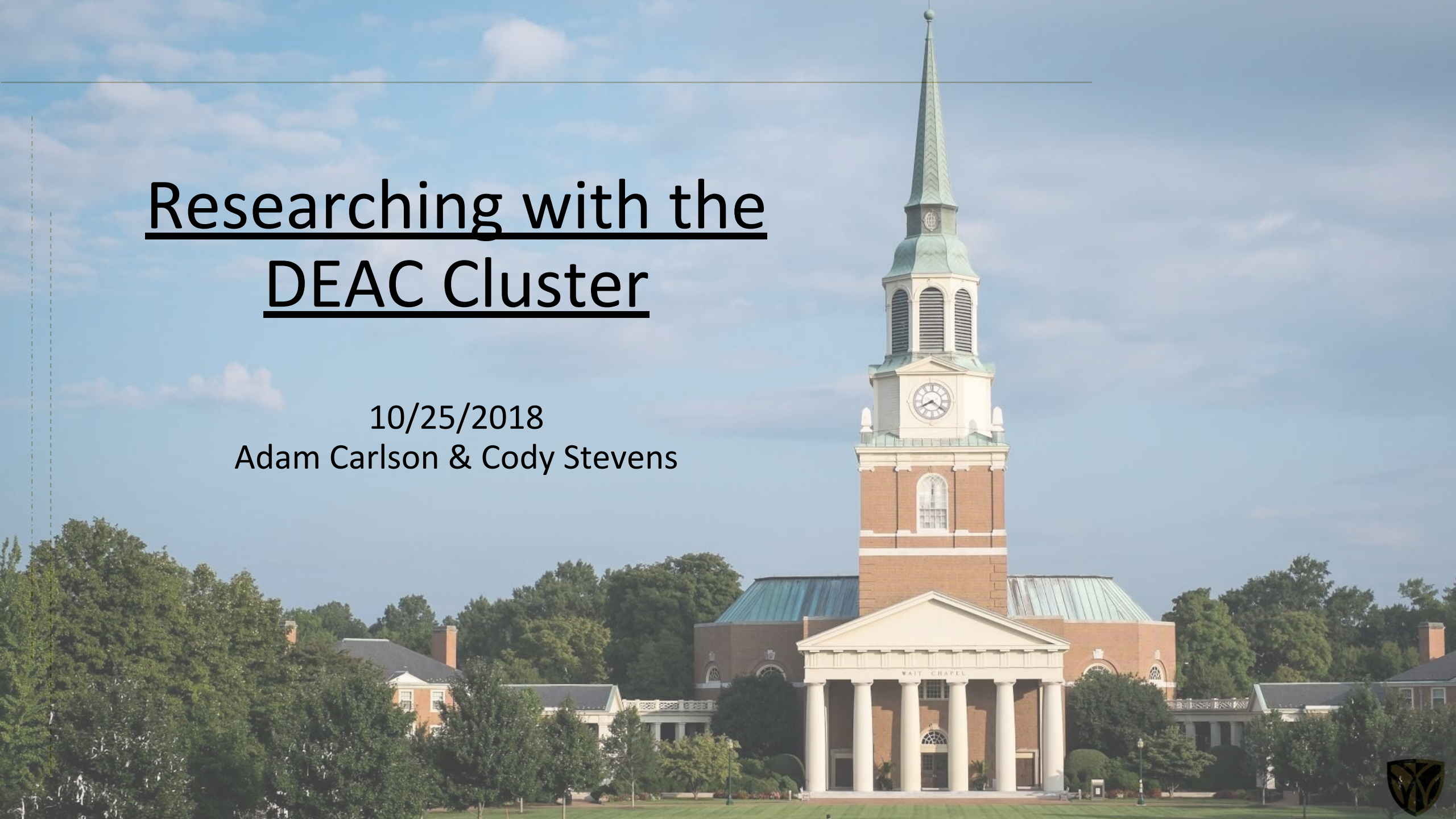
Researching with the DEAC Cluster

10/25/2018

Adam Carlson & Cody Stevens



DEAC Cluster Hardware Overview

- 14 Chassis
- 106 Cisco B-Series Blades - 3,728 cores, 19.61TB RAM
  - 27 Ivy Bridge Blades with 20 cores -- 128GB RAM
  - 24 Haswell Blades with 32 cores -- 128GB RAM
  - 43 Broadwell Blades with 44 cores -- 256GB RAM
  - 12 Skylake Blades with 44 cores -- 192GB RAM
- 2 - UCS C240 Nodes
  - 2 P100 GPU cards per node
  - 44 cores per node
  - 256GB RAM per node
- https://wiki.deac.wfu.edu/user/Cluster:Hardware_Configuration

