How do I configure GPU compute jobs for the cluster?

Tell Me

If you would like to utilize the NVIDIA GPUs on the cluster for your compute job, below are some tips to help your job do so.

- Make sure you ask the scheduler for a GPU in your job request (submit script). You append the GPU request on the #PBS directive in which you ask for CPUs, for example:

  ```bash
  #PBS -l nodes=1:ppn=1:gpus=1,mem=16GB
  ```

- Unless your code has built-in GPU support (for example, Matlab), you may want to load one of the available CUDA Toolkit modules; currently we offer 3: cuda/7.5, cuda/8.0, or cuda/9.0. You can load one of the 3 available by adding a "module load..." line to your submit script. You can also issue a "module list" command to display what modules are currently loaded. The CUDA binaries (like nvcc) and libraries should now be available to your compute job:

  ```bash
  module load cuda/8.0
  module list
  ```

  ```bash
  Currently Loaded Modulefiles:
  1) pymods/2.7.5    2) perlmods/5.16.3    3) cuda/8.0
  ```

  ```bash
  nvcc --version
  nvcc: NVIDIA (R) Cuda compiler driver
  Copyright (c) 2005-2016 NVIDIA Corporation
  Built on Tue_Jan_10_13:22:03_CST_2017
  Cuda compilation tools, release 8.0, V8.0.61
  ```

- If your code depends on The NVIDIA CUDA Deep Neural Network (cuDNN) GPU-accelerated library, you must load an available cuDNN module to set up your $LD_LIBRARY_PATH. There are several cudnn modules to choose from, depending on what cudnn version "and" what CUDA Toolkit version you require. Please use the command "module avail cudnn" to see what's available.

  ```bash
  module load cudnn/6.0-cuda8
  module list
  ```

  ```bash
  Currently Loaded Modulefiles:
  1) pymods/2.7.5    2) perlmods/5.16.3    3) cuda/8.0    4) cudnn/6.0-cuda8
  ```

- If you would like to target a specific model of GPU, you can add a "feature" tag to your request. For example the following directive requests one node with one traditional computing core and one GTX-1080Ti GPU. There is also a "k80" tag for requesting one of the existing Telsa K80 GPUs. The following directive requests one node with one traditional computing core and one K80 GPU:

  ```bash
  ### If you prefer an NVIDIA Tesla GTX-1080ti, specify the "gtx1080ti" feature tag:
  #PBS -l nodes=1:ppn=1:gpus=1:gtx1080ti
  ### If you prefer an NVIDIA Tesla K80, specify the "k80" feature tag:
  #PBS -l nodes=1:ppn=1:gpus=1:k80
  ```

Related FAQs

- Where can I find Copperhead user notes?
- How do I configure GPU compute jobs for the cluster?
- Can I run parallel jobs on the cluster?
- Can I log in to Research Computing clusters from home?
- HPC Cluster Questions