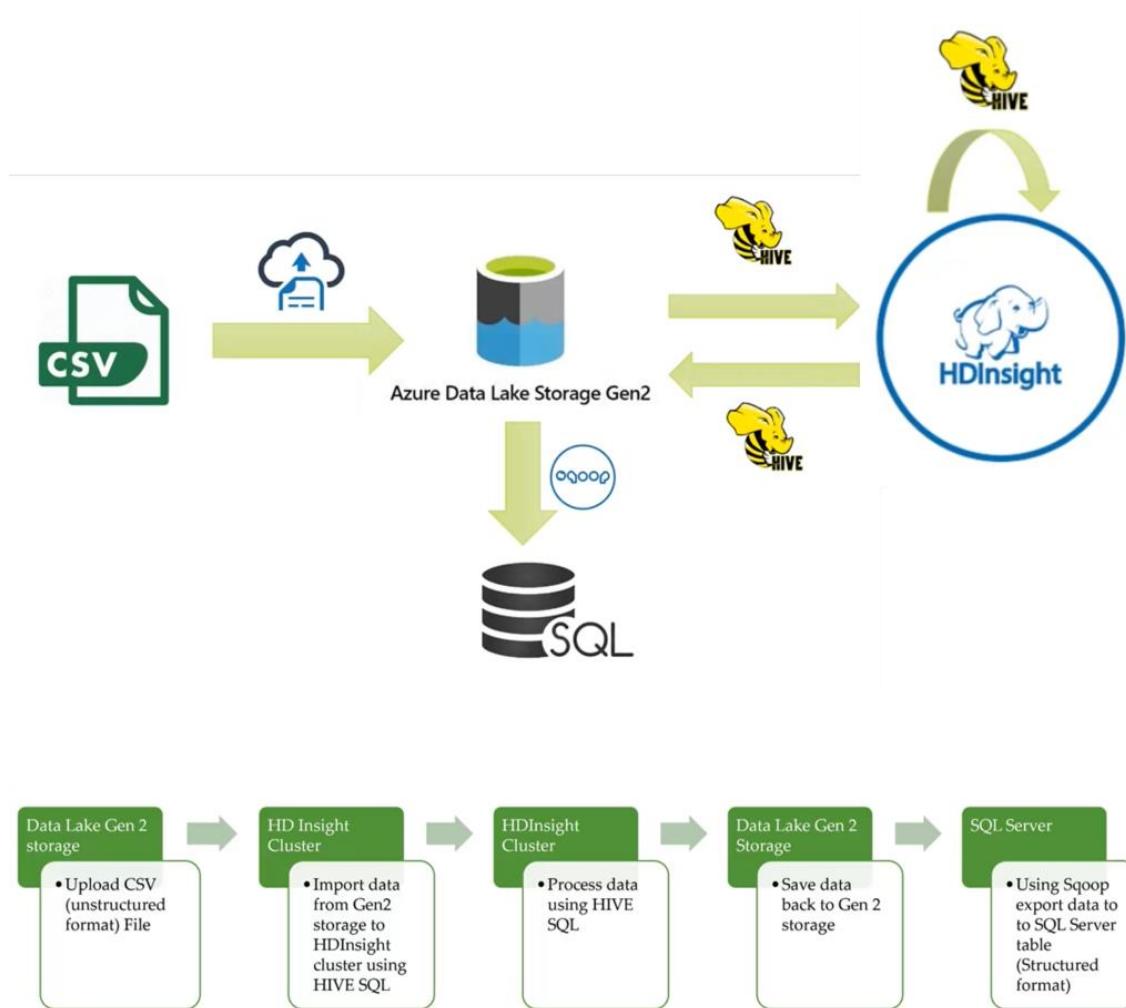


Practica 9. Solucion ETL para batch processing con Hadoop en ambiente Azure con los servicios **HDInsight + Azure Data Lake Storage Gen 2 + HIVE+ SQL Database + Sqoop**



Crear un grupo de recursos:

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 * ⓘ

Practica09

Resource details

Region * ⓘ

(Europe) Italy North

[Review + create](#)

[< Previous](#)

[Next : Tags >](#)

Create a resource group

...

✓ Validation passed.

Basics Tags Review + create

Basics

Subscription	Azure for Students
Resource group	Practica09
Region	Italy North

Tags

None

[Create](#)

[< Previous](#)

[Next >](#)

[Download a](#)

Click en Create.

Crear Source: Azure Data Lake Storage Gen2

Empezamos por la Storage Account

The screenshot shows the Azure services dashboard. On the left, there's a sidebar with 'Create a resource', 'Resource groups', 'Management groups', 'Education', and 'Storage accounts' (which is selected and highlighted). Below that is a 'Resources' section with 'Recent' and 'Favorite' tabs, listing resources like 'Practica09', 'vault-m06yk4lo', 'desarrollo', and 'Azure for Students'. On the right, the 'Storage accounts' blade is open, showing a summary card with a storage icon, the title 'Storage accounts', a 'Create' button, and a 'View' button. Below the card, there's a 'Description' section with a link to 'See more', followed by 'Free training from Microsoft' with links to 'Create an Azure Storage account', 'Connect an app to Azure Storage', and 'Secure your Azure Storage account'. A 'Last Viewed' sidebar on the right lists items viewed recently.

Página principal

The screenshot shows the 'Storage accounts' blade in the Azure portal. At the top, there's a toolbar with 'Create', 'Restore', 'Manage view', 'Refresh', 'Export to CSV', 'Open query', 'Assign tags', and 'Delete'. Below that is a search bar and a filter bar with options like 'Subscription equals all', 'Resource group equals all', and 'Location equals all'. A message says 'Showing 0 to 0 of 0 records.' The main area has columns for 'Name', 'Type', 'Kind', and 'Resource group'. At the bottom, there's a large icon of a storage account, a message saying 'No storage accounts to display', and a 'Create storage account' button.

Click en Create.

Home > Storage accounts >

Create a storage account

Basics

Advanced

Networking

Data protection

Encryption

Tags

Review + create

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 *

Azure for Students

Resource group *

Practica09

[Create new](#)

Instance details

Storage account name * ⓘ

Region * ⓘ

(Europe) Italy North

[Deploy to an Azure Extended Zone](#)

Primary service ⓘ

Select a primary service

Performance * ⓘ

Standard: Recommended for most scenarios (general-purpose v2 account)

Premium: Recommended for scenarios that require low latency.

Redundancy * ⓘ

Locally-redundant storage (LRS)

[Previous](#)

[Next](#)

[Review + create](#)

En Basics, debe estar lleno con lo siguiente:

Create a storage account

...

Basics

Advanced

Networking

Data protection

Encryption

Tags

Review + create

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 *

Azure for Students

Resource group *

Practica09

[Create new](#)

Instance details

Storage account name * ⓘ

gen2storagepractica19

Region * ⓘ

(Europe) Italy North

[Deploy to an Azure Extended Zone](#)

Primary service ⓘ

Azure Blob Storage or Azure Data Lake Storage Gen 2

Performance * ⓘ

Standard: Recommended for most scenarios (general-purpose v2 account)

Premium: Recommended for scenarios that require low latency.

Redundancy * ⓘ

Locally-redundant storage (LRS)

[Previous](#)

[Next](#)

[Review + create](#)

Click en Next

En Advanced lo llenamos con lo siguiente:

Create a storage account

Basics **Advanced** Networking Data protection Encryption Tags Review + create

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 ⓘ

i 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
- Cold: Optimized for rarely accessed data and backup scenarios

Azure Files

Enable large file shares ⓘ

[Previous](#)

[Next](#)

Review + create

Click en Next

En Networking, se llena así:

Home > Storage accounts >

Create a storage account ...

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](#)

Private endpoint

Create a private endpoint to allow a private connection to this resource. Additional private endpoint connections can be created within the storage account or private link center.

 Add private endpoint

Name	Subscription	Resource g...	Region	Target sub...	Subnet	Private DN...
<i>Click on add to create a private endpoint</i>						

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 * (i)

- Microsoft network routing
- Internet routing

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

Click en Next

Data Protection

Home > Storage accounts >

Create a storage account ...

Basics Advanced Networking **Data protection** Encryption Tags Review + create

Recovery

Protect your data from accidental or erroneous deletion or modification.

- Enable point-in-time restore for containers

Use point-in-time restore to restore one or more containers to an earlier state. If point-in-time restore is enabled, then versioning, change feed, and blob soft delete must also be enabled. [Learn more](#)

- Enable soft delete for blobs

Soft delete enables you to recover blobs and directories that were previously marked for deletion. [Learn more](#)

Days to retain deleted blobs ⓘ

7

- Enable soft delete for containers

Soft delete enables you to recover containers that were previously marked for deletion. [Learn more](#)

Days to retain deleted containers ⓘ

7

- Enable soft delete for file shares

Soft delete enables you to recover file shares that were previously marked for deletion. [Learn more](#)

Days to retain deleted file shares ⓘ

7

Tracking

Manage versions and keep track of changes made to your blob data.

- Enable versioning for blobs

Use versioning to automatically maintain previous versions of your blobs. [Learn more](#)

Consider your workloads, their impact on the number of versions created, and the resulting costs. Optimize costs by automatically managing the data lifecycle. [Learn more](#)

- Enable blob change feed

Keep track of create, modification, and delete changes to blobs in your account. [Learn more](#)

Access control

- Enable version-level immutability support

Allows you to set time-based retention policy on the account-level that will apply to all blob versions. Enable this feature to set a default policy at the account level. Without enabling this, you can still set a default policy at the container level or set policies for specific blob versions. Versioning is required for this property to be enabled. [Learn more](#)

Previous

Next

Review + create

Click en Next

Encryption

Home > Storage accounts >

Create a storage account ...

Basics Advanced Networking Data protection **Encryption** Tags Review + create

Encryption type * ⓘ Microsoft-managed keys (MMK) Customer-managed keys (CMK)

Enable support for customer-managed keys ⓘ Blobs and files only All service types (blobs, files, tables, and queues)
⚠ This option cannot be changed after this storage account is created.

Enable infrastructure encryption ⓘ

Click en NExt

Tags, lo podemos dejar en blanco

Home > Storage accounts >

Create a storage account ...

Basics Advanced Networking Data protection Encryption **Tags** Review + create

Name	Value	Resource
<input type="text"/> :	<input type="text"/>	All resources selected

Click en Next

Review + Create: click en Create

Home > **gen2storagepractica19_1731266818027 | Overview** ⚡ ...

Deployment

Search Delete Cancel Redeploy Download Refresh

Overview

Your deployment is complete

Deployment name: gen2storagepractica19_1731266818027
Subscription: Azure for Students
Resource group: Practica09

Start time: 11/10/2024, 8:27:10 PM
Correlation ID: c18c1d1e-7aa6-4495-80a1-b2cb28fe006f

Deployment details
Next steps

Go to resource

Give feedback
Tell us about your experience with deployment

Vamos a crear el servicio de SQL Server (Data Destination):

The screenshot shows the Azure portal's search interface. The search bar at the top contains the text 'SQL'. Below the search bar, there are three tabs: 'All', 'Services (32)', and 'Marketplace (31)'. The 'Services' tab is selected. Under the 'Services' tab, the 'SQL databases' item is highlighted with a red box. Other items listed include 'SQL servers', 'SQL elastic pools', and 'SQL managed instances'. To the right of the search results, there are sections for 'Marketplace' and 'Last Viewed' resources.

The screenshot shows the 'SQL databases' blade in the Azure portal. At the top, there is a navigation bar with 'Microsoft Azure' and a search bar. Below the navigation bar, there are buttons for 'Create', 'Reservations', 'Manage view', 'Refresh', 'Export to CSV', 'Open query', 'Assign tags', and 'Delete'. There are also filter buttons for 'Subscription equals all', 'Resource group equals all', and 'Location equals all'. A message 'Showing 0 to 0 of 0 records.' is displayed. At the bottom, there is a large 'SQL' logo and a message 'No SQL databases to display'. Below the message, there is a sub-message: 'Utilize a fully managed relational database service, perfect for accelerating application development and simplifying management tasks.' and a 'Create SQL database' button.

Click en Create.

