



# Training Workshops & Seminars

*April, 2018*

Henrietta Harvey Building, HH-2030  
Memorial University  
St. John's NL, A1C 5S7  
[info@ace-net.ca](mailto:info@ace-net.ca)  
[www.acceleratediscovery.ca](http://www.acceleratediscovery.ca)

Orientation .....	4
How Can Advanced Computing Expand My Research? (1 hour) .....	4
Basics .....	4
Introduction to HPC with ACENET & Compute Canada (1 hour).....	4
Introduction to Using New National Compute Canada Systems (1 hour) .....	4
Introduction to Linux (1 hour) .....	5
Introduction to Shell Scripting (1 hour) .....	5
Job Scheduling With Slurm (1 hour) .....	5
Job Scheduling with Grid Engine (1 hour).....	5
Big Data & Data Analytics (1 hour) .....	6
Git, Github or Gitlab? An Overview of Version Control Tools (1 hour).....	6
Software & Data Carpentry Bootcamps .....	6
Software Carpentry (2 days) .....	6
Data Carpentry (2 days) .....	6
Cloud Computing .....	7
Introduction to Cloud Computing (1 hour) .....	7
Cloud Powering Up Humanities & Social Sciences (2 days) .....	7
Data Management.....	7
Data Management Planning (1.5 hours) .....	7
ACENET Focus Seminars.....	8
Job Scheduling .....	8
Managing Many Jobs (2 hours).....	8
Parallel Programming .....	8
Introduction to Parallel Programming (1 hour) .....	8
Lightning Introduction to MPI Programming (1 hour) .....	8
Introduction to Shared-Memory (OpenMP) Programming (1 hour).....	8
Advanced OpenMP Programming (2 hours) .....	8
Parallel Computing School (3.5 days).....	8
Tools & Techniques .....	9
Make and Makefiles (1 hour).....	9
Code Optimization - Tips and Tricks (1 hour) .....	9
Troubleshooting Session (1 hour) .....	9
Molecular Dynamics Workshop (full day) .....	9

Coding .....	9
Introduction to Coding Using C++ (1 hour) .....	9
Introduction to C++ (4 hours) .....	9
Introduction to C (1 hour) .....	9
Further Topics in C (1 hour) .....	9
Introduction to Fortran 77/90 (4 hours) .....	10
Introduction to Python (2 hours) .....	10
Scientific Python (2 hours) .....	10
Python – Profiling & Optimizing Code (1.5 hours) .....	10
Python Imaging (2 hours) .....	10
Introduction to R (2 hours) .....	10
R Hands-On (2 hours) .....	10
R - Big Data and Parallel Programming (2 hours) .....	10
R - Programming and Building Modules (2 hours) .....	11

# Orientation

## **How Can Advanced Computing Expand My Research? (1 hour)**

Do you have large data sets that you would like to mine and analyze in innovative ways? Are you trying to model something that's too complex for your desktop? Do you want to look for patterns in visual imagery, reveal trends in spatial data, or perform quantitative analysis of digitized texts? Perhaps you want to build a web-based research or analytical tool and don't know where to start. This orientation session will introduce you to the terms and concepts around high performance (HPC), supercomputing and big data analytics, the resources available within Compute Canada and ACENET, what they do and how they apply to various types of research. It will also provide some illustrative examples of how researchers across the sciences and humanities are using advanced computing, and how to get started.

# Basics

These hour-long sessions introduce the essentials of advanced computing at ACENET. It is strongly recommended that all new Compute Canada/ACENET users take the *Introduction to HPC With ACENET & Compute Canada*, *Introduction to Linux*, *Introduction to Shell Scripting* and *Job Scheduling with Slurm* sessions.

Experienced users may find the *Introduction to Shell Scripting* and *Job Scheduling with Slurm* sessions useful in order to gain greater efficiency from the computing clusters.

## **Introduction to HPC with ACENET & Compute Canada (1 hour)**

What is high performance computing (HPC) and what can it do for me? How can ACENET help? Used by researchers across many disciplines to tackle analyses too large or complex for a desktop, or to achieve improved efficiency over a desktop, this session takes participants through the preliminary stages of learning about HPC and computing clusters, and how to get started with this type of computing. It then reviews software packages available for applications, data analysis, software development and compiling code. Finally, participants will be introduced to the concept of parallel computing to achieve much faster results in analysis. This session is designed for those with no prior experience in HPC, and are looking for an introduction and overview.

