Just IT

B2W group

Apprenticeships | Training | Recruitment

Data Technician

Name: Amnah Bibi

Course Date: 24/02/2025

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?	Cloud computing is useful for: – Data storage and backup on remote servers allowing easy access without relying on physical storage devices – SaaS - Software as a Service which offers ready-to-use applications like Google Docs, Microsoft 365, and Salesforce, which can be accessed from any device, allowing for more flexibility and collaboration. – Collaboration: Teams can access and work on the same files or applications simultaneously from different locations, promoting real-time collaboration and productivity. – Streaming & Entertainment: Services like Netflix, Spotify, and YouTube store massive amounts of media in the cloud, delivering entertainment across the world. – AI & Machine Learning: Cloud computing provides the infrastructure for running AI models and processing huge datasets, making it accessible to businesses without needing massive on-site computing power.
How can it benefit a business?	Cloud computing offers a range of advantages for businesses: 1. It is cost efficient - Instead of purchasing and maintaining expensive hardware, businesses pay for cloud services as they go, which lowers capital expenses and allows more flexibility in spending.

	2. Scalability: Cloud solutions can easily scale to meet demand, whether you're expanding your team or handling a traffic spike on your website. 3. Remote Access & Flexibility: Employees can access company resources from anywhere, fostering remote work and allowing businesses to attract talent from anywhere. 4. Reliability & Uptime: Cloud providers have data centers in multiple locations, ensuring redundancy, high availability, and minimal downtime for businesses. 5. Security: Many cloud providers invest heavily in robust security features, including encryption, firewalls, and multi-factor authentication to protect data. 6. Innovation & Agility: With cloud computing, businesses can quickly test and deploy new technologies and solutions, making them more agile and able to innovate faster.
What's the alternative to cloud computing?	The main alternative to cloud computing is on-premises computing (also called on-premises IT infrastructure). Involves businesses buying and maintaining their own hardware, software, and networking infrastructure. This requires significant upfront investment and ongoing maintenance costs, as well as dedicated IT staff. The pros of this are complete control over data security and infrastructure, potentially lower long-term costs (for larger enterprises with predictable IT needs) and greater control over privacy since data stays in-house. However, cloud computing is generally the preferred choice for businesses looking for flexibility, reduced upfront costs, and easier scaling.
What cloud providers can we use, what are their features and functions?	***Amazon Web Services (AWS)***

- Offers a wide range of services like computing power, storage, machine learning, AI, analytics, and security.
- Most comprehensive and mature cloud provider.
- Offers EC2 (Elastic Compute Cloud) for virtual servers, S3 (Simple Storage Service) for scalable storage, and many more services.

Target Audience: Enterprises, startups, and developers looking for a robust, customizable cloud platform.

Microsoft Azure

- Strong integration with Microsoft products like Windows Server, Office 365, and Active Directory.
- Offers cloud computing, analytics, storage, networking, and AI tools.
- Supports hybrid cloud, allowing businesses to run workloads both on-premises and in the cloud.

Target Audience: Enterprises already using Microsoft products, businesses requiring hybrid-cloud solutions, and developers.

Google Cloud Platform (GCP)

- Known for high-performance computer services and advanced AI/ML tools like TensorFlow.
- Strong offerings for big data, IoT, and machine learning with tools like BigQuery for data analysis and Google Kubernetes Engine for container orchestration.
- Seamless integration with other Google services (e.g., Google Workspace).

Target Audience: Businesses looking for data analytics, machine learning capabilities, and integration with Google's ecosystem.

IBM Cloud

- Focuses on AI, blockchain, and enterprise-level infrastructure.
- Offers a wide range of hybrid cloud solutions and strong security features.
- Specializes in integrating existing enterprise workloads with cloud solutions.

Target Audience: Enterprises with complex IT needs, particularly those using AI or needing hybrid cloud deployments.

Oracle Cloud

- Excellent for database management, offering Oracle Autonomous Database and Oracle Exadata Cloud for high-performance database workloads.
- Strong for enterprise resource planning (ERP) and customer relationship management (CRM).

Target Audience: Large enterprises with a focus on database and enterprise software.

Alibaba Cloud

- Leading cloud provider in China, with strong offerings in big data, computing, and storage.
- Offers cloud computing, elastic computing, data analytics, security, and AI solutions.
- Known for ECS (Elastic Compute Service) and OSS (Object Storage Service).

Target Audience: Businesses with a focus on the Asia-Pacific region or those doing business in China.

A quick summary =

- AWS: Best for flexibility, extensive tools, and global presence.
- Azure: Ideal for businesses relying on Microsoft products or requiring hybrid cloud.
- Google Cloud: Optimal for data-heavy businesses, machine learning, and Google integrations.

