## SCHOOL OF COMPUTING

# BIG DATA PROGRAM

THE UNIVERSITY OF UTAH

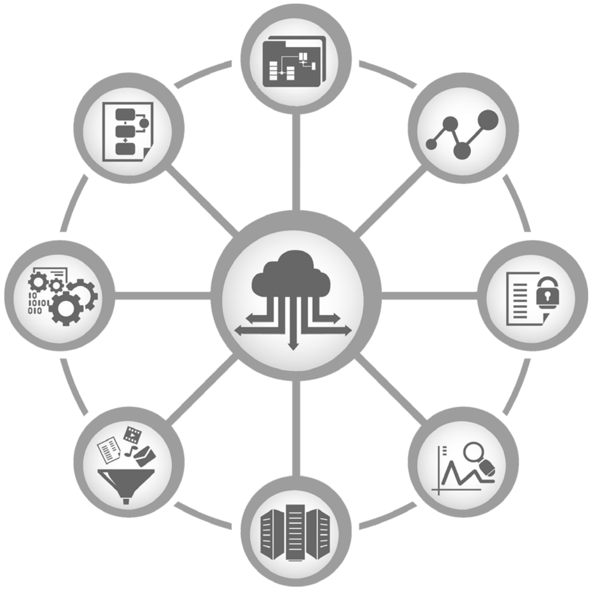
# BIG DATA PROGRAM

In the big data program in the School of Computing at the University of Utah, students will take classes from tenure-track professors actively developing the new techniques for these emerging challenges of big data. And students will learn by doing. They will work on real data, building on the techniques they learn in class under the guidance of the professors. The classes are hands on, and project-focused allowing them to interact with modern software tools and data processing techniques.

Moreover, students will see not just how to use these tools, but will learn the computer science behind how they work. To understand how emerging techniques are best used and can scale, student will learn the fundamental analysis behind them and will code some core aspects. To understand which machine learning algorithm to use or which data mining technique to apply, students will study how they work, and think through the design choices. Knowledge of these deep concepts is essential when faced with real data, and one needs to confront situations when the standard approach is too slow or the answers do not seem to make sense. Students will harness new approaches in visual analytics to quickly investigate, interpret, and present these enormous complex data.

Graduates of this program will have acquired the broad and requisite technical skills and experience to be lead analysts at data rich companies, or to join top software companies at the forefront of this revolution.

In addition, student will have the opportunities to interact with the active research going on in the department through a wide variety of regular research seminars, discussions, and colloquium. Active students will be exposed to new techniques and tools, before they reach the cutting edge!



# OVERVIEW

Big Data is impacting many areas of science, engineering, and industry; from analyzing troves of weather data to modeling traffic patterns to processing millions of online customers, it is the enormous data which is creating new opportunities and challenges.

To tackle these challenges, one must have the training to store, manage, process and analyze data at these scales. But the challenges are beyond scale alone, the complexity of the data requires new powerful analytical techniques. Finally, it is crucial to have skills in communicating and interpreting the results of this analysis. A person trained in all of these skills is a *big data scientist*.

# ACCESSABILITY

The core courses associated with the program are usually scheduled in consecutive blocks and/or in the late afternoons and evenings two days a week.

The classes will be video recorded, with videos made available soon after each class. These recordings can be live-streamed based on demand. This will help professional students on a busy schedule from falling behind if they miss a class, and will allow all students to easily review the material as presented in the lecture. We still require students to occasionally directly interact in class for exams and project presentations.

# CORE CLASSES

### *Advanced Algorithms*

The formal study of the scalable design of algorithms. As datasize grows this analysis and way of thinking becomes essential to ensure the underlying processing engine can keep up. After reviewing basic concepts of divide and conquer, greedy algorithms, and dynamic programming, will also cover advanced and modern approaches including flows, linear programming, hardness reductions, randomization, and approximation.

### *Database Systems*

How large data is stored and managed, focusing on how the back ends of classic and emerging database engines work internally. The 5000-level version prepares students in the theory of how to represent data relationally and form logical queries, how to implement these concepts in SQL, and how to build an advanced application on top of these concepts. The 6000-level version covers how the internals of a full database engine are built by actually building part of such a system in C.

### *Data Mining*

The study of efficiently finding structures and patterns in large data sets, including similarity search, clustering, regression, dimensionality reduction, and graph analysis. With focus on converting from messy and noisy data sets to structured abstract ones, applying scalable and probabilistic analysis on abstract data sets, and the formal modeling and understanding trade-offs.

### *Machine Learning*

Techniques for developing computer programs that can acquire new knowledge automatically or adapt their behavior over time. Topics include several algorithms for supervised and unsupervised learning, decision trees, online learning, linear classifiers, empirical risk minimization, computational learning theory, ensemble methods, Bayesian methods, clustering and dimensionality reduction.

### *Visualization*

The principles, methods, and techniques for effective visual analysis of data, including techniques for both spatial (eg. gridded data from simulations and scanning devices) and non-spatial data (eg. graphs, text, high-dimensional tabular data). Begins with an overview of principles from perception and design, continues with a framework for discussing, critiquing, and analyzing visualizations, and then focuses on visualization techniques and methods for a broad range of data types.

*Upon completion of each class, students should be able to read, understand, and implement many research papers in the area.*

# ELECTIVES

## Algorithmics

Computational Geometry
Computational Topology
Models of Computation for Massive Data

## Analytics

Advanced Scientific Computing
Artificial Intelligence
Image Processing
Natural Language Processing
Probabilistic Modeling

## Management

Advanced Computer Networks
High Performance Parallel Computing
Network Security
Parallel Programming for GPUs/Many Cores/Multi-Cores

| | |
|---|---|
| **Big Data Certificate** | **CORE CLASSES** |
| **Big Data Masters**<br>*(MS in Computing)* | **CORE CLASSES**<br>+ 5 electives (or)<br>+ 3 electives and masters project or thesis |
| **Big Data PhD**<br>*(PhD in Computing)* | **CORE CLASSES**<br>+ 3 electives and PhD thesis |

# HOW TO APPLY

**For MS and PhD degrees:**

- Following instructions on the SoC Graduate Admissions page
  *http://www.cs.utah.edu/graduate/admissions/*
- Apply for the MS or PhD in Computing
  Select "Data Management and Analysis" as an emphasis

**For Certificate in Big Data:**

- If enrolled in another graduate degree program: register for a required classes
- Or apply as a Non-Degree Seeking Student:
  *http://www.cs.utah.edu/bigdata/apply*
- If one switches to MS or PhD degree, up to 9 credit hours (3 classes) can transfer



# WHY UTAH?

The Salt Lake Valley and surrounding area, known as the Silicon Slopes (http://siliconslopes.com) is home to a surprisingly large number of technology companies (including Adobe, Overstock.com, and Ancestory.com), is consistently ranked as a top location for entrepreneurs, and is the home of many data centers, including for Ebay, C7 Data Centers, and the new massive NSA datacenter. The University of Utah is a world-class university with a rich history in computing, and big data is an emerging strength of the School of Computing. Moreover, Utah has unparalleled outdoors activities, with the campus nestled in mountains, located 10 minutes from downtown Salt Lake, 45 minutes from some of the best skiing on earth, and within 5 hours of at least 7 national parks including Yellowstone, Arches, and Zion.

**SCHOOL OF COMPUTING**
UNIVERSITY OF UTAH
50 S. Central Campus Drive RM 3190
Salt Lake City, UT 84112

## www.cs.utah.edu/bigdata