

# Computational Data Science (M.S.)

---

## About The Program:

The M.S. in Computational Data Science is designed for students interested in developing expertise in data science with a specialization in computational analytics. The goal is to enable students to analyze large quantities of data to discover new knowledge and facilitate decision making. To accomplish this, the program provides students with a strong foundation in big data management and analysis; algorithmic, computational, and statistical thinking; and an understanding of computer systems. Data science is a multifaceted, interdisciplinary discipline that employs techniques and theories drawn from the broad areas of computer and information science, mathematics, and statistics, and applies them to a wide range of data-rich domains such as biomedical science, business, education, engineering, geoscience, physical science, and social science.

**Career Options:** Graduates gain the necessary skills to find positions in the data science field requiring a mix of data analysis skills, the ability to deal with large quantities of data, and a strong foundation in computer science. Graduates are also prepared to undertake doctoral studies, either to deepen their overall data science expertise or learn how to better use their analytics skills in a particular data-rich domain.

**Prerequisites for Admission:** One year programming and data structures using C++ or Java. One year theoretical calculus. Applicants without a bachelor's degree in a related discipline will need additional prerequisites.

**Areas of Specialization:** Research interests of faculty include:

- Analysis of algorithms
  - Artificial intelligence
  - Communication and networks
  - Computer architecture
  - Data analytics
  - Digital forensics
  - Flexible and intelligent manufacturing systems
  - Graphics
  - High-performance computing
  - Information security and assurance
  - Intelligent CAI systems
  - Management information and database systems
  - Natural language processing
  - Network security
  - Parallel and distributive processing and operating systems
  - Programming languages
  - Sensory and image processing
  - Software engineering
  - Theory of automata and computation
  - Wired and wireless networks
-

## Requirements of Programs:

- **Total Credit Hours:** 30
- **Culminating Events:** Students complete a capstone project in data science as the culminating event. [CIS 9995](#) is taken for 3 credits under the close supervision of CIS Graduate Faculty.

## Core Courses

**Programming Techniques** – A more formalized view of data structures. Stacks, trees, tables, lists, multilinked structures, strings and files are considered. These are viewed in terms of their general usefulness in the construction of algorithms and in their efficient implementation. Both theoretical results and programming techniques will be stressed.

**Design and Analysis of Algorithms** – The course objective is to provide students with an understanding of the principles and techniques used in the design and analysis of efficient algorithms. The main topics cover Greedy Algorithms, Divide and Conquer, Dynamic Programming, Network Flow, and Approximation Algorithms. Theoretical results related to NP-completeness will also be discussed.

**Machine Learning** – The goal of the field of machine learning is to build computer systems that learn from experience and are able to adapt to their environments. This introductory machine learning course will present modern machine learning algorithms for supervised and unsupervised learning. It will provide the basic intuition behind the algorithms as well as a more formal understanding of how and why they work. Students will learn how to apply machine learning algorithms on a range of real-life problems and how to evaluate their performance.

*Select one course from the following:*

**Operating Systems** – Basic principles of operating systems; multi-tasking systems; control and coordination of tasks; deadlocks; synchronization, mutual exclusion, sharing; memory management, virtual memories, segmentation, paging; protection; file systems; resource management; evaluation and prediction of performance; design and implementation of operating systems in high-level languages.

**Principles of Data Management** – This course covers fundamental concepts in constructing database management systems, including relational query languages, such as SQL and relational algebra, file organizations, storage management, system architectures, query processing, query optimization, transaction management, recovery, and concurrency control. Additional topics may include distributed databases, NoSQL databases and data integration.

**Computer Architecture** – Since 1951, there have been thousands of new computers using a wide range of technologies and having widely varying capabilities. Dramatic changes that have occurred in just over 50 years. After adjusting for inflation, price/performance has improved by almost 100 billion in 55 years, or about 58% per year. Another way to say it is we've seen a factor of 10,000 improvement in cost and a factor of 10,000,000 improvement in performance. This course covers the recent developments in modern computer architectures and the emerging design methods for high performance computing.