- IBM Cloud: Great for enterprises with specialised needs like AI or blockchain.
- Oracle Cloud: Suitable for enterprises focused on databases and ERP systems.
- Alibaba Cloud: Best for businesses targeting the Asian market, particularly China.

Day 1: Task 2

Please research the below cloud offerings, explain what they are and examples of use cases.

Cloud Offerings	Explain what it is	When / how might you use this service in the real-world?
IaaS (Infrastructure as a service)	It provides virtualised computing resources over the internet. It delivers infrastructure components like virtual machines (VMs), networking, storage, and other basic computing resources on a pay-per-use basis. With IaaS, you don't need to own or maintain physical	Examples are Amazon Web Services (AWS), Microsoft Azure and Google Cloud. Use Cases: 1. **Hosting Websites/Applications**: If you're running a website or web application that needs scalable infrastructure, IaaS allows you to quickly provision the right resources (e.g., VMs, load balancers). 2. **Big Data & Analytics**: Companies can rent scalable computing resources for data processing without needing to buy dedicated hardware. For example, running

	hardware—it's all hosted in the cloud. – Virtual machines – Storage solutions – Networking infrastructure (e.g., load balancers, firewalls) – Scalability on demand – API-driven management	large-scale data processing jobs on a cloud VM. 3. **Development and Testing**: Developers can create and tear down environments for testing without worrying about physical hardware. With IaaS, they can quickly spin up servers, databases, and networks as needed for testing. 4. **Disaster Recovery**: IaaS can offer cloud-based disaster recovery solutions where you back up your data and systems to a secure cloud environment in case of a system failure.
PaaS (Platform as a service)	It provides a platform that allows developers to build, deploy, and manage applications without worrying about the underlying infrastructure. It abstracts the hardware and software layers, providing a framework and tools needed to build applications (like databases, development tools, and middleware).	Examples are Google App Engine, Microsoft Azure App Service, Heroku, IBM Cloud Foundry. Use cases: 1. **Web App Development**: Developers can focus on writing code and logic for their web applications, while PaaS handles hosting, scaling, and security. 2. **Mobile App Backend**: You can use PaaS for hosting the backend services of a mobile app. For example, databases, user authentication, and other server-side logic. 3. **Enterprise Applications**: Businesses can use PaaS to quickly develop and deploy custom enterprise applications, taking advantage of built-in scalability and maintenance features.

SaaS (Software as a service)	Delivers software applications over the internet on a subscription or pay-per-use basis. With SaaS, users can access the software via web browsers without needing to install or maintain it on their own hardware. SaaS providers manage everything, including updates, security, and infrastructure. – No installation required on user devices – Accessible from any device with an internet connection – Automatic updates and patches – Scalable user licensing – Subscription-based pricing	Examples are: Google Worskspace, Salesforce, Slack, Dropbox, Zoom. User cases: 1. **Office Productivity**: Tools like Google Workspace and Microsoft 365 let users collaborate on documents, spreadsheets, and presentations in real-time from anywhere. 2. **Customer Relationship Management (CRM)**: Platforms like Salesforce help businesses manage customer interactions, sales pipelines, and customer service from a single interface, all hosted on the cloud. 3. **Communication & Collaboration**: SaaS platforms like Slack or Microsoft Teams provide seamless communication for teams, allowing instant messaging, file sharing, and video calls. 4. **Data Storage and Sharing**: Services like Dropbox or Google Drive offer cloud-based storage, allowing users to save, share, and access files from anywhere.

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 cloud computing environment where the services and infrastructure are provided by a third-party vendor over the public internet. These services are available to anyone who wants to purchase or lease them, typically on a pay-as-you-go model. – Resources like storage, computing power, and networking are shared with other organizations (multi-tenant). – Managed by the cloud service provider (e.g., AWS, Microsoft Azure, Google Cloud). – Highly scalable and cost-effective due to its shared nature. Appropriate for: – Small to medium-sized businesses or startups that need flexible, low-cost solutions without investing in their own hardware. – Organisations that don't have strict security or regulatory requirements that would require a private infrastructure. – Businesses looking for scalability and flexibility for growing or fluctuating workloads. For example, a web development company that needs scalable hosting for client websites might use Amazon Web Services (AWS) or Microsoft Azure to avoid the costs of maintaining physical servers, like Revolut a fintech company.
Private Cloud	A cloud environment dedicated to a single organisation. Unlike the public cloud, the infrastructure is either hosted on-premises or by a third-party provider, but the resources are not shared with other organisations.

