Just IT

B2W group

Apprenticeships | Training | Recruitment

Data Technician

Name:

Course Date:

Table of contents

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?	In the real world, cloud computing allows us to: **Access Data & Apps Anywhere**: Use services like email (Gmail, Outlook), file storage (Google Drive, Dropbox), and software from any internet-connected device, instead of being tied to one specific computer. **Use Services On-Demand**: Stream movies and music (Netflix, Spotify), use social media (Facebook, Instagram), collaborate on documents (Google Docs), and get navigation (Google Maps) instantly. **Enable Businesses**: Companies use the cloud for flexible website hosting, online shops, customer management (CRM), accounting software, data backups, and enabling remote work with tools like Zoom or Microsoft Teams. **Save Costs & Scale Easily**: Reduces the need for expensive physical hardware. Users and businesses can pay for what they use and easily increase or decrease resources (like storage or computing power) as needed. **Access Powerful Technology**: Makes advanced tools like AI, machine learning, and big data analytics more accessible without needing supercomputers on-site.
How can it benefit a business?	Cloud computing offers numerous significant benefits for businesses, regardless of their size: **Reduced Costs**: Lowers spending on hardware, servers, and IT maintenance staff. Businesses pay for the resources they use (pay-as-you-go), shifting costs from large upfront investments (CapEx) to more manageable operating expenses (OpEx). **Scalability & Flexibility**: Allows businesses to easily increase or decrease computing resources (storage,

	power, applications) based on demand. This supports growth, handles busy periods efficiently, and allows quick adaptation to market changes without hardware constraints. **Enhanced Collaboration & Remote Work**: Enables employees to access data and applications securely from anywhere with an internet connection, facilitating teamwork, remote/hybrid work models, and access to a wider talent pool. **Improved Reliability & Disaster Recovery**: Cloud providers offer robust infrastructure with high uptime guarantees and built-in backup/recovery options, ensuring business continuity and protecting data from local hardware failures or disasters. **Increased Security**: Major cloud providers often invest heavily in security measures, potentially offering better protection against cyber threats than many businesses could afford on their own. **Access to Advanced Technology**: Makes powerful tools like data analytics, Artificial Intelligence (AI), and Machine Learning (ML) more accessible, allowing businesses to gain insights, innovate faster, and compete more effectively. **Focus on Core Business**: By outsourcing IT infrastructure management to the cloud provider, businesses can focus their time, effort, and resources on their primary activities and strategic goals
What's the alternative to cloud computing?	The main alternative to cloud computing is On-Premises Computing (often called traditional IT infrastructure or local hosting). This involves: **Owning and Housing Hardware**: Businesses purchase and physically keep their own servers, storage devices, and networking equipment within their own facilities (e.g., in a server room or a private data centre). **Direct Management**: The business's own IT team is fully responsible for installing, configuring, managing, maintaining, securing, and upgrading all the hardware and software.

	Local Access: Resources are typically accessed over the company's internal network rather than primarily over the public internet. **Upfront Investment**: Requires significant capital expenditure (CapEx) to buy the infrastructure.

What cloud providers can we use, what are their features and functions?	**1. Amazon Web Services (AWS)**
	
	• **Overview:** The market leader with the most extensive portfolio of services and the largest global infrastructure footprint. Known for its maturity, reliability, and vast ecosystem.
	• **Key Features & Functions:**
	  ○ **Compute:** EC2 (Elastic Compute Cloud - Virtual Machines), Lambda (Serverless Computing), EKS (Kubernetes Service).
	  ○ **Storage:** S3 (Simple Storage Service - Object Storage), EBS (Elastic Block Store - for EC2), EFS (Elastic File System).
	  ○ **Networking:** VPC (Virtual Private Cloud), Route 53 (DNS), CloudFront (CDN), ELB (Elastic Load Balancing).
	  ○ **Databases:** RDS (Relational Database Service - supports various SQL engines), DynamoDB (NoSQL), Redshift (Data Warehousing).
	  ○ **AI & Machine Learning:** SageMaker (Platform for ML), Lex (Chatbots), Polly (Text-to-Speech), Rekognition (Image/Video Analysis).
	  ○ **Analytics:** EMR (Elastic MapReduce - Big Data Processing), Kinesis (Real-time Data Streaming), QuickSight (Business Intelligence).
	  ○ **Developer Tools:** CodeCommit, CodeBuild, CodeDeploy, CodePipeline (CI/CD services).
	  ○ **Security & Identity:** IAM (Identity and Access Management), KMS (Key Management Service), Shield (DDoS Protection).
	
