CATS Workshop
Friday, April 11

Training Students to Extract Value from Big Data

Workshop Overview
John Lafferty & Raghu Ramakrishnan
Deepak Agarwal, LinkedIn
Corinna Cortes, Google
Jeff Dozier, University of California, Santa Barbara
Anna Gilbert, University of Michigan
Patrick Hanrahan, Stanford University
Rafael Irizarri, Harvard University
Robert Kass, Carnegie Mellon University
John Lafferty, University of Chicago (co-chair)
Prabhakar Raghavan, Google
Raghu Ramakrishnan, Microsoft (co-chair)
Nathanaial Schenker, Centers for Disease Control and Prevention
Ion Stoica, University of California, Berkeley
Special thanks to:

Neal Glassman
Michelle Schwalbe
Scott Weidman
1. The need for training—experiences and case studies (“Why”)

Why is traditional training inadequate? How do big data problems arise in different organizations? How do people from different backgrounds come together to approach the problems?

2. Principles for working with Big Data (“What”)

What should we teach at various levels? Can we identify the principles that should form the core of the curriculum for big data training?
3. Courses, curricula, and interdisciplinary programs (“How”)

How should the principles be delivered? What are some sample courses that have been taught in the past? How can courses be tied together into a curriculum that will serve a broad interdisciplinary audience?

4. Shared resources (“How”)

A key ingredient in training for big data is ... big data. What are best practices for collecting, managing, and sharing massive datasets in the cloud? What possibilities exist (and will exist in the near future) for large-scale computational resources accessible in academic environments?
5. Panel discussion

Summarize the workshop, and allow for discussion of topics that were missed in the main sessions.
Session 1 (8:40–10:00)
The Need for Training: Experiences and Case Studies

Co-chairs: Raghu Ramakrishnan and John Lafferty

Speakers:

Rayid Ghani, University of Chicago
Guy Lebanon, Amazon Corporation

Break (10:00–10:15)
Session 2 (10:15–12:45)
Principles for Working with Big Data

Chair: Brian Caffo, Johns Hopkins University

Speakers:
- Jeff Ullman, Stanford University
- Alex Gray, Skytree Corp.
- Duncan Temple Lang, University of California, Davis
- Juliana Freire, New York University

Lunch (12:45–1:45)
Session 3 (1:45–4:15)
Courses, Curricula, and Interdisciplinary Programs

Chair: Jim Frew, University of California, Santa Barbara

Speakers:
Bill Howe, University of Washington
Peter Fox, Rensselaer Polytechnic Institute
Joshua Bloom, University of California, Berkeley

Break (4:15–4:30)
Discussion (4:30–5:00)
Session 4 (8:30–11:00)  
Shared Resources

Chair: Deepak Agarwal, LinkedIn

Speakers:

Christoper Re, Stanford University  
Bill Cleveland, Purdue University

Panel Discussion:

Roger Barga, Microsoft Corporation  
Ron Brachman, Yahoo!  
Mark Ryland, Amazon Corporation
Session 5: 11:15–1:00
Panel Discussion: Workshop Lessons

Chair: Rob Kass, Carnegie Mellon University

Adjourn (1:00)