

# MDM y calidad de datos para el data warehouse

*Decisiones concisas rigurosas, en el momento adecuado e informes precisos gracias a datos de referencia fiables*

EXECUTIVE BRIEF



Este documento contiene información comercial secreta, confidencial y propiedad de Informatica Corporation ("Información confidencial") y no se puede copiar, distribuir, duplicar o reproducir en modo alguno sin el consentimiento previo por escrito de Informatica.

Aunque se ha intentado por todos los medios garantizar la precisión y compleción de la información aquí presentada, este documento puede contener errores tipográficos o imprecisiones técnicas. Informatica no asume la responsabilidad por ningún tipo de pérdida derivada del uso de la información incluida en este documento. La información incluida en este documento está sujeta a cambios sin previo aviso.

La incorporación de los atributos de los productos mencionados en este documento en cualquier versión o actualización de cualquier producto de software de Informatica, así como la fecha de dichas versiones o actualizaciones, quedan sujetas al criterio exclusivo de Informatica.

Protegido por una o más de las siguientes patentes de EE. UU.: 6,032,158; 5,794,246; 6,014,670; 6,339,775; 6,044,374; 6,208,990; 6,208,990; 6,850,947; 6,895,471; o por las siguientes patentes en trámite de EE. UU.: 09/644,280; 10/966,046; 10/727,700.

Edición publicada en septiembre de 2010

## Contenidos

Resumen ejecutivo . . . . .	2
Los resultados de su negocio (y de su business intelligence) dependen de la calidad de los datos . . . . .	3
Por qué sigue siendo necesario controlar los datos para poder controlar la business intelligence y el negocio . . . . .	4
El desafío técnico: la gestión de datos en un data warehouse . . . . .	5
MDM reduce las tareas de integración de datos . . . . .	6
El papel de la calidad de datos en la BI . . . . .	7
Resumen . . . . .	8

## Resumen ejecutivo

Muchas empresas han invertido en sofisticados sistemas de business intelligence (BI) para optimizar los procesos de negocio o cumplir los requisitos normativos de creación de informes, pero no logran alcanzar los resultados esperados. ¿Por qué? Estos sistemas no contienen datos críticos para el negocio que sean fiables sobre clientes, productos, socios de canal, proveedores y empleados, también llamados datos maestros o de referencia. Esto provoca que los directivos de empresas tomen decisiones poco acertadas. Los encargados de la conformidad normativa encuentran serias dificultades para cumplir con los requisitos normativos de creación de informes. Los analistas de negocio, que preparan informes para sus gerentes, dedican demasiado tiempo a la búsqueda y la conciliación manuales de la información pertinente. Además, los valiosos recursos de IT se malgastan en resolver problemas de calidad de datos. Todos estos problemas pueden atribuirse a la introducción de datos poco fiables en el data warehouse que sirve de base para la BI. En este informe se explica cómo la combinación de la gestión de datos maestros (MDM) y la calidad de datos (DQ) mejora considerablemente la precisión y la fiabilidad de los datos, lo que facilita la toma de decisiones fiables, en el momento adecuado, así como la elaboración de informes precisos.

## Los resultados de su negocio (y de su business intelligence) dependen de la calidad de los datos

Cuando los sistemas de BI y de elaboración de informes no proporcionan datos precisos, todo el negocio se ve afectado. Las consecuencias de utilizar datos poco fiables para alimentar los sistemas de BI afectan a la mayoría de los empleados. Estos son algunas:

- **Directivos:** Los informes de dirección imprecisos llevan a tomas de decisiones poco acertadas. Los malentendidos sobre cuáles son los 100 clientes principales en función de los ingresos y qué productos son los más rentables pueden suponer pérdidas de ingresos y beneficios.
- **Responsables del cumplimiento de normativas:** Las iniciativas de cumplimiento como la Ley Sarbanes-Oxley y Basilea II obligan a las empresas a ofrecer transparencia y auditabilidad en sus informes financieros y de conformidad. Los informes imprecisos exponen a la empresa a graves riesgos como elevadas sanciones pecuniarias y una imagen negativa en el mercado, sin mencionar el riesgo de condenas a prisión para los responsables de la empresa.
- **Analistas de negocio:** La productividad de los analistas de negocio se ve mermada cuando dedican excesivo tiempo a la búsqueda y la conciliación manuales de información en diversos sistemas de BI para actualizar y corregir informes. La falta de eficacia de esta tarea incide directamente en los costes y la rentabilidad.

