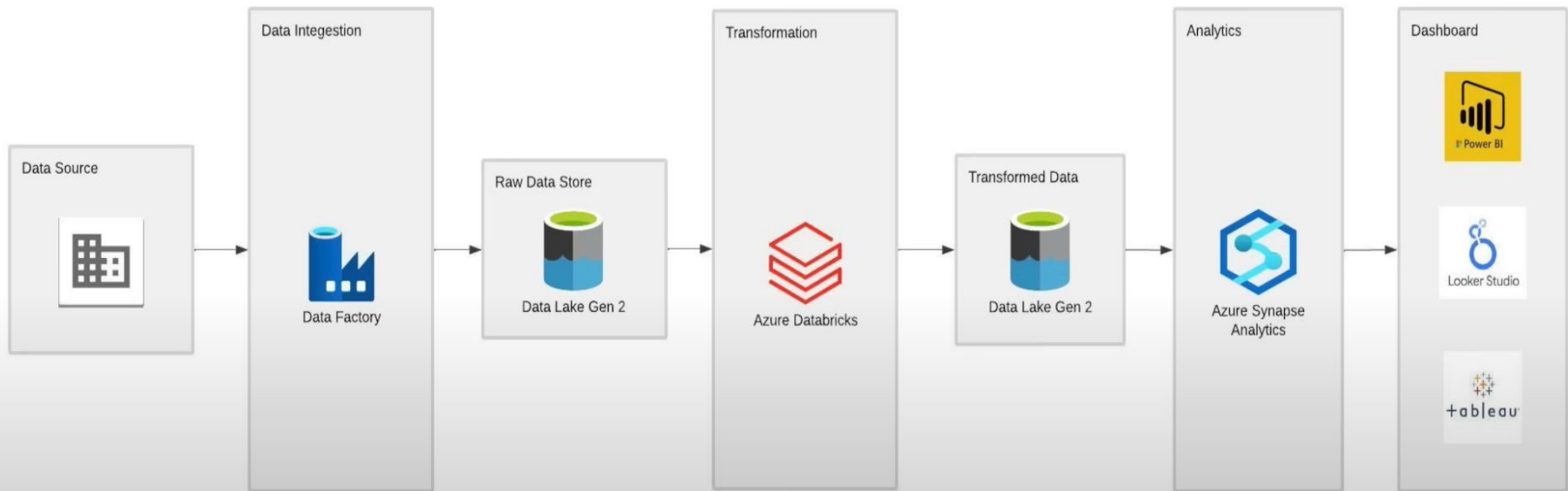


# Olympic Data Analytics

Using Microsoft Azure

# Architecture



# Storage Account

Microsoft Azure

Upgrade

Search resources, services, and docs (G+)

Copilot

3

mahendrachinchkhede1...  
DEFAULT DIRECTORY (MAHENDR...

Home > actokyoolympic\_1747899648009 | Overview >

actokyoolympic

Storage account

Search

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Partner solutions

Resource visualizer

Data storage

- Containers
- File shares
- Queues
- Tables

Security + networking

Upload

Open in Explorer

Delete

Move

Refresh

Open in mobile

CLI / PS

Feedback

Essentials

Resource group (move) : ac-data-engineering-projects

Location : westindia

Primary/Secondary Loca... : Primary: West India, Secondary: South India

Subscription (move) : Azure subscription 1

Subscription ID : 19c1b034-45a8-43d9-9dc6-229aa180787a

Disk state : Primary: Available, Secondary: Available

Tags (edit) : Add tags

Performance : Standard

Replication : Read-access geo-redundant storage (RA-GRS)

Account kind : StorageV2 (general purpose v2)

Provisioning state : Succeeded

Created : 5/22/2025, 1:10:57 PM

JSON View

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

Security

Require secure transfer for REST API operations : Enabled

Storage account key access : Enabled

Minimum TLS version : Version 1.2

Infrastructure encryption : Disabled


# Container


Microsoft Azure


Upgrade


Search resources, services, and docs (G+)


Copilot



5








mahendrachinchhede1...  
DEFAULT DIRECTORY (MAHENDR...

Home > actokyoolympic\_1747899648009 | Overview > actokyoolympic | Containers >

tokyo-olympic-data

...

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

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

Location: tokyo-olympic-data

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state	
<input type="checkbox"/> raw-data	22/05/2025, 13:30:57					-	...
<input type="checkbox"/> transformed-data	22/05/2025, 13:31:06					-	...

Add or remove favorites by pressing Ctrl+Shift+F

# DataFactory & Pipeline

The screenshot displays the Microsoft Azure Data Factory portal interface. At the top, the header bar includes the Microsoft Azure logo, the text 'Data Factory', and the specific factory name 'actokyo-olympic-df', which is highlighted with a red box. A search bar for 'Search factory and documentation' is also present. On the right side of the header, there are notification icons, a user profile icon, and the email address 'mahendrachinchkhede10@gmail.com' with the text 'DEFAULT DIRECTORY' below it.

Below the header, the main navigation area shows a list of items, with 'data-ingestion' selected and highlighted by a red box. To the left of this area is a sidebar menu with icons for Home, Add, Monitor, and Learn. The 'Activities' section is expanded, showing a list of activity types: Move and transform, Synapse, Azure Data Explorer, Azure Function, Batch Service, Databricks, Data Lake Analytics, General, HDInsight, Iteration & conditionals, Machine Learning, and Power Query.

