

**Data Technician**

|  |
| --- |
|  |

|  |
| --- |
| Name: |
| Course Date: |
|  |

**Table of contents**

[Day 1: Task 1 3](#_Toc1514505634)

[Day 1: Task 2 3](#_Toc2125927931)

[Day 1: Task 3 4](#_Toc1965065538)

[Day 2: Task 1 5](#_Toc2141651249)

[Day 3: Task 1 9](#_Toc1229522580)

[Day 3: Task 2 10](#_Toc508692763)

[Day 3: Task 3 11](#_Toc1233463339)

[Day 4: Task 1 12](#_Toc1556426903)

[Day 4: Task 2 13](#_Toc214156810)

[1. Scenario Background 13](#_Toc248266112)

[2. Data Laws and Regulations 14](#_Toc1387014804)

[3. Azure Service Recommendations 14](#_Toc1952198484)

[4. Data Types and Data Modelling 14](#_Toc1792190821)

[5. Data Storage Formats and Structures in Azure 14](#_Toc385598743)

[6. Additional Considerations 15](#_Toc561077662)

[Submission Guidelines: 15](#_Toc1001523541)

[Course Notes 17](#_Toc977988415)

[Additional Information 17](#_Toc1081373283)

# Day 1: Task 1

Please research and complete the below questions relating to key concepts of cloud.

Be prepared to discuss the below in the group following this task.

|  |  |
| --- | --- |
| What can cloud computing do for us in the real-world? | Cloud computing allows us to run applications, store data, and perform complex tasks using the internet. We do not need powerful local computers because everything is processed on remote servers.  Real-world examples:   * Storing photos in Google Drive or iCloud. * Playing cloud games like Google Stadia or Xbox Cloud. * Editing documents online (Google Docs). * Processing big data in science or healthcare. |
| How can it benefit a business? | Businesses receive several key benefits:   * Cost saving — no need to buy expensive hardware. * Scalability — resources can be increased or reduced when needed. * Remote access — employees can work from anywhere. * Reliability — automatic backups protect data. * Security — cloud providers often offer strong data protection. |
| What’s the alternative to cloud computing? | The alternative is using local infrastructure (on-premises), when a company keeps all data and runs applications on its own physical servers, which are located in its office or in a private data centre. The company must buy the equipment, maintain it, and take care of security, updates, and backups.  Advantages: full control over the system.  Disadvantages: expensive, requires IT staff, difficult to scale. |
| What cloud providers can we use, what are their features and functions? | * Amazon Web Services (AWS) - Very wide range of services: databases, computing, storage. * Microsoft Azure - Good for companies using Windows/Office * Google Cloud (GCP) - Strong in data analytics and machine learning * IBM Cloud - Suitable for enterprise-level solutions * Dropbox / Google Drive / OneDrive - Simple file storage and collaboration |

# Day 1: Task 2

Please research the below cloud offerings, explain what they are and examples of use cases.

|  |  |  |
| --- | --- | --- |
| Cloud Offerings | Explain what it is | When / how might you use this service in the real-world? |
| IaaS (Infrastructure as a service) | IaaS is like renting a powerful computer or server over the internet.  You don't need to buy real hardware (a physical server), install it on-premises, configure cooling, power, etc.  Instead, you connect to a ready-made "virtual computer" located in a provider's data center (e.g., Microsoft Azure, Amazon AWS, Google Cloud). | A company may use IaaS to quickly scale their server capacity during high-demand periods, such as during a big sale for an online shop. Developers can also use IaaS to test applications without buying physical servers.  What's included in IaaS:   * Physical servers * Network infrastructure and security * Data storage (virtual disks) * Hardware maintenance   What the user does:   * Installs the desired operating system (e.g., Windows or Linux) * Installs programs and applications * Stores their files and databases * Manages all the contents of their virtual server. |
| PaaS (Platform as a service) | PaaS offers a complete environment for developing, testing and deploying applications. It includes infrastructure, operating systems, development tools and databases, all managed by the provider. Users only manage their applications and data. | Useful for developers who want to build and deploy applications quickly without worrying about setting up servers or installing databases. For example, creating a web app using Azure App Service or Google App Engine.  With PaaS (such as Azure App Service or Google App Engine):   * you simply log into the platform, * upload your code, * click "launch," * and the site starts working.   Without PaaS, you would have to:   * rent or buy a server, * install Windows/Linux on it, * install a database (e.g., MySQL), * configure everything manually, * monitor updates and security. |
| SaaS (Software as a service) | SaaS provides ready-to-use software applications over the internet. Users access the software through a web browser or app, without installing or managing it. The provider handles everything, including updates and security. | Commonly used for email (e.g., Gmail), office tools (Microsoft 365), customer management (Salesforce) or video conferencing (Zoom). Users simply log in and use the service from any device. |

