BPP UNIVERSITY SCHOOL OF TECHNOLOGY

Level 5

Data Engineer

Higher Apprenticeship

Data Fundamentals Module Handbook
2024

MODULE OVERVIEW

Data Fundamentals

Module	Phase	Weekly Topics	Description
Data Fundamentals	Month 1	• **Fundamentals of the data-driven enterprise**: estimating and monitoring the business value of Big Data and small data; data ecosystems in modern organisations: data fabric and data mesh. • **Introduction to data quality:** data standards, data policies, data principles, structured and unstructured data types and formats, data lineage, identifying relevant metrics. • **Introduction to managing data projects and products**: collaborating with stakeholders in Agile and Lean data teams; optimising data discovery through data catalogues, data dictionaries and data stewardship principles; overview of data product types and the data lifecycle. Sustainability and net-zero considerations; • **Introduction to reliable data architectures:** Demonstrate familiarity with UML, ERD, and custom architecture diagrams, including layered and microservices designs. Explain why architecture modelling requires various lenses to minimise the risk of miscommunication and help governance by incorporating data catalogues and	This exciting one-month module exposes our Apprentices to the rich tapestry of key Data concepts. As the first stage of the 'spiral curriculum', we provide the learners with an inspiring panorama of what Data Engineering is really about. Deep-dives are provided in later modules, as we first model a real-world use case scenario that will be revisited throughout the course by gradually enriching it with detail. You are taught how the business generates value from data and the role played by Data Engineering in driving that function. You will learn fundamental Data Engineering best practices and standards, mastering essential skills for managing data projects and products, and understanding how to work with diverse stakeholders who use various types of data in a controlled way. Learners will work practically with modern data ecosystems and begin building their own toolbox of Big Data skills. Gaining insight into cutting-edge data architectures, from the very beginning of your programme, you will be introduced to governance and regulatory issues, data risks and key environmental considerations.

		dictionaries, ensuring robust metadata management	

Outcomes

After finishing this module, you will be able to meet the following outcomes and KSBs:

- **Recognise** fundamental data types and data source types, key Data Engineering technological standards and best practices, and relevant regulations.
  - K9, K10, K11, K16, K19, K21, K28
  - S13, S29,
  - B6

- **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.
  - K4, K5, K7, K29, K30,
  - S12, S22, S26, S27,
  - B1, B2, B4, B6

- **Model** standardised Big Data ecosystems and architectures for enterprise data using visual approaches, **demonstrate the use of an approach** to designing a rudimentary data product, and **explain** how data products add value to the organisation.
  - K20, K24, K25, K26, K27,
  - S1, S2, S24,
  - B6

- **Explain** the importance of assessing, improving, and maintaining data quality using frameworks and methodologies to ensure data accuracy, completeness, and consistency.
  - K3, K4, K5, K26
  - S6, S26

Introduction

This module serves as an essential introduction to the field of Data Engineering, providing you with a thorough understanding of the key concepts, practices, and tools that form the basis of your future success. The one-month module exposes our Apprentices to the rich tapestry of key Data concepts. Deep-dives are provided in later modules, as we first model a real-world use case scenario that will be revisited throughout the course by gradually enriching it with detail.

In this module, you are taught how the business generates value from data and the role played by Data Engineering in driving that function. You will start mastering essential skills for managing data projects and products, and understanding how to work with diverse stakeholders who use various types of data in a controlled way. Gaining insight into cutting-edge data architectures, from the very beginning of your programme, you will be introduced to governance and regulatory issues, data risks and key environmental considerations.

Throughout this module, you'll explore crucial topics such as data-driven enterprises, data quality, data project and product management, and reliable data architecture design. You'll gain an understanding of how businesses use data to drive value and your role as a Data Engineer in making that happen. You'll learn about key concepts of data ecosystems and meshes, as well as the importance of data quality related policies and metrics. The module also focuses on the soft skills you'll need for effective collaboration and communication within data teams, introducing you to Agile and Lean methodologies for managing projects and working with diverse stakeholders.

On the technical side, you'll dive into the fundamentals of designing reliable data architectures using industry-standard modelling techniques like UML. You'll learn to create architecture diagrams that effectively communicate the structure and relationships of data systems and explore key architectural patterns such as layered and microservices designs.By the end of this module, you'll have a solid foundation in the core concepts and practices of Data Engineering, ready to tackle the data landscape, assess data quality, manage data projects, and design reliable data architectures as you progress through the program.

