

Web server when lifting Docker compose for airflow:

Error when using the DB Upgrade command (deprecated) replacing it with DB Migrate.
And then install Flask-session==0.5.0 to avoid incompatibility with error-throwing airflow dependencies:

```
2024-02-16 10:33:17 _create_flask_session_tbl(engine.url) 2024-02-16 10:33:17 File
"/home/airflow/.local/lib/python3.9/site-packages/airflow/utils/db.py", line 711, on
_create_flask_session_tbl 2024-02-16 10:33:17 db = _get_flask_db(sql_database_uri) 2024-02-16
10:33:17 File "/home/airflow/.local/lib/python3.9/site-packages/airflow/utils/db.py", line 700, on
_get_flask_db 2024-02-16 10:33:17 AirflowDatabaseSessionInterface(app=flask_app, db=db,
table="session", key_prefix="") 2024-02-16 10:33:17 TypeError: __init__() 6 required positional
arguments are missing: 'sequence', 'schema', 'bind_key', 'use_signer', 'permanent' and 'sid_length'
```

```
[+] Building 0.0s (0/0)
[+] Running 2/2
✓ Container footballrvmdatengineering-postgres-1  Running
✗ Container footballrvmdatengineering-webserver-1  Error
dependency failed to start: container footballrvmdatengineering-webserver-1 exited (1)
(venv) PS C:\Users\jrver\PycharmProjects\Python\FootballRVMDataEngineering> pip install Flask-Session==0.5.0
Collecting Flask-Session==0.5.0
  Obtaining dependency information for Flask-Session==0.5.0 from https://files.pythonhosted.org/packages/3e/d4/b374183251054928ddb5e72f9a0d3d764d0f2af0638fbfdf205df26e55e3/flask_session-0.5.0-py3-none-any.whl.metadata (1.4 kB)
  Downloading flask_session-0.5.0-py3-none-any.whl.metadata (1.4 kB)
Requirement already satisfied: flask>=2.2 in c:\users\jrver\pycharmprojects\python\footballrvmdatengineering\venv\lib\site-packages (from Flask-Session==0.5.0) (2.2.5)
Requirement already satisfied: cachelib in c:\users\jrver\pycharmprojects\python\footballrvmdatengineering\venv\lib\site-packages (from Flask-Session==0.5.0) (0.9.0)
Requirement already satisfied: Werkzeug>=2.2.2 in c:\users\jrver\pycharmprojects\python\footballrvmdatengineering\venv\lib\site-packages (from flask>=2.2->Flask-Session==0.5.0) (2.2.3)
```

The name of the table in the class tag put only "wikitable sorted" because if you put the full name that appears in source code html you won't find it

The screenshot illustrates a comparison between a browser's rendered view of a table and its corresponding HTML source code.

Browser View (Left):

- A green header row contains the text "association football stadium in the world."
- An orange row contains a blue note: "Denotes the largest association football stadium in or area".
- A white row contains the table's class information: "table.wikitable.sortable.jquery-tablesorter" and its dimensions: "999.92 × 68480.1".
- The main table has a yellow header row labeled "List of".
- Two rows of data are visible:
 - Rank 376: Huizhou Olympic Stadium, Seating capacity 40,000, Region East /
 - Rank 375: Wuhu Olympic Stadium, Seating capacity 40,000, Region East /

Source Code (Right):

```
<div id="mw-content-text" class="mw-body-content">
<div class="mw-content-ltr mw-parser-output" lang="en" dir="ltr">
    <p></p>
    <style data-mw-deduplicate="TemplateStyles:r1033289096">...</style>
    <div role="note" class="hatnote navigation-not-searchable">...</div>
    <style data-mw-deduplicate="TemplateStyles:r1097763485">...</style>
    <table class="box-More_citations_needed plainlinks metadata ambox ambox-content ambox-R...
        <p>...</p>
        <ul>...</ul>
        <table class="wikitable">...</table>
        <p>...</p>
    <table class="wikitable sortable jquery-tablesorter"> == $0
        <caption>List of association football (soccer) stadiums by capacity </caption>
        <thead>
            <tr>
                <th class="headerSort headerSortDown" tabindex="0" role="columnheader button" title="Sort ascending"...
                <th class="headerSort" tabindex="0" role="columnheader button" title="Sort ascending"...
        </thead>
        <tbody>
            <tr>
                ...
            </tr>
            <tr>
                ...
            </tr>
        </tbody>
    </table>
</div>
</div>
```

```
[2024-02-16, 22:23:55 UTC] {task_command.py:423} INFO - Running <TaskInstance: wikipedia_flow.transform_wikipedia_data manual__2024-02-16T22:23:45.499018+00:00 [running]> on host a3c816970f1c
[2024-02-16, 22:23:55 UTC] {taskinstance.py:2480} INFO - Exporting env vars: AIRFLOW_CTX_DAG_OWNER='Rafa Vera' AIRFLOW_CTX_DAG_ID='wikipedia_flow' AIRFLOW_CTX_TASK_ID='transform_wikipedia_data' AIRFLOW_CTX_E
[2024-02-16, 22:23:55 UTC] {base.py:396} INFO - Received an HTTP error (403): <html>
<head>
<title>Access blocked</title>
</head>
<body>
<h1>Access blocked</h1>
<p>You have been blocked because you have violated the
<a href="https://operations.osmfoundation.org/policies/nominatim/">usage policy</a>
of OSM's Nominatim geocoding service. Please be aware that OSM's resources are
limited and shared between many users. The usage policy is there to ensure that
the service remains usable for everybody.</p>
<p>Please review the terms and make sure that your
software adheres to the terms. You should in particular verify that you have set a
<b>custom HTTP referrer or HTTP user agent</b> that identifies your application, and
that you are not overusing the service with massive bulk requests.</p>
<p>If you feel that this block is unjustified or remains after you have adopted
your usage, you may contact the Nominatim system administrator at
nominatim@openstreetmap.org to have this block lifted.</p>
</body>
</head>
[2024-02-16, 22:23:55 UTC] {taskinstance.py:2698} ERROR - Task failed with exception
Traceback (most recent call last):
  File "/home/airflow/.local/lib/python3.9/site-packages/geopy/geocoders/base.py", line 368, in _call_geocoder
    result = self.adapter.get_json(url, timeout=timeout, headers=req_headers)
  File "/home/airflow/.local/lib/python3.9/site-packages/geopy/adapters.py", line 472, in get_json
    resp = self.request(url, timeout=timeout, headers=headers)
```

To avoid an error at the end of the video part 1, type from in the section #handle duplicates:

```
# Handle the duplicates  
duplicates = stadiums_df[stadiums_df.duplicated(['location'])],
```

Line 96.

Although it is mentioned and corrected in the video later.

Airflow DAGs Cluster Activity Datasets Security Browse Admin Docs 23:00 UTC AA

Duration Feb 16, 22:23

00:03:36

00:01:48

00:00:00

extract_data_from_wikipedia transform_wikipedia_data write_wikipedia_data

DAG wikipedia_flow / Run 2024-02-16, 22:48:27 UTC / Task write_wikipedia_data

Clear task Mark state as... Filter Tasks

Details Graph Gantt Code Logs XCom

Layout: Left -> Right

React Flow

```
graph LR; A[extract_data_from_wikipedia] --> B[transform_wikipedia_data]; B --> C[write_wikipedia_data]
```

	country (5)	city (6)	images (7)	home_team (8)	location (9)
1	country	city	images	home_team	location
2	North Korea	Pyongyang	https://upload.wikimedia.org/wikipedia/commons/thumb/8/81/Arirang_Mass_Games_12_IDC%22Any_Anniversary_Mass_Games_in_Pyongyang_Korea_DPR.jpg	Korea DPR national football team, Korea DPR women's national football team	[39.0496732, 125.7753215119]
3	Spain	Barcelona, Catalonia	https://upload.wikimedia.org/wikipedia/commons/thumb/0/0e/Camp_Nou_Exterior_%28_Ant_Kuman%29_Turneve_1992.jpg	FC Barcelona	[41.38089905, 2.1229225008]
4	Mexico	Mexico City	https://upload.wikimedia.org/wikipedia/commons/thumb/c/ca/Panor%C3%A1mica_aerog%C3%A1tica_de_la_Ciudad_de_M%C3%A9jico_en_2010.jpg	Club Am%C3%A9rica, Cruz Azul, Mexico national football team	[19.3029794, -99.1504699526]
5	South Africa	Johannesburg, Gauteng	https://upload.wikimedia.org/wikipedia/commons/thumb/c/c0/First_game_of_the_2010_FIFA_World_Cup%28_South_African_national_football_team_vs_Kaizer_Chiefs%29.jpg	South Africa national football team, Kaizer Chiefs	[-26.23470625, 27.9826439004]
6	Egypt	New Administrative Capital	https://upload.wikimedia.org/wikipedia/commons/thumb/e/e0/The_New_administrative_stadium_of_Egypt.jpg	Egypt national football team	[29.8921738, 31.6732269881]
7	United States	Pasadena, California	https://upload.wikimedia.org/wikipedia/commons/thumb/d/d2/Rose_Bowl%2C_Pasadena.jpg	UCLA Bruins	[34.16134915, -118.1676673983]
8	United States	Dallas, Texas	https://upload.wikimedia.org/wikipedia/commons/thumb/5/52/2019_State_Fair_Classic_at_Rosie_Rosie.jpg		[32.77961185, -96.759602529]
9	England	London, England	https://upload.wikimedia.org/wikipedia/commons/thumb/b/bb/Post_L2_Play-off_Final.jpg	England national football team	[51.55606945, -0.2796033938]
10	Qatar	Lusail	https://upload.wikimedia.org/wikipedia/commons/thumb/1/1d/Qatar_National_Football_Team_Logo.jpg	Qatar national football team	[25.4209062, 51.4904783538]

Microsoft Azure

Search resources, services, and docs (G+)

JoseRafael.Vera@TAJAMAR (TAJAMAR)

Azure services

Create a resource Storage accounts Azure Database for.. Azure Database for.. Virtual machines Cost Management ... Cost Management Firewall Manager Azure Network Function... More services

Resources

Recent Favorite

Name	Type	Last Viewed
	No resources have been viewed recently	View all resources

Navigate

Subscriptions Resource groups All resources Dashboard

Tools

Microsoft Azure

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

JoseRafael.Vera
TAJAMAR (TAJAMAR)

Home >

Storage accounts

Tajamar (tajamar365.com)

+ Create ⌂ Restore ⚙ Manage view ⏪ Refresh ⏴ Export to CSV ⚡ Open query | 🗂 Assign tags 🗑 Delete

Filter for any field... Subscription equals all Resource group equals all Location equals all Add filter

Showing 0 to 0 of 0 records.

