



2023

Training Guide

Data Engineering Learning Journeys



Our Promise to You

- › Learn Today, Apply Tomorrow
- › Authority From the Field, Like No Other
- › Learn by Doing, in Hands-On Labs
- › From Team to Boardroom
- › Learn Your Way, at Your Pace

[Go to the academy](#)



Learn Today, Apply Tomorrow, and Stay in the Driver's Seat

Are the advancements in cloud and ready-to-use technologies such as Stable Diffusion and GPT-3 making data engineering and data science skills obsolete? Far from it. Without knowing what is going on under the hood, extracting maximum performance is impossible. Understanding, controlling, and utilizing the full power of these tools makes having up-to-date data and AI skills more relevant than ever.

There was a time, not so long ago, when gathering enough data to train a deep learning model — and then figure out how to train it on a budget — was an essential skill of data professionals.

Then pre-trained models and communities around them — such as HuggingFace — made it not only easy, but quick and affordable.

This begs the question: Is being proficient in data science and engineering still relevant when you can outsource the job to someone else? The answer is yes — because everything is changing again.

With all the advantages of offloading our algorithms to “someone else”, it’s more important than ever to stay in the driver’s seat. Here are three reasons why:

(1) The lack of transparency and apparent biases in algorithmic decision-making has fuelled public outcry on numerous occasions — the Netherlands made the world news with discrimination against dual nationality families and child benefits. You can only control the algorithm if you can tweak and modify it to address the ethical issues around it. Understanding what happens under the hood, even when using AI-as-a-Service, allows you to stay in Control.

(2) In 2023, we will continue to generate TBs of data per person every month. Storing it all would be far too expensive, but we can apply machine learning to the devices that generate it and only keep the insights. More and more systems at the edge, including mobile phones, have frameworks to help you out.

(3) Transfer learning takes a highly sophisticated algorithm and adapts it to your use case and data. As it requires less training samples and computational power, it is affordable to do it in-house.

These three trends make data science and engineering skills more relevant than ever.

So, it is my pleasure to present Xebia’s all-new training curriculum. Developed and taught by the very best professionals in the field with pragmatism in mind — you can apply tomorrow what you learn today. Happy learning!

Giovanni Lanzani
Managing Director Data
Xebia



[Go to the academy](#)



Table of Contents

5 Xebia Academy

6 Data and AI Learning Journey

7 Data Engineering Learning Journey

- 9 dbt Learn
- 10 Python for Data Engineers
- 11 Create Data Science Products
- 12 Data Processing at Scale
- 13 Docker and Kubernetes
- 14 Certified Apache Airflow
- 15 Getting Started with Apache Airflow
- 16 Optimizing Apache Spark and Tuning Best Practices
- 18 Streaming Architecture at Scale

20 In-company Data and AI Learning Journeys

- 21 From Team to Boardroom – Data and AI Courses for Your Entire Organization
- 23 Supporting bol.com as a Frontrunner in Data and AI
- 24 Data and AI Accelerator Program for Enterprises
- 25 Securing Budget for your Training Course
- 26 Big Enough to Deliver, Small Enough to Care
- 27 Trusted by global industry leaders

28 Trainer Overview

30 Xebia Academy Team for Data and AI Learning Journeys

31 Training Courses Overview

Xebia Academy

Our Promise to You

At Xebia, we believe professional development goes hand-in-hand with staying happy, motivated, confident, and relevant in your job. You trust in us to help you improve your data and AI skills, and we take that very seriously.

So we strive to provide you with the best learning experience possible, by adhering to five promises:

(1) Learn Today and Apply Tomorrow

We design our programs so you can apply your newly acquired knowledge right out of the classroom. That way you can immediately increase your business value. We push the content of every training far beyond textbooks and theory. Applicability is one of our core values.

(2) Authority From the Field, Like No Other

We don't come from an abstract background; we are practitioners as much as we are teachers. All of our trainers work as consultants in the data and AI field, supporting top enterprises like Booking.com, ING, bol.com, Randstad, and Heineken. They solve challenges like yours every day, so you benefit directly from their experience.

(3) Learn by Doing, in Hands-On Labs

You learn best by doing. That's why, in every training, you can develop your skills and craft in our hands-on labs. We provide both the theory and context to get you up and running fast.

All of our courses have a 50/50 split between theory and hands-on labs.

(4) From Team to Boardroom

Built through years of working with the top enterprises in Europe, Xebia has the expertise to turn your data-driven ambition into reality. But becoming data-driven impacts your whole organization. That's why we deliver a wide range of programs suitable for all shapes and sizes—from individual teams to global workforces, as well as the boardroom.

(5) Learn Your Way at Your Pace

We deliver our curriculum through various training formats—classroom, in-company, online, or a combination. You choose the format that fits your purpose and preferred method of learning—or ask one of our academy advisors to guide you.

Xebia is an international consulting and training company, specialized in digital transformation. Xebia employs over 6,000 consultants worldwide. Xebia's values:

- > People First
- > Sharing Knowledge
- > Customer Intimacy
- > Quality Without Compromise



Data and AI Learning Journeys

A training course should not just be a single moment of development. It works best when it's part of a broader program. At Xebia, we call these programs Learning Journeys.

Learning Journeys guide you through our curriculum and ensure the courses meet your needs and provide an effective development path. For example, our three Python courses take participants from beginner to expert, after which you can choose a specialization course to further hone their skills.

The Necessity of Learning Journeys

Without guidance, it can be challenging to invest time and resources in an optimum course for professional development — one that is relevant to your job and aligned with your organization's business objectives.

Instead, it's tempting to follow the latest buzz or ignore the relevance of what they learn. Will you become more effective thanks to what you've learned, or does your development end once the training is over?

Such an ad-hoc approach to continuous professional development creates various issues, such as problematic progress tracking and gap measurement. Learning Journeys address these problems.

Benefits of Learning Journeys

The journeys are designed to overcome the drawbacks of working independently with little guidance:

(1) Applicability is high. Learning Journeys are designed by our expert consultants with working experience in the field, to ensure relevance to the participant's role and business needs.

(2) Learned material is applied and tested in practice through use cases. For example, an aspiring data scientist will be working on relevant use cases.

(3) Measuring impact on the workforce is much easier since employees follow similar paths. Small variations average out, and a clear picture emerges.

(4) Designed by the business to satisfy current and future needs, Learning Journeys are a valuable investment in individual careers and in the company as a whole.