The main workspace area is divided into two sections. The top section contains buttons for 'Validate', 'Debug', and 'Add trigger'. The bottom section is a tabbed interface with 'Parameters', 'Variables', 'Settings', and 'Output' tabs. The 'Parameters' tab is currently active, showing a '+ New' button to add a new parameter.

# data-ingestion Pipeline

Microsoft Azure | Data Factory | actokyo-olympic-df

Search factory and documentation

Validate all | Publish all 16

Preview experience | Off

data-ingestion

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 | Debug | Add trigger

Copy data

Athletes

Coaches

EntriesGender

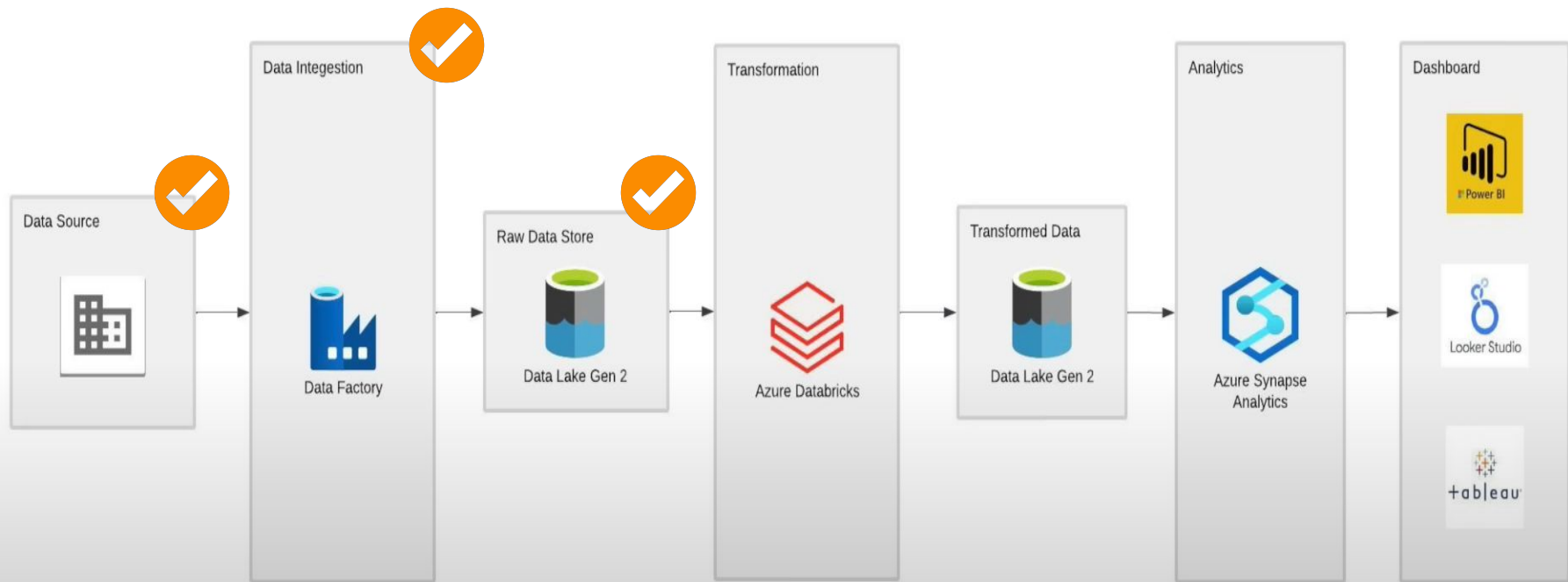
Medals

Teams

Parameters | Variables | Settings | Output

Showing 1 - 5 of 5 items

Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime	User prop...	Activity run ID
Teams	✓ Succeeded	Copy data	5/22/2025, 3:05:05 PM	13s	AutoResolveIntegrationRuntime (Central India)		dad94317-948b
Medals	✓ Succeeded	Copy data	5/22/2025, 3:04:50 PM	14s	AutoResolveIntegrationRuntime (Central India)		4fa21bc9-07c0-
EntriesGender	✓ Succeeded	Copy data	5/22/2025, 3:04:34 PM	16s	AutoResolveIntegrationRuntime (Central India)		a21874f5-531c-
Coaches	✓ Succeeded	Copy data	5/22/2025, 3:04:18 PM	15s	AutoResolveIntegrationRuntime (Central India)		4c6a53d4-db2b
Athletes	✓ Succeeded	Copy data	5/22/2025, 3:04:03 PM	14s	AutoResolveIntegrationRuntime (Central India)		1e7b9f73-916d



# Azure Databricks

+ New

 Workspace

⌚ Recents

 Catalog Workflows

 Compute

Marketplace

SQL

 SQL Editor

## Queries

 Dashboards

 Genie

 Alerts

 Query History

SQL Warehouses

## Data Engineering

☰ Job Runs

## Data Ingestion

## Pipelines

Machine Learning

# Welcome to Databricks

Q Search data, notebooks, recents, and more...

CTRL + P



## Set up your workspace

Follow this [step-by-step guide](#) that walks you through setting up the workspace for your new Databricks account.

