UCI Extension
Information Technologies Programs

Big Data Specialized Studies

Accelerate Your Career
extension.uci.edu/bigdata

Offered in partnership with

University of California, Irvine
Improve Your Career Options with a Professional Certificate

University of California, Irvine

Extension’s professional certificate and specialized studies programs help you increase or enhance your current skills or prepare for a new career. Courses are highly practical and instructors are qualified leaders in their field. Convenient online courses make it easy to learn on your own time, in your own way. A certificate bearing the UC seal signifies a well-known, uncompromising standard of excellence.
Big Data Specialized Studies

Every two days there is as much new data generated as was created in the first 2000 years of our history. The “Internet of Things” and related trends will further increase the amount of data generated every day. Successful businesses throughout the world have become increasingly reliant on a rapidly expanding amount of data to remain competitive. This “unstructured” data, which often consists of web activity, video and audio recordings, photographs, e-mails, and tweets, can be combined with “structured” data, such as financial specifics of a customer order, to provide the organization with the critical intelligence and insight required to succeed in a highly competitive global marketplace. A drop in the price of hardware required to deal with these large data sets coupled with improved functionality of a wide variety of big data software tools has created a perfect storm with skyrocketing demand for individuals with the skills to manipulate and analyze these data sets. The Big Data certificate program provides individuals with the background needed to effectively collect and manage big data and perform data-driven discovery, prediction, and extract value and competitive intelligence for their organizations.

Who Should Enroll

This program is designed for individuals within a wide range of professional or academic backgrounds who desire to learn how to better collect, store, analyze, and act on large data sets to drive business results.

Certificate Eligibility and Requirements

Students must complete 11 credit units (4 required and 7 elective credit units) with a grade of “C” or higher in each course. Courses must be completed within 5 years.

To receive your certificate, please complete all program requirements then submit a Specialized Studies Certificate Request form (along with a non-refundable fee of $35) online or download the form at extension.uci.edu/forms.

Students not pursuing a specialized studies award are welcome to take as many individual courses as they wish.

Program Benefits

- Turn existing large, structured and un-structured data sets into cohesive information and extract valuable insight to solve business issues.
- Use industry standard analytic tools including, Hadoop, Teradata, Knime, Aster, IBM-Netiza, HP-Veritca, Statistica, R, Python and Crystal Ball
- Develop the strategies and skills needed to effectively collect and manage big data and perform data-driven discovery, prediction, and prescription
- Develop big data architectural strategies for your organization
- Improve business processes efficiency and customer satisfaction
- Understand data architecture software including Hadoop, Teradata and Aster, and related tools (Java, SQL and MRSQL)
- Integrate powerful and traditionally untapped sources of unstructured big data including social media data and other web generated information

Program Fees

Individual course fees along with the costs of textbooks and other supplementary materials are subject to change without notice. For budgeting purposes, you may use the following estimates:

<table>
<thead>
<tr>
<th>Cost Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course fees (11 units)</td>
<td>$3,475</td>
</tr>
<tr>
<td>Candidacy fee</td>
<td>$35</td>
</tr>
<tr>
<td>Textbooks</td>
<td>$325</td>
</tr>
<tr>
<td><strong>Total Estimated Cost</strong></td>
<td><strong>$3,835</strong></td>
</tr>
</tbody>
</table>

For more information:
Julie Pai
julie.pai@uci.edu
(949) 824-6333
Curriculum

Required Courses (4 units)

**Introduction to Big Data**  
I&C SCI X425.80 (2 units)  
Web 2.0 has changed the way we conduct business. It affects consumer interactions, information sharing, success measurement in terms of business revenue and customer wallet share, and brand management. Most importantly, it has created a revenue channel like none other. Personalization of products and services by enterprises have increased data volumes and the velocity of data production while creating a variety of data formats. We are now well and truly in the world of big data. The key value of this data is the vast amount of intelligence that can be found when data is modeled with geographic and demographic information. This introductory course will help students navigate through the complex layers of big data and data warehousing while providing information on how to effectively think about using these technologies and architectures to design the next-generation data warehouse. Concepts covered include an introduction to big data, discussion of big data processing architectures, explanation of the integration of big data and data warehouses, and fundamentals of big data analytics. Examples will be used to step students through an implementation of a big data solution.