	– Exclusive use by one organisation (single-tenant). – Can be hosted either in the company's data center (on-premises) or by a third-party vendor. – Provides more control over security, privacy, and customisation. Appropriate for: – Organisations with strict security or regulatory compliance needs, like financial institutions or healthcare companies. – Businesses that require complete control over their infrastructure. – Large enterprises that need the scalability of the cloud but want the security and control of their own infrastructure. For example, banks like HSBC/Barclays or healthcare organisations like the NHS that need to store sensitive customer data in compliance with GDPR regulations. These organisations could implement a private cloud to ensure data privacy and meet compliance requirements. For instance, a hospital might use a private cloud to securely store patient records and ensure they meet healthcare privacy standards.
Hybrid Cloud	Combines both private and public clouds, allowing data and applications to be shared between them. This enables businesses to move workloads between private and public clouds as needed for greater flexibility, optimisation, and scalability. – Offers the flexibility of using both on-premises infrastructure (private cloud) and external cloud resources (public cloud). – Helps organisations manage and scale workloads dynamically, using the public cloud for less-sensitive data while keeping more sensitive information on the private cloud. – Enables businesses to take advantage of cost-effective, scalable public cloud resources while

	retaining control over critical applications and data. Appropriate for: – Large enterprises or businesses with complex workloads that need to balance cost, security, and scalability. – Organisations that want to leverage the public cloud for certain workloads (e.g., testing, development) but need to keep sensitive data and mission-critical systems within a private cloud. – Businesses that need flexibility and want to be able to move workloads based on performance or cost considerations. For example, retail companies like ASDA or Tesco that use a hybrid cloud approach. These companies might use a public cloud to scale their website or e-commerce platform during peak shopping seasons (like Black Friday) but store customer and financial data in a private cloud to meet compliance standards and ensure security. This approach allows them to manage large traffic surges without compromising sensitive data security.
Community Cloud	A cloud computing environment shared by several organisations with common interests or requirements (e.g., compliance, security). The infrastructure is shared between these organisations, which could be from a specific industry, such as healthcare, government, or education. – Shared between several organisations that have common goals, such as similar regulatory or compliance needs. – It can be managed by a third-party provider or one of the organisations within the community. – Offers a balance between private and public clouds, providing shared resources while maintaining control over security and compliance. Appropriate for:

- When multiple organisations from the same industry or with similar needs want to pool resources for cost efficiency but need to ensure their cloud environment meets specific regulatory or security standards.
- Government agencies or educational institutions with shared needs for secure data processing or storage could benefit from community clouds.

For example, Government entities (like the UK Ministry of Defence) or academic institutions that require a cloud platform to collaborate and share resources securely while adhering to regulatory standards. For example, multiple universities could share a community cloud to store research data or manage student records, ensuring compliance with privacy laws while benefiting from shared infrastructure.

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 (Section 1)	This section criminalises accessing computer systems or data without permission. It covers situations where someone gains access to a	E.g. Hacking into someone's online account, such as an email or social media account, without their permission.

	computer or network without the consent of the owner or operator.	
Unauthorised Access with Intent to Commit or Facilitate the Commission of Further Offenses (Section 2)	This area extends to cases where unauthorised access is gained with the intention to commit additional crimes, such as theft, fraud, or espionage.	E.g. A hacker accessing a company's financial database to steal sensitive information or perform fraudulent transactions.
Unauthorised Modification of Computer Material (Section 3)	This section deals with unauthorized changes to computer systems or data, such as altering, deleting, or inserting data without consent. It includes actions that could harm or disrupt the normal functioning of a computer system.	E.g. Introducing malware or viruses to damage or disrupt a network, or altering data to falsify information for fraudulent purposes.

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
Expanded Powers for Law Enforcement to Seize Equipment (Section 37) - **The Police and Justice Act gives law enforcement expanded powers to seize computers, storage devices, and data that may be involved in or related to computer misuse offenses. This power aids in investigations, allowing**

authorities to take preventive actions against potential misuse and secure evidence. - **For example: If someone is suspected of possessing malware or hacking tools, the police can seize their computer or digital devices to examine them for evidence of criminal activity.**
Increase in Penalties for Offenses - **The Act increased the maximum penalties for certain offenses under the Computer Misuse Act 1990, particularly for unauthorised access with the intent to commit further offenses (Section 2) and for unauthorised modification of computer material (Section 3). These changes were designed to reflect the seriousness of cybercrimes and deter offenders.** - **For example: The maximum prison sentence for someone found guilty of unauthorised access with intent to commit further offenses was increased from 5 years to 10 years.**
Clarification of Unauthorized Access to Data and Access with Intent to Modify (Section 35) - **The Act clarifies that unauthorised access to data, whether or not it results in modification, is an offense. It also specifically criminalises gaining unauthorised access with the intent to modify data, providing more comprehensive coverage against computer misuse and ensuring clearer legal definitions.** - **For example: If a person hacks into a company's network and accesses sensitive data, even if they don't alter it, the act of unauthorised access alone is now considered a criminal offense. Additionally, accessing data with the intent to alter it (such as changing financial records) would be criminal.**

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.
Any accident connected to work
Any disciplinary action
Tax code

Give three more examples of data that an employer can only store if they first get the employee's permission.
Biometrics for example fingerprints for ID
Race and ethnicity
Religion

Conduct further research to answer the below questions.

