

# Guía de aprendizaje de integración de datos: Integrar datos

Siga esta guía de aprendizaje para transformar datos almacenados en tres orígenes de datos externos con el caso de uso de integración de datos de la prueba de entramado de datos. Su objetivo es utilizar DataStage para transformar los datos y luego distribuir los datos transformados a un único archivo de salida. Si completó el tutorial *Virtualizar datos externos*, entonces realizó muchas de las mismas tareas utilizando Data Virtualization que este tutorial realiza utilizando DataStage.

La historia de la guía de aprendizaje es que Golden Bank necesita cumplir con un nuevo reglamento por el que no puede realizar préstamos a los solicitantes de préstamos que no estén cualificados. Como ingeniero de datos en Golden Bank, actualmente utiliza DataStage para agregar sus datos de solicitudes de hipotecas anónimas a la información de identificación personal de los solicitantes de hipotecas. Sus prestamistas utilizan esta información para ayudarles a decidir si deben aprobar o denegar solicitudes de hipoteca. Su liderazgo agregó algunos analistas de riesgo que calculan diariamente qué tasa de interés recomiendan ofrecer a los prestatarios en cada rango de puntuación de crédito. Es necesario integrar esta información en la hoja de cálculo que comparte con los prestamistas. La hoja de cálculo incluye información de puntuación de crédito para cada solicitante, la deuda total del solicitante y una tabla de búsqueda de tipos de interés. Por último, cargue los datos en un archivo CSV de salida de destino.

Utilizará DataStage para unirse a los datos del solicitante y de la solicitud, filtrar por estado, unirse a las puntuaciones de crédito del solicitante, calcular la deuda total, buscar el tipo de interés de la hipoteca para ofrecer en función de los rangos de puntuación de crédito y generar el resultado en un archivo CSV.

## Vista previa de la guía de aprendizaje

En esta guía de aprendizaje, realizará estas tareas:

- [Configure los requisitos previos.](#)
- [Tarea 1: Ejecutar un flujo de DataStage existente](#)
- [Editar el flujo de DataStage para:](#)
  - [Tarea 2: Especificar una columna clave para la etapa Unión](#)
  - [Tarea 3: Añadir datos de puntuación de crédito de una base de datos PostgreSQL](#)

- [Tarea 4: Añadir una etapa Join para unir los datos de puntuación de crédito con los datos del solicitante y de la solicitud](#)
  - [Tarea 5: Añadir una etapa Transformer para calcular la deuda total](#)
  - [Tarea 6: Añadir datos de tipo de interés desde una base de datos MongoDB](#)
  - [Tarea 7: Añadir una etapa Búsqueda para buscar tipos de interés para los candidatos](#)
- [Tarea 8: Editar el nodo de archivo secuencial y ejecutar el flujo de DataStage](#)
- [Tarea 9: Crear un catálogo para almacenar el elemento de datos publicado](#)
- [Tarea 10: Ver la salida y publicar en un catálogo](#)
- [Limpieza \(opcional\)](#)

## Requisitos previos

### Registro en Cloud Pak for Data as a Service

Debe registrarse en Cloud Pak for Data as a Service y suministrar los servicios necesarios para el caso de uso de integración de datos.

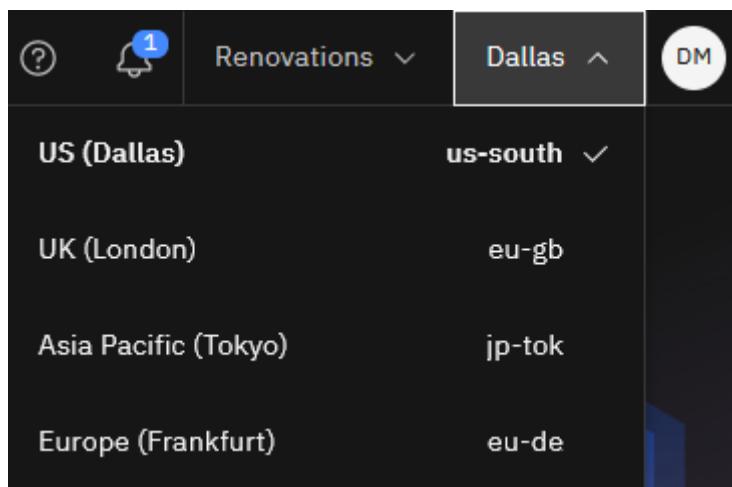
- Si tiene una cuenta de Cloud Pak for Data as a Service existente, puede empezar con esta guía de aprendizaje. Si tiene una cuenta de plan Lite, sólo un usuario por cuenta puede ejecutar esta guía de aprendizaje.
- Si aún no tienes una cuenta, [suscríbete a una prueba de tejido de datos](#).

## Verificar los servicios suministrados necesarios

Siga estos pasos para verificar o suministrar los servicios necesarios:

1. En Cloud Pak for Data, verifique que está en la región de Dallas o Frankfurt. Si no es así, haga clic en el menú desplegable de la región y

seleccione **Dallas** o **Fráncfort**.