Estos problemas empresariales se derivan de la falta de una versión única de la realidad por lo que respecta a los datos críticos para el negocio sobre clientes, productos, socios de canal y proveedores. Estos datos, también llamados datos maestros o de referencia, se recopilan, se alojan y se administran en los distintos sistemas que se ocupan de cada proceso de negocio. Eso implica que los datos de referencia que se solapan y son contradictorios deben converger correctamente en una versión única de la realidad para ofrecer una perspectiva útil y tangible. Muchas organizaciones cuentan con decenas, en ocasiones centenares, de aplicaciones distintas con los mismos objetos de referencia principal en decenas o centenares de bases de datos, con atributos que se solapan. Esta complejidad hace que resulte muy difícil para las empresas responder a preguntas como “¿Cuáles son los 100 clientes principales en función de los ingresos?” o “¿Cuántas de estas piezas hemos vendido en los dos últimos trimestres?”.

El objeto de la BI es elaborar informes sobre los datos existentes en diversos sistemas sin hacer distinciones. Puede realizar alguna agregación para el análisis dimensional, pero no está diseñada ni equipada para crear una versión única de la realidad. La incoherencia de los datos de clientes o productos procedentes de silos de aplicaciones puede tener consecuencias negativas sobre la fiabilidad de los análisis que se llevan a cabo en data warehouses.

## Por qué sigue siendo necesario controlar los datos para poder controlar la business intelligence y el negocio

La gestión de datos maestros es necesaria para reconocer datos distintos, hacerlos converger en una versión única de la realidad y relacionarlos con otros datos para extraer algún significado. De hecho, las fases iniciales de muchos proyectos de MDM se han centrado en crear datos más fiables para elaborar informes de BI más precisos.

MDM es el proceso controlado por el que los datos maestros se crean y mantienen como el sistema de registro de la empresa. MDM se implementa para garantizar que los datos maestros se validen como correctos, coherentes y completos, y puedan gestionarse de forma centralizada y dinámica en jerarquías. Posteriormente, estos datos maestros relacionados y fiables pasan del sistema MDM al data warehouse, en el que los analistas y los directivos de empresas pueden emplearlos para la BI y la elaboración de informes.

En el ejemplo siguiente, ilustrado en la figura 1, se destaca el valor de MDM para la BI. Los directivos de una empresa internacional de fabricación de sistemas de imagen necesitan saber cuáles son sus 100 clientes principales para resolver un problema del servicio de atención al cliente que surgió tras dejar de fabricar un producto. Para responder a esta solicitud, los analistas de negocio prepararon un informe de sus 100 clientes principales en función de los ingresos con ayuda del sistema de BI. Lamentablemente, no habían implementado un sistema MDM, y el data warehouse que sirve de base para el sistema de BI tenía dos entradas para uno de los clientes principales. General Electric compró numerosos sistemas de imagen en dos transacciones distintas: una transacción por valor de 400 millones de dólares figuraba en el sistema de contabilidad de Estados Unidos bajo el nombre de The General Electric Company; la segunda transacción, por valor de 300 millones de dólares, figuraba en el sistema de contabilidad europeo bajo el nombre de GE. El data warehouse no sabía que ambas eran la misma empresa, y contabilizó las dos transacciones por separado como pertenecientes a distintas empresas. Por lo tanto, The General Electric Company no aparecía como uno de los 100 clientes principales. La información de este informe es poco fiable y las decisiones que se hubieran tomado teniendo en cuenta sus conclusiones habrían sido erróneas.

