

**Data Technician**

|  |
| --- |
|  |

|  |
| --- |
| Name: Rash Paul Wadhwa |
| Course Date: |
|  |

**Table of contents**

[Day 1: Task 1 3](#_Toc204784403)

[Day 1: Task 2 4](#_Toc204784404)

[Day 1: Task 3 5](#_Toc204784405)

[Day 1: Task 4 6](#_Toc204784406)

[Day 2: Task 1 7](#_Toc204784407)

[Day 3: Task 1 11](#_Toc204784408)

[Day 3: Task 2 12](#_Toc204784409)

[Day 3: Task 3 14](#_Toc204784410)

[Day 4: Task 1 15](#_Toc204784411)

[Day 4: Task 2 16](#_Toc204784412)

[Day 4: Task 2 (Optional) 17](#_Toc204784413)

[**1. Scenario Background** 17](#_Toc204784414)

[**2. Data Laws and Regulations** 17](#_Toc204784415)

[**3. Azure Service Recommendations** 17](#_Toc204784416)

[**4. Data Types and Data Modelling** 17](#_Toc204784417)

[**5. Data Storage Formats and Structures in Azure** 18](#_Toc204784418)

[**6. Additional Considerations** 18](#_Toc204784419)

[**Submission Guidelines:** 18](#_Toc204784420)

[**Course Notes** 21](#_Toc204784421)

[**Additional Information** 22](#_Toc204784422)

# 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 provides us with on-demand access to computing resources, such as servers, storage, databases, networking, and software, over the internet. It enables businesses and individuals to utilise technology without needing to own and maintain the physical hardware themselves. |
| How can it benefit a business? | Cloud computing provides businesses with numerous notable benefits, making it a valuable tool for growth and efficiency. For example, Cost savings, Scalability, improved security and disaster recovery, and collaboration. |
| What’s the alternative to cloud computing? | The primary alternative to cloud computing is on-premises hosting. It involves a business owning, operating, and maintaining all of its infrastructure, including servers, storage, networking, and software in its physical data centre. |
| What cloud providers can we use, what are their features and functions? | The major cloud providers are Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP).  **1. Amazon Web Services (AWS)**   * **Features**: AWS is the largest and most comprehensive cloud platform, offering over 200 fully-featured services. It is known for its wide range of services, including compute (EC2), storage (S3), and databases (RDS). * **Functions**: It's highly scalable and flexible, making it ideal for everything from startups to large enterprises. AWS is popular for hosting websites, running big data analytics, and building complex applications.   **2. Microsoft Azure**   * **Features**: Azure offers a mix of services, with a strong emphasis on hybrid and multi-cloud environments. Its key strength is seamless integration with other Microsoft products, like Windows Server and Microsoft 365. * **Functions**: It's a great choice for businesses already using Microsoft software. Azure is widely used for enterprise applications, machine learning, and data analytics.   **3. Google Cloud Platform (GCP)**   * **Features**: GCP excels in data analytics, AI, and machine learning. It's the same infrastructure that powers Google's services like Search and YouTube. * **Functions**: GCP is known for its innovative services in big data, such as BigQuery and its advanced AI tools. It is also a leader in container management with Kubernetes |

# Day 1: Task 2

Please research the below cloud offerings, explain what they are and provide 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) | IaaS is a cloud offering that gives you virtualised computing resources over the internet. I rent the fundamental building blocks of a data centre, such as servers, storage, and networking, from a cloud provider. I’m responsible for managing the operating system, applications, and data, while the cloud provider manages the physical hardware. | 1. **Website Hosting**: A developer can use IaaS to host a website on a virtual server, giving them full control over the server's configuration and the software installed. 2. **Big Data Analysis**: Companies can use IaaS to provision large clusters of virtual servers for processing and analysing massive datasets without having to buy and maintain physical hardware. 3. **Disaster Recovery**: Businesses can use IaaS to create a backup of their entire IT infrastructure on the cloud, so if their primary location experiences an outage, they can quickly restore operations. |
| PaaS (Platform as a Service) | PaaS provides a platform for developing, running, and managing applications without the complexity of building and maintaining the underlying infrastructure. The cloud provider handles the operating system, server software, and networking, allowing developers to focus solely on writing and deploying their code. | 1. **Web Application Development**: A software developer can use PaaS to deploy a web application without worrying about setting up the web server, databases, or operating system. The platform handles all of the back-end infrastructure. 2. **Application Testing:** Developers can utilise a PaaS environment to test new software in a consistent and scalable setting before releasing it to the public. 3. **Building APIs**: Companies can use PaaS to quickly build and deploy application programming interfaces (APIs) for their mobile and web applications. |
| SaaS (Software as a Service) | SaaS delivers fully-functional software applications over the internet, typically on a subscription basis. With SaaS, the cloud provider manages everything—the application, data, runtime, and the underlying infrastructure. Users access the software through a web browser or a mobile app, eliminating the need to install or manage it. | 1. **Email Services**: Services like Gmail and Microsoft 365 are examples of SaaS, where users can access their email, calendar, and documents from any device with an internet connection. 2. **Customer Relationship Management (CRM)**: A sales team uses a cloud-based CRM tool like Salesforce to track customer interactions, manage leads, and automate sales processes from their web browsers. |

