

**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 transformed industries and everyday life by offering scalable, flexible, and cost-effective solutions to real-world challenges. Here’s how it impacts individuals, businesses, and society:   * **Businesses** can avoid upfront infrastructure costs (like servers) and pay only for what they use (e.g., AWS, Azure, Google Cloud). * Tools like Microsoft Teams, Zoom, and Google Workspace rely on the cloud to enable real-time collaboration (shared documents, video conferencing) and secure access to data from any device, anywhere. * Automatic backups: Cloud providers like AWS or Backblaze store data redundantly across global servers. So there is no chance to lose your work or data * **Innovation**: Experiment with AI, IoT, machine learning, and big data analytics without investing heavily in specialised hardware. * **Accessibility**: Access applications and data from anywhere with an internet connection, enabling remote work and global collaboration.   In short, cloud computing powers innovation across nearly every sector by making advanced technology **accessible**, **affordable**, and **scalable**. Whether you’re streaming a movie, collaborating with colleagues, the cloud is the invisible engine driving progress. |
| How can it benefit a business? | * Cost Savings: Reduces the need for upfront investments in hardware and maintenance, as businesses pay only for what they use. * Scalability: Quickly scale resources up or down to match business needs, such as during seasonal demand or rapid growth. * **Remote Work Enablement**: Supports remote and hybrid work by providing access to data and applications from anywhere. * Enhances collaboration * Innovate faster with AI, big data, and IoT. * Ensure security and compliance. * **Global Reach**: Enables businesses to serve customers and deploy applications worldwide with minimal delay. |
| What’s the alternative to cloud computing? | The primary alternative to cloud computing is on-premises computing (also known as traditional IT infrastructure), where a business hosts and manages its own hardware, software, and networking infrastructure locally, rather than relying on a third-party cloud provider.   * Infrastructure: Businesses purchase, maintain, and upgrade their own servers, storage, and network hardware. * Control: you have complete control over data, applications, and systems * Costs: Involves high upfront costs for hardware and software, plus ongoing costs for maintenance, power, and cooling. * Scalability: Scaling up requires purchasing and installing new hardware, which can be slow and expensive.     There are also other alternatives:   * Hybrid Cloud: Combines on-premises infrastructure with cloud services for greater flexibility. * Edge Computing: Processes data closer to its source, such as IoT devices, reducing latency for certain use cases. * Colocation: Renting physical space in a third-party data centre to house privately owned servers. |
| What cloud providers can we use, what are their features and functions? | Each cloud provider has unique strengths and is best suited for different industries or use cases. Businesses often choose based on factors like scalability, cost, specific tools, or regional needs.  **Amazon Web Services (AWS)**   * Wide range of services including compute (EC2), storage (S3), and databases (RDS, DynamoDB). * Advanced AI/ML tools like SageMaker. * Serverless computing via AWS Lambda. * Elastic scalability and pay-as-you-go pricing.   **Microsoft Azure**   * Deep integration with Microsoft tools like Office 365 and Active Directory. * Hybrid cloud capabilities via Azure Arc. * Comprehensive AI, IoT, and machine learning offerings. * Azure DevOps for continuous integration and deployment (CI/CD). * Security and compliance features tailored for enterprises.   **Google Cloud Platform (GCP)**   * Superior tools for AI/ML, including TensorFlow and Vertex AI. * Kubernetes expertise through Google Kubernetes Engine (GKE). * Scalable storage options via BigQuery for data warehousing. * Focus on sustainability and energy-efficient data centres.   **IBM Cloud**   * **Strong focus on AI and machine learning through IBM Watson.** * **Hybrid cloud solutions with open-source Kubernetes (OpenShift).** * **Robust enterprise-grade security and compliance.** * **Blockchain as a service for businesses.**   **Oracle Cloud Infrastructure (OCI)**   * High-performance databases like Oracle Autonomous Database. * Robust support for enterprise workloads. * AI, analytics, and machine learning tools. * Cost-effective storage and compute options.   **Alibaba**   * Dominates in Asia with localised services. * Advanced security offerings for compliance. * Scalable storage, AI, and IoT services. |

