

Python for Data Validation, ETL & Reporting

Duration: 5 Days (40 Hours)

Target Audience: Data Engineers, ETL Developers, BI / QA Teams, Data Validation & Reporting Professionals

Day 1 – Database Connectivity & File System Handling (8 Hours)

Module 1: Python Environment & Database Connectivity

- Python environment setup (Anaconda / venv)
- Python DB connectivity concepts
- Connecting Python with:
 - Oracle Database
 - PL/SQL schemas
 - Teradata
 - SQL Developer (via Oracle DB)
- Libraries:
 - cx_Oracle / oracledb
 - pyodbc
 - sqlalchemy
- Secure credential handling
- Connection pooling & performance basics

Hands-On Labs

- Connect Python to Oracle & Teradata
- Execute SELECT, INSERT, UPDATE queries
- Fetch data into Pandas DataFrames

Module 2: File System Handling

- Working with different file formats:
 - Excel (xlsx, xls)
 - CSV (Delimited & Fixed-width)
 - TXT
 - XML
 - JSON
- Python libraries:
 - pandas
 - openpyxl
 - csv
 - json
 - xml.etree
- Handling large files
- Directory traversal & automation
- **Hands-On Labs**
- Read & parse different file formats

- Convert files into Pandas DataFrames
 - Handle fixed-width and malformed files
-

Day 2 – Data Loading & ETL Foundations (8 Hours)

Module 3: Data Loading using Python

- File-to-Database loading strategies
- Bulk insert techniques
- Handling data types & schema mapping
- Error handling during data loads
- Logging & audit columns
- Performance optimization for large datasets

Hands-On Labs

- Load CSV / Excel data into Oracle tables
 - Load JSON/XML data into relational tables
 - Implement rollback & exception handling
-

Module 4: ETL Design using Python

- ETL vs ELT concepts
- Extract → Transform → Load using Python
- Data cleansing & transformations
- Scheduling-ready scripts (enterprise readiness)

Mini Project – Phase 1

- Build a Python ETL script:
 - Extract from files
 - Transform data
 - Load into database
-

Day 3 – Data Validation Techniques (8 Hours)

Module 5: Data Validation Scenarios

- Validation approaches & best practices
- Validation types:
 - Database to Database
 - File to Database
 - Database to File
 - Database to Reports (MicroStrategy-style datasets)
- Row count validation
- Column-level validation
- Aggregate & checksum validation

Hands-On Labs

- Validate source vs target tables
 - Validate file data against database
 - Automate validation checks
-

Module 6: Cross-System Validation

- Connecting to multiple systems in one script
- Oracle ↔ Teradata validation
- File ↔ Database validation
- Generating validation summary reports

Mini Project – Phase 2

- End-to-end validation framework for:
 - File → Database
 - Database → Database
-

Day 4 – Data Quality Validation & Reporting (8 Hours)

Module 7: Data Quality Validation

- Data quality dimensions:
 - Completeness
 - Accuracy
 - Consistency
- Handling:
 - Null
 - Not Null
 - N/A
 - Unknown
 - Undefined
- Calculating:
 - % of invalid values
 - % of nulls per column
- Rule-based data quality checks

Hands-On Labs

- Python program for data quality metrics
 - Column-wise & table-wise DQ reports
-

Module 8: Report Generation on Validation Results

- Generating reports using:
 - Pandas
 - Matplotlib / Seaborn / Plotly

- Dashboards for:
 - ETL status
 - Validation summary
 - Data quality scorecards
- Exporting reports to:
 - Excel
 - PDF
 - HTML

Hands-On Labs

- Build a validation dashboard
 - Generate automated Excel & HTML reports
-

Day 5 – AI-Assisted ETL & Capstone Project (8 Hours)

Module 9: Python AI Libraries for ETL & Reporting

- Introduction to AI-assisted development
- Using AI for:
 - SQL query generation
 - Validation logic generation
 - Automated documentation
- Integrating Python with AI APIs
- Use cases in ETL & BI teams

Hands-On Labs

- Auto-generate SQL using AI
 - Generate validation scripts using AI prompts
 - Create automated result summaries
-

Module 10: Capstone Project (End-to-End)

Project Scope:

- Extract data from files (CSV/Excel/JSON)
- Load into Oracle / Teradata
- Perform:
 - Data validation
 - Data quality checks
- Generate:
 - Validation reports
 - Dashboards
- AI-assisted documentation generation