# Day 1: Task 3

Please research the below terms and explain what they are, when they would be appropriate and a real-world example of where it could be implemented (i.e. what type of organisation).

|  |  |
| --- | --- |
| Public Cloud | A public cloud is managed by companies like Amazon Web Services (AWS), Google Cloud Platform (GCP), or Microsoft Azure. It is used by many organizations that want low cost and flexibility.  Example: A startup that builds a mobile app and stores data on AWS.  A public cloud is like a public bus. Anyone can use it — you just buy a ticket (pay for what you use). You don’t own the bus, but it’s cheap and available for everyone. |
| Private Cloud | A private cloud is built and used by one company only.  It gives more control, privacy, and security.  Example: A bank (like Barclays or Sberbank) uses a private cloud to store customer financial data.  A private cloud is like a company’s own bus. Only the company’s employees can use it. It’s more secure and controlled, but also more expensive because the company owns and maintains it. |
| Hybrid Cloud | A hybrid cloud mixes public and private clouds.  Some apps or data stay private, and others run on a public cloud.  Example: A hospital stores patient records in a private cloud but uses Microsoft Azure for appointment booking.  A hybrid cloud is like a mix of public and private buses.  The company uses its own buses for important routes (private) but also rents public buses during busy hours (public). |
| Community Cloud | A community cloud is shared by several organizations with common goals or rules. It helps reduce costs and improve collaboration.  Example: Several universities share a community cloud for research and data storage.  A community cloud is like several transport companies sharing buses together. They have similar needs (routes, passengers, rules) and share the costs and resources. |

# Day 2: Task 1

Describe, with examples, the **three** major areas that the Computer Misuse Act deals with.

|  |  |  |
| --- | --- | --- |
| Area | Description | Example |
| Unauthorised access to computer systems | This means logging into someone’s computer or account without permission. | A person hacks into another person’s email. |
| Unauthorised access with intent to commit further crimes | This is when someone gets access to a computer to do something illegal, like fraud or stealing data. | A hacker enters a bank system to steal money or customer information. |
| Unauthorised modification of data | This includes changing, deleting, or damaging data without permission. It also covers spreading viruses or malware. | Someone installs a virus that deletes files on a company’s computers. |

The computer misuse act 1990 is an act where an individual can be criminalised because of computer related offense. Describe three extra powers that the Police and Justice Act 2006 (Computer Misuse) has added.

|  |
| --- |
| Description |
| Higher penalties for serious computer crimes.  The maximum sentence for serious offences (like hacking with intent to cause damage) increased to up to 10 years in prison. |
| Denial of Service (DoS) attacks became illegal.  It is now a crime to make a computer system go offline or stop working by overloading it with traffic. |
| Making or supplying tools for hacking became a crime  It is illegal to create, share or sell software that is designed to hack or damage computer systems. |

Look at the below website to answer the questions:

<https://www.gov.uk/personal-data-my-employer-can-keep-about-me>

|  |
| --- |
| Write down three items of data which a company can store about an employee. |
| 1. Name |
| 1. Address |
| 1. National Insurance number |

|  |
| --- |
| Give three more examples of data that an employer can only store if they first get the employee’s permission. |
| 1. Race and ethnicity |
| 1. Religion |
| 1. Health and medical conditions |

Conduct further research to answer the below questions.

|  |  |
| --- | --- |
| Question | Answer |
| Provide one example of: Copyright infringement | **Downloading and sharing a movie from the internet without the permission of the copyright owner.** |
| Provide one example of: Plagiarism | **Copying a paragraph from someone else’s essay or article and submitting it as your own work without giving credit.** |
| What are two consequences of copyright infringement and software piracy? | 1. **Legal action – the person may face fines or even imprisonment.** 2. **Financial loss – companies and creators lose money because their work is used without payment.** |
| Give three possible consequences for individuals when using pirated software | 1. **The computer may get viruses or malware.** 2. **The user may not receive updates or technical support.** 3. **The person could face legal penalties or fines for using illegal software.** |

Listed below are some laws which we have covered today:

1. Computer Misuse Act 1990

2. Police and Justice Act 2006 (Computer Misuse)

3. Copyright, Designs and Patents Act 1988

4. Copyright (Computer Programs) Regulations 1992

5. The Health and Safety (Display Screen Equipment) Regulations 1992

6. Data Protection Act 2018

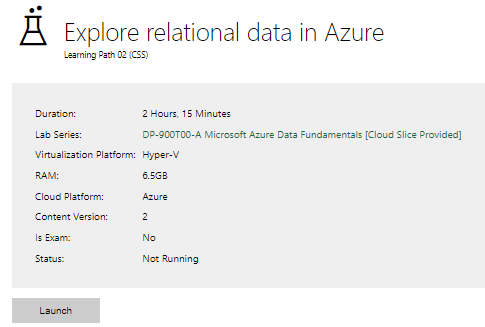
7. Consumer Rights Act 2015

* Insert a number in the first column of each row to match each of the statements with one of the above Acts.
* One of statements is incorrect and not illegal. For this statement, write ‘Not illegal’.

