

# Gestión de datos maestros y migración de datos

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 la integridad 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 noviembre de 2014

## Contenidos

<b>Motivos de la relevancia del MDM en la migración de datos. . . . .</b>	<b>2</b>
<b>Problema n.º 1: comparación con términos equivalentes . . . . .</b>	<b>2</b>
<b>Lanzamiento a tiempo: paso intermedio. . . . .</b>	<b>3</b>
<b>Caso práctico: tropecientos productos. . . . .</b>	<b>4</b>
<b>Problema n.º 2: importancia de la calidad . . . . .</b>	<b>5</b>
<b>Caso práctico: unificación de sistemas corporativos y locales . . . . .</b>	<b>6</b>
<b>Migración de datos como pasarela al valioso MDM . . . . .</b>	<b>6</b>

En este documento, se describe la importancia cada vez mayor que adquiere la gestión de datos maestros (MDM) en los proyectos de migración de datos, su utilidad y las opciones de implantación más apropiadas, incluidos casos prácticos pertinentes.

## Motivos de la relevancia del MDM en la migración de datos

Todos los sistemas nuevos necesitan datos para arrancar la máquina de la actividad; la mayoría de ellos, una gran cantidad. En la actualidad, se migran a los sistemas nuevos datos procedentes de multitud de fuentes heredadas, aparte de enriquecerlos con fuentes externas como los archivos de direcciones postales. Con el conjunto de herramientas de migración de datos de Informática, se implanta una serie de tecnologías y procesos de mejores prácticas que permite resolver los diversos retos que surgen en los actuales ámbitos de migración.

No obstante, contemplemos primero la situación. Por muy integrador que sea el sistema de destino, cabe la posibilidad de que las fuentes de datos sean soluciones independientes cuyo diseño gira en torno a un proceso de negocio distinto con una visión diferente. Aunque el sistema de destino exija uniformidad, el entorno heredado suele ser incoherente tanto en estructura como en contenido. Eso plantea dos problemas, pero, en ambos casos, resulta adecuada la aplicación de técnicas de MDM.

## Problema n.º 1: comparación con términos equivalentes

Veamos como ejemplo la instalación de una aplicación nueva para planificar la producción. Ofrece garantías de mejorar la gestión de la planta e incrementar la eficiencia vinculando las aplicaciones de contabilidad y de recursos humanos. Sin embargo, cada uno de estos departamentos, y los sistemas que los ejecutan, tienen una visión diferente de exactamente los mismos elementos del mundo real. Para el contable, el mundo se resume en centros de costes, centros de beneficios, depreciación, activos de capital y activos operacionales. El ingeniero de producción, en cambio, define el mismo espacio físico conforme a estos atributos: procesos automatizados, semiautomatizados o manuales, flujos de trabajo, calendarios de mantenimiento y cuotas de producción. En cuanto al departamento de RR. HH., contempla ese escenario como una cuestión de personal interno, personal externo, requisitos de formación, grado de conocimientos, convenios salariales y seguridad e higiene en el trabajo.

Todos miran lo mismo, pero cada uno ve una cosa distinta.

Por eso, no resulta sorprendente que, a la hora de elegir y diseñar sus sistemas, los modelen de forma tan diferente. Ninguno está equivocado, sino que, por fuerza, no guardan coherencia. La consecuencia es que, cuando llega la hora de migrar los datos, nos encontramos con una comparación cuyos términos distan de ser equivalentes, sino todo lo contrario. Para que la operación sea correcta, hacen falta los mismos términos.

No vale con desechar una visión o la otra. En sentido estricto, ninguna de esas realidades distintas es errónea. Por otro lado, aunque se considere más apropiada la visión del departamento de producción (dado que, en este caso, es la que motiva el cambio), no cabe insistir en que se modifiquen los otros almacenes de datos heredados para ajustarlos al modelo de dicho departamento de producción sin recurrir a una reestructuración de escala semejante, en potencia, a la propia migración.

