



Data Technician

Name: Callum Laval

Course Date: 16/12/2024

Table of contents

Day 1: Task 1	3
Day 1: Task 2	5
Day 1: Task 3	7
Day 2: Task 1	9
Day 3: Task 1	12
Day 3: Task 2	14
Day 3: Task 3	15
Day 4: Task 1	16
Day 4: Task 2	17
1. Scenario Background	17
2. Data Laws and Regulations	17
3. Azure Service Recommendations	17
4. Data Types and Data Modelling	18
5. Data Storage Formats and Structures in Azure	18
6. Additional Considerations	18
Submission Guidelines:	19
Course Notes	28
Additional Information	29



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?	<p>It allows us to store our files online which can make them more accessible, can allow for more storage and provides more security for your files. It allows cloud-based communication so you can access all your messages just with an internet connection. Most social media is cloud based and so it allows people to make connections with other people and share information with them.</p> <p>Source - 9 Common Uses of Cloud Computing</p>
How can it benefit a business?	<p>Businesses can do big data analysis more affordably since cloud computing takes on more of the computing resources often for better prices than if you did the analysis using your own resources. It allows businesses to build and model disaster recovery solutions in the cloud, this means that a business can more easily relaunch their services in the event of a disaster. It allows businesses to develop and test their software in a way that is cheaper and faster allow them to more easily keep up with the market.</p> <p>Source - 9 Common Uses of Cloud Computing</p>
What's the alternative to cloud computing?	<p>Fog Computing – A decentralised infrastructure that performs a portion of computing between the data source and origin server.</p> <p>Edge Computing – A distributed IT architecture where data processing happens as close to the source of data as possible.</p> <p>Mesh Computing – An architecture where infrastructure nodes are connected directly and non-hierarchically to</p>



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

as many other nodes as possible, allowing numerous routes for data to travel.

Bare Metal Cloud – Enables companies to rent physical servers and deploy them in a similar way to cloud computing.

On-Prem Hosting – An on-site server room is set up and equipped with hardware.

Colocation Hosting – A third-party facility provides space for privately-owned servers and computing hardware and this space is rented by the customers.

Source - [Alternatives to Cloud Computing {6 Options}](#)

AWS – Has highly scalable public cloud offerings, private cloud, hybrid cloud and multi-cloud.

Azure – Has secure storage, scalability, reliability, diverse data handling and advanced analytics capabilities.

Google cloud platform – Has cloud storage, Cloud dataproc, cloud pub/sub, cloud SQL, google app engine, advanced compute engine, network operations center and BigQuery.

Sources - [13 Top Cloud Service Providers Globally \(UPDATED 2024\)](#)

[Microsoft Azure Features and Benefits | Use Cases](#)
[Features of Google Cloud Platform - GeeksforGeeks](#)



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)	<p>It is a type of cloud computing service offers resources on demand on a pay-as-you-go basis. Each resource is paid separately and is paid for only as long as you need it. The provider will manage the infrastructure whilst the purchaser will manage your own software.</p> <p>Source - What is IaaS? Infrastructure as a Service Microsoft Azure</p>	<p>This service may be used when a business wants to make changes such as scaling up, increasing their performance, enhancing their security or reducing their running costs but do not want to refactor their underlying architecture.</p>
PaaS (Platform as a service)	<p>It is a service that offers complete development and deployment environment in the cloud designed to support the complete web application lifecycle. The resources are purchased from a cloud service provider on a pay-as-you-go basis and are accessed through an internet connection. This includes infrastructure like IaaS but also provides things like</p>	<p>A business may use this service to help them with developing applications. This service provides built-in software components and customizable cloud-based applications to support developers. It also provides tools for analytics or business intelligence allowing organisations to make more informed business decisions. PaaS offers the same advantages as IaaS but has additional features that can</p>