## Electives (15 Credits Worth)

*Select at most three Big Data courses from the following:*

**Operating Systems** – Basic principles of operating systems; multi-tasking systems; control and coordination of tasks; deadlocks; synchronization, mutual exclusion, sharing; memory management, virtual memories, segmentation, paging; protection; file systems; resource management; evaluation and prediction of performance; design and implementation of operating systems in high-level languages.

**Principles of Data Management** – This course covers fundamental concepts in constructing database management systems, including relational query languages, such as SQL and relational algebra, file organizations, storage management, system architectures, query processing, query optimization, transaction management, recovery, and concurrency control. Additional topics may include distributed databases, NoSQL databases and data integration.

**Data-Intensive and Cloud Computing** – This course will expose students to recently emerged and fast moving technology of big data and cloud computing. It will cover a spectrum of topics from core techniques in data management and analysis to highly-scalable data processing using parallel database systems. Students will be introduced to big data ecosystems such as Hadoop, Spark, Storm and MapReduce; cloud technologies such as Amazon EC2, Microsoft Azure and Google Cloud; data management tailored to cloud and big data such as No SQL, Google Big Table/Apache HBase, and introductory applications to Big Data and cloud environment. Students will work directly with a selected set of these platforms, compare and contrast their relative strengths and weaknesses, and characterize the problems they are designed to solve. Note: Students may not receive credit for both CIS 5517 and CIS 4517.

**Topics in Computer Science** – Current topics and issues in Computer Sciences are covered. This course is repeatable for credit.

**Computer Architecture** – Since 1951, there have been thousands of new computers using a wide range of technologies and having widely varying capabilities. Dramatic changes that have occurred in just over 50 years. After adjusting for inflation, price/performance has improved by almost 100 billion in 55 years, or about 58% per year. Another way to say it is we've seen a factor of 10,000 improvement in cost and a factor of 10,000,000 improvement in performance. This course covers the recent developments in modern computer architectures and the emerging design methods for high performance computing.

**Emerging Storage Systems and Technologies** – Storage systems are of increasing importance because of ever-growing volume, velocity, and heterogeneity of data produced by a wide variety of computer systems. This course will provide a comprehensive coverage of storage and file systems that underlie bigdata systems with respect to both technological and application related challenges.

**Distributed Systems** – We consider a distributed computer system that consists of multiple autonomous processors that do not share primary memory but cooperate by sending messages over a communication network. Discussion of special problems related to distributed control such as election and mutual exclusion, routing, data management Byzantine agreement, and deadlock handling.

**Advanced Topics in Data Base Systems** – Survey of recent developments in database systems with an emphasis on object-oriented databases (OODB's). Prototype and operational OODB systems will be

analyzed. Applications of OODB's to computer-aided software engineering (CASE) environments, integrated application development environments, and geographical information systems.

*Select at most three Data Analysis courses from the following:*

**Knowledge Discovery and Data Mining** – Basic concepts and techniques for the automated extraction of interesting patterns in large databases. Topics covered include: association-rule mining, sequence mining, web and text mining, data warehousing, information filtering, classification and clustering analysis, Bayesian and neural networks, classification and regression trees, hypotheses evaluation, feature extraction, dimensionality reduction, singular value decomposition, data compression and reconstruction, visualization of large data sets, fractals in databases, and indexing methods that support efficient data mining and queries by content. Special emphasis is given in multimedia, business, scientific, and medical databases. Note: Students may not receive credit for both CIS 5523 and CIS 4523.

**Analysis and Modeling of Social and Information Networks** – This course will include methods for analyzing and modeling the following aspects of social networks: the small-world network models, centralized and decentralized social network search algorithms, power-laws and preferential attachment, diffusion and information propagation in social networks, influence maximization in social networks, community detection in social networks, models of network cascades, models of evolving social networks, links and attributes prediction. Note: Students may not receive credit for both CIS 5524 and CIS 4524.

**Neural Computation** – Neural networks provide powerful techniques to model and control nonlinear and complex systems. The course is designed to provide an introduction to this interdisciplinary topic. The course is structured such that students from computer science, engineering, physics, mathematics, statistics, cognitive sciences, and other disciplines learn the main principles of this area as well as have an opportunity to explore promising research topics through hands-on experience with neural network simulators applied to classification and prediction problems ranging from biomedical sciences to finance and business.