**Unified Big Data Architecture**  
I&C SCI X425.82 (2 units)  
The volume, velocity, and variety of typical big data initiatives coupled with the fact that the data is usually stored in many disparate locations within the organization make it difficult to extract information and business insight. In this class students will learn a systematic approach and Unified Big Data Architecture that will help make this problem more tractable. Learn to first identify and evaluate the type of data and analysis to be used, the sources of that data and any related data processing techniques. Next, learn a dimensions-based approach to assess the viability of a big data solution. You will then develop a set of logical layers to the solution that define the critical components of a big data solution. This logical layering begins with the acquisition of the data from various data sources and build out a logical layer for each step in the process with a continued focus on the end results of deriving business insight. The course will also cover many of the most common and recurring big data problems and solutions and the use of patterns to enable business and technical users to apply a structured approach to establishing the scope and defining the high level solution for a big data problem.

Elective Courses (7 units)

**Big Data Analytics**  
I&C SCI X425.70 (2 units)  
Very large databases of more than 100 petabytes are becoming more common with data storage and reporting technology becoming faster and easier. Successful users of big data, like Wal*Mart, Amazon, Google and others harness data in ways simply impossible 15 years ago. This course will begin with a review of some case histories of big data use. Basic elements of successful big data implementation will be covered as well as a review of commercially available tools and technology. A focus will be placed on value opportunities in online marketing, search optimization and site performance modeling. Other topics covered in the course will include database architecture as related to knowledge discovery and evaluation of the relative merits of normalized relational databases, report-oriented dimensional designs, and unstructured SQL-free systems. All students across the business, technical, and analytic spectrum are welcome.

**Big Data Tools & Applications**  
I&C SCI X425.81 (2 units)  
Learn how to leverage big data using industry standard tools and by developing analytical models to solve real commercial problems. Discover how to extract meaningful insights from massive volumes of data residing in advanced big data repositories using tools such as Hadoop and Teradata. Understand how to identify key business goals and define an efficient and structured solution. Through a series of case studies you will also learn how to select the best analytics solution and use a variety of industry standard tools such Hadoop, Teradata, IBM Netiza, HP Vertica, R, WEKA, Knime and Python. Key real-world issues such as working in cross-functional, geographically dispersed teams of professionals and presenting highly technical results to business-oriented audiences will also be discussed.

extension.uci.edu/bigdata
Hadoop: In Theory and Practice  
I&C SCI X425.18 (3 units)  
Today, organizations in every industry are being showered with imposing quantities of new information. Along with traditional sources, many more data channels and categories now exist. Collectively, these vastly larger information volumes and new assets are known as big data. Enterprises are using technologies such as MapReduce and Hadoop to extract value from big data. This course provides an in-depth overview of Hadoop and MapReduce, the cornerstones of big data processing. To crystallize the concepts behind Hadoop and MapReduce, you will work through a series of short/ focused exercises; you will configure and install a Hadoop cluster, write basic MapReduce programs, gain familiarity with advanced MapReduce programming practices, and utilize interfaces such as Pig and Hive to interact with Hadoop. You will also learn about real-world situations were MapReduce techniques can be used.

Effective Data Preparation  
I&C SCI X425.62 (2 units)  
Broadly speaking, data preparation for data mining consists of three (3) elements: 1. Data Mining Process delineation (understanding the overall process) 2. Data Understanding (data cannot be properly prepared without first understanding it) 3. Data Pre-processing (transforming data into a form compatible with data mining) This intensive hands-on course gives students the skills necessary to extract stored data elements, understand what they mean in the company, transform their formats and derive new relationships among them to produce a dataset suitable for analytical modeling. Learn how to produce a fully processed data set compatible for building powerful predictive models that can be deployed to increase business profitability. Prerequisite: I&C SCI X425.61  

Teradata: Foundation and Principles  
I&C SCI X425.83 (2 units)  
This course will provide a basic understanding of how Teradata works. Course topics include Teradata architecture, AMPs, PEs, BYNET, data modeling, primary key, primary index, data layout, hashing, and Row ID. This course also covers a base understanding of features such as RAID, secondary indexes, spool files, partitioning, data protection, and Teradata utilities.

Introduction to Data Science  
I&C SCI X426.60 (3 units)  
Learn the data design, management and manipulation tools and processes commonly used by data scientists and understand how to apply this knowledge to yield measurable business value. Gain an overview of the basic techniques of data science, including data analysis, statistical modeling, data engineering, relational databases, SQL and NoSQL, manipulation of data at scale (big data), algorithms for data mining, data quality, remediation and consistency operations.

Corporate Training  
UCI Extension’s Corporate Training specialists can deliver this program or a customized one that fits your company’s specific needs. Visit extension.uci.edu/corporate or call (949) 824-1847 for information.
Academic Management

Dave Dimas, Ph.D., Director, Engineering, Sciences and Information Technologies

Big Data
Specialized Studies

UCI Extension

Julie Pai  ●  (949) 824-6333  ●  julie.pai@uci.edu

extension.uci.edu/bigdata