

**Data Technician**

|  |
| --- |
|  |

|  |
| --- |
| Name: |
| Course Date: |
|  |

**Table of contents**

[Day 1: Task 1 3](#_Toc1514505634)

[Day 1: Task 2 3](#_Toc2125927931)

[Day 1: Task 3 4](#_Toc1965065538)

[Day 2: Task 1 5](#_Toc2141651249)

[Day 3: Task 1 9](#_Toc1229522580)

[Day 3: Task 2 10](#_Toc508692763)

[Day 3: Task 3 11](#_Toc1233463339)

[Day 4: Task 1 12](#_Toc1556426903)

[Day 4: Task 2 13](#_Toc214156810)

[1. Scenario Background 13](#_Toc248266112)

[2. Data Laws and Regulations 14](#_Toc1387014804)

[3. Azure Service Recommendations 14](#_Toc1952198484)

[4. Data Types and Data Modelling 14](#_Toc1792190821)

[5. Data Storage Formats and Structures in Azure 14](#_Toc385598743)

[6. Additional Considerations 15](#_Toc561077662)

[Submission Guidelines: 15](#_Toc1001523541)

[Course Notes 17](#_Toc977988415)

[Additional Information 17](#_Toc1081373283)

# Day 1: Task 1

Please research and complete the below questions relating to key concepts of cloud.

Be prepared to discuss the below in the group following this task.