Este fabricante logró resolver el problema, cuyas consecuencias habrían sido costosas, mediante la implementación de un sistema MDM que identificó correctamente que The General Electric Company y GE son la misma empresa. A continuación, introdujo esta información en el data warehouse, que entonces calculó correctamente el importe total de ambas transacciones (700 millones de dólares). Tras poner en marcha el sistema MDM, The General Electric Company apareció correctamente en el informe como uno de los 100 clientes principales en función de los ingresos. La información de este informe es fiable, y las decisiones de negocio que se tomen teniendo en cuenta sus conclusiones serán acertadas.

## Motivos que avalan la necesidad de MDM para la fiabilidad de los datos dentro de una empresa

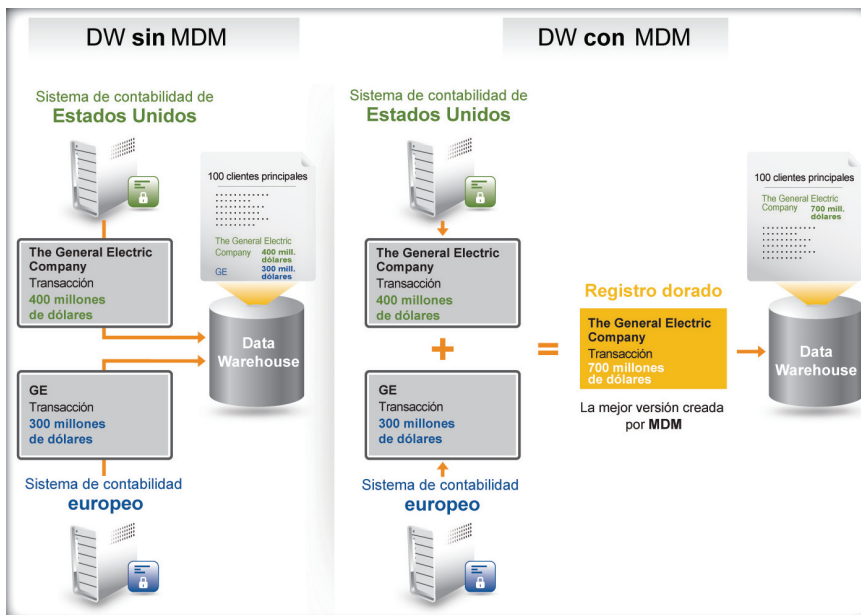


Figura 1. Dos escenarios de data warehousing: a la izquierda, el riesgo de elaborar informes imprecisos cuando no se utiliza un sistema MDM para crear datos maestros fiables antes de introducirlos en el data warehouse; a la derecha, cómo un sistema MDM facilita la elaboración de informes precisos resolviendo el problema de los datos antes de que los datos maestros se introduzcan en el data warehouse.

## El desafío técnico: la gestión de datos en un data warehouse

Para tomar decisiones rigurosas y en el momento adecuado, los directivos deben tener acceso a datos fiables, que pueden lograrse mediante la integración de MDM con un data warehouse. La principal tarea a la hora de crear un data warehouse consiste en la gestión de datos de referencia, lo que supone mucho trabajo para los desarrolladores de data warehouses, ya que deben codificar y mantener una compleja lógica para integrar datos de referencia, en numerosas ocasiones redundantes y a menudo incorrectos, procedentes de todo tipo de fuentes posibles.

Mantener esta lógica codificada de forma personalizada resulta lento y costoso. Añade una carga de trabajo considerable al proceso de integración de datos. Los datos duplicados deben consolidarse. Los conflictos deben evaluarse para determinar la representación de datos más fiable. Esta consolidación no sólo garantiza la fiabilidad de los datos de referencia, sino también la fiabilidad de las claves y dimensiones utilizadas para agregar datos transaccionales de forma precisa. Esta interdependencia, que es la piedra angular de la precisión de los datos en el data warehouse, elimina la información errónea y garantiza la obtención de un valioso conocimiento mediante los sistemas de BI.

