

**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? | Services like OneDrive, Google Drive, and Amazon Photos allow you to store and access files and photos, allowing you to save important files securely so they're not lost if the device is damaged. It also frees up local storage.  Streaming services like Netflix and Spotify use the cloud to store videos and songs and are able to access them instantly for watchers. Downloading them for offline use is an added positive to the capabilities of the cloud. |
| How can it benefit a business? | * Businesses need space for their cloud storage, which circumvents the need for physical hard drives or servers and gives a cost-effective option for them. * Collaboration is also an important benefit, as shared files that can be edited and being able to video call employees who might not be able to get to work or are simply long-distance, making travel less stressful and more cost effective. (Google Drive or Microsoft Teams would be effective for it) * Flexibility is also important for a company, if a business grows and need more storage, cloud computing allows for scalability. |
| What’s the alternative to cloud computing? | **Destruction:**   * Tape Backups * Combining on-site servers with cloud backups * Multiple servers * Local software * External hard drives * USBs   **Privacy and Control (e.g., NHS' health records or a bank's financial records):**   * Encryption tools |
| What cloud providers can we use, what are their features and functions? | Amazon Web Services   * AWS allows for scalability and has a wide range of tools. * AWS is good for general-purpose use and start-ups to enterprises.   Microsoft Azure   * Offers a hybrid cloud and offers various Microsoft tools for usage. * Microsoft Azure is good for businesses using Microsoft products.   Google Cloud Platform   * Allows for data analytics and uses AI tools. * GCP is most useful for AI-driven and data-heavy applications. |

# 

# 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) | A service where you rent IT infrastructure, such as servers, storage and networks) online. The user controls most of the system but the provider manages the hardware. | It's used for flexible and scalable project resources without the need of on-site physical servers. For example, a start-up company would need to rent servers depending on the demand, rather than buying costly equipment. |
| PaaS (Platform as a service) | A service that provides a platform for developers to test, build and debug applications (such as Google App Engine). | The developer focuses on coding while the provider manages the rest of the platform, making it a useful resource for developers. A game developer can use the platform to quickly build and release an MMORPG game. |
| SaaS (Software as a service) | SaaS is a cloud computing model where the provider handles everything, including maintenance, updates and infrastructure. The provider keeps ownership while the client only uses it. | It is useful for clients who don't require custom development, for example; Zoom, Google Workspace, Teams  A business uses SaaS for customer relationship management (CRM) software to track sales and leads without needing IT expertise. |

# 

# 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 | Public cloud is where the third-party providers display resources like storage or hosting a pay-as-you-go internet environment. It's good for cost-effecting and flexible scaling without managing a physical infrastructure.  Start-ups typically use AWS, Azure, or Google Cloud for cost and scalability purposes. Dropbox and Slack also rely on public clouds to host their services. |
| Private Cloud | Private cloud is an environment which is dedicated to a single organisation. It offers more control and security than the public cloud, it's good for businesses that manage personal information and must protect it.  Banks (Metro), Healthcare (i.e.- NHS) and Finance Organisations (Responsible Finance) use it to manage said sensitive and personal information. |
| Hybrid Cloud | Hybrid cloud is a mix of public and private cloud environments which allows flexibility with the data and applications being shared.  It is often used for businesses with changing workloads and stock (demands). retailers such as Walmart or Target would find benefits from using hybrid cloud environment, seasonal demand for certain products would require the public cloud, while sales and sensitive data would require the private clous. Government Agencies also benefit from non-sensitive data on a public cloud and sensitive data on a public cloud. |
| Community Cloud | Community Cloud is a shared cloud environment for organisation for purposes of security and compliance. It would be managed internally or by a third party. Healthcare, Government and Education sectors benefit greatly from community cloud environments' shared resource potential.  Universities, for example, would need to collaborate on research and publicly accessible data. |

# 

# Day 2: Task 1

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

|  |  |  |
| --- | --- | --- |
| Area | Description | Example |
| “Unauthorized Access to Computer Material” | Accessing a computer system without expressed permission | A person trying to log into another’s email without consent |
| “Unauthorized Access with Intent to Commit or Facilitate Further Offenses” | Accessing a system to commit another crime, such as fraud or theft. | Hacking into a bank’s database to steal customers’ financial information |
| “Unauthorized Modification of Computer Material” | Deliberately altering, deleting, or introducing malicious software into computer systems. | Spreading a trojan virus to disrupt operations or encrypt files for a ransomware attack |

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 |
| * Criminalizing Denial-of-Service (DoS) and (DDoS) attacks |
| * Increased penalties for cybercrime |
| * Illegal distribution of hacking tools |

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. |
| Personal ID:   * D.O.B * Name * Address * National Insurance Number * Bank Details |
| Work Performance Report:   * Feedback * Productivity * Completed Tasks * Work Ethic and Contribution * Strength and Weaknesses |
| Health Information:   * Medical Records and Accommodations:   + Mental Health   + Physical Health   + Disabilities   + Allergies * Sick Leaves * Health and Safety Assessments |

|  |
| --- |
| Give three more examples of data that an employer can only store if they first get the employee’s permission. |
| - Medical information |
| - Biometric data |
| - Personal social media accounts |

Conduct further research to answer the below questions.

