

# Componentes de Azure Databricks

En este artículo se presentan los componentes fundamentales que debe comprender para poder usar Azure Databricks de forma eficaz.

## Cuentas y áreas de trabajo

En Azure Databricks, un *área de trabajo* es una implementación de Azure Databricks en la nube que funciona como un entorno para que el equipo acceda a los recursos de Databricks. La organización puede elegir tener múltiples áreas de trabajo o solo una, en función de sus necesidades.

Una *cuenta* de Azure Databricks representa una sola entidad que puede incluir varias áreas de trabajo. Las cuentas habilitadas para [Unity Catalog](#) se pueden usar para administrar usuarios y su acceso a los datos de forma centralizada en todas las áreas de trabajo de la cuenta.

## Facturación: unidades de Databricks (DBU)

Azure Databricks cobra en función de las unidades de Databricks (DBUs), que son unidades de capacidad de procesamiento por hora basadas en el tipo de instancia de máquina virtual (VM).

Consulte la [página de precios de Azure Databricks](#).

## Autenticación y autorización

En esta sección se describen los conceptos que debe conocer cuando administre identidades de Azure Databricks y su acceso a los recursos de Azure Databricks.

### Usuario

Persona única que tiene acceso al sistema. Las identidades de usuario se representan mediante direcciones de correo electrónico. Consulte [Administrar usuarios](#).

### Entidad de servicio

Una identidad de servicio para usarla con trabajos, herramientas automatizadas y sistemas, como scripts, aplicaciones y plataformas de CI/CD. Las entidades de servicio se representan mediante un id. de aplicación. Consulte [entidades de servicio](#).

### Grupo

Colección de identidades. Los grupos simplifican la administración de identidades, lo que facilita la asignación de acceso a áreas de trabajo, datos y otros objetos protegibles. Todas las identidades de Databricks se pueden asignar como miembros de grupos. Consulte [Grupos](#).

### Lista de control de acceso (ACL)

Lista de permisos asociados al área de trabajo, el clúster, el trabajo, la tabla o el experimento. Una ACL especifica a qué usuarios o procesos del sistema se les concede acceso a los objetos, así como qué operaciones se permiten en los recursos. Cada entrada de una ACL típica especifica un sujeto y una operación. Consulte las [Listas de control de acceso](#).

### Un token de acceso personal (PAT)

Un token de acceso personal es una cadena que se usa para autenticar llamadas a la API REST, [socios tecnológicos](#) y conexiones con otras herramientas. Consulte [Autenticación con tokens de acceso personal \(heredado\) de Azure Databricks](#).

También se pueden usar tokens de Microsoft Entra ID para autenticarse en la API de REST.

## Interfaces de Azure Databricks

En esta sección se describen las interfaces para acceder a los recursos en Azure Databricks.

### Interfaz de usuario

La interfaz de usuario de Azure Databricks es una interfaz gráfica para interactuar con distintas características, como carpetas del área de trabajo y objetos que contienen, objetos de datos y recursos de cálculo.

### REST API

La API de REST de Databricks proporciona puntos de conexión para modificar o solicitar información sobre objetos de área de trabajo y cuenta de Azure Databricks. Consulte [cuenta de referencia](#) y [área de trabajo](#).

### API de REST SQL

La API de REST de SQL le permite automatizar tareas en objetos SQL. Consulte [API DE SQL](#).

### Interfaz de línea de comandos (CLI)

La [CLI de Databricks](#) se hospeda en [GitHub](#). La CLI se basa en la API de REST de Databricks.

## Administración de datos

En esta sección se describen las herramientas y los objetos lógicos que se usan para organizar y controlar los datos en Azure Databricks. Consulte [Objetos de base de datos en Azure Databricks](#)

### Catálogo de Unity

Unity Catalog es una solución de gobernanza unificada para los recursos de inteligencia artificial y datos en Azure Databricks que proporciona funcionalidades centralizadas de control de acceso, auditoría, linaje y detección de datos en áreas de trabajo de Databricks. Consulte [¿Qué es Unity Catalog?](#)

### Catálogo

Los catálogos son el contenedor de nivel más alto para organizar e aislar datos en Azure Databricks. Puede compartir catálogos entre áreas de trabajo dentro de la misma región y cuenta. Vea [¿Qué son los catálogos en Azure Databricks?](#).

### Esquema

Los esquemas, también conocidos como bases de datos, se incluyen en catálogos y proporcionan un nivel más granular de organización. Contienen objetos de base de datos y recursos de IA, como volúmenes, tablas, funciones y modelos. Consulte [¿Qué son los esquemas en Azure Databricks?](#).

### Tabla

Las tablas organizan y rigen el acceso a los datos estructurados. Las tablas se consultan con Apache Spark SQL y las API de Apache Spark. Consulte [Tablas de Azure Databricks](#).

