

FULL CURRICULUM • 2026

Data Engineering

Build the pipelines that every AI and analytics team depends on.

Every module, every lesson — the complete syllabus, end to end.

47

MODULES

259

LESSONS

8

MONTHS

₹33,000

FULL PROGRAMME

What this course covers

1

Level 1 • Beginner

Introduction To Data Engineering · Setting Up the System · Introduction to SQL Queries

2 months · 14 modules · 77 lessons · ₹8,000

2

Level 2 • Intermediate

Python Collections for Data Engineering · Data Processing using Pandas · Project1: File Format Converter

3 months · 15 modules · 122 lessons · ₹11,000

3

Level 3 • Advanced

Aggregations using Spark SQL · Joins using Spark SQL · Sorting using Spark SQL

3 months · 18 modules · 60 lessons · ₹14,000

The whole journey

This is the complete syllabus for Data Engineering — 47 modules and 259 lessons, in the order you will learn them. The level markers show how far each stage carries you, but the curriculum runs as one continuous path from the first lesson to the last.

1

LEVEL 1

Beginner

Introduction To Data Engineering · Setting Up the System · Introduction to SQL Queries

2 months · 14 modules · 77 lessons · ₹8,000

- ### 1. Introduction To Data Engineering

6 lessons

 - Introduction to Data engineering
 - What is SQL
 - Overview of Application Architecture and RDBMS
 - Overview of Database Technologies
 - Purpose Built Database
 - RDBMS vs Database Technologies
- ### 2. Setting Up the System

3 lessons

 - Setting up VS code and Python
 - Setting up MySQL Server and PostgreSQL
 - Setting up Jupyter Lab online
- ### 3. Introduction to SQL Queries

3 lessons

 - Creating the table in SQL
 - Relationship Model In SQL
 - Starting with SQL
- ### 4. Manipulation Commands in SQL

7 lessons

 - Insert Command
 - Update part1
 - Update part2
 - Delete command
 - Drop and Truncate
 - Alter Part1
 - Alter Part2
- ### 5. Retrieving Data in SQL

4 lessons

 - Retrieving data using Select and where
 - Sorting and using Alias
 - Limit and Select with Expression
 - Logical Operators
- ### 6. Aggregation in SQL

2 lessons

 - Aggregation function part1
 - Aggregation Function part2

○	7. Joins in SQL	7 lessons
	• Types of joins • Inner join • Left Join • Right Join • Full Outer Join • Self join • Cross join	
○	8. Set Operators in SQL	2 lessons
	• Set Operator Part1mp4 • Set Operator Part2mp4	
○	9. Subqueries and Case Statements	7 lessons
	• Subqueries Part4 • Subqueries Part1 • Subqueries Part2 • Subqueries Part3 • Case Statements Part1 • Case Statements Part2 • Case Statements Part3	
○	10. CTEs and Views	4 lessons
	• CTEs Part1mp4 • CTEs Part2 • Views Part1 • View Part2	
○	11. Advanced Topics	8 lessons
	• Window Function • RowNumber(), Rank(), DenseRank() • NTILE(), Lead(), Lag() • Window Functions examples • Stored Procedure • Index • Mini Project Part1 • Mini Project Part2	
○	12. Introduction to Python Programming	5 lessons
	• Introduction to Python • Variables and Data Types • Input and Output • Basic Operator • Constants and Type casting	
○	13. Data Structures in Python	8 lessons
	• Conditional Statements • Loop Part1 • Loop Part2 • Loop Part3 • List • Tuple and Dictionary • Sets • Mini project	



14. Functions, Modules & File Handling

11 lessons

- Functions Part1
- Functions Part2
- Functions Part3
- Project Calculator Part1
- Project Calculator Part2
- Modules Part1
- Modules part2
- Module Part3
- File Handling Part1
- File Handling Part2
- Error Handling

2

LEVEL 2

Intermediate

Python Collections for Data Engineering · Data Processing using Pandas · Project1: File Format Converter

3 months · 15 modules · 122 lessons · ₹11,000



15. Python Collections for Data Engineering

10 lessons

- File I_O using Python
- Read Data from CSV into python
- Python Collectionsmp4
- Processing Python List
- Lambda Functions
- Filter data in Python
- Manipulating List using set()mp4
- Sort() & Overview of json strings
- Reading data from json files
- Extracting column from json file



16. Data Processing using Pandas

10 lessons

- Overview of pandas for Data Processing
- Pandas Dataframe functionalities
- Aggregation using pandas
- Joining data using pandas
- Summarizing Joins
- Aggregation on Join Results Part1
- Aggregation on Join Results Part2
- Aggregation on Join Results Part3
- Sort data using pandas dataframe
- Exporting data to JSON files