2. En el menú Navegación, seleccione **Servicios > Instancias de servicio**.
3. Utilice la lista desplegable **Producto** para determinar si existe una instancia de servicio de DataStage .
4. Si necesita crear una instancia de servicio de DataStage, pulse **Añadir servicio**.
5. Seleccione **DataStage**.
  - a. Para la región, seleccione **Dallas** o **Frankfurt**.
  - b. Seleccione el plan **Lite**.
  - c. Pulse **Crear**.
6. Repita estos pasos para verificar o suministrar los siguientes servicios adicionales:
  - o IBM watsonx.data intelligence
  - o Cloud Object Storage

#### Comprueba tu progreso

La imagen siguiente muestra las instancias de servicio suministradas:

The screenshot shows the 'Service instances' section of the IBM Cloud Pak for Data interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data', 'All', a search bar ('Search in your workspaces'), 'Buy', 'Renovations', and a user profile icon. Below the navigation is a message: 'To upgrade a service plan, first [upgrade](#) your IBM Cloud account. Then choose [Upgrade service](#) or [Manage in IBM Cloud](#) from the menu in the service's row.' A filter bar allows filtering by 'Resource Groups', 'Locations' (set to 'None'), 'Product' (set to 'None'), and 'Service plan'. A search input 'Find service instances' is present. A blue button 'Add service +' is at the top right. The main table lists three service instances:

Name	Group	Location	Product	Plan	Status
CloudObjectStorage	Default	Global	Cloud Object Storage	Lite	Active
DataStage-rv	Default	Dallas	DataStage	Lite	Active
KnowledgeCatalog	Default	Dallas	Watson Knowledge Catalog	Lite	Active

## Crear el proyecto de ejemplo

Si ya tiene el proyecto de ejemplo para esta guía de aprendizaje, vaya a la [Tarea 1](#).

De lo contrario, siga estos pasos:

1. Acceda al [proyecto de ejemplo Integración 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 las conexiones y el flujo de DataStage .

**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.

Name	Last modified	
Data Integration DataStage flow	4 weeks ago Modified by you	⋮
Monitor data replication Notebook from local system	1 month ago Modified by Service	⋮
CREDIT_SCORE application/x-ibm-rel-table	1 month ago Modified by Service	⋮
Data Fabric Trial - Db2 on Cloud - Source Connection	1 month ago Modified by Service	⋮
Data Fabric Trial - MongoDB Connection	1 month ago Modified by Service	⋮
Data Fabric Trial - Db2 Warehouse Connection	1 month ago Modified by Service	⋮
Data Fabric Trial - Databases for PostgreSQL Connection	1 month ago Modified by Service	⋮

## Tarea 1: Ejecutar un flujo de DataStage existente

Empiece con un flujo de DataStage básico que se une a los solicitantes de hipotecas y a los conjuntos de datos de aplicaciones de hipotecas y, a continuación, genera el resultado en un archivo CSV en el proyecto. Siga estos pasos para ejecutar el flujo de DataStage :

1. Inicie el proyecto **Integración de datos** . Si no tiene el proyecto abierto, siga estos pasos:
  - a. En el **menú Navegación**, seleccione **Proyectos > Ver todos los proyectos**.
  - b. Abra el proyecto **Integración de datos** .
2. Pulse el separador **Activos** para ver todos los activos del proyecto.
3. Pulse **Flujos > Flujos de DataStage**.

**Sugerencia:** Si no ve ningún flujo de DataStage , vuelva a ver las instancias de servicio para verificar que la instancia de DataStage se ha suministrado correctamente.

4. Pulse el flujo **Integración de datos** en la lista para abrirlo. Este flujo se une a las tablas *Solicitantes de hipoteca* y *Aplicaciones de hipoteca* que están almacenadas en Db2 Warehouse, filtra los datos a esos registros desde el estado de California y crea un archivo secuencial en formato CSV como salida.
5. Haga clic en el icono **de acercamiento** y el icono **de alejamiento** de la barra de herramientas para configurar la vista que prefiera del lienzo.

6. Efectúe una doble pulsación en el nodo **MORTGAGE\_APPLICATIONS\_1** para ver los valores.

  - a. Expanda la sección **Propiedades**.
  - b. Desplácese hacia abajo y, a continuación, pulse **Vista previa de datos**. Este conjunto de datos incluye información que se captura en una aplicación de hipoteca.
  - c. Pulse **Cerrar**.
7. Efectúe una doble pulsación en el nodo **MORTGAGE\_APPLICANTS\_1** para ver los valores.

  - a. Expanda la sección **Propiedades**.
  - b. Desplácese hacia abajo y pulse **Vista preliminar de los datos**. Este conjunto de datos incluye información sobre los solicitantes de hipotecas que solicitaron un préstamo.
  - c. Opcional: Visualice los datos.
    - a. Pulse el panel **Gráfico**.
    - b. En la lista *Columnas a visualizar*, seleccione **STATE**.
    - c. Pulse **Visualizar datos** para ver un gráfico circular que muestra la distribución de los datos por estado.
    - d. Pulse el icono **Mapa de árbol** para ver los mismos datos en un gráfico de mapa de árbol.
  - d. Pulse **Cerrar**.
8. Efectúe una doble pulsación en el nodo **Join\_on\_ID** para ver los valores.

  - a. Expanda la sección **Propiedades**.

- b. Tenga en cuenta que la clave de unión es la columna ID.

The screenshot shows the configuration of a 'Join' stage in a DataStage flow. The stage is named 'Join\_on\_ID'. It has three tabs: 'Stage' (selected), 'Input', and 'Output'. Under 'Properties', there is a section titled 'Join Keys (required)' with a table. The table has two columns: 'Key' and 'Case sensitive'. There is one row with the value 'ID' in the 'Key' column. An 'Edit' button is located at the top right of the table.

Key	Case sensitive
ID	

- c. Pulse **Cancelar** para cerrar los valores.

9. Haga clic en el icono **Registros** de la barra de herramientas para poder ver el progreso del flujo.
10. Pulse **Compilar** y luego **Ejecutar**. De forma alternativa, puede pulsar **Ejecutar**, que compila y, a continuación, ejecuta el flujo de DataStage. La ejecución puede tardar aproximadamente un minuto en completarse.
11. Vea los registros. Puede utilizar el total de filas y filas/seg para cada paso del flujo para verificar visualmente que el filtro funciona según lo esperado.
12. Cuando la ejecución finalice correctamente, haga clic en **Integración de datos** en la ruta de navegación para volver al proyecto.

