HPC Training Offerings

Work in Progress
We are still working on this page. Information may be incomplete.

- Tutorials
  - HPC and Big Data in-class Tutorials
    - A Hands-on introduction to Unix/Linux and the Shell on the NYU HPC cluster
    - Getting started in the NYU HPC Environment
    - Using slurm on the NYU Prince cluster
    - MapReduce
    - Using Hive
    - Intro to Spark
  - Related in-class Tutorials
    - MATLAB
    - R
    - STATA
    - Research Data Management (RDM)
- Workshops and Events
  - Big Data Tutorial 1: MapReduce workshop @ Dibner
- Past Workshops and Events
  - Colfax Developer Training: Parallel Programming and Optimization for Intel® Architecture
  - HPC Overview Presentation
- Learning Resources
  - Cornell Virtual Workshop
  - Lynda.com
  - Safari Books Online
  - Books24x7

Tutorials

The NYU HPC team offers in-class tutorials in a variety of topics and levels. The complete schedule of the tutorials can be found on the Classes Calendar. Several upcoming tutorials are listed on the table at the right of the page.

Tutorial registrations open two weeks prior to the tutorial. To register for an offered class, select a tutorial on the Data Services Calendar.

Tutorials are offered free of charge.

Access to Bobst Library
Most of the tutorials take place on the 6th floor of the NYU Bobst Library (room 617). Please take a look at the information on how to access Bobst Library and note that registering for a tutorial does not guarantee entry into the building.

HPC and Big Data in-class Tutorials

Do I need an HPC User Account?
All HPC tutorials include on-hands exercises using the HPC cluster or the Hadoop cluster. If you are planning to attend an HPC or Big data tutorial, it is highly recommended to get an HPC user account (if you do not have one already). You may apply for an HPC account as described at High Performance Computing at NYU (see “Getting and Renewing Access”).

Please note: A valid NYU Netid is required to apply for an NYU HPC user account.
A Hands-on introduction to Unix/Linux and the Shell on the NYU HPC cluster

A hands-on introduction to using the Unix command line interface. Aimed at prospective HPC users with little to no Linux experience. This class provides a general introduction to the operating system used on the NYU HPC.

To find out more about this tutorial, click here to expand...

**Prerequisites:**

- NYU HPC User Account
- Access to Bobst Library

**Tutorial Schedule and Registration:**

To register for an upcoming tutorial see the calendar on the right of this page or visit the [Data Services Classes calendar](#).

**Class Materials:** *A Hands-On introduction to Unix/Linux and the Shell*

**Learning Resources:** The following online resources provide similar content to the material covered in the tutorial

- Lynda.com Video Tutorials (NYU community only)
  - Learn the Linux Command Line: The Basics
- Cornell Virtual Workshop
  - An Introduction to Linux
- Books24x7
  - The Linux Command Line: A Complete Introduction

**Getting started in the NYU HPC Environment**

A hands-on introduction to the NYU HPC Environment. Aimed at NYU HPC users who are already comfortable with the Linux command-line environment, but new to HPC.

The goals of this tutorial are:

- Become familiar with the main components of the NYU HPC Cluster (Computing resources, Data storage systems, Software)
- Prepare, submit and monitor a simple batch job
Prerequisites:

- Comfortable with the Linux command-line environment (see tutorial: A Hands-on introduction to Unix/Linux and the Shell)
- NYU HPC User Account
- Confirm you can SSH to the NYU HPC Cluster
- Access to Bobst Library

Tutorial Schedule and Registration:

To register for an upcoming tutorial see the calendar on the right of this page or visit the Data Services Classes calendar

Class Materials: Getting started in the NYU HPC Environment

Learning Resources: The following online resources provide similar content to the material covered in the tutorial

- Lynda.com Video Tutorials (NYU community only)
  - Linux Bash Shell and Scripts
- Safari Books Online
  - Learning the Unix operating System, 5th Edition
  - Learning the bash Shell, 3rd Edition
  - Learning GNU Emacs, 3rd Edition
- Books24x7
  - Shell Scripting: Expert Recipes for Linux, Bash and more

Using slurm on the NYU Prince cluster

Work in Progress

Using slurm on the Prince cluster is a new tutorial. We are now in the process of developing this tutorial.

A hands-on introduction to using the slurm workload manager on the Prince cluster. Aimed at HPC users with knowledge of Linux.

Prerequisites:

- NYU HPC User Account
- Confirm you can SSH to the NYU HPC Cluster (i.e., Prince)
- Access to Bobst Library

Tutorial Schedule and Registration:

To register for an upcoming tutorial see the calendar on the right of this page or visit the Data Services Classes calendar

Class Materials: IN PROGRESS

Learning Resources: The following online resources provide similar content to the material covered in the tutorial

- Cornell Virtual Workshop

MapReduce

This class will provide a brief overview of what Hadoop is and the various components that are involved in the Hadoop ecosystem. There will be a hands-on showcase for the users on how to use the dumbo(Hadoop) cluster to run basic map-reduce jobs. Various hands-on exercises have been incorporated for the users to get a better understanding.
Prerequisites:
- NYU HPC User Account
- Confirm you can SSH to the NYU Hadoop Cluster (i.e., Dumbo)
- Access to Bobst Library

Tutorial Schedule and Registration:
To register for an upcoming tutorial see the calendar on the right of this page or visit the Data Services Classes calendar.

