

**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? | Remote work, collaboration, storage, development and deployment, CRM, entertainment, AI, Healthcare, IoT, payments and banking, education |
| How can it benefit a business? | Data storage, app development and deployment, customer management, transaction handling, collaboration |
| What’s the alternative to cloud computing? | On-Premises physical servers and infrastructure |
| What cloud providers can we use, what are their features and functions? | AWS – largest cloud provider, trustworthy. Best for large scale apps (Netflix, Airbnb, NASA) for high scalability and performance  Microsoft Azure – second largest, popular in enterprise IT and hybrid cloud setups. Strong integration with microsoft products.  Google Cloud – known for big data, AI and Kubernetes. Runs google services like YouTube, Search and Gmail. Good for data heavy apps (spotify, twitter, snapchat) and AI and machine learning workloads  IBM Cloud – strong in AI, blockchain and hybrid cloud for enterprises. Specialises in watson AI, blockchain services and IoT. Best for banks, healthcare, government agencies needing AI, security and compliance  Oracle Cloud – specialises in databases and enterprise apps. Used by finance, healthcare and large corporations relying on oracle databases  Alibaba Cloud – Largest in China and leader in Asia-Pacific. Competes with AWS, Azure and Google Cloud in e-commerce and AI. For businesses operation in China and Asia needing strong cloud presence |

# 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) | Infrastructure as a Service – servers, storage, networking over the cloud. Users manage their own OS, apps and data. | Gives the business full control without having to maintain physical hardware. AWS EC2 or Azure Virtual machines – used by startups and enterprises to run their own apps without buying servers |
| PaaS (Platform as a service) | Platform as a Service – deployment platform with tools, runtime and frameworks allowing devs to build, test and deploy apps without managing infrastructure | Easy to use environment to focus on coding without worrying about infrastructure or system maintenance. Ideal for app development and deployment. Google App Engine, Microsoft Azure App Service, AWS Elastic Beanstalk – used by software devs and startups to create and deploy apps quickly |
| SaaS (Software as a service) | Provides fully managed software apps over the internet. Users access the software via a browser without installation or maintenance | Ready to use apps ideal for email, collaboration tools and CRM software. Google workspace (Gmail, Google Drive), Microsoft 365, Salesforce, Dropbox – used by businesses for communication, document storage and customer management |

# 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 services hosted by third party providers (AWS, Azure, Google CLoud). Resources shared among multiple users. Due to its scalability and cost effectiveness as there is no maintenance burden, it is ideal for startups, e-commerce and non-sensitive workloads. An example is Netflix using AWS for streaming content |
| Private Cloud | Dedicated to one organisation which can be hosted on premises or by a third party provider. It's more secure and customisable and used by organisations that need high security, control and compliance e.g. banks, government and large corporations. An example is the banking sector e.g. HSBC, Barclays for secure financial transactions |
| Hybrid Cloud | A mix of public and private allowing data to move between them. It is flexible and therefore useful for companies that need both security and scalability handling both sensitive and non-sensitive data e.g. healthcare (NHS) storing patient data. |
| Community Cloud | Shared by multiple organisations with similar needs (e.g. government agencies). Resources are divided among members. It’s ideal for organisations in regulated industries (education, healthcare, government) that need shared infrastructure but strict security and compliance. Examples include Universities and Research institutions. Multiple Universities sharing a cloud for research and collaboration while having secure data storage. |

# 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 | Accessing data or systems without permission | Guessing someone else's password or using someone else’s login details to view personal data |
| Unauthorised access with intent to commit further offenses | Gaining access to facilitate crimes like fraud, data theft or system damage | Stealing login credentials to transfer money from bank accounts. |
| Unauthorised modification of computer material | Deleting, altering or corrupting data/systems (e.g. malware, ransomware) | Deploying ransomware to encrypt a business’s files for extortion. A disgruntled employee deleting critical databases before resigning |

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 unauthorised access – the maximum sentence was increased from 6 months to 2 years |
| Criminalising the production and distribution of hacking tools – illegal to create, supply or obtain hacking tools such as malware, keyloggers or password crackers |
| Introducing a New Offense of Denial-of-Service (DoS) Attacks – deliberate DoS is now an offense. Max prison sentence is 10 years |

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 |
| DoB |

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

Conduct further research to answer the below questions.

|  |  |
| --- | --- |
| Question | Answer |
| Provide one example of: Copyright infringement | Reselling of copyright content without permission |
| Provide one example of: Plagiarism | Copying content from papers and other work and putting it into your own academic work without referencing it |
| What are two consequences of copyright infringement and software piracy? | Fines, lawsuits and imprisonment if caught. |
| Give three possible consequences for individuals when using pirated software | Using pirated software comes with security risks (malware, viruses, spyware, etc). It also means missing out on official updates and support making systems more vulnerable. |

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, 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 |
| legal | 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 |
| 1,2 | 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 | Completed but did not take a screenshot |

# 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 | Completed but did not take a screenshot |

# 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 | Lab faulty |

# 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 | 68% - Sorry due to computer and health issues I fell behind and only had time for one attempt. |