# 

# Day 1: Task 3

Pricing Calculator: **In this exercise, you use the Pricing Calculator to estimate the cost of running a basic web application on Azure.**

|  |  |
| --- | --- |
| Cost of a basic web application | Completed |

TCO Calculator: **In this exercise, you use the Total Cost of Ownership (TCO) Calculator to compare the cost of running a sample workload in your datacentre versus on Azure.**

|  |  |
| --- | --- |
| Datacentre vs Azure | Completed |

# Day 1: Task 4

Please research the terms below 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 environment where a third-party provider, such as Amazon Web Services (AWS) or Microsoft Azure, makes computing resources (like servers, storage, and applications) available to the general public over the internet. These resources are shared among multiple users, but their data and workloads are isolated from one another.  **Appropriate when:**   * An organisation needs to scale resources quickly and affordably. * The workload is not highly sensitive or subject to strict regulatory compliance (e.g., a public-facing website). * A business wants to reduce upfront capital expenditure on hardware.   **Example:** A small e-commerce startup uses a public cloud provider to host its website and process payments. The company gains from the provider's extensive scalability, enabling it to manage traffic spikes during sales events without investing in its own physical servers. |
| Private Cloud | A **private cloud** is a cloud computing environment dedicated exclusively to a single organisation. It can be physically located in the organisation's own data centre or hosted by a third-party provider. The key is that the infrastructure and services are used only by that organisation.  **Appropriate when:**   * An organisation has strict security, privacy, and regulatory compliance requirements. * The organisation needs complete control over its data and infrastructure. * The business has a predictable, stable demand for computing resources.   **Example:** A large financial institution uses a private cloud to store and process sensitive customer financial data. This model allows the institution to meet stringent industry regulations and have full control over security, ensuring that no other entity can access its data. |
| Hybrid Cloud | A **hybrid cloud** is a mixed computing environment that combines a public cloud, a private cloud, and/or on-premises infrastructure. It allows data and applications to be shared between these environments. The goal is to leverage the strengths of each model.  **Appropriate when:**   * An organisation has a mix of sensitive and non-sensitive workloads. * The organisation needs the flexibility to scale out to the public cloud during peak times while keeping core operations on a private cloud. * A business is in the process of migrating to the cloud and needs to maintain some on-premises infrastructure.   **Example:** A healthcare organisation uses a hybrid cloud. It stores sensitive patient records on a private cloud to ensure compliance with privacy regulations, such as HIPAA, while utilising a public cloud for less-sensitive data, including a public-facing patient portal or a billing application. This allows for scalability and cost efficiency without compromising on security. |
| Community Cloud | A **community cloud** is a collaborative cloud infrastructure shared by multiple organisations that have common goals, security requirements, or regulatory needs. It is similar to a private cloud but is shared among a specific group of users.  **Appropriate when:**   * Organisations within a specific industry need to collaborate on projects while maintaining data security and compliance. * A group of organisations wants to pool resources to reduce costs.   **Example:** A consortium of research universities uses a community cloud. This allows the universities to share a common infrastructure for data storage and processing, enabling collaborative research projects on massive datasets. The community cloud ensures that all member organisations adhere to the same security standards and research protocols, while also providing cost savings compared to each university building its private cloud. |