No grouping List view

Name ↑↓	Type ↑↓	Kind ↑↓	Resource group ↑↓	Location ↑↓	Subscription ↑↓
---------	---------	---------	-------------------	-------------	-----------------



No storage accounts to display

Create a storage account to store up to 500TB of data in the cloud. Use a general-purpose storage account to store object data, use a NoSQL data store, define and use queues for message processing, and set up file shares in the cloud. Use the Blob storage account and the hot or cool access tiers to optimize your costs based on how frequently your object data is accessed.

[Create storage account](#)

[Learn more](#)

Microsoft Azure

Home > Storage accounts >
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 new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription *

Resource group *

Instance details

Storage account name

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

Region

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 >

Microsoft Azure

Home > Storage accounts >
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 new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription *

Resource group *

Instance details

Storage account name

Region

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 >

Microsoft Azure

Home > Storage accounts >
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 new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription *

Resource group *

Instance details

Storage account name

Region

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.

Microsoft Azure

Home > Storage accounts >

Create a storage account ...

Basics Advanced Networking Data protection Encryption Tags Review

Security

Configure security settings that impact your storage account.

Require secure transfer for REST API operations

Allow enabling anonymous access on individual containers

Enable storage account key access

Default to Microsoft Entra authorization in the Azure portal

Minimum TLS version

Permitted scope for copy operations (preview)

Hierarchical Namespace

Hierarchical namespace, complemented by Data Lake Storage Gen2 endpoint, enables file and directory semantics, accelerates big data analytics workloads, and enables access control lists (ACLs). [Learn more](#)

Enable hierarchical namespace

Access protocols

Blob and Data Lake Gen2 endpoints are provisioned by default. [Learn more](#)

Enable SFTP

Enable network file system v3

Blob storage

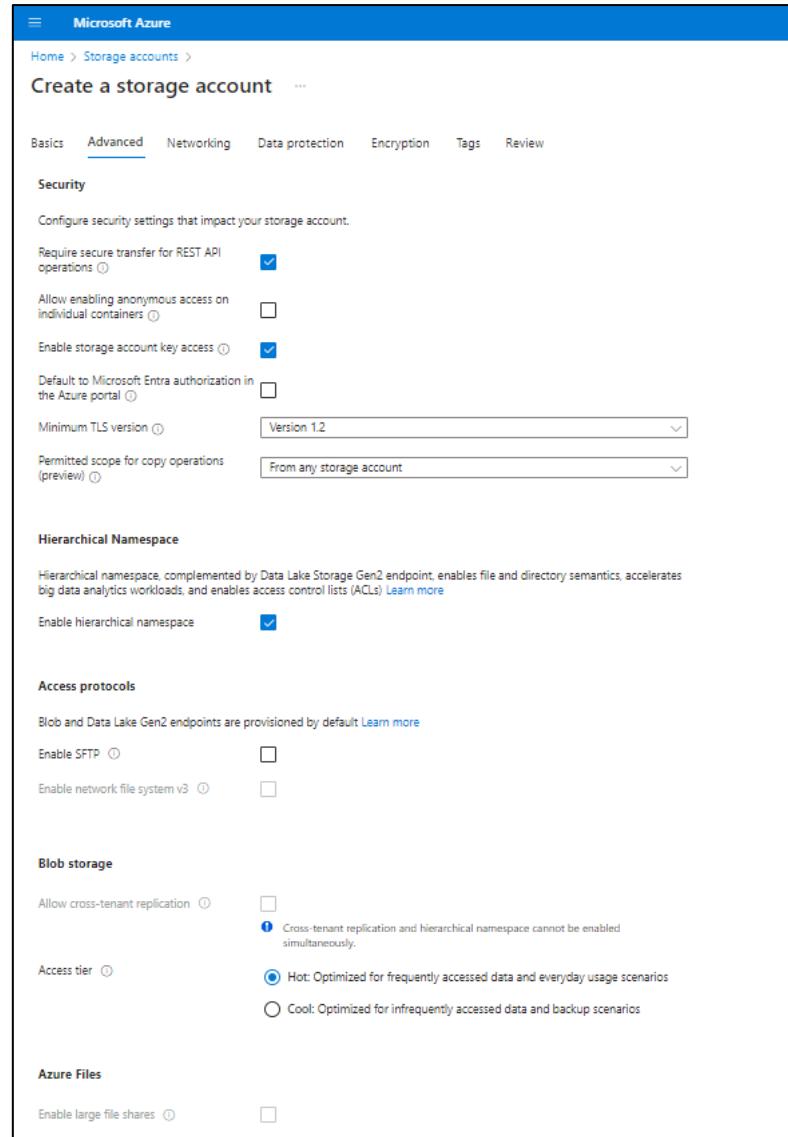
Allow cross-tenant replication

ⓘ Cross-tenant replication and hierarchical namespace cannot be enabled simultaneously.

Access tier Hot: Optimized for frequently accessed data and everyday usage scenarios
 Cool: Optimized for infrequently accessed data and backup scenarios

Azure Files

Enable large file shares



Create a storage account

Basics Advanced Networking Data protection Encryption Tags Review

Network connectivity

You can connect to your storage account either publicly, via public IP addresses or service endpoints, or privately, using a private endpoint.

Network access *

- Enable public access from all networks
 - Enable public access from selected virtual networks and IP addresses
 - Disable public access and use private access
- ⓘ Enabling public access from all networks might make this resource available publicly. Unless public access is required, we recommend using a more restricted access type. [Learn more](#)

Network routing

Determine how to route your traffic as it travels from the source to its Azure endpoint. Microsoft network routing is recommended for most customers.

Routing preference ⓘ *

- Microsoft network routing
- Internet routing

Microsoft Azure Search resources, services, and docs (G+) Home > Storage accounts > Create a storage account Initializing deployment...

Deploying...

Basics Advanced Networking Data protection Encryption Tags Review

Allow storage account key access Enabled

Default to Microsoft Entra authorization in the Azure portal Disabled

Blob anonymous access Disabled

Minimum TLS version Version 1.2

Permitted scope for copy operations (preview) From any storage account

Data protection

Point-in-time restore Disabled

Blob soft delete Enabled

Blob retention period in days 7

Container soft delete Enabled

Container retention period in days 7

File share soft delete Enabled

File share retention period in days 7

Versioning Disabled

Blob change feed Disabled

Version-level immutability support Disabled

Encryption

Encryption type Microsoft-managed keys (MMK)

Enable support for customer-managed keys Blobs and files only

Enable infrastructure encryption Disabled

Microsoft Azure

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

JoseRafael.Vera@TAJAMAR.ITAJAM...

Home > rvmfootballdataengsa_1708339491574 | Overview

Deployment

Search Delete Cancel Redeploy Download Refresh

Overview Inputs Outputs Template

Deployment is in progress

Deployment name: rvmfootballdataengsa_1708339491574 Start time: 2/19/2024, 11:45:24 AM
Subscription: Azure for Students Correlation ID: 9847eb17-2a56-4d5a-ac39-dfdccf835a2

Deployment details

Resource	Type	Status	Operation details
No results.			

Give feedback

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 providers

Home >

rvmfootballdataengsa_1708339491574 | Overview

Deployment

Search Delete Cancel Redeploy Download Refresh

Overview

Inputs

Outputs

Template

Your deployment is complete

Deployment name: rvmfootballdataengsa_1708339491574
Subscription: Azure for Students
Resource group: rvmfootballdataeng-rg

Start time: 2/19/2024, 11:45:24 AM
Correlation ID: 9847eb17-2a56-4d5a-ac39-dfdbccf835a2

Deployment details

Next steps

Go to resource

rvmfootballdataengsa

Storage account

<input type="text"/> Search
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
Shared access signature
Encryption
Microsoft Defender for Cloud
Data management
Redundancy
Data protection
Blob inventory
Static website

Upload Open in Explorer Delete Move Refresh Open in mobile CLI / PS Feedback

Essentials

Resource group (move)	:	rvmfootballdataeng-rg	Performance	:	Standard
Location	:	westeurope	Replication	:	Read-access geo-redundant storage (RA-GRS)
Primary/Secondary Location	:	Primary: West Europe, Secondary: North Europe	Account kind	:	StorageV2 (general purpose v2)
Subscription (move)	:	Azure for Students	Provisioning state	:	Succeeded
Subscription ID	:	9e3c713b-0d42-41e5-81ca-d778fd001da7	Created	:	19/2/2024, 11:45:26
Disk state	:	Primary: Available, Secondary: Available			

[Tags \(edit\)](#) : Add tags

Properties Monitoring Capabilities (5) Recommendations (0) Tutorials Tools + SDKs

Data Lake Storage

Hierarchical namespace	Enabled
Default access tier	Hot
Blob anonymous access	Disabled
Blob soft delete	Enabled (7 days)
Container soft delete	Enabled (7 days)
Versioning	Disabled
Change feed	Disabled
NFS v3	Disabled
SFTP	Disabled

File service

Large file share	Disabled
Identity-based access	Not configured
Default share-level permissions	Disabled
Soft delete	Enabled (7 days)
Share capacity	5 TiB

Security

Require secure transfer for REST API operations	Enabled
Storage account key access	Enabled
Minimum TLS version	Version 1.2
Infrastructure encryption	Disabled

Networking

Allow access from	All networks
Number of private endpoint connections	0
Network routing	Microsoft network routing
Access for trusted Microsoft services	Yes
Endpoint type	Standard

Microsoft Azure

Search resources, services, and docs (G+/-)

Home > rvmfootballdataengsa_1708339491574 | Overview > rvmfootballdataengsa

rvmfootballdataengsa | 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

Search containers by prefix

Name Last modified Anonymous access level

\$logs 2/19/2024, 11:45:49 AM Private

New container

Name * rvmfootballdataeng

Anonymous access level Private (no anonymous access)

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

Advanced

Name	Last modified	Anonymous access level
\$logs	2/19/2024, 11:45:49 AM	Private

Microsoft Azure

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

Home > rvmfootballdataengsa_1708339491574 | Overview > rvmfootballdataengsa | Containers >

rvmfootballdataeng ...

Container

Search

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

Successfully added directory
Successfully added directory 'output'.

