
bash HelloApp.sh Jerome 'Dum City'
bash HelloApp.sh Wilbur 'Cempa City'
bash HelloApp.sh Roberta 'Sykourion City'
bash HelloApp.sh Spencer 'Chiny City'
bash HelloApp.sh Hector 'Eusserthal City'
bash HelloApp.sh Bradford 'Pirogovskiy City'
bash HelloApp.sh Tara 'Pamplona City'
```

GNU Parallel Version

```
$ parallel --link -j2 bash HelloApp.sh {1} {2} :::: names.txt :::: places.txt

Hello Lucy, how is the weather in Pareja City?
Hello Dale, how is the weather in Talitay City?
Hello Caroline, how is the weather in Jefferson City?
Hello Jerome, how is the weather in Dum City?
Hello Wilbur, how is the weather in Cempa City?
Hello Roberta, how is the weather in Sykourion City?
Hello Spencer, how is the weather in Chiny City?
Hello Hector, how is the weather in Eusserthal City?
Hello Bradford, how is the weather in Pirogovskiy City?
Hello Tara, how is the weather in Pamplona City?
```

GNU Parallel

```
[n0003 hello-example]$ cat -n HelloApp.sh
     1  #!/bin/bash
     2  echo Hello ${1}, how is the weather in ${2}?
```

```
[n0003 hello-example]$ cat -n names.txt
     1  Lucy
     2  Dale
     3  Caroline
     4  Jerome
     5  Wilbur
     6  Roberta
     7  Spencer
     8  Hector
     9  Bradford
    10  Tara
```

```
[n0003 hello-example]$ cat -n places.txt
     1  Pareja City
     2  Talitay City
     3  Jefferson City
     4  Dum City
     5  Cempa City
     6  Sykourion City
     7  Chiny City
     8  Eusserthal City
     9  Pirogovskiy City
    10  Pamplona City
```

```
$ module load parallel/20200222
$ parallel --link -j2 bash HelloApp.sh {1} {2} :::: names.txt :::: places.txt

Hello Lucy, how is the weather in Pareja City?
Hello Dale, how is the weather in Talitay City?
Hello Caroline, how is the weather in Jefferson City?
Hello Jerome, how is the weather in Dum City?
Hello Wilbur, how is the weather in Cempa City?
Hello Roberta, how is the weather in Sykourion City?
Hello Spencer, how is the weather in Chiny City?
Hello Hector, how is the weather in Eusserthal City?
Hello Bradford, how is the weather in Pirogovskiy City?
Hello Tara, how is the weather in Pamplona City?
```

GNU Parallel

```
#!/bin/bash
#SBATCH --account scs --partition lr6 --qos lr_normal --time 0:10:0
#SBATCH --nodes 1
cd
/global/scratch/users/elam3/LRC101_2023_09/02-gnu-parallel/hello-example
module load parallel/20200222

parallel --link -j2 bash HelloApp.sh {1} {2} :::: names.txt :::: places.txt
```

```
[n0003 hello-example]$ sbatch submit.sh
Submitted batch job 64524809
```

```
[n0003 hello-example]$ cat slurm-64524809.out
Hello Lucy, how is the weather in Pareja City?
Hello Dale, how is the weather in Talitay City?
Hello Caroline, how is the weather in Jefferson City?
Hello Jerome, how is the weather in Dum City?
Hello Wilbur, how is the weather in Cempa City?
Hello Roberta, how is the weather in Sykourion City?
Hello Spencer, how is the weather in Chiny City?
Hello Hector, how is the weather in Eusserthal City?
Hello Bradford, how is the weather in Pirogovskiy City?
Hello Tara, how is the weather in Pamplona City?
```

GPU Jobs

Matlab

GPU Jobs

Requesting One GPU

```
#!/bin/bash
#SBATCH --job-name=mandelbrot
#SBATCH --account=account_name
#SBATCH --partition=es1
#SBATCH --qos=es_normal
#SBATCH --time=1:00:00
#SBATCH --cpus-per-task=2

