

# PROFESSIONAL CERTIFICATE IN DATA ENGINEERING

Get cutting-edge skills to advance your data engineering career.

Delivered in collaboration with



# Overview

As technology proliferates, data evolves into a true strategic asset, and the demand for data engineers and their specialized expertise grows in tandem. In fact, data engineering was the fastest growing tech occupation in 2020, according to the Dice Tech Job Report. Why? Because before data scientists can glean useful information from the mountains of data possessed by today's organizations, the data must be architected, warehoused, and accessible, and data engineers are responsible for building this infrastructure.

The MIT xPRO Professional Certificate in Data Engineering is an immersive 6-month program that's designed to provide job-ready, in-demand data engineering skills and a competitive edge in the marketplace. Through an exploration of core concepts, tools, techniques, and best practices, participants will learn data engineering essentials, from building effective data architectures and data warehouses to designing data models, streamlining data processing, automating data pipelines, data wrangling, and big data engineering. Participants can also take advantage of personalized feedback, live weekly office hours with course leaders, and the opportunity to develop a GitHub portfolio for potential employers.

MIT xPRO's online learning programs showcase industry-aligned content from world-renowned experts to make learning accessible anytime, anywhere, and solve this challenge for developing technical professionals.

## Price

USD 6,950

## Duration

6 months, online  
15-20 hours per week

# \$121,938

The average annual pay for a data engineer in the United States.

(Source: ZipRecruiter 2021)



"Data engineers build the 'nervous system' of the company. Without it, the company cannot react to changes in the external business environment or within the organization. They build the software and hardware systems that power the company's vision and are masters not just of software, but of hardware, networks, and analytic apps that are changing everyday data."

- John R. Williams, MIT Professor of Information Engineering

# Program Highlights

---



Earn a certificate from MIT xPRO and 36 Continuing Education Units (CEUs) from MIT



Insights and coding demos from renowned MIT faculty



Market-ready data engineering skills in a high-growth market



Build a GitHub portfolio of your projects to share with potential employers

## Services offered by Emeritus

---



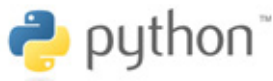
Live weekly office hours with course leaders followed by a Q&A



Personalized feedback, support, career guidance, and network development

# Tools and Resources in the Program

The Professional Certificate in Data Engineering program employs the latest industry tools and resources, including:



# Who Is This Program For?



**Career launchers:** Recent STEM graduates/post-graduates/interns looking to start a career in this high-growth field by gaining exposure to data engineering.



**Career builders:** Early career software engineers/technology professionals seeking to train in the latest data engineering tools and techniques and advance their careers.



**Career switchers:** Mid-career professionals aiming to switch to data engineering from IT, analytics, finance, project management, supply chain, or another technical field.

Program prerequisites: There are no prerequisites for this program, but prior coding experience (in any language) and a minimum of a bachelor's degree are recommended.

## Prepare for these potential job titles:

- Data Engineer
- Big Data Engineer
- Data Software Engineer
- Data Analytics Engineer
- Data Science Engineer
- Data Platform Engineer
- Python Data Engineer
- Cloud Data Engineer
- Data Warehouse Engineer
- Data Integration Engineer
- Data Infrastructure Engineer
- Data Systems Engineer
- Business Intelligence Engineer



"Data engineering really is a core component of today's data infrastructure. And because organizations can't function without data, it's also a career with a great deal of opportunity and incredibly interesting work as well."

- Abel Sanchez, Research Scientist and Executive Director  
of MIT's Geospatial Data Center



# Key Takeaways

---

This program is designed to prepare you with the skills you will need to start or continue your career in data engineering. High-level learning outcomes for this program include:

- Learn to code in Python
- Use SQL to create databases
- Wrangle and analyze millions of pieces of data using databases in Python
- Understand how networks work, including IPs, security, and servers
- Manage big data using data warehousing and workflow management platforms
- Use cutting-edge data engineering platforms and tools to manage data
- Explore artificial intelligence and machine learning concepts, including reinforcement learning and deep neural networks



# Program Schedule

---

This program is organized into three main sections:

## Section 1

As you begin your learning journey in data engineering, you will begin with learning (or refreshing your knowledge of) the Python programming language, how to work with relational databases using SQL, and how to work with Python to create databases and server pipelines.

### Module 0:

#### Orientation

The first week is an orientation module. You will have access to the learning platform from the program start date. You will also install the software and tools you will use in the program on your own computer. There is no teaching, all the content is pre-recorded.

