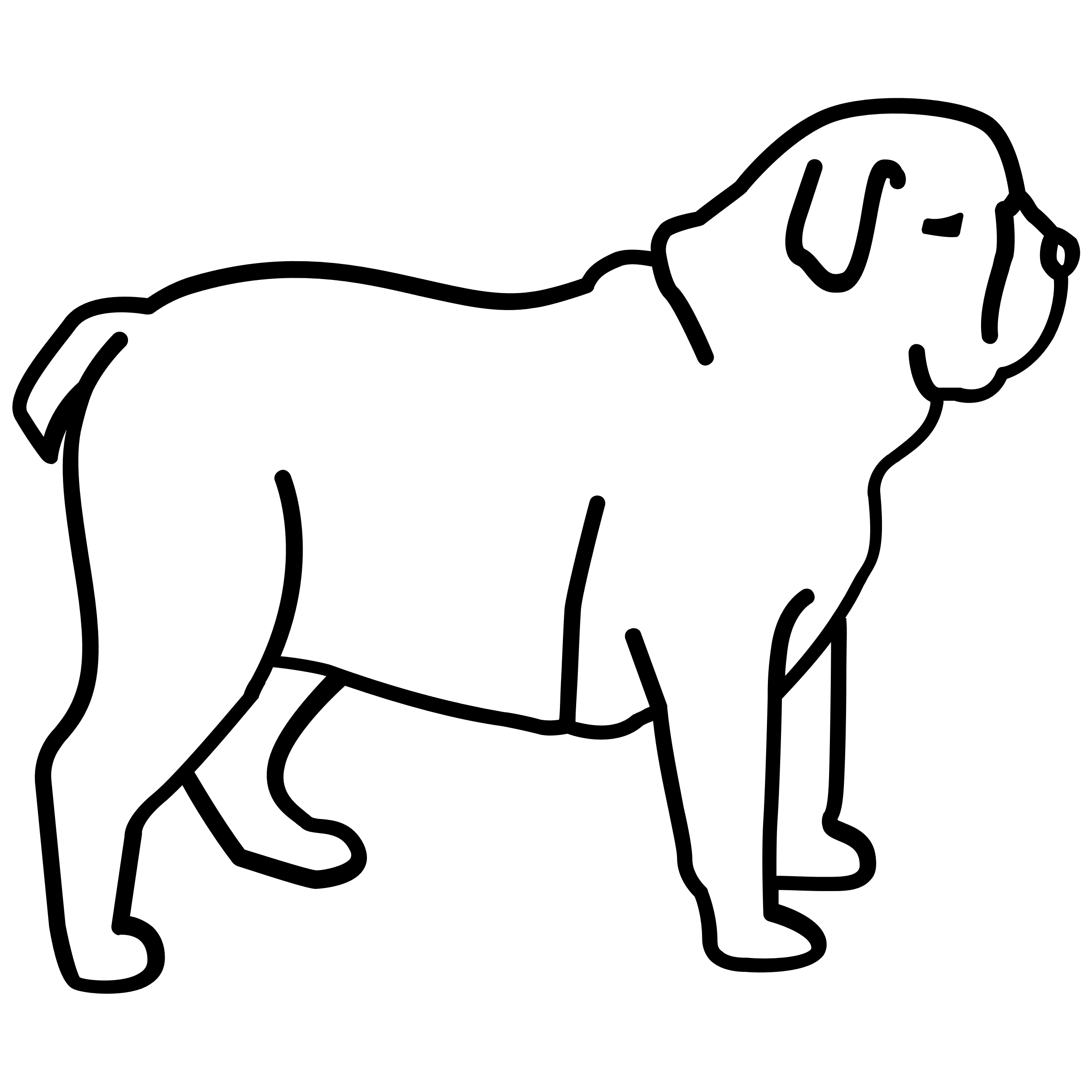
Ptolemy User’s Guide for Python Programmers

1.4



**Table of Contents**

[1. Prerequisites 3](#_Toc162340738)

[2. Gaining Access 3](#_Toc162340739)

[3. Steps for Signing Up 3](#_Toc162340740)

[4. Using Ptolemy 4](#_Toc162340741)

[4.1. Logging into Ptolemy Staging Server with SSH 5](#_Toc162340742)

[4.2. File Structure 6](#_Toc162340743)

[4.3. Modules on Ptolemy 6](#_Toc162340744)

[4.4. Creating a Python Environment 6](#_Toc162340745)

[4.5. Scheduling a Job 7](#_Toc162340746)

[4.6. Running the Job 7](#_Toc162340747)

[5. Appendix 1: slurm.sh 9](#_Toc162340748)

[6. Appendix 2: hello.py 10](#_Toc162340749)

[7. Appendix 3: Frequently Asked Questions 11](#_Toc162340750)

