

Guía de aprendizaje de gobierno de datos: Curar datos de alta calidad

Sigue este tutorial para aprender a preparar datos de confianza con el caso de uso Data intelligence de la prueba data fabric. Su objetivo es crear activos de datos de confianza enriqueciendo los datos y ejecutando el análisis de calidad de datos.

La historia de la guía de aprendizaje corresponde a Golden Bank, que tiene varios departamentos que necesitan acceso a datos de alta calidad sobre hipotecas de cliente. Como encargado de datos en el equipo de gobierno, debe ordenar y organizar los datos de la empresa para proporcionar activos de datos protegidos y de alta calidad que los consumidores de datos pueden encontrar fácilmente en un catálogo de autoservicio.

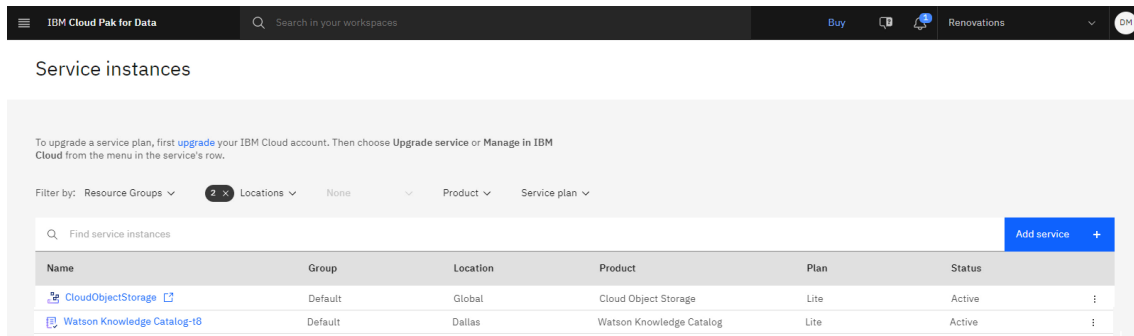
Verificar los servicios suministrados necesarios

Siga estos pasos para verificar o suministrar los servicios necesarios:

1. En el **menú de navegación** , seleccione **Servicios > Instancias de servicio**.
2. Utilice la lista desplegable **Producto** para determinar si existe una instancia del servicio IBM watsonx.data intelligence.
3. Si necesita crear una instancia del servicio IBM watsonx.data intelligence, haga clic en **Añadir servicio**.
 - a. Seleccione **IBM watsonx.data intelligence**.
 - b. Seleccione el plan **Lite**.
 - c. Pulse **Crear**.
4. Repita estos pasos para verificar o suministrar el servicio Cloud Object Storage .

Comprueba tu progreso

La imagen siguiente muestra las instancias de servicio suministradas:



Crear el proyecto de ejemplo

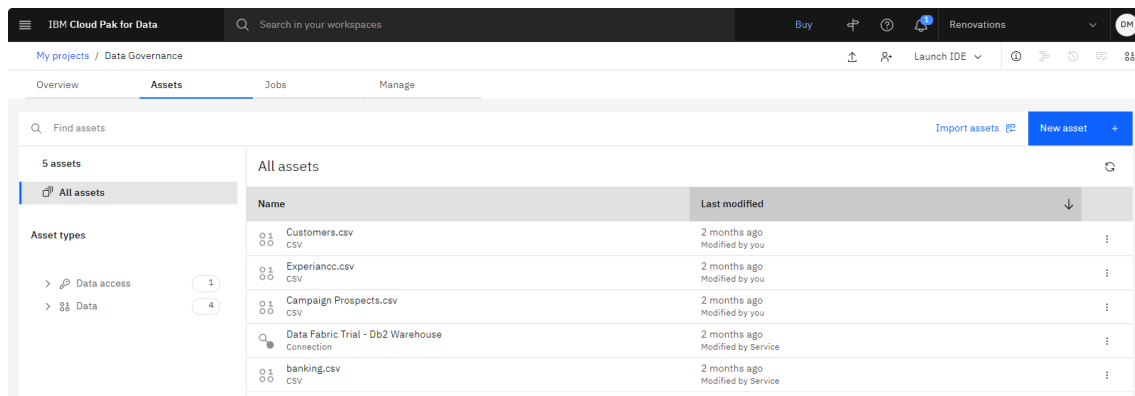
Si todavía no ha creado el proyecto de ejemplo para esta guía de aprendizaje, siga estos pasos:

1. Acceda al [proyecto de ejemplo de Gobernanza de datos](#) en el Centro de recursos.
2. Pulse **Crear proyecto**.
3. Si se le solicita que asocie el proyecto a una instancia de Cloud Object Storage, seleccione una instancia de Cloud Object Storage de la lista.
4. Pulse **Crear**.
5. Espere a que finalice la importación del proyecto y, a continuación, pulse **Ver proyecto nuevo** para verificar que el proyecto y los activos se han creado correctamente.
6. Pulse el separador **Activos** para ver los activos del proyecto.
7. En el menú **Desbordamiento** al final de la fila *Banking.csv* del activo de datos, seleccione **Descargar** y guárdelo en su ordenador. Utilizará ese archivo en un paso posterior.

Nota: Es posible que vea una visita guiada que muestra las guías de aprendizaje que se incluyen con este caso de uso. Los enlaces de la visita guiada abrirán estas instrucciones del tutorial.

Comprueba tu progreso

La imagen siguiente muestra el separador Activos en el proyecto de ejemplo. Ahora está preparado para iniciar la guía de aprendizaje.



Tarea 1: Crear un catálogo

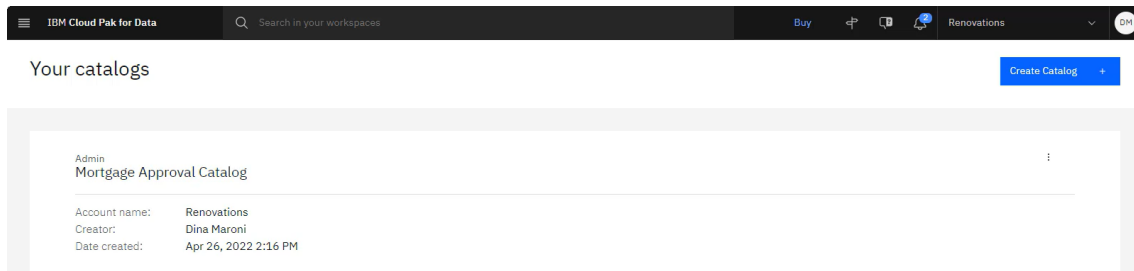
Antes de empezar a trabajar con datos, cree un catálogo donde publicará los datos para compartirlos con su organización. Con el plan IBM watsonx.data intelligence Lite, sólo puede crear dos catálogos. Si ya tiene un catálogo, puede omitir este paso. De lo contrario, siga estos pasos para crear un catálogo:

Sugerencia: Si esta ocasión es la primera vez que accede a un catálogo, verá una visita guiada preguntando si desea realizar una visita guiada a los catálogos. Por ahora, pulse **Quizás más tarde**.

1. En el **menú de navegación**, seleccione **Catálogos > Ver todos los catálogos**.
2. Si ve un catálogo en la página *Catálogos*, vaya a [Tarea 2: Crear una categoría](#). De lo contrario, siga estos pasos para crear un catálogo nuevo:
3. Pulse **Crear catálogo**.
4. Para el *Nombre*, copie y pegue el nombre de catálogo exactamente como se muestra sin espacios iniciales ni finales:
5. Mortgage Approval Catalog
6. Seleccione **Imponer reglas de protección de datos**, confirme la selección y acepte los valores predeterminados para los otros campos.
7. Pulse **Crear**.