(5) Current and potential employees can advance their careers through carefully crafted learning paths.

We have designed Learning Journeys for analytics translators, data scientists, and data engineers. You can find each Learning Journey in this guide:

- » Analytics Engineering Learning Journey – Page 7
- » Analytics Translation Learning Journey – Page 11
- » Data Analyst Learning Journey – Page 16
- » Data Scientist Learning Journey – Page 21
- » Machine Learning Engineering Journey – Page 43
- » Data Engineer Learning Journey – Page 59

[Go to the academy](#)



Advice and Inquiries

Need advice on which course is right for you? Curious about setting up the perfect data and AI in-company program?

Our director of sales, Marcel van Denderen, is available to answer your questions. Get in touch at +31 6 3623 7853 (phone or WhatsApp), or send an email: marcel.vandenderen@xebia.com

Data Engineering Learning Journey

Data Engineers

Learn how to take data and AI concepts from concept to prototype and to production-ready application. Acquire the skills to develop and run Data and AI solutions at an enterprise-scale with ease! Take part in a specific training or advance through the entire journey. Learn how to build secure data platforms and reliable AI applications that are engineered for scale.

Learning Journey

How do you become a data engineering expert? Start here! We've put together a carefully crafted learning journey for data engineers. Knowing engineers love to figure things out on their own, we packed the program with opportunities to learn, hands-on, by solving real-life situations. Plus, there's plenty of practical philosophy, too.

We'll teach you how to leverage Docker to ease your deployments and navigate code written by data scientists (Advanced Python and Data Science in Production). You will learn to use Apache Airflow, Apache Spark, and Kafka like a forklift to move data around. And we won't shy away from proven technologies either, like Elasticsearch. We also remain on the cutting edge with others, like Apache Flink.

Learning Goals of the Fundamentals for Data Engineers

- › Writes correct and clean code with guidance
- › Participates in the technical design of features with guidance
- › Knows how to integrate CI/CD concepts into their daily coding
- › Able to create simple pipelines without guidance
- › Knows how containerization works, and what it simplifies
- › Can write and push containers

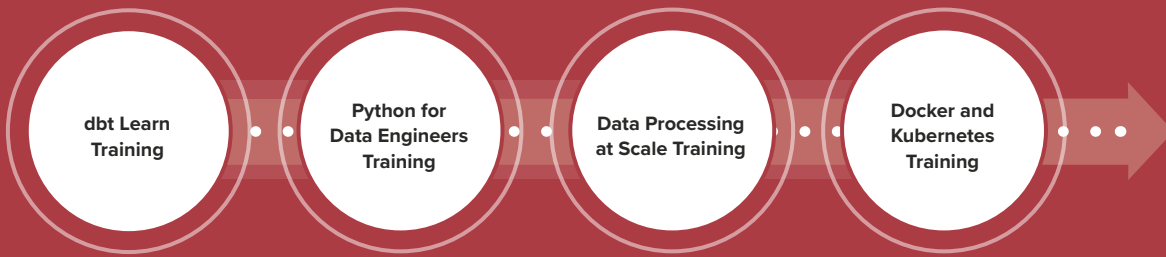
Learning Goals for Advanced Data Engineers

- › Understands and makes well-reasoned design decisions and trade-offs in their area
- › Able to quickly get familiar with larger codebases
- › Able to create complex pipelines without guidance

Learning Goals for Data Engineers who want to specialize

- › Go-to expert in one area; understands the broad architecture of the entire system
- › Provides technical advice and weighs in on technical decisions that impact other teams or the company at large

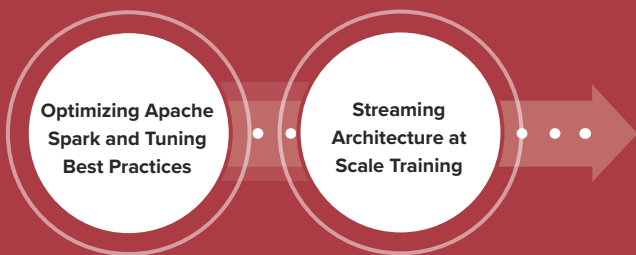
Fundamentals for Data Engineers Learning Journey



Advanced Data Engineer Learning Journey



Specialization for Data Engineers Learning Journey



Cloud Data Engineer Learning Journey



Data Engineering Learning Journey

dbt Learn

 3 days
  Fundamentals


Learn everything you need to know to improve your dbt (data build tool) skillset.

In partnership with dbt Labs, Xebia delivers dbt Learn - a training course that runs over (3) half-days. With this open-source piece of magic, you can transform data in warehouses simply with the power of SQL select statements. Learn about modeling strategies, automated testing, and performance optimization.

This training IS for you if...

- ✓ You (or your team) are using or about to start using dbt.
- ✓ You want to learn the hottest tool for modeling data.
- ✓ You want ownership when it comes to working with your data.
- ✓ You want to be added to gain the dbt Fundamentals badge automatically.

This training is NOT for you if...

- ✗ You want to learn how to use Python to explore data (check out our Python for Data Analysis training instead).
- ✗ You have no intention of ever working with dbt.
- ✗ You have no experience with SQL.

After the training, you'll be able to

- ✓ Use SQL select statements to build robust models for your data
- ✓ Use jinja to create scalable and advanced SQL statements
- ✓ Take part in the dbt community by sharing knowledge

Course Topic Details**dbt Foundations**

- ✓ Build models to shape your data from raw to transformed
- ✓ Configure and run tests on your data to meet your expectations
- ✓ Write, generate, and view documentation as you develop

dbt Applications

- ✓ Leverage Jinja and Macros to write DRY code
- ✓ Create and implement existing packages to make robust scalable projects
- ✓ Deploy your dbt Project to run on a schedule with dbt Cloud

Program**Day 1**

- ✓ The role of an analytics engineer
- ✓ Distributed dbt Learn norms
- ✓ Fundamentals of dbt and Project Design

Day 2

- ✓ Test your transformations and source data
- ✓ Document your transformations to improve collaboration
- ✓ Deploying your dbt projects
- ✓ Polishing your dbt project

Day 3

- ✓ Using Jinja within dbt
- ✓ How to use packages and projects
- ✓ When and where to use dbt Materializations
- ✓ Incremental Models: What they are and where you can use them
- ✓ Trade Techniques: 10 Advanced Tips to continue learning

[More info and dates](#)


Python for Data Engineers



 2 days  Advanced

Python is a flexible and powerful language, which gives developers lots of freedom — but that freedom can lead to unmaintainable, wheel-reinventing, spaghetti-like code.