Basic:

Create SQL Database

Microsoft

Basics Networking Security Additional settings Tags Review + create

Create a SQL database with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize. [Learn more](#)

Want to try Azure SQL Database for free? Create a free serverless database with the first 100,000 vCore seconds, 32GB of data, and 32GB of backup storage free per month for the lifetime of the subscription. [Learn more](#)

[Apply offer \(Preview\)](#)

SQL Database Hyperscale: Low price, high scalability, and best feature set. [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 * ⓘ

Practica09



[Create new](#)

Database details

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name *

dbpractica09



Server * ⓘ

Select a server



[Create new](#)

En Server, dar click en Create New:

Create SQL Database Server

Microsoft

Enter required settings for this server, including providing a name and location. This server will be created in the same subscription and resource group as your database.

Server name * dbserverpractica09 .database.windows.net

Location * (Europe) Italy North

Authentication

i Azure Active Directory (Azure AD) is now Microsoft Entra ID. [Learn more](#)

Select your preferred authentication methods for accessing this server. Create a server admin login and password to access your server with SQL authentication, select only Microsoft Entra authentication [Learn more](#) using an existing Microsoft Entra user, group, or application as Microsoft Entra admin [Learn more](#), or select both SQL and Microsoft Entra authentication.

Authentication method Use Microsoft Entra-only authentication Use both SQL and Microsoft Entra authentication Use SQL authentication

Set Microsoft Entra admin * Not Selected Set admin

Server admin login * admin01

Password * *****

Confirm password * ***** Password and confirm password must match.

OK

Click en Set admin, usa tu email de azure for students:

Microsoft Entra ID

Microsoft Entra ID

X

i Try changing or adding filters if you don't see what you're looking for.

Search

3 results found

All Users Groups Enterprise applications

Selected item (1)

i Reset

 Juan Jose Rodriguez Balza juanjose.rodriguez@tajamar365.com i d

Name Type Details

<input checked="" type="checkbox"/>	 Juan Jose Rodriguez Balza	User	juanjose.rodriguez@tajamar365.com
-------------------------------------	---	------	-----------------------------------

Select

Click en Select.

Debería verse algo similar a esto:

Create SQL Database Server

Microsoft

Enter required settings for this server, including providing a name and location. This server will be created in the same subscription and resource group as your database.

Server name *	dbserverpractica09	.database.windows.net
Location *	(Europe) Italy North	
Authentication		
<p>i Azure Active Directory (Azure AD) is now Microsoft Entra ID. Learn more</p> <p>Select your preferred authentication methods for accessing this server. Create a server admin login and password to access your server with SQL authentication, select only Microsoft Entra authentication Learn more using an existing Microsoft Entra user, group, or application as Microsoft Entra admin Learn more, or select both SQL and Microsoft Entra authentication.</p>		
Authentication method	<input type="radio"/> Use Microsoft Entra-only authentication <input checked="" type="radio"/> Use both SQL and Microsoft Entra authentication <input type="radio"/> Use SQL authentication	
Set Microsoft Entra admin	juanjose.rodriguez@tajamar365.com Admin Object/App ID: 7548ddbf-c124-41b9-b62a-3e3e0a7c2f45 Set admin	
Server admin login *	admin01	
Password *	*****	
Confirm password *	*****	

Click en Ok.

Create SQL Database

Microsoft

Basics Networking Security Additional settings Tags Review + create

Create a SQL database with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize. [Learn more](#)

Want to try Azure SQL Database for free? Create a free serverless database with the first 100,000 vCore seconds, 32GB of data, and 32GB of backup storage free per month for the lifetime of the subscription. [Learn more](#)

[Apply offer \(Preview\)](#)

SQL Database Hyperscale: Low price, high scalability, and best feature set. [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.

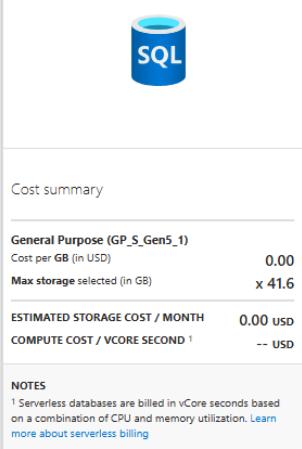
Subscription * [?](#)

Azure for Students

Resource group * [?](#)

Practica09

[Create new](#)



Database details

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name *

dbpractica09



Server * [?](#)

(new) dbserverpractica09 (Italy North)



[Create new](#)

Want to use SQL elastic pool? [?](#)

Yes No

Workload environment

Development

Production

Default settings provided for Development workloads. Configurations can be modified as needed.

Compute + storage * [?](#)

General Purpose - Serverless

Standard-series (Gen5), 1 vCore, 32 GB storage, zone redundant disabled

[Configure database](#)

Backup storage redundancy

Choose how your PITR and LTR backups are replicated. Geo restore or ability to recover from regional outage is only available when geo-redundant storage is selected.

Backup storage redundancy [?](#)

Locally-redundant backup storage

Zone-redundant backup storage

Click en Configure database:

Home > SQL databases >

Create SQL Database

Microsoft

Basics Networking Security Additional settings Tags Review + create

Create a SQL database with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize. [Learn more](#)

Want to try Azure SQL Database for free? Create a free serverless database with the first 100,000 vCore seconds, 32GB of data, and 32GB of backup storage free per month for the lifetime of the subscription. [Learn more](#)

[Apply offer \(Preview\)](#)

SQL Database Hyperscale: Low price, high scalability, and best feature set. [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.

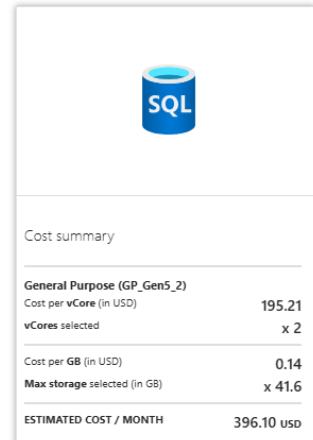
Subscription * ⓘ

Azure subscription 1

Resource group * ⓘ

(New) rg_db

[Create new](#)



Database details

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name *

db_102



Server * ⓘ

(new) dbs102 (West Europe)



[Create new](#)

Want to use SQL elastic pool? ⓘ

Yes No

Workload environment

Development

Production

i Default settings provided for Production workloads. Configurations can be modified as needed.

Compute + storage * ⓘ

General Purpose

Standard-series (Gen5), 2 vCores, 32 GB storage, zone redundant disabled

[Configure database](#)

Click en Configure database:

Configure

[Feedback](#)

Service and compute tier

Select from the available tiers based on the needs of your workload. The vCore model provides a wide range of configuration controls and offers Hyperscale and Serverless to automatically scale your database based on your workload needs. Alternately, the DTU model provides set price/performance packages to choose from for easy configuration. [Learn more](#)

SQL Database Hyperscale: Low price, high scalability, and best feature set. [Learn more](#)

Service tier

General Purpose (Most budget friendly)

[Compare service tiers](#)

Compute tier

Provisioned - Compute resources are pre-allocated. Billed per hour based on vCores configured.

Serverless - Compute resources are auto-scaled. Billed per second based on vCores used.

Compute Hardware

Select the hardware configuration based on your workload requirements. Availability of compute optimized, memory optimized, and confidential computing hardware depends on the region, service tier, and compute tier.

Hardware Configuration

Standard-series (Gen5)

up to 80 vCores, up to 240 GB memory

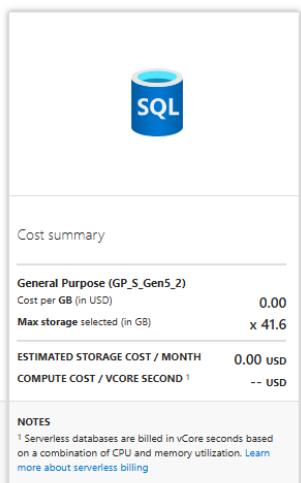
[Change configuration](#)

Max vCores

2

Min vCores

0.5 vCores



2.05 GB MIN MEMORY 6 GB MAX MEMORY

Auto-pause delay

The database automatically pauses if it is inactive for the time period specified here, and automatically resumes when database activity recurs. Alternatively, auto-pausing can be disabled.

Enable auto-pause

Days Hours Minutes

0 1 0

Data max size (GB)

10

3 GB LOG SPACE ALLOCATED

Would you like to make this database zone redundant?

Yes No

Apply

Click en Apply

Next: Networking

Home > SQL databases >

Create SQL Database

Microsoft

Basics Networking Security Additional settings Tags Review + create

Configure network access and connectivity for your server. The configuration selected below will apply to the selected server 'dbs102' and all databases it manages. [Learn more](#)

Network connectivity

Choose an option for configuring connectivity to your server via public endpoint or private endpoint. Choosing no access creates with defaults and you can configure connection method after server creation. [Learn more](#)

Connectivity method * No access Public endpoint Private endpoint

Connection policy

Configure how clients communicate with your SQL database server. [Learn more](#)

Connection policy Default - Uses Redirect policy for all client connections originating inside of Azure (except Private Endpoint connections) and Proxy for all client connections originating outside Azure Proxy - All connections are proxied via the Azure SQL Database gateways Redirect - Clients establish connections directly to the node hosting the database

Encrypted connections

This server supports encrypted connections using Transport Layer Security (TLS). For information on TLS version and certificates, refer to connecting with TLS/SSL. [Learn more](#)

Minimum TLS version

Cost summary

General Purpose (GP_S_Gen5_1)	
Cost per GB (in USD)	0.14
Max storage selected (in GB)	x 13
ESTIMATED STORAGE COST / MONTH 1.78 USD	
COMPUTE COST / VCORE SECOND ¹ 0.000159 USD	

NOTES
¹ Serverless databases are billed in vCore seconds based on a combination of CPU and memory utilization. [Learn more about serverless billing](#)

Review + create [< Previous](#) [Next : Security >](#)

Next: Security

Home > SQL databases >

Create SQL Database

Microsoft

Basics Networking Security Additional settings Tags Review + create

Microsoft Defender for SQL

Protect your data using Microsoft Defender for SQL, a unified security package including vulnerability assessment and advanced threat protection for your server. [Learn more](#)

Get started with a 30 day free trial period, and then 15 USD/server/month.

Enable Microsoft Defender for SQL * Start free trial Not now

Ledger

Ledger cryptographically verifies the integrity of your data and detects any tampering that might have occurred. [Learn more](#)

Ledger [Configure ledger](#)

Server identity

Use system assigned and user assigned managed identities to enable central access management between this database and other Azure resources. [Learn more](#)

Server identity [Configure identities](#)

Cost summary

General Purpose (GP_S_Gen5_1)	
Cost per GB (in USD)	0.14
Max storage selected (in GB)	x 13
ESTIMATED STORAGE COST / MONTH 1.78 USD	
COMPUTE COST / VCORE SECOND ¹ 0.000159 USD	

NOTES
¹ Serverless databases are billed in vCore seconds based on a combination of CPU and memory utilization. [Learn more about serverless billing](#)

Transparent data encryption key management

Transparent data encryption encrypts your databases, backups, and logs at rest without any changes to your application. To enable encryption, go to each database. Database level settings if enabled, will override the server level setting. [Learn more](#)

Server level key ⓘ	Service-managed key selected Configure transparent data encryption
Database level key ⓘ	Not configured Configure transparent data encryption

Always encrypted

Always encrypted is a family of industry-leading data protection features that provide a separation between those who own the data and can view it, and those who manage the data but should have no access, on-premises database administrators, cloud database operators, or other high-privileged but unauthorized users. [Learn more](#)

Enable secure enclaves ⓘ [ON](#) [OFF](#)

Click en: Next: Additional Settings

Create SQL Database ⋮

Microsoft

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

Customize additional configuration parameters including collation & sample data.

Data source

Start with a blank database, restore from a backup or select sample data to populate your new database.

Use existing data * [None](#) [Backup](#) [Sample](#)

Database collation

Database collation defines the rules that sort and compare data, and cannot be changed after database creation. The default database collation is SQL_Latin1_General_CI_AS. [Learn more](#)

Collation * ⓘ [Find a collation](#)

Maintenance window

Select a preferred maintenance window from the drop-down. During maintenance, databases remain available, but some updates may require a failover. The system default maintenance window (5pm to 8am) limits most activities to this time, but urgent updates may occur outside of it. To ensure all updates occur only during the maintenance window, select a non-default option. [Learn more](#)

Maintenance window



Cost summary

General Purpose (GP_S_Gen5_1)	0.14
Cost per GB (in USD)	
Max storage selected (in GB)	x 13
ESTIMATED STORAGE COST / MONTH	1.78 usd
COMPUTE COST / VCORE SECOND ¹	0.000159 usd

NOTES

¹ Serverless databases are billed in vCore seconds based on a combination of CPU and memory utilization. [Learn more about serverless billing](#)

Next: Tags, dejar en blanco

Create SQL Database

Microsoft

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

Tags are name/value pairs that enable you to categorize 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"/> 2 selected <input type="button" value="▼"/>



Cost summary

General Purpose (GP_S_Gen5_2)

Cost per GB (in USD)	0.00
Max storage selected (in GB)	x 13

ESTIMATED STORAGE COST / MONTH	0.00 USD
COMPUTE COST / VCORE SECOND ¹	-- USD

NOTES

¹ Serverless databases are billed in vCore seconds based on a combination of CPU and memory utilization. [Learn more about serverless billing](#)

Next: Review + Create

Create SQL Database ...