|  |  |
| --- | --- |
| What can cloud computing do for us in the real-world? | Cloud computing has a wide range of real-world applications that impact everyday life, businesses, and technology infrastructure. It simply means using the internet to access computing power, storage, and services — instead of owning physical hardware or servers yourself. Here’s how that plays out in everyday life and across industries:  Scalability:   * + Companies can quickly increase or decrease computing resources (like storage or servers) based on demand — for example, e-commerce sites scaling up during holiday sales.   Cost Savings:   * + Instead of buying expensive hardware, businesses pay only for what they use (pay-as-you-go model).   Remote Collaboration:   * + Cloud tools like Google Workspace, Microsoft 365, and Slack allow employees to work together from anywhere.   Data Backup & Recovery:   * + Automatically stores backups in the cloud, protecting against data loss from hardware failure or disasters.   Faster Development:   * + Developers can deploy apps quickly using cloud platforms like AWS, Azure, or Google Cloud. |
| How can it benefit a business? | 1. Cost Savings   * No upfront hardware costs: Businesses don’t need to buy expensive servers or networking gear. * Pay-as-you-go model: You only pay for the computing power, storage, and services you use. * Lower maintenance costs: Cloud providers handle server maintenance, updates, and security patches.   2. Scalability & Flexibility   * Businesses can scale up or down instantly depending on demand. * Perfect for industries with seasonal spikes — like e-commerce or travel. * Cloud platforms can deploy new servers or storage within minutes, instead of days or weeks.   3. Remote Work & Collaboration   * Cloud-based tools (e.g., Microsoft 365, Google Workspace, Slack) allow employees to work from anywhere. * Documents, data, and applications are accessible 24/7 from any device with internet access. * Improves teamwork and communication, especially for distributed teams.   4. Enhanced Security & Data Protection   * Major cloud providers invest heavily in cybersecurity, encryption, and compliance. * Data is often stored in multiple data centres, protecting against local failures. * Built-in tools help manage identity access and monitor threats.   5. Business Continuity & Disaster Recovery   * Cloud backups make it easier to recover from system failures, cyberattacks, or natural disasters. * Automatic replication ensures data isn’t lost even if one data centre fails. |
| What’s the alternative to cloud computing? | The Main Alternative is On-Premises (or On-Prem) Computing  Instead of renting servers or storage from cloud providers, a company owns and manages its own physical hardware — usually located in an on-site data centre or office.  Advantages of On-Premises   1. Full Control 2. Data Security & Privacy 3. Customization 4. No Internet Dependency   Disadvantages of On-Premises   1. High Upfront Costs 2. Maintenance Burden 3. Limited Scalability 4. Less Flexibility   Other Alternatives are:  Colocatio:   * A company rents space in a third-party data centre but still owns and manages its hardware. * Often used as a middle ground between full on-prem and cloud.   Private Cloud:   * Uses cloud-like technologies (e.g., virtualization) but runs on hardware owned by one organization. * Provides scalability and control, but still expensive to manage. |
| What cloud providers can we use, what are their features and functions? | |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | | **Feature / Function** | **AWS (Amazon Web Services)** | **Azure (Microsoft Azure)** | **GCP (Google Cloud Platform)** | | --- | --- | --- | --- | | **Launched** | 2006 | 2010 | 2011 | | **Market Position** | Largest and most mature cloud provider | Strong enterprise integration | Strong in data analytics and AI | | | |  |  |  |  | | --- | --- | --- | --- | | **Compute (Servers / VMs)** | EC2 (Elastic Compute Cloud) for scalable virtual machines | Azure Virtual Machines for Windows/Linux workloads | Compute Engine for custom VMs with per-second billing | | **Serverless Computing** | AWS Lambda (event-driven execution) | Azure Functions | Cloud Functions | | **Containers / Kubernetes** | Elastic Kubernetes Service (EKS), ECS | Azure Kubernetes Service (AKS) | Google Kubernetes Engine (GKE) — originally developed by Google | | **Scalability** | Auto Scaling and Elastic Load Balancer | Azure Scale Sets and Load Balancer | Autoscaler and Cloud Load Balancing | | |  | | |  |  |  |  | | --- | --- | --- | --- | | **Databases** | RDS, DynamoDB, Aurora | SQL Database, Cosmos DB | Cloud SQL, Firestore, Bigtable | | | |  |  |  |  | | --- | --- | --- | --- | | **Virtual Networks** | VPC (Virtual Private Cloud) | Virtual Network | VPC (Virtual Private 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) | Infrastructure as a Service (IaaS) is a cloud computing model where a provider delivers virtualized computing resources over the internet.  Key Characteristics of IaaS:   1. On-demand resources 2. Pay-as-you-go pricing 3. Managed infrastructure 4. Flexibility | Startups and Small Businesses:   * A startup wants to launch a web application but doesn’t have the budget for physical servers. * They can rent vigrows anders and storage on-demand, scale as their user base grows, and only pay for what they use. |
| PaaS (Platform as a service) | Platform as a Service (PaaS) is a cloud computing model that provides a ready-to-use platform for developing, running, and managing applications without dealing with the underlying infrastructure.  Key Characteristics of PaaS:   * Pre-configured environments. * Managed infrastructure. * Built-in scalability. * Integration tools. | Collaborative Development:   * A distributed team of developers is working on the same application. * Cloud-hosted development environment allows multiple developers to code, test, and deploy collaboratively in real time. |
| SaaS (Software as a service) | Software as a Service (SaaS) is a cloud computing model where software applications are hosted by a cloud provider and made available to users over the internet.  The provider handles infrastructure, security, updates, and maintenance, while users focus on using the application.  Key Characteristics of SaaS:   1. Accessible via Internet 2. Subscription-based Pricing 3. Automatic Updates 4. Multi-Tenant Architecture | Collaboration & Communication:   * A remote team needs video conferencing and team messaging. * Tools like Zoom, Slack, or Microsoft Teams allow seamless communication without installing complex software on each device.   File Storage and Backup:   * A business wants secure, accessible cloud storage. * Dropbox or Google Drive provides file storage and sharing with automatic backups and version control. |

# Day 1: Task 3

Please research the below terms and explain what they are, when they would be appropriate and a real-world example of where it could be implemented (i.e. what type of organisation).

