**Application accessing data hosted on ADLS Gen2**

Develop an application that securely and efficiently accesses, retrieves, and manages data stored in Azure Data Lake Storage (ADLS) Gen2. The application should handle large-scale data operations, ensure compliance with data access policies, and provide robust performance and error handling mechanisms.

|  |
| --- |
| **Data** |

Data is crucial for AI as it drives model training, ensuring accuracy and relevance.   
Without high-quality, abundant data, AI systems cannot learn effectively or make informed decisions.

We need to understand, the velocity of data, volume of data, data sources, do we want to process data in batches or do we want to process data as it comes? And we need some place to store data?

The initial raw form of data may not be that useful  
We need to take our data, clean it, convert into a form in which we can analyze it further

In the end, we need to create reports and be able to visualize our data.

|  |
| --- |
| **The Different data formats** |

The simplest way in which we can store data is CSV format

Then we have data that can be represented in semi-structured way and JSON is a classic example of this. Data is split into name- value pairs

Then we take data and host as tables in a relational database.

Those tables have row and columns of information.

|  |
| --- |
| **Different Types of Storage services in Azure** |

Azure offers several storage services, each designed to meet different needs. Here are the main types:

**Azure Blob Storage:**

Purpose: Store unstructured data like documents, images, and backups.

Types: Block blobs, Append blobs, and Page blobs.

**Azure File Storage:**

Purpose: Provide file shares that can be accessed via the SMB protocol.

Use Cases: Lift-and-shift applications, shared file storage.

**Azure Queue Storage:**

Purpose: Store and manage large numbers of messages for asynchronous processing.

Use Cases: Decoupling components in a distributed application.

**Azure Table Storage:**

Purpose: Store structured NoSQL data with a key-attribute store.

Use Cases: Storing large amounts of structured data that require high availability.

**Azure Disk Storage:**

Purpose: Provide durable, high-performance disks for Azure Virtual Machines.

Types: Standard HDD, Standard SSD, Premium SSD, and Ultra Disk.

**Azure Data Lake Storage (ADLS) Gen2:**

Purpose: Provide scalable and secure storage for big data analytics.

Use Cases: Data warehousing, big data analytics, and data lakes.

**Azure Archive Storage:**

Purpose: Store infrequently accessed data at a lower cost.

Use Cases: Long-term backup and archival storage.

**Azure NetApp Files:**

Purpose: Provide high-performance, enterprise-grade file storage.

Use Cases: High-performance applications and database workloads.

Each of these services is optimized for different use cases, allowing you to choose the best option based on performance, cost, and access needs.

|  |
| --- |
| **Example of Application accessing data** |

A simple program in Python that can be used to access the data that is stored in cloud.

1. Uploading data to ADLS Gen2
2. We can read the CSV file using the Pandas library.