

Python for Data Engineering (3 Months)

Duration: 3 Months (~120 Hours)

Mode: Live Online / Classroom

Designation after completion: Python Data Engineer

Tools & Technologies: Python, SQLAlchemy, PySpark, Pandas, AWS SDK (Boto3), Airflow, Kafka

Certifications prepared: AWS Certified Data Engineer, Databricks Data Engineer

Detailed Syllabus (12 Weeks)

Week 1

- ETL basics
- Assignment → Parse CSV → JSON

Week 2

- SQL basics
- DB connection
- Assignment → CRUD ops

Week 3

- SQLAlchemy ORM
- Assignment → Employee DB queries

Week 4

- Batch processing with Pandas
- Assignment → Dataset cleanup
- Mock Interview 1

Week 5

- PySpark basics
- Assignment → Read big dataset

Week 6

- PySpark transformations
- Assignment → Transformation pipeline

Week 7

- PySpark SQL
- Assignment → Sales joins

Week 8

- Airflow basics
- DAGs & scheduling
- Assignment → ETL DAG
- Mock Interview 2

Week 9

- AWS Boto3 basics
- Assignment → Store files in S3

Week 10

- Kafka basics
- Assignment → Stream consumer

Week 11

- End-to-End ETL pipeline
- Assignment → Full ETL run

Week 12

- Capstone Project: Cloud ETL
- Mock Interview 3

Mini Projects

End-to-End ETL Pipeline
Cloud-based ETL Project