Topic 4 – Introduction to reliable data architectures

This document is the handbook for Topic 4 – **Introduction to reliable data architectures**– within Module 1 – **Data Fundamentals**.

The purpose of this document is to guide your learning throughout this topic and help you to maximise the value you get from the materials provided by the BPP School of Technology

Context

This handbook is for one of 4 topics for this Module.

Every topic contributes towards the ultimate learning objectives for the Module, which you will be assessed on at the end of the term.

Module Learning Outcomes

On successfully completing this module, you will be able to:

- **Recognise** fundamental data types and data source types, key Data Engineering technological standards and best practices, and relevant regulations
- **Apply** fundamental principles for fostering a data-driven culture through collaborating with diverse stakeholders in Agile and Lean teams, avoiding waste and respecting organisational processes for data quality and data project management
- **Model** standardised Big Data ecosystems and architectures for enterprise data using visual approaches and comprehend how to design rudimentary data products, and how they add value to the organisation

Module Assessment

The Level 5 Data Engineer EPA has two assessment methods, each with its own mapping of KSBs. The Assessment plan and assessment guidance documents above list the criteria and KSBs that are assessed. The criteria group the KSBs and describe what the apprentice needs to do to achieve a pass or distinction for that assessment method.

Both assessment methods need to be passed by the candidate:

(1) Project with report

The learner will complete a project and write a report of 3500 words. Project brief submitted at gateway:

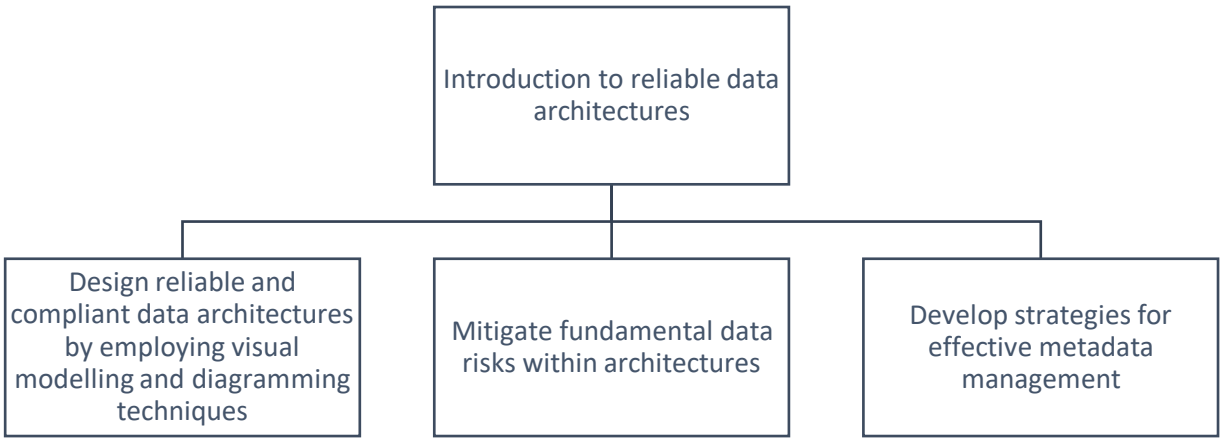
- Learners will have 10 weeks to complete the project and submit the report to the EPAO
- Learners also need to prepare and give a presentation to an independent assessor on their project
- The presentation with questions will last at least 50 minutes. The independent assessor will ask at least 6 questions about the project and presentation
- The project has to have real business application and benefit. Candidates are expected to showcase the use of appropriate standards for sustainability, privacy and security, thoroughly document their data pipeline designs, explain the choice of relevant tooling and demonstrate operational awareness of deployment, access control, risks, and how other stakeholders may be impacted positively and negatively

(2) Professional discussion underpinned by a portfolio of evidence

- Learners will have a professional discussion with an independent assessor. It will last 80 minutes.
- They will be asked at least 10 questions about Data Engineering.
- The portfolio of evidence will be used to help answer the questions.
- We expect the candidates to demonstrate examples of working with data teams on data projects and data products, showcase ideas for future-proofing data, be clear on applying problem-solving skills, show regulatory awareness, and sensitivity towards data quality, data governance and areas for continuous improvement, both personal and organisational.

