EXPLORE AI
ACADEMY

Data Engineering Prospectus

12 months, part-time

Contents

ExploreAI Academy overview

ExploreAI is a global data and AI solutions provider. We build AI-driven software for utilities, banks, insurers, and telcos. Within it is the ExploreAI Academy, whose mission is to transform the lives of talented African youth by equipping them with modern, relevant skills that enable them to find meaningful work.

We teach students the skills the global market is demanding, but that traditional education institutions are not producing. Our course content is curated and created by scientists with practical experience in the industry.

Our philosophy is to teach our students how to solve problems in the real world. We emphasise teamwork, collaboration and working within constraints and for deadlines. Although we cover the theory, we are not a theoretical institution. We are a 'practical, hands-on, roll-up-your-sleeves-and-get-stuff-done' kind of institution.

ExploreAI launched in 2013 and has since taught thousands of students and solved hundreds of problems for businesses across the world. We're reinventing education and invite you to join us.

Why data engineering?

A surge in demand for data engineers is in effect driven by four major points.

Big data - all in the cloud

Vast, and increasing, volumes of data are being generated every minute. Global providers store it all in the cloud where data engineers can make use of it.

Enable organisations

Data engineers enable organisations to transition to become data-centric by making data accessible and usable.

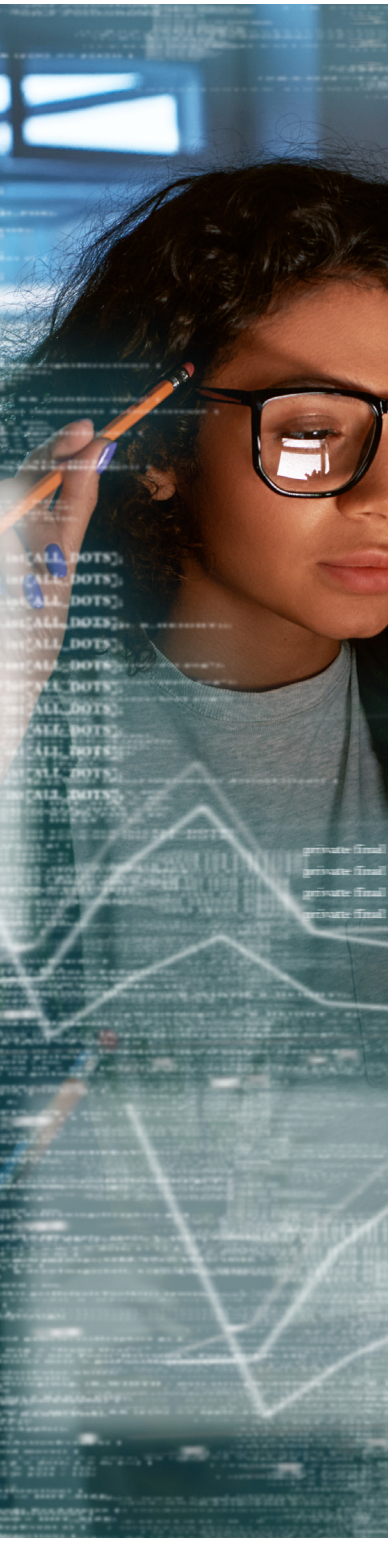
Drive important decisions

Business decisions depend on the quality and availability of data in the business environment, both of which depend on the work of data engineers in supplying it.

Lifeblood of analytics

Data engineering enables you to build data pipelines and environments which support model creation and analytics. Reliable, consistent, and accurate data are more valuable than any model.

Who is this course for?



You should consider doing this course if the following applies to you.

You want to improve your skills

You come from an adjacent or unrelated industry and want to futureproof your skillset. Or, you're in tech already but you feel your skills are out of date.

You want to learn new techniques in tech

The expert scientists at ExploreAI solve complex problems for big global companies. Their learnings are in turn used to refresh our Academy's course content at high velocity. The skills you'll learn here are thus modern, relevant, and used in production worldwide.

You prefer to have a support team when you learn

Our long courses are facilitated by subject matter experts who are available throughout the course duration.

You want to solve real-world problems

We use an agile, project-based approach that immerses our students in the world of problem-solving and prepares them for the real world. You'll learn theory in each lesson, but quickly proceed to apply it.

Who are you?

You're a newcomer to data engineering and you want to close the gaps in your analytical skills and knowledge. You have a passion for hacking things together, common practice in data engineering when integrating the variety of tools on offer.