Mode of Assessment

END POINT ASSESSMENT (EPA)

It is **important** to read all the assessment guide documents contained in the Programme Handbook, as they contain important details.

Reminder: Refer to the programme handbook for further guidance.

Core Reading:

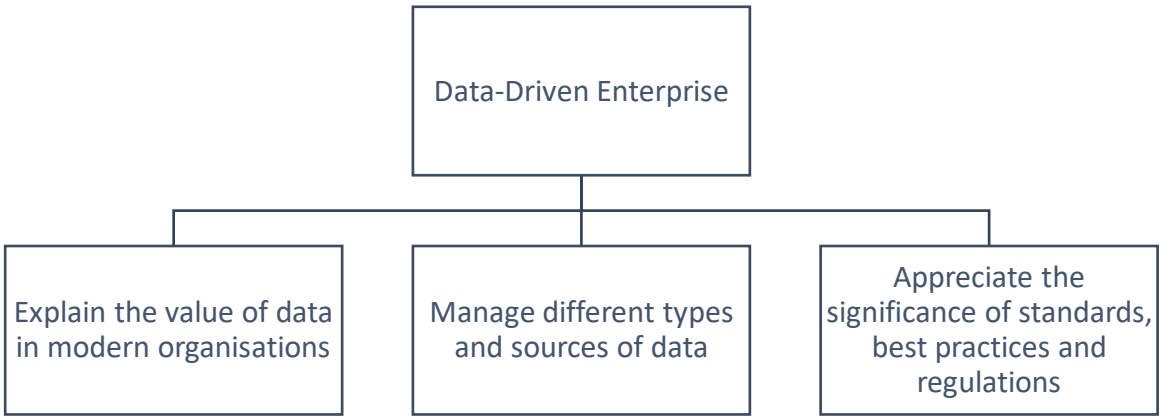
- Programme Handbook: Students will be provided with a single handbook by BPP which will summarise the key expectations for each topic in the course. The handbook is reviewed annually to include the latest and most appropriate academic resources.
- Haskell, C. (2024). Driving Data Projects: A comprehensive guide. British Computer Society [20 hours]
- King, T., Schwarzenbach, J., (2020). Managing Data Quality: A practical guide. British Computer Society [15 hours]
- Reis, J., Housley, M. (2022), Fundamentals of Data Engineering: Plan and Build Robust Data Systems, O'Reilly [25 hours]

TOPIC BREAKDOWN

Topic 1 – Data-Driven Enterprise

Topic Learning Outcomes

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



Introduction

In the modern, data-driven business landscape, mastering data fundamentals is pivotal for organisations to harness the true power of their data assets. This is especially crucial for data engineers, whose roles are centred around designing, building, and maintaining the robust data infrastructures that form the bedrock of data-driven decision-making.

By gaining a comprehensive grasp of key data engineering concepts covered in this topic - from data types, formats and measurement units to data collection, cleansing, transformation and integration techniques - data engineers are empowered to navigate the ever-increasing complexities of managing data at scale.

Moreover, adhering to relevant industry standards, regulations and governance frameworks is paramount to ensuring compliance and fostering trust among stakeholders. An in-depth understanding of data engineering principles like scalability, reliability and security enables practitioners to architect systems capable of handling immense data volumes while maintaining performance and resilience. Throughout this topic, case studies illuminate how organisations are leveraging these fundamentals to drive tangible business impact.

A prime example is the media streaming giant Netflix, which has established itself as a pioneer in harnessing data to optimise decision-making across its operations. By meticulously collecting, analysing and drawing insights from vast troves of subscriber data – spanning viewing habits, preferences and feedback – Netflix's data engineers have equipped the company to make informed decisions on content licensing, recommendation engines and product enhancements, thereby delivering unparalleled customer experiences.

Upon completing this topic, you will possess a comprehensive skillset encompassing data visualisation tools, techniques for extracting business value from data, and strategies for fostering a data-driven organisational culture. Armed with this robust foundation, you will be poised to contribute effectively to transformative, data-centric projects, collaborate seamlessly with cross-functional teams, and spearhead groundbreaking data-driven initiatives that yield sustainable competitive advantages for organisations.

Structure

Topics for this programme follow a Prepare-Collaborate-Apply 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.

The e-learning for this topic covers building a data-driven culture, fundamentals of data types and sizes, standards and best practices in data engineering, working with different data sources and types through the data management lifecycle, and utilisation of visualisation tools.

