UNIVERSITY OF CHICAGO
DEPARTMENT OF COMPUTER SCIENCE
PRESENTS:

"Enabling Data Science for the Majority"

Aditya Parameswaran
University of Illinois at Urbana

Abstract:
Despite great strides in the generation, collection, storage, and processing of data at scale, data science is either out of reach, or, at the very least, extremely inconvenient for the majority of the population. The driving goal of our research is to help individuals and teams--regardless of programming or analysis ability--manage, analyze, make sense of, and draw insights from large datasets. Over the past three years, we've been building (with collaborators at MIT, UMD, Illinois, and UChicago) a number of tools that empower individuals and teams to perform data science more effectively and effortlessly.

These tools span the spectrum of data science or analysis needs, all the way from extracting data into a form amenable to analysis, to exploration and derivation of insights, to recording and sharing of datasets and insights. These tools include DataSpread, a "big data" spreadsheet tool that combines the benefits of spreadsheets and databases; ZenVisage, a visual exploration tool that facilitates the rapid discovery of trends or patterns; and Orpheus, a collaborative data analytics tool that enables the efficient recording and retrieval of dataset versions at various stages of analysis. All of our tools are open-source, and have witnessed usage in fields such as engineering, genomics, astrophysics, marketing analytics, and ad analytics.

In my talk, I will argue that the development of such tools needs to (i) crucially minimize the effort, time, and complexity on the part of the human analyst, (ii) draw on techniques from multiple disciplines--databases, data mining, and interaction, and (iii) revisit the design of all layers of the software stack, from interfaces and interactions, to query languages and APIs, to query execution and optimization, and finally to representation, storage and indexing. Drawing on examples from the tools that we've developed, I will describe how a first-principles approach can lead to solutions that yield practical benefits in terms of scalability, interactivity, usability, and accuracy, while also providing theoretical guarantees. I will finally outline a future research agenda for tool development to truly democratize data science, with the ultimate goal of allowing everyone to tap into the hidden potential in their datasets at scale.

Bio:
Aditya Parameswaran is an Assistant Professor in Computer Science at the University of Illinois (UIUC). He spent a year as a PostDoc at MIT CSAIL following his PhD at Stanford University (2013), before starting at Illinois in August 2014. He develops systems and algorithms for interactive or "human-in-the-loop" data analytics, synthesizing techniques from database systems, data mining, and human computation. Aditya received the NSF CAREER Award (2017), the IEEE TCDE Early Career Award (2017), the C. W. Gear Junior Faculty Award from Illinois (2017), multiple "best" Doctoral Dissertation Awards (from SIGMOD, SIGKDD, and Stanford in 2014), an "Excellent" Instructor award from Illinois (2016), a Google Faculty award (2015), and five best-of-conference citations (from conferences like VLDB, KDD, and ICDE, 2010-17). He is an associate editor of SIGMOD Record and serves on the steering committee of the HILDA (Human-In-the-Loop Data Analytics) Workshop. His research group is supported with funding from the NSF, the NIH, Adobe, Toyota, the Siebel Energy Institute, and Google.

Thursday, January 25, 2018
3:30 pm
Ryerson 251
Host: Aaron Elmore