

Data Technician

Name:

Course Date:

Table of contents

Day 1: Task 1.....	3
Day 1: Task 2.....	10
Day 1: Task 3.....	13
Day 2: Task 1.....	15
Day 3: Task 1.....	20
Day 3: Task 2.....	21
Day 3: Task 3.....	25
Day 4: Task 1.....	30
Day 4: Task 2.....	32
1. Scenario Background	32
2. Data Laws and Regulations.....	32
3. Azure Service Recommendations	32
4. Data Types and Data Modelling.....	32
5. Data Storage Formats and Structures in Azure	33
6. Additional Considerations	33
Submission Guidelines:.....	33
Course Notes.....	38
Additional Information.....	46



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.

<p>What can cloud computing do for us in the real-world?</p>	<ul style="list-style-type: none">❑ Store and Access Data from Anywhere You can save files, photos, or documents online (in the cloud) and access them anytime, from any device. Google Drive, Dropbox, iCloud. Real-world use: Students can access their homework from school and home; ❑ Run Applications Without Installing Them Apps like Microsoft 365, Zoom, and Netflix run in the cloud — no need to install heavy software or buy powerful computers. ❑ Enable Remote Work and Collaboration Cloud platforms allow teams to work together in real time from different locations ❑ Scale Business Resources Easily Companies can quickly increase or reduce their computing power without buying physical hardware. Online stores like Amazon or eBay handle more traffic during sales using cloud servers that scale up automatically. ❑ Secure Backup and Disaster Recovery Cloud storage can automatically back up data and restore it in case of hardware failure or cyberattack.
---	---



	<p>Hospitals, banks, and schools rely on cloud backups to protect important records.</p> <p> Support Artificial Intelligence and Big Data</p> <p>Cloud computing powers AI tools, data analytics, and machine learning by providing the needed storage and processing power.</p> <p>Companies use cloud-based AI to analyze customer behavior, recommend products, or detect fraud.</p>
<p>How can it benefit a business?</p>	<ol style="list-style-type: none"> 1. Cost Savings No need to buy expensive servers or hardware. Pay only for what you use (like a utility bill). Example: A startup can launch its app without buying servers — just rent space from AWS or Azure. 2. Flexibility and Scalability Easily increase or reduce computing power depending on demand. Supports growth without delays. Example: An online shop can handle Black Friday traffic spikes by scaling up its cloud resources temporarily. 3. Remote Work and Collaboration Employees can access files and systems from anywhere. Teams can work together in real-time using tools like Google Workspace or Microsoft 365. Example: A marketing team in London and New York can work on the same presentation at the same time. 4. Data Backup and Recovery Automatic backups protect against data loss.



Quick recovery in case of accidents or cyberattacks.

Example: A company hit by ransomware can restore clean data from the cloud.

5. Faster Deployment Set up software and systems in minutes, not days or weeks.

Launch new projects quickly without waiting for physical equipment.

6. Security Leading cloud providers offer high-level security tools and updates.

Protects data better than many in-house systems.

Example: A small business gets enterprise-grade cybersecurity from using cloud platforms.

7. Automatic Updates Software and security updates happen automatically, reducing IT workload.
8. Sustainability Cloud data centers use energy more efficiently than traditional office servers, reducing a business's carbon footprint.

Benefit	How It Helps
Saves money	No need for costly hardware
Improves flexibility	Easy to scale resources
Enables remote work	Access from anywhere
Protects data	Backup & disaster recovery
Boosts speed	Launch and update projects faster

What's the alternative to cloud computing?

1. On-Premises Computing (Traditional IT Infrastructure)

In this model, a company buys and manages its **own servers and hardware** in its building (on-site).



Characteristics:

- Data stored locally (not online)
- IT team maintains everything: servers, updates, security
- Higher upfront cost for hardware and software
- More control, but less flexible

Example:

A hospital or bank may choose on-premises systems for full control over sensitive data.

2. Hybrid Cloud

A mix of **cloud and on-premises** systems — businesses use the cloud for some services and keep sensitive data in-house.

Characteristics:

- Offers flexibility and control
- Often used during cloud migration or for compliance reasons

Example:

A company might store client data on local servers but use the cloud for email or collaboration tools.

3. Private Cloud

A **cloud-like system** that's built specifically for one organization, either on-premises or hosted by a provider.





Characteristics:

- More secure and customizable than public cloud
- More expensive, but good for companies with strict security needs

Comparison Table

Feature	Cloud Computing	On-Premises	Hybrid/Private Cloud
Cost	Low upfront, pay-as-you-go	High upfront (servers)	Medium
Flexibility	Very high	Limited	High
Control	Less (managed by provider)	Full control	Mixed
Security	High (shared)	High (local)	Highest (custom)
Maintenance	Provider handles it	Handled in-house	Shared

□ □ 1. Amazon Web Services (AWS)

❑ Largest and most widely used cloud platform

Key Features:

- **Compute:** EC2 (virtual machines), Lambda (serverless)
- **Storage:** S3 (object storage), EBS (block storage)
- **Databases:** RDS, DynamoDB
- **AI & Machine Learning:** Amazon SageMaker
- **Scalability & Reliability**
- Global data centers (availability zones)

Used by:

Netflix, NASA, Airbnb, and many startups

□ □ 2. Microsoft Azure

❑ Strong integration with Microsoft tools (e.g., Office, Windows)

Key Features:

What cloud providers can we use, what are their features and functions?



- **Virtual Machines (VMs)**
- **Azure SQL Database**
- **Azure Active Directory** (for identity and access)
- **Azure AI & Cognitive Services**
- **Hybrid cloud support**

Used by:

Government, healthcare, finance, and enterprises that already use Microsoft products

□ □ 3. Google Cloud Platform (GCP)

☒ Known for data analytics and AI/ML tools

Key Features:

- **Compute Engine, App Engine**
- **BigQuery** (data warehouse)
- **Cloud Storage**
- **TensorFlow and AI Platform**
- Strong support for Kubernetes (GKE)

Used by:

Spotify, Snapchat, and companies doing advanced analytics or machine learning

🕒 Cloud Providers Comparison Table

Cloud Provider	Key Features	Advantages	Disadvantages
AWS	- EC2 (virtual servers) - S3 (object storage) - Lambda (serverless) - RDS (database)	✓ Largest market share ✓ Wide range of services ✓ Global infrastructure	✗ Complex pricing ✗ Can be overwhelming for beginners
Microsoft Azure	- Azure Virtual Machines - Azure Blob Storage - Azure SQL Database - Active Directory	✓ Strong integration with Microsoft tools ✓ Good hybrid cloud support	✗ Less intuitive interface ✗ Some services not as mature as AWS
Google Cloud (GCP)	- Compute Engine - BigQuery (analytics) - Cloud Storage - Kubernetes Engine (GKE)	✓ Excellent for data analytics & AI ✓ Strong Kubernetes support	✗ Smaller global reach ✗ Fewer enterprise clients than AWS/Azure