#SBATCH --ntasks=1
#SBATCH --gres=gpu:1

## Command(s) to run (example):
module load matlab/r2022a
matlab -nosplash -nojvm -nodisplay -batch test_mandelbrot_cuda
```

```
[n0002 04_mandelbrot_cuda]$ sbatch submit-01-gpus.sh
Submitted batch job 64391528

[n0002 04_mandelbrot_cuda]$ cat slurm-64391528.out
GPU Device: NVIDIA GeForce RTX 2080 Ti GPU selected.
GPU Device Count: 1.
cpuTime: 7.742 s
gpuCUDAKernelTime: 0.162secs (GPU CUDAKernel) = 47.8x faster
```

```
[n0002 04_mandelbrot_cuda]$ sbatch submit-02-gpus.sh
Submitted batch job 64391529

[n0002 04_mandelbrot_cuda]$ cat slurm-64391529.out
GPU Device: NVIDIA GeForce RTX 2080 Ti GPU selected.
GPU Device Count: 2.
cpuTime: 7.742 s
gpuCUDAKernelTime: 0.149secs (GPU CUDAKernel) = 51.8x faster
```

GPU Jobs

Requesting Two GPUs

```
#!/bin/bash
#SBATCH --job-name=mandelbrot
#SBATCH --account=account_name
#SBATCH --partition=es1
#SBATCH --qos=es_normal
#SBATCH --time=1:00:00
#SBATCH --cpus-per-task=2

#SBATCH --ntasks=2
#SBATCH --gres=gpu:2

## Command(s) to run (example):
module load matlab/r2022a
matlab -nosplash -nojvm -nodisplay -batch test_mandelbrot_cuda
```

```
[n0002 04_mandelbrot_cuda]$ sbatch submit-01-gpus.sh
Submitted batch job 64391528

[n0002 04_mandelbrot_cuda]$ cat slurm-64391528.out
GPU Device: NVIDIA GeForce RTX 2080 Ti GPU selected.
GPU Device Count: 1.
cpuTime: 7.742 s
gpuCUDAKernelTime: 0.162secs (GPU CUDAKernel) = 47.8x faster
```

```
[n0002 04_mandelbrot_cuda]$ sbatch submit-02-gpus.sh
Submitted batch job 64391529

[n0002 04_mandelbrot_cuda]$ cat slurm-64391529.out
GPU Device: NVIDIA GeForce RTX 2080 Ti GPU selected.
GPU Device Count: 2.
cpuTime: 7.742 s
gpuCUDAKernelTime: 0.149secs (GPU CUDAKernel) = 51.8x faster
```

GPU Jobs

Requesting GPU Types: V100 Graphics Card (Volta Architecture)

```
#!/bin/bash
#SBATCH --job-name=mandelbrot
#SBATCH --account=account_name
#SBATCH --partition=es1
#SBATCH --qos=es_normal
#SBATCH --time=1:00:00
#SBATCH --cpus-per-task=2

#SBATCH --ntasks=1
#SBATCH --gres=gpu:1
#SBATCH --constraint=es1_v100

## Command(s) to run (example):
module load matlab/r2022a
matlab -nosplash -nojvm -nodisplay -batch test_mandelbrot_cuda
```

```
[n0003 04_mandelbrot_cuda]$ cat slurm-64408546.out
cpuTime =
   7.8002
cpuTime: 7.800 s
GPU Device: Tesla V100-SXM2-16GB GPU selected.
GPU Device Count: 1.
cpuTime: 7.800 s
gpuCUDAKernelTime: 0.181secs (GPU CUDAKernel) = 43.0x faster
```

GPU Jobs

Requesting GPU Types: A40 Graphics Card (Ampere Architecture)

```
#!/bin/bash
#SBATCH --job-name=mandelbrot
#SBATCH --account=account_name
#SBATCH --partition=es1
#SBATCH --qos=es_normal
#SBATCH --time=1:00:00
#SBATCH --cpus-per-task=2

