

*IBM SPSS Modeler Entity Analytics 17*  
*Guía del usuario*

**IBM**

**Nota**

Antes de utilizar esta información y el producto al que da soporte, lea la información del apartado "Avisos" en la página 59.

**Información de productos**

Esta edición se aplica a la versión 17, release 0, modificación 0 de IBM(r) SPSS(r) Modeler y a todos los releases y modificaciones posteriores hasta que se indique lo contrario en nuevas ediciones.

# Contenido

<b>Prefacio</b> . . . . .	<b>v</b>	Correlación de campos de entrada a características (nodo Generación de análisis de entidad) . . . . .	29
<b>Capítulo 1. Análisis de entidades</b> . . . . .	<b>1</b>	Visualización de las correlaciones de campos y orígenes de datos (nodo Generación de análisis de entidad) . . . . .	30
Acerca del análisis de entidades . . . . .	1	Resultados del nodo Generación de análisis de entidad . . . . .	31
Análisis de entidades y análisis predictivo . . . . .	2	Uso de IBM SPSS Modeler Entity Analytics con otros productos IBM SPSS . . . . .	32
<b>Capítulo 2. Entity Analytics con IBM SPSS Modeler</b> . . . . .	<b>5</b>	Tareas administrativas . . . . .	32
Uso de Entity Analytics con IBM SPSS Modeler. . . . .	5	Configuración de asignaciones de puertos . . . . .	33
Fase 1: Leer el origen de datos en SPSS Modeler . . . . .	6	Gestión de las credenciales de administrador para la base de datos del repositorio . . . . .	33
Fase 2: Crear el repositorio . . . . .	6	Desplazamiento del repositorio a un directorio de almacenamiento distinto . . . . .	34
Fase 3: Conectar SPSS Modeler con el repositorio . . . . .	7	Definir las propiedades de ruta para los campos fecha/hora e indicación de fecha y hora . . . . .	34
Fase 4: Correlacionar los campos de entrada con características del repositorio . . . . .	7	Ajustar los valores de tiempo de espera . . . . .	35
Fase 5: Exportar datos al repositorio y resolver identidades . . . . .	7	Ejecución de IBM SPSS Modeler Entity Analytics con el cliente de SPSS Modeler y SPSS Modeler Server en el mismo sistema Windows. . . . .	35
Fase 6: Analizar las identidades resueltas . . . . .	9	Purga de un repositorio de entidad . . . . .	35
Fase 7: Resolver nuevos casos con el repositorio . . . . .	10	Supresión de orígenes de datos no utilizados de un repositorio . . . . .	36
Fase 8: Generar alertas. . . . .	11	Eliminación de un repositorio de entidad . . . . .	36
<b>Capítulo 3. Tareas de análisis de entidades</b> . . . . .	<b>13</b>	Eliminación de un repositorio cuando no se puede conectar a él. . . . .	37
Acerca de las tareas . . . . .	13	<b>Capítulo 4. El análisis de entidades en acción</b> . . . . .	<b>39</b>
Creación de un repositorio de entidad (nodo de exportación de EA). . . . .	13	Acerca de este ejemplo . . . . .	39
El repositorio de entidad . . . . .	13	El modelo original . . . . .	39
Conexión con un origen de datos . . . . .	14	Con análisis de entidades. . . . .	42
Creación del repositorio . . . . .	14	Volcado de los datos de origen en el repositorio . . . . .	42
Correlación de campos de entrada a características (nodo de exportación de EA). . . . .	16	Lectura de las identidades resueltas . . . . .	43
Visualización de las correlaciones de campos (nodo de exportación de EA) . . . . .	18	Comparación de los resultados del análisis de entidades con el modelo original . . . . .	49
Configuración de un repositorio de entidad . . . . .	18	Resumen . . . . .	53
Visualización de las correlaciones de orígenes de datos . . . . .	19	<b>Apéndice. Propiedades de los scripts para IBM SPSS Modeler Entity Analytics</b> . . . . .	<b>55</b>
Mantenimiento de las características del repositorio. . . . .	20	Creación de scripts con IBM SPSS Modeler Entity Analytics . . . . .	55
Adición o modificación de una característica . . . . .	21	Propiedades comunes . . . . .	55
Anonimización de características del repositorio . . . . .	22	Propiedades entityanalytics_exportnode . . . . .	55
Mantenimiento de los tipos de entidad . . . . .	23	Propiedades entityanalytics_sourcenode . . . . .	56
Establecimiento del umbral para coincidencia de entidades. . . . .	25	Propiedades entityanalytics_processnode . . . . .	56
Reutilización de configuración de repositorio . . . . .	25	<b>Avisos</b> . . . . .	<b>59</b>
Guardar sus cambios de configuración . . . . .	26	Marcas comerciales . . . . .	60
Cerrar la ventana de configuración . . . . .	26	<b>Índice</b> . . . . .	<b>63</b>
Análisis de las identidades resueltas [nodo de origen Entity Analytics(EA)]. . . . .	26		
Selección de un origen de datos . . . . .	26		
Cambio de nombre de campos de datos . . . . .	27		
Establecimiento de la información de tipos para campos de datos. . . . .	27		
Adición de nodos a la ruta . . . . .	28		
Comparación de nuevos casos con el repositorio (nodo Generación de análisis de entidad) . . . . .	28		



---

## Prefacio

IBM® SPSS Modeler es el conjunto de programas de minería de datos de IBM Corp. orientado a las empresas. SPSS Modeler ayuda a las organizaciones a mejorar la relación con sus clientes y los ciudadanos a través de la comprensión profunda de los datos. Las organizaciones utilizan la comprensión que les ofrece SPSS Modeler para retener a los clientes más rentables, identificar las oportunidades de venta cruzada, atraer a nuevos clientes, detectar el fraude, reducir el riesgo y mejorar la prestación de servicios del gobierno.

La interfaz visual de SPSS Modeler invita a la pericia empresarial específica de los usuarios, lo que deriva en modelos predictivos más eficaces y la reducción del tiempo necesario para encontrar soluciones. SPSS Modeler ofrece muchas técnicas de modelado tales como predicciones, clasificaciones, segmentación y algoritmos de detección de asociaciones. Una vez que se crean los modelos, IBM SPSS Modeler Solution Publisher permite su distribución en toda la empresa a los encargados de tomar las decisiones o a una base de datos.

### Acerca de IBM Business Analytics

El software IBM Business Analytics ofrece información completa, coherente y precisa en la que confían los encargados de la toma de decisiones para mejorar el rendimiento comercial. Un conjunto integral de inteligencia empresarial, análisis predictivo, rendimiento financiero y gestión de estrategias y aplicaciones de análisis que ofrece una perspectiva clara, inmediata e interactiva del rendimiento actual y la capacidad de predecir resultados futuros. En combinación con extensas soluciones sectoriales, prácticas probadas y servicios profesionales, las organizaciones de cualquier tamaño pueden conseguir el máximo de productividad, automatizar las decisiones de forma fiable y alcanzar mejores resultados.

Como parte de esta familia, el software de análisis predictivo de IBM SPSS ayuda a las organizaciones a predecir eventos futuros y actuar proactivamente según esa información para lograr mejores resultados comerciales. Los clientes comerciales, gubernamentales y académicos de todo el mundo confían en la tecnología de IBM SPSS como ventaja ante la competencia para atraer, retener y hacer crecer a los clientes, reduciendo al mismo tiempo el fraude y el riesgo. Al incorporar el software de IBM SPSS en sus operaciones diarias, las organizaciones se convierten en empresas predictivas, capaces de dirigir y automatizar decisiones para alcanzar los objetivos comerciales y lograr una ventaja considerable sobre la competencia. Para obtener más información o contactar con un representante, visite <http://www.ibm.com/spss>.

### Asistencia técnica

Hay asistencia técnica disponible para los clientes de mantenimiento. Los clientes podrán ponerse en contacto con el servicio de asistencia técnica si desean recibir ayuda sobre la utilización de los productos de IBM Corp. o sobre la instalación en los entornos de hardware admitidos. Para ponerse en contacto con el servicio de asistencia técnica, consulte el sitio web de IBM Corp. en <http://www.ibm.com/support>. Tenga a mano su acuerdo de asistencia y esté preparado para identificarse a sí mismo y a su organización al solicitar ayuda.



---

# Capítulo 1. Análisis de entidades

---

## Acerca del análisis de entidades

IBM SPSS Modeler Entity Analytics incorpora una dimensión adicional al análisis predictivo de IBM SPSS Modeler. Mientras que el análisis predictivo trata de predecir comportamientos futuros a partir de datos del pasado, el análisis de entidades se centra en mejorar la coherencia de los datos actuales mediante la resolución de conflictos de identidades dentro de los propios registros. La identidad de un individuo, una organización, un objeto o cualquier otra entidad puede estar expuesta a ambigüedades. La resolución de identidades puede ser vital en diversos campos, entre los que se incluyen la gestión de la relación con el cliente, la detección de fraudes, la lucha contra el blanqueo de dinero y la seguridad nacional e internacional.

Imagínese que dispone de los siguientes registros de cliente de dos orígenes distintos, y que no está seguro de si hacen referencia a la misma o a distintas personas.

### Origen 1

N.º de registro: 70001  
Nombre: Jon Smith  
Dirección: 123 Main Street  
Referencia de impuestos: 555-00-1111  
Permiso de conducir: 0001133107  
Tarjeta de crédito: 10229127

### Origen 2

N.º de registro: 9103  
Nombre: JOHNATHAN Smith  
Fecha de nacimiento: 06/17/1934  
Teléfono: 555-1212  
Tarjeta de crédito: 10229128  
Correo electrónico: jls@mail.com  
Dirección IP: 9.50.18.77

No hay ninguna coincidencia exacta en los datos de los dos registros. Sin embargo, si introducimos un tercer origen, encontramos algunos atributos en común.

### Origen 3

N.º de registro: 6251  
Nombre: Jon Smith  
Teléfono: 555-1212  
Permiso de conducir: 0001133107  
Tarjeta de crédito: 10229132

El número del permiso de conducir une los registros de los orígenes 1 y 3, mientras que el número de teléfono une los orígenes 2 y 3. Por tanto, podemos estar bastante seguros de que los tres orígenes se refieren a la misma persona.

Pero, ¿qué sucede si la distinción no es tan sencilla? Podría haber muy poca información sobre la que basar la decisión. Considere los dos registros siguientes.

### Origen 4

N.º de registro: S45286  
Nombre: John T Smith Jr  
Dirección: 456 Main Street  
Teléfono: 703-555-2000  
Fecha de nacimiento: 03/12/1984

N.º de registro: S45287  
Nombre: John T Smith  
Dirección: 456 Main Street  
Teléfono: 703-555-2000  
Permiso de conducir: 009900991

Evidentemente, este Sr. Smith no es el mismo de los registros anteriores: las diferencias son suficientes para descartarlo. Sin embargo, sigue habiendo un problema. Dos registros diferentes, del mismo origen de datos, hacen referencia aparentemente a la misma persona. ¿Son registros duplicados? No podemos estar seguros a menos que encontremos otro registro relacionado que nos proporcione más información, quizás de un origen distinta.

#### Origen 5

N.º de registro: 769582-2  
Nombre: John T Smith Sr  
Dirección: 456 Main Street  
Teléfono: 703-555-2000  
Permiso de conducir: 009900991  
Fecha de nacimiento: 06/25/1959

Esto resuelve el problema. Los dos registros de el origen 4 no son duplicados; en realidad, son padre e hijo con el mismo nombre, que viven en la misma dirección y que tienen el mismo número de teléfono. Mediante un sistema manual, llevaría semanas encontrar el registro que resuelve el problema de identidad. Con un sistema de análisis de entidades automatizado, el tiempo de resolución se reduce considerablemente

---

## Análisis de entidades y análisis predictivo

Si todos sus datos procediesen de un único origen de registros completo y sin ambigüedades, IBM SPSS Modeler lo tendría bastante fácil a la hora de resolver conflictos de identidad. Solo con el análisis predictivo, podría leer sus datos en IBM SPSS Modeler, realizar sus procesos y obtener resultados fiables.

Sin embargo, las situaciones reales son bastante distintas. Normalmente, los datos no están completos, con frecuencia son ambiguos y están repartidos por diferentes orígenes de datos que registran muchos atributos distintos con campos que se solapan. Parte del valor del análisis de entidades reside en que reúne datos de orígenes distintos en una única área central de almacenamiento, conocida como **repositorio**. El sistema de análisis de entidades examina a continuación los datos minuciosamente para resolver conflictos, y añade un identificador exclusivo a los registros que provienen de la misma persona u organización.

La siguiente tabla ilustra las diferencias entre los dos tipos de análisis.

*Tabla 1. Diferencias entre análisis predictivo y de entidades.*

Característica	Análisis predictivo	Análisis de entidades
Tipos de datos de entrenamiento	Basado en conjuntos y rangos numéricos relativamente pequeños	Puede explotar conjuntos grandes (campos sin tipo) como, por ejemplo, nombres y direcciones
Tamaño de datos de entrenamiento	Normalmente ignora los conjuntos grandes (campos sin tipo)	Todos los datos

Tabla 1. Diferencias entre análisis predictivo y de entidades (continuación).

Característica	Análisis predictivo	Análisis de entidades
Generalización	Un algoritmo generaliza todos los datos de entrenamiento para generar un modelo conciso	Los datos permanecen en estructuras disponibles para la coincidencia de entidades y la detección de relaciones
Detección de fraude	Los registros se marcan como potencialmente fraudulentos si tienen características típicas de solicitudes fraudulentas	Los registros se marcan como potencialmente fraudulentos si guardan relación con registros fraudulentos o si proceden del mismo sujeto pero con identidades distintas



---

## Capítulo 2. Entity Analytics con IBM SPSS Modeler

---

### Uso de Entity Analytics con IBM SPSS Modeler

Sospecha que puede tener algún problema de identidad con los datos. Por ejemplo, algunos individuos pueden aparecer más de una vez o puede parecer que distintos individuos se han fusionado o faltan. ¿Cómo puede ayudarle IBM SPSS Modeler Entity Analytics a solucionarlo? A continuación se sugiere un procedimiento, aunque es posible que tenga que hacer variaciones para ajustarlo a sus necesidades.

- Lea los datos de origen en IBM SPSS Modeler
- Cree un repositorio preparado para almacenar los datos
- Conecte IBM SPSS Modeler al repositorio
- Correlacione los campos de datos a características del repositorio
- Exporte los datos al repositorio y resuelva las identidades
- Analice las identidades resueltas
- Resuelva los casos nuevos con el repositorio
- Genere las alertas necesarias (lote o tiempo real)

En este punto, debe conocer algo del funcionamiento de IBM SPSS Modeler. IBM SPSS Modeler es una herramienta muy intuitiva, que se basa en la representación gráfica de una ruta de datos que fluyen a través de un número de nodos. Cada nodo representa una fase particular del flujo de trabajo.

IBM SPSS Modeler ofrece una gran variedad de nodos, que cubren todas las funciones de minería de datos convencionales. IBM SPSS Modeler Entity Analytics añade nodos para utilizarlos específicamente en análisis de entidades. Son el nodo de exportación de EA, el nodo de origen Entity Analytics(EA) y el nodo de proceso Generación de análisis de entidad.

La figura a continuación ilustra el proceso.

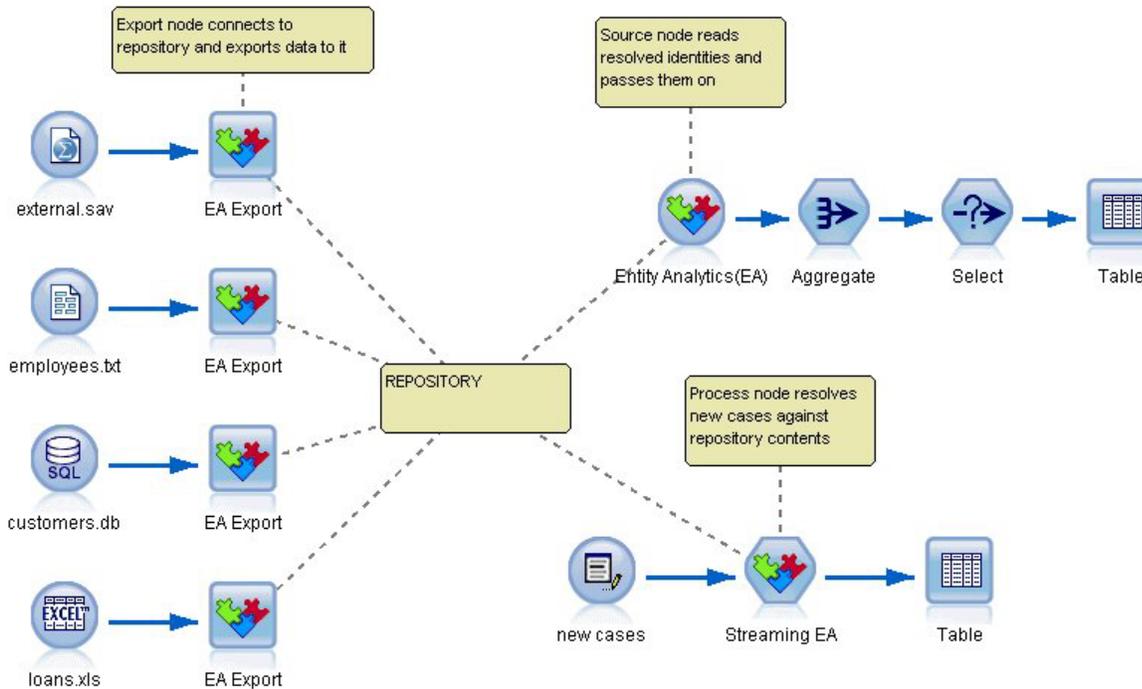


Figura 1. Proceso de análisis de entidades

## Fase 1: Leer el origen de datos en SPSS Modeler

Su primera tarea es leer sus datos en SPSS Modeler a través de uno o más nodos de origen, representado en SPSS Modeler por icono redondo.

Los datos pueden estar en cualquier formato compatible con SPSS Modeler, como archivos de texto, tablas de bases de datos, hojas de cálculo, archivos XML, etc., pero cada formato requiere un nodo de origen de SPSS Modeler correspondiente. La ilustración muestra un nodo de origen de base de datos.

Cada archivo de origen de datos debe tener un campo que identifique exclusivamente cada registro. Si un origen de datos no tiene dicho campo, puede añadir uno fácilmente en SPSS Modeler. Consulte el tema “Adición de un identificador de registro exclusivo” en la página 14 para obtener más información.

Consulte el tema “Conexión con un origen de datos” en la página 14 para obtener más información.

**Nota:** Los datos de caracteres no latinos no están soportados. Si los datos están formados por una mezcla de registros en juegos de caracteres latinos (de Europa Occidental, por ejemplo) y no latinos, sólo se resolverán las entradas de los datos latinos.

## Fase 2: Crear el repositorio

El punto central de sus esfuerzos de análisis de entidades es el repositorio, el área de almacenamiento central donde reúne todos sus registros de datos.

Para crear un repositorio, conecte el origen de datos a un nodo de exportación de EA, representado por el icono cuadrado.

Desde el nodo de exportación, puede crear un repositorio nuevo (o seleccionar uno existente) que esté preparado para recibir los datos exportados.

El proceso de creación de un repositorio se describe con detalle más adelante. Consulte el tema “Creación de un repositorio de entidad (nodo de exportación de EA)” en la página 13 para obtener más información.

**Nota:** Si lo va a ejecutar en modalidad de servidor remoto, debe crear el repositorio en la máquina de Modeler Server (es decir, debe estar conectado a Modeler Server desde Modeler Client al crear el repositorio, de modo que el repositorio de EA se cree en la máquina servidor).

Una vez creado el repositorio, puede mantener su contenido de diversas formas. Consulte el tema “Configuración de un repositorio de entidad” en la página 18 para obtener más información.

### **Fase 3: Conectar SPSS Modeler con el repositorio**

Una vez creado el repositorio, conéctelo a la ruta de SPSS Modeler.

Consulte el tema “Opciones de repositorio de entidad” en la página 16 para obtener más información.

