

THINGS TO ANALYSE

Step 1: Data Preparation

1. Structured Data: Gather structured data in tabular formats like CSV files, which could include user profiles, sales transactions, or inventory records.	https://www.kaggle.com/datasets/nelgiryewithana/top-spotify-songs-2023
2. Semi-Structured Data: Collect semi-structured data in formats such as JSON or XML. Examples include API responses, configuration files, or log files with varying schemas.	json
3. Unstructured Data: Obtain unstructured data, which can encompass text documents, images, audio files, or videos.	audio: https://intmusic.net/262550/peggy-gou-it-goes-like-nanana-2023 Image: https://ra.co/reviews/35539

Step 2: Azure Storage creation

Azure Data Lake Storage Account: If you don't already have one, create an Azure Data Lake Storage account through the Azure portal.

The screenshot displays the Microsoft Azure portal interface. At the top, the navigation bar includes the 'Microsoft Azure' logo, an 'Upgrade' button, a search bar, and user account information. Below the navigation bar, the 'Azure services' section features icons for various services: 'Create a resource', 'Quickstart Center', 'Virtual machines', 'App Services', 'Storage accounts', 'SQL databases', 'Azure Cosmos DB', 'Kubernetes services', 'Function App', and 'More services'. The 'Resources' section shows a table with columns for 'Name', 'Type', and 'Last Viewed', but it is currently empty with the message 'No resources have been viewed recently'. The 'Navigate' section includes links to 'Subscriptions', 'Resource groups', 'All resources', and 'Dashboard'. Below this, a 'Marketplace' modal is open, displaying search results for 'storage account'. The search results are filtered by 'Pricing: All', 'Operating System: All', 'Publisher Type: All', 'Product Type: All', and 'Publisher name: All'. The results list several storage-related services, including 'Storage account' (Microsoft), 'Storage Account Using ARM Template' (FortuneCloud LLC), 'Azure Storage Mover' (Microsoft), 'Storage Account Using ARM' (DIGISTORM LTD.), 'Azure Storage solution for Sentinel' (Microsoft Sentinel), 'Storage Account Using ARM' (DIGISTORM LTD.), 'S3 API for Azure Blob' (FortuneCloud LLC), 'APEX Protection Storage for' (Dell), 'FileMaaS SFTP / FTP & Web' (FortuneCloud LLC), and 'FileMaaS SFTP / FTP & Web' (FortuneCloud LLC). Each result card includes a 'Create' button and a heart icon for favorites.

Create a storage account



Basics Advanced Networking Data protection Encryption Tags Review

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more about Azure storage accounts](#)

Project details

Select the subscription in which to create the storage account. You can also select a resource group to organize and manage your storage account together with other resources.

Subscription *

Resource group *

A resource group is a container that holds related resources for an Azure solution.

Name *

spotify

OK

Cancel

Instance details

Review

< Previous

Next : Advanced >

Give feedback

Create a storage account



Basics Advanced Networking Data protection Encryption Tags Review

[Create new](#)

Instance details

Storage account name ⓘ *

data

Region ⓘ *

(Europe) UK West

[Deploy to an edge zone](#)

Performance ⓘ *

☒ **Standard:** Recommended for most scenarios (general-purpose v2 account)

☐ **Premium:** Recommended for scenarios that require low latency.

Redundancy ⓘ *

Geo-redundant storage (GRS)

☒ Make read access to data available in the event of regional unavailability.

