Gerência de Dados Mestres e de Referência

Em qualquer organização, certos dados são comuns entre diferentes áreas de negócio, processos e sistemas. O compartilhamento de dados comuns (ex: lista de servidores; lista de alunos; lista de cursos; centros de custo; códigos de localização geográfica; etc) dentre as unidades de negócio é algo que beneficia tanto a organização quanto os seus clientes, visto que minimiza os riscos de inconsistências. Usuários de dados geralmente assumem a existência de um certo nível de consistência, até que se deparam com divergências entre fontes distintas (DAMA-DMBOK, 2017).

Na maioria das organizações, os sistemas evoluem de forma mais orgânica do que os profissionais de gerenciamento de dados gostariam. Particularmente em grandes organizações, vários projetos e iniciativas, fusões e aquisições e outras atividades de negócios resultam em vários sistemas executando essencialmente as mesmas funções, isolados uns dos outros. Essas condições inevitavelmente levam a inconsistências na estrutura de dados e valores de dados entre sistemas. Essa variabilidade aumenta os custos e os riscos, que podem ser reduzidos através da Gerência de Dados Mestres e Dados de Referência (DAMA-DMBOK, 2017).

Dados de Referência

Dados de Referência, ou Dados Referenciais, são qualquer dado utilizado para caracterizar ou classificar outro dado, ou para relacionar dados com informações externas à organização. Os Dados de Referência mais básicos consistem em códigos e descrições, mas alguns podem ser mais complexos e incorporar mapeamentos e hierarquias. Dados de referência existem em praticamente todos os armazenamentos de dados. Classificações e categorias podem incluir status ou tipos (por exemplo, Status do pedido: Novo, Em andamento, Fechado, Cancelado) (DAMA-DMBOK, 2017).

De acordo com Barbieri (2019), Dados Referenciais são atributos, normalmente associados aos Dados Mestres e que merecem pela sua volatilidade uma certa gerência especial. Por exemplo: CEP (atributo de endereço de alguém ou de alguma coisa), Código Internacional de Doenças (atributo fundamental do Dado Mestre Doenças em um ambiente de sistemas de saúde, por exemplo). São normalmente obtidos de fontes externas definidas por entidades oficiais (CID, CEP, código de aeroportos, códigos de cidades, de estados, de países, etc.), mas podem ser produzidos internamente, de acordo com o negócio da empresa/organização. Têm forte associação com os Dados Mestres, na maioria das vezes, codificando algumas de suas propriedades.

Dados de Referência e Dados Mestre compartilham propósitos conceitualmente semelhantes. Ambos fornecem contexto para a criação e uso de dados transacionais (Dados de Referência também fornecem contexto para Dados Mestres). Eles permitem que os dados sejam compreendidos de forma significativa.

O objetivo da Gerência de Dados de Referência (*Reference Data Management* - RDM) é garantir que os Dados de Referência sejam consistentes e atuais em diferentes funções e que os dados sejam acessíveis à toda a organização.

Dados Mestres

Segundo a definição do Gartner (2022), Dados Mestres pode ser definido como um conjunto consistente e uniforme de identificadores e atributos que descrevem as principais entidades da organização, como por exemplo: alunos, cursos, colaboradores, estrutura administrativa, fornecedores, hierarquias, planos de conta, etc.

De acordo com Barbieri (2019) Dados Mestres são os dados base ou pilares da instituição. Os Dados Mestres tendem a ser mais estáveis e não muito relacionados com o tempo e sustentam as grandes transações institucionais. São chamados dados de fundação (*foundational*) e através deles são produzidos os dados transacionais. Por exemplo, um cliente do WalMart compra produtos em uma loja. Veja que há três dados mestres (produto, loja e cliente) se relacionando em um ato de Compra, que é um dado transacional.

Dados Mestres exigem a identificação e/ou desenvolvimento de uma versão confiável da verdade para cada instância de entidade conceitual, como aluno, curso, unidade organizacional, pessoa ou organização, e a manutenção da validade dessa versão. O principal desafio é a resolução de entidade, o processo de discernir e gerenciar associações entre dados de diferentes sistemas e processos.

A Gerência de Dados Mestres (*Master Data Management* - MDM) reduz os riscos de tomada de decisão incorreta e perda de oportunidades por meio de uma representação consistente das entidades críticas para o negócio da organização (DAMA-DMBOK, 2017).

Gerência de Dados Mestres

A Gerência de Dados Mestres (*Master Data Management* - MDM) envolve o controle sobre valores e identificadores de Dados Mestres de modo a permitir o uso consistente entre os sistemas, dos dados mais precisos e atualizados sobre as entidades essenciais para o negócio (DAMA-DMBOK, 2017).

De acordo com o Gartner (2022), Gerência de Dados Mestres é uma disciplina na qual a área de negócios e a área de TI trabalham juntas para garantir uniformidade, precisão, curadoria (*stewardship*), consistência semântica e responsabilidade (*accountability*) dos ativos de Dados Mestres compartilhados da organização.

