

**Data Technician**

|  |
| --- |
|  |

|  |
| --- |
| Name: Nelesh Parmar |
| Course Date: 12 May 2025 |
|  |

**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? | File sharing.  Online school classes.  Streaming services.  Data backup.  IOT (Internet of Things).  Connect from anywhere (internet required) |
| How can it benefit a business? | Reduce cost.  Better performance, as you can scale resources when needed.  Remote access globally, as long as you have internet availability.  Scalability on demand.  Flexibility.  Machine Learning to help identify trends.  Storing data in the cloud protects it from hardware failures and cyber threats.  Software as service. |
| What’s the alternative to cloud computing? | On-premises infrastructure.  Hybrid cloud - Combination of On-premise and Cloud resources.  Private Cloud - Vmware, Openstack.  Co-location - Business own hardware but not data centre.  Edge computing - Processing data closer to its source (like IoT devices) instead of relying on a centralized cloud. |
| What cloud providers can we use, what are their features and functions? | Amazon Web Services (AWS) – Offers a vast array of services, including computing power, storage, databases, AI, and machine learning. Known for its scalability and global reach.  Microsoft Azure – Provides enterprise-grade solutions, including AI, analytics, security, and hybrid cloud capabilities. Ideal for businesses integrating with Microsoft products.  Google Cloud Platform (GCP) – Specializes in AI, machine learning, and data analytics. Offers strong support for Kubernetes and open-source technologies.  IBM Cloud – Focuses on AI-driven solutions, hybrid cloud, and enterprise security. Great for businesses needing advanced data processing.  Oracle Cloud – Best for database management and enterprise applications. Offers high-performance computing and security features.  Alibaba Cloud – A leading provider in Asia, offering cloud computing, AI, and big data solutions. |

# 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) | Provides virtualized computing resources like servers and storage. | When businesses need scalable, cost-effective computing resources without managing physical hardware.  In the real-world: **Betfair** and **Netflix** use IaaS to support its operations for computing, storage, and networking.  Big Data & Analytics – Companies use IaaS to process massive datasets efficiently, leveraging cloud-based computing power. E-commerce & Web Hosting – Online retailers and websites rely on IaaS for reliable hosting and traffic management.  Disaster Recovery – Businesses store backup data in the cloud, ensuring quick recovery in case of system failure. |
| PaaS (Platform as a service) | Offers a platform for developing, running, and managing applications. | When businesses or developers need a ready-to-use platform for building, deploying, and managing applications without handling the underlying infrastructure. It streamlines development, making it faster and more efficient.  In the real-world:  **Netflix** use PaaS to provides developers with a structured environment to build, deploy, and scale applications seamlessly. This agility is vital for a service that delivers a vast array of content globally. |
| SaaS (Software as a service) | Delivers software applications over the internet without requiring local installation. | Ideal when businesses or individuals need software that is accessible from anywhere, without the hassle of installation or maintenance. It’s commonly used for collaboration, data management, and automation.  In the real-world:  **Salesforce** and **HubSpot** to manage customer interactions and sales pipelines.  **Netflix** and **Spotify** deliver entertainment through cloud-based SaaS models |

# 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 | Cloud computing model where IT infrastructure (servers, networking, storage) is provided by a third-party vendor and accessible over the internet. It is shared among multiple users and organizations.  Ideal for businesses that need scalability and cost-efficiency without managing physical infrastructure. Suitable for businesses who require on-demand resources.  Netflix uses public cloud services to stream content globally. It relies on cloud providers like AWS or Google Cloud to scale its infrastructure based on demand. |
| Private Cloud | Cloud environment dedicated to a single organization, either hosted on-premises or by a third-party provider. It offers greater control, security, and customization.  Best for organizations handling sensitive data or requiring strict compliance (e.g., healthcare, finance).  Suitable for enterprises that need customized infrastructure and enhanced security.  JPMorgan Chase (Bank) uses a private cloud to store and process financial transactions securely while maintaining regulatory compliance. |
| Hybrid Cloud | A hybrid cloud combines public and private cloud environments, allowing data and applications to move between them for flexibility, scalability, and security.  Ideal for businesses that need both security and scalability.  Suitable for organizations with legacy systems that need cloud integration.  Mayo Clinic (Healthcare provider) uses a hybrid cloud to store patient records securely in a private cloud while leveraging public cloud services for analytics and research. |
| Community Cloud | Is a shared cloud infrastructure used by multiple organizations with common interests, regulations, or security requirements.  Best for industries with shared compliance needs, such as government agencies or research institutions.  Suitable for organizations that need collaborative environments while maintaining security.  A government agency like the European Union uses a community cloud to share data securely among member states while complying with strict regulations. |

# 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 material. | This refers to entering a computer system without permission. | Hacking into computer system. |
| Unauthorised access to computer materials with intent to commit a further crime. | This refers to entering a computer system to steal data or destroy a device or network. | Planting a virus. |
| Unauthorised modification of data. | This refers to modifying or deleting data, and also covers the introduction of *malware* or *spyware* onto a computer. | Installing malware |

