

**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 individuals and organisations to **store, process, and access data and applications over the internet** instead of relying on local servers or personal computers.  In real-world use:   * **File storage and sharing:** Services like Google Drive or OneDrive let users store and access files from any device. * **Streaming and entertainment:** Netflix, Spotify, and YouTube use the cloud to deliver media to millions of users simultaneously. * **Education and remote work:** Platforms like Microsoft Teams, Zoom, and Google Classroom run on cloud infrastructure. * **Smart devices and IoT:** Cloud systems connect devices like home assistants, smart thermostats, and wearable tech. * **Healthcare and finance:** Hospitals and banks use the cloud for secure data storage, analytics, and real-time access to client or patient information. |
| How can it benefit a business? | Cloud computing offers multiple advantages for businesses of all sizes:   * **Cost savings:** Reduces the need for physical hardware and IT maintenance costs. You pay only for what you use. * **Scalability:** Businesses can easily scale resources up or down depending on demand (e.g., seasonal sales). * **Remote access:** Employees can access company systems and data from anywhere, improving flexibility and productivity. * **Data security and backup:** Cloud providers offer built-in redundancy and disaster recovery options. * **Collaboration:** Teams can work on shared documents and projects in real time. * **Innovation and speed:** Businesses can deploy new apps and services quickly using cloud infrastructure |
| What’s the alternative to cloud computing? | The main alternative is **on-premises (on-prem)** or **local infrastructure**:   * Data, software, and servers are hosted **within a company’s own facilities** instead of online. * The business is responsible for **managing, updating, and securing** all equipment and systems.   **Advantages of on-premises:**   * Full control over data and systems. * May meet strict data privacy or regulatory needs.   **Disadvantages:**   * Higher upfront costs for hardware and maintenance. * Limited scalability and flexibility. * More risk of data loss without professional backup systems. |
| What cloud providers can we use, what are their features and functions? | |  |  |  | | --- | --- | --- | | **ovider** | **Main Features** | **Key Services / Functions** |  |  |  |  | | --- | --- | --- | | **Amazon Web Services (AWS)** | Largest and most flexible cloud platform. Offers a wide range of tools for computing, storage, AI, analytics, and IoT. | - EC2 (virtual servers)  - S3 (data storage)  - RDS (databases)  - Lambda (serverless computing) |  |  |  |  | | --- | --- | --- | | **Microsoft Azure** | Strong integration with Microsoft Office and Windows Server. Ideal for hybrid cloud setups. | - Azure Virtual Machines  - Azure SQL Database  - Azure AI and Machine Learning  - Azure DevOps |  |  |  |  | | --- | --- | --- | | **Google Cloud Platform (GCP)** | Known for data analytics and machine learning capabilities. Integrates well with Google Workspace. | - BigQuery (data analytics)  - Compute Engine (VMs)  - Cloud Storage  - TensorFlow AI tools |  |  |  |  | | --- | --- | --- | | **IBM Cloud** | Focuses on enterprise-level AI, hybrid cloud, and data security. | - IBM Watson (AI)  - Cloud Kubernetes Service  - Blockchain platform |  |  |  |  | | --- | --- | --- | | **Oracle Cloud** | Strong in database management and enterprise software. | - Autonomous Database  - Cloud Infrastructure (OCI)  - Data analytics tools |  |  |  |  | | --- | --- | --- | | **Alibaba Cloud** | Leading provider in Asia, strong in e-commerce and international business operations. | - Elastic Compute Service (ECS)  - Object Storage Service (OSS)  - DataWorks (data integration) | |

