



TOKYO 2020

BESTHQWALLPAPERS.COM



Abdallah Amr Khamies

TOKYO 2020



In this Project, I analyzed Olympic data using various tools and technologies, including Azure Data Factory, Data Lake Gen 2, Synapse Analytics, and Azure Databricks and PowerBI


In this project

These steps on the Tokyo Olympics project



Azure services


Create a resource


Azure Synapse Analytics


Azure Databricks


App registrations


Data factories


Storage accounts


Quickstart Center








Virtual machines


App Services


More services

Resources


Recent Favorite

Name	Type	Last Viewed
 tokyoolympicdataabdallah	Storage account	49 minutes ago
 tokyo-olympicsa	Synapse workspace	2 hours ago
 tokyoolympic2021-DF	Data factory (V2)	2 hours ago
 Tokyo-Olympic	Resource group	2 hours ago
 tokyo-olympic-dbricks	Azure Databricks Service	5 hours ago
 Azure subscription 1	Subscription	23 hours ago

See all

Create a storage account first that includes the container that contains two files, a data file in which the data is loaded. I integrated data from github, which consists of 5 Excel file files through Data Factory, creating data pipes, and push Data Lake.

Home > [tokyoolympicdataabdallah](#) | Containers >

 **tokyo-olympic-data** ...

Container

«

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

Authentication method: Access key ([Switch to Microsoft Entra user account](#))

Location: tokyo-olympic-data

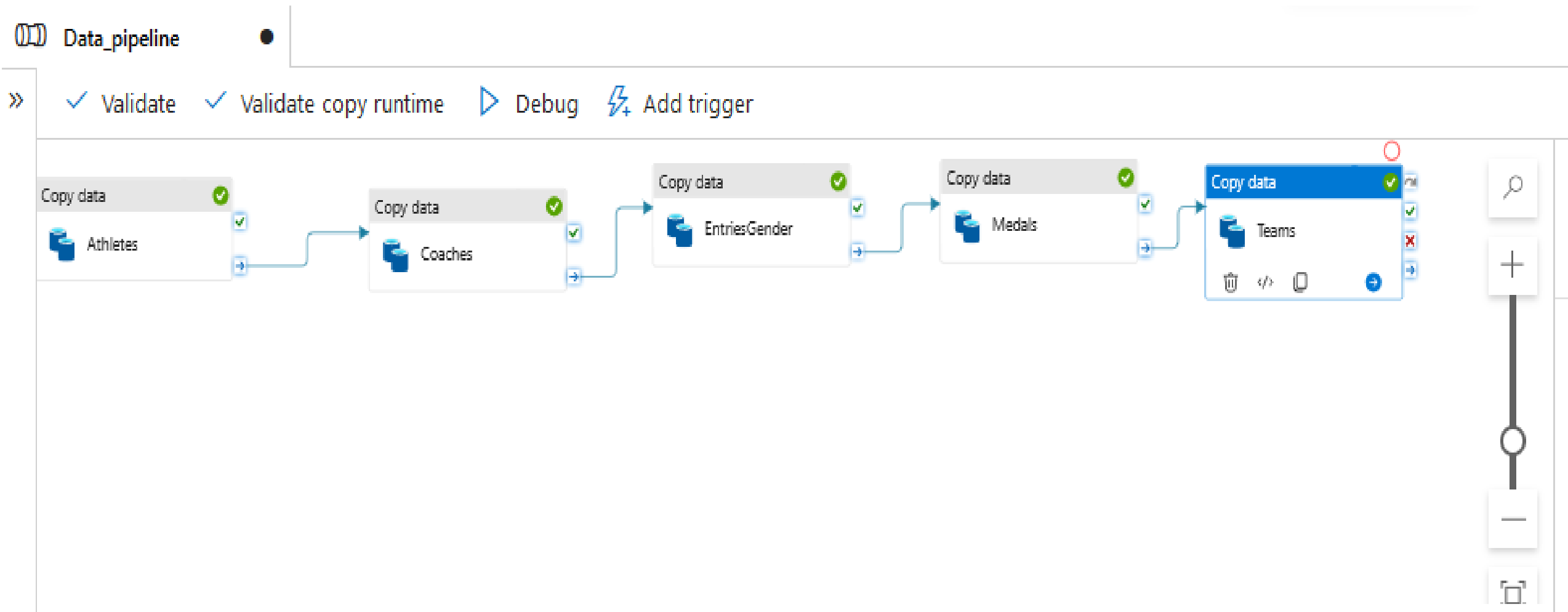
☐ Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size
<input type="checkbox"/> Raw-DataSRC					
<input type="checkbox"/> Transformed-Data					