	<p>middleware, development tools, BI services and database management systems. The purchaser will manage the applications and services they develop whilst the provider will manage everything else.</p> <p>Source - What is PaaS? Platform as a Service Microsoft Azure</p>	<p>reduce coding time, support remote working and efficiently manage the application lifecycle.</p>
SaaS (Software as a service)	<p>It is service that allows the renting of the use of cloud-based apps on a pay-as-you-go basis that is accessed through the internet. The underlying infrastructure, middleware, app software and app data will be in the provider's data center and the provider will manage all the hardware and software along with the availability and security of the app and your data.</p> <p>Source - What is SaaS? Software as a Service Microsoft Azure</p>	<p>This allows businesses to use sophisticated apps that they do not need to maintain which can benefit the business. They are also often cheaper since the business does not need to maintain them and only pays for how much they use the app. The service provider often makes sure the apps can be run on different types of computers and since all you need is an internet location it is easy for an entire workforce to access and use these apps regardless of where they are located.</p>

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>A type of computing where third-party providers offer computing services through the internet to organisations and people who wants to use and purchase them. The providers are the ones responsible for managing and maintaining the infrastructure, hardware and software.</p> <p>Public cloud would be used by businesses looking to scale up or reduce their costs quickly and affordable and do not need much customisation.</p> <p>Examples of public cloud are AWS, Microsoft Azure and Google Cloud which are used by individuals as well as organisations of varying sizes.</p> <p>Source - Public Cloud vs Private Cloud vs Hybrid Cloud Microsoft Azure</p> <p>What Is a Public Cloud? Google Cloud</p>
Private Cloud	<p>A type of computing where resources are only available to a single organisation and the data center is manged on-premises or off-site. The services and infrastructure are maintained on a private network and the hardware and software are solely dedicated to the organisation.</p> <p>Private cloud would be used by large businesses who have unique or specific IT requirements e.g. a government agency or a financial institution.</p> <p>Both Oracle and IBM offer private cloud services.</p> <p>Source - What Is Private Cloud Storage? Definition, Types, Examples, and Best Practices - Spiceworks</p>

	Public Cloud vs Private Cloud vs Hybrid Cloud Microsoft Azure
Hybrid Cloud	<p>Hybrid cloud contains elements of both public cloud and private cloud. It allows data and apps to move between on premises infrastructure and infrastructure available through the internet.</p> <p>Hybrid cloud would be used by businesses who are wanting to scale their computing resources and get some of the benefits of public cloud but still want to keep control of their sensitive data.</p> <p>Netflix, Hulu, Uber and Airbnb all use hybrid cloud to scale their data uses during peak-use periods.</p> <p>Source - Public Cloud vs Private Cloud vs Hybrid Cloud Microsoft Azure</p> <p>Hybrid Cloud Computing: 16 Examples to Know Built In</p>
Community Cloud	<p>Community Cloud is an infrastructure where multiple organisations share resources and services based on common operational and regulatory requirements.</p> <p>It can be used by organisations that have common business requirements and may want to share data, share services or share industry regulations. An example would be a government body wanting to communicate with a different government body.</p> <p>Source - What Is Community Cloud? Definition, Architecture, Examples, and Best Practices - Spiceworks</p>

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	Also called hacking, this is when a person accesses another person's computer without their permission.	If I made a piece of software, I cannot install that software on some else's computer without their permission.
Unauthorised access to computer materials with intent to commit further crime.	This is when someone hacks another person's computer to do a malicious act such as steal data or destroy the device.	This makes things like computer viruses, computer worms, trojan horses, ransomware and spyware illegal since they are often installed without the user's permission for malicious reasons.
Unauthorised modification of data.	This is when you change or delete another person's data, e.g. computer files, without their permission.	If a running a piece of software involves modifying computer files, then the software provider would need to get permission from the user first before the software can run.

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
It made the making, owning or distribution of hacking tools illegal if they were to be used for computer misuse.
Denial of service attacks were made illegal.



Creating or spreading malicious software such as viruses, worms, trojans or ransomware was made illegal.

Look at the below website to answer the questions:
<https://www.gov.uk/personal-data-my-employer-can-keep-about-me>

Write down three items of data which a company can store about an employee.

Name

Address

Date of Birth

Give three more examples of data that an employer can only store if they first get the employee's permission.

Race and Ethnicity

Religion

Sexual History or Orientation

Conduct further research to answer the below questions.

Question	Answer
Provide one example of: Copyright infringement	If I upload a movie to YouTube that is not in the public domain and is copyrighted, e.g. Star Wars, then I would have committed copyright infringement. Since the movie is copyrighted, the owners have certain exclusive rights to that movie which involves control over who can copy and distribute the film. Hence since I uploaded to film on a public and easily accessible platform like YouTube without their permission I have copied and distributed the film without their permission.
Provide one example of: Plagiarism	If I publish a poem that I claimed to have written but I found the poem in a poetry book by a different author, that would be an example of plagiarism.