# 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 provides **virtual computing resources over the internet** such as servers, storage, and networking instead of having to buy and manage physical hardware.  You rent the infrastructure and manage your own software on top. | Used by IT teams, developers, and businesses that need **full control of their systems** but want to avoid buying physical servers.  Example:   * **Amazon EC2 (AWS)** or **Microsoft Azure Virtual Machines** let businesses run their own operating systems and applications in the cloud. * A company launching a new website could use IaaS to host it without owning any hardware. |
| PaaS (Platform as a service) | PaaS provides a **ready-made environment for developers** to build, test, and deploy applications without managing the underlying servers or infrastructure. | Used when building apps or software quickly.  Example:   * **Google App Engine** or **Heroku** lets developers write and deploy code directly in the cloud. * A start-up can build and test a new mobile app using PaaS, saving time on setup and maintenance. |
| SaaS (Software as a service) | SaaS delivers **fully functional software applications** over the internet. Users simply log in and use the service no installation or maintenance needed. | **Real-world use:**  Used daily by businesses and individuals for communication, storage, and productivity.  Examples:   * **Google Workspace**, **Microsoft 365**, **Zoom**, **Salesforce**, or **Netflix**. * A school using Google Workspace for assignments or a business using Microsoft Teams for collaboration. |

# 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 owned and operated by a third-party provider (like **Amazon Web Services**, **Google Cloud**, or **Microsoft Azure**) and delivers computing services over the internet to multiple users.  All hardware, storage, and infrastructure are managed by the provider, and users share the same resources securely.  **When appropriate:**   * When cost efficiency, flexibility, and scalability are important. * Ideal for businesses that don’t need strict data privacy or custom infrastructure.   **Real-world example:**   * **Small businesses or start-ups** using **Microsoft Azure** to host websites, run applications, or store data without buying physical servers. |
| Private Cloud | **What it is:**  A **private cloud** is used exclusively by one organisation. It can be hosted **on-site** (in the company’s own data centre) or by a third-party provider but with dedicated infrastructure.  It offers more control, security, and customisation.  **When appropriate:**   * When organisations handle **sensitive data** or must meet **strict security and compliance** standards. * Common in regulated industries.   **Real-world example:**   * **Banks, government departments, or hospitals** using private clouds to store confidential financial or patient data securely. |
| Hybrid Cloud | **What it is:**  A **hybrid cloud** combines **public and private clouds**, allowing data and applications to move between them.  This approach offers both flexibility and control public cloud for scalability and private cloud for sensitive information.  **When appropriate:**   * When a business needs to keep some data private but still wants to use the public cloud for efficiency or backups.   **Real-world example:**   * A **retail company** using a private cloud for customer payment data and a public cloud for running its online store and promotions. |
| Community Cloud | **What it is:**  A **community cloud** is shared by several organisations that have **similar goals, security requirements, or compliance needs**.  It’s jointly managed and used by the community members.  **When appropriate:**   * When multiple organisations in the same sector want to share infrastructure and costs while maintaining a level of privacy.   **Real-world example:**   * **Universities or healthcare organisations** sharing a community cloud to store and access research data securely. |

# Day 2: Task 1

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

|  |  |  |
| --- | --- | --- |
| **Area** | **Description** | **Example** |
| **1. Unauthorised Access to Computer Material** | Accessing a computer, account, or data without permission. This includes logging into someone else’s account or viewing files you are not allowed to. | A person guessing another user’s password to check their emails or access confidential documents. |
| **2. Unauthorised Access with Intent to Commit or Facilitate Further Offences** | Gaining access to a system to carry out another crime, such as fraud, theft, or data misuse. | A hacker accessing a company’s database to steal customer payment details. |
| Unauthorised Modification of Computer Material | Changing, deleting, or corrupting data or programs without permission, often to cause damage or disruption. | Spreading a virus or ransomware that locks or deletes files on someone’s computer. |

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 |
| **1. Making All Forms of Hacking Illegal** The 2006 update made it a crime to make, supply, or obtain hacking tools or software that could be used to break into computer systems.  Example: Creating or selling software designed to steal passwords or access private networks is now illegal, even if you did not use it yourself. |
| **2. Stronger Penalties for Offenders** The punishment for hacking and other computer crimes became much more serious. Offenders can now face up to 10 years in prison for hacking into systems and up to 14 years for crimes that cause serious damage or risk to national security.  Example: Someone who hacks into hospital systems or power networks could receive a long prison sentence because it puts lives or public safety at risk. |
| **3. New Offence: Impairing Computer Operations** A new offence was added for anyone who intentionally damages, deletes, or disrupts computer systems or data even if they did not gain anything from it.  Example: Launching a DDoS attack to crash a website or deleting files on a company server is now a criminal act under this section. |