Get started 

⌚ Recents

☆ Favorites

~ Popular

✦ Mosaic AI

 What's new

## Start your journey

Try the "New" menu, where you can upload or connect to data and then explore it in a notebook or dashboard.

+ New



# Compute in databricks

The screenshot shows the Databricks web interface for configuring a new cluster. The left sidebar contains navigation links for Workspace, Recents, Catalog, Workflows, Compute (selected), Marketplace, SQL, SQL Editor, Queries, Dashboards, Genie, Alerts, Query History, SQL Warehouses, Data Engineering, Job Runs, Data Ingestion, Pipelines, and Machine Learning. The top header shows 'Microsoft Azure' and 'databricks' logos, a search bar, and the user profile 'tokyo-olympic-db'. The main content area is titled 'Compute > Simple form: OFF' and displays the configuration for 'Mahendra Chinchkhede's Cluster'. The configuration is organized into sections: Policy (Unrestricted), Access mode (Dedicated, Single user or group access: Mahendra Chinchkhede), Performance (Databricks Runtime Version: 12.2 LTS, Use Photon Acceleration: unchecked), Node type (Standard\_D4ds\_v5, 16 GB Memory, 4 Cores), and Tags (No custom tags). A 'Summary' panel on the right provides a quick overview of the cluster specifications. Buttons for 'Terminate' and 'Edit' are located at the top right of the configuration area.

Microsoft Azure | databricks

Search data, notebooks, recents, and more... CTRL + P

tokyo-olympic-db

+ New

- Workspace
- Recents
- Catalog
- Workflows
- Compute**
- Marketplace
- SQL
- SQL Editor
- Queries
- Dashboards
- Genie
- Alerts
- Query History
- SQL Warehouses
- Data Engineering
- Job Runs
- Data Ingestion
- Pipelines
- Machine Learning

Compute > Simple form: OFF

## Mahendra Chinchkhede's Cluster

Terminate Edit

Send feedback

**Configuration** Notebooks (1) Libraries Event log Spark UI Driver logs Metrics Apps Spark compute UI - Master

Policy ⓘ

Unrestricted

☐ Multi node ☒ Single node

Access mode ⓘ

Dedicated (formerly: Single user) Single user or group access ⓘ

Mahendra Chinchkhede

**Performance**

Databricks Runtime Version

12.2 LTS (includes Apache Spark 3.3.2, Scala 2.12)

☐ Use Photon Acceleration ⓘ

Node type ⓘ

Standard\_D4ds\_v5 16 GB Memory, 4 Cores

☒ Terminate after 120 minutes of inactivity ⓘ

**Tags** ⓘ

No custom tags

### Summary

1 Driver	16 GB Memory, 4 Cores
Runtime	12.2.x-scala2.12
Unity Catalog	Standard_D4ds_v5
1 DBU/h	

# Loading data and basic Transformation in Notebook

# Notebook – connection & mounting

The screenshot shows a Databricks notebook interface. The top bar includes the Microsoft Azure logo, the Databricks logo, a search bar, and the text 'tokyo-olympic-db'. The notebook title is 'Tokyo Olympic Transformation', and it is written in Python. The left sidebar contains a navigation menu with options like 'New', 'Workspace', 'Recents', 'Catalog', 'Workflows', 'Compute', 'Marketplace', 'SQL', 'SQL Editor', 'Queries', 'Dashboards', 'Genie', 'Alerts', 'Query History', 'SQL Warehouses', 'Data Engineering', 'Job Runs', 'Data Ingestion', 'Pipelines', and 'Machine Learning'. The main area displays a code cell with the following code:

```
configs = {"fs.azure.account.auth.type": "OAuth",
"fs.azure.account.oauth.provider.type": "org.apache.hadoop.fs.azurebfs.oauth2.ClientCredsTokenProvider",
"fs.azure.account.oauth2.client.id": "2f1b8b1c-0a9b-4f4e-af3b-b06f0f9a185b", clientid
"fs.azure.account.oauth2.client.secret": "h7K8Q~L0k_8D2XJ_X1amKwmUH53DZrTonZT34dxq", secretid
"fs.azure.account.oauth2.client.endpoint": "https://login.microsoftonline.com/60f33b66-293e-41e0-bf30-9c97f1c81e2a/oauth2/token"}

dbutils.fs.mount(Container name@storage name
source = "abfss://tokyo-olympic-data@actokyoolympic.dfs.core.windows.net", # container@storageacc name
mount_point = "/mnt/tokyoolympic",
extra_configs = configs) Mount name
```

Below the code cell, there are instructions: [Shift+Enter] to run and move to next cell, [Ctrl+Shift+P] to open the command palette, and [Esc H] to see all keyboard shortcuts.

For connection purpose

- create App Registration
- get clientid and tenantid
- in certificates&secrets create secretid

+ New

Workspace

Recents

Catalog

Workflows

Compute

Marketplace

SQL

SQL Editor

Queries

Dashboards

Genie

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Pipelines

Machine Learning

## Tokyo Olympic Transformation Python

Tabs: OFF



File Edit View Run Help Last edit was now

Run all

Mahendra Chinchkhed...