Overview

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

Diagnose and solve problems

Access Control (IAM)

Search blobs by prefix (case-sensitive)

Show deleted objects

Settings

Shared access tokens

Manage ACL

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/> data					-	
<input type="checkbox"/> output					-	

Microsoft Azure

Home > Resource groups > rvmfootballdataeng-rg > rvmfootballdataengsa

rvmfootballdataengsa | Access keys

Storage account

Search

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Containers

File shares

Queues

Tables

Networking

Access keys

Shared access signature

Encryption

Set rotation reminder Refresh Give feedback

Access keys authenticate your applications' requests to this storage account. Keep your keys in a secure location like Azure Key Vault, and replace them often with new keys. The two keys allow you to replace one while still using the other.

Remember to update the keys with any Azure resources and apps that use this storage account.
[Learn more about managing storage account access keys](#)

Storage account name: rvmfootballdataengsa

key1 Rotate key
Last rotated: 19/2/2024 (0 days ago)
Key: Show

Connection string: Show

key2 Rotate key
Last rotated: 19/2/2024 (0 days ago)
Key: Show

Connection string: Show

localhost:8080/dags/wikipedia_flow/grid?dag_run_id=manual_2024-02-19T12%3A37%3A56.619568%2B00%3A00&task_id=write_wikipedia_data&tab=logs

APIs AWS BBDD Empleo Forecasting GITHUB Learning My Web NOISE PROJECT OPENAI Python Snowflake Spark Projects TFM Tsystems Utilities Water Project Historial FLO Articulos

17 f? 12:44 UTC AA

DAG: wikipedia_flow

Schedule: None Next Run: None

Grid Graph Calendar Task Duration Task Tries Landing Times Gantt Details Code Audit Log

19/02/2024 12:38:11 25 All Run Types All Run States Clear Filters Auto-refresh

Press shift + / for Shortcuts

Duration: 00:03:44

00:01:52

00:00:00

extract_data_from_wikipedia
transform_wikipedia_data
write_wikipedia_data

DAG wikipedia_flow / Run 2024-02-19, 12:37:56 UTC / Task write_wikipedia_data

Details Graph Gantt Code Logs XCom

by attempts:

1

All Levels All File Sources Wrap Download See More

x-ms-client-request-id: '3afe2400-cf24-11ee-bc1b-0242ac130004'
x-ms-version: 'REDACTED'
x-ms-content-crc64: 'REDACTED'
x-ms-request-server-encrypted: 'REDACTED'
Date: 'Mon, 19 Feb 2024 12:41:49 GMT'
[2024-02-19, 12:41:41 UTC] {universal.py:511} INFO - Request URL: "https://rvmfootballdataeng.blob.core.windows.net/rvmfootballdataeng/data/stadium_cleaned2024-02-19_12_41_40.398458.csv?comp=REDACTED"
Request method: 'PUT'
Request headers:
Content-Length: '91'
x-ms-meta-is_directory: 'REDACTED'
x-ms-version: 'REDACTED'
Content-Type: 'application/xml'
Accept: 'application/xml'
x-ms-date: 'REDACTED'
x-ms-client-request-id: '3b11123a2-cf24-11ee-bc1b-0242ac130004'
User-Agent: 'azsdk-python-storage-blob/12.18.2 Python/3.9.18 (Linux-5.15.133.1-microsoft-standard-WSL2-x86_64-with-glibc2.36)'
Authorization: 'REDACTED'
A body is sent with the request
[2024-02-19, 12:41:41 UTC] {universal.py:550} INFO - Response status: 201
Response headers:
Content-Length: '0'
Last-Modified: 'Mon, 19 Feb 2024 12:41:50 GMT'
ETag: '"0x8D314824C1BAC2"'
Server: 'Windows-Azure-Blob/1.0 Microsoft-HTTPAPI/2.0'
x-ms-request-id: '04e7c3f8-101e-002f-5431-637798000000'
x-ms-client-request-id: '3b11123a2-cf24-11ee-bc1b-0242ac130004'
x-ms-version: 'REDACTED'
x-ms-content-crc64: 'REDACTED'
x-ms-request-server-encrypted: 'REDACTED'
Date: 'Mon, 19 Feb 2024 12:41:49 GMT'
[2024-02-19, 12:41:41 UTC] {python.py:201} INFO - Done. Returned value was: None
[2024-02-19, 12:41:41 UTC] {taskinstance.py:1138} INFO - Marking task as SUCCESS. dag_id=wikipedia_flow, task_id=write_wikipedia_data, execution_date=20240219T123756, start_date=20240219T124140, end_date=20240219T124141
[2024-02-19, 12:41:41 UTC] {local_task_job_runner.py:234} INFO - Task exited with return code 0
[2024-02-19, 12:41:41 UTC] {taskinstance.py:3281} INFO - 0 downstream tasks scheduled from follow-on schedule check

Microsoft Azure

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

Home > rvmfootballdataengsa_1708339491574 | Overview > rvmfootballdataengsa | Containers >

rvmfootballdataeng ...

Container

Search

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

Overview

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: rvmfootballdataeng / data

Diagnose and solve problems

Access Control (IAM)

Search blobs by prefix (case-sensitive)

Settings

Shared access tokens

Manage ACL

Access policy

Properties

Metadata

Name	Modified	Access tier	Archive status
[..]	2/19/2024, 1:41:50 PM	Hot (Inferred)	
stadium_cleaned2024-02-19_12_41_40.398458.csv	2/19/2024, 1:41:50 PM	Hot (Inferred)	

Microsoft Azure

Search resources, services, and docs (G+)

JoseRafael.Vera
TAJAMAR (TAJAMAR)

Home >

Data factories

Tajamar (tajamar365.com)

+ Create Manage view Refresh Export to CSV Open query Assign tags

Filter for any field... Subscription equals all Type equals all Resource group equals all X Location equals all X Add filter

Showing 0 to 0 of 0 records.

No grouping List view

Name ↑↓	Type ↑↓	Subscription ↑↓	Resource group ↑↓	Location ↑↓
[Empty table area]				



Microsoft Azure

Home > Data factories >

Create Data Factory

Basics Git configuration Networking Advanced Tags Review + create

[View automation template](#)

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

Basics

Subscription	Azure for Students
Resource group	rvmfootballdataeng-rg
Name	rvmfootballdataengineering-df
Region	East US
Version	V2

Networking

Connect via	Public endpoint
-------------	-----------------

[Previous](#) [Next](#) [Create](#)

Microsoft Azure

Home >

Microsoft.DataFactory-20240219134813 | Overview

Deployment

Search Delete Cancel Redeploy Download Refresh

Overview Inputs Outputs Template

Your deployment is complete

Deployment name : Microsoft.DataFactory-20240219134813
Subscription : Azure for Students
Resource group : rvmfootballdataeng-rg

Deployment details Next steps

Go to resource

Give feedback

Tell us about your experience with deployment

Microsoft Azure

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

Home > Microsoft.DataFactory-20240219134813 | Overview >

rvmfootballdataengineering-df ...

Data factory (V2)

Search Delete

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Resource group (move) : rvmfootballdataeng_rg

Status : Succeeded

Location : East US

Subscription (move) : Azure for Students

Subscription ID : 9e3c713b-0d42-41e5-81ca-d778fd001da7

Type : Data factory (V2)

Getting started : [Quick start](#)

Essentials

Networking

Managed identities

Properties

Locks

Getting started

Quick start

Monitoring

Alerts

Metrics

Diagnostic settings

Logs

Quick Starts

Tutorials

Template Gallery

Training Modules

Azure Data Factory Studio

Launch studio

Monitoring

Microsoft Azure | Data Factory > rvmfootballdataengineering-df

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to learn more.

Search factor...

Home Author Monitor Manage Learning Center

Data Factory Validate all Publish all

General

Factory settings

Connections

Linked services

Integration runtimes

Microsoft Purview

Source control

Git configuration

ARM template

Author

Triggers

Global parameters

Data flow libraries

Security

Credentials

Customer managed key

Outbound rules

Managed private endpoints

Workflow orchestration manager

Apache Airflow

Linked services

Linked service defines the connection information to a data store or comp...

+ New

Filter by name Annotations : Any

Linked service defines the connection information to a data store or comp...

+ New

Filter by name Annotations : Any

Microsoft Azure | Data Factory > rvmfootballdataengineering-df

Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory Validate all Publish all

Home Author **Factory Resources**

Filter resources by name +

- Pipelines
- Change Data Capture (preview)
- Datasets
- Data flows
- Power Query

Pipeline

Change Data Capture (preview)

Dataset

Data flow

Power Query

Copy Data tool

Template gallery

Import from pipeline template

Learning Center

Microsoft Azure | Data Factory ▶ rvmfootballdataengineering-df

Search factory and documentation

1

i Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Home Author Monitor Manage Learning Center

Data Factory Validate all Publish all 1

Factory Resources

Filter resources by name +

Pipelines 1

pipeline1

Change Data Capture (preview) 0

Datasets 0

Data flows 0

Power Query 0

Activities

Validate Validate copy runtime Debug Add trigger

Search activities

Move and transform

Copy data

Data flow

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Power Query

pipeline1

Copy data

Copy data1

General Source¹ Sink¹ Mapping Settings User properties

Name * Copy data1 Learn more ▾

Description

Activity state Activated Deactivated

Timeout 0.12:00:00

Microsoft Azure | Data Factory > rvmfootballdataengineering-df

Search factory and documentation

1 Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory Validate all Publish all

Home Author Monitor Manage Learning Center

Factory Resources

Pipelines 1 pipeline1

Change Data Capture (preview) 0 Datasets 0 Data flows 0 Power Query 0

Activities

Move and transform

Copy data Data flow

Synapse Azure Data Explorer Azure Function Batch Service Databricks Data Lake Analytics General HDInsight Iteration & conditionals Machine Learning Power Query

pipeline1

Validate Validate copy runtime Debug Add trigger

Copy data

Copy data1

General Source¹ Sink¹ Mapping Settings User properties

Source dataset * Select... + New

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and data within a data store. [Learn more](#)

Select a data store

lake

All Azure Database File Generic protocol NoSQL Service

Azure Data Lake Storage Gen1	Azure Data Lake Storage Gen2	Azure Databrake Lake
Microsoft Fabric Lakehouse Files	Microsoft Fabric Lakehouse Table	Snowflake

Continue

Microsoft Azure | Data Factory > rvmfootballdataengineering-df

Search factory and documentation

1 Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory < Validate all Publish all 1

Home Author Monitor Manage Learning Center

Factory Resources < Pipeline1 Activities <

Filter resources by name +

Pipelines 1 pipeline1

Change Data Capture (preview) 0 Datasets 0 Data flows 0 Power Query 0

Move and transform

Copy data Data flow Synapse Azure Data Explorer Azure Function Batch Service Databricks Data Lake Analytics General HDInsight Iteration & conditionals Machine Learning Power Query

Copy data

Copy data1

Source dataset * Select... + N

Select format

Choose the format type of your data

 Avro	 Binary	 Delimited
 Excel	 JSON	 ORC
 Parquet	 XML	

Continue Back

Microsoft Azure | Data Factory > rvmfootballdataengineering-df

Search factory and documentation

Set properties

Name: rvm_stadiums_cleaned_timestamp

Linked service *: Select...