En todo caso, es más probable colegir que cada almacén de datos heredado es correcto en la medida en que modela su propio ámbito, es decir, el departamento de producción tiene razón desde el punto de vista de la producción; el departamento de RR. HH., desde el punto de vista de los recursos humanos; y así sucesivamente. Por consiguiente, hace falta un modelo capaz de conciliar todas las visiones. Resulta curioso que este reto sea idéntico al proyecto para proporcionar el sistema de destino. Entonces, ¿por qué no esperar a que esté listo y realizar con él el análisis de las carencias?

## Lanzamiento a tiempo: paso intermedio

Cuando el primer problema consiste en conciliar el panorama de datos fragmentados, conviene retroceder un segundo y reflexionar sobre el cronograma de una migración normal.

Suponiendo que se tarde un año en terminar el diseño y la instalación del sistema nuevo, el sistema de destino no puede estar disponible, como mínimo, hasta el octavo mes, y solo con un formato preliminar. (Y eso dando por sentado el empleo de una metodología de proyecto estándar que incluya las actividades preparativas del proyecto, el análisis de la situación actual, la configuración del sistema nuevo, la reestructuración de los procesos de negocio, etc.). Dada nuestra experiencia, cabría esperar la entrega de un sistema de destino estable en el décimo mes e, incluso así, se producirían cambios hasta la propia migración.

Según nuestra experiencia, apenas hay dos meses para realizar el análisis de las carencias, ejecutar las extracciones, las transformaciones y las cargas, escribir y probar las secuencias de carga, etc. Por desgracia, ese tiempo no basta para analizar con la máxima calidad las diferencias estructurales entre los almacenes de datos heredados, ni mucho menos para producir los elementos de datos compuestos imprescindibles en los archivos de carga.

Como solución a este problema de plazos, es posible crear un modelo intermedio (que denominamos prototipo de migración). Registramos las diferencias entre cada almacén de datos heredado y el prototipo y, a continuación, damos comienzo a las actividades de limpieza y preparación de datos. Luego, una vez entregado el sistema de destino definitivo, buscamos las diferencias entre el prototipo y el sistema de destino. Puesto que las transformaciones entre los almacenes de datos heredados y el prototipo son conocidas, podemos aislar las diferencias entre el prototipo y el sistema de destino final, lo cual simplifica en gran medida el trabajo en el momento de mayor actividad y presión.

Es una buena técnica. Primero, se crea un modelo de migración conjeturando el aspecto del sistema de destino, se analizan las diferencias entre los modelos heredado y migrado y se ejecutan las transformaciones, el enriquecimiento de datos, las mejoras, etc. Después, cuando aparezca el auténtico sistema de destino, se ajustan las transformaciones conforme a los problemas de granularidad que surjan en la última fase. De este modo, si el prototipo es correcto solo al 80% (y, por nuestra experiencia, lo es bastante más), significa que ya está diseñado el 80% de la lógica de transformación antes de que se echen encima las estresantes últimas semanas del proyecto.

Pero, ¿por qué sirve de ayuda una solución de gestión de datos maestros como parte de este modelo?

Según consta en la entrada en inglés de Wikipedia:

“El MDM tiene como objetivo facilitar procesos para la recopilación, la agregación, la comparación, la consolidación, la comprobación de la calidad, la conservación y la distribución de datos en toda una organización a fin de garantizar la coherencia y el control en el mantenimiento continuo y el uso por parte de las aplicaciones de dicha información”.

Desde el punto de vista de los datos, se crean instancias de las visiones de los tres ámbitos del ejemplo (contable, ingeniero de producción y profesional de RR. HH.) con valores codificados en los almacenes de datos pertinentes. Por eso, tenemos gran interés en facilitar un proceso para “la recopilación, la agregación, la comparación, la consolidación [y] la comprobación de la calidad”. En nuestro proyecto, no nos afectan directamente “la conservación y la distribución de los datos en toda una organización” por los motivos que ya hemos explicado. Sin embargo, “la coherencia y el control en el mantenimiento continuo y el uso por parte de las aplicaciones de dicha información” sí constituyen una prioridad principal del proyecto de migración. En definitiva, adoptamos el enfoque de MDM para la migración de datos, aunque es poco probable que cerremos el bucle con los sistemas que utilizan esos elementos de datos, ya que nuestro trabajo consiste en sustituir sistemas, no en mejorarlos.