This training empowers you to create simple, beautiful and truly pythonic code. Through both theory and hands-on exercises, you will learn all the advanced features of Python so you can make your code more readable, maintainable and scalable.

This training IS for you if...

- ✓ Data scientists, analysts and data engineers who spend their working hours developing Python-based solutions.

This training is NOT for you if...

- ✗ You don't have basic Python experience, which is required (check out the Python Essentials course instead).

After the training, you'll be able to

- ✓ How to write a decorators
- ✓ How to use context managers
- ✓ How to properly write Python objects
- ✓ How to properly package a Python app via setup.py
- ✓ How to write proper fixtures for unittesting (Flask and regular apps)
- ✓ How to run Flask from within a Docker container
- ✓ How to combine Flask and unicorn
- ✓ How to add commands to the Flask CLI

Course Topic Details

- ✓ Setting up Python
- ✓ List comprehensions
- ✓ Iterators and generators
- ✓ Functions
- ✓ Type hinting
- ✓ Object oriented structures
- ✓ Decorators
- ✓ Context managers
- ✓ OS Filesystem
- ✓ Command line
- ✓ Unit testing
- ✓ Flask applications

Program

Day 1

- ✓ Setting up Python
- ✓ List comprehensions and iterators and generators
- ✓ Functions and Type hinting
- ✓ Object oriented structures

Day 2

- ✓ Decorators and Context managers
- ✓ OS Filesystem
- ✓ Command line and Unit testing
- ✓ Flask applications

More info and dates



Data Engineering Learning Journey

Create Data Science Products



2 days Advanced

Learn how to efficiently productionize data science models.

Creating data science products out of code requires understanding its multiple variations, without being an expert on any particular machine learning model. In this course you'll learn how to take data science models into production. The course covers pandas, scikit-learn APIs and pipelines to effectively structure Python code. You will also learn about serializing data and monitoring models.

This training IS for you if...

- ✓ You are a data or machine learning engineer who deals with productionizing data science models.

This training is NOT for you if...

- ✗ You don't have basic experience with Python, which is required (check out the Python for Data Engineers course instead).

After the training, you'll be able to

- ✓ Restructure a notebook into a Python package
- ✓ Choose the right file format for storing your data
- ✓ Create testable pipelines from code written by data scientists
- ✓ Monitor models running in production

Course Topic Details

- ✓ Pandas dataframes
- ✓ Sklearn pipelines
- ✓ Data Serialization
- ✓ Hands-on code refactoring exercise

Program

Day 1

- ✓ Pandas dataframes
- ✓ Sklearn pipelines

Day 2

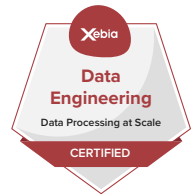
- ✓ Data Serialization
- ✓ Hands-on code refactoring exercise

[More info and dates](#)



Data Processing at Scale

 2 days  Fundamentals



Learn to use Apache Spark to process large sets of data.

Data is knowledge and knowledge is power. But processing data efficiently can be challenging as it scales up. This training dives deep into one of the most popular and scalable tools for large-data transformation: Apache Spark. In this course, you will learn everything you need to know about how Apache Spark works. Through a combination of theory and hands-on exercises, you will also gain the skills you need to write efficient ETL Spark jobs to process large sets of data.

This training IS for you if...

- ✓ You are a data or machine learning engineer who deals with the transformation of large volumes of data.

This training is NOT for you if...

- ✗ You don't have basic experience with Python (required). Check out Python for Data Engineers instead.

After the training, you'll be able to

- ✓ Use Apache Spark and its advanced features
- ✓ Write efficient ETL jobs
- ✓ Use the API to transform data at a basic and advanced level
- ✓ Think in terms of distributed systems when writing Spark jobs

Course Topic Details

- ✓ Inner-workings of Apache Spark
- ✓ Loading data from various formats
- ✓ Basic and advanced dataframe operations
- ✓ Window and user-defined functions
- ✓ Unit testing
- ✓ Hands-on exercise to analyze large-scale logs to find trending topics

Program

Day 1

- ✓ Inner-workings of Apache Spark
- ✓ Loading data from various formats
- ✓ Basic and advanced dataframe operations
- ✓ Window and user-defined functions

Day 2

- ✓ Unit testing
- ✓ Hands-on exercise to analyze large-scale logs to find trending topics

[More info and dates](#)



Data Engineering Learning Journey

Docker and Kubernetes

 3 days
  Specialty


Learn how to containerize your solutions and deploy them in a container-based cluster.

Containerized solutions are here to stay — but what exactly are containers? And how do you build, use and run them? This course covers those questions and more. In three days, you will learn exactly how Docker containers work — how to build them, how to place your solutions in them, and most importantly, how to run them. You will also learn how to deploy your containerized solutions in a Kubernetes cluster and manage all aspects of its life cycle.

This training IS for you if...

- ✓ You are an IT engineer/architect or curious data scientists.
- ✓ You are currently working in a DevOps related area.
- ✓ You are involved in a microservice type of architecture.
- ✓ You want to learn how to better structure and run your solution in a cluster.
- ✓ Bonus if you have some basic knowledge of *nix, though not essential.

This training is NOT for you if...

- ✗ You are looking to know how containers can help your organization but do not plan on building or operating them 4:41.
- ✗ You have a deep understanding of containers and are only interested in Kubernetes.

After the training, you'll be able to

- ✓ Containerize your solution
- ✓ Interact with the container
- ✓ Properly structure the containerized solution
- ✓ Deploy your container(s) to Kubernetes
- ✓ Manage the solution and the cluster

Course Topic Details**Docker**

- ✓ How to work with volumes and containers
- ✓ How Docker builds layers and how this knowledge improves your Dockerfile
- ✓ Interacting with the containers
- ✓ The Joy of automation with Docker-compose
- ✓ How to automatically lint Dockerfiles

Kubernetes

- ✓ Kubernetes core concepts
- ✓ Configuration
- ✓ Multi-container pods
- ✓ Observability
- ✓ Pod design
- ✓ Services and Networking
- ✓ State persistence
- ✓ Introduction to Helm

Program**Day 1**

- ✓ How to work with volumes and containers
- ✓ How Docker builds layers and how this knowledge
- ✓ Interacting with the containers
- ✓ How to automatically lint Dockerfiles

Day 2

- ✓ Kubernetes core concepts
- ✓ Configuration and Multi-container pods
- ✓ Observability and Pod design

Day 3

- ✓ Services and Networking
- ✓ State persistence
- ✓ Introduction to Helm

[More info and dates](#)


Certified Apache Airflow



2 Days



Advanced



Develop workflows with Apache Airflow.