+ New

Copy data

Copy data1

General Source¹ Sink¹ Mapping Settings User properties

Source dataset *: Select...

OK Back

Home Author Monitor Manage Learning Center

Data Factory Validate all Publish all 1

Factory Resources

Filter resources by name +

Pipelines 1

pipeline1

Activities

Search activities

Validate Validate copy runtime Debug Add trigger

Move and transform

Copy data

Data flow

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Power Query

Microsoft Azure | Data Factory > rvmfootballdataengineering-df

Search factory and documentation

New linked service

Azure Data Lake Storage Gen2 [Learn more](#)

Name *
AzureDataLakeStorage1

Description

Connect via integration runtime * ⓘ
AutoResolveIntegrationRuntime

Authentication type
Account key

Account selection method ⓘ
 From Azure subscription Enter manually

Azure subscription ⓘ
Select all

Storage account name *
rvmfootballdataengsa

Test connection ⓘ
 To linked service To file path

Annotations
+ New

Parameters

Advanced ⓘ

Create Cancel

Home Author Monitor Manage Learning Center

Data Factory Validate all Publish all 1

Factory Resources Filter resources by name +

Pipelines 1 pipeline1

Activities Search activities

Move and transform

Copy data

Data flow

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Power Query

pipeline1

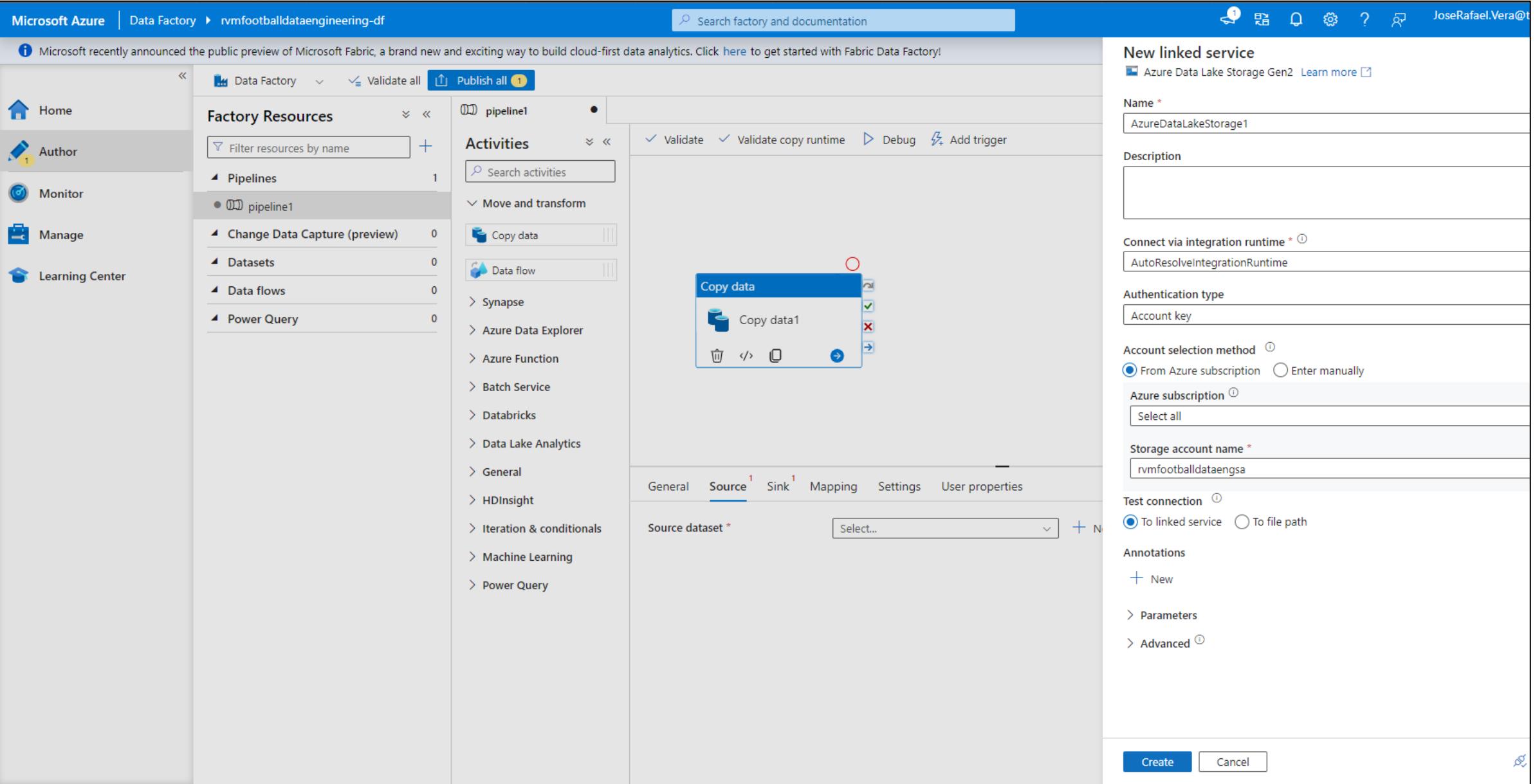
Validate Validate copy runtime Debug Add trigger

Copy data

Copy data1

General Source 1 Sink 1 Mapping Settings User properties

Source dataset * Select... + N



Microsoft Azure | Data Factory > rvmfootballdataengineering-df

Search factory and documentation

Set properties

Name: rvm_stadiums_cleaned_timestamp

Linked service *: AzureDataLakeStorage1

File path: rvmfootballdataeng / data / stadium_cleaned2

First row as header:

Import schema:

- From connection/store
- From sample file
- None

> Advanced

Copy data

Copy data1

General Source¹ Sink¹ Mapping Settings User properties

Source dataset *: Select... + New

OK Back

Home Author Monitor Manage Learning Center

Data Factory Validate all Publish all 1

Factory Resources pipeline1

Pipelines 1

pipeline1

Activities

Move and transform

- Copy data
- Data flow

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Power Query

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the navigation bar includes Home, Author (selected), Monitor, Manage, and Learning Center. The main workspace displays a pipeline named 'pipeline1'. The 'Activities' section shows a 'Copy data' activity named 'Copy data1'. The 'Sink' tab is selected, showing a dropdown menu for 'Sink dataset *' with the option 'Select...'. To the right, a sidebar titled 'New dataset' lists 'Azure Data Lake Storage Gen1' and 'Azure Data Lake Storage Gen2' as available data stores. A search bar at the top right says 'Search factory and documentation'.

Microsoft Azure | Data Factory > rvmfootballdataengineering-df

Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory Validate all Publish all 2

Home Author Monitor Manage Learning Center

Factory Resources Filter resources by name +

Pipelines 1

- pipeline1
- Change Data Capture (preview) 0
- Datasets 1
- rvm_stadiums_cleaned_timestamp
- Data flows 0
- Power Query 0

Activities Search activities

Move and transform

- Copy data
- Data flow
- Synapse
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

Validate Validate copy runtime Debug Add trigger

Copy data

Copy data1

General Source Sink¹ Mapping Settings User properties

Sink dataset * Select... + N

Select format

Choose the format type of your data

 Avro	 Binary	 Delimited
 JSON	 ORC	 Parquet

Continue Back

Microsoft Azure | Data Factory > rvmfootballdataengineering-df

Search factory and documentation

Set properties

Name: rvm_stadium_cleaned

Linked service*: AzureDataLakeStorage1

File path: rvmfootballdataeng / output / File name

First row as header:

Import schema:

- From connection/store
- From sample file
- None

> Advanced

Copy data

Copy data1

General Source Sink¹ Mapping Settings User properties

Sink dataset*: Select...

OK Back

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Home Author Monitor Manage Learning Center

Data Factory Validate all Publish all 2

Factory Resources pipeline1

Pipelines 1

- pipeline1

Activities

Search activities

Move and transform

- Copy data

Data flow

- Synapse
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

Validate Validate copy runtime Debug Add trigger

The screenshot shows the Microsoft Azure Data Factory interface. On the left, there's a navigation bar with Home, Author (selected), Monitor, Manage, and Learning Center. The main area shows a pipeline named 'pipeline1' with a single 'Copy data' activity. The 'Sink' tab is currently selected in the activity configuration pane. A 'Set properties' dialog box is open on the right, containing fields for Name (rvm_stadium_cleaned), Linked service (AzureDataLakeStorage1), File path (rvmfootballdataeng / output / File name), and Import schema (None). There are also checkboxes for First row as header and a link to Advanced settings.

Microsoft Azure | Data Factory > rvmfootballdataengineering-df

Search factory and documentation

JoseRafael.Vera

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory Validate all Publish all 3

Preview experience Off

Factory Resources pipeline1

Activities

Validate Validate copy runtime Debug Add trigger

Copy data

Properties

General Related

Name * pipeline1

Description

Annotations + New

Your pipeline has been validated.

No errors were found.

Close

Home Author Monitor Manage Learning Center

Pipelines 1 pipeline1

Change Data Capture (preview) 0

Datasets 2 rvm_stadium_cleaned rvm_stadiums_cleaned_timestamp

Data flows 0

Power Query 0

Move and transform

Copy data

Data flow

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Power Query

General Source Sink Mapping Settings

Sink dataset * rvm_stadium_cleaned

Open + New Learn more

Copy behavior Select...

