



CosmosDB – Laboratorio

Nombre del Laboratorio:
CosmosDB

Preparado para:
Microsoft

Fecha:
[12/06/2019]

Versión: [3.0]

Preparado por:
David Sánchez Almanza
oss@bside.com.mx

bSide
Expertos en Soluciones TI, S.C.



Contenido

Metas y objetivos del Laboratorio.....	3
Alcance del Laboratorio	3
Desarrollo del Laboratorio.....	4
Creación del recurso de Cosmos DB en Azure	4
Ejecución de consultas a Azure Cosmos DB empleando el API de SQL	13
Ejecución de consultas.....	42
Limpieza de los recursos del laboratorio	47
Conclusiones	47
Contacto	47



Metas y objetivos del Laboratorio

Esta sección describe las metas y los objetivos que se pretende lograr con este laboratorio.

A continuación, se enlistan las metas

- Creación de los recursos en Azure.
- Ejecución de consultas utilizando la herramienta de Data Explorer desde el Portal de Azure.
- Exploración de las capacidades en el manejo de dato con Cosmos DB.
- Exploración de las capacidades en cuanto a versatilidad y eficiencia en el acceso a datos a través del API de SQL en Cosmos DB.
- Automatización de flujos de datos.
- Conocer las capacidades de cumplimiento de Cosmos DB como componente de soluciones para ambientes de producción.

A continuación, se enlistan los objetivos:

- Que los participantes conozcan las bondades y facilidad de despliegue de soluciones con una base de datos multi plataforma sin esquema con capacidades virtualmente ilimitadas.

Alcance del Laboratorio

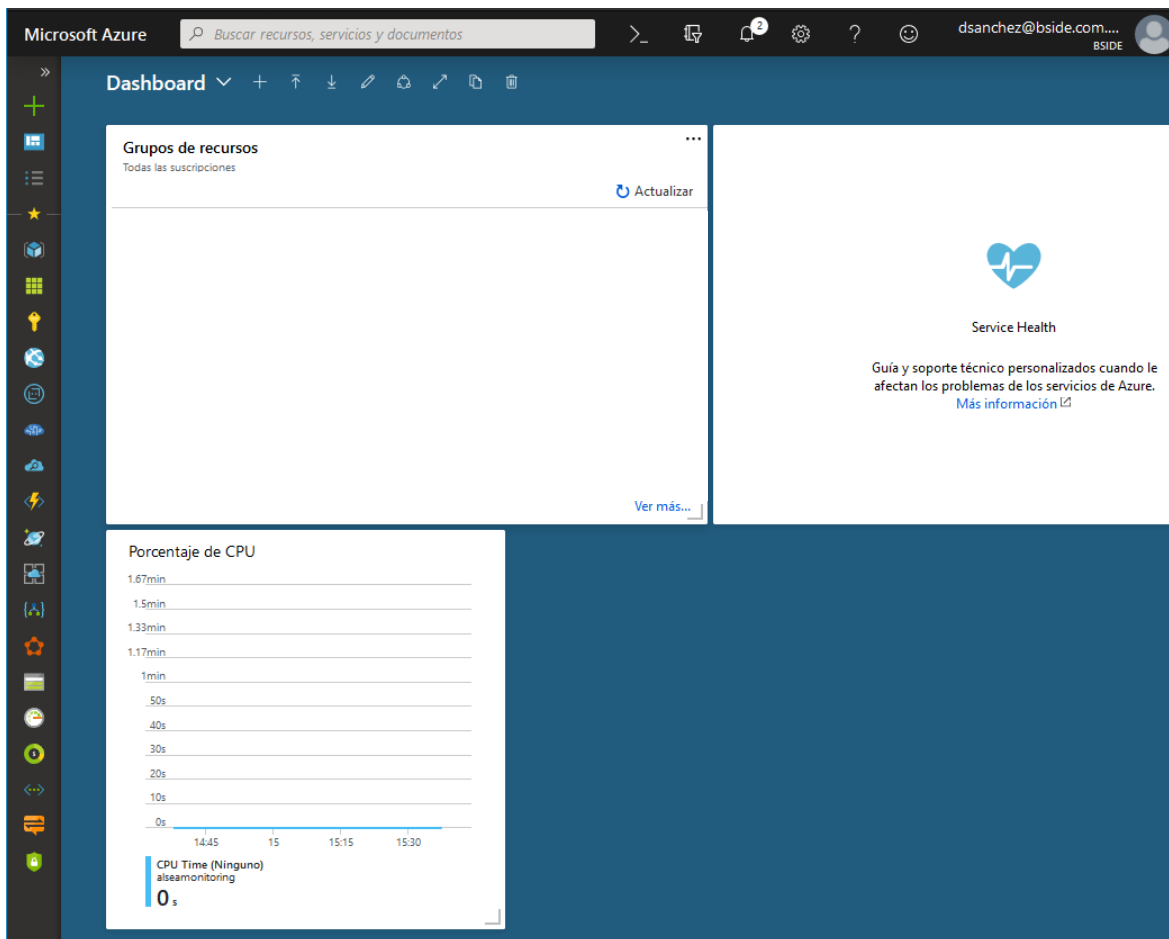
Como parte de esta demo solo se cubre:

1. Creación y configuración de base de datos Cosmos DB en Azure.
2. Creación y configuración de un Storage Account.
3. Montaje de un archivo como BLOB.
4. Creación de un servicio de Data Factory para la generación de un *Pipeline* o flujo de datos.
5. Migrar el archivo a un esquema de base de datos de SQL.
6. Realizar consultas a la colección de SQL para obtener las RU's utilizadas.
7. Conclusiones.

Desarrollo del Laboratorio

Creación del recurso de Cosmos DB en Azure

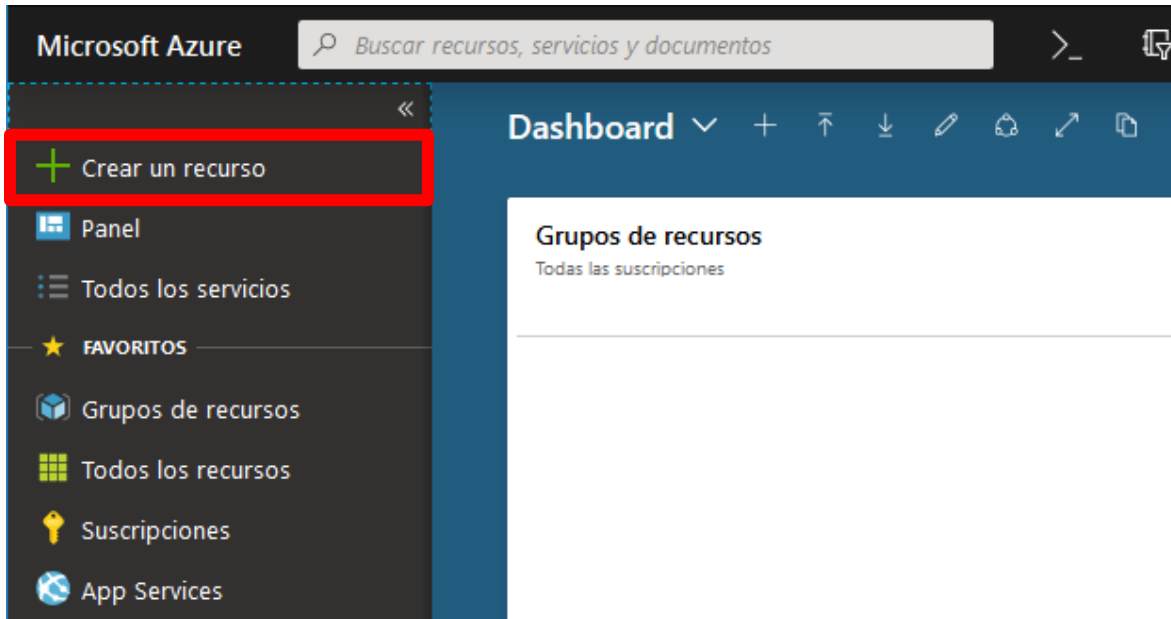
Acceder a Azure en la liga: <http://portal.azure.com/>



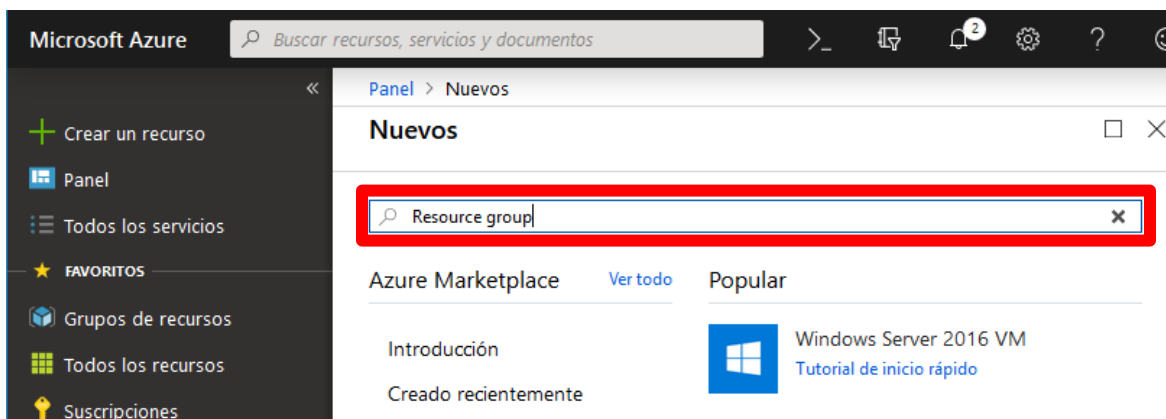
4

Si es la primera vez que utiliza el portal de Azure se desplegará una guía de inicio, hacer clic en saltar este paso.

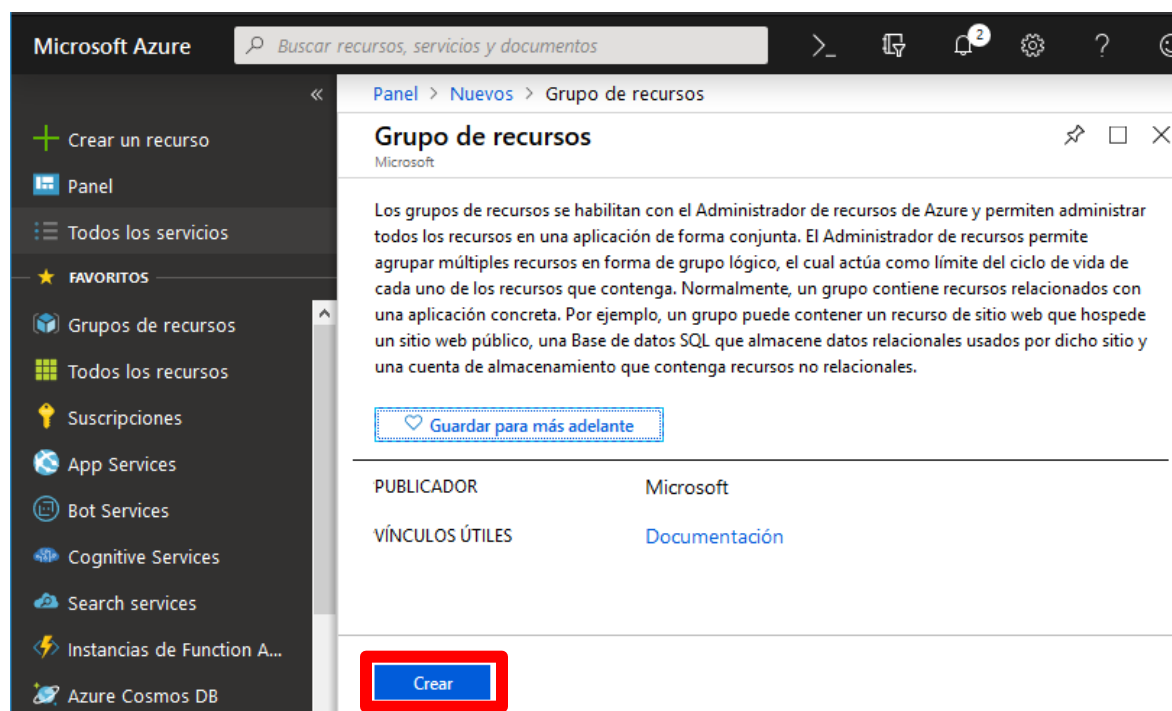
Crear un nuevo Grupo de Recursos haciendo clic en el botón de “+ Crear un recurso” en la barra de herramientas de la izquierda.



