Teaching Big Data to MSBA Students

2018 IT Teaching Workshop
Agenda

- Coverage of big data in MSBA curriculum
  - Elective course, “Big Data Management”
  - Some big data topics in core courses
  - Co-curricular instruction
- Resources and Infrastructure
- Synergy with MBA and UG programs
What is Big Data

Volume

Scale of Data

It’s estimated that 2.5 quintillion bytes (2.5 trillion gigabytes) of data are created each day.

Most companies in the U.S. have at least 100 terabytes of data stored.

6 billion people have cell phones.

The New York Stock Exchange captures 1 TB of trade information during each trading session.

Modern cars have close to 100 sensors that monitor items such as tire pressure and the temperature.

Velocity

Analysis of Streaming Data

By 2015, it is projected there will be 18.9 billion network connections, or almost 2.5 connections per person on earth.

By 2017, 4.4 million IT jobs will be created specifically to support big data with 2.6 million in the United States.

Variety

Different Forms of Data

30 billion pieces of content are shared on Facebook every month.

4 billion hours of video are watched on YouTube each month.

420 million wearable, wireless health monitors.

Veracity

Uncertainty of Data

27% of respondents have concerns of how much of their data may not be accurate.

UCI Paul Merage School of Business
Big Data Management Elective

- Database and SQL refresher
- Big Data frameworks (HDFS, MapReduce, etc.)
- Hadoop
- Spark
- Non-RDBMS (Key-value stores, NoSQL, Cassandra)
- Document stores (Twitter API, Web Crawling, IoT)
- Cloud database systems (Aurora, RedShift)

→ Co-taught by CS faculty from UC Irvine and UC Riverside
→ Pre-requisites: Knowledge of SQL and Python
Big Data in Customer & Social Analytics Course

- Data from Yelp Challenge
  - Second half of 2017
  - 7.5 GB, excluding photos
  - 11 tables (data on businesses, reviews, check-ins, Friends, Elites)
  - MySQL and Mongo DB formats
- Goal is to take R Studio from laptop to AWS
- R Studio Server implemented in EC2 instance --- one server per student group
- Use DBI and RMySQL packages on R Studio server
- Assignment includes data visualization, Poisson and logistic regressions
Co-Curricular Coverage

- Workshops on using AWS
- Video tutorials on big data topics
- Capstone projects (several companies provided data on AWS)
Resources

- Instructors and TA’s from CS Department
- Big Data consultant employed part time (15 hours per week)
- Joint AWS account for the MSBA program, with educational grant from Amazon
  - Sub-accounts for each student team
- Third-party online labs (AWS, QwikLab)
Topics for Discussion

- How much big data coverage?
- Outsource to CS, or hire faculty with CS background?
- Support staff, full-time or part-time?
- Environment?
  - AWS, Azure, on-premise servers?
- Synergy with MBA and UG programs