### **Fase 4: Correlacionar los campos de entrada con características del repositorio**

Los orígenes de datos pueden contener diversos tipos de información de la entidad. Ciertos tipos de información son comunes para la mayoría de orígenes de datos, mientras que otros son específicos de un origen de datos particular. En un repositorio de entidad, estos diferentes tipos de información se llaman **características**. El repositorio ofrece un número de características estándar, pero también puede crear las suyas propias.

Una característica de repositorio es un tipo de información individual que se puede utilizar con un origen de datos de entidad. Algunas características (por ejemplo, Nombre, Apellidos, Fecha de nacimiento, etc.) se pueden utilizar con muchos orígenes de datos diferentes, mientras que otras características son específicas de un origen de datos concreto. En términos generales, una característica equivale a un campo de un registro de datos o a una columna de una tabla de base de datos.

Una vez creado el repositorio y conectado a él, designe un campo de sus datos de entrada como el campo de **clave exclusiva**, que se utilizará en los análisis subsiguientes. Correlacione también los campos de datos de entrada a sus características correspondientes en el repositorio. La correlación con características predefinidas indica al repositorio de entidad qué campos debe comparar y, muy importante, cómo compararlos. El nodo de exportación de EA dispone de una tabla de correlaciones donde puede crear las correlaciones.

Consulte el tema “Correlación de campos de entrada a características (nodo de exportación de EA)” en la página 16 para obtener más información.

### **Fase 5: Exportar datos al repositorio y resolver identidades**

Cada nodo de origen de datos necesita su propio nodo de exportación de EA, de modo que si sus datos están dispersos en diferentes orígenes, puede que tenga una ruta con varios orígenes de datos, cada uno de ellos conectado a un nodo Exportar EA distinto. Consulte el tema “Uso de Entity Analytics con IBM SPSS Modeler” en la página 5 para obtener más información.

Puede elegir entre leer los registros desde una, algunas o todos los orígenes de datos si tiene más de una. El sistema Entity Analytics analiza los registros seleccionados y añade un campo de identificador denominado \$EA\_ID a cada uno. Donde dos o más registros relacionados con identidades previamente ambiguas se pueden resolver, los identificadores añadidos a esos registros son exclusivos en todo el repositorio. El sistema también añade un campo para mostrar el origen de datos desde el que se origina el registro.

Conecte cada nodo de origen de datos a su propio nodo de exportación de EA, correlacione los campos de entrada a las características de repositorio y, a continuación, ejecute la ruta para exportar datos desde SPSS Modeler al repositorio para resolver conflictos de identidades, en una sola operación. A modo de ilustración, imagine que tiene los siguientes registros en cuatro orígenes de datos distintos.

#### Datos externos

Tabla 2. Datos externos

Nombre	Teléfono	Riesgo de crédito
Mike	555-1234	560
Joe	555-4567	780

#### Empleados

Tabla 3. Empleados

Nombre	Dirección	Teléfono
Michael	1234 5th Street	555-1234
Fred	543 1st Avenue	555-9876

#### Clientes

Tabla 4. Clientes

Nombre	Dirección	Ahorros
Susan	1234 5th Street	\$1234
Joe	777 Oak Street	5€

#### Préstamos

Tabla 5. Préstamos

Nombre	Dirección	Teléfono	Préstamo
Sue	1234 5th Street	555-1234	10.000 USD
Joseph	777 Oak Street	555-4567	\$50,000

Como hemos podido ver, usted exporta cada origen de datos en el repositorio por turnos. Mientras lo hace, el repositorio actualiza la resolución de cada registro. En el repositorio, cada registro va precedido de un campo de identificador (llamado \$EA-ID) y un campo de indicador de origen (llamado \$EA-SRC), que muestra el origen de datos donde se originó el registro. Así que en nuestro ejemplo, una vez exportadas los cuatro orígenes de datos, el contenido del repositorio tiene este aspecto.

Tabla 6. Ejemplo de contenido del repositorio tras la fase de exportación.

\$EA-ID	\$EA-SRC	Nombre	Teléfono	Dirección	Riesgo de crédito	Ahorros	Préstamo
1	Empleados	Michael	555-1234	1234 5th St			
1	External	Mike	555-1234		560		
2	Clientes	Joe		777 Oak St		5€	
2	External	Joe	555-4567		780		
2	Préstamos	Joseph	555-4567	777 Oak St			\$50,000
3	Empleados	Fred	555-9876	543 1st Ave			

Tabla 6. Ejemplo de contenido del repositorio tras la fase de exportación (continuación).

\$EA-ID	\$EA-SRC	Nombre	Teléfono	Dirección	Riesgo de crédito	Ahorros	Préstamo
4	Clientes	Susan		1234 5th St		\$1234	
4	Préstamos	Sue	555-1234	1234 5th St			10.000 USD

El sistema Entity Analytics ha determinado que *Mike* en el conjunto de datos *Externos* es la misma persona que *Michael* en el conjunto de datos *Empleados* basándose en el número de teléfono común, y le ha asignado el ID 1.

El caso de *Joe* en el conjunto de datos *Externos* es un poco más complicado. ¿Es la misma persona que *Joe* en *Clientes*? No se puede saber solo con esos dos orígenes de datos, pero hay un tercer origen, *Préstamos*, en la que hay un *Joseph*. Ahora tenemos una coincidencia: el número de teléfono de *Joseph* es el mismo que el de *Joe* en el conjunto de datos *Externos*. Teniendo esto en cuenta, el sistema determina que son la misma persona y les otorga el identificador 2.

No hay registros múltiples para *Fred*, por lo que recibe el ID 3. *Susan* de *Clientes* es identificada como la misma persona que *Sue* de *Préstamos* puesto que tienen la misma dirección, así que recibe el ID 4.

*Note:* este es un ejemplo de una coincidencia optimista con fines ilustrativos. Podría elegir un conjunto de reglas más pesimista, de modo que un simple nombre y un número de teléfono o una dirección por sí mismos no constituirían una coincidencia exacta y no se asignaría a ambos registros el mismo identificador.

## Fase 6: Analizar las identidades resueltas

Una vez resueltos los conflictos de identidad en el repositorio, es posible continuar analizando y procesando los resultados. Por ejemplo, si la existencia de registros duplicados para la misma identidad le hace sospechar de posibles actividades fraudulentas, puede que quiera generar un informe con todas las duplicidades.

Lo primero es crear un nodo de origen Entity Analytics (EA) y enlazarlo con el repositorio.

La salida común del nodo consta de los campos siguientes.

- El campo de identificador añadido por el sistema (*\$EA-ID* en el ejemplo de la fase 5)
- El campo de indicador de origen añadido por el sistema (*\$EA-SRC* en el ejemplo de la fase 5)
- El campo de clave exclusiva designado en la fase 4

Además, si está viendo las relaciones, se produce la salida siguiente. Consulte el tema “Selección de un origen de datos” en la página 26 para obtener más información.

- El grado de separación entre entidades (*\$EA-DEGREE*)
- El campo padre (*\$EA-PARENT*)
- El campo hijo (*\$EA-CHILD*)
- La regla que identifica la relación (*\$EA-RULE*)

Para ver el resultado en SPSS Modeler, puede adjuntar un nodo de resultado de SPSS Modeler, como un nodo Tabla o un nodo Informe, y ejecutar esta parte de la ruta. Si debe resumir el resultado, que puede ser muy extenso, puede incluir todas las operaciones con registros como Agregar o Seleccionar nodos.

El nodo de origen Entity Analytics(EA) se describe con detalle más adelante. Consulte el tema “Análisis de las identidades resueltas [nodo de origen Entity Analytics(EA)]” en la página 26 para obtener más información.

## Fase 7: Resolver nuevos casos con el repositorio

Ha resuelto los conflictos de identidad en todos los registros de sus orígenes de datos. ¿Pero qué sucede cuando quiere comparar un conjunto de nuevos registros para comprobar cómo se relacionan con lo que ya conoce y así obtener una mejor puntuación? Aquí entra en juego el nodo Generación de análisis de entidad.

En primer lugar, añada un nuevo nodo de origen de datos de SPSS Modeler para leer los nuevos datos en la ruta. A continuación, conecte este nodo de origen a un nodo Generación de análisis de entidad. Para ver los resultados, debe añadir un nodo de Tabla como antes.

Al ejecutar esta parte de la ruta, el nodo Generación de análisis de entidad lee cada registro nuevo y lo compara con el contenido del repositorio. Si encuentra registros coincidentes en el repositorio, el nodo Generación de análisis de entidad genera resultados de todos los registros coincidentes junto con el registro nuevo, al que añade los campos de ID e indicador de origen. Si no se encuentra ninguna coincidencia, el nodo de proceso genera únicamente el registro nuevo, con los campos de ID e indicador de origen añadidos.

Para ilustrar esto, imagine que el repositorio consta actualmente del contenido generado por el nodo de origen Entity Analytics (EA). Consulte el apartado Tabla 6 en la página 8.

A continuación, se reciben los siguientes registros nuevos. ¿Hacen referencia a alguien que ya conozcamos?

Tabla 7. Nuevos registros para almacenar.

Nombre	Dirección	Teléfono	Préstamo
Suzan	1234 5th Street	555-1234	100.000 \$
Mark	888 9th Ave	555-9999	60.000 dólares

Al comparar los nuevos datos con el contenido existente en el repositorio, el nodo Generación de análisis de entidad empareja el primer registro nuevo con la persona que tiene el identificador 4 en los registros existentes. Sin embargo, no hay coincidencias para el segundo registro nuevo, por lo que recibe un nuevo identificador exclusivo: 5.

El nodo Generación de análisis de entidad añade los campos de identificador e indicador de origen, y genera los resultados de los nuevos registros junto con todos los registros coincidentes. El resultado será el siguiente.

Tabla 8. Resultados del nodo Generación de análisis de entidad.

\$EA-ID	\$EA-SRC	Nombre	Teléfono	Dirección	Riesgo de crédito	Ahorros	Préstamo
4	Cliente	Susan		1234 5th St		\$1234	
4	Préstamo	Sue	555-1234	1234 5th St			10.000 USD
4	Nuevo préstamo	Suzan	555-1234	1234 5th Street			100.000 \$
5	Nuevo préstamo	Mark	555-9999	888 9th Ave			60.000 dólares

Esta salida puede agregarse mediante el identificador del análisis de entidades como clave de agregación y transferirse a otros nodos posteriores en la ruta para un proceso adicional.

El nodo Generación de análisis de entidad se describe con detalle más adelante.

## Fase 8: Generar alertas

De nuevo, la actividad potencialmente sospechosa puede volverse aparente. En este caso, la persona con el identificador 4 ya tiene un préstamo de 10.000 \$ y ahora solicita otro diez veces mayor, con un nombre ligeramente distinto. Evidentemente, es algo perfectamente aceptable y no tiene por qué haber intenciones fraudulentas. Sin embargo, si dicha actividad resulta sospechosa de acuerdo con las reglas de su empresa, puede que merezca la pena comprobarlo.

Podría, por ejemplo, adjuntar y ejecutar un nodo Tabla o un nodo Informe de SPSS Modeler e imprimir el contenido de la ventana de resultados para que alguien lo lea y genere alertas manualmente. Otra posibilidad sería analizar el resultado del nodo Generación de análisis de entidad con un modelo de evaluación de riesgos previamente creado con IBM SPSS Modeler, para conseguir un conjunto de resultados que reflejen las reglas de su empresa con mayor precisión. Asimismo, puede exportar el resultado a una base de datos o algún otro medio de resultados para su procesamiento. Con IBM SPSS Modeler, dispone de una amplia gama de opciones que se ajustan a sus necesidades concretas.



---

## Capítulo 3. Tareas de análisis de entidades

---

### Acerca de las tareas

En esta sección se describen las siguientes tareas de análisis de entidades.

- Creación de un repositorio de entidad
- Configuración de un repositorio de entidad
- Análisis las identidades resueltas
- Resolución de casos nuevos con el repositorio de entidad
- Purga de un repositorio de entidad
- Eliminación de un repositorio de entidad
- Uso del análisis de entidades con otros productos IBM SPSS
- Administración del análisis de entidades

---

### Creación de un repositorio de entidad (nodo de exportación de EA)

El proceso de creación de un repositorio de entidad consta de las tareas siguientes.

1. Conectar con un origen de datos. Consulte el tema “Conexión con un origen de datos” en la página 14 para obtener más información.
2. Crear el repositorio. Consulte el tema “Creación del repositorio” en la página 14 para obtener más información.
3. Correlacionar campos de entrada en el origen de datos a características del repositorio. Consulte el tema “Correlación de campos de entrada a características (nodo de exportación de EA)” en la página 16 para obtener más información.

Una vez creadas las correlaciones, puede mostrarlas en el origen de datos actual o en todos los orígenes de datos conocidos para el repositorio. Consulte el tema “Visualización de las correlaciones de campos (nodo de exportación de EA)” en la página 18 para obtener más información.

**Nota:** A partir de la versión 16, SPSS Entity Analytics admite repositorios con el producto IBM DB2. Dado que un repositorio es específico de una versión de SPSS Modeler y no puede importarse de una versión anterior, si tiene un repositorio existente y va a actualizar a SPSS Entity Analytics versión 16, tendrá que volver a crear ese repositorio en la nueva base de datos de DB2.

### El repositorio de entidad

El repositorio ofrece un área de almacenamiento central que actúa como caché de datos de toda la información de la entidad. Dado que el repositorio está activo, su estado es único, y por tanto no existe el concepto de versión para un repositorio de entidad. El repositorio contiene el estado actual de todos los datos de entrada y puede alcanzar un gran tamaño.

Los contenidos del repositorio se pueden mantener mediante una interfaz gráfica fácil de utilizar. Consulte el tema “Configuración de un repositorio de entidad” en la página 18 para obtener más información.

*Importante:* a partir de la versión 16, IBM SPSS Modeler Entity Analytics da soporte a repositorios en el producto IBM DB2; las versiones anteriores de SPSS Entity Analytics dan soporte a repositorios alojados en IBM solidDB. Si tiene un repositorio existente de solidDB deberá volver a crearlo en la nueva base de datos DB2 si actualiza a SPSS Entity Analytics versión 16 o posterior.

*Nota:* la versión de IBM SPSS Modeler Entity Analytics proporcionada con IBM SPSS Modeler Premium sólo da soporte a un único repositorio alojado en el producto IBM DB2 que se encuentra en el mismo paquete que SPSS Entity Analytics. Con esta versión, debe eliminar el repositorio existente antes de crear uno nuevo. Tiene a su disposición una actualización con licencia independiente de SPSS Entity Analytics (conocido como IBM SPSS Modeler Entity Analytics Unleashed) que permite que más de un repositorio coexistan en el mismo sistema; cada repositorio puede contener más de 10 millones de filas y utilizar más de cuatro núcleos de procesadores. Póngase en contacto con su representante local de asistencia técnica de IBM para obtener información detallada.

## Conexión con un origen de datos

En primer lugar, se leen los datos de origen en SPSS Modeler mediante un nodo de origen.

Para conectar con un origen de datos

1. En la pestaña Orígenes, en la paleta de nodos de la parte inferior de la ventana principal de SPSS Modeler, pulse dos veces en el icono correspondiente al tipo de datos de origen. De este modo se añade un nodo de origen al lienzo de pantalla.
2. En el lienzo de pantalla, pulse dos veces en el icono para abrir su cuadro de diálogo.
3. En el campo Archivo, introduzca la ubicación y el nombre del archivo de datos de origen.
4. Complete el resto del cuadro de diálogo según sea necesario (pulse en el botón Ayuda para más información) y pulse Aceptar.
5. Si el archivo de datos de origen no tiene un campo que identifique exclusivamente cada registro, añada uno mediante el nodo Derivar. Consulte el tema “Adición de un identificador de registro exclusivo” para obtener más información.

*Nota:* Los datos de caracteres no latinos no están soportados. Si los datos están formados por una mezcla de registros en juegos de caracteres latinos (de Europa Occidental, por ejemplo) y no latinos, sólo se resolverán las entradas de los datos latinos.

## Adición de un identificador de registro exclusivo

Cada archivo de origen de datos introducido en el repositorio de entidad debe tener un campo que identifique exclusivamente cada registro. Si un archivo de origen de datos no cuenta con dicho campo, puede añadirlo mediante el nodo Derivar de SPSS Modeler.

Para añadir un identificador de registro exclusivo a un archivo de origen de datos

1. En el lienzo de pantalla, pulse en el nodo de origen que añadió en la tarea anterior.
2. En la pestaña **Operaciones con campos** de la paleta de nodos, pulse dos veces en el icono **Derivar** para adjuntar un nodo Derivar al nodo de origen.
3. En el lienzo de pantalla, pulse dos veces en el nodo Derivar para abrir su cuadro de diálogo.
4. En el campo **Derivar**, cambie el nombre predeterminado por un nombre distintivo (p. ej.: **ID**) para el campo de identificador que está añadiendo.
5. Asegúrese de que el campo **Derivar como** está establecido como **Fórmula**
6. Establezca el **tipo de Campo** como **Continuo**.
7. En el cuadro de texto **Fórmula**, introduzca **@INDEX** y pulse en **Aceptar**.

## Creación del repositorio

Es necesario crear un repositorio para almacenar todos los datos entrantes.

*Nota:* Si lo va a ejecutar en modalidad de servidor remoto, debe crear el repositorio en la máquina de Modeler Server (es decir, debe estar conectado a Modeler Server desde Modeler Client al crear el repositorio, de modo que el repositorio de EA se cree en la máquina servidor).

Para crear un repositorio

1. En la pestaña Exportar de la paleta de nodos de SPSS Modeler, ponga un nodo de exportación de EA en el lienzo de rutas.

*Note:* si es la primera vez que crea un repositorio, utilice un nodo de exportación de EA y conéctelo al nodo de origen de SPSS Modeler que contiene los datos que desea introducir en el repositorio (o al nodo Derivar, si ha añadido uno para obtener un campo de identificador exclusivo). Para conectar los nodos, haga lo siguiente.

- a. Pulse con el botón derecho en el nodo de origen de SPSS Modeler.
  - b. Seleccione Conectar.
  - c. Pulse en el nodo de exportación de EA.
2. Pulse dos veces en el nodo de exportación de EA para abrir su cuadro de diálogo.
  3. Pulse en la lista del **repositorio de entidad**.
  4. Pulse en <Examinar...> para que se muestre el cuadro de diálogo Repositorios de entidad.
  5. En el cuadro de diálogo Repositorios de entidad, pulse en el campo Nombre del repositorio.
  6. Elija <Crear repositorio nuevo...> para que se muestre el asistente de creación de repositorios.

## Asistente de creación de repositorios

### Paso 1

Aquí debe seleccionar si desea crear un repositorio local, mediante el producto IBM DB2 que se encuentra en el mismo paquete que IBM SPSS Modeler Entity Analytics, o utilizar una base de datos externa para el repositorio.

**Crear repositorio local.** Especifique los detalles de nombre de usuario y contraseña del administrador para la base de datos IBM DB2 que alojará el repositorio que está creando. Confirme la contraseña y pulse en **Siguiente**.

**Nota:** No puede utilizar un guión ni un carácter de subrayado en el nombre de usuario.

Las credenciales que deben utilizarse en la base de datos IBM DB2 dependen del sistema operativo. Los usuarios de UNIX deben utilizar el nombre de usuario g2user y la contraseña G2password.

Las tareas de administración de repositorio en los nodos Entity Analytics (por ejemplo, la creación o destrucción de un repositorio) necesitan permisos adicionales. En UNIX, el usuario que ha iniciado sesión en IBM SPSS Modeler Server debe ser el usuario root o g2user y pertenecer al grupo db2iadm1. En Windows el usuario que inicia sesión en IBM SPSS Modeler Server también debe ser miembro del grupo DB2ADMNS para poder llevar a cabo la administración de repositorio.

Si posteriormente necesita cambiar las credenciales de administrador, puede hacerlo usando el editor de línea de comandos de la base de datos. Consulte el tema “Gestión de las credenciales de administrador para la base de datos del repositorio” en la página 33 para obtener más información.

**Nota:** Solo es posible una combinación de nombre de usuario y contraseña. Todos los usuarios que inician sesión en el repositorio comparten el mismo nombre de usuario y contraseña.

**Añadir repositorio externo.** Utilice esta opción si desea utilizar una base de datos externa para alojar el repositorio. Escriba la ubicación del archivo .ini de la base de datos en el campo **Seleccione el archivo .ini del repositorio** y pulse en **Siguiente**.

### Paso 2

**Nombre de repositorio nuevo.** Escriba un nombre no utilizado anteriormente para el nuevo repositorio.

**Importar configuración de.** (únicamente repositorio local) Si desea basar la configuración en la de un repositorio existente, elija aquí el repositorio; de lo contrario, elija **Valor predeterminado**. Consulte el tema “Configuración de un repositorio de entidad” en la página 18 para obtener más información.

Si elige un repositorio existente, introduzca los detalles de conexión si difieren de los que introdujo en la pantalla anterior.

Pulse en **Aceptar** para crear el repositorio nuevo y que se muestre el cuadro de diálogo Casos de resolución de entidades, desde el que puede conectarse al repositorio.

## Opciones de repositorio de entidad

El cuadro de diálogo Repositorios de entidad contiene varias opciones para crear, conectar, configurar y mantener un repositorio de entidad.

**Conectar al repositorio.** Utilice estas opciones para crear un nuevo repositorio de entidad o para conectar con uno existente.

- **Nombre del repositorio.** Muestra el repositorio de entidad actual si existe alguno. Si hay más de un repositorio, para elegir uno distinto selecciónelo de la lista.  
Para crear un repositorio nuevo, seleccione <**Crear repositorio nuevo...**>. Se iniciará un asistente que le guiará a través del proceso de creación.
- **Nombre de usuario.** Especifique un nombre de usuario válido para el repositorio seleccionado.
- **Contraseña.** La contraseña para este nombre de usuario.
- **Conectar.** Pulse para conectarse al repositorio actual.

**Gestionar repositorio.** La tabla enumera los orígenes de datos cargados en el repositorio actual (al que está conectado) y muestra el número de registros en cada origen de datos.

- **Actualizar.** Actualiza el origen de datos y el tamaño de información en la tabla (por ejemplo, si ha añadido un nuevo origen de datos o ha cambiado el tamaño de una existente).
- **Purgar todo.** Elimina todos los datos de origen del repositorio, pero mantiene todos los detalles de configuración. Puede utilizar esta opción si la información de configuración aún le es útil pero desea eliminar todos los registros de datos del repositorio. Consulte el tema “Purga de un repositorio de entidad” en la página 35 para obtener más información.
- **Suprimir no utilizados.** Elimina los datos de origen resaltados del repositorio pero mantiene todos los detalles de configuración. Consulte el tema “Supresión de orígenes de datos no utilizados de un repositorio” en la página 36 para obtener más información.
- **Cambiar nombre de origen.** Abre un cuadro de diálogo en el que se puede cambiar el nombre del origen de datos resaltado.

**Nota:** Cambia el nombre del origen de datos dentro del repositorio; deberá volver a seleccionar este nuevo nombre de origen de datos en cualquier nodo de generación o exportación existente desde donde se hace referencia a él.

**Destruir el repositorio completo.** Elimina completamente los contenidos actuales del repositorio y los detalles de configuración. Consulte el tema “Eliminación de un repositorio de entidad” en la página 36 para obtener más información.

**Configurar repositorio.** Muestra una ventana donde puede configurar el repositorio actual. Consulte el tema “Configuración de un repositorio de entidad” en la página 18 para obtener más información.

## Correlación de campos de entrada a características (nodo de exportación de EA)

El repositorio ofrece varias características predefinidas estándar. Orígenes de datos distintos pueden utilizar nombres de campo distintos (por ejemplo, **Dirección1** o **Datos de dirección 1**) para tipos de

información que corresponden a la misma característica. Para evitar duplicidades, es necesario correlacionar campos de orígenes de datos entrantes a características del repositorio específicas. No es necesario correlacionar todos los campos en el conjunto de datos, solo aquellos que es probable que correspondan a la misma característica en otro conjunto de datos.

Si un origen de datos utiliza campos correspondientes a otros tipos de información no predefinidos en el repositorio, puede crear nuevas características en la ventana Configuración de repositorio. Consulte el tema “Configuración de un repositorio de entidad” en la página 18 para obtener más información.

Para correlacionar campos de entrada a características

1. Adjunte un nodo de exportación de EA a un nodo de origen de datos en el lienzo de rutas. Cada nodo de origen de datos que utilice deberá estar adjuntado a su propio nodo de exportación de EA.
2. Abra el nodo de exportación de EA para que se muestre la pestaña Entradas, que contiene opciones para correlacionar los campos de entrada. Consulte el tema “Opciones de entrada del repositorio para la correlación” para obtener más información.
3. En el nodo de exportación de EA, seleccione la pestaña Repositorio para ver las asignaciones de correlaciones del origen de datos actual o, si utiliza más de uno, de todos los orígenes de datos.
4. Para guardar un conjunto de asignaciones de correlaciones (por ejemplo, para utilizar un nodo de exportación para un origen de datos diferente), pulse en **Exportar correlación**.

Cuando haya terminado de correlacionar el primer nodo de origen de datos, repita el proceso con el resto de nodos de origen de datos que desee utilizar.

### Opciones de entrada del repositorio para la correlación

La pestaña Entradas contiene las opciones para correlacionar campos de orígenes de datos a características del repositorio listas para exportar al repositorio. Cree las asignaciones de correlaciones en esta pestaña (opcionalmente, puede pulsar en la pestaña Repositorio para ver la correlación de otros orígenes de datos) y pulse en **Ejecutar** para exportar los datos al repositorio.

Si ya ha guardado un conjunto de correlaciones en un archivo XML, puede usarlas pulsando en **Importar correlación**.

**Moda.** Deje la selección predeterminada, **Añadir al repositorio**, si desea añadir los registros del archivo de origen al contenido existente del repositorio. Si desea borrar el contenido del repositorio pero mantener la información de configuración antes de añadir los registros de origen, seleccione **Purgar repositorio antes de exportar**.

**Repositorio de entidad.** Muestra el repositorio de entidad actual si existe alguno. Si hay más de un repositorio, para elegir uno distinto selecciónelo de la lista. Para crear un nuevo repositorio, elija **<Examinar...>** para ver un cuadro de diálogo en el que puede crear el repositorio. Consulte el tema “Opciones de repositorio de entidad” en la página 16 para obtener más información.

**Correlacionar a tipo de entidad.** Una lista de tipos de entidades (es decir, conjuntos de características) definidos en el repositorio. Elija uno de la lista o **<Añadir nuevo tipo de entidad...>** para ver la ventana de configuración de repositorio, donde puede definir un nuevo tipo de entidad. Consulte el tema “Configuración de un repositorio de entidad” en la página 18 para obtener más información.

**Etiqueta de origen.** Una lista de etiquetas que indica los orígenes de datos conocidos actualmente en el repositorio. Elija una de la lista o **<Añadir nueva etiqueta de origen...>** para crear una etiqueta de un nuevo origen de datos.

**Clave exclusiva.** (requerida) El campo de entrada que se utiliza para los identificadores exclusivos de los registros de datos.

**Tabla de correlaciones.** En esta tabla puede correlacionar cada campo de entrada a una característica correspondiente en el repositorio. Si no existe ninguna característica adecuada en el tipo de entidad seleccionado, puede crear una característica nueva aquí.

- **Campo.** El conjunto de campos de entrada en el origen de datos seleccionado. Cada campo tiene un icono que indica el nivel de medición (es decir, el tipo de datos) del campo.
- **Correlacionado a característica.** Para correlacionar un campo a una característica, pulse dos veces en esta columna (o pulse la barra espaciadora) en la fila de campo y elija una característica de la lista. Si no se muestra ninguna característica adecuada, elija <Añadir nueva característica...> para ver la ventana de configuración de repositorio, donde puede definir una característica nueva para este tipo de entidad. Consulte el tema “Configuración de un repositorio de entidad” para obtener más información.
- **Uso.** Indica el contexto de un campo particular, en el que puede haber más de un contexto (por ejemplo, el número de teléfono de casa y el del trabajo). Hay tipos de uso predefinidos disponibles para las características ADDRESS y PHONE, y puede crear sus propios tipos de uso para todas las características. Para establecer un uso distinto al predeterminado (**Auto**), pulse en la fila que desea de esta columna y elija uno de los tipos de uso (si los hubiera) o pulse en <Añadir uso...> para crear uno nuevo. Consulte el tema “Mantenimiento de los tipos de entidad” en la página 23 para obtener más información.

**Importar correlación.** Importa un conjunto previamente exportado de correlaciones de campo a característica desde un archivo XML externo. Puede resultar útil si tiene diferentes orígenes de datos con los mismos requisitos de correlación, ya que evita tener que redefinir las mismas correlaciones para los distintos orígenes.

**Exportar correlación.** Exporta a un archivo XML externo el conjunto de correlaciones de campo a característica que se muestran en la tabla de correlaciones.

## Visualización de las correlaciones de campos (nodo de exportación de EA)

En la pestaña Repositorio, pulse en el botón **Actualizar** para ver las características del repositorio que tienen campos de entrada correlacionados a ellas. Puede ver las del origen de datos actual (el que está controlado por el nodo de origen asociado a este nodo de exportación) o de todos los orígenes de datos.

**Mostrar entradas para.** Elija una opción para mostrar las correlaciones para el origen de datos actual o para todos los orígenes de datos conocidos por el repositorio.

**Actualizar.** Actualiza la visualización de la opción de entrada seleccionada.

**Características.** Una lista de todas las características que tienen correlaciones en los orígenes de datos mostrados. Las características no correlacionadas no se muestran.

<Origen de datos>. Cada columna enumera los campos correlacionados en un origen de datos concreto para cada característica para la que se ha definido una correlación.

---

## Configuración de un repositorio de entidad

Puede mantener los contenidos del repositorio desde la ventana Configuración de repositorio, que ofrece una interfaz visual fácil de utilizar para el repositorio al completo.

si va a utilizar más de un repositorio con configuraciones idénticas o similares, puede crear una configuración básica y exportarla a un archivo que podrá importar desde otros repositorios. Consulte el tema “Reutilización de configuración de repositorio” en la página 25 para obtener más información.

**Nota:** A partir de la versión 16, SPSS Entity Analytics admite repositorios con el producto IBM DB2. Dado que un repositorio es específico de una versión de SPSS Modeler y no puede importarse de una

versión anterior, si tiene un repositorio existente y va a actualizar a SPSS Entity Analytics versión 16, tendrá que volver a crear ese repositorio en la nueva base de datos de DB2.

**PRECAUCIÓN:**

**Si modifica y guarda la configuración de un repositorio que ya contiene datos, es posible que se le pida que purgue el contenido del repositorio y vuelva a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.**

Para crear una configuración de repositorio

1. Abra cualquier nodo Entity Analytics.
2. Pulse en la lista del **repositorio de entidad**.
3. Pulse en **<Examinar...>** para que se muestre el cuadro de diálogo Casos de resolución de entidades.
4. En el cuadro de diálogo Casos de resolución de entidades, pulse en la lista **Nombre del repositorio**.
5. Seleccione el repositorio para el que quiere crear la configuración.
6. Si no está conectado, introduzca el nombre de usuario de administrador y la contraseña y pulse en **Conectar**.
7. Cuando el botón **Configurar repositorio** esté activado, púlselo para ver la ventana Configuración de repositorio.
8. Cree los detalles de configuración como se explica en las secciones siguientes.

El panel de navegación en la parte izquierda de la ventana Configuración de repositorio contiene una estructura de árbol desde la que puede gestionar las distintas características del repositorio.

*Tabla 9. Elementos principales de la ventana Configuración de repositorio.*

Sección	Descripción	
Orígenes de datos	Muestra las correlaciones de todos los orígenes de datos de las diversas características del repositorio.	Consulte el tema “Visualización de las correlaciones de orígenes de datos” para obtener más información.
Características	Crea una nueva característica o duplica, edita o elimina una existente.	Consulte el tema “Mantenimiento de las características del repositorio” en la página 20 para obtener más información.
Tipos de entidad	Crea un nuevo tipo de entidad, o gestiona los existentes (duplicar, cambiar nombre, adjuntar o eliminar características, borrar).	Consulte el tema “Mantenimiento de los tipos de entidad” en la página 23 para obtener más información.
Reglas de resolución	Establecer umbral para coincidencia de entidades.	Consulte el tema “Establecimiento del umbral para coincidencia de entidades.” en la página 25 para obtener más información.

## Visualización de las correlaciones de orígenes de datos

En la sección Orígenes de datos de la ventana Configuración de repositorio, la entrada Todos los orígenes muestra las correlaciones en formato de solo lectura de todos los orígenes de datos de las diversas características del repositorio.

Pulse en **Actualizar** para actualizar la lista si se han añadido nuevos orígenes de datos al repositorio.

*Note:* No puede añadir un origen de datos al repositorio aquí. Los orígenes de datos solo se pueden añadir creando un nodo de origen de SPSS Modeler y conectándolo al nodo de exportación de Entity Analytics. Consulte el tema “Conexión con un origen de datos” en la página 14 para obtener más información.

## Mantenimiento de las características del repositorio

Una característica de repositorio es un tipo de información individual que se puede utilizar con un origen de datos de entidad. Algunas características (por ejemplo, Nombre, Apellidos, Fecha de nacimiento, etc.) se pueden utilizar con muchos orígenes de datos diferentes, mientras que otras características son específicas de un origen de datos concreto. Una característica puede contener uno o más elementos; en términos generales, cada elemento equivale a un campo de un registro de datos o a una columna de una tabla de base de datos.

En la sección Características de la ventana Configuración de repositorio, la entrada Todas las características proporciona los medios para mantener todas las características del repositorio. Puede hacer lo siguiente.

- Crear una nueva característica
- Duplicar una característica existente (por ejemplo, para crear una característica nueva basada en una existente)
- Modificar una característica existente
- Eliminar una característica existente

Puede encontrar las instrucciones de estas tareas más adelante en esta sección.

La lista de características muestra todas las características que se han definido en este repositorio. Las columnas de la lista muestran las diversas propiedades que puede tener una característica.

**Característica.** El nombre de la característica. Un símbolo de candado junto al nombre de una característica indica que la característica está bloqueada. Las características que están bloqueadas no se pueden suprimir ni duplicar, el único cambio realizado en ellas que se puede guardar es el cambio del atributo de anonimización.

**Frecuencia.** Indica cuántas entidades pueden tener el mismo valor para esta característica. Los valores válidos son **Una** (por ej. para un número de pasaporte), **Pocas** (por ej. para una dirección) o **Muchas** (por ej. para una fecha de nacimiento).

**Exclusividad.** Indica que una entidad debe tener normalmente solo un tipo de característica. Por ejemplo, una fecha de nacimiento o un número de identidad nacional tendría el valor **Sí** aquí, mientras que una dirección o número de tarjeta de crédito tendría el valor **No** (dado que una entidad puede tener varias direcciones o tarjetas de crédito).

**Estabilidad.** Indica el valor de estabilidad de esta característica (es decir, si es *poco probable* que cambie durante la vida útil de una entidad). Por ejemplo, una característica de fecha de nacimiento tendría el valor **Sí**, ya que nunca cambia, pero una característica de dirección tendría **No**, ya que es bastante probable que cambie y, por lo tanto, es menos estable. *Note:* el valor de género normalmente es estable a lo largo de la vida pero, dado que con frecuencia se especifica de manera incorrecta debido a datos erróneos, la configuración predeterminada le da el valor **No**.

**Anonimizar.** Indica si la característica se ha anonimizado. Las entradas son **Sí** o **No**. Para obtener más información, consulte el tema “Anonimización de características del repositorio” en la página 22.

Para crear una nueva característica

1. Realice una de las siguientes acciones.
  - Pulse en el botón Crear nueva característica (el botón superior en la parte derecha de la pantalla).
  - Pulse con el botón derecho en **Todas las características** en el panel de navegación a la izquierda de la pantalla, y elija **Nueva característica**.
2. Complete el cuadro de diálogo Añadir o modificar característica. Consulte el tema “Adición o modificación de una característica” en la página 21 para obtener más información.

Para duplicar una característica existente

1. En la columna **Característica** de la tabla en la parte derecha de la pantalla, seleccione la característica que quiere duplicar.
2. Pulse en el botón Duplicar característica seleccionada (el segundo botón en la parte derecha de la pantalla).
3. Complete el cuadro de diálogo Añadir o modificar característica. Consulte el tema “Adición o modificación de una característica” para obtener más información.

Para modificar una característica existente

**PRECAUCIÓN** Si edita, suprime o anonimiza una característica o un elemento de característica cuando el repositorio ya contiene datos, posteriormente debe purgar el repositorio y volver a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.

1. En la columna **Característica** de la tabla en la parte derecha de la pantalla, seleccione la característica que quiere modificar. *Note:* puede modificar solamente las características que haya creado, no las proporcionadas por el sistema.
2. Pulse en el botón Modificar característica seleccionada (el tercer botón en la parte derecha de la pantalla).
3. Complete el cuadro de diálogo Añadir o modificar característica. Consulte el tema “Adición o modificación de una característica” para obtener más información.

Para eliminar una característica existente

**PRECAUCIÓN** Si edita, suprime o anonimiza una característica o un elemento de característica cuando el repositorio ya contiene datos, posteriormente debe purgar el repositorio y volver a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.

1. En la columna **Característica** de la tabla en la parte derecha de la pantalla, seleccione la característica que quiere eliminar. *Note:* puede eliminar solamente las características que haya creado, no las proporcionadas por el sistema.
2. Realice una de las siguientes acciones.
  - Pulse en el botón Eliminar característica seleccionada (el botón inferior en la parte derecha de la pantalla).
  - Pulse con el botón derecho en **Todas las características** en el panel de navegación a la izquierda de la pantalla, y elija **Eliminar**.
3. Pulse en **Continuar** para confirmar la eliminación de la característica.

**PRECAUCIÓN:**

**esta eliminación no se puede deshacer.**

## Adición o modificación de una característica

**PRECAUCIÓN** Si edita, suprime o anonimiza una característica o un elemento de característica cuando el repositorio ya contiene datos, posteriormente debe purgar el repositorio y volver a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.

En el cuadro de diálogo Añadir o modificar característica, puede crear una nueva característica de repositorio, o duplicar o modificar una característica existente.

*Nota:* si una característica existente está bloqueada, no puede editar sus detalles en este cuadro de diálogo.

**Tipo de característica.** Una etiqueta que indica el tipo de información con el que está relacionada la característica. La etiqueta constituye la primera parte del identificador de la característica.

**Descripción.** Un texto breve que describe el tipo de característica, con fines informativos únicamente.

**Frecuencia.** Indica cuántas entidades pueden tener el mismo valor para esta característica. Los valores válidos son **Una** (por ej. para un número de pasaporte), **Pocas** (por ej. para una dirección) o **Muchas** (por ej. para una fecha de nacimiento).

**Exclusividad.** Indica que una entidad debe tener normalmente solo un tipo de característica. Por ejemplo, una fecha de nacimiento o un número de identidad nacional tendría el valor **Sí** aquí, mientras que una dirección o número de tarjeta de crédito tendría el valor **No** (dado que una entidad puede tener varias direcciones o tarjetas de crédito).

**Estabilidad.** Indica el valor de estabilidad de esta característica (es decir, si es *poco probable* que cambie durante la vida útil de una entidad). Por ejemplo, una característica de fecha de nacimiento tendría el valor **Sí**, ya que nunca cambia, pero una característica de dirección tendría **No**, ya que es bastante probable que cambie y, por lo tanto, es menos estable. *Note:* el valor de género normalmente es estable a lo largo de la vida pero, dado que con frecuencia se especifica de manera incorrecta debido a datos erróneos, la configuración predeterminada le da el valor **No**.

**Tabla de elementos.** Una lista de los elementos que comprende esta característica.

- **Elemento.** Nombre del elemento.
- **Descripción.** Una descripción breve de lo que proporciona el elemento.
- **Tipo de datos.** El tipo de los datos que se pueden utilizar para este elemento. Los tipos disponibles son: Cadena, Entero, Real y Fecha.

**Botón Añadir nuevo elemento.** Añade una fila nueva a la tabla de elementos para poder definir un elemento nuevo.

**Botón Eliminar elemento.** Elimina una fila seleccionada de la tabla de elementos. Esta acción no se puede deshacer.

**PRECAUCIÓN** Si edita, suprime o anonimiza una característica o un elemento de característica cuando el repositorio ya contiene datos, posteriormente debe purgar el repositorio y volver a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.

**Anonimizar.** Para la protección de los datos, puede elegir anonimizar los datos cuando se añaden a un repositorio; para permitir esta opción para una característica, seleccione **Sí**. Consulte el tema “Anonimización de características del repositorio” para obtener más información.

## Anonimización de características del repositorio

Como parte de la seguridad de los datos es posible que desee anonimizar los datos cuando se añaden al repositorio para reducir el riesgo de que la información de identificación personal se pueda divulgar inadvertidamente.

Cuando se exportan datos anonimizados a un repositorio es necesario un método de anonimización que siga permitiendo la resolución de entidades con los datos anonimizados. Por ejemplo, si dos registros de datos para los detalles de la tarjeta de crédito de una persona se anonimizan como "anon\_s21" y "anon\_s9271", pierden su relación; sin embargo, si se utiliza un enlace interno en segundo plano entre los registros, el sistema todavía puede comprender que un nombre es una abreviación del otro.

Los enlaces en segundo plano y los identificadores que permiten enlazar los datos anonimizados se generan cuando se crea un repositorio y son exclusivos del repositorio. Los datos cifrados se almacenan internamente y se leen a continuación cuando una ruta se conecta a un repositorio.

Cuando configure el repositorio, puede especificar si cada característica individual se va a anonimizar o no. Si una característica se anonimiza, se anonimizan todos sus elementos y siempre está anonimizado independientemente de su tipo de uso. Consulte el tema “Adición o modificación de una característica” en la página 21 para obtener más información.

**Nota:** Asegúrese de no anonimizar todos los campos para SPSS Entity Analytics o no podrá identificar los datos que se devuelven. Se recomienda dejar al menos un campo (aunque sólo sea un número de fila) sin anonimizar para poder controlar posteriormente la nueva fusión con los datos originales.

Una columna en la lista de características de la ventana Configuración de repositorio muestra qué características se han definido para anonimizar. Las entradas son **Sí** o **No**.

**Nota:** Si un repositorio existente contiene datos antes de anonimizar características, debe purgar todos los datos primero; de lo contrario, no se producirá ninguna coincidencia entre las características anonimizadas y no anonimizadas.

## Mantenimiento de los tipos de entidad

Un **tipo de entidad** es un conjunto nombrado de características de repositorio que pertenecen al mismo grupo de manera lógica. Por ejemplo, un tipo de entidad pensado para su uso con un conjunto de datos de clientes puede constar de características como Nombre, Fecha de nacimiento, Sexo, Dirección, Número de teléfono, etc.

El repositorio de IBM SPSS Modeler Entity Analytics se proporciona con un conjunto estándar de tipos de entidad y puede añadir sus propios tipos de entidad.

La sección Tipos de entidad de la ventana Configuración de repositorio enumera los distintos tipos de entidad que se han creado. Puede hacer lo siguiente.

- Crear un nuevo tipo de entidad
- Duplicar un tipo de entidad existente (por ejemplo, para crear un tipo de entidad nuevo basado en uno existente)
- Adjuntar características a un tipo de entidad
- Eliminar características de un tipo de entidad
- Cambiar nombre a un tipo de entidad
- Eliminar un tipo de entidad

**Tipo de entidad.** El nombre del tipo de entidad seleccionado.

**Característica.** La lista de características válidas que comprende este tipo de entidad.

**Tipo de uso.** (Opcional) Indica distintos contextos en los que se podría usar esta característica. Pulse dos veces en esta columna para añadir o modificar un tipo de uso, separando los tipos de uso con una coma y un espacio. Los valores que especifique aquí definen los valores mostrados en el nodo de exportación de EA o el nodo Generación de análisis de entidad cuando un usuario pulsa en la columna Uso para una característica en la pestaña Entradas. Consulte el tema “Opciones de entrada del repositorio para la correlación” en la página 17 para obtener más información.

Información general acerca de los tipo de uso:

- Los tipos de uso puede tener cualquier etiqueta.
- Puede crear un tipo de uso a partir de prácticamente cualquier tipo de entrada de texto; sin embargo, no debe introducir espacios ni caracteres no válidos.
- El tipo se cambia automáticamente a mayúsculas una vez introducido.
- Puede tener tantos tipos de uso como desee.

- Los tipos de uso no necesitan tener un nombre significativo, pero le ayudará posteriormente al establecer correlaciones si utiliza una convención de nomenclatura que tenga sentido para usted y para otros usuarios.
- Al establecer correlaciones, se le avisará con un mensaje en rojo si ha correlacionado algunos elementos con un tipo de uso y otros con uno distinto.

Normalmente se muestra un error si se intenta correlacionar dos campos con el mismo elemento de característica. Los tipos de uso con una forma de correlacionar dos o más campos con el mismo elemento de característica y que puedan coincidir entre ellos.

Por ejemplo, si definió dos características independientes: *DIRECCIÓNCASA* y *DIRECCIÓNTRABAJO*, no habrá coincidencias entre ellas. Si una entidad tiene una *DIRECCIÓNCASA* que es la misma que la *DIRECCIÓNTRABAJO* de otra entidad, no existe coincidencia porque hay características distintas. Sin embargo, si reutiliza una característica única con distintos tipos de uso, la resolución entiende que *DIRECCIÓNTRABAJO* es lo misma que *DIRECCIÓNCASA*.

Puede reutilizar tipos de uso para distintas características o tener tipos distintos; por ejemplo, *CS* y *TB* para el teléfono y *CASA* y *TRABAJO* para la dirección. No tiene relevancia porque no buscamos coincidencias de teléfonos con direcciones; sin embargo, si mantiene la coherencia podrá identificar y agrupar campos posteriormente.

Si se alimenta un solo repositorio con múltiples tipos de entrada, siempre que use la misma característica, no importa cuáles sean los tipos de uso. Por ejemplo, si define *TB* y *HM* como tipos de uso de *DIRECCIÓN* para el tipo de entidad *EMPRESA*, podrá coincidir con *TRABAJO* y *CASA* como tipos de uso de *DIRECCIÓN* para *PERSONA*.

Para crear un nuevo tipo de entidad

1. Pulse con el botón derecho en **Tipos de entidad** en el panel de navegación a la izquierda de la pantalla.
2. Elija **Nuevo tipo de entidad**.
3. Introduzca un nombre no utilizado anteriormente para el tipo de entidad y pulse en Aceptar.
4. Adjunte características al tipo de entidad (ver sección siguiente).

Para adjuntar características a un tipo de entidad

1. Seleccione el tipo de entidad en el panel de navegación a la izquierda de la pantalla.
2. Pulse en el botón Adjuntar característica (el botón superior en la parte derecha de la pantalla).
3. De la lista de características disponibles, elija una o varias (deje pulsada la tecla Ctrl mientras pulsa en las distintas características) y pulse en Aceptar.

Para eliminar características de un tipo de entidad

1. Seleccione el tipo de entidad en el panel de navegación a la izquierda de la pantalla.
2. Seleccione una o más características de la tabla de características adjuntadas en la parte derecha de la pantalla. Mantenga pulsada la tecla Ctrl para elegir más de una característica.
3. Pulse en el botón Desvincular característica (el botón inferior en la parte derecha de la pantalla).

Para duplicar un tipo de entidad existente

1. En el panel de navegación de la parte izquierda de la pantalla, pulse con el botón derecho en el tipo de entidad que desea duplicar.
2. Elija **Duplicar tipo de entidad**.
3. Introduzca un nombre no utilizado anteriormente para el nuevo tipo de entidad y pulse en Aceptar.
4. Adjunte o desvincule características del tipo de entidad según sea necesario (ver instrucciones anteriores).

Para cambiar el nombre a un tipo de entidad

**PRECAUCIÓN** Si edita, suprime o anonimiza una característica o un elemento de característica cuando el repositorio ya contiene datos, posteriormente debe purgar el repositorio y volver a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.

1. En el panel de navegación de la parte izquierda de la pantalla, pulse con el botón derecho en el tipo de entidad al que desea cambiar el nombre.
2. Elija **Cambiar nombre**.
3. Introduzca el nombre nuevo para el tipo de entidad y pulse en **Aceptar**.

Para eliminar un tipo de entidad

**PRECAUCIÓN** Si edita, suprime o anonimiza una característica o un elemento de característica cuando el repositorio ya contiene datos, posteriormente debe purgar el repositorio y volver a cargar los datos. Al hacerlo, evitará que el repositorio quede en un estado incoherente.

1. En el panel de navegación de la parte izquierda de la pantalla, pulse con el botón derecho en el tipo de entidad que desea eliminar.
2. Seleccione **Eliminar**.
3. Pulse en **Aceptar** para confirmar la eliminación del tipo de entidad.

**PRECAUCIÓN:**  
esta eliminación no se puede deshacer.

## Establecimiento del umbral para coincidencia de entidades.

En la sección Reglas de resolución de la ventana Configuración de repositorio, puede elegir el umbral al que se producirá la coincidencia de entidades.

Al crear el repositorio, la coincidencia está definida en el umbral predeterminado.

Elija **Conjunto para resolución agresiva** si no encuentra suficientes coincidencias en sus registros para llevar a cabo una resolución de entidad.

Elija **Conjunto para resolución predeterminada** para regresar al umbral predeterminado desde una de las otras configuraciones.

Elija **Conjunto para resolución cautelosa** si recupera demasiadas coincidencias.

Para crear un repositorio para ambas entidades y relaciones, seleccione **Incluir relaciones**. Tenga en cuenta que esta opción sólo está disponible si tiene la actualización con una licencia aparte conocida como IBM SPSS Modeler Entity Analytics Unleashed.

## Reutilización de configuración de repositorio

Si ya ha creado una configuración y desea utilizarla en otro repositorio, puede exportar la configuración existente a un archivo XML e importarlo desde el otro repositorio (destino). Esto sólo es posible dentro de una instalación existente. Por ejemplo, no puede migrar una configuración de repositorio de una versión de IBM SPSS Modeler a otra, o de un tipo de base de datos a otro.

Para reutilizar una configuración existente

1. Abra la ventana Configuración de repositorio del repositorio cuya configuración desea utilizar. Consulte el tema “Configuración de un repositorio de entidad” en la página 18 para obtener más información.
2. En el menú de esa ventana, elija **Configuración > Exportar configuración**.

3. En el cuadro de diálogo Guardar como, elija el nombre y la ubicación del archivo XML de exportación.
4. Abra la ventana Configuración de repositorio del repositorio de destino.
5. En el menú de esa ventana, elija **Configuración > Importar configuración**.
6. En el cuadro de diálogo Abrir, elija el nombre y la ubicación del archivo XML exportado anteriormente y pulse en **Abrir**.

## Guardar sus cambios de configuración

Para guardar los cambios de configuración

En el menú de la ventana Configuración de repositorio, elija

**Archivo > Guardar**.

## Cerrar la ventana de configuración

Para salir de la ventana de configuración

En el menú de la ventana Configuración de repositorio, elija

**Archivo > Salir**.

Si hay cambios sin guardar en la configuración, pulse en **Aceptar** para guardar los cambios y salir, o en **Cancelar** para salir sin guardar.

---

## Análisis de las identidades resueltas [nodo de origen Entity Analytics(EA)]

Después de exportar los datos al repositorio, puede utilizar el nodo de origen Entity Analytics(EA) para pasar las identidades resueltas a otros nodos de IBM SPSS Modeler para un posterior análisis o procesamiento como, por ejemplo, crear un informe que liste las identidades resueltas.

Para analizar las identidades resueltas

1. Añada un nodo de origen Entity Analytics(EA) a una ruta.
2. Abra el nodo de origen Entity Analytics(EA).
3. En la pestaña Datos, seleccione el repositorio de entidad y una o más de sus orígenes de datos de entrada (pulse en **Actualizar** para actualizar los recuentos de registros). Consulte el tema “Selección de un origen de datos” para obtener más información.
4. Añada más nodos a la ruta para realizar los procesos que desee. Consulte el tema “Adición de nodos a la ruta” en la página 28 para obtener más información.

## Selección de un origen de datos

En la pestaña Datos, seleccione al menos un origen de datos en el repositorio en el que se realizarán los próximos procesos. Para actualizar los recuentos de registros para los orígenes de datos listados, pulse en **Actualizar**.

**Repositorio de entidad.** Muestra el repositorio de entidad actual si existe alguno. Si hay más de un repositorio, para elegir uno distinto, selecciónelo de la lista. Para crear un nuevo repositorio, elija **<Examinar...>** para ver un cuadro de diálogo en el que puede crear el repositorio. Consulte el tema “Opciones de repositorio de entidad” en la página 16 para obtener más información.

**Incluir registros de orígenes de datos.** Esta tabla enumera los distintos orígenes de datos que se han introducido en el repositorio, junto con el número de registros en cada origen. Seleccione la casilla de verificación **Incluir** de aquellos orígenes de datos que desee utilizar para realizar análisis y procesos posteriores. Para seleccionar o deseleccionar todos los orígenes de datos, pulse en **Incluir todos** o **Excluir todos** respectivamente.

**Relaciones.** Seleccione el tipo de relación que se va a incluir en el repositorio. Tenga en cuenta que esta opción sólo está disponible si tiene la actualización con una licencia aparte conocida como IBM SPSS Modeler Entity Analytics Unleashed y si el repositorio está configurado para incluir relaciones.

- **Ninguna relación.** No se utilizan detalles de relaciones.
- **Relaciones próximas.** Selecciona únicamente las entidades con relaciones próximas. La proximidad de una relación depende de muchas variables posibles como, por ejemplo, las propiedades de las características correlacionadas, de qué características se comparten y de si la resolución se ha establecido como conservadora o agresiva.
- **Todas las relaciones.** Selecciona todas las entidades relacionadas.

**Máx. grado de separación.** Esta opción sólo está disponible si se ha seleccionado **Relaciones próximas** o **Todas las relaciones**. Seleccione el número de grados de separación que utilizarán para identificar una relación. Por ejemplo, si Ann y Bob no se conocen, pero John conoce a Ann y Bob, Ann y Bob están relacionados con dos grados de separación.

**Tipo de entidad de salida.** De forma predeterminada, si el repositorio contiene detalles, éste muestra el primer tipo de entidad listado en el repositorio. Si el repositorio tiene más de un tipo de entidad, la selección de un tipo de entidad aquí cambia las características mostradas en la pestaña Filtro para listar las características para dicho tipo. Puede seleccionar cualquiera de los tipos de entidad utilizados en el repositorio.

## Cambio de nombre de campos de datos

Puede usar la pestaña Filtro para cambiar el nombre de los campos de identidades resueltas que pasan hacia abajo para procesos posteriores. Puede que quiera cambiar el nombre de un campo de identidad resuelta, por ejemplo, para mantener la compatibilidad del nombre de campo en la fusión hacia abajo con otro conjunto de datos.

Los campos con sus nombres originales son los siguientes.

*Tabla 10. Campos de identidades resueltas*

Campo	Descripción
\$EA-ID	Identificador de entidad
\$EA-SRC	Etiqueta de origen que identifica el origen de datos donde se crearon los registros
\$EA-KEY	Campo designado como clave exclusiva en un archivo de origen de datos

*Note:* aunque también puede usar la pestaña Filtro para excluir campos, no debería hacerlo ya que los campos de identidades resueltas son el mínimo necesario para el procesamiento de análisis de entidades.

## Establecimiento de la información de tipos para campos de datos

En la pestaña Tipos, puede ver o cambiar diversas propiedades de los campos de identidades resueltas que pasan hacia abajo para procesos posteriores.

Las propiedades que puede cambiar son las mismas que las de la pestaña Tipos de un nodo Tipo de SPSS Modeler normal, y son las siguientes.

Tabla 11. Propiedades de Tipo para campos

Propiedad	Descripción
Medidas	El nivel de medición (es decir, el tipo de datos), usado para describir características de los datos en el campo.
Valores	Proporciona opciones para la lectura de valores de datos desde el conjunto de datos.
No se encuentra	Se utiliza para especificar cómo se tratan los valores perdidos de un campo.
Comprobar	Opciones de validación para asegurar que los valores de campo se ajustan a los valores o rangos especificados.
Función	Especifica el modo en que se usará el campo si se transfieren los datos a un nodo de modelado o un nugget de modelo.

## Adición de nodos a la ruta

Puede añadir diversos nodos de SPSS Modeler a la ruta para realizar operaciones de análisis o procesos de los resultados del nodo de origen Entity Analytics(EA). Por ejemplo, puede añadir uno o más de los siguientes.

- Nodo Agregar o Distinguir para resumir la salida, que puede ser muy voluminosa
- Nodo Seleccionar para seleccionar un subconjunto de los resultados
- Nodo Tabla para ver el resultado desde el nodo de origen Entity Analytics(EA)
- Nodo Informe para imprimir los resultados en un informe
- Nodo de exportación de SPSS Modeler para exportar el resultado en un formato diferente, como una hoja de cálculo o base de datos

Si desea más información, consulte las secciones sobre los nodos Operaciones con registros, Resultado y Exportar en la guía Nodos de origen, proceso y resultado de *IBM SPSS Modeler*.

---

## Comparación de nuevos casos con el repositorio (nodo Generación de análisis de entidad)

Una vez realizada alguna resolución de identidad en el repositorio, puede utilizar el nodo Generación de análisis de entidad para comparar casos nuevos que encuentre a partir de entonces con el contenido del repositorio. Este nodo procesa registros desde un nuevo origen de datos, los compara con las entidades resueltas que ya se encuentran en el repositorio y pasa cualquier registro coincidente para su posterior procesamiento. Se puede establecer que las coincidencias sean exactas o relacionadas más libremente con entidades existentes.

Al igual que el nodo de exportación de EA, el nodo Generación de análisis de entidad toma un único nodo de origen de SPSS Modeler como entrada. Sin embargo, el nodo Generación de análisis de entidad difiere en lo siguiente. Mientras que el nodo de exportación genera registros para todas las entidades relacionadas con sus registros de entrada, el nodo Generación de análisis de entidad únicamente genera registros para aquellas entidades relacionadas con las entidades ya resueltas en el repositorio. Consulte el tema “Resultados del nodo Generación de análisis de entidad” en la página 31 para obtener más información.

Para comparar casos nuevos con el repositorio

1. Conéctese al origen de datos que contiene los nuevos registros que desee comparar con las entidades existentes. Consulte el tema “Conexión con un origen de datos” en la página 14 para obtener más información.
2. En la pestaña Operaciones con registros, adjunte un nodo Generación de análisis de entidad al nodo de origen de datos.
3. Pulse dos veces en el nodo de exportación de Entity Analytics para abrir su cuadro de diálogo.

4. Pulse en la lista del **repositorio de entidad**.
5. Pulse en <Examinar...> para que se muestre el cuadro de diálogo Repositorios de entidad.
6. En el cuadro de diálogo Repositorios de entidad, pulse en el campo Nombre del repositorio.
7. Pulse en el nombre del repositorio que desea usar.
8. Especifique el nombre de usuario y contraseña de este repositorio y pulse en **Conectar**. Pulse en **Aceptar** cuando el repositorio esté conectado.
9. En el cuadro de diálogo Generación de análisis de entidad, seleccione el tipo de entidad al que desea correlacionar. Consulte el tema “Mantenimiento de los tipos de entidad” en la página 23 para obtener más información.
10. Correlacione los campos de entrada en el origen de datos a características del repositorio. Consulte el tema “Correlación de campos de entrada a características (nodo Generación de análisis de entidad)” para obtener más información.
11. De forma opcional, puede actualizar los registros del repositorio en tiempo real a medida que puntúa los datos. Consulte el tema “Correlación de campos de entrada a características (nodo Generación de análisis de entidad)” para obtener más información.
12. Pulse en la pestaña **Resultados** para ver los detalles de los distintos orígenes de datos entrados en el repositorio y establezca el criterio de selección para recuperar entidades existentes. Consulte el tema “Visualización de las correlaciones de campos y orígenes de datos (nodo Generación de análisis de entidad)” en la página 30 para obtener más información.
13. Pulse en la pestaña **Filtro** para ver los detalles de los campos de entrada y las características que se almacenan en el repositorio. Las características que no se han correlacionado en el nodo se filtran de forma predeterminada; sin embargo, puede cambiarlo si es necesario.
14. Pulse en **Aceptar** cuando haya configurado el nodo correctamente.
15. Conecte un nodo Tabla al nodo Generación de análisis de entidad y ejecute la ruta.

La ventana de resultados del nodo Tabla enumera todas las entidades recuperadas que coinciden con los nuevos registros del origen de datos. Los campos de resultados tienen el prefijo **\$EA-** añadido. Consulte el tema “Resultados del nodo Generación de análisis de entidad” en la página 31 para obtener más información.

**Nota:** es posible que detecte un error de tipo **Se ha detectado un número incorrecto de campos en el modelo de datos del servidor**. al ejecutar el nodo Generación de análisis de entidad. Esto puede ocurrir si ha editado la configuración de repositorio desde que creó el nodo Generación de análisis de entidad. Editar la configuración en estas circunstancias puede tener el efecto de cambiar el número y los nombres de los resultados de los campos del nodo. Para resolver el problema, abra el nodo Generación de análisis de entidad y pulse el botón **Actualizar**. De este modo vuelven a calcularse el número y los nombres de los campos de resultados.

## Correlación de campos de entrada a características (nodo Generación de análisis de entidad)

La pestaña Entradas contiene las opciones para correlacionar campos de la entrada de este nodo a las características del repositorio. Configure las asignaciones de correlación en esta pestaña, o seleccione la pestaña **Ver** para ver los detalles de todos los orígenes de datos del repositorio y pulse **Aceptar**.

Si ya ha guardado un conjunto de correlaciones en un archivo XML, puede usarlas pulsando en **Importar correlación**.

**Repositorio de entidad.** Muestra el repositorio de entidad actual si existe alguno. Si hay más de un repositorio, para elegir uno distinto selecciónelo de la lista. Para crear un nuevo repositorio, elija <Examinar...> para ver un cuadro de diálogo en el que puede crear el repositorio. Consulte el tema “Opciones de repositorio de entidad” en la página 16 para obtener más información.

**Correlacionar a tipo de entidad.** Una lista de tipos de entidades (es decir, conjuntos de características) definidos en el repositorio. Elija uno de la lista o <Añadir nuevo tipo de entidad...> para ver la ventana de configuración de repositorio, donde puede definir un nuevo tipo de entidad. Consulte el tema “Configuración de un repositorio de entidad” en la página 18 para obtener más información.

**Persistir búsquedas.** Si desea actualizar los registros del repositorio en tiempo real a medida que puntúa los datos, seleccione esta opción.

**Etiqueta de origen.** Sólo está disponible cuando se selecciona **Persistir búsquedas**. Una lista de etiquetas que indica los orígenes de datos conocidos actualmente en el repositorio. Elija una de la lista o <Añadir nueva etiqueta de origen...> para crear una etiqueta de un nuevo origen de datos.

**Clave exclusiva.** Sólo está disponible cuando se selecciona **Persistir búsquedas**. El campo de entrada que se utiliza para los identificadores exclusivos de los registros de datos.

**Tabla de correlaciones.** En esta tabla puede correlacionar cada campo de entrada a una característica correspondiente en el repositorio. Si no existe ninguna característica adecuada en el tipo de entidad seleccionado, puede crear una característica nueva aquí.

- **Campo.** El conjunto de campos de entrada en el origen de datos seleccionado. Cada campo tiene un icono que indica el nivel de medición (es decir, el tipo de datos) del campo.
- **Correlacionado a característica.** Para correlacionar un campo a una característica, pulse dos veces en esta columna (o pulse la barra espaciadora) en la fila de campo y elija una característica de la lista. Si no se muestra ninguna característica adecuada, elija <Añadir nueva característica...> para ver la ventana de configuración de repositorio, donde puede definir una característica nueva para este tipo de entidad. Consulte el tema “Configuración de un repositorio de entidad” en la página 18 para obtener más información.
- **Uso.** Indica el contexto de un campo particular, en el que puede haber más de un contexto (por ejemplo, el número de teléfono de casa y el del trabajo). Consulte el tema “Mantenimiento de los tipos de entidad” en la página 23 para obtener más información.

**Importar correlación.** Importa un conjunto previamente exportado de correlaciones de campo a característica desde un archivo XML externo. Puede resultar útil si tiene diferentes orígenes de datos con los mismos requisitos de correlación, ya que evita tener que redefinir las mismas correlaciones para los distintos orígenes.

**Exportar correlación.** Exporta a un archivo XML externo el conjunto de correlaciones de campo a característica que se muestran en la tabla de correlaciones.

## Visualización de las correlaciones de campos y orígenes de datos (nodo Generación de análisis de entidad)

En la pestaña Resultados puede ver los detalles de los distintos orígenes de datos que se han entrado en el repositorio. Estas son los orígenes de datos con los que se procesan las entradas este nodo, para buscar y recuperar entidades coincidentes. Pulse en **Actualizar** para actualizar los recuentos de registros.

**Incluir coincidencias de orígenes de datos.** Esta tabla lista los distintos orígenes de datos disponibles en el repositorio, junto con el número de registros en cada origen.

**Coincidencias.** Estas opciones especifican la precisión con que la información de correlaciones de campo a característica que ha especificado en la pestaña Entradas se hará coincidir con los registros de candidato (es decir, todos los contenidos del repositorio). Cuanto más preciso sea el criterio de búsqueda, menos entidades serán devueltas.

*Nota:* si se encuentran más de 20 coincidencias, solo se devolverán las 20 primeras.

- **Incluir solo resultados exactos.** Éste es el criterio de coincidencia más preciso y con el que se seleccionan resultados en menos registros. Utilice esta opción cuando desee que se devuelvan únicamente estas entidades consideradas como coincidencias exactas.
- **Incluir coincidencias posibles.** Utilice esta configuración cuando desee que se devuelvan las entidades coincidentes y las entidades que compartan los mismos identificadores (aquellos cuyas características se hayan configurado con un valor de frecuencia de Uno; por ejemplo, números de identificación fiscal o números de tarjeta de crédito coincidentes, etc.).
- **Incluir todas las coincidencias.** Utilice esta opción si desea ver el número de entidades más amplio posible en el repositorio con características compartidas. Este es el criterio de coincidencia menos preciso y con el que se seleccionan resultados en más registros. Esta opción devuelve coincidencias exactas y entidades que comparten casi cualquier característica (normalmente aquellas con un valor de frecuencia de Uno o Pocos). Por ejemplo, se incluirán las entidades con el mismo número de identificación fiscal y las entidades con direcciones similares.

**Relaciones.** Sólo está disponible si el repositorio está configurado para incluir relaciones. Para configurar el repositorio para que incluya relaciones, debe tener la actualización con una licencia aparte conocida como IBM SPSS Modeler Entity Analytics Unleashed. Seleccione el tipo de relación que se incluirá en la salida.

- **Ninguna relación.** No se utilizan detalles de relaciones.
- **Relaciones próximas.** Selecciona únicamente las entidades con relaciones próximas. La proximidad de una relación depende de variables posibles como, por ejemplo, las propiedades de las características correlacionadas, de qué características se comparten y de si la resolución se ha establecido como conservadora o agresiva.
- **Todas las relaciones.** Selecciona todas las entidades relacionadas.

**Máx. grado de separación.** Esta opción sólo está disponible si se ha seleccionado **Relaciones próximas** o **Todas las relaciones**. Seleccione el número de grados de separación que utilizarán para identificar una relación. Por ejemplo, si Ann y Bob no se conocen, pero John conoce a Ann y Bob, Ann y Bob están relacionados con dos grados de separación.

**Tipo de entidad de salida.** De forma predeterminada, si el repositorio contiene detalles, éste muestra el primer tipo de entidad listado en el repositorio. Si el repositorio tiene más de un tipo de entidad, la selección de un tipo de entidad aquí cambia las características mostradas en la pestaña Filtro para listar las características para dicho tipo. Puede seleccionar cualquiera de los tipos de entidad utilizados en el repositorio.

## Resultados del nodo Generación de análisis de entidad

Los resultados del nodo Generación de análisis de entidad constan de los siguientes campos para cada registro que se recupera.

Campo	Descripción
<i>Campo1[, Campo2[, ... CampoN]]</i>	Campos del origen de datos que contiene los nuevos registros.
\$EA-ID	Identificador de entidad para este registro en el repositorio.
\$EA-SRC	Etiqueta de origen que identifica el origen de datos donde se creó este registro.
\$EA-KEY	Valor de la clave exclusiva de este registro en el archivo de origen de datos.
\$EA-SC	Campo que indica la cercanía de la coincidencia entre este registro y una entidad observada en el repositorio; un valor de 1,0 (mala coincidencia) a 10,0 (buena coincidencia).

Campo	Descripción
\$EA-Característica1[, \$EA-Característica2[, ... \$EA-CaracterísticaN]]	Valores de las características correlacionadas para este registro en el repositorio.

Si los campos de relaciones están habilitados en el repositorio y el grado de separación es superior a cero en la pestaña Resultados, la salida del nodo Generación de análisis de entidad también contiene los campos siguientes para cada registro que se recupera.

Campo	Descripción
\$EA-DEGREE	Grado de separación.
\$EA-PARENT	Identificador del registro desde el que se calcula la separación.
\$EA-CHILD	Identificador del registro hasta el que se calcula la separación.
\$EA-RULE	

---

## Uso de IBM SPSS Modeler Entity Analytics con otros productos IBM SPSS

Los instaladores pueden permitirle utilizar IBM SPSS Modeler Entity Analytics con los siguientes productos:

- IBM SPSS Collaboration and Deployment Services
- IBM SPSS Modeler Batch para Windows
- IBM SPSS Modeler Solution Publisher

Deberá ejecutar estos instaladores antes de poder utilizar las características de IBM SPSS Modeler Entity Analytics con estos productos. Si desea obtener más información, consulte la guía de *instalación de IBM SPSS Modeler Premium*.

Después de la instalación, debe utilizar el cliente de IBM SPSS Collaboration and Deployment Services Deployment Manager para crear una definición de servidor de repositorio de Entity Analytics. Esto es necesario para utilizar una ruta de IBM SPSS Modeler que contiene un nodo Entity Analytics en un trabajo de IBM SPSS Collaboration and Deployment Services (es decir, para ejecutar rutas de Entity Analytics en IBM SPSS Collaboration and Deployment Services). La definición de servidor debe coincidir con el nombre de repositorio en la ruta; esta definición se utiliza para indicar a la ruta dónde se encuentra el repositorio y proporcionarle la información de conexión que necesita.

---

## Tareas administrativas

Para los repositorios creados en Entity Analytics, se crea un nuevo servicio de base de datos utilizando el producto IBM DB2. Hay algunas tareas administrativas asociadas a DB2; estas tareas las realiza normalmente el administrador de la base de datos o del sistema y son las siguientes:

- Configuración de asignaciones de puertos
- Gestión de las credenciales de administrador para la base de datos del repositorio

Otras tareas administrativas que quizás deban aplicarse a todos los repositorios son las siguientes:

- Desplazamiento del repositorio a un directorio de almacenamiento distinto
- Definir las propiedades de ruta para los campos fecha/hora e indicación de fecha y hora
- Ajuste de los valores de tiempo de espera

- Ejecución de IBM SPSS Modeler Entity Analytics con el cliente de SPSS Modeler y SPSS Modeler Server en el mismo sistema Windows
- Purga de un repositorio de entidad
- Eliminación de un repositorio de entidad
- Eliminación de repositorio cuando no se puede conectar a él

## Configuración de asignaciones de puertos

A cada servicio de base de datos DB2 se le debe asignar un puerto que no se puede asignar a otros servicios que se ejecuten en la máquina. Los servicios de bases de datos residen en la misma máquina en la que se ejecuta IBM SPSS Modeler Server (o, si IBM SPSS Modeler se utiliza sin una conexión a IBM SPSS Modeler Server, la máquina en la que se ejecuta IBM SPSS Modeler).

De forma predeterminada, Entity Analytics asigna puertos en el rango de 1320 a 1520, comenzando por el puerto 1320 para el primer repositorio creado. En caso de conflicto, puede configurar la asignación de puertos editando el archivo `<ruta de instalación del servidor de modeler>/ext/bin/pasw.entityanalytics/ea.cfg` y definiendo valores apropiados para `min_port` y `max_port`. Los contenidos predeterminados de este archivo se muestran a continuación:

```
# configuración de rango de puertos para análisis de entidades
#
#     este rango de puertos controla los puertos que las bases de datos DB2
#     (creado para almacenar repositorios de Entity Analytics)
#     pueden utilizar. Configure este valor si el rango de puertos predeterminado
#     va a crear un conflicto en el sistema.
#
# valor predeterminado: min_port = 1320
# valor predeterminado: max_port = 1520
min_port, 1320
max_port, 1520
```

## Gestión de las credenciales de administrador para la base de datos del repositorio

El nombre de usuario y contraseña del administrador para la base de datos DB2 que aloja un repositorio de entidad se definen al crear el repositorio. Si conoce las credenciales actuales, puede cambiar estos detalles mediante el editor de SQL de DB2.

Para iniciar el editor de SQL de DB2

1. En un equipo cliente, abra una ventana de símbolo de sistema.
2. Escriba:
 

```
cd directorio_instalación_modeler\ext\bin\pasw.entityanalytics\DB2\bin
```

 donde `directorio_instalación_modeler` es el directorio de instalación de SPSS Modeler.
3. Escriba:

```
solsql -c "C:\Documents and Settings\All Users\Application Data\IBM\SPSS\Modeler\versión\EA\
repositories\nombre_repositorio
```

donde *versión* es el número de versión de la instalación de SPSS Modeler y *nombre\_repositorio* es el nombre del repositorio.

4. En el indicador que se muestra, especifique el nombre de usuario y la contraseña del administrador de bases de datos actual para mostrar el indicador `solsql>`.

Para cambiar la contraseña de administrador de base de datos

1. En el indicador `solsql>`, especifique:

```
alter user nombre_usuario identified by contraseña;
commit work;
```

donde *nombre\_usuario* es el nombre de usuario del administrador de base de datos actual y *contraseña* es la nueva contraseña.

2. Salga para cerrar el editor.
3. Reinicie el cliente de SPSS Modeler.

Para obtener información para realizar otras tareas administrativas relacionadas con la base de datos DB2, consulte la documentación para la versión apropiada de IBM DB2 en <http://publib.boulder.ibm.com/>.

## Desplazamiento del repositorio a un directorio de almacenamiento distinto

De forma predeterminada, los archivos del repositorio se almacenan en un directorio denominado *EA* en las siguientes ubicaciones:

- C:\Documents and Settings\All Users\ApplicationData\IBM\SPSS\Modeler\versión\EA (sistemas Windows)
- *directorio\_de\_instalación\_de\_modeler*/ext/bin/pasw.entityanalytics/EA (sistemas UNIX)

Como los archivos utilizados para almacenar el repositorio pueden llegar a ser muy grandes, puede que necesite moverlos a un disco o partición diferente para tener más espacio disponible.

Para mover el repositorio a un directorio distinto:

1. Salga de SPSS Modeler.
2. Mueva el directorio *EA* desde su ubicación original (indicada anteriormente) a una nueva ubicación. Por ejemplo, en Windows puede que desee moverlo a una nueva ubicación como *F:\data\EA*.
3. Edite el archivo `<ruta de instalación del servidor de modeler>/ext/bin/pasw.entityanalytics/ea.cfg` para añadir la siguiente opción:

```
repository_data_directory, nueva_ubicación
```

donde *nueva\_ubicación* es el directorio al que ha movido el directorio *EA*, por ejemplo, *F:\data\EA*.

## Definir las propiedades de ruta para los campos fecha/hora e indicación de fecha y hora

Si sus datos de origen incluyen campos que contienen datos de fecha/hora o indicación horaria, asegúrese de establecer las propiedades de ruta correspondientes con un formato reconocido por IBM SPSS Modeler Entity Analytics.

Para definir el formato de propiedad de ruta:

1. En el menú principal de SPSS Modeler, seleccione:  
**Herramientas > Propiedades de ruta > Opciones.**
2. Seleccione **Fecha/hora**.
3. Establezca **Formato de fecha** como **AAAA-MM-DD**.

4. Establezca **Formato de hora** como **HH:MM:SS**.
5. Pulse en **Aceptar**.

## Ajustar los valores de tiempo de espera

En sistemas que son lentos o que están muy cargados, si se producen errores al crear repositorios o al acceder a los mismos, es posible que deba incrementar los valores de tiempo de espera para iniciar y detener el motor de análisis de entidades o el servidor de bases de datos de análisis de entidades.

Para ajustar el tiempo de espera para el motor de análisis de entidades:

1. Salga de SPSS Modeler.
2. Edite el archivo *<ruta de instalación del servidor de modeler>/ext/bin/pasw.entityanalytics/ea.cfg* para incrementar el valor de la siguiente opción:

`timeout, valor`

donde *valor* es el valor de tiempo de espera, expresado en segundos, para el motor de análisis de entidades (el valor predeterminado es 60).

Para ajustar el tiempo de espera para el servidor de bases de datos de análisis de entidades (sólo DB2)

1. Salga de SPSS Modeler.
2. Edite el archivo *<ruta de instalación del servidor de modeler>/ext/bin/pasw.entityanalytics/ea.cfg* para incrementar el valor de la siguiente opción:

`timeout, valor`

donde *valor* es el valor de tiempo de espera, expresado en segundos, para el servidor de bases de datos DB2 de análisis de entidades (el valor predeterminado es 100).

## Ejecución de IBM SPSS Modeler Entity Analytics con el cliente de SPSS Modeler y SPSS Modeler Server en el mismo sistema Windows

Si ha instalado IBM SPSS Modeler Entity Analytics en el cliente de SPSS Modeler y SPSS Modeler Server en el mismo sistema Windows, tanto el cliente como el servidor compartirán el mismo repositorio de forma predeterminada. Si desea que utilicen repositorios separados, deberá editar el archivo de configuración *ea.cfg* en **uno** de los sistemas y configurarlo para que utilice un rango de puertos y una carpeta de repositorio diferentes.

*Note:* en particular, si utiliza un cliente de SPSS Modeler de 32 bits y un SPSS Modeler Server de 64 bits (o viceversa), deberá realizar este procedimiento.

1. Abra el archivo *<ruta de instalación [del servidor] de modeler>/ext/bin/pasw.entityanalytics/ea.cfg* para su edición.
2. Cambie los valores de `min_port` y `max_port` para utilizar puertos distintos del otro sistema. Consulte el tema “Configuración de asignaciones de puertos” en la página 33 para obtener más información.
3. Cambie el valor de `repository_data_directory` para utilizar un directorio distinto del otro sistema.
4. Guarde y cierre el archivo *ea.cfg*.

## Purga de un repositorio de entidad

Si desea borrar los registros de datos de un repositorio de entidad pero desea mantener la información de configuración, puede purgar los datos del repositorio.

Para purgar todos los datos de un repositorio:

1. Abra un nodo Entity Analytics.
2. Pulse en la lista del **repositorio de entidad**.
3. Pulse en **<Examinar...>** para que se muestre el cuadro de diálogo Casos de resolución de entidades.

4. En el cuadro de diálogo Casos de resolución de entidades, pulse en la lista **Nombre del repositorio**.
5. Seleccione el repositorio que desea purgar.
6. Si no está conectado, introduzca el nombre de usuario de administrador y la contraseña y pulse en **Conectar**.
7. Cuando el botón **Purgar todo** esté activado, púlselo.
8. En el cuadro de diálogo Purgar todos los orígenes de datos pulse **Purgar** para confirmar la purga del repositorio.

## Supresión de orígenes de datos no utilizados de un repositorio

Si tiene un origen de datos que ya no se utiliza o necesita en un repositorio de entidad, puede suprimir el origen del repositorio. Puede seleccionar uno o más orígenes de datos para supresión.

Para suprimir un origen de datos seleccionado de un repositorio:

1. Abra un nodo Entity Analytics.
2. Pulse en la lista del **repositorio de entidad**.
3. Pulse en <Examinar...> para que se muestre el cuadro de diálogo Casos de resolución de entidades.
4. En el cuadro de diálogo Casos de resolución de entidades, pulse en la lista **Nombre del repositorio**.
5. Seleccione el repositorio del que desea suprimir un origen de datos.
6. Si no está conectado, introduzca el nombre de usuario de administrador y la contraseña y pulse en **Conectar**.
7. En la lista **Gestionar repositorio**, seleccione el origen de datos que desea suprimir. Si es necesario, utilice Ctrl-pulsar para seleccionar orígenes de datos adicionales.
8. Cuando el botón **Suprimir no utilizados** esté activado, púlselo.
9. En el cuadro de diálogo Suprimir orígenes de datos no utilizados, pulse **Suprimir** para confirmar la purga del repositorio.

## Eliminación de un repositorio de entidad

Si ya no necesita un repositorio, puede eliminarlo completamente.

*Atención:* esta acción hace exactamente o que dice. **Esta acción no se puede deshacer.** Si no está seguro, utilice el botón **Purgar** para eliminar todos los datos de origen. Así no se eliminar la configuración del repositorio. Consulte el tema “Purga de un repositorio de entidad” en la página 35 para obtener más información.

*Note:* el siguiente procedimiento asume que puede conectarse al repositorio desde SPSS Modeler y que conoce el nombre de usuario y la contraseña de administrador de la base de datos donde se aloja el repositorio. Si no es el caso, siga el procedimiento para eliminar un repositorio cuando no se puede conectar a él. Consulte el tema “Eliminación de un repositorio cuando no se puede conectar a él” en la página 37 para obtener más información.

Para eliminar un repositorio

1. Abra un nodo Entity Analytics.
2. Pulse en la lista del **repositorio de entidad**.
3. Pulse en <Examinar...> para que se muestre el cuadro de diálogo Casos de resolución de entidades.
4. En el cuadro de diálogo Casos de resolución de entidades, pulse en la lista **Nombre del repositorio**.
5. Seleccione el repositorio que desea eliminar.
6. Si no está conectado, introduzca el nombre de usuario de administrador y la contraseña y pulse en **Conectar**.
7. Cuando el botón **Eliminar el repositorio completo** esté activado, púlselo.
8. Pulse en **Eliminar** para confirmar la eliminación del repositorio.

9. Pulse en **Aceptar** para aceptar que se ha eliminado correctamente.

## Eliminación de un repositorio cuando no se puede conectar a él

Use el siguiente procedimiento si desea eliminar un repositorio de entidad pero no puede conectarse al mismo, ya sea por problemas de conectividad con SPSS Modeler o porque no recuerda el nombre de usuario o la contraseña.

Realice este procedimiento en el equipo donde se aloja la base de datos del repositorio.

### Sistemas de Windows

1. Abra una ventana de Símbolo de sistema.
2. Escriba:

```
cd directorio_instalación_modelador
cd ext\bin\pasw.entityanalytics
delete_repository.bat nombre_repositorio
```

donde *directorio\_instalación\_modeler* es el directorio de instalación de SPSS Modeler y *nombre\_repositorio* es el nombre del repositorio.

**Nota:** El nombre de repositorio distingue entre mayúsculas y minúsculas.

3. Continúe desde "Completar el procedimiento" (más adelante en esta sección).

### Sistemas UNIX

1. Abra una shell.
2. Escriba:

```
cd directorio_instalación_servidor_modeler
cd ext/bin/pasw.entityanalytics
./delete_repository.sh nombre_repositorio
```

donde *directorio\_instalación\_servidor\_modeler* es el directorio de instalación de SPSS Modeler Server y *nombre\_repositorio* es el nombre del repositorio.

**Nota:** El nombre de repositorio distingue entre mayúsculas y minúsculas.

### Completar el procedimiento (todos los sistemas)

1. En el mensaje que se muestra, confirme la eliminación del repositorio especificando Y.
2. Elimine el directorio que tenga el mismo nombre que el repositorio que ha eliminado. Si no puede eliminar el directorio, reinicie el equipo e inténtelo de nuevo.



---

## Capítulo 4. El análisis de entidades en acción

---

### Acerca de este ejemplo

En este ejemplo, veremos cómo la adición de análisis de entidades puede mejorar todavía más los impresionantes resultados que pueden obtenerse mediante el uso de IBM SPSS Modeler.

Este ejemplo usa la ruta *loan\_entity\_analytics.str*, que hace referencia al archivo de datos *loan\_applications.csv*. Estos archivos están disponibles en el directorio *Demos* de cualquier instalación de IBM SPSS Modeler que tenga instalado también IBM SPSS Modeler Entity Analytics. Puede acceder al directorio *Demos* desde el grupo de programas de IBM SPSS Modeler en el menú Inicio de Windows. El archivo *loan\_entity\_analytics.str* se encuentra en el directorio *Entity\_Analytics*.

*Note:* Antes de que pueda ejecutar esta ruta de ejemplo, necesita crear un repositorio en su sistema. Hágalo antes de continuar con este ejemplo. Consulte el tema “Creación del repositorio” en la página 14 para obtener más información.

Comencemos por una situación conocida: a los ejecutivos de un banco les interesa saber si es probable que los clientes con una solicitud de préstamo pendiente no puedan pagar dichos préstamos. El departamento de TI del banco utiliza SPSS Modeler desde hace tiempo, por lo que su personal ya ha creado una ruta y ha generado un modelo predictivo a partir de los datos existentes acerca de 700 préstamos concedidos en el pasado. Puede que estos préstamos hayan sido devueltos o no por los clientes.

---

### El modelo original

Así generó el personal del banco su modelo y ésta es la información que obtuvieron.

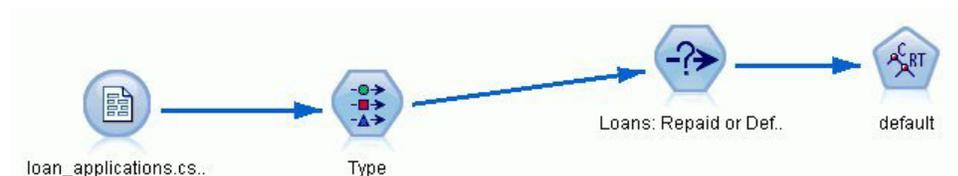


Figura 2. Ruta inicial con nodo de modelado

Además de detalles de préstamos anteriores, el conjunto de datos *loan\_applications.csv* incluye detalles de 150 clientes cuyas solicitudes de préstamos aún están pendientes, lo que daría un total de 850 registros.

No todos los campos del conjunto de datos resultan útiles en la predicción (los campos de nombre, por ejemplo, se pueden ignorar). El nodo Tipo filtra los campos que se deben ignorar estableciendo su rol como **Ninguno**. Los campos que se van a usar en la predicción tienen su rol establecido como **Entrada**, y el campo cuyo valor está tratando de predecir el modelo tiene el rol establecido como **Objetivo**.

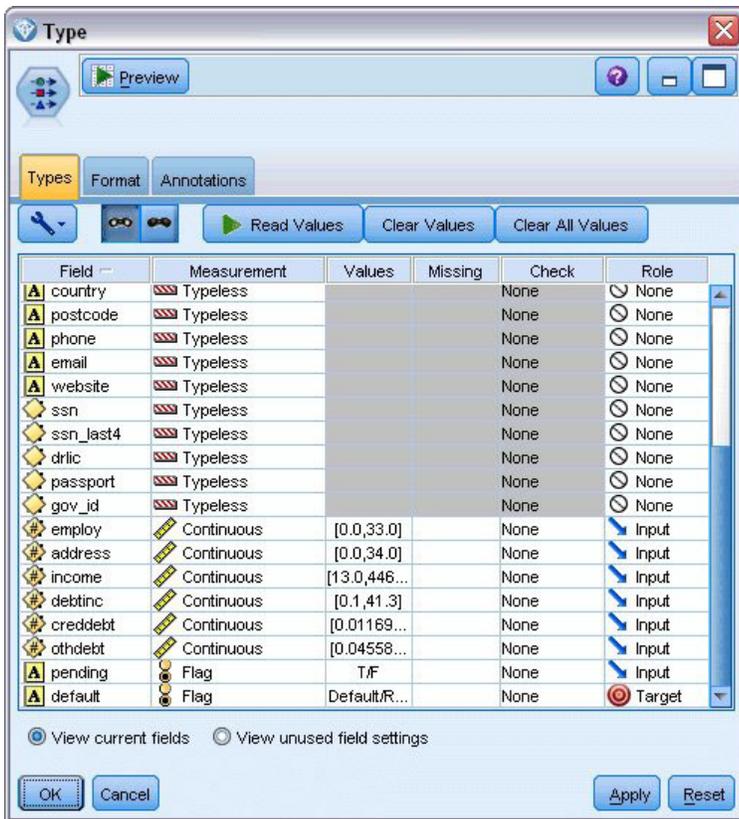


Figura 3. Roles de los campos establecidos en el nodo Tipo

Dado que el modelo debe realizar las predicciones basándose tan solo en datos pasados, la ruta incluye un nodo Seleccionar que incluye solo los préstamos *no* marcados como Pendientes, y por tanto descarta los 150 préstamos pendientes.

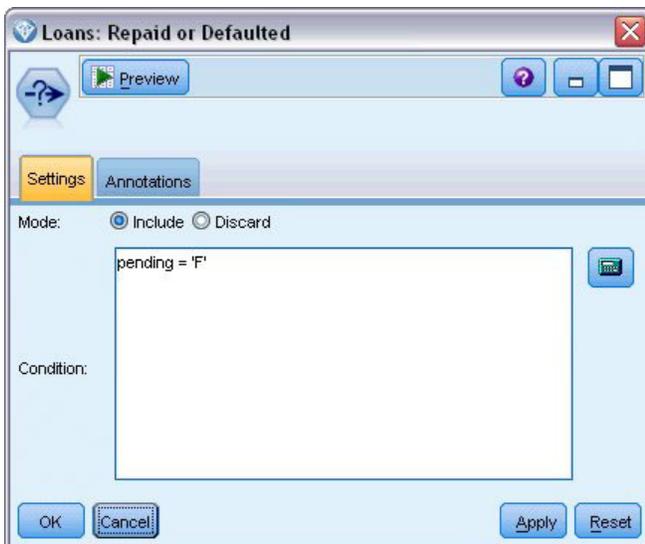


Figura 4. Descarte de las solicitudes de préstamo pendientes

Con los préstamos pendientes descartados, solo los 700 préstamos restantes, tanto los que se devolvieron como los que no, pasaron al nodo de modelado. El banco podría haber utilizado varios algoritmos de SPSS Modeler para generar un buen modelo. En este caso, han utilizado un nodo Árbol C&R, que se

empleará para generar un modelo que prediga quiénes son posibles morosos basándose en el comportamiento anterior de los clientes del banco.

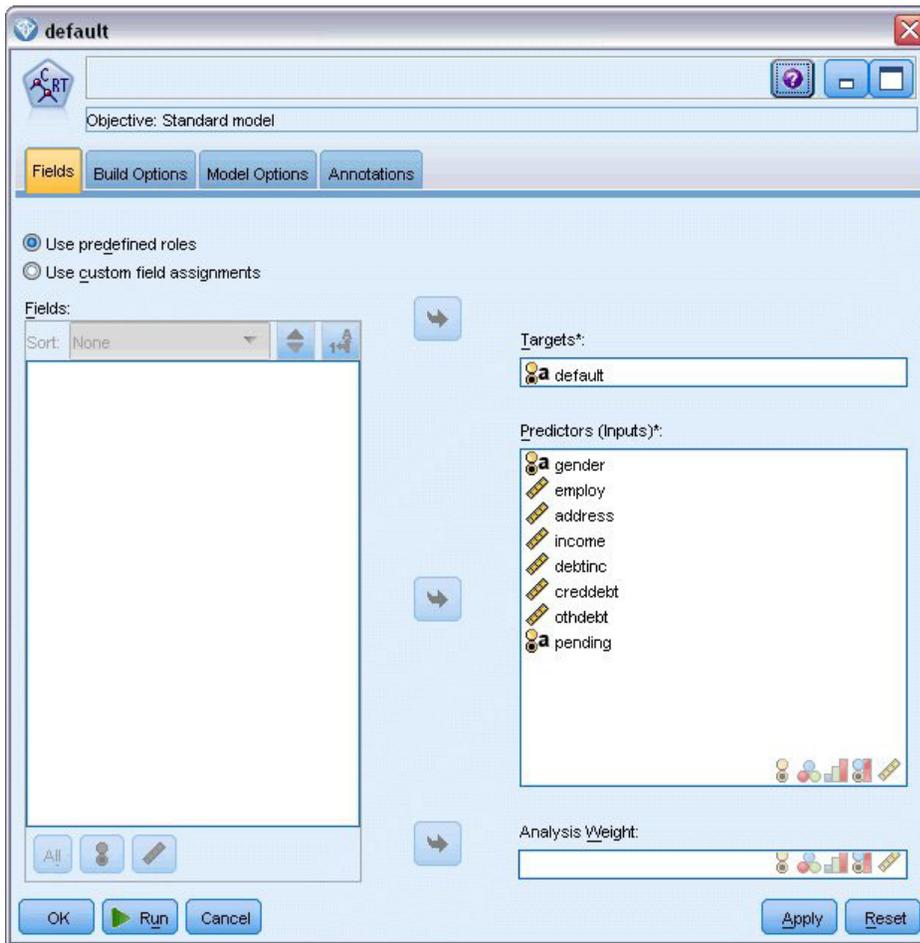


Figura 5. Asignación de campos objetivos y predictores

Los campos usados para hacer la predicción son designados como campos predictores, y el campo cuyo valor intenta predecir el modelo, **impago** en este caso, se establece como campo objetivo, como se definieron anteriormente en el nodo Tipo.

Al ejecutar esta ruta se genera un nugget de modelo, que contiene el modelo creado a partir de los campos predictores.

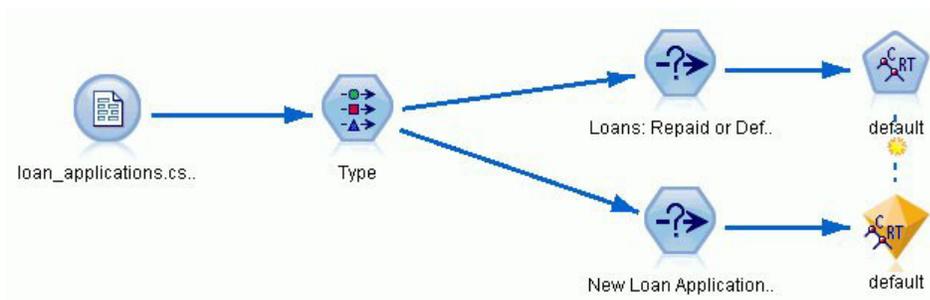


Figura 6. Ruta con nugget de modelo añadido

Ahora, el analista del banco puede utilizar el modelo para comenzar a predecir si es probable que los clientes con pagos pendientes no paguen. Utilizando el conjunto de datos original, el analista inserta un nodo Seleccionar que esta vez incluye solamente los 150 registros de préstamos marcados como pendientes, en lugar de descartarlos. El analista pasa estos registros directamente al modelo, y añade un nodo Distribución que representa visualmente las predicciones del modelo.

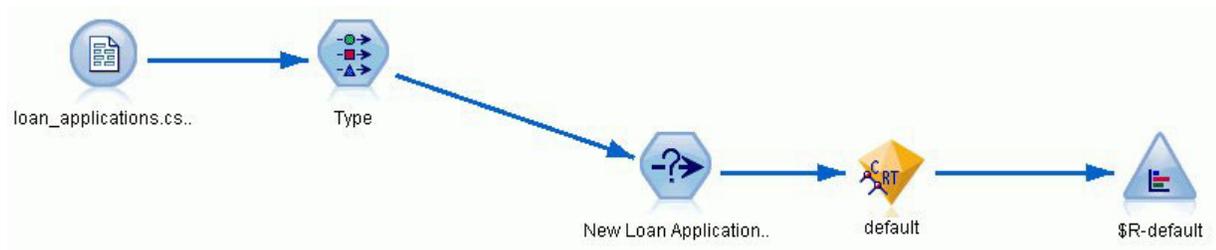


Figura 7. Ruta de selección de solicitudes de préstamos nuevas con nodo Distribución añadido

El nodo Distribución muestra la distribución de los valores del campo \$R-default en el modelo. Al ejecutarlo, el nodo Árbol C&R añade este campo al modelo de datos. El campo contiene la predicción sobre si el nuevo solicitante devolverá o no el préstamo, y posteriormente utilizaremos este campo para comparar el efecto de añadir nuevos análisis de entidades.

Al ejecutar esta parte de la ruta, los resultados del nodo Distribución indican al analista que es previsible que 137 de los 150 solicitantes nuevos devuelvan sus préstamos. Se prevé que los 13 restantes no pagarán sus préstamos, por lo que el analista probablemente recomendará al banco que rechace sus solicitudes.

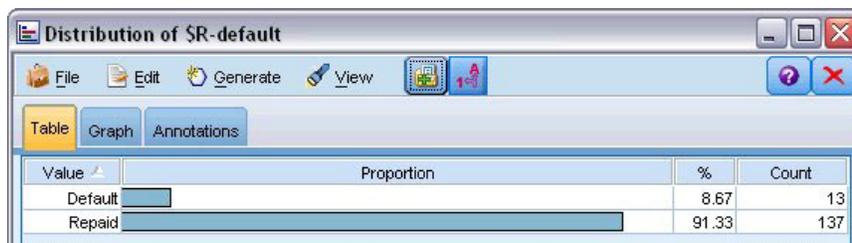


Figura 8. Resultados del nodo Distribución sin análisis de entidades

## Con análisis de entidades

Veamos ahora si la situación puede mejorarse añadiendo análisis de entidades a la ecuación. Imagine que usted es un especialista en análisis de entidades, y que el banco ha contratado sus servicios para investigar posibles entradas fraudulentas en el registro de clientes de los datos de origen. Puede que haya registros duplicados por errores en la entrada de datos, pero también es posible que el solicitante de un préstamo esté intentando falsear su identidad. En cualquier caso, el banco debe saber la verdad.

Para el propósito de este ejemplo, asumiremos que ya se ha creado un repositorio de entidad. Consulte el tema "Creación del repositorio" en la página 14 para obtener más información.

## Volcado de los datos de origen en el repositorio

En primer lugar, añada un nodo de exportación de EA al nodo de origen de datos para que pueda exportar los datos de origen al repositorio de la entidad.

Antes de poder exportar los datos, debe correlacionar los campos del origen de datos a características del repositorio de la entidad. Esto es necesario porque diferentes orígenes de datos pueden utilizar nombres de campos distintos para el mismo tipo de información. El repositorio de entidad proporciona un conjunto estándar de tipos de información (conocido como "características") para evitar duplicidades.

En el nodo Exportar EA, configure los detalles sobre el repositorio: detalles de conexión, etiqueta de origen (para identificar el origen de datos, **TEST** en este caso), tipo de entidad (conjunto de características que estamos utilizando, denominado **PERSON**) y el campo de clave exclusiva (para identificar cada registro unívocamente). En este caso, utilice el campo **clave** como clave exclusiva.

Ahora puede establecer las correlaciones. En el conjunto de características que está utilizando, hay características que corresponden a los campos *fname*, *mname*, *lname*, *generation*, *dob*, *gender*, *addr1*, *city*, *country*, *postcode*, *phone*, *email*, *ssn*, *drlic* y *passport*.

Establezca en primer lugar la correlación para *nombre*. Pulse dos veces en la columna **Correlacionado a característica** en la tabla en la fila *nombre*, desplácese hacia abajo hasta la entrada **NAME.GIVEN\_NAME** y pulse en ella para crear la correlación.

Ahora asigne los campos restantes que tienen características correspondientes, de modo que todo el conjunto de correlaciones tenga el siguiente aspecto.

Tabla 12. Campos correlacionados a características de repositorio.

Campo	Correlacionado a característica
<i>fname</i>	NAME.GIVEN_NAME
<i>mname</i>	NAME.MIDDLE_NAME
<i>lname</i>	NAME.SUR_NAME
<i>generation</i>	NAME.NAME_GEN
<i>dob</i>	DOB.DOB
<i>gender</i>	GENDER.GENDER
<i>addr1</i>	ADDRESS.ADDR1
<i>city</i>	ADDRESS.CITY
<i>country</i>	ADDRESS.COUNTRY
<i>postcode</i>	ADDRESS.POSTAL_CODE
<i>phone</i>	PHONE.PHONE_NUM
<i>email</i>	EMAIL_ADDR.ADDR
<i>ssn</i>	SSN.ID_NUM
<i>drlic</i>	DRLIC.ID_NUM
<i>pasaporte</i>	PASSPORT.ID_NUM

Pulse en **Ejecutar** para exportar los datos al repositorio. Este proceso tardará un poco, así que cuando se cierre el cuadro de diálogo Comentarios de la ejecución, la exportación se habrá completado.

## Lectura de las identidades resueltas

Al exportar los datos al repositorio, el sistema de análisis de entidades comienza a resolver posibles conflictos de identidad, asignando un identificador exclusivo de entidad, que posteriormente verá como el campo **\$EA-ID**. (Nota: no es lo mismo que el campo Clave exclusiva del nodo Exportar EA, ya que dicho campo se utiliza para identificar de forma exclusiva los registros de origen de datos.)

El primer paso para leer las identidades resueltas consiste en añadir un nodo de origen Entity Analytics(EA) a la ruta. Este nodo de origen no debe estar conectado a nada en esta fase.

Abra el nodo de origen Entity Analytics(EA) y configure los detalles del Repositorio de entidad. Se mostrará una lista con los orígenes de datos exportados al repositorio: en este caso solo hay una.

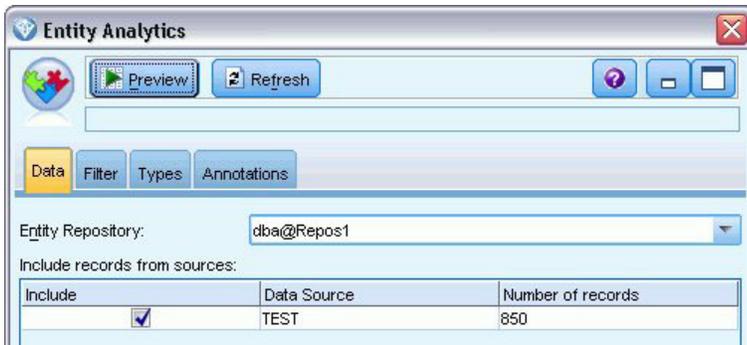


Figura 9. Selección de un origen de datos en el repositorio

Marque la casilla de verificación del origen de datos **TEST** y pulse en Aceptar.

Veamos lo que el sistema de análisis de entidades ha hecho con los datos. Adjunte un nodo Tabla al nodo de origen Entity Analytics(EA), abra el nodo Tabla y pulse en **Ejecutar** para ver la ventana de resultados del nodo Tabla.

	\$EA-ID	\$EA-SRC	\$EA-KEY
1	1	TEST	REF00000001
2	2	TEST	REF00000002
3	3	TEST	REF00000003
4	4	TEST	REF00000004
5	5	TEST	REF00000005
6	6	TEST	REF00000006
7	7	TEST	REF00000007
8	8	TEST	REF00000008
9	9	TEST	REF00000009
10	10	TEST	REF00000010
11	11	TEST	REF00000011
12	12	TEST	REF00000012
13	13	TEST	REF00000013
14	14	TEST	REF00000014
15	15	TEST	REF00000015
16	16	TEST	REF00000016
17	17	TEST	REF00000017
18	18	TEST	REF00000018
19	19	TEST	REF00000019
20	20	TEST	REF00000020

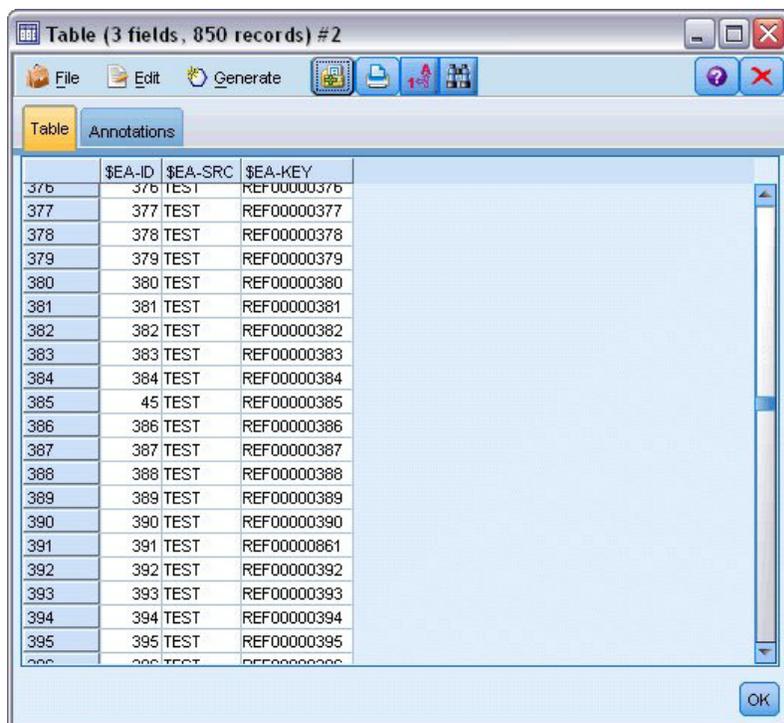
Figura 10. Resultados del nodo Tabla

Solo uno de los campos resulta familiar: el que tiene la etiqueta *\$EA-KEY*. En realidad, es el campo de *clave* de los datos de origen y aparece aquí porque lo seleccionó como campo Clave exclusiva en el nodo Exportar EA.

Sin embargo, el sistema ha añadido otros dos campos. El campo *\$EA-ID* es el identificador exclusivo, no de los registros de origen, sino de las identidades resueltas. Veremos la diferencia en un instante. El campo *\$EA-SRC* identifica de dónde proceden los datos: aquí indica **TEST** porque es la etiqueta de origen asignada en el nodo Exportar EA.

¿Qué ha ocurrido con el resto de campos en los datos de origen? No se preocupe, siguen en el repositorio: lo que ocurre es que, por razones de rendimiento, el nodo de origen Entity Analytics(EA) pasa solamente el conjunto mínimo de campos hacia abajo para su posterior procesamiento.

A continuación, desplácese hacia abajo en los resultados del nodo Tabla hasta la fila 385.



The screenshot shows a window titled "Table (3 fields, 850 records) #2". The window contains a table with three columns: "\$EA-ID", "\$EA-SRC", and "\$EA-KEY". The rows are numbered from 376 to 395. The "\$EA-ID" column shows a sequence of numbers from 376 to 384, followed by a jump to 45 at row 385, and then continues from 386 to 395. The "\$EA-SRC" column contains the word "TEST" for all rows. The "\$EA-KEY" column contains reference numbers like "REF00000376" through "REF00000395".

	\$EA-ID	\$EA-SRC	\$EA-KEY
376	376	TEST	REF00000376
377	377	TEST	REF00000377
378	378	TEST	REF00000378
379	379	TEST	REF00000379
380	380	TEST	REF00000380
381	381	TEST	REF00000381
382	382	TEST	REF00000382
383	383	TEST	REF00000383
384	384	TEST	REF00000384
385	45	TEST	REF00000385
386	386	TEST	REF00000386
387	387	TEST	REF00000387
388	388	TEST	REF00000388
389	389	TEST	REF00000389
390	390	TEST	REF00000390
391	391	TEST	REF00000861
392	392	TEST	REF00000392
393	393	TEST	REF00000393
394	394	TEST	REF00000394
395	395	TEST	REF00000395

Figura 11. Diferencias entre filas de resultados de Tabla y números de \$EA-ID

Observe cómo el número de \$EA-ID parece que no pertenece a esta secuencia. El sistema de análisis de entidades ha determinado que el registro REF00000385 hace referencia a la persona identificada como entidad 45, que también tiene el registro REF00000045. Si se sigue bajando en los resultados, hay más números que no siguen la secuencia, por ejemplo en las filas 485, 517, 520, etc. Será mejor que lo analicemos detenidamente.

En primer lugar, vamos a resaltar el hecho de que el campo \$EA-KEY contiene los datos del campo de clave en los datos de origen, cambiándole el nombre a clave. Adjunte un nodo Filtrar al nodo de origen Entity Analytics(EA) y abra el nodo Filtrar. Efectúe una doble pulsación en la cadena \$EA-KEY en la segunda columna **Campo** y escriba clave.

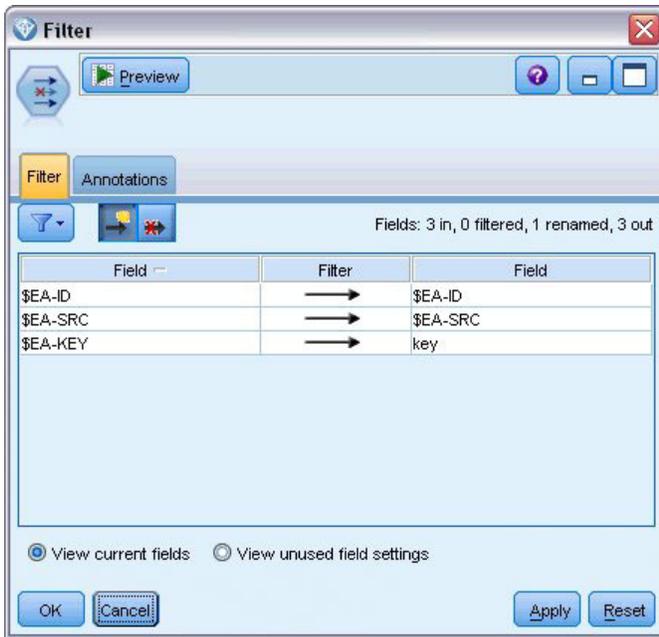


Figura 12. Cambio de nombre del campo \$EA-KEY

Pulse en **Aceptar** para cerrar el nodo Filtrar.

Ahora es necesario ordenar los ID de entidad de \$EA-ID en orden ascendente. Conecte un nodo Clasificar al nodo Filtrar. Abra el nodo Clasificar, pulse en el botón superior junto a la tabla **Ordenar por**, seleccione **\$EA-ID** y pulse en **Aceptar**.



Figura 13. Clasificación de los ID de entidad en orden ascendente

Deje el orden de clasificación como **Ascendente** y pulse en **Aceptar**.

Ahora debe crear un campo adicional que indique si un registro es exclusivo o está duplicado. Conecte un nodo Derivar al nodo Clasificar. Abra el nodo Derivar y establezca el nombre del campo **Derivar** en IsDuplicate. En la lista **Derivar como**, elija **Marca**, que además establece el **tipo de Campo** como **Marca**. Establezca el campo del **valor Verdadero** como **Duplicado** y el campo del **valor Falso** como **Exclusivo**.

Para encontrar registros duplicados, utilizará una función de secuencia especial, denominada @OFFSET, que se proporciona con SPSS Modeler.

Escriba lo siguiente en el campo If:

```
'$EA-ID' = @OFFSET('$EA-ID',1) o '$EA-ID' = @OFFSET('$EA-ID',-1))
```



Figura 14. Establecimiento de la condición en el nodo Derivar

Una vez clasificados los ID de entidad en orden ascendente, la función @OFFSET comprueba si dos ID de entidad adyacentes son idénticos, en cuyo caso los dos registros son duplicados. En ese caso, el valor *Es duplicado* se establece como *Duplicado*; de lo contrario se establece como *Exclusivo*.

Pulse en **Aceptar** para cerrar el nodo.

Para ver el efecto del nodo Derivar, adjunte un nodo Tabla al nodo Derivar, abra el nodo Tabla y pulse en **Ejecutar**. Desplácese hacia abajo en la ventana de resultados del nodo Tabla hasta la fila 45.

	\$EA-ID	\$EA-SRC	key	IsDuplicate
39	39	TEST	REF00000039	Unique
40	40	TEST	REF00000040	Unique
41	41	TEST	REF00000041	Unique
42	42	TEST	REF00000042	Unique
43	43	TEST	REF00000043	Unique
44	44	TEST	REF00000044	Unique
45	45	TEST	REF00000045	Duplicate
46	45	TEST	REF00000385	Duplicate
47	46	TEST	REF00000046	Unique
48	47	TEST	REF00000047	Unique
49	48	TEST	REF00000048	Unique
50	49	TEST	REF00000049	Unique
51	50	TEST	REF00000050	Unique
52	51	TEST	REF00000051	Unique
53	52	TEST	REF00000052	Unique
54	53	TEST	REF00000053	Unique
55	54	TEST	REF00000054	Unique
56	55	TEST	REF00000055	Unique
57	56	TEST	REF00000056	Unique
58	57	TEST	REF00000057	Unique

Figura 15. Resultados del nodo Derivar

Recuerde que hemos visto los resultados directamente desde el nodo de origen Entity Analytics(EA). El sistema ya ha identificado que el registro REF00000385 hace referencia a la misma persona que la entidad 45. Ahora hemos ido un poco más lejos y hemos marcado que los registros REF00000045 y REF00000385 son duplicados, ya que ambos hacen referencia a la entidad 45.

Siga bajando en la ventana de resultados y podrá ir viendo el resto de registros marcados como duplicados.

Para obtener un informe que enumere los registros duplicados, conecte un nodo Informe (en la pestaña Resultados de la paleta de nodos) al nodo Derivar *Es duplicado*. Abra el nodo Informe, copie el siguiente texto en el campo de entrada de la pestaña Plantilla y pulse en **Ejecutar**.

```
<html>
<h1>Lista de registros de cliente duplicados.

<h2>Se ha generado este informe: [@TODAY]

<h2>Registros duplicados
<table>
  <tr>
    <td>ID de entidad</td>
    <td>Clave</td>
  </tr>

#WHERE IsDuplicate = "Duplicate"

  <tr>
    <td>['$EA-ID']</td>
    <td>[key]</td>
  </tr>
```

```
#  
</table>  
</html>
```

Esto genera el siguiente resultado.

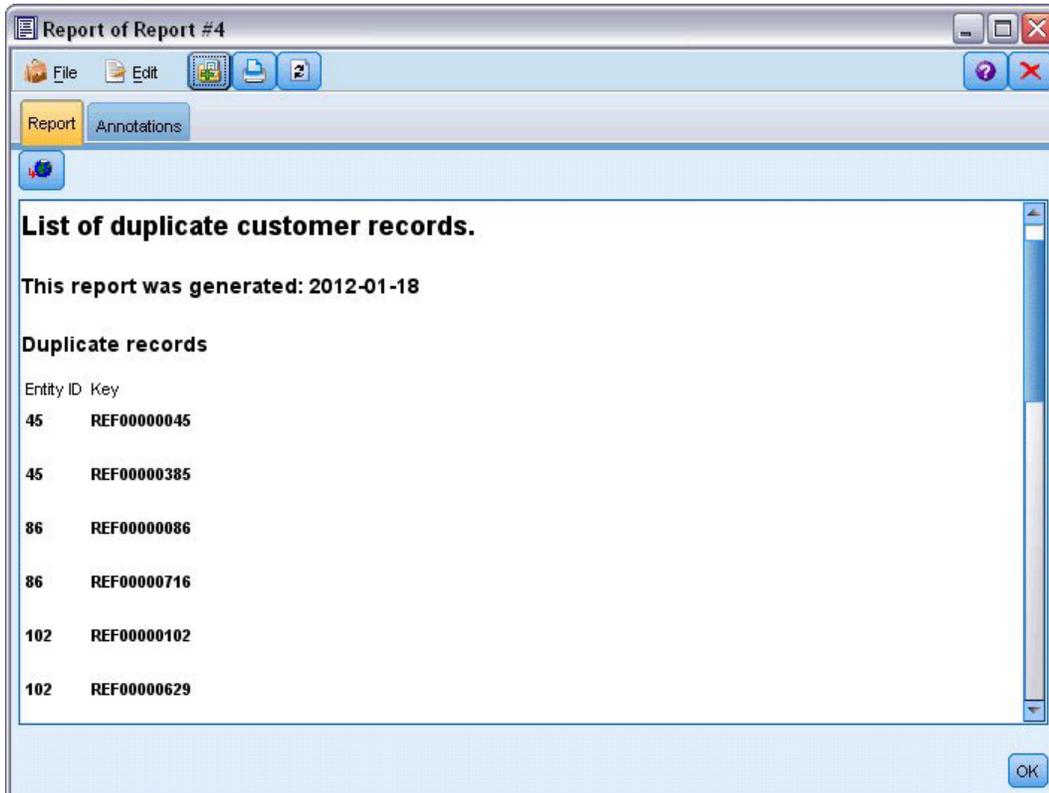


Figura 16. Resultado del nodo Informe

El informe utiliza formato HTML en este caso, aunque también puede utilizar XML o ASCII.

## Comparación de los resultados del análisis de entidades con el modelo original

La última fase de este ejemplo consiste en comprobar si al añadir el análisis de entidades se ve afectado la predicción original del banco. Recuerde que el modelo original predijo que 13 de los 150 solicitantes no pagarían. Va a utilizar un nodo Fundir para fundir los resultados del modelo con información acerca de los registros duplicados del análisis de entidades y comprobar si cambia la predicción.

En primer lugar, asegúrese de que los nuevos campos añadidos por el análisis de entidades incluyen los tipos de datos correctos, o los *niveles de medición*, como se conocen en SPSS Modeler. Adjunte un nodo Tipo al nodo Derivar **Es duplicado**, abra el nodo Tipo y pulse en el botón **Leer valores**.

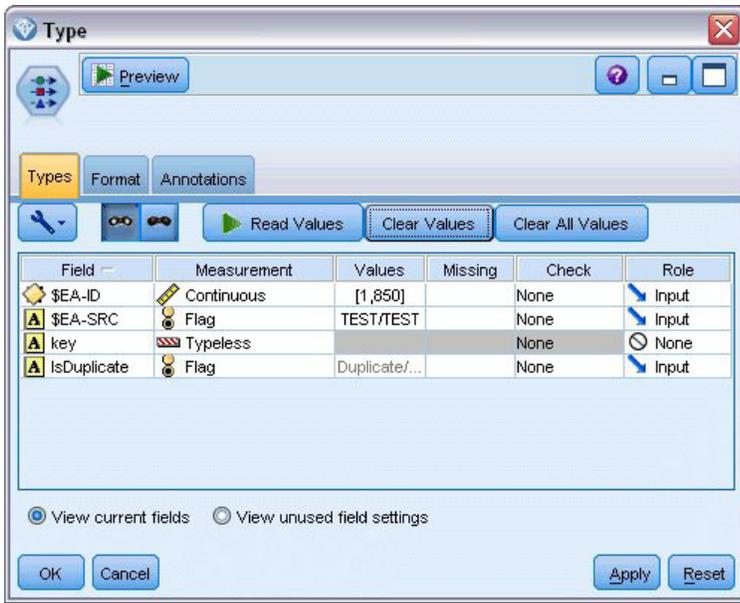


Figura 17. Configuración del nodo Tipo

Ahora puede añadir el nodo Fundir. Conéctelo al nodo Tipo y al nugget dorado que contiene el modelo original. Para ello, pulse con el botón derecho en el nugget dorado, elija **Conectar** y pulse en el nodo Fundir, que ahora debería tener dos flechas de entrada.

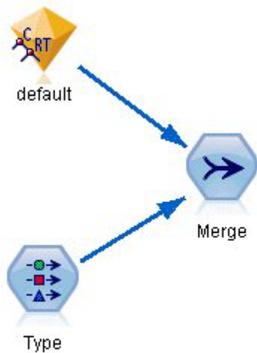


Figura 18. Entradas al nodo Fundir

Abra el nodo Fundir, en **Método de fusión** elija **Claves**, y pulse el botón de la flecha derecha para mover el campo **clave** de **Claves posibles** a **Claves para fusión** y después pulse en **Aceptar**.