Question	Answer
Provide one example of: Copyright infringement	**Using copyrighted music in a YouTube video without permission** If someone uploads a video to YouTube and includes a song that is copyrighted (without permission from the copyright holder), they are infringing on the music's copyright. Even if the video is not monetised, using the song without authorisation is a violation.
Provide one example of: Plagiarism	**Submitting a Research Paper Written by Someone Else** Turning in a research paper or essay that was written by another person, whether purchased online, written by a friend, or downloaded from a paper-sharing website, and claiming it as your own work.
What are two consequences of copyright infringement and software piracy?	1) Security risks such as malware and viruses: Pirated software often comes from unverified sources and may contain malware, spyware, or viruses. Using such software can expose systems to data breaches, loss of sensitive information, and additional security vulnerabilities. 2) Damage to Reputation: Both individuals and companies found guilty of copyright infringement or software piracy may suffer significant damage to their reputation. This can lead to loss of business, damaged

	relationships with clients, and negative publicity.
Give three possible consequences for individuals when using pirated software	1) **Fines and Penalties**: In many countries, using pirated software is a violation of copyright law, and individuals may face significant fines. These fines can range from hundreds to thousands of pounds depending on the severity of the infringement. 2) **Unreliable Software**: Pirated versions of software may be incomplete, unstable, or prone to crashes. This can disrupt work or personal activities and cause frustration due to poor performance. 3) **Criminal Prosecution**: In the UK, under the Copyright, Designs and Patents Act 1988, knowingly using or distributing pirated software can lead to criminal prosecution, with penalties including imprisonment and fines.

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

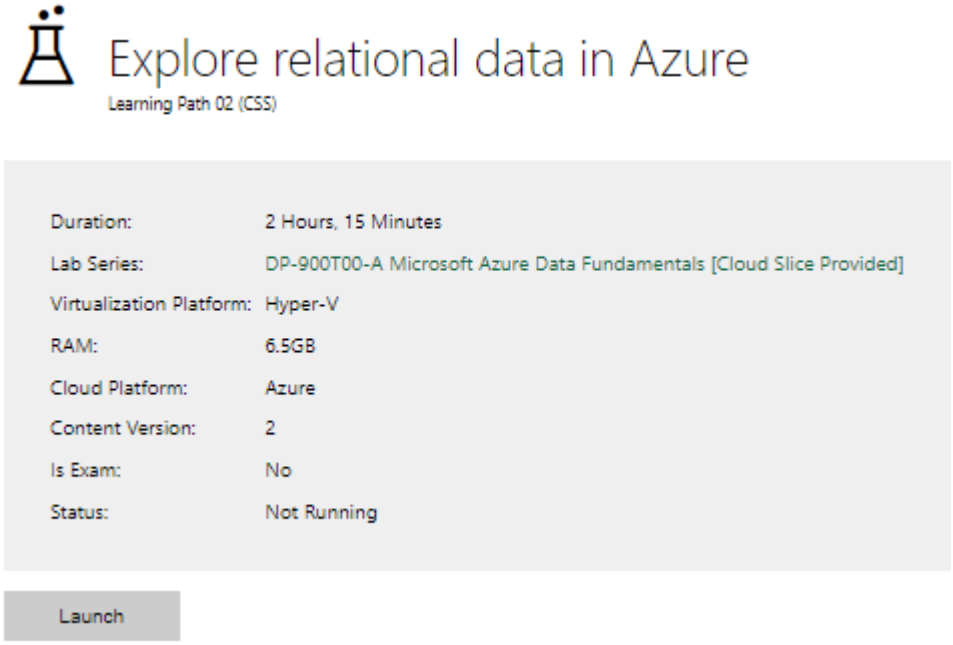
- Insert a number in the first column of each row to match each of the statements with one of the above Acts.