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. |
| * education and qualifications |
| Sex |
| Name |

|  |
| --- |
| Give three more examples of data that an employer can only store if they first get the employee’s permission. |
| * race and ethnicity |
| Genetics |
| Religion |

Conduct further research to answer the below questions.

|  |  |
| --- | --- |
| Question | Answer |
| Provide one example of: Copyright infringement | Downloading and sharing a movie or song online without permission from the creator or copyright holder. |
| Provide one example of: Plagiarism | Copying information or text from a website or another student’s work and presenting it as your own without giving credit. |
| What are two consequences of copyright infringement and software piracy? | 1. **Legal action or fines** – You could face court charges and have to pay large fines for breaking copyright law. 2. **Loss of reputation and trust** – Businesses or individuals can lose credibility if caught using or distributing pirated content. |
| Give three possible consequences for individuals when using pirated software | * **Malware and viruses** – Pirated software often contains harmful code that can steal data or damage your computer. * **No updates or support** – You cannot get security updates or help from the official software company. * **Legal punishment** – You could face fines or prosecution 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** |
| 4 | With some exceptions, it is illegal to use unlicensed software |
| 7 | Any product, digital or otherwise, must be fit for the purpose it is supplied for |
| 1 | Unauthorised modification of computer material is illegal |
| Not illegal | It is illegal to create or use a hacking tool for penetration testing |
| 6 | Personal data may only be used for specified, explicit purposes |
| 5 | Employers must provide their computer users with adequate health and safety training for any workstation they work at |
| 2 | It is illegal to distribute hacking tools for criminal purposes |
| 3 | It is illegal to distribute an illicit recording |
| 6 | Personal data may not be kept longer than necessary |
| 1 | Gaining unauthorised access to a computer system is illegal |
| 5 | Employers must ensure that employees take regular and adequate breaks from looking at their screens |
| 2 | 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 | 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 | I enabled Map view several time, still wont work |