[7.1. How do I move files to Ptolemy? 11](#_Toc162340751)

[7.2. What if I need help? 11](#_Toc162340752)

[7.3. Who wrote this document? 11](#_Toc162340753)

# Prerequisites

To use Ptolemy you must have:

1. A Mississippi State University Net ID and password
2. A Smartphone Running the DuoMobile app. Logging into the GPU system via SSH or SFTP will require receiving and consenting to a push notification or typing in a code generated by the DuoMobile app.
3. Either a wired connection to the MSU network or a VPN connection. See this site to get a VPN client set up on your machine: <https://servicedesk.msstate.edu/TDClient/45/Portal/KB/?CategoryID=80>

# Requesting an account on Ptolemy

Access to Ptolemy is currently granted for students via class accounts, and for faculty and graduate students working on approved research projects.

* Instructors teaching classes with a need for GPU resources can request accounts for their students by sending an email to [help@hpc.msstate.edu](mailto:help@hpc.msstate.edu).
* Faculty and graduate students working on research projects should contact Michael Navicky ([navicky@hpc.msstate.edu](mailto:navicky@hpc.msstate.edu)).

# Setting up your account

Once your account has been created, you will receive an email with instructions for setting up your account. The steps are roughly as follows:

1. Click on the link in the email to fill out request and set up your temporary password.
2. Wait for a response via email.
3. A screenshot of a computer

   Description automatically generatedVisit <https://taps.hpc.msstate.edu> and login with your Net ID and temporary password. The screen should look something like this:
4. You will notice that the Training Status for “IT Security Training” and “Insider Threat Training” says “Incomplete” in red for both. Click “Start Training” on each of these courses and you will complete a training course.
5. Once your training is complete, the portal will look like this:  
     
   A screenshot of a computer

   Description automatically generated
6. You are now ready to use the system.

# Using Ptolemy

Using Ptolemy requires connecting to a login node on Ptolemy, transferring your data and code, then scheduling your job to run on the GPU cluster using SLURM, a popular HPC job scheduler.

## Connecting to the Ptolemy login node with SSH

Follow these steps:

Using the command line or a graphical SSH application, connect to the host:

ptolemy-login.arc.msstate.edu.

A screenshot of a computer error

Description automatically generatedYou will enter your username and password and then complete two-factor authentication using the DuoMobile app. The session looks like this.

## File Structure

When first logging into ptolemy-login.arc.msstate.edu you will likely be in your home directory. This directory is /home/NETID and it has a limited quota (allowed storage). You should use your home directory to store code and other tools that you will want to retain for future use.

Due to the limited space in your home directly, when utilizing Ptolemy for compute and/or data analysis jobs, you will want to keep your data in either /work/ptolemy/users/NETID/ or /scratch/ptolemy/users/NETID/.

It is a usually a good idea to create a subdirectory for each project or job you will be running. Otherwise, the working directory may become cluttered with unrelated files, or you could run the risk of subsequent jobs overwriting existing files that may have a common file name.

The scratch directories are typically used for storing temporary or intermediate files that may result from a particular data analysis. These directories should be used when running jobs that will generate a lot of intermediate files that may not all be needed later.

The work directories are typically used for more important files related to a project. This directory can be used to store files that are too large to fit in your home directory but may be necessary to recreate a specific data run or to perform further analysis.

**NOTE: Files placed under** /scratch/Ptolemy/users **are not backed up, and they may be wiped if they have been on the system for more than a couple of months without being accessed. To prevent data loss, periodically copy important files back into your** /home/ **or** /work/ **directories.**

Try not to copy non-essential files or files that are not necessary for subsequent data analysis from scratch to work or home. These files could potentially be regenerated by running the job over again and would otherwise simply take up valuable storage space under /home/ or /work/.

## Modules on Ptolemy

The server ptolemy-login-1.arc.msstate.edu uses a piece of software called “ml” – module loader – to load modules that will enable you to run analysis jobs on the machine.

For a python programmer, the key modules you will load are Python itself, CUDA (to enable the use of GPUs), and Miniconda (to enable loading additional Python packages). You can do this by hand in an interactive shell (e.g., “ml python/3.9.15” will load Python) but you will eventually want to capture all the modules you will be loading in a SLURM shell script. Feel free to load modules and play around.

You can see all available modules by typing “ml avail”.

