DATA SCIENCE

The part-time Data Science program balances theory and applications so that you can advance your career long term.

The rigorous curriculum focuses on the fundamentals of computer science, statistics, and applied mathematics, while incorporating real-world examples. By learning from practicing engineers and data scientists, graduates are prepared to succeed in specialized jobs involving everything from the data pipeline and storage to statistical analysis and eliciting the story the data tells.

Program Committee

James C. Spall, Program Co-Chair
Principal Professional Staff
JHU Applied Physics Laboratory
Research Professor, Department of Applied Mathematics and Statistics
JHU Whiting School of Engineering

Benjamin Rodriguez, Program Co-Chair
Principal Professional Staff
JHU Applied Physics Laboratory

Anthony N. Johnson, Program Manager
Senior Professional Staff
JHU Applied Physics Laboratory

Tamás Budavári
Associate Professor, Department of Applied Mathematics and Statistics
Associate Professor, Computer Science
JHU Whiting School of Engineering

Stephyn Butcher
Senior Software Engineer
ThreatGRID, a Division of Cisco

Beryl Castello
Senior Lecturer, Department of Applied Mathematics and Statistics
JHU Whiting School of Engineering

Eleanor Boyle Chlan
Senior Professional Staff (retired)
JHU Applied Physics Laboratory

Joel Coffman
Associate Professor
United States Air Force Academy

Stacy D. Hill
Senior Professional Staff
JHU Applied Physics Laboratory

J. Miller Whisnant
Principal Professional Staff
JHU Applied Physics Laboratory

Brian Wilt
Senior Manager, Data Science
Facebook

Programs

- Data Science, Graduate Certificate (https://e-catalogue.jhu.edu/engineering/engineering-professionals/data-science/data-science-graduate-certificate/)
- Data Science, Master of Science (https://e-catalogue.jhu.edu/engineering/engineering-professionals/data-science/data-science-master/)
- Data Science, Post-Master's Certificate (https://e-catalogue.jhu.edu/engineering/engineering-professionals/data-science/data-science-post-masters-certificate/)

Courses

EN.685.621. Algorithms for Data Science. 3 Credits.
This course offers an in-depth journey through the algorithmic concepts vital for mastering the intricacies of data science. It begins with an intensive examination of algorithm analysis, with a special focus on understanding the runtime complexities essential for addressing real-world data problems. The curriculum encompasses thorough training in data preprocessing, along with foundational knowledge in probability and statistics, equipping students to proficiently clean and interpret data. The course introduces key mathematical transformations such as Eigen decomposition, FFT, DCT, and Wavelets. These tools are crucial for unearthing underlying patterns in data by creating innovative feature spaces. Students will explore a seamless blend of diverse algorithm types, including intelligent algorithms, statistical algorithms, optimization algorithms, graph algorithms, and learning algorithms. This comprehensive approach, enriched with optimization techniques, forms a holistic toolkit for the contemporary Data Scientist. Moving beyond theoretical concepts, the course delves into practical aspects of analysis, visualization, and understanding of complexity classes. Occasional forays into algorithmic proofs enhance the theoretical grounding of students, bridging theory with practical application. The course culminates in modules focused on data modeling and visualization, enabling students to adeptly apply algorithmic techniques to produce insightful and meaningful data representations. Upon completing this course, students will be thoroughly equipped with both practical and theoretical algorithmic strategies, preparing them to confidently address a wide array of challenges in the data science field. Students can only earn credit for one of EN.605.620, EN.605.621, or EN.685.621.

EN.685.648. Data Science. 3 Credits.
This course will cover the core concepts and skills in the interdisciplinary field of data science. These include problem identification and communication, probability, statistical inference, visualization, extract/transform/load (ETL), exploratory data analysis (EDA), linear and logistic regression, model evaluation and various machine learning algorithms such as random forests, k-means clustering, and association rules. The course recognizes that although data science uses machine learning techniques, it is not synonymous with machine learning. The course emphasizes an understanding of both data (through the use of systems theory, probability, and simulation) and algorithms (through the use of synthetic and real data sets). The guiding principles throughout are communication and reproducibility. The course is geared towards giving students direct experience in solving the programming and analytical challenges associated with data science. The assignments weight conceptual (assessments) and practical (labs, problem sets) understanding equally. Prerequisite(s): A working knowledge of Python scripting and SQL is assumed as all assignments are completed in Python.
Prerequisite(s): EN.685.652 Data Engineering Principles and Practice or equivalent course.