En la barra de búsqueda escribir **Resource group** y seleccionar el recurso del mismo nombre.



Una vez seleccionado el recurso hacer clic en el botón de **Crear**.

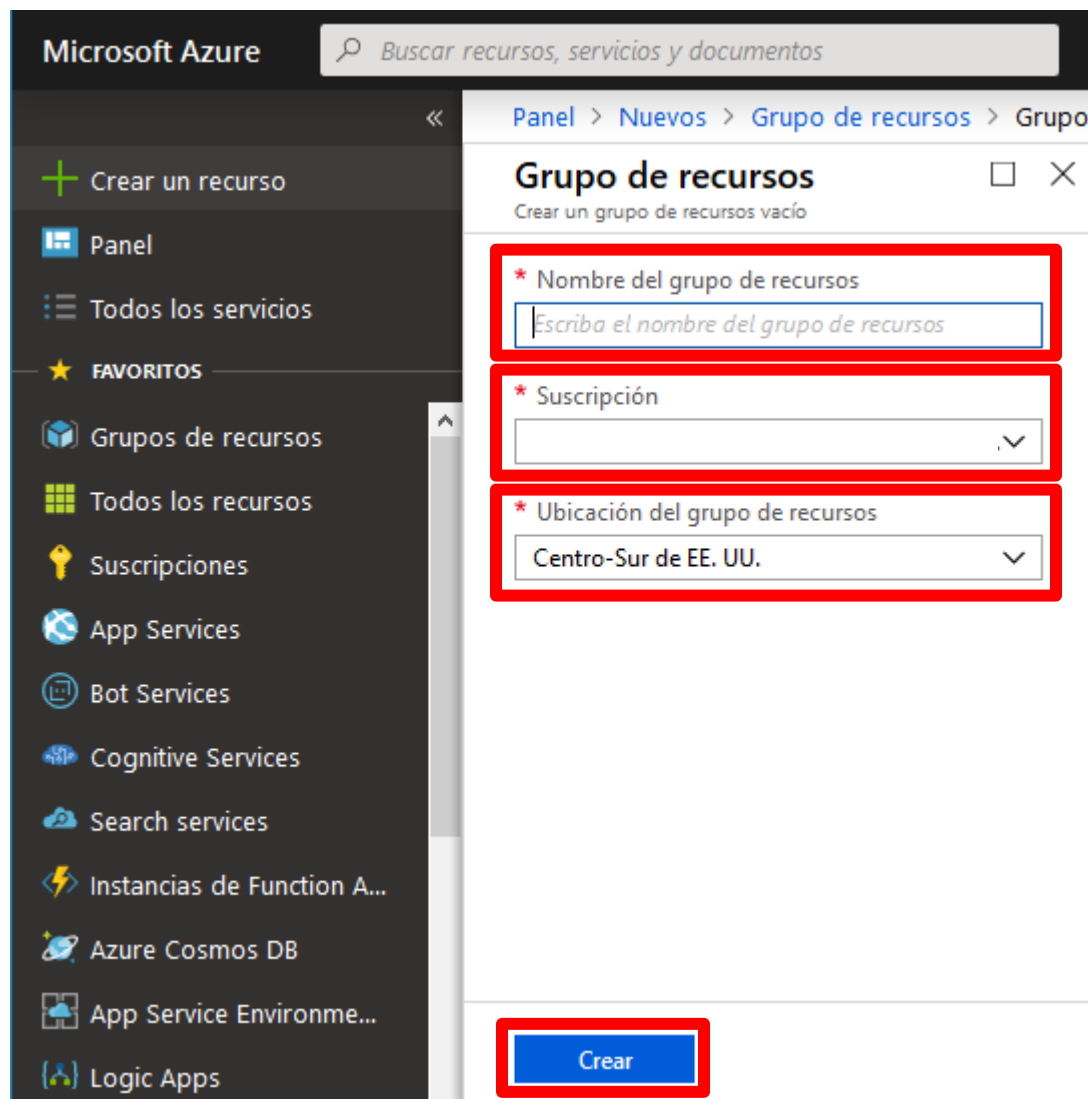


The screenshot shows the Microsoft Azure portal interface. On the left is a navigation pane with options like 'Crear un recurso', 'Panel', 'Todos los servicios', and 'FAVORITOS'. The main area displays the 'Grupo de recursos' page for Microsoft. It includes a description of resource groups and a 'Guardar para más adelante' button. At the bottom, there is a 'PUBLICADOR' section with 'Microsoft' and a 'VÍNCULOS ÚTILES' section with a link to 'Documentación'. A red box highlights the 'Crear' button at the bottom left of the main content area.

La creación del recurso contempla tres campos solicitados:

- Nombre del grupo de recursos.
- Suscripción (Seleccionar la suscripción en la cual se desea colocar el grupo de recursos).
 - Recordar que los grupos de recursos son agrupadores lógicos.
- Ubicación del grupo de recursos
 - Se recomienda que los recursos creados, sino son globales, se creen en la ubicación geográfica más cercana a la ubicación de acceso de los usuarios.

Hacer clic en crear para instanciar el **Grupo de Recursos**.



Microsoft Azure

Buscar recursos, servicios y documentos

Panel > Nuevos > Grupo de recursos > Grupo

Grupo de recursos

Crear un grupo de recursos vacío

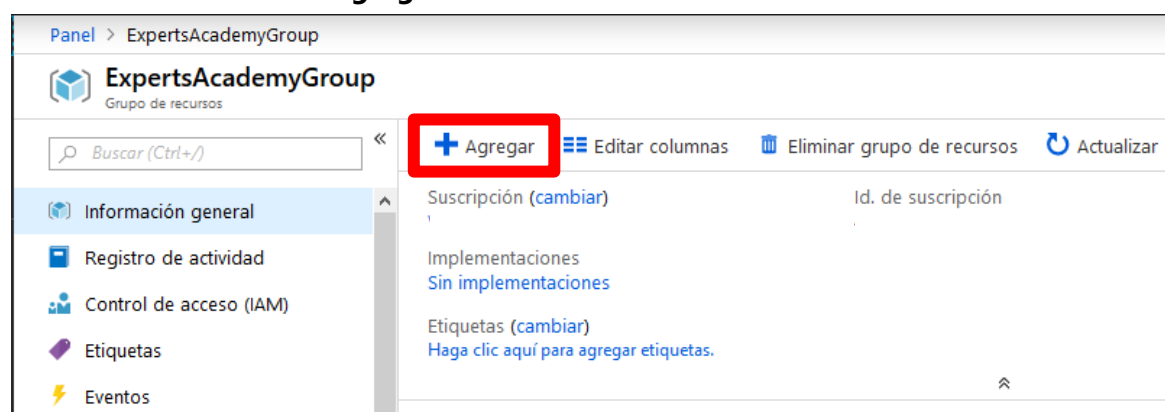
* Nombre del grupo de recursos
Escriba el nombre del grupo de recursos