IBM Cloud Pak for Data

My projects / Data Integration

13. En la pestaña **Activos**, pulse **Datos > Activos de datos**.
14. Abra el archivo **MORTGAGE\_DATA.CSV**. Puede ver que este archivo contiene las columnas de los solicitantes de hipoteca y los conjuntos de datos de solicitudes de hipoteca.

**Comprueba tu progreso**

La siguiente imagen muestra el archivo CSV resultante. La siguiente tarea es editar el flujo DataStage.

ID	INCOME	APPLIEDONLINE	RESIDENCE	YRS_AT_CURRENT_ADDRESS	YRS_WITH_CURRENT_EMPLOYER	NUMBER_OF_CARDS	CREDITCARD
100288	44956	YES	Owner Occu...	21	21	2	806
100310	43972	NO	Private Renti...	0	13	1	340
100338	45813	YES	Owner Occu...	10	13	1	59
100354	57718	NO	Owner Occu...	25	16	2	1555
100381	53115	NO	Owner Occu...	12	1	2	674
100389	59036	NO	Owner Occu...	7	6	1	495
100390	53559	NO	Owner Occu...	7	19	2	2646
100423	49007	YES	Living with p...	14	18	1	13360
100465	44048	YES	Owner Occu...	0	6	1	222
100537	45081	YES	Owner Occu...	14	15	2	713
100562	45103	YES	Owner Occu...	21	14	1	2327
100568	55752	NO	Owner Occu...	5	11	2	2048
100586	44378	NO	Private Renti...	9	18	2	306
100593	50445	NO	Public Housi...	10	8	1	237
100670	56470	NO	Private Renti...	31	15	2	390
100676	44597	NO	Owner Occu...	10	12	2	468
100702	59201	NO	Owner Occu...	12	25	1	12164
100753	53550	YES	Public Housi...	11	23	1	544
100764	44787	NO	Owner Occu...	9	4	2	982
100766	57376	NO	Private Renti...	21	27	2	684

## Visión general: Editar el flujo de DataStage

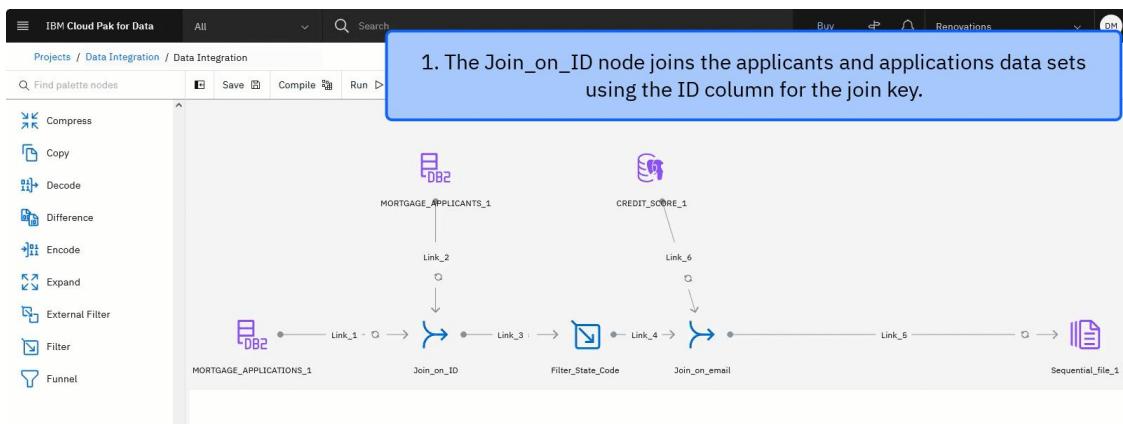
Ahora que se ha unido al solicitante de hipoteca y a los datos de solicitud, está preparado para editar el flujo de DataStage para:

- Tarea 2: Especificar una columna clave para la etapa Join.
- Tarea 3: Añadir datos de puntuación de crédito de una base de datos PostgreSQL .
- Tarea 4: Añadir una etapa de unión para unir los datos de puntuación de crédito con el solicitante y los datos de solicitud.
- Tarea 5: Añadir una etapa Transformer para calcular la deuda total.
- Tarea 6: Añadir datos de tipo de interés de una base de datos MongoDB .
- Tarea 7: Añadir una etapa de búsqueda para buscar tasas de interés para los solicitantes en función de sus puntuaciones de crédito y los rangos de tasas de interés diarias de Golden Bank.

## Tarea 2: Especificar la columna clave para la etapa Join

La identificación de una columna clave indica a DataStage que la columna contiene valores exclusivos. El nodo **Join\_on\_ID** se une a los solicitantes de hipoteca y a los conjuntos de datos de aplicaciones de hipoteca utilizando la columna de ID para la clave de unión. La siguiente fase es unir el conjunto de datos resultante con los datos de puntuación de crédito. Más adelante, unirá los datos filtrados resultantes con el conjunto de datos de puntuación de crédito. La segunda unión utilizará la columna **EMAIL\_ADDRESS** como clave de unión. En esta tarea, edite el flujo de DataStage para especificar la columna **EMAIL\_ADDRESS** como columna clave para el conjunto de datos resultante cuando se une con los datos de puntuación de crédito.

La siguiente imagen animada proporciona una representación visual como alternativa a la descripción de los dos nodos de unión. Pulse la imagen para ver una imagen más grande.



