

Data Engineering

Learning Plan & Quotation

Target Role : Data Engineer

Goal : To acquire the necessary Python programming, data manipulation, and analytical skills to transition into a Data Engineering role.

Learning Philosophy

This plan is structured to provide a comprehensive and practical learning experience. Each level builds upon the previous, ensuring a solid foundation before moving to more complex topics. Hands-on exercises, real-world datasets, and mini-projects are central to reinforcing learning and building a portfolio. We will emphasize how software tools interact with and leverage underlying infrastructure, providing a unique perspective on data engineering.

Program Structure: Quick Start Data Engineering Essentials

Total Sessions : 8 sessions — 2 hours each

Session Rate : RM120 per hour

Cost per Session : RM240 per session

One-time fees : RM 200

Total Investment : RM 2,120

Detailed Module Breakdown

Module 1: Python & Pandas Fundamentals for Data (2 Sessions)

Session 1: Python Essentials - The DE Tool

1. Focus: Rapid Python basics (variables, data types, operations, I/O, functions)
2. Emphasis on scripting for automation and development environment setup
3. Hardware/Cloud Link: Python as an "instruction set" for CPU and OS resources

Session 2: Pandas - The Data Engineer's Spreadsheet

1. Focus: DataFrame fundamentals, loading CSV/JSON data
2. Essential operations: selection, filtering, data inspection (.info(), .describe(), etc.)
3. Goal: Proficiency in getting data into DataFrames and performing initial analysis

Module 2: Use Case 1 - Web Scraping for Data Acquisition (2 Sessions)

Session 3: Fetching & Inspecting Web Data

1. Focus: HTTP basics and GET requests using Python requests library
2. Critical skill: Browser developer tools for HTML element identification
3. Mini-Project Task 1: Successfully fetch target URLs and identify elements

Session 4: Parsing HTML & Initial Extraction

1. Focus: BeautifulSoup fundamentals (find, find_all, attribute selection)
2. Extract data points into structured formats (lists of dictionaries)
3. Save results to JSON/CSV files
4. Mini-Project 1 Outcome: Working web scraper for structured data

Module 3: Use Case 2 - Data Cleansing & Basic Transformation (2 Sessions)

Session 5: Identifying & Handling Messy Data

1. Focus: Common data quality issues (missing values, duplicates, inconsistent)
2. pandas methods for data cleaning (.dropna(), .fillna(), duplicate removal)
3. Mini-Project Task 2: Load and assess messy dataset quality

Session 6: Fixing & Transforming Data

1. Focus: Data type conversions and string manipulation in pandas
2. Conditional transformations and data standardization
3. Mini-Project 2 Outcome: Complete data cleansing script

Module 4: Use Case 3 - Cloud Storage, Simple Pipeline, Git & Next Steps (2 Sessions)

Session 7: Local "Cloud Storage" & Database Concepts

1. Focus: Cloud object storage concepts (S3/Blob Storage) and data lakes
2. Simulate "landing zone" with local directories
3. Introduction to relational databases and basic SQL (CREATE TABLE, INSERT)
4. DE infrastructure mapping to hardware and cloud services

Session 8: Building a Micro Pipeline & Professional Wrap-up

1. Focus: Connecting Python to SQLite database (sqlite3 module)
2. Integration: Build simple ETL script (Extract, Transform, Load)
3. Git version control introduction
4. Discussion: Next steps in DE journey (advanced Python, cloud services)
5. Mini-Project 3 Outcome: Complete data pipeline from raw file to database

Investment Summary

Component	Sessions	Cost per Session (RM)	Total Cost (RM)
Module 1: Python & Pandas Fundamentals	2	RM240	RM480
Module 2: Web Scraping for Data Acquisition	2	RM240	RM480
Module 3: Data Cleansing & Transformation	2	RM240	RM480
Module 4: Cloud Storage & Pipeline Development	2	RM240	RM480
One-time fees:			RM 200
Total Investment:			RM 2,120

Program Timeline

Estimated Duration: 8 weeks (one 2-hour session per week)

Flexible Scheduling: Sessions can be scheduled based on availability

What's Included

1. 8 intensive, 1-on-1 tutoring sessions
2. Comprehensive learning materials and practical code examples
3. Hands-on mini-projects with real-world applications
4. Digital resources including datasets and code templates
5. Session-focused support and guidance

Critical Success Factors

This accelerated program is designed for motivated learners who can commit to:

1. Independent practice between sessions
2. Self-study to reinforce concepts
3. Active engagement with mini-projects
4. Leveraging existing hardware background for deeper understanding

Next Steps

Upon completion of this program, client will have:

1. Practical experience with core data engineering tools
2. A portfolio of mini-projects demonstrating their skills
3. Clear understanding of next steps in DE journey
4. Foundation for more advanced topics
(cloud platforms, orchestration tools, advanced databases)

Ready to begin your Data Engineering journey?

Please confirm if this structure aligns with your learning goals and schedule preferences.

Contact Information:

[Ahmad Najmi bin Ariffin/MyKetapang Group]

[60134748674]

[myketapang@gmail.com]

[<https://anajmiariffin.netlify.app/>]