17. Project1: File Format Converter

12 lessons

- Overview of the files to be processed
- Working with regular expressions
- Getting Column names from the data
- Generating File paths and writing dataframe
- Recap of the Project
- Writing pandas dataframe to JSON file
- Converting csv to json
- Modularize File Format Converter
- Wrapping up the Project
- Runtime Arguments part1
- Runtime Arguments part2
- Environment variables

- 18. Project2: File to Database Loader** 5 lessons

 - Configuring sql to python
 - Connecting to the database
 - Validate pandas and SQL integration
 - Write CSV data to tables in chunks
 - Project ready for deployment

- 19. Troubleshoot and Debugging Python Issues** 6 lessons

 - Troubleshooting and Debugging
 - Troubleshooting Database connectivity
 - Troubleshooting Credentials for Database connectivity
 - Troubleshooting Errors in Python
 - Overview of SDLC
 - Unit Testing & Debugging

- 20. Overview of Performance Tunning in Python** 9 lessons

 - Introduction to Performance of Python Applications
 - Setting up Database Loader
 - Cleaning the data and Performance Tunning
 - Performance Tunning
 - Loading Multi Files in Python
 - Develop applications for Multiprocessing
 - Multiprocessing using python
 - Introduction to Parallel Processing
 - Executing Parallel Processing

- 21. Getting started with GCP** 5 lessons

 - Introduction to Getting started with GCP
 - Signing Up for GCP Account
 - Overview of Google cloud platform
 - Overview of google cloud shell
 - Overview of Analytics services on GCP

- 22. Overview of Big Data and Data Lakes** 9 lessons

 - Database and Their Types
 - Technologies for Different Databases
 - Usecase of different databases
 - Volume of different database
 - Overview of Big data and its evaluation
 - Data Lake using Hadoop
 - Overview of Modern Data Lakes on Cloud
 - Implementation of Modern Data Lakes
 - Implementation + Advantages of Modern Data Lakes

- 23. Overview of Spark and Spark Architecture** 8 lessons

 - Overview of Data Processing
 - Setting up the Environment
 - Code Examples of different Libraries
 - Difference between libraries and Distributed Computing
 - Example of Distributed Computing
 - Overview of Apache Spark Documentation
 - Spark Key Features and Infrastructure
 - Overview of Spark clusters and Key terms

- **24. Setup Databricks Environment using GCP** 5 lessons
 - Overview of Databricks on GCP
 - Signing up for Databricks on GCP
 - Resolving Workspace errors and creating cluster
 - Setup Databricks
 - Databricks Features and Architecture

- **25. Basic Transformation using Spark SQL** 7 lessons
 - Processing Data in databricks using spark
 - Getting started with Spark SQL
 - Create views and retrieve them using Spark SQL
 - Spark SQL Query to compute Daily Product Revenue
 - Overview of pyspark and Reading json files
 - Get Schema details using pyspark
 - Writing a function using Pyspark Python

- **26. Create Delta tables using Spark SQL** 4 lessons
 - Understanding delta tables and Database
 - Creating Delta tables
 - Copy data from external file
 - Performing operations in delta tables

- **27. Pre-defined Functions in Spark SQL** 29 lessons
 - Overview of Functions
 - Describe Functions in Spark SQL
 - Case Conversion Functions
 - Extracting Data using Functions
 - Trimming and Padding Functions
 - Reverse and Concatenation functions
 - Date Manipulation Functions
 - Beginning Date or Time Functions
 - Extracting information using date_format Part1
 - Extracting information using date_format Part2
 - Calendar Functions
 - Dealing with Unix Timestamp
 - Mathematical Functions Part1
 - Mathematical Functions Part2
 - Miscellaneous Functions Part1
 - Miscellaneous Functions Part2
 - Miscellaneous Functions Part3
 - CASE Statements Part1
 - CASE Statements Part2
 - Word count Query part1
 - Word count query Part2
 - Practice Exercise1
 - Practice Exercise2
 - Practice Exercise3
 - Practice Exercise4
 - Practice Exercise5
 - Practice Exercise6
 - Practice Exercise7
 - Practice Exercise8-

- **28. Setup Spark Tables for Basic Transformations** 1 lessons
 - Introduction to transformation using Spark SQL

- **29. Filtering Data using Spark SQL** 2 lessons
 - Filtering using Spark SQL part1
 - Filtering using Spark SQL Part2