**Probabilistic Graph Models** – Probabilistic graphical models are very important machine learning tools for knowledge representation and reasoning under uncertainty. They have been widely used in machine learning and related fields, such as computer vision, natural language processing, data mining, bioinformatics and even computer network research. This course aims to make a comprehensive introduction over the most important theories, algorithms, and applications of probabilistic graphical models, and facilitate the advanced research within the computer & information sciences department and related disciplines outside.

**Text Mining and Language Processing** - This course will cover a broad overview of problems and techniques in text mining and natural language processing. It will also provide in-depth coverage of the latest natural language processing research in selected topics. The in-depth part of the course will focus on the latest research in unsupervised information extraction. This part of the course will cover such techniques as pointwise mutual information, pattern-matching, bootstrapping, Hidden Markov Models, Conditional Random Fields, and language modeling techniques, among others.

**Computer Vision** – The objective of the course is to introduce the theory and application of computer vision. The theoretic part introduces the analysis of visual patterns and the generative models behind

them. The application part uses real world tasks to help students to learn practical computer vision technologies. The course covers the following topics: image formation (camera model, color space, illumination model, etc.), low level vision processing (edge detection, intensity based segmentation, etc.), popular research tools in computer vision, visual matching and registration, visual recognition, image and category classification, scene understanding, object detection, visual tracking, activity and action analysis, and selected advanced topics. In addition to course lectures, the course uses homework assignments, in-class discussions and course projects.

**Artificial Intelligence** – Since today's AI is mostly learning and inference, we will focus on both subjects. This course is designed as the first graduate course in learning and inference. It introduces the basic concepts by focusing on their intuitive understanding and algorithmic perspective. It is intended to prepare beginning graduate students for further graduate courses in machine learning, data mining, robotics, and computer vision. The course project will focus on programming and practical experiments with our high end robot PekeeII from Wany Robotics.

**Independent Study** – Independent research supervised by a Computer and Information Sciences faculty member. NOTE: Only six credits of independent study can be counted towards any MS degree.

---

## Courses:

Click [HERE](#) for more information on the courses below.

- Comp-Based Appl Prog
- Database Design & Programming
- Networking & Operating Systems
- Programming and Data Structure
- System Software and Operating Systems
- Discrete Structure of Computer Science
- Scripting for Sciences and Business
- Data Structures and Objects
- Operating Systems and Architecture
- IT Process Management
- System Development Processes
- Comp Systems Security & Privacy
- Emerging Technologies
- Knowledge Management
- Seminar in Information Science and Technology
- Software Quality Assurance and Testing
- Software Project Management
- Advanced Database Management Systems
- Usability Engineering
- Network Technologies
- Software Engineering
- Introduction to Digital Forensics
- Advanced Seminar in Information Science and Technology
- Ethical Hacking and Intrusion Forensics
- Audit and Compliance for Security and Digital Forensics
- Programming Techniques
- Operating Systems
- Automata and Formal Languages
- Design and Analysis of Algorithms
- Principles of Data Management
- Data-Intensive and Cloud Computing
- Knowledge Discovery and Data Mining
- Analysis and Modeling of Social and Information Networks
- Neural Computation
- Machine Learning
- Data Warehousing, Filtering and Mining
- Probabilistic Graph Models
- Text Mining and Language Processing
- Computer Vision
- Topics in Computer Science
- Artificial Intelligence
- Computer Networking and Communication

- Energy Management in Data Centers and Beyond
- Security in Cyber-Physical Systems
- Ad Hoc Networks
- Network & Information Security
- Wireless Network and Communication
- Computer Architecture
- Emerging Storage Systems and Technologies
- Distributed Systems
- Seminars in Computer and Information Science
- Seminar in Advanced Topics in Computer Science
- Computer Graphics and Image Processing
- User Interface Design and Systems Integration
- Web Applications Development
- Artificial Intelligence, Heuristic Models, and Education
- Advanced Topics in Data Base Systems
- Advanced Networks and Client-Server Computing
- Design and Development of E-Commerce Systems
- Distributed and Parallel Computer Systems
- Master's Research Projects
- Preliminary Examination Preparation
- Capstone Project
- Master's Thesis Research