### Ver

Una vista es un objeto de solo lectura derivado de una o varias tablas y vistas. Las vistas guardan consultas que se definen con relación a las tablas. Consulte [¿Qué es una vista?](#).

## Volumen

Los volúmenes representan un volumen lógico de almacenamiento en una ubicación de almacenamiento de objetos en la nube y organizan y rigen el acceso a datos no tabulares. Databricks recomienda usar volúmenes para administrar todo el acceso a datos no tabulares en el almacenamiento de objetos en la nube. Consulte [¿Qué son los volúmenes de Unity Catalog?](#).

### Tablas delta

De forma predeterminada, todas las tablas creadas en Azure Databricks son tablas delta. Las tablas delta se basan en el [proyecto de código abierto Delta Lake](#) <sup>↗</sup>, un marco para el almacenamiento de tablas ACID de alto rendimiento a través de almacenes de objetos en la nube. Una tabla Delta almacena datos como un directorio de archivos en el almacenamiento de objetos en la nube y registra metadatos de tabla en el metastore dentro de un catálogo y esquema.

### Tienda de metadatos


Unity Catalog proporciona un metastore de nivel de cuenta que registra metadatos sobre datos, inteligencia artificial y permisos sobre catálogos, esquemas y tablas. Consulte [Metastore](#).

Azure Databricks proporciona un metastore de Hive heredado para los clientes que no han adoptado el catálogo de Unity. Consulte [Control de acceso a la tabla de metastore de Hive \(heredado\)](#).

### Explorador de catálogo

El Explorador de catálogos permite explorar y administrar datos y recursos de IA, incluidos esquemas (bases de datos), tablas, modelos, volúmenes (datos no tabulares), funciones y modelos de ML registrados. Puede usarlo para buscar objetos de datos y propietarios, comprender las relaciones de datos entre tablas y administrar permisos y uso compartido. Consulte [¿Qué es el Explorador de catálogos?](#).

### Raíz de DBFS

 **Importante**

El almacenamiento y el acceso a datos mediante la raíz de DBFS o montajes de DBFS es un patrón en desuso y no recomendado por Databricks. En su lugar, Databricks recomienda usar el catálogo de Unity para administrar el acceso a todos los datos. Consulte [¿Qué es Unity Catalog?](#)

La raíz de DBFS es una ubicación de almacenamiento disponible para todos los usuarios de forma predeterminada. Consulte [¿Qué es DBFS?](#)

## Administración de cálculos

En esta sección se describen los conceptos que debe conocer para ejecutar cálculos en Azure Databricks.

### Clúster

Conjunto de recursos de cálculo y configuraciones en el que se ejecutan cuadernos y trabajos. Hay dos tipos de clústeres: de uso general y de trabajo. Consulte [Computación](#).

- Los *clústeres de uso general* se crean mediante la interfaz de usuario, la CLI o la API REST. Puede finalizar y reiniciar manualmente un clúster de uso general. Varios usuarios pueden compartir dichos clústeres para llevar a cabo análisis interactivos colaborativos.
- El programador de trabajos de Azure Databricks crea *un clúster de trabajos* cuando se ejecuta un trabajo en un *clúster de trabajos nuevo* y lo finaliza cuando se completa el trabajo. *No* puede reiniciar un clúster de trabajos.

### piscina

Conjunto de instancias inactivas y listas para usar que reduce los tiempos de inicio y escalado automático del clúster. Cuando se adjunta a un grupo, un clúster asigna sus nodos de controlador y de trabajo desde el grupo. Consulte [Referencia de configuración de la piscina](#).

Si el grupo no tiene suficientes recursos inactivos para dar cabida a la solicitud del clúster, el grupo se expande asignando nuevas instancias del proveedor de instancias. Cuando finaliza un clúster asociado, las instancias que ha usado se devuelven al grupo y otro clúster puede reutilizarlas.

### Entorno de ejecución de Databricks

Conjunto de componentes principales que se ejecutan en los clústeres administrados por Azure Databricks. Consulte [Computación](#). Azure Databricks tiene los siguientes runtimes:

- [Databricks Runtime](#) incluye Apache Spark, pero también agrega una serie de componentes y actualizaciones que mejoran considerablemente la facilidad de uso, el rendimiento y la seguridad en el análisis de macrodatos.
- [Databricks Runtime para aprendizaje automático](#) se compila en Databricks Runtime y proporciona una infraestructura de aprendizaje automático prediseñada que se integra con todas las capacidades del área de trabajo de Azure Databricks. Contiene varias bibliotecas populares, como TensorFlow, Keras, PyTorch y XGBoost.