# 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 that provides virtualised computing resources over the internet. It offers the fundamental infrastructure needed to run applications, including virtual machines, storage, networking, and operating systems, without the need for businesses to invest in and manage physical hardware. | **Infrastructure as a Service (IaaS)** is widely used in the real world for various scenarios where businesses or individuals require flexible, scalable, and cost-effective computing infrastructure without investing in physical hardware. For example:  **Hosting Websites and Applications**  IaaS helps with this because they could use virtual machines to host the website, and they can easily scale resources up or down based on website traffic you can manage goal availability through multiple data centres. An example of this would be a startup holding their e-commerce site on Azure virtual machines.  **Big Data Analytics**  The company wants to process and analyse massive data sets IaaS helps by providing high performance virtual machines for processing big data workloads. It also uses scalable storage solutions for data lakes. For example, a retailer analysing customer purchase data using AWS to forecast trends. |
| PaaS (Platform as a service) | **PaaS** is a cloud computing model that provides a platform and environment for developers to build, deploy, and manage applications without needing to worry about managing the underlying infrastructure, such as servers, storage, and networking. It sits above Infrastructure as a Service (IaaS) in the cloud stack and is designed to simplify the development process. | **Application Development and Deployment**  If I company needs to build and deploy a web or mobile application quickly PaaS helps by providing a preconfigured development environment with tools libraries and frameworks. It allows developers to write code test undeployed directly to the cloud.  For example, a startup using Google app engine to launch a scalable social media app.  **IoT Applications**  If a business needs to manage analyse and respond to data from IoT devices. PaaS can help by providing tools for processing io T data and integrating it with other applications it also includes features for data storage analytics and visualisation. For example, a smart agriculture company using Azure IoT central to collect and analyse data from sensors.  In summary, PaaS is ideal in scenarios where developers want to reduce time-to-market, avoid infrastructure management, and focus on innovation. It provides the tools, frameworks, and scalability needed for modern application development across various industries. |
| SaaS (Software as a service) | **Software as a Service (SaaS)** is a cloud computing model where software applications are delivered over the internet on a subscription basis. Users access the software through a web browser without needing to download, install, or manage it. The service provider handles all infrastructure, maintenance, updates, and security. | **Email and Communication Tools**  If a business needs e-mail hosting, team communication, and collaboration tools. SaaS helps by providing e-mail services and collaborative platforms with no need for on premises servers. It's enables real time communication and file sharing from anywhere. For example, Microsoft 365 outlook for e-mail teams for communication and OneDrive for file sharing.  **Streaming Services**  Netflix for on demand video streaming and entertainment.  In summary, SaaS simplifies workflows across industries by providing on demand access to software tools and services, eliminating the need for local installation and maintenance. |

# 

# 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 services like storage, compute power, and applications are provided by third-party vendors over the internet and shared among multiple users. It is cost-effective, scalable, and accessible globally.  **Microsoft Azure**: A popular public cloud platform offering services like virtual machines, databases, and AI tools for businesses and developers. |
| Private Cloud | A **private cloud** is a cloud computing model dedicated to a single organisation, providing exclusive access to infrastructure, resources, and services. It offers greater control, security, and customisation compared to public clouds.  **IBM Cloud Private**: A private cloud solution used by enterprises to deploy and manage applications securely within their own data centres. |
| Hybrid Cloud | A **hybrid cloud** is a cloud computing model that combines both public and private cloud infrastructures, allowing data and applications to be shared between them. It provides flexibility, allowing businesses to scale workloads on the public cloud while maintaining sensitive data on the private cloud.  **Microsoft Azure Stack**: A hybrid cloud solution that allows organisations to run Azure services in their own data centres, seamlessly integrating with Microsoft Azure’s public cloud. |
| Community Cloud | A **community cloud** is a cloud computing model where the infrastructure is shared by several organisations with common interests, such as security, compliance, or specific business needs. It allows for collaborative sharing of resources while maintaining a certain level of control and privacy.  **Government Cloud**: A cloud infrastructure shared by government agencies, like the **U.S. Federal Government's Cloud Infrastructure** or **FedRAMP**, which ensures that the cloud services meet strict security and compliance requirements. |

