UTSA
Tech Solutions

Arc Onboarding Training

Zhiwei Wang, Ph.D
Research Computing Support Group
University of Texas at San Antonio

Today's Coverage:

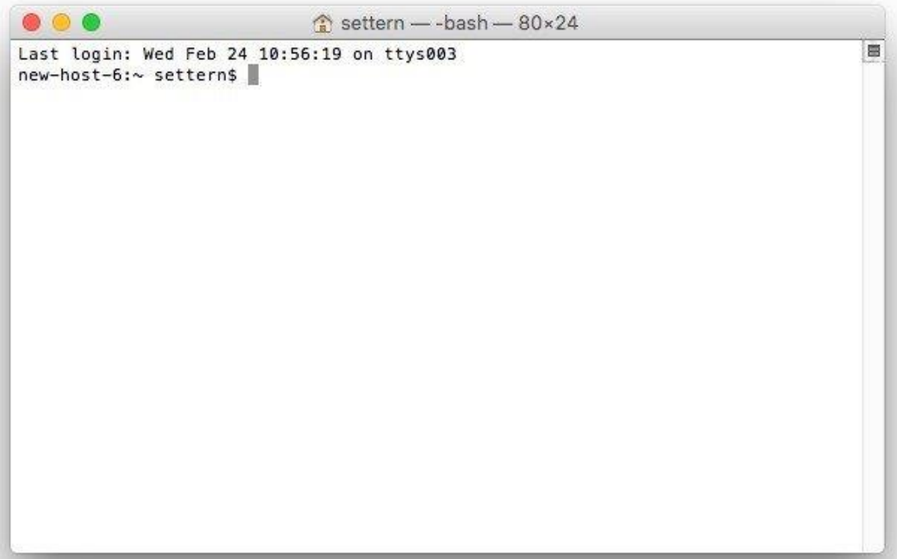
- Logging into Arc:
  - Using MobaXterm for Windows
  - SSH client for Mac and Linux
  - Web-based On-Demand portal for GUI applications
- Introduction to Arc storage
- Running jobs on Arc
  - Running applications interactively
  - Utilizing the correct parameters to access a compute node
  - Submitting batch jobs:
    - Slurm job script basics
    - Submitting and monitoring jobs
    - Receiving notifications and viewing results
- Handling long tasks

Copy The Sample Code to your home directory:
$ cp -r /work/trainings/onboarding .

Arc – the HPC Cluster at UTSA

- 167 total compute/GPU nodes and 2 login node
- 30 GPU nodes - two CPUs/40 cores, 384GB RAM, and each including one V100 Nvidia GPU
- 5 GPU nodes - two CPUs/40 cores, 384GB RAM, and each including two V100 Nvidia GPUs
- 2 GPU nodes - two CPUs/40 cores, 384 GB RAM, and 4 V100 GPUs
- 2 GPU nodes - two CPUs, one A100 80 GB GPU, and 1 TB RAM
- 2 large-memory nodes - four CPUs/80 cores, 1.5TB of RAM
- Two Lustre filesystems:
- Home: 110 TBs capacity
- Work: 1.1 PB of capacity

How to Access Arc

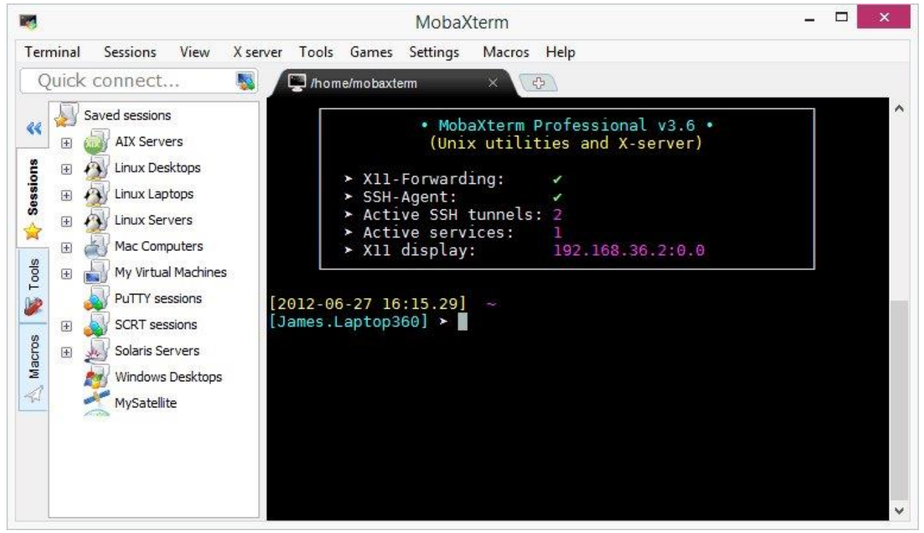


From Linux and Mac

ssh abc123@arc.utsa.edu
ssh –X abc123@arc.utsa.edu

From Windows

Install a SSH client program, such as MobaXterm.



How to access Arc

From Web Portal : portal.arc.utsa.edu

Arc Storage

- Home Directory
  - 100GB per user
  - Daily backups
  - Free up space regularly to ensure normal functioning of various applications
- Work Directory
  - No disk quota (currently)
  - No backup system
  - 30-day data purging policy
- Vault
  - 1TB per user for long-term data storage
  - Daily backups for disaster recovery
  - Accessible only from the login nodes

The Module System on Arc

What are the Environment Modules?
The Environment Modules package is a tool that lets users easily modify their environment during a session.

