



# CENTER FOR GEOSPATIAL INFORMATION SCIENCE

## Winter 2017 Skill-Set Course

The skill-set course sequence at the UMD Center for Geospatial Information Science is designed to provide a comprehensive overview of current state-of-the-art in geospatial information science (GIS) and computational social science (CSS), through intensive training over a short time frame. The skill-set courses provide advanced training in key aspects of GIS and CSS, alongside hands-on lab-based exercises across a diversity of devices, software libraries, exercises, and data environments that emphasize “learning by doing”.



Courses are taught by faculty and staff in the Department of Geographical Sciences and Center for Geospatial Information Science at the University of Maryland, who have extensive experience in building and applying geographic information systems for use in research and professional environments.

Students will also have opportunities to work hands-on with our state-of-the-art facilities, including our high-performance computing cluster and connections to commercial cloud computing providers.

**This Winter, we are offering course on:**

### ◆ GEOG788C: GIS and Hadoop for Big Data (1 credit)

This course is tailored for students and professionals who are newcomers to spatial data sciences and high performance computing, especially those who are interested in Hadoop implementation in processing aspatial and spatial data..

**Course description and schedule are available at**  
<http://geospatial.umd.edu/education/landingtopic/winter-workshops>.

### Enrollment

Current UMD graduate and undergraduate students, Students from other universities and colleges, Professionals, and Life-long learners are welcome to enroll. Undergraduates may contact Dr. Ruibo Han ([ruibo@umd.edu](mailto:ruibo@umd.edu)) for permission.

To register, please visit <http://oes.umd.edu/winter-term>.

### Tuition Remission

Don't forget to utilize your tuition remission to attend these courses. E.g. Full-time Graduate Assistants are eligible for 4 credits of tuition remission for Winter Term, and Part-time Graduate Assistants are eligible for 2 credits. More info. is available at <https://uhr.umd.edu/benefits/tuition-remission>.



### Contact Us

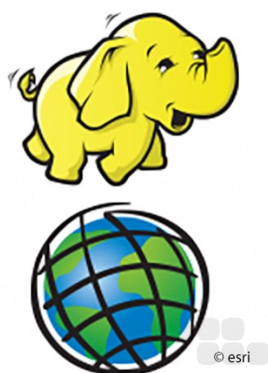
Ruibo Han, PhD  
Director of Programs, Center for GIS  
1111B LeFrak Hall  
College Park, MD 20742  
E-mail: [ruibo@umd.edu](mailto:ruibo@umd.edu)  
Tel: (301) 314-1343



**Center for Geospatial Information Science**  
Department of Geographical Sciences  
University of Maryland, College Park  
Winter 2017

# GIS and Hadoop for Big Data (1 credit)

January 3 to 6, 2017, 9 a.m. - 12:45 p.m



## Overview:

In a world of unstructured data, one of the few common structuring attributes is geography. This places geographic information and spatial data models front and center in the ongoing development of data management and access resources for big-data silos and the systems that rely on them. This course will focus on training students on the latest knowledge and techniques for processing spatial information embedded in big data. Students will work with the Apache Hadoop framework, which is evolving as one of the standard architectures for big data systems used in research and industry. The course will introduce basic concepts and structures of high-performance computing environments atop Apache Hadoop, with detailed instructions on deploying such an environment. Students in this course will develop proficiency in Open sources toolkits, such as GIS Tools for Hadoop, and will be introduced to the various pathways available to leverage the Hadoop framework to conduct spatial and related analyses on big data.

Upon a successful completion of the course the students will be able to:

- Design a solution and architecture on a Hadoop platform for a spatial analysis of big data on a cluster of machines;
- Apply MapReduce extensions, such as SpatialHadoop, to work with spatial data;
- Automate geospatial big data processing using Hive and GIS Tools for Hadoop;
- Run spatial operations/analysis on billions of spatial data records inside Hadoop;
- Visualize analysis results of big data via cartographic and geovisual media.

## Suggested background:

The prerequisites for this course include an introductory course for Geographic Information Systems (GIS). Students taking the course must be familiar with geographic data structures, basic GIS concepts, and demonstrate basic understanding of geospatial analysis.

## Hands-on training:

- Install and configure Virtual Machine with Ubuntu;
- Install and configure Hadoop with Pseudo-Distributed Mode;
- Demo word count example to introduce basic Hadoop mechanism;
- Install and configure Hive on current Hadoop, install postgre if needed;
- Compare performance of SQL querying and aggregation with postgre and Hive;
- Install Hadoop with Fully-Distributed mode on AWS via Cloudera.