Max concurrent connections

Block size (MB)

Metadata + New

Quote all text

Microsoft Azure | Data Factory > rvmfootballdataengineering-df

Search factory and documentation

Home Author Monitor Manage Learning Center

Validate all Publish all 3

Factory Resources

Filter resources by name

Pipelines

- pipeline1

Change Data Capture (preview)

Datasets

- rvm_stadium_cleaned
- rvm_stadiums_cleaned_timestamp

Data flows

Power Query

Activities

Search activities

Move and transform

- Copy data
- Data flow

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Power Query

pipeline1

Validate Cancel options Add trigger

Copy data

Copy data1

Properties

Name * pipeline1

Description

Annotations

New

Parameters Variables Settings Output

Pipeline run ID: 626f76da-7211-44a9-9e8d-4625bce29da1

Pipeline status In progress

All status

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration	Integral
Copy data1	Queued	Copy data	2/19/2024, 2:22:32 PM	Less than 1s	

Microsoft Azure | Data Factory > rvmfootballdataengineering-df

Search factory and documentation

JoseRa

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Preview experience

Data Factory < Validate all Publish all 3

Home Author Monitor Manage Learning Center

Factory Resources < Filter resources by name +

Pipelines 1 pipeline1

Change Data Capture (preview) 0

Datasets 2 rvm_stadium_cleaned rvm_stadiums_cleaned_timestamp

Data flows 0

Power Query 0

Activities < Validate Debug Add trigger

Search activities

Move and transform

Copy data

Data flow

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Power Query

Properties

Name * pipeline1

Description

Annotations

+ New

pipeline1

Copy data

Copy data1

Parameters Variables Settings Output

Pipeline run ID: 626f76da-7211-44a9-9e8d-4625bce29da1

Pipeline status ✓ Succeeded

View debug run consumption

All status

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration	Integral
Copy data1	✓ Succeeded	Copy data	2/19/2024, 2:22:32 PM	13s	AutoRe:

[Home](#) > [rvmfootballdataengsa_1708339491574 | Overview](#) > [rvmfootballdataengsa | Containers](#) >

rvmfootballdataeng

Container

 Search

<<

 [Upload](#) [Add Directory](#) [Refresh](#) | [Rename](#) [Delete](#) [Change tier](#) [Acquire lease](#) [Break lease](#) [Give feedback](#) [Overview](#) [Diagnose and solve problems](#) [Access Control \(IAM\)](#)

Settings

 [Shared access tokens](#) [Manage ACL](#) [Access policy](#) [Properties](#) [Metadata](#)**Authentication method:** Access key ([Switch to Microsoft Entra user account](#))**Location:** rvmfootballdataeng / output

Search blobs by prefix (case-sensitive)

Name	Modified
<input type="checkbox"/> [.]	
<input type="checkbox"/> stadium_cleaned2024-02-19_12_41_40.398458.csv	2/19/2024, 2:22:44 PM

Azure | Data Factory ▶ rvmfootballdataengineering-df Search factory and documentation

Details Refresh

Learn more on copy performance details from here.

Activity run id: 26fab862-0812-4bdc-ad21-5a672109e99d

 Azure Data Lake Storage Gen2
Region: West Europe Succeeded  Azure Data Lake Storage Gen2
Region: West Europe

Data read: ① 95,688 KB
Files read: ① 1
Peak connections: ① 3

Data written: ① 95,688 KB
Files written: ① 1
Peak connections: ① 1

Copy duration 00:00:14
Throughput: ① 47,844 KB/s

✓ Azure Data Lake Storage Gen2 → Azure Data Lake Storage Gen2

Start time 2/19/2024, 2:50:46 PM
Used DIUs ① 4
Used parallel copies ① 1

✓ Duration 00:00:14

Details	Working duration	Total duration						
✓ Queue ①		00:00:08						
✓ Transfer ①	<table border="1"><tr><td>Listing source ①</td><td>00:00:00</td></tr><tr><td>Reading from source ①</td><td>00:00:00</td></tr><tr><td>Writing to sink ①</td><td>00:00:00</td></tr></table>	Listing source ①	00:00:00	Reading from source ①	00:00:00	Writing to sink ①	00:00:00	00:00:02
Listing source ①	00:00:00							
Reading from source ①	00:00:00							
Writing to sink ①	00:00:00							

Data consistency verification ① Not verified

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Home Author Monitor Manage Learning Center

Factory Resources < >

Pipelines pipeline1

Change Data Capture (preview) 0

Datasets 2

- rvm_stadium_cleaned
- rvm_stadium_cleaned_timestamp

Data flows 0

Power Query 0

Validate Publishing 3

Validate Validate copy runtime Debug Add trigger

Copy data

Copy data1

General Source Sink Mapping Settings User properties

Source dataset * rvm_stadium_cleaned_timestamp Open New Preview

File path type File path in dataset Wildcard file path List of files

Filter by last modified Start time (UTC) End time (UTC)

Recursively

Enable partitions discovery

Max concurrent connections

Skip line count

Publish Cancel

Publish all

You are about to publish all pending changes to the live environment.

Pending changes (3)

NAME	CHANGE	EXIST
Pipelines		
pipeline1	(New)	-
Datasets		
rvm_stadium_cleaned_timestamp	(New)	-
rvm_stadium_cleaned	(New)	-

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory Validate all Publish all 3



Home



Author



Monitor



Manage



Learning Center

Factory Resources

Filter resources by name

Pipelines 1

pipeline1

Change Data Capture (preview) 0

Datasets 2

rvm_stadium_cleaned

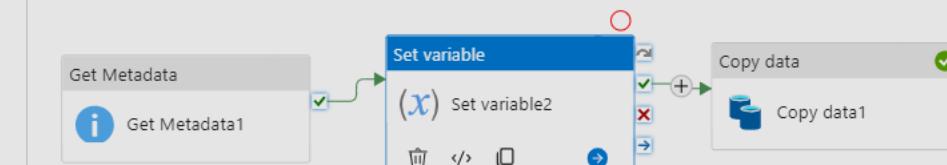
rvm_stadiums_cleaned_timestamp

Data flows 0

Power Query 0

pipeline1

Validate Debug Add trigger

General Settings¹ User propertiesVariable type ① Pipeline variable Pipeline return value

Name *

orderbycreated

New

Value *

Add dynamic content [Alt+Shift+D]

Pipeline expression builder

Add dynamic content below using any combination of expressions, functions and system variables

@last(activity('Get Metadata1').output.childItems).name

Clear contents

Activity outputs Parameters System variables Functions Variables

Search

Get Metadata1

Get Metadata1 activity output

Get Metadata1

Get Metadata1 pipeline return value

Get Metadata1 childItems

List of subfolders and files in the given folder

Get Metadata1 exists

Whether a file, folder, or table exists

Get Metadata1 itemName

Name of the file or folder

Get Metadata1 itemType

Type of the file or folder. Returned value is File or Folder

Get Metadata1 lastModified

Last modified datetime of the file or folder

OK

Cancel

Microsoft Azure Search resources, services, and docs (G+) JoseRafael.Vera

Home > Azure Synapse Analytics ...

Tajamar

Create Manage view Refresh Export to CSV Open query | Assign tags

Filter for any field... Subscription equals all Resource group equals all Location equals all Add filter

Showing 0 to 0 of 0 records. No grouping List view

Name ↑	Type ↑↓	Resource group ↑↓	Location ↑↓	Subscription ↑↓
--------	---------	-------------------	-------------	-----------------



No Azure Synapse Analytics to display

Synapse Analytics is a fully-managed service to build modern data warehouses for enterprises. Synapse Analytics brings together SQL, Apache Spark, Orchestration, and Ingestion into a single workspace, dramatically reducing the time to build an analytics solution.

[Create Synapse workspace](#)

[Service overview](#)

Home > Azure Synapse Analytics >
Create Synapse workspace ...

* Basics *Security Networking Tags Review + create

Create a Synapse workspace to develop an enterprise analytics solution in just a few clicks.

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all of your resources.

Subscription *

Azure for Students

The Synapse and SQL resource providers are now registered with this subscription.

Resource group *

rvmfootballdataeng-rg

[Create new](#)

Managed resource group

Enter managed resource group name

Workspace details

Name your workspace, select a location, and choose a primary Data Lake Storage Gen2 file system to serve as the default location for logs and job output.

Workspace name *

rvmfootballdataengsaw

Region *

West Europe

Select Data Lake Storage Gen2 *

From subscription Manually via URL

Account name *

rvmfootballdataengsa

[Create new](#)

File system name *

rvmfootballdataeng

[Create new](#)

Assign myself the Storage Blob Data Contributor role on the Data Lake Storage Gen2 account to interactively query it in the workspace.

We will automatically grant the workspace identity data access to the specified Data Lake Storage Gen2 account, using the [Storage Blob Data Contributor](#) role. To enable other users to use this storage account after you create your workspace, perform these tasks:

[Review + create](#)

< Previous

Next: Security >

Microsoft Azure

Home > Azure Synapse Analytics >

Create Synapse workspace

* Basics * **Security** Networking Tags Review + create

Configure security options for your workspace.

Authentication

Choose the authentication method for access to workspace resources such as SQL pools. The authentication method can be changed later on. [Learn more](#)

Authentication method ⓘ

- Use both local and Microsoft Entra ID authentication
 Use only Microsoft Entra ID authentication

SQL Server admin login * ⓘ

sqladminuser

SQL Password ⓘ

Enter server password ✓

Confirm password

Confirm the above password ✓

System assigned managed identity permission

Select to grant the workspace network access to the Data Lake Storage Gen2 account using the workspace system identity. [Learn more](#)

Allow network access to Data Lake Storage Gen2 account. ⓘ

! The selected Data Lake Storage Gen2 account does not restrict network access using any network access rules, or you selected a storage account manually via URL under Basics tab. [Learn more](#)

Workspace encryption

! Double encryption configuration cannot be changed after opting into using a customer-managed key at the time of workspace creation.

Choose to encrypt all data at rest in the workspace with a key managed by you (customer-managed key). This will provide double encryption with encryption at the infrastructure layer that uses platform-managed keys. [Learn more](#)

Double encryption using a customer-managed key Enable Disable

[Review + create](#)

< Previous

Next: Networking >

Home > Azure Synapse Analytics >

Create Synapse workspace

* Basics * Security Networking Tags Review + create

Configure networking options for your workspace.

Managed virtual network

Choose whether to set up a dedicated Azure Synapse-managed virtual network for your workspace. [Learn more ↗](#)

Managed virtual network ⓘ

Enable Disable

i To control public network access to your Synapse workspace, you must enable managed virtual network.

Firewall rules

⚠ Azure Synapse Studio and other client tools will only connect to the workspace endpoints if this setting is selected. Connections from specific IP addresses or all Azure services can be allowed or disallowed after the workspace is provisioned.

Allow connections from all IP addresses to your workspace's endpoints. You can restrict these permissions to just Azure datacenter IP addresses and/or specific IP address ranges after creating the workspace.