Collaborate

This is where you will receive guidance from our expert tutors and coaches to shape and refine your understanding through in-depth explanation, discussion, testing and carrying out more advanced practical and realistic tasks. This also helps to develop valuable team-working skills.

The webinar for this topic will focus on building a data-driven culture, data fundamentals like types/sizes, key standards and best practices in data engineering, a practical lab, and summarised the core learning concepts.

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 your learning journal in GitHub

This task is about creating a learning journal on GitHub to document your learning journey throughout the course. It is divided into the following stages:

- **Stage A:** Creating a GitHub account and explore the platform
- **Stage B:** Setting up a directory structure for your repository by creating folders for each module
- **Stage C:** Creating a markdown file within the "Data Fundamentals" folder to compile your notes from the first webinar. Use proper formatting, headings, lists, and visual aids

- **Stage D:** Structure your learning from the webinar by listing key objectives, capturing main concepts, noting questions, self-researching answers, and reflecting on real-world applications

The task focuses on the benefits of keeping a learning journal, such as deepening understanding, boosting memory, tracking progress, and identifying knowledge gaps. It also provides tips for success, like committing changes regularly, seeking feedback, and keeping notes organised.

Further guidance can be found here: L5DE 1.1 Apply (OTJ)

Task 2 brief: Data types and data sources used within your organisation

For this task you should have at least one meeting with an experienced Data Engineer (or Software Engineer) to discuss and identify the key data types and data sources used within your organisation's data systems.

Following this meeting, you should:

- Outline (e.g. in an email) an individual approach for learning the standards, best practices and regulations which inform the use and management of data systems within the organisation
- Cross-validate your findings with other Data Engineers (e.g., in a meeting, or email). Document your learning in your learning journal

Link

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

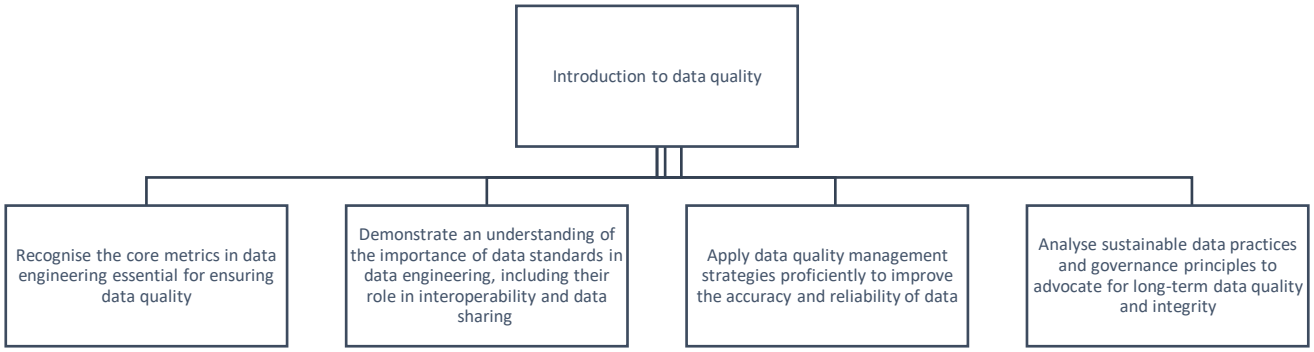
The sequence of topics in this module is carefully designed so that your knowledge and skills will develop as you progress.

The next topic is **Introduction to data quality**.

Topic 2 – Introduction to data quality

Topic Learning Outcomes

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



Introduction

For data engineers, mastering data quality concepts and techniques is an indispensable skill. In today's data-rich world, organisations rely heavily on the accuracy, integrity, and reliability of their data assets to drive informed decision-making, optimise processes, and maintain a competitive edge. This topic is tailored to equip you with a comprehensive understanding of data quality principles and best practices, preparing them to tackle real-world challenges effectively.

Real-world examples of the value delivered by strong data quality practices abound. Consider the case of a leading retail company that faced significant challenges with inaccurate product data, leading to customer dissatisfaction and operational inefficiencies. By implementing a robust data quality testing framework, including data validation, continuous monitoring, and data profiling, the company successfully improved the accuracy and reliability of its product data, resulting in enhanced customer satisfaction and streamlined operations.

Another notable example is from the healthcare industry, where data quality is paramount for ensuring patient safety and coordinated care. By adhering to data quality principles, such as maintaining data lineage and implementing unique identifiers, healthcare organisations can trace the origin of patient records, ensuring compliance with privacy regulations and preventing errors like duplicate or misidentified records.