## **Introduction to Using New National Compute Canada Systems (1 hour)**

This is an introductory session to using the new national systems, Graham and Cedar. Participants will learn about their capabilities, file systems, available software, the Slurm scheduler, and how to get up and running with them. This session provides an overview and quick-start guidance.

### **Introduction to Linux (1 hour)**

Linux is the terminal interface used to enable you to use the ACENET and Compute Canada HPC clusters from your desktop. It's the tool to get your data on the clusters, run your programs, and get your data back. In this session, learn how to get started with Linux, how to create and navigate directories for your data, load files, manage your storage, run programs on the computing clusters, and set file permissions. This workshop is designed for those with no prior experience in working with a terminal interface. Participants are encouraged to bring a laptop.

### **Introduction to Shell Scripting (1 hour)**

Participants will learn how to use shell scripting to exercise the power of the command line. Shell scripting helps you save time, automate file management tasks, and better use Linux. This session teaches you how to name, locate and set permissions for executable files, taking input and producing output. Learn about job scripts, shell variables and looping commands. This workshop is designed for either new HPC users who are familiar with working in a Linux environment, but have not had experience with shell scripting, or for experienced users seeking to get more out of shell scripting. In order to get the most from the session, participants are strongly encouraged to have a Compute Canada/ACENET account and to bring a laptop to do the exercises.

### **Job Scheduling With Slurm (1 hour)**

This session teaches participants how to use Compute Canada's queuing environment on the new national systems (Cedar and Graham), using the job scheduler Slurm. Learn how the scheduler works, how it allocates jobs, what are reasonable requests to minimize wait time, how to make the best use of the resources to be more efficient, how to get more throughput, how to get more jobs running at the same time, and how to troubleshoot and deal with crashes. This workshop is designed for either new HPC users familiar with Linux and Shell Scripting, but who have not had experience with using Slurm, or, for experienced users transitioning to Slurm or seeking to improve efficiency with the scheduler. In order to get the most from the session, participants are strongly encouraged to have a Compute Canada/ACENET account and to bring a laptop to do the exercises.

### **Job Scheduling with Grid Engine (1 hour)**

This session teaches job scheduling on ACENET systems that use the Grid Engine scheduler. Learn how the scheduler works, how it allocates jobs, what are reasonable requests to minimize wait time, how to make the best use of the resources to be more efficient, how to get more throughput, how to get more jobs running at the same time, and how to troubleshoot and deal with crashes. This workshop is designed for either new HPC users familiar with Linux and Shell Scripting, but who have not had experience using Grid Engine, or, for experienced users seeking to improve efficiency with the scheduler. In order to get the most from the session, participants are strongly encouraged to have a Compute Canada/ACENET account and to bring a laptop to do the exercises.

### **Big Data & Data Analytics (1 hour)**

Data analysis and processing has been around for a long time – what is new is the variety, volume, and velocity at which things can now be quantified. Today, data is available from a myriad of sources such as, tweets, GPS coordinates, images, and videos from mobile devices, and massive on-line archives of digitized manuscripts to name a few. The amount and variety of data make it difficult to analyze and we as a society are still working out the best ways to use this data to improve our understanding and make better decisions. This seminar will appeal to people from humanities to natural sciences and in-between, and will help you determine how your data fits into the big data ecosystem and introduce some tools to help you manage and analyze your data.

### **Git, Github or Gitlab? An Overview of Version Control Tools (1 hour)**

Version control has been a key concept in software engineering for decades, but it's now becoming known and used by more than just hard-core programmers. The most popular tool is currently Git, but it's part of a large "ecosystem" of web services (GitHub, BitBucket, GitLab, ...) and competing tools (Subversion, Mercurial, Perforce...). What do these all mean? How do you choose a combination of version control tools and services for your work?

## **Software & Data Carpentry Bootcamps**

Four hour-long lectures introduce the essentials of advanced computing at ACENET. A certificate of completion will be offered to those who attend all four.

### **Software Carpentry (2 days)**

[Software Carpentry](#) teaches basic computational research skills. Sessions include program design, version control, data management and task automation. This is a hands-on workshop, where collaboration is encouraged and participants are asked to apply their learning to individual research problems. This workshop is of interest to: current computational researchers and their teams; anyone considering a research project that requires computational research; or students who want to gain these skills to enhance their career choices. A laptop is required.