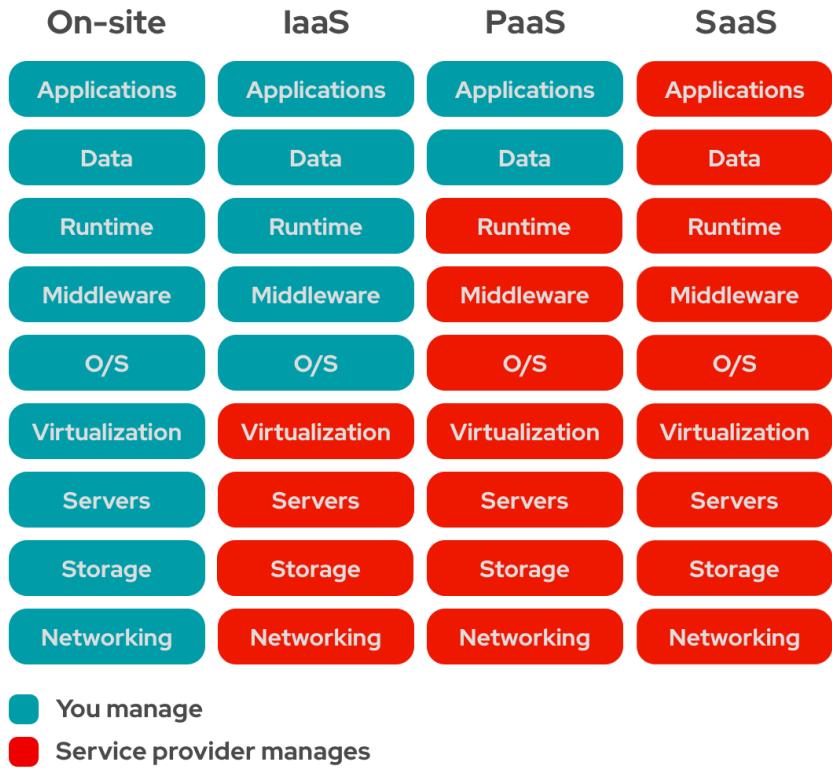


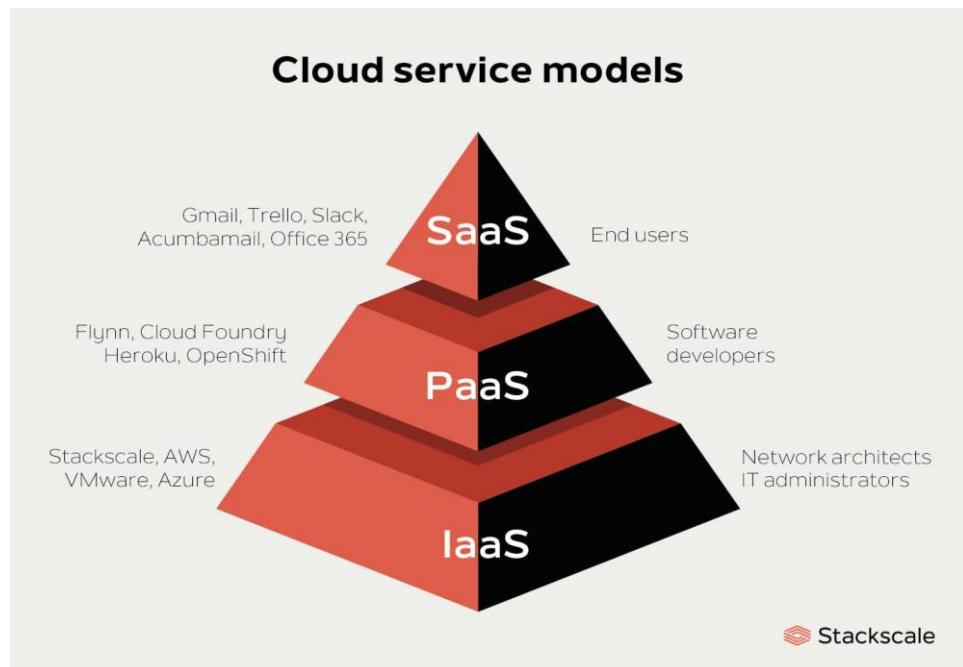
Day 1: Task 2

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 virtual hardware : servers, storage, networking. You manage OS, apps, and data.	<i>Example:</i> A web developer sets up a custom server on AWS EC2 to host a website. Microsoft Azure VMs Google Compute Engine
PaaS (Platform as a service)	ready-to-use platform with tools for building, testing, and deploying apps. You manage your code; the provider manages the rest. Virtual machines with pre-installed operating systems and software/database environments	<i>Example:</i> A startup uses Azure App Service to deploy an app without worrying about infrastructure.
SaaS (Software as a service)	Delivers ready-to-use software over the internet . No installation or maintenance needed	<i>Example:</i> A company uses Google Workspace or Microsoft 365 for email, docs, and collaboration. <ul style="list-style-type: none">• Dropbox• Zoom• Slack• Canva







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	<p>Cloud services are offered to everyone over the internet by providers like AWS, Azure, or GCP.</p> <p>When It's Appropriate</p> <ul style="list-style-type: none"><input checked="" type="checkbox"/> General-purpose computing<input checked="" type="checkbox"/> Startups, testing, scalability <p>Real-World Example</p> <ul style="list-style-type: none">Startups hosting mobile or web apps on AWS/GCPE-commerce stores using Shopify or BigCommerce (SaaS on public cloud)Universities using Microsoft 365 for students
Private Cloud	<p>Cloud infrastructure is used by one organisation only, either on-site or hosted.</p> <ul style="list-style-type: none"><input checked="" type="checkbox"/> For sensitive data or strict security needs<input checked="" type="checkbox"/> Full control <p>Real-World Example</p> <ul style="list-style-type: none"><input checked="" type="checkbox"/> Banks securing customer transactions<input checked="" type="checkbox"/> Pharmaceutical companies protecting research data<input checked="" type="checkbox"/> Government departments storing sensitive records
Hybrid Cloud	<p>A mix of public and private clouds working together with shared data and apps.</p> <ul style="list-style-type: none"><input checked="" type="checkbox"/> Balancing cost & control<input checked="" type="checkbox"/> Gradual migration to the cloud <p>Real-World Example</p>



	<p>Netflix:</p> <p>Netflix uses a hybrid cloud approach for content storage, content delivery, application development, and big data analytics. They store their extensive library of movies and TV shows on AWS public cloud servers for scalability and cost-effectiveness. They also use a private cloud for sensitive customer information like payment details and device information.</p> <p>Healthcare:</p> <p>Healthcare organizations can use a hybrid cloud to store patient data securely on a private cloud for privacy and regulatory compliance, while leveraging the public cloud for data analysis and research.</p> <p>Finance:</p> <p>Financial institutions may use a private cloud for core banking applications and a public cloud for non-production use and data analytics.</p>
Community Cloud	<p>Shared cloud built for organisations with similar requirements or regulations.</p> <ul style="list-style-type: none"> ❑ Shared missions, budgets, compliance goals <p>Real-World Example</p> <p>Healthcare:</p> <p>Sharing patient data:</p> <p>Hospitals and laboratories frequently share patient records, making community clouds ideal for secure storage and transmission of sensitive information.</p> <p>Government:</p> <p>Secure communication:</p> <p>Government agencies can use community clouds to communicate securely between departments, such as the U.S. Department of Defense leveraging community cloud models.</p>



Education:
Online learning:
During the pandemic, schools and universities transitioned to community cloud platforms to support online learning, sharing resources and reducing costs.