Comprueba tu progreso

La imagen siguiente muestra el catálogo. Ahora está preparado para compartir activos con su organización.



Tarea 2: Crear una categoría

Necesita una categoría para contener los términos empresariales que importará en la siguiente tarea. Las categorías actúan como carpetas para organizar los artefactos de gobernabilidad y las personas que pueden crear y gestionar dichos artefactos. Siga estos pasos para crear una categoría:

1. En el Cloud Pak for Data menú de navegación , seleccione **Gobernanza > Categorías**.
2. Pulse **Añadir categoría > Nueva categoría**.
3. Para el nombre, escriba

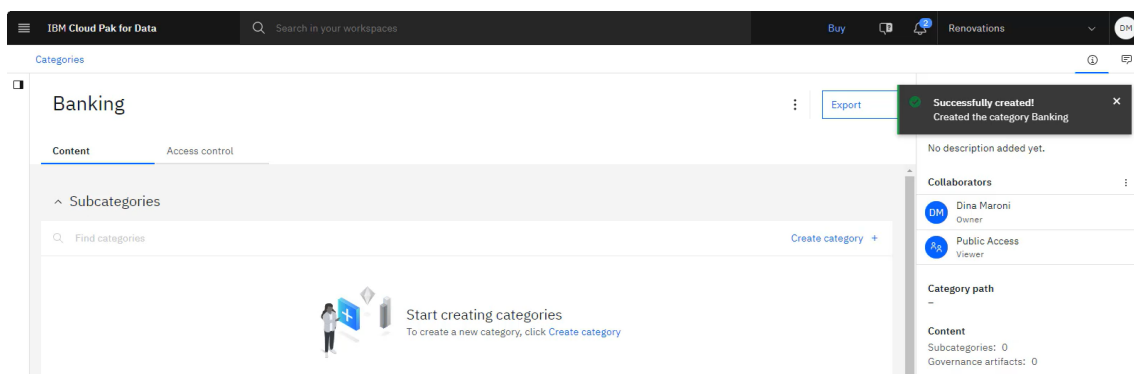
Banking

.

4. Pulse **Crear**.

Comprueba tu progreso

La imagen siguiente muestra la categoría Banca. Ahora está preparado para importar términos de negocio.



Tarea 3: Añadir términos de negocio

Ahora importe los términos de negocio a la nueva categoría. Los utilizará para enriquecer los activos de datos en un paso posterior. Los términos de negocio son definiciones estandarizadas de conceptos de negocio para que los datos se describan de forma uniforme y fácil de comprender en toda la empresa. Siga estos pasos para importar los términos empresariales de un archivo:

1. En el Cloud Pak for Data menú de navegación , seleccione **Gobernanza > Términos comerciales**.
2. Pulse **Añadir término empresarial > Importar desde archivo**.
3. Pulse **Arrastrar y soltar archivo aquí o cargar**.
 - a. Seleccione el archivo **banking.csv** que ha descargado anteriormente.
 - b. Pulse **Abrir**.
4. Pulse **Siguiente**.
5. Seleccione **Sustituir todos los valores** y pulse **Siguiente**.
6. Pulse **Ir a tarea** para ver los términos empresariales de borrador. Si no recibe la notificación, seleccione en el Cloud Pak for Data menú de navegación la opción **Gobernanza > Bandeja de entrada de tareas**.
7. Marque el recuadro de selección **Publicar términos empresariales** y, a continuación, pulse **Publicar**. Pulse **Publicar** para confirmar.
8. En el Cloud Pak for Data menú de navegación , seleccione **Gobernanza > Condiciones comerciales** para ver las condiciones comerciales publicadas.

Comprueba tu progreso

La imagen siguiente muestra los términos de negocio importados. Ahora está preparado para importar los datos a un proyecto y, a continuación, enriquecerlos con los términos de negocio importados.