Siga estos pasos para cambiar los valores del nodo Join:

1. Haga clic en **Integración de datos** en la ruta de navegación para volver al proyecto.

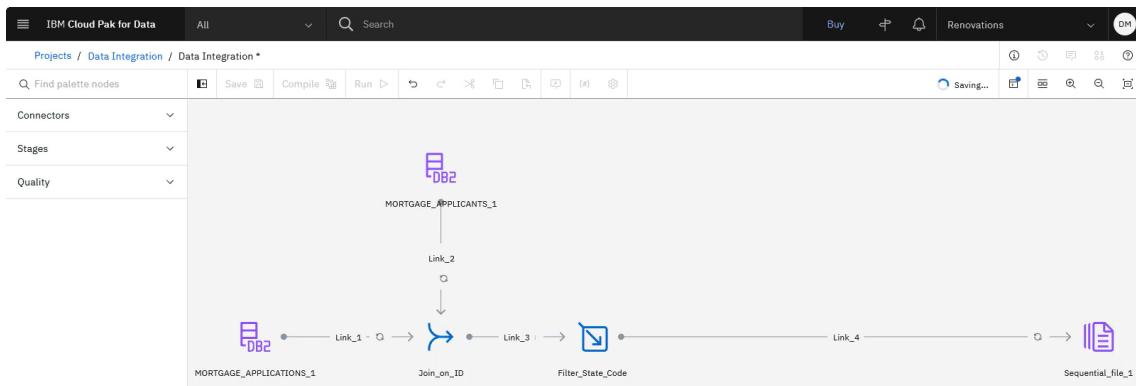


2. En el separador **Activos**, pulse **Flujos > Flujos de DataStage**.
3. Abra el flujo **Integración de datos**.
4. Efectúe una doble pulsación en el nodo **Join\_on\_ID** para editar los valores.
5. Pulse la pestaña **Salida** y expanda la sección **Columnas** para ver una lista de las columnas del conjunto de datos unido.

6. Pulse **Editar**.
7. Para el nombre de columna *EMAIL\_ADDRESS*, seleccione **Clave**.
8. Pulse **Aplicar y volver** para volver a los valores del nodo de *Join\_on\_ID*.
9. Pulse **Guardar** para guardar los valores del nodo de *Join\_on\_ID*.

### Comprueba tu progreso

La imagen siguiente muestra el flujo de DataStage con la etapa *Join\_on\_id* editada. Ahora que ha identificado la columna *EMAIL\_ADDRESS* como la columna clave, puede añadir los datos de PostgreSQL que contienen las puntuaciones de crédito de los solicitantes.



### Tarea 3: Añadir datos de puntuación de crédito de una base de datos PostgreSQL

Siga estos pasos para añadir los datos de puntuación de crédito almacenados en una base de datos PostgreSQL al flujo de DataStage :

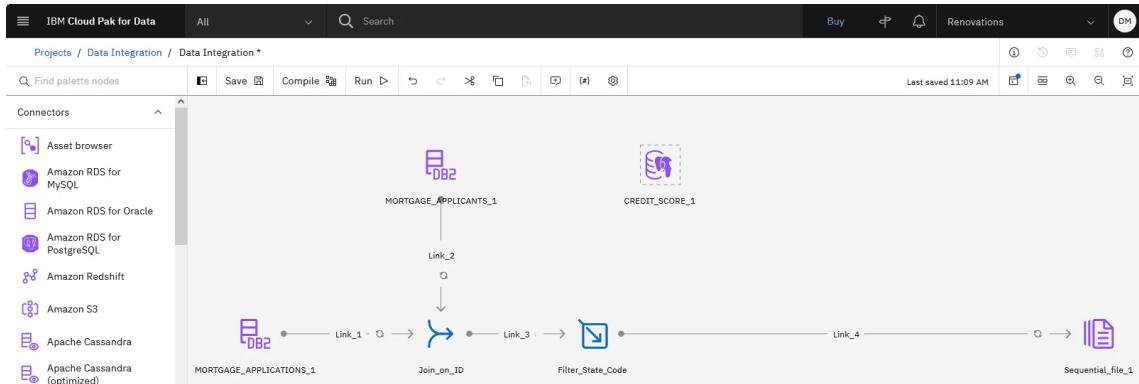
1. En la paleta de nodos, expanda la sección **Conectores**.
2. Arrastre el conector **Navegador de activos** al lienzo junto al nodo *MORTGAGE\_APPLICANTS\_1* .
3. Localice el activo seleccionando **Conexión > Data Fabric Trial- Databases for PostgreSQL > BANKING > CREDIT\_SCORE**.

**Nota:** Pulse el nombre de conexión o esquema en lugar del recuadro de selección para expandir la conexión y el esquema.

4. Haga clic en el icono **Vista previa** para obtener una vista previa de los datos de la puntuación crediticia de cada solicitante.
5. Pulse **Añadir**.

### Comprueba tu progreso

La imagen siguiente muestra el flujo de DataStage con el activo de puntuación de crédito añadido. Ahora que ha añadido los datos de puntuación de crédito al lienzo, debe unirse a los datos de solicitante, solicitud y puntuación de crédito.



