

Gestión de datos maestros (MDM)

El papel fundamental de la integración de datos en el marco de una estrategia de gestión de datos maestros (MDM)

¿QUÉ SON LOS DATOS MAESTROS?

Un dato maestro es un registro único que sirve de referencia para toda la empresa.

Por ejemplo, el nombre de un cliente, el código de un producto o un número de cuenta son datos de referencia.

DEFINICIÓN DE MASTER DATA MANAGEMENT (MDM):

MDM es un método de tratamiento de los datos, lanzado por la consultora Meta Group, para calificar y uniformizar el modo de describir la información a fin de garantizar su correcta integración. Este método abarca el conjunto de datos "de referencia" (o datos maestros) de un repositorio. Por tanto, dicho repositorio servirá de modelo a la hora de actualizar algún sistema o base de datos.

Esta gestión se basa en:

- La unificación de los datos en un repositorio único.
- La puesta al día de los datos gracias a la centralización de las actualizaciones.
- La fiabilidad y la limpieza de los datos
- La validación de los datos en conformidad con los derechos y las normas de seguridad de la organización.
- La disponibilidad de los datos y su divulgación por toda la organización.

El concepto de un repositorio que albergue todos los datos de referencia (o datos maestros) de la empresa no es nuevo. Tras años de informática distribuida, proliferación e imbricación de sistemas heterogéneos que dan soporte a las actividades de negocio y a la gestión de la organización, todos los responsables de los sistemas de información sueñan con poder disponer de una base centralizada y depositaria exclusiva de "la verdad". Un sueño que ha sido elevado al rango de necesidad urgente debido a las nuevas exigencias de respuesta en contextos de fusiones y adquisiciones, gestión de rendimiento, conformidad normativa, etc. La falta de gestión unificada de los datos de referencia se traduce en pérdidas cotidianas de eficacia operativa que ejercen un impacto directo sobre el rendimiento global de la organización.

Partiendo de la base de que es imposible pasar por alto la diversidad, la heterogeneidad y la complejidad de los sistemas, aplicaciones y procesos existentes, el establecimiento de una gestión eficaz de los datos de referencia plantea una serie de cuestiones difíciles de resolver sin el método y las herramientas adecuados: ¿Cómo extraer los datos de referencia del lugar en que se crean? ¿Cómo garantizar su calidad y homogeneidad cuando se trata de datos procedentes de sistemas distintos o de instancias diferentes de un mismo sistema? ¿Cómo actualizar los datos de referencia y, una vez modificados, divulgar los cambios por toda la organización?



Numerosos ámbitos de aplicación...

Una problemática de integración de datos

A la vista de las cuestiones mencionadas anteriormente, es fácil comprender por qué cualquier iniciativa de MDM también constituye (quizá, ante todo) una problemática de integración de datos. Si bien existen múltiples enfoques para gestionar los datos de referencia, todos ellos se basan en la construcción de un repositorio (único o por categoría de datos de referencia) y tienen, fundamentalmente, las mismas necesidades de accesibilidad, disponibilidad, calidad, coherencia, auditabilidad y seguridad de los datos.

¿Qué diferencia hay entre data warehousing y MDM?

Un data warehouse agrupa los datos procedentes de varias fuentes para alimentar aplicaciones de inteligencia empresarial, creación de informes y análisis. Si bien unifica los datos procedentes de los sistemas fuente, el data warehouse no ha sido concebido para devolver los datos modificados a estas fuentes. Es decir, un data warehouse emplea un proceso monodireccional, mientras que el MDM necesita un proceso bidireccional que garantice la sincronización de los datos entre el repositorio y los sistemas de origen y de destino asociados.

En uno de sus artículos*, el analista David Waddington de Ventana Research consideraba que "la plataforma de integración de datos de Informatica es especialmente adecuada para responder a estas necesidades", sobre todo en los contextos de despliegue de arquitecturas orientadas a servicios (SOA). En efecto, PowerCenter ofrece una plataforma unificada que proporciona servicios de datos compartidos, utilizables en múltiples procesos y en entornos de datos heterogéneos, así como funciones de desarrollo colaborativo que el autor califica como "elementos decisivos para dar soporte al MDM".