#SBATCH --ntasks=1
#SBATCH --gres=gpu:1
#SBATCH --constraint=es1_a40

## Command(s) to run (example):
module load matlab/r2022a
matlab -nosplash -nojvm -nodisplay -batch test_mandelbrot_cuda
```

```
[n0003 04_mandelbrot_cuda]$ cat slurm-64408548.out
cpuTime =
    5.4120
cpuTime: 5.412 s
GPU Device: NVIDIA A40 GPU selected.
GPU Device Count: 1.
cpuTime: 5.412 s
gpuCUDAKernelTime: 0.112secs (GPU CUDAKernel) = 48.3x faster
```

Python Example

Package Installation Without Dependencies

Python Examples

Packages Without Dependencies

Most Common Use Case:

If there is a python package that you'd like to use, and it doesn't come bundled with dependencies, then you may consider using the --user flag when installing the package.

If the package of interest does require additional packages to be installed, then you can quickly find yourself in a messy environment as you install more packages.

Example:

module load python/3.11.4

pip install --user spython

```
[n0000 spython]$ pip install --user spython
Collecting spython
  Using cached spython-0.3.0-py3-none-any.whl (109 kB)
Installing collected packages: spython
  WARNING: The script spython is installed in
'/global/home/users/elam3/.local/bin' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this
warning, use --no-warn-script-location.
Successfully installed spython-0.3.0
```

Add to Path

```
[n0000 spython]$ export PATH=$HOME/.local/bin:$PATH
```

Add to ~/.bashrc

```
export PATH=$HOME/.local/bin:$PATH
```

Python Example

Virtual Environments

Python Examples

Virtual Environment Use Case

Pros

- Isolated environment to avoid potential conflicts with package dependencies

Cons

- Can take up a lot of storage space
- Potentially cause a delay when logging in if your ~/.bashrc file loads a conda environment by default

The example coming up will make use of a python virtual environment.

Python Examples

iPHoP: What is it?

iPHoP is a bioinformatic tool for computational prediction of host taxonomy from phage genomes.

https://bitbucket.org/srouxjgi/iphop/src/main/

Python Examples

iPHoP: Installation

```
 # Pick a version of Python
module load python/3.11.4

# Application Setup
conda create -n iphop_env python=3.8 mamba
source activate iphop_env
mamba install -c conda-forge -c bioconda iphop       # ~4GB

# Downloading test files and database
$ mkdir iphop_db
$ iphop download --db_dir iphop_db/ -dbv iPHoP_db_rw_for-test    # ~15GB
$ wget
https://bitbucket.org/srouxjgi/iphop/raw/d27b6bbdcd39a6a1cb8407c44ccbcc800d2b4f78/test/test_input_phages.fna
$ mkdir iphop_test_results
```

More Details: https://bitbucket.org/srouxjgi/iphop/src/main/

Python Examples

iPHoP: SLURM Batch Script

```
#!/bin/bash
#SBATCH --job-name=iphop
#SBATCH --account=account_name
#SBATCH --partition=lr6
#SBATCH --qos=lr_normal
#SBATCH --time=1:00:00

#SBATCH --nodes=1

## Command(s) to run (example):
cd /global/scratch/users/elam3/LRC101_2023_09/05-python-examples/iphop

module purge
module load python/3.11.4

source activate iphop_env

iphop predict --fa_file test_input_phages.fna \
  --db_dir iphop_db/Test_db_rw/  \
  --out_dir iphop_test_results/test_input_phages_iphop
```

Containerization with Apptainer

Putting a python application in a container

Containerization with Apptainer

Using a Container Registry like Docker Hub

This is the lowest hanging fruit, when the developers of the software you want to use already provides the container on a public registry.

For example, iPHoP:

```
$ apptainer build iphop.sif docker://simroux/iphop:latest
```

OR,

(Visit quay.io for image tag)