|  |  |
| --- | --- |
| Public Cloud | A public cloud is a cloud computing model where computing resources—such as servers, storage, networking, and applications—are owned and operated by a third-party provider and made available to multiple customers over the internet.  These resources are shared in a multitenant environment, meaning multiple organizations use the same physical infrastructure, but each customer's data and workloads are logically isolated to ensure privacy and security.  A public cloud is ideal when there is need for:   * Cost Efficiency * Scalability * Speed and Agility * Managed Services. * Global Reach   Real- world Example:  A startup launching an e-commerce platform might choose a public cloud like Amazon Web Services (AWS) or Microsoft Azure. These platforms offer scalable computing resources, storage, and databases that can grow with the business. During peak shopping seasons, the startup can quickly scale up resources to handle increased traffic and scale down afterward to save costs. |
| Private Cloud | A private cloud is a cloud computing environment dedicated to a single organization. Unlike a public cloud, all computing resources—servers, storage, and networks—are used exclusively by one organization, either hosted on-premises or by a third-party provider.  **Key Characteristics:**   1. Exclusive Use 2. Enhanced Security & Privacy 3. Customizable Infrastructure 4. Managed Internally or Externally   **Private Cloud Appropriate for:**   * Highly Regulated Industries * Custom Infrastructure Needs * Consistent Workloads * Control Over Data and Policies   **Real-World Example:**  A bank wants to host its core banking systems and customer financial data in a secure environment.  The bank sets up a private cloud using VMware or Azure Stack within its own data centres.  The bank maintains strict security and compliance, controls all infrastructure, and ensures reliable access to critical applications for employees and customers. |
| Hybrid Cloud | A hybrid cloud is a computing environment that combines both public and private clouds, allowing data and applications to be shared between them.  Organizations can run sensitive workloads in a private cloud while taking advantage of the scalability and cost-efficiency of a public cloud for less critical workloads.  **Key Characteristics:**   1. Integrated Environment 2. Workload Flexibility 3. Scalability 4. Centralized Management   **Hybrid Cloud Appropriate when handling:**   * Sensitive & Non-Sensitive Data * Variable Workloads * Business Continuity & Disaster Recovery * Gradual Cloud Migration   **Real-World Example:**  A hospital wants to keep patient records secure in a private cloud for compliance but run analytics on non-sensitive data in a public cloud to reduce costs.  Critical Electronic Health Records stay in a private cloud; research and reporting workloads use AWS or Azure public cloud.  This combines security, compliance, scalability, and cost efficiency in one solution. |
| Community Cloud | A community cloud is a cloud infrastructure shared by several organizations that have common concerns, such as security, compliance, or industry requirements. Unlike a public cloud, the cloud is not open to everyone but shared among a specific community of users.  **Key Characteristics:**   1. Shared Among Organizations 2. Common Goals 3. Cost Sharing 4. Managed Internally or Externally:   **Community Cloud is Appropriate when there is:**   * Industry Collaboration * Regulatory Compliance * Cost Efficiency * Joint Projects or Research   **Real-World Example:**  Several government agencies want to share data and applications securely while complying with strict security regulations.  A FedRAMP-compliant cloud is established as a community cloud, allowing agencies to host applications and data accessible only to authorized members.  This ensures security, compliance, cost-sharing, and collaboration across government departments. |

# 

# 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 a computer or system without permission, even if no damage is done. | Logging into a company system using someone else’s password without permission. |
| Unauthorised access with intent to commit further offences | Gaining access to a computer or system to commit a crime, such as fraud or theft. | Hacking into a bank’s system intending to steal money. |
| Unauthorised modification of computer material | Changing, deleting, or damaging data without permission, including introducing malware. | Uploading a virus or ransomware to a company server to corrupt files. |

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 |
| The Police and Justice Act 2006 strengthened the Computer Misuse Act 1990 by introducing additional powers to tackle cybercrime more effectively. Three key extra powers are:   1. Offence of creating, supplying, or obtaining malware    * It is now illegal to make, distribute, or possess tools (like viruses, keyloggers, or hacking software) intended for committing computer misuse. 2. Targeting Denial-of-Service (DoS) attacks    * The Act explicitly makes DoS and DDoS attacks criminal offences, addressing cases where hackers overload systems to disrupt services. 3. Increased penalties for serious computer crimes    * Provides longer prison sentences and higher fines for severe offences, especially when substantial financial loss or disruption occurs. |
|  |
|  |

Look at the below website to answer the questions:

<https://www.gov.uk/personal-data-my-employer-can-keep-about-me>

|  |
| --- |
| Write down three items of data which a company can store about an employee. |
| * Name * Address * Date Of Birth |
|  |
|  |

|  |
| --- |
| Give three more examples of data that an employer can only store if they first get the employee’s permission. |
| * Race and Ethnicity * Religion * Political membership or opinions |
|  |
|  |

Conduct further research to answer the below questions.

|  |  |
| --- | --- |
| Question | Answer |
| Provide one example of: Copyright infringement | Downloading or sharing music, movies, or software without the creator’s permission |
| Provide one example of: Plagiarism | Copying text from a website or book and presenting it as your own work without giving credit to the original author. |
| What are two consequences of copyright infringement and software piracy? | 1. Legal consequences   * Individuals or organizations can face fines, lawsuits, or even criminal charges for using copyrighted material without permission.   2. Financial and reputational damage   * Companies or individuals may have to pay compensation and could lose trust or credibility. |
| Give three possible consequences for individuals when using pirated software | * Legal consequences * Security risks * Lack of support and updates |