* Suscripción
▼

* Ubicación del grupo de recursos
Centro-Sur de EE. UU. ▼

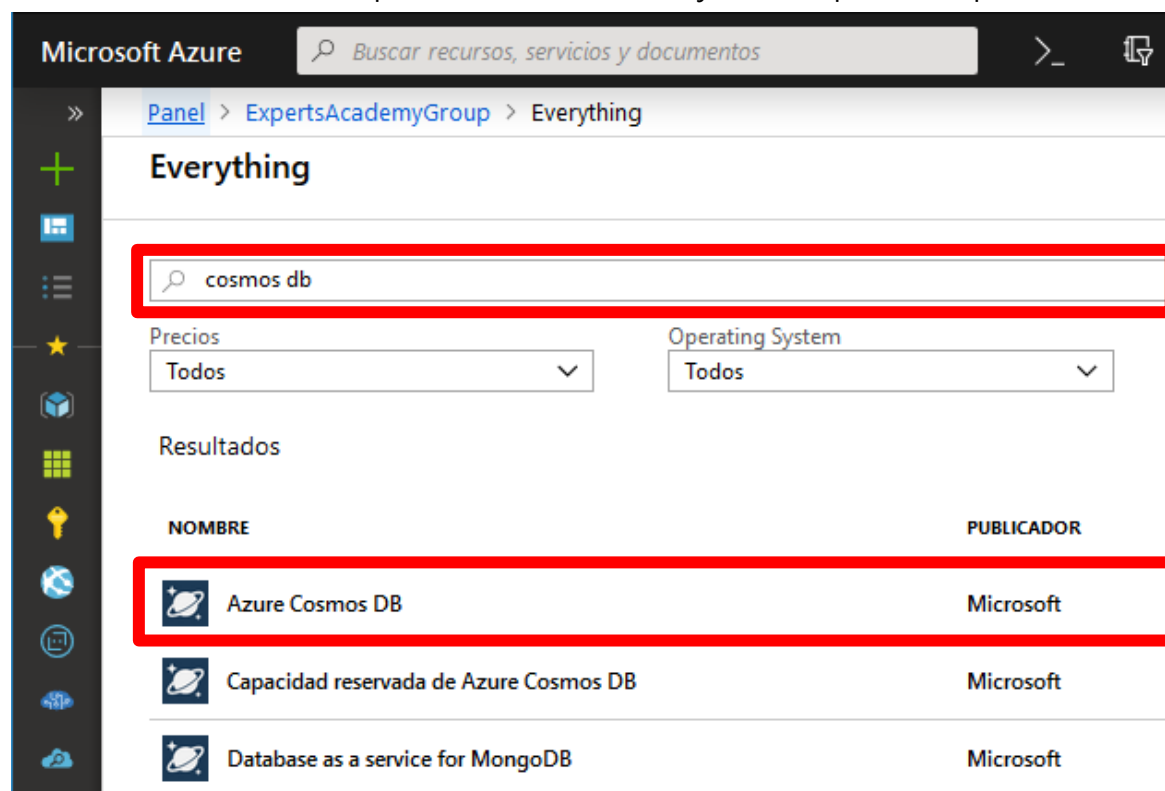
Crear

Hace clic en el botón de **Agregar**

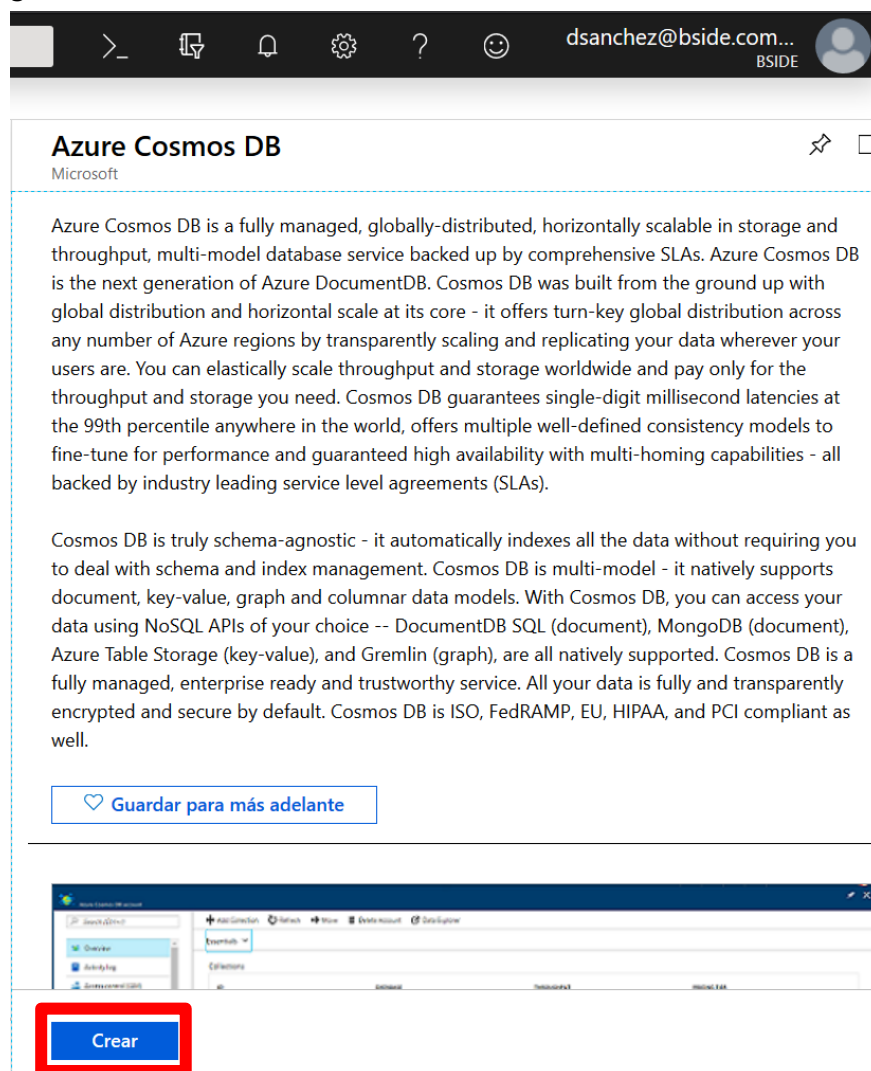


Se desplegará una hoja de búsqueda de recursos.

En la barra de búsqueda escribir **Cosmos** o **Cosmos DB** para filtrar y seleccionar la opción de **Azure Cosmos DB** en el panel de resultados debajo del campo de búsqueda.



Una vez seleccionado el recurso de **Azure Cosmos DB**, hacer clic en el botón de **Crear** para generar el recurso.



Azure Cosmos DB
Microsoft

Azure Cosmos DB is a fully managed, globally-distributed, horizontally scalable in storage and throughput, multi-model database service backed up by comprehensive SLAs. Azure Cosmos DB is the next generation of Azure DocumentDB. Cosmos DB was built from the ground up with global distribution and horizontal scale at its core - it offers turn-key global distribution across any number of Azure regions by transparently scaling and replicating your data wherever your users are. You can elastically scale throughput and storage worldwide and pay only for the throughput and storage you need. Cosmos DB guarantees single-digit millisecond latencies at the 99th percentile anywhere in the world, offers multiple well-defined consistency models to fine-tune for performance and guaranteed high availability with multi-homing capabilities - all backed by industry leading service level agreements (SLAs).

Cosmos DB is truly schema-agnostic - it automatically indexes all the data without requiring you to deal with schema and index management. Cosmos DB is multi-model - it natively supports document, key-value, graph and columnar data models. With Cosmos DB, you can access your data using NoSQL APIs of your choice -- DocumentDB SQL (document), MongoDB (document), Azure Table Storage (key-value), and Gremlin (graph), are all natively supported. Cosmos DB is a fully managed, enterprise ready and trustworthy service. All your data is fully and transparently encrypted and secure by default. Cosmos DB is ISO, FedRAMP, EU, HIPAA, and PCI compliant as well.

[Guardar para más adelante](#)

Crear

La creación del servicio de **Azure Cosmos DB** tiene cuatro *tabs* de configuración: Basics, Network y Tags.

Para la sección de **Basics**, se deben configurara los siguientes campos:

Subscripción. Seleccionar la subscripción correspondiente.

Resource Group: Seleccionar del combo la opción correspondiente al grupo de recursos creado al inicio de esta sección del laboratorio.



Instance details

Account name: Es nombre único de la cuenta que será expuesto en una URL bajo el sufijo [nombre]documents.azure.com

API: Para este laboratorio seleccionar la opción de Core (SQL).

Location: La más cercana a la ubicación de los usuarios y/o la ubicación de los servicios conectados.

Geo-Redundancy: Habilitada.

Multi-region Writes: Deshabilitada.

Por el momento no será necesario configurar las opciones restantes.

Hacer clic en el botón de **Review + create**.

Panel > Grupos de recursos > ExpertsAcademyGroup > Todo > Azure Cosmos DB > Create Azure Cosmos DB Account

Create Azure Cosmos DB Account

Basics Network Tags Summary

Azure Cosmos DB is a fully managed globally distributed, multi-model database service, transparently replicating your data across any number of Azure regions. You can elastically scale throughput and storage, and take advantage of fast, single-digit-millisecond data access using your favorite API among SQL, MongoDB, Apache Cassandra, Tables, or Gremlin, backed by 99.999 SLA. [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

* Resource Group

ExpertsAcademyGroup

[Crear nuevo](#)

INSTANCE DETAILS

* Account Name

expertacademy2018

documents.azure.com

* API ⓘ

Core (SQL)

* Location

Centro-Sur de EE. UU.

Geo-Redundancy ⓘ

Enable

Disable

Multi-region Writes ⓘ

Enable

Disable

Review + create

Previous

Next: Network

Una vez validadas las configuraciones, hacer clic en el botón **Create**.

[Panel](#) > [Grupos de recursos](#) > [ExpertsAcademyGroup](#) > [Todo](#) > [Azure Cosmos DB](#) > [Create Azure Cosmos DB Account](#)

Create Azure Cosmos DB Account

✓ Validation Success

[Basics](#) [Network](#) [Tags](#) [Summary](#)

BASICS

Subscription

Resource Group

ExpertsAcademyGroup

Location

Centro-Sur de EE. UU.

Account Name

(new) expertacademy2018

API

Core (SQL)

Geo-Redundancy

Enable

Multi-region Writes

Disable

Create

Previous

Next

[Download a template for automation](#)

11

Se desplegará una hoja del detalle del despliegue, el cual puede tardar hasta diez minutos.

Panel > Microsoft.Azure.CosmosDB-20181205203210 - Información general

Microsoft.Azure.CosmosDB-20181205203210 - Información general

Implementación

Buscar (Ctrl+/)

Eliminar Cancelar Volver a implementar Actualizar

Información general

Entradas

Salidas

Plantilla

... La implementación está en curso

Compruebe el estado de la implementación, administre los recursos o solucione problemas de implementación encontrándola fácilmente la próxima vez que quiera usarla.

Nombre de implementación: Microsoft.Azure.CosmosDB-20181205203210

Suscripción:

Grupo de recursos: ExpertsAcademyGroup

DETALLES DE IMPLEMENTACIÓN (Descargar)

Hora de inicio: 5/12/2018 20:55:13

Duración: 2 minutos 15 segundos

Id. de correlación: 3617ae4f-4a43-4c29-a413-38f9d624f8f0

RECURSO	TIPO	ESTADO
expertacademy2018	Microsoft.DocumentDb/database...	OK

Notificaciones

Más eventos en el registro de actividad → Descartar todo --

*** Implementación en curso... En ejecución ×

Se está realizando la implementación en el grupo de recursos 'ExpertsAcademyGroup'.

a minute ago



Has completado de forma exitosa esta sección del laboratorio.

[Panel](#) > [Microsoft.Azure.CosmosDB-20181205210249](#) - Información general

Microsoft.Azure.CosmosDB-20181205210249 - Información general
Implementación

[Eliminar](#) [Cancelar](#) [Volver a implementar](#) [Actualizar](#)

Información general

Entradas

Salidas

Plantilla

✓ Se completó la implementación

[Ir al recurso](#)

Nombre de implementación: Microsoft.Azure.CosmosDB-20181205210249
Suscripción: [Visual Studio Enterprise – MPN \(4cefb6fc-47e8-4e28-86bf-a3380102a4d6\)](#)
Grupo de recursos: [ExpertsAcademyGroup](#)

DETALLES DE IMPLEMENTACIÓN [\(Descargar\)](#)
Hora de inicio: 5/12/2018 22:32:49
Duración: 1 minuto 47 segundos
Id. de correlación: 4bc41bcf-989d-42a0-9b8f-61b677b60747

RECURSO	TIPO	ESTADO	DETALLES DE LA OPERACIÓN
✓ expertacademy2018	Microsoft.DocumentDb/d...	OK	Detalles de la operación

Ejecución de consultas a Azure Cosmos DB empleando el API de SQL

Esta sección del laboratorio realizarás consultas sobre una base de datos Cosmos DB utilizando el lenguaje SQL. Utilizaras algunas características del lenguaje SQL como proyecciones y filtros. De igual forma realizarás la ejecución de consultas a través de características únicas del API de SQL de Azure Cosmos DB como proyección sobre JSON, JOINS dentro de documentos y filtrado por rangos de partición.

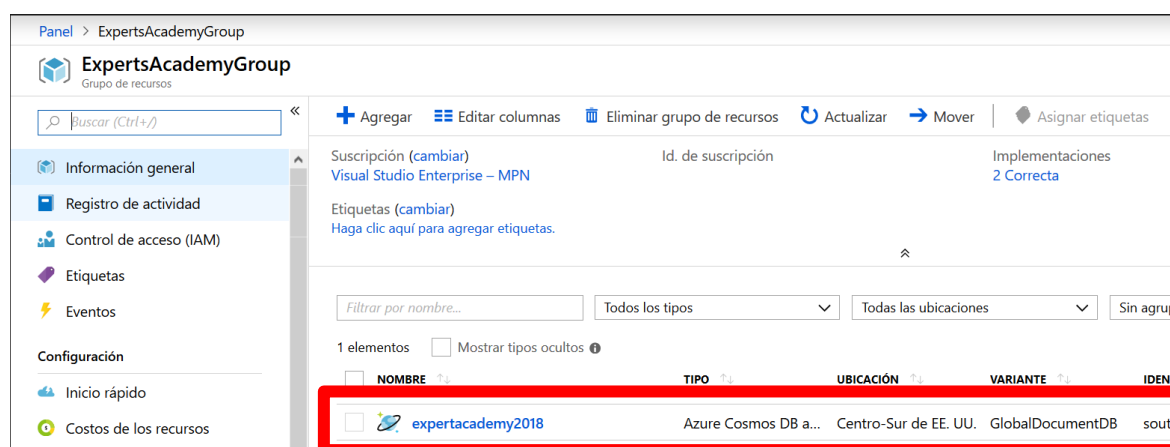
Antes de iniciar es necesario contar con una base de datos de Cosmos DB y una colección que estaremos utilizando para completar este laboratorio.

Tip: Visual Studio Code es un editor de texto que facilita la administración de los APIs disponibles en Cosmos DB (<https://code.visualstudio.com/download>).

Tip: Visual Studio Code corre en Linux, Windows y MacOS.