### Tarea 4: Añadir una etapa de unión para unir los datos de puntuación de crédito con el solicitante y los datos de solicitud

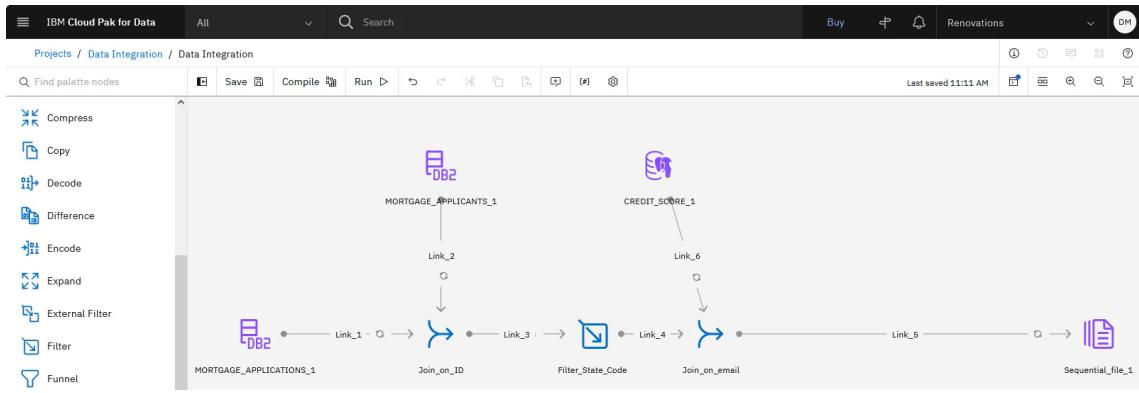
Siga estos pasos para añadir otra etapa Join para unirse a la solicitud de hipoteca filtrada y el solicitante de hipoteca se unió a los datos con los datos de puntuación de crédito en el flujo de DataStage :

1. En la paleta de nodos, expanda la sección **Etapas**.
2. Arrastre la etapa **Unión** al lienzo y suelte el nodo en la línea de enlace entre los nodos *Filter\_State\_Code* y *Sequential\_file\_1* .
3. Pase el ratón por encima del conector **CREDIT\_SCORE\_1** para ver la flecha. Conecte la flecha a la etapa **Unión**.

4. Efectúe una doble pulsación en el nodo **CREDIT\_SCORE\_1** para editar los valores.
  - a. Pulse la pestaña **Salida** y expanda la sección **Columnas** para ver una lista de las columnas del conjunto de datos unido.
  - b. Pulse **Editar**.
  - c. Para los nombres de columna *EMAIL\_ADDRESS* y *CREDIT\_SCORE*, seleccione **Clave**.
  - d. Pulse **Aplicar y volver** para volver a los valores del nodo de *CREDIT\_SCORE\_1*.
  - e. Pulse **Guardar** para guardar los valores del nodo de *CREDIT\_SCORE\_1*.
5. Efectúe una doble pulsación en el nodo **Join\_1** para editar los valores.
  - a. Expanda la sección **Propiedades**.
  - b. Pulse **Añadir clave**.
    - a. Vuelva a pulsar **Añadir clave**.
    - b. Seleccione **EMAIL\_ADDRESS** en la lista de claves posibles.
    - c. Haga clic en **Aplicar**.
    - c. Pulse **Aplicar y volver** para volver a los valores del nodo de *Join\_1*.
    - d. Cambie el nombre de nodo de *Join\_1* por  
Join\_on\_email  
.
  - e. Pulse **Guardar** para guardar los valores del nodo de *Join\_1*.

### **Comprueba tu progreso**

La imagen siguiente muestra el flujo de DataStage con una segunda etapa Join añadida. Ahora que se ha unido a los datos de solicitud, solicitante y puntuación de crédito, debe añadir una etapa Transformer para calcular la deuda total de cada solicitante.



## Tarea 5: Añadir una etapa Transformer para calcular la deuda total

Siga estos pasos para añadir una etapa Transformer que cree una nueva columna sumando las columnas LOAN\_AMOUNT y CREDITCARD\_DEBT:

1. En la sección **Etapas**, arrastre la etapa **Transformer** al lienzo y suelte el nodo en la línea de enlace entre los nodos **Join\_on\_email** y **Sequential\_file\_1**.
2. Efectúe una doble pulsación en el nodo **Transformador** para editar los valores.
3. Pulse la pestaña **Salida**.
  - a. Pulse **Añadir columna**.
  - b. Desplácese hacia abajo en la lista de columnas para ver la nueva columna.
  - c. Nombre de la columna **TOTAL\_DEBT**.
  - d. Haga clic en el icono **Editar** en la columna *Derivación* de la fila.
  - e. Haga clic en el icono **de la calculadora** en la columna *Derivación* para abrir el generador de expresiones.
  - f. Busque **LOAN\_AMOUNT** y efectúe una doble pulsación en el nombre de columna para añadirlo a la expresión. Tenga en cuenta que el número de enlace se añade al nombre de columna.
  - g. Escriba un signo más **+**.
  - h. Busque **CREDITCARD\_DEBT**

y, a continuación, efectúe una doble pulsación en el nombre de columna para añadirlo a la expresión. Tenga en cuenta que el número de enlace se añade al nombre de columna.

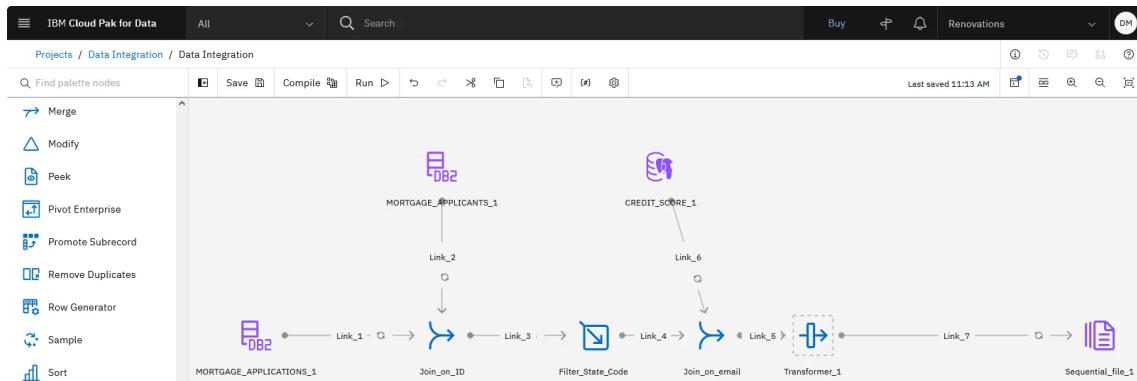
- i. Verifique que la expresión final es Link\_7.LOAN\_AMOUNT + Link\_7.CREDITCARD\_DEBT.

