Seminars

Research Computing Seminars

Class Notes

Any member of the Lehigh community who would like to learn more about Research Computing and High-Performance Computing (HPC), but cannot easily attend our seminar series, is welcome to review our Research Computing Notes at the following location. Note that you must be on the campus network or VPN to access these notes:

Research Computing Seminar Notes

These notes will be continuously updated during each Seminar Series listed below. Research Computing is characterized by a wide diversity of different projects and goals. If you would like to discuss your specific goals with our staff, schedule a “Research” consultation here.

Spring 2024 Seminars

In the Spring, 2024 semester we will continue our high-performance computing (HPC) seminar series. This series contains five parts which cover a wide breadth of high-performance computing fundamentals with an emphasis on surveying many different techniques that may help with your research. Experienced HPC users are welcome to attend any individual sessions that interest them, while we recommend that novice users attend the first three sessions, namely the quickstart, parallelism, and interactive computing sessions.

While there are no formal prerequisites, this series does not cover Linux fundamentals and the use of version control (typically git). To acquire these skills before the HPC seminar series, researchers of all skill levels can benefit from joining LinkedIn Learning, which is free for the Lehigh community. This provides access to two critical courses. First, the “Learning Linux Command Line” course will introduce you the Linux command line, which is essential for interacting with almost all high-performance computing platforms. Second, the “Git Essentials” course will teach you the skills required to collaboratively develop and share your code and electronic lab notebooks.

Notes from the Quickstart are available online (you must be on the Lehigh network to access this).

This series is held Tuesdays from 3/19-4/16 at 2PM in the EWFM Digital Media Studio.

<table>
<thead>
<tr>
<th>Date</th>
<th>Title</th>
<th>Description</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 19</td>
<td>Quickstart</td>
<td>This seminar provides an introduction to high-performance computing (HPC) resources at Lehigh along with a walkthrough of the essential ingredients required to build research computing projects. Participants will learn the best practices for designing small, exploratory projects so that they can be scaled up to answer questions that require big data and parallel computations. This course will teach newcomers how to use Linux and Python to complete a simple but powerful research computing project.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ryan Bradley</td>
</tr>
<tr>
<td>March 26</td>
<td>Parallelism</td>
<td>Many exploratory research computing projects will eventually grow to require larger amounts of computational resources, either because they need to demonstrate that they are correct and reproducible, or to answer larger, more complicated research questions. This seminar will teach you to use the Lehigh high-performance computing (HPC) cluster to build parallel scientific software for answering larger questions using simple Linux tools along with the HPC scheduler (named SLURM).</td>
<td>Ryan Bradley</td>
</tr>
<tr>
<td>April 2</td>
<td>Interactive Computing</td>
<td>Research computing projects of all sizes can benefit from interactive tools which make it easy to develop, debug, and visualize code and data on our local high-performance computing (HPC) systems. This seminar will demonstrate the use of a web portal to interact with your data and code, with a particular emphasis on connecting Jupyter notebooks to very large datasets and parallel computations.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ryan Bradley</td>
</tr>
<tr>
<td>April 9</td>
<td>Containers</td>
<td>Research computing projects can have a large impact if they are portable and easy to share with new collaborators. In this seminar, you will learn how to package your scientific computations in a Linux-based container system (Docker and Appatiner). Containers can make it easy to take your work to larger public supercomputing centers, and improve the reproducibility and openness of your scientific software.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ryan Bradley</td>
</tr>
<tr>
<td>April 16</td>
<td>Data Structures</td>
<td>When a research computing project evolves to answer more nuanced questions, it inevitably requires more computational power and generates more data. This seminar will teach you the best practices for storing large, complex datasets so they can easily be analyzed and mined for new insights. These methods will apply to any research questions that depend on big data.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ryan Bradley</td>
</tr>
</tbody>
</table>

Fall 2023 Seminars

In the fall we will resume our high-performance computing (HPC) seminar series. We begin with two identical sessions of the HPC Quickstart session, and then learn about parallelism, interactive computing, containers, and data structures. Notes from the Quickstart are available online (you must be on the Lehigh network to access this).