What are two consequences of copyright infringement and software piracy?	In the UK if you sell a recording of copyrighted material without the original owners consent you can be sentenced to 6 months in jail or be fined.
Give three possible consequences for individuals when using pirated software	The software owner could sue you for monetary damages. Pirated software is often less safe and secure so you could get a virus. The software is also more likely to not run as well and not eligible for updates.

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

33%

1 of 3 required activities complete



Explore relational data in Azure (Expected Duration 2 hours, 15 minutes)
DP-900T00-A Microsoft Azure Data Fundamentals [Cloud Slice Provided], Learning Path 02 (CSS)

Required: Yes

Started: Wednesday, January 29, 2025 11:16 AM (GMT Standard Time)

Ended: Wednesday, January 29, 2025 11:50 AM (GMT Standard Time)

Launch

Cancel

8 of 10 launch attempts remaining



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

Completed
lab



2



Explore non-relational data in Azure (Expected Duration 2 hours, 15 minutes)
DP-900T00-A Microsoft Azure Data Fundamentals [Cloud Slice Provided], Learning Path 03 (CSS)

Required: Yes

Started: Wednesday, January 29, 2025 2:02 PM (GMT Standard Time)

Ended: Wednesday, January 29, 2025 3:18 PM (GMT Standard Time)

Launch

Cancel

9 of 10 launch attempts remaining



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.



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

**Completed
lab**

Was unable to fully progress in 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

Your practice assessment results | Microsoft Learn - Personal - Microsoft Edge
https://learn.microsoft.com/en-us/credentials/certifications/azure-data-fundamentals/practice/results?assessmentId=24&practice-assessment-type=certification&snapshotId=724be280-7211-43ee-898b-33b4...

Learn / Credentials / Browse Credentials / Microsoft Certified Azure Data Fundamentals /

Practice Assessment Results: January 30, 2025

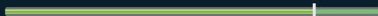
 Practice Assessment for Exam DP-900: Microsoft Azure Data Fundamentals

It took you 39 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: 82%



[Show My Answers](#)



4



Practice Assessment: DP-900T00-A Microsoft Azure Data Fundamentals [Details](#)

Required: No

Started: Thursday, January 30, 2025 11:57 AM (GMT Standard Time)

[Launch](#)



Day 4: Task 2

1. Scenario Background

"Paws & Whiskers" is a growing pet shop that aims to improve its business by analysing sales, customer information, and inventory data. Currently, the data is collected manually or stored in spreadsheets. Management is interested in transitioning to Microsoft Azure to streamline data storage, analysis, and reporting, enabling them to make data-driven decisions.

2. Data Laws and Regulations

Identify and explain the data laws and regulations relevant to handling customer data within the proposal. Ensure you cover the following points:

- **GDPR Compliance:** Highlight the importance of adhering to the General Data Protection Regulation (GDPR), particularly as it relates to storing and processing customer information.
- **Data Protection Act (DPA) 2018:** Outline how the DPA 2018 may affect the way "Paws & Whiskers" collects and stores data, ensuring compliance with UK laws on data privacy.
- **Other Industry Standards:** Research any additional data protection standards or regulations that may apply to pet shop data, particularly if they involve sensitive or payment information.

3. Azure Service Recommendations

Recommend Microsoft Azure services that would suit the company's data analysis needs and explain why these services are suitable. Your recommendations should include:



- **Data Storage:** Identify suitable storage options, such as **Azure Blob Storage** or **Azure SQL Database**, and discuss the benefits of each for storing large datasets, including inventory, sales transactions, and customer details.
- **Data Analysis Tools:** Recommend tools such as **Azure Machine Learning** for customer behaviour analysis or **Azure Synapse Analytics** for analysing sales trends.
- **Data Integration and Automation:** Explain how services like **Azure Data Factory** could automate data collection and integration processes, improving efficiency.

