

Name: Rohan Vinayak Chaudhari

Batch: Data Engineering

Date: 15/02/2024

Topic: Azure DataBricks

Solution:

1. Azure Databricks (Storage account):

Data lake is central repository that stores raw data in its native format.

* **DDL**

* **Delta Lake table:**

- ↳ Delta live table
- ↳ Delta lake
 - ↳ It is an open source storage layer that brings reliability to data lakes.
 - ↳ Delta lake provide ACID transaction, scalable metadata handling & unified streaming & batch data processing.
 - ↳ Delta lake run on top of lake.

Process:

```
graph LR; DeltaTable[Delta table] --> DatabricksBox[Upsparki  
Databricks  
Delta lake]; DatabricksBox --> StorageLayer[Storage layer]; StorageLayer --> DataLake[Data lake];
```

* **Data lake:-**

- ↳ It is a repository of data from different sources.
- ↳ Data lakes has the ability to store variety of data eg:- JSON, BSON, CSV, TSV, ORC, Parquet.
- ↳ It is similar to DW but it is used in ELT.
- ↳ In DW we can keep only transformed data.
- ↳ cheaper storage.
- ↳ AWS S3, Azure or storage Gen3, Google cloud stor.

Page No. _____
Date _____

* Database vs Data Warehouse vs Delta Lake vs Data Lake

* DB: ↳ collectⁿ of data & info. (OLTP)

↳ DBMS stores data in DB & enable user to interact with it

↳ DB → Database + DBMS

↳ characteristic → ACID, Flexible, Full text search, optimized

SQL editor → queries → run query → store in DB

* Data warehouse:-

↳ It is a system that stores highly structured information from various sources.

↳ It stores current & historical data

↳ It is a giant DB optimize for analytics

↳ eg Amazon Redshift, Azure Synapse, Google BigQuery, IBM DB

* OLAP + DW + Data Lake:-

↳ Both DW & DL are meant to support OLAP

↳ OLAP are used to collect data from variety of sources then it used for range of Analytical use cases & BI & reporting

↳ DW characteristic:-

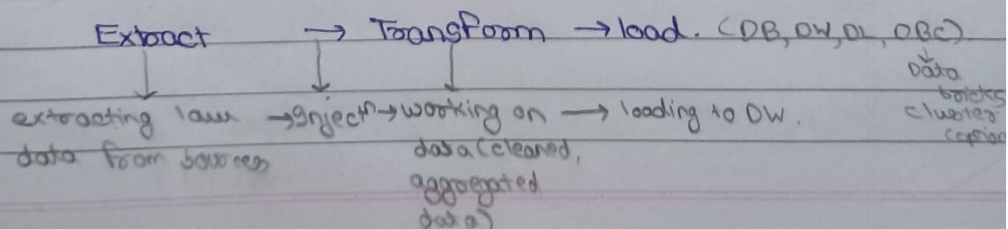
Injection

↳ ETL process data from its original source to DW

↳ DW have predefined or fix schema

↳ Business Analyst connect DW with BI

ETL



* The key Diff. between DW, DB & DL:-

- ↳ A DB stores current data required to power an app
- ↳ A DW stores current & historical data from one or more systems with fixed schema.
- ↳ A DL stores current & historical data from one or more system in its raw form.

	DB	DL	DW
Workload	operational & transactional	Analytical	Analytical
Schema	Rigid or Flexible schema	No schema	pre defined or fixed schema
Data	Real time	May be not up to date based on ETL Req.	same as DL
Datatype	structured or semi structured	both & unstructured	structured or semi structured

similar to E3 in Amazon.

* ^{color} Storage account:-

↳ To upload data

↳ Storage Acct

↳ container [create new]

↳ upload file

To get access in local

Shared access ↓ Download Storage Explorer

↓ VAST taken

copy blob service url

You can upload

Also from Storage Explorer also.

Creating Storage account:

Microsoft Azure

Search resources, services, and docs (G+)

azuser1076_mml.local@...
BHT (BHT@ONMICROSOFT.COM)

Home >

hexadeb1076_1707991428414 | Overview

Deployment

Search

«

Delete

Cancel

Redeploy

Download

Refresh

Overview

Inputs

Outputs

Template

✓ Your deployment is complete

Deployment name: hexadeb1076_1707991428414

Subscription: Azure subscription 1

Resource group: rg-azuser1076_mml.local-di1ea

Start time: 15/02/2024, 15:33:52

Correlation ID: 8494dbdf-8065-43cf-a66b-3d7514001916

Deployment details

Next steps

Go to resource

Give feedback

Tell us about your experience with deployment

Cost Management

Get notified to stay within your budget and prevent unexpected charges on your bill.