**Nota:** El número de enlace puede ser diferente.

- j. Pulse **Aplicar y volver** para volver a la página *Transformador*.
  - k. Para el nombre de columna *CREDIT\_SCORE*, seleccione **Clave**.
4. Haz clic en la pestaña **Escenario**.
  - a. Seleccione la página **Avanzado**.
  - b. Cambia el *Modo de ejecución* a **Secuencial**.
5. Pulse **Guardar y volver** para volver al lienzo.

### Comprueba tu progreso

La imagen siguiente muestra el flujo de DataStage con la etapa Transformer añadida. Ahora que ha calculado la deuda total de cada solicitante, debe añadir la tabla de tipos de interés para ofrecer en función de los rangos de puntuación de crédito.



### Tarea 6: Añadir datos de tipo de interés de una base de datos MongoDB

Siga estos pasos para incluir los tipos de interés en el flujo añadiendo un conector de activo de datos a una base de datos MongoDB :

1. En la paleta de nodos, expanda la sección **Conectores**.

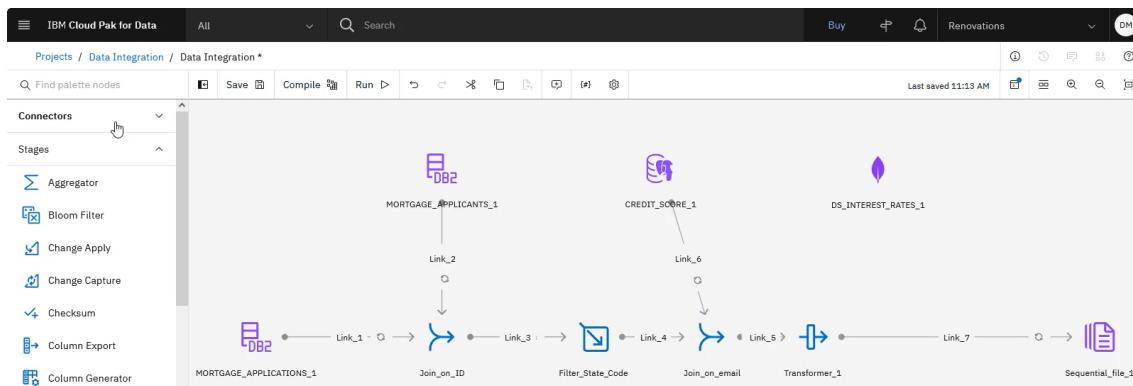
2. Arrastre el conector **Navegador de activos** al lienzo junto al nodo **CREDIT\_SCORE\_1**.
3. Localice el activo seleccionando **Conexión > Data Fabric Trial-Mongo DB > DOCUMENT > DS\_INTEREST\_RATES**.
4. Haga clic en el ícono **Vista previa** para obtener una vista previa de los tipos de interés para cada rango de puntuación crediticia.

Puede utilizar los valores de las columnas **STARTING\_LIMIT** y **ENDING\_LIMIT** para buscar el tipo de interés adecuado en función de la puntuación crediticia del solicitante. La columna ID no es necesaria, por lo que suprimirá esa columna en el paso siguiente.

5. Pulse **Añadir**.

### Comprueba tu progreso

La imagen siguiente muestra el flujo de DataStage con el activo de datos de tipos de interés añadido desde el origen externo MongoDB . Ahora que ha añadido la tabla de tipos de interés, puede buscar la tasa de interés adecuada para cada solicitante.



### Tarea 7: Añadir una etapa de búsqueda para buscar tipos de interés para los candidatos

Basándose en la puntuación de crédito de cada solicitante, desea buscar la tasa de interés apropiada. Siga estos pasos para añadir una etapa de búsqueda y

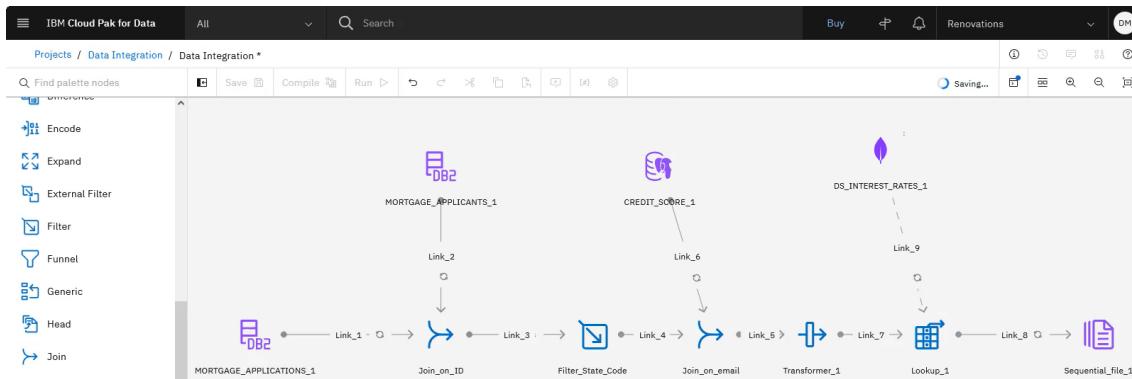
especificar el rango para los límites de puntuación de crédito inicial y final para cada tipo de interés:

1. En la sección *Etapas*, arrastre la etapa **Búsqueda** al lienzo y suelte el nodo en la línea de enlace entre los nodos *Transformer\_1* y *Sequential\_file\_1*.
2. Conecte el conector **DS\_INTEREST\_RATES\_1** a la etapa **Lookup\_1**.
3. Efectúe una doble pulsación en el nodo **DS\_INTEREST\_RATES\_1** para editar los valores.
4. Pulse la pestaña **Salida**.
  - a. Expanda la sección **Columnas** y pulse **Editar**.
  - b. Seleccione la columna **\_ID**.
  - c. Haga clic en el icono **Eliminar** para eliminar la columna **\_ID**.
  - d. Pulse **Aplicar y volver** para volver a los valores del nodo *DS\_INTEREST\_RATES\_1*.
  - e. Pulse **Guardar** para guardar los cambios en el nodo *DS\_INTEREST\_RATES\_1*.
5. Efectúe una doble pulsación en el nodo **Lookup\_1** para editar los valores.
6. Expanda la sección **Propiedades**.
  - a. Para el campo *Aplicar rango a columnas*, seleccione **CREDIT\_SCORE**. Se muestran los campos *Enlaces de referencia*, *Operador* y *Columna de rango*.
  - b. Para *Enlaces de referencia*, seleccione **Link\_9**.
- Nota:** El número de enlace puede ser diferente.
  - c. Para el primer *Operador*, seleccione **<=**.
  - d. Para la primera *Columna de rango*, seleccione **ENDING\_LIMIT**.
  - e. Para el segundo *Operador*, seleccione **>=**.
  - f. Para la segunda *Columna de rango*, seleccione **STARTING\_LIMIT**.
7. Pulse la pestaña **Salida**.
  - a. Expanda la sección **Columnas** y pulse **Editar**.
  - b. Seleccione las columnas **STARTING\_LIMIT** y **ENDING\_LIMIT**.
  - c. Haga clic en el icono **Eliminar** para eliminar estas columnas **STARTING\_LIMIT** y **ENDING\_LIMIT** innecesarias.

- d. Pulse **Aplicar y volver** para volver a los valores del nodo de *Lookup\_1*.
- e. Pulse **Guardar** para guardar los cambios en el nodo *Lookup\_1*.

### Comprueba tu progreso

La imagen siguiente muestra que el flujo de DataStage con la etapa Lookup se ha añadido. Ahora el flujo de DataStage se ha completado. La última tarea antes de ejecutar el flujo es especificar el nombre del archivo de salida.



### Tarea 8: Editar el nodo de archivo secuencial y ejecutar el flujo de DataStage

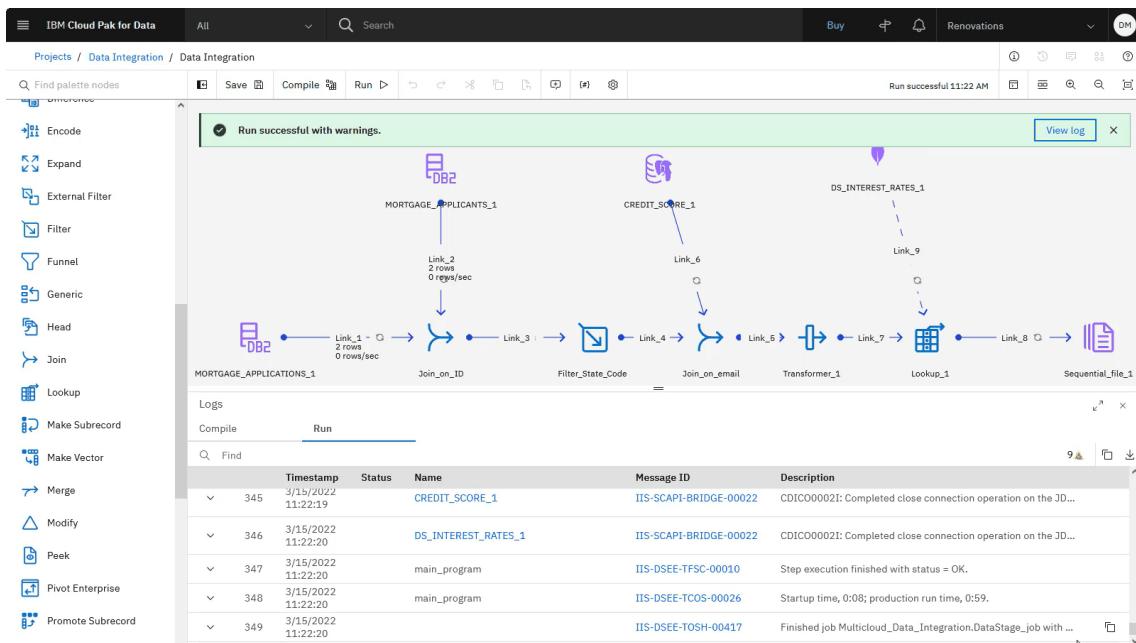
Siga estos pasos para editar el nodo de archivo secuencial para crear un archivo de salida final como un activo de datos en el proyecto y, a continuación, compile y ejecute el flujo de DataStage :

1. Efectúe una doble pulsación en el nodo **Sequential\_file\_1** para editar los valores.
2. Pulse la pestaña **Salida**.
3. Expanda la sección **Propiedades**.
4. Para el *Archivo de destino*, copie y pegue  
MORTGAGE\_APPLICANTS\_INTEREST\_RATES.CSV  
para el nombre de archivo.
5. Seleccione **Crear activo de datos**.
6. En el campo *La primera línea son los nombres de columna* , seleccione **True**.
7. Pulse **Guardar**.
8. Pulse **Ejecutar** , que compila y, a continuación, ejecuta el flujo de DataStage . El trabajo tarda aproximadamente 1 minuto en completarse.