Structure

Topics for this programme follow a Prepare-Collaborate-Apply 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.

This e-learning for this topic covers a comprehensive range of topics related to data quality, equipping learners with the knowledge and skills essential for assessing and enhancing the accuracy, completeness, consistency, and timeliness of data.

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.

The webinar for this topic will focus on data quality concepts, standards, handling issues, strategies, automation, sustainability considerations, and practical exercises related to data quality management for data engineers.

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: Exploring real-world datasets

The task aims to gain hands-on experience in exploring real-world datasets. For this task you should complete the following steps:

1. Find the "IBM HR Analytics Employee Attrition & Performance" dataset on Kaggle.
2. Explore the dataset's metadata, including assessing its usability score, credibility issues, number of columns, license, and usage rights.
3. Examine the dataset's description, data types (e.g., sc:Integer, sc:Boolean), and compression implications.
4. Download the dataset and identify potential data quality issues, such as column interpretation challenges, inconsistencies, and missing/invalid values.
5. Practice basic data cleaning and documentation tasks, including:

- Retaining specific columns
- Creating a "Metadata" sheet with document details
- Adding a "Schema" sheet to define column properties
- Adding a "Provenance" sheet to document the dataset's origin

6. Upload the cleaned and documented dataset to your GitHub repository.
7. Review the Knowledge, Skills, and Behaviours (KSBs) covered in this exercise and document them in your learning journal.

The goal is to develop skills in navigating data repositories, analysing metadata, identifying data quality issues, and performing basic cleaning and documentation tasks essential for data engineers.

Further guidance can be found here: L5DE 1.2 Apply

Task 2 brief: Exploring automated data quality management tools

The task aims to gain hands-on experience with automated data quality management tools by researching, exploring, and documenting findings in your learning journal.

For this task you should:

- Identify and research at least two automated data quality management tools or libraries used in your organization or relevant to your field of interest
- Arrange a shadowing session with a data engineer or subject matter expert to observe how they leverage these tools in their workflow. Document insights on use cases, benefits, challenges, integration, and best practices
- In your learning journal, provide an overview of the tools researched, key features and capabilities, examples of usage in your organisation/industry, potential limitations, and best practice recommendations
- Share your learning journal entry with your coach or mentor for feedback and guidance
- Log your on-the-job training hours using the provided file

The task emphasises practical exposure, documentation, and knowledge-sharing to enhance your understanding and skills in ensuring data quality and reliability using automated tools and techniques.

Further guidance can be found here: L5DE 1.2 Apply

Link

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

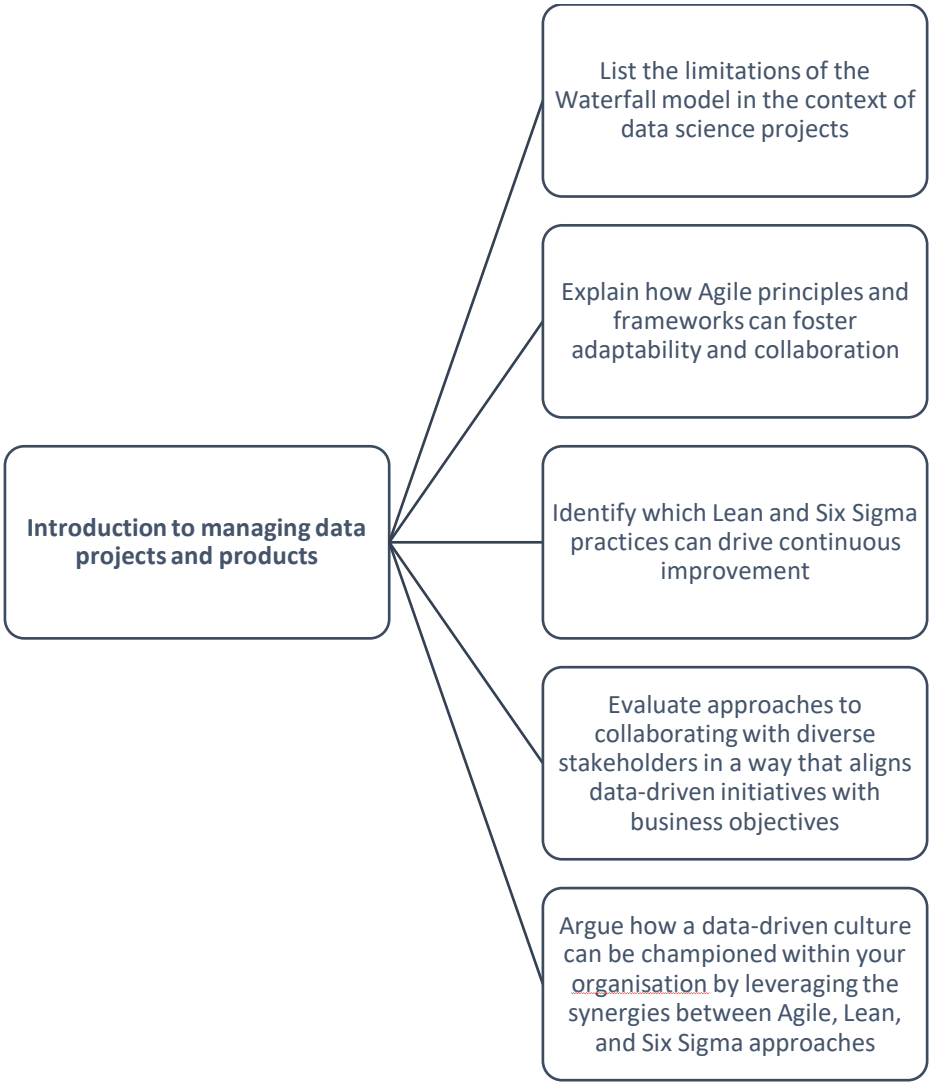
The sequence of topics in this module is carefully designed so that your knowledge and skills will develop as you progress.

