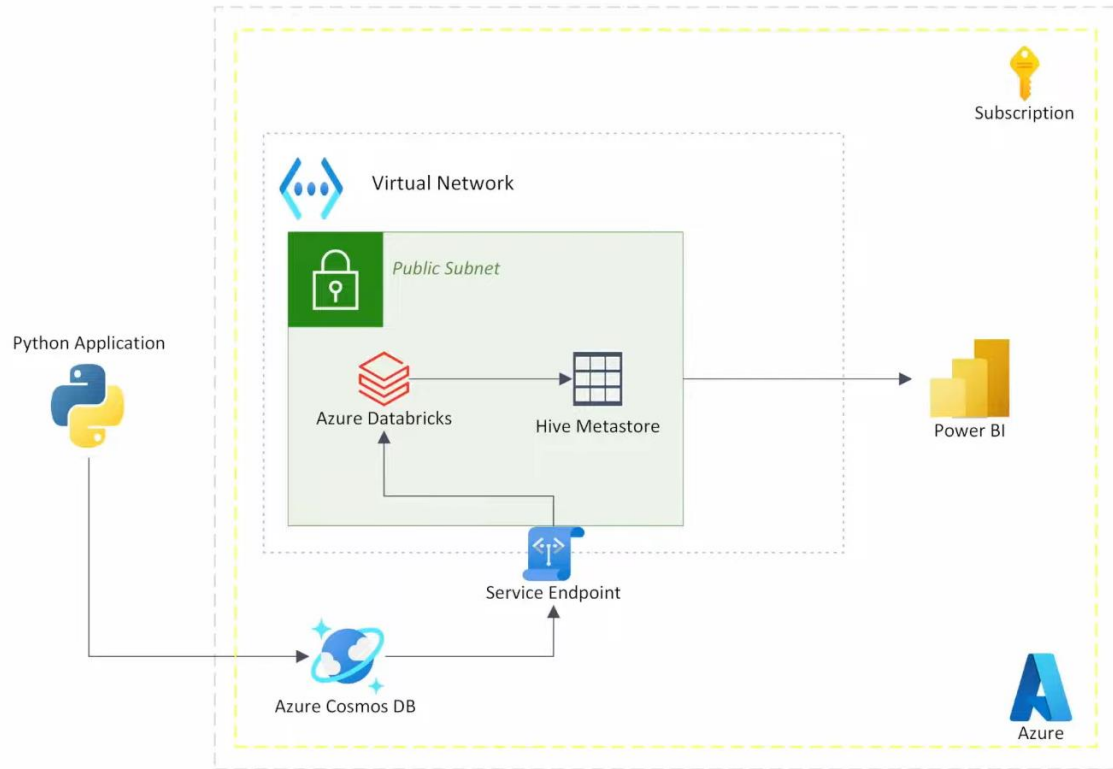


Azure Cosmos DB API para MongoDB utilizando Spark y Databricks con PowerBI mediante Partner Connect

Arquitectura:



Creamos el grupo de recursos

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

Create a resource group ...

[Basics](#) [Tags](#) [Review + create](#)

Resource group - A container that holds related resources for an Azure solution. The resource group can include all the resources for the solution, or only those resources that you want to manage as a group. You decide how you want to allocate resources to resource groups based on what makes the most sense for your organization. [Learn more](#)

Project details

Subscription * ⓘ

Azure for Students

Resource group * ⓘ

MondoDB

Resource details

Region * ⓘ

(Europe) France Central

Azure Cosmos DB

Home > Azure Cosmos DB Tajamar

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

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

Showing 0 to 0 of 0 records.

Name	Status	Subscription	Write region
------	--------	--------------	--------------

No Azure Cosmos DB accounts to display

Create a globally distributed, multi-model, fully managed database using API of your choice. Or try it for free, up to 20k RU/s, for 30 days with unlimited renewal.

Create Azure Cosmos DB account

Try now

Click en create

Home > Azure Cosmos DB > Create an Azure Cosmos DB account

Which API best suits your workload?

Azure Cosmos DB is a fully managed NoSQL and relational database service for building scalable, high performance applications. [Learn more](#)

To start, select the API to create a new account. The API selection cannot be changed after account creation.

Recommended APIs Others

Azure Cosmos DB for NoSQL

Azure Cosmos DB's core, or native API for working with documents. Supports fast, flexible development with familiar SQL query language and client libraries for .NET, JavaScript, Python, and Java.

Create Learn more

Azure Cosmos DB for MongoDB

Fully managed database service for apps written for MongoDB. Recommended if you have existing MongoDB workloads that you plan to migrate to Azure Cosmos DB.

Create Learn more

Give Feedback

Help improve this page

Escoger Azure Cosmos DB for MongoDB luego create

Create Azure Cosmos DB Account - Choose Architecture

Which type of resource?

Azure Cosmos DB for MongoDB offers two resource types with different architectures. Request unit (RU) database accounts and vCore clusters. [See documentation to learn more.](#)

To start, select the type to create a resource. The resource selection cannot be changed after creation.

Request unit (RU) database account

- Industry-leading 99.999% availability
- Instantaneous, granular autoscaling
- Serverless accounts
- [See documentation and supported features](#)

Create

vCore cluster (Recommended)

- Familiar architecture
- High-capacity vertical and horizontal scaling
- Ideal for long-running queries and complex aggregation pipelines
- [See documentation and supported features](#)

Create

Por ahora, seleccionamos el Serverless (Request unit – RU database account)

Basics Global distribution Networking Backup Policy Encryption Tags Review + create

Azure Cosmos DB is a fully managed NoSQL and relational database service for building scalable, high performance applications. Go to production starting at \$24/month per database, multiple containers included. [Learn more](#)

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

Subscription * Azure for Students

Resource Group * MondoDB
[Create new](#)

Instance Details

Account Name * cosmosdbwithmongodb

Configure availability zone settings for your account. You cannot change these settings once the account is created.

Availability Zones ☐ Enable ☒ Disable

Location * (Europe) France Central
Available locations are determined by your subscription's access and availability zone support (if that is enabled). If you don't see or cannot select your desired location, please open a support request for region access.
[Click here for more details on how to create a region access request](#)

Capacity mode ☐ Provisioned throughput ☒ Serverless
[Learn more about capacity mode](#)

Version 7.0

[Review + create](#) [Previous](#) [Next: Global distribution](#)

Next: Global distribution

Create Azure Cosmos DB Account - Azure Cosmos DB for MongoDB

Basics **Global distribution** Networking Backup Policy Encryption Tags Review + create

Multiple regions are not supported with serverless capacity mode.

Geo-Redundancy ☐ Enable ☒ Disable

Multi-region Writes ☐ Enable ☒ Disable

[Review + create](#) [Previous](#) [Next: Networking](#)

Disable, luego Next: Networking

Basics Global distribution **Networking** Backup Policy Encryption Tags Review + create

Network connectivity

You can connect to your Azure Cosmos DB account either publically, via public IP addresses or service endpoints, or privately, using a private endpoint.

Connectivity method * ☒ All networks
☐ Public endpoint (selected networks)
☐ Private endpoint

All networks will be able to access this CosmosDB account. <http://aka.ms/network-security>

Connection Security Settings

Minimum Transport Layer Security Protocol

i This account only accepts this protocol. [Learn more](#)


[Review + create](#) [Previous](#) [Next: Backup Policy](#)

All networks, luego Next: Backup Policy

Create Azure Cosmos DB Account - Azure Cosmos DB for MongoDB

Basics Global distribution Networking **Backup Policy** Encryption Tags Review + create


Azure Cosmos DB provides three different backup policies. You will not be able to switch to Periodic mode once you adopt Continuous mode. [Learn more](#) about the differences of the backup policies and pricing details.

Backup policy 


☒ Periodic
Backup is taken at periodic interval based on your configuration

☐ Continuous (7 days)
Provides backup window of 7 days / 168 hours and you can restore to any point of time within the window. This mode is available for free.

☐ Continuous (30 days)
Provides backup window of 30 days / 720 hours and you can restore to any point of time within the window. This mode has cost impact.

Backup interval 

60-1440

Backup retention 

8-720

Copies of data retained

Backup storage redundancy *

☐ Geo-redundant backup storage

☐ Zone-redundant backup storage

☒ Locally-redundant backup storage

Periodic

Locally-redundant backup storage

Lo demás lo que se muestra por defecto. Luego click en Next: Encryption

Basics Global distribution Networking Backup Policy **Encryption** Tags Review + create

Data Encryption

Azure Cosmos DB encryption protects your data at rest by seamlessly encrypting your data as it's written in our datacenters, and automatically decrypting it for you as you access it.

By default your Azure Cosmos DB account is encrypted at rest using service-managed keys. At the moment, you will not be able to switch back to service-managed key after opting into using custom-managed key while creating your account. [Learn More](#)

Data Encryption *

☒ Service-managed key

☐ Customer-managed key (CMK)

Service-managed key, luego click en Review + Create

Create Azure Cosmos DB Account - Azure Cosmos DB for MongoDB ...