Allow connections from all IP addresses

Encrypted connections

Azure Synapse Analytics applies TLS1.2 encryption protocol for all network connections. [Learn more ↗](#)

Minimum TLS version

TLS 1.2 ▼

[Review + create](#)

[< Previous](#)

[Next: Tags >](#)

Create Synapse workspace

* Basics *Security Networking Tags Review + create

Tags are name/value pairs that enable you to categorize resources and view consolidated billing by applying the same tag to multiple resources and resource groups. [learn more](#)

Note that if you create tags and then change resource settings on other tabs, your tags will be automatically updated.

Name ⓘ	Value ⓘ	Resource
<input type="text"/>	:	<input type="text"/> Synapse workspace

Review + create

< Previous

Next: Review + create >



Create Synapse workspace ...

Validation succeeded

[* Basics](#) [* Security](#) [Networking](#) [Tags](#) [Review + create](#)

Product Details

Azure Synapse Analytics workspace by Microsoft Serverless SQL est. cost/TB ⓘ
5.00 USD
[Terms of use](#) | [Privacy policy](#)

Terms

By clicking Create, I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. For additional details see [Azure Marketplace Terms](#).

Basics

Subscription	Azure for Students
Resource group	rvmfootballdataeng-rg
Region	West Europe
Workspace name	(new) rvmfootballdataengsaw
Data Lake Storage Gen2 account	<input type="text" value="rvmfootballdataeng"/>
Data Lake Storage Gen2 file system	
Managed resource group	None
Role assignments	The Storage Blob Data Contributor role will be assigned on the specified Data Lake Storage Gen2 account to both the workspace managed identity and the current user.

Security

Authentication method	Use both local and Microsoft Entra ID authentication
SQL Server admin login	sqladminuser
SQL Password	Auto-generated
Double encryption	No

Networking

Managed virtual network	No
Allow connections from all IP addresses	Yes

[Create](#)[< Previous](#)[Next >](#)[Download a template for automation](#)

Home >

Microsoft.Azure.SynapseAnalytics-20240219145813 | Overview

Deployment

Search Delete Cancel Redeploy Download Refresh

Overview

Your deployment is complete

Deployment name : Microsoft.Azure.SynapseAnalytics-20240219145813
Subscription : Azure for Students
Resource group : rvmfootballdataeng-rg

Start time : 2/19/2024, 3:07:46 PM
Correlation ID : [REDACTED]

Inputs

Outputs

Template

Deployment details

Resource	Type	Status	Operation details
✓ StorageRoleDeploymentResource	Deployment	OK	Operation details
✓ rvmfootballdataengsaw/allowAll	Microsoft.Synapse/workspaces/	OK	Operation details
✓ rvmfootballdataengsaw	Synapse workspace	OK	Operation details
✓ rvmfootballdataengsaw	Synapse workspace	OK	Operation details

Next steps

[Go to resource group](#)

Give feedback

Tell us about your experience with deployment

Microsoft Azure

Search resources, services, and docs (G+)

Home > Microsoft.Azure.SynapseAnalytics-20240219145813 | Overview >

rvmfootballdataeng-rg

Resource group

Search

Create Manage view Delete resource group Refresh Export to CSV Open query Assign tags Move Delete Export template Open in mobile

Overview

^ Essentials

Subscription (move) : Azure for Students Deployments : 4 Succeeded
Subscription ID : [REDACTED] Location : West Europe
Tags (edit) : Add tags

Activity log Access control (IAM) Tags Resource visualizer Events

Resources Recommendations

Filter for any field... Type equals all × Location equals all × + Add filter

Showing 1 to 3 of 3 records. Show hidden types ⓘ

	Name ↑↓	Type ↑↓
<input type="checkbox"/>	rvmfootballdataengineering-df	Data factory (V2)
<input type="checkbox"/>	rvmfootballdataengsa	Storage account
<input type="checkbox"/>	rvmfootballdataengsaw	Synapse workspace

Deployments Security Deployment stacks Policies Properties Locks

[Home](#) > [Microsoft.Azure.SynapseAnalytics-20240219145813 | Overview](#) > [rvmfootballdataeng-rg](#) >

rvmfootballdataengsaw

Synapse workspace

Search

+ New dedicated SQL pool + New Apache Spark pool + New Data Explorer pool (preview) Refresh Reset SQL admin password Delete

Overview

- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems

Settings

- Microsoft Entra ID
- Properties
- Locks

Analytics pools

- SQL pools
- Apache Spark pools
- Data Explorer pools (preview)

Security

- Encryption
- Networking
- Identity
- Private endpoint connections
- Approved Microsoft Entra

Essentials

Resource group (move)	: rvmfootballdataeng-rg	Networking
Status	: Succeeded	Primary ADLS Gen2 acco
Location	: West Europe	Primary ADLS Gen2 file s
Subscription (move)	: Azure for Students	SQL admin username
Subscription ID	: [REDACTED]	SQL Microsoft Entra adm
Managed virtual network	: No	Dedicated SQL endpoint
Managed Identity object ...	: bb9582fa-9ac0-4894-96ce-c8869813aacd	Serverless SQL endpoint
Workspace web URL	: https://web.azure.synapse.net?workspace=%2Brvmfootballdataeng-rg	Development endpoint
Tags (edit)	: Add tags	

Getting started



Open Synapse Studio

Start building your fully-integrated analytics solution and unlock new insights.



Read documentation

Learn how to be productive quickly. Explore concepts, tutorials, and samples.

[Learn more](#)

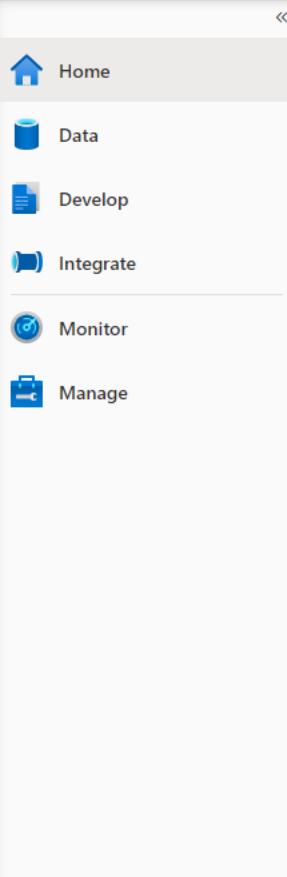
Analytics pools

Search to filter items...

Name	Type
------	------

SQL pools

 We use optional cookies to provide a better experience. [Learn more](#)



Synapse Analytics workspace rvmfootballdataengsaw

New 

-  SQL script
-  KQL script
-  Notebook
-  Data flow
-  Apache Spark job definition
-  Pipeline
-  Import



Explore and analyze 

Learn how to get insights from your data.



Visualize

Build interactive reports with Power BI capabilities.

Discover more



Knowledge center



Browse partners

Recent resources

Microsoft Azure | Synapse Analytics > rvmfootballdataengsa2

We use optional cookies to provide a better experience. [Learn more](#)

Synapse live | Validate all | Publish all (1)

Home Data Develop Integrate Monitor Manage

Data

Workspace

SQL database

Filter resources by name

SQL script 1

Other users in your workspace may have access to modify this item. Do not use

Published Query plan Connect to Built-in Use d

Link

Workspace

SQL database

Lake database

Data Explorer database (preview)

Linked

Connect to external data

Integration dataset

Browse gallery

The screenshot shows the Microsoft Azure Synapse Analytics Data workspace interface. On the left, there's a navigation menu with options like Home, Data, Develop, Integrate, Monitor, and Manage. The 'Data' option is selected. In the main area, under the 'Data' section, there's a 'Workspace' tab where a new resource is being created. A modal dialog is open, prompting the user to choose a resource type. The 'SQL database' option is selected and highlighted. Other options visible in the dropdown include 'Lake database', 'Data Explorer database (preview)', 'Connected', 'Integration dataset', and 'Browse gallery'. At the top of the workspace, there are buttons for 'Synapse live', 'Validate all', and 'Publish all' (with one item pending). There's also a search bar at the top right.

Microsoft Azure | Synapse Analytics > rvmfootballdataengsaw

We use optional cookies to provide a better experience. Learn more [Accept](#) [Reject](#) [More](#)

Synapse live Validate all Publish all 2

Home Data Workspace Linked

SQL script 1 rvmStadiumDB

Table Map data Publish

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

Tables

Lake database 1

rvmStadiumDB

Properties

General Related (0)

Choose a name for This name can be used for a limited time until it is published.

Name * rvmStadiumDB

Description

Storage settings for

Linked service * rvmfootballdataengsaw

Input folder * rvmfootballdataeng/rvr

Data format * Delimited Text

Get started

Select tables to build your database.

Microsoft Azure | Synapse Analytics > rvmfootballdataengsaw

We use optional cookies to provide a better experience. [Learn more](#)

Accept

Synapse live Validate all Publish all

Data

Workspace Linked

Filter resources by name

Lake database 1

rvmStadiumDB

SQL script 1

Table Map data Publish

Custom From template From data lake

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

This screenshot shows the Microsoft Azure Synapse Analytics workspace interface. The left sidebar lists 'Data' resources, including 'Lake database' (rvmStadiumDB) and 'SQL script 1'. A modal dialog is open in the center, titled 'Table', which is part of a larger process for creating or modifying a table. The dialog includes options like 'Map data' and 'Publish'. A tooltip message at the top right of the dialog states: 'Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.' The top navigation bar includes links for 'Synapse live', 'Validate all', and 'Publish all', along with a search bar and various status indicators.

We use optional cookies to provide a better experience. [Learn more](#)

Synapse live Validate all

Data

Workspace

Linked

Filter resources by name

Lake database 1

rvmStadiumDB

SQL script 1 rvmStadiumDB
+ Table Map data

Tables

Filter by keyword



Get started

Select tables to build your database.