Class Materials: Big Data Tutorial 1: MapReduce

Learning Resources: The following online resources provide similar content to the material covered in the tutorial:
- Lynda.com Video Tutorials (NYU community only)
  - Hadoop fundamentals
  - Data Analysis on Hadoop
- Cornell Virtual Workshop
- MapReduce
- Books24x7
  - Large Scale and Big Data: Processing and Management

Using Hive
This tutorial provides a basic understanding of Apache Hive and its usage in the Hadoop eco-system. There will be hands-on examples on how to use Apache Hive and a step by step instructional on how to run Hive jobs using NYU’s Dumbo (Hadoop) Cluster.

Class Materials: https://wikis.nyu.edu/display/NYUHPC/Big+Data+Tutorial+2%3A+Hive

Intro to Spark
This tutorial provides a basic understanding of Apache Spark and its usage in the Hadoop eco-system. There will be hands-on examples on how to use Apache Spark and a step by step instructional on how to run Spark jobs using NYU’s Dumbo (Hadoop) Cluster.

Class Materials: https://wikis.nyu.edu/display/NYUHPC/Big+Data+Tutorial+3%3A+Introduction+to+Spark

Related in-class Tutorials
The following pre-scheduled in-class tutorials are offered by Data Services and may be of interest to NYU HPC users.

Data Services Classes
For a compete list of tutorials offered by Data Services please see: Data Services Classes

MATLAB
For a complete list of MATLAB tutorials, schedule, prerequisites, and to register for one of the tutorials, please visit: http://guides.nyu.edu/matlab
Getting Started with MATLAB

MATLAB is a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numerical computation. This introductory tutorial is designed to cover the basics of getting started with MATLAB as well as demonstrate some of the capabilities of MATLAB.

MATLAB: Beyond Basics

This tutorial is designed for users who have about a year or two of experience with MATLAB and interested in learning to write more efficient MATLAB code.

R

For a complete list of R tutorials, and to register for one of the tutorials, please visit: http://guides.nyu.edu/R

Introduction to R

An introduction to R, the open source data analysis package. Topics include: installing R, creating and editing data sets, computing and recoding variables, basic commands and some of the fundamentals of programming.

Data Wrangling in R

Prior experience with R required!! The tutorial is designed to cover data management tools such as merging and reshaping datasets, conditional statements, and beginner loops and user written functions. Attendees are expected to have understanding of basic concepts of R objects and functions and should have used R on their own previously.

Creating Graphics in R

Use R to create vivid, intriguing and informative statistical graphics.

STATA

For a complete list of STATA tutorials, and to register for one of the tutorials, please visit: http://guides.nyu.edu/stata

Introduction to STATA

Stata is a software package for statistical analysis of data. This tutorial will introduce you to the basics of Stata and will cover creating and editing data sets, computing and recoding variables and provide an overview of basic commands.

Data Wrangling with STATA

In this tutorial, data management tools such as merging, reshaping, conditional statements, and beginner loops will be covered. Attendees should already understand basic Stata syntax and should have used Stata on their own previously.

Research Data Management (RDM)

For Research Data Management resources, please visit: http://guides.nyu.edu/data_management/resources
Introduction to Research Data Management

Research data management allows researchers the ability to more quickly, efficiently, and accurately find, access, and understand their own or others' data. This tutorial will provide an introduction to managing research data throughout its lifecycle in an effort to streamline and make the research process more efficient.

Writing a Data Management Plan

This tutorial covers the basics of writing a successful data management plan for federal funding agencies such as the NSF, NIH, NASA, and others. During this introductory tutorial, you will learn about the different requirements funding agencies have for your research data as well as how to best meet those obligations within your lab or research group.

Workshops and Events

Big Data Tutorial 1: MapReduce workshop @ Dibner

Wednesday October 26, 2016 11:00a - 01:00p
Dibner Library Room LC433

Colfax Developer Training: Parallel Programming and Optimization for Intel® Architecture

A free seminar and Web-based training offered by Colfax Research, on parallel programming and optimization for Intel Xeon and Intel Xeon Phi processors with original programming exercises and remote access to training servers.

http://colfaxresearch.com/events/nyu-16-10/
Thursday October 20, 2016 10:00a - 4:00p
Bobst Library Room 619
Slide deck is available here

HPC Overview Presentation

Wednesday September 28, 2016
Dibner Library
Eric please fill here.
Learning Resources

Cornell Virtual Workshop

Cornell Virtual Workshop℠ learning platform is designed to enhance the computational science skills of researchers, accelerate the adoption of new and emerging technologies, and broaden the participation of underrepresented groups in science and engineering.

- List of Topics

Lynda.com

Lynda.com is an online service for most NYU students, faculty and staff, that provides access to video tutorials on a number of different topics. It is available through the Academics and Work tabs in NYUHome and at www.nyu.edu/lynda.

See: lynda.com general information and getting help Knowledge Base article

Sample video tutorials:

- Learn the Linux Command Line: The Basics
- Linux Bash Shell and Scripts
- AWK Essential Training
- Hadoop fundamentals
- Data Analysis on Hadoop

Safari Books Online

Books24x7

For more info