The next topic is **Introduction to managing data projects and products**.

Topic 3 – Introduction to managing data projects and products

Topic Learning Outcomes

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



Introduction

In today's fast-paced, data-driven business landscape, effective management of data projects and products has become a critical skill for data engineers. As organisations increasingly rely on data to inform decision-making and drive innovation, the ability to collaborate with diverse stakeholders, leverage agile methodologies, and champion a data-driven culture is paramount.

This topic is designed to equip you with the essential skills and knowledge needed to successfully navigate the complexities of managing data projects and products. You will explore the limitations of traditional Waterfall approaches and discover how Agile principles and frameworks, such as Scrum and Kanban, can foster adaptability and collaboration within data engineering teams.

Furthermore, you will delve into the world of Lean and Six Sigma practices, learning how these methodologies can drive continuous improvement and enhance data quality. By understanding the

synergies between Agile, Lean, and Six Sigma, you will be well-positioned to champion a data-driven culture within your organisation.

Throughout this topic, you will also develop critical skills in stakeholder engagement, learning how to effectively collaborate with diverse stakeholders to align data-driven initiatives with overarching business objectives.

By mastering these skills, you will become a valuable asset to any data engineering team, capable of delivering high-quality data products that drive real business value.

Structure

Topics for this programme follow a Prepare-Collaborate-Apply 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.

This e-learning for this topic covers a comprehensive range of topics including Agile methodologies and their benefits over the traditional Waterfall approach, especially in the context of data engineering projects. It also covers how integrating Lean and Six Sigma principles with Agile can further enhance data quality, streamline processes, and drive continuous improvement in data engineering initiatives.

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.

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: Leveraging the power of user stories in Agile

As a Level 5 Data Engineering apprentice, you're on a journey to master the intricacies of managing data projects and products. Today's activity is a pivotal step in this journey, focusing on harnessing the agility of Agile methodologies through the art of crafting user stories.

In this exercise, you'll step into the shoes of a Product Owner, tasked with translating customer needs into actionable user stories.

Preparation

One of the most important aspects of moving to Agile is understanding “stories”. It takes practice to write good stories, and this exercise allows you this practice. As the Product Owner, you must deliver your customer’s or stakeholder’s perspective and share with the project team what is needed and why.

A user story must provide value to some user. An Agile process is driven by the completion of stories, each of which provides tangible, demonstrable value to the user/customer/stakeholder. A sprint consists of a set of conscientiously prioritised stories. Experience will show that it’s best to use a format for each story that identifies who the user is, what they need, and for what purpose (the why). Such stories are written in this format:

“As a ____, I need a ____ in order to ____”.

The **who** in a user story could be someone with a particular functional role, who holds a certain title, comes from the perspective of a persona, or embodies the needs and behaviours of a hypothetical user.

The **what** in a user story details in specific terms the need, feature, or functionality desired by the who. This is what your project team will build into the product or service.

The **why** in a user story states the value. It presents the needs of your users and customers up front and centre.

Here’s an example of a user story that clearly defines the who, what, and why: “As a jazz fan, I need a tuning knob in order to find a jazz station on the radio that I will enjoy listening to.”