Microsoft

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

Product details

SQL database by Microsoft **Estimated cost** Storage cost 0.00 USD / month + Compute cost -- USD / vCore second
[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	Practica09
Region	Italy North
Database name	dbpractica09
Server	(new) dbserverpractica09
Authentication method	SQL and Microsoft Entra authentication
Server admin login	admin01
Microsoft Entra Admin	juanjose.rodriguez@tajamar365.com
Compute + storage	General Purpose - Serverless: Standard-series (Gen5), 2 vCores, 10 GB storage, zone redundant disabled
Backup storage redundancy	Locally-redundant backup storage

Networking

Allow Azure services and resources to access this server	No
Private endpoint	None
Minimum TLS version	1.2
Connection Policy	Default

Security

Identity	Not enabled
Transparent data encryption (Server level)	Service-managed key selected
Database level customer-managed key	Not configured
Database level user assigned managed identity	Not configured
Advanced data security	Not now
Always encrypted with secure enclaves	Not configured
Sql Ledger(Database)	Disabled
Digest Storage	Disabled

Additional settings

Use existing data	Blank
Collation	SQL_Latin1_General_CI_AS
Maintenance window	System default (5pm to 8am)

Tags

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

Click en Create

Microsoft.SQLDatabase.newDatabaseNewServer_395d2728486f47229db49 | Overview

Deployment name : Microsoft.SQLDatabase.newDatabaseNewServer_395d2728486f47229db49
Subscription : Azure subscription 1
Resource group : rg_db

Start time : 11/6/2024, 10:33:39 PM
Correlation ID : 84538db5-694b-42e9-bd9c-e5cab3a24d83

Your deployment is complete

Deployment details

Next steps

Go to resource

Give feedback

Tell us about your experience with deployment

Click en Go to resource:

Home > Microsoft.SQLDatabase.newDatabaseNewServer_e9fb2df74944a408d61e | Overview >

dbpractica09 (dbserverpractica09/dbpractica09) | Overview

SQL database

Overview

Mirror databases in Microsoft Fabric. Easily replicate your existing databases in Fabric, and help your team achieve streamlined ETL and operational analytics goals. [Learn more](#)

Essentials

Resource group (move) : [practica09](#)
Status : Online
Location : Italy North
Subscription (move) : [Azure for Students](#)
Subscription ID : d0bbb60-aded-45eb-97f4-8628a5bd63a9

Tags (edit) : Add tags

Tags (edit)

Getting started Monitoring Properties Features Notifications (0) Integrations Tutorials

Start working with your database

Connect to your database and start working with data with a few simple steps. [Learn more](#)

Configure access
Configure network access to your SQL server. [Learn more](#)

Connect to application
Use connection strings to connect to your SQL database from your applications and favorite tools.

Start developing
Work in your database by using tools to add, modify and query data. [Compare tools](#)

Mirror database in Fabric
Replicate existing databases in Fabric, and help your team achieve streamlined ETL and operational analytics goals. [Learn more](#)

Configure **See connection strings** **Open Azure Data Studio** **Open in Visual Studio** **Open in Visual Studio Code**

Click en Query editor (preview):

dbpractica09 (dbserverpractica09/dbpractica09) | Query editor (preview)

SQL database

Query editor (preview)

Query editor (preview) is a tool to run SQL queries against Azure SQL Database in the Azure portal. It is designed for lightweight querying and object exploration in your database. For more information and troubleshooting, [Learn more](#)

Welcome to SQL Database Query Editor

SQL server authentication

Login *
OR
Microsoft Entra authentication

Reason: An instance-specific error occurred while establishing a connection to SQL Server. Connection was denied since Deny Public Network Access is set to Yes (<https://docs.microsoft.com/azure/sql/database/connectivity-settings#deny-public-network-access>). To connect to this server, use the Private Endpoint from inside your virtual network (<https://docs.microsoft.com/azure/sql-database/sql-database-private-endpoint-overview#how-to-set-up-private-link-for-azure-sql-database>).

OK **Continue as juanjo.rodriguez@taja...**

Para resolver el error nos vamos a Home y luego click en All resources:

Azure services

+ Create a resource SQL databases Storage accounts Resource groups Quotas Subscriptions All resources HDInsight clusters Microsoft Entra ID More services

Resources

Recent Favorite

Name	Type	Last Viewed
db_102 (dbs102/db_102)	SQL database	5 minutes ago
rg_db	Resource group	8 minutes ago
gen2storage102	Storage account	36 minutes ago
rg_gen2	Resource group	40 minutes ago
lab-01-hadoop-hdinsight146689808001	Resource group	an hour ago
Practicas-HDInsight	Resource group	an hour ago
DtlLab-01-Hadoop-HDInsight	Virtual network	9 hours ago
Lab-01-Hadoop-HDInsight	Dev/Test Lab	9 hours ago
Azure subscription 1	Subscription	12 hours ago

See all

Navigate

Subscriptions

Resource groups

All resources

All resources

*

+ Create View

Tools

Home >

All resources ⚡ ...

Tajamar

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

Filter for any field...

Subscription equals all

Resource group equals all

Type equals all

L

0 Unsecure resources

0 Recommendations

3 Changed resources

Name ↑↓

Type ↑↓

dbpractica09 (dbserverpractica09/dbpractica09)

SQL database

dbserverpractica09

SQL server

gen2storagepractica19

Storage account

Click en dbserverpractica09 (SQL Server).

dbserverpractica09

SQL server

Overview

Activity log

Access control (IAM)

Tags

Quick start

Diagnose and solve problems

Settings

Data management

Security

- Networking
- Microsoft Defender for Cloud
- Transparent data encryption
- Identity
- Auditing

Intelligent performance

Automatic tuning

Resource group (move) : Practica09

Status : Available

Location : Italy North

Subscription (move) : Azure for Students

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

Tags (edit) : Add tags

Notifications (0)

Features (6)

All Security (4) Performance (1) Recovery (1)

Microsoft Entra admin
Allows you to centrally manage identity and access to your Azure SQL databases.
CONFIGURED

Microsoft Defender for SQL
Vulnerability Assessment and Advanced Threat Protection.
NOT CONFIGURED

Automatic tuning
Monitors and tunes your database automatically to optimize performance.
CONFIGURED

Auditing
Track database events and writes them to an audit log in Azure storage.
NOT CONFIGURED

Click en Networking:

dbserverpractica09 | Networking

Overview

Activity log

Access control (IAM)

Tags

Quick start

Diagnose and solve problems

Settings

Data management

Security

Networking

Microsoft Defender for Cloud

Public access

Private access

Connectivity

Public network access

Public Endpoints allow access to this resource through the internet using a public IP address. An application or resource that is granted access with the following network rules still requires proper authorization to access this resource. [Learn more](#)

Public network access

Disable

Selected networks

Connections from the IP addresses configured in the Firewall rules section below will have access to this database. By default, no public IP addresses are allowed. [Learn more](#)

Please save public network access value before adding new virtual networks.

Virtual networks

Allow virtual networks to connect to your resource using service endpoints. [Learn more](#)

Activar Selected networks

Virtual networks

Allow virtual networks to connect to your resource using service endpoints. [Learn more](#)

+ Add a virtual network rule

Rule	Virtual network	Subnet	Address range	Endpoint status	Resource group	Subscription	State
------	-----------------	--------	---------------	-----------------	----------------	--------------	-------

Firewall rules

Allow certain public internet IP addresses to access your resource. [Learn more](#)

+ Add your client IPv4 address (2a0c5a8152094300fd2747ec4e27f03) + Add a firewall rule

Rule name	Start IPv4 address	End IPv4 address
-----------	--------------------	------------------

Exceptions

Allow Azure services and resources to access this server

Save **Discard**

En Firewall rules, dar click en Add your client IPv4 address...

[+](#) Add your client IPv4 address (2a0c:5a81:5209:4300:fd27:47e:c4e2:7f03)

Rule name	Start IPv4 address	End IPv4 address
ClientIPAddress_2024-11-10_21-7-9	2a0c:5a81:5209:4300:fd27:47e:c4e2:7f03	2a0c:5a81:5209:4300:fd27:47e:c4e2:7f03

✖ IP address must be a valid IPv4 address.

✖ IP address must be a valid IPv4 address.

Si falla lo anterior , dar click en:

[+](#) Add a firewall rule

Hay que Introducir manualmente la IP, escribir en el navegador whats myip y anotar la dirección ip y llenar. Luego Ok.

Add a firewall rule

Rule name	Start IP	End IP
myip	72.14.201.63	72.14.201.65

OK **Cancel**

Firewall rules

Allow certain public internet IP addresses to access your resource. [Learn more](#)

[+](#) Add your client IPv4 address (2a0c:5a81:5209:4300:fd27:47e:c4e2:7f03) [+](#) Add a firewall rule

Rule name	Start IPv4 address	End IPv4 address
myip	72.14.201.63	72.14.201.65

Exceptions

Allow Azure services and resources to access this server [\(1\)](#)

Save

Discard

Click en Save.

Volver a sql database (dbpractica09) y click en Query editor (preview):

dbpractica09 (dbserverpractica09/dbpractica09) | Query editor (preview)

Overview Activity log Tags Diagnose and solve problems Query editor (preview) Mirror database in Fabric (preview) Settings Data management Integrations Power Platform Security Intelligent performance Monitoring Automation Help

Welcome to SQL Database Query Editor

SQL server authentication

Login * admin01

Password *

OK

OR

Microsoft Entra authentication

Cannot open server 'dbserverpractica09' requested by the login. Client with IP address '188.26.209.102' is not allowed to access the server. To enable access, use the Azure Management Portal or run sp_set_firewall_rule on the master database to create a firewall rule for this IP address or address range. It may take up to five minutes for this change to take effect. Allowlist IP 188.26.209.102 on server dbserverpractica09

Continue as juanjose.rodriguez@taja...

Click en allowlist IP ... on server dbserverpractica09

Continue as [tuemail@tajamar365.com](#)

dbpractica09 (dbserverpractica09/dbpractica09) | Query editor (preview)

Overview Activity log Tags Diagnose and solve problems Query editor (preview) Mirror database in Fabric (preview) Settings Data management Integrations Power Platform Security Intelligent performance Monitoring Automation Help

dbpractica09 (juanjose.rodriguez@...)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Query 1

Run Cancel query Save query Export data as Show only Editor Open Copilot

1

Results Messages

Search to filter items...

Añadir Managed Identity a Gen2 y a DataBase. ¿ Cuál es el propósito de usar Managed Identity?

Managed identities

Copilot

Azure services

All Services (49) Documentation (99+)

Services

See more

Managed Identities External Identities Workload Identities Microsoft Entra Privileged Identity Management

Application HDInsight clusters Azure Databricks More services

Resources

Recent Favorite

Documentation

See more

Managed identities in Azure Container Apps

Click en Create

The screenshot shows the Azure portal's 'Managed Identities' page. At the top, there are navigation links for 'Home' and 'Managed Identities'. Below that is a search bar with placeholder text 'Filter for any field...' and several filter buttons: 'Subscription equals all', 'Resource group equals all', 'Location equals all', and an 'Add filter' button. The main area displays a message 'Showing 0 to 0 of 0 records.' Below this are three sorting options: 'Name ↑↓', 'Type ↑↓', and 'Resource group ↑↓'. A large, semi-transparent watermark of a person's head is visible across the page.

Basics

Create User Assigned Managed Identity

The screenshot shows the 'Create User Assigned Managed Identity' wizard, Step 1: Basics. It has three tabs: 'Basics' (selected), 'Tags', and 'Review + create'. Under 'Project details', it says 'Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.' The 'Subscription' dropdown is set to 'Azure for Students'. The 'Resource group' dropdown is set to 'Practica09', with a 'Create new' option available. Under 'Instance details', the 'Region' dropdown is set to 'Italy North' and the 'Name' input field contains 'mi-practica09'. At the bottom are 'Previous' and 'Next' buttons, and a prominent blue 'Review + create' button.