	2. Microsoft Azure
	
	• **Overview:** The second-largest provider, strong in the enterprise space, particularly for organizations already heavily invested in Microsoft products (like Windows Server, Office 365). Offers strong hybrid cloud capabilities.
	• **Key Features & Functions:**
	  ○ **Compute:** Azure Virtual Machines, Azure Functions (Serverless), AKS (Azure Kubernetes Service).
	  ○ **Storage:** Blob Storage (Object Storage), Disk Storage (for VMs), Azure Files (Managed File Shares).

- **Networking:** Azure Virtual Network (VNet), Azure DNS, Azure CDN, Load Balancer.
- **Databases:** Azure SQL Database, Azure Cosmos DB (Globally Distributed NoSQL), Azure Synapse Analytics (Data Warehousing).
- **AI & Machine Learning:** Azure Machine Learning, Cognitive Services (Vision, Speech, Language APIs), Bot Service.
- **Analytics:** Azure Databricks, HDInsight (Big Data Service), Power BI (Business Intelligence - often integrated).
- **Developer Tools:** Azure DevOps (includes Repos, Pipelines, Boards), Visual Studio Integration.
- **Security & Identity:** Azure Active Directory (Azure AD - Identity Management), Key Vault, Azure Sentinel (SIEM).

3. Google Cloud Platform (GCP)

- **Overview:** Known for its strength in data analytics, artificial intelligence (AI), machine learning (ML), and container orchestration (Kubernetes, which originated at Google). Often appeals to tech-savvy companies and developers.
- **Key Features & Functions:**
  - **Compute:** Compute Engine (Virtual Machines), Cloud Functions (Serverless), GKE (Google Kubernetes Engine).
  - **Storage:** Cloud Storage (Object Storage), Persistent Disk (for VMs), Filestore (Managed File Storage).
  - **Networking:** VPC Network, Cloud DNS, Cloud CDN, Cloud Load Balancing.
  - **Databases:** Cloud SQL (Managed MySQL, PostgreSQL, SQL Server), Bigtable (NoSQL), Spanner (Globally Distributed Relational Database), BigQuery (Data Warehousing/Analytics).
  - **AI & Machine Learning:** AI Platform, Vision AI, Speech-to-Text, Natural Language AI, AutoML.
  - **Analytics:** Dataproc (Managed Spark/Hadoop), Dataflow (Stream/Batch Processing), Looker (Business Intelligence - acquired by Google).
  - **Developer Tools:** Cloud Source Repositories, Cloud Build, Cloud Deploy.
  - **Security & Identity:** Cloud Identity, Key Management Service (KMS), Security Command Center.

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)	Provides fundamental building blocks for cloud IT. Gives access to computing resources like virtual servers (VMs), storage, and networks over the internet on a pay-as-you-go basis. **You manage:** Operating systems, middleware, applications, data. **Provider manages:** Underlying physical hardware.	**Migrating Existing Servers:** Moving physical servers from your office to virtual servers in the cloud ("lift and shift"). **Hosting Websites/Apps:** Running a website or application where you need full control over the server environment. **Disaster Recovery:** Setting up backup servers in the cloud in case your primary ones fail. **Testing & Development:** Creating temporary virtual machines to test software. **High-Performance Computing:** Renting powerful computing resources for complex calculations (e.g., scientific simulations).
PaaS (Platform as a service)	Provides an environment for developing, testing, delivering, and managing software applications. It hides the complexity of managing the underlying infrastructure (hardware, OS, networking). **You manage:** Applications, data. **Provider manages:**	**Web Application Development:** Quickly building and deploying a custom web application without worrying about server setup, patching, or OS management. **Database Services:** Using a managed database (like Cloud SQL, Azure SQL Database) where the provider handles updates, backups, and scaling. **API Development:** Creating and hosting Application Programming Interfaces (APIs) for others to use. **Business Analytics:** Using a platform to analyse data and build reports without managing the underlying infrastructure.