Have you ever scheduled a chain of cron jobs only to discover down the line that data was missing? Apache Airflow eliminates that problem. In this training, you will learn to author, schedule, and monitor workflows through hands-on experience with this leading open-source platform. After successfully passing the exam at the end of the training, you'll receive an Astronomer Airflow certification. This certification demonstrates your knowledge of Airflow's core concepts and your ability to make wise architectural decisions, understand applied use cases, and design data pipelines.

This training IS for you if...

- ✓ You want to schedule and monitor your data pipelines in a user-friendly way.
- ✓ You have complex data pipelines with a lot of dependencies.
- ✓ You want to schedule your data pipelines on external systems.

This training is NOT for you if...

- ✗ You currently don't run any batch data pipelines.
- ✗ You want to use Airflow as an execution engine.
- ✗ You want to learn how to set up Airflow for production.

After the training, you'll be able to

- ✓ Use basic Airflow concepts, such as DAGs, Operators, and Hooks
- ✓ Design and create task workflows using Airflow DAGs
- ✓ Schedule Airflow DAGs and use concepts like backfilling and catchup
- ✓ Use Airflow Operators to execute predefined tasks which integrate with many different systems
- ✓ Use the Airflow UI to get insights into the status of your workflows

Course Topic Details

- ✓ Theory and Practice
- ✓ The essential components of Apache Airflow
- ✓ Running and managing workflows
- ✓ Creating dynamic workflows with Jinja templating
- ✓ Sharing state between tasks with XComs

Program

Day 1

- ✓ Introduction to the Apache Airflow user interface
- ✓ How to create and monitor DAGs
- ✓ The basics of using the most critical operators
- ✓ Creating dynamic workflows with Jinja templating

Day 2

- ✓ How to create dynamic workflows with branching
- ✓ How to communicate with external systems, using hooks and connections
- ✓ How to trigger your workflows with sensors
- ✓ Sharing state between tasks with XComs

More info and dates



Getting Started with Apache Airflow

Machine learning engineer and consultant Julian de Ruiter teaches the Apache Airflow course. He recently used Apache Airflow at a bank to collect data from several external systems into a central data lake. The bank uses the data lake to perform analyses and train machine-learning models. We asked Julian to share some of his challenges with the project and to explain how he uses and teaches Apache Airflow.

“One of the main challenges with such a project is being able to talk to different systems and ensure that data is fetched and processed on time,” he explained. “Apache Airflow makes this process much easier by providing a flexible and extendable framework for defining data workflows. Combined with strong scheduling semantics, it’s relatively easy to use it to build complex workflows that run regularly to ensure your data lake stays up-to-date,” he said.

Julian also explained how he teaches the course. “We start from scratch and work on several different use cases, everyone at their own level,” he said. “The training begins with building a simple example workflow, but by the end of the day, everyone will have built more realistic workflows that automate extraction and loading of data from a database into cloud-based storage that’s suitable for further processing,” he explained.

Testing Workflows

“Many people struggle with setting up a development environment and testing their workflows in Apache Airflow,” explained Julian, “so we spend considerable time on this subject. Everyone learns how to set up a local development environment, how to build custom Airflow components, and how to approach automated testing in Airflow.”

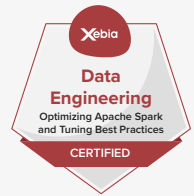
In one of Julian’s recent training sessions, an analyst was using Google’s Big Query to analyze data from a particular source. She wanted to automate the process of getting her data into Airflow. “I thought it was really cool to see how quickly she managed to get this process up-and-running in Apache Airflow, after less than a day of training,” he smiled.

Interested in the Apache Airflow training?

[More info and dates](#)



Optimizing Apache Spark and Tuning Best Practices



2 days Specialty

Learn how to optimize Apache Spark, one of the most popular analytics engines available.

As data scales up, processing it efficiently becomes more challenging. Building on our experience as one of the largest Apache Spark users in the world, this 2-day course provides an in-depth overview of the do's and don'ts of one of the most popular analytics engines available.

This training IS for you if...

- ✓ Are looking to optimize your Spark applications.
- ✓ You are a data engineer or machine learning engineer who deals with transforming large volumes of data and you need production-quality code.
- ✓ You are an expert data scientist who wants to learn simple tweaks to dramatically increase Spark performance.

This training is NOT for you if...

- ✗ If you don't use Python with Spark (PySpark).

After the training, you'll be able to

- ✓ Understand how Apache Spark works internally
- ✓ Use best practices when you write performant code
- ✓ Debug and tweak your Spark applications
- ✓ Explain the Spark fundamentals, including the execution model: Driver/Executors
- ✓ Work with caching, shuffle service, and fair scheduling
- ✓ Troubleshoot optimization problems and issues

Course Topic Details

Fundamentals

- ✓ Spark execution model: Driver/Executors
- ✓ Spark resource managers (YARN, MESOS, K8s)
- ✓ Understanding RDDs/DataFrames APIs and bindings
- ✓ Difference between Actions and Transformations
- ✓ How to read the Query plan (Physical/Logical)

Spark Internals

- ✓ Spark Memory model
- ✓ Understanding persistence (caching)
- ✓ Catalyst optimizer and Tungsten project
- ✓ Shuffle service and how is shuffle operation executed
- ✓ Concept of fair scheduling and pools
- ✓ Java and Kryo serializer
- ✓ Step into the JVM world: what you need to know about GC when running Spark applications

Spark optimization: main problems and issues

- ✓ The most common memory problems
- ✓ Benefit of using early filtering
- ✓ Understanding partition and predicate filtering
- ✓ Join optimization
- ✓ Combating data skew (preprocessing, broadcasting, salting)
- ✓ Understanding shuffle partitions: how to tackle memory/disk spill
- ✓ Downside of using UDF's
- ✓ Executor idle timeout
- ✓ Data formats examples

Moving to production

- ✓ Debugging / troubleshooting
- ✓ Productionizing your Spark application
- ✓ Dynamic allocation and dynamic partitioning
- ✓ (REMOVED) Profiling your Spark application (Sparklint)
- ✓ JVM Profiler