Click here to visit the LTS website for dates, times, and registration information.
<table>
<thead>
<tr>
<th>Date</th>
<th>Title</th>
</tr>
</thead>
</table>
| Sep 19 or Sep 26 | **Description:** This seminar provides an introduction to high-performance computing (HPC) resources at Lehigh along with a walkthrough of the essential ingredients required to build research computing projects. Participants will learn the best practices for designing small, exploratory projects so that they can be scaled up to answer questions that require big data and parallel computations. This course will teach newcomers how to use Linux and Python to complete a simple but powerful research computing project.  
Instructor: Ryan Bradley |
| Oct 3         | **Description:** Many exploratory research computing projects will eventually grow to require larger amounts of computational resources, either because they need to demonstrate that they are correct and reproducible, or to answer larger, more complicated research questions. This seminar will teach you to use the Lehigh high-performance computing (HPC) cluster to build parallel scientific software for answering larger questions using simple Linux tools along with the HPC scheduler (named SLURM).  
Instructor: Ryan Bradley |
| Oct 10        | **Description:** Research computing projects of all sizes can benefit from interactive tools which make it easy to develop, debug, and visualize code and data on our local high-performance computing (HPC) systems. This seminar will demonstrate the use of a web portal to interact with your data and code, with a particular emphasis on connecting Jupyter notebooks to very large datasets and parallel computations.  
Instructor: Ryan Bradley |
| Oct 17        | **Description:** Research computing projects can have a large impact if they are portable and easy to share with new collaborators. In this seminar, you will learn how to package your scientific computations in a Linux-based container system (Docker and Apptainer). Containers can make it easy to take your work to larger public supercomputing centers, and improve the reproducibility and openness of your scientific software.  
Instructor: Ryan Bradley |
| Oct 24        | **Description:** When a research computing project evolves to answer more nuanced questions, it inevitably requires more computational power and generates more data. This seminar will teach you the best practices for storing large, complex datasets so they can easily be analyzed and mined for new insights. These methods will apply to any research questions that depend on big data.  
Instructor: Ryan Bradley |

**Summer 2023 Seminars**

For the summer session, the research computing group will offer a single **High-Performance Computing Quickstart** training session, designed for small groups to learn more about completing their research computing projects on Lehigh University high-performance computing infrastructure.

This seminar reviews the minimum essential ingredients required to build research computing projects on Linux platforms. Students will learn the basics of Linux, use shell scripts, Python, and the scheduler on our high-performance computing (HPC) cluster to perform simple calculations. This course will outline the best practices for scaling and repeating many kinds of calculations. All students, faculty, and staff with an interest in learning more about HPC and Linux are welcome to attend.

Registration is available here for any one of the following dates.

1. Tuesday, 6/13/23 from 1-2:30PM (in person, Digital Media Studio Lab)  
2. Wednesday, 6/21/23 1-2:30PM (Zoom)  
3. Tuesday, 6/27/23 from 1-2:30PM (in person, Digital Media Studio Lab)  
4. Wednesday, 7/5/23 from 1-2:30PM (Zoom)

**Spring 2023 Seminars**

The research computing seminars offered in the Spring semester are designed to lead participants through a survey of the wide range of available research computing techniques that they can apply to their training and research. Over the course of this seminar series, students will be expected to imagine how their projects can be formulated to take advantage of both small and large scale computation. We will begin with the end in mind, by first creating a map of the high-performance computing (HPC) resources available to the research community at Lehigh. After this introduction we will study the Linux basics required to effectively work with code and data, and then tackle the scheduler (SLURM) that we use to share a large HPC platform among many users. The final three sessions will provide a deeper dive into interactive computing with the Open OnDemand portal, building your own highly-portable software containers, and using Python for advanced computation on the cluster. Attendees are welcome to bring their own research problems and questions to these sessions.

These sessions cooperate with a larger body of seminars hosted by Library and Technology Services. Of particular interest are the Python and R seminars, since the skills that you learn in these classes can translate into much larger problems that require high-performance computing.

Click here to visit the LTS website for dates, times, and registration information. This semester, we are offering both in-person and videoconference versions of these seminars.