### **Data Carpentry (2 days)**

[Data Carpentry](#) trains researchers in the core data skills for efficient, shareable, and reproducible research practices covering the full lifecycle of data-driven research. Through a two-day hands-on approach, the focus is on the introductory computational skills needed for data management and analysis. The workshops are domain-specific, and include life and physical sciences, and humanities and social sciences. Sessions build on the existing knowledge of participants to enable them to quickly apply the skills learned to their own research. This workshop is of interest to: current researchers and their teams using large data sets; anyone considering a research project that involves large data sets; or students who want to gain these skills to enhance their career choices. A laptop is required.

# Cloud Computing

## **Introduction to Cloud Computing (1 hour)**

This session provides an introduction to the Compute Canada cloud which is used to create and manage virtual machines. Virtual machines allow great flexibility but require knowledge and effort to configure them for your specific needs. Virtual machines can be used for diverse work flows, from processing particle physics data to running humanities and social sciences websites. Learn how to create a virtual machine and how to start using it.

## **Cloud Powering Up Humanities & Social Sciences (2 days)**

Cloud computing has become very popular. Part of the reason for this popularity is that it provides great flexibility allowing complete control of the computing environment. In addition the environment can be copied, backed up, created and recreated in an automated way. In these lessons we will start you on the path towards making use of the great flexibility and power of cloud computing. This two day workshop is designed to provide an introduction to using cloud computing, as well as to teach the basics in getting started. Participants will learn how to create a virtual machine, apply updates, create a web server, a Wordpress site, a self-signed SSL Certificate, and will receive a demonstration of how to use Heat to create a mediawiki site. No experience necessary, but participants are strongly encouraged to open a Compute Canada account prior to the course and to bring a laptop.

# Data Management

## **Data Management Planning (1.5 hours)**

This workshop provides a brief overview of Research Data Management, with a focus on the development of Data Management Plans (DMPs). DMPs are a key component of the data management process that touch on all aspects of RDM, describing how data are collected, formatted, preserved, and shared. Importantly, DMPs also promote the consideration of the costs and challenges associated with managing research data. Participants will learn about RDM in the context of the research data lifecycle, gain an understanding of the basic components of a DMP, create a DMP for their research project using the Portage Network's [DMP Assistant tool](#), and think critically about data management challenges and how to evaluate DMPs. No software or prior data management training required. Before the workshop, participants should create an account on the Portage Network DMP Assistant tool and come prepared with information about a specific project for which the DMP can be written (e.g., a project workplan or research abstract). Participants must also bring their own laptop.

# ACENET Focus Seminars

Focus Seminars dive more deeply into topics that may be more advanced, or of more specialized interest than the Basics series.

## [Job Scheduling](#)

### **Managing Many Jobs (2 hours)**

Learn how to handle dozens or hundreds of jobs without typing “qsub” every time, using templates, loops, and task arrays. This workshop is most useful for those who already have an ACENET account and have completed one of the Basics workshops on job scheduling. Bring a laptop to participate in the exercises.

## [Parallel Programming](#)

### **Introduction to Parallel Programming (1 hour)**

This workshop introduces the terminology and concepts of parallel programming. This includes types of parallel programs, design methodologies, and performance measures. Appropriate for those with programming experience and who have taken the ACENET Basics workshops.

### **Lightning Introduction to MPI Programming (1 hour)**

This is a one-hour crash course in the primary tool for writing message-passing parallel programs. It covers the basic concepts of MPI, including sending and receiving messages, coordination, and data synchronization. This workshop assumes that you have some programming experience with one or more languages and have taken the ACENET Basics seminars.

### **Introduction to Shared-Memory (OpenMP) Programming (1 hour)**

Participants will learn about the primary tool for writing shared-memory parallel programs. The session will cover OpenMP and provide a short introduction to POSIX threads. You will learn about the most common techniques, such as parallel for loops, barriers and critical sections. This workshop assumes that you have programming experience with one or more languages and have taken the ACENET Basics sessions.

### **Advanced OpenMP Programming (2 hours)**

This workshop assumes that you have some experience with OpenMP or some other shared-memory programming paradigm. It covers more advanced topics like controlling work distributions, creating and using task pools and several other areas.

### **Parallel Computing School (3.5 days)**

The school is designed for participants familiar with the Linux command line and who have some level of programming experience. Completion of the ACENET Basics Series, or equivalent experience, is strongly recommended. The mornings will consist of lectures, with the afternoons following a lab format, where participants will be given exercises, or can bring specific problems to instructors related to their research. Topics include general parallel computing, OpenMP, GPGPU, and Message Passing Interface (MPI).