✓ Validation Success

BasicsGlobal distributionNetworkingBackup PolicyEncryptionTagsReview + create

Creation Time

Estimated Account Creation Time (in minutes)2

ⓘ

The estimated creation time is calculated based on the location you have selected

Basics

Subscription	Azure for Students
Resource Group	MondoDB
Location	France Central
Account Name	(new) cosmosdbwithmongodb
API	Azure Cosmos DB for MongoDB
Capacity mode	Serverless
Availability Zones	Disable

Backup Policy

Backup policy	Periodic
Backup storage redundancy	Locally-redundant backup storage

Encryption

Data Encryption	Service Managed
-----------------	-----------------

Networking

Connectivity method	All networks
Minimum TLS Protocol	TLS 1.2

Create

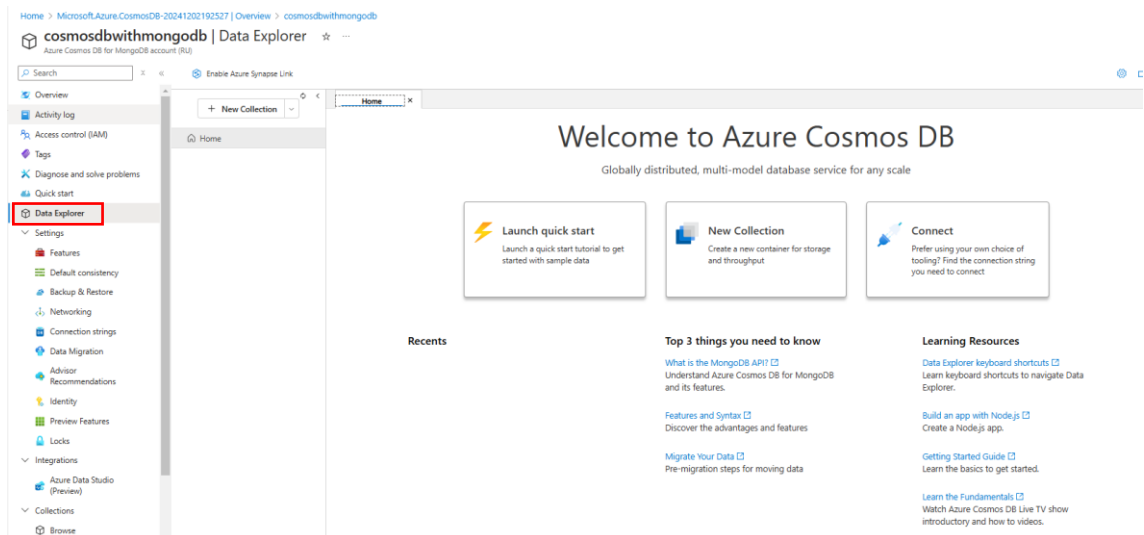
Previous

Next

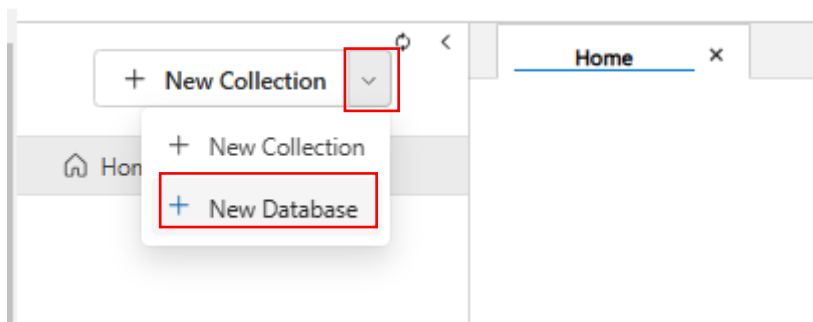
[Download a template for automation](#)

Create

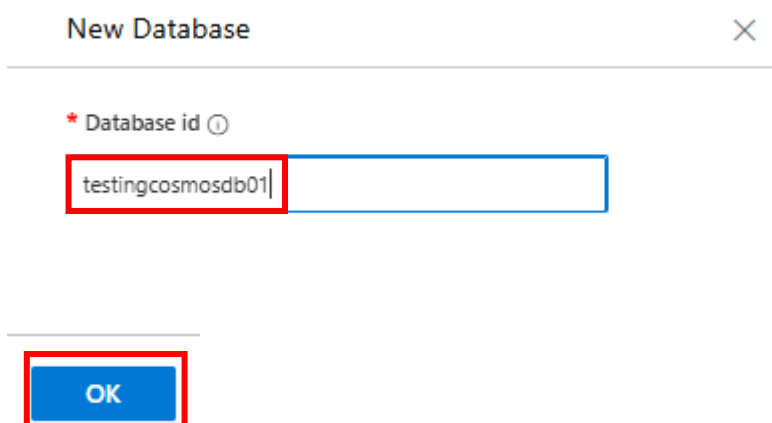
Una vez efectuado el deploy ir a go to ressources luego click Data Explorer



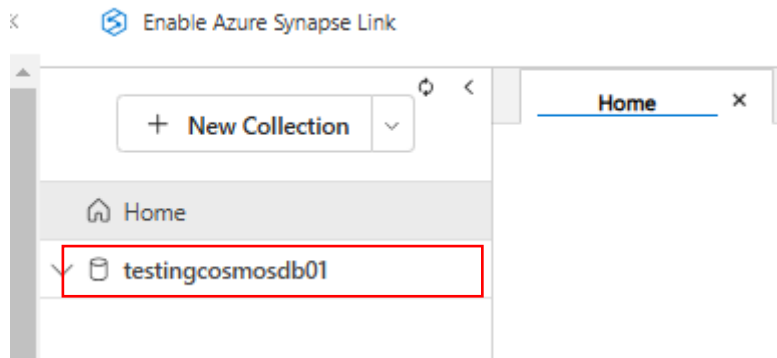
Luego click en +New Collection y seleccionar New Database



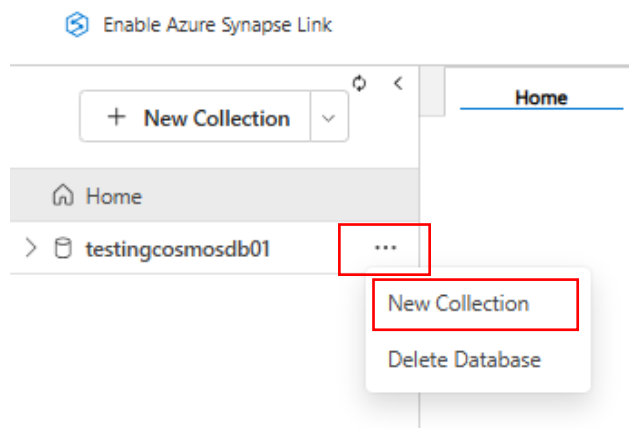
En Database id asignamos un nombre, luego OK



Una vez creado debes verlo así:



Click en los tres puntos y luego new collection:



Use existing, añadir un nombre en Collection id, seleccionar Unsharded

New Collection

* Database name ⓘ

☐ Create new ☒ Use existing

testingcosmosdb01

* Collection id ⓘ

pruebadatabrickcosmosdb

* Sharding ⓘ

☒ Unsharded (20GB limit) ☐ Sharded

Analytical store ⓘ

☐ On ☒ Off

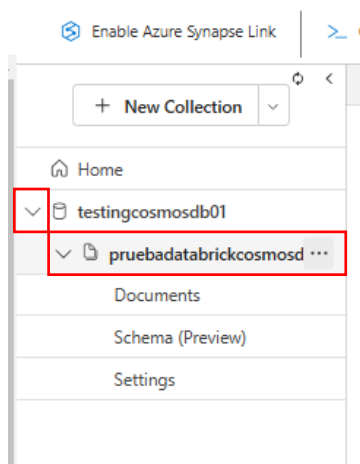
Azure Synapse Link is required for creating an analytical store collection. Enable Synapse Link for this Cosmos DB account. [Learn more](#)

Enable

> Advanced

OK

Una vez creado debería verse esto:



Entra al sitio web [OpenWeather](#) y regístrate.

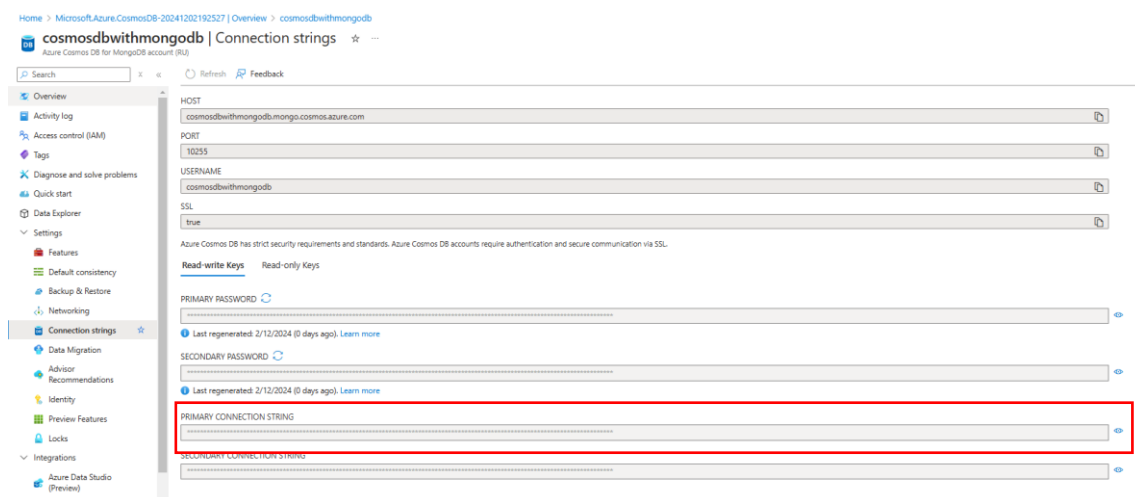
Accede a la sección de API Keys en tu perfil después de registrarte.

Copia la API Key que se te proporciona

Reemplaza tu_api_key en el [script](#) con la Key que has copiado.

```
# Configuración de OpenWeather API
API_KEY = "8b73515bcec2ae54c33e6d422f504f2a" # Reemplaza con tu API Key
CITY = "Toronto" # Cambia por la ciudad deseada
WEATHER_URL = f"http://api.openweathermap.org/data/2.5/weather?q={CITY}&appid={API_KEY}&units=metric"
```

En Azure, en tu CosmosDB , ve a Connection strings y copia PRIMARY CONNECTION STRING:



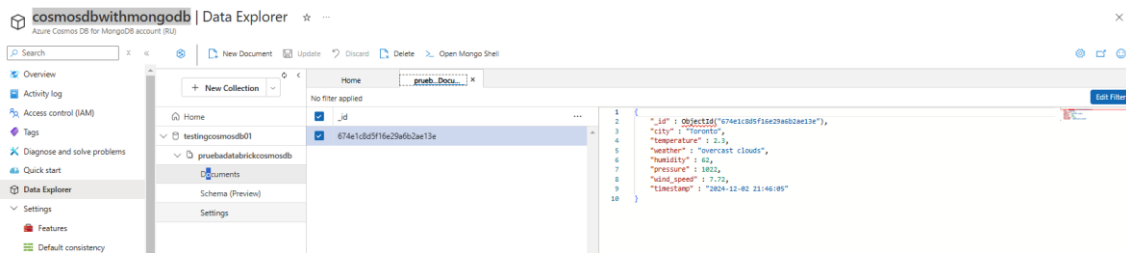
Pegar tu PRIMARY CONNECTION STRING en el [script](#) (weather_to_cosmodb.py)

```
# Configuración de Cosmos DB
DB_NAME = "testingcosmosdb01" # Nombre de tu base de datos
COLLECTION_NAME = "pruebadatabrickcosmosdb" # Nombre de tu colección
CONNECTION = "mongodb://cosmosdbwithmongodb:fUTd69EFQvowiFX0yn4sBTJ3e0yJoMaRAC3hUr1FSZnAKIheVJwezrgr01Jw5GVnFjLWnho3YkRcACDbPhcU3g==@cosmosdbwithr"
```