Program**Day 1****Fundamentals**

- ✓ Spark execution model and Spark resource managers
- ✓ Understanding RDDs/DataFrames APIs and bindings
- ✓ How to read the Query plan (Physical/Logical)

Spark internals

- ✓ Spark Memory model and Understanding persistence
- ✓ Catalyst optimizer and Tungsten project
- ✓ Shuffle service and how is shuffle operation executed
- ✓ Concept of fair scheduling and pools
- ✓ Java and Kryo serializer

Day 2**Spark optimisation: main problems and issues**

- ✓ Understanding partition and predicate filtering
- ✓ Combating Data skew and Understanding shuffle partitions
- ✓ Downside of using UDF's and Executor idle timeout

Moving to production

- ✓ Debugging / troubleshooting
- ✓ Productionizing your Spark application
- ✓ Dynamic allocation and dynamic partitioning
- ✓ Profiling your Spark application (Sparklint) and JVM profiler

[More info and dates](#)

Streaming Architecture at Scale



 2 days  Specialty

While some apps enjoy the luxury of processing in a batch oriented fashion, others, as in the IoT ecosystem, expect events to be ingested and processed as they occur. This training focuses on two key players on the streaming-side of data processing: Apache Kafka and Apache Spark.

This training IS for you if...

- ✓ You are an IT engineer/architect, who deals with data stream processing architectures.

This training is NOT for you if...

- ✗ You don't have basic experience with Python and Apache Spark, which is required. If you're not quite there yet, we recommend the Python for Data Engineers and Data Processing at Scale courses as preparation for this training.

After the training, you'll be able to

- ✓ Understand the fundamentals of queue messaging systems
- ✓ Understand the fundamentals of the Kafka architecture
- ✓ Understand the fundamentals of Spark Streaming, with concept as checkpointing, watermarking, streaming windows and more
- ✓ Route real-time incoming events with Apache Kafka
- ✓ Process events in real-time with Apache Spark

Course Topic Details

Kafka

- ✓ How streaming topics work
- ✓ The basics of messaging systems
- ✓ Watermarks
- ✓ The concept of topics
- ✓ Design considerations for a messaging system
- ✓ Run a Kafka cluster as docker-compose

Spark

- ✓ Set Spark as a consumer for Kafka
- ✓ Process incoming events real-time

Program

Day 1

- ✓ How streaming topics work
- ✓ The basics of messaging systems
- ✓ Watermarks
- ✓ Run a Kafka cluster as docker-compose

Day 2

- ✓ Set Spark as a consumer for Kafka
- ✓ Process incoming events real-time

More info and dates





In-company Data and AI Learning Journeys

Do you want to tackle data issues with the people who understand them best? With our in-company programs, you learn from expert trainers and your colleagues. We give your organization the most effective training programs so your employees can be even more successful data professionals.

Our Approach

A successful in-company program depends on more than the course itself. With our four-step approach, we ensure that your organization achieves maximum value.

(1) Intake

We conduct an intake to discuss the project, challenge, or issues you want to tackle. During the intake, we also explore any ideas you may already have about the execution of the in-company program.

(2) Advice and offer

Based on your needs, our team gets to work in creating a comprehensive program. Together, we refine and adjust it.

(3) Execution

We discuss the specifics of the program with the trainer and also send an intake form to all participants so we know their knowledge level and learning goals. Our academy coordinator provides the necessary course material and participants improve their skills and gain knowledge through theory, hands-on labs, and discussions.

(4) Follow-up

After the training, we conduct an extensive evaluation and use this input to improve the program. We continuously update our courses to ensure that they are current with the latest information, techniques, and technologies.

After the program, we can support your organization and your employees in getting even more out of the courses. This can be achieved through coaching, follow-up days, or more in-depth and specialized additional courses.

The 7 Benefits of In-Company Courses

(1) Adapted to your situation

With an in-company program, we can adjust the course material to your needs, challenges, and specific situation. This makes the course even more applicable.

(2) Extensive offering of over 25 courses

All our public classes are available as in-company courses. Next to that, we have a host of courses exclusively available for in-company programs. Do you have a data and AI challenge but don't see the right course? With our varied expertise, we can shape the right course for you.

(3) Best trainers

All our trainers are practitioners in the field. Through their work as consultants, they solve challenges like yours every day.

(4) You decide on the right place and time

You decide what works best for your employees — in-house or external. Want a specific date? No problem, it's up to you.

(5) Financially attractive

Our in-company programs are available starting at six participants, making them cost-efficient.

(6) High quality, with an average score of 8.7

Participants in our courses are highly satisfied with the quality of their learning.

(7) Intensive guidance throughout the process, from start to finish

We collaborate closely to shape the program you need. Our skills advisors, academy coordinators, and trainers support you every step of the way.

From Team to Boardroom – Data and AI Courses for Your Entire Organization

Embedding data-driven decision-making in your organization by leveraging AI products needs more than just a good data team. You cannot become fully data-driven without executives sponsoring use cases and leading by example. Projects fall short when the business challenge is not connected to the technical one by data science product owners. Your general workforce requires to be open to data and AI and knowledgeable about the changes they bring. That's why we provide programs for your entire workforce.

C-Level and Executives

There can be no project without buy-in, and to have buy-in, your C-level and other executives need to understand the possibilities and pitfalls of AI. Our programs provide them with a deep understanding of data and AI, but also what it takes to generate successful use cases and their role in sponsoring these. This way, AI solutions can be developed more quickly, but also with a better focus on value generation.

Data Engineers and Data Scientists

We deepen and widen the skills of your data engineers and data scientists. With our programs focused on Python, Spark, Airflow, and Docker, they will increase their ability to effectively and efficiently use these critical techniques.

With courses such as Deep Learning, A/B Testing, Time-Series, and Recommenders, they acquire more specialized knowledge and skills.

Data Science Product Owners

Whether you call them data science product owners, analytics translators, business translators, or AI consultants—they all support the connection between data teams and the business.

We teach these product owners how to generate AI use cases, check the value of a project every step of the way, and ensure the business fully adopts and uses AI solutions.

General Workforce

Your general workforce might be the most critical part of your data-driven transformation. Without support from the business, there can be no success in data and AI. They are the people that will drive the adoption of the data products and ultimately change the way you do business.