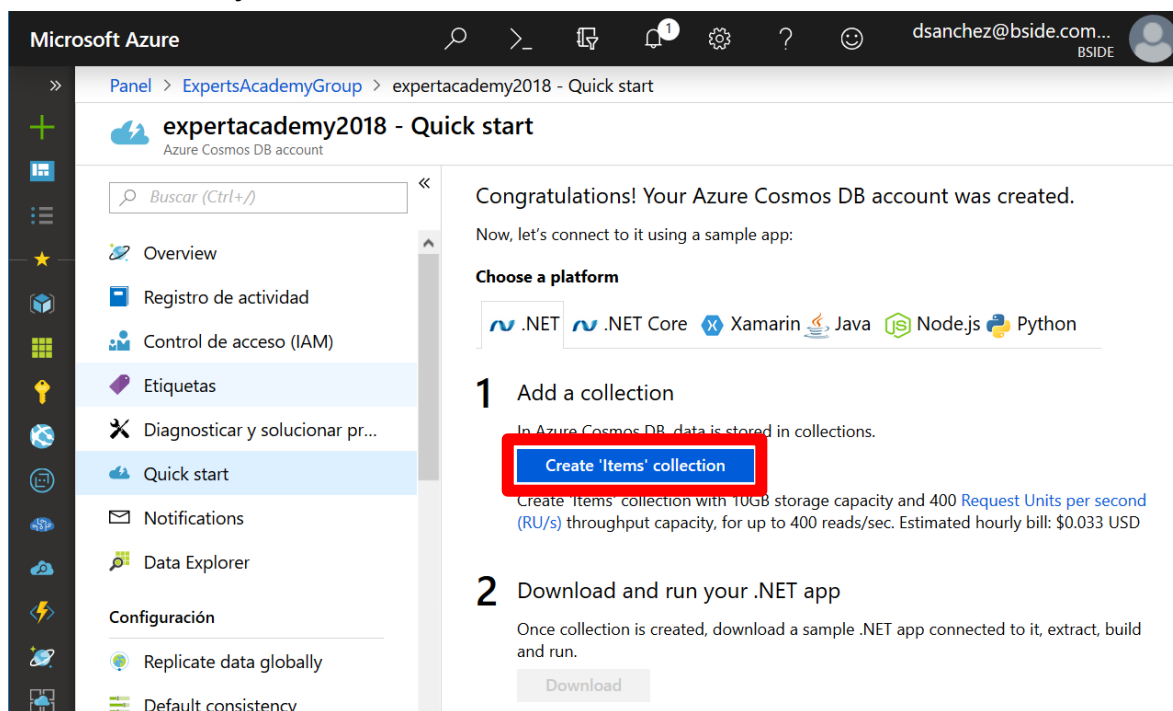
Creación de una colección de Azure Cosmos DB

1. Seleccionar la base de datos de Cosmos DB.



The screenshot shows the Azure portal interface for the 'ExpertsAcademyGroup' resource. The left sidebar contains navigation options like 'Información general', 'Registro de actividad', 'Control de acceso (IAM)', 'Etiquetas', 'Eventos', and 'Configuración'. The main area displays the 'Información general' tab, which includes details about the subscription (Visual Studio Enterprise - MPN) and the 'expertacademy2018' database. The database is highlighted with a red box, and its details (Azure Cosmos DB a..., Centro-Sur de EE. UU., GlobalDocumentDB, south) are visible.

- Al abrir por primera vez el recurso de Cosmos DB se muestra la hoja de inicio rápido, en esta hoja se hace clic en el botón **Create 'Items' collection**.



Microsoft Azure

Panel > ExpertsAcademyGroup > expertacademy2018 - Quick start

expertacademy2018 - Quick start

Azure Cosmos DB account

Buscar (Ctrl+J)

- Overview
- Registro de actividad
- Control de acceso (IAM)
- Etiquetas
- Diagnosticar y solucionar pr...
- Quick start**
- Notifications
- Data Explorer

Configuración

- Replicate data globally
- Default consistency

Congratulations! Your Azure Cosmos DB account was created.

Now, let's connect to it using a sample app:

Choose a platform

.NET .NET Core Xamarin Java Node.js Python

1 Add a collection

In Azure Cosmos DB, data is stored in collections.

Create 'Items' collection

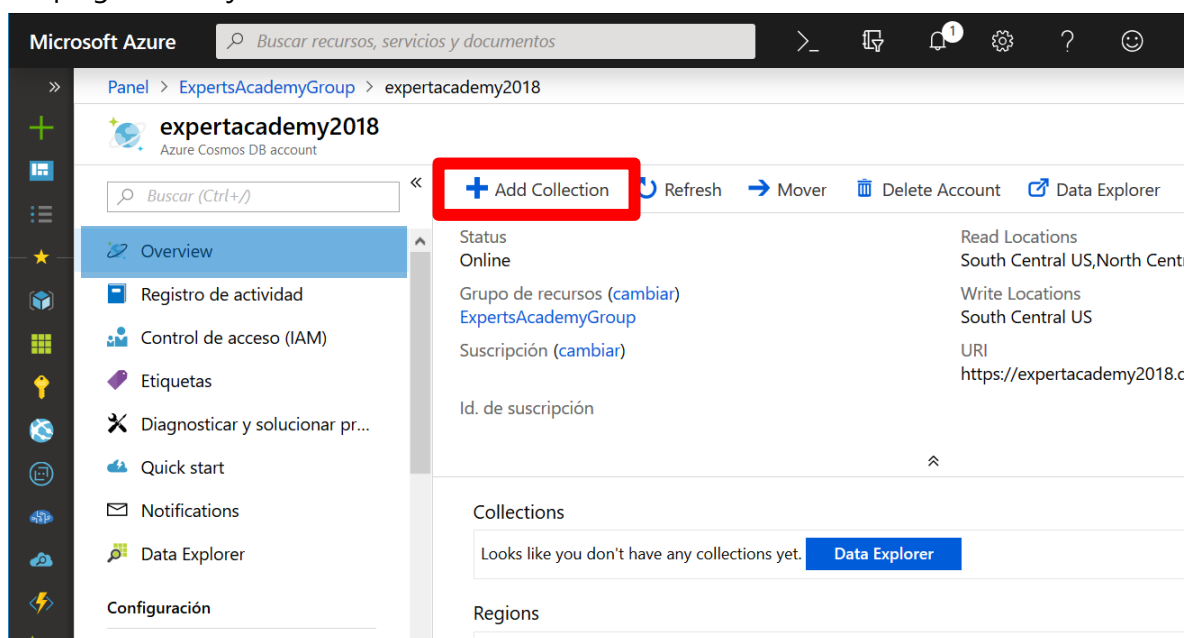
Create 'Items' collection with 10GB storage capacity and 400 Request Units per second (RU/s) throughput capacity, for up to 400 reads/sec. Estimated hourly bill: \$0.033 USD

2 Download and run your .NET app

Once collection is created, download a sample .NET app connected to it, extract, build and run.

Download

Otra opción para la creación de una colección es acceder a la sección de **Overview**, una vez desplegada la hoja de esta sección, hacer clic en el botón de **Add Collection**.



Microsoft Azure

Buscar recursos, servicios y documentos

Panel > ExpertsAcademyGroup > expertacademy2018

expertacademy2018

Azure Cosmos DB account

Buscar (Ctrl+J)

+ Add Collection Refresh Mover Delete Account Data Explorer

Overview

- Registro de actividad
- Control de acceso (IAM)
- Etiquetas
- Diagnosticar y solucionar pr...
- Quick start
- Notifications
- Data Explorer

Configuración

Status: Online

Read Locations: South Central US, North Cent

Write Locations: South Central US

URI: https://expertacademy2018.c

Grupo de recursos (cambiar): ExpertsAcademyGroup

Suscripción (cambiar):

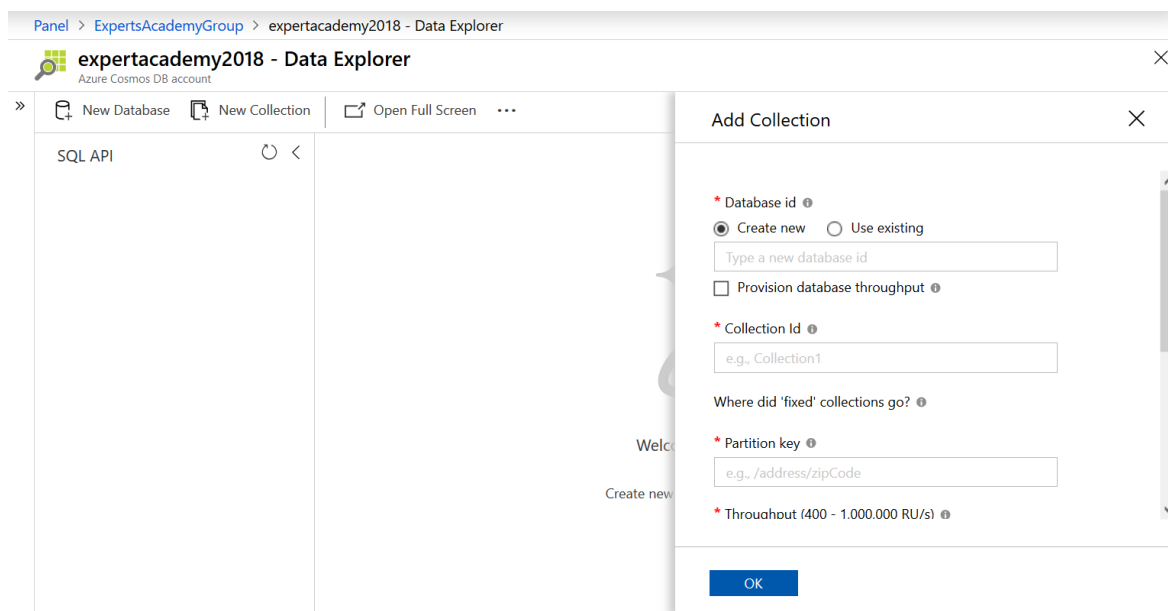
Id. de suscripción:

Collections

Looks like you don't have any collections yet. **Data Explorer**

Regions

- Se abrirá una hoja nueva con una ventana del lado derecho con el título **Add Collection**.



Para este laboratorio rellenar los campos de la siguiente manera:

Database id: Seleccionar la opción **Create new** y colocar el nombre **bdUniversidad**.

Provision database throughput: Dejar este campo sin seleccionar.

Collection id: Escribir el valor **ColeccionDeEstudiantes**.

Partition key: Escribir el valor **/enrollmentYear**.

Throughput: Escribir el valor de **400**.

Unique keys: Seleccionar la opción de **+ Add Unique Key**. Esta opción despliega campo de entrada de texto, escribir **/studentAlias** en este último.

Una vez rellenados los campos necesarios hacer clic en el botón **OK**.

Por defecto todas las colecciones creadas en Cosmos DB tienen la característica de tener capacidad de almacenamiento ilimitada.

A partir del 2018 el costo base de consumo de una instancia base del servicio de Cosmos DB es de \$ 24.00 USD.

Add Collection

* Database id ⓘ

☒ Create new ☐ Use existing

bdUniversidad

☐ Provision database throughput ⓘ

* Collection Id ⓘ

ColeccionDeEstudiantes

Where did 'fixed' collections go? ⓘ

* Partition key ⓘ

/anioDeInscripcion

* Throughput (400 - 1,000,000 RU/s) ⓘ

OK

Add Collection

* Partition key ⓘ

/anioDeInscripcion

* Throughput (400 - 1,000,000 RU/s) ⓘ

11000

–

+

Estimated spend (USD): \$0.88 hourly / \$21.12 daily.

