Khoury College of Computer Sciences and the Department of Electrical and Computer Engineering (ECE) jointly offer a new interdisciplinary Master of Science program in data science. This program is designed to give students a comprehensive framework for reasoning about data. Students will engage in extensive course work intended to develop depth in data collection, storage, retrieval, manipulation, visualization, modeling and interpretation. Students will also be able to choose elective courses from a variety of offerings in Khoury, the College of Engineering (COE), and throughout the campus to explore areas that generate data, or specialized data science applications. Successful program graduates will be well positioned to attain data scientist and data engineer positions in a fast-growing field or to progress into doctoral degrees in related disciplines.

**Prerequisite Courses**
The Master of Science in Data Science curriculum is tailored toward technically or mathematically trained students. To ensure that all students have the foundation necessary to be successful in this program, each incoming student, regardless of his or her background, must take two introductory courses in programming and mathematics. The introductory courses are not counted as credit towards the degree, but are included in the student’s cumulative GPA. The two introductory courses are Introduction to Programming for Data Science (DS 5010) and Introduction to Linear Algebra and Probability for Data Science (DS 5020)

If a student believes that he or she has sufficient academic or professional experience in these areas, they may opt to take two placement exams administered one week prior to the beginning of the semester. The two exams cover fundamentals of computer science and programming skills and basic statistics, probability, and linear algebra. If a student scores a B or higher, he or she is waived out of the corresponding course.

**Learning Outcomes**
Students who successfully complete the MS degree will be able to:

- Collect data from numerous sources (databases, files, XML, JSON, CSV, and Web APIs) and integrate them into a form in which the data is fit for analysis
- Use R and Python to explore data, produce summary statistics, perform statistical analyses; use standard data mining and machine-learning models for effective analysis
- Select, plan, and implement storage, search, and retrieval components of large-scale structure and unstructured repositories
- Retrieve data for analysis, which requires knowledge of standard retrieval mechanisms such as SQL and XPath, but also retrieval of unstructured information such as text, image, and a variety of alternate formats
- Manage, process, analyze, and visualize data at scale. This outcome allows students to handle data where the conventional information technology fail
- Match the methodological principles and limitations of machine learning and data mining methods to specific applied problems and communicate the applicability and the advantages/disadvantages of the methods in the specific problem to nondata experts
- Carry out the full data analysis workflow, including unsupervised class discovery, supervised class comparison, and supervised class prediction; Summarize, interpret, and communicate the analysis of results
- Organize visualization of data for analysis, understanding, and communication; choose appropriate visualization method for a given data type using effective design and human perception principle
- Develop methods for modeling, analyzing, and reasoning about data arising in one or more application domains such as social science, health informatics, web and social media, climate informatics, urban informatics, geographical information systems, business analytics, bioinformatics, complex networks, public health, and game design

**Program Requirements**
Complete all courses and requirements listed below unless otherwise indicated.

Students should refer to the course numbering table for graduate course leveling (http://catalog.northeastern.edu/graduate/academic-policies-procedures/records-transcripts/).

**Core Requirements**
A cumulative GPA of 3.000 or higher is required in the following core courses:

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS 5800</td>
<td>Algorithms</td>
<td>4</td>
</tr>
<tr>
<td>EECE 7205</td>
<td>Fundamentals of Computer Engineering</td>
<td></td>
</tr>
<tr>
<td>DS 5110</td>
<td>Introduction to Data Management and Processing</td>
<td>4</td>
</tr>
<tr>
<td>DS 5220</td>
<td>Supervised Machine Learning and Learning Theory</td>
<td>4</td>
</tr>
<tr>
<td>DS 5230</td>
<td>Unsupervised Machine Learning and Data Mining</td>
<td>4</td>
</tr>
<tr>
<td>DS 5500</td>
<td>Information Visualization: Applications in Data Science</td>
<td>4</td>
</tr>
</tbody>
</table>

**Electives**
Complete 12 semester hours from the following:

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS 5100</td>
<td>Foundations of Artificial Intelligence</td>
<td></td>
</tr>
<tr>
<td>CS 5200</td>
<td>Database Management Systems</td>
<td></td>
</tr>
<tr>
<td>CS 5340</td>
<td>Computer/Human Interaction</td>
<td></td>
</tr>
<tr>
<td>CS 5610</td>
<td>Web Development</td>
<td></td>
</tr>
<tr>
<td>CS 6120</td>
<td>Natural Language Processing</td>
<td></td>
</tr>
<tr>
<td>CS 6200</td>
<td>Information Retrieval</td>
<td></td>
</tr>
<tr>
<td>CS 6240</td>
<td>Large-Scale Parallel Data Processing</td>
<td></td>
</tr>
<tr>
<td>CS 6350</td>
<td>Empirical Research Methods</td>
<td></td>
</tr>
<tr>
<td>CS 6620</td>
<td>Fundamentals of Cloud Computing</td>
<td></td>
</tr>
<tr>
<td>CS 7180</td>
<td>Special Topics in Artificial Intelligence</td>
<td></td>
</tr>
</tbody>
</table>
CS 7280  Special Topics in Database Management

**College of Engineering**
- CIVE 7100  Time Series and Geospatial Data Sciences
- CIVE 7388  Special Topics in Civil Engineering
- EECE 5639  Computer Vision
- EECE 5640  High-Performance Computing
- EECE 7337  Information Theory
- EECE 7370  Advanced Computer Vision
- EECE 7397  Advanced Machine Learning
- IE 7280  Statistical Methods in Engineering

**College of Social Sciences and Humanities**
- ECON 5140  Applied Econometrics
- PPUA 5261  Dynamic Modeling for Environmental Decision Making
- PPUA 5262  Big Data for Cities
- PPUA 5263  Geographic Information Systems for Urban and Regional Policy
- PPUA 5266  Urban Theory and Science
- PPUA 7237  Advanced Spatial Analysis of Urban Systems
- POLS 7200  Perspectives on Social Science Inquiry
- POLS 7201  Research Design
- POLS 7202  Quantitative Techniques

**D’Amore-McKim School of Business**
- BUSN 6320  Business Analytics Fundamentals

**College of Science**
- PHYS 5116  Complex Networks and Applications
- PHYS 7305  Statistical Physics
- PHYS 7321  Computational Physics
- PHYS 7331  Network Science Data

**Bouvé College of Health Sciences**
- PTHH 5202  Introduction to Epidemiology
- PTHH 5210  Biostatistics in Public Health
- PTHH 6224  Social Epidemiology

**College of Arts, Media and Design**
- GSND 5110  Game Design and Analysis
- GSND 6350  Data-Driven Player Modeling

**Program Credit/GPA Requirements**
32 total semester hours required
Minimum 3.000 GPA required

Students taking electives worth less than 4 credits (i.e., Bouvé, CSSH courses) should enroll for an accompanying data science project course in the same semester to bring the cumulative credits to 4. In order to earn this additional credit, students are expected to work with faculty to design an additional project in line with the curricular aims of their chosen elective and the data science core learning outcomes.