|  |  |
| --- | --- |
| **Act number** | **Clause** |
| 3 | 3. Copyright, Designs and Patents Act 1988.  With some exceptions, it is illegal to use unlicensed software.  This law protects creators’ work (software, music, films). Using software without a proper license is usually illegal. |
| 7 | 7. Consumer Rights Act 2015.  Any product, digital or otherwise, must be fit for the purpose it is supplied for.  This law protects buyers: products (including digital) must work as described and be suitable for their purpose. |
| 1 | 1. Computer Misuse Act 1990  Unauthorised modification of computer material is illegal.  This law criminalises unauthorised changes to computer data or programs (hacking, planting malware). |
| 2 | 2. Police and Justice Act 2006 (Computer Misuse)  It is illegal to create or use a hacking tool for penetration testing.  Makes it illegal to create or use hacking tools for criminal purposes (amendment to the Computer Misuse Act). |
| 6 | 6. Data Protection Act 2018  Personal data may only be used for specified, explicit purposes (no hidden purposes). |
| 5 | 5. The Health and Safety (Display Screen Equipment).  Employers must provide their computer users with adequate health and safety training for any workstation they work at.  Employers must train staff to use computers safely and comfortably. |
| 2 | 2. Police and Justice Act 2006 (Computer Misuse)  It is illegal to distribute hacking tools for criminal purposes.  Sharing hacking tools for illegal use is a crime under this act. |
| Not illegal | It is illegal to distribute an illicit recording. |
| 6 | 6. Data Protection Act 2018  Personal data may not be kept longer than necessary  Organisations must delete or anonymise personal data when it’s no longer needed. |
| 1 | 1. Computer Misuse Act 1990  Gaining unauthorised access to a computer system is illegal.  It is illegal to enter a computer system or network without permission. |
| 5 | 5. The Health and Safety (Display Screen Equipment) Regulations 1992  Employers must ensure that employees take regular and adequate breaks from looking at their screens.  Employees should take breaks to protect their eyes and posture. |
| 2 | 2. Police and Justice Act 2006 (Computer Misuse)  It is illegal to prevent or hinder access (e.g. by a denial-of-service attack) to any program or data held in any computer. |
| 6 | 6. Data Protection Act 2018  Personal data must be accurate and where necessary kept up to date.  Personal data must be correct and updated when needed. |

# Day 3: Task 1

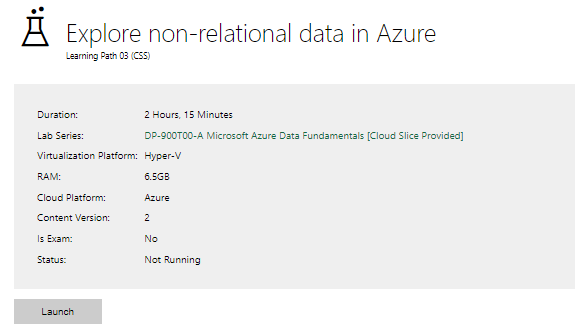
Please complete the below lab (3) *‘Explore relational data in Azure’* and paste evidence of the completed lab in the box provided.



|  |  |
| --- | --- |
| Completed lab |  |

# Day 3: Task 2

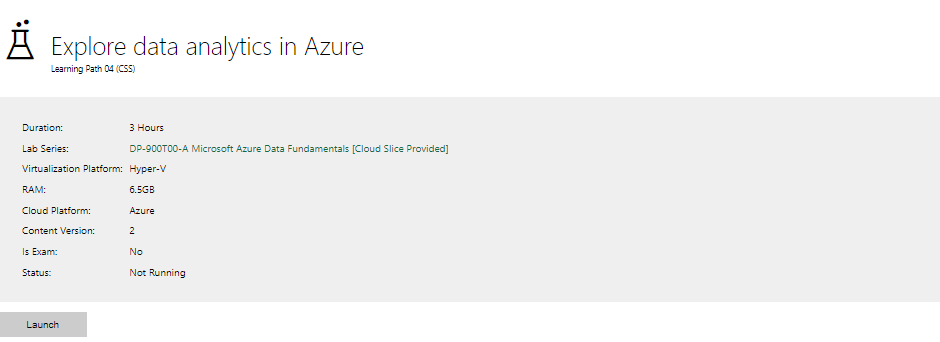
Please complete the below lab (4) *‘Explore non-relational data in Azure’* and paste evidence of the completed lab in the box provided.