<table>
<thead>
<tr>
<th>Date</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Title</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Jan 31 & Feb 2 | **Description:** This seminar provides an overview of the Research Computing resources available to the Lehigh research community. We introduce high-performance computing (HPC) and provide a guide for gaining access to research computing resources, specifically the hardware, software, and training required to support new research. We also provide an overview of external computing resources provided by the National Science Foundation (NSF).  
**Instructor:** Ryan Bradley | | | |
| Feb 7 & Feb 9 | **Description:** Linux is a free and open source operating system that is the operating system of choice of the world’s leading supercomputers as well as LTS high-performance computing (HPC) clusters and some computer labs at Lehigh. This session will provide an introduction to the Linux environment, using the command line, logging into remote systems, and transferring data, and the use of terminal-based text editors. This session is designed for researchers who want to use Linux resources to support their work.  
**Instructor:** Ryan Bradley | Download the slides here. | | |
| Feb 14 & Feb 16 | **Description:** Lehigh uses the SLURM scheduler to ensure that many researchers can effectively share our high-performance computing (HPC) resources. This session provides a hands-on introduction to SLURM so that participants can learn to translate their research questions into hardware requests on an HPC platform by formulating, submitting, and monitoring batch jobs. This session may also be useful for researchers who also leverage external HPC platforms in their work.  
**Instructor:** Ryan Bradley | | | |
| Feb 21 & Feb 23 | **Description:** Open OnDemand (OOD) is an NSF-funded open-source high-performance computing (HPC) portal developed by the Ohio Supercomputing Center. The goal of OOD is to provide an easy way for system administrators to provide web access to their HPC resources. This seminar introduces OOD to access interactive applications such as MATLAB, Jupyter Lab/Notebooks, and RStudio on Lehigh's HPC cluster. The portal provides a more fully-featured user interface to the cluster, complementing the standard, terminal-based method for remotely accessing the cluster.  
**Instructor:** Ryan Bradley | | | |
| Feb 28 & Mar 2 | **Description:** Container systems provide the tools for researchers to more easily build and migrate their software to diverse computing platforms. Singularity is an open-source container engine designed to bring operating system-level virtualization to scientific and high-performance computing. In this seminar, we’ll provide an overview of Singularity and teach you to build your own software in a container system for use on Sol. These techniques can improve the durability and reproducibility of your research.  
**Instructor:** Ryan Bradley | | | |
| Mar 7 & Mar 9 | **Description:** This session will provide a tour of advanced Python methods that can be useful on the cluster. Python often serves as a glue for binding high-performance computation to complex data structures and intricate workflows. This seminar will briefly survey the NumPy and SciPy mathematical libraries, using HDF5 to generate complex data files, and the method for extending Python with C extension modules.  
**Instructor:** Ryan Bradley | | | |

**Spring 2022 HPC Seminars**

When: Thursday @ 2PM

Where: Zoom
### March 3
**Description:** In this seminar, you will learn the basics of R, including language fundamentals, data types, functions and basic programming including File I/O.

**Prerequisites:** Programming background is beneficial but not required.

**Instructor:** Jeremy Mack

### March 10
**Description:** Open OnDemand (OOD) is an NSF-funded open-source HPC portal developed by the Ohio Supercomputing Center. The goal of OOD is to provide an easy way for system administrators to provide web access to their HPC resources. This seminar introduces OOD to access interactive applications such as MATLAB, Jupyter Lab/Notebooks, and RStudio on Lehigh’s HPC cluster.

**Prerequisites:** An HPC account with an active allocation and a web browser. A Lehigh IP or VPN required.

**Instructor:** Alex Pacheco

### March 17
**Description:** This seminar provides a hands on introduction to Data Visualization using the Python programming language.

**Prerequisites:** Programming background in Python is required.

**Instructor:** Sachin Joshi

### March 24
**Description:** This seminar provides a hands on introduction to Data Visualization using the R programming language

**Prerequisites:** Programming background in R is required.

**Instructor:** Jeremy Mack

### March 31
**Description:** Singularity is an open-source container engine designed to bring operating system-level virtualization to scientific and high-performance computing. In this seminar, we’ll provide an overview of Singularity and how you can Build Your Own Software in a singularity container for use in your research on Sol.

**Prerequisites:** A Linux system with singularity installed (check your package manager if your distribution provides singularity or see installation instructions).