Set up cost alerts >

Microsoft Defender for Cloud

Secure your apps and infrastructure

Go to Microsoft Defender for Cloud >

Free Microsoft tutorials

Start learning today >

Work with an expert

Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.

Find an Azure expert >

Creating Container:

Microsoft Azure

Search resources, services, and docs (G+)

azuser1076_mml.local@...
BHT (BHT@ONMICROSOFT.COM)

Home > hexadeb1076_1707991428414 | Overview > hexadeb1076

hexadeb1076 | Containers

Storage account

Search

«

+ Container

Change access level

Restore containers

Refresh

Delete

Give feedback

Search containers by prefix

Show deleted containers

Name	Last modified	Anonymous access level	Lease state	
<input type="checkbox"/> \$logs	15/02/2024, 15:34:23	Private	Available	...
<input type="checkbox"/> firstcontainer1076	15/02/2024, 15:35:36	Private	Available	...

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Storage Mover

Data storage

Containers

File shares

Queues

Tables

✓ Successfully created storage container

Successfully created storage container 'firstcontainer1076'.

Uploading file:

Microsoft Azure

Search resources, services, and docs (G+)

azure1076 mm1.local@...
BHT (BHTLONMICROSOFT.COM)

Home > hexadeb1076_1707991428414 | Overview > hexadeb1076 | Containers >

firstcontainer1076
Container

Search

Upload Change access level Refresh Delete Change tier Acquire lease

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: firstcontainer1076
Search blobs by prefix (case-sensitive)
Add filter

Name	Modified	Access tier	Archive status
No results			

Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Access policy
Properties
Metadata

Upload blob

1 file(s) selected: output_file.csv
Drag and drop files here or Browse for files

☐ Overwrite if files already exist
Advanced

Upload Give feedback

File uploaded:

Microsoft Azure

Search resources, services, and docs (G+)

azure1076 mm1.local@...
BHT (BHTLONMICROSOFT.COM)

Home > hexadeb1076_1707991428414 | Overview > hexadeb1076 | Containers >

firstcontainer1076
Container

Search

Upload Change access level Refresh Delete Change tier Acquire lease Break lease View snapshots Create snapshot Give feedback

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: firstcontainer1076
Search blobs by prefix (case-sensitive) Show deleted blobs
Add filter

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/> output_file.csv	15/02/2024, 15:37:21	Hot (Inferred)		Block blob	61.42 KiB	Available ***

Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Access policy
Properties
Metadata

Viewing File:

The screenshot shows the Microsoft Azure portal interface. The left sidebar displays the navigation menu for the 'firstcontainer1076' container, including options like Overview, Upload, Change access level, and Search. The main content area shows the 'output_file.csv' blob. The blob's details are displayed, including its name, location, and a list of rows. The rows contain data for passengers, including their names, ages, and other details. The interface also shows a 'Preview' button for the blob.

PassengerId	Survived	Pclass	Name	Sex	Age	Sibsp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.25		S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38.0	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925		S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.05		S
6	0	3	Moran, Mr. James	male	0.0	0	0	330877	8.4583		Q
7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.075		S
9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333		S
10	1	2	Hassner, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708		C
11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.766		S
12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.55	C103	S
13	0	3	Saunderscock, Mr. William Henry	male	20.0	0	0	A/5. 2151	8.05		S
14	0	3	Andersson, Mr. Anders Johan	male	39.0	1	5	347082	31.275		S
15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0	0	350406	7.8542		S
16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55.0	0	0	248706	16.0		S
17	0	3	Rice, Master. Eugene	male	2.0	4	1	382652	29.125		Q

Uploading from Storage explorer:

The screenshot shows the 'Connect to Azure Storage' dialog box. The 'Select Connection Method' step is active, showing three options: 'Sign in using OAuth', 'Shared access signature URL (SAS)', and 'Anonymously (my blob container allows public access)'. The 'Shared access signature URL (SAS)' option is selected. The dialog box also includes a 'Back' button, a 'Next' button, and a 'Cancel' button.

Select Connection Method

Select Resource > **Select Connection Method** > Enter Connection Info > Summary

How will you connect to the blob container?

