

Data Engineer, Bootcamp (13 weeks)

by DataScientest

Practical course - 57d - 399h00 - Ref. 4II

Price : 7490 € E.T.

Devenez expert en analyse de données avec ce parcours proposé par notre partenaire DataScientest. Un Data Engineer a pour mission de concevoir des outils et solutions qui vont permettre de traiter et d'analyser de grands volumes de données. Cette formation certifiante se déroule à distance dans un format hybride mêlant temps d'échanges synchrones avec un formateur expert, exercices pratiques et modules E-learning. Basée sur la pédagogie Learning By Doing, vous réaliserez un projet fil rouge en équipe afin de mettre en pratique vos connaissances. Lors de votre inscription, vous serez rattaché à l'une des promotions DataScientest. A l'issue de cette formation, vous obtiendrez un co-certificat « Data Engineer » des Mines Paris - PSL Executive et de DataScientest ainsi que les blocs de compétence 2 et 3 de la certification RNCP « Data Engineer ». Contactez-nous dès maintenant pour connaître les prochaines dates !

Teaching objectives

At the end of the training, the participant will be able to:

- ✓ Develop a technical architecture for data management.
- ✓ Deploy a massive data analysis solution integrating artificial intelligence.

Intended audience

People with an interest in programming and data manipulation.

Prerequisites

Un diplôme ou un titre de niveau bac+3 et des connaissances en Python, SQL, Linux.

Pour les candidats ne présentant pas le niveau de qualification requis, une dérogation est possible sur dossier.

PARTICIPANTS

People with an interest in programming and data manipulation.

PREREQUISITES

Un diplôme ou un titre de niveau bac+3 et des connaissances en Python, SQL, Linux.

Pour les candidats ne présentant pas le niveau de qualification requis, une dérogation est possible sur dossier.

TRAINER QUALIFICATIONS

The experts leading the training are specialists in the covered subjects.

They have been approved by our instructional teams for both their professional knowledge and their teaching ability, for each course they teach. They have at least five to ten years of experience in their field and hold (or have held) decision-making positions in companies.

ASSESSMENT TERMS

The trainer evaluates each participant's academic progress throughout the training using multiple choice, scenarios, hands-on work and more.

Participants also complete a placement test before and after the course to measure the skills they've developed.

Certification

Pour clôturer la formation, l'équipe pédagogique évaluera le projet fil rouge de l'apprenant à l'aide d'un rapport écrit et d'une soutenance à distance. La validation des compétences développées au cours de la formation Data Engineer vous permettra d'obtenir : • Un co-certificat « Data Engineer » des Mines Paris - PSL Executive et de DataScientest • Les blocs de compétence 2 et 3 de la certification RNCP de niveau 7 "Data Engineer" enregistrée au RNCP sous le n°RNCP38919.

Practical details

Digital activities

Online courses and exercises, group masterclasses, question/answer sessions, support classes, e-mail coaching, red thread projects, individualized career coaching, social learning.

Mentoring

Un formateur expert accompagne l'apprenant tout au long de sa formation. Il échange régulièrement avec lui sur son projet fil rouge et l'accompagne lors de points de mentorat (individuel). Plusieurs formateurs animent également les différentes masterclass (classes collectives) et répondent aux questions des apprenants à tout moment depuis un forum dédié. En complément, de nombreuses séances de questions-réponses peuvent être organisées pour aider les apprenants.

Pedagogy and practice

Lors de l'inscription, l'apprenant est affecté à une promotion (dates à définir lors de l'inscription) et reçoit son calendrier de formation. Le parcours de formation est découpé en « Sprint » de plusieurs semaines sur une thématique dédiée. Chaque semaine l'apprenant est convié à un temps d'échange avec le formateur qui se présente sous la forme de masterclass (classe collective) ou de points de mentorat (individuel). Pendant 80% du temps, l'apprenant travaille en autonomie sur la plateforme d'enseignement. Tous les modules intègrent des exercices pratiques permettant de mettre en œuvre les concepts développés en cours. L'apprenant doit également travailler en binôme ou trinôme sur un projet fil rouge tout au long de la formation. Cela lui permettra de développer et faire reconnaître ses compétences. En complément, des événements et ateliers thématiques sont régulièrement proposés pour permettre aux apprenants de découvrir les dernières innovations en matière de Data Science. Afin de suivre efficacement la formation, nous estimons le temps travail nécessaire entre 35 et 40 heures par semaine.