# Day 2: Task 1

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

|  |  |  |
| --- | --- | --- |
| Area | Description | Example |
| Hacking – Unauthorised Access to Computer Material. | This section of the Computer Misuse Act criminalises the act of gaining access to a computer system or data without permission. | An example of this would be a hacker bypassing security measures to access a company’s database without consent. Or someone using another person’s login credentials without authorisation to access that e-mail or personal accounts. |
| Unauthorised Access with Intent to Commit or Facilitate the Commission of Further Offences | This refers to accessing a computer system with the intent to commit a further crime such as theft fraud or damage to data. | And the example would be a person gaming unauthorised access to a network and then using that access to steal sensitive information like credit card details or personal identification data for financial gain. |
| Unauthorised Modification of Computer Material | This offence involves intentionally altering, damaging or deleting data on a computer system without permission. | An example of this would be a cybercriminal deploying malware like a virus or ransomware to corrupt or delete files or even altering financial records in a way that homes the company or its stakeholders. |

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 |
| Offense of "Possession of Articles for the Purpose of Committing a Computer Misuse Offense"  This provision criminalises the possession of tools or software that are specifically designed to commit Computer Misuse offences, even if the tools have not been used yet.  An example of this would be a person found in possession of hacking tools such as a software that could be used to bypass security systems can be prosecuted even if they have not yet used these tools to hack into a computer system. |
| Extended Powers for Law Enforcement to Investigate Computer Misuse Crimes  This act grants to police extended powers to seize and examine computer equipment in the investigation of Computer Misuse crimes this helps law enforcement to investigate digital evidence more effectively.  For example, in cases involving cybercrime the police can seize computers, hard drives or mobile devices suspected of being used in crimes like hacking or the distribution of malware, allowing for detailed forensic analysis to gather evidence. |
| Offense of "Accessing Data Without Authorisation" for the Purpose of Viewing  Act also introduced an offence related to accessing data on the computer without authorisation and with the intent to view it even if no further modification or damage occurs.  For example a person who gains unauthorised access to a company's internal system and use confidential files (such as employee records or financial reports) can be criminally charged even if no changes all damages were made to the data.  These additions strengthen the legal framework surrounding the Computer Misuse and give law enforcement the tools needed to better investigate and prosecute cybercrimes in an increasingly digital world. |

Look at the below website to answer the questions:

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

|  |
| --- |
| Write down three items of data which a company can store about an employee. |
| Personal Identification Information   * Example: Name, address, date of birth, and contact details. * This data helps the company identify the employee and maintain communication. |
| Employment Details   * Example: Job title, department, salary, and employment start date. * This data is necessary for managing employment records, payroll, and other HR processes. |
| Performance and Attendance Records   * Example: Performance appraisals, attendance history, and sick leave records. * This data helps in managing the employee's performance, monitoring attendance, and making decisions about promotions or disciplinary actions. |

|  |
| --- |
| Give three more examples of data that an employer can only store if they first get the employee’s permission. |
| Sensitive Health Information   * Example: Medical history, disability status, or health conditions. * Employers can store this data only if the employee provides explicit consent, especially if it is necessary for accommodations or health and safety purposes. |
| Biometric Data   * Example: Fingerprints, facial recognition data, or iris scans used for security or timekeeping purposes. * This type of data is considered sensitive and requires explicit consent from the employee to be collected and stored. |
| Political Opinions or Religious Beliefs   * Example: Information about an employee's political affiliations, religious practices, or beliefs. * This data is considered sensitive and can only be stored with explicit consent, as it is related to the employee's personal life and may have a higher risk of misuse. |

Conduct further research to answer the below questions.