|  |  |
| --- | --- |
| Completed lab |  |

# Day 3: Task 3

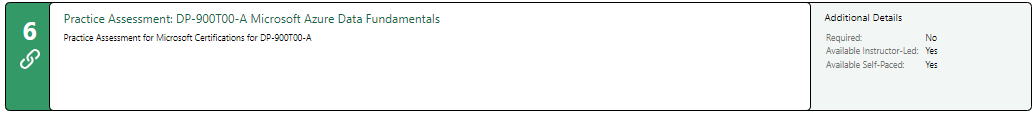
Please complete the below lab (5) ‘Explore data analytics in Azure’ and paste evidence of the completed lab in the box provided.



|  |  |
| --- | --- |
| Completed lab |  |

# Day 4: Task 1

In your teams, complete the Azure DP-900 practice exam and paste your result below – this is open book and please research and discuss your answers as a team.



|  |  |
| --- | --- |
| Result |  |

# Day 4: Task 2

#### **1. Scenario Background**

"Paws & Whiskers" is a growing pet shop that aims to improve its business by analysing sales, customer information, and inventory data. Currently, the data is collected manually or stored in spreadsheets. Management is interested in transitioning to Microsoft Azure to streamline data storage, analysis, and reporting, enabling them to make data-driven decisions.

#### **2. Data Laws and Regulations**

Identify and explain the data laws and regulations relevant to handling customer data within the proposal. Ensure you cover the following points:

* **GDPR Compliance**: Highlight the importance of adhering to the General Data Protection Regulation (GDPR), particularly as it relates to storing and processing customer information.
* **Data Protection Act (DPA) 2018**: Outline how the DPA 2018 may affect the way "Paws & Whiskers" collects and stores data, ensuring compliance with UK laws on data privacy.
* **Other Industry Standards**: Research any additional data protection standards or regulations that may apply to pet shop data, particularly if they involve sensitive or payment information.

#### **3. Azure Service Recommendations**

Recommend Microsoft Azure services that would suit the company’s data analysis needs and explain why these services are suitable. Your recommendations should include:

* **Data Storage**: Identify suitable storage options, such as **Azure Blob Storage** or **Azure SQL Database**, and discuss the benefits of each for storing large datasets, including inventory, sales transactions, and customer details.
* **Data Analysis Tools**: Recommend tools such as **Azure Machine Learning** for customer behaviour analysis or **Azure Synapse Analytics** for analysing sales trends.
* **Data Integration and Automation**: Explain how services like **Azure Data Factory** could automate data collection and integration processes, improving efficiency.

#### **4. Data Types and Data Modelling**

Define the types of data "Paws & Whiskers" will need to work with and describe your approach to data modelling:

* **Data Categories**: Identify key data types, such as customer demographics, transaction history, pet inventory, and product categories.
* **Data Modelling Approach**: Outline how you would structure this data using a relational model or a data warehouse approach, considering factors like tables, entities, relationships, and primary keys.

#### **5. Data Storage Formats and Structures in Azure**

Discuss how you would store data within Azure and the formats you would recommend:

* **Data Formats**: Specify recommended formats (e.g., CSV for raw data imports, JSON for structured data, Parquet for analytics) and explain why these formats are suitable for specific data types.
* **Data Security and Encryption**: Include recommendations for securing data using Azure’s built-in encryption features and access controls to ensure compliance with data privacy regulations.

#### **6. Additional Considerations**

Provide any other considerations that might enhance data handling and efficiency in Azure, such as:

* **Backup and Disaster Recovery**: Outline a backup plan using **Azure Backup** or **Azure Site Recovery** to safeguard against data loss.
* **Data Visualisation**: Discuss potential use of **Power BI** within Azure for creating dashboards that provide management with real-time insights into sales and customer trends.
* **Future Scalability**: Comment on how Azure services can scale as the business grows, accommodating larger datasets and more complex analyses.