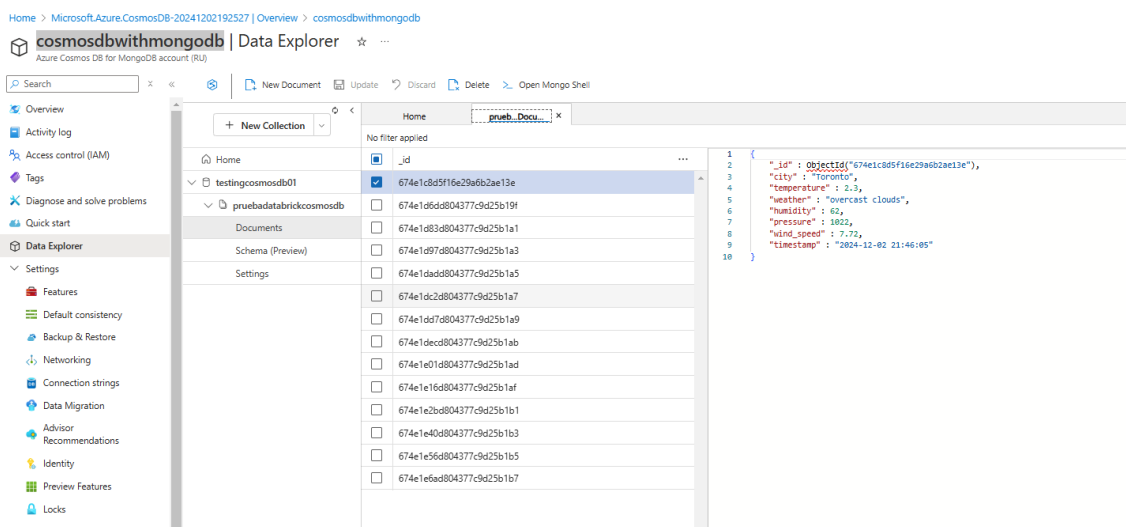
En DB_NAME cambiar por el nombre de tu base de datos.

En COLLECTION_NAME cambiar por el nombre de colección.

Guardar y ejecutar el script, debería imprimir los datos del clima obtenidos y confirmar que se han insertado en tu base de datos Cosmos DB



Cada 20 segundos está capturando datos del clima y enviado a tu Cosmos DB:



Para detener la captura de datos escribimos CTRL + C en el terminal

```
Datos insertados exitosamente en Cosmos DB
Datos obtenidos del clima:
{'city': 'Toronto', 'temperature': 2.06, 'weather': 'overcast clouds', 'humidity': 63, 'pre
Documento insertado con _id: 674e204cd804377c9d25b1e5
Datos insertados exitosamente en Cosmos DB
Traceback (most recent call last):
  File "C:\Users\juanj\OneDrive\Escritorio\weather_to_cosmosdb.py", line 60, in <module>
    time.sleep(20)
KeyboardInterrupt
PS C:\Users\juanj\OneDrive\Escritorio>
```

Crear una red virtual.

En el buscador de Azure escribimos Virtual Networks:

Virtual networks

AllServices (50)Marketplace (7)More (4)

Services

Virtual networks

Virtual networks (classic)

Virtual network gateways

Virtual Network Managers

Marketplace

Home >

Virtual networks

Tajamar

+ Create

Manage view

Refresh

Export to CSV

Open query

Assign tags

Filter for any field...Subscription equals allResource group equals allLocation equals allAdd filter

Showing 0 to 0 of 0 records.

Name

Resource group

Location

No virtual networks to display

Create a virtual network to securely connect your Azure resources to each other. Connect your virtual network to your on-premises network using an Azure VPN Gateway or ExpressRoute.

Create virtual network

Learn more

Click en Create.

Create virtual network ...

[Basics](#) [Security](#) [IP addresses](#) [Tags](#) [Review + create](#)

Azure Virtual Network (VNet) is the fundamental building block for your private network in Azure. VNet enables many types of Azure resources, such as Azure Virtual Machines (VM), to securely communicate with each other, the internet, and on-premises networks. VNet is similar to a traditional network that you'd operate in your own data center, but brings with it additional benefits of Azure's infrastructure such as scale, availability, and isolation.

[Learn more.](#) 

Project details

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

Subscription *	<input type="text" value="Azure for Students"/>
Resource group *	<input type="text" value="MondoDB"/>

[Create new](#)

Instance details

Virtual network name *	<input type="text" value="virtualnetworkdatabrickscosmosdb"/>
Region * ⓘ	<input type="text" value="(Europe) France Central"/>


[Deploy to an Azure Extended Zone](#)

[Previous](#) [Next](#) [Review + create](#)


Darle nombre a tu Virtual network, luego Next:

Create virtual network ...

Basics Security IP addresses Tags Review + create


Enhance the security of your virtual network with these additional paid security services. [Learn more](#) 

Virtual network encryption

Enable Virtual network encryption to encrypt traffic traveling within the virtual network. Virtual machines must have accelerated networking enabled. Traffic to public IP addresses is not encrypted. [Learn more](#) 


Virtual network encryption ☐

Azure Bastion

Azure Bastion is a paid service that provides secure RDP/SSH connectivity to your virtual machines over TLS. When you connect via Azure Bastion, your virtual machines do not need a public IP address. [Learn more](#) 


Enable Azure Bastion ⓘ ☐

Azure Firewall

Azure Firewall is a managed cloud-based network security service that protects your Azure Virtual Network resources. [Learn more](#) 

Enable Azure Firewall ⓘ ☐

Azure DDoS Network Protection

Azure DDoS Network Protection is a paid service that offers enhanced DDoS mitigation capabilities via adaptive tuning, attack notification, and telemetry to protect against the impacts of a DDoS attack for all protected resources within this virtual network. [Learn more](#) 

Enable Azure DDoS Network Protection ⓘ ☐

Previous

Next

Review + create

Dejar todas las casillas desactivadas, luego click en Next:

Create virtual network ...

Basics Security IP addresses Tags Review + create

Configure your virtual network address space with the IPv4 and IPv6 addresses and subnets you need. [Learn more](#)

Define the address space of your virtual network with one or more IPv4 or IPv6 address ranges. Create subnets to segment the virtual network address space into smaller ranges for use by your applications. When you deploy resources into a subnet, Azure assigns the resource an IP address from the subnet. [Learn more](#)

+ Add a subnet

10.0.0.0/16 [Delete address space](#)

10.0.0.0 /16

10.0.0.0 - 10.0.255.255 65,536 addresses

Subnets	IP address range	Size	NAT gateway
default	10.0.0.0 - 10.0.0.255	/24 (256 addresses)	- edit delete

Add IPv4 address space |

[Previous](#) [Next](#) [Review + create](#)

Dejar lo que aparece por defecto, luego Review + Create

Create virtual network ...

Basics Security IP addresses Tags Review + create

[View automation template](#)

Basics

Subscription	Azure for Students
Resource Group	MondoDB
Name	virtualnetworkdatabrickscosmosdb
Region	France Central

Security

Azure Bastion	Disabled
Azure Firewall	Disabled
Azure DDoS Network Protection	Disabled

IP addresses

Address space	10.0.0.0/16 (65,536 addresses)
Subnet	default (10.0.0.0/24) (256 addresses)

Tags



Previous

Next



Create






Create


[Home](#) >


**virtualnetworkdatabrickscosmosdb-1733174796231** | Overview  ...


Deployment


 


 Delete  Cancel  Redeploy  Download  Refresh


 Overview


 Inputs


 Outputs

 Template

 **Your deployment is complete**


 Deployment name : virtualnetworkdatabrickscosmosdb-1733174796231
Subscription : [Azure for Students](#)
Resource group : [MondoDB](#)

 Deployment details

 Next steps

[Go to resource](#)

Give feedback

 [Tell us about your experience with deployment](#)

Click en Go to resource

virtualnetworkdatabrickscosmosdb
Virtual network

Move Delete Refresh Give feedback

Overview

Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Settings
Address space
Connected devices
Subnets
Bastion
DDoS protection
Firewall
Microsoft Defender for Cloud
Network manager
DNS servers
Peerings

Essentials

Resource group (move) : MondoDB
Location (move) : France Central
Subscription (move) : Azure for Students
Subscription ID : dc6bbb60-aded-4...

Tags (edit) : Add tags

Topology
Properties
Capabilities (5)
Recc

DDoS protection

Configure additional protection from distributed denial of service attacks.

Not configured

Click en Address space

virtualnetworkdatabrickscosmosdb-173174796231 | Overview

virtualnetworkdatabrickscosmosdb

virtualnetworkdatabrickscosmosdb | Address space

Search

Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Settings
Address space
Connected devices
Subnets
Bastion

The address space for a virtual network is composed of one or more non-overlapping address ranges that are specified in CIDR notation. IP Address Management (IPAM) is recommended to simplify address management and avoid overlapping address space. When not using IPAM, it is recommended to use an address range that is not globally routable, such as 172.16.0.0/12, or a range defined in RFC 1918 or RFC 6598. [Learn more](#)

Address space	Address range	Address count
10.0.0.0/16	10.0.0.0 - 10.255.255	65,536
Add additional address range		

Peered virtual network address space

Peering name	Peered to	Address space	Address range
No results.			

