

# Gerenciamento de Dados Mestres e Migração de Dados

Este documento contém informações confidenciais, exclusivas e segredos comerciais ("Informações confidenciais") da Informatica Corporation que não podem ser copiadas, distribuídas, duplicadas ou de outra forma reproduzidas, não importa como, sem o consentimento prévio por escrito da Informatica.

Apesar de todos os esforços terem sido realizados para assegurar que as informações deste documento sejam precisas e completas, alguns erros tipográficos ou imprecisões técnicas podem existir. A Informatica não aceita responsabilidade por nenhum tipo de perda resultante do uso das informações contidas neste documento. As informações contidas neste documento estão sujeitas a alterações sem aviso prévio.

A incorporação dos atributos de produto discutidos nesses materiais em qualquer versão ou atualização de qualquer produto de software da Informatica – bem como o momento de tal versão ou atualização – será a critério exclusivo da Informatica.

Protegido por uma ou mais das seguintes patentes norte-americanas: 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; ou pelas seguintes patentes pendentes norte-americanas: 09/644,280; 10/966,046; 10/727,700.

Edição publicada em novembro de 2014

## Índice

<b>Por que o MDM é fundamental para a migração de dados. . . . .</b>	<b>2</b>
<b>Questão no 1: Comparação de situações aparentemente semelhantes . . .</b>	<b>2</b>
<b>Lançamento no tempo certo: Uma etapa intermediária. . . . .</b>	<b>3</b>
<b>Estudo de caso: Um zilhão de produtos . . . . .</b>	<b>4</b>
<b>Questão no 2: Qualidade é fundamental . . . . .</b>	<b>4</b>
<b>Estudo de caso: A união de sistemas corporativos e locais . . . . .</b>	<b>5</b>
<b>Migração de dados como gateway para o MDM de alto valor . . . . .</b>	<b>6</b>

Este documento descreve a crescente importância do Gerenciamento de Dados Mestres em projetos de migração de dados, sua utilidade e as melhores opções de implantação, incluindo estudos de casos relevantes.

## Por que o MDM é fundamental para a migração de dados

Cada sistema novo precisa de dados para incrementar as atividades. A maioria dos sistemas requer uma quantidade absurda de dados. Atualmente, migramos dados de uma infinidade de fontes legadas para novos sistemas, além de abastecer esses sistemas a partir de fontes externas, como arquivos de endereçamento postal. O kit de ferramentas de migração de dados da Informatica implanta um conjunto de tecnologias e processos de práticas recomendadas para solucionar toda uma gama de desafios resultantes dos cenários atuais de migração de dados.

Mas antes vamos definir o cenário. Embora nosso sistema de destino seja bem abrangente, nossas fontes de dados acabam sendo apenas soluções específicas, cada uma projetada em torno de um processo de negócios diferente, com uma visão de mundo distinta. Ainda que o sistema de destino exija consistência, o ambiente legado geralmente apresenta inconsistências na estrutura e no conteúdo. Há duas questões neste contexto que precisamos resolver, e a utilização de técnicas de MDM é apropriada aos dois casos.

## Questão no 1: Comparação de situações aparentemente semelhantes

Vamos usar como exemplo a instalação de um novo aplicativo de planejamento de produção. Ele promete melhorar o gerenciamento do chão de fábrica e aumentar a eficiência vinculando os nossos aplicativos de contabilidade e recursos humanos. Mas observamos também que cada um desses departamentos e os sistemas responsáveis por sua execução têm uma visão diferente dos mesmos elementos da vida real. O universo de um contador é composto de centros de custos, centros de lucro, depreciação, ativos de capital e ativos operacionais. Já o engenheiro de produção define o mesmo espaço físico com as seguintes características: processos automatizados, semiautomatizados e manuais, fluxos de trabalho, programas de manutenção e cotas de produção. E o departamento de recursos humanos enxerga a mesma cena com outros componentes: quadro de funcionários interno, quadro de funcionários externo, requisitos de treinamento, níveis de habilidade, acordos salariais e questões de assistência médica.

Eles estão vendo os mesmos elementos, mas de ângulos diferentes.

Consequentemente, não é de se espantar que, ao escolher e planejar seus sistemas, façam isso de forma bem distinta. Isso não significa que algum deles está errado; no entanto, há claramente uma inconsistência. Quando chegar a hora de fazer a migração de dados, o resultado disso é a constatação de que, na verdade, não estamos comparando elementos semelhantes. Estamos sim comparando elementos bem distintos. Para que tudo dê certo, queremos manter apenas um tipo de elemento.