Or, you're a professional working in any of a range of business areas, including but not limited to marketing, sales, finance, and operations. You want to learn how to prepare, manage, and transfer data better such that the rest of your organisation can make use of it and operate more smoothly.

Or, you're new to the world of work and are intrigued by the technical nature of this course and the range of possibilities for a career in tech. How exciting!

How you'll learn

The course is broken up into manageable, weekly units called lessons.

- Work through downloadable content and online instructional material.
- Interact with your peers and facilitators through real-time chat platforms, the ExploreAI forum, and regular live webinars.
- Enjoy a wide range of interactive content, including video lectures, coding challenges, hackathons, and presentations.
- Investigate real-world case studies.
- Apply what you learn each week in quizzes, coding challenges, and ongoing project submissions, sharpening your ability to solve real-world problems.

Requirements

This course and its subject matter are technical in nature. It is recommended that you have a basic understanding of mathematics and statistics. Basic knowledge of at least one programming language is recommended but not required.

1 Basic requirements

You'll need to make sure you have:

- Basic computer literacy (using a web browser, operating an email account, spreadsheets, etc.).
- A current email account.
- Access to a computer, the internet, and PDF reader software.
- Access to the Google office productivity apps (Docs, Sheets, Slides – freely accessible to anyone with a Google account) or Microsoft's Office apps (Word, Excel, PowerPoint).
- Google Chrome to access the learning management system, though any popular browser should suffice.

2 Technical requirements

- OS: Windows 10 recommended (Windows 7 minimum), in order to use Power BI; MacOS running Parallels for Windows will also suffice.
- Processor: Minimum i3, with a minimum clock speed of 2 GHz.
- RAM: Minimum 4 GB.
- Internet: A 10 Mbps line speed and 20 GB of data per month.
- Communication hardware: Webcam and microphone.

3 Additional requirements

Please note that Google, Vimeo, YouTube, Udemy, and DataCamp may be used in our course delivery, and if these services are blocked in your jurisdiction or on your device, you may have difficulty accessing course content.

Please check with us before registering for this course if you have any concerns about access restrictions affecting your experience with our learning management system.

EXPLORE AI ACADEMY

Your success team

A range of experienced faculty members is at hand during working hours (8 am to 5 pm, Central African Time, Monday to Friday) to assist you throughout your learning journey.

Administrative support

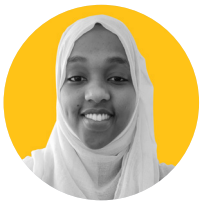
To address your technical and administrative queries. Your support team includes, among others:

Jamie Snyders

Jamie, a leading facilitator at EA, holds a BSc in theoretical physics and went on to do honours in applied mathematics. He's perfectly suited to help deliver our technical course content, giving students the best chance of success.

Chris Barnett

Chris hails from a rich academic background and holds BSc, MSc, and PhD degrees in chemistry. His rigorous research career positions him perfectly to deliver and provide support in our technical data qualifications.



Maryam Hassan

Maryam is an expert in delivering technical education. She holds BSc and Msc degrees in Medical Bioscience and Bioinformatics, respectively, and boasts four years of tutoring and teaching assistance experience at a university level.

Course convenors

Data science subject matter experts applied, and continue to apply, their practical knowledge and real-world experience to build the course content.

Jaco Jansen van Rensburg

Jaco is the curriculum director of the ExploreAI Academy. He holds an MBA and a PhD in mechanical engineering, for which he focused on mathematical modelling and optimisation.

Vincent Le Roux

Vince is a technical lead in the enterprise division of ExploreAI. Busy with his PhD in engineering and with a deep knowledge of big data, Vince ensures our data engineering content is of the highest quality.

Nthikeng Letsoalo

Nthikeng is a computer science master's student turned data scientist. He brings a deep knowledge of supervised and unsupervised learning to our EA courses.

Andries van der Walt

Andries holds an MSc in bioinformatics. He joined ExploreAI as a data scientist and is an alumnus of our data science qualification. He is now a Senior Data Engineer in ExploreAI's enterprise division.

EXPLORE AI ACADEMY

Curriculum overview

This course provides students with the knowledge, skills, and experience to work as a data engineer. We'll look at how to identify and state problems clearly, learn how to code effectively in SQL and Python, and get to grips with a variety of AWS tools used to work with data in the cloud. We'll do a deep dive into all the technicalities and tools used to handle big data. We'll analyse and identify the storage and processing requirements for data, gain the knowledge to build end-to-end data pipelines, and develop an understanding of how to deal with large datasets in a distributed manner.