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

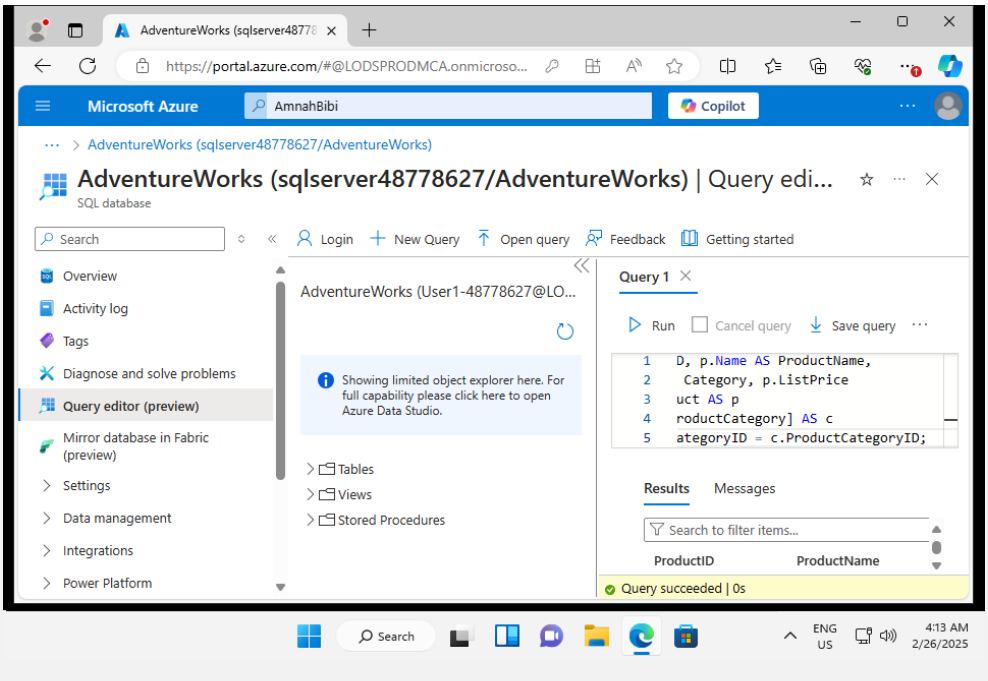
Act number	Clause
4	With some exceptions, it is illegal to use unlicensed software
7	Any product, digital or otherwise, must be fit for the purpose it is supplied for
1	Unauthorised modification of computer material is illegal
Not illegal	It is illegal to create or use a hacking tool for penetration 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.