Advanced

Aggregations using Spark SQL · Joins using Spark SQL · Sorting using Spark SQL

3 months · 18 modules · 60 lessons · ₹14,000

- **30. Aggregations using Spark SQL** 2 lessons
 - Aggregation using Spark SQL part1
 - Aggregation using Spark SQL part2
- **31. Joins using Spark SQL** 2 lessons
 - Joins using Spark SQL part1
 - Joins using Spark SQL part2
- **32. Sorting using Spark SQL** 1 lessons
 - Sorting using Spark SQL
- **33. Copy Query Results into Spark Tables** 3 lessons
 - Copy Query Results using CTAS and INSERT
 - Copy Query Results using MERGE
 - Design Pipeline using CTAS and INSERT
- **34. Ranking using Spark SQL Windowing Functions** 1 lessons
 - Ranking using Spark SQL
- **35. Processing JSON like data using Spark SQL** 4 lessons
 - Creating Tables with Array type Columns
 - Creating and Dealing with Array of Struct Type columns
 - Overview of Functions to Process Data in Spark SQL
 - Processing Delimited strings using Spark SQL
- **36. Getting Started with Pyspark DataFrame APIs** 4 lessons
 - Introduction and Setup of Pyspark
 - Transforming and Processing Schemas
 - Process schema details in json using pyspark
 - Convert CSV to Parquet with schema
- **37. Create spark data frames using Pyspark** 4 lessons
 - Creating Spark DataFrame and Processing json-like data
 - Advantages and Transformations of DataFrame
 - Overview of Data Processing and DataFrame concepts
 - using withColumn() and Writing Dataframe to Delta Files

- **38. Basic Transformation using Pyspark** 6 lessons
 - Overview of Transformations and withColumn usecase
 - Filtering Data using PySpark DataFrame
 - Aggregations using Spark DataFrame
 - Aggregation using Spark DataFrame APIs
 - Sorting data using PySpark DataFrame
 - Dealing with Nulls while sorting the data
- **39. Joining Data using Spark DataFrame** 2 lessons
 - Inner Join over the Spark dataframe APIs
 - Left and Right Join using Spark dataframe APIs
- **40. Ranking using Pyspark** 4 lessons
 - Overview of ranking along with OrderBy functionality
 - Filter Based on rank using Spark dataframe APIs
 - Filter based on Rank per Partition using spark Dataframe
 - Difference Between Rank() and Dense_Rank()
- **41. Integration of Spark SQL and Pyspark data frame** 3 lessons
 - Introduction to Integration of Spark SQL and PySpark Dataframe APIs
 - Run Spark SQL queries and Create Spark Metastore Tables using DataFrames
 - Dataframe Transformations over Metastore Tables
- **42. ELT Data Pipelines using Databricks** 3 lessons
 - Overview of Databricks Workflows
 - Multi-Task Workflows using Databricks jobs
 - Working on Multi-Task job
- **43. Performance Tuning of Spark-Catalyst Optimizer** 4 lessons
 - Getting Started with Performance Tuning and Spark Catalyst Optimizer
 - Overview of Plans using Spark UI
 - Understanding of Spark Architecture with Example
 - Filter, Broadcast Joins & Aggregations
- **44. Performance Tuning of Spark Cluster Configuration** 3 lessons
 - Introduction to Databricks Cluster Configuration and its Types
 - Setting up All-Purpose Cluster and Overview of Auto-Scaling
 - Cluster Performance Tuning using Auto Scaling
- **45. Performance Tuning while inferring schema from csv or json files** 3 lessons
 - Overview of Inferring Schema from CSV or JSON files
 - Overview of CSV_JSON files
 - Overhead of Inferring Schema

○ 46. Performance Tuning using Columnar File Format and Partitioning Strategy 2 lessons

- Introduction of Performance tuning while storing data and Side effects of CSV files in Data Lake
- Restructure CSV Data and performance Boosting

○ 47. Introduction to Linux Commands for Data Engineering 9 lessons

- Introduction to linux and Basic Navigation Commands
- File and Folder Management Commands
- Commands for Viewing and Editing Files
- Commands to make modification inside a Vi Editor
- Understanding of SED Command
- Understanding of GREP Command
- Understanding of AWK Command
- Understanding of Find Command
- Disk, Process & System Monitoring

Certificates & what comes next

You receive a certificate at the end of every level you finish. Join for a single level or the full programme — the choice is yours. Learners who stand out may be invited to work on live client projects as interns; these places are merit-based, limited, and depend on projects being available.