Typically, module loading can alter or set shell environment variables such as PATH, MANPATH, etc. Modules can be loaded and unloaded.

```
[iqr224@login-0-0 bin]$ module avai

------------------------------------------- /cm/local/modulefiles ---------------
cluster-tools/7.3 dot               gcc/6.1.0         module-git        null
cmd               freeipmi/1.5.2    ipmitool/1.8.17   module-info       openldap

------------------------------------------- /cm/shared/modulefiles --------------
abaqus/2017                          hwloc/1.11.3
abaqus/6.12                          infernal/1.1.2
acml/gcc/64/5.3.1                    intel/13/64bit
acml/gcc/fma4/5.3.1                  intel/15/ifort
acml/gcc/mp/64/5.3.1                 intel/16/64bit
acml/gcc/mp/fma4/5.3.1               intel-tbb-oss/ia32/44_20160526oss
acml/gcc-int64/64/5.3.1              intel-tbb-oss/intel64/44_20160526oss
acml/gcc-int64/fma4/5.3.1            iozone/3_434
acml/gcc-int64/mp/64/5.3.1           java/1.8.0_131
acml/gcc-int64/mp/fma4/5.3.1         lammps/20160725
acml/open64/64/5.3.1                 lapack/gcc/64/3.6.0
acml/open64/fma4/5.3.1               lapack/open64/64/3.6.0
acml/open64/mp/64/5.3.1              lis/1.5.66
acml/open64/mp/fma4/5.3.1            ls-bsr/1.0
acml/open64-int64/64/5.3.1           lsdyna/r910
acml/open64-int64/fma4/5.3.1         lsdyna7/r610
acml/open64-int64/mp/64/5.3.1        lumerical/device
acml/open64-int64/mp/fma4/5.3.1      lumerical/fdtd
adina/9.2                            mace/1.1
anaconda/4.2.0                       mace/2.1
anaconda2/4.3.1                      macs/1.4.2
apbs/1.4.2.1                         mathematica/11.0.1
aragorn/1.2.38                       matlab/R2013a
barrnap/0.6                          matlab/R2015a
bcftools/1.3.1                       matlab/R2016a
bcl2fastq2/2.17                      matlab/R2016b
bedtools/2.25.0                      matlab/R2017a
bioperl/1.007001                     mkl/10.0.3.020
blacs/openmpi/gcc/64/1.1patch03      mkl/2013.0.079
```

Module System on Arc

- list all currently loaded modules:
module list

- list all available modules:
module avail

- load a module:
module load name

- unload a module:
module unload name

Use Arc Interactively

- Use the following command to obtain a compute node
    - srun -p compute1 -n 1 -t 00:05:00 -c 40 --pty bash

“-p compute1” or “--partition=compute1” to select the partition compute1

“-n 1” or “--ntasks=1” to specify one task, which is typical for interactive sessions.

“-t 00:05:00” or “--time=00:05:00” to specify the duration of the interactive session (up to 72 hours for compute1)

“-c 1” or “--cpus-per-task=1” to request one CPU core (up to 80 for computer1)

Do not run intensive programs on login nodes!!!
Stay on the login node if you want to compile your code or submit batch jobs.

Use Arc Interactively – An Example

```
$ module load R
$ R


R version 4.4.1 (2024-06-14) -- "Race for Your Life"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu


R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.


  Natural language support but running in an English locale


R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.


Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
```

Batch Job – Job Script

```
#!/bin/bash
#SBATCH --job-name=test # Change to your job
name
##SBATCH --output=out.txt
#SBATCH --partition=compute1
#SBATCH --ntasks=1
#SBATCH --nodes=1
#SBATCH --time=1:00:00
#SBATCH --cpus-per-task=40
#SBATCH --mail-type=ALL
#SBATCH --mail-user=zhiwei.wang@utsa.edu

module load anaconda3
python hello.py
```

Batch Job – Submit a Batch Job

- Submit the job to a queue

$ sbatch jobscript

- Check the status of the job

$ squeue -u abc123

- Delete a job from the queue

$ scancel jobID

Batch Job – Slurm Commands

Show the partition info:
sinfo
Show the node status
sinfo -N -l

Show jobs status:
squeue
Show jobs status of specified user:
squeue -u abc123
Show jobs in waiting:
squeue --start
Show jobs on a specified node:
squeue --nodelist=gpu03

Show jobs in a specified partition:
squeue -p gpu1v100

Show status of a specify job:
squeue -j jobID

R - job is running on compute nodes
PD - job is waiting on compute nodes
CG - job is completing
Cancel a job:
scancel 12345.

Handling Long Tasks on Arc

- Avoid Interactive Jobs for Long Tasks
  - Optimize parallelization to reduce runtimes.
  - Splitting large tasks.
- Check resource limits and allocate appropriately
  - 72-hour runtime limit for most of the partitions on Arc
  - 10-day runtime limit for compute2 partition
  - Request the correct wall time based on previous experience.
- Use checkpointing to save progress and restart from failures.
- Monitor job status regularly and use notifications.
- Test on smaller models to catch issues early.

Useful Linux Commands

To find files in specified directory:
find /home/username/ -name "*.err”
To list the files in current directory:
ls
To delete a file :
rm filename
To show the path of the current directory:
pwd
To change directory
cd new-directory-path

Arc User-Guide Site:
https://hpcsupport.utsa.edu/foswiki/bin/view/ARC/WebHome
Create A Ticket for Technical Support:
https://www.utsa.edu/techsolutions/service-now/