ARCC: Advanced Research Computing Center

What are the core workflow elements?

Prerequisites:
- You've watched the 'Intro to ARCC Services' and 'Intro to HPC Clusters' videos.

Workflow: Pipeline

Sequence of steps which you will work through from the start to the end of your project.

Typically a pattern you will repeat over and over.

How do you perform and implement your workflow?

The choice is yours.

Core Elements

The core elements we will introduce are:

- Getting your **data on/off the cluster**.

- **Setting up your environment** to be able to perform your research.

- Running and **submitting your jobs** on the cluster.

Getting your data on/off the cluster:

Where are you storing it?

Alcova:
- Linux/Mac platform: SMB
- Windows: file explorer
- Within a browser: Globus

On the cluster: /home, /project and/or /gscratch folders:
- Data/File Transfer Tools: **scp**, **sftp**, or **rsync**

Other Options:
- Command line: **wget**, or **git pull** to retrieve your data.

Setting up your environment:

Setting up your environment: Wiki: LMOD

```
[@tlog1 ~]$ module spider r
----------------------------
  r:
----------------------------
     Versions:
        r/3.4.4
        r/3.5.0-py27
        r/3.5.0
        r/3.5.1s
        r/3.5.2-py27
        r/3.5.2
        r/3.5.3-py27
        r/3.5.3
        r/3.6.1-intel
        r/3.6.1-py27
        r/3.6.1
        r/4.0.0-py27
        r/4.0.2-intel
        r/4.0.2-py27
        r/4.0.5-py27
```

```
[@tlog1 ~]$ module spider r/4.0.2-py27
----------------------------------------------
  r: r/4.0.2-py27
----------------------------------------------
You will need to load all module(s) on any one of
the lines below before the "r/4.0.2-py27" module is
available to load.

    swset/2018.05  gcc/7.3.0
```

```
[@tlog1 ~]$ module load gcc/7.3.0 r/4.0.2-py27
[@tlog1 ~]$ R --version
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
```

Running and submitting your job to the cluster:

1. Allocates access to appropriate computer nodes specific to your requests.

2. Framework for starting, executing, monitoring, and even canceling your jobs.

3. Queue management and job state notification.

```
[@tlog1 ~]$ sbatch run.sh
Submitted batch job 3027333

[@tlog1 ~]$ squeue -u salexan5
          JOBID PARTITION       NAME     USER ST      TIME  NODES NODELIST(REASON)
        3027333  inv-arcc        spa salexan5  R      0:01      1 mtest2
```

Wiki: Slurm Workload Manager:

Interactive Session

```
[salexan5@tlog1 ~]$ salloc --account=arcc --time=10:00
salloc: Granted job allocation 3027387

[salexan5@mtest2 ~]$ squeue -u salexan5
             JOBID PARTITION      NAME     USER ST      TIME  NODES NODELIST(REASON)
           3027387  inv-arcc interact salexan5  R       0:11      1 mtest2

[salexan5@mtest2 ~]$ exit
exit
salloc: Relinquishing job allocation 3027387

[salexan5@tlog1 ~]$
```
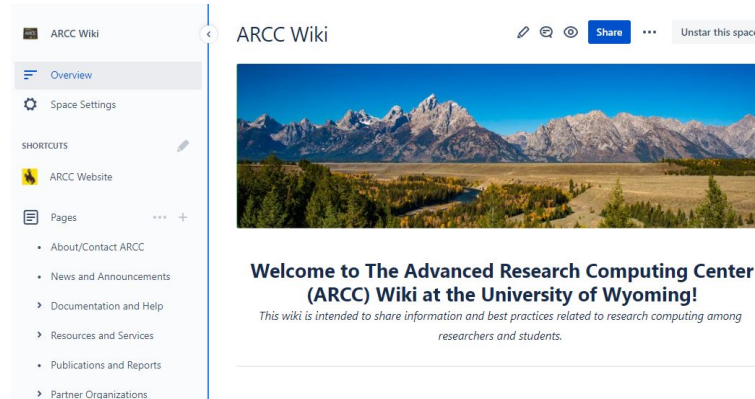
Do NOT run applications on the login nodes.

Summary

Three core elements that you will need to consider within your workflow:

- Getting your data on/off the cluster.

- Setting up your environment to be able to perform your research.

- Submitting your job to the cluster.