	Infrastructure, operating systems, middleware (like databases, messaging queues), runtime environments.	**IoT Backends:** Building the server-side logic for Internet of Things devices.
SaaS (Software as a service)	Delivers ready-to-use software applications over the internet, usually on a subscription basis. The provider manages all aspects: infrastructure, platform, application software, updates, and maintenance. **You manage:** Your user data within the application. **Provider manages:** Everything else.	**Email & Communication:** Using services like Gmail, Microsoft Outlook 365, Slack, or Zoom. **Office Productivity:** Using Google Workspace (Docs, Sheets) or Microsoft 365 online apps. **Customer Relationship Management (CRM):** Using Salesforce, HubSpot, or similar CRM software. **Accounting Software:** Using online accounting tools like Xero or QuickBooks Online. **File Storage:** Using Dropbox, Google Drive, or OneDrive. **Streaming Services:** Using Netflix, Spotify, etc. (These are consumer SaaS examples).

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	**What:** Computing resources (servers, storage) owned by providers (like AWS, Azure, GCP) and shared by multiple users over the internet. Pay-as-you-go. **Use:** For scalability, cost savings, variable workloads, standard applications, development/testing. **Example Org:** Startups, e-commerce sites, users of online email/CRM.
Private Cloud	**What:** Computing resources dedicated solely to one organization, offering high control and privacy. Can be hosted on-premises or by a third party. **Use:** For strict security/compliance needs, sensitive data, legacy systems, or when full environmental control is essential. **Example Org:** Banks, government agencies, healthcare providers.
Hybrid Cloud	**What:** A combination of public and private clouds (or on-premises infrastructure) working together, allowing data/apps to move between them. **Use:** To balance security (private) with scalability/cost (public), for gradual cloud migration, disaster recovery, or handling peak loads ("cloud bursting"). **Example Org:** Large enterprises, businesses with varying IT needs (e.g., retail).
Community Cloud	**What:** Cloud infrastructure shared by several organizations with common goals or requirements (e.g., industry-specific compliance, security needs). **Use:** For collaboration within a specific sector, meeting shared compliance standards, or cost-sharing specialized resources. **Example Org:** University research groups, government agency consortiums, industry associations.

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	This covers the basic act of accessing computer material (programs or data) when you do not have permission to do so, or exceeding the permission you do have. It's essentially illegal "hacking" or "Browse" – gaining entry to a system or data you shouldn't, even if you don't intend to do anything else. The act of simply gaining access is the offence.	A person guessing their colleague's password and logging into their email account just to look at their messages, without changing anything or sending emails. Another example is a student exploring a university's network and accessing student record files they are not authorised to view.
Unauthorised Access with Intent to Commit or Facilitate Further Offences	This is a more serious offence. It involves committing the Section 1 offence (unauthorised access) but with the additional *intent* to commit, or help someone else commit, further serious crimes (typically those punishable by imprisonment, such as theft, fraud, or blackmail). The further crime doesn't actually have to be successfully carried out; the intent at the time of the unauthorised access is sufficient.	An individual hacking into an online banking system with the specific intention of transferring funds illegally to their own account. Another example is accessing a company's confidential customer database without permission with the intent to copy the data and sell it to competitors or use it for identity theft.
Unauthorised Acts with Intent to Impair, or with Recklessness as to Impairing,	This offence deals with interfering with or damaging computer systems or data. It involves doing an unauthorised act (like deleting files, introducing malware, or changing data) with the intent to impair the operation of any computer, prevent or hinder	A disgruntled former employee deleting critical files from the company server before leaving. Someone knowingly sending an email attachment containing a virus or ransomware designed to disrupt the recipient's computer or encrypt their files. Launching a