## MDM reduce las tareas de integración de datos

MDM ofrece una solución sin código y de mantenimiento bajo para la integración y la gestión de datos de referencia. Cuando el sistema MDM se convierte en la fuente definitiva de datos dimensionales para el data warehouse, permite que éste se centre en los objetivos de gestión de volumen y entrega de datos.

En resumen, la gestión de datos maestros proporciona estas capacidades:

- Lógica de correspondencia y fusión para identificar y consolidar registros duplicados de uno o varios sistemas de origen
- Amplio historial y linaje a nivel de celda, que ofrecen una pista de auditoría detallada del contenido de los datos
- Un repositorio central para todos los datos de relaciones en todas las fuentes y aplicaciones

Estas capacidades reducen considerablemente las tareas generales de desarrollo y mantenimiento asociadas a los data warehouses, como se explica a continuación:

- **Dimensiones conformadas de MDM**

El primer paso a la hora de diseñar un data warehouse, ya se trate de un almacén de datos empresariales (EDW), un almacén de datos distribuido (DDW), o una arquitectura de bus de data mart, es adaptarse a las dimensiones del data warehouse. Ralph Kimball, fundador de Kimball Group, es un innovador y formador en el campo del data warehousing. Según viene documentado en varios artículos y en Wikipedia, Kimball define una dimensión como adaptada “cuando dos copias de las dimensiones son o bien exactamente iguales (incluidos los valores de las claves y todos los atributos), o bien una dimensión es un subconjunto perfecto de la otra”.

El sistema MDM se convierte en la única fuente de datos conformados, integrados, limpios y estandarizados para las dimensiones conformadas del data warehouse. Este sistema elimina gran parte del esfuerzo de transformación que supone alimentar estructuras de datos conformados, para que los recursos de los data warehouses se dediquen a responder preguntas de negocios en vez de a integrar datos.

- **Dimensiones lentamente cambiantes de MDM**

Las dimensiones lentamente cambiantes constituyen el método más eficaz y más utilizado para mantener un historial de cambios en las dimensiones. Los desarrolladores de data warehouses emiten un nuevo registro de dimensión para cada registro de dimensión que sufre un cambio en uno de sus atributos de segmentación de datos. La desventaja de este planteamiento es que las tablas de dimensiones crecen de forma continuada, lo cual ralentiza la carga y las consultas. El problema del crecimiento se agrava porque el enfoque de las dimensiones lentamente cambiantes se utiliza para todo tipo de cambios de un registro de dimensión, incluidos los cambios en los atributos como el nombre del cliente, el estado civil, o el nombre del director de la sucursal.

El sistema MDM resuelve este problema de rendimiento mediante una opción de seguimiento de historial. Se pueden registrar todos los cambios en una entidad de datos de referencia, liberando así al data warehouse para que se concentre en el seguimiento de los cambios de segmentación de datos como dimensiones lentamente cambiantes. Cuando las dimensiones del data warehouse crecen más lentamente, afectan menos a la velocidad de la carga y las consultas. Cuando los usuarios quieren consultar el historial de una dimensión que no cambia lentamente de un registro, pueden buscar en las tablas del historial del sistema MDM.

- **MDM y linaje de datos**

Las soluciones de metadatos y linaje de data warehouse suelen limitarse a ofrecer linaje estructural y de procesos. Por ejemplo, una implementación estándar de data warehouse debería poder responder a preguntas como “¿Qué campos de qué sistemas son fuentes para el campo Nombre del cliente en el data warehouse?”, “¿Qué secuencia de comandos de carga actualiza el campo Nombre del cliente en el data warehouse?”, e incluso “¿Qué reglas de negocio o transformaciones afectan al campo Nombre del cliente en el data warehouse?”.