**Instructor:** Alex Pacheco

### April 7
**Description:** Object Oriented Programming or OOP is a programming paradigm which provides a means of structuring programs (combining data and functionality) so that properties and behaviors are bundled into individual objects. In this seminar, you’ll learn the basic concepts of OOP in Python: Python Classes, Object Instances, Defining and Working with Methods and OOP Inheritance.

**Prerequisites:** Programming background in Python is required. Knowledge of OOP’s concepts is beneficial but not required.

**Instructor:** Sachin Joshi

### April 14
**Description:** Shiny Apps, interactive web applications built using R programming language, have grown in popularity. This session will provide attendees experience using the Shiny package and the idea of reactive programming, which forms the basis of building an interactive web application.

**Prerequisites:** Programming background in R is required.

**Instructor:** Jeremy Mack

### Archived Seminars

**Title**

**Description:** This training provides an overview of Research Computing resources available to the Lehigh research community.

**Prerequisites:**

- Familiarity with Linux/Unix environment, basic command and *nix editors such as vi or emacs is mandatory.

**Description:** This training provides a hands on introduction to using the SLURM scheduler to submit and monitor jobs. SLURM is the scheduler on Lehigh’s HPC resource, Sol, and national supercomputing resources including XSEDE and NERSC.

**Prerequisites:** An account on Sol or an account on national supercomputing resources (XSEDE, DOE, etc) that uses SLURM.

**Description:** Open OnDemand (OOD) is an NSF-funded open-source HPC portal developed by the Ohio Supercomputing Center. The goal of OOD is to provide an easy way for system administrators to provide web access to their HPC resources. This tutorial introduces OOD to access Lehigh’s Sol cluster.

**Prerequisites:** An account on Sol with an active allocation and a web browser. A Lehigh IP or VPN required.
| **Description:** This seminar is geared towards users who wish to bring or build their own software stack to use on Sol and Hawk (or even local Linux systems). Topics to be covered include best practices for installing packages using make, cmake and configure, SPACK package manager, and Singularity.  
**Prerequisites:** Some familiarity with compilers, and Linux environment is required. |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A Brief Introduction to Linux</strong></td>
<td></td>
</tr>
<tr>
<td>Slides</td>
<td>Recordings</td>
</tr>
</tbody>
</table>
| **Description:** Linux is a free and open source operating system that is the OS of choice of the world's leading supercomputers as well as LTS HPC clusters and some computer labs at Lehigh. This session will provide an introduction to the Linux/unix environment, command line basics, logging in to remote system, transferring data, vi/emacs editors etc to get started with using a Linux/Unix based computer. This training is geared towards researchers who want to learn or need to learn how to use a Linux/Unix based resource.  
**Prerequisites:** Some familiarity with compilers, and Linux environment is required. |
| **Basic Shell Scripting** |
| Slides | Recordings |
| **Advanced Shell Scripting** |
| Slides | Recordings |
| **Description:** In this tutorial, you will learn the basics of R, including language fundamentals, data types and functions and basic programming including File I/O.  
**Prerequisites:** Programming background is beneficial but not required. |
| **Description:** This tutorial provides a hands on introduction to Data Visualization using the R programming language  
**Prerequisites:** Programming background in R is required. |
| **Description:** Shiny Apps, interactive web applications built using R programming language, have grown in popularity. This session will provide attendees experience using the Shiny package and the idea of reactive programming, which forms the basis of building an interactive web application.  
**Prerequisites:** Programming background in R is required. |
| **Description:** In this seminar, you will learn the basics of Python, including language fundamentals and basic programming.  
**Prerequisites:** Programming background is beneficial but not required. |
| **Description:** Data Structures are the fundamental constructs around which you build your programs. Python is a high-level, interpreted, interactive and object-oriented scripting language using which we can study the fundamentals of data structure in a simpler way as compared to other programming languages. In this seminar we are going to study a short overview of some frequently used data structures in general and how they are related to some specific Python data types.  
**Prerequisites:** Programming background in Python is required. Knowledge of data structure concepts is beneficial but not required. |
| **Description:** In this seminar you will learn the basics of Python, including language fundamentals and basic programming.  
**Prerequisites:** Programming background is beneficial but not required. |
| **Description:** Object Oriented Programming or OOP is a programming paradigm which provides a means of structuring programs (combining data and functionality) so that properties and behaviors are bundled into individual objects. In this seminar, you'll learn the basic concepts of OOP in Python: Python Classes, Object Instances, Defining and Working with Methods and OOP Inheritance.  
**Prerequisites:** Programming background in Python is required. Knowledge of OOP's concepts is beneficial but not required. |
| **Description:** Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence (AI) based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. In this two part seminar you'll learn the basic machine learning concepts: supervised and unsupervised learning, classification and regression algorithms. This is a two part seminar.  
**Prerequisites:** Basic knowledge of general CS concepts such as data structures, linear algebra and design of algorithm is required. |
| **Description:** Text mining, also known as Text Analytics is the process of deriving high-quality information from text. It is the process of examining large collections of written resources to generate new information and to transform the unstructured text into structured data for use in further analysis. In this seminar you'll learn the basic introduction to text analysis, different steps of text analysis; gathering textual data, cleaning and preparing textual data, analyzing the data (ETL) and visualizing textual data using tools from the HathiTrust Research Center (HTRC).  
**Prerequisites:** Programming background is required. Basic knowledge of general CS concepts such as data structures, linear algebra and design of algorithms is required. |
| Slides (Part I) | Recordings (Part I, Part II) |
Description: MATLAB is a programming platform designed specifically for engineers and scientists. MATLAB is a special-purpose language that is an excellent choice for writing moderate-size programs that solve problems involving the manipulation of numbers. MATLAB is easy to learn, versatile and the design of the language makes it possible to write a powerful program in a few lines. Using MATLAB, you can: Analyze data, Develop algorithms and Create models and applications. In this seminar you’ll learn the basics of MATLAB, including language fundamentals and basic programming that will help you achieve the above functionalities.

