

# Data Engineer with Python

Start your journey to becoming a data engineer and gain the in-demand data engineering skills companies need. In this track, you'll discover how to build an effective data architecture, streamline data processing, and maintain large-scale data systems.

In addition to working with Python for data engineering tasks, you'll also grow your language skills as you work with Shell, SQL, and Scala, to create data engineering pipelines, automate common file system tasks, and build a high-performance database.

Through hands-on exercises, you'll add cloud and big data tools, such as AWS Boto, PySpark, Spark SQL, and MongoDB, to your data engineering toolkit to help you create and query databases, wrangle data, and configure schedules to run your pipelines.

By the end of this track, you'll have mastered the critical database, scripting, and process skills you need to progress your career and have a firm grasp of Python for data engineering.

Please note this track assumes a fundamental knowledge of Python and SQL.

✓ Completed

Python 73 hours 19 Courses 2 Skill Assessments



## Understanding Data Engineering

Discover how data engineers lay the groundwork that makes data science possible. No coding involved!



Share

Replay Course

### TEST YOUR NEWLY ACQUIRED SKILLS...



SKILL ASSESSMENT

**Python Programming**

**ADVANCED** Score: 154 | Percentile: 96%



## Introduction to Data Engineering

Learn about the world of data engineering in this short course, covering tools and topics like ETL and cloud computing.



Share

Replay Course



### Streamlined Data Ingestion with pandas

Learn to acquire data from common file formats and systems such as CSV files, spreadsheets, JSON, SQL databases, and ...

[Share](#)[Replay Course](#)

### Writing Efficient Python Code

Learn to write efficient code that executes quickly and allocates resources skillfully to avoid unnecessary overhead.

[in](#)[Practice Now](#)

### Writing Functions in Python

Learn to use best practices to write maintainable, reusable, complex functions with good documentation.

[Share](#)[Replay Course](#)

### Introduction to Shell

The Unix command line helps users combine existing programs in new ways, automate repetitive tasks, and run programs ...

[Share](#)[Replay Course](#)



## Data Processing in Shell

Learn powerful command-line skills to download, process, and transform data, including machine learning pipeline.

[in Share](#)[Replay Course](#)

## Introduction to Bash Scripting

Bash scripting allows you to build analytics pipelines in the cloud and work with data stored across multiple files.

[in Share](#)[Replay Course](#)

## Unit Testing for Data Science in Python

Learn how to write unit tests for your Data Science projects in Python using pytest.

[in Share](#)[Replay Course](#)

## Object-Oriented Programming in Python

Dive in and learn how to create classes and leverage inheritance and polymorphism to reuse and optimize code.

[in](#)[Practice Now](#)



### Introduction to Airflow in Python

Learn how to implement and schedule data engineering workflows.



Share

Replay Course



### Introduction to PySpark

Learn to implement distributed data management and machine learning in Spark using the PySpark package.



Share

Replay Course



### Introduction to AWS Boto in Python

Learn about AWS Boto and harnessing cloud technology to optimize your data workflow.



Share

Replay Course

#### TEST YOUR NEWLY ACQUIRED SKILLS...



SKILL ASSESSMENT

### Data Analysis in SQL (PostgreSQL)

**INTERMEDIATE** Score: 125 | Percentile: 79%



SQL

## Introduction to Relational Databases in SQL

Learn how to create one of the most efficient ways of storing data - relational databases!



Share

Replay Course



SQL

## Database Design

Learn to design databases in SQL.



Share

Replay Course



## Introduction to Scala

Begin your journey with Scala, a popular language for scalable applications and data engineering infrastructure.



Share

Replay Course



## Big Data Fundamentals with PySpark

Learn the fundamentals of working with big data with PySpark.



Share

Replay Course



## Cleaning Data with PySpark

Learn how to clean data with Apache Spark in Python.



 Share

Replay Course



## Introduction to MongoDB in Python

Learn to manipulate and analyze flexibly structured data with MongoDB.



 Share

Replay Course