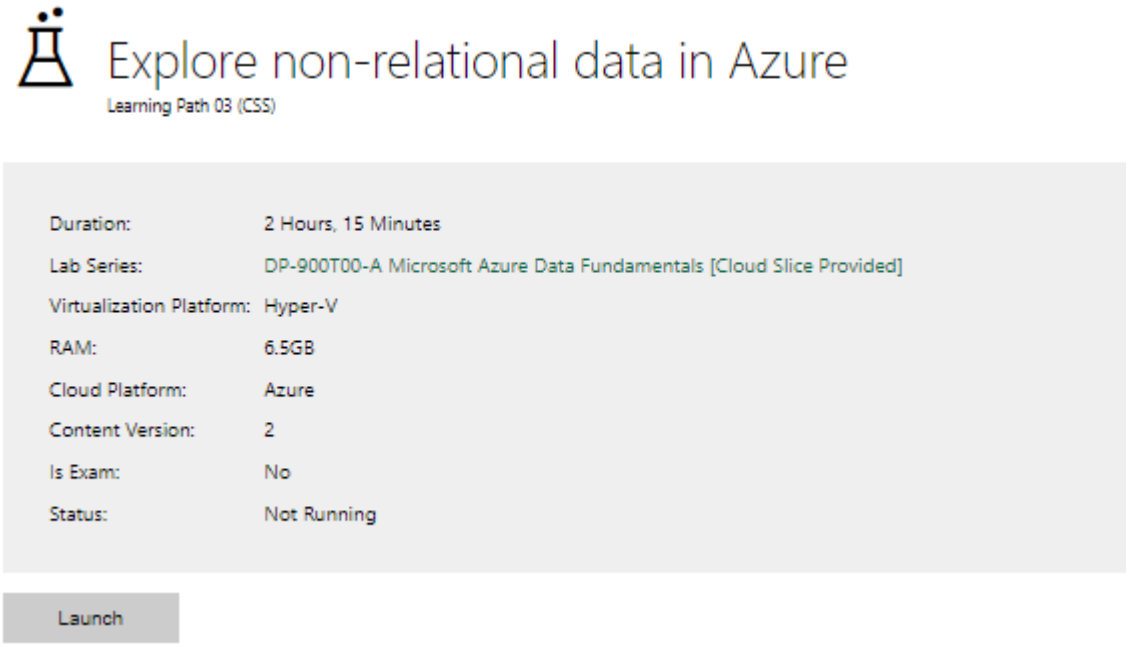


Completed lab

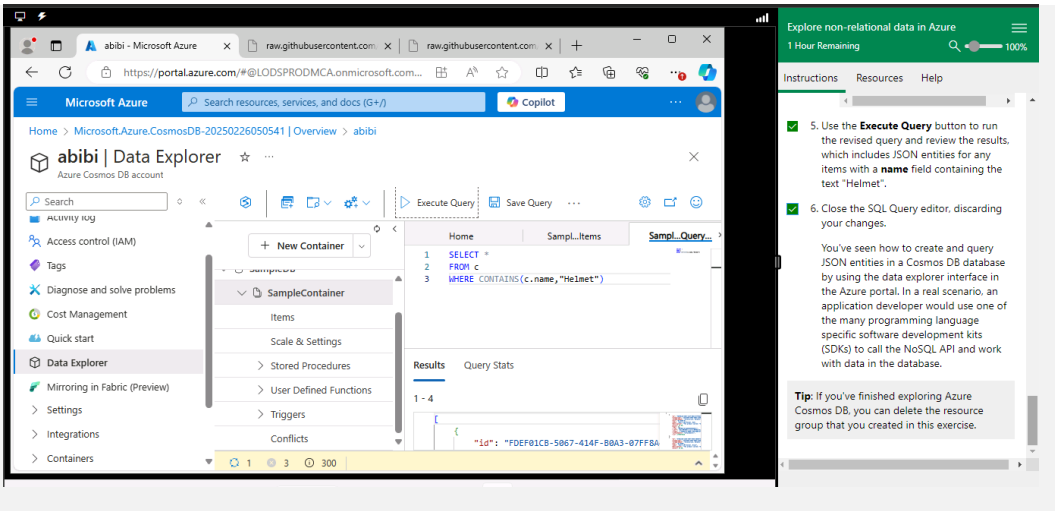


Day 3: Task 2

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

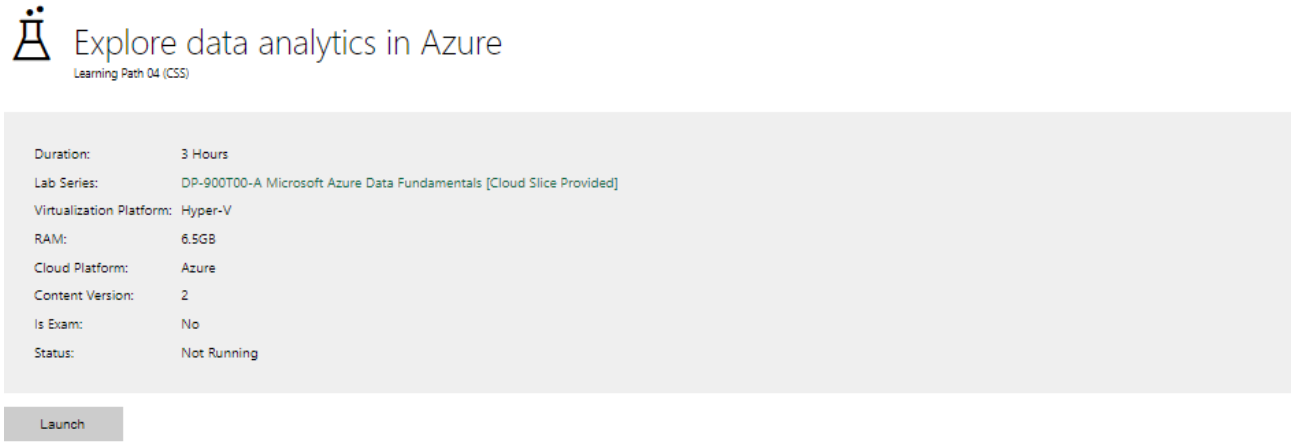


Completed lab



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.

6

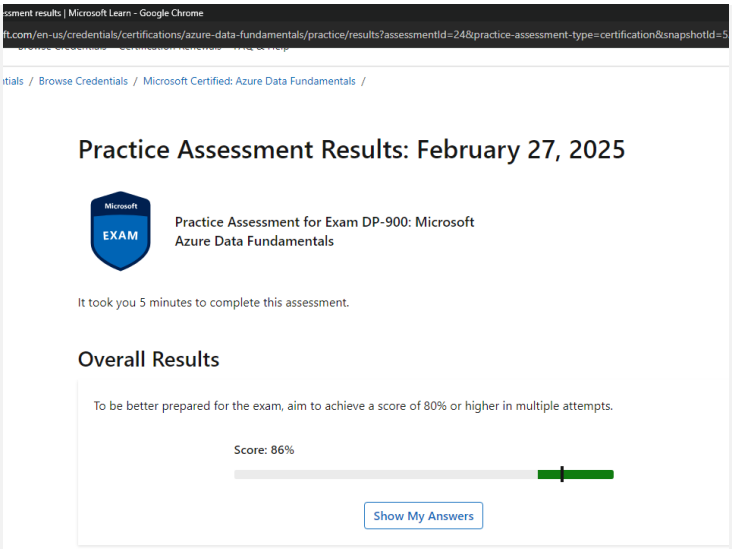
Practice Assessment: DP-900T00-A Microsoft Azure Data Fundamentals

