

**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? | * **Access resources on demand:** Use computing power, storage, and software only when we need it, like electricity from a utility. * **Scale easily:** Quickly increase or decrease our computing resources as our needs change, without huge upfront investments. * **Reduce costs:** Avoid the expense of buying and maintaining physical hardware and infrastructure. * **Access data and applications from anywhere:** Work and collaborate from any device with an internet connection. * **Improve reliability and availability:** Benefit from robust infrastructure with built-in redundancy and disaster recovery. * **Foster innovation:** Easily access advanced technologies like AI, machine learning, and big data analytics. * **Enable collaboration:** Share data and work together seamlessly across different locations. * **Increase agility and speed:** Deploy applications and services faster and respond more quickly to changing market demands.   Essentially, cloud computing provides flexible, scalable, cost-effective, and accessible IT resources that empower individuals, businesses, and organisations to do more with technology, regardless of their size or location. |
| How can it benefit a business? | Cloud computing empowers businesses to be more agile, efficient, and competitive by offering several key benefits: cost reduction through lower upfront investment and pay-as-you-go models; scalability and flexibility to easily adjust resources and deploy quickly; increased accessibility and collaboration enabling remote work and teamwork; improved reliability and business continuity via high availability and disaster recovery; reduced IT burden allowing focus on core business; access to advanced technologies fostering innovation; enhanced security through provider investments; and faster deployment accelerating time to market. Ultimately, it provides adaptable, cost-efficient, innovative, and resilient IT infrastructure. |
| What’s the alternative to cloud computing? | Owning and managing your own physical IT infrastructure: This includes servers, networking equipment, storage devices, and data centres located within your organisation's physical premises. |
| What cloud providers can we use, what are their features and functions? | **Major Cloud Providers:**   * **Amazon Web Services (AWS):** The largest and most mature cloud provider. * **Microsoft Azure:** The second-largest, deeply integrated with Microsoft's existing ecosystem. * **Google Cloud Platform (GCP):** Known for its strengths in data analytics, machine learning, and containerisation   Cloud providers offer a broad range of core services: **Compute** (virtual machines, containers, serverless), **Storage** (object, block, file, data lakes), **Networking** (virtual private clouds, load balancing, DNS, CDNs), **Databases** (managed relational and NoSQL, data warehousing), **Analytics** (data processing, big data analytics, BI), **Machine Learning & AI** (ML platforms, pre-trained services), **Security** (IAM, encryption, firewalls, compliance), and **Management & Governance** (monitoring, cost management, automation). These services provide the fundamental building blocks for running applications and managing data in the cloud. |

# 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) | It is a cloud computing service model that provides users with access to fundamental computing resources – such as computing (virtual machines), storage, and networking – over the internet on demand and on a pay-as-you-go basis | 1. Startups and Small Businesses: Instead of buying servers, storage, and networking equipment, a startup can quickly provision virtual machines (VMs), storage buckets, and virtual networks on an IaaS platform. Web Hosting and Application Deployment: Businesses can deploy their applications on VMs or container services provided by IaaS. |
| PaaS (Platform as a service) | is a cloud computing model that provides developers with a complete environment to build, run, and manage applications without the complexity of managing the underlying infrastructure. | when you want to:   * Accelerate application development and deployment. * Reduce the operational burden of managing infrastructure. * Easily scale your applications based on demand. * Leverage pre-configured services and tools. * Focus your team's efforts on writing code and business logic.   Examples: 1. Developing and Deploying Web Applications 2. Building Mobile Backend Applications 3. Creating and Managing APIs. 4. Running Databases and Data Stores |
| SaaS (Software as a service) | It is a software distribution model where a third-party provider hosts applications and makes them available to customers over the Internet. | 1. Email: Gmail and Outlook.com, are all SaaS email services. 2. Document Creation and Collaboration: Google Docs, Microsoft Word Online. 3. Accounting and Finance: Businesses use SaaS accounting software like Xero and QuickBooks Online. |