Business terms		
Published		
Find business terms	Sort by: Name	Show: All
Account Balance	Banking	Last modified: Apr 28, 2022
Account Holder	Banking	Last modified: Apr 28, 2022
Account Holder ID	Banking	Last modified: Apr 28, 2022

Tarea 4: Importar datos a un proyecto

El proyecto de ejemplo incluye una conexión a una instancia de Db2 Warehouse , que contiene los activos de hipoteca. Puede importar metadatos técnicos asociados con los activos de datos en un proyecto o un catálogo para realizar el inventario, evaluar y catalogar estos activos. Los metadatos técnicos describen la estructura de los objetos de datos.

Siga estos pasos para importar los activos de datos:

1. En el **menú de navegación** , seleccione **Proyectos > Ver todos los proyectos**.
2. Pulse el proyecto **Gobierno de datos** .
3. Pulse la pestaña **Activos**.
4. Hacer clic **Nuevo activo > Importar metadatos para activos de datos** .
5. Para el nombre, copie y pegue el texto siguiente:
6. Mortgage data - metadata import
7. Pulse **Siguiente** para continuar.
8. En la página *Definir objetivos*, siga estos pasos:
 - a. Seleccione **Importar metadatos de activos** para importar y ver activos de varios tipos en un proyecto o catálogo.
 - b. Para *el Destino*, seleccione **Añadir a este proyecto**.
 - c. Pulse **Siguiente** para continuar.
9. En la página *Seleccionar fuente y alcance*, siga estos pasos:
 - a. Junto a *Conexión*, haga clic en **Seleccionar**.

- b. Seleccione la conexión **Data Fabric Trial - Db2 Warehouse** y haga clic en **Seleccionar**.
 - c. Junto a *Ámbito*, haga clic en **Seleccionar > Seleccionar activos**.
 - d. Seleccione el esquema **WKC_MORTGAGE**.
 - e. Seleccione las tablas siguientes:
 - CLIENTE_COMERCIAL
 - CREDIT_SCORE
 - HOUSE_PRICE
 - MORTGAGE_APPLICANTS
 - MORTGAGE_APPLICATION
 - f. Revise la lista de activos en el panel lateral y, a continuación, pulse **Seleccionar**.
10. Haga clic en **Siguiente** para continuar en la página *Definir detalles del trabajo*. Puede ejecutar manualmente la importación de metadatos, así que mantenga la programación desactivada.
 11. Haga clic en **Siguiente** para continuar en la página *Establecer opciones avanzadas*.
 12. Acepte los valores predeterminados en la página *Opciones avanzadas* y pulse **Siguiente** para continuar con la revisión.
 13. Revise el resumen de la importación y pulse **Crear**. Se inicia el trabajo de importación de metadatos.
 14. Haga clic en el icono **Actualizar** para ver cómo cambia el estado de *En cola* a *En curso* y, finalmente, a *Importado*. Cuando se complete la ejecución del trabajo, verá los cinco activos listados.

Comprueba tu progreso

La imagen siguiente muestra la importación de metadatos completada. La siguiente tarea es enriquecer los activos de datos importados con los términos empresariales importados.

IBM Cloud Pak for Data

Search in your workspaces

Buy Renovations

Projects / Data Governance / Metadata Import

Metadata Import complete. 5 assets were imported successfully.

Metadata Import

Metadata import

Imported assets

5 assets

Find assets

Name	Type	Context	Last imported	Status
HOUSE_PRICE	Relational table	WKC_MORTGAGE/HOUSE_PRICE	Apr 28, 2022, 03:09 PM	Imported
MORTGAGE_APPLICANTS	Relational table	WKC_MORTGAGE/MORTGAGE_APPLICANTS	Apr 28, 2022, 03:09 PM	Imported
CREDIT_SCORE	Relational table	WKC_MORTGAGE/CREDIT_SCORE	Apr 28, 2022, 03:09 PM	Imported
MORTGAGE_APPLICATION	Relational table	WKC_MORTGAGE/MORTGAGE_APPLICATION	Apr 28, 2022, 03:09 PM	Imported
COMMERCIAL_CLIENT	Relational table	WKC_MORTGAGE/COMMERCIAL_CLIENT	Apr 28, 2022, 03:09 PM	Imported