Añadimos otra subred para los servicios de Databricks:

The address space for a virtual network is composed of one or more non-overlapping address ranges that are specified in CIDR IPAM, it is recommended to use an address range that is not globally routable, such as 172.16.0.0/12, or a range defined in RFC

Address space	Address range
10.0.0.0/16	10.0.0.0 - 10.0.255.255
10.179.0.0/16	10.179.0.0 - 10.179.255.255
Add additional address range	

⚠ Updating the address space of a virtual network that has peers will cause the peered virtual networks to not be able to connect to the contributor permissions on the peered virtual networks. [Learn more](#)

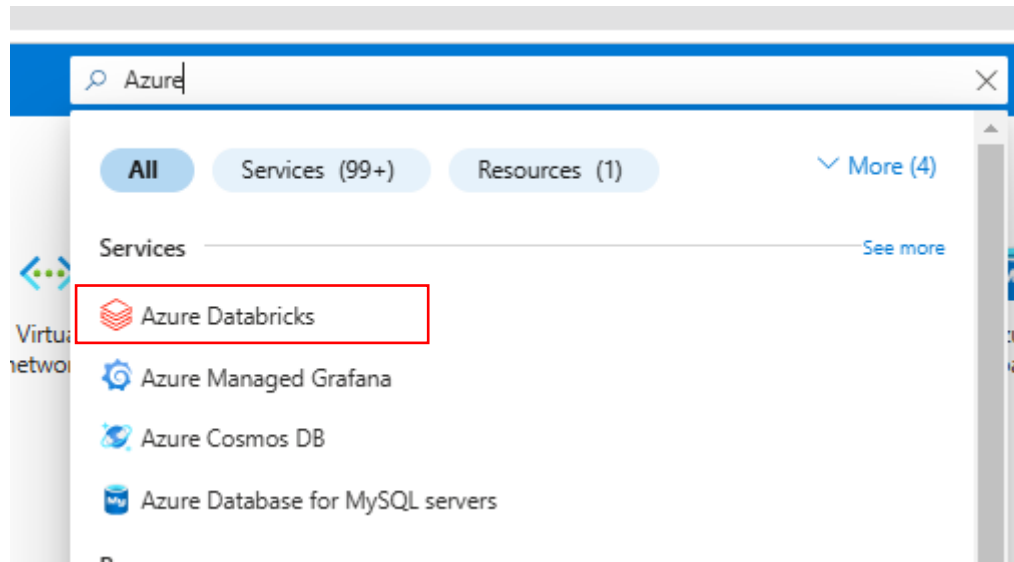
Peered virtual network address space

Peering name	Peered to	Address space
No results.		

Save Cancel

Click en Save

Crear un recursos de Azure Databricks:



En Pricing Tier seleccionar Premium para poder usar Power BI:

Create an Azure Databricks workspace ...

Basics Networking Encryption Security & compliance Tags Review + create

Project Details

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

Subscription * ⓘ	<div>Azure for Students</div>
Resource group * ⓘ	<div>MondoDB</div> <div>Create new</div>

Instance Details

Workspace name *	<div>cosmosdbdatabricks02</div>
Region *	<div>France Central</div>

Pricing Tier * ⓘ	<div>Premium (+ Role-based access controls)</div>
------------------	---

i We selected the recommended pricing tier for your workspace. You can change the tier based on your needs.

Managed Resource Group name	<div>Enter name for managed resource group</div>
-----------------------------	--

Review + create

< Previous

Next : Networking >

Click en Next: Networking

Create an Azure Databricks workspace ...

Basics Networking Encryption Security & compliance Tags Review + create

Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP) ☐ Yes ☒ No

①

Deploy Azure Databricks workspace in your own Virtual Network (VNet) ☒ Yes ☐ No

Para VNet seleccionar Yes.

Two new subnets will be created in your Virtual Network

Implicit delegation of both subnets will be done to Azure Databricks on your behalf

Public Subnet Name *	<input type="text" value="public-subnet"/>	✓
Public Subnet CIDR Range * ①	<input type="text" value="10.179.64.0/18"/>	✓
Private Subnet Name *	<input type="text" value="private-subnet"/>	✓
Private Subnet CIDR Range * ①	<input type="text" value="10.179.0.0/18"/>	✓
Allow Public Network Access ①	<input checked="" type="radio"/> Enabled <input type="radio"/> Disabled	
Required NSG Rules ①	<input type="text" value="All Rules"/>	✓

Private endpoints

Create a private endpoint to allow private connection to this resource. [Learn more](#)

Name	Subscription	Resource group
------	--------------	----------------

Click on add to create a private endpoint

+ Add

Review + create

< Previous

Next : Encryption >

Click en Next: Encryption

Create an Azure Databricks workspace ...

Basics Networking Encryption Security & compliance Tags Review + create

Data Encryption

For additional control of your data, you can add your own key to protect and control access to some types of data. Enabling customer-managed key encryption for Managed Services or Managed Disks is an irreversible action. The key, key vault, and key version may be updated but the features cannot be disabled after being enabled.

Managed Disks

Use your own key ⓘ ☐

Managed Services

Use your own key ⓘ ☐

Double encryption for DBFS root

In addition to your choice of the default encryption or your own managed key encryption, Azure Databricks DBFS root can also be encrypted with a second layer of encryption called infrastructure encryption using platform-managed key to achieve Double Encryption for DBFS root.

Enable Infrastructure Encryption ⓘ ☐

⚠ This feature cannot be changed after this workspace is created.

Review + create

< Previous

Next : Security & compliance >

No habilitar ninguna opción. Click en Next: Security...

Create an Azure Databricks workspace ...

Basics Networking Encryption Security & compliance Tags Review + create

Enhanced Security & Compliance

Enhanced Security and Compliance Add-On helps simplify the complexity of meeting security and regulatory requirements.

Enable compliance security profile ⓘ

☐

⚠ This feature cannot be disabled once it is enabled.

Enable enhanced security monitoring ⓘ

☐

Enable automatic cluster update ⓘ

☐

Review + create

< Previous

Next : Tags >

No habilitar ninguna opción, luego click en Review + Create

Create an Azure Databricks workspace ...

 Validation Succeeded

- Basics
- Networking
- Encryption
- Security & compliance
- Tags
- Review + create

Summary

Basics

Workspace name	cosmosdbdatabricks02
Subscription	Azure for Students
Resource group	MondoDB
Region	France Central
Pricing Tier	premium
Managed Resource Group name	

Networking

Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP)	No
Deploy Azure Databricks workspace in your own Virtual Network (VNet)	Yes
Virtual Network	virtualnetworkdatabrickscosmosdb
Public Subnet Name	public-subnet
Public Subnet CIDR Range	10.179.64.0/18
Private Subnet Name	private-subnet
Private Subnet CIDR Range	10.179.0.0/18
Allow Public Network Access	Enabled
Required NSG Rules	All Rules

Encryption

Enable Infrastructure Encryption	No
Enable CMK for Managed Disks	No
Enable CMK for Managed Services	No

Security & compliance

Create

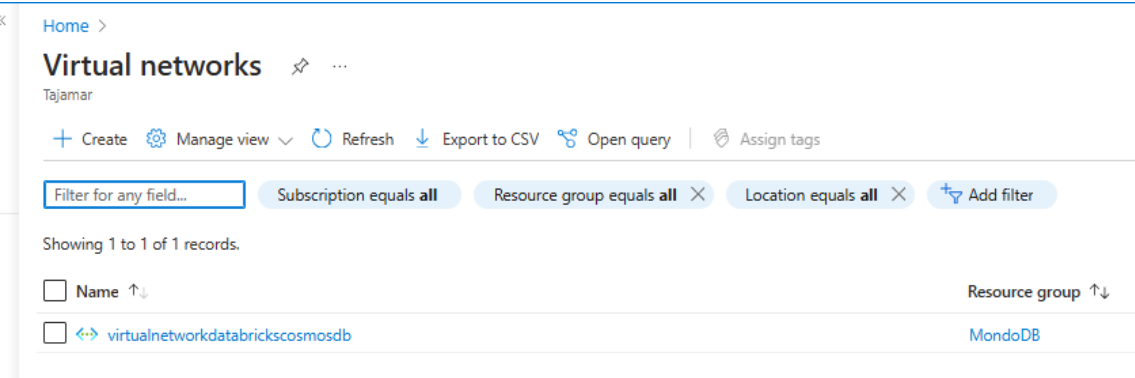
< Previous

[Download a template for automation](#)


Click en Create

Crear un Service Endpoint para Cosmos DB

Vamos a nuestra virtual network:



Luego clic en subnets:

 virtualnetworkdatabrickscosmosdb

Virtual network

Search

Move Delete Refresh Gir

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Address space

Connected devices

Subnets

Bastion

DDoS protection

Firewall

Microsoft Defender for Cloud

Network manager

DNS servers

Peerings

Service endpoints

Private endpoints

Properties

Locks

Essentials

Resource group (move) : MondoDB

Location (move) : France Central

Subscription (move) : Azure for Students

Subscription ID : dc6bbb60-aded-45e

Tags (edit) : Add tags

Topology Properties Capabilities (5)

DDoS protection

Configure additional protection from distributed denial of service attacks.

Not configured

Private endpoints

Privately access Azure services without sending traffic across internet.

Not configured

Click en public-subnet:

Search subnets			
Name ↑↓	IPv4 ↑↓	IPv6 ↑↓	Available IPs ↑↓
default	10.0.0.0/24	-	251
private-subnet	10.179.0.0/18	-	more than 10000
public-subnet	10.179.64.0/18	-	more than 10000

En Services Endpoint seleccionar Microsoft AzureCosmosDB

Service Endpoints

Create service endpoint policies to allow traffic to specific azure resources from your virtual network over service endpoints. [Learn more](#)

Services	Remove service endpoint
<div>Select a service endpoint</div>	

Service Endpoints

Create service endpoint policies to allow traffic to specific azure resources from your virtual network over service endpoints. [Learn more](#)




Services	Remove service endpoint
Microsoft.AzureCosmosDB	
<div>Select a service endpoint</div>	

Save

Cancel

Click en Save.