Day 2: Task 1

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

Area	Description	Example
Unauthorized access to computer material	Gaining access to a computer system or data without permission. This is commonly known as hacking.	A person guessing someone else's password and logging into their account without consent.
Unauthorized access with intent to commit further offense	Accessing a system with the intention of committing additional crimes, such as fraud or theft.	A hacker breaking into a bank's system to steal customer financial data and commit fraud.
Unauthorized modification of computer material	Altering, deleting, or introducing harmful software into a system without permission. This includes spreading viruses or malware.	A person creating and distributing a virus that corrupts files on other people's computers.

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
1. Increased Penalties for Unauthorized Access – The maximum sentence for hacking was doubled from five years to ten years, making penalties more severe for offenders.
2. Criminalization of Denial-of-Service (DoS) Attacks – The Act explicitly made DoS attacks illegal, addressing a previous legal loophole where such attacks were not clearly classified as offenses.
3. Ban on the Creation and Distribution of Hacking Tools – It became an offense to make, supply, or obtain tools intended for committing cybercrimes, ensuring that malicious software and hacking utilities could not be legally distributed.

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.

1. Personal Information – This includes details like full name, date of birth, address, and contact information.
2. Employment Records – Information such as job title, salary, work history, performance reviews, and disciplinary records.
3. Payroll and Financial Data – Bank account details for salary payments, tax records, pension contributions, and benefits information.

Give three more examples of data that an employer can only store if they first get



the employee's permission.

1. Biometric Data – Information like fingerprints, facial recognition, or retina scans used for security or attendance systems.
2. Medical Records – Any health-related information, including disability status, sickness reports, or vaccination records.
3. Political or Religious Beliefs – Details about an employee's affiliations, which could be relevant for specific accommodations or rights but must be stored only with permission.

Conduct further research to answer the below questions.

Question	Answer
Provide one example of: Copyright infringement	Illegally downloading and sharing movies without permission from the copyright holder. For instance, if someone uploads a newly released film to a torrent site and distributes it, this violates copyright law, as the creators and producers are not compensated for their work.
Provide one example of: Plagiarism	Copying and submitting someone else's research paper as your own without proper attribution. For instance, if a student takes an academic article from a journal, changes a few words, and presents it as their own work, they are committing plagiarism.
What are two consequences of copyright infringement and software piracy?	1. Legal Penalties – Violators may face fines or lawsuits from copyright holders. In severe cases, individuals or businesses could be



	<p>prosecuted and sentenced to imprisonment.</p> <p>2. Financial Losses – Companies and creators suffer revenue loss due to unauthorized distribution. This can impact businesses, reduce funding for innovation, and even lead to job losses in affected industries.</p> <p>Protecting intellectual property rights ensures fair compensation for creators and encourages continued development in software and digital content. Let me know if you need more details!</p>
Give three possible consequences for individuals when using pirated software	<ol style="list-style-type: none"> 1. Legal Risks – Piracy is illegal, and individuals may face fines or legal action for using unauthorized software, especially if caught distributing it. 2. Security Threats – Pirated software often lacks security updates and may contain viruses or malware, putting personal data at risk of theft or corruption. 3. Lack of Support and Functionality – Unauthorized software doesn't receive official updates, leading to compatibility issues, bugs, and loss of essential features over time.

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
2	It is illegal to prevent or hinder access (e.g. by a denial-of-service attack) to any program or data held in any computer
6	Personal data must be accurate and where necessary kept up to date

Day 3: Task 1

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

Explore relational data in Azure

Learning Path 02 (CSS)

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

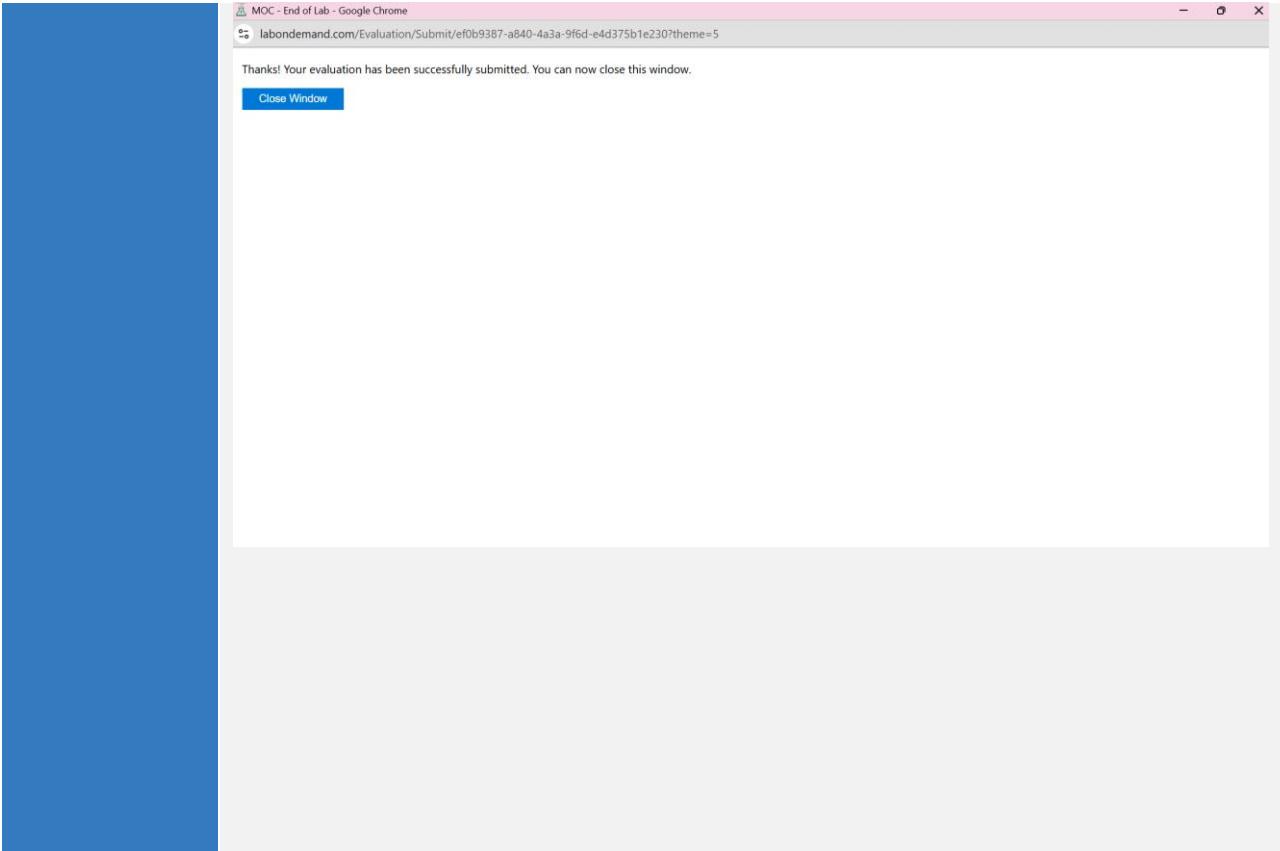
The screenshot shows the Microsoft Azure portal interface. On the left, there's a sidebar with 'Completed lab' text. The main area has a title 'Explore Azure SQL Database' and a sub-section '29 Minutes Remaining'. It includes 'Instructions', 'Resources', and 'Help' tabs. A checked checkbox says '14. Now try the following query, which uses a JOIN to get the category name from the SalesLT.ProductCategory table:' followed by a query code block. Another checkbox '15. Close the query editor pane, discarding your edits.' is also present. The central part of the screen shows a browser window with the URL 'https://portal.azure.com/#@LODSPRODMCA.onmicrosoft.com'. The browser title is 'AdventureWorks (sqlserver52120902/AdventureWorks) | Query editor (preview)'. The query editor pane contains a query to join Product and ProductCategory tables:

```

1 SELECT p.ProductID, p.Name AS ProductName,
2      c.Name AS Category, p.ListPrice
3 FROM SalesLT.Product AS p
4 JOIN [SalesLT].[ProductCategory] AS c
5 ON p.ProductCategoryID = c.ProductCategoryID;
    
```

The results pane shows a table with columns 'ProductID', 'ProductName', and 'Category'.





