

CyberInfrastructure Comprehensive, Applied, and Tangible Summer School CiberCATSS

CiberCATSS is a training program that will teach you some of the skills needed for (advanced) computing that is applicable to your research domain.

- CiberCATSS is **Comprehensive** in that we will teach an entire stack of skills starting from the basics: Python → Algorithms → Machine Learning with some HPC mixed in for good measure.
- CiberCATSS is **Applied** in that you will use your skills for a computing problem in your research domain.
- CiberCATSS is **Tangible** in that you will produce a working research product by the end of the 7 week training course.

Leadership:

- Philip Chang (Professor of Physics, PI of CiberCATSS)
- Mahsa Dabagh (Assistant Professor of Biomedical Engineering, CoI of CiberCATSS)
- Susan McRoy (Professor and Chair of Computer Science, CoI of CiberCATSS)
- Marjorie Piechowski (Assessment Lead)

Instructors:

1. Susan McRoy: Introduction to Data and Data Science
2. Apoorv Prasad: Introduction to Python
3. Philip Chang: Numerical and Statistical Algorithms in Python – subset and superset of Physics 801
4. Mahsa Dabagh: Introduction to Unix and HPC systems
5. Sammie Omranian: Introduction to Machine Learning

Teaching Assistants:

- Akshath Polkampally
- Saad Mirza
- Pushyan Jhaveri

The breakdown of the summer school is 3 weeks of instruction and 4 weeks of project time. The 3 weeks of instruction will start out slow, but the pace will be relentless. In the beginning, most of things will seem trivial, but it will ramp up and no prisoners will be taken. The TA's will ensure that no person gets left behind and the instructor will ensure that no TA's will be left behind. We anticipate that you will learn something valuable during your time here. During the 4 weeks of project time, we expect you to meet with your project mentor on a regular basis to ensure that you are making good progress. The TA's will also meet with you twice or thrice weekly (virtually, but in-person is optional) to ensure that you are making good progress and to answer any questions that you might have. Questions that stump them will percolate up the chain until someone knows the answer or no-one knows the answer.

Platform: We are test-driving a new platform from a company known as CamberCloud which is basically a jupyterhub with cloud based storage, but with hooks to allow you to startup a MPI cluster or an Apache Spark cluster. It was started by a Bay Area former astrophysicist who is a leader in data analytics. We will ensure that the platform works for us. You are welcome to use it for your project if you so desire. We think it is exceptionally useful for data analytic/machine learning projects at this point. They will be zooming in on Wednesday to say "hi" to everyone.

Daily Plan: During the instructional period, we will meet each day from 10am-4pm here in LEC 214. We'll go from 10-12:30, then break for lunch from 12:30-1:30 and then go again from 1:30-4pm. The morning and afternoon sessions will be broken down to lecture + lab exercises. The exceptions will be the afternoon of June 7 and (possibly) the day of June 15th. We will figure it out the plan during this interval.

Communications: During the instructional phase, we can just ask. But we have also setup a slack channel. We will send the invite shortly. Be sure to register. Don't be after to **ask** and **answer** questions. Think of this as a smaller version of stack overflow.

Schedule: This is subject to change.

Week	Date	Day	10-12:30	1:30-4:00
1	30-May	Tues	Intro/Setup; Intro to Data	Intro to Python
	31-May	Wed	Intro to Python	Intro to Python
	1-Jun	Thurs	Intro to Python	Intro to Python
	2-Jun	Friday	Intro to Python	Intro to Python
	5-Jun	Mon	Intro to Python	Intro to Python
2	6-Jun	Tues	Numerical programming	Numerical programming
	7-Jun	Wed	Numerical programming	Numerical programming
	8-Jun	Thurs	Numerical programming	Numerical programming
	9-Jun	Friday	Numerical programming	Numerical programming
	12-Jun	Mon	Unix and HPC	Unix and HPC
3	13-Jun	Tues	Machine Learning	Machine Learning
	14-Jun	Wed	Machine Learning	Machine Learning
	15-Jun	Thurs	Machine Learning	Machine Learning
	16-Jun	Friday	Machine Learning	Machine Learning

Project Phase: Following the 3-week instructional phase, we will embark on a 4 week project phase where you will apply some of the skills you learn to your proposed project. Failure is not an option. You will produce something that you can present at the end of the course.

Github: The git repository for the school is
<https://github.com/philchang/cibercatss.git>

To clone it, you can either use the git option in camber cloud or
git clone <https://github.com/philchang/cibercatss.git>

Every morning you will need to do a "git pull" to make sure everything is updated.

A word of warning: This is the first time something like this has been done and it is being run on a shoestring budget (to ensure that you get an optimal stipend). As a result expect some degree of chaos as we try to figure everything out. You are the beta testers for this program and as a result, don't expect a perfectly smooth ride. That being said, we selected you base mainly on your previous experience which will help ensure that we can recover quickly from any snafus and curveballs thrown our way.