Brief for Level 5 Data Engineer - Portfolio Piece 3

"Mini Project"

Word count: up to 1,500 words (+/- 10%)

Objective: *Demonstrate a comprehensive understanding of data ingestion architectures and to apply this knowledge in a practical and coherent manner.*

Deliverables: You must submit the following:

- A revised architecture diagram reflecting improvements and refinements.
- A written reflection and explanation (up to 1,250 words) detailing changes made, your rationale, and lessons learned.
- At least two supporting items of evidence (e.g., revised screenshots, pseudocode, process documentation) with short captions and in-text references.

Task Breakdown:

You will begin by reviewing the feedback you received on your original mini-project submission. Then, you are to revise your architecture and write a structured reflection and explanation that includes the following sections:

1. Executive Summary

Briefly summarise the revised project, including:

- The original objective.
- Key feedback received.
- The major changes made in response to that feedback.
- Your key learning outcomes and final recommendations.

2. Introduction

Set the context for your mini-project:

- What was the use case or scenario chosen? (e.g., retail inventory analytics, financial data processing, healthcare data integration, or your own organisation).
- Why was this scenario chosen?
- What business problem or data challenge does the ingestion architecture aim to address?

3. Revised Architecture

Present your updated design:

- Include a new architecture diagram that clearly illustrates the changes made.
- Describe the components in your revised pipeline (e.g., data sources, staging, transformation, validation, ingestion, storage).
- Justify the architectural decisions made, including technology choices and data flow adjustments.

Use the following prompts to reflect more heavily on your process

Describe how your design anticipates and handles failure or downtime. What tools help you manage incidents?

Compares and contrasts different types of data store (K1, S7)   Ask: "Explain the data store(s) chosen. How do they compare in terms of performance, availability and suitability for your use case?"

Evaluates the impact of new technologies on practices (S25, S28)   Ask: "What new tools or techniques did you consider? What would be the impact of adopting them in a live environment?"

Evaluates how tech development impacts personal growth (S29)   Add a concluding reflection: "How have developments in data engineering tools shaped your professional learning journey?"

4. Reflection and Continuous Improvement

This section should offer a critical reflection on the strengths and limitations of your final architecture design, with consideration for future adaptability, clarity, and industry relevance. Your aim is not only to describe what you built but to thoughtfully evaluate its effectiveness and opportunities for further refinement.

5. Recommendations

Suggest ways you or others could improve this work further:

- Are there additional features you would consider?
- Could you implement more automation or error handling?
- Would you integrate any observability or logging tools in a future version?

Distinction Level Criteria

If you wish to be considering a distinction, consider the following reflective questions as a part of your work:

- Describe how your design anticipates and handles failure or downtime. What tools help you manage incidents?
- Explain the data store(s) chosen. How do they compare in terms of performance, availability and suitability for your use case?
- What new tools or techniques did you consider? What would be the impact of adopting them in a live environment?

Submission Requirements:

- Submit a single PDF or Word document containing all written sections and embedded visuals.

- Label and number all images (e.g., Fig. 1 – Revised Architecture Diagram).

- Reference images and evidence clearly in your writing (e.g., "As shown in Fig. 1…").

- Save and upload your document to the Hub by the module deadline.

Assessment Criteria:

You will be assessed informally against the following criteria:

- Evidence of meaningful revision and incorporation of feedback.
- Clarity and professionalism in your revised diagram.
- Accurate application of relevant learning outcomes.
- Evidence of practical skills in data collection, cleaning, and ingestion.
- Critical thinking and reflective insight into your own development.
- Adherence to submission format and word count.

KSBs

These are the KSBs you will be focusing on as part of this assessment.

Strong evidence

KSB	Descriptor
K1	Processes to monitor and optimise the performance of the availability, management and performance of data product.
K5	The inherent risks of data such as incomplete data, ethical data sources and how to ensure data quality.
K10	Concepts of data governance, including regulatory requirements, data privacy, security, and quality control. Legislation and its application to the safe use of data.
K11	Data and information security standards, ethical practices, policies and procedures relevant to data management activities such as data lineage and metadata management.
K18	How to use streaming, batching and on-demand services to move data from one location to another.
S7	Work with different types of data stores, such as SQL, NoSQL, and distributed file system.
S13	Use data systems securely to meet requirements and in line with organisational procedures and legislation.
S15	Optimise data ingestion processes by making use of appropriate data ingestion frameworks such as batch, streaming and on-demand.

Weaker evidence

KSB	Descriptor
K3	Data normalisation principles and the advantages they achieve in databases for data protection, redundancy, and inconsistent dependency.
K22	Technology and service management best practice including configuration, change and incident management.
S15	Optimise data ingestion processes by making use of appropriate data ingestion frameworks such as batch, streaming and on-demand.