4. Data Types and Data Modelling

Define the types of data "Paws & Whiskers" will need to work with and describe your approach to data modelling:

- **Data Categories:** Identify key data types, such as customer demographics, transaction history, pet inventory, and product categories.
- **Data Modelling Approach:** Outline how you would structure this data using a relational model or a data warehouse approach, considering factors like tables, entities, relationships, and primary keys.

5. Data Storage Formats and Structures in Azure

Discuss how you would store data within Azure and the formats you would recommend:

- **Data Formats:** Specify recommended formats (e.g., CSV for raw data imports, JSON for structured data, Parquet for analytics) and explain why these formats are suitable for specific data types.
- **Data Security and Encryption:** Include recommendations for securing data using Azure's built-in encryption features and access controls to ensure compliance with data privacy regulations.

6. Additional Considerations

Provide any other considerations that might enhance data handling and efficiency in Azure, such as:

- **Backup and Disaster Recovery:** Outline a backup plan using **Azure Backup** or **Azure Site Recovery** to safeguard against data loss.
- **Data Visualisation:** Discuss potential use of **Power BI** within Azure for creating dashboards that provide management with real-time insights into sales and customer trends.
- **Future Scalability:** Comment on how Azure services can scale as the business grows, accommodating larger datasets and more complex analyses.



Submission Guidelines:

1. **Structure:** Ensure your report is well-organised, with sections for each task (e.g., Data Laws, Azure Services, Data Types, etc.).
2. **Formatting:** Include headings, bullet points where appropriate, and any visuals or diagrams that support your explanations.
3. **References:** Cite any resources or regulations referenced in the report.
4. **Length:** Aim for 1500-2000 words.

Data Laws and Regulations:

Data Law or Regulation	Description	Application to situation
GDPR	The GDPR is a European Union regulation that governs information privacy in the European union (EU) and the European Economic Area (EEA) and the transfer of personal data outside of those areas. It lists seven protection and accountability principles for processing data, it states that you are required to handle data securely which can involve two-factor authentication or end-to-end encryption, and it lists the instances where it's legal to process personal data among many things.	Paw & Whiskers must make sure that the data they store is for an adequate reason that falls under the GDPR and when they analyse their data, they must also follow the GDPR rules. They must also make sure that their current and their future databases have adequate security. In the case there is a data leak they must tell a supervisory authority within 72 hours and must inform the individuals whose data got leaked.
Data Protection Act	The data protection act limits and regulates what organisations, businesses and governments can do with your data. They must follow certain rules such as making sure the data is used fairly and for specific limited purposes, the data is kept up to date, the data is kept no longer than needed and data is handled in a way the ensures appropriate security. There are stronger protections for people's sensitive information such as religion, race	Paws & Whiskers are planning to move their current data and do further analysis on the data. Since some of this data contains people's personal information e.g. customer information, they need to keep these people informed about what they are doing and only used their data for that explicit purpose to comply with UK laws. They must also make sure that when they move the information the information remains secure, and their new database is secure. They

	and health and people also have the right to know what information is stored about them and have the to have that data deleted.	must also be prepared for customers to not want their data moved and analysed; to have their data deleted at any time and ask to see the data they store about them at any time.
Freedom of Information Act	The Freedom of Information act is an act of UK parliament that gives members of the public the right to access to information held by public authorities and obliges public authorities to publish certain information about their activities. Public authorities include government departments, NHS, state schools and police forces and the information may include printed documents, computer files, emails, letters or videos.	Since the pet shop may be holding some personal data on the customers, they must be prepared to show what data they hold on to the person if the person requests it.

Azure Service Recommendations:

Service	Service Category	Description	Application to situation
Azure Blob Storage	Data Storage	This service allows users to store and manage large amounts of unstructured data in the cloud. Usually, data is stored as a file with some sort of file hierarchy but for large files or unstructured files it can be easier to store them as "blobs" which this service helps store. These blobs cover a diverse range of data types and can include documents, images, videos and backups.	<p>Paws & Whiskers may have a very large amount of data to move and store hence this service can be very useful since it its designed to store large amounts of data. Benefits of this storage method may include the following:</p> <ul style="list-style-type: none"> • It can be more affordable since it is designed for this purpose. • The storage can be easily scale if more data needs to be added. • Blob storage is designed for high throughput and low