Operation of Computer, etc	access to any program or data, or impair the reliability of data. It also covers doing such an act recklessly (i.e., knowing there's a risk of impairment and doing it anyway).	Distributed Denial of Service (DDoS) attack against a website to overwhelm its servers and make it inaccessible 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
The 2006 Act significantly raised the maximum prison sentences available to the courts for offences committed under the CMA 1990. For example, the maximum sentence for the Section 3 offence (unauthorised acts with intent to impair, or recklessness as to impairment) was increased substantially (from 5 years up to 10 years, and potentially life imprisonment if Section 3ZA applies following later amendments, though the main increase to 10 years was key in 2006). Sentences for Section 1 and 2 were also increased.
A significant new power was added by introducing Section 3A into the CMA. This made it illegal to make, adapt, supply, offer to supply, or obtain any 'article' (which includes computer programs or data) intending it to be used, or knowing/believing it is likely to be used, to commit or assist in offences under Section 1 (unauthorised access) or Section 3 (impairment).
The 2006 Act amended the wording of the Section 3 offence (unauthorised acts causing impairment). It specifically clarified that acts intended to prevent or hinder access to any program or data held in any computer are covered. It also adjusted the required mental element (intent or recklessness).

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
Health and medical conditions

Conduct further research to answer the below questions.

Question	Answer
Provide one example of: Copyright infringement	A common example of copyright infringement is **uploading a copyrighted movie or TV show to a file-sharing website without the permission of the copyright owner**. This act reproduces and distributes the content unlawfully, violating the exclusive rights of the copyright holder.
Provide one example of: Plagiarism	An example of plagiarism is **copying a published article or essay from the internet and presenting it as one's own work without proper citation**. This misrepresents someone else's ideas and writing as original.
What are two consequences of copyright infringement and software piracy?	**Legal action and financial penalties**: Offenders may face lawsuits, substantial fines, or criminal charges depending on the severity and jurisdiction. **Reputational damage:** Individuals or businesses caught infringing copyrights may suffer long-term damage to their reputation, reducing trust among customers and partners.
Give three possible consequences for individuals when using pirated software	**Security vulnerabilities**: Pirated software often lacks proper updates and may come bundled with malware, increasing the risk of data breaches or system infections.

	Lack of support and updates: Users may miss out on essential technical support, bug fixes, and security patches, leading to performance issues. **Legal risks**: Being caught using or distributing pirated software can result in fines or other legal actions, impacting one's personal and professional life.

Listed below are some laws which we have covered today:

1. Computer Misuse Act 1990
2. Police and Justice Act 2006 (Computer Misuse)
3. Copyright, Designs and Patents Act 1988
4. Copyright (Computer Programs) Regulations 1992
5. The Health and Safety (Display Screen Equipment) Regulations 1992
6. Data Protection Act 2018
7. Consumer Rights Act 2015

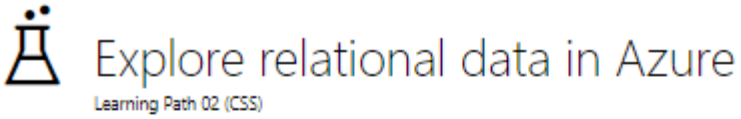
- Insert a number in the first column of each row to match each of the statements with one of the above Acts.

- One of statements is incorrect and not illegal. For this statement, write 'Not illegal'.

Act number	Clause
3	With some exceptions, it is illegal to use unlicensed software
Not illegal	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.



