Introduction to Scientific Computing (BSR1015) Syllabus

1 Credit (15 hours), Graded, Fall 2015
Three hours/week for five weeks
Course Director: Patricia Kovatch
Course Instructor: Anthony Costa, PhD
Large classroom in the Levy Library

Course Description

This course provides an introduction to Sinai's scientific computing environment to enable effective use of our computational and data resources including the Minerva supercomputer. Assuming no prior computing experience, the course will introduce modern compute environments including UNIX fundamentals, the architectures of high performance computing, queuing systems, shell scripting, and programming in Python. Students will gain significant understanding and experience with Sinai's Minerva supercomputer. Emphasis will be placed on real-world practicality by motivating study with examples and tasks relevant to bioinformatics and structural biology. The student will develop both a solid conceptual foundation and experience solving real problems by the end of the class.

Course Policies

Four homeworks and one final exam will be given, on each of the first day of class in a given week. Homework is due the following week on the first day of class, unless otherwise indicated in the syllabus. Letter grades will be given. Late homework will be accepted with a penalty but not beyond the final exam's due date of October 19. Office hours to assist students with homework will be available each week, and course instructors will be available outside of class for 1-on-1.

Week #1

September, 14, 16; 3:00-4:30 PM

Introduction to UNIX

- Filesystem layout, navigation, communication, multi-user environment, permissions, file sharing, environment variables, paths, libraries, shells, pipes, interactive command and script execution, background/foreground, etc.

Week #2

September 21, 23; 3:00-4:30 PM

HPC Architecture and Introduction to the Minerva Supercomputer

- Von Neumann architecture, parallel processing, shared/distributed memory, vector processing, MIMD/SIMD, massively parallel processing, accelerator computing, parallel numerical libraries, motivated examples from computational biology and bioinformatics, the GATK pipeline as a reference on parallel methods.
- Minerva hardware architecture, software environment, module system, LSF and queueing systems, strategies for submitting jobs, interactive vs. batch mode, from interactive command execution to job execution on a cluster.

Week #3

September 28, 30; 3:00-4:30 PM

Shell Scripting

- Scripting vs. compiled languages, scripting control flow and basic constructs, awk, sed, sort, uniq, advanced UNIX pipes.

Week #4

October 5, 7; 3:00-4:30 PM

Advanced Shell Scripting

- Data handling and processing, screen, interactive jobs, advanced job submission with complex scripts.

Python Fundamentals

- Introducing Python, Python vs. other languages (e.g., Perl), variables, operators, data structures.

Week #5

October 12, 14; 3:00-4:30 PM

Continuing Python and Hands on in Bioinformatics Supercomputing

- Decisions and loops, file I/O, functions, and exception handling.
- Introduction to Python modules and packages (scipy, numpy, pandas, etc.).

Final Exam Assigned October 12, Due October 19