Click en Next

Tags, dejar en blanco. Click en Next.

Create User Assigned Managed Identity ...

Basics 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 about tags](#)

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"/>	:	Managed Identity

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

Click en Next

Create User Assigned Managed Identity

Basics Tags Review + create

[View automation template](#)

Basics

Subscription	Azure for Students
Resource group	Practica09
Region	Italy North
Name	mi-practica09

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

Click en create.

The screenshot shows the 'Deployment' blade for a resource named 'Microsoft.ManagedIdentity-2024110211553'. The 'Overview' tab is selected, displaying the message 'Your deployment is complete'. Deployment details are listed: Deployment name: Microsoft.ManagedIdentity-2024110211553, Subscription: Azure for Students, Resource group: Practica09. The deployment started at 11/10/2024, 9:19:39 PM and has a Correlation ID: eb4ae8c6-cfd8-aae6-df8e3d8b708c. Below the details, there are sections for 'Deployment details' and 'Next steps', and a 'Go to resource' button.

Una vez creado nos vamos a Home y buscamos gen2storagepractica09

Resources

Recent Favorite

Name

Type

dbpractica09 (dbserverpractica09/dbpractica09)	SQL database
dbserverpractica09	SQL server
Practica09	Resource group
gen2storagepractica19	Storage account
Azure for Students	Subscription

[See all](#)

Click en Access Control (IAM)

gen2storagepractica19

Storage account

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

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

Resource group (move) : Practica09

Location : italynorth

Subscription (move) : Azure for Students

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

Disk state : Available

Tags (edit) : Add tags

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

Click en Add, luego add role assignment

Home > gen2storagepractica19

gen2storagepractica19 | Access Control (IAM)

Storage account

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Add Download role assignments Edit columns Refresh Delete Feedback

Check access Role assignments Roles Deny assignments Classic administrators

My access

View my level of access to this resource.

View my access

Check access

Review the level of access a user, group, service principal, or managed identity has to this resource. [Learn more](#)

Check access

Creamos uno que se llama Storage Blob Data Owner.

Storage Blob Data Contributor	Allows for read, write, and delete access to Azure Storage blob containers and data	BuiltInRole	Storage	View
Storage Blob Data Owner	Allows for full access to Azure Storage blob containers and data, including assigning POSIX access control.	BuiltInRole	Storage	View
Storage Blob Data Reader	Allows for read access to Azure Storage blob containers and data	BuiltInRole	Storage	View
Storage Blob Delegator	Allows for generation of a user delegation key which can be used to sign SAS tokens	BuiltInRole	Storage	View
Storage File Data Privileged Contributor	Customer has read, write, delete and modify NTFS permission access on Azure Storage file shares.	BuiltInRole	None	View
Storage File Data Privileged Reader	Customer has read access on Azure Storage file shares.	BuiltInRole	None	View
Storage File Data SMB Share Contributor	Allows for read, write, and delete access in Azure Storage file shares over SMB	BuiltInRole	Storage	View
Storage File Data SMB Share Elevated Contributor	Allows for read, write, delete and modify NTFS permission access in Azure Storage file shares over SMB	BuiltInRole	Storage	View

Click en Next

Add role assignment

Role **Members** • Conditions Review + assign

Selected role Storage Blob Data Owner

Assign access to User, group, or service principal Managed identity

Members [+ Select members](#)

Name	Object ID
No members selected	

Description

Seleccionar Managed Identity, luego Select members:

Select managed identities

X

⚠ Some results might be hidden due to your ABAC condition.

Subscription *

Azure for Students

Managed identity

User-assigned managed identity (1)

Select i

Search by name

Selected members:



mi-practica09

/subscriptions/dc6bbb60-aded-45eb-87f4-8628a5bd63a9/resourceGroups...

[Remove](#)

Select

Close

Click en Select

[Role](#) [Members](#) [Conditions](#) [Review + assign](#)

Selected role Storage Blob Data Owner

Assign access to User, group, or service principal

Managed identity

Members [+ Select members](#)

Name	Object ID	Type
mi-practica09	768f843f-272c-4e5e-8ca9-9c22129d8880	Managed Identity i

Description

Review + assign

Previous

Next

Click en Next.

Conditions: Dejar como aparece por defecto

Add role assignment ...

Role Members Conditions Review + assign

Selected role Storage Blob Data Owner

Condition ⓘ + Add condition

None

Review + assign

Previous

Next

Click en Next

Add role assignment ...

Role Members Conditions Review + assign

Role Storage Blob Data Owner

Scope /subscriptions/dc6bbb60-aded-45eb-87f4-8628a5bd63a9/resourceGroups/Practica09/providers/Microsoft.Storage/storageAccounts/gen2storagepractica19

Members	Name	Object ID	Type
	mi-practica09	768f843f-272c-4e5e-8ca9-9c22129d8880	Managed Identity ⓘ

Description No description

Condition None

Review + assign

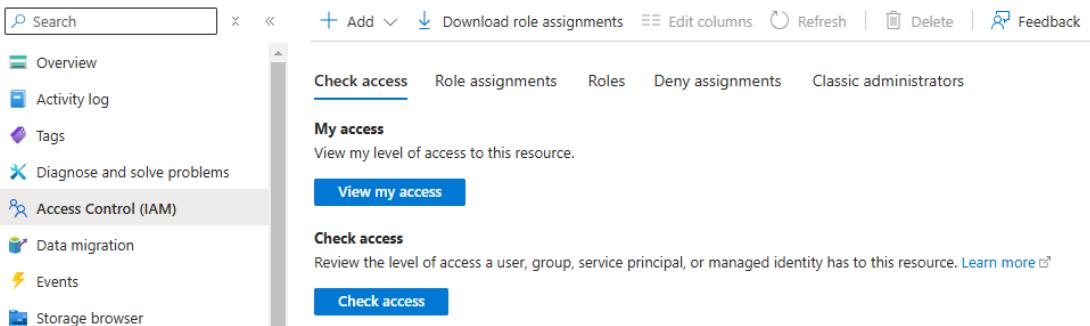
Previous

Next

Click en Review + assign

Home > gen2storagepractica19

 **gen2storagepractica19 | Access Control (IAM)** ⚡ ...
Storage account



Search + Add Download role assignments Edit columns Refresh Delete Feedback

Overview Activity log Tags Diagnose and solve problems Access Control (IAM) Data migration Events Storage browser

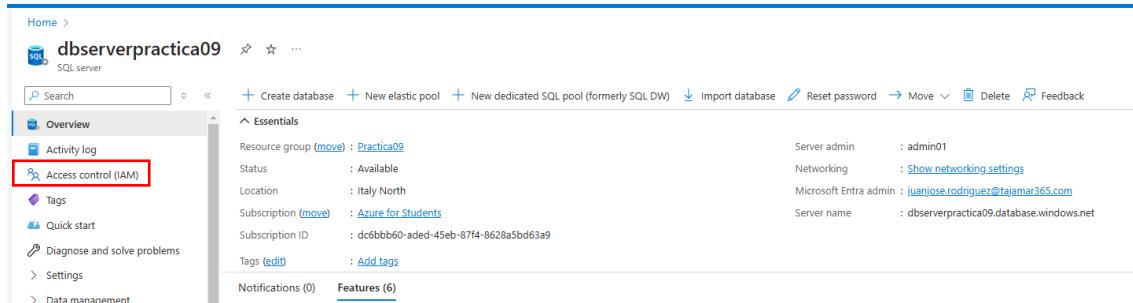
Check access Role assignments Roles Deny assignments Classic administrators

My access
View my level of access to this resource.
View my access

Check access
Review the level of access a user, group, service principal, or managed identity has to this resource. [Learn more](#)

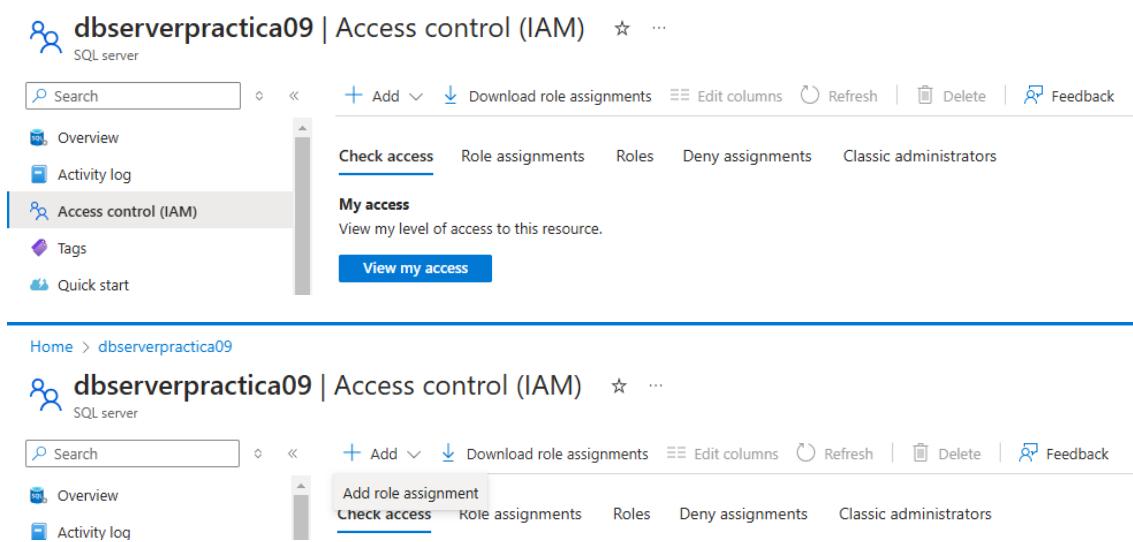
Check access

Nos vamos a Home y buscamos dbserverpractica09 (SQL Server) , luego Access Control (IAM):



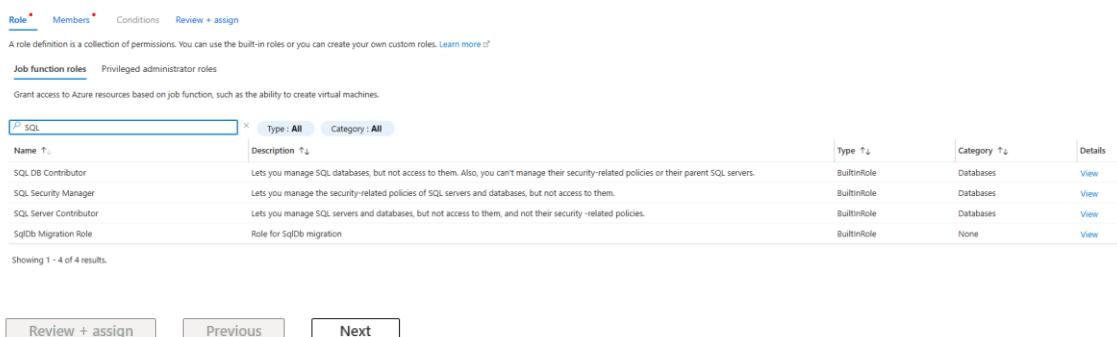
The screenshot shows the Azure portal interface for the dbserverpractica09 SQL server. The left sidebar has a 'Access control (IAM)' link highlighted with a red box. The main content area displays various resource settings like location, subscription, and server admin.

Click en Add, luego add role assignment



The first screenshot shows the 'My access' section with a 'View my access' button highlighted. The second screenshot shows the 'Add role assignment' section with the 'Check access' tab selected.

Seleccionamos SQL DB Contributor:



The screenshot shows the 'Add role assignment' page with a search bar containing 'SQL'. A table lists four roles: SQL DB Contributor, SQL Security Manager, SQL Server Contributor, and SqlDb Migration Role. The 'SQL DB Contributor' row is selected. At the bottom, there are 'Review + assign', 'Previous', and 'Next' buttons.

Click en Next

Add role assignment

Add role assignment ...

Role **Members*** Conditions Review + assign

Selected role SQL DB Contributor

Assign access to User, group, or service principal Managed identity

Members + Select members

Name	Object ID
No members selected	

Description Optional

Managed Identity y luego Select members:

Select managed identities ×

⚠ Some results might be hidden due to your ABAC condition.

Subscription * Azure for Students

Managed identity User-assigned managed identity (1)

Select (i) Search by name

Selected members:
 mi-practica09 </subscriptions/dc6bbb60-aded-45eb-87f4-8628a5bd63a9/resourceGroups...> Remove

Select **Close**

Click en Select

Add role assignment ...