Vamos al servicio de Azure Databricks y damos click en Launch Workspace:

Home > cosmosdbdatabricks02   

Azure Databricks Service

Search

Delete

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Virtual Network Peerings

Encryption

Networking

Security & compliance

Properties

Locks

Monitoring

Diagnostic settings

Essentials

Status : Active

Resource group : MondoDB

Location : France Central

Subscription : Azure for Students

Subscription ID : dc6bbb60-aded-45eb-87f4-8628a5bd63a9

Tags (edit) : Add tags

See more


Managed Resource Group : databricks-rg-c

URL : https://adb-321

Pricing Tier : Premium (+ R)

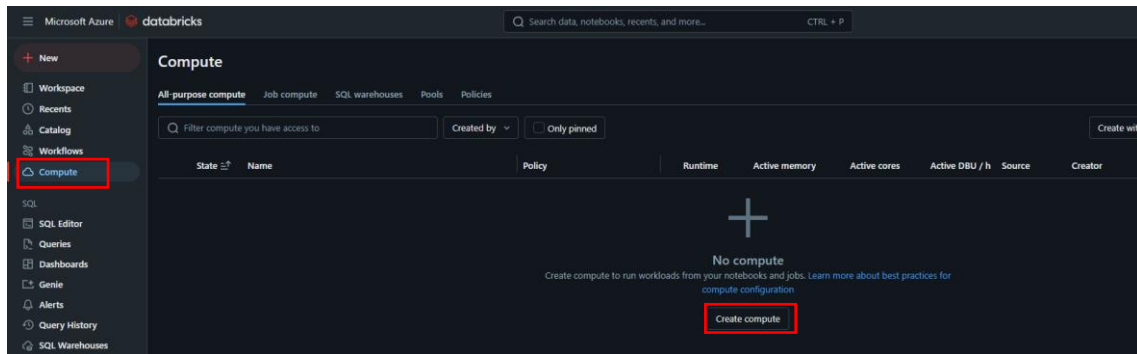
Virtual Network : virtualnetwork

Private Subnet Name : private-subnet




Launch Workspace


Creamos un cluster, click en Compute y luego Create Compute:





Importante, Single node



Juan Jose Rodriguez Balza's Cluster

Policy 


Unrestricted 


☐ Multi node ☒ **Single node**


Access mode  **Single user access** 


Single user  Juan Jose Rodriguez Balza 



Performance


Databricks runtime version 

Runtime: 15.4 LTS (Scala 2.12, Spark 3.5.0) 

☒ Use Photon Acceleration 

Node type 

Standard_DS3_v2 14 GB Memory, 4 Cores  

☒ Terminate after minutes of inactivity 

Click en Advanced options:

En Spark config escribimos:

```
spark.master local[*], 4]
spark.databricks.cluster.profile singleNode
spark.mongodb.output.uri=AQUITUCADENADECONEXIONUSADAENELSCRIPT
spark.mongodb.input.uri= AQUITUCADENADECONEXIONUSADAENELSCRIPT
```

```
# Configuración de Cosmos DB
DB_NAME = "testingcosmosdb01" # Nombre de tu base de datos
COLLECTION_NAME = "pruebadatabrickcosmosdb" # Nombre de tu colección
CONNECTION = "mongodb://cosmosdbwithmongodb:fUTd69EFQvoWifX0yn4sBTJ3e0yJoMaRAC3hUriFSZnAKlheVJwezrgr01Jw5GVnFjLWnho3YkRcACDbPhcU3g==@cosmosdbwith
```

▼ Advanced options

Azure Data Lake Storage credential passthrough ⓘ

☐ Enable credential passthrough for user-level data access

Spark Logging Init Scripts SSH

Spark config ⓘ

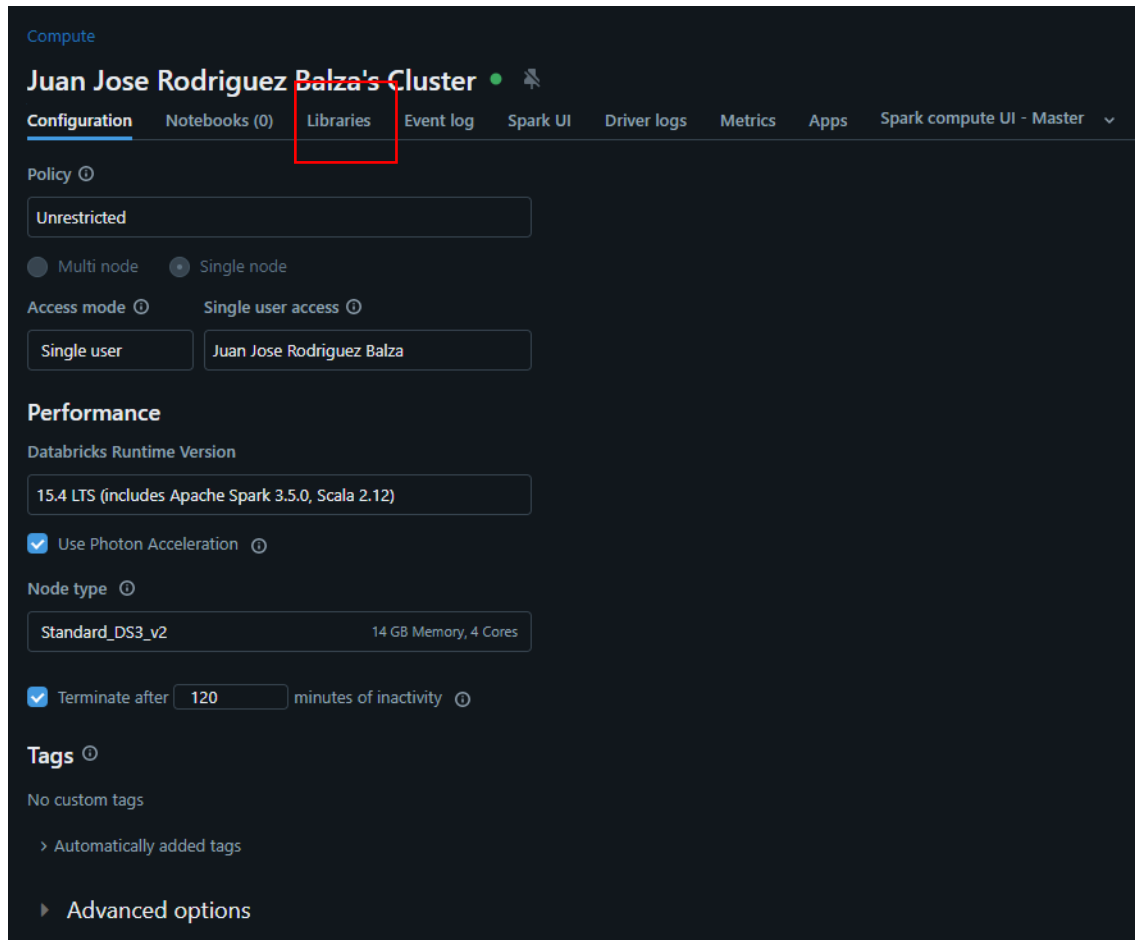
```
spark.master local[*], 4]
spark.databricks.cluster.profile singleNode
spark.mongodb.output.uri=mongodb://cosmosdbwithmongodb:fUTd69EFQvoWifX0yn4sBTJ3e0y
JoMaRAC3hUriFSZnAKlheVJwezrgr01Jw5GVnFjLWnho3YkRcACDbPhcU3g%40cosmosdbwithmon
godb.mongo.cosmos.azure.com:10255/?
```

Environment variables ⓘ

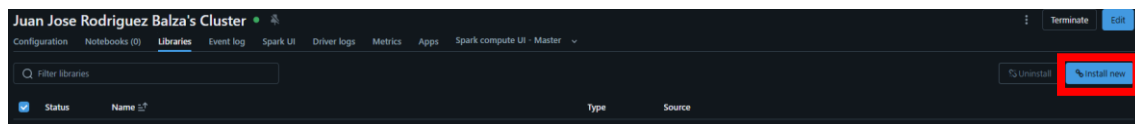
```
PYSPARK_PYTHON=/databricks/python3/bin/python3
```

Create compute Cancel

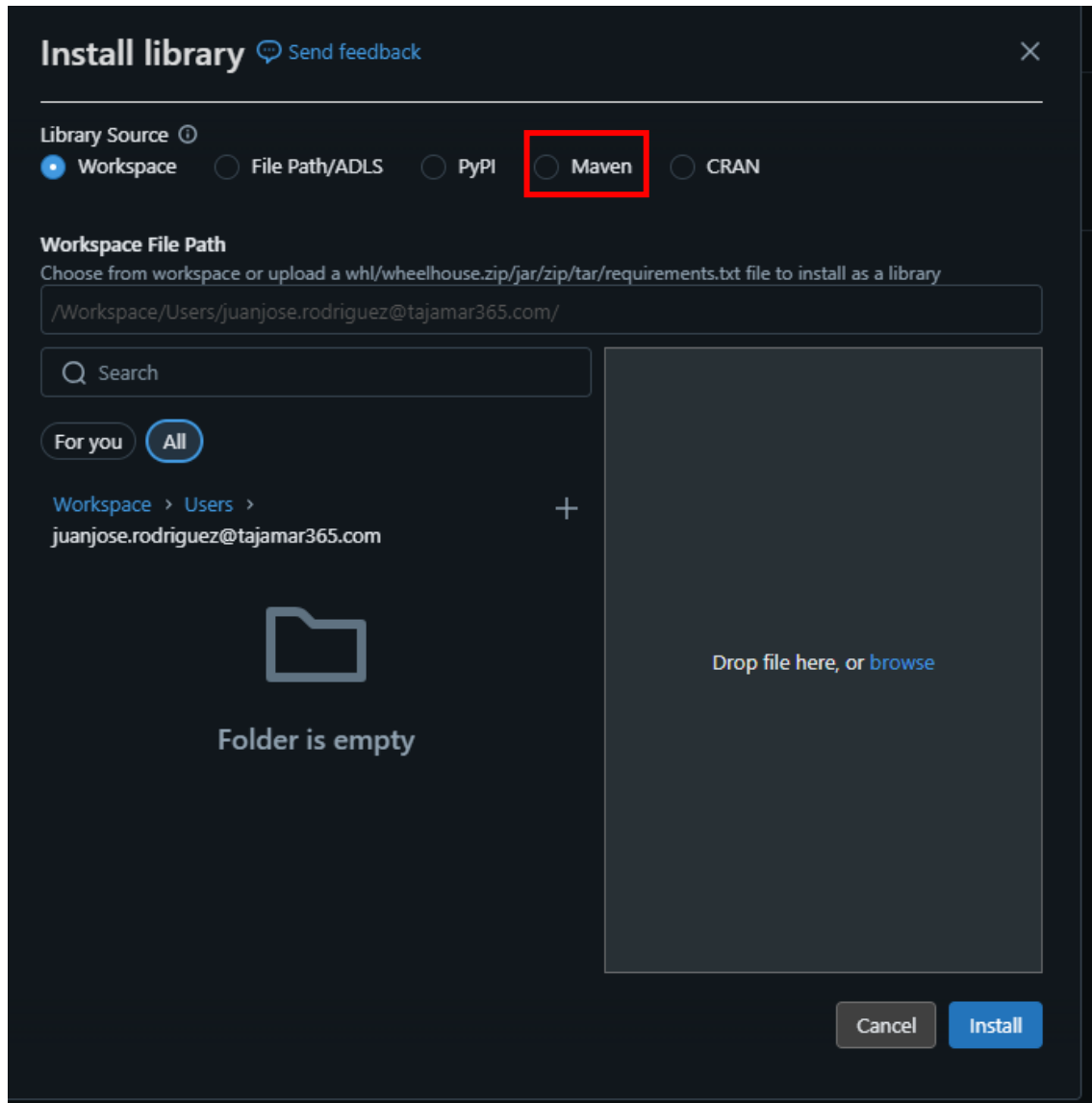
Una vez creado el cluster damos click en Libraries:



Click en install new:



Click en Maven:



Click en Search Packages:

Install library [Send feedback](#)

Library Source ⓘ

☐ Workspace ☐ File Path/ADLS ☐ PyPI ☒ Maven ☐ CRAN

Coordinates

Maven Coordinates (com.databricks:spark-csv_2.10:1.0.0)

Search Packages

Repository ⓘ

Optional

Exclusions

Dependencies to exclude (log4j:log4j; junit:junit)

Cancel

Install

Spark Packages, escribir mongo y seleccionar el conector de spark oficial:

Search packages [Send feedback](#)

Spark Packages

mongo

Name	Organization	Description	Releases	Options
spark-mongodb	Stratio	MongoDB data source for Spark SQL	0.12.0	Select
mongo-spark	mongodb	The official MongoDB Spark Connector	3.0.1	Select

Click en install

Install library [Send feedback](#)

Library Source ⓘ

☐ Workspace ☐ File Path/ADLS ☐ PyPI ☒ Maven ☐ CRAN

Coordinates

org.mongodb.spark:mongo-spark-connector_2.12:3.0.1

Search Packages

Repository ⓘ

Optional

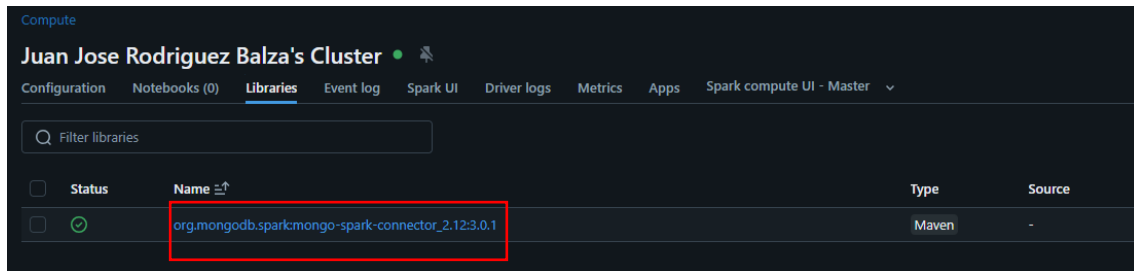
Exclusions

Dependencies to exclude (log4j:log4j; junit:junit)

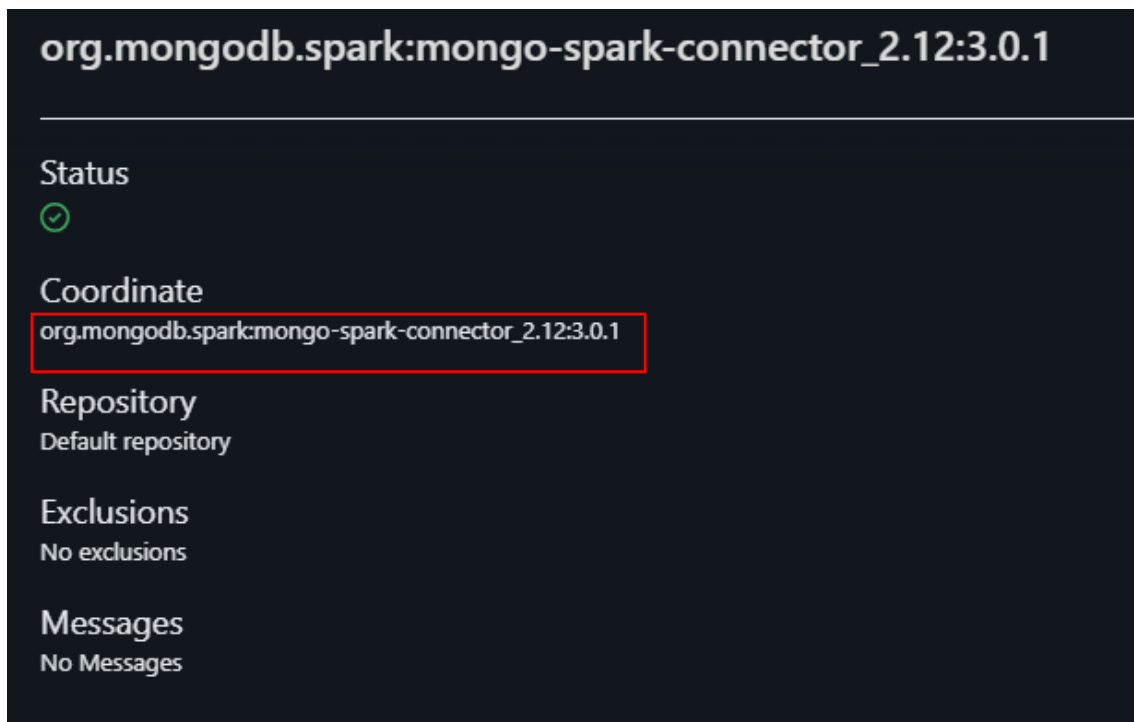
Cancel

Install

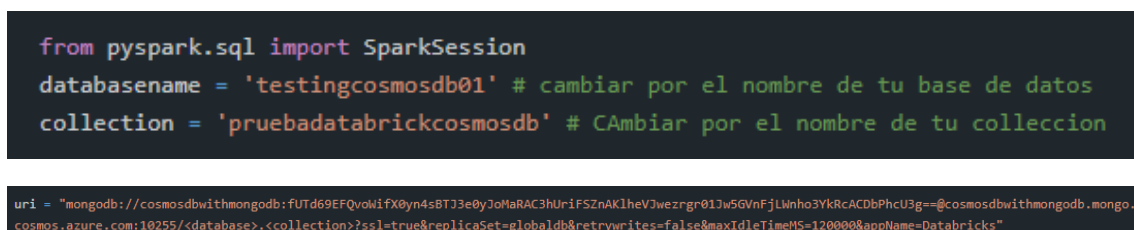
Una vez instalado damos click en el conector:



Copiamos el Coordinate:



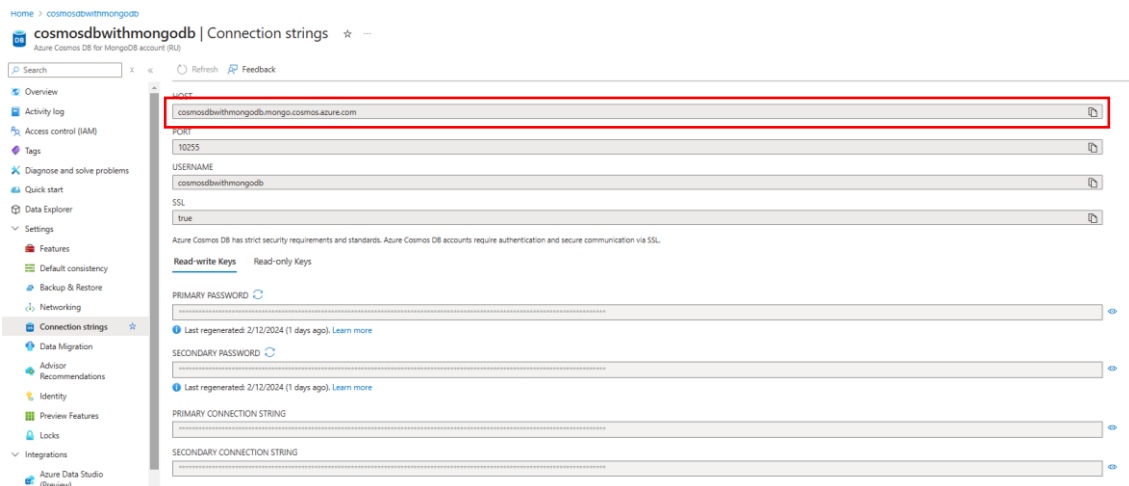
Abrimos un [notebook](#) y escribimos:



Cambia la uri del [notebook](#) por tu propia URI. Esta tiene una estructura como:

mongodb://<username>:<password>@**cosmosdbwithmongodb**.mongo.cosmos.azure.com:10255

Lo que va después del @ (que también es el USERNAME) lo encuentras haciendo click en connection string en tu base de datos de Cosmos DB:



Mientras que:

Usuario y contraseña serían:

- **USERNAME:** cosmosdbwithmongodb
- **PASSWORD:** Debes usar el valor del **Primary Password** visible en la misma sección.

Luego en otra celda escribes:

```
spark = SparkSession.builder.appName("TestingApp").config('spark.jars.packages', 'org.mongodb.spark:mongo-spark-connector_2.12:3.0.1').getOrCreate()
```

Pegar tu 'Coordinate'

Ahora creamos un dataframe:

```
df = spark.read.format("com.mongodb.spark.sql.DefaultSource") \
    .option("uri", uri) \
    .option("database", databasename) \
    .option("collection", collection) \
    .load()

(1) Spark Jobs

df: pyspark.sql.dataframe.DataFrame = [_id: struct, city: string ... 6 more fields]
```


Luego `df.show()`

df.show()

