

# DATA SCIENCE ENGINEERING PROGRAM

# BIG DATA

**Area Director**  
**Prof. John Cho**

## **BIG DATA ENGINEERING PROGRAM** (John Cho)

The exponential growth of data generated by machines and humans present unprecedented challenges and opportunities. From the analysis of this “big data”, businesses can learn key insights about their customers to make informed business decisions. Scientists can discover previously unknown patterns hidden deep inside the mountains of data. In this program, students will learn key techniques used to design and build big data systems and gain familiarity with data-mining and machine-learning techniques that are the foundations behind successful information search, predictive analysis, smart personalization, and many other technology-based solutions to important problems in business and science.

## **CS143 Database Systems** (John Cho)

Information systems and database systems in enterprises. File organization and secondary storage structures. Relational model and relational database systems. Network, hierarchical, and other models. Query languages. Database design principles. Transactions, concurrency, and recovery. Integrity and authorization.

## **CS145 Introduction to Data Mining** (Wei Wang)

Introductory survey of data mining (process of automatic discovery of patterns, changes, associations, and anomalies in massive databases), knowledge engineering, and wide spectrum of data mining application areas such as bioinformatics, e-commerce, environmental studies, financial markets, multimedia data processing, network monitoring, and social service analysis.

## **EE131A Probability and Statistics** (Lara Dolecek)

Introduction to basic concepts of probability, including random variables and vectors, distributions and densities, moments, characteristic functions, and limit theorems. Applications to communication, control, and signal processing. Introduction to computer simulation and generation of random events.

## **CS246 Web Information Systems** (John Cho)

Designed for graduate students. Scale of Web data requires novel algorithms and principles for their management and retrieval. Study of Web characteristics and new management techniques needed to build computer systems suitable for Web environment. Topics include Web measuring techniques, large-scale data mining algorithms, efficient page refresh techniques, Web-search ranking algorithms, and query processing techniques on independent data sources.

**CS260**  
**Machine Learning Algorithms**  
(*Ameet Talwalkar*)

Problems of identifying patterns in data. Machine learning allows computers to learn potentially complex patterns from data and to make decisions based on these patterns. Introduction to fundamentals of this discipline to provide both conceptual grounding and practical experience with several learning algorithms. Techniques and examples used in areas such as healthcare, financial systems, commerce, and social networking

**CS249**  
**Big Data Systems**  
(*Tyson Condie*)

This course is designed to introduce students to the foundations of Big Data Systems, focusing on subjects such as distributed file-systems, parallel and distributed database design. It is designed for students who have taken CS111 and CS143 (or equivalent). Classes will consist of lectures and discussions based on readings from the parallel and distributed database literature.

**CS262A**  
**Learning and Reasoning with Bayesian Networks**  
(*Adnan Darwiche*)

An in-depth exposition of knowledge representation and reasoning under uncertainty using the framework of Bayesian networks. Both theoretical underpinnings and practical considerations will be covered. Special emphasis on constructing Bayesian networks from system design, learning Bayesian networks from data, inference algorithms, and applications to reasoning about systems (e.g., diagnosis, monitoring and reliability). More advanced topics may include embedded inference algorithms, sensitivity analysis and reasoning about text and images.

**CS240A**  
**Databases and Knowledge Bases**  
(*Carlo Zaniolo*)

Theoretical and technological foundation of Intelligent Database Systems, that merge database technology, knowledge-based systems, and advanced programming environments. Rule-based knowledge representation, spatio-temporal reasoning, and logic-based declarative querying/programming are salient features of this technology. Other topics include object-relational systems and data mining techniques.

**CS249**  
**Big Data Analytics**  
(*Wei Wang*)

This course will provide a survey of fundamental concepts in data mining and will provide in-depth coverage of advanced topics in big data analytics from a collection of research papers published in current years. This course is designed to expose students to cutting edge research and new frontiers in this fast growing field.

**Program Contact Information**  
7440 Boelter Hall Box 951601  
Los Angeles, CA 90095-1601  
Phone (310) 825-6542  
Fax (310) 825-3081  
Email: [admissions@seas.ucla.edu](mailto:admissions@seas.ucla.edu)

**For more information, please contact:**  
Dr. John Cho  
Co-Area Director  
(310) 794-5444  
[cho@cs.ucla.edu](mailto:cho@cs.ucla.edu)

Prof. Jenn-Ming Yang  
Associate Dean, International Initiatives and Online Program  
(310) 825-2758  
[jyang@seas.ucla.edu](mailto:jyang@seas.ucla.edu)

**Program Cost**  
The cost for this 9-course program is \$34,650. Payment of fees is on a course-by-course basis, \$3850 per course.

\*The faculty listed is subject to availability, and substitute instructors with equal caliber will be provided.

Numerical background data image by DARPA.