|  |  |
| --- | --- |
| Question | Answer |
| Provide one example of: Copyright infringement | Uploading a media to YouTube or similar sites without the owner's consent |
| Provide one example of: Plagiarism | Copying and pasting paragraphs from a news article and, without credit or citation, presenting it as one's own work |
| What are two consequences of copyright infringement and software piracy? | Legal penalty and financial loss |
| Give three possible consequences for individuals when using pirated software | Loss of revenue for the original creators, impact on the economy and increased risk of data theft and scams to collect personal information via unsafe sites and links |

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** |
| 1 | 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 |
| 1 but not always 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 |
| 1 | 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 |
| 1 | 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 |
| Using/creating hacking tools for penetration testing | \*Not illegal when done lawfully (i.e.- ethical penetration testing) |

# 

# 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 | Inaccessible |

# 

# 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.



|  |  |
| --- | --- |
| Results |  |

# 

# 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.

|  |
| --- |
| “Paws & Whiskers" is a pet shop that wishes to modernise its operations by transitioning from manual data management (including spreadsheets) to a streamlined cloud-based system. With Microsoft Azure, the business aims to enhance its capabilities in data storage and analysis, allowing for informed decision-making.  They'd have to adhere to data laws and regulations, for example:  The General Data Protection Regulation (GDPR) is important for ensuring customer data privacy and security within the EU and UK.  Requirements that "Paws & Whiskers" would need to fulfil is:   * Obtaining explicit consent from customers for collecting and processing their personal data. * Only to collect data necessary for business operations, such as contact information for loyalty programs. * Only to collect data necessary for business operations, such as contact information for loyalty programs. * Enabling customers to access, correct, or delete their data upon request. * Notifying relevant authorities within 72 hours if a data breach occurs.   The Data Protection Act is an act that complements GDPR within the UK, which means that "Paws & Whiskers" need to uphold:   * Accountability and Governance to maintain records of data processing activities. * Establishing a clear purpose for collecting data, such as enhancing customer service or managing inventory. * Defining a timeframe for storing personal data and securely dispose of outdated records.   As they are a company, they’d need to uphold non-data regulations so they   * PCI DSS Compliance is required for handling payment card information, ensuring secure payment processing. * Consumer Rights Act 2015 would need to be upheld to encourage transparency in how customer data is used and shared. * Animal Welfare Data would be important, such as pet-specific data (e.g., breed, health records). It should have been anonymised if analysed for broader trends.   What would be recommended from Azure service is data blob storage, which is ideal for storing unstructured data (raw inventory files and feedback), and allows for scalability and a cost-effective solution to store data, whereas Azure SQL Database is ideal for storing structured data (customer demographics, sales transactions, and product categories) and allows for security.  Azure Machine learning offers advanced analytics, enabling predictions about customer purchasing patterns and inventory requirements. Azure Synapse Analytics, however, allows for big data and data warehousing and provides for powerful tools which analyse sale trends and generate insights which allows "Paws and Whiskers" to act on said insights.  Azure Data Factory automates data integration from multiple sources and allows for real-time updates across inventory and sales systems and logic apps would streamline workflows and would present assistance for restocking or promotional campaigns.  The data types that should be included would be:   * Customer demographics (i.e.- Names, addresses, contact details and other personal information) * Transaction history (i.e.- sales records, payment * Pet inventory (i.e. - species, breeds, stock levels, and pricing.) * Customers' Pet Information (i.e.- breed, age, health records, adoption/sale status. * Product Categories (i.e. - food, tags and identification, treats, leads/collars, dog tags and IDs, accessories, pet toys, pet chews and grooming supplies, bird and reptile supplies, beds and [].)   Using Azure SQL Database will create tables that allow for well-defined relationships with primary keys such as:   * Customers Table: Unique identifiers, names, contact details. * Transactions Table: Transaction IDs, product IDs, and customer IDs. * Products Table: Product IDs, descriptions, categories, and inventory levels.   Relationships that can be used would be one-to-many and many-to-many.  They can also consolidate historical data for analytics. Azure Synapse Analytics allows for a central repository for queries and analytics for large datasets.  Using the star schema with fact tables such as sales data and dimension tables which are for entities like products and customers.  Data storage formats and structure that can be used in Azure are:   * CSV for raw data imports from spreadsheets, it's suitable due to its simplicity and compatibility. * JSON file for structured data which allows for integration with APIs * Parquet is ideal for analytics workloads in Azure Synapse Analytics. It is efficient in storage and query performance. * Encryption at rest and in transit ensures data is encrypted using Azure’s built-in encryption which allows for securing sensitive data. * Role-based access control restricts access to sensitive data. * Azure Key Vault works to safeguard encryption keys and secrets. It manages encryption keys and certificates securely.   Additional considerations include automatically back-ups for data, ensuring proper recovery in case of accidental deletion or corruption.  Azure Site Recovery protects important applications with support during outages. It allows the business to continue by copying systems to back-up Azure regions.  Power BI Integration allows for real-time dashboards to monitor sales and customer trends and is good for customizable reports with interactive dashboards to visualise sale trends, customer preferences and inventory in real time for management decision-making.  Services like Azure Cosmos DB and Azure pay-as-you-go models accommodates growing datasets and handling large volumes of datasets. It will support scalability and integrate additional analytics capabilities. |

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