Review

< Previous

Next : Advanced >

Give feedback

Step 3: Data Upload

- **Structured Data Upload:**
 - Use Azure Storage Explorer, Azure Data Factory, or AzCopy to upload structured data (CSV files) into a designated container within your Azure Data Lake Storage account.
- **Semi-Structured Data Upload:**
 - Similarly, use Azure Storage Explorer, Azure Data Factory, or AzCopy to upload semi-structured data (e.g., JSON or XML files) to a separate container within Azure Data Lake Storage.
- **Unstructured Data Upload:**
 - For unstructured data, use Azure Storage Explorer or relevant Azure SDKs to upload the data to its respective container.

spotifydata123 | Containers

Storage account

Search

+ Container Change access level Restore containers Refresh Delete Give feedback

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Data storage

Containers

File shares

Queues

Tables

Security + networking

Networking

Access keys

Search containers by prefix

Name	Last modified	Anonymous s
<input type="checkbox"/> \$logs	01/10/2023, 20:40:18	Private

New container

Name *

spotify

Anonymous access level ⓘ

Private (no anonymous access)

The access level is set to private because anonymous access is disabled on this storage account.

Advanced

Create

Give feedback

spotify

Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease

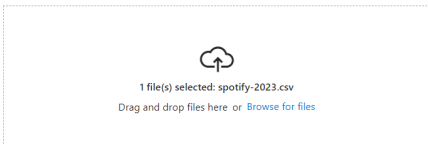
Authentication method: Access key (Switch to Azure AD User Account)

Location: spotify

Search blobs by prefix (case-sensitive)

Name	Modified	Access tier	Arch
No results			

Upload blob



Overwrite if files already exist

Advanced

Upload

Give feedback

spotifymusic

Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Authentication method: Access key (Switch to Azure AD User Account)

Location: spotifymusic

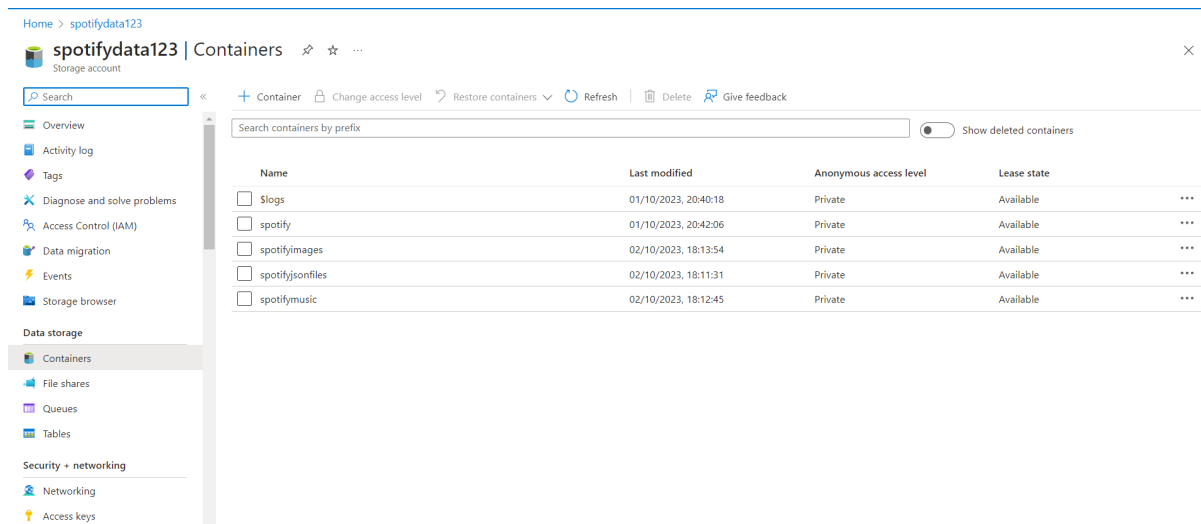
Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state	
<input type="checkbox"/> Peggy Gou - (It Goes Like) Nanana (Edit) Lyrics Ge...	02/10/2023, 18:13:15	Hot (Inferred)		Block blob	97.43 KiB	Available	...

Step 4: Verification

Data Validation: Confirm that the data has been successfully uploaded by using Azure Storage Explorer to inspect the presence of files within the containers.



Step 5: Access Control

Establish access controls and permissions to ensure that only authorised individuals or applications can access and manipulate the data stored in Azure Data Lake Storage. Here I am giving permission to a member.