Topic Learning Outcomes

As a step towards build your skills towards the final module assessment, the learning outcomes for this topic are:



Introduction

In the rapidly evolving data landscape, designing reliable and robust data architectures is a critical competency for data engineers. This topic equips you with the skills to visually model and diagram custom data architectures, ensuring that they adhere to industry standards and best practices.

Moreover, as data continues to proliferate across organisations, ensuring compliance with regulations and maintaining data integrity through rigorous validation processes becomes paramount. This topic delves into these crucial aspects, empowering you to build secure and trustworthy data architectures.

Metadata plays a pivotal role in data management, and this topic explores the risks associated with metadata shortages and provides strategies to mitigate them.

Additionally, you will gain insights into identifying and managing other fundamental data risks, such as data breaches, data quality issues, and infrastructure failures, ensuring that your data architectures are resilient and reliable.

Real-world examples illustrate the importance of these skills. A major healthcare provider implemented robust data validation and compliance measures, preventing costly data breaches and ensuring adherence to stringent regulations like HIPAA. Another company in the finance sector overhauled its data architecture through visual modelling, resulting in improved data governance, reduced technical debt, and enhanced data-driven decision-making.

By mastering the concepts covered in this topic, you will be well-equipped to design reliable data architectures that support organisational goals, mitigate risks, and foster trust in data-driven initiatives.

Structure

Topics for this programme follow a Prepare-Collaborate-Apply-Consolidate structure:

Prepare

This is the stage where you build the knowledge to underpin your learning. This might involve completing interactive e-learning packages, watching videos, or working through reading materials.

It is essential that you make the most of the learning materials provided before attending webinars, as this will allow you to test your knowledge and stretch you understanding further.

e-Learning outline

Here is an outline of what is covered:

Visual Modelling and Diagramming for Data Architectures

- Introduction to visual modelling techniques (e.g., UML, Entity Relationship Diagrams)

- Best practices for diagramming data architectures
- Tools and software for visual modelling

Data Compliance and Regulations

- Overview of data compliance regulations
- Data privacy and security considerations
- Implementing compliance frameworks

Data Validation Strategies

- Importance of data validation
- Data validation techniques (e.g., schema validation, constraint checking)
- Automating data validation processes

Metadata Management and Risks

- Role of metadata in data architectures
- Metadata standards and best practices
- Risks associated with metadata shortages and mitigation strategies

Fundamental Data Risks

- Data breaches and security risks
- Data quality risks and impact
- Infrastructure risks and disaster recovery planning

Collaborate

This is where you interact with our expert tutors and coaches to shape and refine your understanding through discussion, testing and carrying out more advanced practical and realistic tasks. This also helps to develop valuable team-working skills.

Collaborate webinar outline

Data Architecture Design Principles

- Modularisation and scalability
- Performance optimisation

- Security and access control

Data Compliance and Governance

- Interpreting and adhering to data regulations
- Establishing data governance frameworks
- Auditing and reporting compliance

Data Validation Best Practices

- Data profiling and anomaly detection
- Continuous monitoring and alerting
- Integration with data pipelines

Metadata Management Strategies

- Metadata cataloguing and discovery
- Metadata lineage and impact analysis
- Metadata governance and quality control

Risk Management in Data Architectures

- Risk identification and assessment methodologies
- Disaster recovery and business continuity planning
- Incident response and remediation

Practical Activity: Architecture Review

- Peer review of data architecture designs
- Identify potential improvements and risk mitigation strategies
- Receive feedback and recommendations from instructors

Apply

You now apply the knowledge you have developed to real-world tasks.

Off-the-job learning tasks

This stage is all about ensuring you truly grasp and retain what you've learned. Through completion of off-the-job (OTJ) revision tasks and tests, you'll get plenty of practice applying your knowledge. Plan to dedicate 6-8 hours each week to guided study and portfolio work, with sessions typically on the same day each week.