```
$ apptainer build iphop.sif
docker://quay.io/biocontainers/iphop:1.3.2--pyhdfd78af_0
```

Containerization with Apptainer

Potential Obstacles: Disk Quota

Disk Quota Exceeded

FATAL: While performing build: conveyor failed to get: initializing **source** oci:/global/home/users/elam3/.apptainer/cache/blob:8f425a96be7671ce1de1d61ca47f53baab829294e776903320ef960df3518dcb: writing blob: **sync** /global/home/users/elam3/.apptainer/cache/blob/oci-put-blob2936598077: disk quota exceeded

Overwrite the default cache and tmp directories for Apptainer

```
[n0000 iphop]$ mkdir APPTAINER_TMPDIR APPTAINER_CACHEDIR
[n0000 iphop]$ export APPTAINER_TMPDIR=$PWD/APPTAINER_TMPDIR
[n0000 iphop]$ export APPTAINER_CACHEDIR=$PWD/APPTAINER_CACHEDIR
```

Containerization with Apptainer

SLURM Batch Script: Side-By-Side View of iPHoP

Conda Virtual Environment

```
#!/bin/bash
#SBATCH --job-name=iphop
#SBATCH --account=account_name
#SBATCH --partition=lr6
#SBATCH --qos=lr_normal
#SBATCH --time=1:00:00

#SBATCH --nodes=1

## Command(s) to run (example):
cd /global/scratch/users/elam3/LRC101_2023_09/05-python-examples/iphop

module purge
module load python/3.11.4

source activate iphop_env

iphop predict --fa_file test_input_phages.fna \
  --db_dir iphop_db/Test_db_rw/  \
  --out_dir iphop_test_results/test_input_phages_iphop
```

Apptainer

```
#!/bin/bash
#SBATCH --job-name=iphop
#SBATCH --account=account_name
#SBATCH --partition=lr6
#SBATCH --qos=lr_normal
#SBATCH --time=1:00:00

#SBATCH --nodes=1

## Command(s) to run (example):
cd /global/scratch/users/elam3/LRC101_2023_09/05-python-examples/iphop

apptainer run iphop.sif \
    predict --fa_file test_input_phages.fna \
  --db_dir iphop_db/Test_db_rw/  \
  --out_dir iphop_test_results/test_input_phages_iphop
```

Containerization with Apptainer

GPU & Containerization

Containerization with Apptainer

Using GPUs with Apptainer

NVIDIA NGC Catalog

NVIDIA provides gpu containers that are available for use

https://docs.nvidia.com/ngc/index.html

https://catalog.ngc.nvidia.com/orgs/nvidia/containers/pytorch

```
[n0000 gpu]$ mkdir APPTAINER_CACHEDIR APPTAINER_TMPDIR;
[n0000 gpu]$ export APPTAINER_CACHEDIR=$PWD/APPTAINER_CACHEDIR
[n0000 gpu]$ export APPTAINER_TMPDIR=$PWD/APPTAINER_TMPDIR
[n0000 gpu]$ apptainer build pytorch-23.08.sif docker://nvcr.io/nvidia/pytorch:23.08-py3
INFO:    Starting build...
Getting image source signatures
Copying blob 3153aa388d02 done
Copying blob 9ac855545fa9 done
Copying blob e9225f7ab660 done
Copying blob 0ec682bf9971 done
Copying blob 0a77dcbd0e64 done
```

Containerization with Apptainer

Using GPUs with Apptainer

```
#!/bin/bash
#SBATCH --job-name=pytorch
#SBATCH --account=account_name
#SBATCH --partition=es1
#SBATCH --qos=es_normal
#SBATCH --time=1:00:00
#SBATCH --cpus-per-task=2

#SBATCH --ntasks=1
#SBATCH --gres=gpu:1

## Command(s) to run (example):
apptainer exec --nv pytorch-23.08.sif \
python -c 'import torch; print(torch.cuda.is_available())'
```

Questions?

Weekly Office Hours on Weds

Time: 10:30 am - Noon
Location: In-person at Bldg 50-3209,
or join virtually: https://go.lbl.gov/scienceit-officehours-zoom

Contact Us

Email: **hpcshelp@lbl.gov**

Thank You