# 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.

|  |
| --- |
| **1. Scenario Background** As a data analyst, I am recommending an Azure-based architecture that ensures compliance with UK data laws, improves efficiency, and provides a scalable foundation for future data-driven initiatives. **Benefits of the Azure Solution**  * Centralised and secure data storage replacing manual spreadsheets. * Automated pipelines that reduce errors and time spent on manual entry. * Real-time Power BI dashboards that support faster and data-driven decisions. * Compliance with GDPR, DPA 2018, and PCI DSS. * Scalable cloud infrastructure that grows with business needs  **Architecture Summary** **Data Flow:**   1. Point-of-sale, website, and supplier systems export data daily. 2. Azure Data Factory ingests and transforms it into clean CSV or Parquet formats. 3. Data is stored in Azure SQL Database (structured) and Data Lake (raw + historical). 4. Power BI connects to both sources for visual reporting and analytics. 5. Azure Backup and Key Vault protect data and manage encryption.  **2. Data Laws and Regulations****GDPR Compliance** Under the UK General Data Protection Regulation (GDPR), any customer information such as names, addresses, emails, or purchase history must be collected and processed lawfully and transparently. Paws & Whiskers must:   * Obtain consent for marketing communications. * Store personal data securely with access controls and encryption. * Provide customers the right to access, correct, or delete their data. * Retain data only for as long as necessary for business and legal reasons.   These measures protect customer trust and reduce the risk of data breaches or ICO penalties. **Data Protection Act 2018 (DPA)** The DPA complements GDPR by detailing how UK organisations must handle personal data. Paws & Whiskers must ensure:   * Clear accountability for data handling. * Secure storage and processing across systems (especially cloud environments). * Employee awareness and training on privacy practices.  **Other Relevant Standards** If the store accepts card payments online or in-store, it must comply with **PCI DSS** to protect payment card details.  Marketing activities must follow **PECR** (Privacy and Electronic Communications Regulations) for email or SMS marketing consent. **Azure Service Recommendations****Data Storage**  1. **Azure SQL Database** – Best for structured data such as sales transactions, customer details, and product inventory. It supports relational modelling, indexing, and fast queries for analytics. 2. **Azure Data Lake Storage Gen2** – Ideal for storing large or raw datasets such as historical CSVs, receipts, and supplier files. It is cost-effective, scalable, and integrates with Power BI and Azure Synapse.   Together, these provide both **structured (SQL)** and **semi-structured (Data Lake)** storage options, supporting future data warehouse expansion. **Data Integration and Automation** **Azure Data Factory (ADF)** will automate data collection and transfer between sources (e.g., Excel, POS exports, or website databases).   * Automatically schedule daily or hourly data loads. * Perform basic transformations like cleaning and validation. * Reduce manual handling errors and delays.  **Data Analysis Tools**  * **Azure Synapse Analytics** for querying large datasets, building analytical models, and detecting sales trends. * **Power BI** for interactive dashboards showing real-time sales, top-performing products, and customer insights. * **Azure Machine Learning** (future use) to predict seasonal demand, repeat purchases, or stock shortages  **Data Types and Data Modelling****Data Categories** The business will work with several key data types:   * **Customer Data:** Name, contact, loyalty status, purchase history. * **Sales Data:** Product ID, quantity, price, store location, payment method, date/time. * **Inventory Data:** Product name, category, supplier, cost, stock level, reorder threshold. * **Supplier Data:** Supplier ID, name, product range, delivery lead time.  **Data Modelling Approach** A **star schema** is recommended because it supports efficient analysis and is Power BI–friendly.  It separates **dimensions** (descriptive data) from **facts** (numerical measures), enabling quick aggregations. **Example Schema Structure** **Dimensions**   * **DimCustomer:** CustomerID, Name, Email, Location, LoyaltyStatus * **DimProduct:** ProductID, ProductName, Category, SupplierID, Price * **DimStore:** StoreID, StoreName, Region * **DimDate:** DateID, Date, Month, Year   **Fact Tables**   * **FactSales:** SaleID, ProductID, CustomerID, StoreID, DateID, Quantity, TotalAmount * **FactInventory:** InventoryID, ProductID, StockLevel, ReorderPoint, DateID   **Relationships**   * DimCustomer → FactSales (CustomerID) * DimProduct → FactSales & FactInventory (ProductID) * DimStore → FactSales (StoreID) * DimDate → All Facts (DateID)  **Data Storage Formats and Structures in Azure****Data Formats**  * **CSV:** For raw imports from spreadsheets or POS exports. Simple and human-readable. * **JSON:** For structured data coming from APIs or web applications. * **Parquet:** For analytical storage. It compresses efficiently and supports faster queries in Synapse or Power BI.  **Data Security and Encryption** Azure provides strong built-in protection:   * **Encryption at rest:** Azure SQL and Data Lake encrypt data automatically. * **Encryption in transit:** All data transfers use HTTPS or private links. * **Role-Based Access Control (RBAC):** Ensures only authorised users can access sensitive tables. * **Azure Key Vault:** Manages encryption keys and secrets securely.   These measures ensure compliance with GDPR and the DPA 2018 **Additional Considerations****Backup and Disaster Recovery**  * **Azure Backup:** Automatically backs up databases and virtual machines on a schedule. * **Geo-redundant Storage (GRS):** Replicates data to a secondary region for disaster recovery. * **Point-in-Time Restore:** Enables recovery of a SQL database to a previous state if errors occur.  **Data Visualisation** Power BI can directly connect to Azure SQL Database and Data Lake for dynamic dashboards:   * Track daily sales, profit margins, and top-selling categories. * Visualise stock levels to predict restocking needs. * Segment customers by loyalty status or region.  **Scalability** Azure services scale automatically:   * **Data Lake and Synapse** can handle growing data volumes. * **SQL Database** supports vertical or elastic scaling. * **Power BI Premium** can accommodate larger datasets and more concurrent users as the business expands.  **Conclusion** Migrating Paws & Whiskers to Microsoft Azure will transform its data management by consolidating customer, sales, and inventory information into one secure, automated, and scalable environment.  This proposal ensures compliance with UK data laws, enhances data quality, and empowers management with real-time insights through Power BI dashboards.  The result is a data platform that not only supports current reporting needs but also sets the foundation for predictive analytics and 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.**