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.





Explore non-relational data in Azure

Learning Path 03 (CSS)

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

Explore Azure Storage

Explore blob storage

The screenshot shows a Microsoft Azure Storage blob container named 'data'. The container has one blob named 'product_data'. The blob details show its name, modified date, access tier, and archive status. The blob is virtual, as indicated by the 'product_data' name.

Instructions: 1. In the 'Upload to folder' box, enter 'product_data' and select the 'Upload' button. 2. Close the 'Upload blob' panel if it's still open, and verify that a 'product_data' virtual folder has been created in the 'data' container. 3. Select the 'product_data' folder and verify that it contains the 'product.json' blob you uploaded. 4. On the left side, in the 'Data storage' section, select 'Containers'. 5. Open the 'data' container, and verify that the 'product_data' folder you created is listed. 6. Select the '...' icon at the right-end of the folder, and note that it doesn't display any options. Folders in a flat namespace blob container are virtual, and can't be managed. 8. Use the 'X' icon at the top right in the 'data' page to close the page and return to the 'Containers' page.

Completed
lab

Explore Azure Data Lake Storage Gen2



Explore non-relational data in Azure - Google Chrome

data - Microsoft Azure

https://portalazure.com/#view/Microsoft_Azure_Storage/Containers

Microsoft Search resources, services, and docs (G+)

User1-S2 (39665@LODS...@LODS-PROD-MCA (LODSPROD.M...))

Home > week5task2 | Containers > data

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Authentication method: Access key (Switch to Microsoft Entra user account)

Location: data / product_data

Search blobs by prefix (case-sensitive)

Add filter

Name	Modified	Access tier	Archive status
product.json	6/11/2025, 1:05:00 PM	Hot (Inferred)	
product2.json	6/11/2025, 1:12:23 PM	Hot (Inferred)	

Add or remove favorites by pressing Ctrl+Shift+F

Search

ENG US 1:13 PM 6/11/2025

Instructions Resources Help 100%

1 Hr 50 Min Remaining

Explore non-relational data in Azure

7. In the Upload blob panel, select the product2.json file you saved on your local computer. Then select the Upload button.

8. Close the Upload blob panel if it's still open, and verify that a product_data folder now contains the product2.json file.

9. On the left side, in the Data storage section, select Containers.

10. Open the data container, and verify that the product_data folder you created is listed.

11. Select the ... icon at the right-end of the folder, and note that with hierarchical namespace enabled, you can perform configuration tasks at the folder-level, including renaming folders and setting permissions.

12. Use the X icon at the top right in the data page to close the page and return to the Containers page.

Explore Azure Files

Azure Files provides a way to create cloud-based file shares.

Next >

Explore Azure Files

Explore non-relational data in Azure - Google Chrome

Connect - Microsoft Azure

https://portalazure.com/#view/Microsoft_Azure_FileStorage/FileShares

Microsoft Search resources, services, and docs (G+)

User1-S2 (39665@LODS...@LODS-PROD-MCA (LODSPROD.M...))

Home > week5task2 | File shares > New file share > files

files

SMB File share

Search

Connect

Enable Backup

Overview

Diagnose and solve problems

Access Control (IAM)

Browse

Operations

Secure transfer required: 'Secure transfer required' is enabled on the storage account. SMB clients connecting to this share must support SMB protocol version 3 or higher in order to handle the encryption requirement. Click here to learn more.

Windows Linux macOS

To connect to this Azure file share from Windows, choose from the following authentication methods and run the PowerShell commands from a normal (not elevated) PowerShell terminal:

Drive letter: Z

Authentication method:

Give feedback

Instructions Resources Help 100%

1 Hr 46 Min Remaining

Azure Files provides a way to create cloud-based file shares.

1. In the Azure portal page for your storage container, on the left side, in the Data storage section, select File shares.

2. In the File shares page, select + File share and add a new file share named files, using the Transaction optimized tier.

3. Select Next: Backup > and disable backup. Then select Review + create.

4. In the File shares, open your new files share.

5. At the top of the page, select Connect. Then in the Connect pane, note that there are tabs for common operating systems (Windows, Linux, and macOS) that contain scripts you can run to connect to the shared folder from a client computer.

6. Close the Connect pane and then close the files page to return to the File shares page for your Azure storage account.

Explore Azure Tables

Next >

Explore Azure Tables



Explore non-relational data in Azure - Google Chrome

labclient.labondemand.com/LabClient/4987d443-9fd6-4d43-bcb2-37dece0543ff

week5task2 - Microsoft

week5task2 | Storage browser

Storage account

Home > week5task2

Add entity Refresh Delete Edit columns

tables > products

Authentication method: Access key (Switch to Microsoft Entra user account)

Add filter Advanced filters

Showng all 2 items

Timestamp	Name	Price	Discontinued
2025-06-11T20:20:02.52...		true	
2025-06-11T20:19:25.18...	Widget	2.99	

View all

Next >

Explore Azure Cosmos DB

View and create items

Explore non-relational data in Azure - Google Chrome

labclient.labondemand.com/LabClient/4987d443-9fd6-4d43-bcb2-37dece0543ff

tsemenchuk-db-task-2 - Microsoft

tsemenchuk-db-task-2 | Data Explorer

Azure Cosmos DB account

Home > Microsoft.Azure.CosmosDB-20250611132557 | Overview > tsemenchuk-db-task-2

New Item Update Discard Upload Item

SELECT * FROM c

Type a query predicate (e.g., WHERE cid='1'), or choose one from the drop down list, or leave empty to query all documents.

Apply Filter

id	/categoryId
4E4B38CB-0D82-43E5-89A8-...	75BF1ACB-168D-469C-9AA3-1...
5B5E90B8-FE42-4D6C-B728-E...	75BF1ACB-168D-469C-9AA3-1...
B267655B-A7C1-41E3-9682-...	AA28AE74-D57C-4B23-B5F7-F...
5089E32E-8A60-4117-AA98-5...	86F3CBAB-97A7-4D01-BABB-...
0A7E57DA-C73F-467F-954F-1...	4F34E180-384D-42FC-AC10-F...
1A176FD8-D9A8-488B-BDD9-...	86F3CBAB-97A7-4D01-BABB-...

Load more

0 0 0 298 ▲ 0 |

Instructions Resources Help 100%

With a unique id and other properties.