Please note: Examples of a User Story and a template is provided at the end of this document

The keys to developing a valuable user story

- Product Owners must have courage to ask for what they believe their users/customers/stakeholders really want
- A story must have value to someone. It must make the product better in some way
- The story when complete will make a real-world task faster, better, easier to understand, have fewer steps, or collect better info
- The high priority stories affect the most users or procure the highest value data
- Avoid exotic/one-off stories (i.e. edge cases)
- “Clean up the bugs we introduced in the last sprint” is NOT a user story because it does not add anything to the product

Remember, according to the INVEST model! Good user stories are:

- Independent
- Negotiable
- Valuable
- Estimable
- Small and
- Testable

Your instructions for completing this activity

Grab a fellow student (or several) and use one of your current projects (or choose one from the list below). As Product Owner, you want to be able to communicate to the development team what users need. You don’t need to be a technical person to do this, you just need to know what a user wants and why.

These user stories will provide the team with starting points to discuss how they might accomplish something. Instead of saying “I need an event calendar” you might say “A user needs to be notified of upcoming events that are related to topics of interest in her user profile so that she engages with the

community about things that matter to her.” This story is more descriptive and gives the team a better understanding of the goal so they can base their solution on intended outcomes.

Believe it or not, one of the critical technologies for this exercise is either index cards (3×5 or 4×6) or Post-It notes. Part of the reason you use paper technology is so you can easily move stories around, reorganise and reprioritise them, and throw them away when done. The small size of the cards and notes ensures that you will not write too much into each story.

A story is a promise to have a conversation later between the end-user and developers. Your goal in writing stories is not to work out details, but to discover the most important goals for your project and to organise a project into discrete, testable chunks.

Potential project goals / directives that will help you practice story-writing include:

1. Re-arranging office space

2. Make a website that informs people about dietary impact on breast cancer

3. Developing a mobile app that periodically checks your calendar on your phone and informs you if the weather is likely to impact any of your plans

4. Develop a dashboard that shows patient health and insurance information

5. Publishing a how-to guide for planting a garden

Choose a project goal for the workshop from the list above. If you have an existing project or goal already in mind, feel free to use that instead.

Activities

Everyone writes stories for 15 minutes that will advance the project goal.

As new stories are written, you may discover ways to improve previously written stories. You may realise that many small, specific stories can be rolled up into a bigger story; or a big story might be split into two or three pieces.

During this time, stories can be rewritten or reworded to make each story as self-contained as possible. This means that a developer should be able to read the story, understand what a user is hoping to do, and create a feature that enables the user to do it. Additionally, a self-contained story should be understandable when read back to the user, without a lot of explanation needed.

The stories are prioritised into an absolute order.

There can only be one #1, with the highest priority at the top. There are no ties; a specific order must be chosen. Open discussion is allowed, but in the end, you as the Product Owner have authority to set the actual priorities. Things might change as you discover new information, and the backlog can be re-ordered — but at any one time it is an absolute ranking.

Check the format of the user stories.

Remember to keep them in the correct format: “As a ____, I need a ____ in order to ____”.

Think about grouping the stories into sprints.

Now that your stories are prioritised, do you see how you might group them into sprints? The grouping is usually done by the project team, but it’s a good exercise for you to review now to get an idea of how complete your stories are and to recognise if there are gaps._____

Outcomes

- It may feel overwhelming to create user stories to describe everything you are imagining for this project. Start from where you are and remember your product backlog will continue to grow throughout the life of the project

You will get better at creating user stories over time, but this exercise should produce enough user stories that a sprint could be planned.

Task 2 brief: Agile GDPR backlog refinement

You are a Data Protection Officer (DPO) of Cosmetics International Limited based in London. Your organisation introduced a new make-up line *Shine-N-B-Fine* that can be sourced based on a variable colour palette. The initial sales of the product are lower than prior forecasts.

A meeting commissioned by your Marketing department has just concluded, where the Sales team produced a product features' analysis and a recommendation that the product be broken down into geographical variations that should be marketed differently in different regions, also suggesting that this will maximise sales revenue to better hit the forecasts.