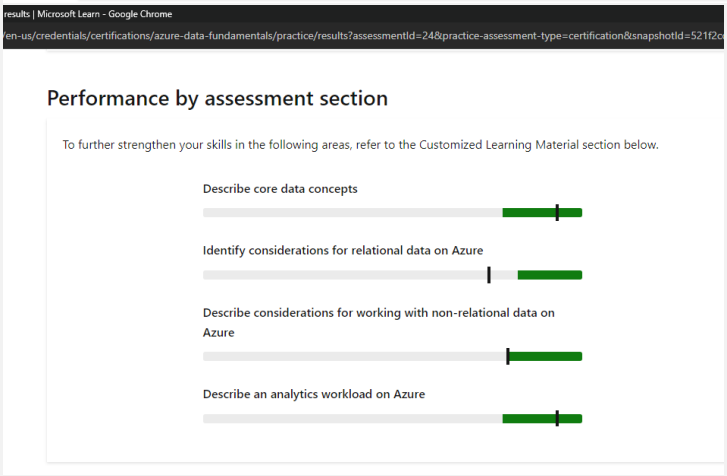
Practice Assessment for Microsoft Certifications for DP-900T00-A

Additional Details

Required:	No
Available Instructor-Led:	Yes
Available Self-Paced:	Yes

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.

Paws & Whiskers is a growing pet shop aiming to analyse sales, customer information, and inventory data to make data-driven decisions. They want to transition to Microsoft Azure for better data storage, analysis, and reporting, enabling them to make data-driven decisions.

Data Laws and Regulations:

GDPR Compliance - The General Data Protection Regulation (GDPR) is the European Union law designed to protect individuals' personal data. It affects any business handling data and although the UK has left the EU, the UK has incorporated GDPR into its national law through the UK GDPR.

- How this can be helpful for Paws & Whiskers:
    - ***Customer Data:*** All customer personal data (e.g., names, addresses, email, phone numbers) must be stored securely.
    - ***Data Processing:*** "Paws & Whiskers" needs to ensure transparency in how they process customer data and allow customers to access, correct, or delete their information.
    - ***Consent:*** They must obtain explicit consent from customers for data collection and processing.
    - ***Data Breach:*** Any data breach must be reported to the Information Commissioner's Office (ICO) within 72 hours if it affects personal data.
    - ***Data Minimisation***: The shop should only collect data necessary for the purpose and limit the retention period.

Data Protection Act (DPA) 2018 - The Data Protection Act (DPA) 2018 is the UK's implementation of the GDPR. It outlines how personal data must be processed, stored, and protected in the UK.

- How this can be helpful for Paws & Whiskers:
    - ***Security***: Customer data must be stored securely, and appropriate technical measures such as encryption and access controls should be applied to protect the data from unauthorised access.
    - ***Rights of Individuals:*** Customers have the right to request access to the data held about them, the right to rectify inaccuracies, and the right to request the erasure of their data.
    - ***Data Transfers:*** If Paws & Whiskers transfers customer data outside the UK, it must ensure that the data is protected to the same standard as required by the UK GDPR, including using Azure services with appropriate compliance certifications.

Other Industry Standards -

- ***PCI-DSS (Payment Card Industry Data Security Standard)***:
    - If Paws & Whiskers accepts card payments, it must comply with PCI-DSS to protect customers' payment information. Azure offers services like Azure Key

Vault to store encryption keys and Azure Security Centre to ensure compliance with security standards.

- ***ISO 27001***:
  - This standard helps organisations implement robust information security management systems. As Paws & Whiskers stores sensitive customer and payment data, using a cloud service like Azure, which is ISO 27001 certified, ensures compliance with this industry-standard framework.

Azure Service Recommendations:

Storage Option	Best For	Benefits	Key Considerations
Azure Blob Storage	Unstructured data (e.g., logs, large files)	Scalable, cost-effective	Limited querying capabilities
Azure SQL Database	Structured data (e.g., customer data)	Strong relational model, SQL support	More complex, costlier than Blob
Azure Synapse Analytics	Big data analysis and data warehousing	Advanced analytics, integration	Requires higher expertise for setup

Data Storage -

- ***Azure SQL Database***:
  A fully managed relational database service that can securely store customer data, sales transactions, and inventory information. It supports built-in security features, including Transparent Data Encryption (TDE), to ensure data protection.