### **Submission Guidelines:**

1. **Structure**: Ensure your report is well-organised, with sections for each task (e.g., Data Laws, Azure Services, Data Types, etc.).
2. **Formatting**: Include headings, bullet points where appropriate, and any visuals or diagrams that support your explanations.
3. **References**: Cite any resources or regulations referenced in the report.
4. **Length**: Aim for 1500-2000 words.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Section 1: Introduction.  "Paws & Whiskers" is a small but growing pet shop that sells pet food, accessories, and live animals such as fish, hamsters, and birds. The company wants to improve its business decisions by analyzing customer behavior, sales performance, and inventory levels. At the moment, most data is stored in Microsoft Excel spreadsheets or collected manually by staff. This creates several problems: files can be lost or duplicated, it is difficult to find information quickly, and there is no clear way to generate accurate reports or forecast future sales. As the company grows, Excel becomes inefficient and cannot support data analysis at a larger scale.  Management is interested in moving to Microsoft Azure, a cloud-based platform that provides tools for data storage, automation, security, and analytics. Using Azure will allow the company to centralize all data in one place, automate data updates, protect customer information in line with legal requirements, and use tools such as Power BI or Azure Synapse Analytics to create reports and dashboards. This move from Excel to the cloud is an important step towards digital transformation and data-driven decision-making.  In this report, I will analyze the legal responsibilities of handling customer data, recommend Azure services for storage and analysis, describe key data types, and explain how to model data correctly for better reporting. I will pay special attention to data migration from Excel to Azure and explain how proper data modeling improves business efficiency. Finally, I will discuss additional considerations such as backup, security, data visualization, and scalability.  Section 2: Data Laws and Regulations.  When moving customer data to the cloud, "Paws & Whiskers" must follow data protection laws to ensure that personal information is stored and processed legally and safely. Three important laws and standards apply to this case: GDPR, the Data Protection Act 2018, and additional rules for financial and sensitive data.   * **GDPR (General Data Protection Regulation): (Legal compliance)**   GDPR is a European regulation that protects customer privacy. The company must collect data only for a clear purpose (for example, processing sales or sending order confirmations). Customers must agree to the storage of their personal details, such as name, address, and contact information. GDPR also gives customers the right to request access to their data, correct it, or ask for it to be deleted (“right to be forgotten”). Azure supports GDPR compliance through encryption, access control, and secure storage options.   * **Data Protection Act (DPA) 2018: (Legal compliance)**   The DPA 2018 is the UK version of GDPR and is legally required for businesses operating in the UK. It explains how businesses must store data securely and prevent unauthorized access. For example, customer data must not be shared with third parties without permission. Azure provides role-based access control (RBAC), meaning only authorized employees can view sensitive customer records.   * **PCI-DSS (Payment Card Industry Data Security Standard): (Industry standard)**   In addition to GDPR and the DPA 2018, the company should also be aware of industry-specific standards. For example, if payment card information is held, a recognized standard like PCI-DSS (Payment Card Industry Data Security Standard) may apply to ensure secure payment processing. While this standard may not always be required for every retail business, it is good practice for any shop that accepts card payments.   * **Azure AD (Azure Active Directory): (Azure compliance support)**   If customer data includes sensitive data such as health information about pets (e.g., medical treatments), this data must be stored securely using encryption and access limitations. Azure provides built-in encryption and access control, such as Azure Active Directory (Azure AD), which limits access to authorized employees only.  By following these laws, the company can avoid legal penalties, protect customer trust, and ensure safe use of cloud services. Section 3. Azure Service Recommendations To move from Excel files to an intelligent cloud-based system, the “Paws & Whiskers” pet shop needs services that can:   * store data * analyse it * update it automatically * protect it from loss and cyberattacks * present it in dashboards and charts   Microsoft Azure provides all of these tools in one platform. **1. Data Storage services**  * **Azure Blob Storage** – used to store large files such as Excel exports, product images, receipts, and animal photos. It works like a cloud folder with low cost and the ability to store large amounts of files. * **Azure SQL Database** – a relational database used to store structured data such as customers, products, sales, stock levels, and transactions. It works like an advanced version of Excel with formulas, relationships, queries, and reports. * **Azure Table Storage** – used to store simple semi-structured data such as product categories (for example: Cats, Dogs, Birds). It is often used for quick lookup information.   