|  |  |
| --- | --- |
| Question | Answer |
| Provide one example of: Copyright infringement | **Copying and distributing a copyrighted database without permission**   * If someone copies a database that is protected by copyright (such as a proprietary list of business contacts, research data, or a collection of articles) and distributes it without the permission of the copyright holder, that constitutes copyright infringement. The creator or owner of the database holds exclusive rights to reproduce and distribute it, and using the data without permission violates those rights. |
| Provide one example of: Plagiarism | **Copying text from a website or book and presenting it as your own work**   * If a student copies a passage or several sentences from a website, book, or article and submits it in a research paper or essay without properly citing the source, they are committing plagiarism. This is a form of intellectual dishonesty, as it involves using someone else's work or ideas without giving proper credit. |
| What are two consequences of copyright infringement and software piracy? | **Legal Penalties**  Individuals or companies found guilty of copyright infringement or software piracy may face legal action including hefty fines or even imprisonment for example distributing pirated software can lead to lawsuits or criminal charges with fines ranging from thousands to millions of dollars depending on the severity and the scale of the infringement.  **Reputational damage**  Being associated with illegal activities can damaged trust with customers, partners, and stakeholders leading to loss of business partnerships and long-term credibility in market. |
| Give three possible consequences for individuals when using pirated software | **Legal Consequences**  This can result in costly legal fees and penalties or even imprisonment  **Security Risks**  Pirated software often lacks proper security updates and patches, making it vulnerable to malware, viruses, and other cyberattacks. Using pirated software can expose personal and financial data to hackers, leading to identity theft, data breaches, or system compromises.  **Loss of Software Support and Updates**  When using pirated software, individuals miss out on legitimate software updates, bug fixes, and customer support. This can cause the software to become outdated, unstable, or incompatible with other systems, reducing its effectiveness and usability. |

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 |
| 1 | It is illegal to create or use a hacking tool for penetration testing |
| 6 | Personal data may only be used for specified, explicit purposes |
| 5 | Employers must provide their computer users with adequate health and safety training for any workstation they work at |
| 2 | It is illegal to distribute hacking tools for criminal purposes |
| 3 | It is illegal to distribute an illicit recording |
| 6 | Personal data may not be kept longer than necessary |
| 1 | Gaining unauthorised access to a computer system is illegal |
| 5 | Employers must ensure that employees take regular and adequate breaks from looking at their screens |
| 2 | It is illegal to prevent or hinder access (e.g. by a denial-of-service attack) to any program or data held in any computer |
| 6 | Personal data must be accurate and where necessary kept up to date |

# Day 3: Task 1

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



|  |  |
| --- | --- |
| Completed lab |  |

# Day 3: Task 2

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



|  |  |
| --- | --- |
| Completed lab |  |

# Day 3: Task 3

Please complete the below lab (5) ‘Explore data analytics in Azure’ and paste evidence of the completed lab in the box provided.