## Caso práctico: tropecientos productos

Para dejarlo todo más claro, veamos un ejemplo de la vida real.

La migración de datos en cuestión se iba a llevar a cabo en una gran empresa de telecomunicaciones. Para hacerse una idea de la escala, contaban sus clientes por decenas de millones y las instalaciones por centenares de millones. Aparte de esas cifras, también había un número elevado de productos distintos, hasta cien mil. Por si aún faltara interés, las empresas de telecomunicaciones basan algunos productos en otros productos, por lo que nos enfrentábamos a una proliferación propia de diversas partes.

Otro reto inherente a las empresas de telecomunicaciones es el número de sistemas heredados que poseen. Debido a que la ordenación, el diseño, el suministro y la facturación abarcan varias capas de actividades físicas y lógicas, se han acostumbrado a crear soluciones de uso puntual para cada paso del camino. El resultado es que, de una posible cartera de más de 400 sistemas heredados, por lo general, gestionamos unos treinta por instalación.

Por supuesto, cada uno de estos legados poseía su propia visión del mundo —lógica, física o financiera— en relación con su posición en el flujo de procesos.

En el marco de este documento, solo nos fijaremos en el problema de la estructura de los productos y en la obligatoriedad del uso de una solución basada en MDM.

Con cien mil productos, algunos de ellos derivados de otros productos, el número de reglas era inmenso. Donde unas aplicaciones heredadas veían una simple instalación de telefonía, otro sistema veía un teléfono, el tono de marcación, la marcación de llamadas salientes, la restricción de llamadas entrantes, la devolución de llamadas, un contestador automático, etc.: en suma, si había hasta una docena de productos para un simple teléfono particular, ni se imagina para las redes complejas de voz y datos. Parecía como si cada sistema del proceso de suministro viera el teléfono desde una óptica diferente. Hacía falta coherencia, y cuanto antes. Era imposible esperar al sistema de destino.

La solución consistió en crear un hub de gestión de datos maestros para los productos y los derivados de productos. Con una actualización semanal, se cubrían las cambiantes necesidades del sector, puesto que se diseñaban y añadían sin parar productos nuevos. El hub nos permitió comprobar si coincidían productos en los almacenes de datos heredados, así como mantener la visión de las diferencias entre las representaciones de los distintos puntos de la cadena de entrega. A medida que se sucedían las fases de la migración, aparecían las diferencias entre las estructuras de los sistemas de origen y el sistema de destino, y podíamos volver a codificar los mappings y las transformaciones en consecuencia.

## Problema n.º 2: importancia de la calidad

A continuación, veamos la segunda clase de problema: el valor de los datos o del contenido. Al igual que los almacenes de datos heredados pueden tener varias estructuras para el mismo objeto de negocio, podemos tener varios valores. Veamos unos cuantos ejemplos habituales en este sentido.

La pesadilla de todos los departamentos de marketing son los duplicados que pueblan las listas de clientes extraídas de los diversos almacenes de datos que forman su interior. Algunos son fáciles de detectar y eliminar. Otros, en cambio, adolecen del temido problema de la homonimia o la sinonimia: ¿José García es el mismo usuario que J. García, que puede figurar como Jose García o incluso Pepe García en otra parte? Al realizar la migración, todos experimentamos los mismos problemas, que también se pueden producir dentro de un almacén de datos si existen varias copias de la misma persona.

A menudo, esos problemas se agravan por otros problemas de tipo estructural. Por ejemplo, en los entornos B2B, siempre existen dudas en cuanto a la identidad del cliente. ¿Se debe considerar como cliente la entidad jurídica principal? (O sea, la empresa que se personaría ante los tribunales si se llegara a ese extremo). ¿O es el cliente la rama comercial con la que se hacen negocios? ¿O bien la tienda o el almacén local? También en este caso, la visión varía según la posición en el negocio. El equipo de logística dirige su atención a los puntos de suministro. El equipo de facturación debe conocer tanto los puntos de suministro como los datos de las direcciones de facturación. Es posible que disponga de equipos de ventas que operan según la zona geográfica, pero que incluyan un grupo de gestores de relaciones estratégicas para grandes clientes internacionales. En suma, aunque exista una sola entidad jurídica, puede haber centenares de puntos locales de pedido, entrega y facturación.