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 |
| Increased Penalties for Unauthorized Access – The maximum sentence for unauthorized access to computer material was increased from six months to two years |
| Criminalization of Denial-of-Service (DoS) Attacks – The Act explicitly made DoS attacks illegal, addressing a previous legal loophole |
| Offense of Making or Supplying Hacking Tools – It became an offense to create, supply, or obtain hacking tools intended for committing cybercrimes |

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. |
| Name |
| Address |
| Date of birth |

|  |
| --- |
| Give three more examples of data that an employer can only store if they first get the employee’s permission. |
| Biometrics, for example if your fingerprints are used for identification |
| Health and medical conditions |
| Race and ethnicity |

Conduct further research to answer the below questions.

|  |  |
| --- | --- |
| Question | Answer |
| Provide one example of: Copyright infringement | AI companies using copyrighted data to train their models. In recent years, several tech giants have faced legal action for allegedly scraping copyrighted content, such as articles, books, and images without permission to develop AI systems. |
| Provide one example of: Plagiarism | Researchers or organizations copy and present scientific data from another study, as their own findings, leading to ethical violations and retraction. |
| What are two consequences of copyright infringement and software piracy? | Legal Penalties – Offenders may face fines, lawsuits, or even imprisonment, depending on the severity.  Financial Losses – Businesses and creators suffer revenue loss, as piracy reduces legitimate sales and discourages innovation. |
| Give three possible consequences for individuals when using pirated software | 1. Individuals caught using pirated software may face fines or legal action. 2. Pirated software often lacks security updates, making users vulnerable to malware, viruses, and hacking attempts. 3. Users of pirated software cannot access official technical support or updates, leading to potential system failures and compatibility issues. |

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** |
| 1988 | With some exceptions, it is illegal to use unlicensed software |
| 2015 | Any product, digital or otherwise, must be fit for the purpose it is supplied for |
| 1990 | Unauthorised modification of computer material is illegal |
| Not illegal | It is illegal to create or use a hacking tool for penetration testing |
| 2018 | Personal data may only be used for specified, explicit purposes |
| 1992 | Employers must provide their computer users with adequate health and safety training for any workstation they work at |
| 2006 | It is illegal to distribute hacking tools for criminal purposes |
| 1998 | It is illegal to distribute an illicit recording |
| 2018 | Personal data may not be kept longer than necessary |
| 1990 | Gaining unauthorised access to a computer system is illegal |
| 1992 | Employers must ensure that employees take regular and adequate breaks from looking at their screens |
| 1990 | 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 |
| 2018 | Personal data must be accurate and where necessary kept up to date |