Assign Role-Based Access Control (RBAC) Roles:

- Navigate to your Azure Data Lake Storage account in the Azure portal.
- Under "Access control (IAM)," add role assignments.
- Select Members, Choose member, then Next
- Then click on the button Review + Sign button

Home > spotifydata123

spotifydata123 | Access Control (IAM)

☆ ...

Storage account

access control

+ Add

Download role assignments

Edit columns

Refresh

Remove

Feedback

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Data storage

Containers

File shares

Queues

Tables

Security + networking

Networking

Access keys

Shared access signature

Encryption

Check access

Role assignments

Roles

Deny assignments

Classic administrators

My access

View my level of access to this resource.

View my access

Check access

Review the level of access a user, group, service principal, or managed identity has to this resource. [Learn more](#)

Check access

Grant access to this resource

Grant access to resources by assigning a role. [Learn more](#)

Add role assignment

View access to this resource

View the role assignments that grant access to this and other resources. [Learn more](#)

View

View deny assignments

View the role assignments that have been denied access to specific actions at this scope. [Learn more](#)

View

Home > spotifydata123 | Access Control (IAM)

Add role assignment

Role

Members

Review + assign

Assign access to

User, group, or service principal

Managed identity

Members

+ Select members

Name

Object ID

Type

No members selected

Description

Optional

Review + assign

Previous

Next

Feedback

Step 6: Documentation

Maintain comprehensive documentation that includes details about data sources, container structures, access policies, and any other relevant information about the project.

Project title	spotifydata123
Data sources	Structured data- data file sourced from a CSV file. Semi-structured data-track details in json format Unstructured data-audio music downloaded from online Unstructured data-album cover image downloaded from online

Data format	Structured data-csv Semi structured json Unstructured data-mp4 Unstructured data-.jpeg
Contain structure	Structured data-stored in "spotify" container Semi structured data-stored in "spotifyjsonfiles" container Unstructured data-stored in "spotifymusic" container Unstructured data -stored in "spotifyimages" container
Access policies	Structured data-Read access for analysts, write access for data engineers. Semi structured data-Read and write access for the data engineering team. Unstructured data-Read access granted to NLP researchers Unstructured data- Read access granted to NLP researchers
Data upload procedures	Structured Data -Uploaded using Azure Storage Explorer with "spotify.csv." Semi-Structured Data -details fetched and stored as JSON files using Azure Data Factory. Unstructured Data -downloaded audio data was downloaded and uploaded via Azure Storage Explorer. Unstructured Data -downloaded image data was downloaded and uploaded via Azure Storage Explorer.
Verification Process	Structured Data-Verified by confirming the presence of "spotify.csv" in the "spotify" container. Semi-Structured Data - Checked for the presence of JSON files in the "spotifyjsonfiles" container. Unstructured Data -Validated by inspecting the uploaded audio document in the "spotifymusic" container. Unstructured Data -Validated by inspecting the uploaded image document in the "spotifyimages" container.
Security Measures	Structured Data-Token-based authentication used Semi-Structured Data-Token-based authentication used Unstructured Data -audio data was anonymized to protect user identities.

	Unstructured Data -image data was anonymized to protect user identities.
Challenges and Solutions	Challenge - Varying JSON structures in API responses. Solution-Developed a flexible parser to handle different JSON structures.
Dependencies and Technologies	Tools -Azure Storage Explorer, Azure Data Factory, Python (for data processing), Azure Data Lake Storage.
Contact Information	Project Lead Athy Email Athymol.com
Additional contacts for access requests	Data analyst zara Email Zara.com Data engineer Ryan Email Ryan.com Data scientist Mark Email Mark.com
Future Considerations Tasks	Implement data processing pipelines to extract insights from semi-structured Spotify data. Enhance security with data encryption at rest. Explore advanced analytics on music review sentiment using Azure tools.