Prerequisites: Programming background is beneficial but not required.

Version Control with GIT          Slides
Document Creation with LaTeX      Slides
Using Virtualized Software at Lehigh Slides
Storage Options at Lehigh         Slides
Research Data Management          Slides
Enhancing Research Impact         Slides

Guest Seminars

1. "Using Overleaf for Writing", Graduate Writing Retreat, Lehigh University, December 5, 2020

Online HPC Training

Please feel free to attend the training programs offered by other Universities and HPC centers.

<table>
<thead>
<tr>
<th>Center</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extreme Science and Engineering Discovery Environment (XSEDE)</td>
<td><a href="https://portal.xsede.org/course-calendar">https://portal.xsede.org/course-calendar</a></td>
</tr>
<tr>
<td>Oklahoma Supercomputing Center</td>
<td><a href="http://www.oscer.ou.edu/education.php">http://www.oscer.ou.edu/education.php</a></td>
</tr>
<tr>
<td>National Institute of Computational Sciences</td>
<td><a href="https://www.nics.tennessee.edu/hpc-seminar-series">https://www.nics.tennessee.edu/hpc-seminar-series</a></td>
</tr>
<tr>
<td>Louisiana State University</td>
<td><a href="http://www.hpc.lsu.edu/training/tutorials.php">http://www.hpc.lsu.edu/training/tutorials.php</a></td>
</tr>
<tr>
<td>Texas Advanced Computing Center</td>
<td><a href="https://portal.tacc.utexas.edu/training">https://portal.tacc.utexas.edu/training</a></td>
</tr>
<tr>
<td>Cornell University Center for Advanced Computing</td>
<td><a href="https://cvw.cac.cornell.edu/topics">https://cvw.cac.cornell.edu/topics</a></td>
</tr>
<tr>
<td>National Center for Supercomputing Applications</td>
<td><a href="https://www.citutor.org">https://www.citutor.org</a></td>
</tr>
</tbody>
</table>

Self-paced HPC Training

The HPC University (HPCU) is a virtual organization whose primary goal is to provide a cohesive, persistent, and sustainable on-line environment to share educational and training materials for a continuum of high performance computing environments that span desktop computing capabilities to the highest-end of computing facilities offered by HPC centers. Please visit the Training Roadmap for a collection of self-paced tutorials from various HPC centers on various topics ranging from basic concepts to programming and data visualization.