Job Summission

```
#!/bin/bash
#SBATCH --job-name=gpu-%j
#SBATCH --output=output-%j.o
#SBATCH --error=error-%j.e
#SBATCH --mail-type=BEGIN,END,FAIL
#SBATCH --mail-user=stevca9@wfu.edu
#SBATCH --account=generalGrp
#SBATCH --partition=gpu
#SBATCH --nodes=1
#SBATCH --tasks-per-node=1
#SBATCH --mem=1gb
#SBATCH --gres=gpu:1
#SBATCH --time=0-00:30:00

cd /deac/generalGrp/usershare/stevenca/gpu/

module load cuda/9

/usr/local/cuda/samples/1_Utilities/deviceQuery/deviceQuery
```

Requesting Resources

- 4 main resources to consider when submitting a job
  - **Nodes**
    - **Determined by the amount of CPUs and Memory requested.**
  - CPU Cores
    - How many CPU cores will my job use?
  - Memory (RAM)
    - How much memory will my job consume?
  - Time
    - How long will it take for my job to complete?

```
#SBATCH --partition=gpu
#SBATCH --nodes=1
#SBATCH --tasks-per-node=1
#SBATCH --mem=1gb
#SBATCH --gres=gpu:1
#SBATCH --time=0-00:30:00
```

- Nodes and Time determine the partition.
- Best to overestimate Memory and Time when submitting.

Requesting Resources

- 4 main resources to consider when submitting a job
  - Nodes
    - Determined by the amount of CPUs and Memory requested.
  - **CPU Cores**
    - **How many CPU cores will my job use?**
  - Memory (RAM)
    - How much memory will my job consume?
  - Time
    - How long will it take for my job to complete?

```
#SBATCH --partition=gpu
#SBATCH --nodes=1
#SBATCH --tasks-per-node=1
#SBATCH --mem=1gb
#SBATCH --gres=gpu:1
#SBATCH --time=0-00:30:00
```

- Nodes and Time determine the partition.
- Best to overestimate Memory and Time when submitting.

Requesting Resources

- 4 main resources to consider when submitting a job
  - Nodes
    - Determined by the amount of CPUs and Memory requested.
  - CPU Cores
    - How many CPU cores will my job use?
  - **Memory (RAM)**
    - **How much memory will my job consume?**
  - Time
    - How long will it take for my job to complete?

```
#SBATCH --partition=gpu
#SBATCH --nodes=1
#SBATCH --tasks-per-node=1
#SBATCH --mem=1gb
#SBATCH --gres=gpu:1
#SBATCH --time=0-00:30:00
```

- Nodes and Time determine the partition.
- Best to overestimate Memory and Time when submitting.

Requesting Resources

- 4 main resources to consider when submitting a job
  - Nodes
    - Determined by the amount of CPUs and Memory requested.
  - CPU Cores
    - How many CPU cores will my job use?
  - Memory (RAM)
    - How much memory will my job consume?
  - **Time**
    - **How long will it take for my job to complete?**

```
#SBATCH --partition=gpu
#SBATCH --nodes=1
#SBATCH --tasks-per-node=1
#SBATCH --mem=1gb
#SBATCH --gres=gpu:1
#SBATCH --time=0-00:30:00
```

- Nodes and Time determine the partition.
- Best to overestimate Memory and Time when submitting.

SLURM Partitions

- 4 partitions are available on the DEAC Cluster
  - SMALL (Highest Priority)
    - Jobs must run on 1 Node
    - Jobs must run for less than 1 Day
  - MEDIUM
    - Jobs must run on 1 Chassis (8 Nodes)
    - Jobs must run for less than 7 Days
  - LARGE (Lowest Priority)
    - Jobs must run for less than 365 Days
  - GPU (Only used when requesting GPU resources)
    - Jobs must run for less than 365 Days
- Small jobs are favored over large jobs.

Job Chaining

- A way to submit jobs that may be dependent on one another
  - *sbatch --dependency=afterany:<JOB_ID>*
  - after
    - Job begins once the dependent job has started
  - afterok
    - Job begins once the dependent job has successfully terminated without error
  - afternotok
    - Job begins once the dependent job has failed
  - afterany
    - Job begins after the dependent job has terminated

- https://wiki.deac.wfu.edu/user/SLURM:Quick_Start_Guide#Job_Chaining_and_Dependencies

Useful SLURM Commands

- Get information about the cluster
  - *sinfo -Nel -p small*
  - *sinfo -p gpu*
- See which nodes are currently IDLE
  - *sinfo -Nel -p small -t idle*
- See which jobs are running
  - *squeue*
- See which of your jobs are running
  - *squeue -u <USERNAME>*
- See max memory consumption for a job
  - *slurm_mem_report* (custom DEAC script)

What's New on the DEAC Cluster?

- Partitions gpu and rhel7test
  - All software is being recompiled for RHEL7
- GCC
  - 6.2.0 (RHEL6 and RHEL7)
  - 8.2.0 (RHEL7)
- Python
  - 2.7.12 (RHEL6)
  - 3.5.2 (RHEL6)
  - 3.6.6 (RHEL7)
  - 3.7.0 (RHEL7)
- CUDA 9.0 and 10.0

Questions?