Empowering them with knowledge of data and AI and informing them about the possibilities creates analytics ambassadors throughout your entire organization. We do this with online courses, workshops, and board games, to name a few of the options. Employees will be able to generate powerful ideas on what data and AI products can enhance their work and how to work with data science product owners on generating use cases.



In-company Courses and Programs

Are you looking for a course to teach your data engineers more about Apache Airflow? Or do you need a full program to take your organization's data scientists from juniors to seniors?

At Xebia, we have the right expertise to offer the flexibility you need. We are happy to show you some examples of our work.



Data Engineering and Data Science Accelerator – ING

For ING, we executed the full data engineering and data science learning journeys described earlier in this guide. With a 13-month/13-modules program, ING's data professional went through an entire journey from junior to senior. Each of the 13 modules featured one day of theory and one day of hands-on labs. The program's breadth covered the full spectrum of relevant knowledge for data engineers and data scientists.



Data Science Specialization Training – ABN AMRO

ABN AMRO's wide variety of products and departments creates a training challenge. To ensure applicability and benefits for all of its data scientists, the training program was set up as a set of single-day courses focusing on several data science specializations, including reinforcement learning, recommender systems, and data science engineering.

Data Discovery Workshops – Schiphol and ING

To educate C-level executives, managers, and the general workforce on the "art of the possible" for data and AI, Xebia regularly hosts Data Discovery Workshops. In these workshops, participants get to work on creating data and AI use cases they can apply in their work. For ING, we organized these workshops as a worldwide program for their C-level executives. At Schiphol, we joined the internal innovation week, showing managers and employees what they can achieve with data and AI.



International Data Science Program – DSM

Chemical enterprise DSM asked Xebia to design an international program to enhance its internal data science capabilities. With courses taught in the Netherlands, the United States, and Switzerland, the program contained foundational, intermediate, and advanced courses on Python. Some of the classes came from Xebia's standard offerings, while others were created especially for DSM.



Apache Airflow Ask Me Anything – Bosch

Bosch Germany was searching for possibilities to improve Apache Airflow expertise within the company. As Bosch's data engineers already possessed solid knowledge of Apache Airflow, they did not just want a training course. Bosch asked Xebia for the best approach, and together we decided to send our most senior Apache Airflow expert. This expert was on-site at Bosch for two days to answer all questions about Apache Airflow and give hands-on support on active projects.



Advanced Apache Spark Users – Schiphol

In addition to training programs, Xebia worked with Schiphol on a data engineering consultancy. During this consultancy, Schiphol employees learned the finer details of Apache Spark and were eager to further improve their skills. Xebia, therefore, provided Schiphol's data engineers with an Advanced Apache Spark course, going far beyond the standard Data Science with Spark course.

Supporting bol.com as a Frontrunner in Data and AI

Customer Challenge To enhance in-house data engineering and data science capabilities and remain a frontrunner in data and AI.

Our Solution An extensive in-company training program comprised of 15 to 20 courses per year, running since 2018.

Outcome Professionals working at bol.com maintain up-to-date skills with the latest data engineering and data science technologies. They can apply the knowledge they acquire in training to their day-to-day work, supporting bol.com's business.

Data for Decision-Making First

Founded in 1999 as a pure internet company, bol.com doesn't have the baggage of traditional organizations in becoming data-driven. The company's advanced analyst Melissa Perotti explained, "Decision-making at bol.com is based on data first. The data doesn't confirm someone's hunch; it tells us where to go."

As the largest e-commerce platform in The Netherlands and Belgium, bol.com has vast amounts of data for decision-making at its fingertips. "We use data science in many places—our search engine and recommender systems, forecasting of sales and customer service, and our chatbot. We have a mature data landscape and are proud of what we have achieved so far—and we're ready to take it all the next level."

Over the past two years, bol.com's data science capability quickly matured and now the shopping platform's next challenge is to extend this beyond its data engineers and data scientists.

Great Fit with Xebia

For the data engineering and data science training program, bol.com looked for a partner to support their journey. The two main criteria for its program were applicability and being up-to-date with the newest theories and models. After comparing different partners, bol.com chose Xebia. "Xebia understood how we saw our data science capability, and where we wanted to go," said Perotti. "It was also really

flexible in setting up the programs, so the courses were perfectly tailored to our situation, and the material was very applicable in our work."

A Wide Host of Courses

The in-company training program offered a wide range of courses to data professionals at bol.com. Courses included programs for data science starters like Python for Data Analysts, but also more advanced courses such as Deep Learning.

Most of the courses are part of the curriculum Xebia offers and were adapted to fit bol.com's needs. For example, the A/B Testing and Experiments course was split into a technical and a non-technical training. The technical training focused on data scientists and ways they could apply experimenting in their day-to-day work. The non-technical training was aimed at managers, to give them an insight in the possibilities.

Extending the Training Program

After providing several courses for bol.com in 2018 and 2019, the program has been extended into 2020 with several additional courses.

Xebia has been heavily involved in setting up and supporting the program. "What is great about the program with bol.com is that we really work together on the execution," he explained. "After every training, we conduct an extensive evaluation to see if there are things we can improve from either side."

bol.com®



Data and AI Accelerator Program for Enterprises

Transforming into a Data-Driven Global Enterprise

As an enterprise, how do you become fully data- and AI-driven? Ingraining data and AI into the DNA of your employees' way of working is an immense challenge. We can help you climb that mountain.

Gain Hands-On Experience with Data and AI

In collaboration with you, GoDataDriven shapes and executes this program, from baseline assessment to delivering training and executing AI use cases.

What makes the program so effective and unique is the combination of learning methods. It's more than a classroom or online course. All participants go hands-on with data and AI and create their own use cases.

For the Entire Workforce

The Data and AI Accelerator Program is specially designed for enterprise organizations that want to upskill their entire workforce. It offers learning journeys, courses, and certifications for every type of employee.

Whether it's C-level executives, data scientists, or employees who rarely touch data, this program covers everyone's training needs.

Why Enterprises Love the Data and AI Accelerator Program

Program Design from Start to Finish

We design the full Data and AI Accelerator Program with you. The program's design is completely tailored toward your organization's goals, challenges, and capabilities.

Authority From the Field, Like No Other

All our trainers work in the field as consultants. They experience and solve challenges like yours every day. Your organization will benefit enormously from this experience. We don't come from an abstract background — we are practitioners as much as we are teachers.

High Satisfaction Rate

The Data and AI Accelerator Program at ING was an overwhelming success. Participants rated their training courses with an 8.5.