			<p>latency so it may have better performance.</p> <ul style="list-style-type: none"> • It automatically replicates data ensuring durability and availability. • It can easily integrate with web applications.
Azure SQL Database	Data Storage	This service is a fully managed relational database service. It is built on SQL server technology and offers many features for users and developers	Data such as inventory data have a natural structure to them so it may be more beneficial to the pet shop to store it in a relational database as it will be easier to query and analyse the data in the future. SQL databases provide ACID (Atomicity, Consistency, Isolation, Durability) properties which are essential for their inventory data and financial data since it guarantees data validity despite error or power failures.
Azure Machine Learning	Data Analysis	This is a cloud-based service that enables users to build, deploy, and manage machine learning models. It is designed to support the entire machine learning lifecycle, from data preparation to model training to deployment and monitoring.	<p>This service has the following features which can be used to analyse customer behaviour:</p> <ul style="list-style-type: none"> • By analysing historical data, it can predict future customer behaviour which can be used to recommend products to customers. • Models can segment customers into different groups based on their behaviour, preferences, and demographics meaning you can tailor marketing

			<p>campaigns to different groups</p> <ul style="list-style-type: none"> • It has real-time analytics meaning the pet shop can respond quickly to customer actions.
Azure Synapse Analytics	Data Analysis	<p>This service is a comprehensive analytics service that brings together big data and data warehousing. It provides a unified platform to ingest, prepare, manage, and serve data for immediate business intelligence and machine learning needs.</p>	<p>This service has the following features which can be used to analyse sales trends:</p> <ul style="list-style-type: none"> • You can integrate and ingest sales data from various sources such as CRM systems, e-commerce platforms, and ERP systems into Azure Synapse. This centralizes your data, making it easier to analyse. • You can clean and preprocess your sales data. This might include handling missing values, removing duplicates, and transforming data into a suitable format for analysis. • By storing and analysing historical sales data, you can identify long-term trends and patterns which can help in forecasting future sales, identifying seasonality, and understanding the impact of past marketing campaigns.
Azure Data Factory	Data Integration	<p>This is a cloud-based data integration service that allows you to create,</p>	<p>This service can automate collection and integration</p>

	and Automation	schedule, and manage data workflows that move and transform data from various sources to a centralized location for analysis and reporting. It also helps automate the movement of data between different systems in a scalable and efficient manner.	<p>processes in the following ways</p> <ul style="list-style-type: none"> • It enables seamless integration of data from various sources, such as on-premises databases, cloud storage, SaaS applications, and APIs. By creating pipelines, you can automate the process of extracting, transforming, and loading data. • You can schedule data workflows to run at specific times or in response to certain events. This means that data collection and integration tasks can be executed automatically at regular intervals, reducing the need for manual triggers and ensuring timely updates. • It supports data transformation activities, allowing you to clean, aggregate, and reformat data as it moves through the pipeline. This automated transformation ensures that data is in the right format and quality for analysis, without the
--	----------------	---	---

			need for manual data wrangling.
--	--	--	---------------------------------

Data Types and Modelling:

In a relational model we would have the following tables and entities with the data types noted in brackets:

Customers:

- CustomerID (Primary Key)
- Name (Text)
- Age (Number)
- Gender (Categorical or Text)
- Location (Text)
- Income (Currency)
- Occupation (Text)

Transactions:

- TransactionID (Primary Key)
- CustomerID (Foreign Key)
- Date and Time of Transaction (Date and Time)
- ProductID (Foreign Key)
- Amount Spent (Currency)
- Payment Method (Categorical or Text)

Pets:

- PetID (Primary Key)
- Species (Categorical or Text)
- Breed (Categorical or Text)
- Age (Number)
- Availability Status (Categorical or Text)
- Price (Currency)

Products:

- ProductID (Primary Key)
- Category (Categorical)
- Product Name (Categorical or Text)
- Description (Text)
- Price (Currency)
- Inventory Count (Number)

There are two relationships in this database:

- Customers to Transactions - One-to-Many Relationship since each customer can have multiple transactions, but each transaction is linked to a single customer.

- Transactions to Products - Many-to-Many Relationship since each transaction can include multiple products, and each product can be part of multiple transactions.