Based on this information, vice president VP Marketing added a user story to your Jira Backlog that says the following:

Epic – Shine-N-B-Fine soft launch (100 points)

User Story (8 points)

As: VP Marketing, I want: an analytics dashboard to classify Shine-N-B-Fine user numbers geographically, so that: I can optimise marketing plan for the product

Acceptance criteria:

The solution should be self-service

- This should be accessible by Marketing and Sales teams
- Should take into account all the potential buyers that we have data about in the Enterprise Data Warehouse (EDW) + partner data through ingestion Application Programming Interface (API)
- The data points required are age range, gender, region, postcode, disposable income, is this a previous buyer of Cosmetics International ranges, skin tone (or estimate if absent)
- Should feed into Automated Sales and Distribution Support System (ASDSS)

A rumour ensued that this user story may potentially cause additional overhead. You are called in to a backlog refinement meeting to answer the following questions:

1. Will any GDPR restrictions apply?
2. What rights can the data subjects exercise? Enumerate and explain in detail.
3. Is a DPIA required? Justify why or why not.
4. What other recommendations do you have?
5. Do you include the story in this sprint?
6. Do you suggest any additional acceptance criteria?
7. Review the User Story points.

Task 3 brief: Championing a data-driven culture through Agile-Lean-Six Sigma approaches

Drawing insights from the "Introduction to managing data projects and products" topic, you will undertake the following research exercise to develop your understanding and skills in this area.

Outcome 1: Evaluate approaches to collaborating with diverse stakeholders in a way that aligns data-driven initiatives with business objectives.

You can meet this outcome by completing the following:

- Identify the key stakeholders (e.g., business leaders, IT, data analysts, end-users) who are typically involved in data-driven initiatives in your organisation
- Analyse effective strategies and best practices for collaborating with this diverse set of stakeholders to ensure alignment between data-driven projects and your organisation's overarching goals
- Document your findings in a stakeholder engagement plan, highlighting key considerations for facilitating communication, managing expectations, and maintaining alignment throughout projects

Outcome 2: Argue how a data-driven culture can be championed within an organisation by leveraging the synergies between Agile, Lean, and Six Sigma approaches in your organisation.

You can meet this outcome by completing the following:

- Examine how the principles and methodologies of Agile, Lean, and Six Sigma can be integrated to drive cultural change and continuously improve data-driven decision-making
- Develop a proposal outlining how you would introduce and implement these integrated methodologies within your organisation to champion a data-driven culture

Address the following in your proposal:

- How can you secure buy-in from leadership and cross-functional teams?
- What specific Agile, Lean, and Six Sigma practices would you implement, and how would they complement each other?
- How can you measure the success of a data-driven culture initiative, and what key performance indicators (KPIs) would you track?

Why complete this apply activity?

This research exercise will allow you to develop your understanding of the concepts covered in the "Introduction to managing data projects and products" topic and apply them to real-world challenges faced by data professionals.

By exploring stakeholder collaboration strategies and proposing a data-driven culture initiative, you will enhance your skills in leveraging the synergies between Agile, Lean, and Six Sigma methodologies to drive organisational transformation.

Link

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

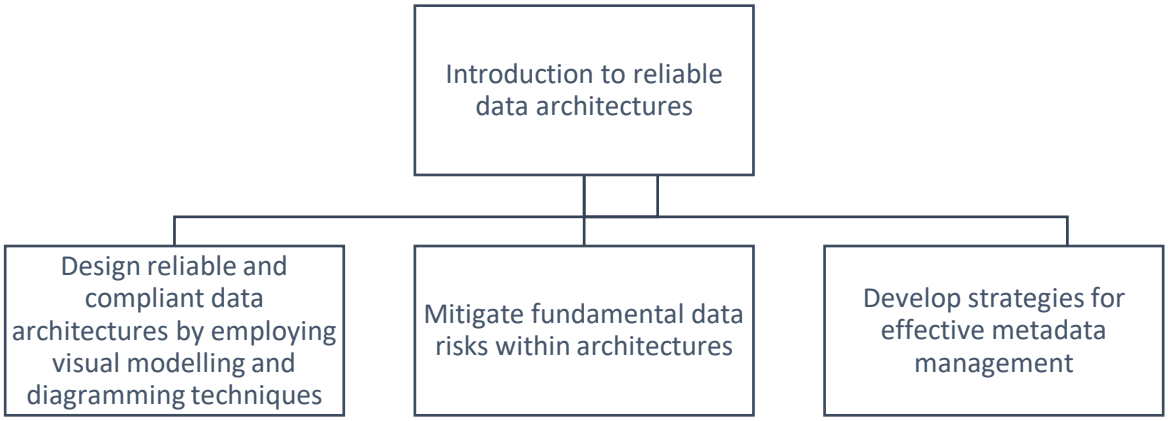
The sequence of topics in this module is carefully designed so that your knowledge and skills will develop as you progress.