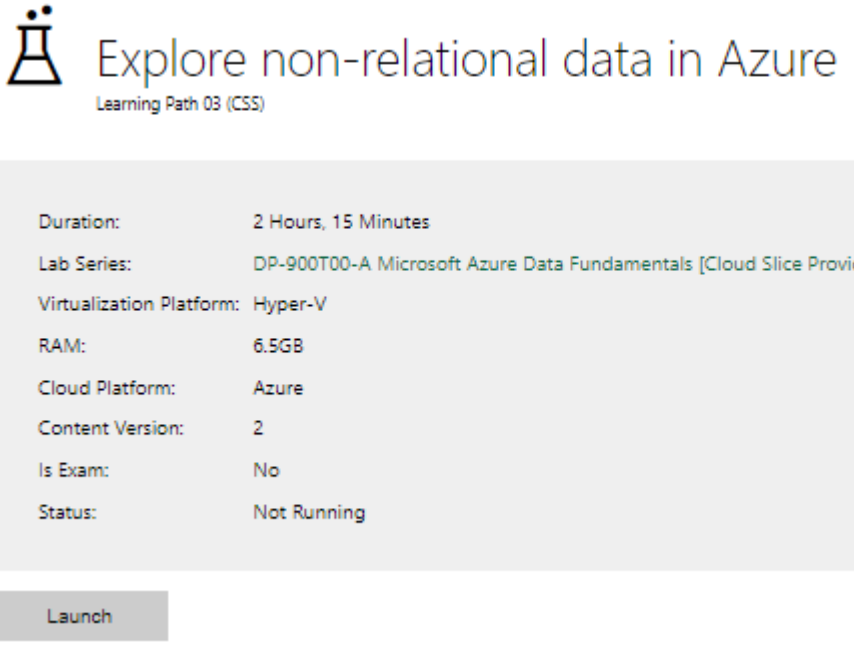
Duration:	2 Hours, 15 Minutes
Lab Series:	DP-900T00-A Microsoft Azure Data Fundamentals [Cloud Slice Provided]
Virtualization Platform:	Hyper-V
RAM:	6.5GB
Cloud Platform:	Azure
Content Version:	2
Is Exam:	No
Status:	Not Running

Launch

Completed lab	

Day 3: Task 2

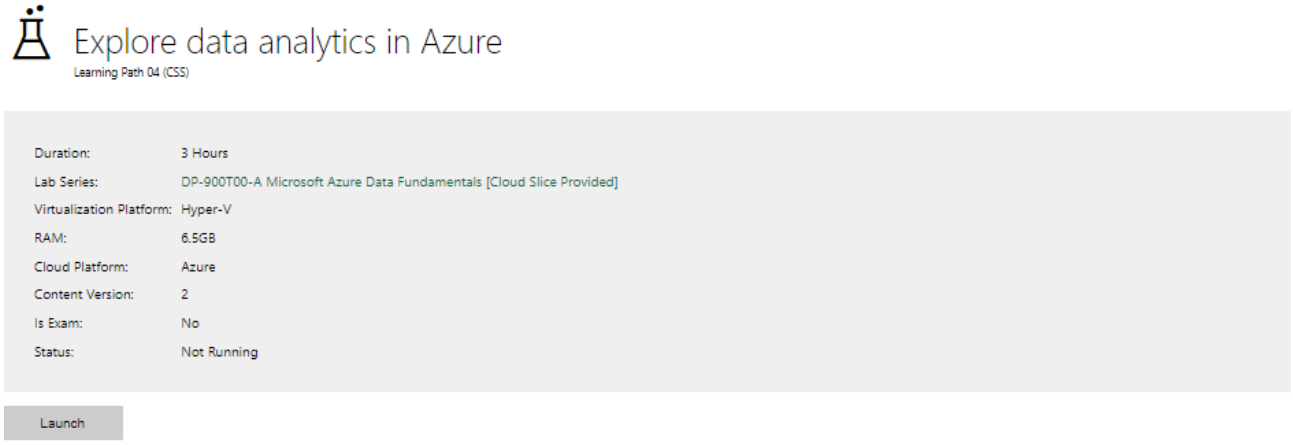
Please complete the below lab (4) *'Explore non-relational data in Azure'* and paste evidence of the completed lab in the box provided.



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 Strategy Proposal for “Paws & Whiskers”

1. Introduction

“Paws & Whiskers” is a growing pet shop aiming to harness technology for streamlined data storage, in-depth analytics, and informed decision-making. At present, the company maintains customer, sales, and inventory data manually or via spreadsheets. However, to propel business growth and achieve operational efficiency, management plans to transition to Microsoft Azure—a robust cloud platform capable of meeting dynamic business needs while ensuring data security and regulatory compliance.

The purpose of this report is to outline the relevant data laws and regulations that govern customer data handling, recommend Azure services to meet analysis and storage needs, discuss the types of data the business will process and model, and finally, propose storage formats, security measures, and additional considerations for long-term success.

2. Data Laws and Regulations

When processing customer data within the Azure environment, “Paws & Whiskers” must adhere to several key data protection laws and industry standards. This section outlines the requirements and explains the impact of each regulation.