Schedule

Share



```
03:08 PM (15s) 1 Python

configs = {"fs.azure.account.auth.type": "OAuth",
           "fs.azure.account.oauth.provider.type": "org.apache.hadoop.fs.azurebfs.oauth2.ClientCredsTokenProvider",
           "fs.azure.account.oauth2.client.id": "2f1b8b1c-0a9b-4f4e-af3b-b06f0f9a185b",
           "fs.azure.account.oauth2.client.secret": 'h7K8Q~L0k_8D2XJ_X1amKwmUH53DZrTonZT34dxq',
           "fs.azure.account.oauth2.client.endpoint": "https://login.microsoftonline.com/60f33b66-293e-41e0-bf30-9c97f1c81e2a/oauth2/token"}

dbutils.fs.mount(
    source = "abfss://tokyo-olympic-data@actokyoolympic.dfs.core.windows.net", # contrainer@storageeacc name
    mount_point = "/mnt/tokyoolympic",
    extra_configs = configs)
```

Out[1]: True

```
Just now (11s) 2

%fs
ls "/mnt/tokyoolympic"
```


Mounted datalake to this location

Table



	path	name	size	modificationTime
1	dbfs/mnt/tokyoolympic/raw-data/	raw-data/	0	1747900857000
2	dbfs/mnt/tokyoolympic/transformed-dat...	transformed-dat...	0	1747900866000



Microsoft Azure

 databricks

Search data, notebooks, recents, and more...

CTRL + P

tokyo-olympic-db



New

Workspace

Recents

Catalog

Workflows

Compute

Marketplace

SQL

SQL Editor

Queries

Dashboards

Genie

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Tokyo Olympic Transformation

Python

Tabs: OFF

☆

File Edit View Run Help

Last edit was 1 minute ago

Run all

Mahendra Chinchkhed...

Schedule

Share

2 rows | 10.74s runtime

refreshed 10 minutes ago

03:18 PM (<1s)

3

Python

spark

SparkSession - hive

SparkContext

[Spark UI](#)

Version

v3.3.2

Master

local[\*, 4]

AppName

Databricks Shell

Spark session active

1 minute ago (<1s)

4

athletes = spark.read.format("csv").option("header", "true").load("/mnt/tokyoolympic/raw-data/athletes.csv")

(1) Spark Jobs

athletes: pyspark.sql.dataframe.DataFrame = [PersonName: string, Country: string ... 1 more field]

Loading athletes.csv file

3 minutes ago (<1s)

5

Python

athletes.show()

(1) Spark Jobs

Records fetched successfully

PersonName	Country	Discipline
AALERUD Katrine	Norway	Cycling Road
ABAD Nestor	Spain	Artistic Gymnastics
ABAGALE Giovanni	Italy	Rowing
ABALDE Alberto	Spain	Basketball
ABALDE Tamara	Spain	Basketball
ABALO Luc	France	Handball
ABARCA Cesar	Chile	Rowing
ABAS Abubakar	Sudan	Swimming
ABASALI Hamideh	Islamic Republic of Iran	Karate
ABBASOV Islam	Azerbaijan	Wrestling
ABBINGH Lois	Netherlands	Handball
ABBOT Emily	Australia	Rhythmic Gymnastics
ABBOTT Ronica	United States of America	Baseball/Softball
ABDALLA Abubakar	Qatar	Athletics
ABDALLA Maryam	Egypt	Artistic Swimming
ABDALLAH Shahid	Egypt	Artistic Swimming
ABDULRASOOL Mohamed	Sudan	Judo
ABDEL LATIF Radwan	Egypt	Shooting



Just now (<1s)

`entriesgender.show()`

▶ (1) Spark Jobs

Discipline	Female	Male	Total
3x3 Basketball	32	32	64
Archery	64	64	128
Artistic Gymnastics	98	98	196
Artistic Swimming	105	0	105
Athletics	969	1072	2041
Badminton	86	87	173
Baseball/Softball	90	144	234
Basketball	144	144	288
Beach Volleyball	48	48	96
Boxing	102	187	289
Canoe Slalom	41	41	82
Canoe Sprint	123	126	249
Cycling BMX Frees...	10	9	19
Cycling BMX Racing	24	24	48
Cycling Mountain ...	38	38	76
Cycling Road	70	131	201
Cycling Track	90	99	189
Diving	72	71	143



1 minute ago (<1s)

`entriesgender.printSchema()`

root

```
-- Discipline: string (nullable = true)
-- Female: string (nullable = true)
-- Male: string (nullable = true)
-- Total: string (nullable = true)
```

For entriesgender table, column Female, Male & Total does not have correct datatype



▶ ✓ 03:45 PM (<1s) 11

```
#We dont have Columns in proper datatypes, so we need to change.
from pyspark.sql.functions import col
from pyspark.sql.types import IntegerType, DateType, DoubleType, BooleanType
```

▶ ✓ Just now (<1s) 12

```
entriesgender = entriesgender.withColumn("Female",col("Female").cast(IntegerType()))\
    .withColumn("Male",col("Male").cast(IntegerType()))\
    .withColumn("Total",col("Total").cast(IntegerType()))
```

▶ entriesgender: pyspark.sql.dataframe.DataFrame = [Discipline: string, Female: integer ... 2 more fields]