## Tools & Techniques

### **Make and Makefiles (1 hour)**

Many programs come as source code and a mysterious Makefile, with instructions like “make all; make install”. Here’s how it works, what can go wrong, how to fix it, and maybe even how to write your own. Participants must have some programming experience and have taken the ACENET Basics sessions.

### **Code Optimization - Tips and Tricks (1 hour)**

This workshop assumes that you have some experience with one or more programming languages. It covers some basic optimization techniques that are general in nature, as well as several tips for specific languages.

### **Troubleshooting Session (1 hour)**

For those with some experience using the Compute Canada and ACENET systems, this session is an informal opportunity to address individual difficulties, areas of concern, or questions.

### **Molecular Dynamics Workshop (full day)**

This workshop introduces researchers to the theory, key ideas, and techniques of Molecular Dynamics. A practical application targeted at biosimulations is introduced using the GROMACS package.

## Coding

### **Introduction to Coding Using C++ (1 hour)**

This workshop is designed for those who have no experience in coding, who want to understand the general principles of coding and have taken the ACENET Basics series.

### **Introduction to C++ (4 hours)**

This is a more in-depth hands-on coding workshop for those who are interested in coding in C++. It is designed for 8-10 participants who have ACENET accounts. Participants must have some rudimentary programming experience and have taken the ACENET Introduction to Coding and the ACENET Basics series. The workshop format is two 45-minute lectures, two 30-minute breaks, and two 45-minute student programming sessions.

### **Introduction to C (1 hour)**

This workshop will cover an introduction to the C programming language, including syntax, variables and data structures. While not necessary, some experience with programming concepts would be helpful. By the end of the workshop, you should be able to read C programs and be able to write and compile simple ones of your own.

### **Further Topics in C (1 hour)**

This workshop assumes that you already know the basics of the C programming language. We will move on to deeper topics like memory management, pointer arithmetic and file I/O.

### **Introduction to Fortran 77/90 (4 hours)**

This is a hands-on coding workshop for those who are interested in coding in Fortran 77/90. Designed for 8-10 participants who have ACENET accounts. Participants must have some rudimentary programming experience and have taken the ACENET introduction to coding and the ACENET Basics series. The workshop format is two 45-minute lectures, two 30-minute breaks, and two 45-minute student programming sessions.

### **Introduction to Python (2 hours)**

This workshop will cover an introduction to the Python programming language, including syntax, variables and data structures. While not necessary, some experience with programming concepts would be helpful. By the end of the workshop, you should be able to read Python programs and be able to write simple ones of your own.

### **Scientific Python (2 hours)**

This workshop assumes that you have some experience with the Python programming language. It will cover several packages, including numpy and scipy, that are useful in doing scientific computations. Several different examples will be discussed, including solving PDEs, solving systems of equations, and even doing symbolic computations.

### **Python – Profiling & Optimizing Code (1.5 hours)**

This workshop assumes that you have some previous Python experience. You will learn many of the tools available to profile your code and find the trouble spots. Once located, the second half of the workshop presents a series of tips and tricks that may be able to help you speed up the execution of your program.

### **Python Imaging (2 hours)**

This workshop assumes that you have some experience with the Python language. It will cover the external packages that are useful in doing image processing. This includes techniques that may be helpful in biomedical imaging, space based earth imaging, or anything else image related.

### **Introduction to R (2 hours)**

This workshop will cover the basics of R, how to do basic statistics and plotting. It will cover how to install and use extra packages, but not how to write your own. By the end of the workshop, you will be able to do basic statistical analysis, and know how to access the tools to do more advanced analysis.

### **R Hands-On (2 hours)**

This workshop provides a hands-on session where attendees are taken through several tasks in order to become familiar with how R works.

### **R - Big Data and Parallel Programming (2 hours)**

This workshop assumes that you have some experience with R. It will cover some of the packages that can help when dealing with very large data sets. It will also cover some of the packages and techniques that are useful when trying to do parallel programming in R, including issues in trying to do parallel work on Windows machines.

## **R - Programming and Building Modules (2 hours)**

This workshop assumes that you have some experience with R. It will cover using R as a full programming language, allowing you to write your own code. It will also cover the basics of creating your own packages for when you are ready to share your code with other users.