GDPR Compliance

Overview:
The General Data Protection Regulation (GDPR) is a stringent data privacy standard applicable across the European Union. Although “Paws & Whiskers” is a pet shop (presumed to be based in or serving customers in regions where GDPR applies), the regulation’s principles are globally influential.

Key Points:

- **Lawfulness, Fairness, and Transparency:** Customer data must be processed in a lawful and transparent manner with clear communication about its use.
- **Purpose Limitation:** Data should only be collected for specified, explicit purposes and not processed in a manner that is incompatible with those purposes.
- **Data Minimisation:** Only the data necessary for the intended purpose should be collected.
- **Accuracy:** Personal data must be accurate and kept up to date.
- **Storage Limitation:** Data should not be held longer than necessary.
- **Integrity and Confidentiality:** Appropriate security measures must protect data from unauthorized access or loss.

Impact on “Paws & Whiskers”:
When storing customer information—including contact details, preferences, or payment information—the pet shop must obtain clear consent, secure data during transfer and storage (e.g., through encryption), and maintain transparent data processing policies. Regular audits and data protection impact assessments (DPIAs) might be required to ensure ongoing compliance.

Data Protection Act (DPA) 2018

Overview:
The UK's Data Protection Act 2018 supplements the GDPR by incorporating specific provisions applicable within the UK context. It outlines the rights of data subjects and imposes duties on organizations regarding how personal data is collected, processed, and shared.

Key Points:

- **Enhanced Subject Rights:** Individuals have the right to access, correct, or delete their personal data.
- **Accountability:** Organizations must demonstrate compliance with data protection principles, maintain records of processing activities, and appoint a Data Protection Officer (DPO) if necessary.
- **Sensitive Information Handling:** Particular categories of data, such as payment details or identifiers, require higher levels of protection.
- **Data Breach Notifications:** In the event of a breach, the organization must notify both the Information Commissioner's Office (ICO) and the affected individuals within set timeframes.

Impact on "Paws & Whiskers":
In practice, this means that when customer data is integrated into Azure-based systems, robust role-based access control and encryption must be implemented. Additionally, privacy-by-design principles should be built into the data management framework to comply with both GDPR and the DPA 2018.

Other Industry Standards

While GDPR and the DPA 2018 form the primary legal framework, additional standards may be relevant, especially when handling sensitive or financial information:

- **PCI DSS (Payment Card Industry Data Security Standard):**
  If "Paws & Whiskers" collects payment information from customers, compliance with PCI DSS is critical. This set of requirements helps protect cardholder data during transactions and storage.
- **ISO/IEC 27001:**
  This international standard for information security management systems (ISMS) provides a comprehensive framework that can enhance trust and reduce data breach risks. Certification against ISO 27001 may be considered as a way to demonstrate robust security practices.
- **Other Local Regulations:**
  Depending on the geographical areas in which the pet shop operates, there may be additional local data protection and consumer protection laws that govern data processing and electronic transactions. It is advisable to perform a legal review to ensure comprehensive compliance.

3. Azure Service Recommendations

Migrating to Microsoft Azure offers "Paws & Whiskers" the scalability, reliability, and integrated services required to handle large volumes of data efficiently. Below are the Azure services

recommended across three critical areas: data storage, data analysis tools, and data integration/automation.

Data Storage Options

Azure Blob Storage:

- **Use Case:** Ideal for storing large amounts of unstructured data such as raw CSV files, images, and backups.
- **Benefits:**
  - Highly scalable, cost-effective storage solution.
  - Supports multiple redundancy options to safeguard data.
  - Easy integration with other Azure analytics and data processing services.

Azure SQL Database:

- **Use Case:** Best suited for structured data such as customer records, transactional data, and inventory details.
- **Benefits:**
  - Managed relational database service that minimizes administrative overhead.
  - Provides built-in security features (encryption, threat detection) to meet data compliance standards.
  - Scalability options and dynamic performance tuning to handle varying workloads.

By using both storage options, the pet shop can balance the need for long-term archive storage (via Blob Storage) with the need for transactional operations and query performance (via SQL Database).