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

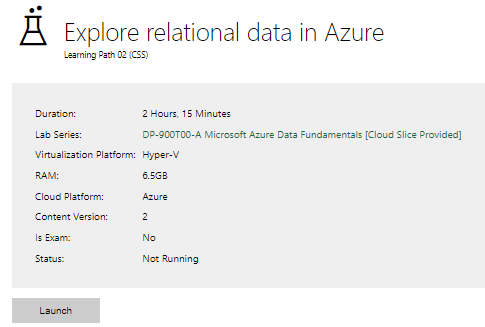
7. Consumer Rights Act 2015

* 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

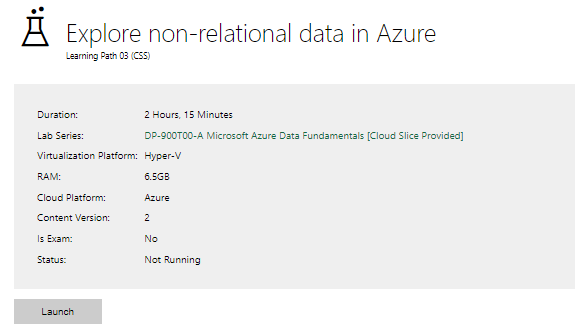
Please complete the below lab (3) *‘Explore relational data in Azure’* and paste evidence of the completed lab in the box provided.



|  |  |
| --- | --- |
| Completed lab | Provision of Azure SQL Database:    Writing queries in Query editor: |

# Day 3: Task 2

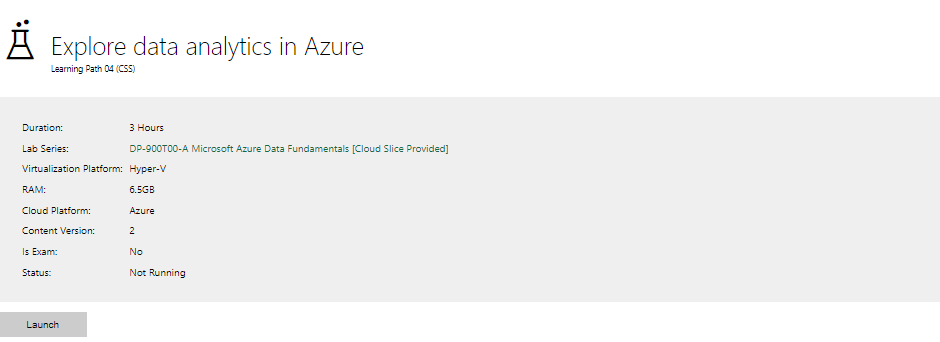
Please complete the below lab (4) *‘Explore non-relational data in Azure’* and paste evidence of the completed lab in the box provided.



|  |  |
| --- | --- |
| Completed lab | Provisioning storage account  Created container and uploaded json file:    Error Message when trying to upgrade to storage account with Azure Data Lake Gen2    Azure files |
|  | Data storage as Tables |
|  |  |
|  | Created Cosmo DB account and created Sample Database |
|  | In SQL query editor, query to select the item that have the name “Helmet” |

# 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.

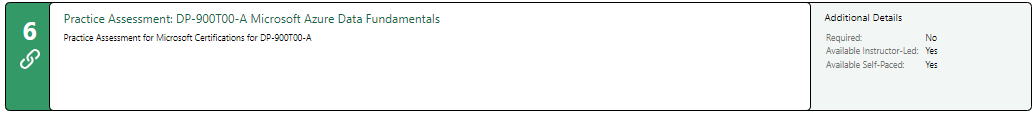


|  |  |
| --- | --- |
| Completed lab | Ingested sample data in Lakehouse    Used SQL Query by changing Lakehouse view to SQL Analytics endpoint to retrieve data from the ingested data    Created workspace, Lakehouse and Ingested Data |
|  | Created event stream and ingested real time data(yellow taxi) |
|  | Created event house and stored the captured data in the table |
|  | Query the table to see the captured data |
|  | Added Table to Sales Report |
|  | Changed Table to column chart |
|  | Drilled down to see individual products in the category |
|  | Added Pie chart to the report with the category and Quantity |
|  | Map added in report to show revenue by city |
|  | Published sales report in Power BI service |