# 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 they could be implemented (i.e. what type of organisation).

|  |  |
| --- | --- |
| Public Cloud | It is a type of cloud computing where a third-party provider (like AWS, Azure, or GCP) owns and operates a massive infrastructure of servers and data centres. These resources are then made available to multiple customers over the Internet on a pay-as-you-go basis. You share the underlying hardware and infrastructure with other organisations, but your data and applications are logically isolated and secure.  It offers a flexible, scalable, and cost-effective IT infrastructure that is appropriate for a wide variety of use cases, as exemplified by companies like Netflix that require massive scale and global reach. A Real-World Example:  Netflix. |
| Private Cloud | It is a cloud computing environment where the infrastructure and services are dedicated to a single organisation,  offering greater control and security compared to a public cloud. Real-World Example:  A large multinational financial institution. |
| Hybrid Cloud | It is a computing environment that combines on-premises infrastructure (your own data centre or private cloud) with one or more public clouds (like AWS, Azure, or GCP). These environments are connected, allowing data and applications to be shared and moved between them. A Real-World Example:  Consider a large financial institution based in London, England: |
| Community Cloud | It is a cloud infrastructure that is provisioned for exclusive use by a specific community of users from organizations that have shared concerns (e.g., security requirements, compliance policies, mission). It can be owned, managed, and operated by one or more of the organisations in the community, a third party, or a combination of them, and it may exist on or off-premises. A Real-World Example:  Consider a group of healthcare providers (hospitals, clinics, research institutions) within a specific region of the UK, focused on sharing anonymised patient data for research on a particular disease. |

# 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 | Gaining access to any program or data held in a computer without authorisation. The key element is the lack of permission to access the specific material. It doesn't matter if the person intends to commit a further crime or cause damage; simply gaining unauthorised entry is the offence. | A student guesses a classmate's password and logs into their university account to read their emails. |
| Unauthorised Access with Intent to Commit or Facilitate the Commission of Further Offences | Gaining unauthorised access to computer material with the intention of committing or facilitating the commission of a more serious crime. | A hacker gains unauthorised access to a bank's customer database with the intention of stealing personal and financial information to commit fraud. |
| Unauthorised Acts with Intent to Impair the Operation of a Computer | It involves performing unauthorised acts that are intended to impair the operation of a computer, prevent or hinder access to data stored on a computer, or impair the reliability of such data. This often involves the deliberate introduction of viruses, malware, or denial-of-service attacks. | A cybercriminal launches a Distributed Denial of Service (DDoS) attack against a company's website, flooding it with traffic and making it unavailable to legitimate users. |

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. New Offence of Making, Supplying or Obtaining Articles for Use in Computer Misuse (Section 3ZA): This created a specific offence targeting the creation, provision, or acquisition of tools (software, hardware, or even information) intended to be used for committing Computer Misuse Act offences. This allows law enforcement to go after individuals who facilitate cybercrime, even if they haven't directly committed an unauthorised access or impairment themselves. For example, someone creating and selling sophisticated hacking tools could be prosecuted under this section. |
| 1. Increased Penalties for Existing Offences: The 2006 Act increased the maximum penalties for the existing offences under the CMA 1990, particularly for the more serious offences under Section 2 (unauthorised access with intent to commit further offences) and Section 3 (unauthorised acts with intent to impair operation). This reflects the growing recognition of the potential harm caused by cybercrime and provides a stronger deterrent. For instance, the maximum prison sentence for Section 3 offences was significantly raised. |
| 1. Extraterritorial Jurisdiction for Certain Offences: The 2006 Act extended the jurisdiction of the Computer Misuse Act to cover certain offences committed outside the UK, provided there is a sufficient connection to the UK. This is crucial in tackling cybercrime, which often transcends national borders. For example, if a cybercriminal located abroad targets a computer system within the UK, they can now be prosecuted under the CMA in the UK, provided certain conditions are met regarding their presence or the significant impact in the UK. |