Data Analysis Tools

Azure Machine Learning:

- **Use Case:** For advanced analytics and predictive modelling, particularly in understanding customer behavior, forecasting sales trends, and personalizing marketing efforts.
- **Benefits:**
  - Facilitates the creation, training, and deployment of machine learning models at scale.
  - Integrates with other Azure services, including Azure Data Factory, enabling automated pipelines.
  - Offers a collaborative workspace for data scientists and analysts.

Azure Synapse Analytics:

- **Use Case:** Supports deep, enterprise-level data analysis across large datasets, combining big data and data warehousing.
- **Benefits:**
  - Integrated analytics service that provides a unified experience for ingesting, preparing, managing, and serving data.
  - Accelerates time to insight through powerful SQL-based analytical capabilities.

- Can handle complex join operations between diverse data sources, which is ideal for combining customer, sales, and inventory datasets.

Data Integration and Automation

Azure Data Factory:

- **Use Case:** Automates and orchestrates data movement and integration between various data sources and storage repositories.
- **Benefits:**
  - Supports data ingestion from on-premises systems, spreadsheets, and third-party applications.
  - Enables the creation of robust ETL (Extract, Transform, Load) pipelines.
  - Facilitates near-real-time data processing to keep dashboards and reporting tools up to date.

Overall, these Azure services can be combined to design a data ecosystem that supports both operational needs and strategic analytics while ensuring that performance, cost, and compliance objectives are met.

4. Data Types and Data Modelling

The business's daily operations and decision-making processes depend on several key data types. A well-defined data modelling approach will ensure that the data is structured effectively for querying, analysis, and reporting.

Data Categories

1. Customer Demographics:

- **Types of Data:** Names, contact details, addresses, preferences, and consent statuses.
- **Usage:** Targeted marketing, loyalty programs, and customer service improvements.

2. Transaction History:

- **Types of Data:** Sales records, payment amounts, methods, and purchase dates.
- **Usage:** Analysis of sales trends, payment processing, and revenue forecasting.

3. Pet Inventory:

- **Types of Data:** Lists of pets available for adoption or sale, including breeds, ages, health records, and care details.
- **Usage:** Inventory management, sales analysis, and product stocking decisions.

4. Product Categories:

- **Types of Data:** Information on pet supplies (food, toys, accessories, etc.), pricing, supplier information, and inventory levels.
- **Usage:** Stock management, supplier negotiations, and promotional planning.

Data Modelling Approach

Given the variety of data types, a combination of the following data models may be considered:

Relational Model:

- **Structure:** Data is organised into tables with rows and columns, where each table represents an entity (customers, transactions, inventory, products).
- **Relationships:**
  - **Primary Keys:** Unique identifiers for each table (e.g., CustomerID, TransactionID).
  - **Foreign Keys:** Fields that reference primary keys in related tables (e.g., linking customer records to transactions).
- **Benefits:**
  - Well-suited for transactional data that requires normalization (reducing redundancy).
  - SQL-based queries support complex joins and filters for detailed analysis.

Data Warehouse Approach:

- **Structure:** A centralized repository built on a star or snowflake schema to support analytics and reporting.
- **Components:**
  - **Fact Tables:** Contain metrics, such as sales figures and inventory counts.
  - **Dimension Tables:** Provide context to the facts, including dates, products, and customer demographics.
- **Benefits:**
  - Optimised for read-intensive operations and complex aggregations.
  - Ideal when combined with services like Azure Synapse Analytics, enabling faster queries across large datasets.

By combining both the relational model for day-to-day operations and a data warehouse schema for historical analysis, “Paws & Whiskers” can ensure both transactional efficiency and robust analytical capability.

5. Data Storage Formats and Structures in Azure

Once the data model is established, selecting the correct data formats and storage structures is vital. These choices impact both performance and compliance.

Recommended Data Formats

CSV (Comma-Separated Values):

- **Usage:**
  - Ideal for raw data imports from spreadsheets or external systems.
  - Simple, lightweight, and widely supported.
- **Benefits:**
  - Easy to generate and parse.
  - Suitable for one-off data loads or batch processing.

JSON (JavaScript Object Notation):

- **Usage:**
  - Suited for structured or semi-structured data that might change over time.
  - Commonly used for API integrations and when the schema is flexible.
- **Benefits:**
  - Readable and easily integrated into Azure services like Azure Functions or Logic Apps.
  - Supports nested data structures and dynamic changes.