⋮  
▼ ▶ ✓ Just now (<1s) 13

```
entriesgender.printSchema()
```

Converted Female,Male,Total Column  
datatype into integer

Python



root

```
-- Discipline: string (nullable = true)
-- Female: integer (nullable = true)
-- Male: integer (nullable = true)
-- Total: integer (nullable = true)
```



3 minutes ago (2s)

4

Python

```
athletes = spark.read.format("csv").option("header", "true").load("/mnt/tokyoolympic/raw-data/athletes.csv")
coaches = spark.read.format("csv").option("header", "true").load("/mnt/tokyoolympic/raw-data/coaches.csv")
entriesgender = spark.read.format("csv").option("header", "true").load("/mnt/tokyoolympic/raw-data/entriesgender.csv")
medals = spark.read.format("csv").option("header", "true").option("inferSchema", "true").load("/mnt/tokyoolympic/raw-data/medals.csv")
teams = spark.read.format("csv").option("header", "true").option("inferSchema", "true").load("/mnt/tokyoolympic/raw-data/teams.csv")
```

(7) Spark Jobs

- athletes: pyspark.sql.dataframe.DataFrame = [PersonName: string, Country: string ... 1 more field]
- coaches: pyspark.sql.dataframe.DataFrame = [Name: string, Country: string ... 2 more fields]
- entriesgender: pyspark.sql.dataframe.DataFrame = [Discipline: string, Female: string ... 2 more fields]
- medals: pyspark.sql.dataframe.DataFrame = [Rank: integer, TeamCountry: string ... 5 more fields]
- teams: pyspark.sql.dataframe.DataFrame = [TeamName: string, Discipline: string ... 2 more fields]

Instead of manually formatting schema, we can use inferSchema function in spark - it will automatically identify datatype and convert accordingly



# Find top countries with the highest number of gold medals

The screenshot shows the Databricks interface for a notebook titled "Tokyo Olympic Transformation". The notebook is written in Python and contains a query to find the top countries with the highest number of gold medals. The query is as follows:

```
#Find the top countries with the highest number of gold medals

highest_gold_contries = medals.orderBy("Gold", ascending=False).select("TeamCountry", "Gold").show()
```

The output of the query is displayed as a table with two columns: TeamCountry and Gold. The table shows the top 20 countries by gold medals.

TeamCountry	Gold
United States of ...	39
People's Republic...	38
Japan	27
Great Britain	22
ROC	20
Australia	17
Netherlands	10
France	10
Germany	10
Italy	10
Canada	7
Brazil	7
New Zealand	7
Cuba	7
Hungary	6
Republic of Korea	6
Poland	4
Czech Republic	4

# Calculate average number of entries by gender for each discipline

Microsoft Azure databricks

Search data, notebooks, recents, and more... CTRL + P

tokyo-olympic-db

New

Workspace

Recents

Catalog

Workflows

Compute

Marketplace

SQL

SQL Editor

Queries

Dashboards

Genie

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Pipelines

Machine Learning

## Tokyo Olympic Transformation

Python Tabs: OFF

File Edit View Run Help [Last edit was now](#)

Run all Mahendra Chinchkhed... Schedule Share

5 minutes ago (<1s) 19

Python

```
#Calculate the average number of entries by gender for each discipline

avg_entries_by_gender = entriesgender.withColumn(
    'Avg_Female', entriesgender['Female'] / entriesgender['Total']
).withColumn(
    'Avg_Male', entriesgender['Male'] / entriesgender['Total']
).show()
```

(1) Spark Jobs

Discipline	Female	Male	Total	Avg_Female	Avg_Male
3x3 Basketball	32	32	64	0.5	0.5
Archery	64	64	128	0.5	0.5
Artistic Gymnastics	98	98	196	0.5	0.5
Artistic Swimming	105	0	105	1.0	0.0
Athletics	969	1072	2041	0.4747672709456149	0.5252327290543851
Badminton	86	87	173	0.49710982658959535	0.5028901734104047
Baseball/Softball	90	144	234	0.38461538461538464	0.6153846153846154
Basketball	144	144	288	0.5	0.5
Beach Volleyball	48	48	96	0.5	0.5
Boxing	102	187	289	0.35294117647058826	0.6470588235294118
Canoe Slalom	41	41	82	0.5	0.5
Canoe Sprint	123	126	249	0.4939759036144578	0.5060240963855421
Cycling BMX Frees...	10	9	19	0.5263157894736842	0.47368421052631576
Cycling BMX Racing	24	24	48	0.5	0.5
Cycling Mountain ...	38	38	76	0.5	0.5

Microsoft Azure databricks

Search data, notebooks, recents, and more... CTRL + P

tokyo-olympic-db

+ New

Workspace

Recents

Catalog

Workflows

Compute

Marketplace

SQL

SQL Editor

Queries

Dashboards

Genie

Alerts

Query History

### Tokyo Olympic Transformation

Python Tabs: OFF

File Edit View Run Help Last edit was 1 minute ago

Run all Mahendra Chinchkhed... Schedule Share

```
#Writing Transformed Data
```

For no file parts To save changes made in same file

```
athletes.repartition(1).write.mode("overwrite").option("header",True).csv("/mnt/tokyoolympic/transformed-data/athletes")
```

Mount location File name

(2) Spark Jobs

Start typing or generate with AI (Ctrl + I)...