Unique keys ⓘ

/aliasEstudiante ⓘ

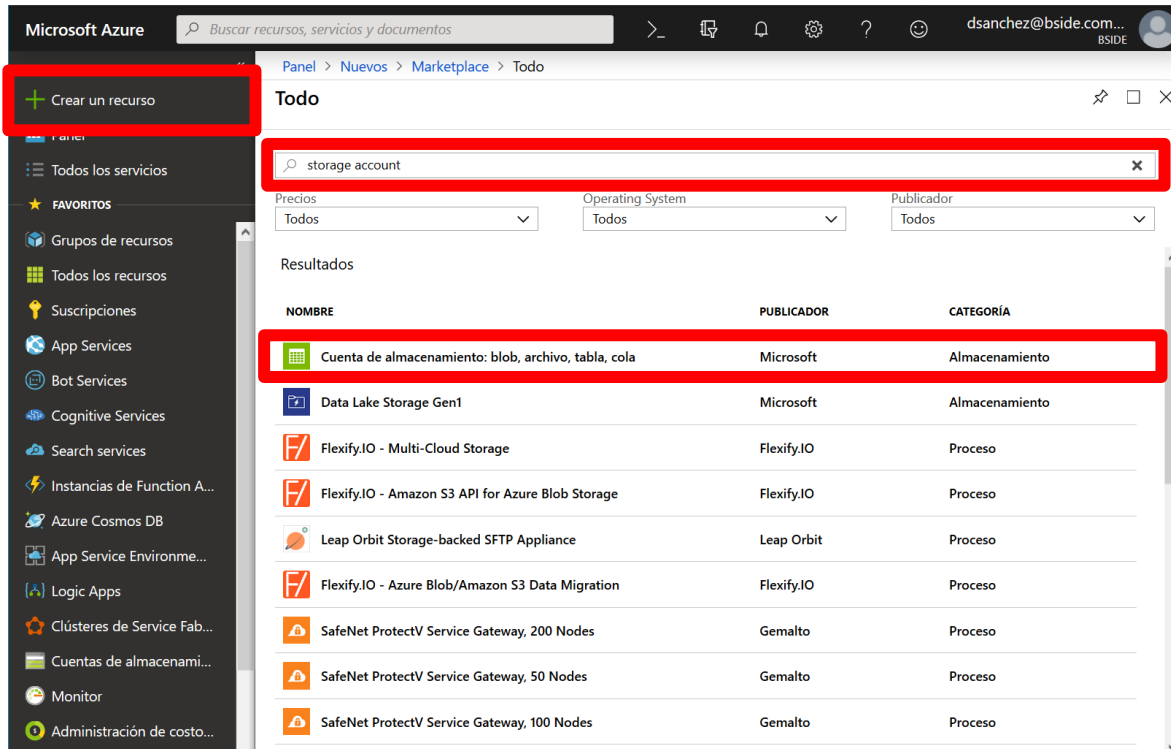
🗑

+ Add unique key

OK

Carga del JSON de base de datos a un blob de Storage Account




1. Hacer clic en el botón de **+ Crear un recurso**.
2. Buscar **storage account** en la barra de búsqueda de recursos.
3. Seleccionar **Cuenta de almacenamiento: blob, archivo, tabla, cola** del panel de resultados inferior.



The screenshot shows the Microsoft Azure portal interface. On the left sidebar, the '+ Crear un recurso' button is highlighted with a red box. The main area shows a search for 'storage account' in the Marketplace. The search results table is also highlighted with a red box, showing the following data:


NOMBRE	PUBLICADOR	CATEGORÍA
Cuenta de almacenamiento: blob, archivo, tabla, cola	Microsoft	Almacenamiento
Data Lake Storage Gen1	Microsoft	Almacenamiento
Flexify.IO - Multi-Cloud Storage	Flexify.IO	Proceso
Flexify.IO - Amazon S3 API for Azure Blob Storage	Flexify.IO	Proceso
Leap Orbit Storage-backed SFTP Appliance	Leap Orbit	Proceso
Flexify.IO - Azure Blob/Amazon S3 Data Migration	Flexify.IO	Proceso
SafeNet ProtectV Service Gateway, 200 Nodes	Gemalto	Proceso
SafeNet ProtectV Service Gateway, 50 Nodes	Gemalto	Proceso
SafeNet ProtectV Service Gateway, 100 Nodes	Gemalto	Proceso

4. Hacer clic en el botón **Crear**.

Cuenta de almacenamiento: blob, archivo, tabla, cola   

Microsoft

Microsoft Azure proporciona soluciones escalables y duraderas de almacenamiento en nube, de copia de seguridad y de recuperación para cualquier tipo de datos, grandes o pequeños. Funciona con la infraestructura de la que usted ya dispone para mejorar la rentabilidad de las aplicaciones existentes y la estrategia de continuidad del negocio. Además proporciona el almacenamiento que necesitan las aplicaciones en la nube, lo que incluye texto no estructurado o datos binarios, tales como vídeo, audio e imágenes.

 [Guardar para más adelante](#)

PUBLICADOR	Microsoft
VÍNCULOS ÚTILES	Documentación Información general del servicio Precio

Crear

5. Completar los campos con la siguiente información:

Suscripción: Seleccionar la suscripción en donde se encuentra la base de datos de Azure Cosmos DB.

Grupo de recursos: Seleccionar el grupo de recursos creado en la primera sección de este laboratorio.

Nombre de la cuenta de almacenamiento: Escribir **datosuniversidad**.

Ubicación: Seleccionar **Centro-Sur de EE. UU.**

Rendimiento: Seleccionar la opción **Estándar**.

Tipo de cuenta: Seleccionar **StorageV2 (uso general v2)**.

Replicación: Seleccionar **Almacenamiento con redundancia local (LRS)**.

Nivel de acceso: Seleccionar la opción **Frecuente**.

Una vez rellenados los campos, hacer clic en el botón **Revisar y crear**.

Crear cuenta de almacenamiento

[Datos básicos](#) [Opciones avanzadas](#) [Etiquetas](#) [Revisar y crear](#)

Azure Storage es un servicio administrado por Microsoft que proporciona almacenamiento en la nube altamente disponible, seguro, duradero, escalable y redundante. Azure Storage incluye Azure Blob (objetos), Azure Data Lake Storage Gen2, Azure Files, Azure Queues y Azure Tables. El costo de una cuenta de Storage depende del uso y de las opciones que elija a continuación. [Más información](#)

DETALLES DEL PROYECTO

Seleccione la suscripción para administrar recursos implementados y los costes. Use los grupos de recursos como carpetas para organizar y administrar todos los recursos.

* Suscripción

* Grupo de recursos [Crear nuevo](#)

DETALLES DE INSTANCIA

El modelo de implementación predeterminado es el de Resource Manager, que admite las últimas características de Azure. Como alternativa, puede elegir el modelo de implementación clásica. [Elegir el modelo de implementación clásica](#)

* Nombre de la cuenta de almacenamiento ⓘ ✓

* Ubicación ▼

Rendimiento ⓘ ☒ Estándar ☐ Premium

Tipo de cuenta ⓘ ▼

Replicación ⓘ ▼

Nivel de acceso (predeterminado) ⓘ ☐ Esporádico ☒ Frecuente

Revisar y crear

Anterior

Siguiente: Opciones avanzadas >

6. Una vez validados los campos hacer clic en el botón **Crear** de la parte inferior de la hoja.

Crear cuenta de almacenamiento

✓ Validación superada

[Datos básicos](#) [Opciones avanzadas](#) [Etiquetas](#) [Revisar y crear](#)

DATOS BÁSICOS

Suscripción	
Grupo de recursos	ExpertsAcademyGroup
Ubicación	Centro-Sur de EE. UU.
Nombre de la cuenta de almacenamiento	datosuniversidad
Modelo de implementación	Resource Manager
Tipo de cuenta	StorageV2 (uso general v2)
Replicación	Almacenamiento con redundancia local (LRS)
Rendimiento	Estándar
Nivel de acceso (predeterminado)	Frecuente

OPCIONES AVANZADAS

Se requiere transferencia segura	Habilitado
Permitir acceso desde	Todas las redes
Espacio de nombres jerárquico	Deshabilitado

Crear

Anterior

Siguiente

[Descargar una plantilla para la automatización](#)

7. Una vez creada la cuenta de almacenamiento, acceder al recurso.

Microsoft.StorageAccount-20181206002447 - Información general

Implementación

[Eliminar](#)
[Cancelar](#)
[Volver a implementar](#)
[Actualizar](#)

Información general


Entradas

Salidas

Plantilla

✓ **Se completó la implementación**

Ir al recurso

 Nombre de implementación: Microsoft.StorageAccount-20181206002447


Grupo de recursos: [ExpertsAcademyGroup](#)

DETALLES DE IMPLEMENTACIÓN [\(Descargar\)](#)

Hora de inicio: 6/12/2018 0:55:24
 Duración: 32 segundos
 Id. de correlación: f47c1fcc-530c-455a-8545-98c6fd5a05ab

RECURSO	TIPO	ESTADO	DETALLES DE L...
✓ datosuniversidad	Microsoft.Stora...	OK	Detalles de la ope...

8. Dentro de la cuenta de almacenamiento, seleccionar la opción **Blobs** dentro de la lista de **Servicios** en la sección de **Información general**.

 **datosuniversidad**
 Cuenta de almacenamiento

[Abrir en el Explorador](#)
[Mover](#)
[Eliminar](#)
[Actualizar](#)

Información general

Registro de actividad

Control de acceso (IAM)

Etiquetas

Diagnosticar y solucionar pr...

Eventos

Explorador de Storage (versi...

Configuración

Claves de acceso

CORS

Configuración

Cifrado

Firma de acceso compartido

Firewalls y redes virtuales

Protección contra amenazas...

Suscripción ([cambiar](#))

Visual Studio Enterprise – MPN

Id. de suscripción
4cefb6fc-47e8-4e28-86bf-a3380102a4d6

Etiquetas ([cambiar](#))

[Haga clic aquí para agregar etiquetas.](#)

Servicios

Blobs

Almacenamiento de objetos basado en REST para datos no estructurados

[Explorar datos con la versión preliminar de Azure AD](#)

[Más información](#)

Archivos

Recursos compartidos de archivos que usan el protocolo SMB 3.0 estándar

[Más información](#)

Tablas

Almacenamiento de datos tabulares

[Más información](#)

Colas

Escalar eficazmente aplicaciones según el tráfico

[Explorar datos con la versión preliminar de Azure AD](#)

[Más información](#)

9. Seleccionar la opción **+ Contenedor** en la barra de herramientas superior.

10. En la ventana de configuración del contenedor, rellenar los siguientes campos:

Nombre: Escribir **estudiantes**.

Nivel de acceso público: Blob (acceso de lectura anónimo solo para blobs).

Al terminar de rellenar los campos hacer clic en **Aceptar**.

11. Seleccionar el contenedor de blobs creado.

12. Seleccionar la opción **Cargar** de la barra de herramientas superior del contenedor **estudiantes**.



Ubicación: estudiantes

Buscar blobs por prefijo (distingue mayúsculas c

NOMBRE

No se encontró ningún blob.

13. La selección despliega una ventana en la derecha de la venta del navegador con un botón para la carga de archivos local.







14. Una vez seleccionado hacer clic en **Cargar**.

Este paso requiere haber descargado el archivo **estudiantes.json** del repositorio de GitHub compartido.

La opción **Sobrescribir los archivos si ya existen** se utiliza cuando se suben versiones distintas de archivos con el mismo nombre o en el caso en que se están subiendo varios archivos dejar deseleccionada para evitar subir duplicados.




15. El archivo cargado se presenta de la siguiente forma en el contenedor.

 Cargar
  Actualizar
  Eliminar
  Adquirir concesión
  Interrumpir concesión
  Más

Ubicación: estudiantes

Buscar blobs por prefijo (distingue mayúsculas de minúsculas)

☐ Mostrar blobs eliminados


NOMBRE	MODIFICADO	NIVEL DE...	TIPO DE ...	TAMAÑO	ESTADO ...
 estudiantes.json	6/12/2018 1:15:47 a....	Frecuen...	Blob en ...	4.3 MiB	Disponible ...

Creación de Azure Data Factory para la migración del archivo de JSON a Azure Cosmos DB

1. Seleccionar la opción + **Crear un recurso** de la barra de herramientas de la izquierda.
2. Buscar **Data Factory** en el campo de búsqueda y seleccionar la opción del mismo nombre.