Las etapas del proceso

Una iniciativa completa de gestión y mantenimiento de datos maestros comprende las etapas siguientes:

- Designación de un responsable para cada categoría de datos maestros (clientes, productos, proveedores, estructuras organizativas, etc.). Este responsable se convierte en el garante de la calidad y de la actualización de los datos ante todos los sistemas, procesos y personas que utilizan este recurso compartido.
- Extracción a partir de los diversos sistemas operativos, transaccionales y analíticos de los datos de referencia de cada ámbito para cargarlos en repositorios por campo o en un *hub* central.
- Aplicación de las normas de calidad de datos para obtener un conjunto de datos limpios (sobre todo, mediante la deduplicación de registros).
- Definición de las reglas de reconciliación y racionalización de los datos maestros.

El objetivo es obtener en cada campo una lista/jerarquía óptima y comprensible para los usuarios, ya sean éstos individuos o aplicaciones.

- Sincronización de los sistemas operativos y de creación de informes con los datos gestionados en el/los repositorio(s) de datos maestros para garantizar que todos los sistemas utilizan en todo momento los datos adecuados (mismo valor y misma versión).
- Seguimiento de las actualizaciones y/o modificaciones de los datos maestros en los sistemas participantes y reconducción sistemática de las 5 etapas anteriores del proceso para mantener la exhaustividad, la calidad y la frescura de los datos del recurso compartido.

Características principales de una solución de MDM

Habida cuenta de las etapas descritas anteriormente, una solución de MDM (tanto si se trata de una plataforma genérica que incluya todas las categorías de datos de referencia como de una solución específica a una sola categoría) debe:

- Permitir la definición de las funciones y los derechos de acceso individuales para cada etapa del proceso de gestión de los datos maestros.
- Ofrecer funciones completas de ETL para extraer los datos de referencia de las diversas fuentes y cargarlos en el repositorio.
- Prestar servicios de limpieza de datos para comparar y deduplicar los registros.
- Ofrecer capacidades de colaboración para

coordinar las decisiones de reconciliación y de racionalización de los datos maestros.

- Ocuparse de la detección de los cambios, la sincronización bidireccional y la replicación de los datos para trasladar a los sistemas afectados cualquier cambio efectuado en el repositorio.
- Permitir el control de versión y la validación de los cambios en el conjunto del sistema. En especial, la solución debe permitir la recreación de un estado anterior de los datos y el modelado de estados futuros.

Recomendación

En el artículo citado*, Ventana Research recomienda a los clientes actuales de Informatica considerar PowerCenter como una extensión necesaria de sus esfuerzos de integración de datos para dar soporte a la gestión de datos maestros. Las organizaciones deben organizar su estrategia de MDM de manera que puedan incluir en ella la integración de datos y asegurarse de que cubre todas sus necesidades de migración, replicación y sincronización de los datos de referencia. Aquellas empresas que busquen una solución de integración de datos flexible deberían considerar a Informatica como la base de soporte para su arquitectura de información y como elemento de apoyo para la gestión de datos maestros.

ACERCA DE INFORMATICA

Informatica Corporation es un proveedor líder de soluciones de integración de datos empresariales. Gracias a los productos de Informatica, organizaciones de todo tipo pueden acceder, integrar, migrar y consolidar los datos generados y utilizados por el conjunto de sus sistemas, procesos y colaboradores, para reducir así la complejidad de la información, garantizar su coherencia y aumentar el rendimiento global de sus actividades.

* VentanaMonitor™: "Informatica addresses Master Data Management", David Waddington, Ventana Research, 18 de abril de 2006.

INFORMATICA
The Data Integration Company™

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

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

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

© 2008 Informatica Corporation. Todos los derechos reservados. Impreso en España. Informatica, el logotipo de Informatica y Data Explorer son marcas comerciales o marcas comerciales registradas de Informatica Corporation en los Estados Unidos y en jurisdicciones de todo el mundo. Todos los demás nombres de compañías y productos son nombres comerciales o marcas comerciales registradas de sus respectivos dueños.

(12/02/2008)