Los data warehouses se quedan cortos a la hora de ofrecer un linaje adecuado del “contenido”. Por ejemplo, la mayoría de data warehouses no podrían responder la pregunta “¿Por qué ‘Fredd Jones’ es el valor del campo Nombre del cliente para este registro?”. Para responder este tipo de preguntas normalmente es necesario, por lo menos, hacer consultas en las tablas intermedias para todos los sistemas de origen posibles a fin de averiguar qué registro de origen facilitó dicho valor.

MDM elimina la necesidad de llevar ese seguimiento de datos lento y manual, ya que mantiene un linaje de datos detallado para todos los campos en todos los registros (es decir, se lleva un seguimiento del linaje de datos a nivel de celda). Para cada celda de datos en un objeto de base, el sistema MDM puede identificar qué sistema de origen facilitó dicho valor de celda, y concretamente qué registro del sistema de origen facilitó el valor, además de cuándo fue la última vez que se actualizó el valor. El sistema MDM también lleva un seguimiento del historial de todas las fusiones de registros.

## El papel de la calidad de datos en la BI

La integración de reglas y actividades de calidad de datos (perfilado, limpieza y supervisión) con procesos de MDM es crítica para mejorar la precisión y el valor de los activos de datos. Antes de emprender cualquier proyecto de MDM, debe comprender el contenido, la calidad y la estructura de los datos de origen. El perfilado de datos en origen permite a los administradores de datos y administradores de data warehouses detectar y analizar rápidamente todas las anomalías en todas las fuentes de datos antes de que éstos pasen al sistema MDM. Este proceso acelera considerablemente la creación de valor de la implementación de MDM.

Dado que la limpieza de datos mejora la precisión de los datos, ofrece integridad y promueve la fiabilidad de los datos en origen, también mejora la coherencia de los datos en el sistema MDM. Una vez que un dato de origen se incorpora al sistema MDM, se somete a un procesamiento de calidad de datos, incluidas la validación, la corrección y la estandarización. El sistema MDM almacena el historial completo de los datos antes y después los limpia, para que así los desarrolladores no necesiten llevar un seguimiento del linaje de los datos en el data warehouse.

Por último, las métricas de calidad de datos permiten a los administradores de data warehouses supervisar mejor la calidad de los datos de referencia y garantizar el uso continuo de datos de alta calidad a largo plazo.

Así, desde el punto de vista técnico, la implementación de un sistema MDM, junto con DQ, como la fuente definitiva de los datos maestros para un data warehouse simplifica la integración de datos del proceso de extracción, transformación y carga (ETL). Este enfoque reduce considerablemente el total de tareas de desarrollo y mantenimiento asociadas a los data warehouses. Mediante el establecimiento de métricas de calidad de datos y la definición de objetivos de calidad, los administradores de data warehouses y los administradores de datos pueden supervisar mejor la calidad de los datos de referencia y garantizar el uso continuo de datos de alta calidad en toda la empresa a largo plazo. MDM simplifica el manejo de las actualizaciones de las dimensiones del data warehouse porque toda la lógica para determinar qué ha cambiado está encapsulada en el sistema MDM.

Además, el sistema MDM puede eliminar gran parte de la carga que supone el seguimiento del historial del data warehouse, al dejar que éste gestione sólo aquellos cambios que necesita controlar para llevar a cabo la agregación. Las consecuencias de este sistema son la reducción de las dimensiones de data warehouse y una mejora considerable en la velocidad de la carga y las consultas. En última instancia, aprovechar el sistema MDM y DQ reduce las tareas de integración de datos y aumenta la calidad de los conocimientos obtenidos de la BI y los informes, de modo que el valor y el retorno de la inversión (ROI) en la iniciativa de data warehouse que sirve de base a la BI quedan garantizados.

## Resumen

Cuando los sistemas de BI no utilizan datos fiables, los directivos pierden confianza y toman decisiones poco acertadas, los responsables del cumplimiento de normativas se arriesgan a elaborar informes normativos imprecisos, y los analistas de negocio dedican demasiado tiempo a la búsqueda y la conciliación de los datos para actualizar y corregir informes. MDM y la calidad de datos mejoran considerablemente la precisión y la fiabilidad de los datos del data warehouse, lo que permite tomar decisiones rigurosas, en el momento adecuado y elaborar informes precisos.