Duration: 12 months
Pre-requisite skills: Basic analytical background
Course difficulty: Advanced
Tools learned: Python, Jupyter Notebooks, MySQL, AWS, Apache Spark

Phase	Module	Duration (Weeks)	Recommended time (Hours)
Fundamentals	Explore101	2	30
	Python	9	135
	SQL	6	90
	Cloud computing	7	105
Machine learning	Storing big data	7	105
	Moving big data	7	105
	Processing big data	7	105
Consolidation	Integrated exams and certification requirements	3	45

Breaks in the delivery schedule vary to accommodate major public holidays and recovery between each 2-3 modules delivered.

EXPLORE AI ACADEMY

Module 1

Explore101

Duration: 2 weeks

Recommended time: 30 hours

What is covered in this module:

Orientation

- Setting up your learning environment
- ExploreAI teaching philosophy and educational support framework
- An introduction to modern data science

Problem-solving

- Mutually exclusive and collectively exhaustive statements and decisions
- Design thinking and the scientific method
- Introduction to solution-oriented communication

The ExploreAI Data Product Framework

- Defining a problem statement
- Understanding the solution landscape and equation of value
- The basics of project management

Ways of work

- Business and systems theory
- Professional workplace conduct and practitioner meetings
- An introduction to agile development and delivery

Module 2

Python

Duration: 9 weeks

Recommended time: 135 hours

What is covered in this module:

Python programming basics

- Working in a Notebook environment
- Pseudo code and debugging concepts
- Working with primitive data types – variables, strings, integers, floating points, booleans

Functions and control flow

- Creating and working with functions
- Conditional statements
- For loops and while loops

Data structures

- Lists, tuples, sets, and dictionaries
- Working with DataFrames
- Plots and graphs

Exploratory data analysis

- Statistical measures, probabilities, and hypotheses
- Algorithms and algorithmic complexity
- Advanced interactive visual analysis

EXPLORE AI ACADEMY

Module 3

SQL

Duration: 6 weeks

Recommended time: 90 hours

What is covered in this module:

Introduction to SQL

- Working with databases
- Basic SQL data types and calculations
- Aggregating, sorting, and grouping data

Relational database design

- SQL schemas and entity relationships
- Table normalisation, primary and foreign keys
- Common table expressions and views

SQL in practice

- Set theory and SQL joins
- Nested and subqueries
- Improving query performance

Data manipulation

- Cleaning and analysing data
- Working with numeric, time, and string data types
- Data transformations and anomalies

Module 4

Cloud computing

Duration: 7 weeks

Recommended time: 105 hours

What is covered in this module:

Cloud computing basics

- Introduction to cloud computing concepts
- Pros and cons of cloud computing
- Popular cloud service providers

Introduction to Amazon Web Services

- Overview of AWS services
- Networking and content delivery
- Economics and billing

Storage and compute resources

- Databases and object storage
- Virtual machines
- Serverless compute resources

Cloud best practice

- Security, identity, and compliance
- Cloud architecture framework
- Automatic scaling and monitoring

EXPLORE AI ACADEMY

Module 5

Storing big data

Duration: 7 weeks

Recommended time: 105 hours

What is covered in this module:

Databases

- Relational and non-relational databases
- Data warehouses
- OLTP and OLAP

Storage

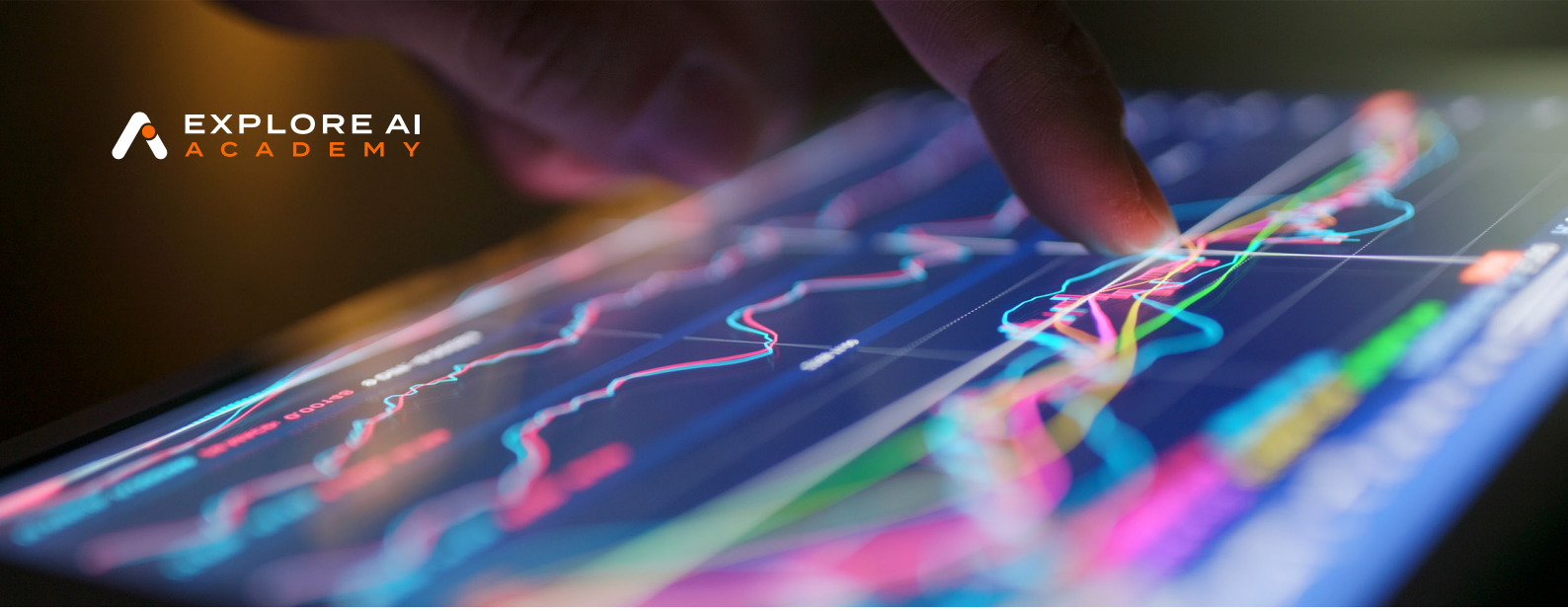
- Block storage and caching
- Data lakes
- Legacy file stores

Static data

- APIs and IoT devices
- File transfer protocols
- Connecting on-premise and cloud-based solutions

Streaming

- Why stream data?
- Retrieving data from source systems
- Making data available: queues and streams



Module 6

Moving big data

Duration: 7 weeks

Recommended time: 105 hours

What is covered in this module:

Pipelines

- Orchestrating pipelines
- Sources and dependencies
- Run pipelines based on specific events

Triggers

- Setting up pipeline triggers
- Batch and continuous processing
- Event metadata and parameters

Logging

- Pipeline health metrics
- Setting up, storing, and accessing logs
- Reporting pipeline runs and downstream impact

Monitoring and alerts

- Why and when to alert end users
- Cloudwatch Logs Insights
- Build an alert dashboard

EXPLORE AI ACADEMY

Module 7

Processing big data

Duration: 7 weeks

Recommended time: 105 hours

What is covered in this module:

Spark, Hive, and Hadoop

- Efficient ways of storing and accessing big data
- Overview of popular big data frameworks
- Set up a cluster that is “big data ready”

Data transformations

- Massively parallel processing (MPP)
- Partitioning and indexing big data
- Optimising workloads on Spark clusters

Data lineage and quality

- Governance documentation and frameworks
- Managing and tracking data
- Surfacing data quality

Data security

- High availability
- Data masking, encryption in transit and rest
- Disaster recovery

Module 8

Integrated exams and certification requirements

Duration: 3 weeks

Recommended time: 45 hours

What is covered in this module:

Review

- Programme recap
- Opportunity to review content in preparation for exams
- Understanding the final assessment plan

Integrated examination

- Consolidated theory exam
- Practical programming assessment
- Applied big data exam

Portfolio of evidence

- Compile evidence to illustrate competence
- Finalise assessment and moderation
- Provide feedback on the programme

Certificate admin

- Confirmation to be assessed
- Declaration of authenticity
- Understanding the appeals procedure

EXPLORE AI ACADEMY

Start a new career today by enrolling in one of our data science or data engineering courses.

Admissions
admissions@explore.ai

General
admissions@explore.ai

ExploreAI Academy
explore-datascience.net

ExploreAI Academy is a Level 3 B-BBEE company and a MICT SETA-accredited learning institute. Registration number ACC/2017/01/007.

Corporate sales enquiries
enterprise@explore.ai