# Day 2: Task 1

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

|  |  |  |
| --- | --- | --- |
| Area | Description | Example |
| 1. Unauthorised Access to Computer Material | This offence involves simply gaining unauthorised access to a computer system or data. It is the basic act of hacking without any further criminal intent. | An employee uses their colleague's password to look at their emails without permission. They don't copy or alter any information; the act of unauthorised viewing is a crime. |
| 2. Unauthorised Access with Intent to Commit or Facilitate a Further Offence | This is a more serious crime where someone gains unauthorised access to a computer to commit another crime. | A person hacks into a company's database to steal customer credit card details, with the clear intent of using those details for fraud. The initial hacking is a crime, but the intent to commit fraud elevates the offence |
| 3. Unauthorised Acts with Intent to Impair the Operation of a Computer | This offence covers actions that intentionally disrupt, damage, or impair the function of a computer or data. | A disgruntled former employee installs a virus on their old company's network to delete files and shut down the servers. This intentional act of destruction is a crime |

The Computer Misuse Act 1990 is an act under which an individual can be criminalised for a computer-related offence. Describe three extra powers that the Police and Justice Act 2006 (Computer Misuse) has added.

|  |
| --- |
| Description |
| * 1. Making, owning or distributing hacking tools was made illegal if the person intends or believes it will be used for computer misuse.   2. Making, owning or supplying hacking tools, including creating or spreading malicious software that can be used in computer misuse offences such as viruses, worms, trojans, ransomware, or denial of service attacks.   3. Especially, Denial of Service attacks were made illegal in the Police and Justice Act, 2006, where a person can face 10 years in prison. |
|  |
|  |

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. |
| Employers can keep the following data about their employees without their permission:   * name * address * date of birth * sex * education and qualifications * work experience |
|  |
|  |

|  |
| --- |
| Give three more examples of data that an employer can only store if they first get the employee’s permission. |
| Employers need their employees’ permission to keep certain types of ’sensitive’ data, including:   * race and ethnicity * religion * political membership or opinions * trade union membership * genetics * biometrics, for example, if your fingerprints are used for identification * health and medical conditions * sexual history or orientation |
|  |
|  |

Conduct further research to answer the below questions.

|  |  |
| --- | --- |
| Question | Answer |
| Provide one example of: Copyright infringement | **Unauthorised use of a photograph.** A person finds a professional photographer's photo online and uses it on their website or in a commercial advertisement without obtaining permission or a license. This is a violation of the photographer's exclusive right to reproduce and distribute their work. |
| Provide one example of: Plagiarism | **Copying and pasting text from a source.** A student copies a paragraph from an article or a blog and pastes it directly into their essay without using **quotation marks or citing** the source. They present the words as their own, which is a form of academic dishonesty. |
| What are two consequences of copyright infringement and software piracy? | 1. **Legal Action and Fines:** Both individuals and companies can face civil lawsuits from the copyright holder, leading to substantial fines and monetary damages. These penalties are often designed to compensate the creator for lost profits. 2. **Criminal Charges:** Depending on the severity and scale of the infringement, individuals can face criminal charges, which may result in probation, community service, or even imprisonment. |
| Give three possible consequences for individuals when using pirated software | 1. **Exposure to Malware and Viruses:** Pirated software often comes from untrustworthy sources and may be bundled with viruses, malware, or ransomware that can damage your computer or steal personal data. 2. **Lack of Support and Updates:** Users of pirated software are not entitled to official technical support, security patches, or software updates. This leaves the software vulnerable to new security threats and can cause it to become unstable or unusable over time. 3. **Legal Liability:** If caught, an individual can be held personally liable for copyright infringement, which can lead to significant financial penalties. This could also affect their reputation and future employment opportunities. |

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

# Day 3: Task 1

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



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

# Day 3: Task 2: Skillable

Complete below exercises in lab environment using AdventureWorks DB.

**-- 1. Customers with last names starting with 'A'**

-- Scenario: The marketing department is preparing a campaign targeting customers whose last names begin with the letter 'A'. They need a list of these customers to personalize outreach messages.

A screenshot of a computer

AI-generated content may be incorrect.

**-- 2. Customers with first name starting with 'A' and last name ending with 'a'**

-- Scenario: The CRM team is analyzing customer name patterns for personalization strategies. They're particularly interested in users with a first name starting with 'A' and a last name ending in 'a'.

A close up of a text

AI-generated content may be incorrect.

**-- 3. Count of customers with first name starting with 'A' and last name including 'o'**

-- Scenario: A data analyst is generating statistics for a name segmentation report and needs to count how many customers meet both criteria: first name starts with 'A' and last name contains the letter 'o'.