<

>


Creating Data Pipeline



Transformed Data (databricks)


made some transformation by Databricks (pyspark)


Microsoft Azure

 databricks

CTRL + P

tokyo-olympic-dbricks





abdallahamr3398@gmail.com

New

Workspace

Recents

Catalog

Workflows

Compute

SQL

SQL Editor

Queries

Dashboards

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Delta Live Tables

Machine Learning

Tokyo-Olympic Transformation

Python

☆

File

Edit

View

Run

Help

Last edit was now

Provide feedback

Run all

Terminated

Schedule

Share

1 %fs

2 ls "/mnt/tokyoolymic"

Table

+

New result table: OFF

	path	name	size	modificationTime
1	dbfs:/mnt/tokyoolymic/Raw-DataSRC/	Raw-DataSRC/	0	1698949599000
2	dbfs:/mnt/tokyoolymic/Transformed-Data/	Transformed-Data/	0	1698949632000

2 rows | 8.29 seconds runtime

Refreshed 2 hours ago

Command took 8.29 seconds -- by abdallahamr3398@gmail.com at 11/3/2023, 5:53:41 PM on Abdallah Amr's Cluster

Cmd 5

Python

1 athletes =spark.read.format("csv").option("header","true").option("inferSchema","true").load("/mnt/tokyoolymic/Raw-DataSRC/Athletes.csv")

2 Coaches =spark.read.format("csv").option("header","true").option("inferSchema","true").load("/mnt/tokyoolymic/Raw-DataSRC/Coaches.csv")

3 EntriesGender =spark.read.format("csv").option("header","true").option("inferSchema","true").load("/mnt/tokyoolymic/Raw-DataSRC/EntriesGenderlink.csv")

4 Medals =spark.read.format("csv").option("header","true").option("inferSchema","true").load("/mnt/tokyoolymic/Raw-DataSRC/Medals.csv")

5 Teams =spark.read.format("csv").option("header","true").option("inferSchema","true").load("/mnt/tokyoolymic/Raw-DataSRC/Teams.csv")

6

(10) Spark Jobs

Analyzed by Synapse Analytics and load data from Transformed data

Microsoft Azure | Synapse Analytics ▶ tokyo-olympicsa

Search

>>

Synapse live

Validate all

Publish all

Home

Database

Document

Query

Analytics

Tools

Data

Workspace

Linked

Filter resources by name

Lake database

1

TokyoOlympicDB

Tables

...

TB_athletes

TB_Coaches

TB_EntriesGender

TB_Medals

TB_Teams

TokyoOlympicDB

New SQL script

tokyo-olympic-

Name

athletes

Coaches

EntriesGender

Medals

Teams

»

Synapse live

Validate all

Publish all 1

Home

Database

Workspace

Filter resources by name

Lake database 1

TokyoOlympicDB

Tables

TB_athletes

TokyoOlympicDB

SQL script 1

Run

Undo

Publish

Query plan

Connect to Built-in

Use database TokyoOlympicDB

SQL script 1

```
1 SELECT TOP (100) [PersonName]
2 , [Country]
3 , [Discipline]
4 FROM [TokyoOlympicDB].[dbo].[TB_athletes]
```

Results Messages

View Table Chart

Export results

Search

PersonName	Country	Discipline
AALERUD Katrine	Norway	Cycling Road
ABAD Nestor	Spain	Artistic Gymnastics
ABAGNALE Giovanni	Italy	Rowing
ABALDE Alberto	Spain	Basketball
ABALDE Tamara	Spain	Basketball
ABALO Luc	France	Handball

00:00:04 Query executed successfully.