### Module 1:

#### Introduction to Python

You will learn the basic Python syntax, including fundamental data types and constructs. You will practice using Python in multiple assignments.

### Module 2:

#### Python: NumPy

You will be introduced to the Numpy library and its functions for array manipulation and probability, distributions, and interpreting histograms. You will also learn how to use the Matplotlib library to create data visualizations.

### Module 3:

#### Python: pandas

This module will introduce you to the pandas library and its basic data types: series and dataframes. The foundations of data cleaning and preparation and of dataframe manipulation are also explored.

### Module 4:

#### Databases: Basic SQL Statements

You will learn how to conceptually and logically design a database using SQL and the MySQL platform. This module also includes an introduction to physical database design.

### Module 5:

#### Databases: Basic SQL Statements

In this module, you will learn more about physical database design using SQL queries, logical operators, and regular expressions.

### Module 6:

#### Databases: Advanced SQL Statements

You will have an opportunity to perform data cleaning and visualize data in SQL. You will be introduced to the client-server interface and how to connect to a driver using MySQL and Python.

### Module 7-8:

#### Portfolio Projects

Build and submit your first portfolio projects to demonstrate your knowledge of the topics learned in Module 1.

## Section 2

In this section of the program, you will deepen your knowledge of SQL and pandas and how to integrate these tools with the help of Python. Examine the security of data, create related APIs, and gain a deeper understanding of data wrangling.

### Module 9:

#### Python Primer and Development Environment

This module begins with a review of Python. You will then learn how to work in GitHub and how to integrate and deploy applications using Docker.

### Module 13-15:

#### Databases

These modules will teach you more about querying data and about the Data Manipulation Language (DML). You will also learn about architectures and their scalability and how to use Python and pandas together with databases and SQL.

### Module 10-12:

#### Software Skills

In these modules, you will learn how to use Docker to connect to a network and how to make sure those containers are part of the same network. Introduction to distributed coding, HTTP, Ports and API and cloud architectures.

### Module 16:

#### Data Wrangling

In this module you will work on a portfolio project by transforming data to a different format using data wrangling as you prepare it for data analytics.

## Section 3

In this section of the program, you will explore big data and data warehousing. You will discover the connections between artificial intelligence (AI), machine learning, and data engineering.

### Module 17-18:

#### Data Warehouses

In these modules, you will learn about Extract Transform Load (ETL) – a process to extract, transform, and load data from multiple sources to a data warehouse or other unified data repository. You will then use the Apache Airflow platform to programmatically author, schedule, and monitor workflows.

### Module 21-22:

#### Streaming Data and Parallel Computing

In these modules, you will learn about Python parallel computing, file creating and analysis using DASK, and web sockets. You will use the Message Queuing Telemetry Transport (MQTT), a lightweight, publish-network protocol, to communicate and transports messages between devices.

### Module 19-20:

#### Big Data

These modules will teach you more about Hadoop, an open-source software platform for distributed storage and distributed processing of very large data sets on computer clusters built from commodity hardware. You will also be introduced to Spark: a scalable open-source data processing engine to increase performance around speed, using APIs in multiple languages and platforms.

### Module 23-24:

#### Introduction to AI and ML

These final modules introduce basic ML and AI topics, including gradient descend, Naïve Bayes, and Guassian Naïve Bayes. You will also learn about K-means and reinforcement learning, neural networks, and deep learning.

**Note:** Break weeks are included to cover project assignment work and prepare for the upcoming module.



# Assignments and Portfolio Projects

---

Each module includes engaging assignments and culminates in at least one GitHub portfolio project that you'll complete based on what you have learned in that portion of the program.

## Assignments



Peer discussions



Interactive activities



Practice exercises



Knowledge checks

## Coding Exercises

Coding exercises are integrated into various modules through simple activities using Jupyter Notebook. They allow you to practice building composite skills to prepare you for the assignments and portfolio projects.

## Portfolio Projects

- Build a Predictive Machine Learning Model involving Feature Selection for Linear Regression
- Build a Reinforcement Learning Model for Robot Navigation (from scratch in Python)
- Run TensorFlow for a Deep Neural Network Model (Deep Dream in Colab)
- Build a Producer/Subscriber Broker for Visualizing Streaming MQTT Sensor Data (integrating ThingsBoard and Firebase)
- Stream load 100 million lines of data and create and write 20 files in parallel using DASK
- Protect your Web Server using JSON Web Tokens

You will receive personalized feedback from your course leaders to include in your GitHub repositories, securing a market-ready portfolio that's ready to share with potential employers.