DAMA (2017) menciona que, infelizmente, o acrônimo MDM é muitas vezes referenciado como sistemas ou produtos utilizados para gerenciar Dados Mestres. Embora existam aplicações que facilitem esta gerência, elas não garantem que os Dados Mestres serão gerenciados de modo a atender as necessidades organizacionais.

A avaliação dos requisitos de MDM de uma organização inclui identificar DAMA-DMBOK (2017) :

- Quais funções, organizações, lugares e coisas são referenciadas repetidamente;
- Quais dados são usados para descrever pessoas, organizações, lugares e coisas;
- Como os dados são definidos e estruturados, incluindo a sua granularidade;
- Onde os dados são criados/originados, armazenados, disponibilizados e acessados;

- Como os dados mudam à medida que se movem pelos sistemas dentro da organização;
- Quem usa os dados e para que finalidades;
- Quais critérios são usados para entender a qualidade e confiabilidade dos dados e suas fontes.

DAMA-DMBOK (2017) complementa que a Gerência de Dados Mestres é desafiadora, e ilustra um desafio fundamental: “as pessoas escolhem maneiras diferentes de representar conceitos semelhantes e a reconciliação entre essas representações nem sempre é direta; tão importante quanto, as informações mudam ao longo do tempo e contabilizar sistematicamente essas mudanças requer planejamento, conhecimento sobre os dados e habilidades técnicas. Resumindo, é muito trabalhoso.“

Uma organização que percebe a necessidade da Gerência de Dados Mestres provavelmente já possui um arcabouço complexo de sistemas, com múltiplas formas de captura e armazenamento de dados que representam entidades do mundo real. Devido ao crescimento orgânico ao longo do tempo, ou fusões e aquisições, os sistemas que forneceram entrada para o processo de MDM podem ter definições diferentes das próprias entidades e muito provavelmente possuem padrões diferentes sobre a definição de qualidade de dados. Devido a essa complexidade, é recomendado abordar a Gerência de Dados Mestres um domínio de negócios por vez. Ou seja, comece com algumas entidades e atributos e evolua com o tempo, de modo incremental.

Dentre as atividades críticas para o sucesso da Gerência de Dados Mestres mencionadas por DAMA-DMBOK (2017), destaco:

- Reconciliar e consolidar dados entre fontes para fornecer um registro mestre ou a melhor versão da verdade.
- Provisionamento de acesso a dados confiáveis entre os sistemas, seja por meio de leituras diretas, serviços de dados, *Data Warehouses* e outros meios de armazenamento analítico.
- Impor o uso de Dados Mestres dentro da organização. Esse processo requer governança e gerenciamento de mudanças para garantir uma perspectiva corporativa compartilhada.

A Figura 1 apresenta as principais etapas de processamento necessárias para a Gerência de Dados Mestres (MDM). Inclui as etapas de gerência do modelo de dados; aquisição de dados; validação padronização e enriquecimento de dados; resolução de entidades; e administração e compartilhamento de dados. Em um ambiente abrangente de MDM, o modelo de dados lógicos será instanciado fisicamente em várias plataformas. Este modelo orienta a implementação da solução de MDM, fornecendo a base para os serviços de integração.



Figura 1: Processos-chave de etapas para a Gerência de Dados Mestres. Fonte: DAMA-DMBOK (2017)

Note que as etapas da Figura 1 são condizentes com as etapas que ocorrem em uma solução de *Data Warehouse* / *Business Intelligence*, da modelagem ao compartilhamento e visualização de dados.

Sistema de Registro e Sistema de Referência

Quando existem versões potencialmente diferentes da “verdade”, é necessário distingui-las. Para isso, é preciso saber de onde os dados se originam ou são acessados e quais dados foram preparados para usos específicos. Um Sistema de Registro é um sistema onde os dados são criados/capturados e/ou mantidos por meio de um conjunto definido de regras (por exemplo, um sistema ERP pode ser o Sistema de Registro para clientes de venda). Um Sistema de Referência é um sistema onde os consumidores de dados podem obter dados confiáveis para apoiar transações e análises, mesmo que a informação não tenha origem no sistema de referência. Sistemas MDM, Data Hubs e Data Warehouses geralmente servem como sistemas de referência (DAMA-DMBOK, 2017).

Fonte confiável

Uma fonte confiável (*Trusted Source*) pode ser definida como uma fonte com “a versão mais precisa da verdade”, baseada na combinação de regras automatizadas e gerência manual (curadoria) do conteúdo dos dados. Todo sistema de MDM/RDM deve ser gerido para ser esta fonte confiável institucional (DAMA-DMBOK, 2017).

Revisão #20
Criado 2022-03-18 11:33:18 UTC por FLAVIO LOPES DE MORAIS
Atualizado: 2024-10-08 11:17:48 UTC por FLAVIO LOPES DE MORAIS