

CyberTraining: Big Data + HPC + Atmospheric Physics
Homework 0 — Initial Parallel Programming
Spring 2019 — Matthias K. Gobbert

This homework should be worked on by *each individual participant*. Submit as a PDF file to the *Assignments* area in the Blackboard course management system by **Thursday, January 24, 2019**, which is one day before our meeting for Module 0. Remember to list your name on top of the report.

Some general notes: Start each problem on a new page. Each programming problem needs to start out with text explaining what you did; computer code, tables, or plots should *never* be the first page of any problem! This text should verbally introduce, explain, and interpret all other material including tables, plots, and computer code.

1. Access the Blackboard site for our training, labeled IS698/MATH700/PHYS650 Special Topics in Information Systems Spring 2019, and familiarize yourself with its structure. Material will be all under the link for *Course Materials*; this area contains an item with useful links (for instance, to Piazza) and two main folders named “Module summaries and associated postings” and “Contact information and preparatory material on Linux, C, Python, etc.” The *Assignments* area is used to submit homeworks.

Read information that is available in Blackboard. In particular, read material posted with Module 0 (and get a head-start on reading for Module 1, namely by at least reading the assignment of Homework 1 before we meet), and work through the preparatory material by following links to YouTube videos and other material on Linux, C, etc.

We will be using Piazza as a Discussion Board for communications among all participants, TAs, and instructors. The link to the course in Piazza is posted in an item in Blackboard. In particular, this is for posting questions by teams on homework! The idea is that questions in the Discussion Board are just like questions in a face-to-face class, namely for all others to hear and help, and for instructors and TAs to address in public.

- (a) Report on what you find and any problems you encounter. If problems such as broken links or similar, report immediately (by e-mail to instructors), so they can be fixed!
 - (b) Introduce yourself to all of us (participants, instructors, and TAs) by posting to the Discussion Board. Tell everybody about your background and explain what your goals for participating in this training are.
2. (a) Go to the webpage for the cluster at hpcf.umbc.edu that we will be using for this class. You should have received information about your UMBC account by e-mail with instructions how to create a username. UMBC only assigned a Campus ID automatically, which has the form AB12345, but for a Linux computer, you need a username of 8 alphanumeric characters, which you need to get to choose yourself following these instructions; I suggest to base your choice on your name (i.e., do not use the Campus ID as username). To use your HPCF account, you

must fill out and submit the Account Request Form under the Forms tab. Please submit this form as soon as possible; put CyberTraining into fields for title and abstract, and list Dr. Jianwu Wang (e-mail jianwu@umbc.edu, no other contact details needed) as sponsor.

Start reading information under the Resources tab very carefully. I realize that there is more material than possible to understand right now; one point is to learn how to determine the most relevant material. Try out your own account, following some information in “Using your account” page as well as following some of the YouTube videos on Linux. Report what you read and what you understood, and what questions remain.

- (b) Are you familiar with the `make` utility? Are you familiar with the `tar` and `gzip` utilities? Are you familiar with `ssh` and `scp`? How do you plan to access the HPCF cluster? Do you intend to use any other parallel computing systems? What type of systems are they (hardware, operating system, MPI version, etc.)?
3. Download the sample codes from the HPCF webpage in the “How to Compile” and the “How to Run” Tutorials into your own account on the machine. Experiment with the codes, for instance, by changing the number of nodes and the number of processes per node (how to change? How large can you make them?), and briefly report on your various experiments. The point of this assignment is to ensure that everything works properly for you, with respect to environment variables, paths, etc. If you find things not working, contact me as soon as possible, so that we will have a functioning setup for everyone as soon as possible!

What to submit: Write a report that describes what you did and why. For instance, which files did you download, how did you compile them, how did you run them? How did you change which files? Submit some output copied-and-pasted from the Linux command line and/or the stdout file `slurm.out` so that another user could reproduce your results exactly, and discuss if everything worked correctly. *Please realize the purposes of this problem: On the one hand, you should test your computer account. But on the other hand, this is about practicing how to put a brief, but meaningful report quickly that includes some original documentation, but not more than sensible. The standard for details and information provided is that others should be able to reproduce your results exactly.*