Mas também não podemos simplesmente descartar uma visão ou a outra. Nenhuma dessas várias verdades está "errada" no sentido absoluto. Por outro lado, mesmo decidindo que a visão do departamento de produção é a mais apropriada (considerando que esse é nosso fator determinante para a mudança nesse caso), não podemos insistir na modificação dos outros repositórios de dados legados para a adequação ao modelo do departamento de produção, sem que eles sejam reestruturados numa escala potencialmente igual à própria migração.

Seja qual for o caso, o mais provável é adotarmos uma visão que estabeleça que cada repositório de dados legados está certo na medida em que ele está reproduzindo o seu próprio domínio. Ou seja, o departamento de produção está certo conforme uma perspectiva da produção, o RH conforme a perspectiva do próprio RH, e assim por diante. Portanto, precisamos de um modelo que acomode todas as visões. Curiosamente, esse desafio é exatamente o mesmo que envolve o projeto de apresentar um sistema de destino. Por que não esperar o destino ficar pronto e executar nossa análise de lacunas desse sistema?

## Lançamento no tempo certo: Uma etapa intermediária

Se a primeira questão envolve a reconciliação de um cenário de dados fragmentado, vamos retroceder por um momento e pensar sobre o cronograma de uma migração normal.

Vamos supor que o planejamento e a instalação de um sistema novo levem um ano. Sendo assim, o sistema de destino não estará disponível pelo menos até o oitavo mês e, depois disso, ainda estará apenas no formato inicial. (E isso pressupõe uma metodologia de projeto padrão, que inclui atividades de start-up de projeto, análise do estado atual, nova configuração do sistema, reestruturação do processo de negócios etc.) Por conta da nossa experiência, a nossa expectativa é que a apresentação de um sistema de destino estável seja feita no décimo mês e, mesmo assim, ainda haveria a necessidade imediata de fazer alterações.

Com base na nossa experiência, provavelmente teríamos pouco mais de dois meses para realizar nossa análise de lacunas, desenvolver nossas extrações, transformações e carregamentos, criar e testar os nossos scripts de carregamento etc. Infelizmente, se isso não é tempo suficiente para executar uma análise de alta qualidade das diferenças estruturais entre os nossos repositórios de dados legados, imagine se é suficiente para produzir os itens de dados compostos que precisamos ter em nossos arquivos de carregamento.

Uma solução para esta questão de timing é criar um modelo intermediário (que chamamos de protótipo de migração). Nós registramos as diferenças entre cada repositório de dados legados e o protótipo, e começamos nossas atividades de limpeza e preparação de dados. Depois, quando o sistema de destino é finalmente apresentado, analisamos as diferenças entre o protótipo e o destino. Como já sabemos quais são as transformações entre os repositórios de dados legados e o protótipo, podemos isolar as diferenças entre o nosso protótipo e o destino final, o que simplifica bastante a nossa atividade na fase mais movimentada e sob pressão possível.

Eis uma dica bem legal: elabore um modelo de criação que é o melhor palpite de como o sistema de destino se parecerá, analise as diferenças entre o modelo legado e o de migração, e faça as transformações, aprimoramento dos dados, melhorias etc. Depois, quando o destino real se materializar, nós ajustamos essas transformações de acordo com os problemas de granularidade mais sutis que surgem no final. Dessa forma, se conseguirmos um protótipo com apenas 80% de acertos (e a experiência nos diz que teremos um resultado bem melhor do que esse), já teremos planejado 80% da nossa lógica de transformação antes de atingirmos o período mais difícil das últimas semanas do projeto.

Mas por que uma solução de Gerenciamento de Dados Mestres ajudará como parte desse modelo?

Segundo a Wikipédia

**"O MDM tem o objetivo de fornecer processos para coletar, agregar, corresponder, consolidar, garantir a qualidade, persistir e distribuir esses dados em uma organização para assegurar a consistência e o controle na manutenção contínua e na utilização dessas informações"**

Sob uma perspectiva dos dados, as três visões de domínio do nosso exemplo (o contador, o engenheiro de produção e o profissional de RH) serão representadas em valores codificados dentro dos repositórios de dados relevantes. Portanto, temos muito interesse em fornecer um processo para "reunir, agregar, corresponder, consolidar, [e] garantir a qualidade". Em benefício do nosso projeto, não estamos diretamente interessados em "persistir e distribuir esses dados em uma organização" pelos motivos já descritos. Mas assegurar a "consistência e o controle na manutenção contínua e na aplicação dessas informações" é uma questão primordial para o nosso projeto de migração. Em outras palavras, adotamos uma abordagem MDM em migrações de dados, a menos que – já que o nosso trabalho é substituir e não melhorar os sistemas – não seja possível finalizar o processo de volta aos sistemas que usam esses itens de dados.