### Interfaz de usuario de trabajos y & canalizaciones

La interfaz de usuario del área de trabajo de **Jobs & Pipelines** proporciona acceso a las interfaces de usuario de Jobs, Lakeflow Spark Declarative Pipelines, y Lakeflow Connect, que son herramientas que permiten orquestar y programar flujos de trabajo.

### Trabajos

Un mecanismo no interactivo para orquestar y programar cuadernos, bibliotecas y otras tareas. Consulte [Trabajos de Lakeflow](#).

### Tuberías / Oleoductos / Canalizaciones

Las canalizaciones declarativas de Spark de Lakeflow proporcionan un marco declarativo para crear canalizaciones de procesamiento de datos confiables, fáciles de mantener y probar. Vea [las canalizaciones declarativas de Lakeflow Spark](#).

### Carga de trabajo

La carga de trabajo es la cantidad de capacidad de procesamiento necesaria para realizar una tarea o un grupo de tareas. Azure Databricks identifica dos tipos de cargas de trabajo: ingeniería de datos (trabajo) y análisis de datos (uso general).

- Ingeniería de datos:** una carga de trabajo (automatizada) se ejecuta en un *clúster de trabajos* que el programador de trabajos de Azure Databricks crea para cada carga de trabajo.
- Análisis de datos:** una carga de trabajo (interactiva) se ejecuta en un *clúster de uso general*. Las cargas de trabajo interactivas suelen ejecutar los comandos dentro de un [cuaderno](#) de Azure Databricks. Sin embargo, la ejecución de un *trabajo* en un *clúster de uso general existente* también se trata como una carga de trabajo interactiva.

### Contexto de ejecución

Estado de un entorno read–eval–print loop (REPL) para cada lenguaje de programación admitido. Los lenguajes admitidos son Python, R, Scala y SQL.

## Ingeniería de datos

Las herramientas de ciencia de datos e ingeniería permiten la colaboración entre científicos de datos, ingenieros de datos, analistas de datos e ingenieros de aprendizaje automático.

### Área de trabajo

Un [área de trabajo](#) es un entorno para acceder a todos los recursos de Azure Databricks. El área de trabajo organiza los objetos (cuadernos, bibliotecas, paneles y experimentos) en [carpetas](#) y proporciona acceso a los objetos de datos y los recursos de cálculo.

### Ordenador portátil

Interfaz basada en web para crear flujos de trabajo de ciencia de datos y aprendizaje automático que pueden contener comandos ejecutables, visualizaciones y texto narrativo. Consulte [Cuadernos de Databricks](#).

## Biblioteca

Paquete de código disponible para el cuaderno o el trabajo que se ejecuta en el clúster. Los entornos de ejecución de Databricks incluyen muchas bibliotecas y puede cargar las suyas propias. Consulte [Instalación de bibliotecas](#).

## Carpeta Git (anteriormente Repos)

Una carpeta cuyo contenido se co versiona conjuntamente mediante la sincronización con un repositorio de Git remoto. [Carpetas de Git de Databricks](#) integran con Git para proporcionar control de código fuente y de versiones para tus proyectos.

## Inteligencia artificial y aprendizaje automático

Databricks proporciona un entorno integrado de un extremo a otro con servicios administrados para desarrollar e implementar aplicaciones de inteligencia artificial y aprendizaje automático.

## IA de mosaicos

El nombre de marca de productos y servicios de Databricks Mosaic AI Research, un equipo de investigadores e ingenieros responsables de los mayores avances de Databricks en la inteligencia artificial generativa. Los productos de IA de mosaicos incluyen las características de aprendizaje automático e inteligencia artificial en Databricks. Vea [Investigación de mosaicos](#) <sup>↗</sup>.

## Entorno de ejecución de aprendizaje automático

Para ayudarlo a desarrollar modelos de ML y AI, Databricks proporciona Databricks Runtime para Machine Learning, que automatiza la creación de cálculos con una infraestructura de aprendizaje automático y aprendizaje profundo prediseñada que incluye las bibliotecas de ML y DL más comunes. También tiene compatibilidad integrada preconfigurada con GPU, incluidos los controladores y las bibliotecas auxiliares. Vaya a información sobre las versiones en tiempo de ejecución más recientes de [notas de la versión de Databricks Runtime y las versiones de compatibilidad](#).

## Experimento

Una colección de [ejecuciones de MLflow](#) para entrenar un modelo de aprendizaje automático. Consulte [Organización de ejecuciones de entrenamiento con experimentos de MLflow](#).

## Características

Las características son un componente importante de los modelos de ML. Un almacén de características permite el uso compartido y la detección de características en toda la organización y también garantiza que se usa el mismo código de cálculo de características para el entrenamiento y la inferencia de modelos. Consulte [Almacén de características de Databricks](#).