|  |  |
| --- | --- |
| Completed lab |  |

# 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.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data Laws**  When handling customer data within the “Paws and Whiskers” proposal, several data laws and regulations must be considered to ensure compliance with legal and ethical standards. These include:  **General data protection regulation (GDPR) (EU/UK)**  GDPR applies to how they collect, store and process customer data.  They should make sure to:   * Obtain explicit consent from customers before collecting personal data * Ensure shared data security and protection against breaches * Use data minimisation - only collect necessary data * Implement data protection policies and appoint a data protection officer if necessary.   **UK Data Protection Act 2018**  For businesses operating in the UK, this law enforcers DPR principles while including specific provisions tailored to the UK such as: extra protection for processing sensitive data.  **Payment Card Industry Data Security Standard (PCI DSS) (Global)**  Since Paws and Whiskers deals with sales and customer payments they must comply with PCI DSS when handling credit slash debit card transactions. Which include:   * secure storage of payment data * encrypt card holder information * regular security assessments   **Consumer Rights Act 2015 (UK)**  This law impacts how customer information is used in transactions and data analytics. So, it will ensure fair trading and protection of customer rights.  **Animal Welfare Act 2006 (UK**)   * Governs the treatment and sale of animals in the UK. * Requires businesses to ensure animals are kept in appropriate conditions. * Customers' data related to pet ownership (e.g., medical history, breed info) must be handled securely under GDPR.   **Veterinary Medicines Regulations 2013 (UK)**   * If selling pet medicines or treatments, the business must comply with strict regulations on storage and sales. * Customer prescription data should be handled according to GDPR and Data Protection Act 2018.   **Pet Animals Act 1951 (UK)**   * Requires businesses selling pets to have a pet shop license. * Ensures animals are sourced responsibly and kept in good conditions. * If collecting customer data related to pet ownership (for vaccinations, microchipping, or purchases), GDPR still applies.   **The Licensing of Activities Involving Animals Regulations 2018 (UK)**   * Covers pet sales, dog breeding, and animal boarding businesses. * Pet shops must be licensed and follow ethical trading standards. * Any data collected on pet owners, pet purchases, or veterinary details should be securely stored and compliant with data laws.   **Azure Service Recommendations**  **Data Storage**  That would be a mix of structured data and unstructured for example pet images.  **Azure Blob Storage**  Azure blob storage would be the best for pet images and videos and PDFs and other unstructured data  Benefits:   * its scalable and cost effective for storing large amounts of images * can be integrated with AI tools for image recognition for example, tagging pet breeds in the images. * Supports secure access control for privacy and compliance * and works seamlessly with Azure synapse analytics for analysis * you can use hot tier for quick access   **Azure SQL Database (For Structured Data)**  Stores sales, customer records, and inventory. You would not need a data lake as we do not have large sets of data for this small pet store.  You could store the structured data in Azure BLOB storage by saving it in formats like CSV, JSON, or Parquet.  However, there are some limitations for example:   * No querying abilities unlike SQL BLOB storage doesn't support SQL queries so if you need to run fast queries joins or relational operations you'd have to load it into another system. * No transactions SQL ensures data consistency with acid transactions which BLOB storage doesn't support. * No indexing SQL databases index data for fast lookups but BLOB storage requires scanning entire files.   **Data Analysis**  I would choose this because it analyses both structured (sales, inventory) and unstructured (pet images, customer feedback) data. For example, if pet images are tagged with their breed or age synapse could analyse patterns in pet sales.  **Power BI**  As it helps visualise both structured and unstructured data insights. We could create visuals like best-selling pet breeds, customer preferences, and seasonal trends.  **Data Integration & Automation**  **Azure Data Factory**  I would recommend this as it automates data flow from spreadsheets, databases, and even image uploads. It can extract metadata from images (e.g. pet breed, age) and store it in SQL.  **Azure Logic Apps**  It can automate workflows like sending adoption confirmation emails with pet images.   |  |  |  | | --- | --- | --- | | **Recommended Service** | **Purpose** | **Why this Service?** | | **Azure SQL database** | Stores customer and sales data. | Supports fast queries, indexing, and transactions for efficient data management. | | **Azure BLOB storage** | Stores pet images, videos and documents. | Scalable storage for images and documents, cost-effective for large datasets. | | **Azure synapse analytics** | Process is sales and pet trends. | Processes large-scale sales & inventory trends, supports both SQL and unstructured data. | | **Power BI** | Visualisation & reporting | Creates interactive dashboards for insights into sales and pet trends. | | **Azure data factory** | Move and transforms data from multiple sources. | Automates data migration from spreadsheets to SQL & Data Lake. | | **Azure logic apps (optional)** | Sends alerts, automates process. | Triggers alerts for low stock, sends customer emails, and automates reports. |   **Data Types and Data Modelling**  "Paws & Whiskers" will handle a variety of structured, semi-structured, and unstructured data across different business operations.   * Customer data (structured and semi structured) * Transaction history (structured data) * Pet inventory data (structured and unstructured data) * product catalogue data (structured and unstructured data) * Sales and business analytics data (semi structured and unstructured data) * System and automation data (unstructured and semi structured data)     **Data Storage Formats and Structures in Azure**  **Data storage formats**  JSON (JavaScript Object Notation) is ideal for semi-structured data (e.g. Pet Images, Product Descriptions), especially when dealing with customer information, product catalogues, or logs. It provides flexibility with nested data structures, making it suitable for representing complex data such as customer preferences or pet details.  JSON is also good for metadata, and Blob Storage for binary files (e.g. images, PDFs).  JSON is great for storing metadata such as descriptions, tags, or classifications associated with pet images or products.  **Structured and Analytical Data (e.g. Sales Analytics, Pet Categories)**  **Recommended Format**: **Parquet**  Parquet is a columnar storage format that’s highly optimised for large-scale analytics. It provides high compression and supports efficient query performance on large datasets, which is ideal for data warehousing and reporting in Azure Synapse Analytics or Azure Data Lake.  In the pet shop you could use it for Storing aggregated sales data, pet categories, or inventory levels that require frequent querying and analytics.  It is good because it provides columnar storage, making it efficient for analytics and aggregation tasks. It also reduces storage costs by compressing data more effectively than row-based formats like CSV.    **Encryption at Rest**  **Azure Storage Encryption (ASE)**:  Built-in Encryption: Azure automatically encrypts data stored in Azure Blob Storage, Azure Data Lake Storage, Azure SQL Database, and Azure Cosmos DB using encryption at rest. This ensures that all stored data, whether it's structured, unstructured, or semi-structured, is protected without the need for additional configuration.  Encryption Algorithms: Azure uses Advanced Encryption Standard (AES) with 256-bit keys, which is compliant with industry standards for data protection. This is a critical step in ensuring compliance with privacy regulations like GDPR or CCPA, as data must be encrypted to protect personally identifiable information (PII) and financial data.  **Encryption in Transit**  **Azure Storage Encryption in Transit:**  Data transferred between Azure Blob Storage, Azure SQL Database, or other Azure services is encrypted using TLS/SSL protocols to prevent interception by unauthorized parties. This ensures that data remains secure while being transferred over the internet or between systems.  **Access Control**  **Role-Based Access Control (RBAC):**  Use RBAC to assign permissions to users, groups, and applications, ensuring that only authorised users can access or modify specific data.  Define roles such as Data Engineer, Data Analyst, or Admin, and assign appropriate levels of access (read, write, delete).  This is an effective way to enforce least privilege access, ensuring that users can only perform actions that are necessary for their roles.  **Additional Considerations**  **Backup & Disaster Recovery**  Azure Backup provides a cost-effective and automated way to protect critical data against accidental deletion, corruption, or cyber threats.   * **Point-in-Time Restore** – Continuous - Allows quick recovery of database changes. * **Azure Blob Snapshot** – Daily - Protects product images, documents, and pet photos from accidental deletion. * **Azure Backup Vault** – Weekly - Ensures historical sales data is recoverable. * **Azure Site Recovery** – Continuous - Ensures business applications can be quickly restored.   **Azure Site Recovery (ASR) for Disaster Recovery** - replicates virtual machines (VMs), databases, and key systems to another Azure region to ensure minimal downtime in case of hardware failure or cyber threats. This means there will be Zero data loss and near instant failover if Azure’s primary region goes down.  **Data visualisation: Power BI for insights**  Power BI is an essential tool for creating interactive dashboards that provide valuable business insights.   * Sales Performance: Monitor daily, weekly, and monthly sales trends. * Customer Insights: Analyse repeat customers, preferences, and demographics. * Inventory Tracking: Detect low stock levels and automate reorder triggers.   Power BI stands out due to its ability to deliver real-time insights by seamlessly connecting to Azure SQL Database and Azure Data Lake for live data streaming. Power Bi also includes:   * Seamless Data Integration – Supports Excel, CSV, JSON, and SQL databases, making data consolidation easy. * Automated Reporting – Schedule reports for daily or weekly analysis, reducing manual effort. * AI-Driven Analytics – Uses artificial intelligence to detect trends, anomalies, and customer behaviour patterns, enhancing decision-making.   **Future Scalability: Expanding with Azure**  As Paws & Whiskers grows, Azure ensures seamless scalability to accommodate larger datasets, more locations, and advanced AI analytics.  **Azure scales with business growth**  The storage scalability is there a BLOB storage and is your data lake can grow infinitely handling thousands of pet images customer profiles and historical data sets  Also, Azure Storage is highly scalable, and you can scale up during peak seasons (e.g., Christmas or Black Friday sales) to handle increased data loads, then scale down when demand decreases.  Pay-as-You-Go Model – You only pay for what you use, so costs increase during peak times but decrease when demand drops.  **Performance Tiers for Cost Optimisation**   * **Hot Tier**: Fast access for frequently used data (e.g., customer transactions during sales). * **Cool Tier**: Cost-effective storage for less frequently accessed data. * **Archive Tier**: Cheapest option for long-term historical data.   You can use only what is necessary for you to access quickly.  By implementing Azure Backup, leveraging Power BI for insights, and scaling with Azure services, "Paws & Whiskers" can ensure data protection, real-time decision-making, and future-proof operations as the business grows. All whilst keeping things cost effective. |

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