2. Select any of the items in the list to see a JSON representation of the item data.

3. At the top of the page, select **New Item** to create a new blank item.

4. Modify the JSON for the new item as follows, and then select **Save**.

```
json
T {
  "name": "Road Helmet,45",
  "id": "123456789",
  "categoryID": "123456789",
  "SKU": "AB-1234-56",
  "description": "The product called \"Road Helmet",
  "price": 48.74,
  "_rid": "OTxKAO7Hyg==/docs/0",
  "_self": "dbs/OTxKAO7Hyg==/colls/OTxKAO7Hyg==/docs/0",
  "_etag": "1*ec0036f8-0000-1100-0000-6849e8f00000",
  "_attachments": "attachments",
  "_ts": 1749674224
}
```

5. After saving the new item, notice that additional metadata properties are added automatically.

Query the database

Previous End >

Query the database



The screenshot shows the Azure Data Explorer interface. On the left, there's a sidebar with options like Quick start, Data Explorer (which is selected), Mirroring in Fabric (Preview), Resource visualizer, Settings, Integrations, Containers, Monitoring, Automation, and Help. The main area has tabs for Home, Connect, Sample_1_items, and Sample_Query (which is currently selected). A SQL query is written in the Sample_Query tab:

```
1 SELECT *
2 FROM c
3 WHERE CONTAINS(c.name, "Helmet")
```

The Results tab displays the query results as a table:

	id	category	SKU	name	description	price
1	"47ED1C1B-C285-4507-94E8-3860A744B261"	"4A1A05D-59EA-48E3-A1B9-67807778380E"	"HL-US09"	"Sport-100 Helmet, Black"	"The product called \"Sport-100 Helmet, Black\"."	48.74
2	"47ED1C1B-C285-4507-94E8-3860A744B261"	"4A1A05D-59EA-48E3-A1B9-67807778380E"	"HL-US09"	"Sport-100 Helmet, Black"	"The product called \"Sport-100 Helmet, Black\"."	48.74
3	"47ED1C1B-C285-4507-94E8-3860A744B261"	"4A1A05D-59EA-48E3-A1B9-67807778380E"	"HL-US09"	"Sport-100 Helmet, Black"	"The product called \"Sport-100 Helmet, Black\"."	48.74
4	"47ED1C1B-C285-4507-94E8-3860A744B261"	"4A1A05D-59EA-48E3-A1B9-67807778380E"	"HL-US09"	"Sport-100 Helmet, Black"	"The product called \"Sport-100 Helmet, Black\"."	48.74

To the right of the main interface, there's a sidebar titled "Explore non-relational data in Azure" with a progress bar showing "1 Hr 13 Min Remaining". It contains instructions and a "Query the database" section with numbered steps:

1. In the Data Explorer page, select the New SQL Query icon.
2. In the SQL Query editor, review the default query `SELECT * FROM c` and use the Execute Query button to run it.
3. Review the results, which includes the full JSON representation of all items.
4. Modify the query as follows:

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.





Explore data analytics in Azure

Learning Path 04 (CSS)

Duration:	3 Hours
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

Explore data analytics in Microsoft Fabric

My settings are in Ukrainian, so you might see some labels in Ukrainian. I hope that's not a problem:)

Create a workspace

Completed
lab



Головна сторінка

Провідник

Пошук у схемах і таблицях

taxi_rides

1	2	lpepPickup...	lpepDropoff...	123	passengerC...	12	tripDistance	abc	puLocationId	abc	dolocationId	12	pi
1	2	1/23/2019 10:17...	1/23/2019 10:17...	1	0	74	74						
2	2	1/23/2019 11:07...	1/23/2019 11:07...	1	0	130	130						
3	2	1/23/2019 3:02:1...	1/23/2019 3:02:2...	1	0	28	28						
4	2	1/23/2019 4:33:2...	1/23/2019 4:33:3...	1	0	264	192						
5	2	1/23/2019 7:13:0...	1/23/2019 7:13:0...	1	0	264	152						
6	2	1/23/2019 11:54:...	1/23/2019 11:55:...	1	0	223	223						
7	2	1/23/2019 11:06:...	1/23/2019 11:06:...	1	0	225	225						
8	2	1/23/2019 11:47:...	1/23/2019 11:47:...	1	0	210	210						
9	2	1/24/2019 1:42:4...	1/24/2019 1:42:5...	1	0	10	10						
10	2	1/24/2019 7:54:1...	1/24/2019 7:54:1...	1	0	264	171						
11	2	1/23/2019 6:36:1...	1/23/2019 6:36:1...	1	0	264	42						
12	2	1/23/2019 10:06:...	1/23/2019 10:06:...	1	0	196	95						

Виконано (10 sec 208 ms)

Стовпці 23 Рядки 1,000

Query data in a lakehouse

```

SELECT DATENAME(dw,lpepPickupDatetime) AS Day,
       AVG(tripDistance) AS AvgDistance
  FROM taxi_rides
 GROUP BY DATENAME(dw,lpepPickupDatetime)
    
```

abc Day	AvgDistance
Monday	2.8768393894496
Wednesday	2.82952699686273
Sunday	3.0794462803924
Tuesday	2.82074476088204

Visualize data in a lakehouse

Файл Подання Подання читання Макет для мобільних пристрів Відкрити модель даних Copilot

Фільтри

Візуалізації

Дані



Create an eventstream

taxi-data

Дані

- ✓ Джерела
- taxi
- ✓ Пункти призначен...

Попередній перегляд даних

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_c...
2	2022-06-01 00:27:47	2022-06-01 00:35:17	1.0
1	2022-06-01 00:06:24	2022-06-01 00:10:01	1.0
2	2022-06-01 00:27:51	2022-06-01 00:48:29	2.0
2	2022-06-01 00:25:14	2022-06-01 00:30:39	1.0
2	2022-06-01 00:45:13	2022-06-01 00:53:16	2.0
2	2022-06-01 00:05:46	2022-06-01 01:00:42	1.0
2	2022-06-01 00:28:17	2022-06-01 00:36:05	1.0
2	2022-06-01 00:42:22	2022-06-01 00:54:21	1.0

Create an eventhouse

my_event_house

Eventhouse

Загальні відомості про систему

Eventhouse storage

Original size	0 B
Compressed size	0 B

Storage resources

Eventhouse storage (compressed)	0 B
---------------------------------	-----

Eventhouse Details

Region	UK South
Query URI	Copy URI
Last ingestion	N/A
Ingestion URI	Copy URI
Minimum consu...	On demand (Chan...)
Python plugin	Not installed



Database details

- Compressed: OB
- Original: OB

OneLake

Availability: Enabled

When enabled, tables in this database are available in OneLake.

Overview

Автор: Tetiana Semenchuk
Створено: June 12, 2025
Регіон: UK South

Дані

- Джерела: taxi
- Пункти призначення: 8d69f9fa-9a7c-4d2...

