Cloud Computing and Operations Research
Spring 2017

Instructor Section 001: Hardeep Johar (hj2203@columbia.edu)
Teaching Assistant: TBA
Class hours: Thursdays 4:10 pm - 6:40 pm
Office hours: TBA

Course Description
Cloud computing refers to the processing and storage of data on remote servers. Cloud computing enables the analysis of large data sets (aka 'big data') by making it possible to use multiple machines for processing (parallelism) without having to buy expensive hardware. Cloud computing also provides mechanisms for storing and managing large datasets across many servers.

The goal of this course is to introduce you to the programming issues around working with clouds for data analytics. While we will learn how to work with the infrastructure of cloud platforms, and talk a bit about distributed computing, the focus of the course is on programming. Topics covered will include MapReduce, parallelism, the rewriting of algorithms (statistical, OR, and machine learning) for the cloud, and the basics of porting applications so that they run on the cloud. We will mostly work in Python (2 or 3) so prior familiarity with Python is a must.

Prerequisites
The only prerequisite for this class is IEOR 4572 Python-Based Data Analytics or equivalent. It will be helpful if you’re also cross registered for one of Data Mining or Machine Learning but that is not essential.

Evaluation and learning components

Home assignments (30%): We’ll have a few home assignments as well. Like the quizzes, assignments are not meant to be diagnostic but rather to help you practice and learn so they will be very lightly graded. You can consult with others, ask me questions, google for help, but do try them on your own first. Because it is important that you do the assignments, I will accept late assignments. Assignments
submitted within a week of the due date will be penalized 25% and assignments submitted more than a week late will be penalized 50%.

**Project (40%)**: You will be required to choose a data set (you need to find your own!) and implement the analysis and reporting on a cloud platform. Final submission will include an in-class “speed-date” presentation and demonstration.

**Participation (10%)**: Demonstrate engagement in the course by asking questions. I’ll respond to every question, either online or, if the response is of general interest, in the classroom.

**Exams (20%)**: One in-class, closed book, exam