The next topic is **Introduction to reliable data structures**.

Topic 4 – Introduction to reliable data architectures

Topic Learning Outcomes

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



Introduction

In today's data-driven world, designing robust and efficient data architectures is crucial for organisations to make informed decisions, drive innovation, and maintain a competitive edge. This topic, "Introduction to Reliable Data Architectures," equips you with the fundamental skills and knowledge needed to navigate the complex landscape of data architecture design.

Throughout this journey, you will explore visual modelling and diagramming techniques, such as Unified Modelling Language (UML) diagrams, Entity-Relationship Diagrams (ERDs), and custom diagrams, to effectively communicate complex data architectures to both technical and non-technical stakeholders. You will also learn about key architectural patterns and principles, like layered architecture and microservices architecture, which help organisations build flexible, scalable, and resilient data systems.

Furthermore, this topic emphasises the importance of designing data products that align with business objectives and drive tangible value. You will learn how to visually prototype data products, ensuring they meet end-users' needs and contribute to the organisation's bottom line. Finally, you will explore the critical role of data governance and architectural frameworks like TOGAF in ensuring the reliability and compliance of data architectures.

Throughout this topic, you will have the opportunity to apply your learning through practical exercises, case studies, and collaborative activities. By mastering the concepts and techniques covered, you will be well-equipped to lead the charge in building robust, efficient, and future-proof data architectures that drive innovation and success.

Structure

Topics for this programme follow a Prepare-Collaborate-Apply 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.

This e-learning for this topic covers a comprehensive range of topics including visual modelling and diagramming techniques (UML, ERD, custom diagrams), key architectural patterns and principles (layered architecture, microservices), designing data products aligned with business objectives, data governance, and architectural frameworks like TOGAF.

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.

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 Database 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

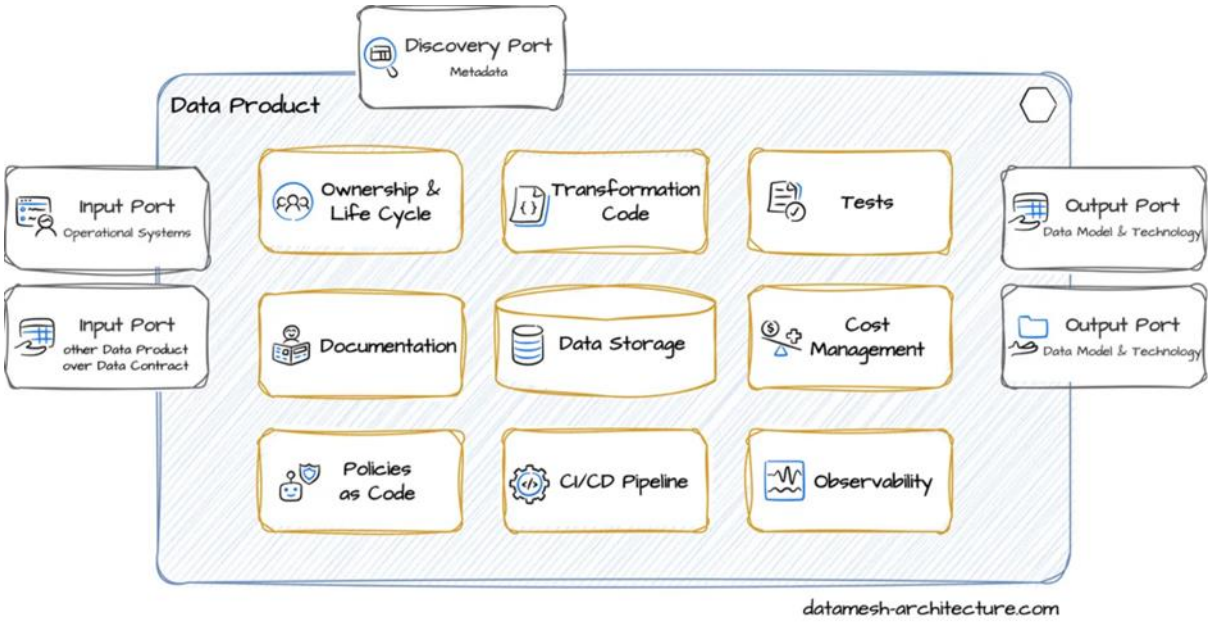
Link to LucidChart:

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 4 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.

Link

This topick 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.