Відомості **Попередній перегляд даних**

VendorID	ttrip_pickup_datetime	ttrip_dropoff_datetime	passenger_s
1	2022-06-01T12:51:11Z	2022-06-01T12:58:34Z	3
1	2022-06-01T12:04:12Z	2022-06-01T12:47:54Z	1
2	2022-06-01T12:48:51Z	2022-06-01T11:30:00Z	1
2	2022-06-01T12:54:09Z	2022-06-01T11:30:55Z	1
2	2022-06-01T12:44:44Z	2022-06-01T11:25:04Z	2
1	2022-06-01T12:57:22Z	2022-06-01T11:23:24Z	0
1	2022-06-01T12:39:41Z	2022-06-01T12:47:52Z	0
2	2022-06-01T10:38:15Z	2022-06-01T11:00:09Z	2

Query the captured data

Explorer

- Tables: taxi
- Materialized View: my_event_house
- Functions: my_event_house

```

4 // SQL - T-SQL conversions - https://aka.ms/sqlcheatsheet
5 //*****
6
7 // Use "take" to view a sample number of records in the table and check the data.
8 taxi
9 | take 100
10
11 // See how many records are in the table.
12 YOUR_TABLE_HERE
  
```

Таблиця 1

VendorID	ttrip_pickup_datetime	ttrip_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store
> 2	2022-06-01 13:55:51.0000	2022-06-01 14:00:49.0000	5	0.25	1 N	
> 2	2022-06-01 13:59:15.0000	2022-06-01 14:13:30.0000	1	2.23	1 N	
> 1	2022-06-01 14:00:17.0000	2022-06-01 14:13:12.0000	1	1.1	1 N	
> 2	2022-06-01 14:01:33.0000	2022-06-01 14:16:43.0000	1	1.57	1 N	



The screenshot shows the Microsoft Fabric Query Workbench interface. The top navigation bar includes tabs for Home, Help, and Copilot, along with a search bar and various status indicators. The main area has tabs for Explorer, Tab, and a code editor. The code editor contains the following SQL query:

```

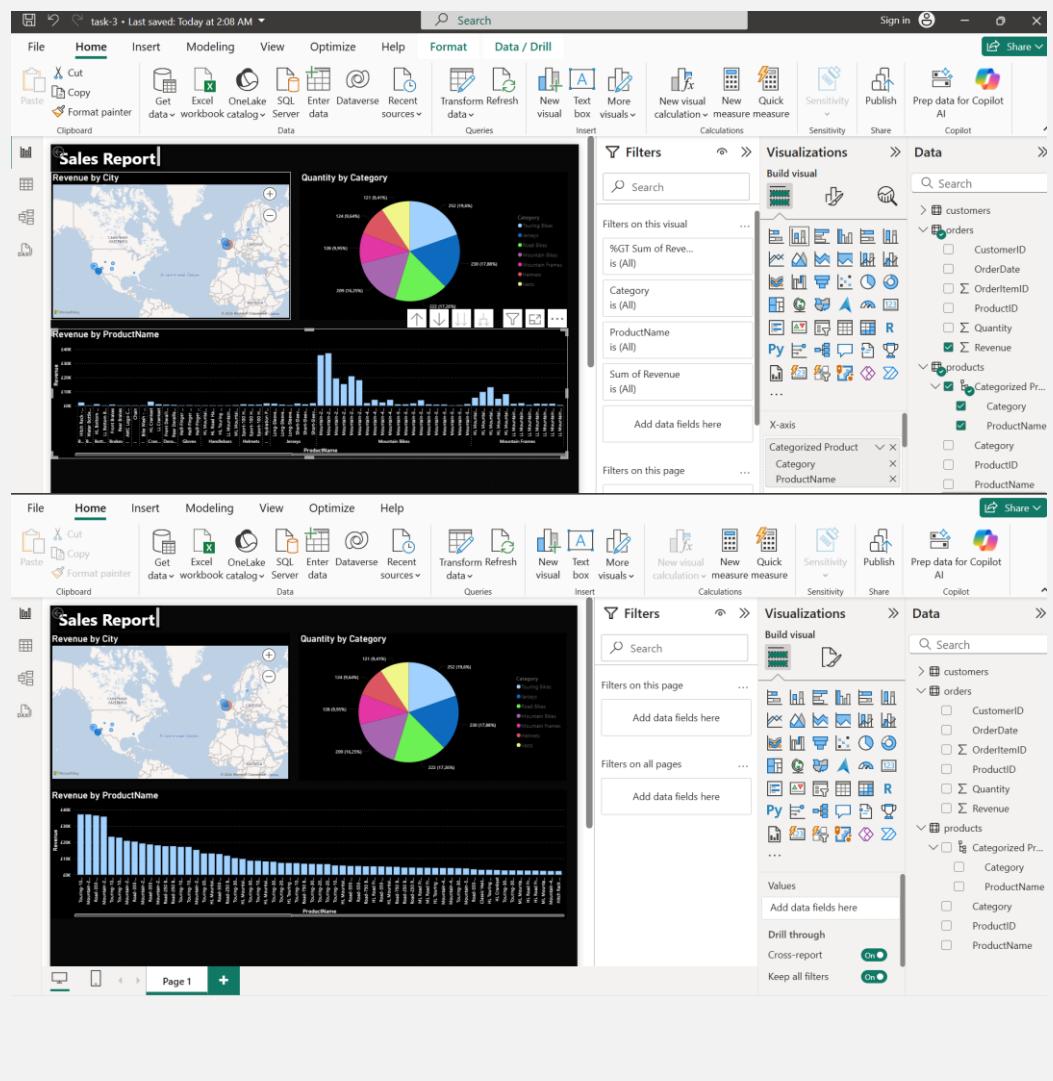
4 // SQL - KQL conversions - https://aka.ms/sqlcheatsheet
5 /**
6 // Use "take" to view a sample number of records in the table and check the data.
7 taxi
8 | summarize PickupCount = count() by bin(todatetime(tpep_pickup_datetime), 1h)
9
10 // See how many records are in the table.
11 VMLR.TARIF.HPFB

```

Below the code editor is a table titled "Таблица 1" (Table 1) with columns "tpep_pickup_datetime" and "PickupCount". The data shows the following distribution of pickup times and counts:

tpep_pickup_datetime	PickupCount
2022-06-01 06:00:00.0000	4
2022-06-01 10:00:00.0000	1
2022-06-01 11:00:00.0000	68
2022-06-01 12:00:00.0000	4454
2022-06-01 13:00:00.0000	6699

Power BI Desktop



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

Practice Assessment Results: June 12, 2025

 Practice Assessment for Exam DP-900: Microsoft Azure Data Fundamentals

It took you 28 minutes to complete this assessment.

Overall Results

To be better prepared for the exam, aim to achieve a score of 80% or higher in multiple attempts.

Score: 86%

[Show My Answers](#)



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: Data-Driven Transformation Using Microsoft Azure

1. Data Laws and Regulations

GDPR Compliance

The General Data Protection Regulation (GDPR) is a critical regulation that mandates how businesses handle personal data within the EU and UK. As "Paws & Whiskers" begins storing and analysing customer data (names, addresses, purchase history, and preferences), it must ensure:

- Data is collected and processed lawfully, fairly, and transparently.
- Customers are informed about what data is being collected and why.
- Personal data is kept secure and only used for specified purposes.
- Customers have the right to access, correct, and delete their data.

Failure to comply can result in significant penalties and damage to the company's reputation.

Data Protection Act (DPA) 2018

The DPA 2018 complements the GDPR in the UK and imposes specific rules on how personal data must be processed:

- It requires that data must be accurate and up-to-date.
- It establishes lawful bases for data processing, especially for marketing purposes.
- It mandates the implementation of technical and organisational measures to protect data integrity and confidentiality.

For "Paws & Whiskers," this means having clear data policies, regular data audits, and staff training on data privacy.