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. |
| Contact Information: This includes name, address, phone number, and email address. |
| Employment History within the Company |
| National Insurance Number |

|  |
| --- |
| Give three more examples of data that an employer can only store if they first get the employee’s permission. |
| * Race and ethnicity * Religion |
| * Political membership or opinions |
| * biometrics, for example, if your fingerprints are used for identification |

Conduct further research to answer the below questions.

|  |  |
| --- | --- |
| Question | Answer |
| Provide one example of: Copyright infringement | A small company in London uses a well-known photograph that they found online on their company website and in their promotional brochures without obtaining permission or a license from the photographer who owns the copyright to the image. This unauthorised reproduction and distribution of the copyrighted photograph constitutes copyright infringement under UK law. |
| Provide one example of: Plagiarism | A university student submits an essay that includes several paragraphs directly copied word-for-word from a published academic journal article without using quotation marks and without properly citing the original source in their bibliography. This act of presenting someone else's work as their own constitutes plagiarism. |
| What are two consequences of copyright infringement and software piracy? | 1. Legal Action and Financial Penalties: Copyright holders in the UK can pursue legal action against infringers. This can result in court orders to cease the infringing activity and financial penalties, including damages to compensate the copyright holder for their losses and potentially additional "flagrant infringement" damages if the infringement is deemed particularly egregious.  2. Security Risks and Malware: Downloading and using pirated software often exposes users to significant security risks. These illegal copies frequently contain malware, viruses, Trojans, and other malicious software that can compromise computer systems, steal personal data (including financial information), and disrupt operations |
| Give three possible consequences for individuals when using pirated software | **1. Legal Prosecution and Fines:** Software piracy is illegal under UK copyright law. Individuals caught using or distributing pirated software can face legal action from copyright holders or be prosecuted by law enforcement. This can result in significant fines, which can be substantial depending on the scale of the infringement and the value of the software.  **2. Exposure to Malware and Security Risks:** Pirated software is often bundled with or contains malicious software (malware) such as viruses, Trojans, spyware, and ransomware. Downloading and installing pirated software can compromise your computer system, leading to data theft (including financial information and personal details), system instability, and even identity theft. These risks can have serious financial and personal repercussions.  **3. Lack of Support and Updates:** Users of pirated software do not receive legitimate licenses, meaning they are not entitled to technical support from the software vendor. Furthermore, they will not receive crucial software updates and security patches. This leaves their systems vulnerable to newly discovered security flaws and may result in compatibility issues with other software or operating system updates, ultimately hindering the functionality and security of their devices. |

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 |
| 2 | 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 |
| 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 |

# 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 | The lab is 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 |  |