A screenshot of a computer

AI-generated content may be incorrect.

**-- 4. Top 3 black products of size 'S' ordered by list price**

-- Scenario: The merchandising team is curating a list of 'Black' colored products in size 'S' for a limited-time promotion. They want to showcase the top 3 most affordable options first.

A screenshot of a computer code

AI-generated content may be incorrect.

**-- 5. Average standard cost of products**

-- Scenario: The finance team is calculating the average standard cost across all products to assess baseline production expenses.

**A screenshot of a computer

AI-generated content may be incorrect.**

**-- 6. Difference between average standard cost and average list price**

-- Scenario: The pricing strategy team is examining the average markup between the standard cost and the list price to evaluate profitability trends.

A close up of a text

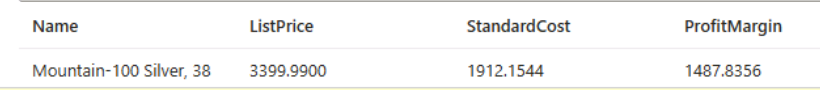
AI-generated content may be incorrect.

**-- 7. Product with highest profit margin**

-- Scenario: Management wants to identify the most profitable product by finding the item with the highest difference between list price and standard cost.

A screenshot of a computer

AI-generated content may be incorrect.



**-- 8. Number of products per category**

-- Scenario: Inventory control needs a summary of how many products exist in each product category to manage stock levels effectively.

SELECT

  c.Name,

  COUNT(p.ProductID) AS NumberOfProducts

FROM

  SalesLT.Product AS p

JOIN

  SalesLT.ProductCategory AS c

ON

  p.ProductCategoryID = c.ProductCategoryID

GROUP BY

  c.Name

ORDER BY

  c.Name;

**-- 9. Average list price by category**

-- Scenario: The product pricing team is reviewing average list prices by category to adjust pricing strategies and identify outliers.

SELECT

  c.Name AS CategoryName,

  AVG(p.ListPrice) AS AverageListPrice

FROM SalesLT.Product AS p

JOIN SalesLT.ProductCategory AS c

ON p.ProductCategoryID = c.ProductCategoryID

GROUP BY c.Name

ORDER BY AverageListPrice DESC;

**-- 10. Orders with customer information**

-- Scenario: The operations team needs a comprehensive report of customer orders, including order details and basic customer info, to review sales performance.

SELECT

  soh.SalesOrderID,

  soh.OrderDate,

  c.FirstName,

  c.LastName,

  c.EmailAddress

FROM SalesLT.SalesOrderHeader AS soh

JOIN SalesLT.Customer AS c

ON soh.CustomerID = c.CustomerID

ORDER BY soh.OrderDate DESC;

**-- 11. Products priced above average list price**

-- Scenario: The sales team wants to identify premium products that are priced above the average list price for highlighting in high-end marketing campaigns.

A screenshot of a computer

AI-generated content may be incorrect.

**-- 12. Total quantity sold per product**

-- Scenario: The business intelligence team is evaluating product performance by analysing the total quantity sold for each product across all orders.

A screenshot of a computer

AI-generated content may be incorrect.

**-- 13. Total sales per customer**

-- Scenario: The customer success team wants to calculate the total value of orders placed by each customer to identify high-value clients for loyalty programs.

A screenshot of a computer

AI-generated content may be incorrect.

# Day 3: Task 3: Skillable

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



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

# Day 4: Task 1: MS Fabric using Just IT

Please follow the link below to complete the lab using your Just IT account in MS Fabric.

There are 3 modules to complete.

[Data Factory end-to-end tutorial introduction and architecture - Microsoft Fabric | Microsoft Learn](https://learn.microsoft.com/en-us/fabric/data-factory/tutorial-end-to-end-introduction)

# Day 4: Task 2: Skillable

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 (Optional)

#### **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 relevant data laws and regulations for 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.