Desde el punto de vista técnico, la implementación de un sistema MDM junto con la calidad de los datos como la fuente definitiva de datos dimensionales para un data warehouse simplifica la complejidad de la integración de datos del proceso de ETL. Esta implementación reduce considerablemente el total de tareas de desarrollo y mantenimiento asociadas a los data warehouses.

Con una buena calidad de datos, el conocimiento empresarial que se obtiene a través de BI se puede poner en práctica de forma más rápida. Gracias a este aumento de confianza en los datos, los ejecutivos, gerentes y empleados pueden reconocer los nuevos patrones y tendencias (oportunidades de crecimiento y posibles problemas a los que se debe prestar atención) y actuar en consecuencia. La gestión integral de la calidad de datos también aumenta la auditabilidad y la visibilidad en cuanto a la elaboración de informes de BI, algo que resulta especialmente valioso para la conformidad normativa y la gestión de riesgos.

## Acerca de Informatica MDM

Gracias a Informatica® MDM™, las empresas pueden mejorar sus operaciones al permitir el acceso de los usuarios de negocio a datos consolidados y fiables que son críticos para el negocio sobre clientes, productos, socios de canal, proveedores y empleados, además de a las relaciones entre los datos. El sistema de gestión de datos maestros multidominio de Informatica, flexible y probado, ofrece un amplio soporte a todos los requisitos del proyecto de MDM (integración, perfilado y calidad de datos y gestión de datos maestros) en la misma plataforma. Las empresas líderes eligen Informatica MDM porque les permite empezar de forma modesta con una rápida implementación para abordar cualquier problema urgente relacionado con MDM y, más adelante, ampliar el alcance de la solución a otros problemas de la empresa, reduciendo el tiempo de creación de valor y el TCO, por un lado, y mejorando el ROI, por otro. Visite [www.informatica.com/mdm](http://www.informatica.com/mdm).

## Más información

Obtenga más información sobre la plataforma de Informatica. Visítenos en [www.informatica.com/es](http://www.informatica.com/es) o llame al 902 882 062.

## Acerca de Informatica

Informatica Corporation (NASDAQ: INFA) es el número uno mundial de los proveedores independientes de software de integración de datos. Empresas de todo el mundo aumentan su ventaja competitiva en la actual economía global de la información con datos oportunos, pertinentes y fiables para cumplir sus imperativos de negocio principales. Más de 4.200 empresas de todo el mundo recurren a Informatica para consultar, integrar y poder confiar en los activos de información guardados en los sistemas tradicionales de la empresa, en sistemas externos y en cloud.



Informatica en España: Albasanz 15, 28037 Madrid · Constitución 1, 08960 Sant Just Desvern, Barcelona  
Teléfono: 902 882 062. Fax: 933 714 895. [www.informatica.com/es](http://www.informatica.com/es)

Sede central mundial: 100 Cardinal Way, Redwood City, CA 94063 (Estados Unidos)

Teléfono: + (1) 650.385.5000. Fax: 650.385.5500. Número gratuito en Estados Unidos: 1.800.653.9871 [www.informatica.com](http://www.informatica.com)

Sucursales de Informatica en el mundo: Alemania · Australia · Bélgica · Canadá · China · Corea · España · Estados Unidos · Francia · Irlanda · Japón · Países Bajos · Portugal · Reino Unido · Singapur · Suiza

© 2009 Informatica Corporation. Reservados todos los derechos. Impreso en EE.UU. Informatica, el logotipo de Informatica y The Data Integration Company son marcas comerciales o marcas registradas de Informatica Corporation en Estados Unidos y en las jurisdicciones de todo el mundo. Todos los demás nombres de compañías y productos pueden ser nombres comerciales o marcas comerciales de sus respectivos propietarios.

7209ES (20/09/10)