# 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.

|  |
| --- |
| By utilizing Azure's extensive range of data storage, analysis, integration, and reporting services, "Paws & Whiskers" can establish a robust and scalable data infrastructure. This infrastructure will empower the company to streamline its data management processes, perform in-depth analysis of key business metrics, generate insightful reports, and make informed, data-driven decisions that will fuel future growth and success in the competitive pet retail market. The implementation of Azure will move "Paws & Whiskers" from a reactive, intuition-based approach to a proactive, data-informed operational model.  **Data Laws and Regulations**  Handling customer data requires strict adherence to relevant legal frameworks to protect individual privacy and ensure ethical data management practices. For "Paws & Whiskers," the following data laws and regulations are paramount:  **· GDPR Compliance:**  The General Data Protection Regulation (GDPR) is a cornerstone of data protection law in the European Economic Area (EEA), including the UK (even post-Brexit, as it has been incorporated into UK law). Compliance with GDPR is not merely a legal obligation but a fundamental aspect of building customer trust and ensuring responsible data handling. For "Paws & Whiskers," GDPR implications are significant in how they collect, store, process, and protect the personal data of their customers.  Key GDPR principles that "Paws & Whiskers" must adhere to include:   * **Lawfulness, Fairness, and Transparency:** Personal data must be processed lawfully, fairly, and in a transparent manner in relation to the data subject. This means providing clear and concise information to customers about how their data will be used. * **Purpose Limitation:** Personal data must be collected for specified, explicit, and legitimate purposes and not further processed in a manner that is incompatible with those purposes. "Paws & Whiskers" must clearly define the reasons for collecting customer data (e.g., processing sales, marketing communications with consent, loyalty programs). * **Data Minimisation:** Personal data collected should be adequate, relevant, and limited to what is necessary in relation to the purposes for which it is processed. "Paws & Whiskers" should avoid collecting excessive or unnecessary customer information. * **Accuracy:** Personal data must be accurate and, where necessary, kept up to date. Steps should be taken to ensure the accuracy of customer records. * **Storage Limitation:** Personal data should be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed. "Paws & Whiskers" needs to establish clear data retention policies. * **Integrity and Confidentiality:** Personal data must be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction, or damage, using appropriate technical or organisational measures. Encryption and access controls in Azure will be crucial here.   Furthermore, GDPR grants several rights to data subjects (customers), including:   * **The right to be informed:** Customers have the right to know what personal data is being collected and how it will be used. * **The right of access:** Customers can request access to their personal data held by "Paws & Whiskers." * **The right to rectification:** Customers can request that inaccurate personal data be corrected. * **The right to erasure ('right to be forgotten'):** Customers can request the deletion of their personal data under certain circumstances. * **The right to restrict processing:** Customers can request the restriction of processing their personal data under certain circumstances. * **The right to data portability:** Customers can request to receive their personal data in a structured, commonly used, and machine-readable format and have the right to transmit that data to another controller. * **The right to object:** Customers have the right to object to the processing of their personal data under certain circumstances, including for direct marketing purposes.   "Paws & Whiskers" must implement appropriate technical and organisational measures within their Azure environment to ensure these principles and rights are upheld. This includes implementing robust security measures, establishing clear data processing procedures, and having mechanisms in place to respond to data subject requests.  **· Data Protection Act (DPA) 2018:**  For "Paws & Whiskers," the DPA 2018 reinforces the principles and obligations outlined in the GDPR. It clarifies aspects such as the role of the Information Commissioner's Office (ICO) as the UK's independent supervisory authority for data protection and outlines the potential penalties for non-compliance.  