Properties

General Related (0)

Name *
SQL script 1

Description

Type
.sql script

Size
97 bytes

Results settings per query ⓘ
☒ First 5000 rows (default)
☐ All rows

🏠

Data

+

▾

⏪

Workspace

Linked

🔍 Filter resources by name

📁 Lake database 1

📁 TokyoOlympicDB

📁 Tables

📁 TB_athletes

▶ 📁 Columns

▶ 📁 TB_Coaches

▶ 📁 TB_EntriesGender

▶ 📁 TB_Medals

▶ 📁 TB_Teams

TokyoOlympicDB

SQL script 1

•

▶ Run

↶ Undo

▾

📄 Publish

🗑️ Query plan

Connect to

✔ Built-in ▾

Use database

TokyoOlympicDB ▾

↺

⚙️

⋮

```
1  --count Athletes of each country
2  select country, COUNT(*) as total_Athletes
3  from TB_athletes
4  GROUP BY Country
5  ORDER BY total_Athletes DESC
6
```

Results

Messages

^

View

Table

Chart

↗ Export results ▾

🔍 Search

country	total_Athletes
United States of America	615
Japan	586
Australia	470
People's Republic of China	401
Germany	400

Properties

General

Related (0)

Name *

SQL script 1

Description

Type

.sql script

Size

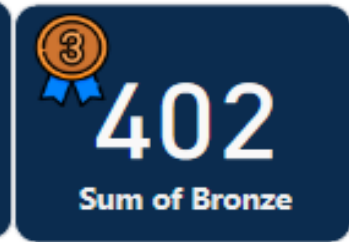
97 bytes

Results settings per query ⓘ

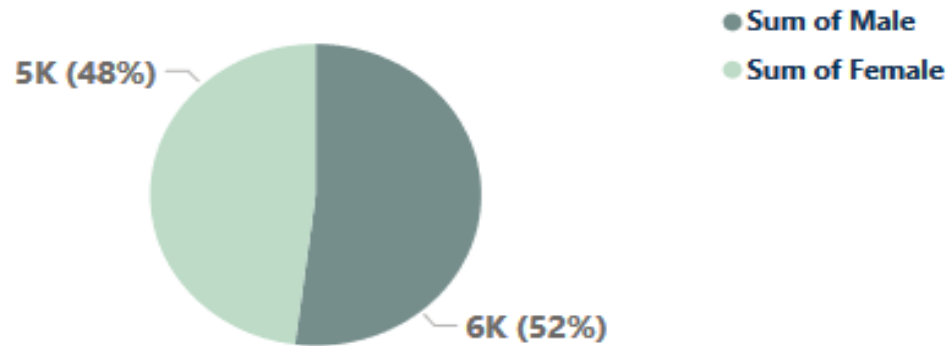
☒ First 5000 rows (default)

☐ All rows

Created a Dashboard and Report for Tokyo Olympic 2020



Sum of Male and Sum of Female



Sum of Total by Country



Team_Country	Sum of Gold	Sum of Bronze	Sum of Silver	Sum of Total	Sum of Rank	Sum of Rank by Total
Venezuela	1	0	3	4	46	47
Uzbekistan	3	2	0	5	32	42
United States of America	39	33	41	113	1	1
Ukraine	1	12	6	19	44	16
Uganda	2	1	1	4	36	47
Turkmenistan	0	0	1	1	77	77
Turkey	2	9	2	13	35	20
Tunisia	1	0	1	2	58	66
Thailand	1	1	0	2	59	66
Syrian Arab Republic	0	1	0	1	86	77
Switzerland	3	6	4	13	24	20
Sweden	3	0	6	9	23	26
Spain	3	6	8	17	22	17
South Africa	1	0	2	3	52	60
Slovenia	3	1	1	5	31	42
Slovakia	1	1	2	4	50	47
Serbia	3	5	1	9	28	26
Saudi Arabia	0	0	1	1	77	77
San Marino	0	2	1	3	72	60
Romania	1	0	3	4	46	47
ROC	20	23	28	71	5	3
Total	340	402	338	1080	4309	4045