## Creating a Python Environment

We recommend interactively creating a conda environment that you can later load in your SLURM script to run your Python code. Follow these steps:

1. First, load the Python and Miniconda modules like this:  
     
   ml python/3.9.15  
   ml miniconda3/4.10.3
2. If CUDA is required for your job, it can be setup with the following command:

ml cuda

If the program fails, then you may also need to run the following:

ml contrib cuda-libs

1. Because of disk quotas, it is important that your conda packages get installed in the scratch drive. Configure conda to use a dir in your scratch drive this way:  
     
   conda config --add pkgs\_dirs /work/ptolemy/users/netid/packages
2. Then, create a Conda environment file and save it as environment.yml. For example:

|  |
| --- |
| channels:  - anaconda  dependencies:  - numpy~=1.19.2  - pandas==1.4.1  - tensorflow-gpu |

This environment file would load numpy, pandas, and tensorflow to support a Python script capable of building a neural network. Unfortunately, this collection of packages is not the latest Tensorflow, but it will get you started.

1. You can now see if you have any miniconda environments already created by typing:  
     
   conda env list
2. Then, to create the environment that you described in your environment.yml file, run the following. Note that you can change “myname” to a name of your choice.  
     
   conda env create --name myname -f environment.yml --force
3. When conda is finished creating your environment, run:  
     
   conda env list  
     
   to see the successfully created environment now available.  
     
   Note that any modules you load only last as long as your session. Thus, if you want to interrupt your work, no problem, just go back to step 2, load Python and Miniconda, and your environments will still be there waiting for you (conda env list) because they are persisted in your directories.
4. Sidenote: You can manually remove an environment (for example, if you run out of space) by executing:  
     
   conda env remove --name myname

## Scheduling a Job

Now that you have created a conda environment, it’s time to schedule the running of your job. The job will be triggered by a scheduler named “SLURM”, and you will control it with a script.

For purposes of illustration, we will schedule the running of a Python program (hello.py) that will train a neural network to recognize handwritten digits in the mnist data set.

Our script, saved as slurm.sh, is in Appendix 1 below. Lines you will want to customize are highlighted in light grey. Our python program, saved as hello.py is in Appendix 2 below. The program depends upon data in the form of a file, mnist.npz. You can download all these files necessary to try out Ptolemy here:

<https://github.com/docbarlow/ptolemy_demo>

NOTE: Compute nodes (nodes allocated by SLURM) cannot see the internet. If your job requires external data to be downloaded, then this should be done from the login or data-transfer nodes prior to submitting your SLURM script.

## Running the Job

Now that your slurm.sh script is created, you can ask SLURM to run your job. Using the following command:

sbatch ./slurm.sh

You can monitor the execution of your jobs by using:

squeue -u jhb11

And, armed with the job id, you can cancel a job using:

scancel <job\_id>

# Appendix 1: slurm.sh

#!/bin/bash

## This line requests that the job be allocated one node

#SBATCH -N 1

## This line requests one task for the job

#SBATCH -n 1

## This specifies the amount of memory required for the job.

## Here, it's set to 10 Gigabytes. This memory is allocated per node.

#SBATCH --mem=10G

## This sets the partition (or queue) to gpu-a100

## Partitions are defined by the system administrators and

## determine the set of nodes the job can run on.

#SBATCH -p gpu-a100

## This specifies the account for job charging or accounting.

## Here, it's set to an account named 'test'. To find out

## which account you should use, run “sacctmgr show association where user=jhb11” and it

## will show the various accounts available to you. If you get an empty list, be sure

## to email [help@hpc.msstate.edu](mailto:help@hpc.msstate.edu) to request the association of your account with an

## appropriate “billing account”

#SBATCH -A test

## This sets the time limit for the job.

## Here, the job can run for up to 1 hour.

## If the job exceeds this limit, SLURM will terminate it.

#SBATCH -t 1:00:00

## This requests a specific generic resource (GRES), in this case,

## a GPU. Specifically, it requests one Nvidia A100 GPU with 10 GB of memory.

#SBATCH --gres=gpu:a100\_1g.10gb:1

## This sets the name of the job to "jhb11-GPU-mnist".

## This name will appear in the queue and in various SLURM reports,

## helping you to identify your job.

#SBATCH --job-name "jhb11-GPU-mnist"

## This specifies that all the output from the job (stdout) will

## be written to the file output.out. This file will be created in the

## directory from which the SLURM script is submitted.

#SBATCH --output=output.out

## each of these lines uses the "module load" command

## to load the modules our python script will need to use

ml cuda

ml python/3.9.15

ml miniconda3/4.10.3

## change directories to your user scratch location where your

## script is

cd /scratch/ptolemy/users/jhb11/hello\_world

## now create your conda environment – NOTE: This will not work from the compute nodes since

## since it requires internet access to download new packages. The conda environment must be

## setup from the login node prior to submitting this script.

## conda env create --file ./environment.yml --name jhb11-GPU-mnist --force

## activate conda environment

source activate jhb11-GPU-mnist

## Set a path you can use inside your python script to find your data files

export DATA\_FILE\_PATH="/scratch/ptolemy/users/jhb11/hello\_world"