Parquet:

- **Usage:**
  - Optimised for analytics on large-scale datasets, especially when using services like Azure Synapse Analytics.
  - Columnar storage format that accelerates read performance on aggregated data.
- **Benefits:**
  - Provides efficient data compression and encoding schemes.
  - Reduces storage costs and increases query performance on large datasets.

Data Security and Encryption

Data protection in Azure is paramount, not only to comply with GDPR and DPA 2018 but also to safeguard the business's reputation and customer trust.

- **Encryption at Rest:**
  - Azure provides built-in encryption mechanisms for data stored in Blob Storage, SQL Database, and other services. Data should be encrypted using Azure Storage Service Encryption (SSE) or Transparent Data Encryption (TDE) for SQL databases.
- **Encryption in Transit:**
  - Ensure that data moving between Azure services and external clients is transmitted over secure channels (TLS/SSL).
- **Access Controls and Identity Management:**
  - Implement role-based access control (RBAC) to restrict data access to authorised personnel only.
  - Use Azure Active Directory (AD) for centralized identity and access management.
- **Audit and Monitoring:**
  - Leverage Azure Security Center and Azure Monitor to track access patterns, detect anomalies, and alert administrators in case of suspicious activity.

6. Additional Considerations

Beyond the core components of storage, analysis, and regulatory compliance, several additional measures can further enhance operational efficiency and resilience.

Backup and Disaster Recovery

Azure Backup and Azure Site Recovery:

- **Backup:**

- Utilize **Azure Backup** to automatically create regular backups of databases and storage blobs. This service ensures that data is recoverable in the event of accidental deletion, corruption, or cyber incidents.

- **Disaster Recovery:**
  - Implement **Azure Site Recovery** to replicate virtual machines and services to a secondary region, ensuring business continuity during major outages.
- **Benefits:**
  - Minimises downtime.
  - Provides rapid restoration of services and data.

Data Visualisation

Power BI Integration:

- **Usage:**
  - Integrate Power BI with Azure SQL Database, Azure Synapse Analytics, or directly with Blob Storage to generate real-time dashboards and reports.
- **Benefits:**
  - Provides management with intuitive visual insights into sales trends, customer behaviors, and inventory levels.
  - Enhances decision-making by transforming raw data into interactive, graphical representations.
- **Considerations:**
  - Ensure data refresh rates are compatible with business requirements.
  - Use embedded analytics features if sharing interactive dashboards with external stakeholders.

Future Scalability

Azure Scalability Features:

- **Elasticity:**
  - Azure services are designed to scale up or out quickly. As "Paws & Whiskers" grows, the company can upgrade computing resources (e.g., more powerful SQL Database tiers or expanded Blob Storage capacity) without service disruption.
- **Serverless Architectures:**
  - Consider serverless options such as **Azure Functions** for certain processing tasks. This allows the business to pay only for actual usage while responding to variable workloads.
- **Global Distribution:**
  - If the pet shop's customer base expands internationally, Azure's global data centre network can provide low latency and high availability across multiple regions.
- **Cost Management:**
  - Utilize Azure's cost management and budgeting tools to monitor spending and optimise resource allocation in real time.

7. Conclusion

This report has provided a detailed roadmap for "Paws & Whiskers" to transition its data operations onto Microsoft Azure. Adhering to data protection laws such as GDPR and the DPA 2018 is fundamental to building trust with customers while ensuring compliance with legal

obligations. By leveraging a suite of Azure services—from scalable storage options like Azure Blob Storage and Azure SQL Database to advanced analytical tools like Azure Synapse Analytics and Azure Machine Learning—the pet shop can unlock actionable insights from its data.

Moreover, a well-considered data modelling strategy, thoughtful choices of storage formats, and robust security measures will not only address current data needs but also position "Paws & Whiskers" for future growth. Incorporating additional considerations such as backup and disaster recovery, data visualisation through Power BI, and scalable architectures ensures that the company remains agile in an ever-changing market environment.

Overall, this integrated approach will facilitate a smoother transition from manual data handling to a more automated, secure, and insightful cloud-based system that drives business improvement and customer satisfaction.

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:

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.