# Career Preparation and Guidance

Stepping into a career in data engineering requires a variety of both hard and soft skills. This program offers you guidance on navigating a career path into tech, including crafting your elevator pitch and giving you tips on communication. These services are provided by Emeritus, our learning collaborator for this program. The program support team includes course leaders to help you reach your learning goals. Eligible participants may receive introductions to our hiring partners; however, job placement is not guaranteed.

## Elements of the career preparation aspects of this course include:



Crafting your elevator pitch



Navigating your job search



Optimizing your LinkedIn profile



Learning interview tips and preparation



Writing resumes/cover letters



Negotiating salary

## Career exercises focused on launching a career as a data engineer:



Building your personal brand and promoting your skills



Communicating technical concepts to non-technical colleagues



Understanding the roles and workflow of Agile development



Reflecting on your skills to learn how to troubleshoot and learn more quickly



Job search and interviewing for data engineering positions



"Companies today are being forced to respond faster and with more precision than ever. Data engineers are responsible for innovating and building pipelines that can turn raw data into business advantage. Companies like AirBnB, Uber, and Robinhood are masters of leveraging real-time data, and many others are trying to understand how to do this. Data engineers are the problem solvers of the cyber-info world."

– John R. Williams, MIT Professor of Information Engineering

# Program Faculty



## JOHN R. WILLIAMS

Professor of Information Engineering in The MIT  
Department of Civil and Environmental Engineering

John Williams holds a BA in physics from Oxford University, an MS in physics from UCLA, and a PhD. in numerical methods from the University of Wales, Swansea. His research focuses on the application of large-scale computation to problems in cyber-physical security and energy studies. He is director of MIT's Geospatial Data Center, and, from 2006–2012, was director of the MIT Auto-ID Laboratory, which invented the Internet of Things.

He is author or co-author of over 250 journal and conference papers, as well as the book, RFID Technology and Applications. He contributed to the 2013 report for the UK Office for Science Foresight Project – The Future of Manufacturing.

Alongside Bill Gates and Larry Ellison, he was named as one of the 50 most powerful people in computer networks. He consults for organizations including Accenture, Schlumberger, SAP Research, Microsoft Research, Kajima Corp, US Lincoln Laboratory, Sandia National Laboratories, US Intelligence Advanced Research Projects Activity, Motorola, Phillip-Morris Inc., Ford Motor Company, Exxon-Mobil, Shell, Total, and ARAMCO.

His international collaborations include HKUST and PolyU (Hong Kong), the University of Cambridge and Imperial College of Science and Technology (UK), Malaysia University of Science and Technology (MUST), KACST (Saudi Arabia), and Masdar Institute of Science and Technology (Abu Dhabi).

He organized the first Cyber-Physical Security Conference in the UK (2011), and, along with Dr. Sanchez, he runs the MIT Applied Cybersecurity Professional Education summer program. At MIT, he teaches courses in Architecting Software Systems (MIT 1.125) and Engineering Computation and Data Science (MIT 1.00/1.001).

# Program Faculty



## ABEL SANCHEZ

Research Scientist and Executive Director of MIT's Geospatial Data Center

Dr. Abel Sanchez holds a PhD. from the Massachusetts Institute of Technology (MIT). He is the executive director of MIT's Geospatial Data Center, architect of the Internet of Things global network, and architect of data analytics platforms for SAP, Ford, Johnson & Johnson, Accenture, Shell, Exxon Mobil, and Altria. In cybersecurity, Dr. Sanchez architected impact analysis of large-scale cyber attacks, designing Cyber Ranges for the Department of Defense (DOD).

In password security, Dr. Sanchez led the design of a password firewall (negative authentication) for the Intelligence Advanced Research Projects Activity (IARPA) agency. In machine learning, addressing fraud detection, Dr. Sanchez designed a situational awareness framework that exploits different perspectives of the same data and assigns risk scores to entities for Accenture.

He led the design of a global data infrastructure simulator, modeling follow-the-sun engineering, to evaluate the impact of competing architectures on the performance, availability, and reliability of the system for the Ford Motor Company. He has been involved in developing e-learning software for Microsoft via their I-Campus Program and establishing the Accenture Technology Academy, an online resource for over 200,000 employees.

He has 10 years of experience with learning management systems and has made deployments in America, Asia, and Europe. He teaches MIT courses on cybersecurity, engineering computation, and data science, and he has produced over 150 educational videos.