The program enabled ING to establish best practices and a standardized way of working, letting ING tackle AI challenges better than ever.

Learn Today, Apply Tomorrow

Applicability of our training courses is a core GoDataDriven value. Our programs are designed so your employees can apply their new knowledge immediately. In the Data and AI Accelerator Program, every trainee participates in data and AI use cases, giving them direct hands-on experience and a deeper understanding.

Five Elements of the Data and AI Accelerator Program

The Data and AI Accelerator Program is designed using a set framework. This allows for both quick implementation and flexibility to tailor the program to your needs.

(1) Assessment and Goal-Setting

We assess the baseline capabilities of your workforce, determine the personas, and set goals for the desired end result.

(2) Learning Journey Design

With the start and end points determined, a learning journey is designed for each persona. This learning journey includes all types of training, examination, and certification, together with governance and training logistics.

(3) Curriculum Creation

We review and update our existing material and create new material if required. To ensure the material fits with your employees, we field-test it with selected colleagues.

(4) Program Execution

Our trainers work on upskilling your employees to become truly data- and AI-driven. The program includes a mix of classroom and online training, meetups, and hackathons. We constantly update the course material, so your employees will always have the latest information.

(5) Evaluation and Improvements

After each session, participants fill out an evaluation. Based on their feedback, we evaluate the session and improve future sessions.

Securing Budget for Your Training Course

Throughout this guide, we have emphasized the necessity of staying up-to-date for data professionals. The Internet of Things, the growing power of cloud providers, ever-changing technologies, specialization—these are all valid and urgent reasons to develop your skills. However, these reasons are also rather generic. If you want to free up a training budget at your organization, what should be your approach? How do you make your case clear and convincing?

Need a helping hand in preparing your case? Get in touch with our director of sales, Marcel van Denderen, at marcel.vandenderen@xebia.com or +31 6 3623 7853.

Talk to HR and Your Colleagues

The best place to start is HR. They will be able to tell you whether there is a policy for the training budget and any requirements. Sometimes it's as simple as putting in a request or picking one of the options in your company's training portal.

Does your organization not offer a standard training budget, or is there no policy? In that case, ask your colleagues. If they have attended a course in the past, they can tell you what their approach was.

Do Your Research

If freeing up training budget requires you to convince your manager, it's vital to conduct a proper investigation into the different options. You may feel you will benefit from a three-day course, costing around €1,800, but your manager might have a very different view.

Research the various options, and be open-minded about what you find. A €20 Udemy course might not be a great option, but clarify for yourself why it's not. This will help you convincingly present your case.

Prepare Your Case

It's essential to present a well-prepared case that shows why your preferred training option is best. Start your case with the conclusion—explain the value your proposed course will bring to the

company. How much money will it save, how much faster will you be able to deploy a model? Of course, your manager is interested in your development, but they also need to validate the expense. This is why it's important to emphasize its value for the company.

In addition to demonstrating the value, it's wise to show alternatives and how they relate to your preferred option. It may sound counterintuitive, but it's likely that your manager or HR will ask about potential alternatives. Being prepared and presenting sound reasoning for your choice only strengthens your case.

Show the Value

Validating your chosen course doesn't end when you receive approval. Before attending the course, have a clear plan for what you want to learn and how you intend to apply it. When returning to work, put your learnings into practice and show your manager the results you have generated and the value it has delivered for your organization.

Big Enough to Deliver, Small Enough to Care

At Xebia, we have been delivering data training courses since 2010. Over the past decade, we have dedicated a lot of time in laying the foundation for our unique curriculum and refining it with every iteration. Based on our own experiences, we are able to bring the practical skills professionals need to add value

every day. We deliver training world-wide for global enterprises, but make sure to provide a personal experience for every participant. Not only with content and delivery methods that seamlessly fits their needs, but also by delivering training in many popular languages.

Our training numbers



460

companies served



1,412

days of training



18,080+

people trained



8.9/10

satisfaction rating

Our Track-Record

We are full stack

We organize, we build, we train. Everybody is talking about data and AI. Our team has been implementing scalable data solutions since 2009.

In-depth knowledge

We contribute to Apache Airflow, Apache Kafka, Keras, Scikit-learn, pandas, Apache Avro, Apache Parquet, Apache Flink, Apache Spark and many more.

Learn today, apply tomorrow

Battle-tested during more than 10,000 hours of classroom training, the extensive, hands-on curriculum has been developed with applicability in mind.

Each trainer is a consultant

We bring real knowledge from the trenches. Xebia has the expertise — built through years of working with the top enterprises in Europe.

Our Global Training Delivery Capabilities

Our international group of expert trainers delivers classes in the most popular European languages. Our trainers are native in Dutch, English, France, German, Italian, Spanish.



Trusted by Global Leaders



Uber



Booking.com



AIRBUS

MERCK

TOMTOM



AIR FRANCE KLM

randstad



bol.com



ASML



rbi reed business information



Amsterdam University of Applied Sciences



Trainer Overview



Alexander Bij



Cor Zuurmond



Daniel van der Ende



Herbert van Leeuwen



James Hayward



Juan Perafan Rodriguez



Julian de Ruiter



Kris Geusebroek



Lucy Sheppard



Lysanne Van Beek



Manolis Manousogiannis



Marysia Winkels



Nelli Gofman



Pádraic Slattery



Razvan Vacaru



Rens Dimmendaal



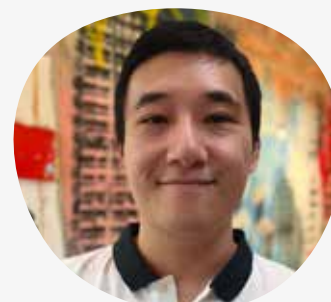
Roel Bertens



Rogier van der Geer



Steven van Duin



Tennyson Wu



Thomas van Latum



Timo Uelen



Travis Dent



Vadim Nelidov



Yke Rusticus

Xebia Academy Team for Data and AI Learning Journeys

Giovanni Lanzani – Managing Director Data

As managing director data of Xebia's Academy, Giovanni is responsible for the content of our courses and learning journeys. He also leads Xebia's own development efforts.

Marcel van Denderen – Sales Director

Marcel is responsible for all sales of the Xebia Data and AI Academy. If you want to improve the data and AI skills of your employees with one of our in-company programs, Marcel can advise you on the best setup.