## Estudo de caso: Um zilhão de produtos

Mas vamos ver um exemplo real para tornar tudo isso mais claro.

A migração de dados deste caso estava acontecendo em uma grande empresa de telecomunicações. Para se ter uma ideia do escopo, eles numeravam seus clientes em dezenas de milhões e o número de instalações em centenas de milhões. Além da quantidade envolvida, o número de produtos distintos também era grande, pelo menos centenas de milhares de produtos individuais. E para tornar tudo muito mais interessante, as empresas de telecomunicações criam produtos em cima de produtos. Ou seja, tínhamos que lidar com a versão da explosão de peças.

As empresas de telecomunicações também são complexas por conta de seus inúmeros de sistemas legados. Como fazer o pedido, planejar, fornecer e cobrar são ações que abrangem várias camadas de atividades físicas e lógicas, há uma tradição de criar soluções pontuais para cada etapa do caminho. O resultado é que, entre um pool potencial de mais de 400 sistemas legados, normalmente estamos lidando com cerca de 30 em cada instalação.

Cada um desses legados obviamente tem a sua visão de mundo relacionada com sua posição no fluxo de processos e com o fato da visão ser lógica, física ou financeira.

Para fins deste documento, só analisaremos o problema da estrutura de produtos e por que uma solução com base em MDM é fundamental.

Com centenas de milhares de produtos, alguns desenvolvidos a partir de outros, havia um número excessivo de regras. Onde um aplicativo legado enxergava uma instalação única de telefone, outro sistema enxergava o fone, um tom de discagem, discagem de chamada externa, bloqueio de chamada interna, retornar chamada, atender telefone etc. Em outras palavras, se havia pelo menos uma dúzia de produtos para um simples telefone doméstico, imagine redes complexas de voz e dados. Parecia que cada sistema no processo de provisionamento enxergava o telefone de maneira diferente. Precisávamos de consistência e precisávamos trabalhar nisso o quanto antes. Esperar pelo sistema de destino simplesmente não era uma opção.

A solução era criar um hub de Gerenciamento de Dados Mestres para produtos e para criações de produtos. Uma renovação semanal abordou os requisitos de uma indústria muito dinâmica, onde produtos novos são desenvolvidos e melhorados o tempo todo. O hub permitiu que verificássemos os repositórios de dados legados para produtos correspondentes, enquanto observávamos as diferenças entre as representações em pontos distintos da cadeia de abastecimento. À medida que cada fase da migração ocorria, sabíamos as diferenças entre as estruturas dos nossos sistemas de origem e de destino, e podíamos recodificar os nossos mapeamentos e transformações de forma apropriada.

## Questão no 2: Qualidade é fundamental

Agora vamos examinar a segunda questão: o problema do conteúdo ou do valor dos dados. Da mesma forma que os nossos repositórios podem ter várias estruturas para o mesmo objeto de negócios, também podemos ter vários valores. Vamos ilustrar essa situação com alguns exemplos bem comuns.

As duplicidades encontradas nas listas de clientes obtidas a partir de vários repositórios de dados que se encontram em áreas remotas são um pesadelo para todos os departamentos de marketing. Algumas serão fáceis de identificar e eliminar. Outras serão enredadas no terrível problema de homônimos ou sinônimos. Será que John Smith é o mesmo que J Smith, que pode ser J P Smith ou até mesmo Jonnie Smith em outro lugar? Quando fazemos uma migração, enfrentamos os mesmos problemas. Tudo isso é muito provável de acontecer dentro do mesmo repositório de dados, onde podem existir várias cópias da mesma pessoa.

Esses problemas geralmente são compostos por questões estruturais. Por exemplo, em ambientes Business-to-Business (B2B), geralmente a confusão envolve o cliente e quem ele é de fato. Será que deveríamos considerar o cliente como a pessoa jurídica principal? (Ou seja, a empresa que enfrentaríamos nos tribunais, se chegássemos a esse ponto.) Ou seria o cliente o ponto de apoio nas negociações? Ou será um depósito local ou ainda uma loja? Mais uma vez, sua visão mudará dependendo de qual for sua posição na empresa. A equipe de logística analisará os pontos de fornecimento; a equipe de faturamento precisará saber quais são os pontos de fornecimento e os detalhes do endereço de cobrança. Talvez você tenha equipes de vendas que atuam geograficamente, mas com uma camada de gerentes com relacionamentos estratégicos no caso de clientes maiores. Portanto, legalmente, pode haver apenas uma pessoa jurídica, mas podem existir centenas de pontos de pedido, entrega e cobrança.