Panel > Nuevos > Marketplace > Todo

Todo

 data factory

Precios

Todos

Operating System

Todos

Resultados

NOMBRE

PUBLICADOR



Data Factory

Microsoft

3. Una vez seleccionado, hacer clic en el botón **Crear**.

Data Factory

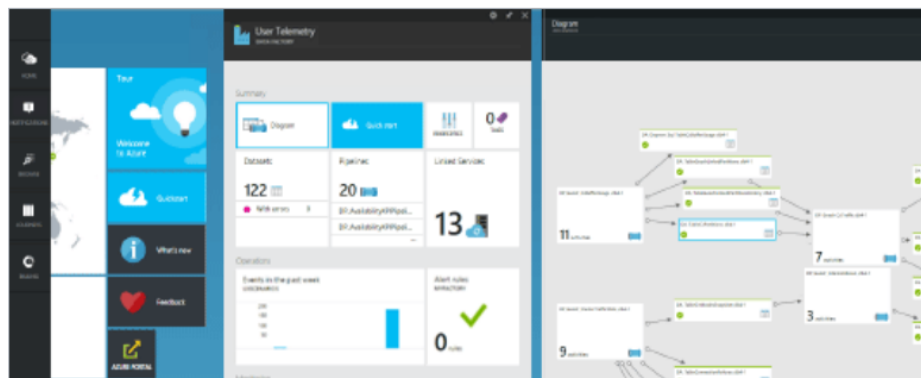
Microsoft



Microsoft Azure Data Factory es un servicio de integración de datos basado en la nube que automatiza el movimiento y la transformación de datos. Puede crear, implementar, programar y supervisar canalizaciones de flujo de datos de alta disponibilidad tolerantes a errores. Mueva y transforme datos de todas las formas y tamaños, y entregue los resultados a un rango de servicios de almacenamiento de destino. Supervise todas las canalizaciones de datos y el estado de servicio de un vistazo con una experiencia visual enriquecida. Consuma fácilmente los datos producidos con BI, herramientas de análisis y otras aplicaciones para controlar de forma confiable la información de la empresa clave y las decisiones.

- Componga servicios de almacenamiento, movimiento y procesamiento de datos en canalizaciones de flujo de datos
- Integración de HDInsight mejorada incluida HCAT y administración de clústeres a petición
- Programe canalizaciones de datos con control optimizado
- Nuevos conectores de datos para orígenes de datos locales y en la nube
- Integración con Microsoft Azure Machine Learning y Azure Batch
- Movimiento de datos implementado globalmente como servicio
- Cree, edite e implemente canalizaciones de datos con un complemento de Visual Studio

 Guardar para más adelante



Crear

Suscripción: Seleccionar la suscripción en donde se encuentran los recursos creados en este laboratorio.

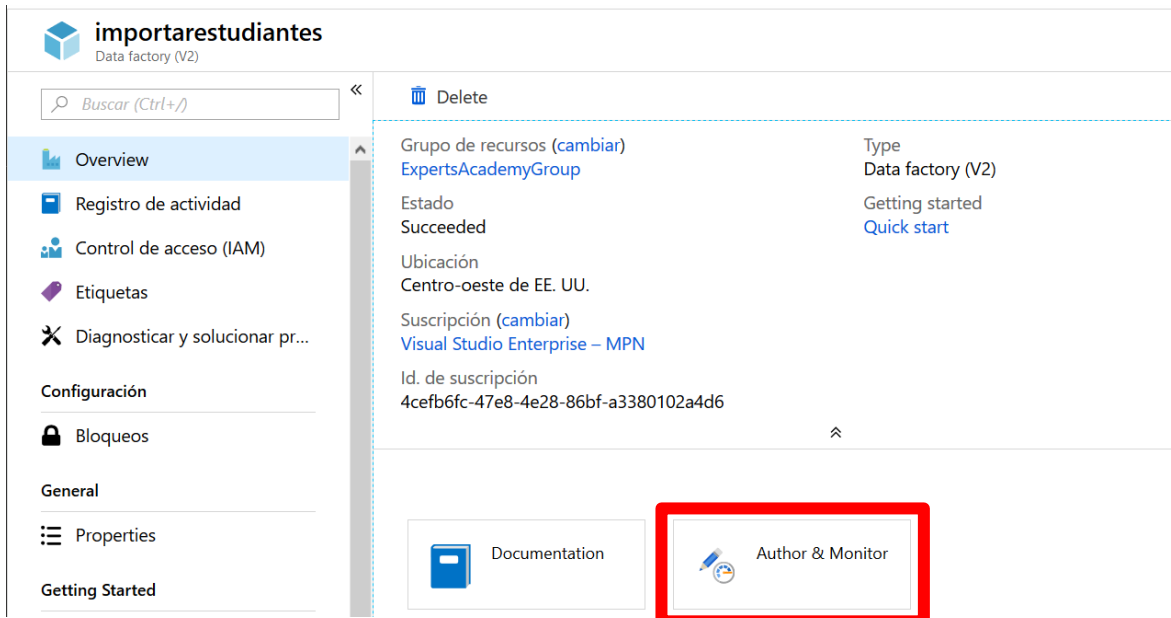
Versión: Seleccionar la versión **V2**.

Ubicación: Seleccionar el centro de datos más cercano a la ubicación geográfica de la cuenta de almacenamiento.

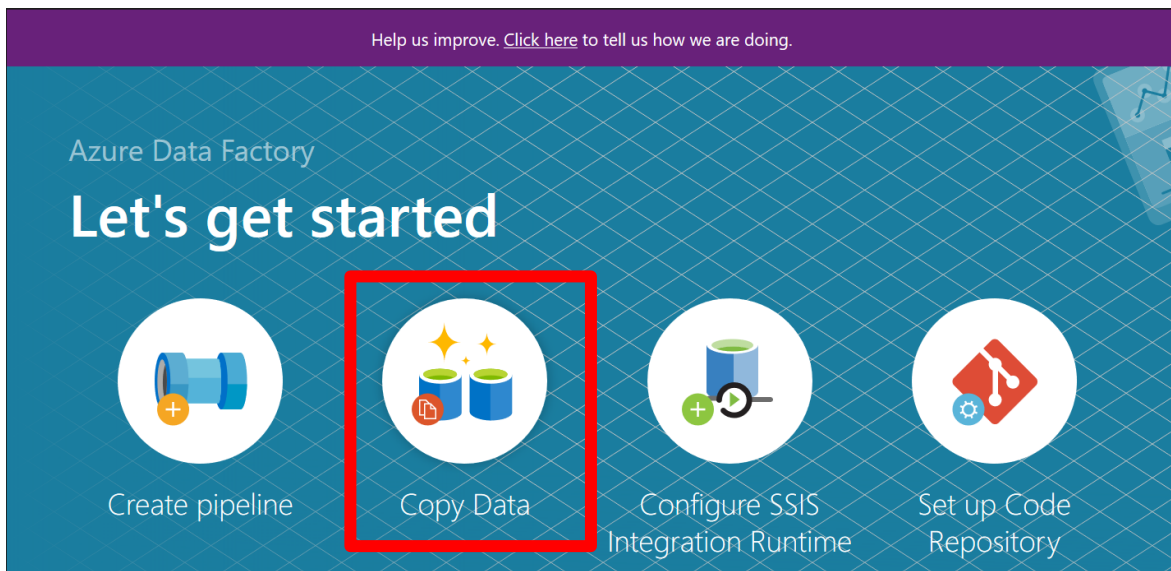
Al finalizar el relleno de los campos, hacer clic en el botón **Crear**.

26

- Una vez creado el servicio de **Azure Data Factory**. Seleccionar la opción de **Author & Monitor** que se encuentra en el panel de **Overview** del servicio.



- Al seleccionar esta opción se abre una nueva pestaña del navegador apuntando a un sitio web del servicio de ADF. En este sitio web seleccionar la opción de **Copy Data**.

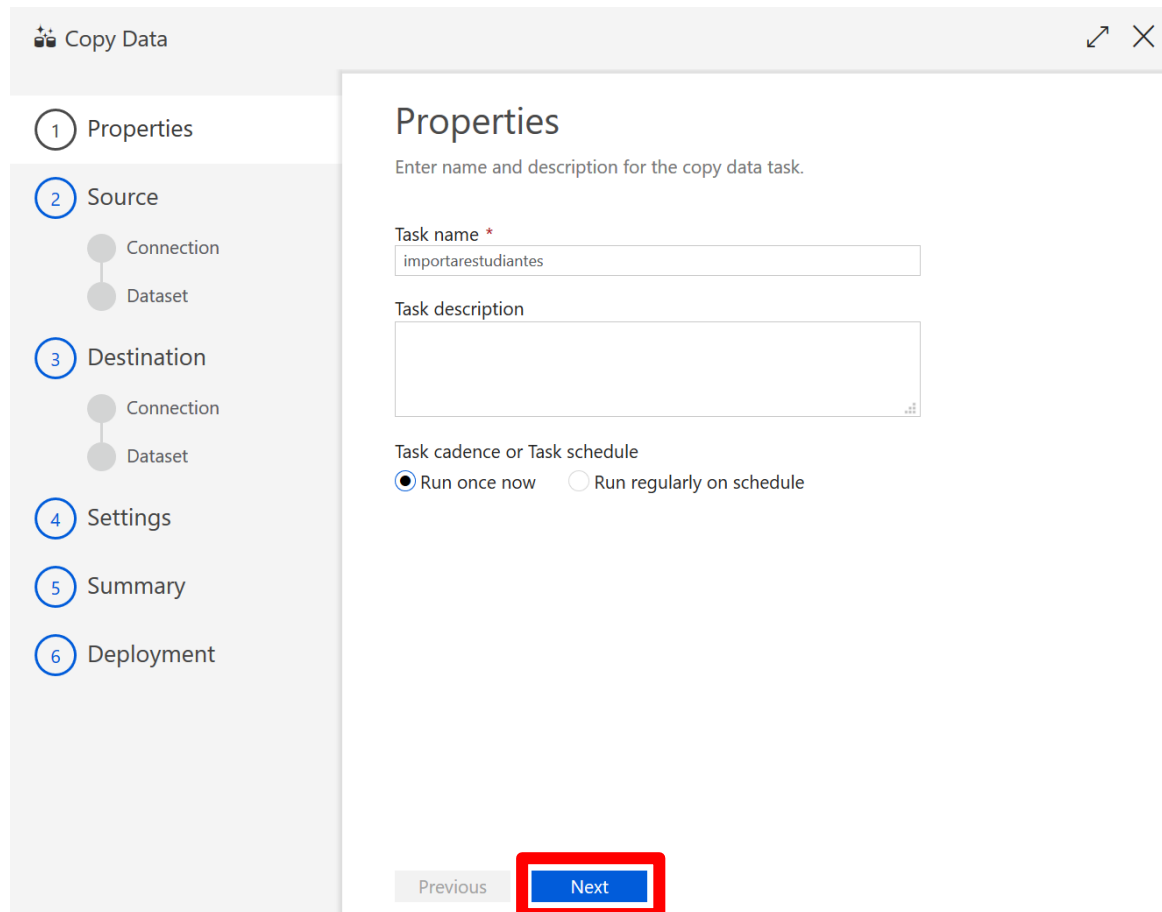


7. Al seleccionar la opción de despliega una ventana modal de configuración para la copia de datos. En la sección de **Properties** completar los campos con la siguiente información:

Task name: Escribir **importarestudiantes**.

Task description: No es un campo requerido para el servicio y el laboratorio.

Task cadence or Task schedule: Seleccionar la opción **Run once now**.



Copy Data

1 Properties

2 Source

- Connection
- Dataset

3 Destination

- Connection
- Dataset

4 Settings

5 Summary

6 Deployment

Properties

Enter name and description for the copy data task.

Task name *
importarestudiantes

Task description

Task cadence or Task schedule

☒ Run once now ☐ Run regularly on schedule

Previous Next

- Una vez completada la configuración de propiedades, hacer clic en **Next**, lo cual nos abre el panel de configuración del origen de datos, en donde se selecciona la opción **+ Create new connection**.

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.

All Azure Database File Generic Protocol NoSQL Services and apps

All ▼

Filter by name

+ Create new connection



No connection to display.

Try changing your filters if you don't see what you're looking for.