# 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.

|  |
| --- |
| 2.  GDPR: If P&W serve European Customers, they must adhere to the rules laid out by the GDPR. If they wish to collect data about their customers, they require their consent and let them know how this information will be used. And only the minimum amount of data, i.e. what is necessary, should be collected. Customers must be able to access and have full control over their data and that it is secure. In case of a breach, P&W must inform their affected customers and the relevant authorities within 72 hours.  DPA: This si the UK’s version of the GDPR. P&W needs to be aware of sensitive Data and similarly to the GDPR, allow customers to access and control their data among other things. In addition, P&W need to document how they handle that data. A data Protection Officer (DPO) should be assigned.  Other Industry Standards and Regulations:  If P&W process credit card transactions, they must comply with the Payment Card Industry Security Standard (PCI DSS). Data needs to be handled securely to reduce risk of fraud.  EPrivacy Directive (Cookie Law): Similarly to GDPR, they need to allow users to agree or opt out of having their data collected using cookies (Since P&W are likely running an online business and will collect data online).  Animal Welfare Act 2006: Since P&W are a pet shop, they need to align with animal welfare laws, including how data regarding the sale of animals is stored and shared. Since owner history, among other data, could intersect with human data protection, this feeds back to GDPR and DPA.  Azure offers a range of compliance certifications (including GDPR, etc) to ensure that the platform meets regulatory requirements for data protection.  3. Azure Service Recommendations  Azure Blob Storage would be the first choice for P&W for their unstructured data needs due to its scalability and cost-effectiveness as well as Analytics Integration. Azure Synapse Analytics (ASA) and Azure Machine Learning (AML) would provide seamless data flow from storage to analysis. AML would allow them to analyse customer behaviour patterns, predict future sales trends, and even create recommendation systems for customers. ASA would enable them to analyse large sales datasets, it can help identify trends, customer segments, and business opportunities. It allows for complex analysis and business intelligence reporting  For customer data, Azure SQL Database would provide a fully managed relational database service with robust performance, security and again, scalability. It’s ideal for customer details, sales records and inventory management.  With Azure Data Factory, they would obtain automated Data Pipelines for real-time updates and to avoid manual data entry. This in turn would further improve data integration with ASA and other elements and improve scheduling and monitoring.  4. Data Types and Data Modelling  Customer Demographics: Name, email, address, phone number, shipping address, DoB  Transaction History: Date of Transaction, total amount, products purchased, payment method, Transaction ID, Customer ID  Pet Inventory: Species, Breed, Age, Health, Pet ID, supplier, price  Product Categories: Name, Description, Category, Quantity, Price, Supplier  Sales and Marketing: Discounts, Promotions, engagement  My recommendation would be to store the Data in tables in a relational database connecting them through primary key and foreign keys. The Customer Table could be the main table with the primary Key being Customer\_ID for example. This would then be linked to the Transactions table via Transaction\_ID (one-tomany), the Products table via Product\_ID (many-to-many), the supplier table with Supplier\_ID (one-to-many) and the Pets table via Pet\_ID (one-to-many).  5. Data Storage Formats and Structures in Azure  CSV is widely used and supported by most applications and it’s ideal for importing external data. It would be great for storing raw sales transaction data, customer details or inventory records for initial import into Azure Blob storage.  JSON is highly flexible and compatible especially with cloud-based services such as Azure. It could be used for customer information, transactional metadata or product details where the data can include various attributes.  Parquet is highly efficient for for big data processing particularly in systems like Azure Data Lake Storage and ASA for sales trends and inventory analysis. It can be easily compressed and split (reducing storage space requirements) and is highly flexible when it comes to adapting the dataset over time (e.g. adding new products).  Data Warehousing with DeltaLake (which is built on top of Parquet) for insuring Data integrity and consistency, where high quality data processing and querying are critical.  Data Security and Encryption in Azure  Azure Storage Service Encryption (SSE): For enhanced Security and control, P&W can use customer-managed keys via Azure Key Vault for data at rest. This ensures automatic encryption without any action from the user.  Transport Layer Security (TLS): For Encryption in Transit, P&W should ensure that all applications, APIs and services that interact with Azure, enforce the use of HTTPS/TLS to protect data during transmission.  Role-Based Access Control (RBAC): Users can only access data based on their roles and permissions. Use Azure Active Directory to manage centralised identity and access management. Use the principle of least privilege by granting only the minimum permissions necessary for users to perform their roles.  5. Additional Considerations  Backup and Disaster Recovery:  Azure Backup allows for automated cloud-based backup of data including Azure virtual machines, databases and storage accounts. Data is backed up locally or geo-redundantly both at rest and in transit which ensures high availability and durability (including encryption).  Azure Site Recovery (ASR): Replicates workloads and virtual machines (e.g. the inventory system or sales application of P&W) to a secondary region to allow P&W to continue operations seamlessly in case of a regional outage.  Data Visualisation: With Power Bi P&W will be able to visualise their complex data into intuitive actionable charts and reports. It’ll allow real-time insights into their business performance including sales trends, inventory levels and customer behaviours. Fortunately, it integrates seamless with Azure allowing multiple interacting users to pull data from multiple sources into a single view.  Azure is scalable in every aspect, from Data Storage and processing to Machine Learning and AI. If P&W start out with Azure Blob Storage or Data Lae, they will have all the necessary flexibility when it is time to scale up later on. It also allows them to save money in the short term by avoiding on-premises installation of their own hardware. |

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