Nico de Jonge – Head of Business Development

Nico is frequently in touch with organizations all over the world about their challenges in data literacy. If you have questions about upskilling your organization, team or self, Nico is happy to help you on the best way to move forward.

Diego Teunissen – Learning and Development Consultant

As a Learning and Development Consultant, Diego is responsible for the public courses on data and AI at Xebia Academy. The data field is dynamic like no other. Keeping your skills up to date is key to staying ahead. If you need advice on which training course fits your learning goals, Diego is always there to help you or your team further.



Giovanni Lanzani



Marcel van Denderen



Nico de Jonge



Diego Teunissen

Frequently Asked Questions

How can I get advice before booking a course?

Diego Teunissen, Learning and Development Consultant, or Marcel van Denderen, Director of Sales, will happily answer your questions and advise you on our courses.

You can reach them via email, or by phone:

Diego Teunissen,
diego.teunissen@xebia.com,
+31 6 1591 4440

Marcel van Denderen,
marcel.vandenderen@xebia.com,
+31 6 3623 7853

I'm interested in attending a course that is not on the current schedule. What do I do?

We plan training courses throughout the entire year, so if you don't see the course of your choice on the schedule, there's still a chance we could add it. Get in touch with Marcel van Denderen to discuss the possibilities via email, marcel.vandenderen@xebia.com, or by phone, +31 6 3623 7853.

Can I cancel a booked course?

If you are unable to attend a booked course, please send an email to academy@xebia.com. If you cancel up to one month before the first training day, there is no cancellation charge. Cancellations between one month and two weeks of the first day pay 50%. If you cancel a course within two weeks of the first training day, we charge the full price.

Do I receive a certificate after attending a course?

Yes, you will receive a Xebia Certificate of completion after completing a Xebia course.

Do I need to bring a laptop and install any software?

All of our courses include extensive hands-on sections requiring a laptop. The need to install software differs between courses. Please refer to the specific course information for details. You will also receive a confirmation email shortly before the training date that lists any requirements.

Training	Duration	Topic	Level	Page	More info at
A/B Testing and Experiments	2 days	Data Science	Fundamentals	27	gdd.li/ab-testing-exp
Advanced Analytics Translation	2 days	Analytics Translation	Advanced	14	gdd.li/advanced-analytics-translation
Advanced Data Science with Python	2 days	Data Science	Advanced	31	gdd.li/adv-ds
Advanced Power BI	2 days	Data Analysis	Specialty	20	gdd.li/da-advanced
Analytics for Executives	4 hours	Analytics Translation	Fundamentals	71	gdd.li/analytics-for-executives
AWS Cloud Fundamentals	1 day	Machine Learning Engineering	Fundamentals	45	gdd.li/cf-aws
Bayesian Statistics Training	2 days	Data Science	Fundamentals	28	gdd.li/bayesian
Big Data on AWS	3 days	Machine Learning Engineering	Advanced	14	gdd.li/aws-big-data
Certified Analytics Translation	2 days	Analytics Translation	Fundamentals	13	gdd.li/analytics-translation
Certified Apache Airflow	2 days	Data Engineering	Advanced	66	gdd.li/apache-airflow
Create Data Science Products	2 days	Data Engineering	Advanced	63	gdd.li/ds-products
Data Processing at Scale	3 days	Data Engineering	Fundamentals	48	gdd.li/data-scale
Data Science with Python (Certified)	3 days	Data Science and Machine Learning Engineering	Fundamentals	25/48	gdd.li/python-certified
Data Science with Spark	3 days	Data Science	Specialty	38	gdd.li/ds-spark
Data Visualization and Storytelling	2 days	Analytics Engineering and Data Science	Advanced	10/24	gdd.li/data-visualization
dbt Learn	3 half-days	Analytics Engineering and Data Engineering	Fundamentals	8/61	gdd.li/dbt-learn
Deep Learning	3 days	Data Science	Specialty	36	gdd.li/deep-learning
Deep Learning for Natural Language Processing	2 days	Data Science	Specialty	42	gdd.li/deep-learning-nlp
Docker	1 day	Machine Learning Engineering	Fundamentals	52	gdd.li/docker
Docker and Kubernetes	3 days	Data Engineering	Fundamentals	65	gdd.li/dockerkubernetes
Google Cloud Platform Fundamentals: Big Data and Machine Learning	1 day	Machine Learning Engineering	Advanced	53	gdd.li/gcp-big-data-ml
Google Cloud Fundamentals: Core Infrastructure	1 day	Machine Learning Engineering	Fundamentals	46	gdd.li/gcf-core-infrastructure
Innovation Through Experiments	1 day	Analytics Translation	Fundamentals	12	gdd.li/innovation-experiments
Kubernetes	1 day	Data Science and Machine Learning	Advanced	55	gdd.li/kubernetes

Machine Learning Explainability	1 day	Data Science	Advanced	34	gdd.li/ml-exp
Certified Microsoft Azure Fundamentals (AZ-900)	1 day	Machine Learning Engineering	Fundamentals	50	gdd.li/cf-azure
Microsoft Certified: Azure Data Scientist Associate	3 days	Machine Learning Engineering	Advanced	54	gdd.li/azure-data-science
Microsoft PL-300	3,5 days	Data Analysis	Advanced	18	gdd.li/power-bi-pl-300
MLOps Training	2 days	Machine Learning Engineering	Advanced	56	gdd.li/mlops-training
Optimizing Apache Spark and Tuning Best Practices	2 days	Data Science and Data Engineering	Specialty	40/68	gdd.li/optimizing-spark
Power BI in a Day	1 day	Analytics Engineering and Data Analysis	Fundamentals	9/17	gdd.li/da-power-bi
Production-Ready Machine Learning	2 days	Data Science and Machine Learning Engineering	Advanced	33/51	gdd.li/prod-ml
Python for Data Analysis	2 days	Data Science	Fundamentals	23	gdd.li/python-da
Python for Data Engineers	2 days	Data Engineering	Advanced	62	gdd.li/python-engineer
Streaming Architecture at Scale	2 days	Data Engineering	Specialty	70	gdd.li/arch-scale
Time Series Analysis and Forecasting	2 days	Data Science	Advanced	29	gdd.li/time-series



RIGHT
UGLY

10010
0000

01
000
010
100
010



RHYMES
Z-Z-Z





Questions about our data and
AI learning journeys?
marcel.vandenderen@xebia.com
+31 6 3623 7853