Course schedule

1 Upcoming session dates

- October 2025: Start date 07/10/25
- November 2025: Start date 04/11/25
- December 2025: Start date 02/12/25

2 Programming

- Python: variables, types, operators, loops, functions, classes, modules.
- Python: multithreading and multiprocessing on Python, asynchronous function, MyPy library.
- Web Scraping: BeautifulSoup, HTML document navigation and data identification.

TEACHING AIDS AND TECHNICAL RESOURCES

- The main teaching aids and instructional methods used in the training are audiovisual aids, documentation and course material, hands-on application exercises and corrected exercises for practical training courses, case studies and coverage of real cases for training seminars.
- At the end of each course or seminar, ORSYS provides participants with a course evaluation questionnaire that is analysed by our instructional teams.
- A check-in sheet for each half-day of attendance is provided at the end of the training, along with a course completion certificate if the trainee attended the entire session.

TERMS AND DEADLINES

Registration must be completed 24 hours before the start of the training.

ACCESSIBILITY FOR PEOPLE WITH DISABILITIES

Do you need special accessibility accommodations? Contact Mrs. Fosse, Disability Manager, at psh-accueil@orsys.fr to review your request and its feasibility.

3 Advanced tools

- Git: introduction, Git repository, branch, tag and merge.
- GitHub: introduction to the platform, fork, pull request, issues, pull and push, Github Actions.
- Linux systems and Bash scripting: Linux systems, using a terminal, Bash scripting.

4 Big Data Variety

- SQL: relational databases, SQL language, in-depth analysis and application.
- Elasticsearch: search engine, index, Mapping, Ingest node, Text Analyzer.
- MongoDB: presentation, MongoDB queries.
- Neo4j: graph-oriented data, Cypher query, data loading, Python client for Neo4J.
- Hbase: column-oriented databases, data modification using Python and Happybase.

5 Batch & streaming

- PySpark: distributed computing, RDD and Dataframe APIs, distributed data processing, distributed Machine Learning.
- Kafka: architecture and benefits, settings management, Consumers settings.
- Streaming with Spark: real-time data processing, mini-batch streaming, Structured Streaming, pipeline.

6 Practical data storage

- Snowflake: Data Warehousing with robust security, SQL data analysis for the cloud, platform optimization.
- Data Warehousing with DBT (ELT): transformations, high-quality datasets, automated execution.

7 Cloud AWS

- AWS Solution Architect: best practices, architecture design, continuous improvement and automation.
- AWS Solution Architect: presentation of the AWS cloud, key services on the AWS platform.

8 Machine Learning

- Statistics: numerical variables, categorical variables, relationships between variables.
- Data Visualization: different types of graphs with Matplotlib, creation of Dash applications.
- Machine Learning: pre-processing, Machine Learning algorithms (regression, classification, clustering).
- ML Flow: MLFlow architecture, MLFlow Tracking, MLFlow Projects, MLFlow Models, MLFlow Registry, lifecycle.

9 DevOps - Virtualization

- APIs: microservices architectures, HTTP methods, FastAPI and Flask libraries, OpenAPI specification, API management.
- Docker: concept of containerization, images and containers, communication, persistence, Dockerhub, docker-compose.
- API security: API Keys, HTTP Basic authentication, JSON Web Token and HTTPS.
- Kubernetes: deploying and managing containers, initialization and architecture, APIs with Kubernetes.

10 CI/CD and Monitoring

- Airflow: orchestration concept, acyclic directed graph or DAG, operators, task management, monitoring.
- Unit testing with Python: unit testing with Pytest, integration testing, benefits of testing, integration.
- GitLab: installation, initialization, adding and deleting, Git Blame, Tag, repository status, conflict management.
- Prometheus & Grafana: the benefits of monitoring, Prometheus Query Language, Dashboard with Grafana, integration.