Continue

Create external table from data lake

External table details

Select the storage location where the files containing the data is staged. Currently Storage (ADLS) Gen2 and Azure Blob Storage are supported. [Learn more](#)

External table name *

stadiums

Linked service *

rvmfootballdataengsa-WorkspaceDefaultStorage(rvmfootballdataengsa)

Input file or folder *

rvmfootballdataeng/output/stadium_cleaned2024-02-19_12_41_40.398458.csv

Microsoft Azure | Synapse Analytics > rvmfootballdataengsaw

Search

We use optional cookies to provide a better experience. Learn more

Accept Reject

Synapse live Validate all Publish all 2

Data Workspace Linked

Filter resources by name

Lake database 1

rvmStadiumDB

Tables 1

Filter by keyword

Others 1

stadiums

stadiums

rank

stadium

capacity

region

country

city

images

home_team

location

See less ^

Properties

General

Name * rvmStadiumDB

Description

Storage settings for table

Linked service rvmfootballdataengsaw

Input folder * rvmfootballdataengsaw

Data format * Delimited Text

Storage settings for table

Microsoft Azure | Synapse Analytics > rvmfootballdataengsaw

Search

We use optional cookies to provide a better experience. [Learn more](#)

Accept

Synapse live Validate all Publish all 1

Publisher Successfully pu

Data

Workspace Linked

Filter resources by name

Lake database 1

rvmStadiumDB

Tables

Filter by keyword

Others 1

stadiums

stadiums

rank

stadium

capacity

region

country

city

images

home_team

location

See less ^

General Columns Relationships

Filter by keyword

+ Column

Clone Convert type Delete

	Name	Keys	Description	Nullability	Data type	Format / Length
Standard column (9)						
	rank	PK		Null	long	
	stadium	PK		Null	string	8000
	capacity	PK		Null	long	

Microsoft Azure | Synapse Analytics > rvmfootballdataengsaw

Search

We use optional cookies to provide a better experience. Learn more

Synapse live Validate all Publish all 1

Data

Workspace Linked

Filter resources by name

Lake database 1

rvmStadiumDB

Tables

stadiums

Columns

rank (long)
stadium (string)
capacity (long)
region (string)
country (string)
city (string)
images (string)
home_team (string)
location (string)

SQL script 1

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

Run Undo Publish Query plan Connect to Built-in Use database master master rvmStadiumDB

1

Microsoft Azure | Synapse Analytics > rvmfootballdataengsaw

We use optional cookies to provide a better experience. [Learn more](#)

Synapse live Validate all Publish all 1

Data Workspace Linked Filter resources by name

Lake database 1

- rvmStadiumDB
 - Tables
 - stadiums
 - Columns
 - rank (long)
 - stadium (string)
 - capacity (long)
 - region (string)
 - country (string)
 - city (string)
 - images (string)
 - home_team (string)
 - location (string)

SQL script 1 Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the wor

Run Undo Publish Query plan Connect to Built-in Use database rvmStadiumDB

```
11
12
13 WITH CapacityPercentiles AS (
14     SELECT
15         stadium,
16         capacity,
17         PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY capacity) OVER () AS P75Capacity
18     FROM stadiums
19 )
20 SELECT
21     stadium,
22     capacity,
23     P75Capacity,
24     capacity - P75Capacity AS AboveP75
25 FROM CapacityPercentiles
26 WHERE capacity > P75Capacity;
27
28
29
30
31
32
```

Results Messages

View Table Chart Export results

Search

stadium	capacity	P75Capacity	AboveP75
Basra International Stadium	65227	65056.75	170.25
Olympic Stadium	65255	65056.75	198.25
Hard Rock Stadium	65326	65056.75	269.25



Conectar

Buscar datos

Tableau Server

A un archivo

Microsoft Excel

Archivo de texto

Archivo JSON

Microsoft Access

Archivo PDF

Archivo espacial

Archivo estadístico

Más...

A un servidor

Microsoft SQL Server

MySQL

Oracle

Amazon Redshift



Buscar

Azure Synapse Analytics

General SQL inicial

Servidor

Base de datos

Opcional

Autenticación

Azure Active Directory mediante OAuth

URL de instancia de OAuth

https://login.microsoftonline.com/common

 Requiere SSL Leer datos no vinculados

Iniciar sesión

Installed Connectors

Actian Vector

Alibaba AnalyticDB pa

Alibaba MaxCompute

Amazon Athena

Amazon Aurora para I

Amazon EMR Hadoop

Amazon Redshift

Análisis de mar de da

Anaplan (obsoleto)

Apache Drill

Azure Data Lake Stora

Azure SQL Database

Azure Synapse Analyt

Box

Cloudera Hadoop

Conector de datos web (obsoleto)

Conector de procesamiento analítico en
línea Teradata

Customer Data Platform (obsoleto)

Intuit QuickBooks Online (obsoleto)

Kyvos

LinkedIn Sales Navigator (obsoleto)

Presto

Progress OpenEdge

Qubole Presto (obsoleto)

Intelligen

o)

er

er Analysis Services

ector

oint Online

pleto)

Database

Salesforce Data Cloud

SAP ASE

SAP ASE IQ

SAP HANA

SAP NetWeaver Business

ServiceNow ITSM (obso

SingleStore

Snowflake

Spark SQL

Splunk

Teradata

Vertica

Virtualización de datos T

Otras bases de datos (JD

Otras bases de datos (O

Conecadores adiciona

Microsoft Azure

Search resources, services, and docs (G+)

Home > Microsoft.Azure.SynapseAnalytics-20240219145813 | Overview > rvmfootballdataeng-rg >

rvmfootballdataengsaw

Synapse workspace

Search

+ New dedicated SQL pool + New Apache Spark pool + New Data Explorer pool (preview) Refresh Reset SQL admin password Delete

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Subscription (move) : Azure for Students

Subscription ID : [REDACTED]

Managed virtual network : No

Managed Identity object ID : bb9582fa-9ac0-4894-96ce-c8869813aacd

Workspace web URL : <https://web.azuresynapse.net?workspace=%2bsubscriptions%2f9e3c713b-0d42-41e5-81ca-d778fd001da7%2fresource...>

Tags (edit) : Add tags

Networking : Show firewall settings

Primary ADLS Gen2 account : <https://rvmfootballdataengsa.dfs.core.windows.net>

Primary ADLS Gen2 file system : rvmfootballdataeng

SQL admin username : sqladminuser

SQL Microsoft Entra admin : [REDACTED]

Dedicated SQL endpoint : rvmfootballdataengsaw.sql.azuresynapse.net

Serverless SQL endpoint : rvmfootballdataengsaw-on-demand.sql.azuresynapse.net

Development endpoint : <https://rvmfootballdataengsaw.dev.azuresynapse.net>

Copied

Getting started

Open Synapse Studio

Start building your fully-integrated analytics solution and unlock new insights.

Read documentation

Learn how to be productive quickly. Explore concepts, tutorials, and samples.

Open

Conectar

Buscar datos
Tableau Server

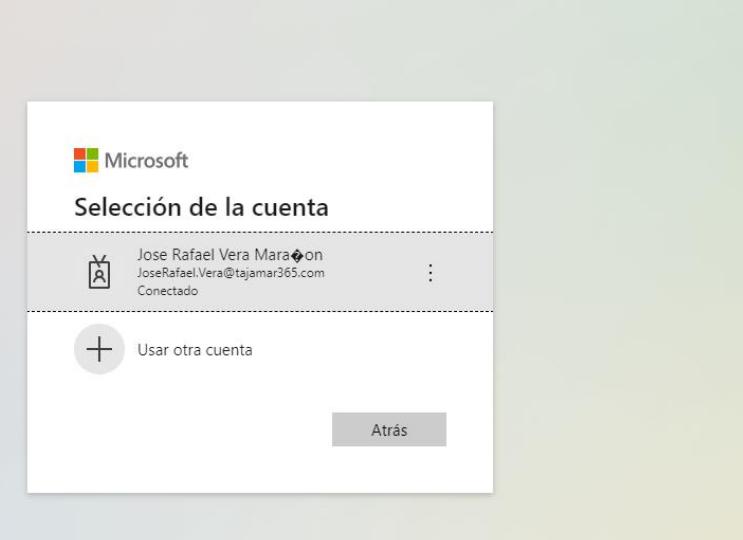
A un archivo
Microsoft Excel
Archivo de texto
Archivo JSON
Microsoft Access
Archivo PDF
Archivo espacial
Archivo estadístico
Más...

A un servidor
MySQL
Oracle
Amazon Redshift
Azure Data Lake Storage Gen2
Más...

Fuentes de datos guardadas
Indicadores Mundiales

Installed Connectors (71)

- Denodo (obsoleto)
- Dremio
- Marketing Cloud
- Alibaba AnalyticDB para MySQL
- Dropbox
- Marketo (obsoleto)
- Alibaba MaxCompute
- Esri
- MarkLogic (obsoleto)
- Amazon Athena
- Exasol
- Microsoft SQL Server
- Amazon Aurora para MySQL
- Firebird 3
- MonetDB
- Amazon EMR Hadoop Hive
- Google Analytics
- MongoDB BI
- Amazon Redshift
- Análisis de mar de datos de Alibaba
- Anaplan (obsoleto)
- Hortonworks Hadoop Hive
- Oracle Eloqua
- Apache Drill
- Azure Data Lake Storage Gen2
- IBM DB2
- Oracle Essbase
- Azure SQL Database
- IBM PDA (Netezza)
- Pivotal Greenplum
- Azure Synapse Analytics
- Impala
- PostgreSQL
- Box



Conectar a Azure Data Lake Storage Gen2

Sesión iniciada como JoseRafael.Vera@tajamar365.com cerrar sesión

Examinar URL del archivo

URL del archivo Extensión de archivo: xls

Cerrar Conectar



Microsoft Azure | Synapse Analytics > rvmfootballdataengsaw

We use optional cookies to provide a better experience. [Learn more](#)

Accept Reject More options

Synapse live Validate all Publish all

Home Data Workspace Linked

Filter resources by name

Azure Data Lake Storage Gen2 2

rvmfootballdataengsaw (Primary - ...)

rvmfootballdataeng (Primary)

(Attached Containers)

SQL script 1 rvmfootballdataeng

New SQL script New notebook New data flow New integration dataset Upload Download New folder Select all More

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

stadium_cleaned2024-02-19_12_41_40.398458... 19/2/2024, 14:51:01

Rename Manage access Properties Delete Refresh

Size 93.4

The screenshot shows the Microsoft Azure Synapse Analytics Data workspace interface. On the left, a navigation bar includes Home, Data, Develop, Integrate, Monitor, and Manage. The Data section is selected and displays a linked Azure Data Lake Storage Gen2 container named 'rvmfootballdataengsaw (Primary)'. Inside this container, there is a single file named 'stadium_cleaned2024-02-19_12_41_40.398458...'. A context menu is open over this file, with the 'Properties' option highlighted. The top navigation bar also features Synapse live, Validate all, Publish all, and various workspace management icons.