Mais uma vez, pode não ser viável corrigir essas irregularidades nos sistemas de origem, especialmente se esses sistemas não estiverem errados. (Obviamente, haverá duplicidades evidentes provocadas por processos falhos, nos quais o mesmo cliente foi criado duas vezes. Esses estão errados mesmo nos termos dos repositórios de dados legados e podem ser corrigidos dentro deles.)

Então, por onde começar e como uma solução MDM pode ajudar? Desnecessário dizer que o MDM é a solução perfeita para dominar nossas principais entidades. Aqui, executaremos praticamente todas as funções de um MDM completo, incluindo a garantia de "consistência e controle na manutenção e aplicação contínuas", ao menos no que se refere às duplicidades que não deveriam estar em nosso conjunto de dados legados.

## Sobre a Informatica

A Informatica Corporation (Nasdaq: INFA) é a fornecedora independente líder mundial de software de integração de dados. Empresas do mundo todo contam com a Informatica para maximizar seu potencial de informações e impulsionar seus principais imperativos de negócios. A Informatica Vibe, primeira e única máquina de dados virtual (VDM) incorporável, aciona os recursos exclusivos do "Map Once. Deploy Anywhere." da plataforma Informatica. Em todo o mundo, mais de 5.500 empresas usam a tecnologia da Informatica para potencializar seus ativos de informação, desde dispositivos móveis e dados sociais até grandes volumes de dados hospedados na empresa, na nuvem e nas redes sociais. Para obter mais informações, ligue para +55 11 3013 5446 (São Paulo, Brasil) ou visite [www.informatica.com](http://www.informatica.com).

## Estudo de caso: A união de sistemas corporativos e locais

Outro exemplo do mundo real ilustrará o desafio. Um banco de médio porte precisava de uma migração. Suas filiais estavam regionalmente espalhadas e, dentro de cada filial, havia uma representação discreta dos sistemas corporativos, vinculados por uma série de processos em lote até a matriz. Temos dois conjuntos de problemas. Havia variações locais em torno de interpretações estruturais e havia duplicidades por todas as agências. Com a ajuda da tecnologia MDM, poderíamos criar uma visão única da verdade e estabelecer um vínculo de volta aos sistemas de origem relevantes. Meta: trazer os dados relevantes para uma estrutura consistente e eliminar as duplicidades entre as agências e dentro delas em tempo real. Nesse caso, o MDM continuou atuando como a única visão da verdade da empresa após a migração ter sido concluída.

## Migração de dados como gateway para o MDM de alto valor

A persistência do MDM é um dos principais benefícios de usar um evento interessante de uma migração de dados significativa como o trampolim para tornar seu MDM real. Durante a preparação dos dados para migração, precisaremos remover as duplicidades e solucionar as questões semânticas em torno das diferenças estruturais. Antigamente, costumávamos criar um aplicativo descartável para acomodar uma visão única da verdade enquanto o projeto precisasse dela. Se usarmos uma solução MDM contra falhas, a lógica de levar esse trabalho adiante para concluir a criação de um MDM e sua integração completa à estrutura de sua arquitetura será muito mais convincente. Inclusive, você já terá mostrado o valor da tecnologia MDM. E você já conhecerá a tecnologia e as questões semânticas associadas que estão relacionadas à adoção do MDM.



Sede mundial, 100 Cardinal Way, Redwood City, CA 94063, EUA Fone: 650.385.5000 Fax: 650.385.5500  
Ligação gratuita nos EUA: 1.800.653.3871 [informatica.com](http://informatica.com) [linkedin.com/company/informatica](https://www.linkedin.com/company/informatica) [twitter.com/InformaticaCorp](https://twitter.com/InformaticaCorp)

© 2014 Informatica Corporation. Todos os direitos reservados. Informatica® e Put potential to work™ são marcas comerciais ou marcas comerciais registradas da Informatica Corporation nos Estados Unidos e em diferentes jurisdições por todo o mundo. Todos os outros nomes de empresas e produtos podem ser nomes comerciais ou marcas comerciais.