De nuevo, quizá no sea factible corregir esas anomalías en los sistemas de origen, sobre todo, si estos no son incorrectos. (Por supuesto, existirán duplicados claros causados por procesos predeterminados en los que el mismo cliente se ha creado dos veces. Estos son incorrectos incluso en los almacenes de datos heredados, donde se pueden corregir).

Así pues, ¿por dónde conviene empezar y en qué ayuda una solución de MDM? Sobra decir que el MDM constituye la solución perfecta para dominar las entidades clave. Permite llevar a cabo casi todas las funciones de un MDM completo, entre otras, garantizar “la coherencia y el control en el mantenimiento continuo y el uso por parte de las aplicaciones”, al menos en lo concerniente a los duplicados que se deben descartar del conjunto de datos heredado.

Así, ya habrá quedado demostrada la utilidad de la tecnología de MDM y conocerá tanto la tecnología como los problemas semánticos asociados a la adopción del MDM.

## Acerca de Informatica

Informatica Corporation (Nasdaq: INFA) es el proveedor independiente líder mundial de software y servicios de integración de datos. Empresas de todo el mundo confían en Informatica para aprovechar su potencial de información y cumplir los principales imperativos de negocio. Informatica Vibe, la primera y única máquina virtual de datos (VDM) integrable del mercado, impulsa las capacidades exclusivas de “Map Once. Deploy Anywhere.” de la Plataforma de Informatica. Más de 5.500 empresas de todo el mundo dependen de Informatica para aprovechar al máximo sus activos de información, desde datos en dispositivos a móviles, datos sociales, Big Data en entornos locales, en el cloud y en redes sociales. Para obtener más información, llame al 902 882 062 o visite [www.informatica.com/es](http://www.informatica.com/es).

## Caso práctico: unificación de sistemas corporativos y locales

Este reto se aclara con otro ejemplo del mundo real. Una entidad bancaria mediana necesitaba una migración. Tenía oficinas dispersas por la región y, en cada una de ellas, existían instancias independientes de los sistemas corporativos, vinculadas entre sí y con la sede central por una serie de procesos de batch. Se planteaban ambos problemas: existían tanto variaciones locales en la interpretación de la estructura como duplicados en las mismas sucursales o en distintas sucursales. Con la ayuda de la tecnología de MDM, conseguimos establecer una visión única de la realidad y enlazarla a los sistemas de origen pertinentes. El objetivo consistía en reunir los datos pertinentes en una estructura coherente y eliminar los duplicados existentes entre las sucursales o dentro de ellas en el momento de la puesta en marcha. En este caso, el MDM siguió funcionando como visión única de la realidad de la empresa incluso después de haber terminado la migración.

## Migración de datos como pasarela al valioso MDM

La persistencia del MDM constituye uno de los beneficios fundamentales por los que servirse del evento obligatorio de una migración de datos de envergadura como trampolín para hacer realidad el MDM. Al preparar los datos para la migración, es preciso eliminar los duplicados y resolver los problemas semánticos que rodean las diferencias estructurales. Antes, solíamos crear una aplicación desechable que contuviese una visión única de la realidad mientras fuera necesaria para el proyecto. Ahora, con el uso de una solución de MDM apropiada para el sector, resulta mucho más lógico realizar este trabajo con el fin de completar la creación del MDM y su plena integración en el tejido de la arquitectura.



Informatica en España: José Echegaray 8, edif. 3, PB3, 28232 Las Rozas, 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) [linkedin.com/company/informatica](https://www.linkedin.com/company/informatica) [twitter.com/InformaticaCorp](https://twitter.com/InformaticaCorp)

© 2014 Informatica Corporation. Todos los derechos reservados. Informatica® y Put potential to work™ son marcas comerciales o marcas comerciales registradas de Informatica Corporation en Estados Unidos y en jurisdicciones de todo el mundo. Todos los demás nombres de compañías y productos pueden ser nombres comerciales o marcas comerciales.