# 

# Day 4: Task 1

In your teams, complete the Azure DP-900 practice exam and paste your result below – this is open book and please research and discuss your answers as a team.



|  |  |
| --- | --- |
| Result |  |

# Day 4: Task 2

#### **1. Scenario Background**

"Paws & Whiskers" is a growing pet shop that aims to improve its business by analysing sales, customer information, and inventory data. Currently, the data is collected manually or stored in spreadsheets. Management is interested in transitioning to Microsoft Azure to streamline data storage, analysis, and reporting, enabling them to make data-driven decisions.

#### **2. Data Laws and Regulations**

Identify and explain the data laws and regulations relevant to handling customer data within the proposal. Ensure you cover the following points:

* **GDPR Compliance**: Highlight the importance of adhering to the General Data Protection Regulation (GDPR), particularly as it relates to storing and processing customer information.
* **Data Protection Act (DPA) 2018**: Outline how the DPA 2018 may affect the way "Paws & Whiskers" collects and stores data, ensuring compliance with UK laws on data privacy.
* **Other Industry Standards**: Research any additional data protection standards or regulations that may apply to pet shop data, particularly if they involve sensitive or payment information.

#### **3. Azure Service Recommendations**

Recommend Microsoft Azure services that would suit the company’s data analysis needs and explain why these services are suitable. Your recommendations should include:

* **Data Storage**: Identify suitable storage options, such as **Azure Blob Storage** or **Azure SQL Database**, and discuss the benefits of each for storing large datasets, including inventory, sales transactions, and customer details.
* **Data Analysis Tools**: Recommend tools such as **Azure Machine Learning** for customer behaviour analysis or **Azure Synapse Analytics** for analysing sales trends.
* **Data Integration and Automation**: Explain how services like **Azure Data Factory** could automate data collection and integration processes, improving efficiency.

#### **4. Data Types and Data Modelling**

Define the types of data "Paws & Whiskers" will need to work with and describe your approach to data modelling:

* **Data Categories**: Identify key data types, such as customer demographics, transaction history, pet inventory, and product categories.
* **Data Modelling Approach**: Outline how you would structure this data using a relational model or a data warehouse approach, considering factors like tables, entities, relationships, and primary keys.

#### **5. Data Storage Formats and Structures in Azure**

Discuss how you would store data within Azure and the formats you would recommend:

* **Data Formats**: Specify recommended formats (e.g., CSV for raw data imports, JSON for structured data, Parquet for analytics) and explain why these formats are suitable for specific data types.
* **Data Security and Encryption**: Include recommendations for securing data using Azure’s built-in encryption features and access controls to ensure compliance with data privacy regulations.

#### **6. Additional Considerations**

Provide any other considerations that might enhance data handling and efficiency in Azure, such as:

* **Backup and Disaster Recovery**: Outline a backup plan using **Azure Backup** or **Azure Site Recovery** to safeguard against data loss.
* **Data Visualisation**: Discuss potential use of **Power BI** within Azure for creating dashboards that provide management with real-time insights into sales and customer trends.
* **Future Scalability**: Comment on how Azure services can scale as the business grows, accommodating larger datasets and more complex analyses.