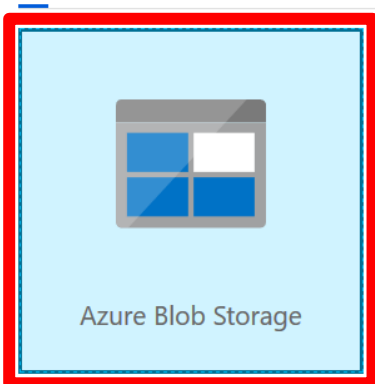
+ Create new connection

- La selección de esta opción despliega una ventana del lado derecho para la selección del tipo del origen de datos, en la barra de búsqueda filtrar por **blob** y seleccionar el servicio conectado **Azure Blob Storage**, una vez seleccionado hacer clic en el botón **Continue**.

New Linked Service



All Azure Database File Generic Protocol NoSQL Services and apps



Cancel

Continue

30

- En la ventana de configuración del servicio conectado, rellenar los campos con la siguiente información:

Name: Escribir **jsonestudiantes**.

Descripción: Saltar este campo.

Connect via integration runtime: Seleccionar **AutoResolveIntegrationRuntime**.


Authentication method: Seleccionar **Use SAS URI**.

SAS URL: Se obtiene ingresando al recurso de **cuenta de almacenamiento**, después al contenedor **estudiantes**, y seleccionando la opción **Generar SAS** en las opciones desplegadas al hacer clic en los tres puntos en el extremo derecho del archivo **estudiantes.json**.

Seleccionar de la barra de herramientas de la cuenta de almacenamiento la opción **Firma de acceso compartido**.

Seleccionar un rango de fechas en el cual será válida la firma de acceso, una vez estipulado el rango de tiempo accesible, hacer clic en **Generar URL y token de SAS de blob**. Para las otras opciones dejar las indicadas por defecto.

Como resultado, el servicio devuelve el **SAS URI** del blob y el **Token de SAS**. Copiar ambos valores en los campos correspondientes de la configuración de la fuente de datos en el servicio de ADF.


datosuniversidad - Firma de acceso compartido
Cuenta de almacenamiento

Información general

Registro de actividad

Control de acceso (IAM)

Etiquetas

Diagnosticar y solucionar pr...

Eventos

Explorador de Storage (versi...

Configuración

Claves de acceso

CORS

Configuración

Cifrado

Firma de acceso compartido

Firewalls y redes virtuales

Protección contra amenazas ...

Sitio web estático (versión pr...

Propiedades

Bloqueos

Script de automatización

Blob service

Blobs

Dominio personalizado

Eliminación temporal

CDN de Azure

Agregar Azure Search

Administración del ciclo de v...

Servicio Archivo

Archivos

Una firma de acceso compartido (SAS) es un identificador URI que concede derechos de acceso limitados a recursos de Azure Storage. Puede proporcionar una firma de acceso compartido a clientes a los que no se les debe confiar la clave de la cuenta de almacenamiento pero en los que desea delegar el acceso a determinados recursos de la cuenta de almacenamiento. Al distribuir un URI de firma de acceso compartido a estos clientes, les concede acceso a un recursos durante un período de tiempo especificado.

Una SAS de nivel de cuenta puede delegar el acceso a varios servicios de almacenamiento (por ejemplo, blob, archivo, cola o tabla). Tenga en cuenta que actualmente no se admiten directivas de acceso almacenadas para una SAS de nivel de cuenta.

[Más información](#)

Servicios permitidos

☒ Blob ☒ Archivo ☒ Cola ☒ Tabla

Tipos de recursos permitidos

☒ Servicio ☒ Contenedor ☒ Objeto

Permisos permitidos

☒ Lectura ☒ Escribir ☒ Eliminar ☒ Lista ☒ Agregar ☒ Crear ☒ Actualizar ☒ Proceso

Fecha y hora de inicio y caducidad

Inicio

2018-12-06 2:56:55

Finalizar

2019-12-06 10:56:55

(UTC-06:00) --- Zona de horaria actual ---

Direcciones IP permitidas

por ejemplo, 168.1.5.65 o 168.1.5.65-168.1.5.70

Protocolos permitidos

☒ HTTPS solamente ☐ HTTPS y HTTP

Clave de firma

key1

Generar la cadena de conexión y SAS

Cadena de conexión

BlobEndpoint=https://datosuniversidad.blob.core.windows.net/QueueEndpoint=https://datosuniversidad.queu ...

Token de SAS

?sv=2018-03-28&ss=bfqt&srt=sco&sp=rwdlacup&se=2019-12-06T16:56:55Z&st=2018-12-06T08:56:55Z&spr=...

URL de SAS de Blob service

https://datosuniversidad.blob.core.windows.net/?sv=2018-03-28&ss=bfqt&srt=sco&sp=rwdlacup&se=2019-12-...

* Para este escenario, en caso necesario, se proveerá acceso compartido al servicio de blob.



Los campos de configuración del servicio conectado en ADF deben corresponder a los de la imagen, una vez rellenados hacer clic en **Test connection** para validar la conexión con el servicio, una vez hecha la prueba de conexión se muestra un icono de paloma color verde y el mensaje **Connection successful**.

Una vez aprobada la conexión, hacer clic en el botón **Finish**.

Edit Linked Service (Azure Blob Storage)

Name *

jsonestudiantes

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Authentication method

Use SAS URI

SAS URI

Azure Key Vault

SAS URL *

https://datosuniversidad.blob.core.windows.net/

SAS Token *

Annotations

+ New

Cancel

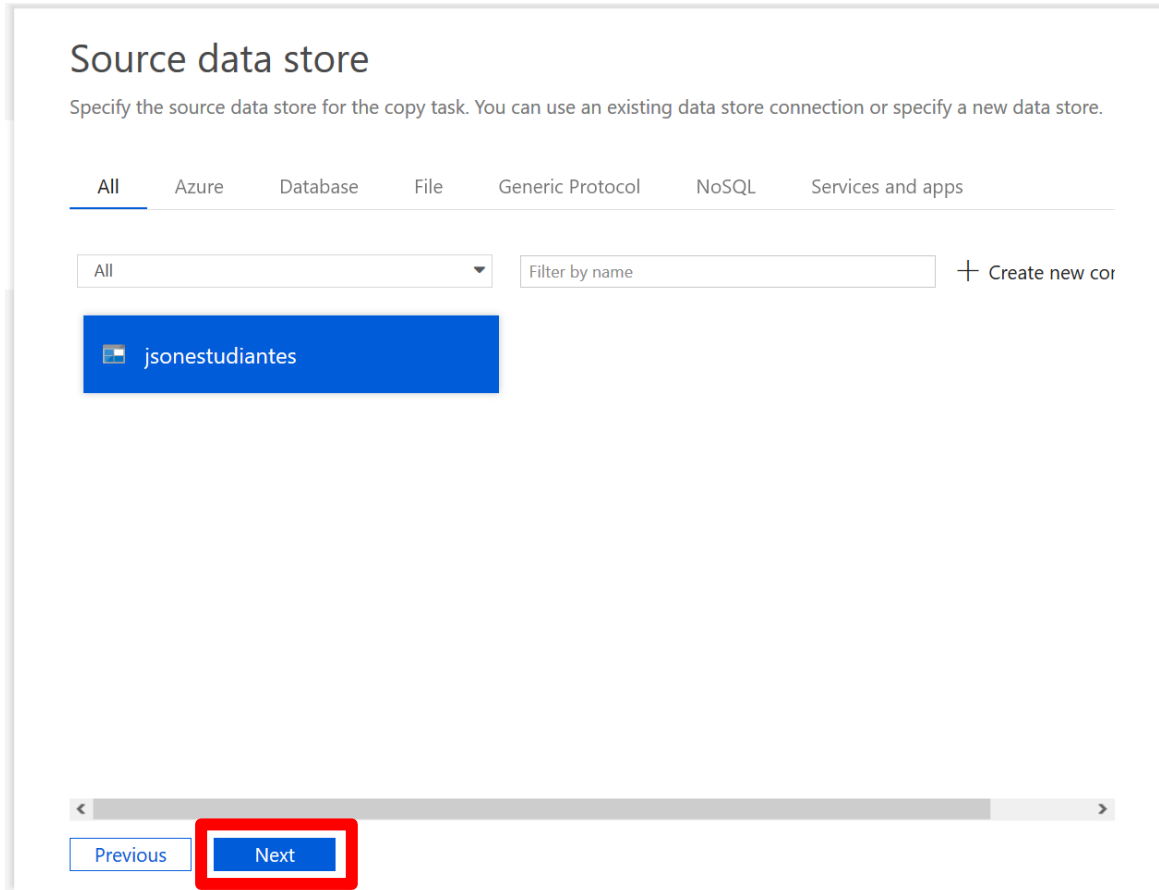
Test connection

Finish

✓ Connection successful

11. El origen de datos configurado, en este caso el blob **estudiantes.json** debe mostrarse como se presenta en la imagen.

Hacer clic en el botón **Next** en la parte inferior del modal de configuración.



12. En la ventana de selección de archivo de entrada, seleccionar el archivo **estudiantes** cargado por defecto en la vista de directorios. Una vez seleccionado, hacer clic en **Choose**, para cargar el archivo y una vez cargado hacer clic en **Next**.

Choose the input file or folder

Select a source folder or file to be copied to the destination data store.

File or folder * [Browse](#)

↑ 📁 >

📁 estudiantes

Cancel **Choose**

Choose the input file or folder

Select a source folder or file to be copied to the destination data store.

File or folder * [Browse](#)

Copy file recursively ☐ ⓘ

Binary Copy ☐ ⓘ

Compression Type

[Previous](#) **Next**

13. En la ventana **File format settings**, ADF detecta por defecto el formato del archivo de entrada, sin embargo, es posible generar la configuración de forma manual en caso contrario. Hacer clic en **Next**.

File format settings

File format

JSON format

File pattern

Array of objects

☐ Export as-is to JSON files or Cosmos DB collection

JSON Path settings

Cross-apply nested JSON array

None

☐ Edit

	COLUMN NAME	JSONPATH EXPRESSION	
≡	firstName	\$.['firstName']	🗑
≡	lastName	\$.['lastName']	🗑
≡	studentAlias	\$.['studentAlias']	🗑
≡	homeEmailAddress	\$.['homeEmailAddress']	🗑
≡	enrollmentYear	\$.['enrollmentYear']	🗑

Preview

Schema

firstName	lastName	studentAlias	homeEmailAddress	enrollmentYear	projectYear
Marcos	Steuber	marcossteuber000000	Marcos36@yahoo.com	2014	2023
Dayana	Pfeffer	dayanapfeffer000002	Dayana.Pfeffer@yahoo.com	2013	2018

Previous

Next

14. En la sección **Destination data store**, hacer clic en **+ Create new connection**.

Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

All
Azure
Database
File
Generic Protocol
NoSQL
Services and apps

Create new connection

15. De la lista de servicio conectados, seleccionar **Azure Cosmos DB** y hacer clic en **Continue**.

New Linked Service

All
Azure
Database
File
Generic Protocol
Services and apps

Azure Blob Storage

Azure Cosmos DB

Azure Data Lake Storage Gen1

Azure Data Lake Storage Gen2 (Preview)

Azure File Storage

Azure SQL Data Warehouse



16. En la ventana de configuración de la conexión con el servicio de Azure Cosmos DB, colocar los siguientes valores en los campos correspondientes:

Name: Escribir **cosmosdb**.

Descripción: Saltar este campo.

Connect via integration runtime: Seleccionar **AutoResolveIntegrationRuntime**.

Seleccionar **Connection String** e ir seleccionando los campos correspondientes en los combos.

Al terminar de rellenar y seleccionar los valores necesarios, hacer clic en **Test connection**.

Una vez validada la conexión, hacer clic en **Finish**.