For long-term reporting and analytical queries, Azure SQL Database is the best option because it supports relationships between tables. **2. Analytics Services (help to understand and predict performance)**  * **Azure Synapse Analytics** – analyses large amounts of historical sales data and helps discover trends (e.g., which products sell more during December: toys or food). * **Azure Machine Learning** – helps to create predictions, such as: “Customers who buy cat food may buy toys after two weeks.” * **Power BI** (connected to Azure SQL Database) – visualises data in dashboards and charts. For example, it can show monthly sales, the most popular products, or inventory levels.   Example: The manager can open a dashboard and see: “40% of January sales are dog food, profit dropped by 10%, we need a promotion.” **3. Data Integration and Automation services (instead of manual uploading)**  * **Azure Data Factory** – automatically collects Excel files and loads them into Azure SQL on a schedule. For example, every evening the “Sales.xlsx” file can be transformed into a structured ‘Sales’ table. * **Azure Logic Apps** – runs automatic actions or sends notifications. For example: “If product stock is less than 10 units, send an email to the supplier.”   Before Azure, employees manually exported Excel files. Now the system automates this process and reduces errors. 4**. Security and Access Control services (who can access which data)**  * **Azure Active Directory (Azure AD)** – controls which employees can access specific data. For example, cashiers can only view sales, while managers can access all business information. * **Azure Key Vault** (optional for future expansion) – securely stores sensitive passwords and encryption keys to protect login details or system connections.   These services help protect customer information and support compliance with GDPR and the DPA 2018. Final Business Outcome After implementing Azure, the pet shop gains:   * one central database instead of many separate Excel files * reduced data loss and human errors * automatic forecasts and analysis * visual reports in Power BI (or Tableau) * protection of customer information * and the ability to scale as the business grows.   Section 4. Data Types and Data Modelling.  The “Paws & Whiskers” pet shop works with different types of data that are important for daily operations and business analysis. These include customer demographics, transaction history, pet and product inventory, and product categories.  **1. Data types and data categories**  Table 1. Customer demographics (example data types)   |  |  |  | | --- | --- | --- | | Column | Description | Data Type | | CustomerID | Unique customer ID | Integer | | Name | Customer name | Text (String) | | City | City of residence | Text (String) | | Email | Email address | Text (String) | | PetPreference | Favourite pet type (cats, dogs, etc.) | Text (String) | | PhoneNumber | Customer phone | Text (String) |   Data types used here: text (String) and numeric identifier (Integer).  Table 2. Transaction / Sales history   |  |  |  | | --- | --- | --- | | Column | Description | Data Type | | SaleID | Sale number | Integer | | CustomerID | Link to customer | Integer | | ProductID | Link to product | Integer | | SaleDate | Date of purchase | Date | | TotalAmount | Total sale amount | Decimal |   Data types used: Integer, Date, Decimal (floating-point number).  Table 3. Product categories and details   |  |  |  | | --- | --- | --- | | Column | Description | Data Type | | ProductID | Unique product ID | Integer | | ProductName | Name of product | Text | | Category | Cats, dogs, birds, etc. | Text | | Price | Product price | Decimal |   Data types used: Text and Decimal.  Table 4. Product / Pet inventory   |  |  |  | | --- | --- | --- | | Column | Description | Data Type | | ProductID | Product reference | Integer | | StockLevel | Quantity in stock | Integer | | ReorderLevel | Minimum stock threshold | Integer | | InStock | Availability (yes/no) | Boolean |   Data types used: Integer and Boolean (True/False). **2. Problems with Excel** Currently, data is stored in different Excel files, such as “Customers.xlsx”, “Sales.xlsx” and “Inventory.xlsx”. This causes several issues:   * Data is duplicated and may contain errors. * If one file is updated, others may become outdated. * There are no relationships between customers, sales, and products. * It is difficult to analyse sales by customer or product category. * Reporting requires a lot of manual work.   **3. Data migration process to Azure (step-by-step)**  Table 5. Migration process   |  |  |  |  | | --- | --- | --- | --- | | № | Step | What happens | Service used | | 1 | Excel Files | Raw data (Sales.xlsx, Customers.xlsx) | Local source files | | 2 | Azure Blob Storage | Files stored in the cloud | Central storage | | 3 | Azure Data Factory | Data extraction and cleaning | Automated data transfer | | 4 | Azure SQL Database | Data stored as structured tables | Tables: Customers, Products, Sales | | 5 | Power BI / Tableau | Data visualisation | Reports, trends, profit, inventory | | 6 | Business Insights | Analytics → actions | Forecasts, marketing, reordering |  **Explanation of the process:**  1. Upload Excel files to Azure Blob Storage as original raw data. 2. Use Azure Data Factory (ADF) to extract the data (ADF is a tool that “takes” data from Excel and prepares it for loading). 3. Transform the data into structured tables.   Excel files may contain extra labels or empty rows and must be cleaned.  Azure Data Factory can:   * remove empty rows * split columns (e.g., “Customer Name” → FirstName + LastName) * rename fields (e.g., “Qty” → “Quantity”) * correct missing values  1. Load them into Azure SQL Database as structured tables.   Now data from Excel is converted into rows and goes into SQL tables.  Examples:   * Customer demographics.xlsx → Customers table * Sales history.xlsx → Sales table * Product categories and details.xlsx → Products table * Product / Pet inventory.xlsx → Inventory   Each table already has clear columns: CustomerID, Name, Email, etc.   1. Create relationships between tables using foreign keys What this means:   In a relational database, we connect tables using keys to understand which customer made a purchase and which product was sold.   * The Sales table contains a CustomerID field, which is linked to CustomerID in the Customers table. This allows the system to identify which customer completed each sale. * The Sales table also includes a ProductID field, which is connected to ProductID in the Products table. This helps the database understand which product was bought in that specific sale. * The Inventory table is linked to the Products table through ProductID, which allows tracking the remaining stock after each sale.   Thanks to these connections, the database can, for example, show that Sale №10 was made by a customer named Anna, who bought dog food. After the sale, the stock level of this product is reduced in the Inventory table.  This is an example of data modelling through relationships.   1. Connect Power BI or Tableau to build interactive dashboards and reports.   **4. Data Modelling (Relational Model)**  Table 6. Steps of data modelling   |  |  |  | | --- | --- | --- | | № | Step | Explanation | | 1 | Identify entities | Find real business objects (e.g., customers, products, sales, inventory). | | 2 | Define attributes | Choose which fields describe each entity (e.g., Name, Email, Price). | | 3 | Assign primary keys | Give each table a unique identifier (e.g., CustomerID, ProductID). | | 4 | Define relationships | Understand how tables are connected (e.g., one customer → many sales). | | 5 | Add foreign keys | Insert keys that connect tables (e.g., CustomerID in Sales links to Customers). | | 6 | Normalize the data | Remove duplication and ensure each value is stored only once. | | 7 | Build the logical model | Combine tables and connections into a full design (ER diagram). | | 8 | Implement the physical database | Create tables in Azure SQL Database with primary and foreign keys. |   To improve analysis and avoid duplicated data, a relational model is used. In this model, data is organised into separate tables and connected using keys.  The main tables in the “Paws & Whiskers” database are:   * **Customers** (CustomerID, Name, Email, City, PetPreference, PhoneNumber) * **Sales** (SaleID, CustomerID, ProductID, SaleDate, TotalAmount) * **Products** (ProductID, ProductName, Category, Price) * **Inventory** (ProductID, StockLevel, ReorderLevel, InStock)   Each table has a primary key (e.g., CustomerID, ProductID, SaleID) to uniquely identify records. Connections between tables are made using foreign keys such as CustomerID (in Sales) and ProductID (in both Sales and Inventory).  Table 6. Primary keys   |  |  | | --- | --- | | Table | Primary Key | | Customers | CustomerID | | Products | ProductID | | Sales | SaleID | | Inventory | ProductID |   Table 7. Foreign keys   |  |  |  |  | | --- | --- | --- | --- | | Table | Foreign Key | Refers to a table | Refers to the field | | Sales | CustomerID | Customers | CustomerID | | Sales | ProductID | Products | ProductID | | Inventory | ProductID | Products | ProductID |   Relationships:   * One customer can make many sales (1-to-many relationship). One product can appear in many sales. * Each product has one stock record in the Inventory table.  Optional extension for future growth:In a more advanced model, the database can be expanded by adding a separate SalesDetails table to support sales with multiple products, or a Suppliers table to manage product sourcing. This shows the model is scalable.**5. Benefits of data modelling**  * Reduces data duplication (e.g., customer names are stored only once in the Customers table). * Increases reporting accuracy and performance. * Makes it easier to run queries like “Which products are most popular with dog owners?” * Supports future use of analytical tools such as Azure Machine Learning for sales forecasting. * Allows the system to be easily expanded as the business grows.   5. Data Storage Formats and Structures in Azure  When moving data to Azure, it is important to choose suitable file formats for storage, analytics, and reporting. Different formats are used for different purposes, depending on how structured the data is and how it will be processed. **5.1 Recommended data formats**  * **CSV (Comma-Separated Values)** – This format is ideal for importing raw data from Excel files when it is first uploaded to Azure Blob Storage. * Stores data in rows and columns, where values are separated by commas (or semicolons). * Lightweight and readable. * Suitable for initial imports and transfers between systems.   Example of a line in CSV:  1, Anna Wilson, anna@mail.com, London, Cats   * **JSON (JavaScript Object Notation)** – Useful for semi-structured data, especially when storing customer preferences (favorite breed, preferred pet food brand, purchase frequency) or product attributes that may vary. JSON can store that data. JSON can be used in Azure services such as Azure Cosmos DB or APIs. * **Parquet** – A column-based format, specially designed for analysing large amounts of data. * It stores data by columns instead of rows, which makes it much faster for analytical queries such as “find the total sales by category.” * It provides excellent compression, so it uses less storage space. * It is commonly used in Azure Synapse Analytics, Azure Data Lake, and Power BI for data analysis and reporting. * **SQL tables in Azure SQL Database** – A structured table format with clear columns, defined data types, and relationships between tables.      * Used for daily business operations such as adding a new customer, recording a sale, or checking stock levels. * Supports relationships using keys, as well as SQL queries, filtering, and report generation. * Forms the foundation for building dashboards in Power BI.     Why it is used in our case:  After the CSV data is cleaned and transformed, it is loaded into SQL tables (such as Customers, Sales, Products, Inventory) and used to create business reports and sales forecasts.  Table 8. Data formats in the analytics process   |  |  |  |  |  | | --- | --- | --- | --- | --- | | № | Stage | What happens | Data format | Why it is used | | 1. | Data collection | Managers upload Excel files with sales and customer data | CSV | Easy to import and read | | 2. | Additional information | Customer online forms store pet preferences | JSON | Suitable for flexible and semi-structured data | | 3. | Processing and analytics | Data is cleaned and converted into a format for analysis | Parquet | Works fast with large datasets | | 4. | Storage and reporting | Final data is stored in Azure SQL Database | SQL | Supports queries, joins, and reporting |  * CSV — is the raw data import. * JSON — is used in parallel for additional data. * Parquet — is an intermediate analytical format. * SQL — is the final storage from which reports and Power BI dashboards are built.  **5.3 Data Security and Encryption** When storing customer and transactional data in the cloud, security is very important. Azure provides built-in tools to protect data at every stage: during storage, during transfer, and during access. **Security features and recommendations:**  * **Encryption at rest** – Azure automatically encrypts data when it is stored in Blob Storage, Azure SQL Database or Data Lake. This protects files even if someone gets physical access to the storage. * **Encryption in transit (HTTPS/TLS)** – All data transferred between Azure services or when users connect to the database is protected using secure HTTPS connections. * **Role-Based Access Control (RBAC)** – Using Azure Active Directory (Azure AD), the company can assign roles such as “Manager”, “Sales staff” or “Data analyst”, so that each employee can only access the data they are allowed to see. * **Firewall rules and network restrictions** – Azure SQL Database supports IP whitelisting, which allows access only from trusted devices or company networks. * **Optional: Azure Key Vault** – If the company needs to securely manage encryption keys or passwords, Azure Key Vault can be used as a safe storage for secrets.   These features help “Paws & Whiskers” follow GDPR and Data Protection Act (DPA 2018) requirements and protect customer trust. Section 6: Additional Considerations In addition to data storage, modelling and security, there are several other important factors that can improve data efficiency and business performance in Azure. **6.1 Backup and Disaster Recovery** To protect the company from data loss due to technical failures or accidental deletion, a backup strategy is necessary.   * Azure Backup can automatically create daily or weekly backups of Azure SQL Database. * Azure Site Recovery can be used to restore data and continue operations in case of a major system failure. This ensures business continuity and protects customer and sales records.  **6.2 Data Visualisation with Power BI** For managers to make fast and informed decisions, visual dashboards are useful.   * Power BI, connected to Azure SQL Database, can display real-time charts showing sales by product category, customer trends, or stock levels. * For example, a dashboard could show that "dog food sales increased by 15% this month", helping management plan promotions or stock orders. This improves understanding of business performance.  **6.3 Future Scalability** As “Paws & Whiskers” grows, the amount of data from customers, products and sales will increase.   * Azure services such as Azure SQL Database and Azure Data Lake can scale automatically to handle larger datasets.   The company can also expand the database by adding new tables in the future (for example, Suppliers, Marketing Campaigns or Detailed Sales). Scalability ensures that the system remains fast and efficient even when the business expands. Conclusion The transition from Excel-based data storage to Microsoft Azure provides “Paws & Whiskers” with a more reliable, scalable and intelligent data management system. By storing data in Azure SQL Database, using Azure Data Factory for automation and applying a relational model, the company can reduce human errors, improve reporting accuracy and support future analysis.  Different data formats such as CSV, JSON, Parquet and SQL are used at various stages to ensure efficient storage and fast analytics. Security features in Azure, including encryption, role-based access control (RBAC) and backup options, help the company comply with GDPR and DPA regulations while protecting customer trust.  With Power BI dashboards and the ability to scale in the future, the company will be able to better understand customer behaviour, track inventory levels and make informed, data-driven decisions. Overall, Azure enables “Paws & Whiskers” to modernise its data environment and support long-term business growth. |

|  |
| --- |
| **Course Notes** |

It is recommended to take notes from the course, use the space below to do so, or use the revision guide shared with the class:

|  |
| --- |
| **Additional Information** |

We have included a range of additional links to further resources and information that you may find useful, these can be found within your revision guide.

**END OF WORKBOOK**

**Please check through your work thoroughly before submitting and update the table of contents if required.**

**Please send your completed work booklet to your trainer.**