### **Submission Guidelines:**

1. **Structure**: Ensure your report is well-organised, with sections for each task (e.g., Data Laws, Azure Services, Data Types, etc.).
2. **Formatting**: Include headings, bullet points where appropriate, and any visuals or diagrams that support your explanations.
3. **References**: Cite any resources or regulations referenced in the report.
4. **Length**: Aim for 1500-2000 words.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1. Scenario Background**  "Paws & Whiskers" is a growing pet shop that currently collects data manually or stores it in spreadsheets. Management recognizes the need to transition to a more robust and efficient digital system to improve business decision-making. By leveraging Microsoft Azure, the company can streamline storage, automate data collection, enhance analysis, and enable actionable insights. Migrating to Azure offers scalable solutions for sales analysis, customer insights, inventory management, and future business growth.  **2. Data Laws and Regulations**  Handling customer data requires adherence to various data protection laws and industry standards. Ensuring compliance protects customers’ privacy, maintains trust, and prevents legal or financial penalties.  **a) GDPR Compliance**  The General Data Protection Regulation (GDPR) is a European Union regulation that governs the collection, processing, and storage of personal data. Key requirements for "Paws & Whiskers" include:   * **Lawful Processing**: Personal data (e.g., customer names, addresses, and purchase history) must be processed lawfully, transparently, and fairly. * **Purpose Limitation**: Data should only be collected for explicit, legitimate purposes such as marketing communications, loyalty programs, or order fulfilment. * **Data Minimization**: Only the necessary data should be collected, avoiding unnecessary personal information. * **Storage Limitation**: Personal data should be stored only as long as required to fulfil business or legal obligations. * **Rights of Data Subjects**: Customers have rights to access, correct, or delete their data. * **Data Security**: Adequate measures must be taken to prevent unauthorized access, data breaches, or misuse.   For a business like "Paws & Whiskers," GDPR compliance ensures that customer data is handled responsibly, particularly when integrating cloud-based storage and analytics solutions in Azure.  **b) Data Protection Act (DPA) 2018**  The UK Data Protection Act 2018 complements GDPR, incorporating its principles into UK law and providing additional guidance. Key considerations include:   * **Legal Basis for Processing**: The company must identify lawful bases for collecting personal information (e.g., consent, contractual necessity). * **Enhanced Protections for Sensitive Data**: While pet shop data is generally not highly sensitive, payment information or health data for pets may require additional safeguards. * **Accountability and Documentation**: The company should document processing activities and demonstrate compliance. * **Data Transfers**: If storing data in international Azure data centres, the company must ensure transfers comply with UK regulations.   By adhering to the DPA 2018, "Paws & Whiskers" ensures that its data handling meets UK-specific legal requirements.  **c) Other Industry Standards**  Other relevant standards include:   * **PCI DSS (Payment Card Industry Data Security Standard)**: If the company processes credit/debit card payments, it must comply with PCI DSS requirements to secure financial data. * **ISO/IEC 27001**: This international standard for information security management provides a framework for protecting business and customer data in the cloud. * **Local Privacy Policies**: Any marketing campaigns or loyalty programs must follow opt-in consent rules for electronic communications.   Implementing these standards ensures comprehensive protection across customer, transaction, and business data.  **3. Azure Service Recommendations**  To optimize data storage, analysis, and integration, Microsoft Azure provides a range of services suitable for "Paws & Whiskers."  **a) Data Storage**  Azure offers several storage options depending on the type and volume of data:   * **Azure Blob Storage**:   + Ideal for unstructured data like sales records, product images, and receipts.   + Benefits include cost-efficiency for large datasets, scalability, and integration with analytics services. * **Azure SQL Database**:   + Suitable for structured relational data such as customer information, transaction history, and inventory details.   + Supports indexing, querying, and relational integrity with primary and foreign keys.   + Offers built-in security, automated backups, and high availability. * **Azure Data Lake Storage**:   + Optimized for analytics workloads, large-scale storage, and multi-format data (CSV, JSON, Parquet).   + Enables integration with machine learning and data transformation tools.   **Recommendation:**  Use Azure SQL Database for structured business data and AzureBlob/Data Lake Storage for raw, unstructured, or analytics-focused datasets.  **b) Data Analysis Tools**  For deriving insights from data, Azure provides robust analytical tools:   * **Azure Machine Learning**:   + Allows predictive analysis of customer behaviour, such as purchase trends, product recommendations, or churn prediction.   + Can be integrated with historical sales and loyalty program data. * **Azure Synapse Analytics**:   + Enables comprehensive analysis of large datasets, such as sales trends across product categories and seasonal demand patterns.   + Integrates data from multiple sources, including SQL databases, Blob Storage, and external APIs.   **Recommendation:**  Utilize Azure Synapse Analytics for sales and inventory analysis and Azure Machine Learning for predictive customer behaviour modelling.  **c)Data Integration and Automation**  Azure Data Factory (ADF**)** is a cloud-based data integration service that enables automated movement and transformation of data from diverse sources into a centralized repository. Key features and benefits include:   * **Automated Data Pipelines**   + ADF allows "Paws & Whiskers" to create automated workflows (pipelines) that extract data from multiple sources, such as spreadsheets, e-commerce platforms, or supplier databases. * **Data Transformation (ETL/ELT)**   + Data can be transformed before storage, such as cleaning, formatting dates, standardizing product codes.   + ADF supports both ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform) workflows, allowing flexibility based on analytics needs. * **Integration Across Platforms**   + ADF integrates seamlessly with Azure services like Azure SQL Database, Blob Storage, and Synapse Analytics, enabling centralized storage and analytics. * **Error Handling and Monitoring**   + Automated pipelines include built-in monitoring and error-handling capabilities. Notifications can be set up for failed data transfers, ensuring quick resolution without manual checks. * **Efficiency Gains**   + Automation reduces the time staff spend on data entry and consolidation, minimizes human errors, and ensures that analytics and reports always reflect the most recent data.   **Recommendation:**  Implement Azure Data Factory to streamline data ingestion and integration, reducing the need for manual spreadsheet handling.  **4. Data Types and Data Modelling**  **a) Data Categories**  "Paws & Whiskers" will need to manage several key data types:   * **Customer Data**:   + Customer ID, Name, contact information, preferences, loyalty program membership, and purchase history. * **Transaction Data**:   + Transaction ID, Sales records, timestamps, product ID, quantities, and payment methods. * **Product Data**:   + Product ID, Product details (e.g., pet food, accessories), stock levels, suppliers, and reorder dates. * **Product Categories**:   + Category ID, Classification of items (e.g., dogs, cats, birds, small animals), pricing, and promotional offers.   **b) Data Modelling Approach**  To manage these data categories efficiently, a structured data model is necessary. Two complementary approaches are recommended: relational model for transactional data and data warehouse model for analytics and reporting.  **1. Relational Model**  A relational model stores structured data in tables with defined relationships. This approach is ideal for day-to-day operations such as sales tracking and inventory management.  **Proposed Tables and Keys:**   | **Table Name** | **Primary Key** | **Key Fields / Description** | | --- | --- | --- | | Customers | Customer ID | Name, email, phone, address, DOB, loyalty status | | Products | Product ID | Name, category, price, supplier ID, stock quantity | | Transactions | Transaction ID | Customer ID, date, total amount, payment method | | ProductCategories | Category ID | Category name, description, Price |   **Relationships:**   * Customers → Transactions: One-to-many (a customer can have multiple transactions). * Transactions → TransactionItems → Products: Many-to-many via junction table (a transaction can include multiple products). * Products → ProductCategories: Many-to-one (each product belongs to one category).   **Benefits:**   * Ensures data integrity through primary and foreign keys. * Supports transactional operations like sales processing and inventory updates. * Reduces redundancy and maintains consistency.   **2. Data Warehouse Approach (Analytical Database)**  A data warehouse centralizes data from multiple sources, optimized for analysis and reporting rather than day-to-day transactions. For "Paws & Whiskers," a star schema is recommended.  **Star Schema Design:**   * Fact Table: SalesFact   + Fields: Transaction ID, Customer ID, Product ID, DateID, Quantity, Total Amount. * Dimension Tables:   + CustomerDim: Customer ID, Name, Contact, Loyalty Status, Demographics   + ProductDim: Product ID, Name, Category ID, Price, SupplierI D   + CategoryDim: Category ID, Category Name   + DateDim: DateID, Day, Month, Quarter, Year   **Recommendations:**  By combining a relational model for operational efficiency with a data warehouse for analytics, "Paws & Whiskers" can ensure:   * Accurate day-to-day transaction processing. * Efficient inventory and customer management. * Powerful insights for strategic decision-making and growth.   **5. Data Storage Formats and Structures in Azure**  Azure provides a range of options for “Paws & Whiskers” for storing, encrypting, and managing data in compliance with regulatory standards such as GDPR and the Data Protection Act 2018.  **a) Data Formats:**   | **Format** | **Recommended Use Case** | **Benefits** | | --- | --- | --- | | **CSV (Comma-Separated Values)** | Ideal for raw data imports from spreadsheets, point-of-sale systems, and supplier data uploads. | - Easy to create and read by both humans and systems. - Widely compatible across databases and analytical tools. - Suitable for data ingestion into Azure Data Factory or Azure SQL Database. | | **JSON (JavaScript Object Notation)** | Suitable for customer profiles, inventory data, or e-commerce transactions coming from online platforms. | - Supports nested and complex data. - Integrates well with REST APIs and NoSQL databases like Azure Cosmos DB. - Ideal for exchanging structured data between applications. | | **Parquet** | Best for large-scale analytics and data warehouse queries in Azure Synapse Analytics or Azure Data Lake. | - Highly efficient for querying and aggregating data. - Reduces storage costs through compression. - Enables faster analytical performance compared to CSV. |   **b) Data Security and Encryption**  Protecting sensitive customer and business data is vital for regulatory compliance and maintaining trust. Azure offers a range of built-in security features to safeguard data at every stage — at rest, in transit, and during processing.  **1. Encryption at Rest**  Azure automatically encrypts stored data using Storage Service Encryption (SSE) with AES-256-bit encryption, one of the strongest encryption standards available.   * All data stored in Azure SQL Database, Blob Storage, or Data Lake will be encrypted by default.   **2. Encryption in Transit**  To prevent unauthorized interception of data as it moves between users and Azure services:   * All communication uses Transport Layer Security (TLS 1.2 or higher). * Data transfers between Azure services and applications are encrypted end-to-end.   **3. Access Control and Authentication**  Access should be limited to authorized users and systems:   * Role-Based Access Control (RBAC): Assigns permissions based on user roles * Azure Active Directory (Azure AD): Manages user authentication and supports Multi-Factor Authentication (MFA) to enhance security.   **5. Compliance and Monitoring**  Azure provides continuous monitoring and compliance management tools:   * **Microsoft Defender for Cloud:** Detects potential threats and provides security recommendations. * **Azure Policy:** Ensures that data governance and compliance rules (GDPR, DPA 2018, PCI DSS) are consistently applied. * **Audit Logs:** Track all data access and administrative activities for accountability and reporting.   **6. Backup Security**  Even backup data is encrypted automatically using Azure’s encryption standards, ensuring full protection in case of recovery scenarios.  **Recommendation:**  Use encryptions for all sensitive data, implement RBAC, and regularly audit access logs.  **Additional Considerations:**  Beyond core storage, analysis, and integration capabilities, several additional Azure features can significantly enhance data handling, reliability, and long-term efficiency for “Paws & Whiskers.” These include backup and disaster recovery planning, data visualisation for insights, and future scalability to accommodate business growth.  **a) Backup and Disaster Recovery**  To prevent data loss, corruption, or system failure, a robust backup and disaster recovery plan is essential.   * **Azure Backup**:   + Provides automated backups for SQL databases, Blob Storage, and virtual machines.   **Key Features:**   * + Backups can be configured to run daily, weekly, or monthly.   + Data is replicated to a secondary Azure region, ensuring protection against regional outages. * **Azure Site Recovery**:   + Enables replication of critical systems to secondary regions for disaster recovery. * **Recommendation:**   Schedule daily backups and test recovery processes regularly.  **b) Data Visualisation**   * Turning data into actionable insights is key to supporting management decisions. Azure seamlessly integrates with Power BI, a powerful data visualisation and business intelligence tool:   + Connects directly to Azure SQL Database, Synapse Analytics, or Data Lake.   + Supports interactive dashboards for sales trends, inventory levels, and customer behaviour.   + Enables real-time reporting and alerts for low-stock items or high-performing products.   + Dashboards can be securely shared across departments or with external partners.   **c) Future Scalability**  Azure’s cloud infrastructure ensures that the company can seamlessly handle larger datasets, more complex analyses, and additional business applications without disruption.   * **Elastic Scaling**:   + Azure SQL and Synapse can scale compute and storage as dataset size increases. * **Serverless Options**:   + Azure Functions can automate tasks without maintaining dedicated servers. * **Integration Flexibility**:   + Easily integrate with new applications, online sales platforms, or IoT devices as business needs expand. * **Cost Efficiency:**   + Azure’s flexible pricing allows the business to pay only for the resources used, scaling up or down as the business evolves.   **Future Analytics and AI Integration**  As the company collects more data, services like **Azure Machine Learning** and **Cognitive Services** can provide advanced capabilities such as demand forecasting, automated customer segmentation, and sentiment analysis from customer feedback.  Azure ensures that “Paws & Whiskers” remains future-ready—able to scale storage, processing power, and analytical capacity in line with business expansion, without needing to redesign its data infrastructure.  **References:**   * European Union. (2016). *General Data Protection Regulation (GDPR)*. * UK Government. (2018). *Data Protection Act 2018*. |

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