## change directories to your user scratch location where your script is

cd /scratch/ptolemy/users/jhb11/hello\_world

## run the script

python hello\_world.py

# Appendix 2: hello.py

import numpy as np

import pandas as pd

import random

import tensorflow as tf

import os

from tensorflow.keras.models import Sequential, model\_from\_json

from tensorflow.keras.layers import Flatten, Conv2D, Dense, MaxPooling2D

from tensorflow.keras.optimizers import SGD

from tensorflow.keras.utils import to\_categorical

from tensorflow.keras.datasets import mnist

# see if we have multiple gpus

print("Num GPUs Available: ", len(tf.config.experimental.list\_physical\_devices('GPU')))

# this is convenient – your job will fail if the tensor flow version

# you’re using isn’t built to use gpu / cuda

if not tf.test.is\_built\_with\_cuda():

print("Not built with cuda… exiting")

exit(1)

# set path to mnist file

# path comes from environment variable set in our slurm script

file\_path = os.getenv('DATA\_FILE\_PATH') + "/mnist.npz"

# load the mnist data

(x\_train, y\_train), (x\_test, y\_test) = mnist.load\_data(file\_path)

# reshape to set up grayscale color channel for each 28x28 image

x\_train = x\_train.reshape((x\_train.shape[0], 28, 28, 1))

# normalize color between 0 and 1

x\_train = x\_train.astype ('float32') / 255.0

# create label for it

y\_train = to\_categorical (y\_train)

# setup the model

model = Sequential([

# Input shape is 28x28x1 like our images

# There will be 32 convolutions (little parts of the image) that are 3x3

# The activation function will be RELU; we talked about it in class

Conv2D(32, (3, 3), activation='relu', strides=1, input\_shape=(28, 28,1)),

# The next layer should pool together all the stuff we learned in the previous

# layer - (TODO: explain pooling layers and why they are helpful)

MaxPooling2D((2, 2)),

# Get a flat tensor - a continuous list of values

Flatten(),

# TODO: explain this one

Dense(100, activation='relu'),

# This is our last or "output layer". Notice it has 10 values like our labels (0-9).

# Softmax gives us a probability distribution so we can see which digit is most likely

#

Dense(10, activation='softmax')

])

optimizer = SGD(learning\_rate=0.01, momentum=0.9)

# compile the model

model.compile(

optimizer=optimizer,

loss='categorical\_crossentropy',

metrics=[ 'accuracy' ]

)

# fit the model

history = model.fit(x\_train, y\_train, epochs=10, batch\_size=32)

# serialize model to JSON

model\_json = model.to\_json()

with open("model1.json", "w") as json\_file:

json\_file.write(model\_json)

# serialize weights to HDF5

model.save\_weights("model1.h5")

print("Saved model to disk")

# Reshape all the test data

x\_test = x\_test.reshape((x\_test.shape[0], 28, 28, 1))

x\_test = x\_test.astype ('float32') / 255.0

y\_test = to\_categorical(y\_test)

# Now let's test the model

score = model.evaluate(x\_test, y\_test, verbose=1 )

print("%s: %.2f%%" % (model.metrics\_names[1], score[1]\*100))

# Appendix 3: Frequently Asked Questions

## How do I move files to Ptolemy?

Data transfer to and from Ptolemy must be conducted using a secure (encrypted) file transfer program. Download a secure file transfer client for Mac or Windows (from a reputable source) that supports encrypted file transfers using sftp, scp, or rsync protocols.

If you are more comfortable on the command line, then you can use scp, or rsync:

scp -r /path-to-local-files/ [netid@ptolemy-login-1.arc.msstate.edu:/path-to-remote-files/](mailto:netid@ptolemy-login-1.arc.msstate.edu:/path-to-remote-files/)

rsync -avE /path-to-local-files/ netid@ptolemy-login-1.arc.msstate.edu:/path-to-remote-files/

Or, if you prefer a more interactive (but still text-based) tool, you can use sftp.

sftp netid@ptolemy-login-1.arc.msstate.edu

NOTE: Ptolemy cannot connect to the Internet. If your analysis job depends upon data that needs to be retrieved from the Internet or your local machine, then you should download/transfer the data into your /scratch/ptolemy or /work/ptolemy directory prior to submitting the SLURM script.

## What if I need help?

If you need technical support, please email Anup Zope: [adz8@hpc.msstate.edu](mailto:adz8@hpc.msstate.edu) or Eric Collins: [emc@hpc.msstate.edu](mailto:emc@hpc.msstate.edu).

## Who wrote this document?

The initial draft was written by Jonathan Barlow: [barlow@datascience.msstate.edu](mailto:barlow@datascience.msstate.edu) to help MSU users begin using the new cluster. Feel free to send any corrections or suggestions his way. Additional edits by Eric Collins: [emc@hpc.msstate.edu](mailto:emc@hpc.msstate.edu) (v1.4).