Provide a better experience. Learn more

Synapse live Validate all Publish all

Data

SQL script 1 rvmfootballdataeng Other users in your workspace may have access to modify this

New SQL script New notebook New data flow New integration dataset Upload Download

← → ↑ rvmfootballdataeng > output

Azure Data Lake Storage Gen2 2

rvmfootballdataengsa (Primary - ...)

rvmfootballdataeng (Primary)

(Attached Containers)

Name Last Modified

stadium_cleaned2024-02-19_12_41_40.398458... 19/2/2024, 14:51:01

Properties

Name

URL

ABFSS Path

Last modified

Cache Control

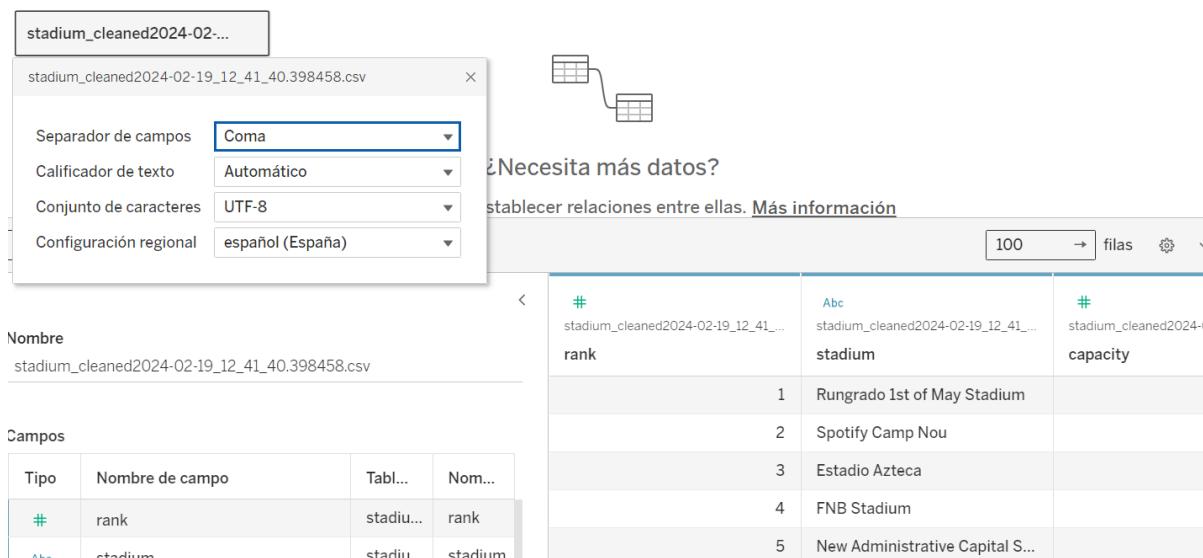
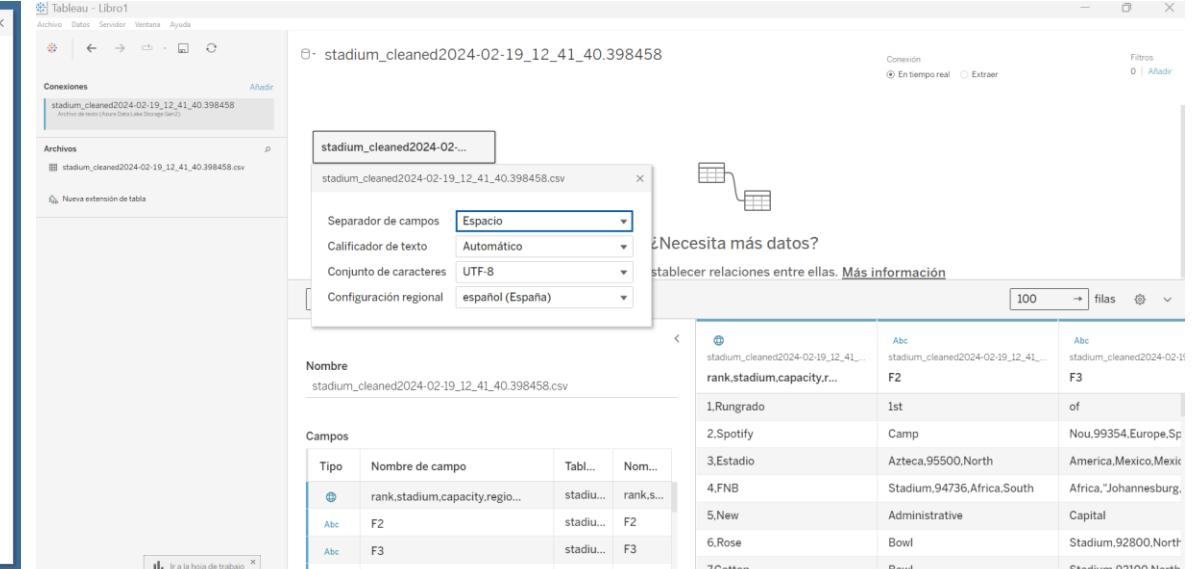
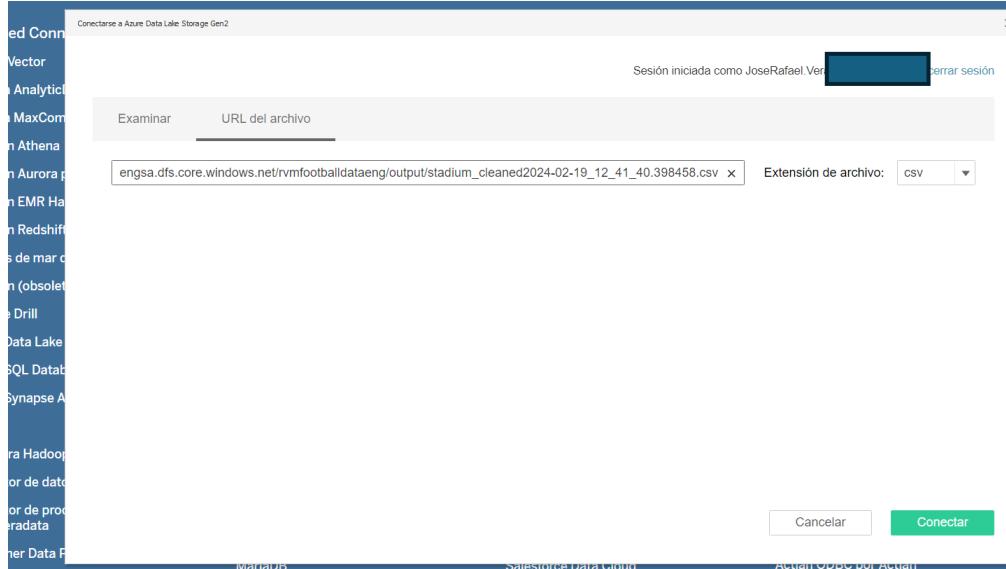
Content Type

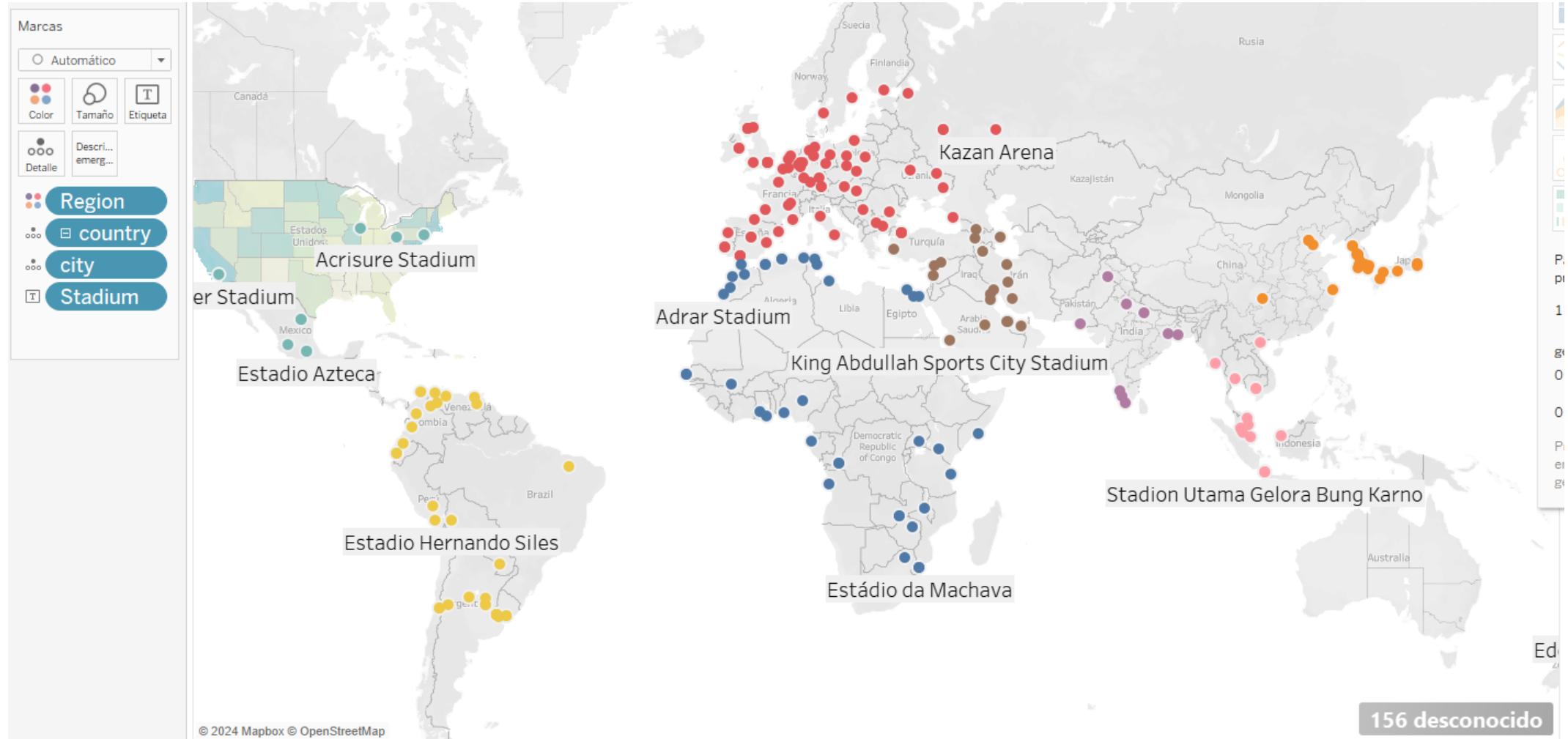
Content Disposition

Content Encoding

Content Language

User Properties





Geographical Distribution of Stadiums