# 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 | 15/05/2025 – Lab not working. |

# 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 | Need more practise 😊 |

# 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.

|  |
| --- |
| Proposal for transitioning to Microsoft Azure **Background**  "Paws & Whiskers" are a growing pet shop aiming to enhance business operations by leveraging Microsoft Azure for data storage, analysis, and reporting.  Currently, data is manually collected or stored in spreadsheets, limiting efficiency and scalability. Using Azure will streamline data management, improve decision-making, and enhance customer insights.   1. **Data Laws and Regulations**   When collecting or handling data, we must abide to legal data laws and regulations that are in place to protect customer privacy and maintain trust. The following laws and regulations have been identified:  The **General Data Protection Regulation** (GDPR) mandates strict guidelines that governs how personal data is collected, stored, and processed.  Customers will receive a questionnaire detailing what types of personal data we will be collecting and the purpose (enhance customer experience and marketing).  We will ensure that all data must be collected and used transparently and only necessary data will be stored securely.  As part of our security, we need to restrict and control privileged access rights, keep logs of user access, and regularly review and adjust access rights based on employee roles and changes, so only the specific users can view the data and encryption will be used to prevent sensitive information such as, financial details and personal data from being exposed to hackers.  Encryption safeguards communications, ensuring that personal transactions remain confidential and ensures that data is not altered or tampered with while being transmitted or stored.  Customers can request access, correction, or deletion of their data.  Customers should provide explicit consent for data use, especially for marketing.  The **Data Protection Act** (DPA) **2018** aligns with GDPR but includes UK-specific provisions.  It affects Paws & Whiskers by:  Defining lawful data processing.  Establishing stricter rules for sensitive data or special category data (e.g. payment details, racial or ethnic origin, political opinions, health data, data concerning sex life or sexual orientation).  Requiring businesses to implement security measures, such as access controls and data is stored and processed securely.  Consent logs must be maintained.  ***Other Industry Standards***  The Payment Card Industry Data Security Standard (PCI DSS) applies when handling credit/debit card data and requires secure transmission and encryption.  Veterinary data protection for pet medical records.   1. **Azure Service Recommendations**   Microsoft Azure offers many data storage services for both relational and non-relational databases depending on the business requirements.  Being in the cloud has many benefits, such as:   1. Scalability - Adjust resources based on demand without having to buy new hardware. 2. Cost-effectiveness - Only pay for what you use, reducing upfront capital expenses. 3. Accessibility - Access data and applications from anywhere with an internet connection. 4. Security - Cloud providers invest heavily in security measures to protect data.   **Data Storage**  We must consider that there will be a mixture of data from customer profiles, inventory, and sales to images of pets/products that would suit the company’s data analysis needs.  Recommend the following:  a. Azure SQL Database  For structured data like customer profiles, inventory, and sales.  Fully managed and supports relational queries.  Provides built-in security, automatic backups, and scalability.  b. Azure Blob Storage  For unstructured data like images of pets/products and logs.  Cost-effective and scalable.  Supports multiple formats (CSV, JSON, Parquet).  **Data Analysis Tools**  Microsoft Azure offers many tools to assist the organisation improve decision-making, driving innovation and enhance customer insights.  Recommended the following:   1. **Azure Synapse Analytics**: A powerful analytics service that helps businesses process, analyse, and visualize large volumes of data. It's used for various tasks such as big data processing, business intelligence, machine learning, and real-time analytics.   Provides advanced analytics for historical sales trends, inventory turnover, and customer behaviour.   1. **Azure Machine Learning**: Empowers businesses to harness AI for predictive analytics, automation, and intelligent decision-making   Provides predictive analytics such as customer behaviours, product recommendations and purchasing trend detections.   1. **Data Integration and Automation**: Allows businesses to harness AI for predictive analytics, automation, and intelligent decision-making. To improve efficiency, automate data collection and integration processes.   **Azure Data Factory**: Automates data extraction, transformation, and loading (ETL). Supports integration from Excel, on-prem sources, or other SaaS tools.  Scheduled pipelines ensure up-to-date reports.  Provides reduced manual input, real-time or batch data processing and centralised data governance.   1. **Data Types and Data Modelling**   We have considered the business data types and categorized them, below:  **Data Categories**   1. Customer Data: Name, contact, pet details, preferences 2. Transaction Data: Purchase history, payment method, date/time 3. Product Inventory: Item name, category, supplier, stock levels 4. Pet Inventory (e.g., live animals): Species, breed, age, health records   **Data Modelling Approach**  To structure the data, recommend using a relational model, as majority of the data will be structured and entities and relationship will be maintained.  Relational model in Azure SQL Database to handle the following data categories  Customers (CustomerID, Name, Email, Phone)  Sales (TransactionID, CustomerID, ProductID, Date, Quantity, Revenue)  Products (ProductID, CategoryID, Price, Quantity)  Pets (PetID, CustomerID, Breed, Age)  Entity Relationship Diagram:    Tables for customers, transactions, inventory, pets.  Primary keys for unique identification.  Relationships linking purchases to customers and pets to customers.  Data Warehouse option can be considered in the future, as the business grows and for analytical workloads, create a star schema in Synapse Analytics for faster querying.   1. **Data Storage Formats and Structures in Azure**   To store data in Azure, the following **Data Formats** are recommended:   1. CSV for raw data imports, as widely supported and is human readable. 2. JSON for structured data exchange of customer/ sales/ pet data, its lightweight and ideal for API’s. 3. Parquet for efficient analytics (reports), for large queries.   Blob Storage structure, supports all the above formats.  **Data Security and Encryption**  To secure the data, Azure offers built-in encryption ensures data protection.  Role-based access control (RBAC) restricts unauthorized access. Grant access by role (e.g. Analyst, Admin, Manager, Staff).  Encryption in Transit: HTTPS ensures secure data movement.  Private Endpoints: Limit data access to internal networks.  Azure Key Vault: Manage secrets, API keys, certificates securely.   1. **Additional Considerations**   **Backup and Disaster Recovery**  To mitigate against data loss, we should consider Azure resources for backup of data and disaster recovery.  **Azure Backup** ensures data protection against loss. Backs up VMs, SQL Databases, and file shares.  Automatic daily snapshots with retention policies.  **Azure Site Recovery** provides disaster recovery for business continuity. Ensures business continuity by replicating critical workloads to secondary locations.  **Data Visualisation**  To help the business interpret the data, we can use data visualisations, such as Power BI (Integrated with Azure) which enables real-time dashboards for:  Sales trends - helps to determine which products are and are not performing.  Customer trends - evolving patterns and behaviours of consumers in purchasing goods and services influenced by factors such as technological advancements, economic conditions, and cultural shifts.  Inventory alerts - helps the business to know which products are running low on stock.  Connects directly to Azure SQL and Synapse.  Benefits are easy-to-use interface, mobile access and custom report sharing with stakeholders.  **Future Scalability**  As the business grows, we must consider how to scale and Azure services offer a highly scalable and available environment. Azure scale vertically (more resources) or horizontally (more instances).  Serverless options (e.g. Azure Functions) reduce operational overhead.  Azure Cost Management tools help optimise resource usage.    Azure services scale dynamically, accommodating larger datasets and complex analyses.  **Conclusion**  Migrating "Paws & Whiskers" to Microsoft Azure will enhance operational efficiency, ensure data compliance, and unlock powerful insights from their data. By using secure storage, automated data pipelines, and scalable analytics tools, the business can evolve into a modern, data-driven retailer. |

|  |
| --- |
| **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.**