# Financing Options

---

We want to make sure that the Professional Certificate in Data Engineering is an affordable option for all. This is why we offer you many different ways to pay for the program.

## Loan Partners (For US Residents)

### **Sallie Mae** Monthly payments starting at USD 25

- Visit the [Sallie Mae application portal](#)
- Fill in your basic details and proceed to the loan application page.
- At the time of loan application, please select '**Student and Career Training School**' when prompted
- Choose from fixed repayment, interest-only repayment, and deferred payment options and submit your application
- Our program advisors will contact you for a confirmation on your loan application\*
- After confirmation, we will certify your loan. You will receive a welcome email with login instructions from [notifications@instructure.com](mailto:notifications@instructure.com) within 3 business days

---

### **Ascent Funding** Monthly payments starting at USD 40

- Visit the [Ascent Funding application portal](#)
- Enter your email address and select the 'Professional Certificate in Data Engineering' option from the dropdown list
- Choose from immediate repayment, interest-only repayment, and deferred payment options and submit your application
- Our program advisors will contact you with a confirmation on your loan application\*
- After confirmation, we will certify your loan and you will sign the final disclosures.
- You will then receive a welcome email with login instructions from [notifications@instructure.com](mailto:notifications@instructure.com) within 3 business days

## Flexible Payment Options (For All Countries)

- Choose to make your payment in two, three, or six [installments](#) for higher flexibility. Monthly payments start as low as US\$1,166
- Complete your application for the [Professional Certificate in Data Engineering](#) and enroll for the program

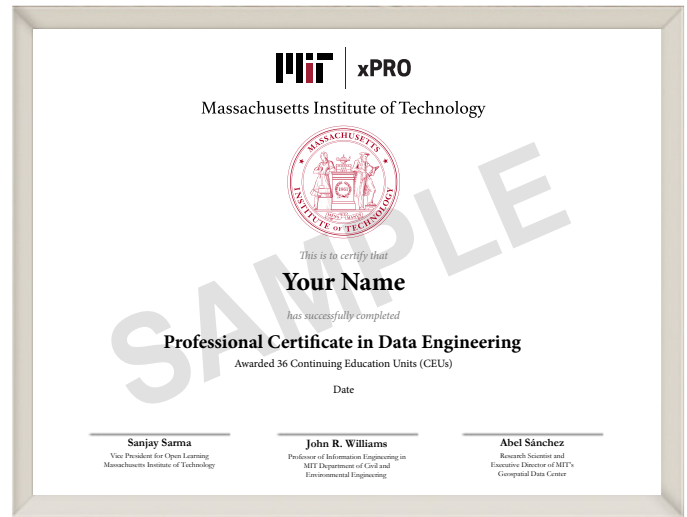
You can opt for any one of the financing options to cover up to the full cost of the program tuition. If you are considering financing your program through one of our partners, the enrollment process can only be completed with the assistance of your program advisor or by calling +315-538-6867.

\*Due to processing time, the loan application should be submitted no later than four business days prior to the enrollment deadline.

# Certificate

Get recognized! Upon successful completion of this program, MIT xPRO grants a certificate of completion to participants and 36 Continuing Education Units (CEUs) from MIT. This program is graded as a pass or fail; participants must receive 75% to pass and obtain the certificate of completion.

After successful completion of the program, your verified digital certificate will be emailed to you, at no additional cost, with the name you used when registering for the program. All certificate images are for illustrative purposes only and may be subject to change at the discretion of MIT.



# About MIT xPRO

MIT xPRO's online learning programs leverage vetted content from world-renowned experts to make learning accessible anytime, anywhere. Designed using cutting-edge research in the neuroscience of learning, MIT xPRO programs are application focused, helping professionals build their skills on the job. To explore the full catalog of MIT xPRO courses and programs, visit: [xpro.mit.edu](https://xpro.mit.edu).

# About Emeritus

MIT xPRO is collaborating with online education provider Emeritus to deliver this online course through a dynamic, interactive, digital learning platform. This course leverages MIT xPRO's thought leadership in engineering and management practice developed over years of research, teaching, and practice.



Easily schedule a call with a program advisor from Emeritus to learn more about this MIT xPRO program.

**SCHEDULE A CALL**

You can apply for the program here

**APPLY**



**Connect with a  
program advisor**

---

**Email:** [mit@emeritus.org](mailto:mit@emeritus.org)

**Phone:** +1-315-538-6867

Delivered in collaboration with