Items per page: 20 1-5 of 5 items 1 of 1 pages

Tarea 5: Enriquecer los datos importados

Puede enriquecer los activos de datos con información que ayude a los usuarios a encontrar datos más rápidamente para decidir si los datos son adecuados para la tarea en cuestión, si pueden confiar en los datos y cómo trabajar con los datos. Esta información incluye, por ejemplo, términos que definen el significado de los datos, reglas que documentan la propiedad o determinan estándares de calidad, o reseñas. Siga estos pasos para enriquecer los datos importados:

1. Haga clic en el nombre del proyecto **de gobernanza de datos** en la ruta de navegación.



2. Sobre el *Activos* pestaña, haga clic **Nuevo activo > Enriquecer activos de datos con metadatos**.
3. Para el nombre, copie y pegue el texto siguiente:
4. Mortgage data - metadata enrichment
5. Pulse **Siguiente** para continuar.
6. Pulse **Seleccionar datos del proyecto**.
 - a. Seleccione **Metadata import**.

b. Pulse el recuadro de selección situado junto a **Datos de hipoteca-importación de metadatos**. Este activo incluye los activos siguientes:

- COMMERCIAL_CLIENT
- CREDIT_SCORE
- HOUSE_PRICE
- MORTGAGE_APPLICANTS
- MORTGAGE_APPLICATION

c. Pulse **Seleccionar**.

7. Pulse **Siguiente** para continuar con el objetivo de enriquecimiento.

8. Seleccione todos los objetivos de enriquecimiento:

- Datos de perfil
- Asignar términos
- Ejecutar análisis de calidad básico

9. Para *Categorías*, pulse **Seleccionar categorías**.

a. Seleccione solo **[sin categorizar]** y **Banca**.

b. Pulse **Seleccionar**.

10. Para *Muestreo*, seleccione **Básico**.

11. Pulse **Siguiente** para continuar con la planificación. Puede ejecutar manualmente la importación, por lo que debe mantener la planificación desactivada.

12. Pulse **Siguiente** para continuar con la revisión.

13. Pulse **Crear**.

14. Se muestra el activo de enriquecimiento de metadatos, pero el trabajo puede tardar varios minutos en completarse. Haga clic en el icono **Actualizar** para ver cómo cambia el estado de *No analizado* a *En curso* y, finalmente, a *Finalizado*. Cuando se complete la ejecución del trabajo, verá los cinco activos listados.

Comprueba tu progreso

La imagen siguiente muestra el enriquecimiento de metadatos completado. Ahora puede explorar los activos de datos enriquecidos.

IBM Cloud Pak for Data

Search in your workspaces

Buy

Renovations

DM

Projects / Data Governance

Tarea 6: Ver los resultados del enriquecimiento de metadatos

Una vez completada la ejecución del enriquecimiento de metadatos, siga estos pasos para ver los datos enriquecidos:

1. En la pantalla *Enriquecimiento de metadatos de datos de hipoteca*, pulse la pestaña **Columnas**.
2. En la lista de *Columnas*, localice la columna **EMAIL_ADDRESS** para el activo *MORTGAGE_SOLICITANTES*.
 - a. Al final de la fila *EMAIL_ADDRESS* para *MORTGAGE_APPLICANTS*, haga clic en el menú **Desbordamiento** y seleccione **Ver detalles de la columna**.
 - b. En el panel lateral de la pestaña *Detalles*, verá información de perfilado como: Formato, Distribución de frecuencia, Estadísticos.
 - c. En el panel lateral, pulse la pestaña **Gobierno**. Esta pestaña incluye las clases de datos y los términos empresariales que se asignaron automáticamente durante el enriquecimiento de metadatos. También puede ver los términos de negocio sugeridos y las clases de datos, y asignarlos manualmente.
 - d. Revise los términos de negocio sugeridos o las clases de datos y asígneles manualmente. Por ejemplo, puede ver *Dirección* como un término empresarial sugerido.
 - a. Pulse **Términos empresariales sugeridos**.
 - b. Para *Dirección*, pulse **Asignar**.