Role Members Conditions Review + assign

Selected role SQL DB Contributor

Assign access to User, group, or service principal Managed identity

Members + Select members

Name	Object ID	Type
mi-practica09	768f843f-272c-4e5e-8ca9-9c22129d8880	Managed Identity ⓘ

Review + assign Previous Next

Click en Next.

Role Members Conditions Review + assign

Role SQL DB Contributor

Scope /subscriptions/dc6bbbb6-aded-45eb-87fa-8628a5bd63a9/resourceGroups/Practica09/providers/Microsoft.Sql/servers/dbserverpractica09

Members

Name	Object ID	Type
mi-practica09	768f843f-272c-4e5e-8ca9-9c22129d8880	Managed Identity ⓘ

Description No description

Review + assign Previous

Click en review + assign

Home > dbserverpractica09

dbserverpractica09 | Access control (IAM) ⭐ ...

SQL server

Search Add Download role assignments Edit columns Refresh Delete Feedback

Overview Activity log Access control (IAM) Tags Quick start Diagnose and solve problems Settings

Check access Role assignments Roles Deny assignments Classic administrators

My access
View my level of access to this resource.
View my access

Check access
Review the level of access a user, group, service principal, or managed identity has to this resource. [Learn more](#)

Ahora, vamos a crear con esas identidades asignadas un Cluster SQL Interactivo con HDInsight. Nos vamos a Home y escribimos: HDInsight clusters

All services >

HDInsight clusters

Default Directory

+ Create 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.

Name	Cluster type	Status	Resource group
------	--------------	--------	----------------

No HDInsight clusters to display

Create an HDInsight cluster to process massive amounts of data using popular open-source frameworks such as Hadoop, Spark, Hive, LLAP, Kafka, Storm, ML Services, and more.

[Create HDInsight cluster](#)

[Learn more about HDInsight](#)

Click en create:

Basics

Create HDInsight cluster ...

Basics Storage Security + networking Configuration + pricing Tags Review + create

New to HDInsight? Get started with our [training resources](#).

Create a managed HDInsight cluster. Select from Spark, Kafka, Hadoop, Storm, and more. [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 *

Practica09

[Create new](#)

Cluster details

Name your cluster, pick a region, and choose a cluster type and version. [Learn more](#)

Cluster name *

clusterpractica09



Region *

Italy North



Cluster type *

Select cluster type

Version

▼

Click en Select cluster type:

Select cluster type



Hadoop

Petabyte-scale processing with Hadoop components like MapReduce, Hive (SQL on Hadoop), Pig, Sqoop and Oozie.

[Select](#)

Spark

Fast data analytics and cluster computing using in-memory processing.

[Select](#)

Kafka

Build a high throughput, low-latency, real-time streaming platform using a fast, scalable, durable, and fault-tolerant publish-subscribe messaging system.

[Select](#)

HBase

Fast and scalable NoSQL database. Available with both standard and premium (SSD) storage options.

[Select](#)

Interactive Query

Build Enterprise Data Warehouse with in-memory analytics using Hive (SQL on Hadoop) and LLAP (Low Latency Analytical Processing). Note that this feature requires high memory instances.

[Select](#)

Escoger el que dice Interactive Query (Select)

Cluster details

Name your cluster, pick a region, and choose a cluster type and version. [Learn more](#)

Cluster name *

clusterpractica09



Region *

Italy North



Availability zone ⓘ

▼

Cluster type *

Interactive Query

[Change](#)

Version *

Interactive Query 3.1.2 (HDI 5.1)



Asignar login y password para el Cluster

Cluster credentials

Enter new credentials that will be used to administer or access the cluster.

Cluster login username *	admin01	✓
Cluster login password *	*****	✓
Confirm cluster login password *	*****	✓ >Password and confirm password must match.
Secure Shell (SSH) username *	sshuser	✓
Use cluster login password for SSH	<input checked="" type="checkbox"/>	

[Review + create](#)

[« Previous](#)

[Next: Storage »](#)

Click en Next: Storage >>

Storage

Create HDInsight cluster ...

[Basics](#) [Storage](#) [Security + networking](#) [Configuration + pricing](#) [Tags](#) [Review + create](#)

Select or create storage accounts that will be used for the cluster's logs, job input, and job output. Configure the cluster's access to these accounts, if needed.

Primary storage

Select or create a storage account that will be the default location for cluster logs and other output.

Primary storage type *	Azure Data Lake Storage Gen2	▼
Primary storage account *	gen2storagepractica19	▼
Filesystem *	filesystemhive	✓ The container name must be at least two characters in length and can contain digits, lowercase letters, and/or hyphens. It must not begin or end with a hyphen, and it cannot contain two consecutive hyphens.
<input checked="" type="checkbox"/> Enable secure channel	○	

Identity

Select a user-assigned managed identity to represent the cluster for Azure Data Lake Gen2 Storage account access. Only identities with access to the selected storage account are listed. Assign the managed identity to the 'Storage Blob Data Owner' role on the storage account. [Learn More](#)

User-assigned managed identity * ⓘ

mi-practica09



Additional Azure Storage

Link additional Azure Storage accounts to the cluster.

[Add Azure Storage](#)

Custom Ambari DB

Use an external Ambari database for greater flexibility, control, and customization. [Learn More](#)

SQL database for Ambari ⓘ



External metadata stores

To store your Hive and Oozie metadata outside of this cluster, select a SQL database. [Learn More](#)

SQL database for Hive ⓘ



SQL database for Oozie ⓘ



[Review + create](#)

[« Previous](#)

[Next: Security + networking »](#)

Click en Next: Security + Networking

Dejar lo que aparece por defecto:

Create HDInsight cluster ...

TLS

Select the minimum TLS version supported for your cluster. [Learn More](#)

Minimum TLS version ⓘ

1.2



Network settings

Resource provider connection ⓘ

Inbound



Connect this cluster to a virtual network. [Learn More](#)

Virtual network ⓘ



Encryption in transit

Configure encryption in transit settings. [Learn More](#)

Enable encryption in transit ⓘ

Encryption at rest

Configure disk encryption settings. [Learn More](#)

Provide your own key from key vault ⓘ

Enable encryption at host on temp data disk ⓘ

Identity

Select a user-assigned service identity to represent your cluster for enterprise security package or disk encryption. [Learn More](#)

User-assigned managed identity ⓘ

- I don't want to provide a managed service identity



Review + create

[« Previous](#)

[Next: Configuration + pricing »](#)

Click en Next: Configuration + pricing >>

Create HDInsight cluster

Basics Storage Security + networking Configuration + pricing Tags Review + create

Configure cluster performance and pricing. [Learn More](#)

Node configuration

Configure your cluster's size and performance, and view estimated cost information.

The cost estimate represented in the table does not include subscription discounts or costs related to storage, networking, or data transfer.

 Sku family standardav2family quota is not enough, available value: 34, additional cores: 4; There are not enough cores available to support the selected number of nodes. Please adjust the number of nodes selected, pick a different region, or open a support case to request additional HDInsight cores.
[View cores usage](#)
[Open an HDInsight quota increase support case](#)

[+ Add application](#)

Node type	Node size	Number of ...	Estimated cost/h...
Head node	E8 V3 (8 Cores, 64 GB RAM), 0.76 USD/hour	2	1.52 USD
Zookeeper node	A2 v2 (2 Cores, 4 GB RAM), 0.12 USD/hour	3	0.36 USD
Worker node	A8m v2 (8 Cores, 64 GB RAM), 0.68 USD/hour	4	2.73 USD

Enable autoscale
[Learn More](#)

Total estimated cost/hour 4.61 USD

En el worker node cambiar el 4 por 1

Node configuration

Configure your cluster's size and performance, and view estimated cost information.

The cost estimate represented in the table does not include subscription discounts or costs related to storage, networking, or data transfer.

 This configuration will use 30 of 34 available cores in the Italy North region.
[View cores usage](#)
[Open an HDInsight quota increase support case](#)

[+ Add application](#)

Node type	Node size	Number of ...	Estimated cost/h...
Head node	E8 V3 (8 Cores, 64 GB RAM), 0.76 USD/hour	2	1.52 USD
Zookeeper node	A2 v2 (2 Cores, 4 GB RAM), 0.12 USD/hour	3	0.36 USD
Worker node	A8m v2 (8 Cores, 64 GB RAM), 0.68 USD/hour	1	0.68 USD

Enable autoscale
[Learn More](#)

Total estimated cost/hour 2.56 USD

Script actions

Use script actions to run custom PowerShell or Bash scripts on cluster nodes during cluster provisioning. [Learn about script actions](#)

[Add script action](#)

[Review + create](#)

[« Previous](#)

[Next: Tags »](#)

Click en Next: Tags >>

Dejar lo que aparece por defecto

Create HDInsight cluster [...](#)

Basics

Storage

Security + networking

Configuration + pricing

Tags

[Review + create](#)

Name 

Value 

Resource

:

HDInsight cluster

[Review + create](#)

[« Previous](#)

[Next: Review + create »](#)

Click en Next: Review + create >>

Create HDInsight cluster ...

✓ Validation succeeded.

Basics Storage Security + networking Configuration + pricing Tags **Review + create**

⚠️ Encryption in transit is enabled. Default node-node communication is using internet protocol security, you can opt-out to plain text protocol based on the job latency needs.

Interactive Query 3.1.2 (HDI 5.1)

2.56 USD Total estimated cost/hour

This estimate does not include subscription discounts or costs related to storage, networking, or data transfer.

Basics

Subscription	Azure for Students
Resource group	Practica09
Region	Italy North
Cluster name	(new) clusterpractica09
Cluster type	Interactive Query 3.1.2 (HDI 5.1)
Cluster login username	admin01
Secure Shell (SSH) username	sshuser
Use cluster login password for SSH	Enabled

Security + networking

Minimum TLS version	1.2
Resource provider connection	Inbound
Encryption at rest	Disabled
Encryption in transit	Enabled
Encryption at host on temp data disk	Disabled

Create

« Previous

Next »

Download a template for automation

Click en Create. Dependiendo de la conexión, tardará unos minutos.

Home >

 **HDInsight_2024-11-10T20.56.52.412Z | Overview** ⚡ ...

Deployment

Search X <> Delete Cancel Redeploy Download Refresh

Overview

Your deployment is complete

Deployment name: HDInsight_2024-11-10T20.56.52.412Z
Subscription: Azure for Students
Resource group: Practica09

Start time: 11/10/2024, 9:56:48 PM
Correlation ID: 7282b214-23f7-4802-9cdf-51b20508d4b9

Deployment details

Next steps

Setup autoscale Recommended

Go to resource

Give feedback

Tell us about your experience with deployment

“Data Ingestion”.

Ingestar nuestros datos al Data Lake Storage

Buscamos el archivo FlightDelayData.csv.

Luego en Azure nos vamos a gen2storagepractica09

Home > **gen2storagepractica19** Storage account

Overview

Essentials

Resource group (move)	:	Practica09	Performance	:	Standard
Location	:	italynorth	Replication	:	Locally-redundant storage (LRS)
Subscription (move)	:	Azure for Students	Account kind	:	StorageV2 (general purpose v2)
Subscription ID	:	dcc6bb60-aded-45eb-87f4-8628a5bd63a9	Provisioning state	:	Succeeded
Disk state	:	Available	Created	:	10/11/2024, 20:27:13

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

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

Data Lake Storage

Hierarchical namespace	Enabled	Require secure transfer for REST API operations	Enabled
Default access tier	Cold	Storage account key access	Enabled
Blob anonymous access	Disabled	Minimum TLS version	Version 1.2
Blob soft delete	Enabled (7 days)	Infrastructure encryption	Disabled
Container soft delete	Enabled (7 days)		
Versioning	Disabled		
Change feed	Disabled		
NFS v3	Disabled		
SFTP	Disabled		
Storage tasks assignments	None		

Security

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

Networking

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

File service

Click en Data Lake Storage:

Home > **gen2storagepractica19** Storage account

Containers

Search containers by prefix:

Name	Last modified	Anonymous access level	Lease state
Slogs	11/10/2024, 8:27:55 PM	Private	Available
filesystemhive	11/10/2024, 9:56:53 PM	Private	Available

Activity log Tags Diagnose and solve problems Access Control (IAM) Data migration Events Storage browser Partner solutions Data storage Containers

Click en Storage Browser



Click en Blob Containers y luego en filesystemhive

Name	Last modified	Access tier	Blob type	Size	Lease state
HdiSamples	10/11/2024, 22:15:11				...
ams	10/11/2024, 21:57:41				...
amshbase	10/11/2024, 21:57:41				...
app-logs	10/11/2024, 21:57:41				...
apps	10/11/2024, 21:57:41				...
atshistory	10/11/2024, 21:57:41				...
custom-scriptaction-logs	10/11/2024, 22:15:38				...
example	10/11/2024, 22:14:51				...
hbase	10/11/2024, 21:57:41				...
hdp	10/11/2024, 21:57:41				...
hive	10/11/2024, 21:57:41				...
mapred	10/11/2024, 21:57:41				...
mr-history	10/11/2024, 21:57:41				...
tmp	10/11/2024, 21:57:41				...
user	10/11/2024, 21:57:41				...
warehouse	10/11/2024, 21:57:44				...
yarn	10/11/2024, 22:10:36				...