Other Industry Standards

In addition to GDPR and DPA 2018, the following standards are relevant:

- **PCI DSS (Payment Card Industry Data Security Standard):** If the business handles credit card payments, compliance with PCI DSS is essential to protect cardholder data.
- **ISO/IEC 27001:** A leading standard for information security management systems (ISMS), useful as the business grows and manages more sensitive data.



2. Azure Service Recommendations

Data Storage

- **Azure SQL Database:** Ideal for structured data such as customer profiles, sales transactions, and product information. It supports relational data models and offers built-in scalability and security.
- **Azure Blob Storage:** Suitable for unstructured data like images (e.g., pet photos), documents, and large CSV datasets. Blob Storage is cost-effective and integrates easily with analytics services.

Data Analysis Tools

- **Azure Synapse Analytics:** A powerful analytics service that integrates data ingestion, storage, and analysis. Ideal for identifying sales trends, seasonal demand, and customer purchase behaviour.
- **Azure Machine Learning:** Useful for building predictive models, such as recommending pet products or forecasting demand.

Data Integration and Automation

- **Azure Data Factory:** Automates the process of collecting, transforming, and loading (ETL) data from spreadsheets and other sources. This ensures timely and accurate data for decision-making.

3. Data Types and Data Modelling

Data Categories

- **Customer Demographics:** Names, addresses, email, pet preferences.
- **Transaction History:** Purchase dates, items bought, total amount.
- **Pet Inventory:** Types of pets, breed, health status, availability.
- **Product Categories:** Food, toys, accessories, grooming items.

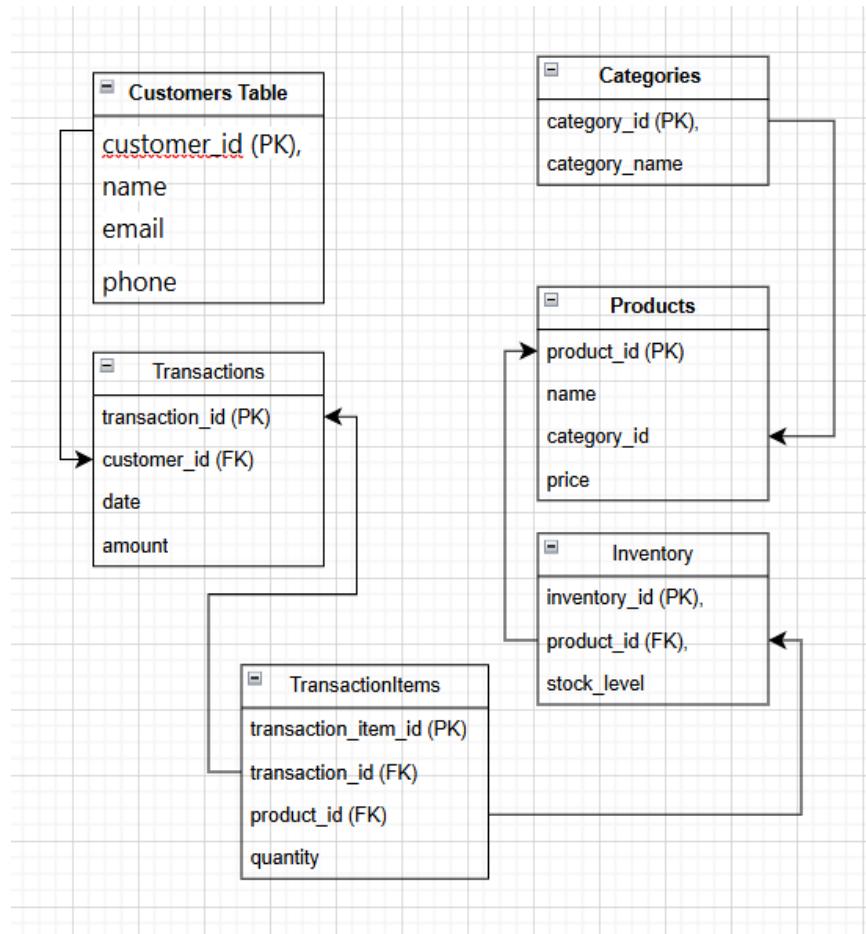
Data Modelling Approach

A **relational model** is most suitable:

- **Customers Table:** customer_id (PK), name, email, phone
- **Transactions Table:** transaction_id (PK), customer_id (FK), date, amount
- **Products Table:** product_id (PK), name, category_id, price
- **Inventory Table:** inventory_id (PK), product_id (FK), stock_level
- **Categories Table:** category_id (PK), category_name



- **TransactionItems Table:** transaction_item_id (PK), transaction_id (FK), product_id (FK), quantity



4. Data Storage Formats and Structures in Azure

Data Formats

- **CSV:** For importing sales and inventory data from spreadsheets.
- **JSON:** For storing structured data from web forms and APIs.
- **Parquet:** Efficient, columnar format for analytics workloads in Azure Synapse.

Data Security and Encryption

- Enable **Transparent Data Encryption (TDE)** for Azure SQL.
- Use **Storage Service Encryption (SSE)** for Blob Storage.
- Apply **Role-Based Access Control (RBAC)** to restrict access.



- Enable **Private Endpoints** and **Virtual Networks** to secure data communication.

5. Additional Considerations

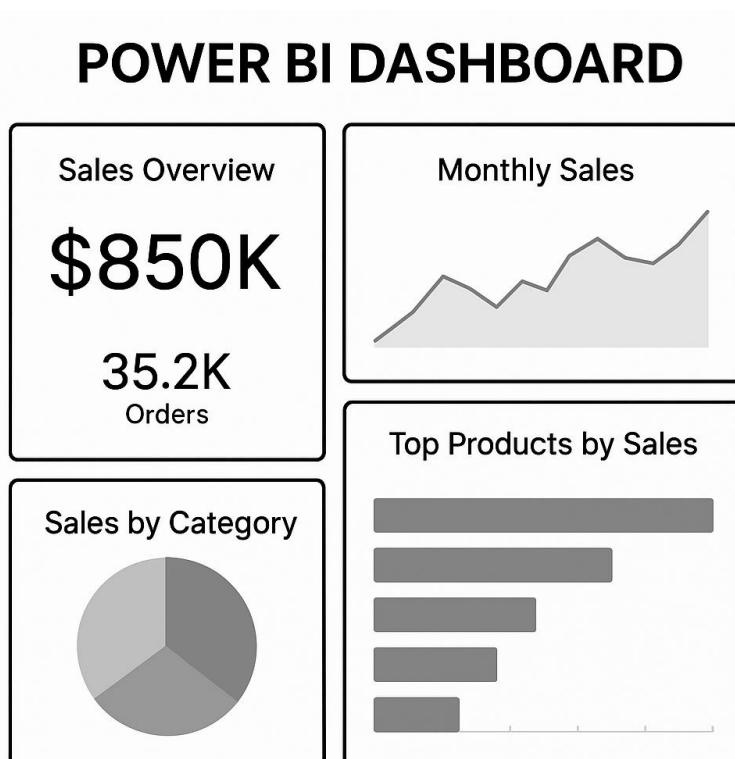
Backup and Disaster Recovery

- Use **Azure Backup** for automatic backups of databases and file systems.
- Consider **Azure Site Recovery** for business continuity in the event of a system failure.

Data Visualisation

- **Power BI:** Connects seamlessly to Azure SQL and Synapse Analytics to create dashboards showing sales performance, customer trends, and inventory levels.