Microsoft Azure Upgrade Search resources, services, and docs (G+V) Copilot

Home > actkyoolympic | Containers >

### tokyo-olympic-data

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

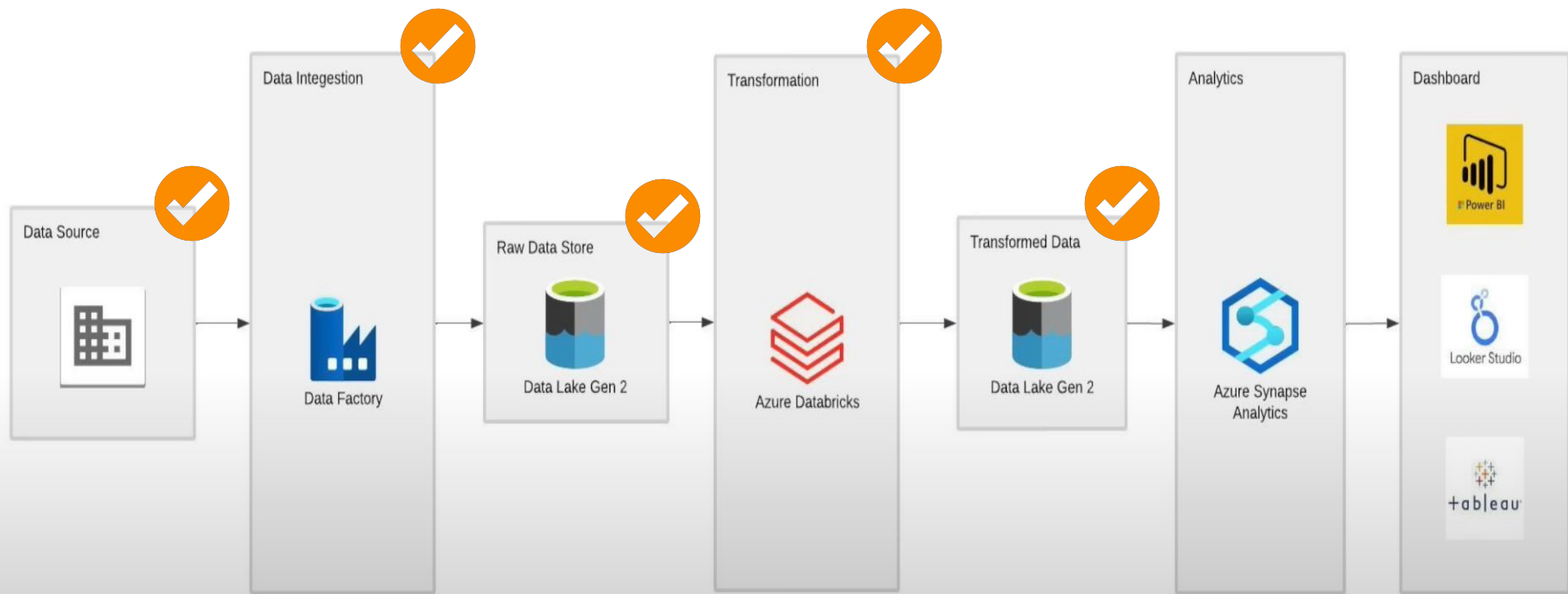
Authentication method: Access key (Switch to Microsoft Entra user account)

Location: tokyo-olympic-data / transformed-data / athletes

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/> _committed_7644537169421703333	23/05/2025, 19:23:38	Hot (Inferred)		Block blob	112 B	Available ***
<input type="checkbox"/> _started_7644537169421703333	23/05/2025, 19:23:37	Hot (Inferred)		Block blob	0 B	Available ***
<input type="checkbox"/> _SUCCESS	23/05/2025, 19:23:38	Hot (Inferred)		Block blob	0 B	Available ***
<input type="checkbox"/> part-00000-tid-7644537169421703333-0ad54187-...	23/05/2025, 19:23:37	Hot (Inferred)		Block blob	397.91 KiB	Available ***

Transformed data saved to desired location successfully



# Azure Synapse Analytics

Home >

## Microsoft.Azure.SynapseAnalytics-20250523193846 | Overview

Deployment

Search

Delete Cancel Redeploy Download Refresh

### Overview

Inputs

Outputs

Template

## ✓ Your deployment is complete

Deployment name : Microsoft.Azure.SynapseAnalytics-20250523193846  
Subscription : [Azure subscription 1](#)  
Resource group : [ac-data-engineering-projects](#)

Start time : 23/05/2025, 19:44:16  
Correlation ID : 4c021ab2-b5f5-4808-85ab-e33be3cd8221

> Deployment details

✓ Next steps

[Go to resource group](#)

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 >](#)