Dando click en Add Directory, creamos una carpeta con nombre demo y luego FlightDelayData.

Click en Save

The screenshot shows the Azure Blob Storage interface for a filesystemhive. At the top, there are navigation links: '+ Add Directory', 'Upload' (with a red box around it), 'Refresh', 'Delete', 'Copy', 'Paste', and 'Rename'. Below the navigation is a breadcrumb trail: 'Blob containers > filesystemhive > demo'. A search bar says 'Search blobs by prefix (case-sensitive)'. The main area displays 'Showing all 1 items' with a table:

	Name	Last modified
<input type="checkbox"/>	[.]	
<input type="checkbox"/>	FlightDelayData	10/11/2024, 22:39:41

Entramos a FlightDelayData y creamos una carpeta con nombre InputData

The screenshot shows the same Azure Blob Storage interface, but now the user has navigated to the 'FlightDelayData' folder. In the top right corner, there is a form to create a new folder:

Name *	<input type="text" value="InputData"/> ✓
--------	---

Below the form, the breadcrumb trail shows: 'Blob containers > filesystemhive > demo > FlightDelayData'. The search bar and table from the previous screenshot are also present.

Click en Save.

Entramos a la carpeta InputData y subimos el fichero FlightDelayData.

The screenshot shows the 'InputData' folder within 'FlightDelayData'. The 'Upload' button is highlighted with a red box. The breadcrumb trail is: 'Blob containers > filesystemhive > demo > FlightDelayData > InputData'. The search bar and table are present.

Showing all 0 items

	Name	Last modified	Access tier
<input type="checkbox"/>	[.]		

Upload blob

1 file(s) selected: FlightDelayData.csv
Drag and drop files here or [Browse for files](#)

Overwrite if files already exist

▼ Advanced

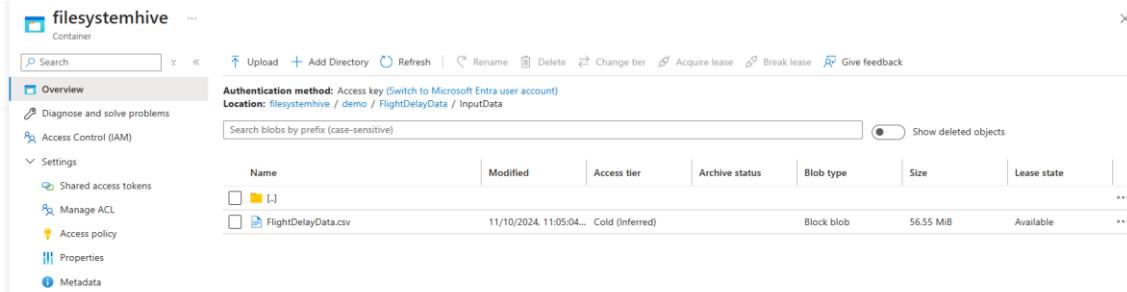
[Upload](#) [Give feedback](#)

Click en Upload

Current uploads Dismiss: [Completed](#) [All](#)

FlightDelayData.csv  0 / 56.55 MiB

Esperar hasta completarse.

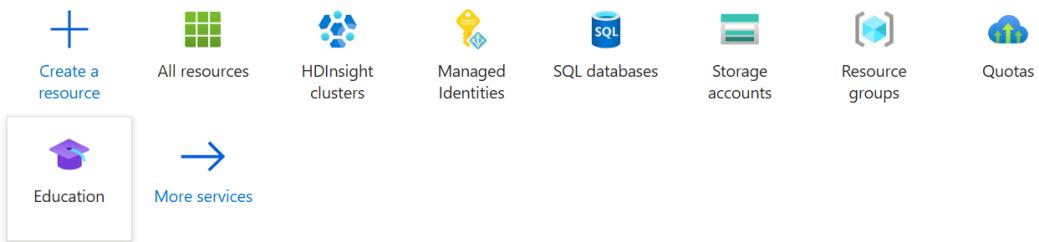


The screenshot shows the Azure Storage Blob container 'filesystemhive'. The left sidebar has 'Overview' selected. The main area displays a table of blobs:

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
FlightDelayData.csv	11/10/2024, 11:05:04...	Cold (Inferred)		Block blob	56.55 MiB	Available

“Data Extraction” con Hive

Desde Home, damos click en nuestro cluster (clusterpractica09):



Resources

Recent Favorite

Name	Type	Last Viewed
gen2storagepractica19	Storage account	3 minutes ago
clusterpractica09	HDInsight cluster	6 minutes ago
Practica09	Resource group	an hour ago
dbserverpractica09	SQL server	2 hours ago
dbpractica09 (dbserverpractica09/dbpractica09)	SQL database	2 hours ago
Azure for Students	Subscription	3 weeks ago

Click en Ambari home:

This screenshot shows the 'clusterpractica09' HDInsight cluster details page. The left sidebar includes options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings (Cluster size, Quota limits, SSH + Cluster login, Data Lake Storage Gen1, Storage accounts), and Help. The main panel displays the 'Essentials' section with cluster details: Status (Running), Location (Italy North), Subscription (Azure for Students), Subscription ID (dc6bbb60-aded-45eb-87f4-8628a5bd63a9), and Tags (None). It also shows the 'Overview' tab selected under 'Dashboards' with links to Ambari home, Zeppelin notebook, Grafana, Tez, and Yarn.

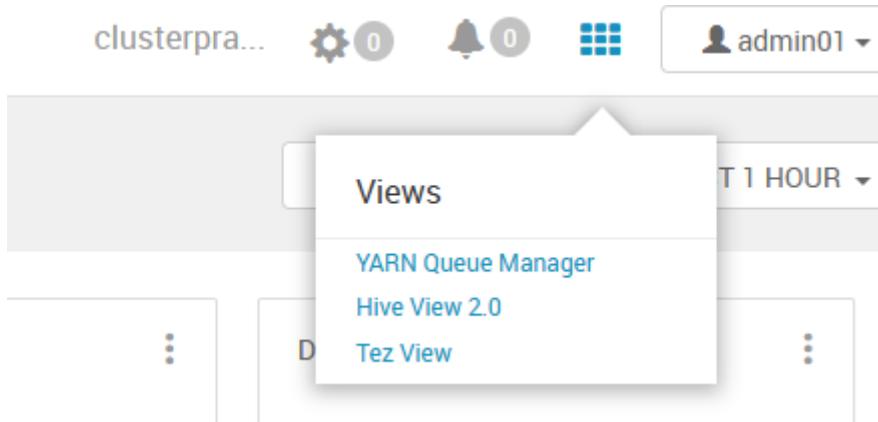
Te pedirá el usuario y clave que creaste al momento de levantar el Clúster.

This screenshot shows the Ambari Metrics dashboard for the 'clusterpractica09' cluster. The URL is 'clusterpractica09.azurehdinsight.net/main/dashboard/metrics'. The dashboard has tabs for METRICS, HEATMAPS, and CONFIG HISTORY. The METRICS tab is selected, showing a graph and some data points. The top navigation bar shows the cluster name and user 'admin01'.

Una vez dentro haz click en:



Escoger Hive View 2.0



Después de validar el hdfs se mostrará el editor de ‘queries’ de hive:

Copiar y pegar el contenido del fichero query-1.txt.

```

1 DROP TABLE flightdelay_raw;
2
3 CREATE EXTERNAL TABLE flightdelay_raw(
4   YEAR string,
5   FL_DATE string,
6   UNIQUE_CARRIER string,
7   CARRIER string,
8   FL_NUM string,
9   ORIGIN_AIRPORT_ID string,
10  ORIGIN string,
11  ORIGIN_CITY_NAME string,
12  ORIGIN_CITY_NAME_TEMP string,
13  ORIGIN_STATE_ABR string,
14  DEST_AIRPORT_ID string,

```

Luego click en execute

Ir a Tables

The screenshot shows the Apache Hive interface. At the top, there are tabs for QUERY, JOBS, TABLES, SAVED QUERIES, UDFs, and SETTINGS. On the right, there are buttons for '+ NEW JOB', '+ NEW TABLE', and 'NOTIFICATIONS'. Below the tabs, a search bar says 'Select or search database/schema' with 'default' selected. Under 'TABLES | 3', there is a list of tables: 'hivesamplable', 'flighthdelay_raw' (which is highlighted), and 'flighthdelay'. The 'flighthdelay_raw' table is expanded to show its columns. The column details are as follows:

COLUMN NAME	COLUMN TYPE
year	string
fl_date	string
unique_carrier	string
carrier	string
fl_num	string
origin_airport_id	string
origin	string
origin_city_name	string
origin_city_name_temp	string
origin_state_abr	string
dest_airport_id	string
dest	string
dest_city_name	string
dest_city_name_temp	string

Volvemos a Query

The screenshot shows the Apache Hive interface with a new query worksheet. The top navigation bar is identical to the previous screenshot. The main area has a 'CreateTables' tab selected. A '+' button is visible. The 'DATABASE' dropdown is set to 'default'. The query editor contains the following DDL code:

```
1 DROP TABLE flighthdelay_raw;
2
3 CREATE EXTERNAL TABLE flighthdelay_raw(
4     YEAR string,
5     FL_DATE string,
6     UNIQUE_CARRIER string,
7     CARRIER string,
8     FL_NUM string,
9     ORIGIN_AIRPORT_ID string,
10    ORIGIN string,
11    ORIGIN_CITY_NAME string,
12    ORIGIN_CITY_NAME_TEMP string,
13    ORIGIN_STATE_ABR string,
14    DEST_AIRPORT_ID string,
```

Below the code, there are buttons for 'Execute', 'Save As', 'Insert UDF', and 'Visual Explain'. The bottom navigation bar includes 'RESULTS' (selected), 'LOG', 'VISUAL EXPLAIN', 'TEZ UI', and 'Filter columns'.

Abrimos un nuevo worksheet en el símbolo +:

The screenshot shows the Apache Hive interface. At the top, there's a navigation bar with tabs for QUERY, JOBS, TABLES, SAVED QUERIES, UDFs, and SETTINGS. Below the navigation bar, there are two tabs: 'CreateTables' and 'Worksheet2'. The 'Worksheet2' tab is active and highlighted in black. A '+' button is located to the right of the tabs. Underneath the tabs, there's a 'DATABASE' section with a dropdown menu set to 'default'. The main area is a large text editor where a single digit '1' is visible at the top left. At the bottom of the editor, there are several buttons: 'Execute' (green background), 'Save As', 'Insert UDF', and 'Visual Explain'.

Hagamos una consulta simple:

```
1 select * from flightdelay limit 5;
```

Click en Execute

This screenshot is identical to the one above, showing the Apache Hive interface with the 'Worksheet2' tab active. In the main query editor area, the previously entered query 'select * from flightdelay limit 5;' is displayed. The 'Execute' button at the bottom is now highlighted in green, indicating it has been clicked. The other buttons ('Save As', 'Insert UDF', 'Visual Explain') are still visible below it.

La salida de la consulta se puede ver en Results:

flightdelay.year	flightdelay.flight_date	flightdelay.unique_carrier	flightdelay.carrier	flightdelay.flight_num	flightdelay.origin_airport_id	flightdelay.origin_airport_code	flight
YEAR	FL_DATE	P_UNIQUE_CARRIER	P_CARRIER_AIRLINE_ID	P_CARRIER_FL_NUM	ORIGIN_AIRPORT_ID	ORIGIN_AIRPORT_CODE	Flight
2019	1/19/2019	E	0363	245	13487	SP	Minne
2019	1/20/2019	E	0363	245	13487	SP	Minne
2019	1/21/2019	E	0363	245	13487	SP	Minne
2019	1/22/2019	E	0363	245	13487	SP	Minne

En esta sección hemos realizado una ingestión desde Data Lake Storage Gen 2 hasta un lugar dentro del cluster utilizando T-SQL mediante Hive.