EN.685.652. Data Engineering Principles and Practice. 3 Credits.
Data Engineering is the ingestion, transformation, storage and serving of data in ways that enable data scientists or applications to use and derive insights from data. In this course, we will look at various file-based data formats, data collection, data cleansing, data transformation, and data modeling for both relational and NoSQL databases. The course will also cover movement of data into data warehouses and/or data lakes using pipelines and workflow automation. Finally, we will discuss data security, governance, and compliance. The format of this course will be a mix of lectures, hands-on demos, and labs. Upon completing this course, students will have a deeper understanding of what a data engineer does and the various technologies that make up data engineering, along with hands-on experience working with various tools and processes.

EN.685.662. Data Patterns and Representations. 3 Credits.
This course will explore the practical application of data visualization and representation, employing lenses such as personas, to understand the different purposes of visualizations. Data visualization plays a crucial role in the entire data science process, serving multiple purposes such as communicating results and insights in a clear and understandable way, facilitating preliminary data exploration, and analyzing outcomes from physics-based or machine learning models and simulations. The course will introduce various tools and equip students with the knowledge to effectively choose the most suitable tool for a given problem. We will also explore various essential tools for data visualization, including Microsoft Excel, Python plotting libraries like matplotlib and plotly, Python graphical interfacing libraries such as streamlit, and Tableau, among others. As a Data Scientist, you will often need to collaborate in cross-functional teams of varying levels of technical expertise and with role-specific requirements. To prepare you for a well-rounded career in Data Science, the course project will focus on connecting stakeholders with appropriate visualization methods and techniques, the aim of which is to enhance your skills in data visualization to effectively communicate insights to diverse audiences.

EN.685.701. Data Science: Modeling and Analytics. 3 Credits.
This course advances the design of data modeling as it applies to the field of data science while leveraging key concepts from AI, machine learning, and statistics. Data modeling is a combination of various fields which allow the processing of various data types, and representing the data in an expressive way that shows the relationships between data points and intrinsic patterns. The course will show how to identify, design, and implement the modeling process by outlining the framework, determining the appropriate model type, evaluating the model, and representing the outputs in an explainable way. The models used will be based on intelligent algorithms (reasoning, optimization, and pattern recognition), machine learning algorithms (supervised and unsupervised), and statistical methods (descriptive statistics, inferential statistics, multi-variate, and regression). The focus will be developing and applying models using Python-based frameworks to datasets from online resources such as Kaggle, Data.gov, and open-source repositories.

EN.685.795. Capstone Project in Data Science. 3 Credits.
This course permits graduate students in data science to work with a faculty mentor to explore a topic in depth or conduct research in selected areas. Requirements for completion include submission of a significant paper or project. Prerequisite(s): Seven data science graduate courses including two courses numbered 605.7xx or 625.7xx or admission to the post-master's certificate program. Students must also have permission of a faculty mentor, the student's academic advisor, and the program chair.

EN.685.801. Independent Study in Data Science I. 3 Credits.
This course permits graduate students in data science to work with a faculty mentor to explore a topic in depth or conduct research in selected areas. Requirements for completion include submission of a significant paper suitable to be submitted for publication. Prerequisite(s): Seven data science graduate courses including two courses numbered 605.7xx or 625.7xx or admission to the post-master's certificate program. Students must also have permission of a faculty mentor, the student's academic advisor, and the program chair.

EN.685.802. Independent Study in Data Science II. 3 Credits.
Students wishing to take a second independent study in data science should sign up for this course. Prerequisite(s): EN.605.801 Independent Study in Data Science I and permission of a faculty mentor, the student's academic advisor, and the program chair. Course Note(s): Students may not receive credit for both EN.685.795 Capstone Project in Data Science and EN.685.802.