- ***Azure Blob Storage***:
  Ideal for storing unstructured data such as images of pets, large product catalogs, and backups. Blob Storage can also store logs, sales reports, and bulk inventory data in a scalable, flexible and cost-effective manner.

Data Analysis Tools -

- ***Azure Machine Learning***:
  Azure ML allows for advanced analysis of customer behaviour and purchasing trends. Paws & Whiskers can build predictive models to forecast sales, personalise customer recommendations, and optimise inventory management.

- ***Azure Synapse Analytics***:
  Azure Synapse combines big data and data warehousing solutions. It can analyse large datasets efficiently, enabling Paws & Whiskers to examine sales patterns, inventory trends, and customer behaviour to inform business decisions, in real time.

Data Integration and Automation -

- ***Azure Data Factory***:
  Azure Data Factory automates the ETL (Extract, Transform, Load) process, making it easier to integrate data from multiple sources. It can automate the collection of sales data, customer information, and inventory levels, improving overall efficiency and reducing manual data entry errors.

Data Types and Data Modelling:

Data Categories -

- ***Customer Data**:* Demographic information (e.g., address, age, name, gender), contact information (e.g., email, phone number), and transaction history.
- ***Transaction Data***: Sales details (e.g., product purchased, price, amounts, payment method), transaction dates, and discounts applied.
- ***Inventory Data**:* Product details (e.g., pet types, category, quantity in stock, price), suppliers, and restocking details.
- ***Product Categories**:* Types of products sold, such as pet food, grooming items, pet accessories, and toys.

Data Modelling Approach -

- ***Relational Model**:*
  - **Entities**: Key entities include Customer, Sales, Products, and Inventory.
  - **Relationships**: A Customer may place multiple Sales (one-to-many relationship), a Product belongs to a Product Category (many-to-one relationship), and Inventory tracks stock levels.
  - **Primary Keys**: Each entity should have a unique identifier, such as CustomerID, SalesID, ProductID, and InventoryID.
  - **Normalisation**: The data should be normalised to reduce redundancy. For example, customer data should be in a Customer table, while sales data should be stored separately in a Sales table linked via CustomerID.

Data Storage Formats and Structures in Azure:

Data Formats -

- ***CSV***: Suitable for raw data imports, such as bulk customer or transaction data. It's simple, but not ideal for storing complex or large datasets.
- ***JSON***: Ideal for structured data that requires flexibility, such as storing customer preferences or product details in a hierarchical format.
- ***Parquet***: A columnar storage format suitable for large-scale data analytics. It provides efficient storage and faster query performance, making it ideal for use with tools like Azure Synapse Analytics for sales trend analysis.

Data Security and Encryption -

- ***Azure Encryption**:* Use Transparent Data Encryption (TDE) for Azure SQL Databases to automatically encrypt data at rest. For data in transit, use SSL/TLS encryption to ensure secure communication.
- ***Azure Active Directory (AAD)**:* Implement Role-Based Access Control (RBAC) to ensure that only authorised personnel can access or modify sensitive data.

- ***Azure Key Vault***: Used to manage and store sensitive keys and secrets.
- ***Data Masking***: Sensitive customer data, like credit card numbers, can be masked to provide limited visibility to unauthorized users.

Additional Considerations:

Backup and Disaster Recovery -

- ***Azure Backup:*** Regular, automated backups of databases and other critical data should be scheduled to ensure business continuity. Azure Backup ensures that even if data is lost or corrupted, it can be restored from a previous backup.
- ***Azure Site Recovery:*** For disaster recovery, Azure Site Recovery can replicate virtual machines and applications to secondary regions, ensuring minimal downtime in the event of a disaster.

Data Visualisation -

- ***Power BI:***
  Enables management to create interactive dashboards and reports. For example, real-time sales trends, customer purchase behaviour, and inventory levels can be displayed in easy-to-read visuals to support decision-making.

Future Scalability -

- *Scalable Solutions:* Azure services such as Azure SQL Database and Azure Synapse Analytics support auto-scaling to handle growing data volumes. Azure Data Lake can also handle big data workloads efficiently and Azure Machine Learning can scale to support more complex models as the business needs evolve. As Paws & Whiskers expands, these services can scale without the need for additional infrastructure.

References:

- **General Data Protection Regulation (GDPR)**
  https://en.wikipedia.org/wiki/General_Data_Protection_Regulation
- **UK Data Protection Act (DPA) 2018**
  https://www.gov.uk/data-protection
- **PCI-DSS Compliance**
- **ISO 27001** Certification
- **Microsoft Azure Documentation**
  https://learn.microsoft.com/en-us/azure/?product=popular

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.