“Data Transformation” con Hive

Abrimos un nuevo worksheet y ejecutamos la siguiente query (query-3.txt)

```

HIVE
clusterpra... 0
QUERY JOBS TABLES SAVED QUERIES UDFs SETTINGS
Worksheet1 * Worksheet1 *
DATABASE default
Select or search database/schema
1 INSERT OVERWRITE DIRECTORY '/demo/FlightDelayData/output'
2 ROW FORMAT DELIMITED FIELDS TERMINATED BY(',')
3 SELECT regexp_replace(origin_city_name, ' ', ''),
4 avg(weather_delay)
5 FROM flightdelay
6 WHERE weather_delay IS NOT NULL
7 GROUP BY origin_city_name;
Execute Save As Insert UDF Visual Explain

```

Explicar paso a paso que hace la query-3 de arriba.

Nos vamos a Azure, hasta gen2storagepractica09

Click en containers y luego click a filesystemhive:

filesystemhive Container

Search | Upload | Add Directory | Refresh | Rename | Delete | Change tier

Overview

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

Name	Modified	Access tier
ams	11/10/2024, 9:57:41 ...	
amshbase	11/10/2024, 9:57:41 ...	
app-logs	11/10/2024, 9:57:41 ...	
apps	11/10/2024, 9:57:41 ...	
atshistory	11/10/2024, 9:57:41 ...	
custom-scriptaction-logs	11/10/2024, 10:15:38...	
demo	11/10/2024, 10:39:14...	
example	11/10/2024, 10:14:51...	
hbase	11/10/2024, 9:57:41 ...	
.../11/10/2024 10:45:44		

Click en demo:

filesystemhive Container

Search | Upload | Add Directory | Refresh | Rename | Delete | Change tier

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: filesystemhive / demo

Search blobs by prefix (case-sensitive)

Name	Modified
FlightDelayData	11/10/2024, 10:39:41...

Click en FlightDelayData y luego output:

Home > gen2storagepractica19 | Containers >

filesystemhive Container

Search Upload Add Directory Refresh Rename Delete Change tier

Overview

Authentication method: Access key ([Switch to Microsoft Entra user account](#))
Location: filesystemhive / demo / FlightDelayData / output

Search blobs by prefix (case-sensitive)

Name	Modified	Access tier
.hive-staging_hive_2024-11-10_23-21-45_387_7...	11/11/2024, 12:21:45...	
000000_0	11/11/2024, 12:21:47...	Cold (Inferred)
000001_0	11/11/2024, 12:21:47...	Cold (Inferred)
000002_0	11/11/2024, 12:21:47...	Cold (Inferred)

La información que tenemos en output no tiene una estructura SQL, hive guardó la salida en chunks. Para tener nuestros datos en un formato SQL usaremos Sqoop.

“Data Export” mediante Sqoop.

Antes de usar sqoop vamos a nuestra base de datos desde azure yendo a dbpractica09.

Home

Azure services

[Create a resource](#) [All resources](#) [HDInsight clusters](#) [Managed Identities](#) [SQL databases](#) [Storage accounts](#) [Resource groups](#) [Quotas](#) [Education](#) [More services](#)

Resources

Recent Favorite

Name	Type	Last Viewed
dbpractica09 (dbserverpractica09/dbpractica09)	SQL database	7 minutes ago
gen2storagepractica19	Storage account	20 minutes ago
clusterpractica09	HDInsight cluster	an hour ago
Practica09	Resource group	3 hours ago
dbserverpractica09	SQL server	3 hours ago
Azure for Students	Subscription	3 weeks ago

[See all](#)

Click en dbpractica09 (SQL Database)

Home > dbpractica09 (dbserverpractica09/dbpractica09) ⚡ ⭐ ...

SQL database

Search Copy Restore Export Set server firewall Delete Connect with... Feedback

Overview Activity log Tags Diagnose and solve problems Query editor (preview) Mirror database in Fabric (preview) Settings Data management Integrations Power Platform Security

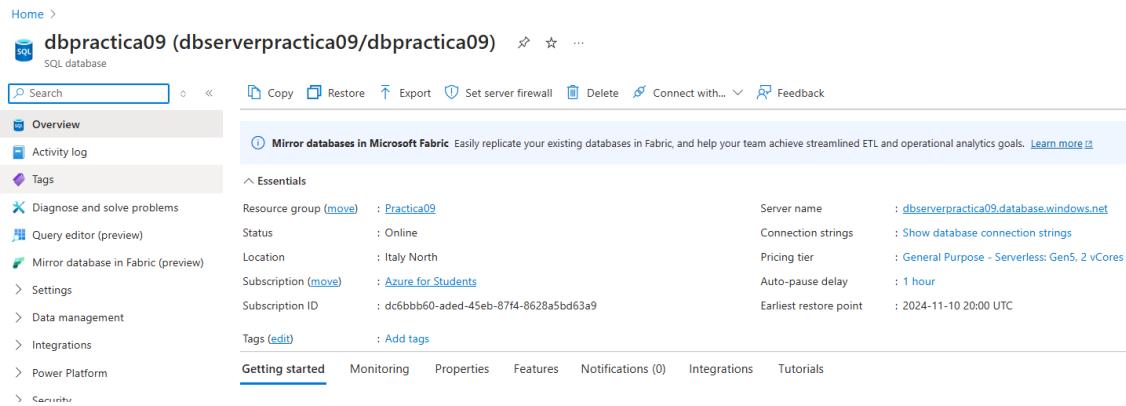
Mirror databases in Microsoft Fabric Easily replicate your existing databases in Fabric, and help your team achieve streamlined ETL and operational analytics goals. [Learn more](#)

Essentials

Resource group (move)	: Practica09	Server name	: dbserverpractica09.database.windows.net
Status	: Online	Connection strings	: Show database connection strings
Location	: Italy North	Pricing tier	: General Purpose - Serverless: Gen5, 2 vCores
Subscription (move)	: Azure for Students	Auto-pause delay	: 1 hour
Subscription ID	: dc6bbb60-aded-45eb-87f4-8628a5bd63a9	Earliest restore point	: 2024-11-10 20:00 UTC

Tags (edit) : Add tags

Getting started Monitoring Properties Features Notifications (0) Integrations Tutorials



Click en Query editor (preview)

Home > dbpractica09 (dbserverpractica09/dbpractica09)

dbpractica09 (dbserverpractica09/dbpractica09) | Query editor (preview) ⭐ ...

SQL database

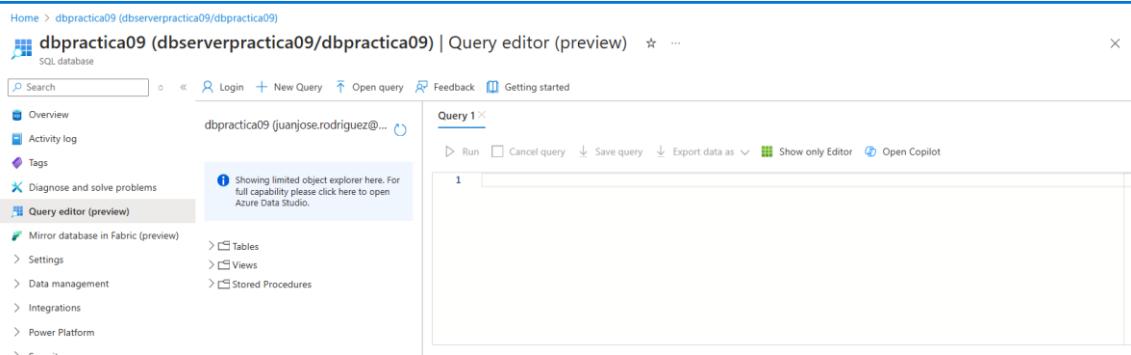
Search Login New Query Open query Feedback Getting started

Overview Activity log Tags Diagnose and solve problems Query editor (preview) Mirror database in Fabric (preview) Settings Data management Integrations Power Platform

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Query 1 ×

Run Cancel query Save query Export data as Show only Editor Open Copilot



En el editor creamos nuestra tabla mediante la query:

Query 1 ×

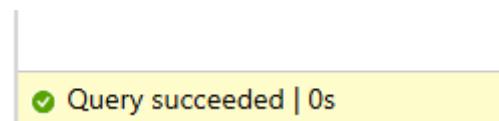
Run Cancel query Save query Export data as Show only Editor Open Copilot

```
1 CREATE TABLE [dbo].[delays](
2     [origin_city_name] [nvarchar](50) NOT NULL,
3     [weather_delay] float,
4     CONSTRAINT [PK_delays] PRIMARY KEY CLUSTERED
5     ([origin_city_name] ASC)
6     GO
7
8     SELECT * FROM information_schema.tables
9     GO
10    
```



Click en run:

Query succeeded | 0s



The screenshot shows the Azure Data Studio interface. On the left, the Object Explorer displays a tree structure with 'Tables' expanded, showing 'dbo.delays' with columns 'origin_city_name' and 'weather_delay'. Other nodes include 'Views' and 'Stored Procedures'. On the right, the 'Query 1' tab contains a T-SQL script:

```

1 CREATE TABLE [dbo].[delays](
2     [origin_city_name] [nvarchar](50) NOT NULL,
3     [weather_delay] float,
4     CONSTRAINT [PK_delays] PRIMARY KEY CLUSTERED
5     ([origin_city_name] ASC)
6     GO
7
8 SELECT * FROM information_schema.tables
9 GO
10

```

The 'Results' tab shows a table with the following data:

TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	TABLE_TYPE
dbpractica09	sys	database_firewall_rules	VIEW
dbpractica09	dbo	delays	BASE TABLE

Ahora, vamos de nuevo a nuestro clúster:

The screenshot shows the Azure portal's HDInsight cluster overview page for 'clusterpractica09'. The left sidebar includes options like 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', 'Diagnose and solve problems', and 'Settings' (with sub-options for 'Cluster size', 'Quota limits', 'SSH + Cluster login', 'Data Lake Storage Gen1', 'Storage accounts', and 'Applications'). The main area displays cluster status, dashboards for Ambari home, Zeppelin notebook, Grafana, Tez, and Yarn, and recommended features for 'Auto scale', 'Applications', and 'Script actions'.

Click en SSH + Cluster login

The screenshot shows the 'SSH + Cluster login' section of the HDInsight cluster overview. The left sidebar has 'SSH + Cluster login' selected. The main area contains instructions for connecting via SSH ('Connect to cluster using secure shell (SSH)') and cluster login ('Connect to cluster using Cluster Login'), both with 'Learn More' links. It also shows a 'Cluster login username' field with 'admin01' entered and a 'Reset credential' button.

En Hostname seleccionamos el ssh:

Connect to cluster using secure shell (SSH)

You can securely connect to the below endpoints in the HDInsight cluster with an SSH client. [Learn More](#)

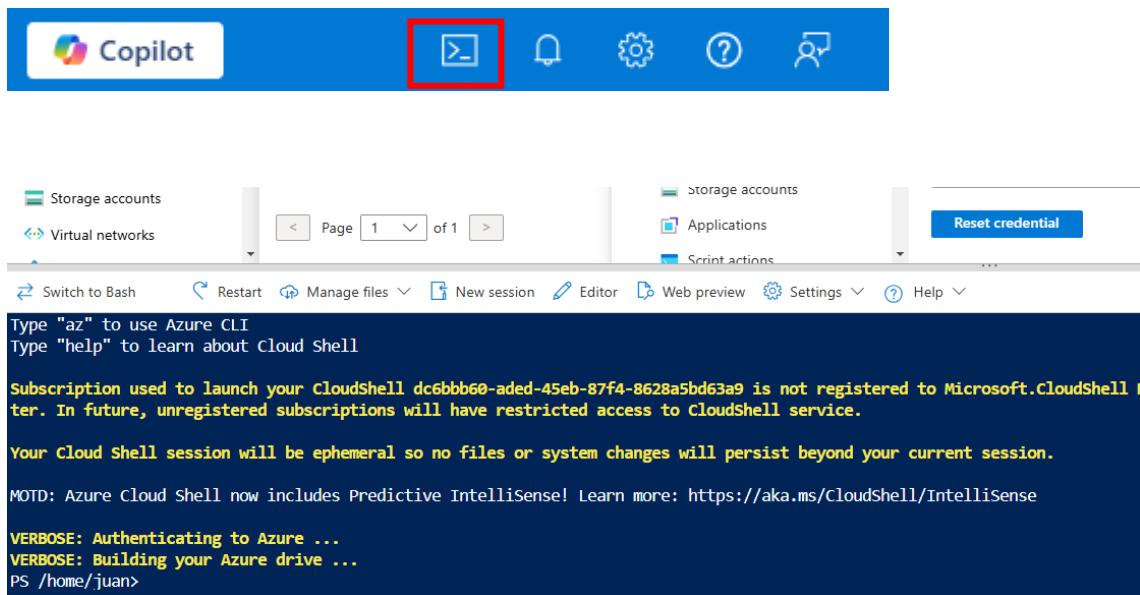
Hostname

clusterpractica09-ssh.azurehdinsight.net

ssh sshuser@clusterpractica09-ssh.azurehdinsight.net

Copiamos el texto.

