XCCELERATED
proudly part of Xebia Group

Data Engineering Program

Data engineers are the pillars on which every data-driven company is built. We believe software engineering is the foundation to become a data engineer, so we put an emphasis on consolidating principles and best practices rather than tools. Of course, you'll deep dive into the most effective technologies as well.

Participants learn through classroom lectures, short assignments, hackathons, and collaborative project assignments with colleagues from our Machine Learning Engineering or Cloud programs. Emanuel, our data engineering lead, chooses the best training approach for the topic, depending on the needs of the groups. The following is an overview of the program;

First Month: Bootcamp

The Xccelerated Data Engineering Program kicks off with a full-time bootcamp. This intensive training focuses on writing deployable code with Python, containerization, data processing at large scale with Apache Spark and Kafka, and cloud fundamentals. Participants learn through classroom lectures, short assignments, hackathons, and collaborative project assignments with our Machine Learning Engineering or Cloud programs. Emanuel, our Data Engineering Lead, chooses the best training approach for the topic, depending on the needs of the groups.

Beyond Bootcamp

After the bootcamp, you spend a full year working on-site four days a week at one of our partner companies and returning one day a week to our Xccelerated office in Amsterdam. During these "return" days, you split your time between theory and practice—gaining project guidance and coaching from our supervising consultants to support you in the partner company's projects, while also diving deeper into training topics not covered in the bootcamp. It's always about continuous learning. Besides the technical aspect, we'll focus on your personal development as well.

Learning Modules

Module 1 Bootcamp Training Program

The training begins with a full-time bootcamp that covers multiple topics in a classroom setting, such as:

1. How to write deployable code: pythonic code; testing methodologies; principles behind scalable and robust code and the functional paradigm.
2. Containerization with docker and various deployment strategies.
3. Cloud fundamentals: infrastructure as code; network and security and serverless solutions.
4. Data lakes: how to process data at large scale with Apache Spark and use Kafka as a streaming system.

The bootcamp classroom training is followed by a one- or two-day hackathon, where participants put newly acquired skills and knowledge to the test.

Module 2 Scaling Applications Through Concurrency and Async

A short response time and the ability to scale with demand is the new norm for applications developed today. Abandoning the single flow of execution in favor of multiple flows of executions is one way to achieve both. We will discuss CPU-bound and IO-bound operations, which paradigms to use and when, plus common pitfalls and how to avoid them.

Module 3 Workflow Scheduling with Apache Airflow

Apache Airflow is one of the most popular data pipeline orchestration tools. In this module, participants learn how to author, schedule and monitor data pipelines using the Airflow's built-in operators, as well as how to build custom operators.

Module 4 Faster Deployments and the Path to Microservices

Scaling and deploying smaller individual units of work independently can speed up the production process. Unfortunately, that can also increase operating costs. This module teaches you how to use Kubernetes to avoid those costs while alleviating the pain of a slowing process.

Module 5 Model Deployment Hackathon

The hackathon pits small teams of data engineers and scientists against each other in a real-life machine-learning simulation. All teams compete in the same virtual world. The challenge? To quickly determine whether or not to accept individual loans. Loan requests come in continuously, and only a single team can handle each loan. If a team's server is down or slow, it impedes their ability to accept loans. The teams must deal with shifts in data patterns over several years of collected loans and make smart choices about when and how to retrain their models. In the end, the team that earns the most money wins.

Module 6 Developing Near Real-Time Systems

Streaming systems allow you to do more with fewer resources in less time than batch systems. However, streaming requires that you pay more attention to data consistency. How do you handle late data? How do you maintain stability? Can you process the same record multiple times without an undesirable effect? This module covers these questions and more, as participants learn how to work with streaming systems for near real-time processing and to handle telemetry data collected from different IoT devices.

Module 7 Scaling Horizontally aka Distributed Systems

Scaling applications through distributed systems makes sense on the surface. Adding more machines to do the work distributes your load, making it easier to handle higher and higher demands. However, since the communication is done through a network, the state of the system is constantly in complete chaos. The question is not if you have failures, but when. This module teaches engineers to design systems that isolate failures and to also design recovering processes.

Module 8 Where is Your Data? aka Storage Systems

Data is the most valuable asset for data-driven companies—after the people who handle it. Learn how to combine different technologies to make data accessible and processable in a secure, efficient, reliable and reproducible way.

Module 9 Finding Similar Items Hackathon

In this project, small teams of data engineers and data scientists collaborate to build a scalable data product—a search system that efficiently finds similar images to an input image. Defining what it means for images to be similar and how to efficiently search for items in a highdimensional search space containing tens of millions of elements are just the tip of the iceberg.

XCCELERATED
proudly part of Xebia Group

www.xccelerated.io
hello@xccelerated.io
+31 (0) 20 76 09 842

Scan QR-code to sign-up for the program