▶ (1) Spark Jobs

	_id	city	humidity	pressure	temperature	timestamp	weather	wind_speed
	{674e1c8d5f16e29a...	Toronto	62	1022	2.3	2024-12-02 21:46:05	overcast clouds	7.72
	{674e1d6dd804377c...	Toronto	62	1022	2.3	2024-12-02 21:49:49	overcast clouds	7.72
	{674e1d83d804377c...	Toronto	62	1022	2.3	2024-12-02 21:50:11	overcast clouds	7.72
	{674e1d97d804377c...	Toronto	62	1022	2.19	2024-12-02 21:50:31	overcast clouds	7.72
	{674e1dadd804377c...	Toronto	62	1022	2.19	2024-12-02 21:50:53	overcast clouds	7.72
	{674e1dc2d804377c...	Toronto	62	1022	2.19	2024-12-02 21:51:14	overcast clouds	7.72
	{674e1dd7d804377c...	Toronto	62	1022	2.19	2024-12-02 21:51:35	overcast clouds	7.72
	{674e1dec804377c...	Toronto	62	1022	2.19	2024-12-02 21:51:56	overcast clouds	7.72
	{674e1e01d804377c...	Toronto	62	1022	2.19	2024-12-02 21:52:17	overcast clouds	7.72
	{674e1e16d804377c...	Toronto	62	1022	2.19	2024-12-02 21:52:38	overcast clouds	7.72
	{674e1e2bd804377c...	Toronto	62	1022	2.19	2024-12-02 21:52:59	overcast clouds	7.72
	{674e1e40d804377c...	Toronto	62	1022	2.19	2024-12-02 21:53:20	overcast clouds	7.72
	{674e1e56d804377c...	Toronto	62	1022	2.19	2024-12-02 21:53:42	overcast clouds	7.72
	{674e1e6ad804377c...	Toronto	62	1022	2.3	2024-12-02 21:54:02	overcast clouds	7.72
	{674e1e7fd804377c...	Toronto	62	1022	2.3	2024-12-02 21:54:23	overcast clouds	7.72
	{674e1e94d804377c...	Toronto	62	1022	2.3	2024-12-02 21:54:44	overcast clouds	7.72
	{674e1ea9d804377c...	Toronto	62	1022	2.3	2024-12-02 21:55:05	overcast clouds	7.72
	{674e1ebed804377c...	Toronto	62	1022	2.19	2024-12-02 21:55:26	overcast clouds	7.72

Vamos a usar un Partner Connect de Databricks para poder utilizar Power BI:

En Databricks, vamos ahora a Partner Connect. Para ello vamos a almacenar el dataframe de arriba en un 'Hive Metastore':

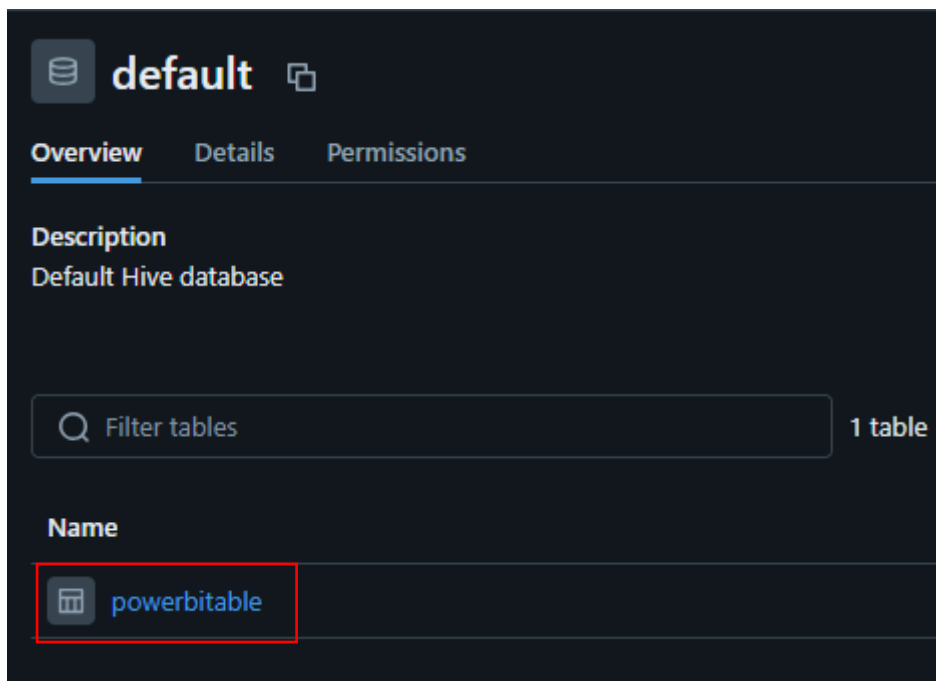
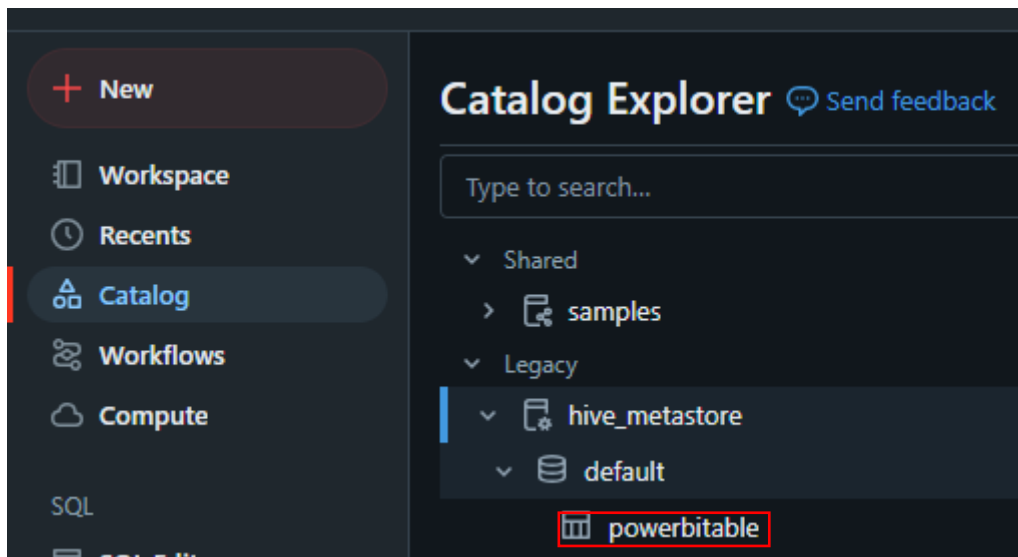
▶ Just now (24s)

```
df.write.saveAsTable("PowerBITable")
```

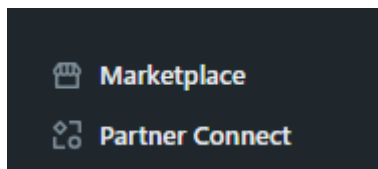
▼ (4) Spark Jobs

- ▶ Job 2 [View](#) (Stages: 1/1)
- ▶ Job 3 [View](#) (Stages: 1/1)
- ▶ Job 4 [View](#) (Stages: 1/1)
- ▶ Job 5 [View](#) (Stages: 1/1, 1 skipped)

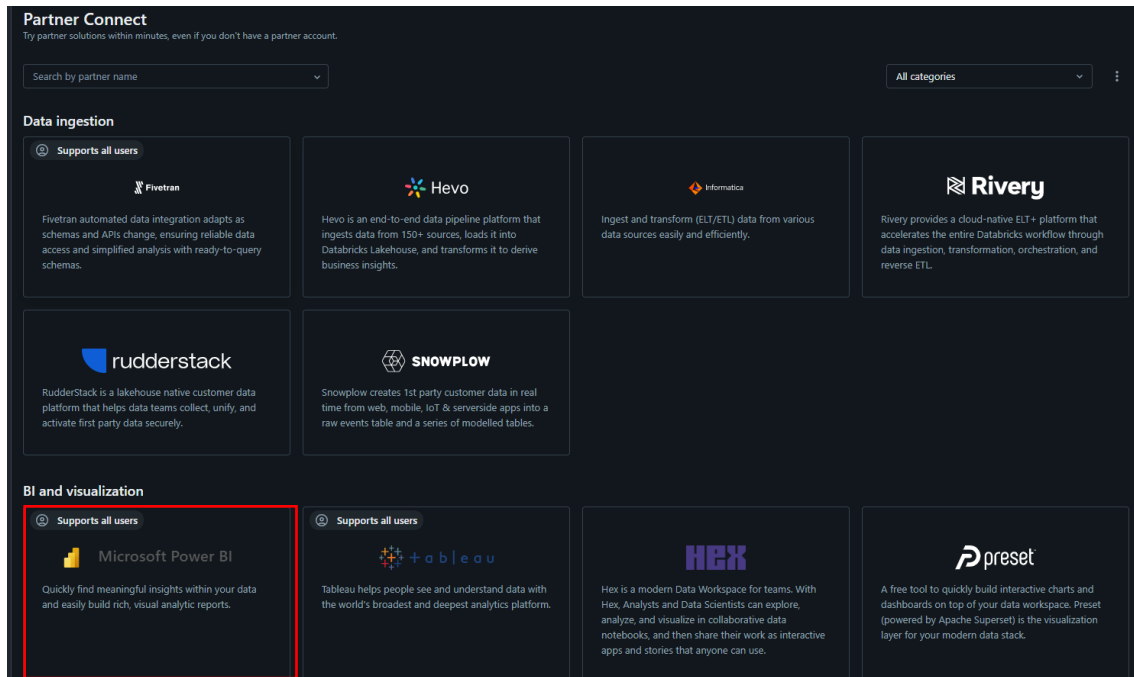
Si damos click en Catalog encontramos la tabla:



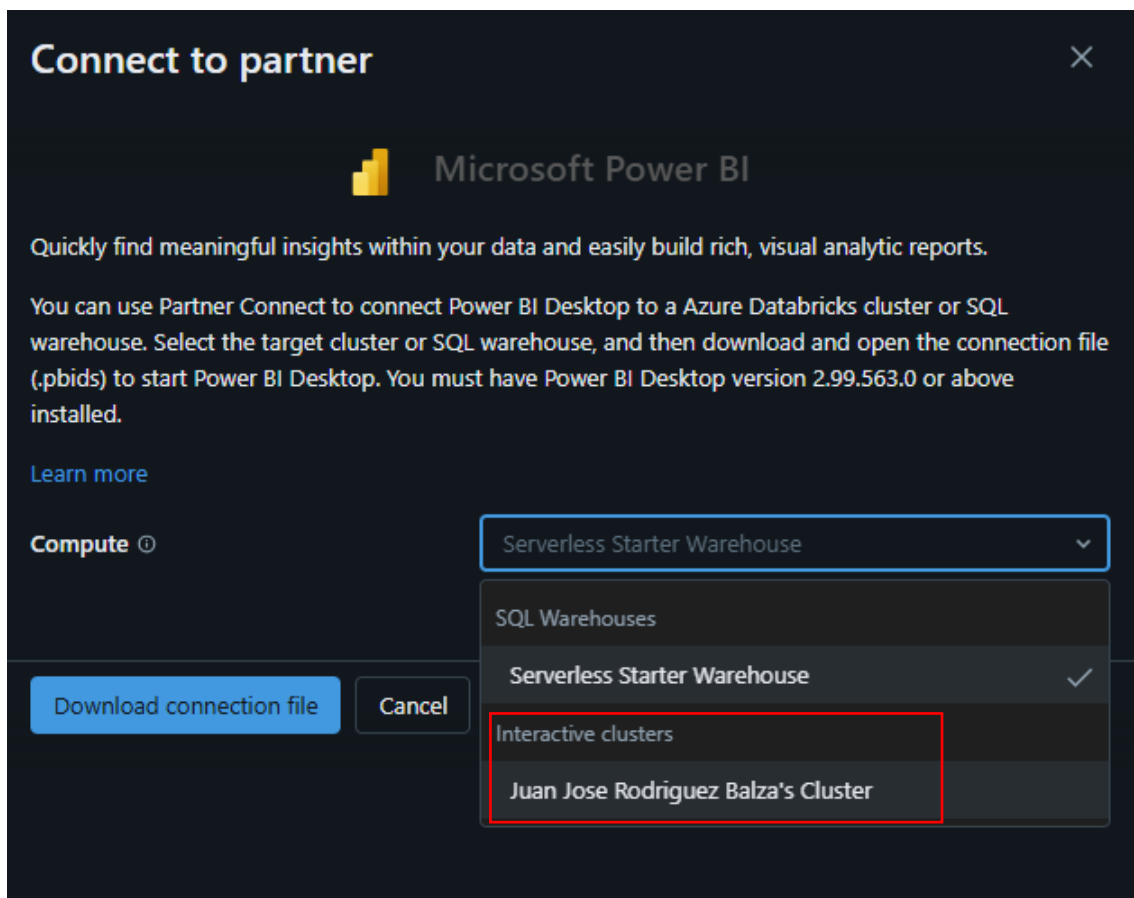
Ahora nos vamos a Partner Connect:



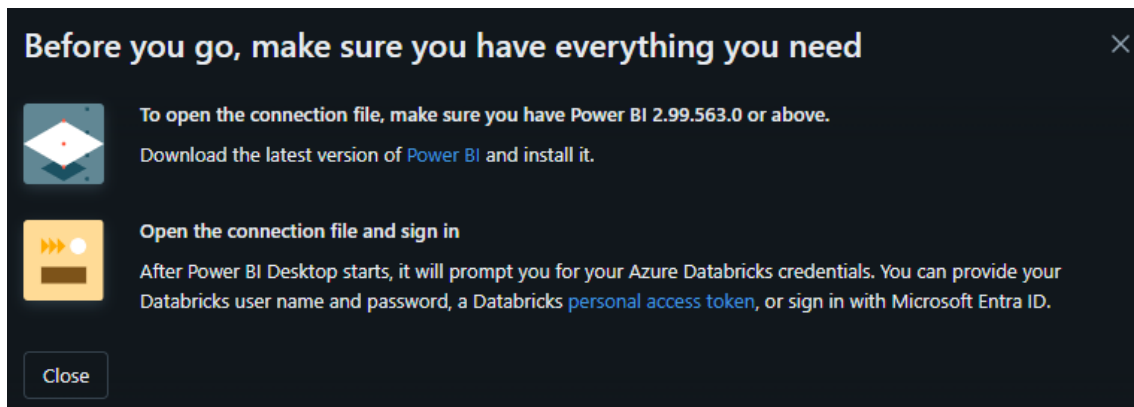
Seleccionamos Power BI:



Seleccionamos el cluster:

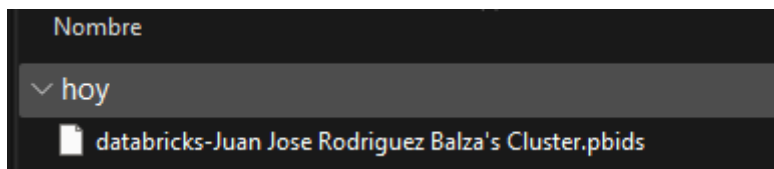


Click en Download connection file



Verifica que tienes Power BI instalado en el equipo con la versión que se especifica arriba.

Una vez que has verificado la instalación de Power BI ve a tu carpeta de descargas y buscas el 'archivo de conexión' (connection file) que descargaste desde el Partner connect.





El archivo tiene extensión .pbids.


Abrir el archivo con Power BI Desktop. Power BI te pedirá clave y contraseña, selecciona Azure Active Directory e inicias sesión con el correo que tienes registrado en Azure, luego click en Conectar:


Navegador


□ ×


Opciones de presentación ▾ 

 adb-3214202235140714.14.azuredatabricks.ne...

 hive_metastore [1]

 default [1]

☐  powerbitable

 samples [3]

No hay elementos seleccionados para vista previa

Cargar

Transformar datos

Cancelar

Seleccionamos powerbitable, luego Cargar:

Navegador

Opciones de presentación ▾

adb-3214202235140714.14.azuredatabricks.ne...

hive_metastore [1]

default [1]

☒ powerbitable

samples [3]

powerbitable

_id	city	humidity	pressure	tempera
{"oid":"674e1c8d5f16e29a6b2ae13e"}	Toronto	62	1022	
{"oid":"674e1d6dd804377c9d25b19f"}	Toronto	62	1022	
{"oid":"674e1d83d804377c9d25b1a1"}	Toronto	62	1022	
{"oid":"674e1d97d804377c9d25b1a3"}	Toronto	62	1022	
{"oid":"674e1dadd804377c9d25b1a5"}	Toronto	62	1022	
{"oid":"674e1dc2d804377c9d25b1a7"}	Toronto	62	1022	
{"oid":"674e1dd7d804377c9d25b1a9"}	Toronto	62	1022	
{"oid":"674e1decd804377c9d25b1ab"}	Toronto	62	1022	
{"oid":"674e1e01d804377c9d25b1ad"}	Toronto	62	1022	
{"oid":"674e1e16d804377c9d25b1af"}	Toronto	62	1022	
{"oid":"674e1e2bd804377c9d25b1b1"}	Toronto	62	1022	
{"oid":"674e1e40d804377c9d25b1b3"}	Toronto	62	1022	
{"oid":"674e1e56d804377c9d25b1b5"}	Toronto	62	1022	
{"oid":"674e1e6ad804377c9d25b1b7"}	Toronto	62	1022	
{"oid":"674e1e7fd804377c9d25b1b9"}	Toronto	62	1022	
{"oid":"674e1e94d804377c9d25b1bb"}	Toronto	62	1022	
{"oid":"674e1ea9d804377c9d25b1bd"}	Toronto	62	1022	
{"oid":"674e1ebed804377c9d25b1bf"}	Toronto	62	1022	
{"oid":"674e1ed3d804377c9d25b1c1"}	Toronto	62	1022	
{"oid":"674e1ee8d804377c9d25b1c3"}	Toronto	62	1022	
{"oid":"674e1efdd804377c9d25b1c5"}	Toronto	62	1022	
{"oid":"674e1f12d804377c9d25b1c7"}	Toronto	62	1022	
{"oid":"674e1f27d804377c9d25b1c9"}	Toronto	62	1022	

Cargar




Transformar datos












































Cancelar

En el panel principal ya se puede ver la tabla:

Visualizaciones

Compilar visual





Datos

Buscar

powerbitable

☐ _id

☒ city

☐ Σ humidity

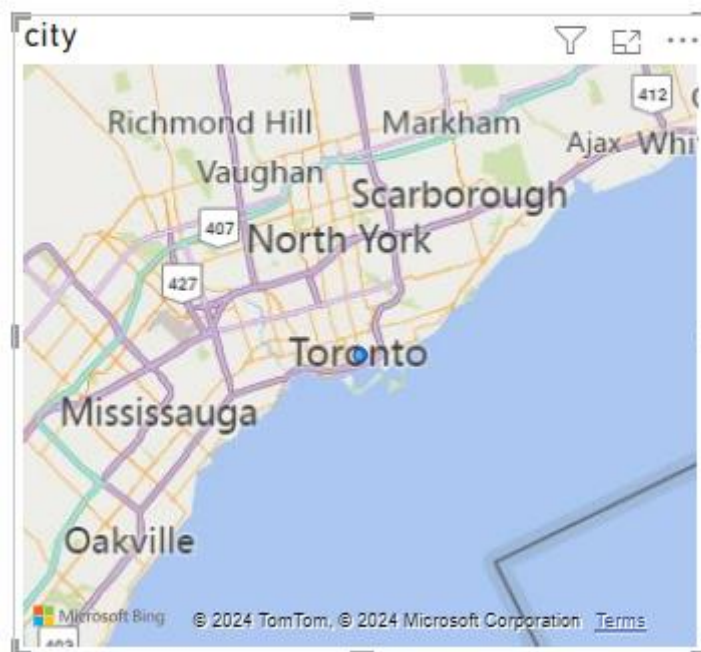
☐ Σ pressure

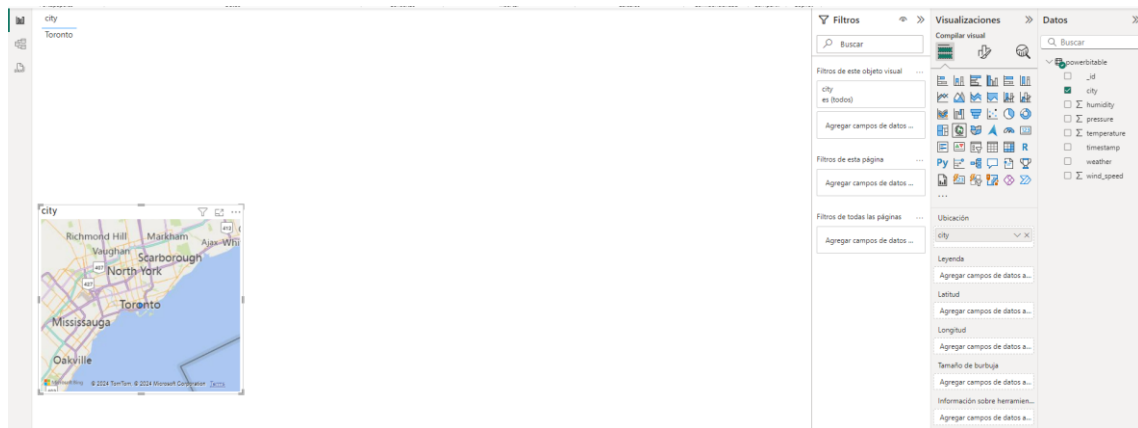
☐ Σ temperature

☐ timestamp

☐ weather

☐ Σ wind_speed





Dedica un buen tiempo ha interaccionar con Power BI haciendo varios plots.

Una vez finalizado eliminar todos los recursos que has creado para evitar cargos adicionales.