Nos vamos al powershell de Azure.



Pegar el string (ssh) que tenemos copiado:

```
VERBOSE: Authenticating to Azure ...
VERBOSE: Building your Azure drive ...
PS /home/juan> ssh sshuser@clusterpractica09-ssh.azurehdinsight.net
The authenticity of host 'clusterpractica09-ssh.azurehdinsight.net (72.146.16.2)' can't be established.
ED25519 key fingerprint is SHA256:MLpn3z5lb1mLiqbgoqhX0ec14iZIT6jmkbicxd3tQJc.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? 
```

Yes

```
VERBOSE: Authenticating to Azure ...
VERBOSE: Building your Azure drive ...
PS /home/juan> ssh sshuser@clusterpractica09-ssh.azurehdinsight.net
The authenticity of host 'clusterpractica09-ssh.azurehdinsight.net (72.146.16.2)' can't be established.
ED25519 key fingerprint is SHA256:MLpn3z5lb1mLiqbgoqhX0ec14iZIT6jmkbicxd3tQJc.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'clusterpractica09-ssh.azurehdinsight.net' (ED25519) to the list of known hosts.
Authorized uses only. All activity may be monitored and reported.
sshuser@clusterpractica09-ssh.azurehdinsight.net's password: 
```

Introducir cluster password:

To see these additional updates run: apt list --upgradable

*** /dev/sda1 will be checked for errors at next reboot ***

Welcome to LLAP on HDInsight.

The programs included with the Ubuntu system are free software; the exact distribution terms for each program are described in the individual files in /usr/share/doc/*/*copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by applicable law.

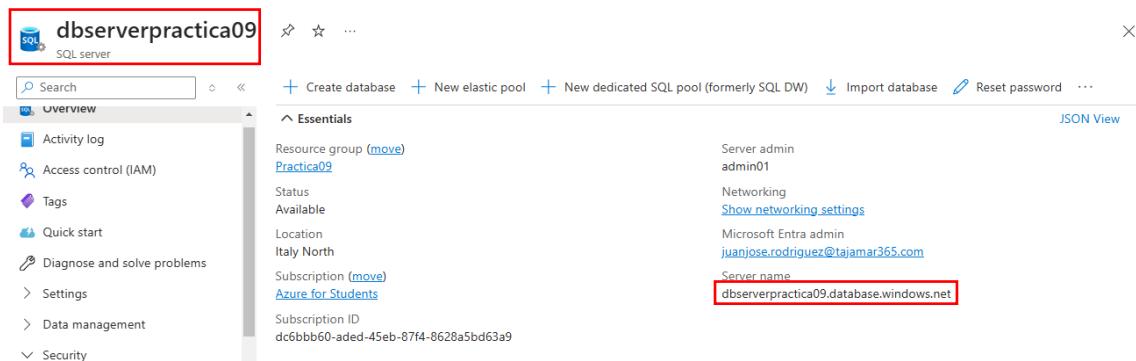
To run a command as administrator (user "root"), use "sudo <command>". See "man sudo_root" for details.

sshuser@hn0-cluste:~\$

Escribimos:

```
sqoop list-databases --connect jdbc:sqlserver://dbserverpractica09.database.windows.net:1433 --username admin01 -P
```

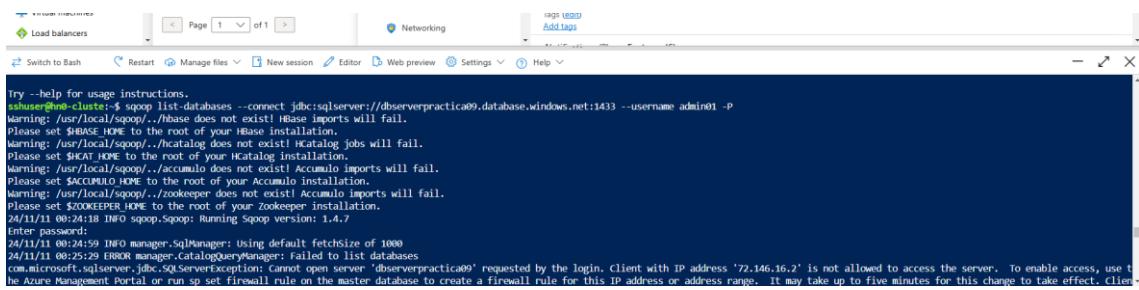
Donde: dbserverpractica09.database.windows.net es el nombre de tu servidor



The screenshot shows the Azure portal interface for managing a SQL database. The top navigation bar includes 'SQL' and 'dbserverpractica09'. Below the navigation, there's a search bar and several action buttons: '+ Create database', '+ New elastic pool', '+ New dedicated SQL pool (formerly SQL DW)', 'Import database', 'Reset password', and '...'. On the left, a sidebar lists navigation items like 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', 'Quick start', 'Diagnose and solve problems', 'Settings', 'Data management', and 'Security'. The main content area is titled 'Essentials' and displays the following details for the database:

Resource group (move)	Practical09	Server admin	admin01
Status	Available	Networking	Show networking settings
Location	Italy North	Microsoft Entra admin	juanjoose.rodriguez@tajamar365.com
Subscription (move)	Azure For Students	Server name	dbserverpractica09.database.windows.net
Subscription ID	dc6bbb60-aded-45eb-87f4-8628a5bd63a9		

Si no esta instalado Sqoop en tu HDInsight hay buscar en la documentación o en ChatGPT como hacerlo.



```
Try --help for usage instructions.
sshuser@hn0-cluste:~$ sqoop list-databases --connect jdbc:sqlserver://dbserverpractica09.database.windows.net:1433 --username admin01 -P
Warning: /usr/local/sqoop../.hbase does not exist! Hbase imports will fail.
Please set $HBASE_HOME to the root of your Hbase installation.
Warning: /usr/local/sqoop../.hadoop does not exist! Accumulo imports will fail.
Please set $HADOOP_HOME to the root of your Hadoop installation.
Warning: /usr/local/sqoop../.accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /usr/local/sqoop../.zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
24/11/11 00:24:18 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
Enter password:
24/11/11 00:24:59 INFO manager.SqlManager: Using default fetchSize of 1000
24/11/11 00:25:29 ERROR manager.CatalogManager: Failed to list databases
com.microsoft.sqlserver.jdbc.SQLServerException: Cannot open server 'dbserverpractica09' requested by the login. Client with IP address '72.146.16.2' is not allowed to access the server. To enable access, use t
he Azure Management Portal or run sp_set_firewall_rule on the master database to create a firewall rule for this IP address or address range. It may take up to five minutes for this change to take effect. Client-
```

Te pide la contraseña de tu base de datos. Si da error es seguramente por el firewall. Copiamos la IP que se muestra en powershell y luego vamos a nuestro sql server (dbserverpractica09):

The screenshot shows the Azure portal interface for the dbserverpractica09 SQL server. The top navigation bar includes options like 'Create database', 'New elastic pool', 'Import database', and 'Reset password'. The left sidebar lists various service integrations and monitoring features. The main content area displays essential information such as the Resource group (Practica09), Status (Available), Location (Italy North), and Subscription (Azure for Students). It also shows the Server admin (admin01), Networking settings, Microsoft Entra admin (juanjose.rodriguez@tajamar365.com), and the Server name (dbserverpractica09.database.windows.net).

Click en Networking:

The screenshot shows the Networking blade for the dbserverpractica09 SQL server. The left sidebar has a 'Networking' section selected. The main area displays 'Firewall rules' which allow certain public internet IP addresses to access the resource. It includes a table for managing rules and a checkbox for allowing Azure services to access the server.

Click en Add a firewall rule:

The screenshot shows the 'Add a firewall rule' dialog box. It has fields for 'Rule name' (containing 'myip'), 'Start IP' (72.14.201.63), and 'End IP' (72.14.201.65). At the bottom are 'OK' and 'Cancel' buttons.

Añadimos la ip copiada que muestra el error en powershell:

Add a firewall rule

Rule name	Start IP	End IP
PowerShell	72.146.16.2	72.146.16.2
<input type="button" value="OK"/> <input type="button" value="Cancel"/>		

Click en OK

Firewall rules

Allow certain public internet IP addresses to access your resource. [Learn more](#)

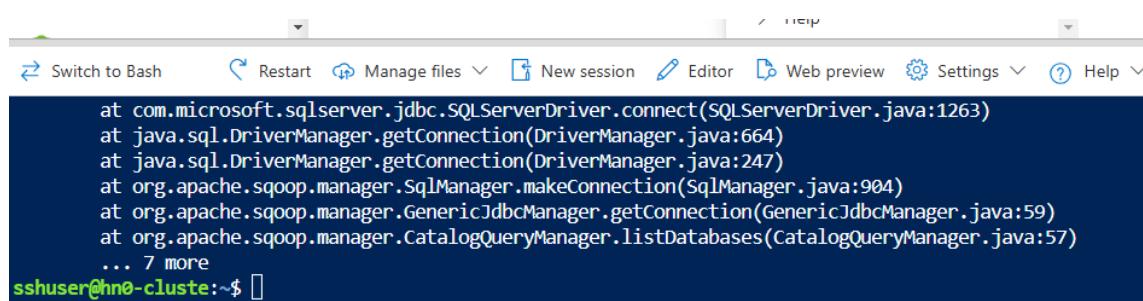
+ Add your client IPv4 address (2a0c:5a81:5209:4300:fd27:47e:c4e2:7f03) + Add a firewall rule

Rule name	Start IPv4 address	End IPv4 address
PowerShell	72.146.16.2	72.146.16.2
mylp	72.14.201.63	72.14.201.65
query-editor-9222b3	188.26.209.102	188.26.209.102

Exceptions

Click en Save

Nos vamos de nuevo a PowerShell



```
at com.microsoft.sqlserver.jdbc.SQLServerDriver.connect(SQLServerDriver.java:1263)
at java.sql.DriverManager.getConnection(DriverManager.java:664)
at java.sql.DriverManager.getConnection(DriverManager.java:247)
at org.apache.sqoop.manager.SqlManager.makeConnection(SqlManager.java:904)
at org.apache.sqoop.manager.GenericJdbcManager.getConnection(GenericJdbcManager.java:59)
at org.apache.sqoop.manager.CatalogQueryManager.listDatabases(CatalogQueryManager.java:57)
... 7 more
sshuser@hn0-cluste:~$
```

Volvemos a escribir el comando de sqoop list -databases...

```
sqoop list-databases --connect jdbc:sqlserver://dbserverpractica09.database.windows.net:1433 --username admin01 -P

24/11/11 00:43:36 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
Enter password:
24/11/11 00:43:42 INFO manager.SqlManager: Using default fetchSize of 1000
master
dbpractica09
sshuser@hn0-cluste:~$
```

Una vez dentro, escribimos:

```
sqoop export --connect 'jdbc:sqlserver://dbserverpractica09.database.windows.net:1433;database=dbpractica09' --username admin01 -P --table 'delays' --export-dir '/demo/FlightDelayData/output' --fields-terminated-by ',' -m 1
```

El proceso tardará unos 5 o 10 minutos.

Una vez que se termina la ejecución, nos vamos a nuestra base de datos y verificamos que la tabla se ha llenado:

The screenshot shows a SQL Server Object Explorer interface. On the left, there's a tree view with 'Tables' expanded, showing 'dbo.delays' selected. The main area has a query window with the following content:

```
Run Cancel query Save query Export data as json Export data as .csv Export data as .xml
1  SELECT TOP (1000) * FROM [dbo].[delays]
```

Below the query window, the results pane is visible, titled 'Results'. It contains a table with the following data:

origin_city_name	weather_delay
Aberdeen	41.5
Abilene	21.78125
Adak Island	0
Aguadilla	3.71428571428571
Akron	28