Specifically, the DPA 2018 may affect how "Paws & Whiskers" collects and stores data by:   * **Defining Processing Conditions:** The Act provides further detail on the conditions for lawful processing of personal data. * **Specifying Exemptions:** While GDPR provides for some exemptions, the DPA 2018 outlines specific exemptions relevant to the UK, though these are unlikely to significantly impact standard customer and sales data processing for a pet shop. * **Outlining Enforcement Powers:** The DPA 2018 details the ICO's powers to investigate and impose fines for breaches of data protection law. "Paws & Whiskers" needs to be aware of these potential consequences of non-compliance. * **Addressing Special Category Data:** If "Paws & Whiskers" were to collect any special category data (e.g., health information about pets linked to a customer), both GDPR and the DPA 2018 would impose stricter requirements for processing, typically requiring explicit consent or another specific legal basis.   Compliance with the DPA 2018 necessitates that "Paws & Whiskers" not only adheres to the GDPR principles but also understands any specific interpretations and requirements under UK law. Implementing a data protection framework within Azure that aligns with both GDPR and the DPA 2018 is crucial.  **· Other Industry Standards:**  "Paws & Whiskers" needs to consider regulations that might apply based on the nature of the data they handle, particularly if they involve sensitive or payment information:   * **Payment Card Industry Data Security Standard (PCI DSS):** If "Paws & Whiskers" directly processes, stores, or transmits credit or debit card information, they must comply with PCI DSS. This standard outlines a set of security requirements designed to protect cardholder data. Implementing secure payment processing within their systems and ensuring that any cardholder data stored (if absolutely necessary and minimized) is done so in a PCI DSS compliant manner within Azure is vital. * **Consumer Protection Laws:** General consumer protection laws in the UK also dictate how customer information should be handled fairly and transparently. Misleading or mishandling customer data could lead to breaches of these laws. * **Marketing Regulations (e.g., Privacy and Electronic Communications Regulations - PECR):** If "Paws & Whiskers" engages in electronic marketing (email, SMS), they must comply with PECR, which governs the use of cookies and electronic communications. Obtaining explicit consent for marketing, communications and providing clear opt-out mechanisms are key requirements.   **3. Azure Service Recommendations**  To effectively address "Paws & Whiskers'" data analysis needs, the following Microsoft Azure services are recommended:  **· Data Storage:**   * **Azure Blob Storage:** This service is ideal for storing large volumes of unstructured data, such as digital images of pets for identification, potentially audio or video logs of customer interactions (with appropriate consent and privacy considerations), and archived raw data files (e.g., CSV exports from point-of-sale systems).   + **Benefits:** Highly scalable and cost-effective for large datasets. Offers different access tiers (Hot, Cool, Archive) to optimise costs based on data access frequency. Provides robust security features, including encryption at rest and in transit. Integrates seamlessly with other Azure services for data processing and analysis. Its object storage nature is well-suited for storing diverse file types without rigid schema requirements. * **Azure SQL Database:** This fully managed relational database service is recommended for storing structured data such as customer details (name, contact information, purchase history), product inventory (SKU, name, price, quantity), and sales transaction records (order ID, date, items purchased, price, customer ID).   + **Benefits:** Provides a familiar relational database model for structured data, enabling efficient querying and data manipulation using SQL. Offers built-in security features, including encryption at rest and in transit, and advanced threat protection. Ensures data integrity and consistency through transactional support. Highly scalable to accommodate growing data volumes and performance demands. Integrates well with Azure Data Factory for data ingestion and transformation and with Power BI for reporting and visualisation.   **· Data Analysis Tools:**   * **Azure Synapse Analytics:** This is a comprehensive analytics service that brings together data warehousing and big data analytics. It is highly suitable for analyzing sales trends, inventory performance, and customer purchase patterns across large datasets.   + **Benefits:** It offers a scalable data warehouse with MPP (Massively Parallel Processing) architecture for fast query performance on large volumes of historical sales data. Includes built-in data integration capabilities (Azure Data Factory pipelines). Integrates with Power BI for advanced reporting and dashboarding. Supports both SQL and Spark for diverse analytical workloads. Its ability to handle both structured and semi-structured data makes it versatile for analyzing various aspects of "Paws & Whiskers'" data. * **Azure Machine Learning:** This cloud-based environment can be leveraged for more advanced customer behaviour analysis. By training machine learning models on historical purchase data, "Paws & Whiskers" can gain insights into customer segmentation (identifying different groups of customers with similar purchasing habits), predict future purchase patterns, and potentially personalize marketing efforts (with appropriate consent).   + **Benefits:** Provides a collaborative environment for data scientists and machine learning engineers to build, train, and deploy models. Offers a range of pre-built algorithms and tools for various machine learning tasks. Integrates with Azure Data Lake Storage for accessing large datasets. Supports automated machine learning (AutoML) to simplify model development. Can be used to develop predictive models for inventory forecasting, helping to optimize stock levels and reduce waste.   **· Data Integration and Automation:**   * **Azure Data Factory:** This is a fully managed, serverless data integration service that can automate the collection, transformation, and loading of data from various sources into the chosen Azure storage and analytics services.   + **Benefits:** Enables the creation of data pipelines to automate the process of extracting data from existing systems (e.g., point-of-sale system, spreadsheets), transforming it into the required formats, and loading it into Azure Blob Storage and Azure SQL Database. Improves efficiency by eliminating manual data handling. Provides monitoring and alerting capabilities for data pipeline execution. Supports a wide range of connectors to various data sources, both on-premises and in the cloud. Can be scheduled to run data integration processes at regular intervals, ensuring that the Azure data platform is always up-to-date.   **4. Data Types and Data Modelling**  "Paws & Whiskers" will need to work with several key data types:  **· Data Categories:**   * **Customer Demographics:** Information about customers, such as name, address, contact details (phone, email), and potentially opt-in preferences for marketing. * **Transaction History:** Records of all sales transactions, including order ID, date and time of purchase, items purchased (linked to product inventory), quantities, prices, payment method, and associated customer ID. * **Pet Inventory:** Details about the products in stock, including SKU, product name, category (e.g., food, toys, accessories), brand, supplier, cost price, selling price, quantity in stock, and potentially information about the types of pets the product is suitable for. * **Product Categories:** A classification system for the pet shop's offerings, allowing for categorization and analysis of sales by product type. * **Supplier Information:** Details about the companies that supply products to "Paws & Whiskers," including name, contact information, and product lines. * **Marketing Interactions (Optional):** Records of customer interactions with marketing campaigns, such as email opens, clicks, and website visits (if applicable). * **Loyalty Program Data (If Implemented):** Information related to a customer loyalty program, such as points earned, rewards redeemed, and membership status.   **· Data Modelling Approach:**  A **relational data model** implemented within Azure SQL Database is the most suitable approach for structuring the core transactional and inventory data. This model excels at handling structured data with well-defined relationships and ensures data integrity through normalization.  The relational model would involve creating several tables, each representing an entity:   * **Customers Table:** Columns for CustomerID (Primary Key), Name, Address, Phone, Email, RegistrationDate, MarketingOptIn. * **Products Table:** Columns for ProductID (Primary Key), SKU (Unique Key), Name, CategoryID (Foreign Key), Brand, SupplierID (Foreign Key), CostPrice, SellingPrice. * **ProductCategories Table:** Columns for CategoryID (Primary Key), CategoryName. * **Suppliers Table:** Columns for SupplierID (Primary Key), SupplierName, ContactPerson, Phone, Email. * **Orders Table:** Columns for OrderID (Primary Key), CustomerID (Foreign Key), OrderDate, TotalAmount. * **OrderItems Table:** Columns for OrderItemID (Primary Key), OrderID (Foreign Key), ProductID (Foreign Key), Quantity, UnitPrice.   **Relationships:**   * A one-to-many relationship between Customers and Orders (one customer can place multiple orders). * A one-to-many relationship between Orders and OrderItems (one order can contain multiple order items). * A many-to-one relationship between OrderItems and Products (multiple order items can refer to the same product). * A one-to-many relationship between Products and ProductCategories (one product belongs to one category). * A one-to-many relationship between Products and Suppliers (one product is supplied by one supplier).   **Primary Keys:** Each table will have a primary key to uniquely identify each record. Foreign keys will be used to establish and enforce relationships between tables, ensuring data consistency.  For the less structured data, such as digital images or archived raw files, Azure Blob Storage will serve as the primary storage location. Metadata can be associated with these blobs to provide context and facilitate searching.  While a full-fledged data warehouse using Azure Synapse Analytics is recommended for advanced analytics, the initial data modelling within Azure SQL Database will lay a strong foundation for future expansion and integration with the data warehouse. The relational model in SQL Database will be optimized for transactional processing and efficient querying of current data, while Synapse Analytics will be optimized for analytical queries across large historical datasets.  **5. Data Storage Formats and Structures in Azure**  The choice of data formats and structures within Azure will be crucial for efficiency, cost-effectiveness, and compatibility with different services:  **· Data Formats:**   * **CSV (Comma Separated Values):** Recommended for initial raw data imports from existing spreadsheets or legacy systems into Azure Blob Storage or Azure Data Factory. CSV is a simple and widely supported format for transferring tabular data.   + **Why Suitable:** Easy to generate and parse by various tools and services. Human-readable for initial data inspection. * **JSON (JavaScript Object Notation):** Suitable for storing semi-structured data, such as potentially product descriptions with varying attributes or customer preferences if they are not strictly relational. Can also be used as an intermediary format for data transformation within Azure Data Factory.   + **Why Suitable:** Flexible and can represent complex data structures. Widely supported by web applications and APIs. * **Parquet:** Highly recommended for storing large analytical datasets within Azure Blob Storage that will be queried by Azure Synapse Analytics or Azure Machine Learning.   + **Why Suitable:** Columnar format allows for efficient retrieval of specific columns needed for analysis, reducing I/O and processing time. Supports data compression, reducing storage costs. Integrates seamlessly with Azure Synapse Analytics and Spark.   **· Data Security and Encryption:**  Securing data at rest and in transit is paramount for GDPR and DPA 2018 compliance. Azure provides several built-in features to achieve this:   * **Encryption at Rest:**   + **Azure Storage Service Encryption (SSE):** Azure Storage automatically encrypts data at rest when it is written to Blob Storage, Queue Storage, Table Storage, and Azure Files. Microsoft manages the encryption keys by default, but "Paws & Whiskers" can also choose to manage their own keys using Azure Key Vault for enhanced control. This should be enabled for all storage services used.   + **Transparent Data Encryption (TDE) for Azure SQL Database:** TDE automatically encrypts the data at rest in Azure SQL Database, including the database files, log files, and backups. This should be enabled for the "Paws & Whiskers" database. * **Encryption in Transit:**   + **Secure Transfer Required:** For Azure Storage accounts, the "Secure transfer required" setting should be enabled. This ensures that all requests to the storage account are made over secure protocols (HTTPS).   + **TLS/SSL for Azure SQL Database:** Connections to Azure SQL Database should always use TLS (Transport Layer Security) to encrypt data in transit between the client application and the database. * **Access Controls:**   + **Role-Based Access Control (RBAC):** Azure RBAC should be used to manage access to Azure resources, including the storage account and SQL Database. Assigning least privilege to users and applications ensures that they only have the specific permissions they need to perform their tasks. For example, data analysts should have read access to the database but may not need write access to production tables. \* **Azure Active Directory (Azure AD) Integration:** Integrate Azure Storage and Azure SQL Database with Azure AD for centralized identity management. This allows using Azure AD credentials for authentication and authorization, enhancing security and simplifying management. \* **Network Security:** Implement Azure Firewall or Network Security Groups (NSGs) to control network traffic to and from the Azure resources, restricting access to authorized IP addresses or virtual networks. \* **Azure Private Link:** Consider using Azure Private Link to establish private connectivity to Azure Storage and Azure SQL Database from within the "Paws & Whiskers" virtual network, eliminating exposure over the public internet. \* **Shared Access Signatures (SAS) with Caution:** If SAS tokens are used to grant temporary access to specific blobs or files, ensure they are generated with the principle of least privilege and have appropriate expiry times. User Delegation SAS tokens, secured with Azure AD credentials, are generally preferred over Service SAS tokens secured with storage account keys.   **Data Masking and Tokenization (If Necessary):** For sensitive personal data that may be used in non-production environments or for specific analytical purposes where the raw data is not required, consider implementing Azure SQL Database Dynamic Data Masking or other tokenization techniques to pseudonymize or anonymize the data. \* **Regular Security Audits:** Conduct regular security audits and vulnerability assessments of the Azure environment to identify and address any potential security weaknesses.  By implementing these data security and encryption measures, "Paws & Whiskers" can significantly enhance the protection of their customer and business data, ensuring compliance with GDPR, the DPA 2018, and other relevant regulations.  **6. Additional Considerations**  To further enhance data handling and efficiency within the Azure environment, "Paws & Whiskers" should consider the following:  **· Backup and Disaster Recovery:**  A robust backup and disaster recovery (DR) plan is crucial to safeguard against data loss due to accidental deletion, hardware failures, or other unforeseen events. Azure offers several services that can be leveraged:   * **Azure Backup:** This service provides a cost-effective and secure solution for backing up data from various Azure resources, including Azure SQL Database and Azure Blob Storage.   + **Azure SQL Database Backups:** Azure SQL Database has built-in automated backups that are retained for a configurable period. "Paws & Whiskers" should review and configure the retention policy based on their recovery point objectives (RPOs). Long-term retention can be configured to store backups in Azure Blob Storage for compliance or archival purposes.   + **Azure Blob Storage Backups:** Azure Backup can be configured to create backups of Azure Blob Storage accounts. This allows for point-in-time recovery of blobs and containers. Consider implementing a backup schedule that aligns with the data change frequency and retention requirements. * **Azure Site Recovery:** This service provides disaster recovery capabilities by replicating workloads running on Azure VMs to a secondary Azure region. While the initial focus is on data storage and analysis, if "Paws & Whiskers" were to host critical applications on Azure VMs in the future, Site Recovery would be essential for ensuring business continuity in the event of a regional outage. * **Regular Backup Testing:** It is crucial to regularly test the backup and recovery processes to ensure their effectiveness and to familiarize the team with the recovery procedures. This will minimize downtime and data loss in a real disaster scenario.   **· Data Visualisation:**  "Paws & Whiskers" can leverage Power BI to create interactive dashboards and reports that provide management with real-time insights into key business metrics:   * **Sales Trends:** Visualize daily, weekly, and monthly sales performance, track top-selling products, identify seasonal trends, and analyse sales by product category or customer demographics. * **Customer Trends:** Analyse customer acquisition rates, churn rates, average order value, customer lifetime value, and identify key customer segments. * **Inventory Management:** Create dashboards to monitor stock levels, identify slow-moving or fast-selling items, track inventory turnover, and optimize reordering processes. * **Marketing Performance:** If marketing data is integrated, Power BI can be used to analyse the effectiveness of different marketing campaigns. By providing easily digestible visualisations, Power BI can empower "Paws & Whiskers" management to quickly understand business performance, identify areas for improvement, and make data-driven decisions without needing to delve into raw data. Power BI can connect directly to Azure SQL Database and Azure Synapse Analytics, allowing for real-time or near real-time data updates in the dashboards.   **· Future Scalability:** One of the key advantages of adopting Microsoft Azure is its inherent scalability. As "Paws & Whiskers" grows and their data volumes increase, the chosen Azure services can easily scale to accommodate these changes:   * **Azure Blob Storage:** Designed for massive scalability, Blob Storage can handle petabytes of data without requiring significant upfront planning. Storage costs are typically pay-as-you-go, so costs scale with usage. * **Azure SQL Database:** It offers various service tiers with different compute and storage capacities. "Paws & Whiskers" can start with a suitable tier and easily scale up or down as their database size and performance requirements evolve. Elastic Pools can further optimize costs for multiple databases. * **Azure Synapse Analytics:** Built for big data analytics, Synapse Analytics can scale its compute resources (Data Warehouse Units for the SQL pool) up or down on demand, allowing "Paws & Whiskers" to handle increasingly complex analyses and larger datasets without significant infrastructure changes. Serverless SQL pool options also provide cost-effective querying for ad-hoc analysis. * **Azure Data Factory:** As data integration needs become more complex and involve more data sources, Azure Data Factory can handle increased pipeline complexity and data volumes through its serverless architecture.   The scalability of these Azure services ensures that "Paws & Whiskers"' data infrastructure can grow seamlessly alongside their business, providing a future-proof solution for their data analysis needs. This eliminates the need for significant upfront investments in on-premises infrastructure and provides the flexibility to adapt to changing business demands.  "Paws & Whiskers" can build a comprehensive and efficient data platform in Azure that not only meets their current needs but also positions them for continued growth and data-driven success in the future.  **References**:  www.ampersand-world.com  www.ulster.ac.uk  www.lawsociety.org.uk  www.parliament.scot  www.sogrape.com Microsoft Azure Documentation cloud.google.com/discover |

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

​