← New Linked Service (Azure Cosmos DB)

×

Name *

cosmosdb

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Connection String

Azure Key Vault

Account selection method

From Azure subscription

Azure subscription

Cosmos DB account name *

expertacademy2018

Database name *

bdUniversidad

Additional connection properties

+ New

Annotations

+ New

Parameters

Advanced ⓘ

Cancel

Test connection

Finish

✓ Connection successful

17. Seleccionar el servicio de conexión de destino para los datos, es decir, el servicio de **Azure Cosmos DB**. Una vez seleccionado, hacer clic en **Next**.

Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

AllAzureDatabaseFileGeneric ProtocolNoSQLServices and apps

AllFilter by name+ C

cosmosdb

jsonestudiantes

Previous

Next

18. En la sección de **Table mapping** seleccionar el destino **ColeccionDeEstudiantes**, el cual configuramos como colección.

Seleccionar la opción **Skip column mapping for all tables**.

Hacer clic en **Next** para generar el mapa de tablas.

Table mapping

For each table you have selected to copy in the source data store, select a corresponding table in the destination data store or specify the stored procedure to run at the destination.

Source	Destination
Azure Blob Storage file	→ ColeccionDeEstudiantes ▼ ↺ ▼

☒ Skip column mapping for all tables

Previous

Next

19. En la sección de **Column mapping**, dejar los campos configurados por defecto y hacer clic en **Next**.

Column mapping

Choose how source and destination columns are mapped

Table mappings (1)

<input checked="" type="checkbox"/> Source
Azure Blob Storage file
Destination
ColeccionDeEstudiantes

Azure Cosmos DB sink properties

Nesting separator ?

Write behavior

Insert

Write batch size

10000

Previous **Next**

20. La sección **Settings** no requiere ajustes adicionales a los presentes por defecto, hacer clic en el botón **Next**.

Settings

More options for data movement

Fault tolerance settings

Fault tolerance Abort activity on first incompatible row ?

Performance settings

Enable Staging ☐ ?

▶ Advanced settings

Previous **Next**



21. En la sección de **Summary**, hacer clic en **Next**.
22. Una vez completado el despliegue, hacer clic en **Monitor**.

Ejecución de consultas

En esta sección del laboratorio se hará uso de la herramienta Data Explorer que permite la ejecución de consultas a base de datos desde el Portal de Azure, sin la necesidad de contar con acceso a un servidor de base de datos o acceso DBMS instalado en el equipo.

Validación de la importación de datos

1. Acceder a los Grupos de recursos en la barra de herramientas de la izquierda en el Portal de Azure.
2. Acceder al Grupo de recursos creado al inicio del laboratorio.
3. Acceder al recurso de Azure Cosmos DB.
4. Seleccionar la opción **Data Explorer** del panel de opciones del lado izquierdo.
5. Expandir el recurso que se encuentra en la ventana de **SQL API** llamado **dbUniversidad**.
6. Seleccionar la colección **ColeccionDeEstudiantes**.
7. Seleccionar **Documents**.
8. Explorar los datos.

Ejecución de consultas

1. Hacer clic en **New SQL Query** localizada en la barra de herramientas superior.
2. En la tab de consulta, reemplazar el contenido del editor de consultas por la siguiente consulta SQL:

```
SELECT * FROM students s WHERE s.enrollmentYear = 2017
```

Esta primera consulta seleccionará todas las propiedades de todos los documentos de la colección donde los estudiantes se matricularon en 2017. Estamos utilizando el alias s para hacer referencia a la colección.

3. Hacer clic en el botón **Execute Query** para ejecutar la consulta.
4. Debajo se muestra el panel de resultados de la busca, se puede observar que el contenido devuelto por la consulta es un arreglo JSON.
5. En el editor de consultas, reemplazar las consultas conforme se vayan realizando:

```
SELECT * FROM students WHERE students.enrollmentYear = 2017
```

En esta consulta, empleamos el alias s y usamos la fuente Students. Cuando ejecutamos esta consulta, debemos ver los mismos resultados que la consulta anterior.



```
SELECT * FROM arbitraryname WHERE arbitraryname.enrollmentYear = 2017
```

En esta consulta, probaremos que el nombre utilizado para el origen puede ser cualquier nombre seleccionado. Vamos a utilizar el nombre arbitraryname para el origen. Cuando ejecutamos esta consulta, debemos ver los mismos resultados que la consulta anterior.

```
SELECT s.studentAlias FROM students s WHERE s.enrollmentYear = 2017
```

Volviendo a s como un alias, ahora crearemos una consulta donde sólo seleccionaremos la propiedad studentAlias y devolveremos el valor de esa propiedad en nuestro conjunto de resultados.

```
SELECT VALUE s.studentAlias FROM students s WHERE s.enrollmentYear = 2017
```

En algunos escenarios, es posible que tenga que devolver una matriz plana como resultado de la consulta. Esta consulta utiliza la palabra clave VALUE para acoplar la matriz tomando la única propiedad devuelta (String) y creando una matriz de strings.

[Ejecutar implícitamente una consulta de partición cruzada](#)

El explorador de datos le permitirá crear una consulta de particiones cruzadas sin necesidad de realizar configuraciones manuales. Usaremos el editor de consultas en el explorador de datos para realizar consultas de una o varias particiones

De vuelta en la tab de consultas, reemplace el contenido del editor de consultas con la siguiente consulta SQL:

```
SELECT * FROM students s WHERE s.enrollmentYear = 2016
```

Como sabemos que nuestra clave de partición es/enrollmentYear, sabemos que cualquier consulta que se dirige a un único valor válido para la propiedad enrollmentYear será una única consulta de partición.

Observe el cargo de solicitud (en RU/s) para la consulta ejecutada.

```
SELECT * FROM students s
```

Si queremos ejecutar una consulta global que se fan-out a todas las particiones, simplemente podemos dejar nuestra cláusula WHERE que se filtra en un único valor válido para nuestra ruta de la clave de partición.

Observe el cargo de solicitud (en RU/s) para la consulta ejecutada, la carga es relativamente mayor.



```
SELECT * FROM students s WHERE s.enrollmentYear IN (2015, 2016, 2017)
```

Observe el cargo de solicitud (en RU/s) para la consulta ejecutada, la carga es mayor que una sola partición, pero mucho menos que un *fan-out* en todas las particiones.

Utilizar Built-In Functions

Hay una gran variedad de built-in functions disponibles en la sintaxis de SQL Query para la API de SQL en Azure Cosmos DB. Nos centraremos en una sola función en esta tarea, sin embargo, hay más información disponible en la siguiente liga:

<https://docs.microsoft.com/Azure/Cosmos-dB/SQL-API-SQL-Query-Reference>

```
SELECT s.studentAlias FROM students s WHERE s.enrollmentYear = 2015
```

Nuestro objetivo es obtener la dirección de correo electrónico emitida por la escuela para todos los estudiantes que se matricularon en 2015. Podemos emitir una simple consulta para empezar que devolverá el alias de inicio de sesión para cada alumno.

```
SELECT CONCAT(s.studentAlias, '@contoso.edu') AS email FROM students s  
WHERE s.enrollmentYear = 2015
```

Para obtener la dirección de correo electrónico emitida por la escuela, necesitaremos concatenar el string @contoso.edu hasta el final de cada alias. Podemos realizar esta acción utilizando la función CONCAT.

```
SELECT VALUE CONCAT(s.studentAlias, '@contoso.edu') FROM students s WHERE  
s.enrollmentYear = 2015
```

En la mayoría de las aplicaciones de cliente, es probable que sólo necesite una matriz de strings en lugar de una matriz de objetos. Podemos utilizar la palabra clave VALUE aquí para aplanar nuestro conjunto de resultados.

Proyección de los resultados de una consulta

En algunos casos de uso, es posible que necesitemos remodelar la estructura de nuestro conjunto de resultados JSON a una estructura que nuestras bibliotecas o APIs de terceros pueden analizar. Nos centraremos en una sola consulta y volveremos a configurar los



resultados en varios formatos utilizando las funcionalidades JSON nativas en la sintaxis de la consulta SQL.

```
SELECT
    CONCAT(s.firstName, " ", s.lastName),
    s.academicStatus.warning,
    s.academicStatus.suspension,
    s.academicStatus.expulsion,
    s.enrollmentYear,
    s.projectedGraduationYear
FROM students s WHERE s.enrollmentYear = 2014
```

En esta primera consulta, queremos determinar el estado actual de cada estudiante que se inscribió en 2014. Nuestro objetivo aquí es eventualmente tener una visión aplanada, simple de entender de cada estudiante y su estado académico actual.

El valor que representa el nombre del alumno, mediante la función CONCAT, tiene un nombre de propiedad de marcador de marcadores en lugar de una cadena simple.

```
SELECT
    CONCAT(s.firstName, " ", s.lastName) AS name,
    s.academicStatus.warning,
    s.academicStatus.suspension,
    s.academicStatus.expulsion,
    s.enrollmentYear,
    s.projectedGraduationYear
FROM students s WHERE s.enrollmentYear = 2014
```

Actualizaremos nuestra consulta anterior nombrando nuestra propiedad que utiliza una built-in function.

```
SELECT {  
  "name": CONCAT(s.firstName, " ", s.lastName),  
  "isWarned": s.academicStatus.warning,  
  "isSuspended": s.academicStatus.suspension,  
  "isExpelled": s.academicStatus.expulsion,  
  "enrollment": {  
    "start": s.enrollmentYear,  
    "end": s.projectedGraduationYear  
  }  
} AS studentStatus  
FROM students s WHERE s.enrollmentYear = 2014
```

Otra forma alternativa de especificar la estructura de nuestro documento JSON es utilizar las llaves de arreglo de JSON. En este punto, estamos definiendo la estructura del resultado JSON directamente en nuestra consulta.

Se debe notar que nuestro objeto JSON todavía está "envuelto" en otro objeto JSON. Esencialmente, tenemos una matriz del tipo padre con una propiedad denominada `studentStatus` que contiene los datos reales que deseamos.

46

```
SELECT VALUE {  
  "name": CONCAT(s.firstName, " ", s.lastName),  
  "isWarned": s.academicStatus.warning,  
  "isSuspended": s.academicStatus.suspension,  
  "isExpelled": s.academicStatus.expulsion,  
  "enrollment": {  
    "start": s.enrollmentYear,  
    "end": s.projectedGraduationYear  
  }  
}  
FROM students s WHERE s.enrollmentYear = 2014
```

Si queremos "desenvolver" nuestros datos JSON y aplanarlos a una simple matriz de objetos similares, necesitamos usar la palabra clave `VALUE`.

Limpieza de los recursos del laboratorio

* Con la finalidad de hacer buen uso de los recursos disponibles en la suscripción de Azure (Azure Pass) y como buena práctica, se recomienda que una vez concluido el uso de los recursos se haga una limpieza de estos a través de la eliminación en el Portal de Azure, CLI o PowerShell.

1. Seleccionar el Grupo de recursos creado para este laboratorio y seleccionar el botón **Eliminar** en la barra superior de herramientas, se abrirá una ventana a la derecha solicitando el nombre del grupo, escribir el nombre del grupo de recursos para habilitar el botón de **Eliminar**, una vez desbloqueado hacer clic en el mismo.

El proceso de eliminación del grupo de recursos y los recursos contenidos en este puede llevar unos minutos, en las notificaciones se le alertará sobre el proceso.

Conclusiones

En esta documentación técnica se incluye los detalles que se realizaron para cubrir los requerimientos de Microsoft. En esto se incluye:

- Creación de los recursos en Azure.
- Ejecución de consultas utilizando la herramienta de Data Explorer desde el Portal de Azure.
- Exploración de las capacidades en el manejo de dato con Cosmos DB.
- Exploración de las capacidades en cuanto a versatilidad y eficiencia en el acceso a datos a través del API de SQL en Cosmos DB.
- Automatización de flujos de datos.
- Conocer las capacidades de cumplimiento de Cosmos DB como componente de soluciones para ambientes de producción.

Contacto