3. Al final de la columna *EMAIL_ADDRESS* de la fila de activos *MORTGAGE_APPLICANTS*, haga clic en el menú **Desbordamiento** y seleccione **Ver detalles de calidad de los datos**.
 - a. Ver la información de calidad de datos. IBM watsonx.data intelligence genera automáticamente una puntuación de calidad de los datos para cada columna y activo de datos analizando cada valor de cada registro según dimensiones preestablecidas.
 - b. Pulse la **X** para cerrar la ventana *Calidad de datos* .
4. En la columna *CITY* del activo *CREDIT_SCORE*, haga clic en el menú **desplegable** y seleccione **Marcar como revisado**.
5. Pulse la pestaña **Activos**.
6. En la lista de *activos*, para el activo *MORTGAGE_APPLICANTS*, haga clic en el menú **desplegable** y seleccione **Ver detalles del activo**.
 - a. En el panel lateral, pulse la pestaña **Gobierno** para ver la asignación automática de términos de negocio.
 - b. Haga clic en el icono **Editar** para asignar manualmente los términos comerciales.
 - c. Busque social. Si no ve ningún resultado, asegúrese de que la lista desplegable esté establecida en **Todos los términos** en lugar de *Términos sugeridos*.
 - d. Seleccione **Número de seguridad social**.
 - e. Pulse **Asignar**.

Comprueba tu progreso

La imagen siguiente muestra los activos de datos revisados y enriquecidos. El paso siguiente es publicar los datos enriquecidos en un catálogo para compartirlos con su organización.

IBM Cloud Pak for Data

Search in your workspaces

Buy Renovations

Projects / Data Governance

Mortgage data - metadata enrichment

Default settings Edit enrichment

Business term assigned
Social Security Number was assigned to MORTGAGE_APPLICANTS

Assets (5)	Columns (44)				
Find assets (search by name, source)	Last refreshed: 4/28/22, 3:17 PM				
Assets	Source	Business terms	Data quality	Review status	Enrichment status
COMMERCIAL_CLIENT	Data Fabric Trial - Db2 Warehouse	Commercial C...	100%	Finished	Apr 28, 2022, 3:12 PM
CREDIT_SCORE	Data Fabric Trial - Db2 Warehouse	Credit Score	93%	Finished	Apr 28, 2022, 3:12 PM
HOUSE_PRICE	Data Fabric Trial - Db2 Warehouse	House Price	91%	Finished	Apr 28, 2022, 3:12 PM
MORTGAGE_APPLICANTS	Data Fabric Trial - Db2 Warehouse	Mortgage App... + 2 more	96%	Finished	Apr 28, 2022, 3:12 PM
MORTGAGE_APPLICATION	Data Fabric Trial - Db2 Warehouse	Mortgage App... + 1 more	92%	Finished	Apr 28, 2022, 3:12 PM

Items per page: 50 1-5 of 5 items 1 of 1 page

Details Governance

Business terms 3

Mortgage Applicant
Banking
Confidence score: 100%

Mortgage Application
Banking
Confidence score: 100%

Social Security Number
Banking
Assigned by Dina Maroni on Apr 28, 2022

Tarea 7: Publicar datos en un catálogo

Ahora que ha enriquecido los datos, desea publicar esos activos de datos en un catálogo para que los científicos de datos y los analistas de datos puedan utilizar los activos de datos enriquecidos. Siga estos pasos para almacenar los activos de datos enriquecidos en un catálogo para que otros tengan acceso a los datos de confianza:

1. Pulse el nombre del proyecto **Gobierno de datos** en la pista de navegación.
2. Pulse la pestaña **Activos**.
3. Seleccione **Datos > Activos de datos**.
4. Seleccione los activos de datos **COMMERCIAL_CLIENT**, **HOUSE_PRICE**, **MORTGAGE_SOLICITANTE** y **MORTGAGE_APPLICATION** de la lista y pulse **Publicar en catálogo**.
 - a. Para el catálogo *Destino*, seleccione **Catálogo de aprobación de hipotecas** y pulse **Siguiente**.
 - b. Para las *Etiquetas*, escriba la etiqueta,
trusted
, pulse + (signo más) y, a continuación, pulse **Siguiente**.
 - c. Revise los activos y pulse **Publicar**.
5. Borre todos los activos seleccionados y, a continuación, seleccione el recuadro de selección situado junto al activo **CREDIT_SCORE** de la lista y pulse **Publicar en catálogo**.

a. Para el catálogo *Destino* , seleccione **Catálogo de aprobación de hipotecas** y pulse **Siguiente**.

b. Para las *Etiquetas*, escriba la etiqueta

confidential

y pulse + (signo más).

c. Escriba la etiqueta

trusted

y pulse + (signo más) en una segunda etiqueta.

d. Seleccione la opción **Ir al catálogo después de publicarlo** y pulse **Siguiente**.

e. Revise los activos y pulse **Publicar**.

6. Filtre los activos en el *Catálogo de aprobación de hipotecas*.

a. Haga clic en el icono **Filtro**.

b. Expanda la sección **Etiqueta** .

c. Seleccione **de confianza** y pulse **Aplicar**.

d. Verifique que los cinco activos de datos se han añadido al catálogo.

7. Cambie el nombre del activo de datos *HIPOTECAGE_SOLICITANTES* .

a. Abra el activo **HIPOTECAGE_SOLICITANTES** .

b. Haga clic en el icono **Editar** nombre.

c. Cambie el nombre por:

MORTGAGE_APPLICANTS_TRUST

e. Haga clic en **Aplicar**.

Comprueba tu progreso

La imagen siguiente muestra los activos de datos enriquecidos publicados en un catálogo. Ahora tiene datos de confianza disponibles a través del catálogo de su empresa.

IBM Cloud Pak for Data

Search in your workspaces

Buy

Renovations

DM

Catalogs

Add to catalog

Mortgage Approval Catalog

Assets

Access control

Settings

Find assets

Filter by: Any asset type

1 x Tag

Clear all

Show featured assets

trusted x

Showing 6 of 6 items

<input type="checkbox"/>	Name	Owner	Tags	Business terms	Asset type	Date added	
<input type="checkbox"/>	COMMERCIAL_CLIENT	DM Dina Maroni	trusted	Commercial Client	Data	May 05, 2022	:
<input type="checkbox"/>	CREDIT_SCORE	DM Dina Maroni	trusted confidential	Credit Score	Data	May 05, 2022	:
<input type="checkbox"/>	Data Fabric Trial - Db2 Warehouse	DM Dina Maroni	trusted		Connection	May 05, 2022	:
<input type="checkbox"/>	HOUSE_PRICE	DM Dina Maroni	trusted	House Price	Data	May 05, 2022	:
<input type="checkbox"/>	MORTGAGE_APPLICANTS_TRUST	DM Dina Maroni	trusted	Mortgage Application + 2 more	Data	May 05, 2022	:
<input type="checkbox"/>	MORTGAGE_APPLICATION	DM Dina Maroni	trusted	Mortgage Application + 1 more	Data	May 05, 2022	:

Como encargado de datos en el equipo de gobierno, ha aprendido a ordenar y organizar los datos de la empresa para proporcionar activos de datos protegidos y de alta calidad que los consumidores de datos pueden encontrar fácilmente en un catálogo de autoservicio.