Task 1 brief: Creating Diagrams with Lucidchart

Review the following PDF guidance documents

- **Step 1:** Choose a real-world scenario or case study that involves a complex data ecosystem, such as an e-commerce platform or a healthcare analytics system, or select from any of the following:

Scenario 1: NHS Patient Data Management System

The National Health Service (NHS) requires a centralised system to manage patient data across multiple healthcare facilities, including hospitals, clinics, and general practitioners. The system should facilitate the secure storage, retrieval, and sharing of patient information, such as medical history, test results, and treatment plans. It should also enable the integration of data from various sources, including electronic health records (EHRs), medical devices, and wearables.

Scenario 2: Financial Institution's Customer Data Platform

A large financial institution aims to develop a comprehensive customer data platform to enhance its services and personalise customer experiences. The platform should integrate data from multiple sources, including core banking systems, credit card transactions, investment portfolios, and customer interactions across various channels (e.g., online banking, mobile apps, and branch visits). The system should enable the bank to gain a 360-degree view of each customer, support targeted marketing campaigns, and facilitate risk assessment and fraud detection.

Scenario 3: E-commerce Platform's Order Management System

A growing e-commerce company needs a scalable and efficient order management system to handle the increasing volume of online transactions. The system should manage the entire order lifecycle, from placement to fulfilment and delivery. It should integrate with various subsystems, such as product catalogues, inventory management, payment gateways, and shipping providers. The system should also support real-time order tracking, customer notifications, and analytics to optimise the supply chain and improve customer satisfaction.

Apply your skills!

Step-by-step guide

- **Step 1:** Identify the main entities, relationships, and data flows within the selected scenario
- **Step 2:** Sign up for a Lucidchart account (if you don't have one already) and familiarise yourself with the platform's interface and features

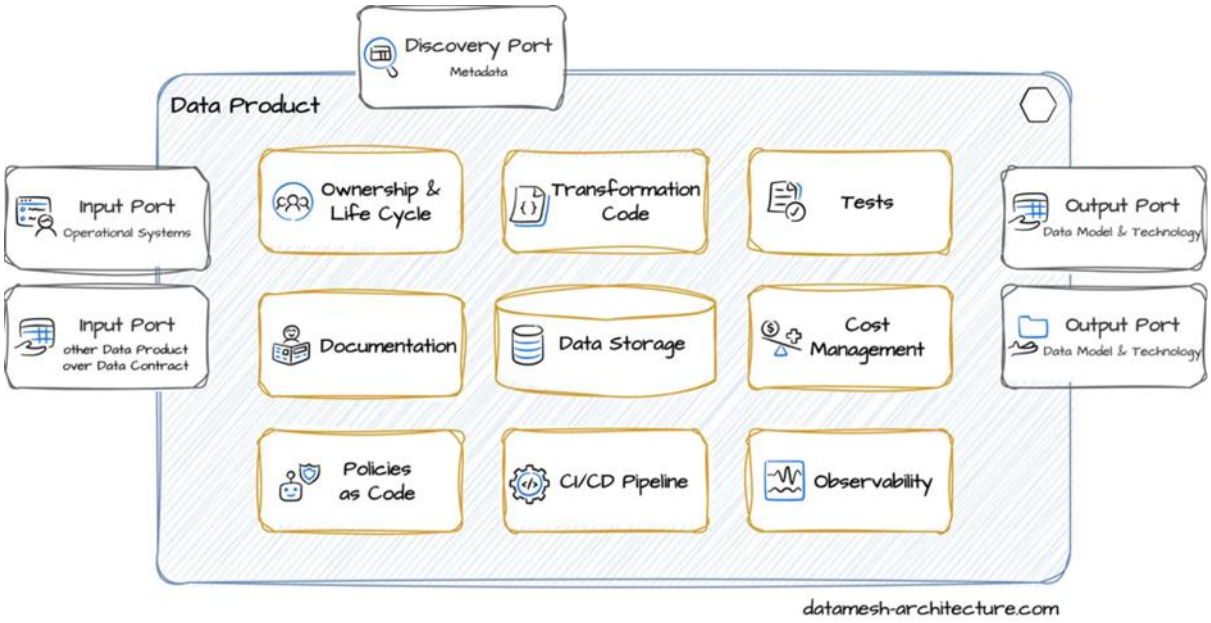
Link to LucidChart:

https://www.lucidchart.com/pages/use-cases/creating-technical-diagrams

Step 3: Create a new document in Lucidchart and select the template for Use Case Diagram as seen in the picture below:

- **Step 4:** Begin adding actors to your diagram, representing each entity as a persona and labelling them appropriately.
- **Step 5:** Define the use cases and the relationships between use cases using connector lines
- **Step 6:** Add any relevant information as needed. You can open the shape library manager by pressing the M key and then search for “UML”
- **Step 7:** Include multiple containers for complex use cases.
- **Step 8:** Annotate the diagram with comments or notes to explain key elements, data considerations, and their impact.
- **Step 9:** Share your completed diagram with peers or mentors and gather feedback on its clarity, completeness, and adherence to best practices.
- **Step 10:** Download a local version of your diagram (an image) and update your learning journal with this new deliverable.

Task 2 brief: Designing a Data Product



Step 1: Identify a specific business problem or opportunity within your organisation that could be addressed by a data product. This could be related to improving decision-making, optimising processes, or enhancing customer experiences.

Step 2: Conduct stakeholder interviews to gather requirements and understand the desired outcomes of the data product. Identify the key users, their needs, and the insights they expect to gain from the product.

Step 3: Based on the requirements, design a rudimentary data product that leverages the organization's data assets to provide actionable insights or enable data-driven decision-making. Consider the data sources, processing requirements, and visualization techniques.

Step 4: Create wireframes or mockups of the data product's user interface and key functionalities using tools like Balsamiq, Sketch, or Figma. Focus on usability, clarity, and the effective presentation of insights.

Step 5: Develop a proposal document that outlines the value proposition, target users, and expected benefits of the data product. Explain how the product aligns with organizational goals and drives business value.

Step 6: Present your data product concept to relevant stakeholders, using the wireframes and proposal document to communicate your vision. Be prepared to answer questions and gather feedback for further refinement.

Task 3 brief: Applying TOGAF Principles to Data Governance

Step 1: Research the key principles and components of the TOGAF framework, focusing on its relevance to data governance. Understand the Architecture Development Method (ADM) and its phases.

Link to the OpenGroup website: TOGAF | www.opengroup.org

Step 2: Select a data domain or system within your organisation that could benefit from improved governance practices. This could be related to master data management, data quality, or data security.

Step 3: Review the existing data governance practices and identify areas for improvement based on TOGAF principles. Consider aspects such as data ownership, data quality standards, and compliance requirements.

Step 4: Develop a high-level data governance plan based on TOGAF principles. Define the governance structure, roles and responsibilities, decision-making processes, and policies for managing data assets effectively.

Step 5: Create a visual representation of the proposed governance framework using tools like Microsoft Visio or Lucidchart. Use flowcharts, mind maps, or other diagrams to illustrate the key components and their relationships.

Step 6: Be prepared to defend your visual representation design choices as required.

Consolidate

This stage ensures learners embed their knowledge so that it can be stored in their long-term memory. It also provides plenty of opportunity to practise retrieving and applying that knowledge by completing revision tasks and tests.

Consolidation Activities

Consolidation Activities are recommended to reinforce concepts.

- **Reading:** Explore industry publications and case studies on data architecture design, compliance, and risk management
- **Video:** Watch recorded webinars or tutorials on visual modelling techniques, data validation, and metadata management

Quiz: Complete a self-assessment quiz to reinforce key concepts and identify areas for improvement.

Extension Activities

Extension Activities generally contain ***optional resources*** for students that seek to explore concepts in more depth.

- **Research Paper:** Investigate emerging trends and best practices in data architecture design, and document your findings in a research paper
- **Presentation:** Prepare a presentation on the importance of data compliance and risk management, and share it with your peers or a local industry group
- **Architecture Redesign:** Identify an existing data architecture (e.g., from a previous project or open-source) and propose enhancements to improve reliability, compliance, and risk mitigation

Link

This handbook is for one of 4 topics for this Module.

The sequence of topics in this module is carefully designed so that your knowledge and skills will develop as you progress.

This is the last topic in the module.