## Modelos de IA generativa

Databricks admite la exploración, el desarrollo y la implementación de modelos de IA generativa, entre los que se incluyen los siguientes:

- Área de juegos de IA, un entorno similar al chat en el área de trabajo donde puede probar, preguntar y comparar los LLM. Vea [Conversar con LLMs y prototipar aplicaciones de IA generativa utilizando AI Playground](#).
- Un conjunto integrado de modelos de base preconfigurados que puede consultar:
  - Vea [API de modelos de base de pago por token](#).
  - Vea [\[Recomendado\] Implementación de modelos de base desde Unity Catalog](#) para obtener modelos de base que puede ofrecer con un solo clic.
- LLM hospedados de terceros, denominados [modelos externos](#). Estos modelos están diseñados para usarse tal cual.
- Funcionalidades para personalizar un modelo de base a fin de optimizar su rendimiento para una aplicación específica (lo que se suele denominar ajuste preciso). Consulte [Ajuste del modelo de fundación](#).

## Registro de modelos

Databricks proporciona una versión hospedada del Registro de modelos de MLflow en Unity Catalog. Los modelos registrados en el Catálogo de Unity heredan el control de acceso centralizado, el linaje y la detección y el acceso entre áreas de trabajo. Vea [Administración del ciclo de vida del modelo en Unity Catalog](#).

## Servicio de modelos

Mosaic AI Model Serving proporciona una interfaz unificada para desplegar, gobernar y consultar modelos de IA. Cada modelo que sirva está disponible como una API de REST que puede integrar en la aplicación web o cliente. Con Mosaic IA Model Serving, puede implementar sus propios modelos, modelos básicos o modelos de terceros hospedados fuera de Databricks. Consulte [Implementación de modelos mediante el Servicio de Modelos de IA de Mosaic](#).

## Almacenamiento de datos

El almacenamiento de datos hace referencia a la recopilación y el almacenamiento de datos de varios orígenes para que se pueda acceder rápidamente a ellos para obtener información empresarial e informes. Databricks SQL es la colección de servicios que proporcionan funcionalidades de almacenamiento de datos y rendimiento a los lagos de datos existentes. Consulte [Arquitectura de almacenamiento de datos](#).

## Consultar

Una consulta es una instrucción SQL válida que permite interactuar con los datos. Puede crear consultas mediante el [editor de SQL](#) incluido en la plataforma o puede utilizar un [conector, controlador o API de SQL](#). Consulte [Acceso y administración de consultas guardadas](#) para aprender más sobre cómo trabajar con consultas.

## Almacén de SQL

Un recurso de cálculo en el que se ejecutan las consultas SQL. Hay tres tipos de almacenes de SQL: Clásico, Pro y Sin servidor. Azure Databricks recomienda usar almacenes sin servidor donde estén disponibles. Consulte [Tipos de almacenamiento de SQL](#) para comparar las características disponibles para cada tipo de almacenamiento.

## Historial de consultas

Una lista de consultas ejecutadas y sus características de rendimiento. El historial de consultas permite supervisar el rendimiento de las consultas, lo que le ayuda a identificar cuellos de botella y optimizar los entornos de ejecución de consultas. Consulte [historial de consultas](#).

## Visualización

Presentación gráfica del resultado de ejecutar una consulta. Consulte [Visualizaciones en cuadernos de Databricks y el editor de SQL](#).

## Panel

Presentación de visualizaciones de datos y comentarios. Puede usar paneles para enviar informes automáticamente a cualquier usuario de la cuenta de Azure Databricks. Use Databricks Assistant para ayudarlo a crear visualizaciones basadas en avisos de lenguaje natural. Consulte [Paneles](#). También puede crear un panel desde un cuaderno. Consulte [Paneles en cuadernos](#).

Para consultar los paneles antiguos, vea [paneles antiguos](#).

### 📘 Importante

Databricks recomienda utilizar paneles de IA o BI (anteriormente paneles de Lakeview). Las versiones anteriores de los paneles, anteriormente conocidas como paneles sql de Databricks, ahora se denominan paneles heredados.

**fin del cronograma de soporte:**

- 12 de enero de 2026:** Los paneles y las API heredados ya no son accesibles directamente. Sin embargo, usted todavía puede convertirlos en paneles de IA/BI. La página de migración está disponible hasta el 2 de marzo de 2026.

Convierta paneles heredados mediante la herramienta de migración o la API de REST. Consulte [Clonar un panel heredado en un panel de IA/BI](#) para obtener instrucciones sobre cómo usar la herramienta de migración integrada. Consulte Uso de las [API de Azure Databricks para administrar paneles](#) para ver tutoriales sobre cómo crear y administrar paneles mediante la API REST.