# Azure Synapse Analytics

Microsoft Azure | Synapse Analytics | actokyo-olympic-sa

Search

Synapse live Validate all Publish all 1

Home Data Develop Integrate Monitor Manage

Data

Workspace Linked

Filter resources by name

Lake database 1

TokyoOlympicDB

Tables

athletes

Columns

TokyoOlympicDB SQL script 1

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

```
1 SELECT * FROM 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
ARAI DF Tamara	Spain	Basketball

00:00:01 Query executed successfully.

Properties

General Related (0)

Name \*  
SQL script 1

Description

Type  
.sql script

Size  
94 bytes

Results settings per query ①

☒ First 5000 rows (default)

☐ All rows

- Created Lake Database (TokyoOlympicDB)
- Imported table athletes from transformed data folder
- performed SQL queries on data

Microsoft Azure | Synapse Analytics | actokyo-olympic-sa

Search

Synapse live Validate all Publish all

Data

Workspace Linked

Filter resources by name

Lake database 1

TokyoOlympicDB

Tables

- athletes
- coaches
- entriesgender
- medals
- teams

TokyoOlympicDB SQL script 1

Run Undo Publish Query plan

Connect to Built-in Use database TokyoOlympicDB

```
1 SELECT * FROM coaches;
```

Results Messages

View Table Chart Export results

Name	Country	Discipline	Event
ABDELMAGID Wael	Egypt	Football	(NULL)
ABE Junya	Japan	Volleyball	(NULL)

00:00:01 Query executed successfully.

Properties

General Related (0)

Name \* SQL script 1

Description

Type .sql script

Size 94 bytes

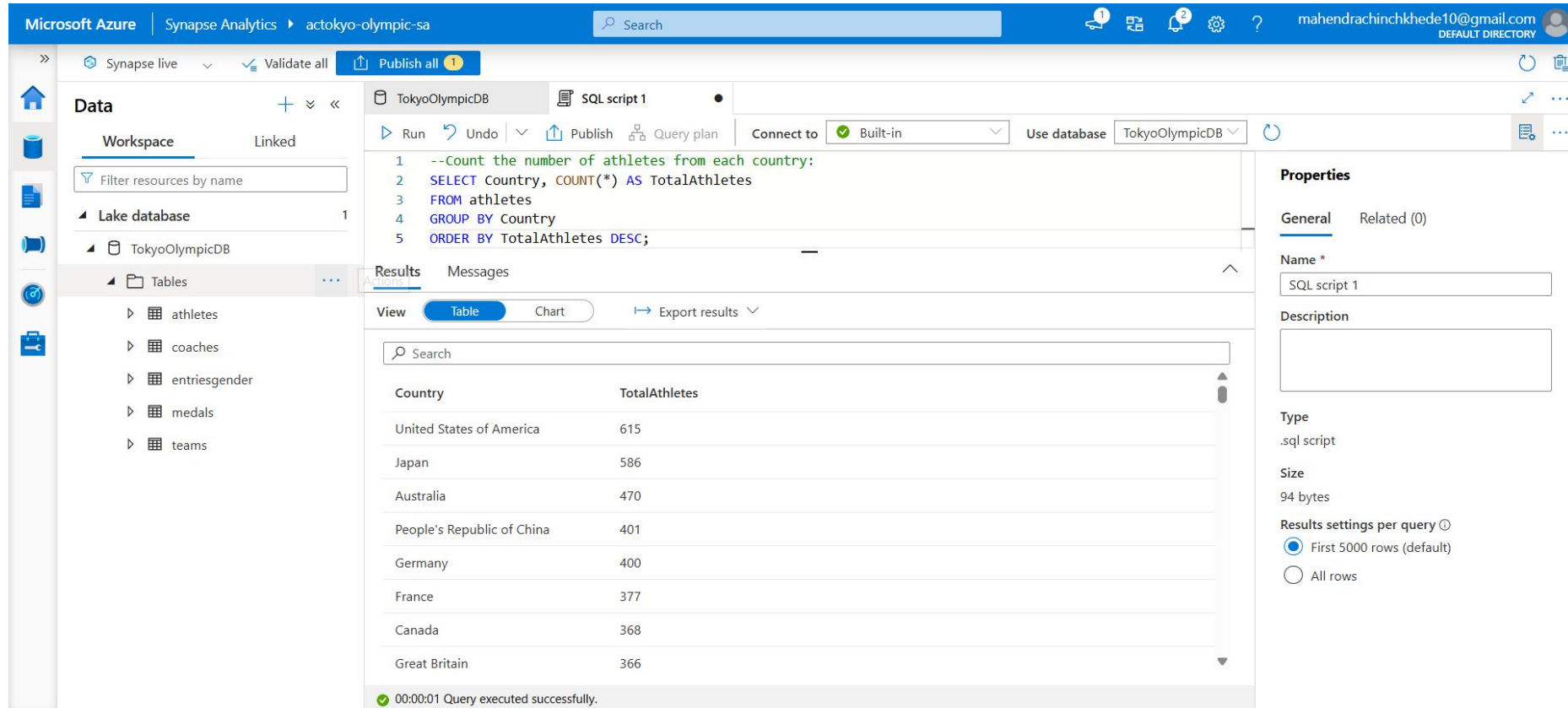
Results settings per query

- ☒ First 5000 rows (default)
- ☐ All rows

Similarly imported other tables



# Count the number of athletes from each country



The screenshot shows the Microsoft Azure Synapse Analytics interface. The top bar indicates the user is logged in as mahendrachinchkhede10@gmail.com. The workspace is named 'actokyo-olympic-sa'. The left sidebar shows the 'Data' section with a 'Workspace' tab. The 'Lake database' section is expanded, showing the 'TokyoOlympicDB' database. The 'Tables' section is also expanded, showing a list of tables: athletes, coaches, entriesgender, medals, and teams. The 'athletes' table is selected. The main area displays the SQL script 'SQL script 1' with the following query:

```
1 --Count the number of athletes from each country:
2 SELECT Country, COUNT(*) AS TotalAthletes
3 FROM athletes
4 GROUP BY Country
5 ORDER BY TotalAthletes DESC;
```

The query is executed successfully, and the results are displayed in a table. The table has two columns: 'Country' and 'TotalAthletes'. The results are as follows:

Country	TotalAthletes
United States of America	615
Japan	586
Australia	470
People's Republic of China	401
Germany	400
France	377
Canada	368
Great Britain	366

The status bar at the bottom indicates '00:00:01 Query executed successfully.' The right sidebar shows the 'Properties' section for the 'SQL script 1' file, with fields for Name, Description, Type, Size, and Results settings per query.

SQL query Operation

# Calculate total medals won by each country

The screenshot displays the Microsoft Azure Synapse Analytics interface. The top navigation bar shows 'Microsoft Azure | Synapse Analytics | actokyo-olympic-sa'. The left sidebar contains a 'Data' section with a 'Workspace' tab and a 'Lake database' section. The 'Lake database' section shows a folder named 'TokyoOlympicDB' containing several tables: 'athletes', 'coaches', 'entriesgender', 'medals', and 'teams'. The main area shows a SQL script named 'SQL script 1' with the following code:

```
-- Calculate the total medals won by each country:
SELECT TeamCountry,
SUM(Gold) Total_Gold,
SUM(Silver) Total_Silver,
SUM(Bronze) Total_Bronze
FROM medals
GROUP BY TeamCountry
ORDER BY Total_Gold DESC;
```

The 'Results' tab is selected, showing a table with the following data:

TeamCountry	Total_Gold	Total_Silver	Total_Bronze
United States of America	39	41	33
People's Republic of China	38	32	18
Japan	27	14	17
Great Britain	22	21	22
ROC	20	28	23
Australia	17	7	22

The bottom status bar indicates '00:00:00 Query executed successfully.' The right sidebar shows the 'Properties' section for the 'SQL script 1' file, including fields for 'Name', 'Description', and 'Type' (set to '.sql script'). The 'Results settings per query' section shows 'First 5000 rows (default)' selected.

# Calculate the average number of entries by gender for each discipline

The screenshot displays the Microsoft Azure Synapse Analytics workspace. The top navigation bar shows the user is logged in as 'mahendrachinchkhede10@gmail.com'. The workspace is named 'actokyo-olympic-sa'. The left sidebar shows the 'Data' section with a 'Workspace' tab. Under 'Lake database', the 'TokyoOlympicDB' database is selected, and the 'entriesgender' table is visible. The main area shows a SQL script named 'SQL script 1' with the following query:

```
17
18 -- Calculate the average number of entries by gender for each discipline:
19 SELECT Discipline,
20        AVG(Female) Avg_Female,
21        AVG(Male) Avg_Male
22 FROM entriesgender
23 GROUP BY Discipline;
24
```

The query is executed, and the results are displayed in a table view. The table has three columns: 'Discipline', 'Avg\_Female', and 'Avg\_Male'. The results are as follows:

Discipline	Avg_Female	Avg_Male
3x3 Basketball	32	32
Archery	64	64
Artistic Gymnastics	98	98
Artistic Swimming	105	0
Athletics	969	1072
Badminton	86	87

The status bar at the bottom indicates '00:00:02 Query executed successfully.' The right sidebar shows the 'Properties' section for the script, including the name 'SQL script 1' and the type '.sql script'.

PowerBI

[Home](#) > [Resource groups](#) >

## Resource groups

Default Directory (mahendrachinchkhede10gmail....)

[+ Create](#) [...](#) [Group by none](#) ▾

You are viewing a new version of Browse experience. Some features may be missing. [Click here to access the old experience.](#)

- ☐ Name ↑
- ☐ ac-data-engineering-projects
- ☐ databricks-rg-tokyo-olympic-db-jc
- ☐ NetworkWatcherRG
- ☐ synapseworkspace-managedrg-bf



## ac-data-engineering-projects

Resource group



### Overview

Activity log

Access control (IAM)

Tags

Resource visualizer

Events

&gt; Settings

&gt; Cost Management

&gt; Monitoring

&gt; Automation

&gt; Help

### Essentials

#### Resources

Recommendations (1)

[JSON View](#)

Type equals all ×

Location equals all ×

Add filter

Showing 1 to 4 of 4 records. ☐ Show hidden types ⓘ

No grouping ▾

List view ▾

<input type="checkbox"/> Name ↑↓	Type ↑↓	Location ↑↓	
<input type="checkbox"/> actokyo-olympic-df	Data factory (V2)	Southeast Asia	...
<input type="checkbox"/> actokyo-olympic-sa	Synapse workspace	Central India	...
<input type="checkbox"/> actokyoolympic	Storage account	West India	...
<input type="checkbox"/> tokyo-olympic-db	Azure Databricks Service	Central India	...

&lt; Previous

Page

1 ▾

of 1

Next &gt;

Give feedback

Add or remove favorites by pressing Ctrl+Shift+F