- ☐ Sign in using OAuth
- ☒ Shared access signature URL (SAS)
- ☐ Anonymously (my blob container allows public access)

Back Next Cancel

Entered Blob service url:

Connect to Azure Storage

Enter Connection Info

Select Resource > Select Connection Method > Enter Connection Info > Summary

Display name:

firstcontainer1076

Service URL:

https://hexadeb1076.blob.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=sco&sp=rwdlacupiytx&se=2024-02-15T18:14:05Z&st=2024-02-15T10:14:05Z&spr=https&sig=0%2FNW%2BSwQsKghswXBk5jhB9rllBwfBCQaANhRwjBJIC0%3D

Back

Next

Cancel

Storage account connected:

Microsoft Azure Storage Explorer

File Edit View Help

EXPLORER

Search for resources

Collapse all

Refresh all

Quick Access

Emulator & Attached

Storage Accounts

Emulator & Attached

Emulator - Default Ports (Key)

firstcontainer1076 (SAS)

firstcontainer1076 (SAS)

Blob Containers

firstcontainer1076

File Shares

Queues

Tables

firstcontainer1076

Get Started

Upload

Download

Open

Preview

New Folder

Select All

Copy

Paste

Clone

Properties

Delete

Undelete

Manage History

Folder Statistics

Refresh

Active blobs (default)

firstcontainer1076

Name	Access Tier	Access Tier Last Modified	Last Modified	Blob Type	Content Ty
output_file.csv	Hot (inferred)		2/15/2024 3:37 PM	Block Blob	text/csv

Showing 1 to 1 of 1 cached items

Actions

Properties

URL

Custom Domain

Type

HNS Enabled

Lease State

Lease Status

Public Read Access

Last Modified

https://hexadeb1076.blo

Blob Container

false

available

unlocked

off

Thursday, February 15

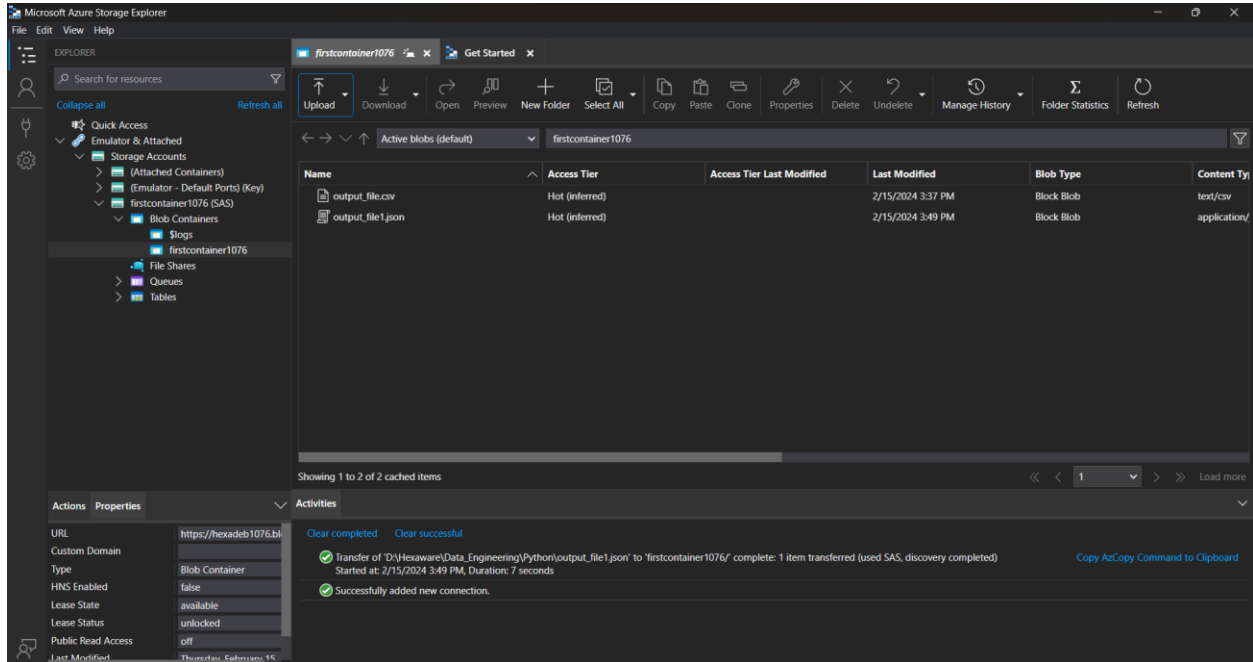
Activities

Clear completed

Clear successful

Successfully added new connection.

Uploaded New file from Explorer:



Viewing the uploaded file from Azure:

