

IBM InfoSphere Information Server
Versión 11 Release 3

*Guía de casos de ejemplo de
integración*



IBM InfoSphere Information Server
Versión 11 Release 3

*Guía de casos de ejemplo de
integración*



Nota

Antes de utilizar esta información y el producto al que da soporte, lea la información del apartado "Avisos y marcas registradas" en la página 29.

Contenido

Casos de ejemplo de integración de InfoSphere Information Server 1

Creación de un caso de ejemplo de aplicación de integración de datos 4

Modernización de un depósito de datos y descripción de un caso de ejemplo sobre la calidad de datos . . . 6

Integración de la supervisión y la calidad de los datos 7

IBM Industry Models 8

Preguntas comunes que se deben abordar 9

Entradas de caso de ejemplo clave. 10

Procesos de entrada de caso de ejemplo clave . . 11

Salidas de caso de ejemplo clave 16

Apéndice A. Accesibilidad de los productos 21

Apéndice B. Cómo ponerse en contacto con IBM 23

Apéndice C. Acceso a la documentación del producto 25

Apéndice D. Cómo aportar comentarios sobre la documentación del producto . 27

Avisos y marcas registradas 29

Índice 35

Casos de ejemplo de integración de InfoSphere Information Server

La integración de información es una actividad compleja que afecta a cada parte de una organización. Para abordar los problemas de integración empresarial más comunes, estos casos de ejemplo de integración muestran cómo desplegar y utilizar los componentes de IBM® InfoSphere Information Server e InfoSphere Foundation Tools conjuntamente de manera integrada. Los casos de ejemplo de integración se centran en la calidad de datos dentro de una implementación de depósito de datos.

Retos de la integración de datos

Hoy día, las organizaciones se enfrentan a una amplia gama de retos relacionados con la información: problemas variados, y a menudo desconocidos, de la calidad de los datos, disputas sobre el significado y el contexto de la información, gestión de varias transformaciones complejas, aprovechamiento de procesos de integración existentes en vez de duplicar el esfuerzo, cantidades de datos en constante crecimiento, empequeñecimiento de ventanas de proceso y la creciente necesidad de supervisión y seguridad con el fin de garantizar el cumplimiento de la ley nacional e internacional.

Las organizaciones deben optimizar y conectar la información y los sistemas en distintos dominios de empresa con una infraestructura de información integrada. La información desconectada impide que las organizaciones de TI puedan responder rápidamente a nuevas solicitudes de información de usuarios y ejecutivos de empresa. También resulta difícil para las empresas, que disponen de pocas herramientas o recursos para hacer un seguimiento de la difusión de la información, supervisar la calidad de los datos y aplicar de forma coherente reglas empresariales. Por consiguiente, la información permanece esparcida en la empresa bajo una multitud de categorías desorganizadas y descripciones incompatibles.

Algunos de los temas de integración de datos clave son los siguientes:

- Los metadatos de origen de aplicación empresarial no se ensamblan fácilmente en un solo lugar para comprender qué hay disponible realmente. La combinación también puede incluir orígenes de legado, que a menudo no hacen disponibles los metadatos a través de una interfaz de programación de aplicaciones (API) estándar, si los hacen disponibles en absoluto.
- Los datos de referencia maestros, los nombres y las direcciones de los proveedores y clientes, y los números de piezas y las descripciones difieren en las aplicaciones y orígenes duplicados de estos datos.
- Deben escribirse cientos de trabajos de extracción, transformación y carga (ETL) para mover los datos de todos los orígenes a la nueva aplicación de destino.
- Las transformaciones de datos son necesarias antes de cargar los datos para que se ajusten a las nuevas estructuras del entorno.
- La capacidad de gestionar grandes cantidades de datos que se puedan ejecutar a través del proceso y terminen a tiempo es esencial. Las empresas necesitan una infraestructura que soporte la ejecución de cualquiera de las rutinas de transformación y coincidencia de datos bajo demanda.
- No hay una vista consolidada de la calidad de datos para toda la organización.

Solución de integración de InfoSphere Information Server

Los componentes de InfoSphere Information Server e InfoSphere Foundation Tools están específicamente diseñados para ayudar a las organizaciones a abordar los retos de la integración de datos y a crear una robusta arquitectura de información que aproveche las inversiones en TI existentes. La solución ofrece un enfoque comprobado para identificar información esencial; especificar cómo, cuándo y dónde debe hacerse disponible; determinar los procesos de gestión de datos y las prácticas de gobierno; y alinear el uso de la información para adaptarla a la estrategia empresarial de una organización.

Los componentes de InfoSphere Foundation Tools ayudan a las organizaciones a crear perfiles de información, así como a modelar, definir, supervisar y gobernar la información. Mediante la integración de las soluciones proporcionadas por los componentes de InfoSphere Foundation Tools, las organizaciones pueden descubrir y diseñar su infraestructura de información y empezar a crear información de confianza en la organización.

La plataforma IBM InfoSphere Information Server consta de varios módulos de productos que se pueden desplegar juntos o individualmente en la infraestructura de integración de la empresa, como se muestra en la Figura 1. InfoSphere Information Server está diseñado para integrarse de manera flexible con procesos de integración de datos organizativos existentes con objeto de abordar el ciclo continuo de descubrimiento, diseño y gobierno en soporte de los proyectos de empresa.

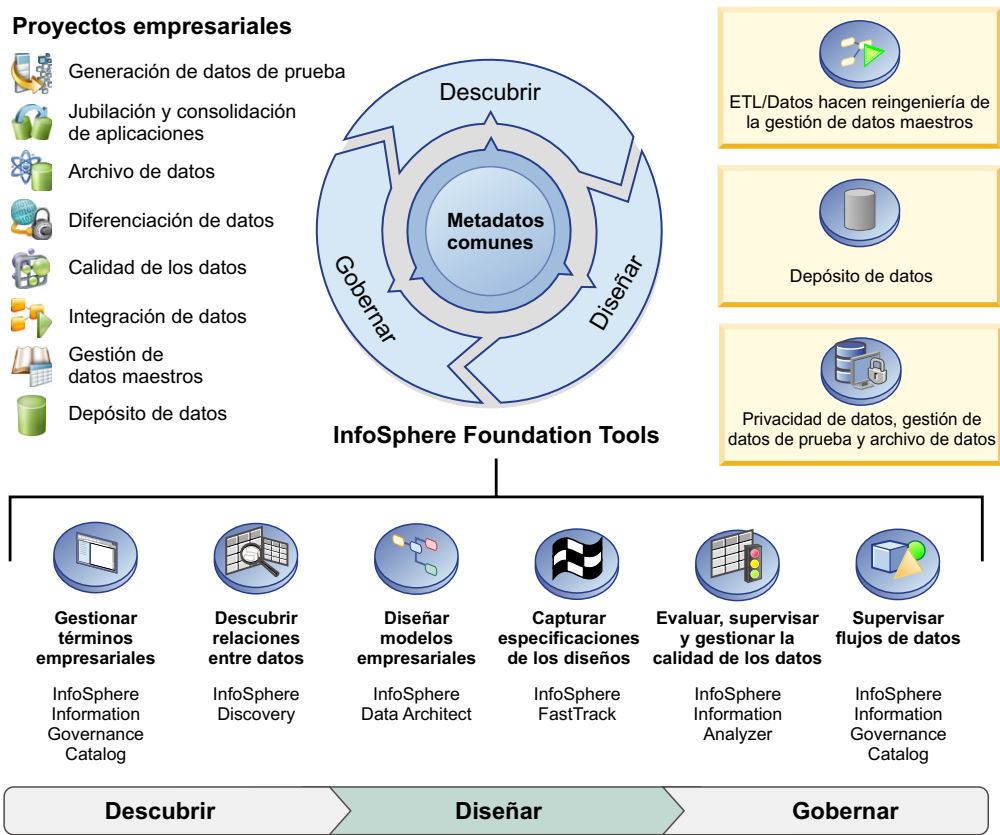


Figura 1. La plataforma InfoSphere Information Server soporta los procesos de integración de datos.

En la Figura 2 se ilustran los componentes y los metadatos que se generan, consumen y comparten.

Normalmente, el proceso empieza con la definición de modelos de datos. Una organización puede importar información de IBM Industry Data Models (disponibles en InfoSphere Data Architect), que incluye un glosario y un modelo de datos físicos y lógicos. Los modelos de glosario contienen miles de términos de estándar de la industria que se pueden utilizar para rellenar previamente InfoSphere Information Governance Catalog. Las organizaciones pueden modificar y ampliar los IBM Industry Data Models para adaptarlos a sus requisitos empresariales específicos.

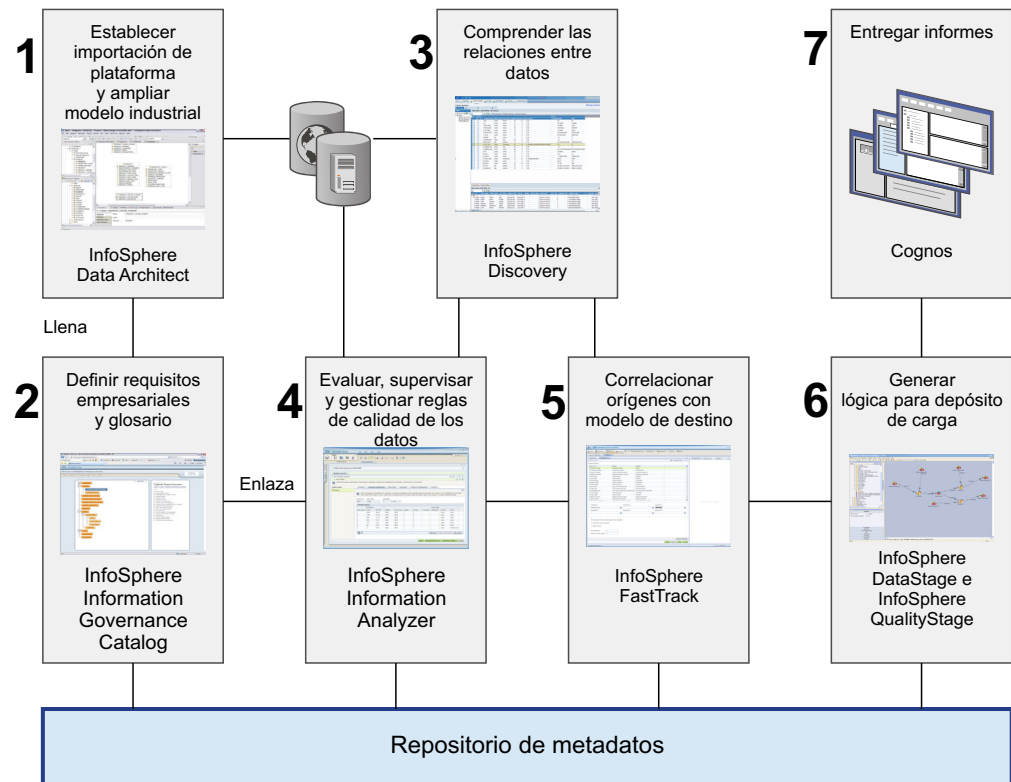


Figura 2. Módulos del producto InfoSphere Information Server

Una vez definidos los modelos de datos y aplicado el contexto empresarial, el analista ejecuta un proceso de descubrimiento de datos en los sistemas de origen que se utilizarán para rellenar el nuevo modelo de datos de destino. Durante el proceso de descubrimiento, el analista puede identificar relaciones clave, reglas de transformación y objetos empresariales que pueden mejorar el modelo de datos, si IBM Industry Data Models no ha definido previamente estos objetos empresariales.

A partir de la información descubierta, el analista puede ampliar el trabajo para centrarse en la valoración de la calidad de los datos y asegurarse de que se documenten las anomalías, se creen las tablas de referencia y se definan las reglas de calidad de los datos. El analista pueda enlazar el contenido de los datos con los términos de glosario establecidos para garantizar que el contexto y el linaje de los datos sean apropiados, entregar resultados analíticos y modelos inferidos a los desarrolladores y probar y desplegar las reglas de calidad de los datos. Cuando las

reglas de calidad de datos se aplican a los datos de los sistemas de origen, el seguimiento de las excepciones a las reglas puede realizarse en IBM InfoSphere Data Quality Console.

Ahora el analista estará preparado para crear las especificaciones de correlación, que son la entrada a los trabajos de ETL para la nueva aplicación. Mediante los resultados del contexto empresarial, de la información descubierta y de la valoración de la calidad de los datos, el analista define las reglas de transformación específicas que son necesarias para convertir los orígenes de datos al formato correcto para el destino de IBM Industry Data Model. Durante este proceso, el analista no sólo define las reglas de transformación empresarial específicas, sino que también puede definir la relación directa entre los términos empresariales y su representación en las estructuras físicas. Después, estas relaciones se pueden publicar en InfoSphere Information Governance Catalog para su consumo y para permitir una mejor comprensión de las relaciones de los activos.

La especificación empresarial sirve ahora como documentación histórica, así como entrada directa para la generación de trabajos de extracción, transformación y carga de IBM InfoSphere DataStage. Las reglas empresariales definidas se incluyen directamente en el trabajo de extracción, transformación y carga como tareas de código o tareas anotadas pendientes que el desarrollador debe completar. Cuando el trabajo de InfoSphere DataStage está preparado, el desarrollador también puede decidir desplegar el mismo proceso por lotes como un componente SOA utilizando IBM InfoSphere Information Services Director.

A lo largo de este proceso, se generan y mantienen metadatos como una consecuencia natural de utilizar cada uno de los módulos de InfoSphere Information Server. La plataforma InfoSphere Information Server comparte los metadatos relevantes con cada uno de los roles específicos de usuario a lo largo de todo el proceso de integración. Debido a esta exclusiva arquitectura, la gestión de metadatos requiere poco mantenimiento manual. Sólo los metadatos de terceros requieren tareas de administración como por ejemplo definir las relaciones con los objetos de metadatos de InfoSphere Information Server. Los administradores y desarrolladores que necesitan ver los activos de metadatos de InfoSphere Information Server y de terceros pueden utilizar InfoSphere Information Governance Catalog para realizar consultas, analizar y notificar esta información desde el repositorio común.

Creación de un caso de ejemplo de aplicación de integración de datos

IBM InfoSphere Information Server ofrece una suite unificada de módulos de productos que están diseñados para optimizar el proceso de creación de una aplicación de integración de datos.

La plataforma InfoSphere Information Server ofrece una arquitectura completa e integrada, creada en un solo repositorio de metadatos compartidos que permite compartir la información perfectamente entre las tareas de integración de datos de proyecto. Puede utilizar reglas de validación de la información, de acceso y de proceso empresarial en varios proyectos, lo que supone un mayor grado de coherencia, mayor control sobre los datos y una mejora de la eficacia. En la Figura 3 en la página 5 se ilustran las funciones: comprender, limpiar, transformar, entregar y realizar una gestión de metadatos unificada.

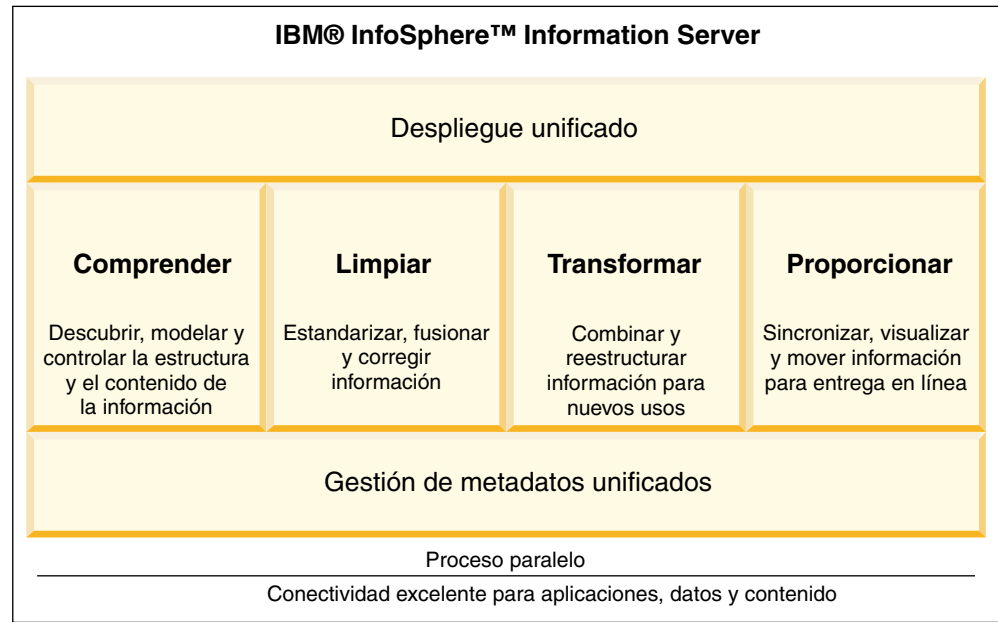


Figura 3. Funciones de integración de InfoSphere Information Server

InfoSphere Information Server permite realizar cinco funciones de integración clave:

- *Comprender los datos.* InfoSphere Information Server ayuda a descubrir, modelar, definir y gobernar automáticamente el contenido y la estructura de la información, así como a comprender y analizar el significado, las relaciones y el linaje de la información. Con estas funciones podrá comprender mejor los orígenes y las relaciones de los datos y definir las reglas empresariales que eliminan el riesgo de utilizar o proliferar datos erróneos.
- *Limpiar los datos.* InfoSphere Information Server da soporte a la calidad y la coherencia de la información mediante la estandarización, la validación, el establecimiento de coincidencias y la fusión de los datos. La plataforma puede ayudarle a crear una única vista completa y precisa de la información estableciendo coincidencias de registros entre orígenes de datos o dentro de cada origen de datos.
- *Transformar los datos en información.* InfoSphere Information Server transforma y enriquece la información para ayudar a garantizar que se encuentra en el contexto adecuado para nuevos usos. También proporciona funciones de transformación y movimiento de grandes volúmenes de datos complejos que puede utilizarse para casos de ejemplo ETL (extraer, transformar y cargar) autónomos o como motor de proceso de datos en tiempo real para aplicaciones o procesos.
- *Proporcionar la información adecuada en el momento adecuado.* InfoSphere Information Server permite virtualizar, sincronizar o mover información para las personas, procesos o aplicaciones que la necesitan. También soporta las arquitecturas críticas orientadas a servicios (SOA), ya que permite el despliegue de reglas de transformación y su reutilización como servicios en varias aplicaciones empresariales.
- *Realizar una gestión de metadatos unificada.* InfoSphere Information Server está creado sobre una infraestructura de metadatos unificados que permite el uso compartido entre los distintos roles de usuario implicados en un proyecto de integración de datos, incluidos los dominios empresariales, operativos y técnicos. Esta infraestructura común gestionada ayuda a reducir el tiempo de desarrollo y

proporciona un registro permanente que puede mejorar la confianza en la información al tiempo que ayuda a eliminar los esfuerzos de coordinación manual.

Modernización de un depósito de datos y descripción de un caso de ejemplo sobre la calidad de datos

En este caso de ejemplo se describen métodos para aprovechar los componentes de IBM InfoSphere Information Server con el fin de abordar la calidad de los datos en un entorno de depósito de datos.

Existen tres casos de uso típicos en los que la calidad de los datos se valora o supervisa en asociación con los depósitos de datos.

Inicio desde cero

Crea un nuevo depósito con actividades que incluyen descubrimientos, terminología, linaje, calidad de datos, modelado de datos, correlación, transformación de datos y limpieza.

Modernización

Modifica o añade a un depósito existente actividades que incluyen terminología, descubrimientos, análisis de impacto, calidad de datos, modelado de datos, correlación, transformación de datos y limpieza.

Gobierno

Gestiona y administra el depósito existente con actividades que incluyen calidad de datos, representación, terminología, linaje y análisis de impacto.

En cada caso de uso hay un rango de actividades que contribuyen a la solución global. Cada actividad utiliza uno o más módulos de productos en el contexto de una metodología y un proceso más amplios que componen la iniciativa. Para cada actividad o fase existen determinadas entradas en el proceso, determinadas tareas que realizar dentro y fuera de un producto y determinadas salidas del proceso que se utilizan en actividades o fases posteriores. La calidad de datos es sólo una de esas actividades.

Las actividades con estos casos de uso no son necesariamente una secuencia rígida de sucesos. A menudo se trata de actividades repetitivas que con frecuencia ocurren en paralelo, donde los descubrimientos de una actividad influirán en otra y, por lo tanto, requieren trabajo adicional.

Por ejemplo, un depósito de datos contiene datos de clientes y de cuentas. Sin embargo, es necesaria una vista más grande del impacto de las ventas y el marketing en los clientes y en sus hábitos de compra. Se selecciona un grupo de orígenes de gestión de ventas para añadirlo al depósito de datos existente. El trabajo de descubrimiento inicial busca y correlaciona cuatro sistemas de gestión de ventas para su inclusión. No obstante, la revisión de la calidad de datos encuentra temas importantes al validar los dominios de los cuatro sistemas, e indica que muchos campos están sin rellenar o contienen comentarios, en vez de datos utilizables. Una revisión de la terminología empresarial descubre que existe confusión respecto al uso del sistema y que se necesitan dos sistemas más. El proceso de descubrimiento introduce los términos empresariales para correlacionar las relaciones con los cuatro sistemas anteriores. A continuación, la revisión de la calidad de datos valida que estas son las tablas necesarias. Después se proporcionan deducciones de la revisión de la calidad de datos a los arquitectos de datos con el fin de mejorar el modelado de las nuevas tablas de depósito de datos.

Existen varios *puntos débiles* comunes que son inherentes a estos casos de uso. Son los siguientes:

- *Terminología poco clara* al gestionar la información del depósito. Por ejemplo, saber dónde está la información de los ingresos o si el contenido del depósito coincide con las expectativas de la empresa.
- *Impacto del cambio desconocido* que puede romper procesos existentes y trastornar el negocio.
- *Linaje de la información desconocida* que repercute negativamente sobre la confianza en los datos. Por ejemplo, ¿estoy utilizando los orígenes de datos correctos para la información solicitada?
- *La calidad de datos desconocida* es una de las razones principales por las que los usuarios de una empresa no confían en sus datos.
- *Representación desconocida*, en la que no está claro quién comprende los datos, quién garantiza el mantenimiento de la calidad y quién gobierna el acceso a los datos.

En este caso de ejemplo, un depósito de datos, que puede ser un depósito de IBM o de cualquier otro proveedor de depósito establecido, como Teradata o Oracle, se amplía o *moderniza*. Este depósito concreto ya contiene diversa información financiera que permite una eficaz creación de informes, pero ahora es necesario añadir datos de clientes para proporcionar información analítica más amplia. Como ocurre con la mayoría de organizaciones, el depósito pasa a ser un lugar importante que es necesario mantener y hay que gestionar la combinación de información financiera, de clientes y de ventas para realizar el análisis.

Integración de la supervisión y la calidad de los datos

Una estrategia de valoración y supervisión de la calidad de los datos aborda los temas relacionados con la calidad y la integridad de la información. Además, se deben establecer procedimientos de calidad de datos para abordar estos temas en un almacén de datos.

IBM InfoSphere Information Governance Catalog, IBM InfoSphere Discovery, IBM InfoSphere Data Architect, IBM InfoSphere Information Analyzer, IBM InfoSphere FastTrack e IBM InfoSphere Data Quality Console utilizan activos de información existentes para suministrarlos a un depósito de datos a través de la integración de información según varios requisitos de inteligencia empresarial basados potencialmente en un modelo o estándar del sector.

Caso de uso del depósito de datos

InfoSphere Information Analyzer puede identificar problemas relacionados con la calidad y la integridad de la información, así como la creación de procedimientos o reglas de calidad de los datos en un entorno multiusuario para supervisar la calidad de los datos a lo largo del tiempo. Este caso de ejemplo puede existir de forma autónoma o ser parte de una iniciativa más amplia, como el depósito de datos, que incorpora la calidad de datos.

Productos clave

- InfoSphere Information Analyzer
- InfoSphere Information Governance Catalog
- InfoSphere Data Architect
- InfoSphere Data Quality Console
- InfoSphere FastTrack
- InfoSphere Discovery

Flujo de trabajo

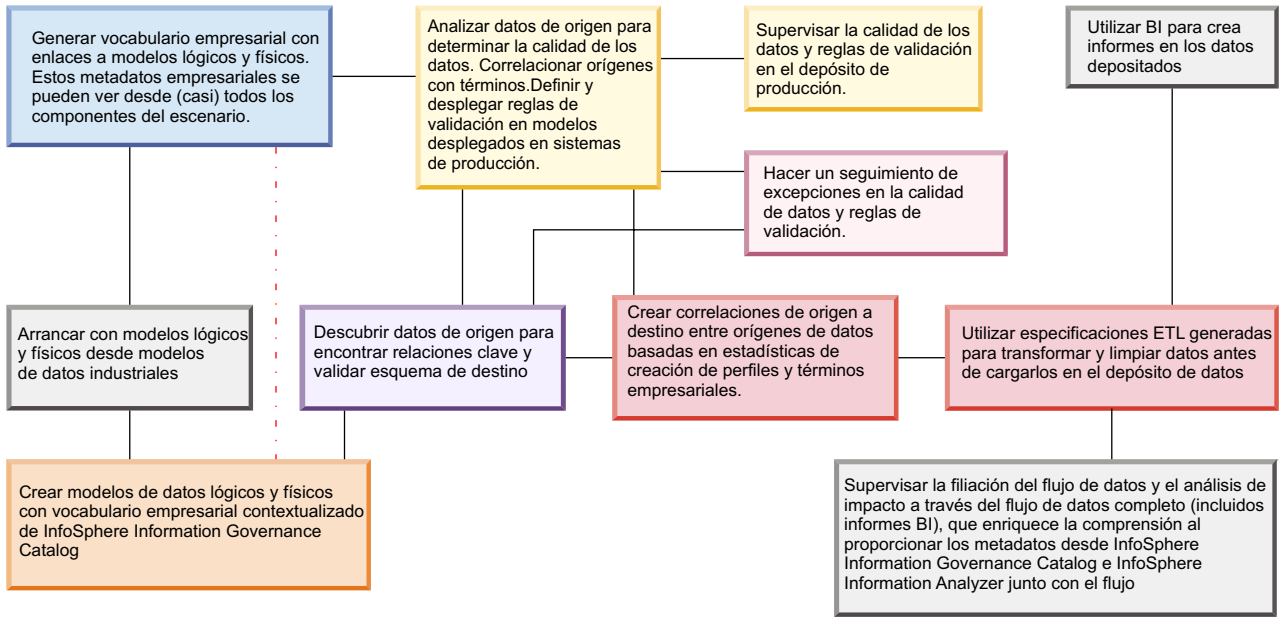


Figura 4. Flujo de trabajo de integración para el depósito de datos

IBM Industry Models

Para ayudar a las organizaciones a lograr resultados más rápidamente, IBM ha empaquetado los conocimientos adquiridos tras años de experiencia trabajando en proyectos de información de sectores específicos en IBM Industry Models.

Estos modelos proporcionan un modelo de datos de empresa completo totalmente atribuido junto con plantillas para la creación de informes que resumen los indicadores clave de rendimiento, las métricas y las cuestiones de conformidad de cada industria. IBM proporciona estos modelos para seis industrias: banca, mercados financieros, asistencia sanitaria, seguros, minoristas y distribución, así como telecomunicaciones.

Estos modelos actúan como aceleradores clave para la migración y la integración de proyectos, y proporcionan una plantilla de eficacia probada específica de la industria que puede ser consultada por todos los participantes de proyectos. Los datos del sistema de origen se pueden cargar directamente en InfoSphere Information Server y proporcionan estructuras de datos de destino y glosarios de empresa predefinidos para acelerar el proceso de desarrollo. Además, las Business Solution Templates (plantillas de soluciones de negocio) proporcionan plantillas para cubos de informes y datos de Cognos 8 Business Intelligence. Gracias a IBM Industry Models, las organizaciones pueden acelerar drásticamente sus proyectos y

reducir los riesgos, así como superar los problemas de organización tradicionales a los que se enfrentan generalmente al integrar información, ya que proporciona un modelo base neutral de eficacia probada.

Preguntas comunes que se deben abordar

Como ocurre con cualquier iniciativa, existen varias preguntas comunes que se pueden formular en relación a la calidad de los datos de una iniciativa del depósito de datos.

Entre estas preguntas se encuentran las siguientes:

Definición de empresa y metadatos frente a la realidad de los datos

La manera de entender las cosas de un usuario de empresa, un experto en el tema y los recursos de TI (arquitectos, modeladores y desarrolladores) puede ser considerablemente diferente. También puede haber diferencias significativas entre distintos grupos empresariales respecto a cómo entienden la terminología empresarial común.

- ¿Se comprende claramente la terminología empresarial?
- ¿Se comprende claramente la relación de la terminología empresarial con la calidad de los datos de origen reales que deben validarse o certificarse?
- ¿Existen vacíos entre la terminología empresarial y los metadatos (el modelo de metadatos de origen o el modelo del depósito de destino)?
- ¿Se utilizan datos y metadatos reales para descubrir y verificar la semántica empresarial y la calidad de los datos?

Impacto del cambio

Al modernizar un sistema existente, es importante comprender la implicación de los diversos cambios planificados.

- ¿Cuáles son los sistemas en sentido ascendente, es decir, los que se suministran al sistema que se va a cambiar, y los sistemas en sentido descendente, es decir, las aplicaciones o los sistemas que consumen la información, en los que repercute el cambio?
- ¿Quiénes son los representantes?
- ¿Existen procesos o aplicaciones, como los informes de inteligencia empresarial, que se puedan interrumpir y deban modificarse?

Énfasis sobre los datos

La información añadida a un depósito de datos generalmente proviene de varios orígenes de datos divergentes o de orígenes de datos divididos entre muchas tablas.

- ¿Se centra la atención en los sistemas o tablas principales, en tipos o clases específicas de datos o en atributos o dominios específicos?
- ¿Hay algún sistema, origen o entidad que se considere el “origen del registro”?
- ¿Cómo se abordarán o resolverán los conflictos o la coherencia entre orígenes?

Validación y entrega de información

Con frecuencia, las iniciativas dejan poco tiempo para las consideraciones iniciales sobre la calidad de los datos. No obstante, cuando se entregan datos y no se logra la calidad de los mismos, el coste de la corrección y de la recuperación de la confianza es alto.

- ¿Cuándo se analiza la calidad de los datos? ¿Se aborda únicamente después del descubrimiento y la revisión inicial, o a lo largo del ciclo de vida de la integración de datos?
- ¿Qué métricas son críticas para validar la calidad de los datos?
- ¿Cómo se entregarán los resultados de la información para la calidad de los datos?

Entradas de caso de ejemplo clave

Cuando el depósito de datos es nuevo o se modifica para añadir nuevos objetos empresariales (como datos de clientes) u orígenes de datos (como información de cliente del sistema de pedidos de venta) existen requisitos para ampliar el glosario con el fin de incorporar nueva terminología, el modelo para incluir todas las nuevas entidades y atributos, y los metadatos para incluir los nuevos orígenes con información analítica asociada (tanto de relaciones como de calidad de datos).

Uno de los principios básicos de la integración de datos es la capacidad de rellenar y almacenar información sobre cada proceso como metadatos almacenados.

Modelos

Para ayudarle a iniciarse en este proceso, IBM ofrece lo que se conoce como modelos de datos *tercera forma normal* (3NF) para integrar grandes conjuntos de datos de detalles así como modelos analíticos predefinidos llamados TDW Reports (informes TDW) o Business Solution Templates (plantillas de soluciones de negocio) (BST) que constan de medidas y dimensiones para resúmenes y creación de informes. IBM también desarrolla modelos de metadatos que contienen diccionarios de términos empresariales que se utilizan para establecer glosarios y abarcar grandes proyectos. Todos los modelos de IBM se correlacionan entre sí para el diseño de la aceleración y el linaje. Los modelos se entregan en InfoSphere Data Architect y en CA ERwin Data Modeler.

A través de la tecnología MetaBroker existente, los modelos físicos de InfoSphere Data Architect se pueden cargar como contenido de metadatos en el repositorio de metadatos de IBM InfoSphere Information Server.

Terminología

Los términos empresariales sirven de clave para describir los tipos de datos con los que trabaja y en el idioma que sea comprensible para su empresa. Este tipo de definición de terminología puede incluir no sólo términos sobre los sistemas de destino, sino también información clave sobre los factores que impulsan el negocio, como los indicadores clave de rendimiento (KPI) o los beneficios esperados. Un ejemplo de este depósito sería el historial de la rentabilidad o de las adquisiciones. Comprender lo que eso significa fomenta la colaboración, por lo tanto, recuerde que todo está relacionado en este lenguaje común.

Una vez importada la información, se comparte con la empresa a través de la colaboración con el fin de aprovecharla. Puede gestionar metadatos para capturar términos empresariales estándar del sector y descripciones que reflejen el lenguaje de los usuarios de empresas. Las organizaciones que institucionalizan la gestión de datos formal o el gobierno de datos pueden publicar estos términos como una forma de garantizar que todos los usuarios de la empresa comprendan claramente la información disponible en la empresa basada en definiciones empresariales estándar.

IBM InfoSphere Governance Catalog proporciona la base para crear semántica basada en las empresas, que incluye categorías y términos.

Si utiliza otras herramientas para importar activos al glosario empresarial, puede utilizar InfoSphere Governance Catalog para asignar un activo a un término. Generalmente, se utiliza InfoSphere Governance Catalog para asignar grandes cantidades de archivos a términos. Puesto que el contenido del glosario se almacena en el repositorio de metadatos de InfoSphere Information Server, puede interactuar con contenido del glosario con los demás componentes de la suite de InfoSphere Information Server.

Procesos de entrada de caso de ejemplo clave

Los temas siguientes describen los procesos de entrada de caso de ejemplo clave.

Descubrimiento de metadatos de origen

Para este caso de ejemplo ya conoce los requisitos de la empresa y la terminología utilizada para describir dichos requisitos. Ahora comprende términos como cliente o ubicación, y puede utilizarlos para referirse a toda la organización. También conoce la estructura del depósito de datos de destino. Debe aumentar o mejorar sus conocimientos sobre la estructura y el contenido de los datos entrantes reales.

IBM InfoSphere Discovery

IBM InfoSphere Discovery se utiliza para identificar las reglas de transformación que se han aplicado a un sistema de origen para rellenar un destino, como un depósito de datos o un almacén de datos operativo. Una vez definidos exactamente, estos objetos empresariales y reglas de transformación proporcionan la entrada esencial a proyectos basados en la información, como la integración de datos, IBM InfoSphere Master Data Management (MDM) y el archivado.

InfoSphere Discovery analiza los valores y patrones de datos de uno o más orígenes para capturar estas correlaciones ocultas y ponerlas claramente a la vista. InfoSphere Discovery aplica algoritmos heurísticos y sofisticados para realizar una completa gama de técnicas de análisis de datos: solapamiento de datos y análisis de relaciones de un solo origen y entre varios orígenes, descubrimiento avanzado de coincidencias clave, transformación de descubrimiento lógico, etc. Posee el rango más amplio de orígenes de datos empresariales: bases de datos relacionales, bases de datos jerárquicas y cualquier origen de datos estructurado que se represente en formato de archivo de texto.

Uso de InfoSphere Discovery para comprender las relaciones de datos a través de la búsqueda de valores y patrones de datos

En este caso de ejemplo combinará tres orígenes de datos distribuidos para descubrir información relacionada para los nombres, las direcciones y los identificadores fiscales de los clientes.

Utilice InfoSphere Discovery para:

1. Crear un proyecto de descubrimiento de datos de origen. En los proyectos de descubrimiento de datos de origen se revisan los solapamientos de los valores de los datos entre las tablas de uno o más conjuntos de datos. Además, se puede crear un esquema unificado y descubrir claves foráneas primarias de esquema unificado. Los resultados incluyen un resumen del número total de tablas y columnas en los conjuntos de datos, número de columnas exclusivas,

porcentaje de solapamiento de valores, número de tablas y columnas que contienen datos solapados y más estadísticas detalladas, incluidas vistas de los propios datos solapados reales.

2. Importar un conjunto de tablas de cada origen de datos.
3. Crear conjuntos de datos. Un conjunto de datos es una colección de tablas de base de datos y archivos de texto que se deben procesar, analizar o correlacionar. Puede contener tantas tablas de base de datos y archivos de texto delimitados o posicionales como sea necesario, de tantas conexiones ODBC como se desee.
4. Ejecutar el descubrimiento de datos. Algunas tablas de base de datos contienen definiciones de tipos de datos correctos en metadatos. No obstante, otras tablas de base de datos contienen metadatos incorrectos o incompletos, y los archivos de texto no definen los tipos de datos.

En este paso, InfoSphere Discovery calcula y muestra estadísticas de los datos en los conjuntos de datos, y también muestra información de los metadatos. InfoSphere Discovery también examina todas las series VARCHAR para determinar si contienen valores de fecha/hora o valores numéricos y los cambia por el tipo de datos apropiado.

Después de ejecutar el descubrimiento de datos, debe revisar los datos de análisis de columnas. Verifique los tipos de datos y realice las modificaciones que sean necesarias, como cambiar la longitud de una columna o corregir un tipo de dato erróneo definido en los metadatos. Marque las columnas que sean importantes para su proyecto, como los elementos de datos críticos (CDE). Siempre que modifique las tablas de un conjunto de datos debe volver a ejecutar el análisis de columnas.

5. Descubrir claves primarias foráneas (PF). InfoSphere Discovery importa automáticamente relaciones clave PF cuando están definidas en los metadatos de una tabla. Cuando las relaciones no están definidas, InfoSphere Discovery busca coincidencias de columnas examinando los datos reales. Las coincidencias de columnas con la selectividad y los índices de coincidencia más altos se designan automáticamente como claves PF.

Después de realizar la tarea Descubrir claves PF, verifique la exactitud de los resultados. Si define o modifica las coincidencias de columnas descubiertas o claves PF, ejecute de nuevo la tarea Descubrir claves PF.

6. Descubrir objetos de datos. Un objeto de datos es una manera conceptual de considerar las relaciones de tablas en un conjunto de tablas. Un objeto de datos representa un grupo de tablas relacionadas mediante claves PF. Un conjunto de datos contiene muchos objetos de datos, y cada objeto de datos consta de muchas tablas o sólo de una tabla. Una tabla puede ser tanto un padre como un hijo, por lo tanto, la misma tabla puede aparecer en varios objetos de datos. Generalmente, estos objetos de datos se relacionan con conceptos clave del glosario o con las nuevas entidades que se añaden o modernizan en el depósito de datos.
7. Descubrir solapamientos. El paso relativo a los solapamientos proporciona una imagen clara de los datos que se solapan en los orígenes. Revise los datos de columnas para verificar que los solapamientos descubiertos son útiles y válidos. Suprima los desajustes obvios en Detalles de solapamiento de valor. Si existe alguna duda sobre los datos en un solapamiento determinado, utilice Resumen de columnas y Solapamientos de columnas para mostrar los datos reales. Defina nuevos solapamientos y utilice InfoSphere Discovery para determinar las estadísticas.

Marque las columnas que sean importantes para su proyecto, como los elementos de datos críticos (CDE).

Una vez completada la tarea de descubrimiento de solapamientos, puede empezar a definir un esquema unificado. Aplique este enfoque cuando las estructuras del modelo de datos de origen actual se deban y puedan utilizar de nuevo, cuando sea necesario crear prototipos rápidamente o cuando no haya requisitos específicos relacionados con el modelo de datos de destino que requiera un diseño muy exclusivo. Incluso en el último caso, puede resultar ventajoso crear un esquema unificado como punto intermedio para los datos estandarizados y alineados de los orígenes de datos antes de la consolidación y carga en el depósito de datos.

8. Crear una tabla de destino para un esquema unificado. Defina el esquema de la tabla de destino. Rellene la tabla con tantas columnas de destino como sea necesario. Puede modificar los tipos de datos y cambiar el orden de la secuencia de columnas.
9. Definir una correlación de origen. Correlacione las columnas de la tabla con las columnas de destino y cree filtros para cada columna de la tabla hasta que cada columna de destino contenga los datos apropiados de las tablas relevantes. Si es necesario, vuelva al separador Esquema de la tabla de destino para definir de nuevo las características de las columnas de tabla.
10. Realizar un análisis de esquema unificado. Consolide los datos de distintos orígenes en la tabla de destino agregada. Las vistas previas de estadísticas y datos de esta pantalla se utilizan para definir condiciones coincidentes para su uso en el siguiente paso, el análisis de coincidencias y fusión.
Revise las estadísticas resultantes. Las estadísticas se muestran para las columnas agregadas en la tabla de destino, así como para cada columna de tabla individual utilizada en la tabla de destino.
Si es necesario, refine la tabla de destino volviendo a la ventana Esquema de la tabla de destino y añadiendo o eliminando columnas de la tabla de destino. También puede cambiar la correlación de origen.
11. Realizar un análisis de coincidencias y fusión. Defina las condiciones coincidentes para cada columna de destino. Defina también las reglas de detección de conflictos que determinan si los valores de cada grupo se consideran o no conflictos. Fusione los resultados y defina reglas de resolución de conflictos que seleccionen el mejor valor de las alternativas de registros duplicados. El esquema unificado está terminado.
12. Exportar datos de InfoSphere Discovery al repositorio de metadatos. Ahora puede exportar los resultados del esquema unificado al repositorio de metadatos utilizando el Gestor de importación/exportación en DBM Metabroker para poner los resultados a disposición de InfoSphere Information Analyzer.

Valoración de la calidad de los datos

Comprender la relación entre las tablas y columnas de datos no es suficiente para garantizar su eficaz utilización en el depósito de datos. Debe valorar los datos para entregarlos en sentido descendente de forma eficaz.

Durante esta valoración, analizará detalladamente el contenido para revelar todas las anomalías de la información. Esta valoración de la calidad permite anotar los errores de los datos para llevar a cabo la acción pertinente en sentido descendente. Estas anotaciones se comparten a nivel de metadatos. Al encontrar y anotar estos problemas de los datos puede incorporar supervisores de datos para seguir buscando estos patrones en los datos. Este proceso garantiza que los datos de alta calidad sigan pasando de los sistemas de origen 'tal cual' al sistema que 'será' de destino.

Aunque el descubrimiento de datos proporciona un punto de partida, para garantizar que los criterios de calidad de datos se cumplan y se aborden tanto en la carga inicial del depósito de datos con esta nueva información como en la administración continua del depósito de datos, es aconsejable llevar a cabo una valoración detallada de la calidad de los datos.

El centro de interés tiene dos aspectos:

- En primer lugar, debe revisar y comunicar la información sobre anomalías, tablas de referencia clave para valores válidos, alineación de orígenes con terminología empresarial y requisitos para condiciones de validación de datos.
- En segundo lugar, debe diseñar y probar las reglas de validación de datos, los valores de referencia y las medidas que se aplicarán al proceso de carga del depósito de datos y al propio depósito de datos una vez cargado.

Análisis de la calidad de los datos de origen

Este es el paso inicial para valorar, revisar y comunicar importante información de la calidad de datos acerca de los orígenes para suministrarla al depósito de datos.

Uso de IBM InfoSphere Information Analyzer para ejecutar la calidad de datos y definir datos sensibles

El esquema del análisis de InfoSphere Discovery lo puede consumir InfoSphere Information Analyzer para proporcionar un despliegue acelerado y una mayor exactitud de las reglas de calidad de datos para el análisis, supervisión y gestión de la calidad a lo largo del tiempo.

Para este caso de ejemplo, los resultados del proceso de descubrimiento proporcionan la información de origen de datos inicial con metadatos acerca de los nombres, las direcciones y los identificadores fiscales de los clientes.

Utilice InfoSphere Information Analyzer para:

1. Crear o abrir un proyecto existente de InfoSphere Information Analyzer.
2. Añadir los orígenes de datos para los que InfoSphere Discovery ha realizado un análisis en el proyecto de Information Analyzer.
3. Importar el esquema de InfoSphere Discovery del caso de ejemplo anterior, Descubrimiento de metadatos de origen, del repositorio de metadatos a InfoSphere Information Analyzer como un destino para las reglas de calidad de datos.
4. Revisar los resultados del análisis de resumen de InfoSphere Discovery bajo la tarea Análisis de columnas. Ignore los campos que sean irrelevantes para el uso en el depósito de datos. Para los campos en los que la calidad de los datos sea esencial para la evaluación, ejecute el análisis de columnas. Si la validación de los datos confidenciales es importante, compruebe la opción de clasificación de datos mejorada cuando se ejecute.
5. Verificar las clases de datos para una clasificación de datos precisa. Al visualizar y verificar los resultados del análisis de columnas, puede aceptar o sobrescribir la clase de datos inferida por el sistema para la columna seleccionada. Estas clases de datos son componentes clave para realizar un análisis posterior.
6. Analizar los resultados del análisis de detalles según clasificaciones de datos y prestando atención a condiciones específicas y anotaciones apropiadas de los campos.
 - **Identificadores** - Busca duplicados y formatos no válidos.

- **Indicadores** - Busca valores de distintivo o desvíos de datos no válidos e incoherentes; genera tablas de referencia de valores válidos.
 - **Códigos** - Busca valores no válidos y de código predeterminado; genera tablas de referencia de valores válidos.
 - **Cantidades** - Busca datos fuera de los rangos válidos; genera tablas de referencia de rangos válidos.
 - **Fechas** - Busca datos fuera de los rangos válidos; genera tablas de referencia de rangos válidos.
 - **Texto** - Identifica campos de texto que son necesarios para el depósito de datos; busca formatos problemáticos; utiliza reglas de datos para explorar condiciones conocidas.
 - **Datos confidenciales** - Busca condiciones no válidas.
 - **Otras condiciones** - Utiliza reglas de datos para ampliar el análisis a otras condiciones definidas para empresas como las combinaciones de valores válidos, el orden de los datos, las expresiones de cálculo, el formato complejo de datos, etc.
7. Durante el análisis, asignar columnas a los términos empresariales identificados como se establece en IBM InfoSphere Governance Catalog. Esta vista centrada en los datos ayuda a validar y garantizar que los datos correctos se enlazarán con los conceptos empresariales correctos.
 8. Notificar y revisar los resultados del análisis de detalles. Este puede ser un ciclo de revisión repetitivo con expertos en el tema. Entre las condiciones típicas que se encuentran en este paso se incluyen las siguientes:
 - Espacios en los datos (por ejemplo, los datos encontrados no se alinean con los requisitos del depósito de datos o no existen datos que cumplan un requisito específico. Los espacios pueden requerir abordar orígenes adicionales y un posible ciclo adicional a través de IBM InfoSphere Discovery).
 - Lagunas de conocimientos (por ejemplo, nadie sabe qué significa un código determinado o qué se considera válido)
 - Problemas en los datos (por ejemplo, existen problemas en los datos que se deben abordar y documentar, o deben establecerse reglas)
 9. Publicar los resultados del análisis. Puede ver los resultados de un análisis y publicarlos en el repositorio de metadatos. Esta publicación actualiza el repositorio de metadatos con los detalles adicionales encontrados a través de la valoración de la calidad de los datos, y los desarrolladores que trabajan con IBM InfoSphere DataStage o IBM InfoSphere QualityStage pueden utilizarla directamente para establecer los procesos de carga correctos desde los orígenes de datos identificados al esquema de destino definido por IBM InfoSphere Discovery o el depósito de datos.
 10. Diseñar nuevas reglas y métricas de validación de la calidad de los datos, o reutilizar las ya existentes, en función de los descubrimientos del análisis de datos. No se seleccionarán todos los campos para la validación continua.
 11. Probar y revisar las nuevas reglas y métricas de datos utilizando los datos de origen planificado. Estas validaciones de reglas estarán dirigidas a las necesidades del depósito de datos. Los resultados formarán parte del ciclo de revisión repetitivo con expertos en el tema, y se utilizarán para establecer valores de referencia iniciales. Las validaciones de reglas ampliarán o resolverán los elementos anotados en el paso "Notificar y revisar los resultados del análisis de detalles" descrito anteriormente.

12. Publicar o desplegar las reglas y métricas de datos. Estas reglas y métricas de datos proporcionan la base para la supervisión continua de la calidad de los datos.

InfoSphere Information Analyzer cumple un rol crítico en el proceso de integración. Para facilitar las especificaciones de correlación más completas y precisas, se puede acceder directamente a los resultados de los perfiles de InfoSphere Information Analyzer desde IBM InfoSphere FastTrack, donde se definen y documentan las especificaciones. Estas especificaciones se convierten en los requisitos de entrada para los trabajos de limpieza y de extracción, transformación y carga de InfoSphere DataStage e IBM InfoSphere QualityStage que soportan la aplicación empresarial que se llevará a cabo. Cuanta más información tenga el analista sobre las estructuras y el contenido de datos verdaderos más exactos serán los requisitos para los desarrolladores en sentido descendente.

Información relacionada:

Enfoques y métodos de reglas

Salidas de caso de ejemplo clave

El análisis de relaciones con IBM InfoSphere Discovery y el análisis de la calidad de los datos con IBM InfoSphere Information Analyzer producen varias salidas que facilitan la modernización del depósito de datos, incluida una mejor correlación e integración en las nuevas estructuras de datos, tablas de referencia clave para la validación de la carga y controles de calidad para la supervisión continua de la calidad de los datos.

Las cuatro salidas clave proporcionadas por el proceso de valoración de la calidad de los datos son: metadatos compartidos, incluidos los resultados analíticos y las tablas de referencia; metadatos compartidos, incluidos los enlaces a terminología empresarial; y reglas de validación de datos para la supervisión continua de la calidad.

Uso de metadatos compartidos para comprender el ámbito y el impacto del cambio

IBM InfoSphere Metadata Workbench proporciona análisis de impacto de los sistemas que están en el ámbito de la iniciativa de depósito. Los sistemas en sentido ascendente y en sentido descendente afectados, así como las rutinas que los conectan, son fáciles de identificar. El análisis de impacto ayuda a reducir el riesgo de que falten cambios necesarios y de que se interrumpan los flujos de datos.

La modernización de un depósito de datos puede ocasionar cambios en la estructura del depósito de datos existente, incluidas las tablas actualizadas, los hechos o nuevas relaciones de datos. Si los cambios son adiciones, lo que significa que no existían previamente, el impacto en sentido descendente deberá ser mínimo. Si los cambios modifican la estructura del depósito de datos, el análisis de impacto será crítico tanto a nivel empresarial como de integración.

El linaje de negocio permite a los consumidores en sentido descendente comprender los impactos en las despendas de datos y los informes suministrados desde el depósito de datos, particularmente si cambia la terminología o el contenido de los datos.

El linaje de datos permite a los desarrolladores comprender los impactos y los cambios necesarios en los trabajos de integración de datos que se suministran a las despensas de datos en sentido descendente, a los sistemas o las aplicaciones.

Uso de metadatos compartidos en la correlación y el desarrollo

Comprende perfectamente la información que tiene en el entorno. Conoce la forma, el tamaño y los términos empresariales del origen y los destinos. Comprende perfectamente la valoración y ha descubierto las relaciones de los datos.

Después, debe decidir cómo correlacionar los orígenes de datos nuevos o actualizados con el destino del depósito de datos. Empiece con los conocimientos que acaba de adquirir y utilícelos para ayudar a proporcionar la información correcta a la empresa. El enfoque aquí consiste en tomar los datos de origen y asegurarse de que se ajustan a la forma y el tamaño del destino.

Para facilitar el trabajo de desarrollo más completo y preciso, se puede acceder directamente a los resultados analíticos de InfoSphere Discovery e InfoSphere Information Analyzer desde InfoSphere FastTrack, donde se definen y documentan las especificaciones, así como desde InfoSphere DataStage, InfoSphere QualityStage e InfoSphere Business Catalog.

Las especificaciones de correlación en InfoSphere FastTrack pueden incorporar directamente los orígenes de datos identificados en el depósito de datos como información de origen. El esquema unificado, si se utiliza, o un modelo físico generado a partir de los modelos de depósito de datos, se puede incorporar como información de destino. Si revisa la información analítica, incluidas las anotaciones y los valores identificados, puede identificar las reglas de transformación específicas que se deben utilizar para alinear adecuadamente los datos del origen con los del destino. Se pueden indicar tablas de referencia para los procesos de búsqueda.

Si hay huecos o falta información entre los orígenes y el destino quizá deba realizar las siguientes acciones:

- Validar las diferencias en la terminología asociada con los datos de origen y de destino
- Solicitar orígenes de datos adicionales
- Regresar al ciclo de descubrimiento inicial
- Revisar información adicional relacionada con los requisitos de calidad de los datos
- Regresar al ciclo de valoración de la calidad de los datos

Estas especificaciones se convierten en los requisitos de entrada para los trabajos de ingeniería de datos, limpieza, extracción, transformación y carga de InfoSphere DataStage e InfoSphere QualityStage que soportan las modificaciones del depósito de datos que se llevarán a cabo. Cuanta más información tenga sobre el contenido de datos verdaderos más exactos serán los requisitos para los desarrolladores en sentido descendente.

Uso de metadatos compartidos en el glosario empresarial

Cuando la información y los orígenes nuevos se introduzcan en el depósito de datos, es posible que deba considerar cuestiones sobre cómo utilizar esa información.

Entre las preguntas se pueden incluir:

- ¿Qué ocurre en este entorno si añado un sistema y después tengo que cambiar o entregar información a la empresa lo más rápido posible?
- ¿Y si los ejecutivos desean comprender de dónde provienen todos los datos de riesgos financieros o cuándo se han actualizado por última vez?

El proceso requiere que la notificación y el gobierno de datos gestionen los cambios, entreguen los informes y permitan que todo el mundo comparta la misma información. Mediante los metadatos compartidos comunes donde la terminología empresarial está enlazada a los activos de datos reales (tanto los orígenes de datos como el depósito de datos), la información se puede comunicar rápidamente en lo que respecta a estos tipos de preguntas.

Tanto si utiliza IBM InfoSphere Governance Catalog para comprender el linaje de negocio del origen al depósito como para revisar los detalles del entorno de metadatos, puede encontrar términos de enlace de información compartida a orígenes de datos o información analítica manteniendo los nuevos componentes de depósito y proporcionando información a las personas que trabajan con el nuevo contenido en el depósito.

Supervisión continua de la calidad de los datos

Gracias a las métricas y las reglas de calidad de datos desarrolladas y desplegadas desde la valoración de la calidad de los datos puede revisar la calidad de la información, así como hacer un seguimiento de esta, en el depósito de datos a lo largo del tiempo. La entrega y la disponibilidad de la información son factores clave en las decisiones de esta etapa. Existen distintas opciones de entrega desde las que hacer un seguimiento de la calidad de los datos y supervisarlos.

Uso de IBM InfoSphere Information Analyzer para supervisar la calidad de los datos

En este caso de ejemplo establecerá las actividades de supervisión de la calidad de los datos.

1. Identifique el enfoque preferido para la entrega y supervisión de la calidad de los datos en torno a estas reglas y métricas identificadas. Las opciones incluyen la interfaz de usuario, informes o la publicación de resultados a través de interfaces de línea de mandatos (CLI) o interfaces de programación de aplicaciones (API). La API se puede utilizar para incorporar las soluciones específicas de creación de informes de la organización, incluido el propio depósito de datos.
2. Identifique la planificación de ejecución apropiada para las reglas y métricas identificadas y sus informes asociados. Cuando los procesos se llevan a cabo a través de la interfaz o los informes, estos procesos se pueden planificar a través del planificador de trabajos de IBM InfoSphere Information Analyzer o la consola de informes de IBM InfoSphere Information Server. Si los procesos se llevan a cabo a través de CLI o API, debe ponerlos en un script o planificarlos externamente.
3. Suponiendo que se realice algún nivel de supervisión a través de los informes y la interfaz de usuario de InfoSphere Information Analyzer, debe revisar los resultados resumidos de las métricas y las reglas como medidos contra valores de referencia (umbrales) establecidos.
4. Cuando se observen discrepancias, obtenga una vista detallada de las reglas o métricas específicas. Estos resultados incluyen estadísticas resumidas de cada ejecución de la regla o métrica desde donde puede obtener detalles de excepción.

5. Entregue los informes o exportaciones de datos a expertos en la materia. Surgirán condiciones nuevas e inesperadas en los datos. Estas condiciones deberán ser revisadas y resueltas. Los procesos o datos de origen quizás tengan que modificarse. Es posible que la información del depósito de datos se deba corregir o que las reglas se deban actualizar y modificar posteriormente.
6. Si es necesario, vuelva al proceso de valoración de la calidad de los datos para analizar de nuevo los datos de origen o modificar las reglas y métricas para resolver las condiciones cambiantes.

Uso de IBM InfoSphere Data Quality Console para hacer un seguimiento de la calidad de los datos

Puede utilizar la consola de calidad de datos para hacer un seguimiento de las excepciones que genera InfoSphere Information Analyzer o InfoSphere Discovery. Las excepciones incluyen cualquier entidad que requiere información o investigación adicional, como por ejemplo un registro que no cumplía con una regla de calidad de datos en InfoSphere Information Analyzer o coincidencias ambiguas de InfoSphere Discovery.

Cuando se hace un seguimiento de excepciones en la consola de calidad de datos, puede encontrar conjuntos de información sobre excepciones utilizando criterios de búsqueda. Por ejemplo, suponga que un supervisor es el responsable de la calidad de datos de los recursos de datos implementados desde el departamento de Finanzas. El supervisor puede utilizar los criterios de búsqueda para ver únicamente información sobre excepciones asociadas con bases de datos, tablas u otros recursos del departamento de Finanzas. La información de las excepciones puede supervisarse a lo largo del tiempo y se utiliza para hacer un seguimiento y mejorar la calidad de los datos.

Comprender mejor su negocio a partir de datos de alta calidad

En última instancia, las adiciones al depósito de datos tienen la finalidad de proporcionar un conocimiento más profundo de la propia empresa. La información de cliente que se añade encima de la información financiera en el depósito de datos puede proporcionar mucha información sobre los hábitos de compra o los gastos de los clientes, las ventas de alto valor o que generan muchos ingresos, etc. Sin embargo, esta información sólo es tan buena como la calidad de los datos proporcionados en el depósito de datos. Incluso cuando la calidad de los datos de los orígenes se considera alta, la falta de terminología coherente, modelos bien definidos, problemas con la alineación de los datos, valores incoherentes o relaciones mal correlacionadas pueden generar problemas en sentido descendente. Resaltar el linaje resultante de datos no es suficiente para garantizar una utilización eficaz y la calidad de la información. Si se pone el énfasis adecuado en el descubrimiento de datos, la valoración de la calidad de los datos y la supervisión continua de la calidad, se puede aumentar significativamente la confianza en la información y el análisis resultantes.

Conceptos relacionados:

Caso de ejemplo: seguimiento de excepciones

Información relacionada:

Enfoques y métodos de reglas

Apéndice A. Accesibilidad de los productos

Puede obtener información sobre el estado de accesibilidad de los productos de IBM.

Los módulos de producto y las interfaces de usuario de IBM InfoSphere Information Server no son totalmente accesibles.

Para obtener información sobre el estado de accesibilidad de los productos de IBM, consulte la información de accesibilidad de productos de IBM en http://www.ibm.com/able/product_accessibility/index.html.

Documentación sobre accesibilidad

Se proporciona documentación accesible para los productos en IBM Knowledge Center. IBM Knowledge Center presenta la documentación en formato XHTML 1.0, que se puede ver en la mayoría de navegadores web. Dado que IBM Knowledge Center utiliza XHTML, puede establecer preferencias de visualización en el navegador. Esto también le permite utilizar lectores de pantalla y otras tecnologías de asistencia para acceder a la documentación.

La documentación que está en IBM Knowledge Center se proporciona en archivos PDF, que no son totalmente accesibles.

IBM y la accesibilidad

Consulte el sitio web IBM Human Ability and Accessibility Center para obtener más información sobre el compromiso de IBM con la accesibilidad.

Apéndice B. Cómo ponerse en contacto con IBM

Puede ponerse en contacto con IBM para obtener soporte al cliente, servicios de software, información sobre productos e información general. También puede facilitar comentarios a IBM sobre los productos y la documentación.

En la tabla siguiente se listan los recursos para soporte al cliente, servicios de software, formación e información sobre productos y soluciones.

Tabla 1. Recursos de IBM

Recurso	Descripción y ubicación
Portal de soporte de IBM	Puede personalizar la información de soporte eligiendo los productos y los temas que le interesen en www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server
Servicios de software	Puede encontrar información sobre servicios de software, de tecnologías de la información y de consultoría de negocio en el sitio de soluciones, en www.ibm.com/businesssolutions/
Mi IBM	Puede gestionar enlaces a sitios web de IBM y a información que satisfaga sus necesidades específicas de soporte técnico creando una cuenta en el sitio Mi IBM en www.ibm.com/account/
Formación y certificación	Puede obtener información sobre formación técnica y servicios de educación diseñados para personas, empresas y organizaciones públicas, a fin de adquirir, mantener y optimizar sus habilidades de TI en http://www.ibm.com/training
Representantes de IBM	Puede contactar con un representante de IBM para obtener información sobre soluciones en www.ibm.com/connect/ibm/us/en/

Apéndice C. Acceso a la documentación del producto

La documentación se proporciona en diversos formatos: en el IBM Knowledge Center en línea, en un centro de información opcional instalado localmente y como manuales PDF. Puede acceder a la ayuda en línea o instalada localmente directamente desde las interfaces de cliente del producto.

IBM Knowledge Center es el mejor lugar para encontrar la información más actualizada de InfoSphere Information Server. IBM Knowledge Center contiene ayuda para la mayoría de las interfaces del producto, así como documentación completa para todos los módulos de producto de la suite. Puede abrir IBM Knowledge Center desde el producto instalado o desde un navegador web.

Cómo acceder a IBM Knowledge Center

Existen varias maneras de acceder a la documentación en línea:

- Pulse el enlace **Ayuda** en la parte superior derecha de la interfaz de cliente.
- Pulse la tecla F1. Normalmente, la tecla F1 abre el tema que describe el contexto actual de la interfaz de cliente.

Nota: La tecla F1 no funciona en clientes web.

- Escriba la dirección en un navegador web, por ejemplo, cuando no tenga iniciada una sesión en el producto.

Escriba la siguiente dirección para acceder a todas las versiones de la documentación de InfoSphere Information Server:

<http://www.ibm.com/support/knowledgecenter/SSZJPZ/>

Si desea acceder a un tema concreto, especifique el número de versión con el identificador de producto, el nombre del plug-in de documentación y la vía de acceso al tema en el URL. Por ejemplo, el URL para la versión 11.3 de este tema es el siguiente. (El símbolo \Rightarrow indica una continuación de línea):

http://www.ibm.com/support/knowledgecenter/SSZJPZ_11.3.0/=>com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html

Consejo:

El Knowledge Center tiene también un URL corto:

<http://ibm.biz/knowctr>

Para especificar un URL corto a una página de producto, versión o tema específico, utilice un carácter de almohadilla (#) entre el URL corto y el identificador de producto. Por ejemplo, el URL corto a toda la documentación de InfoSphere Information Server es el siguiente URL:

<http://ibm.biz/knowctr#SSZJPZ/>

Y el URL corto al tema anterior para crear un URL ligeramente más corto es el siguiente URL (El símbolo \Rightarrow indica una continuación de línea):

http://ibm.biz/knowctr#SSZJPZ_11.3.0/com.ibm.swg.im.iis.common.doc/=>common/accessingiidoc.html

Cambiar los enlaces de ayuda para que hagan referencia a la documentación instalada localmente

IBM Knowledge Center contiene la versión más actualizada de la documentación. Sin embargo, puede instalar una versión local de la documentación como un centro de información y configurar los enlaces de ayuda para que apunten a él. Un centro de información local es útil si su empresa no proporciona acceso a Internet.

Siga las instrucciones de instalación que vienen con el paquete de instalación del centro de información para instalarlo en el sistema que elija. Después de instalar e iniciar el centro de información, puede utilizar el mandato **iisAdmin** en el sistema de la capa de servicios para cambiar la ubicación de la documentación a la que hacen referencia la tecla F1 y los enlaces de ayuda del producto. (El símbolo ⇒ indica una continuación de línea):

Windows

```
vía_instalación_IS\ASBServer\bin\iisAdmin.bat -set -key ⇒  
com.ibm.iis.infocenter.url -value http://<host>:<puerto>/help/topic/
```

AIX Linux

```
vía_instalación_IS/ASBServer/bin/iisAdmin.sh -set -key ⇒  
com.ibm.iis.infocenter.url -value http://<host>:<puerto>/help/topic/
```

Donde <host> es el nombre del sistema donde está instalado el centro de información y <puerto> es el número de puerto para el centro de información. El número de puerto predeterminado es 8888. Por ejemplo, en un sistema llamado server1.example.com que utilice el puerto predeterminado, el valor del URL sería <http://server1.example.com:8888/help/topic/>.

Obtener la documentación en PDF y en copia impresa

- Los manuales en archivos PDF están disponibles en línea y puede accederse a ellos desde este documento de soporte: <https://www.ibm.com/support/docview.wss?uid=swg27008803&wv=1>.
- También puede solicitar publicaciones de IBM en formato impreso en línea o a través de su representante local de IBM. Para solicitar publicaciones en línea, vaya al Centro de Publicaciones de IBM en <http://www.ibm.com/e-business/linkweb/publications/servlet/pbi.wss>.

Apéndice D. Cómo aportar comentarios sobre la documentación del producto

Puede aportar valiosos comentarios en relación a la documentación de IBM.

Sus comentarios ayudarán a IBM a ofrecer información de calidad. Puede utilizar cualquiera de los métodos siguientes para enviar sus comentarios:

- Para proporcionar un comentario acerca de un tema del IBM Knowledge Center que está alojado en el sitio web de IBM, inicie la sesión y pulse el botón **Añadir comentario** en la parte inferior del tema. Los comentarios enviados de esta manera serán visibles para todos los usuarios.
- Para enviar un comentario acerca de un tema del IBM Knowledge Center a IBM y que ningún otro usuario pueda ver, inicie la sesión y pulse en el enlace **Comentarios** en la parte inferior del IBM Knowledge Center.
- Envíe sus comentarios utilizando el formulario de comentarios del lector que encontrará en www.ibm.com/software/awdtools/rcf/.
- Envíe sus comentarios por correo electrónico a comments@us.ibm.com. Incluya el nombre y el número de versión del producto, así como el nombre y el número de pieza de la información (si es pertinente). Si su comentario es sobre un texto específico, incluya la ubicación del texto (por ejemplo, un título, un número de tabla o un número de página).

Avisos y marcas registradas

Esta información ha sido desarrollada para productos y servicios ofrecidos en los Estados Unidos. Este material puede estar disponible en IBM en otros idiomas. Sin embargo, es posible que deba tener una copia del producto o de la versión del producto en ese idioma para poder acceder al mismo.

Avisos

Es posible que IBM no ofrezca en otros países los productos, servicios o características que se describen en este documento. Póngase en contacto con el representante local de IBM para obtener información acerca de los productos y servicios que actualmente están disponibles en su localidad. Cualquier referencia a un producto, programa o servicio de IBM no implica ni establece que sólo se pueda utilizar dicho 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 verificar el funcionamiento de cualquier producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o solicitudes de patente pendientes de aprobación que cubran temas tratados en este documento. La posesión de este documento no confiere ninguna licencia sobre dichas patentes. Puede enviar consultas sobre licencias, por escrito, a la siguiente dirección:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785 EE.UU.

Para realizar consultas relativas a la información de juego de caracteres de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe las consultas, por escrito, a:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokio 103-8510, Japón

El párrafo siguiente no se aplica en el Reino Unido ni en ningún otro país en el que las disposiciones en él expuestas 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 SIN LIMITARSE A ELLAS, LAS GARANTÍAS IMPLÍCITAS DE NO INFRACCIÓN, COMERCIALIZACIÓN E IDONEIDAD PARA UN PROPÓSITO DETERMINADO. Algunas legislaciones no contemplan la declaración de limitación de garantías, ni implícitas ni explícitas, en determinadas transacciones, por lo que cabe la posibilidad de que esta declaración no se aplique en su caso.

Esta información puede contener imprecisiones técnicas o errores tipográficos. La información incluida en este documento está sujeta a cambios periódicos, que se

incorporarán en nuevas ediciones de la publicación. IBM puede realizar en cualquier momento mejoras o cambios en los productos o programas descritos en esta publicación sin previo aviso.

Cualquier referencia hecha en esta información a sitios web que no sean de IBM se proporciona únicamente para su comodidad y no debe considerarse en modo alguno como una aprobación de dichos sitios web. Los materiales de estos sitios web no forman parte de los materiales de este producto de IBM y el uso que haga de estos sitios web es de la entera responsabilidad del usuario.

IBM puede utilizar o distribuir cualquier información que se le facilite de la manera que considere adecuada, sin contraer por ello ninguna obligación con el remitente.

Los licenciatarios de este programa que deseen obtener información acerca del mismo con el fin de: (i) intercambiar la información entre los programas creados independientemente y otros programas (incluido éste) y (ii) utilizar mutuamente la información que se ha intercambiado, deben ponerse en contacto con:

IBM Corporation
J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003 EE.UU.

Esta información puede estar disponible, sujeta a los términos y condiciones pertinentes, incluido en algunos casos el pago de una cantidad determinada.

IBM proporciona el programa bajo licencia descrito en este documento, y todo el material bajo licencia disponible para el mismo, bajo los términos del Acuerdo de cliente de IBM, el Acuerdo acuerdo internacional de licencia de programa de IBM o cualquier otro acuerdo equivalente entre las partes.

Los datos de rendimiento incluidos en este documento se determinaron en un entorno controlado. Por tanto, los resultados obtenidos en otros entornos operativos pueden variar significativamente. Algunas mediciones pueden haberse efectuado en sistemas a nivel de desarrollo, y no existe ninguna garantía de que dichas mediciones sean las mismas en sistemas de disponibilidad general. Además, es posible que algunas mediciones se hayan estimado mediante extrapolación. Los resultados reales pueden variar. Los usuarios de este documento deben verificar los datos aplicables a su entorno específico.

La información relacionada con productos no de IBM se ha obtenido de los suministradores de dichos productos, de sus anuncios publicados o de otras fuentes de información pública disponibles. IBM no ha probado dichos productos y no puede confirmar la precisión del rendimiento, la compatibilidad ni ninguna otra afirmación relacionada con productos que no son de IBM. Las consultas acerca de las prestaciones de los productos que no son de IBM deben dirigirse a los suministradores de tales productos.

Todas las declaraciones relativas a la dirección o intención futura de IBM están sujetas a cambios o anulación sin previo aviso y representan únicamente metas y objetivos.

Esta información se suministra sólo con fines de planificación. La presente información esta sujeta a cambios antes de que los productos que en ella se describen estén disponibles.

Esta información contiene ejemplos de datos e informes utilizados en las operaciones de negocios diarias. Para ilustrarlos de la forma más completa posible, los ejemplos incluyen nombres de personas, empresas, marcas y productos. Todos estos nombres son ficticios y cualquier similitud con los nombres y direcciones utilizados por una empresa real es totalmente casual.

LICENCIA DE COPYRIGHT:

Esta información contiene programas de aplicación de ejemplo en idioma de origen, que ilustra las técnicas de programación en diversas plataformas operativas. Puede copiar, modificar y distribuir estos programas de ejemplo de cualquier forma, sin pagar a IBM, con la finalidad de desarrollar, utilizar, comercializar o distribuir programas de aplicación que se ajusten a la interfaz de programación de aplicaciones para la plataforma operativa para la que se han escrito los programas de ejemplo. Estos ejemplos no se han probado bajo todas las condiciones posibles. Por lo tanto, IBM no puede garantizar ni dar por sentada la fiabilidad, capacidad de servicio o funcionamiento de esos programas. Los programas de ejemplo se suministran "TAL CUAL", sin garantía de ninguna clase. IBM no se hará responsable de los daños que puedan derivarse del uso de los programas de ejemplo.

Cada copia, parcial o completa, de estos programas de ejemplo o cualquier trabajo derivado, debe incluir un aviso de copyright como el siguiente:

© (el nombre de su empresa) (año). Partes de este código provienen de programas de ejemplo de IBM Corp. © Copyright IBM Corp. _escriba el año o años_. Reservados todos los derechos.

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

Consideraciones sobre la política de privacidad

Los productos de software de IBM, incluidas las soluciones de software como servicio, ("Ofertas de software"), pueden utilizar cookies u otras tecnologías para recopilar información sobre el uso de productos, para ayudar a mejorar la experiencia del usuario final, para personalizar las interacciones con el usuario final o para otros fines. En muchos casos, las Ofertas de software no recopilan información de identificación personal. Algunas de nuestras Ofertas de software pueden ayudarle a recopilar información de identificación personal. Si esta Oferta de software utiliza cookies para recopilar información de identificación personal, la información específica sobre el uso de cookies por parte de esta oferta se expone más abajo.

Dependiendo de las configuraciones desplegadas, esta Oferta de software puede utilizar cookies de sesión o persistentes. Si un producto o componente no está en la lista, ese producto o componente no utiliza cookies.

Tabla 2. Uso de cookies de los productos y componentes de InfoSphere Information Server

Módulo de producto	Componente o característica	Tipo de cookie que se utiliza	Recopilar estos datos	Finalidad de los datos	Inhabilitación de las cookies
Cualquiera (parte de la instalación de InfoSphere Information Server)	Consola web de InfoSphere Information Server	<ul style="list-style-type: none"> Sesión Persistente 	Nombre de usuario	<ul style="list-style-type: none"> Gestión de sesiones Autenticación 	No se pueden inhabilitar

Tabla 2. Uso de cookies de los productos y componentes de InfoSphere Information Server (continuación)

Módulo de producto	Componente o característica	Tipo de cookie que se utiliza	Recopilar estos datos	Finalidad de los datos	Inhabilitación de las cookies
Cualquiera (parte de la instalación de InfoSphere Information Server)	InfoSphere Metadata Asset Manager	<ul style="list-style-type: none"> Sesión Persistente 	Ninguna información de identificación personal	<ul style="list-style-type: none"> Gestión de sesiones Autenticación Usabilidad de usuario mejorada Configuración de inicio de sesión único 	No se pueden inhabilitar
InfoSphere DataStage	Etapas Big Data File	<ul style="list-style-type: none"> Sesión Persistente 	<ul style="list-style-type: none"> Nombre de usuario Firma digital ID de sesión 	<ul style="list-style-type: none"> Gestión de sesiones Autenticación Configuración de inicio de sesión único 	No se pueden inhabilitar
InfoSphere DataStage	Etapas XML	Sesión	Identificadores internos	<ul style="list-style-type: none"> Gestión de sesiones Autenticación 	No se pueden inhabilitar
InfoSphere DataStage	Consola de operaciones de IBM InfoSphere DataStage and QualityStage	Sesión	Ninguna información de identificación personal	<ul style="list-style-type: none"> Gestión de sesiones Autenticación 	No se pueden inhabilitar
InfoSphere Data Click	Consola web de InfoSphere Information Server	<ul style="list-style-type: none"> Sesión Persistente 	Nombre de usuario	<ul style="list-style-type: none"> Gestión de sesiones Autenticación 	No se pueden inhabilitar
InfoSphere Data Quality Console		Sesión	Ninguna información de identificación personal	<ul style="list-style-type: none"> Gestión de sesiones Autenticación Configuración de inicio de sesión único 	No se pueden inhabilitar
InfoSphere QualityStage Standardization Rules Designer	Consola web de InfoSphere Information Server	<ul style="list-style-type: none"> Sesión Persistente 	Nombre de usuario	<ul style="list-style-type: none"> Gestión de sesiones Autenticación 	No se pueden inhabilitar
InfoSphere Information Governance Catalog		<ul style="list-style-type: none"> Sesión Persistente 	<ul style="list-style-type: none"> Nombre de usuario Identificadores internos Estado del árbol 	<ul style="list-style-type: none"> Gestión de sesiones Autenticación Configuración de inicio de sesión único 	No se pueden inhabilitar
InfoSphere Information Analyzer	Etapas Reglas de datos en el cliente del Diseñador de InfoSphere DataStage and QualityStage	Sesión	ID de sesión	Gestión de sesiones	No se pueden inhabilitar

Si las configuraciones desplegadas para esta Oferta de software le ofrecen como cliente la posibilidad de recopilar información de identificación personal de los usuarios finales mediante cookies y otras tecnologías, debe buscar asesoramiento jurídico sobre la legislación aplicable a dicha recopilación de datos, incluidos los requisitos de notificación y consentimiento.

Para obtener más información sobre el uso de diversas tecnologías, incluidas las cookies, para estos fines, consulte la Política de privacidad de IBM en <http://www.ibm.com/privacy>, la sección “Cookies, balizas web y otras tecnologías” de la Declaración de privacidad en línea de IBM en <http://www.ibm.com/privacy/details> y la “Declaración de privacidad de productos de software y software como servicio de IBM” (en inglés) en <http://www.ibm.com/software/info/product-privacy>.

Marcas registradas

IBM, el logotipo de IBM e [ibm.com](http://www.ibm.com) son marcas comerciales o marcas registradas de International Business Machines Corp., registradas en muchas jurisdicciones en todo el mundo. Otros nombres de productos y servicios pueden ser marcas registradas de IBM u otras empresas. Encontrará una lista actual de las marcas registradas de IBM en el sitio web www.ibm.com/legal/copytrade.shtml.

Los términos siguientes son marcas comerciales o marcas registradas de otras empresas:

Adobe es una marca registrada de Adobe Systems Incorporated en los Estados Unidos y/o en otros países.

Intel e Itanium son marcas comerciales o marcas registradas de Intel Corporation o sus filiales en los Estados Unidos y otros países.

Linux es una marca registrada de Linus Torvalds en los Estados Unidos y/ en otros países.

Microsoft, Windows y Windows NT son marcas registradas de Microsoft Corporation en los Estados Unidos y/o en otros países.

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

Java[™] y todas las marcas registradas y logotipos basados en Java son marcas comerciales o marcas registradas de Oracle y/o sus filiales.

El Servicio de correos de Estados Unidos (United States Postal Service) es propietario de las siguientes marcas registradas: CASS, CASS Certified, DPV, LACS^{Link}, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS y United States Postal Service. IBM Corporation es un licenciataria no exclusivo de DPV y LACS^{Link} del Servicio de correos de Estados Unidos.

Otros nombres de empresas, productos y servicios pueden ser marcas comerciales o marcas de servicio de terceros.

Índice

A

accesibilidad de los productos
 accesibilidad 21
ámbito de depósito 16
aplicación de integración de datos 4
avisos legales 29

C

calidad de datos
 caso de ejemplo de depósito de
 datos 6
caso de ejemplo de depósito de datos
 calidad de datos 6
 entradas de caso de ejemplo clave 10
 metadatos compartidos en el glosario
 empresarial 17
 metadatos compartidos en la
 correlación y el desarrollo 17
 preguntas a plantear 9
 proceso de supervisión de la calidad
 de los datos 18

D

documentación del producto
 acceder 25

E

entradas de caso de ejemplo clave 10
 caso de ejemplo de depósito de
 datos 10

I

IBM Industry Models 8
impacto de los cambios sobre los
 depósitos 16
incorporación
 depósito de datos 7
integración de la supervisión y la calidad
 de los datos
 depósito de datos 7

M

marcas registradas
 lista de 29
metadatos compartidos en el glosario
 empresarial 17
 caso de ejemplo de depósito de
 datos 17
metadatos compartidos en la correlación
 y el desarrollo 17
 caso de ejemplo de depósito de
 datos 17

P

preguntas comunes
 caso de ejemplo de depósito de
 datos 9
proceso de depósito de datos
 caso de ejemplo de evaluación de
 calidad de datos 13
 proceso de descubrimiento de
 metadatos de origen 11
proceso de descubrimiento de metadatos
 de origen
 proceso de depósito de datos 11
proceso de evaluación de la calidad de
 los datos 13
 proceso de depósito de datos 13
proceso de supervisión de la calidad de
 los datos 18
 caso de ejemplo de depósito de
 datos 18
procesos de entrada de caso de ejemplo
 clave 11

S

servicios de software
 contactar 23
soporte
 cliente 23
soporte al cliente
 contactar 23

V

visión general de los casos de ejemplo de
 integración 1



Impreso en España

SC43-1200-00