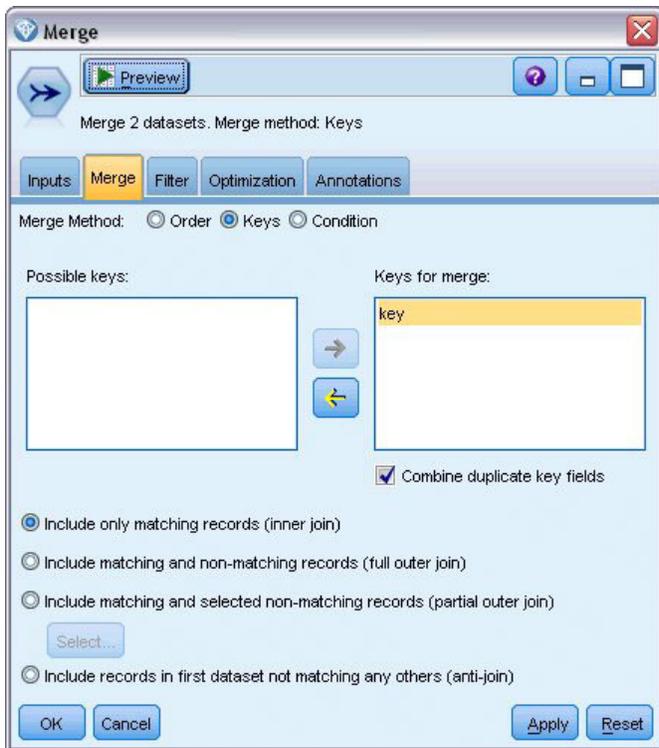


Figura 19. Especificación del campo clave para la operación de fusión

Ya está todo prácticamente preparado para realizar la comparación. No obstante, si adjuntase un nodo Distribución y lo ejecutase en este punto, no notaría ninguna diferencia con respecto a la predicción original. Pese a que la ruta funde ahora los resultados del nugget de modelo original con los nuevos campos creados por el análisis de entidades, el propio campo de predicción (*\$R-default*) en el modelo de datos no se ha actualizado con la nueva información.

Para ello, utilice un nodo Rellenar, que puede sustituir valores de campos. Conecte un nodo Rellenar al nodo Fundir y abra el nodo Rellenar.

Pulse en el botón superior a la derecha de **Rellenar campos**, desplácese hacia la parte inferior de la lista, elija **\$R-default** y pulse en **Aceptar**. Este es el campo cuyos valores se cambiarán si se cumple la condición especificada en el resto del cuadro de diálogo.

Para especificar la condición, asegúrese de que **Reemplazar** está fijado en **En función de una condición**, y en el campo **Condición** introduzca:

```
default /="default" e IsDuplicate = "Duplicate"
```

En el campo **Reemplazar con**, introduzca:

```
"default"
```



Figura 20. Especificación de la condición para reemplazar valores de campo

Esta configuración requiere una breve explicación. La condición establece que, en los registros en que el valor del campo *valor predeterminado* del conjunto de datos original no coincida con el **Valor predeterminado** y el registro haya sido marcado como duplicado, el valor del campo *\$R-default* en el modelo se establezca **valor predeterminado**.

El campo *\$R-default* es el campo del modelo que contiene la predicción de si es probable que un cliente no devuelva el préstamo. De este modo, los clientes con registros duplicados se añaden al modelo como morosos en potencia.

Pulse en **Aceptar** para cerrar el nodo Rellenar.

Finalmente, ya puede ver la diferencia que supone el análisis de entidades. En la paleta Gráficos, conecte un nodo Distribución al nodo Rellenar y abra el nodo Distribución. Pulse en la lista **Campo** y seleccione **\$R-default**.

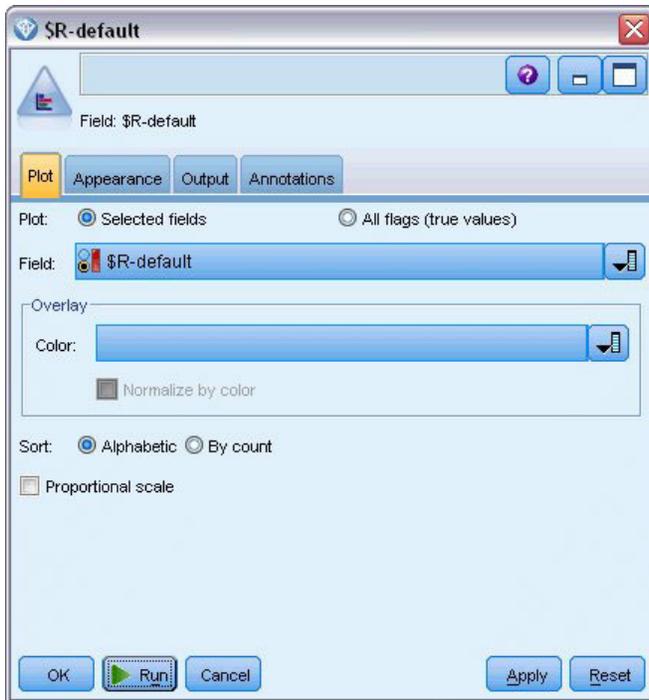


Figura 21. Configuración del nodo Distribución

Pulse en **Ejecutar** para generar el gráfico de la nueva predicción.

Value	Proportion	%	Count
Default		10.67	16
Repaid		89.33	134

Figura 22. Resultados del nodo Distribución tras el análisis de entidades

Ahora hay 16 solicitudes de riesgo en lugar de 13. Estas solicitudes adicionales podrían haber resultado muy costosas de no haber pagado, de modo que puede demostrar gráficamente al banco las ventajas de añadir el análisis de entidades a sus operaciones de evaluación de riesgos.

## Resumen

Este ejemplo ha mostrado cómo se puede eliminar la duplicidad de registros en los datos de personas u organizaciones mediante el análisis de entidades, lo que proporciona una mejora en la calidad de las predicciones.

*Note:* se recomienda eliminar los registros duplicados antes de llevar a cabo cualquier otro proceso. Puede continuar este proceso utilizando el nodo de preparación automática de datos (ADP) para analizar sus datos e identificar los valores fijos, cribar los campos problemáticos o que no serán útiles, derivar nuevos atributos cuando sea necesario y mejorar el rendimiento mediante técnicas de cribado inteligente.

La combinación del análisis de entidades y la preparación automática de datos puede ayudarle a garantizar que está trabajando con datos lo más limpios posible.



---

## Apéndice. Propiedades de los scripts para IBM SPSS Modeler Entity Analytics

---

### Creación de scripts con IBM SPSS Modeler Entity Analytics

El procesamiento en IBM SPSS Modeler Entity Analytics es una herramienta potente para automatizar procesos en la interfaz de usuario. Los scripts pueden realizar los mismos tipos de acciones que se realizan con el ratón o el teclado y se utilizan para automatizar tareas que resultarían extremadamente repetitivas o llevarían mucho tiempo si se realizaran manualmente. Para obtener una explicación del uso de los scripts, consulte la guía *ScriptingAutomation.pdf* que se incluye con IBM SPSS Modeler.

---

### Propiedades comunes

En la siguiente tabla se indican las propiedades que son comunes para los nodos de IBM SPSS Modeler Entity Analytics. En los apartados siguientes se ofrece información específica de los nodos.

Tabla 13. Propiedades comunes

Nombre de la propiedad	Tipo de datos	Descripción de la propiedad
entity_repository	['campo','campo', ... , 'campo']	Serie de conexión del repositorio. Formato: [ 'nombre_repositorio', 'nombre_usuario', 'contraseña' ]  Ejemplo: entity_repository = [ 'repos1', 'dba', 'psw1' ]
entity_type	cadena	Tipo de entidad (conjunto de características) que debe utilizarse.  Ejemplo: entity_type = 'PERSON'

---

### Propiedades entityanalytics\_exportnode



El nodo de exportación de EA es un nodo terminal que lee datos de entidades desde un origen de datos y exporta los datos a un repositorio para la resolución de entidades.

Tabla 14. Propiedades entityanalytics\_exportnode

Propiedades entityanalytics_exportnode	Tipo de datos	Descripción de la propiedad
mode	Add PurgeFirst	Modo de exportación. Add añade registros de archivo de origen a contenido existente del repositorio; PurgeFirst elimina el contenido existente antes de exportar.
source_tag	cadena	Identificador del origen de datos.  Ejemplo: source_tag = 'CUST'

Tabla 14. Propiedades entityanalytics\_exportnode (continuación)

Propiedades entityanalytics_exportnode	Tipo de datos	Descripción de la propiedad
unique_key_field	cadena	Campo de entrada que se utiliza para identificadores exclusivos para registros de datos.  Ejemplo: unique_key_field = 'ID'
field_mapping	[ [ 'nombre_campo' 'característica.elemento' 'tipo_uso' ]... ]	Correlaciona campos de entrada con la característica correspondiente del repositorio.  Ejemplo: field_mapping = [[ 'fname' 'NAME.GIVEN_NAME' '' ] [ 'addr1' 'ADDRESS.ADDR1' 'PRIMARY' ]] <i>Nota:</i> para establecer el <i>tipo_uso</i> en el equivalente de "(Auto)", utilice '' como en el primer ejemplo.

## Propiedades entityanalytics\_sourcenode



El nodo de origen Entity Analytics(EA) lee las entidades resueltas del repositorio y pasa estos datos a la ruta para procesos adicionales (por ejemplo, darles formato en un informe).

Tabla 15. Propiedades entityanalytics\_sourcenode

Propiedades entityanalytics_sourcenode	Tipo de datos	Descripción de la propiedad
source_tags	lista	Lista de códigos para orígenes de datos que deben extraerse del repositorio.  Ejemplo: source_tags=[ 'LOANS', 'CUSTOMERS' ]
relationships	None Close All	Criterios de coincidencia para recuperar detalles de relaciones del repositorio. None no devuelve ninguna relación. Close devuelve coincidencias próximas según detalles como, por ejemplo, el grado de separación. All devuelve todas las relaciones posibles.
max_degree_separation	entero	Mínimo 0, máximo 3.
output_entity_type	cadena	Lista de tipos de entidad que se utilizan en el repositorio.

## Propiedades entityanalytics\_processnode



El nodo Generación de análisis de entidad compara nuevos casos con los datos de entidad del repositorio.

Tabla 16. Propiedades entityanalytics\_processnode

Propiedades entityanalytics_processnode	Tipo de datos	Descripción de la propiedad
match	Exact ByIdentifier All	Criterios de coincidencia para recuperar entidades del repositorio. Exact devuelve solamente coincidencias exactas. ByIdentifier devuelve coincidencias exactas y entidades que comparten los mismos identificadores. All devuelve todas las coincidencias posibles.
save_search_records	booleano	
relationships	None Close All	Criterios de coincidencia para recuperar detalles de relaciones del repositorio. None no devuelve ninguna relación. Close devuelve coincidencias próximas según detalles como, por ejemplo, el grado de separación. All devuelve todas las relaciones posibles.
max_degree_separation	entero	Mínimo 0, máximo 3.
output_entity_type	cadena	Lista de tipos de entidad que se utilizan en el repositorio.



---

## Avisos

Esta información se ha desarrollado para los productos y servicios ofrecidos en todo el mundo.

Es posible que IBM no ofrezca los productos, servicios o características tratados en este documento en otros países. Consulte al representante local de IBM para obtener información sobre los productos y servicios que se pueden adquirir actualmente en su zona geográfica. Las referencias a un programa, producto o servicio de IBM no pretenden afirmar ni implicar que solo se pueda utilizar el producto, programa o servicio de IBM. En su lugar, se puede utilizar cualquier producto, programa o servicio funcionalmente equivalente que no infrinja ninguno de los derechos de propiedad intelectual de IBM. Sin embargo, es responsabilidad del usuario evaluar y comprobar el funcionamiento de todo producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o solicitudes de patentes pendientes que cubran el tema principal descrito en este documento. Este documento no le otorga ninguna licencia para estas patentes. Puede enviar preguntas acerca de las licencias, por escrito, a:

IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785  
Estados Unidos

Para consultas sobre licencias relacionadas con información de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe sus consultas, por escrito, a:

Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan Ltd.  
1623-14, Shimotsuruma, Yamato-shi  
Kanagawa 242-8502 Japón

El párrafo siguiente no se aplica al Reino Unido ni a ningún otro país donde estas disposiciones sean incompatibles con la legislación local: INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL" SIN GARANTÍAS DE NINGÚN TIPO, NI EXPLÍCITAS NI IMPLÍCITAS, INCLUYENDO PERO NO LIMITÁNDOSE A ELLAS, LAS GARANTÍAS IMPLÍCITAS DE NO VULNERACIÓN, COMERCIALIZACIÓN O IDONEIDAD PARA UN FIN CONCRETO. Algunos estados no permiten la renuncia a expresar o a garantías implícitas en determinadas transacciones, por lo tanto, esta declaración no se aplica a usted.

Esta información puede incluir imprecisiones técnicas o errores tipográficos. Periódicamente, se efectúan cambios en la información aquí y estos cambios se incorporarán en nuevas ediciones de la publicación. IBM puede efectuar mejoras y/o cambios en los productos y/o programas descritos en esta publicación en cualquier momento y sin previo aviso.

Cualquier referencia a sitios Web que no sean de IBM en esta información solamente es ofrecida por comodidad y de ningún modo sirve como aprobación de esos sitios Web. La información de esos sitios web no forma parte de la información del presente producto de IBM y la utilización de esos sitios web se realiza bajo la responsabilidad del usuario.

IBM puede utilizar o distribuir la información que proporciona el usuario de la manera que IBM crea apropiada sin incurrir en ninguna obligación con el usuario.

Los propietarios de licencia de este programa que deseen tener información sobre el mismo con el fin de: (i) intercambiar información entre programas creados de forma independiente y otros programas (incluido éste) y (ii) utilizar mutuamente la información que se ha intercambiado, deberán ponerse en contacto con:

IBM Software Group  
ATTN: Licensing  
200 W. Madison St.  
Chicago, IL; 60606  
Estados Unidos

Esta información estará disponible, bajo las condiciones adecuadas, incluyendo en algunos casos el pago de una cuota.

El programa bajo licencia que se describe en este documento y todo el material bajo licencia disponible, los proporciona IBM de acuerdo con los términos de las Condiciones Generales de IBM, Acuerdo Internacional de Programas Bajo Licencia de IBM o cualquier acuerdo equivalente entre las partes.

Cualquier dato de rendimiento mencionado aquí ha sido determinado en un entorno controlado. Por lo tanto, los resultados obtenidos en otros entornos operativos pueden variar de forma significativa. Es posible que algunas mediciones se hayan realizado en sistemas en desarrollo y no existe ninguna garantía de que estas medidas sean las mismas en los sistemas comerciales. Además, es posible que algunas mediciones hayan sido estimadas a través de extrapolación. Los resultados reales pueden variar. Los usuarios de este documento deben consultar los datos que corresponden a su entorno específico.

Se ha obtenido información acerca de productos que no son de IBM de los proveedores de esos productos, de sus publicaciones anunciadas o de otros orígenes disponibles públicamente. IBM no ha probado dichos productos y no puede confirmar la precisión en cuanto al rendimiento, la compatibilidad u otras características relacionadas con productos que no son de IBM. Las preguntas acerca de las aptitudes de productos que no sean de IBM deben dirigirse a los proveedores de dichos productos.

Todas las declaraciones sobre el futuro del rumbo y la intención de IBM están sujetas a cambio o retirada sin previo aviso y representan únicamente metas y objetivos.

Esta información contiene ejemplos de datos e informes utilizados en operaciones comerciales diarias. Para ilustrarlos lo máximo posible, los ejemplos incluyen los nombres de las personas, empresas, marcas y productos. Todos esos nombres son ficticios y cualquier parecido con los nombres y direcciones utilizados por una empresa real es pura coincidencia.

Si está viendo esta información en copia electrónica, es posible que las fotografías y las ilustraciones en color no aparezcan.

---

## **Marcas comerciales**

IBM, el logotipo de IBM e [ibm.com](http://ibm.com) son marcas registradas o marcas comerciales registradas de International Business Machines Corp., registrada en muchas jurisdicciones en todo el mundo. Otros nombres de productos y servicios pueden ser marcas registradas de IBM o de otras empresas. Hay una lista actual de marcas registradas de IBM disponible en la web en "Copyright and trademark information" en [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Intel, el logotipo de Intel, Intel Inside, el logotipo de Intel Inside, Intel Centrino, el logotipo de Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium y Pentium son marcas comerciales o marcas registradas de Intel Corporation o sus filiales en Estados Unidos y otros países.

Linux es una marca registrada de Linus Torvalds en Estados Unidos, otros países o ambos.

Microsoft, Windows, Windows NT, y el logotipo de Windows son marcas comerciales de Microsoft Corporation en Estados Unidos, otros países o ambos.

UNIX es una marca registrada de The Open Group en Estados Unidos y otros países.

Java y todos los logotipos y marcas registradas basadas en Java son marcas registradas de Oracle y/o sus filiales.

Otros nombres de productos y servicios pueden ser marcas registradas de IBM o de otras empresas.



---

# Índice

## A

- análisis de entidades
  - comparado con análisis predictivo 2
  - definido 1
  - utilizar con IBM SPSS Modeler 5
  - utilizar con otros productos IBM SPSS 32
- anonimización de características
  - repositorio de entidad 22
- anonimizar características
  - repositorio de entidad 22
- asignaciones de puertos
  - configuración para análisis de entidades 33

## C

- cambio del nombre
  - campos de datos para análisis de entidades 27
- campos de correlaciones
  - a las características del repositorio de entidad 7, 16, 17, 18, 19, 29, 30
- características
  - repositorio de entidad 7, 16, 17, 18, 19, 20, 21, 29, 30
- claves exclusivas
  - análisis de entidades 7
  - repositorio de entidad 17
- coincidencia de entidades,
  - establecimiento de umbral 25
- configuración
  - repositorio de entidad 18, 25, 26
- credenciales de administrador
  - gestionar para análisis de entidades 33

## E

- eliminación
  - repositorio de entidad 36, 37
- etiquetas de origen
  - repositorio de entidad 17
- exportación
  - datos a un repositorio de entidad 7

## I

- identidades resueltas, analizar con análisis de entidades 26
- información de tipos, establecimiento para análisis de entidades 27

## N

- nodo de exportación de EA, análisis de entidades 13
- Nodo de origen Entity Analytics(EA) 26

- nodo Generación de análisis de entidad, análisis de entidades 28
- nodos
  - adición a una ruta de análisis de entidades 28
- nodos de exportación
  - análisis de entidades 7, 13
- nodos de origen
  - análisis de entidades 9, 26
- nodos de proceso
  - análisis de entidades 10, 28
- nuevos casos, comparar con repositorio de análisis de entidades 28

## O

- origen de datos, selección para análisis de entidades 26
- orígenes de datos
  - que conectan con el análisis de entidades 6, 14
  - visualización del análisis de entidades 16, 30

## P

- propiedades
  - scripts 55
- propiedades de ruta
  - definir para análisis de entidades 34
- purga
  - repositorio de entidad 35

## R

- reglas de resolución, análisis de entidades 25
- repositorio
  - administración del análisis de entidades 32
  - análisis de entidades 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 28, 29, 30, 35, 36, 37
  - directorio de almacenamiento para análisis de entidades, cambiar 34
  - repositorio de entidad 13
  - anonimización 22
  - características 21
  - comparar nuevos casos 28
  - conectar con IBM SPSS Modeler 7
  - configuración 18, 25, 26
  - configuración de asignaciones de puertos 33
  - creación 6, 13, 14, 15
  - definir propiedades de ruta 34
  - eliminación 36, 37
  - gestionar las credenciales de administrador 33
  - mantenimiento 20

- repositorio de entidad (*continuación*)
  - mover a un directorio de almacenamiento distinto 34
  - opciones 16
  - purga 35
  - supresión de datos no utilizados 36
  - tareas administrativas 32
- resolución de identidades, análisis de entidades 7
- resolver identidades, análisis de entidades 7
- resultados
  - de análisis de entidades 31

## S

- scripts
  - propiedades 55
- supresión de datos no utilizados
  - repositorio de entidad 36

## T

- tipos de entidad
  - análisis de entidades 23
  - repositorio de entidad 17
- tipos de uso, análisis de entidades 23

## U

- umbral para coincidencia de entidades, análisis de entidades 25







Impreso en España