Data Storage Formats:

Data Storage Format	Uses	Application to Situation
JSON	Ideal for storing hierarchical data and for applications that require flexible schemas.	Storing customer demographics and pet inventory, which may have nested structures.
CSV	Suitable for tabular data and widely used for data interchange.	Storing transaction history and product categories in a simple, flat structure.
Parquet	Optimized for performance and storage efficiency, especially for large datasets and is Suitable for analytical workloads.	Storing large volumes of transaction history and customer demographics for analysis.
Avro	Ideal for data serialization, schema evolution, and compatibility. Often used in data streaming and messaging systems.	Storing real-time transaction data and pet inventory updates.
ORC	Provides efficient storage and query performance for big data. It is optimized for analytics.	Storing historical sales data and customer demographics for reporting and analysis.
SQL Database	Structured data with well-defined schemas and relationships. Ideal for transactional workloads.	Storing all tables in a relational database format.

Data Security and Encryption:

Azure has the following features for security and encryption:

- Azure Storage Service Encryption - Automatically encrypts data when it is written to Azure Storage.
- Azure Disk Encryption - Encrypts the OS and data disks of your virtual machines using BitLocker for Windows VMs and DM-Crypt for Linux VMs.



- Transport Layer Security (TLS) - Ensure that all data transmitted between your applications and Azure services is encrypted using TLS 1.2 or later. This protects data as it moves across the network.
- Multi-Factor Authentication - Adds an extra layer of security for user logins. This ensures that only authorized users can access sensitive data.
- Azure Active Directory - Used for identity and access management, providing secure user authentication and authorization.
- Audit Logs and Alerts: Enable audit logging and set up alerts to monitor access to sensitive data. This helps detect and respond to any unauthorized access attempts.
- Data Residency: Choose the geographical region where your data will be stored to comply with data residency requirements. Azure provides transparency regarding the location of your data and the controls in place to protect it.
- Compliance Certifications: Ensure that your data handling practices comply with industry standards and regulations, such as GDPR.

Backup and Disaster Recovery:

Azure Backup:

- Set up Recovery Services Vaults in Azure to manage and organize your backups.
- Define backup policies specifying the frequency and retention of backups.
- Backup your Azure Virtual Machines including OS and data disks.
- Backup managed disks without impacting production application performance.
- Backup Azure file shares to ensure data is protected.
- Use Backup Center to manage backups across multiple vaults and subscriptions.
- Generate reports to monitor backup status and performance.

Azure Site Recovery:

- Set up replication for on-premises virtual machines, Azure virtual machines, and physical servers to Azure.
- Define replication policies to meet your Recovery Time Objective and Recovery Point Objective.
- Group machines based on dependencies and recovery order.
- Add scripts or manual actions to automate recovery tasks.
- Regularly test failover to ensure disaster recovery processes work as expected.
- Plan and execute failback to restore services to the primary site after a disaster.

Data Visualisation:

PowerBI can be used in the following ways to help with data visualisations:



- Power BI can be used to create detailed reports that highlight key metrics such as sales performance, customer demographics, and inventory levels.
- It can create interactive visualizations like charts, graphs, and maps to represent your data visually. This makes it easier for management to understand trends and patterns.
- Reports can be published to the Power BI Service, making them accessible to your team and management.
- Dashboards can be made to provide a consolidated view of your key metrics and can be customized to show real-time data.
- Alerts can be configured to notify management of significant events or changes in key metrics, helping in making timely decisions.

Future Scalability:

Azure has the following scalability features for growing businesses:

- Auto Scaling which automatically adjust the number of virtual machines based on demand. This ensures that your applications have the necessary resources during peak times and scales down during low usage periods.
- Virtual Machine Scale Sets which can create and manage a group of identical virtual machines, allowing you to scale out applications seamlessly.
- Azure allows you to deploy applications closer to users to reduce latency and improve performance.
- Can automatically scale compute and storage resources based on workload demands. This ensures that your database can handle increased data volumes and complex queries without manual intervention.
- Azure provides a fully managed SQL Server instance with built-in high availability and disaster recovery.
- Can store petabytes of data in Azure Data Lake Storage, making it suitable for big data scenarios.
- Can use data partitioning and indexing to optimize query performance and handle large datasets efficiently.
- Can use DevOps tools for continuous integration and deployment, ensuring that your applications are always up-to-date and scalable.



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.