- Pulse **Registros** en la barra de herramientas para ver el progreso del flujo. Es normal ver avisos durante la ejecución y, a continuación, ver que el flujo se ha ejecutado correctamente.

### Comprueba tu progreso

La imagen siguiente muestra que el flujo de DataStage se ha ejecutado correctamente. Ahora que el flujo de DataStage ha creado el archivo de salida, debe crear el catálogo donde publicará el archivo de salida.



### Tarea 9: Crear un catálogo para almacenar el elemento de datos publicado

Otros ingenieros de datos y analistas de negocios en Golden Bank necesitan acceso a las tasas de interés hipotecario. Con el plan IBM watsonx.data intelligence Lite, puede crear dos catálogos. Si ya tiene un catálogo, omita este paso. De lo contrario, realice los pasos siguientes para crear un catálogo en el que pueda publicar el conjunto de datos de tipos de interés.

- En el menú Navegación, seleccione **Catálogos > Ver todos los catálogos**.
- Si ve un *Catálogo de aprobación de hipotecas* en la página **Catálogos**, vaya a [Tarea 10: Ver la salida y publicar en un catálogo](#). De lo contrario, siga estos pasos para crear un catálogo nuevo:
- 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 **Impponer 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 que ya existe el catálogo de aprobación de hipoteca, puede publicar el archivo de salida en el catálogo.

The screenshot shows the IBM Cloud Pak for Data interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data', a search bar, and various account and workspace options. Below the header, the title 'Mortgage Approval Catalog' is displayed. Underneath the title, there are three tabs: 'Assets' (which is selected), 'Access control', and 'Settings'. A central message says 'You haven't added any assets yet.' followed by 'To add a new asset, click [Add to catalog](#)'. There's also a small circular icon with a plus sign and a document inside.

### **Tarea 10: Ver la salida y publicar en un catálogo**

Siga estos pasos para ver el archivo de salida en el proyecto y, a continuación, publíquelo en un catálogo:

1. En el **menú Navegación**, seleccione **Proyectos > Ver todos los proyectos**.
2. Abra el proyecto **Integración de datos**.
3. En la pestaña **Activos**, pulse **Datos > Activos de datos**.
4. Abra el archivo **MORTGAGE\_APPLICANTS\_INTEREST\_RATES.CSV**.
5. Desplácese para ver todas las columnas del conjunto de datos integrado con tipos de interés al final de cada entrada de datos.
6. Pulse **Integración de datos** en la ruta de navegación para volver al proyecto.
7. En la pestaña **Activos**, haga clic en el menú **Desbordamiento** al final de la fila del **MORTGAGE\_APPLICANTS\_INTEREST\_RATES.CSV** archivo y seleccione **Publicar en el catálogo**.

- a. Seleccione el **Catálogo de aprobación de hipoteca** (o su nombre de catálogo) en la lista y pulse **Siguiente**.
  - b. Seleccione la opción **Ir al catálogo después de publicarlo** y pulse **Siguiente**.
  - c. Revise los activos y pulse **Publicar**.
8. En el catálogo, busque Mortgage.
  9. Abra el archivo **MORTGAGE\_APPLICANTS\_INTEREST\_RATES.CSV**.
  10. Pulse el separador **Activo** para ver los datos.

### Comprueba tu progreso

La imagen siguiente muestra MORTGAGE\_APPLICANTS\_INTEREST\_RATES.CSV en el catálogo. Los datos que los prestamistas necesitan para tomar decisiones hipotecarias ya están disponibles.

The screenshot shows the IBM Cloud Pak for Data interface with the 'Mortgage Approval Catalog' selected. The 'Asset' tab is active, showing the schema and a preview of the data. The schema details 25 columns and 86 rows. The data preview shows various columns such as ID, INCOME, APPLIEDON, RESIDEN, YRS\_AT\_CURRENT\_ADD, YRS\_WITH\_CURRENT\_EMPL, NUMBER\_OF\_CARS, CREDITCARD, LOANS, LOAN\_AMO, and NAME, along with their respective data values for multiple rows.

ID	INCOME	APPLIEDON	RESIDEN	YRS_AT_CURRENT_ADD	YRS_WITH_CURRENT_EMPL	NUMBER_OF_CARS	CREDITCARD	LOANS	LOAN_AMO	NAME
110087	48799	YES	Public Housing	12	12	2	1490	1	8284	Abdi
110556	49150	NO	Owner Occupier	11	12	2	1041	1	7163	Alex
110533	49133	YES	Owner Occupier	11	12	2	1063	1	7218	Ansl
110050	48772	YES	Owner Occupier	12	12	2	1525	1	8372	Anto
110130	48831	YES	Living with paren	12	12	2	1449	1	8181	Audi
110226	48903	NO	Owner Occupier	12	12	2	1357	1	7952	Beni
110528	49129	YES	Owner Occupier	11	12	2	1068	1	7230	Bobi
110192	48878	NO	Owner Occupier	12	12	2	1389	1	8033	Brae
110355	49000	NO	Owner Occupier	11	12	2	1234	1	7643	Bran
110235	48910	NO	Private Renting	12	12	2	1348	1	7930	Carn
110433	49058	YES	Public Housing	11	12	2	1159	1	7457	Char
110421	49049	YES	Owner Occupier	11	12	2	1171	1	7485	Char
110009	48741	YES	Living with paren	12	12	2	1564	1	8470	Delfi
110048	48770	YES	Public Housing	12	12	2	1527	1	8377	Delti

Como ingeniero de datos de Golden Bank, ha integrado el solicitante de la hipoteca, la solicitud, la calificación crediticia y la información de puntuación de crédito, y ha publicado esos datos en un catálogo.