Dashboard example:



Future Scalability

- Azure services like SQL Database and Synapse scale automatically to handle more data and users.
- Pay-as-you-go pricing allows the company to expand infrastructure only when needed.



Conclusion

Transitioning to Microsoft Azure will modernise how "Paws & Whiskers" manages its data — making storage secure, analysis powerful, and reporting insightful. With built-in compliance features, automation tools, and scalable services, Azure provides a future-proof foundation for the company's data-driven growth.

References

- General Data Protection Regulation (GDPR) – <https://gdpr-info.eu>
- Data Protection Act 2018 –
<https://www.legislation.gov.uk/ukpga/2018/12/contents>
- Microsoft Azure Documentation – <https://learn.microsoft.com/azure>
- PCI DSS Standards – <https://www.pcisecuritystandards.org>

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:

<https://learn.microsoft.com/en-us/credentials/certifications/azure-data-fundamentals/?practice-assessment-type=certification>



Global deployment means the ability to launch applications, services, and data infrastructure **across multiple geographic regions** almost instantly using the cloud — ensuring low latency, data redundancy, and global reach.

Spotify

Use of Cloud:

Spotify uses **Google Cloud Platform (GCP)** to distribute its services across global regions, ensuring that users in Europe, Asia, and the Americas all get fast, reliable music streaming.

Benefit from Global Deployment:

- Delivers high performance and fast load times to users worldwide
- Meets local data compliance rules (e.g. GDPR in the EU)
- Keeps services running even if one region faces downtime (failover support)

Core values

- British Values
- Safeguarding
- Prevent

Here's a more detailed look at the challenges:

1. Burnout and Stress:

Fast-paced environment:

Cloud work often requires quick turnaround times and continuous availability, leading to a sense of pressure and overwhelm.

Overload and long hours:

The constant connectivity of cloud computing can blur the lines between work and personal life, leading to longer hours and a decrease in work-life balance.

2. Isolation and Lack of Social Interaction:



Remote work:

While offering flexibility, remote cloud work can also lead to feelings of isolation and a lack of social interaction with colleagues.

Difficulty building relationships:

The reliance on virtual communication can make it harder to build strong relationships with colleagues, leading to feelings of loneliness and disconnection.

3. Security and Privacy Concerns:

Data breaches:

Cloud computing involves storing sensitive data in a shared environment, making it vulnerable to data breaches and security threats.

Privacy risks:

The ease with which cloud services can access and collect data raises privacy concerns, especially in industries like healthcare where patient information is highly sensitive.

4. Network Dependence:

Reliability issues:

Cloud services are dependent on a stable internet connection, which can be problematic in areas with poor connectivity or power outages.

Latency and performance issues:

Inadequate internet bandwidth can lead to slow performance and delays in accessing cloud services.

5. Lack of Control and Flexibility:



Limited control over data:

Cloud users may have limited control over their data ownership and access, especially in situations where cloud providers have the ability to access or retain data.

Vendor lock-in:

Cloud users can become reliant on specific providers, making it difficult to switch providers or migrate to on-premise solutions.

6. Managing Multiple Clouds:

Increased complexity:

Organizations using multiple cloud providers may face challenges in managing and monitoring resources across different platforms.

Security challenges:

Managing security and compliance across multiple cloud environments can be complex and resource-intensive

□ 1. Resource Provider

A **Resource Provider** is a **service in Azure** that supplies a set of resources you can manage.

□ Example:

- Microsoft.Compute – handles resources like Virtual Machines.
- Microsoft.Storage – handles storage accounts.



- `Microsoft.Sql` – handles SQL databases.

When you create a resource (e.g., a VM), you are **using a resource provider** (e.g., `Microsoft.Compute`) to provision it.

Think of a **resource provider** as a "factory" that creates and manages specific Azure services.

□ 2. Resource

A **Resource** is an **individual instance of a service** that you create in Azure.

□ Examples:

- A **Virtual Machine** (`myVM`)
- A **Storage Account** (`mystorage123`)
- An **SQL Database** (`students-db`)
- A **Web App** (`mywebsite123`)

Each resource has its **own settings, pricing, location, and identity**.

□ 3. Resource Group

A **Resource Group** is a **container** that holds related resources for an Azure solution.

□ Features:

- Logical grouping of resources.
- All resources in a resource group **share the same lifecycle** (create, update, delete).
- Helps organize resources by **project, department, environment**, etc.

□ Example:

A resource group called `myProject-RG` might contain:

- 1 Virtual Machine
- 1 Storage Account
- 1 Virtual Network
- 1 SQL Database

You can manage, monitor, and apply policies to all the resources **together** via the resource group.



□ Summary Table:

Term	What It Is	Example
Resource Provider	Backend service for a type of Azure resource	Microsoft.Compute
Resource	An individual Azure service instance	A Virtual Machine named VM01
Resource Group	A logical container for related resources	dev-team-resources

□ Data Engineer

□ Main Role:

Builds and maintains the **data infrastructure** (pipelines, storage, databases) so that data is collected, stored, and made accessible for analysis.

□□ Key Responsibilities:

- Design and build **data pipelines** (ETL/ELT)
- Develop and manage **data warehouses** or **data lakes**
- Ensure **data quality, reliability, and security**
- Integrate data from various sources (APIs, logs, databases)
- Work with tools like **Azure Data Factory, Databricks, SQL, Spark, Python**

□ Example Tasks:

- Create an Azure pipeline to pull data from an app and store it in Azure SQL
- Optimize large data sets for fast querying
- Automate data ingestion from external sources

□ Data Analyst



□ Main Role:

Uses data to generate **insights, reports, and visualizations** to help businesses make informed decisions.

□ Key Responsibilities:

- Analyze trends, KPIs, and metrics
- Build **dashboards and reports** (e.g., Power BI, Excel)
- Write **queries** to extract data (SQL)
- Present findings to non-technical stakeholders
- Support business teams with **data-driven decisions**

□ Example Tasks:

- Create a Power BI report showing monthly workshop attendance
- Use SQL to pull data on user registration trends
- Identify why website traffic dropped in a certain month

□ How They Work Together:

| **Data Engineer** → builds → | **data platform** → for → | **Data Analyst** → to analyze → | **business insights** |

1. Azure Data Factory (ADF)

Purpose: Data movement and transformation.

What it does:

- Orchestrates and automates **data workflows** (called pipelines).
- Moves data between services (e.g., from SQL Server to Data Lake).
- Transforms data using **data flows** (code-free) or **custom code**.
- Good for **ETL/ELT processes** (Extract, Transform, Load).

2. Azure Databricks

Purpose: Advanced data analytics and machine learning.



What it does:

- Provides a **collaborative Apache Spark-based environment**.
- Used by data engineers, analysts, and data scientists.
- Processes **large-scale data** using languages like Python, Scala, SQL, R.
- Trains **machine learning models** and handles **real-time data analytics**.

3. Azure Synapse Analytics

Purpose: Unified data analytics platform.

What it does:

- Combines **data warehousing** and **big data analytics**.
- Allows you to query data using **SQL** (on-demand or dedicated), **Spark**, and **T-SQL**.
- Integrates easily with **Power BI**, **Data Lake**, and **ADF**.
- Great for building **enterprise-scale BI solutions and dashboards**.

Fabric.microsoft.com

<https://app.diagrams.net/>



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.