|  |
| --- |
| This report outlines a strategic proposal for "Paws & Whiskers" to migrate its sales, customer, and inventory data to Microsoft Azure. By leveraging Azure's robust suite of cloud services, the pet shop can move beyond manual spreadsheets to a scalable, secure, and automated data platform. This will enable management to gain deep, data-driven insights to improve business operations and customer satisfaction.  **Data Laws and Regulations**  Handling customer data responsibly is paramount. As a UK-based business, "Paws & Whiskers" must adhere to a strict legal framework. The transition to Azure must be managed with these regulations in mind.  **GDPR Compliance**  The General Data Protection Regulation (GDPR) is a comprehensive data privacy and security law that regulates how businesses handle personal data of individuals within the European Union (EU) and the UK. "Paws & Whiskers" must comply with key principles of GDPR, including:   * Lawful basis for Processing * Data Minimisation * Rights of the data subject * Data security   **Data Protection Act (DPA) 2018**  The DPA 2018 is the UK's implementation of the GDPR. It outlines specific requirements for data processing in the UK and works in tandem with UK GDPR. Key aspects relevant to "Paws & Whiskers" include:  **Information Commissioner's Office (ICO):** The DPA 2018 clarifies the powers of the ICO, the UK's independent authority for data protection. Non-compliance can lead to significant penalties, so it's critical to ensure all data-handling practices, including those on Azure, are aligned with ICO guidance.  The use of a major cloud provider like **Azure**, which has dedicated security teams and compliance certifications, simplifies the process of meeting **DPA 2018** requirements compared to managing on-premises data manually.  **Other Industry Standards**  Beyond the major data laws, "Paws & Whiskers" must consider other industry-specific regulations, particularly concerning payment information.   * **Payment Card Industry Data Security Standard (PCI DSS):** If the business handles credit card data—even if it's just the final four digits of a card number—it must comply with PCI DSS. While third-party payment processors typically handle the most sensitive data, "Paws & Whiskers" still has a responsibility to ensure its systems are secure. Azure provides a safe, compliant environment that helps businesses meet PCI DSS requirements, particularly for data in transit and at rest.   **3. Azure Service Recommendations**  To successfully transition from spreadsheets to a powerful data platform, "Paws & Whiskers" should leverage a combination of Azure services for storage, analysis, and automation.  **Data Storage**  The choice of storage is critical. I recommend a combination of two services to handle different data types and access patterns.   * **Azure SQL Database:** This is an excellent choice for structured data like customer details, sales transactions, and product information. As a relational database-as-a-service (DBaaS), it provides a familiar structure with tables, rows, and columns.   + **Benefits:** It ensures data integrity and consistency through normalisation and is perfect for Online Transactional Processing (OLTP), which is essential for a point-of-sale system. It's fully managed, meaning Azure handles patching, backups, and scaling, freeing up the "Paws & Whiskers" team. * **Azure Blob Storage:** This is ideal for storing unstructured data that doesn't fit neatly into a database, such as receipts, images of products, or raw data files.   + **Benefits:** It's a highly scalable and cost-effective object storage solution. Raw data from different sources (e.g., website logs, supplier CSV files) can be ingested here for later processing and analysis.   **Data Analysis Tools**  Once the data is stored, these tools will unlock its potential.   * **Azure Synapse Analytics:** This is a powerful, unified platform that brings together data warehousing, big data analytics, and data integration. It's the ideal tool for management to analyse sales trends and inventory data.   + **Benefits:** It allows for complex queries across large datasets, enabling management to identify trends like seasonal sales patterns, best-selling products, or correlations between product sales. It integrates seamlessly with Power BI for visualisation. * **Azure Machine Learning:** For more advanced insights, Azure Machine Learning can be used to predict customer behaviour.   + **Benefits:** By analysing transaction history, a model could be built to predict which customers are likely to churn (stop shopping at "Paws & Whiskers") or recommend products they might like. This could power targeted marketing campaigns and loyalty programmes.   **Data Integration and Automation**  Moving data from its current sources into Azure can be automated.   * **Azure Data Factory:** This is a cloud-based ETL (Extract, Transform, Load) service that automates the movement and transformation of data.   + **Benefits:** It can be used to create data pipelines that automatically ingest data from various sources (e.g., current spreadsheets, new point-of-sale systems, website logs) and load it into Azure SQL Database or Blob Storage on a scheduled basis. This eliminates manual effort and ensures data is always up-to-date and ready for analysis.   **Data Types and Data Modelling**  A clear understanding of the data is the first step toward effective analysis. The following data categories are crucial for "Paws & Whiskers."  **Data Categories**   * **Customer Demographics:** Information like name, contact details, date of birth, and postcode. This data is essential for customer segmentation and marketing. * **Transaction History:** Details of every sale, including transaction date, time, total amount, payment method, and a link to the specific customer and products. * **Pet Inventory:** Real-time stock levels, product names, categories (e.g., dog food, cat toys), supplier information, and cost. * **Product Categories:** A hierarchical classification of products to enable analysis by category (e.g., "Dog Food" -> "Dry Food" -> "Senior Dog Food").   **Data Modelling Approach**  A **data warehouse approach** using a star schema would be highly effective for "Paws & Whiskers" analytics needs. This model is optimised for quick querying and reporting, which is exactly what management requires.   * **Fact Table:** A central SalesFact table would contain the key metrics of the business, such as sale\_id, total\_price, quantity, and foreign keys to link to other tables (e.g., customer\_id, product\_id, date\_key). * **Dimension Tables:** These tables provide descriptive context for the data in the fact table.   + CustomerDim: Stores customer details, with customer\_id as the primary key.   + ProductDim: Contains product information, including product name, category, and supplier details, with product\_id as the primary key.   + DateDim: A time-based table with details for each day, month, and year, to easily analyse trends over time.   This structure allows for fast, powerful queries. For instance, to find the total sales of dog food in a specific month, you only need to join the SalesFact table with the ProductDim and DateDim tables, which is far more efficient than querying a single, massive spreadsheet.  **5. Data Storage Formats and Structures in Azure**  To ensure efficiency, cost-effectiveness, and compliance, data must be stored in appropriate formats within Azure.  **Data Formats**   * **CSV (Comma-Separated Values):** This is a simple and universal format. It is perfect for storing raw data exports from legacy systems or spreadsheets before it is processed. Azure Data Factory can easily ingest CSV files from Blob Storage for processing. * **JSON (JavaScript Object Notation):** This format is ideal for semi-structured data, such as website log files or API data. Azure's services can natively handle JSON, making it a good choice for data that doesn't fit a strict relational schema. * **Parquet:** This is a columnar storage format, which is highly recommended for storing large datasets used for analytics.   + **Benefits:** Parquet organises data by columns, allowing queries to read only the necessary columns, significantly speeding up performance and reducing costs. It also has superior compression, which further lowers storage expenses. The data warehouse tables in Azure Synapse Analytics would benefit from being stored in this format.   **Data Security and Encryption**  Azure provides a multi-layered security model to protect data.   * **Encryption at Rest:** All data stored in Azure SQL Database and Azure Blob Storage is encrypted by default using Microsoft-managed keys. For highly sensitive data, "Paws & Whiskers" can use Azure Key Vault to manage their own encryption keys, giving them full control. * **Encryption in Transit:** All data moving between the business's systems and Azure is encrypted using SSL/TLS protocols. * **Access Controls:** Azure Role-Based Access Control (RBAC) should be implemented to ensure that only authorised personnel have access to specific data. For example, a marketing team member might only need read-only access to customer demographics, while a finance manager might need access to sales transaction data.   **6. Additional Considerations**  **Backup and Disaster Recovery**  A robust plan is essential to protect against data loss from accidental deletion, hardware failure, or cyber-attacks.   * **Azure Backup:** This service can be used to automatically back up the Azure SQL Database and other critical data. It offers a simple, cost-effective, and highly available solution for data recovery. * **Azure Site Recovery:** This service provides disaster recovery for on-premises systems and virtual machines. While "Paws & Whiskers" is starting with data, this service would be vital if they were to migrate their full IT infrastructure to Azure in the future.   **Data Visualisation**  Data is only valuable if it can be easily understood.   * **Power BI:** This business analytics service seamlessly integrates with Azure Synapse Analytics. It would allow management to create interactive dashboards that visualise key performance indicators (KPIs) like daily sales, top-selling products, and customer trends in real time. These dashboards can be accessed from any device, providing insights at a glance.   **Future Scalability**  One of Azure's biggest advantages is its ability to scale.   * **Elasticity:** As "Paws & Whiskers" grows, Azure's services can be scaled up or down to meet changing demands. For example, the compute resources for Azure Synapse Analytics can be increased during a major sales event and then scaled back down to save costs. This eliminates the need for expensive hardware upgrades and ensures the data platform can grow with the business. * **Integration:** Azure's ecosystem is vast, allowing for easy integration of new services as needs evolve, such as implementing a customer loyalty application or a supplier portal. |

|  |
| --- |
| **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:

|  |
| --- |
| **nformation** |

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.**